

# Nonparametric drift estimation for diffusions from noisy data

Emeline Schmisser

► **To cite this version:**

Emeline Schmisser. Nonparametric drift estimation for diffusions from noisy data. *Statistics & Decisions*, 2011, 28 (2), pp.119-150. <10.1524/stnd.2011.1063>. <hal-00419954>

**HAL Id: hal-00419954**

**<https://hal.archives-ouvertes.fr/hal-00419954>**

Submitted on 25 Sep 2009

**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 drift estimation for diffusions from noisy data

Emeline Schmisser \*

*\*Laboratoire MAP5, Université Paris Descartes.  
45 rue des Saints Pères, 75270 Paris Cedex 06, France.  
E-mail: emeline.schmisser@mi.parisdescartes.fr*

## Abstract

Let us consider a diffusion process  $(X_t)_{t \geq 0}$ , with drift  $b(x)$  and diffusion coefficient  $\sigma(x)$ . This process is assumed to be strictly stationary,  $\beta$ -mixing and ergodic. At discrete times  $t_k = k\delta$  for  $k$  from 1 to  $M$ , we have at disposal noisy data of the sample path,  $Y_{k\delta} = X_{k\delta} + \varepsilon_k$ . The random variables  $(\varepsilon_k)$  are i.i.d, centred and independent of  $(X_t)$ . In order to reduce the noise effect, we split data into groups of equal size  $p$  and build empirical means. The group size  $p$  is chosen such that  $\Delta = p\delta$  is small whereas  $n\Delta = N\delta$  is large. Then, we estimate the drift function  $b$  in a compact set  $A$  in a nonparametric way by a penalized least squares approach. We obtain a bound for the risk of the resulting adaptive estimator. We also provide several examples of diffusions satisfying our assumptions and realise various simulations. Our simulation results illustrate the theoretical properties of our estimators.

Running title : Estimation for noisy diffusions

**Keywords** : drift; model selection; noisy data; nonparametric estimation; stationary distribution.

## 1 Introduction

Consider a one-dimensional diffusion process  $(X_t)$  satisfying the stochastic differential equation (SDE):

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = \eta, \quad (1)$$

where  $b, \sigma : \mathbb{R} \rightarrow \mathbb{R}$  are unknown functions,  $\eta$  is a real valued random variable and  $(W_t)$  a Wiener process independent of  $\eta$ . The process  $(X_t)$  is assumed to be strictly stationary, ergodic and  $\beta$ -mixing. At discrete times  $t_k = k\delta$ ,  $k = 0, 1, \dots, M$ , we have at disposal noisy data of the sample path, i.e, we observe

$$Y_{k\delta} = X_{k\delta} + \varepsilon_k \quad (2)$$

where  $(\varepsilon_k, k \geq 0)$  are i.i.d. centred random variables independent of  $(X_t)$ . Our aim is to realize non parametric estimation of the drift function  $b$  over a compact interval  $A = [a_0, a_1]$  ( $A = [0, 1]$  for instance), under the asymptotic framework that  $M \rightarrow +\infty$ ,  $\delta = \delta_M \rightarrow 0$  and  $M\delta_M \rightarrow +\infty$ .

Several papers deal with nonparametric drift estimation for non noisy data, *i.e.* for the direct observation  $(X_{k\delta})$ . (see *e.g.* Hoffmann (1999), Comte *et al.* (2007), Schmisser (2009) for multidimensional diffusions). However, in practice, it is often the case that one cannot observe exactly  $(X_{k\delta})$ . This may be due to either measurements devices or to what is called microstructure noise for financial data. For instance, Zhang *et al.* (2005) and Jacod *et al.* (2009) study the estimation of the integrated volatility  $\int_0^1 \sigma_t^2 dt$  within a fixed-length time interval (*i.e.*  $M\delta_M = 1$ ). In the same context, Gloter and Jacod (2001) study the estimation of unknown parameters in the diffusion coefficient from noisy high-frequency data. With data within of fixed time interval, it is well known that the drift cannot be estimated. The high frequency data context for diffusions observed with noise requires specific tools to reduce the noise effect while keeping enough information to estimate the coefficients of the diffusion. In order to reduce the noise effect, we split data into groups of equal size  $p$  and build empirical means as follows. Assuming that  $M = (n + 2)p$ , we set  $N = np$ ,  $\Delta = p\delta$  and for  $k = 0, 1, \dots, n + 2$ ,

$$\bar{Y}_{k\Delta} = \bar{X}_{k\Delta} + \bar{\varepsilon}_k, \quad (3)$$

with

$$\bar{X}_{k\Delta} = \frac{1}{p} \sum_{j=1}^p X_{k\Delta+j\delta} \quad \text{and} \quad \bar{\varepsilon}_k = \frac{1}{p} \sum_{j=1}^p \varepsilon_{kp+j}. \quad (4)$$

The group size  $p$  is chosen such that  $\Delta = p\delta$  is small whereas  $n\Delta = N\delta$  is large. Then, based on the means sample  $(\bar{Y}_{k\Delta}, k = 0, 1, \dots, n + 2)$ , we apply the method of Comte *et al.* (2007). First, we find an adequate regression-type equation:

$$\frac{\bar{Y}_{(k+1)\Delta} - \bar{Y}_{k\Delta}}{\Delta} = b(\bar{Y}_{(k-1)\Delta}) + \text{noise} + \text{remainder}.$$

(The lag is here to avoid cumbersome correlations.) Then, a penalised least-square approach is used to build a non parametric adaptive estimator of  $b(\cdot)$  (without knowledge of  $\sigma(\cdot)$ ). We introduce a family of finite-dimensional subspaces  $(S_m)$  of  $L^2(A)$  and define a collection  $(\hat{b}_m)$  of estimators of  $b_A = b\mathbb{1}_A$ . Then, introducing a penalty, we select, by a data-driven procedure, an adaptive estimator  $\hat{b}_{\hat{m}}$  among the collection. We prove that the adaptive estimator risk achieves the usual optimal non parametric rate of the non noisy data case under the assumption of sub-Gaussian noises, which include Gaussian or bounded noises.

In Section 2, we specify the model and the assumptions. Section 3 describes the approximation spaces. In Section 4, we precise the regression-type equation and the construction of the collection  $(\hat{b}_m)$  of estimators. Theorem 1 gives the risk bound of an estimator  $\hat{b}_m$  for fixed  $m$ . Theorem 2 and 3 give the risk bound of the adaptive estimator  $\hat{b}_{\hat{m}}$ . In Section 5, we propose examples of models with different noises distribution and implement the estimation method on simulated data. Proofs are gathered in Section 6 and in the Appendix.

## 2 Model and assumptions

Recall that  $A = [a_0, a_1]$  is the compact set where  $b$  is estimated and consider the following assumptions:

**Assumption 1.**

*Functions  $\sigma(x)$  and  $b(x)$  are globally Lipschitz.*

We denote by  $b_L$  and  $\sigma_L$  the Lipschitz constants of  $b$  and  $\sigma$ .

**Assumption 2.**

*There exist constants  $r > 0$  and  $\alpha \geq 1$  such that*

$$\exists M_0 \in \mathbb{R}^+, \quad \forall x, |x| \geq M_0, \quad xb(x) \leq -r|x|^\alpha.$$

**Assumption 3.**

*The diffusion coefficient  $\sigma$  is bounded from below and above:*

$$\exists \sigma_1^2, \sigma_0^2 > 0, \quad \forall x \in \mathbb{R}, \quad \sigma_1^2 \leq \sigma^2(x) \leq \sigma_0^2.$$

Under Assumptions 1-3, there exists a unique invariant density  $\pi$  such that

$$\pi(x) \propto \frac{1}{\sigma^2(x)} \exp\left(2 \int_0^x \frac{b(u)}{\sigma^2(u)} du\right). \quad (5)$$

**Assumption 4.**

*The process is stationary:*

$$\eta \sim \pi$$

Under Assumptions 2-3,  $\eta$  has moments of any order. Moreover, we derive from equation (5) that there exists a positive constant  $\pi_1$  such that

$$\forall x \in \mathbb{R}, \pi(x) \leq \pi_1 \quad (6)$$

and, over  $I = [a_0 - 2, a_1 + 2]$ , there exists a positive constant  $\pi_0$  such that

$$\forall x \in I, \pi(x) \geq \pi_0. \quad (7)$$

According to Pardoux and Veretennikov (2001), Proposition 1 p.1063, under Assumptions 2-3, the process  $(X_t)$  is exponentially  $\beta$ -mixing: there exist positive constants  $C, \theta$  such that, for all positive  $t$ ,

$$\beta_X(t) \leq Ce^{-\theta t}, \quad (8)$$

where  $\beta_X(t)$  is the  $\beta$ -mixing coefficient of  $(X_t)$ . Moreover, Proposition A p.226 of Gloter (2000) implies that, under Assumptions 1 and 4, for any integer  $k \geq 1$ , there exists a positive constant  $c(k)$  such that

$$\forall t > 0, \quad \forall h, 0 \leq h \leq 1, \quad \mathbb{E} \left( \sup_{s \in [t, t+h]} |b(X_s) - b(X_t)|^k \right) \leq c(k)h^{k/2}. \quad (9)$$

**Assumption 5.**

*The random variables  $\varepsilon_k$  have density  $f$ , are centred and have moments of order 4.*

For fixed  $p$  and  $\delta$ , the processes  $(\bar{X}_{k\Delta})$  and  $(\bar{Y}_{k\Delta})$  defined in (3)-(4) are strictly stationary. Their invariant densities are respectively denoted  $\bar{\pi}_{p,\delta}$  and  $\tilde{\pi}_{p,\delta}$ .

**Assumption 6.**

There exist positive constants  $\bar{\pi}_0, \bar{\pi}_1$  independent of  $p$  and  $\delta$  such that:

$$\forall x \in \mathbb{R}, \bar{\pi}_{p,\delta}(x) \leq \bar{\pi}_1 \quad \text{and} \quad \forall x \in B, \bar{\pi}_{p,\delta}(x) \geq \bar{\pi}_0$$

where  $B = [a_0 - 1, a_1 + 1]$ .

We prove this assumption under sufficient conditions on  $b$  and  $\sigma$ .

**Proposition 1.**

If

- (a) the process  $(X_t)_{t \geq 0}$  is an Ornstein-Uhlenbeck process.
- (b) the functions  $b$  and  $\sigma$  are  $\mathcal{C}^3$ ,  $|b|, |b'|, |\sigma|, |\sigma'|, |\sigma''|$  are bounded and  $\sigma$  is bounded from below by  $\sigma_1 > 0$ ,

Assumption 6 is satisfied.

In fact, it is necessary to bound the function  $\tilde{\pi}$ . The following proposition will be very useful in proves:

**Proposition 2.**

Under Assumptions 5 and 6, there exist positive constants  $\tilde{\pi}_0$  and  $\tilde{\pi}_1$  such that, for all  $\delta$ , if  $p \geq p_0$  (constant depending only on the law of  $\varepsilon_1$ ):

$$\forall x \in \mathbb{R}, \tilde{\pi}_{p,\delta}(x) \leq \tilde{\pi}_1 \quad \text{and} \quad \forall x \in A, \tilde{\pi}_{p,\delta}(x) \geq \tilde{\pi}_0.$$

### 3 Approximation spaces

Our aim is to estimate the drift function  $b$  over an interval  $A = [a_0, a_1]$  of  $\mathbb{R}$ . For simplicity, we choose  $A = [0, 1]$ . Below, we construct a family of nested linear subspaces  $(S_m)_{m \in \mathcal{M}_n}$  of  $L^2(A)$  where  $\mathcal{M}_n$  is the index set of the collection:

$$\mathcal{M}_n(r) = \mathcal{M}_n = \{m, D_m := \dim(S_m) \leq \mathcal{D}_n\}, \quad (10)$$

where the maximal dimension  $\mathcal{D}_n$  will be specified later. For any  $m \in \mathcal{M}_n$ , an estimator  $\hat{b}_m$  of  $b_A = b\mathbb{1}_A$  belonging to  $S_m$  is computed. Then the ‘‘best’’ possible estimator is chosen by introducing a penalty function  $pen(m)$ .

Spline functions are used in order to construct the spaces  $S_m$ . The spline function of degree  $r$ , denoted  $g_r$ , is the convolution  $(r+1)$  times of the indicator function of  $[0, 1]$ . It is  $\mathcal{C}^{r-1}$  and supported on  $[0, r+1]$ . The linear subspaces  $S_m$  are generated by translating/dilating  $g_r$ :  $S_m = \text{Vect} \{(f_{m,k}), k = -r, \dots, 2^m - 1\}$  where

$$f_{m,k}(x) = 2^{m/2} g_r(2^m x - k) \mathbb{1}_{[0,1]}(x).$$

The following proposition will be proved in the Appendix:

**Proposition 3.**

Functions  $f_{m,k}$ ,  $k \in \{-r, -r+1, \dots, 2^m - 1\}$  are linearly independent.

The dimension of  $S_m$  is then  $D_m = 2^m + r$ . The infinite norm and the  $L^2$  norm are connected: according to Schmisser (2009), Proposition 3 p.6, there exists a positive constant  $\phi_1$  such that, for any function  $t \in S_m$ :

$$\|t\|_\infty^2 \leq \phi_1^2 D_m \|t\|_{L^2}^2 \quad (11)$$

where  $\|t\|_\infty^2 = \sup_{t \in [0,1]} |t(x)|$  and  $\|t\|_{L^2}^2 = \int_0^1 t^2(x) dx$ .

**Proposition 4.**

*We can construct an orthonormal basis  $(\psi_\lambda)$  of  $S_m$  such that:*

$$\exists \phi_2, \forall \lambda, \text{card}(\{\lambda', \|\psi_\lambda \psi_{\lambda'}\|_\infty \neq 0\}) \leq \phi_2.$$

This proposition will be later proved. According to Meyer (1990), spline functions generate a multiresolution analysis of  $L^2(\mathbb{R})$ , and, according to Proposition 4, p.50, for any function  $t$  in a Besov space  $B_{2,\infty}^\alpha([0,1])$ , with  $r \geq \alpha$ , there exists a positive constant  $C$  such that the orthogonal projection ( $L^2$ ) of  $t$  over  $S_m$  denoted by  $t_m$  satisfies:

$$\|t - t_m\|_{L^2} \leq 2^{-m\alpha} C. \quad (12)$$

## 4 Estimation

Let us introduce the normalised increment

$$T_{k\Delta} = \frac{\bar{Y}_{(k+1)\Delta} - \bar{Y}_{k\Delta}}{\Delta} \quad (13)$$

which satisfies the equation

$$T_{k\Delta} = b(\bar{Y}_{(k-1)\Delta}) + A_{k\Delta} + B_{k\Delta} + R_{k\Delta} + I_{k\Delta} + Z_{k\Delta}, \quad (14)$$

where

$$\begin{aligned} Z_{k\Delta} &= \frac{1}{p\Delta} \sum_{j=1}^p \int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta} \sigma(X_s) dW_s, \\ A_{k\Delta} &= \left( \frac{1}{p} \sum_{j=1}^p b(X_{k\Delta+j\delta}) \right) - b(\bar{X}_{(k-1)\Delta}), \\ I_{k\Delta} &= \frac{1}{p\Delta} \sum_{j=1}^p \int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta} (b(X_s) - b(X_{k\Delta+j\delta})) ds \end{aligned} \quad (15)$$

are only functions of the non noisy process  $(X_t)$ . The other terms  $B_{k\Delta}$  and  $R_{k\Delta}$  contain the noises:

$$R_{k\Delta} = \frac{1}{\Delta} (\bar{\varepsilon}_{k+1} - \bar{\varepsilon}_k) \quad \text{and} \quad B_{k\Delta} = b(\bar{X}_{(k-1)\Delta}) - b(\bar{Y}_{(k-1)\Delta}). \quad (16)$$

For  $t$  belonging to  $S_m$ , we set:

$$\gamma_n(t) = \frac{1}{n} \sum_{k=1}^n [T_{k\Delta} - t(\bar{Y}_{(k-1)\Delta})]^2 \quad (17)$$

and define the estimator  $\hat{b}_m$  as:

$$\hat{b}_m = \arg \min_{t \in S_m} \gamma_n(t).$$

*Remark 1.* We can always find a function minimising  $\gamma_n$ , but it may be not unique. Let us denote by  $v^*$  the transpose of the vector  $v$ . Setting  $T = (T_\Delta, \dots, T_{n\Delta})^*$ , the random vector  $(\hat{b}_m(\bar{Y}_0), \dots, \hat{b}_m(\bar{Y}_{(n-1)\Delta}))^* = \Pi_m(T)$ , where  $\Pi_m$  is the Euclidean projection over the subspace  $\{t(\bar{Y}_0), \dots, t(\bar{Y}_{(n-1)\Delta}), t \in S_m\}$ , is always uniquely defined.

This is why, as in Comte *et al.* (2007), we choose the risk function as the expectation of an empirical norm:

$$\mathcal{R}(\hat{b}_m) = \mathbb{E} \left( \left\| \hat{b}_m - b_A \right\|_n^2 \right)$$

where  $\|t\|_n^2 = \frac{1}{n} \sum_{k=1}^n t^2(\bar{Y}_{(k-1)\Delta})$  and  $b_A = b \mathbb{1}_A$ .

#### 4.1 Risk of the non-adaptive estimator

Let us introduce the asymptotic assumptions and constraints on  $p$ ,  $\delta$ ,  $n$ . We assume that  $p = p_n$ ,  $\delta = \delta_n$  depend on  $n$ , but to ease notations, we omit the subscript  $n$ .

##### Assumption 7.

- (i)  $n \rightarrow \infty, \quad p \rightarrow \infty, \quad \delta \rightarrow 0$
- (ii)  $p\delta = \Delta \rightarrow 0, \quad N\delta = n\Delta \rightarrow \infty$
- (iii)  $\frac{\ln(n)}{n\Delta} \rightarrow 0 \quad \mathcal{D}_n \leq c_0 \frac{n\Delta}{\ln^2(n)}$

where  $\mathcal{D}_n$  is the maximal dimension (see (10)). It is also assumed that  $\Delta \leq 1$ .

##### Theorem 1.

Under Assumptions 1-7, if  $np\Delta^2 \rightarrow \infty$ , the risk of the estimator  $\hat{b}_m$  satisfies:

$$\mathbb{E} \left( \left\| \hat{b}_m - b_A \right\|_n^2 \right) \leq 3\tilde{\pi}_1 \|b_m - b_A\|_n^2 + cD_m \left( \frac{\sigma_0^2}{N\delta} + \frac{\tau^2}{Np^2\delta^2} \right) + \frac{c\sigma_0^2}{N\delta} + \frac{c\tau^2}{Np^2\delta^2} + cp\delta + \frac{c}{p}$$

where  $c$  is a constant.

It remains to optimise the risk with respect to  $p$ . We have to minimise the main variance term, that is  $cD_m \left( \frac{\sigma_0^2}{N\delta} + \frac{\tau^2}{Np^2\delta^2} \right)$ , so we have to take  $p \sim \delta^{-1/2}$ .

##### Corollary 1.

If Assumptions 1-7 are satisfied, the minimum of the risk is obtained for a  $p \sim \delta^{-1/2}$  and satisfies:

$$\mathbb{E} \left( \left\| \hat{b}_m - b_A \right\|_n^2 \right) \leq 3\tilde{\pi}_1 \|b_m - b_A\|_{L^2}^2 + c(\sigma_0^2 + \tau^2) \frac{D_m}{N\delta} + \frac{K}{N\delta} + C\delta^{1/2}.$$

with  $c, C, K$  constants.

## 4.2 Optimisation of the dimension space

For given  $(N, \delta)$ , we wish to select  $m$  in order to obtain the best compromise between the bias term,  $\|b_A - b_m\|_{L^2}^2$ , and the main variance term,  $D_m(N\delta)^{-1}$ . In a first step, the regularity of  $b_A$  is known, that is  $b_A$  belongs to a Besov space  $B_{2,\infty}^\alpha([0, 1])$  and  $\|b\|_{B_{2,\infty}^\alpha}^2 \leq 1$ . According to (12), it is known that  $\|b_m - b_A\|_{L^2}^2 \leq C2^{-2m\alpha}$ , so  $m$  has to satisfy the equation

$$m = \frac{\log_2(N\delta)}{1 + 2\alpha}.$$

and we derive from Corollary 1:

$$\mathbb{E} \left( \left\| \hat{b}_m - b \right\|_n^2 \right) \leq K (N\delta)^{-2\alpha/(2\alpha+1)} + \frac{c}{N\delta} + c'\delta^{1/2}.$$

The risk of a non-noisy process satisfies the same inequality except for the last term which is  $\delta$  instead of  $\delta^{1/2}$ .

*Remark 2.* Let us set  $\delta \sim N^{-\beta}$ , with  $0 < \beta < 1$ . We have

$\beta$	main term	convergence rate
$0 < \beta < \frac{4\alpha}{(6\alpha+1)}$	$\delta^{1/2}$	$N^{-\beta/2}$
$\frac{4\alpha}{(6\alpha+1)} < \beta < 1$	$(N\delta)^{-2\alpha/(2\alpha+1)}$	$N^{-2\alpha(1-\beta)/(2\alpha+1)}$
$2/3 \leq \beta < 1$		

The convergence rate of the risk is bounded below by  $N^{-1/3}$ , and is optimum for  $N\delta^{3/2} = O(1)$ . We recall that, for a non-noisy process, the rate is optimum for  $N\delta^2 = O(1)$ , and is bounded below by  $N^{-1/2}$ .

## 4.3 Adaptive estimation: sub-Gaussian case

Our aim is to select automatically  $\hat{m}$ , without any knowledge of the regularity of  $b$ . Let us introduce a penalty function  $pen(m)$ , depending on the dimension  $D_m$ , the number of observations  $N$  and the discretization step  $\delta$ :  $\hat{m}$  is defined as

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} \left[ \gamma_n(\hat{b}_m) + pen(m) \right],$$

and the resulting estimator is denoted  $\hat{b}_{\hat{m}}$ . We search  $pen(m)$  such that, ideally,

$$\mathbb{E} \left( \left\| \hat{b}_{\hat{m}} - b_A \right\|_n^2 \right) \leq C \inf_{m \in \mathcal{M}_n} \left( \|b_A - b_m\|_{L^2}^2 + pen(m) \right).$$

An additional assumption is needed.

### Assumption 8.

*The variables  $\varepsilon_k$  are sub-Gaussian: there exists a constant  $v$  such that:*

$$\forall \lambda \in \mathbb{R}, \quad \mathbb{E} \left( e^{\lambda \varepsilon_0} \right) \leq e^{v^2 \lambda^2 / 2}.$$



*Remark 3.* Gaussian and bounded variables satisfy this assumption. For instance, a uniform law on  $[-v, v]$  is sub-Gaussian of parameter  $v^2$ . Moreover, a random variable of law  $f(x) \propto x^{2k} e^{-x^2/(2\sigma^2)}$ , where  $k$  is an integer, is sub-Gaussian of parameter  $v^2 = (2k + 1)\sigma^2$ .

**Theorem 2.**

*Under Assumptions 1-8, there exists a universal constant  $\kappa$  such that, if we set*

$$\text{pen}(m) \geq \frac{\kappa(\sigma_0^2 + v^2) D_m}{n\Delta},$$

*for  $p \sim \delta^{-1/2}$ , the risk of the adaptive estimator is bounded by*

$$\mathbb{E} \left( \|\hat{b}_{\hat{m}} - b_A\|_n^2 \right) \leq C \inf_{m \in \mathcal{M}_n} (\|b_m - b_A\|_{L^2}^2 + \text{pen}(m)) + \frac{C(\sigma_0^2 + \tau^2)}{N\delta} + C'\delta^{1/2}.$$

The parameter  $v$  is assumed to be known. On the contrary,  $\sigma_0^2$  is unknown, but we can replace it by a rough estimator (see section 5.2 of Comte *et al.* (2007)). The adaptive estimator automatically realizes the bias-variance compromise: whenever  $b_A$  belongs to some Besov ball, if  $r \geq \alpha$ ,  $\hat{b}_{\hat{m}}$  achieves the optimal corresponding nonparametric rate.

#### 4.4 Adaptive estimation: general case

If Assumption 8 is not satisfied, the inequality of Theorem 1 cannot be generalised to the adaptive case. Nevertheless, it is possible to use a weaker inequality.

**Lemma 1.**

*Under Assumptions 1-7, if  $p\Delta^2 = p^3\delta^2 \rightarrow \infty$ , the risk of the estimator  $\hat{b}_m$  belonging to  $S_m$  satisfies:*

$$\mathbb{E} \left( \|\hat{b}_m - b_A\|_n^2 \right) \leq 3\tilde{\pi}_1 \|b_m - b_A\|_n^2 + c\sigma_0^2 \frac{D_m}{N\delta} + \frac{c\sigma_0^2}{N\delta} + \frac{c\tau^2}{p^3\delta^2} + cp\delta.$$

In order to minimise this inequality, one has to minimise  $\frac{c\tau^2}{p^3\delta^2} + cp\delta$ , that is to take  $p \sim \delta^{-3/4}$ . We obtain that

$$\mathbb{E} \left( \|\hat{b}_m - b_A\|_n^2 \right) \leq 3\tilde{\pi}_1 \|b_m - b_A\|_{L^2}^2 + c\sigma_0^2 \frac{D_m}{N\delta} + \frac{C}{N\delta} + C'\delta^{1/4}.$$

The term  $\delta^{1/2}$  of Corollary 1 is replaced by  $\delta^{1/4}$ , the other terms are the same.

**Theorem 3.**

*Under Assumptions 1-7, there exists a universal constant  $\kappa$  such that, if we set*

$$\text{pen}(m) \geq \frac{\kappa\sigma_0^2 D_m}{N\delta},$$

*the risk of the adaptive estimator with  $p \sim \delta^{-3/4}$  satisfies:*

$$\mathbb{E} \left( \|\hat{b}_{\hat{m}} - b_A\|_n^2 \right) \leq C \inf_{m \in \mathcal{M}_n} (\|b_m - b_A\|_{L^2}^2 + \text{pen}(m)) + \frac{C\sigma_0^2}{N\delta} + C'\delta^{1/4}.$$

## 5 Numerical simulations on examples

We use various methods to simulate variables  $(X_{k\delta})$ . The following table presents, for each model, the drift and the diffusion functions, as well as the method of simulation.

model	$b(x)$	$\sigma(x)$	simulation
Model 1	-2	1	exact
Model 2	$-\frac{x}{\sqrt{1+x^2}}$	1	Beskos
Model 3	$-\frac{\sinh(x)}{\cosh^2(x)} \left(1 + \frac{1}{2\cosh(x)}\right)$	$\frac{1}{\cosh(x)}$	Beskos
Model 4	$-2x + 3 \sin(x)$	1	Euler
Model 5	$-x^3 + 2x$	1	Euler

Model 1 is an Ornstein-Uhlenbeck process, so it can be exactly simulated using Gaussian variables. Models 2 and 3 are detailed in Comte *et al.* (2007). They can be exactly simulated by the retrospective algorithm of Beskos and Roberts (2005) (actually, Model 3 is a transformation of Model 2). Models 4 and 5 are simulated using an Euler scheme. Models 3 and 5 do not satisfy all our assumptions (the diffusion coefficient  $\sigma(x)$  of Model 3 is not bounded from below, and the drift of Model 5 is not Lipschitz).

We use three different laws to compute the noise vector  $(\varepsilon_k)$ : a normal law  $\mathcal{N}(0, 1)$ , a uniform law over  $[-1, 1]$  and a Laplacian law. We recall the density of a Laplacian law (sometimes called double exponential):  $f(x) = \frac{\lambda}{2} e^{-\lambda|x|}$ . Here, we choose  $\lambda = 1$ . This noise is not sub-Gaussian.

For every  $(m, r) \in \mathcal{M}_n \times \{1, \dots, 6\}$ , we compute the estimator  $\hat{b}_{m,r} := \hat{b}_m$  of the drift  $b$ . Then, introducing the penalty function

$$pen(m, r) := pen(m) = \kappa (\sigma_0^2 + v^2) \frac{2^m + r}{n\Delta}$$

(we recall that  $D_m = 2^m + r$ ), we select the adaptive estimator  $\hat{b}_{\hat{m}, \hat{r}}$ . The constant  $\kappa$  is chosen equal to 3 by numerical calibration (see Comte and Rozenholc (2004) for a complete discussion). The error between  $\hat{b}_{\hat{m}, \hat{r}}$  and  $b_A$  is measured by the empirical norm  $error = \|\hat{b}_{\hat{m}, \hat{r}} - b_A\|_n^2$ . In order to check that the algorithm is adaptive, we compute  $emin = \min_{m,r} \{\|\hat{b}_{r,m} - b_A\|_n^2\}$  and the oracle  $error/emin$ .

Figures correspond to an estimation of the drift on the interval  $[-2, 2]$  with  $N = 10^6$ ,  $\delta = 10^{-4}$  and  $p = 100$ .

In tables below, for each triplet  $(N, p, \delta)$ , we have simulated 50 samples of  $\{(X_{k\delta}), k = 1, \dots, M\}$ . We give the empirical risk of the estimator, “ $ris = \text{mean}(error)$ ” over the fifty estimations, and the mean of the oracle “ $or = \text{mean}(error/emin)$ ”. The values selected by the algorithm are denoted  $\hat{m}$  and  $\hat{r}$ . In the tables, we compute their means. The minima of  $ris$  and  $or$  are set in bold.

Comments: The empirical error and the oracle are in general better for the uniform noise, and worse for the Laplacian noise, nevertheless, this is mainly because the noises have different variances: 1/3 for the uniform noise, 1 for the Gaussian noise, and 2 for the Laplacian noise. The results for the Laplacian

noise are very similar to the other results. Actually, even if the variables  $(\varepsilon_k)$  are not sub-Gaussian, according to the central limit theorem, variables  $(\bar{\varepsilon}_k)$  are nearly Gaussian.

If the drift is linear, most of the time, the estimators are linear. The contrary is not true (see Models 2 and 3). Moreover, the smaller  $\Delta = p\delta$  and  $(N\delta)^{-1}$ , the smaller the error, nevertheless, the risk is not proportional to this variation. This can be due to the constants involved in the computation of the risk. The best possible estimators are in general obtained for  $N = 10^{-6}$ ,  $\delta = 10^{-4}$ , either for  $p = 100$ , either for  $p = 10^3$ .

## 6 Proofs

### 6.1 Proof of Theorem 1

We recall the Rosenthal inequality (see Hall and Heyde (1980) theorem 2.12 p.23).

#### The Rosenthal inequality.

Let  $(\eta_1, \dots, \eta_n)$  be centred and independent variables such that  $\mathbb{E}(|\eta_i|^p) < \infty$ . Then, there exists a positive constant  $r_p$  such that

$$\mathbb{E} \left( \left| \sum_{i=1}^n \eta_i \right|^p \right) \leq r_p \left[ \sum_{i=1}^n \mathbb{E} |\eta_i|^p + \left( \sum_{i=1}^n \mathbb{E} (\eta_i^2) \right)^{p/2} \right].$$

The following lemma will be proved in the Appendix. Processes  $A_{k\Delta}$ ,  $I_{k\Delta}$ ,  $B_{k\Delta}$ ,  $R_{k\Delta}$  and  $Z_{k\Delta}$  are defined in (15) and (16).

#### Lemma 2.

$$\begin{aligned} \mathbb{E} (A_{k\Delta}^2 + I_{k\Delta}^2) &\leq cb_L^2 c(2)\Delta, \quad \mathbb{E} (B_{k\Delta}^2) \leq \frac{cb_L^2 \tau^2}{p} \quad \text{and} \quad \mathbb{E} (R_{k\Delta}^2) \leq \frac{c\tau^2}{p\Delta^2}, \\ \mathbb{E} (I_{k\Delta}^4 + A_{k\Delta}^4) &\leq cb_L^4 c(4)\Delta^2, \quad \mathbb{E} (B_{k\Delta}^4) \leq \frac{cb_L^4 \tau^4}{p^2} + \frac{cb_L^4 \mathbb{E}(\varepsilon_1^4)}{p^3} \\ \text{and} \quad \mathbb{E} (R_{k\Delta}^4) &\leq \frac{c\tau^4}{p^2\Delta^4} + \frac{c\mathbb{E}(\varepsilon_1^4)}{p^3\Delta^4}, \\ \mathbb{E} (Z_{k\Delta}^2) &= \left( \frac{2}{3} + \frac{1}{p} \right) \frac{\mathbb{E}(\sigma^2(X_0))}{\Delta} \quad \text{and} \quad \mathbb{E} (Z_{k\Delta}^4) \leq \frac{c\mathbb{E}(\sigma^4(X_0))}{\Delta^2}, \end{aligned}$$

where  $c$  is a universal constant,  $\tau^2 = \mathbb{E}(\varepsilon^2)$ ,  $b_L$  is the Lipschitz constant of  $b$  and  $c(2)$ ,  $c(4)$  are defined in (9).

We have

$$\begin{aligned} \gamma_n(t) - \gamma_n(b) &= \frac{1}{n} \sum_{k=1}^n (b(\bar{Y}_{(k-1)\Delta}) - t(\bar{Y}_{(k-1)\Delta}))^2 \\ &+ \frac{2}{n} \sum_{k=1}^n (I_{k\Delta} + R_{k\Delta} + Z_{k\Delta} + A_{k\Delta} + B_{k\Delta}) (b(\bar{Y}_{(k-1)\Delta}) - t(\bar{Y}_{(k-1)\Delta})). \end{aligned}$$

Let us denote by  $b_m$  the orthogonal projection ( $L^2$ ) of  $b_A$  over  $S_m$  and set, for  $t \in S_m$ ,

$$\begin{aligned}\nu_n(t) &= \frac{1}{n} \sum_{k=1}^n Z_{k\Delta} t(\bar{Y}_{(k-1)\Delta}), \quad \rho_n(t) = \frac{1}{n} \sum_{k=1}^n R_{k\Delta} t(\bar{Y}_{(k-1)\Delta}), \\ E_n(t) &= \frac{1}{n} \sum_{k=1}^n (A_{k\Delta} + B_{k\Delta} + I_{k\Delta}) t(\bar{Y}_{(k-1)\Delta}),\end{aligned}\tag{18}$$

(see (15)- (16)). By definition ,  $\gamma_n(\hat{b}_m) - \gamma_n(b) \leq \gamma_n(b_m) - \gamma_n(b)$ , so:

$$\left\| \hat{b}_m - b \right\|_n^2 \leq \|b_m - b\|_n^2 + 2\nu_n(\hat{b}_m - b_m) + 2\rho_n(\hat{b}_m - b_m) + 2E_n(\hat{b}_m - b_m)$$

and, as  $\hat{b}_m$  and  $b_m$  are  $A$ -supported:

$$\left\| \hat{b}_m - b_A \right\|_n^2 \leq \|b_m - b_A\|_n^2 + 2\nu_n(\hat{b}_m - b_m) + 2\rho_n(\hat{b}_m - b_m) + 2E_n(\hat{b}_m - b_m).$$

Let us consider the norm  $\|t\|_{\tilde{\pi}}^2 = \int_A t^2(x) \tilde{\pi}(x) dx$  where  $\tilde{\pi} = \tilde{\pi}_{p,\delta}$  is the density of  $\bar{Y}_\Delta$ . As the random variables  $\bar{Y}_{k\Delta}$  have density  $\tilde{\pi}$ , we have that  $\mathbb{E} \left( \|t\|_n^2 \right) = \|t\|_{\tilde{\pi}}^2$ . Let us introduce the set

$$\Omega_n = \left\{ \omega, \forall t \in \bigcup_{m,m'} S_m + S_{m'}, \quad \left| \frac{\|t\|_n^2}{\|t\|_{\tilde{\pi}}^2} - 1 \right| \leq \frac{1}{2} \right\}$$

in which norms  $\|\cdot\|_n$  and  $\|\cdot\|_{\tilde{\pi}}$  are equivalent: on  $\Omega_n$ , we have

$$\|t\|_{\tilde{\pi}}^2 \leq 2 \|t\|_n^2 \leq 3 \|t\|_{\tilde{\pi}}^2.\tag{19}$$

In a first step, we will bound the risk over  $\Omega_n$ . Let us set  $\mathcal{B}_m = \{t \in S_m, \|t\|_{\tilde{\pi}} \leq 1\}$ . According to the Cauchy-Schwarz inequality, we have that, on one hand,

$$\begin{aligned}2(\nu_n + \rho_n)(\hat{b}_m - b_m) &\leq 2 \left\| \hat{b}_m - b_m \right\|_{\tilde{\pi}} \sup_{t \in \mathcal{B}_m} |\nu_n(t) + \rho_n(t)| \\ &\leq \frac{1}{14} \left\| \hat{b}_m - b_m \right\|_{\tilde{\pi}}^2 + 28 \sup_{t \in \mathcal{B}_m} (\nu_n^2(t) + \rho_n^2(t))\end{aligned}$$

and on the other hand,

$$2E_n(\hat{b}_m - b_m) \leq \frac{3}{28} \left\| \hat{b}_m - b_m \right\|_n^2 + \frac{28}{n} \sum_{k=1}^n (A_{k\Delta}^2 + B_{k\Delta}^2 + I_{k\Delta}^2).$$

As, according to (19), on  $\Omega_n$ ,

$$\left\| \hat{b}_m - b_m \right\|_{\tilde{\pi}}^2 \leq 2 \left\| \hat{b}_m - b_m \right\|_n^2 \quad \text{and} \quad \left\| \hat{b}_m - b_m \right\|_n^2 \leq 2 \left\| \hat{b}_m - b_A \right\|_n^2 + 2 \|b_A - b_m\|_n^2.$$

By collecting terms, we obtain:

$$\left\| \hat{b}_m - b_A \right\|_n^2 \leq 3 \|b_m - b_A\|_n^2 + 56 \sup_{t \in \mathcal{B}_m} \nu_n^2(t) + \frac{56}{n} \sum_{k=1}^n (A_{k\Delta}^2 + B_{k\Delta}^2 + I_{k\Delta}^2 + R_{k\Delta}^2).$$

Moreover,  $\mathbb{E} \left( \|b_m - b_A\|_n^2 \right) = \|b_m - b_A\|_{\tilde{\pi}}^2 \leq \tilde{\pi}_1 \|b_m - b_A\|_{L^2}^2$ . Using Lemma 2, we get:

$$\mathbb{E} \left( \left\| \hat{b}_m - b_A \right\|_n^2 \mathbf{1}_{\Omega_n} \right) \leq 3\tilde{\pi}_1 \|b_m - b_A\|_{L^2}^2 + 56\mathbb{E} \left( \sup_{t \in \mathcal{B}_m} \nu_n^2(t) + \rho_n^2(t) \right) + \frac{c\tau^2}{p} + c'\Delta.$$

It remains to bound  $\mathbb{E} \left( \sup_{t \in \mathcal{B}_m} \nu_n^2(t) \right)$  and  $\mathbb{E} \left( \sup_{t \in \mathcal{B}_m} \rho_n^2(t) \right)$ . Let  $(\varphi_\lambda, \lambda \in \Lambda_m)$  be an orthonormal basis for the  $L_{\tilde{\pi}}^2$ -norm of  $S_m$ . Any function  $t \in S_m$  can be written  $t = \sum_{\lambda \in \Lambda_m} a_\lambda \varphi_\lambda$  and

$$\|t\|_{\tilde{\pi}} \leq 1 \Leftrightarrow \sum_{\lambda \in \Lambda_m} a_\lambda^2 \leq 1.$$

Hence,

$$\mathbb{E} \left[ \sup_{t \in \mathcal{B}_m} \nu_n^2(t) \right] = \mathbb{E} \left[ \sup_{\sum_{\lambda} a_\lambda^2 \leq 1} \nu_n^2 \left( \sum_{\lambda} a_\lambda \varphi_\lambda \right) \right] \leq \sum_{\lambda \in \Lambda_m} \mathbb{E} [\nu_n^2(\varphi_\lambda)].$$

According to the Cauchy-Schwartz inequality:

$$\mathbb{E} [\nu_n^2(\varphi_\lambda)] \leq \frac{1}{pn^2\Delta^2} \sum_{j=1}^p \mathbb{E} \left[ \left( \sum_{k=1}^n \varphi_\lambda(\bar{Y}_{(k-1)\Delta}) \int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta} \sigma(X_s) dW_s \right)^2 \right].$$

Let us consider the following filtrations

$$\mathcal{F}_t = \sigma(\eta, W_s, s \leq t) \quad \text{and} \quad \mathcal{G}_t = \sigma(\eta, W_s, s \leq t, \varepsilon_j, j\delta \leq t). \quad (20)$$

For any  $k$ ,  $\bar{Y}_{(k-1)\Delta}$  is  $\mathcal{G}_{(k-1)\Delta}$ -measurable and  $\left( \int_0^t \sigma(X_s) dW_s, t \geq 0 \right)$  is a  $(\mathcal{G}_t)$ -martingale with quadratic variation  $\int_0^t \sigma^2(X_s) ds$ . So

$$\begin{aligned} \mathbb{E} [\nu_n^2(\varphi_\lambda)] &\leq \frac{1}{pn^2\Delta^2} \sum_{j=1}^p \sum_{k=1}^n \mathbb{E} \left[ \varphi_\lambda^2(\bar{Y}_{(k-1)\Delta}) \mathbb{E} \left[ \left( \int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta} \sigma(X_s) dW_s \right)^2 \middle| \mathcal{G}_{kp\delta} \right] \right] \\ &+ \frac{2}{pn^2\Delta^2} \sum_{j=1}^p \sum_{1 \leq l < k \leq n} \mathbb{E} [\varphi_\lambda(\bar{Y}_{(k-1)\Delta}) \varphi_\lambda(\bar{Y}_{(l-1)\Delta}) \\ &\quad \int_{lp\delta+j\delta}^{(l+1)p\delta+j\delta} \sigma(X_s) dW_s \mathbb{E} \left( \int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta} \sigma(X_s) dW_s \middle| \mathcal{G}_{kp\delta} \right)] \\ &\leq \frac{1}{pn^2\Delta^2} \sum_{j=1}^p \sum_{k=1}^n \mathbb{E} \left[ \varphi_\lambda^2(\bar{Y}_{(k-1)\Delta}) \int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta} \sigma^2(X_s) ds \right]. \end{aligned}$$

Thanks to Assumption 3, we obtain

$$\mathbb{E} [\nu_n^2(\varphi_\lambda)] \leq \frac{\sigma_0^2}{n\Delta} \mathbb{E} [\|\varphi_\lambda\|_n^2].$$

As  $\mathbb{E} [\|\varphi_\lambda\|_n^2] = \|\varphi_\lambda\|_{\tilde{\pi}}^2 = 1$ , we have that

$$\mathbb{E} \left[ \sup_{t \in \mathcal{B}_m} \nu_n^2(t) \right] \leq \sigma_0^2 \frac{D_m}{n\Delta}. \quad (21)$$

In the same way, we have that

$$\begin{aligned} \mathbb{E} \left[ \sup_{t \in \mathcal{B}_m} \rho_n^2(t) \right] &\leq \sum_{\lambda \in \Lambda_m} \mathbb{E} [\rho_n^2(\varphi_\lambda)] \\ &\leq \frac{2}{n^2 \Delta^2} \sum_{\lambda \in \Lambda_m} \mathbb{E} \left[ \left( \sum_{k \text{ even}} \varphi_\lambda(\bar{Y}_{(k-1)\Delta}) (\bar{\varepsilon}_{k+1} - \bar{\varepsilon}_k) \right)^2 \right] \\ &\quad + \frac{2}{n^2 \Delta^2} \sum_{\lambda \in \Lambda_m} \mathbb{E} \left[ \left( \sum_{k \text{ odd}} \varphi_\lambda(\bar{Y}_{(k-1)\Delta}) (\bar{\varepsilon}_{k+1} - \bar{\varepsilon}_k) \right)^2 \right]. \end{aligned}$$

The splitting into odd and even indexes has the following advantage: each sum is composed of uncorrelated variables with identical distributions. Moreover, since  $\bar{Y}_{(k-1)\Delta} = \bar{X}_{(k-1)\Delta} + \bar{\varepsilon}_{k-1}$ , thanks to the lag,  $\varphi_\lambda(\bar{Y}_{(k-1)\Delta})$  is independent of  $(\bar{\varepsilon}_{k+1} - \bar{\varepsilon}_k)$ . We can write:

$$\begin{aligned} \mathbb{E} \left[ \sup_{t \in \mathcal{B}_m} \rho_n^2(t) \right] &\leq \frac{2}{n \Delta^2} \sum_{\lambda \in \Lambda_m} \left( \mathbb{E} [\varphi_\lambda^2(\bar{Y}_{(k-1)\Delta})] \mathbb{E} [(\bar{\varepsilon}_{k+1} - \bar{\varepsilon}_k)^2] \right) \\ &\leq 2 \frac{D_m \tau^2}{np \Delta^2}. \end{aligned}$$

So

$$\mathbb{E} \left( \left\| \hat{b}_m - b_A \right\|_n^2 \mathbb{1}_{\Omega_n^c} \right) \leq 3\tilde{\pi}_1 \|b_m - b_A\|_{L^2}^2 + c\sigma_0^2 \frac{D_m}{n\Delta} + c\tau^2 \frac{D_m}{np\Delta^2} + \frac{c\tau^2}{p} + c'\Delta. \quad (22)$$

It remains to bound the risk over  $\Omega_n^c$ . The following lemma will be proved in the Appendix:

**Lemma 3.**

$$\mathbb{P}(\Omega_n^c) \leq \frac{1}{n^2}.$$

Let us set  $e = (e_\Delta, \dots, e_{n\Delta})^*$ , where  $e_{k\Delta} = T_{k\Delta} - b(\bar{Y}_{(k-1)\Delta})$  where  $T_{k\Delta}$  is defined by (13) and  $\Pi_m T = \Pi_m(T_\Delta, \dots, T_{n\Delta})^* = (\hat{b}_m(\bar{Y}_0), \dots, \hat{b}_m(\bar{Y}_{(n-1)\Delta}))^*$  where  $\Pi_m$  is the orthogonal projection over  $S_m$ . Then

$$\begin{aligned} \left\| \hat{b}_m - b_A \right\|_n^2 &= \left\| \Pi_m T - b_A \right\|_n^2 \leq 2 \|b_A - \Pi_m b_A\|_n^2 + 2 \|\Pi_m b_A - \Pi_m T\|_n^2 \\ &\leq 2 \|b_A\|_n^2 + 2 \|\Pi_m e\|_n^2 \leq 2 \|b_A\|_n^2 + 2 \|e\|_n^2. \end{aligned}$$

According to the Cauchy-Schwarz inequality and Lemma 3,

$$\mathbb{E} \left( \|b_A\|_n^2 \mathbb{1}_{\Omega_n^c} \right) \leq \left( \mathbb{E} [b_A^4(\bar{Y}_0)] \mathbb{P}[\Omega_n^c] \right)^{1/2} \leq \frac{c}{n} \sqrt{\mathbb{E} [b_A^4(\bar{Y}_0)]}.$$

According to Assumption 6,  $\mathbb{E} [b_A^4(\bar{Y}_0)] = \int b_A^4(x) \tilde{\pi}(x) dx \leq \tilde{\pi}_1 \|b_A\|_\infty^4$ . Using (14),

$$\mathbb{E} \left( \|e\|_n^2 \mathbb{1}_{\Omega_n^c} \right) \leq \frac{c}{n} \left[ \mathbb{E} (I_{k\Delta}^4) + \mathbb{E} (A_{k\Delta}^4) + \mathbb{E} (B_{k\Delta}^4) + \mathbb{E} (R_{k\Delta}^4) + \mathbb{E} (Z_{k\Delta}^4) \right].$$

By Lemma 2, we obtain:

$$\mathbb{E} \left( \left\| \hat{b}_m - b_A \right\|_n^2 \mathbb{1}_{\Omega_n^c} \right) \leq \frac{1}{n} \left( C + C\Delta + \frac{C}{p} + \frac{\tau^2}{p\Delta^2} + \frac{c\sigma_0^2}{\Delta} \right).$$

Hence, as  $\Delta \rightarrow 0$  and  $p \rightarrow \infty$ :

$$\mathbb{E} \left( \left\| \hat{b}_m - b_A \right\|_n^2 \mathbb{1}_{\Omega_n^c} \right) \leq K \left( \frac{1}{n\Delta} + \frac{1}{np\Delta^2} \right). \quad (23)$$

Collecting terms, we obtain (see (22) and (23)):

$$\begin{aligned} \mathbb{E} \left( \left\| \hat{b}_m - b_A \right\|_n^2 \right) &\leq 3\tilde{\pi}_1 \|b_m - b_A\|_{L^2}^2 + cD_m \left( \frac{\sigma_0^2}{N\delta} + \frac{\tau^2}{Np^2\delta^2} \right) \\ &\quad + K \left( \frac{1}{N\delta} + \frac{1}{Np^2\delta^2} \right) + C \left( p\delta + \frac{1}{p} \right). \end{aligned}$$

## 6.2 Proof of Theorem 2

Let us consider the ball  $\mathcal{B}_{m,m'} = \{t \in S_m + S_{m'}, \|t\|_{\tilde{\pi}} \leq 1\}$ . By definition,

$$\gamma_n(\hat{b}_{\hat{m}}) + pen(\hat{m}) \leq \gamma_n(b_m) + pen(m),$$

so we have that, on  $\Omega_n$  :

$$\begin{aligned} \left\| \hat{b}_{\hat{m}} - b_A \right\|_n^2 \mathbb{1}_{\Omega_n} &\leq 3 \|b_m - b_A\|_n^2 + 2pen(m) - 2pen(\hat{m}) \\ &\quad + 56 \sup_{t \in \mathcal{B}_{m,\hat{m}}} (\nu_n^2(t) + \rho_n^2(t)) + \frac{1}{n} \sum_{k=1}^n (A_{k\Delta}^2 + B_{k\Delta}^2 + I_{k\Delta}^2). \end{aligned}$$

Using Lemma 2, we find that:

$$\begin{aligned} \mathbb{E} \left( \left\| \hat{b}_{\hat{m}} - b_A \right\|_n^2 \mathbb{1}_{\Omega_n} \right) &\leq \inf_{m \in \mathcal{M}_n} 3\tilde{\pi}_1 \|b_m - b_A\|_{L^2}^2 + 2\mathbb{E}((pen(m) - pen(\hat{m})) \mathbb{1}_{\Omega_n}) \\ &\quad + 56\mathbb{E} \left( \sup_{t \in \mathcal{B}_{m,\hat{m}}} (\nu_n^2(t) + \rho_n^2(t)) \mathbb{1}_{\Omega_n} \right) + C \left( \Delta + \frac{1}{p} \right). \end{aligned}$$

It remains to bound the supremum of  $\nu_n^2(t)$  and  $\rho_n^2(t)$  over the random ball  $\mathcal{B}_{m,\hat{m}}$  (depending on  $\hat{m}$ ). Let us introduce a function  $p(m, m')$  such that

$$p(m, m') = \kappa_1 \sigma_0^2 \frac{D_m + D_{m'}}{n\Delta} + \kappa_2 v^2 \frac{D_m + D_{m'}}{np\Delta^2} := p_1(m, m') + p_2(m, m')$$

and the associated penalty function:

$$pen(m) = 28\kappa_1 \sigma_0^2 \frac{D_m}{n\Delta} + 28\kappa_2 v^2 \frac{D_m}{np\Delta^2} := pen_1(m) + pen_2(m).$$

We can write that

$$\begin{aligned} E &:= \mathbb{E} \left[ \left( 24 \sup_{t \in \mathcal{B}_{m,\hat{m}}} \nu_n^2(t) + (pen_1(m) - pen_1(\hat{m})) \right) \mathbb{1}_{\Omega_n} \right] \quad (24) \\ &\leq \mathbb{E} \left[ 24 \left( \sup_{t \in \mathcal{B}_{m,\hat{m}}} \nu_n^2(t) - p_1(m, \hat{m}) \right) \mathbb{1}_{\Omega_n} \right. \\ &\quad \left. + (pen_1(m) - pen_1(\hat{m}) + 28p_1(m, \hat{m})) \mathbb{1}_{\Omega_n} \right]. \end{aligned}$$

As  $\text{pen}_1(m) - \text{pen}_1(\hat{m}) + 28p_1(m, \hat{m}) \leq 2\text{pen}_1(m)$ , it remain to bound the term

$$\mathbb{E} \left[ \left( \sup_{t \in \mathcal{B}_{m, \hat{m}}} \nu_n^2(t) - p_1(m, \hat{m}) \right) \mathbb{1}_{\Omega_n} \right].$$

In the same way, we have that

$$\begin{aligned} F &:= \mathbb{E} \left[ \left( 28 \sup_{t \in \mathcal{B}_{m, \hat{m}}} \rho_n^2(t) + (\text{pen}_2(m) - \text{pen}_2(\hat{m})) \right) \mathbb{1}_{\Omega_n} \right] \\ &\leq \mathbb{E} \left[ \left( \sup_{t \in \mathcal{B}_{m, \hat{m}}} \rho_n^2(t) - p_2(m, \hat{m}) \right) \mathbb{1}_{\Omega_n} \right] + 2\text{pen}_2(m). \end{aligned} \quad (25)$$

The following lemma will be later proved:

**Lemma 4.**

Under Assumptions 1-7, we have:

$$\mathbb{P} \left( \nu_n(t) \geq \eta, \|t\|_n^2 \leq \zeta^2 \right) \leq 2 \exp \left( -n\Delta \frac{\eta^2}{8\sigma_0^2 \zeta^2} \right).$$

The function  $\nu_n(t)$  is defined by (18).

**Lemma 5.**

If Assumptions 1-8 are satisfied, then

$$\mathbb{P} \left( \rho_n(t) \geq \eta, \|t\|_n^2 \leq \zeta^2 \right) \leq 2 \exp \left( -np\Delta^2 \frac{\eta^2}{8v^2 \zeta^2} \right)$$

where  $\rho_n(t)$  is defined by (18).

**Lemma 6.**

Let us consider a function  $f_n : S_{m, m'} \rightarrow \mathbb{R}$ ,  $f_n(t) = \sum_{k=1}^n a_k t (\bar{Y}_{(k-1)\Delta})$ , with  $(a_k, k = 1, \dots, n)$  real numbers. Let us assume that

$$\mathbb{P} \left( |f_n(t)| \geq \eta, \|t\|_n^2 \leq \zeta^2 \right) \leq K \exp \left( -\frac{n\eta^2}{a\zeta^2} \right).$$

Then, denoting  $D_{m, m'} = D_m + D_{m'}$ , there exists a constant  $\kappa > 0$  such that

$$\mathbb{E} \left[ \sup_{t \in \mathcal{B}_{m, m'}} f_n^2(t) - \frac{\kappa a D_{m, m'}}{n} \right]_+ \leq 5.2K \frac{\kappa a e^{-D_{m, m'}}}{n}.$$

Moreover, for any function  $f_n : S_{m+m'} \rightarrow \mathbb{R}$

$$\left[ \sup_{t \in \mathcal{B}_{m, \hat{m}}} f_n^2(t) - \kappa a \left( \frac{D_m + D_{\hat{m}}}{n} \right) \right]_+ \leq \sum_{m' \in \mathcal{M}_n} \left[ \sup_{t \in \mathcal{B}_{m, m'}} f_n^2(t) - \kappa a \left( \frac{D_m + D_{m'}}{n} \right) \right]_+.$$

Applying the Lemmas, as  $\sum_m e^{-D_m} < 1$ , we get that there exist two constants  $\kappa_1$  and  $\kappa_2$  such that

$$\mathbb{E} \left[ \sup_{t \in \mathcal{B}_{m, \hat{m}}} \nu_n^2(t) - p_1(m, \hat{m}) \right]_+ \leq 10.4 \frac{\kappa_1 \sigma_0^2}{n\Delta} \quad (26)$$



and

$$\mathbb{E} \left[ \sup_{t \in \mathcal{B}_{m, \hat{m}}} \rho_n^2(t) - p_2(m, \hat{m}) \right]_+ \leq 10.4 \frac{\kappa_2 v^2}{n\Delta}.$$

So, according to (24) and (25), we have that

$$E \leq 300\kappa_1 \frac{\sigma_0^2}{n\Delta} + 2pen_1(m) \quad \text{and} \quad F \leq 300\kappa_2 \frac{v^2}{np\Delta^2} + 2pen_2(m)$$

For the study on  $\Omega_n^c$ , the end of the proof of Theorem 1 can be used. Collecting terms, we obtain:

$$\mathbb{E} \left( \|\hat{b}_{\hat{m}} - b_A\|_n^2 \right) \leq \inf_{m \in \mathcal{M}_n} (3\tilde{\pi}_1 \|b_m - b_A\|_{L^2}^2 + 4pen(m)) + \frac{C}{n\Delta} + \frac{C'}{np\Delta^2} + C^2\Delta + \frac{C^3}{p}.$$

Setting  $p = \delta^{-1/2}$ , we obtain the result.

### 6.2.1 Proof of Lemma 4

Let us denote

$$H_s^{(1)} = \sum_{k \text{ even}, k \leq n} t (\bar{Y}_{(k-1)\Delta}) \mathbb{1}_{]k\Delta, (k+2)\Delta]}(s) \left( \frac{1}{p} \sum_{j=1}^p \mathbb{1}_{]k\Delta+j\delta, (k+1)\Delta+j\delta]}(s) \sigma(X_s) \right)$$

and

$$H_s^{(2)} = \sum_{k \text{ odd}, k \leq n} t (\bar{Y}_{(k-1)\Delta}) \mathbb{1}_{]k\Delta, (k+2)\Delta]}(s) \left( \frac{1}{p} \sum_{j=1}^p \mathbb{1}_{]k\Delta+j\delta, (k+1)\Delta+j\delta]}(s) \sigma(X_s) \right)$$

The processes

$$M_t^{(1)} = \int_0^t H_s^{(1)} dW_s, \quad M_t^{(2)} = \int_0^t H_s^{(2)} dW_s$$

are two  $(\mathcal{G}_t)$ -martingales. We have that

$$M_{(n+2)\Delta}^{(1)} = \sum_{k \text{ even}, k \leq n} t (\bar{Y}_{(k-1)\Delta}) \left( \frac{1}{p} \sum_{j=1}^p \int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta} \sigma(X_s) dW_s \right)$$

and

$$\langle M^{(1)} \rangle_{(n+2)\Delta} \leq \sum_{k \text{ even}, k \leq n} \sigma_0^2 \Delta t^2 (\bar{Y}_{(k-1)\Delta}).$$

We obtain similar results for  $M_{(n+2)\Delta}^{(2)}$ . Moreover, for any  $t \geq 0$ ,  $i \in \{1, 2\}$ , we have that  $\langle M^{(i)} \rangle_t \leq \sigma_0^2 n \Delta \|t\|_n^2$ , so  $\exp\left(\lambda M_t^{(i)} - \frac{\lambda^2}{2} \langle M^{(i)} \rangle_t\right)$  are  $\mathcal{G}_t$ -martingales with respect to the filtration  $\mathcal{G}_t$ . Therefore, for any  $\lambda > 0$ ,  $\eta, \zeta > 0$ ,  $i \in \{1, 2\}$ :

$$\begin{aligned} \mathbb{P} \left( M_s^{(i)} \geq \eta, \langle M^{(i)} \rangle_s \leq \zeta \right) &\leq \mathbb{P} \left( \exp \left( \lambda M_s^{(i)} - \frac{\lambda^2}{2} \langle M^{(i)} \rangle_s \right) \geq \exp \left( \lambda \eta - \frac{\lambda^2}{2} \zeta \right) \right) \\ &\leq \exp \left( -\lambda \eta + \frac{\lambda^2}{2} \zeta \right). \end{aligned}$$

Minimising with respect to  $\lambda$ , we obtain that

$$\mathbb{P}\left(M_s^{(i)} \geq \eta, \langle M^{(i)} \rangle_s \leq \zeta\right) \leq \exp\left(-\frac{\eta^2}{2\zeta}\right).$$

As

$$\mathbb{P}\left(\nu_n(t) \geq \eta, \|t\|_n^2 \leq \zeta\right) \leq \sum_{i=1,2} \mathbb{P}\left(M_{(n+2)\Delta}^{(i)} \geq \frac{n\Delta\eta}{2}, \langle M^{(i)} \rangle_{(n+2)\Delta} \leq \sigma_0^2 n\Delta\zeta\right),$$

we deduce that

$$\mathbb{P}\left(\nu_n(t) \geq \eta, \|t\|_n^2 \leq \zeta\right) \leq 2 \exp\left(-n\Delta \frac{\eta^2}{8\sigma_0^2\zeta^2}\right).$$

### 6.2.2 Proof of theorem 5

Let us consider the two martingales

$$N_k^{(1)} = \sum_{l=1}^k t(\bar{Y}_{(l-1)\Delta}) \bar{\varepsilon}_l \quad N_k^{(2)} = \sum_{l=1}^k t(\bar{Y}_{(l-1)\Delta}) \bar{\varepsilon}_{l+1}.$$

We can notice that  $N_k^{(1)}$  is a  $\mathcal{G}_{(k+1)\Delta}$ -martingale whereas  $N_k^{(2)}$  is a  $\mathcal{G}_{(k+2)\Delta}$ -martingale. Moreover

$$\mathbb{E}(e^{\lambda \bar{\varepsilon}_1}) = \left(\mathbb{E}(e^{\lambda \varepsilon_0/p})\right)^p \leq \exp\left(\frac{\lambda^2 v^2}{2p}\right),$$

so, for  $i \in \{1, 2\}$ , we have

$$\mathbb{E}\left(\exp(\lambda N_{k+1}^{(i)}) \middle| \mathcal{G}_{k\Delta}\right) \leq \exp\left(\lambda N_k^{(i)}\right) \exp\left(\frac{\lambda^2 v^2 t^2(\bar{Y}_{k\Delta})}{2p}\right). \quad (27)$$

According to the Tchebitchev inequality, for any  $\lambda > 0$ ,

$$\begin{aligned} \mathbb{P}\left(N_n^{(i)} \geq cn, \|t\|_n^2 \leq c'\right) &\leq \mathbb{P}\left[\exp\left(\lambda N_n^{(i)} - \frac{\lambda^2 v^2 n \|t\|_n^2}{2p}\right) \geq \exp\left(\lambda cn - \frac{\lambda^2 v^2 c' n}{2p}\right)\right] \\ &\leq \exp\left(-\lambda cn + \frac{\lambda^2 v^2 c' n}{2p}\right) \mathbb{E}\left[\exp\left(\lambda N_n^{(i)} - \frac{\lambda^2 v^2 n \|t\|_n^2}{2p}\right)\right]. \end{aligned}$$

According to (27), we have that

$$\begin{aligned} \mathbb{E}\left[\exp\left(\lambda N_n^{(i)} - \frac{\lambda^2 v^2 n \|t\|_n^2}{2p}\right)\right] &\leq \mathbb{E}\left(\exp\left(-\frac{\lambda^2 v^2 n \|t\|_n^2}{2p}\right) \mathbb{E}\left[\exp\left(\lambda N_n^{(i)}\right) \middle| \mathcal{G}_{n\Delta}\right]\right) \\ &\leq \mathbb{E}\left(\exp\left(\lambda N_{n-1}^{(i)} - \frac{\lambda^2 v^2}{2p} \sum_{k=1}^{n-1} t^2(\bar{Y}_{(k-1)\Delta})\right)\right). \end{aligned}$$

We deduce by induction that  $\mathbb{E}\left(\exp\left(\lambda N_n^{(i)} - \lambda^2 v^2 n \|t\|_n^2 / (2p)\right)\right) \leq 1$ , therefore,

$$\mathbb{P}\left(N_n^{(i)} \geq cn, \|t\|_n^2 \leq c'\right) \leq \exp\left(-\lambda cn + \frac{\lambda^2 v^2 c' n}{2p}\right).$$

Minimising with respect to  $\lambda$ , we obtain that

$$\mathbb{P}\left(N_n^{(i)} \geq cn, \|t\|_n^2 \leq c'\right) \leq \exp\left(-\frac{npc^2}{2v^2c'}\right).$$

As

$$\mathbb{P}\left(\rho_n(t) \geq \eta, \|t\|_n^2 \leq \zeta^2\right) \leq \sum_{i=1,2} \mathbb{P}\left(N_n^{(i)} \geq \frac{\Delta n \eta}{2}, \|t\|_n^2 \leq \zeta^2\right),$$

we obtain that

$$\mathbb{P}\left(\rho_n(t) \geq \eta, \|t\|_n^2 \leq \zeta^2\right) \leq 2 \exp\left(-np\Delta^2 \frac{\eta^2}{8v^2\zeta^2}\right).$$

### 6.2.3 Proof of Lemma 6

We have that

$$\begin{aligned} \mathbb{E}\left[\sup_{t \in \mathcal{B}_{m,m'}} f_n^2(t) - \frac{\kappa a D}{n}\right]_+ &= \int_0^\infty \mathbb{P}\left[\sup_{t \in \mathcal{B}_{m,m'}} f_n^2(t) - \frac{\kappa a D}{n} \geq x\right] dx \\ &= \kappa a \int_0^\infty \mathbb{P}\left[\sup_{t \in \mathcal{B}_{m,m'}} f_n^2(t) \geq \frac{\kappa a(y+D)}{n}\right] dy. \end{aligned}$$

The remainder of the proof is done in Baraud *et al.* (2001) by a  $L^2$ -chaining technique (see their Section 7, p. 44-47, Lemma 7.1, with  $s^2 = a$  and  $\nu_n = f_n$ ).

### 6.3 Proof of Theorem 3.

We assume that  $p\Delta^2 \rightarrow \infty$  and we set  $e_{n,m} = E_n(\hat{b}_m - b_m) + \rho_n(\hat{b}_m - b_m)$ .

Then

$$\left\|\hat{b}_m - b_A\right\|_n^2 \leq \|b_m - b_A\|_n^2 + 2\nu_n(\hat{b}_m - b_m) + 2e_{n,m}.$$

We will first bound this inequality on  $\Omega_n$ . We obtain that, on  $\Omega_n$ ,

$$\left\|\hat{b}_m - b_A\right\|_n^2 \leq 3\|b_m - b_A\|_n^2 + 48 \sup_{t \in \mathcal{B}_m} \nu_n^2(t) + \frac{48}{n} \sum_{k=1}^n (A_{k\Delta}^2 + I_{k\Delta}^2 + B_{k\Delta}^2 + R_{k\Delta}^2).$$

So, according to Lemma 2 and (26),

$$\mathbb{E}\left(\left\|\hat{b}_m - b_A\right\|_n^2 \mathbb{1}_{\Omega_n}\right) \leq 3\tilde{\pi}_1 \|b_m - b_A\|_{L^2}^2 + c\sigma_0^2 \frac{D_m}{n\Delta} + c'\Delta + \frac{c''}{p\Delta^2}.$$

On  $\Omega_n^c$ , we use (3), and, as  $p\Delta^2 \rightarrow \infty$ , we obtain:

$$\mathbb{E}\left(\left\|\hat{b}_m - b_A\right\|_n^2 \mathbb{1}_{\Omega_n^c}\right) \leq \frac{c\sigma_0^2}{n\Delta}.$$

So, collecting terms, we have:

$$\mathbb{E}\left(\left\|\hat{b}_m - b_A\right\|_n^2\right) \leq 3\tilde{\pi}_1 \|b_m - b_A\|_{L^2}^2 + c\sigma_0^2 \frac{D_m}{n\Delta} + \frac{K\sigma_0^2}{n\Delta} + C\Delta + \frac{C'}{p\Delta^2}$$

which proves Lemma 1.

Now let us define a function

$$p(m, m') = \frac{\kappa \sigma_0^2 (D_m + D_{m'})}{n\Delta}$$

and the associated penalty

$$\text{pen}(m) = \frac{24\kappa\sigma_0^2 D_m}{n\Delta}.$$

Let us set

$$E' := \mathbb{E} \left[ \left( 24 \sup_{t \in \mathcal{E}_{m, \hat{m}}} \nu_n^2(t) + (\text{pen}(m) - \text{pen}(\hat{m})) \right) \mathbb{1}_{\Omega_n} \right].$$

According to the previous proof, we obtain that there exists  $\kappa$  such that  $E' \leq 250\kappa \frac{\sigma_0^2}{n\Delta} + 2\text{pen}(m)$ . Moreover, we know that on  $\Omega_n^c$ ,

$$\mathbb{E} \left( \|\hat{b}_{\hat{m}} - b_A\|_n^2 \mathbb{1}_{\Omega_n^c} \right) \leq C \frac{\sigma_0^2}{n\Delta}.$$

Then, collecting terms, we have:

$$\mathbb{E} \left( \|\hat{b}_{\hat{m}} - b_A\|_n^2 \right) \leq \inf_{m \in \mathcal{M}_n} (3\bar{\pi}_1 \|b_m - b_A\|_{L^2}^2 + 4\text{pen}(m)) + C \frac{\sigma_0^2}{n\Delta} + C' \Delta + \frac{C^2}{p\Delta^2}.$$

Using that  $N = pn$  and  $\Delta = p\delta$ , we obtain that

$$\mathbb{E} \left( \|\hat{b}_{\hat{m}} - b_A\|_n^2 \right) \leq \inf_{m \in \mathcal{M}_n} (3\bar{\pi}_1 \|b_m - b_A\|_{L^2}^2 + 4\text{pen}(m)) + C \frac{\sigma_0^2}{N\delta} + C' p\delta + \frac{C^2}{p^3\delta^2}.$$

This inequality is minimum if  $p\delta \sim 1/(p^3\delta^2)$ , that is if  $p \sim \delta^{-3/4}$ .

## A Additional proofs

### A.1 Proof of Lemma 2

As  $b$  is Lipschitz, we derive from (9) that:

$$\begin{aligned} \mathbb{E} [A_{k\Delta}^2] &= \mathbb{E} \left( \left[ \left( \frac{1}{p} \sum_{j=1}^p b(X_{k\Delta+j\delta}) \right) - b(\bar{X}_{(k-1)\Delta}) \right]^2 \right) \\ &\leq \frac{b_L^2}{p} \sum_{j=1}^p \mathbb{E} \left( [X_{k\Delta+j\delta} - \bar{X}_{(k-1)\Delta}]^2 \right) \leq 2b_L^2 c(2)\Delta \end{aligned}$$

and  $\mathbb{E}(A_{k\Delta}^4) \leq 4b_L^4 c(4)\Delta^2$ . The same arguments are used to prove the following inequalities:

$$\begin{aligned} \mathbb{E} (I_{k\Delta}^2) &\leq \frac{1}{p\Delta} \sum_{j=1}^p \mathbb{E} \left[ \int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta} (b(X_s) - b(X_{k\Delta+j\delta}))^2 ds \right] \\ &\leq b_L^2 c(2)\Delta/2 \end{aligned}$$

and  $\mathbb{E}(I_{k\Delta}^4) \leq b_L^4 c(4)\Delta^2/3$ . Using that  $b$  is Lipschitz and the Rosenthal inequality, we obtain:

$$\begin{aligned}\mathbb{E}[B_{k\Delta}^2] &= \mathbb{E}\left([b(\bar{X}_{(k-1)\Delta}) - b(\bar{X}_{(k-1)\Delta} + \bar{\varepsilon}_{k-1})]^2\right) \\ &\leq b_L^2 \mathbb{E}[\bar{\varepsilon}_{k-1}^2] \leq \frac{b_L^2 \tau^2}{p}.\end{aligned}$$

As, moreover,

$$\mathbb{E}(B_{k\Delta}^4) = \mathbb{E}\left[(b(\bar{X}_{(k-1)\Delta}) - b(\bar{X}_{(k-1)\Delta} + \bar{\varepsilon}_{k-1}))^4\right] \leq b_L^4 \mathbb{E}[\bar{\varepsilon}_{k-1}^4],$$

we have that  $\mathbb{E}(B_{k\Delta}^4) \leq cp^{-3}\mathbb{E}(\varepsilon_1^4) + c\tau^4 p^{-2}$ . We derive from (9) that

$$\mathbb{E}(R_{k\Delta}^2) \leq 4\mathbb{E}\left[\left(\frac{1}{\Delta}\bar{\varepsilon}_k\right)^2\right] = \frac{4\tau^2}{p\Delta^2}$$

and, from the Rosenthal inequality:

$$\mathbb{E}(R_{k\Delta}^4) \leq \frac{c}{\Delta^4} [\mathbb{E}(\bar{\varepsilon}_1^4)] \leq \frac{c\mathbb{E}(\varepsilon_1^4)}{p^3\Delta^4} + c\frac{\tau^4}{p^2\Delta^4}.$$

We have that

$$\mathbb{E}(Z_{k\Delta}^2) = \left(\frac{2}{3} + \frac{1}{p^2}\right) \frac{\mathbb{E}(\sigma_0^2(X_0))}{\Delta}.$$

The Burkholder-Davis-Gundy inequality allows us to write:

$$\begin{aligned}\mathbb{E}(Z_{k\Delta}^4) &\leq \frac{c}{p\Delta^4} \sum_{j=1}^p \mathbb{E}\left[\left(\int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta} \sigma^2(X_s) ds\right)^2\right] \\ &\leq \frac{c}{\Delta^2} \mathbb{E}(\sigma^4(X_0)).\end{aligned}$$

## A.2 Proof of Lemma 3

Let us assume that  $n = 2p_n q_n$ , with  $q_n = \left\lceil \frac{3\ln(n)}{\theta\Delta} \right\rceil + 1$  ( $\theta$  is defined in(8)).

According to Assumption 7,  $\frac{\ln(n)}{n\Delta} \rightarrow 0$  so  $p_n \rightarrow \infty$ .

### Lemma 7.

There exists variables  $\bar{Y}_{k\Delta}^*$  such

1. for  $l = 1, \dots, p_n$ , the random vectors

$$\vec{V}_{l,1}^* = \left(\bar{Y}_{2(l-1)q_n\Delta}^*, \dots, \bar{Y}_{(2l-1)q_n-1\Delta}^*\right) \quad \text{and} \quad \vec{V}_{l,1} = \left(\bar{Y}_{2(l-1)q_n\Delta}, \dots, \bar{Y}_{(2l-1)q_n-1\Delta}\right)$$

have same law, as vectors

$$\vec{V}_{l,2}^* = \left(\bar{Y}_{(2l-1)q_n\Delta}^*, \dots, \bar{Y}_{(2lq_n-1)\Delta}^*\right) \quad \text{and} \quad \vec{V}_{l,2} = \left(\bar{Y}_{(2l-1)q_n\Delta}, \dots, \bar{Y}_{(2lq_n-1)\Delta}\right).$$

2. for each  $a \in \{1, 2\}$ , random vectors  $\vec{V}_{1,a}^*, \dots, \vec{V}_{p_n,a}^*$  are independent.

3.  $\mathbb{P}(\exists k, \bar{Y}_{k\Delta}^* \neq \bar{Y}_{k\Delta}, k = 1, \dots, n) \leq p_n \beta_X(q_n \Delta)$ .

To prove this lemma, we use a Berbee coupling method like in Proposition 5.1 of Viennet (1997). We know that there exists variables  $X_{k\delta}^*$  such that

- For  $l = 1, \dots, p_n$ , the random vectors

$$\vec{U}_{l,1} = (X_{[2(l-1)pq_n+1]\delta}, \dots, X_{(2l-1)pq_n\delta}) \quad \text{and} \quad \vec{U}_{l,1}^* = (X_{[2(l-1)pq_n+1]\delta}^*, \dots, X_{(2l-1)pq_n\delta}^*)$$

have same law, as the vectors

$$\vec{U}_{l,2} = (X_{[(2l-1)pq_n+1]\delta}, \dots, X_{2lpq_n\delta}) \quad \text{and} \quad \vec{U}_{l,2}^* = (X_{[(2l-1)pq_n+1]\delta}^*, \dots, X_{2lpq_n\delta}^*).$$

- For  $l = 1, \dots, p_n$ ,  $a \in \{1, 2\}$ ,  $\mathbb{P}(\vec{U}_{l,a} \neq \vec{U}_{l,a}^*) \leq \beta_X(q_n p \delta) = \beta_X(q_n \Delta)$ .
- For each  $a \in \{1, 2\}$ , random vectors  $\vec{U}_{1,a}^*, \dots, \vec{U}_{p_n,a}^*$  are independent.

As the  $\varepsilon_k$  are independent, variables  $\bar{Y}_{k\Delta}^* = \frac{1}{p} \sum_{j=1}^p X_{k\Delta+j\delta}^* + \varepsilon_{kp+j}$  satisfy (1) and (2). Let us set

$$\Omega^* = \{X_{k\delta} = X_{k\delta}^*, k = 1, \dots, N\} \subset \{\bar{Y}_{k\Delta}^* = \bar{Y}_{k\Delta}\}.$$

We have that  $\mathbb{P}(\Omega^{*c}) \leq \mathbb{P}(\exists k, X_{k\delta} \neq X_{k\delta}^*) \leq p_n \beta_X(q_n \Delta)$ , which proves (3). The end of the proof is done for example in Comte *et al.* (2007), section 7 as long as  $\ln(n)/(n\Delta) \rightarrow 0$  and  $\mathcal{D}_n \leq c_0 \frac{n\Delta}{\ln^2(n)}$ .

### A.3 Proof of Proposition 1

As  $(X_t)_{t \geq 0}$  is stationary,  $(\bar{X}_{k\Delta})_{k \geq 0}$  is stationary for  $p, \delta$  fixed. Since the initial variable  $\eta$  has a density  $\pi$  and  $(X_t)$  has transition densities, the random variables  $\bar{X}_{k\Delta}$  have a density  $\bar{\pi} := \bar{\pi}_{p,\delta}$ . Assumption 6 requires precisions on this density.

If  $(X_t)$  is an Ornstein-Uhlenbeck process, it is centred and Gaussian. Its covariance function is  $(s, t) \rightarrow \sigma^2/2b \exp(-|t-s|)$ . Therefore, the random variable  $\bar{X}_\Delta$  is a centred Gaussian, with variance

$$\bar{\sigma}_\Delta^2 = \frac{\sigma^2}{2p^2b} \sum_{i,j=1}^p e^{-b\delta|j-i|} = \frac{\sigma^2}{2b} \left( \frac{2}{p(1-e^{-b\delta})} - \frac{2e^{-b\delta}(1-e^{-bp\delta})}{p^2(1-e^{-b\delta})^2} - \frac{1}{p} \right).$$

If we set  $x = e^{-b\delta}$ , we find that

$$\bar{\sigma}_\Delta^2 = \frac{\sigma^2}{2b} \left( \frac{1}{p} + \frac{2}{p} \sum_{k=1}^{p-1} x^k \left( 1 - \frac{k}{p} \right) \right)$$

and then

$$\bar{\sigma}_\Delta^2 = \frac{\sigma^2}{2b} \left( 1 + \sum_{j=1}^{\infty} \frac{b^j \delta^j}{j!} \left( \frac{2}{p} \sum_{k=1}^{p-1} k^j - \frac{k^{j+1}}{p} \right) \right).$$

We have that

$$\frac{2}{9} \frac{p^j}{3^j} \leq \frac{2}{p} \sum_{k=1}^{p-1} k^j - \frac{k^{j+1}}{p} \leq 2p^j,$$

so

$$\frac{\sigma^2}{9b}e^{b\Delta/3} \leq \bar{\sigma}_\Delta^2 \leq \frac{\sigma^2}{b}e^{b\Delta}.$$

Then, the invariant density of the variables  $\bar{X}_{k\Delta}$ ,

$$f(x) = \frac{1}{\sqrt{2\pi\bar{\sigma}_\Delta}} \exp\left(-\frac{x^2}{2\bar{\sigma}_\Delta^2}\right),$$

is bounded from below and above. Assumption 6 is satisfied.

We will now prove the second condition. In a first step, it is assumed that  $X_0 = x$ . The process  $y_u = \sqrt{\Delta}^{-1}(X_{u\Delta} - 1)$ , for  $0 \leq u \leq 1$ , is solution of the SDE

$$dy_t = \tilde{b}(y_t)dt + \tilde{\sigma}(y_t)d\tilde{W}_t, \quad y_0 = 0,$$

where  $\tilde{b}(t) = \sqrt{\Delta}b(t\sqrt{\Delta} + x)$ ,  $\tilde{\sigma}(t) = \sigma(t\sqrt{\Delta} + x)$  and  $\tilde{W}_u = W_{u\Delta}/\sqrt{\Delta}$ . Functions  $\tilde{b}(x)$  and  $\tilde{\sigma}(x)$  and their derivatives are bounded and  $\tilde{\sigma}(x)$  is bounded from below by  $\sigma_0$ . Processes  $(X_t)$  and  $(y_t)$  are linked by the equations:

$$(\bar{X}_\Delta, X_\Delta) = \left(x + \sqrt{\Delta}U, x + \sqrt{\Delta}V\right) \quad \text{where} \quad U = \frac{1}{p} \sum_{k=1}^p y_{k/p} \quad \text{and} \quad V = y_1.$$

The following lemma is exactly Theorem 4 p.109 in Gloter and Gobet (2008)

**Lemma 8.**

*Let us consider a process  $(x_t)$  solution of the stochastic differential equation*

$$dx_t = \bar{b}(x_t) + \bar{\sigma}(x_t)dW_t, \quad \text{and} \quad x_0 = x$$

*with  $\bar{\sigma}(x)$  and  $\bar{b}(x)$  are  $\mathcal{C}^3$ ,  $\bar{\sigma}(x)$ ,  $\bar{b}(x)$  and their derivatives are bounded and there exist  $\bar{\sigma}_0 > 0$  such that  $\bar{\sigma}(x) \geq \bar{\sigma}_0$ . Let  $\mu$  be a probability measure on  $[0, 1]$  such that  $\mu([0, 1]) > 0$ . Then, if we set*

$$(\bar{U}, \bar{V}) = \left(\int_0^1 x_s d\mu(s), x_1\right),$$

*there exist constants  $c_1$  and  $c_2$  depending only on the bounds of  $\bar{b}$ ,  $\bar{\sigma}$  and their derivatives such that the density of  $(\bar{U}, \bar{V})$ ,  $\bar{p}_x(\bar{u}, \bar{v})$  satisfies:*

$$c_1^{-1}e^{-c_1(\bar{u}^2 + \bar{v}^2)} \leq \bar{p}_x(\bar{u}, \bar{v}) \leq c_2^{-1}e^{-c_2(\bar{u}^2 + \bar{v}^2)}.$$

As  $\mu(du) = p^{-1} \sum_{j=1}^p \delta_{j/p}(du)$  is a probability measure, there exists constants  $c_1$  and  $c_2$  such that the density  $p_x(u, v)$  of  $(U, V)$  satisfies:

$$c_1^{-1}e^{-c_1(u^2 + v^2)} \leq p_x(u, v) \leq c_2^{-1}e^{-c_2(u^2 + v^2)}.$$

Integrating this result with respect to  $v$ , we easily obtain bounds for the density  $p_x(u)$  of  $U$ : there exists constants  $K_1$  and  $K_2$  depending only on the bounds of  $\bar{b}$ ,  $\bar{\sigma}$  and their derivatives such that

$$K_1^{-1} \exp(-c_1 u^2) \leq p_x(u) \leq K_2^{-1} \exp(-c_2 u^2).$$

Let us denote by  $\bar{\pi}_x(y)$  the density of  $\bar{X}_\Delta$  given  $X_0 = x$ . As  $\bar{X}_\Delta = x + \sqrt{\Delta}U$ ,  $\bar{\pi}_x(y) = \Delta^{-1/2}p_x((y-x)\Delta^{-1/2})$  and, substituting  $u$  by  $(y-x)\Delta^{-1/2}$ , we obtain:

$$\frac{K_1^{-1}}{\sqrt{\Delta}} \exp\left(-c_1 \frac{(y-x)^2}{\Delta}\right) \leq \bar{\pi}_x(y) \leq \frac{K_2^{-1}}{\sqrt{\Delta}} \exp\left(-c_2 \frac{(y-x)^2}{\Delta}\right).$$

The invariant density of  $\bar{X}_\Delta$ ,  $\bar{\pi}$ , satisfies  $\bar{\pi}(y) = \int \bar{\pi}_x(y)\pi(x)dx$ . Integrating the previous inequality, we have:

$$K_1^{-1} \int_{\mathbb{R}} \exp(-c_1 z^2) \pi(y + z\sqrt{\Delta}) dz \leq \bar{\pi}(y) \leq K_2^{-1} \int_{\mathbb{R}} \exp(-c_2 z^2) \pi(y + z\sqrt{\Delta}) dz.$$

Then, we derive from (6) that there exists a constant  $C_2$  depending only on functions  $b$  and  $\sigma$  such that, for any  $y \in \mathbb{R}$ ,

$$\bar{\pi}(y) \leq C_2 \pi_1.$$

In the same way, we deduce from (7) that there exists a constant  $C_1$  such that, for any  $y \in [a_0 - 1, a_1 + 1]$ ,

$$\bar{\pi}(y) \geq C_1 \pi_0.$$

#### A.4 Proof of Proposition 2

*Proof.* Let us denote by  $f_p$  the density of  $\bar{\varepsilon}_0$ . Since  $\int f_p = 1$ , for all  $x \in \mathbb{R}$ ,

$$\tilde{\pi}_{p,\delta}(x) := \tilde{\pi}(x) = \int_{\mathbb{R}} \bar{\pi}(y) f_p(x-y) dy \leq \bar{\pi}_1 \int_{\mathbb{R}} f_p(x-y) dy \leq \bar{\pi}_1.$$

According to the law of large numbers,

$$\lim \int_{\mathbb{R} \setminus [-1,1]} f_p(x) dx = \lim_{p \rightarrow \infty} \mathbb{P}(|\bar{\varepsilon}_0| \geq 1) = 0.$$

Then, there exists a constant  $p_0$  such that, for any  $p \geq p_0$ , we have  $\int_{-1}^1 f_p(x) dx \geq 1/2$ . So, for all  $x \in A$ , for any integer  $p$  larger than  $p_0$ , we have:

$$\tilde{\pi}(x) = \int_{\mathbb{R}} \bar{\pi}(y) f_p(x-y) dy \geq \int_{a_0-1}^{a_1+1} \bar{\pi}(y) f_p(x-y) dy \geq \bar{\pi}_0 \int_{-1}^1 f_p(y) dy \geq \frac{\bar{\pi}_0}{2}. \quad \square$$

#### A.5 Proof of Proposition 3

Let us denote by  $\mathcal{P}_r$  the set of piecewise polynomials on  $[k, k+1]$ ,  $k \in \mathbb{Z}$ , of degree  $r$ ,  $\mathcal{C}^{r-1}$ , defined on  $[-r-1, 1]$  and null on  $[-r-1, -r]$ .

##### Lemma 9.

*The set  $\mathcal{P}_r$  is generated by the spline functions of degree  $r$ ,  $\{(g_r(x+r-k)), 0 \leq k \leq r\}$ .*

*Proof.* Let us consider  $P \in \mathcal{P}_r$ . For any  $z \in \mathbb{Z}$ , we can write  $P_{[z-1,z]}(x) = \sum_{j=0}^r a_j (x-z)^j$  et  $P_{[z,z+1]}(x) = \sum_{j=0}^r b_j (x-z)^j$ . As this polynomial is  $\mathcal{C}^{r-1}$  at point  $z$ , for any  $j < r$ ,  $P_{[z-1,z]}^{(j)}(z) = j!a_j = P_{[z,z+1]}^{(j)}(z) = j!b_j$ . Then



$P_{[z, z+1]}(x) = \sum_{j=0}^r a_j (x-z)^j + (b_r - a_r)(x-z)^r$ , and there exists a constant  $c$  such that

$$P_{[z, z+1]}(x) = P_{[z-1, z]}(x) + c(x-z)^r. \quad (28)$$

Now, let  $R$  and  $Q$  be two functions of  $\mathcal{P}_r$  equal on  $[z-1, z]$ . According to (28), there exists a constant  $c'$  such that

$$R_{[z, z+1]}(x) - Q_{[z, z+1]}(x) = c'(x-z)^r. \quad (29)$$

Moreover, the spline function of degree  $r$ , denoted by  $g_r$ , belongs to  $\mathcal{P}_r$  and, on  $[0, 1]$ , is equal to  $x^r/r!$ . We prove by induction on  $k$  that, over  $[-r-1, r+k]$ , there exist some real numbers  $(c_j, 0 \leq j \leq k-1)$  such that

$$P(x) = \sum_{j=0}^{k-1} c_j g_r(x+r-j).$$

Let us denote  $f_k(x) = \sum_{j=0}^{k-1} c_j g_r(x+r-j)$ . As  $P_{[-r-1, -r]}(x) = 0$ , the induction hypothesis is true for  $k=0$ . Let us assume that it is satisfied for  $k$ . Functions  $f_k$  and  $P$  belongs to  $\mathcal{P}_r$  and are equal on  $[-r+k-1, -r+k]$ , so, according to (29), there exists a constant  $c_k$  such that, on  $[-r+k, -r+k+1]$ ,

$$P(x) - f_k(x) = c_k(x+r-k)^r.$$

Hence, on  $[-r+k, -r+k+1]$ ,  $P(x) = \sum_{j=0}^k r!c_j g_r(x+r-j) = f_{k+1}(x)$ . As the function  $x \rightarrow c_j g_r(x+r-k)$  is equal to 0 on  $[-\infty, -r+k]$ , on the interval  $[-r-1, -r+k+1]$ ,  $P(x) = f_{k+1}(x)$ . There exist some real numbers  $(c_j, 0 \leq j \leq r)$  such that  $P(x) = \sum_{j=0}^r c_j g_r(x+r-j)$ . This property is true for any  $P \in \mathcal{P}_r$ . □

**Lemma 10.**

*Let  $Q$  be a polynomial of degree lower or equal to  $r$  on  $[0, 1]$ . There exists a polynomial  $P \in \mathcal{P}_r$  such that the restriction of  $P$  on  $[0, 1]$  is  $Q$ .*

*Proof.* Let  $P$  be a function of  $\mathcal{P}_r$ . It is known that  $P_{[-r-1, -r]}(x) = 0$ , so according to (28), there exists  $a_0$  such that  $P_{[-r, -r+1]}(x) = a_0(x-r)^r$ . By induction, there exist some real numbers  $a_j, 0 \leq j \leq r$  such that

$$P_{[0, 1]}(x) = \sum_{j=0}^r a_j (x+r-j)^r = \sum_{k=0}^r x^k C_r^k \sum_{j=0}^r a_j (r-j)^{r-k}.$$

The polynomial  $Q$  can be written  $Q(x) = \sum_{k=0}^r b_k x^k$ . Polynomials  $P$  and  $Q$  are equal if and only if for any  $k, 0 \leq k \leq r$ , they satisfy  $\sum_{j=0}^r a_j (r-j)^{r-k} = b_k/C_r^k$ . The real numbers  $(a_j, 0 \leq j \leq r)$  exist if the matrix  $C = \left( (r-j)^{r-k} \right)_{0 \leq k \leq r, 0 \leq j \leq r}$  is invertible. We have:

$$C = \begin{pmatrix} r^r & (r-1)^r & \dots & 1 & 0 \\ r^{r-1} & (r-1)^{r-1} & \dots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r & r-1 & \dots & 1 & 0 \\ 1 & 1 & \dots & 1 & 1 \end{pmatrix},$$

then its determinant is proportional to a Vandermonde's. Therefore, the matrix  $C$  is invertible.  $\square$

Let  $Q$  be a polynomial of degree lower or equal to  $r$  over  $[0, 1]$ . According to Lemma 10, there exists a polynomial  $P \in \mathcal{P}_r$  such that  $P_{[0,1]}(x) = Q(x)$ . According to Lemma 9, there exist constants  $(c_j)_{0 \leq j \leq r}$  such that  $P(x) = \sum_{k=0}^r c_k g_r(x+k)$ . Then

$$Q(x) = \sum_{k=0}^r c_k g_r(x+k) \mathbb{1}_{[0,1]}(x).$$

The space of polynomials of degree lower or equal to  $r$  on  $[0, 1]$  is generated by the spline functions  $\{(f_{0,k}(x) = g_r(x+r-k) \mathbb{1}_{[0,1]}(x)), 0 \leq k \leq r\}$ . The dimension of this space is  $r+1$ , we have  $r+1$  generating functions, functions  $f_{0,k}(x)$ , for  $k = -r, \dots, 0$ , are linearly independent.

Let us now consider a function  $h \in S_m$ , equal to 0. We can write  $h(x) = \sum_{k=-r}^{2^m-1} \alpha_k f_{m,k}(x)$ . The function  $h(x)$  is null on  $[0, 1/2^m]$  so we have

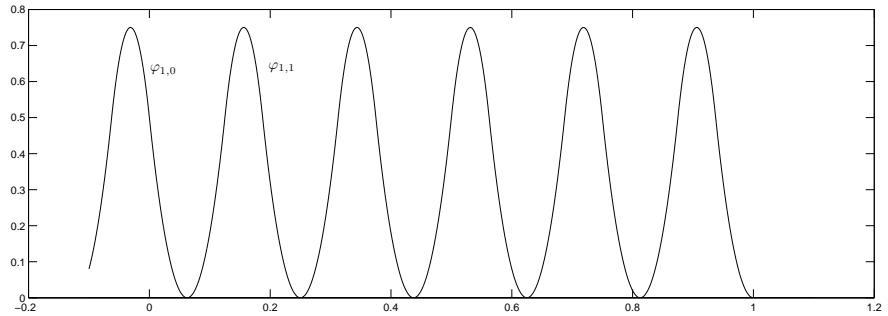
$$h(x) \mathbb{1}_{[0,1/2^m]} = \sum_{k=-r}^0 \alpha_k g_r(2^m x - k) \mathbb{1}_{[0,1/2^m]} = 0.$$

Hence,  $\sum_{k=-r}^0 \alpha_k g_r(y-k) \mathbb{1}_{[0,1]} = 0$ . We know that functions  $g_r(x-k) \mathbb{1}_{[0,1]}$  are linearly independent, so for  $k = -r, \dots, 0$ ,  $\alpha_k = 0$  and  $h(x) = \sum_{k=0}^{2^m-1} \alpha_k f_{m,k}(x)$ . On  $[1/2^m, 2/2^m]$ , the function  $h(x)$  is equal to  $\alpha_1 f_{m,1}(x)$  so  $\alpha_1 = 0$ . We prove by induction on  $k$  that for  $k = -r, \dots, 2^m-1$ ,  $\alpha_k = 0$ . Therefore, functions  $(f_{m,k}, k = -r, \dots, 2^m-1)$  are linearly independent.

## A.6 Proof of Proposition 4

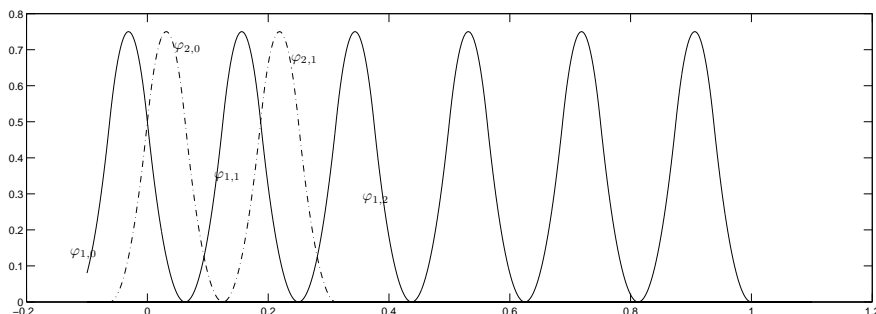
Our aim is to construct an orthonormal basis of  $S_m$ . Let us denote  $\varphi_{a,k}(x) = f_{m,r(k-1)+a}$  for  $k = 0, \dots, [(2^m-1)/r]$  and  $a = 1, \dots, r$ .

The supports of functions  $\varphi_{1,k}$  are  $I_{1,k} = 2^{-m} [r(k-1)+1, rk+1] \cap [0, 1]$ . They are disjoint, so the functions  $\varphi_{1,k}$  are orthogonal, and their length are smaller than  $2^{-m}r$ . Let us set  $\psi_{1,k} = \varphi_{1,k} / \int \varphi_{1,k}^2$ . Those function are orthonormal.



The functions  $\varphi_{2,k}$  are orthogonal to every function  $\psi_{1,k}$  except  $\psi_{1,k}$  and  $\psi_{1,k+1}$ . Using the Gram-Schmidt orthogonalisation procedure, we can construct a function  $\psi_{2,k} = a\varphi_{1,k} + b\varphi_{1,k+1} + c\varphi_{2,k}$  orthogonal to  $\psi_{1,k}$  and  $\psi_{1,k+1}$ .

such that  $\int \psi_{2,k}^2 = 1$ . The functions  $\psi_{2,k}$  et  $\psi_{2,l}$  are orthogonal if  $k \neq l$ . The supports of functions  $\psi_{2,k}$  are  $I_{2,k} = 2^{-m}[r(k-1)+1, r(k+1)+1] \cap [0, 1]$ . The functions  $(\psi_{1,k}, \psi_{2,k})$  are an orthonormal family of  $S_m$ . They generate the same space as the functions  $(\varphi_{1,k}, \varphi_{2,k})$ .



By induction on  $a$ , we can construct an orthonormal basis  $(\psi_{a,k}, a = 1, \dots, r, k = 0, \dots, [(2^m - 1)/r])$  of  $S_m$  such that the supports of the functions  $\psi_{a,k}$  are included in  $2^{-m}[r(k-1)+1, r(k+1)+a-1] \cap [0, 1]$ .

## Références

- Baraud, Y., Comte, F. and Viennet, G. (2001) Model selection for (auto-)regression with dependent data. *ESAIM Probab. Statist.*, 5 pp. 33–49 (electronic).
- Beskos, A. and Roberts, G.O. (2005) Exact simulation of diffusions. *Ann. Appl. Probab.*, 15 (4) pp. 2422–2444.
- Comte, F. and Rozenholc, Y. (2004) A new algorithm for fixed design regression and denoising. *Ann. Inst. Statist. Math.*, 56 (3) pp. 449–473.
- Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007) Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13 (2) pp. 514–543.
- Gloter, A. (2000) Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Statist.*, 4 pp. 205–227 (electronic).
- Gloter, A. and Gobet, E. (2008) LAMN property for hidden processes : the case of integrated diffusions. *Ann. Inst. Henri Poincaré Probab. Stat.*, 44 (1) pp. 104–128.
- Gloter, A. and Jacod, J. (2001) Diffusions with measurement errors. II. Optimal estimators. *ESAIM Probab. Statist.*, 5 pp. 243–260 (electronic).
- Hall, P. and Heyde, C.C. (1980) *Martingale limit theory and its application*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York. Probability and Mathematical Statistics.
- Hoffmann, M. (1999) Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, 79 (1) pp. 135–163.

- Jacod, J., Li, Y., Mykland, P.A., Podolskij, M. and Vetter, M. (2009) Microstructure noise in the continuous case : The pre-averaging approach. *Stochastic Processes and their Applications*.
- Meyer, Y. (1990) *Ondelettes et opérateurs. I*. Actualités Mathématiques. [Current Mathematical Topics]. Hermann, Paris. Ondelettes. [Wavelets].
- Pardoux, E. and Veretennikov, A.Y. (2001) On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29 (3) pp. 1061–1085.
- Schmisser, E. (2009) Penalized nonparametric drift estimation for a multidimensional diffusion process. Preprint 2009-02, MAP5, Université Paris Descartes.
- Viennet, G. (1997) Inequalities for absolutely regular sequences : application to density estimation. *Probab. Theory Related Fields*.
- Zhang, L., Mykland, P.A. and Aït-Sahalia, Y. (2005) A tale of two time scales : determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.*, 100 (472) pp. 1394–1411.

FIG. 1 – Model 1 : Ornstein-Uhlenbeck

$$b(x) = -2x$$

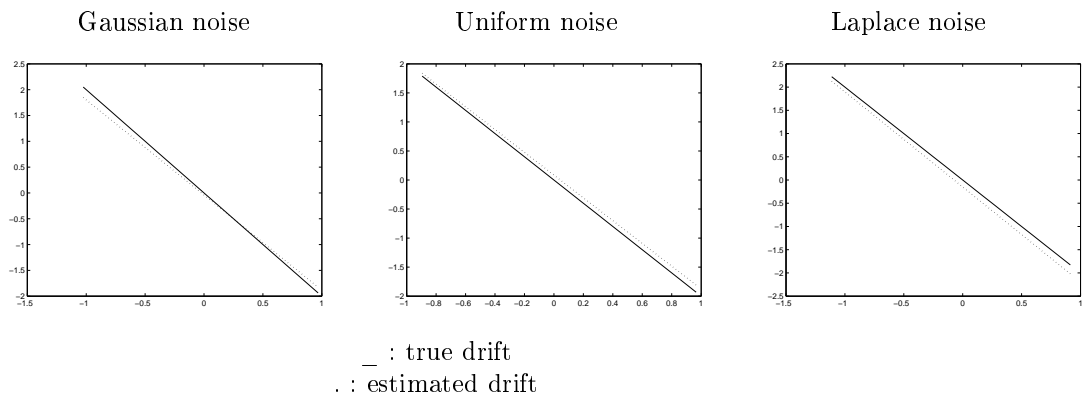


FIG. 2 – Model 2

$$b(x) = -\frac{x}{\sqrt{1+x^2}}$$

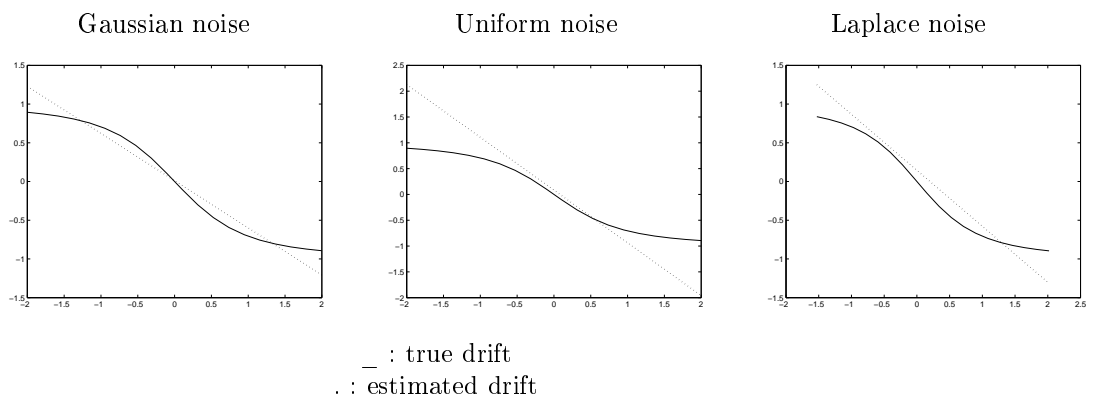


FIG. 3 – Model 3

$$b(x) = \frac{\sinh(x)}{\cosh^2(x)} \left( 1 + \frac{1}{2 \cosh(x)} \right)$$

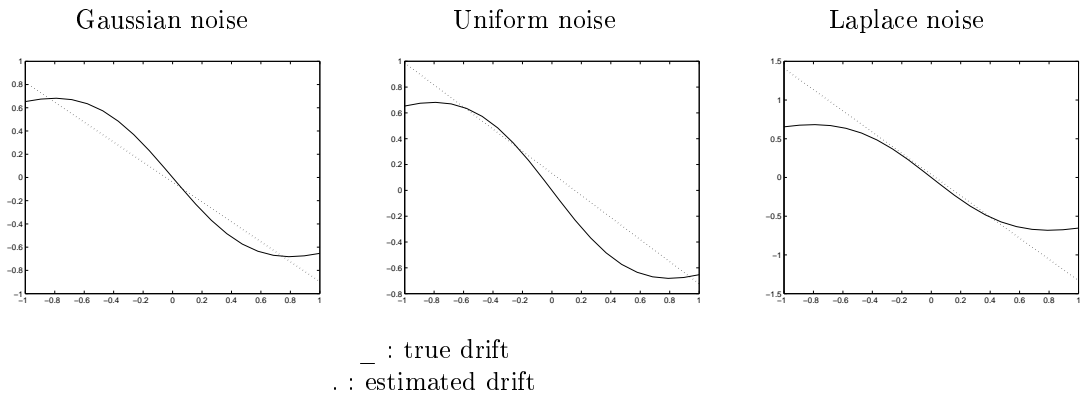
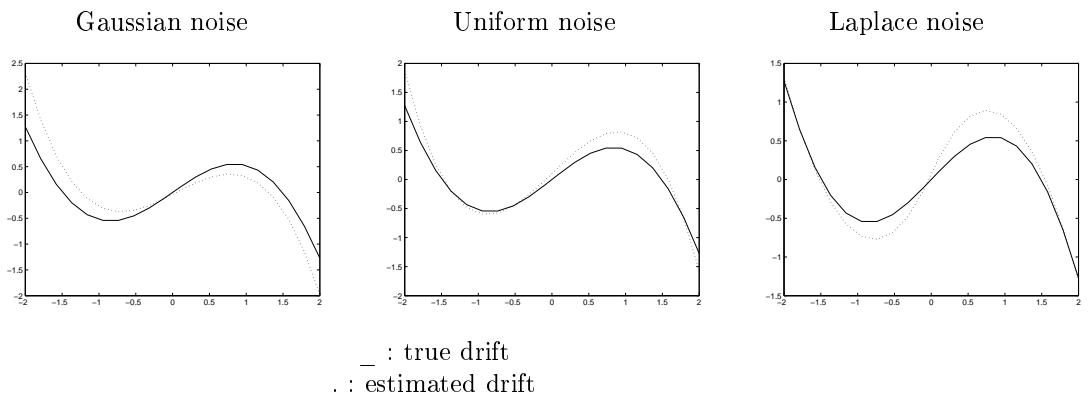


FIG. 4 – Model 4

$$b(x) = -2x + 3 \sin(x)$$



TAB. 1 – Model 1 : Ornstein-Uhlenbeck

$$b(x) = -2x.$$

*ris* : empirical error, *or* : oracle

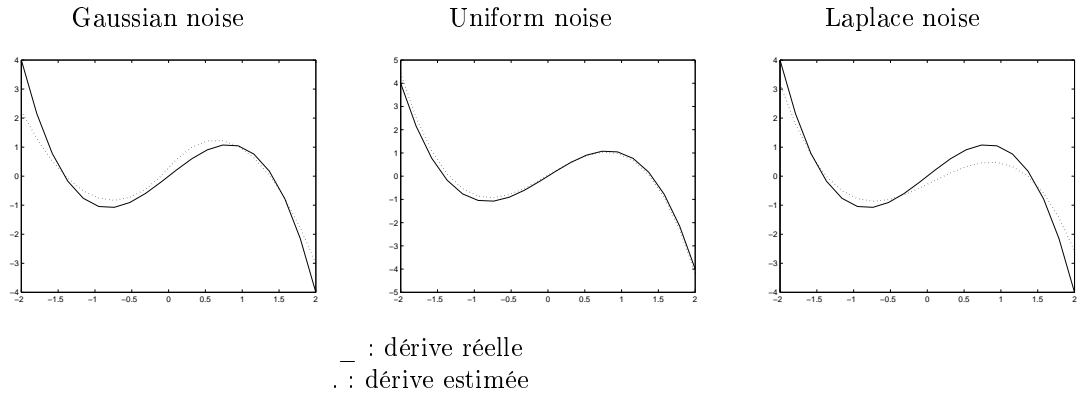
$\delta$	$N$	$p$	Gaussian noise				Uniform noise			
			<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$	<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$
$10^{-4}$	$10^7$	$\delta^{-1/2}$	<b>0.0057</b>	1.14	0.04	1.08	<b>0.0031</b>	1.08	0.04	1.02
$10^{-4}$	$10^6$	$\delta^{-1/2}$	0.024	1.19	0.02	1.04	0.023	1.11	0.02	1.02
$10^{-4}$	$10^5$	$\delta^{-1/2}$	0.27	1.14	0.02	1.02	0.22	1	0	1
$10^{-3}$	$3.10^6$	$\delta^{-1/2}$	0.038	1.01	0.08	1.1	0.013	1.02	0.04	1.12
$10^{-3}$	$3.10^5$	$\delta^{-1/2}$	0.049	1.21	0.1	1.1	0.017	1.03	0.02	1
$10^{-3}$	$3.10^4$	$\delta^{-1/2}$	0.12	1.06	0	1.02	0.10	1.41	0.1	1.06
$10^{-2}$	$10^5$	$\delta^{-1/2}$	0.26	1.01	0.12	1.16	0.10	1	0	1.02
$10^{-2}$	$10^4$	$\delta^{-1/2}$	0.28	1.13	0.18	1.18	0.12	1.05	0.06	1.04
$10^{-2}$	$10^3$	$\delta^{-1/2}$	0.70	1.65	0.24	1.14	0.28	1.15	0.04	1.02
$10^{-6}$	$3.10^7$	$\delta^{-3/4}$	0.065	1	0	1	0.067	1.04	0.02	1
$10^{-4}$	$10^6$	$\delta^{-3/4}$	0.050	1	0	1	0.049	1	0	1
$10^{-4}$	$10^5$	$\delta^{-3/4}$	0.19	1	0	1	0.19	1	0	1

$\delta$	$N$	$p$	Laplace noise			
			<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$
$10^{-4}$	$10^7$	$\delta^{-1/2}$	<b>0.011</b>	1.02	0.02	1
$10^{-4}$	$10^6$	$\delta^{-1/2}$	0.086	2.74	0.2	1.18
$10^{-4}$	$10^5$	$\delta^{-1/2}$	0.24	1	0	1
$10^{-3}$	$3.10^6$	$\delta^{-1/2}$	0.095	1.05	0.36	1.22
$10^{-3}$	$3.10^5$	$\delta^{-1/2}$	0.13	1.48	0.44	1.32
$10^{-3}$	$3.10^4$	$\delta^{-1/2}$	0.32	2.02	0.12	1.18
$10^{-2}$	$10^5$	$\delta^{-1/2}$	0.58	1.026	0.42	1.5
$10^{-2}$	$10^4$	$\delta^{-1/2}$	0.55	1.21	0.34	1.46
$10^{-2}$	$10^3$	$\delta^{-1/2}$	1.58	1.70	0.3	1.14
$10^{-6}$	$3.10^7$	$\delta^{-3/4}$	0.064	1	0	1
$10^{-4}$	$10^6$	$\delta^{-3/4}$	0.051	1	0	1
$10^{-4}$	$10^5$	$\delta^{-3/4}$	0.19	1	0	1

FIG. 5 – Model 5

$$b(x) = -x^3 + 2x$$



TAB. 2 – Model 2

$$b(x) = -\frac{x}{\sqrt{1+x^2}}$$

*ris* : empirical error, *or* : oracle

$\delta$	$N$	$p$	Gaussian noise				Uniform noise			
			<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$	<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$
$10^{-4}$	$10^6$	$\delta^{-1/2}$	0.038	1.31	0.08	1.1	0.037	1.31	0.08	1.08
$10^{-3}$	$3 \cdot 10^4$	$\delta^{-1/2}$	0.16	1.72	0.1	1.1	0.12	1.56	0.16	1.06
$10^{-2}$	$10^3$	$\delta^{-1/2}$	0.29	2.28	0.16	1.18	0.29	1.37	0.08	1.06
$10^{-4}$	$10^6$	$\delta^{-3/4}$	<b>0.034</b>	1.10	0	1.02	<b>0.035</b>	1.14	0	1.04
$10^{-4}$	$10^5$	$\delta^{-3/4}$	0.28	<b>1.01</b>	0	1	0.30	<b>1.07</b>	0.02	1.02
$\delta$	$N$	$p$	Laplace noise							
			<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$				
$10^{-4}$	$10^6$	$\delta^{-1/2}$	0.049	1.44	0.04	1.08				
$10^{-3}$	$3 \cdot 10^4$	$\delta^{-1/2}$	0.42	4.81	0.32	1.24				
$10^{-2}$	$10^3$	$\delta^{-1/2}$	1.28	3.00	0.24	1.28				
$10^{-4}$	$10^6$	$\delta^{-3/4}$	<b>0.034</b>	1.09	0	1				
$10^{-4}$	$10^5$	$\delta^{-3/4}$	0.28	<b>1.01</b>	0	1				



TAB. 3 – Model 3

$$b(x) = \frac{\sinh(x)}{\cosh^2(x)} \left( 1 + \frac{1}{2 \cosh(x)} \right)$$

*ris* : empirical error, *or* : oracle

$\delta$	$N$	$p$	Gaussian noise				Uniform noise			
			<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$	<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$
$10^{-4}$	$10^6$	$\delta^{-1/2}$	0.065	2.61	0.06	1.04	0.060	2.68	0.14	1.12
$10^{-3}$	$3.10^4$	$\delta^{-1/2}$	0.086	1.31	0	1.06	0.088	1.59	0.1	1.08
$10^{-2}$	$10^3$	$\delta^{-1/2}$	0.34	1.59	0.06	1.1	0.21	1.35	0.04	1.04
$10^{-4}$	$10^6$	$\delta^{-3/4}$	<b>0.063</b>	2.29	0	1.06	<b>0.060</b>	2.21	0.04	1.1
$10^{-4}$	$10^5$	$\delta^{-3/4}$	0.19	<b>1.07</b>	0	1	0.19	<b>1.07</b>	0	1
$\delta$	$N$	$p$	Laplace noise							
			<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$				
$10^{-4}$	$10^6$	$\delta^{-1/2}$	0.087	2.41	0.06	0.9				
$10^{-3}$	$3.10^4$	$\delta^{-1/2}$	0.080	2.34	0.18	1.14				
$10^{-2}$	$10^3$	$\delta^{-1/2}$	0.68	2.34	0.16	1.14				
$10^{-4}$	$10^6$	$\delta^{-3/4}$	<b>0.063</b>	2.33	0	1				
$10^{-4}$	$10^5$	$\delta^{-3/4}$	0.19	<b>1.07</b>	0	1				

TAB. 4 – Model 4

$$b(x) = -2x + 3 \sin(x)$$

*ris* : empirical norm, *or* : oracle

$\delta$	$N$	$p$	Gaussian noise				Uniform noise			
			<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$	<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$
$10^{-4}$	$10^6$	$\delta^{-1/2}$	<b>0.11</b>	1.96	0.48	2	<b>0.058</b>	<b>1.18</b>	0.54	2.38
$10^{-3}$	$3.10^4$	$\delta^{-1/2}$	0.31	1.76	0.14	1.2	0.28	1.68	0.24	1.5
$10^{-2}$	$10^3$	$\delta^{-1/2}$	0.79	1.92	0.2	1.2	0.44	1.34	0.04	1
$10^{-4}$	$10^6$	$\delta^{-3/4}$	0.14	2.55	0.34	1.7	0.096	1.69	0.54	2.04
$10^{-4}$	$10^5$	$\delta^{-3/4}$	0.43	<b>1.19</b>	0	1	0.48	1.32	0.04	1.08
$\delta$	$N$	$p$	Laplace noise							
			<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$				
$10^{-4}$	$10^6$	$\delta^{-1/2}$	<b>0.16</b>	2.25	0.4	1.9				
$10^{-3}$	$3.10^4$	$\delta^{-1/2}$	0.55	1.83	0.28	1.38				
$10^{-2}$	$10^3$	$\delta^{-1/2}$	0.69	3.09	0.3	1.3				
$10^{-4}$	$10^6$	$\delta^{-3/4}$	0.22	3.75	0.1	1.22				
$10^{-4}$	$10^5$	$\delta^{-3/4}$	0.44	<b>1.21</b>	0	1				

TAB. 5 – Model 5

$$b(x) = -x^3 + 2x$$

*ris* : empirical error, *or* : oracle

$\delta$	$N$	$p$	Gaussian noise				Uniform noise			
			<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$	<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$
$10^{-4}$	$10^6$	$\delta^{-1/2}$	<b>0.076</b>	1.28	0.7	2.3	<b>0.061</b>	1.25	0.62	2.38
$10^{-3}$	$3.10^4$	$\delta^{-1/2}$	0.58	1.72	0.36	1.84	0.29	1.38	0.36	2.06
$10^{-2}$	$10^3$	$\delta^{-1/2}$	1.38	1.73	0.14	1.08	1.05	1.59	0.12	1.02
$10^{-4}$	$10^6$	$\delta^{-3/4}$	0.26	<b>1.06</b>	0.9	2.04	0.26	<b>1.06</b>	0.9	2.04
$10^{-4}$	$10^5$	$\delta^{-3/4}$	0.80	1.69	0.02	1.04	0.77	1.64	0.04	1.08
$\delta$	$N$	$p$	Laplace noise							
			<i>ris</i>	<i>or</i>	$\hat{m}$	$\hat{r}$				
$10^{-4}$	$10^6$	$\delta^{-1/2}$	<b>0.11</b>	<b>1.08</b>	0.62	2.38				
$10^{-3}$	$3.10^4$	$\delta^{-1/2}$	1.07	1.80	0.22	1.26				
$10^{-2}$	$10^3$	$\delta^{-1/2}$	1.73	1.39	0.12	1.04				
$10^{-4}$	$10^6$	$\delta^{-3/4}$	0.29	1.13	0.8	2.06				
$10^{-4}$	$10^5$	$\delta^{-3/4}$	0.83	1.71	0	1				