Mixing and convergence rates for a family of Markov processes approximating SDEs

S. A. Klokov* and A. Yu. Veretennikov †

November 4, 2003

Abstract

We study a class of Markov processes of the type $X_{n+1,h} = X_{n,h} + F(X_{n,h})h + \sqrt{h}\,\xi_{n+1}$, where $F : \mathbb{R}^d \to \mathbb{R}^d$ is a bounded continuous function, (ξ_n) are i.i.d. random variables with zero mean, and $t = nh$ understood as "macro-time". Such processes are approximations to the SDE, $dX_t = F(X_t)\,dt + dW_t$. Upper estimates for β-mixing and convergence rates to invariant measure are established under certain assumptions on smoothness of F, the density of ξ_n and some recurrence conditions. The estimates are analogous to those for the limiting SDE.

Keywords: Stochastic Difference Equation – Mixing – Markov Process – Euler Scheme – Malliavin Calculus.

1 Introduction

In this paper we continue studying of β-mixing and convergence rates to invariant distribution for models treated in [7, 8, 9, 3]. The main goal is to find an answer to the following question: if one approximates a stochastic

*School of Mathematics, University of Leeds (UK) & Omsk Branch of Institute of Mathematics (Russia); email: klokov@iitam.omsk.net.ru

†School of Mathematics, University of Leeds (UK) & Institute of Information Transmission Problems (Russia); email: veretenn@maths.leeds.ac.uk

differential equation by Euler scheme, what are conditions which ensure a uniform mixing and convergence to equilibrium rates? The term "uniform" relates to different discretisation step sizes, and not to, e.g., initial data. We consider here only the case of bounded coefficients. Notice that the question of convergence of approximation schemes to the limiting solution is not discussed; it would be interesting to investigate relationships between these two problems. We consider approximations by not necessarily Gaussian noise ξ_n, however our approach requires that the "drift" function F is smooth enough, and ξ_n possesses a density with certain nice properties, depending on dimension. The latter dependence may look strange, but it relates naturally to the extensive use of the Bismut approach of Malliavin calculus, although the name is usually applied for continuous time case. Notice that for "strong approximations" (that is, with ξ_n Gaussian), smoothness assumptions on F may be dropped, due to the Harnack inequality technique, which case will be treated separately in another paper, because of an entirely different approach. Also notice that in principle, a diffusion coefficient, at least non-degenerate, may be added to the scheme. We did not do it here, in order not to overload the presentation.

Consider a family of Markov processes of the following form,

$$X_{n+1,h} = X_{n,h} + F(X_{n,h})h + \sqrt{h}\,\xi_{n+1}, \quad X_{0,h} = x \in \mathbb{R}^d, \tag{1}$$

where $F : \mathbb{R}^d \to \mathbb{R}^d$ is a bounded continuous function, the sequence (ξ_n) is a sequence of i.i.d. random variables with $\mathbb{E}\xi_n = 0$ and $\mathbb{E}|\xi_n|^m < \infty$ for some $m \geq 2$. Letting $h = 1$ we get "the discrete case model" from [7, 3]. Consider $t = nh$ as "macro-time" and pass to the limit $h \to 0$. Formally, we look at the process as an approximation to the stochastic differential equation with constant diffusion,

$$dX_t = F(X_t)\,dt + \sigma\,dW_t, \quad X_0 = x \in \mathbb{R}^d, \tag{2}$$

where (W_t) is a Wiener process in $\mathbb{R}^d$ and σ is the covariance matrix of ξ_n. Processes satisfying equation (2) were studied in [7, 8] and we call this setting by "the continuous case model". So, the family of processes (1) is a link between the discrete and continuous models, and studying uniform properties of $X_{n,h}$ for all h small enough seems important.

Denote the distance in total variation between two probability measures μ and ν by $\|\mu - \nu\|_{TV} = \mathrm{var}\,(\mu - \nu)$. The β-mixing coefficient is defined by

$$\beta_x(t) = \sup_{s \geq 0} \mathbb{E}_x \operatorname*{var}_{B \in \mathcal{F}^X_{\geq t+s}} (\mathbb{P}(B \mid \mathcal{F}^X_{\leq s}) - \mathbb{P}(B)),$$

where $\mathcal{F}_I^X = \sigma\{X_{n,h} : nh \in I\}$ and $\mathbb{E}_x$ denotes expectation of the process starting from x. It is not emphasized in the notation but $\beta_x(t)$ depend on the value of h.

Consider the family of processes $(X_{n,h})$ for all $h > 0$ small enough. Under certain assumptions, each process $X_{n,h}$ (h is fixed) possesses an invariant measure, and there are convergence to this measure and β-mixing [7, 3]. Our goal is to establish *uniform* upper estimates for convergence and β-mixing rates on the "macro-time" scale as $t = nh \to \infty$.

In [7, 3] some polynomial and sub-exponential bounds for the rates were established for both discrete and continuous models. Methods of proof were similar but not identical. The method of [7, 3] for the discrete model was essentially based on two ideas: (i) estimating of moments of hitting-times of a compact set (see the discussion in [7] and relations (8)–(10)) and (ii) the coupling method for the process on the compact set through the local Doeblin condition (see condition (3) below). An attempt to apply this method here is straightforward but reveals that the latter condition is difficult to verify for our model while the former one is easy to check.

The rest of the paper is organized as follows. Basic assumptions on the process $(X_{n,h})$ and our main results are formulated in Section 2. Some properties of the process are studied in Section 3 where we show that the coupling method works in our settings; the use of Malliavin calculus is explained from scratch, for the reader convenience. Section 4 deals with two different forms of recurrence conditions. The main results of the paper are established in Section 5.

2 Basic assumptions and main results

We will make some assumptions on the function F and the distribution of the sequence of "noise" (ξ_n) from difference equation (1). It is convenient to formulate all of them right here, although particular results in the sequel may require only some part of them. There are three main assumptions, namely, recurrence condition (F_2) which ensures an attraction to the origin, a moment assumption (either (D_4), or (D_5)), and condition (D_1) on the density. Other assumptions, (F_1), (D_2) and (D_3), are technical ones, they ensure the work of the Malliavin calculus approach, and also provide the local Doeblin condition (3) below. Condition (3) is a version of a standard one in ergodic theory for Markov processes since 50s, cf. [1]. Although assumptions (D_1)–(D_5) have different nature, — positiveness, smoothness, and moment properties, — all of them are features of the random variables ξ_n; this is why we use the same notation (D) with indices for them.

(F_1) the function F and its derivatives of orders up to $d+2$ are bounded;

(F_2) there exist $R_0 > 0$, $0 \le p < 1$ and $r > 0$ such that

$$\langle F(x), x\rangle \le -r|x|^{1-p}, \quad |x| > R_0;$$

(D_1) the distribution of ξ_n has density $p(x)$ which is positive everywhere;

(D_2) the derivatives of $p(x)$ of orders up to $d+1$ are absolute integrable;

(D_3) Fisher information is finite, $\displaystyle\int_{\mathbb{R}^d} \frac{|p'(x)|^2}{p(x)}\,dx < \infty$;

(D_4) $\mathbb{E}|\xi_n|^m < \infty$, $m \ge 2$;

(D_5) $\mathbb{E}\exp\{\kappa|\xi_n|^\alpha\} < \infty$ with $0 < \kappa < K$ and $0 < \alpha \le 1$.

We introduce one more condition of local mixing, see [11, 7, 3]. Let (X_n) be a Markov process, B a compact set, $\tau_1 = \inf\{n \ge 0 : X_n \in B\}$, $\tau_{k+1} = \inf\{n > \tau_k : X_n \in B\}$ hitting times of B. Define "the process on B", $X^B_n = X_{\tau_n}$, with n-step transition probabilities $P^B(n,x,dy)$. We say that the process (X_n) satisfies the *local Doeblin condition*, if for any $R > 0$ large enough and $B = B(0,R)$ there is an integer $n_0 = n_0(R)$ such that

$$\inf_{x,x'\in B} \int \left(1 \wedge \frac{P^B(n_0,x,dy)}{P^B(n_0,x',dy)}\right) P^B(n_0,x',dy) =: q(R,n_0) > 0, \qquad (3)$$

where $P(dy)/P'(dy)$ denotes the derivative of absolute continuous part of P with respect to P' (singular part may be non-trivial). The condition may look sophisticated but its point is to provide non-singularity of the measures within the ball B. This form of non-singularity, rather than a "petite sets" type condition, is used because it gives better constants uder the exponential. The coefficient of ergodicity q in a case of uniform bound was suggested by Dobrushin.

Considering a family of stochastic processes depending on a parameter h, we have to adjust the local Doeblin condition for this family. We say that the *local Doeblin condition* is satisfied for the *family of processes* $(X_{n,h})$ depending on the parameter h, if for any $R > 0$ large enough there exist reals $T = T(R) > 0$, $h_0 > 0$, and an integer-valued function $N = N(T, h)$ such that $N(T, h)h \le T$ and

$$\inf_{h\le h_0} \inf_{x,x'\in B} \int \left(1 \wedge \frac{P^B(N(T,h),x,dy)}{P^B(N(T,h),x',dy)}\right) P^B(N(T,h),x',dy) =: q(R) > 0. \tag{4}$$

Theorem 1 (mixing and convergence rates) *Let the family of processes* $(X_{n,h})$ *satisfy conditions* (F_1), (F_2), *and* (D_1)–(D_3) *with some parameters* R_0, p, r; *and in addition, either* m *in* (D_4), *or* K *and* α *in* (D_5). *In each of the cases 1–3 below,*

(a) the family of processes $(X_{n,h})$ *satisfies the local Doeblin condition (4);*

(b) for each h *small enough there exists the invariant measure* $\mu = \mu_h$ *(the measures may be different for different values of* h *but all constants and functions in the following estimates can be chosen uniformly in* h*);*

(c) marginal distributions $\mu_x(t) = \mathcal{L}(X_{n,h}|X_{0,h} = x)$ *converge to* μ *at some specific rate on the scale of macro-time* $t = nh$ *and* β*-mixing holds with the same rate.*

1. *If* $p = 0$ *and* (D_5) *holds with* $\alpha = 1$ *and some* $K > 0$, *then*

$$\begin{aligned} \|\mu_x(t) - \mu\|_{TV} &\le C(x)\exp\{-c(1+t)\}, \\ \beta_x(t) &\le C(x)\exp\{-c(1+t)\}, \end{aligned}$$

with some $c > 0$ *and positive function* $C(x)$.

2. *If* $0 < p < 1$ *and* (D_5) *holds with* $0 < \alpha < 1 - p$ *and some* $K > 0$, *then*

$$\begin{aligned} \|\mu_x(t) - \mu\|_{TV} &\le C(x)\exp\{-c(1+t)^\delta\}, \\ \beta_x(t) &\le C(x)\exp\{-c(1+t)^\delta\}, \end{aligned}$$

with any $0 < \delta < \alpha/(1+p)$ *and some* $c > 0$ *and a positive function* $C(x)$ *depending on* δ.

3. If $p = 1$ *and* (D_4) *holds with* $m > 4$ *and* $r > (m+1)\mathbb{E}|\xi_1|^m/2$*, then*

$$\begin{aligned} \|\mu_x(t) - \mu\|_{TV} &\le C(1+|x|^m)(1+t)^{-k}, \\ \beta_x(t) &\le C(1+|x|^m)(1+t)^{-k}, \end{aligned}$$

where $C > 0$ *and any* $k < (m-2)/2$ *can be used.*

Proof of the Theorem We give the proof in Section 5

Remark 1 The assumption $m > 4$ in the case 3 is needed for integrating with respect to μ only. The value $\|\mu_x(t) - \mu_{x'}(t)\|_{TV}$ can be estimated from above via $C(1+|x|^m+|x'|^m)(1+t)^{-k}$ for $x, x' \in \mathbb{R}^d$, under $m > 2$ and $k < m/2$. See the remark after the proof of Theorem 1 in [7].

Remark 2 Since the family $(X_{n,h})$ approximates the continuous process (2), it is possible that the sequence of invariant measures (μ_h) converges to the invariant measure for SDE (2) in some sense. Indeed, under certain conditions it can be shown. However, we will not use this here.

3 Uniform properties of $(X_{n,h})$ on a finite macro-time interval

Let $[0, T]$ be a period of macro-time. We choose the value of T later. Define

$$N = N(T, h) = \sup\{n \ge 0 : nh \le T\}. \tag{5}$$

For any fixed value of T we have $N = N(T,h) \to \infty$ and $Nh \sim T$ as $h \to 0$.

Fix $R > 0$ and the ball $B = B(0, R)$. Let D be an open subset in B. It is well-known that the solution to SDE (2) has the property,

$$\inf_{x \in B} \mathbb{P}_x(X_T \in D) > 0.$$

One can write a lower bound depending on the set D only, if the ball B and the value of T are fixed. It seems plausible that random events $\{X_{N,h} \in D\}$ have probabilities bounded away from zero, too, for all h small enough. A reference on weak convergence of approximations to the SDE solution, and on some extension of the Donsker–Prokhorov principle might be given, but we propose another simple proof, for reader's convenience and in order to explore the capacity of the approach.

Lemma 1 *Let $D \subset B = B(0, R)$ be an open set and assume that the CLT holds true for the sequence (ξ_n),*

$$\sqrt{h} \sum_{k=1}^{N} \xi_k \xrightarrow{d} \mathcal{N}(0, T\sigma\sigma^*), \quad h \to 0,$$

where the covariance matrix $\sigma\sigma^*$ is non-degenerate. Then there is $h_0 > 0$ such that

$$\inf_{h \in (0,h_0]} \inf_{x \in B} \mathbb{P}_x(X_{N,h} \in D) > 0.$$

Proof It follows from (1) that

$$X_{n,h} = X_{0,h} + h \sum_{k=0}^{n-1} F(X_{k,h}) + \sqrt{h} \sum_{k=0}^{n-1} \xi_{k+1}. \tag{6}$$

For the sake of simplicity we assume that $B(0, \rho_0) \subset D$ with some $\rho_0 > 0$ (the construction in the general case is similar). Denote $M = \sup |F(x)|$, fix an integer L so large that

$$\rho = \frac{MT}{L} \vee \frac{R}{L} \leq \frac{\rho_0}{3},$$

and consider the balls $B_k = B(0, k\rho)$. Obviously, $B_3 \subset D$ and $x \in B_L$. One represents $X_{N,h}$ as the sum of L blocs,

$$X_{N,h} = x + \sum_{j=0}^{L-1} Y_j,$$

where each bloc Y_j contains $[N/L]$ or $[N/L] + 1$ summands,

$$Y_j = \sum_{k=[jN/L]}^{[(j+1)N/L]-1} \Big(F(X_{k,h})h + \sqrt{h}\, \xi_{k+1} \Big).$$

We show that starting from $x \in B_L$, the process $X_{n,h}$ will hit the sets B_{L-1}, B_{L-2}, ..., B_3, as $n = [N/L]$, $[2N/L]$, ..., N, correspondingly, with probabilities bounded away from 0. Due to the CLT,

$$\sqrt{h} \sum_{k=[jN/L]}^{[(j+1)N/L]-1} \xi_{k+1} \xrightarrow{d} \mathcal{N}(0, L^{-1}T\sigma\sigma^*), \quad h \to 0. \tag{7}$$

The drift arising from F is small enough,

$$\left|h \sum_{k=[jN/L]}^{[(j+1)N/L]-1} F(X_{k,h})\right| \le M([N/L]+1)h \sim MT/L \le \rho. \tag{8}$$

Suppose that $X_{[jN/L]} \in B_i$ for some i. Due to (7), the random events

$$\left\{X_{[jN/L]} + \sqrt{h} \sum_{k=[jN/L]}^{[(j+1)N/L]-1} \xi_{k+1} \in B_{i-2}\right\} \tag{9}$$

have probabilities bounded away from zero. Indeed, a Gaussian random variable η with $\mathbb{E}\eta = 0$ and the covariance matrix $L^{-1}T\sigma\sigma^*$ satisfies the inequality

$$\min_{3\le i\le L} \inf_{x\in B_i} \mathbb{P}(\eta \in B_{i-2} - x) \ge 2\varepsilon$$

for some $\varepsilon > 0$. Due to weak convergence (7), one can take ε as a lower bound for probabilities of events (9) for all h small enough. By virtue of (8), the drift is bounded by ρ. Hence, $X_{[(j+1)N/L]} \in B_{i-1}$. Similar calculations show that $\mathbb{P}(X_{[(j+1)N/L]} \in B_3 \mid X_{[jN/L]} \in B_3)$ is bounded away from zero too, so that

$$\inf_{h\in(0,h_0]} \inf_{x\in B} \mathbb{P}_x(X_{N,h} \in D) \ge \inf_{h\in(0,h_0]} \inf_{x\in B} \mathbb{P}_x(X_{N,h} \in B_3) \ge \varepsilon^L > 0. \qquad \square$$

The Lemma 1 is an important step towards the local Doeblin condition.

Lemma 2 *Let $p_N(x,y)$ be the density of $X_{N,h}$ given $X_0 = x$, and for some $0 < h_1 \le h_0$ the following formula hold true (the value of h_0 was introduced in the Lemma 1): $\forall\varepsilon > 0,\ \exists\delta > 0$, such that $\forall h \in (0,h_1],\ \forall x \in B$,*

$$|y - y'| < \delta \Rightarrow |p_N(x,y) - p_N(x,y')| < \varepsilon. \tag{10}$$

Then there exists a set $B_0 \subset B$ with a positive Lebesgue measure such that

$$\inf_{h\in(0,h_1]} \inf_{x,x'\in B} \inf_{y\in B_0} p_N(x,y) \wedge p_N(x',y) > 0. \tag{11}$$

Proof Denote $P(D) = \inf_{h\in(0,h_0]} \inf_{x\in B} \mathbb{P}_x(X_{N,h} \in D)$, fix an open set $B_1 \subset B$ with $\operatorname{mes} B_1 > 0$, and use the Lemma 1,

$$\mathbb{P}_x(X_{N,h} \in B_1) = \int_{B_1} p_N(x,y)\, dy \ge P(B_1) > 0.$$

By virtue of (10), there exist $x_1 \in B_1$ and a ball $B_2 = B(x_1, \rho_1) \subset B_1$ such that

$$\inf_{y \in B_2} p_N(x, y) \geq \frac{P(B_1)}{2 \operatorname{mes} B_1}.$$

Similarly, for any other point $x' \in B$,

$$\int_{B_2} p_N(x', y)\, dy \geq P(B_2) > 0,$$

and there are a point $x_2 \in B_2$ and a ball $B_3 = B(x_2, \rho_2)$ such that

$$\inf_{y \in B_3} p_N(x', y) \geq \frac{P(B_2)}{2 \operatorname{mes} B_2}\,.$$

Taking $B_0 = B_2 \cap B_3 \neq \varnothing$, we have

$$\inf_{x, x' \in B} \inf_{y \in B_0} p_N(x, y) \wedge p_N(x', y) \geq \frac{P(B_1)}{2 \operatorname{mes} B_1} \wedge \frac{P(B_2)}{2 \operatorname{mes} B_2},$$

where the lower bound does not depend on h. □

One way to establishing formula (10) is given in the next lemma.

Lemma 3 *Let $p_N(x, y)$ and $f_N(x, \lambda)$ be the density and the characteristic function of $X_{N,h}$, given $X_{0,h} = x$. Assume that for some $h_1 > 0$,*

$$\sup_x \sup_{h \leq h_1} \int_{\mathbb{R}^d} |f_N(x, \lambda)|\, d\lambda = K < \infty,$$

$$\sup_x \sup_{h \leq h_1} \int_{|\lambda| > L} |f_N(x, \lambda)|\, d\lambda \to 0, \quad L \to \infty. \tag{12}$$

Then the formula (10) *is true.*

Proof Since characteristic functions $f_N(x, \lambda)$ are integrable, and the random variables $X_{N,h}$ have densities, we may write the inversion formula,

$$p_N(x, y) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle \lambda, y \rangle} f_N(x, \lambda)\, d\lambda.$$

Let ε be an arbitrary positive number. Using (12), fix L so large that

$$\sup_x \sup_{h \leq h_1} (2\pi)^{-d} \int_{|\lambda| > L} 2\, |f_N(x, \lambda)|\, d\lambda < \varepsilon/2.$$

One estimates,

$$|p_N(x,y)-p_N(x,y')| \le (2\pi)^{-d}\int_{|\lambda|\le L}\left|e^{-i\langle\lambda,y\rangle}-e^{-i\langle\lambda,y'\rangle}\right||f_N(x,\lambda)|\,d\lambda+\varepsilon/2.$$

Since $|e^{iz}-1|\le|z|$, $z\in\mathbb{R}$, the integral is bounded by $(2\pi)^{-d}KL|y-y'|$. Hence, (10) is true with $\delta=(2\pi)^d\varepsilon/(2KL)$. □

Further results will be obtained with the help of the Bismut approach to Malliavin calculus; they imply that Lemmas 2 and 3 are applicable. The idea is to introduce a family of random processes depending on a parameter and differentiate several times certain integral identities, prototypes of Girsanov's formulae. In the one-dimensional case we get estimates,

$$|\mathbb{E}_x f''(X_{N,h})|\le C\sup_z|f(z)|,$$

where f is any smooth bounded complex-valued function, and the constant C is chosen uniformly in h and independently of f. Taking $f(y)=\exp(i\lambda y)$ implies that the characteristic function of $X_{N,h}$ decreases at least as C/λ^2 as $\lambda\to\infty$. It gives integrability and the estimates (12) required in Lemma 3. This approach is implemented in the rest of the section.

From now we assume that the condition (D_1) holds, $F(x)$ and $p(x)$ are smooth enough (to be defined later) and their derivatives are bounded. Then the mapping $I+Fh:\mathbb{R}^d\to\mathbb{R}^d$ is invertible, differentiable, and distributions of $X_{n,h}$ have densities for $n\ge1$, as long as h is small enough.

Let $\varepsilon\in\mathbb{R}^d$ be a parameter. Introduce the family of random processes,

$$X^\varepsilon_{n+1,h}=X^\varepsilon_{n,h}+F(X^\varepsilon_{n,h})h+\sqrt{h}(\xi_{n+1}+\varepsilon\sqrt{h}),\quad X^\varepsilon_{0,h}=X_{0,h}. \tag{13}$$

Then $X^0_{n,h}=X_{n,h}$, and

$$X^\varepsilon_{n,h}=X_{0,h}+h\sum_{k=0}^{n-1}F(X^\varepsilon_{k,h})+\sqrt{h}\sum_{k=0}^{n-1}\xi_{k+1}+\varepsilon nh, \tag{14}$$

that is, the value $X^\varepsilon_{n,h}$ can be calculated as a function of $X_{0,h}$ and $\xi_1,\ldots,\xi_n$. Denote by g^ε_n corresponding mappings,

$$\begin{aligned} g^\varepsilon_1(x,y_1) &= x+F(x)h+\sqrt{h}(y_1+\varepsilon\sqrt{h}),\\ g^\varepsilon_2(x,y_1,y_2) &= g^\varepsilon_1(g^\varepsilon_1(x,y_1),y_2),\\ &\cdots\\ g^\varepsilon_n(x,y_1,\ldots,y_n) &= g^\varepsilon_1(g^\varepsilon_{n-1}(x,y_1,\ldots,y_{n-1}),y_n). \end{aligned} \tag{15}$$

Rewrite $\mathbb{E}_x f(X_{n,h})$ for any smooth function $f : \mathbb{R}^d \to \mathbb{C}$, through change of measure,

$$\begin{aligned}
\mathbb{E}_x f(X_{n,h}) &= \int \ldots \int f(g_n^0(x, y_1, \ldots, y_n)) \prod_{k=1}^n p(y_k)\, dy_1 \ldots dy_n \\
&= \int \ldots \int f(g_n^\varepsilon(x, z_1, \ldots, z_n)) \prod_{k=1}^n \frac{p(z_k + \varepsilon\sqrt{h})}{p(z_k)} \prod_{k=1}^n p(z_k)\, dz_1 \ldots dz_n \\
&= \mathbb{E}_x f(X_{n,h}^\varepsilon) \gamma_{n,h}^\varepsilon,
\end{aligned}$$

where

$$\gamma_{n,h}^\varepsilon = \prod_{k=1}^n \frac{p(\xi_k + \varepsilon\sqrt{h})}{p(\xi_k)} \tag{16}$$

is a random density and $\gamma_{n,h}^0 = 1$. The identity holds true for each ε,

$$\mathbb{E}_x f(X_{n,h}^\varepsilon) \gamma_{n,h}^\varepsilon = \mathbb{E}_x f(X_{n,h}). \tag{17}$$

Remark 3 If $X_{n,h} = \psi_n(X_{0,h}, \xi_1, \ldots, \xi_n)$, where ψ_n are Borel functions and $X_{n,h}^\varepsilon$ is calculated in the same way except using variables $\xi_1 + \varepsilon\sqrt{h}, \ldots, \xi_n + \varepsilon\sqrt{h}$ instead of $\xi_1, \ldots, \xi_n$, then identity (17) is true, too.

Case $d = 1$. We start with this case because calculations for $d > 1$ are technically involved, though the basic idea remains the same.

We will often suppress lower indices in calculations to simplify formulas and write X, X^ε, γ^ε, ... instead of $X_{n,h}$, $X_{n,h}^\varepsilon$, $\gamma_{n,h}^\varepsilon$, ... respectively. Also we will often suppress upper zero indices, corresponding to $\varepsilon = 0$.

Differentiate (17) with respect to ε and substitute $\varepsilon = 0$,

$$0 = \mathbb{E}_x f'(X) \frac{\partial X^\varepsilon}{\partial \varepsilon}\Big|_{\varepsilon=0} + \mathbb{E}_x f(X) \frac{\partial \gamma^\varepsilon}{\partial \varepsilon}\Big|_{\varepsilon=0}. \tag{18}$$

From formula (16) we deduce,

$$\frac{\partial \gamma^\varepsilon}{\partial \varepsilon} = \sqrt{h} \sum_{j=1}^n \frac{p'(\xi_j + \varepsilon\sqrt{h})}{p(\xi_j + \varepsilon\sqrt{h})} \prod_{k=1}^n \frac{p(\xi_k + \varepsilon\sqrt{h})}{p(\xi_k)},$$

whence,

$$\frac{\partial \gamma^\varepsilon}{\partial \varepsilon}\Big|_{\varepsilon=0} = \sqrt{h} \sum_{j=1}^n \frac{p'(\xi_j)}{p(\xi_j)}.$$

Denote $Y^\varepsilon = \partial X^\varepsilon / \partial \varepsilon$. By (13), the processes $Y^\varepsilon_{n,h}$ satisfy equations

$$Y^\varepsilon_{n+1,h} = (1 + F'(X^\varepsilon_{n,h})h)Y^\varepsilon_{n,h} + h, \quad Y^\varepsilon_{0,h} = 0. \tag{19}$$

We need an auxiliary result about solutions to equations of the type (19).

Lemma 4 *Let $h > 0$ be small enough, $0 \le n \le N = N(T,h)$, (a_n) and (b_n) sequences of real numbers. Consider (u_n) as the solution to equations*

$$u_{n+1} = (1 + ha_n)u_n + hb_n, \quad u_0 = 0. \tag{20}$$

If $\sup_n |a_n| \le M$ and $\sup_n |b_n| \le M$, then $|u_n| \le C$ with some real C which can be chosen uniformly for all h small enough.

If $\sup_n |a_n| \le M$ and $b_n = 1$, then u_n, $n \ge 1$, are positive numbers and $u_N^{-1} \le C$ with some $C > 0$, for all h small enough.

Proof By virtue of the triangle inequality,

$$|u_{n+1}| \le (1 + h|a_n|)|u_n| + h|b_n| \le (1 + Mh)|u_n| + Mh.$$

Since $|u_1| \le Mh$, we get by iterations,

$$|u_n| \le Mh(1 + (1+Mh) + (1+Mh)^2 + \ldots + (1+Mh)^{n-1}) = (1+Mh)^n - 1.$$

If $n = N(T,h)$ and $h \to 0$, then $nh \to T$ and $(1+Mh)^n - 1 \to e^{TM} - 1$, and we can take $C = e^{TM}$. For $n < N(T,h)$, the same bound holds true.

To prove the second part of the lemma assume $h < 1/M$. We have $u_1 = h > 0$, and $u_{n+1} \ge (1 - Mh)u_n + h > 0$. Again, by iterations,

$$u_n \ge h[1 + (1-Mh) + (1-Mh)^2 + \ldots + (1-Mh)^{n-1}] = (1 - (1-Mh)^n)/M$$

and the lower bound tends to $(1 - e^{-TM})/M$ as $h \to 0$. Therefore $u_N^{-1} \le 2M/(1 - e^{-TM})$ for all h small enough. □

Calculations below show how can we remove $Y = Y^0 = (\partial X^\varepsilon / \partial \varepsilon)|_{\varepsilon=0}$ from (18) and get an important estimate. By Lemma 4, $Y^\varepsilon_{n,h}$ and $(Y^\varepsilon_{N,h})^{-1}$ are bounded. Since $Y^\varepsilon_{n,h}$ is computed through $X^\varepsilon_{n,h}$ (see equations (19)), accordingly to Remark 3 one writes the identity (with lower indices N, h),

$$\mathbb{E}_x f(X)\frac{1}{Y} = \mathbb{E}_x f(X^\varepsilon)\frac{1}{Y^\varepsilon}\gamma^\varepsilon. \tag{21}$$

Denote $Z^\varepsilon = \partial Y^\varepsilon / \partial \varepsilon$. The process $Z^\varepsilon_{n,h}$ is the solution to equations,

$$Z^\varepsilon_{n+1,h} = (1 + hF'(X^\varepsilon_{n,h}))Z^\varepsilon_{n,h} + hF''(X^\varepsilon_{n,h})(Y^\varepsilon_{n,h})^2, \quad Z_{0,h} = 0.$$

Use the Lemma 4 with $a_n = F'(X^\varepsilon_{n,h})$, $b_n = F''(X^\varepsilon_{n,h})(Y^\varepsilon_{n,h})^2$, and obtain that $|Z^\varepsilon|$ are bounded for all h small enough, if F is twice differentiable with bounded derivatives.

Now differentiate (21) and set $\varepsilon = 0$,

$$\begin{aligned} 0 &= \mathbb{E}_x f'(X) Y \frac{1}{Y} + \mathbb{E}_x f(X) \frac{-Z}{(Y)^2} \\ &\quad + \mathbb{E}_x \left(f(X) \frac{1}{Y} \sqrt{h} \sum_{k=1}^N \frac{p'(\xi_k)}{p(\xi_k)} \right), \end{aligned}$$

or, equivalently,

$$\mathbb{E}_x f'(X) = \mathbb{E}_x f(X) \frac{Z}{Y^2} - \mathbb{E}_x \left(f(X) \frac{1}{Y} \sqrt{h} \sum_{k=1}^N \frac{p'(\xi_k)}{p(\xi_k)} \right). \tag{22}$$

We will estimate terms in the right hand side of (22). Those ones with f, $Y = Y^0_{N,h}$ and $Z = Z^0_{N,h}$ are bounded. Hence,

$$|\mathbb{E}_x f'(X_{N,h})| \le C \sup_x |f(x)| \left(1 + \mathbb{E} \left| \sqrt{h} \sum_{k=1}^N \frac{p'(\xi_k)}{p(\xi_k)} \right| \right), \tag{23}$$

Assume (D_3) holds, then the random variables $p'(\xi_k)/p(\xi_k)$ have zero mean, and their independence gives one,

$$\mathbb{E} \left(\sqrt{h} \sum_{k=1}^N \frac{p'(\xi_k)}{p(\xi_k)} \right)^2 = h \sum_{k=1}^N \mathbb{E} \left(\frac{p'(\xi_k)}{p(\xi_k)} \right)^2 \le T \mathbb{E} \left(\frac{p'(\xi_1)}{p(\xi_1)} \right)^2. \tag{24}$$

Applying Cauchy–Bounyakovskii–Schwarz inequality, we get from (23),

$$|E_x f'(X_{N,h})| \le C \sup_x |f(x)| \left(1 + \sqrt{T \mathbb{E} \left(\frac{p'(\xi_1)}{p(\xi_1)} \right)^2} \right). \tag{25}$$

Thus, there exists a new constant C such that

$$|\mathbb{E}_x f'(X_{N,h})| \le C \sup_x |f(x)|.$$

Remark 4 Using different functions f, we obtain several useful inequalities. Particularly, let $f(x)=e^{i\lambda x}$ to have $|i\lambda \mathbb{E}_x e^{i\lambda X_{N,h}}| \le C$. So, for the characteristic function of $X_{N,h}$ we get $|\mathbb{E}_x e^{i\lambda X_{N,h}}| \le 1 \wedge (C/(1+|\lambda|))$.

We start computations giving us estimates involving higher derivatives of f. Let $g(\xi)$ denote an expression which can be evaluated using the random variables of $\xi_1, \ldots, \xi_N$. Accordingly to Remark 3, one may write the identity,

$$\mathbb{E}_x f(X) g(\xi) \frac{1}{Y} = \mathbb{E}_x f(X^\varepsilon) g(\xi^\varepsilon) \frac{1}{Y^\varepsilon} \gamma^\varepsilon,$$

where $g(\xi^\varepsilon)$ is computed in the same way as $g(\xi)$ except using variables $\xi_1+\varepsilon\sqrt{h}, \ldots, \xi_N+\varepsilon\sqrt{h}$ instead of $\xi_1, \ldots, \xi_N$. Differentiating with respect to ε gives,

$$\begin{aligned} 0 &= \mathbb{E}_x f'(X^\varepsilon) g(\xi^\varepsilon)\gamma^\varepsilon + \mathbb{E}_x f(X^\varepsilon) \frac{\partial g(\xi^\varepsilon)}{\partial \varepsilon} \frac{1}{Y^\varepsilon}\gamma^\varepsilon \\ &\quad + E_x f(X^\varepsilon) g(\xi^\varepsilon) \frac{-Z^\varepsilon}{(Y^\varepsilon)^2}\gamma^\varepsilon + \mathbb{E}_x f(X^\varepsilon) g(\xi^\varepsilon) \frac{1}{Y^\varepsilon}\frac{\partial \gamma^\varepsilon}{\partial \varepsilon}. \end{aligned}$$

Whence,

$$\begin{aligned} \mathbb{E}_x f'(X^\varepsilon) g(\xi^\varepsilon)\gamma^\varepsilon &= \mathbb{E}_x f(X^\varepsilon)\left[-\frac{\partial g(\xi^\varepsilon)}{\partial \varepsilon}\frac{1}{Y^\varepsilon}\gamma^\varepsilon \right. \\ &\quad \left. + g(\xi^\varepsilon)\frac{Z^\varepsilon}{(Y^\varepsilon)^2}\gamma^\varepsilon - g(\xi^\varepsilon)\frac{1}{Y^\varepsilon}\frac{\partial\gamma^\varepsilon}{\partial\varepsilon}\right]. \quad (26) \end{aligned}$$

Thus, we can replace differentiating of f by applying a certain differential operator to $g(\xi^\varepsilon)\gamma^\varepsilon$. Its result stands in the brackets in (26).

Suppose, we would like to estimate $\mathbb{E}_x f''(X_{N,h})$ by iterating of (26). Use the formula with $g(\xi^\varepsilon)=1$ and f' instead of f to obtain,

$$\mathbb{E}_x f''(X^\varepsilon)\gamma^\varepsilon = \mathbb{E}_x f'(X^\varepsilon)\left[\frac{Z^\varepsilon}{(Y^\varepsilon)^2}\gamma^\varepsilon - \frac{1}{Y^\varepsilon}\frac{\partial\gamma^\varepsilon}{\partial\varepsilon}\right],$$

this is just the relation (22) with f'' and f' before we set $\varepsilon=0$. Apply (26) again with $g(\xi^\varepsilon)$ given by the expression in the brackets. We get,

$$\mathbb{E}_x f''(X^\varepsilon)\gamma^\varepsilon = \mathbb{E}_x f(X^\varepsilon) R(\xi^\varepsilon), \quad (27)$$

where R denotes a sum of rational expressions which has powers of Y^ε only in denominators, and values of processes Z^ε, and $V^\varepsilon = \partial Z^\varepsilon/\partial\varepsilon$, i.e.

$$V^\varepsilon_{n+1,h} = (1+F'(X^\varepsilon_{n,h})h)V^\varepsilon_{n,h} + h\left(3F''(X^\varepsilon_{n,h})Y^\varepsilon_{n,h}Z^\varepsilon_{n,h} + F'''(X^\varepsilon_{n,h})(Y^\varepsilon_{n,h})^3\right),$$

and derivatives $\partial\gamma^\varepsilon/\partial\varepsilon$ and $\partial^2\gamma^\varepsilon/\partial\varepsilon^2$.

Remark 5 We can continue these iterations as long as all the expressions in (26) are differentiable. In order to perform l iterations, F has to be $l+1$ times differentiable with bounded derivatives and we have to estimate $\mathbb{E}_x|(\partial^j\gamma/\partial\varepsilon^j)|_{\varepsilon=0}|$, $1 \le j \le l$.

At the end of the second iteration we set $\varepsilon = 0$. It remains to estimate the right hand side of (27). By the Lemma 4, the values $|V^\varepsilon|$ are bounded, provided F, F', F'' and F''' are. Bounds for the values of processes X, Y and Z have been established earlier. Hence with some $C > 0$ we have $|\mathbb{E}_x f''(X)| \le C \sup_x |f(x)|$, if one could show that

$$\mathbb{E}_x \left(\frac{\partial\gamma^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0} \right)^2 < \infty, \quad \mathbb{E}_x \left| \frac{\partial^2\gamma^\varepsilon}{\partial\varepsilon^2}\Big|_{\varepsilon=0} \right| < \infty.$$

The former inequality is true due to (24). From the definition of γ^ε given by (16) we deduce,

$$\frac{\partial^2\gamma^\varepsilon}{\partial\varepsilon^2}\Big|_{\varepsilon=0} = h\sum_{k=1}^N \frac{p''(\xi_k)p(\xi_k) - (p'(\xi_k))^2}{p^2(\xi_k)} + \left(\sqrt{h} \sum_{k=1}^N \frac{p'(\xi_k)}{p(\xi_k)} \right)^2 .$$

Using independence of ξ_k and (24), one has,

$$\begin{aligned}
\mathbb{E}_x \left| \frac{\partial^2\gamma^\varepsilon}{\partial\varepsilon^2}\Big|_{\varepsilon=0} \right| &\le Nh\mathbb{E}\left| \frac{p''(\xi_1)}{p(\xi_1)} \right| + 2Nh\mathbb{E}\left(\frac{p'(\xi_1)}{p(\xi_1)} \right)^2 \\
&\le T\left(\mathbb{E}\left| \frac{p''(\xi_1)}{p(\xi_1)} \right| + 2\mathbb{E}\left(\frac{p'(\xi_1)}{p(\xi_1)} \right)^2 \right) < \infty,
\end{aligned}$$

if finiteness of $\mathbb{E}|p''(\xi_1)/p(\xi_1)|$ is assumed, i.e. (D_2) holds.

We have arrived to the final result of the section $d = 1$.

Lemma 5 *Let $d = 1$ and $X_{n,h}$ be defined by* (1), *and the basic assumptions* (F_1), (D_1)–(D_3) *hold true. Then there is a constant $C > 0$ such that*

$$\sup_x |\mathbb{E}_x f''(X_{N,h})| \le C \sup_z |f(z)|,$$

for all h small enough and any twice continuously differentiable bounded complex-valued function f.

The previous lemma is sufficient for our purposes for the one-dimensional case, but we discuss also estimates of $\mathbb{E}_x|(\partial^j\gamma^\varepsilon/\partial\varepsilon^j)|_{\varepsilon=0}|$, $2 \le j \le l$, which will be helpful in the case $d > 1$, with $l = d+1$. What follows is a general way of estimation suitable for any j. Differentiating j times of γ^ε can be represented as multiplying γ^ε by $h^{j/2}$ and a multiple sum of terms like

$$\frac{p^{(j_1)}(\xi_{k_1}+\varepsilon\sqrt{h})}{p(\xi_{k_1}+\varepsilon\sqrt{h})}\frac{p^{(j_2)}(\xi_{k_2}+\varepsilon\sqrt{h})}{p(\xi_{k_2}+\varepsilon\sqrt{h})}\dots,$$

where the lower indices k_1, k_2, ... are all different, the total sum $j_1 + j_2 + \ldots = j$, and the total number of summands is N^j. At the end of iterations we set $\varepsilon = 0$ and γ^ε vanishes and the multiple sum and $h^{j/2}$ remain only. After taking expectations a lot of terms in the sum become zeros. Indeed, if a summand contains a first derivative, then we use independence of ξ_k and notice that

$$\mathbb{E}\frac{p'(\xi_k)}{p(\xi_k)} = \int_{\mathbb{R}} p'(x)\,dx = 0,$$

if $p(x) \to 0$ as $x \to \infty$. Hence we may calculate the number of possibly non-zero summands and estimate each of them by

$$\mathbb{E}\left|\frac{p^{(j_1)}(\xi_{k_1})}{p(\xi_{k_1})}\right|\mathbb{E}\left|\frac{p^{(j_2)}(\xi_{k_2})}{p(\xi_{k_2})}\right|\dots \tag{28}$$

We assume that $p^{(j)}(x)$ are absolute integrable for $2 \le j \le l$ and notice that the number of multipliers is at most $[j/2]$. Any expression (28) is bounded by a constant depending on j, while the number of summands of type (28) for large N has the order $N^{j/2}$, at most. So, the total multiple sum can be estimated by $Ch^{j/2}N^{j/2} \le CT^{j/2} < \infty$.

Notice that the last problem is analogous to the well-known fact (cf., e.g., 3.5.15–3.5.16 in [6]) that for i.i.d. random variables η_i with $\mathbb{E}\eta_i = 0$ and $b > 1$,

$$\mathbb{E}\Big|\sum_{i=1}^m \eta_i\Big|^b \le Cm^{b/2}\mathbb{E}|\eta_1|^b,$$

and the proof is, indeed, similar (for integer b). We have obtained the following generalization of Lemma 5.

Lemma 6 *Let $X_{n,h}$ be a stochastic process defined by* (1) *for* $d = 1$. *Assume that for some positive integer l the function F and its derivatives*

F', ..., $F^{(l+1)}$ are bounded on $\mathbb{R}$, the derivatives of $p(x)$ of orders up to l are absolute integrable, and basic assumptions (D_1) and (D_3) are true. Then there is a constant $C > 0$ such that

$$\sup_x |\mathbb{E}_x f^{(l)}(X_{N,h})| \leq C \sup_z |f(z)|$$

for all h small enough and any l times continuously differentiable bounded complex-valued function f.

Case $d > 1$. We will adapt previous computations for the multi-dimensional case.

Let the parameter $\varepsilon \in \mathbb{R}^d$ changes along a line in $\mathbb{R}^d$. It is more convenient then instead of ε write εv, where now $\varepsilon \in \mathbb{R}$ and v is a fixed arbitrary unit vector giving the direction of the line.

The following lemma is an analog of Lemma 4 for the multi-dimensional case. Define a matrix norm by $\|A\| = \sup\{|Av| : v \in \mathbb{R}^d, |v| = 1\}$.

Lemma 7 *Let $h > 0$ be small enough, $0 \leq n \leq N = N(T, h)$, (A_n) a sequence of $d \times d$-matrices and (b_n) a sequence of vectors in $\mathbb{R}^d$. Consider (u_n) as the vector solution to equations,*

$$u_{n+1} = (I + hA_n)u_n + hb_n, \quad u_0 = 0.$$

If $\sup_n \|A_n\| \leq M$ and $\sup_n |b_n| \leq M$, then $|u_n| \leq C$ with some real C which can be chosen uniformly for all h small enough.

If $T < \pi/(2M)$, $\sup_n \|A_n\| \leq M$ and $b_n \equiv v$, $|v| = 1$, then

$$|\langle v, u_N \rangle| \geq c > 0,$$

where the constant c can be chosen uniformly for all h small enough.

Proof The first part is proved in the same way as in Lemma 4 with obvious changes. Let $b_n \equiv v$, $|v| = 1$ and $h < 1/M$. We have by iterations,

$$u_n = h \left(\sum_{j=1}^{n} \prod_{k=j}^{n-1} (I + hA_k) \right) v,$$

where the product over the empty set of indices is set to I. Consider $I + hA_k$ as a linear transformation acting on a vector w. One gets,

$$(1 - Mh)|w| \leq |(I + hA_k)w| \leq (1 + Mh)|w|,$$

and if α is the angle between w and $(I + hA_n)w$, then $\sin\alpha \le Mh$. Since h is small enough, we may find some $\delta > 0$ such that $\alpha \le (1+\delta)Mh$. It implies that the angle between v and any summand in the expression for u_n is at most $(1+\delta)Mnh$. The assertion $T < \pi/(2M)$ allows choosing δ in such a way that $(1+\delta)MNh \le (1+\delta)MT < \pi/2 - \delta$. Consider the projection of $h\left(\prod_{k=j}^{N-1}(I + hA_k)\right)v$ on the line given by the direction of v. It has the same orientation as v and its length is at least $h(1-Mh)^{N-j}\sin\delta$. By arguments from the proof of Lemma 4 the vector u_N has its projection length at least $M^{-1}\left(1-(1-Mh)^N\right)\sin\delta$ and for all h small enough we can take $c = (2M)^{-1}(1-e^{-TM})\sin\delta$. $\square$

For an arbitraty fixed $\lambda \in \mathbb{R}^d$ we write the following identity,

$$\mathbb{E}_x f(\langle\lambda, X\rangle)\frac{1}{\langle\lambda, Y\rangle}g(\xi) = \mathbb{E}_x f(\langle\lambda, X^\varepsilon\rangle)\frac{1}{\langle\lambda, Y^\varepsilon\rangle}g(\xi^\varepsilon)\gamma^\varepsilon,$$

where $f : \mathbb{R} \to \mathbb{C}$ is a bounded function, $g(\xi)$ is an expression which can be computed using values of $\xi_1, \ldots, \xi_N$, $g(\xi^\varepsilon)$ is evaluated in the same way as $g(\xi)$ but using variables $\xi_1 + \varepsilon\sqrt{h}\,v, \ldots, \xi_N + \varepsilon\sqrt{h}\,v$, and $Y^\varepsilon = \partial X^\varepsilon/\partial\varepsilon$, $Y = Y^0$. We are to show that the values in the denominators are bounded away from zero. Achieving this, one takes the vector $v = \lambda/|\lambda|$ and applies Lemma 7 with $A_n = F'(X^\varepsilon_{n,h})$, $b_n = v$. Differentiating with respect to ε gives,

$$\begin{aligned}\mathbb{E}_x f'(\langle\lambda, X^\varepsilon\rangle)g(\xi^\varepsilon)\gamma^\varepsilon \;=\; \mathbb{E}_x f(\langle\lambda, X^\varepsilon\rangle)&\left[\frac{\langle\lambda, Z^\varepsilon\rangle}{\langle\lambda, Y^\varepsilon\rangle^2}g(\xi^\varepsilon)\gamma^\varepsilon - \frac{1}{\langle\lambda, Y^\varepsilon\rangle}\frac{\partial g(\xi^\varepsilon)}{\partial\varepsilon}\gamma^\varepsilon\right.\\ &\left.-\frac{1}{\langle\lambda, Y^\varepsilon\rangle}g(\xi^\varepsilon)\frac{\partial\gamma^\varepsilon}{\partial\varepsilon}\right], \end{aligned} \tag{29}$$

where $Z^\varepsilon = \partial Y^\varepsilon/\partial\varepsilon$. Formula (29) is a clear analog of (26) suitable for iterations as pointed out in Remark 5. We start with $g(\xi^\varepsilon) = 1$ and $f^{(l-1)}$ instead of f and perform l iterations letting $g(\xi^\varepsilon)$ equal to expressions in brackets for the next iterations. Setting $\varepsilon = 0$ in the end, we get,

$$\mathbb{E}_x f^{(l)}(\langle\lambda, X\rangle) = \mathbb{E}_x f(\langle\lambda, X\rangle)R(\xi), \tag{30}$$

where R denotes a sum of rational expressions which has powers of $\langle\lambda, Y^\varepsilon\rangle$ only in denominators, includes also values of processes $\partial^j X^\varepsilon/\partial\varepsilon^j$ and derivatives $\partial^j\gamma^\varepsilon/\partial\varepsilon^j$ for $1 \le j \le l$. Substitute λ by $|\lambda|v$ in (30) and take out $|\lambda|$ from scalar products. By Lemma 7 terms $\langle v, Y^\varepsilon\rangle$, $\langle v, Z^\varepsilon\rangle$, $\langle v, V^\varepsilon\rangle$, $\ldots$ are bounded provided function F has $l+1$ bounded derivatives. In total, we obtain $|\lambda|^l$ in the denominator similarly to the one-dimensional case adapted for characteristic functions.

It remains to prove that the values $\mathbb{E}|(\partial^j\gamma^\varepsilon/\partial\varepsilon^j)|_{\varepsilon=0}$ are finite for all $1 \le j \le l$. We write,

$$\frac{\partial\gamma^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0} = \sqrt{h}\sum_{k=1}^{N}\frac{\langle p'(\xi_k), v\rangle}{p(\xi_k)},$$

and similarly to the one-dimensional case

$$\left(\mathbb{E}\left|\frac{\partial\gamma^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0}\right|\right)^2 \le \mathbb{E}\left(\frac{\partial\gamma^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0}\right)^2 \le T\mathbb{E}\left(\frac{|p'(\xi_1)|}{p(\xi_1)}\right)^2.$$

Finally, we employ our argumentations having led to Lemma 6 and get the following result.

Lemma 8 *Let $X_{n,h}$ be a stochastic process defined by* (1)*. Assume that basic assumptions (F_1), (D_1)–(D_3) are true and $l \le d+1$. Then there is a constant $C > 0$ such that*

$$\sup_x |\mathbb{E}_x f^{(l)}(\langle\lambda, X_{N,h}\rangle)| \le C|\lambda|^{-l}\sup_z |f(z)|$$

for all h small enough and any l times continuously differentiable bounded complex-valued function f.

We have established the desired estimate for characteristic functions of $X_{N,h}$. One lets $f(z) = e^{iz}$ and obtains that characteristic functions of $X_{N,h}$ decrease in absolute value as $C/|\lambda|^{d+1}$ as $\lambda \to \infty$, Lemma 3 is applicable and the coupling method works uniformly for all h small enough.

4 Recurrence conditions

In [7] two recurrence conditions were introduced separately for discrete and continuous cases. The condition for the latter case was (F_2). The recurrence condition for the discrete case (see also [3]) can be generalized naturally,

$$|x + F(x)h| \le \begin{cases} C_0, & |x| \le R_0, \\ |x|\left(1 - rh/|x|^{1+p}\right), & |x| > R_0, \end{cases} \qquad (31)$$

with some positive constants r, R_0 and C_0. In both cases there is an "attraction" to the origin outside the ball $\{x : |x| \le R_0\}$. The parameters $0 \le p \le 1$ and $r > 0$ regulate the force of the attraction. Condition (F_2) is more simple and hence, natural; however, (31) was used in all proofs in [7, 3] which will be employed in the last section.

Lemma 9 *If* (31) *is true for all h small enough, then* (F_2) *holds with the same constants r and R_0.*

If (F_2) *holds and F is bounded, then for every $\varepsilon > 0$ there exists $h_0 > 0$ such that* (31) *is satisfied with $r - \varepsilon$ instead of r.*

Proof Inequalities

$$|x + F(x)h| \le |x|(1 - rh/|x|^{1+p})$$

and

$$\langle x, F(x)\rangle \le -r|x|^{1-p} + \frac{1}{2}(r^2|x|^{-2p} - |F(x)|^2)h$$

are equivalent.

In the first case we pass to the limit as $h \to 0$ and get (F_2).

In the second case denote $M = \sup |F(x)|$ and choose $h_0 \in (0, 2\varepsilon/M^2]$. Then for every $h \in (0, h_0]$ we get,

$$\langle x, F(x)\rangle \le -(r-\varepsilon) - \varepsilon \le -(r-\varepsilon) + \left(\frac{(r-\varepsilon)^2}{2|x|^{2p}} - \frac{|F(x)|^2}{2}\right) h.$$

The equivalence above shows that condition (31) is true with $r - \varepsilon$ instead of r. □

5 Proof of Theorem 1

Proof We will show how the proofs from [7, 3, 2] can be modified to obtain our main results. Below the scheme of the method is briefly described and some detailed calculus is given.

Let $R > 0$ and $\tau_h = \tau_h(R) = \inf\{nh : |X_{n,h}| \le R,\ n \ge 1\}$ is the hitting (macro-) time of the ball $B = B(0, R)$.

If the process $(X_{n,h})$ were inside B permanently, the local Doeblin condition together with the coupling method would provide exponential rates of convergence and mixing [1]. But it is not true for our model because

$(X_{n,h})$ may have long excursions away from the origin where the local Doeblin condition is not satisfied and mixing may be very slow or absent at all. Suppose, we have shown that, say, $\mathbb{E}_x\tau_h^k \le C(1+|x|^m)$ for some $k>1$. It means the probability $\mathbb{P}(\tau_h > t)$ decreases at a polynomial rate in t and the process does visit B from time to time. Polynomial tails of the distribution of τ_h imply the excursions are not very long. Polynomial rates can be achieved as a superposition of this and exponential mixing inside B. Finite sub-exponential or exponential moments of τ_h guarantee even more regular visits to B and, therefore, higher (sub-exponential or exponential) rates of convergence and mixing.

Although we have formulated results for three different kinds of rates let us focus on one of them. We choose the polynomial case as in [7] for illustrating. Two others are treated similarly.

The first required estimate is

$$\begin{aligned}\mathbb{E}_x|X_{n+1,h}|^m\mathbf{1}((n+1)h<\tau_h) \;\le\; & \mathbb{E}_x|X_{n,h}|^m\mathbf{1}(nh<\tau_h)\\ & - ch\mathbb{E}_x|X_{n,h}|^{m-2}\mathbf{1}(nh<\tau_h) \qquad (32)\end{aligned}$$

with $m>2$ and some $c>0$ which does not depend on h. The function in the last term (here it is $|x|^{m-2}$; in other cases they are different) must tend to infinity as $|x|\to\infty$, be finite in B and bounded away from zero outside B. The inequality implies $\mathbb{E}_x|X_{n,h}|^m\mathbf{1}(nh<\tau_h)\le 1+|x|^m$ by iterations and $\mathbb{E}_x\tau_h < C(1+|x|^m)$ by summing over n. The multiplier h cancels out together with n and gives correct bounds on the scale of macro-time $t=nh$ where τ_h were defined. This will be an analog of Lemma 1 in [7].

The next step is done by deducing from (32) that $\mathbb{E}_x\tau_h^k < C(1+|x|^m)$. This is done exactly in the same way as in Lemma 3 in [7], i.e. through multiplying (32) by some polynomials and summing over n.

Further, we consider two independent copies of our process, $(X_{n,h})$ and $(X'_{n,h})$, and the stopping time $\gamma_h = \inf\{nh : |X_{n,h}|\vee|X'_{n,h}|\le\widetilde{R},\ n\ge 1\}$. We prove an analog of (32),

$$\begin{aligned}&\mathbb{E}_{x,x'}(|X_{n+1,h}|^m+|X'_{n+1,h}|^m)\mathbf{1}((n+1)h<\gamma_h)\\ &\quad\le\; \mathbb{E}_{x,x'}(|X_{n,h}|^m+|X'_{n,h}|^m)\mathbf{1}(nh<\gamma_h)\\ &\qquad - \widetilde{c}h\mathbb{E}_{x,x'}(|X_{n,h}|^{m-2}+|X'_{n,h}|^{m-2})\mathbf{1}(nh<\gamma_h), \qquad (33)\end{aligned}$$

with $m>2$ and some $\widetilde{c}>0$. In the same way as for τ_h (see [7, Lemmas 4 and 5] also) we prove that $\mathbb{E}_{x,x'}\gamma_h^k < C(1+|x|^m+|x'|^m)$.

Finally, exactly identical calculations as in Theorems 1–3 in [7] are used to obtain polynomial bounds for convergence and β-mixing.

Thus, relation (32) has to be checked only, and it is done in the next lemma. $\square$

Lemma 10 *Let* $r_m = \mathbb{E}|\xi_1|^m < \infty$*,* $m \geq 2$*, and condition* (31) *hold with* $r > (m-1)r_2/2$*. Then* $R \geq R_0$ *can be chosen in such a way that for* $|x| > R$

$$\mathbb{E}_x|X_{n+1,h}|^m \mathbf{1}((n+1)h < \tau_h) \leq \mathbb{E}_x(|X_{n,h}|^m - ch|X_{n,h}|^{m-2})\mathbf{1}(nh < \tau_h)$$

with some $c > 0$.

Proof We provide this fairly simple proof, indeed, for the reader's convenience, as well as for completeness of the text. Let $u, v \in \mathbb{R}^d$, $t \in \mathbb{R}$, $m > 2$, and the real-valued function $f(t) = |u + tv|^m = \langle u + tv, u + tv\rangle^{m/2}$. Write the Taylor's formula for $f(t)$ and substitute $t = 1$:

$$\begin{aligned} |u+v|^m &= |u|^m + m|u|^{m-2}\langle u, v\rangle \\ &\quad + \frac{m(m-2)}{2}|u+sv|^{m-4}\langle u+sv, v\rangle^2 + \frac{m}{2}|u+sv|^{m-2}|v|^2, \end{aligned} \tag{34}$$

where $s \in [0, 1]$. Since $\langle u + sv, v\rangle^2 \leq |u + sv|^2|v|^2$, the sum of two last terms do not exceed $m(m-1)|u + sv|^{m-2}|v|^2/2$.

For each $m > 0$ and $\varepsilon > 0$ there are a positive constant $C = C(\varepsilon, m)$ such that

$$(x+y)^m \leq (1+\varepsilon)x^m + Cy^m, \quad x, y \geq 0. \tag{35}$$

From (34) and (35) it follows that

$$|u+sv|^{m-2} \leq (|u| + |v|)^{m-2} \leq (1+\varepsilon)|u|^{m-2} + C|v|^{m-2}$$

and

$$\begin{aligned} |u+v|^m &\leq |u|^m + m|u|^{m-2}\langle u, v\rangle \\ &\quad + \frac{m(m-1)}{2}\Big((1+\varepsilon)|u|^{m-2}|v|^2 + C|v|^m\Big). \end{aligned} \tag{36}$$

If $m = 2$, then the calculations are much simpler,

$$|u+v|^2 = |u|^2 + 2\langle u, v\rangle + |v|^2,$$

and one takes $\varepsilon = 0$ and $C = 0$ to satisfy (36).

We start deducing the desired inequality. For the sake of briefness denote $X = X_{n,h}$, $g = X_{n,h} + F(X_{n,h})h$ and $V = \sqrt{h}\,\xi_{n+1}$, then $X_{n+1,h} = g + V$. Substitute in (36) $u = g$, $v = V$ and write the estimate,

$$\begin{aligned}
\mathbb{E}\{|X_{n+1,h}|^m \mathbf{1}((n+1)h < \tau_h) \mid \mathcal{F}_n\} & \\
&\le \mathbf{1}(nh < \tau_h)\mathbb{E}\{|g+V|^m \mid \mathcal{F}_n\} \\
&\le \mathbf{1}(nh < \tau_h)\Big(|g|^m + m|g|^{m-2}\mathbb{E}\{\langle g, V\rangle \mid \mathcal{F}_n\} \\
&\qquad + (1+\varepsilon)\frac{m(m-1)}{2}|g|^{m-2}\mathbb{E}\{|V|^2 \mid \mathcal{F}_n\} \\
&\qquad + C\mathbb{E}\{|V|^m \mid \mathcal{F}_n\}\Big),
\end{aligned}$$

where $\mathcal{F}_n = \sigma\{X_{k,h} : k \le n\}$. Due to independence of σ-fields $\sigma\{\xi_{n+1}\}$ and $\mathcal{F}_n$ the summand with the scalar product contributes nothing. Using restrictions on the noise and the recurrent condition ($|g| \le |X|$ as $nh < \tau_h$), one concludes that the right hand side of the last inequality does not exceed

$$\mathbf{1}(nh < \tau_h)\left(|g|^m + \frac{(1+\varepsilon)m(m-1)r_2 h}{2}|X|^{m-2} + Cr_m h^{m/2}\right). \tag{37}$$

Assume the condition $nh < \tau_h$ is true in computations below. It implies $|X| > R$. For $m > 2$ we achieve $Cr_m h^{m/2} \le (\varepsilon/2)m(m-1)r_2 h|X|^{m-2}$, increasing R or regarding only small enough values of h, if needed. For $m = 2$ the constants ε and C are zero, and the inequality is true also. Further,

$$|g|^m \le |X|^m\left(1 - \frac{rh}{|X|^2}\right)^m \le |X|^m\left(1 - \frac{mr'h}{|X|^2}\right),$$

where $r' < r$, but it may be chosen arbitrary close to r by increasing R, hence expression (37) is estimated from above by

$$|X|^m\mathbf{1}(nh < \tau_h) + h|X|^{m-2}\left(\frac{(1+2\varepsilon)r_2 m(m-1)}{2} - mr'\right)\mathbf{1}(nh < \tau_h).$$

One deduces,

$$\begin{aligned}
\mathbb{E}_x|g+V|^m\mathbf{1}(nh < \tau_h) &\le \mathbb{E}_x|X|^m\mathbf{1}(nh < \tau_h) - h\mathbb{E}_x|X|^{m-2}\mathbf{1}(nh < \tau_h) \\
&\quad \times \left(mr' - \frac{(1+2\varepsilon)r_2 m(m-1)}{2}\right).
\end{aligned}$$

Since $r > (m-1)r_2/2$, values ε and r' can be chosen in such a way that the difference in brackets is positive. Denote it by c. The required inequality is proved. □

Acknowledgement

Both authors wish to thank for financial support the research grants EP-SRC GR/R40746/01 (UK) and NFGRF 2301863 (University of Kansas at Lawrence, USA); the second author also thanks the grants INTAS-99-0590 and RFBR-00-01-22000.

References

[1] Doob, J. L., *Stochastic processes.* (John Wiley & Sons, Inc., New York; Chapman & Hall, Limited, London, 1953) viii+654 p.

[2] Gulinskiĭ, O. V., Veretennikov, A. Yu. *The mixing rate and the averaging principle for recursive stochastic procedures.* Avtomat. i Telemekh. (Russian) no. 6, (1990), 68–78; translation in Automat. Remote Control **51**, (1990), no. 6, part 1, 779–788.

[3] Klokov, S. A., Veretennikov, A. Yu. *Sub-exponential mixing rate for Markov processes.* Submitted

[4] Meyn, S. P., Tweedie, R. L. *Markov Chains and Stochastic Stability.* (Berlin, Springer-Verlag, 1993) 550 p.

[5] Nummelin, E. *General irreducible Markov chains and non-negative operators.* (Cambridge, University Press, 1984) xi,156 p.

[6] Petrov, V. V. *Summy nezavisimyh sluchajnyh velichin.* (Nauka, Moscow, 1972); translation: *Sums of independent random variables.* (Berlin, New York, Springer-Verlag, 1975) x, 345 p.

[7] Veretennikov, A. Yu. *On polynomial mixing and the rate of convergence for stochastic differential and difference equations.* Teor. Veroyatnost. i Primenen. (Russian) **44** (1999), no. 2, 312–327; translation in Theory Probab. Appl. **44** (2000), no. 2, 361–374.

[8] Veretennikov, A. Yu. *On polynomial mixing bounds for stochastic differential equations.* Stochastic Process. Appl. **70** (1997), no. 1, 115–127.

[9] Veretennikov, A. Yu. *Estimates of the mixing rate for stochastic equations.* Teor. Veroyatnost. i Primenen. (Russian) **32** (1987), no. 2, 299–308; translation in Theory Probab. Appl. **32** (1987), no. 2, 273–281.

[10] Veretennikov, A. Yu. *A probabilistic approach to hypoellipticity.* Russian Math. Surveys **38**, vol. 3, 1983, 127–140.

[11] Veretennikov, A. Yu. *Coupling method for Markov chains under integral Doeblin type condition.*
`www.mathpreprints.com/math/Preprint/veretenn/20010723/1/`