APPENDIX B: CURVE REGISTRATION BY NONPARAMETRIC GOODNESS-OF-FIT TESTING

ANONYMOUS AUTHOR 1, ANONYMOUS AUTHOR 2

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. Some numerical experiments on synthetic datasets are reported as well.

Introduction

Boosted by applications in different areas such as biology, medicine, computer vision and road traffic forecasting, the problem of curve registration and, more particularly, some aspects of this problem related to nonparametric and semiparametric estimation, have been explored in a number of recent statistical studies. In this context, the model used for deriving statistical inference assumes that the input data consist of a finite collection of noisy signals possessing the following feature: 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. In what follows, we will refer to this as the "deformed mean template" 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 that was used to generate the observed data are unknown.

While the problems of estimating the mean template, the deformations and some other related objects have been thoroughly investigated in recent years, the question of the adequacy of modeling the available data by the aforementioned semiparametric model has 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 deformed mean template 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.

In full generality, the mathematical formulation of the "deformed mean template" model is the following. We are given a sample of size n of noisy signals $\{Y_m; m = 1, \dots, n\}$ having common structural pattern f, that is

$$dY_m(x) = f(\phi(x, \tau_m))\, dx + \sigma_m\, dW_m(x), \quad x \in [0, 1]^d, \quad m = 1, \ldots, n, \tag{1}$$

where ϕ is a known function determining the type of the deformation and τ_m is a finite-dimensional parameter allowing to instantiate the deformation. Typical examples are:

(a) Shifted curve model $\phi(x, \tau) = x - \tau$, where $\tau \in \mathbb{R}^d$ is the shift parameter,

(b) Periodic signal model $\phi(x, \tau) = \tau x$, where the signal f is a 1-periodic univariate function, x is one-dimensional and $\tau \in \mathbb{R}$ is the period of the noise-free signal,

(c) Rigid deformation model $\phi(x, \tau) = s(\mathtt{R}x + t)$, where $\tau = (s, \mathtt{R}, t)$ with $s > 0$ being the scale, $\mathtt{R}$ being the rotation and $t \in \mathbb{R}^d$ being the translation.

Starting from Golubev [23] and Kneip and Gasser [30], semiparametric and nonparametric estimation in different instances of problem (1) have been intensively investigated, see for instance [40, 16, 19, 14, 12, 4, 42, 11] for the shifted curve model and [25, 9, 43] for a slightly extended case of affine transforms of shifted curves, [13, 10] for the periodic signal model and [6] for the rigid deformation model. More general deformations have been considered in [37, 38, 21, 29, 7] with applications to image warping [22, 5].

Let us assume now that a collection of sample curves $\{Y_m; m = 1, \ldots, n\}$ is available. Prior to estimating the common template, the deformations or any other object involved in (1), it is natural to check the appropriateness of model (1). The aim of the present work is to develop a theoretically justified approach for carrying out such kind of tests. To achieve this goal, we consider the particular case of shifted curve model or the slightly more general affinely transformed shifted curve model with $n = 2$, *i.e.*, the case where two functions Y and $Y^\#$ are observed such that

$$dY(x) = f(x)dx + \sigma dW(x), \qquad dY^\#(x) = f^\#(x)dx + \sigma^\# dW^\#(x), \qquad \forall x \in [0, 1], \tag{2}$$

where W and $W^\#$ are two independent Brownian motions, f and $f^\#$ are two unknown 1-periodic signals and $\sigma, \sigma^\# > 0$ are positive parameters representing the noise magnitude. The hypothesis we wish to test is that the curves f and $f^\#$ coincide, up to a scale change, a shift of the argument and to a vertical translation:

$$H_0: \text{ there exists some } (a^*, b^*, \tau^*) \in \mathbb{R}^2 \times [0, 1] \text{ s. t. } f(x) = a^* f^\#(x + \tau^*) + b^*, \ \forall x \in [0, 1]. \tag{3}$$

If the null hypothesis H_0 is accepted, then we are in the setting of model (1) for the particular case of deformation given by a shift. Even if the shifted curve model seems to be a very narrow subclass of models given by (1), it plays a central role in several applications. To cite a few of them:

ECG interpretation:: An electro-cardiogram (ECG) can be seen as a collection of replica of nearly the same signal, up to a time shift. Significant informations about heart malformations or diseases can be extracted from the mean signal if we are able to align the available curves, while the deflections would not be so correctly identified if we simply consider the mean of the shifted curves. For more details we refer to [42], where random shifts are considered, and they are estimated along with their common distribution in the asymptotics of a growing number of curves.

Road traffic forecast:: In [32], 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. The mean of the unshifted curves is more significant of a given behavior than the mean of the shifted ones, and hence provides more efficient predictions.

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 their local neighborhoods. More precisely, the regression function of the magnitude of the gradient over the direction of the gradient of the image restricted to a given neighborhood is considered as a local descriptor (cf. the SIFT descriptor [33]). 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.

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. Indeed, both the null hypothesis and the alternative in the context of the present study are nonparametric, *i.e.*, 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 [36] and [26] analyze 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 [20]. 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 [20] 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. Furthermore, their theoretical framework comprises a condition on the sup-norm-entropy of the null hypothesis, which is irrelevant in our set-up and may be violated.

We adopt, in this work, the approach based on the Generalized Likelihood Ratio (GLR) tests, cf. [17] 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 [18].

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. In this relation, let us quote [17]: "While we have observed the Wilks' phenomenon and demonstrated it for a few useful cases, it is impossible for us to verify the phenomenon for all nonparametric hypothesis testing problems. The Wilks' phenomenon needs to be checked for other problems that have not been covered in this paper. In addition, most of the topics outlined in the above discussion remains open and are technically and intellectually challenging. More developments are needed, which will push the core of statistical theory and methods forward."

The rest of the paper is organized as follows. After a brief presentation of the model, we introduce the GLR framework in Section 1. The main results characterizing the asymptotic behavior of the proposed testing procedure, based on GLR testing for a large variety of shrinkage weights, are stated in Section 3. Some numerical examples illustrating the theoretical results are included in Section 4. The proofs of the lemmas and of the theorems are postponed to the Appendix.

1. Model and notation

In the following, we consider the curve registration problem in which the data $\{Y(x) : x \in [0,1]\}$ and $\{Y^{\#}(x) : x \in [0,1]\}$ are available, generated by the Gaussian white noise model

$$dY(x) = f(x)\,dx + \sigma dW(x), \qquad dY^{\#}(x) = f^{\#}(x)\,dx + \sigma^{\#} dW^{\#}(x), \tag{4}$$

where $(W, W^{\#})$ is a two-dimensional Brownian motion. (It is implicitly assumed that f and $f^{\#}$ are squared integrable, which makes model (4) sensible.) This model is often seen as a prototype of nonparametric statistical model, since it is asymptotically equivalent to many other statistical models [8, 34, 24, 15, 39] and it captures main theoretical difficulties of the statistical inference. Let us consider, for the moment, that the noise magnitudes σ and $\sigma^{\#}$ are known and focus on the hypotheses testing problem stemming from the curve registration set-up. Prior to switching to the definition of the generalized likelihood ratio tests, let us recall that model (4) is equivalent to the Gaussian sequence model obtained by projecting the processes $Y(\cdot)$ and $Y^{\#}(\cdot)$ onto the Fourier basis:

$$Y_j = c_j + \sigma\epsilon_j, \qquad Y_j^{\#} = c_j^{\#} + \sigma^{\#}\epsilon_j^{\#}, \qquad j = 0, 1, 2, \dots, \tag{5}$$

where $c_j = \int_0^1 f(x)\, e^{2\mathrm{i}j\pi x}\,dx$ and $c_j^{\#} = \int_0^1 f^{\#}(x)\, e^{2\mathrm{i}j\pi x}\,dx$ 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 $\{c_j; j = 1, 2, \dots\}$ and $\{c_j^{\#}; j = 1, 2, \dots\}$, respectively. We are interested in testing the hypothesis (3), which translates in the Fourier domain to

$$H_0: \qquad \exists\,(a^*, \bar\tau^*) \in \mathbb{R} \times [0, 2\pi[\quad \text{s.t.} \quad c_j = a^* e^{-\mathrm{i}j\bar\tau^*} c_j^{\#} \quad \forall j = 1, 2, \dots. \tag{6}$$

Indeed, one easily checks that the projection onto the functions $e^{2\mathrm{i}j\pi x}$ cancels the term b^* in (3), resulting in (6) with $\bar\tau^* = 2\pi\tau^*$. Furthermore, if (6) is verified, then b^* can be recovered by the formula $b^* = c_0 - a^* c_0^{\#}$. 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 $\boldsymbol{c} = (c_0, c_1, \dots)$ and $\boldsymbol{c}^{\#} = (c_0^{\#}, c_1^{\#}, \dots)$ belong to some Sobolev ball

$$\mathcal{F}_{s,L} = \Big\{ \boldsymbol{u} = (u_0, u_1, \dots) : \sum_{j=1}^{+\infty} j^{2s} |u_j|^2 \le L^2 \Big\},$$

where the positive real numbers s and L stand for the smoothness and the radius of the class $\mathcal{F}_{s,L}$. It is also possible to consider other smoothness classes, as for instance Besov bodies, in which case it would be more appropriate to project not onto the Fourier basis but onto the wavelet basis, as it is done in [41].

2. Penalized Likelihood Ratio Test

In order to convey the main ideas underlying the GLR test analyzed in the present work, we focus on the case where the null hypothesis corresponds to the spatially shifted curve model. This means that in the rest of this section we assume that

$$\begin{cases} H_0: & \exists\,\bar\tau^* \in \mathbb{R}\times[0,2\pi[\quad \text{s.t.} \quad c_j = e^{-\mathrm{i}j\bar\tau^*}c_j^{\#} \quad \forall j=1,2,\ldots\,. \\ H_1: & \inf_\tau \sum_{j=1}^{+\infty} |c_j - e^{-\mathrm{i}j\tau}c_j^{\#}|^2 \geq \rho \end{cases} \tag{7}$$

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.

Because of the Gaussian nature of the noise, the negative log-likelihood of the parameters $\boldsymbol{u}^{\bullet,\#} = (\boldsymbol{u}, \boldsymbol{u}^{\#})$ given the data $\boldsymbol{Y}^{\bullet,\#} = (\boldsymbol{Y}, \boldsymbol{Y}^{\#})$ is

$$\ell(\boldsymbol{Y}^{\bullet,\#}, \boldsymbol{u}^{\bullet,\#}) = \frac{1}{2\sigma^2}\|\boldsymbol{Y} - \boldsymbol{u}\|_2^2 + \frac{1}{2(\sigma^{\#})^2}\|\boldsymbol{Y}^{\#} - \boldsymbol{u}^{\#}\|_2^2. \tag{8}$$

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 $\boldsymbol{u}^{\bullet,\#}$. 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 set

$$p\ell(\boldsymbol{Y}^{\bullet,\#}, \boldsymbol{u}^{\bullet,\#}) = \frac{1}{2\sigma^2}\Big(\|\boldsymbol{Y} - \boldsymbol{u}\|_2^2 + \sum_{j\geq 1}\omega_j|u_j|^2\Big) + \frac{1}{2(\sigma^{\#})^2}\Big(\|\boldsymbol{Y}^{\#} - \boldsymbol{u}^{\#}\|_2^2 + \sum_{j\geq 1}\omega_j|u_j^{\#}|^2\Big). \tag{9}$$

The penalized likelihood ratio test is based on the test statistic

$$\Delta(\boldsymbol{Y}^{\bullet,\#}) = \min_{\boldsymbol{u}^{\bullet,\#}:H_0 \text{ is true}} p\ell(\boldsymbol{Y}^{\bullet,\#}, \boldsymbol{u}^{\bullet,\#}) - \min_{\boldsymbol{u}^{\bullet,\#}} p\ell(\boldsymbol{Y}^{\bullet,\#}, \boldsymbol{u}^{\bullet,\#}). \tag{10}$$

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 functional (9) is an easy exercise and leads to

$$\min_{\boldsymbol{u}^{\bullet,\#}} p\ell(\boldsymbol{Y}^{\bullet,\#}, \boldsymbol{u}^{\bullet,\#}) = \frac{1}{2\sigma^2}\sum_{j\geq 1}\frac{\omega_j}{1+\omega_j}\,|Y_j|^2 + \frac{1}{2(\sigma^{\#})^2}\sum_{j\geq 1}\frac{\omega_j}{1+\omega_j}\,|Y_j^{\#}|^2.$$

Similar but a bit more involved computations lead to the following simple expression:

$$\Delta(\boldsymbol{Y}^{\bullet,\#}) = \frac{\sigma^2 + (\sigma^{\#})^2}{2(\sigma\sigma^{\#})^2}\min_{\tau\in[0,2\pi]}\sum_{j=1}^{+\infty}\frac{|Y_j - e^{-\mathrm{i}j\tau}Y_j^{\#}|^2}{1+\omega_j}. \tag{11}$$

From now on, it will be more convenient to use the notation $\nu_j = 1/(1+\omega_j)$. The elements of the sequence $\boldsymbol{\nu} = \{\nu_j;\ j \geq 1\}$ 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 $(e^{-\mathrm{i}\tau}, e^{-2\mathrm{i}\tau}, \ldots)$. The test statistic can then be written as:

$$\Delta(\boldsymbol{Y}^{\bullet,\#}) = \frac{\sigma^2 + (\sigma^\#)^2}{2(\sigma\sigma^\#)^2} \min_{\tau\in[0,2\pi]} \|\boldsymbol{Y} - \boldsymbol{e}(\tau) \circ \boldsymbol{Y}^\#\|_{2,\boldsymbol{\nu}}^2 \,, \tag{12}$$

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) $\nu_1 = 1, \qquad \text{and} \qquad \nu_j = 0,\ \forall j > N_\sigma,$

(B) for some positive constant $\underline{c}$, it holds that $\sum_{j\geq 1} \nu_j^2 \geq \underline{c} N_\sigma.$

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

(C) $\exists\, \overline{c} > 0$, 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:

$$\nu_j = \begin{cases} \mathbb{1}_{\{j\leq N_\sigma\}}, & & \text{(projection weight)} \\ \big\{1 + \big(\frac{j}{\kappa N_\sigma}\big)^\mu\big\}^{-1} \mathbb{1}_{\{j\leq N_\sigma\}}, & \kappa > 0,\ \mu > 1, & \text{(Tikhonov weight)} \\ \big\{1 - \big(\frac{j}{N_\sigma}\big)^\mu\big\}_+, & \mu > 0. & \text{(Pinsker weight)} \end{cases} \tag{13}$$

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. *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:*

$$\frac{\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#}) - 4\|\boldsymbol{\nu}\|_1}{4\|\boldsymbol{\nu}\|_2} \xrightarrow[\sigma\to 0]{\mathscr{D}} \mathcal{N}(0, 1). \tag{14}$$

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

$$\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#}) \geq 4\|\boldsymbol{\nu}\|_1 + 4z_{1-\alpha}\|\boldsymbol{\nu}\|_2, \tag{15}$$

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 (15) 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:

$$\frac{\|\boldsymbol{\nu}\|_1}{2\|\boldsymbol{\nu}\|_2^2}\,\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#}) \overset{\mathcal{D}}{\approx} \chi^2_{2\|\boldsymbol{\nu}\|_1^2/\|\boldsymbol{\nu}\|_2^2}\,, \qquad \text{as} \qquad \sigma \to 0.$$

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 (15).

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.,*

$$T_\sigma \xrightarrow{P} +\infty, \qquad \textit{as} \qquad \sigma \to 0.$$

In other words, the result above claims that the power of the test defined via (15) 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 A.3). In fact, in most situations $\min\{j \geq 1; j < \bar{c}\}$ is on 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 (15) 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 $\boldsymbol{\nu}$. Following the philosophy of reproducible research, we intend to make our code available for free download on the authors homepages.

4.1. **Convergence of the test under H_0 and the influence of the shrinkage weights.** In order to illustrate the convergence of the test (15) 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[1] $N_\sigma = 50\sigma^{-1/2}$,
- 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 $\{\xi_j\}$,
- 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 (13) and compare these values with the threshold for $\alpha = 5\%$.

We denote by $p_{\text{accept}}^{\text{proj}}(\sigma)$, $p_{\text{accept}}^{\text{Tikh}}(\sigma)$ and $p_{\text{accept}}^{\text{Pinsk}}(\sigma)$ 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 $k \mapsto p_{\text{accept}}^{\text{proj}}(\sigma_k)$, $k \mapsto p_{\text{accept}}^{\text{Tikh}}(\sigma_k)$ and $k \mapsto p_{\text{accept}}^{\text{Pinsk}}(\sigma_k)$, 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.

[1] This value of N_σ satisfies the assumptions required by our theoretical results.

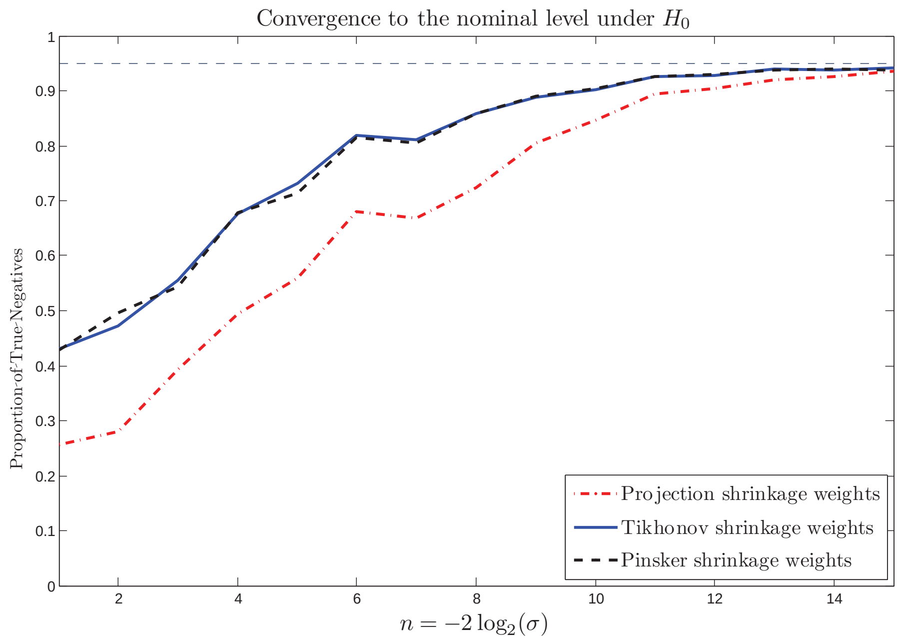

FIGURE 1. The proportion of true negatives in the experiment described in Section 4.1 as a function of $\log_2 \sigma^{-2}$ for three different shrinkage weights: projection (Left), Tikhonov (Middle) and Pinsker (Right). One can observe that for $\sigma = 2^{-15/2} \approx 5 \times 10^{-3}$, the proportion 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.

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.

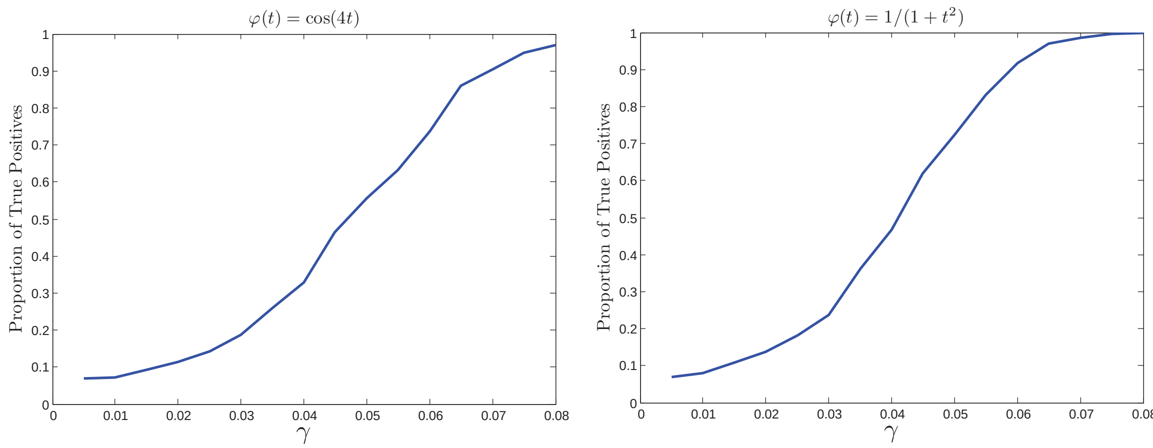

FIGURE 2. The proportion of true positives in the experiment described in Section 4.2 as a function of the parameter γ measuring the distance between the true parameter and the set of parameters characterizing the null hypothesis. The main observation is that both curves tend to 1 very rapidly.

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. It can be interesting, however, to develop a direct inference in these models. In particular, the model of spatial Poisson processes (cf. [27]) can be of special interest because of its applications in image analysis.

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 [28] 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. The adaptive choice of the tuning parameter N_σ is another central issue that has not been answered in the present paper. We envisage to tackle this issue in a future work.

Appendix A. Proofs of the theorems

The proof of Wilks' phenomenon is divided into several parts. First we assume that H_0 is true and study the convergence of the pseudo-estimator $\hat\tau$ (of the shift $\bar\tau^*$) defined as the maximizer of the log-likelihood over the interval $[\bar\tau^*-\pi,\bar\tau^*+\pi]$. Here, $\bar\tau^*$ is an element of $[0,2\pi[$ such that $c_j = e^{-\mathrm{i}j\bar\tau^*}c_j^{\#}$, for all $j\geq 1$.

A.1. **Maximizer of the log-likelihood.**

Proposition 1. *Let H_0 be satisfied, $\boldsymbol{c}\in\mathcal{F}_{1,L}$ and $|c_1|>0$. If the shrinkage weights ν_j satisfy conditions* **(A)** *and* **(B)**, *then the solution $\hat\tau$ to the optimization problem*

$$\hat\tau = \arg\max_{\tau:|\tau-\bar\tau^*|\leq\pi} M(\tau),\qquad \textit{with}\qquad M(\tau)=\sum_{j\geq 1}\nu_j \operatorname{Re}(e^{ij\tau}Y_j\overline{Y_j^{\#}})$$

satisfies the asymptotic relation

$$|\hat\tau-\bar\tau^*| = \sigma\sqrt{\log N_\sigma}\big(1+\sigma N_\sigma^{3/2}\big)O_P(1),\qquad \textit{as}\quad \sigma\to 0.$$

Proof of Proposition 1. If we set $\eta_j = e^{-\mathrm{i}j\bar\tau^*}\epsilon_j$ and $\eta_j^{\#}=\epsilon_j^{\#}$, we can write the decomposition

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

where

$$\begin{aligned}
\mathbf{E}[M(\tau)] &= \sum_{j\geq 1}\nu_j|c_j|^2\cos[j(\tau-\bar\tau^*)],\\
S(\tau) &= \sum_{j\geq 1}\nu_j \operatorname{Re}\Big(e^{\mathrm{i}j\tau}(\overline{c_j}\eta_j + c_j\overline{\eta_j^{\#}})\Big),\\
D(\tau) &= \sum_{j\geq 1}\nu_j \operatorname{Re}\big(e^{\mathrm{i}j\tau}\eta_j\overline{\eta_j^{\#}}\big).
\end{aligned}$$

On the one hand, using the assumption $|c_1|>0$ along with condition (A), we get that

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

Therefore,

$$\begin{aligned}
M(\tau)-M(\bar\tau^*) &= \mathbf{E}[M(\tau)]-\mathbf{E}[M(\bar\tau^*)] + \sigma\big[S(\tau)-S(\bar\tau^*)\big] + \sigma^2\big[D(\tau)-D(\bar\tau^*)\big]\\
&\leq -C\,|\tau-\bar\tau^*|^2 + \sigma|\tau-\bar\tau^*|\cdot\|S'\|_\infty + \sigma^2|\tau-\bar\tau^*|\cdot\|D'\|_\infty\\
&= |\tau-\bar\tau^*|\big\{\sigma\|S'\|_\infty + \sigma^2\|D'\|_\infty - C|\tau-\bar\tau^*|\big\}.
\end{aligned}$$

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

$$\begin{aligned}\mathbf{P}\big(|\hat\tau - \bar\tau^*| > a\big) &\le \mathbf{P}\Big\{\sup_{|\tau-\bar\tau^*|>a} M(\tau) - M(\bar\tau^*) \ge 0\Big\}\\ &\le \mathbf{P}\Big\{\sup_{|\tau-\bar\tau^*|>a} \big[\sigma\|S'\|_\infty + \sigma^2\|D'\|_\infty - C|\tau-\bar\tau^*|\big] \ge 0\Big\}\\ &\le \mathbf{P}\Big\{\sigma\|S'\|_\infty + \sigma^2\|D'\|_\infty \ge Ca\Big\}.\end{aligned}$$

Choosing $a = \sigma\sqrt{\log N_\sigma}\big(2+\sigma N_\sigma^{3/2}\big)z$, we get

$$\begin{aligned}\mathbf{P}\big(|\hat\tau - \bar\tau^*| > \sigma\sqrt{\log N_\sigma}\big(1+\sigma N_\sigma^{3/2}\big)z\big) \le \mathbf{P}\big(\|S'\|_\infty \ge 2Cz\sqrt{\log N_\sigma}\big)&\\ + \mathbf{P}\big(\|D'\|_\infty \ge Cz\sqrt{N_\sigma^3 \log N_\sigma}\big)&.\end{aligned}$$

On the other hand, since

$$S'(t) = \sum_{j\ge1} j|c_j|\nu_j \operatorname{Re}\big(e^{\mathrm{i}j\tau}\zeta_j\big),$$

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 1. *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*

$$\mathbf{P}(\|S\|_\infty \ge \|\boldsymbol{s}\|_2 x) \le (K+1)e^{-x^2/2}, \qquad \forall x > 0.$$

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

Using this lemma and the fact that $N_\sigma \ge 2$, 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 3 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 on the order $\sigma\sqrt{\log N_\sigma}\big(1 + \sigma N_\sigma^{3/2}\big)$. □

A.2. **Proof of Theorem 1.** One can check that, under H_0,

$$\Delta_\sigma(\mathbf{Y}^{\bullet,\#}) = \frac{1}{\sigma^2}\min_{\tau\in[0,2\pi[}\left[\sum_{j=1}^{+\infty}\nu_j\big|Y_j - e^{-\mathrm{i}j\tau}Y_j^{\#}\big|^2\right] = \frac{1}{\sigma^2}\min_{|\tau|\le\pi}\big\{D_\sigma(\tau) + 2C_\sigma(\tau) + P_\sigma(\tau)\big\}, \tag{16}$$

where we have used the notation:

$$D_\sigma(\tau) = \sum_{j=1}^{+\infty} \nu_j |c_j|^2 \big|1 - e^{-\mathrm{i}j(\tau-\bar\tau^*)}\big|^2, \qquad \text{(deterministic term)}$$

$$C_\sigma(\tau) = \sigma \sum_{j=1}^{+\infty} \nu_j \operatorname{Re}\big[c_j\big(1 - e^{-\mathrm{i}j(\tau-\bar\tau^*)}\big)\big(\overline{\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^{\#}}\big)\big], \qquad \text{(cross term)}$$

$$P_\sigma(\tau) = \sigma^2 \sum_{j=1}^{+\infty} \nu_j \big|\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^{\#}\big|^2. \qquad \text{(principal term)}$$

(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 $j \geq 1$.) We denote by $\hat\tau$ the pseudo-estimator of $\bar\tau^*$ defined as the minimizer of the RHS of (16) and study the asymptotic behavior of the terms D_σ, C_σ and P_σ separately.

- For the deterministic term, it holds that

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

- Let us turn now to the cross term. It holds that:

$$\begin{aligned} C_\sigma(\tau) = \sigma \sum_{j=1}^{+\infty} \nu_j \Big\{ &(1 - \cos[j(\tau - \bar\tau^*)]) \operatorname{Re}\Big[c_j\big(\overline{\epsilon_j - e^{-\mathrm{i}j\bar\tau^*}\epsilon_j^{\#}}\big)\Big] \\ &+ \sin[j(\bar\tau^* - \tau)] \operatorname{Im}\Big[c_j\big(\overline{\epsilon_j + e^{-\mathrm{i}j\bar\tau^*}\epsilon_j^{\#}}\big)\Big]\Big\}. \end{aligned}$$

  Thus, as $C_\sigma(\bar\tau^*) = 0$, we have

$$|C_\sigma(\hat\tau)| \leq |\hat\tau - \bar\tau^*| \cdot \|C'_\sigma\|_\infty.$$

  By arguments similar to those used in the proof of Proposition 1, we check that $\|C'_\sigma\|_\infty$ is on 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, $P_\sigma(\tau) = \sigma^2 \sum_{j=1}^{+\infty} \nu_j \big|\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^{\#}\big|^2$, 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 $P_\sigma(\tau) = \sigma^2 \sum_{j=1}^{+\infty} \nu_j \big|\eta_j - e^{-\mathrm{i}j(\tau-\bar\tau^*)}\eta_j^{\#}\big|^2$. We wish to prove 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

$$T_\sigma(\hat\tau) = \frac{P_\sigma(\hat\tau) - 4\sigma^2 \sum_{k\geq 1}\nu_k}{4\sigma^2(\sum_{k\geq1}\nu_k^2)^{1/2}} \xrightarrow[\sigma\to 0]{\mathscr{D}} \mathcal{N}(0,1).$$

  To check this property, we decompose the principal term as follows:

$$T_\sigma(\hat\tau) = T_\sigma(\bar\tau^*) + \underbrace{\frac{P_\sigma(\hat\tau) - P_\sigma(\bar\tau^*)}{4\sigma^2(\sum_{k\geq1}\nu_k^2)^{\frac12}}}_{R_\sigma(\hat\tau)}.$$

- We start by witting $T_\sigma(\bar\tau^*)$ as

$$T_\sigma(\bar\tau^*) = \sum_{j=1}^{N_\sigma} X_{j,\sigma}, \text{ with } X_{j,\sigma} = \frac{\nu_j(|\eta_j - \eta_j^{\#}|^2 - 4)}{4(\sum_{k\geq 1}\nu_k^2)^{\frac12}},$$

and applying the Berry-Esseen inequality [35, Theorem 5.4], which is possible since the $X_{j,\sigma}$'s are independent random variables with mean 0 and finite third moment. Furthermore, we have $B_\sigma = \sum_{j=1}^{N_\sigma} \mathbf{V}ar\big(X_{j,\sigma}\big) = 1$ and $L_\sigma = B_\sigma^{-\frac32}\sum_{j=1}^{N_\sigma}\mathbf{E}|X_{j,\sigma}|^3 \leq C\,N_\sigma^{-\frac12}$. Therefore, the Berry-Esseen inequality yields

$$\sup_x |F_\sigma(x) - \Phi(x)| \leq K\,L_\sigma,$$

where $F_\sigma(x) = \mathbf{P}\big(B_\sigma^{-\frac12}\sum_{j=1}^{N_\sigma} X_{j,\sigma} < x\big)$, Φ is the c.d.f. of the standard Gaussian distribution and K is an absolute constant. Hence

$$T_\sigma(\bar\tau^*) \xrightarrow[\sigma\to 0]{\mathscr{D}} \mathcal{N}(0,1).$$

- It remains now to prove that R_σ tends to 0 in probability, which—in view of Slutski's lemma—will be sufficient for completing the proof. It holds that

$$R_\sigma(\tau) = \sum_{j=1}^{+\infty} \frac{\nu_j}{2(\sum_{j=1}^{+\infty}\nu_j^2)^{\frac12}} \operatorname{Re}\eta_j\overline{\eta_j^{\#}}\,(e^{\mathrm{i}j(\tau-\bar\tau^*)} - 1) = \sum_{j=1}^{N_\sigma} \frac{j\nu_j(\tau-\bar\tau^*)}{2(\sum_{j=1}^{+\infty}\nu_j^2)^{\frac12}} \operatorname{Re}\big(e^{\mathrm{i}jt}\eta_j\overline{\eta_j^{\#}}\big),$$

with t some real number between τ and $\bar\tau^*$. Then, by virtue of Lemma 3,

$$|R_\sigma(\hat\tau)| \leq \frac{|\hat\tau - \bar\tau^*|}{2(\sum_{j=1}^{+\infty}\nu_j^2)^{\frac12}} \sup_{t\in[0,2\pi]}\Big|\sum_{j=1}^{N_\sigma} j\nu_j\operatorname{Re}\big(e^{\mathrm{i}jt}\eta_j\overline{\eta_j^{\#}}\big)\Big| = \{\sigma(1+\sigma N_\sigma^{3/2})N_\sigma\log N_\sigma\}\cdot O_P(1).$$

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

A.3. **Proof of Theorem 2.** Let us 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. It holds true that:

$$\begin{aligned}
\Delta_\sigma(\boldsymbol{Y}^{\bullet,\#}) &= \frac{1}{\sigma^2}\min_{\tau\in[0,2\pi]}\sum_{j=1}^{+\infty}\nu_j\big|Y_j - e^{-\mathrm{i}j\tau}Y_j^{\#}\big|^2\\
&= \frac{1}{\sigma^2}\min_{\tau\in[0,2\pi]}\sum_{j\geq 1}\nu_j\Big|\big(c_j - e^{-\mathrm{i}j\tau}c_j^{\#}\big) + \sigma(\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^{\#})\Big|^2\\
&\geq \frac{1}{\sigma^2}\min_{\tau\in[0,2\pi]}\Big\{\sum_{j\geq 1}\nu_j|c_j - e^{-\mathrm{i}j\tau}c_j^{\#}|^2\Big\} - \frac{2}{\sigma}\max_{\tau\in[0,2\pi]}\Big\{\sum_{j\geq 1}\nu_j|c_j - e^{-\mathrm{i}j\tau}c_j^{\#}|\cdot|\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^{\#}|\Big\}.
\end{aligned}$$

Let us focus on the first term. Denoting $\delta_\sigma = \min\{j \geq 1, \nu_j < \overline{c}\}$, we get by condition **(C)** that $\delta_\sigma \to +\infty$, which implies

$$\begin{aligned}
\min_{\tau\in[0,2\pi]} \sum_{j\geq 1} \nu_j |c_j - e^{-\mathrm{i}j\tau} c_j^{\#}|^2 &\geq \overline{c} \min_{\tau\in[0,2\pi]} \sum_{j=1}^{\delta_\sigma} |c_j - e^{-\mathrm{i}j\tau} c_j^{\#}|^2 \\
&\geq \overline{c}\Big(\min_{\tau\in[0,2\pi]} \sum_{j=1}^{+\infty} |c_j - e^{-\mathrm{i}j\tau} c_j^{\#}|^2 - 4L\delta_\sigma^{-2}\Big) \\
&\geq \overline{c}\big(\rho - 4L\delta_\sigma^{-2}\big).
\end{aligned}$$

Now, the second term satisfies

$$\begin{aligned}
\max_{\tau\in[0,2\pi]} \sum_{j\geq 1} \nu_j \big|c_j - e^{-\mathrm{i}j\tau} c_j^{\#}\big| \cdot \big|\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^{\#}\big| &\leq \max_{j=1,\ldots,N_\sigma} (|\epsilon_j| \vee |\epsilon_j^{\#}|) \sum_{j\geq 1}^{N_\sigma} \Big(|c_j| + |c_j^{\#}|\Big) \\
&= O_P(\sqrt{\log N_\sigma}) \sum_{j\geq 1}^{N_\sigma} \Big(|c_j| + |c_j^{\#}|\Big) \\
&\leq O_P(\sqrt{\log N_\sigma}) \Big(\sum_{j\geq 1}^{N_\sigma} j^{-2}\Big)^{1/2} \Big(\sum_{j\geq 1}^{N_\sigma} j^2 \Big(|c_j| + |c_j^{\#}|\Big)^2\Big)^{1/2} \\
. &= O_P(\sqrt{\log N_\sigma}).
\end{aligned}$$

Putting it all together, we get

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

Appendix B. Bounds for the maxima of random sums

In this section, we will give some technical lemmas which will be useful in the proofs of this paper. We are interested in bounding the maximum of the sum of a growing number of terms, so that the non-asymptotic result given in [3] will be useful:

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*

$$\mathbf{P}\bigg(\sup_{[a,b]} \sum_{j=1}^n g_j(t)\xi_j \geq x\bigg) \leq \frac{L_0}{2\pi} e^{-\frac{x^2}{2}} + \int_x^{+\infty} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\,dt, \quad \textit{with} \quad L_0 = \int_a^b \bigg[\sum_{j=1}^n g_j'(t)^2\bigg]^{1/2} dt.$$

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 2. *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 $\|\boldsymbol{s}\|_\infty = \max_{j=1,\dots,N} |s_j|$ *and* $\|\boldsymbol{s}\|_q^q = \sum_{j=1}^N |s_j|^q$.

Proof. This is a direct consequence of [31, Lemma 1]. $\square$

Lemma 3. *Let* N *be some positive integer and let* η_j, $\eta_j^\#$, $j = 1,\dots,N$ *be independent complex valued random variables such that their real and imaginary parts are independent standard Gaussian variables. Let* $\boldsymbol{s} = (s_1,\dots,s_N)$ *be a vector of real numbers. Denote* $S(t) = \sum_{j=1}^N s_j \operatorname{Re}\left(e^{ijt}\eta_j\eta_j^\#\right)$ *for every* t *in* $[0,2\pi]$ *and* $\|S\|_\infty = \sup_{t\in[0,2\pi]} |S(t)|$. *Then,*

$$\mathbf{P}\Big\{\|S\|_\infty > \sqrt{2}x\big(\|\boldsymbol{s}\|_2 + y\|\boldsymbol{s}\|_\infty\big)\Big\} \le (N+1)e^{-x^2/2} + e^{-y^2/2}, \qquad \forall x, y > 0.$$

Proof. First note that we can not directly use Berman's formula, since the summands are not Gaussian. However, they are conditionally Gaussian if the conditioning is done, for example, with respect to the sequence $(\eta_j^\#)$. Indeed,

$$\sum_{j=1}^N s_j \operatorname{Re}\left(e^{\mathrm{i}j\tau}\eta_j\eta_j^\#\right) \Big| (\eta_j^\#) \sim \sum_{j=1}^N s_j|\eta_j^\#|\big(\cos(j\tau)\xi_j - \sin(j\tau)\xi_j'\big) \text{ with } \xi_j, \xi_j' \overset{iid}{\sim} \mathcal{N}(0,1).$$

It follows by Lemma 1 that

$$\mathbf{P}\bigg(\sup_{[0,2\pi]}\Big|\sum_{j=1}^N s_j \operatorname{Re}\left(e^{\mathrm{i}j\tau}\eta_j\eta_j^\#\right)\Big| \ge x\Big(\sum_{j=1}^N s_j^2|\eta_j^\#|^2\Big)^{\frac{1}{2}}\Big|(\eta_j^\#)\bigg) \le (N+1)\exp\Big(-\frac{x^2}{2}\Big).$$

Let us now denote by ζ the square root of the random variable $\sum_{j=1}^N s_j^2|\eta_j^\#|^2$. It is clear that for all $a > 0$,

$$\begin{aligned}\mathbf{P}(\|S\|_\infty \ge ax) &= \mathbf{P}(\|S\|_\infty \ge ax;\ \zeta \le a) + \mathbf{P}(\|S\|_\infty \ge ax;\ \zeta > a)\\ &\le \mathbf{P}(\|S\|_\infty \ge x\zeta) + \mathbf{P}(\zeta > a)\\ &\le (N+1)e^{-x^2/2} + \mathbf{P}(\zeta > a).\end{aligned}$$

To complete the proof, it suffices to replace a by $\sqrt{2}(\|\boldsymbol{s}\|_2 + y\|\boldsymbol{s}\|_\infty)$ and to apply Lemma 2 along with the inequalities $\|\boldsymbol{s}\|_2 + \|\boldsymbol{s}\|_\infty y = (\|\boldsymbol{s}\|_2^2 + 2\|\boldsymbol{s}\|_\infty\|\boldsymbol{s}\|_2 y + \|\boldsymbol{s}\|_\infty^2 y^2)^{1/2} \ge (\|\boldsymbol{s}\|_2^2 + \sqrt{2}\|\boldsymbol{s}\|_4^2 y + \|\boldsymbol{s}\|_\infty^2 y^2)^{1/2}$. $\square$

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. Application to testing qualitative hypotheses on a regression function. *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. ISSN 0882-0287.

[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. ISSN 0924-9907. doi: 10.1007/s10851-009-0146-1. URL `http://dx.doi.org/10.1007/s10851-009-0146-1`.

[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] J. Bigot, J.-M. Loubès, and M. Vimond. Semiparametric estimation of shifts on compact lie groups for image registration. *Probab. Theory Related Fields*, To appear, 2011.

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

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

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

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

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

[13] I. Castillo, C. Lévy-Leduc, and C. Matias. Exact adaptive estimation of the shape of a periodic function with unknown period corrupted by white noise. *Math. Methods Statist.*, 15(2):146–175, 2006.

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

[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] 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.

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

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

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

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

[21] D. Gervini and T. Gasser. Nonparametric maximum likelihood estimation of the structural mean of a sample of curves. *Biometrika*, 92(4):801–820, 2005.

[22] 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.

[23] 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.

[24] I. Grama and M. Nussbaum. Asymptotic equivalence for nonparametric generalized linear models. *Probab. Theory Related Fields*, 111(2):167–214, 1998.

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

[26] 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.

[27] Yu. I. Ingster and Yu. A. Kutoyants. Nonparametric hypothesis testing for intensity of the Poisson process. *Math. Methods Statist.*, 16(3):217–245, 2007.

[28] 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.

[29] A. B. Juditsky, O. V. Lepski, and A. B. Tsybakov. Nonparametric estimation of composite functions. *Ann. Statist.*, 37(3):1360–1404, 2009.

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

[31] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, 28(5):1302–1338, 2000.

[32] 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.

[33] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. ISSN 0920-5691.

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

[35] V. V. Petrov. *Limit theorems of probability theory*, volume 4 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1995. ISBN 0-19-853499-X. Sequences of independent random variables, Oxford Science Publications.

[36] C. Pouet. An asymptotically optimal test for a parametric set of regression functions against a non-parametric alternative. *J. Statist. Plann. Inference*, 98(1-2):177–189, 2001.

[37] J. O. Ramsay and X. Li. Curve registration. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60 (2):351–363, 1998.

[38] C. Reilly, P. Price, A. Gelman, and S. A. Sandgathe. Using image and curve registration for measuring the goodness of fit of spatial and temporal predictions. *Biometrics*, 60(4): 954–964, 2004.

[39] M. Reiß. Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann. Statist.*, 36(4):1957–1982, 2008.

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

[41] V. G. Spokoiny. Adaptive hypothesis testing using wavelets. *Ann. Statist.*, 24(6):2477–2498, 1996.

[42] 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.

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