Recursive Estimation for Stochastic Damping Hamiltonian Systems

Patrick Cattiaux, José R. Leon, Clémentine Prieur

RECURSIVE ESTIMATION FOR STOCHASTIC DAMPING HAMILTONIAN SYSTEMS.

PATRICK CATTIAUX ♠ , **JOSÉ R. LEÓN** ♣ ,
AND **CLÉMENTINE PRIEUR** ♢

♠ Institut de Mathématiques de Toulouse
♣ Escuela de Matemática UCV, Caracas
♢ Laboratoire Jean Kuntzmann, Grenoble

Abstract. In this paper, we complete our previous works (Cattiaux, Leon and Prieur (2014-a), Cattiaux, Leon and Prieur (2014-b), Cattiaux, Leon and Prieur (2014-c)) on the (non-parametric) estimation of the characteristics (invariant density, drift term, variance term) of some ergodic hamiltonian systems, under partial observations. More precisely, we introduce recursive estimators using the full strength of the ergodic behavior. We compare the theoretical results obtained with these estimators to the results we proved for the ones we have introduced previously.

MSC 2010 : 62G07, 60H10, 60F05.

Keywords : hypoelliptic diffusion, nonparametric estimation, recursive estimation, partial observations.

1. INTRODUCTION.

Let $\big(Z_t := (X_t, Y_t) \in \mathbb{R}^{2d}\,,\ t \geq 0\big)$ be governed by the following Ito stochastic differential equation :

$$\begin{aligned} dX_t &= Y_t dt \\ dY_t &= \sigma\, dW_t - (c(X_t, Y_t)Y_t + \nabla V(X_t))dt. \end{aligned} \tag{1.1}$$

Each component Y^i $(1 \leq i \leq d)$ is the velocity of a particle i with position X^i. Function c is called the damping force and V the potential, σ is some symmetric positive definite constant matrix and W is a standard Brownian motion.
We shall assume that c and V are regular enough for the existence and uniqueness of a non explosive solution of (1.1). We shall also assume that the process is ergodic with a unique invariant probability measure μ, and that the convergence in the ergodic theorem is quick enough. Some sufficient conditions will be discussed below.
These models are important due to their physical relevance. They have a long history. We refer to Wu (2001) and Cattiaux et al. (2014-a) for a detailed bibliography. We have chosen the terminology "damping Hamiltonian systems" in reference to Wu. Such systems are also called "kinetic diffusions" by several authors.

Once the probabilistic picture is well understood, it is particularly relevant to build statistical tools for these models. In a series of previous papers, we have built several estimators for the characteristics of such processes: non parametric estimators for the

Date: October 3, 2014.

invariant density in Cattiaux, Leon and Prieur (2014-a), for the drift term in Cattiaux, Leon and Prieur (2014-b), and various estimators for the noise or the diffusion term in Cattiaux, Leon and Prieur (2014-c). The most important fact is that all these estimators are built under partial observations, i.e. we assume that the position only can be observed, the velocity is then estimated. This is in accordance with the concrete situations where these processes can be used as a pertinent model.
The second common fact in these papers is that we assume high frequency observations, i.e. the process X_t is observed at times kh_n ($k \in \mathbb{R}^+$) for some parameter h_n going to 0 as n goes to infinity (but of course $nh_n \to +\infty$).

In the present paper, we propose recursive estimators.
Recursive estimation is of course particularly well adapted to the ergodic situation, where one can observe the process on a long time interval. To be more precise, denote by p_s the invariant density (see the next section for its existence) and by b the drift term

$$b(x,y) = -c(x,y)y - \nabla V(x)\,.$$

We introduce the two natural estimations of Y and dY,

$$\Delta_1(X_t,h) = \frac{1}{h}\,(X_{t+h} - X_t) \quad \text{ and } \quad h\,\Delta_2(X_t,h) = \frac{1}{h}\,(X_{t+h} + X_{t-h} - 2X_t)\,. \tag{1.2}$$

We define the following three estimators

$$\hat{p}_{r,n}(x,y) := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{b_{1,i}^d\, b_{2,i}^d}\, K\left(\frac{x - X_i}{b_{1,i}}, \frac{y - \Delta_1(X_i,h_i)}{b_{2,i}}\right)\,, \tag{1.3}$$

for the invariant density,

$$\hat{p}_{r,n}(x,y)\,\hat{b}_{r,n}(x,y) := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{b_{1,i}^d\, b_{2,i}^d}\, K\left(\frac{x - X_i}{b_{1,i}}, \frac{y - \Delta_1(X_i,h_i)}{b_{2,i}}\right)\,\Delta_2(X_{i+2h_i},h_i)\,, \tag{1.4}$$

for the drift term, and finally

$$\hat{\sigma}_{r,n} := \frac{3}{2n}\sum_{i=1}^{n} \frac{1}{h_i}\,\Delta_2(X_{i+h_i},h_i) >< \Delta_2(X_{i+h_i},h_i)\,, \tag{1.5}$$

for the (constant) diffusion matrix, where $A = x >< x$ denotes the symmetric matrix with entries $A_{j,k} = x^j\, x^k$, x^j denoting the jth coordinate of x.
The recursion formulae are thus

$$\begin{aligned}
\hat{p}_{r,n}(x,y) &= \frac{n-1}{n}\,\hat{p}_{r,(n-1)}(x,y) + \frac{1}{n\, b_{1,n}^d\, b_{2,n}^d}\, K\left(\frac{x - X_n}{b_{1,n}}, \frac{y - \Delta_1(X_n,h_n)}{b_{2,n}}\right)\,,\\
\hat{p}_{r,n}(x,y)\,\hat{b}_{r,n}(x,y) &= \frac{n-1}{n}\,\hat{p}_{r,(n-1)}(x,y)\,\hat{b}_{r,(n-1)}(x,y) +\\
&\quad + \frac{1}{n\, b_{1,n}^d\, b_{2,n}^d}\, K\left(\frac{x - X_n}{b_{1,n}}, \frac{y - \Delta_1(X_n,h_n)}{b_{2,n}}\right)\,\Delta_2(X_{n+2h_n},h_n)\,,\\
\hat{\sigma}_{r,n} &= \frac{n-1}{n}\,\hat{\sigma}_{r,(n-1)} + \frac{3}{2nh_n}\,\Delta_2(X_{n+h_n},h_n) >< \Delta_2(X_{n+h_n},h_n)\,.
\end{aligned}$$

We intend to give sufficient conditions on the discretization steps h_n, bandwidth $b_{1,n}$ and $b_{2,n}$ and kernels K, for these estimators to be consistent and provide some confidence

intervals.
In the whole paper K will be a smooth kernel with compact support, satisfying

$$\int K(x,y)dxdy = 1\,.$$

It is worth noticing that the observations will be only locally high frequency, i.e. what is needed is the joint observation of $(X_n, X_{n+h_n}, X_{n+2h_n}, X_{n+3h_n})$ and we will choose h_n going to 0 and assume that $3h_n < 1$ for all n. In the case of complete observation, we only need to observe the position process at integer times.

Of course, all these estimators are the "recursive" form of the estimators we have introduced in our previous works.

Recursive estimation for the density of stationary sequences has a long history too, starting presumably with Wegman and Davis (1979). One can mention among others a series of papers by E. Masry (Masry (1986), Masry (1987), Masry (1989), Masry and Györfi (1987)) and more recently Liang and Baek (2004) and Amiri (2009). We refer to Amiri (2010) for a more complete bibliography.
Though these papers do not study explicitly diffusion processes, the results therein can be applied to the case of complete observation of some stationary diffusion processes. However we decided to give here complete new proofs for two major reasons: first the proofs of the Central Limit Theorem given in these papers, based on an intricate block decomposition can and will be simplified, second we are dealing with the partial observation case. In addition, some key assumption in all these papers is not clearly satisfied for our model, though it is presumably true. A precise description and comparison will be given later (see Section 5.2 for further discussion).

The plan of the paper is the following: after recalling some properties of the model, we study recursive estimators for the invariant density and the drift under complete observation. Then, as in Cattiaux et al. (2014-a) and Cattiaux et al. (2014-b), we show that looking at estimators with partial observation, introduce a small perturbation of the fully observed case, at least for well chosen h_n. Finally we study the diffusion coefficient in the spirit of the final section of Cattiaux et al. (2014-c).
Large parts of the proofs being similar to what we have done in our previous works, we only indicate the modifications or simplifications in the present setting.

The main results of the paper can be summarized as follows: choosing $h_n = n^{-\beta}$ for some $\beta > 1/2$ as close as $1/2$ we want, one can find kernels K and bandwidths $b_{n,j}$ such that the recursive estimators simultaneously converge to the invariant density, drift term and diffusion term. In addition we are able to find confidence intervals for all these quantities, of respective size $n^{-\theta_p}$ for $\theta_p < 1/2$, $n^{-\theta_b}$ for $\theta_b < 1/4$ and $n^{-1/2}$.

2. The model and its properties.

We are obliged to recall some facts on the model. A more detailed discussion is contained in Cattiaux et al. (2014-a).

We shall first give some results about non explosion and long time behaviour. In a sense, coercivity can be seen in this context as some exponential decay to equilibrium.

Let $|\cdot|$ denote the euclidean norm in $\mathbb{R}^d$.

Let us first introduce some sets of assumptions:

Hypothesis $\mathcal{H}_1$:

(i) the potential V is lower bounded, smooth over $\mathbb{R}^d$, V and its first three partial derivatives have polynomial growth at infinity and there exists $v > 0$ such that

$$+\infty \geq \liminf_{|x|\to+\infty} \frac{x.\nabla V(x)}{|x|} \geq v > 0\,,$$

the latter being often called "drift condition",

(ii) the damping coefficient $c(x,y)$ is smooth and bounded, its first two derivatives are bounded by some polynomial, and there exist $c, L > 0$ so that $c^s(x,y) \geq cId > 0$, $\forall(|x| > L, y \in \mathbb{R}^d)$, where $c^s(x,y)$ is the symmetrization of the matrix $c(x,y)$, given by $\frac{1}{2}(c_{ij}(x,y) + c_{ji}(x,y))_{1\leq i,j\leq d}$,

(iii) σ is uniformly elliptic, i.e. here σ^{-1} exists.

These conditions ensure that there is no explosion, and that the process is positive recurrent with a unique invariant probability measure μ. In some of our previous papers we assumed that $\sigma = Id$, but of course multiplying Y by σ^{-1} we are immediately reduced (for the probabilistic properties) to the case $\sigma = Id$. Furthermore μ admits some exponential moment, hence polynomial moments of any order.

We will denote by $P_t f(z) = \mathbb{E}_z(f(Z_t))$ which is well defined for all bounded function f, P_t extends as a contraction semi-group on $\mathbb{L}^p(\mu)$ for all $1 \leq p \leq +\infty$.

Another key feature is that the process is actually α-mixing, i.e.

Proposition 2.1. *There exist some constants $C > 0$ and $\rho < 1$ such that:*

$$\forall\, g, f \in \mathbb{L}^\infty(\mu)\,, \quad \forall\, t \geq 0,$$

$$|\mathrm{Cov}_\mu\,(f(Z_t), g(Z_0))| \leq C\,\rho^{t/2}\ \left\|g - \int g d\mu\right\|_\infty \left\|f - \int f d\mu\right\|_\infty . \tag{2.2}$$

i.e., in the stationary regime, $(Z_t,\, t \geq 0)$ is α-mixing with exponential rate.

Actually, using the semi-group property and the Riesz-Thorin interpolation theorem, one can extend the previous mixing condition in the following way: if $g, f \in \mathbb{L}^q(\mu)$ for some $q \geq 2$, then for $t > 0$,

$$|\mathrm{Cov}_\mu\,(f(Z_t), g(Z_0))| \leq C\,\rho^{t(q-2)/2q}\ \left\|g - \int g d\mu\right\|_q \left\|f - \int f d\mu\right\|_q . \tag{2.3}$$

As explained in section 2.2 of Cattiaux et al. (2014-a), the infinitesimal generator L is hypo-elliptic, which implies that

$$\mu(dz) = p_s(z)\,dz$$

for some smooth function p_s.

At the same time, for all $t > 0$ the law of the process has a density $p(t,z,.)$ with respect to Lebesgue measure, which is smooth, and satisfies for any $t_0 > 0$ and for all compact subsets A and A',

$$\sup_{z\in A\,,\, z'\in A'}\ \sup_{t\geq t_0} |p(t,z,z')| = Q(A,A') < +\infty\,. \tag{2.4}$$

One can relax the C^∞ assumption on the coefficients into a C^k assumption, for a large enough k, but this is irrelevant.
Furthermore it can be shown that p_s is everywhere positive.

One can relax some assumptions and still have the same conclusions:

Hypothesis $\mathcal{H}_2$:

(a) One can relax the boundedness assumption on c in $\mathcal{H}_1$, assuming that for all $N > 0$: $\sup_{|x|\leq N, y\in\mathbb{R}^d} \|c(x,y)\|_{H.S.} < +\infty$, where $H.S.$ denotes the Hilbert-Schmidt norm of a matrix; but one has to assume in addition conditions (3.1) and (3.2) in Wu (2001). An interesting example (the Van der Pol model) in this situation is described in Wu (2001) subsection 5.3.

(b) The most studied situation is the one when c is a constant matrix. Actually almost all results obtained in Wu (2001) or Bakry et al. (2008) in this situation extend to the general bounded case.
Nevertheless we shall assume now that c is a constant matrix.
In this case a very general statement replacing $\mathcal{H}_1$ (i) is given in Theorem 6.5 of Bakry et al. (2008). Tractable examples are discussed in Example 6.6 of the same paper. In particular one can replace the drift condition on V by

$$\liminf_{|x|\to+\infty} |\nabla V|^2(x) > 0 \quad \text{and} \quad \|\nabla^2 V\|_{H.S.} \ll |\nabla V| \,.$$

Notice that one can relax the repelling strength of the potential, and obtain, no more exponential but sub-exponential or polynomial decay (see the discussion in Bakry et al. (2008)).

From now on in the whole paper we will assume that Hypothesis $\mathcal{H}_1$ (or $\mathcal{H}_2$) is fulfilled. In all the proofs of the paper C denotes some constant which may vary from line to line. We also use the expression in the stationary regime when the process is stationary, i.e. when Z_0 is distributed according to μ.

3. Estimation of the invariant density and of the drift term in the stationary regime: complete observation case.

To begin with, we look at the recursive estimators under complete observation. This is not our main goal, but as in Cattiaux et al. (2014-a), it is a first step for our study.

3.1. **Invariant density.**

Introduce

$$\tilde{p}_{r,n}(x,y) := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{b_{1,i}^d\, b_{2,i}^d}\, K\left(\frac{x-X_i}{b_{1,i}}, \frac{y-Y_i}{b_{2,i}}\right) . \tag{3.1}$$

We have the following first result

Lemma 3.2. *Assume that $b_{1,n}$ and $b_{2,n}$ are non-increasing sequences of real numbers (say between 0 and 1), going to 0 as $n \to +\infty$. Then for all (x,y), as $n \to +\infty$,*

$$\mathbb{E}_\mu(\tilde{p}_{r,n}(x,y)) \to p_s(x,y) \,.$$

If in addition $nb_{1,n}^d\, b_{n,2}^d \to +\infty$ *as* $n \to +\infty$*, then*

$$\mathrm{Var}_\mu(\tilde p_{r,n}(x,y)) \le \frac{C}{n\, b_{1,n}^d\, b_{n,2}^d}\,,$$

hence $\tilde p_{r,n}(x,y)$ *converges to* $p_s(x,y)$ *in* $\mathbb{L}^2(\mu)$ *norm, as* $n \to +\infty$*.*

Proof. A simple change of variables shows that

$$\mathbb{E}_\mu\left(\frac{1}{b_{1,i}^d\, b_{2,i}^d}\, K\left(\frac{x-X_i}{b_{1,i}}, \frac{y-Y_i}{b_{2,i}}\right)\right) = \int K(u,v)\, p_s(x-ub_{1,i}, y-vb_{2,i})\, du\, dv$$

so that using the smoothness of p_s and the fact that K has compact support, it follows that

$$\mathbb{E}_\mu\left(\frac{1}{b_{1,i}^d\, b_{2,i}^d}\, K\left(\frac{x-X_i}{b_{1,i}}, \frac{y-Y_i}{b_{2,i}}\right)\right) \to p_s(x,y)$$

as i goes to infinity. The first result then follows from the Cesaro rule.

For the variance we have the decomposition $\mathrm{Var}_\mu(\tilde p_{r,n}(x,y)) = V_1(n) + 2V_2(n)$ with

$$\begin{aligned} V_1(n) &= \frac{1}{n^2}\sum_{i=1}^n \frac{1}{b_{1,i}^{2d}\, b_{2,i}^{2d}}\, \mathrm{Var}_\mu\left(K\left(\frac{x-X_0}{b_{1,i}}, \frac{y-Y_0}{b_{2,i}}\right)\right) \qquad (3.3)\\ &= \frac{1}{n^2}\left(\sum_{i=1}^n \frac{1}{b_{1,i}^d\, b_{2,i}^d}\int K^2(u,v)\, p_s(x-ub_{1,i}, y-vb_{2,i})\, du\, dv\right)\\ &\qquad -\frac{1}{n^2}\sum_{i=1}^n \frac{1}{b_{1,i}^{2d} b_{2,i}^{2d}}\left(b_{1,i}^d b_{2,i}^d \int K(u,v) p_s(x-ub_{1,i}, y-vb_{2,i}) du dv\right)^2\\ &\le \frac{C}{n\, b_{1,n}^d\, b_{2,n}^d} \end{aligned}$$

where $C = \| p_s \|_{\mathbb{L}^\infty(suppK+1)} \int K^2(u,v) du dv$, and

$$\begin{aligned} V_2(n) &= \frac{1}{n^2}\sum_{i=1}^n \sum_{j>i} \frac{1}{b_{1,i}^d\, b_{2,i}^d}\, \frac{1}{b_{1,j}^d\, b_{2,j}^d} \qquad (3.4)\\ &\qquad \mathrm{Cov}_\mu\left(K\left(\frac{x-X_0}{b_{1,i}}, \frac{y-Y_0}{b_{2,i}}\right), K\left(\frac{x-X_{j-i}}{b_{1,j}}, \frac{y-Y_{j-i}}{b_{2,j}}\right)\right). \end{aligned}$$

In order to bound the previous quantity, we have two options.
On one hand we may write

$$\mathbb{E}_\mu\left(K\left(\frac{x-X_0}{b_{1,i}}, \frac{y-Y_0}{b_{2,i}}\right) K\left(\frac{x-X_{j-i}}{b_{1,j}}, \frac{y-Y_{j-i}}{b_{2,j}}\right)\right) = b_{1,i}^d\, b_{2,i}^d\, b_{1,j}^d\, b_{2,j}^d \int\int K(u,v)\, K(u',v')$$

$$p_s(x-ub_{1,i}, y-vb_{2,i})\, p(j-i, (x-ub_{1,i}, y-vb_{2,i}), (x-u'b_{1,j}, y-v'b_{2,j}))\, du dv du' dv'$$

so that, using (2.4),

$$\frac{1}{b_{1,i}^d\, b_{2,i}^d}\, \frac{1}{b_{1,j}^d\, b_{2,j}^d}\left|\mathbb{E}_\mu\left(K\left(\frac{x-X_0}{b_{1,i}}, \frac{y-Y_0}{b_{2,i}}\right) K\left(\frac{x-X_{j-i}}{b_{1,j}}, \frac{y-Y_{j-i}}{b_{2,j}}\right)\right)\right| \le C\,.$$

At the same time,

$$\left|\mathbb{E}_\mu\left(K\left(\frac{x-X_0}{b_{1,i}},\frac{y-Y_0}{b_{2,i}}\right)\right)\right| = b_{1,i}^d\, b_{2,i}^d \left|\int K(u,v)\, p_s(x-ub_{1,i}, y-vb_{2,i})\, dudv\right| \le C\, b_{1,i}^d\, b_{2,i}^d$$

so that

$$\frac{1}{b_{1,i}^d\, b_{2,i}^d}\,\frac{1}{b_{1,j}^d\, b_{2,j}^d}\left|\mathrm{Cov}_\mu\left(K\left(\frac{x-X_0}{b_{1,i}},\frac{y-Y_0}{b_{2,i}}\right), K\left(\frac{x-X_{j-i}}{b_{1,j}},\frac{y-Y_{j-i}}{b_{2,j}}\right)\right)\right| \le C\,.$$

On the other hand, we may use (2.3) in order to get for any $2 \le q \le +\infty$,

$$\left|\mathrm{Cov}_\mu\left(K\left(\frac{x-X_0}{b_{1,i}},\frac{y-Y_0}{b_{2,i}}\right), K\left(\frac{x-X_{j-i}}{b_{1,j}},\frac{y-Y_{j-i}}{b_{2,j}}\right)\right)\right| \le C\, b_{1,i}^{d/q}\, b_{2,i}^{d/q}\, b_{1,j}^{d/q}\, b_{2,j}^{d/q}\, \rho^{(j-i)(q-2)/2q}\,.$$

Using roughly the monotonicity of the sequences $b_{.,i}$, we thus get

$$\begin{aligned} V_2(n) &\le \frac{C}{n^2}\sum_{i=1}^n\sum_{j>i} \min\left(1, \frac{\rho^{(j-i)(q-2)/2q}}{b_{1,n}^{2d(1-\frac{1}{q})}\, b_{2,n}^{2d(1-\frac{1}{q})}}\right) \\ &\le \frac{C}{n^2}\sum_{k=1}^n (n-k)\, \min\left(1, \frac{\rho^{k(q-2)/2q}}{b_{1,n}^{2d(1-\frac{1}{q})}\, b_{2,n}^{2d(1-\frac{1}{q})}}\right)\,. \end{aligned}$$

The sum splits into two terms, depending on which term reaches the minimal value. But roughly, for all $1 \le k_n \le n$, we have, with $q = +\infty$ for simplicity, a bound of the form

$$V_2(n) \le C\left(\frac{k_n}{n} + \frac{\rho^{k_n/2}}{n\, b_{1,n}^{2d}\, b_{2,n}^{2d}}\right) = o\left(\frac{1}{n\, b_{1,n}^d\, b_{2,n}^d}\right)$$

provided

$$\ln(1/b_{1,n}) + \ln(1/b_{2,n}) \ll k_n \ll \left(\frac{1}{b_{1,n}^d\, b_{2,n}^d}\right)\,,$$

hence the result. □

The previous proof gives us the feeling of the right normalization for a Central Limit Theorem. Let us state the result

Theorem 3.5. *Assume that $\mathcal{H}_1$ or $\mathcal{H}_2$ is fulfilled. Assume in addition that the following assumption $\mathcal{H}_K$ is satisfied*

($\mathcal{H}_K$) there exists $m \in \mathbb{N}^*$ such that for all multi-index $(m_1, m_2) \in \mathbb{N}^d \times \mathbb{N}^d$ such that $m \ge |m_1| + |m_2| \ge 1$, $\int u^{m_1} v^{m_2} K(u,v) dudv = 0$.

Let $b_{1,n}$ and $b_{2,n}$ be non increasing sequences of real numbers satisfying

(1a) *$b_{1,n}$ and $b_{2,n}$ go to 0 as $n \to +\infty$,*

(2a) *$nb_{1,n}^d b_{2,n}^d \to +\infty$ as $n \to +\infty$,*

(3a) *there exists $\Sigma > 0$ such that $\lim_{n\to+\infty} \frac{1}{n}\left(\sum_{i=1}^n \frac{b_{1,n}^d b_{2,n}^d}{b_{1,i}^d b_{2,i}^d}\right) = \Sigma$.*

(4a) *$\sqrt{nb_{1,n}^d b_{2,n}^d}\,\left(\frac{1}{n}\sum_{i=1}^n \max(b_{1,i}, b_{2,i})^{m+1}\right) \to 0$ as $n \to +\infty$.*

Then, in the stationary regime

$$\sqrt{nb_{1,n}^d b_{2,n}^d}\,(\tilde{p}_{r,n}(x,y) - p_s(x,y)) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0,\, \Sigma\, p_s(x,y) \int K^2(s,t)dsdt\right).$$

Remark 3.6. Typical allowed bandwidths are $b_{j,n} = n^{-\alpha_j}$ for $j = 1, 2$ with $\alpha_j > 0$, $d(\alpha_1 + \alpha_2) < 1$ and then choosing m such that $1 < d(\alpha_1 + \alpha_2) + 2(m+1)\min(\alpha_1, \alpha_2)$. In this case $\Sigma = 1/(1 + d(\alpha_1 + \alpha_2))$.
These conditions are satisfied in particular once

$$\frac{1}{2(d+m+1)} < \min(\alpha_1, \alpha_2) \le \max(\alpha_1, \alpha_2) < \frac{1}{2d},$$

in which case the rate we obtain is at most n^τ with $\tau < \frac{m+1}{2(d+m+1)}$. ♢

Remark 3.7. It is always possible to build some kernel K satisfying $\mathcal{H}_K$ for a given m. First we may take $K(z) = \Pi_{j=1}^{2d} K_1(z_j)$ and choose K_1 satisfying $\mathcal{H}_K$ in dimension 1.
To build K_1 choose for instance some even function ψ which is C^∞ and compactly supported in $[-1, +1]$, positive on its support. Then consider the space of polynomials $\mathbb{R}(u)$, equipped with the $\mathbb{L}^2(\psi(u)du)$ norm, and its completion $H(\psi)$. It is not difficult to see that, for all $k \in \mathbb{N}^*$, the family $1, u, u^2, ..., u^k$ is an independent family of vectors for the euclidean norm induced by the $\mathbb{L}^2(\psi(u)du)$ norm. We may thus find some polynomial M with degree m which is orthogonal to all the u^j for $j = 1, ..., m$ and satisfies $< M, 1 >= 1$ with $< \cdot, \cdot >$ the usual scalar product in $\mathbb{L}^2(\psi(u)du)$. It remains to choose $K_1 = M\psi$.
The previous shows that we may choose m as large as we want. ♢

Proof. First, the bias term $B_n = \sqrt{n\, b_{1,n}^d\, b_{2,n}^d}\,(\mathbb{E}_\mu(\tilde{p}_{r,n}(x,y)) - p_s(x,y))$ can be treated exactly as in Cattiaux et al. (2014-a) section 5 step 3. It is for this term that the conditions on K and (4a) are required.
Next we show that $n\, b_{1,n}^d\, b_{2,n}^d\, \mathrm{Var}_\mu(\tilde{p}_{r,n}(x,y))$ converges as $n \to +\infty$. According to the proof of the previous Lemma, we see that this convergence amounts to the convergence of

$$n\, b_{1,n}^d\, b_{2,n}^d\, V_1(n).$$

But

$$n\, b_{1,n}^d\, b_{2,n}^d\, V_1(n) = V_{11}(n) + V_{12}(n)$$

with

$$V_{11}(n) = \frac{1}{n}\left(\sum_{i=1}^n \frac{b_{1,n}^d b_{2,n}^d}{b_{1,i}^d b_{2,i}^d}\right) p_s(x,y) \int K^2(u,v)dudv,$$

while

$$V_{12}(n) = \frac{1}{n}\left(\sum_{i=1}^n \frac{b_{1,n}^d b_{2,n}^d}{b_{1,i}^d b_{2,i}^d} \int K^2(u,v)(p_s(x - ub_{1,i}, y - vb_{2,i}) - p_s(x,y))\, dudv\right).$$

Using the regularity of p_s (here the local Lipschitz property) we see that

$$V_{12}(n) \le C\, \frac{b_{1,n}b_{2,n}}{n}\left(\sum_{i=1}^n \frac{b_{1,n}^{d-1} b_{2,n}^{d-1}}{b_{1,i}^{d-1} b_{2,i}^{d-1}}\right) \le C\, b_{1,n}b_{2,n}$$

thanks to the monotonicity of the sequences, and the latter goes to 0. Finally

$$\lim_{n\to+\infty} n\, b_{1,n}^d\, b_{2,n}^d\, \mathrm{Var}_\mu(\tilde{p}_{r,n}(x,y)) = \Sigma\, p_s(x,y) \int K^2(u,v)dudv\,. \tag{3.8}$$

To prove the Central Limit Theorem it remains to use the Lindeberg method. As shown in Bardet et al. (2008) Theorem 1, the proof amounts to check two conditions. Define

$$U_{i,n} = \frac{\sqrt{b_{1,n}^d\, b_{2,n}^d}}{\sqrt{n}\, b_{1,i}^d\, b_{2,i}^d}\, K\left(\frac{x-X_i}{b_{1,i}}, \frac{y-Y_i}{b_{2,i}}\right)$$

then what is required is

(1) there exists some $\delta > 0$ such that

$$\lim_{n\to+\infty} A_n := \lim_{n\to+\infty} \sum_{i=1}^n \mathbb{E}_\mu(|U_{i,n}|^{2+2\delta}) \;=\; 0\,,$$

(2) for all $a \in \mathbb{R}$,

$$\lim_{n\to+\infty} T_n := \lim_{n\to+\infty} \sum_{i=1}^n \left|\mathrm{Cov}_\mu\left(e^{ia(U_{1,n}+\ldots+U_{i-1,n})}, e^{iaU_{i,n}}\right)\right| = 0\,.$$

We have

$$\begin{aligned} A_n &= \frac{1}{n^{1+\delta}} \sum_{i=1}^n \frac{b_{1,n}^{d(1+\delta)} b_{2,n}^{d(1+\delta)}}{b_{1,i}^{2d(1+\delta)} b_{2,i}^{2d(1+\delta)}}\, \mathbb{E}_\mu\left(K^{2+2\delta}\left(\frac{x-X_0}{b_{1,i}}, \frac{y-Y_0}{b_{2,i}}\right)\right) \\ &\le \frac{C}{n^{1+\delta}} \sum_{i=1}^n \frac{b_{1,n}^{d(1+\delta)} b_{2,n}^{d(1+\delta)}}{b_{1,i}^{d(1+2\delta)} b_{2,i}^{d(1+2\delta)}} \\ &\le \frac{C}{n^\delta\, b_{1,n}^{d\delta} b_{2,n}^{d\delta}} \left(\frac{1}{n}\sum_{i=1}^n \frac{b_{1,n}^d b_{2,n}^d}{b_{1,i}^d b_{2,i}^d}\right) \end{aligned}$$

so that, under our assumptions, $A_n \to 0$ as $n \to +\infty$.

As in Bardet et al. (2008), we write

$$\left|\mathrm{Cov}_\mu\left(e^{ia(U_{1,n}+\ldots+U_{i-1,n})}, e^{iaU_{i,n}}\right)\right| = \left|\sum_{j=1}^{i-1} \mathrm{Cov}_\mu\left(e^{ia(U_{1,n}+\ldots+U_{j,n})} - e^{ia(U_{1,n}+\ldots+U_{j-1,n})}, e^{iaU_{i,n}}\right)\right| \tag{3.9}$$

(as usual empty sums are equal to 0). Consider a random variable $U_{i,n}^*$ independent from $(U_{1,n},\ldots,U_{j-1,n})$, with same distribution as $U_{i,n}$. Then, following Bardet et al. (2008), we note that each element in the sum of the right hand term of Equality (3.9) is equal to

$$\left|\mathrm{Cov}_\mu\left(e^{ia(U_{1,n}+\ldots+U_{j,n})} - e^{ia(U_{1,n}+\ldots+U_{j-1,n})}, \left(e^{iaU_{i,n}} - e^{iaU_{i,n}^*}\right)\right)\right|$$

which is bounded, arguing as in the proof of Lemma 3.2, by

$$\begin{aligned} &C(a)\min\left(\mathbb{E}_\mu(|U_{j,n}||U_{i,n}|), \rho^{i-j/2}\frac{b_{1,n}^d\, b_{2,n}^d}{n\, b_{1,j}^d\, b_{2,j}^d b_{1,i}^d\, b_{2,i}^d}\right) \\ &\qquad \le C(a)\,\frac{b_{1,n}^d\, b_{2,n}^d}{n}\,\min\left(1, \frac{\rho^{i-j/2}}{b_{1,i}^d b_{2,i}^d b_{1,j}^d b_{2,j}^d}\right) \quad . \end{aligned}$$

$$\begin{aligned} T_n &\leq C(a) \sum_{i=1}^{n} \sum_{j=1}^{i-1} \frac{b_{1,n}^d\, b_{2,n}^d}{n} \min\left(1, \frac{\rho^{i-j/2}}{b_{1,i}^d b_{2,i}^d b_{1,j}^d b_{2,j}^d}\right) \\ &\leq C(a) \sum_{k=1}^{n} (n-k) \min\left(\frac{b_{1,n}^d\, b_{2,n}^d}{n}\,,\ \frac{\rho^{k/2}}{n b_{1,n}^d b_{2,n}^d}\right) \\ &\leq C(a)(k(n)\, b_{1,n}^d\, b_{2,n}^d + \frac{\rho^{k(n)}}{b_{1,n}^d b_{2,n}^d}) \end{aligned}$$

so that choosing again

$$\frac{1}{b_{1,n}^d\, b_{2,n}^d} \gg k(n) \gg \ln(1/b_{1,n}) + \ln(1/b_{2,n})$$

we obtain that $T_n \to 0$. □

Remark 3.10. As we have seen $\mathcal{H}_K$ is only required to control the bias term.
Lemma 3.2 is essentially contained in Amiri (2010) chapter 2, where a still more complicated estimator is studied. For the Central Limit Theorem, usual proofs in the recursive framework are using intricate block decompositions. Here we follow the beautiful and now classical Lindeberg method (using Rio's decomposition). The calculations are done in Bardet et al. (2008) proof of Proposition 4.2 for the case of the classical kernel estimator of the density, under weak dependence. It is adapted here to our framework. ♢

3.2. **Drift term.** The most difficult part in our previous study is presumably the estimation of the drift term in Cattiaux et al. (2014-b) Proposition 4.2. Similarly to what we have done in this paper introduce

$$\tilde{H}_{r,n}(x,y) := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{b_{1,i}^d\, b_{2,i}^d} K\left(\frac{x-X_i}{b_{1,i}}, \frac{y-Y_i}{b_{2,i}}\right) \Delta_2(X_{i+2h_i}, h_i)\,. \tag{3.11}$$

As for the density, using Cesaro's rule, it is easy to see that $\tilde{H}_{r,n}$ is an asymptotically unbiased estimator of $b(x,y)p_s(x,y)$, provided the bandwidths $b_{.,n}$ and h_n go to 0 at infinity. Next it is enough to closely follow the proof of Proposition 4.2 in Cattiaux et al. (2014-b) keeping in mind that if the times ih_n are replaced by i, this does not introduce any modification since what is important is the size of the windows defining $\Delta_j(X, h_i)$. Since some things will differ (in particular the normalization of the kernels), we will give a proof (indicating the differences with Cattiaux et al. (2014-b)) of the following statement

Theorem 3.12. *Assume that $\mathcal{H}_1$ or $\mathcal{H}_2$, and $\mathcal{H}_K$ are fulfilled.*
Let h_n, $b_{1,n}$ and $b_{2,n}$ be non increasing sequences of real numbers satisfying, as $n \to +\infty$,

(1b) $h_n < 1/3$ *for all* n, h_n, $b_{1,n}$ *and* $b_{2,n}$ *go to* 0,
(2b) $n\, h_n\, b_{1,n}^d b_{2,n}^d \to +\infty$,
(3b) *there exists* $\theta > 0$ *such that* $\frac{1}{n}\left(\sum_{i=1}^{n} \frac{h_n\, b_{1,n}^d b_{2,n}^d}{h_i\, b_{1,i}^d b_{2,i}^d}\right) \to \theta$,
(4b) $\sqrt{n b_{1,n}^d b_{2,n}^d h_n}\ \left(\frac{1}{n} \sum_{i=1}^{n} \max(b_{1,i}, b_{2,i})^{m+1}\right) \to 0$,
(5b) *there exists some* $r > 1$ *such that* $\sqrt{n b_{1,n}^d b_{2,n}^d h_n}\ \left(\frac{1}{n} \sum_{i=1}^{n} h_i\, \left(b_{1,i}^d b_{2,i}^d\right)^{\frac{1}{r}-1}\right) \to 0\,,$

(6b) *there exists some* $p>1$ *such that* $\frac{h_n b_{1,n}^d b_{2,n}^d}{n}\sum_{i=1}^n h_i^2\,(b_{1,i}^d b_{2,i}^d)^{\frac{1}{p}-2}\to 0$,

Then, in the stationary regime

$$\sqrt{nb_{1,n}^d b_{2,n}^d h_n}\left(\tilde{H}_{r,n}(x,y)-b(x,y)p_s(x,y)\right)\xrightarrow[n\to+\infty]{\mathcal{D}}\sigma\mathcal{N}\left(0,\,\frac{\theta}{3}\,p_s(x,y)(\int K^2(s,t)dsdt)\,Id\right).$$

Remark 3.13. Assume that h_n is such that $\sum_{i=1}^n h_i^2<+\infty$. Then for (6) to be satisfied, it suffices that $\frac{h_n(b_{1,n}^d b_{2,n}^d)^{1/p}}{n\,b_{1,n}^d b_{2,n}^d}\to 0$, which is automatic, thanks to (2). This situation is not the most interesting.
Now assume that $b_{j,n}=n^{-\alpha_j}$ and $h_n=n^{-\beta}$ for some positive α_j and β. It is not difficult to see that, choosing p as close to 1 as necessary, (6) is again automatically satisfied, whatever the positive value of β. For (1) to (5) to be satisfied, it is then enough to assume

$$\beta+d(\alpha_1+\alpha_2)<1<\beta+d(\alpha_1+\alpha_2)+2\min(\beta,(m+1)\min(\alpha_1,\alpha_2))\,,$$

and in this case we have $\theta=1/(1+\beta+d(\alpha_1+\alpha_2))$.
As we have seen in Remark 3.7, we may choose m as large as we want, so that the previous conditions amounts to

$$\beta+d(\alpha_1+\alpha_2)<1<3\beta+d(\alpha_1+\alpha_2)\,. \tag{3.14}$$

Hence we may choose the α_j's as small as possible and β larger than but close to $1/3$ to get the quasi optimal rate $n^{1/3}$. ♢

Proof. We give the main modifications w.r.t. the proof of Proposition 4.2 in Cattiaux et al. (2014-b). We decompose

$$\begin{aligned}\mathcal{S}_n &:= \sqrt{n\,b_{1,n}^d b_{2,n}^d\,h_n}\big(\tilde{H}_{r,n}(x,y)-b(x,y)p_s(x,y)\big)\\ &= \sqrt{n\,b_{1,n}^d b_{2,n}^d h_n}\big(\tilde{H}_n(x,y)-\mathbb{E}\tilde{H}_{r,n}(x,y)+\mathbb{E}\tilde{H}_{r,n}(x,y)-b(x,y)p_s(x,y)\big)\\ &=: \mathcal{I}_{1n}+\mathcal{I}_{2n}.\end{aligned}$$

Define

$$I_i:=\int_{i+2h_i}^{i+3h_i}\int_{i+2h_i}^{t}b(X_s,Y_s)dsdt+\int_{i+h_i}^{i+2h_i}\int_{t}^{i+2h_i}b(X_s,Y_s)dsdt\,,$$

and

$$\mathfrak{W}_i:=\int_{i+2h_i}^{i+3h_i}(W_s-W_{i+2h_i})ds+\int_{i+h_i}^{i+2h_i}(W_{i+2h_i}-W_s)ds\,.$$

The vector $\sqrt{nb_{1,n}^d b_{2,n}^d h_n}\tilde{H}_{r,n}(x,y)$ can be decomposed in two terms: the one driving the bias in the central limit theorem

$$S_{n,1}(x,y):=\sqrt{nb_{1,n}^d b_{2,n}^d h_n}\;\frac{1}{n}\sum_{i=1}^n\frac{1}{b_{1,i}^d b_{2,i}^d}\,K\left(\frac{x-X_i}{b_{1,i}},\frac{y-Y_i}{b_{2,i}}\right)\frac{1}{h_i^2}I_i\,,$$

and the one driving the variance

$$S_{n,2}(x,y) := \sqrt{nb_{1,n}^d b_{2,n}^d h_n}\ \frac{1}{n}\sum_{i=1}^n \frac{1}{b_{1,i}^d b_{2,i}^d} K\left(\frac{x-X_i}{b_{1,i}}, \frac{y-Y_i}{b_{2,i}}\right) \frac{1}{h_i^2}\,\sigma\,\mathfrak{W}_i.$$

Notice that $\mathbb{E}S_{n,2}(x,y)=0$. We thus have

$$\mathcal{I}_{2n} = \mathbb{E}_\mu S_{n,1}(x,y) - \sqrt{n\, b_{1,n}^d b_{2,n}^d h_n}\, b(x,y)p_s(x,y)\,, \tag{3.15}$$

while

$$\mathcal{I}_{1n} = (S_{n,1} - \mathbb{E}_\mu S_{n,1}(x,y)) + S_{n,2}(x,y)\,. \tag{3.16}$$

First step: Showing that $\mathcal{I}_{2n} \to 0$.

We define

$$\mathfrak{P}_i := \int_{i+2h_i}^{i+3h_i}\int_{i+2h_i}^t (P_s b(X_i,Y_i) - b(X_i,Y_i))dsdt + \int_{i+h_i}^{i+2h_i}\int_t^{i+2h_i} (P_s b(X_i,Y_i) - b(X_i,Y_i))dsdt\,.$$

Thanks to stationarity, it holds

$$\mathcal{I}_{2n} = \sqrt{nb_{1,n}^d b_{2,n}^d h_n}\ \mathbb{E}_\mu\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{b_{1,i}^d b_{2,i}^d} K\left(\frac{x-X_0}{b_{1,i}}, \frac{y-Y_0}{b_{2,i}}\right)\ \frac{1}{h_i^2}\,\mathfrak{P}_0\right) +$$

$$+\sqrt{nb_{1,n}^d b_{2,n}^d h_n}\,\mathbb{E}_\mu\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{b_{1,i}^d b_{2,i}^d} K\left(\frac{x-X_0}{b_{1,i}}, \frac{y-Y_0}{b_{2,i}}\right) b(X_0,Y_0)\ -\ b(x,y)p_s(x,y)\right).$$

The second summand in the above expression can be treated as for the density and goes to 0 as soon as

$$\sqrt{nb_{1,n}^d b_{2,n}^d h_n}\ \left(\frac{1}{n}\sum_{i=1}^n \max(b_{1,i},b_{2,i})^{m+1}\right) \to 0\,.$$

The first summand can be decomposed into a sum $\sqrt{nb_{1,n}^d b_{2,n}^d h_n}\ \sum_{i=1}^n (A_{1i}+A_{2i})$ with

$$A_{1i} = \frac{1}{h_i^2}\int\int_{2h_i}^{3h_i}\int_{2h_i}^t (P_s b(u,v) - b(u,v))\ \frac{1}{b_{1,i}^d b_{2,i}^d} K\left(\frac{x-u}{b_{1,i}}, \frac{y-v}{b_{2,i}}\right) ds\,dt\,\mu(du,dv)\,,$$

A_{2i} being similar just changing $\int_{2h_i}^{3h_i}\int_{2h_i}^t$ into $\int_{h_i}^{2h_i}\int_t^{2h_i}$. We thus only study A_{1i} for which we have ($|.|$ denoting the norm in $\mathbb{R}^d$) as in Cattiaux et al. (2014-b)

$$|A_{1i}| \le CM_p\, h_i\, \frac{1}{b_{1,i}^d b_{2,i}^d}\ \left(\int K^r\left(\frac{x-u}{b_{1,i}}, \frac{y-v}{b_{2,i}}\right)\ \mu(du,dv)\right)^{1/r} \le CM_p\, h_i\ \left(b_{1,i}^d b_{2,i}^d\right)^{\frac{1}{r}-1}\ ,$$

with $p>1\,,\ r<+\infty \in \mathbb{N}^*$ such that $\frac{1}{p}+\frac{1}{r}=1$. Accordingly the first summand goes to zero as n tends to infinity, provided

$$\sqrt{nb_{1,n}^d b_{2,n}^d h_n}\ \left(\frac{1}{n}\sum_{i=1}^n h_i\ \left(b_{1,i}^d b_{2,i}^d\right)^{\frac{1}{r}-1}\right) \to 0\,.$$

Second step: Study of $S_{n,2}$

We now consider the term driving the variance. As in the proof of Proposition 4.2 in Cattiaux et al. (2014-b) one can study the behavior of the characteristic function $\Phi_n(t)$ of $S_{n,2}$, recall that

$$S_{n,2}(x,y) := \sqrt{nb_{1,n}^d b_{2,n}^d h_n}\frac{1}{n}\sum_{i=1}^n \frac{1}{b_{1,i}^d b_{2,i}^d} K\left(\frac{x-X_i}{b_{1,i}}, \frac{y-Y_i}{b_{2,i}}\right)\frac{\sigma\,\mathfrak{W}_i}{h_i^2}\,.$$

and that $\mathbb{E}S_{n,2}(x,y)=0$.
If $<\cdot,\cdot>$ denotes the usual scalar product in $\mathbb{R}^d$, introduce the random function $f_n(t)$ $(t\in\mathbb{R}^d)$

$$f_n(t) = e^{-\frac{1}{6n}|\sigma t|^2 b_{1,n}^d b_{2,n}^d h_n \sum_{k=1}^n \frac{1}{b_{1,k}^{2d} b_{2,k}^{2d} h_k} K^2\left(\frac{x-X_k}{b_{1,k}}, \frac{y-Y_k}{b_{2,k}}\right)}.$$

One can follow the method used in Cattiaux et al. (2014-b) to show that the convergence of Φ_n amounts to the convergence in Probability of f_n.
Now define

$$Z_n = b_{1,n}^d b_{2,n}^d\, h_n\, \frac{1}{n}\sum_{k=1}^n \frac{1}{b_{1,k}^{2d} b_{2,k}^{2d}\, h_k} K^2\left(\frac{x-X_k}{b_{1,k}}, \frac{y-Y_k}{b_{2,k}}\right).$$

Let

$$A = \theta\, p_s(x,y)\int K^2(u,v)dudv\,,$$

where

$$\theta = \lim_{n\to+\infty}\frac{1}{n}\sum_{k=1}^n \frac{b_{1,n}^d b_{2,n}^d\, h_n}{b_{1,k}^d b_{2,k}^d\, h_k}\,.$$

Then

$$\begin{aligned}
\mathbb{E}_\mu(|Z_n - A|) &\le \frac{1}{n}\sum_{k=1}^n \frac{b_{1,n}^d b_{2,n}^d\, h_n}{b_{1,k}^d b_{2,k}^d\, h_k}\int K^2(u,v)\,|p_s(x-ub_{1,k}, y-vb_{2,k}) - p_s(x,y)|dudv \\
&\quad + \left(\frac{1}{n\theta}\sum_{k=1}^n \frac{b_{1,n}^d b_{2,n}^d\, h_n}{b_{1,k}^d b_{2,k}^d\, h_k} - 1\right) A \\
&\le \frac{1}{n}\sum_{k=1}^n \int K^2(u,v)\,|p_s(x-ub_{1,k}, y-vb_{2,k}) - p_s(x,y)|dudv \\
&\quad + \left(\frac{1}{n\theta}\sum_{k=1}^n \frac{b_{1,n}^d b_{2,n}^d\, h_n}{b_{1,k}^d b_{2,k}^d\, h_k} - 1\right) A\,,
\end{aligned}$$

using the monotonicity properties of all sequences. This proves, using Cesaro's rule again that $\mathbb{E}_\mu(|Z_n - A|)\to 0$.
Thus $Z_n \to \theta\, p_s(x,y)\int K^2(u,v)dudv$, in $L^1(\mu)$. Using the bounded convergence theorem, we deduce that

$$f_n(t) \xrightarrow{\mathbb{P}} e^{-\frac{\theta|\sigma t|^2}{6} p_s(x,y)\int K^2(u,v)dudv}\,,$$

so that we get that

$$S_{n,2}(x,y) \xrightarrow[n\to+\infty]{\mathcal{D}} \sigma\,\mathcal{N}\left(0,\ \frac{\theta}{3}\,p_s(x,y)(\int K^2(s,t)dsdt)\,Id\right).$$

Third step: showing that $S_{n,1}(x,y)) - \mathbb{E}_\mu(S_{n,1}(x,y)) \to 0$

Let us denote by $\mathfrak{P}_i^k$ the k^{th} coordinate of the vector $\mathfrak{P}_{i,}$. Defining

$$\Gamma^k(i,x,y,X,Y) = \frac{1}{b_{1,i}^d b_{2,i}^d}\,K(\frac{x-X}{b_{1,i}},\frac{y-Y}{b_{2,i}})\frac{1}{h_i^2}\mathfrak{P}_i^k,$$

we have

$$S_{n,1}^k = \sqrt{nh_n b_{1,n}^d b_{2,n}^d}\ \frac{1}{n}\sum_{i=1}^n \Gamma^k(i,x,y,X_i,Y_i),$$

so that

$$\begin{aligned}&\tfrac{n}{b_{1n}^d b_{2n}^d h_n}\mathrm{Var}_\mu(S_{n,1}^k)\\ &\quad= \left(\sum_{i=1}^n \mathrm{Var}_\mu(\Gamma^k(i,x,y,X_i,Y_i)) + \sum_{i\neq l} Cov_\mu(\Gamma^k(i,x,y,X_i,Y_i),\Gamma^k(l,x,y,X_l,Y_l))\right).\end{aligned}$$

To bound the above expression we first bound as we did for the first step

$$\mathbb{E}_\mu((\Gamma^k(i,x,y,X_i,Y_i))^2) \leq$$

$$\begin{aligned}&\leq \frac{C}{h_i^4 b_{1,i}^{2d} b_{2,i}^{2d}}\\ &\quad \mathbb{E}_\mu\left[K^2\left(\frac{x-X_0}{b_{1,i}},\frac{y-Y_0}{b_{2,i}}\right)\left(\int_0^{h_i}(h_i-s)\,(P_s b(X_0,Y_0)-b(X_0,Y_0))\,ds\right)^2\right]\\ &\leq \frac{C}{h_i b_{1,i}^{2d} b_{2,i}^{2d}}\int\int_0^{h_i} K^2\left(\frac{x-u}{b_{1,i}},\frac{y-v}{b_{2,i}}\right)(P_s b(u,v)-b(u,v))^2\,ds\,d\mu\\ &\leq \frac{C\,h_i}{b_{1,i}^{2d} b_{2,i}^{2d}}\int_0^{h_i}\int K^2\left(\frac{x-u}{b_{1,i}},\frac{y-v}{b_{2,i}}\right)\left(\frac{P_s b(u,v)-b(u,v)}{s}\right)^2 d\mu\,ds\end{aligned}$$

We may argue as in the first step, this time using Hölder inequality for some conjugate pair (p,q) to conclude that

$$\mathbb{E}_\mu((\Gamma^k(i,x,y,X_i,Y_i))^2) \leq C\,h_i^2\,(b_{1,i}b_{2,i})^{(d/p)-2d}. \tag{3.17}$$

It follows

$$\frac{h_n b_{1,n}^d b_{2,n}^d}{n}\sum_{i=1}^n \mathrm{Var}_\mu(\Gamma^k(i,x,y,X_i,Y_i)) \ \leq\ C\,\frac{h_n b_{1,n}^d b_{2,n}^d}{n}\sum_{i=1}^n h_i^2\,(b_{1,i}^d b_{2,i}^d)^{\frac{1}{p}-2},$$

the latter going to 0 thanks to (6).

Let us now compute the covariances.

One has thanks to stationarity and mixing inequality (2.2) (remark that $b_{1,i}^d b_{2,i}^d \Gamma^k(i,.)$ is bounded)

$$\frac{b_{1,n}^d b_{2,n}^d\,h_n}{n}\sum_{i\neq l}\mathrm{Cov}_\mu\left(\Gamma^k(i,x,y,X_i,Y_i),\Gamma^k(l,x,y,X_l,Y_l)\right) \leq$$

$$\begin{aligned}
&\leq C\,\frac{b_{1,n}^d b_{2,n}^d\, h_n}{n}\sum_{i=1}^{n}\sum_{j=1}^{i-1}\min\left(1,\frac{\rho^{i-j/2}}{b_{1,i}^d b_{2,i}^d b_{1,j}^d b_{2,j}^d}\right)\\
&\leq Ch_n\sum_{j=1}^{n-1}\min\left(b_{1,n}^d b_{2,n}^d,\frac{\rho^{j/2}}{b_{1,n}^d b_{2,n}^d}\right)\\
&\leq C\,h_n\left(k_n b_{1,n}^d b_{2,n}^d,\frac{\rho^{k_n/2}}{b_{1,n}^d b_{2,n}^d}\right)
\end{aligned}$$

which goes to 0 for an appropriate choice of k_n. □

Putting together the previous two theorems, we may deduce exactly as in Cattiaux et al. (2014-b) proof of theorem 4.1 the following

Corollary 3.18. *Assume that all the assumptions of Theorem 3.5 and Theorem 3.12 are satisfied. Then, defining $\tilde{b}_{r,n}=\tilde{H}_{r,n}/\tilde{p}_{r,n}$ we have in the stationary regime*

$$\sqrt{nb_{1,n}^d b_{2,n}^d h_n}\left(\tilde{b}_{r,n}(x,y)-b(x,y)\right)\xrightarrow[n\to+\infty]{\mathcal{D}}\sigma\mathcal{N}\left(0,\ \frac{\theta}{3p_s(x,y)}(\int K^2(s,t)dsdt)\,Id\right).$$

Remark 3.19. In the setting of Remark 3.13 the quasi optimal rate is still $n^{1/3}$. ♢

4. Estimation of the invariant density and of the drift term in the stationary regime: partial observation case.

We turn to the study of the estimators $\hat{p}_{r,n}$ and $\hat{b}_{r,n}$ defined in (1.3) and (1.4).

4.1. **Estimation of the invariant density.**

If we can only observe the position process, we have to consider instead of Y_i its natural approximation $\frac{X_{i+h_i}-X_i}{h_i}$.
Recall our estimator

$$\hat{p}_{r,n}(x,y):=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{b_{1,i}^d b_{2,i}^d}K\left(\frac{x-X_i}{b_{1,i}},\frac{y-\frac{X_{i+h_i}-X_i}{h_i}}{b_{2,i}}\right).$$

Theorem 4.1. *Assume that $\mathcal{H}_1$ (or $\mathcal{H}_2$) and $\mathcal{H}_K$ are fulfilled. Assume in addition that conditions (1a) up to (6a) are satisfied, where (1a) up to (4a) are given in Theorem 3 and*

(5a) $\lim_{n\to+\infty}\sqrt{nb_{1,n}^d b_{2,n}^d}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\sqrt{h_i}}{b_{2,i}^{d+1}}\right)=0$,

(6a) *for some $p>1$,* $\lim_{n\to+\infty}\sqrt{nb_{1,n}^d b_{2,n}^d}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{h_i}{b_{2,i}^{d+1}b_{1,i}^{d-(d/p)}}\right)=0$.

Then in the stationary regime

$$\sqrt{nb_{1,n}^d b_{2,n}^d}\left(\hat{p}_{r,n}(x,y)-p_s(x,y)\right)\xrightarrow[n\to+\infty]{\mathcal{D}}\mathcal{N}\left(0,\ \Sigma\,p_s(x,y)\int K^2(s,t)dsdt\right).$$

Proof. We want to evaluate the expectation of the difference :

$$\begin{aligned} & \sqrt{nb_{1,n}^d b_{2,n}^d}\,(\hat{p}_s(x,y)-\tilde{p}_s(x,y)) \qquad (4.2)\\ = & \frac{\sqrt{b_{1,n}^d b_{2,n}^d}}{\sqrt{n}}\sum_{i=1}^n \frac{1}{b_{1,i}^d b_{2,i}^d}\left(K\left(\frac{x-X_i}{b_{1,i}},\frac{y-\frac{X_{i+h_i}-X_i}{h_i}}{b_{2,i}}\right)-K\left(\frac{x-X_i}{b_{1,i}},\frac{y-Y_i}{b_{2,i}}\right)\right),\\ = & \frac{\sqrt{b_{1,n}^d b_{2,n}^d}}{\sqrt{n}}\sum_{i=1}^n \frac{1}{b_{1,i}^d b_{2,i}^d}\,A_i\,. \end{aligned}$$

We will closely follow Step 2 in the proof of Theorem 3.3 of Cattiaux et al. (2014-a).
Introduce $M_u=\frac{1}{h}\int_i^u (Y_s-Y_i)ds$, defined for $i\le u\le i+h_i$.
Then we may write

$$A_i \;=\; \frac{-1}{h_i b_{2,i}}\int_i^{i+h_i}\nabla_y K\left(\frac{x-X_i}{b_{1,i}},\frac{y-Y_i-M_u}{b_{2,i}}\right).(Y_u-Y_i)du\,.$$

Arguing as in the referred proof we can show that for all conjugate pair (p,q) with $1<p<+\infty$,

$$\mathbb{E}_\mu(|A_i|)\le \frac{C}{h_i b_{2,i}}\left(h_i^{3/2}\,b_{1,i}^d+b_{1,i}^{d/p}\,h_i^2\right).$$

Hence

$$\begin{aligned} \mathbb{E}_\mu\left(\sqrt{nb_{1,n}^d b_{2,n}^d}\,|\hat{p}_{r,n}(x,y)-\tilde{p}_{r,n}(x,y)|\right) &\le \frac{\sqrt{b_{1,n}^d b_{2,n}^d}}{\sqrt{n}}\sum_{i=1}^n \frac{1}{b_{1,i}^d b_{2,i}^d}\,\mathbb{E}_\mu(|A_i|)\\ &\le \frac{C\sqrt{b_{1,n}^d b_{2,n}^d}}{\sqrt{n}}\sum_{i=1}^n \frac{h_i^{3/2}\,b_{1,i}^d+b_{1,i}^{d/p}\,h_i^2}{b_{1,i}^d b_{2,i}^{d+1}h_i}\,. \end{aligned}$$

It follows that $\sqrt{nb_{1,n}^d b_{2,n}^d}\,(\hat{p}_{r,n}(x,y)-\tilde{p}_{r,n}(x,y))$ goes to 0 in $\mathbb{L}^1$ as soon as (5a) and (6a) are satisfied. The proof follows from Slutsky's lemma and Theorem 3. □

Remark 4.3. Assume that $b_{j,n}=n^{-\alpha_j}$ and $h_n=n^{-\beta}$. For (1a) up to (4a) to be satisfied, we have seen in Remark 3.6 that the condition

$$d(\alpha_1+\alpha_2)<1<d(\alpha_1+\alpha_2)+2(m+1)\min(\alpha_1,\alpha_2) \qquad (4.4)$$

is sufficient and in Remark 3.7 that we can choose m as large as we want, so that the previous amounts to $d(\alpha_1+\alpha_2)<1$.
Since we can choose p as close to 1 as we want, we easily see that (6a) is automatically satisfied as soon as (5a) is satisfied. The latter implies

$$1+\alpha_2(d+2)<\beta+d\alpha_1\,. \qquad (4.5)$$

It follows that, if we are only interested in estimating the invariant density, we may take β large enough for the latter to be satisfied. But of course we are interested in the simultaneous estimation of both the invariant density and the drift term, so that (4.5) will become some restrictive condition, we shall study in more details at the end of the next subsection. ♢

4.2. **Estimation of the drift.**

Introduce

$$\hat{H}_{r,n}(x,y) := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{b_{1,i}^d\, b_{2,i}^d}\, K\left(\frac{x-X_i}{b_{1,i}}, \frac{y-\frac{X_{i+h_i}-X_i}{h_i}}{b_{2,i}}\right) \Delta_2(X_{i+2h_i}, h_i)\,. \tag{4.6}$$

We shall first prove the analogue of Theorem 3.12.

Theorem 4.7. *Assume that $\mathcal{H}_1$ (or $\mathcal{H}_2$) and $\mathcal{H}_K$ are fulfilled. Assume in addition that conditions (1b) up to (8b) are satisfied, where (1b) up to (6b) are given in Theorem 3.12 and*

(7b) *for some $\varepsilon \geq 1$, $\lim_{n\to+\infty} \sqrt{nh_n b_{1,n}^d b_{2,n}^d}\left(\frac{1}{n}\sum_{i=1}^n \frac{\sqrt{h_i}}{b_{2,i}^{d+1} b_{1,i}^{d\varepsilon/(1+\varepsilon)}}\right) = 0$,*

(8b) *for some $\varepsilon > 0$ and some $r > 1+\varepsilon$,*

$$\lim_{n\to+\infty} \sqrt{nh_n b_{1,n}^d b_{2,n}^d}\left(\frac{1}{n}\sum_{i=1}^{n} \frac{h_i}{b_{2,i}^{d+1} b_{1,i}^{d(r-1)/r}}\right) = 0\,.$$

Then in the stationary regime

$$\sqrt{nh_n b_{1,n}^d b_{2,n}^d}\left(\hat{H}_{r,n}(x,y) - p_s(x,y)b(x,y)\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \sigma\mathcal{N}\left(0,\, \frac{\theta}{3}\, p_s(x,y)\Big(\int K^2(s,t)dsdt\Big)\, Id\right)\,.$$

Proof. Starting from Theorem 3.12, as in the previous subsection, it remains to consider D_n defined by

$$D_n := \sqrt{nh_n b_{1,n}^d b_{2,n}^d}\left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{b_{1,i}^d\, b_{2,i}^d}\, A_i\, \Delta_2(X_{i+2h_i}, h_i)\right)\,,$$

where

$$A_i = \left(K\left(\frac{x-X_i}{b_{1,i}}, \frac{y-Y_i}{b_{2,i}}\right) - K\left(\frac{x-X_i}{b_{1,i}}, \frac{y-\frac{X_{i+h_i}-X_i}{h_i}}{b_{2,i}}\right)\right)\,,$$

as in the previous subsection.
As before we decompose $\Delta_2(X_{i+2h_i}, h_i) = \frac{1}{h_i^2}(I_i + \mathfrak{M}_i)$, I_i being the bounded variation part and $\mathfrak{M}_i$ the martingale part. More precisely we define

$$\mathfrak{M}_i = \sigma\left(\int_{i+2h_i}^{i+3h_i} (W_s - W_{i+2h_i})ds + \int_{i+h_i}^{i+2h_i} (W_{i+2h_i} - W_s)ds\right)\,,$$

$$I_i = \int_{i+2h_i}^{i+3h_i}\int_{i+2h_i}^{t} g(X_s,Y_s)dsdt + \int_{i+h_i}^{i+2h_i}\int_{t}^{i+2h_i} g(X_s,Y_s)dsdt\,.$$

We want now to control

$$M_n = \mathbb{E}_\mu\left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i^2\, b_{1,i}^d\, b_{2,i}^d}|A_i||I_i|\right) \text{ and } N_n = \mathbb{E}_\mu\left(\frac{1}{n}\left|\sum_{i=1}^{n} \frac{1}{h_i^2 b_{1,i}^d\, b_{2,i}^d}\, A_i\, \mathfrak{M}_i\right|\right)\,.$$

For the first term, we use Hölder inequality for some $\varepsilon > 0$,

$$\mathbb{E}_\mu(|A_i\, I_i|) \leq \mathbb{E}_\mu\left(|A_i|^{1+\varepsilon}\right)^{1/(1+\varepsilon)}\, \mathbb{E}_\mu\left(|I_i|^{(1+\varepsilon)/\varepsilon}\right)^{\varepsilon/(1+\varepsilon)}\,.$$

Similarly to what we have done in the previous proof we may show on one hand that

$$\mathbb{E}_\mu(|A_i|^{1+\varepsilon}) \le \frac{C(\varepsilon)}{h_i\, b_{2,i}^{1+\varepsilon}} \left(h_i^{(3+\varepsilon)/2} b_{1,i}^d + h_i^{2+\varepsilon} b_{1,i}^{d/p}\right). \tag{4.8}$$

On the other hand, according to the proof in Cattiaux et al. (2014-b) (see formula (7.7) therein), one has

$$\mathbb{E}_\mu\left(|I_i|^{(1+\varepsilon)/\varepsilon}\right) \le C(\varepsilon,u,v)\, h_i^{2(1+\frac{1}{\varepsilon})}.$$

Finally,

$$\begin{aligned}
\mathbb{E}_\mu(|A_i\, I_i|) &\le C(\varepsilon,u,v)\, \frac{h_i^{\frac{1+2\varepsilon}{1+\varepsilon}}}{b_{2,i}} \left(h_i^{(3+\varepsilon)/2} b_{1,i}^d + h_i^{2+\varepsilon} b_{1,i}^{d/p}\right)^{\frac{1}{1+\varepsilon}} \\
&\le C(\varepsilon,u,v) \left(\frac{h_i^{5/2} b_{1,i}^{d/(1+\varepsilon)}}{b_{2,i}} + \frac{h_i^3 b_{1,i}^{d/p(1+\varepsilon)}}{b_{2,i}} \right).
\end{aligned}$$

Hence $\sqrt{nh_n b_{1,n}^d b_{2,n}^d}\, M_n \to 0$ according to (7b) and (8b).

It remains now to bound N_n.
To this end we compute the expectation of the square. But the terms $A_i\, \mathfrak{M}_i$ are centered and uncorrelated. It follows (similarly to Cattiaux et al. (2014-b)) that

$$\begin{aligned}
N_n^2 &\le \frac{C}{n^2} \sum_{i=1}^n \frac{1}{h_i^4 b_{1,i}^{2d} b_{2,i}^{2d}}\, h_i^3\, \mathbb{E}_\mu(|A_i|^2) \\
&\le \frac{C}{n^2} \sum_{i=1}^n \frac{1}{h_i^4 b_{1,i}^{2d} b_{2,i}^{2d}}\, h_i^3\, \frac{1}{h_i\, b_{2,i}^2} \left(h_i^2 b_{1,i}^d + h_i^3 b_{1,i}^{d/p}\right)
\end{aligned}$$

where we have used (4.8) with $\varepsilon = 1$, and finally

$$N_n \le \frac{C}{n} \left(\sum_{i=1}^n \frac{1}{b_{1,i}^{d/2} b_{2,i}^{d+1}} + \sum_{i=1}^n \frac{\sqrt{h_i}}{b_{2,i}^{d+1} b_{1,i}^{d(1-(1/2p))}} \right).$$

Again $\sqrt{nh_n b_{1,n}^d b_{2,n}^d}\, N_n \to 0$ according to (7b) and (8b). □

Remark 4.9. If $\lim_{n\to+\infty} \sqrt{nh_n b_{1,n}^d b_{2,n}^d} \left(\frac{1}{n}\sum_{i=1}^n \frac{h_i^{\kappa_0}}{b_{2,i}^{\kappa_2} b_{1,i}^{\kappa_1}}\right) = 0$ for some positive κ_j's, then the same holds for all η_j's with $\eta_0 \ge \kappa_0$, $\eta_1 \le \kappa_1$ and $\eta_2 \le \kappa_2$. Hence it is easy to see that (7b) and (8b) amounts to
(7-8-b') there exists $r > \frac{1}{2}$ such that

$$\lim_{n\to+\infty} \sqrt{nh_n b_{1,n}^d b_{2,n}^d} \left(\frac{1}{n} \sum_{i=1}^n \frac{\sqrt{h_i}}{b_{2,i}^{d+1} b_{1,i}^{rd}} \right) = 0.$$

Now assume as usual that $b_{j,n} = n^{-\alpha_j}$ and $h_n = n^{-\beta}$ for some positive α_j and β.
(7-8-b') is satisfied as soon as

$$1 + (d+2)\alpha_2 < 2\beta.$$

In particular, for all the conditions in Theorem 4.7 and Theorem 4.1 to be satisfied it is sufficient that both conditions (3.14) and $1 + (d+2)\alpha_2 < \beta + \beta \wedge d\alpha_1$ are fulfilled.

Unfortunately, as for our estimators in Cattiaux et al. (2014-b), both conditions are incompatible.
If we do not want to multiply the observations, we have to choose the same discretization step h_n in both $\hat{H}_{r,n}$ and $\hat{p}_{r,n}$. But as we did in Cattiaux et al. (2014-b) Theorem 4.1 and proposition 4.4, we may choose different bandwidths for both estimators, denoted by b^p and b^H for respectively the density estimator and the drift estimator.
The required conditions are then, on one hand

$$d(\alpha_1^p + \alpha_2^p) < 1 \quad \text{and} \quad 1 + \alpha_2^p(d+2) < \beta + \beta \wedge d\alpha_1^p\,, \tag{4.10}$$

and, on the other hand

$$\beta + d(\alpha_1^H + \alpha_2^H) < 1 < 3\beta + d(\alpha_1^H + \alpha_2^H) \quad \text{and} \quad 1 + \alpha_2^H(d+2) < 2\beta + d\alpha_1^H\,. \tag{4.11}$$

We also want $\beta + d(\alpha_1^H + \alpha_2^H)$ to be as small as possible. But it is easily seen that this quantity has to be larger than $1/2$. We will describe a typical situation in the next Corollary. ♢

According to the previous discussion we have

Corollary 4.12. *Assume that $\mathcal{H}_1$ (or $\mathcal{H}_2$) and $\mathcal{H}_K$ are fulfilled.*
Assume in addition that $b_{j,n}^H = n^{-\alpha_j^H}$ (resp. $b_{j,n}^p = n^{-\alpha_j^p}$) and $h_n = n^{-\beta}$ for some positive α_j's and β. Denote by $\hat{H}_{r,n}$ (resp. $\hat{p}_{r,n}$) the estimators built with the bandwidths $b_{j,n}^H$ (resp. $b_{j,n}^p$) and the discretization step h_n. Finally assume that all the parameters satisfy (4.10) *and* (4.11) *and that*

$$m\,\min(\alpha_1^H, \alpha_2^H, \alpha_1^p, \alpha_2^p) > 1\,.$$

Then, defining $\hat{b}_{r,n} = \hat{H}_{r,n}/\hat{p}_{r,n}$ we have in the stationary regime

$$\sqrt{nb_{1,n}^d b_{2,n}^d h_n}\left(\hat{b}_{r,n}(x,y) - b(x,y)\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \sigma\mathcal{N}\left(0,\, \frac{\theta}{3p_s(x,y)}(\int K^2(s,t)dsdt)\, Id\right)\,,$$

where $\theta = 1/(1+\beta+d(\alpha_1^H+\alpha_2^H))$.

Remark 4.13. In the previous Corollary we may choose any $\beta > 1/2$, then $d\alpha_1^p > 1/2$, then $d\max(\alpha_1^H, \alpha_2^H, \alpha_2^p)$ sufficiently small and finally m large enough for all the assumptions of the previous Corollary to be satisfied. It yields a quasi optimal rate of order $n^{1/4}$.

Notice that if we want to estimate simultaneously p_s and b, we may use two estimators of p_s, say $\hat{p}_{r,n}^p$ and $\hat{p}_{r,n}^H$, both with $h_n = n^{-\beta}$ for some $\beta > 1/2$, the first one with bandwidths α_j^p, the second one with α_j^H and look at $(\hat{p}_{r,n}^H, (\hat{H}_{r,n}/\hat{p}_{r,n}^p))$ simultaneously. ♢

5. Additional results and comments on the previous sections.

5.1. **Non stationary case.** This subsection will be very short and can be reduced to the following sentence :

"All the results of the two previous sections are still true if the initial distribution is either a Dirac mass or is absolutely continuous with respect to μ."

Indeed the situation here is very simple: if we replace $\sum_{i=1}^n$ by $\sum_{i=l_n}^n$ with $(l_n/n) \to 0$, we obtain exactly the same results as the first l_n terms go to 0 starting from any initial

distribution. Next we have to compare the law of Z_{l_n} to μ. But thanks to geometric ergodicity (or exponential mixing see Proposition 2.1), it is known that the total variation distance between both is less than $C\rho^{l_n}$ provided the initial distribution satisfies the conditions we have mentioned (see e.g. Bakry et al. (2008)). The conclusion is then straightforward using Markov property and $l_n = c\log(n)$.

5.2. **Comparison with existing results.** The rate of convergence for the recursive estimators are much better than the ones we obtain in Cattiaux et al. (2014-a) and Cattiaux et al. (2014-b) (see e.g Proposition 4.4 in Cattiaux et al. (2014-b)). For the density the main reason is that we do not need the small time bounds for the transition kernel as in Cattiaux et al. (2014-a) (see Konakov et al. (2010), Cattiaux (1990), Cattiaux (1986)), for the unnormalized drift estimator $\hat{H}_{r,n}$ the gain is not as important.
In comparison with the existing literature on recursive estimation for the density (see e.g. Amiri (2010) for a review), the results in Section 3 are similar. Of course in many quoted references the authors also studied almost sure results in the spirit of the log-log law. Again see Amiri (2010) for a review.
The assumption made in all these references is that the transition kernel $p_{s,t}(.,.)$ satisfies

$$\sup_{|t-s|\geq 1} \| p_{s,t}(.,.) - p_s \otimes p_s \|_\infty < +\infty \,.$$

In our situation, we do not know about such uniform bound, though it is presumably true.

6. Estimation of the variance.

We turn to the estimation of σ. To this end we merely follow Section 4 in Cattiaux et al. (2014-c), though here we only consider the σ constant case, which introduces various simplifications.
Recall that

$$\hat{\sigma}_{r,n} := \frac{3}{2n}\sum_{i=1}^{n} \frac{1}{h_i}\,\Delta_2(X_{i+h_i},h_i) >< \Delta_2(X_{i+h_i},h_i)\,.$$

Theorem 6.1. *Assume that $\mathcal{H}_1$ (or $\mathcal{H}_2$) and $\mathcal{H}_K$ are fulfilled. Also assume that σ is symmetric (i.e. we take the symmetric square root of $\sigma^*\sigma$). Assume in addition that, as $n \to +\infty$,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h_i \;\to\; 0\,. \tag{6.2}$$

Then, in the stationary regime or in the situation of subsection 5.1

$$\sqrt{n}\,(\hat{\sigma}_{r,n} - \sigma^2) \xrightarrow[n\to+\infty]{\mathcal{D}} \sigma\mathcal{N}_{d,d}\sigma\,,$$

where $\mathcal{N}_{d,d}$ is a symmetric gaussian random matrix whose entries $\mathcal{N}_{d,d}^{k,l}$ are $\mathcal{N}(0, 1+\delta_{k,l})$, all entries for $k \geq l$ being independent.

Remark 6.3. Condition (6.2) is satisfied when $h_n = n^{-\beta}$ with $\beta > 1/2$. ♢

Proof. In the sequel the indices (k,l) denote respectively the index of the row and of the column of a matrix.

Using Ito's formula we have

$$
\begin{aligned}
h^2\ (\Delta_2(X_t,h) >< \Delta_2(X_t,h))_{k,l} = \\
&= \int_{t-h}^{t+h} (h-|u-t|)\,(H_u^k\,(\sigma dW_u)^l + H_u^l\,(\sigma dW_u)^k) + \\
&\quad + \int_{t-h}^{t+h} (h-|u-t|)\,(H_u^k\,b^l(Z_u) + H_u^l\,b^k(Z_u))du \\
&\quad + \int_{t-h}^{t+h} (h-|u-t|)^2\,\sigma_{k,l}^2\,du\,,
\end{aligned}
\tag{6.4}
$$

with

$$
H_u = \int_{t-h}^{u} (h-|s-t|)\,(\sigma dW_s + b(Z_s)ds)\,.
$$

As explained in Cattiaux et al. (2014-c) formula (4.5) applied in our situation, it holds for all $j\in\mathbb{N}$ and s between $t-h$ and $t+h$,

$$
\mathbb{E}_\mu(|H_s|^{2j}) \le C(j)\,(s-(t-h))^j\,h^{2j}\,. \tag{6.5}
$$

Now we can write

$$
\begin{aligned}
\hat\sigma_{r,n} - \sigma^2 = \\
&= \frac{3}{2n}\sum_{i=1}^n \frac{1}{h_i}\left\{\Delta_2(X_{i+h_i},h_i) >< \Delta_2(X_{i+h_i},h_i) - \frac{1}{h_i^2}\int_i^{i+2h_i}(h_i-|s-(i+h_i)|)^2\sigma^2 ds\right\} \\
&= \frac{3}{2n}\sum_{i=1}^n (SS_i + SV_i + VV_i)
\end{aligned}
$$

where

$$
\begin{aligned}
SS_i^{k,l} &= \frac{1}{h_i^3}\int_i^{i+2h_i}(h_i-|u-i-h_i|)\int_i^u (h_i-|s-i-h_i|)((\sigma dW_s)^k\,(\sigma dW_u)^l + (\sigma dW_s)^l\,(\sigma dW_u)^k) \\
SV_i^{k,l} &= \frac{1}{h_i^3}\int_i^{i+2h_i}(h_i-|u-i-h_i|)\int_i^u (h_i-|s-i-h_i|)((\sigma dW_s)^k\,b^l(Z_u)du + (\sigma dW_s)^l b^k(Z_u)du) \\
&\quad + \frac{1}{h_i^3}\int_i^{i+2h_i}(h_i-|u-i-h_i|)\int_i^u (h_i-|s-i-h_i|)((\sigma dW_u)^k\,b^l(Z_s)ds + (\sigma dW_u)^l b^k(Z_s)ds) \\
VV_i^{k,l} &= \frac{1}{h_i^3}\int_i^{i+2h_i}(h_i-|u-i-h_i|)\int_i^u (h_i-|s-i-h_i|)(b^k(Z_s)ds\,b^l(Z_u)du + b^l(Z_s)ds\,b^k(Z_u)du)\,.
\end{aligned}
$$

It is immediate that the random matrices $(M_i = \frac{3}{2}\,\sigma^{-1}SS_i\sigma^{-1})_i$ are independent, with centered entries such that $\mathrm{Var}_\mu(M_i^{k,l}) = 1+\delta_{k,l}$. Hence according to the usual Central Limit Theorem,

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^n M_i \xrightarrow[n\to+\infty]{\mathcal{D}} \sigma\mathcal{N}_{d,d}\sigma\,.
$$

It remains to prove that the remaining terms go to 0 in $\mathbb{L}^1$.
First, it is not hard to see that

$$\mathbb{E}_\mu(|VV_i^{k,l}|) \leq C\, h_i$$

so that $\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}_\mu(|VV_i^{k,l}|) \to 0$ according to (6.2).
Next, to control SV_i we are led to control two types of terms: either martingale terms

$$N_i^{k,l} = \frac{1}{h_i^3} \int_i^{i+2h_i} (h_i - |u-i-h_i|)\, \beta_i^l(u)\, (\sigma dW_u)^k$$

or bounded variation terms

$$B_i^{k,l} = \frac{1}{h_i^3} \int_i^{i+2h_i} (h_i - |u-i-h_i|)\, \gamma_i^l(u)\, b^k(Z_u) du\,.$$

The martingale terms are centered and to prove the convergence of their normalized sum to 0 in $\mathbb{L}^1(\mu)$, it is enough to prove the convergence to 0 of the variance. But for $i \neq j$ the covariance $\mathrm{Cov}_\mu(N_i^{k,l}, N_j^{k,l}) = 0$ so that the variance

$$\mathrm{Var}_\mu\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n N_i^{k,l}\right) \leq \frac{C}{n} \sum_{i=1}^n \frac{1}{h_i^6} \int_i^{i+2h_i} (h_i - |u-i-h_i|)^2\, \mathbb{E}_\mu((\beta_i^l(u))^2)\, du \leq \frac{C}{n} \sum_{i=1}^n h_i$$

goes to 0 provided h_n goes to 0, hence we get convergence to 0 in $\mathbb{L}^1$. Here we have used

$$\mathbb{E}_\mu\left[\left(\int_i^{i+2h_i} \theta(u) du\right)^2\right] \leq C\, h_i\, \mathbb{E}_\mu\left(\int_i^{i+2h_i} \theta^2(u) du\right)\,.$$

For the bounded variation term we use the same trick as in Cattiaux et al. (2014-c) writing $b^k(Z_u) = (b^k(Z_u) - b^k(Z_i)) + b^k(Z_i)$. The terms

$$B_{i,1}^{k,l} = \frac{1}{h_i^3} \int_i^{i+2h_i} (h_i - |u-i-h_i|)\, \gamma_i^l(u)\, b^k(Z_i) du\,,$$

are thus centered and satisfy $\mathrm{Cov}_\mu(B_{i,1}^{k,l}, B_{j,1}^{k,l}) = 0$ for $i \neq j$. As before it follows

$$\mathrm{Var}_\mu\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n B_{i,1}^{k,l}\right) \leq \frac{C}{n} \sum_{i=1}^n \frac{1}{h_i^5} \int_i^{i+2h_i} (h_i - |u-i-h_i|)^2\, \mathbb{E}_\mu((\gamma_i^l(u) b^k(Z_i))^2)\, du\,.$$

But

$$\mathbb{E}_\mu((\gamma_i^l(u) b^k(Z_i))^2) \;=\; \mathbb{E}_\mu\left((b^k(Z_i))^2 \int_i^u (h_i - |s-i-h_i|)^2\, c(\sigma) ds\right) \leq C h_i^3\,,$$

so that again $\mathrm{Var}_\mu\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n B_{i,1}^{k,l}\right) \leq \frac{C}{n} \sum_{i=1}^n h_i$ goes to 0, and we get convergence to 0 in $\mathbb{L}^1$.
It remains to look at $\frac{1}{\sqrt{n}}\, \mathbb{E}_\mu\left(\sum_{i=1}^n |B_{i,2}^{k,l}|\right)$ where

$$B_{i,2}^{k,l} = \frac{1}{h_i^3} \int_i^{i+2h_i} (h_i - |u-i-h_i|)\, \gamma_i^l(u)\, (b^k(Z_u) - b^k(Z_i)) du\,.$$

But, according to what we have done before

$$\mathbb{E}_\mu(\gamma_i^l(u)\, (b^k(Z_u) - b^k(Z_i))) \leq (\mathbb{E}_\mu((\gamma_i^l(u))^2))^{1/2}\, (\mathbb{E}_\mu((b^k(Z_u) - b^k(Z_i))^2))^{1/2} \leq C h_i^{3/2} h_i^{1/2}\,.$$

It follows that

$$\frac{1}{\sqrt{n}}\,\mathbb{E}_\mu\left(\sum_{i=1}^{n}|B_{i,2}^{k,l}|\right)\leq\frac{1}{\sqrt{n}}\sum_{i=1}^{n}h_i\,,$$

which goes to 0 thanks to (6.2). □

References

A. Amiri (2009). Sur une famille paramétrique d'estimateurs séquentiels de la densité pour un processus fortement mélangeant. *C. R. Acad. Sci. Paris, Ser. I.* **347**, p. 309-314.

A. Amiri (2010). Estimateurs fonctionnels récursifs et leurs applications à la prévision. *PHD Thesis. Université d'Avignon.*

D. Bakry, P. Cattiaux, and A. Guillin (2008). Rate of convergence for ergodic continuous Markov processes : Lyapunov versus Poincaré. *J. Func. Anal.* **254**, p. 727-759.

J.M. Bardet, P. Doukhan, G. Lang and N. Ragache (2008). Dependent Lindeberg Central Limit Theorem and some applications. *E*SAIM: Probability and Statistics **12**, p. 154-172.

P. Cattiaux (1986). Hypoellipticité et hypoellipticité partielle pour les diffusions avec une condition frontière. *Ann. Institut Henri Poincaré.* **22**, p. 67-112.

P. Cattiaux (1990). Calcul stochastique et opérateurs dégénérés du second ordre. I Résolvantes, théorème de Hörmander et applications. *Bull. Sc. Math.* **114** p. 421-462.

P. Cattiaux, D. Chafaï and A. Guillin (2012). Central limit theorems for additive functionals of ergodic Markov diffusions processes. *ALEA Lat. Am. J. Prob. Math. Stat.* **9 (2)**, p. 337-382.

P. Cattiaux, J. R. León and C. Prieur (2014-a). Estimation for Stochastic Damping Hamiltonian Systems under Partial Observation. I. Invariant density. *Stochastic Processes Appl.* **124 (3)**, p. 1236-1260.

P. Cattiaux, J. R. León and C. Prieur (2014-b). Estimation for Stochastic Damping Hamiltonian Systems under Partial Observation. II. Drift term. To appear in ALEA.

P. Cattiaux, J. R. León and C. Prieur (2014-c). Estimation for Stochastic Damping Hamiltonian Systems under Partial Observation. III. Diffusion term. `http://hal.archives-ouvertes.fr/hal-01044611`

V. Konakov, S. Menozzi and S. Molchanov (2010). Explicit parametrix and local limit theorems for some degenerate diffusion processes. *Ann. Institut Henri Poincaré.* **46 (4)**, p. 908-923.

Liang, H.Y. and Baek, J. (2004). Asymptotic normality of recursive density estimates under some dependence assumptions. *Metrika.* **60**, p. 155-166.

E. Masry (1986). Recursive probability density estimation for weakly dependent stationary processes. *IEEE Trans. Inform. Theory.* **32 (2)**, p. 254-267.

E. Masry (1987). Almost sure convergence of recursive density estimators for stationary mixing processes. *Statist. Probab. Lett.* **5**, p. 249-254.

E. Masry (1989). Nonparametric estimation of conditional probability density and expectation of stationary processes; strong consistency and rates. *Stochastic Process. Appl.* **32**, p. 109-127.

E. Masry and L. Györfi (1987). Strong consistency and rates for recursive probability density estimators of stationary processes. *J. Multivariate Anal.* **22**, p. 79-93.

A. Samson and M. Thieullen (2012). Contrast estimator for completely or partially observed hypoelliptic diffusion. *Stochastic Processes Appl.* **122 (7)**, p. 2521-2552.

E.J. Wegman and H.I. Davis (1979). Remarks on some recursive estimators of a probability density. *Ann. Statist.* **7**, p. 316-327.

L. Wu (2001). Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic Processes Appl.* **91**, p. 205-238.

Patrick CATTIAUX,, Institut de Mathématiques de Toulouse. Université de Toulouse., CNRS UMR 5219.,, 118 route de Narbonne, 31062 Toulouse cedex 09, France.
E-mail address: cattiaux@math.univ-toulouse.fr

José LEÓN,, Escuela de Matemática, Facultad de Ciencias, Universidad Central de Venezuela., A.P. 47197, Los Chaguaramos, Caracas 1041-A, Venezuela.
E-mail address: jose.leon@ciens.ucv.ve

Clémentine PRIEUR,, Université Joseph Fourier, LJK/MOISE BP 53., 38041 Grenoble Cedex, France.
E-mail address: clementine.prieur@imag.fr