



HAL
open science

Nonparametric estimation in hidden Markov models

Thierry Dumont, Sylvain Le Corff

► **To cite this version:**

Thierry Dumont, Sylvain Le Corff. Nonparametric estimation in hidden Markov models. 2012. hal-00727526v2

HAL Id: hal-00727526

<https://hal.science/hal-00727526v2>

Preprint submitted on 10 Sep 2012 (v2), last revised 10 Aug 2015 (v5)

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

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

Nonparametric estimation in hidden Markov models

Thierry Dumont ^{*†} and Sylvain Le Corff ^{‡§}

September 10, 2012

Abstract

This paper outlines a new procedure to perform nonparametric estimation in hidden Markov models. It is assumed that a Markov chain $\{X_k\}_{k \geq 0}$ is observed only through a process $\{Y_k\}_{k \geq 0}$, where Y_k is a noisy observation of $f_*(X_k)$. We propose a maximum likelihood based procedure to estimate the function f_* using a block of observations $Y_{0:2n-1}$. This paper shows the identifiability of the model under several assumptions on the Markov chain and on the function f_* . We also provide a proof of the consistency of the estimator of f_* as the number of observations grows to infinity. This consistency result relies on the Hellinger consistency of an estimator of the likelihood of the observations. Finally, we provide numerical experiments to highlight the performance of the estimator.

1 Introduction

A bivariate stochastic process $\{(X_k, Y_k)\}_{k \geq 0}$ is said to be a hidden Markov model (HMM) if the state sequence $\{X_k\}_{k \geq 0}$ is a Markov chain, if the observations $\{Y_k\}_{k \geq 0}$ are independent conditionally on $\{X_k\}_{k \geq 0}$ and if the conditional distribution of Y_k given the state sequence depends only on X_k . These models can be applied in a large variety of disciplines such as financial econometrics ([19]), biology ([6]) or speech recognition ([15]).

In this paper, the state-space of the Markov chain $\{X_k\}_{k \geq 0}$ is assumed to be a compact subset of \mathbb{R}^m homeomorphic to a convex subset of \mathbb{R}^m with a Lipschitz boundary. This Markov chain is a random walk with increment distribution known up to a scaling factor a_* . The observations are given, for any $k \geq 0$, by $Y_k = f_*(X_k) + \epsilon_k$, where f_* is a function on K taking values in \mathbb{R}^ℓ and the measurement noise $\{\epsilon_k\}_{k \geq 0}$ is an i.i.d. sequence of i.i.d. Gaussian

*Université Paris-Sud, 91405 Orsay Cedex, France

†ID Services, 22/24 rue Jean Rostand, 91400 Orsay, France.

‡LTCI, TELECOM ParisTech and CNRS, 46 rue Barrault 75634 Paris Cedex 13, France

§This work is partially supported by the French National Research Agency, under the programs ANR-08-BLAN-0218 BigMC and ANR-07-ROBO-0002

on \mathbb{R}^ℓ with known covariance matrix. The aim of this paper is to estimate the function f_\star and the parameter a_\star using only the observations $\{Y_k\}_{k \geq 0}$.

In regression models such as *errors-in-variables* models, the variables $\{X_k\}_{k \geq 0}$ are observed through a sequence $\{Z_k\}_{k \geq 0}$ given by $Z_k \stackrel{\text{def}}{=} X_k + \eta_k$, where the random variables $\{\eta_k\}_{k \geq 0}$ are i.i.d. with known distribution. Many solutions have been proposed to solve this regression problem using an estimation of the probability density of X_0 (this is the *deconvolution* problem), see [4], [5] and [14] for an estimation based on kernel density estimators; see also [7] for an estimation based on the minimization of a penalized contrast. Nevertheless, all these works rely on the assumption that the process $\{X_k\}_{k \geq 0}$ is directly observed, which is not the case in our model.

When $\{X_k\}_{k \geq 0}$ is a Markov chain, [17] proposed an estimation of the density of the invariant probability and of the Markov kernel of $\{X_k\}_{k \geq 0}$ when the chain is observed. The estimation procedure amounts to minimizing a penalized contrast in order to minimize the empirical L_2 -norm of the error. [16] provided an extension of this work in the HMM framework when the observations are given by

$$Y_k = X_k + \epsilon_k ,$$

where the random variables $\{\epsilon_k\}_{k \geq 0}$ are i.i.d. with known distribution. These works provide estimation procedures of the Markov chain $\{X_k\}_{k \geq 0}$ but there does not exist any result on the nonparametric estimation problem studied in this paper.

This problem is motivated by an application to localization using radio measurements (see [11]). In this case, at each time step k , a mobile device observes the power of signals transmitted by ℓ antennas; this measurement is denoted by Y_k . The localization of the device is denoted by X_k and is assumed to be a Markov chain on a subset of \mathbb{R}^2 . The problem consists in estimating the localizations $\{X_k\}_{k \geq 0}$ only observing the signal powers $\{Y_k\}_{k \geq 0}$. In this application, f_\star represents the average propagation model, which means that the variable Y_k follows the normal distribution on \mathbb{R}^ℓ , $\mathcal{N}(f_\star(X_k), \sigma^2 I_\ell)$. An accurate estimation of the positions $\{X_k\}_{k \geq 0}$, using particle filtering for instance, relies on a good estimation of f_\star .

The main result of this paper is the identifiability of the model. We assume that the Markov chain $\{X_k\}_{k \geq 0}$ is stationary with known (up to a scaling factor a_\star) transition kernel, and that f_\star is a diffeomorphism on its image (which necessarily implies that $m \leq \ell$). We assume in addition that f_\star is smooth in the sense that it belongs to some Sobolev space $W^{s,p}$ (see (5)). Provided that f is continuously differentiable and is such that $(f(X_0), f(X_1))$ and $(f_\star(X_0), f_\star(X_1))$ have the same distribution we show that there exists an isometric transformation ϕ on the state-space K such that $f = f_\star \circ \phi$. A key step is to show that $(f_\star)^{-1} \circ f$ is necessarily bijective, which is done using algebraic topology and measure theoretic arguments.

Our estimator \widehat{f}_n is defined as a maximizer of a penalized pairwise likelihood on the Sobolev space $W^{s,p}$. The parameters s and p of the Sobolev space are assumed to satisfy $s > m/p + 1$ and K is assumed to be compact to allow the

use of classical Sobolev embeddings into the space of continuously differentiable functions on K . This estimator of f_\star is associated to an estimator \widehat{p}_n of the marginal distribution of a pair of observations (see (12)). We prove that the Hellinger distance between \widehat{p}_n and the true distribution of a pair of observations under (f_\star, a_\star) vanishes as the number of observations grows to infinity. More precisely, we prove that the rate of convergence of \widehat{p}_n , in Hellinger distance, can be chosen as close as possible to $n^{-1/2}$. The consistency of $(\widehat{f}_n, \widehat{a}_n)$ follows as a consequence together with the identifiability result and continuity properties. To analyze the asymptotic properties of our estimators, we need, as it is now well understood, deviation inequalities for the empirical process of the observations. To that purpose, we use the concentration inequality for additive functionals of Markov chains proved in [1] and the maximal inequality for dependent processes of [10] to have a control on the supremum of a function-indexed empirical process.

Our results are supported by numerical experiments: in the case where the scaling parameter a_\star is known and $m = 1$, we provide an Expectation-Maximization based algorithm to compute \widehat{f}_n , see [8]. We show that the estimation procedure can be solved using a differential equation. We provide several simulations that show the efficiency of our method.

In Section 2 the model, the estimators and the assumptions are presented. The main results are displayed in Section 3: the identifiability of the model in Section 3.1 and the consistency of the estimator along with a rate of convergence in Section 3.2. The algorithm and numerical experiments are displayed in Section 4. Section 5 gathers important proofs on the identifiability and consistency needed to state the main results. Additional technical results are provided in the appendices and in the supplement paper [12].

2 Model and definitions

Let ℓ and m be positive integers and K be a subset of \mathbb{R}^m . The main statistical problem considered in this paper is the estimation of an unknown target function $f_\star : K \rightarrow \mathbb{R}^\ell$ when observing a process $\{Y_k\}_{k \in \mathbb{N}}$ such that for any $k \geq 0$, Y_k belongs to \mathbb{R}^ℓ and satisfies

$$Y_k \stackrel{\text{def}}{=} f_\star(X_k) + \epsilon_k .$$

$\{\epsilon_k\}_{k \in \mathbb{N}}$ is assumed to be an i.i.d Gaussian process with common known distribution $\mathcal{N}(0, \sigma^2 I_\ell)$, I_ℓ being the identity matrix of size ℓ and σ^2 a fixed positive parameter. Denote by φ the probability distribution of ϵ_0 , *i.e.*

$$\forall z \in \mathbb{R}^\ell, \varphi(z) \stackrel{\text{def}}{=} (2\pi\sigma^2)^{-\ell/2} \exp \left\{ -\frac{\|z\|^2}{2\sigma^2} \right\} ,$$

where $\|\cdot\|$ is the euclidean norm on \mathbb{R}^m (we use the same notation for the euclidian norm on \mathbb{R}^ℓ). $\{X_k\}_{k \in \mathbb{N}}$ is assumed to be a non observed Markov chain, taking its values in K and independent of $\{\epsilon_k\}_{k \in \mathbb{N}}$. In the sequel, all the

density functions are with respect to the Lebesgue measure on K , denoted by μ . For any $a \in \mathbb{R}_+^*$, denote by q_a the transition density on K defined, for all $x, x' \in K$, by

$$q_a(x, x') \stackrel{\text{def}}{=} C_a(x)q\left(\frac{\|x' - x\|}{a}\right), \quad (1)$$

where q is a known, positive, continuous and strictly monotone function on \mathbb{R}_+ and where

$$C_a(x) \stackrel{\text{def}}{=} \left(\int_K q\left(\frac{\|x' - x\|}{a}\right) dx' \right)^{-1}, \quad (2)$$

where dx' is a shorthand notation for $\mu(dx')$. In our numerical application in Section 4.2, the Gaussian kernel $q(x) = \exp(-x^2/2)$ is chosen. The Markov transition kernel associated with q_a is denoted by Q_a . Assume the existence of an unknown parameter $a_* > 0$ such that

H1 $\{X_k\}_{k \in \mathbb{Z}}$ is a stationary Markov chain with transition kernel Q_{a_*} .

It follows from H1 that $\{Y_k\}_{k \in \mathbb{N}}$ is stationary. Assume the following statement on the set K :

- H2** (i) K is a compact subset of \mathbb{R}^m .
(ii) K is homeomorphic to a convex subset of \mathbb{R}^m .
(iii) K has a local Lipschitz boundary.

K has a local Lipschitz boundary if, for any x in the boundary ∂K of K , there exists a neighbourhood V of x in ∂K which is the graph of a Lipschitz function. As an immediate consequence of the compactness of K and of the positivity of q , there exists $0 < \sigma_-(a) < \sigma_+(a) < +\infty$ such that, for all $x, x' \in K$,

$$\sigma_-(a) \leq q_a(x, x') \leq \sigma_+(a). \quad (3)$$

For any $a > 0$, Q_a is a ψ -irreducible and recurrent Markov kernel and then, it has a unique invariant probability distribution (see [21, Theorem 10.0.1]). By the symmetry of the kernel $(x, x') \rightarrow q\left(\frac{\|x-x'\|}{a}\right)$, the finite measure on K with density function $x \mapsto C_a^{-1}(x)$ is Q_a -invariant. Therefore, the unique invariant probability of Q_a has a density given by

$$\forall x \in K, \nu_a(x) \stackrel{\text{def}}{=} \frac{\int_K q\left(\frac{\|x'-x\|}{a}\right) dx'}{\int_{K^2} q\left(\frac{\|x'-x''\|}{a}\right) dx' dx''}. \quad (4)$$

Let $p \geq 1$, define

$$\mathbb{L}^p \stackrel{\text{def}}{=} \left\{ f : K \rightarrow \mathbb{R}^\ell ; \|f\|_{\mathbb{L}^p}^p = \int_K \|f(x)\|^p dx < \infty \right\}.$$

For any m -tuple $\alpha \stackrel{\text{def}}{=} \{\alpha_i\}_{i=1}^m$ of non-negative integers, we write $|\alpha| \stackrel{\text{def}}{=} \sum_{i=1}^m \alpha_i$. For any $f : K \rightarrow \mathbb{R}^\ell$ and any $j \in \{1, \dots, \ell\}$, the j^{th} component of f is denoted

by f_j . Let $s \in \mathbb{N}$, define $W^{s,p}$ be the Sobolev space on K with parameters s and p , i.e.,

$$W^{s,p} \stackrel{\text{def}}{=} \{f \in L^p; D^\alpha f \in L^p, \alpha \in \mathbb{N}^m \text{ and } |\alpha| \leq s\}, \quad (5)$$

where $D^\alpha f : K \rightarrow \mathbb{R}^\ell$ represents here the vector of partial derivatives of order α , in the sense of distributions, of the components f_j , for $j \in \{1, \dots, \ell\}$. $W^{s,p}$ is equipped with the norm $\|\cdot\|_{W^{s,p}}$ defined, for any $f \in W^{s,p}$, by

$$\|f\|_{W^{s,p}} \stackrel{\text{def}}{=} \left(\sum_{0 \leq |\alpha| \leq s} \|D^\alpha f\|_{L^p}^p \right)^{1/p}. \quad (6)$$

For any subset Ω_0 of \mathbb{R}^m , and any $k \geq 0$, let $\mathcal{C}^k(\Omega_0)$ be the vector space of all the functions $f : \Omega_0 \rightarrow \mathbb{R}$ such that there exists an open neighbourhood Ω of Ω_0 (if Ω_0 is open we can take $\Omega = \Omega_0$) in \mathbb{R}^m and a function $\bar{f} : \Omega \rightarrow \mathbb{R}$ such that the restriction $\bar{f}|_{\Omega_0}$ of \bar{f} on Ω_0 satisfies $\bar{f}|_{\Omega_0} = f$ and \bar{f} is \mathcal{C}^k -regular on Ω , which means that \bar{f} and all its partial derivatives $D^\alpha \bar{f}$ are continuous on Ω . Define, for any x in Ω_0 , $D^\alpha f(x) = D^\alpha \bar{f}(x)$. Let $\|\cdot\|_{\mathcal{C}^k(\Omega_0)}$ be the norm on $\mathcal{C}^k(\Omega_0)$ defined by $\|f\|_{\mathcal{C}^k(\Omega_0)} \stackrel{\text{def}}{=} \sup_{|\alpha| \leq k} \|D^\alpha f\|_\infty$. We also define $\mathcal{C}^k(\Omega_0, \mathbb{R}^\ell)$ by $\mathcal{C}^k(\Omega_0, \mathbb{R}^\ell) = \mathcal{C}^k(\Omega_0)^\ell$.

Remark 2.1. i) By H2(iii) and the Stein Theorem [2, Theorem 5.24], there exists a positive constant C such that any bounded function f in $\mathcal{C}^1(\overset{\circ}{K})$ can be extended by a function \bar{f} in $\mathcal{C}^1(\mathbb{R}^m)$, with $\|\bar{f}\|_{\mathcal{C}^1(\mathbb{R}^m)} \leq C\|f\|_{\mathcal{C}^1(\overset{\circ}{K})}$.

ii) Note that, for any $j \in \{1, \dots, \ell\}$ and $f \in W^{s,p}$, f_j belongs to $W^{s,p}(K, \mathbb{R})$, the Sobolev space of real-valued functions with parameters s and p . Let $k \geq 0$, by [2, Theorem 6.3], assuming that K satisfies H2(i) and H2(iii) and $s > m/p + k$, $W^{s,p}(K, \mathbb{R})$ is compactly embedded into the subspace of bounded functions in $\left(\mathcal{C}^k(\overset{\circ}{K}), \|\cdot\|_{\mathcal{C}^k(\overset{\circ}{K})}\right)$. Provided that $s > m/p + 1$, and arguing component by component, $W^{s,p}$ is compactly embedded into the subspace of bounded functions $\mathcal{C}^1(\overset{\circ}{K}, \mathbb{R}^\ell)$. Moreover, the identity function $id : W^{s,p} \rightarrow \mathcal{C}^1(\overset{\circ}{K}, \mathbb{R}^\ell)$ being linear and continuous, there exists a positive coefficient κ such that, for any $f \in W^{s,p}$,

$$\|f\|_{\mathcal{C}^1(\overset{\circ}{K}, \mathbb{R}^\ell)} \leq \kappa \|f\|_{W^{s,p}}, \quad (7)$$

thus f is a bounded function in $\mathcal{C}^1(\overset{\circ}{K}, \mathbb{R}^\ell)$ and, by i), can be extended by a function in $\mathcal{C}^1(K, \mathbb{R}^\ell)$ shortly denoted by \mathcal{C}^1 , and

$$\|f\|_{\mathcal{C}^1} \leq \kappa \|f\|_{W^{s,p}}. \quad (8)$$

H3 $s > m/p + 1$.

For any $f \in \mathcal{C}^1$ and any $x \in K$, the Jacobian of f at x , is defined by

$$J_f^2(x) \stackrel{\text{def}}{=} \text{Det} [D_f(x)^T D_f(x)] ,$$

where $D_f(x)$ is the $\ell \times m$ gradient matrix of f at x defined, for any $j \in \{1, \dots, \ell\}$ and any $i \in \{1, \dots, m\}$, by

$$D_f(x)_{j,i} \stackrel{\text{def}}{=} \frac{\partial f_j}{\partial x_i}(x) .$$

For any sets E and F , $f : E \mapsto F$, denote by $\text{Im}(f)$ the image in F of f , $\text{Im}(f) \stackrel{\text{def}}{=} f(E)$.

- H4** (i) $f_\star \in W^{s,p}$.
(ii) $f_\star : K \rightarrow \text{Im}(f_\star)$ is a diffeomorphism.

Remark 2.2. i) We say that a function $f : K \rightarrow \text{Im}(f)$ is a diffeomorphism if there exists an open neighbourhood V of K in \mathbb{R}^m and a diffeomorphism $\bar{f} : V \rightarrow \text{Im}(\bar{f})$ such that $\bar{f}|_V = f$.

- ii) By H4(ii), for any x in K , the linear application $D_{f_\star}(x)$ is injective and thus, $m \leq \ell$.

We now give the definition of the estimators (\hat{f}_n, \hat{a}_n) of (f_\star, a_\star) given $2n$ observations $\{Y_k\}_{k=0}^{2n-1}$. For practical reasons (see proof of Proposition 3.6), we assume that $a_\star \in [a_-, +\infty[$, for a known $a_- > 0$. For all integer $n \geq 1$, define (\hat{f}_n, \hat{a}_n) by

$$(\hat{f}_n, \hat{a}_n) \stackrel{\text{def}}{=} \underset{f \in W^{s,p}, a \geq a_-}{\text{argmax}} \left\{ \frac{1}{n} \sum_{k=0}^{n-1} \ln p_{f,a}(Y_{2k}, Y_{2k+1}) - \lambda_n^2 I^2(f) \right\} , \quad (9)$$

where, for all y_0, y_1 in \mathbb{R}^ℓ ,

$$p_{f,a}(y_0, y_1) \stackrel{\text{def}}{=} \int \varphi(y_0 - f(x_0)) \varphi(y_1 - f(x_1)) \nu_a(x_0) q_a(x_0, x_1) dx_0 dx_1 \quad (10)$$

and, for some positive v ,

$$I^2(f) \stackrel{\text{def}}{=} \|f\|_{W^{s,p}}^{v+1} . \quad (11)$$

Remark 2.3. By the dominated convergence theorem, the function

$$(f, a) \mapsto \frac{1}{n} \sum_{k=0}^{n-1} \ln p_{f,a}(Y_{2k}, Y_{2k+1})$$

is continuous on $\mathcal{C}^1 \times [a_-, \infty[$, thus, by (9), (11) and Remark 2.1, \hat{f}_n exists and belongs to \mathcal{C}^1 .

Consider the following assumption on v .

H5 $v > 2\ell$.

Note that $(f, a) \rightarrow \frac{1}{n} \sum_{k=0}^{n-1} \ln p_{f,a}(Y_{2k}, Y_{2k+1})$ does not represent the likelihood of the observations $\{Y_k\}_{k=0}^{2n-1}$ but what we call the pairwise pseudo-likelihood of the observations.

By (9), \hat{a}_n could be equal to ∞ so that we shall extend our definitions to this case. By the dominated convergence theorem, for any $x_0, x_1 \in K$, any $y_0, y_1 \in \mathbb{R}^\ell$ and any measurable function f , $q_a(x_0, x_1)$, $\nu_a(x_0)$ and $p_{f,a}(y_0, y_1)$ converge as $a \rightarrow \infty$ to $q_\infty(x_0, x_1)$, $\nu_\infty(x_0)$ and $p_{f,\infty}(y_0, y_1)$, defined by:

$$\begin{aligned} \nu_\infty(x_0) &\stackrel{\text{def}}{=} \mu(K)^{-1}, \quad q_\infty(x_0, x_1) \stackrel{\text{def}}{=} \mu(K)^{-1}, \\ p_{f,\infty}(y_0, y_1) &\stackrel{\text{def}}{=} \mu(K)^{-2} \int \varphi(y_0 - f(x_0)) dx_0 \int \varphi(y_1 - f(x_1)) dx_1. \end{aligned}$$

Let \hat{p}_n denote the maximum penalized likelihood estimator (MLE) of the density on $\mathbb{R}^{2\ell}$ of (Y_0, Y_1) , defined by

$$\hat{p}_n \stackrel{\text{def}}{=} p_{\hat{f}_n, \hat{a}_n}. \quad (12)$$

The convergence properties of this estimator will be analyzed with the Hellinger metric, defined, for any probability densities p_1 and p_2 on $\mathbb{R}^{2\ell}$, by

$$h(p_1, p_2) \stackrel{\text{def}}{=} \left[\frac{1}{2} \int \left(p_1^{1/2}(y) - p_2^{1/2}(y) \right)^2 dy \right]^{1/2}. \quad (13)$$

Remark 2.4. The reason we use the Sobolev framework instead of directly considering the space \mathcal{C}^1 is, first of all, computational. Indeed, as we will see in Section 4, the Sobolev norm chosen in penalty (11) can be easily manipulated compared with the \mathcal{C}^1 norm. Moreover, Theorem 3.5 ensures that $\|\hat{f}_n\|_{W^{s,p}}$ stays bounded and thus, by Remark 2.1, that $\{\hat{f}_n\}_{n \geq 1}$ lies in a compact subset of \mathcal{C}^1 . This plays a key role in the proof of Theorem 3.7.

Section 3 provides the main results of the paper. Theorem 3.1 establishes the identifiability of our model. Then, the Hellinger consistency of the MLE (12) is shown in Theorem 3.5. This result does not imply, *a priori*, the consistency of the estimators (\hat{f}_n, \hat{a}_n) defined by (9). However, by Theorem 3.1, whenever the MLE is consistent, so is (\hat{f}_n, \hat{a}_n) up to an isometric transformation on the state space K . The consistency of (\hat{f}_n, \hat{a}_n) is given by Theorem 3.7.

3 Main results

3.1 Identifiability

We denote by \mathcal{I} the set of all the isometries of K . For any functions f and h defined on K we write $f \stackrel{\mathcal{I}}{\sim} h$ and say that f and h are in the same equivalence

class modulo the isometric transformations of K , if and only if there exists an isometry ϕ on K such that $f = h \circ \phi$. In the sequel, for any random variables X and Y , we write $X \stackrel{\mathcal{D}}{=} Y$ if X and Y have the same distribution.

Theorem 3.1. *Assume H1-2 and H4. Let $f : K \rightarrow \mathbb{R}^\ell$ be \mathcal{C}^1 and $0 < b \leq \infty$. Assume also that $h(p_{f,b}, p_{f_\star, a_\star}) = 0$ where $p_{f,b}$ and p_{f_\star, a_\star} are defined by (10). Then, $b = a_\star$ and $f \stackrel{\mathcal{L}}{\sim} f_\star$.*

Proof. The proof of the intermediate lemmas are postponed to Section 5.1. Let $0 < b \leq \infty$ and $f \in \mathcal{C}^1$ such that $h(p_{f,b}, p_{f_\star, a_\star}) = 0$. Let $\{X'_k\}_{k \geq 0}$ be a Markov chain with initial distribution ν_b and transition kernel Q_b . Consider also $\{\epsilon'_k\}_{k \geq 0}$ a sequence of independent $\mathcal{N}(0, \sigma^2 I_\ell)$ random variables, independent from $\{X'_k\}_{k \geq 0}$. Define, for any $k \geq 0$, $Y'_k = f(X'_k) + \epsilon'_k$. If $h(p_{f,b}, p_{f_\star, a_\star}) = 0$, then, for any $k \geq 0$, $(Y_k, Y_{k+1}) \stackrel{\mathcal{D}}{=} (Y'_k, Y'_{k+1})$. The density φ being known, this yields

$$(f(X'_k), f(X'_{k+1})) \stackrel{\mathcal{D}}{=} (f_\star(X_k), f_\star(X_{k+1})). \quad (14)$$

(14) and the irreducibility of the Markov chains $\{X_k\}_{k \geq 0}$ and $\{X'_k\}_{k \geq 0}$ imply that $\text{Im}(f) = \text{Im}(f_\star)$. By H4, f_\star is a diffeomorphism. Let $(f_\star)^{-1}$ denotes its inverse function and define

$$\phi \stackrel{\text{def}}{=} (f_\star)^{-1} \circ f. \quad (15)$$

Since f_\star is a diffeomorphism and $f \in \mathcal{C}^1$, $\phi \in \mathcal{C}^1$. The purposes of the following lemmas is to prove that ϕ is bijective on K and that, for any x in K , $J_\phi(x) > 0$ which is showed in Lemma 3.2.

Lemma 3.2. *Assume H2(i) and H4. For all $x \in K$, $J_\phi(x) > 0$, where ϕ is defined by (15).*

Then, we show that ϕ is necessarily a covering map of K (see definition below) and that, under H2(ii), any covering map of K is a one to one function. These results are established in Lemma 3.3 and Lemma 3.4.

$\phi : K \rightarrow K$ is said to be a covering map if and only if (see [18, Chapter 11])

- (i) ϕ is continuous.
- (ii) ϕ is surjective.
- (iii) For every $y \in K$, there exists an open neighbourhood V of y and a family $(O_i)_{i \in I}$ of disjoint open subsets of K such that $\phi^{-1}(V) = \bigcup_{i \in I} O_i$, with O_i mapped homeomorphically onto V by ϕ , for all $i \in I$.

Lemma 3.3. *Assume H2(i) and H4. Then, the function ϕ defined by (15) is a covering map.*

Lemma 3.4. *Assume H2(ii). Then, every covering map $\phi : K \rightarrow K$ is a one to one function.*

By Lemma 3.3 and Lemma 3.4, ϕ , defined by (15) is bijective, denote by ϕ^{-1} the inverse function of ϕ . By Lemma 3.2, $J_\phi > 0$ on K and thus $\phi^{-1} \in C^1$. By (14), for all $x \in K$ and all positive measurable function h on K ,

$$\begin{aligned} Q_{a_\star}(x, h) &= \mathbb{E}[h(X_k)|X_{k-1} = x] = \mathbb{E}[h \circ \phi(X'_k)|X'_{k-1} = \phi^{-1}(x)] , \\ &= Q_b(\phi^{-1}(x), h \circ \phi) . \end{aligned}$$

Moreover,

$$\begin{aligned} Q_b(\phi^{-1}(x), h \circ \phi) &= \int_K h \circ \phi(u) Q_b(\phi^{-1}(x), u) du \\ &= \int_K h(u) Q_b(\phi^{-1}(x), \phi^{-1}(u)) |J_{\phi^{-1}}(u)| du . \end{aligned}$$

Then, by continuity, for all $(x, x') \in K^2$,

$$Q_{a_\star}(x, x') = Q_b(\phi^{-1}(x), \phi^{-1}(x')) |J_{\phi^{-1}}(x')| .$$

This equation directly leads to $b < \infty$. Indeed, if $b = \infty$, the left side of the equation depends on x (since $a_\star < \infty$) whereas the right side does not. We can now suppose $0 < b < \infty$. By (1),

$$C_{a_\star}(x) q\left(\frac{\|x' - x\|}{a_\star}\right) = C_b(\phi^{-1}(x)) q\left(\frac{\|\phi^{-1}(x') - \phi^{-1}(x)\|}{b}\right) |J_{\phi^{-1}}(x')| . \quad (16)$$

Therefore, for all $x \in K$, applying (16) with $x' = x$ yields

$$|J_{\phi^{-1}}(x)| = \frac{C_{a_\star}(x)}{C_b(\phi^{-1}(x))} . \quad (17)$$

By H2(i-ii), Schauder's theorem (see [24]) states that there exists $x_0 \in K$ such that $\phi^{-1}(x_0) = x_0$. By (16), there exists a constant C such that, for all $x \in K$

$$|J_{\phi^{-1}}(x)| = C \frac{q\left(\frac{\|x - x_0\|}{a_\star}\right)}{q\left(\frac{\|\phi^{-1}(x) - \phi^{-1}(x_0)\|}{b}\right)} .$$

Plugging this expression and (17) in (16) yields

$$\begin{aligned} &q\left(\frac{\|x - x_0\|}{a_\star}\right) q\left(\frac{\|x' - x\|}{a_\star}\right) q\left(\frac{\|\phi^{-1}(x') - \phi^{-1}(x_0)\|}{b}\right) \\ &= q\left(\frac{\|\phi^{-1}(x) - \phi^{-1}(x_0)\|}{b}\right) q\left(\frac{\|\phi^{-1}(x') - \phi^{-1}(x)\|}{b}\right) q\left(\frac{\|x' - x_0\|}{a_\star}\right) . \end{aligned} \quad (18)$$

Applied with $x' = x_0$, we have, for all $x \in K$,

$$q\left(\frac{\|x - x_0\|}{a_\star}\right) = q\left(\frac{\|\phi^{-1}(x) - \phi^{-1}(x_0)\|}{b}\right)$$

and then, since q is a one to one function by assumption,

$$\frac{\|x - x_0\|}{a_\star} = \frac{\|\phi^{-1}(x) - x_0\|}{b}.$$

Considering the supremum of the last inequality for $x \in K$ yields $b = a_\star$. Then, (18) gives, for all $x, x' \in K$

$$\|\phi^{-1}(x') - \phi^{-1}(x)\| = \|x' - x\|.$$

Therefore, ϕ is an isometry and $f = f_\star \circ \phi$ which concludes the proof. \square

3.2 Convergence results

Theorem 3.5 states the Hellinger consistency of the MLE \hat{p}_n and ensures that the Sobolev norm of the estimator \hat{f}_n is bounded. Theorem 3.1 and Theorem 3.5 lead to the second main result, Theorem 3.7, which guarantees that (\hat{f}_n, \hat{a}_n) is also consistent. The proof of Theorem 3.5 uses the same classical proof scheme as in the independent case, see [25, Section 10.2] for an illustration of such a proof. This proof relies on the control of the empirical process, it requires both a result on the concentration of the empirical process and a maximal inequality. Unfortunately, the tools used in the independent case such as the Bernstein or the Hoeffding inequalities do not hold in our model and similar results in the dependent case have to be used, see [1]. Denote by \mathbb{P}_\star the distribution of $\{Y_k\}_{k \geq 0}$ under the true parameters (f_\star, a_\star) . For any sequence of random variables $\{Z_n\}_{n \geq 0}$ and any sequence of positive numbers $\{\alpha_n\}_{n \geq 0}$, we write $Z_n = O_{\mathbb{P}_\star}(\alpha_n)$ if

$$\lim_{T \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}_\star \{|Z_n| > T\alpha_n\} = 0.$$

Theorem 3.5. *Assume H1-3, H4(i) and H5. Let (\hat{f}_n, \hat{a}_n) be defined by (9) and $I(f)$ by (11). Then, provided that*

$$\lambda_n \xrightarrow{n \rightarrow +\infty} 0 \text{ and } \lambda_n^2 n^{1/2} \xrightarrow{n \rightarrow +\infty} \infty, \quad (19)$$

we have

$$h^2(\hat{p}_n, p_{f_\star, a_\star}) = O_{\mathbb{P}_\star}(\lambda_n^2) \quad \text{and} \quad I^2(\hat{f}_n) = O_{\mathbb{P}_\star}(1). \quad (20)$$

Sketch of proof. The proof relies on a *basic inequality* which controls the Hellinger risk $h^2(\hat{p}_n, p_{f_\star, a_\star})$ and the complexity of the estimator $I^2(\hat{f}_n)$ by the empirical process, see [25]. The control of this empirical process will be done in Proposition 3.6. We set, for any density function p on $\mathbb{R}^{2\ell}$,

$$g_p \stackrel{\text{def}}{=} \frac{1}{2} \ln \frac{p + p_{f_\star, a_\star}}{2p_{f_\star, a_\star}}. \quad (21)$$

Let \mathbb{P}_n be the empirical distribution based on the observations $\{Y_{2k}, Y_{2k+1}\}_{k=0}^{n-1}$, *i.e.*, for any measurable set A of $\mathbb{R}^{2\ell}$,

$$\mathbb{P}_n(A) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_A(Y_{2k}, Y_{2k+1}).$$

By (9) and (12), the basic inequality of [25, Lemma 10.5], states that:

$$h^2(\widehat{p}_n, p_{f_*, a_*}) + 4\lambda_n^2 I^2(\widehat{f}_n) \leq 16 \int g_{\widehat{p}_n} d(\mathbb{P}_n - \mathbb{P}_*) + 4\lambda_n^2 I^2(f_*) . \quad (22)$$

Therefore, a control of the term $\int g_{\widehat{p}_n} d(\mathbb{P}_n - \mathbb{P}_*)$ in the right hand side of (22) will provide simultaneously a bound on the growth of $h^2(\widehat{p}_n, p_{f_*, a_*})$ and $I^2(\widehat{f}_n)$.

The empirical process indexed by $W^{s,p}$ is defined, for any $f \in W^{s,p}$ and any $a \geq a_-$, by

$$\nu_n(g_{p_{f,a}}) \stackrel{\text{def}}{=} \sqrt{n} \int g_{p_{f,a}} d(\mathbb{P}_n - \mathbb{P}_*) ,$$

where $g_{p_{f,a}}$ is defined by (21). Proposition 3.6 provides a deviation inequality for the supremum of the normalized empirical process.

Proposition 3.6. *Assume H1-3, H4(i) and H5. There exist some positive constants K , Σ and T such that, for any $x > 0$,*

$$\mathbb{P}_* \left\{ \sup_{f \in W^{s,p}, a \geq a_-} \frac{|\nu_n(g_{p_{f,a}})|}{I^2(f) \vee 1} \geq T + x \right\} \leq K e^{-\Sigma x} . \quad (23)$$

Proposition 3.6 is proved in Section 5.2 below. It ensures that

$$\sup_{f \in W^{s,p}, a \geq a_-} \frac{|\int g_{p_{f,a}} d(\mathbb{P}_n - \mathbb{P}_*)|}{I^2(f) \vee 1} = O_{\mathbb{P}_*}(n^{-1/2}) .$$

Plugging this bound into (22) gives

$$(4 + O_{\mathbb{P}_*}(n^{-1/2} \lambda_n^{-2})) I^2(\widehat{f}_n) \leq 4I^2(f_*) + O_{\mathbb{P}_*}(n^{-1/2} \lambda_n^{-2}) .$$

By (19), this establishes the second statement of (20). Combining this result with (22) gives:

$$h^2(\widehat{p}_n, p_{f_*, a_*}) \leq O_{\mathbb{P}_*}(n^{-1/2}) + O_{\mathbb{P}_*}(\lambda_n^2)$$

which proves the first statement of (20) and concludes the proof of Theorem 3.5. \square

Equations (19) and (20) give a rate of convergence of $h^2(\widehat{p}_n, p_{f_*, a_*})$. This rate of convergence is slower than $n^{-1/2}$ but can be chosen as close as wanted to $n^{-1/2}$, *e.g.* we can choose $\lambda_n^2 = n^{-1/2} \ln n$.

On the other hand, $I^2(\widehat{f}_n) = O_{\mathbb{P}_*}(1)$ and the Sobolev embedding described in Remark 2.1 ensures that \widehat{f}_n belongs to some compact subset of \mathcal{C}^1 with probability converging to 1 as n tends to ∞ . Let $d_{\mathcal{C}^1}$ denotes the distance function on \mathcal{C}^1 associated with the norm $\|\cdot\|_{\mathcal{C}^1}$. Let also \mathcal{F}_* be the set of all the functions f in the same equivalence class as f^* modulo the isometric transformations of K , *i.e.*

$$\mathcal{F}_* \stackrel{\text{def}}{=} \{f; f \stackrel{\mathcal{I}}{\sim} f_*\} . \quad (24)$$

Theorem 3.7. *Assume H1-5. Let (\hat{f}_n, \hat{a}_n) be defined by (9) and $I(f)$ by (11). Then, provided that*

$$\lambda_n \xrightarrow[n \rightarrow +\infty]{} 0 \text{ and } \lambda_n^2 n^{1/2} \xrightarrow[n \rightarrow +\infty]{} \infty ,$$

we have,

$$d_{\mathcal{C}^1}(\hat{f}_n, \mathcal{F}_\star) \xrightarrow[n \rightarrow +\infty]{} 0 \text{ and } \hat{a}_n \xrightarrow[n \rightarrow +\infty]{} a_\star \text{ in } \mathbb{P}_\star - \text{probability} , \quad (25)$$

where \mathcal{F}_\star is defined by (24).

Proof. We prove (25) introducing the Alexandroff compactification $[a_-, \infty]$ of $[a_-, \infty[$ and a distance function on this set such that $[a_-, \infty]$ is compact and metric. Moreover, for any $I > 0$, the set $\mathcal{B}_{W^{s,p}}(0, I)$, defined as the closure in \mathcal{C}_1 of $\{f \in W^{s,p}; I(f) \leq I\}$, is a compact subset of \mathcal{C}^1 . Thus, $\mathcal{B}_{W^{s,p}}(0, I) \times [a_-, \infty]$ is a compact subset of $\mathcal{C}^1 \times [a_-, \infty]$ and (25) will result from Theorem 3.5 and continuity arguments on the function $(f, a) \mapsto h^2(p_{f,a}, p_{f_\star, a_\star})$.

By Theorem 3.5, for any $\gamma > 0$, there exist $\epsilon > 0$ and $I > 0$ such that:

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_\star \{h^2(\hat{p}_n, p_{f_\star, a_\star}) > \epsilon \lambda_n^2\} \leq \frac{\gamma}{2} , \quad (26)$$

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_\star \{I(\hat{f}_n) > I\} \leq \frac{\gamma}{2} . \quad (27)$$

Denote by d the distance on $\mathcal{C}^1 \times [a_-, \infty]$ defined, for all $((f, a), (f', a')) \in (\mathcal{C}^1 \times [a_-, \infty])^2$ by

$$d((f, a), (f', a')) = d_{\mathcal{C}^1}(f, f') + |\arctan(a) - \arctan(a')| ,$$

with $\arctan(\infty) = \frac{\pi}{2}$. The distance on $[a_-, \infty]$ defined for any a and a' in $[a_-, \infty]$ by $|\arctan(a) - \arctan(a')|$ ensures its compactness. Therefore $E \stackrel{\text{def}}{=} \mathcal{B}_{W^{s,p}}(0, I) \times [a_-, \infty]$ is a compact subset of $(\mathcal{C}^1 \times [a_-, \infty], d)$. We also set

$$d((f, a), (\mathcal{F}_\star, a_\star)) = \inf_{f' \in \mathcal{F}_\star} d((f, a), (f', a_\star)) .$$

For any $\eta > 0$, denote by E_η the following set

$$E_\eta \stackrel{\text{def}}{=} E \setminus \bigcup_{f' \in \mathcal{F}_\star} \{(f, a) \in \mathcal{C}^1 \times [a_-, \infty]; d((f, a), (f', a_\star)) < \eta\} ,$$

E_η is a non-empty and closed subset of E which is compact in $\mathcal{C}_1 \times [a_-, \infty]$, thus E_η is also a compact subset of $\mathcal{C}_1 \times [a_-, \infty]$. By the dominated convergence theorem, the function defined on E , by

$$(f, a) \mapsto h^2(p_{f,a}, p_{f_\star, a_\star})$$

is continuous relatively to the topology defined by the distance d on $\mathcal{C}^1 \times [a_-, \infty]$. The compactness of E_η implies that $h^2(p_{f,a}, p_{f_\star, a_\star})$ reaches its minimum on E_η .

Let ϵ_η be this minimum. By Theorem 3.1 and since, for any f in $\mathcal{B}_{W^{s,p}}(0, I)$, $h^2(p_{f,\infty}, p_{f_*, a_*}) > 0$, $\epsilon_\eta > 0$. Moreover,

$$\begin{aligned} \mathbb{P}_* \{d((\hat{f}_n, \hat{a}_n), (\mathcal{F}_*, a_*)) > \eta\} &\leq \mathbb{P}_* \{I(\hat{f}_n) > I\} + \mathbb{P}_* \{h^2(\hat{p}_n, p_{f_*, a_*}) > \epsilon \lambda_n^2\} \\ &+ \mathbb{P}_* \left\{ I(\hat{f}_n) \leq I, h^2(\hat{p}_n, p_{f_*, a_*}) \leq \epsilon \lambda_n^2, d\left(\left(\hat{f}_n, \hat{a}_n\right), (\mathcal{F}_*, a_*)\right) > \eta \right\}. \end{aligned}$$

However, if $I(\hat{f}_n) \leq I$ and $d\left(\left(\hat{f}_n, \hat{a}_n\right), (\mathcal{F}_*, a_*)\right) > \eta$, then \hat{f}_n belongs to E_η and $h^2(\hat{p}_n, p_{f_*, a_*}) \geq \epsilon_\eta$. Choosing n big enough such that $\epsilon \lambda_n^2 < \epsilon_\eta$,

$$\mathbb{P}_* \left\{ I(\hat{f}_n) \leq I, h^2(\hat{p}_n, p_{f_*, a_*}) \leq \epsilon \lambda_n^2, d\left(\left(\hat{f}_n, \hat{a}_n\right), (\mathcal{F}_*, a_*)\right) > \eta \right\} = 0.$$

and, by (26) and (27),

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_* \left\{ d\left(\left(\hat{f}_n, \hat{a}_n\right), (\mathcal{F}_*, a_*)\right) > \eta \right\} \leq \gamma.$$

Since γ can be chosen arbitrarily small, for any $\eta > 0$,

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_* \left\{ d_{\mathcal{C}_1}(\hat{f}_n, \mathcal{F}_*) + |\arctan(\hat{a}_n) - \arctan(a_*)| > \eta \right\} = 0,$$

and $\lim_{n \rightarrow \infty} d_{\mathcal{C}_1}(\hat{f}_n, \mathcal{F}_*) = 0$ in probability. Moreover, the function \tan being continuous on $[0, \frac{\pi}{2}[$ and since $\arctan(a_*) \neq \frac{\pi}{2}$, $\lim_{n \rightarrow \infty} |\hat{a}_n - a_*| = 0$ in probability. □

4 Numerical experiments

In this section, we suppose the parameter a_* to be known and illustrate the performance of the estimator \hat{f}_n defined by (9). For practical considerations, we choose $v = 1$ in (11) and $p = 2$. The theoretical results provided in Section 3 rely on the assumption that $v > 2\ell$. However, choosing $v = 1$ allows to define an algorithm easy to implement with good convergence behavior. Using $v > 1$ would imply more involved numerical procedures to obtain parameter estimates. Let n be a positive integer, in this section, we denote by \hat{f} the estimator defined by (9) that maximizes the function T defined by

$$\begin{aligned} T &: W^{s,2} \rightarrow \mathbb{R} \\ f &\mapsto \frac{1}{n} \sum_{k=0}^{n-1} \ln p_{f, a_*}(Y_{2k}, Y_{2k+1}) - \lambda_n^2 \|f\|_{W^{s,2}}^2. \end{aligned}$$

The HMM framework suggests to use an Expectation-Maximization (EM) type procedure, see [8]. This algorithm iteratively produces a sequence of estimates $\{\hat{f}^p\}_{p \geq 0}$. Assume the current parameter estimate is given by \hat{f}^p . The estimate

\widehat{f}^{p+1} is defined as one of the maximizer of the function Q defined by

$$f \mapsto Q(f, \widehat{f}^p) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E}_{\widehat{f}^p} [\ln p_{f, a_\star}(X_{2k}, Y_{2k}, X_{2k+1}, Y_{2k}) | Y_{2k}, Y_{2k+1}] - \lambda_n^2 \|f\|_{W^{s,2}}^2,$$

where $\mathbb{E}_{\widehat{f}^p}[\cdot]$ denotes the expectation under the law of the stationary HMM parameterized by \widehat{f}^p and where

$$p_{f, a_\star}(x, y, x', y') = \nu_{a_\star}(x) q_{a_\star}(x, x') \varphi(y - f(x)) \varphi(y - f(x')).$$

The differential of $f \mapsto Q(f, \widehat{f}^p)$ is given, for any $f, h \in W^{s,2}$, by

$$d_f Q(\cdot, \widehat{f}^p)(h) = S_{n,1}(\widehat{f}^p, f, h) + S_{n,2}(\widehat{f}^p, f, h) - 2\lambda_n^2 \sum_{0 \leq |\alpha| \leq s} \langle D^\alpha f, D^\alpha h \rangle_{L_2},$$

where

$$S_{n,1}(\widehat{f}^p, f, h) \stackrel{\text{def}}{=} \frac{1}{n\sigma^2} \sum_{k=0}^{n-1} \mathbb{E}_{\widehat{f}^p} [\langle h(X_{2k}), f(X_{2k}) - Y_{2k} \rangle | Y_{2k:2k+1}],$$

$$S_{n,2}(\widehat{f}^p, f, h) \stackrel{\text{def}}{=} \frac{1}{n\sigma^2} \sum_{k=0}^{n-1} \mathbb{E}_{\widehat{f}^p} [\langle h(X_{2k+1}), f(X_{2k+1}) - Y_{2k+1} \rangle | Y_{2k:2k+1}].$$

\widehat{f}^{p+1} is then defined as the function $f \in W^{s,2}$ such that for any $h \in W^{s,2}$, $d_f Q(\widehat{f}^p, \cdot)(h) = 0$. In the sequel, we choose $s = 2$ and $K = [0, 1]$, therefore, this implies, for any $h \in W^{2,2}([0, 1], \mathbb{R})$,

$$S_{n,1}(\widehat{f}^p, f, h) + S_{n,2}(\widehat{f}^p, f, h) - 2\lambda_n^2 \sum_{\alpha=0}^2 \langle f^{(\alpha)}, h^{(\alpha)} \rangle_{L_2} = 0. \quad (28)$$

This equation can be applied to any function h in $W_0^{2,2} \stackrel{\text{def}}{=} \{h \in W([0, 1], \mathbb{R}); h(0) = h(1) = 0\}$. Using integration by parts, this yields, for any component f_j and any $x \in [0, 1]$,

$$\left(1 + \frac{1}{2n\lambda_n^2\sigma^2} \sum_{k=0}^{n-1} \left\{ \phi_{2k|2k:2k+1}^{\widehat{f}^p, a}(x) + \phi_{2k+1|2k:2k+1}^{\widehat{f}^p, a}(x) \right\} \right) f_j(x) - f_j^{(2)}(x) + f_j^{(4)}(x) = \frac{1}{2n\lambda_n^2\sigma^2} \sum_{k=0}^{n-1} \left\{ Y_{2k} \phi_{2k|2k:2k+1}^{\widehat{f}^p, a}(x) + Y_{2k+1} \phi_{2k+1|2k:2k+1}^{\widehat{f}^p, a}(x) \right\}, \quad (29)$$

where $\phi_{2k|2k:2k+1}^{\widehat{f}^p, a_\star}$ and $\phi_{2k+1|2k:2k+1}^{\widehat{f}^p, a_\star}$ are the filtering distributions defined by

$$\begin{aligned}\phi_{2k|2k:2k+1}^{\widehat{f}^p, a_\star}(x) &\stackrel{\text{def}}{=} \frac{\int \nu_{a_\star}(x) q_{a_\star}(x, x') \varphi(Y_{2k} - \widehat{f}^p(x)) \varphi(Y_{2k+1} - \widehat{f}^p(x')) dx'}{p_{\widehat{f}^p, a_\star}(Y_{2k}, Y_{2k+1})}, \\ \phi_{2k+1|2k:2k+1}^{\widehat{f}^p, a_\star}(x') &\stackrel{\text{def}}{=} \frac{\int \nu_{a_\star}(x) q_{a_\star}(x, x') \varphi(Y_{2k} - \widehat{f}^p(x)) \varphi(Y_{2k+1} - \widehat{f}^p(x')) dx}{p_{\widehat{f}^p, a_\star}(Y_{2k}, Y_{2k+1})}.\end{aligned}$$

4.1 Numerical approximations

Let $N \geq 1$ be an integer. The differential system (29) is solved using a discretization of the state space $[0, 1]$ by $\{\frac{i}{N}\}_{i=0}^N$. The filtering distributions $\phi_{2k|2k:2k+1}^{\widehat{f}^p, a_\star}$ and $\phi_{2k+1|2k:2k+1}^{\widehat{f}^p, a_\star}$ are approximated by piecewise constant functions $\underline{\phi}_k^{\widehat{f}^p, a_\star}$ and $\overline{\phi}_k^{\widehat{f}^p, a_\star}$, defined by

$$\underline{\phi}_k^{\widehat{f}^p, a_\star}(x) \stackrel{\text{def}}{=} \sum_{i=0}^{N-1} \mathbf{1}_{[\frac{i}{N}, \frac{i+1}{N}[}(x) \underline{\varphi}_{i,k}^{\widehat{f}^p} \quad \text{and} \quad \overline{\phi}_k^{\widehat{f}^p, a_\star}(x) \stackrel{\text{def}}{=} \sum_{i=0}^{N-1} \mathbf{1}_{[\frac{i}{N}, \frac{i+1}{N}[}(x) \overline{\varphi}_{i,k}^{\widehat{f}^p},$$

where, for any $i \in \{0, \dots, N-1\}$, $\underline{\varphi}_{i,k}^{\widehat{f}^p}$ (resp. $\overline{\varphi}_{i,k}^{\widehat{f}^p}$) is the approximation of $\underline{\phi}_k^{\widehat{f}^p, a_\star}(\frac{i}{N})$ (resp. $\overline{\phi}_k^{\widehat{f}^p, a_\star}(\frac{i}{N})$) obtained with an Euler scheme. The equation (29) is solved on each interval $[\frac{i}{N}, \frac{i+1}{N}[$, $i \in \{0, \dots, N-1\}$, which is straightforward since the coefficients are constant and the equation is linear. For any $i \in \{0, \dots, N-1\}$ and any $j \in \{0, \dots, \ell\}$, the solution $f_{j,i}$ on the interval $[\frac{i}{N}, \frac{i+1}{N}[$ belongs to some affine space of dimension 4. Thus, $4N$ parameters have to be chosen to uniquely determine the solution $\widehat{f}_j^{p+1} = \sum_{i=0}^{N-1} \mathbf{1}_{[\frac{i}{N}, \frac{i+1}{N}[}(f_{j,i})$. The \mathcal{C}^3 -regularity conditions for each boundary provides $4(N-1)$ equations and solving (28) with $h(x) = 1$, $h(x) = x$, $h(x) = x^2$ and $h(x) = x^3$ leads to four other linear equations which conclude the computation of \widehat{f}_j^{p+1} . The procedure is displayed in Algorithm 1. The numerical approximations and the computations of all the constants are detailed in the supplement paper [12, Section 3]

4.2 Experimental results

The Algorithm 1 is applied with the Gaussian kernel ($a_\star = 1$):

$$\forall x \in \mathbb{R}, q(x) = \exp\left\{-\frac{1}{2}x^2\right\}.$$

The aim is first to estimate the function (in this case $\ell = 3$)

$$\begin{aligned}f_\star &: [0, 1] \rightarrow \mathbb{R}^3 \\ x &\mapsto (3x, 30(x-1/4)(x-1/2)(x-3/4), 2\cos(5x)) ,\end{aligned}$$

Algorithm 1 One iteration of the algorithm

Require: $N, \hat{f}^p, a_*, Y_{0:2n-1}$.
Ensure: \hat{f}^{p+1}

```

for  $i \in \{0, \dots, N\}$  do
  for  $k \in \{0, \dots, n-1\}$  do
    Compute  $\underline{\varphi}_{i,k}^{\hat{f}^p}$  and  $\overline{\varphi}_{i,k}^{\hat{f}^p}$ .
  end for
end for
for  $j \in \{1, \dots, \ell\}$  do
  for  $i \in \{0, \dots, N-1\}$  do
    Compute  $f_{j,i}$  by solving (29).
  end for
  Set  $\hat{f}_j^{p+1} = \sum_{i=0}^{N-1} \mathbf{1}_{[\frac{i}{N}, \frac{i+1}{N}]}[f_{j,i}$ .
end for

```

We use $\sigma^2 = 1$ and $N = 50$ to sample observations from the discretized model. The estimation is started with the estimate

$$\begin{aligned} \hat{f}^0 &: [0, 1] \rightarrow \mathbb{R}^3 \\ x &\mapsto (x, 0, 0) . \end{aligned}$$

The Algorithm 1 is run with $\lambda_n^2 = \lfloor c \ln(n) / \sqrt{n} \rfloor$. Figure 1 displays the estimate after 1, 2, 3 and 25 iterations with $n = 50000$ observations along with the true functions for each coordinate. Figure 1 shows that after few iterations of the algorithm, the estimate can recover the curvature of the function f_* , even with a flat initial estimate.

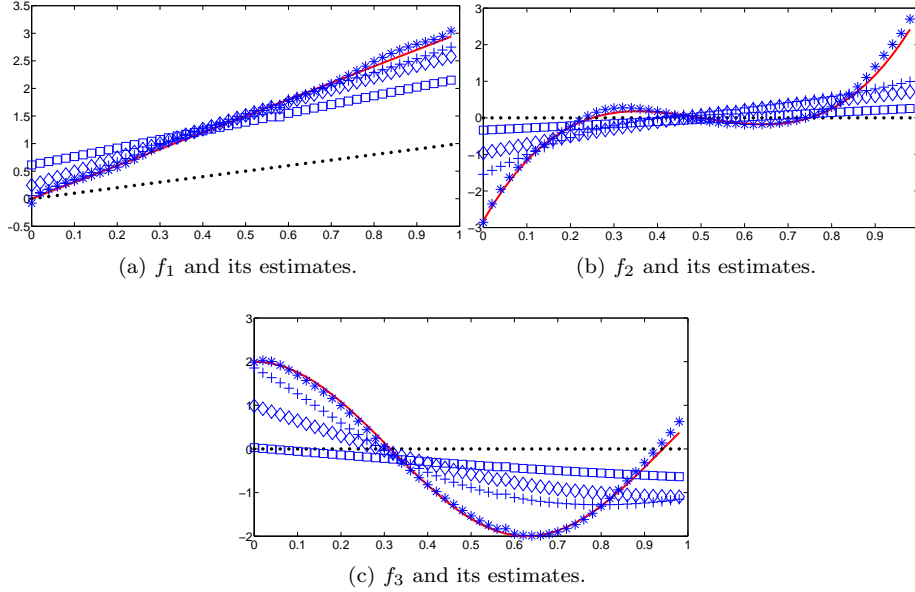


Figure 1: Estimation of f_1 , f_2 and f_3 after 25 iterations of the algorithm. The true function (bold line) and the initial estimate (dots) are displayed along with the estimates after 1 (squares), 2 (diamonds), 3 (crosses) and 25 (stars) iterations.

Figure 2 gives the evolution of the error as a function of the number of observations. We consider the L_2 -error and the L_∞ -error respectively defined, for $h_1, h_2 : [0, 1] \rightarrow \mathbb{R}$, by

$$\|h_1 - h_2\|_2 \stackrel{\text{def}}{=} \left(\frac{1}{N} \sum_{i=1}^N \left| h_1 \left(\frac{i}{N} \right) - h_2 \left(\frac{i}{N} \right) \right|^2 \right)^{1/2},$$

$$\|h_1 - h_2\|_\infty \stackrel{\text{def}}{=} \sup_{1 \leq i \leq N} \left| h_1 \left(\frac{i}{N} \right) - h_2 \left(\frac{i}{N} \right) \right|.$$

For each number of observations 50 independent Monte Carlo runs are used to compute the L_2 -error after 25 iterations of the algorithm. Figure 2 shows the median and the lower and upper quartiles over the 50 independent Monte Carlo runs.

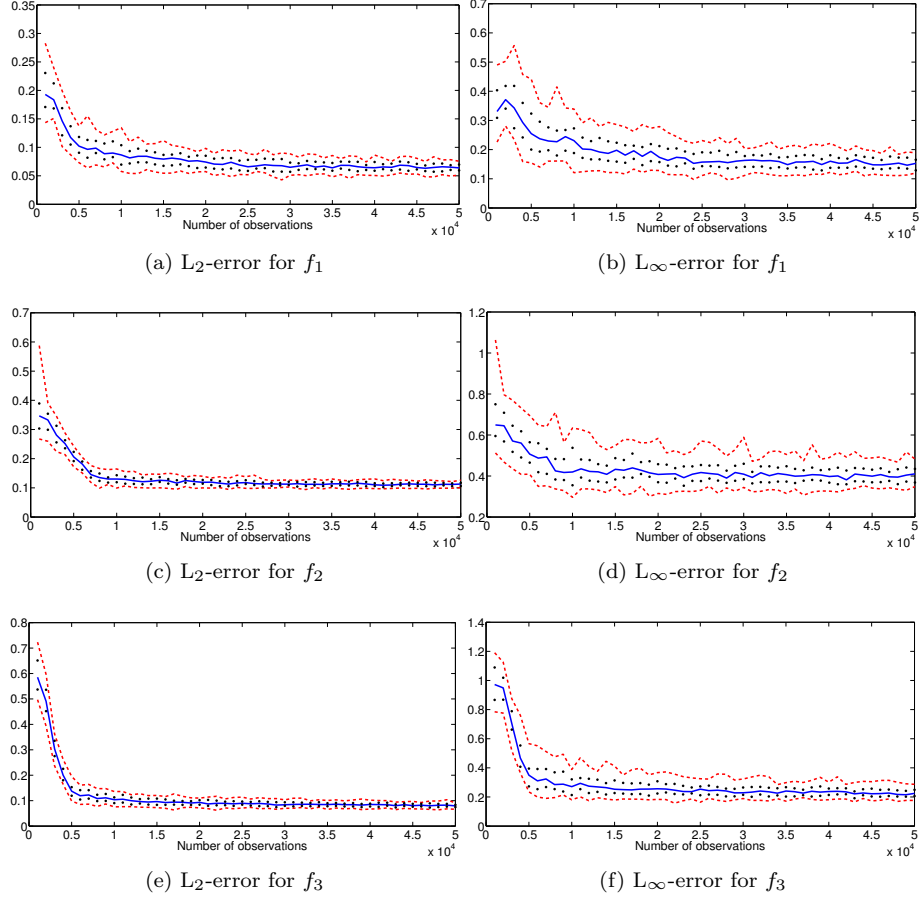


Figure 2: L₂ (left) and L_∞ (right) errors for each coordinate. The median (bold line), .25 and .75 quantiles (dotted lines) and .05 and .95 quantiles (balls) over 50 independent Monte Carlo runs are represented.

5 Proofs

5.1 Identifiability

Lemma (Lemma 3.2). *Assume H2(i) and H4. For all $x \in K$, $J_\phi(x) > 0$, where ϕ is defined by (15).*

Proof. By (15), (14) becomes

$$(\phi(X'_0), \phi(X'_1)) \stackrel{\mathcal{D}}{=} (X_0, X_1).$$

We now give an expression of the density of these two random vectors on $K \times K$.

Let h be a bounded measurable function of $K \times K$. We have

$$\mathbb{E}[h(\phi(X'_0), \phi(X'_1))] = \int h(\phi(x_0), \phi(x_1)) \nu_b(x_0) q_b(x_0, x_1) dx_0 dx_1 . \quad (30)$$

We introduce the set

$$A \stackrel{\text{def}}{=} \{z \in K; \forall x \in K \text{ s.t. } \phi(x) = z, J_\phi(x) > 0\} .$$

Let assume h is of the form

$$h(x_0, x_1) \stackrel{\text{def}}{=} h_2(x_0, x_1) \mathbf{1}_A(x_0) \mathbf{1}_A(x_1) , \quad (31)$$

where h_2 is any bounded measurable function. We have

$$\begin{aligned} \mathbb{E}[h(\phi(X'_0), \phi(X'_1))] &= \int h_2(\phi(x_0), \phi(x_1)) \nu_b(x_0) q_b(x_0, x_1) \\ &\quad \times \mathbf{1}_A(\phi(x_0)) \mathbf{1}_A(\phi(x_1)) dx_0 dx_1 \\ &= \int h_2(\phi(x_0), \phi(x_1)) \frac{\nu_b(x_0) q_b(x_0, x_1)}{J_\phi(x_0) J_\phi(x_1)} \mathbf{1}_A(\phi(x_0)) \mathbf{1}_A(\phi(x_1)) \\ &\quad \times J_\phi(x_0) J_\phi(x_1) dx_0 dx_1 . \end{aligned}$$

By [13, Theorem 2, p.99] and the area formula, for almost every $z \in K$, $\phi^{-1}(\{z\})$ is at most countable and we can apply the change of variable $z_0 = \phi(x_0)$, $z_1 = \phi(x_1)$.

$$\begin{aligned} \mathbb{E}[h(\phi(X'_0), \phi(X'_1))] &= \int h_2(z_0, z_1) \mathbf{1}_A(z_0) \mathbf{1}_A(z_1) \\ &\quad \times \sum_{\substack{x_0 \in \phi^{-1}(\{z_0\}) \\ x_1 \in \phi^{-1}(\{z_1\})}} \frac{\nu_b(x_0) q_b(x_0, x_1)}{J_\phi(x_0) J_\phi(x_1)} dz_0 dz_1 . \end{aligned}$$

Moreover,

$$\mathbb{E}[h(X_0, X_1)] = \int h_2(z_0, z_1) \nu_{a_*}(z_0) q_{a_*}(z_0, z_1) \mathbf{1}_A(z_0) \mathbf{1}_A(z_1) dz_0 dz_1 .$$

Therefore, for almost any $(z_0, z_1) \in K \times K$,

$$\nu_{a_*}(z_0) q_{a_*}(z_0, z_1) \mathbf{1}_A(z_0) \mathbf{1}_A(z_1) = \mathbf{1}_A(z_0) \mathbf{1}_A(z_1) \sum_{\substack{x_0 \in \phi^{-1}(\{z_0\}) \\ x_1 \in \phi^{-1}(\{z_1\})}} \frac{\nu_b(x_0) q_b(x_0, x_1)}{J_\phi(x_0) J_\phi(x_1)} .$$

By Sard Theorem (see [3]), since ϕ is \mathcal{C}^1 ,

$$\mu(\{z \in K; \exists x \in K, \phi(x) = z \text{ and } J_\phi(x) = 0\}) = 0 ,$$

Therefore, the function $z \mapsto \mathbf{1}_A(z)$ equals 1 almost everywhere in K . Finally, for almost any $(z_0, z_1) \in K \times K$,

$$\nu_{a_*}(z_0)q_{a_*}(z_0, z_1) = \sum_{\substack{x_0 \in \phi^{-1}(\{z_0\}) \\ x_1 \in \phi^{-1}(\{z_1\})}} \frac{\nu_b(x_0)q_b(x_0, x_1)}{J_\phi(x_0)J_\phi(x_1)}. \quad (32)$$

Let us assume that there exists $x_0 \in K$ such that $J_\phi(x_0) = 0$. There also exists $x \in K$ such that $J_\phi(x) > 0$ (otherwise $\mu(K) = \mu(\phi(J_\phi^{-1}(\{0\}))) = 0$ by Sard Theorem). By the mean value theorem, for all large enough $k \in \mathbb{N}^*$, there exists $x_k \in K$ such that $J_\phi(x_k) = \frac{1}{k}$. By the inverse function theorem, there exists U_k neighbourhood of x_k in K such that $\phi|_{U_k}$ is a diffeomorphism. J_ϕ being continuous, there also exists a neighbourhood V_k such that, for all $x \in V_k$, $|J_\phi(x) - J_\phi(x_k)| \leq \frac{1}{2k}$. Therefore, for all x in V_k ,

$$\frac{3}{2k} \geq J_\phi(x) \geq J_\phi(x_k) - \frac{1}{2k} = \frac{1}{2k}.$$

Let $W_k = U_k \cap V_k$, $\phi|_{W_k}$ is a diffeomorphism and $\mu(\phi(W_k)) > 0$. Therefore, there exists $(z_{k,0}, z_{k,1}) \in \phi(W_k) \times \phi(W_k)$ such that (32) is true. We denote by $x_{k,0}$ and $x_{k,1}$ the unique elements of W_k such that $z_{k,0} = \phi(x_{k,0})$ and $z_{k,1} = \phi(x_{k,1})$. Then,

$$\begin{aligned} \nu_{a_*}(z_{k,0})q_{a_*}(z_{k,0}, z_{k,1}) &= \sum_{\substack{x_0 \in \phi^{-1}(\{z_{k,0}\}) \\ x_1 \in \phi^{-1}(\{z_{k,1}\})}} \frac{\nu_b(x_0)q_b(x_0, x_1)}{J_\phi(x_0)J_\phi(x_1)} \\ &\geq \frac{\nu_b(x_{k,0})q_b(x_{k,0}, x_{k,1})}{J_\phi(x_{k,0})J_\phi(x_{k,1})} \geq \frac{2k}{3} \nu_b(x_{k,0})q_b(x_{k,0}, x_{k,1}). \end{aligned}$$

By H2(i), $(x_0, x_1) \mapsto \nu_c(x_0)q_c(x_0, x_1)$ is bounded for any $0 < c \leq \infty$: there exists $0 < C_c^- < C_c^+$ such that, for any $(x_0, x_1) \in K^2$, $0 < C_c^- \leq \nu_c(x_0)q_c(x_0, x_1) \leq C_c^+$, we have, for any $k \geq 1$ large enough,

$$C_a^+ \geq \nu_{a_*}(z_{k,0})q_{a_*}(z_{k,0}, z_{k,1}) \geq C_b^- \frac{2k}{3},$$

which is absurd and concludes the proof. \square

Lemma (Lemma 3.3). *Assume H2(i) and H4. Then, the function ϕ defined by (15) is a covering map.*

Proof. (i) comes from the continuity of $(f_*)^{-1}$ and f and (ii) is true since $\text{Im}(f_*) = \text{Im}(f)$. For (iii), let $z \in K$ and assume the set $\phi^{-1}(\{z\}) \stackrel{\text{def}}{=} \{x \in K; \phi(x) = z\}$ is infinite. By Lemma 3.2, ϕ is of full rank, then, by the inverse function theorem, for each $x \in \phi^{-1}(\{z\})$, there exists an open neighborhood V_x of x such that the function $\phi : V_x \rightarrow \phi(V_x)$ is a diffeomorphism and such that the $\{V_x\}_{x \in \phi^{-1}(\{z\})}$ are pairwise disjoint. By H2(i), there exists $n \in \mathbb{N}$ such that $\bigcup_{x \in \phi^{-1}(\{z\})} V_x$ can be covered with only n subsets of the form V_{x_i} ,

$x_i \in \phi^{-1}(\{z\})$, $i \in \{1, \dots, n\}$. Therefore, for any $x \in \phi^{-1}(\{z\})$, there exists $i \in \{1, \dots, n\}$ such that $x \in V_{x_i}$. If $x \notin \{x_i\}_{i=1}^n$, then $x \in V_x \cap V_{x_i}$ which is absurd since V_x and V_{x_i} are disjoint. Therefore, $\phi^{-1}(\{z\}) \stackrel{\text{def}}{=} \{x \in K; \phi(x) = z\}$ is finite and denoted by $\{x_i\}_{i=1}^n$, $n \in \mathbb{N}^*$. Let $\{V_i\}_{i=1}^n$ be disjoint open subsets of K such that $x_i \in V_i$ and define $V \stackrel{\text{def}}{=} \bigcap_{i=1}^n \phi(V_i)$, $O_i = \phi_{|V_i}^{-1}(V)$. Then, V is an open neighborhood of z which concludes the proof. \square

Lemma (Lemma 3.4). *Assume H2(ii). Then, every covering map $\phi : K \rightarrow K$ is a one to one function.*

Proof. Assume there exist x_1 and x_2 in K such that $x_1 \neq x_2$ and $\phi(x_1) = \phi(x_2) = y$. By H2(ii), K is path-connected and there exists a continuous path $\gamma : [0, 1] \rightarrow K$ such that $\gamma(0) = x_1$ and $\gamma(1) = x_2$. Then $\phi \circ \gamma$ is a continuous path taking values in K such that $\phi \circ \gamma(0) = \phi \circ \gamma(1) = y$. If $\tilde{\gamma}$ denotes the path defined by, for all $t \in [0, 1]$, $\tilde{\gamma}(t) = y$, then $\phi \circ \gamma$ and $\tilde{\gamma}$ are two paths in K with the same initial and terminal values. By H2(ii), K is simply connected and $\phi \circ \gamma$ and $\tilde{\gamma}$ are path homotopic (see [18, p.151]). The function $u : [0, 1] \rightarrow K$ such that, for all $t \in [0, 1]$, $u(t) = x_1$ is a lift (see [18, p.237]) of $\tilde{\gamma}$ for the covering map ϕ . Moreover, γ is a lift of $\phi \circ \gamma$ for the covering map ϕ . By the homotopy lifting property (see [18, Proposition 11.11, p.238]), since $u(0) = \gamma(0) = x_1$, then u and γ are path homotopic and have the same extremity: $x_1 = x_2$. This is absurd. \square

5.2 Proof of Proposition 3.6

Proposition 3.6 provides a deviation inequality on the empirical process renormalized by $I^2(f)$. First of all, for any $M \geq 1$ the Sobolev ball of radius M centred in 0 is denoted by $W_M^{s,p}$. Define the following collections of functions on $\mathbb{R}^{2\ell}$:

$$\mathcal{G}_M \stackrel{\text{def}}{=} \{g_{p,f,a}; f \in W_M^{s,p}, a > a_-\} \text{ and } \bar{\mathcal{G}}_M \stackrel{\text{def}}{=} \{g - \mathbb{E}_\star [g(Y_0, Y_1)]; g \in \mathcal{G}_M\} ,$$

where \mathbb{E}_\star is the expectation under the distribution \mathbb{P}_\star .

The first step of the proof establishes a deviation inequality on the empirical process restricted to the Sobolev balls $W_M^{s,p}$, $\sup_{g \in \mathcal{G}_M} |\nu_n(g)|$. The dependency in M of this inequality allows the determination of a lower bound on ν in the penalty (11) sufficient to establish Proposition 3.6. The second step and conclusion of the proof consists in using the peeling device with the decomposition:

$$W^{s,p} = W_1^{s,p} \cup \bigcup_{k \geq 0} \{W_{2^{k+1}}^{s,p} \setminus W_{2^k}^{s,p}\} ,$$

in order to apply the deviation inequalities on $\sup_{g \in \mathcal{G}_M} |\nu_n(g)|$, to each band $\{W_{2^{k+1}}^{s,p} \setminus W_{2^k}^{s,p}\}$. Proposition 5.1 gives a concentration inequality on the restricted empirical processes.

Proposition 5.1. *Assume H1, H2(i)(iii), H3 and H4(i). There exist some positive constants K_1, K_2, C and c , depending on f_\star and a_\star such that, for any $M \geq 1$, any $n \geq 1$ and any $t \geq Cn^{-1/2}$,*

$$\mathbb{P}_\star \left\{ \sup_{g \in \mathcal{G}_M} |\nu_n(g)| \geq c \mathbb{E}_\star \left[\sup_{g \in \mathcal{G}_M} |\nu_n(g)| \right] + Mt \right\} \leq K_1 \left(e^{-K_2 t^2} + e^{-K_2 t} \right). \quad (33)$$

The proof of Proposition 5.1 is given in the supplement paper [12, Section 2] and relies on the concentration results of [1]. It remains to control $\mathbb{E}_\star \left[\sup_{g \in \mathcal{G}_M} |\nu_n(g)| \right]$ for any $M \geq 1$.

Proposition 5.2. *Assume H1, H2(i)(iii), H3, H4(i) and H5. There exists a positive constant K depending on v , such that, for any $M \geq 1$,*

$$\mathbb{E}_\star \left[\sup_{g \in \mathcal{G}_M} |\nu_n(g)| \right] \leq KM^{v+1}. \quad (34)$$

The proof of Proposition 5.2 is given in Appendix B. We now combine Proposition 5.1 and Proposition 5.2 to obtain a deviation inequality on the empirical process restricted to the truncated collection of functions \mathcal{G}_M . Let $\eta > 0$. There exist K_1, K_2 and K_3 such that for any $M \geq 1$, any $n \geq 1$ and any $t \geq \frac{C}{\sqrt{n}}$,

$$\mathbb{P}_\star \left\{ \sup_{g \in \mathcal{G}_M} |\nu_n(g)| \geq K_3 M^{v+1} + Mt \right\} \leq K_1 \left(e^{-K_2 t^2} + e^{-K_2 t} \right). \quad (35)$$

Proposition 3.6 is obtained applying the peeling device as in [25, Lemma 5.14]. Let $\{x_k\}_{k \in \mathbb{N}^*}$ be some chosen weights such that,

$$\sum_{k \geq 1} e^{-x_k} < +\infty \quad \text{and, for any } k \geq 1, \quad C \vee 1 \leq x_k \leq 2^{kv}.$$

Let $k \geq 0$, for any positive x , if $t \stackrel{\text{def}}{=} x + x_k$, for any $n \geq 1$, we have $t \geq C \geq \frac{C}{\sqrt{n}}$. Since $t \geq x_k \geq 1$ and $x_k \leq 2^{kv}$, we have $e^{-K_2 t^2} \leq e^{-K_2 t}$ and $t \leq 2^{kv}(x + 1)$. Plugging these relations into (35) leads to

$$\begin{aligned} \mathbb{P}_\star \left\{ \sup_{f \in W_{2^k}^{s,p}, a \geq a_-} \left| \int g_{p_{f,a}} d(\mathbb{P}_n - \mathbb{P}_\star) \right| \geq \frac{2^{k(v+1)}}{\sqrt{n}} (K'_3 + x) \right\} \\ \leq K'_1 e^{-K_2(x+x_k)}, \quad (36) \end{aligned}$$

where $K'_1 \stackrel{\text{def}}{=} 2K_1$ and $K'_3 \stackrel{\text{def}}{=} K_3 + 1$. If $T \stackrel{\text{def}}{=} 2^{v+1}K'_3$,

$$\begin{aligned} & \mathbb{P}_\star \left\{ \sup_{f \in W^{s,p}, a \geq a_-} \frac{|\int g_{p_f,a} d(\mathbb{P}_n - \mathbb{P}_\star)|}{I^2(f) \vee 1} \geq \frac{T+x}{\sqrt{n}} \right\} \\ & \leq \mathbb{P}_\star \left\{ \sup_{f \in W_1^{s,p}, a \geq a_-} \left| \int g_{p_f,a} d(\mathbb{P}_n - \mathbb{P}_\star) \right| \geq \frac{T+x}{\sqrt{n}} \right\} \\ & + \sum_{k=0}^{\infty} \mathbb{P}_\star \left\{ \sup_{f \in W_{2^{k+1}}^{s,p}, a \geq a_-} \left| \int g_{p_f,a} d(\mathbb{P}_n - \mathbb{P}_\star) \right| \geq 2^{k(v+1)} \frac{T+x}{\sqrt{n}} \right\}. \end{aligned}$$

However, since $T \geq K'_3$ and $x_1 \geq 0$, by (36) applied with $k = 0$,

$$\mathbb{P}_\star \left\{ \sup_{f \in W_1^{s,p}, a \geq a_-} \left| \int g_{p_f,a} d(\mathbb{P}_n - \mathbb{P}_\star) \right| \geq \frac{T+x}{\sqrt{n}} \right\} \leq K'_1 e^{-K_2 x}.$$

Therefore, by the definition of T and by (36),

$$\begin{aligned} & \mathbb{P}_\star \left\{ \sup_{f \in W^{s,p}, a \geq a_-} \frac{|\int g_{p_f,a} d(\mathbb{P}_n - \mathbb{P}_\star)|}{I^2(f) \vee 1} \geq \frac{T+x}{\sqrt{n}} \right\} \\ & \leq K'_1 e^{-K_2 x} + \sum_{k=0}^{\infty} \mathbb{P}_\star \left\{ \sup_{f \in W_{2^{k+1}}^{s,p}, a \geq a_-} \left| \int g_{p_f,a} d(\mathbb{P}_n - \mathbb{P}_\star) \right| \right. \\ & \qquad \qquad \qquad \left. \geq (2^{k+1})^{v+1} \frac{K'_3 + x/2^{v+1}}{\sqrt{n}} \right\} \\ & \leq K'_1 e^{-K_2 x} + \sum_{k=0}^{\infty} K'_1 e^{-K_2(x/2^{v+1} + x_{k+1})}. \end{aligned}$$

This last equation ensures the existence of some positive constants K and Σ such that

$$\mathbb{P}_\star \left\{ \sup_{f \in W^{s,p}, a \geq a_-} \frac{|\int g_{p_f,a} d(\mathbb{P}_n - \mathbb{P}_\star)|}{I^2(f) \vee 1} \geq \frac{T+x}{\sqrt{n}} \right\} \leq K e^{-\Sigma x}.$$

A

For the sake of simplicity, for any $f \in W^{s,p}$ and $\mathbf{x} = (x_0, x_1) \in K \times K$, we set, for any $a > 0$,

$$\mathbf{f}(\mathbf{x}) \stackrel{\text{def}}{=} (f(x_0), f(x_1)) \in \mathbb{R}^{2\ell} \quad \text{and} \quad \boldsymbol{\nu}_a(\mathbf{x}) \stackrel{\text{def}}{=} \nu_a(x_0) q_a(x_0, x_1). \quad (37)$$

This appendix is devoted to the proof of an intermediate lemma on the envelope functions of the sets \mathcal{G}_M and $\bar{\mathcal{G}}_M$ defined, for any $\mathbf{y} \in \mathbb{R}^{2\ell}$, by

$$G_M(\mathbf{y}) \stackrel{\text{def}}{=} \sup_{g \in \mathcal{G}_M} g(\mathbf{y}) \quad \text{and} \quad \bar{G}_M(\mathbf{y}) \stackrel{\text{def}}{=} \sup_{g \in \bar{\mathcal{G}}_M} g(\mathbf{y}).$$

Lemma A.1. *Assume H2(i), H2(iii), H4(i) and H3. There exists a constant $C_G > 0$ such that, for any $\mathbf{y} \in \mathbb{R}^{2\ell}$,*

$$G_M(\mathbf{y}) \leq C_G (1 + M\|\mathbf{y}\|) .$$

Proof. For any $\mathbf{y} \in \mathbb{R}^{2\ell}$, any $f \in W_M^{s,p}$ and any $a \geq a_-$,

$$\begin{aligned} g_{p_{f,a}}(\mathbf{y}) &= \frac{1}{2} \ln \frac{1}{2} + \frac{1}{2} \ln \left(1 + \frac{p_{f,a}(\mathbf{y})}{p_{f_*,a_*}(\mathbf{y})} \right) \\ &\leq \frac{1}{2} \ln \frac{1}{2} + \frac{1}{2} \ln \left(1 + \sup_{\mathbf{x} \in K^2} \frac{\nu_a(\mathbf{x}) \exp(-\|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|^2/2\sigma^2)}{\nu_{a_*}(\mathbf{x}) \exp(-\|\mathbf{f}^*(\mathbf{x}) - \mathbf{y}\|^2/2\sigma^2)} \right) . \end{aligned}$$

By H2(i), (3), (4) and (37), there exists a constant $c_\nu > 1$ such that

$$\sup_{\mathbf{x} \in K^2} \frac{\nu_a(\mathbf{x})}{\nu_{a_*}(\mathbf{x})} \leq c_\nu .$$

Therefore,

$$g_{p_{f,a}}(\mathbf{y}) = \frac{1}{2} \ln \frac{1}{2} + \frac{1}{2} \ln \left(1 + c_\nu \sup_{\mathbf{x} \in K^2} \frac{\exp(-\|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|^2/2\sigma^2)}{\exp(-\|\mathbf{f}^*(\mathbf{x}) - \mathbf{y}\|^2/2\sigma^2)} \right) .$$

By H3 and H4(i) f_* is bounded and there exists a constant c such that

$$\frac{\exp(-\|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|^2/2\sigma^2)}{\exp(-\|\mathbf{f}^*(\mathbf{x}) - \mathbf{y}\|^2/2\sigma^2)} \leq \exp(c(1 + \|\mathbf{f}(\mathbf{x})\| \cdot \|\mathbf{y}\|)) .$$

Then, there exists a constant c such that

$$g_{p_{f,a}}(\mathbf{y}) \leq c(1 + \|\mathbf{f}(\mathbf{x})\| \cdot \|\mathbf{y}\|) ,$$

and the proof is concluded by (8). \square

Lemma A.1 implies that there exists a constant $C > 0$ such that, for any $\mathbf{y} \in \mathbb{R}^{2\ell}$,

$$\bar{G}_M(\mathbf{y}) \leq C (1 + M\|\mathbf{y}\|) . \quad (38)$$

B

We prove Proposition 5.2 using entropy with bracketing arguments on the class of functions \mathcal{G}_M . Define the class of function

$$\mathcal{P}_M \stackrel{\text{def}}{=} \{p_{f,a} : f \in W_M^{s,p}, a \geq a_-\} .$$

Let $\|\cdot\|$ be a norm on \mathcal{G} , the entropy with bracketing for the norm $\|\cdot\|$ is defined as follows:

Definition B.1. Let \mathcal{G} be some class of functions. For any positive δ , let $N_{\square}(\delta, \mathcal{G}, \|\cdot\|)$ be the smallest N such that there exist a set of brackets $\{[g_i^L, g_i^U]\}_{i=1}^N$ for which $\|g_i^U - g_i^L\| \leq \delta$ for all $i \in \{1, \dots, N\}$, and for any g in \mathcal{G} , there exist $i \in \{1, \dots, N\}$ such that

$$g_i^L \leq g \leq g_i^U .$$

$N_{\square}(\delta, \mathcal{G}, \|\cdot\|)$ is called the δ -number with bracketing of \mathcal{G} , and $H_{\square}(\delta, \mathcal{G}, \|\cdot\|) = \ln N_{\square}(\delta, \mathcal{G}, \|\cdot\|)$ is the δ -entropy with bracketing of \mathcal{G} .

Let $\mathbf{Y} \stackrel{\text{def}}{=} \{\mathbf{Y}_k\}_{k \in \mathbb{Z}}$ be the observations process defined, for all $k \in \mathbb{Z}$, by $\mathbf{Y}_k \stackrel{\text{def}}{=} (Y_{2k}, Y_{2k+1})$. A way of measuring the dependency of the process \mathbf{Y} is the determination of its β -mixing coefficients defined, for any $n \geq 1$,

$$\beta_n \stackrel{\text{def}}{=} \sup_{u > 0} \sup_{A \in \mathcal{G}_{u+n}^{\mathbf{Y}}} |\mathbb{P}_{\star}(A | \mathcal{H}_u^{\mathbf{Y}}) - \mathbb{P}_{\star}(A)| , \quad (39)$$

where $\mathcal{H}_u^{\mathbf{Y}} \stackrel{\text{def}}{=} \sigma(\mathbf{Y}_k, k \leq u)$ and $\mathcal{G}_{u+n}^{\mathbf{Y}} \stackrel{\text{def}}{=} \sigma(\mathbf{Y}_k, k \geq u+n)$. Let $\{\beta_n\}_{n \geq 1}$ be defined by (39), then, by combining [23, Chapter 9] and the results on the control of the ergodicity of Markov chains by coupling techniques of [9], it can be proved that there exist β in $(0, 1)$ and $C > 0$ such that, for any $n \geq 1$,

$$\beta_n \leq C\beta^n . \quad (40)$$

Define the mixing rate function $\beta(\cdot)$, by $\beta(t) \stackrel{\text{def}}{=} \beta_{\lfloor t \rfloor}$ if $t \geq 1$ and $\beta(t) = 1$ otherwise. For any numerical function g , we denote by \mathcal{Q}_g the quantile function of $|g(\mathbf{Y}_0)|$ and define the norm $\|g\|_{2,\beta}$ as in [10] by

$$\|g\|_{2,\beta} \stackrel{\text{def}}{=} \left[\int_0^1 \beta^{-1}(u) [\mathcal{Q}_g(u) du]^2 \right]^{1/2} ,$$

where β^{-1} denotes the càdlàg inverse of the function $\beta(\cdot)$. We also denote by $\mathcal{L}_{2,\beta}(\mathbb{P}_{\star})$ the class of numerical functions g such that $\|g\|_{2,\beta} < \infty$.

Proposition B.2. Assume H2(i), H2(iii) and H3. For any $p' > 1$, $s' > 2\ell/p'$, any integer $r > 1$ and any even number b such that $b > s' + 2\ell(1 - 1/p')$, there exists a positive constant C such that:

$$\forall \epsilon > 0, M \geq 1, \quad H_{\square}(\epsilon, \mathcal{G}_M, \|\cdot\|_{2,\beta}) \leq C \left(\frac{M^{s'+b+\frac{2}{p'}\ell}}{\epsilon^{2r}} \right)^{2\ell/s'} .$$

The proof of Proposition B.2 is given in Appendix C. Proposition B.2 allows to apply [10, Theorem 3] to the class of functions \mathcal{G}_M . Let B be the function defined on \mathbb{R}_+ by $B(x) \stackrel{\text{def}}{=} \int_0^x \beta^{-1}(t) dt$ and, for any $\epsilon > 0$, $\delta_M(\epsilon) \stackrel{\text{def}}{=} \sup_{t \leq \epsilon} \mathcal{Q}_{G_M}(t) \sqrt{B(t)}$. The following lemma is an application of [10, Lemma 2] it

allows to bound the $\|\cdot\|_{2,\beta}$ -norm by $\|\cdot\|_{L_{2r}}$ for all $r > 1$. For any g in $\mathcal{L}_{2,\beta}$ and any $r > 1$,

$$\|g\|_{2,\beta} \leq \|g\|_{L_{2r}(\mathbb{P}_*)} \sqrt{\int_0^1 u^{-1/r} \beta^{-1}(u) du}. \quad (41)$$

Moreover, by [20, Lemma 7.26], for any natural number $r > 1$ there exist a positive constant C such that for any f in $W^{s,p}$ and $a > 0$,

$$\|g_{p_f,a}\|_{L_{2r}(\mathbb{P}_*)}^{2r} \leq Ch(p_{f,a}, p_{f^*,a_*}). \quad (42)$$

The Hellinger distance being bounded, (42) and (41) state the existence of a positive number d such that $\|g_{p_f,a}\|_{2,\beta} \leq d$ for all f in $W^{s,p}$ and $a \geq a_-$. Define for any $M \geq 1$, $\varphi_M \stackrel{\text{def}}{=} \int_0^d \sqrt{H_{\square}(u, \mathcal{G}_M, \|\cdot\|_{2,\beta})} du$. Thus, by [10, Theorem 3], provided that $\delta_M(\epsilon) \xrightarrow{\epsilon \rightarrow 0} 0$, there exists a constant C such that

$$\mathbb{E}_* \left[\sup_{g \in \mathcal{G}_M} |\nu_n(g)| \right] \leq C \varphi_M \left(1 + \frac{\delta_M(1 \wedge \epsilon_{n,M})}{d} \right), \quad (43)$$

where $\epsilon_{n,M}$ is the unique solution on \mathbb{R}_+ of the equation:

$$\frac{x^2}{B(x)} = \frac{\varphi_M^2}{nd^2}.$$

In the sequel, we control the quantities appearing in (43). By Proposition B.2 and the definition of φ_M , for any $p' > 1$, $s' > 2\ell/p'$, $r > 1$ and any even number b such that $b > s' + 2\ell(1 - 1/p')$, there exists a constant C depending on p' , s' , r and b , such that

$$\varphi_M \leq C \left(M^{s'+b+\frac{2}{p'}\ell} \right)^{\ell/s'} \int_0^d u^{-2r\ell/s'} du, \quad (44)$$

with $\int_0^d u^{-2r\ell/s'} du < \infty$ whenever $s' > 2r\ell$. If b is the unique even number such that $s' + 2\ell(1 - 1/p') < b \leq [s' + 2\ell(1 - 1/p')] + 1$ and if s' tends to infinity in (44), then $\left(M^{s'+b+\frac{2}{p'}\ell} \right)^{\ell/s'} \xrightarrow{s' \rightarrow \infty} M^{2\ell}$ and it follows that,

$$\forall \eta > 0, \exists C > 0, \forall M \geq 1, \varphi_M \leq CM^{2\ell+\eta}.$$

By H5, there exists a constant C such that

$$\varphi_M \leq CM^v. \quad (45)$$

Lemma B.3. *Assume H4(i) and H3. There exists $C > 0$, such that, for any $M \geq 1$ and any $t \in (0, 1)$,*

$$\mathcal{Q}_{G_M}(t) \leq CM \left(1 + \ln^{1/2} \left(\frac{1}{t} \right) \right).$$

Proof. Set $\boldsymbol{\epsilon}_0 = (\epsilon_0, \epsilon_1)$, set $u > C_G$, where C_G is defined in Lemma A.1,

$$\begin{aligned} \mathbb{P}_\star \{G_M(\mathbf{Y}_0) \geq u\} &\leq \mathbb{P}_\star \{C_G(1 + M\|\mathbf{Y}_0\|) \geq u\} \\ &\leq \mathbb{P}_\star \left\{ \|\mathbf{Y}_0\| \geq \frac{u/C_G - 1}{M} \right\} \\ &\leq \mathbb{P}_\star \left\{ \|\mathbf{f}^\star(\mathbf{X}_0)\| + \|\boldsymbol{\epsilon}_0\| \geq \frac{u/C_G - 1}{M} \right\} \\ &\leq \mathbb{P}_\star \left\{ \|\boldsymbol{\epsilon}_0\| \geq \frac{u/C_G - 1}{M} - c_\infty \right\}, \end{aligned}$$

where $\|\mathbf{f}^\star(\mathbf{x})\| \leq c_\infty$ for all \mathbf{x} in K^2 (f_\star is bounded by H3 and H4(i)). Using Cirelson-Ibragimov-Sudakov inequality, see [20, Section 1.2.1], for any $x > 0$

$$\mathbb{P}_\star \left\{ \frac{1}{\sigma} (\|\boldsymbol{\epsilon}_0\| - \mathbb{E}(\|\boldsymbol{\epsilon}_0\|)) \geq x \right\} \leq e^{-\frac{x^2}{2}}.$$

Hence,

$$\begin{aligned} \mathbb{P}_\star \{G_M(\mathbf{Y}_0) \geq u\} &\leq \exp \left(-\frac{\left(\frac{u/C_G - 1}{M} - c_\infty - \mathbb{E}(\|\boldsymbol{\epsilon}_0\|) \right)^2}{2\sigma^2} \right) \\ &= \exp \left(-\frac{\left(\frac{c_1 u - 1}{M} - c_2 \right)^2}{2\sigma^2} \right), \end{aligned}$$

where $c_1 \stackrel{\text{def}}{=} \frac{1}{C_G}$ and $c_2 \stackrel{\text{def}}{=} c_\infty + \mathbb{E}_\star[\|\boldsymbol{\epsilon}_0\|]$. Setting $1 \geq t > 0$, let u be such that $t = \mathbb{P}_\star \{G_M(\mathbf{Y}_0) \geq u\}$, then,

$$\exp \left(-\frac{\left(\frac{c_1 u - 1}{M} - c_2 \right)^2}{2\sigma^2} \right) \geq t$$

implies

$$u \leq \frac{1}{c_1} \left(M c_2 + 1 + M \left(2\sigma^2 \ln \left(\frac{1}{t} \right) \right)^{1/2} \right),$$

which concludes the proof. \square

By (40), there exists a constant $C > 0$ such that,

$$\forall x \in (0, 1), B(x) \leq Cx \left(1 + \ln \left(\frac{1}{x} \right) \right). \quad (46)$$

Lemma B.4. *Assume H4(i) and H3. There exists $C > 0$ such that for any $0 < \epsilon \leq 1$,*

$$\delta_M(\epsilon) \leq CM \left(\epsilon^{1/2} \ln \left(\frac{1}{\epsilon} \right) \mathbf{1}_{\epsilon \leq e^{-2}} + \mathbf{1}_{\epsilon > e^{-2}} \right) \leq CM.$$

Proof. By Lemma B.3,

$$\mathcal{Q}_{G_M}(t) \leq CM \left(1 + \ln^{1/2} \left(\frac{1}{t} \right) \right) .$$

Therefore, by (46)

$$\mathcal{Q}_{G_M}(t) \sqrt{B(t)} \leq CM \left(1 + \ln^{1/2} \left(\frac{1}{t} \right) \right) \sqrt{t \left\{ 1 + \ln \left(\frac{1}{t} \right) \right\}} .$$

For $t \geq e^{-1}$, $\ln(t^{-1}) \leq 1$ and $\mathcal{Q}_{G_M}(t) \sqrt{B(t)} \leq CM$.
 For $t \leq e^{-1}$, $\ln(t^{-1}) \geq 1$, this yields, for $t \leq e^{-1}$,

$$\mathcal{Q}_{G_M}(t) \sqrt{B(t)} \leq CM \ln \left(\frac{1}{t} \right) \sqrt{t} .$$

The proof is concluded upon noting that the function $t \mapsto t^{1/2} \ln(t^{-1})$ reaches its maximum at e^{-2} . \square

Finally, Lemma B.4 ensures that $\delta_M(\epsilon) \xrightarrow{\epsilon \rightarrow 0} 0$ for any $M \geq 1$, and Proposition 5.2 results from (43), Lemma B.4, and (45).

C

The aim of this appendix is to prove Proposition B.2. The computation of $H_{\square}(\epsilon, \mathcal{G}_M, \|\cdot\|_{2,\beta})$ is not an easy task as the dependency of $\|g\|_{2,\beta}$ in g only appears through the quantile function \mathcal{Q}_g . Moreover, the dependency in M of the entropy $H_{\square}(\epsilon, \mathcal{G}_M, \|\cdot\|_{2,\beta})$ is not straightforward. The next lemma allows to control the bracketing entropy of \mathcal{G}_M relatively to the $\|\cdot\|_{2,\beta}$ -norm by the entropy of \mathcal{P}_M relatively to the $\|\cdot\|_{L_1(\mathbb{R}^{2\ell})}$ -norm.

Lemma C.1. *For any integer $r > 1$, there exists a constant C such that:*

$$H_{\square}(\epsilon, \mathcal{G}_M, \|\cdot\|_{2,\beta}) \leq CH_{\square}(\epsilon^{2r}, \mathcal{P}_M, \|\cdot\|_{L_1(\mathbb{R}^{2\ell})}) .$$

Proof. The function \ln being increasing, if $[P_U, P_L]$ is a bracket for \mathcal{P}_M , then $[g_{P_U}, g_{P_L}]$ is a bracket for \mathcal{G}_M . Moreover, by [20, Lemma 7.26], there exists a positive constant C such that

$$\|g_{P_U} - g_{P_L}\|_{L_{2r}(\mathbb{P}_*)}^{2r} \leq C \|\sqrt{P_U} - \sqrt{P_L}\|_{L_2(\mathbb{R}^{2\ell})}^2 .$$

Moreover it is straightforward that $\|\sqrt{P_U} - \sqrt{P_L}\|_{L_2(\mathbb{R}^{2\ell})}^2 \leq \|P_U - P_L\|_{L_1(\mathbb{R}^{2\ell})}$. The proof is concluded using (41). \square

[22] provides results on the entropy rates for function classes of Besov or Sobolev-type. Therefore, to control the entropy rate of \mathcal{P}_M we prove that it is included in some weighted Sobolev Space. Define the polynomial weighting

function $\langle \mathbf{y} \rangle^b \stackrel{\text{def}}{=} (1 + \|\mathbf{y}\|^2)^{b/2}$ parametrized by $b \in \mathbb{R}$ where $\mathbf{y} \in \mathbb{R}^{2\ell}$. Furthermore, define for $p' \geq 1$, and $s' > 2\ell/p'$ the weighted Sobolev space

$$W^{s',p'}(\mathbb{R}^{2\ell}, \langle \mathbf{y} \rangle^b) \stackrel{\text{def}}{=} \left\{ f : f \cdot \langle \mathbf{y} \rangle^b \in W^{s',p'}(\mathbb{R}^{2\ell}, \mathbb{R}) \right\}.$$

Lemma C.2. *Assume H2(i), H2(iii) and H3. For any $p' \geq 1$, $s' > 2\ell/p'$ and any even and positive number b , there exists a positive constant C such that*

$$\forall f \in W^{s,p}, \forall a \geq a_-, \|p_{f,a} \cdot \langle \mathbf{y} \rangle^b\|_{W^{s',p'}(\mathbb{R}^{2\ell}, \mathbb{R})} \leq C(1 \vee \|f\|_{W^{s,p}})^{s'+b+\frac{2}{p'}\ell}.$$

Proof. Let f be a function in $W^{s,p}$, for any $a \geq a_-$,

$$\|p_{f,a} \cdot \langle \mathbf{y} \rangle^b\|_{W^{s',p'}(\mathbb{R}^{2\ell}, \mathbb{R})}^{p'} = \sum_{|\alpha| \leq s'} \|D^\alpha (p_{f,a} \cdot \langle \mathbf{y} \rangle^b)\|_{L_{p'}}^{p'}.$$

Applying the general Leibniz rule component by component, for any $\alpha \in \mathbb{N}^{2\ell}$,

$$D^\alpha (p_{f,a} \cdot \langle \mathbf{y} \rangle^b) = \sum_{\alpha' \leq \alpha} \binom{\alpha}{\alpha'} D^{\alpha'} (\langle \mathbf{y} \rangle^b) D^{\alpha - \alpha'} (p_{f,a}), \quad (47)$$

where $\binom{\alpha}{\alpha'} \stackrel{\text{def}}{=} \prod_{j=1}^{2\ell} \binom{\alpha_j}{\alpha'_j}$. Thus, Lemma C.2 results from the control of $\|D^{\alpha^{(1)}} (\langle \mathbf{y} \rangle^b) D^{\alpha^{(2)}} (p_{f,a})\|_{L_{p'}}$ for any given $\alpha^{(1)}$ and $\alpha^{(2)}$ in $\mathbb{N}^{2\ell}$. It is straightforward that, for any α in $\mathbb{N}^{2\ell}$, there exists a polynomial function P_α whose degree does not exceed $|\alpha|$ such that, for any $\mathbf{y} \in \mathbb{R}^{2\ell}$,

$$D^\alpha p_{f,a}(\mathbf{y}) = \int_{\mathbf{x} \in K^2} P_\alpha(\mathbf{f}(\mathbf{x}) - \mathbf{y}) \exp\left\{-\frac{\|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|^2}{2\sigma^2}\right\} \nu_a(\mathbf{x}) d\mathbf{x}. \quad (48)$$

Moreover, since b is an even number, that for any $\alpha \in \mathbb{R}^{2\ell}$ such that $|\alpha| \leq b$, $D^\alpha \langle \mathbf{y} \rangle^b$ is a polynomial function denoted by $P_{b,\alpha}$ whose degree does not exceed $b - |\alpha|$. In the case where $|\alpha| > b$, $D^\alpha \langle \mathbf{y} \rangle^b = 0$. Since $P_{\alpha^{(2)}}$ and $P_{b,\alpha^{(1)}}$ are both polynomial functions, and since (8) ensures that, for any \mathbf{x} in K^2 , $\|\mathbf{f}(\mathbf{x})\| \leq \sqrt{2}\kappa\|f\|_{W^{s,p}} \leq \sqrt{2}\kappa(1 \vee \|f\|_{W^{s,p}})$, there exist a constant C depending on $\alpha^{(1)}, \alpha^{(2)}$ and b such that, for any \mathbf{y} in $\mathbb{R}^{2\ell}$ and any \mathbf{x} in K^2 ,

$$\begin{aligned} & |P_{b,\alpha^{(1)}}(\mathbf{y}) P_{\alpha^{(2)}}(\mathbf{f}(\mathbf{x}) - \mathbf{y})| \\ & \leq C(1 + \|\mathbf{y}\|)^{b-|\alpha^{(1)}|} \mathbf{1}_{|\alpha^{(1)}| \leq b} \times \left(\sqrt{2}\kappa(1 \vee \|f\|_{W^{s,p}}) + \|\mathbf{y}\| \right)^{|\alpha^{(2)}|}. \end{aligned}$$

Define the following subset of $\mathbb{R}^{2\ell}$

$$A_f \stackrel{\text{def}}{=} \left\{ \mathbf{y} \in \mathbb{R}^{2\ell}; \|\mathbf{y}\| \leq \sqrt{2}\kappa(1 \vee \|f\|_{W^{s,p}}) \right\}.$$

$\|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|$ can be lower bounded by 0 when \mathbf{y} belongs to A_f and by $|\sqrt{2}\kappa(1 \vee \|f\|_{W^{s,p}}) - \|\mathbf{y}\||$ when \mathbf{y} belongs to A_f^c . Therefore, uniformly in $\mathbf{x} \in K^2$,

$$\exp\left\{-\frac{\|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|^2}{2\sigma^2}\right\} \leq \mathbf{1}_{A_f}(\mathbf{y}) + \mathbf{1}_{A_f^c}(\mathbf{y}) e^{-\frac{1}{2\sigma^2}(\sqrt{2}\kappa(1 \vee \|f\|_{W^{s,p}}) - \|\mathbf{y}\|)^2}.$$

Thus, there exists a constant $C > 0$, independent from a , such that, for any \mathbf{y} in $\mathbb{R}^{2\ell}$,

$$\begin{aligned} & \left| D^{\alpha^{(1)}}(\langle \mathbf{y} \rangle^b) D^{\alpha^{(2)}}(p_{f,a})(\mathbf{y}) \right| \\ & \leq C(1 \vee \|f\|_{W^{s,p}})^{\alpha^{(2)}} \cdot (1 + \|\mathbf{y}\|)^{b-|\alpha^{(1)}|} \left(1 + \frac{\|\mathbf{y}\|}{\sqrt{2\kappa}(1 \vee \|f\|_{W^{s,p}})} \right)^{|\alpha^{(2)}|} \\ & \quad \times \left[\mathbf{1}_{A_f}(\mathbf{y}) + \mathbf{1}_{A_f^c}(\mathbf{y}) e^{-\frac{1}{2\sigma^2}(\sqrt{2\kappa}(1 \vee \|f\|_{W^{s,p}}) - \|\mathbf{y}\|)^2} \right]. \end{aligned}$$

Therefore, for any $p' \geq 1$,

$$\|D^{\alpha^{(1)}}(\langle \mathbf{y} \rangle^b) D^{\alpha^{(2)}}(p_{f,a})\|_{L^{p'}}^{p'} \leq C(1 \vee \|f\|_{W^{s,p}})^{p'\alpha^{(2)}} (I_1 + I_2),$$

where,

$$\begin{aligned} I_1 & \stackrel{\text{def}}{=} \int_{A_f} (1 + \|\mathbf{y}\|)^{p'(b-|\alpha^{(1)}|)} \left(1 + \frac{\|\mathbf{y}\|}{\sqrt{2\kappa}(1 \vee \|f\|_{W^{s,p}})} \right)^{p'|\alpha^{(2)}|} d\mathbf{y}, \\ I_2 & \stackrel{\text{def}}{=} \int_{A_f^c} (1 + \|\mathbf{y}\|)^{p'(b-|\alpha^{(1)}|)} \left(1 + \frac{\|\mathbf{y}\|}{\sqrt{2\kappa}(1 \vee \|f\|_{W^{s,p}})} \right)^{p'|\alpha^{(2)}|} \\ & \quad \times e^{-\frac{p'}{2\sigma^2}(\sqrt{2\kappa}(1 \vee \|f\|_{W^{s,p}}) - \|\mathbf{y}\|)^2} d\mathbf{y}. \end{aligned}$$

By applying the change of variable $\mathbf{y}' = \frac{1}{\sqrt{2\kappa}(1 \vee \|f\|_{W^{s,p}})} \mathbf{y}$ in I_1 and I_2 , and noting that $e^{-\frac{p'\sqrt{2\kappa}(1 \vee \|f\|_{W^{s,p}})}{2\sigma^2}(1 - \|\mathbf{y}'\|)^2} \leq e^{-\frac{\sqrt{2\kappa}p'}{2\sigma^2}(1 - \|\mathbf{y}'\|)^2}$, there exists a constant C such that

$$\|D^{\alpha^{(1)}}(\langle \mathbf{y} \rangle^b) D^{\alpha^{(2)}}(p_{f,a})\|_{L^{p'}}^{p'} \leq C(1 \vee \|f\|_{W^{s,p}})^{p'(|\alpha^{(2)}| - |\alpha^{(1)}| + b) + 2\ell}. \quad (49)$$

Using (49) in (47) with $\alpha^{(1)} = \alpha'$ and $\alpha^{(2)} = \alpha - \alpha'$ for any $|\alpha| \leq s'$ and $\alpha' \leq \alpha$ concludes the proof. \square

Hence Lemma C.2 ensures that, for any $p' \geq 1$, $s' > 2\ell/p'$ any even integer b , the renormalized classes of functions $\mathcal{P}_M/M^{s'+b+\frac{2}{p'}\ell}$, $M \geq 1$ belong to the same bounded subspace of $W^{s',p'}(\mathbb{R}^{2\ell}, \langle \mathbf{y} \rangle^b)$. By [22, Corollary 4], for any $p' \geq 1$, and any $s' > 2\ell/p'$, provided that $b > s' + 2\ell(1 - \frac{1}{p'})$, there exists a constant C such that

$$\forall \epsilon > 0, H_{\square} \left(\epsilon, \mathcal{P}_M/M^{s'+b+\frac{2}{p'}\ell}, \|\cdot\|_{L^1(\mathbb{R}^{2\ell})} \right) \leq C\epsilon^{-2\ell/s'}. \quad (50)$$

Lemma C.1 and (50) conclude the proof of Proposition B.2.

Acknowledgments

The authors are grateful to Elisabeth Gassiat and Eric Moulines for their fruitful remarks. They also wish to thank Aurélien Poiret and Frédéric Leroux for their advice.

References

- [1] R. Adamczak and W. Bednorz. Exponential concentration inequalities for additive functionals of markov chains. arXiv:1201.3569v1, Jan 2012.
- [2] R.A. Adams and J.J.F. Fournier. *Sobolev Spaces*. Number vol. 140 in Pure and Applied Mathematics. Academic Press, 2003.
- [3] T. Bröcker and L. Lander. *Differentiable Germs and Catastrophes*. London Mathematical Society Lecture Note Series. Cambridge University Press, 1975.
- [4] R.J. Carroll and P. Hall. Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, pages 1184–1186, 1988.
- [5] R.J. Carroll and L.A. Stefanski. Deconvolving kernel density estimators. *Statistics*, 21:169–184, 1990.
- [6] G. Churchill. Hidden Markov chains and the analysis of genome structure. *Computers & Chemistry*, 16(2):107–115, 1992.
- [7] F. Comte and M.-L. Taupin. Nonparametric estimation of the regression function in an errors-in-variables model. *Statistica sinica*, 17(3):1065–1090, 2007.
- [8] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B*, 39(1):1–38 (with discussion), 1977.
- [9] R. Douc, E. Moulines, and J. Rosenthal. Quantitative bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.*, 14(4):1643–1665, 2004.
- [10] P. Doukhan, P. Massart, and E. Rio. Invariance principle for absolutely regular processes. *Annales de l’Institut Henri Poincaré*, 31:393–427, 1995.
- [11] T. Dumont and S. Le Corff. Simultaneous localization and mapping problem in wireless sensor networks. Technical report, 2012.
- [12] T. Dumont and S. Le Corff. Supplement paper to “nonparametric estimation using partially observed markov chains”. Technical report, 2012.
- [13] L.C. Evans and R.F. Gariepy. *Measure Theory and Fine Properties of Functions*. Studies in Advanced Mathematics. CRC Press, 1992.
- [14] J. Fan and Y.K. Truong. Nonparametric regression with errors in variables. *Ann. Statist.*, 21:1900–1925, 1993.
- [15] B. Juang and L. Rabiner. Hidden Markov models for speech recognition. *Technometrics*, 33:251–272, 1991.

- [16] C. Lacour. Adaptive estimation of the transition density of a particular hidden markov chain. *Journal of Multivariate Analysis*, 99(5):787–814, 2008.
- [17] Claire Lacour. Nonparametric estimation of the stationary density and the transition density of a markov chain. *Stochastic Processes and their Applications*, 118(2):232 – 260, 2008.
- [18] J.M. Lee. *Introduction to Topological Manifolds*. Graduate Texts in Mathematics. Springer, 2000.
- [19] R.S. Mamon and R.J. Elliott. *Hidden Markov Models in Finance*, volume 104 of *International Series in Operations Research & Management Science*. Springer, Berlin, 2007.
- [20] P. Massart and J. Picard. *Concentration inequalities and model selection: Ecole d’Eté de Probabilités de Saint-Flour XXXIII - 2003*. Number vol. 1896 in *Ecole d’Eté de Probabilités de Saint-Flour*. Springer-Verlag, 2007.
- [21] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Communications and control engineering. Cambridge University Press, 2009.
- [22] R. Nickel and B.M. Potscher. Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov and Sobolev type. *J. Theor. Probab.*, 20:177–199, 2001.
- [23] E. Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*. Springer, 1990.
- [24] D.R. Smart. *Fixed point theorems*. Cambridge University Press, 1980.
- [25] S.A. Van De Geer. *Empirical Processes in M-Estimation*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2009.