



**HAL**  
open science

# Sparse Spikes Super-resolution on Thin Grids II: the Continuous Basis Pursuit

Vincent Duval, Gabriel Peyré

► **To cite this version:**

Vincent Duval, Gabriel Peyré. Sparse Spikes Super-resolution on Thin Grids II: the Continuous Basis Pursuit. *Inverse Problems*, 2017, 33 (9), 10.1088/1361-6420/aa7fce . hal-01389956v2

**HAL Id: hal-01389956**

**<https://hal.science/hal-01389956v2>**

Submitted on 1 Jun 2017

**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 Super-resolution on Thin Grids II: the Continuous Basis Pursuit

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

<sup>1</sup> INRIA, MOKAPLAN and CEREMADE, Université Paris-Dauphine

E-mail: `vincent.duval@inria.fr`

<sup>2</sup> CNRS and DMA, École Normale Supérieure

E-mail: `gabriel.peyre@ens.fr`

**Abstract.** This article analyzes the performance of the Continuous Basis Pursuit (C-BP) method for sparse super-resolution. The C-BP has been recently proposed by Ekanadham, Tranchina and Simoncelli as a refined discretization scheme for the recovery of spikes in inverse problems regularization. One of the most well known discretization scheme, the Basis Pursuit (BP, also known as LASSO) makes use of a finite dimensional  $\ell^1$  norm on a grid. In contrast, the C-BP uses a linear interpolation of the spikes positions to enable the recovery of spikes between grid points. When the sought-after solution is constrained to be positive, a remarkable feature of this approach is that it retains the convexity of the initial  $\ell^1$  problem. The present paper shows how the C-BP is able to recover the spikes locations with sub-grid accuracy in the favorable case. We also prove that this regime generally breaks when the grid is too thin, and we describe precisely the artifacts that appear: each spike is approximated by a pair of Dirac masses. We show numerical illustrations of these phenomena, and evaluate numerically the validity of the technical assumptions of our analysis.

## 1. Introduction

This article studies a specific convex optimization approach, the Continuous Basis-Pursuit (C-BP) for sparse super-resolution. A detailed review of the literature on super-resolution, and in particular variational regularization technics, can be found in the companion paper [12], which is dedicated to the Basis Pursuit (BP, or LASSO) method.

In the following, we focus for simplicity on the compact 1-D domain  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$  (i.e. an interval with periodic boundary conditions), but the algorithms considered (LASSO and C-BP) can be extended to higher dimensional settings (see Section 1.6).

### 1.1. Super-resolution

We formalize the sparse super-resolution problem mathematically as the question of recovering an unknown Radon measure  $m_0 \in \mathcal{M}(\mathbb{T})$  defined on the torus  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$  from noisy linear observations in a separable Hilbert space  $\mathcal{H}$ , which we write as

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

where  $w \in \mathcal{H}$  is some measurement noise. The terminology “super-resolution” refers here to the fact that  $\Phi : \mathcal{M}(\mathbb{T}) \rightarrow \mathcal{H}$  is a bounded linear map defined through the integration against a *continuous* kernel function  $\varphi : \mathbb{T} \rightarrow \mathcal{H}$

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

The continuity of  $\varphi$  is crucial, since it allows one to pose the problem over the space of measures, and thus enables the possible recovery of highly localized solutions (e.g. sums of Dirac masses). In order to avoid unnecessary technicalities, we only consider smooth impulse responses (typically  $\varphi \in \mathcal{C}^5(\mathbb{T}, \mathcal{H})$ ).

A typical class of such problems are the so-called deconvolution problems, when the operator  $\Phi$  is translation invariant. This corresponds to using  $\mathcal{H} = L^2(\mathbb{T})$ ,  $\varphi(x) : x' \mapsto \tilde{\varphi}(x' - x)$  for some smooth kernel  $\tilde{\varphi}$  defined on  $\mathbb{T}$ . This deconvolution setup is equivalent to computing the measurements over the Fourier domain. When the highest measured frequency  $f_c$  is finite, this corresponds to using  $\mathcal{H} = \mathbb{C}^{2f_c+1}$  and  $\varphi(x) = (\hat{\varphi}_k e^{2ikx\pi})_{k=-f_c}^{f_c}$  for some weights  $\hat{\varphi}_k \in \mathbb{C}$  which are the Fourier coefficients of  $\tilde{\varphi}$ .

### 1.2. Sparse Regularization

Since  $\Phi$  is usually ill-posed, a regularization scheme is needed to perform its inversion. Following several recent works (see for instance [4, 9, 6]), we study here sparse regularizations induced by the total variation of the measure (i.e. the total absolute mass)

$$|m|(\mathbb{T}) \stackrel{\text{def.}}{=} \sup \left\{ \int_{\mathbb{T}} \psi(x) dm(x) ; \psi \in \mathcal{C}(\mathbb{T}), \|\psi\|_{\infty} \leq 1 \right\}. \quad (3)$$

For discrete measures defined on a finite grid

$$\mathcal{G} \stackrel{\text{def.}}{=} \{z_i ; i \in \llbracket 0, G-1 \rrbracket\} \subset \mathbb{T} \quad (4)$$

of points, which take the form

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

the total variation is equal to the usual discrete  $\ell^1$  norm

$$|m_{a,\mathcal{G}}|(\mathbb{T}) = \|a\|_1 = \sum_i |a_i|,$$

which is at the heart of the celebrated Basis Pursuit [7] method (also known as the LASSO [22]). The total variation is thus the natural way to extend these intrinsically discrete and finite-dimensional methods to the infinite-dimensional class of Radon measures, to enable a grid-free regularization.

This total variation is used as a regularizer to define the Beurling LASSO (BLASSO) [9]

$$\min_{m \in \mathcal{M}(\mathbb{T})} \frac{1}{2} \|y - \Phi(m)\|_{\mathcal{H}}^2 + \lambda |m|(\mathbb{T}). \quad (5)$$

which is an infinite dimensional convex optimization problem. In the noiseless setting,  $w = 0$ , taking the limit  $\lambda \rightarrow 0$  leads to the following constrained problem

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

### 1.3. From BP to C-BP

The simplest way to compute approximate solutions to (3) is to restrict the solution to live in the finite dimensional sub-space of measures supported on the grid  $\mathcal{G}$  defined in (4). This is exactly equivalent to computing the weights  $a$  of the measure by solving the BP problem

$$\min_{a \in \mathbb{R}^G} \frac{1}{2} \|y - \Phi_{\mathcal{G}} a\|_{\mathcal{H}}^2 + \lambda \|a\|_1 \quad (7)$$

where we defined the restriction of  $\Phi$  to the discretization grid as

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

To obtain a better approximation of the infinite dimensional problem, [14] proposes to perform a first order approximation of the kernel,

$$\sum_{i=0}^{G-1} a_i \varphi(z_i + s_i) \approx \sum_{i=0}^{G-1} a_i \varphi(z_i) + \sum_{i=0}^{G-1} a_i s_i \varphi'(z_i). \quad (9)$$

We introduce the derivative operator restricted to the sampling grid

$$\Phi'_{\mathcal{G}} b \stackrel{\text{def.}}{=} \Phi'(m_{b,\mathcal{G}}) = \sum_{i=0}^{G-1} b_i \varphi'(z_i), \quad (10)$$

and to ease the exposition, we consider a uniform grid  $\mathcal{G} \stackrel{\text{def.}}{=} \{i/G ; i \in \llbracket 0, G-1 \rrbracket\}$  of  $G$  points, so that the grid size is  $h \stackrel{\text{def.}}{=} 1/G$ . The C-BP method of [14] assumes that the unknown measure is nonnegative, and, after a change of variables, solves

$$\min_{(a,b) \in \mathbb{R}^G \times \mathbb{R}^G} \frac{1}{2} \|y - \Phi_{\mathcal{G}} a - \Phi'_{\mathcal{G}} b\|_{\mathcal{H}}^2 + \lambda \|a\|_1 \quad \text{subject to} \quad |b| \leq \frac{h}{2} a, \quad (11)$$

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 nonnegative measures. This is a convex optimization problem, which can be solved using traditional conic optimization methods. As detailed in Section 2.2, this problem can also be re-cast as a LASSO in dimension  $2G$  with positivity constraints (see Section 2.2).

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 (11), one recovers an output discrete measure defined by

$$m^* = \sum_{a_i^* \neq 0} a_i^* \delta_{x_i^*} \quad \text{where} \quad x_i^* \stackrel{\text{def.}}{=} ih + \frac{b_i^*}{a_i^*}, \quad (12)$$

where we set  $\frac{b_i^*}{a_i^*} = 0$  whenever  $a_i^* = 0$ . The rationale behind (11) 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 [14] 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. Previous Works

There is a large body of literature on the study of the recovery performance of the LASSO problem (7). We refer to the companion paper [12] for a detailed overview. Let us however insist on the fact that only a small part of these previous works tackles the super-resolution case where the operator  $\Phi$  is highly “correlated”, which can be informally understood as cases associated to a continuous underlying kernel  $\varphi$ . As highlighted above, these cases are naturally paired with a continuous infinite dimensional problem (5). The theoretical analysis of this setting is recent, and we refer in particular to the works [6] which showed that some minimal separation between the spikes (the so-called “Rayleigh limit”) is necessary and sufficient to achieve exact recovery in the absence of noise, for deconvolution problems. Several related works have studied the impact of noise on this recovery, see for instance [4, 5, 15, 1]. In two recent papers, including the initial work [11] and the companion paper [12], we have shown how to transfer support stability properties of the infinite dimensional BLASSO (5) to the discretized LASSO problem (7). Most notably, these results show that solutions of the discrete LASSO problem estimate in general twice the number of spikes of the input measure. It is important to note that in the special case where  $m_0$  is a nonnegative measure (which is the setup considered by the C-BP method) and  $\Phi$  is a convolution with a typical kernel such as Dirichlet or Gauss,  $m_0$  is always a solution to (6) (see for instance [9, 20] and [10] for a refined analysis of the stability to noise in this special case).

#### 1.5. Contributions and organization of the paper

The main purpose of this paper is to evaluate the behavior of the C-BP and its reconstruction artifacts, in the same spirit as the results in the companion paper [12] for the basic LASSO.

First of all, Section 2 shows that the C-BP problem (11) can be recast as a LASSO under positivity constraints. It proposes an abstract analysis of support

stability properties of such a class of problems, under a non-degeneracy condition (hypothesis (34)), see Theorem 1. The framework is general, and the results hold when  $\Phi_{\mathcal{G}}, \Phi'_{\mathcal{G}}$  are replaced with any linear operator  $A$  and  $B$ .

In Section 3, we turn to the C-BP problem as a method for solving continuous inverse problems (*e.g.* deconvolution) with the help of a grid. First, we highlight a regime in which the C-BP recovers the spikes locations with sub-grid accuracy. This is the ideal situation. But we notice that this regime generally breaks if the grid is too thin. The rest of the paper is devoted to describe the situation in that less favorable setting. We show the  $\Gamma$ -convergence of Problem (11), as the grid gets thinner and thinner, towards the BLASSO (5) with positivity constraints: the solutions of (11) converge in the weak\* topology towards solutions of (5).

Section 4 is devoted to a fine analysis of the support of the solutions using thin grids. First, for fixed  $\lambda > 0$  and under a mild assumption, we prove that each spike of the solution to (5) is approximated with one or two consecutive spikes, and those spikes are “blocked” at half-grid points. It is remarkable that, although the situation is *a priori* considerably richer than with the LASSO (see Table 1), the C-BP yields a similar behavior in practice.

One may object that those artifacts are due to too large values of  $\lambda$  or the noise level. We study the low noise regime ( $\lambda$  and  $\|w\|_{\mathcal{H}}$  small, *i.e.* of order  $O(h^3)$ ) associated with those thin grids, and we show that, under some non-degeneracy condition (the “Twice Non-Degenerate Source Condition”, see Definition 5), the recovered spikes still appear in pairs (see Theorem 3), one of them being “free”

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

and we provide a closed form expression for  $\varepsilon$ , which depends on some corresponding *natural shift* intrinsic to the measure. The corresponding low noise regime is characterized by  $\frac{\|w\|_{\mathcal{H}}}{\lambda} = O(1)$  and  $\lambda = O(h^3)$ , which should be compared with the one for the LASSO:  $\frac{\|w\|_{\mathcal{H}}}{\lambda} = O(1)$  and  $\lambda = O(h)$ . Let us mention that this result holds for unknown spikes that are not necessarily on the grid, but in its neighborhood (*i.e.* at distance at most  $O(h^{3/2})$  from some point of the grid), that is a situation which is *a priori* favorable to C-BP (as the Taylor expansion (9) is a good approximation).

Let us also stress the fact that the results of Section 4 do not follow from the study in the companion paper [12], as they crucially depend on the structure of the operator ( $\Phi_{\mathcal{G}}, \Phi'_{\mathcal{G}}$ ) and the cone constraint. As a result, the conditions for the low noise regime and the quantities involved in the natural shift are different from the LASSO. Moreover, those results crucially depend on the structure of the C-BP (two operators, the latter being the derivative of the former).

Eventually, we illustrate in Section 5 these theoretical results with numerical experiments comparing LASSO and C-BP. We first highlight in Section 5.1 the improvements brought by C-BP over LASSO for spikes that are isolated and that fall off-the-grid. Then, we explore the validity of the Twice Non-Degenerate Source Condition,

and we highlight the doubling phenomenon. We display the evolution of the solution path  $\lambda \mapsto a_\lambda$  (a solution of (7)) and  $\lambda \mapsto (a_\lambda, b_\lambda)$  (a solution of (11)). These paths are piecewise-affine, and our contributions (Theorem 3) precisely characterize the first affine segment of these paths, which perfectly matches the numerical observations.

To summarize, this paper shows how the C-BP is able to reach sub-grid accuracy in favorable situation, outperforming the LASSO. It also proves, however, that there is a regime where artifacts appear. A natural question is whether those artifacts appear on the typical grids used by practitioners. Our numerical experiments suggest that their appearance depends not only on the grid size, but also on the complexity of the measure to recover (are the spikes well separated or densely clustered?). For simple input measures (*i.e.* not densely clustered), the phenomenon only appears on *very* thin grids, so that the practitioner usually obtains the correct number of spikes, with a fairly precise estimation of the locations. However, if the measure has dense clusters, the phenomenon is likely to appear even on grids that are not so thin, preventing a faithful estimation of the support.

The limitation of the LASSO and C-BP are due to the polyhedral nature of the finite-dimensional approximation functional (related to the  $\ell^1$  norm). Improving support recovery requires to use more advanced approximation, for instance SDP relaxation as introduced in [9, 6, 11]. These methods are truly off-the-grid, and thus do not perform a polyhedral approximation of the continuous problem. They suffer however from other difficulties (large computational cost, difficult analysis of support recovery in more than 1 dimension, ...).

### 1.6. Extensions

While we restrict here the exposition to 1-D problems, the C-BP formulation (11) 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

$$\Phi'_G b \stackrel{\text{def.}}{=} \sum_{i \in \mathcal{G}} \sum_{k=1}^d b_{i,k} \partial_k \varphi(x_i) \in \mathcal{H}$$

where  $\partial^k$  denote the differential operator with respect to the  $k^{\text{th}}$  direction in  $\mathbb{R}^d$ . The optimality conditions stated in this paper should readily extend to this setting, but the fine analysis of the support seems nontrivial to adapt.

The paper [14] 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 [13].

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 (12) 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 [16], and shows promising practical performance for spectrum estimation.

### 1.7. Notations and preliminaries

The set of Radon measures (resp. nonnegative Radon measures) is denoted by  $\mathcal{M}(\mathbb{T})$  (resp.  $\mathcal{M}^+(\mathbb{T})$ ). Endowed with the total variation norm (3),  $\mathcal{M}(\mathbb{T})$  is a Banach space. Another useful topology on  $\mathcal{M}(\mathbb{T})$  is the weak\* topology: a sequence of measures  $(m_n)_{n \in \mathbb{N}}$  weakly\* converges towards  $m \in \mathcal{M}(\mathbb{T})$  if and only if for all  $\psi \in \mathcal{C}(\mathbb{T})$ ,  $\lim_{n \rightarrow +\infty} \int_{\mathbb{T}} \psi dm_n = \int_{\mathbb{T}} \psi dm$ . 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  $m_{a, \mathcal{G}}$ ), and that the  $x_{0,\nu}$ 's are pairwise distinct.

The space  $\mathcal{H}$  denotes a separable Hilbert space. As explained in the companion paper [12, Lemma 1], the operator  $\Phi : \mathcal{M}(\mathbb{T}) \rightarrow \mathcal{H}$  defined in (2) is weak-\* to weak continuous, and so are the operators involving the derivatives, for  $k \in \mathbb{N}$ ,

$$\forall m \in \mathcal{M}(\mathbb{T}), \quad \Phi^{(k)}(m) = \int_{\mathbb{T}} \varphi^{(k)}(x) dm(x).$$

Their respective adjoint operators,  $\Phi^{(k),*} : \mathcal{H} \rightarrow \mathcal{C}(\mathbb{T})$ , are given by  $(\Phi^{(k),*}q)(t) = \langle q, \varphi^{(k)}(t) \rangle$  for all  $q \in \mathcal{H}$ ,  $t \in \mathbb{T}$ . Moreover,  $\Phi^{(k),*}q$  has the regularity of  $\varphi$  and  $\frac{d^k}{dt^k}(\Phi^{*}q)(t) = (\Phi^{(k),*}q)(t)$ . The  $\infty, \mathcal{H}$ -operator norm of  $\Phi^* : \mathcal{H} \rightarrow \mathcal{C}(\mathbb{T})$  is defined as  $\|\Phi^*\|_{\infty, \mathcal{H}} \stackrel{\text{def.}}{=} \sup \{ \|\Phi^*w\|_{\infty} ; w \in \mathcal{H}, \|w\|_{\mathcal{H}} \leq 1 \}$  (and the  $\infty, \mathcal{H}$  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  $\mathbb{R}^N \rightarrow \mathcal{H}$ , with

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

It may also be seen as the restriction of  $\Phi$  to measures supported on the set  $\{x_{0,\nu} ; \nu \in \llbracket 1, N \rrbracket\}$ . A similar notation is adopted for  $\Phi'_{x_0}$ , resp.  $\Phi_{x_0}^{(k)}$  (replacing  $\varphi(x_{0,\nu})$  with  $\varphi'(x_{0,\nu})$ , resp.  $\varphi^{(k)}(x_{0,\nu})$ ). 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 rely on the notion of set convergence. Given a sequence  $(C_n)_{n \in \mathbb{N}}$  of subsets of



$\mathbb{T}$ , we define

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

$$\liminf_{n \rightarrow +\infty} C_n = \left\{ x \in \mathbb{T} ; \limsup_{n \rightarrow +\infty} d(x, C_n) = 0 \right\} \quad (14)$$

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

$$\lim_{n \rightarrow +\infty} C_n = C. \quad (15)$$

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 [19] for more detail about set convergence. We shall also use this notion in Hilbert spaces, with obvious adaptations.

## 2. 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 3, since this variant turns out to be the Continuous Basis-Pursuit when the degradation operator is the integration of an impulse response and its derivative.

### 2.1. Notations

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

$$\mathcal{C}_h \stackrel{\text{def.}}{=} \left\{ (c, d) \in \mathbb{R} \times \mathbb{R} ; c \geq 0 \quad \text{and} \quad -c\frac{h}{2} + |d| \leq 0 \right\}. \quad (16)$$

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

Now, given a vector  $(a_0, b_0) \in \mathcal{C}_h^G$  (i.e.  $\forall k \in \llbracket 0, G-1 \rrbracket, a_{0,k} \geq \frac{2}{h}|b_{0,k}|$ ), we observe  $y_0 \stackrel{\text{def.}}{=} Aa_0 + Bb_0$ , where  $A : \mathbb{R}^G \rightarrow \mathcal{H}$  and  $B : \mathbb{R}^G \rightarrow \mathcal{H}$  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:

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

and for  $\lambda = 0$ ,

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

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

$$I \stackrel{\text{def.}}{=} \text{supp}(a) \stackrel{\text{def.}}{=} \{i \in \llbracket 0, G-1 \rrbracket ; a_i > 0\} \quad (17)$$

$$= I^{(\mathbf{r})} \cup I^{(1)} \quad (18)$$

$$\text{where } I^{(\mathbf{r})} \stackrel{\text{def.}}{=} \left\{ i \in I ; a_i + \frac{2}{h}b_i > 0 \right\}, \quad I^{(1)} \stackrel{\text{def.}}{=} \left\{ i \in I ; a_i - \frac{2}{h}b_i > 0 \right\}. \quad (19)$$

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 1.* The notation  $I^{(\mathbf{r})}$ ,  $I^{(1)}$  shall become clearer in the next section. 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  $(ih - \frac{h}{2}, ih + \frac{h}{2})$  (see Figure 1).

## 2.2. 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, writing for all  $i \in \llbracket 0, G-1 \rrbracket$ ,

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \stackrel{\text{def.}}{=} \begin{pmatrix} 1 & 1 \\ \frac{h}{2} & -\frac{h}{2} \end{pmatrix} \begin{pmatrix} r_i \\ l_i \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} r_i \\ l_i \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{h} \\ \frac{1}{2} & -\frac{1}{h} \end{pmatrix} \begin{pmatrix} a_i \\ b_i \end{pmatrix} \quad (20)$$

In the following, we define the linear map

$$H_h : \begin{pmatrix} r \\ l \end{pmatrix} \mapsto \begin{pmatrix} a \\ b \end{pmatrix} \quad (21)$$

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

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

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

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

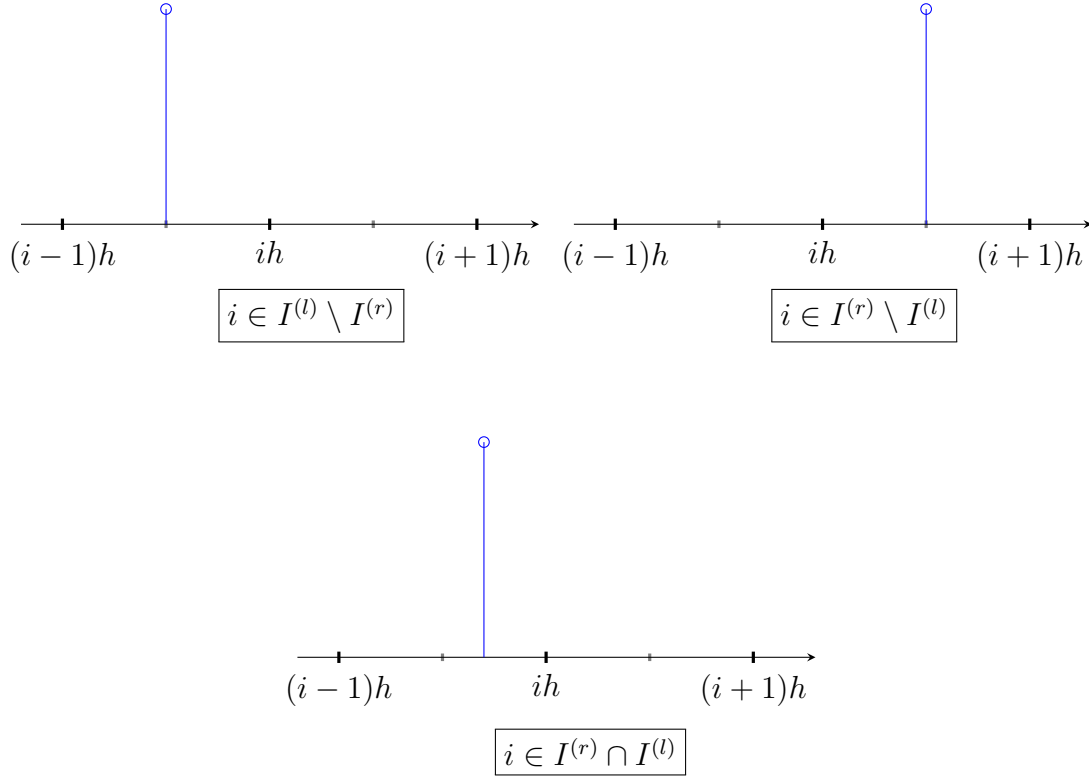


Figure 1: Depending on  $i$  being in  $I^{(r)}$  or  $I^{(l)}$ , the spikes of the solutions to (11) are supported on half-grid points or in the interval  $((i - 1/2)h, (i + 1/2)h)$ .

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

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

The “support recovery” of  $(a_0, b_0)$  through  $(\mathcal{Q}_\lambda(y))$  is equivalent to the support recovery of  $(r_0, l_0)$  through  $(\tilde{\mathcal{Q}}_\lambda(y))$ . As described 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 of the classical LASSO described in the companion paper [12, Section 2].

Since the regularization term is of the form

$$F(r, l) \stackrel{\text{def.}}{=} \sum_{i=0}^{G-1} f(r_i) + \sum_{i=0}^{G-1} f(l_i), \quad \text{with } f(x) \stackrel{\text{def.}}{=} \sup \{qx ; q \leq 1\} = \begin{cases} x & \text{if } x \geq 0, \\ +\infty & \text{otherwise,} \end{cases}$$

its subdifferential is the product of the subdifferentials  $\partial f(r_i)$  and  $\partial f(l_i)$  for  $1 \leq i \leq$

$G - 1$ , where

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

That is 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. Hence, one may adapt the standard results for the LASSO to the LASSO with positivity constraint, simply by 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  $(\mathcal{Q}_\lambda(y))$ .

### 2.3. Optimality conditions

The optimality conditions for  $(\tilde{\mathcal{Q}}_\lambda(y))$  and  $(\tilde{\mathcal{Q}}_0(y_0))$ , written in terms of  $(a_\lambda, b_\lambda)$ , yield the following results.

**Proposition 1.** *Let  $y \in \mathcal{H}$ ,  $(a_\lambda, b_\lambda) \in \mathcal{C}_h^G$ , 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  $p_\lambda \in \mathcal{H}$  such that*

$$\max \left( (A^* + \frac{h}{2}B^*)p_\lambda \right) \leq 1, \quad \text{and} \quad \max \left( (A^* - \frac{h}{2}B^*)p_\lambda \right) \leq 1, \quad (22)$$

$$(A_{I^{(r)}}^* + \frac{h}{2}B_{I^{(r)}}^*)p_\lambda = \mathbf{1}_{I^{(r)}}, \quad \text{and} \quad (A_{I^{(l)}}^* - \frac{h}{2}B_{I^{(l)}}^*)p_\lambda = \mathbf{1}_{I^{(l)}}, \quad (23)$$

$$\lambda \begin{pmatrix} A^* \\ B^* \end{pmatrix} p_\lambda + \begin{pmatrix} A^* \\ B^* \end{pmatrix} (Aa_\lambda + Bb_\lambda - y) = 0. \quad (24)$$

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

$$\max \left( (A^* + \frac{h}{2}B^*)p \right) \leq 1, \quad \text{and} \quad \max \left( (A^* - \frac{h}{2}B^*)p \right) \leq 1, \quad (25)$$

$$(A_{I^{(r)}}^* + \frac{h}{2}B_{I^{(r)}}^*)p = \mathbf{1}_{I^{(r)}}, \quad \text{and} \quad (A_{I^{(l)}}^* - \frac{h}{2}B_{I^{(l)}}^*)p = \mathbf{1}_{I^{(l)}}, \quad (26)$$

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

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

**Proposition 2.** *Under the hypotheses of Proposition 1, if  $\left( (A + \frac{h}{2}B)_{I^{(r)}} \quad (A - \frac{h}{2}B)_{I^{(l)}} \right)$  has full rank and if  $p_\lambda$  (resp.  $p$ ) satisfies*

$$\forall k \in I^c, \quad (A^*p_\lambda)_k + \frac{h}{2}|(B^*p_\lambda)_k| < 1, \quad (27)$$

$$\forall i \in I^{(l)} \setminus I^{(r)}, \quad ((A^* + \frac{h}{2}B^*)p_\lambda)_i < 1, \quad (28)$$

$$\forall i \in I^{(r)} \setminus I^{(l)}, \quad ((A^* - \frac{h}{2}B^*)p_\lambda)_i < 1, \quad (29)$$

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

One may interpret the optimality conditions of Proposition 1 as the primal-dual relations between the solutions of  $(\mathcal{Q}_\lambda(y))$  and  $(\mathcal{Q}_0(y_0))$  with the solutions of their respective dual problems

$$\begin{aligned} \inf_{p \in D} \left\| \frac{y}{\lambda} - p \right\|_{\mathcal{H}}^2 & \quad (\mathcal{E}_\lambda(y)) \\ \sup_{p \in D} \langle y, p \rangle & \quad (\mathcal{E}_0(y)) \\ \text{where } D \stackrel{\text{def.}}{=} \left\{ p \in \mathcal{H} ; \max_{k \in \llbracket 0, G-1 \rrbracket} (A^*p)_k + \frac{h}{2} |(B^*p)_k| \leq 1 \right\}. & \quad (30) \end{aligned}$$

Conversely,  $p \in \mathcal{H}$  is a solution to  $(\mathcal{E}_\lambda(y))$  (resp.  $(\mathcal{E}_0(y))$ ) if and only if  $p \in D$  and  $p$  satisfies the conditions of Proposition 1.

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

The theorem of Fuchs [17] for the LASSO describes an almost necessary and sufficient condition for the support stability of the problem at low noise. Its adaptation to the positive LASSO is straightforward. However, the criterion that it provides is not satisfied in general, and the support at low noise is strictly larger than  $(I^{(\mathbf{r})}(a_0, b_0), I^{(1)}(a_0, b_0))$ . We provide below a finer description of that support by studying the minimal norm certificate.

**Definition 1** (Minimal norm certificate). *Let  $(a_0, b_0) \in \mathcal{C}_h^G$ . Its minimal norm certificate is  $\bar{\eta}_0 \stackrel{\text{def.}}{=} \begin{pmatrix} A^* + \frac{h}{2} B^* \\ A^* - \frac{h}{2} B^* \end{pmatrix} p_0$  where  $p_0 \in \mathcal{H}$  is the solution to  $(\mathcal{E}_0(y))$  with minimal norm.*

*The extended support is  $\text{ext}^{(\mathbf{r}1)}(a_0, b_0) = (\text{ext}^{(\mathbf{r})}(a_0, b_0), \text{ext}^{(1)}(a_0, b_0))$ , where*

$$\text{ext}^{(\mathbf{r})}(a_0, b_0) = \left\{ j \in \llbracket 0, G-1 \rrbracket ; ((A^* + \frac{h}{2} B^*)p_0)_j = 1 \right\}, \quad (31)$$

$$\text{ext}^{(1)}(a_0, b_0) = \left\{ j \in \llbracket 0, G-1 \rrbracket ; ((A^* - \frac{h}{2} B^*)p_0)_j = 1 \right\}. \quad (32)$$

From the optimality conditions, if  $(a_0, b_0)$  is a solution of  $(\mathcal{Q}_0(y_0))$  then  $I^{(\mathbf{r})} \subset \text{ext}^{(\mathbf{r})}(a_0, b_0)$  and  $I^{(1)} \subset \text{ext}^{(1)}(a_0, b_0)$  (where  $I = I(a_0, b_0)$ ), and  $p_0$  can be characterized as

$$p_0 = \underset{p \in \mathcal{H}}{\text{argmin}} \left\{ \|p\|_{\mathcal{H}} ; \begin{pmatrix} A^* + \frac{h}{2} B^* \\ A^* - \frac{h}{2} B^* \end{pmatrix} p \in \partial F(r_0, l_0) \right\}. \quad (33)$$

It is now possible to describe the behavior of  $(\mathcal{Q}_\lambda(y))$  at low noise in the generic case. We state the following theorem for problems of the form  $(\mathcal{Q}_\lambda(y))$  and  $(\mathcal{Q}_0(y_0))$  as it is of independent interest. But in Section 4.2 below (see Theorem 3 and Corollary 1), we establish a more precise version of this result, which describes the constants and the extended support, in the special case of the C-BP for a continuous inverse problem.

**Theorem 1.** Let  $(a_0, b_0) \in \mathcal{C}_h^G \setminus \{0\}$  be an identifiable signal,  $(J^{(\mathbf{r})}, J^{(1)}) \stackrel{\text{def.}}{=} \text{ext}^{(\mathbf{r}1)}(a_0, b_0)$  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

$$\begin{pmatrix} u_{J^{(\mathbf{r})}} \\ v_{J^{(1)}} \end{pmatrix} \stackrel{\text{def.}}{=} -(\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} s \quad \text{where} \quad s \stackrel{\text{def.}}{=} (1, \dots, 1)^* \in \mathbb{R}^{|J^{(\mathbf{r})}| + |J^{(1)}|},$$

and assume that the following non-degeneracy condition holds

$$\forall j \in J^{(\mathbf{r})} \setminus I^{(\mathbf{r})}, u_j > 0, \quad \text{and} \quad \forall j \in J^{(1)} \setminus I^{(1)}, v_j > 0. \quad (34)$$

Then, there exists constants  $C^{(1)} > 0$ ,  $C^{(2)} > 0$ ,  $C^{(3)} > 0$  such that for

$$C^{(1)} \|w\|_{\mathcal{H}} + C^{(2)} \lambda \leq \min \left( \left\{ \frac{1}{2} a_{0,i} + \frac{1}{h} b_{0,i}; i \in I^{(\mathbf{r})}(a_0, b_0) \right\} \right. \quad (35)$$

$$\left. \cup \left\{ \frac{1}{2} a_{0,i} - \frac{1}{h} b_{0,i}; i \in I^{(1)}(a_0, b_0) \right\} \right) \quad (36)$$

and  $\|w\|_{\mathcal{H}} \leq C^{(3)} \lambda$ , 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 (21).

Some comments about the proofs of Theorem 1 can be found in [Appendix B](#).

### 3. Continuous-Basis Pursuit on (thin) grids

Now, we turn to the ‘‘continuous’’ inverse problem described in the introduction, and we assume that each  $\alpha_\nu$  ( $1 \leq \nu \leq N$ ) is positive. We aim at recovering  $m_0$  using the Continuous Basis-Pursuit (C-BP) proposed in [14]. Given a grid  $\mathcal{G}$  with stepsize  $h > 0$ , the goal is to reconstruct a measure  $m = \sum_{i=0}^{G_n-1} a_i \delta_{ih_n+t_i}$  where  $t_i \in [-\frac{h}{2}, \frac{h}{2}]$  which estimates  $m_0$ . Applying a Taylor expansion to  $\varphi$  and setting  $b_i = t_i a_i$ , the authors of [14] are led to solve

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

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

which are particular instances of  $(\mathcal{Q}_\lambda(y))$  and  $(\mathcal{Q}_0(y_0))$ , with the choice  $(A, B) = (\Phi_{\mathcal{G}}, \Phi'_{\mathcal{G}})$ . To study the behavior of the solutions to these problems, we shall apply the results of the previous section, and in particular Lemma 3.

### 3.1. Motivating example: the ideal regime

Let us first examine a regime in which the C-BP introduced in [14] identifies the spikes locations with sub-grid accuracy, motivating its use for inverse problems instead of the LASSO.

Let  $h > 0$  be fixed, let  $\bar{x}_1, \dots, \bar{x}_N \in \mathcal{G}$  be fixed points and  $\bar{I} \in \llbracket 0, G-1 \rrbracket^N$  such that  $\bar{x} = (\bar{x}_1, \dots, \bar{x}_N) = \bar{I}h$ . Assuming that a solution  $\begin{pmatrix} a \\ b \end{pmatrix}$  to  $(\mathcal{Q}_\lambda(y))$  has support  $(\bar{I}, \bar{I})$ , we deduce by restricting (24) to  $(\bar{I}, \bar{I})$  that

$$\lambda \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} + \begin{pmatrix} \Phi_{\bar{x}}^* \\ \Phi_{\bar{x}}'^* \end{pmatrix} \left( \begin{pmatrix} \Phi_{\bar{x}} & \Phi_{\bar{x}}' \end{pmatrix} \begin{pmatrix} a_{\bar{I}} \\ b_{\bar{I}} \end{pmatrix} - y \right) = 0. \quad (37)$$

Writing  $\Gamma_{\bar{x}} \stackrel{\text{def.}}{=} \begin{pmatrix} \Phi_{\bar{x}} & \Phi_{\bar{x}}' \end{pmatrix}$ , and assuming that  $\Gamma_{\bar{x}}$  has full rank, we get

$$\begin{pmatrix} a_{\bar{I}} \\ b_{\bar{I}} \end{pmatrix} = (\Gamma_{\bar{x}}^* \Gamma_{\bar{x}})^{-1} \Gamma_{\bar{x}}^* y - \lambda (\Gamma_{\bar{x}}^* \Gamma_{\bar{x}})^{-1} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix}. \quad (38)$$

If the input is  $y = \Phi m_0 + w$  with  $m_0 \stackrel{\text{def.}}{=} \sum_{\nu=1}^N \alpha_\nu \delta_{x_{0,\nu}}$  and  $x_{0,\nu} = \bar{x}_\nu + s_\nu$ , then

$$\begin{pmatrix} a_{\bar{I}} \\ b_{\bar{I}} \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Gamma_{\bar{x}}^+ (R_\alpha(s) + w) - \lambda (\Gamma_{\bar{x}}^* \Gamma_{\bar{x}})^{-1} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix}, \quad (39)$$

where  $\beta_\nu = \alpha_\nu s_\nu$ ,  $\Gamma_{\bar{x}}^+ = (\Gamma_{\bar{x}}^* \Gamma_{\bar{x}})^{-1} \Gamma_{\bar{x}}^*$  and  $R_\alpha(s) = \Phi m_0 - \Gamma_{\bar{x}} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = O(\|\alpha\|_\infty \|s\|_\infty^2)$  is the remainder of the Taylor expansion of  $\Phi m_0$  around  $\bar{x}$ . In other words, provided this regime holds, the C-BP is able to retrieve the positions  $\bar{x}_\nu + s_\nu$  up to an error  $O(\|\alpha\|_\infty \|s\|_\infty^2 + \|\Gamma_{\bar{x}}^* w\|_\infty + \lambda)$ , yielding a better precision than the LASSO (which yields a  $O(h)$  error) provided  $w$  and  $\lambda$  are small enough. We call that regime the *ideal regime*.

The existence of such a regime, *i.e.* when the solutions to  $(\mathcal{Q}_\lambda(y))$  have support  $(\bar{I}, \bar{I})$  is discussed in the next theorem (which is proved in Appendix C.1). Before stating it, we define  $\bar{\eta}_V \stackrel{\text{def.}}{=} \Phi^* \Gamma_{\bar{x}}^{+,*} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix}$  provided  $\Gamma_{\bar{x}}$  has full rank.

**Theorem 2.** *Let  $\bar{x} \in \mathcal{G}^N$ , such that  $\Gamma_{\bar{x}}$  has full rank, and assume that*

$$\max_{t \in \mathcal{G} \setminus \{\bar{x}_1, \dots, \bar{x}_N\}} \left( \bar{\eta}_V \pm \frac{h}{2} \bar{\eta}_V' \right) (t) < 1. \quad (\text{IC}_h)$$

*Then there exists constants  $C_1 > 0$  (which only depends on  $\bar{x}$  and  $\varphi$ ),  $C_2 > 0$  (which depends on  $\bar{x}, \varphi$  and  $h$ ), such that for all input measure  $m_0 \stackrel{\text{def.}}{=} \sum_{\nu=1}^N \alpha_\nu \delta_{x_{0,\nu}}$  with  $\alpha_\nu > 0$  and  $x_{0,\nu} = \bar{x}_\nu + s_\nu$ , all  $\lambda > 0$ ,  $w \in L^2(\mathbb{T})$  with*

$$\|s\|_\infty < \frac{h}{2} \left( 1 - C_1 \left( \frac{\|\alpha\|_\infty \|s\|_\infty^2 + \|\Gamma_{\bar{x}}^* w\|_\infty + \lambda}{\min \alpha} \right) \right), \quad (40)$$

$$\text{and } \frac{\|\alpha\|_\infty \|s\|_\infty^2 + \|w\|_{\mathcal{H}}}{\lambda} < C_2, \quad (41)$$

the solution to  $(\mathcal{Q}_\lambda(y))$  is unique, has support  $(\bar{I}, \bar{I})$ , and is given by (39).

*Remark 2.* Condition  $(\mathbf{IC}_h)$  ensures that for all measures which have support in a small neighborhood of  $\bar{x}_1, \dots, \bar{x}_N$ , the C-BP is able to identify the support up to a small error. It is similar to the Non-Degenerate Source Condition (NDSC) introduced in [11] to study the noise stability of the Beurling LASSO and which involves the vanishing derivatives precertificate (see below). Whereas some theoretical results ensure the NDSC for sufficiently separated measures (see [21]), it seems more challenging to ensure a priori the validity of  $(\mathbf{IC}_h)$ .

In the particular case of a single spike and a convolution operator (that is  $\mathcal{H} = L^2(\mathbb{T})$ ,  $\varphi(x) : x' \mapsto \tilde{\varphi}(x' - x)$ ), the following result ensures the existence of the ideal regime (see Appendix C.2 for the proof).

**Proposition 3.** *If  $\Phi$  is a convolution operator with  $\tilde{\varphi} \in \mathcal{C}^4(\mathbb{T}) \setminus \{0\}$  such that  $\tilde{\varphi}$  is not  $T$ -periodic for  $T < 1$ , there exists  $h_1 > 0$  such that for all  $h \in (0, h_1]$ , any  $\bar{x} \in \mathcal{G}$  satisfies the hypotheses of Theorem 2. In particular, for any input  $m_0 = \alpha_1 \delta_{\bar{x}_1 + s_1}$ , the solution is*

$$\begin{pmatrix} a_{i_1} \\ b_{i_1} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_1 s_1 \end{pmatrix} + O(\alpha_1 s_1^2 + \|\Gamma_{\bar{x}}^* w\|_\infty + \lambda).$$

For measures with more Dirac masses, it seems difficult to state any general result. Numerical experiments reported in Section 5.1 suggest that the ideal regime exists for reasonable filters and measures with sufficiently separated spikes. But this regime crucially depends on the stepsize  $h$ . Indeed, as a consequence of Lemma 5, assuming  $(\mathbf{IC}_h)$  for all  $h$  small enough yields strong consequences such as  $\overline{\eta_V^{(3)}}(\bar{x}_\nu) = 0$ , which is not true in general. Therefore we conjecture that for typical signals and filters, Condition  $(\mathbf{IC}_h)$  is only valid in some interval  $h_0 < h < h_1$ , where  $h_0 > 0$ . As the numerical experiments suggest,  $h_0$  and  $h_1$  depend on the repartition of  $\bar{x}_1, \dots, \bar{x}_N$ : the interval  $(h_0, h_1)$  is large for well separated spikes, and small for densely clustered spikes.

### 3.2. The Positive Beurling LASSO

As the C-BP motivation stems from a Taylor expansion (see (9)) which is only valid for small shifts  $s$ , it is natural to expect that the performance of C-BP improves when using medium or small stepsizes, as opposed to larger ones. The situation is in fact more subtle. Now and for the next section, we are interested in the behavior of the model on a thin grid  $\mathcal{G}_n$ ,  $n \in \mathbb{N}$ , with stepsize  $h_n \searrow 0^+$  as  $n \rightarrow +\infty$ . The problems we consider are

$$\begin{aligned} \min_{(a,b) \in \mathcal{C}_{h_n}^{G_n}} \frac{1}{2} \|y - \Phi_{\mathcal{G}_n} a - \Phi'_{\mathcal{G}_n} b\|_{\mathcal{H}}^2 + \lambda \|a\|_1 & \quad (\mathcal{Q}_\lambda^n(y)) \\ \min_{(a,b) \in \mathcal{C}_{h_n}^{G_n}} \|a\|_1 \text{ such that } \Phi_{\mathcal{G}_n} a + \Phi'_{\mathcal{G}_n} b = y_0. & \quad (\mathcal{Q}_0^n(y_0)) \end{aligned}$$



and their dual problems are respectively equivalent to:

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

$$\text{where } D^n \stackrel{\text{def.}}{=} \left\{ q \in \mathcal{H} ; \max_{k \in \llbracket 0, G_n - 1 \rrbracket} (\Phi^* q)(kh_n) + \frac{h_n}{2} |(\Phi^* q)'(kh_n)| \leq 1 \right\}, \quad (42)$$

As we explain below, a natural limit model of C-BP as  $n \rightarrow +\infty$  is the *positive Beurling LASSO*,

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

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:

$$m(\mathbb{T}) + \iota_{\mathcal{M}^+(\mathbb{T})}(m) = \sup \left\{ \int_{\mathbb{T}} \psi(t) dm(t) ; \psi \in \mathcal{C}(\mathbb{T}) \text{ and } \sup_{t \in \mathbb{T}} \psi(t) \leq 1 \right\}. \quad (43)$$

The characterization of optimality, the notions of minimal norm certificates and extended support are straightforward adaptations of those of the Beurling LASSO exposed in the companion paper [12, Section 3]. Again, it essentially amounts to replacing condition  $\|\eta\|_\infty \leq 1$  with  $\sup_{t \in \mathbb{T}} \eta(t) \leq 1$  where  $\eta = \Phi^* p$  for  $p \in \mathcal{H}$ . 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:

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

$$\text{where } D^\infty \stackrel{\text{def.}}{=} \left\{ p \in \mathcal{H} ; \sup_{t \in \mathbb{T}} (\Phi^* p)(t) \leq 1 \right\}. \quad (44)$$

As for the Beurling LASSO, the low noise behavior of  $(\mathcal{Q}_\lambda^\infty(y))$  is governed by the minimal norm solution of  $(\mathcal{E}_0^\infty(y_0))$  (see [12, Section 3]). That solution being difficult to compute in general, one is led to study a “good candidate” for it, the vanishing derivatives precertificate, which can easily be computed by solving a linear system in the least square sense. In this paper, we are not directly interested in the low noise behavior of  $(\mathcal{Q}_\lambda^\infty(y))$  but we shall use this precertificate as an auxiliary quantity, hence we may adopt the following definition.

**Definition 2.** Let  $m_0 = \sum_{\nu=1}^N \alpha_{0,\nu} \delta_{x_{0,\nu}}$  an identifiable measure for the problem  $(\mathcal{Q}_0^\infty(y_0))$  such that  $\Gamma_{x_0} \stackrel{\text{def.}}{=}} \begin{pmatrix} \Phi_{x_0} & \Phi'_{x_0} \end{pmatrix}$  has full rank.

The vanishing derivatives precertificate is defined as

$$\eta_{V,\infty} \stackrel{\text{def.}}{=} \Phi^* p_{V,\infty} \quad \text{where } p_{V,\infty} \stackrel{\text{def.}}{=} \Gamma_{x_0}^{+,*} \begin{pmatrix} s \\ 0 \end{pmatrix}, \quad (45)$$

where  $s = (1, \dots, 1)^T \in \mathbb{R}^N$  and  $\Gamma_{x_0}^{+,*} = \Gamma_{x_0}^* (\Gamma_{x_0} \Gamma_{x_0}^*)^{-1}$ .

### 3.3. The Limit Problem for Thin Grids

Let us recall that we obtain a measure from the vector  $(a, b) \in \mathcal{C}_{h_n}^{G_n}$  by setting

$$m = \sum_{i=0}^{G_n-1} a_i \delta_{ih_n + b_i/a_i} \quad (46)$$

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

In this section we prove the  $\Gamma$ -convergence of  $(\mathcal{Q}_\lambda^n(y))$  towards  $(\mathcal{Q}_\lambda^\infty(y))$ . Since all the solutions to  $(\mathcal{Q}_\lambda^n(y))$  belong to  $X_+ \stackrel{\text{def.}}{=} \{m \in \mathcal{M}^+(\mathbb{T}) ; \lambda|m|(\mathbb{T}) \leq \frac{1}{2} \|y\|_{\mathcal{H}}^2\}$ , we may restrict the problems to  $X_+$ , which is metrizable for the weak\* topology. Hence, we may use the following formulation of  $\Gamma$ -convergence valid in metric spaces (see [8]).

**Definition 3.** *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 (46) with  $(a^{(n)}, b^{(n)}) \in \mathcal{C}_{h_n}^{G_n}$  such that  $m_n$  weakly\* converges towards  $m$ ,

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

- (Limsup inequality) there exists a sequence of measures  $(m_n)_{n \in \mathbb{N}} \in X_+^{\mathbb{N}}$  of the form (46) with  $(a^{(n)}, b^{(n)}) \in \mathcal{C}_{h_n}^{G_n}$  such that  $m_n$  weakly\* converges towards  $m$  and

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

The following proposition, which is proved in [Appendix C.3](#), states the  $\Gamma$ -convergence of the model and its consequences.

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

$$\lim_{n \rightarrow +\infty} (\inf (\mathcal{Q}_\lambda^n(y))) = \inf (\mathcal{Q}_\lambda^\infty(y)). \quad (47)$$

*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  $m_{\lambda,\infty}$ .

## 4. Convergence of the support

Though [Proposition 4](#) 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 1). In this continuous context, a dual certificate is determined by a function  $\eta = \Phi^* p \in \mathcal{C}(\mathbb{T})$  where  $p \in \mathcal{H}$ , and if  $(a, b)$  is a solution to  $(\mathcal{Q}_\lambda^n(y))$ ,

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

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

To sum up, we shall exploit the following observations

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

In the next two paragraphs, we describe the behavior of the support. Those results rely on auxiliary Lemmas in [Appendix D](#).

#### 4.1. Asymptotics of the Support for fixed $\lambda > 0$

The following proposition relies on the convergence of the dual certificates (see Lemma 7 in Appendix). It 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. In the statement below,  $r$  is chosen such that  $0 < r < \frac{1}{2} \min_{\nu \neq \nu'} |x_\nu - x_{\nu'}|$ .

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

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

$$\lim_{n \rightarrow +\infty} (\text{supp}(m_{\lambda, n})) = \{x_1, \dots, x_N\}. \quad (48)$$

*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  $(x_\nu - r, x_\nu + r)$  is a sum of Dirac masses whose configuration is given in Table 1, and if  $(\eta_{\lambda, \infty})^{(3)}(x_\nu) \neq 0$ , then only the cases indicated with (\*) may appear.*

*Remark 3.* Proposition 5 states that the support of the C-BP on thin grids actually depends on the properties of the dual certificate  $\eta_{\lambda, \infty}$  of the (positive) Beurling LASSO. The condition  $\eta_{\lambda, \infty}''(x_\nu) \neq 0$  seems to be overwhelming, or generic, and it is ensured for instance if  $\lambda$  is small and the Non-Degenerate Source Condition holds (see [11]). As for the condition  $\eta_{\lambda, \infty}^{(3)}(x_\nu) \neq 0$ , it also seems to be generic (with the notable exception of

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

Table 1: Number of Dirac masses that may appear if  $\eta''_{\lambda,\infty}(x_\nu) \neq 0$ . 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.

single spike input measures), as there is nothing to impose  $\eta''_{\lambda,\infty}(x_\nu) = 0$  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, if the grid is too thin, 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  $(ih - \frac{h}{2}, (i+1)h - \frac{h}{2})$ .

#### 4.2. Asymptotic of the Low Noise Support

Now, we pursue the discussion of Section 3.1, trying to describe the solutions when  $h$  is too small for  $(IC_h)$  to hold. Again, we consider measure  $m_0 \stackrel{\text{def}}{=} \sum_{\nu=1}^N \alpha_\nu \delta_{x_{0,\nu}}$  where  $x_{0,\nu} = \bar{x}_\nu + s_\nu$ , and we assume that  $\bar{x} \in (\mathcal{G}_n)^N$  for  $n$  large enough (for the sake of simplicity, we consider increasing grids  $(\mathcal{G}_n \subset \mathcal{G}_{n+1})$  and we assume that  $\bar{x} \in \mathcal{G}_0$ ). As the computations in Section 3.1 suggest, we may split the observation as  $y = \bar{y} + (R_\alpha(s) + w)$  where  $\bar{y} \stackrel{\text{def}}{=} \Gamma_{\bar{x}} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$  and  $\beta_\nu = \alpha_\nu s_\nu$  for all  $1 \leq \nu \leq N$ . In other words, we regard the

remainder  $R_\alpha(s)$  of the Taylor expansion as an additional noise term induced by the shift  $s$ .

Whereas the previous subsection considered a fixed value of  $\lambda$  and  $w$ , here, to be able to describe precisely the support based upon intrinsic properties of  $\bar{x}$ , we focus our attention on the simpler situation of a *low noise regime*, that is:

- the parameter  $\lambda > 0$  is small,
- the noise  $w \in \mathcal{H}$  is small,
- the shift  $s \in \mathbb{R}^N$  is small.

The bounds on  $\lambda$ ,  $w$  and  $s$  depend on  $n$  (i.e. on  $h_n$ ), they will be made precise below. In that regime, the problem  $(\mathcal{Q}_\lambda^n(y))$  approximates

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

and the support of the solutions to  $(\mathcal{Q}_\lambda^n(y))$  is governed by the saturations of the minimal norm solutions of the dual problem  $(\mathcal{E}_0^n(\bar{y}))$ . The difficult point here is that the minimal norm solutions of  $(\mathcal{E}_0^n(\bar{y}))$  do not converge towards the minimal norm solution of  $(\mathcal{E}_0^\infty(\bar{y}))$ , and we need introduce a new variational problem to carry the study further.

**Definition 4** (Third derivative precertificate). *Given  $\bar{x} \in (\mathcal{G}_0)^N$ , we define the third derivative precertificate as  $\eta_T \stackrel{\text{def.}}{=} \Phi^* p_T$  where*

$$p_T \stackrel{\text{def.}}{=} \operatorname{argmin}_{p \in \mathcal{H}} \left\{ \|p\|_{\mathcal{H}}; \forall \nu \in \{1, \dots, N\}, (\Phi^* p)(\bar{x}_\nu) = 1, \right. \\ \left. (\Phi^* p)'(\bar{x}_\nu) = 0 \text{ and } (\Phi^* p)^{(3)}(\bar{x}_\nu) = 0 \right\}, \quad (49)$$

whenever the above set is not empty.

Note that  $p_T$  is the solution to a quadratic minimization under linear constraint, and can hence be computed by solving a linear system, provided  $\begin{pmatrix} \Phi_{\bar{x}} & \Phi'_{\bar{x}} & \Phi_{\bar{x}}^{(3)} \end{pmatrix}$  has full rank,

$$p_T = \begin{pmatrix} \Gamma_{\bar{x}}^* \\ \Phi_{\bar{x}}^{(3)*} \end{pmatrix}^+ \begin{pmatrix} \mathbb{1}_N \\ 0_N \\ 0_N \end{pmatrix} = \begin{pmatrix} (\Pi_{(\operatorname{Im} \Phi_{\bar{x}}^{(3)})^\perp} \Gamma_{\bar{x}}^+)^+ \\ (\Pi_{(\operatorname{Im} \Gamma_{\bar{x}})^\perp} \Phi_{\bar{x}}^{(3)})^+ \end{pmatrix} \begin{pmatrix} \mathbb{1}_N \\ 0_N \\ 0_N \end{pmatrix} = (\Pi_{(\operatorname{Im} \Phi_{\bar{x}}^{(3)})^\perp} \Gamma_{\bar{x}}^+)^+ \begin{pmatrix} \mathbb{1}_N \\ 0_N \end{pmatrix}, \quad (50)$$

where  $\Pi_{(\operatorname{Im} \Gamma_{\bar{x}})^\perp}$  (resp.  $\Pi_{(\operatorname{Im} \Phi_{\bar{x}}^{(3)})^\perp}$ ) is the orthogonal projector onto  $(\operatorname{Im} \Gamma_{\bar{x}})^\perp$  (resp.  $(\operatorname{Im} \Phi_{\bar{x}}^{(3)})^\perp$ ), and  $\Gamma_{\bar{x}} = \begin{pmatrix} \Phi_{\bar{x}} & \Phi'_{\bar{x}} \end{pmatrix}$ . The third derivative precertificate is involved in the following technical assumption.

**Definition 5** (Twice Non-Degenerate Source Condition). *Given  $\bar{x} \in (\mathcal{G}_0)^N$ , we say that the Twice Non-Degenerate Source Condition (TNDSC) holds if  $p_T$  in (49) is well defined and if it satisfies, for  $\eta_T = \Phi^* p_T$ ,*

$$\forall t \in \mathbb{T} \setminus \{\bar{x}_1, \dots, \bar{x}_N\}, \quad \eta_T(t) < 1, \\ \forall \nu \in \{1, \dots, N\}, \quad \eta_T''(\bar{x}_\nu) < 0 \quad \text{and} \quad \eta_T^{(4)}(\bar{x}_\nu) > 0.$$

*Remark 4.* In view of Lemma 4 in Appendix C.2, the Twice Non-Degenerate Source Condition ensures that there exists  $n_1 \in \mathbb{N}$  such that for all  $n \geq n_1$ ,  $p_T \in D^n$ . From the optimality conditions (Proposition 1), we deduce that for any  $(\alpha, s) \in (\mathbb{R}_+^*)^N \times \mathbb{R}^N$ , the corresponding vector  $(\bar{a}, \bar{b}) \in \mathbb{R}^{G_n} \times \mathbb{R}^{G_n}$  defined by

$$(\bar{a}_i, \bar{b}_i) = \begin{cases} (\alpha_\nu, \beta_\nu) & \text{if } ih_n = \bar{x}_\nu \\ 0 & \text{otherwise,} \end{cases} \quad (51)$$

is a solution to  $(\mathcal{Q}_0^n(\bar{y}))$ , where  $\bar{y} = \Phi_{\mathcal{G}_n} \bar{a} + \Phi'_{\mathcal{G}_n} \bar{b}$ , provided  $\|s\|_\infty \leq \frac{h_n}{2}$  (i.e.  $(\bar{a}, \bar{b}) \in \mathcal{C}_{h_n}$ ).

We are now in position to describe precisely the corresponding extended support, with a prediction on the location of the neighbor. The proof of this Theorem can be found in Appendix D.2.

**Theorem 3.** *Let  $\bar{x} \in (\mathcal{G}_0)^N$  such that the operator  $\begin{pmatrix} \Phi_{\bar{x}} & \Phi'_{\bar{x}} & \Phi_{\bar{x}}^{(3)} \end{pmatrix}$  has full rank and that the Twice Non Degenerate Source condition (Definition 5) holds. Moreover, assume that all the components of the natural shift*

$$\rho \stackrel{\text{def.}}{=} (\Phi_{\bar{x}}^{(3)*} \Pi_{(\text{Im } \Gamma_{\bar{x}})^\perp} \Phi_{\bar{x}}^{(3)})^{-1} \Phi_{\bar{x}}^{(3)*} \Gamma_{\bar{x}}^{+,*} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} \quad (52)$$

are nonzero. Then, for  $n$  large enough, any vector  $(\bar{a}, \bar{b}) \in \mathcal{C}_{h_n}$  defined by (51), with  $\bar{y} \stackrel{\text{def.}}{=} \Phi_{\mathcal{G}_n} \bar{a} + \Phi'_{\mathcal{G}_n} \bar{b}$ , is a solution to  $(\mathcal{Q}_0^n(\bar{y}))$  and its extended support has the form

$$\begin{aligned} \text{ext}_n^{(\mathbf{r})}(\bar{a}, \bar{b}) &= \{i_1, \dots, i_N\} \cup \{i_\nu - 1; \nu \in \llbracket 1, N \rrbracket \text{ and } \rho_\nu > 0\} \\ \text{ext}_n^{(\mathbf{l})}(\bar{a}, \bar{b}) &= \{i_1, \dots, i_N\} \cup \{i_\nu + 1; \nu \in \llbracket 1, N \rrbracket \text{ and } \rho_\nu < 0\}. \end{aligned}$$

where  $i_\nu \in \llbracket 0, G_n - 1 \rrbracket$  is defined by  $i_\nu h_n = \bar{x}_\nu$ .

*Remark 5.* The extended support is locally constant:  $\text{ext}_n^{(\mathbf{r})}(\bar{a}, \bar{b}) = \text{ext}_n^{(\mathbf{r})}(\bar{a}, 0)$  and  $\text{ext}_n^{(\mathbf{l})}(\bar{a}, \bar{b}) = \text{ext}_n^{(\mathbf{l})}(\bar{a}, 0)$  for all  $s$  such that  $\|s\|_\infty \leq \frac{h_n}{2}$ .

The next result is a specific variant of Theorem 1 which precisely describes the low noise solutions on thin grids. It states that, at low noise, the C-BP reconstructs a measure of the form

$$m_\lambda = \sum_{\nu=1}^N \left( \alpha_{\lambda, \nu}^{(1)} \delta_{\bar{x}_\nu + t_\nu} + \alpha_{\lambda, \nu}^{(2)} \delta_{\bar{x}_\nu + \varepsilon_\nu h/2} \right) \quad \text{where} \quad \begin{cases} \varepsilon_\nu \in \{-1, +1\}, \\ -h/2 < t_\nu < h/2. \end{cases}$$

As in the ideal regime, the model is able to recover the locations of the original spikes (using  $\beta_\nu = s_\nu \alpha_\nu$ ) up to a small error, but two parasitic spikes appear at neighboring half-grid points.

**Corollary 1.** *Under the hypotheses of Theorem 3, there exists constants  $C^{(1)} > 0$ ,  $C^{(2)} > 0$  and  $C^{(3)} > 0$  such that for any input  $y = \Phi m_0 + w$ , with  $m_0 = \sum_{\nu=1}^N \alpha_\nu \delta_{x_{0,\nu}}$  and  $x_{0,\nu} = \bar{x}_\nu + s_\nu$ , the conditions*

- $C^{(1)} (\|\alpha\|_\infty \|s\|_\infty^2 + \|w\|_{\mathcal{H}}) + C^{(2)}\lambda \leq (h_n)^3 \min \alpha$ ,
- $(\|\alpha\|_\infty \|s\|_\infty^2 + \|w\|_{\mathcal{H}}) \leq C^{(3)}\lambda$ ,

imply that the solution to  $(\mathcal{Q}_\lambda^n(y))$  is unique. Defining the vector  $I$  by  $Ih \stackrel{\text{def.}}{=} \bar{x}$ , the nonzero components of the solution are

$$\begin{pmatrix} (a_\lambda)_I \\ (b_\lambda)_I \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + (\Pi_{(\text{Im } \Phi_x^{(3)})^\perp} \Gamma_{\bar{x}})^+ w - \lambda (\Gamma_{\bar{x}}^* \Pi_{(\text{Im } \Phi_x^{(3)})^\perp} \Gamma_{\bar{x}})^{-1} \begin{pmatrix} \mathbf{1}_N \\ \mathbf{0}_N \end{pmatrix} + \begin{pmatrix} O\left(\frac{\Delta}{h^3}\right) \\ O\left(\frac{\Delta}{h^2}\right) \end{pmatrix}, \quad (53)$$

where we denoted  $\Delta \stackrel{\text{def.}}{=} \|w\|_{\mathcal{H}} + \|\alpha\|_\infty \|s\|_\infty^2 + \lambda$ , and, letting  $\varepsilon \stackrel{\text{def.}}{=} -\text{sign}(\rho)$  (see (52)),

$$\begin{pmatrix} (a_\lambda)_{I+\varepsilon} \\ (b_\lambda)_{I+\varepsilon} \end{pmatrix} = \begin{pmatrix} O\left(\frac{\Delta}{h^3}\right) \\ O\left(\frac{\Delta}{h^2}\right) \end{pmatrix} \quad \text{with} \quad (a_\lambda)_{I+\varepsilon} + \frac{2}{h} \text{diag}(\varepsilon)(b_\lambda)_{I+\varepsilon} = 0. \quad (54)$$

The proof is given in [Appendix D.3](#).

## 5. Numerical illustrations

In this section, we illustrate the relevance of our analysis to gain a precise understanding of the recovery performance of  $\ell^1$ -type methods (LASSO and C-BP) for deconvolution. The code to reproduce these numerical experiments is available online<sup>‡</sup>.

### 5.1. Ideal regime for the deconvolution problem

We begin by illustrating the ideal regime highlighted in Section 3.1: C-BP can give a better result than LASSO, and locate spikes with sub-grid accuracy, provided the grid size is not too small.

For this purpose, we consider the deconvolution problem, that is  $\varphi(x) = \tilde{\varphi}(\cdot - x)$ , and we assume that the input measure has the form  $m_0 \stackrel{\text{def.}}{=} \sum_{\nu=1}^N \alpha_\nu \delta_{i_\nu h + s_\nu}$  where  $\alpha_\nu > 0$  are positive amplitudes,  $i_\nu \in \{0, \dots, G-1\}$  are grid indices, and  $s_\nu \in [0, h]$  indicates a shift of the positions outside the grid. We do not specifically add noise, considering observations  $y = \Phi m$ , but let us note that if the spikes do not lie on the grid, this is somehow equivalent to adding a “structured” interpolation noise in the observation. In the following experiments,  $\Phi$  is a Gaussian filter of standard deviation 0.06, with observations  $\Phi m \in \mathbb{R}^P$  sampled on a uniform grid of  $P = 1024$  points.

Figure 2, compares the solution path (evolution of the solution as a function of  $\lambda$ ) of the LASSO (with positivity constraints) and C-BP in the case of a single input spike  $N = 1$ . It is obtained for a grid size  $G = 32$ , but quite remarkably, for this single spike case (and as opposed to the case  $N > 1$  considered below), the same conclusion holds for any grid size  $G$ . This result shows that, while the LASSO always recovers a pair of spikes at quantized indices  $\{i_1, i_1 + 1\}$ , in sharp contrast, C-BP correctly recovers a

‡ <https://github.com/gpeyre/2015-IP-lasso-cbp/>

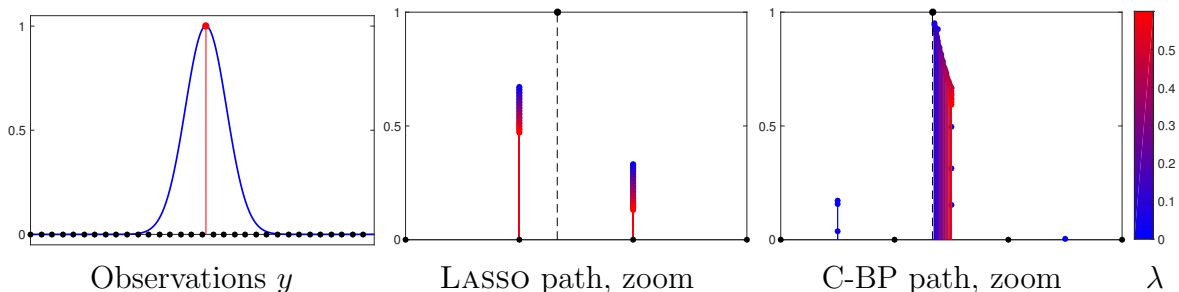


Figure 2: Lasso and C-BP paths are functions of  $\lambda$  (from blue to red) for a single Dirac input measure ( $N = 1$ ).

single Dirac, which approximates the input position  $i_0 + s_0$  for small  $\lambda$ . Note however that if  $\lambda$  is chosen too small, then the “structured” interpolation noise starts having an effect, and the recovered measure becomes degenerate (blue spurious spikes).

Figure 3 shows that the situation is more intricate in the case of  $N = 3$  input spikes (a similar conclusion holds for larger  $N$ ). This illustrates the conclusion of Section 3.1. The behaviour of the LASSO solution is simple to understand, and is always composed (even for very small  $\lambda$ ) of pairs of spikes located at  $\{i_\nu, i_\nu + 1\}$  around each input spike location  $i_\nu h + s_\nu$ . In contrast, for the C-BP, two situations emerge when the grid size  $h = 1/G$  is not too small ( $G = 32$  on the left)

- If a spike location  $i_\nu h + s_\nu$  is sufficiently far away from the other locations (here the leftmost spike), then C-BP performs well, and recovers a single spike, with a location approximating very precisely the input one. For small  $\lambda$  however, the solution degenerates and produces spurious spikes.
- If a spike location  $i_\nu + s_\nu h$  is located too close from another one (the cluster of two spikes on the right part of the domain illustrates this), then C-BP operates very similarly to LASSO, and quantizes the positions, as already seen in Section 5.3 when the spikes are on-the-grid.

If the grid size  $h = 1/G$  is too small ( $G = 64$  on the right of Figure 2), the quantization effects impacts both LASSO and C-BP, which both fails to locate spikes between grid points, as studied in Section 4.1.

## 5.2. Convergence of pre-certificates

In this section and in Section 5.3, we consider the deconvolution problem again, but  $\tilde{\varphi}$  is an ideal filter, i.e. whose Fourier coefficients

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

satisfy  $\hat{\tilde{\varphi}}(k) = 1$  if  $k \in \{-f_c, \dots, f_c\}$  and  $\hat{\tilde{\varphi}}(k) = 0$  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



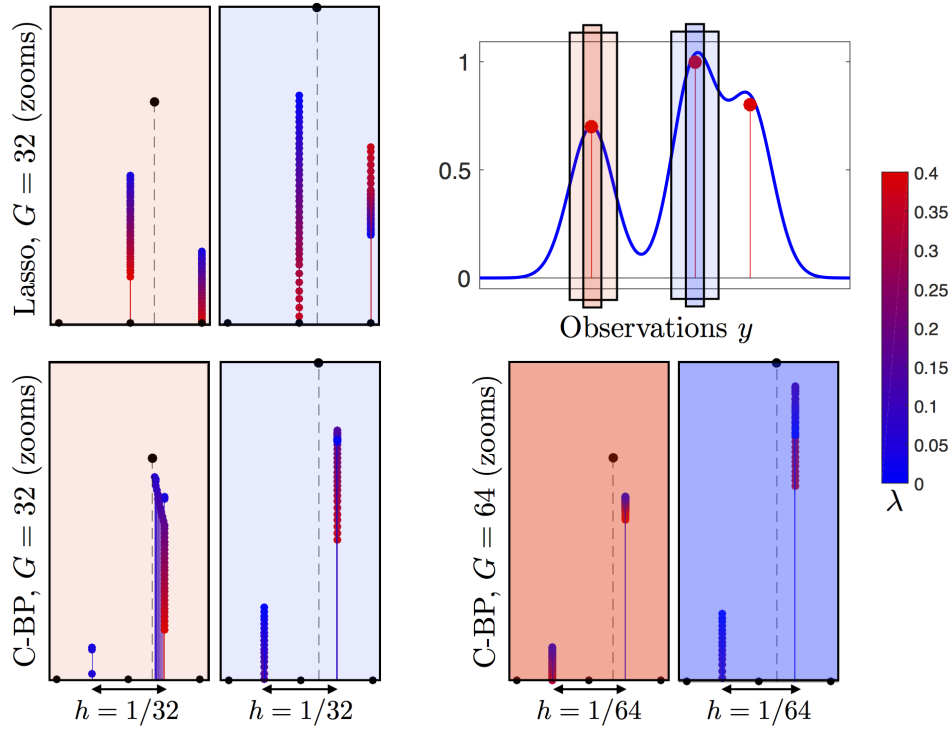


Figure 3: Lasso and C-BP paths are functions of  $\lambda$  (from blue to red) for  $N = 3$  input Diracs.

using a matrix of size  $(Q, P)$  when evaluated on a grid of  $P$  points. In Figures 4 and 5 we used  $f_c = 10$ .

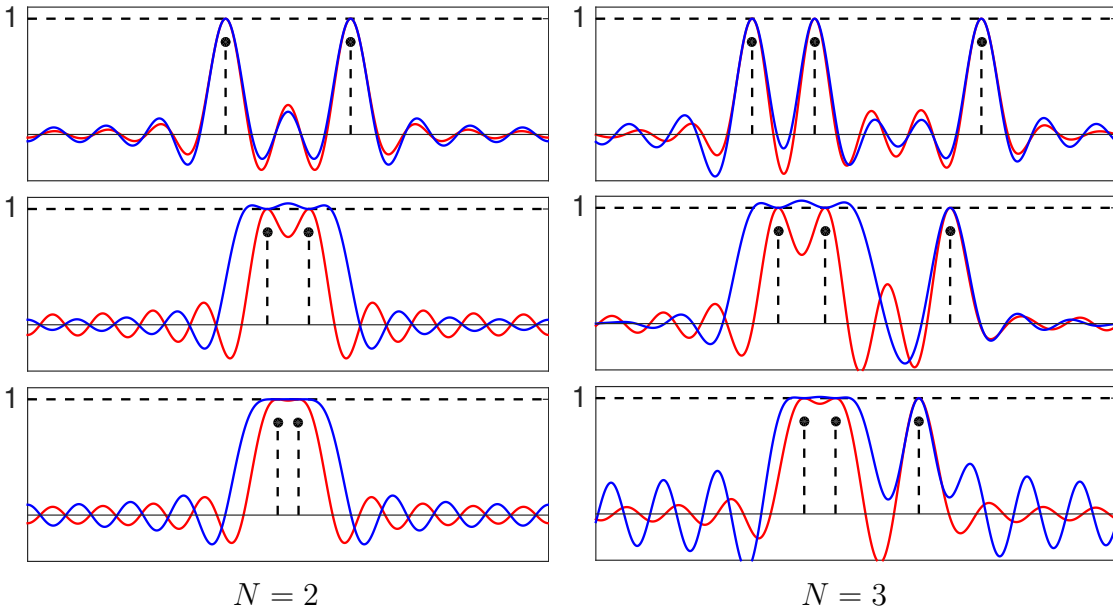


Figure 4: Display of  $\eta_{V, \infty}$  (red) and  $\eta_T$  (blue) pre-certificate for different input positive measures  $m_0$  (showed as black dots to symbolize the position of the Diracs).

Figure 4 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 2) useful to analyze LASSO/BLASSO problems and of the pre-certificate  $\eta_T$  (see Definition 4) 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 2), meaning that one actually has  $\eta_{V,\infty} = \eta_{0,\infty}$  (where the minimal norm certificate  $\eta_{0,\infty}$  is the minimal norm solution of  $(\mathcal{D}_0^\infty(y_0))$  in [12]). 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 [10]. Since  $\eta_{0,\infty}$  is non-degenerate, one can thus apply [12, Theorem 2] to analyze the extended support of the LASSO on a thin grid (see Section 5.3 below 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 5) holds. This means that Theorem 3 can be applied to analyze the extended support of C-BP on a thin grid (see Section 5.3 below for a numerical illustration). But when the Dirac masses are too close (second and third rows), one has  $\|\eta_T\|_\infty > 1$ , so that one cannot ensure the support stability of the C-BP solution with our result.

### 5.3. Extended support for deconvolution on a thin grid

We still consider the case of an ideal low pass filter. Figure 5 displays the evolution, as a function of  $\lambda$  (in abscissa) of the solution  $a_\lambda$  of the LASSO (Eq. (7)) and of the solution  $(a_\lambda, b_\lambda)$  of the C-BP  $(\mathcal{Q}_\lambda^n(y))$ , on a thin grid. 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). To simplify the interpretation, we set the Dirac masses on the grid ( $x_{0,\nu} = \bar{x}_\nu$ ) and we do not add noise ( $w = 0$ ). Each 1-D curve (either plain or dashed) represents the evolution of a single coefficient, e.g.  $(a_\lambda)_i$ , for some index  $i$  (only non-zero coefficients are displayed).

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 pre-certificate  $\eta_{V,\infty}$  (in magenta, see Definition 2) and  $\eta_T$  (in green, see Definition 4). This shows graphically that these two precertificates are non-degenerate (in the sense of [12, Proposition 3] and Definition 5) so that the conclusions of [12, Theorem 2] and Theorem 3 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 [12, Corollary 1], 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

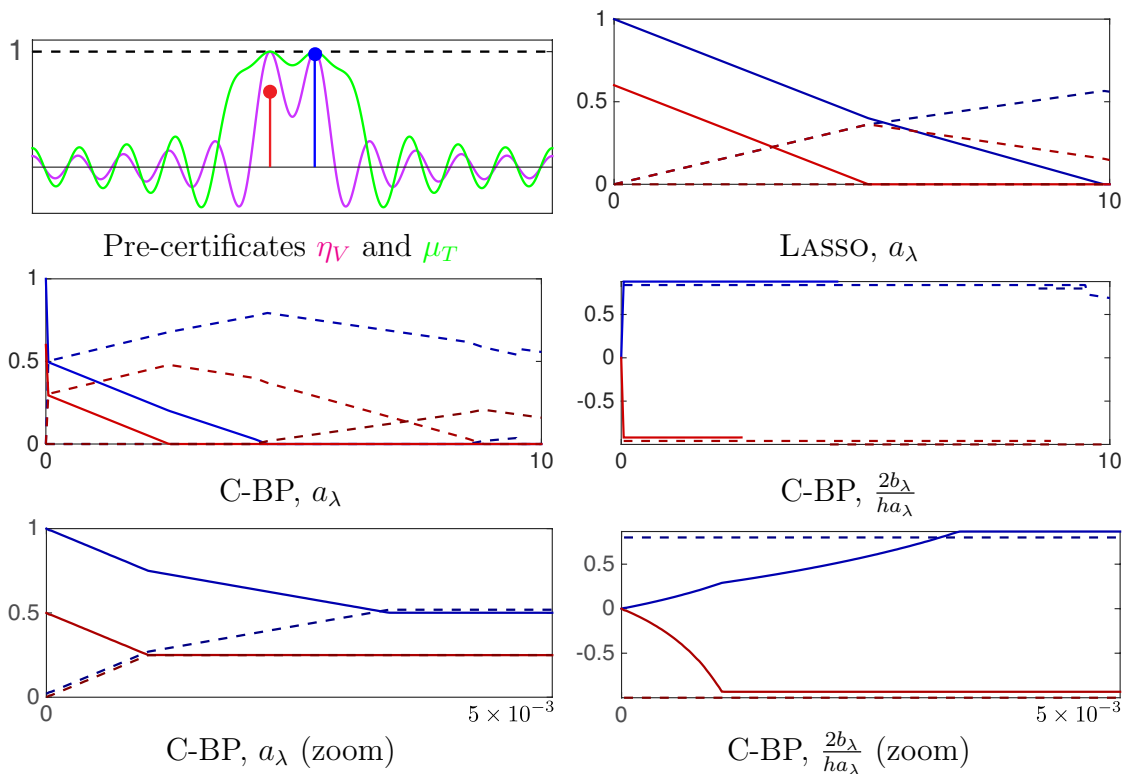


Figure 5: Display of the evolution as a function of  $\lambda$  of the solutions of the LASSO and C-BP problems. Note that dashed curves have been (artificially) slightly shifted to avoid that they overlap with the plain curve.

to the support  $I$  and the dashed curve corresponds to  $J \setminus I$ ). Also, as predicted by [12, 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 Corollary 1 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 3, the support of the solutions (which correspond to the extended support  $J$  described in Theorem 3) 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 5 and Remark 3 which predict that, in the case where  $\eta_{\lambda, \infty}^{(3)}(x_{\lambda, \nu}) \neq 0$ , the C-BP yields either one spike or a pair of spikes with the same shift (the latter case is in fact overwhelming).

## Conclusion

In this paper, we have provided a detailed analysis of the support recovery properties of the C-BP method. We have highlighted the existence of three different regimes. In the first one, the model identifies the spikes locations up to a very small error. This is the ideal situation. Unfortunately, for typical unknown signals, this regime breaks down as the grid size becomes small, and all the sooner as the measure has dense clusters. In the second one, for each unknown spike, the model yields two spikes saturating at consecutive half-grid points: it holds when  $\lambda > 0$  is fixed and the grid is small. In the last one, the model identifies one free spike and an additional one which saturates a half grid neighbor: it holds when the grid size and  $\lambda > 0$  are small, and the unknown spikes are in a neighborhood of the grid points.

These results explain how the C-BP yields better recovery performance than the LASSO for isolated spikes falling off-the-grid. They also explain that a necessary condition for this gain is the use of grids that are not so small. Finding precise conditions ensuring the existence of the ideal regime is an interesting avenue for future works.

## 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. 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. The proof of the following lemma can be found in [12].

**Lemma 1.** *Let  $A: \mathbb{R}^N \rightarrow \mathcal{H}$ ,  $B: \mathbb{R}^N \rightarrow \mathbb{R}^N$  be linear operators such that  $A$  has full rank and  $B$  is invertible. Then  $(AB)^+ = B^{-1}A^+$ .*

**Lemma 2.** *Let  $A, B, C, C_h: \mathbb{R}^N \rightarrow \mathcal{H}$  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  $\Gamma \stackrel{\text{def.}}{=} \begin{pmatrix} A & B \end{pmatrix}$ ,  $\Pi_{(\text{Im}\Gamma)^\perp}$  be the orthogonal projector onto  $(\text{Im}\Gamma)^\perp$ , and let*

$$M_h \stackrel{\text{def.}}{=} \begin{pmatrix} A + \frac{h}{2}B & A - \frac{h}{2}B & A + \frac{h}{2}B \text{diag}(\varepsilon) + h^3 C_h \text{diag}(\varepsilon) \end{pmatrix} \quad (\text{A.1})$$

Then for  $h > 0$  small enough,  $M_h$  has full rank and

$$(M_h^* M_h)^{-1} \begin{pmatrix} \mathbf{1}_N \\ \mathbf{1}_N \\ \mathbf{1}_N \end{pmatrix} = -\frac{1}{h^3} \begin{pmatrix} \text{diag}(-\frac{1+\varepsilon}{2}) \\ \text{diag}(\frac{1-\varepsilon}{2}) \\ \text{diag}(\varepsilon) \end{pmatrix} (C^* \Pi_{(\text{Im } \Gamma)^\perp} C)^{-1} C^* \Gamma^{+,*} \begin{pmatrix} \mathbf{1}_N \\ 0 \\ 0 \end{pmatrix} + o\left(\frac{1}{h^3}\right), \quad (\text{A.2})$$

$$M_h^+ = \frac{1}{h^3} \begin{pmatrix} \text{diag}(-\frac{1+\varepsilon}{2}) \\ \text{diag}(\frac{1-\varepsilon}{2}) \\ \text{diag}(\varepsilon) \end{pmatrix} (\Pi_{(\text{Im } \Gamma)^\perp} C_h)^+ + o\left(\frac{1}{h^3}\right), \quad (\text{A.3})$$

$$\text{but } M_h^{+,*} \begin{pmatrix} \mathbf{1}_N \\ \mathbf{1}_N \\ \mathbf{1}_N \end{pmatrix} = \Gamma^{+,*} \begin{pmatrix} \mathbf{1}_N \\ 0_N \\ 0_N \end{pmatrix} - (\Pi_{(\text{Im } \Gamma)^\perp} C_h)^{+,*} C_h^* \Gamma^{+,*} \begin{pmatrix} \mathbf{1}_N \\ 0_N \\ 0_N \end{pmatrix}. \quad (\text{A.4})$$

*Proof.* Observe that

$$M_h = \begin{pmatrix} \Gamma & C_h \end{pmatrix} Z, \quad \text{where } Z \stackrel{\text{def.}}{=} \begin{pmatrix} I_N & I_N & I_N \\ \frac{h}{2} I_N & -\frac{h}{2} I_N & \frac{h}{2} \text{diag}(\varepsilon) \\ 0 & 0 & h^3 \text{diag}(\varepsilon) \end{pmatrix}.$$

We note that

$$Z^{-1} = \begin{pmatrix} \frac{1}{2} I_N & \frac{1}{h} I_N & \frac{1}{h^3} \text{diag}(-\frac{1+\varepsilon}{2}) \\ \frac{1}{2} I_N & -\frac{1}{h} I_N & \frac{1}{h^3} \text{diag}(\frac{1-\varepsilon}{2}) \\ 0 & 0 & \frac{1}{h^3} \text{diag}(\varepsilon) \end{pmatrix}.$$

Moreover, writing  $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \stackrel{\text{def.}}{=} \begin{pmatrix} \Gamma^* \Gamma & \Gamma^* C_h \\ C_h^* \Gamma & C_h^* C_h \end{pmatrix}$ , we see from the full rank assumption that  $a, T \stackrel{\text{def.}}{=} a - b d^{-1} c$  and  $S \stackrel{\text{def.}}{=} d - c a^{-1} b$  are invertible, at least for  $h > 0$  small enough. The block inversion formula yields (see [18, Section 0.7.3])

$$\begin{pmatrix} \Gamma^* \Gamma & \Gamma^* C_h \\ C_h^* \Gamma & C_h^* C_h \end{pmatrix}^{-1} = \begin{pmatrix} T^{-1} & -a^{-1} b S^{-1} \\ -S^{-1} c a^{-1} & S^{-1} \end{pmatrix} = \begin{pmatrix} T^{-1} & -T^{-1} b d^{-1} \\ -d^{-1} c T^{-1} & S^{-1} \end{pmatrix}$$

with, from straightforward formulations,

$$\begin{aligned} T &= \Gamma^* \Pi_{(\text{Im } C_h)^\perp} \Gamma \\ S &= C_h^* \Pi_{(\text{Im } \Gamma)^\perp} C_h \\ a^{-1} b S^{-1} &= T^{-1} b d^{-1} = \Gamma^+ C_h (C_h^* \Pi_{(\text{Im } \Gamma)^\perp} C_h)^{-1} = (\Gamma^* \Pi_{(\text{Im } C_h)^\perp} \Gamma)^{-1} \Gamma C_h^+ \\ S^{-1} c a^{-1} &= d^{-1} c T^{-1} = (C_h^* \Pi_{(\text{Im } \Gamma)^\perp} C_h)^{-1} C_h \Gamma^+ = C_h^+ \Gamma (\Gamma^* \Pi_{(\text{Im } C_h)^\perp} \Gamma)^{-1} \end{aligned}$$

where  $\Gamma^+ = (\Gamma^* \Gamma)^{-1} \Gamma$  is the pseudo-inverse of  $\Gamma$  and  $\Pi_{(\text{Im } \Gamma)^\perp} = \text{Id}_{\mathcal{H}} - \Gamma (\Gamma^* \Gamma)^{-1} \Gamma^*$  is the orthogonal projector onto  $(\text{Im } \Gamma)^\perp$ , and similarly for  $C_h$ . As a result,

$$(M_h^* M_h)^{-1} = Z^{-1} \begin{pmatrix} \Gamma^* \Gamma & \Gamma^* C_h \\ C_h^* \Gamma & C_h^* C_h \end{pmatrix}^{-1} Z^{*, -1},$$

Since  $Z^{*, -1} \begin{pmatrix} \mathbf{1}_N \\ \mathbf{1}_N \\ \mathbf{1}_N \end{pmatrix} = \begin{pmatrix} \mathbf{1}_N \\ \mathbf{0}_N \\ \mathbf{0}_N \end{pmatrix}$ , we see that

$$(M_h^* M_h)^{-1} \begin{pmatrix} \mathbf{1}_N \\ \mathbf{1}_N \\ \mathbf{1}_N \end{pmatrix} = Z^{-1} \begin{pmatrix} T^{-1} & -a^{-1}bS^{-1} \\ -S^{-1}ca^{-1} & S^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{1}_N \\ \mathbf{0}_N \\ \mathbf{0}_N \end{pmatrix} \quad (\text{A.5})$$

$$= \frac{1}{h^3} \begin{pmatrix} \text{diag} \left( -\frac{1+\varepsilon}{2} \right) \\ \text{diag} \left( \frac{1-\varepsilon}{2} \right) \\ \text{diag}(\varepsilon) \end{pmatrix} (-S^{-1}ca^{-1}) \begin{pmatrix} \mathbf{1}_N \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \frac{1}{2}\text{Id}_N & \frac{1}{h}\text{Id}_N \\ \frac{1}{2}\text{Id}_N & -\frac{1}{h}\text{Id}_N \\ 0 & 0 \end{pmatrix} T^{-1} \begin{pmatrix} \mathbf{1}_N \\ \mathbf{0} \end{pmatrix} \quad (\text{A.6})$$

which yields (A.2).

Moreover, the formula for the Moore-Penrose pseudo-inverse [2, Theorem 1] yields

$$\begin{pmatrix} \Gamma & C_h \end{pmatrix}^+ = \begin{pmatrix} (\Pi_{(\text{Im } C_h)^\perp} \Gamma)^+ \\ (\Pi_{(\text{Im } \Gamma)^\perp} C_h)^+ \end{pmatrix}.$$

From Lemma 1, we deduce that

$$M_h^+ = Z^{-1} \begin{pmatrix} (\Pi_{(\text{Im } C_h)^\perp} \Gamma)^+ \\ (\Pi_{(\text{Im } \Gamma)^\perp} C_h)^+ \end{pmatrix} \quad (\text{A.7})$$

$$= \frac{1}{h^3} \begin{pmatrix} \text{diag} \left( -\frac{1+\varepsilon}{2} \right) \\ \text{diag} \left( \frac{1-\varepsilon}{2} \right) \\ \text{diag}(\varepsilon) \end{pmatrix} (\Pi_{(\text{Im } \Gamma)^\perp} C_h)^+ + \begin{pmatrix} \frac{1}{2}\text{Id}_N & \frac{1}{h}\text{Id}_N \\ \frac{1}{2}\text{Id}_N & -\frac{1}{h}\text{Id}_N \\ 0 & 0 \end{pmatrix} (\Pi_{(\text{Im } C_h)^\perp} \Gamma)^+ \quad (\text{A.8})$$

Considering its adjoint, we obtain

$$M_h^{+,*} \begin{pmatrix} \mathbf{1}_N \\ \mathbf{1}_N \\ \mathbf{1}_N \end{pmatrix} = 0 + (\Pi_{(\text{Im } C_h)^\perp} \Gamma)^+ \begin{pmatrix} \mathbf{1}_N \\ \mathbf{0}_N \end{pmatrix}.$$

□

## Appendix B. Proofs for Section 2

The proof of Theorem 1 is similar to the proofs of Theorem 3 and Corollary 1 which are detailed below, therefore we omit it. Let us mention however that it relies on the following fundamental Lemma (which is a simplified version of [12, Lemma 4]), which characterizes  $\bar{\eta}_0$ .

**Lemma 3.** *Let  $J^{(\mathbf{r})}, J^{(1)} \subset \llbracket 0, G-1 \rrbracket$ , and  $(a_0, b_0) \in \mathcal{C}_h^G$ . Assume that  $(I^{(\mathbf{r})}, I^{(1)}) \stackrel{\text{def.}}{=} (I^{(\mathbf{r})}(a_0, b_0), I^{(1)}(a_0, b_0))$  is such that*

$$I^{(\mathbf{r})} \subset J^{(\mathbf{r})}, \quad I^{(1)} \subset J^{(1)} \quad \text{and} \quad \hat{\mathcal{A}}_h \stackrel{\text{def.}}{=} \left( (A + \frac{h}{2}B)_{J^{(\mathbf{r})}} \quad (A - \frac{h}{2}B)_{J^{(1)}} \right)$$

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

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

$$\bar{\eta}_0 = -\mathcal{A}_h^* \hat{\mathcal{A}}_h (\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} s.$$

## Appendix C. Proofs for Section 3

### Appendix C.1. Proof of Theorem 2

We study when the solutions to  $(\mathcal{Q}_\lambda(y))$  have support  $(\bar{I}, \bar{I})$ . In view of Proposition 1, and assuming that  $\Gamma_{\bar{x}}$  has full rank, the vector  $(a, b)$  defined by (38) is a solution iff

- $r_i > 0$  and  $l_i > 0$  for all  $i \in \bar{I}$  (where  $(r, l)$  is defined from  $(a, b)$  in (20)),
- The vector

$$p_\lambda \stackrel{\text{def.}}{=} \frac{1}{\lambda} \left( y - \Gamma_{\bar{x}} \begin{pmatrix} a_{\bar{I}} \\ b_{\bar{I}} \end{pmatrix} \right) = \frac{1}{\lambda} \Pi_{(\text{Im } \Gamma_{\bar{x}})^\perp} y + \Gamma_{\bar{x}}^{+,*} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix}, \quad (\text{C.1})$$

where  $\Pi_{(\text{Im } \Gamma_{\bar{x}})^\perp}$  is the orthogonal projector onto  $(\text{Im } \Gamma_{\bar{x}})^\perp$ , satisfies

$$\forall \nu \in \{1, \dots, N\}, \quad (\Phi^* p_\lambda)(\bar{x}_\nu) = 1, \quad (\Phi'^* p_\lambda)(\bar{x}_\nu) = 0, \quad (\text{C.2})$$

$$\forall t \in \mathcal{G} \setminus \{\bar{x}_1, \dots, \bar{x}_N\}, \quad (\Phi^* p_\lambda)(t) \pm \frac{h}{2} (\Phi'^* p_\lambda)(t) \leq 1. \quad (\text{C.3})$$

Let us deal with the first item. Let  $\nu \in \llbracket 1, N \rrbracket$  and  $i_\nu \in \llbracket 0, G-1 \rrbracket$  such that  $i_\nu h = \bar{x}_\nu$ . From (39) we deduce that

$$a_{i_\nu} \pm \frac{2}{h} b_{i_\nu} \geq \alpha_\nu \pm \frac{2}{h} \beta_\nu - C_1 (\|\alpha\|_\infty \|s\|_\infty^2 + \|\Gamma_{\bar{x}}^* w\|_\infty + \lambda)$$

where  $C_1$  depends on the norms of  $\Phi_{\bar{x}}^*, \Phi'_{\bar{x}}{}^*$ ,  $(\Gamma_{\bar{x}}^* \Gamma_{\bar{x}})^{-1} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix}$  and  $\varphi''$ , but can be chosen independent from  $0 < h \leq 1$ . Since  $\beta_\nu = \alpha_\nu s_\nu$ , we see that  $a_{i_\nu} \pm \frac{2}{h} b_{i_\nu} > 0$  provided (40) holds.

As for the second item, we notice that (C.2) is satisfied by construction of  $p_\lambda$ . Moreover, setting  $\overline{\eta}_V \stackrel{\text{def.}}{=} \Phi^* \Gamma_{\bar{x}}^{+,*} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix}$ ,  $\xi_{s,w} \stackrel{\text{def.}}{=} \Phi^* \Pi_{(\text{Im } \Gamma_{\bar{x}})^\perp} y$ , we rewrite (C.3) as

$$\left( \overline{\eta}_V \pm \frac{h}{2} \overline{\eta}_V' \right) (t) + \frac{1}{\lambda} \left( \xi_{s,w} \pm \frac{h}{2} \xi_{s,w}' \right) (t) \leq 1.$$

Since  $\xi_{s,w} = \Phi^* \Pi_{(\text{Im } \Gamma_{\bar{x}})^\perp} (R_\alpha(s) + w)$ , there exists some constant  $\tilde{C}_2 > 0$  such that  $\max_{t \in \mathcal{G} \setminus \{\bar{x}_1, \dots, \bar{x}_N\}} \xi_{s,w}(t) \leq \tilde{C}_2 (\|\alpha\|_\infty \|s\|_\infty^2 + \|w\|_{\mathcal{H}})$ . As a result, defining

$$C_2 \stackrel{\text{def.}}{=} \frac{1}{\tilde{C}_2} \left( 1 - \max_{t \in \mathcal{G} \setminus \{\bar{x}_1, \dots, \bar{x}_N\}} \left( \overline{\eta}_V \pm \frac{h}{2} \overline{\eta}_V' \right) (t) \right) > 0, \quad (\text{C.4})$$

we see that (C.3) provided (41). The conclusion follows from Propositions 1 and 2.

### Appendix C.2. Proof of Proposition 3

By the translation invariance of the problem, it is sufficient to prove the result for  $\bar{x} = (0)$ ,  $\bar{I} = (0)$ .

By integration by parts, we observe that for a convolution filter,

$$\forall z \in \mathbb{T}, \quad \langle \varphi^{(k)}(z), \varphi^{(j)}(z) \rangle = \langle \varphi^{(k)}(0), \varphi^{(j)}(0) \rangle \quad (\text{C.5})$$

$$= \begin{cases} (-1)^{\frac{j-k}{2}} \left\| \varphi^{(\frac{j+k}{2})}(0) \right\|_{\mathcal{H}}^2 & \text{if } j+k \text{ is even,} \\ 0 & \text{otherwise.} \end{cases} \quad (\text{C.6})$$

Incidentally,  $\Gamma_{\bar{x}}$  has full rank, and  $\overline{\eta}_V(t) \stackrel{\text{def.}}{=} \Phi^* \Gamma_{\bar{x}}^{+,*} \begin{pmatrix} 1 \\ 0 \end{pmatrix} (t) = \frac{1}{\|\varphi(0)\|_{\mathcal{H}}^2} \langle \varphi(t), \varphi(0) \rangle$ . From the Cauchy-Schwarz inequality and the fact that  $\tilde{\varphi}$  has no period smaller than 1,  $\overline{\eta}_V(t) < 1$  for all  $t \in \mathbb{T} \setminus 0$ . Moreover, from (C.6) we deduce that

$$\overline{\eta}_V'(0) = \overline{\eta}_V^{(3)}(0) = 0, \quad \overline{\eta}_V^{(2)}(0) < 0 \text{ and } \overline{\eta}_V^{(4)}(0) > 0.$$

From Lemma 4 below, we deduce that there exists  $h_1 > 0$  such that for all  $h \in (0, h_1]$ ,

$$\max_{t \in \mathcal{G} \setminus \{0\}} \left( \overline{\eta}_V \pm \frac{h}{2} \overline{\eta}_V' \right) (t) < 1. \quad (\text{C.7})$$

hence, condition (IC<sub>h</sub>) holds. The claimed result follows.

The proof of Proposition 3 relies on the following Lemma which will also be helpful for the study of the low noise support in Section 4.2.

**Lemma 4.** *Let  $\{\bar{x}_1, \dots, \bar{x}_N\} \subset \mathcal{G}$ ,  $\eta \in \mathcal{C}^4(\mathbb{T})$  such that for all  $t \in \mathbb{T} \setminus \{\bar{x}_1, \dots, \bar{x}_N\}$ ,  $\eta(t) < 1$  and*

$$\forall \nu \in \{1, \dots, N\}, \quad \eta(\bar{x}_\nu) = 1, \quad \eta''(\bar{x}_\nu) \neq 0, \quad \eta^{(3)}(\bar{x}_\nu) = 0, \quad \eta^{(4)}(\bar{x}_\nu) > 0.$$



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

 Table C1: Variations of  $\eta$  and its derivatives.

Then, there exists  $h_1 > 0$  such for all stepsize  $h \in (0, h_1]$ ,

$$\max_{t \in \mathcal{G} \setminus \{\bar{x}_1, \dots, \bar{x}_N\}} \left( \eta \pm \frac{h}{2} \eta' \right) (t) < 1.$$

*Remark 6.* The condition  $\eta^{(3)}(\bar{x}_\nu) = 0$  might seem surprising, but it is in fact necessary: if it does not hold, by Lemma 5, for all  $h$  small enough there exists  $t \in \mathcal{G} \setminus \{\bar{x}_1, \dots, \bar{x}_N\}$  such that  $(\eta \pm \frac{h}{2} \eta') (t) \geq 1$ .

*Proof.* Let  $r > 0$  small enough so that the intervals  $(\bar{x}_\nu - r, \bar{x}_\nu + r)$  for  $\nu \in \{1, \dots, N\}$  are pairwise disjoint, and such that  $\eta''(t) < 0$ , and  $\eta^{(4)}(t) > 0$  for all  $t \in (\bar{x}_\nu - r, \bar{x}_\nu + r)$ .

We first prove that  $\eta(kh) + \frac{h}{2} |\eta'(kh)| < 1$  for all  $k$  such that  $kh \in (\bar{x}_\nu - r, \bar{x}_\nu + r) \setminus \{\bar{x}_\nu\}$ . To simplify the notation, we assume without loss of generality that  $\bar{x}_\nu = 0$ . The variations of  $\eta$  and its derivatives are given in Table C1.

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

$$\forall t \in (0, r), \theta'(t) = \frac{1}{2} (\eta'(t) - t\eta''(t)) = \frac{1}{2} \int_0^t \underbrace{(\eta''(u) - \eta''(t))}_{<0} du < 0. \quad (\text{C.8})$$

Hence, for all  $k$  such that  $kh \in (0, r)$ ,

$$\eta(kh) - \frac{h}{2} \eta'(kh) = \underbrace{\eta(kh) - \frac{kh}{2} \eta'(kh)}_{=\theta(kh) < \theta(0)=1} + \underbrace{\frac{(k-1)h}{2} \eta'(kh)}_{<0} < 1. \quad (\text{C.9})$$

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

$$\forall t \in (-r, 0), \theta'(t) = \frac{1}{2} (\eta'(t) - t\eta''(t)) = \frac{1}{2} \int_0^t \underbrace{(\eta''(u) - \eta''(t))}_{<0} du > 0. \quad (\text{C.10})$$

As a consequence, for all  $k$  such that  $kh \in (-r, 0)$ ,

$$\eta(kh) + \frac{h}{2}\eta'(kh) = \underbrace{\eta(kh) - \frac{kh}{2}\eta'(kh)}_{=\theta(kh) < \theta(0)=1} + \underbrace{\frac{(k+1)h}{2}\eta'(kh)}_{\leq 0} < 1. \quad (\text{C.11})$$

Thus we see that  $|\eta(kh) + \frac{h}{2}\eta'(kh)| < 1$  for all  $kh \in (-r, r) \setminus \{0\}$ , and we proceed similarly on all the intervals of the form  $(\bar{x}_\nu - r, \bar{x}_\nu + r)$ .

Now, by a compactness argument, there exists a constant  $\beta < 1$  such that  $\eta(t) \leq \beta$  for all  $t \in \mathbb{T} \setminus \bigcup_{\nu=1}^N (\bar{x}_\nu - r, \bar{x}_\nu + r)$ . For  $h > 0$  small enough, the inequality  $\frac{h}{2}(\sup_{t \in \mathbb{T}} |\eta'(t)|) < 1 - \beta$  holds, and we conclude that  $|\eta(kh) + \frac{h}{2}\eta'(kh)| < 1$  for all  $t \in \mathbb{T} \setminus \bigcup_{\nu=1}^N (\bar{x}_\nu - r, \bar{x}_\nu + r)$ .  $\square$

### Appendix C.3. Proof of Proposition 4

We observe that our embedding of the discrete problem into the continuous one, that is  $m_n = \sum_{i=0}^{G_n-1} a_i^{(n)} \delta_{ih_n + b_i/a_i}$ , yields  $\|a^{(n)}\|_1 = |m_n|(\mathbb{T}) = m_n(\mathbb{T})$ . For the liminf inequality, let  $(m_n)_{n \in \mathbb{N}}$  be of the form (46) which weakly\* converges towards  $m$ . We notice that  $\Phi'_{\mathcal{G}_n} b^{(n)} = \Phi'(\sum_{i=0}^{G_n-1} b_i^{(n)} \delta_{ih_n})$ , where  $\Phi' : m \mapsto \int_{\mathbb{T}} \varphi'(x) dm(x)$  is continuous from  $\mathcal{M}(\mathbb{T})$  to  $\mathcal{H}$  (in the strong topologies). Moreover,

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

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

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

where  $\omega_\psi : t \mapsto \sup_{|x'-x| \leq t} |\psi(x) - \psi(x')|$  is the modulus of continuity of  $\psi$ . As a result,  $\sum_{i=0}^{G_n-1} a_i^{(n)} \delta_{ih_n} - m_n \xrightarrow{*} 0$  so that  $\sum_{i=0}^{G_n-1} a_i^{(n)} \delta_{ih_n}$  weakly\* converges to  $m$ . Hence,  $\Phi_{\mathcal{G}_n} a^{(n)}$  weakly converges towards  $\Phi m$  in  $\mathcal{H}$ . 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 by invoking the lower semi-continuity of both terms:

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

For the limsup inequality, we build a recovery sequence  $m_n$  by choosing  $a_k^{(n)} = m([kh_n, (k+1)h_n])$  and  $b_k^{(n)} = 0$  for all  $k \in \llbracket 0, G_n - 1 \rrbracket$ . Then, for all  $n$ ,  $\|a^{(n)}\|_1 = m(\mathbb{T})$

and

$$\begin{aligned} \left\| \Phi \left( \sum_{i=0}^{G_n-1} a_i^{(n)} \delta_{ih_n} \right) - \Phi m \right\|_{\mathcal{H}} &\leq \sum_{i=0}^{G_n-1} \left\| \int_{[ih_n, (i+1)h_n)} (\varphi(t) - \varphi(ih_n)) dm(t) \right\|_{\mathcal{H}} \\ &\leq m(\mathbb{T}) \omega_\varphi(h_n) \rightarrow 0 \end{aligned}$$

since  $\varphi$  is uniformly continuous on  $\mathbb{T}$ . As a result,  $\Phi m_n - y$  strongly converges towards  $\Phi m - y$  and the limsup inequality is proved.

#### Appendix D. Proofs of Section 4

The following lemma is central in our analysis. It studies sequences of dual certificates  $(\eta_n)_{n \in \mathbb{N}}$ . For  $0 < r < \frac{1}{2} \min_{\nu \neq \nu'} |x_\nu - x_{\nu'}|$ ,  $\nu \in \{1, \dots, N\}$ , it will be useful to consider the sets:

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

**Lemma 5.** *Let  $(x_1, \dots, x_N) \in \mathbb{T}^N$  pairwise distinct, and let  $\{\eta_n\}_{n \in \mathbb{N}} \in (\mathcal{C}^3(\mathbb{T}))^{\mathbb{N}}$  be a sequence of functions which converges uniformly towards some  $\eta_\infty$  (and similarly for the derivatives) such that for all  $\nu \in \{1, \dots, N\}$ ,  $\eta_\infty(x_\nu) = 1$  and for all  $t \in \mathbb{T} \setminus \{x_1, \dots, x_N\}$ ,  $\eta_\infty(t) < 1$ .*

(i) *Then*

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

*In particular for  $r > 0$  small enough, there exists  $n_0 \in \mathbb{N}$  such that for  $n \geq n_0$*

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

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

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

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

(iii) *If  $\eta_\infty^{(3)}(x_\nu) \neq 0$ , then there exists  $n_0 \in \mathbb{N}$  such that for each  $n \geq n_0$ ,  $S_{n,\nu}^{(\mathbf{r})}(r) = \emptyset$  or  $S_{n,\nu}^{(\mathbf{l})}(r) = \emptyset$ .*

(iv) *If  $\eta_\infty^{(4)}(x_\nu) \neq 0$ , 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  $S_{n,\nu}^{(\mathbf{l})}(r) = \{i_n h_n, (i_n + 1)h_n\}$  (with the same  $i_n \in \llbracket 0, G_n - 1 \rrbracket$ ) is finite.*

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

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

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

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

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

where  $\alpha_k$  is defined in (D.6). Now, let  $n \rightarrow +\infty$ . By (D.1),  $ih_n \rightarrow x_\nu$  and  $jh_n \rightarrow x_\nu$ , and using the uniform convergence of  $\eta_n^{(k)}$  towards  $\eta_\infty^{(k)}$ , dividing by  $h_n^3$ , we obtain respectively  $0 \geq -\eta_\infty^{(3)}(x_\nu) \times \frac{1}{12}$  and  $0 \geq \eta_\infty^{(3)}(x_\nu) \times \frac{1}{12}$ , thus  $\eta_\infty^{(3)}(x_\nu) = 0$ .

(iv) Assume, by contradiction, that the mentioned set is infinite. For such  $n$ , a Taylor

expansion at  $ih_n$  yields (we write  $i$  for  $i_n$ ):

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

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  $(\eta_\infty)^{(4)}(x_\nu) = 0$ , a contradiction.  $\square$

This other lemma focusses on the limit of the sets  $D^n$  defined in (42).

**Lemma 6.** *As  $n \rightarrow +\infty$ , the sets  $D^n$  converge towards  $D^\infty$  defined in (44) (in the sense of set convergence).*

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

$$\begin{aligned}
 E^n &\stackrel{\text{def.}}{=} \left\{ p \in \mathcal{H} ; \max_{t \in \mathbb{T}} (\Phi^* p)(t) + \frac{h_n}{2} |(\Phi^* p)'(t)| \leq 1 \right\}, \\
 F^n &\stackrel{\text{def.}}{=} \left\{ p \in \mathcal{H} ; \max_{k \in [0, G_n - 1]} \Phi^* p(kh_n) \leq 1 \right\}
 \end{aligned}$$

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 (see [19, Ex. 4.3]).  $\square$

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

$$D^n \stackrel{\text{def.}}{=} \left\{ p \in \mathcal{H} ; (\Phi^* p)(ih_n) + \frac{h_n}{2} |(\Phi^* p)'(ih_n)| \leq 1 \right\}.$$

Since the set convergence of  $D^n$  (see Lemma 6 above) implies the convergence of the projections onto  $D^n$  (see [19], or [11] for a direct proof in a similar context), we obtain:

**Lemma 7.** *Let  $p_{\lambda,n}$  (resp.  $p_{\lambda,\infty}$ ) be a solution of  $(\mathcal{E}_\lambda^n(y))$  (resp.  $(\mathcal{E}_\lambda^\infty(y))$ ), and  $\eta_{\lambda,n} = \Phi^* p_{\lambda,n}$  (resp.  $\eta_{\lambda,\infty} = \Phi^* p_{\lambda,\infty}$ ). Then*

$$\lim_{n \rightarrow +\infty} p_{\lambda,n} = p_{\lambda,\infty} \text{ strongly in } \mathcal{H},$$

$$\lim_{n \rightarrow +\infty} \eta_{\lambda,n}^{(k)} = \eta_{\lambda,\infty}^{(k)} \text{ in the sense of the uniform convergence, for all } k \in \mathbb{N} \text{ up to the regularity of } \varphi.$$

We are now in position to prove Proposition 5.

*Proof of Proposition 5.* By Lemma 7, we know that the dual certificates  $\eta_{\lambda,n}$  converge towards  $\eta_{\lambda,\infty}$ . By Lemma 5 (i) 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. 4) which ensures that  $m_{\lambda,n}$  converges towards  $m_{\lambda,\infty}$  in 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 5 ensures that the sets  $S_{n,\nu}^{(\mathbf{r})}(r)$  and  $S_{n,\nu}^{(\mathbf{l})}(r)$  are of the form  $\emptyset$ ,  $\{ih_n\}$ , or  $\{ih_n, (i+1)h_n\}$ . Moreover, since  $\lim_{n \rightarrow +\infty} (\text{supp } m_{\lambda,n}) = \{x_1, \dots, x_N\}$  we must have  $S_{n,\nu}^{(\mathbf{r})}(r) \neq \emptyset$  or  $S_{n,\nu}^{(\mathbf{l})}(r) \neq \emptyset$ . Using the fact that  $\max S_{n,\nu}^{(\mathbf{r})}(r) \leq \min S_{n,\nu}^{(\mathbf{l})}(r)$ , one may check that the only possible saturation points of  $\eta_{\lambda,n} + \frac{h_n}{2} \eta'_{\lambda,n}$  and  $\eta_{\lambda,n} - \frac{h_n}{2} \eta'_{\lambda,n}$  are given in Table 1. The optimality conditions of Proposition 1 imply that  $m_{\lambda,n}$  is at most a sum of Dirac masses at those locations.

If  $\eta_{\lambda,\infty}^{(3)}(x_\nu) \neq 0$ , Lemma 5 (iii) implies that for  $n$  large enough,  $S_{n,\nu}^{(\mathbf{r})}(r) = \emptyset$  or  $S_{n,\nu}^{(\mathbf{l})}(r) = \emptyset$  (but not both). Hence there are at most two (successive) saturations, produced either by  $\eta_{\lambda,n} + \frac{h_n}{2} \eta'_{\lambda,n}$  or by  $\eta_{\lambda,n} - \frac{h_n}{2} \eta'_{\lambda,n}$ .  $\square$

#### Appendix D.1. Asymptotics of the Minimal Norm Certificates

We begin with two propositions concerning minimal norm certificates. From Remark 4, given  $\bar{x} \in \mathcal{G}_0$ , if the Twice Non-Degenerate Source Condition (Definition 5) holds, then the vector  $(\bar{a}, \bar{b})$  defined by (51), with  $\bar{y} \stackrel{\text{def.}}{=} \Phi_{\mathcal{G}_n} \bar{a} + \Phi'_{\mathcal{G}_n} \bar{b}$ , is a solution to  $(\mathcal{Q}_0^n(\bar{y}))$  for all  $n$  large enough, and  $\eta_T$  is a valid dual certificate, i.e. a solution to  $(\mathcal{E}_0^n(\bar{y}))$ . In fact the associated minimal norm certificates (which thus exist) converge towards  $\eta_T$ .

**Proposition 6.** *Let  $\bar{x} \in \mathcal{G}_0$  satisfy the Twice Non-Degenerate Source Condition (and  $\eta_T$  the corresponding Third derivative (pre)certificate). Let  $p_{0,n}$  be the minimal norm solution of  $(\mathcal{E}_0^n(\bar{y}))$ , and  $\eta_{0,n} = \Phi^* p_{0,n}$ . Then,*

- $\lim_{n \rightarrow +\infty} p_{0,n} = p_T$  for the  $\mathcal{H}$  strong topology,
- $\lim_{n \rightarrow +\infty} \eta_{0,n}^{(k)} = \eta_T^{(k)}$  in the sense of the uniform convergence, for all  $k \in \mathbb{N}$  up to the regularity of  $\varphi$ .

*Proof.* As mentioned above, the Twice Non-Degenerate Source Condition and Lemma 4 imply  $\eta_T$  is a certificate for  $(\mathcal{Q}_0^n(\bar{y}))$ . As a result,  $\|p_{0,n}\| \leq \|p_T\|$  and the sequence  $(p_{0,n})_{n \in \mathbb{N}}$  is bounded in  $\mathcal{H}$ . We may extract a subsequence  $p_{0,n'}$  which weakly converges towards some  $\tilde{p} \in \mathcal{H}$ , and then  $\|\tilde{p}\| \leq \liminf_{n' \rightarrow +\infty} \|p_{0,n'}\| \leq \|p_T\|$ . Since  $\Phi^*$  and  $\Phi^{(k),*}$  are compact (see [12, Lemma 1]), we obtain that  $\eta_{0,n'}^{(k)} = (\Phi^* p_{0,n'})^{(k)}$  converges toward  $\tilde{\eta}^{(k)} \stackrel{\text{def.}}{=} (\Phi^* \tilde{p})^{(k)}$  in the (strong) topology of uniform convergence. We immediately obtain that  $\tilde{\eta}(t) \leq 1$  for all  $t \in \mathbb{T}$ , and  $\tilde{\eta}(\bar{x}_\nu) = 1$ ,  $\tilde{\eta}(\bar{x}_\nu) = 0$  for all  $\nu \in \{1, \dots, N\}$ .

Moreover, applying Lemma 5 to  $\Phi^* p_{0,n}$  (observing that  $x_\nu \in S_{n,\nu}^{(\mathbf{r})}(r) \cap S_{n,\nu}^{(1)}(r)$ ), we get  $\tilde{\eta}^{(3)}(\bar{x}_\nu) = 0$ . As a result,  $\tilde{p}$  is admissible for (49), hence  $\|p_T\| \leq \|\tilde{p}\|$ . Thus in fact  $\|p_T\| = \|\tilde{p}\|$  and  $p_T = \tilde{p}$ . 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  $\mathcal{H}$  since  $\lim_{n \rightarrow +\infty} \|p_{0,n}\| = \|p_T\|$ .  $\square$

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

**Proposition 7.** *Let  $\bar{x} \in \mathcal{G}_0$ , such that the Twice Non Degenerate Source Condition holds. Then, for  $n$  large enough, any vector  $(\bar{a}, \bar{b}) \in \mathcal{C}_{h_n}$  defined by (51), with  $\bar{y} \stackrel{\text{def.}}{=} \Phi_{\mathcal{G}_n} \bar{a} + \Phi'_{\mathcal{G}_n} \bar{b}$ , is a solution to  $(\mathcal{Q}_0^n(\bar{y}))$  and its extended support is given by:*

$$\text{ext}_n^{(\mathbf{r})}(\bar{a}, \bar{b}) = \bigcup_{\nu=1}^N S_{n,\nu}^{(\mathbf{r})}(r), \quad \text{and} \quad \text{ext}_n^{(1)}(\bar{m}) = \bigcup_{\nu=1}^N S_{n,\nu}^{(1)}(r), \quad (\text{D.2})$$

where

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

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

*Proof.* By Remark 4  $\eta_T$  is a solution to  $(\mathcal{E}_0^n(\bar{y}))$  and  $(\bar{a}, \bar{b})$  is a solution to  $(\mathcal{Q}_0^n(\bar{y}))$ . Applying Lemma 5 to  $\eta_{0,n}$ ,  $\eta_T$ , 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  $\eta_{0,n}$  (solution of  $(\mathcal{E}_0^n(\bar{y}))$ ) and  $(\bar{a}, \bar{b})$  (solution of  $(\mathcal{Q}_0^n(\bar{y}))$ ),  $\bar{x}_\nu \in S_{n,\nu}^{(\mathbf{r})}(r)$  and  $\bar{x}_\nu \in S_{n,\nu}^{(1)}(r)$ . As a consequence  $S_{n,\nu}^{(\mathbf{r})}(r)$  is equal to  $\{\bar{x}_\nu\}$  or  $\{\bar{x}_\nu - h_n, \bar{x}_\nu\}$ , and  $S_{n,\nu}^{(1)}(r)$  is equal to  $\{\bar{x}_\nu\}$  or  $\{\bar{x}_\nu, \bar{x}_\nu + h_n\}$ .

Now, since  $\eta_T^4(0) \neq 0$ , the fourth point of Lemma 5 ensures that for  $n$  large enough, one cannot have simultaneously  $S_{n,\nu}^{(\mathbf{r})}(r) = \{\bar{x}_\nu - h_n, \bar{x}_\nu\}$  and  $S_{n,\nu}^{(1)}(r) = \{\bar{x}_\nu, \bar{x}_\nu + h_n\}$ .  $\square$

*Remark 7.* As Proposition 7 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  $-h_n/2$ .

## Appendix D.2. Proof of Theorem 3

We proceed by building a good candidate for  $\eta_{0,n}$ , making the ansatz that for all  $\nu \in \{1, \dots, N\}$ , its saturation points satisfy

$$\text{if } \rho_\nu > 0, \text{ then } S_{n,\nu}^{(r)}(r) = \{\bar{x}_\nu - h_n, \bar{x}_\nu\}, \quad \text{and} \quad S_{n,\nu}^{(l)}(r) = \{\bar{x}_\nu\}, \quad (\text{D.3})$$

$$\text{if } \rho_\nu < 0, \text{ then } S_{n,\nu}^{(r)}(r) = \{\bar{x}_\nu\}, \quad \text{and} \quad S_{n,\nu}^{(l)}(r) = \{\bar{x}_\nu, \bar{x}_\nu + h_n\}, \quad (\text{D.4})$$

and then, using Lemma 3, we prove that this candidate is indeed the minimal norm certificate.

To comply with the notations of Lemma 3, let us write  $I^{(r)} \stackrel{\text{def.}}{=} I^{(l)} \stackrel{\text{def.}}{=} I \stackrel{\text{def.}}{=} \{i \in \llbracket 0, G_n - 1 \rrbracket ; \bar{a}_i \neq 0\}$ .

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

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

In this particular case where  $I^{(r)} = I^{(l)} = I$ , all  $j$  in  $(J^{(r)} \setminus I) \cup (J^{(l)} \setminus I)$  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  $\begin{pmatrix} u_{J^{(r)}} \\ v_{J^{(l)}} \end{pmatrix} = -(\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} \mathbf{1}_{3N}$  into

$$\begin{pmatrix} u_I \\ v_I \\ \tilde{t}_I \end{pmatrix} = -(\bar{\mathcal{A}}_h^* \bar{\mathcal{A}}_h)^{-1} \begin{pmatrix} \mathbf{1}_N \\ \mathbf{1}_N \\ \mathbf{1}_N \end{pmatrix}, \quad (\text{D.5})$$

where  $\bar{\mathcal{A}}_h \stackrel{\text{def.}}{=} \begin{pmatrix} A_I + \frac{h_n}{2}B_I & A_I - \frac{h_n}{2}B_I & A_{I+\varepsilon} - \frac{h_n}{2}B_{I+\varepsilon} \text{diag}(\varepsilon) \end{pmatrix}$  and  $\tilde{t}_i > 0$  for all  $i \in I$  (more precisely  $\tilde{t}_i = u_{i-1}$  for  $\varepsilon_i = -1$ ,  $\tilde{t}_i = v_{i+1}$  for  $\varepsilon_i = 1$ ). But a Taylor expansion yields

$$A_{I+\varepsilon} - \frac{h_n}{2}B_{I+\varepsilon} \text{diag}(\varepsilon) = \underbrace{\Phi_{\bar{x}}}_{=A_I} + \frac{h_n}{2} \underbrace{\Phi'_x}_{=B_I} \text{diag}(\varepsilon) + (h_n)^3 (\gamma_3 \Phi_{\bar{x}}^{(3)} \text{diag}(\varepsilon) + o(1)),$$

where we defined

$$\gamma_k \stackrel{\text{def.}}{=} \frac{1}{k!} - \frac{1}{(k-1)! \times 2}. \quad (\text{D.6})$$

Hence, we may apply Lemma 2 to  $\Phi_{\bar{x}}$ ,  $\Phi'_x$  and  $(\gamma_3 \Phi_{\bar{x}}^{(3)} + o(1))$  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 (recall that  $\gamma_3 < 0$ ).

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

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

Let us write  $\tilde{p}_n \stackrel{\text{def.}}{=} \hat{\mathcal{A}}_h \begin{pmatrix} u_{J^{(r)}} \\ v_{J^{(l)}} \end{pmatrix}$ . Since  $\begin{pmatrix} u_{J^{(r)}} \\ v_{J^{(l)}} \end{pmatrix} = -(\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} \mathbb{1}_{3N}$ , we get  $\tilde{p}_n = \hat{\mathcal{A}}_h^{+,*} \mathbb{1}_{3N} = \bar{\mathcal{A}}_h^{+,*} \mathbb{1}_{3N}$ , and applying Lemma 2 again, we see that  $\tilde{p}_n$  converges towards  $p_T$  (using (50)).

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

$$\begin{aligned} \forall j \in J^{(r)} \setminus I, \quad (\Phi^* \tilde{p}_n + \frac{h_n}{2} (\Phi^* \tilde{p}_n)')(jh_n) &= 1, \\ \text{and } \forall j \in J^{(l)} \setminus I, \quad (\Phi^* \tilde{p}_n - \frac{h_n}{2} (\Phi^* \tilde{p}_n)')(jh_n) &= 1, \end{aligned} \quad (\text{D.7})$$

which may be summarized as

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

Arguing as in the proof of point (iv) in Lemma 5 (replacing “ $1 = \dots$ ” with “ $1 \geq \dots$ ” and using that  $\eta_T^{(4)}(\bar{x}_\nu) > 0$ ), we may prove that for  $n$  large enough,  $(\Phi^* \tilde{p}_n + \varepsilon_i \frac{h_n}{2} (\Phi^* \tilde{p}_n)')((i - \varepsilon_i)h_n) < 1$ .

Then, by the same argument of compactness and local concavity as in point (ii) of Lemma 5, we observe that

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

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

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

Hence, by Lemma 3,  $\Phi^* \tilde{p}_n$  is the minimal norm certificate  $\eta_{0,n}$  and  $(J^{(r)}h_n, J^{(l)}h_n)$  is the extended support. This concludes the proof.

### Appendix D.3. Proof of Corollary 1

We build a pair of candidate solutions for the primal and dual problems, and we prove that they satisfy the optimality conditions. Again, we change variables so as to deal with the positive LASSO, working with  $(r_\lambda, l_\lambda)$  rather than  $(a_\lambda, b_\lambda)$  (see (20)). In particular, we define  $(\bar{r}, \bar{l})$  from  $(\bar{a}, \bar{b})$  using (20).

Moreover, as the noise  $w$  and the remainder of the Taylor expansion  $R_\alpha(s)$  play the same role, we set  $\tilde{w} \stackrel{\text{def.}}{=} R_\alpha(s) + w$ , to simplify the notation. In the rest of the proof, we assume that there exists  $\kappa \in (0, 1)$  such that  $\|s\|_\infty \leq \kappa \frac{h_n}{2}$ : as we shall see, it is not restrictive since the constraints on the ‘‘noise’’  $\tilde{w}$  impose  $s \ll h_n$ .

Recall the notation  $A \stackrel{\text{def.}}{=} \Phi_{\mathcal{G}_n}$ ,  $B \stackrel{\text{def.}}{=} \Phi'_{\mathcal{G}_n}$  and  $\mathcal{A}_h \stackrel{\text{def.}}{=} \begin{pmatrix} A + \frac{h}{2}B & A - \frac{h}{2}B \end{pmatrix}$ . Defining  $(J^{(\mathbf{r})}, J^{(1)})$  as in the proof of Theorem 3, we let  $\hat{\mathcal{A}}_h \stackrel{\text{def.}}{=} \begin{pmatrix} (A + \frac{h}{2}B)_{J^{(\mathbf{r})}} & (A - \frac{h}{2}B)_{J^{(1)}} \end{pmatrix}$  (again,  $\hat{\mathcal{A}}_h$  has full rank for  $h$  small enough since  $\begin{pmatrix} \Phi_{\bar{x}} & \Phi'_{\bar{x}} & \Phi_{\bar{x}}^{(3)} \end{pmatrix}$  has full rank).

Now, we let

$$\begin{cases} \begin{pmatrix} (r_\lambda)_{J^{(\mathbf{r})}} \\ (l_\lambda)_{J^{(1)}} \end{pmatrix} \stackrel{\text{def.}}{=} \hat{\mathcal{A}}_h^+ y - \lambda (\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} \mathbf{1}_{3N} = \begin{pmatrix} \bar{r}_{J^{(\mathbf{r})}} \\ \bar{l}_{J^{(1)}} \end{pmatrix} + \hat{\mathcal{A}}_h^+ \tilde{w} + \lambda \begin{pmatrix} u_{J^{(\mathbf{r})}} \\ v_{J^{(1)}} \end{pmatrix}, \\ p_\lambda \stackrel{\text{def.}}{=} \frac{1}{\lambda} \left( y - \hat{\mathcal{A}}_h \begin{pmatrix} (r_\lambda)_{J^{(\mathbf{r})}} \\ (l_\lambda)_{J^{(1)}} \end{pmatrix} \right) = \frac{1}{\lambda} \Pi_{(\text{Im } \hat{\mathcal{A}}_h)^\perp} y + \hat{\mathcal{A}}_h^{+,*} \mathbf{1}_{3N}, \end{cases} \quad (\text{D.8})$$

where  $u, v$  are defined as in the proof of Theorem 3 by  $\begin{pmatrix} u_{J^{(\mathbf{r})}} \\ v_{J^{(1)}} \end{pmatrix} \stackrel{\text{def.}}{=} -(\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} \begin{pmatrix} \mathbf{1}_{J^{(\mathbf{r})}} \\ \mathbf{1}_{J^{(1)}} \end{pmatrix} = -(\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} \mathbf{1}_{3N}$ .

Observe that by construction of  $p_\lambda$ ,  $\begin{pmatrix} A + \frac{h}{2}B & A - \frac{h}{2}B \end{pmatrix}^* p_\lambda = \hat{\mathcal{A}}_h^* p_\lambda = \mathbf{1}_{3N}$ . It remains to prove that

- (i) For all  $i \in I$ ,  $r_{\lambda,i} \geq 0$  and  $l_{\lambda,i} \geq 0$ .
- (ii) For all  $j \in J^{(\mathbf{r})} \setminus I$  (resp.  $j \in J^{(1)} \setminus I$ ),  $r_{\lambda,j} \geq 0$  (resp.  $l_{\lambda,j} \geq 0$ ).
- (iii) For  $k \notin J^{(\mathbf{r})}$  (resp.  $k \notin J^{(1)}$ )  $((A^* + \frac{h_n}{2}B^*)p_\lambda)_k < 1$  (resp.  $((A^* - \frac{h_n}{2}B^*)p_\lambda)_k < 1$ )

Regarding the first two points, we apply the same reordering of the columns as in the proof of Theorem 3,

$$\begin{pmatrix} (r_\lambda)_I \\ (l_\lambda)_I \\ (z_\lambda)_I \end{pmatrix} = \begin{pmatrix} \bar{r}_I \\ \bar{l}_I \\ 0_N \end{pmatrix} + \bar{\mathcal{A}}_h^+ \tilde{w} - \lambda (\bar{\mathcal{A}}_h^* \bar{\mathcal{A}}_h)^{-1} \mathbf{1}_{3N}, \quad (\text{D.9})$$

where  $(z_\lambda)_i = (r_\lambda)_{i-1}$  if  $i \in J^{(\mathbf{r})} \setminus I$ , and  $(l_\lambda)_{i+1}$  otherwise (*i.e.*  $i \in J^{(1)} \setminus I$ ). Using the same Taylor expansion (denoting  $\Phi_{\bar{x},h}^{(3)} \stackrel{\text{def.}}{=} \Phi_{\bar{x}}^{(3)} + o(1)$ ), combined with (A.8) and (A.6), we get

$$\begin{aligned} \bar{\mathcal{A}}_h^+ &= \frac{1}{\gamma_3 h^3} \begin{pmatrix} \text{diag} \left( -\frac{1+\varepsilon}{2} \right) \\ \text{diag} \left( \frac{1-\varepsilon}{2} \right) \\ \text{diag}(\varepsilon) \end{pmatrix} \left( \Pi_{(\text{Im } \Gamma_{\bar{x}})^\perp} \Phi_{\bar{x},h}^{(3)} \right)^+ + \begin{pmatrix} \frac{1}{2} \text{Id}_N & \frac{1}{h} \text{Id}_N \\ \frac{1}{2} \text{Id}_N & -\frac{1}{h} \text{Id}_N \\ 0 & 0 \end{pmatrix} \left( \Pi_{(\text{Im } \Phi_{\bar{x},h}^{(3)})^\perp} \Gamma_{\bar{x}} \right)^+ \\ \text{and } (\bar{\mathcal{A}}_h^* \bar{\mathcal{A}}_h)^{-1} \mathbf{1}_{3N} &= -\frac{1}{\gamma_3 h^3} \begin{pmatrix} \text{diag} \left( -\frac{1+\varepsilon}{2} \right) \\ \text{diag} \left( \frac{1-\varepsilon}{2} \right) \\ \text{diag}(\varepsilon) \end{pmatrix} \left( \Phi_{\bar{x},h}^{(3),*} \Pi_{(\text{Im } \Gamma_{\bar{x}})^\perp} \Phi_{\bar{x},h}^{(3)} \right)^{-1} \Phi_{\bar{x},h}^{(3)} \Gamma_{\bar{x}}^+ \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} \\ &\quad + \begin{pmatrix} \frac{1}{2} \text{Id}_N & \frac{1}{h} \text{Id}_N \\ \frac{1}{2} \text{Id}_N & -\frac{1}{h} \text{Id}_N \\ 0 & 0 \end{pmatrix} \left( \Gamma_{\bar{x}}^* \Pi_{(\text{Im } \Phi_{\bar{x},h}^{(3)})^\perp} \Gamma_{\bar{x}} \right)^{-1} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix}. \end{aligned}$$

As a result, the following componentwise inequality holds,

$$\begin{pmatrix} (r_\lambda)_I \\ (l_\lambda)_I \\ (z_\lambda)_{I+\varepsilon} \end{pmatrix} \geq \begin{pmatrix} \bar{r}_I \\ \bar{l}_I \\ 0_N \end{pmatrix} - \|\hat{\mathcal{A}}_h^+\|_{\infty, \mathcal{H}} \|\tilde{w}\|_{\mathcal{H}} - \lambda \|(\bar{\mathcal{A}}_h^* \bar{\mathcal{A}}_h)^{-1} \mathbf{1}_{3N}\|_{\infty}$$

From the above expressions, and using the fact that  $(\Phi_{\bar{x}, h}^{(3)*} \Pi_{(\text{Im } \Gamma_{\bar{x}})^\perp} \Phi_{\bar{x}, h}^{(3)})^{-1} \Phi_{\bar{x}, h}^{(3)} \Gamma_{\bar{x}}^+$   $\begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix} = \rho + o(1)$ , we get the asymptotic equivalents

$$\begin{aligned} \|\hat{\mathcal{A}}_h^+\|_{\infty, \mathcal{H}} &\sim \frac{1}{|\gamma_3|(h)^3} \left\| (\Pi_{(\text{Im } \Gamma_{\bar{x}})^\perp} \Phi_{\bar{x}}^{(3)})^+ \right\|_{\infty, \mathcal{H}}, \\ \text{and } \|(\bar{\mathcal{A}}_h^* \bar{\mathcal{A}}_h)^{-1} \mathbf{1}_{3N}\|_{\infty} &\sim \frac{1}{|\gamma_3|(h)^3} \|\rho\|_{\infty}. \end{aligned}$$

Since  $\min_{i \in I} \bar{r}_i = \min_{\nu \in \{1, \dots, N\}} \left( \frac{1}{2} \alpha_\nu + \frac{1}{h_n} \alpha_\nu s_\nu \right) \geq \frac{(1-\kappa)}{2} \min \alpha$ , and similarly  $\min_{i \in I} \bar{l}_i \geq \frac{(1-\kappa)}{2} \min \alpha$ , we deduce that  $(r_\lambda)_I$  and  $(l_\lambda)_I$  have positive components provided

$$(h_n)^3 \min \alpha \geq c_1 \|\tilde{w}\|_{\mathcal{H}} + c_2 \lambda, \quad (\text{D.10})$$

$$\text{with } c_1 \stackrel{\text{def.}}{=} \frac{4}{|\gamma_3|(1-\kappa)} \left\| (\Pi_{(\text{Im } \Gamma_{\bar{x}})^\perp} \Phi_{\bar{x}}^{(3)})^+ \right\|_{\infty, \mathcal{H}} \quad \text{and} \quad c_2 \stackrel{\text{def.}}{=} \frac{4}{|\gamma_3|(1-\kappa)} \|\rho\|_{\infty}. \quad (\text{D.11})$$

As for  $(z_\lambda)_I$ , *i.e.* the components of  $(r_\lambda)_{J(r) \setminus I}$  and  $(l_\lambda)_{J(l) \setminus I}$ , we note that in fact,

$$(z_\lambda)_I \geq - \left\| \hat{\mathcal{A}}_h^+ \right\|_{\infty, \mathcal{H}} \|\tilde{w}\|_{\mathcal{H}} + \lambda \underbrace{\frac{1}{\gamma_3 h^3} \text{diag}(\varepsilon) (\Phi_{\bar{x}, h}^{(3)*} \Pi_{(\text{Im } \Gamma_{\bar{x}})^\perp} \Phi_{\bar{x}, h}^{(3)})^{-1} \Phi_{\bar{x}, h}^{(3)} \Gamma_{\bar{x}}^+}_{= \frac{1}{|\gamma_3| h^3} (\rho + o(1))} \begin{pmatrix} \mathbf{1}_N \\ 0 \end{pmatrix}.$$

Hence, they are positive provided  $\left\| \hat{\mathcal{A}}_h^+ \right\|_{\infty, \mathcal{H}} \|\tilde{w}\|_{\mathcal{H}} < \frac{\lambda}{|\gamma_3| h^3} \min \{ |\rho_i| + o(1) ; i \in I \}$ . This condition holds for large  $n$  provided  $\|\tilde{w}\|_{\mathcal{H}} / \lambda \leq c_3$ , where

$$c_3 = \frac{\min_{i \in I} |\rho_i|}{2 \left\| (\Pi_{(\text{Im } \Gamma_{\bar{x}})^\perp} \Phi_{\bar{x}}^{(3)})^+ \right\|_{\infty, \mathcal{H}}}. \quad (\text{D.12})$$

To summarize, we have proved the first two points.

For the last point, we have to ensure that

$$\max \left[ \left( (A^* + \frac{h}{2} B^*)_{(J(r))^c} \right) p_\lambda, \left( (A^* - \frac{h}{2} B^*)_{(J(l))^c} \right) p_\lambda \right] < 1$$

Since by construction

$$p_\lambda = \frac{1}{\lambda} \left( \Pi_{(\text{Im } \hat{\mathcal{A}}_h)^\perp} y \right) + \hat{\mathcal{A}}_h^{+,*} \mathbf{1}_{3N} = \frac{1}{\lambda} \left( \Pi_{(\text{Im } \hat{\mathcal{A}}_h)^\perp} \tilde{w} \right) + \hat{\mathcal{A}}_h^{+,*} \mathbf{1}_{3N} = \frac{1}{\lambda} \left( \Pi_{(\text{Im } \hat{\mathcal{A}}_h)^\perp} \tilde{w} \right) + p_{0,n},$$

we are led to check that

$$\left(\eta_{0,n} + \frac{\omega}{\lambda}\right)(jh) + \frac{h}{2} \left(\eta_{0,n} + \frac{\omega}{\lambda}\right)'(jh) < 1 \quad \text{for all } j \in (J^{(r)})^C, \quad (\text{D.13})$$

$$\left(\eta_{0,n} + \frac{\omega}{\lambda}\right)(jh) - \frac{h}{2} \left(\eta_{0,n} + \frac{\omega}{\lambda}\right)'(jh) < 1 \quad \text{for all } j \in (J^{(1)})^C, \quad (\text{D.14})$$

where  $\omega \stackrel{\text{def.}}{=} \Phi^* \Pi_{(\text{Im } \hat{\mathcal{A}}_h)^\perp} w$  and  $\eta_{0,n} \stackrel{\text{def.}}{=} \Phi^* (\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} \mathbf{1}_{3N} = \Phi^* p_{0,n}$  yields the minimal norm certificate

$$\bar{\eta}_{0,n} = \begin{pmatrix} (\eta_{0,n} + \frac{h}{2} \eta'_{0,n})(\mathcal{G}_n) \\ (\eta_{0,n} - \frac{h}{2} \eta'_{0,n})(\mathcal{G}_n) \end{pmatrix} = \begin{pmatrix} (A + \frac{h}{2} B)^* \\ (A - \frac{h}{2} B)^* \end{pmatrix} (\hat{\mathcal{A}}_h^* \hat{\mathcal{A}}_h)^{-1} \mathbf{1}_{3N}.$$

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

$$\tilde{k}_2 \stackrel{\text{def.}}{=} - \sup_{t \in \mathcal{N}(r)} \eta_T''(t) > 0, \quad \text{and} \quad \tilde{k}_4 \stackrel{\text{def.}}{=} \inf_{t \in \mathcal{N}(r)} \eta_T^{(4)}(t) > 0.$$

By compactness,  $\tilde{k}_0 \stackrel{\text{def.}}{=} \sup_{t \in \mathbb{T} \setminus \mathcal{N}(r)} \eta_T(t) < 1$ .

Let us recall that  $\eta_{0,n} \rightarrow \eta_T$  in the sense of the uniform convergence (and similarly for the derivatives). As a result, for  $n \in \mathbb{N}$  large enough,

$$\sup_{t \in \mathbb{T} \setminus \mathcal{N}(r)} \eta_{0,n}(t) \leq \frac{1 + \tilde{k}_0}{2} < 1, \quad \sup_{t \in \mathcal{N}(r)} (\eta_{0,n})''(t) \leq -\frac{\tilde{k}_2}{2} < 0, \quad \inf_{t \in \mathcal{N}(r)} (\eta_{0,n})^{(4)}(t) \geq \frac{\tilde{k}_4}{2} > 0, \quad (\text{D.15})$$

$$\left\| \eta_{0,n}^{(5)} \right\|_\infty \leq \tilde{k}_5 \stackrel{\text{def.}}{=} \left\| \mu_T^{(5)} \right\|_\infty + 1, \quad (\text{D.16})$$

$$\frac{h}{2} \left\| (\eta_{0,n})' \right\|_\infty \leq \frac{1 - \tilde{k}_0}{6} \quad \text{and} \quad \frac{h}{2} \left\| (\eta_{0,n})^{(3)} \right\|_\infty \leq \frac{\tilde{k}_2}{8}.$$

Now, we assume that  $\frac{\|\tilde{w}\|_{\mathcal{H}}}{\lambda}$  is small enough, so that

$$\left\| (\Phi^{(j)})^* \right\|_{\infty, \mathcal{H}} \frac{\|\tilde{w}\|_{\mathcal{H}}}{\lambda} \leq \frac{1 - \tilde{k}_0}{6}, \quad \text{for } j \in \{0, 1\}, \quad (\text{D.17})$$

$$\left\| (\Phi^{(j)})^* \right\|_{\infty, \mathcal{H}} \frac{\|\tilde{w}\|_{\mathcal{H}}}{\lambda} \leq \frac{\tilde{k}_2}{8}, \quad \text{for } j \in \{2, 3\}, \quad (\text{D.18})$$

$$\text{and} \quad \left\| (\Phi^{(4)})^* \right\|_{\infty, \mathcal{H}} \frac{\|\tilde{w}\|_{\mathcal{H}}}{\lambda} \leq \frac{\tilde{k}_4}{4}. \quad (\text{D.19})$$

Then, using the fact that  $|\omega^{(j)}(t)| \leq \left\| (\Phi^{(j)})^* \right\|_{\infty, \mathcal{H}} \|\tilde{w}\|_{\mathcal{H}}$  and  $h \leq 1$ , we obtain

$$\sup_{t \in \mathbb{T} \setminus \mathcal{N}(r)} \left( \eta_{0,n} + \frac{\omega}{\lambda} + \frac{h}{2} \left| (\eta_{0,n} + \frac{\omega}{\lambda})' \right| \right) (t) \leq \frac{1 + \tilde{k}_0}{2} + \frac{1 - \tilde{k}_0}{6} + \frac{1 - \tilde{k}_0}{6} + \frac{1}{2} \left( \frac{1 - \tilde{k}_0}{6} \right) < 1.$$

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

$$\left( \eta_{0,n} + \frac{\omega}{\lambda} + \frac{h}{2} (\eta_{0,n} + \frac{\omega}{\lambda})' \right) (t) < 1 \text{ for } t \in (\bar{x}_\nu - r, \bar{x}_\nu + r) \cap \mathcal{G}_n \setminus S_{n,\nu}^{(\mathbf{r})}(r), \quad (\text{D.20})$$

$$\text{and } \left( \eta_{0,n} + \frac{\omega}{\lambda} - \frac{h}{2} (\eta_{0,n} + \frac{\omega}{\lambda})' \right) (t) < 1 \text{ for } t \in (\bar{x}_\nu - r, \bar{x}_\nu + r) \cap \mathcal{G}_n \setminus S_{n,\nu}^{(\mathbf{l})}(r). \quad (\text{D.21})$$

We first deal with the case  $S_{n,\nu}^{(\mathbf{r})}(r) = \{\bar{x}_\nu\}$ ,  $S_{n,\nu}^{(\mathbf{l})}(r) = \{\bar{x}_\nu, \bar{x}_\nu + h\}$ . Let  $f \stackrel{\text{def}}{=} \frac{1}{\lambda} \omega(\cdot - \bar{x}_\nu) + \eta_{0,n}(\cdot - \bar{x}_\nu)$ . By definition of  $\Pi_{(\text{Im } \hat{A}_h)^\perp}$ ,  $\omega(\bar{x}_\nu) = \omega'(\bar{x}_\nu) = \omega(\bar{x}_\nu + h) - \frac{h}{2} \omega'(\bar{x}_\nu + h) = 0$ , so that

$$f(0) = 1, \quad f'(0) = 0, \quad \text{and} \quad f(h) - \frac{h}{2} f'(h) = 1. \quad (\text{D.22})$$

Moreover, letting  $k_2 = \frac{3}{8} \tilde{k}_2$ ,  $k_4 = \frac{1}{2} \tilde{k}_4$ ,  $k_5 = \tilde{k}_5$ , we see from Eq. (D.15) to (D.19) that  $f$  satisfies the hypotheses of Lemma 8 below. We deduce that the local inequalities (D.20) and (D.21) hold. The symmetric case,  $S_{n,\nu}^{(\mathbf{r})}(r) = \{\bar{x}_\nu - h, \bar{x}_\nu\}$ ,  $S_{n,\nu}^{(\mathbf{l})}(r) = \{\bar{x}_\nu\}$ , is handled similarly, replacing  $f$  with  $f(-\cdot)$ .

To summarize, we have proved the global inequalities (D.13) and (D.14). The (strict) extremality relations hold for  $(r_\lambda, l_\lambda)$  and  $p_\lambda$  and they are unique solutions of the primal and dual problems respectively. The conditions on  $\frac{\|\tilde{w}\|_{\mathcal{H}}}{\lambda}$  stemming from (D.12), and (D.17)-(D.19) define the bound  $C^{(3)}$ . The condition on the signal to noise ratio  $\|\tilde{w}\| / \min \alpha$  and  $\lambda / \min \alpha$  given in (D.10) define the bounds  $C^{(1)}$  and  $C^{(2)}$ .

To finish, using the mapping  $H_h$ , we recover  $(a_\lambda, b_\lambda)$  from (D.9) so as to derive (53). We obtain

$$\begin{aligned} \begin{pmatrix} (a_\lambda)_I \\ (b_\lambda)_I \end{pmatrix} &= \begin{pmatrix} \bar{a}_I \\ \bar{b}_I \end{pmatrix} + (\Pi_{(\text{Im } \Phi_{\bar{x},h}^{(3)})^\perp} \Gamma_{\bar{x}})^+ w + \frac{1}{|\gamma_3|} \begin{pmatrix} \frac{1}{h^3} \text{diag}(\varepsilon) \\ \frac{1}{2h^2} \text{diag}(\varepsilon) \end{pmatrix} (\Pi_{(\text{Im } \Gamma_{\bar{x}})^\perp} \Phi_{\bar{x},h}^{(3)})^+ w \\ &\quad - \lambda (\Gamma_{\bar{x}}^* \Pi_{(\text{Im } \Phi_{\bar{x},h}^{(3)})^\perp} \Gamma_{\bar{x}})^{-1} \begin{pmatrix} \mathbf{1}_N \\ \mathbf{0}_N \end{pmatrix} + \frac{\lambda}{|\gamma_3|} \begin{pmatrix} \frac{1}{h^3} (|\rho| + o(1)) \\ \frac{1}{2h^2} (|\rho| + o(1)) \end{pmatrix}. \end{aligned}$$

and using that  $r_i = 0$  (resp.  $l_i = 0$ ) if  $z_i = l_i$  (resp.  $r_i$ ), we get

$$\begin{pmatrix} (a_\lambda)_{I+\varepsilon} \\ (b_\lambda)_{I+\varepsilon} \end{pmatrix} = -\frac{1}{|\gamma_3|} \begin{pmatrix} \frac{1}{h^3} \text{diag}(\varepsilon) (\Pi_{(\text{Im } \Phi_{\bar{x},h}^{(3)})^\perp} \Gamma_{\bar{x}})^+ w \\ \frac{1}{2h^2} (\Pi_{(\text{Im } \Phi_{\bar{x},h}^{(3)})^\perp} \Gamma_{\bar{x}})^+ w \end{pmatrix} + \frac{\lambda}{|\gamma_3|} \begin{pmatrix} \frac{1}{h^3} (|\rho| + o(1)) \\ \frac{1}{2h^2} \text{diag}(\varepsilon) (|\rho| + o(1)) \end{pmatrix}.$$

**Lemma 8.** *Let  $r > 0$ ,  $f \in \mathcal{C}^5([-r, r])$ ,  $k_2 > 0$ ,  $k_4 > 0$ , and  $k_5 > 0$ . Then, for all  $h > 0$  small enough, the conditions*

$$f(0) = 1, \quad f'(0) = 1, \quad \text{and} \quad f(h) - \frac{h}{2} f'(h) = 1, \quad (\text{D.23})$$

$$\sup_{t \in (-r, r)} f''(t) \leq -k_2, \quad \frac{h}{2} \|f^{(3)}\|_\infty \leq \frac{k_2}{2}, \quad \inf_{t \in (-r, r)} f^{(4)}(t) \geq k_4 \quad \text{and} \quad \|f^{(5)}\|_\infty \leq k_5, \quad (\text{D.24})$$

imply

$$\forall t \in [-r, 0) \cup (h, r], \quad f(t) - \frac{h}{2}f'(t) < 1, \quad (\text{D.25})$$

$$\text{and } \forall t \in [-r, -h] \cup (0, r], \quad f(t) + \frac{h}{2}f'(t) < 1. \quad (\text{D.26})$$

*Proof.* We let  $g_-(t) \stackrel{\text{def.}}{=} f(t) - \frac{h}{2}f'(t)$  and  $g_+(t) \stackrel{\text{def.}}{=} f(t) + \frac{h}{2}f'(t)$ . It suffices to show that  $g_-(t) < 1$  and  $g_+(t) < 1$  in the corresponding intervals.

Observe that

$$\forall t \in (-r, r), \quad g_-''(t) = f''(t) - \frac{h}{2}f^{(3)}(t) \leq -\frac{k_2}{2} < 0$$

so that  $g_-$  is strictly concave. Since  $g_-(0) = g_-(h) = 1$ , we deduce that  $g_-(t) < 1$  for all  $t \in [-r, 0) \cup (h, r]$ .

As for  $g_+$ , we observe similarly that  $g_+$  is strictly concave. Since  $g_+'(0) = \frac{h}{2}f''(0) < 0$  and  $g_+'$  is decreasing, we deduce that  $g_+(t) < g_+(0) = 1$  for all  $t \in (0, r]$ .

We obtain the desired result if we can prove that  $g_+(-h) < 1$ . Now,

$$f(h) + f(-h) = 2 \left( f(0) + h^2 \frac{f''(0)}{2} + h^4 \frac{f^{(4)}(0)}{4!} \right) + h^5 \int_0^1 (f^{(5)}(sh) - f^{(5)}(-sh)) \frac{(1-s)^4}{4!} ds$$

and

$$\frac{h}{2}(-f'(h) + f'(-h)) = h \left( -f''(0)h - h^3 \frac{f^{(4)}(0)}{3!} \right) - h^5 \int_0^1 (f^{(5)}(sh) - f^{(5)}(-sh)) \frac{(1-s)^3}{2 \times 3!} ds.$$

Hence,

$$\begin{aligned} g_-(h) + g_+(-h) &= 2f(0) + 2h^4 f^{(4)}(0) \left( \frac{1}{4!} - \frac{1}{2 \times 3!} \right) \\ &\quad + h^5 \int_0^1 (f^{(5)}(sh) - f^{(5)}(-sh)) \left( \frac{(1-s)^4}{4!} - \frac{(1-s)^3}{2 \times 3!} \right) ds. \end{aligned}$$

Since  $g_-(h) = f(0) = 1$  and  $f^{(4)}(0) \geq k_4 > 0$ , we obtain

$$g_+(-h) \leq 1 - \frac{1}{12}k_4 h^4 + \frac{1}{40} \|f^{(5)}\|_\infty h^5 \quad (\text{D.27})$$

$$\leq 1 - \frac{1}{12}k_4 h^4 + \frac{1}{40}k_5 h^5 < 1 \quad (\text{D.28})$$

for  $h$  small enough. We conclude that the claimed inequality holds.  $\square$

## References

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

- [2] Jerzy K. Baksalary and Oskar Maria Baksalary. “Particular formulae for the Moore–Penrose inverse of a columnwise partitioned matrix”. In: *Linear Algebra and its Applications* 421.1 (Feb. 2007), pp. 16–23. ISSN: 0024-3795 (print), 1873-1856 (electronic).
- [3] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York: Springer-Verlag, 2011.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] G. Dal Maso. *An Introduction to  $\Gamma$ -convergence*. Vol. 8. Progress in nonlinear differential equations and their applications. Boston, MA: Birkhauser, 1993.
- [9] 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.
- [10] Q. Denoyelle, V. Duval, and G. Peyré. “Support Recovery for Sparse Deconvolution of Positive Measures”. In: 1506.08264 (2015). URL: <http://arxiv.org/abs/1506.08264>.
- [11] V. Duval and G. Peyré. “Exact Support Recovery for Sparse Spikes Deconvolution”. In: *to appear in Foundation of Computational Mathematics* (2015).
- [12] V. Duval and G. Peyré. *Sparse Spikes Retrieval on Thin Grids I: the LASSO*. Tech. rep. 2016.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] A. Florescu et al. *A Constrained Optimization Approach For Complex Sparse Perturbed Models*. Tech. rep. Preprint hal-00783298, 2013.

- [17] J.J. Fuchs. “On sparse representations in arbitrary redundant bases”. In: *IEEE Transactions on Information Theory* 50.6 (2004), pp. 1341–1344.
- [18] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. 2nd. New York, NY, USA: Cambridge University Press, 2012. ISBN: 0521548233, 9780521548236.
- [19] R. T. Rockafellar, R. J-B. Wets, and M. Wets. *Variational analysis*. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg, New York: Springer, 1998.
- [20] Geoffrey Schiebinger, Elina Robeva, and Benjamin Recht. “Superresolution without Separation”. In: *CoRR* abs/1506.03144 (2015). URL: <http://arxiv.org/abs/1506.03144>.
- [21] Gongguo Tang and Benjamin Recht. “Atomic decomposition of mixtures of translation-invariant signals”. In: *IEEE CAMSAP* (2013).
- [22] 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.