Asymptotic of Sparse Support Recovery for Positive Measures

Q Denoyelle, V Duval and G Peyré,
CEREMADE, Université Paris-Dauphine
INRIA
CEREMADE, CNRS, Université Paris-Dauphine
E-mail: denoyelle@ceremade.dauphine.fr, vincent.duval@inria.fr,
gabriel.peyre@ceremade.dauphine.fr

Abstract. We study sparse spikes deconvolution over the space of Radon measures 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 when the minimum separation distance of the input measure (the minimum distance between two spikes) tends to zero. We show that when \(|w|_2^2/\lambda, |w|_2^2/t^{2N-1} \) 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 and a regularization parameter small enough with respect to the minimum separation distance, there exists a unique solution to the BLASSO program with exactly the same number of spikes as the original measure. We provide an upper bound on the error 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 \(\lambda \) and \(w \) drop to zero faster than \(t^{2N-1} \).

1. Introduction – Sparse Spikes Deconvolution

Sparse spikes deconvolution aims at recovering a sparse spikes train \( m_{a_0,t_{z_0}} \) with amplitudes \( a_{0,i} \in \mathbb{R} \) and positions \( t_{z_0,i} \) on the torus \( \mathbb{T} \), from low resolution observations \( y_t + w \) where \( y_t \triangleq \Phi m_{a_0,t_{z_0}} \). Here \( \Phi m \triangleq \varphi \ast m \) is a low pass filter against an arbitrary smooth kernel \( \varphi \) and \( w \) is the noise contaminating the observations. This problem is a toy model for many applications, ranging from seismic trace deconvolution to stars field observation in astrophysics and spike sorting in neuro-imaging. The parameter \( t > 0 \) controls the minimum separation distance between the spikes, and we study the case when \( t \) is small. Note that while we expose the problem in 1-D, our analysis carries over in arbitrary dimension.

The recovery is obtained by the BLASSO convex optimization problem over the space \( \mathcal{M} \) of Radon measure on \( \mathbb{T} \):

\[
\min_{m \in \mathcal{M}} \frac{1}{2} \| \Phi m - (y_t + w) \|^2 + \lambda |m|_{(\mathbb{T})}. \\
(P_\lambda(y_t + w))
\]

where \( \lambda > 0 \) is the regularization parameter and \( |m|_{(\mathbb{T})} \) is the total variation of a measure \( m \) on the torus (not to be mistaken for the total variation of an image), that extends the discrete \( \ell^1 \) norm of vectors to measures.
\[ |m|(T) \overset{\text{def.}}{=} \sup \left\{ \int_T \psi dm ; \psi \in C(T), |\psi|_\infty \leq 1 \right\}. \]

Without the presence of noise on the data \( y_t = \Phi m_{a_0,tz_0} \), we rather consider the Basis Pursuit for measures ([11, 3]):

\[
\min_{m \in M, \Phi m = y_t} |m|(T).
\]

(P0(yt))

We refer the reader to [1, 4, 6, 7, 8, 5] for more details about the BLASSO program, the Basis Pursuit for measures and algorithms to compute numerically the solution to this problem.

The recovery performance of (P0(yt)) and (P(\lambda)(yt + w)) (i.e. how \( m \) is close to \( m_{a_0,tz_0} \)) for arbitrary \( a_0 \) has been the focus of several recent contributions. The breakthrough paper [2] shows that stable recovery of \( m_{a_0,tz_0} \) holds provided that the spikes are sufficiently separated, i.e. when \( t \) is large enough. Subsequent contributions analyzed carefully this stability in terms of recovered spikes concentration [6, 7] and exact support recovery [5].

For positive spikes (i.e. \( a_{0,i} > 0 \)), the picture is radically different, since exact recovery without noise (i.e. \( (w, \lambda) = (0, 0) \)) holds for all \( t > 0 \) in the case of the ideal low-pass filter, see for instance [4]. Stability constants however explode as \( t \to 0 \). A recent work [8] 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 exact \( O(1/t^{2N-1}) \) 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 [5] by providing an asymptotic analysis when \( t \to 0 \).

2. Asymptotic of Exact Support Recovery

2.1. Dual Certificate

An approach to study the problem \( P(\lambda)(yt + w) \) is to apply Fermat’s rule. Indeed, a discrete measure \( m_{a,tz} \overset{\text{def.}}{=} \sum_{i=1}^N a_i \delta_{tzi} \) is solution of \( P(\lambda)(yt + w) \) if and only if there exists \( \eta \in C(T) \) such that:

\[
\Phi^*(\Phi m_{a,tz} - yt) + \lambda \eta = 0 \quad \text{and} \quad \eta \in \partial|m_{a,tz}|(T)
\]

where \( \partial|m_{a,tz}|(T) \) is the subdifferential of \( |\cdot|(T) \) at \( m_{a,tz} \) and is given by:

\[
\partial|m_{a,tz}|(T) = \{ \eta \in C(T) ; \ |\eta|_\infty \leq 1 \ \text{and} \ \forall i = 1, \ldots, n, \ \eta(tzi) = \text{sign}(a_i) \}.
\]

In that case, \( \eta \) is what we called a certificate for \( P(\lambda)(yt + w) \) of \( m_{a,tz} \).

Similarly, finding a continuous function \( \eta \) verifying:

\[
\eta \in \text{Im}\Phi^* \ \text{with} \ \eta \in \partial|m_{a_0,tz_0}|(T),
\]

assures that \( m_{a_0,tz_0} \) is a solution of \( P_0(yt) \). \( \eta \) is then a certificate for \( P_0(yt) \) of \( m_{a_0,tz_0} \). It interpolates the sign of the amplitudes at the spikes’ positions and is lower than or equal to 1 in uniform norm.
2.2. Construction of a Solution to $\mathcal{P}_\lambda(y_t + w)$

Our approach to study $\mathcal{P}_\lambda(y_t + w)$ is to proceed in a constructive manner and then find sufficient conditions on the parameters $w, \lambda, t$ (typically a low noise regime with the right behavior of $w$ and $\lambda$ with respect to $t$) to assure that the unique solution of $\mathcal{P}_\lambda(y_t + w)$ is composed of $N$ spikes. With that knowledge we can then provide a study of the properties of this solution (for example how far are the spikes from the original measures with respect to noise level).

So the idea is to build a candidate solution $m_{a,tz}$ and then verify that $\eta_{\lambda,t} = \frac{1}{2} \Phi^*(y_t + w - \Phi m_{a,tz})$ is a certificate, for any $(w, \lambda)$ close from $(0,0)$ with the right behavior with respect to $t$. We do not provide the proofs of the results detailed below due to a lack of space. Interested readers may refer to [9].

For the first step we consider the measures $m_{a,tz}$ composed of $N$ spikes that are solutions of the following two equations:

$$\forall i = 1, \ldots, N, \Phi_{z_i}^*(\Phi m_{a,tz} - (y_t + w)) + \lambda = 0 \quad \text{and} \quad (\Phi_{z_i}^*)'(\Phi m_{a,tz} - (y_t + w)) = 0.$$ 

These equations mean respectively that $\eta_{\lambda,t}(z_i) = 1$ and $\eta_{\lambda,t}'(z_i) = 0$ for all $i$, i.e. $\eta_{\lambda,t}$ interpolates the sign of the amplitudes with zero derivatives. By applying the Implicit Function Theorem with respect to $(\lambda, w)$ around $(0,0)$, we obtain a local parametrization of the measures $m_{a,tz}$ solutions of the equations. Now that we have a candidate solution $m_{a,tz}$ for any $(\lambda, w)$ around $(0,0)$ and $t > 0$, we can move to the second step where it remains to show that $|\eta_{\lambda,t}|_{\infty} \leq 1$ provided sufficient hypothesis on $(\lambda, w)$ with respect to $t$.

Following [5], we introduce below the so called “vanishing derivatives pre-certificate” $\eta_{V,t}$, which is a function defined on $T$ that interpolates the spikes positions and signs (here +1). Note that $\eta_{V,t}$ can be computed in closed form by solving a linear system [5].

**Proposition 1** (Vanishing derivatives pre-certificate). One has the uniform convergence $\eta_{\lambda,t} \to \eta_{V,t}$ as $\frac{w}{\lambda} \to 0$ and $\lambda \to 0$ where $\eta_{V,t} = \Phi^* p_{V,t}$ with

$$p_{V,t} = \text{argmin}_{p \in L^2(T)} \left\{ ||p||_2; (\Phi^* p)(t) = 1 \text{ and } (\Phi^* p)'(t) = 0 \; \forall i \right\}.$$ 

Recall the conditions that define a certificate for $\mathcal{P}_0(y_t)$ that we saw in the previous section, then note that $\eta_{V,t}$ is a good candidate of being one but is not in general (hence its name). It is one if and only if $||\eta_{V,t}||_{\infty} \leq 1$, and then $m_{a_0,tz_0}$ is recovered by the Basis Pursuit ($\mathcal{P}_0(y_t)$).

**Definition 1** (Non-degeneracy of $\eta_{V,t}$). We say that $\eta_{V,t}$ is non-degenerate if $|\eta_{V,t}|_{\infty} < 1$ outside the position $t z_{0,i}$ and $\eta_{V,t}'(t) \neq 0$ for $i = 1, \ldots, N$.

It is shown in [5] that $\eta_{V,t}$ being non-degenerate implies that solving $(\mathcal{P}_\lambda(y_t + w))$ in a certain low noise regime ($w/\lambda \to 0$ and $\lambda \to 0$) gives a unique solution with the correct number $N$ of spikes, i.e. one obtains exact support recovery.

Proposition 1 is a step forward towards the proof of $\eta_{\lambda,t}$ being a certificate. Indeed if we are able to prove that $\eta_{V,t}$ is non-degenerate for all $t > 0$ then we can prove the expected result in a sufficient low noise regime with respect to $t$. So now we investigate the problem when $t \to 0$ by describing how $\eta_{V,t}$ behaves.

Figure 1, top row, shows graphically that $\eta_{V,t}$ converges with $t \to 0$ toward a fixed function $\eta_W$ that does not depend on $z_0$ and $t$. This function is defined in the following proposition, which ensures that the non-degeneracy of $\eta_{V,t}$ can thus be checked independently of $z_0$ and $t$ provided that $t$ is small.

**Proposition 2**. One has the uniform convergence $\eta_{V,t} \to \eta_W = \Phi^* p_W$ as $t \to 0$, where

$$p_W = \text{argmin}_{p \in L^2(T)} \left\{ ||p||_2; (\Phi^* p)(0) = 1, (\Phi^* p)'(0) = 0 \; \forall i \leq 2N - 1 \right\}.$$
\[ t = 0.4 \quad t = 0.2 \quad t = 0.01 \]

\[ N = 1 \quad (\eta_{V,t} = \eta_W) \quad N = 2 \quad N = 3 \]

**Figure 1.** Top row: \( \eta_{V,t} \) for several values of \( t \), showing the convergence toward \( \eta_W \). The filter \( \varphi \) is an ideal low-pass filter with a frequency cutoff \( f_c = 10 \). Bottom row: \( \eta_W \) for several values of \( N \).

Furthermore, if \( |\eta_W| < 1 \) outside 0 and \( \eta_W^{(2N)}(0) \neq 0 \) (we then say that \( \eta_W \) is \((2N - 1)\)-non-degenerate), then \( \eta_{V,t} \) is non-degenerate for small \( t \).

Figure 1, bottom row, shows \( \eta_W \) for several values of \( N \). Note 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 1 below) where signal-to-noise ratio is required to scale like \( 1/t^{2N-1} \).

So one can sum up the convergence of the pre-certificates as follows:

\[ \eta_{\lambda,t} \xrightarrow{\lambda \to 0, t \to 0} \eta_{V,t} \xrightarrow{t \to 0} \eta_W. \]

As a result, we can state as in Proposition 2 that if \( \eta_W \) is \((2N - 1)\)-non-degenerate, then \( \eta_{\lambda,t} \) is a certificate. In fact the next Proposition shows that \( \eta_{\lambda,t} \) verifies the same non-degeneracy properties as \( \eta_{V,t} \) (which implies in particular that \( \eta_{\lambda,t} \) is certificate).

**Proposition 3.** If \( \eta_W \) is \((2N - 1)\)-non-degenerate then for \( \frac{\pi}{2} \), \( \lambda \) and \( t \) small enough, one has \( |\eta_{\lambda,t}| < 1 \) outside the positions \( tz_i \) and \( \eta_{\lambda,t}'(tz_i) \neq 0 \).

**Remark 1.** \( \eta_W \) being \((2N - 1)\)-non-degenerate depends only on the filter \( \varphi \). See Figure 2 for a \((2N - 1)\)-non-degenerate \( \eta_W \) (when \( \varphi \) is the Dirichlet kernel) and a degenerate \( \eta_W \) (a low-pass filter with high extreme Fourier coefficients). However \( \eta_W^{(2N)}(0) \neq 0 \) is true when the family \( (\varphi, \varphi', \ldots, \varphi^{(2N)}) \) is linearly independent in \( L^2(\mathbb{T}) \). In fact one has even \( \eta_W^{(2N)}(0) < 0 \), so that \( |\eta_W| < 1 \) locally around 0. You can check that the latter statement is true in both the plots of \( \eta_W \) of Figure 2, even when \( \|\eta_W\|_\infty > 1 \).

**Figure 2.** \( \eta_W \) for the ideal low-pass filter (left) and a filter with high extreme Fourier coefficients (right).
2.3. Main Result

The results of the last section enable us to state the following theorem which is our main contribution.

**Theorem 1.** If \( \eta_W \) is \((2N - 1)\)-non-degenerate, then there exist positive constants \( t_1, C, M > 0 \) (which only depends on the filter and the initial measure) such that:

\[
\forall t \in (0, t_1), \; \forall (\lambda, w) \in B(0, Ct^2N^{-1}) \quad \text{with} \quad \frac{w}{\lambda} \leq C,
\]

there exists a unique measure \( m_{a,tz} \) solution of the BLASSO composed of exactly \( N \) spikes. Moreover,

\[
\|(a, z) - (a_0, z_0)\| \leq M \left( \frac{\|\lambda\|}{t^{2N-1}} + \|w\|_2 \right).
\]

The theorem provides an upper-bound on the error of the recovered measure \( m_{a,tz} \) with respect to the initial measure \( m_{a_0,tz_0} \). In particular when \( w/t^{2N-1} \) and \( \lambda/t^{2N-1} \) tend to zero, \( m_{a,tz} \) converges towards \( m_{a_0,tz_0} \).

Figure 3 shows the domain, given in Equation (1), in which there is exact support recovery of the initial measure. Note that when the number of spikes increases, the domain gets smaller i.e. the recovery of the support is harder to perform. This is reflected in Figure 1 where \( \eta_W \) gets flatter around 0 when \( N \) increases.

\[\text{Figure 3. Domain of recovery of the support for 1 and 2 spikes.}\]

**Acknowledgment**

This work has been partly supported by the European Research Council (ERC project SIGMA-Vision).

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

[2] E.J. Candès and C. Fernandez-Granda. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013.

[3] E.J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

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

[5] V. Duval and G. Peyré. Exact support recovery for sparse spikes deconvolution. *To appear in Foundation and Computational Mathematics*, 2013.

[6] C. Fernandez-Granda. Support detection in super-resolution. *Proc. Proceedings of the 10th International Conference on Sampling Theory and Applications*, pages 145–148, 2013.

[7] Y. de Castro J.-M. Azaïs and F. Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 2014.

[8] V. I. Morgenshtern and E. J. Candès. Stable super-resolution of positive sources: the continuous setup. *in final preparation*, 2014.

[9] G. Peyré Q. Denoyelle, V. Duval. Asymptotic of sparse support recovery for positive measures. *in preparation*, 2015.

[10] The recoverability limit for superresolution via sparsity. L. demanet and n. nguyen. *preprint*, 2014.

[11] D.L. Donoho S.S. Chen and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998.