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

Jean-Marc Azaïs[1] **Federico Dalmao**[2] **Yohann De Castro** [3,4]

[1] *Institut de Mathématiques de Toulouse*
Université Paul Sabatier, France, e-mail: jean-marc.azais@math.univ-toulouse.fr

[2] *DMEL, CENUR Litoral Norte*
Universidad de la República, Salto, Uruguay, e-mail: fdalmao@unorte.edu.uy

[3] *Institut Camille Jordan, CNRS UMR 5208*
École Centrale Lyon, France, e-mail: yohann.de-castro@ec-lyon.fr

[4] *Institut Universitaire de France (IUF)*

Abstract:
In this article, we introduce the novel concept of the second maximum of a Gaussian random field defined on a Riemannian manifold. This statistic serves as a powerful tool for characterizing the distribution of the global maximum. By employing a tailored Kac-Rice formula, we derive the explicit distribution of the maximum, conditioned on the second maximum and the independent part of the Riemannian Hessian. This approach yields an exact test based on the spacing between these maxima, which we refer to as the spacing test.

We investigate the applicability of this test for detecting sparse alternatives in Gaussian symmetric tensors, continuous sparse deconvolution, and two-layered neural networks with smooth rectifiers. Theoretical results are supported by numerical experiments that illustrate the calibration and power of the proposed procedures. More generally, we provide a framework for applying the spacing test to continuous sparse regression for any Gaussian random field on a Riemannian manifold.

Furthermore, when the variance-covariance function is known only up to a scaling factor, we derive an exact Studentized version of the procedure, coined the t-spacing test. We show that this test is perfectly calibrated under the null hypothesis and exhibits high power for detecting sparse alternatives. As a corollary, we derive an exact, non-asymptotic test for the spiked tensor model without requiring prior knowledge of the noise level.

MSC2020 subject classifications: Primary 62E15, 62F03, 60G15, 62H10; secondary 60E05, 60G10, 62J05.
Keywords and phrases: Gaussian random field, Second maximum, Spacing test, Kernel regression, Tensor PCA.

Version of December 5, 2025

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

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

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 $\mathrm{Var}[X(t)] = 1$ for every $t \in M$. We consider the simple global null hypothesis $\mathbb{H}_0$: "$m(\cdot) = 0$", even when the standard error σ is unknown. The test investigated is the quantile $\widehat{\alpha}$ of the maximum λ_1 under the null defined as

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

and we denote by $t_1 \in M$ its argument maximum.

arXiv: 2406.18397v2

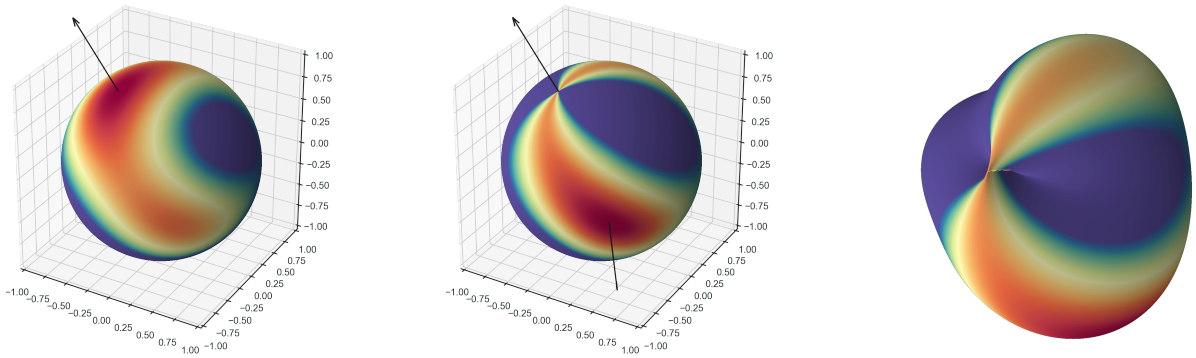

FIGURE 1. **[Spiked tensor PCA, Section 5.2, example 1/4]** *Visualisation of the random fields on the sphere* $\mathbb{S}^2$ *for the Spiked Tensor PCA problem.* ***Left:*** *The Gaussian homogeneous polynomial* $X(\cdot)$*, where the arrow indicates the global maximizer* t_1 *(first eigenvector).* ***Middle:*** *The conditional random field* $X^{|t_1}(\cdot)$ *defined in (1.4), where the arrow indicates the second maximizer* t_2*.* ***Right:*** *A volumetric view of* $X^{|t_1}(\cdot)$ *where the radial height represents the field value.*

This quantile is built from the distribution of λ_1 conditional on the values of the so-called second maximum λ_2 and the so-called independent part of the Riemannian Hessian Ω, under the null. Suppose, in a first step, that σ 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,

$$\widehat{\alpha} := \frac{\displaystyle\int_{\lambda_1/\sigma}^{+\infty} \det(u\mathrm{Id} - \Omega/\sigma)\varphi(u)\,\mathrm{d}u}{\displaystyle\int_{\lambda_2/\sigma}^{+\infty} \det(u\mathrm{Id} - \Omega/\sigma)\varphi(u)\,\mathrm{d}u}\,, \tag{1.3}$$

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

We will show the exactness of this test, meaning that $\widehat{\alpha}$ 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 λ_1 conditional on λ_2 and Ω, under the null hypothesis. Small values of $\widehat{\alpha}$ indicate that the maximum λ_1 is abnormally large compared to the value of the second maximum λ_2, making $\widehat{\alpha}$ interpretable as a p-value. This test is referred to as the "spacing test". The second maximum is defined as

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

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

where $c(\cdot,\cdot)$ is the variance-covariance function of $X(\cdot)$ and $\Lambda_2(t) := \mathrm{Var}[\nabla X(t)]$, 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.

Furthermore, we will show how to achieve the same result when the variance σ^2 is unknown using an estimation $\widehat{\sigma}^2$ 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 $\lambda_1/\widehat{\sigma}$ and $\lambda_2/\widehat{\sigma}$. Note that λ_1 and λ_2 are the same for the spacing and the t-spacing test, and these values can be computed without knowing σ. These results are supported by numerical experiments as illustrated in Figure 2.

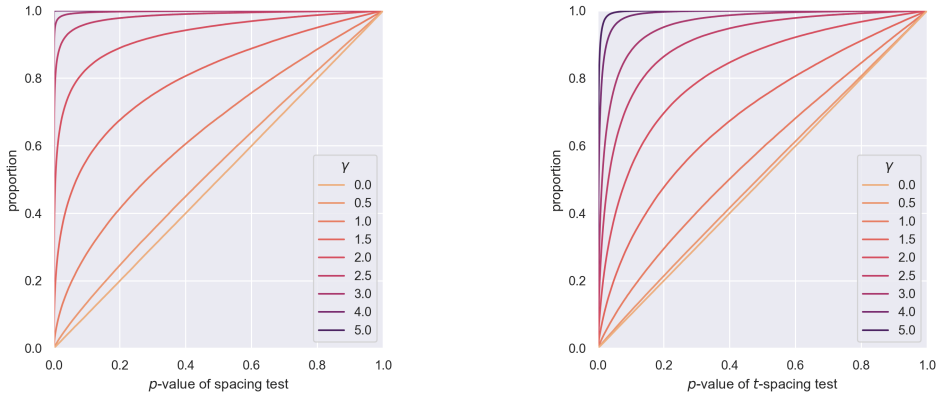

FIGURE 2. **[Spiked tensor PCA, Section 5.2, example 2/4]** *The CDFs of the p-value of (t-)spacing tests over* $250,000$ *Monte-Carlo samples for each value of* γ*. Note that these tests are perfectly calibrated under the null hypothesis, for which* $\gamma = 0$*. The parameter* γ *is a scaling factor 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* σ *has been estimated on* $X^{|t_1}(\cdot)$*.*

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

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

and we denote one-sparse alternatives by, for some $\lambda_0 \neq 0$ and $t_0 \in M$ unknown,

$$\mathbb{H}_1(t_0, \lambda_0) \ : \text{“ } m(\cdot) = \lambda_0 c(\cdot, t_0) \text{”} \,. \tag{1.5}$$

Indeed, as detailed in Section 1.3.3, the maxima λ_1 and λ_2 correspond to *the knots of the Continuous LARS algorithm and are intrinsic to continuous kernel sparse regression*. We illustrate this behavior in Figure 3 using the Spiked Tensor PCA model. In this experiment, the alternative is parameterized by λ_0, where γ scales the signal-to-noise ratio.[1] As illustrated in the top panels, λ_1 stochastically increases with γ, while the distribution of λ_2 remains stable for moderate signal strengths (e.g., $\gamma \geq 2$). Consequently, the spacing $\lambda_1 - \lambda_2$ grows linearly with γ. In this moderate regime, the bottom panels confirm that the maximizer t_1 concentrates around the true parameter t_0 ($t_1 \simeq t_0$) and that λ_1 follows the distribution of $\sigma X(t_0)$ under $\mathbb{H}_1(t_0, \lambda_0)$ (Gaussian with mean λ_0 and variance σ^2). This implies that (t_1, λ_1) weakly converges to the distribution of $(t_0, \sigma X(t_0))$, allowing the (t-)spacing tests to successfully detect the alternative. We leave the formal proof of this specific phenomenon for future work.

Comparison with asymptotic phase transitions. It is crucial to distinguish our contributions from well-established statistical limits in Spiked Tensor PCA. The existing literature, such as Perry et al. (2020, Theorem 1.3), proves that detecting a rank-one tensor below a specific eigenvalue threshold is impossible in the asymptotic regime where the dimension $d \to \infty$. In contrast, the (t-)spacing test proposed here provides an *exact, non-asymptotic* inference procedure valid for any fixed dimension d and sample size. While our numerical experiments (Figure 3) recover the phenomenological behavior of this phase transition—where power drops to zero below the theoretical threshold—our procedure guarantees perfect calibration under the null hypothesis regardless of signal strength or dimension. Furthermore, unlike theoretical thresholds that typically require knowledge of the noise level σ, our t-spacing test *remains exact even when the noise variance is unknown*. To the best of our knowledge, this is the first exact non-asymptotic test for the spiked tensor model with unknown σ. Finally, while Tensor PCA serves

[1] The signal is $\lambda_0 = \gamma \times \sigma\sqrt{3\log 3 + 3\log\log 3}$. Here, $\gamma = 1$ corresponds to the phase transition threshold derived in Perry et al. (2020, Theorem 1.3). Note that in their framework, the threshold is $\sqrt{2\log 3 + 2\log\log 3}$; we use a different normalization of the Gaussian noise, resulting in a statistical threshold of $\sigma\sqrt{d\log k + d\log\log k}$. Our Gaussian noise is presented in (5.5b) and differs by $\sigma\sqrt{d/2}$ from theirs (see (Montanari et al., 2015, Equation 10)), where σ is an unknown noise level.

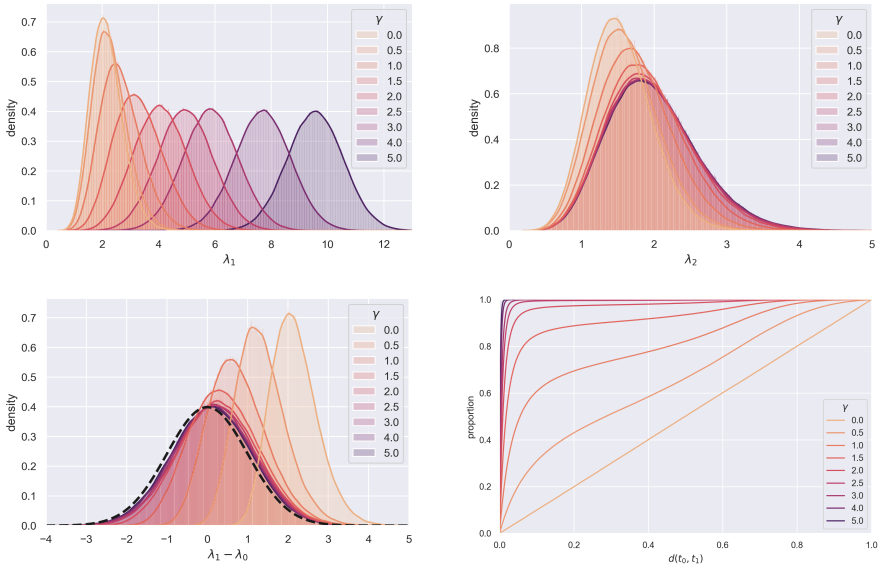

Figure 3. **[Spiked tensor PCA, Section 5.2, example 3/4]** *The PDFs of the maxima λ_1, λ_2 and the CDF of the distance $\mathrm{d}(t_0, t_1)$ over $250,000$ Monte-Carlo samples for each value of the parameter γ. The alternative is given by t_0 fixed and $\lambda_0 = \gamma \times \sigma\sqrt{3\log 3 + 3\log\log 3} \simeq 0.684\gamma\sigma$ 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 $\mathrm{d}(t_0, t_1)$ is normalized so that it is uniformly distributed on $(0, 1)$ if t_1 is uniformly distributed on the sphere (e.g., $\gamma = 0$).*

as a primary motivating example, our framework applies to the broader context of Gaussian random fields on Riemannian manifolds and general continuous sparse kernel regression. Future work may explore the precise links between our method and maximum likelihood detection thresholds (see Section 1.3.2).

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 $c(\cdot, \cdot) : M \times M \to \mathbb{R}$ denotes its variance-covariance function.

Assumption $(\mathbf{A_{1\text{-}4}})$ We assume that

- the *paths of* $X(\cdot)$ *are* $\mathcal{C}^2$ *almost surely*; $(\mathbf{A_1})$
- $\mathbb{E}[X(t)] = 0$ and $\mathrm{Var}[X(t)] = 1$, for every $t \in M$; $(\mathbf{A_2})$
- $\forall s \neq t\,,\ c(s,t) < 1$; $(\mathbf{A_3})$
- for every $t \in M$, the gradient $\nabla X(t)$ has a non-degenerate Gaussian distribution. $(\mathbf{A_4})$

The variance-covariance matrix of the Gaussian tangent vector $\nabla X(t)$ is denoted by $\Lambda_2(t)$, which is invertible by Assumption $(\mathbf{A_4})$.

1.3.2. A continuous regression framework for the spacing test

General framework

We consider the following regression problem given by

$$Y = \lambda_0 \psi_{t_0} + \sigma W\,, \tag{1.6}$$

where $\lambda_0 \in \mathbb{R}$, $t_0 \in \mathrm{M}$ are unknown parameters; σ 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 Id_E. The *feature map* ψ satisfies the following Assumptions ($\mathbf{B_{1\text{-}6}}$):

- ψ is a $\mathcal{C}^3$-function; ($\mathbf{B_1}$)
- $\forall t \in M\,,\ \|\psi_t\|_E^2 = 1$; ($\mathbf{B_2}$)
- $\forall t \in M\,,\ \forall s \in M \setminus \{t\}\,,\ \langle \psi_s, \psi_t\rangle_E < 1$; ($\mathbf{B_3}$)
- $\forall t \in M\,,\ J\psi_t {J\psi_t}^\top$ has full rank d; ($\mathbf{B_4}$)
- the span of $\{\psi_t \ :\ t \in M\}$ has dimension m such that $m > d+1$; ($\mathbf{B_5}$)
- $\forall t \in M\,,\ \exists s \in M$ s.t. $\psi_t = -\psi_s$; ($\mathbf{B_6}$)

where $J\psi_t$ is the Jacobian matrix of ψ_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

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

One can check that ($\mathbf{B_1}$) implies ($\mathbf{A_1}$), ($\mathbf{B_2}$) is equivalent to ($\mathbf{A_2}$), ($\mathbf{B_3}$) is equivalent to ($\mathbf{A_3}$), ($\mathbf{B_4}$) is equivalent to ($\mathbf{A_4}$), and ($\mathbf{B_5}$) is equivalent to the forthcoming Assumption ($\mathbf{A_5}$).

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* $\mathrm{M} = \mathbb{S}^2$*. The Euclidean space of 3-way symmetric tensors of size* $3\times3\times3$ *is denoted by* $(E, \langle\cdot,\cdot\rangle_E)$ *with dimension* $m = 10$*, and*

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

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,* $1 \le i \ne j \le 3$;
- W_{123} *of variance* $1/6$; *there is* 1 *off-diagonal term*.

Maximum likelihood estimator (MLE)

The Maximum Likelihood Estimator (MLE) of the couple (λ_0, t_0) is given by

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

Minimization with respect to λ is straightforward under Assumptions ($\mathbf{B_2}$) and ($\mathbf{B_6}$) leading to

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

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 λ_0) by (1.7), namely

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

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

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

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

Remark 2. *Assumption* ($\mathbf{B_6}$) *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 (λ_i, t_i) of the so-called knots. The first knot (λ_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

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

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

$$\forall t \in \{1, \dots, p\}\,, \quad Z^\lambda(t) := \langle Y - (\lambda_1 - \lambda)\psi_{t_1}, \psi_t\rangle_{\mathbb{R}^n}\,, \tag{1.9b}$$

The second knot is defined by

$$\lambda_1 - \lambda_2 := \inf\left\{\varepsilon > 0 \ : \ \exists t \neq t_1 \text{ s.t. } Z^{\lambda_1-\varepsilon}(t) \ge \lambda_1 - \varepsilon\right\}. \tag{1.9c}$$

The second knot (λ_2, t_2) is built so that t_1 is the unique argument maximum of $Z^\lambda(\cdot)$ for $\lambda_2 < \lambda \le \lambda_1$ 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 A.1 for further details. We introduce the continuous LARS in a general setting as follows. The first knot (λ_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

$$\lambda_1 = \max_{t\in M} |\langle Y, \psi_t\rangle_E| \quad \text{and} \quad t_1 = \arg\max_{t\in M} |\langle Y, \psi_t\rangle_E| \,. \tag{1.10a}$$

Under ($\mathbf{B_6}$), 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 λ_1 given in (1.2) coincides with the aforementioned definition of λ_1. One can choose t_1 being the unique argument maximum of $Z(\cdot)$, by ($\mathbf{B_1}$) and ($\mathbf{B_3}$) (see also (2.1d) in Lemma 1). We define the so-called residual by, for all $\lambda \le \lambda_1$,

$$Z^\lambda(t) := \langle Y - (\lambda_1 - \lambda)\psi_{t_1}, \psi_t\rangle_E = Z(t) - (\lambda_1 - \lambda)c(t, t_1)\,, \tag{1.10b}$$

where $c(s,t) = \langle \psi_s, \psi_t\rangle_E$. Again, under ($\mathbf{B_6}$), 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

$$\lambda_1 - \lambda_2 := \inf\left\{\varepsilon > 0 \ : \ \exists t \neq t_1 \text{ s.t. } Z^{\lambda_1-\varepsilon}(t) \ge \lambda_1 - \varepsilon\right\}. \tag{1.10c}$$

The second knot (λ_2, t_2) is built so that t_1 is the unique argument maximum of $Z^\lambda(\cdot)$ for $\lambda_2 < \lambda \le \lambda_1$ 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 λ,

$$\begin{aligned} Z^\lambda(t) \le \lambda \ &\Leftrightarrow \ Z(t) - \lambda_1 c(t,t_1) \le \lambda(1 - c(t,t_1)) \\ &\Leftrightarrow \ \frac{Z(t) - Z(t_1)c(t,t_1)}{1 - c(t,t_1)} \le \lambda \\ &\Leftrightarrow \ Z^{|t_1}(t) \le \lambda\,, \end{aligned}$$

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 (λ_2, t_2) is the second knot of the LARS where

$$\lambda_2 = \max_{t\in M\setminus\{t_1\}} Z^{|t_1}(t) \quad \text{and} \quad t_2 = \max_{t\in M\setminus\{t_1\}} Z^{|t_1}(t)\,. \tag{1.10d}$$

Note that the definition of λ_2 in (1.4) is equivalent to the aforementioned definition of λ_2 (this is also true for the argument maximum). This paper gives the distribution of λ_1 (first knot of the continuous LARS) conditional on (λ_2, Ω) (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

Main contributions.

We investigate the *second maximum* λ_2 of a $\mathcal{C}^2$ centered Gaussian random field on a compact manifold M, a quantity naturally arising along the continuous LARS path of sparse kernel regression. Classical inference relying only on the global maximum λ_1 typically uses tail or Euler characteristic approximations; here, by conditioning on the richer object (λ_2, Ω)—where Ω denotes the independent part of the Hessian at the (random) maximiser t_1—we obtain *exact, non-asymptotic* distributional identities for λ_1 and devise spacing tests. Our main results are:

(C1) **Geometric characterisation of λ_2.** Definition of λ_2 as the maximum of the singular conditional field $X^{|t_1}(\cdot)$ with a removable first-order structure, and study of its basic properties (Section 2).
(C2) **Helix / eigen-structure link.** A correspondence (Lemma 3) between the eigen-decomposition of Ω and directional limiting values of the rescaled field near t_1, clarifying local geometry around the global maximum.
(C3) **Singular Kac–Rice formula.** An ad-hoc Kac–Rice formula adapted to $X^{|t_1}(\cdot)$ yielding the explicit conditional density of λ_1 given (λ_2, Ω) (Theorem 1, Appendix C).
(C4) **Exact spacing test (known variance).** A finite-sample valid test based on the spacing $\lambda_1 - \lambda_2$ (Theorem 2).
(C5) **Variance estimation.** Construction of a Karhunen–Loève based estimator $\widehat{\sigma}$ for incorporation into the Kac–Rice framework (Section 4.4).
(C6) **Studentised spacing test.** Exact inference using the studentised spacing $\lambda_1/\widehat{\sigma} - \lambda_2/\widehat{\sigma}$ without prior knowledge of σ^2 (Theorem 3, Proposition 7).
(C7) **Empirical validation and applications.** Simulations confirm calibration and improved power for sparse detection (`https://github.com/ydecastro/tensor-spacing/`); links with continuous sparse kernel regression and super-resolution are detailed in Section 5.

Conceptually, the work extends Kac–Rice formula to a controlled singular setting and shows that conditioning on (λ_2, Ω) unlocks exact finite-sample inference for λ_1. Statistically, spacing-based tests can be studentized leading to exact inference without prior knowledge of σ^2. Geometrically, the helix interpretation ties curvature encoded in Ω to directional approach of the conditional field.

Outline.

Section 2 introduces λ_2 and the conditional singular field $X^{|t_1}(\cdot)$. Section 2.3 presents the singular Kac–Rice formula and Theorem 1. Section 3 develops spacing-based tests for known variance. Section 4.4 constructs $\widehat{\sigma}$ and derives the plug-in law leading to the unknown-variance test. Section 5 details connections with continuous sparse kernel regression and super-resolution. Appendices collect technical lemmas, weight computations and further numerical illustrations. A table of list of notation is given in Appendix F.

1.5. Related works

The first paper to consider the Kac-Rice formula on manifolds, specifically the sphere and the Stiefel Manifold, is Azaïs and Wschebor (2005). This theory has been comprehensively developed in the monographs by Adler and Taylor (2009) and Azaïs and Wschebor (2009). In high-dimensional settings, complexity results for random smooth functions on the sphere were derived by Auffinger and Ben Arous (2013). A new set of hypotheses and proofs for the Kac-Rice formula, particularly for the measure of level sets, is given in Armentano et al. (2023).

The Kac-Rice formula is also central to Azaïs et al. (2017) for studying the maximum of Gaussian fields on the torus, with applications to the Super-Resolution problem. This falls under the general theory of continuous sparse regression over the space of measures, which has attracted significant attention in signal processing (Candès and Fernandez-Granda, 2014; Duval and Peyré, 2015; Azaïs et al., 2015), machine learning (De Castro et al., 2021), and optimization (Chizat, 2022). The super-resolution framework aims to recover fine-scale details from low-frequency measurements and has applications in astronomy, medical imaging, and microscopy. The novel aspects of this body of work rely on new statistical and optimization guarantees for sparse regression. Initiated by the work presented in (Azaïs et al., 2020), we investigate the possibility of detecting a sparse object from a test on the mean $m(\cdot)$ of a Gaussian random field under a sparsity assumption.

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* ($\mathbf{A_{1\text{-}4}}$)*, one has*

$$\circ\ c(s,t) = \mathbb{E}[X(s)\,X(t)] \text{ and } c(t,t) = 1; \tag{2.1a}$$
$$\circ\ X(t) \text{ and } \nabla X(t) \text{ are independent;} \tag{2.1b}$$
$$\circ\ \Lambda_2(t) := \mathrm{Var}[\nabla X(t)] = -\mathrm{Cov}[\nabla^2 X(t), X(t)] \tag{2.1c}$$
$$\circ\ \text{the argument maximum of } X(\cdot) \text{ is unique.} \tag{2.1d}$$

Proof. The first property (2.1a) is equivalent to ($\mathbf{A_2}$). 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). □

The Gaussian random field $X^{|t}(\cdot)$

We know that $X(t)$ and $\nabla X(t)$ are independent by (2.1b). 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 ($\mathbf{A_4}$), 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

$$X^{|t}(s) := \frac{X(s) - c(s,t)\,X(t) - \nabla_t c(s,t)^\top \Lambda_2^{-1}(t) \nabla X(t)}{1 - c(s,t)}, \tag{2.2}$$

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

$$\nabla^2 X(t) = -\Lambda_2(t)\,X(t) - \Lambda_3(t) \nabla X(t) + \widetilde{R}(t), \tag{2.3a}$$

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 $\widetilde{R}(t)$ 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

$$R(t) := \Lambda_2^{-\frac{1}{2}}(t)\,\widetilde{R}(t)\,\Lambda_2^{-\frac{1}{2}}(t)\,. \tag{2.3b}$$

Note that $\widetilde{R}(t)$ 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. Second maximum and independent part of the Hessian

First maximum: We define the first maximum λ_1 of $\sigma X(\cdot)$ and its argument maximum t_1 by

$$\lambda_1 := \sigma X(t_1) \quad \text{and} \quad t_1 := \operatorname*{arg\,max}_{t \in M} X(t). \tag{2.4}$$

The argument maximum is almost surely a singleton by (2.1d), hence t_1 is unique almost surely.

Second maximum: We define the second maximum λ_2 of $\sigma X(\cdot)$ by

$$\forall t \in M, \quad \lambda_2^t := \sup_{s \in M \setminus \{t\}} \{\sigma X^{|t}(s)\} \quad \text{and} \quad \lambda_2 := \lambda_2^{t_1}. \tag{2.5}$$

Independent part of the Hessian: We define the independent part of the Hessian Ω as

$$\Omega := \sigma R(t_1) = \sigma \Lambda_2^{-\frac{1}{2}}(t_1)\, \widetilde{R}(t_1)\, \Lambda_2^{-\frac{1}{2}}(t_1). \tag{2.6}$$

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 B. 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 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 εh. Then, under ($\mathbf{A_1}$), ($\mathbf{A_2}$) and ($\mathbf{A_4}$),*

$$\lim_{\substack{\varepsilon \neq 0 \\ \varepsilon \to 0}} X^{|t}(s(\varepsilon)) = \frac{h^\top \widetilde{R}(t) h}{h^\top \Lambda_2(t) h} = \frac{h^\top}{\|h\|_2} R(t) \frac{h}{\|h\|_2}, \ \textit{almost surely}. \tag{2.7a}$$

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

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

Remark 3. *The aforementioned lemma shows that λ_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 (2.7a)*

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

$$X^{|t}(s) = \frac{X(s) - c(s,t)X(t)}{1 - c(s,t)} =: X^t(s). \tag{2.7b}$$

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:*

- $\imath_1 := \mathbb{1}\{t = t_1\}$;
- $\imath_2 := \mathbb{1}\{\forall s \in M \setminus \{t\},\ X(s) \leq X(t)\}$;
- $\imath_3 := \mathbb{1}\{\forall s \in M \setminus \{t\},\ X^t(s) \leq X(s)\}$;
- $\imath_4 := \mathbb{1}\{\nabla X(t) = 0\}\mathbb{1}\{\lambda_2^t \leq X(t)\}$;

then $\imath_1 = \imath_2 = \imath_3 = \imath_4$ 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 λ_1, defined by (2.4), conditional on (λ_2, Ω), and proven in Appendix D.

Theorem 1. *Let $X(\cdot)$ be a Gaussian random field satisfying Assumption ($\mathbf{A_{1\text{-}4}}$). Then, the distribution $\mathcal{D}(\lambda_1 \,|\, \lambda_2, \Omega)$ of the maximum λ_1 conditional on (λ_2, Ω) has a density with respect the Lebesgue measure* Leb *and this conditional density at point $\ell \in \mathbb{R}$ is given by*

$$\frac{\mathrm{d}\mathcal{D}(\lambda_1 \,|\, \lambda_2, \Omega)}{\mathrm{dLeb}}(\ell) = \frac{\det(\ell\, \mathrm{Id} - \Omega)\varphi(\ell/\sigma)}{G_\Omega(\lambda_2)} \mathbb{1}_{\lambda_2 \leq \ell} \quad \textit{with} \quad G_\Omega(\lambda_2) := \int_{\lambda_2}^{+\infty} \det(u\mathrm{Id} - \Omega)\varphi(u/\sigma)\, \mathrm{d}u.$$

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

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

$$Z^{|t_1}(s) := \frac{Z(s) - \lambda_1 c(s,t_1)}{1 - c(s,t_1)}; \tag{3.1b}$$

$$\lambda_2 := \max_{s\in M}\{Z^{|t_1}(s)\} \quad \text{and} \quad t_2 := \arg\max_{s\in M}\{Z^{|t_1}(s)\}; \tag{3.1c}$$

$$\Omega := \Lambda_2^{-\frac{1}{2}}(t_1)\left(\nabla^2 Z(t_1) + \lambda_1 \Lambda_2(t_1)\right)\Lambda_2^{-\frac{1}{2}}(t_1). \tag{3.1d}$$

In the previous section, we assumed that the Gaussian random field $X(\cdot)$ was centered in ($\mathbf{A_2}$). In this section, we give an exact test procedure for the following null hypothesis:

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

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 σ^2 is known.

Theorem 2. *Let $X(\cdot)$ be a Gaussian random field satisfying Assumption ($\mathbf{A_{1\text{-}4}}$). Under $\mathbb{H}_0$, the following test statistic satisfies*

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

where $\mathcal{U}(0,1)$ is the uniform distribution on $(0,1)$,

$$G_{\Omega/\sigma}(\ell) := \int_\ell^{+\infty} \det(u\mathrm{Id} - \Omega/\sigma)\varphi(u)\,\mathrm{d}u\,,$$

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 on $(\lambda_2, \Omega) = (\ell_2, r)$, $G_r(\lambda_1)/G_r(\ell_2) \sim \mathcal{U}(0,1)$. Since the conditional distribution does not depend on (ℓ_2, r), it is also the non-conditional distribution as claimed. □

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 ($\mathbf{A_2}$), for the moment. We consider the order m Karhunen-Loève expansion in the sense that

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

where the equality holds in L^2, uniformly in t, and $(f_1, \ldots, f_m)$ 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 $t_1 \in M$

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

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

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

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

$$\begin{aligned}&\exists (t_{d+2},\dots,t_p) \in (M\setminus\{t_1\})^{p-d-1} \text{ pairwise distinct s.t.}\\ &(X(t_1), \nabla X(t_1), X(t_{d+2}),\dots,X(t_p)) \text{ is non degenerate.}\end{aligned} \tag{ND(∞)}$$

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, its proof is in Appendix E.

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 (KL(m)) implies Assumption (ND(m)), and also Assumption (KL(∞)) implies Assumption (ND(∞)).*

We give some examples below:

- The normalized Brownian motion $W_t/\sqrt{t}$ satisfies Assumption (KL(∞));
- Any Gaussian stationary field with a spectrum that admits an accumulation point satisfies Assumptions (KL(∞)) and (ND(∞)) 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 (KL(∞)) and (ND(∞))
- Note that in the applications of Section 5 all the Gaussian random fields satisfy (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 ($\mathbf{A_{1\text{-}4}}$). In particular, $X(\cdot)$ admits a Karhunen-Loève expansion of order denoted $m_{\mathrm{KL}}(X)$ possibly infinite. We assume the following fifth condition:

$$\kappa := m_{\mathrm{KL}}(X) - d - 1 \geq 1\,. \tag{$\mathbf{A_5}$}$$

We will make use of the notation $m := m_{\mathrm{KL}}(X)$ 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 $m_{\mathrm{KL}}(X)$ 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 (KL(∞)), note that, for every integer $p \geq 2$, from $(\sigma X(t_1),\dots,\sigma X(t_p))$ for conveniently chosen points $t_1,\dots,t_p$ (say uniformly at random for instance), one can build an estimator, say $\widehat{\sigma}^2_{(p)}$, of the variance σ^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 $\widehat{\sigma}^2_{(p)}$ converges almost surely to σ^2 under $\mathbb{H}_0$. Thus the variance σ^2 is theoretically directly observable from the entire path of $\sigma X(\cdot)$ (though in practical applications one will estimate it by a χ^2 with a large number of degrees of freedom). We still denote by $\widehat{\sigma}^2_t$ this observation.

The finite order case

By Proposition 5, Assumption (ND(m)) holds with $m > d+1$. Let $t \in \mathrm{M}$ and let $(t_{d+2}, \dots, t_m) \in M^\kappa$ be as in Assumption (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 Σ is some known matrix and an estimator of σ^2 is

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

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

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

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

$$\widehat{\sigma}_t^2 = \frac{1}{\kappa}\sum_{i=1}^{\kappa} \overline{\zeta}_i^{\,2}\,. \tag{4.3}$$

It shows that $\widehat{\sigma}_t^2$ does not depend on the choice of $t_{d+2}, \dots, t_m$ in (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) $\widehat{\sigma}_t^2$ is well defined and follows a $\sigma^2\chi^2(\kappa)/\kappa$ distribution;
(ii) $\widehat{\sigma}_t^2$ is independent of $(X(t), \nabla X(t))$;
(iii) the random field $X^{|t}(\cdot)/\widehat{\sigma}_t$ is independent of the random variable $\widehat{\sigma}_t$.

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 $\overline{\zeta} = (\overline{\zeta}_1, \cdots, \overline{\zeta}_\kappa)$ and (4.2). □

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 $X(t)\,, \nabla X(t)\,,\ X^{|t}(\cdot)/\widehat{\sigma}_t$ and $\widehat{\sigma}_t$ are mutually independent. By (2.3b), 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 $X^{|t}(\cdot)$ at point t hence the random variables $\Big\{X(t)\,, \nabla X(t)\,,\ \big(X^{|t}(\cdot)/\widehat{\sigma}_t\,,\ R(t)/\widehat{\sigma}_t\big)\,,\ \widehat{\sigma}_t\Big\}$ 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 λ_2^t is defined in (2.5). We consider $t \in \mathrm{M}$ a putative value and λ_1^t build from $\sigma X(\cdot)$ in (2.4). Now, consider the test statistics

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

as in (2.6), and let $\overline{\mu}_t$ be the joint law of $(T_{2,t}, \sigma R(t)/\widehat{\sigma}_t)$.

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

$$f_{\frac{\chi_\kappa}{\sqrt{\kappa}}}(s) = \frac{2^{1-\frac{\kappa}{2}}}{\Gamma\left(\frac{\kappa}{2}\right)}\sqrt{\kappa}\left(s\sqrt{\kappa}\right)^{\kappa-1}\exp\left(-(\kappa\, s^2/2)\right),$$

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 $\mathrm{Leb}_2 \otimes \overline{\mu}_t$ 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, κ and σ. 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 the set of assumptions ($\mathbf{A_{1\text{-}5}}$). Then, under ($\mathbb{H}_0$), the joint distribution of $\big(\lambda_1, \widehat{\sigma}/\sigma, T_2, \Omega/\widehat{\sigma}\big)$ has a density with respect to $\mathrm{Leb}_2 \otimes \overline{\mu}^\star$ at point $(\ell_1, s, t_2, r) \in \mathbb{R}^3 \times \mathbb{S}_d$ equal to*

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

where $\overline{\mu}^\star$ is defined by $\overline{\mu}^\star(\cdot) := \int_M \overline{\mu}_t(\cdot)\, p_{\nabla X(t)}(0) \det \Lambda_2(t)\, \mathrm{d}\nu(t)$.

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 ($\mathbf{A_{1\text{-}5}}$). For all $r \in \mathcal{S}_d$, define H_r as*

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

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 , Ω are given by (4.4).

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

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

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

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

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

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

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

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

and we conclude using the same trick as the one of Theorem 2. □

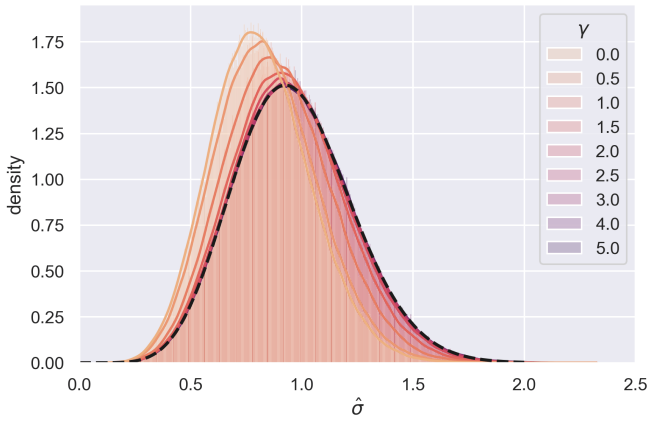

FIGURE 4. **[Spiked tensor PCA, Section 5.2, example 4/4]** *The variance estimator is not distributed according to χ^2-distribution with $\kappa = 7$ degrees of freedom, it underestimates the variance. The dashed black line is the distribution of a $\chi(7)/\sqrt{7}$, and $\sigma = 1$ in these experiments.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

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 $\kappa\,\widehat{\sigma}^2/\sigma^2$ (resp. $(m-1)\,\widehat{\sigma}^2/\sigma^2$) fails to be distributed according to a χ^2-distribution with κ (resp. $m-1$) degrees of freedom. This because t_1 is random.*
- *Second, Figure 4 illustrates that $\widehat{\sigma}^2$ under-estimates the variance σ^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 ψ_{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 $\kappa\,\widehat{\sigma}^2/\sigma^2$ has almost χ^2-distribution with κ 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 $\kappa\widehat{\sigma}^2/\sigma^2 \simeq \kappa\widehat{\sigma}_{t_1}/\sigma^2$ which has χ^2-distribution with κ degrees of freedom.*

5. Examples

Application	Manifold M	Dim. d	Section
Spiked Tensor PCA (Rank-1)	Sphere $\mathbb{S}^2$	2	5.2
Two-Spiked Tensor Model	$\mathbb{S}^1 \times \mathcal{M}_{n,2}$ (Stiefel)	$2n-2$	5.3
Super-Resolution	Torus $\mathbb{T}^2 = [0, 2\pi)^2$	2	A.1
2-Layer Neural Networks	$\mathbb{S}^{r-1} \times (\mathbb{S}^{n-1})^r$	$nr-1$	A.2

TABLE 1
Summary of applications considered in this paper. We specify the underlying manifold M and its dimension d.

5.1. How to deploy our method

We are given an observation $Y \in E$ as in Equation (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 ($\mathbb{H}_0$). The testing framework is depicted in Section 3.1 and we recall that

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

The first and second maximum (and their arguments) are given by (2.4), which are computed using a Riemannian 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 Riemannian 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 σ^2 is known, the testing statistics is given by Theorem 2. In particular, one needs to compute

$$G_{\Omega/\sigma}(\ell) = \int_\ell^{+\infty} \det(u\mathrm{Id} - \Omega/\sigma)\varphi(u)\,\mathrm{d}u\,. \tag{5.2}$$

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

$$H_{\Omega/\widehat{\sigma}}(\ell) = \int_\ell^{+\infty} \det(u\,\mathrm{Id} - \Omega/\widehat{\sigma})\, f_{m-1}\Big(u\sqrt{(m-1)/\kappa}\Big)\,\mathrm{d}u\,, \tag{5.3}$$

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 $H_{\Omega/\widehat{\sigma}}(\ell)$.

As for the variance estimation, we draw κ independent points $(t_{d+2}, \dots, t_m)$ uniformly on M. They generically satisfy Condition (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 Σ is some known matrix and an estimator of σ^2 is given by (4.1) with $t = t_1$ as in (4.4). In our experiments, we have drawn 5 independent samples on κ points and we found the same value for $\widehat{\sigma}$, 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,

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

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

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

The reader may consult https://github.com/ydecastro/tensor-spacing/ for further details on the numerical experiments. Figure 1 illustrates the geometric intuition behind the second maximum in this context. The left panel displays the original random field $X(\cdot)$. The middle and right panels depict the conditional field $X^{|t_1}(\cdot)$, which represents the regression remainder of $X(\cdot)$ with respect to the value and gradient at the maximizer t_1. Crucially, the volumetric view (right panel) reveals that $X^{|t_1}(\cdot)$ exhibits a singularity at t_1, forming a structure we refer to as the *helix*. This singularity arises because the limit of the conditional field at t_1 depends on the direction of approach, a phenomenon theoretically characterized in Lemma 3 (see also Remark 3).

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 $\Lambda_2 = 3\,\mathrm{Id}_2$ (on the tangent space), hence

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

Since the second fundamental form of the unit sphere is $-\mathrm{Id}$, the Riemannian Hessian is equal to the Euclidean Hessian $Z''(t)$ limited to the tangent space $-\mathrm{Id}$ 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 (2.6), it yields

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

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

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

Test statistics

One can compute the statistics (t_1, λ_1) and (t_2, λ_2) using a gradient descent. The expression of Ω has an explicit form by (5.4). Because we deal with 2×2 matrices, we have for the spacing test the following identities,

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

with

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

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

$$\begin{aligned} H_{\Omega/\widehat{\sigma}}(\ell) &= \int_\ell^{+\infty} \det(u\,\mathrm{Id}_2 - \Omega/\widehat{\sigma})\, f_{m-1}\Big(u\sqrt{(m-1)/\kappa}\Big)\,\mathrm{d}u \\ &= \int_\ell^{+\infty} \big[u^2 - \mathrm{Tr}(\Omega/\widehat{\sigma})u + \det(\Omega/\widehat{\sigma})\big]\, f_{m-1}\Big(u\sqrt{(m-1)/\kappa}\Big)\,\mathrm{d}u \\ &= \frac{\kappa\sqrt{m-3}}{(m-2)\sqrt{m-1}} \frac{\Gamma(\frac{m}{2})\Gamma(\frac{m-3}{2})}{\Gamma(\frac{m-1}{2})\Gamma(\frac{m-2}{2})} \sqrt{\frac{\kappa}{m-3}} \times \\ &\quad \Big[\ell\sqrt{\frac{m-3}{\kappa}} f_{m-3}(\ell\sqrt{\frac{m-3}{\kappa}}) + 1 - F_{m-3}(\ell\sqrt{\frac{m-3}{\kappa}})\Big] \\ &\quad - \mathrm{Tr}(\Omega/\widehat{\sigma}) \frac{\kappa\sqrt{m-3}}{(m-2)\sqrt{m-1}} \frac{\Gamma(\frac{m}{2})\Gamma(\frac{m-3}{2})}{\Gamma(\frac{m-1}{2})\Gamma(\frac{m-2}{2})} f_{m-3}(\ell\sqrt{\frac{m-3}{\kappa}}) \\ &\quad + \det(\Omega/\widehat{\sigma})\sqrt{\frac{\kappa}{m-1}}\Big(1 - F_{m-1}(\ell\sqrt{\frac{m-1}{\kappa}})\Big)\,, \end{aligned}$$

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

$$\det(\Omega/\widehat{\sigma}) = (1/9\widehat{\sigma}^2)\det\left[(\mathrm{Id}_3 - \Pi_{t_1})Z''(t_1)(\mathrm{Id}_3 - \Pi_{t_1}) + \Pi_{t_1}\right]$$
$$\mathrm{Tr}(\Omega/\widehat{\sigma}) = (1/3\widehat{\sigma})\big(\mathrm{Tr}(Z''(t_1)) - t_1^\top \sigma X''(t_1)t_1\big)\,.$$

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 λ_1 is stochastically increasing as γ increases while the top right panel shows that the distribution of λ_2 remains unchanged for moderate values of γ, say $\gamma \geq 2$. It illustrates that the spacing between λ_1 and λ_2 grows linearly with γ. In the moderate regime, the bottom left panel shows that λ_1 is distributed as a Gaussian with mean λ_0 and variance σ^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 γ. It illustrates that (t_1, λ_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 Euclidean space E is given by k-way n-dimensional symmetric tensors ($k \geq 3, n \geq 4$) equipped with the dot product
$$\forall\ \mathbf{T}, \mathbf{U} \in (\mathbb{R}^n)^{\otimes k}\,, \quad \langle \mathbf{T}, \mathbf{U}\rangle_E = \sum_{i_1,\ldots,i_k \in [n]} T_{i_1,\ldots,i_k}U_{i_1,\ldots,i_k}\,, \tag{5.5a}$$
with Euclidean or Frobenius norm $\|\mathbf{T}\|^2 = \langle \mathbf{T}, \mathbf{T}\rangle_E$, where $[n] = 1, \ldots, n$;
- the noise tensor $W \in (\mathbb{R}^n)^{\otimes k}$ is defined by
$$W = \frac{1}{k!}\sum_{\pi \in \mathfrak{S}_n} G^\pi\,, \tag{5.5b}$$
where $G_{i_1\cdots i_k}$ for $1 \leq i_1, \ldots, i_k \leq n$ are i.i.d standard Gaussians, $\mathfrak{S}_n$ is the set of permutations of the set $[n]$, and $G^\pi_{i_1\cdots i_k} = G_{\pi(i_1)\cdots\pi(i_k)}$. Note that the entries of W with indices $i_1 < i_2 < \cdots < i_k$ form an i.i.d. collection of Gaussian random variables, namely $\{W_{i_1\cdots i_k}\}_{i_1<i_2<\cdots<i_k}$ 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
$$\mathcal{M}_{n,2} := \{(x, y) \in (\mathbb{S}^{n-1})^2 : x \perp y\}.$$

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

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

So the relevant manifold is $M = \mathbb{S}^1 \times \mathcal{M}_{n,2}$, 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 $\theta = \arccos\left(\nu_{0,1}/\lambda_0\right)$.

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

$$\begin{aligned} c(s,t) &= \mathbb{E}\left(Y(t)Y(s)\right) \\ &= \cos(\theta)\cos(\theta')\langle u^{\otimes k}, x^{\otimes k}\rangle_E + \sin(\theta)\sin(\theta')\langle v^{\otimes k}, y^{\otimes k}\rangle_E \\ &= \cos(\theta)\cos(\theta')\langle u, x\rangle^k + \sin(\theta)\sin(\theta')\langle v, y\rangle^k . \end{aligned} \tag{5.6a}$$

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

$$c((\theta', u, v), (\theta, x, y)) = c((\theta', Uu, Uv), (\theta, Ux, Uy)), \tag{5.6b}$$

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 (5.6b), $\Lambda_2(t)$ depends on θ only. Let us define for distinct $j \neq i, \ell \neq i$:

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

Because of the independence properties within Y, all these variables are independent. We consider the profile likelihood random field at point (θ, 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 (θ, 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 θ, is given by

$$\nabla Z := \begin{pmatrix} \sin(\theta)Y_{[1]} + \cos(\theta)Y_{[2]} & \cdot \\ \frac{\cos(\theta)Y_{[1],2} - \sin(\theta)Y_{[2],1}}{\sqrt{2}} & \cdot \\ \cos(\theta)Y_{[1],3} & \sin(\theta)Y_{[2],3} \\ \vdots & \vdots \\ \cos(\theta)Y_{[1],n} & s\ Y_{[2],n} \end{pmatrix} . \tag{5.7}$$

Hence, we deduce that

$$\Lambda_2(\theta) = \operatorname{diag}\left(1, k/2, \cos^2(\theta), \dots, \sin^2(\theta), \dots\right), \tag{5.8}$$

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 A.3 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 ε-neighborhood of these points in the Kac-Rice formula and passing to the monotonic limit as ε tends to 0.

Let us compute the matrix Ω

Let us use for short c and s for $\cos(\theta)$ and $sin(\theta)$ respectively. We start by computing the Euclidean Hessian at (θ, 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

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

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

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

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

To obtain $\nabla^2 Z(t)$ for a generic point (θ, x, y) belonging to $\mathcal{M}_{n,2}$ we have to perform an isometry as in (5.6b). Using (5.1), we deduce the expression of Ω.

Testing procedures

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

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. (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.

Montanari, A., Reichman, D., and Zeitouni, O. (2015). On the limitation of spectral methods: From the gaussian hidden clique problem to rank-one perturbations of gaussian tensors. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 217–225.

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: Continued examples

A.1. 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); Azaïs et al. (2015); Candès and Fernandez-Granda (2014); Duval and Peyré (2015).

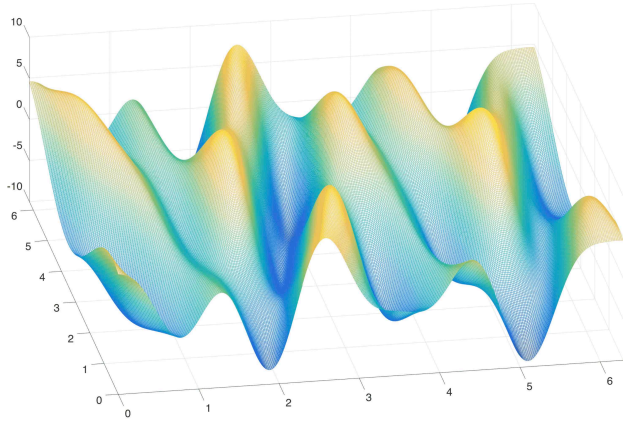

FIGURE 5. **[Super-Resolution, Section A.1]** *The super-resolution random field $Z(\cdot)$ depicts the observation of a point source at location x_0 on the interval $[0, 2\pi)$ (spatial domain on the x-axis) with phase θ_0 given on the y-axis. The z-axis corresponds to value of the profile likelihood $Z(x, \theta)$ at point x with phase θ. This figure is presented in Azaïs et al. (2020, Fig. 2).*

We consider the framework of Azaïs et al. (2020) where one observes $n = 2f + 1$ noisy frequencies between $-f$ and f with $f \geq 1$. The ℓ^{th} Fourier coefficient of a point source (Dirac mass) at location x_0 with amplitude λ_0 and phase θ_0 is $\lambda_0\, e^{\imath\theta_0} e^{-\imath\ell x_0}$. The observation is given by $Y = (y_\ell)_\ell$ where $-f \leq \ell \leq f$, they are complex random variables given by

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

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 $W_\ell = \xi_{\ell,1} + \imath\xi_{\ell,2}$ 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

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

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 (Azaïs et al., 2020, Proposition 10), λ_1 is the maximum of $Z(\cdot)$ and λ_2 the maximum of $Z^{|t_1}(\cdot)$. The t-spacing is explicitly described in Azaïs et al. (2020, Proposition 11).

A.2. Two-layer 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 $h:\mathbb{R}^n\to\mathbb{R}$, standard deviation $\sigma>0$, and w_i standard Gaussian variable.

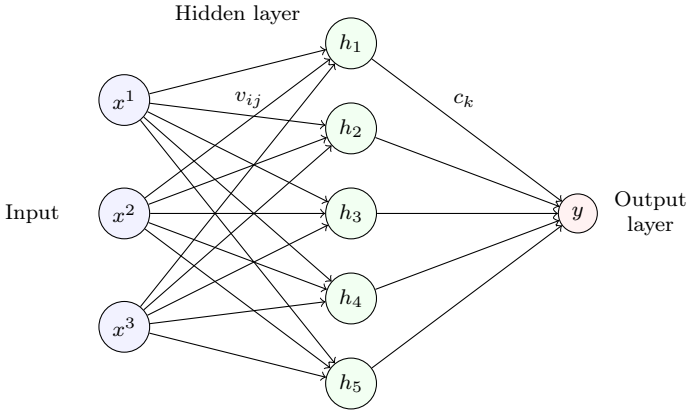

Figure 6. a two-layer neural network with c_k and $v_{i,j}$ real numbers.

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 $v_k\in\mathbb{S}^{n-1}$. Using the trick that the mean of r reals is the least squares estimator, we have to minimize

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

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

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

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

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

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

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

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 $c(s,t)$.

A.3. 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 θ 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 $e_2, \ldots, e_n$. We obtain $Y_{2[1]}, \ldots, Y_{n[1]}$. Then we consider the derivatives in θ of $X(\theta, x_1, y)$ with $y = e_2, \ldots, e_n$ We obtain $Y_{[2]}, \ldots, Y_{[n]}$. All these variables are independent with fixed variance so the density of $F(x)$ is bounded. □

Appendix B: 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 Π 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 ($\mathbf{A_1}$), the Taylor formula of order two gives

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

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

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

by Assumption ($\mathbf{A_2}$). From (B.1), we have

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

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 (B.2). We deduce that

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

From (2.3a) and passing to the limit, we deduce that

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

invoking that $\widetilde{R}$ is continuous by regression formulas and Assumption ($\mathbf{A_1}$), and that $\Lambda_2(t)[h,h]$ is positive by Assumption ($\mathbf{A_4}$).

For the second and last statement, observe that

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

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

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

Note that the exponential map is a local diffeomorphism on a neighborhood of the point t. Hence, there exist a sequence of positive reals ε_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 $\overline{h}$. The Taylor formula gives that

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

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

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

Passing to the limit by continuity yields

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

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

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

which concludes the proof.

Appendix C: 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 λ_2^t as a function of t is difficult to control. For this reason we must used the following tailored argument.

Denote by dist(s, t) the geodesic distance between points s and t. Define

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

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

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

So we can use Armentano et al. (2023, Section 7) for instance to compute the expectation of the left side of (C.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

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

where $\widetilde{X}(t)$ corresponds to the distribution conditional on $X(t_0) = 0$. Our goal is to show the continuity of the distribution of λ_2 in this representation. The conditioned random field $\widetilde{X}(\cdot)$ satisfies

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

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

$$X^{|t}(s) = \widetilde{X}^{|t}(s) + ug^{|t}(s), \quad \mathrm{dist(s, t)} > \epsilon\,,$$

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 D: Proof of Theorem 1

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

$$\begin{aligned}
&\mathbb{P}\left\{(\lambda_1, \lambda_2, \Omega) \in B\right\} \\
&\quad = \mathbb{P}\left\{\exists t \in M \ : \ \nabla X(t) = 0\,,\ (\sigma X(t), \lambda_2^t, \sigma R(t)) \in B\right\} && \text{(D.1a)}\\
&\quad = \mathbb{E}\left[\#\left\{t \in M \ : \ \nabla X(t) = 0\,,\ (\sigma X(t), \lambda_2^t, \sigma R(t)) \in B\right\}\right] && \text{(D.1b)}\\
&\quad = \int_M \mathbb{E}\left[\left|\det\left(-\nabla^2 X(t)\right)\right| \mathbb{1}_{(\sigma X(t), \lambda_2^t, \sigma R(t)) \in B} \middle| \nabla X(t) = 0\right] p_{\nabla X(t)}(0)\, \mathrm{d}\nu(t) && \text{(D.1c)}\\
&\quad = \int_M \mathbb{E}\left[\det\left(-\nabla^2 X(t)\right) \mathbb{1}_{(\sigma X(t), \lambda_2^t, \sigma R(t)) \in B} \middle| \nabla X(t) = 0\right] p_{\nabla X(t)}(0)\, \mathrm{d}\nu(t) && \text{(D.1d)}\\
&\quad = \int_M \mathbb{E}\left[\det(X(t)\Lambda_2(t) - \widetilde{R}(t)) \mathbb{1}_{(\sigma X(t), \lambda_2^t, \sigma R(t)) \in B}\right] p_{\nabla X(t)}(0)\, \mathrm{d}\nu(t) && \text{(D.1e)}\\
&\quad = \int_M \mathbb{E}\left[\det(X(t)\mathrm{Id} - R(t)) \mathbb{1}_{(\sigma X(t), \lambda_2^t, \sigma R(t)) \in B}\right] p_{\nabla X(t)}(0) \det \Lambda_2(t)\, \mathrm{d}\nu(t)\,, && \text{(D.1f)}
\end{aligned}$$

where ν 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 (D.1a);
- we used the uniqueness of the maximum (2.1d) in (D.1b);
- we used a *Kac-Rice formula* to get the third equality (D.1c), a proof is given in Appendix C;
- we used that $\imath_1 = \imath_4$ (see Lemma 4) and that the Hessian $-\nabla^2 X(t)$ is semi-definite positive at a maximum to get (D.1d);
- we used the Hessian regression formula (2.3a) with $\nabla X(t) = 0$ and invoked the independence of $\nabla X(t)$ from $(\sigma X(t), \lambda_2^t, \sigma \widetilde{R}(t))$ to get (D.1e);

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

$$\begin{aligned}
\mathrm{d}\overline{\nu}(t) &:= p_{\nabla X(t)}(0) \det \Lambda_2(t)\, \mathrm{d}\nu(t)\,, && \text{(D.2a)}\\
\mu^\star(\cdot) &:= \int_M \mu_t(\cdot)\, \mathrm{d}\overline{\nu}(t)\,, && \text{(D.2b)}
\end{aligned}$$

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 ($\mathbf{A_2}$). Now, coming back to (D.1f) we have

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

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 (ℓ_1, ℓ_2, ω) on the set $\mathcal{B}$ has a density proportional to $\det(\ell_1 \mathrm{Id}/\sigma - \omega/\sigma)\varphi(\ell_1/\sigma)$ with respect to $\mathrm{Leb} \otimes \mu^\star$. This implies in turn that the density of the maximum λ_1 conditional on (λ_2, Ω) is proportional to $\det(\ell_1 \mathrm{Id} - \omega)\varphi(\ell_1/\sigma)\mathbb{1}_{\ell_2 \le \ell_1}$. □

Appendix E: Proof of Proposition 5

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 $y \in \mathrm{M}$.

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 Φ 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 the vector Ψ given by $\Psi := \left[\partial_1\Phi(t_1)\, \partial_2\Phi(t_1) \cdots \partial_d\Phi(t_1)\, \Phi(t_1)\, \Phi(t_{d+2}) \cdots \Phi(t_p)\right] \in \mathbb{H}^p$, and it holds $\Sigma(V) = \Psi^\star\Psi$ where $\Psi^\star : h \in \mathbb{H} \mapsto (\langle\partial_1\Phi(t_1), h\rangle_{\mathbb{H}}, \cdots, \langle\Phi(t_p), h\rangle_{\mathbb{H}}) \in \mathbb{R}^p$ is the adjoint operator of Ψ.

Using Assumption ($\mathbf{A_2}$), Assumption ($\mathbf{A_4}$) and (2.1b), one gets that $W = (X(t_1), \nabla X(t_1))$ is non-degenerate, 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 $\widetilde{\Psi} := \left[\Phi(t_1)\partial_1\Phi(t_1) \cdots \partial_d\Phi(t_1)\right] \in \mathbb{H}^{d+1}$. 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 $\widetilde{\Psi}$ by $p-d-1$ vectors of the form $\Phi(t)$ to get a p dimensional free system Ψ, otherwise $\mathbb{H}$ would be of dimension less than p. □

Appendix F: Notation

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;
δ_{jk}	Kronecker symbol;
(const)	Positive constant which may change from line to line;
$\mathcal{D}(U \mid V)$	Conditional distribution of U with respect to V;
$\mathrm{Var}(U)$	Variance matrix of a random vector U;
$\mathrm{Cov}(U, V)$	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 ℓ;
$\mathbb{1}_A$ (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;
Random fields	
$Z(\cdot)$	Observed Gaussian random field, indexed by M;
$m(\cdot)$	Mean function of $Z(\cdot)$;
σ	Standard error of the $Z(\cdot)$;
$X(\cdot)$	Centered Gaussian random field with unit variance;
$c(s,t)$	Covariance function of $X(\cdot)$, $c(s,t) = \mathbb{E}(X(s)X(t))$;
Differential geometry	
M	$\mathcal{C}^2$-compact Riemannian manifold of dimension d without boundary;
$\nabla f(t)$	Riemannian gradient at point $t \in M$ of $f : M \to \mathbb{R}$;
$\nabla_t g(s,t)$	Riemannian gradient at point $t \in M$ of $g(s,\cdot)$;
$\nabla^2 f(t)$	Riemannian Hessian at point $t \in M$ of $f : M \to \mathbb{R}$;
Continuous regression	
$(E, \langle \cdot, \cdot \rangle_E)$	Euclidean space;
ψ	Feature map from M to E, $\psi : t \in M \mapsto \psi_t \in E$;
ψ_t	Vector $\psi(t) \in E$;
$J\psi_t$ (resp. $J{\psi_t}^\top$)	(resp. transpose of) Jacobian matrix of ψ 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 2

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