PROBABILISTIC PROPERTIES AND PARAMETRIC INFERENCE OF SMALL VARIANCE NONLINEAR SELF-STABILIZING STOCHASTIC DIFFERENTIAL EQUATIONS

Valentine Genon-Catalot, Catherine Laredo

PROBABILISTIC PROPERTIES AND PARAMETRIC INFERENCE OF SMALL VARIANCE NONLINEAR SELF-STABILIZING STOCHASTIC DIFFERENTIAL EQUATIONS.

VALENTINE GENON-CATALOT[(1)] AND CATHERINE LARÉDO[2]

Abstract. We consider a process (X_t) solution of a one-dimensional nonlinear self-stabilizing stochastic differential equation, with classical drift term $V(\alpha, x)$ depending on an unknown parameter α, self-stabilizing term $\Phi(\beta, x)$ depending on another unknown parameter β and small noise amplitude ε. Self-Stabilization is the effect of including a mean-field interaction in addition to the state-dependent drift. Adding this term leads to a nonlinear or Mac Kean-Vlasov Markov process with transitions depending on the distribution of X_t. We study the probabilistic properties of (X_t) as ε tends to 0 and exhibit a Gaussian approximating process for (X_t). Next, we study the estimation of (α, β) from a continuous observation of $(X_t, t \in [0, T])$. We build explicit estimators using an approximate log-likelihood function obtained from the exact log-likelihood function of a proxi-model. We prove that, for fixed T, as ε tends to 0, α can be consistently estimated with rate ε^{-1} *but not* β. Then, considering n *i.i.d.* sample paths $(X_t^i, i = 1, \ldots, n)$, we build consistent and asymptotically Gaussian estimators of (α, β) with rates $\sqrt{n}\varepsilon^{-1}$ for α and $\sqrt{n}$ for β. Finally, we prove that the statistical experiments generated by (X_t) and the proxi-model are asymptotically equivalent in the sense of the Le Cam Δ-distance both for the continuous observation of one path and for n *i.i.d.* paths under the condition $\sqrt{n}\varepsilon \to 0$, which justifies our statistical method.

Keywords and phrases: Nonlinear diffusion processes, Mac-Kean-Vlasov stochastic differential equation, Small noise, Asymptotic properties, Parametric inference, Approximate likelihood.

AMS Classification. 60J60, 60J99, 62F12, 62M05

1. Introduction

In this paper we study a dynamical system subject to three sources of forcing. First, there is a potential which describes the geometry of the state space of the dynamical

[(1)]: Université de Paris, CNRS, MAP5, UMR 8145, F-75006 Paris, France.
email: valentine.genon-catalot@u-paris.fr,

[(2)]: INRAE, MaIAGE, Université Paris-Saclay, F-78350 Jouy-en-Josas, France & LPSM, UMR 8001, Paris.
email:catherine.laredo@inrae.fr

.

system. Second, small random pertubations occur on this dynamical system via Brownian motion with small noise. Third, there is an additional source of forcing called self-stabilization. Assuming a large population of identical dynamical systems subject to the same evolution rules, this last term characterizes their influence. They act on the individual through an attractive potential averaged over the whole population and results in an additional term in the underlying potential drift.

More precisely, we consider the one-dimensional process

$$dX_t = V(\alpha, X_t)dt - b(\theta, t, \varepsilon, X_t)dt + \varepsilon dW_t, \quad X_0 = \eta \ \sim \mu, \tag{1}$$

$$b(\theta, t, \varepsilon, x) = \int_{\mathbb{R}} \Phi(\beta, x - y) u_t^{\theta,\varepsilon,\mu}(dy), \tag{2}$$

where (W_t) is a Wiener process, η is random variable independent of (W_t) with distribution μ, $u_t^{\theta,\varepsilon,\mu}(dy)$ is the distribution of $X_t = X_t^{\theta,\varepsilon,\mu}$, $V : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $\Phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are deterministic Borel functions and $\theta = (\alpha, \beta) \in \Theta = \Theta_\alpha \times \Theta_\beta \subset \mathbb{R}^2$ is an unknown parameter. A solution of (1) is the couple $(X_t, u_t^{\theta,\varepsilon,\mu}(dy))_{t\geq 0}$ composed of the stochastic process (X_t) and the family of distributions $(u_t^{\theta,\varepsilon,\mu})$. The function $x \to b(\theta, t, \varepsilon, x)$ depends on θ, t, ε and the initial distribution through $u_t^{\theta,\varepsilon,\mu}$. When defined, the process (X_t) is a time-inhomogeneous Markov process known as self-stabilizing diffusion or nonlinear stochastic differential equation.

The integral term involves the process marginal law $u_t^{\theta,\varepsilon,\mu}$. It models the interaction between X_t and a fixed point x in the state space by means of an interaction function $\Phi(\beta, .)$. The interaction between the various dynamical systems is of *Mean Field* type, this effective additional drift term being a measure of the average force exerted by independent copies of itself through the potential $\Phi(\beta, .)$. This model was first described by Mac Kean (1966). These equations arise in the modeling of granular media by interacting particle systems (see Benedetto *et al.* 1997 for the issues in Physics; Snitzman,1991 for a survey). They could also occur when modeling Epidemic Dynamics with two levels of mixing (see Ball *et al.*, 1997; Ball and Sirl, 2020 and Forien and Pardoux , 2020).

The aim of this paper is to study both the asymptotic properties of the process (X_t) solution of (1) as ε tends to 0 and the estimation of θ from a continuous observation of $(X_t, t \in [0, T]))$ on the fixed time interval $[0, T]$ or from n independent paths $((X_t^i, t \in [0, T]), i = 1, \ldots, n)$. From now on, we assume that $\varepsilon \leq 1$.

Our plan is as follows. We recall in Section 2 the existing results for these processes and detail in particular the conditions ensuring existence and unicity of a strong solution of (1) and illustrate these equations on an explicit example. In Section 3, we extend the bounds existing for small pertubations of dynamical systems (see *e.g.* Azencott, 1982, Freidlin and Wentsell, 2012) to these nonlinear diffusion processes and prove the existence of a time-dependent Gaussian process approximating the paths of (X_t) as $\varepsilon \to 0$. We prove the existence of accurate bounds for the remainder terms which are crucial for the inference. Section 4 is devoted to the estimation of the parameters present in the potential term $V(\alpha, .)$ and in the interaction term $\Phi(\beta, .)$ based on a continuous observation of (X_t) on $[0, T]$ with T fixed. We build explicit estimators using an approximate likelihood function obtained from the exact likelihood of a proxi-model consisting in a time dependent diffusion process derived from the results of Section 3. We obtain that only α can be

estimated. We build consistent and asymptotically Gaussian estimators at rate ε^{-1} of α for β known or unknown. Next, we study the case where we observe n independent trajectories $((X_t^i, 0 \le t \le T), i = 1, \dots, n)$. This is reminiscent of the situation of particle systems (see Sznitman (1991)), where, for an infinite number of interacting particles, a result of *Propagation of chaos* is proved for the limit, which corresponds to asymptotically independent self-stabilizing diffusions.

We build consistent and asymptotically Gaussian estimators of (α, β) with different rates of convergence: we prove that α is estimated at rate $\sqrt{n}/\varepsilon$ and β at rate $\sqrt{n}$ as ε tends to 0 and n tends to infinity in such a way that $\varepsilon\sqrt{n}$ tends to 0. Finally, asymptotic equivalence of the experiments generated by (X_t) and the proxi-model in the sense of the Le Cam distance is proved in Section 5. Asymptotic equivalence also holds in the case of n *i.i.d.* trajectories. Concluding remarks and perspectives are given in Section 7. Proofs are gathered in the Appendix.

2. Recap on nonlinear diffusion processes and preliminary results

2.1. **Preliminaries.** Let us first present heuristically the situation where these equations are obtained. They first appear for modeling interacting particle systems.
Consider a system of N equations of N interacting dynamical systems. Then, equations such as (1) occur as limits of the interacting dynamical systems, as the number N of systems in a set of identical ones tends to infinity. Indeed, consider an interaction function $\Phi(\beta, .)$ and let $\Phi(\beta, x - y)$ model the force of interaction between to points x and y in the state space. Suppose that we are given N dynamical systems, with dynamics described by the stochastic differential equation

$$dX_t^{i,N} = V(\alpha, X_t^{i,N})dt - \frac{1}{N}\sum_{j=1}^{N} \Phi(\beta, X_t^{i,N} - X_t^{j,N})dt + \varepsilon dB_t^i, \quad X_0^{i,N} = X_0^i \quad \text{with } 1 \le i \le N.$$

Here the B^i are independent Brownian motions and (X_0^i) are independent identically distributed random variables, independent of $(B^i, i = 1 \dots, N)$. The self-stabilizing effect originates in the global action of the system on a particular one in the limit $N \to \infty$. Under suitable assumptions, the empirical measures $\frac{1}{N}\sum_{j=1}^{N} \delta_{X_t^{j,N}}$ converge to some law $u_t^{\alpha,\beta,\varepsilon}$ for each time t and noise intensity ε, and each dynamical system converges in probability to the solution of

$$dX_t = V(\alpha, X_t)dt - \int \Phi(\beta, X_t - x)u_t^{\alpha,\beta,\varepsilon}(dx)dt + \varepsilon dW_t.$$

Moreover, asymptotic independence of the different dynamical systems is obtained in the sense of the McKean-Vlasov limit (*Propagation of Chaos*, see Sznitman, 1991 for a review).

2.2. **First assumptions.** To ensure existence and unicity of a strong solution to (1) for all $\theta = (\alpha, \beta) \in \Theta \subset \mathbb{R} \times \mathbb{R}$, we consider the following assumptions taken from Herrmann *et al.* (2008).

- [H0] For all α, β, the functions $x \to V(\alpha, x)$ and $x \to \Phi(\beta, x)$ are locally Lipschitz.
- [H1] For all β, the function $x \to \Phi(\beta, x)$ is odd, increasing and grows at most polynomially: there exist $K(\beta) > 0$ and $r(\beta) \in \mathbb{N}$ such that
  $|\Phi(\beta, x) - \Phi(\beta, y)| \le |x - y|(K(\beta) + |x|^{r(\beta)} + |y|^{r(\beta)}),\ x, y \in \mathbb{R}$.

- [H2] For all α, the function $x \to V(\alpha, x)$ is continuously differentiable and there exist $K_V(\alpha) > 0$ and $R_0(\alpha) > 0$ such that

$$\frac{\partial V}{\partial x}(\alpha, x) \leq -K_V(\alpha), \text{ for } |x| \geq R_0(\alpha).$$

Assumption [H2] ensures that there exists $L(\alpha) > 0$ such that

$$\forall x \in \mathbb{R}, \quad \frac{\partial V}{\partial x}(\alpha, x) \leq L(\alpha). \tag{3}$$

Under [H0]-[H1]-[H2], according to Theorem 2.13 in Herrmann *et al.*(2008), if $\mathbb{E}(\eta^{8q^2}) < +\infty$ where $q = [(r(\beta)/2) + 1]$, then, for all θ, there exists a drift term $b(\theta, t, \varepsilon, x) = b^{\mu}(\theta, t, \varepsilon, x)$ such that (1) admits a unique strong solution $(X_t = X_t^{\theta,\varepsilon,\mu})$ satisfying $b(\theta, t, \varepsilon, x) = \int_{\mathbb{R}} \Phi(\beta, x - y) u_t^{\theta,\varepsilon,\mu}(dy)$ and X is the unique strong solution of (1). Moreover, for all $p \in \{1, \ldots, 4q^2\}$, whenever $\mathbb{E}(\eta^{2p}) < +\infty$, $\sup_{t\geq 0} \mathbb{E}(X_t^{2p}) < +\infty$. In particular, if η is deterministic, for all $p \in \mathbb{N}$, $\sup_{t\geq 0} \mathbb{E}(X_t^{2p}) < +\infty$.

Before dealing with the general case, we illustrate the small variance properties on an explicit example.

2.3. **An explicit example.** Consider the model where:

$$V(\alpha, x) = -\alpha x \quad \text{with } \alpha > 0; \quad \Phi(\beta, x) = \beta x \quad \text{with } \beta > 0. \tag{4}$$

We consider successively the case where $X_0 = \eta$ is random and the case $X_0 = x_0$ deterministic.

Case X_0 random: The value $r(\beta)$ of Assumption [H1] is $r(\beta) = 1$. Therefore, according to Herrmann *et al.* (2008), if η satisfies $\mathbb{E}(\eta^8) < +\infty$, Equation (1) has a unique strong solution. As $b(\theta, t, \varepsilon, x) = \beta(x - \mathbb{E}_\theta(X_t))$, Equation (1) writes:

$$dX_t = -\alpha X_t dt - \beta(X_t - \mathbb{E}_\theta(X_t))dt + \varepsilon dW_t, \quad X_0 = \eta.$$

with η independent of (W_t). Obviously, $\mathbb{E}_\theta X_t = \mathbb{E}_\theta X_0 - \alpha \int_0^t \mathbb{E}_\theta X_s ds$.
Hence $\mathbb{E}_\theta(X_t) := m_\theta(t, \varepsilon) = m_\theta(t) = m_\theta(0) \exp(-\alpha t)$. Therefore, we obtain:

$$dX_t = -(\alpha + \beta)X_t dt + \beta m_\theta(0) \exp(-\alpha t))dt + \varepsilon dW_t, \quad X_0 = \eta. \tag{5}$$

This is a time-inhomogeneous Ornstein-Uhlenbeck process. Solving (5) yields

$$X_t = (X_0 - m_\theta(0))e^{-(\alpha+\beta)t} + m_\theta(0)e^{-\alpha t} + \varepsilon g_t(\theta) \text{ with } g_t(\theta) = e^{-(\alpha+\beta)t} \int_0^t e^{(\alpha+\beta)s} dW_s. \tag{6}$$

The process (X_t) has a unique invariant distribution, the Gaussian law $\mathcal{N}(0, \varepsilon^2/2(\alpha + \beta))$. As $\varepsilon \to 0$, X_t converges to $x_t(\theta) = (X_0 - m_\theta(0)) \exp(-(\alpha + \beta)t) + m_\theta(0) \exp(-\alpha t)$ uniformly on $[0, T]$, solution of the ordinary differential equation (ODE) corresponding to $\varepsilon = 0$ in (5).

$$dx_t = -(\alpha + \beta)x_t dt + \beta m_\theta(0) \exp(-\alpha t)dt, \quad x_0 = X_0.$$

Case $X_0 = x_0$ deterministic: In this case $m_\theta(0) = x_0$, $m_\theta(t) = x_0 e^{-\alpha t}$,

$$x_t(\alpha) = x_0 e^{-\alpha t}; \quad X_t = x_0 e^{-\alpha t} + \varepsilon g_t(\theta) \text{ with } g_t(\theta) = e^{-(\alpha+\beta)t} \int_0^t e^{(\alpha+\beta)s} dW_s. \tag{7}$$

The process converges uniformly on $[0,T]$ to $x_t(\alpha)$. Note that β is no longer present in the limit.

3. Properties on a fixed time interval.

Let us consider now the general case. We first study the convergence of the process as ε tends to 0 and prove bounds for its moments which are uniform on a fixed time interval $[0,T]$ under [H0]-[H2] and

- [H3-k] The functions $x \to V(\alpha,x)$ and $x \to \Phi(\beta,x)$ have continuous partial derivatives up to order k and these derivatives have polynomial growth: for all α,β, and all $i, i \le k$, there exist constants $k(\alpha)>0, k(\beta)>0, \gamma(\alpha)\ge 0, \gamma(\beta)\ge 0$, such that

$$\forall x \in \mathbb{R}, \quad |\frac{\partial^i V}{\partial x^i}(\alpha,x)| \le k(\alpha)(1+|x|^{\gamma(\alpha)}), \quad |\frac{\partial^i \Phi}{\partial x^i}(\beta,x)| \le k(\beta)(1+|x|^{\gamma(\beta)}).$$

Consider the equation corresponding to $\varepsilon = 0$ in (1)

$$dx_t(\theta) = V(\alpha, x_t(\theta))dt - b(\theta,t,0,x_t(\theta))dt \quad x_0(\theta)=\eta, \text{ where} \tag{8}$$

$$b(\theta,t,0,x) = \int \Phi(\beta, x-y) u_t^{\theta,0}(dy), \text{ and } u_t^{\theta,0}(dy) \text{ is the distribution of } x_t(\theta), \tag{9}$$

and $u_t^{\theta,0}(dy) = u_t^{\theta,0,\mu}(dy)$. Clearly, Equation (8) is a nonlinear differential equation of the same type as (1) without noise. Thus, it has a unique solution process $(x_t(\theta), u_t^{\theta,0}(dy), t \ge 0)$. Note that $x_t(\theta)$ is deterministic if $x_0(\theta) = x_0$ is deterministic.

Proposition 3.1. *Assume [H0]-[H2] and [H3-2]. Let $x_t(\theta)$ denote the solution of (8). Then,*

$$\forall p \ge 1, \quad \mathbb{E}_\theta (X_t - x_t(\theta))^{2p} \le \varepsilon^{2p} C_{2p}(t) e^{2pL(\alpha)t},$$

where $C_2(t) = t$, $C_{2p}(t)$ is a continuous nondecreasing function of t and $L(\alpha)$ is defined in (3).

Remark 3.1. *The previous proposition extends Lemma 3.1 of Hermann et al (2008) where only the case $\eta = x_0$ deterministic and $p = 1$ is considered.*

When $\eta = x_0$ is deterministic, there is no source of randomness in equation (8) and $x_t(\theta)$ is deterministic. Therefore, $u_t^{\theta,0}(dy)$ is the Dirac mass at $x_t(\theta)$ and this implies

$$b(\theta,t,0,x_t(\theta)) = \int \Phi(\beta, x_t(\theta)-y) u_t^{\theta,0}(dy) = \Phi(\beta, x_t(\theta) - x_t(\theta)) = 0.$$

Therefore, the solution of (8), $x_t(\theta) = x_t(\alpha)$, does no longer depend on β and becomes

$$dx_t(\alpha) = V(\alpha, x_t(\alpha))dt, \; x_0(\alpha) = x_0. \tag{10}$$

In this case, we can prove the stronger result.

Corollary 3.1. *Assume [H0]-[H2], [H3-2], $X_0 = x_0$ deterministic. Then, if $x_t(\alpha)$ satisfies (10),*

$$\sup_{t \le T} \varepsilon^{-2} |\mathbb{E}_\theta (X_t - x_t(\alpha))| = O(1). \tag{11}$$

Moreover, in the special case where $V(\alpha,x) = -\alpha x$, $\mathbb{E}_\theta(X_t) = x_t(\alpha)$.

From now on, we assume

[H4] The initial condition of (X_t) is non random : $X_0 = x_0$ deterministic.

As for classical stochastic differential equation with small variance, (X_t) possesses a Gaussian diffusion approximation. The expansion relies on regularity properties of $b(\theta, t, \varepsilon, x)$ with respect to ε and x at $(\theta, t, 0, x)$.

Lemma 3.1. *Assume [H0]-[H2], [H3-2] and [H4].Then*
(i) For all $\theta, t \geq 0$, $(\varepsilon, x) \to b(\theta, t, \varepsilon, x)$ is continuously differentiable on $[0, +\infty) \times \mathbb{R}$.
(ii) $\lim_{\varepsilon \to 0} b(\theta, t, \varepsilon, x) = \Phi(\beta, x - x_t(\alpha))$.
(iii) At $\varepsilon = 0$, $\frac{\partial b}{\partial \varepsilon}(\theta, t, 0, x) = 0$ and $\frac{\partial b}{\partial x}(\theta, t, 0, x) = \frac{\partial \phi}{\partial x}(\beta, x - x_t(\alpha))$.

Analogously to the expansions obtained in Azencott (1982), Freidlin and Wenzell (2012) for small pertubations of dynamical systems, the following holds.

Theorem 3.1. *Assume [H0]-[H2], [H3-3] and [H4]. Then,*

$$X_t = x_t(\alpha) + \varepsilon g_t(\theta) + \varepsilon^2 R^\varepsilon(t), \tag{12}$$

where $x_t(\alpha)$ satisfies (10)*, $(g_t(\theta))$ is the Ornstein-Uhlenbeck process (recall that $\frac{\partial \Phi}{\partial x}(\beta, 0) \geq 0$)*

$$dg_t(\theta) = \left(\frac{\partial V}{\partial x}(\alpha, x_t(\alpha)) - \frac{\partial \Phi}{\partial x}(\beta, 0) \right) g_t(\theta) + dW_t, \quad g_0(\theta) = 0, \tag{13}$$

and $R^\varepsilon(t)$ defined in (12) *satisfies*

$$\sup_{t \in [0,T], \varepsilon > 0} \mathbb{E}_\theta |R^\varepsilon(t)| = O(1) \text{ and for all } p \geq 1, \sup_{t \in [0,T], \varepsilon > 0} \mathbb{E}_\theta \left((R^\varepsilon(t) - \mathbb{E}_\theta R^\varepsilon(t))^{2p} \right) = O(1).$$

Let us define the two functions,

$$a(\theta, t) = \frac{\partial V}{\partial x}(\alpha, x_t(\alpha)) - \frac{\partial \Phi}{\partial x}(\beta, 0); \quad A(\theta, t) = \int_0^t a(\theta, u) du. \tag{14}$$

Remark 3.2. *Under [H0]-[H2], [H3-3] and [H4], Equation* (13) *can be solved. Using* (14)*, we get*

$$g_t(\theta) = \int_0^t \exp\left(\int_s^t a(\theta, u) du \right) dW_s = e^{A(\theta,t)} \int_0^t e^{-A(\theta,s)} dW_s. \tag{15}$$

The proof of Theorem 3.1 provides an explicit expression of $R^\varepsilon(t)$ that will be used for bounding it further, together with bounds detailed in the following corollary which are crucial for the statistical study. Let

$$D(\theta, t, \varepsilon, x) = b(\theta, t, \varepsilon, x) - \Phi(\beta, x - x_t(\alpha)). \tag{16}$$

Corollary 3.2. *Assume [H0]-[H2], [H3-3] and [H4]. Then $D(\theta, t, \varepsilon, X_t)$satisfies*

$$\sup_{t \in [0,T], \varepsilon > 0} \varepsilon^{-2} \left| \mathbb{E}_\theta D(\theta, t, \varepsilon, X_t) \right| = O(1), \tag{17}$$

$$\forall p \geq 1, \quad \sup_{t \in [0,T], \varepsilon > 0} \varepsilon^{-6p} \mathbb{E}_\theta \left((D(\theta, t, \varepsilon, X_t) - \mathbb{E}_\theta D(\theta, t, \varepsilon, X_t))^{2p} \right) = O(1). \tag{18}$$

Remark 3.3. *Consider Example* (7). *The process* (X_t) *is equal to its Gaussian approximation, i.e. the remainder term* $R^\varepsilon(t)$ *of Theorem 3.1 is null. Moreover,* $D(\theta, t, \varepsilon, x) \equiv 0$.

4. Parametric inference as ε tends to 0.

As it is usual in statistics, we consider the canonical space associated with the observation of $(X_t)_{t\in[0,T]}$, $(\Omega, \mathcal{F}, (\mathcal{F}_t, t \in [0,T]), \mathbb{P}_\theta)$, where $\Omega = C([0,T])$ is the space of continuous real-valued functions defined on $[0,T]$ endowed with the Borel σ-field associated with the uniform convergence on $[0,T]$, $(X_t, t \in [0,T])$ is the canonical process $(X_t(\omega) = \omega(t))$, $(\mathcal{F}_t, t \in [0,T])$ is the canonical filtration and $\mathbb{P}_\theta$ is the distribution of (1) on $C([0,T])$.

We assume that $X_0 = x_0$ is deterministic and known, so that (X_t) converges uniformly on [0,T] to $(x_t(\alpha))$, solution of (10). Let us consider the problem of estimating (α, β) from a continuous observation $(X_t, t \in [0,T])$. In classical stochastic differential equations with small diffusion coefficient, all drift parameters have rate ε^{-1}. Here, the situation is different since we observe that the parameter β is no longer present in the limiting ODE (10). We show that α can be consistently estimated as ε tends to 0, but not β.

4.1. Approximate likelihood.

The exact log-likelihood of $(X_t, t \in [0,T])$ contains the term $b(\theta, t, \varepsilon, X_t)$ which is not explicit. This is why we proceed differently. We have proved in the previous section that, for small ε, $D(\theta, t, \varepsilon, X_t) = b(\theta, t, \varepsilon, X_t) - \Phi(\beta, X_t - x_t(\alpha))$ is small (see Corollary 3.2). Therefore, we consider the process (ξ_t)

$$d\xi_t = [V(\alpha, \xi_t) - \Phi(\beta, \xi_t - x_t(\alpha))]dt + \varepsilon dW_t, \quad \xi_0 = x_0. \tag{19}$$

This process has already been used by Herrmann *et al.*(2008) for proving large deviations results. It is close to (X_t) for small ε. The euristics is that $b(\theta, t, \varepsilon, x) = \int \Phi(\beta, x - y) u_t^{\theta,\varepsilon}(dy)$ is close to $\Phi(\beta, x - x_t(\alpha))$ as the distribution of X_t, $u_t^{\theta,\varepsilon}(dy)$, converges to the Dirac mass at $x_t(\alpha)$.

So, we define the following approximate likelihood or contrast where we replace the true drift function by its approximation for small ε,

$$\Lambda_{\varepsilon,T}(\theta) = \frac{1}{\varepsilon^2}\int_0^T H(\theta, s, X_s)dX_s - \frac{1}{2\varepsilon^2}\int_0^T H^2(\theta, s, X_s)ds, \quad \text{with} \tag{20}$$

$$H(\theta, s, x) = V(\alpha, x) - \Phi(\beta, x - x_s(\alpha)). \tag{21}$$

This contrast is thus explicit and will lead to explicit estimators. The method consisting in plugging the observed process in the likelihood of a proxi-model has been used in many situations, for instance in the case of discretization of diffusions (likelihood of the associated Euler scheme (see *e.g.* Kessler, 1997, 2000; Kessler *et al.* 2012, Chapter 1; Genon-Catalot and Larédo, 2014), or in the case of small variance diffusions (see Genon-Catalot,1990; Gloter and Sørensen, 2009; Guy *et al.*, 2014; Larédo, 2020, Part IV in Britton and Pardoux).

For the statistical study, additional assumptions are needed.

- [S0] The parameter sets $\Theta_\alpha, \Theta_\beta$ are compact intervals. The true value of the parameter, denoted by $\theta_0 = (\alpha_0, \beta_0)$, belongs to $\mathring{\Theta}_\alpha \times \mathring{\Theta}_\beta$.

- [S1] There exist two open sets U_α, U_β with $\Theta_\alpha \subset U_\alpha$, $\Theta_\beta \subset U_\beta$ such that all the derivatives $(\alpha,x) \to \dfrac{\partial^{i+j}V}{\partial x^i \partial \alpha^j}(\alpha,x)$, $\quad (\beta,x) \to \dfrac{\partial^{i+j}\Phi}{\partial x^i \partial \beta^j}(\beta,x)$ exist, are continuous respectively on $U_\alpha \times \mathbb{R}$ (resp. $U_\beta \times \mathbb{R}$) and have polynomial growth with respect to x.

Note that it is not necessary to assume the existence of derivatives of any order but this simplifies the exposition.
Let us first investigate the asymptotic behaviour of the first and second derivatives of $\Lambda_{\varepsilon,T}(\theta)$ with respect to the parameters. Using Definitions (20), (21), the following holds.

Proposition 4.1. *Assume that [H0]-[H2], [H4] and [S1] hold. Then, as $\varepsilon \to 0$, under $\mathbb{P}_\theta$,*

$$\begin{aligned}
\varepsilon \frac{\partial \Lambda_{\varepsilon,T}}{\partial \alpha}(\theta) \quad &\to_{\mathcal{L}} \quad = \int_0^T \left[\frac{\partial V}{\partial \alpha}(\alpha, x_s(\alpha)) + \frac{\partial \Phi}{\partial x}(\beta,0)\frac{\partial x_s}{\partial \alpha}(\alpha)\right] dW_s,\\
\varepsilon^2 \frac{\partial^2 \Lambda_{\varepsilon,T}}{\partial \alpha^2}(\theta) \quad &\to_{\mathbb{P}} \quad = -\int_0^T \left[\frac{\partial V}{\partial \alpha}(\alpha, x_s(\alpha)) + \frac{\partial \Phi}{\partial x}(\beta,0)\frac{\partial x_s}{\partial \alpha}(\alpha)\right]^2 ds.
\end{aligned}$$

This result indicates that α should be consistently estimated with rate ε^{-1}.

Proposition 4.2. *Assume [H0]-[H2], [H4] and [S1]. Then, as $\varepsilon \to 0$, under $\mathbb{P}_\theta$,*

$$\begin{aligned}
\frac{\partial \Lambda_{\varepsilon,T}}{\partial \beta}(\theta) \quad &\to_{\mathbb{P}_\theta} \quad -\frac{\partial^2 \Phi}{\partial \beta \partial x}(\beta,0)\int_0^T g_s(\theta))dW_s,\\
\frac{\partial^2 \Lambda_{\varepsilon,T}}{\partial \beta^2}(\theta) \quad &\to_{\mathbb{P}_\theta} \quad -\left(\frac{\partial^2 \Phi}{\partial \beta \partial x}(\beta,0)\right)^2 \int_0^T g_s^2(\theta)ds - \frac{\partial^3 \Phi}{\partial \beta^2 \partial x}(\beta,0)\int_0^T g_s(\theta)dW_s,
\end{aligned}$$

where $g_s(\theta)$ is defined in (13) *or* (15).

Contrary to α, there is no rate of convergence to estimate β in the sense that, as $\varepsilon \to 0$, $-\frac{\partial^2 \Lambda_{\varepsilon,T}}{\partial \beta^2}(\theta)$ converges to a fixed random variable so that the approximate Fisher information associated with β does not tend to infinity. Therefore, we cannot estimate β as $\varepsilon \to 0$ with fixed T.

4.2. **Estimation of α with known β.** Assume that $\beta = \beta_0$ is known and define the estimator

$$\hat{\alpha}_\varepsilon(\beta_0) = \arg\max_{\alpha \in \Theta_\alpha} \Lambda_{\varepsilon,T}(\alpha, \beta_0). \tag{22}$$

Lemma 4.1. *Assume [H0]-[H2], [H4] and [S1]. Then, for all $\alpha, \alpha_0, \beta_0$, under $\mathbb{P}_{\theta_0}$, as $\varepsilon \to 0$,*

$$\varepsilon^2 \left(\Lambda_{\varepsilon,T}(\alpha,\beta_0) - \Lambda_{\varepsilon,T}(\alpha_0,\beta_0)\right) \to K(\alpha,\alpha_0;\beta_0), \quad \textit{where}$$

$$K(\alpha,\alpha_0;\beta_0) = -\frac{1}{2}\int_0^T \left[V(\alpha, x_s(\alpha_0)) - V(\alpha_0, x_s(\alpha_0)) - \Phi(\beta_0, x_s(\alpha_0) - x_s(\alpha))\right]^2 ds.$$

Hence, $K(\alpha,\alpha_0;\beta_0) \le 0$ and, if $\{\forall s \in [0,T], \quad x_s(\alpha) = x_s(\alpha_0) \Rightarrow \alpha = \alpha_0\}$, then,

$$\{K(\alpha,\alpha_0;\beta_0) = 0\} \Rightarrow \{\alpha = \alpha_0\}.$$

Consequently, we consider the following additional identifiability assumption:

- [S2] $\{\forall s \in [0,T], \quad x_s(\alpha) = x_s(\alpha_0)\} \Rightarrow \{\alpha = \alpha_0\}$.

Corollary 4.1. *Assume [H0]-[H2], [H4] and [S0]-[S2]. Then, $\hat{\alpha}_\varepsilon(\beta_0)$ is consistent and under $\mathbb{P}_{\theta_0}$,*

$$\varepsilon^{-1}(\hat{\alpha}_\varepsilon(\beta_0)-\alpha_0) \to_{\mathcal{D}} \mathcal{N}(0, \frac{1}{I_T(\theta_0)}), \quad \textit{with}$$
$$I_T(\theta_0) = \int_0^T \left[\frac{\partial V}{\partial \alpha}(\alpha_0, x_s(\alpha_0)) + \frac{\partial \Phi}{\partial x}(\beta_0, 0)\frac{\partial x_s}{\partial \alpha}(\alpha_0)\right]^2 ds.$$

By standard tools relying the compactness of the parameter set ([S0]), under [S2], we can prove that the estimator $\hat{\alpha}_\varepsilon(\beta_0)$ is consistent. Then, we have that, under $\mathbb{P}_{\theta_0}$, $\varepsilon^{-1}(\hat{\alpha}_\varepsilon(\beta_0)-\alpha_0) \simeq -\varepsilon\frac{\partial \Lambda_{\varepsilon,T}}{\partial \alpha}(\alpha_0,\beta_0)/\varepsilon^2\frac{\partial^2 \Lambda_{\varepsilon,T}}{\partial \alpha^2}(\alpha_0,\beta_0)$. The proof is achieved using Proposition 4.1 and the consistency result.

Example (7)*(continued).* Since (X_t) is equal to its Gaussian approximation, $\Lambda_{\varepsilon,T}(\theta) = \ell_{\varepsilon,T}(\theta)$ is the exact log-likelihood and

$$\ell_{\varepsilon,T}(\theta) = \frac{1}{\varepsilon^2}\left(\int_0^T [-(\alpha+\beta)X_s + \beta x_0 e^{-\alpha s})]dX_s - \frac{1}{2}\int_0^T [-(\alpha+\beta)X_s + \beta x_0 e^{-\alpha s})]^2 ds\right).$$

Since $\beta = \beta_0$ is known, Condition [S2] is: $\forall t \le T, x_0(e^{-\alpha_0 t} - e^{-\alpha t}) = 0$. It implies either that $x_0 = 0$ or that, if $x_0 \neq 0$, $\alpha = \alpha_0$. We deduce that [S2] holds for $x_0 \neq 0$.

Some computations show that, as ε tends to 0, under $\mathbb{P}_\theta$,

$$\varepsilon^2\frac{\partial^2 \ell_{\varepsilon,T}}{\partial \alpha^2}(\theta) \to_{\mathbb{P}} -\int_0^T x_0^2 e^{-2\alpha s}(1+\beta s)^2 ds, \quad \varepsilon\frac{\partial \ell_{\varepsilon,T}}{\partial \alpha}(\theta) \to_{\mathcal{D}} -\int_0^T x_0 e^{-\alpha s}(1+\beta s)dW_s.$$
$$\frac{\partial^2 \ell_{\varepsilon,T}}{\partial \beta^2}(\theta) = -\int_0^T g_\theta^2(s)ds, \quad \frac{\partial \ell_{\varepsilon,T}}{\partial \beta}(\theta) = \int_0^T g_\theta(s)dW_s.$$

We obtain with this exact log-likelihood the result that, if $x_0 \neq 0$, there is a rate of convergence ε^{-1} for α and no rate for β as $\frac{\partial^2 \ell_{\varepsilon,T}}{\partial \beta^2}(\theta)$ and $\frac{\partial \ell_{\varepsilon,T}}{\partial \beta}(\theta)$ are fixed random variables. Moreover,

$$\varepsilon^{-1}(\hat{\alpha}_\varepsilon(\beta_0)-\alpha_0) \to_{\mathcal{D}} \mathcal{N}(0, I_T^{-1}(\alpha_0,\beta_0)) \quad \text{where } I_T(\alpha,\beta) = \int_0^T x_0^2 e^{-2\alpha s}(1+\beta s)^2 ds. \tag{23}$$

4.3. **Estimation of α with β unknown.** Consider now the estimation of α when β in equation (1) is unknown. The approximate likelihood (20) can no longer be used. We propose a conditional least squares contrast:

$$U_{\varepsilon,T}(\alpha) = \frac{1}{\varepsilon^2}\int_0^T V(\alpha, X_s)dX_s - \frac{1}{2\varepsilon^2}\int_0^T V^2(\alpha, X_s)ds. \tag{24}$$

Lemma 4.2. *Assume [H0]-[H2], [H4] and [S1]. Then, under $\mathbb{P}_{(\alpha_0,\beta)}$,*

$$\varepsilon^2\left(U_{\varepsilon,T}(\alpha) - U_{\varepsilon,T}(\alpha_0)\right) \to -\frac{1}{2}\int_0^T (V(\alpha, x_s(\alpha_0)) - V(\alpha_0, x_s(\alpha_0)))^2 ds.$$

Hence the identifiability assumption associated with $U_{\varepsilon,T}(\alpha)$ is

- [S3] For all α, α_0, $\{V(\alpha, x_t(\alpha_0)) = V(\alpha_0, x_t(\alpha_0)) \quad \forall t \in [0,T]\} \Rightarrow \{\alpha = \alpha_0\}$.

Note that [S3] is stronger than [S2] in the sense that if [S3] holds, [S2] is satisfied.
Define now the associated estimator

$$\tilde{\alpha}_\varepsilon = \arg\max_{\alpha\in\Theta_\alpha} U_{\varepsilon,T}(\alpha). \tag{25}$$

The proof that, under [S3], the estimator $\tilde{\alpha}_\varepsilon$ is consistent follows standard tools and is omitted.
For the asymptotic normality of $\tilde{\alpha}_\varepsilon$, we need the following result.

Proposition 4.3. *Assume [H0]-[H2], [H4] and [S1], [S2]. Then, under $\mathbb{P}_\theta$, as $\varepsilon\to 0$, using definition* (13) *for $g_t(\theta)$,*

$$\begin{aligned} \varepsilon\frac{\partial U_{\varepsilon,T}}{\partial\alpha}(\alpha) &\to_{\mathbb{P}} \int_0^T \frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha))dW_s - \frac{\partial\Phi}{\partial x}(\beta,0)\int_0^T \frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha))g_s(\theta)ds, \\ \varepsilon^2\frac{\partial^2 U_{\varepsilon,T}}{\partial\alpha^2}(\alpha) &\to_{\mathbb{P}} -\int_0^T (\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha)))^2 ds. \end{aligned} \tag{26}$$

Remark 4.1. *Using* (14)*, (15) and interchanging integrals yields that the limit (26) can be written as $\int_0^T f(\theta,s)dW_s$ with*

$$f(\theta,s) = \frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha)) - \frac{\partial\Phi}{\partial x}(\beta,0)e^{-A(\theta,s)}\int_s^T \frac{\partial V}{\partial\alpha}(\alpha,x_u(\alpha))e^{A(\theta,u)}du. \tag{27}$$

With standard tools, this leads to the following corollary.

Corollary 4.2. *Assume [H0]-[H2], [H4] and [S0], [S1], [S3]. Then, $\tilde{\alpha}_\varepsilon$ is consistent and under $\mathbb{P}_{(\alpha_0,\beta_0)}$,*

$$\varepsilon^{-1}(\tilde{\alpha}_\varepsilon-\alpha_0)\to_{\mathcal{D}}\mathcal{N}(0,\frac{1}{J_T(\alpha_0,\beta_0)}), \quad \text{with } J_T(\theta) = \frac{\left(\int_0^T(\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha)))^2ds\right)^2}{\int_0^T f^2(\theta,s)ds}. \tag{28}$$

Example (continued). Consider the case of β unknown in Example (7). Then, $f(\theta,s) = -\frac{x_0}{1+\frac{\beta}{2\alpha}}e^{-\alpha s}(1+\frac{\beta}{2\alpha}e^{-(2\alpha+\beta)(T-s)})$, and $\varepsilon^{-1}(\tilde{\alpha}_\varepsilon-\alpha_0)\to_{\mathcal{D}}\mathcal{N}(0,J_T^{-1}(\alpha_0,\beta_0))$ with

$$J_T(\alpha,\beta) = \frac{x_0^2(1+\frac{\beta}{\alpha})^2(\int_0^T e^{-2\alpha t}dt)^2}{\int_0^T e^{-2\alpha s}(1+\frac{\beta}{2\alpha}e^{-(2\alpha+\beta)(T-s)})^2ds}.$$

The comparison of $I_T(\theta)$ defined in (23) and $J_T(\theta)$ is intricate for T fixed, but simpler for large T. We have that $I_T(\theta) = \frac{x_0^2}{2\alpha}(1+\frac{\beta}{\alpha}+\frac{\beta^2}{2\alpha^2})+\frac{1}{T}O(1)$ and $J_T(\theta) = \frac{x_0^2}{2\alpha}(1+\frac{\beta}{2\alpha})^2+\frac{1}{T}O(1)$. Hence, we obtained the expected result : For all θ, $I_T(\theta) > J_T(\theta)$ for large T.

5. Continuous observation of n paths $(X_t^i, 0\le t\le T, i=1,\dots,n)$

Assume now that the observations consist of n paths $(X_t^i, 0\le t\le T)$, which are independent, identically distributed.

$$dX_t^i = V(\alpha,X_t^i)dt - b^i(\theta,t,\varepsilon,X_t^i)dt + \varepsilon dW_t^i, \quad X_0^i = x_0 \tag{29}$$

$$b^i(\theta,t,\varepsilon,x)=\int_{\mathbb{R}}\Phi(\beta,x-y)u_{t,i}^{\theta,\varepsilon,\mu}(dy),$$

where (W_t^i) are independent Wiener processes and $u_{t,i}^{\theta,\varepsilon,\mu}$ is the distribution of X_t^i. Now, these distributions are identical so that we have for $i=1,\ldots,n$, $u_{t,i}^{\theta,\varepsilon,\mu}=u_t^{\theta,\varepsilon,\mu}$ and

$$b^i(\theta,t,\varepsilon,x)=b(\theta,t,\varepsilon,x)=\int_{\mathbb{R}}\Phi(\beta,x-y)u_t^{\theta,\varepsilon,\mu}(dy), \tag{30}$$

5.1. **Convergence rates.** The approximate loglikelihood is now, using $H(.)$ defined in (21),

$$\Lambda_{\varepsilon,T}^n(\theta)=\frac{1}{\varepsilon^2}\sum_{i=1}^n\left(\int_0^T H(\theta,s,X_s^i)dX_s^i-\frac{1}{2}\int_0^T H^2(\theta,s,X_s^i)ds\right). \tag{31}$$

The following holds.

Proposition 5.1. *Assume that [H0]-[H2], [H4], [S1]. Then, as $\varepsilon\to 0$, $n\to\infty$ in such a way that $n\varepsilon^2\to 0$, under $\mathbb{P}_\theta$,*

$$\begin{aligned}
&\frac{\varepsilon}{\sqrt{n}}\frac{\partial\Lambda_{\varepsilon,T}^n}{\partial\alpha}(\theta) \quad\to_{\mathcal{L}}\quad \int_0^T\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))dW_s=\int_0^T\left[\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha))+\frac{\partial\Phi}{\partial x}(\beta,0)\frac{\partial x_s}{\partial\alpha}(\alpha)\right]dW_s,\\
&\frac{\varepsilon^2}{n}\frac{\partial^2\Lambda_{\varepsilon,T}^n}{\partial\alpha^2}(\theta) \quad\to_{\mathbb{P}}\quad -\int_0^T\left(\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))\right)^2ds=-\int_0^T\left[\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha))+\frac{\partial\Phi}{\partial x}(\beta,0)\frac{\partial x_s}{\partial\alpha}(\alpha)\right]^2ds,
\end{aligned}$$

where W is a Brownian motion.

This result indicates that α should be consistently estimated with rate $\frac{\sqrt{n}}{\varepsilon}$.
Consider now $\frac{\partial\Lambda_{\varepsilon,T}^n}{\partial\beta}(\theta)$. Define, using (14) and (15),

$$\sigma_T^2(\theta):=\int_0^T\mathbb{E}_\theta g_t^2(\theta)dt=\int_0^T dt\int_0^t e^{2[A(\theta,t)-A(\theta,s)]}ds. \tag{32}$$

Proposition 5.2. *Assume [H0]-[H2], [H4], [S1].Then, as $\varepsilon\to 0$, $n\to\infty$ in such a way that $n\varepsilon^2\to 0$, under $\mathbb{P}_\theta$, provided that $\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\neq 0$,*

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\frac{\partial\Lambda_{\varepsilon,T}^n}{\partial\beta}(\theta) \quad\to_{\mathcal{L}}\quad -\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)Z_T \quad \textit{in distribution with } Z_T\sim\mathcal{N}(0,\sigma_T^2(\theta)),\\
&\frac{1}{n}\frac{\partial^2\Lambda_{\varepsilon,T}^n}{\partial\beta^2}(\theta) \quad\to_{\mathbb{P}_\theta}\quad -\left(\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\right)^2\sigma_T^2(\theta).
\end{aligned}$$

We may now study the joint estimation of (α,β). Set

$$D_{\varepsilon,T}=\begin{pmatrix}\frac{\varepsilon}{\sqrt{n}} & 0\\ 0 & \frac{1}{\sqrt{n}}\end{pmatrix},\quad \mathcal{J}_{\varepsilon,T}(\theta)=-\begin{pmatrix}\frac{\partial^2\Lambda_{\varepsilon,T}^n}{\partial\alpha^2}(\theta) & \frac{\partial^2\Lambda_{\varepsilon,T}^n}{\partial\beta\partial\alpha}(\theta)\\ \frac{\partial^2\Lambda_{\varepsilon,T}^n}{\partial\beta\partial\alpha}(\theta) & \frac{\partial^2\Lambda_{\varepsilon,T}^n}{\partial\beta^2}(\theta)\end{pmatrix} \tag{33}$$

We have

$$D_{\varepsilon,T}\mathcal{J}_{\varepsilon,T}(\theta)D_{\varepsilon,T}=-\begin{pmatrix}\frac{\varepsilon^2}{n}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha^2}(\theta) & \frac{\varepsilon}{n}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta\partial\alpha}(\theta)\\ \frac{\varepsilon}{n}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta\partial\alpha}(\theta) & \frac{1}{n}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta^2}(\theta)\end{pmatrix}.$$

In view of Proposition 5.2, we can state the following corollary.

Corollary 5.1. *Assume [H0]-[H2], [H4], [S1] and that* $\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\neq 0$. *Then, if* $\varepsilon\to 0$, $n\to+\infty$ *with* $n\varepsilon^2\to 0$, *under* $\mathbb{P}_\theta$,

$$\begin{pmatrix} \frac{\varepsilon}{\sqrt{n}}\frac{\partial\Lambda^n_{\varepsilon,T}}{\partial\alpha}(\theta) \\ \\ \frac{1}{\sqrt{n}}\frac{\partial\Lambda^n_{\varepsilon,T}}{\partial\beta}(\theta)\end{pmatrix}\to_{\mathcal{L}}\mathcal{N}_2(\mathbf{0},\mathcal{J}(\theta))\quad \textit{with}$$

$$\mathcal{J}(\theta)=\begin{pmatrix}\int_0^T[\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha))+\frac{\partial\Phi}{\partial x}(\beta,0)\frac{\partial x_s}{\partial\alpha}(\alpha,s))]^2ds & 0\\ 0 & \left(\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\right)^2\sigma_T^2(\theta)\end{pmatrix}. \tag{34}$$

Moreover, the matrix $D_{\varepsilon,T}\mathcal{J}_{\varepsilon,T}(\theta)D_{\varepsilon,T}=\mathcal{J}(\theta)+o_P(1)$.

As a consequence of Propositions 5.1, 5.2 and Corollary 5.1, we see that α and β should be estimated at different rates of convergence, $\frac{\sqrt{n}}{\varepsilon}$ for α, $\sqrt{n}$ for β.

Example 7 (continued). Consider again the special case $V(\alpha,x)=-\alpha x$, $\Phi(\beta,x)=\beta x$. Then, for the exact likelihood of this model, as $\varepsilon\to 0$, $n\to+\infty$ under $\mathbb{P}_\theta$,

$$\begin{pmatrix} \frac{\varepsilon}{\sqrt{n}}\frac{\partial\ell^n_{\varepsilon,T}}{\partial\alpha}(\theta) \\ \\ \frac{1}{\sqrt{n}}\frac{\partial\ell^n_{\varepsilon,T}}{\partial\beta}(\theta)\end{pmatrix}\to_{\mathcal{L}}\mathcal{N}_2\left(\mathbf{0},\begin{pmatrix}x_0^2\int_0^T(1+s\beta)^2e^{-2\alpha s}ds & 0\\ 0 & \frac{2(\alpha+\beta)T+e^{-2(\alpha+\beta)T}-1}{4(\alpha+\beta)^2}\end{pmatrix}\right).$$

5.2. **Asymptotic properties of estimators.** Consider the approximate likelihood $\Lambda^n_{\varepsilon,T}$ defined in (31),(21) and let $(\hat\alpha_{\varepsilon,n},\hat\beta_{\varepsilon,n})$ the maximum pseudo-likelihood estimator defined as any solution of

$$(\hat\alpha_{\varepsilon,n},\hat\beta_{\varepsilon,n})=\arg\max_{(\alpha,\beta)\in(\Theta_\alpha\times\Theta_\beta)}\Lambda^n_{\varepsilon,T}(\alpha,\beta). \tag{35}$$

Consider the additional assumption:

- [S4] Identifiability assumption for β: $\{\frac{\partial\Phi}{\partial x}(\beta,0)=\frac{\partial\Phi}{\partial x}(\beta_0,0)\Rightarrow\beta=\beta_0\}$.

Define the two functions:

$$\begin{aligned}\Lambda_1(\alpha,\alpha_0,\beta) &= -\frac{1}{2}\int_0^T\left(V(\alpha,x_s(\alpha_0))-V(\alpha_0,x_s(\alpha_0))-\Phi(\beta,x_s(\alpha_0)-x_s(\alpha))\right)^2ds,\\ \Lambda_2(\alpha_0,\beta,\beta_0) &= -\frac{1}{2}\left(\frac{\partial\Phi}{\partial x}(\beta,0)-\frac{\partial\Phi}{\partial x}(\beta_0,0)\right)^2\sigma_T^2(\theta_0).\end{aligned}$$

Lemma 5.1. *Assume [H1], [H2], [H4] and [S1]. Then, the following holds, under* $\mathbb{P}_{\theta_0}$
(i) $\frac{\varepsilon^2}{n}(\Lambda^n_{\varepsilon,T}(\alpha,\beta)-\Lambda_{\varepsilon,T}(\alpha_0,\beta))\to\Lambda_1(\alpha,\alpha_0,\beta)$ *uniformly with respect to* $(\alpha,\beta)\in\Theta_\alpha\times\Theta_\beta$.
(ii) $\frac{1}{n}(\Lambda^n_{\varepsilon,T}(\alpha_0,\beta)-\Lambda^n_{\varepsilon,T}(\alpha_0,\beta_0))\to\Lambda_2(\alpha_0,\beta,\beta_0)$ *uniformly with respect to* $\beta\in\Theta_\beta$.

Theorem 5.1. *Assume [H0]-[H2], [H4] and [S0]-[S2] and that* $\varepsilon\to 0$, $n\to+\infty$ *in such a way that* $n\varepsilon^2\to 0$. *Then, using* (35), *under* $\mathbb{P}_{\theta_0}$,

(i) $(\hat{\alpha}_{\varepsilon,n}, \hat{\beta}_{\varepsilon,n})$ *converges in* $\mathbb{P}_{\theta_0}$*-probability to* (α_0, β_0).
(ii) *Assume moreover that* $\frac{\partial^2 \Phi}{\partial \beta \partial x}(\beta, 0) \neq 0$*, then,*

$$\begin{pmatrix} \frac{\sqrt{n}}{\varepsilon}(\hat{\alpha}_{\varepsilon,n} - \alpha_0) \\ \sqrt{n}(\hat{\beta}_{\varepsilon,n} - \beta_0) \end{pmatrix} \to_{\mathcal{L}} \mathcal{N}_2(0, \mathcal{J}^{-1}(\theta_0)), \quad \textit{where } \mathcal{J}(\theta) \textit{ is defined in } (34).$$

This is an inference problem with different rates of convergence. The proof of **(i)** requires several steps detailed in the appendix. The proof of **(ii)** follows from Lemma 5.1, **(i)** and Corollary 5.1.

6. An asymptotic equivalence of experiments property.

We have derived the parametric inference from (ξ_t) instead of (X_t) and a natural question that arises is whether we have lost information in this approach. We rely on the Le Cam theory to answer this question (see *e.g.* Le Cam,1986, Le Cam and Yang, 2000).

Using the notations introduced in Section 4, let $\mathbb{P}_\theta^{\varepsilon,T}$ denote the distribution of (X_t) satisfying (1) with $X_0 = x_0$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t, t \in [0,T]))$, and $\mathbb{Q}_\theta^{\varepsilon,T}$ the distribution of (ξ_t) satisfying (19) with $\xi_0 = x_0$. Let $\Theta = \Theta_\alpha \times \Theta_\beta$ and consider the two statistical experiments

$$\mathcal{E}^{\varepsilon,T} = (\Omega, \mathcal{F}, (\mathbb{P}_\theta^{\varepsilon,T})_{\theta\in\Theta}) \quad \text{and} \quad \mathcal{G}^{\varepsilon,T} = (\Omega, \mathcal{F}, (\mathbb{Q}_\theta^{\varepsilon,T})_{\theta\in\Theta}).$$

Let $\Delta(\mathcal{E}^{\varepsilon,T}, \mathcal{G}^{\varepsilon,T})$ be their Le Cam deficiency distance. As the experiments are defined on the same probability space, we have an upper bound for Δ:

$\Delta(\mathcal{E}^{\varepsilon,T}, \mathcal{G}^{\varepsilon,T}) \leq \Delta_0(\mathcal{E}^{\varepsilon,T}, \mathcal{G}^{\varepsilon,T}) = sup_{\theta\in\Theta} ||\mathbb{P}_\theta^{\varepsilon,T} - \mathbb{Q}_\theta^{\varepsilon,T}||_{TV}$, with $||.||_{TV}$ total variation distance.

To study $\Delta_0(\mathcal{E}^{\varepsilon,T}, \mathcal{G}^{\varepsilon,T})$, we rely on the Pinsker inequality (see *e.g.* Tsybakov, 2009) for the total variation distance between two probability measures:

$$||\mathbb{P} - \mathbb{Q}||_{TV} \leq \sqrt{K(\mathbb{P}, \mathbb{Q})/2},$$

where $K(\mathbb{P}, \mathbb{Q})$ is the Kullback- Leibler divergence of $\mathbb{P}$ with respect to $\mathbb{Q}$. We can compute this quantity by means of the Girsanov formula. Set $p(\theta, s, X_s) = V(\alpha, X_s) - b(\theta, \varepsilon, s, X_s)$. Then, using definition (21),

$$\frac{d\mathbb{P}_\theta^{\varepsilon,T}}{d\mathbb{Q}_\theta^{\varepsilon,T}} = \exp\left(\frac{1}{\varepsilon^2}\int_0^T (p(\theta, s, X_s) - H(\theta, s, X_s))dX_s - \frac{1}{2\varepsilon^2}\int_0^T (p^2(\theta, s, X_s) - H^2(\theta, s, X_s)ds\right).$$

Under $\mathbb{Q}_\theta^{\varepsilon,T}$, $dX_s - H(\theta, s, X_s) = \varepsilon dW_s$ and $p(\theta, s, X_s) - H(\theta, s, X_s) = D(\theta, s, \varepsilon, X_s) := D_s$ using (16). Hence, we get that,

$$K(\mathbb{P}_\theta^{\varepsilon,T}, \mathbb{Q}_\theta^{\varepsilon,T}) = \frac{1}{2\varepsilon^2}\mathbb{E}_{\mathbb{P}_\theta^{\varepsilon,T}}\int_0^T (p(\theta, s, X_s) - H(\theta, s, X_s))^2 ds = \frac{1}{2\varepsilon^2}\mathbb{E}_{\mathbb{P}_\theta^{\varepsilon,T}}\int_0^T D_s^2 ds.$$

Now, we have

$$\varepsilon^{-2}\mathbb{E}_{\mathbb{P}_\theta^{\varepsilon,T}}\int_0^T D_s^2 ds \leq \frac{2}{\varepsilon^2}\left(\int_0^T \mathbb{E}_{\mathbb{P}_\theta^{\varepsilon,T}}((D_s - \mathbb{E}_{\mathbb{P}_\theta^{\varepsilon,T}} D_s)^2)ds + \int_0^T (\mathbb{E}_{\mathbb{P}_\theta^{\varepsilon,T}} D_s)^2 ds\right).$$

Using Corollary 3.2 yields that under $\mathbb{P}_\theta^{\varepsilon,T}$,

$$K(\mathbb{P}_\theta^{\varepsilon,T}, \mathbb{Q}_\theta^{\varepsilon,T}) \leq \varepsilon^2\, C(\theta).$$

It remains to prove that $\sup_{\theta\in\Theta} C(\theta) < +\infty$. For this, we strengthen some assumptions.

Proposition 6.1. *Assume that:*

- $\exists L > 0,\ \forall \alpha \in \Theta_\alpha,\ \forall x,\ \frac{\partial V}{\partial x}(\alpha, x) \le L(\alpha) \le L.$
- $\exists K > 0,\ k \ge 0\ \forall \alpha \in \Theta_\alpha, \forall \beta \in \Theta_\beta,\ \ \forall x,\ |\frac{\partial^2 V}{\partial x^2}(\alpha, x)| + |\frac{\partial^3 \Phi}{\partial x^3}(\beta, x)| \le K(1 + |x|^k).$
- $\exists B > 0, \forall \beta \in \Theta_\beta,\ \frac{\partial \Phi}{\partial x}(\beta, 0) \le B.$

Then, there exists a constant $C > 0$ such that

$$\sup_{\theta\in\Theta} K(\mathbb{P}_\theta^{\varepsilon,T}, \mathbb{Q}_\theta^{\varepsilon,T}) \le \varepsilon^2\, C.$$

Note that, the parameter set Θ being compact, the above assumptions are not stringent. Hence, the Le Cam deficiency distance $\Delta(\mathcal{E}^{\varepsilon,T}, \mathcal{G}^{\varepsilon,T})$ between the two experiments converges to 0 as $\varepsilon \to 0$.
In the case on n independent trajectories, consider the statistical experiments

$$\mathcal{E}^{\varepsilon,T,n} = (\Omega^n, \mathcal{F}^{\otimes n}, ((\mathbb{P}_\theta^{\varepsilon,T})^{\otimes n})_{\theta\in\Theta}) \quad \text{and} \quad \mathcal{G}^{\varepsilon,T,n} = (\Omega^n, \mathcal{F}^{\otimes n}, ((\mathbb{Q}_\theta^{\varepsilon,T})^{\otimes n})_{\theta\in\Theta}).$$

Then,

$$K((\mathbb{P}_\theta^{\varepsilon,T})^{\otimes n}, (\mathbb{Q}_\theta^{\varepsilon,T})^{\otimes n}) = nK(\mathbb{P}_\theta^{\varepsilon,T}, \mathbb{Q}_\theta^{\varepsilon,T}).$$

Under the condition $n\varepsilon^2 = o(1)$, the Le Cam deficiency distance between the product experiments tends to 0.

7. Concluding remarks

In this paper, we consider the process (X_t) given by (1), *i.e.* a one-dimensional nonlinear self-stabilizing stochastic differential equation, with classical drift term $V(\alpha, x)$ depending on an unknown parameter α, self-stabilizing term $\Phi(\beta, x)$ depending on another unknown parameter β and small noise amplitude ε. Extending and going further the results of Hermann *et al.* (2008), we study the probabilistic properties of (X_t) as ε tends to 0 concerning the convergence of (X_t) to the solution of the ordinary differential equation corresponding to $\varepsilon = 0$ in (1). In particular, we exhibit an approximating Gaussian process for (X_t) and study with accuracy the remainder terms of the approximation. The statistical inference strongly relies on these bounds.
Next, we consider the estimation of (α, β) from a continuous observation of $(X_t, t \in [0, T])$. To our knowledge, it is the first paper tackling the problem of statistical inference of such models. We propose an approximate log-likelihood function which is obtained from the exact log-likelihood function of a proxi-model (ξ_t), solution of a classical time-inhomogeneous diffusion process. We prove that, for fixed T, as ε tends to 0, α *but not* β can be consistently estimated with rate ε^{-1} and study in this framework the asymptotic properties of estimators of α for β known or unknown. This approximate likelihood is justified by the fact that we prove that the statistical experiments generated by (X_t) and (ξ_t) are asymptotically equivalent in the Le Cam Δ-distance sense.
Next, we consider n *i.i.d.* processes distributed as (1) and study the estimation of (α, β) from continuous observations of $(X_t^i, t \in [0, T], i = 1, \ldots, n)$. This kind of observations is consistent with the underlying idea of interacting particles systems and propagation of chaos. From the statistical point of view, we obtain a two-rate model where α is estimated at rate $\sqrt{n}/\varepsilon$ while β is estimated at rate $\sqrt{n}$.

Several natural extensions are relevant in this set-up. We have considered the case of unidimensional parameters α, β. The results should be still valid for multidimensional parameters.
It will be also interesting to study the case of multidimensional non-linear self-stabilizing stochastic differential equations which may also include a non constant diffusion coefficient. Finally, this work should be extended to a discretized observation of the sample path either under high frequency or low frequency setting.

8. Appendix

8.1. **Proofs of Section 3.**

Proof of Proposition 3.1. Let $Z_t = X_t - x_t(\theta)$. We have, by Ito's formula,

$$\begin{aligned} Z_t^2 &= 2\varepsilon \int_0^t Z_s dW_s + 2\int_0^t Z_s[V(\alpha, X_s) - V(\alpha, x_s(\theta)]ds \\ &-2 \int_0^t Z_s \int \Phi(\beta, X_s - y) u_s^{\theta,\varepsilon}(dy)ds + 2\int_0^t Z_s \int \Phi(\beta, x_s(\theta) - y) u_s^{\theta,0}(dy)ds + \varepsilon^2 t, \end{aligned}$$

where $u_s^{\theta,\varepsilon}(dy) = u_s^{\theta,\varepsilon,\mu}(dy)$. Denote by $\mathbb{E}^{x_0}$ the conditional expectation $\mathbb{E}_\theta(.|\eta = x_0)$ and let $\nu_s^{x_0}$ denote the conditional distribution of $X_s - x_s(\theta)$ given $\eta = x_0$. By the assumptions on $\Phi(\beta, .)$ and using that, given $\eta = x_0$, $x_t(\theta)$ is deterministic,

$$2\mathbb{E}^{x_0}(Z_s b(\theta, s, X_s)) = 2\mathbb{E}^{x_0}\left(Z_s b(\theta, s, Z_s + x_s(\theta))\right) = 2\mathbb{E}^{x_0}(Z_s \int \Phi(\beta, X_s - y) dP_{X_s}(y))$$

$$= 2\mathbb{E}^{x_0}(Z_s \int \Phi(\beta, Z_s - z') dP_{Z_s}(z'))$$

$$= 2\int\int z\Phi(\beta, z - z')\nu_s^{x_0}(dz)\nu_s^{x_0}(dz') = \int\int (z - z')\Phi(\beta, z - z')\nu_s^{x_0}(dz)\nu_s^{x_0}(dz') \geq 0.$$

For the last identity, we use the assumption that $x \to \Phi(\beta, x)$ is odd, which yields

$$\begin{aligned} \int\int -z'\Phi(\beta, z - z')\nu_s^{x_0}(dz)\nu_s^{x_0}(dz') &= \int\int z'\Phi(\beta, z' - z)\nu_s^{x_0}(dz)\nu_s^{x_0}(dz') \\ &= \int\int z\Phi(\beta, z - z')\nu_s^{x_0}(dz)\nu_s^{x_0}(dz'). \end{aligned}$$

Next consider the fourth term of Z_t^2, noting that given $X_0 = x_0$, $u_s^{\theta,0}(dy) = \delta_{x_s(\theta)}(dy)$,

$$\mathbb{E}^{x_0}\left(Z_s \int \Phi(\beta, x_s(\theta) - y) u_s^{\theta,0}(dy)\right) = \mathbb{E}^{x_0}(Z_s \Phi(\beta, 0)) = 0.$$

It follows that, using (3),

$$\mathbb{E}_\theta(Z_t^2) \leq 2\int_0^t \mathbb{E}_\theta(Z_s[V(\alpha, X_s) - V(\alpha, x_s(\theta))])ds + \varepsilon^2 t \leq 2L(\alpha)\int_0^t \mathbb{E}_\theta Z_s^2 ds + \varepsilon^2 t.$$

By the Gronwall lemma, we conclude that $\mathbb{E}_\theta(Z_t^2) \leq \varepsilon^2 t e^{2L(\alpha)t}$.

We can proceed analogously to study $\mathbb{E}_\theta(Z_t^{2p})$. By induction, assume that, for $k = 2, \ldots, p-1$, $\mathbb{E}_\theta Z_t^{2k} \leq C_{2k}(t)\varepsilon^{2k}\exp(2kL(\alpha)t)$, where $C_{2k}(t)$ is a non-decreasing function

and $C_2(t) = t$. Then,

$$\begin{aligned} Z_t^{2p} &= 2p\int_0^t Z_s^{2p-1}\{\varepsilon dW_s + [V(\alpha, X_s) - V(\alpha, x_s(\alpha)]ds\} \\ &- 2p\int_0^t Z_s^{2p-1} b(\theta, s, \varepsilon, X_s)ds + 2p\int_0^t Z_s^{2p-1} b(\theta, s, 0, x_s(\theta))ds + \varepsilon^2 p(2p-1)\int_0^t Z_s^{2p-2}ds. \end{aligned}$$

Analogously, by the same conditional device, we can write:

$$\begin{aligned} 2p\mathbb{E}^{x_0}\int_0^t Z_s^{2p-1} b(\theta, s, \varepsilon, X_s)ds &= 2p\int_0^t\int\int z^{2p-1}\Phi(\beta, z-z')\nu_s^{x_0}(dz)\nu_s^{x_0}(dz')ds \\ &= p\int_0^t\int\int (z^{2p-1} - (z')^{2p-1})\Phi(\beta, z-z')\nu_s^{x_0}(dz)\nu_s^{x_0}(dz')ds \geq 0. \end{aligned}$$

And,

$$\mathbb{E}^{x_0}\left(Z_s^{2p-1}\int \Phi(\beta, x_s(\theta) - y)u_s^{\theta,0}(dy)\right) = \mathbb{E}^{x_0} Z_s^{2p-1}\Phi(\beta, 0) = 0.$$

This yields,

$$\mathbb{E}_\theta(Z_t^{2p}) \leq 2pL(\alpha)\int_0^t \mathbb{E}_\theta(Z_s^{2p})ds + \varepsilon^2 p(2p-1)\int_0^t \mathbb{E}_\theta Z_s^{2p-2}ds.$$

We apply the Gronwall lemma, and after sone elementary computations get:

$$\mathbb{E}_\theta(Z_t^{2p}) \leq \varepsilon^{2p}\exp\left(2pL(\alpha)t\right)C_{2p}(t) \quad \text{with} \quad C_{2p}(t) = p(2p-1)C_{2p-2}(t).$$

□

Proof of Corollary 3.1. We have,

$$\mathbb{E}_\theta(X_t - x_t(\alpha)) = \int_0^t \mathbb{E}_\theta\left(V(\alpha, X_s) - V(\alpha, x_s(\alpha))\right)ds - \int_0^t \mathbb{E}_\theta b(\theta, s, \varepsilon, X_s)ds.$$

Let $(\overline{X}_s)$ be an independent copy of (X_s). Then,

$$\mathbb{E}_\theta b(\theta, s, \varepsilon, X_s) = \mathbb{E}_\theta\int \Phi(\beta, X_s - y)u_s^{\theta,\varepsilon}(dy) = \mathbb{E}_\theta(\Phi(\beta, X_s - \overline{X}_s)) = 0, \tag{36}$$

since $\Phi(\beta, .)$ is odd and since the distribution of $X_s - \overline{X}_s$ is symmetric.
Now, a Taylor expansion at $x_s(\alpha)$ yields

$$\mathbb{E}_\theta(V(\alpha, X_s) - V(\alpha, x_s(\alpha)) = \mathbb{E}_\theta(X_s - x_s(\alpha))\frac{\partial V}{\partial x}(\alpha, x_s(\alpha)) + R_s, \text{ with}$$

$$R_s = \int_0^1 (1-u)\mathbb{E}_\theta\left((X_s - x_s(\alpha))^2\frac{\partial^2 V}{\partial x^2}(\alpha, x_s(\alpha) + u(X_s - x_s(\alpha)))\right)du.$$

Therefore,

$$\mathbb{E}_\theta X_t - x_t(\alpha) = \int_0^t \left(\mathbb{E}_\theta X_s - x_s(\alpha)\right)\frac{\partial V}{\partial x}(\alpha, x_s(\alpha))ds + \int_0^t R_s ds.$$

Now, $f(t) = \mathbb{E}_\theta X_t - x_t(\alpha)$ satisfies $f'(t) = \frac{\partial V}{\partial x}(\alpha, x_t(\alpha))f(t) + R_t,\ f(0) = 0$. Consequently,

$$f(t) = \int_0^t R_s \exp\left(\int_s^t \frac{\partial V}{\partial x}(\alpha, x_u(\alpha))du\right)ds. \tag{37}$$

Using Proposition 3.1 with $p=1$ and $C_2(s)=s$, we get that, if $|\frac{\partial^2 V}{\partial x^2}(\alpha,x)|$ is bounded by $L_2(\alpha)$, $|R_s| \leq L_2(\alpha)\varepsilon^2 s\exp(2L(\alpha)s)$. Using the explicit expression of $f(t)$, we conclude,

$$|\mathbb{E}_\theta X_t - x_t(\alpha)| \leq \varepsilon^2 L_2(\alpha) t^2 e^{2L(\alpha)t}.$$

If $|\frac{\partial^2 V}{\partial x^2}(\alpha,x)| \leq k(\alpha)(1+|x|^{\gamma(\alpha)})$ (Assumption [H3-2]), then

$$|R_s)| \lesssim \mathbb{E}_\theta\left((X_s - x_s(\alpha))^2(1+|x_s(\alpha)|^{\gamma(\alpha)} + |X_s - x_s(\alpha)|^{\gamma(\alpha)})\right) \lesssim \varepsilon^2 K(s), \tag{38}$$

where $K(s)$ is uniformly bounded on $[0,T]$, $\lesssim$ means $\leq$ up to a constant, so that (11) follows.

If $V(\alpha,x) = -\alpha x$, then $\mathbb{E}_\theta X_t = x_0 - \alpha\int_0^t \mathbb{E}_\theta X_s ds$, thus $\mathbb{E}_\theta X_t = x_0\exp(-\alpha t) = x_t(\alpha)$. □

Proof of Lemma 3.1. Only the behaviour at $\varepsilon = 0$ is to be studied. Applying the Taylor formula and integrating w.r.t. $u_t^{\theta,\varepsilon}(dy)$, $D(\theta,t,\varepsilon,x)$ defined in (16) writes:

$$D(\theta,t,\varepsilon,x) = \int (\Phi(\beta,x-y) - \Phi(\beta,x-x_t(\alpha)))\, u_t^{\theta,\varepsilon}(dy) = \mathbb{E}_\theta(x_t(\alpha) - X_t)\frac{\partial\Phi}{\partial x}(\beta,x-x_t(\alpha)) + R_1(t,x),$$

$$\text{with } R_1(t,x) = \int (x_t(\alpha)-y)^2\left(\int_0^1 (1-u)\frac{\partial^2\Phi}{\partial x^2}(\beta,x-x_t(\alpha)+u(x_t(\alpha)-y))du\right) u_t^{\theta,\varepsilon}(dy).$$

Under Assumption [H3-2], $|\frac{\partial^2\Phi}{\partial x^2}(\beta,x)| \leq k(\beta)(1+|x|^{\gamma(\beta)})$. Therefore by Proposition 3.1,

$$|R_1(t,x)| \lesssim \mathbb{E}_\theta\left((x_t(\alpha)-X_t)^2(1+|x-x_t(\alpha)|^{\gamma(\beta)}) + |X_t - x_t(\alpha)|^{\gamma(\beta)}\right) \lesssim \varepsilon^2.$$

By Corollary 3.1, $\mathbb{E}_\theta(x_t(\alpha)-X_t)\frac{\partial\Phi}{\partial x}(\beta,x-x_t(\alpha)) = O(\varepsilon^2)$ and $D(\theta,t,\varepsilon,x)\to 0$ as $\varepsilon\to 0$. Thus, extending $b(\theta,t,\varepsilon,x)$ by continuity at $\varepsilon=0$ yields $b(\theta,t,0,x) = \Phi(\beta,x-x_t(\alpha))$. Moreover, $b(\theta,t,\varepsilon,x)$ is differentiable w.r.t. ε at $\varepsilon=0$ with $\frac{\partial b}{\partial\varepsilon}(\theta,t,\varepsilon,x)|_{\varepsilon=0} = 0$.
For the differentiability of $b(\theta,t,\epsilon,x)$ with respect to x at $(\theta,t,0,x)$, we proceed as above.

$$\frac{\partial b}{\partial x}(\theta,t,\varepsilon,x) = \int \frac{\partial\Phi}{\partial x}(\beta,x-y)u_t^{\theta,\varepsilon}(dy) = \frac{\partial\Phi}{\partial x}(\beta,x-x_t(\alpha)) + R_2(t,x) \quad \text{with}$$

$R_2(t,x) = \int(\frac{\partial\Phi}{\partial x}(\beta,x-y) - \frac{\partial\Phi}{\partial x}(\beta,x-x_t(\alpha)))u_t^{\theta,\varepsilon}(dy)$. Therefore, under Assumption [H3-2], $R_2(t,x) = \int(x_t(\alpha)-y)\left(\int_0^1(1-u)\frac{\partial^2\Phi}{\partial x^2}(\beta,x-x_t(\alpha)+u(x_t(\alpha)-y))du\right)u_t^{\theta,\varepsilon}(dy) \to 0$ as ε tends to 0 by Corollary 3.1. This yields that $\frac{\partial b}{\partial x}(\theta,t,\varepsilon,x) = \frac{\partial\Phi}{\partial x}(\beta,x-x_t(\alpha))$.

Note that, for $\varepsilon=0, x=x_t(\alpha)$, $\frac{\partial b}{\partial x}(\theta,t,0,x_t(\alpha)) = \frac{\partial\Phi}{\partial x}(\beta,0)$. □

Proof of Theorem 3.1 . By (12), we have $R^\varepsilon(t) = \varepsilon^{-2}(X_t - x_t(\alpha) - \varepsilon g_t(\theta))$. Set

$$B(\theta,t,\varepsilon,x) = V(\alpha,x) - b(\theta,t,\varepsilon,x). \tag{39}$$

Lemma 3.1 yields that $B(\theta,t,0,x_t(\alpha)) = V(\alpha,x_t(\alpha))$, and using $a(\theta,t)$ defined in (15),

$$\frac{\partial B}{\partial x}(\theta,t,0,x_t(\alpha)) = \frac{\partial V}{\partial x}(\alpha,x_t(\alpha)) - \frac{\partial\Phi}{\partial x}(\beta,0) = a(\theta,t). \tag{40}$$

Therefore using (13),

$$\begin{aligned} dR^\varepsilon(t) &= \frac{1}{\varepsilon^2}\left(B(\theta,t,\varepsilon,X_t)dt - B(t,0,x_t(\alpha))dt + \varepsilon dW_t - \varepsilon(\frac{\partial B}{\partial x}(\theta,t,0,x_t(\alpha))g_t(\theta)dt + dW_t)\right) \\ &= \frac{1}{\varepsilon^2}\left[(X_t - x_t(\alpha) - \varepsilon g_t(\theta))\,\frac{\partial B}{\partial x}(\theta,t,0,x_t(\alpha))\right]dt + \nu(\theta,t,\varepsilon,X_t) \\ &= a(\theta,t)R_t^\varepsilon dt + \nu(\theta,t,\varepsilon,X_t)dt, \quad R^\varepsilon(0)=0, \text{ where} \end{aligned}$$

$$\nu(\theta,t,\varepsilon,X_t) = \frac{1}{\varepsilon^2}\left(B(t,\varepsilon,X_t) - B(t,0,x_t(\alpha)) - (X_t - x_t(\alpha))\frac{\partial B}{\partial x}(t,0,x_t(\alpha))\right) = T_1(t)+T_2(t).$$

with

$$T_1(t) = \varepsilon^{-2}\left(V(\alpha,X_t) - V(\alpha,x_t(\alpha)) - (X_t - x_t(\alpha))\frac{\partial V}{\partial x}(\alpha,x_t(\alpha))\right) \tag{41}$$

$$T_2(t) = -\varepsilon^{-2}\left(\int \Phi(\beta,X_t - y)u_t^\theta(dy) - \frac{\partial \Phi}{\partial x}(\beta,0)(X_t - x_t(\alpha))\right). \tag{42}$$

The equation satisfied by $R^\varepsilon(t)$ can be solved and we get, using (40) and (15),

$$R^\varepsilon(t) = \int_0^t \nu(\theta,s,\varepsilon,X_s)\exp\Big(\int_s^t a(\theta,u)du\Big)ds. \tag{43}$$

Inequality (3) yields that $a(\theta,t) \le L(\alpha) - \frac{\partial\Phi}{\partial x}(\beta,0) \le L(\alpha)$. Therefore bounds for $R^\varepsilon(t)$ derive from the bounds satisfied by $\nu(\theta,t,\varepsilon,X_t)$.
Let us first study $T_1(t)$. A Taylor expansion at point $x_t(\alpha)$ yields, using Assumption [H3-2],

$$\begin{aligned} T_1(t) &= \varepsilon^{-2}(X_t - x_t(\alpha))^2 \int_0^1 \frac{\partial^2 V}{\partial x^2}(\alpha, x_t(\alpha) + u(X_t - x_t(\alpha)))du \\ |T_1(t)| &\le k(\alpha)\varepsilon^{-2}(X_t - x_t(\alpha))^2(1 + |x_t(\alpha)|^{\gamma(\alpha)} + |X_t - x_t(\alpha)|^{\gamma(\alpha)}). \end{aligned} \tag{44}$$

Hence, by Proposition 3.1, for all $p \ge 1$, $\mathbb{E}_\theta|T_1(t)|^{2p} = O(1)$ uniformly on $t \in [0,T], \varepsilon > 0$.
Let us study $T_2(t)$ defined in (42) and consider first its expectation. Equation (36) yields

$$\mathbb{E}_\theta T_2(t) = \varepsilon^{-2}\frac{\partial\Phi}{\partial x}(\beta,0)(\mathbb{E}_\theta X_t - x_t(\alpha)). \tag{45}$$

Applying Corollary 3.1 yields that $|\mathbb{E}_\theta T_2(t)| = O(1)$.
Coming back to $T_2(t)$, a Taylor expansion at point 0 yields, noting that $\frac{\partial^2\Phi}{\partial x^2}(\beta,0) = 0$,

$$\Phi(\beta,X_t - y) - \frac{\partial\Phi}{\partial x}(\beta,0)(X_t - x_t(\alpha)) = \frac{\partial\Phi}{\partial x}(\beta,0)(x_t(\alpha) - y) + \rho_1(X_t,y),$$

where

$$\rho_1(X_t,y) = \frac{1}{2}(X_t - y)^3\int_0^1 (1-u)^2\frac{\partial^3\Phi}{\partial x^3}(\beta,u(X_t - y))du. \tag{46}$$

Therefore we can write, for $\overline{X}_t$ an independent copy of X_t,

$$\mathbb{E}_\theta\int \rho_1(X_t,y)u_t^{\theta,\varepsilon}(dy) = \frac{1}{2}\mathbb{E}_\theta\left((X_t - \overline{X}_t)^3\int_0^1 (1-u)^2\frac{\partial^3\Phi}{\partial x^3}(\beta,u(X_t - \overline{X}_t))du\right).$$

Under [H3-3], $x \to x^3 \frac{\partial^3 \Phi}{\partial x^3}(\beta, ux)$ is well defined and odd so that

$$\mathbb{E}_\theta \int \rho_1(X_t, y) u_t^{\theta,\varepsilon}(dy) = \mathbb{E}_\theta(\rho_1(X_t, \overline{X}_t)) = 0. \tag{47}$$

This implies that $T_2(t) - \mathbb{E}_\theta T_2(t) = \varepsilon^{-2} \int \rho_1(X_t, y)) u_t^{\theta,\varepsilon}(dy)$. Therefore,

$$\mathbb{E}_\theta(T_2(t) - \mathbb{E}_\theta T_2(t))^{2p} = \varepsilon^{-4p} \mathbb{E}_\theta(\rho_1(X_t, \bar{X}_t)^{2p}).$$

Now, by [H3-3] and (46)

$$\begin{aligned} \mathbb{E}_\theta(\rho_1(X_t, \bar{X}_t)^{2p}) &= 2^{-2p} \mathbb{E}_\theta \left((X_t - \overline{X}_t)^{6p} (\int_0^1 (1-u)^2 \frac{\partial^3 \Phi}{\partial x^3}(\beta, u(X_t - \overline{X}_t)) du)^{2p} \right) \\ &\leq C \mathbb{E}_\theta \left(|X_t - \overline{X}_t|^{6p} (1 + |X_t - \overline{X}_t|^{c}))^{2p} \right). \end{aligned}$$

By splitting $X_t - \overline{X}_t$ into $X_t - x_t(\alpha) + x_t(\alpha)) - \overline{X}_t$ we get that $\mathbb{E}_\theta(\rho_1(X_t, \bar{X}_t)^{2p}) \leq C \mathbb{E}_\theta \left((X_t - x_t(\alpha))^{6p} \right)$. Applying Proposition 3.1 yields that, uniformly on $[0, T]$, $\varepsilon > 0$,

$$\mathbb{E}_\theta \left(T_2(t) - \mathbb{E}_\theta T_2(t) \right)^{2p} \leq \varepsilon^{6p} O(1).$$

Finally,(43) yields that, using the bounds obtained for $\nu(\theta, s, \varepsilon, X_s) = T_1(s) + T_2(s)$,

$$\begin{aligned} |\mathbb{E}_\theta R^\varepsilon(t)| &\leq \int_0^t |\mathbb{E}_\theta \nu(\theta, s, \varepsilon, X_s)| e^{L(\alpha)(t-s)} ds = O(1) \text{ and} \\ \mathbb{E}_\theta(R^\varepsilon(t) - \mathbb{E}_\theta R^\varepsilon(t))^{2p} &\lesssim \int_0^t \mathbb{E}_\theta \left((\nu(\theta, s, \varepsilon, X_s) - \mathbb{E}_\theta \nu(\theta, s, \varepsilon, X_s))^{2p} \right) e^{2pL(\alpha)(t-s)} ds = O(1). \end{aligned}$$

□

Proof of Corollary 3.2. Using (16), $D(\theta, t, \varepsilon, X_t) = \int (\Phi(\beta, X_t - y) - \Phi(\beta, X_t - x_t(\alpha))) u_t^{\theta,\varepsilon}(dy)$. Similarly to the study of $T_2(t)$, a Taylor expansion of $\Phi(\beta, .)$ yields, using (46),

$$\Phi(\beta, X_t - y) - \Phi(\beta, X_t - x_t(\alpha)) = \frac{\partial \Phi}{\partial x}(\beta, 0)(x_t(\alpha) - y) + \rho_1(X_t, y) - \rho_2(X_t), \quad \text{with}$$

$$\rho_2(X_t) = \frac{1}{2}(X_t - x_t(\alpha))^3 \int_0^1 (1-u)^2 \frac{\partial^3 \Phi}{\partial x^3}(\beta, u(X_t - x_t(\alpha))) du. \tag{48}$$

Therefore, $D(\theta, t, \varepsilon, X_t) = \frac{\partial \Phi}{\partial x}(\beta, 0)(x_t(\alpha) - \mathbb{E}_\theta X_t) + \int \rho_1(X_t, y) u_t^{\theta,\varepsilon}(dy) - \rho_2(X_t)$. Using (47),

$$\mathbb{E}_\theta D(\theta, t, \varepsilon, X_t) = \frac{\partial \Phi}{\partial x}(\beta, 0)(x_t(\alpha) - \mathbb{E}_\theta X_t) - \mathbb{E}_\theta \rho_2(X_t).$$

By Proposition 3.1, $\mathbb{E}_\theta |\rho_2(X_t)| \lesssim \varepsilon^3 O(1)$. Applying Corollary 3.1 yields (17).
Moreover, as for the upper bound of $T_2(t)$, $\mathbb{E}_\theta |\rho_1(X_t, \bar{X}_t)|^{2p} \lesssim \mathbb{E}_\theta |X_t - \overline{X}_t|^{6p} \lesssim \varepsilon^{6p}$.
By Proposition 3.1, uniformly on $[0, T]$, $\mathbb{E}_\theta |\rho_2(X_t)|^{2p} \lesssim \mathbb{E}_\theta |X_t - x_t(\alpha)|^{6p} \leq \varepsilon^{6p} O(1)$.
Joining these two inequalities, we get (18). □

8.2. **Proofs of Section 4.** Recall that $H(\theta,s,x)=V(\alpha,x)-\Phi(\beta,x-x_s(\alpha))$. Thus,

$$\begin{aligned}
\frac{\partial H}{\partial\alpha}(\theta,s,X_s) &= \frac{\partial V}{\partial\alpha}(\alpha,X_s)+\frac{\partial\Phi}{\partial x}(\beta,X_s-x_s(\alpha))\frac{\partial x_s}{\partial\alpha}(\alpha,s),\\
\frac{\partial H}{\partial\beta}(\theta,s,X_s) &= -\frac{\partial\Phi}{\partial\beta}(\beta,X_s-x_s(\alpha))\\
\frac{\partial^2 H}{\partial\alpha^2}(\theta,s,X_s) &= \frac{\partial^2 V}{\partial\alpha^2}(\alpha,X_s)+\frac{\partial\Phi}{\partial x}(\beta,X_s-x_s(\alpha))\frac{\partial^2 x_s}{\partial\alpha^2}(\alpha,s)-\frac{\partial^2\Phi}{\partial x^2}(\beta,X_s-x_s(\alpha))\left(\frac{\partial x_s}{\partial\alpha}(\alpha,s)\right)^2\\
\frac{\partial^2 H}{\partial\beta^2}(\theta,s,X_s) &= -\frac{\partial^2\Phi}{\partial\beta^2}(\beta,X_s-x_s(\alpha)),\ \frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,X_s)=\frac{\partial^2\Phi}{\partial x\partial\beta}(\beta,X_s-x_s(\alpha))\frac{\partial x_s}{\partial\alpha}(\alpha,s).
\end{aligned}$$

We rely on two lemmas: Lemma 8.1 is for dealing with $\frac{\partial\Lambda_{\varepsilon,T}}{\partial\alpha}(\theta)$, $\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha^2}(\theta)$ (Proposition 4.1), Lemma 8.2 is for $\frac{\partial\Lambda_{\varepsilon,T}}{\partial\beta}(\theta)$, $\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta^2}(\theta)$ (Proposition 4.2). These Lemmas are also be useful for Section 5.
In proofs, we set for simplicity where there is no ambiguity,

$$D_t=D(\theta,t,\varepsilon,X_t)=b(\theta,t,\varepsilon,X_t)-\Phi(\beta,X_t-x_t(\alpha)).$$

Lemma 8.1. *Assume [H0]-[H2] an [S1]. For* $F(\theta,s,X_s)=\frac{\partial V}{\partial\alpha}(\alpha,X_s)$, $\frac{\partial^2 V}{\partial\alpha^2}(\alpha,X_s)$, $\frac{\partial\Phi}{\partial x}(\beta,X_s-x_s(\alpha))$, $\frac{\partial^2\Phi}{\partial x^2}(\beta,X_s-x_s(\alpha))$, *we have:*

$$\sup_{s\le T}\mathbb{E}_\theta(F(\theta,s,X_s)-F(\theta,s,x_s(\alpha)))^{2p}\le\varepsilon^{2p}C_F(\theta,T), \tag{49}$$

$$\sup_{s\le T}\mathbb{E}_\theta\left|F(\theta,s,X_s)D(\theta,s,\varepsilon,X_s)\right|\le\varepsilon^2C'_F(\theta,T), \tag{50}$$

where $C(\theta,T),C'(\theta,T)$ *do not depend on* ε.

Proof of Lemma 8.1. Consider first $F=\frac{\partial V}{\partial\alpha}$. By the Taylor formula,

$$\frac{\partial V}{\partial\alpha}(\alpha,X_s)-\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha))=(X_s-x_s(\alpha))\int_0^1\frac{\partial^2 V}{\partial\alpha\partial x}(\alpha,x_s(\alpha)+u(X_s-x_s(\alpha))du.$$

Using that $\frac{\partial^2 V}{\partial\alpha\partial x}(\alpha,.)$ has polynomial growth, we get that , for some constant c,

$$\left|\frac{\partial V}{\partial\alpha}(\alpha,X_s)-\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha))\right|\lesssim|X_s-x_s(\alpha))|(1+|x_s(\alpha)|^c+|X_s-x_s(\alpha)|^c).$$

By Proposition 3.1,

$$\sup_{s\le T}\mathbb{E}_\theta(\frac{\partial V}{\partial\alpha}(\alpha,X_s)-\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha))^{2p}\le\varepsilon^{2p}C(\alpha,T).$$

We proceed analogously for $\frac{\partial^2 V}{\partial\alpha^2}(\alpha,X_s)$. For $\frac{\partial\Phi}{\partial x}(\beta,X_s-x_s(\alpha))$, we write

$$\frac{\partial\Phi}{\partial x}(\beta,X_s-x_s(\alpha))-\frac{\partial\Phi}{\partial x}(\beta,0)=(X_s-x_s(\alpha))\int_0^1\frac{\partial^2\phi}{\partial x^2}(\alpha,u(X_s-x_s(\alpha))du.$$

Therefore, we conclude analogouly that (49) holds for $\frac{\partial\Phi}{\partial x}(\beta,X_s-x_s(\alpha))$ and $\frac{\partial^2\Phi}{\partial x^2}(\beta,X_s-x_s(\alpha))$.

To prove (50), we write, $F(\theta,s,X_s)D_s=\sum_{k=1}^4 A_k(s)$, where

$$\begin{aligned}
A_1(s) &= F(\theta,s,x_s(\alpha))(D_s-\mathbb{E}_\theta D_s), \quad A_2(s)=F(\theta,s,x_s(\alpha))\mathbb{E}_\theta D_s,\\
A_3(s) &= (F(\theta,s,X_s)-F(\theta,s,x_s(\alpha)))(D_s-\mathbb{E}_\theta D_s),\\
A_4(s) &= (F(\theta,s,X_s)-F(\theta,s,x_s(\alpha)))\mathbb{E}_\theta D_s.
\end{aligned}$$

We now use Corollary 3.2, to obtain:

$$\begin{aligned}
\sup_{s\leq T}\mathbb{E}_\theta|A_1(s)| &\leq \sup_{s\leq T}|F(\theta,s,x_s(\alpha))|\sup_{s\leq T}(\mathbb{E}_\theta(D_s-\mathbb{E}_\theta D_s)^2)^{1/2}\leq \varepsilon^3 O(1),\\
\sup_{s\leq T}|A_2(s)| &\leq \varepsilon^2 O(1).
\end{aligned}$$

Now,

$$\mathbb{E}_\theta|A_3(s)|\leq\left[\mathbb{E}_\theta(F(\theta,s,X_s)-F(\theta,s,x_s(\alpha)))^2\mathbb{E}_\theta(D_s-\mathbb{E}_\theta D_s)^2\right]^{1/2}.$$

Therefore, by the first part of the Lemma and Corollary 3.2,

$$\sup_{s\leq T}\mathbb{E}_\theta|A_3(s)| \leq \varepsilon\times\varepsilon^3 O(1), \quad \sup_{s\leq T}\mathbb{E}_\theta|A_4(s)|\leq\varepsilon\times\varepsilon^2 O(1).$$

Hence,

$$\varepsilon^{-1}\sup_{s\leq T}\mathbb{E}_\theta|F(\theta,s,X_s)D_s|\leq\varepsilon O(1).$$

□

Proof of Proposition 4.1. Considering the regularity assumptions on V,Φ, we can derive under the stochastic and the ordinary integral. Replacing dX_s by its expression in (1), we get (see (20) and (21)):

$$\begin{aligned}
\varepsilon\frac{\partial\Lambda_{\varepsilon,T}}{\partial\alpha}(\theta) &= \frac{1}{\varepsilon}\left(\int_0^T\frac{\partial H}{\partial\alpha}(\theta,s,X_s)dX_s-\int_0^T H(\theta,s,X_s)\frac{\partial H}{\partial\alpha}(\theta,s,X_s)ds\right),\\
&= \int_0^T\frac{\partial H}{\partial\alpha}(\theta,s,X_s)dW_s-\frac{1}{\varepsilon}\left(\int_0^T\frac{\partial H}{\partial\alpha}(\theta,s,X_s)D_s ds\right). \qquad (51)
\end{aligned}$$

By Lemma 8.1,

$$\int_0^T[\frac{\partial H}{\partial\alpha}(\theta,s,X_s)-\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))]^2 ds\to_{\mathbb{P}_\theta}0,$$

and

$$\frac{1}{\varepsilon}\left(\int_0^T\frac{\partial H}{\partial\alpha}(\theta,s,X_s)D_s ds\right)\to_{\mathbb{P}_\theta}0. \qquad (52)$$

Hence, the first limit of Proposition 4.1. Now,

$$\begin{aligned}
\varepsilon^2\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha^2}(\theta) &= \int_0^T\frac{\partial^2 H}{\partial\alpha^2}(\theta,s,X_s)dX_s-\int_0^T H(\theta,s,X_s)\frac{\partial^2 H}{\partial\alpha^2}(\theta,s,X_s)ds-\int_0^T(\frac{\partial H}{\partial\alpha}(\theta,s,X_s))^2 ds,\\
(53)= &\ \varepsilon\int_0^T\frac{\partial^2 H}{\partial\alpha^2}(\theta,s,X_s)dW_s-\int_0^T\frac{\partial^2 H}{\partial\alpha^2}(\theta,s,X_s)D_s ds-\int_0^T\left(\frac{\partial H}{\partial\alpha}(\theta,s,X_s)\right)^2 ds.
\end{aligned}$$

We have, as ε tends to 0,

$$\begin{aligned}\int_0^T \frac{\partial H^2}{\partial\alpha^2}(\theta,s,X_s)dW_s &\to_{\mathbb{P}_\theta} \int_0^T \frac{\partial H^2}{\partial\alpha^2}(\theta,s,x_s(\alpha))dW_s,\\ \int_0^T \left(\frac{\partial H}{\partial\alpha}(\theta,s,X_s)\right)^2 ds &\to_{\mathbb{P}_\theta} \int_0^T \left(\frac{\partial H}{\partial\alpha}(\theta,s,x_s)(\alpha)\right)^2 ds,\end{aligned}$$

and $\int_0^T \frac{\partial^2 H}{\partial\alpha^2}(\theta,s,X_s)D_s ds = o_P(1)$. Thus, the second limit of Proposition 4.1. □

Lemma 8.2. *Assume [H0]-[H2] an [S1]. For* $F(\theta,s,X_s) = \frac{\partial\Phi}{\partial\beta}(\beta, X_s - x_s(\alpha))$, $\frac{\partial^2\Phi}{\partial\beta^2}(\beta, X_s - x_s(\alpha))$, *we have:*

$$\sup_{s\le T} \mathbb{E}_\theta \left|F(\theta,s,X_s)D(\theta,s,\varepsilon,X_s)\right| \le \varepsilon^3 C''(\theta,T). \tag{54}$$

where $C''(\theta,T)$ *does not depend on* ε.

Proof of Lemma 8.2. Noting that $x \to \frac{\partial\Phi}{\partial\beta}(\beta,x)$ is an odd function yields:

$$\frac{\partial\Phi}{\partial\beta}(\beta,x) = x\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0) + x^2\int_0^1 (1-u)\frac{\partial^3\Phi}{\partial\beta\partial x^2}(\beta,ux)du. \tag{55}$$

Thus, we split into four terms as above,

$$\frac{\partial\Phi}{\partial\beta}(\beta, X_s - x_s(\alpha))D_s = \sum_{i=1}^4 B_i(s)$$

where

$$\begin{aligned}B_1(s) &= \frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)(X_s - x_s(\alpha))(D_s - \mathbb{E}_\theta D_s),\quad B_2(s) = \frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)(X_s - x_s(\alpha))\mathbb{E}_\theta D_s,\\ B_3(s) &= (X_s - x_s(\alpha))^2\int_0^1 (1-u)\frac{\partial^3\Phi}{\partial\beta\partial x^2}(\beta, u(X_s - x_s(\alpha))du\ (D_s - \mathbb{E}_\theta D_s),\\ B_4(s) &= (X_s - x_s(\alpha))^2\int_0^1 (1-u)\frac{\partial^3\Phi}{\partial\beta\partial x^2}(\beta, u(X_s - x_s(\alpha))du\ \mathbb{E}_\theta D_s.\end{aligned}$$

We have

$$\mathbb{E}_\theta|B_1(s)| \le |\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)|\left[\mathbb{E}_\theta(D_s - \mathbb{E}_\theta D_s)^2\mathbb{E}_\theta(X_s - x_s(\alpha))^2\right]^{1/2}.$$

By Proposition 3.1 and Corollary 3.2, we get

$$sup_{s\le T}\mathbb{E}_\theta|B_1(s)| \lesssim \varepsilon\times\varepsilon^3 O(1) \quad \text{and} \quad sup_{s\le T}\mathbb{E}_\theta|B_2(s)| \lesssim \varepsilon\times\varepsilon^2 O(1).$$

Using that $x \to \frac{\partial^3\Phi}{\partial\beta\partial x^2}(\beta,x)$ has polynomial growth, we obtain, proceeding as for the first two terms,

$$sup_{s\le T}\mathbb{E}_\theta|B_3(s)| \lesssim \varepsilon^2\times\varepsilon^3 O(1) \quad \text{and} \quad sup_{s\le T}\mathbb{E}_\theta|B_4(s)| \lesssim \varepsilon^2\times\varepsilon^2 O(1).$$

Finally, the term B_2 is the main one and we get $\varepsilon^{-2}sup_{s\le T}\mathbb{E}_\theta|\frac{\partial\Phi}{\partial\beta}(\beta, X_s - x_s(\alpha))D_s \lesssim \varepsilon O(1)$.

We proceed in the same way for $\frac{\partial^2\Phi}{\partial\beta^2}(\beta, X_s - x_s(\alpha))$ as $\frac{\partial^2\Phi}{\partial\beta^2}(\beta,.)$ is odd. □

Proof of Proposition 4.2 Analogously, we can write:

$$\begin{aligned}\frac{\partial \Lambda_{\varepsilon,T}}{\partial \beta}(\theta) &= \frac{1}{\varepsilon^2}\int_0^T \frac{\partial H}{\partial \beta}(\theta,s,X_s)dX_s - \frac{1}{2\varepsilon^2}\int_0^T H(\theta,s,X_s)\frac{\partial H}{\partial \beta}(\theta,s,X_s)ds \\ &= -\frac{1}{\varepsilon}\int_0^T \frac{\partial \Phi}{\partial \beta}(\beta, X_s - x_s(\alpha))dW_s + \frac{1}{\varepsilon^2}\int_0^T \frac{\partial \Phi}{\partial \beta}(\beta, X_s - x_s(\alpha))D_s ds. \qquad (56)\end{aligned}$$

By Lemma 8.2, the second term above tends to 0. Substituting x by $X_s - x_s(\alpha)) = \varepsilon g_\theta(s) + \varepsilon^2 R^\varepsilon(s)$ in (55) yields using Theorem 3.1 and Lemma 8.1

$$\begin{aligned}\frac{1}{\varepsilon}\int_0^T \frac{\partial \Phi}{\partial \beta}(\beta, X_s - x_s(\alpha))dW_s &= \frac{\partial^2 \Phi}{\partial \beta \partial x}(\beta,0)\left(\int_0^T (g_\theta(s) + \varepsilon R^\varepsilon(s))dW_s\right) \\ &+ \varepsilon \int_0^T \frac{(X_s - x_s(\alpha))^2}{\varepsilon^2}\int_0^1 (1-u)\frac{\partial^3 \Phi}{\partial \beta \partial x^2}(\beta, u(X_s - x_s(\alpha)))du\, dW_s \\ &= \frac{\partial^2 \Phi}{\partial \beta \partial x}(\beta,0)\left(\int_0^T g_\theta(s)dW_s\right) + O_P(\varepsilon).\end{aligned}$$

So we get the first limit of Proposition 4.2. Now

$$\begin{aligned}\frac{\partial^2 \Lambda_{\varepsilon,T}}{\partial \beta^2}(\theta) &= \frac{1}{\varepsilon^2}\int_0^T \frac{\partial^2 H}{\partial \beta^2}(\theta,s,X_s)dX_s - \frac{1}{\varepsilon^2}\int_0^T H(\theta,s,X_s)\frac{\partial^2 H}{\partial \beta^2}(\theta,s,X_s)ds \\ &\quad -\frac{1}{\varepsilon^2}\int_0^T \left(\frac{\partial H}{\partial \beta}(\theta,s,X_s)\right)^2 ds \\ &= -\frac{1}{\varepsilon}\int_0^T \frac{\partial^2 \Phi}{\partial \beta^2}(\beta, X_s - x_s(\alpha))dW_s + \frac{1}{\varepsilon^2}\int_0^T \frac{\partial^2 \Phi}{\partial \beta^2}(\beta, X_s - x_s(\alpha))D_s ds \\ &\quad -\frac{1}{\varepsilon^2}\int_0^T \left(\frac{\partial \Phi}{\partial \beta}(\beta, X_s - x_s(\alpha))\right)^2 ds. \qquad (57)\end{aligned}$$

By Lemma 8.2, we have $\frac{1}{\varepsilon^2}\int_0^T \frac{\partial^2 \Phi}{\partial \beta^2}(\beta, X_s - x_s(\alpha))D_s ds \to_{\mathbb{P}_\theta} 0$. Using that

$$\frac{\partial^2 \Phi}{\partial \beta^2}(\beta,x) = x\frac{\partial^3 \Phi}{\partial \beta^2 \partial x}(\beta,0) + x^2\int_0^1 du(1-u)\frac{\partial^4 \Phi}{\partial \beta^2 \partial x^2}(\beta, u(X_s - x_s(\alpha)), \qquad (58)$$

we can prove by tools previously used that

$$\begin{aligned}-\frac{1}{\varepsilon}\int_0^T \frac{\partial^2 \Phi}{\partial \beta^2}(\beta, X_s - x_s(\alpha))dW_s &\to_{\mathbb{P}_\theta} -\frac{\partial^3 \Phi}{\partial \beta^2 \partial x}(\beta,0)\int_0^T g_s(\theta)dW_s, \\ -\frac{1}{\varepsilon^2}\int_0^T \left(\frac{\partial \Phi}{\partial \beta}(\beta, X_s - x_s(\alpha))\right)^2 ds &\to_{\mathbb{P}_\theta} -\left(\frac{\partial^2 \Phi}{\partial \beta \partial x}(\beta,0)\right)^2\int_0^T g_s^2(\theta)ds.\end{aligned}$$

Thus, the second limit of Proposition 4.2 is obtained.

□

Proof of Lemma 4.1. We have, under $\mathbb{P}_{\theta_0}$, using (21), (16), $\varepsilon^2\Lambda_{\varepsilon,T}(\alpha,\beta_0) = A_1(\alpha,\beta_0) + A_2(\alpha,\beta_0) + A_3(\alpha,\beta_0)$ with

$$\begin{aligned} A_1(\alpha,\beta_0) &= \int_0^T H(\alpha,\beta_0,s,X_s)H(\alpha_0,s,\beta_0,X_s)ds - \frac{1}{2}\int_0^T H^2(\alpha,\beta_0,s,X_s)ds, \\ A_2(\alpha,\beta_0) &= -\int_0^T H(\alpha,\beta_0,s,X_s)D(\alpha_0,\beta_0,s,\varepsilon,X_s)ds, \quad A_3(\alpha,\beta_0) = \varepsilon\int_0^T H(\alpha,\beta_0,s,X_s)dW_s. \end{aligned}$$

By Theorem 3.1, $\int_0^T H(\alpha,\beta_0,s,X_s)dW_s \to_{\mathbb{P}_{\theta_0}} \int_0^T H(\alpha,\beta_0,s,x_s(\alpha_0))dW_s$, which is a bounded random variable. Hence $A_3(\alpha,\beta_0)$ converges in probability to 0.
Using now Proposition 3.1, $D(\alpha_0,\beta_0,s,\varepsilon,X_s) \to D(\alpha_0,\beta_0,s,0,x_s(\alpha_0)) = 0$ uniformly on $[0,T]$ and $\int_0^T H^2(\alpha,\beta_0,s,X_s)ds$ converges to $\int_0^T H^2(\alpha,\beta_0,s,x_s(\theta_0))ds$. Hence,

$$A_2^2(\alpha,\beta_0) \le \int_0^T H^2(\alpha,\beta_0,s,X_s)ds \int_0^T D^2(\alpha_0,\beta_0,s,\varepsilon,X_s)ds \to 0.$$

Finally, $A_1(\alpha,\beta_0) = -\frac{1}{2}\int_0^T (H(\alpha,\beta_0,s,X_s) - H(\alpha_0,\beta_0,s,X_s))^2 ds + \frac{1}{2}\int_0^T H(\alpha_0,\beta_0,s,X_s)^2 ds$. By Theorem 3.1,

$$A_1(\alpha,\beta_0) - A_1(\alpha_0,\beta_0) \to -\frac{1}{2}\int_0^T (H(\alpha,\beta_0,s,x_s(\alpha_0)) - H(\alpha_0,\beta_0,s,x_s(\alpha_0))^2 ds = K(\alpha,\alpha_0;\beta_0).$$

Under [S1], $(\alpha,\beta) \to K(\alpha,\alpha_0;\beta)$ is continuous non positive. Assume now that $K(\alpha,\alpha_0;\beta_0) = 0$. Then, for all $\beta_0 \in \overline{\Theta}_\beta$, $V(\alpha,x_s(\alpha_0)) - \Phi(\beta_0, x_s(\alpha_0) - x_s(\alpha)) - V(\alpha_0,x_s(\alpha_0)) = 0$. Then,

$$\forall \beta_0 \in \Theta_\beta, \quad V(\alpha,x_s(\alpha_0)) - V(\alpha_0,x_s(\alpha_0)) = \Phi(\beta_0, x_s(\alpha_0) - x_s(\alpha)).$$

Therefore, for all $s \in [0,T]$, the function $\beta_0 \to \Phi(\beta_0, x_s(\alpha_0) - x_s(\alpha))$ is constant. As a function of x, Φ is increasing. So, this is possible only if $x_s(\alpha_0) - x_s(\alpha)$ is constant. Using now that $x_0(\alpha) - x_0(\alpha_0) = 0$ yields the result and Assumption [S2] for the identifiability. □

Proof of Lemma 4.2. We have, under $\mathbb{P}_{\theta_0}$, $\varepsilon^2 U_{\varepsilon,T}(\alpha) = B_1(\alpha,\alpha_0) + B_2(\alpha,\theta_0) + B_3(\alpha)$ with

$$\begin{aligned} B_1(\alpha,\alpha_0) &= \int_0^T V(\alpha,X_s)V(\alpha_0,X_s)ds - \frac{1}{2}\int_0^T V^2(\alpha,X_s)ds, \\ B_2(\alpha,\theta_0) &= -\int_0^T V(\alpha,X_s)b(\theta_0,s,\varepsilon,X_s)ds, \quad B_3(\alpha) = \varepsilon\int_0^T V(\alpha,X_s)dW_s. \end{aligned}$$

As above $B_3(\alpha)$ converges in probability to 0 and $B_1(\alpha,\alpha_0) - B_1(\alpha_0,\alpha_0) \to -\frac{1}{2}\int_0^T (V(\alpha,x_s(\alpha_0)) - V(\alpha_0,x_s(\alpha_0)))^2 ds$.

For $B_2(\alpha,\theta_0)$, we use that, for all s, $b(\theta,s,\varepsilon,x) \to \Phi(\beta, x - x_s(\alpha))$ as $\varepsilon \to 0$. Hence $b(\theta_0,s,\varepsilon,X_s) \to 0$ and $B_2(\alpha,\theta_0) = o_P(1)$. □

Proof of Proposition 4.3. Let us first study $\varepsilon\frac{\partial U_{\varepsilon,T}}{\partial\alpha}(\alpha)$. We have

$$\begin{aligned}\varepsilon\frac{\partial U_{\varepsilon,T}}{\partial\alpha}(\alpha) &= \frac{1}{\varepsilon}\int_0^T\frac{\partial V}{\partial\alpha}(\alpha,X_s)dX_s-\frac{1}{\varepsilon}\int_0^T\frac{\partial V}{\partial\alpha}(\alpha,X_s)V(\alpha,X_s)ds\\ &= \int_0^T\frac{\partial V}{\partial\alpha}(\alpha,X_s)dW_s-\frac{1}{\varepsilon}\int_0^T\frac{\partial V}{\partial\alpha}(\alpha,X_s)b(\theta,s,\varepsilon,X_s)ds=A_1(T)+A_2(T).\end{aligned}$$

First

$$A_1(T)=\int_0^T\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha))dW_s+\int_0^T(\frac{\partial V}{\partial\alpha}(\alpha,X_s)dW_s-\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha)))dW_s=A_{11}(T)+A_{12}(T).$$

By Theorem 3.1, $<A_{12}>_T\to 0$. Next, we have, using (16),

$$A_2(T)=-\frac{1}{\epsilon}\int_0^T\frac{\partial V}{\partial\alpha}(\alpha,X_s)[D(\theta,\varepsilon,s,X_s)+\Phi(\beta,X_s-x_s(\alpha))]ds=A_{21}(T)+A_{22}(T).$$

By Lemma 8.1, $A_{21}(T)=\varepsilon O(1)$.
Let us study

$$\begin{aligned}A_{22}(T) &= -\frac{1}{\varepsilon}\int_0^T\frac{\partial V}{\partial\alpha}(\alpha,X_s)\Phi(\beta,X_s-x_s(\alpha))ds=C_2(T)+C_2'(T),\ \text{ with}\\ C_2(T) &= -\frac{1}{\varepsilon}\int_0^T(\frac{\partial V}{\partial\alpha}(\alpha,X_s)-\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha)))\Phi(\beta,X_s-x_s(\alpha))ds,\\ C_2'(T) &= -\frac{1}{\varepsilon}\int_0^T\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha))\Phi(\beta,X_s-x_s(\alpha))ds.\end{aligned}$$

As above, $C_2(T)$ converges to 0. For $C_2'(T)$, we use the Taylor expansion of $\Phi(\beta,x)$ at 0, *i.e.* $\Phi(\beta,x)=x\frac{\partial\Phi}{\partial x}(\beta,0)+x^3\int_0^1\frac{1}{2}(1-u)^2\frac{\partial^3\Phi}{\partial x^3}(\beta,ux)$. Substituting $X_t-x_t(\alpha)$ by its expression in Theorem 3.1, and using the notations defined in (12), in $C_2'(T)$, we get,

$$C_2'(T)=-\int_0^T\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha))g_s(\theta)\frac{\partial\Phi}{\partial x}(\beta,0)ds+\tilde{R}_\varepsilon(T)$$

with

$$\begin{aligned}\tilde{R}^\varepsilon(T) &= -\varepsilon\int_0^T\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha))R^\varepsilon(s)\frac{\partial\Phi}{\partial x}(\beta,0)ds\\ &\quad -\frac{1}{2\varepsilon}\int_0^T\frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha))(X_s-x_s(\alpha))^3\int_0^1\frac{1}{2}(1-u)^2\frac{\partial^3\Phi}{\partial x^3}(\beta,u(X_s-x_s(\alpha)))duds.\end{aligned}$$

By Theorem 3.1, $sup_{t\in[0,T]}\mathbb{E}_\theta|R^\varepsilon(t)|=O(1)$ and by Proposition 3.1, $E_\theta(X_t-x_t(\alpha))^{2p}=\varepsilon^{2p}O(1)$.
Joining these results achieves the proof of the first part of the proposition.

Consider now $\varepsilon^2\frac{\partial^2U_{\varepsilon,T}}{\partial\alpha^2}(\alpha)=C_1+C_2+C_3$ with

$$C_1=\varepsilon\int_0^T\frac{\partial^2V}{\partial\alpha^2}(\alpha,X_s)dW_s,\ C_2=\int_0^T\frac{\partial^2V}{\partial\alpha^2}(\alpha,X_s)b(\theta,\epsilon,s,X_s)ds,\ C_3=-\int_0^T(\frac{\partial V}{\partial\alpha}(\alpha,X_s))^2ds.$$

First $\mathbb{E}_\theta(C_1^2) = \varepsilon^2 \int_0^T (\frac{\partial^2 V}{\partial \alpha^2}(\alpha, X_s))ds \to 0$. Then, using (16), $C_2 = C_{21} + C_{22}$ with

$$C_{21} = \int_0^T \frac{\partial^2 V}{\partial \alpha^2}(\alpha, X_s) D(\theta, s, X_s), \;\; C_{22} = \int_0^T \frac{\partial^2 V}{\partial \alpha^2}(\alpha, X_s))\Phi(\beta, X_s - x_s(\alpha))ds.$$

Using Lemma 8.1, we obtain that C_{21} tends to 0. For C_{22}, we write

$$(C_{22})^2 \leq \int_0^T (\frac{\partial^2 V}{\partial \alpha^2}(\alpha, X_s))^2 ds \int_0^T \Phi(\beta, X_s - x_s(\alpha))^2 ds = \varepsilon^2 O_P(1).$$

Finally, an application of Lemma 8.1 yields that $C_3 \to -\int_0^T (\frac{\partial V}{\partial \alpha}(\alpha, x_s(\alpha)))^2 ds$.
This achieves the proof of the proposition. □

8.3. **Proofs of Section 5.** Now, we deal with the observations of n *i.i.d.* sample paths.

Proof of Proposition 5.1. We have (see (51))

$$\begin{aligned}
\frac{\varepsilon}{\sqrt{n}} \frac{\partial \Lambda^n_{\varepsilon,T}}{\partial \alpha}(\theta) &= \frac{1}{\varepsilon^2} \sum_{i=1}^n \left(\int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, X_s^i) dX_s^i - \int_0^T H(\theta, s, X_s^i) \frac{\partial H}{\partial \alpha}(\theta, s, X_s^i) ds \right), \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, X_s^i) dW_s^i - \frac{1}{\varepsilon\sqrt{n}} \int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, X_s^i) D_s^i ds,
\end{aligned}$$

where $D_s^i = D(\theta, s, \varepsilon, X_s^i)$. By Lemma 8.1, (50),

$$\frac{1}{\varepsilon\sqrt{n}} \left| \mathbb{E}_\theta \sum_{i=1}^n \int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, X_s^i) D_s^i ds \right| \leq \frac{\sqrt{n}}{\varepsilon} \sup_{s \leq T} \mathbb{E}_\theta |\frac{\partial H}{\partial \alpha}(\theta, s, X_s) D_s| \times T \lesssim \varepsilon \sqrt{n} O(1),$$

which tends to 0 by the condition $\varepsilon\sqrt{n} = o(1)$. For the first term,

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, X_s^i) dW_s^i &= \int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha) \frac{1}{\sqrt{n}} \sum_{i=1}^n dW_s^i \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^T (\frac{\partial H}{\partial \alpha}(\theta, s, X_s^i) - \frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha)) dW_s^i.
\end{aligned}$$

Now, $\int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha) \frac{1}{\sqrt{n}} \sum_{i=1}^n dW_s^i$ is distributed as $\int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha) dW_s$ and

$$\mathbb{E}_\theta \frac{1}{n} \sum_{i=1}^n \int_0^T (\frac{\partial H}{\partial \alpha}(\theta, s, X_s^i) - \frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha))^2 ds = \mathbb{E}_\theta \int_0^T (\frac{\partial H}{\partial \alpha}(\theta, s, X_s) - \frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha))^2 ds$$

tends to 0 by Lemma 8.1, (49). Thus, setting $B_s^{(n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n dW_s^i$,

$$\frac{\varepsilon}{\sqrt{n}} \frac{\partial \Lambda^n_{\varepsilon,T}}{\partial \alpha}(\theta) = \int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha) dB_s^{(n)} + o_P(1). \tag{59}$$

The first limit of Proposition 5.1 is obtained. Now, using (53),

$$\begin{aligned}\frac{\varepsilon^2}{n}\frac{\partial^2\Lambda^n_{\varepsilon,T}}{\partial\alpha^2}(\theta) &= \frac{\varepsilon}{\sqrt{n}}\times\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^T\frac{\partial^2 H}{\partial\alpha^2}(\theta,s,X^i_s)dW^i_s-\frac{1}{n}\sum_{i=1}^n\int_0^T\frac{\partial^2 H}{\partial\alpha^2}(\theta,s,X^i_s)D^i_s ds\\ &\quad -\frac{1}{n}\sum_{i=1}^n\int_0^T\left(\frac{\partial H}{\partial\alpha}(\theta,s,X^i_s)\right)^2 ds.\end{aligned}$$

We can prove in the same way as above that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^T\frac{\partial^2 H}{\partial\alpha^2}(\theta,s,X^i_s)dW^i_s\to_{\mathcal{L}}\int_0^T\frac{\partial^2 H}{\partial\alpha^2}(\theta,s,x_s(\alpha))dW_s,$$

and that

$$-\frac{1}{n}\sum_{i=1}^n\int_0^T\left(\frac{\partial H}{\partial\alpha}(\theta,s,X^i_s)\right)^2 ds\to-\int_0^T(\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha)))^2 ds.$$

Using again Lemma 8.1, we get that $\frac{1}{n}\sum_{i=1}^n\int_0^T\frac{\partial^2 H}{\partial\alpha^2}(\theta,s,X^i_s)D^i_s ds$ tends to 0. Hence, the second limit of Proposition 5.1. □

Proof of Proposition 5.2. Using (56), we get

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial\Lambda^n_{\varepsilon,T}}{\partial\beta}(\theta) &= \frac{1}{\sqrt{n}\varepsilon^2}\sum_{i=1}^n\left(\int_0^T\frac{\partial H}{\partial\beta}(\theta,s,X^i_s)dX^i_s-\int_0^T H(\theta,s,X_s)\frac{\partial H}{\partial\beta}(\theta,s,X^i_s)ds\right)\\ &= -\frac{1}{\sqrt{n}\varepsilon}\sum_{i=1}^n\int_0^T\frac{\partial\Phi}{\partial\beta}(\beta,X^i_s-x_s(\alpha))dW^i_s+\frac{1}{\sqrt{n}\varepsilon^2}\sum_{i=1}^n\int_0^T\frac{\partial\Phi}{\partial\beta}(\beta,X^i_s-x_s(\alpha))D^i_s ds.\end{aligned}$$

First note that, by Lemma 8.2,

$$\frac{1}{\sqrt{n}\varepsilon^2}\mathbb{E}_\theta\left|\sum_{i=1}^n\int_0^T\frac{\partial\Phi}{\partial\beta}(\beta,X^i_s-x_s(\alpha))D^i_s ds\right|\le\frac{\sqrt{n}}{\varepsilon^2}\mathbb{E}_\theta\int_0^T|\frac{\partial\Phi}{\partial\beta}(\beta,X_s-x_s(\alpha))D_s|ds\le\varepsilon\sqrt{n}O(1).$$

Therefore, we look at the first term and use again the Taylor expansion (55) and the development $X^i_s=x_s(\alpha)+\varepsilon g^i_s(\theta)+\varepsilon^2R^{\varepsilon,i}_s$:

$$\frac{1}{\sqrt{n}\varepsilon}\sum_{i=1}^n\int_0^T\frac{\partial\Phi}{\partial\beta}(\beta,X^i_s-x_s(\alpha))dW^i_s=\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^T g^i_s(\theta)dW^i_s+R^{(1)}_{n,\varepsilon}+R^{(2)}_{n,\varepsilon}\tag{60}$$

where $(g^i_s(\theta))$ is the Ornstein-Uhlenbeck process of the development of X^i,

$$\begin{aligned}R^{(1)}_{n,\varepsilon} &= \frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\frac{\varepsilon}{\sqrt{n}}\sum_{i=1}^n\int_0^T R^{\varepsilon,i}_s dW^i_s\\ R^{(2)}_{n,\varepsilon} &= \frac{\varepsilon}{\sqrt{n}}\sum_{i=1}^n\int_0^T\varepsilon^{-2}(X^i_s-x_s(\alpha))^2\int_0^1(1-u)\frac{\partial^3\Phi}{\partial\beta\partial x^2}(\beta,u(X^i_s-x_s(\alpha)dudW^i_s.\end{aligned}$$

Applying Lemma 8.1 and Lemma 8.2, we prove that the terms above are $o_P(1)$. Finally, we have obtained:

$$\frac{1}{\sqrt{n}}\frac{\partial\Lambda^n_{\varepsilon,T}}{\partial\beta}(\theta)=\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^T g^i_s(\theta)dW^i_s+o_P(1).\tag{61}$$

The martingale $M_T^n = \sum_{i=1}^n \int_0^T g_s^i(\theta) dW_s^i$ has quadratic variation $\langle M^n \rangle_T = \sum_{i=1}^n \int_0^T (g_s^i(\theta))^2 ds$ satisfying

$$\frac{1}{n}\langle M^n \rangle_T \to \mathbb{E}_\theta \int_0^T (g_s(\theta))^2 ds = \sigma_T^2(\theta).$$

By the central limit theorem for martingales, we deduce

$$\frac{1}{\sqrt{n}} M_T^n \to_{\mathcal{L}} \mathcal{N}(0, \sigma_T^2(\theta)).$$

This achieves the first part of Proposition 5.2.
Now, we study (see (57))

$$\begin{aligned}\frac{1}{n}\frac{\partial^2 \Lambda_{\varepsilon,T}^n}{\partial \beta^2}(\theta) \quad = \quad & -\frac{1}{n\varepsilon}\sum_{i=1}^n \int_0^T \frac{\partial^2 \Phi}{\partial \beta^2}(\beta, X_s^i - x_s(\alpha)) dW_s^i + \frac{1}{n\varepsilon^2}\sum_{i=1}^n \int_0^T \frac{\partial^2 \Phi}{\partial \beta^2}(\beta, X_s^i - x_s(\alpha)) D_s^i ds \\ & -\frac{1}{n\varepsilon^2}\sum_{i=1}^n \int_0^T \left(\frac{\partial \Phi}{\partial \beta}(\beta, X_s^i - x_s(\alpha))\right)^2 ds.\end{aligned}$$

For the first term, we write:

$$\frac{1}{n\varepsilon}\sum_{i=1}^n \int_0^T \frac{\partial^2 \Phi}{\partial \beta^2}(\beta, X_s^i - x_s(\alpha)) dW_s^i = \frac{1}{\sqrt{n}}\frac{1}{\varepsilon\sqrt{n}}\sum_{i=1}^n \int_0^T \frac{\partial^2 \Phi}{\partial \beta^2}(\beta, X_s^i - x_s(\alpha)) dW_s^i.$$

This term is analogous to (60) with $\frac{\partial^2 \Phi}{\partial \beta^2}$ instead of $\frac{\partial \Phi}{\partial \beta}$ but it is now multiplied by $\frac{1}{\sqrt{n}}$ and therefore is $o_P(1)$.
The second term, $\frac{1}{n\varepsilon^2}\sum_{i=1}^n \int_0^T \frac{\partial^2 \Phi}{\partial \beta^2}(\beta, X_s^i - x_s(\alpha)) D_s^i ds$, is proved to be $o_P(1)$ using Lemma 8.2.
Using as above the Taylor expansion the Taylor expansion (55) and the development $X_s^i = x_s(\alpha) + \varepsilon g_s^i(\theta) + \varepsilon^2 R_s^{\varepsilon,i}$, we prove that

$$-\frac{1}{n\varepsilon^2}\sum_{i=1}^n \int_0^T \left(\frac{\partial \Phi}{\partial \beta}(\beta, X_s^i - x_s(\alpha))\right)^2 ds = -\frac{1}{n}\left(\frac{\partial^2 \Phi}{\partial \beta \partial x}(\beta, 0)\right)^2 \sum_{i=1}^n \int_0^T (g_s^i(\theta))^2 ds + o_P(1).$$

which tends to $-(\frac{\partial^2 \Phi}{\partial \beta \partial x}(\beta, 0))^2 \sigma_T^2(\theta)$. Thus we get the second part of Proposition 5.2 whose proof is now complete. □

Proof of Corollary 5.1. ¿From (59) and (61), to get the joint convergence in distribution, it remains to study

$$< \int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha)) \frac{1}{\sqrt{n}}\sum_{i=1}^n dW_s^i, \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^T g_s^i(\theta) dW_s^i >= \frac{1}{n}\sum_{i=1}^n \int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha)) g_s^i(\theta) ds.$$

This term converges to $\mathbb{E}_\theta \int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha)) g_s(\theta) ds = \int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha)) \mathbb{E}_\theta g_s(\theta) ds = 0$.
Hence, the joint convergence in distribution is proved.

Now, for the second part of the corollary, the only remaining term to study is:

$$\begin{aligned}\frac{\varepsilon}{n}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha\partial\beta}(\theta) &= \frac{1}{\varepsilon n}\sum_{i=1}^n\left(\int_0^T\frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,X_s^i)dX_s^i-\int_0^T H(\theta,s,X_s^i)\frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,X_s^i)ds\right)\\ &\quad-\frac{1}{\varepsilon n}\sum_{i=1}^n\left(\int_0^T\frac{\partial H}{\partial\beta}(\theta,s,X_s^i)\frac{\partial H}{\partial\alpha}(\theta,s,X_s^i)ds\right)\\ &= \frac{1}{n}\sum_{i=1}^n\int_0^T\frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,X_s^i)dW_s^i-\frac{1}{\varepsilon n}\sum_{i=1}^n\int_0^T\frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,X_s^i)D(\theta,s,\varepsilon,X_s^i)ds\\ &\quad-\frac{1}{\varepsilon n}\sum_{i=1}^n\left(\int_0^T\frac{\partial H}{\partial\beta}(\theta,s,X_s^i)\frac{\partial H}{\partial\alpha}(\theta,s,X_s^i)ds\right)\end{aligned}$$

where

$$\frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,X_s) = \frac{\partial^2\Phi}{\partial x\partial\beta}(\beta,X_s-x_s(\alpha))\frac{\partial x_s}{\partial\alpha}(\alpha,s),\ \frac{\partial H}{\partial\beta}(\theta,s,X_s)=-\frac{\partial\Phi}{\partial\beta}(\theta,X_s-x_s(\alpha)).$$

We can write:

$$\frac{1}{n}\sum_{i=1}^n\int_0^T\frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,X_s^i)dW_s^i=\frac{1}{\sqrt{n}}\int_0^T\frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,x_s(\alpha))\frac{1}{\sqrt{n}}\sum_{i=1}^n dW_s^i+o_P(1)=o_P(1).$$

We have:

$$\frac{\partial^2\Phi}{\partial x\partial\beta}(\beta,X_s-x_s(\alpha))=\frac{\partial^2\Phi}{\partial x\partial\beta}(\beta,0)+(X_s-x_s(\alpha))\int_0^1\frac{\partial^3\Phi}{\partial x^2\partial\beta}(\beta,u(X_s-x_s(\alpha)))du$$

Therefore, the second term satisfies by Lemma 8.1 :

$$\frac{1}{\varepsilon n}\sum_{i=1}^n\left(\int_0^T\frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,X_s^i)D_s^i ds\right)=\varepsilon O_P(1),$$

There remains the term for which we only give the main terms:

$$\begin{aligned}\frac{1}{\varepsilon n}\sum_{i=1}^n\left(\int_0^T\frac{\partial H}{\partial\beta}(\theta,s,X_s^i)\frac{\partial H}{\partial\alpha}(\theta,s,X_s^i)ds\right) &= \frac{1}{\varepsilon n}\sum_{i=1}^n\left(\int_0^T(X_s^i-x_s(\alpha))\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\frac{\partial H}{\partial\alpha}(\theta,s,X_s^i)ds\right)\\ &+ o_p(1)\\ &= \frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\frac{1}{n}\sum_{i=1}^n\int_0^T g_s^i(\theta)\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))ds+o_P(1)\\ &= \frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\mathbb{E}_\theta\left(\int_0^T g_s(\theta)\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))ds\right)+o_P(1)\\ &= 0+o_P(1).\end{aligned}$$

The proof of Corollary 5.1 is now complete. □

Proof of Lemma 5.1. We have:

$$\frac{\varepsilon^2}{n}\Lambda_{\varepsilon,T}^n(\alpha,\beta)=\frac{1}{n}\sum_{i=1}^n\left(\int_0^T H(\alpha,\beta,s,X_s^i)dX_s^i-\frac{1}{2}H^2(\alpha,\beta,s,X_s^i)ds\right).$$

We replace

$$dX_s^i = \varepsilon dW_s^i + H(\alpha_0,\beta_0,s,X_s^i)ds - D(\alpha_0,\beta_0,s,X_s^i)ds \tag{62}$$

and this yields: $\frac{\varepsilon^2}{n}\Lambda_{\varepsilon,T}^n(\theta) = T_1 + T_2 + T_3$ with

$$\begin{aligned}
T_1 &= \varepsilon\left[\frac{1}{n}\sum_{i=1}^n \int_0^T H(\alpha,\beta,s,x_s(\alpha)))dW_s^i + \frac{1}{n}\sum_{i=1}^n \int_0^T \left(H(\alpha,\beta,s,X_s^i) - H(\alpha,\beta,s,x_s(\alpha)))\right) dW_s^i\right]\\
T_2 &= -\frac{1}{n}\sum_{i=1}^n \int_0^T H(\alpha,\beta,s,X_s^i)D(\alpha_0,\beta_0,s,X_s^i)ds\\
T_3 &= \frac{1}{n}\sum_{i=1}^n \int_0^T \left(H(\alpha,\beta,s,X_s^i)H(\theta_0,s,X_s^i) - \frac{1}{2}H^2(\alpha,\beta,s,X_s^i)\right) ds = \frac{1}{n}\sum_{i=1}^n T_3^i
\end{aligned}$$

By the same tools as in Lemma 8.1, we check that $T_1 = o_P(1), T_2 = o_P(1)$ and the main term

$$\begin{aligned}
T_3^i &= \int_0^T \left(V(\alpha,X_s^i) - \Phi(\beta, X_s^i - x_s(\alpha))(V(\alpha_0,X_s^i) - \Phi(\beta_0, X_s^i - x_s(\alpha_0))\right) ds\\
&- \frac{1}{2}\int_0^T \left(V(\alpha,X_s^i) - \Phi(\beta, X_s^i - x_s(\alpha)\right)^2 ds
\end{aligned}$$

converges to

$$\begin{aligned}
&\int_0^T (V(\alpha,x_s(\alpha_0)) - \Phi(\beta,x_s(\alpha_0) - x_s(\alpha))V(\alpha_0,x_s(\alpha_0)))\, ds\\
&- \frac{1}{2}\int_0^T (V(\alpha,x_s(\alpha_0) - \Phi(\beta,x_s(\alpha_0) - x_s(\alpha)))^2\, ds\\
&= -\frac{1}{2}\int_0^T (V(\alpha,x_s(\alpha_0)) - \Phi(\beta,x_s(\alpha_0) - x_s(\alpha)) - V(\alpha_0,x_s(\alpha_0)))^2\, ds + \frac{1}{2}\int_0^T (V(\alpha_0,x_s(\alpha_0))^2\, ds.
\end{aligned}$$

Therefore,

$$\frac{\varepsilon^2}{n}\Lambda_{\varepsilon,T}^n(\alpha_0,\beta_0) \to \frac{1}{2}\int_0^T (V(\alpha_0,x_s(\alpha_0))^2\, ds.$$

Hence, we have proved the convergence (i). The unformity of this convergence is obtained without difficulty taking into account our regularity assumptions.
The convergence (ii) is more subtle. We directly write the difference: $\frac{1}{n}(\Lambda_{\varepsilon,T}^n(\alpha_0,\beta) - \Lambda_{\varepsilon,T}^n(\alpha_0,\beta_0))$, replace dX_s^i by its expression (62) and develop. In the difference, several terms cancel and there remains:

$$\begin{aligned}
\frac{1}{n}(\Lambda_{\varepsilon,T}^n(\alpha_0,\beta) - \Lambda_{\varepsilon,T}^n(\alpha_0,\beta_0)) &= -\frac{1}{n\varepsilon^2}\sum_{i=1}^n \int_0^T \left(\Phi(\beta, X_s^i - x_s(\alpha_0)) - \Phi(\beta_0, X_s^i - x_s(\alpha_0))\right)^2 ds\\
&- \frac{1}{n\varepsilon}\sum_{i=1}^n \int_0^T \left(\Phi(\beta, X_s^i - x_s(\alpha_0)) - \Phi(\beta_0, X_s^i - x_s(\alpha_0))\right) dW_s^i\\
&+ \frac{1}{n\varepsilon^2}\sum_{i=1}^n \int_0^T \left(\Phi(\beta, X_s^i - x_s(\alpha_0)) - \Phi(\beta_0, X_s^i - x_s(\alpha_0))\right) D(\theta_0,s,X_s^i)ds
\end{aligned}$$

The last term is proved to be $o_P(\varepsilon)$ by the same tools as Lemma 8.2. For the first two terms, we write as done previously

$$\begin{aligned}\Phi(\beta, X_s^i - x_s(\alpha_0)) &= (X_s^i - x_s(\alpha_0))\frac{\partial\Phi}{\partial x}(\beta,0) + \ldots = \varepsilon g_s^i(\theta_0)\frac{\partial\Phi}{\partial x}(\beta,0) + \ldots \\ \Phi(\beta_0, X_s^i - x_s(\alpha_0)) &= (X_s^i - x_s(\alpha_0))\frac{\partial\Phi}{\partial x}(\beta_0,0) + \ldots = \varepsilon g_s^i(\theta_0)\frac{\partial\Phi}{\partial x}(\beta_0,0) + \ldots\end{aligned}$$

Skipping details which have been used repeatly, we obtain:

$$\begin{aligned}\frac{1}{n}(\Lambda_{\varepsilon,T}^n(\alpha_0,\beta) - \Lambda_{\varepsilon,T}^n(\alpha_0,\beta_0)) &= -\frac{1}{2n}\left(\frac{\partial\Phi}{\partial x}(\beta,0) - \frac{\partial\Phi}{\partial x}(\beta_0,0)\right)^2 \sum_{i=1}^n \int_0^T (g_s^i(\theta_0))^2 ds \\ &- \frac{1}{n}\left(\frac{\partial\Phi}{\partial x}(\beta,0) - \frac{\partial\Phi}{\partial x}(\beta_0,0)\right)\sum_{i=1}^n \int_0^T g_s^i(\theta_0) dW_s^i + o_P(\varepsilon)\end{aligned}$$

where the first term tends to $-\frac{1}{2}\left(\frac{\partial\Phi}{\partial x}(\beta,0) - \frac{\partial\Phi}{\partial x}(\beta_0,0)\right)^2 \mathbb{E}_\theta \int_0^T (g_s(\theta_0))^2 ds = \Lambda_2(\alpha_0,\beta,\beta_0)$ and the second term tends to 0. We have thus obtained the limit (ii). The uniformity can be proved without difficulty although after some cumbersome developments. □

Proof of Theorem 5.1. Following Gloter and Sørensen (2009) and Sørensen and Uchida (2003), the proof of consistency for a two-rate model goes in three steps:
step 1: Prove the consistency of $\hat{\alpha}_{\varepsilon,n}$: the consistency follows directly from Lemma 5.1, (i).
step 2: Prove that $\frac{\sqrt{n}}{\epsilon}(\hat{\alpha}_{\varepsilon,n} - \alpha_0)$ is $\mathbb{P}_{\theta_0}$ -tight.
As $\mathbb{P}_{\theta_0}(\hat{\alpha}_{\varepsilon,n} \in \Theta_\alpha) \to 1$, on the set $(\hat{\alpha}_{\varepsilon,n} \in \theta_\alpha)$, we can write:

$$0 = \frac{\partial\Lambda_{\varepsilon,T}^n}{\partial\alpha}(\hat{\alpha}_{\varepsilon,n}, \hat{\beta}_{\varepsilon,n}) = V_{\varepsilon,n} + (\hat{\alpha}_{\varepsilon,n} - \alpha_0)N_{\varepsilon,n}$$

where

$$V_{\varepsilon,n} = \frac{\partial\Lambda_{\varepsilon,T}^n}{\partial\alpha}(\alpha_0, \hat{\beta}_{\varepsilon,n}), \quad N_{\varepsilon,n} = \int_0^1 \frac{\partial^2\Lambda_{\varepsilon,T}^n}{\partial\alpha^2}(\alpha_0 + u(\hat{\alpha}_{\varepsilon,n} - \alpha_0), \hat{\beta}_{\varepsilon,n})du$$

Therefore,

$$\frac{\sqrt{n}}{\varepsilon}(\hat{\alpha}_{\varepsilon,n} - \alpha_0) = -\frac{\frac{\varepsilon}{\sqrt{n}}V_{\varepsilon,n}}{\frac{\varepsilon^2}{n}N_{\varepsilon,n}}.$$

We must prove that $\frac{\varepsilon}{\sqrt{n}}V_{\varepsilon,n}$ and $\frac{\varepsilon^2}{n}N_{\varepsilon,n}$ are tight. This can be done as for $\frac{\varepsilon}{\sqrt{n}}\frac{\partial\Lambda_{\varepsilon,T}^n}{\partial\alpha}(\alpha,\beta)$ and $\frac{\varepsilon^2}{n}\frac{\partial^2\Lambda_{\varepsilon,T}^n}{\partial\alpha^2}(\alpha,\beta)$.

step 3: Prove that $\hat{\beta}_{\varepsilon,n}$ is consistent. For this, it is enough to prove that, under $\mathbb{P}_{\theta_0}$,

$$\frac{1}{n}(\Lambda_{\varepsilon,T}^n(\hat{\alpha}_{\varepsilon,n},\beta) - \Lambda_{\varepsilon,T}^n(\hat{\alpha}_{\varepsilon,n},\beta_0)) \to \Lambda_2(\alpha_0,\beta,\beta_0)$$

uniformly in β. This follows from the previous steps and the same tools as in Lemma 5.1, (ii). □

Proof of Proposition 6.1.
We just have to prove uniformity properties with respect to θ. First, note that, for all $p \geq 1$, the constant $C_{2p}(t)$ of Proposition 3.1 only depends on t and is a continuous non decreasing function of t. By the assumption $L(\alpha) \leq L$, for all $p \geq 1$,

$$\sup_{\theta\in\Theta}\sup_{t\in[0,T]} \mathbb{E}_\theta(X_t - x_t(\alpha))^{2p} \leq \varepsilon^{2p} C_{2p}(T) e^{2pLT}. \tag{63}$$

For (ii), we look at $f(\theta,t) = \mathbb{E}_\theta X_t - x_t(\alpha)$ which satisfies (37). Let $c(\alpha) = \sup\{|x_s(\alpha)|, s \leq T\}$. Under (S1), $c(\alpha) \leq c$. Therefore, using (63), the term R_s which actually depends on θ, ε present in (37) satisfies, $|R_s| \leq K(\mathbb{E}_\theta(X_s - x_s(\alpha))^2(1 + B^k) + \mathbb{E}_\theta|X_s - x_s(\alpha)|^{2+k}) \leq \varepsilon^2 K(T)$, where the constant $K(T)$ does not depend on θ. This implies that

$$\sup_{\theta\in\Theta}\sup_{t\in[0,T]} |f(\theta,t)| = \sup_{\theta\in\Theta}\sup_{t\in[0,T]} |\mathbb{E}_\theta X_t - x_t(\alpha)| \leq \varepsilon^2 K(T) T e^{LT}. \tag{64}$$

Consider now $D(\theta, t, \varepsilon, X_t)$ (see (16)). Looking at the proof of Corollary 3.2, equations (46)-(48) and (63), (64) allow to get that

$$\sup_{\theta\in\Theta}\sup_{t\in[0,T]} |\mathbb{E}_\theta D(\theta,t,\varepsilon,X_t)| \leq \varepsilon^2 C_T,$$

$$\sup_{\theta\in\Theta}\sup_{t\in[0,T]} |\mathbb{E}_\theta D(\theta,t,\varepsilon,X_t)| \leq \varepsilon^2 C_T.$$

□

References

[1] Azencott, R. (1982) Formule de Taylor stochastique et développement asymptotique d'intégrales de Feynman Séminaire de probabilités (Strasbourg), 237-285.

[2] Ball,F., Mollison, D. and Scalia-Tomba G. (1997) Epidemics with two levels of mixing, *The Annals of Applied Probability* **7**, 46-89.

[3] Ball, F. and Sirl D. (2020) Stochastic SIR in Structured Populations, in *Stochastic Epidemic Models with inference*, **Part II** , 123-240. Britton,T. and Pardoux, E., Editors. Lecture Notes in Mathematics 2255, Mathematical Biosciences Subseries. Springer.

[4] Benachour, S., Roynette, B., Talay, D. and Vallois, P. (1998a). Nonlinear self-stabilizing processes - I Existence, invariant probability, propagation of chaos. *Stoch; proc. appl.* **75**, 173-201.

[5] Benachour, S., Roynette, B. and Vallois, P. (1998b). Nonlinear self-stabilizing processes - II Convergence to invariant probability. *Stoch; proc. appl.* **75**, 203-224.

[6] Benedetto, D., Caglioti, E. and Pulverenti, M. (1997). A kinetic equation for granular media. *Mathematical Modelling and Numerical Analysis*, **31** (5), 615-641.

[7] Britton,T. and Pardoux, E., Editors (2020) *Stochastic Epidemic Models with Inference*. Lecture Notes in Mathematics 2255, Mathematical Biosciences Subseries. Springer

[8] Carrillo, I.A., McCann, R.J. and Villani, C. (2006). Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Rat. Mech. Anal.* **179**, 217-263.

[9] Cattiaux, P., Guillin, A. and Malrieu, F. (2008). Probabilistic approach for granular media equations in the non-uniformly convex case, *Probab. Theory Relat. Fields* **140**, 19-40.

[10] Freidlin, M.I. and Wentzell A.D. (2012). *Random perturbations of dynamical sysytems.* Third edition, Springer. Heidelberg.

[11] Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations, *Statistics*, **21**, 99-116.

[12] Genon-Catalot V. and Larédo, C. (2014). Asymptotic equivalence of nonparametric diffusion and Euler scheme experiments, *The Annals of Statistics*, **42**, 1145-1165.

[13] Gloter, A. and Sørensen, M. (2009). Estimation for stochastic differential equations with a small diffusion coefficient. *Stoch. Proc. Appl.* **119**, 679-699.
[14] Guy, R., Larédo, C. and Vergu, E. (2014) Parametric inference for discretely observed multidimensional diffusions with small diffusion coefficient .*Stoch. Proc. Appl.* **124**, 51-80.
[15] Herrmann, S., Imkeller, P. and Peithmann, D. (2008). Large deviations and a Kramers'type low for self-stabilizing diffusions. *The Annals of Applied Probability*, **18**, 1379-1423.
[16] Herrmann, S. and Tugaut, J. (2010). Non uniqueness of stationary measures for self-stabilizing diffusions. *Stoch. Proc. Appl.* **120**, 1215-1246.
[17] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211-229.
[18] Kessler, M. (2000) Simple and explicit estimating functions for a discretely observed diffusion process.*Scandinavian journal of Statistics* **27**(1), 65-82.
[19] Kessler, M., Lindner, A. and Sørensen, M., Editors (2012). *Statistical methods for stochastic differential equations*. CRC press. Taylor & Francis Group. Boca Raton.
[20] Kutoyants, Y.A., (1984). *Parameter estimation for stochastic processes*. Berlin: Heldermann.
[21] Kutoyants, Y.A., (2004). *Statistical inference for ergodic diffusion processes*. Springer, London. L. (2007). Convergence rate of the mle in generalized linear and nonlinear mixed-effects models: Theory and applications. *J. Statist. Plann Inf.* **137**, 1787-1804.
[22] Larédo, C. (1990). A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process. *The Annals of Statistics* **18**, 1158-1171.
[23] Larédo, C. (2020). Statistical Inference for epidemicProcesses in a Homogeneous Community. **Part IV** 363-472. Britton,T. and Pardoux, E., Editors. Lecture Notes in Mathematics 2255, Mathematical Biosciences Subseries. Springer.
[24] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.
[25] Le Cam, L. and Yang, , G. L. (2000). *Asymptotics in Statistics: Some Basic Concepts*, second ed. Springer, New York.
[26] Mac Kean, H.P., Jr (1966). A class of Markov processes associated with nonlinear parabolic equation. Proceedings NatL. Acad. Sci. U.S.A. **56**, 1907-1911.
[27] Malrieu, F. (2003). Convergence to equilibrium for granular media equations and their Euler schemes. *Annals of applied Probability* **13**, 540-560.
[28] Sørensen, M. and Uchida, M. (2003). Small diffusion asymptotics for discretely sampled stochastic differential equations. *Bernoulli* **9**, 1051-1069.
[29] Sznitman, A.-S. (1991). Topics in propagation of chaos.*Ecole d'été de probabilités de Saint-Flour XIX-1989. Lecture Notes in Math.* **1464**, 165-251. Springer, Berlin.