HAL open science

Penalized nonparametric drift estimation for a multidimensional diffusion process

Emeline Schmisser

To cite this version:

Emeline Schmisser. Penalized nonparametric drift estimation for a multidimensional diffusion process. Statistics, 2013, 47 (1), pp 61-84. 10.1080/02331888.2011.591931 . hal-00358410

HAL Id: hal-00358410
https://hal.science/hal-00358410

Submitted on 3 Feb 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.

Penalized nonparametric drift estimation for a multidimensional diffusion process

Emeline Schmisser*

*Laboratoire MAP5, Université Paris Descartes.

Abstract

We consider a multi-dimensional diffusion process $(\mathbf{X}_t)_{t\geq 0}$ with drift vector $\mathbf{b}$ and diffusion matrix Σ. This process is observed at $n+1$ discrete times with regular sampling interval Δ. We provide sufficient conditions for the existence and unicity of an invariant density. In a second step, we assume that the process is stationary, and estimate the drift function $\mathbf{b}$ on a compact set K in a nonparametric way. For this purpose, we consider a family of finite dimensional linear subspaces of $L^2(K)$, and compute a collection of drift estimators on every subspace by a penalized least squares approach. We introduce a penalty function and select the best drift estimator. We obtain a bound for the risk of the resulting adaptive estimator. Our method fits for any dimension d, but, for safe of clarity, we focus on the case $d = 2$. We also provide several examples of two-dimensional diffusions satisfying our assumptions, and realize various simulations. Our results illustrate the theoretical properties of our estimators.

Running title : Estimation for multidimensional diffusions

key words : Drift, Model selection, Multidimensional diffusions, Nonparametric estimation, Stationary distribution.

1 Introduction

Let us consider a d-dimensional diffusion process $(\mathbf{X}_t)_{t\geq 0} = \left(X_t^1, X_t^2, ..., X_t^d\right)_{t\geq 0}$ satisfying the stochastic differential equation (SDE):

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)dt + \Sigma(\mathbf{X}_t)d\mathbf{W}_t, \qquad \mathbf{X}_0 = \eta, \tag{1}$$

where $\mathbf{b}(\mathbf{x}) = (b_i(\mathbf{x}))_{1\leq i\leq d}$ is a d-dimensional vector, $\Sigma(\mathbf{x}) = (\sigma_{ij}(\mathbf{x}))_{1\leq i,j\leq d}$ a (d,d) matrix, η a d-dimensional random vector and $(\mathbf{W}_t)$ a Brownian motion of $\mathbb{R}^d$ independent of η. The process $(\mathbf{X}_t)$ is assumed to be strictly stationary and

ergodic. Our aim is to realize nonparametric estimation of the drift function $\mathbf{b}$, given discretized observations.

Several papers deal with nonparametric drift estimation for one-dimensional processes. In particular, Hoffmann (1999) studies nonparametric adaptive estimators using projections on wavelet bases. However, these estimators are difficult to implement numerically. Comte *et al.* (2007) propose different nonparametric estimators, based on a penalized least squares approach. Their estimators are easily computable and have optimality properties. Our aim is to extend these results to multidimensional diffusions. Statistical inference for multidimensional ergodic diffusions is not often studied. This is partly because the characterization and computation of stationary laws is more difficult than for one-dimensional models. From Pardoux and Veretennikov (2001) and Kent (1978), we provide sufficient conditions for the existence and unicity of an invariant density for the SDE (1) (see also Jacobsen (2001)). The observed process is assumed to be strictly stationary and β-mixing, and the discrete observations $\left(\mathbf{X}_0, \mathbf{X}_\Delta, \ldots, \mathbf{X}_{(n+1)\Delta}\right)$ have a sampling interval Δ. Our asymptotic framework is n tends to infinity, $\Delta = \Delta_n$ tends to 0 and $n\Delta$ tends to infinity. The drift vector is estimated on a compact set $K \subset \mathbb{R}^d$. For each component of the drift, $b_i(\mathbf{x}),\ i = 1, \ldots, d$, we define a collection of nonparametric estimators $(\hat{b}_{i,m})_m$ of b_i belonging to a family of linear subspaces $\left(\bar{S}_m\right)_m$ of $L^2(K)$. Then, introducing a penalty, we select the best estimator $\hat{\mathbf{b}}_{\hat{m}}$: the adaptive estimator risk reaches the usual optimal nonparametric rate. The case $\mathbf{b}(x_1, \ldots, x_d) = \sum_{i=1}^d \mathbf{c}_i(x_i)$ is studied separately. For this case, the optimal rate obtained for the estimator is the same as in dimension $d = 1$.

In Section 2, we specify the model and its assumptions. Section 3 describes approximation spaces. Section 4 presents the estimator and studies its risk like in Comte *et al.* (2007). Section 5 gives some indications for the estimation algorithm. Section 6 proposes examples of multidimensional diffusion processes for which data are simulated and estimators are implemented. Numerical simulations results are convincing, even if some theoretical assumptions are not satisfied. Proofs are given in Section 7.

2 Model and assumptions

We consider a diffusion process $(\mathbf{X}_t)$ satisfying (1). We denote by $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^d x_i y_i$ the usual scalar product of $\mathbb{R}^d$, by $|\mathbf{x}|$ the associated norm and by $|M|_{mat}$ a matricial norm. For a matrix M, M^* denotes its transpose. Let us consider the following assumptions:

Assumption 1.
The functions $\Sigma(\mathbf{x})$ *and* $\mathbf{b}(\mathbf{x})$ *are globally Lipschitz:*

$$\exists L,\ \forall (\mathbf{x},\mathbf{y}) \in \left(\mathbb{R}^d\right)^2 \qquad |\Sigma(\mathbf{x}) - \Sigma(\mathbf{y})|_{mat} + |\mathbf{b}(\mathbf{x}) - \mathbf{b}(\mathbf{y})| \leq L\,|\mathbf{x}-\mathbf{y}|\,.$$

Assumption 2.
There exist constants $r > 0$ *and* $\alpha \geq 1$ *such that*

$$\exists M_0 \in \mathbb{R}^+, \quad \forall \mathbf{x}, |\mathbf{x}| > M_0, \qquad \langle \mathbf{b}(\mathbf{x}), \mathbf{x}\rangle \leq -r\,|\mathbf{x}|^{\alpha}\,.$$

Assumption 3.
(i) The diffusion matrix $A(\mathbf{x}) = \Sigma(\mathbf{x})\,\Sigma^*(\mathbf{x}) = (a_{ij}(\mathbf{x}))_{1\leq i,j\leq d}$ *is bounded and positive. Let* σ_0^2 *be such that*

$$\forall \mathbf{x}, \quad Tr(A(\mathbf{x})) \leq \sigma_0^2.$$

(ii) The matrix A *satisfies:*

$$\exists \lambda_-, \lambda_+ > 0,\ \forall \mathbf{x} \in \mathbb{R}^d, \qquad \lambda_-\,|\mathbf{x}|^2 \leq \langle A(\mathbf{x})\,\mathbf{x}, \mathbf{x}\rangle \leq \lambda_+\,|\mathbf{x}|^2\,.$$

Assumption 4.
(i) $\mathbf{b} \in \mathscr{C}^1\left(\mathbb{R}^d, \mathbb{R}^d\right)$, $A \in \mathscr{C}^2\left(\mathbb{R}^d, \mathbb{R}^d \otimes \mathbb{R}^d\right)$.
(ii) There exists a function $V \in \mathscr{C}^2\left(\mathbb{R}^d, \mathbb{R}\right)$ *satisfying:*

$$\mathbf{b} = \frac{1}{2}\left(\sum_{j=1}^{d} \frac{\partial}{dx_j} a_{ij}\right)_{1\leq i\leq d} - A\nabla V.$$

(iii) $c = \exp(-2V(\mathbf{x})) < +\infty$.

Assumption 1 implies existence and uniqueness of a process $(\mathbf{X}_t)_{t\geq 0}$, solution of (1) (see Karatzas and Shreve, Theorem 2.5, p.281 (1988)). Under Assumptions 2-3, there exists a unique invariant density (see Pardoux and Veretennikov (2001), Veretennikov (1987)). Under Assumption 4, we introduce

$$\pi(\mathbf{x}) = c^{-1}\exp(-2V(\mathbf{x}))\,. \tag{2}$$

Proposition 1.
Under Assumptions 1-4, equation (1) *has a unique invariant density, which is* π.

Let us assume:

Assumption 5.
$\eta \sim \pi$.

Under Assumptions 1-5, according to Pardoux and Veretennikov (2001), the process $(\mathbf{X}_t)$ is strictly stationary and exponentially β-mixing: there exists positive constants C, θ such that, for all $t > 0$,

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

where $\beta_X(t)$ denotes the β-mixing coefficient of $(\mathbf{X}_t)$. Recall that, for a stationary diffusion process,

$$\beta_X(s) = \frac{1}{2}\|P_{(X_0,X_s)} - P_{X_0} \otimes P_{X_0}\|_{TV}$$

where $P_{(X_0,X_s)}$ is the joint law of (X_0, X_s), P_{X_0} the law of X_0 and $\|.\|_{TV}$ the total variation distance. Furthermore,

$$\exists \nu > 0, \quad \mathbb{E}\left[\exp\left(\nu|\eta|\right)\right] < \infty.$$

In particular, η has moments of any order. The following proposition is useful for proofs.

Proposition 2.
Under Assumptions 1-5,
$\forall k \geq 1$, $\exists c(k) \in \mathbb{R}$, $\forall h$, $0 < h \leq 1$, $\forall t \geq 0$*:*

$$\mathbb{E}\left(\sup_{s\in[t,t+h]} |\mathbf{b}(\mathbf{X}_s) - \mathbf{b}(\mathbf{X}_t)|^k\right) \leq c\left(k\right) h^{k/2}.$$

This proposition is proved *e.g.* in Gloter (2000) for a one-dimensional diffusion. The extension to the multidimensional case is straightforward.

3 Approximation spaces

Our aim is to estimate the drift function $\mathbf{b}$ on a compact set K of $\mathbb{R}^d$. Without loss of generality, we consider $K = [0,1]^d$. As the stationary density $\pi\left(\mathbf{x}\right)$ is proportional to $\exp\left(-2V\left(\mathbf{x}\right)\right)$, there exist some constants π_0 and π_1 such that, for all $\mathbf{x} \in K$,

$$0 < \pi_0 \leq \pi\left(\mathbf{x}\right) \leq \pi_1 < +\infty.$$

Below, we construct a family $\left(\bar{S}_m\right)_{m\in\mathscr{M}_n}$ of linear subspace of $L^2\left(K\right)$ with $D_m = \dim(\bar{S}_m)$ and where $\mathscr{M}_n$ is the index set of the collection:

$$\mathscr{M}_n = \mathscr{M}_n(r) = \left\{m,\ D_m \leq N_n\right\}, \tag{3}$$

where the maximal dimension N_n will be later precised. For each $m \in \mathscr{M}_n$, we compute an estimator $\hat{\mathbf{b}}_m$ of $\mathbf{b}$, belonging to $\bar{S}_m$. Then we choose the "best" possible estimator by introducing a penalty function $pen(m)$. For simplicity, we describe the case $d = 2$. The construction of a collection would be exactly the same for any dimension d. We start by constructing subspaces of $L^2([0,1])$. Then we deduce subspaces of $L^2([0,1]^2)$.

3.1 Construction of univariate subspaces

For our construction, we use spline functions. We recall here some of their properties. The spline function of degree r is denoted by g_r, where

$$g_r = \mathbb{1}_{[0,1]} * \mathbb{1}_{[0,1]} * \ldots * \mathbb{1}_{[0,1]} \qquad r+1 \text{ times}$$

is the $(r+1)$ times convolution of the indicator function of $[0,1]$. This function is a piecewise polynomial of degree r and of support $[0, r+1]$ and, for any $r \geq 1$, it belongs to $\mathscr{C}^{r-1}$. According to standard properties of convolution, we obtain by induction that, for any integer r, $\int_{-\infty}^{+\infty} g_r(x)\,dx = 1$ and, for all $x \in \mathbb{R}$, $\sum_{k\in\mathbb{Z}} g_r(x-k) = 1$. Let us fix $r \geq 1$ and denote, for $k \in \mathbb{Z}$,

$$f_{0,k} = g_r(. - k)\mathbb{1}_{[0,1]} \quad \text{and} \quad S_0 = \text{Vect}\left\{(f_{0,k}), k \in \mathbb{Z}\right\}.$$

Every function $g \in S_0$ has support in $[0,1]$ and can be written as

$$g = \sum_{k=-r}^{0} \alpha_k f_{0,k}.$$

Functions $f_{0,k}$, for $k \notin \{-r, \ldots, 0\}$, are identically null. Let us define, for $m \in \mathbb{N}, k \in \mathbb{Z}$,

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

which has support $\left[\frac{k}{2^m} \vee 0, \frac{k+r+1}{2^m} \wedge 1\right]$. Non null functions $f_{m,k}$ correspond to $k \in \{-r, -r+1, \ldots, 2^m - 1\}$. Their supports are not disjoint but these functions are linearly independent. Let us set $S_m = \text{Vect}\left\{(f_{m,k}), k \in \mathbb{Z}\right\}$ the vector space generated by functions $f_{m,k}$. Its dimension is $d_m = 2^m + r$. Any function $g \in S_m$ has a support included in $[0,1]$ and can be written as

$$g = \sum_{k=-r}^{2^m-1} \alpha_{m,k} f_{m,k}.$$

Moreover, as $g_r^2 \leq g_r \leq 1$, we have that $\int_{\mathbb{R}} f_{m,k}^2(x)dx \leq \int_{\mathbb{R}} g_r^2(x)dx \leq 1$.

Proposition 3.
There exists a positive constant ϕ_0 such that, for any $t \in S_m$:

$$\|t\|_\infty^2 \leq \phi_0^2 d_m \|t\|_{L^2}^2 .$$

3.2 Bi-variate spaces

Now, we build subspaces $\bar{S}_m$ of $L^2\left([0,1]^2\right)$. For this purpose, we use two different constructions.

The first one uses tensorial products. It is more general than the second and allows us to approximate any function of $L^2\left([0,1]^2\right)$. Let us set $\bar{S}_m = S_m \otimes S_m$. Any function $g \in \bar{S}_m$ can be written as

$$g(x,y) = \sum_{k,l=-r}^{2^m-1} \alpha_{m,k,l} f_{m,k}(x) f_{m,l}(y) .$$

The family $\bar{F}_m = \{f_{m,k}(x) f_{m,l}(y)\}$ is a basis of $\bar{S}_m$ and we have that $D_m = \dim(\bar{S}_m) = d_m^2 = (2^m + r)^2$. For d-variate spaces, we would have

$$D_m = d_m^d = (2^m + r)^d .$$

The second construction only allows us to estimate drift functions of the form

$$\mathbf{b}(x,y) = \mathbf{c}(x) + \mathbf{e}(y) .$$

In that case, our estimator is of the form $\hat{\mathbf{b}}(x,y) = \hat{\mathbf{c}}(x) + \hat{\mathbf{e}}(y)$. As a consequence, we can consider the family

$$\bar{S}_m = \left\{ (x,y) \mapsto g(x,y), \quad g(x,y) = \sum_{k=-r}^{2^m-1} \alpha_{m,k} f_{m,k}(x) + \sum_{l=-r}^{2^m-1} \beta_{m,l} f_{m,l}(y) \right\} .$$

The dimension of the latter $\bar{S}_m$ is $D_m = 2(r + 2^m)$. In the d-variate case, we would have

$$D_m = d(r + 2^m) .$$

In order to use generic notations, we consider $\bar{F}_m = \left\{ (\varphi_{m,k})_{0 \leq k \leq D_m - 1} \right\}$ a basis of $\bar{S}_m$. According to Meyer (1990), spline functions constitute a multiresolution analysis of $L^2(\mathbb{R})$, and, according to Proposition 4, p.50, we deduce

Proposition 4.
Let t belong to the Besov space $B_{2,\infty}^\alpha\left([0,1]^d\right)$ and let t_m be its orthogonal pro-

jection (L^2) over $\bar{S}_m$, with $r \geq \alpha$. Then

$$\exists C > 0, \quad \|t - t_m\|_{L^2} \leq C 2^{-m\alpha}.$$

Remark 1.
We could as well use anisotropic spaces $\bar{S}_{m_1,m_2} = S_{m_1} \otimes S_{m_2}$ depending on m_1, m_2, r_1 and r_2, generated by $F_{m_1,m_2} = \{f_{m_1,k}(x) f_{m_2,l}(y)\}$, with $D_m = (r_1 + 2^{m_1})(r_2 + 2^{m_2})$. Reduced spaces would be

$$\bar{S}_{m_1,m_2} = S_{m_1} + S_{m_2} = \left\{ (x,y) \to g(x,y) = \sum_{k=-r_1}^{2^{m_1}-1} \alpha_{m_1,k} f_{m_1,k}(x) + \sum_{l=-r_2}^{2^{m_2}-1} \beta_{m_2,l} f_{m_2,l}(y) \right\},$$

with dimension $D_m = r_1 + r_2 + 2^{m_1} + 2^{m_2}$. The following result replaces Proposition 4:

Proposition 5.
Let t belongs to the Besov space $B_{2,\infty}^{\alpha_1,\alpha_2}\left([0,1]^d\right)$ with projection t_{m_1,m_2} on $\bar{S}_{m_1,m_2}$. Assume that $r_1 \geq \alpha_1$ and $r_2 \geq \alpha_2$. We have that

$$\exists C > 0, \quad \|t - t_{m_1,m_2}\|_{L^2} \leq C\left(2^{-m_1\alpha_1} + 2^{-m_2\alpha_2}\right).$$

(see Lacour (2007), Lemma 9.)

4 Drift estimation

4.1 Notations

Remember that $K = [0,1]^d$ and set

$$\mathbf{Y}_{k\Delta} = \frac{\mathbf{X}_{(k+1)\Delta} - \mathbf{X}_{k\Delta}}{\Delta}, \qquad \mathbf{Z}_{k\Delta} = \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta} \Sigma(\mathbf{X}_s) d\mathbf{W}_s, \tag{4}$$

and, for any function t from $\mathbb{R}^d$ to $\mathbb{R}$:

$$I_{k\Delta}(t) = \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta} \left(t(\mathbf{X}_s) - t(\mathbf{X}_{k\Delta})\right) ds.$$

By (1), for $i = 1, \ldots, d$,

$$Y_{k\Delta}^i = b_i(\mathbf{X}_{k\Delta}) + I_{k\Delta}(b_i) + Z_{k\Delta}^i$$

where x^i is the ith component of the vector $\mathbf{x}$. In this equation, $b_i(\mathbf{X}_{k\Delta})$ is the main term, $Z_{k\Delta}^i$ a noise term and $I_{k\Delta}(b_i)$ a remainder term. We estimate each component b_i of the drift $\mathbf{b}$. Let us consider, for $i = 1, \ldots, d$, the contrast

$$\gamma_{n,i}(t) = \frac{1}{n}\sum_{k=1}^{n}\left(Y^i_{k\Delta} - t(\mathbf{X}_{k\Delta})\right)^2 \tag{5}$$

and define the estimator $\hat{b}_{m,i}$:

$$\hat{b}_{m,i} = \arg\min_{t\in\bar{S}_m} \gamma_{n,i}(t).$$

We always can find a function $\hat{b}_{m,i}$ which minimizes $\gamma_{n,i}$, but it may be not unique. On the contrary, setting $Y = \left(Y^i_{\Delta}, Y^i_{2\Delta}, \ldots, Y^i_{n\Delta}\right)$, the random vector $\left(\hat{b}_{m,i}\left(\mathbf{X}_{\Delta}\right), \ldots, \hat{b}_{m,i}\left(\mathbf{X}_{n\Delta}\right)\right) = \Pi_m\left(Y\right)$, where Π_m is the Euclidean projection over the subspace $\left\{\left(t\left(\mathbf{X}_{\Delta}\right), \ldots, t\left(\mathbf{X}_{n\Delta}\right)\right), t \in \bar{S}_m\right\}$, is always uniquely defined. For this reason, as in Comte *et al.* (2007), we choose the risk function equal to

$$\mathscr{R}\left(\hat{b}_{m,i}\right) = \mathbb{E}\left(\left\|\hat{b}_{m,i} - b_{i,K}\right\|_n^2\right),$$

where $\|t\|_n^2 = \frac{1}{n}\sum_{k=1}^{n} t^2(\mathbf{X}_{k\Delta})$ and $b_{i,K} = b_i \mathbb{1}_K$.

4.2 Risk of the non-adaptive estimator

Using equations (4) and (5), we obtain that

$$\gamma_{n,i}\left(t\right) - \gamma_{n,i}\left(b_i\right) = \|t - b_i\|_n^2 + \frac{2}{n}\sum_{k=1}^{n}(b_i - t)(\mathbf{X}_{k\Delta})Z^i_{k\Delta} + \frac{2}{n}\sum_{k=1}^{n}(b_i - t)(\mathbf{X}_{k\Delta})I_{k\Delta}\left(b_i\right).$$

Set

$$\nu_{n,i}(t) = \frac{1}{n}\sum_{k=1}^{n} t(\mathbf{X}_{k\Delta})Z^i_{k\Delta}.$$

The orthogonal projection (L^2) of b_i over $\bar{S}_m$ is denoted $b_{m,i}$. We have

$$\begin{aligned}
\gamma_{n,i}(\hat{b}_{m,i}) &\leq \gamma_{n,i}(b_{m,i}) \\
\gamma_{n,i}(\hat{b}_{m,i}) - \gamma_{n,i}(b_i) &\leq \gamma_{n,i}(b_{m,i}) - \gamma_{n,i}(b_i).
\end{aligned}$$

So we can write

$$\|\hat{b}_{m,i} - b_i\|_n^2 \leq \|b_{m,i} - b_i\|_n^2 + 2\nu_{n,i}(\hat{b}_{m,i} - b_{m,i}) + \frac{2}{n}\sum_{k=1}^{n}(\hat{b}_{m,i} - b_{m,i})(\mathbf{X}_{k\Delta})I_{k\Delta}\left(b_i\right).$$

As the supports of $\hat{b}_{m,i}$ and $b_{m,i}$ are included in K,

$$\|\hat{b}_{m,i} - b_{i,K}\|_n^2 \leq \|b_{m,i} - b_{i,K}\|_n^2 + 2\nu_{n,i}(\hat{b}_{m,i} - b_{m,i}) + \frac{2}{n}\sum_{k=1}^{n}(\hat{b}_{m,i} - b_{m,i})(\mathbf{X}_{k\Delta})I_{k\Delta}\left(b_i\right).$$

Let us introduce the asymptotic framework:

Assumption 6.
(i) $\Delta = \Delta_n \to 0$ *and* $n\Delta_n^2 = O(1)$,
(ii) $\frac{n\Delta_n}{ln^2(n)} \to \infty$ *and* $N_n^2 = O\left(\frac{n\Delta_n}{ln^2(n)}\right)$, *where* N_n *is the maximal dimension (see equation* (3)*).*

Theorem 1.
Under Assumptions 1-6, the risk for the drift estimator $\hat{b}_m$ *belonging to a space* $\bar{S}_m$ *satisfies, for all* $i = 1, \ldots, d$*:*

$$\mathbb{E}(\|\hat{b}_{m,i} - b_{i,K}\|_n^2) \leq C\left(\|b_{m,i} - b_{i,K}\|_{L^2}^2 + \sigma_0^2 \frac{D_m}{n\Delta}\right) + C'\Delta + \frac{C''}{n\Delta} \tag{6}$$

with C, C', C'' *constants.*

4.3 Optimization of the dimension space

For given (n,Δ), we wish to select m in order to obtain the best compromise between the bias term $\|b_{m,i} - b_{i,K}\|_{L^2}^2$, and the main variance term, $\frac{D_m}{n\Delta}$. In a first step, we assume that the regularity is known, *i.e.* that $b_{i,K} \in B_{2,\infty}^{\alpha}$ and $\|b_{i,K}\|_{B_{2,\infty}^{\alpha}}^2 \leq 1$, with $r \geq \alpha$. Thanks to Proposition 4, we have that

$$\|b_{i,K} - b_{m,i}\|_{L^2}^2 \leq C2^{-2m\alpha}.$$

Let us distinguish two cases. If $D_m = (2^m + r)^d$, *i.e.* if D_m is of order 2^{md} (tensorial product), m has to satisfy the equation

$$m = \frac{1}{d + 2\alpha} \log_2 (n\Delta).$$

If $D_m = d\,(2^m + r)$ (reduced basis), D_m is of order 2^m and we must have

$$m = \frac{1}{1 + 2\alpha} \log_2 (n\Delta).$$

Using equation (6), we get for a basis obtained by tensorial product

$$\mathbb{E}(\|\hat{b}_{m,i} - b_{i,K}\|_n^2) \leq K\,(n\Delta)^{-2\alpha/(2\alpha+d)} + C'\Delta + \frac{C''}{n\Delta}$$

or, for a reduced basis,

$$\mathbb{E}(\|\hat{b}_{m,i} - b_{i,K}\|_n^2) \leq K\,(n\Delta)^{-2\alpha/(2\alpha+1)} + C'\Delta + \frac{C''}{n\Delta}.$$

In the latter case, our estimator converges at the same rate as for a one-dimensional model.

Remark 2.
Using anisotropic tensorised bases, we would have that

$$\mathbb{E}(\|\hat{b}_{m,i} - b_{i,K}\|_n^2) \leq K\,(n\Delta)^{-2\bar{\alpha}/(2\bar{\alpha}+d)} + C'\Delta + \frac{C''}{n\Delta},$$

with $\bar{\alpha}$ defined by $\frac{d}{\bar{\alpha}} = \sum_{i=1}^{d} \frac{1}{\alpha_i}$. For anisotropic reduced bases, we would find that

$$\mathbb{E}(\|\hat{b}_{m,i} - b_{i,K}\|_n^2) \leq K\,(n\Delta)^{-2\alpha/(2\alpha+1)} + C'\Delta + \frac{C''}{n\Delta},$$

where $\alpha = \min(\alpha_i)$: in this case, using anisotropic bases does not modify the theoretical convergence rate.

4.4 Adaptive estimation

Since we do not know the regularity of b_i, it is important to construct an algorithm which selects automatically m, without any knowledge about the regularity of b_i. For that purpose, we introduce a penalty function $pen(m)$, depending on the dimension D_m, on the number of observations n and on the discretization step Δ. Then, we define

$$\widehat{m}_i = \arg\min_{m \in \mathcal{M}_n} \left[\gamma_{n,i}(\hat{b}_{m,i}) + pen(m)\right]$$

with the penalty function $pen(m)$ such that

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

We denote by $\tilde{b}_i := \hat{b}_{\hat{m}_i,i}$ the resulting estimator. In our simulations, we used $pen(m) = \kappa\sigma_0^2 \frac{D_m}{n\Delta}$ with $\kappa = 5$. (This constant was chosen by numerical calibration, see Comte and Rozenholc (2002, 2004) for a complete discussion).

Theorem 2.
Under Assumptions 1-6, the risk of the adaptive estimator satisfies, for $i = 1, \ldots, d$,

$$\mathbb{E}\left(\|\tilde{b}_i - b_{i,K}\|_n^2\right) \leq C \inf_{m \in \mathcal{M}_n} \left(\|b_{i,K} - b_{m,i}\|_{L^2}^2 + pen(m)\right) + C'\Delta + \frac{C''}{n\Delta}$$

where C, C', C'' are constants.

The adaptive estimator automatically realizes the bias-variance compromise: whenever $b_{i,K}$ belongs to some Besov ball, if $r \geq \alpha$, $\hat{b}_{\hat{m}}$ achieves the optimal corresponding nonparametric rate.

5 Examples and simulation

5.1 Algorithms

In this section, we set $\mathbf{X}_k = \mathbf{X}_{k\Delta}$ and $Y_k = \mathbf{Y}^i_{k\Delta}$ for the ith component of vector $\mathbf{Y}_{k\Delta}$. Any function $g \in \bar{S}_m$ can be written $g(\mathbf{x}) = \sum_{k=0}^{D_m-1} \alpha_{m,k}\varphi_{m,k}(\mathbf{x})$ and is characterized by the vector $\alpha = (\alpha_{m,k})_{k=0,\ldots,D_m-1}$. For computing the estimator $\hat{b}_{m,i}$, we minimize with respect to α the expression

$$\sum_{i=1}^{n}\left(Y_i - \sum_{k=0}^{D_m-1}\alpha_{m,k}\varphi_{m,k}(\mathbf{X}_i)\right)^2.$$

We have to solve, for $l = 0, \ldots, D_m - 1$:

$$\sum_{j=1}^{n}\sum_{k=0}^{D_m-1}\alpha_{m,k}\left[\varphi_{m,k}(\mathbf{X}_j)\,\varphi_{m,l}(\mathbf{X}_j)\right] = \sum_{j=1}^{n} Y_j\left[\varphi_{m,l}(\mathbf{X}_j)\right].$$

Let us set $P = (\varphi_{m,l}(X_j))_{l=0,\ldots,D_m-1,\ j=1,\ldots,n}$ and $Y = (Y_1, \ldots, Y_n)^*$, and solve the equation $PP^*\alpha = PY$.

We simulate a process $(\mathbf{X}_t)$ by an Euler discretization scheme with sampling interval δ and consider $\mathbf{X}_{k\Delta}$ with $\Delta = p\delta$ and p integer, $k = 1, \ldots, n$. The number n of observations varies from 100 to 50 000 and Δ from 0.01 to 0.1. When $n \geq 10000$, we have chosen $\delta = \Delta$, otherwise, $\delta = 0.01$ and $p = 5, 8, 10$. To have enough points in our estimation compact, we keep 95% of the data points, suppressing 5% of extreme values. To estimate the drift on any rectangle K, we have two solutions: either center and renormalize the points $\mathbf{X}_k$ in order to have values on $[0,1]^2$, or modify the functions $\varphi_{m,k}$. We have modified the points $\mathbf{X}_k$, and given afterwards the drift estimator on the rectangle K. Actually, our algorithm is adaptive with respect to m and r.

We let r vary from 1 to $Rmax$ and m from 0 to $\max(\mathscr{M}_n(r))$. Then, to compute $\hat{b}_{m,i}$, we solve in α the equation $PP^*\alpha = PY$ and compute $\gamma_{n,i}(m,r)$ with $pen(m) = pen(m,r)$. We minimize $\gamma_{n,i}(m,r) + pen(m,r)$ with respect to m and r, and return the obtained estimator $\tilde{b}_i = \hat{b}_{\hat{m}_i,i}$.

5.2 Examples

5.2.1 Constant diffusion matrix

We consider the stochastic differential equation

$$d\mathbf{X}_t = -A\nabla V(\mathbf{X}_t)\,dt + \Sigma d\mathbf{W}_t, \qquad \mathbf{X}_0 = \eta, \tag{7}$$

with $A = \Sigma\Sigma^*$ a constant matrix. According to Proposition 1, if the function $\rho(\mathbf{x}) = \exp(-2V(\mathbf{x}))$ is integrable, the process $(\mathbf{X}_t)_{t\geq 0}$ solution of equation (7) is a reversible process, with stationary density proportional to ρ. The following examples are proposed in Fearnhead *et al.* (2007).

Model 1: Ornstein-Uhlenbeck process

We consider the stochastic process of parameters

$$\mathbf{b}(x,y) = \begin{pmatrix} -0.2 & 0.2 \\ 0.1 & -0.2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Its invariant density $\pi(\mathbf{x})$ is proportional to $\exp\left(-\frac{3}{10}x^2 - \frac{3}{5}y^2 + \frac{4}{5}xy\right)$, *i.e.*

$$\pi \sim \mathscr{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{5}{2}\begin{pmatrix} 6 & 4 \\ 4 & 3 \end{pmatrix}\right).$$

We simulate a path with $n = 10000$ points and $\Delta = 0.1$. Figure 3 shows the estimation of the drift first component b_1 with a reduced basis. Figure 4 represents sections for the same model.

Model 2: Double-well potential

Let us consider

$$\mathbf{b}(x,y) = 4\begin{pmatrix} -4x + 4y \\ 4x - y^3 - 3y \end{pmatrix}, \qquad \Sigma = Id.$$

The invariant density is

$$\pi(\mathbf{x}) \propto \exp\left(-2\left(y^2 - 1\right)^2 - 16\left(y - x\right)^2\right).$$

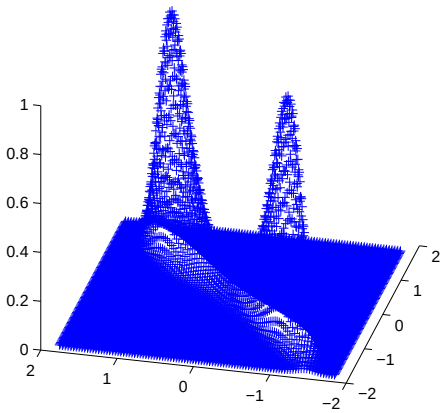
Figure 1 shows its graph (non normalized). This SDE does not satisfy all our assumptions. In particular, the drift is not Lipschitz. We simulate a path with $n = 10000$ points and $\Delta = 0.01$. Figures 5 and 6 show the estimation of the second component b_2, which is a non linear function, with a reduced basis. Risks are computed on the square $[-1.8, 1.8]^2$.

Model 3: Fixman potential

Let us consider

$$\mathbf{b}(x,y) = \begin{pmatrix} -x/2 - 2\cos(2x) - 6\cos(3x) + y/2 \\ x/2 - y/2 \end{pmatrix}, \qquad \Sigma = Id.$$

Figure 1: Stationary density of model 2



The invariant density is

$$\pi(\mathbf{x}) \propto \exp\left(-\frac{1}{2}(y-x)^2 - 2\sin(2x) - 4\sin(3x)\right).$$

Figure 2 shows the graph (non normalized) of this stationary density. We simulate a path with $n = 10000$ points and $\Delta = 0.08$. Figures 7 and 8 represent the estimation of the drift first component b_1, with a reduced basis, on the compact set $[-\pi, \pi]^2$. Risks are computed on the same rectangle.

Figure 2: Stationary density of model 3

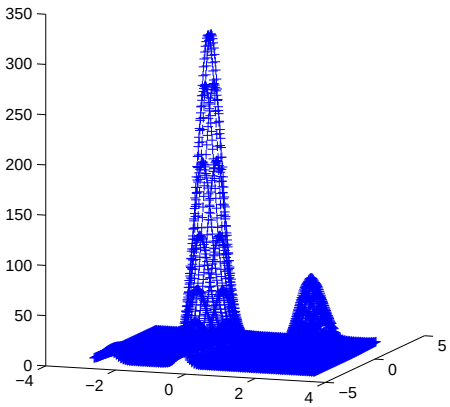
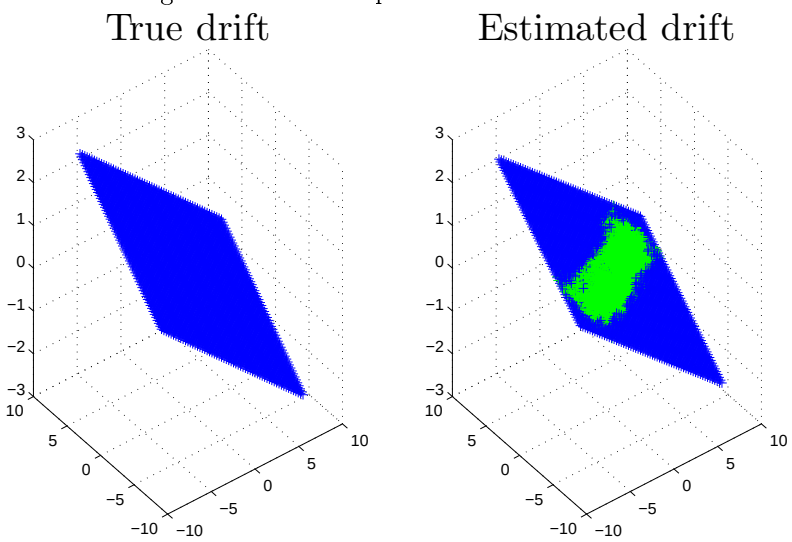
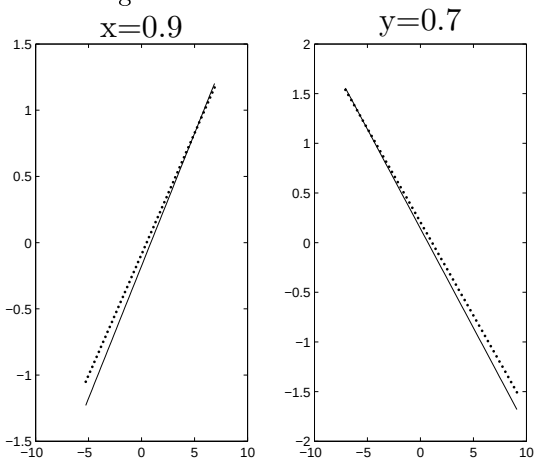


Figure 3: Flows comparison for model 1



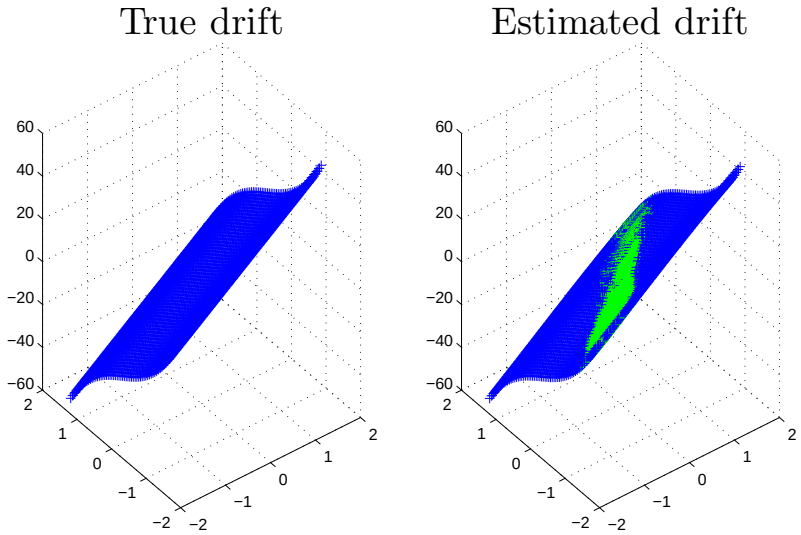
Estimated function: $b_1(x, y) = -0.2x + 0.2y$.
light area: estimated drift at observed points.
dark area: estimated drift on the whole rectangle.

Figure 4: Sections for model 1



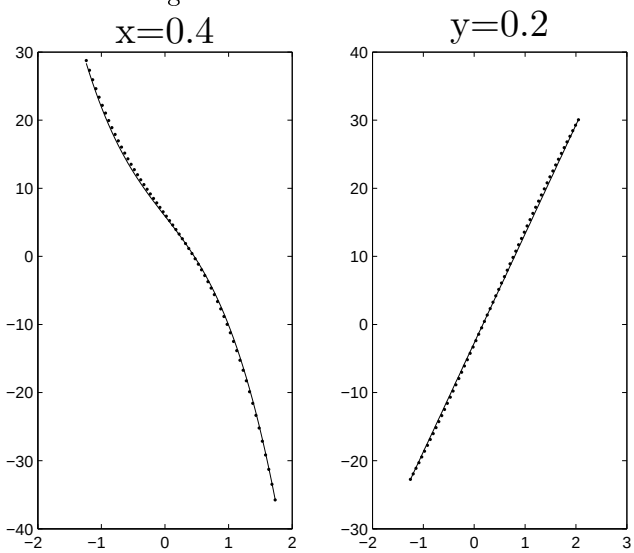
Estimated function: b_1
_: true drift
. : estimated drift

Figure 5: Flows comparison for model 2



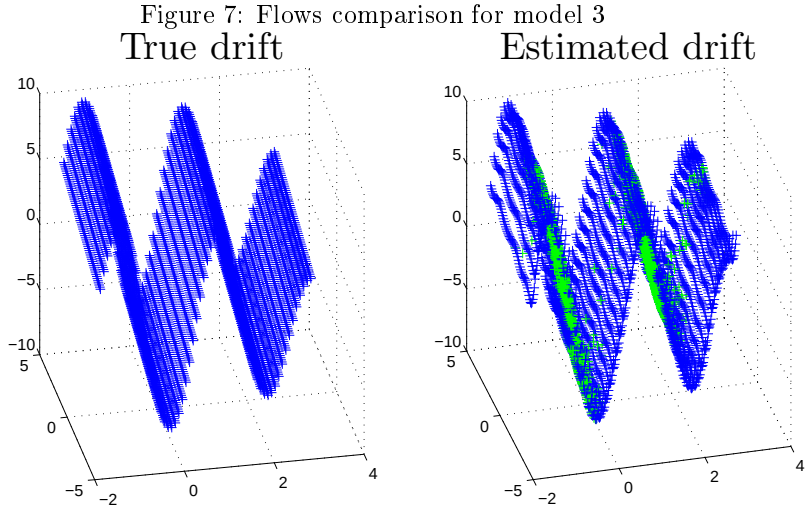
Estimated function: $b_2(x, y) = 16x - 4y^3 - 12y$.
light area: estimated drift at observed points.
dark area: estimated drift on the whole rectangle.

Figure 6: Sections for model 2



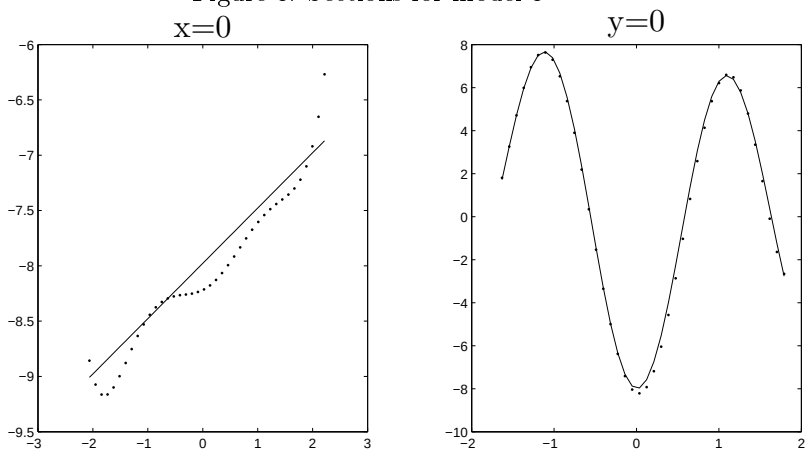
Estimated function: b_2
_: true drift
. : estimated drift

Figure 7: Flows comparison for model 3



Estimated function: $b_1(x, y) = -x/2 - 2\cos(2x) - 6\cos(3x) + y/2$.
light area: estimated drift at observed points.
dark area: estimated drift on the whole rectangle.

Figure 8: Sections for model 3



Estimated function: b_1
_: true drift
. : estimated drift

5.2.2 Non constant diffusion matrix

Model 4: Function of an Ornstein-Uhlenbeck process

We consider the two-dimensional Ornstein-Uhlenbeck process

$$d\mathbf{Z}_t = -B\mathbf{Z}_t dt + \Sigma d\mathbf{W}_t, \qquad \mathbf{Z}_0 = \eta,$$

with $B = (b_{ij})$, $\Sigma = (\sigma_{ij})$ and the vector function $\mathbf{f}(z_1, z_2) = (\exp(z_1), \exp(z_2))^*$. If we set $\mathbf{X}_t = \mathbf{f}(\mathbf{Z}_t)$, according to the Ito formula, $(\mathbf{X}_t)_{t\geq 0}$ satisfies the SDE:

$$\begin{cases} dX_t^1 &= -X_t^1 \left(b_{11} \ln\left(X_t^1\right) + b_{12} \ln\left(X_t^2\right) - \frac{\sigma_{11}^2 + \sigma_{12}^2}{2} \right) + X_t^1 \left(\sigma_{11} dW_t^1 + \sigma_{12} dW_t^2 \right) \\ dX_t^2 &= -X_t^2 \left(b_{21} \ln\left(X_t^1\right) + b_{22} \ln\left(X_t^2\right) - \frac{\sigma_{21}^2 + \sigma_{22}^2}{2} \right) + X_t^2 \left(\sigma_{21} dW_t^1 + \sigma_{22} dW_t^2 \right) \end{cases}$$

The drift function is not the sum of two functions of one variable. Set $B = \begin{pmatrix} 9 & -1 \\ 0 & 7 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 3 & 1 \\ 0 & 2 \end{pmatrix}$. The stationary density associated with $\mathbf{Z}_t$ is $\pi = \mathcal{N}(0, A)$ where $A = \frac{1}{7} \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$, so the invariant density of $\mathbf{X}_t$ is

$$\tilde{\pi}(\mathbf{x}) \propto x^{-\ln(x) + \ln(y) - 1} y^{-2\ln(y) - 1}.$$

Figures 9 and 10 show the estimation of the drift first component b_1 for a tensorised spline basis, for $n = 10000$ observations and $\Delta = 0.1$. Risks are computed on the rectangle $[0, 3.5] \times [0.2, 2.2]$.

Model 5: Multivariate Student invariant distribution

The following example comes from Jacobsen and Sorensen (2001) and Jacobsen (2004). We consider a stochastic process such that

$$d\mathbf{X}_t = -BB^*\mathbf{X}_t dt + B\sqrt{\mathbf{v}(\mathbf{X}_t)} d\mathbf{W}_t, \quad \mathbf{X}_0 = \eta,$$

where B is a constant, symmetric and positive definite matrix, and $\mathbf{v}(\mathbf{x}) = 2(\nu + d - 2)^{-1} \left(\nu + \|\mathbf{x}\|^2\right)$. The associated invariant density is a multivariate Student law with parameter ν, and density

$$\pi_\nu(\mathbf{x}) \propto \left(\nu + \|\mathbf{x}\|^2\right)^{-(\nu+d)/2}.$$

The multivariate Student law with dimension d and degree of freedom ν is the distribution of $\frac{\mathbf{X}}{\sqrt{Y/\nu}}$, where $\mathbf{X}$ has law $\mathcal{N}(\mathbf{0}, Id)$, Y has law $\chi^2(\nu)$ and $\mathbf{X}$ and Y are independent. This model satisfies the equation $\mathbf{b} = -A\nabla V$ with diffusion

matrix $A(\mathbf{x}) = BB^*\mathbf{v}(\mathbf{x})$ and $V(\mathbf{x}) = \frac{\nu+d}{4}\ln\left(\nu + \|\mathbf{x}\|^2\right)$. We choose

$$\nu = 10, \qquad BB^* = \begin{pmatrix} 1 & .9 \\ .9 & 1 \end{pmatrix}$$

and estimate the first component $b_1(x,y) = -x - 0.9y$. Figures 11 and 12 correspond to estimation with reduced spline functions, with $n = 10000$ observations and sample path $\Delta = 0.1$. Risks are computed on the rectangle $[-1.6, 2.4] \times [-2.4, 1.6]$. In this example, $A(\mathbf{x})$ is not bounded. We set $\sigma_0^2 = 8$ for the penalty, which is larger than the maximal value of $Tr(A(\mathbf{x}))$ over the estimation domain.

5.2.3 Model 6: Hamiltonian system

Consider the stochastic process associated with the following SDE:

$$\begin{aligned} dX_t &= Y_t dt \\ dY_t &= -\nabla V(X_t)dt - cY_t dt + \sigma dW_t. \end{aligned}$$

The process parameters are

$$\mathbf{b}(x,y) = \begin{pmatrix} y \\ -\nabla V(x) - cy \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 0 & 0 \\ 0 & \sigma \end{pmatrix}.$$

They do not satisfy Assumptions 1 and 4, but, according to Wu (2001), if the function $\exp(-2V)$ is integrable, the model has a unique invariant density, which is

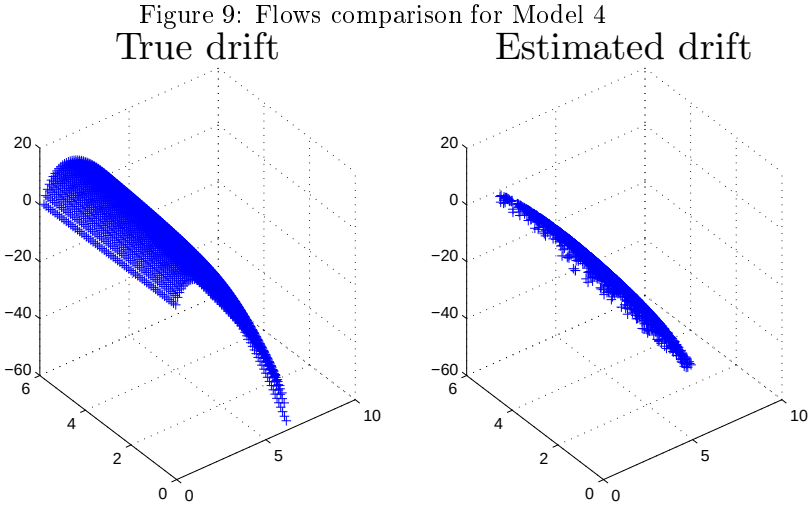
$$\pi(x,y) \propto \exp\left(-\frac{2c}{\sigma^2}\left(\frac{y^2}{2} + V(x)\right)\right).$$

Figure 13 shows the non normalized graph of this function. The infinitesimal generator adjoint L^* can be written, for any function $g \in \mathscr{C}^2$,

$$L^*g = cg - y\partial_x g + (\nabla V(x) + cy)\,\partial_y g + \frac{\sigma^2}{2}\partial_{yy} g.$$

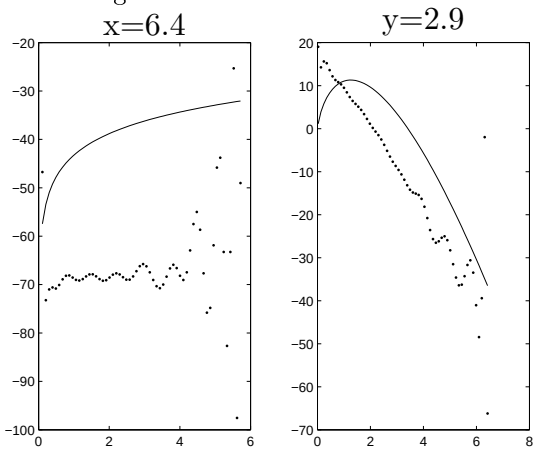
We can easily check than $L^*\pi = 0$. Let us choose $c = 1$, $\sigma = 1$ and $V(x) = x^2 + \sin(3x)$. Figures 14 and 15 show the estimation of the drift second component, b_2, over the compact set $[-\pi, \pi]^2$ for $n = 10000$ points and $\Delta = 0.1$. Risks are computed on the same set.

Figure 9: Flows comparison for Model 4



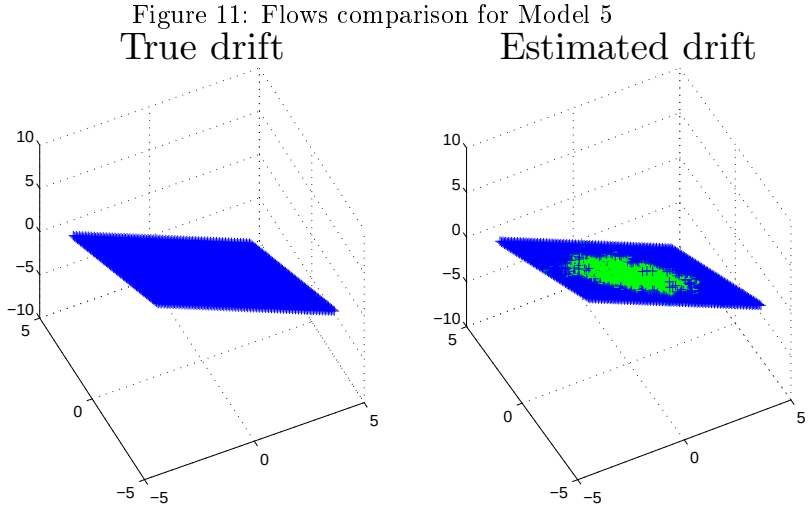
Estimated function: $b_1(x, y) = -9x\ln(x) + \ln(y) + 5x$.
The drift is estimated at simulated points.

Figure 10: Sections for Model 4



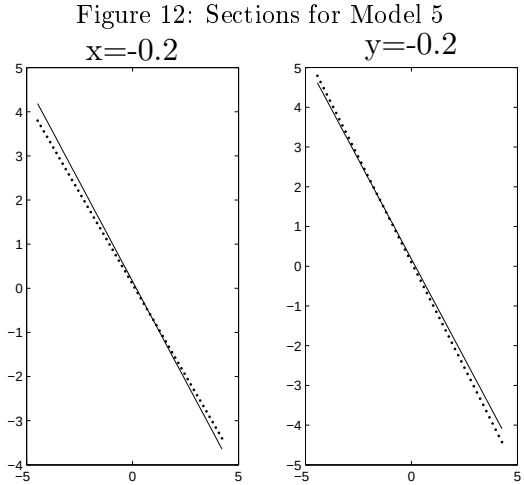
Estimated drift: b_1
_: true drift
. : estimated drift

Figure 11: Flows comparison for Model 5



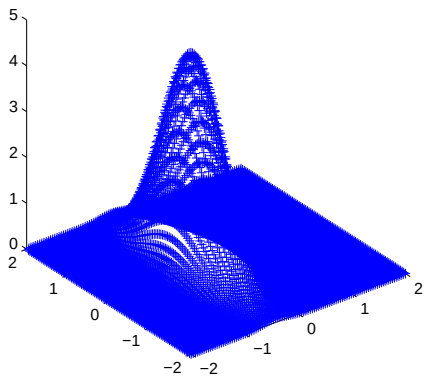
Estimated function: $b_1(x, y) = -x - 0.9y$.
light area: estimated drift at simulated points.
dark area: estimated drift on the whole rectangle.

Figure 12: Sections for Model 5



Estimated function: b_1
_: true drift
. : estimated drift

Figure 13: Stationary density of Model 6



5.3 Results and comments

The values selected by the algorithm are denoted $\hat{m}$ and $\hat{r}$. We compute the error measured by the empirical norm:

$$error = \|\tilde{b}_i - b_{i,K}\|_n^2.$$

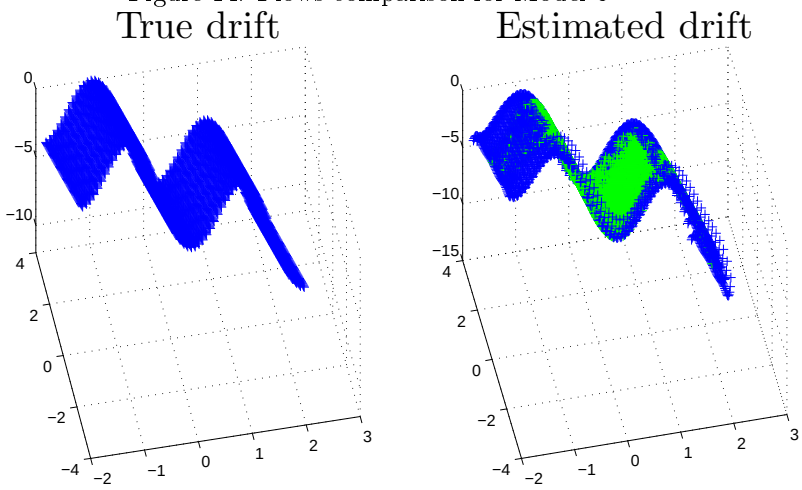
In order to check that the algorithm is adaptive, we also compute

$$emin = \min_m \left\{ \|\hat{b}_{m,i} - b_{i,K}\|_n^2 \right\}.$$

In tables below, we choose a fixed compact $K = K_1 \times K_2$ precised for each example, and compute the mean of $\hat{m}$ and $\hat{r}$. We also compute "ris =mean of $error$" over 50 estimations, and an oracle "or =mean of $error/emin$" over 50 estimations. We used spline functions (reduced sp, tensorised sp), and the piecewise polynomial bases (tensorised poly) described in Comte *et al.* (2007) (for the latter bases, we use $\kappa = 10$).

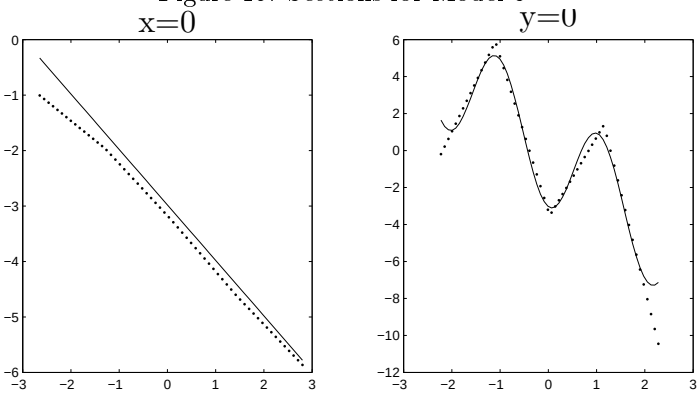
When the drift is linear (Models 1 and 5), our risks are nearly proportional to the product $n\Delta$. Moreover, estimated functions are always linear. Approximating a polynomial function (see Model 2) is a little more difficult and the estimator degree $\hat{r}$ is in general smaller than the true drift one. Moreover, the smaller $n\Delta$, the smaller $\hat{r}$. When the norm of the drift function is large (Model 2), the noise term in Δ becomes important. In Model 4, the drift is a logarithmic function, and, as we use polynomial functions, its estimation is not very good. The drift is not a sum of two functions of one variables, nevertheless, risks computed by reduced bases and tensorised bases are of same order. Estimation

Figure 14: Flows comparison for Model 6



Estimated function: $b_2(x, y) = -2x - 3\cos(x) - y$.
light area: estimated drift at simulated points.
dark area: estimated drift on the whole rectangle.

Figure 15: Sections for Model 6



Estimated function: b_1
_: true drift
. : estimated drift

of trigonometric functions (Models 3 and 6) is not easy. Moreover, unexpectedly, risks decrease when Δ get smaller. In fact, the smaller the discretization path, the smaller the compact K, and the less observed oscillations of the drift function. The better estimations are obtained with reduced spline functions.

In general, we can not compare oracle values for our four different estimators, because these estimators are not chosen over the same functions spaces. We can observe that the oracle for tensorised bases are pretty good, better than for the reduced spline functions. Nevertheless, risks are in general smaller for the reduced spline functions.

6 Proofs

6.1 Proof of Proposition 1

We have to prove that π is an invariant density. Infinitesimal generator L associated with SDE (1) can be written, for any function $f \in \mathscr{C}_c^2(\mathbb{R}^d, \mathbb{R}) := \mathscr{C}_c^2$,

$$Lf = \sum_{i=1}^{d} b_i \partial_i f + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij} \partial_{ij} f,$$

where ∂_i and ∂_{ij} denote partial derivative operators. General form of adjoint L^* of L with respect to $L^2(\mathbb{R}^d) := L^2$ is, for any function $g \in \mathscr{C}^2$:

$$L^* g = -\sum_{i=1}^{d} \partial_i (b_i g) + \frac{1}{2} \sum_{i,j=1}^{d} \partial_{ij} (a_{ij} g).$$

Lemma 1.
If $h \in \mathscr{C}^2$ is a stationary density, then $L^* h = 0$.

Proof. Assume that the process $(\mathbf{X}_t)_{t \geq 0}$ is stationary, with marginal density h. According to the Ito formula, for any function $f \in \mathscr{C}_c^2$:

$$f(\mathbf{X}_t) = f(\mathbf{X}_0) + \int_0^t Lf(\mathbf{X}_s)\, ds + \sum_{i=1}^{d} \int_0^t \partial_i f(\mathbf{X}_s) \sum_{j=1}^{d} \sigma_{ij}(\mathbf{X}_s)\, dW_s^j.$$

Taking expectation, for any $f \in \mathscr{C}_c^2$, we obtain

$$\mathbb{E}(f(\mathbf{X}_t)) = \mathbb{E}(f(\mathbf{X}_0)) + \int_0^t \mathbb{E}(Lf(\mathbf{X}_s))\, ds$$

Table 1: Risks for models with constant variance

Risks for Model 1: $b_1(x, y) = -2x + y$

n	Δ	reduced sp				tensorised sp				tensorised poly			
		ris	or	$\hat{m}$	$\hat{r}$	ris	or	$\hat{m}$	$\hat{r}$	ris	or	$\hat{m}$	$\hat{r}$
10^4	0.1	0.0009	1.05	0	1	0.001	1.00	0	1	0.001	1.00	0	1
10^4	0.01	0.01	1.07	0	1	0.01	1.00	0	1	0.01	1.00	0	1
10^3	0.1	0.01	1.00	0	1	0.01	1.00	0	1	0.01	1.00	0	1
10^3	0.01	0.1	1.7	0	1	0.1	1.00	0	1	0.1	1.11	0	1
10^2	0.1	0.07	1.00	0	1	0.09	1.00	0	1	0.09	1.02	0	1

Risks for Model 2: $b_2(x, y) = 16x - 4y^3 - 12y$

n	Δ	reduced sp				tensorised sp				tensorised poly			
		ris	or	$\hat{m}$	$\hat{r}$	ris	or	$\hat{m}$	$\hat{r}$	ris	or	$\hat{m}$	$\hat{r}$
10^4	0.1	0.009	1.06	0	1	0.01	1.00	0	1	0.01	1.4	0	1
10^4	0.01	0.1	1.03	0	1	0.1	1.04	0	1	0.1	2.2	0	1
10^3	0.1	0.1	1.17	0	1	0.2	1.17	0	1	0.2	3	0	1
10^3	0.01	1.7	1.07	0	1	1.9	1.00	0	1	1.9	46	0	1
10^2	0.1	1.5	1.00	0	1	1.8	1.00	0	1	1.8	47	0	1

Risks for Model 3: $b_2(x, y) = 16x - 4y^3 - 12y$

n	Δ	reduced sp				tensorised sp				tensorised poly			
		ris	or	$\hat{m}$	$\hat{r}$	ris	or	$\hat{m}$	$\hat{r}$	ris	or	$\hat{m}$	$\hat{r}$
5.10^4	0.01	0.02	1.00	0	3	0.05	1.25	0.1	2	0.05	1.58	0	2
10^4	0.01	0.1	1.09	0.1	2.9	0.1	1.00	0	2	0.1	1.28	0.1	1.9
10^3	0.01	1.7	1.77	0	1.3	1.4	1.00	0	1	1.3	1.5	0.2	1
10^3	0.1	7	1.00	0.9	2.1	7	1.01	0.8	1.2	7	1.02	0.9	1

Risks for Model 4: $b_1(x, y) = {}^{x}/_{2} - 2\cos(2x) - 6\cos(3x) + {}^{y}/_{2}$

n	Δ	reduced sp				tensorised sp				tensorised poly			
		ris	or	$\hat{m}$	$\hat{r}$	ris	or	$\hat{m}$	$\hat{r}$	ris	or	$\hat{m}$	$\hat{r}$
10^4	0.08	1.1	1.07	1.5	4.3	6.2	1.00	0.8	2.1	2	1.01	1.7	1.3
10^4	0.05	0.2	1.01	0.2	3.3	1.2	1.01	0.8	1.8	0.6	1.01	1	1.3
10^4	0.01	0.2	2.2	0	1.9	0.3	1.02	0	1.1	0.2	1.04	0	1
10^3	0.08	1.8	1.02	0.6	2.9	3.5	1.00	0.4	1.6	2.1	1.06	1	1.2
10^3	0.05	1.3	1.02	0	1.2	1.4	1.04	0	1.1	1.4	1.10	0	1
10^2	0.01	0.6	1.13	0	1.1	0.7	1.00	0	1	0.7	1.02	0	1

Table 6: Risks for models with non constant variance

Risks for Model 5: $b_1(x,y) = -9x\ln(x) + x\ln(y) + 5x$

n	Δ	reduced sp				tensorised sp				tensorised poly			
		ris	or	$\hat{m}$	$\hat{r}$	ris	or	$\hat{m}$	$\hat{r}$	ris	or	$\hat{m}$	$\hat{r}$
10^4	0.1	43	1.05	2.7	2.9	43	1.05	1.8	2.3	57	1.45	3	1
10^4	0.01	44	1.05	0.5	2.4	45	1.02	0.2	1.6	70	1.61	2	1
10^3	0.1	56	1.03	0.1	1.2	57	1.03	0	1	58	1.05	0.3	1
10^3	0.01	58	1.19	0.3	1.2	58	1.00	0	1	109	2.14	0.7	1
10^2	0.1	63	1.05	0	0	64	1.00	0	1	68	1.19	0.1	1

Risks for Model 6: $b_1(x,y) = -x - 0.9y$

n	Δ	reduced sp				tensorised sp				tensorised poly			
		ris	or	$\hat{m}$	$\hat{r}$	ris	or	$\hat{m}$	$\hat{r}$	ris	or	$\hat{m}$	$\hat{r}$
10^4	0.1	0.01	1.00	0	1	0.01	1.00	0	1	0.01	1.00	0	1
10^4	0.01	0.1	1.00	0	1	0.1	1.00	0	1	0.1	1.00	0	1
10^3	0.1	0.1	1.00	0	1	0.2	1.00	0	1	0.1	1.00	0	1
10^3	0.01	1.3	1.00	0	1	1.6	1.00	0	1	1.6	1.69	0	1
10^2	0.1	1.3	1.02	0	1	1.8	1.00	0	1	1.9	1.87	0	1

Table 9: Risks for Model 7: $b_1(x,y) = -2x - 3\cos(3x) - y$

n	Δ	reduced sp				tensorised sp				tensorised poly			
		ris	or	$\hat{m}$	$\hat{r}$	ris	or	$\hat{m}$	$\hat{r}$	ris	or	$\hat{m}$	$\hat{r}$
10^4	0.08	0.04	1.18	1.6	4.6	1.2	1.00	1.1	1	0.3	1.68	1	2.1
10^4	0.05	0.07	1.08	1.5	3.8	0.8	1.00	1	1	0.3	1.39	1	1.9
10^4	0.01	0.2	1.48	1	2.7	0.6	1.11	0.8	1.1	0.6	1.39	0.9	1
10^3	0.08	0.3	1.28	0.9	2.2	0.6	1.11	0.8	1.1	0.6	1.34	0.8	1.1
10^3	0.05	0.4	1.38	0.7	2.1	0.9	1.27	0.6	1.1	0.8	1.35	0.7	1.1
10^3	0.01	1.2	1.21	0.2	1.2	1.7	1.67	0	1	1.6	1.24	0.2	1
10^2	0.1	1.6	1.39	0.4	1.2	2.4	1.00	0	1	2.0	1.19	0.2	1

and, thanks to stationarity,

$$\int_0^t \mathbb{E}\left(Lf\left(\mathbf{X}_s\right)\right) ds = 0 = t \int Lf(\mathbf{x})h(\mathbf{x})d\mathbf{x} = \langle Lf, h\rangle_{L^2}.$$

As $\langle Lf, h\rangle_{L^2} = \langle f, L^*h\rangle_{L^2}$, we obtain the expected result.

□

Let us explicit the particular form of L^* and the solutions of equation $L^*h = 0$ under Assumption 4.

Lemma 2.
Set $\tilde{b}_i = b_i - \frac{1}{2}\sum_{j=1}^d \partial_j a_{ij}$. Then

$$L^*g = -\sum_{i=1}^d \partial_i\left(\tilde{b}_i g\right) + \frac{1}{2}\sum_{i=1}^d \partial_i\left(\sum_{j=1}^d a_{ij}\partial_j g\right) = -div\left(l(g)\right),$$

where $l(g) = g\tilde{\mathbf{b}} - \frac{1}{2}A\nabla g$. Moreover, under Assumption 4, $\tilde{\mathbf{b}} = -A\nabla V$ and the function

$$h\left(\mathbf{x}\right) \propto \exp\left(-2V\left(\mathbf{x}\right)\right)$$

is solution of the equation $L^*h = 0$.

Proof. We have, for any function $f \in \mathscr{C}_c^2$:

$$Lf = \frac{1}{2}\sum_{i,j=1}^d \partial_i\left(a_{ij}\partial_j f\right) + \sum_{i=1}^d \tilde{b}_i\partial_i f = \frac{1}{2}\mathrm{div}(A\nabla f) + \left\langle \tilde{\mathbf{b}}, \nabla f\right\rangle. \tag{8}$$

By integrating by parts, for any function $g \in \mathscr{C}^2$:

$$\begin{aligned}\langle g, Lf\rangle_{L^2} &= -\sum_{i=1}^d \int_{\mathbb{R}^d} f\partial_i\left(\tilde{b}_i g\right) - \frac{1}{2}\sum_{i,j=1}^d \int_{\mathbb{R}^d} a_{ij}\partial_j f\partial_i g \\ &= -\sum_{i=1}^d \int_{\mathbb{R}^d} f\partial_i\left(\tilde{b}_i g\right) + \frac{1}{2}\sum_{i,j=1}^d \int_{\mathbb{R}^d} f\partial_j\left(a_{ij}\partial_i g\right) = \langle f, L^*g\rangle_{L^2}.\end{aligned}$$

As A is symmetric, we obtain the predicted formula.

Let us solve $l\left(h\right) = 0$ under Assumption 4. As $\tilde{\mathbf{b}} = -A\nabla V$ and A is invertible, we have to solve $2h\nabla V = \nabla h$. We find

$$h\left(\mathbf{x}\right) \propto \exp\left(-2V\left(\mathbf{x}\right)\right).$$

□

Lemma 3.
L is selfadjoint with respect to $L^2_\pi := L^2(\mathbb{R}^d, \pi(\mathbf{x})d\mathbf{x})$. Under Assumptions 1- 4, π is the only invariant density associated with the SDE (1) *(see* (2)*).*

Proof. The adjoint of L with respect to L^2_π is denoted $L^{*\pi}$. We have:

$$L^{*\pi}(g) = \frac{1}{\pi} L^*(\pi g).$$

It is known that $\nabla\pi/2\pi = -\nabla V = A^{-1}\tilde{\mathbf{b}}$. For any function $g \in \mathscr{C}^2_c$,

$$L^{*\pi} g = -\frac{1}{\pi}\mathrm{div}\left(\pi g\tilde{\mathbf{b}} - \frac{1}{2}A\nabla(\pi g)\right).$$

On one hand,

$$\frac{1}{\pi}\mathrm{div}\left(\pi g\tilde{\mathbf{b}}\right) = \mathrm{div}\left(g\tilde{\mathbf{b}}\right) + \left\langle \frac{\nabla\pi}{\pi}, g\tilde{\mathbf{b}}\right\rangle = g\mathrm{div}\left(\tilde{\mathbf{b}}\right) + \left\langle \nabla g, \tilde{\mathbf{b}}\right\rangle + 2g\left\langle A^{-1}\tilde{\mathbf{b}}, \tilde{\mathbf{b}}\right\rangle.$$

On the other hand,

$$\begin{aligned}
\frac{1}{2\pi}\mathrm{div}\left(A\nabla(\pi g)\right) &= \frac{1}{2\pi}\mathrm{div}\left(\pi A\nabla g + gA\nabla\pi\right) \\
&= \frac{1}{2}\mathrm{div}\left(A\nabla g\right) + \frac{1}{2}\left\langle \frac{\nabla\pi}{\pi}, A\nabla g\right\rangle \\
&+ g\frac{1}{2\pi}\mathrm{div}\left(A\nabla\pi\right) + \frac{1}{2}\left\langle \nabla g, A\frac{\nabla\pi}{\pi}\right\rangle.
\end{aligned}$$

As A is symmetric,

$$\left\langle \frac{\nabla\pi}{2\pi}, A\nabla g\right\rangle = \left\langle A^{-1}\tilde{\mathbf{b}}, A\nabla g\right\rangle = \left\langle \tilde{\mathbf{b}}, \nabla g\right\rangle,$$

and the last term is written

$$\left\langle \nabla g, A\frac{\nabla\pi}{2\pi}\right\rangle = \left\langle \tilde{\mathbf{b}}, \nabla g\right\rangle.$$

Furthermore:

$$\frac{1}{2\pi}\mathrm{div}\left(A\nabla\pi\right) = \frac{1}{2\pi}\mathrm{div}\left(2\pi\tilde{\mathbf{b}}\right) = \mathrm{div}\left(\tilde{\mathbf{b}}\right) + \left\langle \frac{\nabla\pi}{\pi}, \tilde{\mathbf{b}}\right\rangle = \mathrm{div}\left(\tilde{\mathbf{b}}\right) + 2\left\langle A^{-1}\tilde{\mathbf{b}}, \tilde{\mathbf{b}}\right\rangle.$$

Collecting terms, we obtain:

$$\frac{1}{2\pi}\mathrm{div}\left(A\nabla(\pi g)\right) = \frac{1}{2}\mathrm{div}\left(A\nabla g\right) + 2\left\langle \tilde{\mathbf{b}}, \nabla g\right\rangle + g\mathrm{div}\left(\tilde{\mathbf{b}}\right) + 2g\left\langle A^{-1}\tilde{\mathbf{b}}, \tilde{\mathbf{b}}\right\rangle,$$

and, using (8),

$$L^{*\pi} g = \left\langle \tilde{\mathbf{b}}, \nabla g\right\rangle + \frac{1}{2}\mathrm{div}\left(A\nabla g\right) = Lg. \quad \square$$

Kent (1978) demonstrates the following result:

Lemma 4.
For a function $h > 0$ of $\mathbb{R}^d$, $L^{*h} = L$ if and only if the transtion density of $(\mathbf{X}_t)$, $p(t, \mathbf{x}, \mathbf{y})$, is h-symmetric, i.e.

$$\forall t, \mathbf{x}, \mathbf{y}, \qquad \frac{p(t, \mathbf{x}, \mathbf{y})}{h(\mathbf{y})} = \frac{p(t, \mathbf{y}, \mathbf{x})}{h(\mathbf{x})}.$$

If h is integrable, the symmetry relation implies that h (normalized) is a stationary density.

As L is selfadjoint with respect to $\pi(\mathbf{x})d\mathbf{x}$, π is the only stationary density.

6.2 Proof of Proposition 3

Any function $t \in S_m$ can be written $t(x) = \sum_{k=-r}^{2^{m-1}} \alpha_{m,k} f_{m,k}(x)$. Consequently, we have

$$\|t\|_{L_2}^2 = 2^m \int_0^1 \left(\sum_{k=-r}^{2^m-1} \alpha_{m,k} g_r \left(2^m x - k\right) \right)^2 dx = \int_0^{2^m} u^2(y) dy,$$

where

$$u(y) = \sum_{k=-r}^{2^{m-1}} \alpha_{m,k} g_r \left(y - k\right) \mathbb{1}_{[0,2^m]}(y).$$

Let us notice that

$$\|t\|_\infty^2 = 2^m \|u\|_\infty^2 \quad \text{and} \quad \|t\|_{L^2}^2 = \|u\|_{L^2}^2. \tag{9}$$

As $d_m = r + 2^m$, we only have to prove that there exists a positive constant c_0 such that $\|u\|_\infty \leq c_0 \|u\|_{L^2}$. The maximum of u is attained at a point $x \in I = [j_0, j_0 + 1]$. We have that

$$\|u\|_\infty = \|u \mathbb{1}_I\|_\infty \quad \text{and} \quad \|u\|_{L^2} \geq \|u \mathbb{1}_I\|_{L^2}.$$

Assume $I = [0, 1]$. We can write

$$u(x) \mathbb{1}_{[0,1]}(x) = \sum_{k=-r}^{0} \alpha_{m,k} g_r \left(x - k\right) \mathbb{1}_{[0,1]}(x).$$

Then $u \mathbb{1}_{[0,1]} \in S_0$, a vector subspace of finite dimension $d_0 = r + 1$. In this subspace, all norms are equivalent, as a consequence we obtain that

$$\exists c_0 > 0, \quad \left\|u \mathbb{1}_{[0,1]}\right\|_\infty \leq c_0 \left\|u \mathbb{1}_{[0,1]}\right\|_{L^2},$$

which, with (9), ends the proof.

6.3 Proof of Theorem 1

All these proofs are adapted of Comte *et al* (2007). Introduce the norm

$$\|t\|_\pi^2 = \int t^2(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}$$

and the set

$$\Omega_n = \left\{\omega, \quad \forall\,(m,m') \in \mathscr{M}_n^2, \quad \forall t \in \bigcup_{m,m'} (\bar{S}_m + \bar{S}_{m'}) \setminus \{0\}\,, \quad \left|\frac{\|t\|_n^2}{\|t\|_\pi^2} - 1\right| \leq \frac{1}{2}\right\}$$

in which norms $\|.\|_n$ and $\|.\|_\pi$ are equivalent: in Ω_n, we have

$$\|t\|_\pi^2 \leq 2\|t\|_n^2 \leq 3\|t\|_\pi^2. \tag{10}$$

Proposition 6.

$$\mathbb{E}(\|\widehat{b}_{m,i} - b_{i,K}\|_n^2 \mathbb{1}_{\Omega_n}) \leq 7\pi_1 \mathbb{E}\left(\|b_{m,i} - b_{i,K}\|_{L^2}^2\right) + 32\sigma_0^2 \frac{D_m}{n\Delta} + 32c\Delta.$$

Proof. We have:

$$\begin{aligned}\|\widehat{b}_{m,i} - b_{i,K}\|_n^2 &\leq \|b_{m,i} - b_{i,K}\|_n^2 + 2\nu_{n,i}(\widehat{b}_{m,i} - b_{m,i}) \\ &+ \frac{2}{n}\sum_{k=1}^n \left(\widehat{b}_{m,i} - b_{m,i}\right)(\mathbf{X}_{k\Delta})\, I_{k\Delta}\,(b_i)\,.\end{aligned}$$

On one hand,

$$\begin{aligned}2\nu_{n,i}(\widehat{b}_{m,i} - b_{m,i}) &\leq 2\|\widehat{b}_{m,i} - b_{m,i}\|_\pi \sup_{t\in\bar{S}_m, \|t\|_\pi = 1} |\nu_{n,i}(t)| \\ &\leq \frac{1}{8}\|\widehat{b}_{m,i} - b_{m,i}\|_\pi^2 + 8 \sup_{t\in\bar{S}_m, \|t\|_\pi = 1} |\nu_{n,i}(t)|^2.\end{aligned}$$

On the other hand, according to the Cauchy-Schwartz inequality,

$$\begin{aligned}\frac{2}{n}\sum_{k=1}^n \left(\widehat{b}_{m,i} - b_{m,i}\right)(\mathbf{X}_{k\Delta})\, I_{k\Delta}\,(b_i) &\leq 2\|\widehat{b}_{m,i} - b_{m,i}\|_n \sqrt{\frac{1}{n}\sum_{k=1}^n I_{k\Delta}\,(b_i)^2} \\ &\leq \frac{1}{8}\|\widehat{b}_{m,i} - b_{m,i}\|_n^2 + \frac{8}{n}\sum_{k=1}^n I_{k\Delta}\,(b_i)^2\,.\end{aligned}$$

Introducing (10), we have that $\|\widehat{b}_{m,i} - b_{m,i}\|_\pi^2 \leq 2\|\widehat{b}_{m,i} - b_{m,i}\|_n^2$ and $\|\widehat{b}_{m,i} -$

$b_{m,i}\|_n^2 \leq 2\|\widehat{b}_{m,i} - b_{i,K}\|_n^2 + 2\|b_{i,K} - b_{m,i}\|_n^2$. Collecting terms, we obtain that

$$\frac{1}{4}\|\widehat{b}_{m,i} - b_{i,K}\|_n^2 \leq \frac{7}{4}\|b_{i,K} - b_{m,i}\|_n^2 + 8 \sup_{t\in\bar{S}_m, \|t\|_\pi=1} |\nu_{n,i}(t)|^2 + \frac{8}{n}\sum_{k=1}^n I_{k\Delta}\left(b_i\right)^2.$$

Hence, we have

$$\|\widehat{b}_{m,i} - b_{i,K}\|_n^2 \mathbb{1}_{\Omega_n} \leq 7\|b_{m,i} - b_{i,K}\|_n^2 + 32 \sup_{t\in\bar{S}_m, \|t\|_\pi=1} |\nu_{n,i}(t)|^2 + \frac{32}{n}\sum_{k=1}^n \left(I_{k\Delta}\left(b_i\right)\right)^2.$$

Thanks to Assumption 2, the last term is easily bounded:

$$\begin{aligned} \mathbb{E}\left[I_{k\Delta}^2\left(b_i\right)\right] &\leq \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta} \mathbb{E}[(b_i(\mathbf{X}_s) - b_i(\mathbf{X}_{k\Delta}))^2]ds \\ &\leq c\Delta. \end{aligned}$$

It remains to bound

$$\mathbb{E}\left(\sup_{t\in\bar{S}_m, \|t\|_\pi=1} \nu_{n,i}^2(t)\right).$$

Vector subspace $\bar{S}_m$ has an orthonormal basis with respect to L_π^2. We denote it by $\{\varphi_\lambda, \lambda \in \Lambda_m\}$, with $\mathrm{card}(\Lambda_m) = D_m$. Every function $t \in \bar{S}_m$ can be written as $t = \sum_{\lambda\in\Lambda_n} a_\lambda \varphi_\lambda$, and its norm is obtained by the formula $\|t\|_\pi^2 = \sum_{\lambda\in\Lambda_m} a_\lambda^2$. So,

$$\mathbb{E}\left(\sup_{t\in\bar{S}_m, \|t\|_\pi=1} \nu_{n,i}^2(t)\right) \leq \sum_{\lambda\in\Lambda_m} \mathbb{E}\left[\nu_{n,i}^2\left(\varphi_\lambda\right)\right].$$

As the process is stationary,

$$\mathbb{E}\left[\varphi_\lambda^2\left(\mathbf{X}_{k\Delta}\right)\right] = \int \varphi_\lambda^2(\mathbf{x})\pi(\mathbf{x})d\mathbf{x} = 1,$$

and we obtain

$$\begin{aligned} \mathbb{E}\left[\nu_{n,i}^2\left(\varphi_\lambda\right)\right] &= \frac{1}{n^2\Delta^2}\sum_{k=1}^n \mathbb{E}\left[\varphi_\lambda^2\left(\mathbf{X}_{k\Delta}\right)\int_{k\Delta}^{(k+1)\Delta} a_{ii}(\mathbf{X}_s)ds\right] \\ &\leq \frac{\sigma_0^2}{n\Delta}, \end{aligned}$$

where σ_0^2 is defined in Assumption 3.

□

Proposition 7.

$$\mathbb{E}(\|\widehat{b}_{m,i} - b_{i,K}\|_n^2 \mathbb{1}_{\Omega_n^c}) \leq \frac{c}{n\Delta}.$$

Proof. It is demonstrated in Comte *et al.* (2007), Lemma 1 p.533 (see Appendix), that

$$\mathbb{P}(\Omega_n^c) \leq \frac{c}{n^2}. \tag{11}$$

Let us set $\varepsilon_{k\Delta} = I_{k\Delta}(b_i) + Z^i_{k\Delta}$, $\varepsilon = (\varepsilon_\Delta, \ldots, \varepsilon_{n\Delta})^*$ and $\Pi_m Y = \Pi_m(Y^i_\Delta, \ldots, Y^i_{n\Delta})^* = \left(\widehat{b}_{m,i}(\mathbf{X}_\Delta), \ldots, \widehat{b}_{m,i}(\mathbf{X}_{n\Delta})\right)^*$. We obtain

$$\begin{aligned}
\|b_{i,K} - \widehat{b}_{m,i}\|_n^2 &= \|b_{i,K} - \Pi_m b_i\|_n^2 + \|\Pi_m b_i - \Pi_m Y^i\|_n^2 \\
&= \|b_{i,K} - \Pi_m b_i\|_n^2 + \|\Pi_m \varepsilon\|_n^2 \\
&\leq \|b_{i,K}\|_n^2 + \|\varepsilon\|_n^2.
\end{aligned}$$

Using the Cauchy-Schwartz inequality, strict stationarity and (11), we have

$$\mathbb{E}(\|b_{i,K}^2\|_n \mathbb{1}_{\Omega_n^c}) \leq \left(\mathbb{E}(b_{i,K}^4(\mathbf{X_0}))\mathbb{P}(\Omega_n^c)\right)^{1/2} \leq \frac{c}{n}.$$

Besides,

$$\mathbb{E}\left(\|\varepsilon\|_n^2 \mathbb{1}_{\Omega_n^c}\right) \leq \left(\mathbb{E}\left[\varepsilon_\Delta^4\right] \mathbb{P}(\Omega_n^c)\right)^{1/2}.$$

Let us compute $\mathbb{E}\left[\varepsilon_\Delta^4\right]$. According to the Burkholder inequality, we know that:

$$\begin{aligned}
\mathbb{E}(\varepsilon_\Delta^4) &\leq C \quad \mathbb{E}\left[(I_{k\Delta}(b_i))^4\right] + \frac{C}{\Delta^4}\mathbb{E}\left[\left(\int_0^\Delta a_{ii}(\mathbf{X}_s)ds\right)^2\right] \\
&\leq \quad C\mathbb{E}\left[\sup_{0\leq s\leq\Delta}\left\{(b_i(\mathbf{X}_s) - b_i(\mathbf{X}_{k\Delta}))^4\right\}\right] ds + \frac{C}{\Delta^2}\mathbb{E}\left[a_{ii}^2(\mathbf{X}_0)\right].
\end{aligned}$$

By stationarity,

$$\mathbb{E}(\varepsilon_\Delta^4) \leq Cc\Delta^2 + \frac{C}{\Delta^2}\mathbb{E}\left[a_{ii}^2(\mathbf{X}_0)\right] \leq \frac{C'}{\Delta^2}.$$

Collecting terms, we have

$$\mathbb{E}(\|\widehat{b}_{m,i} - b_{i,K}\|_n^2 \mathbb{1}_{\Omega_n^c}) \leq \frac{c}{n\Delta}. \qquad \square$$

As the process is stationary, for any function t with support in K,

$$\mathbb{E}\left(\|t\|_n^2\right) = \|t\|_\pi^2 \leq \pi_1 \|t\|_{L^2}^2.$$

Propositions 6 and 7 allow us to conclude the proof of Theorem 1.

6.4 Proof of Theorem 2

We have

$$\|\tilde{b}_i - b_{i,K}\|_n^2 = \|\tilde{b}_i - b_{i,K}\|_n^2 \mathbb{1}_{\Omega_n} + \|\tilde{b}_i - b_{i,K}\|_n^2 \mathbb{1}_{\Omega_n^c}.$$

We obtain, thanks to a proof similar to the previous one, that

$$\mathbb{E}\left(\|\tilde{b}_i - b_{i,K}\|_n^2 \mathbb{1}_{\Omega_n^c}\right) \leq \frac{c}{n\Delta}.$$

By definition of the estimator $\tilde{b}_i$, for any integer m, we have

$$\gamma_{n,i}\left(\tilde{b}_i\right) - \gamma_{n,i}\left(b_i\right) + pen(\hat{m}_i) \leq \gamma_{n,i}\left(b_{m,i}\right) - \gamma_{n,i}\left(b_i\right) + pen(m).$$

As previously, we obtain, for any $m \in \mathscr{M}_n$,

$$\begin{aligned} \|\tilde{b}_i - b_{i,K}\|_n^2 &\leq \|b_{m,i} - b_{i,K}\|_n^2 + 2\nu_{n,i}\left(\tilde{b}_i - b_{m,i}\right) \\ &+ \frac{2}{n}\sum_{k=1}^{n}\left(\tilde{b}_i - b_{m,i}\right)\left(\mathbf{X}_{k\Delta}\right) I_{k\Delta}\left(b_i\right) + pen(m) - pen(\hat{m}_i). \end{aligned}$$

We easily obtain, for any $m \in \mathscr{M}_n$, that

$$\begin{aligned} \|\tilde{b}_i - b_{i,K}\|_n^2 \mathbb{1}_{\Omega_n} &\leq 7\|b_{m,i} - b_{i,K}\|_n^2 + 4\left(pen(m) - pen(\widehat{m}_i)\right)\mathbb{1}_{\Omega_n} \\ &+ 32 \sup_{\|t\|_\pi = 1, t \in \bar{S}_m + \bar{S}_{\hat{m}_i}} |\nu_{n,i}(t)|^2 \mathbb{1}_{\Omega_n} + \frac{32}{n}\sum_{k=1}^{n} I_{k\Delta}\left(b_i\right)^2. \end{aligned}$$

We know that

$$\begin{aligned} \mathbb{E}\left(I_{k\Delta}^2\left(b_{i,K}\right)\right) &\leq c\Delta \\ \mathbb{E}\left(\|b_{m,i} - b_{i,K}\|_n^2\right) &\leq \pi_1 \|b_{m,i} - b_{i,K}\|_{L^2}^2. \end{aligned}$$

Let us set

$$G_m(m') = \sup_{t \in \bar{S}_m + \bar{S}_{m'}, \|t\|_\pi = 1} |\nu_{n,i}(t)|$$

and introduce a penalty function $p(m, m')$ such that

$$p(m, m') = \frac{\kappa_1 \sigma_0^2 (D_m + D_{m'})}{n\Delta}.$$

The following proposition is based on a result of Baraud *et al* (2001b) and an inequality of Bernstein type. We give a sketch of proof in the Appendix.

Proposition 8.

There exists a positive numerical constant κ_1 such that

$$\mathbb{E}\left[\left(G_m^2(m') - p(m,m')\right)\mathbb{1}_{\Omega_n}\right]_+ \leq c\sigma_0^2 \frac{e^{-D_{m'}}}{n\Delta}.$$

Applying this proposition, we choose the penalty function

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

with $\kappa = 8\kappa_1$. We have

$$\begin{aligned} E &:= \mathbb{E}\left[\left(8 \sup_{t\in\bar{S}_m+\bar{S}_{\hat{m}_i}, \|t\|_\pi=1} \nu_{n,i}^2(t) + (pen(m) - pen(\hat{m}_i))\right)\mathbb{1}_{\Omega_n}\right] \\ &\leq \mathbb{E}\left[8\left(G_m^2(\hat{m}_i) - p(m,\hat{m}_i)\right)\mathbb{1}_{\Omega_n} + (pen(m) - pen(\hat{m}_i) + 8p(m,\hat{m}_i))\mathbb{1}_{\Omega_n}\right]. \end{aligned}$$

As

$$\left[\left(G_m^2(\widehat{m}_i) - p(m,\widehat{m}_i)\right)\mathbb{1}_{\Omega_n}\right]_+ \leq \sum_{m'\in\mathcal{M}_n}\left[\left(G_m^2(m') - p(m,m')\mathbb{1}_{\Omega_n}\right],$$

we can write

$$E \leq 8\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left[\left(G_m^2(m') - p(m,m')\mathbb{1}_{\Omega_n}\right] + 2pen(m).$$

Applying Proposition 8 and using the fact that $\sum_m e^{-D_m} < +\infty$, we obtain

$$E \leq 8\sum_{m'\in\mathcal{M}_n} c\sigma_0^2\frac{e^{-D_{m'}}}{n\Delta} + 2pen(m) \leq \frac{c'\sigma_0^2}{n\Delta} + 2pen(m).$$

Then, collecting terms, we have

$$\mathbb{E}\left(\|\tilde{b}_i - b_{i,K}\|_n^2\right) \leq \inf_{m\in\mathscr{M}_n}\left(7\pi_1\|b_{m,i} - b_{i,K}\|_{L^2}^2 + 8pen(m)\right) + \frac{C}{n\Delta} + C'\Delta$$

and the proof of Theorem 2 is complete.

Acknowledgment: the author wishes to thank F. Comte and V.Genon-Catalot for helpful discussions.

References

[1] Baraud, Y., Comte, F., Viennet, G. (2001a) Adaptive estimation in autoregression or beta-mixing regression via model selection. *Ann. Statist.* **29**, no. 3, 839-875.

[2] Baraud, Y., Comte, F., Viennet, G. (2001b) Model selection for (auto-)regression with dependent data. *ESAIM Probab. Statist.* **5**, 33-49

[3] Comte, F., Genon-Catalot, V., Rozenholc, Y. (2007) Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* **13**, no. 2, 514-543.

[4] Comte, F., Rozenholc, Y. (2002) Adaptive estimation of mean and volatility functions in (auto-)regressive models. *Stochastic Process. Appl.* **97**, no. 1, 111-145.

[5] Comte, F., Rozenholc, Y. (2004) A new algorithm for fixed design regression and denoising. *Ann. Inst. Statist. Math.* **56** 449-473.

[6] Fearnhead, P., Papaspiliopoulos, G., O.Roberts, G., Stuart, A. (2007) Filtering systems of coupled stochastic differential equations partially observed at high frequency. available on http://www2.warwick.ac.uk/fac/sci/statistics/crism/research/working_papers/2007/paper07-11/07-11w.pdf

[7] Gloter, A. (2000) Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Statist.* **4**, 205-227

[8] Hoffmann, M. (1999) Adaptive estimation in diffusion processes. *Stochastic Process. Appl.* **79**, no. 1, 135-163

[9] Jacobsen, M (2001) Examples of multivariate diffusions: Time-reversibility, a Cox-Ingersoll-Ross type process. *Preprint 2001-6, Department of Statistics and Operations Research, University of Copenhagen.*

[10] Jacobsen, M., Sorensen, M. (2004) Multivariate diffusions with linear drift and given marginal distribution. *Private communication.*

[11] Karatzas, I., Shreve, S. (1988) E. *Brownian motion and stochastic calculus.* Graduate Texts in Mathematics, 113. Springer-Verlag, New York.

[12] Kent, J. (1978) Time-reversible diffusions. *Adv. in Appl. Probab.* **10**, no. 4, 819-835.

[13] Lacour, C. (2007) Adaptive estimation of the transition density of a Markov chain. *Ann. Inst. H. Poincaré Probab. Statist.* **43**, no. 5, 571-597

[14] Meyer, Y. (1990) *Ondelettes et opérateurs. I. Ondelettes.* Actualités Mathématiques. Hermann, Paris.

[15] Pardoux, E., Veretennikov,A.Yu. (2001) On the Poisson equation and diffusion approximation. *Ann. Probab.* **29**, no. 3, 1061-1085.

[16] Veretennikov, A.Yu.(1987) Bounds for the mixing rates in the theory of stochastic differential equations. *Theory Probab. Appl.* **32** 273-281.

[17] Wu, L. (2001) Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic Process. Appl.* **91**, no. 2, 205-238.

Laboratoire MAP5 (CNRS UMR 8145), UFR de Mathématiques et Informatique, Université Paris Descartes, 45 rue des Saints Pères, 75270 Paris Cedex 06, France.

E-mail : emeline.schmisser@mi.parisdescartes.fr

A Additional proofs

A.1 Sketch of proof of Proposition 8.

We follow the steps of Comte *et al.* and adapt these to dimension d. First we prove the following lemma which is a Bernstein-type inequality:

Lemma 5.
Under the assumptions of Theorem 1, for any function t with support in K,
$\forall \varepsilon > 0, \forall \zeta > 0,$

$$\mathbb{P}\left(\sum_{k=1}^{n} \left|t(\mathbf{X}_{k\Delta})Z^i_{k\Delta}\right| \geq n\varepsilon, \|t\|_n^2 \leq \zeta^2\right) \leq 2\exp\left(-\frac{n\Delta\varepsilon^2}{2\sigma_0^2\zeta^2}\right).$$

Hence, for all $x > 0$,

$$\mathbb{P}_n\left(|\nu_{n,i}(t)| \geq \zeta\sqrt{\frac{2\sigma_0^2 x}{\Delta}}, \|t\|_n^2 \leq \zeta^2\right) \leq 2\exp(-nx).$$

where $\mathbb{P}_n(.) := \mathbb{P}(. \cap \Omega_n)$.

Proof. Let us consider a martingale (M_s) such that $\left(\exp\left(\lambda M_s - \frac{\lambda^2}{2}\langle M\rangle_s\right)\right)$ is also a martingale. We have that

$$\mathbb{E}\left(\exp\left(\lambda M_s - \lambda^2\langle M\rangle_s/2\right)\right) = 1.$$

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

$$\begin{aligned}
\mathbb{P}\left[(M_s \geq c), (\langle M\rangle_s \leq c')\right] &= \mathbb{P}\left[\exp\left(\lambda M_s - \lambda^2\langle M\rangle_s/2\right) \geq \exp\left(\lambda c - \tfrac{\lambda^2 c'}{2}\right)\right] \\
&\leq \exp\left(-\lambda c + \tfrac{\lambda^2 c'}{2}\right)\mathbb{E}\left(\exp\left(\lambda M_s - \tfrac{\lambda^2}{2}\langle M\rangle_s\right)\right) \\
&\leq \exp\left(-\lambda c + \tfrac{\lambda^2 c'}{2}\right).
\end{aligned}$$

Hence, minimizing with respect to λ, we obtain

$$\mathbb{P}\left(M_s \geq c, \langle M\rangle_s \leq c'\right) \leq \inf_{\lambda>0} \exp\left(-\lambda c + \frac{\lambda^2 c'}{2}\right) = \exp\left(-\frac{c^2}{2c'}\right).$$

Let us consider the process

$$(\mathbf{H}_u)_j = \left(\sum_{k=1}^{n} \mathbb{1}_{]k\Delta,(k+1)\Delta]}(u) t(\mathbf{X}_{k\Delta}) \sigma_{ij}(\mathbf{X}_u)\right)_j$$

which satisfies, for all positive real u, the inequality $\mathbf{H}_u.\mathbf{H}_u \leq \sigma_0^2 \|t\|_\infty$ (in order to avoid confusion with the martingale bracket, the euclidean scalar product is denoted by "." in this proof). Let us set $M_s = \int_0^s \mathbf{H}_u.d\mathbf{W}_u$. This process satisfies:

$$\begin{aligned} M_{(n+1)\Delta} &= \sum_{j=1}^{d}\sum_{k=1}^{n} t(\mathbf{X}_{k\Delta}) \int_{k\Delta}^{(k+1)\Delta} \sigma_{ij}(\mathbf{X}_s) dW_s^j = \Delta \sum_{k=1}^{n} t(\mathbf{X}_{k\Delta}) Z_{k\Delta}^i. \\ \langle M\rangle_{(n+1)\Delta} &= \sum_{k=1}^{n} t^2(\mathbf{X}_{k\Delta}) \int_{k\Delta}^{(k+1)\Delta} a_{ii}(\mathbf{X}_s) ds \leq \sigma_0^2 n\Delta \|t\|_n^2. \end{aligned}$$

Moreover

$$\langle M\rangle_s = \int_0^s \mathbf{H}_u.\mathbf{H}_u du \leq n\sigma_0^2 \Delta \|t\|_\infty.$$

Then M_s and $\exp\left(\lambda M_s - \frac{\lambda^2}{2}\langle M\rangle_s\right)$ are martingales. We obtain

$$\begin{aligned} F &:= \mathbb{P}\left[\left(\textstyle\sum_{k=1}^{n} t(\mathbf{X}_{k\Delta}) Z_{k\Delta}^i \geq n\varepsilon\right), \left(\|t\|_n^2 \leq \zeta^2\right)\right] \\ &\leq \mathbb{P}\left[\left(M_{(n+1)\Delta} \geq \Delta n\varepsilon\right), \left(\langle M\rangle_{(n+1)\Delta} \leq \sigma_0^2 n\Delta\zeta^2\right)\right] \\ &\leq \exp\left(-\frac{n\varepsilon^2\Delta}{2\sigma_0^2\zeta^2}\right). \quad \square \end{aligned}$$

To complete the proof of Proposition 8, we use that

$$\mathbb{E}\left[\left(G_m^2(m') - p(m,m')\right)\mathbb{1}_{\Omega_n}\right]_+ = \int_0^\infty \mathbb{P}\left[\left(G_m^2(m') - p(m,m')\right)_+ \mathbb{1}_{\Omega_n} \geq x\right] dx.$$

By substituting $x = \kappa_1\sigma_0^2 \frac{1}{n\Delta}\tau$, and replacing $p(m,m')$ by its expression $p(m,m') = \dfrac{\kappa_1\sigma_0^2(D_m + D_{m'})}{n\Delta}$, we have

$$\mathbb{E}\left[\left(G_m^2(m') - p(m,m')\right)\mathbb{1}_{\Omega_n}\right]_+ \leq \kappa_1 \frac{\sigma_0^2}{n\Delta} \int_0^\infty \mathbb{P}_n\left[G_m^2(m') \geq \kappa_1 \frac{\sigma_0^2}{n\Delta}(\tau + D)\right] d\tau.$$

Lemma 5 and the L^2 chaining technique of Baraud *et al.* (2001b) allow to obtain the announced result.

A.2 Proof of inequality (11)

The following lemma is proved in Baraud *et al.* (2001a).

Lemma 6.
Let us set $n = p_n q_n$ and consider $\bar{\mathscr{S}}_n$ the largest vector space generated by the families of functions F_m, for $m \in \mathscr{M}_n$ (see Section 3). The dimension of $\bar{\mathscr{S}}_n$ is equal to N_n. For all positive Δ, we have

$$\mathbb{P}(\Omega_n^c) \leq 2n\beta_X(q_n\Delta) + 2n^2 \exp\left(-A_0\pi_0^2 \frac{n}{q_n L_n(\phi)}\right),$$

where $L_n(\phi)$ satisfies

$$L_n(\phi) \leq \phi_0^2 N_n^2.$$

Comte *et al.* (2007) use this lemma to end the proof of the inequality (11).