Super optimal rates for nonparametric density estimation via projection estimators

F. Comte\textsuperscript{a,*}, F. Merlevède\textsuperscript{b}

\textsuperscript{a}Université Paris V, MAP5, UMR CNRS 8145, 45 rue des Saints-Pères, 75270 Paris cedex 06, France
\textsuperscript{b}Université Paris VI, Laboratoire de Probabilité et Modèles Aléatoires, UMR CNRS 7599, 4 Place Jussieu, 75252 Paris cedex 05, France

Received 24 November 2003; received in revised form 20 September 2004; accepted 13 December 2004
Available online 6 January 2005

Abstract

In this paper, we study the problem of the nonparametric estimation of the marginal density \( f \) of a class of continuous time processes. To this aim, we use a projection estimator and deal with the integrated mean square risk. Under Castellana and Leadbetter’s condition (Stoch. Proc. Appl. 21 (1986) 179), we show that our estimator reaches a parametric rate of convergence and coincides with the projection of the local time estimator. Discussions about the optimality of this condition are provided. We also deal with sampling schemes and the corresponding discretized processes.

\( \odot \) 2005 Elsevier B.V. All rights reserved.

Keywords: Castellana–Leadbetter’s condition; Continuous time projection estimator; Markov processes; Nonparametric estimation; Local time; Sampling

1. Introduction

1.1. The problem and the framework

Consider a weakly stationary process \( X = (X_t, t \in \mathbb{R}) \) observed either in continuous time for \( t \) varying in \([0, T] \) or in discrete time for \( t = t_1, \ldots, t_n \) and...
denote by \( f \) its common unknown marginal density with respect to Lebesgue measure.

In this paper, we are interested in the problem of giving nonasymptotic risk bounds in terms of the \( L_2 \)-integrated risk for an estimator \( \hat{f} \) of \( f \). Namely we study \( \mathbb{E}\|\hat{f} - f\|^2 \) where \( \|h\| = \left( \int_R h^2(x) \, dx \right)^{1/2} \) is the \( L_2(A) \)-norm and \( A \) is a compact set. We will consider both the process \((X_t, t \in \mathbb{R})\) observed continuously over the time interval \([0, T]\) and the process \( X = (X_t, t \in \mathbb{R}) \) observed at sampling instants \( \delta_n, 2\delta_n, \ldots, n\delta_n \) where \( \delta_n \to 0 \) and \( T_n = n\delta_n \to \infty \). Then we construct the discrete time process \((X_{t_i})_{1 \leq i \leq n}\) by setting \( t_i = i\delta_n \) for \( 1 \leq i \leq n \). In all the sequel, we denote by \( f_{(X_s,X_t)} \) the bivariate density of \((X_s,X_t)\). Moreover we will suppose throughout this paper that the process \( X \) belongs to the class \( \mathcal{X} \) defined as follows:

**Definition 1.** We define \( \mathcal{X} \) as the class of real processes \( X \) with common marginal density \( f \) with respect to Lebesgue measure on \( \mathbb{R} \) and such that the joint density of \((X_s,X_t)\) exists for all \( s \neq t, \) is measurable and satisfies \( f_{(X_s,X_t)} = f_{(X_t,X_s)} = f_{(X_0,X_{t-s})} \) and is denoted by \( f_{[t-s]} \) for all \( s, t \in \mathbb{R} \). We also denote by \( g_u, g_u = f_u - f \otimes f \) where \( (f \otimes f)(x,y) = f(x)f(y) \).

If \( X \) is assumed to belong to \( \mathcal{X} \), then it is weakly stationary. In particular strict stationarity is not required.

Historically, Castellana and Leadbetter [12] first introduced, in the context of processes belonging to \( \mathcal{X} \), the following condition, in order to exhibit a particular order for the rate of convergence of the quadratic risk of a nonparametric estimator of \( f \):

**CL.** \( u \mapsto \|g_u\|_{\infty} \) is integrable on \( ]0, \infty[ \) and \( g_u(\cdot, \cdot) \) is continuous at \((x,x)\) for each \( u > 0 \).

In the following, we use a slightly different condition:

**WCL.** There exists a positive integrable function \( k(\cdot) \) (defined on \( \mathbb{R} \)) such that

\[
\forall x \in \mathbb{R}, \sup_{y \in \mathbb{R}} \int_0^{+\infty} |g_s(x,y)| \, dt \leq k(x).
\]

If \( X \) is restricted to a compact set, then condition CL entails WCL. Condition WCL is in the spirit of Leblanc [31], who imposes in addition that \( k(\cdot) \) is bounded, and shows that this condition is satisfied for a wide class of diffusion processes. Note also that Assumption WCL does not impose continuity either and does not require \( g_s(x,y) \) to be integrable for all \( x, y \in \mathbb{R} \).

### 1.2. Some bibliographic remarks

The problem of estimating the marginal density of a continuous time process has been mainly studied using kernel estimators by Banon [1], Banon and N’Guyen [2], N’Gyen [33], and Bosq [9]. Under some mixing conditions, their pointwise non-integrated \( L_2 \)-risk (namely \( \mathbb{E}[\|\hat{f}_n(x) - f(x)\|^2] \) if \( \hat{f}_n \) denotes their estimator), reaches the standard rate of convergence \( T^{-2/s/(2s+1)} \) when \( f \) belongs to the Hölder class \( C^s \) and \( s \) is known, and these rates are minimax in their context. Castellana and
Leadbetter [12] proved, under the specific assumption on the joint density of \((X_0, X_t)\) described by CL, that the nonintegrated quadratic risk of kernel estimators can reach the parametric rate \(T^{-1}\). To be more precise, they enlightened the fact that if the distribution of \((X_0, X_t)\) is not too close to a singular distribution for \(|t|\) small, then the pointwise quadratic risk of the kernel density estimator can reach the “full rate”: \(T^{-1}\). In fact this can be explained as follows: local irregularities of the sample paths provide some additional information to the statistician.

The work of Castellana and Leadbetter led to a lot of works concerning the problem of estimating the common marginal distribution of a continuous time process. We refer to Bosq [8,10], Cheze-Payaud [13], Kutoyants [28], and Blanke [6], among others, for results of this kind and various examples. Kutoyants [29], in the case of diffusion processes, and Bosq and Davydov [11], in a more general context, have given an alternative to the kernel density estimator, by studying the local time density estimator which has the advantage to be an unbiased estimator of the density. In particular, Bosq and Davydov [11] have studied its properties and showed that the mean square error reaches the “full rate” \(T^{-1}\), under slightly weaker assumptions than Condition CL. Concerning now the study of the integrated risk, Leblanc [31] built a wavelet estimator of \(f\) when \(f\) belongs to some general Besov space and proved that its \(L_p\)-integrated risk converges at rate \(T^{-1}\) as well, provided that the process is geometrically strong mixing, still under a condition like CL. In this paper, we build a projection estimator for which we show that its \(L_2\)-integrated risk attains the parametric rate \(T^{-1}\) under WCL, but without any additional mixing condition. This rate is achieved without knowing the regularity of \(f\). Moreover, we provide counter-examples in order to prove that this rate cannot be attained if some parts of WCL are not fulfilled. However, if data are collected using a sampling scheme, one may ask if some optimal sampling schemes allow to keep parametric rates. Various schemes have been already proposed such as deterministic or randomized ones, see e.g., Masry [32] or Prakasa Rao [35]. In this paper, we consider some discretization schemes in accordance with the sample paths properties of the underlying process. Specifically we suppose that the statistician may dispose of frequent observations during a long time. We explore what kind of sampling schemes allow to recover the parametric rate and what is the influence of the sample step on the rate when we consider an adaptive procedure.

1.3. Outline of the paper

Since our study involves several technical conditions, we start by a general description of our results.

The first part of the work is devoted to the estimation of the density of the process when using continuous time observations. All the tools are available from the literature, in order to prove that the local time estimator, precisely defined below and denoted by \(\hat{f}\), satisfies, under WCL and some other additional conditions:

\[
\mathbb{E}\|\hat{f} - f\|^2 \leq \frac{2}{T} \int k(x) \, dx.
\]
It follows from this bound that this estimator reaches the “super optimal” rate $T^{-1}$.

The present paper states and improves this result in several directions. Indeed, the local time estimator may be uneasy to compute so that it is natural to look for more realistic continuous time estimators. We show, and this is both simple and new, that the projection $\hat{f}_S$ of $\hat{f}$ on a finite dimensional linear sub-space $S$ of $L^2(A)$ coincides with the minimum contrast estimator associated to the projection contrast

$$\gamma_T(h) = \|h\|^2 - \frac{2}{T} \int_0^T h(X_s) \, ds,$$

for $h$ a function belonging to $S$. This result allows to check, by using the known properties of the local time estimator $\hat{f}$, that $\hat{f}_S$ keeps the super optimal rate $T^{-1}$ when considering some standard finite dimensional functional spaces (trigonometric polynomials, piecewise polynomials, wavelets) with large enough dimension and as soon as the function to be estimated belongs to some class of regularity (described by Besov spaces). Another consequence of this result is that by considering directly the projection estimator, we can give a self-contained (namely, without using the local time estimator nor any of the results previously obtained for it) and quite simple proof of the super optimal rate of $\hat{f}_S$, under lighter conditions.

The second part of the work is concerned with discrete time observation of the continuous time process. First, we study the links that can be found between WCL, the condition for obtaining super optimal rates in continuous time, and the standard condition for obtaining standard rates in discrete time, namely a mixing condition. For this purpose, WCL is decomposed into a local irregularity condition WCL1 plus an asymptotic independence condition WCL2. Secondly, it follows from the first part of the work that it is natural to consider the minimum projection contrast estimator based on the discrete time sample. Then we illustrate that the standard discrete time results and rates can be generalized to arbitrary sample step, but do not lead to super optimal rates. The point then is to find what kind of conditions are required in order to replace WCL in discrete time. Such conditions are exhibited and discussed.

The last part of the work is devoted to the illustration of those conditions. Since they are quite technical, some examples of processes satisfying them are given, in classes of general Markov processes or more specifically among diffusion processes. The sharpness of condition WCL (namely of both WCL1 and WCL2) is also studied and counter-examples are provided in cases where either WCL1 or WCL2 is violated. Most of those examples and counter-examples are new and help to understand the meaning and the nonemptiness of some technical assumptions.

The paper is organized as follows. In Section 2, we exhibit the link between the local time estimator and our projection estimator. We explain how the rate of the integrated mean square risk of our projection estimator can be either deduced from the results concerning the local time estimator, or directly proved. Section 3 concerns discretely observed processes and the conditions required to keep parametric rates. Section 4 is devoted to some examples of Markov processes satisfying assumption WCL and to the study of the optimality of WCL. Detailed proofs of our results are postponed to Section 5.
2. The local time and the projection estimators

2.1. The local time estimator

The role of the local time in density estimation has been noticed by N’Guyen and Pham [34] and also by Doukhan and Leon [21]. Later Kutoyants has studied an unbiased estimator based on the local time when the observed process is a stationary diffusion process, whereas a more general context was studied in Bosq [8] and also in Bosq and Davydov [11]. Let us recall the definition of this estimator. First, if \( X \) is observed over \([0, T]\), its occupation measure \( n_T \) is defined by

\[
n_T(B) = \int_0^T 1_B(X_t) \, dt, \quad B \in \mathcal{B}_\mathbb{R},
\]

where \( \mathcal{B}_\mathbb{R} \) denotes the \( \sigma \)-algebra of Borel sets in \( \mathbb{R} \). If \( n_T \) is a.s. absolutely continuous with respect to Lebesgue measure \( \lambda \), then a local time for \( X \) is defined as a measurable random function \( l_T(x, \omega) \) such that \( l_T(\cdot, \omega) \) is a version of \( d n_T / d \lambda \) for almost all \( \omega \) in \( \Omega \). In the following “a.s.” is omitted and we use the notation \( l_T(x) \) instead of \( l_T(x, \omega) \). Obviously the problem of the existence of the local time arises. We refer to Geman and Horowitz [24, 25] for existence criteria. However it is more convenient to work with the conditions introduced in Bosq and Davydov [11] even if they are slightly stronger; namely,

A1. The function \( ((s, t), (x, y)) \mapsto f_{[t-s]}(x, y) \) is defined and measurable over \((D^c \cap [0, T]^2) \times U\) where \( U \) is an open neighborhood of \( D = \{(x, x), \ x \in \mathbb{R}\} \).

A2. The function \( F_T(x, y) = \int_{[0, T]^2} f_{(x, x)}(x, y) \, ds \, dt \) is finite in a neighborhood of \( D \) and is continuous at each point of \( D \) for all \( T \).

Remark 1. It is noteworthy that A1 together with A2 entail the existence of a square integrable local time (see [11]). Furthermore it is useful to notice that if we assume that

A3. \( g_u(\cdot, \cdot) \) is continuous at \( (x, x) \) for each \( u > 0 \),

in addition of WCL, then A1 and A2 hold.

Since \( l_T \) is the density of \( n_T \) it is natural to define a density estimator by setting

\[
\hat{f}(x) = \frac{l_T(x)}{T}, \quad x \in \mathbb{R}.
\]

This estimator is called the local time density estimator. It is shown in Bosq and Davydov [11] that \( \hat{f} \) is an unbiased density estimator which reaches the so-called parametric rate \( T^{-1} \). Indeed, according to Corollary 5.2 in Bosq and Davydov [11], if \( X \) belongs to \( \mathcal{X} \) and A1 and A2 hold, then

\[
\text{Var}(\hat{f}(x)) = \frac{2}{T} \int_0^T \left(1 - \frac{u}{T}\right) g_u(x, x) \, du, \quad x \in \mathbb{R}.
\]
A continuity condition is needed to get (2.2) since the proof is mainly based on the
fact that a kernel estimator is introduced to approximate the local time (see
Proposition 5.1 in Bosq and Davydov [11]). Using (2.2) together with the fact that \(\hat{f}\)
is an unbiased density estimator, we get that

\[
\mathbb{E}[\|\hat{f} - f\|^2] \leq \frac{2}{T} \int k(x) \, dx,
\]

provided that WCL and A3 hold.

2.2. Projection of the local time estimator

Let \(S\) be a linear subspace of \(L^2(A)\) with dimension \(D\), let \(\Pi_S\) denote the
orthogonal projection (in the \(L^2\)-sense) on \(S\) and let \((\varphi_{j,D})\) be an orthonormal
basis of \(S\). Assume that \(X\) is a continuous time process admitting a local time. Then
for any measurable function \(h\),

\[
\int_0^T h(X_t) \, dt = \int h(x) l_T(x) \, dx.
\]

It follows that the projection on \(S\) of \(\hat{f}\), denoted by \(\Pi_S \hat{f} = \hat{f}_S\) satisfies \(\hat{f}_S = \sum_{j=1}^D \hat{a}_{j,D} \varphi_{j,D}\) and since \(\hat{f} - \hat{f}_S\) is orthogonal to \(S\),

\[
\hat{a}_{j,D} = \langle \hat{f}, \varphi_{j,D} \rangle = \frac{1}{T} \int_A l_T(x) \varphi_{j,D}(x) \, dx = \frac{1}{T} \int_0^T \varphi_{j,D}(X_s) \, ds.
\]

It appears that \(\hat{f}_S\) coincides with the minimum contrast estimator associated with the
contrast function:

\[
\gamma_T(h) = \|h\|^2 - \frac{2}{T} \int_0^T h(X_s) \, ds.
\]

This is a new result which allows to consider the local time estimator in a quite
different way. To be more precise, we have that

\[
\hat{f}_S = \text{Argmin}_{h \in S} \gamma_T(h).
\]

Notice that \(\hat{f}_S\) is an unbiased estimator of \(f_S = \Pi_S f\), the orthogonal projection of \(f\)
on \(S\). As a consequence by using Pythagoras Theorem,

\[
\mathbb{E}[\|\hat{f}_S - f\|^2] = \|f - f_S\|^2 + \mathbb{E}[\|\Pi_S (\hat{f} - f)\|^2]
\]

as soon as the local time is square integrable. Therefore since \(\mathbb{E}[\|\Pi_S (\hat{f} - f)\|^2] \leq \mathbb{E}[\|\hat{f} - f\|^2]\), it follows from (2.3) and (2.5) that under WCL and A3,

\[
\mathbb{E}[\|\hat{f}_S - f\|^2] \leq \|f - f_S\|^2 + \frac{2}{T} \int k(x) \, dx.
\]
Inequality (2.6) is most useful since it allows to compute the rate of the estimator $\hat{f}_S$ in a straightforward way.

2.3. Application to Besov spaces

The order of the bias term, $\|f - f_S\|$, which appears in inequality (2.6) is usually given by classical theorems of approximation theory, and three examples of linear subspaces $S$ of $L_2(A)$ with dimension $D$ are standardly developed to make this order precise. Let us recall them.

Tr. Trigonometric spaces: $S$ is generated by $1$, $\cos(2\pi jx)$, $\sin(2\pi jx)$ for $j = 1, \ldots, m$, $A = [0, 1]$ and $D = 2m + 1$.

P. Regular piecewise polynomial spaces: $S$ is generated by $r$ polynomials of degree less or equal to $r - 1$ on each of the $m$ subintervals of equal length of $A$ (e.g., intervals $[(j-1)/m, j/m]$, for $j = 1, \ldots, m$ when $A = [0, 1]$), $D = rm$.

W. Dyadic wavelet generated spaces with regularity $r - 1$, as described e.g., in Donoho and Johnstone [19].

For a precise description of those spaces and their properties, we refer also to Birgé and Massart [4]. The quantity $\|f - f_S\|$ is known to be of order $D^{-\alpha}$ provided that $f$ belongs to some Besov space $B_{2,2}(A)$ with norm denoted by $\|\cdot\|_{2,2}$ (see DeVore and Lorentz [18, Chapter 2, Section 7], for the definition of these spaces and the associated norms) and $S$ is a regular model Tr, P or W. This consideration together with (2.6) leads to the following result.

**Proposition 2.** Consider a model $S$ in Tr, P or W with dimension $D$ and with $r > \alpha > 0$. Assume that the continuous time process $X = (X_t)_{t \in [0,T]}$ belongs to the class $\mathcal{X}$. In addition assume that WCL and A3 hold and let $L > 0$. Then the estimator $\hat{f}_S = \Pi S\hat{f}$ of $f$ defined by (2.4) satisfies

$$\sup_{f \in B_{2,2}(L)} \mathbb{E}\|f - \hat{f}_S\|^2 \leq C(\alpha, L)D^{-2\alpha} + \frac{2}{T} \int_R k(x) \, dx,$$

(2.7) where $B_{2,2}(L) = \{h \in B_{2,2}(A), |h|_{2,2} \leq L\}$, $C(\alpha, L)$ is a constant depending only on $\alpha$ and $L$.

From this result it appears that if the dimension of $S$ is great enough then the integrated quadratic risk of $\hat{f}_S$ reaches the parametric rate: $T^{-1}$. For instance if we consider a model $S$ in Tr, P or W with dimension $D = [T]$ and if $f$ belongs to some Besov space $B_{2,2}(A)$ with $1/2 < \alpha < r$, then

$$\sup_{f \in B_{2,2}(L)} \mathbb{E}\|f - \hat{f}_S\|^2 \leq \left(C(\alpha, L) + 2 \int R k(x) \, dx\right)/T,$$

(2.8) as soon as $X$ belongs to $\mathcal{X}$ and satisfies WCL and A3. The important point here is that, as soon as $\alpha > 1/2$, $D$ can be chosen independently of $\alpha$. 
It follows that, in continuous time, making the dimension of the projection space depend on the regularity of $f$ is not necessary provided that a local assumption on the irregularity of the sample paths is fulfilled. More generally, if $f$ belongs to some Besov space $\mathcal{B}_{a,2,\infty}(A)$, then the integrated $L^2$-risk of our projection estimator reaches the full rate $T^{-1}$ whatever $\alpha > 0$ provided that the dimension of the projection space is great enough ($|e^T|$ suits, even if it is not realistic from a practical point of view, but is independent of $\alpha$, contrary to $[T^{1/2\alpha}]$, which would be the right choice if $\alpha$ were known). This is not really surprising if we compare with what happens when kernel density estimators are considered: it amounts in that case to make the bandwidth as small as possible.

2.4. The direct study of the projection estimator

We want to emphasize here that it is possible to obtain straightforwardly inequality (2.6) without using any property of the local time estimator nor any reference to kernel estimators. We will see in particular, that condition A3 is not necessary to get the parametric rate $T^{-1}$. On the other hand, we cannot deal with all types of bases simultaneously. Indeed, to deal with the variance term in decomposition (2.5), i.e., $\mathbb{E}||\hat{f}_S - f_S||^2$, a localization constraint on the basis of $S$ is required:

A4. Let $\varphi_D := (\varphi_{\lambda,D})_{\lambda \in A}$ be an $\mathbb{L}_2$-orthonormal basis of $S$ with $\text{dim}(S) = D$. Then there exists a finite constant $C_\varphi$ not depending on $D$ but only on the structure of the basis such that

$$\sup_{x \in A} \int_A \left| \sum_{\lambda \in A} \varphi_{\lambda,D}(x) \varphi_{\lambda,D}(y) \right| \, dy \leq C_\varphi.$$

Note that A4 is satisfied by the Haar basis and more generally we can check that:

**Proposition 3.** The spaces $W$ and $P$ satisfy assumption A4.

The proof of Proposition 3 is straightforward and is therefore omitted. We only indicate that to check that the spaces $P$ satisfy assumption A4, we use Eq. (8) in Birgé and Massart [4].

Note also that this localization constraint excludes the trigonometric polynomial spaces for which a direct result would also be easy to prove, as it would be the case for any space generated by orthonormal functions not depending on the dimension $D$ of the space, i.e., such that $\varphi_{j,D} = \varphi_j$. In that case A4 is not required even in a direct computation but some other regularity conditions for the kernel $q_T(x,y) = \int_0^T (1 - u/T) g_u(x,y) \, du$ must be imposed instead.

To see how to study the variance term in (2.5), let us first define the following centered empirical process: for any function $h$,

$$v_T(h) = \frac{1}{T} \int_0^T [h(X_s) - \langle f, h \rangle] \, ds,$$

where $\langle f, h \rangle = \int f(x)h(x) \, dx$. 
Since \( \mathbb{E} \| \hat{f}_S - f_S \|^2 = \sum_{j=1}^{D} \mathbb{E} (\hat{a}_j, D - (f, \varphi_j, D))^2 \), \( \hat{a}_j, D - (f, \varphi_j, D) = v_T(\varphi_j, D) \) and \( v_T(\hat{f}_S - f_S) = \sum_{j=1}^{D} (\hat{a}_j, D - (f, \varphi_j, D))v_T(\varphi_j, D) \), we get that \( \mathbb{E} \| \hat{f}_S - f_S \|^2 = \sum_{j=1}^{D} \mathbb{E} (v_T(\varphi_j, D))^2 = \mathbb{E} (v_T(\hat{f}_S - f_S)) \). It follows that

\[
\mathbb{E}(\| \hat{f}_S - f \|^2) = \| f - f_S \|^2 + \sum_{j=1}^{D} \mathbb{E} (v_T(\varphi_j, D))^2
= \| f - f_S \|^2 + \mathbb{E} (v_T(\hat{f}_S - f_S)).
\] (2.9)

Then from weak stationarity, we have

\[
\mathbb{E}(\| \hat{f}_S - f_S \|^2) = \frac{1}{T^2} \sum_{j=1}^{D} \text{Var} \left( \int_0^T \varphi_j, D(X_s) \, ds \right)
= \frac{1}{T^2} \sum_{j=1}^{D} \int_0^T \int_0^T \text{cov}(\varphi_j, D(X_s), \varphi_j, D(X_u)) \, ds \, du
= \frac{2}{T^2} \sum_{j=1}^{D} \int_A \int_A \varphi_j, D(x)\varphi_j, D(y)
\times \left( \int_0^T (T - v)(f_v(x, y) - f(x)f(y)) \, dv \right) \, dx \, dy.
\] (2.10)

Therefore using WCL and A4, we derive that

\[
\mathbb{E}(\| \hat{f}_S - f_S \|^2) \leq \frac{2}{T} \int_A \int_A \left| \sum_{j=1}^{D} \varphi_j, D(x)\varphi_j, D(y) \right| k(x) \, dx \, dy \leq \frac{2C_0}{T} \int_A k(x) \, dx.
\]

This leads to the following result:

**Proposition 4.** Consider a linear subspace \( S \) of \( L_2(A) \) with dimension \( D \) and satisfying A4. Assume that the continuous time process \( X = (X_t)_{t \in [0, T]} \) belongs to the class \( X \) and that WCL holds. Then the estimator \( \hat{f}_S \) defined by (2.4) satisfies

\[
\mathbb{E}(\| \hat{f}_S - f \|^2) \leq \| f - f_S \|^2 + \frac{C_0 \kappa_A}{T},
\] (2.11)
where \( C_0 \) is defined in A4, \( \kappa_A = 2 \int_A k(x) \, dx \) and \( k(\cdot) \) is defined in WCL.

Note that compared to Section 2.2 we do not need to assume the continuity of \( g_u(\cdot, \cdot) \) at \( (x, u) \) for each \( u > 0 \). Indeed to obtain the bound given by (2.11), the study of the variance term is direct and does not use any preliminary study of the local time estimator, which is rather tedious and requires the use of kernel estimators to approximate the local time.

### 3. Rates for discretized processes

In continuous time, data are often collected by using a sampling scheme. In light of the results of Section 2, it appears that building a discretized projection estimator
must be a simpler idea than taking the projection of any discretization of the local time estimator. For that purpose, we consider the following contrast function associated to the observed process $\left( X_{k\delta_n} \right)_{1 \leq k \leq n}$:

$$\gamma_n^d(h) = \frac{1}{n} \sum_{i=1}^{n} [||h||^2 - 2h(X_{i\delta_n})].$$

Then we define our discretized projection estimator of $f$ as follows:

$$\hat{f}^d_S = \text{Argmin}_{h \in S} \gamma_n^d(h).$$

(3.1)

By introducing the discretized empirical centered process

$$v_n^d(h) = \frac{1}{n} \sum_{i=1}^{n} [h(X_{i\delta_n}) - \langle f, h \rangle],$$

(3.2)

and by proceeding as in the continuous time case, it is easy to see that

$$E(||\hat{f}^d_S - f||^2) = ||f - f_S||^2 + E(||\hat{f}^d_S - f_S||^2)$$

$$= ||f - f_S||^2 + \sum_{j=1}^{D} E((v_n^d)^2(\varphi_{j,D})).$$

3.1. Mixing assumptions and links with WCL

At this point, it is useful to note that Assumption WCL contains both a local irregularity condition and an asymptotic independence condition, respectively, for some $\theta_0 > 0$.

WCL1. There exists a positive integrable function $k(\cdot)$ defined on $\mathbb{R}$ such that, $\forall x \in \mathbb{R}$, $\sup_{y \in \mathbb{R}} \int_{0}^{\theta_0} |g_t(x, y)| \, dt \leq k(x)$.

and

WCL2. There exists a positive integrable function $k(\cdot)$ defined on $\mathbb{R}$ such that, $\forall x \in \mathbb{R}$, $\sup_{y \in \mathbb{R}} \int_{0}^{+\infty} |g_t(x, y)| \, dt \leq k(x)$.

This is illustrated in Section 4. Condition WCL1 means that the information provided by $(X_0, X_t)$ and $X_0$ respectively, differs significantly even if $t$ is small. In addition, as noticed by Bosq [8], Section 4, it also means that sample paths are not smooth.

In some situations (see Theorem 7), we will use the classical absolute regular coefficient, $\beta^*$, which quantifies the degree of inner dependence of the continuous (or discrete) time process $X$ and which is standardly defined as follows.

**Definition 5.** Let $\mathbb{P}_{\mathcal{U} \otimes \mathcal{V}}$ be the unique probability measure on $(\Omega \times \Omega, \mathcal{U} \otimes \mathcal{V})$ characterized by $\mathbb{P}_{\mathcal{U} \otimes \mathcal{V}}(U \times V) = \mathbb{P}(U \cap V)$. We denote by $\mathbb{P}_{\mathcal{U}}$ and $\mathbb{P}_{\mathcal{V}}$ the
restriction of the probability measure $\mathbb{P}$ to $\mathcal{U}$ and $\mathcal{V}'$, respectively. The $\beta$-mixing (or absolute regular) coefficient $\beta(\mathcal{U}, \mathcal{V}')$ of Rozanov and Volkonskii [37] is defined by

$$\beta(\mathcal{U}, \mathcal{V}') = \sup \{ |\mathbb{P}_{\mathcal{U} \otimes \mathcal{V}'}(C) - \mathbb{P}_{\mathcal{U}} \otimes \mathbb{P}_{\mathcal{V}'}(C)| : C \in \mathcal{U} \otimes \mathcal{V}' \}. \quad (3.3)$$

**Definition 6.** Let $\mathcal{F}_u^v$ be the $\sigma$-algebra of events generated by the random variables $\{X_t, u \leq t \leq v\}$. In the case of discrete time processes, $u, t, v$ are taken in a discrete set. A process $\{X_t\}_{t \in \mathbb{R}}$ (or $\mathbb{R}$) is said to be absolutely regular or $\beta^*$-mixing if

$$\beta_u^* := \sup_{t \in \mathbb{R}^+} \beta(\mathcal{F}_{-\infty}^t, \mathcal{F}_{u+t}^\infty) \to 0 \quad \text{as } u \to \infty. \quad (3.4)$$

Moreover the process is said to be arithmetically $\beta^*$-mixing with rate $\theta$ if there exists some $\theta > 0$ such that $\beta_u^* \leq (1 + u)^{-(1+\theta)}$ for all $u \in \mathbb{N}$ or $\mathbb{R}^+$. Similarly it will be said geometrically $\beta^*$-mixing with rate $\theta$ if there exists some $\theta > 0$ such that $\beta_u^* \leq e^{-\theta u}$ for all $u \in \mathbb{N}$ or $\mathbb{R}^+$.

According to (3.3), if we denote by $\beta_u$ the mixing coefficient

$$\beta_u = \sup_{t \in \mathbb{R}^+} \beta(\sigma(X_t), \sigma(X_{t+u})), \quad (3.5)$$

where $\sigma(X_s)$ is the $\sigma$-algebra generated by $X_s$, and if we suppose that the process takes its values in a compact set $A = [-R, R]$ (so that the density is compactly $A$-supported) and that WCL2 holds, we successively get for all $u_0 > 0$,

$$\int_{u_0}^{\infty} \beta_u \, du = \int_{u_0}^{\infty} \sup_{t \in \mathbb{R}^+} \sup_{B \in \mathcal{G}([-R, R]^2)} |\mathbb{P}_{X_t, X_{u+t}}(B) - \mathbb{P}_{X_t} \otimes \mathbb{P}_{X_{u+t}}(B)| \, du$$

$$= \int_{u_0}^{\infty} \sup_{B \in \mathcal{G}([-R, R]^2)} \left| \int_{B} (f_u(x, y) - f(x)f(y)) \, dx \, dy \right| \, du$$

$$\leq 2R \int_{-R}^{R} \left( \sup_{y \in \mathbb{R}} \int_{u_0}^{\infty} |g_u(x, y)| \, du \right) \, dx \leq 2R \int_{-R}^{R} k(x) \, dx < \infty. \quad (3.6)$$

Therefore in that case and since $\{\beta_u\}$ is always bounded by one, WCL2 implies that $\int_{0}^{\infty} \beta_u \, du$ is finite. This illustrates in what sense WCL2 can be viewed as an asymptotic independence condition. However note that WCL2 is not strong enough to derive some properties on the sequence of absolute regular coefficients $(\beta_u^*; u \in \mathbb{R}^+)$ involving the whole past and/or the whole future of the process, which are needed for instance in the results established in Leblanc [31]. Imposing some conditions on the coefficients $(\beta_u^*; u \in \mathbb{R}^+)$ rather than on the coefficients $(\beta_u^*; u \in \mathbb{R}^+)$ is clearly less restrictive since processes can be $\beta$-mixing without being $\beta^*$-mixing.

**3.2. The discrete adaptive procedure rate with small sampling step under mixing condition**

When the sampling step is fixed (i.e., $\delta_n = 1$), several adaptive procedures have been developed (see for instance Tribouley and Viennet [39] or Comte and Merlevède [14], both in the mixing case). They all aim to choose automatically the optimal
dimension of the projection space, without requiring a priori knowledge of the regularity of \( f \). The optimal rate obtained in that case is \( n^{-2a/(2a+1)} \) provided that \( f \) belongs to some Besov space of regularity \( a \). Moreover this rate is known to be minimax (see [20]).

We would like here to enlighten the influence of the mesh \( \delta_n \) of the observations on the optimal rate of convergence. To this aim, we use the same penalization procedure as done in Comte and Merlevède [14]; namely, we consider the penalized estimator

\[
\hat{f}_d = \sum_{m \in \mathcal{M}_n} \tilde{f}_S^m \quad \text{with} \quad \hat{m} = \text{Argmin}_{m \in \mathcal{M}_n} [\gamma_n^d(\tilde{f}_S^m) + \text{pen}(m)],
\]

where \((S_m)_{m \in \mathcal{M}_n}\) is a collection of spaces of the same kind as \( S \) with dimension denoted by \( D_m \) and where pen is a penalty function that happens to be of order \( D_m = n \):

**P1.** For each \( m \) in \( \mathcal{M}_n \), \( S_m \) is a linear subspace of \( L_2(A) \) with dimension \( D_m \) and \( N_n = \max_{m \in \mathcal{M}_n} D_m \) satisfies \( N_n \leq n \).

**P2.** There exists a constant \( \Phi_0 \) such that \( \forall m, m' \in \mathcal{M}_n, \forall h \in S_m \) and \( h' \in S_{m'} \),

\[
\|h + h'\| \leq \Phi_0 \sqrt{\text{dim}(S_m + S_{m'})} \|h + h'\|.
\]

**P3.** For any positive \( c \), \( \sum_{m \in \mathcal{M}_n} \sqrt{D_m} e^{-c\sqrt{D_m}} \leq \Sigma(c) \), where \( \Sigma(c) \) denotes a finite constant depending only on \( c \).

The three examples above (Tr, P, W) fulfill this set of assumptions, when their dimensions vary in the set of integers or of dyadic integers. In addition, it is noteworthy to indicate that, according to Lemma 6 of Birgé and Massart [5], property P2 is equivalent to the following property of any orthonormal basis \( \{\varphi_{j,D_m}\}_{j \in A_m} \) spanning \( S_m \) of dimension \( D_m \),

\[
\left\| \sum_{j \in A_m} \varphi_{j,D_m}^2 \right\|_\infty \leq \Phi_0^2 D_m.
\]

Moreover in order to develop our results, we need the following notations:

\[
A_r \coloneqq \int_0^\infty s^{r-2} \beta^*_s \, ds \quad \text{and} \quad B_r(\delta_n) \coloneqq \sum_{k=0}^\infty (k + 1)^{r-2} \beta^*_k \delta_n,
\]

where \( \{\beta^*_s\}_{s \in \mathbb{R}} \) is defined by (3.4) and provided the integral (or the series) is convergent.

**Theorem 7.** Consider a collection of models satisfying P1–P3. Let \((X_{i\delta_n})_{1 \leq i \leq n}\) be a discrete time sample of the continuous time process \( X = (X_t)_{t \in [0,T_n]} \), with \( n\delta_n = T_n \) and \( \delta_n = n^{-a} \) for some \( 0 < a < 1/2 \). Assume that \( X \) is weakly stationary with common marginal density \( f \) with respect to Lebesgue measure on \( \mathbb{R} \) and such that \( \|f\|_\infty < \infty \). In addition assume that the process is arithmetically \( \beta^* \)-mixing with mixing rate

\[
\theta > 3/(1 - 2a).
\]
Then the estimator $f^d$ defined by (3.7) with $\text{pen}(m) = \kappa \Phi_0^2 B_2(\delta_n) D_m/n$, where $\kappa$ is a universal constant, satisfies

$$
\mathbb{E}(\|f^d - f\|^2) \leq \inf_{m \in \mathcal{A}_n} \left( 3\|f - f_s\|^2 + K(1 + A_2)\Phi_0^2 \frac{D_m}{T_n} \right) + \frac{K'}{T_n},
$$

(3.11)

where $K$ is a numerical constant and $K'$ is a constant depending on $\Phi_0$, $\theta$, $A_2$, $A_3$ and $\|f_\mathcal{A} \mathcal{A}\|_{\infty}$. Moreover the choices $\delta_n = \ln^2(n)$ and $\text{pen}(m) = \hat{\kappa}(1 + A_2^{-2})\Phi_0^2 B_2(\delta_n) D_m/n$ (where $\hat{\kappa}$ is a universal constant) lead also to (3.11) provided that $\theta > 3$.

Note that the condition $\theta > 3$ is the one obtained in Comte and Merlevède [14] for $\delta_n = 1$. Besides, if the mixing is geometrical then $\delta_n = n^{-a}$ for any $a$ in $]0, 1[$ can be chosen, and no condition (3.10) is required.

Theorem 7 shows that the rate obtained does not really depend on the number of observations $n$ but rather on the length of the interval of observations, $T_n$. As soon as $f$ is assumed to belong to some Besov spaces, $\mathcal{B}_{2,2,\infty}$, then we reach the rate $T_n^{-2s/(2s+1)}$. In fact, if no assumption on the local behavior of the sample paths is imposed, then we are not able to improve the rate by considering more observations.

### 3.3. Using assumption WCL in discrete time

From the previous section, it appears that if we want the process $(X_{\delta_n})_{1 \leq i \leq n}$ to reach the parametric rate of convergence, assumptions on the local behavior of the sample paths have to be imposed. We use the following assumptions which are in the spirit of the ones introduced in Bosq [10], Section 4.5.3, in order to specify what he calls an admissible sampling, that is a sequence $(\delta_n)$ such that the super optimal rate for the mean square risk remains valid when the observations are $X_{\delta_n}, X_{2\delta_n}, \ldots, X_{n\delta_n}$ with a minimal sample size $n$.

**B1.** There exists a positive integrable function $M(\cdot)$ defined on $\mathbb{R}$ such that for all $u \in [0, u_0]$, $\forall x \in \mathbb{R}$, $\sup_{y \in \mathbb{R}} f_u(x, y) \leq M(x)u^{-\gamma}, \gamma \in [0, 1[$.

**B2.** There exist positive functions $k(\cdot)$ and $\pi(\cdot, \cdot)$ such that for all $u \in [0, \infty[$,

$$
\forall x \in \mathbb{R}, \sup_{y \in \mathbb{R}} |g_u(x, y)| du \leq k(x) f(x) \pi(u, x).
$$

In addition these functions are such that there exists two conjugate exponents $p, q \geq 1$ such that $\sup_{u \in \mathbb{R}^+} \int_{\mathbb{R}} \pi^p(u,x) f(x) dx < \infty$ and $\int_{\mathbb{R}} k^q(x) f(x) dx < \infty$. Lastly, we assume that $\int_{\mathbb{R}} \pi^p(u,x) f(x) dx$ is an ultimately decreasing function of $u$ which satisfies for $u_1 > 0$, $\int_{u_1}^{\infty} \left( \int_{\mathbb{R}} \pi^p(u,x) f(x) dx \right)^{1/p} du < \infty$.

If $f$ is bounded, these assumptions imply WCL and more precisely B1 implies WCL1 and B2 implies WCL2. They are fulfilled by the Ornstein–Uhlenbeck process and studied for Markov processes in Section 4.1. Note also that if $\sup_{x \in \mathbb{R}} \pi(u,x) \leq \tilde{\pi}(u)$, then B2 holds under the stronger but simpler condition:
SB2. There exists a positive integrable function \( \tilde{k}(\cdot) \) (defined on \( \mathbb{R} \)) such that for all \( u \in [u_0, \infty[ \),
\[
\forall x \in \mathbb{R}, \quad \sup_{y \in \mathbb{R}} |g_u(x,y)| \leq \tilde{k}(x)\tilde{\pi}(u),
\]
where \( \tilde{\pi}(u) \) is a bounded and ultimately decreasing function which satisfies for \( u_1 > 0 \),
\[
\int_{u_1}^{\infty} \tilde{\pi}(u) \, du < \infty.
\]

Some assumptions closely related to B1 and SB2 (but slightly stronger) are made in Blanke and Pumo [7] where optimal discretization is discussed when kernel density estimators and pointwise quadratic risk are considered. Another assumption than B2 allows to find a similar order; namely,

B3. Assumption WCL2 holds and there exists \( u_0 > 0 \) such that the function \( g_u(x,y) = f_u(x,y) - f(x)f(y) \) is Lipschitz as a function of \( u \), uniformly in \( y \in \mathbb{R} \), that is: there exists a positive integrable function \( \ell(\cdot) \) (defined on \( \mathbb{R} \)) such that for all \( u, v \geq u_0 > 0 \) and \( x \in \mathbb{R} \),
\[
\sup_{y \in \mathbb{R}} |g_u(x,y) - g_v(x,y)| \leq \ell(x)|u - v|. \tag{3.12}
\]

Assumption B3 allows to substitute \( \int_{u_0}^{T} |g_u(x,y)| \, dt \) for its discretized counterpart. As an illustration, if we consider a Gaussian stationary process \( (X_t, t \in \mathbb{R}) \) with autocovariance \( \rho(u) \), then Condition B3 is fulfilled as soon as \( \rho(u) \) is a Lipschitz function for all \( u \geq u_0 \). Again, this holds if \( X \) is an Ornstein–Uhlenbeck process.

A property similar to P2 is also required, still equivalent to (3.8) when an orthonormal basis is considered, which can in this simpler case be written:

A5. There exists a constant \( \Phi_0 \) independent of \( D \) such that \( \forall h \in S \), \( \|h\|_{\infty} \leq \Phi_0 \sqrt{D} \|h\| \).

Under assumptions B1 and B2 or B1 and B3, we have the following result for a suitably discretized continuous time process.

Proposition 8. Let \( S \) be a linear subspace of \( L_2(A) \) with dimension \( D \) and satisfying A4 and A5. Let \( (X_{i\theta_n})_{1 \leq i \leq n} \) be a discrete time sample of the continuous time process \( X = (X_t)_{t \in [0,T_n]} \), with \( n \theta_n = T_n \). Assume that \( X \) belongs to the class \( \mathcal{X} \) such that \( \|f\|_{\mathcal{X}} < \infty \) and consider the estimator \( \hat{f}_S^d \) by (3.1) with \( \text{dim}(S) = D \).

1. If B1 and B2 are fulfilled, then
\[
\mathbb{E}(\|\hat{f}_S^d - f\|^2) \leq \left( \|f - f_S\|^2 + \frac{\Phi_0^2 D}{n} \right) + \frac{K}{T_n}, \tag{3.13}
\]

where \( \Phi_0 \) is defined in A5, \( K \) is a constant depending on \( \|f\|_{\mathcal{X}} \), \( \pi(\cdot, \cdot) \), \( M(\cdot) \), \( k(\cdot) \) and \( C_\phi \) (defined in A4),

2. If B1 and B3 are fulfilled, then
\[
\mathbb{E}(\|\hat{f}_S^d - f\|^2) \leq \left( \|f - f_S\|^2 + \frac{\Phi_0^2 D}{n} \right) + \frac{2C_\phi}{T_n} \int_A k(x) \, dx + C_\phi \delta_n \int_A \ell(x) \, dx.
\] \tag{3.14}
As soon as \( f \) belongs to some Besov space \( \mathcal{B}_{2, \infty}(A) \), the classical theorems of approximation theory previously mentioned lead to the following result: under the assumptions of Proposition 8 and if \( S \) is a regular model \( P \) or \( W \) with dimension \( D \) and with \( r > x > 0 \), we get, for any \( f \in \mathcal{B}_{2, \infty}(L) \) with \( \| f \|_\infty < \infty \),
\[
\mathbb{E}[\| f - f^d_{S} \|^2] \leq C(x, L)D^{-2x} + \frac{\Phi_0 D}{n} + \frac{K}{T_n} + K'' \delta_n, \tag{3.15}
\]
where we recall that \( \mathcal{B}_{2, \infty}(L) = \{ h \in \mathcal{B}_{2, \infty}, |h|_{2, 2} \leq L \} \).

**Remark 2.** The above inequality must be compared with inequality (2.7) in Proposition 2 and inequality (2.8) in the discussion following. From the above result, if we choose \( \delta_n \leq T_n^{-1} \) and if \( x > 1/2 \), it appears that the parametric rate \( T_n^{-1} \) remains valid as soon as we consider a model with dimension \( D = [T_n] \). It is noteworthy to mention that the supremum of \( f_{S} \) is uniformly bounded on the Besov ball \( \mathcal{B}_{2, \infty}(L) \) if \( x > 1/2 \). More generally, it appears from (3.15) that if \( D \) and \( \delta_n^{-1} \) are both great enough then the parametric rate is attained. For instance if we take \( D = [e^{T_n}] \) and \( \delta_n \leq e^{-T_n} \), then we reach the full rate whatever the regularity \( x \) of the Besov space. It follows that it is always possible to make a choice of \( \delta_n \) which does not depend on \( x \). If \( x \) is known, the conditions are \( \delta_n \leq T_n^{-1/(2x)} \) and \( D \geq T_n^{1/(2x)} \).

**Remark 3.** If we add to the Assumptions of Theorem 7, Assumptions B1 and B2, then we obtain an adaptive estimator \( f^d \) associated to a penalty function \( \text{pen}(m) = \kappa \Phi_0 D/m \) and satisfying
\[
\mathbb{E}[\| f^d - f \|^2] \leq \inf_{m \in \mathcal{A}_n} \left( 3\| f - f_{S_m} \|^2 + \frac{K \Phi_0^2 D_m}{n} \right) + \frac{K'}{T_n},
\]
where \( K \) is a numerical constant and \( K' \) is a constant depending on \( \Phi_0, \theta, A_2, A_3 \) and \( \| f \|_{\infty} \). In any case, the rate is improved.

Consider now in addition that the mixing is geometric and \( \delta_n = n^{-a} \) with \( a \epsilon ]0, 1[ \). Then, if \( f \) belongs to some Besov space \( \mathcal{B}_{2, \infty} \) with \( 1/2 < \alpha \), we find as usual
\[
\mathbb{E}[\| f^d - f \|^2] \leq C(f_{2, 2}, \Phi_0)n^{-2a/(2x+1)} + \frac{K'}{T_n}. \tag{3.16}
\]
As a consequence, if \( a \geq 1/2 \), then \( n^{-2a/(2x+1)} = (T_n/\delta_n)^{-2a/(2x+1)} \leq T_n^{-1} \) and the super optimal rate is also reached by the adaptive estimator, automatically and without the a posteriori choice of \( D \). But the procedure is more complicated than the one involved by Proposition 8 associated to Remark 2 and requires more assumptions (namely the geometrical mixing assumption), without any obvious gain. Nevertheless, it illustrates that the super optimal rate can be reached if the step of observations \( \delta_n \) is small enough (condition \( a \geq 1/2 \)). Note that this could not be obtained for arithmetical mixing.
4. Condition WCL: discussions and examples

4.1. Condition WCL in case of Markov processes

In this section, we consider a stationary homogeneous Markov process \( X := (X_t, t \in \mathbb{R}) \). Moreover, we assume that \( X \) belongs to the class \( \mathcal{M} \) defined as follows:

**Definition 9.** We define \( \mathcal{M} \) as the class of real stationary homogeneous Markov processes \( X \) such that \( X \) is ergodic with a unique invariant probability measure \( \Pi \) having a density \( f(x) \) and such that the conditional density \( p_u(x, y) \) of \( X_u \) knowing \( X_0 = x \) exists.

Then obviously such a \( X \) possesses a joint density \( f_u(x, y) = f(x)p_u(x, y) \). We also denote by \( P^u(\cdot, \cdot) \) the probability transition associated to \( p_u(\cdot, \cdot) \).

For Markov processes, checking WCL1 or WCL2 amounts to study, for \( u \geq 0 \),

\[
\text{H1. } \int_{\mathbb{R}} (\sup_{y \in \mathbb{R}} \int_0^u f(x)|p_u(x, y) - f(y)| \, du) \, dx < \infty.
\]

\[
\text{H2. } \int_{\mathbb{R}} (\sup_{y \in \mathbb{R}} \int_u^{\infty} f(x)|p_u(x, y) - f(y)| \, du) \, dx < \infty.
\]

First, as soon as the marginal density is bounded, H1 holds provided that \( \sup_{y \in \mathbb{R}} p_u(x, y) \leq C(x)u^{-\gamma} \), with \( C(\cdot) \) an integrable function on \( \mathbb{R} \) and \( \gamma < 1 \). Examples of Markov diffusion processes satisfying such an assumption may be found for instance in Dynkin [22] or more recently in Leblanc [31], Proposition 11. Note that for these processes, \( \gamma = 1/2 \). Besides, in that case, assumption B1 holds as well.

Let us now turn to Assumptions H2 or B2. Several papers address the question to describe the class of ergodic processes satisfying this assumption in terms of “simple” characteristics (see [31, Theorem 3] or [41]). As a consequence of the proof of Theorem 1 in Veretennikov [41], we would like to give conditions for stationary Markov processes to satisfy H2, in terms of assumptions on the rate of convergence of the absolutely regular coefficients. Denote first \( \phi(\lambda) = \mathbb{E}\exp(i\lambda X_t) \) and \( \phi_x(\lambda, t) = \mathbb{E}\exp(i\lambda X_t|X_0 = x) \).

**Proposition 10.** Assume that \( X := (X_t, t \in \mathbb{R}) \) is a stationary homogeneous Markov process with absolute regular coefficient \( (\beta_u)_{u \in \mathbb{R}^+} \), belonging to the class \( \mathcal{M} \). Moreover assume that there exist constants \( \gamma > 1 \) and \( p > \gamma/(\gamma - 1) \) such that

\[
|\phi(\lambda)| \leq C_1(1 + |\lambda|)^{-\gamma} \text{ where } C_1 > 0, \tag{4.1}
\]

and such that

\[
\int_0^{\infty} \beta_u^{1/p} \, du < \infty. \tag{4.2}
\]

In addition assume that for all \( u \geq u_0 \) and any \( x \in \mathbb{R} \), there exists a nonnegative function \( C_2(\cdot) \) satisfying \( \int_{\mathbb{R}} C_2(x)f(x) \, dx < \infty \) and such that

\[
|\phi_x(\lambda, u)| \leq C_2(x)(1 + |\lambda|)^{-\gamma}. \tag{4.3}
\]
Then we have
\[ \int_{\mathbb{R}} \int_{u_0}^{\infty} \sup_{y \in \mathbb{R}} (p_u(x,y) - f(y)) \, du \, dx \leq C_3 \int_{u_0}^{\infty} \beta_u^{1/p} \, du, \tag{4.4} \]
where \( C_3 \) is a positive constant. It follows that assumption H2 holds.

Note that compared to Condition (H3) in Veretennikov [41], we need not assume \( C_2(\cdot) \) to be bounded. Analogously, we can prove that

**Proposition 11.** Assume that \( X := (X_t, t \in \mathbb{R}) \) is a stationary homogeneous Markov process with absolute regular coefficient \( (\beta_u)_{u \in \mathbb{R}} \), belonging to the class \( \mathcal{M} \) and satisfying the assumptions of Proposition 10. In addition assume that the sequence \( (\beta_u)_{u \in \mathbb{R}^+} \) is ultimately decreasing. Then the process \( X \) satisfies assumption B2.

**Example 1.** Consider a homogeneous Markov diffusion process, defined as a solution of the stochastic differential equation
\[ dX_t = b(X_t) \, dt + \sigma(X_t) \, dW_t, \quad t \geq 0, \tag{4.5} \]
where \((W_t, t \in \mathbb{R})\) is a Wiener process. Conditions for such an \( X_t \) to satisfy (4.1)–(4.3) are given in Veretennikov [40,41]. Note that Leblanc [31] also gives conditions for a diffusion process to be geometrically absolutely regular (see her conditions (6), (7) and (13)).

**Example 2.** Here we enlighten the fact that a diffusion Markov process does not need to satisfy the regularity assumptions given in Veretennikov [40,41] to verify H2; indeed, \( b \) and \( \sigma \) are required to admit \( n \geq 2 \) bounded continuous derivatives. With this aim, note first that the conclusion of Proposition 10 still holds if condition (4.2) is replaced by
\[ \int_{\mathbb{R}} \left( \int_{u_0}^{+\infty} \sup_{\lambda \in \mathbb{R}} |\varphi_\lambda(\lambda, u)| - \varphi(\lambda)|^{1/p} \, du \right) f(x) \, dx < +\infty. \tag{4.6} \]

Consider now the process solution of the following stochastic differential equation
\[ dX_t = (a - bX_t) \, dt + \sigma \sqrt{X_t} \, dW_t, \quad t \geq 0, \tag{4.7} \]
where \((W_t, t \in \mathbb{R})\) is a Wiener process, \( a, b \) and \( \sigma \) are positive numbers. Such a model is called the Cox–Ingersoll–Ross model (see [16]) and is used in financial mathematics. In addition for \( X_0 = x > 0 \), this process is known to remain almost surely positive as soon as \( a \geq \sigma^2/2 \). It is clear that the square root function does not satisfy the regularity condition mentioned above. However, by setting
\[ L(u) = \frac{\sigma^2}{4b}(1 - e^{-bu}) \quad \text{and} \quad \zeta_u(x) = \frac{4xb}{\sigma^2(e^{bu} - 1)}, \]
and by using the fact that
\[ \varphi_\lambda(\lambda, u) = \left( \frac{1}{1 - 2i\lambda L(u)} \right)^{2a/\sigma^2} \quad \text{e}^{\frac{i\lambda L(u)\zeta_u(x)}{1 - 2i\lambda L(u)}} \quad \text{and} \quad \varphi(\lambda) = \left( \frac{1}{1 - 2i\lambda L(+\infty)} \right)^{2a/\sigma^2} \]
(see for instance [30], Proposition 6.2.5 p. 130), we easily derive that conditions (4.1) and (4.3) are satisfied with $a = 2a/s^2$ and $\sup_{x \in \mathbb{R}} C_2(x) < +\infty$. In addition simple calculations lead to the fact that there exists $u_0$ such that for all $u \geq u_0$,

$$\sup_{\lambda \in \mathbb{R}} |\phi_x(\lambda, u) - \phi(\lambda)| \leq Ke^{-bu},$$

where $K$ is a positive constant.

Consequently it follows that the process $\{X_t, t \in \mathbb{R}\}$ solution of (4.7) satisfies Assumption H2 as soon as $a > \sigma^2/2$. Nevertheless, $\sigma(x) = \sigma \sqrt{x}$ does not admit bounded derivatives on $\mathbb{R}$.

4.2. Sharpness of condition WCL

In this section, we enlighten the fact that condition WCL is sharp in the sense that if WCL1 or WCL2 is violated then there are continuous time processes for which our projection estimator does not reach the parametric rate.

4.2.1. About WCL1 and the local behavior of the sample paths

The following example shows that, in some sense, making a local assumption on the irregularity of the sample paths is necessary to obtain the super optimal rate $T^{-1}$.

**Example 3.** Let us give an example of process $(X_t)_{t \in [0,T]}$ belonging to the class $\mathcal{X}$ with $\|f\|_{L^2(A)} < \infty$ and fulfilling WCL2 but not WCL1, and for which the estimator $\hat{f}_S$ of $f$ defined by (2.4) satisfies,

$$\lim_{T \to \infty} T \mathbb{E}[\|f - \hat{f}_S\|^2] = \infty.$$  (4.8)

To this aim, consider $X = \{X_t, t \in \mathbb{R}\}$ a real zero mean stationary Gaussian process which is continuous and differentiable in mean square with variance $\sigma^2 > 0$ and positive autocorrelation $\rho(\cdot)$ on $\mathbb{R}^+$ that is integrable over $[u_0, +\infty[$, $u_0 > 0$ and such that $|\rho(u)| < 1$ for $u > 0$. Then this process satisfies WCL2 but not WCL1. The bound (5.10) can be used to see this. If we consider $A = [0, 1]$, and if we build the estimator $\hat{f}_S$ on the linear space $S$ of $L^2(A)$ with $\dim(S) = D_T \to \infty$ and spanned by the orthonormal basis $(\phi_{j,D_T})_{1 \leq j \leq D_T}$, $\phi_{j,D_T}(x) = \sqrt{D_T^{1/2}} [D_{j-1}(j-1)D_{j}^{-1}]$, then we can prove in Section 5.5 that (4.8) holds. More precisely, we obtain that $\lim \inf_{T \to \infty} (T/\ln T) \mathbb{E}[\|f - \hat{f}_S\|^2] > 0$. This proves the sharpness of condition WCL1. Moreover, this example illustrates the fact that, at least for Gaussian processes, WCL1 is closely linked with the irregularity of the sample paths. Indeed if $X$ has differentiable sample paths then they are differentiable in mean square (see [27]), and it follows that the Gaussian process $X$ does not reach the parametric rate.

4.2.2. About WCL2, the integrability condition near infinity

In this section, we give an example showing that if WCL1 holds but not WCL2, then the parametric rate is not necessarily reached. For that purpose, we give a sufficient condition on the joint density of a stationary continuous time process $X$ belonging to the class $\mathcal{X}$ for $X$ to satisfy WCL2. Following Giraitis et al. [26], we consider the following decomposition of the joint density $f_u$ of $(X_0, X_u)$. There exists
$u_0 > 0$ such that for all $u \geq u_0$,

$$f_u(x, y) = f(x)f(y) + \rho(u) f'(x)f'(y) + h_u(x, y), \ x, y \in \mathbb{R}, \quad (4.9)$$

where $\rho(u)$ is the autocovariance of the process $X$ and $h_u(\cdot, \cdot)$ is such that:

$$\forall u \geq u_0, \ |h_u(x, y)| \leq |\rho(u)|k_1(x)k_2(y), \ x, y \in \mathbb{R}, \text{ for some } \varepsilon > 1, \quad (4.10)$$

where $k_1(\cdot)$ and $k_2(\cdot)$ are positive functions defined on $\mathbb{R}$.

**Remark 4.** When $f_u(\cdot, \cdot)$ is a bivariate normal density, $h_u(\cdot, \cdot)$ is the remainder in the Taylor’s expansion and in this case, easy computations lead to the fact that (4.10) is satisfied for $\varepsilon = 2$ and $k_1(x) = k_2(x) = K(x) \exp(-x^2/4)$ where $K(x)$ is a certain univariate polynomial function.

It is clear that if $X$ is a continuous time process belonging to the class $\mathcal{X}$ and satisfying conditions (4.9) and (4.10) with

$$\sup_{y \in A} |f'(y)| < +\infty, \quad \sup_{y \in A} k_2(y) < +\infty \quad \text{and} \quad \int_A k_1(x) \, dx < +\infty,$$

then $WCL2$ holds as soon as

$$\int_{u_0}^{\infty} |\rho(u)| < +\infty. \quad (4.11)$$

Conversely, this consideration together with Remark 4 leads to the conclusion that if $X$ is a Gaussian process such that (4.11) is not satisfied but $\rho(u)$ is square-integrable, then $WCL2$ fails to hold. Since cases where (4.11) fails to hold correspond to the situation of long-range dependence processes, this leads to the following example of a continuous time process that does not satisfy $WCL2$ but that verifies the local integrability condition $WCL1$.

**Example 4.** Some fractional integrals of Gaussian processes have been considered in the literature. The fractional integral of order $\alpha$, $0 < \alpha < 1/2$ of e.g. a stationary Ornstein–Uhlenbeck process has two characteristics (see for instance [15]):

- a fractional index of local regularity namely $\alpha + 1/2$, which is illustrated by the following development, for $u$ in the neighbourhood of 0: $\rho(u) = 1 + cu^{2\alpha + 1} + o(u^{2\alpha + 2})$, where $c$ is a constant.
- a long memory property characterized by the following equivalent for $u$ near $+\infty$: $\rho(u) \sim c' u^{2\alpha - 1}$, where $c'$ is a positive constant.

Using the behaviour of $\rho(u)$ in a neighbourhood of 0, we get that for $u \in [0, u_0]$,  

$$|g_u(x, y)| \leq \tilde{c}(1 + u^{-\frac{\alpha - 1}{2}}), \ x, y \in \mathbb{R}, \text{ where } \tilde{c} \text{ is a positive constant,} \quad (4.12)$$

which entails that $WCL1$ holds since $0 < \alpha < 1/2$.

On the other hand the long memory property of this Gaussian process implies that for $0 < \alpha < 1/4$, $\rho(u)$ is square integrable but not integrable near infinity, and therefore $WCL2$ is not satisfied as soon as $0 < \alpha < 1/4$. When $1/4 < \alpha < 1/2$, the
Taylor expansion as given in (4.9) does not allow any conclusion. Thus for such a process with 0 < z < 1/4, the parametric rate is not reached since we can prove that
\[
\liminf_{T \to \infty} T^{1-2z} \mathbb{E}[f - \hat{f}_S]^2 > 0,
\]
where the estimator \( \hat{f}_S \) of \( f \) is defined by (2.4) on a linear subspace \( S \) of \( \mathbb{L}_2(A) \) with dimension \( D_T \) such that \( \lim_{T \to \infty} D_T = +\infty \) and spanned by an orthonormal basis \( (\phi_j, D_T)_{1 \leq j \leq D_T} \) defined by \( \phi_j, D_T(x) = \sqrt{D_T} \mathbb{E}_{[D_T^{j-1}, D_T]}(x) \) (see Section 5).

5. Proofs

5.1. Proof of Theorem 7

We first show the following auxiliary lemma which is a triangular version of Viennet’s inequality [42].

**Lemma 12.** Let \( (X_{tj})_{1 \leq j \leq n} \) be a discrete time sample of the continuous time process \( X = (X_t)_{t \in \mathbb{R}} \) assumed to be weakly stationary and with absolute regular mixing coefficients \( (\beta_t^i)_{t \geq 0} \). Denote by \( P \) the distribution of \( X_0 \) and by \( \mathbb{E}_p(\psi) = \int_{\mathbb{R}} \psi(x) \, dP(x) \). There exists a sequence of measurable functions \( (b_{k, \delta_n})_{k \geq 0} \) with \( b_{0, \delta_n} = 1, 0 \leq b_{k, \delta_n} \leq 1, \mathbb{E}_p(b_{k, \delta_n}) \leq \beta_{k, \delta_n}^n \) such that for any \( h \in \mathbb{L}_2(P) \) and any positive integer \( n \),

\[
\text{Var} \left( \sum_{i=1}^n h(X_{ih_n}) \right) \leq 4n \int_{\mathbb{R}} \left( \sum_{k=0}^n b_{k, \delta_n} \right) h^2 \, dP.
\]

Moreover for \( 1 \leq p < \infty \), \( \mathbb{E}_p(\sum_{k=0}^\infty b_{k, \delta_n})^p \leq p \sum_{l \geq 0} (l + 1)^{p-1} \beta_{l, \delta_n}^* = pB_{p+1}(\delta_n) \) provided this last series is convergent.

**Proof of Lemma 12.** We first write that by weak stationarity, we have

\[
\text{Var} \left( \sum_{i=1}^n h(X_{ih_n}) \right) \leq 2 \sum_{k=0}^n (n - k) |\text{cov}(h(X_0), h(X_{ih_n}))|.
\]

Next following the proof of Theorem 2.1 of Viennet [42], we infer that there exist two function \( b_{k, \delta_n}' \) and \( b_{k, \delta_n}'' \) from \( \mathbb{R} \) into \([0, 1] \) such that \( \mathbb{E}_p(b_{k, \delta_n}') = \mathbb{E}_p(b_{k, \delta_n}'') \leq \beta_{k, \delta_n}'' \) and that

\[
\text{cov}(h(X_0), h(X_{k, \delta_n})) \leq 2 \mathbb{E}_p\left( b_{k, \delta_n}' h^2 \right) \mathbb{E}_p\left( b_{k, \delta_n}'' h^2 \right).
\]

Thus

\[
\text{Var} \left( \sum_{i=1}^n h(X_{ih_n}) \right) \leq 4n \sum_{k=0}^n \mathbb{E}_p \left( \frac{1}{2} (b_{k, \delta_n}' + b_{k, \delta_n}'') \right) h^2.
\]

\(^1\)Here we consider the absolute regular coefficients \( (\beta_t^i)_{t \geq 0} \) (and not \( (\beta_t)_{t \geq 0} \)) because we need them to be nonincreasing in order to make use of \( \beta_{\delta_n}^i(u) \) in the proof.
Theorem 3.1 (discrete time case) in Comte and Merlevède [14] with some decompositions remains valid when the process is only supposed to be weakly stationary instead of modifications due to Lemma 12. We also notice that the above-mentioned proof the initial sequence by no more than \( \beta_{\delta_n} \). We end the proof by noticing that for any conjugate exponents \( p \) and \( q \), we have
\[
\int b_{\delta_n} \psi \, dP \leq \int_0^1 \beta_{\delta_n}^{-1}(u)Q_\psi(u) \, du.
\]

Using this representation together with (5.2), we derive
\[
\left( \int b_{\delta_n}^p \, dP \right)^{1/p} \leq \sup_{\|\psi\|=1} \left( \int_0^1 Q_\psi^p(u) \, du \right)^{1/q} \int_0^1 \beta_{\delta_n}^{-1}(u)Q_\psi(u) \, du.
\]

Next Holder’s inequality leads to
\[
\int b_{\delta_n}^p \, dP \leq \int_0^1 (\beta_{\delta_n}^{-1})^p(u) \, du.
\]

We end the proof by noticing that \( \int_0^1 (\beta_{\delta_n}^{-1})^p(u) \, du \leq \sum_{i=0}^{l+1} \beta_{\delta_n}^{*} \) (see for instance the bound pages 15–16 in [36]).

We turn now to the proof of Theorem 7. We proceed similarly to the proof of Theorem 3.1 (discrete time case) in Comte and Merlevède [14] with some modifications due to Lemma 12. We also notice that the above-mentioned proof remains valid when the process is only supposed to be weakly stationary instead of being strictly stationary. Let us now describe the methodology. We consider the decompositions
\[
\|f - f^d\|^2 \leq \|f - f_m\|^2 + 2 \nu_n(f^d - f_m) + \text{pen}_d(m) - \text{pen}_d(m'),
\]
and
\[
2|\nu_n^d(f^d - f_m)| \leq 2|\nu_n^d(f - f_m)| - \nu_n^d(f - f_m) + 2|\nu_n^d(f - f_m)|.
\]

where \( \nu_n^d \) denotes the empirical contrast computed on the \( X_{i\delta_n}^* \), where the \( X_{i\delta_n}^* \) are distributed as \( X_{i\delta_n} \) and constructed using Berbee’s lemma [3]. In fact they are such that blocks far from a certain distance, say \( q_n \), are independent and such that the blocks obtained from the \( X_{i\delta_n}^* \) differ in probability from the ones constructed with the initial sequence by no more than \( \beta_{q_n} \). Moreover if we denote by \( B_{m,m'}(0,1) \) the unit ball of the linear space \( S_m + S_{m'} \), then for any function \( p(m, m') \) of \( m \) and \( m' \), we
also consider the following decomposition:

$$
2|v_n^d(f - f_m)| \leq \frac{1}{4} \|f_m - f\|^2 + \frac{1}{4} \|f - \tilde{f}\|^2 + 8 \sum_{m' \in \mathcal{M}_n} W^{d*}(m') + 8p(m, \hat{m}),
$$

where $W^{d*}(m') := \left[\left(\sup_{h \in B_{n,m'}(0,1)} |v_n^d(h)|\right)^2 - p(m, m')\right]_+$. The aim of the proof is then to find $p(m, m')$ such that

$$
\sum_{m' \in \mathcal{M}_n} \mathbb{E}(W^{d*}(m')) \leq CT_n^{-1}.
$$

(5.3)

This is done using concentration inequalities from Talagrand [38].

Let us now indicate the changes compared to the proof given in Comte and Merlevède [14] for $\delta_n = 1$. Eq. (6.16) is now replaced for $N_n \ll n$, by

$$
n \beta_n^* \leq \frac{C}{T_n} = \frac{C}{n \delta_n}
$$

(5.4)

and the function $p(m, m')$ is now taken as

$$
p(m, m') = 8(4 + \xi^2)\phi_0^2 B_2(\delta_n) \frac{D_m + D_{m'}}{n},
$$

where $\xi$ is positive. Using Lemma 12 in the proof given in Comte and Merlevède [14], this choice leads to

$$
\mathbb{E}(W^{d*}(m')) \leq C_0 \left[\frac{\sqrt{B_3(\delta_n)D}}{n} \exp(-C_1 \xi^2 \sqrt{D}) + \frac{q_n^2}{n} \exp\left(-\frac{K_1 \xi}{2} \sqrt{nB_2(\delta_n)}/q_n\right)\right],
$$

where $C_0 = C_0(K_1, \phi_0, \|f\|_\infty^1 A)$ with $K_1$ a universal constant, and

$$
\frac{K_1 \phi_0 A_2}{4\sqrt{2\|f\|_\infty^1 (1 + 2A_2 + A_3)}} \leq C_1 = \frac{K_1 \phi_0 B_2(\delta_n)}{4\sqrt{2\|f\|_\infty^1 B_3(\delta_n)}} \leq \frac{K_1 \phi_0 (1 + A_2)}{4\sqrt{2\|f\|_\infty^1 A_3}}
$$

since from relation (3.9), we have: $\delta_n^{-1} A_2 \leq B_2(\delta_n) \leq 1 + \delta_n^{-1} A_2$ and $\delta_n^{-2} A_3 \leq B_3(\delta_n) \leq 1 + 2\delta_n^{-1} A_2 + \delta_n^{-2} A_3$. It follows that

$$
\mathbb{E}(W^{d*}(m')) \leq C \left[\frac{\sqrt{D}}{T_n} \exp(-C_1 \xi^2 \sqrt{D}) + \frac{q_n^2}{n} \exp\left(-\frac{K_1 A_2 \xi}{2} \sqrt{n\delta_n^{-1}}/q_n\right)\right],
$$

where $C = C(K_1, \phi_0, \|f\|_\infty^1, A_2, A_3)$. Consequently the term $\sum_{m' \in \mathcal{M}_n} \mathbb{E}(W^{d*}(m'))$ is less than $CT_n^{-1}$ as soon as assumption P3 is fulfilled, $q_n = [n^{1/2}]$ and $\delta_n = n^{-a}$ with $a > 0$. Replacing in (5.4) $q_n$ by $[n^{1/2}]$ and using that $\beta_i^* \leq 1/(1 + t)^{1+\theta}$ leads then to the constraint $(1 + \theta)(1/2 - a) > 2 - a$ and therefore to (3.10). If $\delta_n^{-1} \geq \ln^2(n)$, then it suffices to choose $q_n = [n^{1/2}]$ and $\xi = 2/(K_1 A_2)$ to get (5.3). On the other hand (5.4) leads to the constraint $\theta > 3$. We end the proof by taking into account that $\delta_n^{-1} A_2 \leq B_2(\delta_n) \leq 1 + \delta_n^{-1} A_2 \leq (1 + A_2)\delta_n^{-1}$. □
5.2. Proof of Proposition 8

Case 1. First we consider the decomposition

\[
\mathbb{E}(\|f_S^n - f_S\|^2) = \frac{1}{n} \sum_{j=1}^{D} \text{Var} \left( \frac{1}{n} \sum_{k=1}^{n} \phi_{j,D}(X_{k\delta_n}) \right)
\]

\[
= \frac{1}{n} \sum_{j=1}^{D} \text{Var}(\phi_{j,D}(X_0)) + \frac{2}{n} \sum_{j=1}^{D} \sum_{k=1}^{n} \int_A \int_A \phi_{j,D}(x)\phi_{j,D}(y) \left( 1 - \frac{k}{n} \right) g_{k\delta_n}(x,y) \, dx \, dy
\]

\[
= T_{1,n} + 2T_{2,n}.
\]  

(5.5)

Using (3.8) to bound \( T_{1,n} \), we find: \( T_{1,n} \leq \frac{1}{n} \int_A (\sum_{j=1}^{D} \phi_{j,D}(x))^2 f(x) \, dx \leq \Phi_0^2 D/n \). To study the last term in the right-hand side of inequality (5.5), we write the decomposition: \( T_{2,n} = T_{2}^{(1)} + T_{2}^{(2)} + T_{2}^{(3)} \), where

\[
T_{2,n}^{(1)} = \frac{1}{n} \sum_{j=1}^{D} \sum_{k=m_n^{(1)}+1}^{m_n^{(1)}} \int_A \int_A \phi_{j,D}(x)\phi_{j,D}(y) \left( 1 - \frac{k}{n} \right) g_{k\delta_n}(x,y) \, dx \, dy,
\]

where \( m_n^{(0)} = 0 \), \( m_n^{(1)} = [\delta_n^{-1} u_0] \), \( m_n^{(2)} = [\delta_n^{-1} u_1] \) with an arbitrarily large \( u_1 \) and \( m_n^{(3)} = n \). Using first assumptions B2 and A4 combined with Hölder’s inequality, we derive

\[
|T_{2,n}^{(2)}| \leq \frac{C}{n} \sum_{\ell=m_n^{(1)}+1}^{m_n^{(3)}} \left( \int_A k^{q}(x) f(x) \, dx \right)^{1/q} \left( \int_A \pi^{p}(\ell\delta_n, x) f(x) \, dx \right)^{1/p}.
\]  

(5.6)

Next using again assumption B2, we get that there exists a positive finite constant \( K_1 \) such that

\[
|T_{2,n}^{(2)}| \leq \frac{C}{n \delta_n} (u_1 - u_0) \left( \int_A k^{q}(x) f(x) \, dx \right)^{1/q} \left( \sup_{[u_0, u_1]} \int_A \pi^{p}(u, x) f(x) \, dx \right)^{1/p} \leq \frac{K_1}{n \delta_n}.
\]

Similar arguments leading to the bound on (5.6) yield that \( T_{2,n}^{(3)} \) may be bounded as follows \( |T_{2,n}^{(3)}| \leq \frac{C}{n \sum_{\ell=m_n^{(2)}+1}^{m_n^{(3)}}} \left( \int_A k^{q}(x) f(x) \, dx \right)^{1/q} \left( \int_A \pi^{p}(\ell\delta_n, x) f(x) \, dx \right)^{1/p} \). Consequently since \( \int_{\mathbb{R}} \pi^{p}(u, x) f(x) \, dx \) is assumed to be an ultimately decreasing function of \( u \), we get

\[
|T_{2,n}^{(3)}| \leq \frac{C}{n \delta_n} \left( \int_A k^{q}(x) f(x) \, dx \right)^{1/q} \int_{u_1}^{+\infty} \left( \int_A \pi^{p}(u, x) f(x) \, dx \right)^{1/p} \, du.
\]
and then under assumption B2, there exists a positive finite constant $K_2$ such that $|T_{2,n}^{(3)}| \leq K_2 (n\delta_n)^{-1}$.

Thus it remains to treat $T_{2,n}^{(1)}$. To this aim we first write

\[
T_{2,n}^{(1)} = \frac{1}{n} \sum_{j=1}^{D} \sum_{\ell=1}^{m_n^{(1)}} \int_A \int_A \varphi_{j,D}(x) \varphi_{j,D}(y) \left( 1 - \frac{\ell}{n} \right) f_{\ell\delta_n}(x,y) \, dx \, dy
- \frac{1}{n} \sum_{j=1}^{D} \sum_{\ell=1}^{m_n^{(1)}} \int_A \int_A \varphi_{j,D}(x) \varphi_{j,D}(y) \left( 1 - \frac{\ell}{n} \right) f(x)f(y) \, dx \, dy := I + II. \tag{5.7}
\]

By using A4 and with the choice of $m_n^{(1)}$, we easily derive $|II| \leq u_0 (C_\varphi \|f_{\ell \delta_n}\|_\infty)/(n\delta_n)$.

Concerning the first term of decomposition (5.7), Assumption B1 together with A4 yield

\[
|I| \leq \frac{1}{n\delta_n^2} \left( \int_A M(x) \, dx \right) \sum_{\ell=1}^{m_n^{(1)}} \frac{1}{\ell^\gamma} \leq \frac{1}{n\delta_n^2} \left( \int_A M(x) \, dx \right) \left( 1 + \int_1^{m_n^{(1)}} u^{-\gamma} \, du \right)
\leq \frac{u_0^{1-\gamma}}{(1-\gamma)n\delta_n} \left( C_\varphi \int_A M(x) \, dx \right).
\]

Gathering all the bounds gives the result in Case 1. \qed

Case 2: We consider decomposition (5.5) again. $T_{1,n}$ is treated as previously whereas for $T_{2,n}$, we set $m_n^{(1)} = [\delta_n^{-1} u_0]$ and consider this time the decomposition: $T_{2,n} = I_{2,n} + J_{2,n}$, where

\[
I_{2,n} = \frac{1}{n} \sum_{j=1}^{D} \sum_{k=1}^{m_n^{(1)}} \int_A \int_A \varphi_{j,D}(x) \varphi_{j,D}(y) \left( 1 - \frac{k}{n} \right) g_{k\delta_n}(x,y) \, dx \, dy,
\]

\[
J_{2,n} = \frac{1}{n} \sum_{j=1}^{D} \sum_{k=m_n^{(1)}+1}^{n} \int_A \int_A \varphi_{j,D}(x) \varphi_{j,D}(y) \left( 1 - \frac{k}{n} \right) g_{k\delta_n}(x,y) \, dx \, dy.
\]

The term $I_{2,n}$ has already been treated in the proof of Proposition 8, Case 1, and found to be of order $(n\delta_n)^{-1}$ under assumptions B1, A4 and $\|f\|_\infty < \infty$. Concerning $J_{2,n}$, we first notice that

\[
|J_{2,n}| \leq \frac{1}{T_n} \int_A \int_A \left| \sum_{j=1}^{D} \varphi_{j,D}(x) \varphi_{j,D}(y) \right| \left( \delta_n - \sum_{k=m_n^{(1)}+1}^{n} |g_{k\delta_n}(x,y)| \right) \, dx \, dy
\leq \frac{1}{T_n} \int_A \int_A \left| \sum_{j=1}^{D} \varphi_{j,D}(x) \varphi_{j,D}(y) \right| \left( \int_{u_0}^{T_n} |g_s(x,y)| \, ds \right) \, dx \, dy
+ \frac{1}{T_n} \int_A \int_A \left| \sum_{j=1}^{D} \varphi_{j,D}(x) \varphi_{j,D}(y) \right| \left( \int_{u_0}^{T_n} |g_s(x,y) - g_{\delta_n \delta_n}(x,y)| \, ds \right) \, dx \, dy
:= I_n + II_n,
\]
by using $\delta_n \sum_{k=m_n^{(1)} + 1}^{n-1} |g_{k \delta_n}(x, y)| = \int_{u_0}^{T_n} |g_{\delta_n[y/\delta_n]}(x, y)| \, ds$. The first right-hand side term $I_n$ corresponds to the continuous time one and has already been proved to be of the order $C_{\phi} T_n^{-1} \int_A k(x) \, dx$ under WCL2 and A4. To study the second right-hand-side term, we use B3 which yields that for all $x \in \mathbb{R}$,

$$
\sup_{y \in \mathbb{R}} \int_{u_0}^{T_n} |g_s(x, y) - g_{\delta_n[s/\delta_n]}(x, y)| \, ds \\
\leq \ell(x) \int_{u_0}^{T_n} |s - \delta_n[s/\delta_n]| \, ds \\
\leq \ell(x) \sum_{k=m_n^{(1)} + 1}^{n-1} \int_{k \delta_n}^{(k+1) \delta_n} (s - k \delta_n) \, ds \leq \ell(x) \sum_{k=m_n^{(1)} + 1}^{n-1} \int_0^{\delta_n} x \, dx \leq \ell(x) \frac{T_n \delta_n}{2},
$$

which in turn together with A4 leads to $II_n \leq (C_{\phi}/2) \delta_n \int_A \ell(x) \, dx$. Gathering all the bounds gives the result. □

5.3. Proof of Proposition 10

Using the inverse Fourier transform, we have

$$
2\pi(p_u(x, y) - f(y)) = \int_{\mathbb{R}} \exp(-i \lambda y)(\varphi_x(\lambda, u) - \varphi(\lambda)) \, d\lambda.
$$

Then using (4.1) and (4.3), we get for $u \geq u_0$

$$
2\pi |p_u(x, y) - f(y)| \\
\leq 2(C_1 + C_2(x))^{(p-1)/p} \left( \sup_{\lambda \in \mathbb{R}} |\varphi_x(\lambda, u) - \varphi(\lambda)| \right)^{1/p} \int_{\mathbb{R}^+} (1 + \lambda)^{-\gamma(p-1)/p} \, d\lambda,
$$

and since it is assumed that $\gamma(p-1)/p > 1$, we get that there exists a finite constant $K$ such that for all $u > u_0$,

$$
\sup_{y \in \mathbb{R}} (f(x)|p_u(x, y) - f(y)|) \\
\leq K(C_1 + C_2(x))^{(p-1)/p} f(x) \left( \sup_{\lambda \in \mathbb{R}} |\varphi_x(\lambda, u) - \varphi(\lambda)| \right)^{1/p} . \tag{5.8}
$$

By using the fact that $\sup_{\lambda \in \mathbb{R}} |\varphi_x(\lambda, u) - \varphi(\lambda)| \leq \|P(u, \cdot) - II\|_v$, (where $\| \cdot \|_v$ denotes the variation norm), together with (5.8) and Fubini's Theorem, we get

$$
\int_{\mathbb{R}} \int_{u_0}^{\infty} \sup_{y \in \mathbb{R}} (f(x)|p_u(x, y) - f(y)|) \, du \, dx \\
\leq K \int_{u_0}^{\infty} \left( \int_{\mathbb{R}} (C_1 + C_2(x))^{(p-1)/p} \|P(u, \cdot) - II\|_v^{1/p} \, dx \right) \, du.
This combined with H"older’s inequality with respect to the measure $\Pi$, gives

$$\int_{\mathbb{R}} \int_{u_0}^{\infty} \text{sup}(f(x)|p_u(x,y) - f(y)|) \, du \, dx \leq K \int_{u_0}^{\infty} \left( \int_{\mathbb{R}} \|P^u(x,.) - \Pi\|_{\Pi}(dx) \right)^{1/p} \left( \int_{\mathbb{R}} (C_1 + C_2(x)) f(x) \, dx \right)^{(p-1)/p} \, du.$$  

According to Davydov [17], the absolute regular coefficient for Markov chains can also be defined as $\beta_u = \int_{\mathbb{R}} \|P^u(x,.) - \Pi\|_{\pi}(dx)$. Combining this consideration with the fact that $\int_{\mathbb{R}} C_2(x)f(x)(dx) < \infty$, we derive that there exists a finite constant $C_3$ such that

$$\int_{\mathbb{R}} \int_{u_0}^{\infty} \text{sup}(f(x)|p_u(x,y) - f(y)|) \, du \, dx \leq C_3 \int_{u_0}^{\infty} \beta_u^{1/p} \, du. \quad \square$$

5.4. Proof of Proposition 11

According to (5.8) and the inequality after, under (4.1) and (4.3), there exists a finite constant $K$ such that for all $u > u_0$ and all $x \in \mathbb{R}$,

$$\text{sup}_{y \in \mathbb{R}} |g_u(x,y)| \leq K (C_1 + C_2(x))^{(p-1)/p} f(x)(\|P^u(x,.) - \Pi\|_{\pi})^{1/p}. \quad (5.9)$$

Then take $\pi(u, x) := (\|P^u(x,.) - \Pi\|_{\pi})^{1/p}$ and $k(x) := K(C_1 + C_2(x))^{(p-1)/p}$. First it is clear from the condition on $C_2(.)$ that $\int (k(x))^{\rho/(p-1)} f(x) \, dx$ is finite. Now we notice that since $\beta_u = \int_{\mathbb{R}} \pi^u(u, x)f(x) \, dx$ (see [17]), then we both get that $\sup_{u \in \mathbb{R}^+} \int_{\mathbb{R}} \pi^u(u, x)f(x) \, dx$ and $\int_{u_0}^{\infty} (\int_{\mathbb{R}} \pi^u(u, x)f(x) \, dx)^{1/p} \, du$ are finite. The first assertion comes from the fact that the absolutely regular coefficient is uniformly bounded by 1 and the second one from Condition (4.2). Since $(\beta_u)_{u \in \mathbb{R}}$ is assumed to be ultimately decreasing so is the function $\int_{\mathbb{R}} \pi^u(u, x)f(x) \, dx$ in $u$. This achieves the proof. \( \square \)

5.5. Proof of (4.8)

Since $X$ is mean square differentiable then $1 - \rho(u) \sim cu^2$ as $u \to 0$. So there exist constants $c_1$ and $c_2$ such that in a neighborhood of zero, $0 < 1 - c_1 u^2 \leq \rho(u) \leq 1 - c_2 u^2$. In addition, we have for $u \in ]0, u_0[$,

$$f_u(x,y) \geq \frac{u^{-1}}{2\pi\sigma^2\sqrt{2c_1}} \exp\left(-\frac{x^2 + 1}{2\sigma^2(1 + \rho(u))} - \frac{(x-y)^2}{2\sigma^2(1 - \rho^2(u))}\right). \quad (5.10)$$

Then for all $j \in \{1, \ldots, D_T\}$, all $x, y \in [D_T^{-1}(j-1), D_T^{-1}j]$ and all $u \in ]0, u_0[$,

$f_u(x,y) \geq \Psi_u(x,D_T)/u$ with

$$\Psi_u(x,D_T) = \frac{1}{2\pi\sigma^2\sqrt{2c_1}} \exp\left(-\frac{x^2 + 1}{2\sigma^2(1 + \rho(u))} - \rho(u) \frac{D_T^{-2}}{2\sigma^2(1 - \rho^2(u))}\right).$$
Then
\[
\liminf_{T \to \infty} \inf_{u \in [D^{-1}_T, u_0]} \Psi_u(x, D_T) \geq \Psi(x),
\]
(5.11)
where \( \Psi(x) = \frac{1}{2\pi\sigma^2 \sqrt{2c_1}} \exp \left( -\frac{x^2 + 1}{2\sigma^2 (2 - c_1 u_0^2)} - \frac{1}{2\sigma^2 c_2} \right) \), for \( u_0^2 < 2/c_1 \).

Due to (2.9), it suffices to show (4.8) to prove that
\[
\lim_{T \to \infty} T \mathbb{E}(v_T(\hat{f}_S - f_S)) = +\infty.
\]
(5.12)
Using (2.10), we write for \( T \geq 2u_0 \) that \( T \mathbb{E}(v_T(\hat{f}_S - f_S)) = I_T + J_T \), where
\[
I_T = 2 \sum_{j=1}^{D_T} \int_A \int_A \varphi_{j,D_T}(x) \varphi_{j,D_T}(y) \left( \int_{u_0}^{u_T} (1 - \frac{u}{T}) g_u(x, y) \, du \right) \, dx \, dy,
\]
(5.13)
\[ J_T = 2 \sum_{j=1}^{D_T} \int_A \int_A \varphi_{j,D_T}(x) \varphi_{j,D_T}(y) \left( \int_{u_0}^{u_T} (1 - \frac{u}{T}) g_u(x, y) \, du \right) \, dx \, dy. \]

First using condition WCL2 and the definition of the \( \varphi_{j,D_T} \), we easily obtain the bound \( |I_T| \leq 2 \int_A k(x) \, dx \). On the other hand, to treat \( J_T \), we write: \( J_T = J_T^{(1)} + J_T^{(2)} \), with
\[
J_T^{(1)} = -2 \sum_{j=1}^{D_T} \int_A \int_A \varphi_{j,D_T}(x) \varphi_{j,D_T}(y) \left( \int_{u_0}^{u_T} \left( 1 - \frac{u}{T} \right) \, du \right) f(x)f(y) \, dx \, dy \quad \text{and}
\]
\[
J_T^{(2)} = 2 \sum_{j=1}^{D_T} \int_A \int_A \varphi_{j,D_T}(x) \varphi_{j,D_T}(y) \left( \int_{u_0}^{u_T} \left( 1 - \frac{u}{T} \right) f_u(x, y) \, du \right) \, dx \, dy.
\]

For the first term, we use \( ||f||_A || \leq +\infty \) and the definition of the \( \varphi_{j,D_T} \) to derive
\[
|J_T^{(1)}| \leq 2u_0 \sum_{j=1}^{D_T} \int_A \int_A |\varphi_{j,D_T}(x)\varphi_{j,D_T}(y)| f(x)f(y) \, dx \, dy \leq 2u_0 ||f||_A ||\.
\]
Consequently, gathering all the bounds, to prove (5.12) it remains to show that \( \lim_{T \to \infty} J_T^{(2)} = +\infty \). Now \( T \geq 2u_0 \) implies \( 2(1 - u_0 T^{-1}) \geq 1 \). Then by using the definition of the \( \varphi_{j,D_T} \), we derive
\[
J_T^{(2)} \geq D_T \sum_{j=1}^{D_T} \int_{(j-1)/D_T}^{j/D_T} \int_{(j-1)/D_T}^{j/D_T} \left( \int_{D_T^{-1} u}^{u_0} f_u(x, y) \, du \right) \, dx \, dy.
\]
This last inequality together with the lower bound for \( f_u \) entails
\[
J_T^{(2)} \geq \int_A \int_{D_T^{-1} u}^{u_0} \Psi_u(x, D_T) \, dx \, du \geq \left( \int_{D_T^{-1} u}^{u_0} \frac{1}{u} \, du \right) \int_A \inf_{u \in [D_T^{-1} u, u_0]} \Psi_u(x, D_T) \, dx.
\]
We get
\[ \frac{1}{\ln D_T} J^{(2)}_T \geq \frac{\ln u_0 - \ln D_T^{-1}}{\ln D_T} \int_A \inf_{u \in [D_T^{-1}, u_0]} \Psi_u(x, D_T) \, dx. \]

Next Fatou’s Lemma and Inequality (5.11) yield that
\[ \liminf_{T \to \infty} \left( \frac{1}{\ln D_T} J^{(2)}_T \right) \geq \int_A \Psi(x) \, dx > 0. \]

Therefore \( \lim_{T \to \infty} J^{(2)}_T = +\infty \). This ends the proof of (5.12) and of (4.8).

5.6. Proof of (4.13)

As in the proof of (4.8), we use decomposition (5.13). Using decomposition (4.12) together with the behaviour of \( \rho(u) \) in a neighbourhood of 0, we easily derive that \( \lim_{T \to \infty} |J_T| < +\infty \). By using the Taylor’s expansion (4.9), we get
\[
I_T = 2 \sum_{j=1}^{D_T} \int_A \int_A \varphi_{j,D_T}(x) \varphi_{j,D_T}(y) \left( \int_{u_0}^{T} \left( 1 - \frac{u}{T} \right) \rho(u) \, du \right) f'(x) f'(y) \, dx \, dy + 2 \sum_{j=1}^{D_T} \int_A \int_A \varphi_{j,D_T}(x) \varphi_{j,D_T}(y) \left( \int_{u_0}^{T} \left( 1 - \frac{u}{T} \right) h_u(x, y) \, du \right) \, dx \, dy = I^{(1)}_T + I^{(2)}_T.
\]

By using Remark 4, the long memory property, the fact that \( 0 < \alpha < 1/4 \) and the definition of the \( \varphi_{j,D_T} \), we derive that \( \lim_{T \to \infty} |I^{(2)}_T| < +\infty \). Next by using the definition of the \( \varphi_{j,D_T} \), we clearly have that
\[
2 \sum_{j=1}^{D_T} \int_A \int_A \varphi_{j,D_T}(x) \varphi_{j,D_T}(y) f'(x) f'(y) \, dx \, dy = K > 0.
\]

Then Fubini’s Theorem entails that \( I^{(1)}_T = K \int_{u_0}^{T} (1 - u/T) \rho(u) \, du \). By using the long memory property, we derive that \( I^{(1)}_T \sim \{Kc'(2\alpha)(2\alpha + 1)^{-1}\}T^{2\alpha} \). Consequently \( \lim_{T \to \infty} T^{-2\alpha} I^{(1)}_T > 0 \). Gathering all the bounds ends the proof of (4.13).}

References


