Wilks' phenomenon and penalized likelihood-ratio test for nonparametric curve registration

Dalalyan Arnak **Collier Olivier**
ENSAE / CREST and LIGM / Université Paris-Est

Abstract

The problem of curve registration appears in many different areas of applications ranging from neuroscience to road traffic modeling. In the present work, we propose a nonparametric testing framework in which we develop a generalized likelihood ratio test to perform curve registration. We first prove that, under the null hypothesis, the resulting test statistic is asymptotically distributed as a chi-squared random variable. This result, often referred to as Wilks' phenomenon, provides a natural threshold for the test of a prescribed asymptotic significance level and a natural measure of lack-of-fit in terms of the p-value of the χ^2-test. We also prove that the proposed test is consistent, *i.e.*, its power is asymptotically equal to 1. Finite sample properties of the proposed methodology are demonstrated by numerical simulations.

1 Introduction

Boosted by applications in different areas such as biology, medicine, computer vision, the problem of curve registration has been explored in a number of recent statistical studies. The model used for deriving statistical inference represents the input data as a finite collection of noisy signals such that each input signal is obtained from a given signal, termed mean template or structural pattern, by a parametric deformation and by adding a white noise. Hereafter, we refer to this as the *deformed mean template* (DMT) model. The main difficulties for developing statistical inference in this problem are caused by the nonlinearity of the deformations and the fact that not only the deformations but also the mean template used to generate the observed data are unknown.

Appearing in Proceedings of the 15^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2012, La Palma, Canary Islands. Volume XX of JMLR: W&CP XX.

While the problems of estimating the mean template and the deformations was thoroughly investigated in recent years, the question of the adequacy of modeling the available data by the DMT model received little attention. By the present work, we intend to fill this gap by introducing a nonparametric goodness-of-fit testing framework that allows us to propose a measure of appropriateness of a DMT model. To this end, we focus our attention on the case where the only allowed deformations are translations and propose a measure of goodness-of-fit based on the p-value of a chi-squared test.

1.1 Model description

We consider the case of functional data, that is each observation is a function on a fixed interval, taken for simplicity equal to $[0, 1]$. More precisely, assume that two independent samples, denoted $\{X_i\}_{i=1,...,n}$ and $\{X_i^\#\}_{i=1,...,n^\#}$, of functional data are available such that within each sample the observations are independent identically distributed (iid) drifted and scaled Brownian motions. Let f and $f^\#$ be the corresponding drift functions: $f(t) = \frac{d\mathbf{E}[X_1(t)]}{dt}$ and $f^\#(t) = \frac{d\mathbf{E}[X_1^\#(t)]}{dt}$. Then, $X_i(t) = \int_0^t f(u)\,du + sB_i(u)$ where $s > 0$ is the scaling parameter and B_is are independent Brownian motions. An analogous relation with a different parameter $s^\#$ holds for $X_i^\#$s. Since we assume that the entire paths are observed, the scale parameters s and $s^\#$ can be recovered with arbitrarily small error using the quadratic variation. So, in what follows, these parameters are assumed to be known.

The goal of the present work is to provide a statistical testing procedure for deciding whether the curves of the functions f and $f^\#$ coincide up to a translation. Considering periodic extensions of f and $f^\#$ on the whole real line, this is equivalent to checking the null hypothesis

$$H_0 : \exists\tau^* \in [0, 1] \text{ s.t. } f(\cdot) = f^\#(\cdot + \tau^*). \qquad (1)$$

If the null hypothesis is satisfied, we are in the set-up of a DMT model, where $f(\cdot)$ plays the role of the mean template and spatial translations represent the set of possible deformations.

Starting from Golubev [21] and Kneip and Gasser [25], semiparametric and nonparametric estimation in different instances of the DMT model have been intensively investigated [4, 6, 8–11, 13, 14, 18, 22, 30–32] with applications to image warping [5, 20]. However, prior to estimating the common template, the deformations or any other object involved in a DMT model, it is natural to check its appropriateness, which is the purpose of this work.

To achieve this goal, we first note that the pair of sequences of complex-valued random variables $\boldsymbol{Y} = (Y_0, Y_1, \ldots)$ and $\boldsymbol{Y}^\# = (Y_0^\#, Y_1^\#, \ldots)$, defined by

$$\big[Y_j, Y_j^\#\big] = \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \big[X_i(t), X_i^\#(t)\big] e^{2\pi i j t}\, dt,$$

constitutes a sufficient statistic in the model generated by the data $(\{X_i\}_{i=1,\ldots,n}; \{X_i^\#\}_{i=1,\ldots,n^\#})$. Therefore, without any loss of information, the initial (functional) data can be replaced by the transformed data $(\boldsymbol{Y}, \boldsymbol{Y}^\#)$. It is clear that the latter satisfy

$$Y_j = c_j + \frac{s}{\sqrt{n}}\,\epsilon_j, \quad Y_j^\# = c_j^\# + \frac{s^\#}{\sqrt{n^\#}}\,\epsilon_j^\#, \quad j \in \mathbb{N}, \quad (2)$$

where $c_j = \int_0^1 f(x)\, e^{2 i j \pi x}\, dx$, $c_j^\# = \int_0^1 f^\#(x)\, e^{2 i j \pi x}\, dx$ are the complex Fourier coefficients. The complex valued random variables ϵ_j, $\epsilon_j^\#$ are i.i.d. standard Gaussian: $\epsilon_j, \epsilon_j^\# \sim \mathcal{N}_{\mathbb{C}}(0, 1)$, which means that their real and imaginary parts are independent $\mathcal{N}(0, 1)$ random variables. In what follows, we will use boldface letters for denoting vectors or infinite sequences so that, for example, $\boldsymbol{c}$ and $\boldsymbol{c}^\#$ refer to $\{c_j; j = 0, 1, \ldots\}$ and $\{c_j^\#; j = 0, 1, \ldots\}$, respectively.

Under the mild assumption that f and $f^\#$ are squared integrable, the likelihood ratios of the Gaussian process $\boldsymbol{Y}^{\bullet,\#} = (\boldsymbol{Y}, \boldsymbol{Y}^\#)$ is well defined. Using the notation $\boldsymbol{c}^{\bullet,\#} = (\boldsymbol{c}, \boldsymbol{c}^\#)$, $\sigma = s/\sqrt{n}$ and $\sigma^\# = s^\#/\sqrt{n^\#}$, the corresponding negative log-likelihood is given by

$$\ell(\boldsymbol{Y}^{\bullet,\#}, \boldsymbol{c}^{\bullet,\#}) = \frac{\|\boldsymbol{Y} - \boldsymbol{c}\|_2^2}{2\sigma^2} + \frac{\|\boldsymbol{Y}^\# - \boldsymbol{c}^\#\|_2^2}{2(\sigma^\#)^2}. \quad (3)$$

In the present work, we present a theoretical analysis of the penalized likelihood ratio test (PLR) in the asymptotics of large samples, *i.e.*, when both n and $n^\#$ tend to infinity, or equivalently, when σ and $\sigma^\#$ tend to zero. The finite sample properties are examined through numerical simulations. The testing procedure will be described for any pair $(\sigma, \sigma^\#)$, but to keep theoretical developments readable, we will assume in our main results that $\sigma = \sigma^\#$.

1.2 Some motivations

Even if the shifted curve model is a very particular instance of the general DMT model, it plays a central role in several applications. To cite a few of them:

Road traffic forecast: In [26], a road traffic forecasting procedure is introduced. For this, archetypes of the different types of road trafficking behavior on the Parisian highway network are built, using a hierarchical classification method. In each obtained cluster, the curves all represent the same events, only randomly shifted in time.

Keypoint matching: An important problem in computer vision is to decide whether two points in a same image or in two different images correspond to the same real-world point. The points in images are then usually described by the regression function of the magnitude of the gradient over the direction of the gradient of the image restricted to a given neighborhood (cf. [27]). The methodology we shall develop in the present paper allows to test whether two points in images coincide, up to a rotation and an illumination change, since a rotation corresponds to shifting the argument of the regression function by the angle of the rotation.

1.3 Relation to previous work

The problem of estimating the parameters of the deformation is a semiparametric one, since the deformation involves a finite number of parameters that have to be estimated by assuming that the unknown mean template is merely a nuisance parameter. In contrast, the testing problem we are concerned with is clearly nonparametric. The parameter describing the probability distribution of the observations is infinite-dimensional not only under the alternative but also under the null hypothesis. Surprisingly, the statistical literature on this type of testing problems is very scarce. Indeed, while [23] analyzes the optimality and the adaptivity of testing procedures in the setting of a parametric null hypothesis against a nonparametric alternative, to the best of our knowledge, the only papers concerned with nonparametric null hypotheses are [1, 2] and [19]. Unfortunately, the results derived in [1, 2] are inapplicable in our set-up since the null hypothesis in our problem is neither linear nor convex. The set-up of [19] is closer to ours. However, they only investigate the minimax rates of separation without providing the asymptotic distribution of the proposed test statistic, which generally results in an overly conservative testing procedure.

1.4 Our contribution

We adopt, in this work, the approach based on the Generalized Likelihood Ratio (GLR) tests, cf. [16] for a comprehensive account on the topic. The advantage of this approach is that it provides a general framework for constructing testing procedures which asymptotically achieve the prescribed significance level for the first kind error and, under mild conditions, have a power that tends to one. It is worth mentioning that in the context of nonparametric testing, the use of the *generalized* likelihood ratio leads to a substantial improvement upon the likelihood ratio, very popular in parametric statistics. In simple words, the generalized likelihood allows to incorporate some prior information on the unknown signal in the test statistic which introduces more flexibility and turns out to be crucial both in theory and in practice [17]. We prove that under the null hypothesis the GLR test statistic is asymptotically distributed as a χ^2-random variable. This allows us to choose a threshold that makes it possible to asymptotically control the test significance level without being excessively conservative. Such results are referred to as Wilks' phenomena.

2 PLR test statistic

We are interested in testing the hypothesis (1), which translates in the Fourier domain to

$$H_0: \quad \exists\, \bar\tau^* \in [0, 2\pi[\quad \text{s.t.} \quad c_j = e^{-\mathrm{i}j\bar\tau^*} c_j^{\#} \quad \forall j = \mathbb{N}.$$

Indeed, one easily checks that if (1) is true, then[1] $c_j^{\#} = \int_0^1 f(t-\tau^*) e^{2\mathrm{i}j\pi t}\,dt = e^{2\mathrm{i}j\pi\tau^*}\int_0^1 f(z)e^{2\mathrm{i}j\pi z}\,dz = e^{2\mathrm{i}j\pi\tau^*} c_j$ and, therefore, the aforementioned relation holds with $\bar\tau^* = 2\pi\tau^*$. If no additional assumptions are imposed on the functions f and $f^{\#}$, or equivalently on their Fourier coefficients $\boldsymbol{c}$ and $\boldsymbol{c}^{\#}$, the nonparametric testing problem has no consistent solution. A natural assumption widely used in nonparametric statistics is that $\boldsymbol{c} = (c_0, c_1, \ldots)$ and $\boldsymbol{c}^{\#} = (c_0^{\#}, c_1^{\#}, \ldots)$ belong to some Sobolev ball $\mathcal{F}_{s,L} = \big\{\boldsymbol{u} = (u_0, u_1, \ldots) : \sum_{j=0}^{+\infty} j^{2s}|u_j|^2 \le L^2\big\}$, where the positive real numbers s and L stand for the smoothness and the radius of the class $\mathcal{F}_{s,L}$.

Since we will also be interested by establishing the (uniform) consistency of the proposed testing procedure, we need to precise the from of the alternative. It seems that the most compelling form for the null and the alternative is

$$\begin{cases} H_0: & \exists\, \bar\tau^* \quad \text{s.t.} \quad c_j = e^{-\mathrm{i}j\bar\tau^*} c_j^{\#} \quad \forall j \in \mathbb{N}. \\ H_1: & \inf_\tau \sum_{j=0}^{+\infty} |c_j - e^{-\mathrm{i}j\tau} c_j^{\#}|^2 \ge \rho \end{cases} \tag{4}$$

for some $\rho > 0$. In other terms, under H_0 the graph of the function $f^{\#}$ is obtained from that of f by a translation.

[1]We use here the change of the variable $z = t - \tau^*$ and the fact that the integral of a 1-periodic function on an interval of length one does not depend on the interval of integration.

To present the penalized likelihood ratio test, which is a variant of the GLR test, we introduce a penalization in terms of weighted ℓ^2-norm of $\boldsymbol{c}^{\bullet,\#}$. In this context, the choice of the ℓ^2-norm penalization is mainly motivated by the fact that Sobolev regularity assumptions are made on the functions f and $f^{\#}$. For a sequence of non-negative real numbers, $\boldsymbol{\omega}$, we define the weighted ℓ_2 norm $\|\boldsymbol{c}\|_{\boldsymbol{\omega},2}^2 = \sum_{j\ge 0}\omega_j|c_j|^2$. We will also use the standard notation $\|\boldsymbol{u}\|_p = (\sum_j |u_j|^p)^{1/p}$ for any $p > 0$. Using this notation, the penalized log-likelihood is given by

$$p\ell(\boldsymbol{Y}^{\bullet,\#}, \boldsymbol{c}^{\bullet,\#}) = \frac{\|\boldsymbol{Y} - \boldsymbol{c}\|_2^2 + \|\boldsymbol{c}\|_{\boldsymbol{\omega},2}^2}{2\sigma^2} + \frac{\|\boldsymbol{Y}^{\#} - \boldsymbol{c}^{\#}\|_2^2 + \|\boldsymbol{c}^{\#}\|_{\boldsymbol{\omega},2}^2}{2(\sigma^{\#})^2}. \tag{5}$$

The resulting penalized likelihood ratio test is based on the test statistic

$$\Delta(\boldsymbol{Y}^{\bullet,\#}) = \min_{\boldsymbol{c}^{\bullet,\#}: H_0 \text{ is true}} p\ell(\boldsymbol{Y}^{\bullet,\#}, \boldsymbol{c}^{\bullet,\#}) - \min_{\boldsymbol{c}^{\bullet,\#}} p\ell(\boldsymbol{Y}^{\bullet,\#}, \boldsymbol{c}^{\bullet,\#}). \tag{6}$$

It is clear that $\Delta(\boldsymbol{Y}^{\bullet,\#})$ is always non-negative. Furthermore, it is small when H_0 is satisfied and is large if H_0 is violated. The minimization of the quadratic functionals in (6) can be done explicitly and leads to the following result.

Proposition 1. *For any pair of sequences $\boldsymbol{Y}$ and $\boldsymbol{Y}^{\#}$ from ℓ^2, the test statistic $\Delta(\boldsymbol{Y}^{\bullet,\#})$ defined by (6) has the following simplified form:*

$$\Delta(\boldsymbol{Y}^{\bullet,\#}) = \frac{\sigma^2 + (\sigma^{\#})^2}{2(\sigma\sigma^{\#})^2}\min_\tau \sum_{j=0}^{+\infty}\frac{|Y_j - e^{ij\tau}Y_j^{\#}|^2}{1+\omega_j}. \tag{7}$$

Proof. The minimization of the quadratic functional (5) is an easy exercise and leads to

$$\min_{\boldsymbol{c}^{\bullet,\#}} p\ell(\boldsymbol{Y}^{\bullet,\#}, \boldsymbol{c}^{\bullet,\#}) = \frac{1}{2}\sum_{j\ge 0}\frac{\omega_j}{1+\omega_j}\Big[\Big|\frac{Y_j}{\sigma}\Big|^2 + \Big|\frac{Y_j^{\#}}{\sigma^{\#}}\Big|^2\Big]. \tag{8}$$

To compute the first term in the right hand side of (6), remark that it is equal to the minimum with respect to (w.r.t.) $\tau \in [0, 2\pi[$ and $\boldsymbol{c} \in \ell^2$ of the function

$$\sum_{j\ge 0}\Big[\frac{|Y_j - c_j|^2 + \omega_j|c_j|^2}{2\sigma^2} + \frac{|Y_j^{\#} - e^{\mathrm{i}j\tau}c_j|^2 + \omega_j|c_j|^2}{2(\sigma^{\#})^2}\Big].$$

The minimization w.r.t. $\boldsymbol{c}$ is attained when $c_j = (1+\omega_j)^{-1}(\sigma^{-2} + (\sigma^{\#})^{-2})^{-1}(Y_j\sigma^{-2} + e^{-\mathrm{i}j\tau}Y_j^{\#}(\sigma^{\#})^{-2})$ and

$$\min_{\boldsymbol{c}^{\bullet,\#}: H_0 \text{ true}} p\ell(\boldsymbol{Y}^{\bullet,\#}, \boldsymbol{c}^{\bullet,\#}) = \sum_{j\ge 0}\Big[\frac{|Y_j|^2}{2\sigma^2} + \frac{|Y_j^{\#}|^2}{2(\sigma^{\#})^2}\Big] - \min_\tau \sum_{j\ge 0}\frac{\sigma^2(\sigma^{\#})^2}{2(1+\omega_j)(\sigma^2 + (\sigma^{\#})^2))}\Big|\frac{Y_j}{\sigma^2} + \frac{e^{-\mathrm{i}j\tau}Y_j^{\#}}{(\sigma^{\#})^2}\Big|^2.$$

Combining this relation with (8), we get the result stated in the proposition. □

From now on, it will be more convenient to use the notation $\nu_j = 1/(1+\omega_j)$. The elements of the sequence $\boldsymbol{\nu} = \{\nu_j;\ j \geq 0\}$ are hereafter referred to as shrinkage weights. They are allowed to take any value between 0 and 1. Even the value 0 will be authorized, corresponding to the limit case when $w_j = +\infty$, or equivalently to our belief that the corresponding Fourier coefficient is 0. To ease notation, we will use the symbol $\circ$ to denote coefficient-by-coefficient multiplication, also known as the Hadamard product, and $\boldsymbol{e}(\tau)$ will stand for the sequence $(1, e^{-\mathrm{i}\tau}, e^{-2\mathrm{i}\tau}, \ldots)$. The test statistic can then be written as:

$$\Delta(\boldsymbol{Y}^{\bullet,\#}) = \frac{\sigma^2 + (\sigma^\#)^2}{2(\sigma\sigma^\#)^2} \min_{\tau\in[0,2\pi]} \|\boldsymbol{Y} - \boldsymbol{e}(\tau)\circ\boldsymbol{Y}^\#\|^2_{\boldsymbol{\nu},2}\,, \quad (9)$$

and the goal is to find the asymptotic distribution of this quantity under the null hypothesis.

3 Main results

The test based on the generalized likelihood ratio statistic involves a sequence $\boldsymbol{\nu}$, which is completely modulable by the user. However, we are able to provide theoretical guarantees only under some conditions on these weights. To state these conditions, we focus on the case $\sigma = \sigma^\#$ and choose a positive integer $N = N_\sigma \geq 2$, which represents the number of Fourier coefficients involved in our testing procedure. In addition to requiring that $0 \leq \nu_j \leq 1$ for every j, we assume that:

(A) $\nu_1 = 1$ and $\nu_j = 0, \forall j > N_\sigma$,

(B) $\sum_{j\geq 1} \nu_j^2 \geq \underline{c} N_\sigma$ for some constant $\underline{c} > 0$.

Moreover, we will use the following condition in the proof of the consistency of the test:

(C) there exists $\overline{c} > 0$, such that $\min\{j \geq 0, \nu_j < \overline{c}\} \to +\infty$, as $\sigma \to 0$.

In simple words, this condition implies that the number of terms ν_j that are above a given strictly positive level goes to $+\infty$ as σ converges to 0. If $N_\sigma \to +\infty$ as $\sigma \to 0$, then all the aforementioned conditions are satisfied for the shrinkage weights $\boldsymbol{\nu}$ of the form $\nu_{j+1} = h(j/N_\sigma)$, where $h : \mathbb{R} \to [0,1]$ is an integrable function, supported on $[0,1]$, continuous in 0 and satisfying $h(0) = 1$. The classical examples of shrinkage weights include:

$$\nu_j = \begin{cases} \mathbb{1}_{\{j\leq N_\sigma\}}, & \text{(projection weight)} \\ \mathbb{1}_{\{j\leq N_\sigma\}}/\big\{1 + \big(\frac{j}{\kappa N_\sigma}\big)^\mu\big\}, & \text{(Tikhonov weight)} \\ \big\{1 - \big(\frac{j}{N_\sigma}\big)^\mu\big\}_+,, & \text{(Pinsker weight)} \end{cases} \quad (10)$$

with $\kappa, \mu > 0$. Note that condition **(C)** is satisfied in all these examples with $\overline{c} = 0.5$, or any other value in $(0,1)$. Here on, we write $\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#})$ instead of $\Delta(\boldsymbol{Y}^{\bullet,\#})$ in order to stress its dependence on σ.

Theorem 1 (Wilks' phenomenon)**.** *Let* $\boldsymbol{c} \in \mathcal{F}_{1,L}$ *and* $|c_1| > 0$. *Assume that the shrinkage weights* ν_j *are chosen to satisfy conditions* **(A)**, **(B)**, $N_\sigma \to +\infty$ *and* $\sigma^2 N_\sigma^{5/2}\log(N_\sigma) = o(1)$. *Then, under the null hypothesis, the test statistic* $\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#})$ *is asymptotically distributed as a Gaussian random variable:*

$$\frac{\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#}) - 4\|\boldsymbol{\nu}\|_1}{4\|\boldsymbol{\nu}\|_2} \xrightarrow[\sigma\to 0]{\mathscr{D}} \mathcal{N}(0,1). \quad (11)$$

The main outcome of this result is a test of hypothesis H_0 that is asymptotically of a prescribed significance level $\alpha \in (0,1)$. Indeed, let us define the test that rejects H_0 if and only if

$$\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#}) \geq 4\|\boldsymbol{\nu}\|_1 + 4z_{1-\alpha}\|\boldsymbol{\nu}\|_2, \quad (12)$$

where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard Gaussian distribution.

Corollary 1. The test of hypothesis H_0 defined by the critical region (12) is asymptotically of significance level α.

Remark 1. Let us consider the case of projection weights $\nu_j = \mathbb{1}(j \leq N_\sigma)$. One can reformulate the asymptotic relation stated in Theorem 1 by claiming that $\frac{1}{2}\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#})$ is approximately $\mathcal{N}(2N_\sigma, 4N_\sigma)$ distributed. Since the latter distribution approaches the chi-squared distribution, we get:

$$\frac{1}{2}\,\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#}) \overset{\mathcal{D}}{\approx} \chi^2_{2N_\sigma}\,, \qquad \text{as} \qquad \sigma \to 0.$$

In the case of general shrinkage weights satisfying the assumptions stated in the beginning of this section, an analogous relation holds as well: $\frac{\|\boldsymbol{\nu}\|_1}{2\|\boldsymbol{\nu}\|_2^2}\,\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#}) \overset{\mathcal{D}}{\approx} \chi^2_{2\|\boldsymbol{\nu}\|_1^2/\|\boldsymbol{\nu}\|_2^2}$, as $\sigma \to 0$. This type of results are often referred to as Wilks' phenomenon.

Remark 2. The p-value of the aforementioned test based on the Gaussian or chi-squared approximation can be used as a measure of the goodness-of-fit or, in other terms, as a measure of alignment for the pair of curves under consideration. If the observed two noisy curves lead to the data $\boldsymbol{y}^{\bullet,\#}$, then the (asymptotic) p-value is defined as

$$\alpha^* = \Phi\Big(\frac{\Delta_\sigma(\boldsymbol{y}^{\bullet,\#}) - 4\|\boldsymbol{\nu}\|_1}{4\|\boldsymbol{\nu}\|_2}\Big),$$

where Φ stands for the c.d.f. of the standard Gaussian distribution.

So far, we have only focused on the behavior of the test under the null without paying attention on what happens under the alternative. The next theorem fills this gap by establishing the consistency of the test defined by the critical region (12).

Theorem 2. *Let condition* **(C)** *be satisfied and let* $\sigma^4 N_\sigma$ *tend to* 0 *as* $\sigma \to 0$. *Then the test statistic* $T_\sigma = \frac{\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#}) - 4\|\boldsymbol{\nu}\|_1}{4\|\boldsymbol{\nu}\|_2}$ *diverges under* H_1, *i.e.,* $T_\sigma \xrightarrow{P} +\infty$, *as* $\sigma \to 0$.

In other words, the result above claims that the power of the test defined via (12) is asymptotically equal to one as the noise level σ decreases to 0.

Remark 3. The previous theorem tells us nothing about the (minimax) rate of separation of the null hypothesis from the alternative. In other words, Theorem 2 does not provide the rate of divergence of T_σ. However, a rate is present in the proof (cf. Section D). In fact, in most situations $\min\{j \geq 0; j < \bar{c}\}$ is of the order N_σ, in which case we prove that

$$T_\sigma \geq \frac{\bar{c}\rho + O(N_\sigma^{-2}) + O_P(\sigma\sqrt{\log N_\sigma})}{4\sigma^2\sqrt{N_\sigma}}$$

as $\sigma \to 0$. This implies that, for instance, if $N_\sigma \to +\infty$ and satisfies $\sigma\sqrt{N_\sigma} = O(1)$ then T_σ tends to infinity if and only if $\rho/(\sigma\sqrt{\log N_\sigma}) \to \infty$. This argument can be made rigorous to establish that the minimax rate of separation is at least $\sigma^{1/2}(\log\sigma^{-1})^{1/4}$. However, we will not go into the details here since we believe that this rate is not optimal and intend to develop the minimax approach in a future work.

4 Numerical experiments

We have implemented the proposed testing procedure (12) in Matlab and carried out a certain number of numerical experiments on synthetic data. The aim of these experiments is merely to show that the methodology developed in the present paper is applicable and to give an illustration of how the different characteristics of the testing procedure, such as the significance level, the power, etc, depend on the noise variance σ^2 and on the shrinkage weights $\boldsymbol{\nu}$.

4.1 Convergence of the test under H_0

In order to illustrate the convergence of the test (12) when σ tends to zero, we made the following experiment. We chose the function HeaviSine, considered as a benchmark in the signal processing community, and computed its complex Fourier coefficients $\{c_j; j = 0, \ldots, 10^6\}$. For each value of σ taken from the set $\{2^{-k/2}, k = 1, \ldots, 15\}$, we repeated 5000 times the following computations:

- set $N_\sigma = 50\sigma^{-1/2}$,
- generate the noisy sequence $\{Y_j; j = 0, \ldots, N_\sigma\}$ by adding to $\{c_j\}$ an i.i.d. $\mathcal{N}_\mathbb{C}(0, \sigma^2)$ sequence $\{\xi_j\}$,

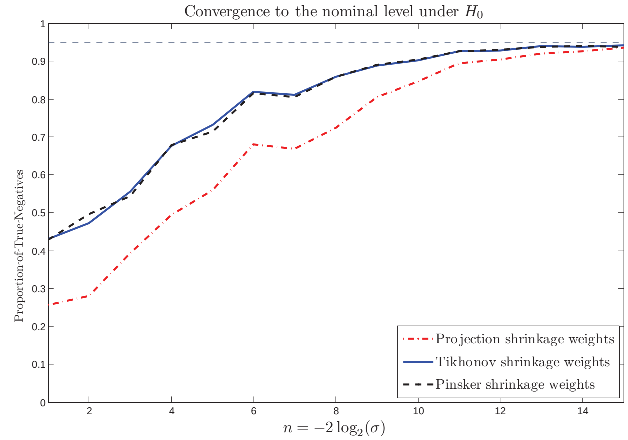

Figure 1: The proportion of true negatives in the first experiment described in Section 4 as a function of $\log_2\sigma^{-2}$ for three different shrinkage weights: projection (Left), Tikhonov (Middle) and Pinsker (Right). One can observe that for $\sigma = 2^{-15/2} \approx 5 \times 10^{-3}$, the proportion of true negatives is almost equal to the nominal level 0.95. Another observation is that the Pinsker and the Tikhonov weights lead to a faster convergence to the nominal significance level.

- randomly choose a parameter τ^* uniformly distributed in $[0, 2\pi]$, independent of $\{\xi_j\}$,
- generate the shifted noisy sequence $\{Y_j^\#; j = 0, \ldots, N_\sigma\}$ by adding to $\{e^{\mathrm{i}j\tau^*}c_j\}$ an i.i.d. $\mathcal{N}_\mathbb{C}(0, \sigma^2)$ sequence $\{\xi_j^\#\}$, independent of $\{\xi_j\}$ and of τ^*,
- compute the three values of the test statistic Δ_σ corresponding to the classical shrinkage weights defined by (10) and compare these values with the threshold for $\alpha = 5\%$.

We denote by $p^{\text{proj}}_{\text{accept}}(\sigma)$, $p^{\text{Tikh}}_{\text{accept}}(\sigma)$ and $p^{\text{Pinsk}}_{\text{accept}}(\sigma)$ the proportion of experiments (among 10^3 that have been realized) leaded to a value of the corresponding test statistic lower than the threshold, *i.e.*, the proportion of experiments leading to the acceptance of the null hypothesis. We plotted in Figure 1 the (linearly interpolated) curves $k \mapsto p^{\text{proj}}_{\text{accept}}(\sigma_k)$, $k \mapsto p^{\text{Tikh}}_{\text{accept}}(\sigma_k)$ and $k \mapsto p^{\text{Pinsk}}_{\text{accept}}(\sigma_k)$, with $\sigma_k = 2^{-k/2}$. It can be clearly seen that for $\sigma = 2^{-7} \approx 8 \times 10^{-3}$, the proportion of true negatives is almost equal to the nominal level 0.95. It is also worth noting that the three curves are quite comparable, with a significant advantage for the curve corresponding to Pinsker's and Tikhonov's weights: this curves converge a faster to the level $1 - \alpha = 95\%$ than the curve corresponding to the projection weights.

4.2 Power of the test

In the previous experiment, we illustrated the behavior of the penalized likelihood ratio test under the null hypothesis. The aim of the second experiment is to show what happens under the alternative. To this end, we still use the HeaviSine function as signal f and define $f^\# = f + \gamma\varphi$, where γ is a real parameter. Two cases are considered: $\varphi(t) = c\cos(4t)$ and $\varphi(t) = c/(1+t^2)$, where c is a constant ensuring that ϕ has an L^2 norm equal to 1. For each of these two pairs of functions $(f, f^\#)$, we repeated 5000 times the following computations:

- set $\sigma = 1$ and $N_\sigma = 50\sigma^{-1/2}$,
- compute the complex Fourier coefficients $\{c_j; j = 0, \ldots, 10^6\}$ and $\{c_j^\#; j = 0, \ldots, 10^6\}$ of f and $f^\#$, respectively,
- generate the noisy sequence $\{Y_j; j = 0, \ldots, N_\sigma\}$ by adding to $\{c_j\}$ an i.i.d. $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ sequence $\{\xi_j\}$,
- generate the shifted noisy sequence $\{Y_j^\#; j = 0, \ldots, N_\sigma\}$ by adding to $\{c_j^\#\}$ an i.i.d. $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ sequence $\{\xi_j^\#\}$, independent of $\{\xi_j\}$,
- compute the value of the test statistic Δ_σ corresponding to the projection weights and compare this value with the threshold for $\alpha = 5\%$.

To show the dependence of the behavior of the test under H_1 when the distance between the null and the alternative varies, we computed for each γ the proportion of true positives, also called the empirical power, among the 5000 random samples we have simulated. The results, plotted in Figure 2 show that even for moderately small values of γ, the test succeeds in taking the correct decision. It is a bit surprising that the result for the case $\varphi(t) = c\cos(4t)$ is better than that for $\varphi(t) = c/(1+t^2)$. Indeed, one can observe that the curve at the right panel approaches 1 much faster than the curve of the left panel.

5 Conclusion

In the present work, we provided a methodological and theoretical analysis of the curve registration problem from a statistical standpoint based on the nonparametric goodness-of-fit testing. In the case where the noise is white Gaussian and additive with a small variance, we established that the penalized log-likelihood ratio (PLR) statistic is asymptotically distribution free, under the null hypothesis. This result is valid for the weighted l^2-penalization under some mild assumptions on the weights. Furthermore, we proved that the test based on the Gaussian (or chi-squared) approximation of the PLR statistic is consistent. These results naturally carry over to other nonparametric models for which asymptotic equivalence (in the Le Cam sense) with the Gaussian white noise has been proven [7, 15, 28].

Some important issues closely related to the present work have not been treated here and will be done in near future. Perhaps the most important one is to determine the minimax rate of separation of the null hypothesis from the alternative. The results we have shown tell us that this rate is not slower than $\sigma^{1/2}(\log\sigma^{-1})^{1/4}$. However, it is very likely that this latter rate is suboptimal. There is a large body of literature on the topic of minimax rates of separation (cf. the book by Ingster and Suslina [24] and the references therein), but they mainly concentrate on the case of a simple null hypothesis. We expect that the composite character of the null hypothesis in our set-up will slow down the rate of convergence at least by a logarithmic factor.

Appendix

In order to respect the space limitations imposed by the conference, only the sketches of the proofs are presented. We refer the interested reader to the technical report [12] for more details.

A Maxima of random sums

In this section, we will give some technical lemmas which will be useful in the proofs of this paper. The proofs of Lemmas can be found in the technical report [12].

Proposition 2 (Berman [3])**.** *Suppose that g_j are continuously differentiable functions satisfying $\sum_{j=1}^n g_j(t)^2 = 1$ for all t, and $\xi_j \overset{iid}{\sim} \mathcal{N}(0,1)$. Then, for every $x > 0$, we have*

$$\mathbf{P}\bigg(\sup_{[a,b]}\sum_{j=1}^n g_j(t)\xi_j \geq x\bigg) \leq \frac{L_0}{2\pi}e^{-\frac{x^2}{2}} + \int_x^{+\infty}\frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\,dt,$$

with $L_0 = \int_a^b \big[\sum_{j=1}^n g_j'(t)^2\big]^{1/2}\,dt$.

We will also use the following fact about moderate deviations of the random variables that can be written as the sum of squares of independent centered Gaussian random variables.

Lemma 1. *Let N be some positive integer and let $\eta_j^\#$, $j = 1, \ldots, N$ be independent complex valued random variables such that their real and imaginary parts are independent standard Gaussian variables. Let $\boldsymbol{s} = (s_1, \ldots, s_N)$ be a vector of real numbers. For any $y \geq 0$, it holds that $\mathbf{P}\big\{\sum_{j=1}^N s_j^2|\eta_j^\#|^2 \geq 2\|\boldsymbol{s}\|_2^2 + 2\sqrt{2}\|\boldsymbol{s}\|_4^2 y + 2\|\boldsymbol{s}\|_\infty^2 y^2\big\} \leq e^{-y^2/2}$, with the standard notations $\|\boldsymbol{s}\|_\infty = \max_{j=1,\ldots,N}|s_j|$ and $\|\boldsymbol{s}\|_q^q = \sum_{j=1}^N |s_j|^q$.*

Lemma 2. *Let N be some positive integer and let η_j, $\eta_j^\#$, $j = 1, \ldots, N$ be independent complex valued random variables such that their real and imaginary parts are independent standard Gaussian variables. Let $\boldsymbol{s} = (s_1, \ldots, s_N)$ be a vector of real numbers. Denote $S(t) =$*

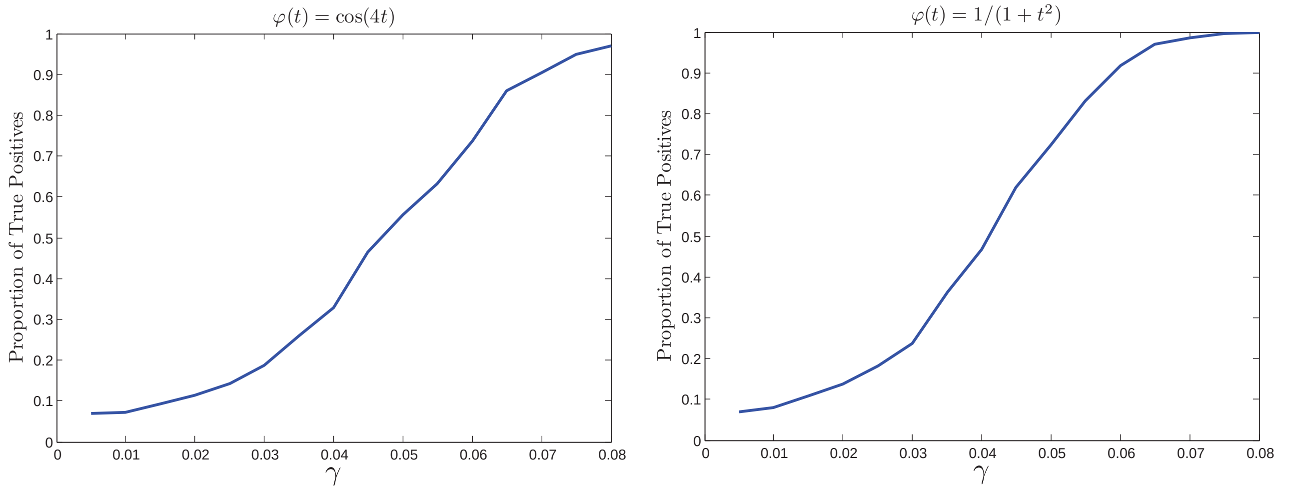

Figure 2: The proportion of true positives in the second experiment described in Section 4 as a function of the parameter γ measuring the distance between the true parameter and the set of parameters characterizing the null hypothesis. The main observation is that both curves tend to 1 very rapidly.

$\sum_{j=1}^N s_j \operatorname{Re}\big(e^{ijt}\eta_j\eta_j^\sharp\big)$ *for every* t *in* $[0,2\pi]$ *and* $\|S\|_\infty = \sup_{t\in[0,2\pi]}|S(t)|$. *Then, for all* $x,y>0$, $\mathbf{P}\big\{\|S\|_\infty > \sqrt{2}x\big(\|\boldsymbol{s}\|_2 + y\|\boldsymbol{s}\|_\infty\big)\big\} \le (N+1)e^{-x^2/2} + e^{-y^2/2}$.

B An auxiliary result

Proposition 3. *Assume that* H_0 *is satisfied. Let* $\boldsymbol{c}\in\mathcal{F}_{1,L}$ *and* $|c_1|>0$. *If the shrinkage weights* ν_j *satisfy conditions* **(A)** *and* **(B)**, *then the solution* $\hat\tau$ *to the optimization problem* $\max_{|\tau-\bar\tau^*|\le\pi} M(\tau)$, *with* $M(\tau) = \sum_{j\ge0}\nu_j \operatorname{Re}(e^{ij\tau}Y_j\overline{Y_j^\sharp})$ *satisfies the relation* $|\hat\tau - \bar\tau^*| = \sigma\sqrt{\log N_\sigma}\big(1+\sigma N_\sigma^{3/2}\big)O_P(1)$, *as* $\sigma\to0$.

Proof. If we set $\eta_j = e^{-ij\bar\tau^*}\epsilon_j$ and $\eta_j^\sharp = \epsilon_j^\sharp$, we can write the decomposition

$$M(\tau) = \mathbf{E}[M(\tau)] + \sigma S(\tau) + \sigma^2 D(\tau+\bar\tau^*),$$

with $\mathbf{E}[M(\tau)] = \sum_{j\ge0}\nu_j|c_j|^2\cos[j(\tau-\bar\tau^*)]$, $S(\tau) = \sum_{j\ge0}\nu_j\operatorname{Re}\big(e^{ij\tau}(\overline{c_j}\eta_j + c_j\overline{\eta_j^\sharp})\big)$ and $D(\tau) = \sum_{j\ge0}\nu_j\times \operatorname{Re}\big(e^{ij\tau}\eta_j\overline{\eta_j^\sharp}\big)$. On the one hand, using the assumption $|c_1|>0$ along with condition (A), we get that

$$\begin{aligned}\frac{\mathbf{E}\big[M(\tau)\big]-\mathbf{E}\big[M(\bar\tau^*)\big]}{(\tau-\bar\tau^*)^2} &\le -\nu_1|c_1|^2\frac{1-\cos(\tau-\bar\tau^*)}{(\tau-\bar\tau^*)^2}\\ &\le -\frac{2|c_1|}{\pi^2} \triangleq C<0.\end{aligned}$$

Therefore, $M(\tau)-M(\bar\tau^*)$ is equal to

$$\mathbf{E}[M(\tau)-M(\bar\tau^*)] + \sigma\big[S(\tau)-S(\bar\tau^*)\big] + \sigma^2\big[D(\tau)-D(\bar\tau^*)\big]$$

which is obviously bounded by

$$|\tau-\bar\tau^*|\big\{\sigma\|S'\|_\infty + \sigma^2\|D'\|_\infty - C|\tau-\bar\tau^*|\big\}.$$

Using this result, for every $a>0$, we get

$$\begin{aligned}\mathbf{P}\big(|\hat\tau-\bar\tau^*|>a\big) &\le \mathbf{P}\Big\{\sup_{|\tau-\bar\tau^*|>a} M(\tau)-M(\bar\tau^*)\ge0\Big\}\\ &\le \mathbf{P}\big\{\sigma\|S'\|_\infty + \sigma^2\|D'\|_\infty \ge Ca\big\}.\end{aligned}$$

The choice $a = \sigma\sqrt{\log N_\sigma}\big(2+\sigma N_\sigma^{3/2}\big)z$ implies that the probability $\mathbf{P}\big(|\hat\tau-\bar\tau^*| > \sigma\sqrt{\log N_\sigma}\big(1+\sigma N_\sigma^{3/2}\big)z\big)$ is bounded by

$$\mathbf{P}\big(\|S'\|_\infty \ge 2Cz\sqrt{\log N_\sigma}\big) + \mathbf{P}\big(\|D'\|_\infty \ge Cz\sqrt{N_\sigma^3\log N_\sigma}\big).$$

On the other hand, since $S'(t) = \sum_{j\ge0} j|c_j|\nu_j\operatorname{Re}\big(e^{ij\tau}\zeta_j\big)$, where ζ_j are i.i.d. complex valued random variable, whose real and imaginary parts are independent $\mathcal{N}(0,2)$ variables, the large deviations of the sup-norm of S' can be controlled by using the following lemma.

Lemma 3. *The sup-norm of the function* $S(t) = \sum_{j=0}^K s_j\{\cos(jt)\xi_j + \sin(jt)\xi_j'\}$, *where* $\{\xi_j\}$ *and* $\{\xi_j'\}$ *are two independent sequences of i.i.d.* $\mathcal{N}(0,1)$ *random variables, satisfies* $\mathbf{P}(\|S\|_\infty \ge \|\boldsymbol{s}\|_2 x) \le (K+1)e^{-x^2/2}$, $\forall x>0$.

Proof. This results is a direct consequence of Berman's inequality that we recall in Section A for the reader's convenience. □

Using this lemma and the fact that $N_\sigma\ge2$, we get that $\mathbf{P}\big(\|S'\|_\infty \ge 2LC\sqrt{2y\log N_\sigma}\big) \le 2N_\sigma^{1-y} \le 2^{2-y}$ for every $y>1$. Finally, the large deviations of the term $\|D'\|_\infty$ are controlled by using Lemma 2 below. Putting these inequalities together, we find that for any $\alpha\in(0,1)$, there exists $z>0$ such that $\mathbf{P}\big(|\hat\tau-\bar\tau^*| > \sigma\sqrt{\log N_\sigma}\big(1+\sigma N_\sigma^{3/2}\big)z\big)\le\alpha$. In conclusion, we get that $\hat\tau-\bar\tau^*$ is, in probability, at most of the order $\sigma\sqrt{\log N_\sigma}\big(1+\sigma N_\sigma^{3/2}\big)$. □

C Proof of Theorem 1

The term $\Delta_\sigma(\mathbf{Y}^{\bullet,\sharp}) = \frac{1}{\sigma^2}\min_\tau\big[\sum_{j=0}^{+\infty}\nu_j\big|Y_j - e^{-ij\tau}Y_j^\sharp\big|^2\big]$ can be written under H_0 as

$$\Delta_\sigma(\mathbf{Y}^{\bullet,\sharp}) = \frac{1}{\sigma^2}\min_{|\tau|\le\pi}\big\{D_\sigma(\tau) + 2C_\sigma(\tau) + P_\sigma(\tau)\big\}, \qquad (13)$$

where we have used the notation:

$$D_\sigma(\tau) = \sum_{j=0}^{+\infty} \nu_j |c_j|^2 \big|1 - e^{-\mathrm{i}j(\tau-\bar\tau^*)}\big|^2,$$
$$C_\sigma(\tau) = \sigma \sum_{j=0}^{+\infty} \nu_j \operatorname{Re}\big[c_j\big(1 - e^{-\mathrm{i}j(\tau-\bar\tau^*)}\big)\big(\overline{\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^\#}\big)\big],$$
$$P_\sigma(\tau) = \sigma^2 \sum_{j=0}^{+\infty} \nu_j \big|\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^\#\big|^2.$$

(Since H_0 is assumed satisfied, there exists $\bar\tau^* \in [0, 2\pi[$ such that $c_j = e^{-\mathrm{i}j\bar\tau^*} c_j^\#$ for all $j \ge 0$.) We denote by $\hat\tau$ the pseudo-estimator of $\bar\tau^*$ defined as the minimizer of the RHS of (13) and study the asymptotic behavior of the terms D_σ, C_σ and P_σ separately. For the deterministic term, it holds that

$$\begin{aligned}|D_\sigma(\hat\tau)| &\le \sum_{j=0}^{+\infty} j^2 \nu_j |c_j|^2 (\hat\tau - \bar\tau^*)^2 \\ &\le L(\hat\tau - \bar\tau^*)^2 = \{\sigma^2(1+\sigma^2 N_\sigma^3)\log N_\sigma\}\, O_p(1).\end{aligned}$$

Let us turn now to the cross term. It holds that $C_\sigma(\tau) = \sigma\sum_{j=0}^{+\infty}\nu_j\big\{(1 - \cos[j(\tau - \bar\tau^*)])\operatorname{Re}\big[c_j\big(\overline{\epsilon_j - e^{-\mathrm{i}j\bar\tau^*}\epsilon_j^\#}\big)\big] + \sin[j(\bar\tau^* - \tau)]\operatorname{Im}\big[c_j\big(\overline{\epsilon_j + e^{-\mathrm{i}j\bar\tau^*}\epsilon_j^\#}\big)\big]\big\}$. Combining this with $C_\sigma(\bar\tau^*) = 0$, we have $|C_\sigma(\hat\tau)| \le |\hat\tau - \bar\tau^*| \cdot \|C'_\sigma\|_\infty$. By arguments similar to those used in the proof of Proposition 3, we check that $\|C'_\sigma\|_\infty$ is of the order $\{\sigma\sqrt{\log N_\sigma}\}$ in probability. Therefore, it holds that $|C_\sigma(\hat\tau)| = \{\sigma^2(1 + \sigma N_\sigma^{3/2})\log N_\sigma\}\, O_p(1)$.

Let us now study the last term, $P_\sigma(\tau) = \sigma^2\sum_{j=0}^{+\infty}\nu_j\big|\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^\#\big|^2$, which will determine the asymptotic behavior of the test statistic. Now denoting $\eta_j = e^{\mathrm{i}j\bar\tau^*}\epsilon_j$ and $\eta_j^\# = \epsilon_j^\#$, we can rewrite this term as $P_\sigma(\tau) = \sigma^2\sum_{j=0}^{+\infty}\nu_j\big|\eta_j - e^{-\mathrm{i}j(\tau-\bar\tau^*)}\eta_j^\#\big|^2$. We wish to prove now that under H_0, if conditions **(A)**, **(B)**, $N\sigma \to +\infty$ and $\sigma^2 N_\sigma^{5/2}\log(N_\sigma) = o_P(1)$ are fulfilled, then

$$T_\sigma(\hat\tau) = \frac{P_\sigma(\hat\tau) - 4\sigma^2\sum_{k\ge0}\nu_k}{4\sigma^2(\sum_{k\ge0}\nu_k^2)^{1/2}} \xrightarrow[\sigma\to0]{\mathscr{D}} \mathcal{N}(0,1).$$

We start by writing $T_\sigma(\bar\tau^*)$ as the sum over $j \in \{1, \dots, N_\sigma\}$ of random variables $X_{j,\sigma} = \nu_j(|\eta_j - \eta_j^\#|^2 - 4)/4\|\boldsymbol{\nu}\|_2$, and applying the Berry-Esseen inequality [29, Theorem 5.4]. Since we have $B_\sigma = \sum_{j=0}^{N_\sigma}\mathbf{V}ar\big(X_{j,\sigma}\big) = 1$ and $L_\sigma = \sum_{j=0}^{N_\sigma}\mathbf{E}|X_{j,\sigma}|^3 \le C\,N_\sigma^{-1/2}$, the Berry-Esseen inequality yields $\sup_x |F_\sigma(x) - \Phi(x)| \le K\,L_\sigma$, where $F_\sigma(x) = \mathbf{P}\big(B_\sigma^{-1/2}\sum_{j=0}^{N_\sigma}X_{j,\sigma} < x\big)$, Φ is the c.d.f. of $\mathcal{N}(0,1)$ and K is an absolute constant. Hence $T_\sigma(\bar\tau^*) \xrightarrow[\sigma\to0]{\mathscr{D}} \mathcal{N}(0,1)$.

It remains now to prove that $R_\sigma = T_\sigma(\hat\tau) - T_\sigma(\bar\tau^*)$ tends to 0 in probability, which—by Slutski's lemma—will be sufficient for completing the proof. It holds that

$$\begin{aligned}R_\sigma(\tau) &= \sum_{j=0}^{+\infty}\frac{\nu_j}{2\|\boldsymbol{\nu}\|_2}\operatorname{Re}\eta_j\overline{\eta_j^\#}\,(e^{\mathrm{i}j(\tau-\bar\tau^*)} - 1) \\ &= \sum_{j=0}^{N_\sigma}\frac{j\nu_j(\tau - \bar\tau^*)}{2\|\boldsymbol{\nu}\|_2}\operatorname{Re}\big(e^{\mathrm{i}jt}\eta_j\overline{\eta_j^\#}\big),\end{aligned}$$

with some $t \in [\tau, \bar\tau^*]$. Then, by virtue of Lemma 2, $|R_\sigma(\hat\tau)|$ is bounded by

$$\begin{aligned}&\frac{|\hat\tau - \bar\tau^*|}{2\|\boldsymbol{\nu}\|_2}\sup_{t\in[0,2\pi]}\Big|\sum_{j=0}^{N_\sigma} j\nu_j\operatorname{Re}\big(e^{\mathrm{i}jt}\eta_j\overline{\eta_j^\#}\big)\Big| \\ &\qquad = \{\sigma(1+\sigma N_\sigma^{3/2})N_\sigma\log N_\sigma\}\cdot O_P(1).\end{aligned}$$

Hence, $R_\sigma(\hat\tau) = o_P(1)$ and the desired result follows.

D Proof of Theorem 2

The aim of this section is to present a proof of Theorem 2. To this end, we study the test statistic $T_\sigma = (\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#}) - 4\|\boldsymbol{\nu}\|_1)/4\|\boldsymbol{\nu}\|_2$, and show that it tends to $+\infty$ in probability under H_1. Actually, the hypothesis H_1 will be supposed to be satisfied throughout this section. The term $\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#})$ is bounded from below by

$$\begin{aligned}&\frac{1}{\sigma^2}\min_{\tau\in[0,2\pi]}\Big\{\sum_{j\ge0}\nu_j|c_j - e^{-\mathrm{i}j\tau}c_j^\#|^2\Big\} \\ &\quad - \frac{2}{\sigma}\max_{\tau\in[0,2\pi]}\Big\{\sum_{j\ge0}\nu_j|c_j - e^{-\mathrm{i}j\tau}c_j^\#|\cdot|\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^\#|\Big\}.\end{aligned}$$

Let us focus on the first term. Denoting $\delta_\sigma = \min\{j \ge 0, \nu_j < \overline{c}\}$, we get by condition **(C)** that $\delta_\sigma \to +\infty$, which implies that for any τ, the term $\sum_{j\ge0}\nu_j|c_j - e^{-\mathrm{i}j\tau}c_j^\#|^2$ can be bounded from below by $\overline{c}\sum_{j=0}^{\delta_\sigma}|c_j - e^{-\mathrm{i}j\tau}c_j^\#|^2 \ge \overline{c}\big(\rho - 4L\delta_\sigma^{-2}\big)$. Now, the second term satisfies

$$\begin{aligned}&\sum_{j\ge0}\nu_j\big|c_j - e^{-\mathrm{i}j\tau}c_j^\#\big|\cdot\big|\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^\#\big| \\ &\qquad \le \max_{j=0,\dots,N_\sigma}\big(|\epsilon_j| \vee |\epsilon_j^\#|\big)\sum_{j\ge0}\big(|c_j| + |c_j^\#|\big).\end{aligned}$$

On the one hand, standard inequalities on the tails of Gaussian random variables combined with the union bound imply that $\max_{j=0,\dots,N_\sigma}\big(|\epsilon_j| \vee |\epsilon_j^\#|\big) = O_P(\sqrt{\log N_\sigma})$. On the other hand, one can also check that $\sum_{j\ge0}\big(|c_j| + |c_j^\#|\big) \le \big(\sum_{j\ge0}j^{-2}\big)^{1/2}\big(\sum_{j\ge0}j^2\big(|c_j| + |c_j^\#|\big)^2\big)^{1/2} = O(1)$. Putting all these together, we get

$$\begin{aligned}T_\sigma &= \frac{\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#}) - 4\|\boldsymbol{\nu}\|_1}{4\|\boldsymbol{\nu}\|_2} \\ &\ge \frac{\overline{c}\rho - 4L\overline{c}\delta_\sigma^{-2} + O_P\big(\sigma\sqrt{\log N_\sigma}\big)}{4\sigma^2\sqrt{N_\sigma}} \xrightarrow{P} +\infty.\end{aligned}$$

Acknowledgements

This work has been partially supported by ANR Callisto.

References

[1] Y. Baraud, S. Huet, and B. Laurent. Adaptive tests of linear hypotheses by model selection. *Ann. Statist.*, 31(1):225–251, 2003.

[2] Y. Baraud, S. Huet, and B. Laurent. Testing convex hypotheses on the mean of a Gaussian vector. *Ann. Statist.*, 33(1):214–257, 2005.

[3] S. M. Berman. Sojourns and extremes of a stochastic process defined as a random linear combination of arbitrary functions. *Comm. Statist. Stochastic Models*, 4(1):1–43, 1988.

[4] J. Bigot and S. Gadat. A deconvolution approach to estimation of a common shape in a shifted curves model. *Ann. Statist.*, 38(4):2422–2464, 2010.

[5] J. Bigot, S. Gadat, and J.-M. Loubes. Statistical M-estimation and consistency in large deformable models for image warping. *J. Math. Imaging Vision*, 34(3):270–290, 2009.

[6] J. Bigot, F. Gamboa, and M. Vimond. Estimation of translation, rotation, and scaling between noisy images using the Fourier-Mellin transform. *SIAM J. Imaging Sci.*, 2(2):614–645, 2009.

[7] L. D. Brown and M. G. Low. Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, 24(6):2384–2398, 1996.

[8] R. J. Carroll and P. Hall. Semiparametric comparison of regression curves via normal likelihoods. *Austral. J. Statist.*, 34(3):471–487, 1992.

[9] I. Castillo. Semi-parametric second-order efficient estimation of the period of a signal. *Bernoulli*, 13(4):910–932, 2007.

[10] I. Castillo. A semiparametric Bernstein-von Mises theorem for Gaussian process priors. *Probab. Theory Related Fields*, To appear, 2011.

[11] I. Castillo and J.-M. Loubes. Estimation of the distribution of random shifts deformation. *Math. Methods Statist.*, 18(1):21–42, 2009.

[12] O. Collier and A. S. Dalalyan. Curve registration by nonparametric goodness-of-fit testing. Technical Report arXiv:1104.4210, 2011.

[13] A. S. Dalalyan. Stein shrinkage and second-order efficiency for semiparametric estimation of the shift. *Math. Methods Statist.*, 16(1):42–62, 2007.

[14] A. S. Dalalyan, G. K. Golubev, and A. B. Tsybakov. Penalized maximum likelihood and semiparametric second-order efficiency. *Ann. Statist.*, 34(1):169–201, 2006.

[15] A. S. Dalalyan and M. Reiß. Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. *Probab. Theory Related Fields*, 137(1-2):25–47, 2007.

[16] J. Fan and J. Jiang. Nonparametric inference with generalized likelihood ratio tests. *TEST*, 16(3):409–444, 2007.

[17] J. Fan, C. Zhang, and J. Zhang. Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.*, 29(1):153–193, 2001.

[18] F. Gamboa, J.-M. Loubes, and E. Maza. Semiparametric estimation of shifts. *Electron. J. Stat.*, 1:616–640, 2007.

[19] G. Gayraud and C. Pouet. Adaptive minimax testing in the discrete regression scheme. *Probab. Theory Related Fields*, 133(4):531–558, 2005.

[20] C. A. Glasbey and K. V. Mardia. A penalized likelihood approach to image warping. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(3):465–514, 2001.

[21] G. K. Golubev. Estimation of the period of a signal with an unknown form against a white noise background. *Problems Inform. Transmission*, 24(4):288–299, 1988.

[22] W. Härdle and J. S. Marron. Semiparametric comparison of regression curves. *Ann. Statist.*, 18(1):63–89, 1990.

[23] J. L. Horowitz and V. G. Spokoiny. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, 69(3):599–631, 2001.

[24] Yu. I. Ingster and I. A. Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2003.

[25] A. Kneip and T. Gasser. Statistical tools to analyze data representing a sample of curves. *Ann. Statist.*, 20(3):1266–1305, 1992.

[26] J.-M. Loubes, E. Maza, M. Lavielle, and L. Rodríguez. Road trafficking description and short term travel time forecasting, with a classification method. *Canad. J. Statist.*, 34(3):475–491, 2006.

[27] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[28] M. Nussbaum. Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.*, 24(6):2399–2430, 1996.

[29] V. V. Petrov. *Limit theorems of probability theory*, volume 4 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1995. Sequences of independent random variables, Oxford Science Publications.

[30] B. B. Rønn. Nonparametric maximum likelihood estimation for shifted curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(2):243–259, 2001.

[31] Y. Ritov U. Isserles and T. Trigano. Semiparametric curve alignment and shift density estimation for biological data. *IEEE Transactions on Signal Processing,*, To appear, 2011.

[32] M. Vimond. Efficient estimation for a subclass of shape invariant models. *Ann. Statist.*, 38(3):1885–1912, 2010.