Spike detection from inaccurate samplings

Jean-Marc Azaïs, Yohann De Castro, Fabrice Gamboa

▶ To cite this version:

HAL Id: hal-00780808

https://hal.archives-ouvertes.fr/hal-00780808v2

Submitted on 25 Feb 2014

SPIKE DETECTION FROM INACCURATE SAMPLINGS

Jean-Marc Azaïs, Yohann de Castro, and Fabrice Gamboa

Abstract. This article investigates the support detection problem using the LASSO estimator in the space of measures. More precisely, we study the recovery of a discrete measure (spike train) from few noisy observations (Fourier samples, moments...) using an ℓ_1-regularization procedure. In particular, we provide an explicit quantitative localization of the spikes.

1. Introduction

1.1. **Super-resolution.** Imaging experiments can be subject to device limitations where one cannot observe enough information in order to recover fine details. For instance, in optical imaging, the physical limitations are evaluated by the resolution. This latter measures the minimal distance between lines that can be distinguished. Hence, the details below the resolution limit seem unreachable. The super-resolution phenomenon is the ability to recover the information beyond the physical limitations. Surprisingly, if the object of interest is simple, e.g. a discrete measure, then it is possible to override the resolution limit. In particular, the reader may think of important questions in applied harmonic analysis such as the problem of source separation. Many companion applications in astronomy, medical imaging and single molecule imaging in 3D microscopy are at stake, see [16, 15, 21] and references therein. Hence, theoretical guarantees of source detection are of crucial importance in practice.

In this paper, we prove quantitative detection guarantees from noisy observations (Fourier samples, moments samples). Furthermore, these quantitative estimates can be computed using a tractable algorithm, called BLASSO.

1.2. **Previous works.** The theoretical analysis of the ℓ_1-regularization in the space of measures was initiated by Donoho [8]. Few years after, rates of convergence in super-resolution have been investigated by P. Doukhan, E. Gassiat and one author of this present paper [9, 12]. They considered the exact reconstruction of a nonnegative measure and derived results when one only knows the values of a finite number of linear functionals at the target measure. Moreover, they study stability with respect to a metric for weak convergence. Likewise, two authors of this paper [7] proved that k spikes trains can be faithfully resolved from $m = 2k + 1$ samples (Fourier samples, Stieltjes transformation, Laplace transform...) by using an ℓ_1-minimization method.

Following recent proposal [3, 4, 17, 19] on inverse problems regularization in Banach spaces, we consider convergence rates in Bregman divergence. On a more general note, inverse problems on the space of measures are now well understood, see [13, 20] for instance. We capitalize on these earlier works to construct our analysis. In particular, we use them to give quantitative localizations of the recovered spikes, which is new.

In the super-resolution frame, the important paper [6] shows that if the spikes are well "seperated" then there exists a dual certificate, i.e. an ℓ_∞-constrained trigonometric polynomial that interpolates the phase of the weights at the spikes locations. This construction provides a quadratic isolation condition, see Definition 2.2, of the spikes which is crucial for ℓ_1-minimization in the space of measures. In a predating paper [7], the authors investigates ℓ_1-minimization with different types of measurements: trigonometric, polynomial, Laplace transform... In view of application, the recent works [5, 22] derive results in ℓ_1 and ℓ_2 prediction, i.e. the estimation of the input frequencies. Moreover, note that noise robustness of support recovery is proved in [10].

A postdating paper [11] bounds the support detection error for a constrained formulation of the ℓ_1-minimization in the space of measures as in Theorem 2.2. Regarding unconstrained formulation, a second postdating paper [22] studies spikes detection for the Fourier sampling under a Gaussian noise model. The authors provide an optimal rate for the ℓ_2-prediction, namely the ℓ_2 distance between the recovered Fourier coefficients and the original ones.

The aforementioned results suggest that the recovered spike locations should be close to the input support. This is investigated, for the first time, in this paper for an unconstrained ℓ_1-minimization problem under a general sampling scheme.

1.3. **General model and notation.** Let $(\mathbb{T}, d)$ be a compact metric space homeomorphic to either the interval $([0, 1], |\,.\,|)$ or the unit circle $\mathbb{S}^1$, which is identified to the metric space $(\mathbb{R} \mod (1), d(\cdot, \cdot))$ via the mapping $z = e^{\mathrm{i}2\pi t}$. In this latter case, the distance d is taken around the circle. Let Δ be a complex measure on $\mathbb{T}$ with discrete support of size s. In particular, the measure Δ has polar decomposition, see [18] for a definition:

$$(1.1) \qquad \Delta = \sum_{k=1}^{s} \Delta_k \exp(\mathrm{i}\,\theta_k)\, \delta_{T_k}\,,$$

where $\Delta_k > 0$, $\theta_k \in \mathbb{R}$, $T_k \in \mathbb{T}$ for $k = 1, \ldots, s$ and δ_x denotes the Dirac measure at point x. Let m be a positive integer and $\mathbf{F} = \{\varphi_0, \varphi_1, \ldots, \varphi_m\}$ be a family of complex continuous functions on $\mathbb{T}$. Assume that the family $\mathbf{F}$ is **a family of orthonormal functions** with respect to a probability measure Π on $\mathbb{T}$.

We recall some basic concepts in the frame of generalized polynomials. Define the k-th generalized moment of a complex measure μ on $\mathbb{T}$ as:

$$c_k(\mu) = \int_{\mathbb{T}} \varphi_k \,\mathrm{d}\mu\,,$$

for $k = 0, 1, \ldots, m$. Suppose we observe $y = (y_k)_{k=0}^{m}$ defined by $y_k = c_k(\Delta) + \varepsilon_k$ for $k = 0, 1, \ldots, m$, where $\varepsilon = (\varepsilon_k)_{k=0}^{m}$ is a complex valued white noise. This can be written as:

$$y = \int_{\mathbb{T}} \Phi \,\mathrm{d}\Delta + \varepsilon\,,$$

where $\Phi = (\varphi_0, \ldots, \varphi_m)$. In this paper, we investigate the recovery of the complex measure Δ from $m + 1$ measurements given by y.

Remark. Along this article, we shall mention examples in the Fourier case (Fourier samples) or in the polynomial case (moment samples), notation are described therein. If not specified, notation are in accordance with the general model.

1.4. **Unconstrained minimization.** Denote by $\mathcal{M}$ the set of all finite complex measures on $\mathbb{T}$ and by $\|\,.\,\|_{TV}$ the total variation norm. We recall that for all $\mu \in \mathcal{M}$,

$$\|\mu\|_{TV} = \sup_{\mathbf{P}} \sum_{E \in \mathbf{P}} |\mu(E)|\,,$$

where the supremum is taken over all partitions $\mathbf{P}$ of $\mathbb{T}$ into a finite number of disjoint measurable subsets. For further details, we refer the reader to [18].

By analogy with the LASSO [23], *Beurling LASSO* (BLASSO) is the process of reconstructing a discrete measure Δ from the samples y by finding a solution to:

$$\text{(BLASSO)} \qquad \hat{\Delta} \in \arg \min_{\mu \in \mathcal{M}} \frac{1}{2} \| \int_{\mathbb{T}} \Phi \, d\mu - y \|_2^2 + \lambda \| \mu \|_{TV} \,,$$

where λ is a tuning parameter.

Remark. For the case of Fourier coefficients and $\varepsilon = 0$, (BLASSO) is called *Generalized Minimal Extrapolation* (GME), see [7] and Eq. (2.1) for a definition. This procedure finds an extrapolation of a function given on a subset of $\mathbb{T}$ with minimal TV-norm of its Fourier transform among all possible extrapolating functions. Moreover, our procedure is an extension of the LASSO estimator of the high-dimensional regression theory.

Remark. On the algorithmic side, observe that the Fenchel dual program of (BLASSO) can be recast into a SDP program when using Fourier samplings or moment samplings, see [6, 22] for instance.

One knows that the extreme points of the unit ball of the TV-norm are the atoms δ_x where $x \in \mathbb{T}$. Therefore, ℓ_1-minimization compels the solutions to be discrete measures. Nevertheless, the TV-norm is innappropriated in measuring the distance between spikes, and it seems difficult to localize BLASSO. Two questions immediately arise:

- How close is the recovered spike train $\hat{\Delta}$ from the target Δ?
- How accurate is the localization of BLASSO in terms of the noise and the amplitude of the recovered/original spike?

To the best of our knowledge, this paper is the first to address these questions in a general frame. In particular, it is the first paper to provide quantitative localization guarantees from the output amplitudes.

1.5. **Contribution.** The present paper is concerned with the problem of the localization of the target spikes. In the general frame, we show that the large recovered spikes are very close to the target spikes, and that the mass of all the recovered spikes far from the original support is small. In the Fourier case and the moment case, we specify that under mild assumptions on the support of the target, the mass of the reconstructed measure is concentrated around large spikes, i.e. spikes above the noise level. In particular, we derive error bounds in terms of the recovered measure and the original measure.

1.6. **Organization of the paper.** The next section present the main results in the general frame. We derive the corresponding results in the Fourier frame and the moment frame in Section 3. The bound on the noise in the space of continuous functions is given in Section 4 using the Rice method. Section 5 recalls some useful optimization tools in order to implement our procedure.

2. QUANTITATIVE LOCALIZATION IN THE GENERAL FRAME

2.1. **Zero-noise problem.** All solution to ℓ_1-regularization satisfies some optimality condition based on the fact that the sub-gradient of the regularization function vanishes at the solution point. Then a sufficient condition for exact recovery is that the target measure satisfies this optimality condition. This analysis has led to the notion of dual certificate, see [7, 6] for instance, and the notion of source condition, see [4].

Definition 2.1 (Dual certificate)**.** *We say that a generalized polynomial $P = \sum_{k=0}^{m} a_k \varphi_k$ is a dual certificate for the measure Δ defined by* (1.1) *if and only if it satisfies the following properties:*

- *phase interpolation: $\forall k \in \{1, \ldots, s\}$, $\quad P(T_k) = \exp(-\mathbf{i}\theta_k)$,*
- *ℓ_∞-constraint: $\|P\|_\infty \leq 1$.*

Indeed, one can prove that the target measure Δ is a solution of the following ℓ_1-minimization:

$$(2.1) \qquad \Delta^{GME} \in \arg\min_{\mu \in \mathcal{M}} \|\mu\|_{TV} \quad \text{s.t.} \int_{\mathbb{T}} \Phi \, \mathrm{d}\mu = \int_{\mathbb{T}} \Phi \, \mathrm{d}\Delta.$$

if and only if Δ has a dual certificate, a proof can be found in [7]. We understand that if Δ has no dual certificate (i.e. no generalized polynomial with infinity norm less than 1 interpolates the phases at support point locations) there is no hope in recovering Δ with an ℓ_1-minimization method.

2.2. **Quantitative localization guarantees from the output amplitudes.** In the presence of noise, one cannot ask for exact recovery and the notion of dual certificate is too loose for establishing stability results. Therefore, we strengthen a bit more this notion so as to derive a quantitative localization of the target spikes from BLASSO.

Definition 2.2 (Quadratic isolation condition)**.** *We say that a finite set $\mathbf{S} = \{T_1, \ldots, T_s\} \subset \mathbb{T}$ satisfies the quadratic isolation condition with parameters $C_a > 0$ and $0 < C_b < 1$, denoted by $\mathrm{QIC}(C_a, C_b)$, if and only if for all $(\theta_k)_{k=1}^{s} \in \mathbb{R}^s$, there exists $P \in \mathrm{Span}(\mathbf{F})$ such that for all $k = 1, \ldots, s$, $P(T_k) = \exp(-\mathbf{i}\theta_k)$, and*

$$\forall x \in \mathbb{T}, \quad 1 - |P(x)| \geq \min_{T \in \mathbf{S}} \{C_a m^2 d(x, T)^2, C_b\}.$$

Before stating the theorem, we would like to point out that, under general conditions [14], the solutions of the convex program (BLASSO) always contain an atomic solution with support of size less than $m + 2$. Moreover, this solution can be computed from the solution of the convex dual problem in the Fourier frame, see [6] for instance. Therefore, we can always consider an atomic solution of BLASSO:

$$\hat{\Delta} = \sum_{k=1}^{n} \hat{\Delta}_k \exp(\mathbf{i}\hat{\theta}_k) \, \delta_{\hat{T}_k}.$$

We begin with a first result showing that, even with a small regularization parameter λ, the large spikes of BLASSO detect the original support points. Thereafter, we will denote

$$\mathcal{N}(c_0, \mathbf{S}) := \left\{ x \in \mathbb{T}; \quad \min_{T \in \mathbf{S}} d(x, T) \leq \frac{c_0}{m} \right\},$$

and $\mathcal{F}(c_0, \mathbf{S})$ its complement, or $\mathcal{N}$ and $\mathcal{F}$ for short.

Theorem 2.1. *Assume that the support $\mathbf{S}$ of Δ satisfies $\mathrm{QIC}(C_a, C_b)$. Let λ be such that $\lambda \geq \|\varepsilon\|_2$. Then, for all $x \in \mathbb{T}$ such that:*

$$|\hat{\Delta}(\{x\})| > \frac{2\lambda}{C_b},$$

there exists $T \in \mathbf{S}$ satisfying:

$$d(x, T) \leq \left[\frac{2\lambda}{C_a |\hat{\Delta}(\{x\})|} \right]^{1/2} \frac{1}{m} \leq \frac{c_0}{m},$$

where $c_0 = \sqrt{C_b / C_a}$. Observe that, if $\min_{k \neq l} d(T_k - T_l) > \frac{2c_0}{m}$ then the aforementioned point T is unique. Moreover,

(1) $\displaystyle\sum_{\hat{T}_k\in\mathcal{N}(c_0,\mathbf{S})}\hat{\Delta}_k\min_{T_l\in\mathbf{S}}d(\hat{T}_k,T_l)^2\leq\frac{2\lambda}{C_am^2}$,

(2) $\displaystyle\sum_{\hat{T}_k\in\mathcal{F}(c_0,\mathbf{S})}\hat{\Delta}_k\leq\frac{2\lambda}{C_b}$.

This result shows that BLASSO puts a small weight far from the original support, see (2). Moreover, it shows that the reconstructed points, with large weights, are close to the true support, see (1). Our result quantitatively bounds the support recovery error in terms of the amplitude of the solution. In actual practice, the result (1) provides a confidence set on the localization of the true support. Furthermore, it shows that this localization is getting better as the recovered amplitude is large.

Remark. Postdating this paper, some important works [11, 22] improve our result, in the Fourier case, providing bounds that depend only on the amplitude of the original spike. An attentive reader can see that the bounds *(ii)* and *(iii)* in Theorem 1.2 of [11] and the bounds *(i)* and *(ii)* in Theorem 2 of [22] are covered by our predating result. Moreover, Theorem 2.1 deals with a regularizing parameter λ that can be of the order of $\|\varepsilon\|_2$. In contrast, the regularizing parameter τ in [22] has to be of the order of $\lambda_0=\|\sum_{k=0}^m\varepsilon_k\varphi_k\|_\infty$ which is roughly of the order of $\|\varepsilon\|_\infty$ in the Gaussian noise model, see 4.2.

2.3. **Quantitative localization guarantees from the input amplitudes.** One knows [2] that any bounded polynomial on a compact of the real line has its derivative upper bounded by a constant times its degree. Hence, a uniform upper bound on the derivatives of the dual certificate P can be given.

Definition 2.3 (Bernstein Isolation Property). *We say that a set $\mathbf{S}$ satisfies the Bernstein Isolation Property with parameters $c_0>0$ and $C_c>0$, denoted by $\mathrm{BIP}(c_0,C_c)$, if and only if*

$$\forall P\in\mathrm{Span}(\mathbf{F}),\quad\forall x\in\mathcal{N}(c_0,\mathbf{S}),\quad|P''(x)|\leq C_cm^2\|P\|_\infty.$$

Using this property, we can specify the result of Theorem 2.1 in terms of the input amplitudes.

Theorem 2.2. *Assume that the support $\mathbf{S}$ of Δ satisfies $\mathrm{BIP}(c_0,C_c)$ and $\mathrm{QIC}(C_a,C_b)$ with $c_0=\sqrt{C_b/C_a}$. Let λ be greater than $\|\sum_{k=0}^m\overline{\varepsilon}_k\varphi_k\|_\infty$. Then*

(3) $\displaystyle|\Delta_j-\sum_{\{k:\,d(\hat{T}_k,T_j)\leq\frac{c_0}{m}\}}\hat{\Delta}_k|\leq C'\lambda$,

where $C'=2+\max\left\{\frac{2(1-C_b)}{C_b},\frac{C_c}{C_a}\right\}$. Moreover for all $T_j\in\mathbf{S}$ corresponding to a weight such that:

$$\Delta_j>C'\lambda,$$

there exists $\hat{T}\in\mathrm{Supp}(\hat{\Delta})$ satisfying:

$$d(T_j,\hat{T})\leq\left[\frac{2\lambda}{C_a(\Delta_j-C'\lambda)}\right]^{1/2}\frac{1}{m}.$$

Observe that this phenomenon has been investigated in the Fourier case, see for instance [11, 22]. Theorem 2.2 shows that it is conducted by the BIP property, namely a control of the second derivative of the dual certificate on a small vicinity of each point of the support $\mathbf{S}$. Hence, it may applies to a more general frame.

2.4. **Examples of families satisfying BIP.** The BIP property is satisfied for a large class of family of measurements, see [2] for instance. We present some standard Bernstein-type inequalities for which the BIP property can be derived.

Fourier samples: Let $\mathbf{F}_F = \{\exp(-\imath 2\pi f_c x), \ldots, 1, \ldots, \exp(\imath 2\pi f_c x)\}$ then

$$(2.2) \qquad \forall P \in \mathrm{Span}(\mathbf{F}_F), \quad \forall x \in [0,1], \quad |P''(x)| \leq \pi^2 (2f_c)^2 \|P\|_\infty.$$

Moment samples: Let $\mathbf{F}_M = \{1, x, \ldots, x^m\}$ then

$$(2.3) \qquad \forall P \in \mathrm{Span}(\mathbf{F}_M), \quad \forall x \in (-1,1), \quad |P''(x)| \leq 4\left(\frac{m}{\sqrt{1-x^2}}\right)^2 \|P\|_\infty.$$

Laplace transform: Let $0 < \lambda_0 < \lambda_1 < \lambda_2 < \cdots$ be any real numbers. Let $\mathbf{F}_L = \{\exp(-\lambda_0 x), \exp(-\lambda_1 x), \exp(-\lambda_2 x), \ldots\}$. From Newman's inequality ([2], p.276), we know that:

$$\forall P \in \mathrm{Span}(\mathbf{F}_L), \quad \forall x \in [0, +\infty), \quad |P''(x)| \leq \left(9 \sum_{i=0}^{\infty} \lambda_i\right)^2 \|P\|_\infty.$$

Müntz polynomials I: Let $0 < \alpha_1 < \cdots < \alpha_m$ be any real numbers. Let $\mathbf{F}_{M\ddot{u}} = \{1, x^{\alpha_1}, \ldots, x^{\alpha_m}\}$ then for all $\eta > 0$, there is a constant c_η such that

$$\forall P \in \mathrm{Span}(\mathbf{F}_{M\ddot{u}}), \quad \forall x \in (0, 1-\eta), \quad |P''(x)| \leq \frac{c_\eta}{x^2} \|P\|_\infty.$$

Müntz polynomials II: Let $1 < \alpha_0 < \cdots < \alpha_m$ be any real numbers. Let $\mathbf{F}_{M\ddot{u}} = \{x^{\alpha_0}, x^{\alpha_1}, \ldots, x^{\alpha_m}\}$ then, by Newman's inequality ([2], p.276),

$$\forall P \in \mathrm{Span}(\mathbf{F}_{M\ddot{u}}), \quad \forall x \in (0,1], \quad |P''(x)| \leq \frac{3^4 (\sum_{i=0}^m \alpha_i)(\sum_{i=0}^m \alpha_i + 1)}{x^2} \|P\|_\infty.$$

These examples shows that the BIP property is a mild assumption in most practical cases.

3. SUPPORT DETECTION FROM NOISY FOURIER/MOMENT SAMPLES

3.1. **Detection from noisy Fourier samples.** In this subsection, we mention the example of Fourier samples to illustrate our results. Recently, much emphasis has been put on the recovery of a discrete measure from noisy band-limited data [5, 11, 22]. In this setting, we observe noisy Fourier samples up until a frequency cut-off $f_c \in \mathbb{N}^*$. We specify notation:

- The number of samples is $2f_c + 1$ hence $m = 2f_c$.
- For sake of simplicity, we place ourselves on $\mathbb{T} = [0,1]$.
- For all $k \in \{-f_c, \ldots, f_c\}$, we set for all $x \in [0,1]$, $\varphi_k(x) = \exp(\mathrm{i} 2\pi k x)$, and $\Phi = (\varphi_{-f_c}, \ldots, \varphi_{f_c})$.
- Assume $(\varepsilon_k)_{k=0}^m$ are random complex Gaussian:

$$\varepsilon_k = \varepsilon_k^{(1)} + \mathrm{i}\, \varepsilon_k^{(2)},$$

where $\varepsilon_k^{(1)}, \varepsilon_k^{(2)}, k \in \{-f_c, \ldots, f_c\}$ are i.i.d. centered Gaussian random variables with standard deviation σ:

$$\varepsilon_k^{(1)} \sim \varepsilon_k^{(2)} \sim \mathcal{N}(0, \sigma^2).$$

We mention that $\varepsilon = (\varepsilon_{-f_c}, \ldots, \varepsilon_{f_c})$.
- Finally, we recall that we observe $y = \int_{\mathbb{T}} \Phi \, \mathrm{d}\Delta + \varepsilon$.

Our results show that if the spikes are sufficiently separated, at least $2.5/f_c$ apart, then one can detect some point sources with a known precision solving a simple convex optimization program.

Definition 3.1 (Minimum separation [6]). *For a family of points $\mathbf{S} \subset \mathbb{T}$, the minimum separation is defined as the closest distance between any two elements from $\mathbf{S}$:*

$$\ell(\mathbf{S}) = \inf_{\substack{(x,x') \in \mathbf{S}^2 \\ x \neq x'}} |x - x'| \, .$$

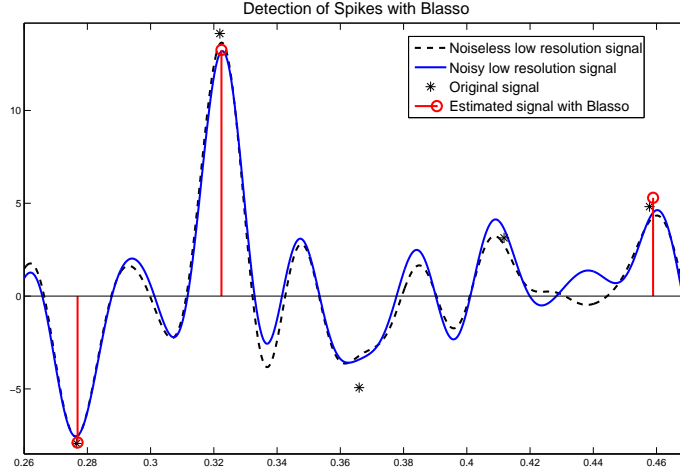*We emphasize that the distance is taken around the circle so that, for example $|5/6 - 1/6| = 1/3$.*

FIGURE 1. The problem is the following: we aim at recovering some spikes of the original signal (stars ∗) from the observation of a corrupted optical device (blue line) which can differ heavily from the true noiseless observation (black dotted line). Our procedure (red circles) provides a close estimate of the location of some spikes.

In this framework, we have the following theorem that quantifies the support detection error of BLASSO.

Corollary 1. *Assume that $f_c \geq 128$. Let Δ be a discrete measure such that:*

$$(3.1) \qquad \qquad \ell(\mathbf{S}) \geq \frac{2.5}{f_c} \, ,$$

where $\mathbf{S}$ denotes the support of Δ. Let $\hat{\Delta}$ be a solution to (BLASSO) with tuning parameter λ such that:

$$\lambda \geq \lambda_F := 2\sigma \sqrt{6 f_c \log(f_c)} \, .$$

then, with probability greater than $1 - 2\exp[-(\log f_c) \lambda^2 / \lambda_F^2]$, the following holds.

*   *For all $x \in [0, 1]$ such that:*

$$(3.2) \qquad \qquad |\hat{\Delta}(\{x\})| \geq 218 \lambda \, ,$$

   there exists a unique $T \in \mathbf{S}$ satisfying:

$$|x - T| \leq \left[\frac{\lambda}{0.1678 \, |\hat{\Delta}(\{x\})|} \right]^{1/2} \frac{1}{f_c} \leq \frac{0.1649}{f_c} \, ,$$

- *For all $T_j \in \mathbf{S}$ corresponding to a weight such that:*

$$\Delta_j > 218\,\lambda\,,$$

there exists $\hat{T} \in \mathrm{Supp}(\hat{\Delta})$ satisfying:

$$|T_j - \hat{T}| \leq \left[\frac{\lambda}{0.1678(\Delta_j - 218\lambda)}\right]^{1/2} \frac{1}{f_c}\,.$$

This result provides a quantitative estimate of the location of spikes in terms of the output amplitudes or the input amplitudes.

Remark. Observe that our procedure do not suppose any knowledge on the total number of spikes. This property is of great importance in actual practice. Only the minimal distance between any pair of atoms is relevant to BLASSO.

Remark. The constant 2.5 in the minimal separation assumption (3.1) is not optimal. However, it can be lowered to 2 using the result [6] but the constants in Corollary 1 seems too large in this case. For instance, the constant 218 in (3.2) would be greater than 10^3. So, we choose to keep 2.5 in (3.1) so as to get not too large constants.

3.2. **Detection from noisy moment samples.** In this section, we investigate the diffraction limit from moment measurements. In particular, we try to understand how BLASSO can super-resolve spikes from noisy moment samples.

- The number of samples is $m + 1$.
- For sake of simplicity, consider that $\mathbb{T} = [-1, 1]$.
- Consider $\mathbf{F}_C = \{\varphi_0, \ldots, \varphi_m\}$ defined by $T_0 = 1$ and for all $k \in \{1, \ldots, m\}$,

(3.3) $$\varphi_k = \sqrt{2}\,T_k\,,$$

  where $T_k(x) = \cos(k \arccos(x))$ is the k-th Chebyshev polynomial of the first kind. Hence, the family $\mathbf{F}_C$ is an orthonormal family with respect to the probability measure $\Pi(\mathrm{d}t) = (1/\pi)\,(1 - t^2)^{-1/2}\,\mathcal{L}(\mathrm{d}t)$ on $[-1, 1]$ where $\mathcal{L}$ denotes the Lebesgue measure.
- Assume ε_k are i.i.d. centered Gaussian random variables with standard deviation σ. We mention that $\varepsilon = (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_m)$.
- Finally, we recall that we observe $y = \int_{\mathbb{T}} \Phi \,\mathrm{d}\Delta + \varepsilon$.

Corollary 2. *Let $m \geq 9$. Let Δ be a discrete measure such that its support $\mathbf{S}$ is $2c_0$ far from the endpoints of $\mathbb{T}$, namely ± 1. Assume that Δ satisfies $\mathrm{QIC}(C_a, C_b)$ with $C_b/C_a \leq c_0^2$. Let $\hat{\Delta}$ be a solution to (BLASSO) with tuning parameter λ such that:*

$$\lambda \geq \lambda_M := \sigma\sqrt{6m \log m}\,.$$

then, with probability greater than $1 - 8\frac{\lambda}{\lambda_M}\sqrt{\log m}\,\exp\left[-(2\frac{\lambda^2}{\lambda_M^2} - 1)\log m\right]$, the following holds.

- *For all $x \in [-1, 1]$ such that:*

$$|\hat{\Delta}(\{x\})| > \frac{2\lambda}{C_b}\,,$$

there exists a unique $T \in \mathbf{S}$ satisfying:

$$|x - T| \leq \left[\frac{2\lambda}{C_a\,|\hat{\Delta}(\{x\})|}\right]^{1/2} \frac{1}{m} \leq \frac{c_0}{m}\,,$$

- *Let $C' = 2 + \max\left\{ \frac{2(1-C_b)}{C_b}, \frac{4}{C_a-C_b} \right\}$. Then, for all $T_j \in \mathbf{S}$ corresponding to a weight such that:*

$$\Delta_j > C'\lambda,$$

there exists $\hat{T} \in \mathrm{Supp}(\hat{\Delta})$ satisfying:

$$|T_j - \hat{T}| \leq \left[\frac{2\lambda}{C_a(\Delta_j - C'\lambda)} \right]^{1/2} \frac{1}{m}.$$

4. RICE METHOD

We use the Rice method as a convenient tool to upper bound the supremum of a random polynomial on a compact set. Proofs can be found in F.

4.1. Polynomial case.

Consider the family $\mathbf{F}_C = \{\varphi_0, \dots, \varphi_m\}$ defined by 3.3. Define the Gaussian process $\{X_m(t), \ t \in [-1,1]\}$ by:

$$\forall t \in [-1,1], \quad X_m(t) = \xi_0\varphi_0(t) + \xi_1\varphi_1(t) + \xi_2\varphi_2(t) + \dots + \xi_m\varphi_m(t),$$

where $\xi_1, \dots, \xi_m$ are i.i.d. standard normal. Its covariance function is

$$r(s,t) = 1 + \varphi_1(t)\varphi_1(s) + \varphi_2(t)\varphi_2(s) + \dots + \varphi_1(m)\varphi_m(s),$$

where the dependence in m has been omitted. Observe that its maximal variance is attained at point 1 and is given by $\sigma_m^2 = 2m + 1$, and its variance function is $\sigma_m^2(t) = 1 + \varphi_1(t)^2 + \varphi_2(t)^2 + \dots + \varphi_1(m)^2$.

Proposition 4.1. *Let $M = \max_{t \in [-1,1]} |X_m(t)|$. Then, for $m \geq 12$ and for $u > \sqrt{1+2m}$,*

$$\mathbb{P}\{M > u\} \leq \frac{4m(1+u)}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{1+2m}\right).$$

4.2. Fourier case.

We consider the trigonometric functions:

$$\varphi_k(t) = \exp(i2\pi kt), \quad t \in [0,1] \text{ and } k \in \mathcal{K} := \{-f_c, \dots, f_c\},$$

and random complex Gaussian errors:

$$\varepsilon_k = \varepsilon_k^{(1)} + i\,\varepsilon_k^{(2)},$$

where the variables $\varepsilon_k^{(1)}, \varepsilon_k^{(2)}, k \in \mathcal{K}$ are independent with standard normal distribution.

Proposition 4.2. *Let $Z(t) = \sum\limits_{k \in \mathcal{K}} \varepsilon_k \varphi_k(t)$. Then, for $u > \sqrt{2}$,*

$$\mathbb{P}\{\sup_{t \in [0,1]} \|Z(t)\| > u\} \leq 4\left(\exp\left(-\frac{u^2}{2(2f_c+1)}\right) + \sqrt{\frac{f_c(f_c+1)}{3}}\exp\left(-\frac{u^2}{4(2f_c+1)}\right)\right).$$

APPENDIX A. PROOF OF THEOREM 2.1

Let $P = \sum\limits_{k=0}^{m} a_k\varphi_k$ given by the QIC condition such that P is a dual certificate of Δ. Set:

$$d = \|\hat{\Delta}\|_{TV} - \|\Delta\|_{TV} - \Re\left(\int_{\mathbb{T}} P\,\mathrm{d}(\hat{\Delta} - \Delta)\right),$$

where $\Re(.)$ denotes the real part. One can check that d belongs to the Bregman divergence of the TV-norm between $\hat{\Delta}$ and Δ. Hence, we know that it is non-negative. Indeed, observe that:

$$\Re\left(\int_{\mathbb{T}} P\,\mathrm{d}(\hat{\Delta} - \Delta)\right) = \Re\left(\int_{\mathbb{T}} P\,\mathrm{d}(\hat{\Delta})\right) - \|\Delta\|_{TV} \leq \|\hat{\Delta}\|_{TV} - \|\Delta\|_{TV}$$

since P is sub-gradient of the TV-norm at point Δ. From the definition of BLASSO, we have:

$$\frac{1}{2}\|\int_{\mathbb{T}} \Phi\, \mathrm{d}\hat{\Delta} - y\|_2^2 + \lambda\|\hat{\Delta}\|_{TV} \leq \frac{1}{2}\|\varepsilon\|_2^2 + \lambda\|\Delta\|_{TV}\,.$$

Therefore:

(A.1) $$\frac{1}{2}\|\int_{\mathbb{T}} \Phi\, \mathrm{d}\hat{\Delta} - y\|_2^2 + \lambda\, d + \lambda\, \Re(\int_{\mathbb{T}} P\, \mathrm{d}(\hat{\Delta} - \Delta)) \leq \frac{1}{2}\|\varepsilon\|_2^2\,.$$

It follows that:

$$\frac{1}{2}\|\int_{\mathbb{T}} \Phi\, \mathrm{d}\hat{\Delta} - y + \lambda\, \overline{a}\|_2^2 + \lambda\, d \leq \frac{1}{2}\|\lambda\, a\|_2^2 + \frac{1}{2}\|\varepsilon\|_2^2 - \lambda\, \langle \varepsilon\,, \overline{a}\rangle\,.$$

Eventually, we get:

$$d \leq \frac{\lambda}{2}\|\overline{a} - \frac{\varepsilon}{\lambda}\|_2^2\,.$$

Using the triangular inequality and Parseval's identity (we recall that $\mathbf{F}$ is an orthonormal family with respect to a probability measure Π on $\mathbb{T}$), it yields:

$$\|\overline{a} - \frac{\varepsilon}{\lambda}\|_2 \leq \|a\|_2 + \|\frac{\varepsilon}{\lambda}\|_2\,,$$
$$= \left(\int_{\mathbb{T}} |P|^2(x)\Pi(\mathrm{d}x)\right)^{1/2} + \frac{\|\varepsilon\|_2}{\lambda}\,.$$

Since $\lambda \geq \|\varepsilon\|_2$ and $\|P\|_\infty \leq 1$, we have:

(A.2) $$\|\overline{a} - \frac{\varepsilon}{\lambda}\|_2 \leq 2\,.$$

So we get that $d \leq 2\lambda$. Moreover, observe that:

$$d = \|\hat{\Delta}\|_{TV} - \Re(\int_{\mathbb{T}} P\, \mathrm{d}\hat{\Delta})\,,$$

(A.3) $$= \sum_{k=1}^n \hat{\Delta}_k \left[1 - |P|(\hat{t}_k)\, \cos(\hat{\theta}_k + \theta_P(\hat{T}_k))\right]\,,$$

(A.4) $$\geq \sum_{k=1}^n \hat{\Delta}_k \min\{C_a m^2 \min_{T\in\mathbf{S}} d(\hat{T}_k, T)^2, C_b\}\,.$$

This shows the results (1) and (2). Observe that the localization estimate in the statement of the theorem is a consequence of the aforementioned inequality.

APPENDIX B. PROOF OF THEOREM 2.2

We begin with a lemma on the optimality conditions of Blasso. For all $a \in \mathbb{C}^{m+1}$ and for all $\mu \in \mathcal{M}$, we denote

$$\langle a\,, \Phi\rangle = \sum_{k=0}^m \overline{a}_k \varphi_k \quad \text{and} \quad c(\mu) = \int_{\mathbb{T}} \Phi\, \mathrm{d}\mu,\,.$$

We recall some standard lemmas. The proofs are given in Appendix C.

Lemma B.1. *Any solution $\hat{\Delta}$ to* (BLASSO) *satisfies the following optimatlity conditions:*
  (1) $\|\langle c(\hat{\Delta}) - y\,, \Phi\rangle\|_\infty \leq \lambda$;
  (2) $\langle y - c(\hat{\Delta})\,, c(\hat{\Delta})\rangle = \lambda\|\hat{\Delta}\|_{TV}$.

Lemma B.2. *Set $v = \hat{\Delta} - \Delta$ then the following inequality holds:*

(B.1) $$\forall P \in \mathrm{Span}(\mathbf{F}), \quad |\int_{\mathbb{T}} P\, \mathrm{d}v| \leq (\lambda + \lambda_0)\|P\|_1\,,$$

where $\lambda_0 = \|\langle \varepsilon\,, \Phi\rangle\|_\infty$.

Using the BIP property and the QIC condition, we prove the following interpolation lemma.

Lemma B.3. *Let $j \in \{1, \ldots, s\}$ then there exists a generalized polynomial $Q_j \in \mathbf{F}$ satisfying the following properties:*

- $\forall k \in \{1, \ldots, s\}, \quad Q_j(T_k) = \delta_{k,j},$
- $\forall x \neq T_j, \quad |Q_j(x)| < 1,$
- $\forall x \in \mathbb{T}$ *such that* $d(x, T_j) \leq c_0/m$, *it holds:*

$$|1 - Q_j(x)| \leq (C_c/2) \, m^2 \, d(x, T_j)^2,$$

- $\forall k \in \{1, \ldots, s\} \setminus \{j\}, \forall x \in \mathbb{T}$ *such that* $d(x, T_k) \leq c_0/m$, *it holds:*

$$|Q_j(x)| \leq (C_c/2) \, m^2 \, d(x, T_k)^2,$$

- $\forall x \in \mathcal{F}(c_0, \mathbf{S}), \quad |Q_j(x)| \leq 1 - C_b.$

Using these interpolating polynomials and (A.4) we get that for all $j \in \{1, \ldots, s\}$:

$$\Big| \sum_{\{k:\, d(\hat{T}_k, T_j) > \frac{c_0}{m}\}} \hat{\Delta}_k \, \exp(i\hat{\theta}_k) \, Q_j(\hat{T}_k) + \sum_{\{k:\, d(\hat{T}_k, T_j) \leq \frac{c_0}{m}\}} \hat{\Delta}_k \, \exp(i\hat{\theta}_k) \, (Q_j(\hat{T}_k) - 1) \Big|,$$

$$\leq \sum_{\{k:\, d(\hat{T}_k, T_j) > \frac{c_0}{m}\}} \hat{\Delta}_k \, |Q_j(\hat{T}_k)| + \sum_{\{k:\, d(\hat{T}_k, T_j) \leq \frac{c_0}{m}\}} \hat{\Delta}_k \, |Q_j(\hat{T}_k) - 1|,$$

$$\leq \sum_{k=1}^{n} \hat{\Delta}_k \min\{(C_c/2) m^2 \min_{T \in \mathbf{S}} d(\hat{T}_k, T)^2, 1 - C_b\},$$

$$\leq \max\left\{ \frac{1 - C_b}{C_b}, \frac{C_c}{2C_a} \right\} \times \sum_{k=1}^{n} \hat{\Delta}_k \min\{C_a m^2 \min_{T \in \mathbf{S}} d(\hat{T}_k, T)^2, C_b\},$$

$$\leq \lambda \max\left\{ \frac{2(1 - C_b)}{C_b}, \frac{C_c}{C_a} \right\}.$$

Now, invoking (B.1) we get:

$$\Big| \Delta_j - \sum_{\{k:\, d(\hat{T}_k, T_j) \leq \frac{c_0}{m}\}} \hat{\Delta}_k \Big| = \Big| \int_{\mathbb{T}} Q_j \, d\Delta - \int_{\mathbb{T}} Q_j \, d\hat{\Delta} + \sum_{\{k:\, d(\hat{T}_k, T_j) > \frac{c_0}{m}\}} \hat{\Delta}_k \, \exp(i\hat{\theta}_k) \, Q_j(\hat{T}_k)$$

$$+ \sum_{\{k:\, d(\hat{T}_k, T_j) \leq \frac{c_0}{m}\}} \hat{\Delta}_k \, \exp(i\hat{\theta}_k) \, (Q_j(\hat{T}_k) - 1) \Big|,$$

$$\leq \left(2 + \max\left\{ \frac{2(1 - C_b)}{C_b}, \frac{C_c}{C_a} \right\} \right) \lambda.$$

This proves (3). From (1) and (2) given by Theorem 2.1, we deduce the last part of the theorem.

Appendix C. Proofs of the auxiliary lemmas

C.1. **Proof of Lemma B.1.** The convex function $f(\nu) = \frac{1}{2} \| \int_{\mathbb{T}} \Phi \, d\nu - y \|_2^2 + \lambda \|\nu\|_{TV}$ is minimized at point $\hat{\Delta}$ then $0 \in \partial f(\hat{\Delta})$ or equivalently, for all $\nu \in \mathcal{M}$,

$$(C.1) \qquad \lambda(\|\nu\|_{TV} - \|\hat{\Delta}\|_{TV}) \geq \langle y - c(\hat{\Delta}), c(\nu - \hat{\Delta}) \rangle.$$

Conversely, we can easily check that if (C.1) holds then $\hat{\Delta}$ is a minimizer. Thus, (C.1) is necessary and sufficient for $\hat{\Delta}$ to minimize $f(\nu)$. Observe that (C.1) leads to:

$$(C.2) \qquad \lambda \|\hat{\Delta}\|_{TV} - \langle y - c(\hat{\Delta}), c(\hat{\Delta}) \rangle \leq \inf_{\nu \in \mathcal{M}} \{\lambda \|\nu\|_{TV} - \langle y - c(\hat{\Delta}), c(\nu) \rangle\}.$$

Now, observe that, for all $z \in \mathbb{C}^{m+1}$,

$$\sup_{\nu \in \mathcal{M}} \{\langle z, c(\nu)\rangle - \|\nu\|_{TV}\} = \sup_{\nu \in \mathcal{M}} \left\{ \int_{\mathbb{T}} \langle z, \Phi\rangle \mathrm{d}\nu - \|\nu\|_{TV}\right\} = \begin{cases} 0 & \text{if } \|\langle z, \Phi\rangle\|_\infty \le 1; \\ \infty & \text{otherwise.} \end{cases}$$

Using this in (C.2), we find that $\hat{\Delta}$ is a minimizer if and only if (1) and (2) hold.

C.2. **Proof of Lemma B.2.** Let $\overline{a}$ be the coefficients of P. It holds

$$\int_{\mathbb{T}} P \, \mathrm{d}\nu = \int_{\mathbb{T}} \langle a, \Phi\rangle \, \mathrm{d}\nu,$$
$$= \langle a, \int_{\mathbb{T}} \Phi \, \mathrm{d}\nu\rangle_{\ell^2(\mathbb{C}^{m+1})},$$
$$= \langle a, c(\hat{\Delta}) - c(\Delta)\rangle_{\ell^2(\mathbb{C}^{m+1})},$$
$$= \langle P, E\rangle_{L^2_{\Pi}(\mathbb{T})},$$

using the Parseval identity where E denotes the trigonometric polynomial of order n such that $E = \langle c(\hat{\Delta}) - x(\Delta), \Phi\rangle$. By Hölder's inequality we have

$$|\int_{\mathbb{T}} P \, \mathrm{d}\nu| \le \|P\|_1 \|E\|_\infty.$$

Now, we have to upper bound $\|E\|_\infty$. By triangular inequality, we have

$$\|E\|_\infty \le \|\langle c(\hat{\Delta}) - y, \Phi\rangle\|_\infty + \|\langle \varepsilon, \Phi\rangle\|_\infty.$$

By Lemma B.1, we know that the first term in the right hand side is upper bounded by λ. By hypothesis, we know that the second term in the right hand side is upper bounded by λ_0.

C.3. **Proof of Lemma B.3.** Without loss of generality, assume that $j = 1$ and set $\theta_1 = 0, \tilde{\theta}_1 = 1, \theta_k = 0$ and $\tilde{\theta}_k = \pi$, for all $k \in \{2, \ldots, s\}$. Invoke QIC to get P and $\tilde{P}$ such that:

- $P(T_1) = \tilde{P}(T_1) = 1$,
- $\forall k \in \{2, \ldots, s\}, \quad P(T_k) = -\tilde{P}(T_k) = 1$,
- $\forall x \in \mathcal{F}(c_0, \mathbf{S}), \quad \max(|P(x)|, |\tilde{P}(x)|) \le 1 - C_b$.

Set $Q_1 = (1/2)(P + \tilde{P})$ then one can check that:

- $\forall k \in \{1, \ldots, s\}, \quad Q_1(T_k) = \delta_{k,1}$,
- $\forall t \ne T_1, \quad |Q_1(t)| < 1$,
- $\forall x \in \mathcal{F}(c_0, \mathbf{S}), \quad |Q_1(x)| \le 1 - C_b$,

where $\delta_{k,1}$ equals 0 if $k \ne 1$ and 1 if $k = 1$. Moreover, since P and $\tilde{P}$ reach their maxima at each point of T, their first derivative vanishes on T. Therefore, Q_1' vanishes on T. Using a Taylor's theorem with explicit remainder and BIP, we conclude the proof.

APPENDIX D. PROOF OF COROLLARY 1

We begin with a key result.

Theorem D.1 (Lemma 2.5 in [6]). *Let* $\mathbf{S} = \{T_1, \ldots, T_s\} \subset [0, 1]$ *be the support of the target measure* Δ. *If* $\ell(\mathbf{S}) \ge 2.5/f_c$ *then there exists* $P \in \mathrm{Span}(\mathbf{F})$ *such that for all* $k = 1, \ldots, s$:

- $P(T_k) = \exp(-\mathbf{i}\theta_k)$,
- *Bound on the Taylor expansion at point* T_k:

$$\forall x \in \left[T_k - \frac{0.3313}{m}, T_k + \frac{0.3313}{m}\right], \quad |P(x)| \le 1 - 0.0839 \, m^2 \, (x - T_k)^2,$$

- *Bound on the complement:*

$$\forall x \in [0,1] \setminus \bigcup_{k=1}^{s} \left[T_k - \frac{0.3313}{m}, T_k + \frac{0.3313}{m} \right], \quad |P(x)| \leq 1 - 0.0092.$$

Hence, the support **S** enjoys QIC$(0.0838, 0.0092)$ and Theorem 2.1 applies. Moreover, the support **S** satisfies BIP$(1, \pi^2)$, see (2.2), and Theorem 2.2 can be invoked. Using the identity, for $a > 0$ and $b > 0$,

(D.1) $$\forall u^2 \geq 2ab, \quad -\frac{u^2}{a} + b \leq -\frac{u^2}{2a},$$

and Propositon 4.2, we get that $\lambda \geq \|\sum_{k=-f_c}^{f_c} \overline{\varepsilon_k} \varphi_k\|_\infty$ with a probability described in the statement of Corollary 1.

APPENDIX E. PROOF OF COROLLARY 2

Observe that **S** satisfies BIP$(c_0, 4/(1 - c_0^2))$ thanks to (2.3). Take

$$u = \sqrt{6m \log m} \, \lambda/\lambda_0,$$

in Propositon 4.1. Hence, we get that $\lambda \geq \|\sum_{k=0}^{m} \overline{\varepsilon_k} \varphi_k\|_\infty$ with a probability described as in the statement of Corollary 2. Therefore, Theorems 2.1 and 2.2 give the result.

APPENDIX F. RICE FORMULA

F.1. **Polynomial case.** By the change of variables $t = \cos\theta$, for all $t \in [-1, 1]$:

$$X_m(t) = X_m(\cos\theta) = \xi_0 + \sqrt{2}\,\xi_1 \cos(\theta) + \sqrt{2}\,\xi_2 \cos(2\theta) + \ldots + \sqrt{2}\,\xi_m \cos(m\theta).$$

Set $T_m(\theta) := X_m(t)$. We recall that its variance function is given by:

$$\sigma_m^2(\theta) = 1 + 2\cos^2(\theta) + 2\cos^2(2\theta) + \ldots + 2\cos^2(m\theta) = \frac{1 + 2m + \mathbf{D}_m(2\theta)}{2},$$

where $\mathbf{D}_m$ denotes the Dirichlet kernel of order m. Observe that:

$$\forall \theta \in \mathbb{R}, \quad \frac{1 + 2m}{3} = \sigma_m^2(3\pi/(2 + 4m)) \leq \sigma_m^2(\theta) \leq 1 + 2m,$$

so that $T_m\left(\frac{3\pi}{2(1+2m)}\right) \sim \mathcal{N}(0, \frac{1+2m}{3})$. By the Rice method [1], for $u > 0$:

$$\mathbb{P}\{M > u\} \leq 2\mathbb{P}\{\max_{\theta \in [0,\pi]} T_m(\theta) > u\},$$

$$\leq 2\mathbb{P}\{T_m(0) > u\} + 2\,\mathbb{E}[U_u([0,\pi])],$$

$$= 2[1 - \Psi\left(\frac{u}{\sqrt{1+2m}}\right)] + 2\int_0^\pi \mathbb{E}\left((T_m'(\theta))^+ \big| T_m(\theta) = u\right) \psi_{\sigma_m(\theta)}(u) \mathrm{d}t,$$

where U_u is the number of crossings of the level u, Ψ is the standard normal distribution, and ψ_σ is the density of the centered normal distribution with standard error σ. First, observe that for $v > 0$, $(1 - \Psi(v)) \leq \exp(-v^2/2)$. Hence,

$$1 - \Psi\left(\frac{u}{\sqrt{1+2m}}\right) \leq \exp\left(-\frac{u^2}{1+2m}\right).$$

Moreover, regression formulas implies that:

$$\mathbb{E}\left(T_m'(\theta) \big| T_m(\theta) = u\right) = \frac{r_{0,1}(\theta, \theta)}{r(\theta, \theta)} u,$$

$$\mathrm{Var}\left(T_m'(\theta) \big| T_m(\theta) = u\right) \leq \mathrm{Var}\left(T_m'(\theta)\right) = r_{1,1}(\theta, \theta),$$

where, for instance, $r_{1,1}(\nu, \theta) = \frac{\partial^2 r(\nu,\theta)}{\partial \nu \partial \theta}$. We recall that the covariance function is given by:

$$r(\nu, \theta) = 1 + 2\cos(\nu)\cos(\theta) + 2\cos(2\nu)\cos(2\theta) + \ldots + 2\cos(m\nu)\cos(m\theta),$$

$$= \frac{1}{2}\left[\mathbf{D}_m(\nu - \theta) + \mathbf{D}_m(\nu + \theta)\right].$$

Observe that:

$$r_{0,1}(\theta, \theta) = \frac{1}{2}\mathbf{D}'_m(2\theta) = -\sum_{k=1}^{m} k\sin(2k\theta),$$

$$\text{and} \quad r_{1,1}(\theta, \theta) = \frac{(2m+1)(m+1)m}{6} - \sum_{k=1}^{m} k^2\cos(2k\theta).$$

On the other hand, if $Z \sim \mathcal{N}(\mu, \sigma^2)$ then

$$\mathbb{E}(Z^+) = \mu\,\Psi\left(\frac{\mu}{\sigma}\right) + \sigma\,\psi\left(\frac{\mu}{\sigma}\right) \le \mu^+ + \frac{\sigma}{\sqrt{2\pi}},$$

where ψ is the standard normal density. We get that:

$$\int_0^\pi \mathbb{E}\left((T'_m(\theta))^+ \,\big|\, T_m(\theta) = u\right)\psi_{\sigma_m(\theta)}(u)\mathrm{d}t \le \int_0^\pi \frac{[\mathbf{D}'_m(2\theta)]^+}{2\,\sigma_m^2(\theta)}\,u\psi_{\sigma_m(\theta)}(u)\mathrm{d}\theta$$

$$+ \frac{1}{\sqrt{2\pi}}\int_0^\pi \left[\frac{(2m+1)(m+1)m}{6} - \sum_{k=1}^{m} k^2\cos(2k\theta)\right]^{1/2} \psi_{\sigma_m(\theta)}(u)\mathrm{d}\theta,$$

$$= A + B.$$

We use the following straightforward relations:

- $\forall\, 0 < \sigma_1 < \sigma_2 < u$, $\quad \psi_{\sigma_1}(u) \le \psi_{\sigma_2}(u)$,
- $\forall\theta$, $\quad [\mathbf{D}'_m(2\theta)]^+ \le \frac{m(m+1)}{2}$,
- $\forall\theta$, $\quad \frac{1+2m}{3} \le \sigma_m^2(\theta)$,

Eventually, we get, for $u > \sqrt{2m+1}$:

$$A \le \frac{3\pi}{2}\, m\, u\psi_{\sqrt{2m+1}}(u),$$

$$B \le \left[\frac{\pi}{12}(2m+1)(m+1)m\right]^{1/2} \psi_{\sqrt{2m+1}}(u),$$

and therefore:

$$\mathbb{P}\{M > u\} \le 2\exp\left(-\frac{u^2}{1+2m}\right) + 2\left[\frac{3\pi\, mu}{2\sqrt{2m+1}} + \left(\frac{\pi}{12}m(m+1)\right)^{1/2}\right]\psi\left(\frac{u}{\sqrt{2m+1}}\right).$$

Morover, for $m \ge 12$, $\pi m(m+1)/12 \le m^2$ and $3\pi\, mu/(2\sqrt{2m+1}) \le mu$, yielding:

$$\mathbb{P}\{M > u\} \le \frac{4m(1+u)}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{1+2m}\right).$$

F.2. **Fourier case.** We have:

$$Z(t) = \varepsilon_0^{(1)} + \sum_{k=1}^{f_c}\left(\varepsilon_k^{(1)} + \varepsilon_{-k}^{(1)}\right)\cos(2\pi kt) + \left(\varepsilon_k^{(2)} - \varepsilon_{-k}^{(2)}\right)\sin(2\pi kt)$$

$$+ \mathrm{i}\left[\varepsilon_0^{(2)} + \sum_{k=1}^{f_c}\left(\varepsilon_k^{(2)} + \varepsilon_{-k}^{(2)}\right)\cos(2\pi kt) + \left(\varepsilon_k^{(1)} - \varepsilon_{-k}^{(1)}\right)\sin(2\pi kt)\right].$$

One can see that $Z(t) = X(t) + \mathrm{i}\,Y(t)$ where $X(t)$ and $Y(t)$ are two independent Gaussian stationary processes with the same auto-covariance function:

$$\Gamma(t) = 1 + 2\sum_{k=1}^{f_c}\cos(2\pi k t) = \mathbf{D}_{f_c}(t)\,,$$

where $\mathbf{D}_{f_c}(t)$ denotes the Dirichlet kernel. Set:

$$\sigma_m{}^2 = \mathrm{Var}(X(t)) = \mathbf{D}_{f_c}(0) = 2f_c + 1\,.$$

We use the following inequalities:

$$\mathbb{P}\{\|Z\|_\infty > u\} \le \mathbb{P}\{\|X\|_\infty > u/\sqrt{2}\} + \mathbb{P}\{\|Y\|_\infty > u/\sqrt{2}\}\,,$$
$$= 2\mathbb{P}\{\|X\|_\infty > u/\sqrt{2}\}\,.$$

and

(F.1) $$\mathbb{P}\{\|X\|_\infty > u/\sqrt{2}\} \le 2\mathbb{P}\{\sup_{t\in[0,1]} X(t) > u/\sqrt{2}\}\,.$$

To give bounds to the right hand side of (F.1), we use the Rice method [1] using the fact that the process $X(t)$ (for example) is periodic with $\Gamma(\frac{1}{2f_c+1}) = 0$:

$$\mathbb{P}\{\sup_{t\in[0,1]} X(t) > u/\sqrt{2}\} = \mathbb{P}\{\forall t \in [0,1]; X(t) > u/\sqrt{2}\} + \mathbb{P}\{U_{u/\sqrt{2}} > 0\}\,,$$

$$\le \left(\overline{\Psi}(u/(\sqrt{2}\sigma_m))\right)^2 + \mathbb{E}(U_{u/\sqrt{2}})\,,$$

where U_v is the number of up-crossings of the level v by the process $X(t)$ on the interval $[0,1]$ and $\overline{\Psi}$ is the tail of the standard normal distribution. By the Rice formula:

$$\mathbb{E}(U_{u/\sqrt{2}}) = \frac{1}{2\pi}\sqrt{\mathrm{Var}(X'(t))}\frac{1}{\sigma_m}\exp(-\frac{u^2}{4\sigma_m{}^2})\,,$$

where:

$$\mathrm{Var}(X'(t)) = -\Gamma''(0) = 2(2\pi)^2\sum_{k=1}^{f_c}k^2 = \frac{4\pi^2}{3}f_c(f_c+1)(2f_c+1)\,.$$

The following inequality is well known : for $v > 0$, $\overline{\Psi}(v) \le \exp(-v^2/2)$, it yields:

$$\mathbb{P}\{\sup_{t\in[0,1]} X(t) > \frac{u}{\sqrt{2}}\} \le \exp(-\frac{u^2}{2(2f_c+1)}) + \sqrt{\frac{f_c(f_c+1)}{3}}\exp\left(-\frac{u^2}{4(2f_c+1)}\right)\,.$$

The result follows.

REFERENCES

[1] J.-M. Azaïs and M. Wschebor. A general expression for the distribution of the maximum of a gaussian field and the approximation of the tail. *Stochastic Processes and their Applications*, 118(7):1190 – 1218, 2008.

[2] P.B. Borwein and T. Erdélyi. *Polynomials and polynomial inequalities*. Springer Verlag, 1995.

[3] K. Bredies and H. Pikkarainen. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 1(1), 2010.

[4] Martin Burger and Stanley Osher. Convergence rates of convex variational regularization. *Inverse problems*, 20(5):1411, 2004.

[5] E.J. Candes and C. Fernandez-Granda. Super-resolution from noisy data. *arXiv preprint arXiv:1211.0290*, 2012.

[6] E.J. Candes and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 2012.

[7] Y. de Castro and F. Gamboa. Exact reconstruction using beurling minimal extrapolation. *Journal of Mathematical Analysis and Applications*, 2012.

[8] D.L. Donoho. Superresolution via sparsity constraints. *SIAM Journal on Mathematical Analysis*, 23(5):1309–1331, 1992.

[9] P. Doukhan and F. Gamboa. Superresolution rates in prokhorov metric. *Canad. J. Math.*, 48(2):316–329, 1996.

[10] V. Duval and G. Peyré. Exact support recovery for sparse spikes deconvolution. *arXiv preprint arXiv:1306.6909*, 2013.

[11] C. Fernandez-Granda. Support detection in super-resolution. *arXiv preprint arXiv:1302.3921*, 2013.

[12] F. Gamboa and E. Gassiat. Sets of superresolution and the maximum entropy method on the mean. *SIAM journal on mathematical analysis*, 27(4):1129–1152, 1996.

[13] B Hofmann, B Kaltenbacher, C Poeschl, and O Scherzer. A convergence rates result for tikhonov regularization in banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987, 2007.

[14] MG Krein and AA Nudelman. The markov moment problem and extremal problems, volume 50 of translations of mathematical monographs. *American Mathematical Society, Providence, Rhode Island*, 1977.

[15] J. Lindberg. Mathematical concepts of optical superresolution. *Journal of Optics*, 14(8):083001–1, 2012.

[16] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine, IEEE*, 20(3):21–36, 2003.

[17] E Resmerita and O Scherzer. Error estimates for non-quadratic regularization and the relation to enhancement. *Inverse Problems*, 22(3):801, 2006.

[18] W. Rudin. Real and complex analysis, mcgraw&hill series in higher mathematics, 1987.

[19] O Scherzer. *Variational methods in imaging*, volume 167. Springer, 2009.

[20] Otmar Scherzer and Birgit Walch. Sparsity regularization for radon measures. In *Scale Space and Variational Methods in Computer Vision*, pages 452–463. Springer, 2009.

[21] V. Studer, J. Bobin, M. Chahid, H.S. Mousavi, E. Candes, and M. Dahan. Compressive fluorescence microscopy for biological and hyperspectral imaging. *Proceedings of the National Academy of Sciences*, 109(26):E1679–E1687, 2012.

[22] G. Tang, B. Narayan Bhaskar, and B. Recht. Near minimax line spectral estimation. *arXiv preprint arXiv:1303.4348*, 2013.

[23] R Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288, 1996.

JMA and FG are with the Institut de Mathématiques de Toulouse (CNRS UMR 5219). Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France.

E-mail address: {jean-marc.azais},{fabrice.gamboa}@math.univ-toulouse.fr
URL: www.math.univ-toulouse.fr/~{azais},{gamboa}

YdC is with the Département de Mathématiques (CNRS UMR 8628), Bâtiment 425, Faculté des Sciences d'Orsay, Université Paris-Sud 11, F-91405 Orsay Cedex, France.

E-mail address: yohann.decastro@math.u-psud.fr
URL: www.math.u-psud.fr/~decastro