A model of Poissonian interactions and detection of dependence

Laure Sansonnet, Christine Tuleau-Malot

A model of Poissonian interactions and detection of dependence

Laure Sansonnet
Laboratoire de Mathématiques, CNRS UMR 8628
Université Paris-Sud
15 rue Georges Clémenceau 91405 Orsay Cedex, France
E-mail: laure.sansonnet@math.u-psud.fr
and
Christine Tuleau-Malot
Laboratoire J. A. Dieudonné, CNRS UMR 7351
Université de Nice Sophia-Antipolis
Parc Valrose 06108 Nice Cedex 2, France
E-mail: malot@unice.fr

January 24, 2013

Abstract: This paper proposes a model of interactions between two point processes, ruled by a reproduction function h, which is considered as the intensity of a Poisson process. In particular, we focus on the context of neurosciences to detect possible interactions in the cerebral activity associated with two neurons. To provide a mathematical answer to this specific problem of neurobiologists, we address so the question of testing the nullity of the intensity h. We construct a multiple testing procedure obtained by the aggregation of single tests based on a wavelet thresholding method. This test has good theoretical properties: it is possible to guarantee the level but also the power under some assumptions and its uniform separation rate over weak Besov bodies is adaptive minimax. Then, some simulations are provided, showing the good practical behavior of our testing procedure.

Keywords: Adaptive tests, interactions model, neurosciences, Poisson process, uniform separation rate, Unitary Events, wavelets, weak Besov bodies.

MSC2010: Primary 62G10; secondary 62G20, 62G30.

1 Introduction

In neurosciences, an important issue lies in a better understanding of the dynamics of cerebral activity in the cortex. In practice it is possible to measure, in vivo and for a specific task, the cerebral activity through the emission of action potentials by several neurons, and the specific interest of the neurobiologists is to understand how these action potentials appear. During a task, the recording of all arrival times of these action potentials (or spikes) on a neuron forms a spike train. From this point of view, the spike train can be modeled by a point process.

Several years ago it was thought that activities of different neurons during a task were independent (for example, see Barlow [3]); this explains why in the studies, the spike trains were usually modeled by independent Poisson processes. Today, thanks to technological advances in terms of recording brain activity, various studies show that this belief is false (for instance, see Gerstein [13] and Lestienne [24]). Thus the recent studies consider neuronal assemblies instead of the separate neuronal activities. For

example activities of pairs of neurons, that have been recorded simultaneously, show that there exists a phenomenon called synchronization (see Grammont and Riehle [14] and Grün *et al.* [16]): the presence of a spike on one of the two spike trains can affect the apparition of a spike, with a delay, on the second spike train. From a biological point of view, such a phenomenon reflects a reality. Indeed, an action potential appears if the neuron is sufficiently excited. However, to obtain a sufficient excitation, two strategies exist: either the frequency of spikes received by a single neuron increases, either the receiving neuron receives less spikes but at the same time from different neurons. This second strategy is precisely the synchronization. From a biological point of view, it is less energy consuming and the reaction is faster. Therefore, the neurobiologists are interested in detecting the synchronization phenomenon. More generally, they want to detect whether or not neurons evolve independently of each other, a dependence being a hint of a functional connection during a task.

To mathematically answer this question, we need a model taking into account the possible interactions between two neurons. In neurosciences, a possible model is the Hawkes process (for instance, see [18] for theoretical aspects and [5, 23, 27, 28] for its introduction in neurosciences). However, the Hawkes process is, theoretically speaking, a very complicated model, thus we consider a simpler version of Hawkes process which is realistic for the possible applications (in neurosciences, or in genomics, ...) and for which it is possible to carry out computations. Our model is the following one. Let N_p and N_c be two point processes with respective conditional intensity

$$\tilde{\lambda}_p : t \longmapsto \mu_p \quad \text{and} \quad \tilde{\lambda}_c : t \longmapsto \mu_c + \int_{-\infty}^{t} h(t-u)\, dN_p(u), \tag{1.1}$$

where μ_p and μ_c are positive parameters describing the spontaneous part (in the context of neurosciences, the spontaneous apparition of spikes) and h is a function which reflects the influence of N_p on N_c. In this model, we have to assume that $supp(h) \subset \mathbb{R}_+^*$, where $supp(h)$ is the support of the function h. Moreover, N_p is a homogeneous Poisson process (for instance, see [22]) and N_c is a special case of Hawkes process. The biological problem which consists in knowing whether N_p influences N_c is equivalent to test the null hypothesis $\mathcal{H}_0$: "$h = 0$" against the alternative $\mathcal{H}_1$: "$h \neq 0$".

The above formulation of $\tilde{\lambda}_c$ is an integral form. However it is possible conditionally on the points of N_p to have a vision in terms of descendants and no more in terms of intensity conditionally on the only past observations. Indeed, given T a positive real number representing the time of record of the neuronal activity and given n an integer, conditionally on the event "the number of points of N_p lying in $[0;T]$ is n", the points of the process N_p obey the same law as a n-sample of uniform random variables on $[0;T]$, denoted $U_1, \ldots, U_n$ and named parents. Thus, conditionally on $U_1, \ldots, U_n$, we can write $\tilde{\lambda}_c(t) = \mu_c + \sum_{i=1}^{n} h(t-U_i)$. This new expression of $\tilde{\lambda}_c$ can be interpreted as follows. Each U_i gives birth independently to a Poisson process N_c^i with intensity the function $t \longmapsto h(t-U_i)$ with respect to the Lebesgue measure on $\mathbb{R}$, to which is added a homogeneous Poisson process N_c^0 with constant intensity μ_c, representing the orphans. We consequently consider the aggregated process

$$N_c = \sum_{i=0}^{n} N_c^i \quad \text{whose intensity is given by the function} \quad t \longmapsto \mu_c + \sum_{i=1}^{n} h(t-U_i) \tag{1.2}$$

and the points of the process N_c are named children. With this interpretation, the goal of the present paper is to test the "influence or not" of the parents on the children, via the reproduction function h. This second writing contains many benefits. First, the assumption $supp(h) \subset \mathbb{R}_+^*$ is not mandatory. With respect to the first formulation, this may appear like a minor difference, but in practice the impact is considerable. Indeed, if we refer to the context of neurosciences, assuming that the support of h is in $\mathbb{R}_+$ means that one favors a sense of interactions, namely N_p affects N_c. However in practice, we do not have this information a priori. Therefore, when the test does not reject $\mathcal{H}_0$, it means that N_p does not seem to influence N_c, but this may be because in reality this is N_c that affects N_p. We must be careful that the initially proposed model is not symmetric in terms of neurons and that a support in $\mathbb{R}_+$ does not really allow to answer the question of dependence. The causality is indeed represented by the fact that a child appears after its parent and therefore h has to be supported in $\mathbb{R}_+$. Heuristically, a

consequence is the following interpretation: if a parent has a child before its own birth, it may represent that the child is the parent and the parent the real child. Looking at both sides of the support makes the procedure in some sense adaptive to the causality of parent/child roles. Another advantage of this second writing is that it allows applications to other disciplines such as genomics where one studies for instance the favored or avoided distances between patterns on a strand of DNA and where it is not always possible to know which pattern rules the other. More details about this application to genomics can be find in Sansonnet [34], where the author proposes an estimation procedure of the function h, assumed to be well localized, based on wavelet thesholding methods, in a very similar model to the one studied here. The interested reader will find other estimation procedures of the function h in this DNA context, by using a Hawkes' model in Gusto and Schbath [17] and Reynaud-Bouret and Schbath [30]. In this paper, we consider the model defined by (1.2). For the simulation study, parents process $(U_i)_i$ is simulated by a homogenous Poisson process of intensity μ_p.

Since the null hypothesis $\mathcal{H}_0$: "$h = 0$" means that conditionally on the total number of points of N_c, the points of the process N_c are i.i.d. (independent and identically distributed) with uniform distribution, a first rather naive approach is to perform a Kolmogorov-Smirnov test (see for instance [8]). But this test is not powerful, as illustrated in the section devoted to the simulations. The aim of this paper is then to build a more powerful and nonparametric test Φ_α with values in $\{0, 1\}$ of $\mathcal{H}_0$: "$h = 0$" against the alternative $\mathcal{H}_1$: "$h \neq 0$", rejecting $\mathcal{H}_0$ when $\Phi_\alpha = 1$, with prescribed probabilities of first and second kind errors. The performance of the test Φ_α is measured by its uniform separation rate (see for instance [1]).

In neurosciences, parametric methods exist to detect such dependence. For instance, the Unitary Event (UE) (see [16]) and the Multiple Tests based on a Gaussian Approximation of the UE (MTGAUE) (see [35]) methods answer partially the problem by considering coincidences (see Section 4 for more details). In the one-sample Poisson process model (that is to say $n = 1$ and $\mu_c = 0$ in our model), many papers deal with different problems of testing the simple hypothesis that an observed point process is a Poisson process with a known intensity. We can cite for instance the papers by Fazli and Kutoyants [10] where the alternative is also a Poisson process with a known intensity, Fazli [9] where the alternatives are Poisson processes with one-sided parametric intensities or Dachian and Kutoyants [7] where the alternatives are self-exciting point processes (namely, Hawkes processes). In the nonparametric framework, Ingster and Kutoyants [20] propose a goodness-of-fit test where the alternatives are Poisson processes with nonparametric intensities in a Sobolev or a Besov ball $\mathcal{B}^{\delta}_{2,q}(R)$ with $1 \leqslant q < \infty$ and known smoothness parameter δ. They establish its uniform separation rate over a Sobolev or a Besov ball and show the adaptivity of their testing procedure in a minimax sense.

In some practical cases like the study of the expression of neuronal interactions or the study of favored or avoided distances between patterns on a strand of DNA, such smooth alternatives (Sobolev or Besov balls) cannot be considered. Indeed, the intensity of the Poisson process N_c in these cases may burst at a particular position of special interest for the neuroscientist or the biologist. So we have to develop a testing procedure able to distinguish a constant function (or here a null function) from a function that has some small localized spikes. These features are not well captured by using classical Besov spaces. Hence we focus in particular on alternatives based on sparsity rather than on alternatives based on smoothness. For this, we are interested in the computation of uniform separation rates over weak versions of Besov balls. Such alternatives have already been considered. For instance, Fromont *et al.* [11] propose non-asymptotic and nonparametric tests of the homogeneity of a Poisson process that are adaptive over various Besov bodies simultaneously and in particular over weak Besov bodies. Another example is Fromont *et al.* [12] which construct non-asymptotic and nonparametric multiple tests of the equality of the intensities of two independent Poisson processes, that are adaptive in the minimax sense over a large variety of classes of alternatives based on classical and weak Besov bodies in particular.

The test Φ_α proposed in this paper consists in a multiple testing procedure obtained by aggregating several single tests based on a wavelet thresholding method as in Fromont *et al.* [11, 12] (they also consider model selection and kernel estimation methods). First, Proposition 2 proves that the multiple test is an α-level test and Theorem 2 gives a condition on the alternative to ensure that our multiple

test has a prescribed second kind error. This result reveals two regimes as in [34]. Indeed our model presents a double asymptotic through the number n of parents and the recording time T (namely, the length of the observations interval), which is not usual. Since N_p is a homogeneous Poisson process with constant intensity μ_p, the number n of points of N_p falling into $[0;T]$ is the realization of a Poisson random variable with parameter $\mu_p T$. As a consequence with very high probability, T is proportional to n and in this case, the uniform separation rates of the multiple test over weak Besov bodies are established by Theorem 3. Thus, our testing procedure is near adaptive in the minimax sense over a class of such alternatives. The proofs of these results are essentially based on concentration inequalities (see [26]) and on exponential inequalities for U-statistics (see [19]). Secondly, some simulations are carried out to validate our procedure, which is compared with the classical Kolmogrov-Smirnov test and a testing procedure proposed by Tuleau-Malot *et al.* [35], which formalized a well-known procedure in neurosciences, namely the UE method (see Grün *et al.* [16]).

The paper is organized as follows. Section 2 deals with the description of our testing procedure. Section 3 is devoted to the general results of the paper. The control of the probability of second kind error is ensured by Theorem 1 for the single testing procedures and by Theorem 2 for the multiple test. The uniform separation rates of the multiple test over weak Besov bodies are provided in Theorem 3. Section 4 presents the simulation study. The proofs of our main theoretical results are finally postponed in Section 6.

To end this section we introduce some notations that will be used along the paper. We denote by dN_p and dN_c the point measures associated with N_p and N_c respectively. We denote by $\mathbb{P}_0$ the distribution of the aggregated process N_c under $\mathcal{H}_0$, $\mathbb{P}_h$ the distribution of N_c whose intensity conditionally on $U_1,\dots,U_n$ is given by the function $t \longmapsto \mu_c + \sum_{i=1}^n h(t-U_i)$ for any alternative h and by $\mathbb{E}_h$ the corresponding expectation. The uniform distribution on $[0;T]$ is named π and $\mathbb{E}_\pi(f(U))$ denotes the expectation of $f(U)$ where $U \sim \pi$ (an independent copy of $U_1,\dots,U_n$) for any measurable function f. For an orthonormal basis $\{\varphi_\lambda, \lambda \in L\}$ of a finite dimensional subspace S_L of $\mathbb{L}_2(\mathbb{R})$, we denote by D_L the dimension of S_L (namely the cardinal of L) and by h_L the orthogonal projection of h onto S_L.

2 Description of our testing procedure

In the sequel, we assume that h is compactly supported (there is a maximal time of synchronization during a task according to the neuroscientists). Without loss of generality, we suppose now that the support of h is strictly included in $[-1;1]$ and that we observe the U_i's (the parents) on $[0;T]$ and realizations of the process N_c (the children) on $[-1;T+1]$. Consequently, h belongs to $\mathbb{L}_1(\mathbb{R})$, $\mathbb{L}_2(\mathbb{R})$ and $\mathbb{L}_\infty(\mathbb{R})$. Then, we can consider the decomposition of h on the Haar basis denoted by $\{\varphi_\lambda, \lambda \in \Lambda\}$:

$$h = \sum_{\lambda\in\Lambda} \beta_\lambda \varphi_\lambda \quad \text{with} \quad \beta_\lambda = \int_{\mathbb{R}} h(x)\varphi_\lambda(x)\,dx,$$

where

$$\Lambda = \{\lambda = (j,k) : j \geqslant -1, k \in \mathbb{Z}\}$$

and for any $\lambda \in \Lambda$ and any $x \in \mathbb{R}$,

$$\varphi_\lambda(x) = \begin{cases} \phi(x-k) & \text{if } \lambda = (-1,k) \\ 2^{j/2}\psi(2^j x - k) & \text{if } \lambda = (j,k) \text{ with } j \geqslant 0 \end{cases},$$

with

$$\phi = \mathbf{1}_{[0;1]} \quad \text{and} \quad \psi = \mathbf{1}_{]\frac{1}{2};1]} - \mathbf{1}_{[0;\frac{1}{2}]}.$$

The functions ϕ and ψ are respectively the father and the mother wavelets. Since the goal is to detect a signal and not to reconstruct it, the Haar basis is suitable in our context. Furthermore from a practical point of view, the use of the Haar basis yields fast algorithms, easy to implement. Nevertheless the theoretical results of the present paper can be generalized to a biorthogonal wavelet basis (see [6] for

a definition of this particular basis) as in [29, 31, 34]. We precise that we can easily extend our results to a function h compactly supported in $[-A;A]$ for any $A>0$ by scaling the data by $\lceil A\rceil+1$.

By considering this wavelet decomposition of h, the null hypothesis $\mathcal{H}_0$: "$h=0$" means that all the coefficients β_λ are null and the alternative hypothesis $\mathcal{H}_1$: "$h\neq 0$" means that there exists at least one non-zero coefficient. Since h is strictly supported in $[-1;1]$, if one coefficient $\beta_{(-1,k)}$ is non-zero, then there exists at least one coefficient $\beta_{(j,k)}$ with $j\geqslant 0$ which is also non-zero. Therefore, we focus only on the coefficients $\beta_{(j,k)}$ with $j\geqslant 0$ and we introduce the following subset Γ of Λ

$$\Gamma=\{\lambda=(j,k)\in\Lambda : j\geqslant 0, k\in\mathcal{K}_j\}$$

with $\mathcal{K}_j=\{k\in\mathbb{Z} : -2^j\leqslant k\leqslant 2^j-1\}$ ($\mathcal{K}_j$ is the set of integers k such that the intersection of the support of φ_λ and $[-1;1]$ is not empty, with $\lambda=(j,k)$).

For every λ in Γ, the coefficient β_λ is estimated by

$$\hat{\beta}_\lambda=\frac{\mathcal{G}(\varphi_\lambda)}{n},\quad\text{with}\quad\mathcal{G}(\varphi_\lambda)=\int_{\mathbb{R}}\sum_{i=1}^{n}\left[\varphi_\lambda(x-U_i)-\frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(x-U))\right]dN_c(x).$$

These estimates, inspired by those proposed in [34] for a simpler model, namely with $\mu_c=0$, are unbiased:

Proposition 1. *For all $\lambda=(j,k)$ in Γ, $\hat{\beta}_\lambda$ is an unbiased estimator of β_λ.*

The proof of Proposition 1 uses the fact that for all λ in Γ, $\int_{-1}^{1}\varphi_\lambda(t)\,dt=0$, avoiding boundary effects (see Section 6.1).

In order to test the null hypothesis $\mathcal{H}_0$: "$h=0$" against $\mathcal{H}_1$: "$h\neq 0$", namely "$\exists\lambda\in\Gamma,\beta_\lambda\neq 0$", we first propose to test for all $\lambda\in\Gamma$, the null hypothesis $\mathcal{H}_0$ against the alternative $\mathcal{H}_1^\lambda$: "$\beta_\lambda\neq 0$". For each $\lambda\in\Gamma$, the associated simple test actually consists in testing "$\beta_\lambda=0$" against "$\beta_\lambda\neq 0$" or more precisely, in testing the absence of variation of the function h on a small interval. Then in a second time, we will aggregate these simple tests to test the nullity of h on its complete support.

2.1 The single testing procedures

Let us fix some $\alpha\in]0;1[$ and $\lambda\in\Gamma$. We want to construct an α-level test of the null hypothesis $\mathcal{H}_0$: "$h=0$" against $\mathcal{H}_1^\lambda$: "$\beta_\lambda\neq 0$", from the observation of the parents $U_1,\dots,U_n$ and the realization of the Poisson process N_c. We notice first that the null hypothesis entails in particular that $\beta_\lambda=0$.

We introduce the testing statistic $\hat{T}_\lambda$ defined by

$$\hat{T}_\lambda=|\hat{\beta}_\lambda|. \tag{2.1}$$

Our single test consists in rejecting the null hypothesis when $\hat{T}_\lambda$ is too large and more precisely, when

$$\hat{T}_\lambda>q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha),$$

where $N_{c,tot}$ is the (random) number of points of the process N_c falling into $[-1;T+1]$ and for any $m\in\mathbb{N}^*$, $q_\lambda^{[U_1,\dots,U_n;m]}(\alpha)$ is the $(1-\alpha)$-quantile conditionally on $U_1,\dots,U_n$ of

$$\hat{T}^0_{\lambda,m}=\frac{1}{n}\left|\sum_{k=1}^{m}\sum_{i=1}^{n}\left[\varphi_\lambda(V_k^0-U_i)-\frac{n-1}{n}\mathbb{E}_\pi\big(\varphi_\lambda(V_k^0-U)\big)\right]\right|, \tag{2.2}$$

with $(V_1^0,\dots,V_m^0)$ a m-sample with uniform distribution on $[-1;T+1]$ (namely a m-sample of the process N_c under $\mathcal{H}_0$). We can easily prove that conditionally on $U_1,\dots,U_n$ and $N_{c,tot}=m$, $\hat{T}_\lambda$ and $\hat{T}^0_{\lambda,m}$ have exactly the same distribution under $\mathcal{H}_0$. Thus, the corresponding test function is defined by

$$\Phi_{\lambda,\alpha}=\mathbf{1}_{\hat{T}_\lambda>q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha)}. \tag{2.3}$$

2.2 The multiple testing procedure

Previously, we have built testing procedures based on each single empirical coefficient $\hat{\beta}_\lambda$. We propose in this subsection to consider a collection of empirical coefficients instead of a single one, and to define a multiple testing procedure by aggregating the corresponding single tests.

Let $\{w_\lambda, \lambda \in \Gamma\}$ be a collection of positive numbers such that $\sum_{\lambda\in\Gamma} e^{-w_\lambda} \leqslant 1$. This set allows us to put weights to empirical coefficients according to their index $\lambda = (j,k) \in \Gamma$. Given $\alpha \in]0;1[$, we consider the test which rejects $\mathcal{H}_0$ when there exists at least one λ in Γ such that

$$\hat{T}_\lambda > q_\lambda^{[U_1,\ldots,U_n;N_{c,tot}]}(u_\alpha^{[U_1,\ldots,U_n;N_{c,tot}]} e^{-w_\lambda}),$$

where

$$\begin{aligned} &u_\alpha^{[U_1,\ldots,U_n;N_{c,tot}]} \\ &= \sup\left\{u>0 : \mathbb{P}\left(\max_{\lambda\in\Gamma}\left(\hat{T}^0_{\lambda,N_{c,tot}} - q_\lambda^{[U_1,\ldots,U_n;N_{c,tot}]}(ue^{-w_\lambda})\right) > 0 \,\middle|\, U_1,\ldots,U_n;N_{c,tot}\right) \leqslant \alpha\right\}. \end{aligned} \tag{2.4}$$

The corresponding test function is defined by

$$\Phi_\alpha = \mathbf{1}_{\max_{\lambda\in\Gamma}\left(\hat{T}_\lambda - q_\lambda^{[U_1,\ldots,U_n;N_{c,tot}]}(u_\alpha^{[U_1,\ldots,U_n;N_{c,tot}]}e^{-w_\lambda})\right)>0}. \tag{2.5}$$

We mention that, since the set Γ is infinite countable, the number of tests to be performed is infinite and this is not a problem from a theoretical point of view. But in practice, we have to perform a finite number of single tests and so, we will fix a maximal resolution level j_0 and we will carry out the single tests $\Phi_{\lambda,\alpha}$ for $\lambda = (j,k)$ in Γ with $j \leqslant j_0$.

In the next section, we study the properties of the single tests $\Phi_{\lambda,\alpha}$ defined by (2.3) and the multiple test Φ_α defined by (2.5), through their probabilities of first and second kind errors.

3 Main theoretical results

3.1 Probability of first kind error

We constructed our single and multiple tests in such a way that the first kind error, which measures the probability that the test wrongly rejects the null hypothesis, is less than α.

Proposition 2. *Let α be a fixed level in $]0;1[$. Then the single test $\Phi_{\lambda,\alpha}$ defined by (2.3) for any $\lambda \in \Gamma$ and the multiple test Φ_α defined by (2.5) are of level α. Furthermore, $u_\alpha^{[U_1,\ldots,U_n;N_{c,tot}]}$ defined by (2.4) satisfies $u_\alpha^{[U_1,\ldots,U_n;N_{c,tot}]} \geqslant \alpha$.*

This result shows that the tests are exactly of level α, which is required for a test from a non-asymptotic point of view (namely n and T are not required to tend to infinity).

3.2 Probability of second kind error

The second kind error, which measures the probability that the test does not wrongly reject the null hypothesis is not fixed by the testing procedure, unlike the first kind error. We have to control the probability of second kind error in such a way that it is close to 0, in order to obtain powerful tests. The following theorem brings out a condition which guarantees that the single tests have a prescribed second kind error.

Since h belongs to $\mathbb{L}_1(\mathbb{R})$ and $\mathbb{L}_\infty(\mathbb{R})$, we introduce R_1 and R_∞ two positive real numbers such that $\|h\|_1 \leqslant R_1$ and $\|h\|_\infty \leqslant R_\infty$.

Theorem 1. *Let α, β be fixed levels in $]0;1[$. Let ζ and κ be positive constants depending on β, μ_c, R_1 and R_∞. For all $\lambda \in \Gamma$, let $\Phi_{\lambda,\alpha}$ be the test function defined by (2.3). Assume that*

$$\begin{aligned} |\beta_\lambda| > & \sqrt{\frac{2\zeta}{\beta}\left(\frac{1}{n}+\frac{1}{T}+\frac{2^{-j}n}{T^2}\right)} \\ & + \kappa\left\{\sqrt{\ln(2/\alpha)}\left(\frac{1}{\sqrt{n}}+\frac{1}{\sqrt{T}}+\frac{2^{-j/2}\sqrt{n}}{T}\right)+\ln(2/\alpha)\left(\frac{2^{j/2}}{n^{3/2}}+\frac{2^{-j/2}}{nT}\right)\right\}, \end{aligned} \tag{3.1}$$

for $\lambda=(j,k)$. Then,

$$\mathbb{P}_h(\Phi_{\lambda,\alpha}=0) \leqslant \beta.$$

Note that the quantity $\frac{1}{n}+\frac{1}{T}+\frac{2^{-j}n}{T^2}$ that appears under the square root of the first term of the right hand side of (3.1) is of the same order as the upper bound of the variance of the estimates $\hat{\beta}_\lambda$ (see Proposition 1 of [34]). Consequently, the right hand side of (3.1) can be viewed as a standard deviation term, since the other terms are not asymptotically larger than the first term if we assume that $2^j \leqslant n^2$, where asymptotic means $n \to +\infty$ or $T \to +\infty$.

Theorem 1 means that if the coefficient β_λ is far enough from 0, then the probability of second kind error is controlled. This result gives a threshold for β_λ from which our associated single testing procedure is able to detect a signal and shows that its power is larger than $1-\beta$. Furthermore, if we consider the regime "T proportional to n" in order to compare our result with known asymptotic rates of testing, Condition (3.1) can be easily obtained for instance if $\beta_\lambda^2 > C/n$ by assuming that $2^j \leqslant n^2$, with C a positive constant.

Now we are interested in the power of the multiple testing procedure and the following theorem gives a condition on the alternative in order to ensure that our multiple test has a prescribed second kind error.

Theorem 2. *Let α, β be fixed levels in $]0;1[$. Let Φ_α be the test function defined by (2.5). Assume that there exists at least one finite subset L of Γ such that*

$$\|h_L\|_2^2 > \left(C_1 D_L + C_2 \sum_{\lambda\in L} w_\lambda\right)\left[\frac{1}{n}+\frac{n}{T^2}\right] + \left(C_3 D_L + C_4 \sum_{\lambda\in L} w_\lambda + C_5 \sum_{\lambda\in L} w_\lambda^2\right)\left[\frac{2^{j_L}}{n^3}+\frac{1}{n^2T^2}\right], \tag{3.2}$$

where $j_L = \max\{j \geqslant 0 : (j,k) \in L \text{ with } k \in \mathcal{K}_j\}$ and C_1, C_2, C_3, C_4 and C_5 are positive constants depending on α, β, μ_c, R_1 and R_∞. Then,

$$\mathbb{P}_h(\Phi_\alpha=0) \leqslant \beta.$$

This theorem means that if there exists one subspace S_L of $\mathbb{L}_2(\mathbb{R})$ such that h_L (the orthogonal projection of h onto S_L) lies outside a small ball around 0, then the probability of second kind error is controlled. This result gives a threshold for the energy of h_L from which our multiple testing procedure is able to detect a signal and shows that its power is larger than $1-\beta$. Furthermore, if we consider the regime "T proportional to n" in order to compare our result with known asymptotic rates of testing, Condition (3.2) can be easily obtained for instance if $\|h_L\|_2^2 > C \times \left(D_L + \sum_{\lambda\in L} w_\lambda + \sum_{\lambda\in L} w_\lambda^2\right)/n$ by assuming that $2^{j_L} \leqslant n^2$, with C a positive constant. Then, the separation rate between the null and the alternative hypotheses is of order D_L/n and this is typical for testing procedures based on a thresholding approach (see [11, 12] for instance). Usually, nested tests (namely based on model selection) achieve a faster rate of separation of order $\sqrt{D_L}/n$ (see [1, 2] for instance). But these latter tests are not adaptive over weak Besov bodies. Consequently, the separation rate established by Theorem 2 leads to sharp upper bounds for the uniform separation rates over such particular classes of alternatives and so, our multiple testing procedure will be proved to be adaptive over particular classes of alternatives, based on weak Besov bodies.

3.3 Uniform separation rates

Given some $\alpha, \beta \in]0;1[$, we have previously built an α-level test Φ_α defined by (2.5), with a probability of second kind error at most equals to β if Condition (3.2) is satisfied. Then, given a class $\mathcal{S}_\delta$ of alternatives h, it is natural to measure the performance of the test via its uniform separation rate $\rho(\Phi_\alpha, \mathcal{S}_\delta, \beta)$ over $\mathcal{S}_\delta$ (see [1]) defined by

$$\rho(\Phi_\alpha, \mathcal{S}_\delta, \beta) = \inf\left\{\rho > 0 : \sup_{h \in \mathcal{S}_\delta, \|h\|_2 > \rho} \mathbb{P}_h(\Phi_\alpha = 0) \leqslant \beta\right\}. \tag{3.3}$$

In order to compare our result with known asymptotic rates of testing, we consider the regime "T proportional to n" in this subsection.

We introduce for $\delta > 0$, $R > 0$ the Besov body

$$\mathcal{B}^\delta_{2,\infty}(R) = \left\{f \in \mathbb{L}_2(\mathbb{R}) : f = \sum_{\lambda \in \Lambda} \beta_\lambda \varphi_\lambda, \quad \forall j \geqslant 0, \sum_{k \in \mathcal{K}_j} \beta^2_{(j,k)} \leqslant R^2 2^{-2j\delta}\right\}.$$

We also consider a weaker version of the above Besov bodies defined for $p > 0$, $R' > 0$ by

$$\mathcal{W}^*_p(R') = \left\{f \in \mathbb{L}_2(\mathbb{R}) : f = \sum_{\lambda \in \Lambda} \beta_\lambda \varphi_\lambda, \quad \sup_{s>0} s^p \sum_{\lambda \in \Gamma} \mathbf{1}_{|\beta_\lambda| > s} \leqslant R'^p\right\}.$$

Whereas the spaces $\mathcal{B}^\delta_{2,\infty}(R)$ constitute an ideal class to measure the regularity of the possible alternatives h, the spaces $\mathcal{W}^*_p(R')$ constitute an ideal class to measure the sparsity of a wavelet decomposed signal h. Indeed, if $f = \sum_{\lambda \in \Lambda} \beta_\lambda \varphi_\lambda \in \mathcal{W}^*_p(R')$, then the associated sequence $\beta = (\beta_\lambda)_{\lambda \in \Gamma}$ satisfies $\sup_{\ell \in \mathbb{N}^*} \ell^{1/p} |\beta|_{(\ell)} < \infty$, where the sequence $(|\beta|_{(\ell)})_\ell$ is the non-increasing rearrangement of β: $|\beta|_{(1)} \geqslant |\beta|_{(2)} \geqslant \ldots \geqslant |\beta|_{(\ell)} \geqslant \ldots$. This condition gives a polynomial control of the decreasing rate of the sequence $(|\beta|_{(\ell)})_\ell$ and smaller p, sparser the signal. There exists an embedding between Besov and weak Besov balls:

$$\mathcal{B}^\delta_{2,\infty}(R) \subset \mathcal{W}^*_{\frac{2}{1+2\delta}}(r),$$

where the radius r of the weak Besov ball depends on δ and R (more precisely, $r = 4^\delta R/\sqrt{2^{2\delta} - 1}$). See [21, 32, 33] for more details and for extensions in a more general setting. So, we consider in this paper such alternatives based on the intersection of Besov and weak Besov bodies, namely sparse functions with some minimal regularity, see below.

To evaluate the uniform separation rates, we choose the following collection of weights $\{w_\lambda, \lambda \in \Gamma\}$ defined by

$$w_\lambda = 2\big(\ln(j+1) + \ln(\pi/\sqrt{6})\big) + \ln|\mathcal{K}_j|, \tag{3.4}$$

for any $\lambda = (j,k) \in \Gamma$, where $|\mathcal{K}_j|$ is the cardinal of $\mathcal{K}_j$ which is of order 2^j. With this choice, the collection of weights satisfies the condition $\sum_{\lambda \in \Gamma} e^{-w_\lambda} \leqslant 1$. The following theorem gives the uniform separation rates over $\mathcal{B}^\delta_{2,\infty}(R) \cap \mathcal{W}^*_{\frac{2}{1+2\gamma}}(R')$, for $\gamma \geqslant \delta$.

Theorem 3. *Let α, β be fixed levels in $]0;1[$. Assume that T is proportional to n. Let Φ_α be the test function defined by (2.5) with the weights w_λ's defined by (3.4). Then, for any $\delta > 0$, $\gamma > 0$, $R > 0$, $R' > 0$, if $2\delta > \gamma/(1+2\gamma)$*

$$\rho(\Phi_\alpha, \mathcal{B}^\delta_{2,\infty}(R) \cap \mathcal{W}^*_{\frac{2}{1+2\gamma}}(R'), \beta) \leqslant C\left(\frac{\ln n}{n}\right)^{\frac{\gamma}{1+2\gamma}},$$

with C a positive constant depending on δ, γ, R, R', α, β, μ_c, R_1 and R_∞.

Note that this result holds for instance with $\delta = 1/4$. It corresponds to the minimal regularity mentioned previously.

Theorem 3 illustrates the optimality of our testing procedure in the minimax setting. Indeed, considering the regime "T proportional to n", uniform separation rates of the test Φ_α match the minimax separation rates obtained by Fromont *et al.* [11, 12], if $2\delta > \gamma/(1+2\gamma)$ and also $\delta < \gamma/2$ and $\gamma > 1/2$ (see Theorem 1 of [11]). Furthermore, the upper bound of uniform separation rates of our test Φ_α over $\mathcal{B}^{\delta}_{2,\infty}(R) \cap \mathcal{W}^*_{\frac{2}{1+2\gamma}}(R')$ has already obtained, up to a logarithmic term, for a wavelet thresholding estimation method proposed by Sansonnet [34] in a very similar context and more precisely, this is equal to the minimax estimation rates of the maxisets of the thresholding estimation procedure (see [21, 29, 33] for more details). This means that it is at least as difficult to test as to estimate over such classes of alternatives. Note that on Sobolev or Besov spaces, testing rates are usually faster than estimation rates.

4 Simulation study

In this section, we study our testing procedure from a practical point of view and we compare it to the conditional Kolmogorov-Smirnov (KS) test and a Gaussian Approximation of the Unitary Events (GAUE) method developed by Tuleau-Malot *et al.*[35].

4.1 Description of the data

We create different data sets that are to a certain extent a reflection of a neurobiological reality. We consider the spike trains of two neurons N_p and N_c which are modeled by two point processes with respective conditional intensity $\tilde{\lambda}_p$ and $\tilde{\lambda}_c$ defined by (1.1).

For real spike trains it is not reasonable to postulate the stationarity of N_p and N_c, i.e. μ_p and μ_c are constant and considering the same function h on the entire recording period $[0;T]$ (see Grün *et al.* [16]). But this assumption is quite feasible on smaller time ranges (see Grün [15] and Grammont and Riehle [14]). However, to date, we have no algorithmic and statistical tool to clearly identify the stationarity ranges. Several methods (UE and MTGAUE, see [35] for instance) propose to perform many tests on different small windows of time and to use multiple testing procedure (see Benjamini and Hochberg [4] for instance) to combine them. Hence those methods can solve, at least in practice, this stationarity problem. The aim of this simulation study is not to show how our testing procedure can be incorporated in a Benjamini and Hochberg's approach, which lies outside the scope of the present paper, but to discuss the advantage of our method on one small window of time. This explains the use of the simulated data described below.

We need therefore to simulate dependence between N_p and N_c on $[0;T]$ and to take into account the major part of the neurobiological reality. So, we simulate processes N_p and N_c whose intensities are respectively given by

$$\tilde{\lambda}_p = 50 \quad \text{and} \quad \tilde{\lambda}_c = 20 + \int_{-\infty}^{t} h(t-u)\, dN_p(u), \text{ with } h = \theta \mathbf{1}_{[\nu;0.01]}.$$

In order to evaluate the performance of different procedures, several parameters θ and ν are tested. The parameter θ represents the influence strength of N_p on N_c: larger the parameter θ and more important the influence of N_p on N_c. The parameter ν introduces a possible minimal delay in the synchronization, i.e. the synchronization of the neuronal activity occurs with a delay δ uniform on $[\nu; 0.01]$. We consider nine different data sets denoted $Data_0$, $Data_{10}$, $Data_{30}$, $Data_{50}$, $Data_{80}$, $Data_{10r}$, $Data_{30r}$, $Data_{50r}$ and $Data_{80r}$. For $k \in \{0, 10, 30, 50, 80\}$, $Data_k$ is simulated with $\theta = k$ and $\nu = 0$ while $Data_{kr}$ is simulated with $\theta = k$ and $\nu = 0.005$.

4.2 The Kolmogorov-Smirnov test

First, we look at the performance of the Kolmogorov-Smirnov test (see Darling [8]) to convince us that this commonly used test is not reliable in this context. Indeed, even if the KS test is not a test of

Data set	Power for KS
$Data_{10}$	0.040
$Data_{30}$	0.051
$Data_{50}$	0.087
$Data_{80}$	0.113
$Data_{10r}$	0.054
$Data_{30r}$	0.059
$Data_{50r}$	0.053
$Data_{80r}$	0.073

Table 1: Power of the KS test with level $\alpha = 0.05$, evaluated for various interactions.

independence, the KS test may answer the problem. Since as said before, under $\mathcal{H}_0$ and conditionally on $U_1, \dots, U_n$ and $N_{c,tot} = m$, the observations of N_c are i.i.d. with common law the uniform distribution on $[-1; T+1]$, looking for the adequation of N_c with this law could be an idea to detect the rejection of $\mathcal{H}_0$. So, the use of the KS test is relevant.

First, we focus on the empirical rate of the type I error which is an approximation of the level of the test. Thus, we simulate 5000 independent realizations of $Data_0$, simulations on which we perform the KS test with level $\alpha = 0.05$. The empirical rate of the type I error evaluated on those data is 0.051, which is as desired.

What about the number of wrong rejections of $\mathcal{H}_1$? We consider the power of the tests which is the proportion of correct rejections of $\mathcal{H}_0$. We simulate 1000 independent realizations of $Data_{10}$, $Data_{30}$, $Data_{50}$, $Data_{80}$, $Data_{10r}$, $Data_{30r}$, $Data_{50r}$ and $Data_{80r}$, data on which we perform the KS test with level $\alpha = 0.05$ and we evaluate the empirical power of the test. Table 1 summarizes the obtained results. The KS test power is extremely close to the expected level. Hence the KS test is not able to clearly detect the dependence.

4.3 The GAUE method adapted to our context

Before comparing both methods, we briefly return to the principle of the GAUE method. The aim of the GAUE method is to detect the dependence on a single window $[0; T]$. This method is based on the coincidences with delay. More precisely for the couple of processes (N_p, N_c), we compute the number of coincidences with delay δ on $[0; T]$, i.e. the variable $X_T = \int_{[0;T]^2} \mathbf{1}_{|x-y| \leqslant \delta} \, dN_p(x) \, dN_c(y)$, that represents the number of pairs (x, y) in $N_p \times N_c$ such that $|x - y| \leqslant \delta$. Let us define $\hat{\lambda}_p = N_p([0; T])/T$ and $\hat{\lambda}_c = N_c([0; T])/T$ where $N_p([0; T])$ and $N_c([0; T])$ denote respectively the number of spikes of N_p and N_c among $[0; T]$. The quantities $\hat{\lambda}_p$ and $\hat{\lambda}_c$ are estimators of $\tilde{\lambda}_p$ and $\tilde{\lambda}_c$.

We reject the null hypothesis $\mathcal{H}_0$: "$h = 0$" when $X_T \geqslant \hat{m}_0 + \hat{\sigma} u_{1-\alpha/2}$, where $\hat{m}_0 = \hat{\lambda}_p \hat{\lambda}_c (2T\delta - \delta^2)$, $\hat{\sigma}^2 = \hat{\lambda}_p \hat{\lambda}_c (2T\delta - \delta^2) + \hat{\lambda}_p \hat{\lambda}_c \left(\hat{\lambda}_p + \hat{\lambda}_c \right) \left(\frac{2}{3}\delta^3 - \frac{1}{T}\delta^4 \right)$ and $u_{1-\alpha/2}$ is the $(1 - \alpha/2)$-quantile of a standard normal. This threshold comes from the theory developed in [35] and is adapted to our context. The quantity $\hat{m}_0$ is a plug-in estimator of the expectation of X_T under $\mathcal{H}_0$ and $\hat{\sigma}^2$ is an estimator of the variance. It can be shown that under the assumptions "N_p and N_c are Poisson processes" and "N_p and N_c are stationary", this test is asymptotically of level α. Further details about the meaning of those different estimators are given in [35].

The GAUE method was developed jointly with a neurophysiologist and it fits in line the UE method developed by Grün and coauthors (for instance, see [15] and [16]), which is a commonly used method in neurosciences. One of its main disadvantage is that δ has to be chosen beforehand. Part of the aim of this work is to propose a more adaptive method.

Data set	our procedure	GAUE
$Data_0$	0.0486	0.0446/0.0510/0.0548

Table 2: Empirical rate of type I error associated to our procedure and the GAUE method. The theoretical level is $\alpha = 0.05$. Since the GAUE method depends on the tuning parameter δ, the given value is the minimum/median/maximum of the empirical rate over all the δ.

4.4 Our procedure in practice

From a theoretical point of view, considering all the resolution levels is not a problem. However, in practice, we have a maximal resolution level, denoted j_0 in the sequel. We choose j_0 quite small for time computational reasons ($j_0 = 3$). Thus, it is better to have the support $[-A; A]$ of h with A close to 1. Nevertheless, if in addition to a global detection, we are interested in a more local detection, i.e. the coefficients $\lambda \in \Gamma$ for which $\Phi_{\alpha,\lambda}$ rejected, j_0 should not be too small. For instance, if $h = \mathbf{1}_{[0;A]}$ and if the order of magnitude of A is 2^{-J} or $1 - 2^{-J}$, with $J > j_0 + 1$, our procedure does not allow to detect locally the jump of h at A. Consequently, taking A close to $1/2$ may appear reasonable. Hence, the data are multiplied by 50 before being treated.

Let us recall that our test rejects $\mathcal{H}_0$ when there exists at least one $\lambda = (j, k)$ in Γ with $j \leqslant j_0$ such that

$$\hat{T}_\lambda > q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(u_\alpha^{[U_1,\dots,U_n;N_{c,tot}]} e^{-w_\lambda}),$$

where $j_0 \geqslant 1$ denotes the maximal resolution level, $u_\alpha^{[U_1,\dots,U_n;N_{c,tot}]}$ is defined by (2.4) and the w_λ's are given by (3.4). Hence, for each observation of the process N_c whose number of points is denoted by $N_{c,tot} = m$, given the points of N_p denoted $U_1, \dots, U_n$, we estimate $u_\alpha^{[U_1,\dots,U_n;m]}$ and the quantiles $q_\lambda^{[U_1,\dots,U_n;m]}$ by classical Monte Carlo methods based on the simulations of B independent sequences $\{V^b, 1 \leqslant b \leqslant B\}$, where $V^b = (V_1^b, \dots, V_m^b)$ is a m-sample of uniform variables on $[-1; T + 1]$ (i.e. the law of N_c under $\mathcal{H}_0$, conditionally on $U_1, \dots, U_n$ and $N_{c,tot} = m$). We fix $B = 20000$ in the sequel since for larger values of B, the gain in precision for the estimates of $u_\alpha^{[U_1,\dots,U_n;m]}$ and $q_\lambda^{[U_1,\dots,U_n;m]}$ becomes negligible. We define for any $\lambda = (j, k)$ in Γ with $j \leqslant j_0$, for $1 \leqslant b \leqslant B$:

$$\hat{T}_{\lambda,m}^{0,b} = \frac{1}{n} \left| \sum_{k=1}^{m} \sum_{i=1}^{n} \left[\varphi_\lambda(V_k^b - U_i) - \frac{n-1}{n} \mathbb{E}_\pi\big(\varphi_\lambda(V_k^b - U)\big) \right] \right|.$$

We compute these $\hat{T}_{\lambda,m}^{0,b}$'s with a cascade algorithm (see Mallat [25]).

Half of the m-samples is used to estimate the quantiles by putting in ascending order the $\hat{T}_{\lambda,m}^{0,b}$'s for any λ. The other half is used to approximate the conditional probabilities occurring in (2.4). Then, $u_\alpha^{[U_1,\dots,U_n;m]}$ is obtained by dichotomy, such that the estimated conditional probability occurring in (2.4) is less than α, but as close as possible to α. Choosing $j_0 = 3$, our procedure considers 15 single tests $\Phi_{\lambda,\alpha}$ involving wavelets whose support length is respectively 0.125, 0.25, 0.5 and 1. This allows us to make detections at the positions $m \times 2^{-3}$ (m in $\{0, \dots, 7\}$) with a range of 2^{-3}. Due to the scaling of the data in our procedure, we need to divide the positions and the range of the possible detections by 50. Consequently, in the real time, the positions and the range become $m \times 0.0025$ (m in $\{0, \dots, 7\}$) and 0.0025.

4.5 Results

We compare our testing procedure and the GAUE method on the different data sets. As for the KS test, we look first at the level of both tests. We simulate 5000 independent realizations of $Data_0$, simulations on which we perform the present method and the GAUE ones with $\alpha = 0.05$. For the GAUE, the tuning parameter δ varies on a regular grid of $[0.001; 0.04]$ of step 0.001. The order of magnitude of δ is similar to the range of the possible detections done with our method. On those data,

Data set	our procedure	GAUE
$Data_{10}$	0.095	0.068/0.1085/0.168
$Data_{30}$	0.478	0.154/0.3795/0.707
$Data_{50}$	0.864	0.278/0.6645/0.953
$Data_{80}$	0.993	0.451/0.9160/0.998
$Data_{10r}$	0.073	0.047/0.0575/0.077
$Data_{30r}$	0.282	0.050/0.1415/0.277
$Data_{50r}$	0.664	0.053/0.2825/0.589
$Data_{80r}$	0.968	0.048/0.4900/0.879

Table 3: Powers associated with our procedure and the GAUE method, evaluated for various interactions. The theoretical level is $\alpha = 0.05$. Since the GAUE method depends on the tuning parameter δ, the given value is the minimum/median/maximum of the empirical rate over all the δ.

we evaluate the empirical rate of type I error. Those results, for both methods, are summarized in Table 2: both testing methods seem to have a correct level in practice. This means that the number of wrong rejections of $\mathcal{H}_0$ is well controlled.

Now we want to see if the number of wrong rejections of $\mathcal{H}_1$ is also controlled. To evaluate the power of both tests, we simulate 1000 independent realizations of $Data_{10}$, $Data_{30}$, $Data_{50}$, $Data_{80}$, $Data_{10r}$, $Data_{30r}$, $Data_{50r}$ and $Data_{80r}$ as for the KS test (see Section 4.2). The results of the empirical power are given by Table 3: both methods are comparable in terms of power when $\nu = 0$ (i.e. for the $Data_k$). However when $\nu \neq 0$, our method seems to have better performance since the power is higher.

Moreover, if both methods are comparable in terms of performance, it remains that the testing procedure proposed in this paper has an advantage over the GAUE method. In fact, our method is statistically adaptive. Indeed, the parameter δ which appears in the GAUE method is not calibrated in practice. In our method, we aggregate the single tests over (j, k). So on one hand, we do not need to specify this parameter but just an upper bound j_0, the maximal resolution level: the method through weights (3.4), adapts to this unspecified parameter (j, k). But on the other hand, by looking at the single tests $\Phi_{\lambda,\alpha}$ that have supported the rejection, we are able to partially recover an important information for the practitioner: the position $(k2^{-j})$ and the range (2^{-j}) of the influence. In fact, by looking only at this single testing procedure, we get an upper value for 0.01 and a lower value for ν on the range of delay δ of synchronization. To obtain more precise estimations of the support of h, we can consider an estimate of h, for instance the one proposed by Sansonnet [34]. The capacity to our method to get an information on ν is due to the fact that for a resolution level j we consider different positions k. This is not possible with the GAUE method. This explains why the results on $Data_{10r}$, $Data_{30r}$, $Data_{50r}$ and $Data_{80r}$ are better with our method.

5 Conclusion

In our paper, we have investigated the influence of a point process on another one. We have built a multiple testing procedure based on wavelet thresholding. The main results of the paper have revealed the optimality of the procedure. Furthermore, our test is adaptive in the minimax sense over classes of alternatives essentially based on weak Besov bodies. Then, from a practical point of view, our method answers several practical questions. However, a number of challenges remain before applying our method on real data. To overcome the problem of stationarity, we could use a Benjamini and Hochberg's approach as for the GAUE method. Finally, we could consider a more sophisticated model that takes into account the phenomenons of spontaneous apparition and self-excitation (as for the complete Hawkes model). But this model raises serious difficulties from the theoretical point of view. This is an exciting challenge.

6 Proofs

All along the proofs, we introduce some positive constants denoted by $C(\xi, \ldots)$ meaning that they may depend on ξ, ... They do not depend on j, n and T (which drive the asymptotic). Furthermore, the values of these constants may vary from line to line.

We recall that $\{\varphi_\lambda, \lambda \in \Lambda\}$ is the Haar basis and consequently, we have:

$$\|\varphi_\lambda\|_1 = 2^{-j/2}, \quad \|\varphi_\lambda\|_2 = 1 \quad \text{and} \quad \|\varphi_\lambda\|_\infty = 2^{j/2}.$$

In the case of a biorthogonal wavelet basis, $\|\varphi_\lambda\|_1$, $\|\varphi_\lambda\|_2$ and $\|\varphi_\lambda\|_\infty$ are of the same order as above, up to a positive constant respectively depending on $\|\psi\|_1$, $\|\psi\|_2$ and $\|\psi\|_\infty$, where ψ is the mother wavelet associated to the considered biorthogonal wavelet basis. Consequently, the same proofs potentially lead to the results on a biorthogonal wavelet basis as well as in [34] for the wavelet thresholding estimation.

6.1 Proof of Proposition 1

We first notice that for any λ in Γ, for any $u \in [0; T]$,

$$\int_{-1}^{T+1} \varphi_\lambda(t-u)\, dt = 0. \tag{6.1}$$

Let $\lambda \in \Gamma$ be fixed. By considering the aggregated process (1.2), we can write

$$\mathcal{G}(\varphi_\lambda) = G^0(\varphi_\lambda) + G(\varphi_\lambda), \tag{6.2}$$

with

$$G^0(\varphi_\lambda) = \int_{\mathbb{R}} \sum_{i=1}^{n} \left[\varphi_\lambda(x - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(x-U)) \right] dN_c^0(x)$$

and

$$G(\varphi_\lambda) = \int_{\mathbb{R}} \sum_{i=1}^{n} \left[\varphi_\lambda(x - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(x-U)) \right] \sum_{j=1}^{n} dN_c^j(x).$$

On the one hand, we notice that $G(\varphi_\lambda)$ is the same quantity as the one defined by equation (2.3) of [34]. Thus, by applying the first part of Proposition 1 of [34], we obtain

$$\mathbb{E}(G(\varphi_\lambda)) = n \int_{\mathbb{R}} \varphi_\lambda(x) h(x)\, dx.$$

On the other hand, we have

$$G^0(\varphi_\lambda) = \int_{\mathbb{R}} \varphi_\lambda(x - U_1)\, dN_c^0(x) + \sum_{i=2}^{n} \int_{\mathbb{R}} \left[\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x-U)) \right] dN_c^0(x).$$

Thus,

$$\mathbb{E}(G^0(\varphi_\lambda) | U_1, \ldots, U_n) = \int_{-1}^{T+1} \varphi_\lambda(x - U_1) \mu_c\, dx + \sum_{i=2}^{n} \int_{-1}^{T+1} \left[\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x-U)) \right] \mu_c\, dx$$

and by using (6.1), we obtain

$$\mathbb{E}(G^0(\varphi_\lambda)) = \sum_{i=2}^{n} \int_{-1}^{T+1} \mathbb{E}\big[\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x-U)) \big] \mu_c\, dx = 0.$$

Finally,

$$\mathbb{E}(\hat{\beta}_\lambda) = \mathbb{E}\left(\frac{\mathcal{G}(\varphi_\lambda)}{n} \right) = \int_{\mathbb{R}} \varphi_\lambda(x) h(x)\, dx = \beta_\lambda,$$

which proves Proposition 1.

6.2 Proof of Proposition 2

Let α be a fixed level in $]0;1[$. Let $\lambda \in \Gamma$ be fixed. First, the probability that the single test defined by (2.3) wrongly detects a signal is

$$\mathbb{P}_0(\Phi_{\lambda,\alpha} = 1) = \mathbb{P}_0\left(\hat{T}_\lambda > q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha)\right).$$

Since conditionally on $U_1,\dots,U_n$ and $N_{c,tot}$, $\hat{T}_\lambda$ and $\hat{T}^0_{\lambda,N_{c,tot}}$ have exactly the same distribution under $\mathcal{H}_0$, $q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha)$ is also the $(1-\alpha)$-quantile of $\hat{T}_\lambda \big| U_1,\dots,U_n; N_{c,tot}$ under $\mathcal{H}_0$. Thus,

$$\mathbb{P}_0(\Phi_{\lambda,\alpha} = 1) \leqslant \alpha$$

and the level of the single test is α.

Then, the probability that the multiple test defined by (2.5) wrongly detects is

$$\mathbb{P}_0(\Phi_\alpha = 1) = \mathbb{P}_0\left(\max_{\lambda\in\Gamma}\left(\hat{T}_\lambda - q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(u_\alpha^{[U_1,\dots,U_n;N_{c,tot}]}e^{-w_\lambda})\right) > 0\right).$$

By definition (2.4) of $u_\alpha^{[U_1,\dots,U_n;N_{c,tot}]}$,

$$\mathbb{P}_0\left(\max_{\lambda\in\Gamma}\left(\hat{T}_\lambda - q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(u_\alpha^{[U_1,\dots,U_n;N_{c,tot}]}e^{-w_\lambda})\right) > 0 \Big| U_1,\dots,U_n; N_{c,tot}\right) \leqslant \alpha,$$

because conditionally on $U_1,\dots,U_n$ and $N_{c,tot}$, $\hat{T}_\lambda$ and $\hat{T}^0_{\lambda,N_{c,tot}}$ have exactly the same distribution under $\mathcal{H}_0$. By taking the expectation over $U_1,\dots,U_n$ and $N_{c,tot}$, we obtain that

$$\mathbb{P}_0(\Phi_\alpha = 1) \leqslant \alpha$$

and the level of the multiple test is α.

Furthermore, by Bonferroni's inequality we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{\lambda\in\Gamma}\left(\hat{T}^0_{\lambda,N_{c,tot}} - q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha e^{-w_\lambda})\right) > 0 \Big| U_1,\dots,U_n; N_{c,tot}\right)\\
&\leqslant \sum_{\lambda\in\Gamma}\mathbb{P}\left(\hat{T}^0_{\lambda,N_{c,tot}} - q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha e^{-w_\lambda}) > 0 \Big| U_1,\dots,U_n; N_{c,tot}\right)\\
&\leqslant \sum_{\lambda\in\Gamma}\alpha e^{-w_\lambda}\\
&\leqslant \alpha
\end{aligned}$$

and consequently $u_\alpha^{[U_1,\dots,U_n;N_{c,tot}]} \geqslant \alpha$ by definition (2.4) of $u_\alpha^{[U_1,\dots,U_n;N_{c,tot}]}$, which concludes the proof of Proposition 2.

6.3 Proof of Theorem 1

Let $\lambda \in \Gamma$ be fixed. Here we want to find a condition which will guarantee that

$$\mathbb{P}_h(\Phi_{\lambda,\alpha} = 0) \leqslant \beta,$$

given $\beta \in]0;1[$.

Let us introduce $q^\alpha_{1-\beta/2}$ the $(1-\beta/2)$-quantile of the conditional quantile $q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha)$. Then for any h,

$$\begin{aligned}
\mathbb{P}_h(\Phi_{\lambda,\alpha} = 0) = &\mathbb{P}_h\left(\hat{T}_\lambda \leqslant q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha)\,,\, q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha) \leqslant q^\alpha_{1-\beta/2}\right)\\
&+ \mathbb{P}_h\left(\hat{T}_\lambda \leqslant q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha)\,,\, q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha) > q^\alpha_{1-\beta/2}\right)
\end{aligned}$$

$$\leqslant \mathbb{P}_h(\hat{T}_\lambda \leqslant q^\alpha_{1-\beta/2}) + \beta/2$$

and a condition which guarantees $\mathbb{P}_h(\hat{T}_\lambda \leqslant q^\alpha_{1-\beta/2}) \leqslant \beta/2$ will be enough to ensure that

$$\mathbb{P}_h(\Phi_{\lambda,\alpha} = 0) \leqslant \beta.$$

The following lemma gives such a condition.

Lemma 6.1. *Let α, β be fixed levels in $]0;1[$. For any $\lambda = (j,k) \in \Gamma$, if*

$$\mathbb{E}_h(\hat{T}_\lambda) > \sqrt{\frac{2\zeta Q_{j,n,T}}{\beta}} + q^\alpha_{1-\beta/2} \tag{6.3}$$

for a particular ζ which is a positive constant depending on μ_c, R_1 and R_∞, where

$$Q_{j,n,T} = \frac{1}{n} + \frac{1}{T} + \frac{2^{-j}n}{T^2},$$

then

$$\mathbb{P}_h(\hat{T}_\lambda \leqslant q^\alpha_{1-\beta/2}) \leqslant \beta/2,$$

so that

$$\mathbb{P}_h(\Phi_{\lambda,\alpha} = 0) \leqslant \beta.$$

The proof of this lemma is postponed in Section 6.6.1.

In order to have an idea of the order of the right hand side of (6.3), we are now interested in the control of $q^\alpha_{1-\beta/2}$, the $(1-\beta/2)$-quantile of $q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha)$. A sharp upper bound for $q^\alpha_{1-\beta/2}$ is given by the following lemma.

Lemma 6.2. *Let α, β be fixed levels in $]0;1[$. For any $\lambda = (j,k) \in \Gamma$, there exists some positive constant κ depending on β, μ_c and R_1 such that*

$$q^\alpha_{1-\beta/2} \leqslant \kappa \left\{ \sqrt{\ln(2/\alpha)} \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2}\sqrt{n}}{T} \right) + \ln(2/\alpha) \left(\frac{2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT} \right) \right\}.$$

The proof of this lemma is postponed in Section 6.6.2.

Now, observe that if Condition (3.1) of Theorem 1 is satisfied, namely

$$|\beta_\lambda| > \sqrt{\frac{2\zeta Q_{j,n,T}}{\beta}} + \kappa \left\{ \sqrt{\ln(2/\alpha)} \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2}\sqrt{n}}{T} \right) + \ln(2/\alpha) \left(\frac{2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT} \right) \right\},$$

then by Lemma 6.2,

$$|\beta_\lambda| > \sqrt{\frac{2\zeta Q_{j,n,T}}{\beta}} + q^\alpha_{1-\beta/2}.$$

We notice by Jensen's inequality that $|\beta_\lambda| = |\mathbb{E}_h(\hat{\beta}_\lambda)| \leqslant \mathbb{E}_h(|\hat{\beta}_\lambda|) = \mathbb{E}_h(\hat{T}_\lambda)$. Thus, Condition (6.3) of Lemma 6.1 is satisfied and by Lemma 6.1,

$$\mathbb{P}_h(\Phi_{\lambda,\alpha} = 0) \leqslant \beta,$$

which concludes the proof of Theorem 1.

6.4 Proof of Theorem 2

Since $u_\alpha^{[U_1,\dots,U_n;N_{c,tot}]} \geqslant \alpha$ (see Proposition 2) and by setting $\alpha_\lambda = \alpha e^{-w_\lambda}$, we have

$$\begin{aligned}
\mathbb{P}_h(\Phi_\alpha = 0) &= \mathbb{P}_h\left(\forall \lambda \in \Gamma, \hat{T}_\lambda \leqslant q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(u_\alpha^{[U_1,\dots,U_n;N_{c,tot}]} e^{-w_\lambda})\right) \\
&\leqslant \mathbb{P}_h\left(\forall \lambda \in \Gamma, \hat{T}_\lambda \leqslant q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha_\lambda)\right) \\
&\leqslant \min_{\lambda\in\Gamma} \mathbb{P}_h\left(\hat{T}_\lambda \leqslant q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha_\lambda)\right) \\
&\leqslant \min_{\lambda\in\Gamma} \mathbb{P}_h(\Phi_{\lambda,\alpha_\lambda} = 0) \\
&\leqslant \beta,
\end{aligned}$$

as soon as there exists λ in Γ such that $\mathbb{P}_h(\Phi_{\lambda,\alpha_\lambda} = 0) \leqslant \beta$.

First, let us give the precise values of the constants that appear in Condition (3.2) of Theorem 2:

$$C_1 = 8\left(\frac{\zeta}{\beta} + 3\kappa^2 \ln(2/\alpha)\right), \quad C_2 = 24\kappa^2, \quad C_3 = 8\kappa^2 \ln^2(2/\alpha), \quad C_4 = 16\kappa^2 \ln(2/\alpha) \quad \text{and} \quad C_5 = 8\kappa^2,$$

where ζ and κ are the constants defined respectively by Lemma 6.1 and Lemma 6.2. We recall that $Q_{j,n,T} = \frac{1}{n} + \frac{1}{T} + \frac{2^{-j}n}{T^2}$ and we denote $R_{j,n,T} = \frac{2^j}{n^3} + \frac{2^{-j}}{n^2T^2}$.

Let us assume that there exists one finite subset L of Γ such that Condition (3.2) of Theorem 2 is satisfied. Thus,

$$\begin{aligned}
\|h_L\|_2^2 > 8 &\left(\left(\frac{\zeta}{\beta} + 3\kappa^2 \ln(2/\alpha)\right) D_L + 3\kappa^2 \sum_{\lambda\in L} w_\lambda\right)\left[\frac{1}{n} + \frac{n}{T^2}\right] \\
&+ \left(8\kappa^2 \ln^2(2/\alpha) D_L + 16\kappa^2 \ln(2/\alpha) \sum_{\lambda\in L} w_\lambda + 8\kappa^2 \sum_{\lambda\in L} w_\lambda^2\right)\left[\frac{2^{j_L}}{n^3} + \frac{1}{n^2T^2}\right].
\end{aligned}$$

Since $\ln(2/\alpha) + w_\lambda = \ln(2/\alpha_\lambda)$,

$$\sum_{\lambda\in L} \beta_\lambda^2 > \sum_{\lambda\in L} \left\{8\left(\frac{\zeta}{\beta} + 3\kappa^2 \ln(2/\alpha_\lambda)\right)\left[\frac{1}{n} + \frac{n}{T^2}\right] + 8\kappa^2 \ln^2(2/\alpha_\lambda)\left[\frac{2^{j_L}}{n^3} + \frac{1}{n^2T^2}\right]\right\}$$

and it implies that there exists one coefficient $\lambda = (j,k)$ in Γ such that

$$\beta_\lambda^2 > 8\left(\frac{\zeta}{\beta} + 3\kappa^2 \ln(2/\alpha_\lambda)\right)\left[\frac{1}{n} + \frac{n}{T^2}\right] + 8\kappa^2 \ln^2(2/\alpha_\lambda)\left[\frac{2^j}{n^3} + \frac{1}{n^2T^2}\right].$$

Seeing that $Q_{j,n,T} \leqslant 2\left[\frac{1}{n} + \frac{n}{T^2}\right]$ and $R_{j,n,T} \leqslant \left[\frac{2^j}{n^3} + \frac{1}{n^2T^2}\right]$, we have:

$$\beta_\lambda^2 > 4\frac{\zeta}{\beta} Q_{j,n,T} + 12\kappa^2 \ln(2/\alpha_\lambda) Q_{j,n,T} + 8\kappa^2 \ln^2(2/\alpha_\lambda) R_{j,n,T}.$$

Since $(\sqrt{a}+\sqrt{b})^2 \leqslant 2(a+b)$ and $(\sqrt{a}+\sqrt{b}+\sqrt{c})^2 \leqslant 3(a+b+c)$ for all a, b, c nonnegative real numbers,

$$\beta_\lambda^2 > 4\frac{\zeta}{\beta} Q_{j,n,T} + 4\kappa^2 \ln(2/\alpha_\lambda)\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2}\sqrt{n}}{T}\right)^2 + 4\kappa^2 \ln^2(2/\alpha_\lambda)\left(\frac{2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT}\right)^2$$

and then,

$$\beta_\lambda^2 > \left(\sqrt{\frac{2\zeta}{\beta} Q_{j,n,T}} + \kappa\left\{\sqrt{\ln(2/\alpha_\lambda)}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2}\sqrt{n}}{T}\right) + \ln(2/\alpha_\lambda)\left(\frac{2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT}\right)\right\}\right)^2.$$

Finally, it is equivalent to

$$|\beta_\lambda| > \sqrt{\frac{2\zeta}{\beta} Q_{j,n,T}} + \kappa\left\{\sqrt{\ln(2/\alpha_\lambda)}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2}\sqrt{n}}{T}\right) + \ln(2/\alpha_\lambda)\left(\frac{2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT}\right)\right\},$$

which is exactly Condition (3.1) of Theorem 1 and we conclude the proof of Theorem 2 by applying Theorem 1.

6.5 Proof of Theorem 3

With T proportional to n, Condition (3.2) of Theorem 2 is satisfied if there exists one finite subset L of Γ such that

$$\|h\|_2^2 \geqslant \|h - h_L\|_2^2 + C(\alpha, \beta, \mu_c, R_1, R_\infty) \left\{ \left(D_L + \sum_{\lambda \in L} w_\lambda \right) \frac{1}{n} + \left(D_L + \sum_{\lambda \in L} w_\lambda + \sum_{\lambda \in L} w_\lambda^2 \right) \frac{2^{j_L}}{n^3} \right\},$$

with $j_L = \max\{j \geqslant 0 : (j,k) \in L \text{ with } k \in \mathcal{K}_j\}$, $\sum_{\lambda \in L} w_\lambda \leqslant C \times (j_L + 1) D_L$ and $\sum_{\lambda \in L} w_\lambda^2 \leqslant C \times (j_L + 1)^2 D_L$. Consequently, Condition (3.2) is satisfied if there exists one finite subset L of Γ such that

$$\|h\|_2^2 \geqslant \|h - h_L\|_2^2 + C(\alpha, \beta, \mu_c, R_1, R_\infty) \frac{(j_L + 1)}{n} D_L, \tag{6.4}$$

with the maximal resolution level j_L such that $2^{j_L} \leqslant n^2 / \ln n$.

Let $J \geqslant 1$ that will be chosen later. We consider the following finite subset Γ_J of Γ

$$\Gamma_J = \{\lambda = (j,k) \in \Gamma : 0 \leqslant j \leqslant J, k \in \mathcal{K}_j\}.$$

We introduce for all integer $D \leqslant |\Gamma_J|$ the subset L of Γ_J such that $\{\beta_\lambda, \lambda \in L\}$ is the set of the D largest coefficients among $\{\beta_\lambda, \lambda \in \Gamma_J\}$. We can notice that

$$\|h - h_L\|_2^2 = \|h - h_{\Gamma_J}\|_2^2 + \|h_{\Gamma_J} - h_L\|_2^2.$$

On the one hand, since h belongs to $\mathcal{B}^\delta_{2,\infty}(R)$,

$$\|h - h_{\Gamma_J}\|_2^2 = \sum_{j > J} \sum_{k \in \mathcal{K}_j} \beta_{(j,k)}^2 \leqslant C(\delta) R^2 2^{-2J\delta}.$$

On the other hand, using equivalent definitions of weak Besov balls given by Lemma 2.2 of [21] and using for instance page 211 of [11], we obtain:

$$\|h_{\Gamma_J} - h_L\|_2^2 \leqslant C(\gamma) R''^{2+4\gamma} D^{-2\gamma},$$

since h belongs to $\mathcal{W}^*_{\frac{2}{1+2\gamma}}(R')$, with R'' an absolute positive constant depending eventually on γ and R'.

Taking

$$J = \lfloor \log_2 (n^\varepsilon) \rfloor + 1$$

for some $0 < \varepsilon < 2$, we obtain that the right hand side of (6.4) is upper bounded by

$$C(\delta, \gamma, R, R', \alpha, \beta, \mu_c, R_1, R_\infty) \left(n^{-2\varepsilon\delta} + D^{-2\gamma} + \frac{\varepsilon D \ln n}{n} \right).$$

Taking $D = \left\lfloor (n / \ln n)^{1/(1+2\gamma)} \right\rfloor$ and $\varepsilon > \gamma / (\delta(1 + 2\gamma))$, we obtain that the right hand side of (6.4) is upper bounded by

$$C(\delta, \gamma, R, R', \alpha, \beta, \mu_c, R_1, R_\infty) \left(\frac{n}{\ln n} \right)^{\frac{-2\gamma}{1+2\gamma}}$$

when $2\delta > \gamma / (1 + 2\gamma)$ and so,

$$\rho(\Phi_\alpha, \mathcal{B}^\delta_{2,\infty}(R) \cap \mathcal{W}^*_{\frac{2}{1+2\gamma}}(R'), \beta) \leqslant C(\delta, \gamma, R, R', \alpha, \beta, \mu_c, R_1, R_\infty) \left(\frac{n}{\ln n} \right)^{\frac{-\gamma}{1+2\gamma}},$$

which concludes the proof of Theorem 3.

6.6 Proof of lemmas

6.6.1 Proof of Lemma 6.1

Let $\lambda \in \Gamma$ be fixed. From Markov's inequality, we have that for any $x > 0$,

$$\mathbb{P}_h\left(\left|\hat{T}_\lambda - \mathbb{E}_h(\hat{T}_\lambda)\right| \geqslant x\right) \leqslant \frac{\mathrm{Var}(\hat{T}_\lambda)}{x^2}. \tag{6.5}$$

Let us control $\mathrm{Var}(\hat{T}_\lambda) = \mathbb{E}_h(\hat{T}_\lambda^2) - \mathbb{E}_h^2(\hat{T}_\lambda)$. We easily obtain by Jensen's inequality and by considering the decomposition (6.2) of $\mathcal{G}(\varphi_\lambda)$:

$$\begin{aligned}
\mathrm{Var}(\hat{T}_\lambda) &\leqslant \mathrm{Var}(\hat{\beta}_\lambda) \\
&\leqslant \frac{1}{n^2}\mathrm{Var}(G^0(\varphi_\lambda) + G(\varphi_\lambda)) \\
&\leqslant \frac{2}{n^2}\big[\mathrm{Var}(G^0(\varphi_\lambda)) + \mathrm{Var}(G(\varphi_\lambda))\big],
\end{aligned}$$

with

$$\mathrm{Var}(G(\varphi_\lambda)) \leqslant C(R_1, R_\infty)\left\{n + \frac{n^2}{T} + \frac{2^{-j}n^3}{T^2}\right\},$$

by applying the second part of Proposition 1 of [34]. It remains to compute $\mathrm{Var}(G^0(\varphi_\lambda))$. For this purpose, we apply the same methodology developed in Section 6.1.2 of [34]. We have the following decomposition of $\mathrm{Var}(G^0(\varphi_\lambda))$ into two terms:

$$\mathrm{Var}(G^0(\varphi_\lambda)) = \mathbb{E}(\mathrm{Var}(G^0(\varphi_\lambda)|U_1, \ldots, U_n)) + \mathrm{Var}(\mathbb{E}(G^0(\varphi_\lambda)|U_1, \ldots, U_n)). \tag{6.6}$$

We start by dealing with the first term of (6.6). We have

$$\begin{aligned}
&\mathrm{Var}(G^0(\varphi_\lambda)|U_1, \ldots, U_n) \\
&= \int_{-1}^{T+1}\left(\sum_{i=1}^n\left[\varphi_\lambda(x - U_i) - \frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(x - U))\right]\right)^2 \mu_c\, dx \\
&= \mu_c\int_{-1}^{T+1}\left(\varphi_\lambda(x - U_1) + \sum_{i=2}^n\left[\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))\right]\right)^2 dx \\
&= \mu_c\int_{-1}^{T+1}\varphi_\lambda^2(x - U_1)\, dx + 2\mu_c\int_{-1}^{T+1}\varphi_\lambda(x - U_1)\sum_{i=2}^n\left[\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))\right] dx \\
&\quad + \mu_c\int_{-1}^{T+1}\sum_{i=2}^n\sum_{k=2}^n\left[\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))\right]\left[\varphi_\lambda(x - U_k) - \mathbb{E}_\pi(\varphi_\lambda(x - U))\right] dx.
\end{aligned}$$

In the first integral, write $y = x - U_1$. So,

$$\begin{aligned}
&\mathrm{Var}(G^0(\varphi_\lambda)|U_1, \ldots, U_n) \\
&= \mu_c\|\varphi_\lambda\|_2^2 + 2\mu_c\int_{-1}^{T+1}\varphi_\lambda(x - U_1)\sum_{i=2}^n\left[\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))\right] dx \\
&\quad + \mu_c\int_{-1}^{T+1}\sum_{i=2}^n\sum_{k=2}^n\left[\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))\right]\left[\varphi_\lambda(x - U_k) - \mathbb{E}_\pi(\varphi_\lambda(x - U))\right] dx.
\end{aligned}$$

Thus,

$$\mathbb{E}(\mathrm{Var}(G^0(\varphi_\lambda)|U_1, \ldots, U_n)) = \mu_c\|\varphi_\lambda\|_2^2 + \mu_c\int_{-1}^{T+1}\sum_{i=2}^n\mathbb{E}\left(\left[\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))\right]^2\right) dx$$

$$= \mu_c \|\varphi_\lambda\|_2^2 + (n-1)\mu_c \int_{-1}^{T+1} \mathrm{Var}_\pi(\varphi_\lambda(x-U))\, dx$$
$$\leqslant \mu_c \|\varphi_\lambda\|_2^2 + (n-1)\mu_c (T+2)\frac{\|\varphi_\lambda\|_2^2}{T}$$
$$\leqslant C(\mu_c) n, \tag{6.7}$$

by using (6.1) and Lemma 6.1 of [34].

Now, we deal with the second term of (6.6). We have

$$\mathbb{E}(G^0(\varphi_\lambda)|U_1,\ldots,U_n) = \int_{-1}^{T+1} \varphi_\lambda(x-U_1)\mu_c\, dx + \sum_{i=2}^{n} \int_{-1}^{T+1} \left[\varphi_\lambda(x-U_i) - \mathbb{E}_\pi(\varphi_\lambda(x-U))\right] \mu_c\, dx$$
$$= \mu_c \sum_{i=2}^{n} \int_{-1}^{T+1} \left[\varphi_\lambda(x-U_i) - \mathbb{E}_\pi(\varphi_\lambda(x-U))\right]\, dx,$$

by using (6.1). Therefore,

$$\mathrm{Var}(\mathbb{E}(G^0(\varphi_\lambda)|U_1,\ldots,U_n)) = \mu_c^2 \mathrm{Var}\left(\sum_{i=2}^{n} \int_{-1}^{T+1} \left[\varphi_\lambda(x-U_i) - \mathbb{E}_\pi(\varphi_\lambda(x-U))\right]\, dx\right)$$
$$= \mu_c^2 (n-1)\mathrm{Var}\left(\int_{-1}^{T+1} \left[\varphi_\lambda(x-U_1) - \mathbb{E}_\pi(\varphi_\lambda(x-U))\right]\, dx\right)$$
$$\leqslant \mu_c^2 (n-1)\mathbb{E}\left[\left(\int_{-1}^{T+1} |\varphi_\lambda(x-U_1)|\, dx\right)^2\right]$$
$$\leqslant \mu_c^2 (n-1)\|\varphi_\lambda\|_1^2$$
$$\leqslant C(\mu_c) 2^{-j} n. \tag{6.8}$$

Finally, by combining inequalities (6.6), (6.7) and (6.8), we obtain:

$$\mathrm{Var}(G^0(\varphi_\lambda)) \leqslant C(\mu_c) n.$$

Thus,

$$\mathrm{Var}(\hat{T}_\lambda) \leqslant \frac{C(\mu_c, R_1, R_\infty)}{n^2}\left\{n + \frac{n^2}{T} + \frac{2^{-j}n^3}{T^2}\right\} \leqslant \zeta Q_{j,n,T},$$

with

$$Q_{j,n,T} = \frac{1}{n} + \frac{1}{T} + \frac{2^{-j}n}{T^2}$$

and ζ a positive constant depending on μ_c, R_1 and R_∞.

Taking $x = \sqrt{2\zeta Q_{j,n,T}/\beta}$ in (6.5) and using the previous inequality leads to

$$\mathbb{P}_h\left(\left|\hat{T}_\lambda - \mathbb{E}_h(\hat{T}_\lambda)\right| \geqslant \sqrt{2\zeta Q_{j,n,T}/\beta}\right) \leqslant \frac{\beta}{2}.$$

Therefore, if $\mathbb{E}_h(\hat{T}_\lambda) > \sqrt{2\zeta Q_{j,n,T}/\beta} + q^\alpha_{1-\beta/2}$, then

$$\mathbb{P}_h(\hat{T}_\lambda \leqslant q^\alpha_{1-\beta/2}) = \mathbb{P}_h\big(\hat{T}_\lambda - \mathbb{E}_h(\hat{T}_\lambda) \leqslant q^\alpha_{1-\beta/2} - \mathbb{E}_h(\hat{T}_\lambda)\big)$$
$$\leqslant \mathbb{P}_h\left(\left|\hat{T}_\lambda - \mathbb{E}_h(\hat{T}_\lambda)\right| \geqslant \mathbb{E}_h(\hat{T}_\lambda) - q^\alpha_{1-\beta/2}\right)$$
$$\leqslant \mathbb{P}_h\left(\left|\hat{T}_\lambda - \mathbb{E}_h(\hat{T}_\lambda)\right| \geqslant \sqrt{2\zeta Q_{j,n,T}/\beta}\right)$$
$$\leqslant \beta/2$$

and so

$$\mathbb{P}_h(\Phi_{\lambda,\alpha} = 0) \leqslant \beta,$$

which concludes the proof of Lemma 6.1.

6.6.2 Proof of Lemma 6.2

We focus first on the control of the conditional quantile $q_\lambda^{[U_1,\dots,U_n;N_{c,tot}]}(\alpha)$. For all $m \in \mathbb{N}^*$, the $(1-\alpha)$-quantile $q_\lambda^{[U_1,\dots,U_n;m]}(\alpha)$ is the smallest real number such that

$$\mathbb{P}\left(\hat{T}^0_{\lambda,m} > q_\lambda^{[U_1,\dots,U_n;m]}(\alpha)\middle| U_1,\dots,U_n; N_{c,tot}=m\right) \leqslant \alpha,$$

where $\hat{T}^0_{\lambda,m}$ is defined by (2.2). Let $m \in \mathbb{N}^*$ be fixed. We write

$$\hat{T}^0_{\lambda,m} = \frac{1}{n}\left|\sum_{k=1}^m S(\varphi_\lambda)(V_k^0)\right|,$$

where $(V_1^0,\dots,V_m^0)$ is a m-sample with uniform distribution on $[-1;T+1]$ and for any $v \in [-1;T+1]$,

$$S(\varphi_\lambda)(v) = \sum_{i=1}^n \left[\varphi_\lambda(v-U_i) - \frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(v-U))\right].$$

Since $\mathbb{E}(\varphi_\lambda(V-U)|U)=0$ for independent random variables U and V uniformly distributed on $[0;T]$ and $[-1;T+1]$ respectively, the $S(\varphi_\lambda)(V_k^0)$'s are centered and independent conditionally on $U_1,\dots,U_n$. Then we apply Bernstein's inequality (for instance, see Proposition 2.9 of [26]) to get that for all $\omega>0$, with probability larger than $1-2e^{-\omega}$,

$$\left|\sum_{k=1}^m S(\varphi_\lambda)(V_k^0)\right| \leqslant \sqrt{2m\mathrm{Var}(S(\varphi_\lambda)(V_1^0)|U_1,\dots,U_n)\omega} + \frac{\omega}{3}\sup_{v\in[-1;T+1]}\left|S(\varphi_\lambda)(v)\right|.$$

Thus, with probability larger than $1-\alpha$,

$$\hat{T}^0_{\lambda,m} \leqslant f(U_1,\dots,U_n;m),$$

with

$$f(U_1,\dots,U_n;m) = \frac{1}{n}\left\{\sqrt{2m\ln(2/\alpha)V_S} + \frac{\ln(2/\alpha)}{3}B_S\right\}, \tag{6.9}$$

where

$$V_S = \mathrm{Var}(S(\varphi_\lambda)(V_1^0)|U_1,\dots,U_n) \quad \text{and} \quad B_S = \sup_{v\in[-1;T+1]}|S(\varphi_\lambda)(v)|.$$

Therefore we have $q_\lambda^{[U_1,\dots,U_n;m]}(\alpha) \leqslant f(U_1,\dots,U_n;m)$ by definition of the quantile $q_\lambda^{[U_1,\dots,U_n;m]}(\alpha)$.

Let us now provide a control in probability of $f(U_1,\dots,U_n;m)$. We control first V_S.

$$\begin{aligned}
V_S &= \mathrm{Var}\left(\sum_{i=1}^n \varphi_\lambda(V_1^0-U_i) - (n-1)\mathbb{E}_\pi\big(\varphi_\lambda(V_1^0-U)\big)\middle| U_1,\dots,U_n\right)\\
&\leqslant \mathbb{E}\left[\Big(\sum_{i=1}^n \varphi_\lambda(V_1^0-U_i) - (n-1)\mathbb{E}_\pi(\varphi_\lambda(V_1^0-U))\Big)^2\middle| U_1,\dots,U_n\right]\\
&\leqslant \frac{1}{T+2}\int_{v=-1}^{T+1}\left(\sum_{i=1}^n \varphi_\lambda(v-U_i) - (n-1)\mathbb{E}_\pi(\varphi_\lambda(v-U))\right)^2 dv\\
&\leqslant \frac{2}{T+2}\int_{v=-1}^{T+1}\left(\sum_{1\leqslant i,k\leqslant n}\varphi_\lambda(v-U_i)\varphi_\lambda(v-U_k) + (n-1)^2\mathbb{E}_\pi^2(\varphi_\lambda(v-U))\right) dv\\
&\leqslant \frac{2}{T+2}\left\{\int_{v=-1}^{T+1}\sum_{i=1}^n \varphi_\lambda^2(v-U_i)\,dv + \int_{v=-1}^{T+1}\sum_{1\leqslant i\neq k\leqslant n}\varphi_\lambda(v-U_i)\varphi_\lambda(v-U_k)\,dv\right.
\end{aligned}$$

$$+\frac{(n-1)^2}{T^2}\int_{v=-1}^{T+1}\left(\int_0^T|\varphi_\lambda|(v-u)\,du\right)^2 dv\Bigg\}$$

$$\leqslant\frac{2}{T+2}\left\{n\|\varphi_\lambda\|_2^2+\int_{v=-1}^{T+1}\sum_{1\leqslant i\neq k\leqslant n}\varphi_\lambda(v-U_i)\varphi_\lambda(v-U_k)\,dv+\frac{(n-1)^2}{T^2}(T+2)\|\varphi_\lambda\|_1^2\right\}$$

$$\leqslant\frac{C}{T}\left\{n+\sum_{1\leqslant i\neq k\leqslant n}\int_{v=-1}^{T+1}\varphi_\lambda(v-U_i)\varphi_\lambda(v-U_k)\,dv+\frac{2^{-j}n^2}{T}\right\},\tag{6.10}$$

with C an absolute positive constant. We have a decomposition of the second term in a sum of degenerate U-statistics of order 0, 1 and 2. Indeed

$$\sum_{1\leqslant i\neq k\leqslant n}\int_{v=-1}^{T+1}\varphi_\lambda(v-U_i)\varphi_\lambda(v-U_k)\,dv=W_0+W_{1,1}+W_{1,2}+W_2,$$

with

$$W_2=\sum_{1\leqslant i\neq k\leqslant n}\int_{v=-1}^{T+1}[\varphi_\lambda(v-U_i)-\mathbb{E}_\pi(\varphi_\lambda(v-U))][\varphi_\lambda(v-U_k)-\mathbb{E}_\pi(\varphi_\lambda(v-U))]\,dv,$$

$$W_{1,1}=\sum_{1\leqslant i\neq k\leqslant n}\int_{v=-1}^{T+1}\varphi_\lambda(v-U_i)\mathbb{E}_\pi(\varphi_\lambda(v-U))\,dv,$$

$$W_{1,2}=\sum_{1\leqslant i\neq k\leqslant n}\int_{v=-1}^{T+1}\mathbb{E}_\pi(\varphi_\lambda(v-U))\varphi_\lambda(v-U_k)\,dv$$

and

$$W_0=-\sum_{1\leqslant i\neq k\leqslant n}\int_{v=-1}^{T+1}\mathbb{E}_\pi^2(\varphi_\lambda(v-U))\,dv.$$

First we control W_0:

$$\begin{aligned}|W_0|&\leqslant\frac{n(n-1)(T+2)}{T^2}\|\varphi_\lambda\|_1^2\\&\leqslant C\frac{2^{-j}n^2}{T},\end{aligned}\tag{6.11}$$

with C an absolute positive constant. Next we deal with the control of $W_{1,1}$ and $W_{1,2}$. We notice that

$$W_{1,1}=W_{1,2}=(n-1)\sum_{i=1}^n\int_{v=-1}^{T+1}\varphi_\lambda(v-U_i)\mathbb{E}_\pi(\varphi_\lambda(v-U))\,dv$$

and consequently we have by using Lemma 6.3 of [34]

$$\begin{aligned}|W_{1,1}|=|W_{1,2}|&\leqslant(n-1)\sum_{i=1}^n\int_{v=-1}^{T+1}|\varphi_\lambda|(v-U_i)\,dv\frac{\|\varphi_\lambda\|_1}{T}\\&\leqslant C\frac{2^{-j}n^2}{T},\end{aligned}\tag{6.12}$$

with C an absolute positive constant.

Now it remains to control W_2, with

$$W_2=\sum_{1\leqslant i<k\leqslant n}g(U_i,U_k),$$

where

$$g(U_i,U_k)=2\int_{v=-1}^{T+1}[\varphi_\lambda(v-U_i)-\mathbb{E}_\pi(\varphi_\lambda(v-U))][\varphi_\lambda(v-U_k)-\mathbb{E}_\pi(\varphi_\lambda(v-U))]\,dv.$$

One can apply Theorem 3.4 of [19] to W_2 and $-W_2$. It implies that there exist absolute positive constants c_1, c_2, c_3 and c_4 such that with probability larger than $1-2\times 2.77e^{-\omega}$,

$$|W_2|\leqslant c_1\mathcal{C}\sqrt{\omega}+c_2D\omega+c_3B\omega^{3/2}+c_4A\omega^2$$

for all $\omega>0$, where

- $A=\|g\|_\infty\leqslant 8\|\varphi_\lambda\|_1\|\varphi_\lambda\|_\infty\leqslant 8$;
- $\mathcal{C}^2=\mathbb{E}(W_2^2)$ and we have

$$\begin{aligned}\mathcal{C}^2&=\sum_{1\leqslant i<k\leqslant n}\mathbb{E}(g^2(U_i,U_k))\\&\leqslant 4n(n-1)\mathbb{E}\left[\left(\int_{v=-1}^{T+1}[\varphi_\lambda(v-U_1)-\mathbb{E}_\pi(\varphi_\lambda(v-U))][\varphi_\lambda(v-U_2)-\mathbb{E}_\pi(\varphi_\lambda(v-U))]\,dv\right)^2\right].\end{aligned}$$

We denote $\mathbb{E}_{(U,U')\sim\pi\otimes\pi}(f(U,U'))$ the expectation of $f(U,U')$ where $U\sim\pi$ and $U'\sim\pi$ are independent and $f^U(v)=f(v-U)$. Hence,

$$\begin{aligned}&\mathcal{C}^2\\&\leqslant 4n(n-1)\mathbb{E}_{(U,U')\sim\pi\otimes\pi}\left[\left(\int_{v=-1}^{T+1}\left[\varphi_\lambda^U(v)-\mathbb{E}_\pi(\varphi_\lambda^U(v))\right]\left[\varphi_\lambda^{U'}(v)-\mathbb{E}_\pi(\varphi_\lambda^{U'}(v))\right]dv\right)^2\right]\\&\leqslant 4n^2\mathbb{E}_{(U,U')\sim\pi\otimes\pi}\left[\left(\int_{v=-1}^{T+1}\varphi_\lambda^U(v)\varphi_\lambda^{U'}(v)\,dv-\mathbb{E}_{U'\sim\pi}\left(\int_{v=-1}^{T+1}\varphi_\lambda^U(v)\varphi_\lambda^{U'}(v)\,dv\right)\right.\right.\\&\qquad\left.\left.-\mathbb{E}_{U\sim\pi}\left(\int_{v=-1}^{T+1}\varphi_\lambda^U(v)\varphi_\lambda^{U'}(v)\,dv\right)+\mathbb{E}_{(U,U')\sim\pi\otimes\pi}\left(\int_{v=-1}^{T+1}\varphi_\lambda^U(v)\varphi_\lambda^{U'}(v)\,dv\right)\right)^2\right]\\&\leqslant Cn^2\left\{\mathbb{E}_{(U,U')\sim\pi\otimes\pi}\left[\left(\int_{v=-1}^{T+1}|\varphi_\lambda^U|(v)|\varphi_\lambda^{U'}|(v)\,dv\right)^2\right]+\left[\mathbb{E}_{(U,U')\sim\pi\otimes\pi}\left(\int_{v=-1}^{T+1}|\varphi_\lambda^U|(v)|\varphi_\lambda^{U'}|(v)\,dv\right)\right]^2\right\},\end{aligned}$$

with C an absolute positive constant. But,

$$\begin{aligned}&\mathbb{E}_{(U,U')\sim\pi\otimes\pi}\left[\left(\int_{v=-1}^{T+1}|\varphi_\lambda^U|(v)|\varphi_\lambda^{U'}|(v)\,dv\right)^2\right]\\&\leqslant\mathbb{E}_{(U,U')\sim\pi\otimes\pi}\left(\int_{v=-1}^{T+1}|\varphi_\lambda^U|^2(v)|\varphi_\lambda^{U'}|(v)\,dv\int_{v=-1}^{T+1}|\varphi_\lambda^{U'}|(v)\,dv\right)\\&=\mathbb{E}_{(U,U')\sim\pi\otimes\pi}\left(\int_{v=-1}^{T+1}|\varphi_\lambda^U|^2(v)|\varphi_\lambda^{U'}|(v)\,dv\right)\|\varphi_\lambda\|_1\\&\leqslant\|\varphi_\lambda\|_2^2\frac{\|\varphi_\lambda\|_1^2}{T}\end{aligned}$$

and

$$\mathbb{E}_{(U,U')\sim\pi\otimes\pi}\left(\int_{v=-1}^{T+1}|\varphi_\lambda^U|(v)|\varphi_\lambda^{U'}|(v)\,dv\right)=\int_{v=-1}^{T+1}\mathbb{E}_\pi(|\varphi_\lambda^U|(v))\mathbb{E}_\pi(|\varphi_\lambda^{U'}|(v))\,dv\leqslant(T+2)\frac{\|\varphi_\lambda\|_1^2}{T^2},$$

by using Lemma 6.3 of [34]. So,

$$\mathcal{C}^2\leqslant Cn^2\left\{\frac{2^{-j}}{T}+\frac{2^{-2j}}{T^2}\right\},$$

with C an absolute positive constant;

- $D = \sup\left\{ \mathbb{E}\left(\sum_{1\leqslant k<i\leqslant n} g(U_i,U_k)a_i(U_i)b_k(U_k)\right) : \mathbb{E}\left(\sum_{i=2}^{n} a_i(U_i)^2\right)\leqslant 1, \mathbb{E}\left(\sum_{k=1}^{n-1} b_k(U_k)^2\right)\leqslant 1\right\}$.
But, with the conditions on the a_i's and the b_k's, we have:

$$
\begin{aligned}
&\mathbb{E}\left(\sum_{1\leqslant k<i\leqslant n} g(U_i,U_k)a_i(U_i)b_k(U_k)\right)\\
&= 2\mathbb{E}\left(\sum_{i=2}^{n}\sum_{k=1}^{i-1}\int_{v=-1}^{T+1}[\varphi_\lambda(v-U_i)-\mathbb{E}_\pi(\varphi_\lambda(v-U))][\varphi_\lambda(v-U_k)-\mathbb{E}_\pi(\varphi_\lambda(v-U))]\,dv a_i(U_i)b_k(U_k)\right)\\
&\leqslant 2\int_{v=-1}^{T+1}\mathbb{E}\Big(\sum_{i=2}^{n}\big|\varphi_\lambda(v-U_i)-\mathbb{E}_\pi(\varphi_\lambda^U(v))\big||a_i(U_i)|\Big)\mathbb{E}\Big(\sum_{k=1}^{n-1}\big|\varphi_\lambda(v-U_k)-\mathbb{E}_\pi(\varphi_\lambda^U(v))\big||b_k(U_k)|\Big)\,dv\\
&\leqslant 2\int_{v=-1}^{T+1}\sqrt{(n-1)\mathrm{Var}_\pi(\varphi_\lambda(v-U))}\mathbb{E}\Big(\sum_{k=1}^{n-1}\big|\varphi_\lambda(v-U_k)-\mathbb{E}_\pi(\varphi_\lambda^U(v))\big||b_k(U_k)|\Big)\,dv\\
&\leqslant 2\sqrt{(n-1)\frac{\|\varphi_\lambda\|_2^2}{T}}\mathbb{E}\left(\sum_{k=1}^{n-1}\int_{v=-1}^{T+1}\big|\varphi_\lambda(v-U_k)-\mathbb{E}_\pi(\varphi_\lambda(v-U))\big||b_k(U_k)|\,dv\right)\\
&\leqslant 2\sqrt{\frac{n-1}{T}}\|\varphi_\lambda\|_2\mathbb{E}\left(2\|\varphi_\lambda\|_1\sum_{k=1}^{n-1}|b_k(U_k)|\right)\\
&\leqslant 4\sqrt{\frac{n-1}{T}}\|\varphi_\lambda\|_2\|\varphi_\lambda\|_1\sqrt{n-1}\\
&\leqslant 4\frac{n-1}{\sqrt{T}}\|\varphi_\lambda\|_1\|\varphi_\lambda\|_2,
\end{aligned}
$$

by using Lemma 6.1 of [34]. Then,

$$D\leqslant C\frac{2^{-j/2}n}{\sqrt{T}},$$

with C an absolute positive constant;

- $B^2 = \sup_u\left(\sum_{k=1}^{n-1}\mathbb{E}(g^2(u,U_k))\right)$, with

$$
\begin{aligned}
&\mathbb{E}(g^2(u,U_k))\\
&= 4\mathbb{E}\left[\left(\int_{v=-1}^{T+1}[\varphi_\lambda(v-u)-\mathbb{E}_\pi(\varphi_\lambda(v-U))][\varphi_\lambda(v-U_k)-\mathbb{E}_\pi(\varphi_\lambda(v-U))]\,dv\right)^2\right]\\
&\leqslant 4\mathbb{E}\left[\int_{v=-1}^{T+1}\big[\varphi_\lambda^u(v)-\mathbb{E}_\pi(\varphi_\lambda^U(v))\big]^2\big|\varphi_\lambda^{U_k}(v)-\mathbb{E}_\pi(\varphi_\lambda^U(v))\big|\,dv\int_{v=-1}^{T+1}\big|\varphi_\lambda^{U_k}(v)-\mathbb{E}_\pi(\varphi_\lambda^U(v))\big|\,dv\right]\\
&\leqslant 8\mathbb{E}\left[\int_{v=-1}^{T+1}\big[\varphi_\lambda(v-u)-\mathbb{E}_\pi(\varphi_\lambda(v-U))\big]^2\big|\varphi_\lambda(v-U_k)-\mathbb{E}_\pi(\varphi_\lambda(v-U))\big|\,dv\right]\|\varphi_\lambda\|_1\\
&\leqslant \frac{16}{T}\int_{v=-1}^{T+1}\big[\varphi_\lambda(v-u)-\mathbb{E}_\pi(\varphi_\lambda(v-U))\big]^2\,dv\|\varphi_\lambda\|_1^2\\
&\leqslant \frac{64}{T}\|\varphi_\lambda\|_1^2\|\varphi_\lambda\|_2^2,
\end{aligned}
$$

by using Lemma 6.3 of [34]. Hence,

$$B^2\leqslant C\frac{2^{-j}n}{T},$$

with C an absolute positive constant.

Finally, we obtain for all $\omega > 0$, with probability larger than $1 - 2 \times 2.77e^{-\omega}$,

$$|W_2| \leqslant C \left\{ \frac{2^{-j/2}n}{\sqrt{T}}\sqrt{\omega} + \frac{2^{-j}n}{T}\sqrt{\omega} + \frac{2^{-j/2}n}{\sqrt{T}}\omega + \frac{2^{-j/2}\sqrt{n}}{\sqrt{T}}\omega^{3/2} + \omega^2 \right\}, \tag{6.13}$$

with C an absolute positive constant.

Thus, by inequalities (6.10), (6.11), (6.12) and (6.13), for all $\omega > 0$, with probability larger than $1 - 2 \times 2.77e^{-\omega}$,

$$V_S \leqslant \frac{C(\omega)}{T} \left\{ n + \frac{2^{-j}n^2}{T} + \frac{2^{-j/2}n}{\sqrt{T}} \right\}. \tag{6.14}$$

Then it remains to compute B_S. We recall that

$$\begin{aligned} B_S &= \sup_{v\in[-1;T+1]} \left| \sum_{i=1}^{n} \left[\varphi_\lambda(v - U_i) - \frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(v-U)) \right] \right| \\ &\leqslant \tilde{B}_S + \frac{1}{T}\|\varphi_\lambda\|_1, \end{aligned}$$

with $\tilde{B}_S = \sup_{v\in[-1;T+1]} \left| \sum_{i=1}^{n} \left[\varphi_\lambda(v - U_i) - \mathbb{E}_\pi(\varphi_\lambda(v-U)) \right] \right|$. Since the Haar basis is considered here, we can write for any $x \in \mathbb{R}$:

$$\varphi_\lambda(x) = 2^{j/2} \left(\mathbf{1}_{(2k+1)2^{-(j+1)} < x \leqslant (k+1)2^{-j}} - \mathbf{1}_{k2^{-j} \leqslant x \leqslant (2k+1)2^{-(j+1)}} \right),$$

with $\lambda = (j,k)$. Thus,

$$\tilde{B}_S \leqslant 2^{j/2} \left(\tilde{B}_S^1 + \tilde{B}_S^2 \right),$$

where

$$\tilde{B}_S^1 = \sup_{v\in[-1;T+1]} \left| \sum_{i=1}^{n} \left[\mathbf{1}_{k2^{-j} \leqslant v - U_i \leqslant (2k+1)2^{-(j+1)}} - \mathbb{E}_\pi(\mathbf{1}_{k2^{-j} \leqslant v-U \leqslant (2k+1)2^{-(j+1)}}) \right] \right|$$

and

$$\tilde{B}_S^2 = \sup_{v\in[-1;T+1]} \left| \sum_{i=1}^{n} \left[\mathbf{1}_{(2k+1)2^{-(j+1)} < v - U_i \leqslant (k+1)2^{-j}} - \mathbb{E}_\pi(\mathbf{1}_{(2k+1)2^{-(j+1)} < v-U \leqslant (k+1)2^{-j}}) \right] \right|.$$

We observe that

$$\tilde{B}_S^1 \leqslant \sup_{B_v, v\in\mathbb{R}} \left| \sum_{i=1}^{n} \left[\mathbf{1}_{B_v}(U_i) - \mathbb{E}_\pi\big(\mathbf{1}_{B_v}(U)\big) \right] \right|,$$

where for any $v \in \mathbb{R}$, $B_v = [v - (2k+1)2^{-(j+1)}; v - k2^{-j}]$. We set $\mathcal{B} = \{B_v, v \in \mathbb{R}\}$ and for every integer n, $m_n(\mathcal{B}) = \sup_{A\subset\mathbb{R}, |A|=n} |\{A \cap B_v, v \in \mathbb{R}\}|$. It is easy to see that

$$m_n(\mathcal{B}) \leqslant 1 + \frac{n(n+1)}{2}$$

and so, the VC-dimension of $\mathcal{B}$ defined by $\sup\{n \geqslant 0, m_n(\mathcal{B}) = 2^n\}$ is bounded by 2 (see Definition 6.2 of [26]). By applying Lemma 6.4 of [26], we obtain:

$$\sqrt{n}\mathbb{E}(\tilde{B}_S^1) \leqslant \frac{K}{2}\sqrt{2},$$

where K is an absolute constant. So, with a similar argument for $\tilde{B}_S^2$, we obtain for any λ in Γ

$$\mathbb{E}(\tilde{B}_S) \leqslant \frac{2^{j/2}}{\sqrt{n}} K\sqrt{2}.$$

Consequently,

$$\mathbb{E}(B_S) \leqslant C\left\{\frac{2^{j/2}}{\sqrt{n}} + \frac{2^{-j/2}}{T}\right\},$$

with C an absolute positive constant and from Markov's inequality, we have that for all $\omega > 0$

$$\mathbb{P}\left(B_S > C(\omega)\left\{\frac{2^{j/2}}{\sqrt{n}} + \frac{2^{-j/2}}{T}\right\}\right) \leqslant e^{-\omega}. \tag{6.15}$$

Thus, by combining inequalities (6.9), (6.14) and (6.15), we obtain for all $\omega > 0$, with probability larger than $1 - (2 \times 2.77 + 1)e^{-\omega}$,

$$f(U_1, \ldots, U_n; m) \leqslant \frac{C(\omega)}{n}\left\{\sqrt{m \ln(2/\alpha)\left(\frac{n}{T} + \frac{2^{-j}n^2}{T^2} + \frac{2^{-j/2}n}{T^{3/2}}\right)} + \ln(2/\alpha)\left(\frac{2^{j/2}}{\sqrt{n}} + \frac{2^{-j/2}}{T}\right)\right\}.$$

Furthermore, $N_{[-1;T+1]} \sim \mathcal{P}((T+2)\mu_c + n\|h\|_1)$. Hence,

$$\mathbb{E}(N_{[-1;T+1]}) \leqslant C(\mu_c, R_1)(n+T).$$

From Markov's inequality, we have that for all $\omega > 0$

$$\mathbb{P}\left(N_{[-1;T+1]} > C(\omega, \mu_c, R_1)(n+T)\right) \leqslant e^{-\omega}. \tag{6.16}$$

Then, we choose ω such that this quantity $(2 \times 2.77 + 2)e^{-\omega}$ is equal to $\beta/2$. So, with probability larger than $1 - \beta/2$,

$$\begin{aligned}
&f(U_1, \ldots, U_n; m) \\
&\leqslant \frac{C(\beta, \mu_c, R_1)}{n}\left\{\sqrt{\ln(2/\alpha)}\sqrt{(n+T)\left(\frac{n}{T} + \frac{2^{-j}n^2}{T^2} + \frac{2^{-j/2}n}{T^{3/2}}\right)} + \ln(2/\alpha)\left(\frac{2^{j/2}}{\sqrt{n}} + \frac{2^{-j/2}}{T}\right)\right\} \\
&\leqslant \frac{C(\beta, \mu_c, R_1)}{n}\left\{\sqrt{\ln(2/\alpha)}\sqrt{n + \frac{n^2}{T} + \frac{2^{-j}n^3}{T^2}} + \ln(2/\alpha)\left(\frac{2^{j/2}}{\sqrt{n}} + \frac{2^{-j/2}}{T}\right)\right\} \\
&\leqslant C(\beta, \mu_c, R_1)\left\{\sqrt{\ln(2/\alpha)}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2}\sqrt{n}}{T}\right) + \ln(2/\alpha)\left(\frac{2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT}\right)\right\}.
\end{aligned}$$

Therefore by definition of $q^{\alpha}_{1-\beta/2}$,

$$q^{\alpha}_{1-\beta/2} \leqslant C(\beta, \mu_c, R_1)\left\{\sqrt{\ln(2/\alpha)}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2}\sqrt{n}}{T}\right) + \ln(2/\alpha)\left(\frac{2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT}\right)\right\},$$

which concludes the proof of Lemma 6.2.

Acknowledgments: The authors are very grateful to Patricia Reynaud-Bouret and Vincent Rivoirard for stimulating discussions and constructive comments. The research of Laure Sansonnet and Christine Tuleau-Malot is partly supported by the french Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration).

References

[1] Baraud, Y. Non asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8:577–606, 2002.

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

[3] Barlow, H. B. Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, 1:371–394, 1972.

[4] Benjamini, Y. and Hochberg, Y. Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 57(1):289–300, 1995.

[5] Brémaud, P. and Massoulié, L. Stability of nonlinear Hawkes processes. *Ann. Probab.*, 24(3):1563–1588, 1996.

[6] Cohen, A., Daubechies, I. and Feauveau, J.-C. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45:485–560, 1992.

[7] Dachian, S. and Kutoyants, Y. A. Hypotheses testing: Poisson versus self-exciting. *Scand. J Stat.*, 33(2):391Ű–408, Jun. 2006.

[8] Darling, D. A. The Kolmogorov-Smirnov, Cramér-von Mises tests. *Ann. Math. Statist.*, 28(4):823–838, Dec. 1957.

[9] Fazli, K. Second order efficient test for inhomogeneous Poisson processes. *Stat. Inference Stoch. Process.*, 2:181–208, 2007.

[10] Fazli, K. and Kutoyants, Y. A. Two simple hypotheses testing for Poisson process. *Far East J. Theor. Stat.*, 15(2):251–290, 2005.

[11] Fromont, M., Laurent, B. and Reynaud-Bouret, P. Adaptive tests of homogeneity for a Poisson process. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(1):176–213, 2011.

[12] Fromont, M., Laurent, B. and Reynaud-Bouret, P. The two-sample problem for Poisson processes: adaptive tests with a non-asymptotic wild bootstrap approach. *arXiv:1203.3572v1*, Mar. 2012.

[13] Gerstein, G. L. Searching for significance in spatio-temporal firing patterns. *Acta Neurobiol Exp (Wars)*, 64(2):203–207, 2004.

[14] Grammont, F. and Riehle, A. Precise spike synchronization in monkey motor cortex involved in preparation for movement. *Exp Brain Res.*, 128(1-2):118–122, 1999.

[15] Grün, S. *Unitary joint-Events in multiple-neuron spiking activity: detection, significance, and interpretation.* Reihe Physik, Band 60. Verlag Harri Deutsch, Thun, Frankfurt/Main, 1996.

[16] Grün, S., Diesmann, M. and Aertsen, A. Unitary event analysis. In *Analysis of parallel spike trains*, volume 7, pages 191–220. Springer Ser. Comput. Neurosci., 2010.

[17] Gusto, G. and Schbath, S. FADO: a statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes' model. *Stat. Appl. Genet. Mol. Biol.*, 4(1), Sep. 2005.

[18] Hawkes, A. G. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[19] Houdré, C. and Reynaud-Bouret, P. Exponential inequalities, with constants, for U-statistics of order two. In *Stochastic Inequalities and Applications*, volume 56, pages 55–69. Progr. Probab., 2003.

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

[21] Kerkyacharian, G. and Picard, D. Thresholding algorithms, maxisets and well-concentrated bases. With discussion and a rejoinder by the authors. *Test*, 9(2):283–345, 2000.

[22] Kingman, J. F. C. *Poisson Processes*. Oxford Sci. Publ., 1993.

[23] Krumin, M., Reutsky, I. and Shoham, S. Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input. *Front Comput Neurosci.*, 4(147), 2010.

[24] Lestienne, R. Determination of the precision of spike timing in the visual cortex of anaesthetised cats. *Biol. Cybern.*, 74:55–61, 1996.

[25] Mallat, S. G. Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$. *Trans. Amer. Math. Soc.*, 315(1):69–87, Sep. 1989.

[26] Massart, P. *Concentration inequalities and model selection – Ecole d'été de Probabilités de Saint-Flour 2003*, volume 1896. Lecture Notes in Math., 2007.

[27] Pernice, V., Staude, B., Cardanobile, S. and Rotter, S. How structure determines correlations in neuronal networks. *PLoS Comput Biol*, 7(5), 2011.

[28] Pernice, V., Staude, B., Cardanobile, S. and Rotter, S. Recurrent interactions in spiking networks with arbitrary topoplogy. *Phys. Rev. E*, 85(3), 2012.

[29] Reynaud-Bouret, P. and Rivoirard, V. Near optimal thresholding estimation of a Poisson intensity on the real line. *Electron. J. Stat.*, 4:172–238, 2010.

[30] Reynaud-Bouret, P. and Schbath, S. Adaptive estimation for Hawkes processes; application to genome analysis. *Ann. Statist.*, 38(5):2781–2822, 2010.

[31] Reynaud-Bouret, P., Rivoirard, V. and Tuleau-Malot, C. Adaptive density estimation: a curse of support? *J. Statist. Plann. Inference*, 141:115–139, 2011.

[32] Rivoirard, V. Maxisets for linear procedures. *Statist. Probab. Lett.*, 67(3):267–275, 2004.

[33] Rivoirard, V. Thresholding procedure with priors based on Pareto distributions. *Test*, 13(1):213–246, 2004.

[34] Sansonnet, L. Wavelet thresholding estimation in a Poissonian interactions model with application to genomic data. *Accepted for publication in the Scand. J Stat.*, 2012.

[35] Tuleau-Malot, C., Rouis, A., Reynaud-Bouret, P. and Grammont, F. Multiple tests based on a Gaussian approximation of the Unitary Events method. *hal-00757323*, Nov. 2012.