



HAL
open science

Estimation for Stochastic Damping Hamiltonian Systems under Partial Observation. I. Invariant density.

Patrick Cattiaux, Jose R. Leon, Clémentine Prieur

► To cite this version:

Patrick Cattiaux, Jose R. Leon, Clémentine Prieur. Estimation for Stochastic Damping Hamiltonian Systems under Partial Observation. I. Invariant density.. Stochastic Processes and their Applications, 2014, 124 (3), pp.1236-1260. 10.1016/j.spa.2013.10.008 . hal-00739136v2

HAL Id: hal-00739136

<https://hal.science/hal-00739136v2>

Submitted on 4 Nov 2013

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

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

**ESTIMATION FOR STOCHASTIC DAMPING HAMILTONIAN
SYSTEMS UNDER PARTIAL OBSERVATION.
I. INVARIANT DENSITY.**

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 study the non-parametric estimation of the invariant density of some ergodic hamiltonian systems, using kernel estimators. The main result is a central limit theorem for such estimators under partial observation (only the positions are observed). The main tools are mixing estimates and refined covariance inequalities, the main difficulty being the strong degeneracy of such processes. This is the first paper of a series of at least two, devoted to the estimation of the characteristics of such processes: invariant density, drift term, volatility

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

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

1. INTRODUCTION.

Let $(Z_t := (X_t, Y_t) \in \mathbb{R}^{2d}, t \geq 0)$ 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 (non-zero) constant and W 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) for a detailed bibliography. We have chosen the terminology “damping

Date: October 29, 2013.

Hamiltonian systems” in reference to Wu. Such systems are also called “kinetic diffusions” by several authors.

The long time behavior of such models has been recently deeply studied, in particular in Villani (2009), since they are the basic examples of hypocoercive models. The Lyapunov approach of Wu, has been developed in Bakry, Cattiaux and Guillin (2008) in connection with hypocoercive models. Particular features are, on one hand, that these models are fully degenerate, but still hypoelliptic, on the other hand, that they still satisfy some coercivity property, but not in the usual sense, due again to full degeneracy. We shall explain both previous sentences in the next section.

Once the probabilistic picture is well understood, it is particularly relevant to build statistical tools for these models. In this paper we are starting a general program of non-parametric estimation for the characteristics of such processes, more precisely we focus on the non-parametric estimation of the invariant density. It is worth noticing that, except for some special situations, an explicit expression for this density is unknown.

This program seems to be new, up to our knowledge. Actually some works have already been done, in the hypo-elliptic context, but in a parametric framework. We refer to the recent work by Samson and Thieullen (2012) and the bibliography therein.

The main result of the present paper is contained in Theorem 3.3 in the stationary case and extended to any initial starting point in Theorem 6.2.

It reads as follows : if p_s denotes the invariant density (see the next section for its existence), then one can find a discretization step h_n , bandwidth $b_{1,n}$ and $b_{2,n}$ and kernels K such that, defining the estimator

$$\hat{p}_s(x, y) := \frac{1}{nb_{1,n}b_{2,n}} \sum_{i=1}^n K \left(\frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - \frac{X_{(i+1)h_n} - X_{ih_n}}{h_n}}{b_{2,n}} \right),$$

corresponding to partial observation, it holds, for all (x, y)

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

We also give explicit examples of allowed choices of h_n , $b_{1,n}$ and $b_{2,n}$ (see Remark 3.4 in the stationary case and the end of section 6 otherwise).

The proof lies on a new version of the Central Limit Theorem for triangular arrays of mixing sequences stated in Theorem 4.3, which is inspired by previous works by Rio, Doukhan and the third named author.

In order to apply this result, we need some upper estimates for the transition kernels of the process, both for *small times* and for *long times*. These estimates, as well as several properties of the process Z , are given in the first section. In particular, the full degeneracy of the infinitesimal generator, yields a non-usual behavior of the transition kernel recently obtained by Konakov, Menozzi and Molchanov (2010) and recalled in Theorem 2.10. We have to slightly extend their result to unbounded coefficients. This is done by using old ideas of the first named author. The long time behavior is connected, as previously said,

to hypo-coercivity.

With these two ingredients, the proof is merely standard but technical.

Actually, the conditions on the parameters h_n and $b_{i,n}$ obtained in Theorem 3.3 should be slightly improved, due to the fact that the explosion of the transition kernel holds *near* the diagonal, but not *on* the diagonal.

It is also interesting to notice that the study of coercivity made by Villani (2009) also uses both small and long time estimates.

We conclude the paper by some simulations on two models. The first one is an harmonic oscillator subject to noise and positive damping. The second one is a general Duffing oscillator as described in Wu (2001).

As we said at the beginning, this is the first part of a general program devoted to non-parametric estimation for these fully degenerate models. Estimation for the diffusion coefficient (volatility), the drift term, crossings ... will be done elsewhere.

Even if part of the results (the one concerned with the fully observed process) can be written for more general diffusion processes (see Remark 3.5), we deliberately decided to write all the main statements for the considered model.

2. THE MODEL AND ITS PROPERTIES.

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

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 ∇V have polynomial growth at infinity and

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

the latter being often called “drift condition”,

- (ii) the damping coefficient $c(x, y)$ is smooth and bounded, 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 $\frac{1}{2}(c_{ij}(x, y) + c_{ji}(x, y))_{1 \leq i, j \leq d}$,

These conditions ensure the existence of a Foster-Lyapunov function Ψ larger than 1 (and actually growing to infinity at infinity) satisfying

$$L\Psi \leq -\alpha\Psi + b\mathbb{1}_K$$

for some $\alpha > 0$ and some compact subset K . Here L denotes the infinitesimal generator

$$L = \frac{\sigma^2}{2} \Delta_y + y \nabla_x - (c(x, y)y + \nabla V) \cdot \nabla_y.$$

Hence, there is no explosion according to Khasminski test, and the process is positive recurrent with a unique invariant probability measure μ . Furthermore, if we 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 $1 \leq p \leq +\infty$.

In addition, there exist $D > 0$ and $\rho < 1$ such that for all z ,

$$\left| P_t f(z) - \int f d\mu \right| \leq D \sup_a \left(\frac{|f(a) - \int f d\mu|}{\Psi(a)} \right) \Psi(z) \rho^t. \quad (2.1)$$

Since Ψ is μ integrable (see Theorem 3.1 in Wu (2001) but this is a very general result for such Lyapunov functions), the previous pointwise convergence becomes a convergence in $\mathbb{L}^1(\mu)$. All these results are contained in Wu (2001) (see in particular Theorem 2.4, Theorem 3.1 and remark 3.2 therein), and follow from a general approach of recurrence via the use of Lyapunov functions described, in the diffusion case, in Down, Meyn and Tweedie (1995).

Notice that, if f is bounded and satisfies $\int f d\mu = 0$, we deduce from (2.1)

$$\int (P_t f)^2 d\mu \leq D \left(\int \Psi d\mu \right) \|f\|_\infty^2 \rho^t. \quad (2.2)$$

As explained in Bakry, Cattiaux and Guillin (2008), in particular Theorem 2.1, another possible way to describe this exponential convergence to equilibrium (under the same assumptions) is: for all bounded f with $\int f d\mu = 0$

$$\int (P_t^* f)^2 d\mu \leq D' \left(\int \Psi d\mu \right) \|f\|_\infty^2 \rho^t, \quad (2.3)$$

where P_t^* denotes the adjoint of P_t in $\mathbb{L}^2(\mu)$. This is a consequence of the convergence in total variation distance obtained in Down et al. (1995): for all z ,

$$\|P_t(z, \cdot) - \mu\|_{TV} \leq D'' \Psi(z) \rho^t, \quad (2.4)$$

where $P_t(z, \cdot)$ denotes the law at time t of the process starting from z . If Z_0 has distribution ν , the law at time t of the process is given by $P_t^* \nu$, and exponential convergence in total variation holds as soon as Ψ is ν integrable.

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 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, Cattiaux and Guillin (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

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

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)).

A still delicate feature is that in many situations, no explicit expression for the invariant measure μ is known. An important exception is the case when the matrix c is constant, and for simplicity equal to cId (for some $c > 0$). Indeed in this case, defining the Hamiltonian $H(x, y) = \frac{1}{2}|y|^2 + V(x)$, the unique invariant measure (up to a numerical constant factor) writes

$$\mu(dx, dy) = \exp\left(-\frac{2c}{\sigma^2}H(x, y)\right) dx dy.$$

Remark 2.5. Note that the invariant measure is not symmetric, so that P_t and P_t^* do not coincide. In particular the Dirichlet form

$$\mathcal{E}(f, g) = - \int Lf g d\mu = \int \nabla_y f \cdot \nabla_y g d\mu$$

does not satisfy the usual property in the symmetric ergodic situation

$$\left(\mathcal{E}(f, f) = 0 \text{ and } \int f d\mu = 0\right) \Rightarrow f = 0.$$

Hence μ cannot satisfy a Poincaré inequality with energy term given by $\mathcal{E}(f, f)$, though we have some exponential decay to equilibrium.

Still more surprising, under some stronger assumptions on V , Villani has shown that the exponential decay in (2.2) is still true for \mathbb{L}^2 functions f , implying that the constant D' is strictly larger than 1. For a precise statement of Villani's result see Theorem 6.1 in Bakry et al. (2008). \diamond

Finally, we will need some tail behaviour of μ . Actually a careful look at Wu (2001) formula (3.3) or Bakry et al. (2008) formula (6.4), show that the Lyapunov function Ψ satisfies the following property

$$\log(\Psi(x, y)) \geq C(|y|^2 + |x|) \quad \text{as } z \text{ goes to infinity, for some well chosen } C > 0.$$

It follows that μ admits some exponential moment, in particular all its polynomial moments are finite.

As remarked in Cattiaux, Chafai and Guillin (2011), a uniform \mathbb{L}^∞ - \mathbb{L}^2 decay is equivalent to some mixing property. Let us state the result

Proposition 2.6. *Assume that (2.2) and (2.3) are satisfied. Then, there exists some constant $C > 0$ such that : $\forall g, f \in \mathbb{L}^\infty(\mu), \forall t \geq 0$,*

$$|\text{Cov}_\mu(f(Z_t), g(Z_0))| \leq C \rho^{t/2} \left\| g - \int g d\mu \right\|_\infty \left\| f - \int f d\mu \right\|_\infty. \quad (2.7)$$

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

Proof. We give the proof for completeness. Assume that $\int f d\mu = \int g d\mu = 0$. Then

$$\begin{aligned} |\text{Cov}_\mu(f(Z_t), g(Z_0))| &= \int g(z) P_t f(z) \mu(dz) \\ &= \int P_{t/2}^* g P_{t/2} f d\mu, \end{aligned}$$

and it remains to apply Cauchy-Schwartz inequality and both (2.2) and (2.3). Now if F is measurable w.r.t. the filtration of the future, just take conditional expectation using the Markov property, to get the statement about mixing. \square

Actually this statement admits a converse: an exponential decay of such covariances implies (2.2) and (2.3) for some ad-hoc $\rho < 1$ (Cattiaux et al. (2011)).

Remark 2.8. Starting from Inequality (2.1), and noting that the Lyapunov function Ψ is μ integrable, one can also deduce that the sequence $(Z_k)_{k \in \mathbb{N}^*}$ is β -mixing with exponential rate.

2.2. Local properties, hypoellipticity. We turn to the study of the hypoellipticity property.

First, since the diffusion coefficient is constant, (1.1) is also written in Stratonovitch form and the generator L can be written in Hörmander form

$$L = \frac{\sigma^2}{2} \sum_{i=1}^d L_i^2 + L_0$$

where the vector fields L_j are defined by :

$$(1) \quad \text{for } 1 \leq i \leq d, L_i = \frac{\partial}{\partial y_i},$$

(2)

$$L_0 = \sum_{k=1}^d y_k \frac{\partial}{\partial x_k} - \sum_{k=1}^d \left((c(x, y)y)_k + \frac{\partial V}{\partial x_k} \right) \frac{\partial}{\partial y_k}.$$

It immediately follows that the Lie bracket

$$[L_i, L_0] = L_i L_0 - L_0 L_i = \frac{\partial}{\partial x_i} - \sum_{k=1}^d \frac{\partial((c(x, y)y)_k)}{\partial y_i} \frac{\partial}{\partial y_k},$$

so that the vector space spanned by $\{L_i, 1 \leq i \leq d; [L_i, L_0], 1 \leq i \leq d\}(z)$ is full (i.e. equal to \mathbb{R}^{2d}) at each z .

According to the famous theorem of the sum of squares of Hörmander, it follows that $\partial_t + L$ and its adjoint in the space of Schwartz distributions are hypoelliptic. As a consequence, for any z and any $t > 0$, the distribution $P_t(z, \cdot)$ of the process Z_t starting from z (i.e. $Z_0 = z$), has a smooth density $p_t(z, \cdot)$ with respect to Lebesgue measure.

Of course the same holds for the invariant measure, i.e.

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

with some smooth function p_s . One can relax the C^∞ assumption on the coefficient into a C^k assumption, for a large enough k , but this is irrelevant.

In the p.d.e. vocabulary, we are in a fully degenerate situation, i.e. brackets with the drift vector field are necessary to span the whole tangent space.

In the sequel we shall need more information on the density $p_t(z, \cdot)$, both for small and for large t 's. Once again, full degeneracy introduces some trouble.

Example 2.9. To understand what happens, let us consider a very simple gaussian situation. For $d = 1$ we consider the case where c and V are equal to 0. Then Z_t is a two dimensional gaussian vector with mean $(x_0 + y_0 t, y_0)$ and covariance matrix given via

$$\text{Var}(X_t) = \frac{t^3}{3}, \text{Var}(Y_t) = t, \text{Cov}(X_t, Y_t) = \frac{t^2}{2}.$$

In particular $p_t(z, z) \simeq c/t^2$ instead of the usual c/t for the brownian motion for instance. Of course this example does not enter the framework of this work, since this process is not positive recurrent.

Actually if we choose $V(x) = a|x|^2$ and c constant, Z_t is still a gaussian vector and one can show that the covariance matrix is similar to the previous one, in particular the behaviour of each term as t goes to 0 is the same (see e.g. Risken (1989) section 10.2.1).

◇

The previous behaviour is actually true in our very general situation, up to one restriction: it has been shown only for bounded, with bounded derivatives, coefficients. This is the main result in Konakov, Menozzi and Molchanov (2010):

Theorem 2.10. (Konakov, Menozzi and Molchanov (2010)). *Consider the following system*

$$\begin{aligned} dX_t &= Y_t dt \\ dY_t &= \sigma dW_t + b(X_t, Y_t) dt, \end{aligned} \tag{2.11}$$

where b is assumed to be smooth, bounded with bounded derivatives. Then for any initial point $z = (x, y)$ and any $t > 0$, the distribution of $Z_t = (X_t, Y_t)$ has a smooth density $q_t(z, \cdot)$ with respect to the Lebesgue measure, which satisfies the following gaussian upper bound : there exist positive constants C and C' depending on $b, \sigma, T > 0$ and the

dimension $2d$, such that for $0 < t < T$,

$$q_t(z, z') \leq C' \frac{1}{t^{2d}} \exp \left(-C \left[\frac{|y - y'|^2}{4t} + \frac{3 \left| x' - x - \frac{t(y+y')}{2} \right|^2}{t^3} \right] \right).$$

In addition, for some $t_0 > 0$, there exists $C'' > 0$ such that for any $0 < t < t_0$,

$$q_t((x, y), (x + ty, y)) \geq C'' \frac{1}{t^{2d}}.$$

Of course, when the drift b is not bounded one cannot hope to get such a uniform result (uniform with respect to z), but similar local results.

Corollary 2.12. *For the system (1.1) with smooth coefficients c and V , for all z and all bounded, open neighborhood U of z , the density $p_t(z, \cdot)$ can be written*

$$p_t(z, \cdot) = q_t(z, \cdot) + r_t(z, \cdot)$$

where $q_t(z, z')$ satisfies the gaussian bound in Theorem 2.10 for $z' \in U$, with C and C' depending in addition on U , and r_t satisfies the following: for any bounded f compactly supported in U ,

$$\int f(z') r_t(z, z') dz' \leq D(U) e^{-\frac{D'(U)}{t}} \|f\|_\infty,$$

for some ad-hoc positive constants $D(U)$ and $D'(U)$.

Proof. Once again the proof is standard. Consider an enlargement U_a of U , for instance $U_a = \{v; d(v, U) \leq a\}$. Let T be the hitting time of U_a^c . Now consider a smooth, bounded with bounded derivatives vector field $b(x', y')$, such that

$$b(x', y') = c(x', y')y' + \nabla V(x') \quad \text{on } U_a.$$

We also consider \bar{Z} the diffusion process solution of (2.11). Starting from $z \in U$, both Z and \bar{Z} coincide up to the (common) stopping time T .

If f is compactly supported in U we may write

$$\begin{aligned} \mathbb{E}_z(f(Z_t)) &= \mathbb{E}_z(f(Z_t) \mathbf{1}_{t>T}) + \mathbb{E}_z(f(Z_t) \mathbf{1}_{t \leq T}) \\ &= \mathbb{E}_z(f(Z_t) \mathbf{1}_{t>T}) + \mathbb{E}_z(f(\bar{Z}_t) \mathbf{1}_{t \leq T}) \\ &= \mathbb{E}_z(f(Z_t) \mathbf{1}_{t>T}) - \mathbb{E}_z(f(\bar{Z}_t) \mathbf{1}_{t>T}) + \mathbb{E}_z(f(\bar{Z}_t)) \end{aligned}$$

and the result follows from the previous theorem since it is well known that

$$\mathbb{P}_z(T < t) \leq C e^{-C' a^2/t}$$

where C and C' only depend on the bounds of b . □

The proof also furnishes the same lower bound on the diagonal than in Konakov et al. (2010).

Actually with some little more effort (using the localization method in Cattiaux (1986)), one can obtain (still with $0 < t < T$)

$$p_t(z, z') \leq q_t(z, z') + c(U) e^{-\frac{c'(U)}{t}}$$

for $z' \in U$. Also notice that these bounds heavily depend on U , in particular the bigger U or its enlargement, the worse the constants $D(U)$ and $D'(U)$.

This localization method can be used to get the following result:

Lemma 2.13. *Let $t_0 > 0$. Then for all fixed z' , $\sup_w p_{t_0}(w, z') = m(t_0, z') < +\infty$.*

This result is contained in Cattiaux (1990) Proposition 1.12.(4), once we have observed that the proof of the latter only requires the boundedness of the coefficients in a neighborhood of z' .

Thanks to the lemma, we have

Proposition 2.14. *For all $t \geq 0$ and all pair (z, z') ,*

$$p_{t_0+t}(z, z') \leq \sup_w p_{t_0}(w, z') < +\infty.$$

Proof. Since $w \mapsto p_{t_0}(w, z')$ is bounded, we may compute

$$P_t(p_{t_0}(\cdot, z'))(z) = \int p_{t_0}(w, z') p_t(z, w) dw = p_{t_0+t}(z, z')$$

the latter equality being an immediate consequence of the Chapman-Kolmogorov equation. The result follows immediately thanks to lemma 2.13 \square

From now on in the whole paper we will assume that Hypothesis \mathcal{H}_1 (or \mathcal{H}_2) is fulfilled.

3. ESTIMATION OF THE INVARIANT DENSITY IN THE STATIONARY REGIME. MAIN RESULTS

In this section we propose two non-parametric estimators for the invariant density p_s . We also assume that we can simulate the chain in the stationary regime, i.e. we assume in the whole section that the distribution of Z_0 is the invariant measure μ . As it is not the case in practice, we will see in Section 6 how to overcome this issue.

First we consider that one can observe the whole process Z at discrete times with discretization step h_n (for typographical reasons we shall use h instead of h_n in various formulae when there is no doubt), i.e we consider

$$\tilde{p}_s(x, y) := \frac{1}{nb_{1,n}^d b_{2,n}^d} \sum_{i=1}^n K \left(\frac{x - X_{ih}}{b_{1,n}}, \frac{y - Y_{ih}}{b_{2,n}} \right). \quad (3.1)$$

Second we consider the partially observed case, where only the position process X can be observed, and we approximate the velocity, i.e. we consider

$$\hat{p}_s(x, y) := \frac{1}{nb_{1,n}^d b_{2,n}^d} \sum_{i=1}^n K \left(\frac{x - X_{ih}}{b_{1,n}}, \frac{y - \frac{X_{(i+1)h} - X_{ih}}{h}}{b_{2,n}} \right). \quad (3.2)$$

In both cases, the kernel K is some C^2 function with compact support A such that $\int_A K(x, y) dx dy = 1$. We may also assume, without loss of generality that A is a bounded

ball. Moreover, we assume that there exists $m \in \mathbb{N}^*$ such that for all non constant polynomial $P(x, y)$ with degree less than or equal to m , $\int P(u, v)K(u, v)dudv = 0$.

Let us state our main result.

Theorem 3.3. *Assume Hypothesis \mathcal{H}_1 or \mathcal{H}_2 are fulfilled. Recall that p_s denotes the density of the invariant measure μ . Assume that the bandwidths $b_{1,n}$, $b_{2,n}$ and the discretization step h_n satisfy assumption \mathcal{H}_3 :*

- (i) $b_{1,n}, b_{2,n}$ and $h_n \rightarrow 0$,
- (ii) $n b_{1,n}^d b_{2,n}^d \rightarrow +\infty$,
- (iii) $\frac{b_{1,n} b_{2,n}}{h_n^2} \rightarrow 0$,
- (iv) *there exists an integer $m > 0$ such that $n b_{1,n}^d b_{2,n}^d \max(b_{1,n}, b_{2,n})^{2(m+1)} \rightarrow 0$.*

Then, in the stationary regime,

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

If in addition

- (v) $nh_n \frac{b_{1,n}^d}{b_{2,n}^{2+d}} \rightarrow 0$,
- (vi) *there exists $1 < p$ such that $nh_n^2 \frac{b_{1,n}^{d(2-p)/p}}{b_{2,n}^{2+d}} \rightarrow 0$.*

Then, in the stationary regime

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

Remark 3.4. In particular, if $h_n = n^{-\gamma}$, $b_{i,n} = n^{-\alpha_i}$, $i = 1, 2$ with $\gamma, \alpha_1, \alpha_2 > 0$, then assumptions (i) to (iv) are equivalent to $\gamma < \frac{\alpha_1 + \alpha_2}{2} < \frac{1}{2d}$ and $m > \frac{1 - (\alpha_1 + \alpha_2)d}{2(\alpha_1 \wedge \alpha_2)}$. Moreover assumptions (i) to (v) are equivalent to $1 - \alpha_1 d + \alpha_2(2 + d) < \gamma < \frac{\alpha_1 + \alpha_2}{2} < \frac{1}{2d}$ and $m > \frac{1 - (\alpha_1 + \alpha_2)d}{2(\alpha_1 \wedge \alpha_2)}$. Then (vi) holds with $p = 1 + \varepsilon$ for any $0 < \varepsilon \leq \frac{\gamma}{2 - \gamma}$. A necessary condition is $\alpha_1 > \frac{2 + \alpha_2(3 + 2d)}{1 + 2d}$. Hence the optimal rate of convergence, for bandwidths $b_{1,n}$, $b_{2,n}$ and discretization step h_n satisfying assumptions in Theorem 3.3, is $n^{-\eta}$ with $\eta < \frac{1}{2(1 + 2d)}$.

Remark 3.5. The first part of the Theorem dealing with the “complete observation” situation can be extended to much more general settings. Since the whole process is observed, we can just use one bandwidth b_n , the dimension of the state space being simply d . A quick look at the proof shows that the only required assumptions are

- (1) the existence of a locally bounded invariant density p_s ,
- (2) the existence for all $t > 0$ of a transition density $p_t(z, z')$ satisfying for each compact subsets K, K' ,

$$\sup_{0 < t \leq 1, z \in K, z' \in K'} p_t(z, z') \leq C(K, K') t^{-\beta d} \quad \text{for some } \beta > 0,$$

(3) the exponential mixing property (2.7) .

In this situation we just have to replace $b_1 b_2$ by b , $2d$ by d , and (iii) has to be changed into $b_n/h_n^\beta \rightarrow 0$ as $n \rightarrow +\infty$.

The situation for partial observation is much more complicated and will heavily depend on the chosen model.

4. A TRIANGULAR CENTRAL LIMIT THEOREM FOR A MIXING SEQUENCE.

In this section we prove a general triangular central limit theorem for a triangular array $(Z_{n,k})_{1 \leq k \leq k_n, n \in \mathbb{N}^*}$ in \mathbb{R}^{2d} , which will be useful for the proof of the main theorem.

Assume that the sequence of integers k_n increases to infinity with n . Let $S_n = Z_{n,1} + \dots + Z_{n,k_n}$, where for each fixed $n \in \mathbb{N}$, $(Z_{n,k})_{k \geq 1}$ is a centered stationary sequence. We assume moreover that there exists a triangular array of positive real numbers $(\alpha_n(j))_{n \in \mathbb{N}^*, 1 \leq j \leq k_n}$ such that $\mathbb{E}(Z_{n,j} Z_{n,0}) \leq \alpha_n(j)$.

Now let $S_{k,n} = Z_{n,1} + \dots + Z_{n,k}$ for $1 \leq k \leq k_n$, we also assume that there exists a constant γ such that

$$\lim_{n \rightarrow \infty} \text{Var} S_n = \gamma^2 > 0 \text{ and defining } v_{n,k} = \text{Var} S_{k,n} - \text{Var} S_{k-1,n}, \quad (4.1)$$

it holds $\exists n_0 \in \mathbb{N}^*$ such that $v_{n,k} > 0 \forall n \geq n_0, 1 \leq k \leq n$.

We shall also set

$$M_n = \sup_{1 \leq k \leq k_n} \|Z_{n,k}\|_\infty, \quad \delta_n = \sup_{1 \leq k \leq k_n} \mathbb{E}(|Z_{n,k}|) \text{ and for } k_n \geq j \geq 1 \quad \Delta_{n,j} = \mathbb{E}(|Z_{n,0} Z_{n,j}|). \quad (4.2)$$

We can state Theorem 4.3 below :

Theorem 4.3. *Assume that the triangular array $(Z_{n,k})_{1 \leq k \leq k_n}$ defined as before satisfies assumption (4.1), then if*

$$(k_n M_n + k_n^{\frac{2}{3}}) M_n \delta_n \rightarrow 0, \quad k_n^{2/3} \sum_{j=1}^{k_n-1} \min(\alpha_n(j), \Delta_{n,j}) \rightarrow 0,$$

and

$$(1 + M_n) \sum_{j=1}^{k_n-1} (k_n - j) \min(\alpha_n(j), (\Delta_{n,j} + \delta_n^2)) \rightarrow 0,$$

as $n \rightarrow \infty$, we obtain

$$S_n \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \gamma^2) .$$

Proof. The proof of Theorem 4.3 is a variation on the Lindeberg method. Up to our knowledge, the Lindeberg method was first used in the setting of kernel density estimation for weakly dependent stationary sequences in Coulon-Prieur and Doukhan (2000). The proof in Coulon-Prieur and Doukhan (2000) was based on a variation of the Lindeberg method proposed by Rio (1995) to prove the central limit theorem for strongly mixing and possibly non-stationary sequences.

For the sake of completeness let us write the complete proof of Theorem 4.3.

Consider a bounded three times differentiable function $f : \mathbb{R} \rightarrow \mathbb{R}$ with continuous and bounded derivatives. Set $C_j = \|f^{(j)}\|_\infty$, for $j = 0, 1, 2, 3$. Also consider $\sigma_n^2 = \text{Var}S_n$.

For some standard Gaussian r.v. η , define

$$\Delta_n(f) = \mathbb{E}(f(S_n) - f(\sigma_n \eta)).$$

Theorem 4.3 will follow from Assumption (4.1), if we prove that

$$\lim_{n \rightarrow \infty} \Delta_n(f) = 0.$$

Recall that $v_{n,k} > 0$ for each k and set $N_{n,k} \sim \mathcal{N}(0, v_{n,k})$.

The random variables $(N_{n,k})_{1 \leq k \leq k_n, n \geq 1}$ are assumed to be independent and independent of the sequence $(Z_k)_{k \geq 1}$.

The idea is to write $\sigma_n \eta$ as the sum of the $N_{n,k}$ and to replace step by step $Z_{n,k}$ by $N_{n,k}$, writing $\Delta_n(f)$ as a ‘‘cascade’’.

More precisely, for $1 \leq k \leq k_n$, we set $T_{k,n} = \sum_{j=k+1}^{k_n} N_{n,j}$, empty sums being as usual equal to 0, and we use Rio’s decomposition

$$\Delta_n(f) = \sum_{k=2}^{k_n} \Delta_{k,n}(f), \quad (4.4)$$

with $\Delta_{k,n}(f) = \mathbb{E}(f(S_{k-1,n} + Z_{n,k} + T_{k,n}) - f(S_{k-1,n} + N_{n,k} + T_{k,n}))$ and $S_{0,n} = 0$.

In order to bound each term, one writes $\Delta_{k,n}(f) = \Delta_{k,n}^{(1)}(f) - \Delta_{k,n}^{(2)}(f)$, with

$$\Delta_{k,n}^{(1)}(f) = \mathbb{E}(f(S_{k-1,n} + Z_{n,k} + T_{k,n})) - \mathbb{E}(f(S_{k-1,n} + T_{k,n})) - \frac{v_{n,k}}{2} \mathbb{E}(f''(S_{k-1,n} + T_{k,n})), \quad (4.5)$$

$$\Delta_{k,n}^{(2)}(f) = \mathbb{E}(f(S_{k-1,n} + N_{n,k} + T_{k,n})) - \mathbb{E}(f(S_{k-1,n} + T_{k,n})) - \frac{v_{n,k}}{2} \mathbb{E}(f''(S_{k-1,n} + T_{k,n})). \quad (4.6)$$

Notice that the function $x \rightarrow f_{k,n}(x) = \mathbb{E}(f(x + T_{k,n}))$ has the same regularity properties as f ; e.g. for $0 \leq j \leq 3$, $\|f_{k,n}^{(j)}\| \leq C_j$. Using independence (recall the definition of $T_{k,n}$) it is not difficult to see that one can write

$$\Delta_{k,n}^{(1)}(f) = \mathbb{E}(f_{k,n}(S_{k-1,n} + Z_{n,k})) - \mathbb{E}(f_{k,n}(S_{k-1,n})) - \frac{v_{n,k}}{2} \mathbb{E}(f_{k,n}''(S_{k-1,n})),$$

$$\Delta_{k,n}^{(2)}(f) = \mathbb{E}(f_{k,n}(S_{k-1,n} + N_{n,k})) - \mathbb{E}(f_{k,n}(S_{k-1,n})) - \frac{v_{n,k}}{2} \mathbb{E}(f_{k,n}''(S_{k-1,n})).$$

• *Bound for $\Delta_{k,n}^{(2)}(f)$.* Taylor expansion yields the existence of some random variable $\tau_{n,k} \in (0, 1)$:

$$\Delta_{k,n}^{(2)}(f) = \mathbb{E}(f'_{k,n}(S_{k-1,n})N_{n,k}) + \frac{1}{2} \mathbb{E}(f''_{k,n}(S_{k-1,n})(N_{n,k}^2 - v_{n,k})) + \frac{1}{6} \mathbb{E}(f_{k,n}^{(3)}(S_{k-1,n} + \tau_{n,k}N_{n,k})N_{n,k}^3).$$

Using independence, we see that the first two terms vanish. In addition since the third derivative of f is bounded we get $|\Delta_{k,n}^{(2)}(f)| \leq \frac{C_3}{6} \mathbb{E}(|N_{n,k}|^3)$, hence

$$|\Delta_{k,n}^{(2)}(f)| \leq \frac{C_3 v_{n,k}^{3/2}}{9\sqrt{2\pi}}.$$

Now $v_{n,k} = \text{Var}Z_{n,k} + 2 \sum_{j=1}^{k-1} \text{Cov}(Z_{n,j}, Z_{n,k})$, hence

$$v_{n,k} \leq M_n \delta_n + 2 \sum_{j=1}^{k-1} \min(\alpha_n(k-j), \Delta_{n,k-j}).$$

For $\Delta_n^{(2)}(f)$ to go to zero, we thus need

$$k_n^{2/3} \left[M_n \delta_n + 2 \sum_{j=1}^{k_n-1} \min(\alpha_n(j), \Delta_{n,j}) \right] \rightarrow_{n \rightarrow \infty} 0. \quad (4.7)$$

• *Bound for $\Delta_{k,n}^{(1)}(f)$.* Set $\Delta_{k,n}^{(1)}(f) = \mathbb{E}(\delta_{k,n}^{(1)}(f))$. Then, using Taylor formula again (with some random $\tau_{k,n} \in (0, 1)$), we may write

$$\delta_{k,n}^{(1)}(f) = f'_{k,n}(S_{k-1,n})Z_{n,k} + \frac{1}{2}f''_{k,n}(S_{k-1,n})(Z_{n,k}^2 - v_{n,k}) + \frac{1}{6} \left(f_{k,n}^{(3)}(S_{k-1,n} + \tau_{k,n}Z_{n,k})Z_{n,k}^3 \right).$$

We analyze separately the terms in the previous expression. The last term can be bounded in the following way

$$|\mathbb{E}(f_{k,n}^{(3)}(S_{k-1,n} + \tau_{k,n}Z_{n,k})Z_{n,k}^3)| \leq C_3 M_n^2 \delta_n. \quad (4.8)$$

The second term can be written as

$$\frac{1}{2} \left(\text{Cov}(f''_{k,n}(S_{k-1,n}), Z_{n,k}^2) - 2\mathbb{E}(f''_{k,n}(S_{k-1,n})) \sum_{i=1}^{k-1} \mathbb{E}(Z_{n,k}Z_{n,i}) \right).$$

On one hand, as $\text{Cov}(f''_{k,n}(S_{k-1,n}), Z_{n,k}^2) = \sum_{j=1}^{k-1} \text{Cov}(f''_{k,n}(S_{j,n}) - f''_{k,n}(S_{j-1,n}), Z_{n,k}^2)$, using the mixing property, it holds

$$|\text{Cov}(f''_{k,n}(S_{k-1,n}), Z_{n,k}^2)| \leq C_3 M_n \sum_{j=1}^{k-1} \min(\alpha_n(k-j), (\Delta_{n,k-j} + \delta_n^2)). \quad (4.9)$$

On the other hand,

$$\mathbb{E}(f''_{k,n}(S_{k-1,n})) \sum_{i=1}^{k-1} \mathbb{E}(Z_{n,k}Z_{n,i}) \leq C_2 \sum_{i=1}^{k-1} \min(\alpha_n(k-i), \Delta_{n,k-i}). \quad (4.10)$$

Finally, write the first order term $f'_{k,n}(S_{k-1,n})Z_{n,k} = \sum_{j=1}^{k-1} [f'_{k,n}(S_{j,n}) - f'_{k,n}(S_{j-1,n})]Z_{n,k}$.

We have

$$|\text{Cov}(f'_{k,n}(S_{j,n}) - f'_{k,n}(S_{j-1,n}), Z_{n,k})| \leq C_2 \min(\alpha_n(k-j), (\delta_n^2 + \Delta_{n,k-j})). \quad (4.11)$$

Then, adding (4.8) to (4.11) one gets :

$$|\Delta_{k,n}^{(1)}(f)| \leq C \left[M_n^2 \delta_n + (1 + M_n) \sum_{j=1}^{k-1} \min(\alpha_n(j), (\Delta_{n,j} + \delta_n^2)) \right]. \quad (4.12)$$

Now we sum up for all k to conclude:

$$\left| \sum_{k=2}^{k_n} \Delta_{k,n}^{(1)}(f) \right| \leq C \left[k_n M_n^2 \delta_n + (1 + M_n) \sum_{j=1}^{k_n-1} (k_n - j) \min(\alpha_n(j), (\Delta_{n,j} + \delta_n^2)) \right]. \quad (4.13)$$

It concludes the proof of Theorem 4.3. \square

5. PROOF OF THEOREM 3.3

The proof of Theorem 3.3 is decomposed in several steps, as we have to consider the discretization error as well as the stochastic error.

Step 1 : In this first step, we consider the fully observed discretization (X_{kh_n}, Y_{kh_n}) , $k = 1, \dots, n$ of $(Z_t, t \geq 0)$. For simplicity we write h instead of h_n in the sequel.

We want to apply the result of Theorem 4.3 to our problem of density estimation, with $k_n = n$.

We thus define $Z_{n,k} = \frac{1}{\sqrt{nb_{1,n}^d b_{2,n}^d}} \left(K \left(\frac{x - X_{kh}}{b_{1,n}}, \frac{y - Y_{kh}}{b_{2,n}} \right) - \mathbb{E} K \left(\frac{x - X_0}{b_{1,n}}, \frac{y - Y_0}{b_{2,n}} \right) \right) = g_n(Z_{kh_n})$ with

$$g_n(s, t) := \frac{1}{\sqrt{nb_{1,n}^d b_{2,n}^d}} K \left(\frac{x - s}{b_{1,n}}, \frac{y - t}{b_{2,n}} \right) - \mathbb{E} \frac{1}{\sqrt{nb_{1,n}^d b_{2,n}^d}} K \left(\frac{x - X_0}{b_{1,n}}, \frac{y - Y_0}{b_{2,n}} \right),$$

so that

$$M_n = \mathcal{O} \left(\frac{1}{\sqrt{nb_{1,n}^d b_{2,n}^d}} \right).$$

Using the covariance inequality (2.7) one gets $\alpha_n(j) = \mathcal{O} (M_n^2 \rho^{(jh_n)/2})$.

Writing

$$v_{n,k} \geq \text{Var} Z_{n,k} - 2 \sum_{j=1}^{k-1} \min(\alpha_n(k-j), \Delta_{n,k-j}),$$

to verify (4.1), it is not hard to see (since of course $b_{i,n} \rightarrow 0$ as n grows to infinity), that it suffices to check

$$n \sum_{j=1}^n \min(\alpha_n(j), \Delta_{n,j}) \rightarrow 0 \quad \text{as } n \rightarrow +\infty.$$

This condition clearly implies the second one in Theorem 4.3.

In order to estimate δ_n , just write the definition

$$\begin{aligned} \delta_n &= \sup_{k \leq n} \mathbb{E}(|Z_{n,k}|) \leq 2 \frac{1}{\sqrt{nb_{1,n}^d b_{2,n}^d}} \mathbb{E} K \left(\frac{x - X_0}{b_{1,n}}, \frac{y - Y_0}{b_{2,n}} \right) \\ &\leq \frac{2}{\sqrt{nb_{1,n}^d b_{2,n}^d}} \int K \left(\frac{x - u}{b_{1,n}}, \frac{y - v}{b_{2,n}} \right) p_s(u, v) du dv \\ &\leq \frac{2 b_{1,n}^{d/2} b_{2,n}^{d/2}}{\sqrt{n}} \int_A K(u, v) p_s(b_{1,n}(x - u), b_{2,n}(y - v)) du dv. \end{aligned} \quad (5.1)$$

Notice that, if $(u, v) \in A$ and $b_{i,n} \leq 1$ for $i = 1, 2$, then $(b_{1,n}(x - u), b_{2,n}(y - v)) \in (x, y) - A$ which is bounded. Hence p_s is bounded on the latter set. It follows

$$\delta_n = \mathcal{O} \left(\frac{b_{1,n}^{d/2} b_{2,n}^{d/2}}{\sqrt{n}} \right).$$

In order to estimate $\Delta_{n,j}$, as for the estimate of δ_n , we can come back to the definition yielding

$$\Delta_{n,j} \leq 3 \delta_n^2 + A_{n,j}$$

where

$$\begin{aligned} A_{n,j} &= \frac{1}{nb_{1,n}^d b_{2,n}^d} \int K \left(\frac{x - u}{b_{1,n}}, \frac{y - v}{b_{2,n}} \right) K \left(\frac{x - u'}{b_{1,n}}, \frac{y - v'}{b_{2,n}} \right) \\ &\quad p_s(u, v) p_{jh}((u, v), (u', v')) du' dv' du dv. \end{aligned} \quad (5.2)$$

Using the same change of variables as in (5.1), the same compactness argument and the estimate in Corollary 2.12 we obtain that

$$A_{n,j} \leq \frac{C}{n} \left(b_{1,n}^d b_{2,n}^d \left(\frac{1}{(jh)^{2d}} + e^{-D/jh} \right) \right).$$

Define now

$$\tilde{p}_s(x, y) := \frac{1}{nb_{1,n}^d b_{2,n}^d} \sum_{i=1}^n K \left(\frac{x - X_{ih}}{b_{1,n}}, \frac{y - Y_{ih}}{b_{2,n}} \right).$$

Applying Theorem 4.3 we get

$$\sqrt{nb_{1,n}^d b_{2,n}^d} (\tilde{p}_s(x, y) - \mathbb{E} \tilde{p}_s(x, y)) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N} \left(0, p_s(x, y) \int \int K^2(s, t) ds dt \right),$$

as soon as , when $n \rightarrow +\infty$,

- (1) $nb_{1,n}^d b_{2,n}^d \rightarrow +\infty$, implying that $M_n \rightarrow 0$, and
- (2) $\sum_{j=1}^{n-1} \min \left(\frac{\rho^{jh/2}}{b_{1,n}^d b_{2,n}^d}, C b_{1,n}^d b_{2,n}^d \left(\frac{1}{(jh)^{2d}} + 1 \right) \right) \rightarrow 0$, for some ad-hoc constant C .

A necessary condition for (2) to hold is that the first term (for $j = 1$) goes to 0 i.e. that

$$\frac{b_{1,n} b_{2,n}}{h_n^2} \rightarrow 0. \quad (5.3)$$

It turns out that this condition is also sufficient.

Indeed, using $\min(a, b + c) \leq \min(a, b) + \min(a, c)$, we have

$$\sum_{j=1}^{n-1} \frac{\rho^{jh/2}}{b_{1,n}^d b_{2,n}^d} \wedge C b_{1,n}^d b_{2,n}^d \left(\frac{1}{(jh)^{2d}} + 1 \right) \leq \frac{C b_{1,n}^d b_{2,n}^d}{h^{2d}} \sum_{j=1}^{n-1} \frac{1}{j^2} + \sum_{j=1}^{n-1} \frac{\rho^{jh/2}}{b_{1,n}^d b_{2,n}^d} \wedge C b_{1,n}^d b_{2,n}^d.$$

If (5.3) holds, the first term in the right hand side goes to 0. For the second one, define $j_n = h^{-2d}$. If (5.3) is satisfied, it holds $b_{1,n} b_{2,n} \leq C' h_n^2$, so that

$$\begin{aligned} \sum_{j=1}^{n-1} \frac{\rho^{jh/2}}{b_{1,n}^d b_{2,n}^d} \wedge C b_{1,n}^d b_{2,n}^d &\leq C b_{1,n}^d b_{2,n}^d j_n + \sum_{j=j_n}^{n-1} \frac{\rho^{jh_n/2}}{b_{1,n}^d b_{2,n}^d} \\ &\leq \frac{C b_{1,n}^d b_{2,n}^d}{h_n^{2d}} + \frac{\rho^{j_n h_n/2}}{b_{1,n}^d b_{2,n}^d (1 - \rho^{h_n/2})} \\ &\leq \frac{C b_{1,n}^d b_{2,n}^d}{h_n^{2d}} + \frac{2 \rho^{j_n h_n/2}}{b_{1,n}^d b_{2,n}^d h_n \log(1/\rho)} \\ &\leq \frac{C b_{1,n}^d b_{2,n}^d}{h_n^{2d}} + \frac{2C' \exp\left(-\log(1/\rho) \frac{C'}{b_{1,n}^{d-\frac{1}{2}} b_{2,n}^{d-\frac{1}{2}}}\right)}{b_{1,n}^{d+\frac{1}{2}} b_{2,n}^{d+\frac{1}{2}} \log(1/\rho)}, \end{aligned}$$

and both terms go to 0 when $n \rightarrow +\infty$ since $b_{i,n} \rightarrow 0$ and $\rho < 1$.

Step 2 :

If we can only observe the position process, we have to consider instead of Y_{ih} its natural approximation $\frac{X_{(i+1)h} - X_{ih}}{h}$.

Recall our estimator

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

We want now to evaluate the following difference :

$$\begin{aligned} &\sqrt{n b_{1,n}^d b_{2,n}^d} (\hat{p}_s(x, y) - \tilde{p}_s(x, y)) \\ &= \frac{1}{\sqrt{n b_{1,n}^d b_{2,n}^d}} \sum_{i=1}^n \left(K \left(\frac{x - X_{ih}}{b_{1,n}}, \frac{y - \frac{X_{(i+1)h} - X_{ih}}{h}}{b_{2,n}} \right) - K \left(\frac{x - X_{ih}}{b_{1,n}}, \frac{y - Y_{ih}}{b_{2,n}} \right) \right). \end{aligned} \quad (5.4)$$

Introduce $M_u = \frac{1}{h} \int_{ih}^u (Y_s - Y_{ih}) ds$, defined for $ih \leq u \leq (i+1)h$. Then we may write

$$\begin{aligned} A_i &= K \left(\frac{x - X_{ih}}{b_{1,n}}, \frac{y - \frac{X_{(i+1)h} - X_{ih}}{h}}{b_{2,n}} \right) - K \left(\frac{x - X_{ih}}{b_{1,n}}, \frac{y - Y_{ih}}{b_{2,n}} \right) \\ &= \frac{-1}{hb_{2,n}} \int_{ih}^{(i+1)h} \nabla_y K \left(\frac{x - X_{ih}}{b_{1,n}}, \frac{y - Y_{ih} - M_u}{b_{2,n}} \right) \cdot (Y_u - Y_{ih}) du. \end{aligned}$$

Recall that K is compactly supported with bounded derivatives, so that for some well chosen constants C and D ,

$$|\nabla_y K(a, b)| \leq C \mathbb{1}_{|a| \leq D} \mathbb{1}_{|b| \leq D}.$$

Hence

$$|A_i| \leq \frac{C \mathbb{1}_{|X_{ih} - x| \leq Db_{1,n}}}{hb_{2,n}} \int_{ih}^{(i+1)h} |Y_u - Y_{ih}| \mathbb{1}_{|Y_{ih} - y + M_u| \leq Db_{2,n}} du. \quad (5.5)$$

We will estimate the expectation of $|A_i|$. To this end, first write

$$Y_u - Y_{ih} = \sigma(W_u - W_{ih}) - \int_{ih}^u (c(X_t, Y_t) Y_t + \nabla V(X_t)) dt.$$

Hence, using the Markov property and since $\nabla V(\cdot)$ increases at infinity with polynomial rate, there exists some $k \in \mathbb{N}$ and some $M > 0$ (that may change from line to line, but depending only on the coefficients and the dimension) such that

$$\begin{aligned} \mathbb{E}(|A_i|) &\leq \int_{ih}^{(i+1)h} \mathbb{E} \left(\frac{C \mathbb{1}_{|X_{ih} - x| \leq Db_{1,n}}}{hb_{2,n}} \mathbb{E}_{Z_{ih}}(|Y_{u-ih} - Y_0|) \right) du \\ &\leq \frac{M}{hb_{2,n}} \int_{ih}^{(i+1)h} \mathbb{E} \left(\mathbb{1}_{|X_{ih} - x| \leq Db_{1,n}} \left(|u - ih|^{1/2} + \int_0^{u-ih} \mathbb{E}_{Z_{ih}}(|Z_s|^k) ds \right) \right) du \\ &\leq \frac{M}{hb_{2,n}} \left(h^{3/2} b_{1,n}^d + \int_{ih}^{(i+1)h} \mathbb{E} \left(\mathbb{1}_{|X_{ih} - x| \leq Db_{1,n}} \int_0^{u-ih} \mathbb{E}_{Z_{ih}}(|Z_s|^k) ds \right) \right) du, \end{aligned}$$

where we used that μ has a locally bounded density. In order to control the second term in the sum, first use Hölder inequality

$$\begin{aligned} \mathbb{E} \left(\mathbb{1}_{|X_{ih} - x| \leq Db_{1,n}} \int_0^{u-ih} \mathbb{E}_{Z_{ih}}(|Z_s|^k) ds \right) &\leq M b_{1,n}^{d/p} \mathbb{E}^{1/q} \left[\left(\int_0^{u-ih} \mathbb{E}_{Z_{ih}}(|Z_s|^k) ds \right)^q \right] \\ &\leq M b_{1,n}^{d/p} (u - ih) \mathbb{E}^{1/q} \left[\int_0^{u-ih} \mathbb{E}_{Z_{ih}}(|Z_s|^{kq}) \frac{ds}{u - ih} \right] \\ &\leq M b_{1,n}^{d/p} (u - ih) \left(\int_0^{u-ih} \mathbb{E}(|Z_s|^{kq}) \frac{ds}{u - ih} \right)^{1/q} \\ &\leq M(q) b_{1,n}^{d/p} (u - ih), \end{aligned}$$

where we have used the existence of all polynomial moments of μ .

Let us come back to (5.4). We have obtained, for all $1 < p < +\infty$ with conjugate q ,

$$\begin{aligned} \mathbb{E} \left(\sqrt{nb_{1,n}^d b_{2,n}^d} |\hat{p}_s(x, y) - \tilde{p}_s(x, y)| \right) &\leq \frac{1}{\sqrt{nb_{1,n}^d b_{2,n}^d}} \sum_{i=1}^n \mathbb{E}(|A_i|) \\ &\leq \frac{\sqrt{n}}{\sqrt{b_{1,n}^d b_{2,n}^d}} \frac{M(q)}{hb_{2,n}} \left(h^{3/2} b_{1,n}^d + b_{1,n}^{d/p} h^2 \right). \end{aligned}$$

It follows that $\sqrt{nb_{1,n}^d b_{2,n}^d} (\hat{p}_s(x, y) - \tilde{p}_s(x, y))$ goes to 0 in \mathbb{L}^1 as soon as

$$\sqrt{nh_n} \frac{b_{1,n}^{d/2}}{b_{2,n}^{1+(d/2)}} \rightarrow 0 \quad \text{and there exists } 1 < p \text{ such that } \sqrt{n} h_n \frac{b_{1,n}^{d(\frac{1}{p}-\frac{1}{2})}}{b_{2,n}^{1+(d/2)}} \rightarrow 0. \quad (5.6)$$

Step 3 :

It remains to consider the bias term

$$B_n = \sqrt{nb_{1,n}^d b_{2,n}^d} |\mathbb{E}(\tilde{p}_s(x, y)) - p_s(x, y)|$$

which is independent of the mixing properties. It can be written

$$\begin{aligned} B_n &= \sqrt{nb_{1,n}^d b_{2,n}^d} \int \frac{1}{b_{1,n}^d b_{2,n}^d} K \left(\frac{x-u}{b_{1,n}}, \frac{y-v}{b_{2,n}} \right) (p_s(u, v) - p_s(x, y)) du dv \\ &= \sqrt{nb_{1,n}^d b_{2,n}^d} \int K(u, v) (p_s(x - b_{1,n}u, y - b_{2,n}v) - p_s(x, y)) du dv. \end{aligned}$$

For the latter to go to 0, using Taylor expansion and standard tools, it is enough to assume that there exists $m \in \mathbb{N}^*$ such that for all polynomial $P(x, y)$ with degree less than or equal to m , $\int P(u, v)K(u, v)dudv = 0$, and $nb_{1,n}^d b_{2,n}^d \max(b_{1,n}, b_{2,n})^{2(m+1)} \xrightarrow[n \rightarrow \infty]{} 0$.

The proof of Theorem 3.3 is completed. \square

6. NON-STATIONARY CASE

In Section 3 we stated the central limit theorem for the invariant density $p_s(x, y)$ in the case where the process is in the stationary regime. Let us now define the new estimator

$$\bar{p}_n(x, y) = \frac{1}{\sqrt{nb_{1,n}^d b_{2,n}^d}} \sum_{k=l_n+1}^{l_n+n} K \left(\frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - \frac{X_{(i+1)h_n} - X_{ih_n}}{h_n}}{b_{2,n}} \right). \quad (6.1)$$

Remark that if $Z_0 \sim \mu(dz)$, then $\bar{p}_n(x, y) \stackrel{\mathcal{L}}{=} \hat{p}_n(x, y) \forall n \in \mathbb{N}^*$.

Theorem 6.2 below states that we can estimate $p_s(x, y)$ by using $\bar{p}_n(x, y)$ with $Z_0 = z_0 = (x_0, y_0)$.

Theorem 6.2. *Assume that assumptions (i) to (vi) in Theorem 3.3 hold true. Then, starting from any initial point $z_0 = (x_0, y_0)$, and assuming that l_n appearing in the definition (6.1) of $\bar{p}_n(x, y)$ satisfies $l_n h_n \xrightarrow{n \rightarrow +\infty} +\infty$, one has*

$$\sqrt{nb_{1,n}^d b_{2,n}^d} (\bar{p}_n(x, y) - p_s(x, y)) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N} \left(0, p_s(x, y) \int K^2(s, t) ds dt \right).$$

Proof of Theorem 6.2 :

Recall that $p_s : \mathbb{R}^{2d} \mapsto \mathbb{R}_+$ denote the invariant density of Z and μ the associated invariant probability measure.

Denote by $\mathcal{C}_b(\mathbb{R})$ the set of bounded continuous functions $h : \mathbb{R} \rightarrow \mathbb{R}$. It is only necessary to prove that, for any $h \in \mathcal{C}_b(\mathbb{R})$, the difference

$$\begin{aligned} \Delta_n(h) &= \mathbb{E} \left[h(\sqrt{nb_{1,n}^d b_{2,n}^d} \hat{p}_s(x, y) | Z_0 \sim \mu) - h(\sqrt{nb_{1,n}^d b_{2,n}^d} \bar{p}_n(x, y) | Z_0 = z_0) \right] \\ &= \mathbb{E} \left[h(\sqrt{nb_{1,n}^d b_{2,n}^d} \bar{p}_n(x, y) | Z_0 \sim \mu) - h(\sqrt{nb_{1,n}^d b_{2,n}^d} \bar{p}_n(x, y) | Z_0 = z_0) \right] \end{aligned}$$

goes to zero as n tends to infinity. Let $h \in \mathcal{C}_b(\mathbb{R})$ and denote $\theta = \|h\|_\infty$.

To evaluate $\mathbb{E}(h(\bar{p}_n(x, y)) | Z_0 \sim \mu) - \mathbb{E}(h(\bar{p}_n(x, y)) | Z_0 = z_0)$. Let us fix $n \in \mathbb{N}^*$. We first make the computations conditionally to $Z_{jh_n}, j \geq l_n + 2$.

Define $h_{Z_{jh_n}, j \geq l_n + 2}(z')$ by

$$h \left(\frac{1}{\sqrt{nb_{1,n}^d b_{2,n}^d}} \left[K \left(\frac{x - x'}{b_{1,n}}, \frac{y - \frac{X_{(l_n+2)h_n} - x'}{h_n}}{b_{2,n}} \right) + \sum_{k=l_n+2}^{l_n+n} K \left(\frac{x - X_{kh_n}}{b_{1,n}}, \frac{y - \frac{X_{(i+1)h_n} - X_{ih_n}}{h_n}}{b_{2,n}} \right) \right] \right).$$

Now, conditionally to $Z_{jh_n}, j \geq l + 2$, one has :

$$\begin{aligned} & \left| \mathbb{E}(h(\bar{p}_n(x, y)) | Z_0 \sim \mu) - \mathbb{E}(h(\bar{p}_n(x, y)) | Z_0 = z_0) \right| \\ &= \left| \int h_{Z_{jh_n}, j \geq l_n + 2}(z') (p_s(z') - q_{(l_n+1)h_n}(z_0, z')) dz' \right| \\ &\leq \theta D \rho^{(l_n+1)h_n} \Psi(z_0) \end{aligned} \tag{6.3}$$

using Inequality (2.1).

Finally, since $0 < \rho < 1$, we can conclude that $\Delta_n(h)$ goes to zero as n tends to infinity as soon as $l_n h_n \xrightarrow{n \rightarrow +\infty} +\infty$, which concludes the proof of Theorem 6.2. \square

7. SIMULATION STUDY

We consider two models for simulations. The first one has been proposed by Pokern et al. (2009). It corresponds to a linear oscillator subject to noise and damping with $\gamma > 0$. The second example is one example of generalized Duffing oscillators described in Wu (2001) subsection 5.2. Both models are of type (1.1) and satisfy assumptions needed to apply our estimation results.

7.1. Model I: harmonic oscillator. Consider an harmonic oscillator that is driven by a white noise forcing:

$$\begin{aligned} dX_t &= Y_t dt \\ dY_t &= \sigma dW_t - (\kappa Y_t + DX_t) dt. \end{aligned} \quad (7.1)$$

with $\sigma > 0$, $\kappa > 0$ and $D > 0$. In the following we choose $D = 2$, $\kappa = 2$ and $\sigma = 1$. For this model we know that the stationary distribution is gaussian, with mean zero and an explicit variance matrix given in Gardiner (1985), e.g. With our choice of parameters, the gaussian invariant density is

$$p_s(x, y) = \frac{2\sqrt{2}}{\pi} \exp(-4x^2 - 2y^2).$$

In the following we make use of the explicit Euler scheme to simulate an approximated discrete sampling $(\tilde{X}_i, \tilde{Y}_i)_{i \in \mathbb{N}}$ of $(X_t, Y_t)_{t \in \mathbb{R}_+}$. For a given step $\delta > 0$, the scheme is defined as

$$\begin{aligned} \tilde{X}_{i+1} - \tilde{X}_i &= \tilde{Y}_i \delta \\ \tilde{Y}_{i+1} - \tilde{Y}_i &= \sigma (W_{(i+1)\delta} - W_{i\delta}) - (\kappa \tilde{Y}_i + D \tilde{X}_i) \delta \end{aligned} \quad (7.2)$$

$(\tilde{X}_0, \tilde{Y}_0) = (0, 0)$.

We now estimate the invariant density p_s on a grid $(z_l)_{l=1, \dots, L} = (x_l, y_l)_{l=1, \dots, L}$. More precisely, for $l = 1, \dots, L$ we consider the estimate

$$\hat{p}_s(x_l, y_l) = \frac{1}{nb_{1,n}b_{2,n}} \sum_{i=1}^n K \left(\frac{x - U_i}{b_{1,n}}, \frac{y - \frac{U_{i+1} - U_i}{h_n}}{b_{2,n}} \right)$$

with K the Epanechnikov kernel defined by $K(u, v) = K_e(u) \times K_e(v)$ with $K_e(w) = \frac{3}{4}(1 - w^2)\mathbf{1}_{|w| \leq 1}$, $b_{1,n} = n^{-\alpha_1}$, $b_{2,n} = n^{-\alpha_2}$, $h_n = n^{-\gamma}$. The sampling $(U_i)_{i=1, \dots, n}$ is simulated using, as in Pokern et al. (2009), the explicit Euler scheme described above with a step $\delta = (1/30)h_n$, and then considering $U_i = \tilde{X}_{30i}$, $i = 1, \dots, n$.

To measure the performance of the method, we estimate on $M = 30$ samplings the relative mean integrated squared error by

$$\frac{\frac{1}{M} \sum_{m=1}^M \frac{1}{L} \sum_{l=1}^L \left(\hat{p}_s^{(m)}(z_l) - p_s(z_l) \right)^2}{\frac{1}{L} \sum_{l=1}^L p_s(z_l)^2}.$$

In the following, we take $\alpha_1 = 0.2$, $\alpha_2 = 0.2$, $\gamma = 0.15$. We take a grid of size $L = 100$. We obtain the following results (see Table 1 below) :

n	100	1000	10000
relative error	0.09	0.03	0.01

TABLE 1. Evolution of the relative mean integrated squared error for $\alpha_1 = 0.2$, $\alpha_2 = 0.2$, $\gamma = 0.15$, $L = 100$ and $M = 30$ and different sizes of sampling

As expected, the relative error decreases with n .

We obtain for a sample size of $n = 10^4$ the following graphics for the discretized theoretical bivariate invariant density (see Figure 1), its estimated version (see Figure 2), and the estimation of the marginal invariant density for the position (see Figure 3) and of the marginal invariant density for the velocity (see Figure 4) when only the position is observed.

[Fig. 1, 2, 3 and 4 about here.]

To measure the influence of the explicit Euler scheme used to simulate the data, we also simulated with an exact scheme the stationary distribution (which is gaussian in this case) as it is done in Samson and Thieullen (2012). We do not show all the results of these simulations here as they were quite similar to the ones obtained with an explicit Euler scheme. We just show on Figure 5 below the estimated bivariate invariant density for a sample of size $n = 10^4$.

[Fig. 5 about here.]

We now want to evaluate the sharpness of assumptions (ii) to (iv) in Theorem 3.3. Assumption (ii) on the bandwidths $b_{1,n}$ and $b_{2,n}$ is usual in kernel density estimation. Assumption (iv) is needed if one wants that the bias term in the central limit theorem vanishes as n tends to infinity. Let us now focus on assumption (iii) which links the bandwidths terms with the discretization step. In order to understand the meaning of this assumption, we consider the convergence of $\tilde{p}_s(x, y)$ to $p_s(x, y)$. We thus work with the sample $(X_{ih_n}, Y_{ih_n})_{i=1, \dots, n}$. The mixing property of this sequence may be lost if the step h_n tends too quickly to zero. In that case the convergence in Theorem 3.3, even when working with \tilde{p}_s instead of \hat{p}_s may fail.

To illustrate this point, we now use the exact scheme to simulate under the stationary distribution. To draw Figures 6 and 7, we chose $\alpha_1 = \alpha_2 = 0.2$, $\gamma = 0.15$, and $n = 10^4$, thus assumptions (ii) to (iv) are satisfied.

[Fig. 6, 7 about here.]

We then consider $n = 10^4$, $\alpha_1 = \alpha_2 = 0.22$ and $\gamma = 0.30$, thus assumption (iii) in Theorem 3.3 is not satisfied. We see on Figures 8 and 9 that the quality of the estimation is lower than previously.

[Fig. 8, 9 about here.]

The deterioration of the convergence in Theorem 3.3 in case assumption (iii) is not satisfied was observed on a large set of simulations, thus strengthening the sharpness of this assumption.

7.2. Model II: Kramers oscillator. We consider the noisy Duffing oscillator known as Kramers oscillator. The system (1.1) writes now

$$\begin{aligned} dX_t &= Y_t dt \\ dY_t &= \sigma dW_t - (\kappa Y_t + \alpha X_t^3 - \beta X_t) dt \end{aligned} \quad (7.3)$$

with σ, κ, α and $\beta > 0$. The potential is then $V(x) = \alpha \frac{x^4}{4} - \beta \frac{x^2}{2}$.

For nonlinear oscillators with cubic restoring force, the stability of the explicit Euler scheme may fail (see e.g., Talay (2002)). Therefore we used the implicit Euler scheme (see Talay (2002) for its definition) to simulate the data. The invariant density is in that case

$$p_s(x, y) = \frac{\sqrt{\kappa}}{\sqrt{\pi}\sigma C} \exp\left(\frac{-2\kappa}{\sigma^2}\left(\frac{\alpha x^4}{4} - \frac{\beta x^2}{2} + \frac{y^2}{2}\right)\right),$$

with C the normalizing constant. In the following, we take $\sigma = \kappa = \alpha = \beta = 1$. We use the same estimation procedure as for the preceding example. For this model, contrarily to the previous example, we do not have any exact simulation scheme for the invariant distribution. Simulations are slower as at each step one has to solve a fixed point problem, but such a scheme is stable for our model.

In the following, we choose the Epanechnikov kernel and we take $\alpha_1 = 0.2$, $\alpha_2 = 0.2$, $\gamma = 0.15$. We take a grid of size $L = 100$.

We obtain for a sample size of $n = 10^4$ the following graphics for the discretized theoretical bivariate invariant density (see Figure 10), its estimated version (see Figure 11), and the estimation of the marginal invariant density for the position (see Figure 12) and of the marginal invariant density for the velocity (see Figure 13) when only the position is observed.

[Fig. 10, 11, 12 and 13 about here.]

Acknowledgements : The authors gratefully acknowledge the ECOS-Nord program for French- Venezuelan Scientific Cooperation (V12M01 project). They also thank Angie Pineda from Universidad Central Venezuela in Caracas for preliminary simulations. We also thank an anonymous referee for his (her) suggestions and questions, who helped us to improve the presentation of the paper.

REFERENCES

- 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.
- 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. Chafai, and A. Guillin (2011). Central limit theorems for additive functionals of ergodic Markov diffusions processes. *ALEA, Lat. Am. J. Probab. Math. Stat.* **9(2)**, p. 337-382, (2012).
- C. Coulon-Prieur and P. Doukhan (2000). A CLT for triangular arrays of weakly dependent sequences. *Statist. Probab. Letters* **47**, p. 61-68 (2000).
- P. Doukhan (1994). Mixing: Properties and examples. *Lecture Notes in Statistics 85*. Springer-Verlag.
- N. Down, S. P. Meyn, and R. L. Tweedie (1995). Exponential and uniform ergodicity of Markov processes. *Ann. Prob.*, **23(4)**, p. 1671-1691.
- C. Gardiner (1985). *Handbook of Stochastic Methods*. Springer Verlag, New York.
- 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.
- Y. Pokern, A. Stuart and P. Wiberg (2009). Parameter estimation for partially observed hypoelliptic diffusions. *J. Roy. Stat. Soc. B* **71**, p. 49-73.
- E. Rio (1995). About the Lindeberg method for strongly mixing sequences. *ESAIM, Probabilités et Statistiques* **1**, p. 35-61.

- E. Rio (2000). Théorie asymptotique des processus aléatoires faiblement dépendants. *Mathématiques & Applications* **31**, Springer.
- H. Risken (1989). The Fokker-Planck equation. Methods of solution and applications. Springer.
- A. Samson and M. Thieullen (2012). Contrast estimator for completely or partially observed hypoelliptic diffusion. *Stochastic Processes and their Applications* **122** (7), p. 2521-2552.
- D. Talay (2002). Stochastic Hamiltonian dissipative systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Processes and Related Fields* **8** (2), p. 163-198.
- C. Villani (2009). Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950).
- L. Wu (2001). Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic Processes and their Applications* **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

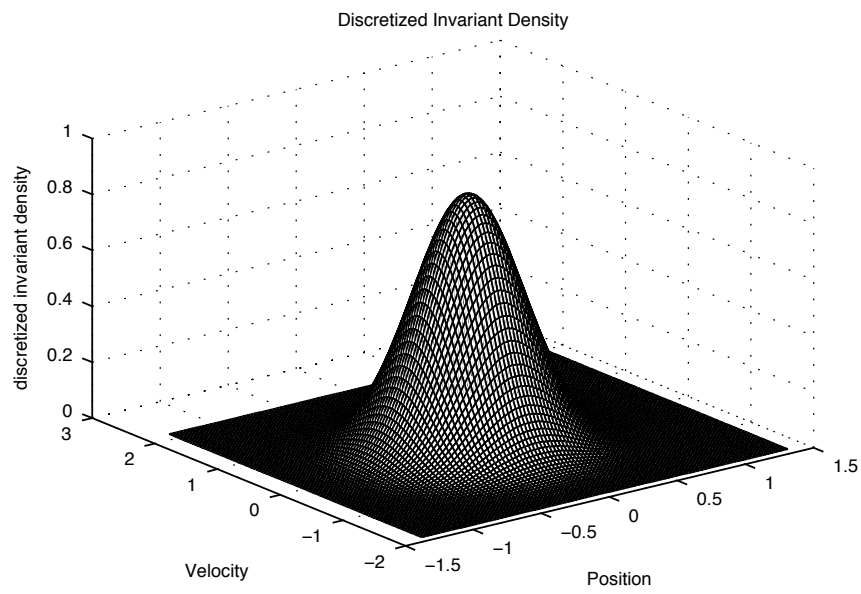


FIGURE 1. Harmonic oscillator, discretized theoretical bivariate invariant density.

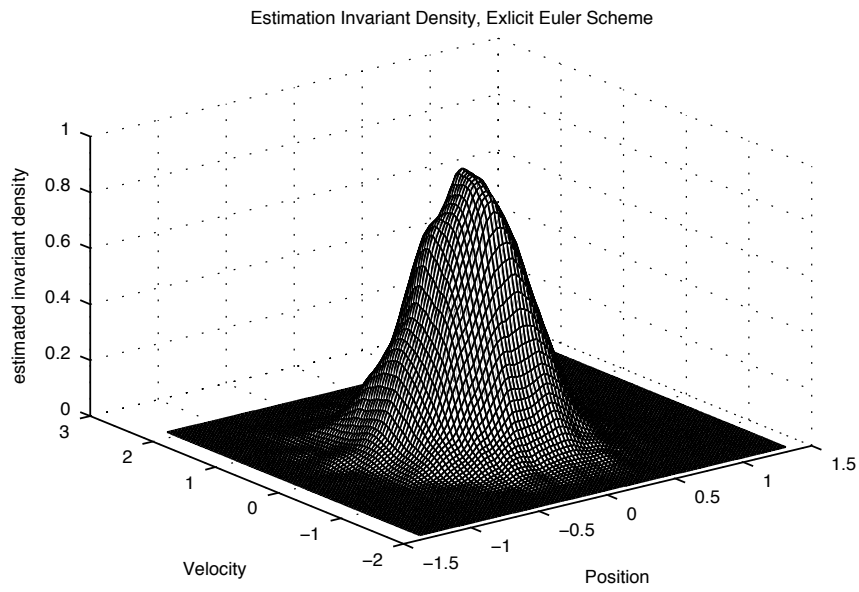


FIGURE 2. Harmonic oscillator, estimated bivariate invariant density.

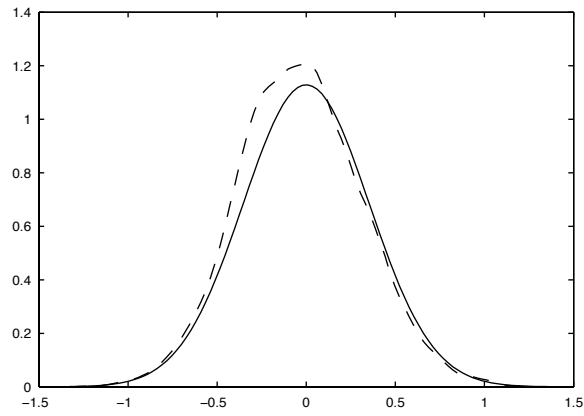


FIGURE 3. Harmonic oscillator, invariant density for the position

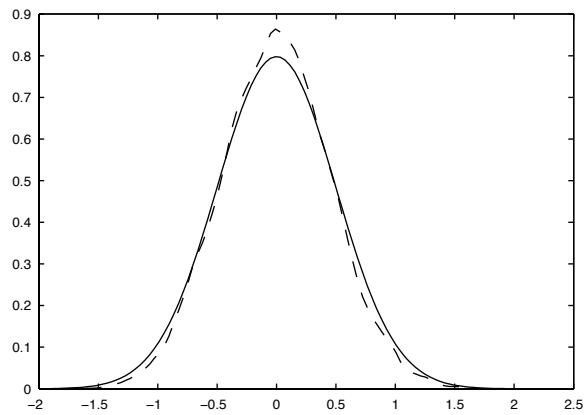


FIGURE 4. Harmonic oscillator, invariant density for the velocity

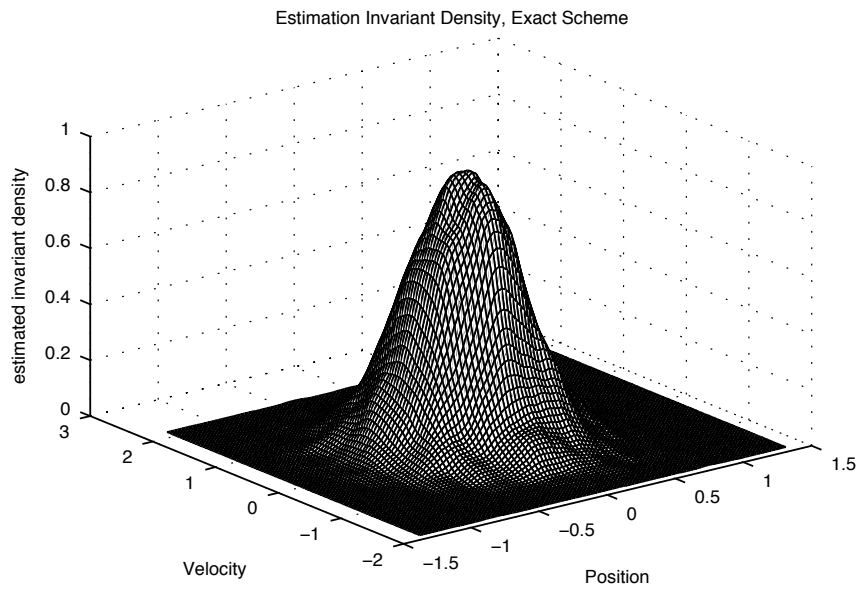


FIGURE 5. Harmonic oscillator, estimated bivariate invariant density, Exact scheme.

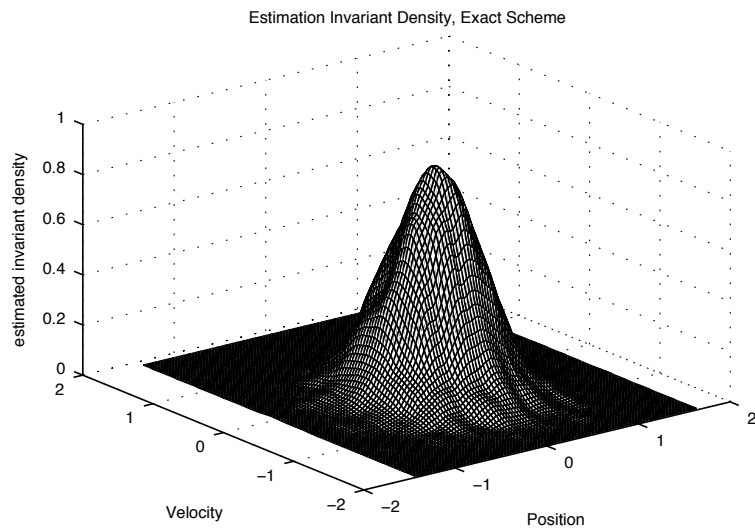


FIGURE 6. Harmonic oscillator, assumptions (ii) to (iv) in Theorem 3.3 satisfied, estimated bivariate stationary density

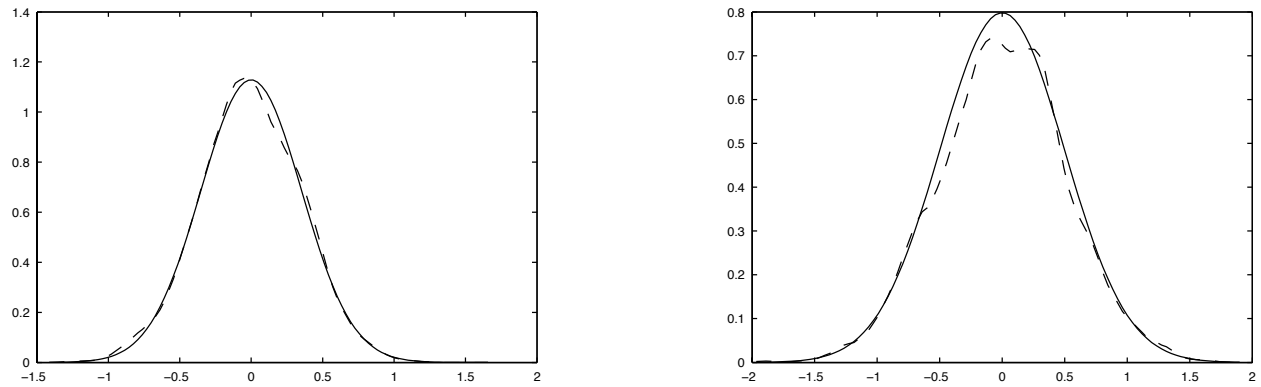


FIGURE 7. Harmonic oscillator, (ii) to (iv) in Theorem 3.3 satisfied, estimated marginals of the stationary density, position (left), velocity (right)

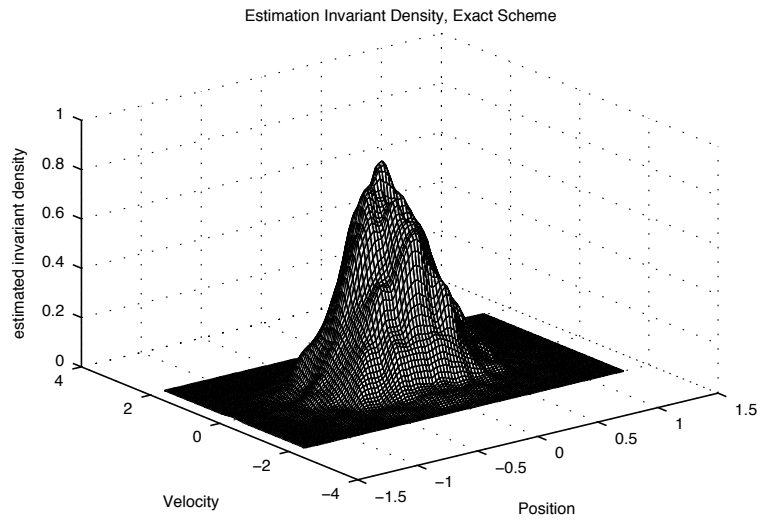


FIGURE 8. Harmonic oscillator, assumption (iii) in Theorem 3.3 not satisfied, estimated bivariate stationary density

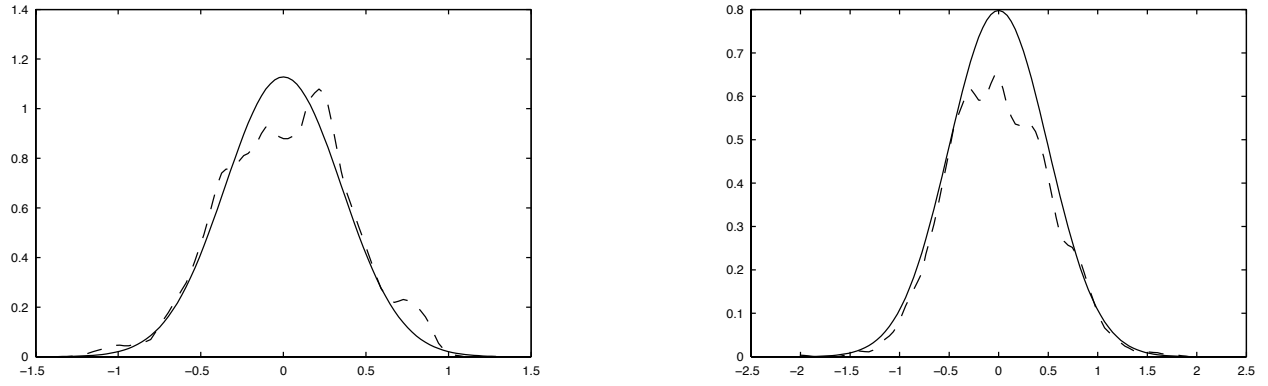


FIGURE 9. Harmonic oscillator, assump. (iii) in Theorem 3.3 not satisfied, estimated marginals of the stationary density, position (left), velocity (right)

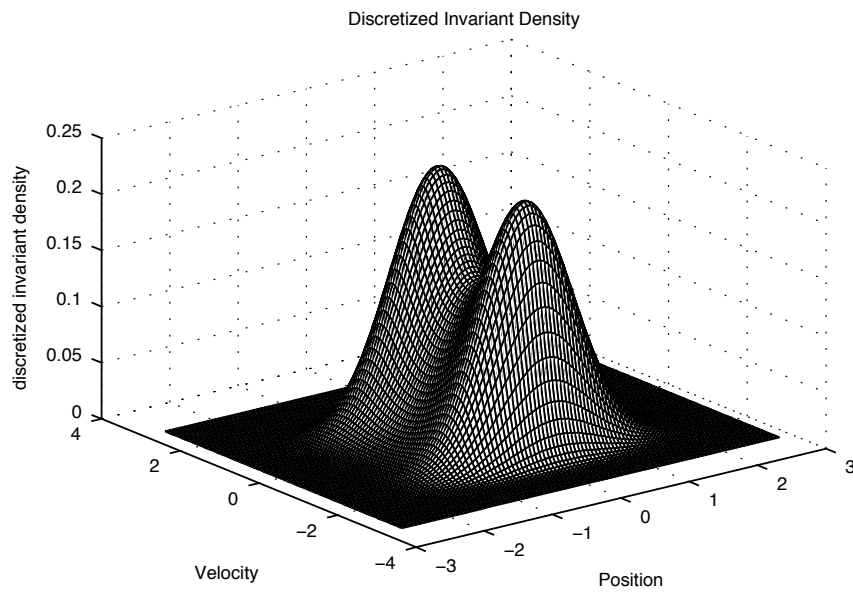


FIGURE 10. Duffing oscillator, discretized theoretical bivariate invariant density.

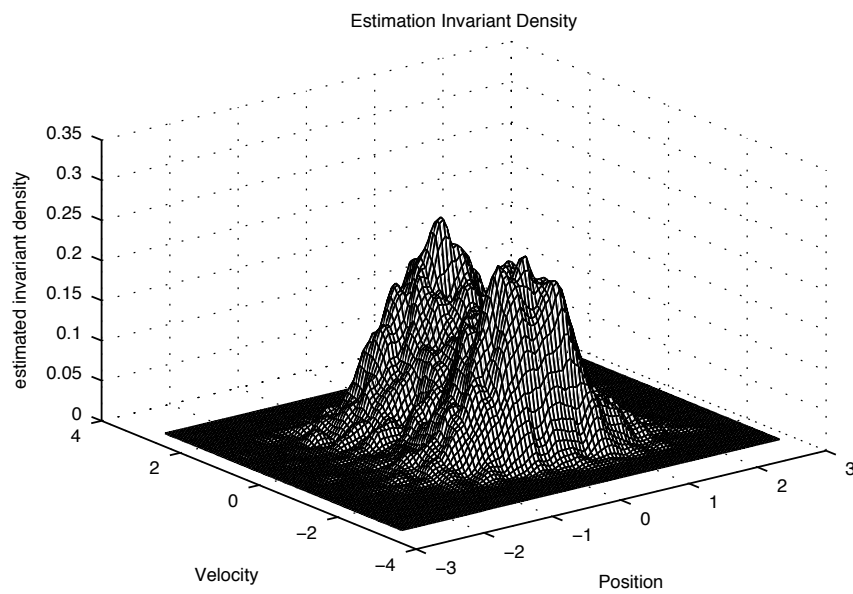


FIGURE 11. Duffing oscillator, estimated bivariate invariant density.

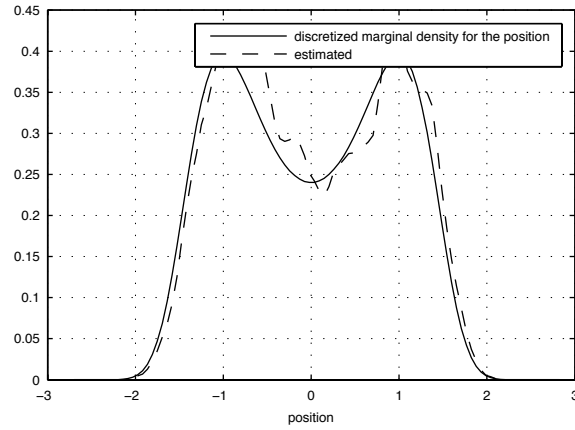


FIGURE 12. Duffing oscillator, invariant density for the position

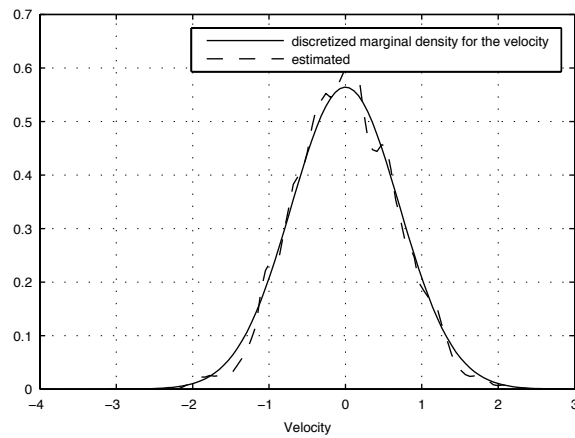


FIGURE 13. Duffing oscillator, invariant density for the velocity.