RATE OF ESTIMATION FOR THE STATIONARY DISTRIBUTION OF STOCHASTIC DAMPING HAMILTONIAN SYSTEMS WITH CONTINUOUS OBSERVATIONS

Sylvain Delattre, Arnaud Gloter, Nakahiro Yoshida

RATE OF ESTIMATION FOR THE STATIONARY DISTRIBUTION OF STOCHASTIC DAMPING HAMILTONIAN SYSTEMS WITH CONTINUOUS OBSERVATIONS

SYLVAIN DELATTRE, ARNAUD GLOTER, AND NAKAHIRO YOSHIDA

Abstract. We study the problem of the non-parametric estimation for the density π of the stationary distribution of a stochastic two-dimensional damping Hamiltonian system $(Z_t)_{t\in[0,T]} = (X_t, Y_t)_{t\in[0,T]}$. From the continuous observation of the sampling path on $[0,T]$, we study the rate of estimation for $\pi(x_0, y_0)$ as $T \to \infty$. We show that kernel based estimators can achieve the rate T^{-v} for some explicit exponent $v \in (0, 1/2)$. One finding is that the rate of estimation depends on the smoothness of π and is completely different with the rate appearing in the standard i.i.d. setting or in the case of two-dimensional non degenerate diffusion processes. Especially, this rate depends also on y_0. Moreover, we obtain a minimax lower bound on the L^2-risk for pointwise estimation, with the same rate T^{-v}, up to $\log(T)$ terms.

1. Introduction

The class of hypo-elliptic diffusion processes, for which the diffusion coefficient is degenerate, has been the subject of many recent works and is used for modeling in many fields, such as mathematical finance, biology, neuro-science, mechanics, ecology,... (see e.g. [8], [11], [7] and references therein). In this paper, we focus on the situation of a bi-dimensional hypo-elliptic process, describing the evolution in time of the couple position/velocity of some quantity. The velocity Y is modeled by a non degenerate one-dimensional diffusion process, while the position X is its integral, and the resulting bi-dimensional process $Z = (X, Y)$ is hypo-elliptic. Our aim is to estimate the density π of the stationary measure of this diffusion process, under the assumption of an ergodic setting.

The problem of non-parametric estimation of the stationary measure of a continuous mixing process is a long-standing problem (see for instance N'Guyen [13], or Comte and Merlevede [3] and references therein). Based on sample of length T of the data Comte and Merlevede [3] find estimators of the stationary measure, converging at rate depending on the smoothness of the stationary measure and slower than $\sqrt{T}$, as it is usual in non-parametric problems.

In the specific context where the continuous time process is a one-dimensional diffusion process, observed continuously on some interval $[0,T]$, the finding is different. It is shown that the rate of estimation of the stationary measure is $\sqrt{T}$ (see Kutoyants [10]). The rate of estimation is thus independent of the smoothness of the object that one estimates, in contrast to the typical non-parametric situation.

Date: 26, january, 2020.

2010 *Mathematics Subject Classification.* Primary 62G07, 62G20; secondary 60J60.

Key words and phrases. hypo-elliptic diffusion, non-parametric estimation, stationary measure, minimax rate.

The optimal estimator is very specific to the diffusive nature of the process as it relies on the local time of the process. Remark that if the process is a diffusion observed discretely on $[0,T]$ with a sufficiently high frequency it is possible to estimate with rate $\sqrt{T}$ also (see [12], [4]).

The case of multi-dimensional non-degenerate diffusions is treated in Dalalyan and Reiß [6] and Srauss [14]. In that case, the local time process is not available, but it is shown that for non degenerate diffusion process of dimension $d = 2$, there exists an estimator of the pointwise values of the stationary measure with rate $\sqrt{T}/\log(T)^2$. This rate of estimation does not depend on the smoothness of the stationary measure for $d = 2$. In the situation of a non degenerate diffusion with dimension $d \geq 3$, they find estimators whose rate is polynomial in T, depends on both the smoothness of the stationary measure and the dimension $d \geq 3$. For $d \geq 3$, the rate is strictly slower than $\sqrt{T}$, but faster than the rate appearing in standard multivariate density estimation from T i.i.d. observations. Hence, for $d \geq 3$, the diffusive structure of the process, enables to get a faster estimation rate than for the i.i.d. case, as well.

In the case of hypo-elliptic processes, fewer results exist for the estimation of the stationary distribution. In [5], the authors consider the case of two-dimensional process $Z = (X, Y)$, where Y is a velocity and X a position. Based on a discrete sampling of the path of size n, they propose an estimator of the stationary measure which converges at a non parametric rate depending on the smoothness of the stationary measure. Assuming that the stationary measure has anisotropic regularity (k_1, k_2), the proposed estimator has a rate depending on the harmonic mean $2/(1/k_1 + 1/k_2)$ as it is for the optimal rate of estimation of distribution on in the case of i.i.d. sequence.

In this paper, we focus on the situation where the process is observed continuously on $[0,T]$ and our goal is to determine what is the optimal rate of estimation of π in this context. Assuming that the stationary density $(x, y) \mapsto \pi(x, y)$ has an anisotropic Hölder regularity with index k_1 with respect to the variable x, and k_2 with respect to y, we construct an estimator of $\pi(x_0, y_0)$ based on $(X_t, Y_t)_{t\in[0,T]}$. This estimator achieves some rate $T^{-v(k_1,k_2)}$ and this rate of convergence depends on the smoothness of π in a very specific way. Indeed, the expression of the rate of estimation involves only k_1 or k_2, depending on the relative positions of these two smoothness indexes. This shows the specificity of the estimation problem for continuous observation of process $Z = (X, Y)$ with a non degenerate diffusive velocity Y, and a degenerate component X. It is noteworthy that we find a rate of estimation slower than in the case of continuous observation of a non-degenerate diffusion in dimension 2. Another interesting finding is that the rate of estimation of $\pi(x_0, y_0)$ depends on the point where one estimates the stationary measure, and is slower for points corresponding to null velocity $y_0 = 0$. A crucial ingredient in the study of the rate of estimation is to derive the variance of a kernel estimator with a choice of bandwidth (h_1, h_2). We show that the variance of the kernel estimator depends in a completely unsymmetrical way on h_1 and h_2. As a consequence, we get that the optimal bandwidth choice for the estimator is such that one of the two bandwidth can go almost arbitrarily fast to 0, while the optimal choice for the other bandwidth depends sharply on the smoothness of the stationary measure.

Also, we show a lower bound for the minimax risk of estimation of $\pi(x_0, y_0)$ on a class of hypo-elliptic diffusion models with stationary measure of Hölder regularity

(k_1, k_2). This proves that it is impossible to estimate uniformly on this class with a rate faster than $T^{-v(k_1,k_2)}$ (up to $\log(T)$ term).

The outline of the paper is the following. In Section 2, we present the model and give assumptions that are sufficient to get an ergodic system with stationary measure admitting a density π. In Section 3, we present the construction of the estimator and states the results on their rate of convergence (in Theorem 1 for $y_0 \neq 0$, and Theorem 2 for $y_0 = 0$). In Section 4, we prove the upper bound on the variance of the kernel estimator. We also illustrate the very specific behaviour of the variance of the kernel estimator for hypo-elliptic diffusion by numerical simulations. In Section 5, we state and prove minimax lower bounds for the risk of estimation. In the Appendix, we prove some technical results used in the proofs of Section 4.

2. Hamiltonian system and mixing property

Let us consider $(\Omega, \mathcal{A}, \mathbb{P})$, some probability space on which a standard one dimensional Brownian motion $(B_t)_{t\geq 0}$ is defined. We assume that the process $(Z_t)_{t\geq 0} = (X_t, Y_t)_{t\geq 0}$ is solution of the stochastic differential equation

$$\mathrm{d}X_t = Y_t \mathrm{d}t \tag{2.1}$$

$$\mathrm{d}Y_t = a(X_t, Y_t)\mathrm{d}B_t - [\beta(X_t, Y_t)Y_t + V'(X_t)]\mathrm{d}t, \tag{2.2}$$

where (X_0, Y_0) is a random variable independent of $(B_t)_t$.

We introduce the following regularity assumption on the coefficients.

Assumption **HReg**:

- The functions $a : \mathbb{R}^2 \to \mathbb{R}$ is a $\mathcal{C}^\infty$ function and $\overline{a} > a(x, y) \geq \underline{a} > 0$, and for some constants $\overline{a} > \underline{a} > 0$.
- The function $\beta : \mathbb{R}^2 \to \mathbb{R}$ is continuously differentiable, and such that $|\beta(x, y)| \leq \overline{\beta}$ for all $(x, y) \in \mathbb{R}^2$, and for some constant $\overline{\beta} > 0$. Moreover, we have $\beta(x, y) > \underline{\beta}$, $\forall x \in [l, \infty), y \in \mathbb{R}$, where $l \geq 0$, $\underline{\beta} > 0$ are two constants.
- The function $V : \mathbb{R} \to \mathbb{R}$ is lower bounded, and with $\mathcal{C}^2$ regularity.

It is shown in [17], that under the assumptions **HReg** the S.D.E. (2.1)–(2.2) admits a weak solution, which satisfies the Markov property, and the associated semi group $(P_t)_{t\geq 0}$ is strongly Feller [17], which in turn implies that the process is strongly Markovian. Let us stress that the sign condition on β for large x together with the existence of lower bound on V are crucial to insure that the solution of (2.1)–(2.2) does not explode in finite time. Of course, if we know that $(x, y) \mapsto a(x, y)$, $(x, y) \mapsto y\beta(x, y)$ and $x \mapsto V'(x)$ are globally Lipschitz, the solutions of the S.D.E. exists in the strong sense.

We now introduce an assumption on the potential V of the system that ensure that the process tends to some equilibrium.

Assumption **HErg**: one has $\lim_{|x|\to\infty} V'(x)\mathrm{sign}(x) = +\infty$.

Is is shown in [17] that under **HReg** and **HErg**, one can construct a Lyapounov function $\Psi \geq 1$, and that a stationary probability π exists and is unique for the process $Z = (X, Y)$, and satisfies $\pi(\Psi) < \infty$. It is shown in [17] (see Theorem 2.4) that for some $D > 0$ and $\rho > 0$,

$$\forall t \geq 0, \quad \forall z \in \mathbb{R}^2, \quad \sup_{|f|\leq\Psi} \left| P_t(f)(z) - \int_{\mathbb{R}^2} f(z')\pi(\mathrm{d}z') \right| \leq D\Psi(z)e^{-\rho t} \tag{2.3}$$

for any function measurable function f such as f/Ψ is bounded on $\mathbb{R}$ and where $(P_t)_{t\geq 0}$ is the semi group of Z,

$$P_t(f)(z) = E[f(X_t, Y_t) \mid (X_0, Y_0) = z].$$

Remark that under **HReg** and **HErg**, it is possible to construct a Lyapounov function such that $\Psi(x,y) \geq \frac{1}{C}\exp\left(\frac{1}{C}[|y|^2 + V(x)]\right)$, for some constant $C > 0$ (see (3.10) in [17]). As a consequence, (2.3) applies to functions f with exponential growth, and the invariant distribution admits finite exponential moments.

Hence, we can state the following proposition

Proposition 1. *Assume that the coefficients of the equation (2.2) satisfy* **HReg** *and* **HErg**, *then there exists a stationary solution $Z = (X, Y)$ to the S.D.E. (2.1)–(2.2), and the stationary distribution is unique and admits some density π. Moreover, there exist constants $D_{erg} > 0$ and $\rho > 0$ such that for any bounded measurable functions f, g, we have*

$$\forall t \geq 0, \quad |cov(f(Z_0), g(Z_t))| \leq D_{erg}\|f\|_\infty \|g\|_\infty e^{-\rho t}. \tag{2.4}$$

The Proposition (1) is a consequence of the results shown in [17]. Indeed, from the fact that the Lyapounov function Ψ is greater than 1 and integrable with respect to the stationary measure, one can check that (2.4) is a consequence of (2.3).

3. Estimator and upper bounds

In this section we introduce the expression for our estimator of the stationary measure π of the S.D.E. (2.1)–(2.2) and prove that the estimator achieves some rate of convergence, depending on the smoothness of π.

Let $\varphi : \mathbb{R} \to \mathbb{R}$ a bounded, compactly supported function. For convenience, we suppose that the support of φ is $[-1, 1]$. We assume that

$$\int_{\mathbb{R}} \varphi(u)\mathrm{d}u = 1, \quad \int_{\mathbb{R}} \varphi(u)u^l \mathrm{d}u = 0, \text{ for } l \in \{1, \dots, L\} \text{ where } L \geq 1. \tag{3.1}$$

We let $h_1(T) > 0$, $h_2(T) > 0$ be two bandwidths which converge to zero as $T \to \infty$, and we consider a kernel estimator of π at the point $(x_0, y_0) \in \mathbb{R}^2$ as

$$\hat{\pi}_T(x_0, y_0) = \frac{1}{T}\int_0^T \varphi_{h_1(T),h_2(T)}(X_s - x_0, Y_s - y_0)\mathrm{d}s, \tag{3.2}$$

where

$$\varphi_{h_1,h_2}(x - x_0, y - y_0) = \frac{1}{h_1 h_2}\varphi(\frac{x - x_0}{h_1})\varphi(\frac{y - y_0}{h_2}). \tag{3.3}$$

We assume that the two bandwidths satisfy,

$$\exists K > 0, \quad h_1(T)^{-1} + h_2(T)^{-1} \leq K(1 + T^K), \ \forall T > 0, \tag{3.4}$$

$$\sqrt{h_1(T)} + h_2(T) \leq K(\log(T))^{-3/2} \wedge 1, \ \forall T > 1. \tag{3.5}$$

The two previous conditions insure that the bandwidths go faster to zero than the logarithmic rate by (3.5), but not faster than any polynomial rates by (3.4). Actually these two bandwidths will be specified later (see equations (3.12), (3.13), (3.14), (3.15) in the proofs of Theorems 1 and 2).

We introduce the class of Hölder functions.

Definition 1. *For $(k_1,k_2)\in(0,\infty)^2$, and $R>0$, we denote $\mathcal{H}^{k_1,k_2}(R)$ the set of functions $f:\mathbb{R}^2\to\mathbb{R}$ such that $x\mapsto f(x,y)$ and $y\mapsto f(x,y)$ are respectively of class $\mathcal{C}^{\lfloor k_1\rfloor}$ and $\mathcal{C}^{\lfloor k_2\rfloor}$, and satisfy the control, $\forall x,y$ in $\mathbb{R}$ and $h\in[-1,1]$,*

$$
\begin{aligned}
&|f(x,y)|\le R,\\
&\left|\frac{\partial^{\lfloor k_1\rfloor}f}{\partial x^{\lfloor k_1\rfloor}}(x+h,y)-\frac{\partial^{\lfloor k_1\rfloor}f}{\partial x^{\lfloor k_1\rfloor}}(x,y)\right|\le R|h|^{k_1-\lfloor k_1\rfloor},\\
&\left|\frac{\partial^{\lfloor k_2\rfloor}f}{\partial x^{\lfloor k_2\rfloor}}(x,y+h)-\frac{\partial^{\lfloor k_2\rfloor}f}{\partial x^{\lfloor k_2\rfloor}}(x,y)\right|\le R|h|^{k_2-\lfloor k_2\rfloor}.
\end{aligned}
$$

We can state the main results on the asymptotic behaviour of the estimator. This behaviour is different according to the fact that we estimate the value of the stationary measure on a point (x_0,y_0) corresponding to a null velocity or not.

Theorem 1. *Assume that $Z=(X,Y)$ is a stationary solution to (2.1)–(2.2) and that Assumptions* **HReg**, **HErg** *hold true. We assume that the stationary distribution π belongs to the set $\mathcal{H}^{k_1,k_2}(R)$ for $k_1>0,k_2>0,R>0$, with $\max(k_1,k_2)\le L$ (recall (3.1)).*

Assume that $y_0\neq 0$. Then, there exist bandwidths $(h_1(T))_T$, $(h_2(T))_T$, depending only on k_1 and k_2, such that the estimator satisfies :

$$
\text{if } k_1<k_2/2,\quad E\left[(\hat\pi_T(x_0,y_0)-\pi(x_0,y_0))^2\right]\le CT^{-\frac{2k_2}{2k_2+1}}, \tag{3.6}
$$

$$
\text{if } k_1\ge k_2/2,\quad E\left[(\hat\pi_T(x_0,y_0)-\pi(x_0,y_0))^2\right]\le CT^{-\frac{2k_1}{2k_1+1/2}}, \tag{3.7}
$$

for some constant C independent of T.

Theorem 2. *Assume that $Z=(X,Y)$ is a stationary solution to (2.1)–(2.2) and that Assumptions* **HReg**, **HErg** *hold true. We assume that the stationary distribution π belongs to the set $\mathcal{H}^{k_1,k_2}(R)$ for $k_1>0,k_2>0,R>0$, with $\max(k_1,k_2)\le L$ (recall (3.1)).*

Assume that $y_0=0$. Then, there exist bandwidths $(h_1(T))_T$, $(h_2(T))_T$, depending only on k_1 and k_2, such that the estimator satisfies :

$$
\text{if } k_1<k_2/3,\quad E\left[(\hat\pi_T(x_0,y_0)-\pi(x_0,y_0))^2\right]\le C\Big(\frac{T}{\log T}\Big)^{-\frac{2k_2}{2k_2+2}}, \tag{3.8}
$$

$$
\text{if } k_1\ge k_2/3,\quad E\left[(\hat\pi_T(x_0,y_0)-\pi(x_0,y_0))^2\right]\le CT^{-\frac{2k_1}{2k_1+2/3}}, \tag{3.9}
$$

for some constant C independent of T.

Remark 1. *The rates of estimation obtained in Theorems 1–2 are completely different with the usual one in several ways. First, they do not depend on the harmonic mean of the smoothness index k_1, k_2, as it is usual in non-parametric setting. Second, the rate depends on the point (x_0,y_0) where the density is estimated. We state in Section 5 a minimax lower bound for the L^2 risk of estimation of $\pi(x_0,y_0)$ with the same rates (up to* log *terms).*

The asymptotic behaviour of the estimator relies on the standard bias variance decomposition. Hence, we need sharp evaluations for the variance of the estimator, that are stated below, and will be proved in Section 4.

Proposition 2. *Assume that $Z=(X,Y)$ is solution to (2.1)–(2.2), that Assumptions* **HReg**, **HErg**, *hold true and $\|\pi\|_\infty\le R$ for some $R>0$.*

Assume that $y_0 \neq 0$. Then, there exists some constant C, such that for all $T > 0$,

$$\mathrm{Var}(\hat{\pi}_T(x_0, y_0)) \leq C\left(\frac{1}{Th_2(T)} \wedge \frac{1}{T\sqrt{h_1(T)}}\right) + C\varepsilon(T, h_1(T), h_2(T)),$$

where

$$\varepsilon(T, h_1(T), h_2(T)) \leq \frac{|\log(h_1(T)h_2(T))|^C}{T}. \tag{3.10}$$

Proposition 3. *Assume that $Z = (X, Y)$ is a solution to (2.1)–(2.2), that Assumptions* **HReg**, **HErg**, *hold true and that $\|\pi\|_\infty \leq R$ for some $R > 0$.*

Assume that $y_0 = 0$. Then, there exists some constant C, such that for all $T > 0$,

$$\mathrm{Var}(\hat{\pi}_T(x_0, y_0)) \leq C\left(\frac{\ln(T)}{Th_2(T)^2} \wedge \frac{1}{Th_1(T)^{2/3}}\right) + C\varepsilon(T, h_1(T), h_2(T)),$$

where

$$\varepsilon(T, h_1(T), h_2(T)) \leq \frac{|\log(h_1(T)h_2(T))|^C}{T}.$$

We can now prove the that our estimator achieves the rates given in Theorems 1–2.

Proof [Proof of Theorem 1.] We write the usual bias-variance decomposition,

$$E[(\hat{\pi}_T(x_0, y_0) - \pi(x_0, y_0))^2] \leq |E(\hat{\pi}_T(x_0, y_0)) - \pi(x_0, y_0)|^2 + \mathrm{Var}(\hat{\pi}_T(x_0, y_0)). \tag{3.11}$$

Using the stationarity of the process, we can upper bound the bias term as

$$\begin{aligned}|E(\hat{\pi}_T(x_0, y_0)) - \pi(x_0, y_0)|^2 &= \left(\int_{\mathbb{R}^2} \varphi(u)\varphi(v)[\pi(x_0 + uh_1(T), y_0 + vh_2(T)) - \pi(x_0, y_0)]\mathrm{d}u\mathrm{d}v\right) \\ &\leq C(h_1(T)^{2k_1} + h_2(T)^{2k_2}),\end{aligned}$$

where in the last line we used $\pi \in \mathcal{H}^{k_1,k_2}(R)$ with (3.1) (see e.g. [16], or Proposition 1 in [2] for details). We now use the results of Proposition 2 on the variance of the estimator and choose the optimal bandwidths $h_1(T)$, $h_2(T)$.

- Case 1: $k_1 < k_2/2$.

Using Proposition 2

$$E[(\hat{\pi}_T(x_0, y_0) - \pi(x_0, y_0))^2] \leq C(h_1(T)^{2k_1} + h_2(T)^{2k_2}) + \frac{C}{Th_2(T)} + C\varepsilon(T, h_1(T), h_2(T)).$$

We now choose to balance $h_2(T)^{2k_2}$ with the main contribution of the variance term and let the contribution of $h_1(T)$ on the bias be smaller. It yields us to set

$$h_2(T) = T^{-1/(2k_2+1)}, \quad h_1(T) = T^{-C_1}, \text{ where } C_1 \geq \frac{k_2}{k_1(2k_2+1)}. \tag{3.12}$$

With these choices and recalling (3.10), we get (3.6).

- Case 2: $k_1 \geq k_2/2$.

We use again Proposition 2:

$$E[(\hat{\pi}_T(x_0,y_0)-\pi(x_0,y_0))^2] \leq C(h_1(T)^{2k_1}+h_2(T)^{2k_2}) + \frac{C}{T\sqrt{h_1(T)}} + C\varepsilon(T,h_1(T),h_2(T)).$$

Balancing the variance and bias terms yields to

$$h_1(T)=T^{-1/(2k_1+1/2)}, \quad h_2(T)=T^{-C_2}, \text{ where } C_2 \geq \frac{k_1}{k_2(2k_1+1/2)}, \tag{3.13}$$

and (3.7) follows. □

Proof [Proof of Theorem 2] We use again the bias/variance decomposition (3.11) and exploit now the results of Proposition 3.

• Case 1: $k_1 < k_2/3$.

We have that,

$$E[(\hat{\pi}_T(x_0,y_0)-\pi(x_0,y_0))^2] \leq C(h_1(T)^{2k_1}+h_2(T)^{2k_2}) + \frac{C\ln(T)}{Th_2(T)^2} + C\varepsilon(T,h_1(T),h_2(T)).$$

We set

$$h_2(T)=\left(\frac{T}{\ln(T)}\right)^{-1/(2k_2+2)}, \quad h_1(T)=T^{-C_1}, \text{ where } C_1 \geq \frac{k_2}{k_1(2k_2+2)}, \tag{3.14}$$

and (3.8) follows.

• Case 2: $k_1 \geq k_2/3$.

We have that,

$$E[(\hat{\pi}_T(x_0,y_0)-\pi(x_0,y_0))^2] \leq C(h_1(T)^{2k_1}+h_2(T)^{2k_2}) + \frac{C}{Th_1(T)^{2/3}} + C\varepsilon(T,h_1(T),h_2(T)).$$

The choice

$$h_1(T)=T^{-1/(2k_1+2/3)}, \quad h_2(T)=T^{-C_2}, \text{ where } C_2 \geq \frac{k_1}{k_2(2k_1+2/3)} \tag{3.15}$$

gives (3.9). □

Remark 2.

- *The optimal choices for the bandwidths are given in (3.12), (3.13), (3.14) and (3.15). In all the situations, we see that one of the two bandwidths $h_1(T)$ or $h_2(T)$ can be chosen "arbitrarily small" (as the constants C_1 and C_2 in (3.12), (3.13), (3.14), (3.15) can be arbitrarily large). In means that the bias induced by the variation of π along one of the two variables x or y can be arbitrarily reduced by the choice of a very thin bandwidth. It explains why the expression of the rate of estimation depends only on one index of smoothness k_1 or k_2.*
- *The fact that one of the two bandwidths can be chosen arbitrarily small is reminiscent to the situation of the estimation of the stationary measure $\pi(z)$ for a one dimensional diffusion process $(Z_t)_{t\in[0,T]}$ observed continuously. In that case, the efficient estimator is based on the local time of the process (see [10]) and the rate is $\sqrt{T}$ independently of the smoothness of π. The*

use of local time is a way to give a rigorous analysis of the quantity $\tilde{\pi}_T(z) = \frac{1}{T}\int_0^T \delta_{\{z\}}(Z_s)ds$, *where* $\delta_{\{z\}}$ *is the Dirac mass located at* z. *We see that* $\tilde{\pi}_T$ *is essentially a kernel estimator with bandwidth* $h = 0$, *for which the bias is reduced to* 0.

4. Variance of the kernel estimator

In this section we prove the crucial upper bounds given in Propositions 2–3. We first need to state two lemmas related to the behaviour of density and semi group of the process, and whose proofs are postponed to Section 6.

Lemma 1 (Corollary 2.12 in [1])**.** *Assume* **HReg**. *Then, the process admits a transition density* $(p_t)_{t>0}$ *which satisfies the following upper bound. For all* K *compact subset of* $\mathbb{R}^2$, $\forall (x,y,x',y') \in K \times K$, $\forall t \in (0,1)$

$$p_t((x,y);(x',y')) \leq p_t^G((x,y);(x',y')) + p_t^U((x,y);(x',y')) \tag{4.1}$$

where

$$p_t^G((x,y);(x',y')) = \frac{C_G}{t^2}\exp\left(-\frac{1}{C_G}\left[\frac{(y-y')^2}{t} + \frac{(x'-x-\frac{y+y'}{2}t)^2}{t^3}\right]\right), \tag{4.2}$$

for some $C_G > 0$ *and* p^U *is a measurable non negative function such that for any compact* $K \subset \mathbb{R}^2$ *and* $(x,y) \in K$ *we have for all* $t \in (0,1)$,

$$\int_{\mathbb{R}^2} p_t^U((x,y);(x',y'))dx'dy' \leq C_U \exp(-t^{-1}C_U^{-1}), \tag{4.3}$$

for $C_U > 0$. *The two constants* C_G *and* C_U *are independent of* $t \in [0,1]$, *but depend on the compact set* K.

The Lemma 1 gives us a control on the short time behaviour of the transition density. For the sequel, we need a control valid for any time. This is the purpose of the following lemma about the semi group of the process.

Lemma 2. *Assume* **HReg**, *and let* K *be a compact subset of* $\mathbb{R}^2$. *Then, there exists a constant* $\widetilde{C}_K$ *such that for all* $0 < D < 1$, $t \geq D$, $\forall z \in \mathbb{R}^2$, *and any* f *measurable bounded function with support on* K,

$$|P_t(f)(z)| \leq \widetilde{C}_K\left[\frac{\|f\|_{L^1(\mathbb{R}^2)}}{D^2} + \|f\|_\infty e^{-1/(\widetilde{C}_K D)}\right]. \tag{4.4}$$

4.1. **Proof or Proposition 2.** Throughout the proof we suppress in the notation the dependence upon T of $h_1(T)$ and $h_2(T)$. The constant C may change from line to line and is independent of T. In the proof, we will use repeatedly Lemmas 1–2. To this end, we consider a compact set K that contains a ball of radius $\sqrt{2}$ centered at (x_0,y_0), and as a result the support of $(x,y) \mapsto \varphi(x-x_0,y-y_0)$ is included in this compact.

To prove the proposition, it is sufficient to prove that the following inequalities holds both, for T large enough,

$$\mathrm{Var}(\hat{\pi}_T(x_0,y_0)) \leq C\frac{1}{Th_2} + C\varepsilon(T,h_1(T),h_2(T)), \tag{4.5}$$

$$\mathrm{Var}(\hat{\pi}_T(x_0,y_0)) \leq C\frac{1}{T\sqrt{h_1}} + C\varepsilon(T,h_1(T),h_2(T)). \tag{4.6}$$

First step: we prove (4.5).
From (3.2), and the stationarity of the process we get that

$$\mathrm{Var}\left(\hat{\pi}_T(x_0,y_0)\right) = \frac{1}{T^2}\int_0^T\int_0^T \kappa(t-s)\mathrm{d}t\mathrm{d}s, \tag{4.7}$$

where

$$\kappa(u) = \mathrm{Cov}\left(\varphi_{h_1,h_2}(X_0-x_0,Y_0-y_0),\varphi_{h_1,h_2}(X_u-x_0,Y_u-y_0)\right).$$

We deduce that

$$\mathrm{Var}\left(\hat{\pi}_T(x_0,y_0)\right) \leq \frac{1}{T}\int_0^T |\kappa(s)|\mathrm{d}s. \tag{4.8}$$

We will find an upper bound for the integral on the right-hand side of the latter expression by splitting the time interval $[0,T]$ into 4 pieces $[0,T]=[0,\delta)\cup[\delta,D_1)\cup[D_1,D_2)\cup[D_2,T]$, where δ, D_1, D_2, will be chosen latter.

• For $s\in[0,\delta)$, we write from (4.7) and using Cauchy-Schwarz inequality and the stationarity of the process,

$$\begin{aligned}|\kappa(s)| &\leq \mathrm{Var}(\varphi_{h_1,h_2}(X_0-x_0,Y_0-y_0))^{1/2}\,\mathrm{Var}(\varphi_{h_1,h_2}(X_s-x_0,Y_s-y_0))^{1/2}\\ &= \mathrm{Var}(\varphi_{h_1,h_2}(X_0-x_0,Y_0-y_0)).\end{aligned}$$

This variance is smaller than

$$\int_{\mathbb{R}^2}\varphi_{h_1,h_2}(x-x_0,y-y_0)^2\pi(x,y)\mathrm{d}x\mathrm{d}y$$

and using that π is bounded and (3.3), we deduce

$$|\kappa_s| \leq \frac{C}{h_1h_2}. \tag{4.9}$$

In turn, we have

$$\int_0^\delta |\kappa(s)|\mathrm{d}s \leq C\frac{\delta}{h_1h_2}. \tag{4.10}$$

• For $s\in[\delta,D_1)$, where $\delta<D_1\leq 1$. We write

$$\begin{aligned}|\kappa(s)| \leq\ & E\left[|\varphi_{h_1,h_2}(X_0-x_0,Y_0-y_0)||\varphi_{h_1,h_2}(X_u-x_0,Y_u-y_0)|\right] +\\ & E\left[|\varphi_{h_1,h_2}(X_0-x_0,Y_0-y_0)|\right]E\left[|\varphi_{h_1,h_2}(X_s-x_0,Y_s-y_0)|\right]\end{aligned}$$

Using that $(X_t,Y_t)_t$ is stationary with marginal law having a bounded density we deduce that $E\left[|\varphi_{h_1,h_2}(X_s-x_0,Y_s-y_0)|\right] \leq C\int_{\mathbb{R}^2}|\varphi_{h_1,h_2}(x-x_0,y-y_0)|\mathrm{d}x\mathrm{d}y \leq C$ from (3.3). This gives,

$$\begin{aligned}|\kappa(s)| \leq & \int_{\mathbb{R}^2}|\varphi_{h_1,h_2}(x-x_0,y-y_0)|\\ & \int_{\mathbb{R}^2}|\varphi_{h_1,h_2}(x'-x_0,y'-y_0)|p_s(x,y;x',y')\mathrm{d}x'\mathrm{d}y'\pi(x,y)\mathrm{d}x\mathrm{d}y + C.\end{aligned} \tag{4.11}$$

Using now equation (4.1) in Lemma 1, we get that

$$|\kappa(s)| \leq \overline{\kappa}^1(s) + \overline{\kappa}^2(s) + C \tag{4.12}$$

with

$$\overline{\kappa}^1(s) := \int_{\mathbb{R}^2} |\varphi_{h_1,h_2}(x-x_0,y-y_0)| \int_{\mathbb{R}^2} |\varphi_{h_1,h_2}(x'-x_0,y'-y_0)| p_s^G(x,y;x',y')\mathrm{d}x'\mathrm{d}y'\pi(x,y)\mathrm{d}x\mathrm{d}y, \tag{4.13}$$

$$\overline{\kappa}^2(s) := \int_{\mathbb{R}^2} |\varphi_{h_1,h_2}(x-x_0,y-y_0)| \int_{\mathbb{R}^2} |\varphi_{h_1,h_2}(x'-x_0,y'-y_0)| p_s^U(x,y;x,y)\mathrm{d}x'\mathrm{d}y'\pi(x,y)\mathrm{d}x\mathrm{d}y.$$

In order to upper bound $\overline{\kappa}^1(s)$ we show that the Gaussian kernel (4.2) appearing in the expression of $\overline{\kappa}^1(s)$ takes small values for $s \in [\delta, D_1)$ as soon as δ is well chosen. Recall that $y_0 \neq 0$, and for simplicity assume that $y_0 > 0$. Then, using that φ is compactly supported on$[-1,1]$ we know that $|\varphi_{h_1,h_2}(x-x_0,y-y_0)\varphi_{h_1,h_2}(x'-x_0,y'-y_0)| \neq 0$ implies that

$$|x-x_0| \leq h_1, |x'-x_0| \leq h_1, |y-y_0| \leq h_2, |y'-y_0| \leq h_2. \tag{4.14}$$

Let us denote $K(h_1,h_2)$ the rectangle of $\mathbb{R}^4$ defined by the conditions (4.14). Then,

$$\overline{\kappa}^1(s) \leq C \int_{K(h_1,h_2)} |\varphi_{h_1,h_2}(x-x_0,y-y_0)| |\varphi_{h_1,h_2}(x'-x_0,y'-y_0)| p_s^G(x,y;x',y')\mathrm{d}x\mathrm{d}y\mathrm{d}x'\mathrm{d}y', \tag{4.15}$$

where we used that π is bounded. On $K(h_1,h_2)$, we have $\frac{y+y'}{2} \geq \frac{y_0}{2} > 0$ if h_2 is small enough, and $|x-x'| \leq 2h_1$. Hence, if we assume that $s \geq \frac{6h_1}{y_0}$ we have $|x'-x| \leq \frac{sy_0}{3}$. It entails, $x'-x-\frac{y+y'}{2}s \leq \frac{sy_0}{3} - \frac{y_0}{2}s = -\frac{sy_0}{6}$, and in turn $\frac{(x'-x-\frac{y+y'}{2}s)^2}{s^3} \geq \frac{y_0^2}{36}\frac{1}{s}$. Plugging in (4.2) this yields to $p_s^G(x,y;x';y') \leq \frac{C}{s^2}\exp(-\frac{1}{Cs})$ for some constant C independent of s. Using (4.15), we deduce

$$\overline{\kappa}^1(s) \leq \frac{C}{s^2}\exp(-\frac{1}{Cs}) \int_{K(h_1,h_2)} |\varphi_{h_1,h_2}(x-x_0,y-y_0)| |\varphi_{h_1,h_2}(x'-x_0,y'-y_0)|\mathrm{d}x\mathrm{d}y\mathrm{d}x'\mathrm{d}y'$$

and hence

$$\overline{\kappa}^1(s) \leq \frac{C}{s^2}\exp(-\frac{1}{Cs}). \tag{4.16}$$

To control $\overline{\kappa}^2(s)$, we use (4.3) and $\|\varphi_{h_1,h_2}(\cdot-x_0,\cdot-y_0)\|_\infty \leq \frac{C}{h_1h_2}$ to get that for all x,y in the compact K containing a ball of radius $\sqrt{2}$ centered at (x_0,y_0), we have the upper bound $\int_{\mathbb{R}^2} |\varphi_{h_1,h_2}(x'-x_0,y'-y_0)| p_s^U(x,y;x',y')\mathrm{d}x'\mathrm{d}y' \leq \frac{C}{h_1h_2}e^{-\frac{1}{Cs}}$. As a consequence,

$$\begin{aligned}\overline{\kappa}^2(s) &\leq C\int_{\mathbb{R}^2} |\varphi_{h_1,h_2}(x-x_0,y-y_0)| \frac{1}{h_1h_2}e^{-\frac{1}{Cs}}\pi(x,y)\mathrm{d}x\mathrm{d}y \\ &\leq \frac{C}{h_1h_2}e^{-\frac{1}{Cs}},\end{aligned} \tag{4.17}$$

where we used again that π is bounded and that the support of $\varphi_{h_1,h_2}(\cdot - x_0, \cdot - y_0)$ is included in K. From (4.12), (4.16)–(4.17), we deduce that for $\frac{6h_1}{y_0} \leq \delta \leq D_1 \leq 1$,

$$\begin{aligned}\int_\delta^{D_1} |\kappa(s)|\mathrm{d}s &\leq \int_\delta^{D_1} [\frac{C}{s^2}\exp(-\frac{1}{Cs}) + \frac{C}{h_1h_2}\exp(-\frac{1}{Cs}) + C]\mathrm{d}s \\ &\leq \int_\delta^{D_1} [\frac{C}{s^2}\exp(-\frac{1}{Cs}) + \frac{C}{h_1h_2s^2}\exp(-\frac{1}{Cs}) + C]\mathrm{d}s \\ &\leq C\exp(-\frac{1}{CD_1})[1+\frac{1}{h_1h_2}] + CD_1, \end{aligned} \tag{4.18}$$

where in the second line we used $D_1 \leq 1$.

• For $s \in [D_1, D_2)$ with $D_1 \leq 1 \leq D_2 < T$, we start from the control (4.11) that we write

$$|\kappa(s)| \leq \int_{\mathbb{R}^2} |\varphi_{h_1,h_2}(x-x_0, y-y_0)| P_s(|\varphi_{h_1,h_2}(\cdot - x_0, \cdot - y_0)|)(x,y)\pi(x,y)\mathrm{d}x\mathrm{d}y + C.$$

Since $\varphi_{h_1,h_2}(\cdot - x_0, \cdot - y_0)$ vanishes outside the compact neighbourhood K of (x_0, y_0) we can use Lemma 2 to upper bound the semi group term. Hence, we get for some constant $C > 0$,

$$\begin{aligned}|\kappa(s)| \leq C \int_{\mathbb{R}^2} |\varphi_{h_1,h_2}(x-x_0, y-y_0)| \times \\ [\frac{\|\varphi_{h_1,h_2}\|_{L^1(\mathbb{R}^2)}}{D_1^2} + \frac{1}{h_1h_2}e^{-1/(CD_1)}]\pi(x,y)\mathrm{d}x\mathrm{d}y + C,\end{aligned}$$

and we deduce $|\kappa(s)| \leq C[\frac{1}{D_1^2} + \frac{1}{h_1h_2}e^{-1/(CD_1)} + 1]$. This yields,

$$\int_{D_1}^{D_2} |\kappa(s)|\mathrm{d}s \leq C[\frac{D_2}{D_1^2} + \frac{D_2}{h_1h_2}e^{-1/(CD_1)} + D_2], \tag{4.19}$$

for some constant $C > 0$.

• For $s \in [D_2, T]$, we use the covariance control (2.4), that allows us to write $|\kappa(s)| \leq C\|\varphi_{h_1,h_2}(\cdot - x_0, \cdot - y_0)\|_\infty^2 e^{-\rho s} \leq C\left(\frac{1}{h_1h_2}\right)^2 e^{-\rho s}$, for $C > 0$ and $\rho > 0$. It entails the upper bound,

$$\int_{D_2}^{T} |\kappa(s)|\mathrm{d}s \leq C\frac{e^{-\rho D_2}}{(h_1h_2)^2}. \tag{4.20}$$

Collecting together (4.8), (4.10), (4.18), (4.19), (4.20) we deduce,

$$\begin{aligned}\mathrm{Var}(\hat{\pi}_T(x_0,y_0)) \leq \frac{C}{T}[\frac{\delta}{h_1h_2} + \exp(-\frac{1}{CD_1})[1+\frac{D_2}{h_1h_2}] + \\ D_1 + D_2 + \frac{D_2}{D_1^2} + \frac{e^{-\rho D_2}}{(h_1h_2)^2}],\end{aligned}$$

for $C > 0$ some constant. We choose $\delta = \frac{6h_1}{y_0}$, $D_1 = \frac{1}{C|\log h_1h_2|}$, $D_2 = \frac{|\ln((h_1h_2)^2)|}{\rho}$. By (3.4)–(3.5) we see that this choice is such that, $\frac{6h_1}{y_0} = \delta < D_1 < 1 < D_2 < T$ for T large enough. And it yields,

$$\mathrm{Var}(\hat{\pi}_T(x_0,y_0)) \leq \frac{C}{T}\big[\frac{6}{y_0}\frac{1}{h_2} + h_1h_2 + D_1 + D_2 + \frac{D_2}{D_1^2} + 1\big].$$

Since $h_1 \to 0$, $h_2 \to 0$, and as the value of C may change from line to line, we can write that

$$\mathrm{Var}(\hat{\pi}_T(x_0,y_0)) \le \frac{C}{T}\left[\frac{1}{h_2} + |\ln(h_1h_2)|^C\right]$$

and we have shown (4.5).

Second step: we prove (4.6).

We use the same decomposition of $\int_0^T |\kappa(s)|\mathrm{d}s$ in four terms as for the proof of (4.5), but we treat in a different way the contribution of the short time correlations $\int_0^\delta |\kappa(s)|\mathrm{d}s$.

• Let us find an upper bound for $\kappa(s)$ for $s \in (0,\delta)$ with $\delta < 1$. We recall that, from Lemma 1, the decomposition (4.12) holds true where $\overline{\kappa}^1(s)$ is given by (4.13) and $\overline{\kappa}^2(s)$ is upper bounded by (4.17). We now study $\overline{\kappa}^1(s)$. To this end, we remark that $p_s^G(x,y;x',y') \le \frac{C}{\sqrt{s}} q_s(x' \mid x,y,y')$ where

$$q_s(x' \mid x,y,y') = \frac{C}{s^{3/2}} \exp\left(-C^{-1}\frac{(x'-x-\frac{y+y'}{2}s)^2}{s^3}\right).$$

Let us stress that

$$\sup_{s\in(0,1)} \sup_{(x,y,y')\in\mathbb{R}^3} \int_{\mathbb{R}} q_s(x' \mid x,y,y')\mathrm{d}x' \le C < \infty. \tag{4.21}$$

Thus, using (4.13), we have

$$\overline{\kappa}^1(s) \le \frac{C}{\sqrt{s}} \int_{\mathbb{R}^2} |\varphi_{h_1,h_2}(x-x_0,y-y_0)|\pi(x,y) \Big(\int_{\mathbb{R}^2} |\varphi_{h_1,h_2}(x'-x_0,y'-y_0)|q_s(x' \mid x,y,y')\mathrm{d}x'\mathrm{d}y'\Big)\mathrm{d}x\mathrm{d}y.$$

By (3.3), we have $|\varphi_{h_1,h_2}(x'-x_0,y'-y_0)| \le \frac{C}{h_1}\frac{1}{h_2}\left|\varphi(\frac{y'-y_0}{h_2})\right|$, and thus, using (4.21), we get

$$\begin{aligned}\int_{\mathbb{R}^2} |\varphi_{h_1,h_2}(x'-x_0,y'-y_0)|q_s(x' \mid x,y,y')\mathrm{d}x'\mathrm{d}y' &\le \frac{C}{h_1}\int_{\mathbb{R}} \frac{1}{h_2}\left|\varphi(\frac{y'-y_0}{h_2})\right| \int_{\mathbb{R}} q_s(x' \mid x,y,y')\mathrm{d}x'\mathrm{d}y' \\ &\le \frac{C}{h_1}\int_{\mathbb{R}} \frac{1}{h_2}\left|\varphi(\frac{y'-y_0}{h_2})\right|\mathrm{d}y' \le \frac{C}{h_1}.\end{aligned}$$

We deduce that $\overline{\kappa}^1(s) \le \frac{C}{\sqrt{s}h_1}\int_{\mathbb{R}^2} |\varphi_{h_1,h_2}(x-x_0,y-y_0)|\pi(x,y)\mathrm{d}x\mathrm{d}y \le \frac{C}{\sqrt{s}h_1}$. Collecting the latter equation with (4.12) and (4.17), it yields

$$\begin{aligned}\int_0^\delta |\kappa(s)|\mathrm{d}s &\le \int_0^\delta [\frac{C}{\sqrt{s}h_1} + C\frac{e^{-\frac{1}{Cs}}}{h_1h_2} + C]\mathrm{d}s \\ &\le \int_0^\delta [\frac{C}{\sqrt{s}h_1} + C\frac{1}{s^2}\frac{e^{-\frac{1}{Cs}}}{h_1h_2} + C]\mathrm{d}s \\ &\le C[\frac{\sqrt{\delta}}{h_1} + \frac{e^{-\frac{1}{C\delta}}}{h_1h_2} + \delta] \qquad (4.22)\end{aligned}$$

where we have used in the second line that $\delta < 1$.

We now gather (4.8), (4.18), (4.19), (4.20), (4.22), to derive,

$$\mathrm{Var}(\hat{\pi}_T(x_0,y_0)) \leq \frac{C}{T}\Big[\frac{\sqrt{\delta}}{h_1}+\frac{e^{-\frac{1}{C\delta}}}{h_1h_2}+\delta+\exp(-\frac{1}{CD_1})[1+\frac{D_2}{h_1h_2}]+ \\ D_1+D_2+\frac{D_1}{D_2^2}+\frac{e^{-\rho D_2}}{(h_1h_2)^2}\Big].$$

We choose the same thresholds as in the first step, $\delta=\frac{6h_1}{y_0}$, $D_1=\frac{1}{C|\log h_1h_2|}$, $D_2=\frac{|\ln((h_1h_2)^2)|}{\rho}$. Recalling (3.4)–(3.5), we have $e^{-(C6h_1/y_0)^{-1}}=O(e^{-\varepsilon\log(T)^3})=o(h_1h_2)$, with some $\varepsilon>0$. We derive that

$$\mathrm{Var}(\hat{\pi}_T(x_0,y_0)) \leq \frac{C}{T}[\frac{1}{\sqrt{h_1}}+|\ln(h_1h_2)|^C],$$

for some $C>0$. The second step of the proposition is proved. □

Remark 3. (1) *The Proposition 2 consists actually in the two upper bounds (4.5)–(4.6) for the variance of the estimator. We see that one of these two bounds is smaller than the other, depending on the relative positions of h_2 or $\sqrt{h_1}$. It explains why the expression for the rate of convergence of the estimator in Theorem 1 depends on the relative positions of k_1 and $k_2/2$, which determines which one of the two bounds (4.5) or (4.6) is used in the bias/variance decomposition of the estimation error (see proof of Theorem 1).*

(2) *The control of $\kappa(s)=\mathrm{Cov}\,(\varphi_{h_1,h_2}(X_0-x_0,Y_0-y_0),\varphi_{h_1,h_2}(X_s-x_0,Y_s-y_0))$ for $s\in[\delta,D_1]$, with $\delta\approx h_1$ and $D_1\leq 1$ depends on the fine structure of the main term (4.2) in the short time expansion of the transition density of the process and on the fact that $y_0\neq 0$. In the situation $y_0\neq 0$, it is impossible to get such a refined result on the covariance, and eventually the bound on the variance of the estimator is larger (see Proposition 3).*

4.2. **Proof of Proposition 3.** We need to prove that the following two inequalities hold true, for T large enough:

$$\mathrm{Var}(\hat{\pi}_T(x_0,y_0)) \leq C\frac{1}{Th_1^{2/3}}+C\varepsilon(T,h_1(T),h_2(T)), \tag{4.23}$$

$$\mathrm{Var}(\hat{\pi}_T(x_0,y_0)) \leq C\frac{\ln(T)}{Th_2^2}+C\varepsilon(T,h_1(T),h_2(T)). \tag{4.24}$$

Again we consider K a compact set of $\mathbb{R}^2$ that contains a ball of radius $\sqrt{2}$ centered at (x_0,y_0).

First step : let us prove (4.23).

We recall the control (4.8) for the variance of $\hat{\pi}_T(x_0,y_0)$ and split the integral in (4.8) into four pieces corresponding to the partition $[0,T]=[0,\delta)\cup[\delta,D_1)\cup[D_1,D_2)\cup[D_2,T]$, where δ, D_1, D_2 will be specified latter. Let us stress that in the proof of Proposition 2, only the control of $|\kappa(s)|$ for $s\in[\delta,D_1)$, uses the fact that $y_0\neq 0$.

- For $s\in[0,\delta)$ with $\delta<1$, we recall the result obtained in (4.22) which states

$$\int_0^\delta |\kappa(s)|\mathrm{d}s \leq C[\frac{\sqrt{\delta}}{h_1}+\frac{e^{-\frac{1}{C\delta}}}{h_1h_2}+\delta].$$

• For $s \in [\delta, D_1]$ with $0 < \delta < D_1 < 1$, exactly with the same proof as in Proposition 2, we have $|\kappa(s)| \leq \overline{\kappa}^1(s) + \overline{\kappa}^2(s) + C$ where $\overline{\kappa}^1(s)$ is given by (4.13) and $\overline{\kappa}^2(s)$ is upper-bounded as in (4.17). We need to find a control on $\overline{\kappa}^1(s)$ in the situation $y_0 \neq 0$. Using, from (4.2), that $p_s^G(x, y; x', y) \leq \frac{C}{s^2}$, $\forall (x, y, x', y') \in K^2$, and the fact that π is bounded, we get

$$\overline{\kappa}^1(s) \leq \frac{C}{s^2} \int_{\mathbb{R}^4} |\varphi_{h_1,h_2}(x - x_0, y - y_0)| \\ |\varphi_{h_1,h_2}(x' - x_0, y' - y_0)| \mathrm{d}x \mathrm{d}y \mathrm{d}x' \mathrm{d}y' \leq \frac{C}{s^2}.$$

We deduce

$$\begin{aligned} \int_\delta^{D_1} |\kappa(s)| \mathrm{d}s &\leq \int_\delta^{D_1} C[\frac{1}{s^2} + \frac{e^{-\frac{1}{Cs}}}{h_1 h_2} + C] \mathrm{d}s \\ &\leq C[\frac{1}{\delta} + \exp(-\frac{1}{CD_1}) \frac{1}{h_1 h_2} + D_1] \\ &\leq C[\frac{1}{\delta} + \exp(-\frac{1}{CD_1}) \frac{1}{h_1 h_2}], \end{aligned} \tag{4.25}$$

where we used $D_1 \leq 1$.

• For $s \in [D_1, D_2)$, with $D_1 < 1 < D_2 < T$, we use the control (4.19).

• For $s \in [D_2, T]$, with $D_2 < T$, we use the control (4.20).

Collecting the four previous controls, we deduce

$$\mathrm{Var}(\hat{\pi}_T(x_0, y_0)) \leq \frac{C}{T}[\frac{\sqrt{\delta}}{h_1} + \frac{e^{-\frac{1}{C\delta}}}{h_1 h_2} + \frac{1}{\delta} + \frac{\exp(-\frac{1}{CD_1})}{h_1 h_2}(1 + D_2) + \\ D_1 + D_2 + \frac{D_2}{D_1^2} + \frac{e^{-\rho D_2}}{(h_1 h_2)^2}]$$

where $C > 0$, $\rho > 0$. We choose δ that balances $\sqrt{\delta}/h_1$ with $1/\delta$, namely $\delta = h_1^{2/3}$ which is smaller than 1 for T large enough, recalling (3.5). Next, we choose $D_2 = \frac{|\ln((h_1 h_2)^2)|}{\rho}$, and $D_1 = \frac{C}{|\log h_1 h_2|}$. We deduce

$$\mathrm{Var}(\hat{\pi}_T(x_0, y_0)) \leq \frac{C}{T}[\frac{1}{h_1^{2/3}} + |\log(h_1 h_2)|^C],$$

for some $C > 0$, and where we have used that, from (3.4)–(3.5), $\exp(-1/(Ch_1^{2/3})) = O(\exp(-\varepsilon \log(T)^2)) = o(h_1 h_2)$, with some $\varepsilon > 0$. Hence (4.23) is proved.

Second step: we show (4.24).

Comparing with the first part of the proposition, we have to modify our upper bound on $\int_0^\delta |\kappa(s)| \mathrm{d}s$ with $\delta < 1$. We split this integral into two parts, $\int_0^{\delta'} |\kappa(s)| \mathrm{d}s + \int_{\delta'}^\delta |\kappa(s)| \mathrm{d}s$. On the first part, we use the control (4.9) and get

$$\int_0^{\delta'} |\kappa(s)| \mathrm{d}s \leq C \frac{\delta'}{h_1 h_2}. \tag{4.26}$$

On the second part, we use (4.12), where $\overline{\kappa}^2(s)$ is bounded in (4.17). We deduce,

$$\begin{aligned}\int_{\delta'}^{\delta}|\kappa(s)|\mathrm{d}s &\leq \int_{\delta'}^{\delta} C[\frac{1}{h_1h_2}e^{-\frac{1}{Cs}}+1]\mathrm{d}s+\int_{\delta'}^{\delta}|\overline{\kappa}^1(s)|\mathrm{d}s\\ &\leq C[\frac{1}{h_1h_2}e^{-\frac{1}{C\delta}}+\delta]+\int_{\delta'}^{\delta}\overline{\kappa}^1(s)\mathrm{d}s. \end{aligned} \tag{4.27}$$

To upper bound $\overline{\kappa}^1(s)$, we use (4.13) and (3.3), and obtain by Fubini's Theorem,

$$\begin{aligned}\overline{\kappa}^1(s)=&\int_{\mathbb{R}^2}\left|\frac{1}{h_1}\varphi(\frac{x-x_0}{h_1})\frac{1}{h_1}\varphi(\frac{x'-x_0}{h_1})\right|\\ &\Big(\int_{\mathbb{R}^2}\left|\frac{1}{h_2}\varphi(\frac{y-y_0}{h_2})\frac{1}{h_2}\varphi(\frac{y'-y_0}{h_2})\right|p_s^G(x,y;x',y')\pi(x,y)\mathrm{d}y\mathrm{d}y'\Big)\mathrm{d}x\mathrm{d}x'.\end{aligned} \tag{4.28}$$

Since π is bounded, and using (4.2), we deduce that the inner integral is lower than

$$\begin{aligned}&\frac{\|\varphi\|_\infty^2}{h_2^2}\int_{\mathbb{R}^2}\frac{C}{s^2}\exp\left(-\frac{1}{C}\left[\frac{(y-y')^2}{s}+\frac{(x'-x-\frac{y+y'}{2}s)^2}{s^3}\right]\right)\mathrm{d}y\mathrm{d}y'\\ =&\frac{\|\varphi\|_\infty^2}{2h_2^2}\int_{\mathbb{R}^2}\frac{C}{s}\exp\left(-\frac{1}{C}\left[w^2+(\frac{x'-x}{s^{3/2}}-\frac{w'}{2})^2\right]\right)\mathrm{d}w\mathrm{d}w'\\ =&\frac{\|\varphi\|_\infty^2}{2h_2^2}\int_{\mathbb{R}^2}\frac{C}{s}\exp\left(-\frac{1}{C}\left[w^2+(\frac{w'}{2})^2\right]\right)\mathrm{d}w\mathrm{d}w',\end{aligned}$$

where we have made the change of variables $w=\frac{y-y'}{\sqrt{s}}$, $w'=\frac{y+y'}{\sqrt{s}}$ in the second line, and used the invariance by translation of the Lebesgue measure in the last one. We deduce that the inner integral in (4.28) is lower than $\frac{C}{h_2^2 s}$ where we stress that C does not depend on (x,x'). In turn,

$$\overline{\kappa}^1(s)\leq\frac{C}{h_2^2 s}\int_{\mathbb{R}^2}\left|\frac{1}{h_1}\varphi(\frac{x-x_0}{h_1})\frac{1}{h_1}\varphi(\frac{x'-x_0}{h_1})\right|\mathrm{d}x\mathrm{d}x'\leq\frac{C}{h_2^2 s}.$$

This yields, using (4.27) to

$$\int_{\delta'}^{\delta}|\kappa(s)|\mathrm{d}s\leq C[\frac{1}{h_1h_2}e^{-\frac{1}{C\delta}}+\delta+\frac{1}{h_2^2}\ln(\frac{\delta}{\delta'})]. \tag{4.29}$$

Collecting (4.26), (4.29), (4.25), (4.19), (4.20), we deduce, for $0<\delta'<\delta<D_1<1<D_2<T$,

$$\begin{aligned}\mathrm{Var}(\hat{\pi}_T(x_0,y_0))\leq\frac{C}{T}\Big[&\frac{\delta'}{h_1h_2}+\frac{e^{-\frac{1}{C\delta}}}{h_1h_2}+\frac{1}{h_2^2}\ln(\frac{\delta}{\delta'})+\frac{1}{\delta}+\frac{\exp(-\frac{1}{CD_1})}{h_1h_2}(1+D_2)+\\ &D_1+D_2+\frac{D_1}{D_2^2}+\frac{e^{-\rho D_2}}{(h_1h_2)^2}\Big].\end{aligned}$$

Now we let $\delta=h_2^2$ and $\delta'=T^{-C'_\delta}$ where $C'_\delta>0$ is such that $\delta'=o(h_1h_2)$ which is possible from the at most polynomial decay of the bandwidths, resorting to (3.4). As in the first step of the proposition, we set $D_1=\frac{C}{|\ln(h_1h_2)|}$, $D_2=\frac{|\ln((h_1h_2)^2)|}{\rho}$. With these choices, we have for T large enough, $0<\delta'<\delta<D_1<1<D_2<T$

and

$$\begin{aligned}\mathrm{Var}(\hat{\pi}_T(x_0,y_0)) &\leq \frac{C}{T}[1+\frac{e^{-\frac{1}{Ch_2^2}}}{h_1h_2}+\frac{1}{h_2^2}\ln(\delta/T^{-C_{\delta'}})+|\ln(h_1h_2)|^C]\\ &\leq \frac{C}{T}[\frac{\ln(T)}{h_2^2}+|\ln(h_1h_2)|^C]\end{aligned}$$

where we used again (3.4)–(3.5) in the last line. This proves (4.24). □

4.3. **Numerical simulations.** In this section, we explore numerically on an example the behaviour of the variance of the estimator as h_1 and h_2 go to 0. Especially, we wonder if the variance of the estimator asymptotically depends, in these simulations, only on the minimum of two quantities related to h_1 and h_2 as suggested by the upper bounds in Propositions 2–3. This is the crucial point in the upper bound of the variance, that makes the choice of the optimal bandwidth very specific, allowing an arbitrary thin bandwidth on one component. We consider the model (2.1)–(2.2) with $\beta(x,y)=0.5$, $V(x)=x^2/2$ and $a(x,y)=1$. From a Monté–Carlo experiment based on 500 replications, we evaluate the variance of $\hat{\pi}(x_0,y_0)$ for $(x_0,y_0)=(0,1.5)$ and with different values of bandwidths h_1 and h_2. We have chosen the simple kernel $\varphi(u)=\frac{1}{2}1_{[-1,1]}(u)$ and $T=200$. Results are given in Figure 1, where each curve corresponds to a choice for the bandwidth h_2, and these curves plot the value of the variance as a function of h_1, using log-scales. We see that, as expected, the variance is increasing as h_1 and h_2 get smaller. Moreover, it appears that when h_2 gets smaller than some threshold depending on h_1 the variance ceases to strictly increase, as all the curves are flat on the left side of the Figure 1. Symmetrically, we see that the right part of the curves for $h_2=10^{-1.8}$ and $h_2=10^{-2.4}$ coincides. It shows for instance that decreasing h_2 below $10^{-1.8}$ does not increase anymore the variance when $h_1\geq 10^{-2}$. Hence, the numerical results are consistent with the upper bound given for the variance, as a function depending on $\min(1/h_2,1/\sqrt{h_1})$. This suggests that the upper bounds given in Propositions 2–3 are fairly sharp.

5. Minimax lower bound

In this section, we show that it is impossible to construct any estimator with a uniform rate better (up to a log term) than the rates obtained in Theorems 1–2.

5.1. **Lower bounds.** For the computation of lower bounds, we introduce the family of S.D.E.

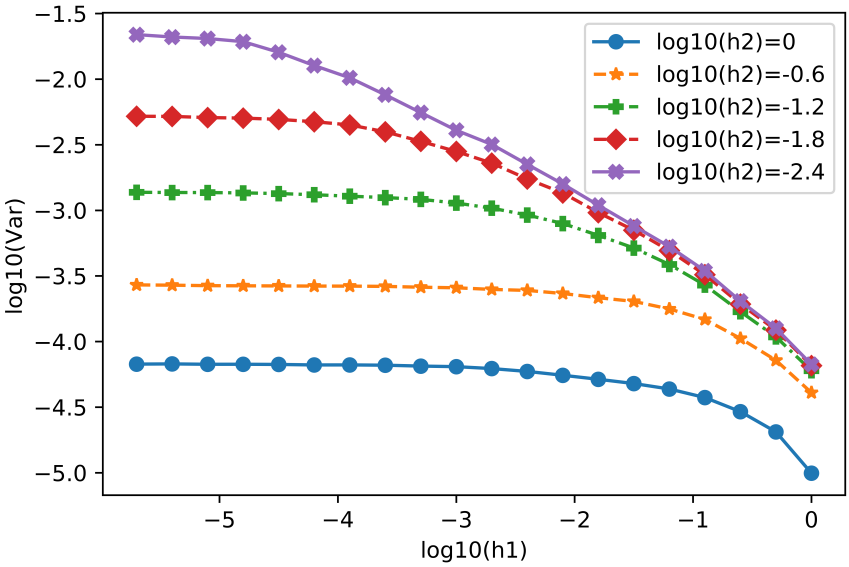
$$\mathrm{d}X_t = Y_t\mathrm{d}t \tag{5.1}$$

$$\mathrm{d}Y_t = 2\sigma\mathrm{d}B_t-[\sigma^2\beta(X_t,Y_t)Y_t+V'(X_t)]\mathrm{d}t, \tag{5.2}$$

where $\sigma>0$, β is a bounded $\mathcal{C}^1$ function lower bounded by a strictly positive number and V is $\mathcal{C}^2$. As the model satisfies the conditions of **HReg** we know that the S.D.E. admits a weak solution. We know that if V satisfies **HErg** then a Lyapounov function exists and the process admits a unique stationary measure, that we note $\pi_{V,\beta}$. In Section 5.2, we make more explicit the connection between $\pi_{V,\beta}$ and the coefficients V, β. Remark that we omit in the notations the dependence on σ, as σ will be fixed in the sequel.

If the stationary measure exists and is unique, we denote $\mathbb{P}_{V,\beta}$ the law of a stationary solution $(X_t,Y_t)_{t\geq 0}$ of (5.1)–(5.2). Here, $\mathbb{P}_{V,\beta}$ is a measure on the space

FIGURE 1. Variance of the estimator



of continuous function $\mathcal{C}([0,\infty),\mathbb{R}^2)$, and we note by $\mathbb{E}_{V,\beta}$ the corresponding expectation. When needed, we will note by $\mathbb{P}^{(T)}_{V,\beta}$ the law of the stationary process $(X_t,Y_t)_{t\in[0,T]}$ solution to the S.D.E. (5.1)–(5.2).

In the sequel, we note again (X,Y) the canonical process on $\mathcal{C}([0,\infty),\mathbb{R}^2)$ or $\mathcal{C}([0,T],\mathbb{R}^2)$.

In order to write down an expression for the minimax risk of estimation, we have to consider a set of solutions to the S.D.E. (5.1)–(5.2), which are stationary and whose stationary measure has a prescribed Hölder regularity. This leads us to the following definition.

Definition 2. *Let $V:\mathbb{R}\to\mathbb{R}$ be a $\mathcal{C}^2$ function satisfying* **HErg**. *We consider $k_1>0$, $k_2>0$, $R>0$, and $R'>1$ real numbers. We define $\Sigma^{k_1,k_2}(V,R,R')$ the set of continuously differentiable functions $\beta:\mathbb{R}^2\to\mathbb{R}$ satisfying the following two conditions :*

1. *$1/R'\le\beta(x,y)\le R'$ for all $(x,y)\in\mathbb{R}^2$,*
2. *the density $\pi_{V,\beta}$ of the stationary measure associated to the S.D.E. (5.1)–(5.2) is such that $\pi_{V,\beta}\in\mathcal{H}^{k_1,k_2}(R)$.*

We introduce the minimax risk for the estimation at some point. Let $(x_0,y_0)\in\mathbb{R}^2$, and V, k_1, k_2, R, R' as in Definition 2. We let

(5.3)
$$R_T(V,k_1,k_2,R,R')=\inf_{\widetilde{\pi}_T}\sup_{\beta\in\Sigma^{k_1,k_2}(V,R,R')}\mathbb{E}_{V,\beta}\left[(\widetilde{\pi}_T(x_0,y_0)-\pi_{V,\beta}(x_0,y_0))^2\right],$$

where the infimum is taken on all possible estimators of $\pi_T(x_0,y_0)$, that is for $\widetilde{\pi}_T=\widetilde{\pi}_T(x_0,y_0)$ ranging in the set of all the measurable functions of $(X_t,Y_t)_{t\in[0,T]}$ with values in $\mathbb{R}$.

Theorem 3. *Let $k_1, k_2, R > 0$ and assume $y_0 \neq 0$ and $\max(k_1, k_2/2) > 1/2$. Then, there exists V satisfying* **HErg** *and $R' > 1$ such that, for some constant $C > 0$, we have :*

$$R_T(V, k_1, k_2, R, R') \geq C T^{-2v(k_1,k_2)}, \quad \forall T > 1, \tag{5.4}$$

with

$$v(k_1, k_2) = \begin{cases} \frac{k_2}{2k_2+1}, & \text{if } k_1 < k_2/2, \\ \frac{k_1}{2k_1+1/2}, & \text{if } k_1 \geq k_2/2. \end{cases}$$

Remark 4.

(1) *Theorem 3 tells us that it is impossible to find an estimator with a rate of estimation, for the pointwise L^2 risk, better than $T^{-v(k_1,k_2)}$ on a the class of diffusions $Z = (X, Y)$ having a with $\mathcal{H}^{k_1,k_2}(R)$ stationary measure. On the other hand the estimator introduced in Section 3 achieves this rate, by Theorem 1, for each diffusion $Z = (X, Y)$ satisfying* **HReg** *and* **HErg** *and with stationary measure in $\mathcal{H}^{k_1,k_2}(R)$.*

(2) *The upper bound given in Theorem 1 is not stated uniformly on the class of all diffusions satisfying* **HReg** *and* **HErg** *and is not a minimax upper bound. To get uniform upper bound, we would need that the mixing control (2.4) holds uniformly on a class of diffusions whose coefficients satisfy uniform versions of the assumptions* **HReg**, **HErg**. *We are not aware of such uniform mixing results, and hence, getting a uniform version of Theorem 1 is left for further research.*

(3) *The condition $\max(k_1, k_2/2) > 1/2$ asserts that π is not too irregular with respect to both variables x and y. Such assumption is weak, as the stationary measure is typically smoother than the coefficients of the S.D.E. (see point 3 of Remark 5 below).*

Theorem 4. *Let $k_1, k_2, R > 0$ and assume $y_0 = 0$ and $\max(k_1, k_2/3) > 2/3$. Then, there exists V satisfying* **HErg** *and $R' > 1$ such that, for some constant $C > 0$, we have :*

$$R_T(V, k_1, k_2, R, R') \geq C T^{-2v'(k_1,k_2)}, \quad \forall T > 1, \tag{5.5}$$

with

$$v'(k_1, k_2) = \begin{cases} \frac{k_2}{2k_2+2}, & \text{if } k_1 < k_2/3, \\ \frac{k_1}{2k_1+2/3}, & \text{if } k_1 \geq k_2/3. \end{cases}$$

Again, the previous result shows that the estimator introduced in Section 3 is rate efficient, up to a log term, in the case where $y_0 = 0$.

5.2. **Explicit link between the drift and the stationary measure.** Recall that from Proposition 1, **HReg** and **HErg** are sufficient for the existence and uniqueness of a stationary probability of the process solution of (5.1)–(5.2) (see [17] also, or see Talay [15] for related conditions too). In this section, we will characterize explicit relations between (V, β) and $\pi_{V,\beta}$.

We need to introduce $A^{\star}_{V,\beta}$ the adjoint on $\mathbf{L}^2(\mathbb{R}^2, \mathrm{d}x\mathrm{d}y)$ of the generator $A_{V,\beta}$ of the process $Z = (X, Y)$ solution to (5.1)–(5.2).

Assume that V is $\mathcal{C}^2$ and that $(x,y) \mapsto y\beta(x,y)$ is of class $\mathcal{C}^1$. Then we define for $g : \mathbb{R}^2 \to \mathbb{R}$ any $\mathcal{C}^2$ function,

$$A^\star_{V,\beta} g(x,y) = 2\sigma^2 \frac{\partial^2 g}{\partial y^2}(x,y) - y\frac{\partial g}{\partial x}(x,y) + [\sigma^2 y\beta(x,y) + V'(x)]\frac{\partial g}{\partial y}(x,y) + \sigma^2 \frac{\partial (y\beta)}{\partial y}(x,y) g(x,y). \tag{5.6}$$

It can be checked that (5.6) is the expression for the adjoint of the generator of the process. If $g : \mathbb{R}^2 \to \mathbb{R}$ is a probability density, of class $\mathcal{C}^2$, solution to $A^\star_{V,\beta} g = 0$, then it is an invariant density for the process. Hence, when the stationary distribution $\pi_{V,\beta}$ is unique, it can be computed as solution of the equation $A^\star_{V,\beta}\pi_{V,\beta} = 0$. From the expression (5.6) it seems impossible to find explicit solutions g to the equation $A^\star_{V,\beta} g = 0$ for any V and β, as one need to solve explicitly some P.D.E. Consequently, it seems impossible to write $\pi_{V,\beta}$ as an explicit expression of (V,β).

On the other hand, using (5.6) it can be seen that if one consider g and V as fixed and β as the unknown variable in the equation $A^\star_{V,\beta} g = 0$, then finding solution in β is simpler as one has to deal with a P.D.E. involving only differentiation with respect to y. As a consequence, it will be possible to express β as a function of the stationary distribution π (for a fixed V). This is the object of the next proposition. We need to introduce some notations first.

For $g \in \mathcal{C}^2$ and $g > 0$, we define for all $(x,y) \in \mathbb{R}^2$,

$$\xi_g(x,y) = \frac{1}{\sigma^2 g(x,y)} \int_0^y \big[z\frac{\partial g}{\partial x}(x,z) - V'(x)\frac{\partial g}{\partial y}(x,z) + 2\sigma^2 \frac{\partial^2 g}{\partial y^2}(x,z)\big]\mathrm{d}z, \tag{5.7}$$

and

$$\beta_g(x,y) = \begin{cases} \dfrac{1}{y}\xi_g(x,y), \text{ for } y \neq 0, \\ \lim\limits_{y\to 0} \dfrac{1}{y}\xi_g(x,y) = \\ \quad \dfrac{1}{\sigma^2 g(x,0)}[-V'(x)\dfrac{\partial g}{\partial y}(x,0) + 2\sigma^2 \dfrac{\partial^2 g}{\partial y^2}(x,0)], \text{ for } y = 0. \end{cases} \tag{5.8}$$

Proposition 4. 1) *Let $V : \mathbb{R} \to \mathbb{R}$ with regularity $\mathcal{C}^2$, $g : \mathbb{R}^2 \to \mathbb{R}$ with regularity $\mathcal{C}^2$ and $g > 0$.*

Then, we have that $(x,y) \mapsto y\beta_g(x,y)$ is a $\mathcal{C}^1$ function and

$$A^\star_{V,\beta_g} g(x,y) = 0, \quad \forall (x,y) \in \mathbb{R}^2. \tag{5.9}$$

Moreover, β_g is the unique function solution to (5.9) such that $(x,y) \mapsto y\beta_g(x,y)$ is $\mathcal{C}^{0,1}$.

2) *Let $V : \mathbb{R} \to \mathbb{R}$ with regularity $\mathcal{C}^2$ and satisfying* **HErg** *and consider $\pi : \mathbb{R}^2 \to \mathbb{R}$ a probability density with regularity $\mathcal{C}^2$ and $\pi > 0$.*

Assume that $1/R' < \beta_\pi < R'$ for some $R' > 1$, where β_π is defined by (5.8).

Then, π is the unique stationary probability of the S.D.E (5.1)–(5.2) with damping coefficient $\beta = \beta_\pi$ and potential V

Proof 1) For $g : \mathbb{R}^2 \to \mathbb{R}$ with regularity $\mathcal{C}^2$ and $g > 0$ and β such that $(x,y) \mapsto$

$y\beta(x,y)$ is of class $\mathcal{C}^1$, we can write the equation $A^\star_{V,\beta}g(x,y)=0$, recalling (5.6), as

$$\frac{\partial(y\beta)}{\partial y}(x,y)+\frac{\frac{\partial g}{\partial y}(x,y)}{g(x,y)}(y\beta(x,y))=i_g(x,y) \tag{5.10}$$

with

$$i_g(x,y)=\frac{1}{\sigma^2 g(x,y)}\left[y\frac{\partial g}{\partial x}(x,y)-V'(x)\frac{\partial g}{\partial y}(x,y)-2\sigma^2\frac{\partial^2 g}{\partial y^2}(x,y)\right].$$

Let us fix $x\in\mathbb{R}$, we then interpret (5.10) as an ordinary differential equation with differentiation variable y, where the unknown parameter is the function $y\mapsto \xi(x,y):=y\beta(x,y)$:

$$\frac{\partial \xi}{\partial y}+\frac{\frac{\partial g}{\partial y}}{g}\xi=i_g. \tag{5.11}$$

A solution of the homogeneous equation $\frac{\partial \xi}{\partial y}+\frac{\frac{\partial g}{\partial y}}{g}\xi=0$ is $\xi(x,y)=1/g(x,y)$. Then, by variation of the constant method, we deduce that the solution to (5.11), has the expression,

$$\xi(x,y)=\frac{1}{g(x,y)}\left[c(x)+\int_0^y g(x,z)i_g(x,z)\mathrm{d}z\right],$$

where $c(x)$ is an integration constant. As $\xi(x,0)=0\times\beta(x,0)=0$, we deduce that $c(x)=0$, $\forall x$. Hence the solution $\xi(x,y)=y\beta(x,y)$ of (5.11) is given by (5.7) and in turn, we deduce that $\beta=\beta_g$ given by (5.8) is the unique solution to (5.10) or equivalently to (5.9).

2) Using Ito's formula, one can check that any π solution to $A^\star_{V,\beta}\pi=0$ is a stationary measure for the process (X,Y) given by (5.1)–(5.2). From the first part of the proposition, π is solution to $A^\star_{V,\beta_\pi}\pi=0$. By Proposition 1, the stationary measure of the equation with damping coefficient β_π is unique, and is thus equal to π. □

Remark 5. (1) *Is is known that the S.D.E.*

$$\begin{aligned}dX_t&=Y_t\,dt\\ dY_t&=2\sigma\,dB_t-[\sigma^2\varepsilon Y_t+V'(X_t)]\,dt.\end{aligned}$$

admits for the stationary measure $\pi(x,y)=C\exp\left(-\frac{\varepsilon}{2}[\frac{y^2}{2}+V(x)]\right)$ *(see e.g. [5]). As expected, if we take* $\pi(x,y)=C\exp\left(-\frac{\varepsilon}{2}[\frac{y^2}{2}+V(x)]\right)$ *and compute* β_π *by the formula (5.8), we find* $\beta_\pi=\varepsilon$.

(2) *Proposition 4 shows how to compute the damping part of the drift in order to get a diffusion with a prescribed stationary measure. However, it is not clear that for a given* π *the corresponding* β_π*, computed with (5.7)–(5.8) satisfies the sign condition* $\beta_\pi>1/R'$ *insuring that the process is indeed ergodic. This is why in part 2) of Proposition 4 we postulate* $\beta_\pi>1/R'$*. However, we will see in Section 5.3 that if* π *is a small deviation of* π_0 *given by* $\pi_0(x,y)=C\exp\left(-\frac{\varepsilon}{2}[\frac{y^2}{2}+V(x)]\right)$*, then the corresponding* β *is a small deviation of* $\beta_0=\varepsilon$ *and thus is positive.*

(3) *The equation (5.7) enables to relate the degree of smoothness of the drift coefficient and of the stationary measure, when the latter exists and is unique. Indeed, from (5.7), we get that if $\pi \in \mathcal{C}^{k_1,k_2}$ and $V \in \mathcal{C}^{k_1}$, then the associated drift of the S.D.E. $\xi_\pi + V'$ is $\mathcal{C}^{k_1-1,k_2-1}$.*

5.3. **Proof of Theorem 3.** The proof of the lower bound is made by a comparison between the minmax risk (5.3) and some Bayesian risk where the Bayesian prior is supported on a set of two elements.

5.3.1. *Construction of the prior.* Let $k_1, k_2 > 0$ and $R > 0$. We set $V_0(x) = |x|^2$ and define

$$\pi_0(x,y) = c_\eta \exp(-\frac{\eta}{2}[\frac{y^2}{2} + x^2]), \quad \beta_0(x,y) = \eta, \quad \xi_0(x,y) = \eta y, \tag{5.12}$$

where $\eta > 0$ and where c_η is the constant that makes π_0 a probability measure. The function π_0 is $\mathcal{C}^\infty$ and it is possible to choose $0 < \eta < 1/2$ small enough such that

$$\pi_0 \in \mathcal{H}^{k_1,k_2}(R/2).$$

We know from Section 5.2 that π_0 is the unique stationary measure for $(X^{(0)}, Y^{(0)})$ solution of

$$\mathrm{d}X_t^{(0)} = Y_t^{(0)}\mathrm{d}t \tag{5.13}$$

$$\mathrm{d}Y_t^{(0)} = 2\sigma \mathrm{d}B_t - [\sigma^2 \eta Y_t^{(0)} + V_0'(X_t^{(0)})]\mathrm{d}t. \tag{5.14}$$

Now, if we set $R' = 2/\eta > 1$, we have, using $\beta_0 = \eta$ and recalling Definition 2,

$$\beta_0 \in \Sigma^{k_1,k_2}(V, R/2, R'/2). \tag{5.15}$$

Let $h : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^\infty$ function with support on $[-1,1]$ and such that

$$h(0) = 1, \ \int_{-1}^{1} h(z)\mathrm{d}z = 0, \ \int_{-1}^{1} z h(z)\mathrm{d}z = 0. \tag{5.16}$$

We set for $T > 0$,

$$\tilde{\pi}_T(x,y) = \pi_0(x,y) + \frac{1}{M_T} h(\frac{x - x_0}{h_1(T)}) h(\frac{y - y_0}{h_2(T)}), \tag{5.17}$$

where M_T, $h_1(T)$, $h_2(T)$ will be calibrated later and satisfy

$$M_T \xrightarrow{T\to\infty} \infty, \ h_1(T) \xrightarrow{T\to\infty} 0, \ h_2(T) \xrightarrow{T\to\infty} 0.$$

From (5.16), we see that $\int_{\mathbb{R}^2} \tilde{\pi}_T(x,y)\mathrm{d}x\mathrm{d}y = \int_{\mathbb{R}^2} \pi_0(x,y)\mathrm{d}x\mathrm{d}y = 1$, and using $\pi_0 > 0$, $1/M_T \to 0$ and that h is compactly supported, we see that $\tilde{\pi}_T > 0$ for T large enough. Hence $\tilde{\pi}_T$ is a smooth probability measure for T large enough. We define

$$\tilde{\beta}_T(x,y) = \beta_{\tilde{\pi}_T}(x,y), \quad \tilde{\xi}_T(x,y) = y\tilde{\beta}_T(x,y) = \xi_{\tilde{\pi}_T}(x,y),$$

where we used the definitions (5.7) and (5.8).

Before proving Theorem 3, we need to state two lemmas. The first lemma shows that the two functions β_0 and $\tilde{\beta}_T$ only differ on some vanishing neighbourhood of (x_0, y_0).

Lemma 3. *1) Let us define the compact set of* $\mathbb{R}^2$

$$K_T = [x_0 - h_1(T), x_0 + h_1(T)] \times [y_0 - h_2(T), y_0 + h_2(T)].$$

Then, for T *large enough, we have for all* $(x,y) \notin K_T$ *:*

$$\beta_0(x,y) = \tilde{\beta}_T(x,y), \quad \xi_0(x,y) = \tilde{\xi}_T(x,y).$$

2) For $(x,y) \in K_T$*, we have the control*

$$\left|\beta_0(x,y) - \tilde{\beta}_T(x,y)\right| \leq \frac{C}{M_T}\left[\frac{h_2(T)}{h_1(T)} + \frac{1}{h_2(T)}\right], \tag{5.18}$$

$$\left|\xi_0(x,y) - \tilde{\xi}_T(x,y)\right| \leq \frac{C}{M_T}\left[\frac{h_2(T)}{h_1(T)} + \frac{1}{h_2(T)}\right], \tag{5.19}$$

where C *is some constant independent of* T*,* $h_1(T)$*,* $h_2(T)$*,* M_T*.*

3) We have

$$\int_{\mathbb{R}^2} \left|\tilde{\xi}_T(x,y) - \xi_0(x,y)\right|^2 dx\,dy \leq \frac{C}{M_T^2}\left[\frac{h_2(T)^3}{h_1(T)} + \frac{h_1(T)}{h_2(T)}\right].$$

Proof 1) We first prove the $\tilde{\xi}_T$ and ξ_0 coincides on K_T^c. With the definition (5.7) in mind, we set for g of class $\mathcal{C}^{1,2}$:

$$\begin{aligned}\mathcal{I}[g](x,y) &= \frac{1}{\sigma^2}\int_0^y \left[z\frac{\partial g}{\partial x}(x,z) - V'(x)\frac{\partial g}{\partial y}(x,z) + 2\sigma^2\frac{\partial^2 g}{\partial y^2}(x,z)\right]\mathrm{d}z \\ &= \mathcal{I}^1[g](x,y) + \mathcal{I}^2[g](x,y) + \mathcal{I}^3[g](x,y),\end{aligned}$$

where

$$\mathcal{I}^1[g](x,y) = \frac{1}{\sigma^2}\int_0^y z\frac{\partial g}{\partial x}(x,z)\mathrm{d}z, \tag{5.20}$$

$$\mathcal{I}^2[g](x,y) = -\frac{V'(x)}{\sigma^2}[g(x,y) - g(x,0)], \tag{5.21}$$

$$\mathcal{I}^3[g](x,y) = 2[\frac{\partial g}{\partial y}(x,y) - \frac{\partial g}{\partial y}(x,0)]. \tag{5.22}$$

Using this notation, we have

$$\tilde{\xi}_T = \frac{1}{\tilde{\pi}_T}\mathcal{I}[\tilde{\pi}_T], \quad \xi_0 = \frac{1}{\pi_0}\mathcal{I}[\pi_0]. \tag{5.23}$$

Let us note

$$d_T = \tilde{\pi}_T - \pi_0 \tag{5.24}$$

and by (5.17), we have

$$d_T(x,y) = \frac{1}{M_T}h(\frac{x-x_0}{h_1(T)})h(\frac{y-y_0}{h_2(T)}). \tag{5.25}$$

Since $g \mapsto \mathcal{I}[g]$ is a linear operator we deduce that

$$\tilde{\xi}_T = \frac{1}{\tilde{\pi}_T}\mathcal{I}[\tilde{\pi}_T] = \frac{1}{\tilde{\pi}_T}\mathcal{I}[\pi_0] + \frac{1}{\tilde{\pi}_T}\mathcal{I}[d_T]. \tag{5.26}$$

If $(x,y) \notin K_T$ we have from (5.24), (5.25) and the fact that the support of h is included in $[-1,1]$ that $\tilde{\pi}_T(x,y) = \pi_0(x,y)$. Thus $\tilde{\xi}_T(x,y) = \frac{1}{\pi_0(x,y)}\mathcal{I}[\pi_0](x,y) +$

$\frac{1}{\pi_0(x,y)}\mathcal{I}[d_T](x,y) = \xi_0(x,y) + \frac{1}{\pi_0(x,y)}\mathcal{I}[d_T](x,y)$. It follows that the equality of $\tilde{\xi}_T$ and $\xi_0$on K_T^c will a consequence of the following fact:

$$\text{for } (x,y) \notin K_T, \text{ we have, } \quad \mathcal{I}[d_T](x,y) = 0. \tag{5.27}$$

Let us check that (5.27) holds true. To this end, it is enough that $\mathcal{I}^i[d_T](x,y) = 0$ for $i = 1,2,3$ and $(x,y) \notin K_T$. Since h is a smooth function with compact support on $[-1,1]$, the function d_T and its derivatives vanishes outside of the compact set K_T by (5.25). Recalling $y_0 \neq 0$, for T large enough and for all $x \in \mathbb{R}$, the point $(x,0)$ does not belong to K_T, thus we deduce from (5.21)–(5.22) that $\mathcal{I}^2[d_T](x,y) = \mathcal{I}^3[d_T](x,y) = 0$ when $(x,y) \notin K_T$. It remains to see that $\mathcal{I}^1[d_T](x,y) = 0$ for $(x,y) \notin K_T$. We have by (5.20) and (5.25),

$$\mathcal{I}^1[d_T](x,y) = \frac{h'((x-x_0)/h_1(T))}{\sigma^2 M_T h_1(T)} \int_0^y z h(\frac{z-y_0}{h_2(T)})\mathrm{d}z. \tag{5.28}$$

For $(x,y) \notin K_T$, a first possibility is $x \notin [x_0 - h_1(T), x_0 + h_1(T)]$ that leads to $\mathcal{I}^1[d_T](x,y) = 0$ as h' vanishes outside $[-1,1]$ and thus $h'((x-x_0)/h_1(T)) = 0$. Otherwise, we must have $y \notin [y_0 - h_2(T), y_0 + h_2(T)]$. For simplicity of the presentation, assume that $y_0 > 0$. Then,

$$\begin{aligned}\int_0^y z h(\frac{z-y_0}{h_2(T)})\mathrm{d}z &= \int_0^y z h(\frac{z-y_0}{h_2(T)}) 1_{[y_0-h_2(T), y_0+h_2(T)]}(z)\mathrm{d}z, \\ &= \begin{cases} 0, & \text{if } y \leq y_0 - h_2(T), \\ \int_{y_0-h_2(T)}^{y_0+h_2(T)} z h(\frac{z-y_0}{h_2(T)})\mathrm{d}z, & \text{if } y \geq y_0 + h_2(T). \end{cases}\end{aligned}$$

But $\int_{y_0-h_2(T)}^{y_0+h_2(T)} z h(\frac{z-y_0}{h_2(T)})\mathrm{d}z = h_2(T)\int_{-1}^1 (y_0 + h_2(T)z)h(z)\mathrm{d}z = 0$ using (5.16). This yields to $\mathcal{I}^1[d_T](x,y) = 0$ for $(x,y) \notin K_T$ and (5.27) is proved. It follows that $\tilde{\xi}(x,y) = \xi_0(x,y)$ for $(x,y) \notin K_T$.

The equality between $\tilde{\beta}_T$ and β_0 outside K_T is a consequence of $\tilde{\beta}_T(x,y) = \tilde{\xi}(x,y)/y$, $\beta_0(x,y) = \xi_0(x,y)/y$ for $y \neq 0$.

2) We will prove (5.19) first. From (5.23) and (5.26), we have

$$\tilde{\xi}_T - \xi_0 = \frac{\pi_0 - \tilde{\pi}_T}{\tilde{\pi}_T}\xi_0 + \frac{1}{\tilde{\pi}_T}\mathcal{I}[d_T]. \tag{5.29}$$

On the set K_T, we see that $\tilde{\pi}_T = \pi_0 + d_T$ is lower bounded away from 0 and that ξ_0 is bounded. Using $\|d_T\|_\infty \leq C/M_T$ we deduce that

$$\forall (x,y) \in K_T, \ \left|\tilde{\xi}_T(x,y) - \xi_0(x,y)\right| \leq C\big[\frac{1}{M_T} + \mathcal{I}[d_T](x,y)\big].$$

Now, $|\mathcal{I}[d_T](x,y)| \leq \left|\mathcal{I}^1[d_T](x,y)\right| + \left|\mathcal{I}^2[d_T](x,y)\right| + \left|\mathcal{I}^3[d_T](x,y)\right|$. From (5.21) and (5.25), $\left|\mathcal{I}^2[d_T](x,y)\right| \leq C\|d_T\|_\infty \leq C/M_T$ and by (5.22), $\left|\mathcal{I}^3[d_T](x,y)\right| \leq C\left\|\frac{\partial d_T}{\partial y}\right\|_\infty \leq C/(h_2(T)M_T)$. Using (5.28), we have

$$\left|\mathcal{I}^1[d_T](x,y)\right| \leq \frac{\|h'\|_\infty}{\sigma^2 M_T h_1(T)} \int_{y_0-h_2(T)}^{y_0+h_2(T)} |z|\mathrm{d}s\|h\|_\infty \leq C\frac{h_2(T)}{M_T h_1(T)}.$$

We deduce that,

$$\forall (x,y) \in K_T, \ \left|\tilde{\xi}_T(x,y) - \xi_0(x,y)\right| \leq \frac{C}{M_T}\big[\frac{h_2(T)}{h_1(T)} + 1 + \frac{1}{h_2(T)}\big],$$

which gives (5.19) as $h_2(T) \to 0$.

Eventually, (5.18) follows from the fact that, for T large enough, K_T does not intersect the axis $y = 0$ since $y_0 \neq 0$ and the relation between $\tilde{\beta}_T(x,y) = \tilde{\xi}(x,y)/y$, $\beta_0(x,y) = \xi_0(x,y)/y$.

3) We have,

$$\int_{\mathbb{R}^2} \left|\tilde{\xi}_T(x,y) - \xi_0(x,y)\right|^2 \mathrm{d}x\mathrm{d}y = \int_{K_T} \left|\tilde{\xi}_T(x,y) - \xi_0(x,y)\right|^2 \mathrm{d}x\mathrm{d}y$$

and the third point of the lemma follows from (5.19) with the fact that the Lebesgue measure of K_T is proportional to $h_1(T)h_2(T)$. □

Lemma 4. *Let $\varepsilon > 0$ and assume that for all T large,*

$$M_T^{-1} \leq \varepsilon h_1(T)^{k_1}, \quad M_T^{-1} \leq \varepsilon h_2(T)^{k_2}, \tag{5.30}$$

and

$$\frac{h_2(T)}{h_1(T)} + \frac{1}{h_2(T)} = o(M_T) \text{ as } T \to \infty.$$

Then, if $\varepsilon > 0$ is small enough, we have

$$\tilde{\beta}_T \in \Sigma^{k_1,k_2}(V_0, R, R'),$$

for all T sufficiently large.

Proof From Lemma 3, we know that $\tilde{\beta}_T = \beta_0$ outside K_T and thus is constant equal to $\eta > 0$ outside K_T. For $(x,y) \in K_T$, we have by (5.18) in Lemma 3, $\tilde{\beta}_T(x,y) = \beta_0(x,y) + O\left(\frac{1}{M_T}\left[\frac{h_2(T)}{h_1(T)} + \frac{1}{h_2(T)}\right]\right) = \eta + o(1)$. where C is some constant. Thus for T sufficiently large we have

$$\forall (x,y), \quad 1/R' = \eta/2 < \tilde{\beta}_T(x,y) < 1 < R'$$

where we recall that $R' = 2/\eta > 1$. As V_0 is $\mathcal{C}^1$ and satisfy **HErg**, we can apply the second point of Proposition 4 and deduce that $\tilde{\pi}_T$ is the unique stationary measure associated to $\tilde{\beta}_T$. Recalling Definition 2, the lemma will be shown as soon as we have,

$$\tilde{\pi}_T \in \mathcal{H}^{k_1,k_2}(R).$$

Let us check the Hölder condition with respect to the variable x, as the condition with respect to the variable y is similar. For all $(x,y) \in \mathbb{R}^2$ and $z \in [-1,1]$,

$$\begin{aligned}
\left|\frac{\partial^{\lfloor k_1 \rfloor}\tilde{\pi}_T}{\partial x^{\lfloor k_1 \rfloor}}(x+z,y) - \frac{\partial^{\lfloor k_1 \rfloor}\tilde{\pi}_T}{\partial x^{\lfloor k_1 \rfloor}}(x,y)\right| &\leq \left|\frac{\partial^{\lfloor k_1 \rfloor}\pi_0}{\partial x^{\lfloor k_1 \rfloor}}(x+z,y) - \frac{\partial^{\lfloor k_1 \rfloor}\pi_0}{\partial x^{\lfloor k_1 \rfloor}}(x,y)\right| + \\
&\quad \left|\frac{\partial^{\lfloor k_1 \rfloor}d_T}{\partial x^{\lfloor k_1 \rfloor}}(x+z,y) - \frac{\partial^{\lfloor k_1 \rfloor}d_T}{\partial x^{\lfloor k_1 \rfloor}}(x,y)\right| \\
&\leq \frac{R}{2}|z|^{k_1 - \lfloor k_1 \rfloor} + \left|\frac{\partial^{\lfloor k_1 \rfloor}d_T}{\partial x^{\lfloor k_1 \rfloor}}(x+z,y) - \frac{\partial^{\lfloor k_1 \rfloor}d_T}{\partial x^{\lfloor k_1 \rfloor}}(x,y)\right| \\
&\leq \frac{R}{2}|z|^{k_1 - \lfloor k_1 \rfloor} + \frac{\|h\|_\infty}{M_T h_1(T)^{\lfloor k_1 \rfloor}}\left|h^{(\lfloor k_1 \rfloor)}\left(\frac{x+z-x_0}{h_1(T)}\right) - h^{(\lfloor k_1 \rfloor)}\left(\frac{x-x_0}{h_1(T)}\right)\right|
\end{aligned}$$

where we have successively used $\tilde{\pi}_T = \pi_0 + d_T$, $\pi_0 \in \mathcal{H}^{k_1,k_2}(R/2)$, and the definition (5.25) of d_T. We now write

$$\left|h^{(\lfloor k_1 \rfloor)}\left(\frac{x+z-x_0}{h_1(T)}\right) - h^{(\lfloor k_1 \rfloor)}\left(\frac{x-x_0}{h_1(T)}\right)\right| \leq$$
$$\left|h^{(\lfloor k_1 \rfloor)}\left(\frac{x+z-x_0}{h_1(T)}\right) - h^{(\lfloor k_1 \rfloor)}\left(\frac{x-x_0}{h_1(T)}\right)\right|^{k_1-\lfloor k_1 \rfloor} * (2\left\|h^{(\lfloor k_1 \rfloor)}\right\|_\infty)^{1-(k_1-\lfloor k_1 \rfloor)}$$

which is smaller than $\left\|h^{(\lfloor k_1 \rfloor+1)}\right\|_\infty^{k_1-\lfloor k_1 \rfloor}\left|\frac{z}{h_1(T)}\right|^{k_1-\lfloor k_1 \rfloor} * (2\left\|h^{(\lfloor k_1 \rfloor)}\right\|_\infty)^{1-(k_1-\lfloor k_1 \rfloor)}$. It implies that

$$\left|\frac{\partial^{\lfloor k_1 \rfloor}\tilde{\pi}_T}{\partial x^{\lfloor k_1 \rfloor}}(x+z,y) - \frac{\partial^{(\lfloor k_1 \rfloor)}\tilde{\pi}_T}{\partial x^{\lfloor k_1 \rfloor}}(x,y)\right| \leq |z|^{k_1-\lfloor k_1 \rfloor}\left[\frac{R}{2} + \frac{c_h}{M_T h_1(T)^{k_1}}\right]$$

where $c_h = \|h\|_\infty \left\|h^{(\lfloor k_1 \rfloor+1)}\right\|_\infty^{k_1-\lfloor k_1 \rfloor}(2\left\|h^{(\lfloor k_1 \rfloor)}\right\|_\infty)^{1-(k_1-\lfloor k_1 \rfloor)}$. If one uses (5.30) with any $\varepsilon < \frac{R}{2c_h}$, we deduce

$$\left|\frac{\partial^{\lfloor k_1 \rfloor}\tilde{\pi}_T}{\partial x^{\lfloor k_1 \rfloor}}(x+z,y) - \frac{\partial^{\lfloor k_1 \rfloor}\tilde{\pi}_T}{\partial x^{\lfloor k_1 \rfloor}}(x,y)\right| \leq R|z|^{k_1-\lfloor k_1 \rfloor}.$$

This is the required Hölder control on the derivatives of $\tilde{\pi}_T$ with respect to x. The lemma follows. □

5.3.2. *Proof of the lower bound* (5.4) *on the minimax risk.* Let us recall some notations. We denote $\mathbb{P}_{V,\beta}$ the law of the stationary solution to (5.1)–(5.2) on the canonical space $\mathcal{C}([0,\infty),\mathbb{R}^2)$ and $\mathbb{E}_{V,\beta}$ the corresponding expectation. We denote by $\mathbb{P}_{V,\beta}^{(T)}$ (resp. $\mathbb{E}_{V,\beta}^{(T)}$) the restrictions of this probability (resp. expectation) on $\mathcal{C}([0,T],\mathbb{R}^2)$.

Let $\widetilde{\pi}_T(x_0,y_0)$ be any measurable function from $\mathcal{C}([0,T],\mathbb{R}^2)$ to $\mathbb{R}$. We will estimate by below, for T large,

$$R(\tilde{\pi}_T(x_0,y_0)) := \sup_{\beta \in \Sigma^{k_1,k_2}(V_0,R,R')} \mathbb{E}_{V_0,\beta}^{(T)}\left[(\widetilde{\pi}_T(x_0,y_0) - \pi_{V_0,\beta}(x_0,y_0))^2\right].$$

Let us assume that the following conditions hold true,

$$M_T^{-1} \leq \varepsilon h_1(T)^{k_1}, \quad M_T^{-1} \leq \varepsilon h_2(T)^{k_2}, \tag{5.31}$$

$$\frac{h_2(T)}{h_1(T)} + \frac{1}{h_2(T)} = o(M_T) \text{ as } T \to \infty, \tag{5.32}$$

where ε is sufficiently small to get the conclusion of Lemma 4. We deduce that for T large enough $\tilde{\beta}_T \in \Sigma^{k_1,k_2}(V_0,R,R')$. From (5.15), we have $\beta_0 \in \Sigma^{k_1,k_2}(V_0,R/2,R'/2) \subset \Sigma^{k_1,k_2}(V_0,R,R')$. It follows

$$R(\tilde{\pi}_T(x_0,y_0)) \geq \frac{1}{2}\mathbb{E}_{V_0,\tilde{\beta}_T}^{(T)}\left[(\widetilde{\pi}_T(x_0,y_0) - \pi_{V_0,\tilde{\beta}_T}(x_0,y_0))^2\right] +$$
$$\frac{1}{2}\mathbb{E}_{V_0,\beta_0}^{(T)}\left[(\widetilde{\pi}_T(x_0,y_0) - \pi_{V_0,\beta_0}(x_0,y_0))^2\right].$$

Using Lemma 5 below, we know that $Z^{(T)} = \frac{\mathrm{d}\mathbb{P}^{(T)}_{V_0,\tilde{\beta}_T}}{\mathrm{d}\mathbb{P}^{(T)}_{V_0,\beta_0}}$ exists, and we can write

$$\begin{aligned}
R(\tilde{\pi}_T(x_0,y_0)) &\geq \frac{1}{2}\mathbb{E}^{(T)}_{V_0,\beta_0}\left[(\widetilde{\pi}_T(x_0,y_0)-\pi_{V_0,\tilde{\beta}_T}(x_0,y_0))^2 Z^{(T)}\right] + \\
&\qquad \frac{1}{2}\mathbb{E}^{(T)}_{V_0,\beta_0}\left[(\widetilde{\pi}_T(x_0,y_0)-\pi_{V_0,\beta_0}(x_0,y_0))^2\right] \\
&\geq \frac{1}{2\lambda}\mathbb{E}^{(T)}_{V_0,\beta_0}\left[(\widetilde{\pi}_T(x_0,y_0)-\pi_{V_0,\tilde{\beta}_T}(x_0,y_0))^2 1_{\{Z^{(T)}\geq\frac{1}{\lambda}\}}\right] + \\
&\qquad \frac{1}{2}\mathbb{E}^{(T)}_{V_0,\beta_0}\left[(\widetilde{\pi}_T(x_0,y_0)-\pi_{V_0,\beta_0}(x_0,y_0))^2 1_{\{Z^{(T)}\geq\frac{1}{\lambda}\}}\right] \\
&= \frac{1}{2\lambda}\mathbb{E}^{(T)}_{V_0,\beta_0}\left[[(\widetilde{\pi}_T(x_0,y_0)-\pi_{V_0,\tilde{\beta}_T}(x_0,y_0))^2+(\widetilde{\pi}_T(x_0,y_0)-\pi_{V_0,\beta_0}(x_0,y_0))^2]1_{\{Z^{(T)}\geq\frac{1}{\lambda}\}}\right]
\end{aligned}$$

for all $\lambda > 1$. As $(\widetilde{\pi}_T(x_0,y_0)-\pi_{V_0,\tilde{\beta}_T}(x_0,y_0))^2 + (\widetilde{\pi}_T(x_0,y_0)-\pi_{V_0,\beta_0}(x_0,y_0))^2 \geq \left(\frac{\pi_{V_0,\tilde{\beta}_T}(x_0,y_0)-\pi_{V_0,\beta_0}(x_0,y_0)}{2}\right)^2$ we deduce,

$$R(\tilde{\pi}_T(x_0,y_0)) \geq \frac{1}{8\lambda}(\pi_{V_0,\tilde{\beta}_T}(x_0,y_0)-\pi_{V_0,\beta_0}(x_0,y_0))^2\mathbb{P}^T_{V_0,\beta_0}\left(Z^{(T)}\geq\frac{1}{\lambda}\right).$$

Since $\pi_{V_0,\tilde{\beta}_T} = \tilde{\pi}_T$, $\pi_{V_0,\beta_0} = \pi_0$ and recalling (5.24), (5.25) with $h(0) = 1$ we deduce $\pi_{V_0,\tilde{\beta}_T}(x_0,y_0) - \pi_{V_0,\beta_0}(x_0,y_0) = d_T(x_0,y_0) = 1/M_T$, and it follows,

$$R(\tilde{\pi}_T(x_0,y_0)) \geq \frac{1}{8\lambda}\frac{1}{M_T^2}\mathbb{P}^T_{V_0,\beta_0}\left(Z_T\geq\frac{1}{\lambda}\right).$$

From Lemma 5 below we know that $\inf_{T\geq 0}\mathbb{P}^T_{V_0,\beta_0}\left(Z_T\geq\frac{1}{\lambda}\right) > 0$ for some $\lambda = \lambda_0$ as soon as

$$\sup_{T\geq 0} T\int_{\mathbb{R}^2}\left|\widetilde{\xi}_T(x,y)-\xi_0(x,y)\right|^2 \mathrm{d}x\mathrm{d}y < \infty.$$

Using the third point of Lemma 3, the latter condition is implied by,

$$\sup_T \frac{T}{M_T^2}\left[\frac{h_2(T)^3}{h_1(T)}+\frac{h_1(T)}{h_2(T)}\right] < \infty. \tag{5.33}$$

We deduce that

$$R(\tilde{\pi}_T(x_0,y_0)) \geq \frac{C}{M_T^2}, \tag{5.34}$$

for $C > 0$, if the conditions (5.31), (5.32) and (5.33) are satisfied. It remains to find the larger choice for $1/M_T^2$, subject to the conditions (5.31), (5.32) and (5.33). The optimal choice depends on k_1 and k_2.

Case 1, $k_1 < k_2/2$:

We set $h_1(T) = h_2(T)^2$, and $h_2(T) = \left(\frac{1}{\varepsilon M_T}\right)^{1/k_2}$. The choice for $h_2(T)$ saturates one the conditions in (5.31). Let us see that the other condition holds also true . Indeed $1/M_T = \varepsilon h_2(T)^{k_2} = \varepsilon h_1(T)^{k_2/2} \leq \varepsilon h_1(T)^{k_1}$ for T large, as $k_1 < k_2/2$ and $h_1(T) \to 0$. Thus (5.31) is satisfied.

Plugging the values of $h_1(T)$ and $h_2(T)$ in (5.33), we obtain the constraint $\frac{T}{M_T^2}\left(\frac{1}{\varepsilon M_T}\right)^{1/k_2} \leq C$ for some $C > 0$, that leads us to the choice $M_T = T^{1/(2+1/k_2)} = T^{\frac{k_2}{2k_2+1}}$. Then, the condition (5.32) is satisfied as $k_2 > 1$, indeed $h_2(T)/h_1(T) + 1/h_2(T) = 2/h_2(T) = 2\left(\varepsilon M_T\right)^{1/k_2} = o(M_T)$.

Eventually, we deduce from the application of (5.34),

$$R(\tilde{\pi}_T(x_0,y_0)) \geq C/M_T^2 = CT^{-\frac{2k_2}{2k_2+1}}. \tag{5.35}$$

Case 2, $k_1 \geq k_2/2$:

We set $h_1(T) = h_2(T)^2$, and $h_1(T) = \left(\frac{1}{\varepsilon M_T}\right)^{1/k_1}$. As $1/M_T = \varepsilon h_1(T)^{k_1} = h_2(T)^{2k_1} \leq h_2(T)^{k_2}$, we see that (5.31) is satisfied. Plugging these choices of bandwidths in (5.33), we obtain the constraint $\frac{T}{M_T^2}\left(\frac{1}{\varepsilon M_T}\right)^{1/(2k_1)} \leq C$ for some $C>0$, that leads us to the choice $M_T = T^{1/(2+1/(2k_1))} = T^{\frac{k_1}{2k_1+1/2}}$. Then, the condition (5.32) is satisfied as $k_1 > 1/2$, indeed $h_2(T)/h_1(T)+1/h_2(T) = 2/h_2(T) = 2\left(\varepsilon M_T\right)^{1/(2k_1)} = o(M_T)$.

Eventually, we deduce from the application of (5.34),

$$R(\tilde{\pi}_T(x_0,y_0)) \geq C/M_T^2 = CT^{-\frac{2k_1}{2k_1+1/2}}. \tag{5.36}$$

Gathering (5.35) and (5.36), we have shown Theorem 3. □

Lemma 5. *1) The measure $\mathbb{P}^{(T)}_{V_0,\tilde{\beta}_T}$ is absolutely continuous with respect to $\mathbb{P}^{(T)}_{V_0,\beta_0}$.*

2) Denote $Z^{(T)} = \frac{d\mathbb{P}^{(T)}_{V_0,\tilde{\beta}_T}}{d\mathbb{P}^{(T)}_{V_0,\beta_0}}$ and assume that,

$$\sup_{T\geq 0} T\int_{\mathbb{R}^2}\left|\widetilde{\xi}_T(x,y)-\xi_0(x,y)\right|^2 dxdy < \infty. \tag{5.37}$$

Then, there exist $\lambda_0,\ C>0$ such that,

$$\mathbb{P}^{(T)}_{V_0,\beta_0}\left(Z^{(T)} \geq \frac{1}{\lambda_0}\right) \geq C$$

for all T large enough.

Proof 1) The absolute continuity $\mathbb{P}^{(T)}_{V_0,\tilde{\beta}_T} \ll \mathbb{P}^{(T)}_{V_0,\beta_0}$ and expression for the ratio $Z^{(T)} = \frac{\mathrm{d}\mathbb{P}^{(T)}_{V_0,\tilde{\beta}_T}}{\mathrm{d}\mathbb{P}^{(T)}_{V_0,\beta_0}}$ is obtained by Girsanov formula, changing the drift of the component Y^0 in (5.14) to the drift appearing in the component $Y^{(T)}$ of the stationary solution of the S.D.E.

$$\begin{aligned}\mathrm{d}X_t^{(T)} &= Y_t^{(T)}\mathrm{d}t\\ \mathrm{d}Y_t^{(T)} &= 2\sigma\mathrm{d}B_t - [\sigma^2\tilde{\beta}_T(Y_t^{(T)})Y_t^{(T)} + V_0'(X_t^{(T)})]\mathrm{d}t.\end{aligned}$$

By classical computations (see Theorem 1.12 in [10]), we have

$$\begin{aligned} Z^{(T)} &= \frac{\mathrm{d}\mathbb{P}^{(T)}_{V_0,\tilde{\beta}_T}}{\mathrm{d}\mathbb{P}^{(T)}_{V_0,\beta_0}}((X_s,Y_s)_{0\leq s\leq T})\\ &= \frac{\tilde{\pi}_T(X_0,Y_0)}{\pi_0(X_0,Y_0)}\exp\left\{-\frac{1}{4}\int_0^T \Delta_T(X_s,Y_s)\mathrm{d}Y_s\right.\\ &\left.-\frac{1}{8\sigma^2}\int_0^T [\sigma^2\xi_0(X_s,Y_s)+\sigma^2\Delta_T(X_s,Y_s)+V_0'(X_s,Y_s)]^2-[\sigma^2\xi_0(X_s,Y_s)+V_0'(X_s,Y_s)]^2\mathrm{d}s\right\},\end{aligned} \tag{5.38}$$

where $\Delta_T(x,y) = \tilde{\xi}_T(x,y) - \xi_0(x,y) = y\tilde{\beta}_T(x,y) - y\beta_0(x,y)$. Let us stress that the ratio $\frac{\tilde{\pi}_T(X_0,Y_0)}{\pi_0(X_0,Y_0)}$ in the expression of $Z^{(T)}$ comes from the fact that the two diffusions $(X^{(0)},X^{(0)})_t$ and $(X^{(T)},X^{(T)})_t$ have different initial laws, since they are both stationary, with the different stationary laws.

2) Let us control by below $\mathbb{P}^{(T)}_{V_0,\beta_0}(Z^{(T)} \geq \frac{1}{\lambda})$ for $\lambda > 0$. Recalling the definition of π_0 and $\tilde{\pi}_T$ (see (5.12), (5.17))), we see that $\tilde{\pi}_T/\pi_0$ is equal to 1 outside some compact set (that can be chosen independent of T), and converges uniformly to 1 on this compact set. Hence it is bounded away from zero if T is large, and $\frac{\tilde{\pi}_T(X_0,Y_0)}{\pi_0(X_0,Y_0)} \geq C > 0$ almost surely.

Hence, we will focus on the exponential term in (5.38), that we note $\mathcal{E}^{(T)} = Z^{(T)}\frac{\pi_0(X_0,Y_0)}{\tilde{\pi}_T(X_0,Y_0)}$. We know that under $\mathbb{P}^{(T)}_{V_0,\beta_0}$ the canonical process $(X,Y)_t$ has the same law as $(X^{(0)},Y^{(0)})_t$ defined in (5.13)–(5.14). Hence, the law of $\log(\mathcal{E}^{(T)})$ is the law of the random variable

$$
\begin{aligned}
&-\frac{1}{4}\int_0^T \Delta_T(X_s^{(0)},Y_s^{(0)})\mathrm{d}Y_s^{(0)} \\
&\quad -\frac{1}{8\sigma^2}\int_0^T [\sigma^2\xi_0(X_s^{(0)},Y_s^{(0)}) + \sigma^2\Delta_T(X_s^{(0)},Y_s^{(0)}) + V_0'(X_s^{(0)},Y_s^{(0)})]^2 \\
&\qquad\qquad - [\sigma^2\xi_0(X_s^{(0)},Y_s^{(0)}) + V_0'(X_s^{(0)},Y_s^{(0)})]^2\mathrm{d}s.
\end{aligned}
$$

This random variable is equal, using (5.14) and after some computations, to

$$
\begin{aligned}
&-\frac{\sigma}{2}\int_0^T \Delta_T(X_s^{(0)},Y_s^{(0)})\mathrm{d}B_s - \frac{\sigma^2}{8}\int_0^T \Delta_T(X_s^{(0)},Y_s^{(0)})^2\mathrm{d}s \\
&:= -M_T - I_T.
\end{aligned}
$$

Using the previous considerations we can write that, for T large enough,

$$
\begin{aligned}
\mathbb{P}^{(T)}_{V_0,\beta_0}(Z^{(T)} \geq \frac{1}{\lambda}) &\geq \mathbb{P}^{(T)}_{V_0,\beta_0}(\mathcal{E}^{(T)} \geq \frac{1}{C\lambda}) \\
&= \mathbb{P}^{(T)}_{V_0,\beta_0}\left(-\log\mathcal{E}^{(T)} \leq \log(C\lambda)\right) \\
&= 1 - \mathbb{P}^{(T)}_{V_0,\beta_0}\left(-\log\mathcal{E}^{(T)} > \log(C\lambda)\right) \\
&\geq 1 - \mathbb{P}^{(T)}_{V_0,\beta_0}\left(\left|\log\mathcal{E}^{(T)}\right| > \log(C\lambda)\right) \\
&= 1 - \mathbb{P}\left(|M_T + I_T| > \log(C\lambda)\right)
\end{aligned}
$$

where in the last line we have used that the law of $\log\mathcal{E}^{(T)}$ under $\mathbb{P}^{(T)}_{V_0,\beta_0}$ is the law of $-M_T - I_T$. Assume now that $\lambda > 1/C$, then using Markov inequality, we can write

$$
\begin{aligned}
\mathbb{P}\left(|M_T + I_T| > \log(C\lambda)\right) &\leq \mathbb{P}\left(|M_T| > \frac{1}{2}\log(C\lambda)\right) + \mathbb{P}\left(|I_T| > \frac{1}{2}\log(C\lambda)\right) \\
&\leq \frac{4}{\log(C\lambda)^2}\mathbb{E}(M_T^2) + \frac{2}{\log(C\lambda)}\mathbb{E}(|I_T|)
\end{aligned}
$$

Since $\mathbb{E}(M_T^2) = 2\mathbb{E}(I_T)$ by Ito's isometry, we see that the condition

$$
\sup_{T\geq 0}\ \mathbb{E}(I_T) < \infty \tag{5.39}
$$

is sufficient to get that there exists λ_0 such that for any T large enough we have,

$$\mathbb{P}^{(T)}_{V_0,\beta_0}(Z^{(T)} \geq \frac{1}{\lambda_0}) \geq 1/2.$$

It remains to check that (5.39) holds true. Recalling that $I_T = \frac{\sigma^2}{8}\int_0^T \Delta_T(X_s^{(0)}, Y_s^{(0)})^2 \mathrm{d}s$ and using that the process $(X_t^{(0)}, Y_t^{(0)})_{t\geq 0}$ is stationary, with invariant law π_0 we have

$$\mathbb{E}(I_T) = T\frac{\sigma^2}{8}\mathbb{E}\big[\Delta_T(X_0^{(0)}, Y_0^{(0)})^2\big] = T\frac{\sigma^2}{8}\int_{\mathbb{R}^2} \Delta_T(x,y)^2\pi_0(x,y)\mathrm{d}x\mathrm{d}y.$$

Since π_0 is a bounded function by (5.12), we deduce

$$\mathbb{E}(I_T) \leq CT\int_{\mathbb{R}^2} \Delta_T(x,y)^2\mathrm{d}x\mathrm{d}y.$$

Recalling that by definition $\Delta_T = \tilde{\xi}_T - \xi_0$ and using the assumption (5.37) in the statement of the lemma, we deduce that (5.39) holds true and the lemma follows. □

5.4. **Proof of Theorem 4.** The scheme of the proof is similar to the proof of Theorem 3. However, one needs some modifications taking into account that $y_0 = 0$.

5.4.1. *Constuction of the prior.* The prior is the same as in the proof of Theorem 3 except that we need to modify slightly the functions V_0 and h. Let us give more details. Let k_1, k_2 and $R > 0$. We choose $V_0 : \mathbb{R} \to \mathbb{R}$ a $\mathcal{C}^\infty$ function such that $V_0(x) = x^2$ for $|x|$ large and $V_0'(x) = 0$ on a neighbourhood of x_0, and we define

$$\pi_0(x,y) = c_\eta \exp(-\frac{\eta}{2}[\frac{y^2}{2} + V_0(x)]), \quad \beta_0(x,y) = \eta, \quad \xi_0(x,y) = \eta y,$$

where $0 < \eta < 1/2$ and where c_η is the constant that make π_0 a probability measure. The function π_0 is $\mathcal{C}^\infty$ and it is possible to choose η small enough such that

$$\pi_0 \in \mathcal{H}^{k_1,k_2}(R/2).$$

We know from Section 5.2 that π_0 is the unique stationary measure of the process $(X^{(0)}, Y^{(0)})$ solution to the stochastic differential equation (5.13)–(5.14). If we set $R' = 2/\eta$, then recalling Definition 2 we have $\beta_0 \in \Sigma^{k_1,k_2}(V, R/2, R'/2)$.

Let $h : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^\infty$ function with support on $[-1,1]$ such that,

$$h(0) = 1,\ h'(0) = 0,\ \int_{-1}^1 h(z)\mathrm{d}z = 0,\ \int_0^1 zh(z)\mathrm{d}z = \int_{-1}^0 zh(z)\mathrm{d}z = 0. \tag{5.40}$$

For $T > 0$ we define the perturbation of π_0, as in Section 5.3.1 by

$$\tilde{\pi}_T(x,y) = \pi_0(x,y) + \frac{1}{M_T}h(\frac{x-x_0}{h_1(T)})h(\frac{y}{h_2(T)}),$$

where $M_T \to \infty$, $h_1(T) \to 0$, $h_2(T) \to 0$ will be calibrated latter. Again $\tilde{\pi}_T$ is a smooth probability measure for T large enough and we define

$$\tilde{\beta}_T(x,y) = \beta_{\tilde{\pi}_T}(x,y), \quad \tilde{\xi}_T(x,y) = y\tilde{\beta}_T(x,y) = \xi_{\tilde{\pi}_T}(x,y),$$

where we used the definitions (5.7) and (5.8).

The following lemma gives an assessment of the difference between β_0 and $\tilde{\beta}_T$.

Lemma 6. *1) Recall the definition of the following compact set of* $\mathbb{R}^2$

$$K_T = [x_0 - h_1(T), x_0 + h_1(T)] \times [-h_2(T), h_2(T)].$$

Then, for T *large enough, we have for all* $(x,y) \notin K_T$ *:*

$$\beta_0(x,y) = \tilde{\beta}_T(x,y), \quad \xi_0(x,y) = \tilde{\xi}_T(x,y).$$

2) For $(x,y) \in K_T$*, we have the control,*

$$\left|\beta_0(x,y) - \tilde{\beta}_T(x,y)\right| \leq \frac{C}{M_T}\left[\frac{h_2(T)}{h_1(T)} + \frac{1}{h_2(T)^2}\right], \tag{5.41}$$

$$\left|\xi_0(x,y) - \tilde{\xi}_T(x,y)\right| \leq \frac{C}{M_T}\left[\frac{h_2(T)^2}{h_1(T)} + \frac{1}{h_2(T)}\right], \tag{5.42}$$

where C *is some constant independent of* T*,* $h_1(T)$*,* $h_2(T)$*,* M_T*.*

3) We have

$$\int_{\mathbb{R}^2} \left|\tilde{\xi}_T(x,y) - \xi_0(x,y)\right|^2 dx\,dy \leq \frac{C}{M_T^2}\left[\frac{h_2(T)^5}{h_1(T)} + \frac{h_1(T)}{h_2(T)}\right].$$

Proof 1) We first prove that $\tilde{\xi}_T$ and ξ_0 coincides on K_T^c. Using the notations and arguments of Lemma 3, we know that $\tilde{\xi}_T(x,y) = \xi_0(x,y)$, for all $(x,y) \notin K_T$ is a consequence of $\mathcal{I}[d_T](x,y) = 0$ for $(x,y) \notin K_T$. We recall that $\mathcal{I} = \sum_{i=1}^3 \mathcal{I}^i$ is given by (5.20)–(5.22) and $d_T(x,y) = \tilde{\pi}_T(x,y) - \pi_0(x,y)$ is given by :

$$d_T(x,y) = \frac{1}{M_T} h\left(\frac{x - x_0}{h_1(T)}\right) h\left(\frac{y}{h_2(T)}\right). \tag{5.43}$$

If $(x,y) \notin K_T$, the first situation is $|x - x_0| > h_1(T)$, then $d_T(x,z) = \frac{\partial d_T}{\partial x}(x,z) = \frac{\partial d_T}{\partial y}(x,z) = 0$ for all $z \in \mathbb{R}$ and we deduce that $\mathcal{I}^i[d_T](x,y) = 0$ for $i = 1,2,3$. The second situation is $|y| > h_2(T)$ and $|x - x_0| \leq h_1(T)$. In that case $\mathcal{I}^2[d_T](x,y) = 0$ for T large enough, by using that from assumption on V_0, $V_0'(x) = 0$ for x in some neighbourhood of x_0. From (5.22), we have $\mathcal{I}^3[d_T](x,y) = \frac{2}{M_T h_2(T)} h\left(\frac{x-x_0}{h_1(T)}\right)[h'\left(\frac{y}{h_2(T)}\right) - h'(0)]$ which is equal to 0 since $|y| > h_2(T)$ and $h'(0) = 0$ by (5.40). In order to check that $\mathcal{I}^1[d_T](x,y) = 0$, let us assume for simplicity that $y > h_2(T)$, as the case $y < -h_2(T)$ is similar. Then,

$$\begin{aligned}\mathcal{I}^1[d_T](x,y) &= \frac{1}{\sigma^2 h_1(T)} \int_0^{h_2} z h\left(\frac{z}{h_2(T)}\right) \mathrm{d}z h'\left(\frac{x - x_0}{h_1(T)}\right) \\ &= \frac{h_2(T)^2}{\sigma^2 h_1(T)} \int_0^1 z h(z) \mathrm{d}z h'\left(\frac{x - x_0}{h_1(T)}\right) = 0\end{aligned}$$

by (5.40). Eventually, this gives that $\mathcal{I}[d_T](x,y) = 0$ for $(x,y) \notin K_T$, and thus $\tilde{\xi}_T(x,y) = \xi_0(x,y)$.

The equality between the functions $\tilde{\beta}_T$ and β_0 on K_T^c is a consequence of $\tilde{\beta}_T(x,y) = \tilde{\xi}(x,y)/y$, $\beta_0(x,y) = \xi_0(x,y)/y$ for $y \neq 0$.

2) We first prove (5.42). Recalling (5.29), the fact that $\|\pi_0 - \tilde{\pi}\|_\infty = \|d_T\|_\infty \leq C/M_T$, and that $\tilde{\pi}_T$ is lower bounded on K_T as soon as T is large enough, we

deduce

$$\forall (x,y)\in K_T,\ \Big|\tilde{\xi}_T(x,y)-\xi_0(x,y)\Big| \le C\big[\frac{|\xi_0(x,y)|}{M_T}+\mathcal{I}[d_T](x,y)\big]$$
$$\le C\big[\frac{|y|}{M_T}+\mathcal{I}[d_T](x,y)\big]. \tag{5.44}$$

Now, we use $|\mathcal{I}[d_T](x,y)| \le \big|\mathcal{I}^1[d_T](x,y)\big| + \big|\mathcal{I}^2[d_T](x,y)\big| + \big|\mathcal{I}^3[d_T](x,y)\big|$. From (5.20), we have $\big|\mathcal{I}^1[d_T](x,y)\big| \le C\big|\int_0^y z\mathrm{d}z\big|\big\|\frac{\partial d_T}{\partial x}\big\|_\infty \le C|y|h_2(T)\frac{C}{M_T h_1(T)}$, for all $(x,y)\in K_T=[x_0-h_1(T),x_0+h_1(T)]\times[-h_2(T),h_2(T)]$, and where we have used the expression (5.43) for d_T. As $x\mapsto V'(x)$ vanishes on a neighbourhood of x_0, we get that for T large enough $\mathcal{I}^2[d_T](x,y)=0$ for $(x,y)\in K_T$. From (5.22), we deduce $\big|\mathcal{I}^3[d_T](x,y)\big| \le C|y|\Big\|\frac{\partial^2 d_T}{\partial y^2}\Big\|_\infty \le C\frac{|y|}{M_T h_2(T)^2}$. Collecting the controls on $\mathcal{I}^i[d_T](x,y)$ for $i=1,2,3$, with (5.44) we get

$$\forall (x,y)\in K_T,\ \Big|\tilde{\xi}_T(x,y)-\xi_0(x,y)\Big| \le C\frac{|y|}{M_T}\big[1+\frac{h_2(T)}{h_1(T)}+\frac{1}{h_2(T)^2}\big].$$

Using that for $(x,y)\in K_T$, we have $|y|\le h_2(T)$ and the last equation implies (5.42). Moreover, from the fact that $\tilde{\xi}_T(x,y)-\xi_0(x,y)=y[\tilde{\beta}_T(x,y)-\beta_0(x,y)]$, it implies (5.41).

3) The third point of the lemma in a consequence of the first two points and the fact that the Lebesgue measure of K_T is proportional to $h_1(T)h_2(T)$. □

We now state a result analogous to Lemma 4, but in the situation $y_0=0$.

Lemma 7. *Let $\varepsilon>0$ and assume that for all T large,*

$$M_T^{-1}\le \varepsilon h_1(T)^{k_1},\quad M_T^{-1}\le \varepsilon h_2(T)^{k_2},$$

and

$$\frac{h_2(T)}{h_1(T)}+\frac{1}{h_2(T)^2}=o(M_T)\ \text{ as } T\to\infty.$$

Then, if $\varepsilon>0$ is small enough, we have

$$\tilde{\beta}_T\in\Sigma^{k_1,k_2}(V_0,R,R'),$$

for all T sufficiently large.

We omit the proof of Lemma 7 as it is similar to the proof of Lemma 4 (except that we use (5.41) instead of (5.18)).

5.4.2. *Proof of the lower bound (5.5) on the minimax risk.* We omit most of the details of the proof as it is very similar to the proof given in Section 5.3.2. Indeed, by repeating the arguments of the proof given in Section 5.3.2, relying on Lemmas 6–7 instead of Lemmas 3–4, we deduce that,

$$R(\widetilde{\pi}_T(x_0,0))\ge \frac{C}{M_T^2}, \tag{5.45}$$

as soon as we can find $\varepsilon > 0$, $h_1(T) \to 0$, $h_2(T) \to 0$ and $M_T \to \infty$ satisfying the conditions

$$M_T^{-1} \leq \varepsilon h_1(T)^{k_1}, \quad M_T^{-1} \leq \varepsilon h_2(T)^{k_2}, \tag{5.46}$$

$$\frac{h_2(T)}{h_1(T)} + \frac{1}{h_2(T)^2} = o(M_T) \text{ as } T \to \infty, \tag{5.47}$$

$$\sup_T \frac{T}{M_T^2}\left[\frac{h_2(T)^5}{h_1(T)} + \frac{h_1(T)}{h_2(T)}\right] < \infty. \tag{5.48}$$

Let us maximise $1/M_T^2$ under these three constraints.

Case 1, $k_1 < k_2/3$: we set $h_1(T) = h_2(T)^3$ and $h_2(T) = \left(\frac{1}{\varepsilon M_T}\right)^{1/k_2}$ and the conditions (5.46) hold true, using $k_1 < k_2/3$. With these choices, the condition (5.48) reduces to the boundedness of $\frac{T}{M_T^2} M_T^{-2/k_2}$, which is implied if we set $M_T = T^{-\frac{k_2}{2+2k_2}}$. Next (5.47) holds true as $\frac{h_2(T)}{h_1(T)} + \frac{1}{h_2(T)^2} = O(\frac{1}{h_2(T)^2}) = O(M_T^{2/k_2}) = o(M_T)$ as $k_2 > 2$. Hence, we can use (5.45) with $M_T = T^{-\frac{1}{2+2k_2}}$, yielding to (5.5) in the case $k_1 < k_2/3$.

Case 2, $k_1 \geq k_2/3$: we set $h_1(T) = h_2(T)^3$ and $h_1(T) = \left(\frac{1}{\varepsilon M_T}\right)^{1/k_1}$ and the conditions (5.46) follows. Now, the condition (5.48) reduces to the boundedness of $\frac{T}{M_T^2} M_T^{-2/(3k_1)}$, which yields to the choice $M_T = T^{-\frac{k_1}{2k_1+2/3}}$. Next (5.47) holds true as $\frac{h_2(T)}{h_1(T)} + \frac{1}{h_2(T)^2} = O(\frac{1}{h_2(T)^2}) = O(M_T^{2/(3k_1)}) = o(M_T)$ as $k_1 > 2/3$. Eventually, we deduce (5.5) from (5.45).

6. Appendix

In this section we prove the technical Lemmas 1–2 on the semi group of the process.

6.1. **Proof of Lemma 1.** This proof is exactly the same as the one of Corollary 2.12 in [1], after remarking that the results of Theorem 2.1. in [9] can be applied to S.D.E. with $\mathcal{C}^1$ coefficients. □

6.2. **Proof of Lemma 2.** We first prove that (4.4) holds for $t = D$. Let us denote $\widetilde{K} = \{z \in \mathbb{R}^2 \mid d(z, K) \leq 1\}$ the compact set of points at distance less than 1 of K. Since $D < 1$, we can apply Lemma 1 with the choice of compact set $\widetilde{K}$, to get that if f has support on $K \subset \widetilde{K}$, and $z \in \widetilde{K}$

$$\begin{aligned} |P_D(f)(z)| &\leq \int_{\mathbb{R}^2} |f(z')| p_D(z; z') dz' \\ &\leq \int_{\mathbb{R}^2} |f(z')| p_D^G(z; z') dz' + \int_{\mathbb{R}^2} |f(z')| p_D^U(z; z') dz' \\ &\leq \frac{C_G}{D^2} \int_{\mathbb{R}^2} |f(z')| dz' + C_U \|f\|_\infty e^{-1/(C_U D)}. \end{aligned} \tag{6.1}$$

Hence, it proves (4.4) for $t = D$ and $z \in \widetilde{K}$.

If $z \notin \widetilde{K}$, we let $T_{\widetilde{K}} = \inf\{t \geq 0 \mid Z_t \in \widetilde{K}\}$ the entrance time in the compact set $\widetilde{K}$, which is a stopping time. As the support of f is included in K, we have by

continuity of the process, $P_D(f)(z) = E_z[f(Z_D)] = E_z[f(Z_D)1_{\{T_{\widetilde{K}} \leq t\}}]$. Using the strong Markov property at the time $T_{\widetilde{K}}$ we deduce,

$$P_D(f)(z) = E_z[P_{D-T_{\widetilde{K}}}(f)(Z_{T_{\widetilde{K}}})1_{\{T_{\widetilde{K}} \leq D\}}]. \tag{6.2}$$

By the continuity of the process, we remark that $d(Z_{T_{\widetilde{K}}}, K) = 1$ on the set $T_{\widetilde{K}} \leq D$, and $D - T_{\widetilde{K}}$ in $(0, D)$ as $z \notin \widetilde{K}$. This lead us to consider for $z' \in \widetilde{K}$ with $d(z', K) = 1$ and $s \in (0, D) \subset (0, 1)$, an upper bound for

$$\begin{aligned} |P_s(f)(z')| &\leq \int_{\mathbb{R}^2} |f(w)| p_s^G(z'; w) dw + \int_{\mathbb{R}^2} |f(w)| p_s^U(z'; w) dw \\ &\leq \int_{\mathbb{R}^2} |f(w)| dw \tilde{C}_G + C_U \|f\|_\infty e^{-1/(C_U s)}, \end{aligned} \tag{6.3}$$

where $\tilde{C}_G = \sup\{p_s^G(z', w) \mid s \in (0, 1), w \in K, z' \in \widetilde{K} \text{ with } d(z', K) = 1\}$, and where we used again Lemma 1. We can see that $\tilde{C}_G$ is finite. Indeed, if $z' = (x', y')$ is such that $d(z', K) = 1$ and $w = (w_1, w_2) \in K$, we have $p_s^G(z', w) \leq \frac{C_G}{s^2} \exp(-\frac{1}{C_G}[\frac{(w_2 - y')^2}{s} + \frac{(w_1 - x' - \frac{w_2 + y'}{2} s)^2}{s^3}])$. Using the inequality $A^2 \leq (A-B)^2(1 + 1/s) + B^2(1 + s)$ for any A, B, that entails $(A - B)^2 \geq A^2 \frac{s}{s+1} - B^2 s$, we deduce $\frac{(w_2 - y')^2}{s} + \frac{(w_1 - x' - \frac{w_2 + y'}{2} s)^2}{s^3} \geq \frac{(w_2 - y')^2}{s} + \frac{(w_1 - x')^2}{s^3} \frac{s}{s+1} - \frac{(w_2 + y')^2}{2s} s$. Using that $s < 1$ and that $|w_2|$ and $|y'|$ are bounded by some constant depending on the compact K, we deduce $\frac{(w_2 - y')^2}{s} + \frac{(w_1 - x' - \frac{w_2 + y'}{2} s)^2}{s^3} \geq d(z', w)^2/(2s) - \tilde{C}_K$ for some constant $\tilde{C}_K$ depending on the compact K only. It gives $p_s^G(z', w) \leq \frac{C_G}{s^2} \exp(-\frac{d(z', w)^2}{C_G 2s}) \exp(\frac{\tilde{C}_K}{C_G})$. As $w \in K$ and $d(w', K) = 1$, we deduce that $p_s^G(z', w) \leq \frac{C_G}{s^2} \exp(-\frac{1}{C_G 2s}) \exp(\frac{\tilde{C}_K}{C_G})$ and thus $\widetilde{C}_G$ is finite. Joining (6.3) and (6.2), we deduce that, for $z \notin \widetilde{K}$

$$\begin{aligned} |P_D(f)(z)| &\leq \widetilde{C}_G \|f\|_{L^1(\mathbb{R}^2)} + C_U \|f\|_\infty E_z[e^{-1/(C_U(D - T_{\widetilde{K}}))} 1_{\{T_{\widetilde{K}} \leq D\}}] \\ &\leq \widetilde{C}_G \|f\|_{L^1(\mathbb{R}^2)} + C_U \|f\|_\infty e^{-1/(C_U D)}. \end{aligned} \tag{6.4}$$

The control (4.4) for $t = D$ is now a consequence of (6.1) and (6.4). Eventually, we prove that (4.4) for $t = D$ is sufficient to deduce the lemma. Let $0 < D < 1$ and $t > D$, then for $z \in \mathbb{R}^2$, we write

$$|P_t(f)(z)| \leq \int_{\mathbb{R}^2} p_{t-D}(z, z') |P_D(f)(z')| dz'$$

and using the estimate (4.4) for $|P_D(f)(z')|$ gives the result for $|P_t(f)(z)|$. □

References

[1] Cattiaux P., Len, J. and Prieur, C. Estimation for stochastic damping Hamiltonian systems under partial observation I. Invariant density. *Stochastic Process. Appl.*, **124**, no. 3, (2014), 12361260.

[2] Comte, F. and Lacour, C. Anisotropic adaptive kernel deconvolution. *Annales de lInstitut Henri Poincar - Probabilits et Statistiques*, Vol. **49**, No. 2, (2013), 569609.

[3] Comte F, and Merlevède, F. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Statist.*, 6:211–238, 2002. New directions in time series analysis (Luminy, 2001).

[4] Comte, F. and Merlevède, F. Super optimal rates for nonparametric density estimation via projection estimators. *Stochastic Process. Appl.*, 115(5):797–826, 2005.

[5] Comte, F., Prieur, C. and Samson, A. Adaptive estimation for stochastic damping Hamiltonian systems under partial observation. *Stochastic Process. Appl.* **127** (2017), no. 11, 3689–3718.

[6] Dalalyan, A. and Reiß, M. Asymptotic statistical equivalence for ergodic diffusions : the multidimensional case *Probab. Theory. Relat. Fields*, **137**(1), 25–47

[7] Ditlevsen, S. and Srensen, M. Inference for Observations of Integrated Diffusion Processes. Scandinavian Journal of Statistics, 31 (2004), 417–429. doi:10.1111/j.1467-9469.2004.02_023.x

[8] Ditlevsen, S. and Samson, A. Hypoelliptic diffusions: discretization, filtering and inference from complete and partial observations, *J Royal Statistical Society B,* 81, (2019) 361–384.

[9] Konakov, V., Menozzi, S. and Molchanov, S. Explicit parametrix and local limit theorems for some degenerate diffusion processes, *Annales de lInstitut Henri Poincar - Probabilits et Statistiques*, Vol. **46**, No. 4, (2010), 908923.

[10] Kutoyants, Y. Statistical inference for ergodic diffusion processes. Springer Series in Statistics. Springer-Verlag London, Ltd., London, 2004. xiv+481 pp. ISBN: 1-85233-759-1.

[11] Len, J.R. and Samson, A. Hypoelliptic stochastic FitzHugh-Nagumo neuronal model: mixing, up-crossing and estimation of the spike rate, *Annals of Applied Probability*, 28, (2018), 2243–2274, .

[12] Nishiyama, Y. Estimation for the invariant law of an ergodic diffusion process based on high-frequency data, *Journal of Nonparametric Statistics*, 23:4, 909-915, DOI: 10.1080/10485252.2011.591397

[13] Nguyen, H-T. Density estimation in a continuous-time stationary Markov process. *Ann. Statist.*, 7(2):341–348, 1979.

[14] Strauch, C. Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. *Ann. Statist.*, 46(6B):3451–3480, 2018.

[15] Talay, D. Stochastic Hamiltonian Systems: Exponential Convergence to the Invariant Measure, and Discretization by the Implicit Euler Scheme, *Markov Processes Relat. Fields*, **8**, (2002), 136.

[16] Tsybakov, A. Introduction to nonparametric estimation. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. Springer Series in Statistics. Springer, New York, 2009. xii+214 pp. ISBN: 978-0-387-79051-0.

[17] Wu, L. Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems, *Stochastic Processes and their Applications*, **91**, (2001), 205238.

Laboratoire de Probabilités Statistique et Modélisation, UMR 7599, Université Denis Diderot, 77251 Paris Cedex 05, France

Laboratoire de Mathématiques et Modélisation d'Evry, CNRS, Univ Evry, Université Paris-Saclay, 91037, Evry, France

Graduate School of Mathematical Sciences, University of Tokyo: 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan.