

# Second Maximum of a Gaussian Random Field and Exact ( $t$ -)Spacing test

Jean-Marc Azaïs<sup>•</sup>, Federico Dalmao<sup>◦</sup> and Yohann De Castro<sup>\*†</sup>

<sup>•</sup>*Institut de Mathématiques de Toulouse  
Université Paul Sabatier, France*

<sup>◦</sup>*DMEL, CENUR Litoral Norte  
Universidad de la República, Salto, Uruguay*

<sup>\*</sup>*Institut Camille Jordan, CNRS UMR 5208  
École Centrale Lyon, France*

<sup>†</sup>*Institut Universitaire de France (IUF)*

**Abstract:** In this article, we introduce the novel concept of the second maximum of a Gaussian random field on a Riemannian submanifold. This second maximum serves as a powerful tool for characterizing the distribution of the maximum. By utilizing an ad-hoc Kac—Rice formula, we derive the explicit form of the maximum’s distribution, conditioned on the second maximum and some regressed component of the Riemannian Hessian. This approach results in an exact test, based on the evaluation of spacing between these maxima, which we refer to as the spacing test.

We investigate the applicability of this test in detecting sparse alternatives within Gaussian symmetric tensors, continuous sparse deconvolution, and two-layered neural networks with smooth rectifiers. Our theoretical results are supported by numerical experiments, which illustrate the calibration and power of the proposed tests. More generally, this test can be applied to any Gaussian random field on a Riemannian manifold, and we provide a general framework for the application of the spacing test in continuous sparse kernel regression.

Furthermore, when the variance-covariance function of the Gaussian random field is known up to a scaling factor, we derive an exact Studentized version of our test, coined the  $t$ -spacing test. This test is perfectly calibrated under the null hypothesis and has high power for detecting sparse alternatives.

**MSC2020 subject classifications:** Primary 62E15, 62F03, 60G15, 62H10, 62H15; secondary 60E05, 60G10, 62J05, 94A08.

**Keywords and phrases:** Gaussian random field, Second maximum, Spacing test, Kernel regression, Tensor PCA.

*Version of July 8, 2024*

## 1. Introduction

### 1.1. The ( $t$ -)spacing test and the second maximum

This paper introduces a test for the mean  $m(\cdot)$  of a Gaussian random field  $Z(\cdot)$ , defined on a  $\mathcal{C}^2$ -compact Riemannian manifold  $M$  of dimension  $d$  without boundary, which can be decomposed as

$$\forall t \in M, \quad Z(t) = m(t) + \sigma X(t), \quad (1.1)$$

where  $m(\cdot)$  is any  $\mathcal{C}^2$ -function,  $\sigma > 0$  is the standard deviation, and  $X(\cdot)$  is a centered Gaussian random field such that  $\text{Var}[X(t)] = 1$  for every  $t \in M$ . We consider the simple global null hypothesis  $\mathbb{H}_0 : “m(\cdot) = 0”$ , even when the standard error  $\sigma$  is unknown. The test investigated is the quantile  $\hat{\alpha}$  of the maximum  $\lambda_1$  under the null defined as

$$\lambda_1 = \max_{t \in M} \{\sigma X(t)\} \quad (1.2)$$

and we denote by  $t_1 \in M$  its argument maximum. This quantile is built from the distribution of  $\lambda_1$  conditional to the values of the so-called second maximum  $\lambda_2$  and the so-called independent part of the Riemannian Hessian  $\Omega$ , under the null. Suppose, in a first step, that  $\sigma$  is known, in such a case we will demonstrate that the cumulative distribution of this conditional law can

be expressed as a ratio, resulting in the following expression derived from an ad-hoc Kac-Rice formula,

$$\hat{\alpha} := \frac{\int_{\lambda_1/\sigma}^{+\infty} \det(u\text{Id} - \Omega/\sigma)\varphi(u) du}{\int_{\lambda_2/\sigma}^{+\infty} \det(u\text{Id} - \Omega/\sigma)\varphi(u) du}, \quad (1.3)$$

where  $\varphi(\cdot)$  is the standard Gaussian density.

We will show the exactness of this test, meaning that  $\hat{\alpha}$  is uniformly distributed on the interval  $(0, 1)$  under the null hypothesis. It demonstrates that one minus the ratio (1.3) represents the cumulative distribution function (CDF) of the law of  $\lambda_1$  conditional to  $\lambda_2$  and  $\Omega$ , under the null hypothesis. Small values of  $\hat{\alpha}$  indicate that the maximum  $\lambda_1$  is abnormally large compared to the value of the second maximum  $\lambda_2$ , making  $\hat{\alpha}$  interpretable as a  $p$ -value. This test is referred to as the “spacing test”. The second maximum is defined as

$$\lambda_2 \in \arg \max_{s \neq t_1 \in M} \{\sigma X^{|t_1}(s)\}, \quad (1.4)$$

$$\text{with, for } s \neq t, X^{|t}(s) := \frac{X(s) - c(s, t)X(t) - \nabla_t c(s, t)^\top \Lambda_2^{-1}(t) \nabla X(t)}{1 - c(s, t)}$$

where  $c(\cdot, \cdot)$  is the variance-covariance function of  $X(\cdot)$  and  $\Lambda_2(t) := \text{Var}[\nabla X(t)]$ , the formal definition will be given in Section 2.1. Note that  $X^{|t}(s)$  is a normalisation of the remainder of the regression of  $X(s)$  with respect to  $(X(t), \nabla X(t))$ . An example of the second maximum in the rank-one tensor detection case, referred to as Spiked tensor PCA, is illustrated in Figure 1 and developed in Section 5.2.

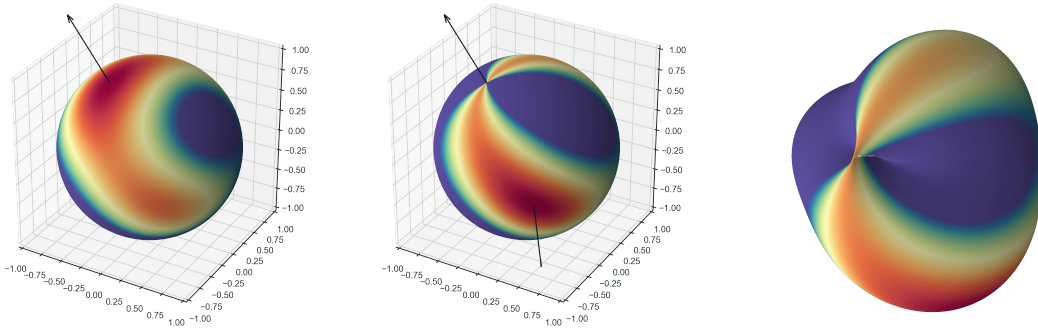


FIGURE 1. [Spiked tensor PCA, Section 5.2, example 1/4] The first eigenvector (resp. first eigenvalue) of a Gaussian symmetric tensor (arrow in the left panel) is the arg maximum (resp. maximum,  $\lambda_1$ ) of some Gaussian homogeneous polynomial  $X(\cdot)$ . The second eigenvector (resp. second maximum,  $\lambda_2$ ), represented by an arrow in the middle panel, is the arg maximum of  $X^{|t_1}(\cdot)$ , some regression of  $X(\cdot)$ . A volumic view of  $X^{|t_1}(\cdot)$  is given in the right panel, the height of the surface around the sphere is given by the value of the random field. We witness that a singularity, referred to as the helix, appears at point  $t_1$ . A thoroughly study of this helix will be given in this paper.

Furthermore, we will show how to achieve the same result when the variance  $\sigma^2$  is unknown using an estimation  $\hat{\sigma}^2$  built from the Karhunen-Loève expansion of  $X^{|t_1}(\cdot)$ , see the right panel of Figure 1 for an illustration of this random field in the case of rank-one tensor detection. We refer to this test as the  $t$ -spacing test which detects abnormally large spacing between  $\lambda_1/\hat{\sigma}$  and  $\lambda_2/\hat{\sigma}$ . Note that  $\lambda_1$  and  $\lambda_2$  are the same for the spacing and the  $t$ -spacing test, and these values can be computed without knowing  $\sigma$ . These results are supported by numerical experiments as illustrated in Figure 2.

## 1.2. Detecting one sparse alternatives in continuous sparse kernel regression

Our test is perfectly calibrated for all  $\mathcal{C}^2$ -mean  $m(\cdot)$ . But, we expect this test to have high power on  $s$ -sparse alternatives of the form

$$m(\cdot) = \sum_{k=1}^s \lambda_{0,k} c(\cdot, t_{0,k}),$$

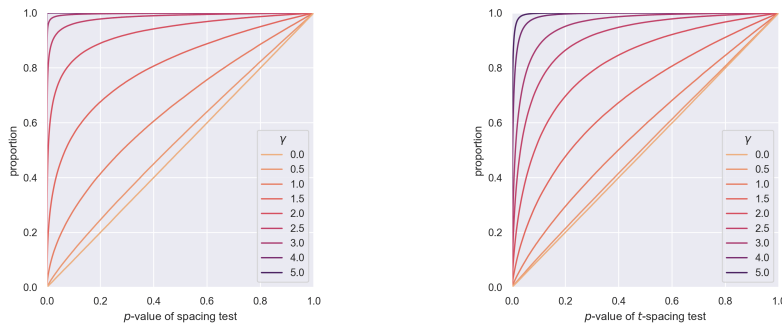


FIGURE 2. [Spiked tensor PCA, Section 5.2, example 2/4] The CDFs of the p-value of ( $t$ -)spacing tests over 250,000 Monte-Carlo samples for each value of  $\gamma$ . Note that these tests are perfectly calibrated under the null hypothesis, for which  $\gamma = 0$ . The parameter  $\gamma$  is a scaling factor of eigenvalue of the rank-one tensor to be detected. The value  $\gamma = 1$  corresponds to the so-called phase transition in Spiked tensor PCA as presented in Perry et al. (2020, Theorem 1.3). In the  $t$ -spacing test, the variance  $\sigma$  has been estimated on  $X^{\lfloor t_1 \rfloor}(\cdot)$ .

and we denote one-sparse alternatives by

$$\mathbb{H}_1(t_0, \lambda_0) : "m(\cdot) = \lambda_0 c(\cdot, t_0)", \quad (1.5)$$

for some  $\lambda_0 \neq 0$  and  $t_0 \in M$  unknown. Indeed, we will show that the maxima  $\lambda_1$  and  $\lambda_2$  correspond to the so-called knots of the Continuous LARS algorithm and are then related to continuous kernel sparse regression, see Section 1.3.3. We illustrate this phenomenon in Figure 3 in the case of Spiked tensor PCA.

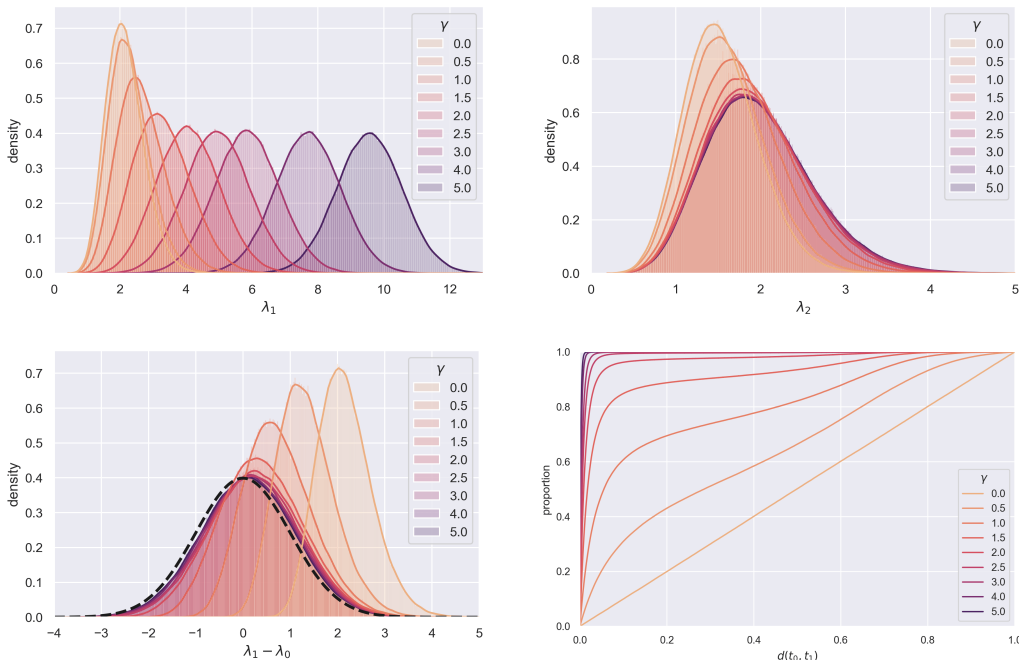


FIGURE 3. [Spiked tensor PCA, Section 5.2, example 3/4] The PDFs of the maxima  $\lambda_1, \lambda_2$  and the CDF of the distance  $d(t_0, t_1)$  over 250,000 Monte-Carlo samples for each value of the parameter  $\gamma$ . The alternative is given by  $t_0$  fixed and  $\lambda_0 = \gamma \times \sigma \sqrt{3 \log 3 + 3 \log \log 3} \simeq 0.684 \gamma \sigma$  where  $\gamma = 1$  corresponds to the so-called phase transition in Spiked tensor PCA as presented in Perry et al. (2020, Theorem 1.3). The distance  $d(t_0, t_1)$  is normalized so that it is uniformly distributed on  $(0, 1)$  if  $t_1$  is uniformly distributed on the sphere (e.g.,  $\gamma = 0$ ).

In Figure 3, the alternative is given by  $\lambda_0 = \gamma \times \sigma \sqrt{3 \log 3 + 3 \log \log 3}$  where  $\gamma = 1$  corresponds to the so-called phase transition in Spiked tensor PCA as presented in Perry et al. (2020, Theorem 1.3). The top left panel shows that  $\lambda_1$  is stochastically increasing as  $\gamma$  increases while the top right panel shows that the distribution of  $\lambda_2$  remains unchanged for moderate values of  $\gamma$ , say  $\gamma \geq 2$ . It illustrates that the spacing between  $\lambda_1$  and  $\lambda_2$  grows linearly with  $\gamma$ . In the

moderate regime, the bottom left panel shows that  $\lambda_1$  is distributed as a Gaussian with mean  $\lambda_0$  and variance  $\sigma^2$  (dashed black line), which is the distribution of  $\sigma X(t_0)$  under  $\mathbb{H}_1(t_0, \lambda_0)$ . The bottom right panel shows that  $t_1 \simeq t_0$ , for moderate values of  $\gamma$ . It illustrates that  $(t_1, \lambda_1)$  is weakly close to the distribution of  $(t_0, \sigma X(t_0))$ , the  $(t)$ -spacing tests detect the alternative  $\mathbb{H}_1(t_0, \lambda_0)$ . We leave the proof of this phenomenon for future work.

### 1.3. A general framework related to continuous kernel regression

#### 1.3.1. The four conditions on the Gaussian random field

We start by giving the assumptions of the Gaussian random field  $X(\cdot)$  that we will invoke along this paper. We consider a real valued Gaussian random field  $X(\cdot)$  defined on a  $\mathcal{C}^2$  compact Riemannian manifold  $M$  of dimension  $d$  without boundary. We recall that  $c(\cdot, \cdot) : M \times M \rightarrow \mathbb{R}$  denotes its variance-covariance function.

**Assumption (A<sub>1-4</sub>)** We assume that

- the paths of  $X(\cdot)$  are  $\mathcal{C}^2$  almost surely; (A<sub>1</sub>)
- $\mathbb{E}[X(t)] = 0$  and  $\text{Var}[X(t)] = 1$ , for every  $t \in M$ ; (A<sub>2</sub>)
- $\forall s \neq t, c(s, t) < 1$ ; (A<sub>3</sub>)
- for every  $t \in M$ , the gradient  $\nabla X(t)$  has a non-degenerate Gaussian distribution. (A<sub>4</sub>)

The variance-covariance matrix of the Gaussian tangent vector  $\nabla X(t)$  is denoted by  $\Lambda_2(t)$ , which is invertible by Assumption (A<sub>4</sub>).

#### 1.3.2. A continuous regression framework for the spacing test

**General framework** We consider the following regression problem given by

$$Y = \lambda_0 \psi_{t_0} + \sigma W, \quad (1.6)$$

where  $\lambda_0 \in \mathbb{R}$ ,  $t_0 \in M$  are unknown parameters;  $\sigma$  is the standard deviation;  $(E, \langle \cdot, \cdot \rangle_E)$  is any Euclidean space;  $\psi : t \in M \mapsto \psi_t \in E$ ;  $W \in E$  is a centered Gaussian vector with variance-covariance matrix  $\text{Id}_E$ . The feature map  $\psi$  satisfies the following Assumptions (B<sub>1-6</sub>):

- $\psi$  is a  $\mathcal{C}^3$ -function; (B<sub>1</sub>)
- $\forall t \in M, \|\psi_t\|_E^2 = 1$ ; (B<sub>2</sub>)
- $\forall t \in M, \forall s \in M \setminus \{t\}, \langle \psi_s, \psi_t \rangle_E < 1$ ; (B<sub>3</sub>)
- $\forall t \in M, J\psi_t J\psi_t^\top$  has full rank  $d$ ; (B<sub>4</sub>)
- the span of  $\{\psi_t : t \in M\}$  has dimension  $m$  such that  $m > d + 1$ ; (B<sub>5</sub>)
- $\forall t \in M, \exists s \in M$  s.t.  $\psi_t = -\psi_s$ ; (B<sub>6</sub>)

where  $J\psi_t$  is the Jacobian matrix of  $\psi_t$  and  $J\psi_t^\top$  its transpose. In the next paragraph, we will see that  $X(t) = \langle W, \psi_t \rangle_E$  and

$$\forall s, t \in M, c(s, t) = \langle \psi_s, \psi_t \rangle_E.$$

One can check that (B<sub>1</sub>) implies (A<sub>1</sub>), (B<sub>2</sub>) is equivalent to (A<sub>2</sub>), (B<sub>3</sub>) is equivalent to (A<sub>3</sub>), (B<sub>4</sub>) is equivalent to (A<sub>4</sub>), and (B<sub>5</sub>) is equivalent to the forthcoming Assumption (A<sub>5</sub>).

**Remark 1** (Spiked tensor PCA, Section 5.2). *In Figures 1, 2 and 3, we considered, as an example, the detection of a rank one 3-way symmetric tensor. We consider the sphere  $M = \mathbb{S}^2$ . The Euclidean space of 3-way symmetric tensors of size  $3 \times 3 \times 3$  is denoted by  $(E, \langle \cdot, \cdot \rangle_E)$  with dimension  $m = 10$ , and*

$$\begin{aligned} \psi_t &:= t^{\otimes 3} = (t_i t_j t_k)_{1 \leq i, j, k \leq 3}, \\ X(t) &:= \langle W, \psi_t \rangle_E = \sum_{ijk} t_i t_j t_k W_{ijk}, \end{aligned}$$

where  $W$  is an order 3 symmetric tensor defined using symmetry and the following independent terms,

- $W_{iii}$  of variance 1; there are 3 principal diagonal terms,  $i \in [3]$ ;
- $W_{iij}$  of variance 1/3; there are 6 sub-diagonals terms,  $1 \leq i \neq j \leq 3$ ;
- $W_{123}$  of variance 1/6; there is 1 off-diagonal term.

**Maximum likelihood estimator (MLE)** The Maximum Likelihood Estimator (MLE) of the couple  $(\lambda_0, t_0)$  is given by

$$\arg \min_{(\lambda, t)} \left\{ \frac{1}{2} \|Y - \lambda \psi_t\|_E^2 \right\}.$$

Minimization with respect to  $\lambda$  is straightforward under Assumptions **(B<sub>2</sub>)** and **(B<sub>6</sub>)** leading to

$$\arg \min_t \min_{\lambda} \left\{ \frac{1}{2} \|Y - \lambda \psi_t\|_E^2 \right\} = \arg \max_t \langle Y, \psi_t \rangle_E. \quad (1.7)$$

Now consider the Gaussian random field  $Z(t) := \langle Y, \psi_t \rangle_E$ , referred to as the profile likelihood. The argument maximum (resp. the maximum) of  $Z(\cdot)$  is exactly the MLE of  $t_0$  (resp. of  $\lambda_0$ ) by (1.7), namely

$$\lambda^{\text{MLE}} = \max Z \quad \text{and} \quad t^{\text{MLE}} = \arg \max Z.$$

By (1.6), note that we uncover the decomposition (1.1) and the alternative hypothesis (1.5) as

$$Z(t) = \langle Y, \psi_t \rangle_E = \lambda_0 \langle \psi_{t_0}, \psi_t \rangle_E + \sigma \langle W, \psi_t \rangle_E = \lambda_0 c(t_0, \cdot) + \sigma X(t), \quad (1.8)$$

where  $X(t) := \langle W, \psi_t \rangle_E$  and  $m(\cdot) = \lambda_0 c(t_0, \cdot)$ .

**Remark 2.** Assumption **(B<sub>6</sub>)** ensures that the profile likelihood is given by  $Z(\cdot)$ , see (1.7).

### 1.3.3. Parallel with the LARS algorithm

**Discrete LARS** The Least Angle Regression (LARS) algorithm has been introduced in Efron et al. (2004) and has been widely used in statistics for variable selection. Given an observation  $Y \in \mathbb{R}^n$  and  $p$  covariates  $(\psi_t)_{t=1}^p$ , the LARS algorithm is a forward stepwise variable selection algorithm giving a sequence  $(\lambda_i, t_i)$  of the so-called knots. The first knot  $(\lambda_1, t_1)$  is the maximum and the argument maximum of the maximal absolute correlation between the observation and the covariates. Hence, the first step of the LARS aims at maximizing

$$\lambda_1 = \max_{1 \leq t \leq p} |\langle Y, \psi_t \rangle_{\mathbb{R}^n}| \quad \text{and} \quad t_1 = \arg \max_{1 \leq t \leq p} |\langle Y, \psi_t \rangle_{\mathbb{R}^n}|. \quad (1.9)$$

We define the so-called residual by, for all  $\lambda \leq \lambda_1$ ,

$$\forall t \in \{1, \dots, p\}, \quad Z^\lambda(t) := \langle Y - (\lambda_1 - \lambda) \psi_{t_1}, \psi_t \rangle_{\mathbb{R}^n}, \quad (1.10)$$

The second knot is defined by

$$\lambda_1 - \lambda_2 := \inf \{ \varepsilon > 0 : \exists t \neq t_1 \text{ s.t. } Z^{\lambda_1 - \varepsilon}(t) \geq \lambda_1 - \varepsilon \}. \quad (1.11)$$

The second knot  $(\lambda_2, t_2)$  is built so that  $t_1$  is the unique argument maximum of  $Z^\lambda(\cdot)$  for  $\lambda_2 < \lambda \leq \lambda_1$  and that there exists a second point, say  $t_2 \neq t_1$ , such that  $Z^{\lambda_2}(t_1) = Z^{\lambda_2}(t_2) = \lambda_2$ .

**Continuous LARS** The continuous LARS has been investigated in Azaïs et al. (2020) for continuous sparse regression from Fourier measurements, see Section 5.4 for further details. We introduce the continuous LARS in a general setting as follows. The first knot  $(\lambda_1, t_1)$  is the maximum and the argument maximum of the maximal absolute correlation between the observation  $Y \in E$  and  $\psi_t \in E$ , the features. Hence, the first step of the LARS aims at maximizing

$$\lambda_1 = \max_{t \in M} |\langle Y, \psi_t \rangle_E| \quad \text{and} \quad t_1 = \arg \max_{t \in M} |\langle Y, \psi_t \rangle_E|. \quad (1.12)$$

Under **(B<sub>6</sub>)**, note that  $\max_{t \in M} |Z(t)| = \max_{t \in M} Z(t)$ , where  $Z(t) = \langle Y, \psi_t \rangle_E$ . Therefore, the definition of  $\lambda_1$  given in (1.2) coincides with the aforementioned definition of  $\lambda_1$ . One can choose  $t_1$

being the unique argument maximum of  $Z(\cdot)$ , by **(B<sub>1</sub>)** and **(B<sub>3</sub>)** (see also (2.4) in Lemma 1). We define the so-called residual by, for all  $\lambda \leq \lambda_1$ ,

$$Z^\lambda(t) := \langle Y - (\lambda_1 - \lambda)\psi_{t_1}, \psi_t \rangle_E = Z(t) - (\lambda_1 - \lambda)c(t, t_1), \quad (1.13)$$

where  $c(s, t) = \langle \psi_s, \psi_t \rangle_E$ . Again, under **(B<sub>6</sub>)**, note that  $\max_{t \in M} |Z^\lambda(t)| = \max_{t \in M} Z^\lambda(t)$ . Moreover, one can check that  $Z^\lambda(t_1) = \lambda$ . The second knot is defined by

$$\lambda_1 - \lambda_2 := \inf \{ \varepsilon > 0 : \exists t \neq t_1 \text{ s.t. } Z^{\lambda_1 - \varepsilon}(t) \geq \lambda_1 - \varepsilon \}. \quad (1.14)$$

The second knot  $(\lambda_2, t_2)$  is built so that  $t_1$  is the unique argument maximum of  $Z^\lambda(\cdot)$  for  $\lambda_2 < \lambda \leq \lambda_1$  and that there exists a second point, say  $t_2 \neq t_1$ , such that  $Z^{\lambda_2}(t_1) = Z^{\lambda_2}(t_2) = \lambda_2$ . Now observe that, for all  $t \neq t_1 \in M$  and all  $\lambda$ ,

$$Z^\lambda(t) \leq \lambda \Leftrightarrow Z(t) - \lambda_1 c(t, t_1) \leq \lambda(1 - c(t, t_1)) \Leftrightarrow \frac{Z(t) - Z(t_1)c(t, t_1)}{1 - c(t, t_1)} \leq \lambda \Leftrightarrow Z^{t_1}(t) \leq \lambda,$$

where  $Z^{t_1}(t)$  is defined as in (1.4), taking into account that the gradient is vanishing at point  $t_1$ . We uncover that  $(\lambda_2, t_2)$  is the second knot of the LARS where

$$\lambda_2 = \max_{t \in M \setminus \{t_1\}} Z^{t_1}(t) \quad \text{and} \quad t_2 = \max_{t \in M \setminus \{t_1\}} Z^{t_1}(t). \quad (1.15)$$

Note that the definition of  $\lambda_2$  in (1.4) is equivalent to the aforementioned definition of  $\lambda_2$  (this is also true for the argument maximum). This paper gives the distribution of  $\lambda_1$  (first knot of the continuous LARS) conditional to  $(\lambda_2, \Omega)$  (second knot of the continuous LARS and independent part of the Hessian at the first knot) under the null hypothesis.

#### 1.4. Contributions and outline

This paper investigates the notion of second maximum  $\lambda_2$  of a Gaussian random field. The definition of second maximum originates from continuous LARS, see Section 1.3.3, and can be used in a Kac-Rice formula to compute the conditional law of the maximum  $\lambda_1$  with respect to  $\lambda_2$  and the so-called independent part of the Hessian  $\Omega$ . The second maximum is the maximum of a random field  $\sigma X^{t_1}(\cdot)$  with a singularity at point  $t_1$ . A first contribution establishes a link between the eigen-decomposition of  $\Omega$  and the directional limiting values of  $\sigma X^{t_1}(\cdot)$  at the singularity point  $t_1$ . see Lemma 3. The second contribution is a Kac-Rice type formula, see Appendix B, that gives the conditional law in an explicit form, see Theorem 1. From this point, we derive an exact test based on the spacing between  $\lambda_1$  and  $\lambda_2$ , see Theorem 2. We introduce an estimation of the variance  $\hat{\sigma}$  in Section 4.2, that can be used in a Kac-Rice formula and lead to the conditional law of  $\lambda_1/\hat{\sigma}$  in Proposition 7. We derive an exact test based on the spacing between  $\lambda_1/\hat{\sigma}$  and  $\lambda_2/\hat{\sigma}$  which can be used when the variance  $\sigma^2$  is unknown, see Theorem 3. The reader may consult <https://github.com/ydecastro/tensor-spacing/> for further details on the numerical experiments of this paper.

The paper is organized as follows. Section 2 introduces the second maximum of a Gaussian random field. Section 2.3 gives an ad-hoc Kac-Rice formula for the conditional law of  $\lambda_1$  with respect to  $\lambda_2$  and  $\Omega$ . Section 3 gives the exact tests based on the spacing between  $\lambda_1$  and  $\lambda_2$  and the spacing between  $\lambda_1/\hat{\sigma}$  and  $\lambda_2/\hat{\sigma}$ . Section 4.4 introduces the estimation of the variance  $\hat{\sigma}$  from the Karhunen-Loève expansion of  $X^{t_1}(\cdot)$ . Section 5 gives the parallel with continuous sparse kernel regression.

#### 1.5. Related works

The first paper which considers Kac-Rice formula on manifolds, in fact the sphere and the Stiefel Manifold, is Azaïs and Wschebor (2004). The subject is developed in Adler and Taylor (2009) and Azaïs and Wschebor (2009). The article Auffinger and Ben Arous (2013) considers the high dimensional sphere. A new set of hypotheses and proofs is also given in Armentano et al. (2023).

The Kac-Rice formula is also used in Azaïs et al. (2017) for the study of the maximum of a Gaussian random field on the torus with application to the so-called Super-Resolution problem. It

falls under the general theory of continuous sparse regression over the space of measures which has recently attracted a lot of attention in signal processing (Candès and Fernandez-Granda, 2014; Duval and Peyré, 2015) among the super-resolution community; in machine learning (De Castro et al., 2021); and in optimization (Chizat, 2022). The super-resolution framework aims to recover fine scale details of an image from just a few low frequency measurements, where ideally the observation is given by a low-pass filter. It has potential applications in astronomy, medical imaging, and microscopy. The novel aspects of this body of work rely on new statistical and optimization guarantees of sparse regression over the space of measures. Initiated by the work presented in (Azaïs et al., 2020), one investigate the possibility of detecting a sparse object from a test on the mean  $m(\cdot)$  of a Gaussian random field under a sparsity-type assumption.

*General notation*

$[a]$	Set of integers $\{1, \dots, a\}$ ;
$A^\top$ (resp. $A_i$ .)	transpose of a matrix $A$ (resp. $i$ -th line);
Id	Identity matrix of dimension $d \times d$ ;
$\mathcal{S}_d$	Space of $d \times d$ symmetric matrices;
$\mathbb{S}^{d-1}$	Euclidean sphere of $\mathbb{R}^d$ ;
$t$ (and $s$ )	Generic value for a vector on the manifold $M$ ;
$\mathcal{C}^k$	Set of $k$ times differentiable functions;
$\delta_{jk}$	Kronecker symbol;
(const)	Positive constant which may change from line to line;
$\mathcal{D}(U V)$	Conditional distribution of $U$ with respect to $V$ ;
$\text{Var}(U)$	Variance matrix of a random vector $U$ ;
$\text{Cov}(U, V)$	Covariance matrix of the random vectors $U$ and $V$ ;
$\varphi(\cdot)$	Standard Gaussian density in $\mathbb{R}$ ;
$p_U(\ell)$	Density function of the random variable/vector $U$ at point $\ell$ ;
$\mathbb{1}_A$ (resp. $\mathbb{1}\{A\}$ )	Indicator function of condition $A$ (resp. event $A$ );
Leb	Lebesgue measure on $\mathbb{R}$ ;
$\mathcal{U}(0, 1)$	Uniform law on $(0, 1)$ ;
$\mu \otimes \nu$	Product of measures;
<b>Random fields</b>	
$Z(\cdot)$	Observed Gaussian random field, indexed by $M$ ;
$m(\cdot)$	Mean function of $Z(\cdot)$ ;
$\sigma$	Standard error of the $Z(\cdot)$ ;
$X(\cdot)$	Centered Gaussian random field with unit variance;
$c(s, t)$	Covariance function of $X(\cdot)$ , $c(s, t) = \mathbb{E}(X(s)X(t))$ ;
<b>Differential geometry</b>	
$M$	$\mathcal{C}^2$ -compact Riemannian manifold of dimension $d$ without boundary;
$\nabla f(t)$	Riemannian gradient at point $t \in M$ of $f : M \rightarrow \mathbb{R}$ ;
$\nabla_t g(s, t)$	Riemannian gradient at point $t \in M$ of $g(s, \cdot)$ ;
$\nabla^2 f(t)$	Riemannian Hessian at point $t \in M$ of $f : M \rightarrow \mathbb{R}$ ;
<b>Continuous regression</b>	
$(E, \langle \cdot, \cdot \rangle_E)$	Euclidean space;
$\psi$	Feature map from $M$ to $E$ , $\psi : t \in M \mapsto \psi_t \in E$ ;
$\psi_t$	Vector $\psi(t) \in E$ ;
$J\psi_t$ (resp. $J\psi_t^\top$ )	(resp. transpose of) Jacobian matrix of $\psi$ at point $t$ ;
$W$	Standard Gaussian random vector of $E$ ;
$\mathbf{S}_\sigma$ (resp. $\mathbf{T}$ )	Generalized spacing test (resp. Generalized $t$ -spacing test);

TABLE 1

*List of notation: Riemannian gradients and Hessians are defined using the Levi-Civita connection. Riemannian Hessians are represented in matrix form.*

## 2. The second maximum of a Gaussian random field

### 2.1. The two Gaussian regression remainders

Some useful properties are presented in the next lemma.

**Lemma 1.** *Under Assumption (A<sub>1-4</sub>), one has*

$$\circ c(s, t) = \mathbb{E}[X(s) X(t)] \text{ and } c(t, t) = 1; \quad (2.1)$$

$$\circ X(t) \text{ and } \nabla X(t) \text{ are independent}; \quad (2.2)$$

$$\circ \Lambda_2(t) := \text{Var}[\nabla X(t)] = -\text{Cov}[\nabla^2 X(t), X(t)] \quad (2.3)$$

$$\circ \text{the argument maximum of } X(\cdot) \text{ is unique.} \quad (2.4)$$

*Proof.* The first property (2.1) is equivalent to (A<sub>2</sub>). The other properties can be deduced by differentiating  $c(t, t) = 1$ . The last statement is a consequence of Tsirelson's theorem, see for instance Lifshits (1983, Theorem 3).  $\square$

**The Gaussian random field  $X^{|t}(\cdot)$**  We know that  $X(t)$  and  $\nabla X(t)$  are independent by (2.2). For a fixed point  $t \in M$ , consider the remainder of Gaussian regression of  $X(s)$  with respect to  $(X(t), \nabla X(t))$  given by the Gaussian random field

$$s \in M \mapsto X(s) - c(s, t) X(t) - \nabla_t c(s, t)^\top \Lambda_2^{-1}(t) \nabla X(t) \in \mathbb{R},$$

where  $\nabla_t c(s, t)$  is the Riemannian gradient of  $t \mapsto c(s, t)$ . By (A<sub>4</sub>), remark that  $\Lambda_2(t)$  is invertible. This Gaussian random field is well defined on  $M$  and independent of  $(X(t), \nabla X(t))$ . Now, for  $s \neq t$ , set

$$X^{|t}(s) := \frac{X(s) - c(s, t) X(t) - \nabla_t c(s, t)^\top \Lambda_2^{-1}(t) \nabla X(t)}{1 - c(s, t)}, \quad (2.5)$$

that is,  $X^{|t}(s)$  is a normalisation of the remainder of the regression of  $X(s)$  w.r.t.  $(X(t), \nabla X(t))$ .

**The regression of the Hessian  $R(t)$**  Now, consider the following regression in the space of Gaussian symmetric matrices

$$\nabla^2 X(t) = -\Lambda_2(t) X(t) - \Lambda_3(t) \nabla X(t) + \tilde{R}(t), \quad (2.6)$$

for some well-defined 3-way tensor  $\Lambda_3(t)$ . It will not be necessary to give the explicit expression of  $\Lambda_3(t)$  for our purposes, this tensor is well defined by Gaussian regression formulas. Thus, one can identify the symmetric matrix  $\tilde{R}(t)$  as the remainder of the regression of  $\nabla^2 X(t)$  on  $(X(t), \nabla X(t))$ . For future use, it is convenient to set

$$R(t) := \Lambda_2^{-\frac{1}{2}}(t) \tilde{R}(t) \Lambda_2^{-\frac{1}{2}}(t). \quad (2.7)$$

Note that  $\tilde{R}(t)$  and  $R(t)$  are symmetric. The following lemma is straightforward.

**Lemma 2.** *For any  $t \in M$ , the Gaussian random field  $X^{|t}(\cdot)$  and the Gaussian random matrix  $R(t)$  are independent of  $(X(t), \nabla X(t))$ .*

### 2.2. The first and second maxima, and the independent part of the Hessian

**First maximum:** We define the first maximum  $\lambda_1$  of  $\sigma X(\cdot)$  and its argument maximum  $t_1$  by

$$\lambda_1 := \sigma X(t_1) \quad \text{and} \quad t_1 := \arg \max_{t \in M} X(t). \quad (2.8)$$

The argument maximum is almost surely a singleton by (2.4), hence  $t_1$  is unique almost surely.

**Second maximum:** We define the second maximum  $\lambda_2$  of  $\sigma X(\cdot)$  by

$$\forall t \in M, \quad \lambda_2^t := \sup_{s \in M \setminus \{t\}} \{\sigma X^{|t}(s)\} \quad \text{and} \quad \lambda_2 := \lambda_2^{t_1}. \quad (2.9)$$



**Independent part of the Hessian:** We define the independent part of the Hessian  $\Omega$  as

$$\Omega := \sigma R(t_1) = \sigma \Lambda_2^{-\frac{1}{2}}(t_1) \tilde{R}(t_1) \Lambda_2^{-\frac{1}{2}}(t_1). \quad (2.10)$$

At this stage, it is not clear that  $\lambda_2^t < \infty$  a.s. and how  $X^{|t}(s)$  is shaped around point  $t$ . The next lemma gives a description of  $s \mapsto X^{|t}(s)$  around point  $t$  which proves that  $\lambda_2^t < \infty$  almost surely. A proof of this lemma can be found in Appendix A. Note that, since  $M$  is compact, by the Hopf–Rinow theorem,  $M$  is geodesically complete and the exponential map exists on the whole tangent space.

**Lemma 3.** *Let  $h$  be a nonzero vector of the tangent space at point  $t$ . For all  $\varepsilon \neq 0$ , let  $s(\varepsilon) := \exp_t(\varepsilon h) \in M$  be the exponential map at  $t$  given by the tangent vector  $\varepsilon h$ . Then, under Assumptions (A<sub>1</sub>), (A<sub>2</sub>) and (A<sub>4</sub>),*

$$\lim_{\substack{\varepsilon \neq 0 \\ \varepsilon \rightarrow 0}} X^{|t}(s(\varepsilon)) = \frac{h^\top \tilde{R}(t) h}{h^\top \Lambda_2(t) h} = \frac{h^\top}{\|h\|_2} R(t) \frac{h}{\|h\|_2}, \text{ almost surely.} \quad (2.11)$$

Furthermore, there exists a unit Euclidean norm tangent vector  $h_0$  at point  $t$  such that

$$\limsup_{s \rightarrow t} X^{|t}(s) = h_0^\top R(t) h_0 = \lambda_{\max}(R(t)) < \infty, \text{ almost surely.}$$

**Remark 3.** *The aforementioned lemma shows that  $\lambda_2^t$  varies in  $(-\infty, \infty)$  almost surely. Also, it shows that  $X^{|t}(\cdot)$  is a **helix** random field (Azaïs and Wschebor, 2005, Lemma 4.1) with pole  $t$ : the paths of the random field need not extend to a continuous function at the point  $t$ ; however, the paths have radial limits at  $t$  and the random field may take the form of a helix around  $t$ . The shape of the helix locally around the singularity is described by the eigen-decomposition of the independent part of the Hessian, as shown by (2.11)*

Besides, for every  $t$  such that  $\nabla X(t) = 0$  it follows that

$$X^{|t}(s) = \frac{X(s) - c(s, t)X(t)}{1 - c(s, t)} =: X^t(s). \quad (2.12)$$

In particular, we have the following identity between random fields  $X^{|t_1}(\cdot) = X^{t_1}(\cdot)$ , which may not be Gaussian (due to the random point  $t_1$ ). The following lemma is straightforward.

**Lemma 4.** *For a fixed  $t \in M$ , consider the following indicator functions:*

- $\nu_1 := \mathbf{1}\{t = t_1\}$ ;
- $\nu_2 := \mathbf{1}\{\forall s \in M \setminus \{t\}, X(s) \leq X(t)\}$ ;
- $\nu_3 := \mathbf{1}\{\forall s \in M \setminus \{t\}, X^t(s) \leq X(s)\}$ ;
- $\nu_4 := \mathbf{1}\{\nabla X(t) = 0\} \mathbf{1}\{\lambda_2^t \leq X(t)\}$ ;

then  $\nu_1 = \nu_2 = \nu_3 = \nu_4$  and  $-\infty < \lambda_2 \leq \lambda_1 < \infty$ , almost surely.

### 2.3. The conditional distribution of the maximum

We have the following key result giving the distribution of  $\lambda_1$ , defined by (2.8), conditional to  $(\lambda_2, \Omega)$ .

**Theorem 1.** *Let  $X(\cdot)$  be a Gaussian random field satisfying Assumption (A<sub>1-4</sub>). Then, the distribution  $\mathcal{D}(\lambda_1 | \lambda_2, \Omega)$  of the maximum  $\lambda_1$  conditional to  $(\lambda_2, \Omega)$  has a density with respect to the Lebesgue measure  $\text{Leb}$  and this conditional density at point  $\ell \in \mathbb{R}$  is given by*

$$\frac{d\mathcal{D}(\lambda_1 | \lambda_2, \Omega)}{d\text{Leb}}(\ell) = \frac{\det(\ell \text{Id} - \Omega) \varphi(\ell/\sigma)}{G_\Omega(\lambda_2)} \mathbf{1}_{\lambda_2 \leq \ell},$$

where  $G_\Omega(\lambda_2) := \int_{\lambda_2}^{+\infty} \det(u \text{Id} - \Omega) \varphi(u/\sigma) du$  and  $\varphi(\cdot)$  is the standard Gaussian density.

*Proof.* Consider the set of parameters

$$\mathcal{B} := \left\{ (\ell_1, \ell_2, r) \in \mathbb{R}^2 \times \mathcal{S}_d : \ell_2 \leq \ell_1 \right\},$$

where  $\mathcal{S}_d$  is the set of  $d \times d$  symmetric matrices. Let  $B_1$  be an open set of  $\mathbb{R} \times \mathcal{S}_d$  and  $b \in \mathbb{R}$ , define

$$B = \left\{ (\ell_1, \ell_2, r) \in \mathbb{R}^2 \times \mathcal{S}_d : \ell_2 < \ell_1, \ell_2 < b, (\ell_1, r) \in B_1 \right\}.$$

Note that  $B$  is an open set of  $\mathcal{B}$ . Note that this class of open sets generates the Borel sigma algebra of  $\mathcal{B}$ . Hence, in order to derive the joint law of the triplet  $(\lambda_1, \lambda_2, \Omega)$ , we compute

$$\begin{aligned} & \mathbb{P} \{ (\lambda_1, \lambda_2, \Omega) \in B \} \\ &= \mathbb{P} \{ \exists t \in M : \nabla X(t) = 0, (\sigma X(t), \lambda_2^t, \sigma R(t)) \in B \} \end{aligned} \quad (2.13)$$

$$= \mathbb{E} [\# \{ t \in M : \nabla X(t) = 0, (\sigma X(t), \lambda_2^t, \sigma R(t)) \in B \}] \quad (2.14)$$

$$= \int_M \mathbb{E} \left[ \left| \det(-\nabla^2 X(t)) \right| \mathbb{1}_{(\sigma X(t), \lambda_2^t, \sigma R(t)) \in B} \left| \nabla X(t) = 0 \right] p_{\nabla X(t)}(0) d\nu(t) \quad (2.15)$$

$$= \int_M \mathbb{E} \left[ \det(-\nabla^2 X(t)) \mathbb{1}_{(\sigma X(t), \lambda_2^t, \sigma R(t)) \in B} \left| \nabla X(t) = 0 \right] p_{\nabla X(t)}(0) d\nu(t) \quad (2.16)$$

$$= \int_M \mathbb{E} \left[ \det(X(t)\Lambda_2(t) - \tilde{R}(t)) \mathbb{1}_{(\sigma X(t), \lambda_2^t, \sigma R(t)) \in B} \right] p_{\nabla X(t)}(0) d\nu(t) \quad (2.17)$$

$$= \int_M \mathbb{E} \left[ \det(X(t)\text{Id} - R(t)) \mathbb{1}_{(\sigma X(t), \lambda_2^t, \sigma R(t)) \in B} \right] p_{\nabla X(t)}(0) \det \Lambda_2(t) d\nu(t), \quad (2.18)$$

where  $\nu$  is the surfacic measure on  $M$  and

- we used that  $(\sigma X(t), \lambda_2^t, \sigma R(t)) \in B$  implies  $\lambda_2^t \leq \sigma X(t)$ , and invoked Lemma 4 in (2.13);
- we used the uniqueness of the maximum (2.4) in (2.14);
- we used a *Kac-Rice formula* to get the third equality (2.15), a proof is given in Appendix B;
- we used that  $\nu_1 = \nu_4$  (see Lemma 4) and that the Hessian  $-\nabla^2 X(t)$  is semi-definite positive at a maximum to get (2.16);
- we used the Hessian regression formula (2.6) with  $\nabla X(t) = 0$  and invoked the independence of  $\nabla X(t)$  from  $(\sigma X(t), \lambda_2^t, \sigma \tilde{R}(t))$  to get (2.17);

Now, we introduce the measures (defined up to normalizing constants)

$$d\bar{\nu}(t) := p_{\nabla X(t)}(0) \det \Lambda_2(t) d\nu(t), \quad (2.19)$$

$$\mu^*(\cdot) := \int_M \mu_t(\cdot) d\bar{\nu}(t), \quad (2.20)$$

where  $\mu_t := \mathcal{D}(\lambda_2^t, \sigma R(t))$  is the distribution of  $(\lambda_2^t, \sigma R(t))$ . Since  $X(t)$  is independent of  $(\lambda_2^t, \sigma R(t))$ , one has

$$\mathcal{D}(\sigma X(t), \lambda_2^t, \sigma R(t)) = \mathcal{N}(0, \sigma^2) \otimes \mu_t,$$

by Assumption (A<sub>2</sub>). Now, coming back to (2.18) we have

$$\begin{aligned} \mathbb{P} \{ (\lambda_1, \lambda_2, \Omega) \in B \} &= \int_M \mathbb{E} \left[ \det(X(t)\text{Id} - R(t)) \mathbb{1}_{(\sigma X(t), \lambda_2^t, \sigma R(t)) \in B} \right] d\bar{\nu}(t) \\ &= \int_M \int_{\mathbb{R}^D} \det(\ell_1 \text{Id} / \sigma - \omega / \sigma) \mathbb{1}_{(\ell_1, \ell_2, \omega) \in B} \varphi(\ell_1 / \sigma) d\ell_1 d\mu_t(\ell_2, \omega) d\bar{\nu}(t) \\ &= \int_{\mathbb{R}^D} \int_M \det(\ell_1 \text{Id} / \sigma - \omega / \sigma) \mathbb{1}_{(\ell_1, \ell_2, \omega) \in B} \varphi(\ell_1 / \sigma) d\ell_1 d\mu_t(\ell_2, \omega) d\bar{\nu}(t) \\ &= \int_{\mathbb{R}^D} \det(\ell_1 \text{Id} / \sigma - \omega / \sigma) \mathbb{1}_{(\ell_1, \ell_2, \omega) \in B} \varphi(\ell_1 / \sigma) d\ell_1 d\mu^*(\ell_2, \omega), \end{aligned}$$

where  $D = (d^2 + d + 4)/2$  is the dimension of  $\mathbb{R}^2 \times \mathcal{S}_d$ . It implies that the joint density of  $(\lambda_1, \lambda_2, \Omega)$  at point  $(\ell_1, \ell_2, \omega)$  on the set  $\mathcal{B}$  has a density proportional to  $\det(\ell_1 \text{Id} / \sigma - \omega / \sigma) \varphi(\ell_1 / \sigma)$  with respect to  $\text{Leb} \otimes \mu^*$ . This implies in turn that the density of the maximum  $\lambda_1$  conditional to  $(\lambda_2, \Omega)$  is proportional to  $\det(\ell_1 \text{Id} - \omega) \varphi(\ell_1 / \sigma) \mathbb{1}_{\ell_2 \leq \ell_1}$ .  $\square$

### 3. Spacing test for the mean of a random field with known variance

#### 3.1. Testing framework

We observe a random field  $Z(t) = m(t) + \sigma X(t)$  and we would like to test the global nullity of its mean  $m(\cdot)$ . We define the statistics

$$\lambda_1 := \max_{t \in M} \{Z(t)\} \quad \text{and} \quad t_1 := \arg \max_{t \in M} \{Z(t)\}; \quad (3.1)$$

$$Z^{|t_1}(s) := \frac{Z(s) - \lambda_1 c(s, t_1)}{1 - c(s, t_1)}; \quad (3.2)$$

$$\lambda_2 := \max_{s \in M} \{Z^{|t_1}(s)\} \quad \text{and} \quad t_2 := \arg \max_{s \in M} \{Z^{|t_1}(s)\}; \quad (3.3)$$

$$\Omega := \Lambda_2^{-\frac{1}{2}}(t_1) (\nabla^2 Z(t_1) + \lambda_1 \Lambda_2(t_1)) \Lambda_2^{-\frac{1}{2}}(t_1). \quad (3.4)$$

In the previous section, we assumed that the Gaussian random field  $X(\cdot)$  was centered in **(A<sub>2</sub>)**. In this section, we give an exact test procedure for the following null hypothesis:

$$\mathbb{E}[Z(\cdot)] = 0, \quad (\mathbb{H}_0)$$

as a consequence  $Z(\cdot) = \sigma X(\cdot)$  and the notations  $\lambda_1, t_1, \lambda_2, t_2, \Omega$  are consistent with Section 2.

#### 3.2. Spacing test

We can now state our main result when the variance  $\sigma^2$  is known.

**Theorem 2.** *Let  $X(\cdot)$  be a Gaussian random field satisfying Assumption **(A<sub>1-4</sub>)**. Under  $\mathbb{H}_0$ , the test statistic*

$$\mathbf{S}_\sigma(\lambda_1, \lambda_2, \Omega) := \frac{G_{\Omega/\sigma}(\lambda_1/\sigma)}{G_{\Omega/\sigma}(\lambda_2/\sigma)} \sim \mathcal{U}(0, 1),$$

where  $\mathcal{U}(0, 1)$  is the uniform distribution on  $(0, 1)$ ,  $G_{\Omega/\sigma}(\ell) := \int_{\ell}^{+\infty} \det(u\text{Id} - \Omega/\sigma) \varphi(u) du$  and  $\varphi(\cdot)$  is the standard Gaussian density.

*Proof.* Without loss of generality, assume that  $\sigma = 1$ . It is well known that, if a random variable  $Q$  has a *continuous* cumulative distribution function  $F$ , then  $F(Q) \sim \mathcal{U}(0, 1)$ . This implies that, conditionally to  $(\lambda_2, \Omega) = (\ell_2, r)$ ,  $G_r(\lambda_1)/G_r(\ell_2) \sim \mathcal{U}(0, 1)$ . Since the conditional distribution does not depend on  $(\ell_2, r)$ , it is also the non-conditional distribution as claimed.  $\square$

### 4. The unknown variance case

#### 4.1. Existence of non-degenerate systems

We consider a real valued centered Gaussian random field  $X(\cdot)$  defined on  $M$  satisfying Assumptions **(A<sub>2</sub>)**, for the moment. We consider the order  $m$  Karhunen-Loève expansion in the sense that

$$\sigma X(t) = \sum_{i=1}^m \zeta_i f_i(t) \quad \text{with} \quad \text{Var}(\zeta_i) = \sigma^2 \quad \text{and} \quad \forall t \in M, \quad \sum_{i=1}^m |f_i(t)|^2 = 1, \quad (\text{KL}(m))$$

where the equality holds in  $L^2$ , uniformly in  $t$ , and  $(f_1, \dots, f_m)$  is a system of non-zero functions orthogonal on  $L^2(M)$ . We say that  $X(\cdot)$  satisfies Assumption **(KL)(m)** if it admits an order  $m$  Karhunen-Loève expansion. Through our analysis, we need to consider the following non-degeneracy Assumption:  $X$  is a.s. differentiable, it holds that  $m > d + 1$  and for all  $t_1 \in M$

$$\begin{aligned} &\exists (t_{d+2}, \dots, t_m) \in (M \setminus \{t_1\})^{m-d-1} \quad \text{pairwise distincts s.t.} \\ &(X(t_1), \nabla X(t_1), X(t_{d+2}), \dots, X(t_m)) \text{ is non degenerate.} \end{aligned} \quad (\text{ND}(m))$$

When  $X(\cdot)$  admits an infinite order Karhunen-Loève expansion in the sense that

$$\sigma X(t) = \sum_{i=1}^{\infty} \zeta_i f_i(t) \quad \text{with} \quad \text{Var}(\zeta_i) = \sigma^2 \quad \text{and} \quad \forall t \in M, \quad \sum_{i=1}^{\infty} |f_i(t)|^2 = 1, \quad (\text{KL}(\infty))$$

we say that  $X(\cdot)$  satisfies Assumption  $(\text{KL}(\infty))$ . In this case, the non-degeneracy condition reads:  $X$  is a.s. differentiable, for all  $p > d + 1$  and for all  $t_1 \in M$

$$\begin{aligned} \exists(t_{d+2}, \dots, t_p) \in (M \setminus \{t_1\})^{p-d-1} \text{ pairwise distincts s.t.} \\ (X(t_1), \nabla X(t_1), X(t_{d+2}), \dots, X(t_p)) \text{ is non degenerated.} \end{aligned} \quad (\text{ND}(\infty))$$

A standard result shows that if the covariance function of  $X(\cdot)$  is  $\mathcal{C}^0(M \times M)$  on  $M$  compact then the Karhunen-Loève expansion exists (of finite or infinite order). The following key result shows that if  $X$  has  $\mathcal{C}^1$  paths almost surely, and satisfies Assumptions  $(\mathbf{A}_2)$  and  $(\mathbf{A}_4)$  then the non-degeneracy condition also holds.

**Proposition 5.** *Let  $X(\cdot)$  be a real valued Gaussian random field having  $\mathcal{C}^1$  paths almost surely, and satisfying Assumptions  $(\mathbf{A}_2)$  and  $(\mathbf{A}_4)$ . Then, for all  $m > d + 1$ , Assumption  $(\text{KL}(m))$  implies Assumption  $(\text{ND}(m))$ , and also Assumption  $(\text{KL}(\infty))$  implies Assumption  $(\text{ND}(\infty))$ .*

*Proof.* Since  $X$  has  $\mathcal{C}^1$  paths almost surely, note that the covariance function  $c(\cdot, \cdot)$  of the Gaussian random field  $X(\cdot)$  has continuous partial derivatives.

Consider the centered Gaussian vector  $V := (\nabla X(t_1), X(t_1), X(t_{d+2}), \dots, X(t_p))$  with variance-covariance matrix  $\Sigma(V) = \mathbb{E}[VV^\top]$  given by

$$\begin{pmatrix} \overbrace{\begin{matrix} \partial_1 \partial_1 c(t_1, t_1) & \partial_1 \partial_2 c(t_1, t_1) & \dots & \partial_1 \partial_d c(t_1, t_1) \\ \partial_2 \partial_1 c(t_1, t_1) & \partial_2 \partial_2 c(t_1, t_1) & \dots & \partial_2 \partial_d c(t_1, t_1) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_d \partial_1 c(t_1, t_1) & \partial_d \partial_2 c(t_1, t_1) & \dots & \partial_d \partial_d c(t_1, t_1) \end{matrix}}^{\nabla X(t_1)} & \overbrace{\begin{matrix} \partial_1 c(t_1, t_1) & \partial_1 c(t_1, t_{d+2}) & \dots & \partial_1 c(t_1, t_p) \\ \partial_2 c(t_1, t_1) & \partial_2 c(t_1, t_{d+2}) & \dots & \partial_2 c(t_1, t_p) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_d c(t_1, t_1) & \partial_d c(t_1, t_{d+2}) & \dots & \partial_d c(t_1, t_p) \end{matrix}}^{(X(t_1), X(t_{d+2}), \dots, X(t_p))} \\ \star & \begin{matrix} c(t_1, t_1) & c(t_1, t_{d+2}) & \dots & c(t_1, t_p) \\ c(t_{d+2}, t_1) & c(t_{d+2}, t_{d+2}) & \dots & c(t_{d+2}, t_p) \\ \vdots & \vdots & \ddots & \vdots \\ c(t_p, t_1) & c(t_p, t_{d+2}) & \dots & c(t_p, t_p) \end{matrix} \end{pmatrix}$$

where  $\partial_i \partial_j c(x, x) := \mathbb{E}[(\partial X / \partial x_i)(x)(\partial X / \partial x_j)(x)]$  and  $\partial_i c(x, y) := \mathbb{E}[(\partial X / \partial x_i)(x)X(y)]$  with  $(\partial X / \partial x_i)(x)$  the partial derivative with respect to  $x_i$  at point  $x = (x_1, \dots, x_d) \in M$  in Riemannian normal coordinates, and  $y \in M$ .

Denote by  $\mathbb{H}$  the RKHS defined by  $c(\cdot, \cdot)$ . By a standard result, see for instance [Steinwart and Christmann \(2008, Corollary 4.36\)](#), it holds that  $\partial_i \Phi(t_j)$  belongs to  $\mathbb{H}$  and

$$\begin{aligned} c(t_i, t_j) &= \langle \Phi(t_i), \Phi(t_j) \rangle_{\mathbb{H}}; \\ \partial_i c(t_1, t_j) &= \langle \partial_i \Phi(t_1), \Phi(t_j) \rangle_{\mathbb{H}}; \\ \partial_i \partial_j c(t_1, t_1) &= \langle \partial_i \Phi(t_1), \partial_j \Phi(t_1) \rangle_{\mathbb{H}}; \end{aligned}$$

where  $\Phi$  is the canonical feature map of  $(\mathbb{H}, \langle \cdot, \cdot \rangle_{\mathbb{H}})$ . We recognize that  $\Sigma(V)$  is the Gram matrix, for the scalar product  $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ , of  $\Psi := [\partial_1 \Phi(t_1) \partial_2 \Phi(t_1) \dots \partial_d \Phi(t_1) \Phi(t_1) \Phi(t_{d+2}) \dots \Phi(t_p)] \in \mathbb{H}^p$ , and  $\Sigma(V) = \Psi^* \Psi$  where  $\Psi^* : h \in \mathbb{H} \mapsto (\langle \partial_1 \Phi(t_1), h \rangle_{\mathbb{H}}, \dots, \dots, \dots, \langle \Phi(t_p), h \rangle_{\mathbb{H}}) \in \mathbb{R}^p$  is the adjoint operator of  $\Psi$ .

Using Assumption  $(\mathbf{A}_2)$ , Assumption  $(\mathbf{A}_4)$  and (2.2), one gets that  $W = (X(t_1), \nabla X(t_1))$  is non-degenerated, and its variance covariance matrix  $\Sigma(W)$  has full rank. This matrix is also the Gram matrix of the system of vectors in  $\mathbb{H}$  given by  $\tilde{\Psi} := [\Phi(t_1) \partial_1 \Phi(t_1) \dots \partial_d \Phi(t_1)] \in \mathbb{H}^{d+1}$ . We deduce that is a free system (i.e., it spans a vector space of dimension  $d + 1$ ). One also knows that the dimension of  $\mathbb{H}$  is exactly the order of the Karhunen-Loève expansion, actually it is standard to prove the Karhunen-Loève expansion from Mercer's theorem. Hence, it is always possible to complete  $\tilde{\Psi}$  by  $p - d - 1$  vectors of the form  $\Phi(t)$  to get a  $p$  dimensional free system  $\Psi$ , otherwise  $\mathbb{H}$  would be of dimension less than  $p$ .  $\square$

We give some examples below:

- The normalized Brownian motion  $W_t / \sqrt{t}$  satisfies Assumption  $(\text{KL}(\infty))$ ;
- Any Gaussian stationary field with a spectrum that admits an accumulation point satisfies Assumptions  $(\text{KL}(\infty))$  and  $(\text{ND}(\infty))$  if differentiable, see for instance [Azaïs and Wschebor \(2009, Exercices 3.4 and 3.5\)](#);

- Any Gaussian random field satisfying conditions of [Azaïs and Wschebor \(2005, Proposition 3.1\)](#) has  $(\text{KL}(\infty))$  and  $(\text{ND}(\infty))$
- Note that in the applications of Section 5 all the Gaussian random fields satisfy  $(\text{KL}(m))$  with  $m$  finite and explicitly given.

#### 4.2. The Karhunen-Loève estimator of the variance

In practical applications, the assumption that the variance is known is too restrictive. In this section, we are supposed to observe  $\sigma X(\cdot)$  where  $\sigma > 0$  is unknown and  $X(\cdot)$  satisfies Assumption  $(\mathbf{A}_{1-4})$ . In particular,  $X(\cdot)$  admits a Karhunen-Loève expansion of order denoted  $m_{\text{KL}}(X)$  possibly infinite. We assume the following fifth condition:

$$\kappa := m_{\text{KL}}(X) - d - 1 \geq 1. \quad (\mathbf{A}_5)$$

We will make use of the notation  $m := m_{\text{KL}}(X)$  when there is no ambiguity.

**Remark 4.** Using Assumption  $(\mathbf{A}_4)$  and the fact that the gradient of  $X(t)$  is independent of  $X(t)$ , it can be shown that  $m_{\text{KL}}(X)$  is greater than or equal to  $d + 1$ , by an argument similar to the proof of Proposition 5. Note Assumption  $(\mathbf{A}_5)$  requires at least  $d + 2$  degrees of freedom, and this extra degree of freedom is the price to estimate the variance.

**The infinite order case** When  $X(\cdot)$  satisfies  $(\text{KL}(\infty))$ , note that, for every integer  $p \geq 2$ , from the observation of  $(\sigma X(t_1), \dots, \sigma X(t_p))$  for conveniently chosen points  $t_1, \dots, t_p$  (say uniformly at random for instance), one can build an estimator, say  $\hat{\sigma}_{(p)}^2$ , of the variance  $\sigma^2$  with chi-squared distribution  $\sigma^2 \chi^2(p-1)/(p-1)$  under  $\mathbb{H}_0$ . Making  $p$  tend to infinity, classical concentration inequalities and Borel-Cantelli lemma prove that  $\hat{\sigma}_{(p)}^2$  converges almost surely to  $\sigma^2$  under  $\mathbb{H}_0$ . Thus the variance  $\sigma^2$  is theoretically directly observable from the entire path of  $\sigma X(\cdot)$  (though in practical applications one will estimate it by a  $\chi^2$  with a large number of degrees of freedom). We still denote by  $\hat{\sigma}_t^2$  this observation.

**The finite order case** By Proposition 5, note that Assumption  $(\text{ND}(m))$  holds with  $m > d + 1$ . Let  $t \in M$  and let  $(t_{d+2}, \dots, t_m) \in M^\kappa$  be as in Assumption  $(\text{ND}(m))$ , then the Gaussian vector  $(\sigma X^{|t}(t_{d+2}), \dots, \sigma X^{|t}(t_m))$  has for variance-covariance matrix  $\sigma^2 \Sigma$  where  $\Sigma$  is some known matrix and an estimator of  $\sigma^2$  is

$$\hat{\sigma}_t^2 := \|\Sigma^{-1/2}(\sigma X^{|t}(t_{d+2}), \dots, \sigma X^{|t}(t_m))\|^2 / \kappa. \quad (4.1)$$

Direct algebra shows that  $X^{|t}(\cdot)$  inherits an order  $\kappa$  Karhunen-Loève expansion from the order  $m$  Karhunen-Loève expansion of  $X(\cdot)$ . More precisely, under  $(\mathbb{H}_0)$ ,

$$\forall s \neq t \in M, \quad \sigma X^{|t}(s) = \sum_{i=1}^{\kappa} \bar{\zeta}_i \bar{f}_i(s) \text{ with } \text{Var}(\bar{\zeta}_i) = \sigma^2, \quad (4.2)$$

where equality holds in  $L^2$ , uniformly in  $s$ , and it holds

$$\hat{\sigma}_t^2 = \frac{1}{\kappa} \sum_{i=1}^{\kappa} \bar{\zeta}_i^2. \quad (4.3)$$

It shows that  $\hat{\sigma}_t^2$  does not depend on the choice of  $t_{d+2}, \dots, t_m$  in  $(\text{ND}(m))$ .

**Proposition 6.** Let  $X(\cdot)$  satisfy Assumptions  $(\mathbf{A}_1)$ ,  $(\mathbf{A}_2)$ ,  $(\mathbf{A}_4)$  and  $(\mathbf{A}_5)$ . Let  $t \in M$ , then the following claims are true under the null hypothesis  $(\mathbb{H}_0)$ .

- (i)  $\hat{\sigma}_t^2$  is well defined and follows a  $\sigma^2 \chi^2(\kappa)/\kappa$  distribution;
- (ii)  $\hat{\sigma}_t^2$  is independent of  $(X(t), \nabla X(t))$ ;
- (iii) the random field  $X^{|t}(\cdot)/\hat{\sigma}_t$  is independent of the random variable  $\hat{\sigma}_t$ .

*Proof.* Statement (i) follows from (4.3) and Statement (ii) follows from (4.1) and the independence of  $X^{|t}(\cdot)$  from  $(X(t), \nabla X(t))$ . The last statement is a direct consequence of the independence between the angle and the norm of  $\bar{\zeta} = (\bar{\zeta}_1, \dots, \bar{\zeta}_\kappa)$  and (4.2).  $\square$

### 4.3. Joint law in the unknown variance case

We present below the use of the estimation  $\widehat{\sigma}_t$  to modify the spacing test. For fixed  $t \in M$ , by Proposition 6, we know that  $X(t), \nabla X(t), X^{|\cdot|}(\cdot)/\widehat{\sigma}_t$  and  $\widehat{\sigma}_t$  are mutually independent. By (2.7), we recall that

$$R(t) := \Lambda_2^{-\frac{1}{2}}(t) \left[ \nabla^2 X(t) + \Lambda_2(t) X(t) + \Lambda_3(t) \nabla X(t) \right] \Lambda_2^{-\frac{1}{2}}(t),$$

As Lemma 3 shows,  $R(t)$  can be expressed as radial limits of  $X^{|\cdot|}(\cdot)$  at point  $t$  hence we get that the random variables  $\left\{ X(t), \nabla X(t), (X^{|\cdot|}(\cdot)/\widehat{\sigma}_t), R(t)/\widehat{\sigma}_t, \widehat{\sigma}_t \right\}$  are mutually independent and, by consequence, the variables

$$X(t), \nabla X(t), \left[ \frac{\lambda_2^t}{\widehat{\sigma}_t}, \frac{R(t)}{\widehat{\sigma}_t} \right], \widehat{\sigma}_t \text{ are mutually independent,}$$

where we recall that  $\lambda_2^t$  is defined in (2.9). We consider  $t \in M$  a putative value and  $\lambda_1^t$  build from  $\sigma X(\cdot)$  in (2.8). Now, consider the test statistics

$$T_{1,t} := \frac{\lambda_1^t}{\widehat{\sigma}_t}, T_1 := T_{1,t_1}, T_{2,t} := \frac{\lambda_2^t}{\widehat{\sigma}_t}, T_2 := T_{2,t_1}, \widehat{\sigma} := \widehat{\sigma}_{t_1} \text{ and } \Omega = \sigma R(t_1) \text{ as in (2.10), (4.4)}$$

and let  $\bar{\mu}_t$  be the joint law of the couple of random variables  $(T_{2,t}, \sigma R(t)/\widehat{\sigma}_t)$ .

Under null  $(\mathbb{H}_0)$ , note that the variable  $\sigma X(t)$  is a centered Gaussian variable with variance  $\sigma^2$  and  $\sqrt{\kappa} \widehat{\sigma}_t/\sigma$  is distributed as a  $\chi$ -distribution with  $\kappa$  degrees of freedom. Hence,  $\widehat{\sigma}_t/\sigma$  has density

$$f_{\frac{\widehat{\sigma}_t}{\sigma}}(s) = \frac{2^{1-\frac{\kappa}{2}}}{\Gamma(\frac{\kappa}{2})} \sqrt{\kappa} (s\sqrt{\kappa})^{\kappa-1} \exp(-(\kappa s^2/2)),$$

under the null  $(\mathbb{H}_0)$ . Then  $(\sigma X(t), \widehat{\sigma}_t/\sigma, T_{2,t}, \sigma R(t)/\widehat{\sigma}_t)$  has a density

$$(\text{const}) s^{\kappa-1} \exp(-(\kappa s^2/2)) \varphi(\ell_1/\sigma),$$

with respect to  $\text{Leb}_2 \otimes \bar{\mu}_t$  at point  $(\ell_1, s, t_2, r) \in \mathbb{R}^3 \times \mathcal{S}_d$  and where the constant (const) may depend on  $m, \kappa$  and  $\sigma$ . Using the same method as for the proof of Theorem 1 we have the following proposition.

**Proposition 7.** *Let  $X(\cdot)$  be a Gaussian random field satisfying Assumption  $(\mathbf{A}_{1-5})$ . Then, under  $(\mathbb{H}_0)$ , the joint distribution of  $(\lambda_1, \widehat{\sigma}/\sigma, T_2, \Omega/\widehat{\sigma})$  has a density with respect to  $\text{Leb}_2 \otimes \bar{\mu}^*$  at point  $(\ell_1, s, t_2, r) \in \mathbb{R}^3 \times \mathcal{S}_d$  equal to*

$$(\text{const}) \det(\ell_1 \text{Id} - \sigma s r) s^{\kappa-1} \exp(-(\kappa s^2/2)) \varphi(\ell_1/\sigma) \mathbf{1}_{\{0 < \sigma s t_2 < \ell_1\}},$$

where  $\bar{\mu}^*$  is defined by  $\bar{\mu}^*(\cdot) := \int_M \bar{\mu}_t(\cdot) d\bar{\nu}(t)$  and  $\bar{\nu}(t)$  by (2.19).

### 4.4. $t$ -Spacing test, an exact test for the unknown variance case

We have the second main result, when the variance is unknown.

**Theorem 3.** *Let  $X(\cdot)$  be a Gaussian random field satisfying Assumption  $(\mathbf{A}_{1-5})$ . For all  $r \in \mathcal{S}_d$ , define  $H_r$  as*

$$\forall \ell \in \mathbb{R}, H_r(\ell) := \int_{\ell}^{+\infty} \det(u \text{Id} - r) f_{m-1}\left(u \sqrt{(m-1)/\kappa}\right) du, \quad (4.5)$$

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

$$\mathbf{T}(T_1, T_2, \Omega/\widehat{\sigma}) := \frac{H_{\Omega/\widehat{\sigma}}(T_1)}{H_{\Omega/\widehat{\sigma}}(T_2)} \sim \mathcal{U}(0, 1),$$

where  $T_1, T_2, \Omega$  are given by (4.4).

*Proof.* First, using Proposition 7 and the change of variable  $t_1 = \ell_1/\sigma s$ , the joint density of the quadruplet  $(T_1, \hat{\sigma}/\sigma, T_2, \Omega/\hat{\sigma})$  at point  $(t_1, s, t_2, r)$  is given by

$$\begin{aligned} & (\text{const}) \det(\sigma s(t_1 \text{Id} - r)) s^\kappa \exp(-(\kappa s^2/2)) \varphi(st_1) \mathbf{1}_{\{0 < t_2 < t_1\}} \\ &= (\text{const}) \det(t_1 \text{Id} - r) s^{m-1} \exp \left[ - \left( s \sqrt{\frac{\kappa}{m-1}} \right)^2 \frac{m-1}{2} \right] \varphi(st_1) \mathbf{1}_{\{0 < t_2 < t_1\}}. \end{aligned}$$

Second, note that if  $U$  and  $V$  are independent with densities  $f_U$  and  $f_V$  then  $W := \frac{U}{V}$  has density

$$f_W(w) = \int_{\mathbb{R}} f_U(wv) v f_V(v) dv.$$

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

$$\begin{aligned} & \int_{\mathbb{R}^+} \varphi(st_1) s^{m-1} \exp \left[ - \left( s \sqrt{\frac{\kappa}{m-1}} \right)^2 \frac{m-1}{2} \right] ds \\ &= (\text{const}) \int_{\mathbb{R}^+} \varphi \left( st_1 \sqrt{\frac{m-1}{\kappa}} \right) s s^{m-2} \exp \left[ - \frac{s^2(m-1)}{2} \right] ds \\ &= (\text{const}) \int_{\mathbb{R}^+} \varphi \left( st_1 \sqrt{\frac{m-1}{\kappa}} \right) s f_{\frac{\chi_{m-1}}{\sqrt{m-1}}}(s) ds \\ &= (\text{const}) f_{m-1} \left( t_1 \sqrt{\frac{m-1}{\kappa}} \right). \end{aligned}$$

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

$$(\text{const}) \det(t_1 \text{Id}_d - r) f_{m-1} \left( t_1 \sqrt{\frac{m-1}{\kappa}} \right) \mathbf{1}_{\{0 < t_2 < t_1\}},$$

and we conclude using the same trick as the one of Theorem 2.  $\square$

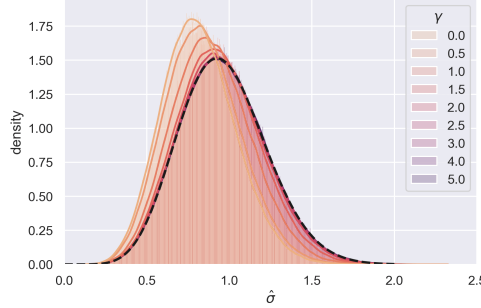


FIGURE 4. [Spiked tensor PCA, Section 5.2, example 4/4] The variance estimator is not distributed according to  $\chi^2$ -distribution with  $\kappa = 7$  degrees of freedom, it underestimates the variance. The dashed black line is the distribution of a  $\chi(7)/\sqrt{7}$ , and  $\sigma = 1$  in these experiments. The probability distribution function is estimated over 250,000 Monte-Carlo samples for each value of  $\gamma$ .

**Remark 5** (On the variance estimator). The formula (4.5) involves the Student density function with  $m - 1$  degrees of freedom which necessitates several comments:

- First, this formula shows that  $\kappa \hat{\sigma}^2/\sigma^2$  (resp.  $(m - 1) \hat{\sigma}^2/\sigma^2$ ) fails to be distributed according to a  $\chi^2$ -distribution with  $\kappa$  (resp.  $m - 1$ ) degrees of freedom. This because  $t_1$  is random.
- Second, Figure 4 illustrates that  $\hat{\sigma}^2$  under-estimates the variance  $\sigma^2$  under the null ( $\gamma = 0$ ) or moderate alternatives ( $\gamma \simeq 1$ ). The reason for this is clear: the chi-squared distribution assumes that the model is pre-specified, not chosen on the basis of  $\sigma X(\cdot)$ . But the Continuous LARS procedure has deliberately chosen the strongest predictor  $\psi_{t_1}$ , with  $t_1 \in M$  maximum of  $Z(\cdot)$ , among all of the available choices, so it is not surprising that it yields a drop in the value of the variance estimate on the residuals  $Z^{|t_1}(\cdot)$ .

- Third, Figure 4 shows that  $\kappa \hat{\sigma}^2 / \sigma^2$  has almost  $\chi^2$ -distribution with  $\kappa$  degrees of freedom for large alternatives ( $\gamma = 5$ ). The reason for this is clear: the mean  $m(\cdot)$  is much larger than noise  $\sigma X(\cdot)$  hence  $t_1 \simeq t_0$  and  $\kappa \hat{\sigma}^2 / \sigma^2 \simeq \kappa \hat{\sigma}_{t_1}^2 / \sigma^2$  which has  $\chi^2$ -distribution with  $\kappa$  degrees of freedom.

## 5. Examples

### 5.1. How to deploy our method

We are given an observation  $Y \in E$  as in Section 1.6 and we compute the corresponding profile likelihood random field given by  $Z(t) = \langle Y, \Psi_t \rangle_E$ . We would like to test the global nullity of its mean  $m(\cdot)$  given by hypothesis  $(\mathbf{H}_0)$ . The testing framework is depicted in Section 3.1 and we recall that

$$\Omega = \Lambda_2^{-\frac{1}{2}}(t_1) (\nabla^2 Z(t_1) + \lambda_1 \Lambda_2(t_1)) \Lambda_2^{-\frac{1}{2}}(t_1). \quad (5.1)$$

The first and second maximum (and their arguments) are given by (2.8), which are computed using a Riemaniann gradient descent algorithm on  $M$ . We will give the expression of  $\Lambda_2(t)$  in each example together with a closed form expression to compute the Riemaniann Hessian from the Euclidean Hessian and the Euclidean gradient. The Euclidean Hessian can be computed using numerical differentiation or, in some cases, is given explicitly.

When the variance  $\sigma^2$  is known, the testing statistics is given by Theorem 2. In particular, one needs to compute

$$G_{\Omega/\sigma}(\ell) = \int_{\ell}^{+\infty} \det(u \text{Id} - \Omega/\sigma) \varphi(u) du. \quad (5.2)$$

When the variance  $\sigma^2$  is unknown, the testing statistics is given by Theorem 3. In particular, one needs to compute

$$H_{\Omega/\hat{\sigma}}(\ell) = \int_{\ell}^{+\infty} \det(u \text{Id} - \Omega/\hat{\sigma}) f_{m-1}\left(u \sqrt{(m-1)/\kappa}\right) du, \quad (5.3)$$

where we recall that  $f_{m-1}(\cdot)$  is the density of the Student  $t$ -distribution with  $m-1$  degrees of freedom. Numerical integration can be done but, in some cases, we give explicit expressions of  $G_{\Omega/\sigma}(\ell)$  and  $H_{\Omega/\hat{\sigma}}(\ell)$ .

As for the variance estimation, we draw  $\kappa$  independent points  $(t_{d+2}, \dots, t_m)$  uniformly on  $M$ . They generically satisfy Condition  $(\mathbf{ND}(m))$ . For a putative point  $t \in M$ , the Gaussian vector  $(X^t(t_{d+2}), \dots, X^t(t_m))$  has for variance-covariance matrix  $\sigma^2 \Sigma$  where  $\Sigma$  is some known matrix and an estimator of  $\sigma^2$  is given by (4.1) with  $t = t_1$  as in (4.4). In our experiments, we have drawn 5 independent samples on  $\kappa$  points and we found the same value for  $\hat{\sigma}$ , as shown by the theory (4.3).

### 5.2. Spiked tensor PCA

We consider a simple example of the detection of a rank one 3-way symmetric tensor observing  $Y = \lambda_0 t_0^{\otimes 3} + \sigma W$  where  $W$  is an order 3 symmetric tensor defined using symmetry and the following independent terms,

- $W_{iii}$  of variance 1; there are 3 principal diagonal terms,  $i \in [3]$ ,
- $W_{ijj}$  of variance 1/3; there are 6 sub-diagonals terms,  $i \neq j \in [3]$ ,
- $W_{123}$  of variance 1/6; there is 1 off-diagonal term.

The profile likelihood is given by  $Z(t) = \langle Y, t^{\otimes 3} \rangle_E = \lambda_0 \langle t_0, t \rangle^3 + \sigma X(t)$  where  $M = S^2$  is the 2-sphere and  $E$  is the Euclidean space of 3-way symmetric tensors of size  $3 \times 3 \times 3$  denoted by  $(E, \langle \cdot, \cdot \rangle_E)$  with dimension  $m = 10$ . Hence, we have

$$\begin{aligned} \psi_t &:= t^{\otimes 3} = (t_i t_j t_k)_{1 \leq i, j, k \leq 3}, \\ c(s, t) &:= \langle s, t \rangle^3, \\ Z(t) &:= \langle Y, \psi_t \rangle_E = \sum_{ijk} t_i t_j t_k Y_{ijk}. \end{aligned}$$

The reader may consult <https://github.com/ydecastro/tensor-spacing/> for further details on the numerical experiments.



**Hessian** To compute gradient and Hessian we can take advantage of the isometry of the problem and consider, without loss of generality, the case where  $t \in M$  is the so-called “north pole”:  $(0, 0, 1)$ . We have

$$Z'(0, 0, 1) = \begin{pmatrix} 3Y_{133} \\ 3Y_{233} \\ 3Y_{333} \end{pmatrix}.$$

This shows that  $\Lambda_2 = 3 \text{Id}_2$  (on the tangent space), hence

$$\Omega = \sigma R(t_1) = \frac{\sigma}{3} \tilde{R}(t_1).$$

Since the second fundamental form of the unit sphere is  $-\text{Id}$ , the Riemannian Hessian is equal to the Euclidean Hessian  $Z''(t)$  limited to the tangent space  $-\text{Id}$  times the normal derivative (oriented outwards). The Euclidean Hessian  $Z''(t)$  limited to the tangent space is given by

$$\begin{pmatrix} 2Y_{113} & Y_{123} \\ Y_{123} & 2Y_{223} \end{pmatrix}.$$

This shows that the gradient is independent from the Hessian. To compute the independent part  $R$  we can use

$$Z(0, 0, 1) = Z'_3(0, 0, 1) = Y_{333} \quad \text{and} \quad Z'_3(0, 0, 1) = Y_{333}.$$

This shows that  $R(t)$  is always equal to the Euclidean Hessian restricted to the tangent space. By (2.10), it yields

$$\Omega = \frac{\sigma}{3} \tilde{R}(t_1) = \frac{\sigma}{3} (\text{Id} - \Pi_{t_1}) Z''(t_1) (\text{Id} - \Pi_{t_1}),$$

where  $\Pi_t := tt^\top$  is the orthogonal projection onto the normal space at point  $t$ , and direct algebra gives

$$\frac{\partial^2 Z}{\partial t^i \partial t^j}(t) = 6 \sum_{k=1}^3 Y_{ijk} t^k. \quad (5.4)$$

**Test statistics** One can compute the statistics  $(t_1, \lambda_1)$  and  $(t_2, \lambda_2)$  using a gradient descent. The expression of  $\Omega$  has an explicit form by (5.4). Because we deal with  $2 \times 2$  matrices, we have for the spacing test the following identities,

$$\begin{aligned} G_{\Omega/\sigma}(\ell) &= \int_{\ell}^{+\infty} \det(u \text{Id}_2 - \Omega/\sigma) \varphi(u) du \\ &= \int_{\ell}^{+\infty} [(u^2 - 1) - \text{Tr}(\Omega/\sigma)u + \det(\Omega/\sigma) + 1] \varphi(u) du \\ &= \ell \varphi(\ell) - \text{Tr}(\Omega/\sigma) \varphi(\ell) + (\det(\Omega/\sigma) + 1)(1 - \Phi(\ell)), \end{aligned}$$

with

$$\begin{aligned} \det(\Omega/\sigma) &= (1/9\sigma^2) \det[(\text{Id}_3 - \Pi_{t_1})Z''(t_1)(\text{Id}_3 - \Pi_{t_1}) + \Pi_{t_1}] \\ \text{Tr}(\Omega/\sigma) &= (1/3\sigma)(\text{Tr}(Z''(t_1)) - t_1^\top Z''(t_1)t_1). \end{aligned}$$

As for the  $t$ -spacing test, one can check that

$$\begin{aligned} H_{\Omega/\hat{\sigma}}(\ell) &= \int_{\ell}^{+\infty} \det(u \text{Id}_2 - \Omega/\hat{\sigma}) f_{m-1}\left(u\sqrt{(m-1)/\kappa}\right) du \\ &= \int_{\ell}^{+\infty} [u^2 - \text{Tr}(\Omega/\hat{\sigma})u + \det(\Omega/\hat{\sigma})] f_{m-1}\left(u\sqrt{(m-1)/\kappa}\right) du \\ &= \frac{\kappa\sqrt{m-3}}{(m-2)\sqrt{m-1}} \frac{\Gamma(\frac{m}{2})\Gamma(\frac{m-3}{2})}{\Gamma(\frac{m-1}{2})\Gamma(\frac{m-2}{2})} \sqrt{\frac{\kappa}{m-3}} \left[ \ell \sqrt{\frac{m-3}{\kappa}} f_{m-3}\left(\ell \sqrt{\frac{m-3}{\kappa}}\right) + 1 - F_{m-3}\left(\ell \sqrt{\frac{m-3}{\kappa}}\right) \right] \\ &\quad - \text{Tr}(\Omega/\hat{\sigma}) \frac{\kappa\sqrt{m-3}}{(m-2)\sqrt{m-1}} \frac{\Gamma(\frac{m}{2})\Gamma(\frac{m-3}{2})}{\Gamma(\frac{m-1}{2})\Gamma(\frac{m-2}{2})} f_{m-3}\left(\ell \sqrt{\frac{m-3}{\kappa}}\right) \\ &\quad + \det(\Omega/\hat{\sigma}) \sqrt{\frac{\kappa}{m-1}} \left(1 - F_{m-1}\left(\ell \sqrt{\frac{m-1}{\kappa}}\right)\right), \end{aligned}$$

with  $m = 10$ ,  $\kappa = 7$ ,  $F_\alpha$  (resp.  $f_\alpha$ ) the cumulative distribution function (resp. distribution density) of the Student  $t$ -distribution with  $\alpha$  degrees of freedom and

$$\begin{aligned}\det(\Omega/\hat{\sigma}) &= (1/9\hat{\sigma}^2) \det [(\text{Id}_3 - \Pi_{t_1})Z''(t_1)(\text{Id}_3 - \Pi_{t_1}) + \Pi_{t_1}] \\ \text{Tr}(\Omega/\hat{\sigma}) &= (1/3\hat{\sigma})(\text{Tr}(Z''(t_1)) - t_1^\top \sigma X''(t_1)t_1).\end{aligned}$$

**The  $t$ -spacing test** In Figures 2 and 3, the alternative is given by  $\lambda_0 = \gamma \times \sigma \sqrt{3 \log 3 + 3 \log \log 3}$  where  $\gamma = 1$  corresponds to the so-called phase transition in Spiked tensor PCA as presented in Perry et al. (2020, Theorem 1.3). In Figure 3, the top right panel shows that  $\lambda_1$  is stochastically increasing as  $\gamma$  increases while the top right panel shows that the distribution of  $\lambda_2$  remains unchanged for moderate values of  $\gamma$ , say  $\gamma \geq 2$ . It illustrates that the spacing between  $\lambda_1$  and  $\lambda_2$  grows linearly with  $\gamma$ . In the moderate regime, the bottom left panel shows that  $\lambda_1$  is distributed as a Gaussian with mean  $\lambda_0$  and variance  $\sigma^2$  (dashed black line), which is the distribution of  $Z(t_0)$  under  $\mathbb{H}_1(t_0, \lambda_0)$ . The bottom right panel shows that  $t_1 \simeq t_0$ , for moderate values of  $\gamma$ . It illustrates that  $(t_1, \lambda_1)$  is weakly close to the distribution of  $(t_0, Z(t_0))$ , the  $(t)$ -spacing tests detect the alternative  $\mathbb{H}_1(t_0, \lambda_0)$ .

### 5.3. Two-spiked tensor model

We consider a generalization to higher dimensions and two-spiked tensors of the preceding example (Section 5.2) by

$$Y = \nu_{0,1}x_0^{\otimes k} + \nu_{0,2}y_0^{\otimes k} + \sigma W,$$

where:

- the Euclidian space  $E$  is given by  $k$ -way  $n$ -dimensional symmetric tensors ( $k \geq 3, n \geq 4$ ) equipped with the dot product

$$\forall \mathbf{T}, \mathbf{U} \in (\mathbb{R}^n)^{\otimes k}, \quad \langle \mathbf{T}, \mathbf{U} \rangle_E = \sum_{i_1, \dots, i_k \in [n]} T_{i_1, \dots, i_k} U_{i_1, \dots, i_k}, \quad (5.5)$$

with Euclidean or Frobenius norm  $\|\mathbf{T}\|^2 = \langle \mathbf{T}, \mathbf{T} \rangle_E$ , where  $[n] = 1, \dots, n$ ;

- the noise tensor  $W \in (\mathbb{R}^n)^{\otimes k}$  is defined by

$$W = \frac{1}{k!} \sum_{\pi \in \mathfrak{S}_n} G^\pi, \quad (5.6)$$

where one considers i.i.d standard Gaussian  $G_{i_1 \dots i_k}$  for indices  $1 \leq i_1, \dots, i_k \leq n$ ,  $\mathfrak{S}_n$  is the set of permutations of the set  $[n]$ , and  $G_{i_1 \dots i_k}^\pi = G_{\pi(i_1) \dots \pi(i_k)}$ . Note that the entries of  $W$  with indices  $i_1 < i_2 < \dots < i_k$  form an i.i.d. collection of Gaussian random variables, namely  $\{W_{i_1 \dots i_k}\}_{i_1 < i_2 < \dots < i_k}$  with distribution  $\mathcal{N}(0, 1/(k!))$ .

- the eigenvectors  $x_0$  and  $y_0$  are normalized and orthogonal, they belong to the Stiefel manifold  $\mathcal{M}_{n,2}$  given by

$$\mathcal{M}_{n,2} := \{(x, y) \in (S^{n-1})^2 : x \perp y\}.$$

Using the framework of Section 1.3.2, we are led to consider the profile likelihood random field

$$\begin{aligned}t &= (\theta, x, y) \in [0, 2\pi) \times \mathcal{M}_{n,2} \\ \psi_t &= \cos(\theta)x^{\otimes k} + \sin(\theta)y^{\otimes k} \\ Z(t) &= \langle \psi_t, Y \rangle_E = \cos(\theta)\langle x^{\otimes k}, Y \rangle_E + \sin(\theta)\langle y^{\otimes k}, Y \rangle_E.\end{aligned}$$

So the relevant manifold is  $M = S^1 \times \mathcal{M}_{n,2}$ , where the circle  $S^1$  is represented by  $[0, 2\pi)$ . The dimension of  $M$  is  $2n - 2$ . Hence we uncover

$$Y = \lambda_0 \psi_{t_0} + \sigma W,$$

with  $\lambda_0 = \sqrt{\nu_{0,1}^2 + \nu_{0,2}^2}$  and  $\theta = \arccos(\nu_{0,1}/\lambda_0)$ .

The covariance function at points  $s = (\theta', u, v)$  and  $t = (\theta, x, y)$  is given by

$$\begin{aligned}c(s, t) &= \mathbb{E}(Y(t)Y(s)) \\ &= \cos(\theta)\cos(\theta')\langle u^{\otimes k}, x^{\otimes k} \rangle_E + \sin(\theta)\sin(\theta')\langle v^{\otimes k}, y^{\otimes k} \rangle_E \\ &= \cos(\theta)\cos(\theta')\langle u, x \rangle^k + \sin(\theta)\sin(\theta')\langle v, y \rangle^k.\end{aligned} \quad (5.7)$$

In particular, it follows that  $c(t, t) = 1$  for all  $t \in M$  and

$$c((\theta', u, v), (\theta, x, y)) = c((\theta', Uu, Uv), (\theta, Ux, Uy)), \quad (5.8)$$

for any orthogonal map  $U$  in  $\mathbb{R}^n$ .

**Let us compute the matrix  $\Lambda_2(t)$ .** Note that because of the partial invariance by isometry given by (5.8),  $\Lambda_2(t)$  depends on  $\theta$  only. Let us define for distinct  $j \neq i, \ell \neq i$ :

$$\begin{aligned} Y_{[i]} &:= Y_{i, \dots, i}, \quad \text{with variance } 1; \\ Y_{[i], j} &:= Y_{i, \dots, i, j} + \dots + Y_{j, i, \dots, i} = kY_{i, \dots, i, j}, \quad \text{with variance } k; \\ Y_{[i], j, \ell} &:= \sum Y_{i_1, i_2, i_3} \quad \text{with } k-2 \text{ occurrences of } i, \text{ one } j \text{ and one } \ell, \quad \text{with variance } k(k-1); \\ Y_{[i], j, j} &\text{ has variance } \frac{k(k-1)}{2}. \end{aligned}$$

Because of the independence properties within  $Y$ , all these variables are independent. We consider the profile likelihood random field at point  $(\theta, e_1, e_2)$  where  $e_1$  and  $e_2$  are the first two elements of the canonical basis. We have

$$Z(\theta, e_1, e_2) = \cos(\theta)Y_{[1]} + \sin(\theta)Y_{[2]}.$$

The Euclidean gradient at  $(\theta, e_1, e_2)$  is given by

$$\begin{aligned} \frac{\partial Z(\theta, e_1, e_2)}{\partial x_1} &= k \cos(\theta)Y_{[1]}, & \frac{\partial Z(\theta, e_1, e_2)}{\partial y_2} &= k \sin(\theta)Y_{[2]}, \\ \frac{\partial Z(\theta, e_1, e_2)}{\partial x_j} &= \cos(\theta)Y_{[1], j}, j \neq 1, & \frac{\partial Z(\theta, e_1, e_2)}{\partial x_j} &= \sin(\theta)Y_{[2], j}, j \neq 2, \end{aligned}$$

and

$$\frac{\partial Z(\theta, e_1, e_2)}{\partial \theta} = -\sin(\theta)Y_{[1]} + \cos(\theta)Y_{[2]}.$$

Using the following orthonormal parameterization of the tangent space of the Stiefel manifold:

$$\begin{pmatrix} 0 & \frac{-\Delta_{21}}{\sqrt{2}} \\ \frac{\Delta_{21}}{\sqrt{2}} & 0 \\ \Delta_{31} & \Delta_{32} \\ \vdots & \vdots \\ \Delta_{n1} & \Delta_{n2} \end{pmatrix},$$

we get that the Riemannian gradient, including  $\theta$ , is given by

$$\nabla Z := \begin{pmatrix} \sin(\theta)Y_{[1]} + \cos(\theta)Y_{[2]} & \cdot \\ \frac{\cos(\theta)Y_{[1],2} - \sin(\theta)Y_{[2],1}}{\sqrt{2}} & \cdot \\ \cos(\theta)Y_{[1],3} & \sin(\theta)Y_{[2],3} \\ \vdots & \vdots \\ \cos(\theta)Y_{[1],n} & s Y_{[2],n} \end{pmatrix}. \quad (5.9)$$

Hence, we deduce that

$$\Lambda_2(\theta) = \text{diag}(1, k/2, \cos^2(\theta), \dots, \sin^2(\theta), \dots), \quad (5.10)$$

where  $\cos^2(\theta)$  and  $\sin^2(\theta)$  are repeated  $n-2$  times.

**Remark 6.** Observe that if  $\theta = k\pi/2$  then the matrix  $\Lambda_2(\theta)$  is singular. We prove in Lemma 8 of Appendix C that almost surely  $\theta_1 \neq k\pi/2$  where  $t_1 = (\theta_1, x_1, y_1)$ , hence  $\Lambda_2(t_1)$  is non-singular.

Also, Theorems 2 or 3 can not be directly applied. But, one can retrieve these results by omitting an  $\varepsilon$ -neighborhood of these points in the Kac-Rice formula and passing to the monotonic limit as  $\varepsilon$  tends to 0.

**Let us compute the matrix  $\Omega$**  Let us use for short  $c$  and  $s$  for  $\cos(\theta)$  and  $\sin(\theta)$  respectively. We start by computing the Euclidian Hessian at  $(\theta, e_1, e_2)$  as follows

$$\begin{aligned} \frac{\partial^2 Z(\theta, e_1, e_2)}{\partial x_1^2} &= ck(k-1)Y_{[1]}, & \frac{\partial^2 Z(\theta, e_1, e_2)}{\partial y_j \partial y_{j'}} &= s Y_{[2],j,j'}, j, j' \neq 2, \\ \frac{\partial^2 Z(\theta, e_1, e_2)}{\partial x_1 \partial x_j} &= c(k-1)Y_{[1],j}, j \neq 1, & \frac{\partial^2 Z(\theta, e_1, e_2)}{\partial \theta \partial x_1} &= ckY_{[1]}, \\ \frac{\partial^2 Z(\theta, e_1, e_2)}{\partial x_j \partial x_{j'}} &= cY_{[1],j,j'}, j, j' \neq 1, & \frac{\partial^2 Z(\theta, e_1, e_2)}{\partial \theta \partial x_j} &= ckY_{[1],j}, j \neq 1, \\ \frac{\partial^2 Z(\theta, e_1, e_2)}{\partial y_2^2} &= s k(k-1)Y_{[2]}, & \frac{\partial^2 Z(\theta, e_1, e_2)}{\partial \theta \partial y_2} &= s kY_{[2]}, \\ \frac{\partial^2 Z(\theta, e_1, e_2)}{\partial y_j \partial y_{j'}} &= s (k-1)Y_{[2],j}, j \neq 2, & \frac{\partial^2 Z(\theta, e_1, e_2)}{\partial \theta \partial y_j} &= s Y_{[2],j}, j \neq 2, \end{aligned}$$

and

$$\frac{\partial^2 Z(\theta, e_1, e_2)}{\partial \theta^2} = -Z.$$

Note that all the cross derivatives between  $x$  and  $y$  vanish. The Riemannian Hessian consists of four parts (it can be checked by an order two Taylor expansion along the tangent space) namely

- The projected Euclidean Hessian: the Euclidean Hessian restricted to the tangent space;
- For each of the three vectors,  $V_1, V_2, V_3$ , normal to the Stiefel manifold, the part of the second fundamental form associated to the vector multiplied by the normal derivative.

The 3 components of the second fundamental form are

$$\begin{aligned} V_1 &= - \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & \text{Id}_{n-2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ V_2 &= - \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \text{Id}_{n-2} \end{pmatrix} \\ V_3 &= - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \text{Id}_{n-2} \\ 0 & \frac{1}{\sqrt{2}} \text{Id}_{n-2} & 0 \end{pmatrix}. \end{aligned}$$

Hence, the expression of the Riemannian Hessian  $\nabla^2 Z(\theta, e_1, e_2)$  is

$$\begin{pmatrix} Z(\theta, e_1, e_2) & A & -s(Y_{[1],j})_{j>2} & c(Y_{[2],j})_{j>2} \\ \star & B & \frac{1}{\sqrt{2}}c(Y_{[1],2,j})_j & -\frac{1}{\sqrt{2}}s(Y_{[2],1,j})_j \\ \star & \star & c((Y_{[1],j,j'})_{j,j'} - kY_{[1]} \text{Id}_{n-2}) & -(cY_{[1],2} + sY_{[2],1}) \text{Id}_{n-2} / 2 \\ \star & \star & \star & s((Y_{[2],j,j'})_{j,j'} - kY_{[2]} \text{Id}_{n-2}) \end{pmatrix},$$

with  $A = \frac{1}{\sqrt{2}}(cY_{[2],1} - sY_{[1],2})$ , and  $B = -\frac{1}{2}(ckY_{[1]} + skY_{[2]}) + \frac{1}{2}(cY_{[1],2,2} + sY_{[2],1,1})$ .

To obtain  $\nabla^2 Z(t)$  for a generic point  $(\theta, x, y)$  belonging to  $\mathcal{M}_{n,2}$  we have to perform an isometry as in (5.8). Using (5.1), we deduce the expression of  $\Omega$ .

**Testing procedures** The test statistics involve the function  $G_{\Omega/\sigma}$  (resp. and  $H_{\Omega/\hat{\sigma}}$ ), recalled in Equation (5.2) (resp. and Equation (5.3)), which is given using an integral. This integration can be done numerically.

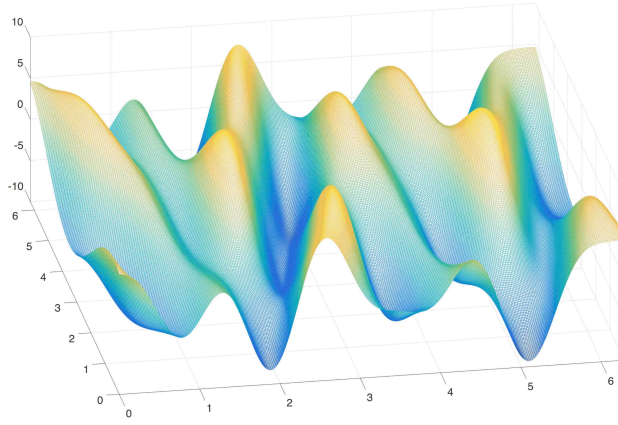


FIGURE 5. **[Super-Resolution, Section 5.4]** The super-resolution random field  $Z(\cdot)$  depicts the observation of a point source at location  $x_0$  on the interval  $[0, 2\pi)$  (spacial domain on the  $x$ -axis) with phase  $\theta_0$  given on the  $y$ -axis. The  $z$ -axis corresponds to value of the profile likelihood  $Z(x, \theta)$  at point  $x$  with phase  $\theta$ . This figure is presented in Azaï's et al. (2020, Fig. 2).

#### 5.4. Super resolution

In Signal processing, the super resolution phenomenon is the ability to distinguish two close sources (Dirac masses) from noisy low frequency measurements given by an optical system. This issue can be tackled using sparse regularization on the space of measures using the so-called Beurling-LASSO, introduced by De Castro and Gamboa (2012); Azaï's et al. (2015); Candès and Fernandez-Granda (2014); Duval and Peyré (2015).

We consider the framework of Azaï's et al. (2020) where one observes  $n = 2f + 1$  noisy frequencies between  $-f$  and  $f$  with  $f \geq 1$ . The  $\ell^{\text{th}}$  Fourier coefficient of a point source (Dirac mass) at location  $x_0$  with amplitude  $\lambda_0$  and phase  $\theta_0$  is  $\lambda_0 e^{i\theta_0} e^{-i\ell x_0}$ . The observation is given by  $Y = (y_\ell)_\ell$  where  $-f \leq \ell \leq f$ , they are complex random variables given by

$$y_\ell = \lambda_0 e^{i\theta_0} e^{-i\ell x_0} + \sigma W_\ell$$

with  $\lambda_0 > 0$  the amplitude,  $t_0 = (x_0, \theta_0) \in [0, 2\pi)^2$  the location and the phase of the source, and  $\sigma > 0$  the standard deviation of the noise. The noise is given by  $W_\ell = \xi_{\ell,1} + i\xi_{\ell,2}$  with  $\xi_{\ell,p}$  independent standard Gaussian variables,  $-f \leq \ell \leq f$  and  $p = 1, 2$ .

The profile likelihood is given by

$$Z(t) = \langle Y, \psi_t \rangle_E = \lambda_0 \cos(\theta - \theta_0) \mathbf{D}_f(x - x_0) + \sigma X(t)$$

with  $t = (x, \theta)$ ,  $M = [0, 2\pi)^2$  is the 2-Torus,  $E = \mathbb{C}^n$  equipped with the standard complex inner product,  $X(\cdot)$  is a (real valued) centered Gaussian random field with covariance function  $c(\cdot, \cdot)$ ,  $\mathbf{D}_f$  is the Dirichlet kernel. The feature map, the Dirichlet kernel and the covariance function are given by

$$\begin{aligned} \psi_t &= e^{i\theta} (e^{ifx}, \dots, e^{-ifx}) \in \mathbb{C}^n, \\ c(t, t') &= \cos(\theta - \theta') \mathbf{D}_f(x - x'), \\ \mathbf{D}_f(u) &= \frac{\sin(fu/2)}{\sin(u/2)}. \end{aligned}$$

An illustration of the super-resolution (profile likelihood) random field  $Z(\cdot)$  is given in Figure 5. One can check that the assumptions of the present article are satisfied by the super-resolution random field. The spacing test is given by

$$\mathbf{S}_\sigma = \frac{\sigma(\alpha_1 \lambda_1 + \alpha_2) \varphi(\lambda_1/\sigma) + (\alpha_1 \sigma^2 - \alpha_3^2) (1 - \Phi)(\lambda_1/\sigma)}{\sigma(\alpha_1 \lambda_2 + \alpha_2) \varphi(\lambda_2/\sigma) + (\alpha_1 \sigma^2 - \alpha_3^2) (1 - \Phi)(\lambda_2/\sigma)},$$

where  $\alpha_1, \alpha_2, \alpha_3$  have explicit forms given in (Azaï's et al., 2020, Proposition 10),  $\lambda_1$  is the maximum of  $Z(\cdot)$  and  $\lambda_2$  the maximum of  $Z^{t_1}(\cdot)$ . The  $t$ -spacing is explicitly described in Azaï's et al. (2020, Proposition 11).

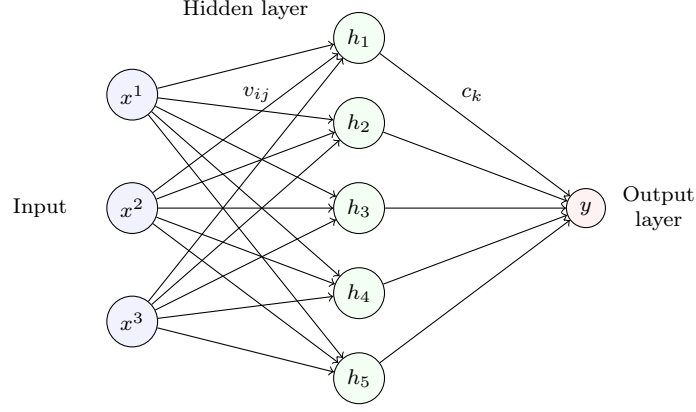


FIGURE 6. a two-layers neural network with  $c_k$  and  $v_{i,j}$  real numbers.

### 5.5. Two-layers neural networks with smooth rectifier

We observe a  $N$ -sample  $(x_i, y_i)_{i \in [N]}$ , where  $x_i \in \mathbb{R}^n$  is the input and  $y_i \in \mathbb{R}$  the output. The output is such that  $y_i = h(x_i) + \sigma w_i$  for some measurable target function  $h : \mathbb{R}^n \rightarrow \mathbb{R}$ , standard deviation  $\sigma > 0$ , and  $w_i$  standard Gaussian variable.

We consider a two layer neural network given by

- **(hidden layer)** a layer of  $r$  neurons with activation a  $C^2$  function  $\rho(\cdot)$ ;
- **(output layer)** and a layer which is simply a mean;

See Figure 6 for an illustration. As a consequence the output is  $\frac{1}{r} \sum_{k=1}^r c_k \rho(\langle v_k, x_i \rangle)$ .

The unknown parameters are the weights  $c_k$  and the directions  $v_k$  of the neurons, for  $k \in [r]$ . We assume that  $v_k \in S^{n-1}$ . Using the trick that the mean of  $r$  reals is the least squares estimator, we have to minimize

$$\frac{1}{N} \sum_{i=1}^N \left( y_i - \frac{1}{r} \sum_{k=1}^r c_k \rho(\langle v_k, x_i \rangle) \right)^2 = \frac{1}{Nr^2} \sum_{i=1}^N \sum_{k=1}^r \left( y_i - c_k \rho(\langle v_k, x_i \rangle) \right)^2,$$

and we recognize the mean-square training error in the left hand side and the Euclidean distance between  $Y = (y_i)_{i,k}$  and  $(c_k \rho(\langle v_k, x_i \rangle))_{i,k}$  on the right hand side in  $E = \mathbb{R}^{N \times r}$  equipped with the dot product

$$\langle (a_{i,k}), (b_{i,k}) \rangle_E = \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{r} \sum_{k=1}^r a_{i,k} \right) \left( \frac{1}{r} \sum_{k=1}^r b_{i,k} \right).$$

The Gaussian regression problem is  $Y = H + \sigma W$  where  $H = (h(x_i))_{i,k}$  and  $W = (W_i)_{i,k}$ . To enter into the framework of Section 1.3.2 we need to perform the following change of variables

$$c_k \rho(\langle v_k, x_i \rangle) = \lambda \frac{c_k V}{\lambda} \frac{\rho(\langle v_k, x_i \rangle)}{V} =: \lambda a_k \frac{\rho(\langle v_k, x_i \rangle)}{V},$$

with  $V^2 := \sum_{i=1}^N \rho^2(\langle v_k, x_i \rangle)$  and  $\lambda^2 := V^2(c_1^2 + \dots + c_r^2)$  are normalizing constants. One can check that  $t = (a_1, \dots, a_r, v_1, \dots, v_r) \in S^{r-1} \times (S^{n-1})^r$  with  $S^{r-1}$  the Euclidean sphere of  $\mathbb{R}^r$  and that  $M = S^{r-1} \times (S^{n-1})^r$ . We have

$$\psi_t := \left( \frac{a_k \rho(\langle v_k, x_i \rangle)}{\sqrt{\sum_{i'} \rho^2(\langle v_k, x_{i'} \rangle)}} \right)_{\substack{k \in [r] \\ i \in [N]}} \quad \text{and} \quad c(s, t) := \langle \psi_s, \psi_t \rangle_E,$$

and the profile likelihood is  $Z(t) = \langle H, \psi_t \rangle_E + \sigma X(t)$  with  $X(t) = \langle W, \psi_t \rangle_E$  having covariance function  $c(s, t)$ .

## References

- Adler, R. J. and Taylor, J. E. (2009). *Random fields and geometry*. Springer Science & Business Media.
- Armentano, D., Azaïs, J.-M., and León, J. R. (2023). On a general Kac-Rice formula for the measure of a level set. *ArXiv preprint*, abs/2304.07424.
- Auffinger, A. and Ben Arous, G. (2013). Complexity of random smooth functions on the high-dimensional sphere. *The Annals of Probability*, 41(6):4214–4247.
- Azaïs, J.-M., De Castro, Y., and Gamboa, F. (2015). Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195.
- Azaïs, J.-M., De Castro, Y., and Mourareau, S. (2017). A rice method proof of the null-space property over the grassmannian. In *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*.
- Azaïs, J.-M., De Castro, Y., and Mourareau, S. (2020). Testing gaussian process with applications to super-resolution. *Applied and Computational Harmonic Analysis*, 48(1):445–481.
- Azaïs, J.-M. and Wschebor, M. (2004). Upper and lower bounds for the tails of the distribution of the condition number of a gaussian matrix. *SIAM Journal on Matrix Analysis and Applications*, 26(2):426–440.
- Azaïs, J.-M. and Wschebor, M. (2005). On the distribution of the maximum of a gaussian field with  $d$  parameters. *The Annals of Applied Probability*, 15(1A):254–278.
- Azaïs, J.-M. and Wschebor, M. (2009). *Level sets and extrema of random processes and fields*. John Wiley & Sons Inc.
- Candès, E. J. and Fernandez-Granda, C. (2014). Towards a mathematical theory of super-resolution. *Communications on pure and applied Mathematics*, 67(6):906–956.
- Chizat, L. (2022). Sparse optimization on measures with over-parameterized gradient descent. *Mathematical Programming*, 194(1):487–532.
- De Castro, Y., Gadat, S., Marteau, C., and Maugis-Rabusseau, C. (2021). Supermix: sparse regularization for mixtures. *The Annals of Statistics*, 49(3):1779–1809.
- De Castro, Y. and Gamboa, F. (2012). Exact reconstruction using beurling minimal extrapolation. *Journal of Mathematical Analysis and applications*, 395(1):336–354.
- Duval, V. and Peyré, G. (2015). Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2):407 – 499.
- Lifshits, M. A. (1983). On the absolute continuity of distributions of functionals of random processes. *Theory of Probability & Its Applications*, 27(3):600–607.
- Perry, A., Wein, A. S., and Bandeira, A. S. (2020). Statistical limits of spiked tensor models. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(1):230 – 264.
- Steinwart, I. and Christmann, A. (2008). *Support vector machines*. Springer Science & Business Media.

# Appendix of Second Maximum of a Gaussian Random Field and Exact ( $t$ -)Spacing test

In this appendix, the Riemannian Hessians are represented by 2-forms.

## Appendix A: The helix random field

This section is a proof of Lemma 3. The argument is inspired from [Azaïs and Wschebor \(2005, Lemma 4.1\)](#). Let  $t \in M$  be a fixed point. On  $L^2$ , let  $\Pi$  be the projector on the orthogonal complement to the subspace generated by  $(X(t), \nabla X(t))$ . Consider  $h$  a vector of the tangent space at point  $t$ . Consider the exponential map  $\varepsilon \mapsto \exp_t(\varepsilon h)$ . This function is well defined on a neighborhood of 0. Hence there exists  $\varepsilon_0 > 0$  such that for all  $\varepsilon \in ]\varepsilon_0, \varepsilon_0[$ , the point  $s(\varepsilon) := \exp_t(\varepsilon h) \in M$  exists. The function  $\varepsilon \mapsto s(\varepsilon)$  is a parametrization of the geodesic starting at point  $t$  with velocity  $h$ . We denote by  $\varepsilon \mapsto h(\varepsilon)$  the parallel transport of  $h$  along this geodesic. By Assumption **(A<sub>1</sub>)**, the Taylor formula of order two gives

$$X(s(\varepsilon)) = X(t) + \varepsilon \langle \nabla X(t), h \rangle + \frac{\varepsilon^2}{2} \nabla^2 X(s(\varepsilon'))[h(\varepsilon'), h(\varepsilon')], \quad (\text{A.1})$$

for some  $\varepsilon' \in [0, \varepsilon]$ , and

$$c(s(\varepsilon), t) = 1 + \frac{\varepsilon^2}{2} \Lambda_2(t)[h, h] + o(\varepsilon^2), \quad (\text{A.2})$$

by Assumption **(A<sub>2</sub>)**. From **(A.1)**, we have

$$\Pi(X(s(\varepsilon))) = \frac{\varepsilon^2}{2} \tilde{R}(s(\varepsilon'))[h(\varepsilon'), h(\varepsilon')],$$

Note also that  $-\Pi(X(s(\varepsilon)))$  is the numerator of  $X^{|t}(s(\varepsilon))$  while the denominator is given by  $-\frac{\varepsilon^2}{2} \Lambda_2(t)[h, h] + o(\varepsilon^2)$  thanks to **(A.2)**. We deduce that

$$X^{|t}(s(\varepsilon)) = \frac{\tilde{R}(s(\varepsilon'))[h(\varepsilon'), h(\varepsilon')]}{\Lambda_2(t)[h, h] + o(1)}$$

From **(2.6)** and passing to the limit, we deduce that

$$\lim_{\varepsilon \rightarrow 0} X^{|t}(s(\varepsilon)) = \frac{\tilde{R}(t)[h, h]}{\Lambda_2(t)[h, h]},$$

invoking that  $\tilde{R}$  is continuous by regression formulas and Assumption **(A<sub>1</sub>)**, and that  $\Lambda_2(t)[h, h]$  is positive by Assumption **(A<sub>4</sub>)**.

For the second and last statement, observe that

$$\limsup_{s \rightarrow t} X^{|t}(s) \geq \sup_{\|h\|_2=1} \lim_{\varepsilon \rightarrow 0} X^{|t}(\exp_t(\varepsilon h)) = \max_{\|h\|_2=1} \frac{\tilde{R}(t)[h, h]}{\Lambda_2(t)[h, h]} = \frac{\tilde{R}(t)[h_0, h_0]}{\Lambda_2(t)[h_0, h_0]}$$

where the vector  $h_0$  exists by continuity of  $h \mapsto \tilde{R}(t)[h, h]/\Lambda_2(t)[h, h]$  on the Euclidean sphere, which is compact. Now, let  $\delta > 0$  and let  $s_n$  be a sequence such that  $s_n \rightarrow t$  and

$$\lim_{n \rightarrow \infty} X^{|t}(s_n) \geq \limsup_{s \rightarrow t} X^{|t}(s) - \delta.$$

Note that the exponential map is a local diffeomorphism on a neighborhood of the point  $t$ . Hence, there exist a sequence of positive reals  $\varepsilon_n$  converging to zero and a sequence  $h_n$  of unit norm tangent vectors such that  $s_n = \exp_t(\varepsilon_n h_n)$ . Since the Euclidean sphere is compact, we can extract a sequence such that  $h_n$  converges to unit norm tangent vector  $\bar{h}$ . The Taylor formula gives that

$$\begin{aligned} X(s_n) &= X(t) + \varepsilon_n \langle \nabla X(t), h_n \rangle + \frac{\varepsilon_n^2}{2} \nabla^2 X(\tilde{s}_n)[\tilde{h}_n, \tilde{h}_n], \\ c(s_n, t) &= 1 + \frac{\varepsilon_n^2}{2} \Lambda_2(t)[h_n, h_n] + o(\varepsilon_n^2), \end{aligned}$$



for some  $\tilde{s}_n$  on the geodesic between  $t$  and  $s_n$  and  $\tilde{h}_n$  the parallel transport at point  $\tilde{s}_n$  of the tangent vector  $h_n$  along this geodesic. By the  $\Pi$ -projection argument above, we deduce that

$$X|t(s_n) = \frac{\tilde{R}(\tilde{s}_n)[\tilde{h}_n, \tilde{h}_n]}{\Lambda_2(t)[h_n, h_n] + o_n(1)}.$$

Passing to the limit by continuity yields

$$\max_{\|h\|_2=1} \frac{\tilde{R}(t)[h, h]}{\Lambda_2(t)[h, h]} \geq \frac{\tilde{R}(t)[\bar{h}, \bar{h}]}{\Lambda_2(t)[\bar{h}, \bar{h}]} = \lim_{n \rightarrow \infty} X|t(s_n) \geq \limsup_{s \rightarrow t} X|t(s) - \delta.$$

Note that the most left hand side term does not depends on  $\delta > 0$ , hence

$$\max_{\|h\|_2=1} \frac{\tilde{R}(t)[h, h]}{\Lambda_2(t)[h, h]} \geq \limsup_{s \rightarrow t} X|t(s),$$

which concludes the proof.

## Appendix B: Ad-hoc Kac-Rice formula

The paper [Azaïs and Wschebor \(2009, Theorem 7.2\)](#) concerns weighted sum of number of roots when the weight is a continuous function of time and of the level. This has been extended, by a monotone convergence argument to the case of lower semi-continuous weights in [Armentano et al. \(2023, Section 7\)](#). However this is not sufficient mainly because the regularity of  $\lambda_2^t$  as a function of  $t$  is difficult to control. For this reason we must used the following tailored argument.

Denote by  $\text{dist}(s, t)$  the geodesic distance between points  $s$  and  $t$ . Define

$$\forall t \in M, \quad \lambda_{2,\epsilon}^t := \sup_{s \in M \text{ s.t. } \text{dist}(s,t) > \epsilon} \sigma X|t(s) \quad \text{and} \quad \lambda_{2,\epsilon} := \lambda_{2,\epsilon}^{t_1}, \quad (\text{B.1})$$

and note that  $\lambda_{2,\epsilon}^t$  is a continuous function of  $t$ . Define a monotone approximation  $\xi_n(\cdot)$  of the indicator function  $\mathbb{1}\{\cdot \in B\}$ . Then

$$\xi_n(\sigma X(t), \lambda_{2,1/n}^t, \sigma R(t)) \uparrow \mathbb{1}\{(\sigma X(t), \lambda_2^t, \sigma R(t)) \in B\}. \quad (\text{B.2})$$

So we can use [Armentano et al. \(2023, Section 7\)](#) for instance to compute the expectation of the left side of (B.2). Indeed, all conditions are clear except Condition c) of [Armentano et al. \(2023, Section 7\)](#) for  $\lambda_{2,\epsilon}$  which is detailed hereunder. And then use monotone convergence theorem gives the result.

**Checking Condition c) of Armentano et al. (2023, Section 7)** Using regression formulas, the distribution of  $X(\cdot)$  under the condition  $X(t_0) = u$  admits the representation

$$X(t) = \tilde{X}(t) + uf(t),$$

where  $\tilde{X}(t)$  corresponds to the distribution conditional to  $X(t_0) = 0$ . Our goal is to show the continuity of the distribution of  $\lambda_2$  in this representation. The conditioned random field  $\tilde{X}(\cdot)$  satisfies

$$\forall \{s \neq t\}, \quad \text{Cor}(X(s), X(t)) < 1.$$

so that, by the Tsirelson theorem, the maximum of  $\tilde{X}(\cdot) + uf(\cdot)$  is a.s. unique. The first consequence is the continuity, as a function of  $u$ , of  $\hat{t}$ . For  $t$  fixed, under  $X(t_0) = u$ ,

$$X|t(s) = \tilde{X}|t(s) + ug|t(s), \quad \text{dist}(s, t) > \epsilon,$$

with obvious notation. The proof of Lemma 2 shows that  $g|t(s)$  is bounded uniformly in  $s$  and  $t$ . This gives the desired continuity, as a function of  $u$ , of the distribution of  $\lambda_{2,\epsilon}^t$  and then of  $\lambda_{2,\epsilon}$ .

## Appendix C: Two-spiked tensor profile likelihood

**Lemma 8.** *With probability 1, it holds that  $\theta_1 \neq k\pi/2$ .*

*Proof.* We prove that  $\theta_1 \neq 0$  a.s., the other cases are equivalent. Note that  $\theta_1 = 0$  implies that there exists a point  $x_1$  such that  $f(x_1)$  is the maximum of  $f(x) := \langle x^{\otimes k}, Y \rangle_E$  on  $\mathbb{S}^{n-1}$ . As a consequence the gradient along  $\mathbb{S}^{n-1}$  is zero. Now, note that the derivative with respect to  $\theta$  at  $\theta = 0$  of  $Z(\theta, x_1, y)$  is zero for every  $y$  orthogonal to  $x_1$ . Choosing an orthonormal basis we obtain  $n - 1$  such derivatives.

Then, we use the Bulinskaya lemma (Azaïs and Wschebor, 2009, Prop 6.11). We denote by  $F(x), x \in \mathbb{S}^{n-1}$  the random field defined on a set of dimension  $n - 1$  with values in  $\mathbb{R}^{2n-2}$  given by the  $2n - 2$  derivatives above. To use the Bulinskaya lemma we need to prove that

- (i) the function  $F(\cdot)$  has  $\mathcal{C}^1$  paths,
- (ii) the density  $p_{F(x)}$  is uniformly bounded.

Then since the dimension of the parameter set is smaller than the dimension of the image set, a.s there is no point  $x$  such that  $F(x) = u$  where  $u$  is any value and in particular the value 0.

Condition (i) is clear. To address Condition (ii), we can use the invariance by isometry and study the density  $p_{F(x)}$  at the particular value  $x = e_1$ . We can consider the derivatives of  $f(x)$  at  $e_1$  along the basis of the tangent space  $e_2, \dots, e_n$ . We obtain  $Y_{2[1]}, \dots, Y_{n[1]}$ . Then we consider the derivatives in  $\theta$  of  $X(\theta, x_1, y)$  with  $y = e_2, \dots, e_n$ . We obtain  $Y_{[2]}, \dots, Y_{[n]}$ . All these variables are independent with fixed variance so the density of  $F(x)$  is bounded.  $\square$