Appl. Comput. Harmon. Anal. 48 (2020) 445–481

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Testing Gaussian process with applications to super-resolution

Jean-Marc Azaïs [a], Yohann De Castro [b,c,*], Stéphane Mourareau [d]

[a] *Institut de Mathématiques de Toulouse, Université Paul Sabatier 118 route de Narbonne, 31062 Toulouse, France*
[b] *Laboratoire de Mathématiques d'Orsay Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France*
[c] *INRIA, Centre de Recherche de Paris, Équipe MoKaPlan 2 rue Simone Iff, 75012 Paris, France*
[d] *Laboratoire d'Analyse et de Mathématiques Appliquées, Univ. Paris-Est Champs sur Marne, 77454 Marne La Vallée, France*

ARTICLE INFO

Article history:
Received 14 November 2017
Received in revised form 2 July 2018
Accepted 4 July 2018
Available online 10 July 2018
Communicated by Guy Nason

MSC:
62E15
62F03
60G15
62H10
62H15
60E05
60G10
62J05
94A08

Keywords:
Hypothesis testing
Gaussian process
Kac–Rice formula
Super-resolution

ABSTRACT

This article introduces exact testing procedures on the mean of a Gaussian process X derived from the outcomes of ℓ_1-minimization over the space of complex valued measures. The process X can be thought as the sum of two terms: first, the convolution between some kernel and a target atomic measure (mean of the process); second, a random perturbation by an additive centered Gaussian process. The first testing procedure considered is based on a dense sequence of grids on the index set of X and we establish that it converges (as the grid step tends to zero) to a randomized testing procedure: the decision of the test depends on the observation X and also on an independent random variable. The second testing procedure is based on the maxima and the Hessian of X in a grid-less manner. We show that both testing procedures can be performed when the variance is unknown (and the correlation function of X is known). These testing procedures can be used for the problem of deconvolution over the space of complex valued measures, and applications in frame of the Super-Resolution theory are presented. As a byproduct, numerical investigations may demonstrate that our grid-less method is more powerful (it detects sparse alternatives) than tests based on very thin grids.

* Corresponding author.
E-mail address: yohann.decastro@math.u-psud.fr (Y. De Castro).

https://doi.org/10.1016/j.acha.2018.07.001
1063-5203/

1. Introduction

1.1. Grid-less spike detection through the "continuous" LARS

New testing procedures based on the outcomes of ℓ_1 minimization methods have attracted a lot of attention in the statistical community. Of particular interest is the so-called "*Spacing test*", that we referred to as S^{ST}, based on the Least-Angle Regression Selection (LARS), that measures the significance of the addition of a new active variable along the LARS path, see [1, Chapter 6] for further details. Specifically, one is testing the relative distance between consecutive "*knots*" of the LARS, for instance $\lambda_{1,P}$ and $\lambda_{2,P}$. The first knot $\lambda_{1,P}$ is the maximal correlation between a response variable $y \in \mathbb{C}^N$ and P predictors. The second knot $\lambda_{2,P}$ is then the correlation between some residuals and $P-1$ predictors, and so on. This approach is now well referenced among the regularized methods of high-dimensional statistics and it can be linked to minimizing the ℓ_1-norm over P coordinates, see for instance [1, Chapter 6].

In this paper, we focus on ℓ_1-minimization over the space of signed measures and we ask for testing procedures based on these solutions. Indeed, in deconvolution problems over the space of measures [2]—*e.g.*, Super-Resolution or line spectral estimation [3–8]—one may observe a noisy version of a convolution of a target discrete measure by some known kernel K and one may be willing to infer on the target discrete measure. In this case, testing a particular measure is encompassed by testing the mean of some "correlation" process Z, see Section 6 for further details.

In general deconvolution problems, remark that there is an uncountable number of predictors with valued in a hilbert space (not necessarily finite)—while there were P predictors previously when inferring on vectors of $\mathbb{R}^N$ in the high-dimensional statistics framework. Indeed, we are looking at correlations $Z(t) = \langle y, k(t)\rangle$ between a response variable y and a "feature map" $k(t)$ indexed by a continuum, say for instance $t \in \mathbb{K} = [0, 2\pi)$. In this case, the set of predictors is uncountable and given by $\{k(t)\,;\ t \in \mathbb{K}\}$. Furthermore, $k(t)$ is an element of the Reproducing Kernel Hilbert Space $\mathcal{H}$ (RKHS) defined by the convolution kernel K—assumed to be symmetric positive definite. In particular, the hilbert space $\mathcal{H}$ can be infinite dimensional. As an example, assume that one observes some Fourier coefficients of some discrete measure on the torus $[0, 2\pi)$ and one is willing to infer on its support. A strategy would be to look at correlations between the response variable $y \in \mathbb{C}^N$ and the Fourier curve $k(t) = (\cos(kt) \pm \imath \sin(kt))_{-f_c \le k \le f_c} \in \mathbb{C}^N$ for some frequency cut-off $f_c \ge 1$ so that $N = 2f_c + 1$. It results in a complex valued correlation process $Z(t) := \langle y, k(t)\rangle = \sum y_k e^{\imath kt}$ indexed by $t \in [0, 2\pi)$. In this case, the RKHS $\mathcal{H}$ has dimension N, the number of observed Fourier coefficients, and the convolution kernel is given by the Dirichlet kernel, see Section 6. As an illustration, we present Fig. 1 where we take $Z_1 = Z$ and the red curve displays the absolute value of the correlation process Z. One can standardly show that $|Z|(t)$ is the likelihood of the model that consists in one spike at point t. Therefore, its maximal value λ_1 can be interpreted as the Maximum Likelihood for models with one spike. Its argument maximal point t_1 is then the Maximum Likelihood Estimator and one may be willing to consider it as a first estimation of the target discrete measure's support. Then one can consider the residuals $Z_2 = Z_1 - a\langle y, k(t_1)\rangle$ where $a \in \mathbb{C}$ is the weight of the estimated signal chosen so that we get the blue curve of Fig. 1, namely a second support point t_2 should enter the model since the residuals $|Z_2|$ achieve their maximal absolute value at two locations, t_1 and t_2. More details can be found in Section 2.4.

In this framework, the LARS algorithm does not return a sequence of entries (among P possible coordinates) and phases as in high-dimensional statistics but rather a sequence of locations $t_1, t_2, \dots$ (among the continuum $\mathbb{K} = [0, 2\pi)$) and phases. In this paper, we invoke the LARS to this framework—we referred to it as "continuous" LARS—for which an uncountable number of active variables may enter the model. We present this extension in Section 2 defining consecutive knots (λ_1, λ_2). One can wonder:

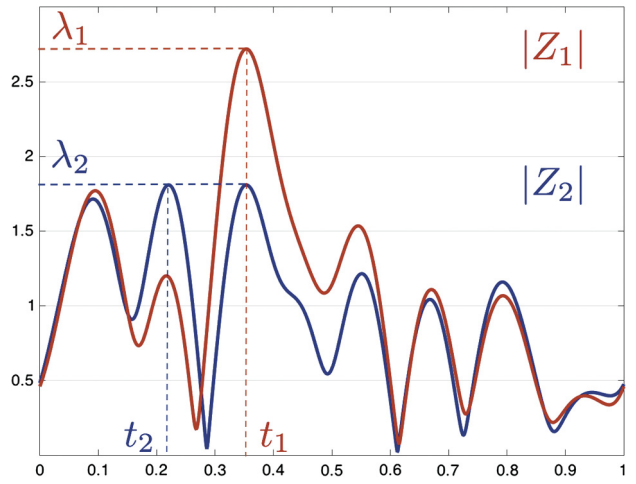

Fig. 1. LARS for Super-Resolution: we fit a Dirichlet kernel (which is the Point Spread Function of Super-Resolution) at the maximal correlation point t_1 until the maximal correlation in the residual is matched at a second point $t_2 \neq t_1$. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

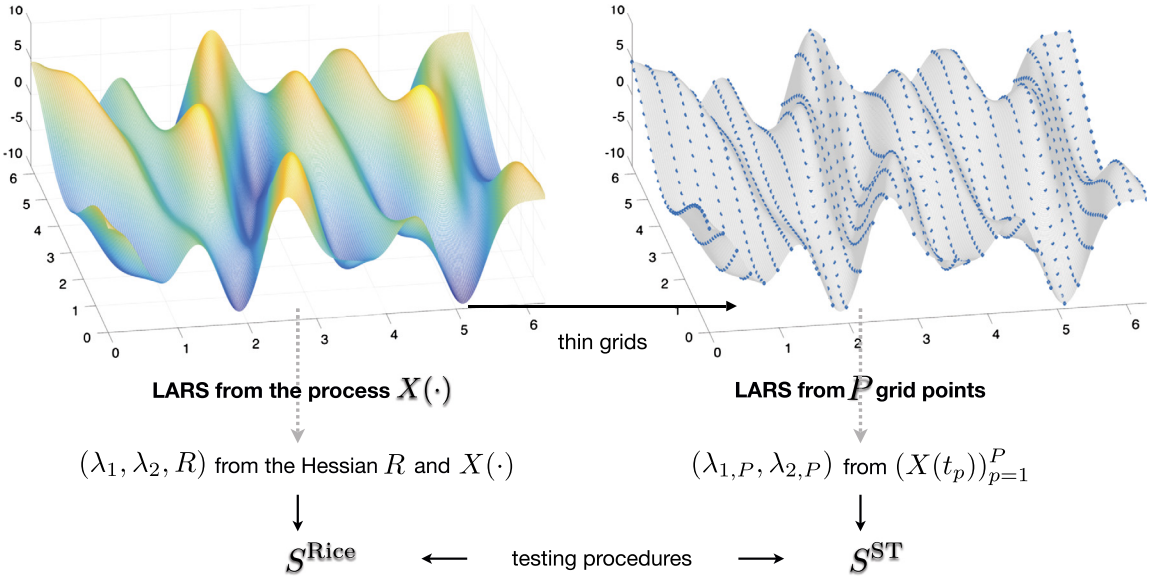

Fig. 2. The grid-less approach uses the Hessian and the first two "knots" (λ_1, λ_2) of the "continuous" LARS to build the test statistics S^{Rice}. We compare it to the grid approach that builds a test statistics S^{ST} using the knots $(\lambda_{1,P}, \lambda_{2,P})$ computed from a P points grid discretization $(X(t_p))_{p=1}^P$ of the continuous process X.

- *Can the Spacing test be used in the frame of Super-Resolution?*
- *Is there a grid-less procedure more powerful, in the sense of detecting spikes, than the Spacing tests constructed on thin grids?*

Interestingly, as we will prove, the answer is no to the first question if no modifications of the test statistic is done. Furthermore, the way that the Spacing test can be fixed to be extended to a "grid-less" frame gives a new testing procedure S^{Rice} that accounts for the distance between consecutive knots (λ_1, λ_2) with respect to value of the Hessian R at some maximal point, see Fig. 2 for a global view on our approach.

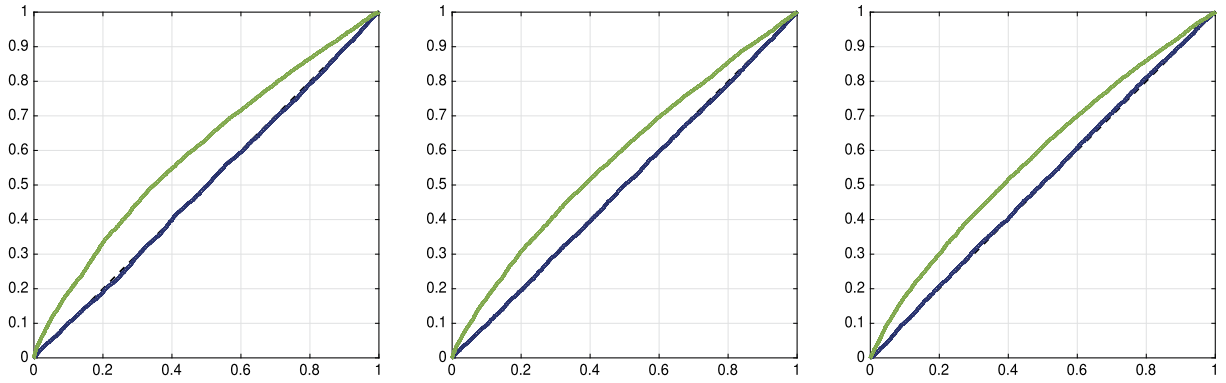

Fig. 3. **[Under the null]** Comparison of the empirical cumulative distribution of the two statistics S^{Rice} (blue line, see Theorem 4) and S^{ST} (green line) under the null hypothesis when applied to the consecutive knots (λ_1, λ_2) given by the "continuous" LARS in both cases. The diagonal (cdf of the uniform) is represented in dotted black line. The model is described by the Super-Resolution framework (see Section 6) with cutoff frequencies $f_c = 3, 5, 7$ from left to right. The new test statistic S^{Rice} is exactly distributed w.r.t. the uniform law on $[0, 1]$ under the null hypothesis.

1.2. A comparative study

When the predictors are normalized, the Spacing test (ST) statistics is given by the expression

$$S^{\mathrm{ST}}(\lambda_{1,P}, \lambda_{2,P}) := \frac{\overline{\Phi}(\lambda_{1,P})}{\overline{\Phi}(\lambda_{2,P})}$$

where $\overline{\Phi} = 1 - \Phi$ is the Gaussian survival function and Φ the standard normal cumulative distribution function. In the framework of high-dimensional statistics, this statistics is exactly distributed w.r.t. a uniform law on $[0, 1]$ under the global null, namely S^{ST} can be considered as the observed significance [9,10]. It is clear that one should not use this testing procedure in the Super-Resolution framework since there is no theoretical guarantees in this case. Yet the practitioner may be tempted to replace $(\lambda_{1,P}, \lambda_{2,P})$ by (λ_1, λ_2) given by the "continuous" LARS. Unfortunately, this paper shows that the resulting test statistics S^{ST} is non conservative in this frame, *i.e.*, it makes too many false rejections and one should avoid using it in practice, see the green line in Fig. 3.

To overcome this disappointing feature, one may be willing to consider thinner and thinner grids and look at the limit as P tends to infinity. In this case, one can show that $\lambda_{1,P}$ tends to the λ_1 of "continuous" LARS, but $\lambda_{2,P}$ does not converge to λ_2, it converges to $\overline{\lambda}_2$ as shown in (11). This results in a limit test that is a randomized version of the Spacing test that we referred to as S^{Grid} and presented in Theorem 1.

The second approach is to take a thin grid and to use S^{ST}. This approach is perfectly valid, this test statistics follows a uniform distribution under the null and it should be compared to our new testing procedure S^{Rice}. This numerical investigation has been performed in the frame of Super-Resolution and it is presented in Fig. 4, more details can be found in Section 6.2. Fig. 4 gives the cumulative distribution functions of the test statistics under "sparse" alternatives, *i.e.,* when true spikes are to be detected. The larger the power, the better the test detects spikes (abscissa represents the level of the test and the ordinate the probability to detect the spike). In these sets of experiments, we can note that

- *The testing procedure S^{Rice} based on some Hessian and the whole process $X(\cdot)$ is uniformly better than the spacing test even if one takes very thin grids.*

One can see that the power (the ability to detect sparse objects, Dirac masses here) of the grid methods seems to present a limit that is always improved by the continuous approach.

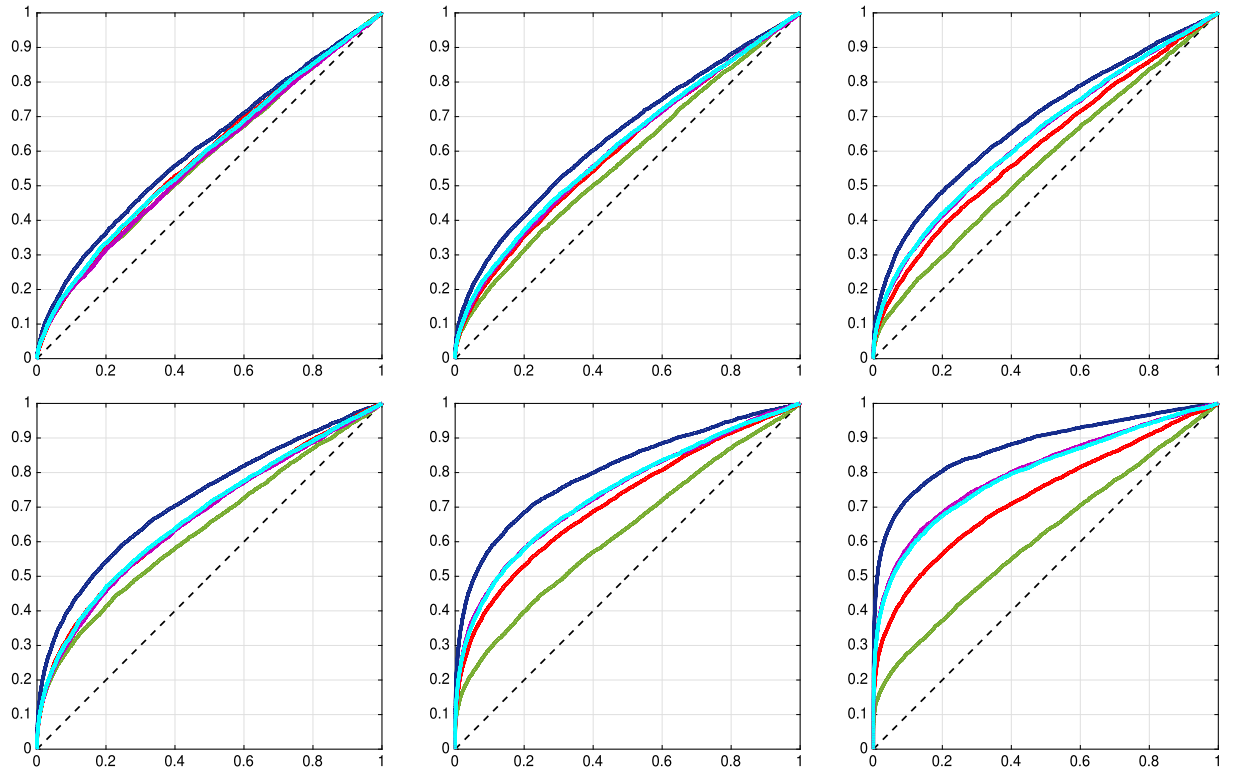

Fig. 4. [**Under the alternative**] Empirical cumulative distribution under the alternative of the Rice test (blue) and the discrete grid tests with size 3^2 (green), 10^2 (red), 32^2 (purple) and 50^2 (cyan). The alternative is defined by a single atom at a random location with a weight $\log(N) = \log(2f_c + 1)$ (first row) or $\sqrt{N}$ (second row). In columns: $f_c = 3, 5, 7$.

1.3. Contribution

For the first time, this paper paves the way to build new testing procedures in the framework of Super-Resolution theory and line spectral estimation. In particular, we prove that we can rightfully construct global null exact testing procedures on the first two *knots* λ_1 and λ_2 of the "continuous" LARS when one has a continuum of predictors, see Theorems 1 and 4 and Fig. 3. These two new procedures offer the ability to test the mean of any stationary Gaussian process with known correlation function Γ and $\mathcal{C}^2$-paths. Furthermore, one of these tests is unbiased, see Theorem 1 and they can be both studentized, see Theorems 2 and 8 and Fig. 6, when variance σ^2 is unknown.

Notations and the formal problem formulation is described in Section 3. In Section 4, we present the test statistic S^{Grid} which is constructed taking the limit of consecutive LARS knots $(\lambda_{1,P}, \lambda_{2,P})$ on thinner and thinner grids (namely the number of predictors P tends to infinity). Section 5 is the theoretical construction of our grid-less test based on consecutive knots (λ_1, λ_2) of the "continuous" LARS. The main result concerning the test statistic S^{Rice} is presented in this section. Applications to spike detection in Super-Resolution are developed in Section 6 and a Github page giving the Python code of the experiments can be found at https://github.com/ydecastro/super-resolution-testing. An appendix with the proofs can be found at the end of the paper.

The general construction of the "continuous" LARS is given in Section 2. This section is independent from the rest of the paper.

2. The "continuous" LARS

2.1. Cameron–Martin type assumption on the mean

The algorithm presented here can be used for a large class of complex processes Z. We consider a complex-valued Gaussian process Z indexed on a compact metric space $\mathbb{K}$ with covariance function K.

Remark 1. Note that this model encompasses our to-be-announced-framework (see Section 3) setting $\mathbb{K} = [0, 2\pi)$ and $K = 2\sigma^2\Gamma$ with Γ the correlation of A_1 defined in Section 3.1. We do not assume that the process is stationary in this section.

We assume that its covariance K is such that there exists $\sigma > 0$ such that

$$\forall s \neq t \in \mathbb{K}, \quad K(t,t) = 2\sigma^2 \text{ and } K(s,t) < 2\sigma^2\,. \tag{1}$$

The scalar 2 accounts for the contribution of the real and the imaginary part of Z and σ^2 is the variance of the real part of Z. We assume that Z has continuous sample paths.

We present here the underlying hypothesis on the mean of the Gaussian processes under consideration when using the LARS algorithm. This hypothesis is of Cameron–Martin type. Indeed, the main drawback that should be avoided is when the mean cannot be represented in the RKHS of the Gaussian process Z. We recall that we can define a reproducing Hilbert space of the covariance K, see [11, Chapter 2.6] for instance. Denote $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ this complex Hilbert space. Also, we can invoke a Karhunen–Loève expansion of the process Z. Namely, there exist i.i.d. complex standard normal variables $(g_j)_{j\geq 1}$, a real orthonormal system $(e_j)_{j\geq 1}$ on $L^2(\mathbb{K})$ and $\sigma_j > 0$ such that

$$Z - \mathbb{E}Z = \sum_j \sigma_j g_j e_j \quad \text{and} \quad \sum_j \sigma_j^2 = 2\sigma^2 < \infty,$$

where the identity holds almost surely in the Banach space of continuous functions on $\mathbb{K}$ equipped with the L^∞-norm. By Mercer's theorem, we know that

$$\forall s,t \in \mathbb{K}, \quad K(s,t) = \sum_j \sigma_j^2\, e_j(s) e_j(t)\,,$$

where the identity holds almost surely in the Banach space of continuous functions on $\mathbb{K}\times\mathbb{K}$ equipped with the L^∞-norm. We recall also that the Hilbert space $\mathcal{H}$ can be defined as

$$\mathcal{H} := \Big\{ \sum_j a_j e_j \;\mid\; \sum_j \frac{|a_j|^2}{\sigma_j^2} < \infty \Big\}$$

with the inner product

$$\Big\langle \sum_j a_j e_j, \sum_j b_j e_j \Big\rangle_{\mathcal{H}} = \sum_j \frac{a_j \overline{b}_j}{\sigma_j^2}\,.$$

We observe Z and we want to estimate its mean $\mathbb{E}\,Z$. Remark that almost surely it holds $Z - \mathbb{E}\,Z \in \overline{\mathcal{H}}$, where $\overline{\mathcal{H}}$ is the closure of $\mathcal{H}$ in the space of continuous functions equipped with the infinity norm, see *e.g.* [11, Corollary 2.6.11]. Remark that $\overline{\mathcal{H}}$ is also closed in $L^2(\mathbb{K})$. Denoting by E the L^2 orthogonal space of $\overline{\mathcal{H}}$, one has $L^2(\mathbb{K}) = \overline{\mathcal{H}} \oplus E$ where the sum is orthogonal. We denote by $\mathcal{P}$ (resp. $\mathcal{P}^\perp$) the orthogonal projection onto $\overline{\mathcal{H}}$ (resp. E). Since almost surely $Z - \mathbb{E}\,Z \in \overline{\mathcal{H}}$, remark that almost surely $\mathcal{P}^\perp(Z) = \mathcal{P}^\perp(\mathbb{E}Z)$ and

this process can be observed and is deterministic. Without loss of generality, we assume that $\mathcal{P}^{\perp}(\mathbb{E}Z) = 0$ subtracting $\mathcal{P}^{\perp}(Z)$ to Z. Also, we assume that

$$\mathcal{P}(\mathbb{E}Z) \in \mathcal{H}\,. \tag{2}$$

Recall that $\mathcal{P}(\mathbb{E}Z) = \mathbb{E}Z$ and Assumption (2) gives that $Z \in \overline{\mathcal{H}}$ using $Z - \mathbb{E}\,Z \in \overline{\mathcal{H}}$.

2.2. Description of the "continuous" LARS

We assume that $Z \in \mathcal{H}$ and, as mentioned above, this assumption is equivalent to Assumption (2). Following standard references, *e.g.,* [1, Chapter 5.6], the Least-Angle Regression Selection (LARS) algorithm can be extended to Gaussian processes. To the best of our knowledge, the LARS for complex Gaussian processes has never been introduced and we present its formulation here for the first time. Actually, the presentation given in this section can be applied to any RKHS setting. It results in a description of the LARS in infinite dimensional feature spaces and this framework has been dealt in [12]. However, note that the paper [12] only concerns real signed measures and their "doubling" dimension trick [12, page 546] cannot be used when dealing with complex measures. In particular, their result cannot be invoked in Super-Resolution where it is of utmost importance to deal with complex measures. This section presents the "continuous" LARS for Super-Resolution.

The LARS is a variable selection algorithm giving a sequence $((\lambda_k, \mu_k))_{k\geq 1}$ where the knots are ordered such that $\lambda_1 \geq \lambda_2 \geq \ldots > 0$ and $\mu_k \in (\mathcal{M}(\mathbb{K}, \mathbb{C}), \|\cdot\|_1)$ is a complex-valued measure. We recall that the space $(\mathcal{M}(\mathbb{K}, \mathbb{C}), \|\cdot\|_1)$ is defined as the dual space of the space of continuous functions on $\mathbb{K}$ equipped with the L^{∞}-norm. A pseudo-code is presented in Algorithm 1 and the technical details are presented below. When defining the "continuous LARS", we assume that

$$K \text{ is (at least) four times differentiable.} \tag{$\boldsymbol{A}_K$}$$

Under this assumption, the process Z is twice differentiable in quadratic mean and once differentiable almost surely.

2.2.1. The first knot

Inspired by the Super-Resolution framework—presented in Section 6, we consider Z as some "correlation process" in the spirit of (18). In particular, the most correlated point can be defined by (9), namely

$$\lambda_1 := \max_{t\in\mathbb{K}} |Z(t)|\,.$$

Under Assumption (1), Proposition 16 shows that almost surely there exists a unique point t_1 such that $\lambda_1 = |Z(t_1)|$. Define the "active set" function $\lambda \mapsto \mathcal{A}(\lambda)$ as

$$\mathcal{A}(\lambda_1) = \mathcal{A}_1 := (t_1)\,,$$

and $\mathcal{A}(\lambda) = \emptyset$ for $\lambda > \lambda_1$. The path $\lambda \mapsto \mathcal{A}(\lambda)$ for $\lambda \leq \lambda_1$ will be defined in the sequel. It is a piecewise continuously differentiable path representing the support of a discrete measure $\mu^{(\lambda)}$ such that

$$\|Z - \int_{\mathbb{K}} K(u,\cdot)\mathrm{d}\mu^{(\lambda)}(u)\|_{\infty} \leq \lambda\,,$$

namely the residual has L^{∞}-norm less than λ. Set the first fitted solution to $\mu_1 = 0$ and the first residual to $Z_1 = Z$ for initialization purposes. Observe that

Algorithm 1: Continous LARS.

Data: A correlation process Z indexed by $\mathbb{K}$ and its variance-covariance function K.
Result: A sequence $((\lambda_k, \mu_k))_{k\geq 1}$ where the knots are ordered such that $\lambda_1 \geq \lambda_2 \geq \ldots > 0$ and $\mu_k \in (\mathcal{M}(\mathbb{K}, \mathbb{C}), \|\cdot\|_1)$ is a complex-valued measure.

`/* We initialize this Forward procedure computing` λ_1 `and` μ_1 `*/`

1 Set $k = 1$, $\lambda_1 := \max_{\mathbb{K}} |Z|$ and $\mu_1 = 0$.

`/* We use an "active set"` $\mathcal{A}_k$ `giving the support of the next solution` μ_{k+1} `*/`

2 Set $t_1 := \arg\max_{\mathbb{K}} |Z|$ and $\mathcal{A}_1 = (t_1)$.

`/* We use a "residual"` Z_k `initialized with */`

3 Set $Z_1 := Z$.

`/* We iterate the next commands until a stopping criterion is met */`

4 Set $k \leftarrow k+1$ `/*` $\mathcal{A}_{k-1} =: (t_1, \ldots, t_{k-1})$ `and` λ_{k-1} `has been defined at the previous step. */`

5 For $\lambda > 0$ and $x = (x_1, \ldots, x_{k-1}) \in \mathbb{R}^{k-1}$ define

$$a(\lambda, x) := (K(x_i, x_j))^{-1} \begin{pmatrix} Z(x_1) - (\lambda/\lambda_{k-1}) Z_{k-1}(t_1) \\ \vdots \\ Z(x_{k-1}) - (\lambda/\lambda_{k-1}) Z_{k-1}(t_{k-1}) \end{pmatrix}$$

$$h_j(\lambda, x) := \frac{\partial}{\partial t}\Big[|Z(t) - \sum_{i=1}^{k-1} a_i(\lambda, x) K(x_i, t)|^2\Big](x_j)$$

and solve $(h_1(\lambda, x), \ldots, h_{k-1}(\lambda, x)) = 0$ starting from $(\lambda, x) = (\lambda_{k-1}, \mathcal{A}_{k-1})$ for $0 < \lambda \leq \lambda_{k-1}$. The solution path is denoted by $x(\lambda) := (t_1(\lambda), \ldots, t_{k-1}(\lambda))$.

6 Set $Z^{(\lambda)}(\cdot) := Z(\cdot) - \sum_{i=1}^{k-1} a_i(\lambda, x(\lambda)) K(t_i(\lambda), \cdot)$ and pick

$$\lambda_k := \max\{\beta > 0\,;\ \exists\, t \notin x(\beta),\ \text{s.t. } |Z^{(\beta)}|(t) = \beta\} \quad \text{and} \quad t_k := \arg\max_{s \notin x(\lambda_k)} |Z^{(\lambda_k)}|(s)\,.$$

7 Set $\mathcal{A}_k = (t_1(\lambda_k), \ldots, t_{k-1}(\lambda_k), t_k)$ and

$$\mu_k = \sum_{i=1}^{k-1} a_i(\lambda_k, x(\lambda_k)) \delta_{t_i(\lambda_k)} \text{ and } Z_k(\cdot) = Z^{(\lambda_k)}(\cdot) = Z(\cdot) - \sum_{i=1}^{k-1} a_i(\lambda_k, x(\lambda_k)) K(t_i(\lambda_k), \cdot)\,.$$

8 Iterate from **4**.

$$\begin{aligned} Z_1(t_1) &= \lambda_1\, e^{\imath\theta_1}, \\ |Z_1(t_1)| &= \lambda_1\,, \\ \forall t \in \mathbb{K}, \quad Z_1(t) &= Z(t) - \int_{\mathbb{K}} K(u,t)\mu_1(\mathrm{d}u)\,, \\ \forall t \neq t_1, \quad |Z_1(t)| &< \lambda_1\,. \end{aligned} \tag{3}$$

2.2.2. The second knot

We want to add an other point t_2 to the active set and define a discrete measure μ_2 supported on $\mathcal{A}_1$ while keeping the above inequalities true. First, we solve the least-squares fit given by

$$a = \arg\min_{c\in\mathbb{C}} \big\|Z_1(\cdot) - cK(t_1, \cdot)\big\|_{\mathcal{H}}^2\,.$$

This program can be solved in closed form and it holds that $a = Z(t_1)/(2\sigma^2)$. Then, for any $0 < \lambda \leq \lambda_1$, define $Z^{(\lambda)}$ by

$$Z^{(\lambda)}(t) = Z(t) + (\frac{\lambda}{\lambda_1} - 1) Z(t_1) \frac{K(t_1, t)}{2\sigma^2}\,,$$

and observe that $|Z^{(\lambda)}(t_1)| = \lambda$. Remark also that $|Z^{(\lambda)}|$ has a local maxima at point $t = t_1$. Indeed, under ($\boldsymbol{A}_K$), the function $X : (t, \theta) \mapsto \mathrm{Re}(e^{-\imath\theta} Z(t))$ is continuously differentiable and it has $t = t_1$ as global maximum by definition of t_1 and λ_1. Il follows from (7) that $\widehat{z} := (t_1, \theta_1)$ is a local maxima of X and therefore t_1 is a local maxima of $|Z^{(\lambda)}|$.

Now, we keep track of the largest value of the "correlation" process $|Z^{(\lambda)}|$ on the complementary set of $\mathcal{A}_1$ while moving λ from λ_1 toward zero. We define λ_2 as the largest value for which there exists a point $t \notin \mathcal{A}_1$ such that $|Z^{(\lambda)}|(t) = \lambda$. Set

$$\begin{aligned} \lambda_2 &:= \max\left\{\beta > 0\,;\ \exists\, t \notin \mathcal{A}_1,\ \text{s.t. } |Z^{(\beta)}|(t) = \beta\right\}, \\ \text{and} \quad t_2 &:= \arg\max_{s \notin \mathcal{A}_1} |Z^{(\lambda_2)}|(s)\,. \end{aligned} \tag{4}$$

If t_2 is not unique, we add all the solutions of (4) to the active set $\mathcal{A}_2$. For sake of readability, we assume that t_2 is the only solution to (4). Then, update

$$\begin{aligned} \mathcal{A}(\lambda) &= \mathcal{A}_1 \quad \lambda_2 < \lambda \leq \lambda_1\,, \\ \mathcal{A}(\lambda_2) &= \mathcal{A}_2 := (t_1, t_2)\,, \\ \mu_2 &= (1 - \lambda_2/\lambda_1) a \delta_{t_1}\,, \\ Z_2(\cdot) &= Z^{(\lambda_2)}(\cdot) = Z_1(\cdot) + (\lambda_2/\lambda_1 - 1) a K(t_1, \cdot)\,, \end{aligned}$$

where, for all $t \in \mathbb{K}$,

$$\begin{aligned} Z_2(t) &= Z_1(t) + (\lambda_1/\lambda_2 - 1) a K(t_1, t) \\ &= Z(t) - \int_{\mathbb{K}} K(u, t) \mu_2(\mathrm{d}u)\,, \end{aligned}$$

is the second residual associated to the second fitted solution μ_2. Remark also that

$$\begin{aligned} \forall t \in \{t_1, t_2\}, \quad |Z_2(t)| &= \lambda_2\,, \\ \forall t \neq \{t_1, t_2\}, \quad |Z_2(t)| &< \lambda_2\,. \end{aligned}$$

2.2.3. The other knots: moving the active set between knots

From this point we proceed iteratively. For $k \geq 3$, we assume that we have found $(\lambda_{k-1}, \mu_{k-1})$ and $\mathcal{A}_{k-1} = (t_1, \ldots, t_{k-1})$ such that

$$\begin{aligned} \mathcal{A}(\lambda_{k-1}) &= \mathcal{A}_{k-1}\,, \\ \forall t \in \mathbb{K}, \quad Z_{k-1}(t) &:= Z(t) - \int_{\mathbb{K}} K(u, t) \mu_{k-1}(\mathrm{d}u)\,, \\ \forall t \in \mathcal{A}_{k-1}, \quad |Z_{k-1}(t)| &= \lambda_{k-1}\,, \\ \forall t \notin \mathcal{A}_{k-1}, \quad |Z_{k-1}(t)| &< \lambda_{k-1}\,. \end{aligned}$$

We want to define the path $\lambda \mapsto \mathcal{A}(\lambda)$ for values $\lambda \leq \lambda_{k-1}$ starting from $\mathcal{A}(\lambda_{k-1}) = \mathcal{A}_{k-1}$. We look for a path $\mathcal{A}(\lambda) = (t_1(\lambda), \ldots, t_{k-1}(\lambda))$ such that $t_i(\lambda)$ are continuously differentiable and there exists $\mu^{(\lambda)}$ supported on $\mathcal{A}(\lambda)$ such that the above inequalities hold true. This path will be defined on $(\lambda_k, \lambda_{k-1}]$ for a value λ_k defined later.

◦ Consider $0 < \lambda \leq \lambda_{k-1}$ and define

$$a(\lambda) = M_{k-1}(\lambda)^{-1} \begin{pmatrix} Z(t_1(\lambda)) - (\lambda/\lambda_{k-1}) Z_{k-1}(t_1(\lambda_{k-1})) \\ \vdots \\ Z(t_{k-1}(\lambda)) - (\lambda/\lambda_{k-1}) Z_{k-1}(t_{k-1}(\lambda_{k-1})) \end{pmatrix}$$

where we denote $M_{k-1}(\lambda) = (K(t_i(\lambda), t_j(\lambda)))_{1 \leq i,j \leq k-1}$ and we assume that $M_{k-1}(\lambda)$ is invertible. If $M_{k-1}(\lambda)$ is not invertible then we stop. The path $\mathcal{A}(\lambda) = (t_1(\lambda), \dots, t_{k-1}(\lambda))$ will be defined later on. Note that $\mathcal{A}(\lambda_{k-1}) = (t_1(\lambda_{k-1}), \dots, t_{k-1}(\lambda_{k-1})) = \mathcal{A}_{k-1}$ for $\lambda = \lambda_{k-1}$.

Remark 2. Note that the function $\sum_{i=1}^{k-1} a_i(\lambda_{k-1}) K(t_i(\lambda_{k-1}), \cdot)$ is the regression of Z onto the finite dimensional space $\mathrm{Span}\{K(t_i(\lambda_{k-1}), \cdot)\,;\ i = 1, \dots, k-1\}$.

◦ Then, for any $0 < \lambda \leq \lambda_{k-1}$, define

$$\mu^{(\lambda)} := \sum_{i=1}^{k-1} a_i(\lambda) \delta_{t_i(\lambda)}\,,$$

$$Z^{(\lambda)}(\cdot) := Z(\cdot) - \sum_{i=1}^{k-1} a_i(\lambda) K(t_i(\lambda), \cdot)\,,$$

and observe that $|Z^{(\lambda)}(t)| = \lambda$ for all $t \in \{t_1(\lambda), \dots, t_{k-1}(\lambda)\}$. Indeed, it holds

$$\begin{aligned} Z^{(\lambda)}(t_j(\lambda)) &= Z(t_j(\lambda)) - \sum_{i=1}^{k-1} a_i(\lambda) K(t_i(\lambda), t_j(\lambda))\,, \\ &= Z(t_j(\lambda)) - a(\lambda)^\top (M_{k-1}(\lambda))(0, \dots, 0, \underbrace{1}_{j\text{th}}, 0, \dots, 0)\,, \\ &= Z(t_j(\lambda)) - Z(t_j(\lambda)) + (\lambda/\lambda_{k-1}) Z_{k-1}(t_j(\lambda_{k-1}))\,, \\ &= \lambda Z_{k-1}(t_j(\lambda_{k-1}))/\lambda_{k-1}\,, \end{aligned}$$

and recall that it holds $|Z_{k-1}(t_j(\lambda_{k-1}))| = \lambda_{k-1}$.

We will enforce that $t_j(\lambda)$ is a local maximum of $|Z^{(\lambda)}|$ imposing that its derivative is zero along the path $\mathcal{A}(\lambda)$ for $\lambda_k < \lambda \leq \lambda_{k-1}$. This can be done invoking the implicit function theorem as follows. Define for $\lambda > 0$ and $x = (x_1, \dots, x_{k-1}) \in \mathbb{R}^{k-1}$

$$F(\lambda, x) := (h_1(\lambda, x), \dots, h_{k-1}(\lambda, x))$$

where

$$a(\lambda, x) := (K(x_i, x_j))^{-1} \begin{pmatrix} Z(x_1) - (\lambda/\lambda_{k-1}) Z_{k-1}(t_1(\lambda_{k-1})) \\ \vdots \\ Z(x_{k-1}) - (\lambda/\lambda_{k-1}) Z_{k-1}(t_{k-1}(\lambda_{k-1})) \end{pmatrix}$$

$$h_j(\lambda, x) := \frac{\partial}{\partial t} \Big[\big| Z(t) - \sum_{i=1}^{k-1} a_i(\lambda, x) K(x_i, t) \big|^2 \Big](x_j) \tag{5}$$

Assume that the jacobian $\frac{\partial F}{\partial x}$ is invertible. If $\frac{\partial F}{\partial x}$ is not invertible then we stop. Therefore, the implicit function theorem implies that there exists a continuously differentiable path $x(\lambda) := (t_1(\lambda), \dots, t_{k_1}(\lambda))$ such

that $F(\lambda, x) = 0$ is equivalent to $x = (t_1(\lambda), \dots, t_{k_1}(\lambda))$ on a neighborhood of $\lambda = \lambda_{k-1}$. On this path, the derivative of $t \mapsto |Z^{(\lambda)}|^2(t)$ at points $t = t_j(\lambda)$ is zero (thanks to (5)) while $|Z^{(\lambda)}|(t_j(\lambda)) = \lambda$. We deduce that there exists a neighborhood of λ_{k-1} on which each point $t_j(\lambda)$ is a local maximum of $|Z^{(\lambda)}|$.

Now, we keep track of the largest value of the "correlation" process $|Z^{(\lambda)}|$ on the complementary set of $\mathcal{A}(\lambda)$ while moving λ from λ_{k-1} toward zero. We define λ_k as the largest value for which there exists a point $t \notin \mathcal{A}(\lambda)$ such that $|Z^{(\lambda)}|(t) = \lambda$. Set

$$\begin{aligned} \lambda_k &:= \max\left\{\beta > 0\,;\ \exists\, t \notin \mathcal{A}(\beta),\ \text{s.t. } |Z^{(\beta)}|(t) = \beta\right\}, \\ \text{and} \quad t_k &:= \arg\max_{s \notin \{t_1(\lambda_k), \dots, t_{k-1}(\lambda_k)\}} |Z^{(\lambda_k)}|(s)\,. \end{aligned} \tag{6}$$

If t_k is not unique, we add all the solutions of (6) to the active set $\mathcal{A}_k$. For sake of readability, we assume that t_k is the only solution to (6).

- Update

$$\begin{aligned} \mathcal{A}(\lambda) &= (t_1(\lambda_k), \dots, t_{k-1}(\lambda_k)) \quad \lambda_k < \lambda \leq \lambda_{k-1}\,, \\ \mathcal{A}(\lambda_k) &= \mathcal{A}_k := (t_1(\lambda_k), \dots, t_{k-1}(\lambda_k), t_k)\,, \\ \mu_k &= \mu^{(\lambda_k)} = \sum_{i=1}^{k-1} a_i(\lambda_k)\delta_{t_i(\lambda_k)}\,, \\ Z_k(\cdot) &= Z^{(\lambda_k)}(\cdot) = Z(\cdot) - \sum_{i=1}^{k-1} a_i(\lambda_k) K(t_i(\lambda_k), \cdot)\,, \end{aligned}$$

where, for all $t \in \mathbb{K}$,

$$Z_k(t) = Z(t) - \int_{\mathbb{K}} K(u, t)\mathrm{d}\mu_k(u)\,,$$

is the kth residual associated to the kth fitted solution μ_k. Remark also that

$$\begin{aligned} \forall t \in \{t_1, \dots, t_k\}, \quad & |Z_k(t)| = \lambda_k\,, \\ \forall t \neq \{t_1, \dots, t_k\}, \quad & |Z_k(t)| < \lambda_k\,, \end{aligned}$$

and update k to $k+1$ to iterate the procedure.

2.3. Equivalent expression of the second knot

First, observe that λ_1 is defined as in (9) and that the two definitions agree. Indeed, recall that $X(t, \theta) = \mathrm{Re}\,(e^{-\imath\theta} Z(t))$ so that $\max X = \max |Z|$ at point $\widehat{z} = (t_1, \theta_1)$ with t_1 as in (3). By optimality, it holds that $\lambda_1 = e^{-\imath\theta_1} Z(t_1)$.

Then, the case $k = 2$ is interesting since λ_2 is a statistic used in the test statistics described in the sequel. We will see that the two definitions agree here again, please refer to Section 3 for notations. For $k = 2$, it holds $Z_1 = Z$ and the least squares direction a is given by $a = Z(t_1)/(2\sigma^2)$ and $Z^{(\lambda)}$ by

$$\begin{aligned} Z^{(\lambda)}(t) &= Z(t) + (\frac{\lambda}{\lambda_1} - 1) Z(t_1) \frac{K(t_1, t)}{2\sigma^2}\,, \\ &= Z(t) + e^{\imath\theta_1}(\lambda - \lambda_1)\frac{K(t_1, t)}{2\sigma^2} \end{aligned}$$

Multiplying by $e^{-\imath\theta}$ and taking the real part, this latter can be equivalently written as

$$\mathrm{Re}\,(e^{-\imath\theta}Z^{(\lambda)}(t)) = X(z) + (\lambda-\lambda_1)\cos(\theta_1-\theta)\frac{K(t_1,t)}{2\sigma^2}\,,$$

where $z=(t,\theta)\in\mathbb{T}$. Now, recall that $\rho(t,\theta) := \Gamma(t)\cos\theta = \cos(\theta)K(0,t)/(2\sigma^2)$ to compute

$$\mathrm{Re}\,(e^{-\imath\theta}Z^{(\lambda)}(t)) = X(z) + (\lambda-\lambda_1)\rho(z-\widehat{z})\,. \tag{7}$$

We deduce that

$$\begin{aligned}\mathrm{Re}\,(e^{-\imath\theta}Z^{(\lambda)}(t)) \le \lambda &\Leftrightarrow X(z) - \lambda_1\rho(z-\widehat{z}) \le \lambda(1-\rho(z-\widehat{z}))\\ &\Leftrightarrow \frac{X(z)-X(\widehat{z})\rho(z-\widehat{z})}{1-\rho(z-\widehat{z})} \le \lambda\\ &\Leftrightarrow X^{\widehat{z}}(z) \le \lambda\end{aligned}$$

showing that the second knot λ_2 is exactly the quantity defined in (10).

2.4. Illustration: the two first knots of super-resolution

The Super-Resolution process is defined in (18). It satisfies Condition ($\mathrm{KL}_Z(N)$) and Condition ($\mathrm{ND}_Z(N)$) of Section 5.2.1 with $N = 2f_c+1$. The first point is given by the maximum of the modulus of Z, see the red curve in Fig. 1. Observe that $Z_1 = Z$ and the maximum satisfies $Z_1(t_1) = \lambda_1\, e^{\imath\theta_1}$. Then, we compute

$$Z^{(\lambda)}(t) = Z_1(t) + (\frac{\lambda}{\lambda_1}-1)Z_1(t_1)\frac{\mathbf{D}_N(t_1-t)}{2N\sigma^2}\,,$$

where $\mathbf{D}_N$ denotes the Dirichlet kernel. For $\lambda>\lambda_2$, the maximum of $|Z^{(\lambda)}|$ is achieved at a unique point, namely t_1. For $\lambda=\lambda_2$, a second point achieves the maximum. This transition defines $Z_2 := Z^{(\lambda_2)}$, see Fig. 1.

From this point, we can iterate fitting the least squares direction on the support $\{t_1,t_2\}$ and decreasing $|Z_2|$ while a third point achieves the maximum. Given the red curve in Fig. 1, it was not obvious that the second knot would have been t_2 since other local maxima seemed more significant than t_2 on the red curve.

3. Notations and problem formulation

3.1. Hypothesis testing problem

In this paper, our purpose is to test the mean value of a stationary complex-valued Gaussian process Z with $\mathcal{C}^2$-paths indexed by $[0,2\pi)$. We assume that $Z = A_1 + \imath A_2$ where A_1 and A_2 are two independent and identically distributed real-valued processes with $\mathcal{C}^2$-paths. Assume that the correlation function Γ of A_1 (and A_2) satisfies

$$\forall t\in(0,2\pi),\quad |\Gamma(t)|<1 \tag{$\mathbf{A}_{\mathrm{norm}}$}$$

and let $\sigma^2 := \mathbb{V}\mathrm{ar}(A_1(\cdot))$ so that

$$\mathbb{C}\mathrm{ov}(A_1(s),A_1(t)) = \sigma^2\Gamma(t-s)\,. \tag{8}$$

We denote by $\mathbb{T} := [0, 2\pi)^2$ the 2-dimensional torus. Assume that we observe a real-valued process $(X(z))_{z\in\mathbb{T}}$ indexed by $\mathbb{T}$ such that

$$\forall z \in \mathbb{T}, \quad X(z) := A_1(t)\cos\theta + A_2(t)\sin\theta = \mathrm{Re}\big(e^{-\imath\theta} Z(t)\big)\,,$$

where $z = (t, \theta)$ and $\mathrm{Re}(\cdot)$ denotes the real part of a complex number. Remark that observing X is equivalent to observe Z since we can recover Z from X and conversely. Furthermore, we may assume that the process $(X(z))_{z\in\mathbb{T}}$ satisfies

$$\text{a.s. there is no point } z \in \mathbb{T} \text{ s.t. } X'(z) = 0 \text{ and } \det(X''(z)) = 0, \tag{$\mathbf{A}_{\mathrm{degen}}$}$$

where $X'(z)$ and $X''(z)$ denote the gradient and the Hessian of X at point z. Note that sufficient conditions for ($\mathbf{A}_{\mathrm{degen}}$) are given by [13, Proposition 6.5] applied to $(X(z))_{z\in\mathbb{T}}$. In particular if the distribution of $X''(t)$ is non degenerated, using [13, Condition (b) of Proposition 6.5], it implies that Assumption ($\mathbf{A}_{\mathrm{degen}}$) is met. Note also that Assumption ($\mathbf{A}_{\mathrm{degen}}$) is referred to as "Morse" process in [14]. Remark that ($\mathbf{A}_{\mathrm{norm}}$) and ($\mathbf{A}_{\mathrm{degen}}$) are mild assumptions ensuring that Z is a non-pathological process with $\mathcal{C}^2$-paths.

This paper aims at testing the following hypotheses:

$$\mathbb{H}_0 : \text{“}Z \text{ is centered”} \quad \text{against} \quad \mathbb{H}_1 : \text{“}Z \text{ is not centered”}\,.$$

Remark that this framework encompasses any testing problem whose null hypothesis is a single hypothesis on the mean of Z, subtracting the mean tested by the null hypothesis. Indeed, remark that Z can always be decomposed into

$$Z = Z^0 + \eta\,,$$

where $Z^0 = \mathbb{E}Z$ is the deterministic noiseless response and η is some centered random additive perturbation of Z^0. Given any function f^0, one might be interested in testing whether $Z^0 = f^0$ or equivalently $Z - f^0$ is centered. Not rejecting this hypothesis means that there is no evidence that the residual $Z - f^0$ is not centered. On the other hand, rejecting the null means that the testing procedure have found some evidence that one should not consider that the residual $Z - f^0$ is centered. Now, the same discussion can be made for X remarking that

$$X(z) = \mathrm{Re}\big(e^{-\imath\theta} Z(t)\big) = X^0(z) + N(z)$$

where we denote by $X^0(z) := \mathrm{Re}\big(e^{-\imath\theta} Z^0(t)\big)$ the deterministic noiseless response part and by $N(z) := \mathrm{Re}\big(e^{-\imath\theta}\eta(t)\big)$ some centered random additive perturbation of X^0.

3.2. The first and second knots of a Gaussian process

As in high-dimensional statistics, we can define the first and second knots (λ_1, λ_2) as follows. If we model some spatial correlation by means of the process X, the most correlated point $\widehat{z} \in \mathbb{T}$ and the maximal correlation λ_1 are respectively the argument maximum and the maximum of X defined by

$$\widehat{z} := \arg\max_{z\in\mathbb{T}} X(z) \quad \text{and} \quad \lambda_1 := X(\widehat{z})\,. \tag{9}$$

Under Assumption ($\mathbf{A}_{\mathrm{norm}}$), one can check that the argument maximum is almost surely a singleton, see Proposition 16.

To construct the second knot, given a fixed $z \in \mathbb{T}$, one can equivalently consider two regressions of $X(y)$, as follows.

• On the one hand, the regression on $X(z)$ that will appear in the grid method of Section 4. Using a convenient normalisation related to the definition of the LARS knots, we set

$$\forall y \in \mathbb{T} \setminus \{z\}, \quad X^z(y) := \frac{X(y) - X(z)\rho(z-y)}{1-\rho(z-y)} = X(z) + \frac{X(y)-X(z)}{1-\rho(z-y)},$$

where

$$\forall z \in \mathbb{T}, \quad \rho(z) := \Gamma(t)\cos\theta,$$

is the correlation function of the stationary Gaussian process X. One can check that X^z is a Gaussian process indexed by $\mathbb{T} \setminus \{z\}$ and independent of $X(z)$.

• On the other hand, the regression on $(X(z), X'(z))$ will be needed for convergence purposes in Section 5. With the convenient normalization, we set

$$\forall y \in \mathbb{T} \setminus \{z\}, \quad X^{|z}(y) := \frac{X(y) - \rho(z-y)X(z) + \langle \rho'(z-y), \widetilde{\Lambda}^{-1}X'(z)\rangle}{1-\rho(z-y)},$$

where ρ' is the gradient of the correlation function ρ and $\widetilde{\Lambda} := -\rho''(0)$ is the variance-covariance matrix of the derivative process of X, namely X'.

Since the derivative at $\widehat{z}$ is zero, note that $X^{\widehat{z}}(\cdot) = X^{|\widehat{z}}(\cdot)$ and we define the second knot λ_2 as

$$\widehat{y} := \arg\max_{y \in \mathbb{T}\setminus\{\widehat{z}\}} X^{\widehat{z}}(y) \quad \text{and} \quad \lambda_2 := X^{\widehat{z}}(\widehat{y}) = X^{|\widehat{z}}(\widehat{y}), \tag{10}$$

where we prove that $(\widehat{y}, \lambda_2)$ are well defined and that $\widehat{y}$ is almost surely unique, see Proposition 16 and Remark 8. Furthermore, the couple $(\widehat{y}, \lambda_2)$ can be equivalently defined using the extension of the LARS to our framework, the interested reader may consult Section 2.3.

4. Passing to the limit, the grid approach

The main idea of this section is to define a sequence of grids $(G_n)_{n\geq 1}$ on $\mathbb{T}$, to construct a sequence of test statistics $(S_n)_{n\geq 1}$ from the values of the process X on G_n as in [10] and to pass to the limit as $n \to \infty$. More precisely, we consider G_n to be the grid with mesh $\Delta_n := (2\pi)2^{-n}$ on $\mathbb{T}$ (corresponding to $P = 2^{2n}$ grid points so that $n = (\log_2 P)/2$),

$$\widehat{z}_n := \arg\max_{z \in G_n} X(z) \quad \text{and} \quad \lambda_{1,n} := \max_{z \in G_n} X(z).$$

It is the maximum of the process X when indexing by the grid. We can also define the maximum of the regression when indexing by the grid, namely

$$\lambda_{2,n} := \max_{y \in G_n \setminus \{\widehat{z}_n\}} X^{\widehat{z}_n}(y).$$

The Hessian at the maximum (9) on $\mathbb{T}$ is denoted by $X'' := X''(\widehat{z})$ (in particular it does not depend on the grid but on the maximum $\widehat{z}$ of X). By Assumption ($\mathbf{A}_{\text{degen}}$), it is a random variable with values in the set of non degenerated negative definite matrices of size 2×2. We can define a non degenerated positive quadratic form (*i.e.,* a metric) on $\mathbb{R}^2$ by $\|v\|_{X''} = -v^\top X'' v$, for $v \in \mathbb{R}^2$. Using this metric, we can consider the corresponding Voronoi tessellation of $\mathbb{Z}^2$. It is a regular partition of $\mathbb{R}^2$ by parallelograms, invariant by

translations $(1,0)$ and $(0,1)$. Denote by $V_o \subset [-1,1]^2$ the Voronoi cell of the origin in this partition and by $\mathcal{U} := \mathcal{U}(V_o)$ the uniform distribution on this cell. We understand the law $\mathcal{U}$ as a conditional law with respect to X'' and, conditionally to X'', this law is taken independent of (λ_1, λ_2), see Lemma 12. Conditionally to X'', define the randomized statistics

$$\overline{\lambda}_2 := \lambda_2 \vee \left\{ \lambda_1 + \sup_{k \in \mathbb{Z}^2 \setminus \{0\}} \frac{k^\top}{\|\widetilde{\Lambda}^{\frac{1}{2}} k\|} X''\Big(\frac{k - 2\mathcal{U}}{\|\widetilde{\Lambda}^{\frac{1}{2}} k\|} \Big) \right\}, \tag{11}$$

where $\widetilde{\Lambda}^{\frac{1}{2}}$ is the square root of $\widetilde{\Lambda} = -\rho''(0)$ and $a \vee b = \max(a,b)$. A proof of the following result is given in Appendix A.1.

Remark 3. Remark that we have taken dyadic grids here. Following the proof in Appendix A.1, one can exhibit how $\overline{\lambda}_2$ depend on the sequence of grids. The key result is Lemma 12 and we borrow its notation in this remark. In the general case where one consider a different type of sequence of grids, one still have independence between $(\widehat{z} - \overline{z}_n)$ and (λ_1, λ_2) but the law of the limit of $\Delta_n^{-1}(\widehat{z} - \overline{z}_n)$ (for some Δ_n that may depend on the grid sequence) may differ from $\mathcal{U}$. We refer to this law (if it exists) as $\mathcal{V}_k$ where $k \in \mathbb{Z}^2$. The dependence in k depicts the fact cells defined by joining adjacent points of the grid might be topologically different (which is not the case in the dyadic case). It results that the definition of $\overline{\lambda}_2$ should be modified changing $\mathcal{U}$ by $\mathcal{V}_k$. It does not change the main message here: the resulting test is randomized and (12) is non-conservative and should be avoided in practice.

Theorem 1. *Under $\mathbb{H}_0$, Assumptions ($\mathbf{A}_{\mathrm{norm}}$) and ($\mathbf{A}_{\mathrm{degen}}$), the randomized test statistics*

$$S^{\mathrm{Grid}} := \frac{\overline{\Phi}(\lambda_1/\sigma)}{\overline{\Phi}(\overline{\lambda}_2/\sigma)} \sim \mathcal{U}([0,1]),$$

where $\overline{\Phi}$ denotes the standard Gaussian survival function. Moreover, the test with p-value S^{Grid} is unbiased: under the alternative $\mathbb{H}_1$, it holds $\mathbb{P}\{S^{\mathrm{Grid}} \leqslant \alpha\} \geqslant \alpha$ for all $\alpha \in (0,1)$.

Theorem 1 shows in particular that the statistics—referred to as the Spacing test statistics in the introduction—given by

$$S^{\mathrm{ST}} = \frac{\overline{\Phi}(\lambda_1/\sigma)}{\overline{\Phi}(\lambda_2/\sigma)} \tag{12}$$

does not follow a $\mathcal{U}([0,1])$ distribution under $\mathbb{H}_0$ ans leads to a non-conservative test. Indeed, observe that almost surely $\lambda_2 \leq \overline{\lambda}_2$ so that $S^{\mathrm{ST}} \geq S^{\mathrm{Grid}}$ almost surely. Note that the two test statistics differ on the event $\{\lambda_2 \neq \overline{\lambda}_2\} = \{\lambda_2 < \overline{\lambda}_2\}$.

Now, when the variance σ^2 is unknown, we can build an estimator $\widehat{\sigma}^2$ defined in (16) and obtain a studentized version of the previous theorem. Please consult Section 5.2.1 for further details on the construction of the estimator $\widehat{\sigma}$ and on Conditions ($\mathrm{KL}_Z(N)$) and ($\mathrm{ND}_Z(N)$).

Theorem 2. *Assume ($\mathbf{A}_{\mathrm{norm}}$), ($\mathbf{A}_{\mathrm{degen}}$), ($\mathrm{KL}_Z(N)$) and ($\mathrm{ND}_Z(N)$) where $2 \leqslant N < \infty$, then the following test statistics T^{Grid} satisfies*

$$T^{\mathrm{Grid}} := \frac{\overline{F}_{m-1}(\lambda_1/\widehat{\sigma})}{\overline{F}_{m-1}(\overline{\lambda}_2/\widehat{\sigma})} \sim \mathcal{U}([0,1])$$

under $\mathbb{H}_0$ where $m = 2N$, F_{m-1} is the Student cumulative distribution function with $m-1$ degrees of freedom, $\overline{F}_{m-1} = 1 - F_{m-1}$ its survival function and $\widehat{\sigma}^2$ is defined by (16).

A proof can be found in Appendix A.2.

Remark 4. Only the first point of ($\mathrm{ND}_Z(N)$) is required for the proof. Moreover, if $m = +\infty$, the Student distribution is to be replaced by a standard normal distribution.

5. The Rice method: a grid-less approach

In this section, we build our test statistic directly on the entire path of the process X in a grid-less manner. We assume that the process X satisfies Assumptions ($\mathbf{A}_{\mathrm{norm}}$) and ($\mathbf{A}_{\mathrm{degen}}$), and is centered (namely $\mathbb{H}_0$). As in the preceding section, we consider λ_1 and λ_2 defined by (9) and (10) respectively.

We denote $X = \sigma\widetilde{X}$ so that the covariance function of $\widetilde{X}$ is the correlation function ρ of X, namely $\widetilde{X}$ is the standardized version of X. Note that, by regression formulas and stationarity, it holds

$$\forall z \in \mathbb{T}, \quad \mathbb{E}\left[\widetilde{X}''(z)\middle|(\widetilde{X}(z), \widetilde{X}'(z))\right] = -\widetilde{\Lambda}\widetilde{X}(z)\,,$$

so that we can define the process $\widetilde{R}$ by the decomposition

$$\widetilde{X}''(z) = -\widetilde{\Lambda}\widetilde{X}(z) + \widetilde{R}(z)$$

where $\widetilde{R}(z)$ and $\widetilde{X}(z)$ are independent for any $z \in \mathbb{T}$ and $\widetilde{\Lambda} = -\rho''(0)$ is the variance-covariance matrix of $\widetilde{X}'(t)$. In particular, observe that

$$X''(\widehat{z}) = -\widetilde{\Lambda}X(\widehat{z}) + R(\widehat{z})\,,$$

where $R(\widehat{z}) = \sigma\widetilde{R}(\widehat{z})$. Using the Rice method of [13, Theorem 7.2] (see also [15]), it follows that the maximum λ_1 has for density w.r.t the Lebesgue measure on $\mathbb{R}^+$ at point $\ell > 0$

$$(\mathtt{cst})(-1)^d\,\mathbb{E}\left[\det(-\widetilde{\Lambda}X(0) + R(0))\mathbb{1}_{A_\ell}\middle|X(0) = \ell, X'(0) = 0\right]\sigma^{-1}\phi(\sigma^{-1}\ell),$$

where ϕ denotes the standard Gaussian density, A_ℓ is the event $\{X(y) \leqslant \ell,\ \forall y \in \mathbb{T}\}$ and ($\mathtt{cst}$), as in the following, denotes a positive constant. The numerical values ($\mathtt{cst}$) may vary from an occurrence to another and it may depend on m and σ which are assumed fixed in our framework.

5.1. The known variance case

We begin by the known variance case. The main observation is that the method of [13, Theorem 7.2] can be extended to compute the joint distribution of $(\lambda_1, \lambda_2, R(\widehat{z}))$ as follows.

- Denote $\mathbb{S}$ the set of symmetric matrices and pick a Borel set B on $\mathcal{D} := \mathbb{R}^2 \times \mathbb{S}$.
- For every $z \in \mathbb{T}$, recall that

$$\forall y \in \mathbb{T} \setminus \{z\}, \quad X^{|z}(y) := \frac{X(y) - \rho(z-y)X(z) + \langle \rho'(z-y), \widetilde{\Lambda}^{-1}X'(z)\rangle}{1 - \rho(z-y)}$$

and define

$$\forall z \in \mathbb{T}, \quad \lambda_2^z := \sup_{y \in \mathbb{T}\setminus\{z\}} X^{|z}(y)\,. \tag{13}$$

Remark that, for fixed $z \in \mathbb{T}$, λ_2^z is a.s. finite by Lemma 9, $X^{|z}(\cdot)$ is independent of $(X(z), X'(z))$ and, by way of consequence, λ_2^z is independent of $(X(z), X'(z))$. Furthermore, note that since $\mathbb{T}$ is without boundary, for $z = \widehat{z}$, one has $X'(z) = 0$ and $\lambda_2^z = \lambda_2$ as defined by (10).

- Observe that on the event $\{\forall y \neq z,\ X(y) < X(z)\}$ one has almost surely that $z = \widehat{z}$, $X(z) = \lambda_1$, $\lambda_2^z = \lambda_2$ and $R(z) = R(\widehat{z})$. Also, a simple computation shows that

$$\forall z \in \mathbb{T} \text{ s.t. } X'(z) = 0, \quad \mathbb{1}_{\{z=\widehat{z}\}} = \mathbb{1}_{\{\forall y \neq z,\ X(y) < X(z)\}} = \mathbb{1}_{\{0 < \lambda_2^z < X(z)\}},$$

almost surely. Hence, by unicity of $\widehat{z}$ and recalling that the set $\{z\,;\ X'(z) = 0\}$ is finite under ($\mathbf{A}_{\text{degen}}$), we deduce that

$$\sum_{z : X'(z)=0} \mathbb{1}_{\{(X(z),\lambda_2^z,R(z)) \in B\} \cap \{0 < \lambda_2^z < X(z)\}} = \mathbb{1}_{\{(\lambda_1,\lambda_2,R(\widehat{z})) \in B\}}.$$

- On $\mathcal{D}$ define smooth lower approximations $\varphi_B^{(n)}$ of the indicator function of B that converge when n goes to infinity i.e.

$$\forall (\ell_1, \ell_2, r) \in \Omega, \quad \varphi_B^{(n)}(\ell_1, \ell_2, r) \longrightarrow \mathbb{1}_{\{(\ell_1,\ell_2,r) \in B\} \cap \{0 < \ell_2 < \ell_1\}}.$$

- Apply Rice formula with weights [13, Theorem 6.4] (see also the proof of [13, Theorem 7.2]) to compute

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{z : X'(z)=0} \varphi_B^{(n)}\big(X(z), \lambda_2^z, R(z)\big) \Big] \\
&\quad = (\texttt{cst}) \int_{\mathbb{T}} \mathbb{E}\Big[|\det(-\widetilde{\Lambda} X(z) + R(z))|\, \varphi_B^{(n)}(X(z), \lambda_2^z, R(z)) \ \Big|\ X'(z) = 0 \Big] \mathrm{d}z \\
&\quad = (\texttt{cst}) \int_{\mathbb{T}} \mathbb{E}\Big[|\det(-\widetilde{\Lambda} X(z) + R(z))|\, \varphi_B^{(n)}(X(z), \lambda_2^z, R(z)) \Big] \mathrm{d}z
\end{aligned}$$

where the last equality relies on the fact that $(X(z), \lambda_2^z, R(z))$ is independent of $X'(z)$.

- Combining the previous observations and passing to the monotone limit as n tends to ∞ in the aforementioned Rice formula with weights, we get that

$$\begin{aligned}
&\mathbb{P}\big\{ \big(\lambda_1, \lambda_2, R(\widehat{z})\big) \in B \big\} \\
&\quad = \mathbb{E}\Big[\sum_{z : X'(z)=0} \mathbb{1}_{\{(X(z),\lambda_2^z,R(z)) \in B\} \cap \{0 < \lambda_2^z < X(z)\}} \Big] \\
&\quad = (\texttt{cst}) \int_{\mathbb{T}} \mathbb{E}\Big[|\det(-\widetilde{\Lambda} X(z) + R(z))| \mathbb{1}_{\{(X(z),\lambda_2^z,R(z)) \in B\} \cap \{0 < \lambda_2^z < X(z)\}} \Big] \mathrm{d}z \\
&\quad = (\texttt{cst})\ \mathbb{E}\Big[|\det(-\widetilde{\Lambda} X(0) + R(0))| \mathbb{1}_{\{(X(0),\lambda_2^0,R(0)) \in B\} \cap \{0 < \lambda_2^0 < X(0)\}} \Big], \\
&\quad = (\texttt{cst})\ \mathbb{E}\Big[\det(-\widetilde{\Lambda} X(0) + R(0)) \mathbb{1}_{\{0 < \lambda_2^0 < X(0)\}} \mathbb{1}_{\{(X(0),\lambda_2^0,R(0)) \in B\}} \Big], \qquad (14)
\end{aligned}$$

by stationarity and using that, on the event $\{0 < \lambda_2^0 < X(0)\}$, the matrix $-X''(0) = \widetilde{\Lambda} X(0) - R(0)$ belongs to the set of positive definite symmetric matrices, namely $\mathbb{S}^+$.

Before stating the key result on the joint density of $(\lambda_1, \lambda_2, R(\widehat{z}))$ we need to introduce a dominating measure. First, recall that $X(0)$ is independent of the pair $(\lambda_2^0, R(0))$. Then, observe that $(\lambda_2^0, R(0)) =$

$\sigma \times (\widetilde{\lambda}_2^0, \widetilde{R}(0))$ where $\widetilde{\lambda}_2^0$ is defined as in (13) for the process $\widetilde{X}$. Denote μ_1 the law of $(\widetilde{\lambda}_2^0, \widetilde{R}(0))$ and note that it does not depend on σ. Denote μ_σ the law of $(\lambda_2^0, R(0))$ and remark that for any Borel set B of $\mathbb{R}\times\mathbb{S}$, it holds $\mu_\sigma(\sigma B) = \mu_1(B)$. Eventually, remark that

$$\text{The law of } (X(0), \lambda_2^0, R(0)) \text{ is dominated by } \mathrm{Leb}(\mathbb{R}) \otimes \mu_\sigma\,, \tag{15}$$

where $\mathrm{Leb}(\mathbb{R})$ denotes the Lebesgue measure on $\mathbb{R}$. As a consequence we can prove the following proposition.

Proposition 3. *Under $\mathbb{H}_0$, the joint law $\mathcal{L}((\lambda_1, \lambda_2, R(\widehat{z})))$ of $(\lambda_1, \lambda_2, R(\widehat{z}))$ satisfies for all $(\ell_1, \ell_2, r) \in \mathbb{R}^2 \times \mathbb{S}$,*

$$\frac{\mathrm{d}\mathcal{L}((\lambda_1, \lambda_2, R(\widehat{z})))}{\mathrm{dLeb}(\mathbb{R}) \otimes \mu_\sigma}(\ell_1, \ell_2, r) = (\mathtt{cst}) \det(-\widetilde{\Lambda}\ell_1 + r)\mathbb{1}_{\{0<\ell_2<\ell_1\}}\sigma^{-1}\phi(\sigma^{-1}\ell_1)\,,$$

where $\mathrm{Leb}(\mathbb{R}) \otimes \mu_\sigma$ is defined by (15) and $\mathbb{S}$ denotes the set of symmetric matrices.

Proof. Observe that the density at point ℓ_1 of $X(0)$ with respect to the Lebesgue measure is $\sigma^{-1}\phi(\sigma^{-1}\ell_1)$ and recall (15). Now, for any Borel set B of $\mathbb{R}^2 \times \mathbb{S}$, note that

$$\begin{aligned}&\mathbb{E}\Big[\det(-\widetilde{\Lambda}X(0) + R(0))\mathbb{1}_{\{(0<\lambda_2^0<X(0)\}}\mathbb{1}_{\{(X(0),\lambda_2^0,R(0))\in B\}}\Big]\\ &\quad= \int_B \det(-\widetilde{\Lambda}\ell_1 + r)\mathbb{1}_{\{0<\ell_2<\ell_1\}}\sigma^{-1}\phi(\sigma^{-1}\ell_1)\mathrm{d}\ell_1\mu_\sigma(\mathrm{d}(\ell_2, r))\end{aligned}$$

thanks to (14), which prove the result. □

We can now state our result when the variance is known.

Theorem 4. *Set*

$$\forall r \in \mathbb{S}^+,\ \forall \ell > 0, \quad \overline{G}_r(\ell) := \int_\ell^{+\infty} \det(-\widetilde{\Lambda}u + r)\phi(u\sigma^{-1})\mathrm{d}u\,,$$

where $\widetilde{\Lambda}$ denotes the Hessian of the correlation function ρ of X at the origin. Under Assumptions ($\mathbf{A}_{\mathrm{norm}}$) and ($\mathbf{A}_{\mathrm{degen}}$), the test statistic

$$S^{\mathrm{Rice}} := \frac{\overline{G}_{R(\widehat{z})}(\lambda_1)}{\overline{G}_{R(\widehat{z})}(\lambda_2)} \sim \mathcal{U}([0,1])$$

under $\mathbb{H}_0$.

Proof. Using Proposition 3, we know that the density of λ_1 at ℓ_1 and conditional to $(\lambda_2, R(\widehat{z})) = (\ell_2, r)$ is equal to

$$(\mathtt{cst})\ \det(-\widetilde{\Lambda}\ell_1 + r)\phi(\sigma^{-1}\ell_1)\mathbb{1}_{\{0<\ell_2<\ell_1\}}.$$

It is well known that, if a random variable Z has for cumulative density function $\mathbb{F}$ then $\mathbb{F}(Z)$ follows an uniform distribution on $[0,1]$. This implies that, conditionally to $(\lambda_2, R(\widehat{z})) = (\ell_2, r)$,

$$\frac{\overline{G}_r(\lambda_1)}{\overline{G}_r(\ell_2)} \sim \mathcal{U}([0,1]).$$

Since the conditional distribution does not depend on (ℓ_2, r), it is also the non conditional distribution and it yields

$$\frac{\overline{G}_{R(\hat{z})}(\lambda_1)}{\overline{G}_{R(\hat{z})}(\lambda_2)} \sim \mathcal{U}([0,1])\,,$$

as claimed. □

5.2. The unknown variance case

5.2.1. Estimating the variance

When the variance σ^2 is unknown in (8), we precise here the assumptions and the estimator we use to estimate the variance. In this section, except for explicit examples, we consider a real valued centered Gaussian process Y not necessarily stationary defined on the 2-dimensional torus $\mathbb{T}$. Let $m \geq 2$ (possibly infinite) and assume that Y admits an order m Karhunen–Loève expansion in the sense that

$$Y = \sum_{i=1}^{m} \zeta_i f_i \text{ with } \mathbb{V}\mathrm{ar}(\zeta_i) = \sigma^2 \text{ and } \forall t \in \mathbb{T},\ \sum_{i=1}^{m} |f_i(t)|^2 = 1\,, \tag{KL(m)}$$

where the equality holds in $\mathbb{L}^2(\Omega)$ and $(f_1, \dots, f_m)$ is a system of non-zero functions orthogonal on $\mathbb{L}^2(\mathbb{T})$. Through our analysis, we need to consider one of the following assumptions.

- If m is finite,

$$\begin{aligned} &\exists (z_1, \dots, z_m) \in \mathbb{T}^m \ \text{ pairwise distincts s.t.} \\ &(Y(z_1), \dots, Y(z_m)) \text{ is non degenerated.} \end{aligned} \tag{ND(m)}$$

- If $m = \infty$,

$$\begin{aligned} &\forall p \in \mathbb{N}^\star,\ \exists (z_1, \dots, z_p) \in \mathbb{T}^p \ \text{ pairwise distincts s.t.} \\ &(Y(z_1), \dots, Y(z_p)) \text{ is non degenerated.} \end{aligned} \tag{ND(∞)}$$

Recall that a Gaussian vector is called non-degenerated if its variance-covariance matrix is non-degenerated, *i.e.,* it has full rank.

Some examples of process Y satisfying (KL(m)) and (ND(m)) with $m = \infty$ are given by the normalized Brownian motion and any Gaussian stationary process with a spectrum that admits an accumulation point, see [16, Page 203]. For instance, the process corresponding to the Super-Resolution problem satisfies (KL(m)) and (ND(m)) with m finite, namely m is twice the number of observed frequencies, see Section 6.

Definition 5. Let Y be a Gaussian process with constant variance $\sigma^2 = \mathbb{V}\mathrm{ar}(Y(\cdot))$ and satisfying Assumptions (KL(m)) and (ND(m)) with m finite. The quantity

$$\widehat{\sigma}^2_{\mathrm{KL}}(Y) := \frac{1}{m} \sum_{i=1}^{m} \zeta_i^2,$$

is called the Karhunen–Loève estimator of σ^2.

Remark 5. An explicit expression of the estimator $\widehat{\sigma}^2_{\mathrm{KL}}$ is always possible from some set of pairwise disjoint points $z_1, \dots, z_{m'}$ with $m' \geqslant m$. We only need to check that the variance-covariance matrix of the $(Y(z_1), \dots, Y(z_{m'}))$ has rank m.

Remark 6. Sufficiency considerations imply that $\widehat{\sigma}^2_{\mathrm{KL}}$ is an optimal unbiased estimator for the mean-squared error by Rao–Blackwell theorem.

Given the aforementioned definition, we are now able to construct variance estimators for the process X. We assume that the complex Gaussian process Z that define X satisfies the following hypotheses for some $N \in \mathbb{N}$.

$$Z \text{ admits a complex Karhunen–Loève expansion of order } N \tag{$\mathrm{KL}_Z(N)$}$$

and satisfies the following non-degeneracy conditions:

$$\begin{aligned} &\forall (t_1, \dots, t_N) \in [0, 2\pi)^N \ \text{ pairwise distincts,} \\ &(Z(t_1), Z(t_2), \dots, Z(t_N)) \text{ is non degenerated and} \\ &(Z(t_1), Z'(t_1), Z(t_3), \dots, Z(t_N)) \text{ is non degenerated.} \end{aligned} \tag{$\mathrm{ND}_Z(N)$}$$

Our aim is to build, for each $z \in \mathbb{T}$, two estimators of the variance σ^2 independently from $X(z)$ or $(X(z), X'(z))$. Indeed, in the following, we will distinguish two kinds of statistics. The first one is the limit of the finite dimensional statistic S^{Grid}, see Section 4. The second one is the case of the maximum over $\mathbb{T}$, see Section 5. Both cases won't use the same estimation of σ^2.

- In the grid situation, we define

$$X^z_{\mathrm{norm}}(y) := \frac{X(y) - \rho(z-y)X(z)}{\sqrt{1-\rho^2(z-y)}},$$

  where y belongs to $\mathbb{T} \setminus \{z\}$, $\rho(\cdot)$ denotes the correlation function of the process X and set

$$\widehat{\sigma}^2_z := \widehat{\sigma}^2_{\mathrm{KL}}\big(X^z_{\mathrm{norm}}(\cdot)\big)$$

  which is well defined, independent of $X(z)$ and with constant variance σ^2. Finally, we consider the variance estimator

$$\widehat{\sigma}^2 = \widehat{\sigma}^2_{\widehat{z}}, \tag{16}$$

  defined at point $\widehat{z}$ given by (9).
- In the continuous case, we define

$$X^{|z}_{\mathrm{norm}}(y) := \frac{X(y) - \rho(z-y)X(z) + \langle \rho'(z-y), \widetilde{\Lambda}^{-1} X'(z) \rangle}{\sqrt{1-\rho^2(z-y) + \langle \rho'(z-y), \widetilde{\Lambda}^{-1} \rho'(z-y) \rangle}},$$

  where y belongs to $\mathbb{T} \setminus \{z\}$ and set

$$\widehat{\sigma}^2_{|z} := \widehat{\sigma}^2_{\mathrm{KL}}\big(X^{|z}_{\mathrm{norm}}(\cdot)\big)$$

which is well defined, independent of $(X(z), X'(z))$ and with constant variance σ^2. Finally, we consider the variance estimator

$$\widehat{\sigma}^2_{|} = \widehat{\sigma}^2_{|\widehat{z}}\,, \tag{17}$$

defined at point $\widehat{z}$ given by (9).

Proposition 6. *Let Z satisfy* ($\mathrm{KL}_Z(N)$) *and* ($\mathrm{ND}_Z(N)$) *and set $z \in \mathbb{T}$ then the following claims are true under $\mathbb{H}_0$.*

(*a*) *$\widehat{\sigma}^2_z$ is well defined and follows a $\frac{\sigma^2\chi^2_{2N-1}}{2N-1}$ distribution.*
(*b*) *$\widehat{\sigma}^2_{|z}$ is well defined and follows a $\frac{\sigma^2\chi^2_{2N-3}}{2N-3}$ distribution.*
(*c*) *The process $X^z_{\mathrm{norm}}(\cdot)/\widehat{\sigma}_z$ is independent of $\widehat{\sigma}_z$, and the process $X^{|z}_{\mathrm{norm}}(\cdot)/\widehat{\sigma}_{|z}$ is independent of the random variable $\widehat{\sigma}_{|z}$.*

Proof. (*a*) Fix $z = (t_1, \theta_1) \in \mathbb{T}$. Since Z satisfies ($\mathrm{ND}_Z(N)$), there exists $(t_2, \dots, t_N) \in [0, 2\pi)^{N-1}$ pairwise different such that $(Z(t_1), Z(t_2), \dots, Z(t_N))$ is non degenerated. Then, considering $z_1 = z$, $z_{N+1} = z + (0, \pi/2)$ and

$$\forall i \in \{2, \dots, N\},\ z_i = (t_i, \theta_1) \ \text{ and } \ z_{N+i} = (t_i, \theta_1 + \pi/2),$$

the vector $V_1 := (X(z_1), \dots, X(z_{2N}))$ satisfies

$$\begin{aligned} 2N &= \mathrm{rank}(X(z_1), X(z_2), \dots, X(z_{2N})) \\ &= \mathrm{rank}(X(z_1), X^z_{\mathrm{norm}}(z_2), \dots, X^z_{\mathrm{norm}}(z_{2N})) \\ &= 1 + \mathrm{rank}(X^z_{\mathrm{norm}}(z_2), \dots, X^z_{\mathrm{norm}}(z_{2N})) \end{aligned}$$

where rank denotes the rank of the covariance matrix of a random vector. Deduce that $X^z_{\mathrm{norm}}(\cdot)$ satisfies $(\mathrm{ND}(2N-1))$. This, in turn, implies that the $2N$ functions

$$g_i(\cdot) = f_i(\cdot) - \rho(z - \cdot) f_i(z)$$

are in fact in a space of dimension $2N-1$ and a Gram–Schmidt orthogonalization in $\mathbb{L}^2(\mathbb{T})$ gives $(\mathrm{KL}(2N-1))$ for the process $X^z_{\mathrm{norm}}(\cdot)$. Finally, from $(X^z_{\mathrm{norm}}(z_2), \dots, X^z_{\mathrm{norm}}(z_{2N}))$, we compute $\widehat{\sigma}^2_z$ that follows the desired distribution.

(*b*) In the case of the regression over $(X(z), X'(z))$, remark that

$$\partial_\theta X(z) = X(t_1, \theta_1 + \pi/2) = X(z_2)$$

and $\partial_t X(z) = \mathrm{Re}(e^{-\imath\theta_1} Z'(t_1))$ where ∂_θ (resp. ∂_t) denote the partial derivative with respect to θ (resp. t). Because of hypothesis ($\mathrm{ND}_Z(N)$), the two vectors V_1 and

$$V_2 := (X(z_1), X(z_2), \mathrm{Re}(e^{-\imath\theta_1} Z'(t_1)), \mathrm{Im}(e^{-\imath\theta_1} Z'(t_1)), X(z_5), \dots, X(z_{2N}))$$

have rank $2N$ so both are invertible functions of $(\mathrm{Re}(\zeta_1), \mathrm{Im}(\zeta_1), \dots, \mathrm{Re}(\zeta_N), \mathrm{Im}(\zeta_N))$. In particular, $\mathrm{Im}(e^{-\imath\theta_1} Z'(t_1))$ is a linear combination of V_2. Let γ_1 and γ_2 be the coefficients associated to $X(z_3)$ and $X(z_4)$. By triangular combination, we deduce that the distribution of

$$(X(z_1), \partial_\theta X(z_1), \partial_t X(z_1), \gamma_1 X(z_3) + \gamma_2 X(z_4), X(z_5), \dots, X(z_{2N}))$$

is non-degenerated and so that $(\gamma_1, \gamma_2) \neq (0,0)$. Setting ψ such that

$$\cos(\psi) = \frac{\gamma_1}{\sqrt{\gamma_1^2+\gamma_2^2}} \quad \text{and} \quad \sin(\psi) = \frac{\gamma_2}{\sqrt{\gamma_1^2+\gamma_2^2}}$$

we get the non-degeneracy of

$$(X(z_1), \partial_\theta X(z_1), \partial_t X(z_1), X(z_{2N+1}), X(z_5), \dots, X(z_{2N}))$$

where $z_{2N+1} = (t_2, \theta_1 + \psi)$. Finally, similarly to the proof of the previous point, regression, scaling and independence prove that the rank of $(X^{|z}_{\text{norm}}(z_5), \dots, X^{|z}_{\text{norm}}(z_{2N+1}))$ is $2N-3$ so that $X^{|z}_{\text{norm}}(\cdot)$ satisfies $\text{KL}(2N-3)$ and $\text{ND}(2N-3)$ and that $\widehat{\sigma}^2_{|z}$ is well defined and distributed as $\frac{\sigma^2 \chi^2_{2N-3}}{2N-3}$.

(c) This is a direct consequence of the independence of the angle and the norm for each marginal Gaussian vector build from X^z_{norm} or $X^{|z}_{\text{norm}}$. □

Remark 7. When the complex process Z admits an infinite Karhunen–Loève decomposition, we need the following modified hypothesis

$$\begin{aligned} &\forall p \in \mathbb{N}^\star,\ \forall (t_1, \dots, t_p) \in [0, 2\pi)^p \ \text{ pairwise distincts,} \\ &(Z(t_1), Z(t_2), \dots, Z(t_p)) \text{ is non degenerated and} \\ &(Z(t_1), Z'(t_1), Z(t_3), \dots, Z(t_p)) \text{ is non degenerated.} \end{aligned} \qquad (\text{ND}_Z(\infty))$$

Indeed, for every enter $p \geq 1$, note that from the observation of the vector $(Z(t_1), Z(t_2), \dots, Z(t_p))$ (resp. $(Z(t_1), Z'(t_1), \dots, Z(t_p))$) for pairwise disjoint points $t_1, \dots, t_p$, we can construct an estimator, say $\widehat{\sigma}^2_{2p}$ (resp. $\widehat{\sigma}^2_{|2p}$), of σ^2 with distribution $\sigma^2 \chi^2_{2p-1}/(2p-1)$ (resp. $\sigma^2 \chi^2_{2p-3}/(2p-3)$) under $\mathbb{H}_0$. Making p tend to infinity, classical concentration inequalities and Borel–Cantelli lemma prove that $\widehat{\sigma}^2_{2p}$ (resp. $\widehat{\sigma}^2_{|2p}$) converges almost surely to σ^2 under $\mathbb{H}_0$. Thus the variance σ^2 is directly observable from the entire path of X. We still denote $\widehat{\sigma}^2_z$ (resp. $\widehat{\sigma}^2_{|z}$) this observation, where $z = z_1 = (t_1, \theta_1)$.

5.2.2. Computing the joint law

Hence, suppose that we observe $X = \sigma \widetilde{X}$ where $\sigma > 0$ is unknown. Assume that Z satisfies ($\text{KL}_Z(N)$) and ($\text{ND}_Z(N)$), and set $m = 2N$. The regression of the Hessian on $(X(z), X'(z))$ reads now

$$\forall z \in \mathbb{T}, \quad X''(z) = -\widetilde{\Lambda} X(z) + \sigma \widetilde{R}(z),$$

because $X'(z)$ is independent of $(X(z), X''(z))$ by stationarity. The variance being unknown, we estimate it using $\widehat{\sigma}^2_{|}$ which is defined by (17). For fixed $z \in \mathbb{T}$, by Claims (b) and (c) of Proposition 6, we know that the following random variables or random processes

$$X(z)\,, X'(z)\,, \ \frac{X^{|z}_{\text{norm}}(\cdot)}{\widehat{\sigma}_{|z}} \text{ and } \ \widehat{\sigma}_{|z}$$

are mutually independent. As $X^{|z}_{\text{norm}}(\cdot) = h_z(\cdot)\, X^{|z}(\cdot)$ where $h_z(\cdot)$ is a deterministic function and as Lemma 15 shows that $R(z)$ can be expressed as radial limits of $X^{|z}(\cdot)$ at point z, we get that

$$X(z)\,, X'(z)\,, \ \Big(\frac{X^{|z}(\cdot)}{\widehat{\sigma}_{|z}}, \frac{R(z)}{\widehat{\sigma}_{|z}}\Big) \text{ and } \ \widehat{\sigma}_{|z} \text{ are mutually independent,}$$

and by consequence

$$X(z)\,, X'(z)\,,\ \left(\frac{\lambda_2^z}{\widehat{\sigma}_{|z}}, \frac{R(z)}{\widehat{\sigma}_{|z}}\right) \text{ and } \widehat{\sigma}_{|z} \text{ are mutually independent.}$$

We turn now to the Rice formula described previously and introduce the notation

$$T_{2,z} := \frac{\lambda_2^z}{\widehat{\sigma}_{|z}} \text{ and } T_2 := T_{2,\widehat{z}}.$$

Denote $\mathrm{Leb}(\mathbb{R}^2)$ the Lebesgue measure on $\mathbb{R}^2$ and let $\overline{\mu}_1$ be the joint law of the couple of random variables $(T_{2,0}, R(0)/\widehat{\sigma}_{|0})$. Under $\mathbb{H}_0$, note that $X(0)$ is a centered Gaussian variable with variance σ^2 and $\widehat{\sigma}_{|0}/\sigma$ is distributed as a *chi*-distribution with $m-3$ degrees of freedom, *i.e.*, the law of density

$$f_{\chi_{m-3}}(s) = \frac{2^{1-\frac{m-3}{2}}}{\overline{\Gamma}\left(\frac{m-3}{2}\right)} s^{m-4} \exp(-s^2/2)$$

where $\overline{\Gamma}$ is the Gamma function. Then the quadruplet $(X(0), \widehat{\sigma}_{|0}/\sigma, T_{2,0}, R(0)/\widehat{\sigma}_{|0})$ has a density with respect to $\mathrm{Leb}(\mathbb{R}^2) \otimes \overline{\mu}_1$ at point $(\ell_1, s, t_2, r) \in \mathbb{R}^3 \times \mathbb{S}$ equal to

$$(\mathtt{cst})\, s^{m-4} \exp\left(-\frac{s^2(m-3)}{2}\right) \sigma^{-1} \phi(\sigma^{-1}\ell_1).$$

Using the same method as for the proof of Proposition 3 we have the following proposition.

Proposition 7. *Assume that Z satisfies ($\mathbf{A}_{\mathrm{norm}}$), ($\mathbf{A}_{\mathrm{degen}}$), ($\mathrm{KL}_Z(N)$) and ($\mathrm{ND}_Z(N)$), and set $m = 2N$. Then, under $\mathbb{H}_0$, the joint distribution of $\left(\lambda_1, \widehat{\sigma}_{|}/\sigma, T_2, R(\widehat{z})/\widehat{\sigma}_{|}\right)$ has a density with respect to $\mathrm{Leb}(\mathbb{R}^2) \otimes \overline{\mu}_1$ at point $(\ell_1, s, t_2, r) \in \mathbb{R}^3 \times \mathbb{S}^+$ equal to*

$$(\mathtt{cst}) \det(-\widetilde{\Lambda}\ell_1 + \sigma s r)\, s^{m-4} \exp\left(-\frac{s^2(m-3)}{2}\right) \phi(\sigma^{-1}\ell_1) \mathbb{1}_{\{0<\sigma s t_2 < \ell_1\}},$$

where $(\mathtt{cst})$ *is a positive constant that may depend on m and σ.*

Consequently, we derive the following result.

Theorem 8. *Assume that Z satisfies ($\mathbf{A}_{\mathrm{norm}}$), ($\mathbf{A}_{\mathrm{degen}}$), ($\mathrm{KL}_Z(N)$) and ($\mathrm{ND}_Z(N)$), and set $m = 2N$. For all $r \in \mathbb{S}^+$, define $\overline{H}_r(\cdot)$ as*

$$\forall \ell > 0, \quad \overline{H}_r(\ell) := \int_{\ell}^{+\infty} \det\left(-\widetilde{\Lambda}t_1 + r\right) f_{m-1}\left(t_1\sqrt{\frac{m-1}{m-3}}\right) \mathrm{d}t_1,$$

where f_{m-1} is the density of the Student distribution with $m-1$ degrees of freedom. Under the null $\mathbb{H}_0$, the test statistic

$$T^{\mathrm{Rice}} := \frac{\overline{H}_{R(\widehat{z})}(T_1)}{\overline{H}_{R(\widehat{z})}(T_2)} \sim \mathcal{U}([0,1]),$$

where $T_1 := \lambda_1/\widehat{\sigma}_{|}$, $T_2 = \lambda_2/\widehat{\sigma}_{|}$ and $\widehat{\sigma}_{|}$ is defined by (17).

Proof. First, using Proposition 7 and the change of variable $t_1 = \frac{\ell_1}{\sigma s}$, the joint distribution of the quadruplet $(T_1, \widehat{\sigma}_|/\sigma, T_2, R(\widehat{z})/\widehat{\sigma}_|))$ at point (t_1, s, t_2, r) is given by

$$\begin{aligned}
&(\mathtt{cst}) \det(\sigma s(-\widetilde{\Lambda} t_1 + r)) s^{m-3} \exp\left(-\frac{s^2(m-3)}{2}\right) \phi(st_1) \mathbb{1}_{\{0<t_2<t_1\}} \\
&= (\mathtt{cst}) \det(-\widetilde{\Lambda} t_1 + r) s^{m-1} \exp\left(-\left(s\sqrt{\frac{m-3}{m-1}}\right)^2 \frac{m-1}{2}\right) \phi(st_1) \mathbb{1}_{\{0<t_2<t_1\}}.
\end{aligned}$$

Second, note that if X and Y are two independent random variables of density f_X and f_Y then the density of X/Y satisfies

$$f_{X/Y}(z) = \int_{\mathbb{R}} f_X(zy) y f_Y(y) \mathrm{d}y.$$

In our case, integrating over s and with the change of variable $s \leftarrow s\sqrt{(m-1)/(m-3)}$, it holds

$$\begin{aligned}
&\int_{\mathbb{R}^+} \phi(st_1) s^{m-1} \exp\left[-\left(s\sqrt{\frac{m-3}{m-1}}\right)^2 \frac{m-1}{2}\right] \mathrm{d}s \\
&= (\mathtt{cst}) \int_{\mathbb{R}^+} \phi\left(st_1\sqrt{\frac{m-1}{m-3}}\right) s s^{m-2} \exp\left[-\frac{s^2(m-1)}{2}\right] \mathrm{d}s \\
&= (\mathtt{cst}) \int_{\mathbb{R}^+} \phi\left(st_1\sqrt{\frac{m-1}{m-3}}\right) s f_{\frac{\chi_{m-1}}{\sqrt{m-1}}}(s) \mathrm{d}s \\
&= f_{m-1}\left(t_1\sqrt{\frac{m-1}{m-3}}\right).
\end{aligned}$$

Putting together, the density of $(T_1, T_2, R(\widehat{z})/\widehat{\sigma})$ at point (t_1, t_2, r) is now given by

$$(\mathtt{cst}) \det(-\widetilde{\Lambda} t_1 + r) f_{m-1}\left(t_1\sqrt{\frac{m-1}{m-3}}\right) \mathbb{1}_{\{0<t_2<t_1\}},$$

and we conclude using the same trick as the one of Theorem 4. □

6. Applications to the super-resolution theory

6.1. Framework and results

Deconvolution over the space of complex-valued Radon measure has recently attracted a lot of attention in the "Super-Resolution" community—and its companion formulation in "Line spectral estimation". A standard aim is to recover fine scale details of an image from few low frequency measurements—ideally the observation is given by a low-pass filter. The novelty in this body of work relies on new theoretical guarantees of the ℓ_1-minimization over the space of Radon measures with finite support. Some recent works on this topic can be found in the papers [7,2,6,3,8,17,18,5] and references therein.

An important example throughout this paper is given by the Super-Resolution problem which can be stated as follows. Let $\nu^0 \in (\mathcal{M}([0, 2\pi), \mathbb{C}), \|\cdot\|_1)$ a complex-valued Radon measure on the one dimensional

torus identified to $[0, 2\pi)$ equipped with the natural circle-wise metric. Note that $||\cdot||_1$ denotes the total variation norm on $\mathcal{M}([0, 2\pi))$. The space $(\mathcal{M}([0, 2\pi), \mathbb{C}), \|\cdot\|_1)$ can be defined as the topological dual space of continuous functions on $[0, 2\pi)$ equipped with the L^∞-norm.

Let $N = 2f_c + 1$ where $f_c \geq 1$ is referred to as the "frequency cut-off". Denote by $\mathbf{D}_N$ the Dirichlet kernel defined by

$$\forall t \in [0, 2\pi), \quad \mathbf{D}_N(t) := \frac{\sin(Nt/2)}{\sin(t/2)}.$$

Consider the linear operator $\mathcal{F}_N : \mathcal{M}([0, 2\pi), \mathbb{C}) \to \mathbb{C}^N$ that maps any complex-valued Radon measure ν onto its Fourier coefficients $c_k(\nu)$ where

$$c_k(\nu) := \int_{\mathbb{T}} \exp(-\imath k x)\nu(\mathrm{d}x)$$

for integers k such that $|k| \leqslant f_c$. Consider $\zeta = (\zeta_k)_k$ where $\zeta_k = \zeta_{1,k} + \imath\zeta_{2,k}$ and $\zeta_{\ell,k}$ are i.i.d. standard Gaussian random variables for $|k| \leqslant f_c$ and $\ell = 1, 2$. In the Super-Resolution frame, we observe a perturbed version of the Fourier coefficients, namely

$$y = \frac{1}{\sqrt{N}}\mathcal{F}_N(\nu^0) + \sigma\zeta\,.$$

Applying $\mathcal{F}_N^\star$—the dual operator of $\mathcal{F}_N$, remark that we observe the trigonometric polynomial

$$Z := \frac{1}{\sqrt{N}}\mathcal{F}_N^\star(y)$$

which reads as

$$\forall t \in [0, 2\pi), \quad Z(t) = \frac{1}{N}\int_{\mathbb{T}} \mathbf{D}_N(t - x)\nu^0(\mathrm{d}x) + \sigma \sum_{k=-f_c}^{f_c} \frac{1}{\sqrt{N}}\zeta_k \exp(\imath k t). \tag{18}$$

Hence, one observes Z and infers on ν^0 assuming that it has finite support. To this purpose, consider the process X defined for all $(t, \theta) \in \mathbb{T}$ by

$$X(t, \theta) := \mathrm{Re}(e^{-\imath\theta}Z(t)) = \cos(\theta)\,\mathrm{Re}(Z(t)) + \sin(\theta)\,\mathrm{Im}(Z(t)), \tag{19}$$

where Re and Im denote the real and imaginary part of a complex number. When $\nu^0 \equiv 0$, remark that the processes $A_1 = \mathrm{Re}(Z)$ and $A_2 = \mathrm{Im}(Z)$ are two independent and identically distributed real-valued processes with $\mathcal{C}^\infty$-paths. An elementary computation shows that X has correlation function ρ and A_1 has correlation function Γ with

$$\rho(z - y) = \cos(\theta - \alpha)\Gamma(t - s) \quad \text{where} \quad \Gamma(t - s) = \mathbf{D}_N(t - s)/N$$

for all $z = (t, \theta)$ and $y = (s, \alpha)$ in $\mathbb{T}$. Remark that ($\mathbf{A}_{\mathrm{norm}}$) holds true for Γ. In this case, we are testing

$$\mathbb{H}_0 : \text{“}\mathcal{F}_N^\star(\mathcal{F}_N(\nu^0)) \equiv 0\text{”} \quad \text{against} \quad \mathbb{H}_1 : \text{“}\exists t \in [0, 2\pi),\ \mathcal{F}_N^\star(\mathcal{F}_N(\nu^0))(t) \neq 0\text{”},$$

or equivalently

$$\mathbb{H}_0 : \text{“}\nu^0 \equiv 0\text{”} \quad \text{against} \quad \mathbb{H}_1 : \text{“}\exists t \in [0, 2\pi),\ \nu^0(t) \neq 0\text{”}.$$

Subtracting the known measure ν^0, remark that this framework encompasses testing problem whose null hypothesis is any single hypothesis $\mathbb{H}_0$: "$\nu^0 \equiv \nu_0$ " against alternatives of the form $\mathbb{H}_1$: "$\exists t \in [0, 2\pi)$, $\nu^0(t) \neq \nu_0(t)$ ".

Furthermore, we have the following propositions. First, we check that we can apply our results to the Super-Resolution process.

Proposition 9. *The process X defined by (19) satisfies Condition (KL(m)) and Condition (ND(m)) with $m = 2N = 4f_c + 2$.*

Then, we derive a first result when the noise level σ is known.

Proposition 10. *Under the null $\mathbb{H}_0$, the test statistic*

$$S_{\mathrm{SR}}^{\mathrm{Rice}} = \frac{\sigma(\alpha_1\lambda_1 + \alpha_2)\phi(\lambda_1/\sigma) + (\alpha_1\sigma^2 - \alpha_3^2)\overline{\Phi}(\lambda_1/\sigma)}{\sigma(\alpha_1\lambda_2 + \alpha_2)\phi(\lambda_2/\sigma) + (\alpha_1\sigma^2 - \alpha_3^2)\overline{\Phi}(\lambda_2/\sigma)} \sim \mathcal{U}([0,1]),$$

where Φ is the standard Gaussian cumulative distribution function , $\overline{\Phi} = 1 - \Phi$ its survival function, ϕ its density function, (λ_1, λ_2) is defined by ((9), (10)) and

$$\begin{cases} \alpha_1 = \frac{1}{3} f_c(f_c + 1), \\ \alpha_2 = \frac{1}{\sqrt{N}} \sum_{k=-f_c}^{f_c} (k^2 - \alpha_1) \times \mathrm{Re}(y_k e^{\imath(k\widehat{t} - \widehat{\theta})}), \\ \alpha_3 = \frac{1}{\sqrt{N}} \sum_{k=-f_c}^{f_c} k \times \mathrm{Re}(y_k e^{\imath(k\widehat{t} - \widehat{\theta})}). \end{cases}$$

Finally, we have the following result when the noise level σ is unknown.

Proposition 11. *Under the null $\mathbb{H}_0$, the test statistic*

$$T_{\mathrm{SR}}^{\mathrm{Rice}} = \frac{\alpha_1\overline{F}_{m-3}(T_1) + (\alpha_1 T_1 + \alpha_2) f_{m-3}(T_1) - \gamma_m^{-1}\alpha_3^2\overline{F}_{m-1}(T_1)}{\alpha_1\overline{F}_{m-3}(T_2) + (\alpha_1 T_2 + \alpha_2) f_{m-3}(T_2) - \gamma_m^{-1}\alpha_3^2\overline{F}_{m-1}(T_2)} \sim \mathcal{U}([0,1]),$$

where F_d is the Student cumulative distribution function with d degrees of freedom, $\overline{F}_d = 1 - F_d$ its survival function, f_d its density function, $T_1 = \lambda_1/\widehat{\sigma}_|$, $T_2 = \lambda_2/\widehat{\sigma}_|$, $\widehat{\sigma}_|$ is defined by (17) and $\gamma_m = \frac{m-3}{m-2}\frac{\overline{\Gamma}(\frac{m}{2})\overline{\Gamma}(\frac{m-3}{2})}{\overline{\Gamma}(\frac{m-1}{2})\overline{\Gamma}(\frac{m-2}{2})}$.

A proof of these propositions can be found in Appendix A.4.

6.2. A numerical study

A Python code (and a Jupyter notebook) illustrating the following numerical experiments can be found at: https://github.com/ydecastro/super-resolution-testing.

6.2.1. Computation of λ_2

To build our test statistic S^{Rice} in the Super-Resolution context (namely $S_{\mathrm{SR}}^{\mathrm{Rice}}$), we need to compute three quantities. The first one is λ_1, the maximum of $X(\cdot)$ over the torus $\mathbb{T}$. Its simple form allow us to use classical optimization routines, for instance **scipy.optimize.minimize** on **Python**, **fminsearch** on **MATLAB** or **optim** on **R** both combined with global resolution options on $\mathbb{T}$. The second one is $R = R(\widehat{z})$ which appears in the test statistic through the coefficients α_1, α_2 and α_3 that are simple functions of the observation y and $\widehat{z}$. Finally, the third one is

$$\lambda_2 = \lambda_2^{\widehat{z}} = \lambda_1 + \max_{y\in\mathbb{T}} \left\{ \frac{X(y) - X(\widehat{z})}{1 - \rho(\widehat{z} - y)} \right\}.$$

Contrary to λ_1, there is some indetermination problem when y is close to $\widehat{z}$. In particular, the approximation of $\widehat{z}$ is by definition not exact and the radial limits of $X^|$ are not numerically achieved. A way to get around that is the integral form of the remainder in Taylor's theorem. In full generality, we compute

$$\lambda_2^{\widehat{z}} - \lambda_1 = \max_{y\in\mathbb{T}} \left\{ \frac{\int_0^1 (1-h)\ (y-\widehat{z})^T X''(\widehat{z} + h(y-\widehat{z}))(y-\widehat{z})\ \mathrm{d}h}{\int_0^1 (1-h)\ (y-\widehat{z})^T \rho''(\widehat{z} + h(y-\widehat{z}))(y-\widehat{z})\ \mathrm{d}h} \right\}.$$

Denote by $r = ||y - \widehat{z}||_2$ the distance between y and $\widehat{z}$. The numerical indetermination occurs for small values of r. But remark that one can factorize r^2 in both the numerator and the denominator. This leads to the expression

$$\lambda_2^{\widehat{z}} - \lambda_1 = \max_{y\in\mathbb{T}} \left\{ \frac{\int_0^1 (1-h)\ (\frac{y-\widehat{z}}{r})^T X''(\widehat{z} + h(y-\widehat{z}))(\frac{y-\widehat{z}}{r})\ \mathrm{d}h}{\int_0^1 (1-h)\ (\frac{y-\widehat{z}}{r})^T \rho''(\widehat{z} + h(y-\widehat{z}))(\frac{y-\widehat{z}}{r})\ \mathrm{d}h} \right\},$$

which is more robust in practice. In the Super-Resolution case, elementary trigonometry identities give the following simpler form of the denominator

$$\sum_{k=-f_c}^{f_c} \left(k\cos(\alpha) - \sin(\alpha)\right)^2 \times \mathrm{sinc}\left(\frac{r(k\cos(\alpha) - \sin(\alpha))}{2}\right)^2$$

where $y - \widehat{z} = (r\cos(\alpha), r\sin(\alpha))$ and **sinc** denote the cardinal sine function, i.e.

$$\mathrm{sinc}(x) = \begin{cases} \dfrac{\sin x}{x} & \text{if } x \neq 0, \\ 1 & \text{if } x = 0, \end{cases}$$

which is a numerically robust function. We conclude the optimization using the same routine as the one of λ_1.

6.2.2. Monte-Carlo experiment

In this section we compare the cumulative distribution of several statistics of test in the case where the variance is known, namely

- The statistics of the Rice test S^{Rice}, given by Theorem 4, are displayed in blue.
- The statistics of the Spacing test S^{ST}, given by (12), are displayed in green.
- The statistics of the Spacing test on grids G_n given by $\overline{\Phi}(\lambda_{1,n})/\overline{\Phi}(\lambda_{2,n})$ are displayed with a color that take the respective values green, red, purple and cyan for sizes equal to $3^2, 10^2, 32^2, 50^2$.

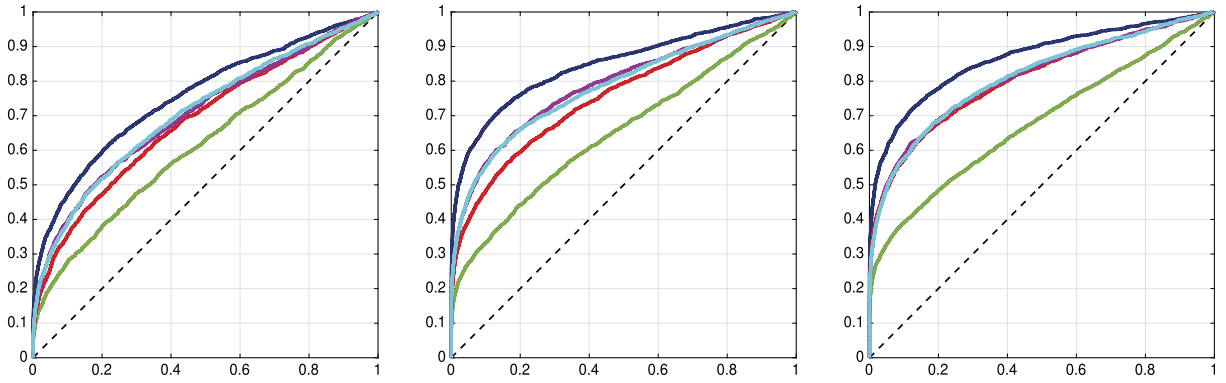

Fig. 5. Same as Fig. 2 except that (a) $f_c = 7$, (b) the alternative is defined by two spikes at random locations (with a constraint of separation) (c) the weights are now from left to right $(\log N, \log N)$; $(\log N, \sqrt{N})$; $(\sqrt{N}, \sqrt{N})$.

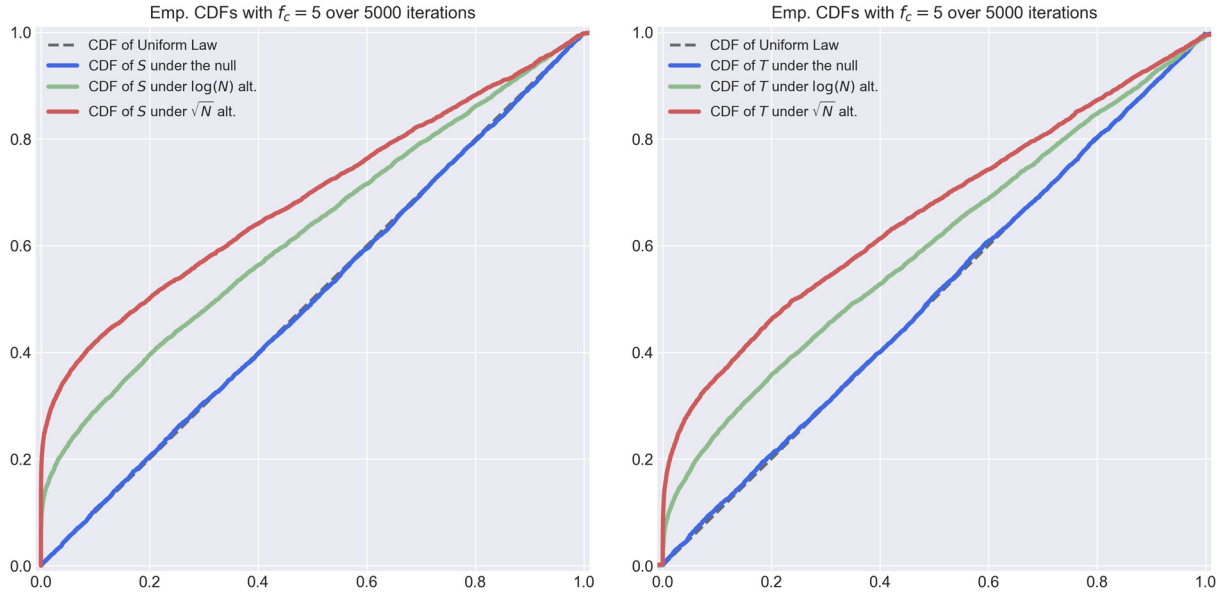

Fig. 6. We compute three empirical cumulative distribution functions: under the null (dashed gray line), under one spike alternative of size $\log N$ (green line) and under one spike alternative of size $\sqrt{N}$. The left panel uses the statistic S (known variance case) and the right panel the statistic T (unknown variance case). We witness a slight loss of power in this later case.

- The grid test, based on S^{Grid} of Theorem 1 can be viewed as the limit of the discrete grid tests above as the size growths to infinity. As one can see in the figures, there is some evidence that this limit is numerically reached for a size $n = 50^2$.

We complete each graph by the diagonal to the cumulative distribution function of the uniform law on $[0, 1]$ displayed in black. All the figures are based on 2000 simulations of the corresponding statistics.

The first figure studies the distribution of S^{Rice} and S^{ST} under the Null. This figure is displayed in the introduction (see Fig. 3). The second figure deals with the grid statistic and S^{Rice} under various alternatives defined by a single spike and compares the power of the Rice test with the discrete grid tests, see Fig. 4. Finally, the third figure performs the same study but with an alternative defined by two atoms, see Fig. 5.

A last set of experiments is devoted to the computation of the testing procedure when the noise level is unknown, see Fig. 6.

These latter numerical experiments were conducted using a Python code. The notebook **testing-super-resolution.ipynb** available at *github.com/ydecastro/super-resolution-testing* allows to reproduce these experiments.

6.3. Discussion

Fig. 3 suggests that the Spacing test is highly non-conservative which is a major drawback. For instance, when $f_c = 7$, the empirical level of the Spacing test at a nominal level of 5% is in fact 11,3%, showing that this test is very non-conservative. For its part, the Rice test is exact as predicted by the theory. This numerical agreement prove that the numerical algorithm described in Section 6.2.1 is efficient.

In Fig. 4 and 5 we see that the power of the discrete grid tests may seem an increasing function of the number of points of the grid. This power seems to converge since the curves associated to 32^2 (purple) and 50^2 (cyan) are almost indistinguishable. This suggests that the Rice test (blue) is always more powerful than the discrete grid test or the limit grid test. Consequently, it seems unbiased for any choice of alternative.

In conclusion the Rice test seems to be the best choice even if we are still not able to prove theoretically that it is unbiased.

Acknowledgments

The authors would like to thank the referees for their useful comments and interesting remarks that have improved the presentation of this paper.

Appendix A. Proofs

We denote for random variables, $X_n = o_P(r_n)$ and $Y_n = O_P(r_n)$ (for $r_n \neq 0$) means that $r_n^{-1}\|X_n\|$ converges to 0 in probability and $r_n^{-1}\|Y_n\|$ is uniformly tight, respectively. Furthermore, we consider the following processes.

- The stationary process $X(z) = X(t, \theta)$ defined on $\mathbb{T}$ with covariance function given by $\mathrm{Cov}(X(y), X(z)) = \sigma^2\rho(z-y)$ where we recall the correlation function is given by $\rho(z-y) = \cos(\theta - \alpha)\Gamma(t-s)$,
- For every $z \in \mathbb{T}$, recall the regressions with respect to $X(z)$

$$\forall y \in \mathbb{T} \setminus \{z\}, \quad X^z(y) = \frac{X(y) - X(z)\rho(z-y)}{1-\rho(z-y)} = X(z) + \frac{X(y)-X(z)}{1-\rho(z-y)}\,,$$
$$X^z_{\mathrm{norm}}(y) = \frac{X(y) - \rho(z-y)X(z)}{\sqrt{1-\rho^2(z-y)}}\,.$$

- For every $z \in \mathbb{T}$, recall the regressions with respect to $(X(z), X'(z))$

$$\forall y \in \mathbb{T} \setminus \{z\}, \quad X^{|z}(y) = \frac{X(y) - \rho(z-y)X(z) + \langle \rho'(z-y), \widetilde{\Lambda}^{-1}X'(z)\rangle}{1-\rho(z-y)}\,,$$
$$X^{|z}_{\mathrm{norm}}(y) = \frac{X(y) - \rho(z-y)X(z) + \langle \rho'(z-y), \widetilde{\Lambda}^{-1}X'(z)\rangle}{\sqrt{1-\rho^2(z-y) + \langle \rho'(z-y), \widetilde{\Lambda}^{-1}\rho'(z-y)\rangle}}\,.$$

In particular, recall that $\widehat{z}$ is defined by (9) so $X'(\widehat{z}) = 0$ and it yields that $X^{\widehat{z}} = X^{|\widehat{z}}$.

A.1. Proof of Theorem 1

Since the variance is known, we consider without loss of generality that $\sigma^2 = 1$. Using the metric given by the quadratic form represented by X'', we can consider the closest point $\overline{z}_n$ of the grid G_n to $\widehat{z}$ by

$$\overline{z}_n = \arg\min_{u\in G_n} \|\widehat{z} - u\|_{X''} = \arg\min_{u\in G_n} \langle \widehat{z} - u, -X''(\widehat{z})(\widehat{z} - u)\rangle \,.$$

The main claim is that, while it holds $\lambda_{1,n} \to \lambda_1$ a.s., we don't have the same result for $\lambda_{2,n}$, see Lemma 14. We begin with the following preliminary result, which is related to the result of Azaïs–Chassan [19].

Lemma 12. *Under $\mathbb{H}_0$ and conditionally to X'', $\Delta_n^{-1}(\widehat{z} - \overline{z}_n)$ follows a uniform distribution on V_0 and this distribution is independent from λ_1 and λ_2.*

Proof. Remark that $\widehat{z}$ has uniform distribution on $\mathbb{T}$ by stationarity and this distribution is independent from λ_1 and λ_2. Let B be a Borelian in $\mathbb{R}^2$. Remark that $\widehat{z} - \overline{z}_n \in \Delta_n V_0$ by definition of $\overline{z}_n$ and note that $\overline{z}_n \in \Delta_n \mathbb{Z}^2$. Conditionally to X'', it holds

$$\begin{aligned}
\mathbb{P}\{\Delta_n^{-1}(\widehat{z} - \overline{z}_n) \in B\} &= \mathbb{P}\{\Delta_n^{-1}(\widehat{z} - \overline{z}_n) \in B \cap V_0\} \\
&= \mathbb{P}\{(\widehat{z} - \overline{z}_n) \in \Delta_n(B \cap V_0)\} \\
&= \mathbb{P}\{\widehat{z} \in \Delta_n(B \cap V_0 + \mathbb{Z}^2)\} \,.
\end{aligned}$$

Since $\widehat{z}$ has uniform distribution on $\mathbb{T}$ and since $V_0 + \mathbb{Z}^2$ is a partition of $\mathbb{R}^2$, it holds that

$$\mathbb{P}\{\widehat{z} \in \Delta_n(B \cap V_0 + \mathbb{Z}^2)\} = \mathbb{P}\{\widehat{z} \in 2\pi(B \cap V_0 + \mathbb{Z}^2)\} = \frac{\mathrm{Leb}(\mathbb{R}^2)(2\pi(B \cap V_0))}{\mathrm{Leb}(\mathbb{R}^2)(2\pi V_0)} \,,$$

where $\mathrm{Leb}(\mathbb{R}^2)$ denotes the Lebesgue measure on $\mathbb{R}^2$. □

Lemma 13. *Under $\mathbb{H}_0$, it holds that*

(a) $X(\widehat{z}_n) - X(\overline{z}_n) = o_P(\Delta_n^2)$.
(b) $\mathbb{P}\{\widehat{z}_n \neq \overline{z}_n\} \to 0$ *as n goes to ∞.*
(c) Let F be any measurable function, then $F(\widehat{z}_n) - F(\overline{z}_n)$ tends to zero in probability at arbitrary speed.
(d) Almost surely, one has $\overline{z}_n \to \widehat{z}$ and $\widehat{z}_n \to \widehat{z}$ as n goes to infinity.

Proof. Let $\varepsilon > 0$. By definition of $\overline{z}_n$ and since $V_0 \subset [-1, 1]^2$, it holds that

$$\|\widehat{z} - \overline{z}_n\| \leqslant \sqrt{2}\Delta_n \,, \tag{A.1}$$

almost surely. Since X has $\mathcal{C}^2$-paths and by Taylor expansion, one has

$$X(\widehat{z}) - X(\overline{z}_n) = (1/2)\|\widehat{z} - \overline{z}_n\|^2_{X''} + o_P(\Delta_n^2) \tag{A.2}$$

Since $-X''$ is positive definite, there exists $M > 0$ sufficiently large such that

$$(1/M)\mathrm{Id}_2 \preccurlyeq -X'' \preccurlyeq M\mathrm{Id}_2$$

where $\preccurlyeq$ denotes the Lowner ordering between symmetric matrices. Then, it holds

$$\forall z \in \mathbb{R}^2, \quad (1/M)\|z\|^2 \leq \|z\|^2_{X''} \leqslant M\|z\|^2 . \tag{A.3}$$

From (A.1), (A.2) and (A.3), we deduce that

$$0 \leqslant X(\widehat{z}) - X(\widehat{z}_n) \leqslant X(\widehat{z}) - X(\overline{z}_n) = O_P(\Delta_n^2) , \tag{A.4}$$

using the optimality of $\widehat{z}$ and $\widehat{z}_n$.

By compactness of $\mathbb{T}$, uniqueness of optimum $\widehat{z} \in \mathbb{R}^2$ and $\mathcal{C}^2$-continuity of X, there exists $\eta > 0$ and a neighborhood $N_0 \subset \mathbb{R}^2$ of $\widehat{z} \in \mathbb{R}^2$ such that $X(\widehat{z}) - \eta \geqslant X(z)$ for any $z \notin N_0$ and

$$\forall z \in N_0, \quad (1/4)||\widehat{z} - z||^2_{X''} \leqslant X(\widehat{z}) - X(z) \leqslant ||\widehat{z} - z||^2_{X''} \tag{A.5}$$

using again a Taylor expansion as in (A.2). Using (A.4), it holds that, on an event of probability at least $1 - \varepsilon/4$ and for n large enough, $0 \leqslant X(\widehat{z}) - X(\widehat{z}_n) \leq \eta/2$ implying that $\widehat{z}_n \in N_0$. Invoke (A.3), (A.4) and (A.5) to deduce that $\widehat{z} - \widehat{z}_n = O_P(\Delta_n)$.

Using Taylor formula again, we get that

$$X(\widehat{z}) - X(\widehat{z}_n) = (1/2)||\widehat{z} - \widehat{z}_n||^2_{X''} + o_P(\Delta_n^2) . \tag{A.6}$$

By optimality of $\widehat{z}_n$ and $\overline{z}_n$ and using (A.2) and (A.6), one gets

$$0 \leqslant X(\widehat{z}_n) - X(\overline{z}_n) \leq (1/2)(||\widehat{z} - \overline{z}_n||^2_{X''} - ||\widehat{z} - \widehat{z}_n||^2_{X''}) + o_P(\Delta_n^2) . \tag{A.7}$$

Observing that $||\widehat{z} - \overline{z}_n||^2_{X''} - ||\widehat{z} - \widehat{z}_n||^2_{X''} \leq 0$, we get (a).

Conditionally to X'' and in the metric defined by $||\cdot||_{X''}$, there exists $\eta' > 0$, such that the η'-neighborhood, denoted by $N_{\eta'}$, of the boundary ∂V_0 of V_0 has relative volume (for the Lebesgue measure) less than $\varepsilon/8$. More precisely, $N_{\eta'}$ denotes the set of points in $V_0 \subset \mathbb{R}^2$ with $||\cdot||_{X''}$-distance less than η' to the boundary of V_0. In particular,

$$\forall k \in \mathbb{Z}^2 \setminus \{0\}, \ \forall z \in V_0 \setminus N_{\eta'}, \quad ||z||_{X''} + \eta' \leqslant ||z - k||_{X''} ,$$

by Cauchy–Schwarz inequality. Using Lemma 12 and by homogeneity, we deduce that it holds

$$\forall g \in \Delta_n\mathbb{Z}^2 \setminus \{0\}, \quad ||\widehat{z} - \overline{z}_n||_{X''} + \eta'\Delta_n \leqslant ||\widehat{z} - g||_{X''} ,$$

with probability at least $1 - \varepsilon/8$. It follows that

$$\forall g \in \Delta_n\mathbb{Z}^2 \setminus \{0\}, \quad ||\widehat{z} - \overline{z}_n||^2_{X''} + (\eta')^2\Delta_n^2 \leqslant ||\widehat{z} - g||^2_{X''} ,$$

using that $(a + b)^2 \geqslant a^2 + b^2$ for $a, b \geqslant 0$. Now, invoke (A.7) to get that

$$0 \leqslant -\frac{(\eta')^2}{2}\Delta_n^2 \ \mathbf{1}_{\{\overline{z}_n \neq \widehat{z}_n\}} + o_P(\Delta_n^2) .$$

On these events, we get that, for n sufficiently large, $\widehat{z}_n$ and $\overline{z}_n$ must be equal except on an event of probability at most $\varepsilon/4 + \varepsilon/8 \leq \varepsilon$. Furthermore, this result holds unconditionally in X''. We deduce that $\limsup \mathbb{P}\{\overline{z}_n \neq \widehat{z}_n\} \leqslant \varepsilon$, proving (b). Note that (c) is a consequence of the fact that, for n sufficiently large, $\widehat{z}_n$ and $\overline{z}_n$ must be equal except on an event of arbitrarily small size. In particular, it shows that $\sup_{k \geqslant n} ||\overline{z}_k - \widehat{z}_k||$ converges towards zero in probability, which is equivalent to almost sure converge of $\overline{z}_n - \widehat{z}_n$ towards zero. Claim (d) follows when remarking that (A.1) proves a.s. convergence of $\overline{z}_n$ towards $\widehat{z}$. □

Lemma 14. *As n tends to infinity, $\lambda_{2,n}$ converges in distribution to $\overline{\lambda}_2$.*

Proof. Let $\beta \in \mathbb{R}$ be such that $0 < \beta < 1/2$, say $\beta = 1/4$. Let $\varepsilon \in (0,1)$. We can write $\lambda_{2,n} = \lambda_{A,n} \vee \lambda_{B,n}$ with

$$\lambda_{A,n} := \max_{u \in G_n \setminus \{\widehat{z}_n\} \text{ s.t. } \|u-\widehat{z}_n\| \leqslant \Delta_n^\beta} X^{\widehat{z}_n}(u) =: \max_{u \in G_{n,A}} X^{\widehat{z}_n}(u),$$
$$\lambda_{B,n} := \max_{u \in G_n \text{ s.t. } \|u-\widehat{z}_n\| > \Delta_n^\beta} X^{\widehat{z}_n}(u) =: \max_{u \in G_{n,B}} X^{\widehat{z}_n}(u).$$

We first prove that $\lambda_{B,n} \to \lambda_2$ as n tends to infinity in distribution. By compactness, remark that there exists a constant $C_r > 0$ such that

$$1 - \rho(u) \geqslant C_r \|u\|^2. \tag{A.8}$$

It also holds that

$$X^{\widehat{z}_n}(u) = X(\widehat{z}_n) + \frac{X(u) - X(\widehat{z}_n)}{1 - \rho(u - \widehat{z}_n)}, \tag{A.9}$$
$$X^{\widehat{z}}(u) = X(\widehat{z}) + \frac{X(u) - X(\widehat{z})}{1 - \rho(u - \widehat{z})}. \tag{A.10}$$

Let us look to the rhs of (A.9) and (A.10). By Claim (d) of Lemma 13 and the continuous mapping theorem, note that $X(\widehat{z}_n)$ converges toward $\lambda_1 = X(\widehat{z})$ a.s. and we can omit these terms. It remains to prove that on $G_{n,B}$ the second terms are equivalent. Because of Lemma 13, $1 - \rho(u - \widehat{z}_n)$ converges to $1 - \rho(u - \overline{z}_n)$ at arbitrary speed. Remember that (A.1) gives $\widehat{z} - \overline{z}_n = O_P(\Delta_n)$ and it holds that $||u - \widehat{z}|| > \Delta_n^\beta$, on $G_{n,B}$. It follows that there exists $C > 0$ such that

$$1 - \rho(u - \widehat{z}_n) \geqslant C\Delta_n^{2\beta} \quad \text{and} \quad 1 - \rho(u - \widehat{z}) \geqslant C\Delta_n^{2\beta}, \tag{A.11}$$

with probability greater than $1 - \varepsilon/2$. As for the numerators, Eqs. (A.4) and (A.5) show that for all $u \in G_{n,B}$

$$\left| \frac{X(u) - X(\widehat{z}_n)}{X(u) - X(\widehat{z})} - 1 \right| = \left| \frac{X(\widehat{z}) - X(\widehat{z}_n)}{X(u) - X(\widehat{z})} \right| = (\text{cst}) \frac{|X(\widehat{z}) - X(\widehat{z}_n)|}{||u - \widehat{z}||^2} = O_P(\Delta_n^{2-2\beta})$$

In this sense, we say that $X(u) - X(\widehat{z}_n)$ is uniformly equivalent to $X(u) - X(\widehat{z})$ on the grid $G_{n,B}$ in probability. Using (A.11) and noticing that for any $u \in \mathbb{T}$

$$|\rho(u - \widehat{z}) - \rho(u - \widehat{z}_n)| \leqslant ||\rho'||_\infty ||\widehat{z} - \widehat{z}_n|| = O_P(\Delta_n)\,,$$

the same result holds for the denominators, namely $1 - \rho(u - \widehat{z}_n)$ is uniformly equivalent to $1 - \rho(u - \widehat{z})$ on the grid $G_{n,B}$ in probability. We deduce that $X^{\widehat{z}_n}(u)$ is uniformly equivalent to $X^{\widehat{z}_n}(u)$ on the grid $G_{n,B}$ in probability and, passing tho their maximum, one can deduce that $\lambda_{B,n}$ converges to λ_2 in probability.

We turn now to the study of the local part $\lambda_{A,n}$. Again, by Claim (c) of Lemma 13 we can replace $\widehat{z}_n$ by $\overline{z}_n$ in the numerator of the r.h.s in (A.9) and we forget the first term which limit is clearly λ_1 almost surely. We perform a Taylor expansion at $\widehat{z}$, it gives that

$$X(u) - X(\widehat{z}) = (1/2)(u - \widehat{z})^\top X''(u - \widehat{z})(1 + o_P(1)),$$

for any $u \in G_{n,A}$. Since $\overline{z}_n - \widehat{z} = O_P(\Delta_n)$, we also get that

$$X(\overline{z}_n) - X(\widehat{z}) = (1/2)(\overline{z}_n - \widehat{z})^\top X''(\overline{z}_n - \widehat{z}) + o_P(\Delta_n^2).$$

As for the denominator, invoke (A.1), (A.8) and Claim (c) of Lemma 13 to get that

$$\begin{aligned} 1 - \rho(u - \overline{z}_n) &\geqslant 2C_r\Delta_n^2, \\ 1 - \rho(u - \widehat{z}_n) &= 1 - \rho(u - \overline{z}_n) + o_P(\Delta_n^2), \\ \text{and} \quad 1 - \rho(u - \widehat{z}_n) &= (1/2)\big((u - \overline{z}_n)^\top \widetilde{\Lambda}(u - \overline{z}_n)\big)(1 + o_P(1)), \end{aligned}$$

where $-\widetilde{\Lambda}$ denotes the Hessian at point 0 of ρ. Putting all together yields

$$\frac{X(u) - X(\widehat{z}_n)}{1 - \rho(u - \widehat{z}_n)} = \frac{(u - \overline{z}_n)^\top X''(u + \overline{z}_n - 2\widehat{z})}{(u - \overline{z}_n)^\top \widetilde{\Lambda}(u - \overline{z}_n)}(1 + o_P(1)),$$

for any $u \in G_{n,A}$. Now we know that, in distribution, $\widehat{z} - \overline{z}_n = \Delta_n \mathcal{U}$ and we know that $u - \overline{z}_n = k\Delta_n$ with k belonging to a certain growing subset of $\mathbb{Z}^2$ which limit is $\mathbb{Z}^2$. Finally, conditionally to X'', we obtain that

$$\max_{u \in G_{n,A} \setminus \{\widehat{z}_n\}} X^{\widehat{z}_n}(u) \longrightarrow \lambda_1 + \sup_{k \in \mathbb{Z}^2 \setminus \{0\}} \frac{k^\top}{\|\widetilde{\Lambda}^{\frac{1}{2}} k\|} X'' \frac{(k - 2\mathcal{U})}{\|\widetilde{\Lambda}^{\frac{1}{2}} k\|},$$

in distribution. □

Eventually, consider the test statistic $S_n := \overline{\Phi}(\lambda_{1,n})/\overline{\Phi}(\lambda_{2,n})$ and keep in mind that $X\big(u + (0,\pi)\big) = -X(u)$ and that if u belongs to G_n, $\big(u + (0,\pi)\big)$ also belongs. So Theorem 1 of [10] applies showing that, under the alternative, $\mathbb{P}\{S_n \leqslant \alpha\} \geqslant \alpha$. It suffices to pass to the limit to get the desired result.

A.2. Proof of Theorem 2

We use the same grid argument as for the proof of Theorem 1.

Let $t_1, t_2, \ldots, t_N$ be pairwise distinct points of $[0, 2\pi)$, $\theta_1 \in [0, 2\pi)$, $m = 2N$ and set

$$z_1 = (t_1, \theta_1), \quad \ldots, \quad z_N = (t_N, \theta_1), \quad z_{N+1} = (t_1, \theta_1 + \pi/2), \quad \ldots, \quad z_m = (t_N, \theta_1 + \pi/2).$$

Because of the first assumption of ($\mathrm{ND}_Z(N)$), the distribution of $(X(z_1), \ldots, X(z_m))$ is non degenerated. Consequently, following the proof of Proposition 6, we know that X^{z_1} satisfies KL(m − 1) and ND(m − 1). Denote $g_1, \ldots, g_{m-1}$ the eigenfunctions of the Karhunen–Loève (KL) representation of $X^{(0,0)}$. Note that $X^{z_1}(\cdot)$ has the same distribution as $X^{(0,0)}(. - z_1)$ (stationarity) and that both are defined on the same space so the KL-eigenfunctions of X^{z_1} are $g_1(. - z_1), \ldots, g_{m-1}(. - z_1)$.

Now consider $A^{z_1} = (A^{z_1}_{i,j})_{1 \leq i,j \leq m-1}$ the matrix with entries $A^{z_1}_{i,j} = g_i(z_{j+1} - z_1)$ which is invertible thanks to KL(m − 1) and ND(m − 1) and build so that

$$\begin{pmatrix} X^{z_1}(z_2) \\ \vdots \\ X^{z_1}(z_m) \end{pmatrix} = A^{z_1} \begin{pmatrix} \zeta_1 \\ \vdots \\ \zeta_{m-1} \end{pmatrix}.$$

One possible explicit expression, among many others, of $\widehat{\sigma}^2_{\mathrm{KL}}(X^{\widehat{z}_n}_{\mathrm{norm}}(G_n))$, the estimator of σ^2 on the grid G_n, is

$$\widehat{\sigma}_n^2 := \widehat{\sigma}_{\mathrm{KL}}^2(X_{\mathrm{norm}}^{\widehat{z}_n}(G_n)) = \frac{1}{m-1}\left\| \left(A^{\widehat{z}_n}\right)^{-1} \begin{pmatrix} X^{(0,0)}(z_2-\widehat{z}_n) \\ \vdots \\ X^{(0,0)}(z_m-\widehat{z}_n) \end{pmatrix} \right\|_2^2,$$

which is a composition of continuous functions of $\widehat{z}_n$. In particular, as $\widehat{z}_n$ converges a.s. to $\widehat{z}$ (see Lemma 13, Claim (b)), we deduce that $\widehat{\sigma}_n^2$ converges a.s. to $\widehat{\sigma}_{\widehat{z}}^2$ as n goes to infinity.

Finally, since the KL estimator is unique, this estimator coincide with the estimator $\widehat{\sigma}_2^2$ of [10] and Theorem 3 of [10] implies that

$$\frac{\overline{F}_{m-1}(\lambda_{1,n}/\widehat{\sigma}_n)}{\overline{F}_{m-1}(\lambda_{2,n}/\widehat{\sigma}_n)} \sim \mathcal{U}([0,1]).$$

Note that λ_1^n converges almost surely to λ_1 and $\lambda_{2,n}$ converges in distribution to $\overline{\lambda}_2$ (see Lemma 14) to complete the proof.

A.3. Proof of Proposition 9

(a). We can assume that Z defined by (18) is centered and, in this case, it holds

$$\forall t \in [0,2\pi) \quad Z(t) = \frac{\sigma}{\sqrt{N}} \sum_{k=-f_c}^{f_c} \zeta_k \exp(\imath k t), \tag{A.12}$$

where we recall that $N = 2fc+1$ and $\zeta_k = \zeta_{k,1} + \imath\zeta_{k,2}$ for $k = -f_c, \dots, f_c$ are independent standard complex Gaussian variables. Formula (A.12) shows that Z satisfies ($\mathrm{KL}_Z(N)$).

(b). Let $(t_1, \dots, t_N) \in [0,2\pi)$ be pairwise differents, $\theta \in [0,2\pi)$ and set

$$\begin{pmatrix} Z(t_1) \\ \vdots \\ Z(t_N) \end{pmatrix} = \begin{pmatrix} \exp(-\imath f_c t_1) & \dots & \exp(\imath f_c t_1) \\ \vdots & & \vdots \\ \exp(-\imath f_c t_N) & \dots & \exp(\imath f_c t_N) \end{pmatrix} \begin{pmatrix} \zeta_1 \\ \vdots \\ \zeta_N \end{pmatrix} =: A_{t_1,\dots,t_N}\zeta,$$

where $A_{t_1,\dots,t_N}$ is a Vandermonde matrix, invertible as soon as $t_i \neq t_j$ for all $i \neq j$. This proves the first point of $\mathrm{ND}_Z(N)$. For the second assertion, consider $h > 0$ such that $h < \min_{1\le i<j\le N-1}(t_i - t_j)$ and the Gaussian vector

$$(Z(t_1), \dots, Z(t_{N-1}), Z(t_1+h))^T =: A_{t_1,\dots,t_{N-1},t_1+h}\zeta,$$

where the covariance matrix $A_{t_1,\dots,t_{N-1},t_1+h}$ satisfies

$$\begin{aligned} \det(A^*_{t_1,\dots,t_1+h}A_{t_1,\dots,t_1+h}) &= \prod_{1\le i<j\le N-1} |\exp(\imath t_i) - \exp(\imath t_j)|^2 \prod_{j=1}^{N-1} |\exp(\imath(t_1+h)) - \exp(\imath t_j)|^2 \\ &= 4^{N(N-1)/2} \prod_{1\le i<j\le N-1} \sin^2\left(\frac{t_i - t_j}{2}\right) \prod_{j=1}^{N-1} \sin^2\left(\frac{t_j - (t_1+h)}{2}\right) \\ &= \sin^2(h/2) \times g_{t_1,\dots,t_{N-1}}(h) \end{aligned}$$

where $g_{t_1,\dots,t_{N-1}}(0) \neq 0$ if $(t_i)_{1\le i\le N-1}$ are pairwise distincts. Finally, denote by R_h the linear transformation involving the first and the last coordinate such that

$$\begin{pmatrix} Z(t_1) \\ \vdots \\ Z(t_{N-1}) \\ \frac{Z(t_1+h)-Z(t_1)}{h} \end{pmatrix} = R_h \begin{pmatrix} Z(t_1) \\ \vdots \\ Z(t_{N-1}) \\ Z(t_1+h) \end{pmatrix}$$

and remark that

$$\lim_{h\to 0} \det(A^*_{t_1,\dots,t_1+h} R_h^* R_h A_{t_1,\dots,t_1+h}) = g_{t_1,\dots,t_{N-1}}(0) \times \lim_{h\to 0} \frac{\sin^2(h/2)}{h^2} = g_{t_1,\dots,t_{N-1}}(0) \times \frac{1}{4} \neq 0$$

giving the desired non degeneracy condition.

A.4. Proof of Proposition 10 and Proposition 11

Easy computations give the following results for $\phi(\cdot)$,

$$\int_\ell^{+\infty} \phi(t)\mathrm{d}t = \overline{\Phi}(\ell), \quad \int_\ell^{+\infty} t\phi(t)\mathrm{d}t = \phi(\ell), \quad \int_\ell^{+\infty} t^2\phi(t)\mathrm{d}t = \ell\phi(\ell) + \overline{\Phi}(\ell),$$

for $f_{m-1}(\cdot)$,

$$\begin{aligned}
&\int_\ell^{+\infty} f_{m-1}\left(t\sqrt{\frac{m-1}{m-3}}\right)\mathrm{d}t = \sqrt{\frac{m-3}{m-1}}\,\overline{F}_{m-1}\left(\ell\sqrt{\frac{m-1}{m-3}}\right), \\
&\int_\ell^{+\infty} t f_{m-1}\left(t\sqrt{\frac{m-1}{m-3}}\right)\mathrm{d}t = \frac{(m-3)\sqrt{m-3}}{(m-2)\sqrt{m-1}}\,\frac{\overline{\Gamma}\left(\frac{m}{2}\right)\overline{\Gamma}\left(\frac{m-3}{2}\right)}{\overline{\Gamma}\left(\frac{m-1}{2}\right)\overline{\Gamma}\left(\frac{m-2}{2}\right)}\, f_{m-3}(\ell), \\
&\int_\ell^{+\infty} t^2 f_{m-1}\left(t\sqrt{\frac{m-1}{m-3}}\right)\mathrm{d}t = \frac{(m-3)\sqrt{m-3}}{(m-2)\sqrt{m-1}}\,\frac{\overline{\Gamma}\left(\frac{m}{2}\right)\overline{\Gamma}\left(\frac{m-3}{2}\right)}{\overline{\Gamma}\left(\frac{m-1}{2}\right)\overline{\Gamma}\left(\frac{m-2}{2}\right)} \times \left(\ell f_{m-3}(\ell) + \overline{F}_{m-3}(\ell)\right),
\end{aligned}$$

and for R,

$$\begin{aligned}
X''(\widehat{z}) &= -\widetilde{\Lambda} X(\widehat{z}) + R(\widehat{z}), \\
&= -\begin{pmatrix} \alpha_1 & 0 \\ 0 & 1 \end{pmatrix} X(\widehat{z}) + \begin{pmatrix} -\alpha_2 & \alpha_3 \\ \alpha_3 & 0 \end{pmatrix},
\end{aligned}$$

where

$$\begin{cases}
\alpha_1 = \frac{1}{3} f_c(f_c+1), \\
\alpha_2 = \frac{1}{\sqrt{N}} \sum_{k=-f_c}^{f_c} (k^2 - \alpha_1) \times \mathrm{Re}(y_k e^{\imath(k\widehat{t} - \widehat{\theta})}), \\
\alpha_3 = \frac{1}{\sqrt{N}} \sum_{k=-f_c}^{f_c} k \times \mathrm{Re}(y_k e^{\imath(k\widehat{t} - \widehat{\theta})}).
\end{cases}$$

To conclude, use Proposition 9 to apply Theorem 4 and Theorem 8.

Appendix B. Auxiliary results

B.1. Regularity of $X^{|z}$ and new expression of $R(z)$

Lemma 15. *$X^{|z}(y)$ admits radials limits as $y \to z$. More precisely for all λ in the unit sphere*

$$\lim_{u\to 0} X^{|z}(z+u\lambda) = \frac{\lambda^\top R(z)\lambda}{\lambda^\top \widetilde{\Lambda}\lambda}.$$

Proof. As u tends to zero

$$1-\rho(u\lambda) = \frac{u^2}{2}(\lambda^\top\widetilde{\Lambda}\lambda + o(1)).$$

Moreover, a Taylor expansion gives

$$X(z+u\lambda) = X(z) + uX'_\lambda(z) + \frac{u^2}{2}X''_\lambda(z) + o_p(u^2),$$

and

$$\rho'_\lambda(u\lambda) = u\rho''_\lambda(0) + o_p(u^2) = -u\widetilde{\Lambda} + o_p(u^2),$$

where $(X'_\lambda, \rho'_\lambda)$ and $(X''_\lambda, \rho''_\lambda)$ are directional derivative and directional Hessian. By consequence,

$$\begin{aligned} X^{|z}(z+u\lambda) &= \frac{\frac{u^2}{2}X(z)\lambda^\top\widetilde{\Lambda}\lambda + \langle \rho'(u\lambda), \widetilde{\Lambda}^{-1}X'(z)\rangle + uX'_\lambda(z) + \frac{u^2}{2}X''_\lambda(z) + o_p(u^2)}{\frac{u^2}{2}(\lambda^\top\widetilde{\Lambda}\lambda + o(1))} \\ &= \frac{\frac{u^2}{2}\left(X(z)\lambda^\top\widetilde{\Lambda}\lambda + X''_\lambda(z) + o_p(1)\right)}{\frac{u^2}{2}(\lambda^\top\widetilde{\Lambda}\lambda + o(1))} \end{aligned}$$

which tends to

$$\frac{\lambda^\top\left(\widetilde{\Lambda}X(z) + X''(z)\right)\lambda}{\lambda^\top\widetilde{\Lambda}\lambda}$$

as u tends to 0 since $X''_\lambda(z) = \lambda^\top X''(z)\lambda$. The result follows from $X''(z) = -\widetilde{\Lambda}X(z) + R(z)$. □

B.2. Maximum of a continuous process

The following result is borrowed from [20, Theorem 3] and [21].

Proposition 16. *Let $\{Y(t)\,;\ t\in T\}$ be a Gaussian process with continuous sample paths defined on a compact metric space T. Suppose in addition that:*

$$\text{There is no two different points } s,t \in T \text{ such that } X(s) = X(t) \text{ a.s.} \tag{B.1}$$

Then almost surely the maximum of X on T is attained at a single point.

Observe that ($\mathbf{A}_{\text{norm}}$) implies (B.1).

Remark 8. Proposition 16 can be applied to the process $X^{|\widehat{z}}$ which is not continuous on a compact set. We use the "pumping method" as follows. Use

(a) a parameterization of $\mathbb{T}$ as $[0, 2\pi)^2$,
(b) polar coordinates for $y \in \mathbb{T} \setminus \{\widehat{z}\}$ with origin at $\widehat{z}$,
(c) the change of parameter

$$y = (\rho, \theta) \mapsto ((\rho + 1), \theta)$$

that transforms the non-compact set $\mathbb{T} \setminus \{\widehat{z}\}$ into a compact set (we have inflated the "hole" $\{\widehat{z}\}$ into a ball centered around $\widehat{z}$ with radius one) on which the process $X^{|\widehat{z}}$ is continuous thanks to Lemma 15.

References

[1] T. Hastie, R. Tibshirani, M. Wainwright, Statistical Learning with Sparsity: The Lasso and Generalizations, CRC Press, 2015.
[2] K. Bredies, H.K. Pikkarainen, Inverse problems in spaces of measures, ESAIM Control Optim. Calc. Var. 19 (01) (2013) 190–218.
[3] E.J. Candès, C. Fernandez-Granda, Towards a Mathematical Theory of Super-resolution, Comm. Pure Appl. Math. 67 (6) (2014) 906–956.
[4] C. Fernandez-Granda, Super-resolution of point sources via convex programming, Inf. Inference (2016) iaw005.
[5] V. Duval, G. Peyré, Exact support recovery for sparse spikes deconvolution, Found. Comput. Math. (2015) 1–41.
[6] G. Tang, B.N. Bhaskar, P. Shah, B. Recht, Compressed sensing off the grid, IEEE Trans. Inform. Theory 59 (11) (2013) 7465–7490.
[7] Y. De Castro, F. Gamboa, Exact reconstruction using beurling minimal extrapolation, J. Math. Anal. Appl. 395 (1) (2012) 336–354.
[8] J.-M. Azaïs, Y. De Castro, F. Gamboa, Spike detection from inaccurate samplings, Appl. Comput. Harmon. Anal. 38 (2) (2015) 177–195.
[9] R.J. Tibshirani, J. Taylor, R. Lockhart, R. Tibshirani, Exact post-selection inference for sequential regression procedures, J. Amer. Statist. Assoc. 111 (514) (2016) 600–620, https://doi.org/10.1080/01621459.2015.1108848.
[10] J.-M. Azaïs, Y. De Castro, S. Mourareau, Power of the Spacing test for Least-Angle Regression, Bernoulli.
[11] E. Giné, R. Nickl, Mathematical Foundations of Infinite-Dimensional Statistical Models, vol. 40, Cambridge University Press, 2015.
[12] S. Rosset, G. Swirszcz, N. Srebro, J. Zhu, ℓ_1 regularization in infinite dimensional feature spaces, Lecture Notes in Comput. Sci. 4539 (2007) 544.
[13] J.-M. Azaïs, M. Wschebor, Level Sets and Extrema of Random Processes and Fields, John Wiley & Sons Inc., 2009.
[14] R.J. Adler, J.E. Taylor, Random Fields and Geometry, Springer Science & Business Media, 2009.
[15] R. Lockhart, J. Taylor, R.J. Tibshirani, R. Tibshirani, A significance test for the lasso, Ann. Statist. 42 (2) (2014) 413.
[16] H. Cramer, M. Leadbetter, Stationary and Related Stochastic Processes, 1967.
[17] C. Fernandez-Granda, Support detection in super-resolution, in: 10th International Conference on Sampling Theory and Applications (SampTA 2013), Bremen, Germany, 2013, pp. 145–148.
[18] T. Bendory, S. Dekel, A. Feuer, Robust recovery of stream of pulses using convex optimization, J. Math. Anal. Appl. 442 (2) (2016) 511–536.
[19] M. Chassan, J.-M. Azais, Discretization Error for the Maximum of a Gaussian Field, Tech. rep., HAL, working paper or preprint, 2017, https://hal.archives-ouvertes.fr/hal-01513724.
[20] M.A. Lifshits, On the absolute continuity of distributions of functionals of random processes, Theory Probab. Appl. 27 (3) (1983) 600–607.
[21] V. Tsirel'Son, The density of the distribution of the maximum of a Gaussian process, Theory Probab. Appl. 20 (4) (1976) 847–856.