

# Support Recovery for Sparse Super-Resolution of Positive Measures

Quentin Denoyelle<sup>1,2</sup> · Vincent Duval<sup>2</sup> · Gabriel Peyré<sup>2,3</sup>

Received: 22 July 2015 / Revised: 6 July 2016 / Published online: 26 September 2016  
© Springer Science+Business Media New York 2016

**Abstract** We study sparse spikes super-resolution over the space of Radon measures on  $\mathbb{R}$  or  $\mathbb{T}$  when the input measure is a finite sum of positive Dirac masses using the BLASSO convex program. We focus on the recovery properties of the support and the amplitudes of the initial measure in the presence of noise as a function of the minimum separation  $t$  of the input measure (the minimum distance between two spikes). We show that when  $w/\lambda$ ,  $w/t^{2N-1}$  and  $\lambda/t^{2N-1}$  are small enough (where  $\lambda$  is the regularization parameter,  $w$  the noise and  $N$  the number of spikes), which corresponds roughly to a sufficient signal-to-noise ratio and a noise level small enough with respect to the minimum separation, there exists a unique solution to the BLASSO program with exactly the same number of spikes as the original measure. We show that the amplitudes and positions of the spikes of the solution both converge toward those of the input measure when the noise and the regularization parameter drops to zero faster than  $t^{2N-1}$ .

**Keywords** Super-resolution · Sparsity · Deconvolution · Random measures · Convex optimization · LASSO · BLASSO

---

Communicated by Massimo Fornasier.

---

✉ Gabriel Peyré  
gabriel.peyre@ceremade.dauphine.fr  
Quentin Denoyelle  
denoyelle@ceremade.dauphine.fr  
Vincent Duval  
vincent.duval@inria.fr

<sup>1</sup> CEREMADE, Univ. Paris-Dauphine, Paris, France

<sup>2</sup> INRIA - Mokaplan project-team, Paris, France

<sup>3</sup> CNRS & CEREMADE, Univ. Paris-Dauphine, Paris, France

## 1 Introduction

### 1.1 Super-Resolution and Sparse Spikes Deconvolution

Super-resolution consists in retrieving fine scale details of a possibly noisy signal from coarse scale information. The importance of recovering the high frequencies of a signal comes from the fact that there is often a physical blur in the acquisition process, such as diffraction in optical systems, wave reflection in seismic imaging or spikes recording from neuronal activity.

In resolution theory, the two-point resolution criterion defines the ability of a system to resolve two points of equal intensities. As a point source produces a diffraction pattern which is centered about the geometrical image point of the point source, it is often admitted that two points are resolved by a system if the central maximum of the intensity diffraction of one point source coincides with the first zero of the intensity diffraction pattern of the other point. This defines a distance that only depends on the system and which is called the Rayleigh Length. In the case of the ideal low-pass filter (meaning that the input signal is convolved with the Dirichlet kernel, see (9) for the exact definition) with cutoff frequency  $f_c$ , the Rayleigh Length is  $1/f_c$ . We refer to [11] for more details about resolution theory. Super-resolution in signal processing thus consists in developing techniques that enable to retrieve information below the Rayleigh Length.

Let us introduce in a more formal way the problem which will be the core of this article. Let  $X$  be the real line  $\mathbb{R}$  or the 1-D torus  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$  and  $\mathcal{M}(X)$  the Banach space of bounded Radon measures on  $X$ , which can be seen as the topological dual of the space  $\mathcal{C}_X$  where  $\mathcal{C}_X$  is either the space of continuous functions on  $\mathbb{R}$  that vanish at infinity when  $X = \mathbb{R}$  or the space of continuous functions on  $\mathbb{T}$  when  $X = \mathbb{T}$ . We consider a given integral operator  $\Phi : \mathcal{M}(X) \rightarrow \mathcal{H}$ , where  $\mathcal{H}$  is a separable Hilbert space, whose kernel  $\varphi$  is supposed to be a smooth function (see Definition 1 for the technical assumptions made on  $\varphi$ ), i.e.

$$\forall m \in \mathcal{M}(X), \quad \Phi m = \int_X \varphi(x) dm(x). \quad (1)$$

$\Phi$  represents the acquisition operator and can for instance account for a blur in the measurements. In the special case of  $\varphi(x) = \tilde{\varphi}(\cdot - x)$ ,  $\Phi$  is a convolution operator. We denote by  $m_{a_0, tz_0} = \sum_{i=1}^N a_{0,i} \delta_{tz_{0,i}}$  our input sparse spikes train where the  $a_{0,i} \in \mathbb{R}_+^*$  are the amplitudes of the Dirac masses at positions  $tz_{0,i} \in X$ . Let  $y_t = \Phi m_{a_0, tz_0}$  be the noiseless observation. The parameter  $t > 0$  controls the minimum separation distance between the spikes, and this paper aims at studying the recovery of  $m_{a_0, tz_0}$  from  $y_t + w$  (where  $w \in \mathcal{H}$  is some noise) when  $t$  is small.

### 1.2 From the LASSO to the BLASSO

#### 1.2.1 LASSO

$\ell^1$  regularization techniques were first introduced in geophysics (see [6, 18, 21]) for seismic prospecting. Indeed, the density changes in the underground can be modeled

as a sparse spikes train.  $\ell^1$  reconstruction property provides solutions with few nonzero coefficients and can be solved efficiently with convex optimization methods. Donoho theoretically studied and justified these techniques in [12]. In statistics, the  $\ell^1$  norm is used in the Lasso method [23] which consists in minimizing a quadratic error subject to a  $\ell^1$  penalization. As the authors remarked it, it retains both the features of subset selection (by setting to zero some coefficients, thanks to the property of the  $\ell^1$  norm to favor sparse solutions) and ridge regression (by shrinking the other coefficients). In signal processing, the basis pursuit method [13] uses the  $\ell^1$  norm to decompose signals into overcomplete dictionaries.

### 1.2.2 BLASSO

Following recent works (see for instance [1,3,5,8,14]), the sparse deconvolution method that we consider in this article operates over a continuous domain, i.e. without resorting to some sort of discretization on a grid. The inverse problem is solved over the space of Radon measures which is a non-reflexive Banach space. This continuous “grid free” setting makes the mathematical analysis easier and allows us to make precise statement about the location of the recovered spikes locations.

The technique that we study in this paper consists in solving a convex optimization problem that uses the total variation norm which is the equivalent of the  $\ell^1$  norm for measures. The  $\ell^1$  norm is known to be particularly well fitted for the recovery of sparse signals. Thus the use of the total variation norm favors the emergence of spikes in the solution.

The total variation norm is defined by

$$\forall m \in \mathcal{M}(X), \quad |m|(X) \stackrel{\text{def.}}{=} \sup_{\psi \in \mathcal{C}_X} \left\{ \int_X \psi \, dm ; \|\psi\|_{L^\infty(X)} \leq 1 \right\}.$$

In particular,

$$|m_{a_0, t z_0}|(X) = \|a_0\|_1,$$

which shows in a way that the total variation norm generalizes the  $\ell^1$  norm to the continuous setting of measures (i.e. no discretization grid is required).

The first method that we are interested in is the classical basis pursuit, defined originally in [13] in a finite dimensional setting, and written here over the space of Radon measures

$$\min_{m \in \mathcal{M}(X)} \{|m|(X) ; \Phi m = y_t\}. \quad (\mathcal{P}_0(y_t))$$

This is the problem studied in [5], in the case where  $\Phi$  is an ideal low-pass filter on the torus (i.e.  $X = \mathbb{T}$ ).

When the signal is noisy, i.e. when we observe  $y_t + w$  instead, with  $w \in \mathcal{H}$ , we may rather consider the problem

$$\min_{m \in \mathcal{M}(X)} \frac{1}{2} \|\Phi m - (y_t + w)\|_{\mathcal{H}}^2 + \lambda |m|(X). \quad (\mathcal{P}_\lambda(y_t + w))$$

Here  $\lambda > 0$  is a parameter that should be adapted to the noise level  $\|w\|_{\mathcal{H}}$ . This problem is coined “BLASSO” in [8].

While this is not the focus of this article, we note that there exist algorithms to solve the infinite dimensional convex problems  $(\mathcal{P}_0(y_t))$  and  $(\mathcal{P}_\lambda(y_t + w))$ , see for instance [3, 5].

### 1.3 Previous Works

#### 1.3.1 LASSO/BLASSO Performance Analysis

In order to quantify the recovery performance of the methods  $\mathcal{P}_0(y_t)$  and  $\mathcal{P}_\lambda(y_t + w)$ , the following two questions arise:

1. Does the solution of  $\mathcal{P}_0(y_t)$  recover the input measure  $m_{a_0, t z_0}$ ?
2. How close is the solution of  $\mathcal{P}_\lambda(y_t + w)$  to the solution of  $\mathcal{P}_0(y_t)$ ?

When the amplitudes of the spikes are arbitrary complex numbers, the answers to the above questions require a large enough minimum separation distance  $\Delta(t z_0)$  between the spikes where

$$\Delta(t z_0) \stackrel{\text{def.}}{=} \min_{i \neq j} d_X(t z_{0,i}, t z_{0,j}). \quad (2)$$

where  $d_X$  is either the canonical distance on  $\mathbb{R}$  i.e.

$$\forall x, y \in \mathbb{R}, \quad d_X(x, y) = |x - y|, \quad (3)$$

when  $X = \mathbb{R}$ , or the canonical induced distance on  $\mathbb{T}$  i.e.

$$\forall x, y \in \mathbb{R}, \quad d_X(x + \mathbb{Z}, y + \mathbb{Z}) = \min_{k \in \mathbb{Z}} |x - y + k|, \quad (4)$$

when  $X = \mathbb{T}$ . The first question is addressed in [5] where the authors showed, in the case of  $\Phi$  being the ideal low-pass filter on the torus [see (9)], i.e. when  $\mathcal{H} = \mathbb{T}$ , that  $m_{a_0, t z_0}$  is the unique solution of  $\mathcal{P}_0(y_t)$  provided that  $\Delta(t z_0) \geq \frac{C}{f_c}$  where  $C > 0$  is a universal constant and  $f_c$  the cutoff frequency of the ideal low-pass filter. In the same paper, it is shown that  $C \leq 2$  when  $a_0 \in \mathbb{C}^N$  and  $C \leq 1.87$  when  $a_0 \in \mathbb{R}^N$ . In [14], the authors show that necessarily  $C \geq \frac{1}{2}$ .

The second question receives partial answers in [3, 4, 9, 15]. In [3], it is shown that if the solution of  $\mathcal{P}_0(y_t)$  is unique then the measures recovered by  $\mathcal{P}_\lambda(y_t + w)$  converge in the weak-\* sense to the solution of  $\mathcal{P}_0(y_t)$  when  $\lambda \rightarrow 0$  and  $\|w\|_{\mathcal{H}}/\lambda \rightarrow 0$ . In [4], the authors measure the reconstruction error using the  $L^2$  norm of an ideal low-pass filtered version of the recovered measures. In [9, 15], error bounds are given on the locations of the recovered spikes with respect to those of the input measure  $m_{a_0, t z_0}$ .

However, those works provide little information about the structure of the measures recovered by  $\mathcal{P}_\lambda(y_t + w)$ . That point is addressed in [14] where the authors show that under the *Non Degenerate Source Condition* (see Sect. 2 for more details), there exists a unique solution to  $\mathcal{P}_\lambda(y_t + w)$  with the exact same number of spikes as the original measure provided that  $\lambda$  and  $\|w\|_{\mathcal{H}}/\lambda$  are small enough. Moreover in that regime, this solution converges to the original measure when the noise drops to zero.

### 1.3.2 LASSO/BLASSO for Positive Spikes

For positive spikes (i.e.  $a_{0,i} > 0$ ), the picture is radically different, since exact recovery of  $m_{a_0, t z_0}$  without noise (i.e.  $(w, \lambda) = (0, 0)$ ) holds for all  $t > 0$ , see for instance [8]. Stability constants however explode as  $t \rightarrow 0$ . A recent work [20] shows however that stable recovery is obtained if the signal-to-noise ratio grows faster than  $O(1/t^{2N})$ , closely matching optimal lower bounds of  $O(1/t^{2N-1})$  obtained by combinatorial methods, as also proved recently [10]. Our main contribution is to show that the same  $O(1/t^{2N-1})$  signal-to-noise scaling in fact guarantees a perfect support recovery of the spikes under a certain non-degeneracy condition on the filter. This extends, for positive measures, the initial results of [14] by providing an asymptotic analysis when  $t \rightarrow 0$ .

### 1.3.3 MUSIC and Related Methods

There is a large body of literature in signal processing on spectral methods to perform spikes location from low frequency measurements. One of the most popular methods is MUSIC (for multiple signal classification) [22] and we refer to [16] for an overview of its use in signal processing for line spectral estimation. In the noiseless case, exact reconstruction of the initial signal is guaranteed as long as there are enough observations compared to the number of distinct frequencies [19]. Stability to noise is known to hold under a minimum separation distance similar to the one of the BLASSO [19]. However, on sharp contrast with the behavior of the BLASSO, numerical observations (see for instance [7]), as well as a recent work of Demanet and Nguyen, show that this stability continues to hold regardless of the sign of the amplitudes  $a_{0,i}$ , as soon as the signal-to-noise ratio scales like  $O(1/t^{2N-1})$ . Note that this matches (when  $w$  is a Gaussian white noise) the Cramer-Rao lower bound achievable by any unbiased estimator [2]. This class of methods are thus more efficient than BLASSO for arbitrary measures, but they are restricted to operators  $\Phi$  that are convolution with a low-pass filter, which is not the case of our analysis for the BLASSO.

## 1.4 Contributions

### 1.4.1 Main Results

From these previous works, one can ask whether exact support estimation by BLASSO for positive spikes is achievable when  $t$  tends to 0. Our main result, Theorem 2, shows that this is indeed the case. It states, under some non-degeneracy condition on  $\Phi$ , that there exists a unique solution to  $\mathcal{P}_\lambda(y_t + w)$  with the exact same number of

spikes as the original measure provided that  $\|w\|_{\mathcal{H}}/\lambda$ ,  $\|w\|_{\mathcal{H}}/t^{2N-1}$  and  $\lambda/t^{2N-1}$  are small enough. Moreover we give an upper bound, in that regime, on the error of the recovered measure with respect to the initial measure. As a by-product, we show that the amplitudes and positions of the spikes of the solution both converge towards those of the input measure when the noise and the regularization parameter tend to zero faster than  $t^{2N-1}$ .

#### 1.4.2 Extensions

We consider in this article the case where all the spikes locations  $tz_0$  cluster near zero. Following for instance [20], it is possible to consider a more general model with several cluster points, where the sign of the Diracs is the same around each of these points. Our analysis, although more difficult to perform, could be extended to this setting, at the price of modifying the definition of  $\eta_W$  (see Definition 3) to account for several cluster points.

Lastly, if the kernel  $\Phi$  under consideration has some specific scale  $\sigma$  (such as the standard deviation of a Gaussian kernel, or  $\sigma = 1/f_c$  for the ideal low-pass filter in the case of the deconvolution on the torus), then it is possible to state our contribution by replacing  $t$  by the dimensionless quantity  $\text{SRF} \stackrel{\text{def.}}{=} t/\sigma$  (called “super-resolution factor” in [20]). It is then possible to extend our proof so show that the signal-to-noise ratio should obey the scaling  $1/\text{SRF}^{2N-1}$ .

#### 1.4.3 Roadmap

The exact statement of Theorem 2 (our main contribution), and in particular the definition of the non-degeneracy condition, requires some more background, which is the subject of Sect. 2. The proof of this result can be found in Sects. 3, 4 and 5. It relies on an independent study of the asymptotic behavior of quantities depending on the operator  $\Phi_{tz_0}$  (such as its pseudo-inverse) when  $t$  tends to zero. This takes place in Sect. 3. Note that a sketch of proof of the main result can be found in Sect. 2.3.

### 1.5 Notations

#### 1.5.1 Measures

We consider  $X = \mathbb{R}$  or  $X = \mathbb{T}$  as the space of Dirac masses positions.  $X$  equipped with the distance  $d_X$  [see (3) and (4)] is a locally compact metric space (compact in the case  $X = \mathbb{T}$ ). We denote by  $\mathcal{M}(X)$  the space of bounded Radon measures on  $X$ . It is the topological dual of the Banach space  $\mathcal{C}_X$  (endowed with  $\|\cdot\|_{L^\infty(X)}$ ) of continuous functions defined on  $X$ , that furthermore are imposed to vanish at infinity in the case  $X = \mathbb{R}$ . The two problems that we study, i.e. the Basis Pursuit for measures  $\mathcal{P}_0(y_t)$  and the BLASSO  $\mathcal{P}_\lambda(y_t + w)$ , are two convex optimization problems on the space  $\mathcal{M}(X)$ .

Let us consider  $a_0 \in (\mathbb{R}_+^*)^N$  and  $z_0 \in \mathbb{R}^N$ . When  $X = \mathbb{T}$ , we make the assumption that  $z_0 \in (-\frac{1}{4}, \frac{1}{4})^N$ . We define

$$\Delta_0 \stackrel{\text{def.}}{=} \Delta(z_0), \tag{5}$$

where  $\Delta(z_0)$  is introduced in (2). We denote by  $\overline{B}(\bar{x}, r)$  (resp.  $B(\bar{x}, r)$ ) the closed (resp. open)  $\ell^\infty$  ball in  $\mathbb{R}^N$  with center  $\bar{x}$  and radius  $r$ . We define neighborhoods of respectively  $z_0$  and  $a_0$  as

$$\overline{B}_{z_0} \stackrel{\text{def.}}{=} \overline{B}\left(z_0, \frac{\Delta_0}{4}\right) \quad \text{and} \quad \overline{B}_{a_0} \stackrel{\text{def.}}{=} \overline{B}\left(a_0, \frac{\min_i(a_{0,i})}{4}\right). \tag{6}$$

Note that when  $X = \mathbb{T}$ , if  $z \in \overline{B}_{z_0}$  then  $z \in (-\frac{1}{2}, \frac{1}{2})^N$  and for all  $t \in (0, 1]$ ,  $tz \in (-\frac{1}{2}, \frac{1}{2})^N$ . As a result, depending on the context, we consider  $tz_0$  and  $tz$  as elements of  $\mathbb{T}^N$  by identifying them with their equivalent classes.

Then, the initial measure we want to recover is of the form

$$m_{a_0, tz_0} \stackrel{\text{def.}}{=} \sum_{i=1}^N a_{0,i} \delta_{tz_{0,i}}.$$

Note that for example, in the case  $X = \mathbb{T}$ , when we write  $\delta_{tz_{0,i}}$ ,  $tz_{0,i}$  is considered as an element of  $\mathbb{T}$ .

We use the parameter  $t \in (0, 1]$  to make the spikes locations tend to 0 in  $X$ , so that the minimum separation distance of  $m_{a_0, tz_0} : \Delta(tz_0) \rightarrow 0$  when  $t \rightarrow 0$ .

### 1.5.2 Kernels

The admissible kernels  $\varphi : X \rightarrow \mathcal{H}$  defining the integral operators  $\Phi$  (modeling the blur of acquisition) are functions, defined on  $X$  and taking values in a separable Hilbert space  $\mathcal{H}$ , satisfying some regularity properties listed in the following Definition.

**Definition 1** (*Admissible kernels*) We denote by  $\text{KER}^k$ , the set of admissible kernels of order  $k$ . A function  $\varphi : X \rightarrow \mathcal{H}$  belongs to  $\text{KER}^k$  if  $\varphi \in \mathcal{C}^k(X, \mathcal{H})$ . When  $X = \mathbb{R}$ ,  $\varphi$  must also satisfies the following requirements:

- For all  $p \in \mathcal{H}$ ,  $\langle \varphi(x), p \rangle_{\mathcal{H}} \rightarrow 0$  when  $|x| \rightarrow +\infty$ .
- For all  $0 \leq i \leq k$ ,  $\sup_{x \in \mathcal{H}} \|\varphi^{(i)}(x)\|_{\mathcal{H}} < +\infty$ .

### 1.5.3 Linear Operators

We consider a linear operator  $\Phi : \mathcal{M}(X) \rightarrow \mathcal{H}$  of the form

$$\forall m \in \mathcal{M}(X), \quad \Phi m \stackrel{\text{def.}}{=} \int_X \varphi(x) dm(x), \tag{7}$$

where  $\varphi : X \rightarrow \mathcal{H}$  belongs to  $\text{KER}^0$ .  $\Phi$  is weak-\* to weak continuous. Its adjoint  $\Phi^* : \mathcal{H} \rightarrow \mathcal{C}_X$  is given by

$$\forall p \in \mathcal{H}, \quad \forall x \in X, \quad (\Phi^* p)(x) = \langle \varphi(x), p \rangle_{\mathcal{H}}.$$

A typical example is a convolution operator, where  $\varphi(x) = \tilde{\varphi}(\cdot - x)$ , for some continuous function  $\tilde{\varphi} : X \rightarrow \mathbb{R}$ ,  $\tilde{\varphi} \in L^2(X) \cap \mathcal{C}_X$ , so that  $(\Phi m)(v) = \int_X \tilde{\varphi}(v - x) dm(x)$ . In particular, in the case  $X = \mathbb{R}$ , the Gaussian filter is defined as

$$\forall x \in \mathbb{R}, \quad \tilde{\varphi}(x) = \tilde{\varphi}_G(x) \stackrel{\text{def.}}{=} e^{-x^2/2}, \quad (8)$$

where  $\tilde{\varphi}_G$  is the Gaussian kernel. In the case  $X = \mathbb{T}$ , a typical example of convolution operator is the *ideal low pass filter* which is defined as

$$\forall x \in \mathbb{T}, \quad \tilde{\varphi}(x) = \tilde{\varphi}_D(x) \stackrel{\text{def.}}{=} \sum_{k=-f_c}^{f_c} e^{2i\pi kx}, \quad (9)$$

for some  $f_c \in \mathbb{N}^*$  called the cutoff frequency.  $\tilde{\varphi}_D$  is the Dirichlet kernel with cutoff frequency  $f_c$ .

*Remark 1* This last example is equivalently obtained (as considered for instance in [5]) by using  $\mathcal{H} = \mathbb{C}^{2f_c+1}$  (endowed with its canonical inner product) in place of  $\mathcal{H} = L^2(\mathbb{T})$ , and defining  $\varphi$  as  $\varphi(x) = (e^{2i\pi kx})_{-f_c \leq k \leq f_c}$ . Note that, to simplify the notation, we consider in this paper real Hilbert spaces  $\mathcal{H}$ , but our analysis readily extends to complex Hilbert spaces.

Given general  $\varphi$  and  $\Phi$  as in (7), and given  $\bar{x} = (x_1, \dots, x_N) \in X^N$ , we denote by  $\Phi_{\bar{x}} : \mathbb{R}^N \rightarrow \mathcal{H}$  the linear operator such that

$$\forall a \in \mathbb{R}^N, \quad \Phi_{\bar{x}}(a) \stackrel{\text{def.}}{=} \sum_{i=1}^N a_i \varphi(x_i),$$

and by  $\Gamma_{\bar{x}} : (\mathbb{R}^N \times \mathbb{R}^N) \rightarrow \mathcal{H}$  the linear operator defined by

$$\Gamma_{\bar{x}} \begin{pmatrix} a \\ b \end{pmatrix} \stackrel{\text{def.}}{=} \sum_{i=1}^N (a_i \varphi(x_i) + b_i \varphi'(x_i)).$$

For  $\varphi \in \text{KER}^k$ , operators involving the derivatives are defined similarly,

$$\forall 0 \leq i \leq k, \quad (\Phi_{\bar{x}})^{(k)} : a \in \mathbb{R}^N \mapsto (\Phi_{\bar{x}})^{(k)}(a) = \sum_{i=1}^N a_i \varphi^{(i)}(x_i).$$

We occasionally write  $\Phi'_{\bar{x}}$  (resp.  $\Phi''_{\bar{x}}$ ) for  $(\Phi_{\bar{x}})^{(1)}$  (resp. for  $(\Phi_{\bar{x}})^{(2)}$ ), and we adopt the following matricial notation

$$\Phi_{\bar{x}} = (\varphi(x_1) \dots \varphi(x_N)) \quad \text{and} \quad \Gamma_{\bar{x}} = (\Phi_{\bar{x}} \Phi'_{\bar{x}}),$$



where the “columns” of those “matrices” are elements of  $\mathcal{H}$ . In particular, the adjoint operator  $\Phi_{\bar{x}}^* : \mathcal{H} \rightarrow \mathbb{R}^N$  is given by

$$\forall p \in \mathcal{H}, \quad \Phi_{\bar{x}}^* p = ((\Phi^* p)(x_i))_{1 \leq i \leq N}.$$

We denote by  $\varphi_k \in \mathcal{H}$  the  $k^{th}$  derivative of  $\varphi$  at 0, i.e.

$$\varphi_k \stackrel{\text{def.}}{=} \varphi^{(k)}(0). \tag{10}$$

In particular,  $\varphi_0 = \varphi(0)$ .

Given  $k \in \mathbb{N}$ , we define

$$\Psi_k \stackrel{\text{def.}}{=} (\varphi_0 \quad \varphi_1 \quad \dots \quad \varphi_k). \tag{11}$$

If  $\Psi_k : \mathbb{R}^{k+1} \rightarrow \mathcal{H}$  has full column rank, we define its pseudo-inverse as  $\Psi_k^+ \stackrel{\text{def.}}{=} (\Psi_k^* \Psi_k)^{-1} \Psi_k^*$ . Similarly, we denote  $\Gamma_{\bar{x}}^+ \stackrel{\text{def.}}{=} (\Gamma_{\bar{x}}^* \Gamma_{\bar{x}})^{-1} \Gamma_{\bar{x}}^*$  provided  $\Gamma_{\bar{x}}$  has full column rank.

### 1.5.4 Linear Algebra

For  $z \in \mathbb{R}^N$ , we let

$$H_z \stackrel{\text{def.}}{=} \begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ z_1 & \dots & z_N & 1 & \dots & 1 \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{(z_1)^{2N-1}}{(2N-1)!} & \dots & \frac{(z_N)^{2N-1}}{(2N-1)!} & \frac{(z_1)^{2N-2}}{(2N-2)!} & \dots & \frac{(z_N)^{2N-2}}{(2N-2)!} \end{pmatrix} \in \mathbb{R}^{2N \times 2N}, \tag{12}$$

so that  $H_z^{*,-1}$  is the matrix of the Hermite interpolation at points  $z_1, \dots, z_N$  when  $\mathbb{R}_{2N-1}[X]$  is equipped with the basis  $(1, X, \dots, \frac{X^{2N-1}}{(2N-1)!})$ .

For each  $N \in \mathbb{N}$ , we define

$$\delta_N \stackrel{\text{def.}}{=} (1, 0, \dots, 0)^T \in \mathbb{R}^N, \tag{13}$$

$$\mathbb{1}_N \stackrel{\text{def.}}{=} (1, 1, \dots, 1)^T \in \mathbb{R}^N. \tag{14}$$

We use the  $\ell^\infty$  norm,  $|\cdot|_\infty$ , for vectors of  $\mathbb{R}^N$  or  $\mathbb{R}^{2N}$ , whereas the notation  $\|\cdot\|$  refers to an operator norm (on matrices, or bounded linear operators).  $\|\cdot\|_{\mathcal{H}}$  is the norm on  $\mathcal{H}$  associated to the inner product  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ .  $\|\cdot\|_{L^\infty(X)}$  denotes the  $L^\infty$  norm for functions defined on  $X$ .

## 2 Asymptotic Analysis of Support Recovery

This section exposes our main contribution (Theorem 2). The hypotheses of this result require an injectivity condition and a non-degeneracy condition, that are explained in Sects. 2.1 and 2.2.

### 2.1 Injectivity Hypotheses

We introduce here an injectivity hypothesis which ensures the invertibility of  $\Phi_{tZ}^* \Phi_{tZ}$  and  $\Gamma_{tZ}^* \Gamma_{tZ}$  for  $t > 0$  small enough.

In the case of the ideal low-pass filter [defined in (9)],  $\Gamma_{\bar{x}}$  has full column rank provided that  $\bar{x} = (x_1, \dots, x_N) \in X^N$  has pairwise distinct coordinates (see [14, Sect. 3.6, Proposition 6]). That property is not true for a general operator  $\Phi$ . However, in this paper we focus on sums of Dirac masses that are clustered around the point  $0 \in X$ , i.e.  $\bar{x} = tZ$  for  $t > 0$  and  $Z \in \mathbb{R}^N$  with pairwise distinct components. The following assumption, which is crucial to our analysis, shall ensure that  $\Gamma_{tZ}$  has full rank at least for small  $t$ .

**Definition 2** Let  $\varphi : X \rightarrow \mathcal{H}$ . For all  $k \in \mathbb{N}$ , we say that the hypothesis  $\mathcal{I}_k$  holds if and only if

$$\varphi \in \text{KER}^k \text{ and } (\varphi_0, \dots, \varphi_k) \text{ are linearly independent in } \mathcal{H}. \quad \mathcal{I}_k$$

See Definition 1 for the definition of the space  $\text{KER}^k$  and Eq. (10) for the definition of  $\varphi_k$ .

If  $\mathcal{I}_k$  holds, then  $\Psi_k^* \Psi_k$  is a symmetric positive *definite* matrix, where  $\Psi_k$  is defined in (11).

To exemplify the meaning of this injectivity hypothesis, Proposition 1 below considers the case  $X = \mathbb{T}$  with  $\Phi$  a convolution operator.

**Proposition 1** Let  $\tilde{\varphi} \in \mathcal{C}^k(\mathbb{T}, \mathbb{R})$  (where  $\varphi(x) = \tilde{\varphi}(\cdot - x)$ ), then  $\mathcal{I}_k$  holds if and only if  $\varphi_0$  has at least  $k + 1$  non-zeros Fourier coefficients. In particular if  $\Phi$  is the ideal low-pass filter with cutoff frequency  $f_c \in \mathbb{N}^*$ ,  $\mathcal{I}_k$  holds if and only if  $k \leq 2f_c$ .

The proof of this proposition is given in Sect. 1.

As we shall see in Sect. 3, the conditions  $\mathcal{I}_{N-1}$  and  $\mathcal{I}_{2N-1}$  imply respectively the invertibility of  $\Phi_{tZ}^* \Phi_{tZ}$  and  $\Gamma_{tZ}^* \Gamma_{tZ}$ , provided that  $t$  is small enough. According to Proposition 1, in the special case of an ideal low-pass filter, these conditions holds if and only if  $f_c$  is large enough with respect to the number  $N$  of spikes.

### 2.2 Vanishing Derivatives Precertificate

Following [14, Sect. 4.1, Definition 6], we introduce below the so called “vanishing derivatives pre-certificate”  $\eta_{V,t}$ , which is a function defined on  $X$  that interpolates the spikes positions and signs (here  $+1$ ). Note that  $\eta_{V,t}$  can be computed in closed form by solving the linear system (15).

**Proposition 2** (Vanishing derivatives precertificate, [14]) *If  $\Gamma_{tz_0}$  has full column rank, there is a unique solution to the problem*

$$\inf \left\{ \|p\|_{\mathcal{H}} ; \forall i = 1, \dots, N, (\Phi^* p)(tz_{0,i}) = 1, (\Phi^* p)'(tz_{0,i}) = 0 \right\}.$$

Its solution  $p_{V,t}$  is given by

$$p_{V,t} = \Gamma_{tz_0}^{+,*} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix}, \tag{15}$$

and we define the vanishing derivatives precertificate as  $\eta_{V,t} \stackrel{\text{def.}}{=} \Phi^* p_{V,t}$ .

As shown in [14] (see Sect. 4.2),  $\eta_{V,t}$  governs the support recovery properties of the BLASSO  $\mathcal{P}_\lambda(y_t + w)$ . More precisely, if  $\eta_{V,t}$  is non-degenerate, i.e.

$$\begin{cases} \forall x \in X \setminus \{tz_{0,1}, \dots, tz_{0,N}\}, & |\eta_{V,t}(x)| < 1, \\ \forall i \in \{1, \dots, N\}, & \eta''_{V,t}(tz_{0,i}) \neq 0, \end{cases} \tag{16}$$

then there exists a low noise regime in which the BLASSO recovers exactly the correct number  $N$  of spikes, and the error on the locations and amplitudes is proportional to the noise level.

The constants involved in the main result of [14] (Theorem 2) depend on the value of  $t > 0$ . The goal of this paper is to precisely determine this dependency, and to show that this support recovery result extends to the setting where  $t \rightarrow 0$ , provided that  $(\lambda, w)$  obey some precise scaling with respect to  $t$ .

Since our focus is actually on the support recovery properties of the BLASSO when  $t \rightarrow 0$ , it is natural to look at the limit of  $p_{V,t}$  as  $t \rightarrow 0$ . This leads us to the  $(2N - 1)$ -vanishing derivatives precertificate defined below.

**Proposition 3** ( $(2N - 1)$ -vanishing derivatives precertificate) *If  $\mathcal{I}_{2N-1}$  holds, there is a unique solution to the problem*

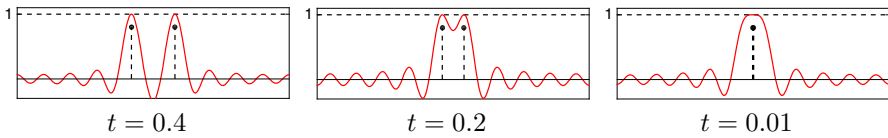
$$\inf \left\{ \|p\|_{\mathcal{H}} ; (\Phi^* p)(0) = 1, (\Phi^* p)'(0) = 0, \dots, (\Phi^* p)^{(2N-1)}(0) = 0 \right\}.$$

We denote by  $p_W$  its solution, given by

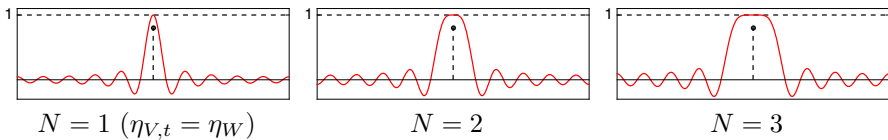
$$p_W = \Psi_{2N-1}^{+,*} \delta_{2N} \tag{17}$$

and we define the  $(2N - 1)$ -vanishing derivatives precertificate as  $\eta_W \stackrel{\text{def.}}{=} \Phi^* p_W$ .

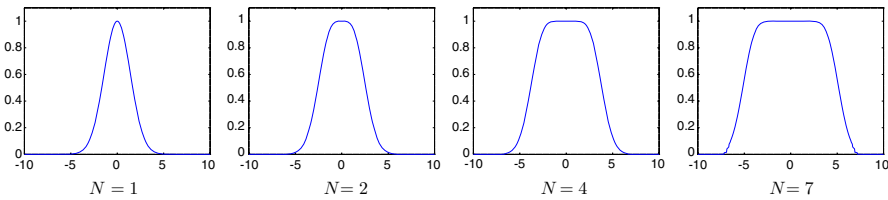
The following Proposition, which is a direct consequence of Lemma 1 in the next section, shows that indeed  $\eta_{V,t}$  converges toward  $\eta_W$ .



**Fig. 1** Top row  $\eta_{V,t}$  for several values of  $t$ , showing the convergence toward  $\eta_W$ . The operator  $\Phi$  is an ideal low-pass filter with a cutoff frequency  $f_c = 10$  [see (9)]



**Fig. 2**  $\eta_W$  for several values of  $N$ . The operator  $\Phi$  is an ideal low-pass filter with a cutoff frequency  $f_c = 10$  [see (9)]



**Fig. 3**  $\eta_W$  for the Gaussian filter ( $x \in \mathbb{R}, \varphi(x) = e^{-\frac{x^2}{2\sigma^2}}$ ) for several numbers of spikes and  $\sigma = 1$ .  $\eta_W$  is  $(2N - 1)$ -non-degenerate. It gets flatter at 0 when the number of spikes increases

**Proposition 4** *If  $\mathcal{I}_{2N-1}$  holds and  $\varphi \in \text{KER}^K$  (for  $K \geq 2N - 1$ ), then for  $t > 0$  small enough  $\Gamma_{t z_0}$  has full column rank. Moreover*

$$\lim_{t \rightarrow 0^+} p_{V,t} = p_W \text{ strongly in } \mathcal{H},$$

$$\lim_{t \rightarrow 0^+} \eta_{V,t}^{(k)} = \eta_W^{(k)} \text{ in the sense of the uniform convergence on } X,$$

for all  $0 \leq k \leq K$ .

Figure 1 shows graphically this convergence of  $\eta_{V,t}$  toward  $\eta_W$  in the case of the deconvolution problem over the 1-D torus with the Dirichlet kernel. Figures 2 and 3 show  $\eta_W$  for several values of  $N$ . Notice how it becomes flatter at 0 as  $N$  increases. This implies that  $\eta_{V,t}$  for small  $t$  gets closer to degeneracy as  $N$  increases. This is reflected in our main contribution (Theorem 2) where the signal-to-noise ratio is required to scale with  $t^{2N-1}$ .

The behavior of  $\eta_{V,t}$  is therefore governed by specific properties of  $\eta_W$  for small values of  $t > 0$ . In particular, as stated by Theorem 1 below, the non-degeneracy of  $\eta_W$  (as defined in Definition 3 below) implies the non-degeneracy of  $\eta_{V,t}$  (as defined in (16)).

**Definition 3** Assume that  $\mathcal{I}_{2N-1}$  holds and  $\varphi \in \text{KER}^{2N}$ . We say that  $\eta_W$  is  $(2N - 1)$ -non-degenerate if  $\eta_W^{(2N)}(0) \neq 0$  and for all  $x \in X \setminus \{0\}, |\eta_W(x)| < 1$ .

**Theorem 1** *Suppose that  $\eta_W$  is  $(2N - 1)$ -non-degenerate (Definition 3). Let  $R_W > 0$ . Then, there exist  $C_W > 0$ ,  $t_W > 0$  such that for all  $t \in (0, t_W)$ , all  $z \in \mathbb{R}^N$  with pairwise distinct coordinates and  $|z|_\infty \leq R_W$ , and all  $\eta \in \mathcal{C}^{2N}(X) \cap W^{2N,\infty}(X)$  satisfying for  $1 \leq i \leq N$ ,  $\eta(tz_i) = 1$  and  $\eta'(tz_i) = 0$ ,*

$$\left( \forall \ell \in \{0, \dots, 2N\}, \quad \left\| \eta^{(\ell)} - \eta_W^{(\ell)} \right\|_{L^\infty(X)} \leq C_W \right) \\ \implies \left( \forall x \in X \setminus \bigcup_i \{tz_i\}, \quad |\eta(x)| < 1 \quad \text{and} \quad \forall 1 \leq i \leq N, \quad \eta''(tz_i) < 0 \right).$$

The proof of this theorem can be found in Sect. 1.

*Remark 2* An important consequence of Theorem 1 is that if  $\eta_W$  is  $(2N - 1)$ -non-degenerate, then, by Proposition 4,  $\eta_{V,t}$  is also non-degenerate for  $t > 0$  small enough. The function  $\eta_{V,t}$  is then the minimal norm certificate for the measure  $m_{a_0,tz_0}$ , and the *Non-Degenerate Source Condition* (see [14]) holds. As a result, for fixed small  $t > 0$ , the BLASSO admits a unique solution in a certain low noise regime corresponding to a large enough signal to noise ratio, with exactly the same number of spikes as the original measure  $m_{a_0,tz_0}$ . For more details on that matter, see [14].

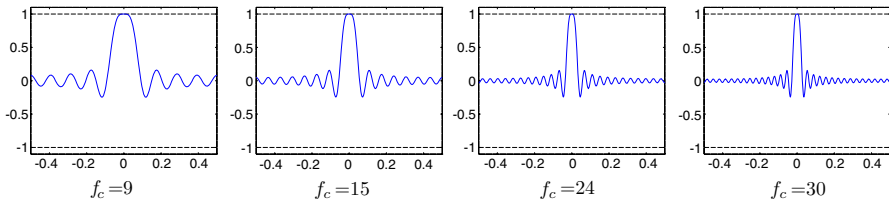
A natural question is whether  $\eta_W$  is indeed  $(2N - 1)$ -non-degenerate. Proposition 5 below (proved in Sect. 1) gives a partial answer in the case of the deconvolution over  $X$ .

**Proposition 5** *Assume that  $\Phi$  is a convolution operator (i.e. for all  $x \in X$ ,  $\varphi(x) = \tilde{\varphi}(\cdot - x)$  and  $\mathcal{H} = L^2(X)$ ) and  $\mathcal{I}_{2N}$  holds. Suppose also, only in the case  $X = \mathbb{R}$ , that for all  $0 \leq i \leq 2N - 1$ ,  $\tilde{\varphi}^{(i)}(x) \rightarrow 0$  when  $|x| \rightarrow +\infty$ . Then  $\eta_W^{(2N)}(0) < 0$ .*

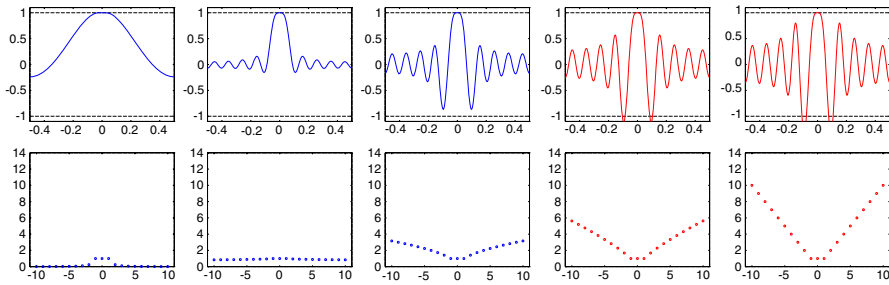
*Remark 3* Thanks to Proposition 5 and the first part of the proof of Theorem 1 (which is given in Sect. 1), note that the following is true: there exist  $C_W > 0$ ,  $t_W > 0$  such that for all  $t \in (0, t_W)$ ,  $z \in \mathbb{R}^N$  with pairwise distinct coordinates and  $|z|_\infty \leq R_W$ , there exists  $r^+ > 0$  with  $r^+ > \max_{1 \leq i \leq N} t_W z_i$  and  $r^- < 0$  with  $r^- < \min_{1 \leq i \leq N} t_W z_i$  such that for all  $\eta \in \mathcal{C}^{2N}(X) \cap W^{2N,\infty}(X)$  satisfying for all  $1 \leq i \leq N$ ,  $\eta(tz_i) = 1$  and  $\eta'(tz_i) = 0$ :

$$\left( \forall \ell \in \{0, \dots, 2N\}, \quad \left\| \eta^{(\ell)} - \eta_W^{(\ell)} \right\|_{L^\infty(X)} \leq C_W \right) \\ \implies \left( \forall x \in (r^-, r^+) \setminus \bigcup_i \{tz_i\}, \quad |\eta(x)| < 1 \quad \text{and} \quad \forall i \in \{1, \dots, N\}, \quad \eta''(tz_i) < 0 \right).$$

Whether the other condition in the definition of the  $(2N - 1)$ -non-degeneracy of  $\eta_W$  holds (i.e. whether  $|\eta_W| < 1$  on  $X \setminus \{0\}$ ) should be checked on a case-by-case basis. Since  $\eta_W$  depends only on the kernel  $\varphi$  and can be computed by simply inverting a linear system, it is easy to check numerically if  $\eta_W$  is  $(2N - 1)$ -non-degenerate. Proposition 6 shows that in the special case of  $X = \mathbb{R}$  and the Gaussian kernel,  $\eta_W$  can be computed in closed form and is indeed  $(2N - 1)$ -non-degenerate.



**Fig. 4**  $\eta_W$  for the ideal low pass filter as  $f_c$  increases, for  $N = 2$



**Fig. 5**  $\eta_W$  for a low pass filter for  $f_c = 10$  with increasing high frequency content. *First row*  $\eta_W$ . *Second row* associated Fourier coefficients of the filter. The curve showing  $\eta_W$  is in blue when it is  $(2N - 1)$ -non-degenerate and in red when it is degenerate

**Proposition 6** Assume that  $X = \mathbb{R}$ ,  $\mathcal{H} = L^2(\mathbb{R})$  and  $\Phi$  is a convolution operator, i.e. for all  $x \in \mathbb{R}$ ,  $\varphi(x) = \tilde{\varphi}(\cdot - x)$ , where  $\tilde{\varphi} : x \in \mathbb{R} \mapsto e^{-x^2/2}$  is a Gaussian. Then the associated  $(2N - 1)$ -vanishing derivatives precertificate is

$$\forall x \in \mathbb{R}, \quad \eta_W(x) = e^{-\frac{x^2}{4}} \sum_{k=1}^N \frac{x^{2k}}{2^{2k} k!}. \tag{18}$$

In particular,  $\eta_W$  is  $(2N - 1)$ -non-degenerate.

The proof of this result can be found in the Appendix 1.4. If we denote by  $\eta_{W,\sigma}$ , the  $(2N - 1)$  vanishing derivatives precertificate associated to the filter  $\varphi_\sigma : x \in \mathbb{R} \mapsto e^{-\frac{x^2}{2\sigma^2}}$ , then  $\eta_{W,\sigma} = \eta_{W,1}(\frac{\cdot}{\sigma})$ . That is why we only consider the case of  $\sigma = 1$  in Proposition 6. Figure 3 shows  $\eta_W$  for the Gaussian filter with an increasing number  $N$  of spikes.

For the deconvolution over the 1-D torus, we observed numerically (as illustrated in Fig. 2) that  $\eta_W$  is  $(2N - 1)$ -non degenerate for the ideal low pass filter for any value  $N$  such that  $N \leq f_c$  (Fig. 4 represents the complementary case where  $N$  is fixed but  $f_c$  increases). However, for some filters, the associated  $\eta_W$  might be degenerate. This is illustrated in Fig. 5 where  $\eta_W$  is illustrated for several filters with increasing complexity i.e. we consider low pass filters with a fixed cutoff frequency, with increasing extreme Fourier coefficients (starting with a slowly varying filter). Remark that the last two  $\eta_W$  (in red) are degenerate, as they correspond to the filters with the higher complexity (the Fourier coefficients increase the most with the frequency).

### 2.3 Main Contribution

We now state our main contribution.

**Theorem 2** *Suppose that  $\varphi \in \text{KER}^{2N+1}$  and that  $\eta_W$  is  $(2N - 1)$ -non-degenerate. Then there exist constants  $(t_1, t_W, C, C_R, M)$  (depending only on  $\varphi, a_0$  and  $z_0$ ) such that for all  $0 < t < \min(t_1, t_W)$ , for all  $(\lambda, w) \in \mathbf{B}(0, C_R t^{2N-1})$  with  $\|\frac{w}{\lambda}\|_{\mathcal{H}} \leq C$ ,*

- *the problem  $\mathcal{P}_\lambda(y_t + w)$  admits a unique solution,*
- *that solution has exactly  $N$  spikes, and it is of the form  $m_{a,tz}$ , with  $(a, z) = g_t^*(\lambda, w)$  (where  $g_t^*$  is a  $\mathcal{C}^{2N}$  function defined on  $\mathbf{B}(0, C_R t^{2N-1}) \subset \mathbb{R} \times \mathcal{H}$ ),*
- *the following inequality holds*

$$|(a, z) - (a_0, z_0)|_\infty \leq M \left( \frac{|\lambda|}{t^{2N-1}} + \frac{\|w\|_{\mathcal{H}}}{t^{2N-1}} \right).$$

Note that the value of constants involved can be found in the proof of the theorem, more precisely,  $(t_1, C)$  are defined in (46),  $C_R$  is defined in Proposition 9,  $t_W$  is defined in Theorem 1 and  $M$  is given in Corollary 1.

The proof of Theorem 2 uses results spanning Sects. 3, 4 and 5. Below, we give a sketch of proof to guide the reader through the remaining of the paper. The elements of the proof are divided in three main steps.

**Step 1** (Sections 4.1 and 4.2) We start with the first order optimality equation that any solution  $m_{a,tz}$ , for fixed  $t > 0$ , of  $\mathcal{P}_\lambda(y_t + w)$  must satisfy i.e.

$$\Gamma_{tz}^* (\Phi_{tz} a - \Phi_{tz_0} a_0 - w) + \lambda \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} = 0.$$

It is obtained by applying Fermat’s rule to the problem  $\mathcal{P}_\lambda(y_t + w)$ . Since the parameters  $(a, z, \lambda, w) = (a_0, z_0, 0_{\mathbb{R}}, 0_{\mathcal{H}})$  are a solution of the equation, the idea is to parametrize, in a neighborhood of  $(\lambda, w) = (0_{\mathbb{R}}, 0_{\mathcal{H}})$ , the amplitudes and positions  $(a, z)$  in terms of  $(\lambda, w)$  by applying the implicit function theorem so that  $(a, z, \lambda, w)$  is a solution of the first order equation. This process is detailed in Sects. 4.1 and 4.2. The rest of the proof consists in proving that the measure  $m_{a,tz}$  is the unique solution of the problem  $\mathcal{P}_\lambda(y_t + w)$ . But before we have to deal with the domain of existence of the above parametrization.

**Step 2** (Section 4.3) The implicit function theorem only provides the existence of a neighborhood in  $(\lambda, w)$  of  $(0_{\mathbb{R}}, 0_{\mathcal{H}})$  where the parametrization holds, but we do not know how its size depends on the parameter  $t$ . This issue is important because one of our aims is to determine the constraints on  $t$  (corresponding roughly to the minimum distance between the spikes of the original measure), on the noise level and on the regularization parameter  $\lambda$  so that the recovery of the support is possible. Section 4.3 is devoted to show that the parametrization, which writes  $(a, z) = g_t^*(\lambda, w)$  (see Eq. (38) for the definition of the implicit function  $g_t^*$ ), of the solutions of the first order optimality equation holds in a neighborhood of  $(0_{\mathbb{R}}, 0_{\mathcal{H}})$  and of size proportional to  $t^{2N-1}$ . This result corresponds to Proposition 9. The proof uses an upper bound of

$dg_t^*$  which is stated in Corollary 1. Proposition 9 relies on asymptotic expansions (of  $\Gamma_{tz}$  for example), when  $t \rightarrow 0$ , gathered and used in the proof of Lemma 5. Section 3 is devoted to these asymptotic expansions and it may be skipped at first reading.

**Step 3** (Section 5) Up to now, we have constructed a candidate solution  $m_{a,tz}$  (composed of  $N$  spikes) where  $(a, z) = g_t^*(\lambda, w)$  is built by parametrizing the solutions of the first order optimality equation. Moreover this parametrization holds for all  $(\lambda, w)$  in a ball of radius proportional to  $t^{2N-1}$ . It remains to prove that  $m_{a,tz}$  is indeed the unique solution of  $\mathcal{P}_\lambda(y_t + w)$ . To prove that  $m_{a,tz}$  is a solution, it is equivalent to check that

$$0 \in \partial \left( m \mapsto \frac{1}{2} \|\Phi m - y_t - w\|_{\mathcal{H}}^2 + \lambda |m|(X) \right) (m_{a,tz}),$$

which reformulates into  $\eta_{\lambda,t} \stackrel{\text{def.}}{=} \frac{1}{\lambda} \Phi^*(y_t + w - \Phi m_{a,tz}) \in \partial |m_{a,tz}|(X)$ . This is done by first showing the convergence of  $\eta_{\lambda,t}$  towards  $\eta_W$  when  $(t, \lambda, w) \rightarrow 0$  in a well chosen domain, see Proposition 11, and then using Theorem 1 and the fact that  $\eta_W$  is ensured to be  $(2N - 1)$ -non-degenerate (which is one of the hypotheses of Theorem 2) to get the non-degeneracy of  $\eta_{\lambda,t}$  and the conclusion.

### 2.3.1 Putting All Together

After this sketch, we now give the detailed proof. It uses Proposition 9 (parametrization of the solution of the first order optimality equation on a ball, for the parameter  $(\lambda, w)$ , of radius proportional to  $t^{2N-1}$ ), Proposition 11 (convergence of  $\eta_{\lambda,t}$  towards  $\eta_W$ ), Theorem 1 (use of the  $(2N - 1)$ -non-degeneracy of  $\eta_W$  to transfer it to  $\eta_{\lambda,t}$ ), and Proposition 10 (upper bound on the error of  $m_{a,tz}$  with respect to  $m_{a_0,tz_0}$ ).

*Proof of Theorem 2* Let us take  $t, \lambda, w$  as in the hypotheses of the Theorem 2. Let  $(a, z) = g_t^*(\lambda, w)$ , where  $g_t^*$  is the function constructed in Sect. 4. Let us define

$$p_{\lambda,t} \stackrel{\text{def.}}{=} \frac{1}{\lambda} (\Phi_{tz_0} a_0 + w - \Phi_{tz} a) \quad \text{and} \quad \eta_{\lambda,t} \stackrel{\text{def.}}{=} \Phi^* p_{\lambda,t}.$$

By Proposition 11 combined with Theorem 1 where we take

$$R_W \stackrel{\text{def.}}{=} \sup\{|z|_\infty ; z \in \bar{B}_{z_0}\}, \tag{19}$$

we have for  $0 < t < \min(t_W, t_1)$ ,

$$\forall x \in X \setminus \bigcup_i \{tz_i\}, |\eta_{\lambda,t}(x)| < 1 \quad \text{and} \quad \forall 1 \leq i \leq N, \eta''_{\lambda,t}(tz_i) < 0, \tag{20}$$

while  $\eta_{\lambda,t}(tz_i) = 1 = \text{sign}(a_i)$  by definition.

We deduce that  $\eta_{\lambda,t}$  is in the subdifferential of the total variation at  $m_{a,tz}$  because

- $\eta_{\lambda,t} \in \mathcal{C}_X$ ,



- $\|\eta_{\lambda,t}\|_{L^\infty(X)} \leq 1$  thanks to Eq. (20),
- $\forall 1 \leq i \leq N, \eta_{\lambda,t}(tz_i) = 1 = \text{sign}(a_i)$  by definition of  $\eta_{\lambda,t}$  (recall that  $(a, z) = g_t^*(\lambda, w)$ ).

As a result  $m_{a,tz}$  is a solution to  $\mathcal{P}_\lambda(y_t + w)$  and  $p_{\lambda,t}$  is the unique solution to the dual problem associated to  $\mathcal{P}_\lambda(y_t + w)$  (see [14, Sect. 2.4] for details on dual certificates and optimality conditions for  $\mathcal{P}_\lambda(y_t + w)$ ).

Let  $m$  be an other solution of  $\mathcal{P}_\lambda(y_t + w)$ . Then the support of  $m$  is included in the saturation points of  $\eta_{\lambda,t} = \Phi^* p_{\lambda,t}$  i.e. in  $\{tz_1, \dots, tz_N\}$ . As a result  $m = m_{a',tz}$  for some  $a' \in \mathbb{R}^N$  and  $m$  satisfies the first order optimality equation  $\Gamma_{tz}^* (\Phi_{tz} a' - \Phi_{tz_0} a_0 - w) + \lambda \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} = 0$ . Hence  $\Phi_{tz}^* \Phi_{tz} a' = \Phi_{tz}^* \Phi_{tz} a$  and since  $\Phi_{tz}$  has full rank (by assumption  $t$  is chosen sufficiently small, see Lemma 1 for the proof),  $\Phi_{tz}^* \Phi_{tz}$  is invertible and  $a' = a$ . So  $m = m_{a,tz}$  and  $\mathcal{P}_\lambda(y_t + w)$  admits a unique solution:  $m_{a,tz}$ .

The bound on the error between  $(a, z)$  and the amplitudes and positions of the initial measure  $(a_0, z_0)$  is a direct consequence of Proposition 10. □

### 2.4 Necessary Condition for the Recovery in the Limit $t \rightarrow 0$

Our main contribution, Theorem 2, states that under a non-degeneracy property which involves  $\eta_W$ , it is possible to perform the recovery of the support of a measure  $m_{a_0,tz_0}$  in the limit  $t \rightarrow 0$  when the data are contaminated by some noise, provided that  $\max(|\lambda|/t^{2N-1}, \|w\|_{\mathcal{H}}/t^{2N-1}, \|w\|_{\mathcal{H}}/\lambda) \leq C$  for some constant  $C > 0$  depending only on the filter  $\varphi$  and  $(a_0, z_0)$ . It is natural to ask whether the non-degeneracy condition on  $\eta_W$ , in order to get the recovery of the support in some low noise regime, is sharp.

The following Theorem shows that the  $(2N - 1)$ -non-degeneracy assumption on  $\eta_W$  is almost sharp in the sense that the recovery of the support in a low noise regime leads to  $\|\eta_W\|_{L^\infty(X)} \leq 1$ .

**Theorem 3** *Suppose that  $\mathcal{I}_{2N-1}$  holds and  $\varphi \in \text{KER}^{2N+1}$ . Suppose also that there exists a sequence  $(t_n)_{n \in \mathbb{N}}$  such that  $t_n \rightarrow 0$  and satisfying*

$$\forall n \in \mathbb{N}, \exists(\lambda_n, w_n), \exists(a_n, z_n) \in \mathbb{R}^N \times \mathbb{R}^N, m_{a_n,t_n z_n} \text{ is solution of } \mathcal{P}_{\lambda_n}(y_{t_n} + w_n),$$

where  $(\lambda_n, w_n) \rightarrow 0$  with  $\frac{\|w_n\|_{\mathcal{H}}}{\lambda_n} \rightarrow 0$ . Then

$$\|\eta_W\|_{L^\infty(X)} \leq 1. \tag{21}$$

The proof of this result can be found in Appendix 1.5.

The remaining sections of the paper, namely Sects. 3, 4 and 5 are devoted to the proof of Theorem 2.

### 3 Preliminaries

Our study relies to a large extent on the asymptotic behavior of quantities built upon  $\Phi_{t_z}$  and  $\Gamma_{t_z}$  for  $t > 0$  small, such as  $(\Phi_{t_z}^* \Phi_{t_z})^{-1}$  or  $(\Gamma_{t_z}^* \Gamma_{t_z})^{-1}$ . In this section, we gather several preliminary results that enable us to control that behavior.

#### 3.1 Approximate Factorizations

Our asymptotic estimates are based on an approximate factorization of  $\Phi_{\bar{x}}$  and  $\Gamma_{\bar{x}}$  using Vandermonde and Hermite matrices. It enables us to study the asymptotic behavior of the optimality conditions of  $\mathcal{P}_\lambda(y_t + w)$  when  $t \rightarrow 0^+$ . In the following, we consider  $z \in \bar{\mathcal{B}}_{z_0}$  [see (6)] and  $t \in (0, 1]$ , so that  $H_{t_z}$  is always invertible. Moreover, we shall always assume that  $\varphi \in \text{KER}^{2N}$ .

**Proposition 7** *The following expansion holds*

$$\Gamma_{t_z} = \Psi_{2N-1} H_{t_z} + \Lambda_{t,z} D_t, \quad (22)$$

where  $\Psi_{2N-1}$  is defined in (11),  $H_{t_z}$  is defined in (12), and where

$$\begin{aligned} \Lambda_{t,z} &\stackrel{\text{def.}}{=} \left( \left( \int_0^1 (z_i)^{2N} \varphi^{(2N)}(stz_i) \frac{(1-s)^{2N-1}}{(2N-1)!} ds \right)_{1 \leq i \leq N}, \right. \\ &\quad \left. \left( \int_0^1 (z_i)^{2N-1} \varphi^{(2N)}(stz_i) \frac{(1-s)^{2N-2}}{(2N-2)!} ds \right)_{1 \leq i \leq N} \right) \\ D_t &\stackrel{\text{def.}}{=} \text{diag}(t^{2N}, \dots, t^{2N}, t^{2N-1}, \dots, t^{2N-1}). \end{aligned}$$

*Proof* This expansion is nothing but the Taylor expansions for  $\varphi$  and  $\varphi'$ :

$$\begin{aligned} \varphi(tz_i) &= \varphi_0 + (tz_i)\varphi_1 + \dots + \frac{(tz_i)^{2N-1}}{(2N-1)!} \varphi_{2N-1} \\ &\quad + (tz_i)^{2N} \int_0^1 \varphi^{(2N)}(stz_i) \frac{(1-s)^{2N-1}}{(2N-1)!} ds, \end{aligned} \quad (23)$$

$$\begin{aligned} \varphi'(tz_i) &= \varphi_1 + (tz_i)\varphi_2 + \dots + \frac{(tz_i)^{2N-2}}{(2N-2)!} \varphi_{2N-1} \\ &\quad + (tz_i)^{2N-1} \int_0^1 \varphi^{(2N)}(stz_i) \frac{(1-s)^{2N-2}}{(2N-2)!} ds. \end{aligned} \quad (24)$$

□

The above expansion yields a useful factorization for  $\Gamma_{t_z}$ ,

$$\Gamma_{t_z} = \Psi_{t_z} H_{t_z} \quad \text{where} \quad \Psi_{t_z} \stackrel{\text{def.}}{=} \Psi_{2N-1} + \Lambda_{t,z} D_t H_{t_z}^{-1}.$$

The rest of this section is devoted to the consequence of that factorization for the asymptotic behavior of  $\Gamma_{t_z}$  and its related quantities. The main ingredient of this analysis is the factorization of  $H_{t_z}$  as

$$H_{tz} = \text{diag}(1, t, \dots, t^{2N-1})H_z \text{diag}\left(1, \dots, 1, \frac{1}{t}, \dots, \frac{1}{t}\right). \tag{25}$$

Let us emphasize that our Taylor expansions are uniform in  $z \in \overline{B}_{z_0}$ . More precisely, given two quantities  $f(z, t), g(z, t)$ , we say that  $f(z, t) = g(z, t) + O(t^k)$  if

$$\limsup_{t \rightarrow 0^+} \sup_{z \in \overline{B}_{z_0}} \left| \frac{f(z, t) - g(z, t)}{t^k} \right| < +\infty.$$

**Lemma 1** *The following expansion holds for  $t \rightarrow 0^+$ ,*

$$\Psi_{tz} = \Psi_{2N-1} + O(t). \tag{26}$$

Moreover, if  $\mathcal{I}_{2N-1}$  holds then  $\Psi_{tz}$  and  $\Gamma_{tz}$  have full column rank for  $t > 0$  small enough and

$$(\Psi_{tz}^* \Psi_{tz})^{-1} = (\Psi_{2N-1}^* \Psi_{2N-1})^{-1} + O(t) \tag{27}$$

$$\Gamma_{tz}^{+,*} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} = \Psi_{2N-1}^{+,*} \delta_{2N} + O(t). \tag{28}$$

*Proof* We begin by noticing that

$$\begin{aligned} \Lambda_{t,z} D_t H_{tz}^{-1} &= t^{2N} \Lambda_{t,z} H_z^{-1} \text{diag}(1, 1/t, \dots, 1/t^{2N-1}) \\ &= \Lambda_{t,z} H_z^{-1} \text{diag}(t^{2N}, t^{2N-1}, \dots, t). \end{aligned}$$

The function  $z \mapsto H_z^{-1}$  is  $\mathcal{C}^\infty$  and uniformly bounded on  $\overline{B}_{z_0}$ , and  $(z, t) \mapsto \Lambda_{t,z}$  is  $\mathcal{C}^0$  on the compact set  $\overline{B}_{z_0} \times [0, 1]$  hence uniformly bounded too. As a result, we get (26).

Assume now that  $\mathcal{I}_{2N-1}$  holds. Since  $\Psi_{2N-1}^* \Psi_{2N-1}$  is invertible, there is some  $R > 0$  such that for every  $A$  in the closed ball  $\overline{B}(\Psi_{2N-1}^* \Psi_{2N-1}, R) \subset \mathbb{R}^{(2N-1) \times (2N-1)}$ ,  $A$  is invertible. By the mean value inequality

$$\begin{aligned} \left\| (\Psi_{2N-1}^* \Psi_{2N-1})^{-1} - A^{-1} \right\| &\leq \sup_{B \in \overline{B}(\Psi_{2N-1}^* \Psi_{2N-1}, R)} \left\| B^{-1} (A - \Psi_{2N-1}^* \Psi_{2N-1}) B^{-1} \right\| \\ &\leq \left( \sup_{B \in \overline{B}(\Psi_{2N-1}^* \Psi_{2N-1}, R)} \left\| B^{-1} \right\| \right)^2 \left\| A - \Psi_{2N-1}^* \Psi_{2N-1} \right\|. \end{aligned}$$

Applying that to  $A = \Psi_{tz}^* \Psi_{tz} = \Psi_{2N-1}^* \Psi_{2N-1} + O(t)$  (since each term in the product is uniformly bounded), we get (27).

Now, for the last point, we infer from  $\Gamma_{tz} = \Psi_{tz} H_{tz}$  and the fact that  $H_{tz}$  is invertible that  $\Gamma_{tz}^{+,*} = \Psi_{tz}^{+,*} H_{tz}^{-1,*}$ . Hence

$$\Gamma_{tz}^{+,*} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} = \Psi_{tz}^{+,*} \delta_{2N} = \Psi_{tz} (\Psi_{tz}^* \Psi_{tz})^{-1} \delta_{2N}$$

where  $\delta_{2N}$  is defined in (13).

Each factor below being uniformly bounded in  $\bar{B}_{z_0} \times [0, 1]$ , we get

$$\begin{aligned} \Gamma_{tz}^{+,*} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} &= (\Psi_{2N-1} + O(t)) \left( (\Psi_{2N-1}^* \Psi_{2N-1})^{-1} + O(t) \right) \delta_{2N} \\ &= \Psi_{2N-1} (\Psi_{2N-1}^* \Psi_{2N-1})^{-1} \delta_{2N} + O(t). \end{aligned}$$

□

### 3.2 Projectors

In this paragraph, we shall always suppose that  $\mathcal{I}_{2N-1}$  holds. Another important quantity in our study is the orthogonal projector  $P_{(\text{Im } \Gamma_{tz})^\perp}$  (resp.  $P_{(\text{Im } \Psi_{2N-1})^\perp}$ ) onto  $(\text{Im } \Gamma_{tz})^\perp$  (resp.  $(\text{Im } \Psi_{2N-1})^\perp$ ). We define

$$\begin{aligned} \Pi_{tz} &\stackrel{\text{def.}}{=} P_{(\text{Im } \Gamma_{tz})^\perp} = \text{Id}_{\mathcal{H}} - \Gamma_{tz} (\Gamma_{tz}^* \Gamma_{tz})^{-1} \Gamma_{tz}^*, \\ \Pi_{2N-1} &\stackrel{\text{def.}}{=} P_{(\text{Im } \Psi_{2N-1})^\perp} = \text{Id}_{\mathcal{H}} - \Psi_{2N-1} (\Psi_{2N-1}^* \Psi_{2N-1})^{-1} \Psi_{2N-1}^*. \end{aligned}$$

Observing that  $P_{(\text{Im } \Gamma_{tz})^\perp} = P_{(\text{Im } \Psi_{tz})^\perp}$ , we immediately obtain from the previous Lemma that  $\Pi_{tz} = \Pi_{2N-1} + O(t)$ .

By construction,  $\Pi_{tz} \Phi_{tz} = \Pi_{tz} \Phi'_{tz} = 0$ , but the following proposition shows that this quantity is also small if we replace  $\Phi_{tz}$  with  $\Phi''_{tz}$ .

**Lemma 2** *There exists a constant  $L_1 > 0$  (which only depends on  $\varphi$ ,  $a_0$  and  $z_0$ ) such that*

$$\|\Pi_{tz} \Phi''_{tz}\|_{\mathcal{H}} \leq L_1 t^{2N-2}$$

uniformly in  $z \in \bar{B}_{z_0}$ .

*Proof* Applying a Taylor expansion to  $\varphi^{(2)}$ , we write

$$\Phi''_{tz} = \Psi_{2N-1} \tilde{V}_{tz} + t^{2N-2} \tilde{\Lambda}_{t,z}$$

where

$$\tilde{V}_{tz} = \begin{pmatrix} 0 & \dots & 0 \\ 0 & \dots & 0 \\ 1 & \dots & 1 \\ \vdots & & \vdots \\ \frac{(tz_1)^{2N-3}}{(2N-3)!} & \dots & \frac{(tz_N)^{2N-3}}{(2N-3)!} \end{pmatrix}, \tag{29}$$

$$\tilde{\Lambda}_{t,z} = \left( (z_i)^{2N-2} \int_0^1 \varphi^{(2N)}(stz_i) \frac{(1-s)^{2N-3}}{(2N-3)!} ds \right)_{1 \leq i \leq N}.$$

Hence

$$\begin{aligned} \Pi_{t,z} \Phi''_{t,z} &= \Pi_{t,z} (\Psi_{t,z} \tilde{V}_{t,z} + (\Psi_{N-1} - \Psi_{t,z}) \tilde{V}_{t,z} + t^{2N-2} \tilde{\Lambda}_{t,z}) \\ &= \Pi_{t,z} (-\Lambda_{t,z} D_t H_{t,z}^{-1} \tilde{V}_{t,z} + t^{2N-2} \tilde{\Lambda}_{t,z}) \text{ since } \Pi_{t,z} \Psi_{t,z} = 0. \end{aligned}$$

Using (25), we see that

$$\begin{aligned} D_t H_{t,z}^{-1} \tilde{V}_{t,z} &= H_z^{-1} \text{diag}(t^{2N}, t^{2N-1}, \dots, t) \tilde{V}_{t,z} \\ &= t^{2N-2} H_z^{-1} \tilde{V}_z, \quad \text{hence} \\ \Pi_{t,z} \Phi''_{t,z} &= t^{2N-2} \Pi_{t,z} (-\Lambda_{t,z} H_z^{-1} \tilde{V}_z + \tilde{\Lambda}_{t,z}). \end{aligned}$$

Since  $\|\Pi_{t,z}\| \leq 1$  and the continuous function  $(z, t) \mapsto -\Lambda_{t,z} H_z^{-1} \tilde{V}_z + \tilde{\Lambda}_{t,z}$  is uniformly bounded on the compact set  $\bar{B}_{z_0} \times [0, 1]$ , we obtain

$$\|\Pi_{t,z} \Phi''_{t,z}\|_{\mathcal{H}} \leq \left( \sup_{(z',t') \in \bar{B}_{z_0} \times [0,1]} \|\Lambda_{t',z'} H_{z'}^{-1} \tilde{V}_{z'} + \tilde{\Lambda}_{t',z'}\|_{\mathcal{H}} \right) t^{2N-2}$$

□

We study further the projector  $\Pi_{t,z}$  when it is not evaluated at the same  $z$  as  $\Gamma_{t,z}$ .

**Lemma 3** *If  $\varphi \in \text{KER}^{2N+1}$ , then there is a constant  $L_2 > 0$  (which only depends on  $\varphi, a_0$  and  $z_0$ ) such that for all  $z \in \bar{B}_{z_0}$ , all  $t \in (0, 1]$ ,*

$$\left\| \Pi_{t,z} \Gamma_{t,z_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix} \right\|_{\mathcal{H}} \leq L_2 t^{2N} |z - z_0|_{\infty}.$$

*Proof* Let us observe that

$$\begin{aligned} \Pi_{t,z} \Gamma_{t,z_0} &= \Pi_{t,z} (\Psi_{2N-1} H_{t,z_0} + \Lambda_{t,z_0} D_t) \\ &= \Pi_{t,z} (\Psi_{t,z} H_{t,z_0} + (\Psi_{2N-1} - \Psi_{t,z}) H_{t,z_0} + \Lambda_{t,z_0} D_t) \\ &= \Pi_{t,z} (-\Lambda_{t,z} D_t H_{t,z}^{-1} H_{t,z_0} + \Lambda_{t,z_0} D_t) \text{ since } \Pi_{t,z} \Psi_{t,z} = 0. \end{aligned}$$

Observing that

$$D_t H_{t,z}^{-1} H_{t,z_0} = t^{2N} H_z^{-1} H_{z_0} \text{diag}(1, \dots, 1, 1/t, \dots, 1/t) = H_z^{-1} H_{z_0} D_t,$$

we get

$$\Pi_{t,z} \Gamma_{t,z_0} = \Pi_{t,z} \left( \Lambda_{t,z_0} (\text{Id}_{2N} - H_z^{-1} H_{z_0}) + (\Lambda_{t,z_0} - \Lambda_{t,z}) H_z^{-1} H_{z_0} \right) D_t.$$

For  $k \in \{2N - 1, 2N\}$ , the function  $(u, s, t) \mapsto u^k \varphi^{(2N)}(stu)$  is defined and  $\mathcal{C}^1$  on the compact sets (since  $\varphi \in \text{KER}^{2N+1}$ )

$$\forall 1 \leq i \leq N, \quad \left[ z_{0i} - \frac{\Delta_0}{4}, z_{0i} + \frac{\Delta_0}{4} \right] \times [0, 1] \times [0, 1],$$

where  $\Delta_0$  is defined in (5). Thus there is a constant  $C > 0$  (which does not depend on  $t$  nor  $z \in \overline{B}_{z_0}$ ) such that

$$\left| \int_0^1 \left( (z_i)^k \varphi^{(2N)}(stz_i) - (z_{0i})^k \varphi^{(2N)}(stz_{0i}) \right) \frac{(1-s)^{k-1}}{(k-1)!} ds \right| \leq C |z_i - z_{0i}|,$$

hence  $\|\Lambda_{t,z_0} - \Lambda_{t,z}\| \leq C |z - z_0|_\infty.$

As a result, since  $\|\Pi_{tz}\| \leq 1$  and  $z \mapsto H_z^{-1}H_{z_0}$  is bounded on  $\overline{B}_{z_0}$ ,

$$\left\| \Pi_{tz}(\Lambda_{tz_0} - \Lambda_{t,z})H_z^{-1}H_{z_0} \right\|_{\mathcal{H}} \leq C \sup_{z' \in \overline{B}_{z_0}} \left\| H_{z'}^{-1}H_{z_0} \right\| |z - z_0|_\infty.$$

As for the left term,  $\Lambda_{tz_0}$  is bounded uniformly in  $t \in [0, 1]$ , and the mapping  $z \mapsto H_z^{-1}H_{z_0}$  is  $\mathcal{C}^1$  on  $\overline{B}_{z_0}$ . As a result, there is a constant  $\tilde{C} > 0$  such that

$$\forall z \in \overline{B}_{z_0}, \quad \left\| \text{Id}_N - H_z^{-1}H_{z_0} \right\| \leq \tilde{C} |z - z_0|_\infty.$$

To conclude, we observe that  $D_t \begin{pmatrix} a_0 \\ 0 \end{pmatrix} = t^{2N} \begin{pmatrix} a_0 \\ 0 \end{pmatrix}$ , and we combine the above inequalities to obtain

$$\left\| \Pi_{tz} \Gamma_{tz_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix} \right\|_{\mathcal{H}} \leq \left( C \sup_{z' \in \overline{B}_{z_0}} \left\| H_{z'}^{-1}H_{z_0} \right\| + \tilde{C} \sup_{t \in [0,1]} \left\| \Lambda_{tz_0} \right\| \right) t^{2N} |z - z_0|_\infty.$$

□

### 3.3 Asymptotics of the Vanishing Derivatives Precertificate

We end this section devoted to the asymptotic behavior of quantities related to  $\Gamma_{tz}$  by studying the second derivative of the vanishing derivatives precertificate  $\eta_{V,t}$  (see Definition 2, and [14] for more details). Theorem 1 ensures that the second derivatives of  $\eta_{V,t}$  do not vanish at  $z_{0,i}$ ,  $1 \leq i \leq N$ . However, it does not provide any estimation of those second derivatives. That is the purpose of the next proposition.

In view of Sect. 4, it will be useful to study those second derivatives not only for the precertificates that are defined by interpolating the sign at  $tz_0$  but more generally for the precertificates that are defined to interpolate the sign at  $tz$  for any  $z \in \overline{B}_{z_0}$ .

**Proposition 8** Assume that  $\varphi \in \text{KER}^{2N+1}$  and that  $\mathcal{I}_{2N-1}$  holds. Then

$$\Phi''_{tz} * \Gamma_{tz}^{+,*} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} = t^{2N-2} \eta_W^{(2N)}(0) d_z + \mathcal{O}(t^{2N-1}), \tag{30}$$

where  $d_z \in \mathbb{R}^N$ ,  $d_{z,i} \stackrel{\text{def.}}{=} \frac{2}{(2N)!} \prod_{j \neq i} (z_i - z_j)^2$  for  $1 \leq i \leq N$ . (31)

*Proof* We proceed as in the proof of Lemma 2 by writing  $\Phi''_{tz} = \Psi_{2N-1} \tilde{V}_{tz} + t^{2N-2} \tilde{\Lambda}_{t,z}$  [see (29)] and  $\Psi_{tz} = \Gamma_{tz} H_{tz}^{-1}$ . We obtain

$$\Phi''_{tz} = \Psi_{tz} \tilde{V}_{tz} + t^{2N-2} \tilde{\Lambda}_{t,z} - t^{2N-2} \Lambda_{t,z} H_z^{-1} \tilde{V}_z.$$

The first term yields

$$\tilde{V}_{tz} * \Psi_{tz} * \Gamma_{tz}^{+,*} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} = \tilde{V}_{tz} * \Psi_{tz}^* \Psi_{tz} (\Psi_{tz}^* \Psi_{tz})^{-1} H_{tz}^{-1,*} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} = \tilde{V}_{tz} * \delta_{2N} = 0. \tag{32}$$

As for the second term, we take the Taylor expansion a little further (using integration by parts),

$$\int_0^1 \varphi^{(2N)}(stz_i) \frac{(1-s)^{2N-3}}{(2N-3)!} ds = \frac{\varphi_{2N}}{(2N-2)!} + tz_i \int_0^1 \varphi^{(2N+1)}(stz_i) \frac{(1-s)^{2N-2}}{(2N-2)!} ds,$$

so as to obtain

$$\tilde{\Lambda}_{t,z} = (\varphi_{2N}, \dots, \varphi_{2N}) \text{diag}(e_z) + \mathcal{O}(t), \text{ where } e_z \stackrel{\text{def.}}{=} \left( \frac{(z_i)^{2N-2}}{(2N-2)!} \right)_{1 \leq i \leq N} \in \mathbb{R}^N$$

and, as usual,  $\mathcal{O}(t)$  is uniform in  $z \in \bar{B}_{z_0}$ . From Lemma 1, we also know that  $\Gamma_{tz}^{+,*} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} = p_W + \mathcal{O}(t)$ , hence

$$\tilde{\Lambda}_{t,z} * \Gamma_{tz}^{+,*} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} = \text{diag}(e_z) \begin{pmatrix} \langle \varphi_{2N}, p_W \rangle_{\mathcal{H}} \\ \vdots \\ \langle \varphi_{2N}, p_W \rangle_{\mathcal{H}} \end{pmatrix} + \mathcal{O}(t) = \eta_W^{(2N)}(0) e_z + \mathcal{O}(t). \tag{33}$$

Now, we proceed with the last term. Similarly, by integration by parts,

$$\Lambda_{t,z} = (\varphi_{2N} \dots \varphi_{2N}) \text{diag}(f_z) + \mathcal{O}(t).$$

where  $f_z$  is defined by

$$f_z \stackrel{\text{def.}}{=} \left( \frac{(z_1)^{2N}}{(2N)!}, \dots, \frac{(z_N)^{2N}}{(2N)!}, \frac{(z_1)^{2N-1}}{(2N-1)!}, \dots, \frac{(z_N)^{2N-1}}{(2N-1)!} \right) \in \mathbb{R}^{2N}. \tag{34}$$

Hence,

$$\Lambda_{t,z}^* \Gamma_{tz}^{+,*} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} = \text{diag}(f_z) \begin{pmatrix} \langle \varphi_{2N}, p_W \rangle_{\mathcal{H}} \\ \vdots \\ \langle \varphi_{2N}, p_W \rangle_{\mathcal{H}} \end{pmatrix} + O(t) = \eta_W^{(2N)}(0) f_z + O(t).$$

To conclude, we study  $\tilde{V}_z^* H_z^{-1,*}$  (which is uniformly bounded on  $\overline{B}_{z_0}$ ). In  $\mathbb{R}_{2N-1}[X]$  endowed with the basis  $\left(1, X, \dots, \frac{X^{2N-1}}{(2N-1)!}\right)$ ,  $H_z^*$  is the matrix of the linear map which evaluates a polynomial and its derivatives at  $\{z_1, \dots, z_N\}$ . On the other hand  $\tilde{V}_z^*$  represents the evaluation of the second derivative at  $\{z_1, \dots, z_N\}$ . Thus,

$$\tilde{V}_z^* H_z^{-1,*} f_z = (P''(z_i))_{1 \leq i \leq N},$$

where  $P$  is the unique polynomial in  $\mathbb{R}_{2N-1}[X]$  which satisfies

$$\forall i = 1, \dots, N, \quad P(z_i) = \frac{(z_i)^{2N}}{(2N)!} \quad \text{and} \quad P'(z_i) = \frac{(z_i)^{2N-1}}{(2N-1)!}.$$

One may check that

$$P(X) = \frac{X^{2N}}{(2N)!} - \frac{1}{(2N)!} \prod_{i=1}^N (X - z_i)^2$$

$$\text{and} \quad P''(z_i) = \frac{z_i^{2N-2}}{(2N-2)!} - \frac{2}{(2N)!} \prod_{j \neq i} (z_i - z_j)^2.$$

As a result,

$$- \tilde{V}_z^* H_z^{-1,*} \Lambda_{t,z}^* \Gamma_{tz}^{+,*} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} = \eta_W^{(2N)}(0) (d_z - e_z) + O(t), \tag{35}$$

where  $O(t)$  is uniform in  $z \in \overline{B}_{z_0}$ . We obtain the claimed result by summing (32), (33) and (35). □

### 4 Building a Candidate Solution

Now that the technical issues regarding the asymptotic behavior of  $\Gamma_{tz}$  have been settled, we are ready to tackle the study of the BLASSO. In this section, we build a candidate solution for  $\mathcal{P}_\lambda(y_t + w)$  by relying on its optimality conditions.



### 4.1 First Order Optimality Conditions

The optimality conditions for  $\mathcal{P}_\lambda(y_t + w)$  (see [14]) state that any measure of the form  $m_{a,tz} = \sum_{i=1}^N a_i \delta_{tz_i}$  is a solution to  $\mathcal{P}_\lambda(y_t + w)$  if and only if the function defined by  $\eta_{\lambda,t} \stackrel{\text{def.}}{=} \Phi^* p_{\lambda,t}$  with  $p_{\lambda,t} \stackrel{\text{def.}}{=} \frac{1}{\lambda} (\Phi_{tz_0} a_0 + w - \Phi_{tz} a)$  satisfies  $\|\eta_{\lambda,t}\|_{L^\infty(X)} \leq 1$  and  $\eta_{\lambda,t}(tz_i) = \text{sign}(a_i)$  for all  $1 \leq i \leq N$ .

Observe that we must have  $\eta'_{\lambda,t}(tz_i) = 0$  for all  $1 \leq i \leq N$ . Moreover, in our case, since we assume that  $a_{0,i} > 0$  we have in fact  $\eta_{\lambda,t}(tz_i) = 1$ .

In order to build such a function  $\eta_{\lambda,t}$ , let us consider the function  $f_t$  defined for some fixed  $t > 0$  on  $(\mathbb{R}^N)^2 \times \mathbb{R} \times \mathcal{H}$  by

$$f_t(u, v) \stackrel{\text{def.}}{=} \Gamma_{tz}^* (\Phi_{tz} a - \Phi_{tz_0} a_0 - w) + \lambda \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} \tag{36}$$

$$\text{where } u = (a, z) \text{ and } v = (\lambda, w). \tag{37}$$

Now, let us write  $u_0 \stackrel{\text{def.}}{=} (a_0, z_0)$ . Notice that  $m_{a,tz}$  is a solution to  $\mathcal{P}_\lambda(y_t + w)$  if and only if  $f_t(u, v) = 0$  and  $\|\eta_{\lambda,t}\|_{L^\infty(X)} \leq 1$ . Our strategy is therefore to construct solutions of  $f_t(u, v) = 0$  and to prove that  $\|\eta_{\lambda,t}\|_{L^\infty(X)} \leq 1$  provided  $(\lambda, w)$  and  $\eta_W$  satisfy certain properties. More precisely we start by parametrizing the solutions of  $f_t(u, v) = 0$ , in a neighborhood of  $(u_0, 0)$ , using the Implicit Function Theorem.

The following Lemma 4 (whose proof is omitted and corresponds to simple computations) shows that  $f_t$  is smooth and gives its derivatives.

**Lemma 4** *If  $\varphi \in \text{KER}^{k+1}$  for some  $k \in \mathbb{N}^*$  then  $f_t$  is of class  $\mathcal{C}^k$  and for all  $(u, v) \in (\mathbb{R}^N)^2 \times (\mathbb{R} \times \mathcal{H})$*

$$\begin{aligned} \partial_u f_t(u, v) &= \Gamma_{tz}^* \Gamma_{tz} J_{ta} + t \begin{pmatrix} 0 & \text{diag}(\Phi_{tz}^* (\Phi_{tz} a - \Phi_{tz_0} a_0 - w)) \\ 0 & \text{diag}(\Phi_{tz}^* (\Phi_{tz} a - \Phi_{tz_0} a_0 - w)) \end{pmatrix} \\ \partial_v f_t(u, v) &= \left( \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix}, -\Gamma_{tz}^* \right) \end{aligned}$$

where  $J_{ta} \stackrel{\text{def.}}{=} \begin{pmatrix} \text{Id}_N & 0 \\ 0 & t \text{diag}(a) \end{pmatrix}$ .

### 4.2 Implicit Function Theorem

Suppose that  $\mathcal{I}_{2N-1}$  holds and  $\varphi \in \text{KER}^{2N+1}$ . By the results of Sect. 3, there exists  $0 < t_0 < 1$  such that for  $0 < t < t_0$  and all  $z \in \bar{\mathcal{B}}_{z_0}$ ,  $\Gamma_{tz}^* \Gamma_{tz}$  is invertible. In the following we shall consider a fixed value of such  $t_0$  provided by Lemma 5 below which also ensures additional properties.

Now, let  $t \in (0, t_0)$  be fixed. By Lemma 4,  $f_t$  is  $\mathcal{C}^{2N}$ ,  $\partial_u f_t(u_0, 0) = \Gamma_{tz_0}^* \Gamma_{tz_0} J_{ta_0}$  is invertible and  $f_t(u_0, 0) = 0$ . Hence by the Implicit Function Theorem, there exists  $V_t$  a neighborhood of 0 in  $\mathbb{R} \times \mathcal{H}$ ,  $U_t$  a neighborhood of  $u_0$  in  $(\mathbb{R}^N)^2$  and  $g_t : V_t \rightarrow U_t$

a  $\mathcal{C}^{2N}$  function such that

$$\forall(u, v) \in U_t \times V_t, \quad f_t(u, v) = 0 \iff u = g_t(v).$$

Moreover, denoting  $dg_t$  the differential of  $g_t$ , we have

$$\forall v \in V_t, \quad dg_t(v) = -(\partial_u f_t(g_t(v), v))^{-1} \partial_v f_t(g_t(v), v).$$

### 4.3 Extension of the Implicit Function $g_t$

Our goal is to prove that  $m_{a,tz}$  is the solution of the BLASSO, where  $(a, z) = u = g_t(v)$ . To this end, we shall exhibit additional constraints on  $v \in V_t$ , such as the scaling of the noise  $\|w\|_{\mathcal{H}}$  or  $\lambda$  with respect to  $t$ , in order to ensure that  $\|\eta_{\lambda,t}\|_{L^\infty(X)} \leq 1$ . However, the size of the neighborhood  $V_t$  provided by the Implicit Function Theorem is *a priori* unknown, and it might implicitly impose even stronger conditions on  $\lambda$  and  $w$  as  $t \rightarrow 0^+$ .

Hence, before studying whether  $\|\eta_{\lambda,t}\|_{L^\infty(X)} \leq 1$ , we show in this section that we may replace  $V_t$  with some ball with radius of order  $t^{2N-1}$  and still have a parametrization of the form  $u = g_t(v)$  satisfying  $f_t(g_t(v), v) = 0$  where  $f_t$  is defined in (36).

Let  $V_t^* = \bigcup_{V \in \mathcal{V}} V$ , where  $\mathcal{V}$  is the collection of all open sets  $V \subset \mathbb{R} \times \mathcal{H}$  such that

- $0 \in V$ ,
- $V$  is star-shaped with respect to 0,
- $V \subset B(0, C_T t^{2N-2})$ , where  $C_T > 0$  is a constant defined by Lemma 5 below,
- there exists a  $\mathcal{C}^{2N}$  function  $g : V \rightarrow (\mathbb{R}^N)^2$  such that  $g(0) = u_0$  and  $f_t(g(v), v) = 0$  for all  $v \in V$ ,
- $g(V) \subset \mathcal{B}_{a_0} \times \mathcal{B}_{z_0}$ .

Observe that  $\mathcal{V}$  is nonempty (by the Implicit Function Theorem in Sect. 4.2) and stable by union, so that  $V_t^* \in \mathcal{V}$ . Indeed, all the properties defining  $\mathcal{V}$  are easy to check except possibly the last two. Let  $V, \tilde{V} \in \mathcal{V}$  and  $g, \tilde{g}$  be corresponding functions. The set  $\{v \in V \cap \tilde{V} ; g(v) = \tilde{g}(v)\}$  is nonempty (because  $g(0) = u_0 = \tilde{g}(0)$ ) and closed in  $V \cap \tilde{V}$ . Moreover, it is open since for any  $v \in V \cap \tilde{V}$ ,  $v \in B(0, C_T t^{2N-2})$  and, by Lemma 5 below, the Implicit Function Theorem applies at  $(g(v), v)$ , yielding an open neighborhood in which  $g$  and  $\tilde{g}$  coincide. By connectedness of  $V \cap \tilde{V}$ ,  $g$  and  $\tilde{g}$  coincide in the whole set  $V \cap \tilde{V}$ . As a result, the function  $g_t^* : V_t^* \rightarrow (\mathbb{R}^N)^2$ , defined by

$$g_t^*(v) \stackrel{\text{def.}}{=} g(v) \text{ if } v \in V, V \in \mathcal{V}, \text{ and } g \text{ is a corresponding function,} \tag{38}$$

is well defined. Moreover,  $g_t^*$  is  $\mathcal{C}^{2N}$  and  $g_t^*(V_t^*) \subset \mathcal{B}_{a_0} \times \mathcal{B}_{z_0}$ .

Before proving that  $V_t^*$  contains a ball of radius of order  $t^{2N-1}$  and studying the variations of  $g_t^*$ , we state Lemma 5 mentioned above.

**Lemma 5** *If  $\mathcal{I}_{2N-1}$  holds and  $\varphi \in \text{KER}^{2N+1}$  then there exists  $t_0 \in (0, 1)$  and  $C_T > 0$  (which only depend on  $\varphi$  and  $u_0$ ) such that for all  $t \in (0, t_0)$ , if*

$$u = (a, z) \in \overline{\mathcal{B}}_{a_0} \times \overline{\mathcal{B}}_{z_0} \quad \text{and} \quad v = (\lambda, w) \in \mathbf{B}\left(0, C_T t^{2N-2}\right), \tag{39}$$

then the matrix

$$G_{tz}(\lambda, w) \stackrel{\text{def.}}{=} \Psi_{tz}^* \Psi_{tz} + t H_{tz}^{*-1} F_{tz} J_{ta}^{-1} H_{tz}^{-1}, \tag{40}$$

$$\text{where } F_{tz} \stackrel{\text{def.}}{=} \begin{pmatrix} 0 & 0 \\ -\text{diag}(\Phi_{tz}^{/'*} q_{tz}) & \end{pmatrix}, \tag{41}$$

$$\text{and } q_{tz} \stackrel{\text{def.}}{=} \lambda \Gamma_{tz}^{*+} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} + \Pi_{tz} w + \Pi_{tz} \Gamma_{tz_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix}, \tag{42}$$

is invertible and the norm of its inverse is less than  $3 \left\| (\Psi_{2N-1}^* \Psi_{2N-1})^{-1} \right\|$ .  
 If, moreover,  $f_t(u, v) = 0$ , then

$$\partial_u f_t(u, v) = H_{tz}^* G_{tz}(\lambda, w) H_{tz} J_{ta}$$

and this is an invertible matrix.

Let us precise that by  $(\lambda, w) \in \mathbf{B}(0, C_T t^{2N-2})$ , we mean that  $|\lambda| < C_T t^{2N-2}$  and  $\|w\|_{\mathcal{H}} < C_T t^{2N-2}$ .

*Proof* We consider  $t_0 \in (0, 1)$  small enough so that for  $0 < t < t_0$  and all  $z \in \overline{\mathcal{B}}_{z_0}$ ,  $\Gamma_{tz}^* \Gamma_{tz}$  is invertible and

$$\left\| (\Psi_{tz}^* \Psi_{tz})^{-1} \right\| \leq 2 \left\| (\Psi_{2N-1}^* \Psi_{2N-1})^{-1} \right\| \text{ by Lemma 1,} \tag{43}$$

$$\left| \lambda \Phi_{tz}^{/'*} \Gamma_{tz}^{*+} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} \right|_{\infty} \leq \frac{4(2R_W)^{2N}}{(2N)!} \left| \lambda \eta_W^{(2N)}(0) \right| t^{2N-2} \text{ by Proposition 8.} \tag{44}$$

In the last equation, we have used the fact that  $\left| \prod_{j \neq i} (z_i - z_j) \right| \leq (2R_W)^{2N}$  (where  $R_W = \sup\{|z|_{\infty}; z \in \overline{\mathcal{B}}_{z_0}\}$  is defined in Eq. (19)).

We also know that for some constants  $L_1, L_2 > 0$  which only depend on  $\varphi$  and  $u_0$ ,

$$\begin{aligned} \left| \Phi_{tz}^{/'*} \Pi_{tz} w \right|_{\infty} &\leq \|w\|_{\mathcal{H}} L_1 t^{2N-2} \text{ by Lemma 2,} \\ \text{and } \left| \Phi_{tz}^{/'*} \Pi_{tz} \Gamma_{tz_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix} \right|_{\infty} &\leq L_1 L_2 \frac{\Delta_0}{4} t^{4N-2} \text{ by Lemmas 2 and 3,} \end{aligned}$$

for all  $z \in \overline{\mathcal{B}}_{z_0} = \overline{\mathbf{B}}\left(z_0, \frac{\Delta_0}{4}\right)$ ,  $t \in (0, t_0)$ , where  $\Delta_0$  is defined in (5).

Combining those inequalities with (44), we see that

$$\begin{aligned} |\Phi''_{tz} q_{tz}|_\infty &= \left| \lambda \Phi''_{tz} \Gamma_{tz}^{*,+} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} + \Phi''_{tz} \Pi_{tz} w + \Phi''_{tz} \Pi_{tz} \Gamma_{tz0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix} \right|_\infty \\ &\leq |\lambda| \frac{4}{(2N)!} \left| \eta_W^{(2N)}(0) \right| t^{2N-2} + \|w\|_{\mathcal{H}} L_1 t^{2N-2} + L_1 L_2 \frac{\Delta_0}{4} t^{4N-2} \end{aligned}$$

On the other hand, since

$$\begin{aligned} H_{tz}^{*, -1} &= \text{diag} \left( 1, \dots, 1/t^{2N-1} \right) H_z^{*, -1} \text{diag} \left( 1, \dots, 1, t, \dots, t \right), \\ \text{and } F_{tz} &= \begin{pmatrix} 0 & 0 \\ 0 & -\text{diag} \left( \Phi''_{tz} q_{tz} \right) \end{pmatrix}, \end{aligned}$$

we get

$$\begin{aligned} H_{tz}^{*, -1} F_{tz} J_{ta}^{-1} H_{tz}^{-1} &= t \text{diag} \left( 1, \dots, \frac{1}{t^{2N-1}} \right) H_z^{*, -1} F_{tz} \\ &\quad J_a^{-1} H_z^{-1} \text{diag} \left( 1, \dots, \frac{1}{t^{2N-1}} \right) \end{aligned}$$

so that

$$\begin{aligned} \left\| t H_{tz}^{*, -1} F_{tz} J_{ta}^{-1} H_{tz}^{-1} \right\| &\leq \frac{|a^{-1}|_\infty}{t^{4N-4}} \left\| H_z^{*, -1} \right\| \left\| H_z^{-1} \right\| \left| \Phi''_{tz} q_{tz} \right|_\infty \\ &\leq C \left( \frac{|\lambda|}{t^{2N-2}} \frac{4(2R)^{2N}}{(2N)!} \left| \eta_W^{(2N)}(0) \right| \right. \\ &\quad \left. + \frac{\|w\|_{\mathcal{H}}}{t^{2N-2}} L_1 + L_2 L_1 \frac{\Delta_0}{4} t^2 \right) \end{aligned}$$

with  $C = \sup_{(a,z) \in \bar{B}_{a_0} \times \bar{B}_{z_0}} \left\| H_z^{*, -1} \right\| \left\| H_z^{-1} \right\| |a^{-1}|_\infty$ .

Possibly choosing  $t_0$  a bit smaller, we may assume that

$$0 \leq C L_2 L_1 \frac{\Delta_0}{4} t_0^2 < \frac{1}{8 \left\| (\Psi_{2N-1}^* \Psi_{2N-1})^{-1} \right\|}.$$

As a consequence, there exists  $C_T > 0$  such that for all  $t \in (0, t_0)$ , and all  $(a, z) \in \bar{B}_{a_0} \times \bar{B}_{z_0}$ ,

$$\begin{aligned} \left( \max \left( \frac{|\lambda|}{t^{2N-2}}, \frac{\|w\|_{\mathcal{H}}}{t^{2N-2}} \right) \right) &\leq C_T \\ \implies \left\| t H_{tz}^{*, -1} F_{tz} J_{ta}^{-1} H_{tz}^{-1} \right\| &\leq \frac{1}{4 \left\| (\Psi_{2N-1}^* \Psi_{2N-1})^{-1} \right\|}. \end{aligned}$$

Then, recalling (43) and setting

$$r \stackrel{\text{def.}}{=} \left\| t (\Psi_{Iz}^* \Psi_{Iz})^{-1} H_{Iz}^{*, -1} F_{Iz} J_{Ia}^{-1} H_{Iz}^{-1} \right\| \leq 2 \left\| (\Psi_{2N-1}^* \Psi_{2N-1})^{-1} \right\| \frac{1}{4 \left\| (\Psi_{2N-1}^* \Psi_{2N-1})^{-1} \right\|} = \frac{1}{2},$$

we see that the matrix (40) is invertible, and

$$\left\| \left( \text{Id}_{2N} + t (\Psi_{Iz}^* \Psi_{Iz})^{-1} H_{Iz}^{*, -1} F_{Iz} J_{Ia}^{-1} H_{Iz}^{-1} \right)^{-1} \right\| \leq \sum_{k=0}^{+\infty} r^k = \frac{1}{1-r} \leq \frac{3}{2}.$$

Eventually, using (43) again, we obtain that the norm of the inverse of (40) is less than  $3 \left\| (\Psi_{2N-1}^* \Psi_{2N-1})^{-1} \right\|$ .

Now, if  $f_t(u, v) = 0$ , then  $\Phi_{Iz}^{*'}(\Phi_{Iz} a - \Phi_{Iz_0} a_0 - w) = 0$ , so that thanks to Lemma 4 we obtain

$$\partial_u f_t(u, v) = \Gamma_{Iz}^* \Gamma_{Iz} J_{Ia} + t \begin{pmatrix} 0 & 0 \\ 0 & \text{diag}(\Phi_{Iz}^{*'}(\Phi_{Iz} a - \Phi_{Iz_0} a_0 - w)) \end{pmatrix}.$$

Moreover,

$$\begin{aligned} \Phi_{Iz} a - \Phi_{Iz_0} a_0 - w &= \Gamma_{Iz} \begin{pmatrix} a \\ 0 \end{pmatrix} - \Gamma_{Iz_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix} - w \\ &= \Gamma_{Iz}^{*, +} \Gamma_{Iz} \Gamma_{Iz_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix} + \Gamma_{Iz}^{*, +} \Gamma_{Iz}^* w - \lambda \Gamma_{Iz}^{*, +} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} \\ &\quad - \Gamma_{Iz_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix} - w \\ &= -\Pi_{Iz} \Gamma_{Iz_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix} - \Pi_{Iz} w - \lambda \Gamma_{Iz}^{*, +} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} = -q_{Iz}. \end{aligned}$$

As a result,

$$\begin{aligned} \partial_u f_t(u, v) &= H_{Iz}^* \Psi_{Iz}^* \Psi_{Iz} H_{Iz} J_{Ia} + t \begin{pmatrix} 0 & 0 \\ 0 & -\text{diag}(\Phi_{Iz}^{*'} q_{Iz}) \end{pmatrix} \\ &= H_{Iz}^* G_{Iz}(\lambda, w) H_{Iz} J_{Ia}, \end{aligned}$$

and  $\partial_u f_t(u, v)$  is invertible. □

We may now study the variations of  $g_t^*$ .

**Corollary 1** *If  $\mathcal{I}_{2N-1}$  holds and  $\varphi \in \text{KER}^{2N+1}$  then there exists  $M > 0$  (which only depends on  $\varphi$  and  $u_0$ ), such that for  $0 < t < t_0$ , for all  $v \in V_t^*$*

$$\|dg_t^*(v)\| \leq \frac{M}{t^{2N-1}}.$$

*Proof* Let us recall that by construction,  $V_t^* \subset B(0, C_T t^{2N-2})$ . Thus, from Lemma 5, we know that for all  $v \in V_t^*$ ,  $\partial_u f_t(g_t(v), v) = H_{tz}^* G_{tz}(\lambda, w) H_{tz} J_{ta}$ , where  $(a, z) = g_t^*(v)$ . Since  $dg_t^*(v) = -(\partial_u f_t(g_t^*(v), v))^{-1} \partial_v f_t(g_t^*(v), v)$ , we get

$$\begin{aligned} dg_t^*(v) &= -J_{ta}^{-1} H_{tz}^{-1} G_{tz}(\lambda, w)^{-1} H_{tz}^{*-1} \left( \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix}, -H_{tz}^* \Psi_{tz}^* \right) \\ &= J_a^{-1} H_z^{-1} \text{diag} \left( 1, \dots, \frac{1}{t^{2N-1}} \right) G_{tz}(\lambda, w)^{-1} (\delta_{2N}, \Psi_{tz}^*), \end{aligned}$$

Using Lemma 5 and the fact that  $J_a^{-1}, H_z^{-1}, \Psi_{tz}^*$  are uniformly bounded on  $\overline{B_{a_0}} \times \overline{B_{z_0}}$ , we obtain the claimed upper bound of  $\|dg_t^*(v)\|$  for all  $v \in V_t^*$ .  $\square$

We are now in position to prove that  $V_t^*$  contains a ball of radius of order  $t^{2N-1}$ .

**Proposition 9** *If  $\mathcal{I}_{2N-1}$  holds and  $\varphi \in \text{KER}^{2N+1}$ , there exists  $C_R > 0$  such that for all  $t \in (0, t_0)$ ,*

$$B\left(0, C_R t^{2N-1}\right) \subset V_t^* \quad \text{with} \quad C_R \geq \min\left(\frac{\Delta_0}{4M}, \frac{\min_i(a_{0,i})}{4M}, \frac{C_T}{t_0}\right).$$

*Proof* Let  $v \in \mathbb{R} \times \mathcal{H}$  with unit norm (i.e.  $\max(\lambda, \|w\|_{\mathcal{H}}) = 1$ ), and define

$$R_v \stackrel{\text{def.}}{=} \sup \{r \geq 0 ; rv \in V_t^*\}.$$

Clearly  $0 < R_v \leq C_T t^{2N-2}$ . Assume that  $R_v < C_T t^{2N-2}$ . Then by Corollary 1,  $g_t^*$  is uniformly continuous on  $V_t^*$ , so that the value of  $g_t^*(R_v v)$  can be defined as a limit, and  $f_t(g_t^*(R_v v), R_v v) = 0$ .

By contradiction, if  $g_t^*(R_v v) \in B_{a_0} \times B_{z_0}$ , then by Lemma 5, we may apply the Implicit Function Theorem to obtain a neighborhood of  $(g_t^*(R_v v), R_v v)$  in which  $g_t^*$  may be extended. This enables us to construct an open set  $V \in \mathcal{V}$  (in particular we may ensure that  $V$  is star-shaped with respect to 0) such that  $V_t^* \subsetneq V$ , which contradicts the maximality of  $V_t^*$ .

Hence,  $g_t^*(R_v v) \in \partial(B_{a_0} \times B_{z_0}) = (\partial(B_{a_0}) \times \overline{B_{z_0}}) \cup (\overline{B_{a_0}} \times \partial(B_{z_0}))$ . Assume for instance that  $g_t^*(R_v v) \in \overline{B_{a_0}} \times \partial(B_{z_0})$  (the other case being similar). Then, for  $(a, z) = g_t^*(R_v v)$ ,

$$\frac{\Delta_0}{4} = |z - z_0|_\infty \leq \int_0^1 |dg_t^*(sR_v v) \cdot R_v v|_\infty ds \leq \frac{M}{t^{2N-1}} R_v,$$

which yields  $R_v \geq \frac{\Delta_0}{4M} t^{2N-1}$ . Similarly, if  $g_t^*(R_v v) \in \partial(B_{a_0}) \times \overline{B_{z_0}}$ , we may prove that  $R_v \geq \frac{\min_i(a_{0,i})}{4M} t^{2N-1}$ .

Eventually, we have proved that for all  $v \in \mathbb{R} \times \mathcal{H}$  with unit norm,

$$R_v \geq \min\left(\frac{\Delta_0}{4M} t^{2N-1}, \frac{\min_i(a_{0,i})}{4M} t^{2N-1}, C_T t^{2N-2}\right),$$

and the claimed result follows.  $\square$

### 4.4 Continuity of $g_t^*$ at 0

Before moving to Sect. 5 and to the proof of  $\|\eta_{\lambda,t}\|_{L^\infty(X)} \leq 1$  (which ensures that  $m_{a,tz}$  is a solution to the BLASSO), we give a first order expansion of our candidate solution  $u = g_t^*(v)$  for all  $v \in B(0, C_R t^{2N-1})$ .

**Proposition 10** *If  $\mathcal{I}_{2N-1}$  holds and  $\varphi \in \text{KER}^{2N+1}$  then for all  $t \in (0, t_0)$ ,  $v \in B(0, C_R t^{2N-1})$ ,*

$$\|g_t^*(v) - g_t^*(0)\| \leq M \left( \frac{|\lambda|}{t^{2N-1}} + \frac{\|w\|_{\mathcal{H}}}{t^{2N-1}} \right). \tag{45}$$

*Proof* To show Eq. (45), it suffices to write

$$g_t^*(v) = g_t^*(0) + \int_0^1 dg_t^*(sv) \cdot v ds,$$

and use Corollary 1 to conclude. □

### 5 Convergence of $\eta_{\lambda,t}$ to $\eta_W$

It remains to prove that  $m_{a,tz}$  where  $(a, z) = g_t^*(v)$  is indeed a solution to  $\mathcal{P}_\lambda(y_t + w)$ . To show this statement, we prove that  $\eta_{\lambda,t}$  converges towards  $\eta_W$  when  $(t, \lambda, w) \rightarrow 0$  in a well chosen domain. This section is devoted to this result. The proof of our main contribution, Theorem 2, which can be found in Sect. 2.3, uses this convergence result and the assumption that  $\eta_W$  is  $(2N - 1)$ -non-degenerate to conclude.

**Proposition 11** *Assume that  $\varphi \in \text{KER}^{2N+1}$  and that  $\mathcal{I}_{2N-1}$  holds, and let  $C_W > 0$  be the constant defined in Theorem 1,  $g_t^*$ ,  $t_0 > 0$  and  $C_R > 0$  be the function and constants defined in Sect. 4.*

*Then there exist constants  $t_1 \in (0, t_0)$  and  $C > 0$  (which depend only on  $\varphi$  and  $u_0$ ) such that for all  $t \in (0, t_1)$  and for all  $(\lambda, w) \in B(0, C_R t^{2N-1})$  with  $\|\frac{w}{\lambda}\|_{\mathcal{H}} \leq C$ , the following inequalities hold*

$$\forall \ell \in \{0, \dots, 2N\}, \quad \left\| \eta_{\lambda,t}^{(\ell)} - \eta_W^{(\ell)} \right\|_{L^\infty(X)} \leq C_W,$$

with  $\eta_{\lambda,t} = \Phi^* \left( \frac{1}{\lambda} (\Phi_{t z_0} a_0 + w - \Phi_{t z} a) \right)$  and  $(a, z) = g_t^*(\lambda, w)$ .

*Proof* Let  $t \in (0, t_0)$ ,  $v \in B(0, C_R t^{2N-1})$ , and  $(a, z) = u = g_t^*(v)$ . Then, using  $f_t(u, v) = 0$  [see (36)], we get

$$\begin{aligned}
 p_{\lambda,t} &\stackrel{\text{def.}}{=} \frac{1}{\lambda} (\Phi_{t z_0} a_0 + w - \Phi_{t z} a) \\
 &= \frac{1}{\lambda} \left( \Gamma_{t z_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix} + w - \Gamma_{t z} \begin{pmatrix} a \\ 0 \end{pmatrix} \right) \\
 &= \Gamma_{t z}^{*,+} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} + \Pi_{t z} \frac{w}{\lambda} + \frac{1}{\lambda} \Pi_{t z} \Gamma_{t z_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix}.
 \end{aligned}$$

Hence,

$$\|p_{\lambda,t} - p_W\|_{\mathcal{H}} \leq \left\| \Gamma_{t z}^{*,+} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} - p_W \right\|_{\mathcal{H}} + \left\| \Pi_{t z} \frac{w}{\lambda} \right\|_{\mathcal{H}} + \left\| \frac{1}{\lambda} \Pi_{t z} \Gamma_{t z_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix} \right\|_{\mathcal{H}}$$

From Lemma 1, there exists  $C_V > 0$  and  $t_V > 0$  (which only depends on  $\varphi, u_0$ ) such that for all  $t \in (0, t_V)$  and all  $z \in \overline{B}_{z_0}$ ,

$$\left\| \Gamma_{t z}^{*,+} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} - p_W \right\|_{\mathcal{H}} \leq C_V t.$$

Moreover, since  $\Pi_{t z}$  is an orthogonal projector,

$$\left\| \Pi_{t z} \frac{w}{\lambda} \right\|_{\mathcal{H}} \leq \left\| \frac{w}{\lambda} \right\|_{\mathcal{H}},$$

and by Lemma 3, there exists  $L_2 > 0$  (which only depends on  $\varphi, u_0$ ) such that, for all  $t \in (0, t_0)$

$$\begin{aligned}
 \left\| \frac{1}{\lambda} \Pi_{t z} \Gamma_{t z_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix} \right\|_{\mathcal{H}} &\leq \frac{L_2}{|\lambda|} t^{2N} |z - z_0|_{\infty} \\
 &\leq \frac{L_2}{|\lambda|} t^{2N} \frac{M}{t^{2N-1}} (|\lambda| + \|w\|_{\mathcal{H}}) \text{ by Proposition 10} \\
 &\leq L_2 M t \left( 1 + \left\| \frac{w}{\lambda} \right\|_{\mathcal{H}} \right).
 \end{aligned}$$

Gathering all these upper-bounds, one obtains

$$\begin{aligned}
 \|p_{\lambda,t} - p_W\|_{\mathcal{H}} &\leq (C_V + L_2 M)t + (1 + L_2 M t) \left\| \frac{w}{\lambda} \right\|_{\mathcal{H}} \\
 &\leq (C_V + L_2 M)t + (1 + L_2 M) \left\| \frac{w}{\lambda} \right\|_{\mathcal{H}}
 \end{aligned}$$

Now, denoting by  $\Phi^{(\ell)} : \mathcal{M}(X) \rightarrow \mathcal{H}$  the operator  $m \mapsto \int_X \varphi^{(\ell)}(x) dm(x)$  and by  $\Phi^{(\ell)*} : \mathcal{H} \rightarrow \mathcal{C}_X$  its adjoint (so that  $\eta_{\lambda,t}^{(\ell)} = \Phi^{(\ell)*} p_{\lambda,t}$  and  $\eta_W^{(\ell)} = \Phi^{(\ell)*} p_W$ ), we let

$$K \stackrel{\text{def.}}{=} \max_{0 \leq \ell \leq 2N} \sup_{x \in X} \left\| \varphi^{(\ell)}(x) \right\|_{\mathcal{H}},$$



which satisfies  $K < +\infty$  because  $\varphi \in \text{KER}^{2N+1}$ . Then, for all  $\ell \in \{0, \dots, 2N\}$ ,

$$\begin{aligned} \left\| \eta_{\lambda,t}^{(\ell)} - \eta_W^{(\ell)} \right\|_{L^\infty(X)} &\leq K \left\| p_{\lambda,t} - p_W \right\|_{\mathcal{H}} \\ &\leq K \left( (C_V + L_2M)t + (1 + L_2M) \left\| \frac{w}{\lambda} \right\|_{\mathcal{H}} \right). \end{aligned}$$

As a consequence, by taking  $t$  smaller than  $\min(t_0, t_V, \frac{C_W}{2K(C_V + L_2M)})$  and for all  $(\lambda, w) \in B(0, C_R t^{2N-1})$  such that

$$(1 + L_2M) \left\| \frac{w}{\lambda} \right\|_{\mathcal{H}} \leq \frac{C_W}{2K},$$

we get

$$\left\| \eta_{\lambda,t}^{(\ell)} - \eta_W^{(\ell)} \right\|_{L^\infty(X)} \leq C_W.$$

□

*Remark 4* The constants involved in Proposition 11 are

$$t_1 \stackrel{\text{def.}}{=} \min \left( t_0, t_V, \frac{C_W}{2K(C_V + L_2M)} \right) \quad \text{and} \quad C \stackrel{\text{def.}}{=} \frac{C_W}{2K(1 + L_2M)}. \tag{46}$$

They only depend on  $\varphi$  and  $u_0$ .

### 6 Conclusion

In this paper, we have proposed a detailed analysis of the recovery performance of the BLASSO for positive measures. We have shown that if the signal-to-noise ratio is of the order of  $1/t^{2N-1}$ , then the BLASSO achieves perfect support estimation. This results nicely matches both Cramer-Rao lower-bounds [2], bounds for combinatorial approaches [10] and practical performances of the MUSIC algorithm [7]. It is also close to the upper-bound obtained by [20] for stable recovery on a grid by the LASSO. We have hence shown that these bounds do not only imply stability, they imply exact support recovery for the BLASSO, if  $\eta_W$  is  $(2N - 1)$ -non-degenerate. We have observed numerically that this condition holds for the ideal low-pass filter and that this is the case for a large class of low-pass filters. We have showed that this is true for the Gaussian kernel by providing the expression of  $\eta_W$ .

**Acknowledgements** We would like to thank Laurent Demanet for his initial questions that motivated us to prove the main result of this paper. We would also like to thank Jean-Marie Mirebeau for stimulating discussions about the intriguing structure of the inverse of checkerboard matrices. This work has been partly supported by the European Research Council (ERC Project SIGMA-Vision).

## Appendix 1: Proof of the Results of Sect. 2

### Appendix 1.1: Proof of Proposition 1

The functions  $(\varphi_0, \dots, \varphi_k)$  are linearly independent in  $L^2(\mathbb{T})$  if and only if their respective Fourier coefficients are linearly independent in  $\ell^2(\mathbb{Z})$ . If  $(c_n[\varphi_0])_{n \in \mathbb{Z}}$  denotes the Fourier coefficients of  $\varphi_0$ , the Fourier coefficients of  $\varphi_j$  are given by  $((2i\pi n)^j c_n[\varphi_0])_{n \in \mathbb{Z}}$  (with the convention that  $0^0 = 1$ ).

If  $\varphi_0$  has  $k + 1$  nonzeros Fourier coefficients corresponding to pairwise distinct frequencies  $(n_0, \dots, n_k)$ , those Fourier coefficients are given by the matrix product

$$\begin{pmatrix} c_{n_0}[\varphi_0] & & & & 0 \\ & \ddots & & & \\ & & \ddots & & \\ 0 & & & c_{n_k}[\varphi_0] & \end{pmatrix} \begin{pmatrix} 1 & (2i\pi n_0) & (2i\pi n_0)^2 & \dots & (2i\pi n_0)^k \\ 1 & (2i\pi n_1) & (2i\pi n_1)^2 & \dots & (2i\pi n_1)^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (2i\pi n_k) & (2i\pi n_k)^2 & \dots & (2i\pi n_k)^k \end{pmatrix}$$

Both the diagonal and the Vandermonde matrices are invertible, hence the family of Fourier coefficients of  $(\varphi_0, \dots, \varphi_k)$  is linearly independent.

Conversely, if  $\mathcal{I}_k$  holds, one can find  $k + 1$  Fourier coefficients, corresponding to some frequencies  $n_0, \dots, n_k$ , such that the matrix  $(c_{n_\ell}[\varphi_j])_{0 \leq \ell, j \leq k}$  is invertible. From the above factorization, we deduce that each  $c_{n_\ell}[\varphi_0]$  must be nonzero for  $0 \leq \ell \leq k$ .

### Appendix 1.2: Proof of Theorem 1

The proof proceeds in two steps. First we show the result locally around 0 in  $X$  thanks to  $\eta_W^{(2N)}(0) < 0$  (because  $\eta_W^{(2N)}(0) \neq 0$  and  $|\eta_W| < 1$  on  $X \setminus \{0\}$ ) and then we extend the result to all  $X$  thanks to  $|\eta_W| < 1$  on  $X \setminus \{0\}$ .

*Locally*

Let us prove that there exist  $C_W > 0, t_W > 0$  such that for all  $t \in (0, t_W), z \in \mathbb{R}^N$  with pairwise distinct coordinates and  $|z|_\infty \leq R_W$ , there exist  $r^+ > 0$  with  $r^+ > \max_{1 \leq i \leq N} t_W z_i$  and  $r^- < 0$  with  $r^- < \min_{1 \leq i \leq N} t_W z_i$  such that for all  $\eta \in \mathcal{C}^{2N}(X) \cap W^{2N, \infty}(X)$  satisfying for all  $1 \leq i \leq N, \eta(tz_i) = 1$  and  $\eta'(tz_i) = 0$ , the following implication holds,

$$\left( \forall \ell \in \{0, \dots, 2N\}, \left\| \eta^{(\ell)} - \eta_W^{(\ell)} \right\|_{L^\infty(X)} \leq C_W \right) \implies \left( \forall x \in (r^-, r^+) \setminus \bigcup_i \{tz_i\}, |\eta(x)| < 1 \text{ and } \forall 1 \leq i \leq N, \eta''(tz_i) < 0 \right).$$

First, we prove that, provided  $C_W > 0, t_W > 0$  are small enough,  $\eta'$  has exactly  $2N - 1$  zeros in  $(r^-, r^+)$ .

Let  $t > 0$  and  $\eta \in \mathcal{C}^{2N}(X) \cap W^{2N,\infty}(X)$  and  $z \in \mathbb{R}^N$  with pairwise distinct coordinates and  $|z|_\infty \leq R_W$  such that for all  $1 \leq i \leq N$ ,  $\eta(tz_i) = 1$  and  $\eta'(tz_i) = 0$ . We suppose that  $z_1 < z_2 < \dots < z_N$ . By Rolle’s Theorem, for all  $1 \leq i \leq N - 1$ , there exists  $c_i(t) \in (tz_i, tz_{i+1})$  such that  $\eta'(c_i(t)) = 0$ . As a result  $\eta'$  has at least  $2N - 1$  zeros in all  $(r^-, r^+)$  satisfying the requirements.

Now, let us assume by contradiction that  $\eta'$  has strictly more than  $2N - 1$  zeros for arbitrarily small values of  $C_W, t_W$  and in all  $(r^-, r^+)$  satisfying the requirements. As a result, there are sequences  $(t_k)_{k \in \mathbb{N}}$  where  $t_k \rightarrow 0$ ,  $(z_k)_{k \in \mathbb{N}}$  (where each  $z_k \in \mathbb{R}^N$  has pairwise distinct coordinates and  $|z_k|_\infty \leq R_W$ ),  $(r_k^+)_{k \in \mathbb{N}}$  and  $(r_k^-)_{k \in \mathbb{N}}$  where for all  $k \in \mathbb{N}$ ,  $r_k^+ > 0, r_k^+ > \max_{1 \leq i \leq N} t_k(z_k)_i, r_k^+ \rightarrow 0$  and  $r_k^- < 0, r_k^- < \min_{1 \leq i \leq N} t_k(z_k)_i$  and  $r_k^- \rightarrow 0$ . And there exists  $(\eta_k)_{k \in \mathbb{N}} \in (\mathcal{C}^{2N}(X) \cap W^{2N,\infty}(X))^{\mathbb{N}}$  such that for all  $k \in \mathbb{N}$

$$\forall i \in \{1, \dots, N - 1\}, \eta_k(t_k(z_k)_i) = 1 \text{ and } \eta'_k(t_k(z_k)_i) = 0,$$

$$\forall \ell \in \{0, \dots, 2N\}, \left\| \eta_k^{(\ell)} - \eta_W^{(\ell)} \right\|_{L^\infty(X)} \leq \frac{1}{k}.$$

and  $\eta'_k$  has at least  $2N$  zeros in  $(r_k^-, r_k^+)$  (we already know that  $\eta'_k$  has at least  $2N - 1$  zeros in  $(r_k^-, r_k^+)$ ). Thus, for all  $k \in \mathbb{N}$ , by applying successively Rolle’s Theorem, we obtain that there exists  $x_k \in (r_k^-, r_k^+)$  such that  $\eta_k^{(2N)}(x_k) = 0$ . Since  $x_k \rightarrow 0$  (because  $r_k^-, r_k^+ \rightarrow 0$ ) and  $\left\| \eta_k^{(2N)} - \eta_W^{(2N)} \right\|_{L^\infty(X)} \leq \frac{1}{k}$ , we deduce that  $\eta_W^{(2N)}(0) = 0$ , which is a contradiction. Hence  $\eta'$  has exactly  $2N - 1$  zeros in some  $(r^-, r^+)$ .

Using the same argument, we may also prove that for all  $i \in \{1, \dots, N\}$ ,  $\eta''(tz_i) \neq 0$ . Let us now observe that, either for all  $1 \leq i \leq N$ ,  $\eta''(tz_i) > 0$  or, for all  $1 \leq i \leq N$ ,  $\eta''(tz_{i+1}) < 0$ . Indeed, assume for instance by contradiction that for some  $1 \leq i \leq N - 1$ ,  $\eta''(tz_i) > 0$  and  $\eta''(tz_{i+1}) < 0$ . Then, there exists  $c_i(t) \in (tz_i, tz_{i+1})$  such that  $\eta(c_i(t)) = 1$ . Applying Rolle’s Theorem on respectively  $(tz_i, c_i(t))$  and  $(c_i(t), tz_{i+1})$ , we obtain that  $\eta'$  vanishes at least twice in  $(tz_i, tz_{i+1})$ . It is a contradiction with the fact that  $\eta'$  has exactly  $2N - 1$  zeros in  $(r^-, r^+)$  for all  $0 < t < t_W$ .

As a result, there are only two possibilities: either  $\eta''(tz_i) < 0$  for all  $i$  (and then for all  $x \in (r^-, r^+) \setminus \bigcup_i \{tz_i\}$   $\eta(x) < 1$ ), or  $\eta''(tz_i) > 0$  for all  $i$  (and then for all  $x \in (r^-, r^+) \setminus \bigcup_i \{tz_i\}$   $\eta(x) > 1$ ). But since  $\eta_W^{(2N)}(0) < 0$ , there is some  $\tilde{x} \in (r^-, r^+)$  and some  $\varepsilon_0 > 0$  such that  $\eta_W(\tilde{x}) < 1 - \varepsilon_0$ . Choosing  $C_W$  small enough so that  $C_W < \varepsilon_0/2$ , we obtain that  $\eta(\tilde{x}) < 1 - \varepsilon_0/2$ . As a consequence, for all  $x \in (r^-, r^+) \setminus \bigcup_i \{tz_i\}$ ,  $\eta(x) < 1$ . Finally we can suppose that  $\eta_W(x) > -1$  on  $(r^-, r^+)$  by imposing  $0 < C_W < \inf_x \eta_W + 1$ .

To sum up, we have proved the following :

$$\forall x \in (r^-, r^+) \setminus \bigcup_i \{tz_i\}, |\eta(x)| < 1 \text{ and } \forall 1 \leq i \leq N, \eta''(tz_i) < 0. \tag{47}$$

Globally

As  $\eta_W$  is non-degenerate, we have  $\sup_{X \setminus (r^-, r^+)} |\eta_W| < 1$ . We can assume that  $0 < C_W < (1 - \sup_{X \setminus (r^-, r^+)} |\eta_W|)/2$  and use  $\|\eta - \eta_W\|_{L^\infty(X)} \leq C_W$  so as to obtain

$$\forall x \in X \setminus (r^-, r^+), \quad |\eta(x)| < 1. \tag{48}$$

Gathering Eqs. (47) and (48), we obtain the claimed result.

### Appendix 1.3: Proof of Proposition 5

The proof of Proposition 5 relies on the study of the structure of the matrices  $\Psi_k^* \Psi_k$ . We start by introducing the notion of checkerboard matrices.

**Definition 4** Let  $n \in \mathbb{N}^*$  and  $A = (a_{i,j})_{i,j} \in \mathbb{R}^{n \times n}$ . We say that  $A$  is a checkerboard matrix if for all  $(i, j)$  such that  $i + j$  is odd,  $a_{i,j} = 0$ .

For an odd  $n$ , such a matrix looks like

$$\begin{pmatrix} a_{1,1} & 0 & a_{1,3} & 0 & \dots & 0 & a_{1,n-2} & 0 & a_{1,n} \\ 0 & a_{2,2} & 0 & a_{2,4} & \dots & a_{2,n-3} & 0 & a_{2,n-1} & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & a_{n-1,2} & 0 & a_{n-1,4} & \dots & a_{n-1,n-3} & 0 & a_{n-1,n-1} & 0 \\ a_{n,1} & 0 & a_{n,3} & 0 & \dots & 0 & a_{n,n-2} & 0 & a_{n,n} \end{pmatrix}.$$

**Lemma 6** The set of checkerboard matrices of size  $n \in \mathbb{N}^*$  is an algebra over  $\mathbb{R}$ . As a result the inverse of a checkerboard matrix is also a checkerboard matrix.

*Proof* The only difficulty is to show that the product of two checkerboard matrices is also a checkerboard matrix. Let  $A = (a_{i,j})_{i,j}$  and  $B = (b_{i,j})_{i,j}$  be two such matrices. Let  $(i, j)$  such that  $i + j$  is odd. Then  $\sum_{k=1}^n a_{i,k} b_{k,j} = 0$  because if  $i + k$  is even then  $(i + j) - (i + k) = j - k$  is odd and  $j - k + 2k = j + k$  is odd, hence  $b_{j,k} = 0$ . On the contrary, if  $i + k$  is odd then  $a_{i,k} = 0$ . So  $AB$  is a checkerboard matrix. The last statement holds because the inverse of any matrix is a polynomial in that matrix.  $\square$

Now we give a more precise result on the structure of some particular checkerboard matrices.

**Lemma 7** Let  $A = (a_{i,j})_{1 \leq i,j \leq 2n+1}$  be a symmetric positive-definite checkerboard matrix of size  $2n + 1$  such that

$$a_{i,j} = (-1)^{\frac{i-j}{2}} a_{\frac{i+j}{2}, \frac{i+j}{2}}$$

for all  $i, j \in \{1, \dots, 2n + 1\}$  such that  $i + j$  is even, and let  $(b_{i,j})_{1 \leq i,j \leq 2n+1}$  denote the entries of  $A^{-1}$ . Then  $b_{2n+1,2n+1}$  and  $b_{1,2n+1}$  are positive.

*Proof* The fact that  $b_{2n+1,2n+1}$  is positive, is a direct consequence of the fact that  $A^{-1}$  is symmetric positive-definite because so is  $A$ .

From the expression of the inverse of a matrix using cofactors, we see that  $b_{1,2n+1} = \det(\hat{A}_{1,2n+1}) / \det(A)$ , where  $\hat{A}_{1,2n+1}$  is the matrix obtained from  $A$  by removing the first row and the last column. Since  $A$  is symmetric positive-definite,  $\det(A) > 0$ , and we need only show that  $\det(\hat{A}_{1,2n+1}) > 0$ . But one can see that  $\hat{A}_{1,2n+1}$  is a skew symmetric matrix of size  $2n$ , hence

$$\det(\hat{A}_{1,2n+1}) = \text{pf}(\hat{A}_{1,2n+1})^2,$$

where  $\text{pf}(\hat{A}_{1,2n+1})$  is the Pfaffian of the matrix  $\hat{A}_{1,2n+1}$ . For more details on the Pfaffian (definition and proof of the result used here), see [17]. As a result,  $\det(\hat{A}_{1,2n+1}) \geq 0$ .

Moreover,  $\hat{A}_{1,2n+1}$  is invertible. Indeed if we denote the columns of  $A$  (resp. the columns  $\hat{A}_{1,2n+1}$ ) by  $C_1, \dots, C_{2n+1}$  (resp.  $\hat{C}_1, \dots, \hat{C}_{2n}$ ), we observe that for  $0 \leq i \leq n - 1$ ,  $\hat{C}_{2i+1} \in E_1$  and  $\hat{C}_{2(i+1)} \in E_2$  where

$$E_1 = \text{Span}\{e_{2j+1}; 0 \leq j \leq n - 1\},$$

$$E_2 = \text{Span}\{e_{2(j+1)}; 0 \leq j \leq n - 1\},$$

with  $(e_i)_{1 \leq i \leq 2n}$  is the canonical base of  $\mathbb{R}^{2n}$ . Since  $E_1$  and  $E_2$  are in direct sum, if  $\hat{A}_{1,2n+1}$  is not invertible it means that either  $(\hat{C}_{2i+1})_{0 \leq i \leq n-1}$  or  $(\hat{C}_{2(i+1)})_{0 \leq i \leq n-1}$  is linearly dependent. But it cannot be  $(\hat{C}_{2(i+1)})_{0 \leq i \leq n-1}$  because it would imply that  $(C_{2(i+1)})_{0 \leq i \leq n-1}$  is also linearly dependent, because

$$C_{2(i+1)} = \begin{pmatrix} 0 \\ \hat{C}_{2(i+1)} \end{pmatrix},$$

which would contradict the invertibility of  $A$ .

So it means that  $(\hat{C}_{2i+1})_{0 \leq i \leq n-1}$  must be linearly dependent. However from the structure of the matrix  $A$ , one can see that for all  $0 \leq i \leq n - 1$

$$\hat{C}_{2i+1} = \begin{pmatrix} 0 \\ (-1)^{i+1} a_{2i+1,1} \\ 0 \\ -(-1)^{i+1} a_{2i+3,1} \\ \vdots \\ 0 \\ (-1)^{i+1} a_{2(i+n)-1,1} \\ 0 \\ -(-1)^{i+1} a_{2(i+n)+1,1} \end{pmatrix} \quad \text{and} \quad \hat{C}_{2(i+1)} = \begin{pmatrix} (-1)^{i+1} a_{2i+1,1} \\ 0 \\ -(-1)^{i+1} a_{2i+3,1} \\ 0 \\ \vdots \\ (-1)^{i+1} a_{2(i+n)-1,1} \\ 0 \\ -(-1)^{i+1} a_{2(i+n)+1,1} \\ 0 \end{pmatrix}.$$

Thus, a linear combination between the elements of  $(\hat{C}_{2i+1})_{0 \leq i \leq n-1}$  gives the same linear combination between the elements of  $(\hat{C}_{2(i+1)})_{0 \leq i \leq n-1}$ , which contradicts for the same reason as before the invertibility of  $A$ .

Hence  $\hat{A}_{1,2n+1}$  is invertible and  $b_{1,2n+1} > 0$ . □

The next result describes the structure of the matrices  $\Psi_k^* \Psi_k$  when  $\Phi$  is a convolution operator (i.e. for all  $x \in \mathbb{R}$ ,  $\varphi(x) = \tilde{\varphi}(\cdot - x)$ ) and  $k \in \mathbb{N}^*$  is odd.

**Lemma 8** *Suppose that  $\Phi$  is a convolution operator. Then for all  $k \in \mathbb{N}^*$ ,  $\Psi_k^* \Psi_k$  is a checkerboard matrix. Moreover if  $\mathcal{I}_{2N}$  holds, then the entries indexed by  $(1, 2N + 1)$  and  $(2N + 1, 2N + 1)$  of  $(\Psi_{2N}^* \Psi_{2N})^{-1}$  are positive.*

*Proof* Let  $(i, j)$  such that  $i + j$  is odd and for example  $i > j$  ( $\Psi_k^* \Psi_k$  is symmetric so it does not matter). The entry  $(i, j)$  of  $\Psi_k^* \Psi_k$  is equal to

$$\langle \varphi_{i-1}, \varphi_{j-1} \rangle_{\mathcal{H}}.$$

Recall that in the setting of Proposition 5,  $\mathcal{H} = L^2(X)$  and  $\langle f, g \rangle_{\mathcal{H}} = \int_X fg$ . Successive integrations by parts (we integrate the left term and derive the right term) yield

$$\begin{aligned} \langle \varphi_{i-1}, \varphi_{j-1} \rangle_{\mathcal{H}} &= (-1)^{\frac{i-j+1}{2}} \left\langle \varphi_{\frac{i+j-1}{2}-1}, \varphi_{\frac{i+j+1}{2}-1} \right\rangle_{\mathcal{H}}, \\ \langle \varphi_{i-1}, \varphi_{j-1} \rangle_{\mathcal{H}} &= (-1)^{\frac{i-j-1}{2}} \left\langle \varphi_{\frac{i+j+1}{2}-1}, \varphi_{\frac{i+j-1}{2}-1} \right\rangle_{\mathcal{H}}. \end{aligned}$$

Hence, by the symmetry of the scalar product, we obtain that

$$\langle \varphi_{i-1}, \varphi_{j-1} \rangle_{\mathcal{H}} = - \langle \varphi_{i-1}, \varphi_{j-1} \rangle_{\mathcal{H}}.$$

This implies that  $\langle \varphi_{i-1}, \varphi_{j-1} \rangle_{\mathcal{H}} = 0$  for odd values of  $i + j$ , so that  $\Psi_k^* \Psi_k$  is a checkerboard matrix (see Definition 4).

Moreover, similar computations show that when  $i + j$  is even

$$\langle \varphi_{i-1}, \varphi_{j-1} \rangle_{\mathcal{H}} = (-1)^{\frac{i-j}{2}} \left\langle \varphi_{\frac{i+j}{2}-1}, \varphi_{\frac{i+j}{2}-1} \right\rangle_{\mathcal{H}}.$$

If  $\mathcal{I}_{2N}$  holds,  $\Psi_{2N}^* \Psi_{2N}$  is symmetric positive-definite. As a result  $\Psi_{2N}^* \Psi_{2N}$  satisfies the assumptions of Lemma 7, so the entries  $(1, 2N + 1)$  and  $(2N + 1, 2N + 1)$  of  $(\Psi_{2N}^* \Psi_{2N})^{-1}$  are positive. □

We can now prove that  $\eta_W^{(2N)}(0) < 0$  when  $\Phi$  is a convolution operator and  $\mathcal{I}_{2N}$  holds.

*Proof of Proposition 5* Since  $\eta_W = \Phi^* p_W$ , we deduce that

$$\eta_W^{(2N)}(0) = \langle \varphi_{2N}, p_W \rangle_{\mathcal{H}} = \left\langle \varphi_{2N}, \Psi_{2N-1} (\Psi_{2N-1}^* \Psi_{2N-1})^{-1} \delta_{2N} \right\rangle_{\mathcal{H}}.$$

Consider the symmetric positive definite matrix

$$\Psi_{2N}^* \Psi_{2N} = \begin{pmatrix} \Psi_{2N-1}^* \Psi_{2N-1} & \Psi_{2N-1}^* \varphi_{2N} \\ [\Psi_{2N-1}^* \varphi_{2N}]^* & \|\varphi_{2N}\|_{\mathcal{H}}^2 \end{pmatrix} \in \mathbb{R}^{2N \times 2N}.$$

Observe that  $\Psi_{2N-1}^* \Psi_{2N-1}$  is invertible, and that

$$S \stackrel{\text{def.}}{=} \|\varphi_{2N}\|_{\mathcal{H}}^2 - \left\langle \varphi_{2N}, \Psi_{2N-1}(\Psi_{2N-1}^* \Psi_{2N-1})^{-1} \Psi_{2N-1}^* \varphi_{2N} \right\rangle_{\mathcal{H}} \neq 0.$$

Indeed,

$$S = \|(\text{Id}_{\mathcal{H}} - \Pi_{2N-1})\varphi_{2N}\|_{\mathcal{H}}^2 > 0$$

since  $(\varphi_0, \dots, \varphi_{2N})$  has full rank.

Thus, we may apply the block inversion formula, and  $(\Psi_{2N}^* \Psi_{2N})^{-1}$  is of the form

$$\begin{aligned} (\Psi_{2N}^* \Psi_{2N})^{-1} &= \frac{1}{\|(\text{Id}_{\mathcal{H}} - \Pi_{2N-1})\varphi_{2N}\|_{\mathcal{H}}^2} \\ &\times \begin{pmatrix} * & * & * & -(\Psi_{2N-1}^* \Psi_{2N-1})^{-1} \Psi_{2N-1}^* \varphi_{2N} \\ * & * & * & 1 \\ * & * & * & 1 \\ -[(\Psi_{2N-1}^* \Psi_{2N-1})^{-1} \Psi_{2N-1}^* \varphi_{2N}]^* & & & 1 \end{pmatrix} \end{aligned}$$

Lemma 8 ensures that the entry  $(2N + 1, 1)$  of  $(\Psi_{2N}^* \Psi_{2N})^{-1}$  is (strictly) positive. This precisely means that

$$-\frac{\langle \varphi_{2N}, \Psi_{2N-1}(\Psi_{2N-1}^* \Psi_{2N-1})^{-1} \delta_{2N} \rangle_{\mathcal{H}}}{\|(\text{Id}_{\mathcal{H}} - \Pi_{2N-1})\varphi_{2N}\|_{\mathcal{H}}^2} > 0,$$

and as a result  $\eta_W^{(2N)}(0) = \langle \varphi_{2N}, p_W \rangle_{\mathcal{H}} < 0$ . □

**Appendix 1.4: Proof of Proposition 6**

Let denote  $(a_k)_{0 \leq k \leq 2N-1}$  the coefficients of the first column of the matrix  $(\Psi_{2N-1}^* \Psi_{2N-1})^{-1}$ . Then we know that for all  $x \in \mathbb{R}$ ,

$$\eta_W(x) = \sum_{k=0}^{N-1} a_{2k} \varphi \star \varphi^{(2k)}(x).$$

Since  $\varphi \star \varphi = x \rightarrow \sqrt{\pi} e^{-x^2/4}$ , we get that for all  $x \in \mathbb{R}$ ,

$$\eta_W(x) = \sqrt{\pi} e^{-x^2/4} \sum_{k=0}^{N-1} a_{2k} \hat{H}_{2k}(x), \tag{49}$$

where  $\hat{H}_{2k}$  is the Hermite polynomial of order  $2k$  associated to  $x \rightarrow e^{-x^2/4}$ .

Now let us show (18) recursively on  $N \in \mathbb{N}^*$ . If  $N = 1$ , we have that  $\eta_W(x) = e^{-x^2/4}$  because  $\Psi_{2N-1}^* \Psi_{2N-1} = (\sqrt{\pi})$ . Suppose that the property is true for some  $N \in \mathbb{N}^*$ , i.e.,

$$\eta_{W,N}(x) = e^{-x^2/4} \sum_{k=0}^{N-1} \frac{x^{2k}}{2^{2k} k!}, \quad (50)$$

where we use the notation  $\eta_{W,N}$  to recall that this is the function  $\eta_W$  for  $N$  spikes. Then form for all  $x \in \mathbb{R}$ ,

$$\kappa(x) \stackrel{\text{def.}}{=} e^{x^2/4} (\eta_{W,N+1}(x) - \eta_{W,N}(x)).$$

$\kappa$  is a polynomial of degree  $2N$  [thanks to (49)], it satisfies  $\kappa(0) = 0$  and for all  $1 \leq i \leq 2N - 1$ ,  $\kappa^{(i)}(0) = 0$ . As a result for all  $x \in \mathbb{R}$ ,

$$\kappa(x) = \lambda x^{2N},$$

for some  $\lambda \in \mathbb{R}$ .

It remains to show that  $\lambda = \frac{1}{2^{2N} N!}$ . Remark that for all  $x \in \mathbb{R}$ ,  $\kappa^{(2N)}(x) = \lambda(2N)!$  and on the other hand  $\kappa^{(2N)}(0) = -\eta_{W,N}^{(2N)}(0)$  because  $\eta_{W,N+1}^{(2N)}(0) = 0$  by definition of  $\eta_{W,N+1}$ . So it is enough to show that,

$$\eta_{W,N}^{(2N)}(0) = -\frac{(2N)!}{2^{2N} N!}. \quad (51)$$

Thanks to the Leibniz formula applied to (50), we have the following formula,

$$\eta_{W,N}^{(2N)}(0) = \sum_{k=0}^{N-1} \binom{2N}{2k} \hat{H}_{2N-2k}(0) \frac{(2k)!}{2^{2k} k!}.$$

Since  $\hat{H}_{2N-2k}(0) = (-1)^{N-k} \frac{(2N-2k-1)!}{2^{2N-2k-1} (N-1-k)!}$ , thanks to Lemma 9 below, we obtain,

$$\begin{aligned} \eta_{W,N}^{(2N)}(0) &= \sum_{k=0}^{N-1} \frac{(2N)!}{(2N-2k)!(2k)!} \cdot (-1)^{N-k} \frac{(2N-2k-1)!}{2^{2N-2k-1} (N-1-k)!} \cdot \frac{(2k)!}{2^{2k} k!} \\ &= -\frac{(2N)!}{2^{2N} N!} \sum_{k=0}^{N-1} \frac{N(N-1)!}{(N-1-k)!k!} \cdot \frac{2 \cdot (-1)^{N-1-k}}{(2N-2k)} \\ &= -\frac{(2N)!}{2^{2N} N!} N \sum_{k=0}^{N-1} \binom{N-1}{k} (-1)^{N-1-k} \int_0^1 x^{N-1-k} dx \\ &= -\frac{(2N)!}{2^{2N} N!} N \underbrace{\int_0^1 (1-x)^{N-1} dx}_{=1/N}. \end{aligned}$$

This ends the recursive proof and thus we have (18).



To conclude that  $\eta_W$  is  $(2N - 1)$ -non-degenerate, it remains to show that for all  $x \in \mathbb{R}^*$ ,  $|\eta_W(x)| < 1$ , since we already know that  $\eta_W^{(2N)}(0) < 0$  thanks to (51). But we have, thanks to (18), that for all  $x \in \mathbb{R}^*$ ,  $0 < \eta_W(x) < 1$  since  $\sum_{k=0}^{N-1} \frac{x^{2k}}{2^{2k}k!}$  is the truncated power series of  $e^{x^2/4}$ .

**Lemma 9** *One has for all  $k \in \mathbb{N}$ ,*

$$\hat{H}_{2k+2}(0) = (-1)^{k+1} \frac{(2k + 1)!}{2^{2k+1}k!}. \tag{52}$$

*Proof* We know that  $H_{k+2}(x) = xH_{k+1}(x) - (k + 1)H_k(x)$  where  $H_k$  is the Hermite polynomial of order  $k$  associated to  $x \mapsto e^{-x^2/2}$ . Thus  $H_{2k+2}(0) = -(2k + 1)H_{2k}(0)$  and then,

$$H_{2k+2}(0) = (-1)^{k+1} \frac{(2k + 1)!}{2^k k!}. \tag{53}$$

Now, remark that  $\hat{H}_k(x) = 2^{-k/2}H_k(x/\sqrt{2})$ , so that together with (53) we get the expected result (52). □

**Appendix 1.5: Proof of Theorem 3**

For all  $n \in \mathbb{N}$ , the solution  $m_n \stackrel{\text{def.}}{=} m_{a_n, t_n z_n}$  of  $\mathcal{P}_{\lambda_n}(y_{t_n} + w_n)$  satisfies the following first order optimality conditions:

$$\Gamma_{t_n z_n}^* (\Phi_{t_n z_n} a_n - \Phi_{t z_0} a_0 - w_n) + \lambda_n \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} = 0. \tag{54}$$

Since  $\eta_{\lambda_n, t_n} = \Phi^* (\frac{1}{\lambda_n} (\Phi_{t z_0} a_0 + w_n - \Phi_{t_n z_n} a_n))$  satisfies  $\|\eta_{\lambda_n, t_n}\|_{L^\infty(X)} \leq 1$ ,  $\eta_{\lambda_n, t_n} \in \text{Im } \Phi^*$  and  $\eta_{\lambda_n, t_n}(t z_n) = \text{sign}(a_n) = 1$ .

As a result, by taking  $n$  large enough if necessary, we know that  $(a_n, z_n) = g_t^*(\lambda_n, w_n)$  with  $g_t^* \in \mathcal{C}^{2N}(V_t^*)$  so that  $(a_n, z_n) \rightarrow (a_0, z_0)$ . Therefore all the asymptotic results that we established are true when applied for  $(a_n, z_n)$  for  $n$  large enough. Observe from (54), that we get for all  $n \in \mathbb{N}$ :

$$p_{\lambda_n, t_n} = \Gamma_{t_n z_n}^{*,+} \begin{pmatrix} \mathbb{1}_N \\ 0 \end{pmatrix} + \Pi_{t_n z_n} \frac{w_n}{\lambda_n} + \frac{1}{\lambda_n} \Pi_{t_n z_n} \Gamma_{t z_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix}.$$

As a consequence:

$$\|p_{\lambda_n, t_n} - p_W\|_{\mathcal{H}} \rightarrow 0 \text{ when } n \rightarrow +\infty,$$

so that for all  $0 \leq l \leq 2N$ ,

$$\left\| \eta_{\lambda_n, t_n}^{(l)} - \eta_W^{(l)} \right\|_{L^\infty(X)} \rightarrow 0 \text{ when } n \rightarrow +\infty. \tag{55}$$

In particular, since  $\|\eta_{\lambda_n, t_n}\| \leq 1$ , we deduce (21):

$$\|\eta_W\|_{L^\infty(X)} \leq 1.$$

## References

1. Bhaskar, B.N., Recht, B.: Atomic norm denoising with applications to line spectral estimation. In: 2011 49th Annual Allerton Conference on Communication, Control, and Computing, pp. 261–268 (2011)
2. Blu, T., Dragotti, P.L., Vetterli, M., Marziliano, P., Coulot, L.: Sparse sampling of signal innovations: theory, algorithms and performance bounds. *IEEE Signal Process. Mag.* **25**(2), 31–40 (2008)
3. Bredies, K., Pikkarainen, H.K.: Inverse problems in spaces of measures. *ESAIM* **19**(1), 190–218 (2013)
4. Candès, E.J., Fernandez-Granda, C.: Super-resolution from noisy data. *J. Fourier Anal. Appl.* **19**(6), 1229–1254 (2013)
5. Candès, E.J., Fernandez-Granda, C.: Towards a mathematical theory of super-resolution. *Commun. Pure Appl. Math.* **67**(6), 906–956 (2014)
6. Claerbout, J.F., Muir, F.: Robust modeling with erratic data. *Geophysics* **38**(5), 826–844 (1973)
7. Condat, L., Hirabayashi, A.: Cadzow denoising upgraded: A new projection method for the recovery of dirac pulses from noisy linear measurements. Preprint hal-00759253 (2013)
8. de Castro, Y., Gamboa, F.: Exact reconstruction using beurling minimal extrapolation. *J. Math. Anal. Appl.* **395**(1), 336–354 (2012)
9. de Castro, Y., Azaïs, J.-M., Gamboa, F.: Spike detection from inaccurate samplings. *Appl. Comput. Harmon. Anal.* **38**(2), 177–195 (2015)
10. Demanet, L., Nguyen, N.: The recoverability limit for superresolution via sparsity (2015). [arXiv:1502.01385](https://arxiv.org/abs/1502.01385)
11. Den Dekker, A.J., Van den Bos, A., et al.: Resolution: a survey. *JOSA A* **14**(3), 547–557 (1997)
12. Donoho, D.L.: Super-resolution via sparsity constraints. *SIAM J. Math. Anal.* **23**(5), 1309–1331 (1992)
13. Donoho, D.L., Chen, S.S., Saunders, M.A.: Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20**(1), 33–61 (1998)
14. Duval, V., Peyré, G.: Exact support recovery for sparse spikes deconvolution. To appear in *Foundation and Computational Mathematics* (2015)
15. Fernandez-Granda, C.: Support detection in super-resolution. In: *Proceedings of the 10th International Conference on Sampling Theory and Applications*, pp. 145–148 (2013)
16. Krim, H., Viberg, M.: Two decades of array signal processing research: the parametric approach. *IEEE Signal Process. Mag.* **13**(4), 67–94 (1996)
17. Lang, S.: *Algebra*. Graduate Texts in Mathematics. Springer, New York (2002)
18. Levy, S., Fullagar, P.: Reconstruction of a sparse spike train from a portion of its spectrum and application to high resolution deconvolution. *Geophysics* **46**(9), 1235–1243 (1981)
19. Liao, W., Fannjiang, A.: MUSIC for single-snapshot spectral estimation: stability and super-resolution (2014). [arXiv:1404.1484](https://arxiv.org/abs/1404.1484)
20. Mörrenshtern, V.I., Candès, E.J.: Super-resolution of positive sources: the discrete setup (2014). [arXiv:1504.00717](https://arxiv.org/abs/1504.00717)
21. Santosa, F., Symes, W.W.: Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. Stat. Comput.* **7**(4), 1307–1330 (1986)
22. Schmidt, R.O.: Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.* **34**(3), 276–280 (1986)
23. Tibshirani, R.: Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **58**, 267–288 (1994)