SPIKE DETECTION FROM INACCURATE SAMPLINGS

JEAN-MARC AZAÏS, YOHANN DE CASTRO, AND FABRICE GAMBOA

Abstract. This article investigates the super-resolution phenomenon using the celebrated statistical estimator LASSO in the complex valued measure framework. More precisely, we study the recovery of a discrete measure (spike train) from few noisy observations (Fourier samples, moments, Stieltjes transformation...). In particular, we provide an explicit quantitative localization of the spikes. Moreover, our analysis is based on the Rice method and provide an upper bound on the supremum of white noise perturbation in the measure space.

1. Introduction

1.1. **Super-resolution.** In some situations, experiments can be subject to device limitations where one cannot observe enough information in order to recover fine details from an image. 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 breaking the diffraction limit of an optical system or the issues arising in source separation. Many companion applications in astronomy, medical imaging and microscopy are at stake [9, 7, 6] and theoretical guarantees of source detection are of crucial importance in practice. For instance, the idea of this paper gives a process to compute a quantitative estimate of the localization of the active molecules in Single Molecule Imaging in 3D Microscopy [15].

This paper offers quantitative detection guarantees from noisy observations (Fourier samples, moments, Stieltjes transformation...). The authors provide a tractable algorithm (BLASSO) and quantitative estimates of a train of complex valued spikes from few noisy observations.

Similarly, P. Doukhan, E. Gassiat and one author of this present paper [10, 12] considered the exact reconstruction of a nonnegative measure. More precisely, they 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 [8] proved that k spikes trains can be faithfully resolved from $m = 2k + 1$ samples (Fourier, Stieltjes transformation, Laplace transform, ...) by using total-variation method.

Last but not least, our analysis involves an estimate of the magnitude of the noise perturbation in the signal domain using the Rice method, see for example

Date: January 28, 2013.

Key words and phrases. Super-resolution; LASSO; Signed measure; Bregman divergence; Semidefinite programming; Compressed Sensing.

[1]. In particular, we derive explicit bounds for the tunning parameter appearing in BLASSO.

1.2. **General model and notation.** Let $\mathbb{T}$ be a compact set homeomorphic to either the interval $[0,1]$ or the unit circle $\mathbb{S}^1$ (which is identified as $\mathbb{R} \mod (2\pi)$ via the mapping $z = e^{it}$). Let Δ be a complex measure on $\mathbb{T}$ with discrete support of (unknown) size s. In particular, Δ has polar decomposition (see [14] for a definition):

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

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 $\mathcal{F} = \{\varphi_0, \varphi_1, \ldots, \varphi_m\}$ be a family of complex continuous functions on $\mathbb{T}$. 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 all the indices $k = 0,1,\ldots,m$. Assume that we observe $y = (y_k)_{k=0}^m$ defined as:

$$\forall\, k \in \{0,1,\ldots,m\}\,,\quad y_k = c_k(\Delta) + \varepsilon_k\,,$$

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 $\Phi = (\varphi_0, \ldots, \varphi_m)$. We aim at reconstructing the complex measure Δ from the $m+1$ 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 would be described therein. If not specified, notation are in accordance with the general model.

1.3. **Beurling LASSO (BLASSO).** Denote by $\mathcal{M}$ the set of finite complex measures on $\mathbb{T}$ and by $\|\,.\,\|_{TV}$ the total variation norm. We recall that for all $\mu \in \mathcal{M}$,

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

where the supremum is taken over all partitions Π of $\mathbb{T}$ into a finite number of disjoint measurable subsets. For further details, we refer the reader to [14].

Remark. We mention that the TV-norm considered in this paper is not the usual TV-norm of signal processing which is essentially the ℓ_1-norm of the ℓ_2-norms of the finite differences at any point. In particular, our model has nothing to do with the Rudin-Osher-Fatemi (ROF) model [13].

By analogy with the LASSO [16], *Beurling LASSO* (BLASSO) is the process of reconstructing a discrete measure Δ from the samples y by finding a solution to:

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

where λ is a tuning parameter.

Remark. For the case of Fourier coefficients and $\varepsilon = 0$, (BLASSO) is simply *Beurling Minimal Extrapolation* [2]. Moreover, in the finite dimension framework (i.e. $\mathbb{T}$ should be viewed as $\mathbb{R}^n$), BLASSO is nothing else than LASSO. BLASSO is named after this remark.

Remark. The factor $1/2$ in the definition of BLASSO plays no role, except for simplifying the proofs.

1.4. **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 spike train (discrete measure) from noisy band-limited data [6]. In this setting, we observe noisy Fourier samples up until a frequency cut-off $f_c \in \mathbb{N}^*$. We shall 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 $x \in [0, 1]$, $\varphi_k(x) = \exp(\mathrm{i}\,2\pi kx)$, and $\Phi = (\varphi_{-f_c}, \ldots, \varphi_{f_c})$.
- Assume ε_k 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 (or atoms) are sufficiently separated, at least $2/f_c$ apart, then one can detect some point sources with a known precision solving a simple convex optimization program.

Definition 1.1 (Minimum separation [7])**.** *For a family of points $S \subset \mathbb{T}$, the minimum separation is defined as the closest distance between any two elements from S:*
$$\ell(S) = \inf_{\substack{(x,x') \in S^2 \\ x \neq x'}} |x - x'|\,.$$
We emphasis that the distance is taken around the circle so that, for example $|5/6 - 1/6| = 1/3$.

In this framework, we have the following theorem that quantifies the BLASSO stability.

Theorem 1.1. *Let Δ be a discrete measure such that:*
$$\ell(\mathrm{Supp}(\Delta)) \geq \frac{2}{f_c}\,,$$
where $\mathrm{Supp}(\Delta)$ denotes the support of Δ. Let $\hat{\Delta}$ be a solution to (BLASSO) with tuning parameter λ such that:
$$\lambda \geq \lambda_F := \sigma\,\sqrt{12\, f_c\, \log(f_c)}\,.$$
then, with probability greater than
$$1 - 8\exp\big(-\lambda^2/\lambda_F^2\big)$$
the following holds. For all $t \in [0, 1]$ such that:
$$|\hat{\Delta}(\{t\})| \geq \frac{2\lambda}{C_b}\,,$$
there exists a unique $T \in \mathrm{Supp}(\Delta)$ satisfying:
$$|t - T| \leq \Big(\frac{2\lambda}{C_a |\hat{\Delta}(\{t\})|}\Big)^{1/2} \frac{1}{f_c} \leq \frac{0.16}{f_c}\,,$$

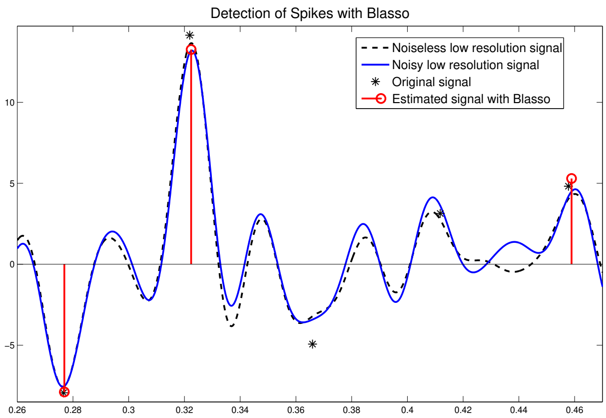

FIGURE 1. The problem is the following: we aim at recovering some spikes of the original signal (fives stars $*$) from the observation through a corrupted optical device (blue line) which can differ heavily from the true noiseless observation (black dotted line). Our procedure (red circles) provides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

where $0 < C_b \leq (0.16)^2 C_a < 1$ *are universal constants.*

Remark. Furthermore, a solution $\hat{\Delta}$ to BLASSO can be efficiently computed using a companion SDP program. We refer the reader to Section 6.2 for further details.

This result is new and interesting because it provides a quantitative estimate of the location of spikes. To the best of our knowledge, this is the first result of this kind in the literature.

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

Remark. Unlike the finite dimension case [5] (where $\mathbb{T}$ should be viewed as $\mathbb{R}^n$ and BLASSO is simply LASSO), we have numerically witnessed to the following fact: the solutions $\hat{\Delta}$ to BLASSO have generally less atoms than the target Δ (i.e. the size of the estimated support is often less or equal than the size of the true support). For instance, Figure 1 shows that the BLASSO solution has support of size 3 while the target has support of size 5. It seems that there is no hope to recover *all* atoms of the target measure. Hence, we can only hope to recover the low hanging fruits, namely the large atoms. Hence, our result quantify the localization of these large atoms.

1.5. **Detection from noisy moment samples.** In the frame of a polynomial system, a companion corollary of Theorem 1.1 can be given using the Rice inequality given in Proposition 5.1 (that gives a way of tunning λ) and the Szegó mapping (as done in [8]). Notice that, in this case the minimal spacing between the support points is no more uniform as the Szegó induced a non linear distortion.

1.6. **Comparison with related work.** The problem of super-resolution without noise has been investigated in numerous articles [9, 10, 12] (this list is not meant to be exhaustive). In particular, we mention [8, 7] which exhibit the notion of dual certificates in the measure framework. Hence, in [8] the authors investigates TV-norm minimization with different types of measurements: trigonometric, polynomial, Laplace transform... Furthermore, the article [7] provides an explicit construction of a tight dual certificate P in the Fourier case and gives an upper bound on the magnitude of P at each point. We mention [4] which aim at approximating the solution by estimating the support of the signal in an iterative fashion.

In the case of noisy measurements, [6] derives a stability result for a weighted ℓ_1 distance between measures (the weight function is given by a high-frequency kernel). In contrast, our result result provides a quantitative localization of the spike train, which is crucial in applications.

To the best of our knowledge, this paper is the first work on a quantitative detection of atoms.

1.7. **Organization of the paper.** The next section present a general definition of measures that can be detected using BLASSO. Section 3 gives the main result and a key lemma on the localization of the solution to BLASSO. Section 4 provides some examples of target measures that can be considered in our framework. Section 5 presents some results on Gaussian processes that are useful in the super-resolution framework. The last part is devoted to the proofs.

2. Separable measures (SM)

As a matter of fact, our framework deals with more general samplings (or observations). We begin with the definition of a dual certificate (see [8, 7]).

Definition 2.1 ((Tight) Dual certificate)**.** *We say that a linear combination*

$$P = \sum_{k=0}^{m} a_k \varphi_k$$

is a dual certificate of a discrete measure which polar decomposition is given by:

$$\mu = \sum_{k=1}^{n} \mu_k \exp(\mathrm{i}\,\theta_k) \delta_{x_k}\,,$$

where $\mu_k > 0$, ***if*** *in addition*

- $\forall\, k \in \{1, \ldots, n\}\,,\ P(x_k) = \exp(-\mathrm{i}\,\theta_k)$,
- $\forall\, x \in \mathbb{T}\,,\ |P(x)| \leq 1$.

Similarly, we say that P is a "tight dual certificate" ***if*** *P is a dual certificate and*

- $\forall\, x \in \mathbb{T} \setminus \{x_1, \ldots, x_n\}\,,\ |P(x)| < 1$.

The set of all the dual certificates (resp. tight dual certificates) of a measure μ is denoted by $\mathcal{P}(\mu)$ (resp. $\mathcal{P}_0(\mu)$).

We use the dual certificate to give the definition of SM measures.

Definition 2.2 ((Tight) Separable Measure ((T)SM))**.** *We say that a discrete measure μ is a separable measure (SM) (resp. tight separable measure (TSM)) with respect to a family $\mathcal{F} = \{\varphi_0, \varphi_1, \ldots, \varphi_m\}$* ***if*** *it has at least a dual certificate (resp. tight dual certificates).*

Remark. Observe that every solution to a total-variation regularization method is SM. Indeed, one can prove that the target Δ is a solution of the following total-variation method:

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

if and only if Δ is SM (a proof can be found in [8]). We understand that if Δ is not SM there is no hope in recovering Δ with a total-variation method. Moreover, one can prove [8] that if the target Δ is TSM then it is the unique solution to (GME).

This remark shows that TSM is a natural assumption in TV-minimization.

- **Assumption:** From now on, we assume that Δ is TSM.

3. Main result

3.1. **Confinement.** The analysis of BLASSO uses the extremal properties of the TV-norm. In particular, it is known that the extreme points of the unit ball of the TV-norm are the atoms δ_x for any $x \in \mathbb{T}$. Thus the TV-norm minimization forces the solutions to be discrete measures. Furthermore, we recall that:

$$\forall \nu \in \mathcal{M}\,, \quad \|\nu\|_{TV} = \sup_{\|f\|_\infty \le 1} \Re\Big(\int_{\mathbb{T}} f \mathrm{d}\mu\Big)\,,$$

where $\Re(\cdot)$ denotes the real part. It follows that the sub-gradient of the TV-norm is given by:

$$\partial\|\cdot\|_{TV}(\nu) := \Big\{ f \in L^\infty(\mathbb{T})\,; \quad \|\mu\|_{TV} - \|\nu\|_{TV} - \Re\Big(\int_{\mathbb{T}} f\, \mathrm{d}(\nu-\mu)\Big) \ge 0\,,\ \forall \mu \in \mathcal{M} \Big\}\,.$$

We mention that this sub-differential can be further characterized as:

$$\begin{aligned}\partial\|\cdot\|_{TV}(\nu) := \Big\{ f \in L^\infty(\mathbb{T})\,; \quad & \Re\Big(\int_{\mathbb{T}} f\, \mathrm{d}\nu\Big) = \|\nu\|_{TV}\,, \\ & \text{and } \Re\Big(\int_{\mathbb{T}} f\, \mathrm{d}\mu\Big) \le \|\mu\|_{TV}\,,\ \forall \mu \in \mathcal{M} \Big\}\end{aligned}$$

Another tool we shall use in our theorems is the notion of Bregman divergence. It is defined as follows.

Definition 3.1 (Bregman divergence)**.** *Let μ and ν be two complex measures such that ν is TSM. Define the Bregman divergence as:*

$$D(\mu,\nu) = \Big\{ \|\mu\|_{TV} - \Re\Big(\int_{\mathbb{T}} P\, \mathrm{d}\mu\Big)\,; \quad P \in \mathcal{P}_0(\nu) \Big\}\,, \tag{3.1}$$

We recall that $\mathcal{P}_0(\nu)$ is the set of the tight dual certificates of ν.

Remark. One can check that the Bregman divergence at point (μ,ν) is an interval of $[0,+\infty[$.

Remark. Strictly speaking, the "Bregman divergence" considered in this paper is not the usual Bregman divergence (see [4] for instance) since we restrict P to $\mathcal{P}_0(\nu)$ while the standard Bregman divergence for the total-variation norm should be:

$$\Big\{ \|\mu\|_{TV} - \Re\Big(\int_{\mathbb{T}} P\, \mathrm{d}\mu\Big)\,; \quad \Re\Big(\int_{\mathbb{T}} P \mathrm{d}\nu\Big) = \|\nu\|_{TV} \text{ and } \|P\|_\infty = 1 \Big\}\,. \tag{3.2}$$

As a matter of fact, (3.1) is the intersection of (3.2) with the set of all possible linear combinations of the family $\mathcal{F}$.

Bregman divergence is not a common distance functional since it does not satisfy the triangle inequality and it is not symmetric. However, Bregman divergence is non-negative and allows us to localize the support of discrete measures. The next lemma shows that measures with small divergence are "close".

Lemma 3.1 (Confinement). *Let μ and ν be two complex measures such that ν is TSM. Assume that μ is discrete with a polar decomposition $\mu = \sum_{k=1}^{n} \mu_k \exp(\mathrm{i}\,\theta_k)\delta_{x_k}$, where $\mu_k > 0$. We have the following properties:*

- *Let d be in $D(\mu,\nu)$ then*

$$d = \sum_{k=1}^{n} \mu_k \left[1 - |P|(x_k)\,\cos(\theta_k + \theta_P(x_k))\right],$$

where

$$P(x) = |P|(x)\exp(\mathrm{i}\,\theta_P(x)),$$

is a polar decomposition of a tight dual certificate P (at point x) of ν.

- *In particular, for all $k = 1,\dots,n$,*

$$x_k \in \bigcap_{P\in\mathcal{P}_0(\nu)} |P|^{-1}(I_k),$$

and

$$\theta_k \in \arccos(I_k) \boxplus \bigcap_{P\in\mathcal{P}_0(\nu)} \{(-\theta_P)(|P|^{-1}(I_k)) \boxplus \{2p\pi;\ p\in\mathbb{Z}\}\},$$

where $I_k = [1-(d/\mu_k),1]$ and $\boxplus$ denotes the Minkowski sum.

A proof of this lemma can be found in Section A.1. In other words, the support $\{x\in\mathbb{T}\,;\ |\mu(\{x\})|\geq\rho\}$ is included in the level set:

$$\bigcap_{P\in\mathcal{P}_0(\nu)} |P|^{-1}([1-(d/\rho),1]),$$

and the support $\{\theta\,;\ \exists\, x\in\mathbb{T} \text{ s.t. } \mu(\{x\}) = \mu_x\exp(\mathrm{i}\,\theta) \text{ and } \mu_x\geq\rho\}$ is included in the level set:

$$\arccos([1-(d/\rho),1]) \boxplus \bigcap_{P\in\mathcal{P}_0(\nu)} \Big\{(-\theta_P)(|P|^{-1}([1-(d/\rho),1])) \boxplus \{2p\pi;\ p\in\mathbb{Z}\}\Big\},$$

where $\boxplus$ denotes the Minkowski sum.

3.2. Main theorem.

Theorem 3.2. *Let Δ be a Tight Separable Measure such that:*

$$P = \sum_{k=0}^{m} a_k\varphi_k,$$

is a tight dual certificate. Let $\hat\Delta$ be a solution to (BLASSO) then it holds:

$$D(\hat\Delta,\Delta) \leq \min\left[\frac{\lambda}{2}\|a\|_2^2 + \frac{1}{\lambda}\Re(\langle\int_{\mathbb{T}}\Phi\,\mathrm{d}(\hat\Delta-\Delta)\,,\,\varepsilon\rangle)\,;\ \frac{\lambda}{2}\|a-\frac{\overline{\varepsilon}}{\lambda}\|_2^2\right]. \tag{3.3}$$

Moreover, if the tuning parameter λ is such that:

$$\lambda \geq \lambda_0 := \|\sum_{k=0}^{m}\overline{\varepsilon_k}\varphi_k\|_\infty.$$

then it holds:

$$D(\hat\Delta,\Delta) \leq \|a\|_2\sqrt{4\lambda\|\Delta\|_{TV}} + \frac{1}{2\lambda}\|\varepsilon\|_2^2. \tag{3.4}$$

A proof of this theorem can be found in Section A.3.

4. Examples

4.1. Nonnegative measures and standard moments. The nonnegative measures whose support has size s not greater than $m/2$ are tight separable measures with respect to the standard moments, i.e. $\varphi_k(x) = x^k$. Indeed, let Δ be a nonnegative measure and $\mathcal{S} = \{T_1, \ldots, T_s\}$ be its support. Following [3], set

$$P(x) = 1 - c\prod_{i=1}^{s}(x - T_i)^2 .$$

Then, for a sufficiently small value of the parameter c, the polynomial P has supremum norm not greater than 1. The existence of such a polynomial shows that the measure Δ is TSM [8].

4.2. Nonnegative measures and the Stieltjes transformation. Any nonnegative measure whose support has size s not greater than $m/2$ are SM [8] with respect to the family

$$\mathcal{F} = \Big\{1, \frac{1}{z_1 - x}, \frac{1}{z_2 - x}, \ldots\Big\} ,$$

where none of the z_k's belongs to $\mathbb{T}$.

4.3. Chebyshev measures and standard moments. Define the Chebyshev polynomials of the first order as:

$$T_k(x) = \cos(k \arccos(x)), \quad \forall x \in [-1, 1] .$$

It is well known (see [3] for instance) that it has supremum norm not greater than 1, and that

- T_k is equal to 1 on $\big\{\cos(2l\pi/k),\ l = 0, \ldots, \lfloor \frac{k}{2} \rfloor\big\}$,
- T_k is equal to -1 on $\big\{\cos((2l+1)\pi/k),\ l = 0, \ldots, \lfloor \frac{k}{2} \rfloor\big\}$,

whenever $k > 0$. Then, any real measure Δ which Jordan decomposition is given by $\Delta = \Delta^+ - \Delta^-$, and such that

- $\mathrm{Supp}(\Delta^+) \subseteq \big\{\cos(2l\pi/k),\ l = 0, \ldots, \lfloor \frac{k}{2} \rfloor\big\}$,
- $\mathrm{Supp}(\Delta^-) \subseteq \big\{\cos((2l+1)\pi/k),\ l = 0, \ldots, \lfloor \frac{k}{2} \rfloor\big\}$,

for some $0 < k \leq n$, is SM.

4.4. Minimum separation measures. Last but not least, E.J. Candès and C. Fernandez-Granda [7, 6] have shown that tight dual certificates (with respect to the Fourier basis) exist for discrete complex measure satisfying a "minimum separation condition". Their Proposition 2.1 [7] and Lemma 2.4 [6] give an explicit construction using the Fejér kernel.

5. Rice method

5.1. Polynomial case. We consider the Gaussian process $X_m(t)$, $t \in [0, 1]$, defined by:

$$X_m(t) = \xi_0 + \xi_1 t + \xi_2 t^2 + \ldots + \xi_m t^m ,$$

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

$$r(s,t) = 1 + st + s^2t^2 + \ldots + s^m t^m ,$$

where the dependence in m has been omitted. Its maximal variance is:

$$\sigma_m^2 = m + 1.$$

Its variance function is:

$$\sigma_m^2(t) = 1 + t^2 + t^4 + \ldots + t^{2m}.$$

Proposition 5.1. *Let $M = \max_{t\in[0,1]} |X_m(t)|$. Then, for $u > 2\sqrt{m+1}$,*

$$\mathbb{P}\{M > u\} \leq 2\Big[\frac{(m+1)\sqrt{\pi}u + m\sqrt{2}}{2\sqrt{\pi m}}\Big]\psi(u/\sqrt{m+1}) + 2(1-\Psi(u)),$$

where ψ and Ψ are respectively the standard normal density and distribution .

For sake of completeness, a proof can be found in A.2.

5.2. **Fourier case.** We consider the trigonometric functions:

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

and random complex Gaussian errors:

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

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

Proposition 5.2. *Let $Z(t) = \sum_{k\in\mathcal{K}} \varepsilon_k \varphi_k(t)$. Then, for $u > \sqrt{2}$,*

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

For sake of completeness, a proof can be found in A.2.

6. Numerical experiments

6.1. **Fenchel dual program.** The usual convex analysis shows that (BLASSO) can be viewed as a Fenchel dual problem (see [17, 4] for a definition). As a matter of fact, any solution to (BLASSO) can be faithfully computed from a companion program that builds a dual certificate of $\hat{\Delta}$.

Proposition 6.1 ([17, 4]). *The problem*

$$\min_{a\in\mathbb{C}^{m+1}} \frac{\|a-y\|_2^2}{2} + \mathbb{I}_{\{a\in\mathbb{C}^{m+1}\,;\; \|\sum_{k=0}^m a_k\varphi_k\|_\infty \leq \lambda\}}(a) \tag{6.1}$$

has its Fenchel dual with the same minimizers as (BLASSO). Here, the indicator $\mathbb{I}_E(v)$ of a set $E \subset \mathbb{C}^{m+1}$ is defined by $\mathbb{I}_E(v) = 0$ if $v \in E$ and $\mathbb{I}_E(v) = +\infty$ otherwise.

Using the predual problem (6.1), it is possible to derive optimality conditions for (BLASSO). Hence, we mention that all solution to (BLASSO) is SM as shown by Proposition 3 in [4] (their analysis extends naturally to the complex field).

Proposition 6.2 ([4]). *The optimization problem (BLASSO) admits at least a solution. Moreover, all solution $\hat{\Delta}$ is SM and it has a dual certificate $\hat{P} = \sum_{k=0}^m \hat{a}_k\varphi_k$ where*

$$\forall k \in \{0,\ldots,m\}, \quad \hat{a}_k = \frac{c_k(\hat{\Delta}) - y_k}{\lambda}. \tag{6.2}$$

Remark. We have an explicit formulation of a dual certificate $\hat{P}$ of $\hat{\Delta}$ using (6.2). Moreover, all solution to (BLASSO) is discrete, SM and satisfies:

$$\{x \in \mathbb{T}\,;\quad |\hat{\Delta}(\{x\})| > 0\} \subseteq \{x \in \mathbb{T}\,;\quad |\hat{P}|(x) = 1\}. \tag{6.3}$$

In other words, the support of $\hat{\Delta}$ is included in the set of the points for which $|\hat{P}|$ is maximal.

On the algorithmic side, the program (6.1) allows us to compute a dual certificate of a solution $\hat{\Delta}$ to (BLASSO). As a matter of fact, it takes the form:

$$\text{(Dual BLASSO)} \qquad \hat{a} \in \arg \min_{a \in \mathbb{C}^{m+1}} \|a - \frac{y}{\lambda}\|_2^2 \quad \text{subject to } \|\sum_{k=0}^{m} a_k \varphi_k\|_\infty \leq 1 \,.$$

By definition of a dual certificate, the support of $\hat{\Delta}$ is located at the points where the dual certificate $\hat{P} = \sum_{k=0}^{m} \hat{a}_k \varphi_k$ has modulus equal to 1, its maximal value. Once the support is estimated accurately, a solution to (BLASSO) can be found by solving a well-posed linear problem.

6.2. **Fourier Case.** In this subsection, we follow notation of the Fourier case described in Section 1.4. At first glance, the program (Dual BLASSO) seems difficult to solve due to the ℓ_∞ norm in the hard constraint. As pointed out by [7], this difficulty can be circumvented using the following lemma.

Lemma 6.3 (Corollary to Theorem 4.24 in [11])**.** *A trigonometric polynomial:*

$$P = \sum_{k=-f_c}^{f_c} a_k \exp(\mathrm{i}\, 2\pi k t)$$

*is bounded by one in magnitude **if and only if** there exists a Hermitian matrix* $Q \in \mathbb{C}^{(m+1)\times(m+1)}$ *satisfying:*

$$\begin{pmatrix} Q & a \\ a^* & 1 \end{pmatrix} \succeq 0 \quad \text{and} \quad \sum_{i=1}^{m+1-j} Q_{i,i+j} = \begin{cases} 1\,, & \text{if } j = 0\,, \\ 0\,, & \text{if } j = 1, \ldots, m\,. \end{cases} \tag{6.4}$$

We deduce that (Dual BLASSO) is equivalent to the following Semi-Definite Program:

$$\hat{a} \in \arg \min_{a \in \mathbb{C}^{m+1}} \|a - \frac{y}{\lambda}\|_2^2 \quad \text{subject to (6.4)}\,. \tag{6.5}$$

This program gives the coefficients of a dual certificate $\hat{P}$ of $\hat{\Delta}$. The level set of the maximum value of $|\hat{P}|$ contains the support of $\hat{\Delta}$, see (6.3). It suffices to solve a regular LASSO on these points to deduce the values of the weights of $\hat{\Delta}$.

Remark. As pointed out by [7], it is not clear that $\hat{P}$ is not constant and hence the level set of the maximum value of $|\hat{P}|$ is discrete. However, we have run several numerical experiments, they all show that $\hat{P}$ is not constant. We devote the theoretical analysis of this fact to future work.

Appendix A. Proofs

A.1. **Lemma 3.1.** The lemma follows from the identity:

$$\begin{aligned} d &= \|\mu\|_{TV} - \Re\Big(\int_{\mathbb{T}} P \,\mathrm{d}\mu\Big)\,, \\ &= \sum_{k=1}^{n} \mu_k \left[1 - |P|(x_k)\, \cos(\theta_k + \theta_P(x_k))\right]\,, \end{aligned}$$

hence

$$|P|(x_k)\, \cos(\theta_k + \theta_P(x_k)) \geq 1 - \frac{d}{\mu_k}\,.$$

We conclude considering the level sets associated with the value $1 - (d/\mu_k)$.

A.2. **Rice formula.**

Polynomial case. By the Rice method [1], for $u > 0$,

$$\begin{aligned}\mathbb{P}\{M > u\} &\leq 2\mathbb{P}\{\max_{t\in[0,1]} X_m(t) > u\}\,,\\ &\leq 2\mathbb{P}\{X_m(0) > u + 2\mathbb{E}(U_u[0,1])\}\,,\\ &= 2(1-\Psi(u)) + 2\int_0^1 \mathbb{E}\big((X_m'(t))^+\big|X_m(t)=u\big)\psi_{\sigma_m(t)}(u)dt\,,\end{aligned}$$

where U_u is the number of crossings of the level u and ψ_σ is the density of the centered normal distribution with standard error σ. Regression formulas implies that:

$$\begin{aligned}\mathbb{E}\big(X_m'(t)\big|X_m(t)=u\big) &= \frac{r_{0,1}(t,t)}{r(t,t)}u\,,\\ \mathrm{Var}\,\big(X_m'(t)\big|X_m(t)=u\big) &\leq \mathrm{Var}\,\big(X_m'(t)\big) = r_{1,1}(t,t)\,,\end{aligned}$$

where, for instance $r_{1,1}(s,t) = \frac{\partial^2 r(s,t)}{\partial s \partial t}$. We have:

$$\begin{aligned}r_{0,1}(t,t) &= t + 2t^3 + \cdots + mt^{2m-1}\,,\\ r_{1,1}(t,t) &= 1 + 4t^2 + \cdots + m^2t^{2m-2}\,.\end{aligned}$$

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

$$\mathbb{E}(Z^+) = \mu\Psi(\mu/\sigma) + \sigma\psi(\mu/\sigma) \leq \mu^+ + \frac{\sigma}{\sqrt{2\pi}}\,.$$

We get that:

$$\begin{aligned}\mathbb{P}\{M > u\} &\leq 2(1-\Psi(u)) + 2\Big(\int_0^1 \frac{t+2t^3+\cdots+mt^{2n-1}}{t^2+\cdots+t^{2m}} u\psi_{\sigma_m(t)}(u)dt\\ &+ \frac{1}{\sqrt{2\pi}}\int_0^1 \sqrt{1+4t^2+\cdots+m^2t^{2n-2}}\,\psi_{\sigma_m(t)}(u)dt\Big)\\ &:= 2(1-\Psi(u)) + 2(A+B)\end{aligned}$$

We use the following straightforward relations:

- for $\sigma_1 < \sigma_2 < u$, $\psi_{\sigma_1}(u) \leq \psi_{\sigma_2}(u)$,
- for $u > 2$ and $\sigma < 1$, $\sigma^{-2}\psi_\sigma(u) \leq \psi(u)$,
- and
$$\sqrt{A_1 + \cdots A_m} \leq \sqrt{A_1} + \cdots + \sqrt{A_m}\,.$$

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

$$\begin{aligned}A &\leq \frac{m+1}{2}u\psi_{\sigma_m}(u)\,,\\ B &\leq \frac{m}{\sqrt{2\pi}}\psi_{\sigma_m}(u)\,.\end{aligned}$$

and we are done. □

Fourier case. We have:

$$\begin{aligned}Z(t) &= \varepsilon_0^{(1)} + \sum_{k=1}^{f_c}\big(\varepsilon_k^{(1)} + \varepsilon_{-k}^{(1)}\big)\cos(2\pi kt) + \big(\varepsilon_k^{(2)} - \varepsilon_{-k}^{(2)}\big)\sin(2\pi kt)\\ &+ \mathrm{i}\Big[\varepsilon_0^{(2)} + \sum_{k=1}^{f_c}\big(\varepsilon_k^{(2)} + \varepsilon_{-k}^{(2)}\big)\cos(2\pi kt) + \big(\varepsilon_k^{(1)} - \varepsilon_{-k}^{(1)}\big)\sin(2\pi kt)\Big].\end{aligned}$$

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:

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

where $D_{f_c}(t)$ denotes the Dirichlet Kernel. Set:

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

We use the following inequalities:

$$\begin{aligned}\mathbb{P}\{\|Z\|_\infty > u\} &\leq \mathbb{P}\{\|X\|_\infty > u/\sqrt{2}\} + P\{\|Y\|_\infty > u/\sqrt{2}\}\,,\\ &= 2\mathbb{P}\{\|X\|_\infty > u/\sqrt{2}\}\,.\end{aligned}$$

and

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

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

$$\begin{aligned}\mathbb{P}\{\sup_{t\in[0,1]} X(t) > u/\sqrt{2}\} &= \mathbb{P}\{\forall t \in [0,1]; X(t) > u/\sqrt{2}\} + \mathbb{P}\{U_{u/\sqrt{2}} > 0\}\,,\\ &\leq \Big(\overline{\Psi}(u/(\sqrt{2}\sigma_m))\Big)^2 + \mathbb{E}(U_{u/\sqrt{2}})\,,\end{aligned}$$

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 $v > 0$, $\overline{\Psi}(v) \leq \exp(-v^2/2)$, it yields:

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

The result follows. □

A.3. **Theorem 3.2.** Let $a = (a_i)_{i=0}^m$ be the coefficients of a dual certificate P of Δ:

$$P = \sum_{k=0}^m a_k\varphi_k := \langle \Phi\,,\,\overline{a}\rangle\,.$$

The corresponding element d of $D(\hat{\Delta},\Delta)$ is given by the expression:

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

From the definition of BLASSO, we know that

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

It holds:

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

• From (A.2), we deduce the following inequalities:

First inequality: Since $y=\int\Phi\,\mathrm{d}\Delta+\varepsilon$, it follows that:

$$\frac{1}{2}\|\int_{\mathbb{T}}\Phi\,\mathrm{d}(\hat{\Delta}-\Delta)\|_2^2-\Re(\langle\int_{\mathbb{T}}\Phi\,\mathrm{d}(\hat{\Delta}-\Delta)\,,\,\varepsilon\rangle)+\lambda\,d$$
$$+\lambda\,\Re(\langle\int_{\mathbb{T}}\Phi\,\mathrm{d}(\hat{\Delta}-\Delta)\,,\,\overline{a}\rangle)\leq 0\,.$$

And so:

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

A simple calculation gives that:

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

Eventually, we get:

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

Second inequality: We have:

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

Eventually, we get:

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

• If $\lambda\geq\lambda_0:=\|\sum_{k=0}^m\overline{\varepsilon_k}\varphi_k\|_\infty$ then we have the following result.

Lemma A.1. *Under the same hypothesis as Theorem 3.2, it holds:*

$$\|\int_{\mathbb{T}}\Phi\,\mathrm{d}(\hat{\Delta}-\Delta)\|_2^2\leq 2\lambda\|\Delta\|_{TV}\,.$$

Proof. From the definition of BLASSO, we know that

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

Since $y=\int\Phi\,\mathrm{d}\Delta+\varepsilon$, it follows that:

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

By linearity, we have:

$$\frac{1}{2}\|\int_{\mathbb{T}}\Phi\,\mathrm{d}(\hat{\Delta}-\Delta)\|_2^2\leq\Re(\int\langle\Phi\,,\,\varepsilon\rangle\,\mathrm{d}(\hat{\Delta}-\Delta))+\lambda(\|\Delta\|_{TV}-\|\hat{\Delta}\|_{TV})\,,$$

where

$$\langle\Phi\,,\,\varepsilon\rangle=\sum_{k=0}^m\overline{\varepsilon_k}\varphi_k\,.$$

Set $\lambda_0=\|\langle\Phi\,,\,\varepsilon\rangle\|_\infty$ then

$$\frac{1}{2}\|\int_{\mathbb{T}}\Phi\,\mathrm{d}(\hat{\Delta}-\Delta)\|_2^2\leq\lambda_0(\|\hat{\Delta}\|_{TV}+\|\Delta\|_{TV})+\lambda(\|\Delta\|_{TV}-\|\hat{\Delta}\|_{TV})\,.$$

Since $\lambda\geq\lambda_0$, it holds:

$$\|\int_{\mathbb{T}}\Phi\,\mathrm{d}(\hat{\Delta}-\Delta)\|_2^2\leq 2\lambda\|\Delta\|_{TV}\,.$$

□

From (A.2), it holds:

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

Using the Cauchy-Schwarz inequality and the previous lemma:

$$d \leq \|a\|_2 \sqrt{2\lambda \|\Delta\|_{TV}} + \frac{1}{2\lambda}\|\varepsilon\|_2^2\,.$$

A.4. **Theorem 1.1.** We begin with a key lemma [7].

Lemma A.2 (Tight dual certificate [7])**.** *Let $T = \{T_1, \ldots, T_s\} \subset [0,1]$ be the support of the target measure Δ. We recall that:*

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

If $\lambda(T) \geq \dfrac{2}{f_c}$ then there exists a tight dual certificate P such that:

$$\forall\, x \in [0,1]\,, \quad P(x) = \sum_{k=-f_c}^{f_c} a_k \exp(\mathbf{i}2\pi k x)\,,$$

satisfying the following properties for all $k = 1, \ldots, s$:

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

$$\forall x \in \Big[T_k - \frac{0.16}{f_c},\, T_k + \frac{0.16}{f_c}\Big]\,, \quad |P(x)| \leq 1 - C_a\, f_c^2\, (x - T_k)^2\,,$$

- *Bound on the complement:*

$$\forall x \in [0,1] \setminus \bigcup_{k=1}^{s} \Big[T_k - \frac{0.16}{f_c},\, T_k + \frac{0.16}{f_c}\Big]\,, \quad |P(x)| \leq 1 - C_b\,,$$

with $0 < C_b \leq 0.16\, C_a < 1$ universal constants.

Proof of Theorem 1.1. The Rice method ensures that $\lambda \geq \lambda_0 := \|\sum_{k=-f_c}^{f_c} \overline{\varepsilon_k} \varphi_k\|_\infty$ with a probability described in the statement of Theorem 1.1. The previous lemma shows that Δ is TSM and gives an explicit upper on the dual certificate P. The hypotheses of Theorem 3.2 are matched and we shall invoke (3.3) to get:

$$D(\hat{\Delta}, \Delta) \leq \frac{\lambda}{2}\|a - \frac{\overline{\varepsilon}}{\lambda}\|_2^2\,,$$

Using the triangular inequality and Parseval's identity, it yields:

$$\begin{aligned}\|a - \frac{\overline{\varepsilon}}{\lambda}\|_2 &\leq \|a\|_2 + \|\frac{\overline{\varepsilon}}{\lambda}\|_2 \\ &= \Big(\int_{[0,1]} |P|^2(x)\mathrm{d}x\Big)^{1/2} + \frac{1}{\lambda}\Big(\int_{[0,1]} |\sum_{k=-f_c}^{f_c} \overline{\varepsilon_k}\varphi_k|^2(x)\mathrm{d}x\Big)^{1/2}\end{aligned}$$

Since $\|\sum_{k=-f_c}^{f_c} \overline{\varepsilon_k}\varphi_k\|_\infty \leq \lambda$ and $\|P\|_\infty \leq 1$, we have:

$$\|a - \frac{\overline{\varepsilon}}{\lambda}\|_2 \leq 2\,. \tag{A.3}$$

Finally, it holds:

$$D(\hat{\Delta}, \Delta) \leq 2\lambda\,.$$

Now, let $t \in [0,1]$ such that:

$$|\hat{\Delta}(\{t\})| \geq \frac{2\lambda}{C_b}.$$

Let $d = D(\hat{\Delta}, \Delta)$ then (A.3) gives:

$$\frac{d}{|\hat{\Delta}(\{t\})|} \leq (0,16)^2 C_a ,$$

where C_a is the same constant as in Lemma A.2. The result follows invoking Lemma 3.1 and Lemma A.2. □

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] A. Beurling. Sur les intégrales de fourier absolument convergentes et leur application à une transformation fonctionnelle. In *Ninth Scandinavian Mathematical Congress*, pages 345–366, 1938.

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

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

[5] P. Bühlmann and S. Van De Geer. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, 2011.

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

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

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

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

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

[11] B. Dumitrescu. *Positive trigonometric polynomials and signal processing applications*. Springer, 2007.

[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] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

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

[15] 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.

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

[17] C. Zălinescu. *Convex analysis in general vector spaces*. World Scientific Publishing Company Incorporated, 2002.

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