



HAL
open science

Nonparametric estimation in a mixed-effect Ornstein-Uhlenbeck model

Charlotte Dion

► **To cite this version:**

Charlotte Dion. Nonparametric estimation in a mixed-effect Ornstein-Uhlenbeck model. 2014. hal-01023300v1

HAL Id: hal-01023300

<https://hal.science/hal-01023300v1>

Preprint submitted on 11 Jul 2014 (v1), last revised 7 Feb 2016 (v5)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nonparametric estimation in a mixed-effect Ornstein-Uhlenbeck model

Charlotte Dion ^{(1),(2)}

(1)*LJK, UMR CNRS 5224, Université Joseph Fourier,
51 rue des Mathématiques, 38041 Grenoble*

(2)*MAP5, UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité,
45 rue des Saints Pères, 75006 Paris*

Abstract

Two adaptive nonparametric procedures are proposed to estimate the density of the random effects in a mixed-effect Ornstein-Uhlenbeck model. First an estimator using deconvolution tools is introduced, which depends on two tuning parameters to be chosen in a data-driven way. The selection of these two parameters is achieved with a Goldenshluger and Lepski's method, extended to this particular case with a new two-dimensional penalized criterion. Then, we propose a kernel estimator of the density of the random effect, with a new bandwidth selection method. For both data driven estimators, risk bounds are provided in term of integrated \mathbb{L}^2 -error. The estimators are evaluated on simulations and show good results. Finally, these non-parametric estimators are applied to a neuronal database of interspike intervals, and are compared with a previous parametric estimation.

Keywords. stochastic differential equations, Ornstein-Uhlenbeck process, mixed-effect model, nonparametric estimation, deconvolution method, kernel estimator, neuronal interspike interval.

1 Introduction

Stochastic differential models have been surveyed a lot in the theoretical literature with continuous observations (e.g Kutoyants, 2004) or discrete observations, in parametric field (e.g Genon-Catalot and Jacod, 1993) or nonparametric field (e.g Comte *et al.*, 2007; Hoffmann, 1999). More recently, stochastic differential equations with random effects have been introduced with various applications such as neuronal modelling or pharmacokinetic (e.g Picchini *et al.*, 2008; Delattre and Lavielle, 2013; Donnet and Samson, 2013). Mixed-effects models are used to analyse repeated measurements with similar functional form but with some variability between experiments (see Pinheiro and Bates, 2000; Diggle *et al.*, 2002; Davidian and Giltinan, 1995). The advantage is that a single estimation procedure is used to fit the overall data simultaneously.

Estimation methods in stochastic differential models with random effects have been proposed, especially in the parametric framework (e.g Donnet and Samson, 2008; Donnet *et al.*, 2010; Picchini *et al.*, 2010; Picchini and Ditlevsen, 2011; Delattre and Lavielle, 2013; Delattre *et al.*, 2013; Donnet and Samson, 2014; Delattre *et al.*, 2014; Genon-Catalot and Larédo, 2013). All these parametric estimation methods are developed assuming the normality of the random effect. However, one can ask if this assumption is reasonable depending on the application context. We focus here on the nonparametric estimation of the density of the random effect. To the best of our knowledge, the only reference in this context is Comte *et al.* (2013).

To simplify the presentation, we focus on the stochastic Ornstein-Uhlenbeck differential model with one additive random effect. More precisely, we consider N real valued stochastic processes $(X_j(t), t \in [0, T])$, $j = 1, \dots, N$, with dynamics ruled by the following SDEs:

$$\begin{cases} dX_j(t) &= \left(\phi_j - \frac{X_j(t)}{\alpha} \right) dt + \sigma dW_j(t) \\ X_j(0) &= x_j \end{cases} \quad (1)$$

where $(W_j)_{1 \leq j \leq N}$ are N independent Wiener processes, and $(\phi_j)_{1 \leq j \leq N}$ are N unobserved independent and identically distributed (*i.i.d.*) random variables taking values in \mathbb{R} , with a common density f . The sequences $(\phi_j)_{1 \leq j \leq N}$ and $(W_j)_{1 \leq j \leq N}$ are independent. Here (x_1, \dots, x_N) are known values. The positive constants σ and α are known. Each process $(X_j(t), 0 \leq t \leq T)$ represents an individual, the variable ϕ_j is the random effect of individual j . Due to the independence of the ϕ_j and the W_j , the $X_j(t)$, $j = 1, \dots, N$ are *i.i.d.* random variables when t is fixed, and the N trajectories $(X_j(t), 0 \leq t \leq T)$, $j = 1, \dots, N$ are *i.i.d.*. Nevertheless, differences between observations are due to the realization of both the W_j and ϕ_j . Indeed, the explicit solution of Equation (1) is given using Itô's formula, by

$$X_j(t) = X_j(0)e^{-t/\alpha} + \phi_j\alpha(1 - e^{-t/\alpha}) + \sigma e^{-t/\alpha} \int_0^t e^{s/\alpha} dW_j(s) \quad (2)$$

and for $t, t' \in [0, T]$,

$$\text{Cov}(X_j(t), X_j(t')) = \alpha^2(1 - e^{-t/\alpha})(1 - e^{-t'/\alpha})\text{Var}(\phi_j) + \frac{\alpha\sigma^2}{2}e^{-(t+t')/\alpha}(e^{\frac{2\min(t,t')}{\alpha}} - 1).$$

The purpose of the present work is to build nonparametric estimators of the random effect density f , considering that the processes are observed on $[0, T]$ with $T > 0$ given. In practice we consider discrete observations of the X_j 's with a very small time step δ . We are able to evaluate the error made by this discretization. The main difficulty is that we do not observe the ϕ_j 's but only the $X_j(k\delta)$'s. Thus the first step is to find an estimator of the random effects ϕ_j and then to estimate f , taking into account the approximation introduced by the estimation of the ϕ_j .

In the context of stochastic differential equations with random effects, Comte *et al.* (2013) propose different nonparametric estimators with good theoretical results. We focus on an estimator built by deconvolution and a kernel estimator. Both estimators have good theoretical properties assuming asymptotics on T . But these models are likely to be applied to experimental data with small T . This is especially the case for the neuronal data we are dealing with. Therefore, we adapt Comte *et al.* (2013) estimator to this small T framework. First we study a new estimator built by a deconvolution method (see Comte *et al.*, 2013; Butucea and Tsybakov, 2007, for example). The introduction of an additional new tuning parameter allows to deal with small T and to control the variance of the noise. We obtain a satisfying collection of estimators depending on two parameters. To select the final estimator among this collection, we extend the Goldenshluger and Lepski method for two dimensional model selection (Goldenshluger and Lepski, 2011). Finally we have a consistent estimator satisfying an oracle inequality, for any value of T .

Secondly, we consider a kernel estimator of f built from estimators of the random effects ϕ_j . This kernel estimator depends on a bandwidth to be chosen from the data. Several methods to select the bandwidth of kernel estimators are known. The originality here is the selection method of the bandwidth we set up: we use a method proposed by Goldenshluger and Lepski (2011) which provides an adaptive estimator. This kind of non-asymptotic result are not usual in this context.

We illustrate the properties of the proposed estimators with a simulation study. Especially, we compare them with standard bandwidth selection method of cross-validation type. Then, the estimators are applied to neuronal data (interspike interval (ISI) measures). Intracellular measures of the neuronal membrane potential between two spikes can be modelled with an Ornstein-Uhlenbeck model with one random effect (1). Each one can be considered as an independent experimental unit with a different realization of the random effect. This database has been surveyed with parametric strategies in Picchini *et al.* (2008) and Picchini *et al.* (2010), where it is assumed that the random

effects are Gaussian, and proved that the Ornstein-Uhlenbeck model with one random effect fits better the data than the model without them (by applying a likelihood ratio test). Our goal is to estimate nonparametrically the density of the random effect. This estimated density could be used in further works to model this phenomenon instead of the Gaussian density systematically.

The paper is organized as follows. In Section 2, we present a new nonparametric estimation method based on deconvolution strategy. Then, in Section 3 we propose a kernel estimator with an original method of bandwidth selection. In Section 4 we calibrate the selection methods and illustrate the good performances of the two estimators on simulated data. In Section 5 we experiment both estimators on the real database. We conclude this article with a discussion in Section 6. Most proofs are deferred to Appendix A, as well as the computation of the error made by discretization in Appendix B.

2 A nonparametric estimator built by deconvolution

2.1 Notations and assumptions

Let us introduce some notations. For two functions f and g in $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, the scalar product is defined by $\langle f, g \rangle = \int_{\mathbb{R}} f(x) \overline{g(x)} dx$ and the associated norm is $\|f\|^2 = \int_{\mathbb{R}} |f(x)|^2 dx$. The Fourier transform of f is $f^*(x) = \int_{\mathbb{R}} e^{ixu} f(u) du$ for all $x \in \mathbb{R}$. Then the convolution product of f and g for all $x \in \mathbb{R}$, is $f \star g(x) = \int_{\mathbb{R}} f(x-y) g(y) dy$. Finally we recall the Plancherel-Parseval's formula: $\forall f \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, $2\pi \|f\|^2 = \|f^*\|^2$.

We assume **(A)** $f \in \mathbb{L}^2(\mathbb{R})$, $f^* \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$.

2.2 Construction of the collection of estimators

The first step is to build estimators of the ϕ_j which are not observed, while we want to estimator their density. For this purpose, we introduce for $j = 1, \dots, N$ and $\tau \in]0, T]$, the following random variables

$$Z_{j,\tau} := \frac{X_j(\tau) - X_j(0) - \int_0^\tau (-\frac{X_j(s)}{\alpha}) ds}{\tau} = \phi_j + \frac{\sigma}{\tau} W_j(\tau). \quad (3)$$

The $(Z_{j,\tau})$ can be considered as estimators or approximations of the ϕ_j and the random variables $Z_{j,\tau} - \phi_j$ are centered. The $(Z_{j,\tau})_{j=1,\dots,N}$ are *i.i.d.* when τ is fixed, due to the independence of $(\phi_j)_{j=1,\dots,N}$ and $(W_j)_{j=1,\dots,N}$.

The deconvolution method is based on the decomposition (3). Indeed the two members of the sum are independent when τ is fixed which implies

$$f_{Z_\tau}(u) = f \star f_{\frac{\sigma}{\tau} W_j(\tau)}(u).$$

Taking the Fourier transform under assumption **(A)** gives the simple product

$$f_{Z_\tau}^*(u) = f^*(u) f_{\frac{\sigma}{\tau} W_j(\tau)}^*(u),$$

with $f_{\frac{\sigma}{\tau} W_j(\tau)}^*(u) = e^{-\frac{u^2 \sigma^2}{2\tau}}$. Thus $f^*(u) = f_{Z_\tau}^*(u) e^{u^2 \sigma^2 / 2\tau}$. Finally the Fourier inversion gives the closed formula, for all $x \in \mathbb{R}$,

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} f_{Z_\tau}^*(u) e^{\frac{u^2 \sigma^2}{2\tau}} du. \quad (4)$$

Then, we estimate $f_{Z_\tau}^*(u)$ by its empirical estimator $\widehat{f}_{Z_\tau}^*(u) = (1/N) \sum_{j=1}^N e^{iuZ_{j,\tau}}$. However, plugging this in formula (4) involves integrability problems. Indeed the integrability of $\widehat{f}_{Z_\tau}^*(u) e^{u^2 \sigma^2 / 2\tau}$ is not ensured. Therefore, we have to introduce a cut-off. The first idea due to Comte *et al.* (2013) is to link the time of the process τ and the cut-off:

$$\widehat{f}_\tau(x) = \frac{1}{2\pi} \int_{-\sqrt{\tau}}^{\sqrt{\tau}} e^{-iux} \frac{1}{N} \sum_{j=1}^N e^{iuZ_{j,\tau}} e^{\frac{u^2 \sigma^2}{2\tau}} du. \quad (5)$$

Nevertheless, when τ is small (which is the case for the real database we investigate), the integration domain is not large enough, and the estimators of f are not satisfactory. To improve the previous estimator, we introduce a new parameter s in the cut-off:

$$\hat{f}_{s,\tau}(x) = \frac{1}{2\pi} \int_{-s\sqrt{\tau}}^{s\sqrt{\tau}} e^{-iux} \frac{1}{N} \sum_{j=1}^N e^{iuZ_{j,\tau}} e^{\frac{u^2\sigma^2}{2\tau}} du.$$

Then to simplify the theoretical study, we replace $s\sqrt{\tau}$ in the integral by a new parameter m . The resulting estimator $\tilde{f}_{m,s}$ is defined when $m^2/s^2 \in]0, T]$, by

$$\tilde{f}_{m,s}(x) = \frac{1}{2\pi} \int_{-m}^m e^{-iux} \frac{1}{N} \sum_{j=1}^N e^{iuZ_{j,m^2/s^2}} e^{\frac{u^2\sigma^2 s^2}{2m^2}} du \quad (6)$$

with m and s in two finite sets \mathcal{M} and \mathcal{S} that we precise later.

Let us emphasize that this estimator $\tilde{f}_{m,s}$ specifically depends on two parameters which have to be selected from the data. This is not usual in the deconvolution setting, where only one cut-off parameter is often introduced. The selection of these two parameters (m, s) among the finite sets \mathcal{M}, \mathcal{S} is thus more difficult. It is even more challenging here because the cut-off m appears both in the integral and in the integrand. But this will induce gains in the rates of the estimators. Before proposing a selection method of (m, s) we start by evaluating the quality of the estimator with the mean integrated squared error (MISE):

$$\mathbb{E} \left[\|\tilde{f}_{m,s} - f\|^2 \right] = \|f - \mathbb{E}[\tilde{f}_{m,s}]\|^2 + \mathbb{E} \left[\|\tilde{f}_{m,s} - \mathbb{E}[\tilde{f}_{m,s}]\|^2 \right].$$

It is interesting to note that $\mathbb{E}[\tilde{f}_{m,s}] = f_m$ where f_m is defined by its Fourier transform

$$f_m^* := f^* \mathbf{1}_{[-m,m]}.$$

It means that the bias does not depend on s . We obtain the following bound on the MISE of $\tilde{f}_{m,s}$.

Proposition 1. *Under (A), the estimator $\tilde{f}_{m,s}$ given by (6) is an unbiased estimator of f_m and we have*

$$\mathbb{E} \left[\|\tilde{f}_{m,s} - f\|^2 \right] \leq \|f_m - f\|^2 + \frac{m}{\pi N} \int_0^1 e^{\sigma^2 s^2 v^2} dv. \quad (7)$$

The proofs are relegated in Section A. The first term of the bound (7) is the bias term. It represents the error resulting from estimating f by f_m and it decreases when m increases, indeed:

$$\|f_m - f\|^2 = \frac{1}{2\pi} \int_{|u| \geq m} |f^*(u)|^2 du.$$

The second term is the variance term, and increases with m and s . This implies that the couple (m, s) minimizing the MISE is the one which realizes a compromise between the two terms.

Now we specify the two sets \mathcal{M} and \mathcal{S} . We notice that the quality of the estimate in the Fourier domain is good on an interval around zero with length related with σ . We choose the following set for s ,

$$\mathcal{S} := \left\{ s_l = \frac{1}{2^l} \frac{2}{\sigma}, 1/2^{P-1} \leq \sigma s_l \leq 2, l = 0, \dots, P \right\}.$$

Moreover with this chosen collection \mathcal{S} , the order of the variance term is m/N . Thus with the idea that m^2/s^2 is homogeneous to a time, we choose m in the finite collection:

$$\mathcal{M} := \left\{ m = \frac{\sqrt{k\Delta}}{\sigma}, k \in \mathbb{N}^*, 0 < m \leq N \right\}$$

with $0 < \Delta < 1$ a small step to be fixed. Finally the collection of couples of parameters is

$$\mathcal{C} := \{(m, s) \in \mathcal{M} \times \mathcal{S}, m^2/s^2 \leq T\}.$$

The final estimator is the best estimator $\tilde{f}_{m,s}$ from the collection $(m, s) \in \mathcal{C}$. Choose the best estimator is not an easy task except if we know the regularity of f . For example, let us assume that f is in the Sobolev space with regularity parameter b , i.e. f belongs to the set $\mathcal{A}_b(L)$ defined by

$$\mathcal{A}_b(L) = \{f \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R}), \int_{\mathbb{R}} |f^*(x)|^2 (1+x^2)^b dx \leq L\}$$

with $b > 0$, $L > 0$. In this situation, the bias term satisfies:

$$\|f_m - f\|^2 = \frac{1}{2\pi} \int_{|u| \geq m} |f^*(u)|^2 du \leq \frac{L}{2\pi} m^{-2b}.$$

Consequently, the \mathbb{L}^2 -risk of $\tilde{f}_{m,s}$ is bounded by,

$$\mathbb{E}[\|\tilde{f}_{m,s} - f\|^2] \leq \frac{L}{2\pi} m^{-2b} + \frac{m}{\pi N} e^{\sigma^2 s^2}.$$

Therefore, the best theoretical choice of s is s_P the smallest s in our collection, and

$$m = m^* = K_b N^{\frac{1}{2b+1}}$$

with $K_b = (bL \exp(-1/(2^{2(P-1)})))^{1/(2b+1)}$. Then we obtain the following asymptotic result.

Corollary 2. *If $f \in \mathcal{A}_b(L)$, and if we choose $s = s_P$ and $m = m^*$, there exists a constant K depending on b, L, P , such that*

$$\mathbb{E}[\|\tilde{f}_{m^*,s_P} - f\|^2] \leq KN^{-\frac{2b}{2b+1}}.$$

The order of the risk in this case is $N^{-2b/(2b+1)}$ for a large N , it is the nonparametric estimation rate of convergence obtained when the observation are N realizations of the variable of interest. Nevertheless, this choice is theoretical because it depends on the regularity b of f which is unknown. The next section provides a data-driven method to select (m, s) .

2.3 Selection of the final estimator

If we look at the risk bound (7), a natural idea is to choose the smallest s and then to select m . But this strategy does not work in practice, thus we need to find a data driven selection method for s . Besides, if we choose m first, the selected m depends on s and then we do not have any criterion to select s . Hence, we select both m and s simultaneously in a data driven way. This is a crucial issue. Indeed, the role of the two parameters is not the same. Thus we propose a new criterion adapted from the Goldenshluger and Lepski (2011) method.

The idea is to select the couple which minimizes the MISE: $\mathbb{E}[\|\tilde{f}_{m,s} - f\|^2]$. As it is unknown, we have to find a computable approximation of this quantity. We define the best couple (m, s) as the one minimizing a criterion defined as the sum of a bias term and a variance term called penalty. We define the penalty function, which has the same order as the variance term (in the risk bound):

$$\text{pen}(m, s) = \kappa \frac{m}{N} e^{\sigma^2 s^2},$$

where κ is a numerical constant to be calibrated. Note that for $m \in \mathcal{M}$ and $s \in \mathcal{S}$, the penalty function is bounded.

To estimate the bias term, we generalize Goldenshluger and Lepski's criterion for a two dimensional index. The method is inspired by the ideas developed for kernel estimators by Goldenshluger and Lepski (2011) and adapted to model selection in one dimension in Comte and Johannes (2012) and in two dimensions by Chagny (2013). The idea is to estimate $\|f - f_m\|^2$ by the \mathbb{L}^2 -distance between two estimators defined in (6). But this induces a bias which has to be corrected by the penalty function. We consider the following estimator of the bias, with $(m', s') \wedge (m, s) := (m' \wedge m, s' \wedge s)$,

$$\Gamma_{m,s} = \max_{(m',s') \in \mathcal{C}} \left(\|\tilde{f}_{m',s'} - \tilde{f}_{(m',s') \wedge (m,s)}\|^2 - \text{pen}(m', s') \right)_+$$

for $(m, s) \in \mathcal{C}$. Finally the selected couple is:

$$(\tilde{m}, \tilde{s}) = \arg \min_{(m,s) \in \mathcal{C}} \{\Gamma_{m,s} + \text{pen}(m, s)\}. \quad (8)$$

Now we are able to prove the following result.

Theorem 3. *Under (A), consider the estimator $\tilde{f}_{\tilde{m}, \tilde{s}}$ given by (6) and (8). There exists κ_0 a numerical constant such that, for all penalty constant $\kappa \geq \kappa_0$,*

$$\mathbb{E}[\|\tilde{f}_{\tilde{m}, \tilde{s}} - f\|^2] \leq C \inf_{(m,s) \in \mathcal{C}} \{\|f - f_m\|^2 + \text{pen}(m, s)\} + \frac{C'(P+1)}{N} \quad (9)$$

where $C > 0$ is a numerical constant as soon as κ is fixed and C' is a constant depending on $\|f\|$, σ , Δ , and $P+1$ the cardinality of \mathcal{S} .

Inequality (9) means that $\tilde{f}_{\tilde{m}, \tilde{s}}$ automatically makes the bias-variance trade-off. Moreover, our result is of non asymptotic nature on N and of oracle type.

One should notice that this new parameter s generalizes the results of Comte *et al.* (2013) even if T is large. We choose the two parameters in an adaptive way, thus this gives more flexibility in the choice of the estimator.

The numerical constant κ is calibrated by simulation experiments. From the theoretical part we get $\kappa_0 = 24$ which is much too large in practice. Besides the cardinality P of the set \mathcal{S} is chosen small in practice ($P = 3$).

In the following, we study another nonparametric estimator of the density f . Indeed, it is a natural idea to define a kernel estimator from the estimators $Z_{j,\tau}$ given in (3).

3 Kernel estimator

First we present the kernel estimator proposed by Comte *et al.* (2013). Then we recall the MISE bound they prove. Finally we propose a new procedure to select the bandwidth.

We focus only on $Z_{j,T}$ among the $(Z_{j,\tau})_{[0,\tau]}$ because when T is large $Z_{j,T}$ clearly approximates ϕ_j without needing to remove the noise as for the deconvolution estimator. The idea of the kernel estimator is thus to use directly the $Z_{j,T}$ as an approximation of the non-observed random effects ϕ_j to estimate their density f .

Thus, for $j = 1, \dots, N$, we consider the $Z_{j,T}$ defined by (3) when $\tau = T$. These N random variables are *i.i.d.* and the resulting kernel estimator of the density of the ϕ_j 's is given for all $x \in \mathbb{R}$, by

$$\hat{f}_h(x) = \frac{1}{N} \sum_{j=1}^N K_h(x - Z_{j,T}) \quad (10)$$

where $h > 0$ is a bandwidth, and the kernel $K : \mathbb{R} \rightarrow \mathbb{R}$ is a \mathcal{C}^2 function such that

$$\int K(u)du = 1, \quad \|K\|^2 = \int K^2(u)du < +\infty, \quad \int (K''(u))^2 du < +\infty. \quad (11)$$

For $h > 0$ we define $f_h := K_h \star f$ with

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right).$$

We denote for all $p \in \mathbb{R}$, $\|f\|_p = (\int |f(x)|^p dx)^{1/p}$ and for $p = 2$ we still use $\|f\|_2 = \|f\|$. Let us notice the relations: $\|K\| = \|K_h\|/\sqrt{h}$ and $\|K_h\|_1 = \|K\|_1$. We recall the result proved in Comte *et al.* (2013) for the MISE.

Proposition 4. *Consider estimator \hat{f}_h given by (10), we have*

$$\mathbb{E}[\|\hat{f}_h - f\|^2] \leq 2\|f - f_h\|^2 + \frac{\|K\|^2}{Nh} + \frac{\sigma^4 \|K''\|^2}{3T^2 h^5}. \quad (12)$$

This bound with three terms comes from the bias-variance decomposition. Indeed:
 $\|\mathbb{E}[\widehat{f}_h] - f\|^2 \leq 2\|f - f_h\|^2 + 2\|\mathbb{E}[\widehat{f}_h] - f_h\|^2$, with

$$\|\mathbb{E}[\widehat{f}_h] - f_h\|^2 \leq \frac{\sigma^4 \|K''\|^2}{3T^2 h^5} \quad (13)$$

(see the proof in Comte *et al.* (2013)). In the bound (12), the first term is a bias term, which decreases when h decreases. The second term is the term of variance which increases when h decreases. Finally, the third term is an unusual error term due to the approximation of the ϕ_j 's by the $Z_{j,T}$ also increasing when h decreases.

Now we focus on how to choose the bandwidth from the data. The best choice of h is the one which minimizes the sum of these three terms. The selection of the bandwidth can be done for example using cross validation, see *e.g.* the R-function *density* which is commonly used. However, the only theoretical results known for cross-validation procedure are asymptotic and to the best of our knowledge there is no adaptive result on the final estimator. In the present work, we propose to adapt a selection method due to Goldenshluger and Lepski (2011) mentioned before, which provides a data driven bandwidth for which we provide non-asymptotic theoretical results.

We denote $\mathcal{H}_{N,T}$ the set of bandwidths h , to be defined later. The best theoretical choice of the bandwidth is the h which minimizes the bound on the MISE given by (12). Nevertheless, in practice, this bound is unknown, and has to be estimated. We denote by V the estimator of the two terms of variance

$$V(h) = \kappa_1 \frac{\|K\|_1^2 \|K\|^2}{Nh} + \kappa_2 \frac{\sigma^4 \|K\|_1^2 \|K''\|^2}{3T^2 h^5}. \quad (14)$$

The estimator of the bias term is built as in Section 2.3:

$$A(h) = \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\widehat{f}_{h,h'} - \widehat{f}_h\|^2 - V(h') \right)_+ \quad (15)$$

with

$$\widehat{f}_{h,h'}(x) := K_{h'} \star \widehat{f}_h(x) = \frac{1}{N} \sum_{j=1}^N K_{h'} \star K_h(x - Z_{j,T}).$$

The bandwidth \widehat{h} is selected as follows

$$\widehat{h} = \operatorname{argmin}_{h \in \mathcal{H}_{N,T}} (A(h) + V(h)) \quad (16)$$

with $\mathcal{H}_{N,T}$ a discrete set of bandwidth h such that $h > 0$, $\frac{1}{Nh} \leq 1$, $\frac{1}{h^5 T^2} \leq 1$, $\forall c > 0, \exists \Sigma(c) < \infty$, $\sum_{h \in \mathcal{H}_{N,T}} h^{-1/2} e^{-c/\sqrt{h}} \leq \Sigma(c)$. Then we can prove the following Theorem.

Theorem 5. Consider estimator \widehat{f}_h given by (10). Then,

$$\mathbb{E}[\|\widehat{f}_h - f\|^2] \leq C_1 \inf_{h \in \mathcal{H}_{N,T}} \left\{ \|f - f_h\|^2 + \frac{\|K\|_1^2 \|K\|^2}{Nh} + \frac{\sigma^4 \|K\|_1^2 \|K''\|^2}{3T^2 h^5} \right\} + \frac{C_2}{N}$$

where C_1, C_2 are two positive constants such that C_1 depends on $\|K\|_1$ and C_2 depends on $\|f\|, \|K\|, \|K\|_1, \|K\|_{4/3}$.

The latest result is an oracle inequality: the bias variance compromise is automatically obtained and in a data-driven and non-asymptotic way.

It is important to notice that if we take for example $\mathcal{H}_{N,T} = \{1/k, k = 1, \dots, N\}$ then the sum $\sum_{h \in \mathcal{H}_{N,T}} h^{-1/2} e^{-c/\sqrt{h}}$ converges, which is a necessary condition for the proof.

4 Simulated study

In the following we compare on simulations the two presented procedures, and we compare our bandwidth selection method with the cross-validation estimator from the R-function *density*.

We simulate data by computing the exact solution of (1) given by formula (2) at discrete times $t_k \in \mathcal{T} := \{k\delta, k \in \{0, \dots, J\}, J\delta = T\}$. For the simulation study, we have to fix $N, \delta, T, \sigma, \alpha$, and the density f . In the following we fix $\alpha = 0.039$ (which is the value obtained in Lansky *et al.* (2006) for the real database, see Section 5). Then we take $\sigma = 0.0135$ or $\sigma = 0.05$. For the time T , we choose $T = 0.3, 10, 100$ with different values of δ the discrete time step at which observations are recorded. Thus the value of J , the number of observations for one trajectory changes between 150 and 5000. All these parameter values are chosen in relation with the parameters of the real database. In this study we hope to highlight the influence of each one. For f , we investigate four different distributions:

- Gaussian distribution $\mathcal{N}(0.278, (0.041)^2)$
- Gamma distribution $\Gamma(1.88, 0.148)$
- mixed Gaussian distribution $0.3\mathcal{N}(0, (0.02)^2) + 0.7\mathcal{N}(1, (0.02)^2)$
- mixed Gamma distribution $0.4\Gamma(3, 0.08) + 0.6\Gamma(30, 0.035)$.

First, we implement the two adaptive estimators: $\tilde{f}_{\tilde{m}, \tilde{s}}$ and $\hat{f}_{\hat{h}}$. We begin by computing the random variables used by both estimators: $Z_{j, \tau}$ given by (3), with Riemann sums approximations (see Appendix B for details). Furthermore, for the kernel estimator given by (10), we choose a Gaussian kernel: $K(u) = (1/\sqrt{2\pi})e^{-u^2/2}$. In this case $\|K\|_1 = 1$, $\|K\|_2^2 = 1/(2\sqrt{\pi})$, $\|K''\|_2^2 = (1 + 1/\sqrt{2})/(\sqrt{2\pi})$. Then, the selected bandwidth \hat{h} is given by Equation (16). Note that for all $(h, h') \in \mathcal{H}^2$,

$$K_{h'} \star K_h(x) = \frac{1}{\sqrt{2\pi}\sqrt{h'^2 + h^2}} e^{-x^2/[2(h'^2 + h^2)]}.$$

We use this relation to compute the $\hat{f}_{h, h'}$. For the deconvolution estimator given by (6) we also use Riemann sums to compute the integral. For the collection of m , we choose $\Delta = \delta$ except when $T = 100$ and $\sigma = 0.0135$ for which we choose $\Delta = \delta/100$ so that the collection is not empty and the values are smaller.

Secondly, we have to calibrate the penalty constants: κ for the deconvolution estimator, and κ_1, κ_2 for the kernel estimator. Classically, the constants are fixed thanks to preliminary simulation experiments. Different functions f have been investigated with different parameter values, and a large number of repetitions. Comparing the MISE obtained as functions of the constants $\kappa, \kappa_1, \kappa_2$ yields to select values making a good compromise over all experiences. Finally we choose $\kappa = 0.3$, $\kappa_1 = 1$, and $\kappa_2 = 0.0001$.

We represent 25 estimators $\tilde{f}_{\tilde{m}, \tilde{s}}$ on Figure 1, and 25 estimators $\hat{f}_{\hat{h}}$ on Figure 2, with the 4 investigated densities f . The beams of estimators are close to the estimated density.

In order to evaluate the performances of each estimator on the different designs, we compare their empirical MISE computed from 100 simulated data sets. Results are presented in Table 1 when $N = 240$. Note that the MISE of the two estimators are close to each other. However we can point out some differences. The first row of the Table corresponds to simulation with the parameters of the real database. In the first column, the Gaussian case, the MISE are 10 times larger than the ones for other cases. This can be easily explained: the value of the estimated density are 10 times larger than others. Nevertheless, on lines 3 and 4 for the Gaussian case, the MISE are very large. This is due to the bad estimation of the ϕ_j by the $Z_{j, T}$ with $\sigma = 0.05$ and $T = 0.3$ ¹. The quality of the estimation is significantly better when we jump to a $\mathcal{N}(0.278, 0.2)$ (0.2 is the variance of the mixed Gaussian for example). In general one can notice that when σ is larger than the standard deviation of the density of the random effects f , the estimation is less precise, which is coherent in term of signal to noise ratio.

¹We insist that this bad estimation is not due to the fact the noise is Gaussian. Indeed even if Fan (1991) proves the rates to be logarithmic in that case, the rates are improved and can be polynomial when the density under estimation is of the same type of the noise (see Lacour (2006), Comte *et al.* (2006)).

Table 1: Empirical MISE computed from 100 simulated data sets, with $N = 240$, various T , δ , σ for the kernel estimator \hat{f}_h and the deconvolution estimator $\tilde{f}_{\tilde{m},\tilde{s}}$

Case	T	δ	σ	Estimator	f			
					Gaussian	gamma	mixed Gaussian	mixed gamma
1	0.3	0.0002	0.0135	\hat{f}_h	0.234	0.034	0.015	0.026
				$\tilde{f}_{\tilde{m},\tilde{s}}$	0.381	0.072	0.021	0.038
2	0.3	0.002	0.0135	\hat{f}_h	0.240	0.034	0.015	0.025
				$\tilde{f}_{\tilde{m},\tilde{s}}$	1.584	0.071	0.019	0.038
3	0.3	0.0002	0.05	\hat{f}_h	2.535	0.095	0.025	0.054
				$\tilde{f}_{\tilde{m},\tilde{s}}$	3.402	0.231	0.078	0.226
4	0.3	0.002	0.05	\hat{f}_h	2.524	0.093	0.026	0.053
				$\tilde{f}_{\tilde{m},\tilde{s}}$	3.269	0.154	0.029	0.232
5	10	0.002	0.0135	\hat{f}_h	0.096	0.036	0.016	0.024
				$\tilde{f}_{\tilde{m},\tilde{s}}$	0.080	0.037	0.035	0.052
6	10	0.02	0.0135	\hat{f}_h	0.105	0.032	0.015	0.023
				$\tilde{f}_{\tilde{m},\tilde{s}}$	0.151	0.048	0.033	0.034
7	10	0.002	0.05	\hat{f}_h	0.111	0.037	0.014	0.026
				$\tilde{f}_{\tilde{m},\tilde{s}}$	0.687	0.061	0.016	0.034
8	10	0.02	0.05	\hat{f}_h	0.121	0.037	0.016	0.026
				$\tilde{f}_{\tilde{m},\tilde{s}}$	0.181	0.053	0.023	0.040
9	100	0.2	0.0135	\hat{f}_h	0.106	0.032	0.015	0.024
				$\tilde{f}_{\tilde{m},\tilde{s}}$	0.123	0.062	0.091	0.046

Table 1 does not prove any clear influence of T as shows case 9. If J is large enough, meaning if δ is small enough, (which is the case even for $J = 150$ when $T = 0.3$) the deconvolution estimator fits well the density. In practice, when T increases, the selected value of s decreases, which could have been predicted. The results are still satisfying for large T . For the kernel estimator, although the theoretical condition $1/h^5 < T^2$ is not satisfied, the numerical results are good. Besides, it is obvious on simulations that the larger N the better the estimator \hat{f}_h . However, the role of N is not as clear for the deconvolution estimator.

The main difference between our two estimators \hat{f}_h and $\tilde{f}_{\tilde{m},\tilde{s}}$ is the computation time: a few seconds for the first one and ten minutes for the second one. The kernel estimator is much faster than the deconvolution one.

We also compute the MISE for the kernel estimator obtained by cross-validation with the R-function *density*. We explored the same scenarios as on Table 1. We can conclude from this study that both kernel estimators are really close most of the time. Thus the results are not shown. Nevertheless it appears that the R-function fits slightly better the Gaussian or Gamma densities, while our estimator \hat{f}_h fits better mixture densities in general (case 4 for example).

5 Application to neuronal data

5.1 Database

Data are measurements along time of the membrane potential in volts of one single neuron of a pig between the spikes (see Yu *et al.*, 2004; Lansky *et al.*, 2006, for details on data acquisition). In this neuronal context, between the $(j - 1)$ th and the j th spike, the depolarization of the membrane

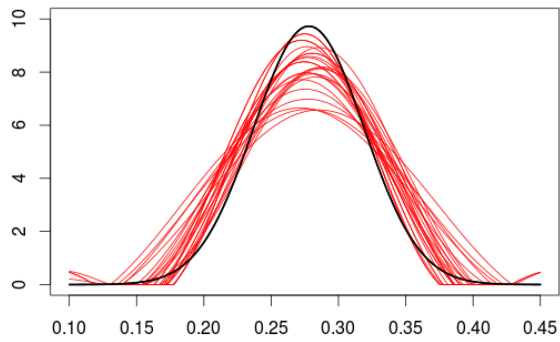
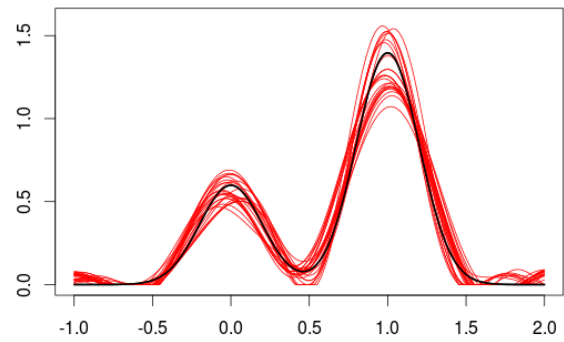
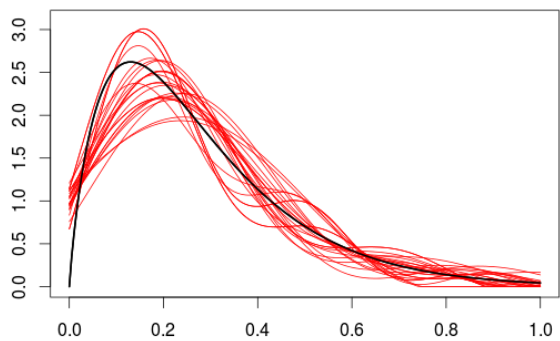
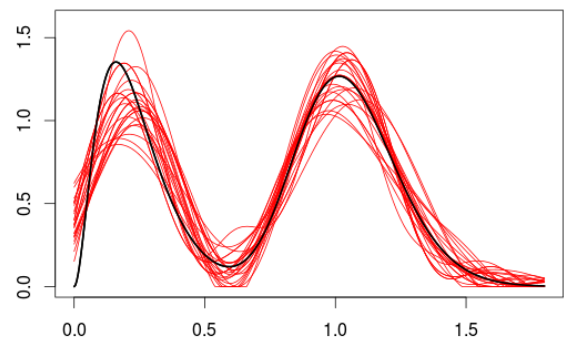
(a) f Gaussian(b) f mixed Gaussian(c) f Gamma(d) f mixed Gamma

Figure 1: Simulated data. In red 25 estimators $\tilde{f}_{\tilde{m}, \tilde{s}}$ with parameters: $N = 240$, $T = 0.3$, $\delta = 0.00015$, $\sigma = 0.0135$, $\alpha = 0.039$ and the bold black line the true density f

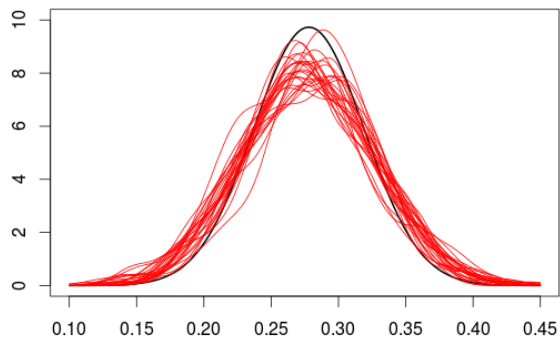
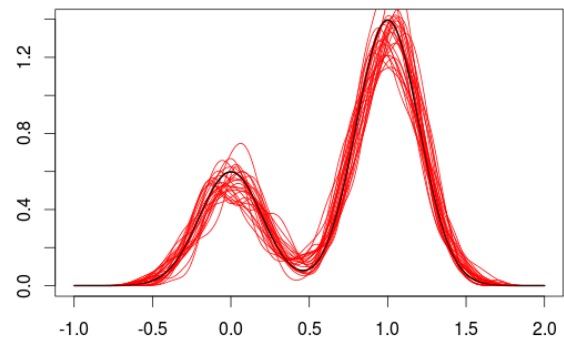
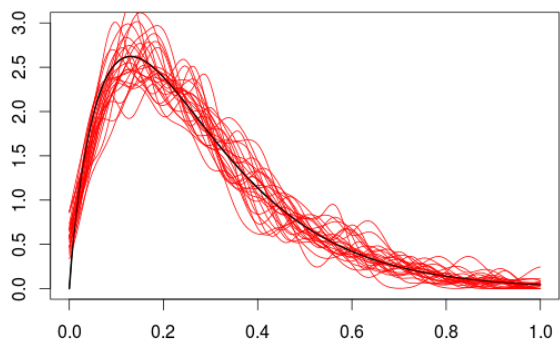
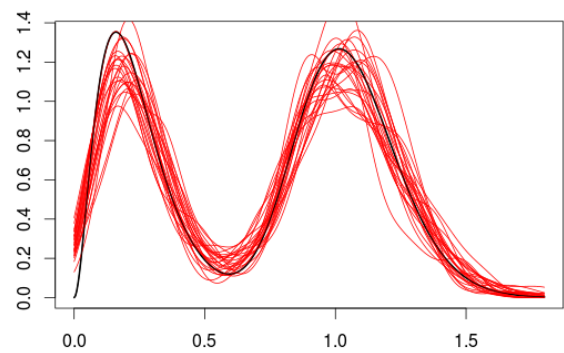
(a) f Gaussian(b) f mixed Gaussian density(c) f Gamma(d) f mixed Gamma density

Figure 2: Simulated data. In red 25 estimators \hat{f}_h with parameters: $N = 240$, $T = 0.3$, $\delta = 0.00015$, $\sigma = 0.0135$, $\alpha = 0.039$ and the bold black line the density true density f

potential receiving a random input, can be described by the Ornstein Uhlenbeck model with one random effect (1). The spikes are not intrinsic to the model but are generated when the voltage reaches for the first time a certain threshold S , then the process is reset. Thus each trajectory is observed on an interval $[0, T_j]$ where $T_j = \inf\{t > 0, X_j(t) \geq S\}$. The initial voltage (the value following a spike) is assumed to be equal to the resting potential and set to zero: $x_j = 0$.

The positive constant parameter α [s] is called the time constant of the neuron (the coefficient of decay in the exponential, when there is no noise). Lansky *et al.* (2006) estimate this parameter on the same database by a regression method based on the first moment of the stochastic process (when the constant diffusion coefficient is fixed $\sigma = 0.0135$ [V/ \sqrt{s}]), they obtain $\alpha = 0.039$ [s]. We keep this value. The ϕ_j represents the local average input that the neuron receives during the j^{th} ISI. We assume that ϕ_j changes from one ISI to another because of other neurons or environment influence, for example. So parameters ϕ and σ characterize the input, while α , x_j (the resting potential), and S (the firing threshold) describe the neuron irrespectively of the incoming signal (Picchini *et al.*, 2008).

Data are composed of $N = 312$ ISIs. For each interval $[0, T_j]$ the time step is the same: $\delta = 0.00015$ [s]. We decide to keep only realizations with more than 2000 observations ($T_j/\delta \geq 2000$). Finally we have $N = 240$ realizations with $J = 2000$ observations and for $j = 1, \dots, N$, $T = T_j = 0.3$ [s]. Also the data are normalized in order to begin with zero at the initial time.

The study of the units of measurement can highlight the collections given in Section 2.2. One can notice that the unit of measurement of u in the integrand must be [s/V] (same unit as $1/Z_{j,\tau}$) such that the exponential terms are without unit. The unit of s is [\sqrt{s}/V], and the choice of \mathcal{M} with the same unit as u seems natural.

In the following we compare our results to the estimation obtained in Picchini *et al.* (2010) under the parametric Gaussian assumption.

5.2 Comparison of estimators

The estimation of the density f obtained by Picchini *et al.* (2010) under the Gaussian assumption on this database are $\mathcal{N}(\mu = 0.278, \eta^2 = (0.041)^2)$. The mean and the standard deviation we obtain using a maximum-likelihood estimator on the $(Z_{j,T})$'s are: 0.270 and 0.046. Thus these two estimations are close to the one of Picchini *et al.* (2010). We use our two nonparametric estimators to see how close to a Gaussian density they are.

On Figure 3 we represent both estimators \hat{f}_h and $\tilde{f}_{\tilde{m}, \tilde{s}}$ applied to the real database and the density $\mathcal{N}(\mu, \eta^2)$. The two estimations are close, and close to the estimation of Picchini *et al.* (2010). However, it is also legitimate to think about a Gamma distribution to model the random parameters ϕ_j 's because it is a non negative local average input. Thus, a Gamma distribution may seem more appropriate than a Gaussian distribution, even if the chosen Gaussian has small probability to be negative. We look for the Gamma distribution which has for mean $\mu = 0.278$ and for variance $\eta = 0.041$. This distribution is $\Gamma(a = 46.3, s = 0.006)$ with associated density $f_{a,s}(x) = x^{a-1}e^{-x/s}/(\Gamma(a)s^a)$. We notice the similarity between the previous Gaussian curve and the new one. Thus this distribution seems also suitable to fit the distribution of the ϕ_j 's as shows Figure 3.

The Gaussian assumption is strong and leads to parametric tractable models. The present work confirms that this approximation is acceptable. However, the nonparametric estimation gives a density for the ϕ_j 's that can be used to simulate the random effect and could be closer to the true one.

Notice that, as mentioned in introduction, Comte *et al.* (2013)'s estimator cannot handle small values of T while our new proposals are successful in such case. One can wonder if the new estimators are robust when increasing T . Indeed, our method works for larger T . Precisely changing volts in millivolts and seconds in milliseconds implies $T = 300$ and $\sigma = 0.426$, and on simulated data with $\Delta = \delta$, we reconstruct well the shape of the density.

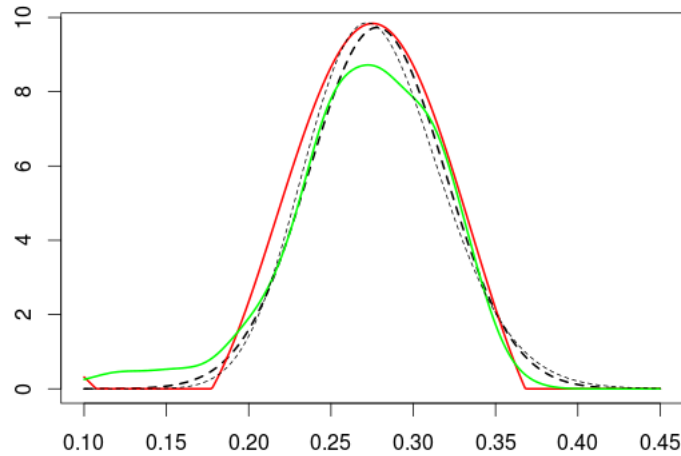


Figure 3: Real data. In green estimator \hat{f}_h , in red $\tilde{f}_{\tilde{m},\tilde{s}}$, the black dotted and bold line the density $\mathcal{N}(\mu, \eta^2)$ from Picchini *et al.* (2010) and the black dotted thin line the density $\Gamma(46.3, 0.006)$

6 Discussion

In this work we study a stochastic differential Ornstein-Uhlenbeck mixed-effects model. We propose two estimators of the density of the random effect. Both estimators are not very sensitive to the effect of the time of observation T . Both are data-driven and satisfy inequality of oracle type.

Besides we study neuronal data (ISI) with nonparametric estimation strategy. Instead of setting parametric assumption for the random effect distribution, we build an estimation of its density. Future works based on this estimation could be more precise and closer to the real neuronal data. The model can be improved by adding another random effect: the time constant of the neuron. This will be surveyed in forgoing works.

Acknowledgements

I would like to thank Fabienne Comte and Adeline Samson for very useful discussions and advice.

A Appendix

A.1 Talagrand's inequality

The following result follows from the Talagrand concentration inequality given in Klein and Rio (2005) and arguments in Birgé and Massart (1998).

Theorem 6. Consider $n \in \mathbb{N}^*$, \mathcal{F} a class at most countable of measurable functions, and $(X_i)_{i \in \{1, \dots, N\}}$ a family of real independent random variables. One defines, for all $f \in \mathcal{F}$,

$$\nu_N(f) = \frac{1}{N} \sum_{i=1}^N (f(X_i) - \mathbb{E}[f(X_i)]).$$

Supposing there are three positive constants M , H and v such that $\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M$,

$\mathbb{E}[\sup_{f \in \mathcal{F}} |\nu_N f|] \leq H$, and $\sup_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N \text{Var}(f(X_i)) \leq v$, then for all $\alpha > 0$,

$$\mathbb{E} \left[\left(\sup_{f \in \mathcal{F}} |\nu_N(f)|^2 - 2(1 + 2\alpha)H^2 \right)_+ \right] \leq \frac{4}{K_1} \left(\frac{v}{N} \exp \left(-K_1 \alpha \frac{NH^2}{v} \right) + \frac{49M^2}{K_1 s^2(\alpha) N^2} \exp \left(-\frac{\sqrt{2} K_1 C(\alpha) \sqrt{\alpha} NH}{7M} \right) \right)$$

with $C(\alpha) = (\sqrt{1 + \alpha} - 1) \wedge 1$, and $K_1 = \frac{1}{6}$.

A.2 Young inequality

This inequality can be found in Briane and Pagès (2006) for example.

Theorem 7. Let f be a function belonging to $\mathbb{L}^p(\mathbb{R})$ and g belonging to $\mathbb{L}^q(\mathbb{R})$, let p, q, r be real numbers in $[1, +\infty]$ and such that

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1.$$

Then,

$$\|f \star g\|_r \leq \|f\|_p \|g\|_q.$$

A.3 Proof of Proposition 1

The bias term is $\|f - \mathbb{E}[\tilde{f}_{m,s}]\|^2$. Let us compute $\mathbb{E}[\tilde{f}_{m,s}]$. As the $Z_{j,\tau}$ are *i.i.d.* when τ is fixed and due to the independence of ϕ_1 and W_1 , we obtain:

$$\begin{aligned} \mathbb{E}[\tilde{f}_{m,s}(x)] &= \frac{1}{2\pi} \int_{-m}^m e^{-iux} \mathbb{E} \left[e^{iuZ_{1,m^2/s^2} + u^2 \sigma^2 s^2 / (2m^2)} \right] du \\ &= \frac{1}{2\pi} \int_{-m}^m e^{-iux} \mathbb{E} \left[e^{iu\phi_1 + iu\sigma W_1(m^2/s^2)s^2/m^2 + u^2 \sigma^2 s^2 / (2m^2)} \right] du \\ &= \frac{1}{2\pi} \int_{-m}^m e^{-iux + u^2 \sigma^2 s^2 / (2m^2)} f^*(u) \mathbb{E} \left[e^{iu\sigma W_1(m^2/s^2)s^2/m^2} \right] du \\ &= \frac{1}{2\pi} \int_{-m}^m e^{-iux + u^2 \sigma^2 s^2 / (2m^2)} f^*(u) e^{-u^2 \sigma^2 s^2 / (2m^2)} du \\ &= \frac{1}{2\pi} \int_{-m}^m e^{-iux} f^*(u) du =: f_m(x). \end{aligned}$$

Therefore this gives $\mathbb{E}[\tilde{f}_{m,s}(x)] = f_m(x)$, and $\|f - \mathbb{E}[\tilde{f}_{m,s}]\|^2 = \|f - f_m\|^2 = \frac{1}{2\pi} \int_{|u| \geq m} |f^*(u)|^2 du$. The variance term is:

$$\begin{aligned} \mathbb{E} \left[\|\tilde{f}_{m,s} - f_m\|^2 \right] &= \frac{1}{2\pi} \mathbb{E} \left[\int_{-m}^m \left| \frac{1}{N} \sum_{j=1}^N e^{iuZ_{j,m^2/s^2}} e^{\frac{u^2 \sigma^2 s^2}{2m^2}} - f^*(u) \right|^2 du \right] \\ &= \frac{1}{2\pi N} \int_{-m}^m e^{\frac{u^2 \sigma^2 s^2}{m^2}} \text{Var} \left(e^{iuZ_{1,m^2/s^2}} \right) du \\ &\leq \frac{1}{2\pi N} \int_{-m}^m e^{\frac{u^2 \sigma^2 s^2}{m^2}} du = \frac{m}{\pi N} \int_0^1 e^{s^2 \sigma^2 v^2} dv. \quad \square \end{aligned}$$

A.4 Proof of Theorem 3

Let us study the term $\|\tilde{f}_{\tilde{m}, \tilde{s}} - f\|^2$. We decompose it into a sum of three terms and the definition of (\tilde{m}, \tilde{s}) (8) implies for all $(m, s) \in \mathcal{C}$

$$\begin{aligned} \|\tilde{f}_{\tilde{m}, \tilde{s}} - f\|^2 &\leq 3 \left(\|\tilde{f}_{\tilde{m}, \tilde{s}} - \tilde{f}_{(\tilde{m}, \tilde{s}) \wedge (m, s)}\|^2 + \|\tilde{f}_{(\tilde{m}, \tilde{s}) \wedge (m, s)} - \tilde{f}_{m, s}\|^2 + \|\tilde{f}_{m, s} - f\|^2 \right) \\ &\leq 3(\Gamma_{m, s} + \text{pen}(\tilde{m}, \tilde{s})) + 3(\Gamma_{\tilde{m}, \tilde{s}} + \text{pen}(m, s)) + 3\|\tilde{f}_{m, s} - f\|^2 \\ &\leq 6\Gamma_{m, s} + 6\text{pen}(m, s) + 3\|\tilde{f}_{m, s} - f\|^2 \end{aligned} \quad (17)$$

Now we study $\Gamma_{m, s}$. First:

$$\|\tilde{f}_{(m, s) \wedge (m', s')} - \tilde{f}_{m', s'}\|^2 \leq 3 \left(\|\tilde{f}_{m', s'} - f_{m'}\|^2 + \|f_{m'} - f_{m \wedge m'}\|^2 + \|f_{m \wedge m'} - \tilde{f}_{(m', s') \wedge (m, s)}\|^2 \right).$$

Thus:

$$\begin{aligned} \Gamma_{m, s} &\leq \max_{(m', s') \in \mathcal{C}} \left(3\|\tilde{f}_{m', s'} - f_{m'}\|^2 + 3\|f_{m'} - f_{m \wedge m'}\|^2 + 3\|f_{m \wedge m'} - \tilde{f}_{(m', s') \wedge (m, s)}\|^2 - \text{pen}(m', s') \right)_+ \\ &\leq 3 \max_{(m', s') \in \mathcal{C}} \left(\|\tilde{f}_{m', s'} - f_{m'}\|^2 - \frac{1}{6}\text{pen}(m', s') \right)_+ \\ &+ 3 \max_{(m', s') \in \mathcal{C}} \left(\|\tilde{f}_{(m', s') \wedge (m, s)} - f_{m \wedge m'}\|^2 - \frac{1}{6}\text{pen}(m', s') \right)_+ \\ &+ 3 \max_{m' \in \mathcal{M}} \|f_{m'} - f_{m \wedge m'}\|^2. \end{aligned}$$

The last maximum can be explicit. If $m' \leq m$, then $\|f_{m'} - f_{m \wedge m'}\|^2 = \|f_{m'} - f_{m'}\|^2 = 0$. Otherwise,

$$\|f_{m'} - f_{m \wedge m'}\|^2 = \|f_{m'} - f_m\|^2 = \int_{m \leq |u| \leq m'} |f^*(u)|^2 du \leq \|f - f_m\|^2.$$

Finally:

$$\max_{m' \in \mathcal{M}} \|f_{m'} - f_{m \wedge m'}\|^2 \leq \|f - f_m\|^2.$$

We get the following bound for $\Gamma_{m, s}$:

$$\begin{aligned} \Gamma_{m, s} &\leq 3 \max_{(m', s') \in \mathcal{C}} \left(\|\tilde{f}_{m', s'} - f_{m'}\|^2 - \frac{1}{6}\text{pen}(m', s') \right)_+ \\ &+ 3 \max_{(m', s') \in \mathcal{C}} \left(\|\tilde{f}_{(m', s') \wedge (m, s)} - f_{m \wedge m'}\|^2 - \frac{1}{6}\text{pen}(m', s') \right)_+ \\ &+ 3\|f - f_m\|^2. \end{aligned} \quad (18)$$

Then we gather Equations (17) and (18):

$$\begin{aligned} \|\tilde{f}_{\tilde{m}, \tilde{s}} - f\|^2 &\leq 6\text{pen}(\tilde{m}, \tilde{s}) + 3\|\tilde{f}_{m, s} - f\|^2 + 18\|f - f_m\|^2 + \max_{(m', s') \in \mathcal{C}} 18 \left(\|\tilde{f}_{m', s'} - f_{m'}\|^2 - \frac{1}{6}\text{pen}(m', s') \right)_+ \\ &+ \max_{(m', s') \in \mathcal{C}} 18 \left(\|\tilde{f}_{(m', s') \wedge (m, s)} - f_{m \wedge m'}\|^2 - \frac{1}{6}\text{pen}(m', s') \right)_+. \end{aligned}$$

We first notice that our penalty function is increasing in s and m , thus we get the following bound for the last term:

$$\begin{aligned}
& \mathbb{E} \left[\max_{(m', s') \in \mathcal{C}} \left(\|\tilde{f}_{(m', s') \wedge (m, s)} - f_{m \wedge m'}\|^2 - \frac{1}{6} \text{pen}((m', s') \wedge (m, s)) \right)_+ \right] \\
& \leq \mathbb{E} \left[\max_{m' \leq m, s' \leq s} \left(\|\tilde{f}_{m', s'} - f_{m'}\|^2 - \frac{1}{6} \text{pen}(m', s') \right)_+ \right] + \mathbb{E} \left[\max_{m \leq m', s \leq s'} \left(\|\tilde{f}_{m, s} - f_m\|^2 - \frac{1}{6} \text{pen}(m, s) \right)_+ \right] \\
& \quad + \mathbb{E} \left[\max_{m \leq m', s' \leq s} \left(\|\tilde{f}_{m, s'} - f_m\|^2 - \frac{1}{6} \text{pen}(m, s') \right)_+ \right] + \mathbb{E} \left[\max_{m' \leq m, s \leq s'} \left(\|\tilde{f}_{m', s} - f_{m'}\|^2 - \frac{1}{6} \text{pen}(m', s) \right)_+ \right] \\
& \leq 4 \sum_{m' \in \mathcal{M}} \sum_{s' \in \mathcal{S}} \mathbb{E} \left[\|\tilde{f}_{m', s'} - f_{m'}\|^2 - \frac{1}{6} \text{pen}(m', s') \right]_+.
\end{aligned}$$

Moreover, according to Proposition 1 and using the inequality $\int_0^1 e^{\sigma^2 s^2 v^2} dv \leq e^{\sigma^2 s^2}$, we obtain, for all $(m, s) \in \mathcal{C}$,

$$\begin{aligned}
\mathbb{E}[\|\tilde{f}_{\tilde{m}, \tilde{s}} - f\|^2] & \leq 5 \times 18 \sum_{m' \in \mathcal{M}} \sum_{s' \in \mathcal{S}} \mathbb{E} \left[\|\tilde{f}_{m', s'} - f_{m'}\|^2 - \frac{1}{6} \text{pen}(m', s') \right]_+ + 6 \text{pen}(m, s) \\
& \quad + 3 \frac{m}{\pi N} e^{\sigma^2 s^2} + 21 \|f - f_m\|^2.
\end{aligned}$$

Then we obtain the announced result with the following Lemma.

Lemma 8. *There exists a constant $C' > 0$ such that for $\text{pen}(m, s)$ defined by $\text{pen}(m, s) = \kappa \frac{m}{N} e^{\sigma^2 s^2}$,*

$$\sum_{m' \in \mathcal{M}} \sum_{s' \in \mathcal{S}} \mathbb{E} \left[\|\tilde{f}_{m', s'} - f_{m'}\|^2 - \frac{1}{6} \text{pen}(m', s') \right]_+ \leq \frac{C'}{N}.$$

According to the Lemma 8, we choose $\text{pen}(m, s) = \kappa \frac{m}{N} e^{\sigma^2 s^2}$, thus, there exist two constants $C, C' > 0$ such that,

$$\begin{aligned}
\mathbb{E}[\|\tilde{f}_{\tilde{m}, \tilde{s}} - f\|^2] & \leq 5 \times 18 \sum_{m' \in \mathcal{M}} \sum_{s' \in \mathcal{S}} \mathbb{E} \left[\|\tilde{f}_{m', s'} - f_{m'}\|^2 - \frac{1}{6} \text{pen}(m', s') \right]_+ + (6\kappa + \frac{3}{\pi}) \frac{m}{N} e^{\sigma^2 s^2} + 21 \|f - f_m\|^2 \\
& \leq C \inf_{(m, s) \in \mathcal{C}} \{ \|f - f_m\|^2 + \frac{m}{N} e^{\sigma^2 s^2} \} + \frac{C'}{N}. \quad \square
\end{aligned}$$

Proof of Lemma 8

For a couple $(m, s) \in \mathcal{C}$ fixed, let us consider the subset $S_m := \{t \in \mathbb{L}^1 \cap \mathbb{L}^2, \text{supp}(t^*) = [-m, m]\}$. For $t \in S_m$,

$$\nu_N(t) = \frac{1}{N} \sum_{j=1}^N (\varphi_t(Z_{j, m^2/s^2}) - \mathbb{E}[\varphi_t(Z_{j, m^2/s^2})])$$

with $\varphi_t(x) := \frac{1}{2\pi} \int \overline{t^*(u)} e^{iux + \sigma^2 u^2 s^2 / (2m^2)} du$, then $\nu_N(t) = \frac{1}{2\pi} \langle t^*, (\tilde{f}_{m, s} - f_m)^* \rangle$. This leads to

$$\|\tilde{f}_{m, s} - f_m\|^2 = \sup_{t \in S_m, \|t\|=1} |\nu_N(t)|^2. \quad (19)$$

We also have by Cauchy-Schwarz inequality

$$\begin{aligned}
\|\varphi_t\|_\infty & \leq \frac{1}{2\pi} \int |t^*(u)| e^{\sigma^2 u^2 s^2 / (2m^2)} du \leq \frac{1}{2\pi} \left(\int_{-m}^m |t^*(u)|^2 du \right)^{1/2} \left(\int_{-m}^m e^{\sigma^2 u^2 s^2 / m^2} du \right)^{1/2} \\
& \leq \frac{\sqrt{2m}}{\sqrt{2\pi}} e^{\sigma^2 s^2 / 2}
\end{aligned}$$

thus

$$\sup_{t \in S_m, \|t\|=1} \|\varphi_t\|_\infty \leq \frac{\sqrt{m}}{\sqrt{\pi}} e^{\sigma^2 s^2 / 2} := M.$$

Then, by Proposition 1,

$$\mathbb{E} \left[\sup_{t \in S_m, \|t\|=1} |\nu_N(t)|^2 \right] = \mathbb{E} \left[\|\tilde{f}_{m,s} - f_m\|^2 \right] \leq \frac{m}{N} \int_0^1 e^{\sigma^2 s^2 v^2} dv \leq \frac{m}{N} e^{\sigma^2 s^2} := H^2.$$

Using Fubini and Cauchy-Schwarz we obtain for all $(m, s) \in \mathcal{C}$:

$$\begin{aligned} 4\pi \sup_{t \in S_m, \|t\|=1} \text{Var}(\varphi_t(Z_{j,m^2/s^2})) &\leq \sup_{t \in S_m, \|t\|=1} \iint t^*(u)t^*(-v) \mathbb{E} \left[e^{i(u-v)Z_{j,m^2/s^2}} \right] e^{(u^2+v^2)\sigma^2 s^2 / (2m^2)} dudv \\ &\leq 2\pi \left(\iint_{[-m,m]^2} |f^*(u-v)|^2 e^{(u^2+v^2)\sigma^2 s^2 / m^2} dudv \right)^{1/2} \\ &\leq 2\pi \left(e^{2\sigma^2 s^2} \iint_{[-m,m]^2} |f^*(u-v)|^2 dudv \right)^{1/2} \\ &\leq 2\pi e^{\sigma^2 s^2} \sqrt{2m} \left(\int_{-2m}^{2m} |f^*(z)|^2 dz \right)^{1/2} \leq 2\sqrt{2m} \sqrt{2\pi} \sqrt{\pi} e^{\sigma^2 s^2} \|f\| =: 4\pi^2 v, \\ v &:= \frac{\sqrt{m} e^{\sigma^2 s^2} \|f\|}{\sqrt{\pi}}. \end{aligned}$$

Finally using that $m \leq N$, $s \leq 2/\sigma$ and $\sum_{s \in \mathcal{S}} s = (4/\sigma)(1 - (1/2)^{P+1}) < 4/\sigma$, the Talagrand's inequality with $\alpha = 1/2$ if $4H^2 \leq \text{pen}(m, s)/6$ implies,

$$\begin{aligned} \sum_{s \in \mathcal{S}} \sum_{m \in \mathcal{M}} \mathbb{E} \left[\|\tilde{f}_{m,s} - f_m\|^2 - \frac{1}{6} \text{pen}(s, m) \right]_+ &\leq \sum_{s \in \mathcal{S}} \sum_{m \in \mathcal{M}} \left(\frac{C_1 \|f\|}{N} e^{\sigma^2 s^2} \sqrt{m} e^{-C_2 \frac{\sqrt{m}}{\|f\|}} + C_3 \frac{m}{N^2} e^{\sigma^2 s^2} e^{-C_4 \sqrt{N}} \right) \\ &\leq \sum_{s \in \mathcal{S}} \frac{C_1 \|f\|}{N} e^{\sigma^2 s^2} \left(\sum_{m \in \mathcal{M}} \sqrt{m} e^{-C_2 \frac{\sqrt{m}}{\|f\|}} \right) + \sum_{s \in \mathcal{S}} \sum_{m \in \mathcal{M}} C_3 e^4 \frac{1}{N} e^{-C_4 \sqrt{m}} \\ &\leq \frac{C_1 \|f\| (P+1) e^4}{N} \left(\sum_{m \in \mathcal{M}} \sqrt{m} e^{-C_2 \frac{\sqrt{m}}{\|f\|}} \right) + C_3 e^4 \frac{P+1}{N} \sum_{m \in \mathcal{M}} e^{-C_4 \sqrt{m}} \\ &\leq \frac{C'(P+1)}{N} \end{aligned}$$

because with the definition of \mathcal{M} , $\sum_{m \in \mathcal{M}} \sqrt{m} e^{-C_2 \frac{\sqrt{m}}{\|f\|}} \leq a_1 \sum_{k \in \mathbb{N}} k^{1/4} e^{-a_2 k^{1/4}} < +\infty$, and $\sum_{m \in \mathcal{M}} e^{-C_4 m^{1/2}} \leq \sum_{k \in \mathbb{N}} e^{-a_3 k^{1/4}} < +\infty$, with a_1, a_2, a_3 three positive constants. Notice that $C' > 0$ depends on $\sigma, \|f\|, \Delta$.

We choose $\text{pen}(m, s) = \kappa m e^{\sigma^2 s^2} / N$ with $\kappa \geq 24$. \square

A.5 Proof of Theorem 5

We denote:

$$V(h) = \kappa_1 \frac{\|K\|_1^2 \|K\|^2}{Nh} + \kappa_2 \frac{\sigma^4 \|K\|_1^2 \|K\|^2}{3T^2 h^5} =: V_1(h) + V_2(h).$$

Using the definition of $A(h)$ and of \hat{h} we obtain

$$\begin{aligned} \|\hat{f}_{\hat{h}} - f\|^2 &\leq 3\|\hat{f}_{\hat{h}} - \hat{f}_{h, \hat{h}}\|^2 + 3\|\hat{f}_{h, \hat{h}} - \hat{f}_h\|^2 + 3\|\hat{f}_h - f\|^2 \\ &\leq 3 \left(A(h) + V(\hat{h}) \right) + 3 \left(A(\hat{h}) + V(h) \right) + 3\|\hat{f}_h - f\|^2 \\ &\leq 6A(h) + 6V(h) + 3\|\hat{f}_h - f\|^2. \end{aligned}$$

Thus,

$$\|\widehat{f}_h - f\|^2 \leq 6\mathbb{E}[A(h)] + 6V(h) + 3\mathbb{E}[\|\widehat{f}_h - f\|^2],$$

hence, we only have to study the term $\mathbb{E}[A(h)]$. We can decompose $\|\widehat{f}_{h,h'} - \widehat{f}_{h'}\|^2$ as follows:

$$\|\widehat{f}_{h,h'} - \widehat{f}_{h'}\|^2 \leq 5\|\widehat{f}_{h,h'} - \mathbb{E}[\widehat{f}_{h,h'}]\|^2 + 5\|\mathbb{E}[\widehat{f}_{h,h'}] - f_{h,h'}\|^2 + 5\|f_{h,h'} - f_{h'}\|^2 + 5\|f_{h'} - \mathbb{E}[\widehat{f}_{h'}]\|^2 + 5\|\mathbb{E}[\widehat{f}_{h'}] - \widehat{f}_{h'}\|^2$$

thus

$$A(h) \leq 5(D_1 + D_2 + D_3 + D_4 + D_5)$$

with:

$$\begin{aligned} D_1 &:= \sup_{h' \in \mathcal{H}_{N,T}} \|f_{h,h'} - f_{h'}\|^2, \\ D_2 &:= \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\widehat{f}_{h'} - \mathbb{E}[\widehat{f}_{h'}]\|^2 - \frac{V_1(h')}{10} \right)_+, \quad D_3 := \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\widehat{f}_{h,h'} - \mathbb{E}[\widehat{f}_{h,h'}]\|^2 - \frac{V_1(h')}{10} \right)_+, \\ D_4 &:= \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\mathbb{E}[\widehat{f}_{h'}] - f_{h'}\|^2 - \frac{V_2(h')}{10} \right)_+, \quad D_5 := \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\mathbb{E}[\widehat{f}_{h,h'}] - f_{h,h'}\|^2 - \frac{V_2(h')}{10} \right)_+. \end{aligned}$$

According to Young inequality (see Theorem 7), we obtain

$$\|f_{h,h'} - f_{h'}\|^2 = \|K_{h'} \star (f_h - f)\|^2 \leq \|K_{h'}\|_1^2 \|f_h - f\|^2 = \|K\|_1^2 \|f_h - f\|^2$$

thus

$$D_1 \leq \|K\|_1^2 \|f_h - f\|^2. \quad (20)$$

Let us study the term D_2 . We define $\nu_{N,h}(t) := \langle t, \widehat{f}_h - \mathbb{E}[\widehat{f}_h] \rangle$, then $|\nu_{N,h}(t)| \leq \|t\| \|\widehat{f}_h - \mathbb{E}[\widehat{f}_h]\|$ thus, the estimator \widehat{f}_h satisfies:

$$\|\widehat{f}_h - \mathbb{E}[\widehat{f}_h]\|^2 = \sup_{t \in \mathbb{L}^2, \|t\|=1} (\nu_{N,h}(t))^2.$$

We can also compute the scalar product which defines $\nu_{N,h}$ and we obtain

$$\nu_{N,h}(t) = \frac{1}{N} \sum_{j=1}^N (t \star K_h^-(Z_{j,T}) - \mathbb{E}[t \star K_h^-(Z_{j,T})]) \quad (21)$$

with $K_h^-(x) := K_h(-x)$. This finally conducts to:

$$\mathbb{E}[D_2] \leq \sum_{h' \in \mathcal{H}_{N,T}} \mathbb{E} \left[\sup_{t \in \mathcal{B}(1)} (\nu_{N,h}(t))^2 - \frac{V_1(h')}{10} \right]_+$$

with $\mathcal{B}(1) = \{f \in \mathbb{L}^2(\mathbb{R}), \|f\| = 1\}$. This bound and Equation (21) leads to apply Talagrand's inequality (6). We have to compute 3 quantities: M , H^2 and v .

First:

$$\begin{aligned} \sup_{t \in \mathcal{B}(1)} \|t \star K_{h'}^-\|_\infty &= \sup_{t \in \mathcal{B}(1)} \sup_{x \in \mathbb{R}} \left| \int t(y) K_{h'}^-(x-y) dy \right| = \sup_{t \in \mathcal{B}(1)} \sup_{x \in \mathbb{R}} |\langle t, K_{h'}^-(\cdot - x) \rangle| \\ &\leq \sup_{t \in \mathcal{B}(1)} \|t\| \|K_{h'}^-\| = \frac{\|K\|}{\sqrt{h'}} := M. \end{aligned} \quad (22)$$

Secondly, Proposition 4 gives

$$\mathbb{E} \left[\sup_{t \in \mathcal{B}(1)} (\nu_{N,h}(t))^2 \right] = \mathbb{E}[\|\widehat{f}_h - \mathbb{E}[\widehat{f}_h]\|^2] \leq \frac{\|K\|^2}{Nh} := H^2. \quad (23)$$

Thirdly:

$$\begin{aligned} \sup_{t \in \mathcal{B}(1)} (\text{Var}(t \star K_{h'}^-(Z_{1,T}))) &\leq \sup_{t \in \mathcal{B}(1)} \mathbb{E}[(t \star K_{h'}^-(Z_{1,T}))^2] \\ &\leq 2 \sup_{t \in \mathcal{B}(1)} \mathbb{E}[(t \star K_{h'}^-(\phi_1))^2] + 2 \sup_{t \in \mathcal{B}(1)} \mathbb{E}[(t \star (K_{h'}^-(Z_{1,T}) - K_{h'}^-(\phi_1)))^2]. \end{aligned}$$

Let us investigate the two terms separately. Young's inequality gives:

$$\mathbb{E}[(t \star K_{h'}^-(\phi_1))^2] = \int (t \star K_{h'}^-(x))^2 f(x) dx \leq \|f\| \|t \star K_{h'}^-\|_4^2 = \frac{\|f\| \|K\|_{4/3}^2}{\sqrt{h'}} := v_1. \quad (24)$$

Then, one can write: $K_{h'}(x - Z_{1,T}) - K_{h'}(x - \phi_1) = (\phi_1 - Z_{1,T}) \int_0^1 (K_{h'})'(x - \phi_1 + u(\phi_1 - Z_{1,T})) du$, thus

$$\begin{aligned} (t \star K_{h'}^-(Z_{1,T}) - t \star K_{h'}^-(\phi_1))^2 &= (\phi_1 - Z_{1,T})^2 \left(\int_0^1 t(x) \int_0^1 (K_{h'})'(x - \phi_1 + u(\phi_1 - Z_{1,T})) du dx \right)^2 \\ &\leq (\phi_1 - Z_{1,T})^2 \int t^2(x) \left(\int_0^1 (K_{h'})'(x - \phi_1 + u(\phi_1 - Z_{1,T})) du \right)^2 dx \\ &\leq (\phi_1 - Z_{1,T})^2 \|t\|^2 \int (K_{h'})'(y)^2 dy = (\phi_1 - Z_{1,T})^2 \|(K_{h'})'\|^2. \end{aligned}$$

With $\mathbb{E}[(\phi_1 - Z_{1,T})^2] = \frac{\sigma^2}{T^2} \mathbb{E}[W_1(T)^2] = \frac{\sigma^2}{T}$, the assumption $T \leq h^{5/2}$ leads to

$$\mathbb{E}[(t \star K_{h'}^-(Z_{1,T}) - t \star K_{h'}^-(\phi_1))^2] \leq \frac{\|K'\|^2 \sigma^2}{h^3 T} \leq \frac{\|K\|^2 \sigma^2}{\sqrt{h'}} := v_2. \quad (25)$$

Finally $v = v_1 + v_2 = A_0/\sqrt{h'}$ with $A_0 = \|f\| \|K\|_{4/3}^2 + \|K\|^2 \sigma^2$.

If $\kappa_1 \|K\|_1^2 \geq 40$, with the assumption $1/(Nh) \leq 1$, Talagrand's inequality (under the assumptions of the Theorem 5) gives

$$\begin{aligned} \mathbb{E} \left(\sup_{t \in \mathcal{B}(1)} (\nu_{N,h'}(t))^2 - \frac{V_1(h')}{10} \right)_+ &\leq \frac{C_1}{N\sqrt{h'}} e^{-C_2/\sqrt{h'}} + C_3 \frac{1}{h'N^2} e^{-C_4\sqrt{N}} \\ &\leq \frac{C_5}{N} \sum_{h' \in \mathcal{H}_{N,T}} \frac{1}{\sqrt{h'}} e^{-C_6/\sqrt{h'}} \leq \frac{C_5 \Sigma(C_6)}{N}. \end{aligned}$$

One can lead the study of D_3 as we have done for D_2 , using the same steps and tools. However $K_h \star K_{h'}$ instead of $K_{h'}$, adds $\|K\|_1$ in M and $\|K\|_1^2$ in H^2 and v .

Then, let us study the term D_4 . If $\kappa_2 \geq 10/(3\|K\|_1^2)$, the bound (13) leads us to

$$\begin{aligned} D_4 &= \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\mathbb{E}[\widehat{f}_{h'}] - f_{h'}\|^2 - \frac{V_2(h')}{10} \right)_+ \\ &\leq \sup_{h' \in \mathcal{H}_{N,T}} \left(\frac{\|K''\|^2 \sigma^4}{3h'^5 T^2} - \frac{\kappa_2 \|K\|_1^2 \|K''\|^2 \sigma^4}{10T^2 h'^5} \right)_+ = 0 \end{aligned}$$

thus $D_4 = 0$. Finally, similarly, if $\kappa_2 \geq 10/3$, we obtain

$$\begin{aligned} D_5 &= \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\mathbb{E}[\widehat{f}_{h,h'}] - f_{h,h'}\|^2 - \frac{V_2(h')}{10} \right)_+ \\ &\leq \sup_{h' \in \mathcal{H}_{N,T}} \left(\frac{\|K''\|^2 \|K\|_1^2 \sigma^4}{3h'^5 T^2} - \frac{\kappa_2 \|K\|_1^2 \|K''\|^2 \sigma^4}{10T^2 h'^5} \right)_+ = 0. \end{aligned}$$

Finally we have shown that for all $h \in \mathcal{H}_{N,T}$:

$$\begin{aligned} \mathbb{E}[\|\widehat{f}_h - f\|^2] &\leq 6\kappa_1 \frac{\|K\|_1^2 \|K''\|^2}{Nh} + 6\kappa_2 \frac{\|K\|_1^2 \|K''\|^2 \sigma^4}{T^2 h^5} + 3 \left(2\|f - f_h\|^2 + \frac{\|K\|^2}{Nh} + \frac{\|K''\|^2 \sigma^4}{3T^2 h^5} \right) \\ &\quad + 30 \left(\|K\|_1^2 \|f - f_h\|^2 + \frac{C}{N} \right) \\ &\leq C_1 \inf_{h \in \mathcal{H}_{N,T}} \left(\|f - f_h\|^2 + \frac{\|K''\|^2}{Nh} + \frac{\|K''\|^2 \sigma^4}{3T^2 h^5} \right) + \frac{C_2}{N}. \end{aligned}$$

where C_1 depends on $\|K\|_1$ and C_2 depends on $\|f\|, \|K\|_1, \|K\|, \|K\|_{4/3}$. \square

B Discretization

We study the error applied by discretization of the $Z_{j,\tau}$. Indeed, Times of observations are the $t_k = k\delta, k = 1, \dots, N$ and $0 < \delta < 1$. Then, for any $0 < m^2/s^2 \leq T$ we use:

$$\widehat{Z}_{j,m^2/s^2} = \frac{s^2}{m^2} \left[X_j(\delta[m^2/(s^2\delta)]) - X_j(0) + \frac{\delta}{\alpha} \sum_{k=1}^{[m^2/(s^2\delta)]} X_j((k-1)\delta) \right] \quad (26)$$

to approximate $Z_{j,m^2/s^2}$ given by (3). The corresponding estimator of f is

$$\widetilde{f}_{m,s}(x) = \frac{1}{2\pi} \int_{-m}^m e^{-iux} \frac{1}{N} \sum_{j=1}^N e^{iu\widehat{Z}_{j,m^2/s^2}} e^{\frac{u^2\sigma^2 s^2}{2m^2}} du. \quad (27)$$

We investigate the error:

$$\mathbb{E}[\|\widehat{f}_{m,s} - f\|^2] \leq 2\mathbb{E}[\|\widehat{f}_{m,s} - \widetilde{f}_{m,s}\|^2] + 2\mathbb{E}[\|\widetilde{f}_{m,s} - f\|^2]$$

where the second term of the right hand side is bounded by Proposition 1. Then, Plancherel-Parseval's Theorem implies:

$$\begin{aligned} \mathbb{E}[\|\widehat{f}_{m,s} - \widetilde{f}_{m,s}\|^2] &\leq \frac{1}{2\pi} \mathbb{E} \left[\int_{-m}^m \frac{1}{N} \sum_{j=1}^N e^{u^2\sigma^2 s^2/m^2} \left| e^{iu\widehat{Z}_{j,m^2/s^2}} - e^{iuZ_{j,m^2/s^2}} \right|^2 du \right] \\ &\leq \frac{1}{2\pi} \int_{-m}^m e^{u^2\sigma^2 s^2/m^2} \mathbb{E} \left[\left| e^{iu\widehat{Z}_{1,m^2/s^2}} - e^{iuZ_{1,m^2/s^2}} \right|^2 \right] du \end{aligned}$$

and

$$\mathbb{E} \left[\left| e^{iu\widehat{Z}_{1,m^2/s^2}} - e^{iuZ_{1,m^2/s^2}} \right|^2 \right] \leq |u|^2 \mathbb{E} \left[\left| \widehat{Z}_{1,m^2/s^2} - Z_{1,m^2/s^2} \right|^2 \right]$$

thus we study the last term. For all $(m, s) \in \mathcal{C}$, $m^2/s^2 \leq T$,

$$\begin{aligned} Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2} &= \frac{s^2}{m^2} (X_j(m^2/s^2) - X_j(\delta[m^2/(s^2\delta)])) \\ &\quad + \frac{s^2}{\alpha m^2} \sum_{k=1}^{[m^2/(s^2\delta)]} \int_{(k-1)\delta}^{k\delta} (X_j(s) - X_j((k-1)\delta)) ds \end{aligned}$$

then by Cauchy-Schwarz's inequality we obtain

$$\begin{aligned} (Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2})^2 &\leq \frac{2s^4}{m^4} (X_j(m^2/s^2) - X_j(\delta[m^2/(s^2\delta)]))^2 \\ &\quad + \frac{2s^4}{\alpha^2 m^4} \left[\sum_{k=1}^{\lceil \frac{m^2}{s^2\delta} \rceil} \int_{(k-1)\delta}^{k\delta} (X_j(s) - X_j((k-1)\delta)) ds \right]^2. \end{aligned}$$

Hölder's inequality yields

$$\begin{aligned} \left[\sum_{k=1}^{\lceil \frac{m^2}{s^2\delta} \rceil} \int_{(k-1)\delta}^{k\delta} (X_j(s) - X_j((k-1)\delta)) ds \right]^2 &\leq \sum_{k=1}^{\lceil \frac{m^2}{s^2\delta} \rceil} \left[\int_{(k-1)\delta}^{k\delta} (X_j(s) - X_j((k-1)\delta)) ds \right]^2 \left[\frac{m^2}{s^2\delta} \right] \\ &\leq \left[\frac{m^2}{s^2\delta} \right] \delta \sum_{k=1}^{\lceil \frac{m^2}{s^2\delta} \rceil} \int_{(k-1)\delta}^{k\delta} (X_j(s) - X_j((k-1)\delta))^2 ds. \end{aligned}$$

Let us study $\mathbb{E}[(X_j(s) - X_j((k-1)\delta))^2]$, for $(k-1)\delta \leq s \leq k\delta$:

$$X_j(s) - X_j((k-1)\delta) = \int_{(k-1)\delta}^s \left(\phi_j - \frac{X_j(u)}{\alpha} \right) du + \int_{(k-1)\delta}^s \sigma dW_j(u)$$

and Cauchy-Schwarz's inequality gives

$$\begin{aligned} \mathbb{E}[(X_j(s) - X_j((k-1)\delta))^2] &\leq 2\mathbb{E} \left[\left(\int_{(k-1)\delta}^s \left(\phi_j - \frac{X_j(u)}{\alpha} \right) du \right)^2 \right] + 2\mathbb{E} \left[\left(\int_{(k-1)\delta}^s \sigma dW_j(u) \right)^2 \right] \\ &\leq 2\mathbb{E} \left[\int_{(k-1)\delta}^s \left(\phi_j - \frac{X_j(u)}{\alpha} \right)^2 du \right] + 2\delta\sigma^2 \\ &\leq 4\delta^2 \left(\mathbb{E}(\phi_j^2) + \frac{1}{\alpha^2} \sup_{s \geq 0} \mathbb{E}[X_j(s)^2] \right) + 2\delta\sigma^2. \end{aligned} \quad (28)$$

Finally, after simplification and using for all $x \in \mathbb{R}^+$, $[x] \leq x$,

$$\begin{aligned} \mathbb{E} \left[(Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2})^2 \right] &\leq \frac{2s^4}{m^4} \mathbb{E}[(X_j(m^2/s^2) - X_j(\delta[m^2/(s^2\delta)]))^2] \\ &\quad + \frac{2}{\alpha^2} \left(4\delta^2 \left(\mathbb{E}(\phi_j^2) + \frac{1}{\alpha^2} \sup_{s \geq 0} \mathbb{E}[X_j(s)^2] \right) + 2\delta\sigma^2 \right) \end{aligned}$$

and we can deal with the term $\mathbb{E}[(X_j(m^2/s^2) - X_j(\delta[m^2/(s^2\delta)]))^2]$ using formula (28) and $m^2/s^2 - \delta[m^2/(s^2\delta)] \leq \delta$. Thus:

$$\mathbb{E} \left[(Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2})^2 \right] \leq \left(\frac{2s^4}{m^4} + \frac{2}{\alpha^2} \right) \left(4\delta^2 \left(\mathbb{E}(\phi_j^2) + \frac{1}{\alpha^2} \sup_{s \geq 0} \mathbb{E}[X_j(s)^2] \right) + 2\delta\sigma^2 \right).$$

Besides, for model (1), Equation (2) implies $\mathbb{E}[X_j(s)^2] \leq 3x_j^2 + 3\alpha^2 \mathbb{E}[\phi_j^2] + 3\sigma^2$, and $0 < \delta < 1$ implies

$$\mathbb{E} \left[(Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2})^2 \right] \leq C\delta \left(\frac{2s^4}{m^4} + \frac{2}{\alpha^2} \right)$$

with C a positive constant which does not depend on δ or m^2/s^2 . Finally,

$$\begin{aligned}\mathbb{E}[\|\widehat{f}_{m,s} - \widetilde{f}_{m,s}\|^2] &\leq C\delta \left(\frac{2s^4}{m^4} + \frac{2}{\alpha^2}\right) \frac{1}{2\pi} \int_{-m}^m u^2 e^{u^2\sigma^2s^2/m^2} du \\ &\leq C'\delta \left(\int_0^1 v^2 e^{v^2\sigma^2s^2} dv\right) \left(\frac{s^4}{m} + \frac{m^3}{\alpha^2}\right).\end{aligned}$$

But $s \leq 2/\sigma$ and $m = \sqrt{k\Delta}/\sigma$, with $k \in \mathbb{N}^*$ and $0 < \Delta < 1$, thus we obtain

$$\mathbb{E}[\|\widehat{f}_{m,s} - \widetilde{f}_{m,s}\|^2] \leq \frac{C'}{\sigma^3} \left(\int_0^1 v^2 e^{v^2\sigma^2s^2} dv\right) \left(2^4\sqrt{k} \left(\frac{\delta}{\sqrt{\Delta}}\right) + \frac{k^{3/2}}{\alpha^2} (\delta\Delta^{3/2})\right).$$

Proposition 9. Under (A), assuming $\mathbb{E}[\phi_j^2] < +\infty$, the estimator $\widehat{f}_{m,s}$ given by (27) satisfies

$$\mathbb{E}[\|\widehat{f}_{m,s} - f\|^2] \leq \|f_m - f\|^2 + \frac{\sqrt{k\Delta}}{\sigma\pi N} e^{\sigma^2s^2} + \frac{C' e^{\sigma^2s^2}}{\sigma^3 2\sigma^2s^2} \left(2^4\sqrt{k} \left(\frac{\delta}{\sqrt{\Delta}}\right) + \frac{k^{3/2}}{\alpha^2} (\delta\Delta^{3/2})\right).$$

Finally if Δ is fixed and δ is small, the error is acceptable. When $\Delta = \delta$ the error is of order $\sqrt{\delta}$, which is the choice we made in practice most of the time.

For study on the kernel estimator we refer to Comte *et al.* (2013).

References

- Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4**, 329–375.
- Briane, M. and Pagès, G. (2006). *Théorie de l'intégration*. Vuibert, Paris.
- Butucea, C. and Tsybakov, A. (2007). Sharp optimality in density deconvolution with dominating bias. II. *Teor. Veroyatnost. i Primenen.* **52**, 336–349.
- Chagny, G. (2013). Warped bases for conditional density estimation. *Mathematical Methods of Statistics* **22**, 253–282.
- Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* **13**, 514–543.
- Comte, F., Genon-Catalot, V. and Samson, A. (2013). Nonparametric estimation for stochastic differential equation with random effects. *Stochastic Processes and their Applications* **7**, 2522–2551.
- Comte, F. and Johannes, J. (2012). Adaptive functional linear regression. *The Annals of Statistics* **40**, 2765–2797.
- Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Can. J. Stat.* **34**, 431–452.
- Davidian, M. and Giltinan, D. (1995). *Nonlinear models for repeated measurement data*. Chapman and Hall.
- Delattre, M., Genon-Catalot, V. and Samson, A. (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scandinavian Journal of Statistics* **40**, 322–343.

- Delattre, M., Genon-Catalot, V. and Samson, A. (2014). Estimation of population parameters in stochastic differential equations with random effects in the diffusion coefficient. *preprint MAP5-2014-07* .
- Delattre, M. and Lavielle, M. (2013). Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Statistics and Its Interface* **6**, 519–532.
- Diggle, P., Heagerty, P., Liang, K. and Zeger, S. (2002). *Analysis of longitudinal data*. Oxford University Press.
- Donnet, S., Foulley, J. and Samson, A. (2010). Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics* **66**, 733–741.
- Donnet, S. and Samson, A. (2008). Parametric inference for mixed models defined by stochastic differential equations. *ESAIM P&S* **12**, 196–218.
- Donnet, S. and Samson, A. (2013). A review on estimation of stochastic differential equations for pharmacokinetic - pharmacodynamic models. *Advanced Drug Delivery Reviews* **65**, 929–939.
- Donnet, S. and Samson, A. (2014). Using PMCMC in EM algorithm for stochastic mixed models: theoretical and practical issues. *Journal de la Société Française de Statistique* **155**, 49–72.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257–1272.
- Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques* **29**, 119–151.
- Genon-Catalot, V. and Larédo, C. (2013). Stationary distributions for stochastic differential equations with random effects and statistical applications. preprint map5 .
- Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics* **39**, 1608–1632.
- Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications* **79**, 135 – 163.
- Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33**, 1060–1077.
- Kutoyants, Y. (2004). *Statistical Inference for Ergodic Diffusion Processes*. Springer, London.
- Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Math. Acad. Sci. Paris* **342**, 877–882.
- Lansky, P., Sanda, P. and He, J. (2006). The parameters of the stochastic leaky integrate-and-fire neuronal model. *Journal of Computational Neuroscience* **21**, 211–223.
- Picchini, U., De Gaetano, A. and Ditlevsen, S. (2010). Stochastic differential mixed-effects models. *Scandinavian Journal of statistics* **37**, 67–90.
- Picchini, U. and Ditlevsen, S. (2011). Practicle estimation of high dimensional stochastic differential mixed-effects models. *Computational Statistics & Data Analysis* **55**, 1426–1444.
- Picchini, U., Ditlevsen, S., De Gaetano, A. and Lansky, P. (2008). Parameters of the diffusion leaky integrate-and-fire neuronal model for a slowly fluctuating signal. *Neural Computation* **20**, 2696–2714.

Pinheiro, J. and Bates, D. (2000). *Mixed-effect models in S and Splus*. Springer-Verlag, New York.

Yu, Y., Xiong, Y., Chan, Y. and He, J. (2004). Corticofugal gating of auditory information in the thalamus: an in vivo intracellular recording study. *The Journal of Neuroscience* **24**, 3060–3069.