

JSPI-D-14-00463 - Point-to-point reply to AE's and Reviewer's comments

Olivier Collier and Arnak Dalalyan

December 4, 2014

1 General comments

We would like to thank the Associate Editor and the Reviewer for their prompt reviews containing insightful comments. We have substantially revised our manuscript taking into account all the issues raised by the AE and the Reviewer. In particular, we have performed additional numerical experiments for illustrating the performance of the noise-level-adaptive test procedure described in Section 3.2.

We have also carefully revised the Matlab code of the experiments reported in Section 4 and, in order to ensure the reproducibility, made it freely available on <https://code.google.com/p/shifted-curve-testing/>.

In addition, we changed the layout of the manuscript using the L^AT_EX style file provided by the publisher.

We believe that these modifications improved significantly the presentation of the material. Specific details answering the comments of each Reviewer separately are provided below.

2 Reply to the AE's comments

1.

1. From a technical point of view, the paper follows a standard routine in Theorem 1 but the result is nice. Theorem 2 is for me quite incomplete since no rates are described here, but the authors argue that the companion paper of [Collier,12] provides the missing results. I have checked this, and thus have carefully read also the companion paper . . . Then, I have been quite confused about the intertwining of the proofs between the two papers. In this submitted paper, the argument are self contained but some of them are already used in the EJS paper. Hence, I would advise the authors to improve the presentation of the technical results, and to underline what are the supplementary technical difficulties here, even if I understand the justifiable repetitions.

For example: the authors use the Berman formula on the supremum of Gaussian processes. This formula is available for the supremum of the sums of the form

$$\sum_{j=1}^n g_j(t)\xi_j,$$

where the functions g_j are smooth functions. But this novelty should be balanced by the fact that it is already used in the paper of [Collier,12].

We agree that several arguments used in the proofs of the present paper are also used in [Collier,12]. However, the perspectives of [Collier,12] and of the present paper are

substantially different. While [Collier,12] studies the limits of testing and provides the (nearly) optimal separation rate between the null and the alternative hypotheses, the present work is concerned with the possibility of establishing Wilks' property for the penalized likelihood ratio test. In this paper, the asymptotic behavior of the test under H_0 is carefully studied, whereas in [Collier,12] only an upper bound on the probability of making an error under H_0 is obtained. We could say that this paper is concerned with a very practical issue, i.e. setting the threshold of the GLR test, while [Collier,12] was a theoretical study where the analysis was carried out up to constants and no clear guidance for choosing the threshold was provided.

Besides, even if we agree that the arguments of our proofs are standard (Berry-Esseen inequality, tails of chi-square, Berman's formula, ...), we feel that this does not make the results less interesting, as they are novel in both papers. So, in a nutshell, the results on Gaussian processes are not the point of this paper, nor were they the point of [Collier,12]. We believe that it is suitable to give them in both papers for the sake of self-containedness. Following the suggestion of the AE, we added the following sentence in the beginning of the Appendix (page 18):

For the sake of self-containedness, we provide all the details of the proofs although some of them—such as the Berman theorem—have been already used in earlier references (see, for instance, [Collier,12]).

2.

The, authors make the assumption that the signal f belongs to $\mathcal{F}_{1,L}$. It is indeed an important assumption regarding the framework of curve registration. In fact, one should expect that when the function f is not regular, it may be easier to estimate shifts and then derive statistical testing procedure. This last point should be at the least included in a remark in the text. What should be wrong is the smoothness parameter s is lower than 1? Standard results are generally obtained for $s > 1/2$... The balance between the size of N_{σ^} and σ^* should be commented and related to the bias and variance tradeoff involved in the Fourier cut-off. Finally, it is not clear to me with in the statements of Theorem 1 it is not possible to consider some values of s lower than 1.*

Following your remark, we have rewritten the proof of Theorem 1 in the case of general $s > 0$. It appears from the proof that the condition we really need is $s > 7/8$. We added the following discussion just after Remark 1 to give the reader some insight on this point. *The proof of Theorem 1 is rather technical and, therefore, is deferred to the Appendix. Let us simply mention here that we present the proof in a slightly more general case $c \in \mathcal{F}_{s,L}$ with a smoothness $s \in (0, 1]$. It appears from the proof that the convergence stated in (12) holds if $s > 7/8$, $N_{\sigma^*} \rightarrow \infty$ and $\sigma_*^2 N_{\sigma_*}^{-2s+9/2} \log(N_{\sigma_*}) \rightarrow 0$, as $\sigma_* \rightarrow 0$. We do not know whether the last condition on N_{σ_*} can be avoided by using other techniques, but it seems that in our proof there is no room for improvement in order to relax this assumption. At a heuristic level, it is quite natural to avoid choosing N_{σ_*} too large. Indeed, large N_{σ_*} leads to undersmoothing in the problem of estimating the quadratic functional $\|c - e(\tau) \circ c^\#\|_2^2$. Therefore, the test statistic corresponds to registration of two curves in which the signal is dominated by noise, which is clearly not a favorable situation for performing curve registration.*

3.

I would explicitly said that the statistic Δ_{σ^} depends on the weight sequence ν .*

Done.

4.

The frequency threshold N_{σ^} used in Theorem 1 plays a significant role. Some discussions should be helpful for non familiar readers.*

Done (cf. the last paragraph of Page 7).

5. | *Page 8, l-8: a comma is missing after 'these assumptions'.*

Indeed, we have added the comma.

6. | *The statement of Theorem 3 is quite long. Maybe authors should try to alleviate it.*

We have now removed the second statement of the theorem, which was overly technical and probably not very interesting. We have also used this opportunity for simplifying the proof and slightly sharpening the result. In particular, the condition imposed on p and N_{σ_*} has been improved.

7. | *The balance between (p, N_{σ_*}) and σ_* should be more discussed, especially regarding the assumption on Theorem 1.*

We completely agree with this point and, therefore, have added the following discussion at the end of Section 3:

*To conclude this section, let us have a closer look at the assumptions of the last theorem. The conditions **(A)**, **(B)** and $\max_{j \neq 0} j^{-2}(1 - \nu_j) \leq c' N_{\sigma_*}^{-2}$ are satisfied for most weights used in practice. Thus, the most important conditions are $(p - N_{\sigma_*})\sigma_*^2 N_{\sigma_*}^{3/2} \rightarrow \infty$ and $\sigma_* N_{\sigma_*} \rightarrow 0$. The first of these two conditions ensures that the error term coming from the estimation of the unknown noise level is small. The second one is a weak version of the condition $\sigma_* N_{\sigma_*}^{5/2} \log N_{\sigma_*} = o(1)$ present in Theorem 1, which ensures that we do not use a strongly undersmoothed test statistic. We manage here to obtain a condition on N_{σ_*} which is weaker than the corresponding condition in Theorem 1 because we do not establish the asymptotic distribution of the test statistic but just an upper bound of the latter. The choice of p and N_{σ_*} is particularly important for obtaining a test with a power close to one, especially in the case of alternatives that are close to the null. However, the investigation of this point is out of scope of the present work. Let us just mention that if we choose $N_{\sigma_*} = \sigma_*^{-\beta}$ and $p = 2\sigma_*^{-\gamma}$, the conditions of Theorem 3 are satisfied if $\beta \leq \min(\gamma, 1)$ and $2\gamma + 3\beta \geq 4$. Of course, these conditions are closely related to the assumption that the unknown signal belongs to the smoothness class of regularity 1.*

8. | *In the numerical experiments, authors are concerned with the HeavySine function, but I would like to see the effect of the smoothness of the unknown functions.*

We have now modified the function used in the experiments by using the smoothed version of the HeavySine function. In order to observe the impact of the smoothness, we performed a new experiment (described in Section 4.2) in which the perturbation function φ is chosen to have a smoothness degree lower than the one of the function $\varphi(t) = c/(1+t^2)$. We see that this results in a much lower power, which is consistent with the existing theoretical results.

9. | *There is a typo page 12 line -7.*

Actually this was not a typo, but a character used by the JRSS B style file for indicating the footnote. Since we use the Elsevier style file now, this character does not appear anymore.

10. | *I do not find in the paper the justification of 4 rings for the LoFT descriptors. I guess it depends on the size of the radius r and the smoothness of the signal. Some details should be helpful.*

There is no theoretical justification for this choice. We just tried several values for the number of rings and the number 4 leaded to the best results.

11. | *Concerning the deviation of the supremum: could the authors illustrate the technical differences reached either using the Berman formula or using the Dudley + Talagrand routine? It would be nice for the reader to know what argument provides the best results.*

In the context of the present work, both approaches lead to optimal results. However, the Berman formula is much easier to apply since it avoids the computation of the covering numbers or other quantities of the same flavor. It should be however emphasized that this is made possible by the fact that the noise is assumed Gaussian and the functions $g_j(t)$ are continuously differentiable. In a more general situation (non Gaussian noise, nonsmooth sample paths, etc.) one would most likely have no other choice than applying the Dudley + Talagrand routine. This discussion is now added to the paper (page 26, first paragraph).

3 Answer to the Reviewer

1.

Equation (7) : after a bit more involved computations, the authors claim to obtain Equation (7). The reader may find computation details in the companion paper [1]. When the model is homoscedastic ($\sigma = \sigma^\#$), it is straightforward to obtain Equation (7). But when $\sigma \neq \sigma^\#$, the minimum of the penalized log-likelihood is reached for linear combination of Fourier coefficients Y_j and $Y_j^\#$ which weights may depend of σ and $\sigma^\#$ respectively. In this case, I do not understand why this weights would disappear in the final result : when I compare naively the terms in $|Y_j|^2$ in Equation (6) and Equation (7), it does not match. Then the test statistic appears to me more as a test which is built from a contrast and not as a penalized likelihood ratio test. Could the authors give more details about their computations ?

[1] A.S. Dalalyan and O. Collier. Wilks' phenomenon and penalized likelihood ratio test for nonparametric curve registration. *Journal of Machine Learning Research - Proceedings Track*, 22: 264-272, 2012.

We added in Section 1 a new lemma, Lemma 1, stating the validity of Equation (7). The proof of this lemma is also presented. This made the section a bit longer, but we agree that these computations are worth being included in the paper, especially that the corresponding result of the paper [1] was stated with a mistake (the factor in front of the min was not correctly computed).

2.

The Fourier coefficient $(Y_0; Y_0^\#)$ disappears while the null hypothesis, Equation (4), takes it account. On the other hand, this would explain why the authors claim that they are able to whether two points in images coincide up to a rotation and an illumination change for keypoint matching application. The authors should clarify this point.

Reviewer 1 is perfectly right, there were an inconsistency between the statement of the null and alternative hypotheses and the definition of the test statistics. This inconsistency is repaired in the revised version by changing the definition of the null and the alternative to the following one:

$$\mathbf{H}_0 : \quad \exists (\tau^*, a^*) \in [0, 1] \times \mathbb{R} \quad \text{such that} \quad f(\cdot) = f^\#(\cdot + \tau^*) + a^*, \quad (1)$$

which means that under the null the functions f and $f^\#$ coincide up to a spatial shift (not necessarily horizontal).

3.

Numerical experiments : the test procedure should take into account a sufficient large number of Fourier coefficients. On Figure 1, the number of Fourier coefficients is fixed according to the level of the noise : $N_{\sigma_*} = 50\sigma_*^{-1/2}$. For large value of sigma, the test nominal level is not close $1 - \alpha = 0.95$. Have the authors investigated for these values the effect of N_{σ_*} in order to get maybe a more satisfactory coverage level? The Tikhonov and Pinsker weights depend on parameters κ and ν which are not set for the numerical experiments. Moreover, it would be desirable to present simulation on the unknown noise level test procedure.

We have carefully checked our code and it turns out that the poor behavior of the Type I error rate for small values of N_{σ_*} (or, equivalently, large values of σ_*) was caused by our implementation. In fact, we compute the minimum over $\tau \in [0, 1]$ by an exhaustive search over a fine grid and, in the previous experiment, the mesh of the grid was not chosen small enough. The new experiments reported in Section 4 of the revised version

are carried out with a smaller mesh and the resulting Type I error rate is always smaller than the nominal level.

We now make precise the parameters of the Pinsker and the Tikhonov weights. Thank you for pointing out this flaw.

Finally, numerical experiments for the noise-level adaptive test procedure are presented in the new version of the manuscript, cf. Section 4 (Figure 1, right panel, and Figure 3).

4. | *page 2, line 8 : notation error for Brownian motion $B_n^\#$.*

Thank you for spotting out this typo. There were two other similar typos in the text. They are all corrected in the new version.

5. | *Equation (2) : the authors define the complex random variables ϵ_j and $\epsilon_j^\#$ as i.i.d. Gaussian $N_{\mathbb{C}}(0, 1)$, which means that their real and imaginary parts are independent $N(0, 1)$ random variables. But from the starting model (first Equation on page 2), we do not get this settings since the L2-norm of the real part and imaginary part of $t \mapsto e^{i2\pi jt}$ is equal to $1/\sqrt{2}$. This could be confusing to the reader since the variance may modify the normalization of the test statistic, Equation (10) ?*

This is a very good remark. In the new version of the paper, we consider separately the cases $j = 0$ and $j > 0$. In the second case, we divide the standard deviation by $\sqrt{2}$ in order to get standard Gaussian noise (cf. Eq. (2)). We have also modified all the subsequent parts of the manuscript in order to make it consistent with this change.