Bernoulli **8**(3), 2002, 367–405

Recursive computation of the invariant distribution of a diffusion

DAMIEN LAMBERTON[1] and GILLES PAGÈS[2]

[1]*Laboratoire d'Analyse et de Mathématiques Appliquées, UMR 8050, Université de Marne-la-Vallée, Cité Descartes, 5 boulevard Descartes, Champs-sur-Marne, F-77454 Marne-la-Vallée Cedex 2, France. E-mail: dlamb@math.univ-mlv.fr*
[2]*Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, Université Pierre et Marie Curie, case 188, 4 place Jussieu, F-75252 Paris Cedex 05, France. E-mail: gpa@ccr.jussieu.fr*

We propose a recursive stochastic algorithm with decreasing step to compute the invariant distribution ν of a Brownian diffusion process, in which we approximate $\nu(f)$ for a wide class of possibly unbounded continuous functions f. We consider a somewhat general setting which includes cases where the diffusion may have several invariant distributions. Our main convergence result contains as a corollary the almost sure central limit theorem. Further, we investigate the weak rate of convergence of the algorithm. We show, in the class of polynomial steps $\gamma_n = n^{-\alpha}$, that it can be at most $n^{1/3}$ when the white noise has third moment zero and $n^{1/4}$ otherwise, where n denotes the number of iterations of the algorithm.

Keywords: almost sure central limit theorem; central limit theorem; diffusion process; invariant distribution; numerical probability; stochastic algorithm

1. Introduction

The aim of this paper is to propose and study a stochastic recursive algorithm for the computation of the invariant distribution ν of a Brownian diffusion process

$$dY_t = b(Y_t)dt + \sigma(Y_t)dW_t, \tag{1}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ is a continuous vector field, σ is continuous on $\mathbb{R}^d$ with values in the set $\mathcal{M}(d \times q)$ of matrices with d rows and q columns, and W is a q-dimensional Brownian motion.

Invariant distributions are crucial in the study of the long-run behaviour of diffusions. We refer the reader to Has'minskii (1980) and Ethier and Kurtz (1986) for background on the stability of stochastic differential systems. In general, the functions b and σ are given by some physical model, and a numerical procedure for the computation of the invariant distribution is needed. A typical situation of this kind is given by randomly perturbed mechanical systems (see Soize 1995). Talay (1990; 2002) was the first to design and analyse such a procedure, and we will briefly discuss below the difference between his approach and ours.

In order to motivate our algorithm, recall that if the diffusion is stationary and ergodic with invariant distribution ν, then, for every ν-integrable function f,

$$\lim_{t\to+\infty}\frac{1}{t}\int_0^t f(Y_s)\mathrm{d}s = \int_{\mathbb{R}^d} f\,\mathrm{d}\nu = \nu(f) \text{ almost surely.}$$

This suggests the use of the average of f along the path of the diffusion as a proxy for $\nu(f)$. However, since the exact simulation of the diffusion Y itself cannot be achieved, we first discretize the diffusion and then compute a discrete analogue of the average. Therefore, we propose an algorithm consisting of two successive phases.

Phase I: Discretization. Compute the Euler discretization of (1) with a *step* γ_n vanishing to 0, i.e.

$$\forall n\in\mathbb{N},\qquad X_{n+1} = X_n + \gamma_{n+1}b(X_n) + \sqrt{\gamma_{n+1}}\sigma(X_n)U_{n+1}, \tag{2}$$

where $X_0 \in L^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P})$ and $(U_n)_{n\in\mathbb{N}^*}$ is an $\mathbb{R}^q$-valued white noise defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, independent of X_0. The *step* sequence $\gamma := (\gamma_n)_{n\in\mathbb{N}}$ satisfies the condition

$$\forall n\in\mathbb{N}^*,\qquad \gamma_n \geqslant 0,\quad \lim_n \gamma_n = 0 \quad\text{and}\quad \lim_n \Gamma_n = +\infty,\qquad \text{where } \Gamma_n := \sum_{k=1}^n \gamma_k. \tag{3}$$

Choosing a vanishing discretization step allows for a more accurate approximation of the diffusion as time goes to infinity.

Phase II: Averaging. Form a weighted empirical measure with the X_n using a weight sequence $\eta := (\eta_n)_{n\in\mathbb{N}^*}$ satisfying the condition

$$\forall n\in\mathbb{N}^*,\qquad \eta_n \geqslant 0 \quad\text{and}\quad \lim_n H_n = +\infty,\qquad \text{where } H_n := \eta_1 + \ldots + \eta_n. \tag{4}$$

Let δ_x denote the Dirac mass at x. For every $n \geqslant 1$ and every $\omega\in\Omega$, set

$$\nu_n^\eta(\omega, \mathrm{d}x) := \frac{\eta_1\delta_{X_0(\omega)} + \ldots + \eta_{k+1}\delta_{X_k(\omega)} + \ldots + \eta_n\delta_{X_{n-1}(\omega)}}{\eta_1 + \ldots + \eta_n}, \tag{5}$$

and use $\nu_n^\eta(\omega, f)$ to approximate $\nu(f)$.

For numerical purposes, it is observed that, for a fixed function f, $\nu_n^\eta(\omega, f)$ can be recursively computed as follows:

$$\nu_{n+1}^\eta(\omega, f) = \nu_n^\eta(\omega, f) + \tilde{\eta}_{n+1}(f(X_n(\omega)) - \nu_n^\eta(\omega, f)) \qquad \text{with } \tilde{\eta}_{n+1} := \frac{\eta_{n+1}}{H_{n+1}}. \tag{6}$$

This recursive form shows that $\nu_n^\eta(f)$ is given by a linear stochastic algorithm with step $\tilde{\eta}_{n+1}$ and pseudo-innovation $f(X_n)$ (at time $n+1$).

Our general almost surely weak convergence result for $\nu_n^\eta(\omega, \mathrm{d}x)$ is Theorem 3, stated in Section 2. Since the full statement of Theorem 3 is somewhat technical, we state a slightly simpler version (with $\eta_n = \gamma_n$) usually sufficient for practical purposes.

Theorem 1. *Assume that there is a* $\mathcal{C}^2$ *function* $V: \mathbb{R}^d \to [v_*, +\infty)$, $v_* > 0$, *satisfying the following conditions:*

$$\|D^2 V\|_\infty := \sup_{x \in \mathbb{R}^d} \|D^2 V(x)\| < +\infty, \quad \textit{and} \quad \lim_{|x| \to +\infty} V(x) = +\infty; \tag{7a}$$

$$|\nabla V|^2 + |b|^2 \leqslant c_V V \quad \textit{for some } c_V > 0, \quad \textit{and} \quad \mathrm{Tr}(\sigma\sigma^*)(x) = o(V(x)) \quad \textit{as } |x| \to +\infty; \tag{7b}$$

$$(\nabla V | b) \leqslant -\alpha V + \beta, \qquad \textit{for some } \alpha \in \mathbb{R}_+^*, \beta \in \mathbb{R}. \tag{7c}$$

If the white noise U_1 *has moments at any order and if* (1) *has a unique invariant measure* ν, *then, for every continuous function* f *satisfying* $f(x) = o(V^k(x))$ *for some* $k \in \mathbb{N}$, $\lim_n \nu_n^\gamma(f) = \nu(f)$ *a.s.*

Assumption (7b) is a no-explosion condition for the diffusion process, whereas (7c) is a stability condition that guarantees the existence of (at least) one invariant distribution. Theorem 3 covers more general situations of interest, in particular the cases $\eta_n \neq \gamma_n$, $\mathrm{Tr}(\sigma\sigma^*) \asymp V$ or U_1 with finitely many finite moments. This theorem also has some interesting theoretical consequences. One of these is that, applied to the Ornstein–Uhlenbeck process with $\gamma_n = \eta_n := 1/n$, it yields the celebrated almost sure central limit theorem for sequences of independent and identically distributed (i.i.d.) random variables (see Brosamler 1988; Lacey and Philip 1990; Berkes and Csáki 2001). This fact is established in Section 5.

In Section 6, we analyse the rate of convergence of our algorithm when the invariant distribution ν is unique. More precisely, we study $\nu_n^\eta(\omega, f)$ for test functions f of the form $f = A\varphi$, where A stands for the infinitesimal generator of (1). Although the choice $\eta = \gamma$ turns out to be optimal in many cases, it is worth mentioning that the results on the rate of convergence for $\nu_n^\gamma(\omega, f)$ rely on the almost sure convergence of auxiliary weighted sums with $\eta \neq \gamma$. Our study shows that the normalized error may weakly converge towards either a Dirac mass or a centred Gaussian measure. Actually, both types of limit can be observed for a given rate, depending on the choice of step sequence (see the beginning of Section 6 for details).

In order to put our results in the context of the recent literature on invariant distribution approximation, let us first recall that the rate of convergence of the semigroup $P_t(x, \mathrm{d}y)$ towards ν has been thoroughly investigated (see Ganidis *et al*. 1999). Actually, it may converge quite fast to ν (in variation) so that $P_t(x, \mathrm{d}y) \approx \nu$ for reasonably large t. Since P_t is generally not explicitly known, it seems natural to approximate P_t by $\mathcal{L}(\overline{Y}_n^h)$, $nh \approx t$, where $\overline{Y}^h$ denotes a discretization scheme with constant step $h > 0$, e.g. the Euler scheme given by

$$\overline{Y}_{n+1}^h = \overline{Y}_n^h + hb(\overline{Y}_n^h) + \sqrt{h}\sigma(\overline{Y}_n^h)U_{n+1}.$$

Finally, one estimates $\mathcal{L}(\overline{Y}_n^h)$ using N trials of a Monte Carlo simulation. This method induces three different errors (d_w denotes a distance for the weak topology): $d_\mathrm{w}(P_t(x, \mathrm{d}y), \nu)$, $d_\mathrm{w}(\mathcal{L}(\overline{Y}_{nh}^h), P_t(x, \mathrm{d}y))$ and the $O(N^{-1/2})$ induced by the MC simulation.

One way to improve on this first approach is to use the ergodic properties of the diffusion (if any). Thus, inferring that, for a small enough step h, the discretization scheme $(\overline{Y}^h_n)_{n\in\mathbb{N}^*}$ is an ergodic Markov chain with invariant distribution ν_h, one has

$$\mathbb{P}\text{-a.s.} \qquad \frac{1}{n}\sum_{k=1}^{n}\delta_{\overline{Y}^h_k} \text{ weakly converges to } \nu^h. \tag{8}$$

In this method, there are only two sources of error: that in (8), which can generally be controlled by a central limit theorem; and that in $d_{\mathrm{w}}(\nu^h, \nu)$, which can be estimated along smooth enough functions f. This method was first designed and studied by Talay (1990). He obtains some explicit rates for $d_{\mathrm{w}}(\nu^h, \nu)$ (but not for the convergence in (8)).

Another alternative is to get rid of the asymptotic methodological error due to the step h of the discretization scheme. To this end one may consider a discretization scheme with a vanishing step. This idea was investigated by Basak *et al*. (1997) and also appears in Pelletier (1998, Theorem 7): it is established, under suitable assumptions, that the recursive sequence $(X_n)_{n\in\mathbb{N}}$ defined by (2) converges weakly towards the unique invariant distribution ν of the diffusion. Here the two sources of error are $d_{\mathrm{w}}(X_n, \nu)$ and the Monte Carlo fluctuations.

Our algorithm combines the advantages of both approaches: the computation of $\nu^\eta_n(f)$ is recursive as in (8) and almost surely does converge towards its true target $\nu(f)$ (at least for a wide class of unbounded continuous functions f). From a technical point of view, we obtain this result under less stringent assumptions on the underlying diffusion: even the uniqueness of the invariant distribution ν is not crucial. Note, however, that Talay (1990) and Basak *et al*. (1997) require strong ergodicity assumptions on the diffusion.

There is a formal analogy between the recursive procedure defined by (2) and (5) and that designed in Fort and Pagès (1998) (see also Pagès 2001) in the framework of regular stochastic approximation. However, in Fort and Pagès (1998) the noise is essentially vanishing so that the asymptotics of the algorithm is described by an ordinary differential equation.

Like any numerical method for computing the invariant distribution of a diffusion, our algorithm is an alternative to partial differential equation methods for solving the stationary Fokker–Planck equation: any invariant distribution ν solves $A^*\mu = 0,\ \mu \geqslant 0,\ \mu(\mathbb{R}^d) = 1$, where A^* denotes the adjoint of the infinitesimal generator of the diffusion. Probabilistic methods are especially efficient in higher-dimensional settings ($d \geqslant 3$) or when A is degenerate.

Let $\mathcal{M}(d \times q, \mathbb{R})$ denote the set of matrices having d rows and q columns. For every $A \in \mathcal{M}(d \times q, \mathbb{R})$, let $A^* \in \mathcal{M}(q \times d, \mathbb{R})$ denote its transpose matrix. For every $d \times q$ matrix, let $\|A\|$ denote the related operator norm defined by $\|A\| := \sup_{|x|=1}|Ax|$.

Let $\mathcal{S}_+(d, \mathbb{R})$ denote the set of symmetric non-negative $d \times d$ matrices. For $S \in \mathcal{S}_+(d, \mathbb{R})$, $\|S\| = \sup_{|x|=1}(x|Sx) = \max_i\{\lambda_i\}$, where λ_i is the ith eigenvalue of S. If S, $S' \in \mathcal{S}(d, \mathbb{R})$, then $S \leqslant S'$ if $S' - S \in \mathcal{S}_+(d, \mathbb{R})$. For every $S \in \mathcal{S}(d, \mathbb{R})$, $\lambda_S := \max(\{\lambda_i\}, 0)$. The real λ_S is the lowest real number satisfying

$$\forall y \in \mathbb{R}^d, \qquad y^*Sy \leqslant \lambda_S|y|^2.$$

One may readily verify that $\lambda_S \leqslant \|S\|$ and that, if $S \leqslant S'$, then $\lambda_S \leqslant \lambda_{S'}$. By extension, if S denotes a *mapping* from $\mathbb{R}^d$ into $\mathcal{S}(d, \mathbb{R})$, then

$$\lambda_S := \sup_{x\in\mathbb{R}^d} \lambda_{S(x)}.$$

For every non-negative symmetric matrix $S \in \mathcal{S}_+(d, \mathbb{R})$, the trace operator defined by $\mathrm{Tr}(S) = \sum_{i=1}^d S_{ii}$ satisfies the inequality $\mathrm{Tr}(S) \leqslant d\|S\| \leqslant d\,\mathrm{Tr}(S)$. For every $A \in \mathcal{M}(d \times q, \mathbb{R})$, $\mathrm{Tr}(A^* A) = \mathrm{Tr}(AA^*)$ is simply the square of the Euclidean norm of A viewed as a vector in $\mathbb{R}^{d\times q}$.

We will make extensive use of the Chow theorem (see Hall and Heyde 1980) for martingale increments. This yields, under suitable assumptions, the almost sure convergence (in $\mathbb{R}^d$) of a martingale having possibly non-square integrable increments.

Theorem 2 (Chow). *Let $(M_n)_{n\in\mathbb{N}^*}$ be a martingale with respect to some filtration $\underline{\mathcal{F}} = (\mathcal{F}_n)_{n\in\mathbb{N}}$. Then*

$$\forall\rho \in (0, 1], \qquad M_n \overset{n\to+\infty}{\longrightarrow} M_\infty \in \mathbb{R} \text{ a.s. on the event } \left\{\sum_{n\geqslant 1} \mathbb{E}(|\Delta M_n|^{1+\rho}/\mathcal{F}_{n-1}) < +\infty\right\},$$

where $\Delta M_n := M_n - M_{n-1}$.

Most of the time, this theorem will be combined with the Kronecker lemma for series (Neveu 1972):

Lemma 1 (Kronecker). *Let $(a_n)_{n\in\mathbb{N}^*}$ and $(b_n)_{n\in\mathbb{N}^*}$ be two sequences of real numbers. If $(b_n)_{n\in\mathbb{N}^*}$ is non-decreasing, positive, with $\lim_n b_n = +\infty$, and $\sum_{n\geqslant 1} a_n/b_n$ converges in $\mathbb{R}$, then*

$$\lim_n \frac{1}{b_n}\sum_{k=1}^n a_k = 0.$$

The rest of this paper is organized as follows. In Section 2, the main almost sure weak convergence theorem is stated in full generality. The step and weight conditions are (successfully) tested in some classical settings ($\gamma_n = n^{-\alpha}$ or $\gamma_n := \ln^{-\alpha} n$, $\eta_n = n^{-\beta}$). Section 3 deals with the almost sure tightness of the empirical (random) measures $\nu_n^\eta(\omega, \mathrm{d}x)$, whereas Section 4 is devoted to the identification of the limit. Section 5 points out the link with the almost sure central limit theorem. Section 6 deals with the weak rate of convergence of ν_n^η towards the invariant measure ν. This leads to optimization of the choice of steps according to whether the third moment $\mathbb{E}(U_1^{\otimes 3})$ is 0 or not. Finally, some recommendations and numerical illustrations are given in Section 7.

2. The main a.s. weak convergence results

The main assumption will be the existence of a *Lyapunov* function for the diffusion.

Assumption $(\mathcal{L}_{V,p})$. There exists a $\mathcal{C}^2$ *function* $V : \mathbb{R}^d \to [v_*, +\infty)$, $v_* > 0$*, satisfying, for some* $p \geqslant 1$,

(*i*) $\|D^2V\|_\infty := \sup_{x\in\mathbb{R}^d}\|D^2V(x)\| < +\infty$ *and* $\lim V_{|x|\to+\infty}(x) = +\infty$,
(*ii*) $|\nabla V|^2 + |b|^2 + \mathrm{Tr}(\sigma\sigma^*) \leqslant c_V V$,
(*iii*) $(\nabla V|b) + \lambda_p \mathrm{Tr}(\sigma\sigma^*) \leqslant -\alpha V + \beta$, *for some* $\alpha \in \mathbb{R}_+^*$ *and* $\beta \in \mathbb{R}$,

where $\lambda_p := \frac{1}{2}\lambda_{D^2V+(p-1)}(\nabla V \otimes \nabla V)/V$.

Conditions (7) in Theorem 1 will be referred to as $(\mathcal{L}_{V,\infty})$. We clearly have $(\mathcal{L}_{V,\infty}) \Rightarrow (\mathcal{L}_{V,p})$ for all $p \geqslant 1$. More generally, since λ_p is a non-decreasing function of p, for $p' \geqslant p \geqslant 1$, $(\mathcal{L}_{V,p'}) \Rightarrow (\mathcal{L}_{V,p})$. Further comments on these assumptions will be made in Section 2.1.

Concerning the white noise, we assume that $(U_n)_{n\in\mathbb{N}^*}$ is a sequence of i.i.d. $\mathbb{R}^q$-valued random vectors, independent of X_0, satisfying $\mathbb{E}|U_1|^2 < +\infty$, $\mathbb{E}(U_1) = 0$ and $\Sigma_{U_1} := [\mathrm{cov}(U_1^i, U_1^j)] = Id_q$. Our results are likely to remain valid for suitably dependent noise, but for the applications that we have in mind dependent noise seems irrelevant.

Henceforth, the filtration $\underline{\mathcal{F}} := (\mathcal{F}_n)_{n\in\mathbb{N}}$ is given by $\mathcal{F}_n := \sigma(X_0, U_1, \ldots, U_n)$, $n \geqslant 0$. Actually, we will also assume that X_0 is deterministic. This entails no loss of generality for our main results, which can be obtained in the non-deterministic case by conditioning on X_0.

We now introduce a condition on the steps γ_n and weights η_n as follows:

$$\sum_{n\geqslant 1}\frac{1}{H_n}\left(\Delta\frac{\eta_n}{\gamma_n}\right)_+ < +\infty \quad \text{and} \quad \lim_{n\to\infty}\frac{1}{H_n}\sum_{k=1}^{n}\left|\Delta\frac{\eta_k}{\gamma_k}\right| = 0. \tag{9}$$

Condition (9) is clearly satisfied if the sequence $(\eta_n/\gamma_n)_{n\in\mathbb{N}}$ is non-increasing.

Theorem 3. *Let* $p \in [1, +\infty)$. *Assume* $(\mathcal{L}_{V,p})$ *and* $\mathbb{E}|U_1|^{2p} < \infty$. *Let* $\rho \in (0, 1]$. *Assume that* γ *and* η *satisfy* (9) *and*

$$\sum_{n\geqslant 1}\left(\frac{\eta_n}{H_n\sqrt{\gamma_n}}\right)^{1+\rho} < +\infty. \tag{10}$$

(*a*) *Then*

$$\sup_{n\in\mathbb{N}} \nu_n^\eta(\omega, V^{p/(1+\rho)}) < +\infty \qquad \mathbb{P}(\mathrm{d}\omega)\text{-}a.s. \tag{11}$$

(*b*) *When* $p \leqslant 1 + \rho$, *assume also that* $\sigma\sigma^* = o(V^{p/(1+\rho)})$ *and* $\sum_{n\geqslant 1}\eta_n\gamma_n/H_n < +\infty$. *Then, with probability 1, any weak limit of the sequence* $(\nu_n^\eta)_{n\in\mathbb{N}}$ *is an invariant distribution for the diffusion* (1).

Remark 1. When $p = 1 + \rho$, the assumption that $\sum_{n\geqslant 1}\eta_n\gamma_n/H_n < +\infty$ turns out to be unnecessary.

Note that (11) implies that $(\nu_n^\eta)_{n\in\mathbb{N}}$ is almost surely tight, since $\lim_{|x|\to+\infty} V(x) = +\infty$. If we add to the assumptions of Theorem 3 the uniqueness of the invariant distribution of the diffusion, we obtain that $\mathbb{P}$-a.s. $\nu_n^\eta \overset{(\mathbb{R}^d)}{\Longrightarrow} \nu$, where ν is the invariant distribution, and, for every $f \in \mathcal{C}(\mathbb{R}^d, \mathbb{R})$,

$$(f(x) = o(V^{p/(1+\rho)}(x)) \text{ as } |x| \to +\infty) \Rightarrow (\mathbb{P}\text{-a.s. } \nu_n^\eta(f) \to \nu(f)). \tag{12}$$

Note that if, in Theorem 3, we have $\eta = \gamma$, condition (10) is satisfied with $\rho = 1$. This shows in particular that Theorem 1 is a corollary of Theorem 3.

Further comments are made below on the main assumptions of Theorem 3.

2.1. The Lyapunov assumptions $(\mathcal{L}_{V,p})$, $p \geqslant 1$

The main assumption we make on the diffusion, i.e. on functions b and σ, is the existence of a positive Lyapunov function V satisfying $(\mathcal{L}_{V,p})$. Actually, this kind of assumption is quite standard as soon as one wishes to investigate the long-run behaviour of a dynamical system, either in the stochastic or in the deterministic world.

Remark 2. Requiring the Lyapunov function V to be bounded away from 0 is not a true restriction since one may always consider $V + 1$ instead of V.

Remark 3. Assumptions (ii)–(iii) in $(\mathcal{L}_{V,p})$ induce a quite stringent constraint on the drift b and on the function V. As a matter of fact (iii) implies that, for every $\lambda > 0$,

$$\lambda |\nabla V|^2 + \frac{|b|^2}{\lambda} \geqslant 2\alpha V - 2\beta.$$

Combined with (ii), this yields that $|\nabla\sqrt{V}|$ is bounded and bounded away from 0 near infinity so that $V(x) = O(|x|^2)$ and $b(x) = O(|x|)$. Actually, in most examples, $V(x) \asymp |x|^2$ and $|b(x)| \asymp |x|$.

Remark 4. In the more specific case where $V(x) = |x|^2 + A$ for every $x \in \mathbb{R}^d$, one can verify that $\lambda_p = 2p - 1$ (so it does not depend upon the real constant A). Then $(\mathcal{L}_{V,p})$ becomes

$$\begin{aligned} &|b|^2(x) + \mathrm{Tr}(\sigma\sigma^*)(x) \leqslant c_V(1 + |x|^2) \qquad \text{for some constant } c_V > 0, \\ &(x|b(x)) + (2p-1)\mathrm{Tr}(\sigma\sigma^*)(x) \leqslant \beta - \alpha|x|^2 \qquad \text{where } \alpha > 0 \text{ and } \beta \in \mathbb{R}, \end{aligned} \tag{13}$$

which we denote by $(\mathcal{L}_{|.|^2,p})$. In particular, $b(x) := -\mu x$, $\mu \in \mathbb{R}_+^*$ and $\sigma(x) := (1 + |x|)\Sigma$, $\Sigma \in \mathcal{S}_+(d, \mathbb{R})$ satisfy $(\mathcal{L}_{|.|^2,p})$ if and only if

$$\mathrm{Tr}(\Sigma\Sigma^*) < \frac{\mu}{2p-1}.$$

Remark 5. Lyapunov functions are typically used to derive the existence of a stationary distribution (see Has'minskii 1980, Chapters 3–4; Ethier and Kurtz 1986). Concerning uniqueness, at least two types of uniqueness criteria are available: uniform ellipticity conditions for diffusions with smooth and bounded coefficients (Has'minskii 1980; Ethier and Kurtz 1986; see also Karatzas and Shreve 1988, Theorem 5.15 and Exercise 5.40, for one-dimensional diffusions); and asymptotic flatness (Basak and Bhattacharya 1992).

2.2. Steps and weights

In this subsection, we will specify the step–weight assumptions of Theorem 3 when $\eta = \gamma$ and when $\eta \neq \gamma$ for two natural parametrized families of steps and weights: polynomial steps and weights on the one hand, and log-polynomial steps and polynomial weights on the other hand.

The case $\eta = \gamma$. The step–weight condition (10) is equivalent to

$$\sum_{n\geqslant 1}\left(\frac{\gamma_n}{\Gamma_n^2}\right)^{(1+\rho)/2} < +\infty.$$

The condition $\sum_{n\geqslant 1}\eta_n\gamma_n/\Gamma_n < +\infty$ becomes $\sum_{n\geqslant 1}\gamma_n^2/\Gamma_n < +\infty$.

For polynomial steps $\gamma_n := n^{-\alpha}$, these conditions become:

$$\sum_{n\geqslant 1}\left(\frac{\gamma_n}{\Gamma_n^2}\right)^{(1+\rho)/2} < +\infty \Leftrightarrow \left(0 < \alpha < \frac{2\rho}{1+\rho} \text{ or } \alpha = \rho = 1\right),$$

$$\sum_{n\geqslant 1}\frac{\gamma_n^2}{\Gamma_n} < +\infty \Leftrightarrow 0 < \alpha \leqslant 1.$$

For log-polynomial steps $\gamma_n := (\ln n)^{-\alpha}$, we obtain:

$$\sum_{n\geqslant 1}\left(\frac{\gamma_n}{\Gamma_n^2}\right)^{(1+\rho)/2} < +\infty \Leftrightarrow \alpha > 0,$$

$$\sum_{n\geqslant 1}\frac{\gamma_n^2}{\Gamma_n} < +\infty \Leftrightarrow \alpha > 1.$$

The case $\eta \neq \gamma$. For polynomial steps and weights, set $\gamma_n := n^{-\alpha}$, $0 < \alpha \leqslant 1$, $\eta_n := n^{-\beta}$, $\beta \leqslant 1$. The assumptions of Theorem 3 are fulfilled if and only if

$$(\alpha, \beta) \in \left(0, \frac{2\rho}{1+\rho}\right) \times (-\infty, 1] \cup \left\{\left(\frac{2\rho}{1+\rho}, 1\right)\right\}.$$

For log-polynomial steps and polynomial weights, set $\gamma_n := (\ln n)^{-\alpha}$, $\eta_n := n^{-\beta}$, with $\alpha, \beta > 0$. Assumption (10) is then satisfied for all values of α and β and $\sum_{n\geqslant 1}\eta_n\gamma_n/\Gamma_n < +\infty$ holds if and only if $\beta = 1$ or $(\alpha, \beta) \in (1, +\infty) \times (0, 1)$.

3. Almost sure tightness of the weighted empirical measures

This section is devoted to the tightness of the random empirical measures $(\nu_n^\eta(\omega, \mathrm{d}x))_{n\in\mathbb{N}^*}$. More precisely, we obtain the almost sure boundedness of the empirical moments of V according to the integrability properties of the noise.

The key inequality is (16), which is a variant of results recently obtained in the field of stochastic approximation (see Pelletier 1999; Basak *et al.* 1997). When $p = 1$ it is the discrete-time version of the decay assumption $AV \leqslant \beta - \alpha V$ which appears in the diffusion setting.

Theorem 4. *Let* $p \in [1, +\infty)$. *Assume* $(\mathcal{L}_{V,p})$ *and* $\mathbb{E}|U_1|^{2p} < +\infty$. *If there exists* $\rho \in (0, 1]$ *such that*

$$\sum_{n\geqslant 1} \frac{1}{H_n}\left(\Delta \frac{\eta_n}{\gamma_n}\right)_+ < +\infty \quad \textit{and} \quad \sum_{n\geqslant 1}\left(\frac{\eta_n}{H_n\sqrt{\gamma_n}}\right)^{1+\rho} < +\infty, \tag{14}$$

then

$$\mathbb{P}(\mathrm{d}\omega)\text{-}a.s. \qquad \sup_{n\in\mathbb{N}^*} \nu_n^\eta(\omega, V^{p/(1+\rho)}) < +\infty.$$

Lemma 2. (a) *If* $(\mathcal{L}_{V,1})$ (i)–(ii) *hold, then, for every* $a \geqslant \frac{1}{2}$,

$$|V^a(X_{n+1}) - V^a(X_n)| \leqslant c_a\sqrt{\gamma_{n+1}}V^a(X_n)(1 + |U_{n+1}|^{2a}). \tag{15}$$

(b) *If, for some real number* $p \geqslant 1$, *we have* $(\mathcal{L}_{V,p})$ *and* $\mathbb{E}|U_1|^{2p} < +\infty$, *then there exist real numbers* $\tilde{\alpha} > 0$ *and* $\tilde{\beta}$ *and* $n_0 \in \mathbb{N}$ *such that*

$$\forall n \geqslant n_0, \qquad \mathbb{E}(V^p(X_{n+1})/\mathcal{F}_n) \leqslant V^p(X_n) + \gamma_{n+1}V^{p-1}(X_n)(\tilde{\beta} - \tilde{\alpha}V(X_n)) \tag{16}$$

and, furthermore,

$$\sup_{n\in\mathbb{N}} \mathbb{E}(V^p(X_n)) < +\infty.$$

Note that part (b), stated with $p = 1$, yields $\sup_{n\in\mathbb{N}} \mathbb{E}V(X_n) < +\infty$.

Proof. (a) It follows from the mean value theorem that

$$V^a(X_{n+1}) - V^a(X_n) = aV^{a-1}(\xi_{n+1})(\nabla V(\xi_{n+1})|\Delta X_{n+1})$$

for some $\xi_{n+1} \in (X_n, X_{n+1})$ (geometric segment). Hence, using $(\mathcal{L}_{V,1})$ (ii),

$$|V^a(X_{n+1}) - V^a(X_n)| \leqslant CV^{a-1/2}(\xi_{n+1})|\Delta X_{n+1}|. \tag{17}$$

Now, $\nabla\sqrt{V}$ being bounded by $(\mathcal{L}_{V,1})$ (ii), $\sqrt{V}$ is Lipschitz. Hence,

$$\begin{aligned} V^{a-1/2}(\xi_{n+1}) &\leqslant (\sqrt{V}(X_n) + [\sqrt{V}]_1|\Delta X_{n+1}|)^{2a-1} \\ &\leqslant 2^{2a-1}(V^{a-1/2}(X_n) + [\sqrt{V}]_1^{2a-1}|\Delta X_{n+1}|^{2a-1}). \end{aligned} \tag{18}$$

We also have

$$\begin{aligned} |\Delta X_{n+1}| &\leqslant \gamma_{n+1}|b(X_n)| + \sqrt{\gamma_{n+1}}\|\sigma(X_n)\||U_{n+1}| \\ &\leqslant C\sqrt{\gamma_{n+1}}\sqrt{V}(X_n)(1 + |U_{n+1}|). \end{aligned} \tag{19}$$

Plugging (18) and (19) into (17) leads to

$$|V^a(X_{n+1}) - V^a(X_n)| \leqslant c_a(\sqrt{\gamma_{n+1}}V^a(X_n)(1+|U_{n+1}|) + V^a(X_n)\gamma_{n+1}^a(1+|U_{n+1}|)^{2a})$$

which, noting that $2a \geqslant 1$, yields the required result.

(b) The Taylor formula applied to V^p between X_n and X_{n+1} yields

$$V^p(X_{n+1}) = V^p(X_n) + pV^{p-1}(X_n)(\nabla V(X_n)|\Delta X_{n+1}) + \tfrac{1}{2}D^2(V^p)(\xi_{n+1}).(\Delta X_{n+1})^{\otimes 2},$$

where $\xi_{n+1} \in (X_n, X_{n+1})$. Now, starting from $D^2(V^p) = pV^{p-1}D^2V + p(p-1) \times V^{p-2}\nabla V \nabla V^*$, the very definition of λ_p implies that

$$D^2(V^p)(\xi_{n+1}).(\Delta X_{n+1})^{\otimes 2} \leqslant 2p\lambda_p V^{p-1}(\xi_{n+1})|\Delta X_{n+1}|^2.$$

Hence

$$V^p(X_{n+1}) \leqslant V^p(X_n) + pV^{p-1}(X_n)(\nabla V(X_n)|\Delta X_{n+1}) + p\lambda_p V^{p-1}(\xi_{n+1})|\Delta X_{n+1}|^2. \qquad (20)$$

At this stage, one needs to investigate successively the cases $p=1$ and $p>1$. Assume first that $p=1$. It follows from (20) and assumption $(\mathcal{L}_{V,1})$ (iii) that

$$\begin{aligned}
\mathbb{E}(V(X_{n+1})/\mathcal{F}_n) &\leqslant V(X_n) + \gamma_{n+1}(\nabla V|b)(X_n) + \lambda_1(\gamma_{n+1}^2|b(X_n)|^2 + \gamma_{n+1}\mathrm{Tr}(\sigma\sigma^*)(X_n)) \\
&\leqslant V(X_n) + \gamma_{n+1}[(\nabla V|b)(X_n) + \lambda_1 \mathrm{Tr}(\sigma\sigma^*)(X_n)] + \lambda_1\gamma_{n+1}^2|b(X_n)|^2 \\
&\leqslant V(X_n) + \gamma_{n+1}(\beta - \alpha V(X_n)) + \lambda_1 c_V \gamma_{n+1}^2 V(X_n). \qquad (21)
\end{aligned}$$

Consequently, there exist $\tilde{\alpha} > 0$ and $n_0 \in \mathbb{N}$, such that, for every $n \geqslant n_0$,

$$\mathbb{E}(V(X_{n+1})/\mathcal{F}_n) \leqslant V(X_n)(1 - \tilde{\alpha}\gamma_{n+1}) + \beta\gamma_{n+1} \quad \text{and} \quad 1 - \tilde{\alpha}\gamma_{n+1} > 0.$$

Then, one shows by induction from (21) that $V(X_n) \in L^1$ for every $n \in \mathbb{N}$. Taking the expectation leads, for $n \geqslant n_0$, to

$$\mathbb{E}(V(X_{n+1})) \leqslant \mathbb{E}(V(X_n))(1 - \tilde{\alpha}\gamma_{n+1}) + \beta\gamma_{n+1}.$$

A simple induction yields

$$\sup_{n \geqslant n_0} \mathbb{E}(V(X_n)) \leqslant \frac{\beta}{\tilde{\alpha}} \vee \mathbb{E}V(X_{n_0})$$

and completes the proof.

Assume now that $p > 1$. The function $\sqrt{V}$ is Lipschitz by $(\mathcal{L}_{V,p})$ (ii) since $\nabla(\sqrt{V})$ is bounded. Consequently, as $\xi_{n+1} \in (X_n, X_{n+1})$,

$$V^{p-1}(\xi_{n+1}) = \sqrt{V}^{2(p-1)}(\xi_{n+1}) \leqslant (\sqrt{V}(X_n) + [\sqrt{V}]_1|\Delta X_{n+1}|)^{2(p-1)}.$$

Then, using the following straightforward inequalities, valid for every $u, v \geqslant 0$,

$$\forall \alpha \in [0,1], \qquad (u+v)^\alpha \leqslant u^\alpha + v^\alpha,$$

$$\forall \alpha \geqslant 1, \qquad (u+v)^\alpha \leqslant u^\alpha + \alpha(u+v)^{\alpha-1}v \leqslant u^\alpha + \alpha 2^{\alpha-1}(u^{\alpha-1}v + v^\alpha),$$

one derives

$$V^{p-1}(\xi_{n+1}) \leqslant \begin{cases} V^{p-1}(X_n) + ([\sqrt{V}]_1 |\Delta X_{n+1}|)^{2(p-1)} & \text{if } 2(p-1) \leqslant 1, \\ V^{p-1}(X_n) + C(V^{(2p-3)/2}(X_n)|\Delta X_{n+1}| + |\Delta X_{n+1}|^{2(p-1)}) & \text{if } 2(p-1) > 1, \end{cases}$$

$$V^{p-1}(\xi_{n+1})|\Delta X_{n+1}|^2 \leqslant V^{p-1}(X_n)|\Delta X_{n+1}|^2 + C|\Delta X_{n+1}|^{2p\wedge 3} \times \begin{cases} 1 & \text{if } 2p \leqslant 3, \\ (V^{(2p-3)/2}(X_n) + |\Delta X_{n+1}|^{2p-3}) & \text{if } 2p > 3. \end{cases} \tag{22}$$

Combining inequality (22) with (19) leads to the following upper bound (using $v_* := \min V > 0$ for the case $2p > 3$):

$$V^{p-1}(\xi_{n+1})|\Delta X_{n+1}|^2 \leqslant V^{p-1}(X_n)|\Delta X_{n+1}|^2 + C\gamma_{n+1}^{p\wedge 3/2} V^p(X_n)(1 + |U_{n+1}|^{2p}). \tag{23}$$

By plugging (23) into (20), one obtains

$$\begin{aligned} \mathbb{E}(V^p(X_{n+1})/\mathcal{F}_n) &\leqslant V^p(X_n) + pV^{p-1}(X_n)\gamma_{n+1}((\nabla V(X_n)|b(X_n)) + \lambda_p \mathrm{Tr}(\sigma\sigma^*(X_n))) \\ &\quad + p\lambda_p V^{p-1}(X_n)\gamma_{n+1}^2 |b(X_n)|^2 + C\gamma_{n+1}^{p\wedge 3/2} V^p(X_n), \\ &\leqslant V^p(X_n) + pV^{p-1}(X_n)\gamma_{n+1}(\beta - \alpha V(X_n)) + CV^p(X_n)(\gamma_{n+1}^2 + \gamma_{n+1}^{p\wedge 3/2}). \end{aligned} \tag{24}$$

As $\gamma_n^{p\wedge 3/2} = o(\gamma_n)$, there exist $\tilde{\alpha} > 0$, $\tilde{\beta} > 0$ and some $n_0 \in \mathbb{N}$ such that, for every $n \geqslant n_0$,

$$\mathbb{E}(V^p(X_{n+1})/\mathcal{F}_n) \leqslant V^p(X_n) + \gamma_{n+1} V^{p-1}(X_n)(\tilde{\beta} - \tilde{\alpha} V(X_n)).$$

The boundedness of the sequence $(\mathbb{E} V^p(X_n))_{n\in\mathbb{N}}$ follows as in the case $p = 1$. □

Lemma 3. *Let $W : \mathbb{R}^d \to \mathbb{R}_+$ be a non-negative Borel function.*

(*a*) *If $\sup_{n\in\mathbb{N}} \mathbb{E} W(X_n) < +\infty$ and $\sum_{n\geqslant 1}(1/H_n)(\Delta(\eta_n/\gamma_n))_+ < +\infty$, then*

$$\overline{\lim_n} -\frac{1}{H_n}\sum_{k=1}^{n} \frac{\eta_k}{\gamma_k} \Delta W(X_k) \leqslant 0 \ a.s.$$

(*b*) *If W is bounded and $\lim_n (1/H_n)\sum_{k=1}^{n} |\Delta(\eta_k/\gamma_k)| = 0$, then*

$$\lim_n -\frac{1}{H_n}\sum_{k=1}^{n} \frac{\eta_k}{\gamma_k} \Delta W(X_k) = 0 \ a.s.$$

Proof. (a) Setting, without loss of generality, $\eta_0/\gamma_0 := 0$, one obtains

$$\begin{aligned} -\frac{1}{H_n}\sum_{k=1}^{n} \frac{\eta_k}{\gamma_k} \Delta W(X_k) &= -\frac{\eta_n}{H_n \gamma_n} W(X_n) + \frac{1}{H_n}\sum_{k=1}^{n} \Delta\left(\frac{\eta_k}{\gamma_k}\right) W(X_{k-1}) \\ &\leqslant \frac{1}{H_n}\sum_{k=1}^{n} \left(\Delta \frac{\eta_k}{\gamma_k}\right)_+ W(X_{k-1}). \end{aligned} \tag{25}$$

The series assumption implies that

$$\sum_{n\geqslant 1}\frac{1}{H_n}\left(\Delta\frac{\eta_n}{\gamma_n}\right)_+ \mathbb{E}W(X_{n-1}) < +\infty.$$

Hence the series

$$\sum_{n\geqslant 1}\frac{1}{H_n}\left(\Delta\frac{\eta_n}{\gamma_n}\right)_+ W(X_{n-1}) \qquad \mathbb{P}\text{-a.s.}$$

converges in $\mathbb{R}^d$. The Kronecker lemma completes the proof.

(b) This follows from (25), once it is noticed that $\lim_n \eta_n/(H_n\gamma_n) = 0$. This follows from the triangle inequality. □

Proof of Theorem 4. From Lemma 2(b), we have, for some $n_0 \in \mathbb{N}$,

$$\forall n \geqslant n_0, \qquad \mathbb{E}(V^p(X_{n+1})/\mathcal{F}_n) \leqslant V^p(X_n) + \gamma_{n+1}V^{p-1}(X_n)(\tilde\beta - \tilde\alpha V(X_n)),$$

or, equivalently,

$$\mathbb{E}\left(\frac{V^p(X_{n+1})}{V^p(X_n)}/\mathcal{F}_n\right) \leqslant 1 + \gamma_{n+1}\frac{\tilde\beta - \tilde\alpha V(X_n)}{V(X_n)}.$$

Now set $p' := p/(1+\rho)$. For $n \geqslant n_0$, we have, using Jensen's inequality and the concavity of $x \mapsto x^{1/(1+\rho)}$,

$$\mathbb{E}\left(\frac{V^{p'}(X_{n+1})}{V^{p'}(X_n)}/\mathcal{F}_n\right) \leqslant \left(1 + \gamma_{n+1}\frac{\tilde\beta - \tilde\alpha V(X_n)}{V(X_n)}\right)^{1/(1+\rho)}$$

$$\leqslant 1 + \frac{\gamma_{n+1}}{1+\rho}\frac{\tilde\beta - \tilde\alpha V(X_n)}{V(X_n)}.$$

The last inequality can be rewritten as follows:

$$\mathbb{E}(V^{p'}(X_{n+1})/\mathcal{F}_n) \leqslant V^{p'}(X_n) + \frac{\gamma_{n+1}}{1+\rho}V^{p'-1}(X_n)(\tilde\beta - \tilde\alpha V(X_n)).$$

Using the fact that, for any $\varepsilon > 0$, $V^{p'-1} \leqslant C_\varepsilon + \varepsilon V^{p'}$ for some constant C_ε, we obtain, for some $\hat\alpha > 0$ and some $\hat\beta$,

$$\mathbb{E}(V^{p'}(X_{n+1})/\mathcal{F}_n) \leqslant V^{p'}(X_n) + \gamma_{n+1}(\hat\beta - \hat\alpha V^{p'}(X_n)),$$

or, equivalently,

$$\forall n \geqslant n_0, \qquad V^{p'}(X_n) \leqslant \frac{V^{p'}(X_n) - \mathbb{E}(V^{p'}(X_{n+1})/\mathcal{F}_n)}{\gamma_{n+1}\hat\alpha} + \frac{\hat\beta}{\hat\alpha}.$$

Consequently, the almost sure finiteness of $\sup_{n\in\mathbb{N}^*}\nu_n^\eta(\omega, V^{p'})$ amounts to showing that, $\mathbb{P}$-a.s.,

$$\sup_n \frac{1}{H_n}\sum_{k=1}^{n}\frac{\eta_k}{\gamma_k}(V^{p'}(X_{k-1}) - \mathbb{E}(V^{p'}(X_k)/\mathcal{F}_{k-1})) < +\infty.$$

Now

$$\sum_{k=1}^{n}\frac{\eta_k}{\gamma_k}(V^{p'}(X_{k-1}) - \mathbb{E}(V^{p'}(X_k)/\mathcal{F}_{k-1}))$$

$$= -\sum_{k=1}^{n}\frac{\eta_k}{\gamma_k}\Delta V^{p'}(X_k) + \sum_{k=1}^{n}\frac{\eta_k}{\gamma_k}(V^{p'}(X_k) - \mathbb{E}(V^{p'}(X_k)/\mathcal{F}_{k-1})).$$

On the one hand, Lemma 3(a) applied to $W := V^{p'}$ implies that

$$\overline{\lim_n} - \frac{1}{H_n}\sum_{k=1}^{n}\frac{\eta_k}{\gamma_k}\Delta V^{p'}(X_k) \leqslant 0.$$

On the other hand, the Kronecker lemma shows that the conclusion will follow from the almost sure convergence of the martingale

$$M_n := \sum_{k=1}^{n}\frac{\eta_k}{H_k\gamma_k}(V^{p'}(X_k) - \mathbb{E}(V^{p'}(X_k)/\mathcal{F}_{k-1})), \qquad n \geqslant 1$$

($M_0 := 0$). In turn, this almost sure convergence will follow, using the Chow theorem, from the convergence of the series

$$\sum_{n\geqslant 1}\left(\frac{\eta_n}{H_n\gamma_n}\right)^{1+\rho}\mathbb{E}|V^{p'}(X_n) - \mathbb{E}(V^{p'}(X_n)/\mathcal{F}_{n-1})|^{1+\rho}.$$

First, observe that

$$\begin{aligned}\|V^{p'}(X_k) - \mathbb{E}(V^{p'}(X_k)/\mathcal{F}_{k-1})\|_{1+\rho} &\leqslant \|V^{p'}(X_k) - V^{p'}(X_{k-1})\|_{1+\rho}\\ &\quad + \|V^{p'}(X_{k-1}) - \mathbb{E}(V^{p'}(X_k)/\mathcal{F}_{k-1})\|_{1+\rho}\\ &\leqslant \|V^{p'}(X_k) - V^{p'}(X_{k-1})\|_{1+\rho}\\ &\quad + \|\mathbb{E}(V^{p'}(X_{k-1}) - V^{p'}(X_k)/\mathcal{F}_{k-1})\|_{1+\rho}\\ &\leqslant 2\|V^{p'}(X_k) - V^{p'}(X_{k-1})\|_{1+\rho}.\end{aligned}$$

Combining this inequality with Lemma 2(a) yields

$$\mathbb{E}|V^{p'}(X_n) - \mathbb{E}(V^{p'}(X_n)/\mathcal{F}_{n-1})|^{1+\rho} \leqslant 2^{1+\rho}C\gamma_k^{(1+\rho)/2}\mathbb{E}(V(X_{k-1})^p(1+|U_k|^{2p})).$$

Consequently the above convergence holds as a consequence of assumption (14) and $\sup_n \mathbb{E}V^p(X_n) < +\infty$. □

4. Identification of the limit

In this section we characterize – $\mathbb{P}(d\omega)$-a.s. – the weak limiting distributions of the tight sequence $\nu_n^\eta(\omega, dx)$. To this end we will essentially establish that any weak limiting distribution $\nu_\infty^\eta(\omega, dx)$ satisfies

$$\forall f \in \mathcal{C}_c^2(\mathbb{R}^d), \qquad \nu_\infty^\eta(\omega, Af) = 0 \tag{26}$$

where

$$A(f)(x) := (b|\nabla f)(x) + \frac{1}{2}\sum_{i,j=1}^{d}(\sigma\sigma^*)_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$$

denotes the infinitesimal generator of the diffusion $(Y_t)_{t\in\mathbb{R}_+}$. It will then follow from the Echeverría–Weiss theorem (see below) that $\nu_\infty^\eta(\omega, dx)$ is an invariant distribution of $(Y_t)_{t\in\mathbb{R}_+}$. In fact, establishing (26) will even prove the existence of at least one invariant distribution (of course, the assumptions we made are not optimal for that purpose, but we are looking for much more than that).

4.1. Some background on the Echeverría–Weiss theorem

The key to the identification of the limit is the Echeverría–Weiss theorem (see Ethier and Kurtz 1986, Chapter 4, Theorem 9.17) for linear operators satisfying a *positive maximum principle*.

Definition 1. *Let E be a locally compact and separable metric space. A linear operator A defined on the subset $\mathcal{D}(A)$ of the set $\mathcal{C}_0(E)$ of continuous functions that vanish at infinity satisfies the positive maximum principle if*

$$\forall f \in \mathcal{D}(A), \qquad \sup_{x\in E} f(x) = f(x_0) \geqslant 0 \Rightarrow Af(x_0) \leqslant 0.$$

(*When E is a compact set, $\mathcal{C}_0(E)$ is defined as $\mathcal{C}_0(E) = \mathcal{C}(E)$*).

Theorem 5 (Echeverría–Weiss). *Let E be a locally compact Polish space and A a linear operator satisfying the positive maximum principle. Assume that its domain $\mathcal{D}(A)$ is an algebra everywhere dense in $(\mathcal{C}_0(E), \|\cdot\|_\infty)$, containing a sequence $(f_n)_{n\in\mathbb{N}}$ satisfying*

$$\sup_{n\in\mathbb{N}}(\|f_n\|_\infty + \|Af_n\|_\infty) < +\infty, \quad \forall x \in E, \quad f_n(x) \to 1 \quad \textit{and} \quad Af_n(x) \to 0. \tag{27}$$

If a distribution ν on $(E, \mathcal{B}(E))$ satisfies $\int_E Af\, d\nu = 0$ for every $f \in \mathcal{D}(A)$, then there exists a stationary solution for the martingale problem (A, ν) (this means that there exists a stationary continuous-time homogeneous Markov process with infinitesimal generator A and invariant distribution ν).

Remark 6. Although assumption (27) is not explicitly stated in Theorem 9.17 of Ethier and Kurtz (1986), it seems to be used implicitly in the proof.

Lemma 4. *If the functions b, σ and V satisfy* $|b|^2 + \mathrm{Tr}(\sigma\sigma^*) + |\nabla V|^2 \leqslant c_V V$, *then the above operator A satisfies the assumptions of the Echeverría–Weiss theorem.*

Proof. It is well known that the generator of a diffusion satisfies the positive maximum principle. In order to check (27), observe that the function V satisfies $V(x) = O(|x|^2)$, which in turn implies that $|b(x)| \leqslant C(1 + |x|)$ and $|(\sigma\sigma^*)_{ij}(x)| \leqslant C(1 + |x|^2)$. Now one sets $f_n(x) := \varphi(x/n)$, where φ is $\mathcal{C}^2$ with compact support and $\varphi(0) = 1$. The fact that $f_n(x)$ goes to 1 while $A(f_n)(x)$ goes to zero is straightforward. The uniform boundedness simply follows from the above bounds for b and $\sigma\sigma^*$. □

4.2. The main identification result

In this subsection we need a kind of 'light' Lyapunov assumption on the functions b and σ, which is that there exists some function $V : \mathbb{R}^d \to \mathbb{R}_+$ such that

$$\lim_{|x|\to+\infty} V(x) = +\infty, \quad |\nabla V|^2 + |b|^2 + \mathrm{Tr}(\sigma\sigma^*) \leqslant cV \quad \text{and} \quad \sup_{n\in\mathbb{N}} \mathbb{E} V(X_n) < +\infty. \tag{28}$$

Theorem 6. *Assume that b and σ are* continuous *functions. Assume that* (28) *is fulfilled for some function V. Assume that* (3) *and* (4) *hold and that*

$$\lim_n \frac{1}{H_n} \sum_{k=1}^{n} \left| \Delta \frac{\eta_k}{\gamma_k} \right| = 0 \quad \text{and} \quad \sum_{n\geqslant 1} \left(\frac{\eta_n}{H_n \sqrt{\gamma_n}} \right)^2 < +\infty. \tag{29}$$

Let $a \geqslant \frac{1}{2}$. *Assume that* $\sup_n \nu_n^\eta(V^a) < +\infty$ $\mathbb{P}$-*a.s. and* $\mathrm{Tr}(\sigma\sigma^*) = o(V^a)$. *If* $a < 1$, *assume furthermore that* $\sum_{n\geqslant 1} \eta_n \gamma_n / H_n < +\infty$.

Then, $\mathbb{P}(\mathrm{d}\omega)$-*a.s., every limiting distribution* $\nu_\infty(\omega, \mathrm{d}x)$ *of the sequence* $(\nu_n^\eta(\omega, \mathrm{d}x))_{n\in\mathbb{N}}$ *is an invariant distribution of a (weak) solution of* (1).

Note that Theorem 3 follows from Theorems 4 and 6.

The proof of Theorem 6 relies on the two technical lemmas below.

Lemma 5. *If condition* (28) *is fulfilled and if the step–weight assumptions* (3), (4) *and* (29) *hold, then, for every bounded Lipschitz continuous function* $f : \mathbb{R}^d \to \mathbb{R}$,

$$\mathbb{P}\text{-}a.s. \qquad \lim_n \frac{1}{H_n} \sum_{k=1}^{n} \frac{\eta_k}{\gamma_k} \mathbb{E}(\Delta f(X_k)/\mathcal{F}_{k-1}) = 0.$$

Proof. The proof is similar to that of Theorem 4. Setting $\eta_0/\gamma_0 := 0$ gives

$$\frac{1}{H_n} \sum_{k=1}^{n} \frac{\eta_k}{\gamma_k} \mathbb{E}(\Delta f(X_k)/\mathcal{F}_{k-1}) = \frac{1}{H_n} \sum_{k=1}^{n} \frac{\eta_k}{\gamma_k} \Delta f(X_k) - \frac{1}{H_n} \sum_{k=1}^{n} \frac{\eta_k}{\gamma_k} (f(X_k) - \mathbb{E}(f(X_k)/\mathcal{F}_{k-1})).$$

As f is bounded, Lemma 3(b) implies that, a.s.,

$$\lim_n \frac{1}{H_n} \sum_{k=1}^{n} \frac{\eta_k}{\gamma_k} \Delta f(X_k) = 0.$$

Then

$$\frac{1}{H_n} \sum_{k=1}^{n} \frac{\eta_k}{\gamma_k} (f(X_k) - \mathbb{E}(f(X_k)/\mathcal{F}_{k-1}))$$

will converge to 0 once the martingale

$$M_n^f := \sum_{k=1}^{n} \frac{\eta_k}{\gamma_k H_k} (f(X_k) - \mathbb{E}(f(X_k)/\mathcal{F}_{k-1}))$$

a.s. converges in $\mathbb{R}$. Now, using assumption (28),

$$\mathbb{E}\langle M^f \rangle_\infty = \sum_{n \geqslant 1} \left(\frac{\eta_n}{H_n \gamma_n} \right)^2 \| f(X_k) - \mathbb{E}(f(X_k)/\mathcal{F}_{k-1}) \|_2^2 \leqslant \sum_{n \geqslant 1} \left(\frac{\eta_n}{H_n \gamma_n} \right)^2 \| f(X_k) - f(X_{k-1}) \|_2^2$$

$$\leqslant [f]_1^2 \sum_{n \geqslant 1} \left(\frac{\eta_n}{H_n \gamma_n} \right)^2 \| \Delta X_k \|_2^2 \leqslant C \sum_{n \geqslant 1} \left(\frac{\eta_n}{H_n \sqrt{\gamma_n}} \right)^2 < +\infty,$$

where

$$[f]_1 := \sup_{x, y \in \mathbb{R}^d, x \neq y} \frac{|f(x) - f(y)|_F}{|x - y|}.$$

□

Lemma 6. *If the assumptions of Theorem 6 hold, then for every twice continuously differentiable function f with compact support,*

$$\lim_n \left(\frac{1}{H_n} \sum_{k=1}^{n} \frac{\eta_k}{\gamma_k} \mathbb{E}(\Delta f(X_k)/\mathcal{F}_{k-1}) - \nu_n^\eta(Af) \right) = 0 \ a.s.$$

Proof. Setting $R_2(x, y) := f(y) - f(x) - (\nabla f(x) | y - x) - \frac{1}{2} D^2 f(x).(y - x)^{\otimes 2}$, one obtains

$$f(X_k) - f(X_{k-1}) = (\nabla f(X_{k-1}) | \Delta X_k) + \tfrac{1}{2} D^2 f(X_{k-1}) \Delta X_k^{\otimes 2} + R_2(X_{k-1}, X_k)$$

for every $k \in \mathbb{N}^*$. Hence, using the fact that $\mathbb{E}(U_k/\mathcal{F}_{k-1}) = 0$,

$$\mathbb{E}(f(X_k) - f(X_{k-1})/\mathcal{F}_{k-1}) - \gamma_k Af(X_{k-1}) = \frac{\gamma_k^2}{2} D^2 f(X_{k-1}).b^{\otimes 2}(X_{k-1}) + \mathbb{E}(R_2(X_{k-1}, X_k)/\mathcal{F}_{k-1}),$$

$$\frac{\eta_k}{\gamma_k}\mathbb{E}(f(X_k) - f(X_{k-1})/\mathcal{F}_{k-1}) - \eta_k Af(X_{k-1}) = \frac{\eta_k\gamma_k}{2} D^2 f(X_{k-1}) b^{\otimes 2}(X_{k-1}) + \frac{\eta_k}{\gamma_k}\mathbb{E}(R_2(X_{k-1}, X_k)/\mathcal{F}_{k-1}).$$

First, note that

$$\left|\frac{1}{H_n}\sum_{k=1}^n \eta_k\gamma_k D^2 f(X_{k-1}) b^{\otimes 2}(X_{k-1})\right| \leqslant \|D^2 f.b^{\otimes 2}\|_\infty \frac{1}{H_n}\sum_{k=1}^n \eta_k\gamma_k \overset{n\to+\infty}{\to} 0.$$

Let us now deal with the conditional expectation involving R_2. Let, for every $x, y \in \mathbb{R}^2$,

$$r_2(x, y) := \tfrac{1}{2}\sup_{t\in(0,1)} \|D^2 f(x + t(y-x)) - D^2 f(x)\|.$$

This defines a bounded continuous function on $\mathbb{R}^d \times \mathbb{R}^d$, satisfying $r_2(x, x) = 0$. One may straightforwardly verify that

$$|R_2(x, y)| \leqslant r_2(x, y)|x-y|^2.$$

Then, as $|\Delta X_k|^2 \leqslant 2\gamma_k(\gamma_k|b(X_{k-1})|^2 + |\sigma(X_{k-1})U_k|^2)$, one obtains

$$\begin{aligned}\frac{\eta_k}{\gamma_k}&|\mathbb{E}(R_2(X_{k-1}, X_k)/\mathcal{F}_{k-1})|\\ &\leqslant 2\eta_k\gamma_k\|r_2\|_\infty|b(X_{k-1})|^2 + 2d\eta_k \mathrm{Tr}(\sigma\sigma^*)(X_{k-1})\mathbb{E}(r_2(X_{k-1}, X_k)|U_k|^2/\mathcal{F}_{k-1}),\\ &\leqslant 2\eta_k\gamma_k\|r_2\|_\infty|b(X_{k-1})|^2 + 2d\eta_k(V^a\theta)(X_{k-1})J(\gamma_k, X_{k-1}),\end{aligned}$$

where

$$\mathrm{Tr}(\sigma\sigma^*)(x) := V(x)^a\theta(x) \qquad \text{with } \lim_{|x|\to+\infty}\theta(x) = 0$$

and

$$J(\gamma, x) := \int_{\mathbb{R}^q} r_2(x, x + \gamma b(x) + \sqrt{\gamma}\sigma(x)u)|u|^2\mu(\mathrm{d}u), \qquad \mu := \mathcal{L}(U_1).$$

Note that J is still a bounded continuous function, on $\mathbb{R}_+ \times \mathbb{R}^d$, and $J(0, x) := 0$.

The term $\|r_2\|_\infty\eta_k\gamma_k|b(X_{k-1})|^2$ can be handled as follows. If $a \geqslant 1$, $\mathbb{P}$-a.s.,

$$\frac{1}{H_n}\sum_{k=1}^n \eta_k\gamma_k|b(X_{k-1})|^2 \leqslant \frac{C}{H_n}\sum_{k=1}^n \eta_k\gamma_k V(X_{k-1}) \overset{n\to+\infty}{\to} 0 \qquad \text{as } \sup_n \nu_n^\eta(V) < +\infty \text{ and } \gamma_n \to 0.$$

If $a \in [\frac{1}{2}, 1)$, the result follows from the Kronecker lemma and $\sum_{n\geqslant 1}\eta_n\gamma_n/H_n < +\infty$.

Finally, it remains to prove that

$$\mathbb{P}\text{-}a.s. \quad \lim_n \frac{1}{H_n}\sum_{k=1}^{n}\eta_k J(\gamma_k,\, X_{k-1})(V^a\theta)(X_{k-1}) = 0.$$

Let $A > 0$ be a temporarily fixed real number. The function J is uniformly continuous on the compact set $[0,\, \|\gamma\|_\infty] \times \overline{B}(0;\, A)$, hence $J(\gamma_k,\, X_{k-1})\mathbf{1}_{\{|X_{k-1}|\leqslant A\}} \overset{\mathbb{P}\text{-a.s.}}{\to} 0$ so that, $V^a\theta$ being bounded on $B(0;\, A)$,

$$\lim_n \frac{1}{H_n}\sum_{k=1}^{n}\eta_k J(\gamma_k,\, X_{k-1})\mathbf{1}_{\{|X_{k-1}|\leqslant A\}}(V^a\theta)(X_{k-1}) = 0 \qquad \mathbb{P}\text{-a.s.}$$

To complete the proof, let $A \to +\infty$ in the inequality

$$\overline{\lim_n} \frac{1}{H_n}\sum_{k=1}^{n}\eta_k J(\gamma_k,\, X_{k-1})(V^a\theta)(X_{k-1})\mathbf{1}_{\{|X_{k-1}|>A\}} \leqslant \sup_{|x|>A}|\theta(x)|\|J\|_\infty \sup_n \nu_n^\eta(V^a).$$

□

Proof of Theorem 6. Combining Lemmas 5 and 6 yields that, if f is $\mathcal{C}^2$ with compact support, $\lim_n \nu_n^\eta(Af) = 0$. As Af is bounded and continuous, we obtain that $\nu_\infty^\eta(Af) = 0$ a.s., and may apply Theorem 5. □

5. Theoretical application: the almost sure central limit theorem

As far as numerical simulations are concerned, the choice of a Gaussian or a Bernoulli white noise for $(U_n)_{n\in\mathbb{N}^*}$ seems more appropriate than a square-integrable sequence of i.i.d. random vectors having no finite higher moments. (In fact the existence of higher moments will be required later in Section 6 to derive some rates of convergence.) However, relaxing the moment conditions to square integrability is of some interest from a theoretical point of view: applied to the Ornstein–Uhlenbeck diffusion process, Theorem 3 yields a natural proof of the almost sure central limit theorem, stated with its minimal moment conditions.

Theorem 7 (Standard almost sure CLT). *Let $(U_n)_{n\in\mathbb{N}^*}$ be a sequence of i.i.d. square-integrable random vectors satisfying $\mathbb{E}(U_1) = 0$ and $\Sigma_{U_1} = I_d$. Then*

$$\mathbb{P}\text{-}a.s. \qquad \frac{1}{\ln n}\sum_{k=1}^{n}\frac{1}{k}\delta_{(U_1+\ldots+U_k)/\sqrt{k}} \overset{(\mathbb{R}^d)}{\Longrightarrow} \mathcal{N}(0,\, I_d).$$

The almost sure (or almost everywhere) central limit theorem was first established independently by Brosamler (1988) and Schatte (1988) under slightly more stringent assumptions: $U_1 \in L^{2+\delta}$, $\delta > 0$, in Brosamler (1988); and $U_1 \in L^3$ in Schatte (1988); see also Fisher (1987). In Brosamler (1988) a functional version of the theorem is proved using Skorokhod embedding. These moment assumptions were relaxed in Lacey and Philip (1990); see also Touati (1995). Recently, this theorem has been generalized in several directions. We mention the extension of the almost sure CLT to vector-valued martingale increments by Chaâbane (1996) and Chaâbane *et al*. (1996) or that to converging recursive stochastic algorithms established by Pelletier (1999; 2000).

In another direction, there has been some discussion of the rate of convergence in the almost sure CLT: see Csörgő and Horváth (1992) for a central limit theorem, Heck (1998) for some large-deviation results, and Chaâbane and Maâouia (2000) for a law of the iterated logarithm, among others. In Section 6, we also deal with the rate of convergence of algorithm (2) which, as a by-product, will allow us to recover the central limit theorem for the almost sure central limit theorem.

Actually, we are naturally led, in our framework, to establish an extension of the almost sure CLT that embodies a slightly wider class of weights than $\eta_n = 1/n$.

Theorem 8 (Weighted almost sure CLT). *Let* $(U_n)_{n\in\mathbb{N}^*}$ *be a sequence of i.i.d. square-integrable random vectors defined on a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *and satisfying* $\mathbb{E}(U_1) = 0$ *and* $\Sigma_{U_1} = I_d$. *Let* $(\eta_n)_{n\in\mathbb{N}^*}$ *be a sequence of weights satisfying*

$$\sum_{n\geqslant 1}\frac{\eta_n}{nH_n} < +\infty, \quad \sum_{n\geqslant 1}\frac{n\eta_n}{H_n}\left|1-\left(1-\frac{1}{n}\right)\frac{\eta_{n-1}}{\eta_n}\right| < +\infty \quad \textit{and} \quad \sum_{n\geqslant 1}\frac{n\eta_n^2}{H_n^2} < +\infty. \tag{30}$$

Then

$$\mathbb{P}\text{-}a.s. \qquad \frac{1}{H_n}\sum_{k=1}^{n}\eta_k\delta_{(U_1+\ldots+U_k)/\sqrt{k}} \overset{(\mathbb{R}^d)}{\Longrightarrow} \mathcal{N}(0, I_d).$$

Note that assumption (30) *holds for the family of weights*

$$\eta_n := \frac{\ln^\beta n}{n}, \qquad \beta \geqslant -1.$$

Remark 7. Assumption (30) is not satisfied by polynomial weights $\eta_n := n^\beta$ except for $\beta := -1$, which is the setting of the standard almost sure CLT.

The starting point of the proof of Theorem 8, as in the original proof of the almost sure CLT, is to write the CLT in a recursive form: set

$$Z_0 := 0, \qquad Z_n := \frac{U_1 + \ldots + U_n}{\sqrt{n}},$$

$$\gamma_n := \frac{1}{n}, \qquad n \geqslant 1.$$

One may easily verify that

$$Z_{n+1} = Z_n - \gamma_{n+1}\frac{Z_n}{2} + \sqrt{\gamma_{n+1}}U_{n+1} + R_{n+1}, \qquad \text{where } R_{n+1} := O\left(\frac{1}{n^2}\right)Z_n. \tag{31}$$

This procedure appears as a perturbation of the general decreasing-step Euler scheme (2) with drift $b(x) := -x/2$ and diffusion coefficient $\sigma(x) := I_d$. This suggests the investigation of the asymptotic behaviour of the difference Δ_n between this perturbed procedure (31) and the standard algorithm defined by

$$\forall n \in \mathbb{N}, \quad X_{n+1} = X_n - \gamma_{n+1}\frac{X_n}{2} + \sqrt{\gamma_{n+1}}U_{n+1}, \qquad X_0 := 0.$$

As $\Delta_n := Z_n - X_n$, it is straightforward that

$$\Delta_{n+1} = \left(1 - \frac{1}{2(n+1)}\right)\Delta_n + R_{n+1}, \qquad \Delta_0 := 0. \tag{32}$$

Lemma 7. *The sequence of random variables* $(\Delta_n)_{n\in\mathbb{N}}$ *defined by* (32) $\mathbb{P}$-*a.s. converges to* 0. *That is,*

$$\|\Delta_n\|_2 = O\left(\frac{1}{\sqrt{n}}\right),$$

$$\mathbb{P}\text{-}a.s. \qquad \Delta_n = O\left(\frac{1}{\sqrt{n}}\right).$$

Proof. One shows by induction that, for every $n \in \mathbb{N}$, $\Delta_n = (1/c_n)\sum_{k=1}^{n} R_k c_k$, where

$$c_n := \prod_{k=1}^{n}\left(1 - \frac{1}{2k}\right)^{-1} = \prod_{k=1}^{n}\frac{2k}{2k-1} \sim \sqrt{\pi n}$$

by the Wallis formula. On the other hand,

$$\|R_n\|_2 = O\left(\frac{1}{n^2}\right)\|Z_n\|_2 \leqslant O\left(\frac{1}{n^2}\right)\|Z_n\|_2 = O\left(\frac{1}{n^2}\right),$$

so that $\sum_{n\geqslant 1} c_n\|R_n\|_2 < +\infty$. This in turn implies that $\sum_{n\geqslant 1} c_n\|R_n\|_1 < +\infty$ and, consequently, that $\sum_{n\geqslant 1} c_n R_n$ $\mathbb{P}$-a.s. converges in $\mathbb{R}^d$. Hence $\Delta_n \sim \lambda/\sqrt{n}$ and $\|\Delta_n\|_2 = O(1/\sqrt{n})$. □

Proof of Theorem 8. It follows from the above lemma that the η-empirical random measures related to the sequences $(Z_n)_{n\in\mathbb{N}}$ and $(X_n)_{n\in\mathbb{N}}$ $\mathbb{P}$-a.s. have the same weak limiting distributions. This is obvious once it is noticed that bounded Lipschitz functions characterize weak convergence of probability measures and that

$$\mathbb{P}\text{-a.s.} \qquad \lim_n \frac{1}{H_n}\sum_{k=1}^{n}\eta_k|Z_k - X_k| = 0.$$

Consequently, one may focus on the sequence $(X_n)_{n\in\mathbb{N}^*}$. We now apply Theorem 3 with $p = 1$, $b(x) := -x/2$ and $\sigma := I_d$, so that the diffusion is the Ornstein–Uhlenbeck process $\mathrm{d}Y_t = -\frac{1}{2}Y_t\mathrm{d}t + \mathrm{d}W_t$ with unique invariant distribution $\mathcal{N}(0, I_d)$. Assumption (28) is clearly satisfied with $V(x) := |x|^2 + 1$. □

6. Rate(s) of convergence

Throughout this section, we assume that diffusion (1) admits a unique invariant distribution ν. It then follows from Theorem 3 that, under some appropriate step–weight assumptions,

$$\mathbb{P}(\mathrm{d}\omega)\text{-a.s.} \qquad \nu_n^\eta(\omega, \mathrm{d}x) \Rightarrow \nu.$$

The aim of this section is to elucidate the rate of this convergence for a given pair (γ, η) of acceptable step and weight sequences. This rate will be evaluated along some smooth 'test functions'

$$f = A\varphi + C, \text{ with } \varphi \text{ at least twice differentiable and } C \text{ a real constant.}$$

In fact $\int A\varphi \mathrm{d}\nu$ is always 0 and one may assume without loss of generality that $C = 0$. We are looking for some *weak* rate, i.e. a sequence $\rho_n \to \infty$, depending only on γ and η, such that $\rho_n \nu_n^\eta(f)$ converges weakly towards a distribution depending on f (and the parameters b, σ of the diffusion). Then, we will be able to recommend some sequences γ and η that maximize the rate of convergence ρ_n for practical use.

A large part of what follows is carried out in the special case $\eta = \gamma$. A posteriori, this apparently restricted setting embodies the best possible rates of convergence when this rate ρ_n is associated with a regular CLT (see below). When it is not, the (slight) improvement induced by the choice of weights $\eta \neq \gamma$ will be clarified.

Our main results in this section are Theorems 9 and 10. In Theorem 9, we fully describe the global structure of the rates of convergence as a function of the step sequence γ: a 'reachable' rate of convergence ρ_n can usually be achieved either with 'fast-decreasing steps' γ_n, leading to a *regular CLT* in which $\rho_n \nu_n^\gamma(\omega, f)$ converges weakly towards a centred Gaussian measure, or with 'slowly decreasing steps', leading to a *convergence in probability* of $\rho_n \nu_n^\gamma(\omega, f)$ towards *a deterministic real constant* $m(f)$.

From a practical point of view, the choice of error type – *bias* or *variance* – is left to the user. There is only one exception: when the rate ρ_n is maximal, both phenomena get mixed and the rate of convergence holds as a *biased CLT*. This optimal rate is achieved at the boundary between fast and slowly decreasing steps, as expected. See Section 7 for more details.

Finally, we observe that it is in the 'slowly decreasing step' setting, when the rate of convergence holds in probability, that choosing some 'heavy' weights η_n different from the step γ_n can slightly increase the speed (only in terms of constant).

Theorem 9 describes what happens in the general setting: as little as possible is assumed about the white noise U_n. For practical simulations, the choice of the noise is left to the user and, among all possible choices, a noise with vanishing third moment such as Gaussian or Bernoulli seems quite appropriate. Now one can easily verify that the limiting parameter $m(f)$ in the above convergence in probability is zero when U_1 has null third moment. So in that case, the rate in probability provided by Theorem 9 is not the real one. The aim of Theorem 10 is to elucidate how the rate structure is modified in that case. It turns out that a regular CLT determines the global rate of convergence for a wider family of steps so that the optimal rate can be substantially improved. More precisely, for polynomial steps, the optimal rates are the following, according to the value of $\mathbb{E}(U_1^{\otimes 3})$: if $\mathbb{E}(U_1^{\otimes 3}) \neq 0$, then

$\eta_n = \gamma_n = n^{-1/2}$, which yields a (biased) CLT with a rate $\sqrt{\Gamma_n}$ proportional to $n^{1/4}$; if $\mathbb{E}(U_1^{\otimes 3}) = 0$, then $\eta_n = \gamma_n = n^{-1/3}$, which yields a (biased) CLT with a rate $\sqrt{\Gamma_n}$ proportional to $n^{1/3}$.

In the special case where $f := b$ (drift), these rates can be substantially improved by using some log-polynomial steps $\gamma_n := \ln^{-\alpha} n$, $\alpha > 0$ (if $U_1 \in L^{2p}$, $p > 1$) or $\alpha > 1$ (if $U_1 \in L^2$): for such steps a CLT holds at a $\sqrt{(n/\ln^\alpha n)}$ rate. These results have been confirmed by numerical experiments; see Bignone (1999) and Section 7. On the way, we retrieve as a by-product the CLT for the almost sure CLT (Section 6.2.2).

The main tool for this study will be the CLT for arrays of martingale increments (see Hall and Heyde 1980). This is the key to the proof of the technical Proposition 2 below. Henceforth, the distribution of U_1 will be denoted by μ.

6.1. Optimal choice of the weights in the CLT

In this subsection, we show why the optimal choice for the weight is always $\eta = \gamma$ so long as the final rate of convergence holds as a CLT. This is based on a very general CLT result obtained for the drift $b = A(I_d)$ of the diffusion which is, in some sense, the simplest non-constant 'test function'.

Summing up the original definition of the algorithm successively leads to

$$X_n - X_0 = \sum_{k=1}^{n} \gamma_k b(X_{k-1}) + \sum_{k=1}^{n} \sqrt{\gamma_k}\sigma(X_{k-1})U_k. \tag{33}$$

Hence

$$\frac{1}{\sqrt{\Gamma_n}} \sum_{k=1}^{n} \gamma_k b(X_{k-1}) = \frac{X_n - X_0}{\sqrt{\Gamma_n}} - \frac{1}{\sqrt{\Gamma_n}} \sum_{k=1}^{n} \sqrt{\gamma_k}\sigma(X_{k-1})U_k,$$

i.e.

$$\sqrt{\Gamma_n}\nu_n^\gamma(b) = \frac{X_n - X_0}{\sqrt{\Gamma_n}} - \frac{1}{\sqrt{\Gamma_n}} \sum_{k=1}^{n} \sqrt{\gamma_k}\sigma(X_{k-1})U_k. \tag{34}$$

Proposition 1. *Let* $p \in [1, +\infty)$. *Assume* $(\mathcal{L}_{V,p})$ *and* $\mathbb{E}|U_1|^{2p} < +\infty$. *Assume that* γ_n *satisfies* (3) *and that* $\sigma\sigma^* = o(V^{p/2})$. *Then*

$$\frac{1}{\sqrt{\Gamma_n}} \sum_{k=1}^{n} \gamma_k b(X_{k-1}) = \sqrt{\Gamma_n}\nu_n^\gamma(b) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \int_{\mathbb{R}^d} \sigma\sigma^* \mathrm{d}\nu\right).$$

Proof. We rely on (34). Since $\lim_{|x|\to+\infty} V(x) = +\infty$ and $\sup_{n\in\mathbb{N}^*} \mathbb{E}V(X_n) < +\infty$, $(X_n)_{n\in\mathbb{N}}$ is tight; then $(X_n - X_0)/\sqrt{\Gamma_n} \xrightarrow{\mathbb{P}} 0$. On the other hand, we derive from Theorem 3 (with $\eta = \gamma$ and $\rho = 1$) that $\sup_{n\in\mathbb{N}} \nu_n^\gamma(V^{p/2}) < +\infty$ a.s. It then follows from Proposition 2 below (with $W = V^{p/2}$) that $(1/\sqrt{\Gamma_n})\sum_{k=1}^{n} \sqrt{\gamma_k}\sigma(X_{k-1})U_k \xrightarrow{\mathcal{L}} \mathcal{N}(0, \int_{\mathbb{R}^d} \sigma\sigma^* \mathrm{d}\nu)$. □

Remark 8. The remarkable feature of this CLT for the drift is that it holds with no extra assumptions. This will not be the case for more general functions, but it will help us optimize the choice of the weights.

We now comment on the rate of convergence. Since there is no additional constraint on the step to obtain this CLT for the drift, one derives that the step sequence $\gamma_n := 1/\ln^\alpha n$, $\alpha > 0$, yields a rate $\sqrt{\Gamma_n} \sim \sqrt{n/\ln^\alpha n}$ which asymptotically gets close to $\sqrt{n}$ when $\alpha \to 0$.

On the other hand, starting from (33), one may also introduce the general weights η_n. A little algebra yields, still using the usual convention $\eta_0/\gamma_0 := 0$,

$$\nu_n^\eta(b) = \frac{\eta_n}{\gamma_n H_n} X_n - \frac{1}{H_n}\sum_{k=1}^n \Delta\left(\frac{\eta_k}{\gamma_k}\right) X_{k-1} - \frac{1}{H_n}\sum_{k=1}^n \frac{\eta_k}{\sqrt{\gamma_k}}\sigma(X_{k-1})U_k. \tag{35}$$

If the auxiliary weight sequence $\tilde{\eta}_n := \eta_n^2/\gamma_n$ and the step sequence γ_n satisfy the 'step–weight' assumptions of Theorem 3, then one shows that

$$\frac{1}{(\sum_{k=1}^n \eta_k^2/\gamma_k)^{1/2}} \sum_{k=1}^n \frac{\eta_k}{\sqrt{\gamma_k}}\sigma(X_{k-1})U_k$$

converges in distribution to $\mathcal{N}(0, \int \sigma\sigma^* \mathrm{d}\nu)$. Now, with obvious notation,

$$\frac{H_n}{\sqrt{\tilde{H}_n}}\nu_n^\eta(b) = \frac{\eta_n}{\gamma_n\sqrt{\tilde{H}_n}} X_n - \frac{1}{\sqrt{\tilde{H}}}\sum_{k=1}^n \Delta\left(\frac{\eta_k}{\gamma_k}\right) X_{k-1} - \frac{1}{\sqrt{\tilde{H}_n}}\sum_{k=1}^n \frac{\eta_k}{\sqrt{\gamma_k}}\sigma(X_{k-1})U_k. \tag{36}$$

Assumptions that would enable the convergence to 0 of the first two terms of the right-hand side of the equality are easy to state using the Kronecker lemma, but a little difficult to handle from a practical point of view. Thus, they hold if X_n is L^1-bounded (e.g. if $\underline{\lim}_{x\to\infty} V(x)/|x| > 0$) and $\gamma_n = n^{-\alpha}$, $0 < \alpha < 1$, $\eta_n = n^\beta$, $\beta \geqslant -1$). When they do hold, one has

$$\frac{H_n}{\sqrt{\tilde{H}_n}}\nu_n^\eta(b) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \int \sigma\sigma^* \mathrm{d}\nu\right).$$

The rate of convergence of the random integrals $\nu_n^\eta(b)$ is then $H_n/\sqrt{\tilde{H}_n}$.

Now the Cauchy–Schwarz inequality shows that

$$H_n = \sum_{k=1}^n \frac{\eta_k}{\sqrt{\gamma_k}}\sqrt{\gamma_k} \leqslant \left(\sum_{k=1}^n \frac{\eta_k^2}{\gamma_k}\right)^{1/2}\left(\sum_{k=1}^n \gamma_k\right)^{1/2},$$

$$\frac{H_n}{\sqrt{\tilde{H}_n}} \leqslant \sqrt{\Gamma_n}.$$

This shows that the fastest convergence in the CLT for the drift b holds for $\eta = \gamma$.

In fact, the same phenomenon occurs for the more general test functions $f = A\varphi$ investigated below: a similar martingale increment term, namely $\varphi'(X_{k-1})\sigma(X_{k-1})U_k$, appears (it is called N_n; see Lemma 8 and Proposition 2). When this term does determine the global rate of convergence of $\nu_n^\eta(f)$, this rate holds as a CLT. It is optimal when $\eta = \gamma$ as well, because the same reasoning as in (36) can be used. So, for this reason and for the sake of simplicity, we will focus on the case $\eta = \gamma$. When the term $\varphi'(X_{k-1})\sigma(X_{k-1})U_k$ no longer determines the global rate of convergence (this occurs for 'slowly decreasing steps'), setting the weights η_n equal to the steps γ_n is no longer optimal: some limited improvement can be obtained by considering very heavy weights (see Section 6.3.1 for some results and comments).

6.2. Rate of convergence for functions $f = A\varphi$

6.2.1. General result

In this section, we will provide a CLT for $\nu_n^\gamma(A\varphi)$ when φ is (at least) twice differentiable with bounded derivatives. The results below are more restrictive concerning the step assumptions than that obtained for the drift $b = A(I_d)$. There is no conflict between these results since Theorem 9 below determines the rate of convergence of $A(\varphi)$ for functions φ with Hessian not identically zero.

Throughout this section, we will use a new notation for partial sums of powers of the step: for every $\alpha > 0$, set $\Gamma_n^{(\alpha)} := \gamma_1^\alpha + \ldots + \gamma_n^\alpha$.

We will make extensive use of the following proposition which is a consequence of the CLT for martingale increments.

Proposition 2. *Assume $\nu_n^\gamma \xrightarrow{\mathcal{L}} \nu$ and $\sup_{n\in\mathbb{N}} \nu_n^\gamma(W) < +\infty$ a.s., where W is a positive continuous function on $\mathbb{R}^d$. Then, for any continuous vector field, $\zeta: \mathbb{R}^d \to \mathbb{R}^d$, satisfying $\lim_{|x|\to+\infty} |\zeta(x)|^2/W(x) = 0$,*

$$\frac{1}{\sqrt{\Gamma_n}}\sum_{k=1}^{n}\sqrt{\gamma_k}(\zeta(X_{k-1})|U_k) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \int_{\mathbb{R}^d}|\zeta|^2 \mathrm{d}\nu\right).$$

Proof. For every $n \in \mathbb{N}^*$ and every positive integer k, $k \leqslant n$, let $\xi_k^{(n)} := \sqrt{\gamma_k}(\zeta(X_{k-1})|U_k)/\sqrt{\Gamma_n}$. We have $\mathbb{E}(\xi_k^{(n)}|\mathcal{F}_{k-1}) = 0$ and

$$\sum_{k=1}^{n}\mathbb{E}(|\xi_k^{(n)}|^2/\mathcal{F}_{k-1}) = \frac{1}{\Gamma_n}\sum_{k=1}^{n}\gamma_k|\zeta(X_{k-1})|^2 = \nu_n^\gamma(|\zeta|^2).$$

Therefore, $\lim_{n\to\infty}\sum_{k=1}^{n}\mathbb{E}(|\xi_k^{(n)}|^2/\mathcal{F}_{k-1}) = \int|\zeta|^2\mathrm{d}\nu$ a.s.

The lemma will follow from the CLT for arrays of square-integrable martingale increments (see Hall and Heyde 1980) once we have checked the Lindeberg condition. For every $\varepsilon > 0$, set

$$K_\varepsilon(x, \Gamma) := \int_{\mathbb{R}^q}|u|^2\mathbf{1}_{\{|(\zeta(x)|u)|\geqslant\varepsilon\sqrt{\Gamma/\|\gamma\|_\infty}\}}\mu(\mathrm{d}u), \qquad x \in \mathbb{R}^d, \Gamma \geqslant 0.$$

We have $K_\varepsilon(x, \Gamma) \leqslant \mathbb{E}|U_1|^2$ and, for every positive number A, $\lim_{\Gamma\to+\infty}\sup_{|x|\leqslant A} K_\varepsilon(x, \Gamma) = 0$. Now let

$$R_n^\varepsilon := \sum_{k=1}^{n}\mathbb{E}\left(|\xi_k^{(n)}|^2\mathbf{1}_{\{|\xi_k^{(n)}|\geqslant\varepsilon\}}/\mathcal{F}_{k-1}\right).$$

We have, for every $A > 0$,

$$\begin{aligned} R_n^{\varepsilon} &\leqslant \sum_{k=1}^{n} \frac{\gamma_k}{\Gamma_n} |\zeta(X_{k-1})|^2 K_\varepsilon(X_{k-1}, \Gamma_n) \\ &\leqslant \frac{1}{\Gamma_n} \sum_{k=1}^{n} \gamma_k |\zeta(X_{k-1})|^2 K_\varepsilon(X_{k-1}, \Gamma_n) \mathbf{1}_{\{|X_{k-1}| \leqslant A\}} \\ &\quad + \frac{1}{\Gamma_n} \sum_{k=1}^{n} \gamma_k |\zeta(X_{k-1})|^2 K_\varepsilon(X_{k-1}, \Gamma_n) \mathbf{1}_{\{|X_{k-1}| > A\}} \\ &\leqslant C\nu_n^{\gamma}(W) \sup_{|x| \leqslant A} K_\varepsilon(x, \Gamma_n) + \sup_{|x|>A} \left(\frac{|\zeta(x)|^2}{W(x)} \right) \nu_n^{\gamma}(W). \end{aligned}$$

Letting $A \to +\infty$ completes the proof. □

Theorem 9. *Assume that* $(\mathcal{L}_{V,\infty})$ *holds and that the sequence* $(\gamma_n)_{n \geqslant 1}$ *is non-increasing.*

(a) Fast-decreasing step. *If* $\lim_n (1/\sqrt{\Gamma_n}) \sum_{k=1}^{n} \gamma_k^{3/2} = 0$ *and* $\mathbb{E}|U_1|^4 < +\infty$, *then, for every* $\mathcal{C}^2$ *function* φ *with* $D^2\varphi$ *bounded and Lipschitz and* $\lim_{|x| \to +\infty} |\sigma^*(x)\nabla\varphi(x)|^2 / V(x) = 0$, *the following CLT holds:*

$$\sqrt{\Gamma_n}\nu_n^{\gamma}(A\varphi) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \int_{\mathbb{R}^d} |\sigma^* \nabla\varphi|^2 \mathrm{d}\nu\right). \tag{37}$$

(b) Slowly decreasing step. *If* $\lim_n (1/\sqrt{\Gamma_n}) \sum_{k=1}^{n} \gamma_k^{3/2} = \tilde{\gamma} \in (0, +\infty]$ *and* $\mathbb{E}|U_1|^6 < +\infty$, *for every* $\mathcal{C}^3$ *function* φ *with* $D^2\varphi$ *and* $D^3\varphi$ *bounded and Lipschitz and* $\sup_{x \in \mathbb{R}^d} |\sigma^*(x)\nabla\varphi(x)|^2 / V(x) < \infty$, *we have*

$$\sqrt{\Gamma_n}\nu_n^{\gamma}(A\varphi) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\tilde{\gamma} m, \int_{\mathbb{R}^d} |\sigma^* \nabla\varphi|^2 \mathrm{d}\nu\right) \qquad \text{if } \tilde{\gamma} < \infty, \tag{38}$$

$$\frac{\Gamma_n}{\Gamma_n^{(3/2)}} \nu_n^{\gamma}(A\varphi) \xrightarrow{\mathbb{P}} m \qquad \text{if } \tilde{\gamma} = +\infty, \tag{39}$$

where

$$m := -\frac{1}{6} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^q} D^3\varphi(x).(\sigma(x)u)^{\otimes 3} \mu(\mathrm{d}u) \right) \nu(\mathrm{d}x).$$

Remark 9. One may slightly relax the moment assumption in (a) by simply assuming that $U_1 \in L^3$, provided the step assumption is strengthened to $\sum_{n \geqslant 1} \gamma_n^{3/2} / \sqrt{\Gamma_n} < +\infty$. The only noticeable change in the proof below lies in the treatment of $Z_n^{(3)}$. Instead of the L^2 argument given in the case $\mathbb{E}|U_1|^4 < +\infty$, we use the Chow theorem and Kronecker lemma to show that $Z_n^{(3)} / \sqrt{\Gamma_n}$ goes to 0 almost surely.

Remark 10. We derive the following results from Theorem 9.

For log-polynomial steps $\gamma_n := \ln^{-\alpha} n,\ \alpha > 1$, we have

$$\lim_n \frac{1}{\sqrt{\Gamma_n}} \sum_{k=1}^{n} \gamma_k^{3/2} = +\infty.$$

However, it is computationally easy to show that this leads to poor rates of convergence.

For polynomial steps $\gamma_n := n^{-\alpha},\ 0 < \alpha \leqslant 1$, some easy computations lead to

$$\frac{1}{\sqrt{\Gamma_n}} \sum_{k=1}^{n} \gamma_k^{3/2} \to \begin{cases} 0 & \text{if } \alpha \in (\frac{1}{2}, 1], \\ 2\sqrt{2} & \text{if } \alpha = \frac{1}{2}, \\ +\infty & \text{if } \alpha \in (0, \frac{1}{2}). \end{cases}$$

For this family of steps, the theorem reads as follows: if $\alpha = 1$, a CLT holds at rate $\sqrt{\Gamma_n} \sim \sqrt{\ln n}$; if $\alpha \in (\frac{1}{2}, 1)$, a CLT holds at rate $\sqrt{\Gamma_n} \sim n^{(1-\alpha)/2}/\sqrt{1-\alpha}$; if $\alpha = \frac{1}{2}$, a *biased* CLT holds, at rate $\sqrt{\Gamma_n} \sim \sqrt{2}n^{1/4}$; if $\alpha \in (0, \frac{1}{2})$, the rate of almost sure convergence is determined by

$$\frac{\Gamma_n}{\Gamma_n^{(3/2)}} \sim \frac{2 - 3\alpha}{2(1-\alpha)} n^{\alpha/2}.$$

This makes the best choice of step a little unclear: the optimal rate of convergence ($n^{1/4}$) is obtained in a situation where there is an unknown bias on the limiting distribution, whereas the range in which a centred CLT holds does not yield the optimal rate.

For the proof of Theorem 9, we first establish the following decomposition of $\nu_n^{\gamma}(A\varphi)$.

Lemma 8. *If φ is a twice continuously differentiable function on $\mathbb{R}^d$, then*

$$\Gamma_n \nu_n^{\gamma}(A\varphi) = \sum_{k=1}^{n} \gamma_k A\varphi(X_{k-1}) = Z_n^{(0)} - (N_n + Z_n^{(1)} + Z_n^{(2)} + Z_n^{(3)} + Z_n^{(4)}) \tag{40}$$

with

$$Z_n^{(0)} := \varphi(X_n) - \varphi(X_0) \quad \textit{and} \quad N_n := \sum_{k=1}^{n} \sqrt{\gamma_k}(\sigma(X_{k-1})U_k|\nabla\varphi(X_{k-1})),$$

$$Z_n^{(1)} := \frac{1}{2}\sum_{k=1}^{n} \gamma_k^2 D^2\varphi(X_{k-1})b^{\otimes 2}(X_{k-1}),$$

$$Z_n^{(2)} := \sum_{k=1}^{n} \gamma_k^{3/2}\langle D^2\varphi(X_{k-1});\, b(X_{k-1}),\, \sigma(X_{k-1})U_k\rangle,$$

$$Z_n^{(3)} := \frac{1}{2}\sum_{k=1}^{n} \gamma_k[D^2\varphi(X_{k-1})(\sigma(X_{k-1})U_k)^{\otimes 2} - \mathbb{E}(D^2\varphi(X_{k-1})(\sigma(X_{k-1})U_k)^{\otimes 2}/\mathcal{F}_{k-1})],$$

$$Z_n^{(4)} := \sum_{k=1}^{n} R_2(X_{k-1},\, X_k),$$

where $R_2(x,\, y) = \varphi(y) - \varphi(x) - (\nabla\varphi(x)|y - x) - \frac{1}{2}D^2\varphi(x).(y - x)^{\otimes 2}$.

Proof. We deduce from the definition of the algorithm that

$$\begin{aligned}\Delta\varphi(X_k) &= (\nabla\varphi(X_{k-1})|\Delta X_k) + \frac{1}{2}D^2\varphi(X_{k-1})(\Delta X_k)^{\otimes 2} + R_2(X_{k-1},\, X_k)\\ &= \gamma_k A\varphi(X_{k-1}) + \sqrt{\gamma_k}(\sigma(X_{k-1})U_k|\nabla\varphi(X_{k-1}))\\ &\quad + \frac{\gamma_k^2}{2}D^2\varphi(X_{k-1})b^{\otimes 2}(X_{k-1}) + \gamma_k^{3/2}\langle D^2\varphi(X_{k-1});\, b(X_{k-1}),\, \sigma(X_{k-1})U_k\rangle\\ &\quad + \frac{\gamma_k}{2}[D^2\varphi(X_{k-1})(\sigma(X_{k-1})U_k)^{\otimes 2} - \mathbb{E}(D^2\varphi(X_{k-1})(\sigma(X_{k-1})U_k)^{\otimes 2}/\mathcal{F}_{k-1})]\\ &\quad + R_2(X_{k-1},\, X_k).\end{aligned}$$

The lemma follows from summing the equality for $k = 1, \ldots, n$ and reordering the terms. □

The next lemma characterizes the behaviour of N_n.

Lemma 9. *Assume that* $(\mathcal{L}_{V,\infty})$ *holds and that, for some* $p \geqslant 1$, $\mathbb{E}|U_1|^{2p} < +\infty$ *and* $\sum_{n\geqslant 1}(\gamma_n/\Gamma_n^2)^{(1+\rho)/2} < +\infty$ *for some* $\rho \in (0, 1]$. *If* $\varphi \in \mathcal{C}^2(\mathbb{R}^d)$ *with*

$$\lim_{|x|\to+\infty} \frac{|\sigma^*(x)\nabla\varphi(x)|^2}{V^{p/(1+\rho)}(x)} = 0,$$

then

$$\frac{1}{\sqrt{\Gamma_n}}\sum_{k=1}^{n} \sqrt{\gamma_k}(\sigma(X_{k-1})U_k|\nabla\varphi(X_{k-1})) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\, \int_{\mathbb{R}^d} |\sigma^*\nabla\varphi|^2 \mathrm{d}\nu\right). \tag{41}$$

Proof. We deduce from Theorem 4 that $\sup_n \nu_n^\gamma(V^{p/(1+\rho)}) < \infty$. We also deduce from Theorem 3 that $\nu_n^\gamma \xrightarrow{\mathcal{L}} \nu$ a.s. It remains to apply Proposition 2 with $\zeta(x) = \sigma^*(x)\nabla\varphi(x)$ and $W = V^{p/(1+\rho)}$. □

For Theorem 9(b) we will need the following lemma.

Lemma 10. *Under the assumptions of Theorem 9(b), we have, in the notation of Lemma 8,*

$$\mathbb{P}\text{-}\lim_{n\to\infty} \frac{Z_n^{(4)}}{\Gamma_n^{(3/2)}} = \frac{1}{6}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} D^3\varphi(x)(\sigma(x)u)^{\otimes 3}\mu(\mathrm{d}u)\nu(\mathrm{d}x).$$

Proof. We have, using Taylor's formula,

$$R_2(x,\ y) = \tfrac{1}{6}D^3\varphi(x)(y-x)^{\otimes 3} + R_4(x,\ y),$$

with

$$|R_4(x,\ y)| \leqslant \frac{[D^3\varphi]_1}{24}|y-x|^4.$$

Hence

$$R_2(X_{k-1},\ X_k) = \tfrac{1}{6}D^3\varphi(X_{k-1})\Delta X_k^{\otimes 3} + r_k,$$

with

$$\begin{aligned}|r_k| &\leqslant \frac{[D^3\varphi]_1}{24}|\Delta X_k|^4 \leqslant C(\gamma_k^4|b(X_{k-1})|^4 + \gamma_k^2|\sigma(X_{k-1})U_k|^4)\\ &\leqslant C\gamma_k^2 V^2(X_{k-1})(1+|U_k|^4).\end{aligned}$$

Since $\mathbb{E}|U_1|^4 < +\infty$, we have, using Lemma 2, $\sup_n \mathbb{E}V^2(X_n) < \infty$. Therefore, $\mathbb{E}\sum_{k=1}^n |r_k| \leqslant C\sum_{k=1}^n \gamma_k^2$. From the assumption $\lim_{n\to\infty}\sum_{k=1}^n \gamma_k^{3/2}/\sqrt{\Gamma_n} = \tilde{\gamma} \in (0, +\infty]$, we deduce that $\lim_{n\to\infty}\sum_{k=1}^n \gamma_k^{3/2} = +\infty$ and $\lim_{n\to\infty}\sum_{k=1}^n \gamma_k^2/\Gamma_n^{(3/2)} = 0$. Therefore,

$$\frac{1}{\Gamma_n^{(3/2)}}\sum_{k=1}^n r_k \xrightarrow{\mathbb{P}} 0.$$

We now prove that

$$\frac{1}{\Gamma_n^{(3/2)}}\sum_{k=1}^n D^3\varphi(X_{k-1})(\Delta X_k)^{\otimes 3} \xrightarrow{\mathbb{P}} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} D^3\varphi(x)(\sigma(x)u)^{\otimes 3}\mu(\mathrm{d}u)\nu(\mathrm{d}x).$$

We have that $D^3\varphi(X_{k-1})(\Delta X_k)^{\otimes 3} = \gamma_k^{3/2}\Theta(\gamma_k,\ X_{k-1},\ U_k)$, where

$$\Theta(\gamma,\ x,\ u) = D^3\varphi(x)(\sqrt{\gamma}b(x) + \sigma(x)u)^{\otimes 3}.$$

Since $\mathbb{E}|U_1|^6 < \infty$, $\sup_n \mathbb{E}V^3(X_n) < \infty$, so that $\sup_n \mathbb{E}|\Theta(\gamma_n,\ X_{n-1},\ U_n)|^2 < \infty$, and we have (since $\lim_{n\to\infty}\Gamma_n^{(3)}/(\Gamma^{(3/2)})^2 = 0$),

$$\frac{1}{\Gamma_n^{(3/2)}}\sum_{k=1}^{n}\gamma_k^{3/2}[\Theta(\gamma_k, X_{k-1}, U_k) - \mathbb{E}(\Theta(\gamma_k, X_{k-1}, U_k)|\mathcal{F}_{k-1})] \xrightarrow{L^2} 0.$$

Observe that $\mathbb{E}(\Theta(\gamma_k, X_{k-1}, U_k)|\mathcal{F}_{k-1}) = J(X_{k-1}) + \sqrt{\gamma_k}\varepsilon(\gamma_k, X_{k-1})$, where J and ε are given by $J(x) := \int_{\mathbb{R}^q} D^3\varphi(x)(\sigma(x)u)^{\otimes 3}\mu(\mathrm{d}u)$ and $\varepsilon(\gamma, x) \leqslant CV^{3/2}(x)$.

We now apply Theorem 3 with $p = 3$, $\rho = 1$ and $\eta_k = \gamma_k^{3/2}$. Note that the sequence $\eta_k/\gamma_k = \gamma_k^{1/2}$ is non-increasing and that

$$\sum_{n\geqslant 1}\left(\frac{\eta_n}{H_n\sqrt{\gamma_n}}\right)^2 = \sum_{n\geqslant 1}\frac{\gamma_n^2}{(\Gamma_n^{(3/2)})^2} \leqslant C\sum_{n\geqslant 1}\frac{\gamma_n^{3/2}}{(\Gamma_n^{(3/2)})^2} < +\infty.$$

Therefore, we can assert that $\nu_n^\eta \xrightarrow{\mathcal{L}} \nu$ a.s. and that $\sup_n \nu_n^\eta(V^{3/2}) < \infty$ with probability 1. Since $J = o(V^{3/2})$, we conclude that $\lim_n \nu_n^\eta(J) = \int J\mathrm{d}\nu$ a.s., and the lemma follows easily. □

Proof of Theorem 9. Using the notation of Lemma 8, we first observe that, for any sequence of positive numbers $(a_n)_{n\geqslant 1}$ such that $\lim_{n\to\infty} a_n = +\infty$, we have $Z_n^{(0)}/a_n \xrightarrow{\mathbb{P}} 0$. Indeed, we know that the sequence $(X_n)_{n\in\mathbb{N}}$ is tight (because $\sup_n \mathbb{E}V(X_n) < \infty$). Since φ is continuous, the sequence $(\varphi(X_n))$ is tight as well.

We also derive from the definition of $Z_n^{(1)}$, $Z_n^{(2)}$ and $Z_n^{(3)}$ the inequalities

$$\mathbb{E}|Z_n^{(1)}| \leqslant C\sum_{k=1}^{n}\gamma_k^2\|D^2\varphi\|_\infty\mathbb{E}V(X_{k-1}), \tag{42}$$

$$\mathbb{E}|Z_n^{(2)}|^2 \leqslant C\sum_{k=1}^{n}\gamma_k^3\|D^2\varphi\|_\infty^2\mathbb{E}V^2(X_{k-1}), \tag{43}$$

$$\mathbb{E}|Z_n^{(3)}|^2 \leqslant C\sum_{k=1}^{n}\gamma_k^2\|D^2\varphi\|_\infty^2\mathbb{E}[V^2(X_{k-1})(1 + |U_k|^4)]. \tag{44}$$

(a) Now assume that $\lim_{n\to\infty}(1/\sqrt{\Gamma_n})\sum_{k=1}^n\gamma_k^{3/2} = 0$. We then have $\lim_{n\to\infty}\sum_{k=1}^n\gamma_k^2/\sqrt{\Gamma_n} = 0$, and it follows from (42) that $Z_n^{(1)}/\sqrt{\Gamma_n} \xrightarrow{L^1} 0$. We also deduce from (43) and (44) that $Z_n^{(j)}/\sqrt{\Gamma_n} \xrightarrow{L^2} 0$, for $j = 2, 3$. Here we use the assumption $\mathbb{E}|U_1|^4 < \infty$, which, due to Lemma 2, implies $\sup_n \mathbb{E}V^2(X_n) < \infty$.

We now study $Z_n^{(4)}$. Note that, due to our assumptions on φ,

$$|R_2(X_{k-1}, X_k)| \leqslant C|\Delta X_k|^3 \leqslant C\gamma_k^{3/2}V^{3/2}(X_{k-1})(1 + |U_k|^3).$$

Since $\mathbb{E}|U_1|^3 < \infty$, we have, using Lemma 2, $\sup_n \mathbb{E}V^{3/2}(X_n) < \infty$. The assumption $\lim_{n\to\infty}\sum_{k=1}^n\gamma_k^{3/2}/\sqrt{\Gamma_n} = 0$ now implies that $Z_n^{(4)}/\sqrt{\Gamma_n} \xrightarrow{L^1} 0$.

Finally, we apply Lemma 9 with $p = 2$ and $\rho = 1$ to obtain

$$\frac{N_n}{\sqrt{\Gamma_n}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \int|\sigma^*\nabla\varphi|^2\mathrm{d}\nu\right),$$

which completes the proof of part (a).

(b) We now assume that $(1/\sqrt{\Gamma_n})\sum_{k=1}^n\gamma_k^{3/2} = \tilde{\gamma} \in (0, +\infty]$. We then have that

$\lim_{n\to\infty} \Gamma_n^{(3/2)} = +\infty$. Therefore, $Z_n^{(0)}/\Gamma_n^{(3/2)} \xrightarrow{\mathbb{P}} 0$. It follows from (42) that $Z_n^{(1)}/\Gamma_n^{(3/2)} \xrightarrow{L^1} 0$, and from (43) and (44) that $Z_n^{(j)}/\Gamma_n^{(3/2)} \xrightarrow{L^2} 0$, for $j = 2, 3$. Applying Lemma 9 with $p = 3$ and $\rho = 1$, we have

$$\frac{N_n}{\sqrt{\Gamma_n}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \int |\sigma^* \nabla\varphi|^2 \mathrm{d}\nu\right). \tag{45}$$

We also know from Lemma 10 that

$$\frac{Z_n^{(4)}}{\Gamma_n^{(3/2)}} \xrightarrow{\mathbb{P}} \frac{1}{6}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} D^3\varphi(x)(\sigma(x)u)^{\otimes 3}\mu(\mathrm{d}u)\nu(\mathrm{d}x) = -m. \tag{46}$$

Now, if $\tilde{\gamma} < +\infty$, we have $Z_n^{(j)}/\sqrt{\Gamma_n} \xrightarrow{\mathbb{P}} 0$, for $j = 0, 1, 2, 3$, and

$$\frac{Z_n^{(4)}}{\sqrt{\Gamma_n}} \xrightarrow{\mathbb{P}} -\tilde{\gamma} m, \tag{47}$$

and (38) follows from (45) and (47).

If $\tilde{\gamma} = +\infty$, we have $Z_n^{(j)}/\Gamma_n^{(3/2)} \xrightarrow{\mathbb{P}} 0$, for $j = 0, 1, 2, 3$, and $N_n/\Gamma_n^{(3/2)} \xrightarrow{\mathbb{P}} 0$, and (39) follows from (46). This completes the proof of Theorem 9. □

6.2.2. An application: a CLT for the (standard) almost sure CLT

The above theorem yields a CLT for the almost sure CLT. It will follow from the fact that the step sequence $\gamma_n := 1/n$, $n \geqslant 1$, satisfies the step assumption of Theorem 9(a) and as a result the rate is $\sqrt{\ln n}$. The rate of convergence in the standard almost sure CLT has been studied by Csörgő and Horváth (1992) for real-valued i.i.d. random variables. Analogous results have been obtained by Chaâbane (1998) for real-valued martingales and by Maâouia (1998) for additive functionals of Markov processes. Chaâbane and Maâouia (2000) examine the rate of convergence for the so-called strong quadratic law of large numbers in the context of vector-valued martingales (see Section 6.3.1).

Proposition 3. *Denote by ν_d the standard d-dimensional Gaussian measure. If $U_1 \in L^3$, then, for every function $f \in \mathcal{C}^2(\mathbb{R}^d)$, such that f, Df and $D^2 f$ are bounded and Lipschitz continuous on $\mathbb{R}^d$,*

$$\sqrt{\ln n}\left(\frac{1}{\ln n}\sum_{k=1}^{n}\frac{1}{k} f\left(\frac{U_1 + \ldots + U_k}{\sqrt{k}}\right) - \int f(u)\nu_d(\mathrm{d}u)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_f^2), \tag{48}$$

with $\sigma_f^2 = \int |\nabla g(x)|^2 \nu_d(\mathrm{d}x)$, where g satisfies $Ag = f$ and $\int_{\mathbb{R}^d} g \mathrm{d}\nu_d = 0$, and A is the Ornstein–Uhlenbeck generator $A\varphi := \frac{1}{2}\Delta\varphi - \frac{1}{2}(x|\nabla\varphi)$.

It should be mentioned that the results of Csörgő & Horváth (1992) for real-valued random variables require less regularity for the function *f*. The proof of this proposition will follow from Theorem 9 and the following lemma.

Lemma 11. *Assume $f \in \mathcal{C}^2(\mathbb{R}^d)$, with f, Df and $D^2 f$ bounded and Lipschitz continuous on*

$\mathbb{R}^d$. *There exists a function* $g \in \mathcal{C}^2(\mathbb{R}^d)$, *with* Dg *and* $D^2 g$ *bounded and Lipschitz continuous, such that* $\int_{\mathbb{R}^d} g \mathrm{d}\nu_d = 0$ *and*

$$f - \int_{\mathbb{R}^d} f \mathrm{d}\nu_d = \frac{1}{2}\Delta g - \frac{1}{2}(x|\nabla g).$$

Proof. We can assume, without loss of generality, that $\int f \mathrm{d}\nu_d = 0$. Let $(T_t)_{t\geqslant 0}$ be the Ornstein–Uhlenbeck semigroup. The infinitesimal generator of $(T_t)_{t\geqslant 0}$ is A. We refer to Nualart (1995) for the basic properties of the Ornstein–Uhlenbeck semigroup. Note that our definition of the Ornstein–Uhlenbeck generator differs (by a factor of $\frac{1}{2}$) from that used in the context of Malliavin calculus. Denote by P_n the orthogonal projection on the nth Wiener chaos. We have $P_0 f = \int_{\mathbb{R}^d} f \mathrm{d}\nu_d = 0$ and

$$T_t f = \sum_{n=1}^{\infty} \mathrm{e}^{-nt/2} P_n f, \qquad t \geqslant 0.$$

It follows that the integral $\int_0^\infty T_t f \mathrm{d}t$ is convergent in L^2. Let $g = -\int_0^\infty T_t f \mathrm{d}t$. The function g may be unbounded but is in $L^2(\nu_d)$. We have

$$g = \sum_{n=1}^{\infty} \frac{-2}{n} P_n f \quad \text{and} \quad Ag = f,$$

and it remains to check the regularity properties of g. Recall Mehler's formula,

$$T_t f(x) = \int_{\mathbb{R}^d} f(\mathrm{e}^{-t/2}x + \sqrt{1-\mathrm{e}^{-t}}y)\nu_d(\mathrm{d}y).$$

One may easily derive from this formula the commutation relations

$$\frac{\partial}{\partial x_i} T_t f = \mathrm{e}^{-t/2} T_t\left(\frac{\partial f}{\partial x_i}\right).$$

Therefore

$$\frac{\partial g}{\partial x_i} = \int_0^\infty \mathrm{e}^{-t/2} T_t\left(\frac{\partial f}{\partial x_i}\right)\mathrm{d}t$$

and $\|\nabla g\|_\infty \leqslant 2\|\nabla f\|_\infty$. A similar argument shows that the second-order derivatives of g are bounded and Lipschitz. □

Proof of Proposition 3. Let g be as in Lemma 11. We use the notation of Section 5. Theorem 9(a) above, together with Remark 9, implies that (48) holds for the sequence X_k instead of $Z_k = (U_1 + \ldots + U_k)/\sqrt{k}$, and $f = Ag$. Now f is Lipschitz, so that

$$\sqrt{\ln n}\mathbb{E}\left|\frac{1}{\ln n}\sum_{k=1}^{n}\frac{1}{k}(f(Y_k) - f(Z_k))\right| \leqslant \frac{[f]_1}{\sqrt{\ln n}}\sum_{k=1}^{n}\frac{\mathbb{E}|\Delta_k|}{k} = O\left(\frac{1}{\sqrt{\ln n}}\right),$$

and the proposition follows easily. □

6.3. Rate of convergence for noise with zero third moment

One interesting feature of Theorem 8(b) is that, whenever $\tilde{\gamma} = +\infty$ and $\mathbb{E}((U_1^\ell)^3) = 0$ for every $\ell \in \{1, \ldots, q\}$, the limiting term is zero, i.e. the (weak) rate of convergence for 'large steps' is $o(\Gamma_n/\Gamma_n^{(3/2)})$. It seems natural, then, to investigate what really happens in this case.

Theorem 10. *Assume that $(\mathcal{L}_{V,\infty})$ holds, that the sequence $(\gamma_n)_{n\in\mathbb{N}}$ is non-increasing, with $\lim_{n\to\infty}(\sum_{k=1}^n \gamma_k^{3/2})/\sqrt{\Gamma_n} = +\infty$, and that $\mathbb{E}|U_1|^6 < +\infty$ and $\mathbb{E}(U_1^{\otimes 3}) = 0$.*

(a) Fast-decreasing step. *If $\lim_n (1/\sqrt{\Gamma_n})\sum_{k=1}^n \gamma_k^2 = 0$ then, for every $\mathcal{C}^3$ function φ with $D^2\varphi$ bounded, $D^3\varphi$ bounded and Lipschitz, and $\sup_{x\in\mathbb{R}^d} |\sigma^*.\nabla\varphi(x)|^2/V(x) < \infty$, we have*

$$\sqrt{\Gamma_n}\nu_n^\gamma(A\varphi) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \int_{\mathbb{R}^d} |\sigma^*\nabla\varphi|^2 \mathrm{d}\nu\right). \tag{49}$$

(b) Slowly decreasing step. *If $\lim_n (1/\sqrt{\Gamma_n})\sum_{k=1}^n \gamma_k^2 = \hat{\gamma} \in (0, +\infty]$ and $\mathbb{E}|U_1|^8 < +\infty$, then, for every $\mathcal{C}^4$ function φ, with $D^2\varphi$ and $D^3\varphi$ bounded, $D^4\varphi$ bounded and Lipschitz, and $\sup_{x\in\mathbb{R}^d} |\sigma^*.\nabla\varphi(x)|^2/V(x) < +\infty$,*

$$\sqrt{\Gamma_n}\nu_n^\gamma(A\varphi) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\hat{\gamma}\hat{m}, \int_{\mathbb{R}^d} |\sigma^*\nabla\varphi|^2 \mathrm{d}\nu\right) \qquad \text{if } \hat{\gamma} < +\infty, \tag{50}$$

$$\frac{\Gamma_n}{\Gamma_n^{(2)}}\nu_n^\gamma(A\varphi) \xrightarrow{\mathbb{P}} \hat{m} \qquad \text{if } \hat{\gamma} = +\infty. \tag{51}$$

where

$$\hat{m} := -\int_{\mathbb{R}^d}\left(\frac{1}{2}D^2\varphi(x)b(x)^{\otimes 2} + \Phi_4(x)\right)\nu(\mathrm{d}x), \tag{52}$$

with

$$\Phi_4(x) = \int_{\mathbb{R}^q}\left(\frac{1}{2}\langle D^3\varphi(x);\, b(x), (\sigma(x)u)^{\otimes 2}\rangle + \frac{1}{24}D^4\varphi(x)(\sigma(x)u)^{\otimes 4}\right)\mu(\mathrm{d}u).$$

Proof. (a) Using (40), (42), (43) and (44) and the assumption $\lim_n \sum_{k=1}^n \gamma_k^2/\sqrt{\Gamma_n} = 0$, we see that $Z_n^{(j)}/\sqrt{\Gamma_n}$ goes to zero in probability, for $j = 0, 1, 2, 3$. Since we know from Lemma 9 that $N_n/\sqrt{\Gamma_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \int |\sigma^*\nabla\varphi|^2\mathrm{d}\nu)$, it remains to show that $Z_n^{(4)}/\sqrt{\Gamma_n} \xrightarrow{\mathbb{P}} 0$.

We have, in the notation of the proof of Lemma 10,

$$R_2(X_{k-1}, X_k) = \frac{1}{6}D^3\varphi(X_{k-1})(\Delta X_k)^{\otimes 3} + r_k,$$

with

$$|r_k| \leqslant C\gamma_k^2 V^2(X_{k-1}(1 + |U_k|^4).$$

We also have

$$D^3\varphi(X_{k-1})(\Delta X_k)^{\otimes 3} = \gamma_k^{3/2}D^3\varphi(X_{k-1})(\sigma(X_{k-1})U_k)^{\otimes 3} + \rho(X_{k-1}, X_k),$$

with

$$|\rho(X_{k-1}, X_k)| \leqslant C\gamma_k^2 V^{3/2}(X_{k-1})(1+|U_k|^2).$$

Using the assumptions $\mathbb{E}|U_1|^6 < +\infty$ and $\mathbb{E}U_1^{\otimes 3} = 0$, we obtain

$$\left\| \sum_{k=1}^n \gamma_k^{3/2} D^3\varphi(X_{k-1})(\sigma(X_{k-1})U_k)^{\otimes 3} \right\|_{L^2} \leqslant C\sqrt{\sum_{k=1}^n \gamma_k^3}.$$

Since $\lim_{n\to\infty} \sum_{k=1}^n \gamma_k^3/\Gamma_n = 0$, we conclude that $\lim_{n\to\infty} \|Z_n^{(4)}/\sqrt{\Gamma_n}\|_{L^2} = 0$. This completes the proof of (a).

(b) If $\lim_n \sum_{k=1}^n \gamma_k^2/\sqrt{\Gamma_n} = \hat{\gamma} \in (0, +\infty]$, we easily deduce from (43) and (44) that $Z_n^{(j)}/\sqrt{\Gamma_n}$ goes to zero in probability, for $j = 0, 2, 3$. We observe that $Z_n^{(1)}/\Gamma_n^{(2)} = \frac{1}{2}\nu_n^\eta(\psi)$, if we set $\eta_k = \gamma_k^2$ and $\psi(x) = (D^2\varphi.b)(x)^{\otimes 2}$. Since $\psi = o(V^{3/2})$, we have

$$\frac{Z_n^{(1)}}{\Gamma_n^{(2)}} \xrightarrow{\mathbb{P}} \frac{1}{2}\int_{\mathbb{R}^d} D^2\varphi(x)b^{\otimes 2}(x)\mathrm{d}\nu(x).$$

We now examine $Z_n^{(4)}$. Using Taylor's formula and $\Delta X_k = \gamma_k b(X_{k-1}) + \sqrt{\gamma_k}\sigma(X_{k-1})U_k$, we derive

$$R_2(X_{k-1}, X_k) = \gamma_k^{3/2} a_3(X_{k-1}, U_k) + \gamma_k^2 a_4(X_{k-1}, U_k) + \bar{\rho}_k,$$

with

$$a_3(x, u) := \frac{1}{6} D^3\varphi(x)(\sigma(x)u)^{\otimes 3},$$

$$a_4(x, u) := \frac{1}{2}\langle D^3\varphi(x);\, b(x),\, (\sigma(x)u)^{\otimes 2}\rangle + \frac{1}{24} D^4\varphi(x)(\sigma(x)u)^{\otimes 4},$$

$$|\bar{\rho}_k| \leqslant C\gamma_k^{5/2} V^{5/2}(X_{k-1})(1+|U_k|^5).$$

It is clear that

$$\frac{1}{\Gamma_n^{(2)}} \sum_{k=1}^n \bar{\rho}_k \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \frac{1}{\sqrt{\Gamma_n}} \sum_{k=1}^n \bar{\rho}_k \xrightarrow{\mathbb{P}} 0.$$

The assumption on the third moments of U_1 implies that

$$\left\| \sum_{k=1}^n \gamma_k^{3/2} a_3(X_{k-1}, U_k) \right\|_{L^2}^2 = \frac{1}{36}\sum_{k=1}^n \gamma_k^3 \mathbb{E}|D^3\varphi(X_{k-1})(\sigma(X_{k-1})U_k)^{\otimes 3}|^2 \leqslant C\sum_{k=1}^n \gamma_k^3 = o(\Gamma_n).$$

Hence

$$\lim_{n\to\infty} \frac{1}{\sqrt{\Gamma_n}} \left\| \sum_{k=1}^n \gamma_k^{3/2} a_3(X_{k-1}, U_k) \right\|_{L^2} = 0.$$

Finally, we must study $\sum_{k=1}^{n}\gamma_k^2 a_4(X_{k-1}, U_k)$. We observe that

$$|a_4(x, u)| \leqslant CV^2(x)(1+|u|^4),$$

so that, using the assumption $\mathbb{E}|U_1|^8 < +\infty$, $\sup_k \mathbb{E}|a_4(X_{k-1}, U_k)|^2 < +\infty$. Therefore,

$$\left\| \sum_{k=1}^{n} \gamma_k^2 [a_4(X_{k-1}, U_k) - \mathbb{E}(a_4(X_{k-1}, U_k)/\mathcal{F}_{k-1})] \right\|_{L^2}^2 \leqslant C \sum_{k=1}^{n} \gamma_k^4$$

and

$$\frac{1}{\sqrt{\Gamma_n}} \sum_{k=1}^{n} \gamma_k^2 [a_4(X_{k-1}, U_k) - \mathbb{E}(a_4(X_{k-1}, U_k)/\mathcal{F}_{k-1})] \xrightarrow{L^2} 0.$$

We have $\mathbb{E}(a_4(X_{k-1}, U_k)/\mathcal{F}_{k-1}) = \Phi_4(X_{k-1})$, with $\Phi_4(x) = \int_{\mathbb{R}^q} a_4(x, u)\mu(\mathrm{d}u)$. We now apply Theorem 3 (with $\eta_k = \gamma_k^2,\ p = 4,\ \rho = 1$) to assert that

$$\lim_{n\to\infty} \frac{1}{\Gamma_n^{(2)}} \sum_{k=1}^{n} \gamma_k^2 \Phi_4(X_{k-1}) = \int_{\mathbb{R}^d} \nu(\mathrm{d}x)\Phi_4(x) \text{ a.s.}$$

Therefore,

$$\frac{Z_n^{(4)}}{\Gamma_n^{(2)}} \xrightarrow{\mathbb{P}} \int_{\mathbb{R}^d} \left(\frac{1}{2} D^2\varphi . b^{\otimes 2} + \Phi_4\right) \mathrm{d}\nu.$$

Using the weak convergence of $N_n/\sqrt{\Gamma_n}$, it is now easy to derive (50) and (51). □

When applying Theorem 10 to polynomial steps, one can easily verify that parts (a) and (b) of the theorem respectively lead, for $\gamma_n := n^{-\alpha}$, $0 < \alpha \leqslant \frac{1}{2}$, to the following weak rates of convergence. In the case $\frac{1}{3} < \alpha \leqslant \frac{1}{2}$, a CLT holds at the rate $\sqrt{\Gamma_n} \sim n^{(1-\alpha)/2}/\sqrt{1-\alpha}$ which enlarges the range of validity of the CLT formerly obtained in Theorem 9(a) for $\frac{1}{2} < \alpha < 1$ when the white noise is standard. For $\alpha = \frac{1}{3}$, a *biased* CLT holds at the rate $\sqrt{\Gamma_n} \sim \sqrt{\frac{3}{2}} n^{1/3}$. For $0 < a < \frac{1}{3}$, a weak $\Gamma_n/\Gamma_n^{(2)} \sim ((1-2\alpha)/(1-\alpha))n^{\alpha}$ rate holds.

6.3.1. Some further improvements

For functions $f = A(\varphi)$, with φ quadratic, one may obtain the improved rates of Theorem 10 even if the third moment of U_1 is not zero. In fact, the rate structure of Theorem 10 holds if $(\mathcal{L}_{V,\infty})$ holds, $\mathbb{E}|U_1|^{2p} < +\infty$ for some $p > 1$, and the sequence $(\gamma_n)_{n\in\mathbb{N}}$ is non-increasing and satisfies

$$\sum_{n\geqslant 1} \left(\frac{\gamma_n}{\Gamma_n^2}\right)^{(1+\rho)/2} < +\infty, \quad \text{for some } \rho \in (0, 1], \quad \text{and} \quad \sum_{n\geqslant 1} \left(\frac{\gamma_n}{\sqrt{\Gamma_n}}\right)^{p\wedge 2} < +\infty. \qquad (53)$$

The proof relies on the fact that, $D^3\varphi$ being identically 0, the third moment of U_1 no longer appears in the expansion. The method of proof is the same as above; details are left to the reader.

If we apply this result with $\varphi(x) = |x|^2 - d$ and $p > 2$, we recover the CLT for the strong quadratic law of large numbers of Chaâbane and Maâouia (2000), in the special case of i.i.d. random variables in L^{2p}. Indeed, in that case, $A\varphi(x) = -\varphi(x) = d - |x|^2$.

If we no longer assume that $\eta = \gamma$, we may slightly improve the rate of convergence for slowly decreasing steps by choosing arbitrarily heavy polynomial weights η_n, as shown by the following proposition.

Proposition 4. *Assume that* $\mathbb{E}(U_1^{\otimes 3}) = 0$. *If* $\gamma_n = cn^{-\alpha}$, $0 < \alpha < \frac{1}{3}$ *(slowly decreasing step) and* $\eta_n = n^{-\beta}$, $\beta \leqslant 1$, *then*

$$\gamma_n^{-1}\nu_n^{\eta}(A\varphi) \xrightarrow{\mathbb{P}} \frac{1-\beta}{1-(\alpha+\beta)}\hat{m},$$

where φ *and* $\hat{m}$ *are as in Theorem 10 and* φ *is bounded. One may verify that*

$$\min_{\beta \leqslant 1}\left|\frac{1-\beta}{1-(\alpha+\beta)}\hat{m}\right| = \lim_{\beta \to -\infty}\left|\frac{1-\beta}{1-(\alpha+\beta)}\hat{m}\right| = |\hat{m}|.$$

Finally, we may increase the rate of convergence still further for a subclass of test functions. In this paragraph we set $d = 1$ for the sake of simplicity. Furthermore, we assume that b and σ are smooth. Let us consider a white noise whose *first four moments* coincide with those of the standard normal distribution, i.e. $\mathbb{E}(U_1) = \mathbb{E}(U_1^3) = 0$, $\mathbb{E}(U_1^2) = 1$ and $\mathbb{E}(U_1^4) = 3$. Then a little algebra yields

$$\hat{m} = -\frac{1}{2}\int_{\mathbb{R}} bA\varphi' + \frac{\sigma^2}{2}A\varphi''\mathrm{d}\nu.$$

Set $A'g := b'g + \sigma\sigma' g''$. Formal computations show that, if there is some function Ψ satisfying $\frac{1}{4}\sigma^2 A'(\varphi') = A\Psi$, then

$$bA\varphi' + \frac{\sigma^2}{2}A\varphi'' = \frac{1}{2}A(\Psi - \Phi), \qquad \text{where } \Phi(x) = \int_0^x A\varphi'(u)\mathrm{d}u.$$

Consequently, $\hat{m} = 0$, which in turn implies that the rate of convergence in probability obtained in Theorem 10 is not the real one for such functions φ. In fact, this means that the CLT still holds for larger steps along this subclass of test functions.

7. Simulations and recommendations

7.1. Operating conclusions

The two schemes below show the rate of convergence in a polynomial scale n^{θ} as a function of the power α of the polynomial step $\gamma_n = n^{-\alpha}$.

- The fastest possible speed of convergence for generic test functions is $n^{1/3}$, achieved in the case $\eta_n = \gamma_n = n^{-1/3}$. This holds as a biased CLT (with unknown parameters).

- For a given rate, it seems that the 'slowly decreasing step' solution (convergence in probability) is more stable than the 'fast-decreasing step' side. This is illustrated by Figure 1.
- When convergence in probability holds, Proposition 4 shows the rate is improved by the use of heavy weights, i.e. $\eta_n := n^{-\beta}$ with $\beta \to -\infty$. This causes no problem for implementation since the recursive form of $\nu_n^\eta(f)$ only uses $\tilde{\eta}_n := \eta_n / H_n \sim \beta/n(\beta - 1)$. Numerical experiments show that the greater $|\beta|$ is, the later the rate improvement becomes significant. So the specification of β depends on the a priori order of the simulation size.

7.2. Some simulations

We choose to illustrate the rate structure obtained in Theorem 10, in particular to compare in a practical simulation how $n \mapsto \nu_n^\gamma(f)$ behaves with the steepness of the step sequence. To this end, we consider the one-dimensional standard Ornstein–Uhlenbeck process $\mathrm{d}Y_t = -\frac{1}{2}Y_t\mathrm{d}t + \mathrm{d}W_t$, and its Euler scheme with (decreasing) step implemented with a Gaussian white noise. The test function selected is

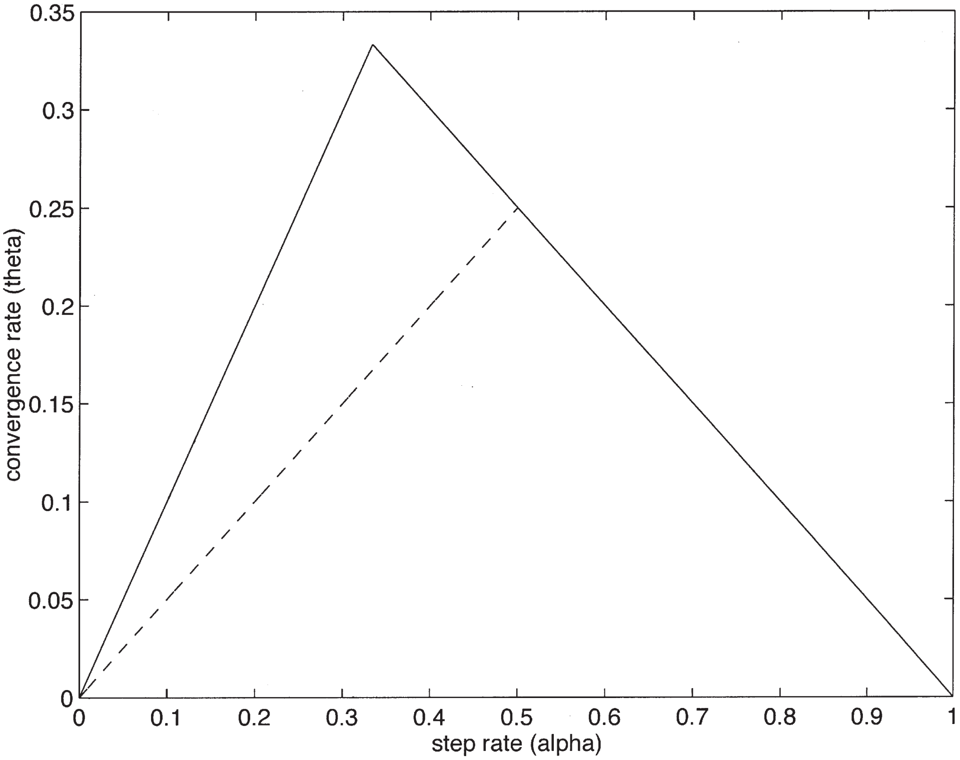

Figure 1. Plot of θ as a function of α from the rate of convergence n^θ when the step size is $\gamma_n = n^{-\theta}$

$$f(x) := A\left(\frac{1}{1+x^2}\right) = \frac{1}{1+x^2} + \frac{2}{(1+x^2)^2} - \frac{4}{(1+x^2)^3}.$$

Then, for a given reachable rate $\theta \in (0, \frac{1}{3}]$, we set

$$\gamma_n^{\text{fast}} := \frac{2\theta}{n^{1-2\theta}} \quad \text{and} \quad \gamma_n^{\text{slow}} := \frac{1-\theta}{1-2\theta}\frac{1}{n^\theta}.$$

In both cases, constants have been set so that the rate of convergence is equivalent to n^θ as $n \to +\infty$.

Numerical simulations have been carried out up to $n = 10^6$ iterations for the theoretical rate $\theta = 0.3$ with both fast (centred CLT) and slowly (bias) decreasing step sequences. Note that $\theta = 0.3$ is close to the optimal rate $\frac{1}{3}$. The same simulation has been processed simultaneously with the optimal polynomial step sequence $\gamma_n = n^{-1/3}$ (biased CLT with rate $n^{1/3}$).

In Figure 2 the thick line is for γ^{slow} with $\theta = 0.3$, the regular line is for γ^{fast} still with $\theta = 0.3$ and the dashed line is for $\gamma_n = n^{-1/3}$. One can verify that for a given theoretical rate of convergence, convergence in probability seems numerically more stable than the

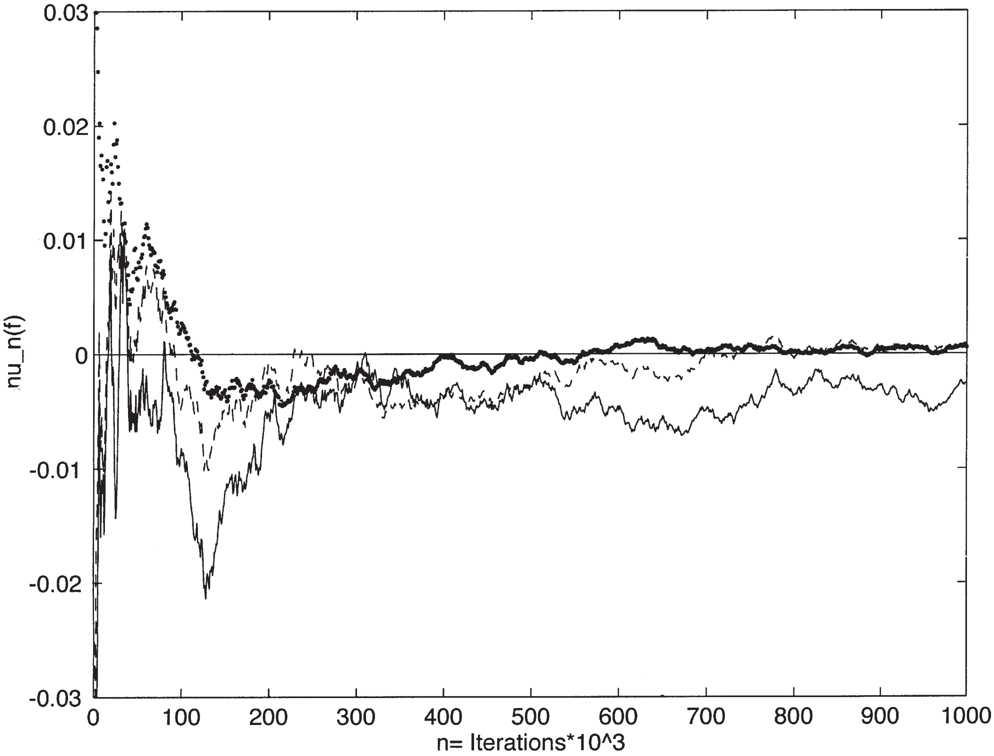

Figure 2. Rates of convergence of $\nu_n^\gamma(f)$ for different polynomial step functions

centred CLT. When $\theta = 0.3$, it competes with the optimal step sequence. This is confirmed by other simulations; see Pagès (2001) when the diffusion has several invariant distributions.

Note added in proof

After this paper was accepted for publication we learned of work by Piccioni and Scarlatti (1994) in which the mean square convergence of a similar algorithm for diffusions on compact Lie groups is studied.

References

Basak, G.K. and Bhattacharya, R.N. (1992) Stability in distribution for a class of singular diffusions. *Ann. Probab.*, **20**, 312–321.

Basak, G.K., Hu, I. and Wei, C.-Z. (1997) Weak convergence of recursions. *Stochastic Process. Appl.*, **68**, 65–82.

Berkes, I. and Csáki, E. (2001) A universal result in almost sure central limit theory. *Stochastic Process. Appl.*, **94**, 105–134.

Bignone, L. (1999) Validation numérique d'un schéma discret d'estimation de lois invariantes de diffusions browniennes pour un processus uni-dimensionnel d'Ornstein-Uhlenbeck. Technical report, Université de Marne-la-Vallée.

Brosamler, G.A. (1988) An almost everywhere central limit theorem. *Math. Proc. Cambridge Philos. Soc.*, **104**, 561–574.

Chaâbane, F. (1996) Version forte du théorème de la limite centrale fonctionnel pour les martingales. *C. R. Acad. Sci. Paris Sér. I Math.*, **323**, 195–198.

Chaâbane, F. (1998) Invariance principles with logarithmic averaging for martingales. Preprint.

Chaâbane, F. and Maâouia, F. (2000) Théorèmes limites avec poids pour les martingales vectorielles. *ESAIM Probab. Statist.*, **4**, 137–190.

Chaâbane, F., Maâouia, F. and Touati, A. (1996) Versions fortes associées aux théorèmes limites en loi pour les martingales vectorielles. Technical report, Université de Bizerte, Tunisia.

Csörgő, M. and Horváth, L. (1992) Invariance principles for logarithmic averages. *Math. Proc. Cambridge Philos. Soc.*, **112**, 195–205.

Ethier, S. and Kurtz, T. (1986) *Markov Processes. Characterization and Convergence*, Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

Fort, J.C. and Pagès, G. (1998) Stochastic algorithms with non constant step: a.s. weak convergence of weighted empirical measures. Preprint, Université Paris 12 Val de Marne.

Fisher, A. (1987) Convex invariant means and a pathwise central limit theorem. *Adv. Math.*, **63**, 213–246.

Ganidis, H., Roynette, B. and Simonot, F. (1999) Convergence rate of some semi-groups to their invariant probability. *Stochastic Process. Appl.*, **79**, 243–264.

Hall, P. and Heyde, C.C. (1980) *Martingale Limit Theory and Its Application*. New York: Academic Press.

Has'minskii, R.Z. (1980) *Stochastic Stability of Differential Equations.* Alphen a/d Rijn: Sijthoff & Noordhoff.

Heck, M.K. (1998) The principle of large deviations for the almost everywhere central limit theorem. *Stochastic Process. Appl.*, **76**, 61–76.

Karatzas, I. and Shreve, S. (1988) *Brownian Motion and Stochastic Calculus*. Berlin: Springer-Verlag.

Lacey, M.T. and Philip, W. (1990) A note on the almost sure central limit theorem. *Statist. Probab. Lett.*, **9**, 201–205.

Maâouia, F. (1998) Principes d'invariance par 'moyennisation' logarithmique pour les processus de Markov. Preprint, Faculté des Sciences, Université de Bizerte, Tunisia.

Neveu, J. (1972) *Martingales à Temps Discret*. Paris: Masson.

Nualart, D. (1995) *The Malliavin Calculus and Related Topics*. New York: Springer-Verlag.

Pagès, G. (2001) Sur quelques algorithmes récursifs pour les probabilités numériques. *ESAIM Probab. Statist.*, **5**, 141–170.

Pelletier, M. (1998) Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. *Ann. Appl. Probab.*, **8**, 10–44.

Pelletier, M. (1999) An almost sure central limit theorem for stochastic algorithms. *J. Multivariate Anal.*, **71**, 76–93.

Pelletier, M. (2000) Asymptotic almost sure efficiency of averaged stochastic algorithms. *SIAM J. Control Optim.*, **39**, 49–72.

Piccioni, M. and Scarlatti, S. (1994) An iterative Monte Carlo scheme for generating Lie group-valued random variables. *Adv. in Appl. Probab.*, **26**, 616–628.

Schatte, P. (1988) On strong versions of the central limit theorem. *Math. Nachr.*, **137**, 249–256.

Soize, C. (1995) Exact steady state solution of Fokker–Planck equations in high dimension for a class of non-linear Hamiltonian dissipative dynamical system excited by Gaussian white noise. In P. Krée and W. Wedig (eds), *Probabilistic Methods in Applied Physics*, Lecture Notes in Phys. 451, pp. 284–309. Berlin: Springer-Verlag.

Talay, D. (1990) Second order discretization of stochastic differential systems for the computation of the invariant law. *Stochastics Stochastics Rep.*, **29**, 13–36.

Talay, D. (2002) Stochastic Hamiltonian dissipative systems with non globally Lipschitz coefficients: exponential convergence to the invariant measure and discretization by the implicit Euler scheme. *Markov Process. Related Fields.* To appear.

Touati, A. (1995) Sur les versions fortes du théorème de la limite centrale. Technical report, Université de Marne-la-Vallée.

Received June 2000 and revised January 2002