The Annals of Applied Probability
2015, Vol. 25, No. 2, 986–1029
DOI: 10.1214/14-AAP1016

A DECREASING STEP METHOD FOR STRONGLY OSCILLATING STOCHASTIC MODELS

BY CAMILO ANDRÉS GARCÍA TRILLOS

Université Nice Sophia Antipolis

We propose an algorithm for approximating the solution of a strongly oscillating SDE, that is, a system in which some ergodic state variables evolve quickly with respect to the other variables. The algorithm profits from homogenization results and consists of an Euler scheme for the slow scale variables coupled with a decreasing step estimator for the ergodic averages of the quick variables. We prove the strong convergence of the algorithm as well as a C.L.T. like limit result for the normalized error distribution. In addition, we propose an extrapolated version that has an asymptotically lower complexity and satisfies the same properties as the original version.

1. Introduction. Consider a system of stochastic equations of the form

$$\begin{cases} X_t^\varepsilon = x_0 + \int_0^t f(X_s^\varepsilon, Y_s^\varepsilon)\,ds + \int_0^t g(X_s^\varepsilon, Y_s^\varepsilon)\,dW_s, \\ Y_t^\varepsilon = y_0 + \varepsilon^{-1}\int_0^t b(X_s^\varepsilon, Y_s^\varepsilon)\,ds + \varepsilon^{-1/2}\int_0^t \sigma(X_s^\varepsilon, Y_s^\varepsilon)\,d\tilde{W}_s, \end{cases} \tag{1}$$

where X_t^ε is a d_x-dimensional process, Y_t^ε a d_y-dimensional process, W and $\tilde{W}$ are two independent Brownian motions of dimensions d_x and d_y, and the functions b, σ, f and g have the right dimensions.

This type of system models the dynamics of two sets of interacting variables evolving in different time scales. The difference between time scales is controlled by the parameter ε. In many domains the most interesting case of study is that of the regime when $\varepsilon \ll 1$, that is, the situation in which X^ε is evolving very slowly compared to Y^ε (for this reason we will frequently denominate them as *slow scale* and *fast scale* variables, resp.). This regime may be studied by singular perturbation techniques as the ones presented in Bensoussan, Lions and Papanicolaou (1978) for deterministic models: instead of looking at the system with a small ε, we study the limit of (1) as $\varepsilon \to 0$ (when it exists) and estimate the error induced by this approximation.

There exist several analytical works with applications in different domains on the described type of approximation for stochastic models. For example in Majda,

Received October 2012; revised January 2014.
MSC2010 subject classifications. Primary 60H35; secondary 65C30.
Key words and phrases. Stochastic approximation, strongly oscillating, multi-scale system, ergodicity, limit distribution.

Timofeyev and Vanden-Eijnden (2001) a climate model is considered and studied on the advection scale (i.e., in the time scale of the slow variable). In Fouque, Papanicolaou and Sircar (2000) and Fouque et al. (2003) a system similar to (1) is presented and studied for pricing derivatives in the context of stochastic volatility models. A complete study with rather general hypothesis on the coefficients of the system is found in Pardoux and Veretennikov (2001) and Pardoux and Veretennikov (2003). In these papers a system with a fast scale ergodic diffusion is considered. More precisely, if

$$Y_t^x = y_0 + \int_0^t b(x, Y_s^x)\, ds + \int_0^t \sigma(x, Y_s^x)\, d\tilde{W}_s, \tag{2}$$

is ergodic with unique invariant measure μ^x, we might define the *effective equation*

$$X_t = x_0 + \int_0^t F(X_s)\, ds + \int_0^t G(X_s)\, dW_s, \tag{3}$$

with coefficients given by

$$\begin{aligned} F(x) &= \int f(x, y)\mu^x(dy), \qquad G(x) = \sqrt{H(x)}, \\ H(x) &= \int h(x, y)\mu^x(dy), \end{aligned} \tag{4}$$

where $h(x, y) = gg^*(x, y)$, and $G(x)$ could be in principle any matrix with square given by H, but we choose it to represent the Cholesky decomposition of the positive semi-definite matrix H. It follows that under appropriate assumptions $X^\varepsilon \xrightarrow{\mathcal{L}} X$ as $\varepsilon \to 0$; cf. Pardoux and Veretennikov (2003). The idea behind this kind of singular perturbation method is that when the difference between scales is large enough, the dynamics of the system behave as if the slow scale would be frozen and the ergodic limit of the fast diffusion would be attained.

However, except for a few particular examples, it is not an easy task to find explicit expressions for the averages (4). Naturally, this leads to the question of designing numerical methods of approximation of the effective equation. Several methods have been developed for a purely deterministic case; see, for example, the review E et al. (2007). Most of them are based on choosing a macro-solver for the slow scale in which some information from the fast scale is added via parameters' introduction to guarantee the correct approximation.

The literature with respect to numerical approximation of the general stochastic case is, to our knowledge, much more restricted. In E, Liu and Vanden-Eijnden (2005) the authors present an algorithm based on the use of an approximation scheme for the slow scale (e.g., the Euler scheme), and at each step of the slow scale another scheme is used to solve for the fast scale contribution; the weak and strong error induced by the scheme is analyzed when considering the particular case of an ODE with random coefficients slow scale equation and a stochastic ergodic fast scale variable [i.e., when $g(x, y) = 0$ in (1)].

In our work we use a similar approach. We focus on approaching numerically equation (3). With this objective in mind, we propose a *Multi-scale Decreasing Step (MsDS)* algorithm defined as a composition of an Euler scheme for the slow scale, the decreasing Euler step algorithm and estimator proposed in Lamberton and Pagès (2002) for the ergodic average approximation at each step, and a Cholesky decomposition for finding the volatility coefficient.

In order to control the total error approximation of this proposed algorithm we need to take into account four effects. First, we need an estimate on the ergodic average approximation at each step. We show that this control is based on the existence, regularity and control of the solution of the Poisson equation associated to the fast scale diffusion

$$\mathcal{L}_y^x \phi_\psi(x, y) = \psi(x, y), \tag{5}$$

where

$$\mathcal{L}_y^x := \frac{1}{2} \sum_{i,j=1}^{d_y} a_{ij}(x, y) \frac{\partial^2}{\partial y_i \, \partial y_j} + \sum_{i=1}^{d_y} b_i(x, y) \frac{\partial}{\partial y_i} \tag{6}$$

with $a := \sigma\sigma^*$, when considering as sources (i.e., the right-hand side functions) the coefficients F and H centered with respect to their respective invariant measures. Second, we need to control the error obtained after performing a Cholesky decomposition. Then, we have to account for discretization errors. Finally, we need to control the error propagation which will be possible under some growth control on the coefficients of the effective equation.

The MsDS algorithm strongly converges to the exact solution and proves to be more efficient than a simple Euler scheme for highly oscillating problems. Moreover, it features a nonstandard C.L.T. property in the sense that the normalized error distribution converges toward the solution of an SDE. The coefficients appearing in this normalized error SDE depend on the solution of the previously mentioned Poisson problem and are, in general, unknown. Nevertheless, the available explicit expression for them is valuable for the estimation of confidence intervals and eases the task of parameter tuning for actual implementation of the algorithm.

We study as well an *extrapolated MsDS (EMsDS)* version of the algorithm, differing from the original one in that it uses a Richardson–Romberg extrapolation of the decreasing step estimator (i.e., a well-chosen linear combination of the decreasing step Euler estimator with appropriate parameters) to approach the ergodic averages. As the MsDS, the EMsDS also features a nonstandard C.L.T. property and shares the same rate of convergence. However, the extrapolated version has lower asymptotic complexity and hence higher asymptotic efficiency than the original one.

1.1. *Outline of the paper.* The organization of the paper is as follows: in Section 2, we describe the algorithm and state the standing hypothesis and our main results (strong convergence, limit distribution). The proof of the main theorem is presented in Section 4 after having reminded some regularity properties of the effective equation and available results on the decreasing Euler estimation algorithm in Section 3. We extend the main results to an extrapolated version of the algorithm that we introduce and study in Section 5. Finally, we perform some numerical studies in Section 6. The paper ends with an Appendix containing the proof of a couple of technical results.

2. The MsDS algorithm. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and W be an $\mathcal{F}$-adapted Brownian motion. Suppose we are given an independent probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ and a family of independent Brownian motions $\tilde{W}^q$, $q \in \mathbb{Q}$ with an associated filtration $\tilde{\mathcal{F}}_t^q := \sigma\{\tilde{W}_s^q, s \leq t\}$. Define the extended space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathcal{F}}_t, \bar{\mathbb{P}})$ by

$$\bar{\Omega} := \Omega \times \tilde{\Omega}, \qquad \bar{\mathbb{P}}(d\omega, d\bar{\omega}) = \mathbb{P}(d\omega)\tilde{\mathbb{P}}(d\bar{\omega}),$$
$$\bar{\mathcal{F}} := \mathcal{F} \otimes \tilde{\mathcal{F}}, \qquad \tilde{\mathcal{F}}_t^q := \bigvee_{q \in \mathbb{Q}; q \leq t} \tilde{F}_\infty^q, \qquad \bar{F}_t := \mathcal{F}_t \vee \tilde{\mathcal{F}}_t^q.$$

Such extended space will be useful for treating independently the noise coming from the Brownian in the effective diffusion and the one related to the approximation of the ergodic diffusion averages. Consider the decreasing step Euler algorithm introduced in Lamberton and Pagès (2002) to approach the invariant measure of a recursive diffusion. Let $\{\gamma_k\}_{k\in\mathbb{N}}$ be a decreasing sequence of steps satisfying:

HYPOTHESIS $(\mathcal{H}_\gamma)$ (On the sequence of steps for the average estimation algorithm).

(i) $\gamma_k > 0$ *for all* k;
(ii) γ_k *is a sequence of decreasing steps with* $\lim_{n\to\infty} \gamma_k = 0$;
(iii) $\lim_{k\to\infty} \Gamma_k = \infty$; *where* $\Gamma_k := \sum_{j=0}^{k} \gamma_j$;
(iv) $\sum_{k=1}^{\infty} (\frac{\gamma_k^2}{\Gamma_k}) < +\infty$.

For any $q \in \mathbb{Q}$, let $\sqrt{\gamma_{k+1}} U_{k+1}^q := \tilde{W}_{\Gamma_{k+1}}^q - \tilde{W}_{\Gamma_k}^q$ so that U_{k+1} is a standard Gaussian vector. Let $y_0 \in \mathbb{R}^{d_y}$. We define the *decreasing step Euler approximation of the ergodic diffusion* by

$$\begin{aligned} &\tilde{Y}_0^{x,q} = y_0, \\ &\tilde{Y}_{k+1}^{x,q} = \tilde{Y}_k^{x,y_0,q} + \gamma_{k+1} b\big(x, \tilde{Y}_k^{x,q}\big) + \sqrt{\gamma_{k+1}} \sigma\big(x, \tilde{Y}_k^{x,q}\big) U_{k+1}^q, \end{aligned} \tag{7}$$

and the *decreasing step average estimator* by

$$\tilde{F}^k(x, q) = \frac{1}{\Gamma_k} \sum_{j=1}^{k} \gamma_j f\big(x, \tilde{Y}_{j-1}^{x,q}\big). \tag{8}$$

The idea behind the particular form of estimator (8) is to take advantage of the ergodicity of the diffusion: the long-term time average approaches the invariant measure of the diffusion. Note that the estimator can also be written recursively as

$$\tilde{F}^0(x,q) = 0; \qquad \tilde{F}^k(x,q) = \tilde{F}^{k-1}(x,q) + \frac{\gamma_k}{\Gamma_k}\bigl(f(x, \tilde{Y}^{x,q}_{k-1}) - \tilde{F}^{k-1}(x,q)\bigr).$$

Evidently, using the same ergodic average argument, it is also possible to use a uniform step estimator of the type $k^{-1}\sum_{j=1}^{k}\gamma_j f(x, \tilde{Y}^{x,q}_{j-1})$ as studied, for example, in Talay (1990). The main difference between both estimators appears in the type of error that they generate. The uniform step estimator induces two types of errors coming from the truncation of the series and the fact that the *ergodic limit of the approached sequence is not the ergodic limit of the original diffusion*. In contrast, the decreasing Euler scheme estimator eliminates the asymptotic gap between the invariant law of the continuous equation and that of its discretization; see Lamberton and Pagès (2002). Moreover, the decreasing step method features a kind of "error expansion" [as shown in Lemaire (2005)] when applied to a certain family of functions. These properties are important to show the limit properties of our algorithm and to deduce the extrapolated version.

We should remark that we have chosen to work with a simplified version of the algorithm in Lamberton and Pagès (2002): its more general version allows the use of different sequences for the Euler scheme step and for the weights in the average.

With this estimator in hand we can define an Euler scheme to approach our effective diffusion. Assuming a time horizon T, for $n \in \mathbb{N}^*$ we put $t_k = Tk/n$, so that the Euler scheme will be given by

$$\check{X}^n_{t_{k+1}} = \check{X}^n_{t_k} + \tilde{F}^{M(n)}(\check{X}^n_{t_k}, t_k)\Delta t_{k+1} + \tilde{G}^{M(n)}(\check{X}^n_{t_k}, t_k)\Delta W_{k+1},$$

where $\tilde{F}^M$ is defined in (8) and $\tilde{G}^M(x,q)$ is defined in two steps: First we find $\tilde{H}^M(x,q)$ using the decreasing step algorithm as in (8) [recall that $h(x,y) = g^*g(x,y)$], and then we perform a Cholesky decomposition on it to find $\tilde{G}^M(x,q) = \sqrt{\tilde{H}^M(x,q)}$. Note that the number of steps in the decreasing Euler estimator, M, is expressed as a function of the number of steps in the Euler scheme for the slow scale n. The form of $M(n)$ will be clear from the main theorems.

It will be easier to work mathematically with a continuous interpolation of the Euler approximation. Let us denote by $\underline{t}(n) = \lfloor nt \rfloor / n$. We will usually omit the explicit dependence on n and write $\underline{t}$ when clear from the context. The continuous Euler approximation is then given by

$$\tilde{X}^n_t = x_0 + \int_0^t \tilde{F}^{M(n)}(\tilde{X}^n_{\underline{s}}, \underline{s})\,ds + \int_0^t \tilde{G}^{M(n)}(\tilde{X}^n_{\underline{s}}, \underline{s})\,dW_s, \tag{9}$$

that is, a linear interpolation from the discrete Euler scheme. Clearly, at times t_k the continuous Euler coincide with the Euler algorithm. All our results will be derived for the continuous version of the algorithm.

2.1. *Standing hypothesis and main result.* Let us introduce the assumptions under which our main results follow.

HYPOTHESIS ($\mathcal{H}_{s.s.}$) (On the slow-scale coefficients).

(i) *Lipschitz in* x: *There exist constants* K, m *such that for all* $x, x' \in \mathbb{R}^{d_x}$ *and* $y \in \mathbb{R}^{d_y}$,

$$|f(x,y) - f(x',y)| + |g(x,y) - g(x',y)| \le K|y|^m|x - x'|;$$

(ii) *regularity*: f, h *belong to* $C^{2,r^y}_{b,p}$ *for some* $r^y > 3$, *where the subindex* b, p *means the derivatives* $\partial^i_x \partial^j_y$ *for* $0 \le i \le 2$ *and* $0 \le j \le r^y - i$ *are bounded in* x *and polynomially bounded in* y;

(iii) *degeneracy*: *either* h *is identically zero*, *or it is uniformly nondegenerate*, *that is*, *there exists* $\lambda'_- \in \mathbb{R}^+_*$ *such that* $\lambda'_- I \le h(x,y)$.

Before giving the standing hypothesis on the fast scale equation, recall that we have defined the matrix $a(x,y) = \sigma\sigma^*(x,y)$.

HYPOTHESIS ($\mathcal{H}_{f.s.}$) (On the fast-scale coefficients).

(i) $a, b \in C^{2,0}_{b,l}$, *that is*, *they are continuous and linearly bounded in* y *and* C^2 *and bounded in* x.

(ii) *The matrix* a *is uniformly continuous and uniformly nondegenerate and bounded*, *that is*, *there exist* $\lambda_-, \lambda_+ \in \mathbb{R}^+_*$ *such that*

$$\lambda_- I \le a(x,y) \le \lambda_+ I;$$

(iii) $\sup_x b(x,y) \cdot y \le -c_1|y|^2 + c_2$, *for some* $c_1 \in \mathbb{R}^*_+, c_2 \in \mathbb{R}$.

The regularity and growth hypothesis contained in ($\mathcal{H}_{s.s.}$) are assumed to control the error propagation. The main goal of imposing conditions on the fast scale diffusion is to guarantee the existence of an invariant limit for any possible fixed value of x and a uniform control on its averages. For this reasons they are quite restrictive: note that ($\mathcal{H}_{f.s.}$)(i) implies $\sup_x |b(x,y)| = O(|y|)$ and ($\mathcal{H}_{f.s.}$)(iii) deduces $\lim_{|y|\to\infty} \sup_x b(x,y) \cdot y = -\infty$, meaning that the drift has at most linear growth in y and that it is mean reverting uniformly in x. In turn, the ellipticity and nondegeneracy assumption ($\mathcal{H}_{f.s.}$)(ii) is helpful to deduce the uniqueness of the invariant measure.

We are ready to state our main Theorem on the MsDS algorithm. Its proof is found in Section 4.

THEOREM 2.1. *Let* $0 < \theta < 1$, $\gamma_0 \in \mathbb{R}^+$ *and* $\gamma_k = \gamma_0 k^{-\theta}$. *Let* M_1 *be a positive constant*. *Assume* ($\mathcal{H}_{f.s.}$) *and* ($\mathcal{H}_{s.s.}$). *Define* $M(n)$ *by*

$$M(n) = \lceil M_1 n^{1/(1-\theta)} \rceil,$$

then:

(i) *ODE with random coefficients case* $[g(x, y) \equiv 0]$:

(a) (*Strong convergence*). *There exists a constant* K *such that*

$$\mathbb{E}\Big[\sup_{0 \le t \le T} |X_t - \tilde{X}_t^n|^2\Big] \le K n^{-2[(1-\theta)\wedge\theta]/(1-\theta)}.$$

(b) (*Limit distribution of the error*). *Assume in addition that* $r^y \ge 7$ *and* $\theta \ge 1/2$. *Then*

$$n(X - \tilde{X}^n) =: \zeta^n \Rightarrow \zeta^\infty,$$

where $\Rightarrow$ *denotes convergence in law, and* ζ^∞ *is the solution of an SDE stated explicitly on Theorem* 4.12.

(ii) *Full SDE case*:

(a) (*Strong convergence*). *There exists a constant* K *such that*

$$\mathbb{E}\Big[\sup_{0 \le t \le T} |X_t - \tilde{X}_t^n|^2\Big] \le K n^{-[(1-\theta)\wedge 2\theta]/(1-\theta)}.$$

(b) (*Limit distribution of the error*). *Assume in addition that* $r^y \ge 7$ *and* $\theta \ge 1/3$. *Then*

$$n^{1/2}(X - \tilde{X}^n) =: \zeta^n \Rightarrow \zeta^\infty,$$

where ζ^∞ *is the solution of an SDE stated explicitly on Theorem* 4.12.

Note that we study the mean square error of our approximation algorithm toward the effective equation. We perform this strong error analysis to guarantee that the algorithm will be used for applications demanding to approach functions that depend on the whole trajectory (as in finance). As will be clear from Theorem 4.12, the SDE defining the limit results both for the fully stochastic and the ODE with random coefficients case are explicitly given in terms of the invariant law of the ergodic diffusion and are consequently unknown. Nevertheless, the key point is that, being explicit, they might be estimated numerically for practical purposes.

We have announced an extrapolated version of the algorithm. Given that its proper introduction requires a further understanding of the basic algorithm, we postpone the presentation to Section 5.

3. Preliminaries. In this section we present the main tools needed to analyze the presented algorithm.

Let us start by stating properly the stochastic approximation theorem we mentioned in the Introduction and that justifies the relation between the effective equation (3) and the original strongly oscillating system (1).

THEOREM 3.1 [Theorem 4 in Pardoux and Veretennikov (2003)]. *Let* b, σ, f, g *be defined as in* (1) *and* $a = \sigma\sigma^*$. *Assume we have a recurrence condition of the type* $\lim_{|y|\to\infty} b(x, y) \cdot y = -\infty$, *and that the matrix "a" is nondegenerate* *and uniformly elliptic. Assume that* $a, b \in C_b^{2,1+\alpha}$, *and that* f, g *are Lipschitz with respect to the* x *variable uniformly in* y *and have at most polynomial growth in* y *and linear growth in* x.

Then, for any $T > 0$, *the family of processes* $\{X_t^\varepsilon, 0 \le t \le T\}_{0<\varepsilon\le 1}$ *is weakly relatively compact in* $C([0,T]; \mathbb{R}^l)$. *Any accumulation point* X *is a solution of the martingale problem associated with the operator* $\bar{\mathcal{L}}$.

If moreover, the martingale problem is well posed, then $X^\varepsilon \xrightarrow{\mathcal{L}} X$, *where* X *is the unique (in law) diffusion process with generator* $\bar{\mathcal{L}}$.

It is worth mentioning that the actual framework of Pardoux and Vertennikov's statement includes the case in which there is an ε^{-1} order term in the slow variable, which complicates the proof with respect to the framework we present here. Note that under the standing hypothesis, the martingale problem is well posed and X in the theorem is the unique solution to (3).

3.1. *A priori estimates*. An important result is related to some a priori estimates valid for general SDEs. Since they are quite standard, we will state the result without giving the details of the proof.

PROPOSITION 3.2. *Let*

$$\vartheta_t = \vartheta_0 + \int_0^t V_1(\vartheta_s, s)\,ds + \int_0^t V_2(\vartheta_s, s)\,dW_s, \tag{10}$$

where V_1, V_2 *are adapted random functions*.

(i) *For all* $\alpha \ge 2$,

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{0\le t\le T} |\vartheta_t|^\alpha\Big] \\
&\quad \le K_\alpha \mathbb{E}\big[|\vartheta_0|^\alpha\big] + K(\alpha, T)\int_0^T \big(\mathbb{E}\big[\big|V_1(\vartheta_s, s)\big|^\alpha\big] + \mathbb{E}\big[\big|V_2(\vartheta_s, s)\big|^\alpha\big]\big)\,ds \\
&\quad \le K_\alpha \mathbb{E}\big[|\vartheta_0|^\alpha\big] \\
&\qquad + K'(\alpha, T)\Big(\sup_{0\le t\le T} \mathbb{E}\big[\big|V_1(\vartheta_t, t)\big|^\alpha\big] + \sup_{0\le t\le T} \mathbb{E}\big[\big|V_2(\vartheta_t, t)\big|^\alpha\big]\Big).
\end{aligned}$$

(ii) *Assume that* $\forall \alpha \ge 2$,

$$\mathbb{E}\big[\big|V_1(\vartheta_t, t)\big|^\alpha\big] + \mathbb{E}\big[\big|V_2(\vartheta_t, t)\big|^\alpha\big] \le K\big(1 + \mathbb{E}[|\vartheta_t|]^\alpha\big).$$

Then:

(a) *for* $t \in [0, T]$ *and* $\alpha \ge 2$, $\mathbb{E}[|\vartheta_t|^\alpha] \le K(\alpha, T)$;

(b) *for* $\alpha \ge 2$, $\mathbb{E}[\sup_{0\le s\le t} |\vartheta_s|^\alpha] \le K(\alpha, T)\ \mathbb{P}(\sup_{0\le s\le t} \tau_r \le t) \le \frac{K'(\alpha,t)}{r^\alpha}$.

3.2. *Cholesky decomposition*. The Cholesky decomposition of a positive definite matrix consists of expressing this matrix as the product of a lower triangular matrix and its conjugate transpose. A stability analysis of this procedure is a key point in our analysis for the SDE case behavior of our algorithm.

Recall that we denote by $|\cdot|$ the induced operator norm. Let us denote by $\|\cdot\|_F$ the Frobenius norm. Recall that if H is a $d \times d$ matrix,

$$|H| \le \|H\|_F \le \sqrt{d}|H|. \tag{11}$$

THEOREM 3.3 [Theorem 1.1 in Sun (1991)]. *Let H be a $d \times d$ positive definite matrix with Cholesky factorization $H = GG^*$. If ΔH is a $d \times d$ symmetrical matrix satisfying $|H^{-1}|\|\Delta H\|_F < 1/2$, then there is a unique Cholesky factorization $H + \Delta H = (G + \Delta G)(G + \Delta G)^*$ and*

$$\frac{\|\Delta G\|_F}{|G|} \le \sqrt{2}\frac{\kappa\kappa_2(H)}{1 + \sqrt{1 - 2\kappa_2(H)\kappa}}, \tag{12}$$

where $\kappa = |\Delta H\|_F |H|^{-1}$ and $\kappa_2(H) = |H||H^{-1}|$.

Theorem 3.3 gives a control on the sensitivity of the Cholesky procedure. In Lemma 3.4 we study the propagation effect at each stage of the Cholesky factorization to say a little bit more on the particular form of the error. Its proof is given in Appendix B.

LEMMA 3.4. *Suppose the hypothesis of Theorem 3.3 holds. Then*

$$\Delta G_{i,i} = \frac{\Delta H_{i,i} - 2\sum_{k=1}^{i-1} \Delta G_{i,k} G_{i,k}}{2G_{i,i}} + O(|\Delta H|^2),$$

$$\Delta G_{i,j} = \frac{\Delta H_{i,j} - G_{i,j}\Delta G_{j,j} \sum_{k=1}^{j-1}(\Delta G_{j,k} G_{i,k} + \Delta G_{i,k} G_{j,k})}{G_{j,j}} + O(|\Delta H|^2)$$

for $i > j$.

Lemma 3.4 gives a first order approximation of the error matrix ΔG knowing the perturbation matrix ΔH. From this lemma, we can deduce on the regularity of the Cholesky approximation. The following corollary follows from the definition of H and Lemma 3.4.

COROLLARY 3.5. *Let $H : \mathbb{R}^d \to M^{d\times d}$ be C_b^2 and nondegenerate [in the sense given in Hypothesis $(\mathcal{H}_{s.s.})$]. Then G is also C_b^2 and nondegenerate.*

3.3. *Decreasing step Euler algorithm.* In this section we present some control and error expansion results valid for the decreasing step Euler algorithm. The results here presented are found in Lamberton and Pagès (2002) or in the Ph.D. thesis of Lemaire (2005).

A first interesting property is that the sequence of estimators defined in (8) converges almost surely to the ergodic average for any fixed x.

PROPOSITION 3.6. *Assume* $(\mathcal{H}_{f.s.})$, *and let* $\psi : \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \to \mathbb{R}$, *and suppose that* $\psi(x, y) \le C(x)(1 + |y|^\pi)$. *Let* $\tilde{\Psi}^M(x, q)$ *be defined as in* (8). *Then, for any* $x \in \mathbb{R}^{d_x}$, $q \in \mathbb{Q}$,

$$\tilde{\Psi}^M(x, q) \xrightarrow{a.s.} \int \psi(x, y)\mu^x(dy) \qquad \text{as } M \to \infty,$$

where μ^x *is the invariant measure of* (2).

PROOF. $(\mathcal{H}_{f.s.})$ imply that $V(y) := 1 + |y|^2$ is a uniformly in x function satisfying the hypothesis of Theorem 1 in Lamberton and Pagès (2002), from which the claim follows. □

We have as well a control on the moments of any order of $\tilde{Y}_k^{x,q}$.

PROPOSITION 3.7. *Let* $\pi > 0$ *and let* $\tilde{Y}_k^{x,q}$ *be given by* (7). *Then there exists a constant* K_π *given only by* π, λ_-, λ_+ *and* γ_0 *such that for all* $x \in \mathbb{R}^{d_x}$ *and* $q \in \mathbb{Q}$,

$$\sup_{i \in \mathbb{N}} \mathbb{E}[|\tilde{Y}_i^{x,q}|^\pi] < K_\pi.$$

Moreover, for every $\pi > 1$,

$$\sup_{M \in \mathbb{N}} \left(\frac{1}{\Gamma_M} \sum_{i=1}^{M} \gamma_i |\tilde{Y}_i^{x,q}|^\pi \right) < +\infty.$$

PROOF. By Lemma 2 in Lamberton and Pagès (2002) given that U_k^q has moments of any order and $V(y) = |y|^2 + 1$ satisfies the needed hypothesis uniformly in x, we get that for any $\pi \ge 1$ and $q \in \mathbb{Q}$,

$$\sup_{i \in \mathbb{N}} \mathbb{E}[|\tilde{Y}_i^{x,q}|^{2\pi}] \le \sup_{i \in \mathbb{N}} \mathbb{E}[V(\tilde{Y}_i^{x,q})^\pi] < K_\pi.$$

The extension to all $\pi > 0$ is straightforward.

The second claim follows from Theorem 3 in Lamberton and Pagès (2002). □

Proposition 3.8 is an adaptation of a result appearing in the Ph.D. thesis Lemaire (2005). The proof comes from performing a Taylor expansion and reordering the terms in a proper way. For the statement, we introduce in addition to the sequence

$\{\gamma_k\}_{\{k\in\mathbb{N}^*\}}$ a new sequence that we denote by $\{\eta_k\}_{\{k\in\mathbb{N}^*\}}$ (that may be taken equal to the former). This added flexibility will be useful in the following, in particular to prove Proposition 3.10. We may interpret Proposition 3.8 as an error expansion result. Indeed if we fix $\eta_k = \gamma_k$ satisfying $(\mathcal{H}_\gamma)$, then we will have an explicit expression for the approximation error of the decreasing Euler algorithm.

PROPOSITION 3.8. *Let $\psi : \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \to \mathbb{R}$. Under the assumptions of Proposition 3.6, suppose that for each $x \in \mathbb{R}^{d_x}$ there exists $\phi_\psi^x : \mathbb{R}^{d_y} \to \mathbb{R}$ solution of the centered Poisson equation*

$$\mathcal{L}_y^x \phi_\psi^x(y) = \psi(x, y) - \int \psi(x, z)\mu^x(dz). \tag{13}$$

Suppose as well for $r \in \mathbb{N}$, $r \geq 2$, that ϕ_ψ^x is C^r in the y-variable uniformly in x, and $D^r\phi_\psi$ is Lipschitz in y uniformly in x. Let γ_k and η_k be two decreasing sequences with $\gamma_k \to 0$, $\eta_k \to 0$, $\Gamma_k = \sum_{1\leq j\leq k} \gamma_k$, $\mathcal{H}_k = \sum_{1\leq j\leq k} \eta_k$. Let $\tilde{Y}_k^{x,q}$ be defined as in (7) (with step sequence γ_k). Then

$$\sum_{k=1}^{M} \eta_k\left(\psi(x, \tilde{Y}_{k-1}^{x,q}) - \int \psi(x, z)\mu^x(dz)\right) = A_{\psi,M}^0 - N_{\psi,M} - \sum_{i=2}^{r} A_{\psi,M}^i - Z_{\psi,M}^r,$$

where

$$A_{\psi,M}^0(x, q) := \sum_{k=1}^{M} \frac{\eta_k}{\gamma_k}\left[\phi_\psi^x(\tilde{Y}_k^{x,q}) - \phi_\psi^x(\tilde{Y}_{k-1}^{x,q})\right], \tag{14}$$

$$N_{\psi,M}(x, q) := \sum_{k=1}^{M} \frac{\eta_k}{\sqrt{\gamma_k}}\langle D_y\phi_\psi^x(\tilde{Y}_{k-1}^{x,q}), \sigma(x, \tilde{Y}_{k-1}^{x,q})U_k^q\rangle, \tag{15}$$

$$\begin{aligned} A_{\psi,M}^2(x, q) := \frac{1}{2}\sum_{k=1}^{M} \eta_k\big[&D^2\phi_\psi^x(\tilde{Y}_{k-1}^{x,q}) \cdot \big(\sigma(x, \tilde{Y}_{k-1}^{x,q})U_k^q\big)^{\otimes 2} \\ &- \operatorname{Tr}(D^2\phi_\psi^x(\tilde{Y}_{k-1}^{x,q})(\sigma^*\sigma(x, \tilde{Y}_{k-1}^{x,q}))\big], \end{aligned} \tag{16}$$

$$A_{\psi,M}^i(x, q) := \sum_{k=1}^{M} \eta_k\gamma_k^{i/2-1} v_\psi^{i,r}(x, \tilde{Y}_{k-1}^{x,q}, U_k^q) \tag{17}$$

for $i = 3, \ldots, r$ with

$$v_\psi^{i,r}(x, y, z) = \sum_{j\geq i/2}^{i\wedge r} \binom{j}{i-j} \frac{1}{j!} D_y^j\phi_\psi^x(y) \cdot \langle b(x, y)^{\otimes(i-j)}, (\sigma(x, y)z)^{\otimes(2j-i)}\rangle$$

and

$$|Z_{\psi,M}^r|(x, q) \leq K \sum_{k=1}^{M} \eta_k\gamma_k^{(r-1)/2}(1 + |\tilde{Y}_{k-1}^{x,q}|^{r+1})(1 + |U_k^q|)^{r+1}. \tag{18}$$

The average of each expansion term will play an important role in our analysis, so that we will present a special notation for them. Indeed, let

$$\begin{aligned}
&\bar{v}_{\psi}^{i,r}(x,y)\\
(19)\quad &:= \mathbb{E}\big[v_{\psi}^{i,r}\big(x,y,U_1^0\big)\big]\\
&= \sum_{j\geq i/2}^{i\wedge r} \binom{j}{i-j} \frac{1}{j!} D_y^j \phi_{\psi}^x(y) \mathbb{E}\big[\big\langle b(x,y)^{\otimes(i-j)}, \big(\sigma(x,y)U_k^q\big)^{\otimes(2j-i)}\big\rangle \big| \tilde{F}_{\Gamma_{k-1}}\big].
\end{aligned}$$

REMARK 3.9. Consider $A_{\psi,M}^{2i+1}$ for $i \leq \lfloor (r-1)/2 \rfloor$. As $2j-2i-1$ is odd for any j integer and given the fact that the odd powers of a centered Gaussian are centered, we deduce $\bar{v}_{\psi}^{2i+1,r} = 0$. Of course this property transfers to $A_{\psi,M}^{2i+1}$ so that $\mathbb{E}[A_{\psi,M}^{2i+1}] = 0$, implying in turn that the terms with an odd index are centered.

Under some additional hypotheses, Proposition 3.8 may be used to obtain an L_2 control on the error of the approximation. For the sake of the presentation, let us denote from now on

$$\Gamma_M^{[r]} = \sum_{k=1}^{M} (\gamma_k)^r. \tag{20}$$

Note we have in particular $\Gamma_M^{[1]} = \Gamma_M$.

PROPOSITION 3.10. *Under the assumptions of Proposition 3.8, let $\alpha \geq 1$. Assume $\{\gamma_k\}$ satisfies $(\mathcal{H}_\gamma)$, and that $\Gamma_M^{[\alpha]} \to \infty$, for $\Gamma_M^{[\alpha]}$ defined as in (20). Assume as well that the solution of the centered Poisson equation ϕ_ψ is in $C_{b,p}^{2,r}$ for $r > 3$. Let $\bar{\Psi} := \int \psi(x,z)\mu^x(dz)$, then*

$$\mathbb{E}\left[\left|\frac{1}{\Gamma_M^{[\alpha]}} \sum_{k=1}^{M} \gamma_k^{\alpha}\big(\psi\big(x,\tilde{Y}_{k-1}^{x,q}\big) - \bar{\Psi}(x)\big)\right|^2\right] \leq K \frac{1 + \Gamma_M^{[2\alpha-1]} + \Gamma_M^{[2\alpha]} + (\Gamma_M^{[\alpha+1]})^2}{(\Gamma_M^{[\alpha]})^2}.$$

PROOF. We recall first some martingale inequalities. Let $\{a_k\}$ be any sequence of random tensors. By Cauchy–Schwarz inequality we have that

$$\mathbb{E}\left[\left|\sum_{k=1}^{M} \gamma_k^p a_k\right|^2\right] \leq \mathbb{E}\left[\Gamma_M^{[p]} \sum_{k=1}^{M} \gamma_k^p |a_k|^2\right] = \Gamma_M^{[p]} \sum_{k=1}^{M} \gamma_k^p \mathbb{E}\big[|a_k|^2\big]. \tag{21}$$

Let $\{b_k\}$ be also a sequence of tensors. If $s_0 < s_1 < \cdots < s_k < \cdots$, the $\{a_k\}$, $\{b_k\}$ are $\tilde{\mathcal{F}}_{s_k}^q$ adapted, and for all k, $\mathbb{E}[a_k|\tilde{\mathcal{F}}_{s_k}^q] = \mathbb{E}[b_k|\tilde{\mathcal{F}}_{s_k}^q] = 0$, we have by martingale properties that

$$\mathbb{E}\left[\left\langle \sum_{k=1}^{M} \gamma_k^p a_k, \sum_{k=1}^{M} \gamma_k^p b_k \right\rangle\right] = \sum_{k=1}^{M} \gamma_k^{2p} \mathbb{E}\big[\langle a_k, b_k\rangle\big] \tag{22}$$

and in particular,

$$\mathbb{E}\left[\left|\sum_{k=1}^{M}\gamma_k^{p}a_k\right|^2\right]=\sum_{k=1}^{M}\gamma_k^{2p}\mathbb{E}[|a_k|^2]. \tag{23}$$

Now, take the error expansion in Proposition 3.8 with $r=3$, and let $\eta_k=\gamma_k^{\alpha}$. By Abel's transformation, using convexity, estimate (21), the regularity properties of ϕ_ψ and Proposition 3.7, we get

$$\begin{aligned}
&\mathbb{E}[|A^0_{\psi,M}(x,q)|^2]\\
&\quad=\mathbb{E}\left[\left|\sum_{k=1}^{M}\gamma_k^{\alpha-1}[\phi_\psi^x(\tilde{Y}_k^{x,q})-\phi_\psi^x(\tilde{Y}_{k-1}^{x,q})]\right|^2\right]\\
&\quad=\mathbb{E}\Bigg[\Bigg|\gamma_M^{\alpha-1}\phi_\psi^x(\tilde{Y}_M^{x,q})-\gamma_0^{\alpha-1}\phi_\psi^x(\tilde{Y}_0^{x,q})\\
&\qquad\quad+\sum_{k=1}^{M-1}[(\gamma_k^{\alpha-1}-\gamma_{k+1}^{\alpha-1})\phi_\psi^x(\tilde{Y}_k^{x,q})]\Bigg|^2\Bigg]\\
&\quad\le 3\mathbb{E}[|\gamma_M^{\alpha-1}\phi_\psi^x(\tilde{Y}_M^{x,q})|^2]+3\mathbb{E}[|\gamma_0^{\alpha-1}\phi_\psi^x(\tilde{Y}_0^{x,q})|^2]\\
&\qquad+3\mathbb{E}\left[\left|\sum_{k=1}^{M-1}[(\gamma_k^{\alpha-1}-\gamma_{k+1}^{\alpha-1})\phi_\psi^x(\tilde{Y}_k^{x,q})]\right|^2\right]\\
&\quad\le K\left[(\gamma_M^{\alpha-1})^2+1+\left(\sum_{k=1}^{M-1}(\gamma_k^{\alpha-1}-\gamma_{k+1}^{\alpha-1})\right)^2\right]\le K.
\end{aligned} \tag{24}$$

Moreover, using the fact that the terms are centered from Remark 3.9, equation (23) and the finite moments of the Brownian increments imply

$$\mathbb{E}[|N_{\psi,M}(x,q)|^2]=\sum_{k=1}^{M}\gamma_k^{2\alpha-1}\mathbb{E}[|\langle\sigma^*D_y\phi_\psi(x,\tilde{Y}_{k-1}^{x,q}),U_k^q\rangle|^2]\le K\Gamma_M^{2\alpha-1}, \tag{25}$$

$$\begin{aligned}
\mathbb{E}[|A^2_{\psi,M}(x,q)|^2]&\le\frac{1}{4}\sum_{k=1}^{M}\gamma_k^{2\alpha}\mathbb{E}[|D_y^2\phi_\psi(x,\tilde{Y}_{k-1}^{x,q})\cdot(\sigma(x,\tilde{Y}_{k-1}^{x,q})U_k^q)^{\otimes 2}|^2]\\
&\le K\Gamma_M^{[2\alpha]}.
\end{aligned} \tag{26}$$

More generally, estimate (23) leads to

$$\mathbb{E}[|A^3_{\psi,M}(x,q)|^2]=\sum_{k=1}^{M}\gamma_k^{2\alpha+1}\mathbb{E}[|v_\psi^{3,r}(x,\tilde{Y}_{k-1}^{x,q},U_k^q)|^2]\le K\Gamma_M^{[2\alpha+1]}, \tag{27}$$

while by virtue of (21), we find as estimate

$$\begin{aligned}\mathbb{E}\big[\big|Z^3_{\psi,M}(x,q)\big|^2\big] &\le K\mathbb{E}\left[\left|\sum_{k=1}^{M}\gamma_k^{\alpha+1}\big(1+|\tilde{Y}^{x,q}_{k-1}|^4\big)\big(1+|U^q_k|\big)^4\right|^2\right]\\ &\le K\big(\Gamma_M^{[\alpha+1]}\big)^2.\end{aligned} \tag{28}$$

On the other hand, from $(\mathcal{H}_\gamma)$ and given that $\Gamma_M^{[\alpha]}\to\infty$, we have for M large enough that, if $i>j$,

$$\frac{\Gamma_M^{[i]}}{\Gamma_M^{[\alpha]}}\le\frac{\Gamma_M^{[j]}}{\Gamma_M^{[\alpha]}}.$$

The claim follows from Proposition 3.8 and (24)–(28). □

3.4. *Ergodic average and Poisson equation.* Being basic to our analysis, we introduce in this section some known properties of the exact averages and the effective diffusion. These results are studied in Pardoux and Veretennikov (2001, 2003).

Let us start by stating a growth control result proved in Veretennikov (1997).

PROPOSITION 3.11. *Let $\alpha>0$, and let Y_t^x be the solution of* (2) *with deterministic initial condition y_0 and coefficients satisfying $(\mathcal{H}_{f.s.})$.*

Then there exists a constant K given only by α, λ_-, λ_+ such that for all $t\ge 0$ and $x\in\mathbb{R}^{d_x}$,

$$\mathbb{E}\big[\big|Y_t^x\big|^\alpha\big]<K\big(1+|y_0|^{\alpha+2}\big).$$

This proposition has a natural corollary.

COROLLARY 3.12. *Under the same hypothesis of the theorem, for any $\alpha>0$ and all $x\in\mathbb{R}^{d_x}$,*

$$\int|y|^\alpha\mu^x(dy)<K.$$

LEMMA 3.13. *Let $\psi(x,y)$ be a function satisfying the regularity and growth conditions in $(\mathcal{H}_{s.s.})$, and let $\Psi(x)=\int\psi(x,y)\mu^x(dy)$, then $\Psi(x)$ is C_b^2.*

PROOF. The claim follows from adapting Theorems 3 and 5 in Veretennikov (2011) to the linear growth case: the needed equivalent results of convergence in total variation and control of expectations may be found in Meyn and Tweedie (1993). □

As it was shown in Proposition 3.8, the centered Poisson equation (13) plays a special role in understanding the error expansion of the decreasing Euler algorithm. Proposition 3.14, which is an adaptation of Theorem 1 in Pardoux and

Veretennikov (2001) and Veretennikov (2011), states some sufficient conditions for having the solution of such an equation when f belongs to a certain family of functions.

PROPOSITION 3.14. *Consider a function $\psi(x, y)$ satisfying the regularity and growth conditions in* $(\mathcal{H}_{s.s.})$(i), (ii) *and such that*

$$\int \psi(x, y)\mu^x(dy) = 0 \qquad \forall x.$$

Assume $(\mathcal{H}_{f.s.})$. Then there exists a function $\phi_\psi(x, y)$, continuous in y and belonging to the class $\bigcap_{p>1} W^2_{p,\mathrm{loc}}$ in y, such that for every $x \in \mathbb{R}^{d_x}$:

(i) $\mathcal{L}^x_y \phi_\psi(x, y) = \psi(x, y)$,
(ii) $\int \phi_\psi(x, y)\mu^x(dy) = 0$,
(iii) $\phi_\psi \in C^{2,r^y}_{b,p}$.

This function is the unique solution up to an additive constant of the Poisson equation on the class of continuous and $\bigcap_{p>1} W^2_{p,\mathrm{loc}}$ functions in y which are locally bounded and grow at most polynomially in $|y|$ as $|y| \to \infty$. Moreover, it has the representation

$$\phi_\psi(x, y) = -\int_0^\infty \mathbb{E}_{x,y}\big(\psi\big(x, Y^x_t\big)\big)\, dt.$$

4. Convergence results for the MsDS algorithm. We focus now on the study of the MsDS algorithm. First, we show that the proposed approximated coefficients (by means of Decreasing Euler step and Cholesky procedures) satisfy a growth control and error control properties. As a consequence, we will conclude on some regularity property of the approximated diffusion (9) and show its strong convergence toward (3). Then we will study the limit error distribution property.

4.1. *Existence*, *uniqueness*, *continuity*. From Hypotheses $(\mathcal{H}_{s.s.})$, $(\mathcal{H}_{f.s.})$, Proposition 3.11 and Proposition 3.2, it follows that there exists a unique solution to equation (3), and that it has a continuous modification. We show the defined approximation has the same properties.

Proposition 4.1 uses the results of Section 3 to show that, under the standing hypothesis, the coefficients of the approximated diffusion have finite moments of any order, and that its error with respect to the exact coefficients decrease as a power of the number of steps n.

PROPOSITION 4.1. *Assume $(\mathcal{H}_{s.s.})$, $(\mathcal{H}_{f.s.})$ and $(\mathcal{H}_\gamma)$. Let $\beta_0 > 0$, and define $M(n)$ implicitly by $\Gamma_{M(n)} = C_0 n^{2\beta_0}$, where C_0 is some constant.*

(i) *There exist ϕ_f and ϕ_h solutions of the centered Poisson equations*:

- $\mathcal{L}^x_y \phi_f(x, y) = f(x, y) - \int f(x, y')\mu^x(dy')$;

- $\mathcal{L}^x_y \phi_h(x,y) = h(x,y) - \int h(x,y')\mu^x(dy')$.

(ii) *Let*

$$\varsigma := \min_{l\geq 4, i=1,\dots,d}(\bar{v}^{l,r^y}_{F^i} \neq 0) \wedge \min_{l\geq 4, i,j=1,\dots,d}(\bar{v}^{l,r^y}_{H^{i,j}} \neq 0) \wedge (r^y+1) \tag{29}$$

[*with the convention that* $\min(\varnothing) = \infty$] *and* $\bar{v}^{l,r}_{F^i}, \bar{v}^{l,r^y}_{H^{i,j}}$ *defined as in* (19) *applied to* $F^1,\dots,F^{d_x}, H^{1,1},\dots,H^{d_x,d_x}$. *Assume the asymptotic expantion*

$$\frac{\Gamma_M^{[\varsigma/2]}}{\Gamma_M} = C_1 n^{-\beta_1} + o(n^{-\beta_1}), \tag{30}$$

for some $\beta_1 > 0$, *and some constant* C_1, *holds. Let*

$$\beta := \beta_0 \wedge \beta_1. \tag{31}$$

Then $\tilde{F}^n$ (*and resp.*, $\tilde{H}^n, \tilde{G}^n := \sqrt{\tilde{H}^n}$) *satisfies for any* $\alpha \in \mathbb{R}^+$ *and* $k = 0,\dots,n$

$$\begin{cases} \mathbb{E}[|\tilde{F}^n(x,t_k)|^\alpha] \leq K, \\ \mathbb{E}[|\tilde{F}^n(x,t_k) - F(x)|^2] \leq K n^{-2\beta}. \end{cases}$$

REMARK 4.2. We should understand ς as marking the first nonzero value in the error expansion of either $\tilde{F}^n$ or $\tilde{H}^n$. It depends exclusively on the coefficients of the effective and ergodic diffusion (in particular it does not depend on n).

REMARK 4.3. Proposition 4.1 means that we have a rate of convergence in norm L_2 for the coefficient estimators of order $O(n^{-\beta})$. Since we *choose* β_0 by taking $M(n)$ as needed, the actual limit to β comes from β_1. But of course, increasing β_0 implies growing M faster as a function of n, increasing the algorithm's cost.

PROOF OF PROPOSITION 4.1. Note first that (i) follows from $(\mathcal{H}_{s.s.})$ and Proposition 3.14.

We prove (ii). By Jensen's inequality and Proposition 3.7, we have for every $\alpha \geq 1$ and n big enough,

$$\begin{aligned} \mathbb{E}[|\tilde{F}^n(x,q)|^\alpha] &= \mathbb{E}\left[\left|\frac{1}{\Gamma_M}\sum_{k=1}^M \gamma_k f(x, \tilde{Y}^{x,q}_{k-1})\right|^\alpha\right] \\ &\leq \mathbb{E}\left[\frac{1}{\Gamma_M}\sum_{k=1}^M \gamma_k |f(x, \tilde{Y}^{x,q}_{k-1})|^\alpha\right] \leq K, \end{aligned}$$

and similarly for every $\alpha \geq 2$,

$$\mathbb{E}[|\tilde{G}^n(x,q)|^\alpha] = \mathbb{E}[|\tilde{H}^n(x,q)|^{\alpha/2}] \leq K,$$

since $|G|^2 = |H|$. The result extends trivially to every $\alpha > 0$.

It remains to prove the error control. We obtain an expansion of order r^y in Proposition 3.8. We can bound the first terms as we did in Proposition 3.10 by taking $\gamma_k = \eta_k$ for all $k = 1, \dots, M$ (i.e., taking $\alpha = 1$ in the statement of Proposition 3.10). More generally, from the definition of ς in (29), we have that for every $l < \varsigma$ or l odd $\bar{v}^{l,r^y}_{F^i}(x, y) = 0$, (23) leads to

$$\mathbb{E}\big[\big|A^l_{F^i,M}(x,q)\big|^2\big] = \sum_{k=1}^{M} \gamma_k^l \mathbb{E}\big[\big|v^{l,r^y}_{F^i}\big(x, \tilde{Y}^{x,q}_{k-1}, U^q_k\big)\big|^2\big] \le K\Gamma^{[l]}_M, \tag{32}$$

while for even l with $l \ge \varsigma$, by virtue of (21), we find as estimate

$$\mathbb{E}\big[\big|A^l_{F^i,M}(x,q)\big|^2\big] \le \Gamma^{l/2}_M \sum_{k=1}^{M} \gamma_k^{l/2} \mathbb{E}\big[\big|v^{l,r^y}_{F^i}\big(x, \tilde{Y}^{x,q}_{k-1}, U^q_k\big)\big|^2\big] \le K\big(\Gamma^{[l/2]}_M\big)^2. \tag{33}$$

Likewise,

$$\begin{aligned}
&\mathbb{E}\big[\big|Z^{r^y}_{F^i,M}(x,q)\big|^2\big] \\
&\quad \le K\mathbb{E}\left[\left|\sum_{k=1}^{M} \gamma_k^{r^y+1/2}\big(1 + \big|\tilde{Y}^{x,q}_{k-1}\big|^{r^y+1}\big)\big(1 + \big|U^q_k\big|\big)^{r^y+1}\right|^2\right] \qquad (34)\\
&\quad \le K\big(\Gamma^{[r^y+1/2]}_M\big)^2.
\end{aligned}$$

Note that estimates (32) and (33) are uniform in x. On the other hand, from $(\mathcal{H}_\gamma)$, we have for M big enough and $l \le r^y$ that

$$1 \ge \frac{\Gamma^{[2]}_M}{\Gamma_M} \ge \frac{\Gamma^{[3]}_M}{\Gamma_M} \ge \cdots \ge \frac{\Gamma^{[l]}_M}{\Gamma_M}.$$

Hence from Proposition 3.8 and equations (24)–(26), (32), (33),

$$\mathbb{E}\big[\big|\tilde{F}^{i;n}(x,q) - F^i(x,q)\big|^2\big] \le \frac{K(\Gamma^{[\varsigma/2]}_M)^2}{(\Gamma_M)^2} + \frac{K}{\Gamma_M} \le K' n^{-2(\beta_0 \wedge \beta_1)},$$

implying our claim for F, $\tilde{F}^n$. Since H satisfies the same properties as F, the claim follows for H, $\tilde{H}^n$. As a final step, we prove the error control for $\tilde{G}^n$. Let $\Delta H^n(x,q) := H(x) - \tilde{H}^n(x,q)$ and $E = \{|\Delta H^n(x,q)| \ge |2H^{-1}|^{-1}\}$. Markov inequality gives us the control

$$\mathbb{P}(E) \le 4\big|H^{-1}(x)\big|^2 \mathbb{E}\big[\big|\Delta H^n(x,q)\big|^2\big] \le Kn^{-2(\beta_0 \wedge \beta_1)},$$

which, in conjunction with Theorem 3.3, deduces

$$\begin{aligned}
&\mathbb{E}\big[\big|G(x) - \tilde{G}^n(x,q)\big|^2\big] \\
&\quad = \mathbb{E}\big[\big|G(x) - \tilde{G}^n(x,q)\big|^2 \mathbf{1}_E\big] + \mathbb{E}\big[\big|G(x) - \tilde{G}^n(x,q)\big|^2 \mathbf{1}_{E^\complement}\big]
\end{aligned}$$

$$\le K'n^{-2(\beta_0\wedge\beta_1)} + \mathbb{E}[|G(x) - \tilde{G}^n(x,q)|^2 \mathbf{1}_{E^\complement}]$$
$$\le K'n^{-2(\beta_0\wedge\beta_1)} + Kn^{-2(\beta_0\wedge\beta_1)} = K''n^{-2(\beta_0\wedge\beta_1)}. \qquad \square$$

We can deduce from Proposition 4.1 and the assumed structure, the following a priori estimates.

COROLLARY 4.4. *Under the hypothesis and notation of Proposition 4.1, for any* $0 \le s \le T$,

$$\mathbb{E}[|\tilde{F}^n(\tilde{X}^n_{\underline{s}}, \underline{s})|^\alpha] \le K \tag{35}$$

and

$$\mathbb{E}[|\tilde{F}^n(\tilde{X}^n_s, \underline{s}) - F(\tilde{X}^n_s)|^2] \le Kn^{-2\beta}. \tag{36}$$

The same bounds hold with $\tilde{F}^n$, F *replaced by* $\tilde{H}^n$, H *and* $\tilde{G}^n$, G.

PROOF. Define

$$\bar{\mathcal{F}}_{t,t^-} := \Bigl(\mathcal{F}_t \vee \bigvee_{q\in\mathbb{Q}, q<t} \tilde{\mathcal{F}}^q_\infty\Bigr) \tag{37}$$

by construction, $\tilde{X}_{\underline{s}}$ is $\bar{\mathcal{F}}_{\underline{s},\underline{s}^-}$ measurable and since $\tilde{F}^n(x,\underline{s}) \perp\!\!\!\perp \bar{\mathcal{F}}_{\underline{s},\underline{s}^-}$ for any deterministic x, we get from Proposition 4.1,

$$\mathbb{E}[|\tilde{F}^n(\tilde{X}_{\underline{s}}, \underline{s})|^\alpha] = \mathbb{E}[\mathbb{E}[|\tilde{F}^n(\tilde{X}_{\underline{s}}, \underline{s})|^\alpha | \bar{\mathcal{F}}_{\underline{s},\underline{s}^-}]] \le \mathbb{E}[K] = K.$$

A similar argument leads to (36), and to the claims for $\tilde{H}^n$, H and $\tilde{G}^n$, G. $\square$

Corollary 4.4 should be understood as an a priori control on the approximated process. From this control, we can deduce, using Proposition 3.2 as in the case of the effective equation, the existence and strong uniqueness of the solution of the approximated diffusion (9). In addition, Proposition 4.5 states that approximation (9) has a continuous modification. The result follows from Proposition 3.11, the estimates in Corollary 4.4 and Kolmogorov's criterion.

PROPOSITION 4.5. *Under the hypothesis and notation of Proposition 4.1, for every* $\alpha \ge 2$,

$$\mathbb{E}[|\tilde{X}^n_t - \tilde{X}^n_s|^\alpha] \le K_{\alpha,T}(t-s)^{\alpha/2}((t-s)^{\alpha/2} + 1).$$

Moreover, the solution of (9) *has a continuous modification.*

4.2. *Strong convergence.* In what follows, we choose $\tilde{X}$ to be continuous in time. We can proceed to show the mean square convergence of $\tilde{X}^n$ toward X.

THEOREM 4.6. *Under* $(\mathcal{H}_{s.s.})$, $(\mathcal{H}_{f.s.})$ *and* $(\mathcal{H}_\gamma)$, *let* X *be defined by* (3) *and* $\tilde{X}^n$ *by* (9). *Let* β *be defined as in* (31). *Then*:

- *if* $g \equiv 0$ (*ODE with random coefficients*), *then* $\mathbb{E}[\sup_{0\le t\le T}|X_t - \tilde{X}_t^n|^2] \le Kn^{-2(1\wedge\beta)}$;
- *under the full SDE case*, $\mathbb{E}[\sup_{0\le t\le T}|X_t - \tilde{X}_t^n|^2] \le Kn^{-(1\wedge 2\beta)}$.

PROOF. We treat the full SDE case. By definition,

$$X_t - \tilde{X}_t^n = \int_0^t [F(X_s) - \tilde{F}^n(\tilde{X}_{\underline{s}}^n, \underline{s})]\,ds + \int_0^t [G(X_s) - \tilde{G}^n(\tilde{X}_{\underline{s}}^n, \underline{s})]\,dW_s.$$

Our plan is to use Proposition 3.2(ii). By convexity,

$$\begin{aligned}
&|F(X_s) - \tilde{F}^n(\tilde{X}_{\underline{s}}^n, \underline{s})|^2 \\
&\quad \le 3|F(X_s) - F(\tilde{X}_s^n)|^2 + 3|F(\tilde{X}_s^n) - F(\tilde{X}_{\underline{s}}^n)|^2 + 3|F(\tilde{X}_{\underline{s}}^n, \underline{s}) - \tilde{F}^n(\tilde{X}_{\underline{s}}^n, \underline{s})|^2.
\end{aligned}$$

By Lipschitz assumption in $(\mathcal{H}_{s.s.})$,

$$\begin{aligned}
&\mathbb{E}[|F(X_s) - F(\tilde{X}_s^n)|^2] \le K\mathbb{E}[|X_s - \tilde{X}_s^n|^2], \\
&\mathbb{E}[|F(\tilde{X}_s^n) - F(\tilde{X}_{\underline{s}}^n)|^2] \le K\mathbb{E}[|\tilde{X}_s^n - \tilde{X}_{\underline{s}}^n|^2] \le Kn^{-1},
\end{aligned} \tag{38}$$

the last inequality being possible for n large enough thanks to Proposition 4.5. Also, by Corollary 4.4, we get

$$\mathbb{E}[|F(\tilde{X}_{\underline{s}}^n, \underline{s}) - \tilde{F}^n(\tilde{X}_{\underline{s}}^n, \underline{s})|^2] \le Kn^{-2\beta}.$$

Therefore,

$$\mathbb{E}[|F(X_s) - \tilde{F}^n(\tilde{X}_{\underline{s}}^n, \underline{s})|^2] \le K(n^{-(1\wedge 2\beta)} + \mathbb{E}[|X_s - \tilde{X}_s^n|^2]). \tag{39}$$

Since we may obtain similar bounds for the terms with G, we also have

$$\mathbb{E}[|G(X_s) - \tilde{G}^n(\tilde{X}_{\underline{s}}^n, \underline{s})|^2] \le K(n^{-(1\wedge 2\beta)} + \mathbb{E}[|X_s - \tilde{X}_s^n|^2]). \tag{40}$$

Now, Proposition 3.2(ii) shows

$$\mathbb{E}[|X_t - \tilde{X}_t^n|^2] \le K\int_0^T (n^{-(1\wedge 2\beta)} + \mathbb{E}[|X_s - \tilde{X}_s^n|^2])\,ds.$$

Therefore, by Gronwall's lemma,

$$\sup_{0\le t\le T}\mathbb{E}[|X_t - \tilde{X}_t^n|^2] \le Kn^{-(1\wedge 2\beta)}.$$

Replacing (39) and (40) we get

$$\sup_{0\le t\le T}\big(\mathbb{E}\big[\big|F(X_s)-\tilde{F}^n(\tilde{X}^n_{\underline{s}},\underline{s})\big|^2\big]+\mathbb{E}\big[\big|G(X_s)-\tilde{G}^n(\tilde{X}^n_{\underline{s}},\underline{s})\big|^2\big]\big)\le Kn^{-(1\wedge 2\beta)}.$$

So that by Proposition 3.2,

$$\mathbb{E}\Big[\sup_{0\le t\le T}\big|X_t-\tilde{X}^n_t\big|^2\Big]\le Kn^{-(1\wedge 2\beta)}.$$

Note that the case $g\equiv 0$ is proven in the same way, but the Euler error (39) is bounded by n^{-2} and $G\equiv 0$. This implies the stated result. □

4.3. *Limit distribution.* In this section we show under slightly stronger regularity assumptions on the coefficients of the diffusion, that we have convergence in the weak (uniform topology) sense toward a limit distribution given as the solution of a particular SDE.

Our plan to prove the limit distribution result is to look at the rescaled error and its associated stochastic differential equation. We prove the joint weak convergence of the terms appearing in that SDE and use the fact that under certain hypothesis the joint convergence of the terms suffices to deduce the weak convergence of the solution of the equation. The reader may find most of the needed material on weak convergence of stochastic integrals and stochastic SDEs in Jakubowski, Mémin and Pagès (1989), Kurtz and Protter (1991a, 1996).

DEFINITION 4.7. Let X^n be a sequence of $\mathbb{R}^d$-valued semimartingales, and let $A^n(\delta)$ be the predictable process with finite variation null at zero and $M^n(\delta)$ the local martingale null at zero appearing in the representation of X^n as

$$X^n_t=X^n_0+A^n_t(\delta)+M^n_t(\delta)+\sum_{s\le t}\Delta X^n_s\mathbf{1}_{\{|\Delta X^n_s|>\delta\}}.$$

We say that the sequence X^n satisfies property $(*)$ if for some $\delta>0$,

$$(*)\qquad \langle M^n(\delta),M^n(\delta)\rangle_T+\int_0^T\big|dA^n(\delta)_s\big|+\sum_{s\le T}\big|\Delta X^n_s\big|\mathbf{1}_{\{|\Delta X^n_s|>\delta\}}$$

is tight. (The notation $\int_0^T|dA|$ denotes the total variation of A on $[0,T]$.)

The importance of property $(*)$ is shown by the following theorem; see Jakubowski, Mémin and Pagès (1989), Jacod and Protter (1998) and Kurtz and Protter (1996).

THEOREM 4.8. *Let X^n be a sequence of $\mathbb{R}^d$-valued semimartingales relative to the filtration $\mathcal{F}_t$. Suppose that X^n weakly converges in the Skorokhod topology $D_{\mathbb{R}^{d_X}}$. Then $(*)$ is necessary and sufficient for goodness: for any sequence H^n of $(\mathcal{F}_t)$-adapted càdlàg processes such that $(H^n,X^n)\Rightarrow(H,X)$ in the Skorokhod* *topology $D_{M^{d_x \times d_x} \times \mathbb{R}^{d_x}}$, then X is a semimartingale w.r.t. the filtration generated by (H, X) and $(H^n, X^n, \int H^n \, dX^n) \Rightarrow (H, X, \int H \, dX)$ in the Skorokhod topology $D_{M^{d_x \times d_x} \times \mathbb{R}^{d_x} \times \mathbb{R}^{d_x}}$.*

Goodness gives us a direct way to show the convergence of sequences of stochastic integrals, and will play a key role for the convergence of sequences of SDEs.

Before proceeding to the main propositions of this section, we cite another useful result concerning weak convergence of sequences of solutions of SDEs, allowing us to compare the limit of two sequences with converging coefficients.

THEOREM 4.9 [Theorem 2.5(b) Jacod and Protter (1998)]. *Consider a sequence of linear SDEs*

$$\vartheta_t^n = P_t^n + \int_0^t \vartheta_{s-}^n Q_t^n \, dJ_t, \tag{41}$$

where the P_t^n are stochastic processes in $\mathbb{R}^d$, Q_t^n are stochastic processes in $\mathbb{R}^{d \times d'}$ and J_t is a semimartingale in $\mathbb{R}^{d'}$, and all processes are in same the filtered probability space. Suppose that we have another sequence of equations like (41) with solution ϑ'^n and coefficients P'^n and Q'^n. If the sequences $\sup_{0 \le s \le T} \|P_s^n\|$ and $\sup_{0 \le s \le T} \|Q_s^n\|$ are tight, and if

$$\sup_{0 \le s \le T} \|P_s^n - P_s'^n\| \xrightarrow{P} 0, \qquad \sup_{0 \le s \le T} \|Q_s^n - Q_s'^n\| \xrightarrow{P} 0,$$

then

$$\sup_{0 \le s \le T} \|\vartheta_s^n - \vartheta_s'^n\| \xrightarrow{P} 0.$$

Proposition 4.10 shows the weak convergence of some tuples appearing in the rescaled error SDE.

PROPOSITION 4.10. *Let $\mathcal{I}$ be a set of indices, and consider a family of independent standard Gaussian variables $\{\nu_{t_k}^{i;n}\}_{n \in \mathbb{N}^*; 0 \le k \le n; i \in \mathcal{I}}$ where for any n, i we have $\nu_{t_k}^{i;n}$ is $\bar{\mathcal{F}}_{t_k}$ measurable.*

Consider the sequence of random processes $A^{0;n}$ (dimension 1), $A^{1;n}$, $B^{0;n}$ (dimension d_x), $B^{2;n}$ (dimension $d_x \times d_x$), $B^{1;n}$ (dimension $|\mathcal{I}|$) and $B^{3;n}$ (dimension $|\mathcal{I}| \times d_x$) defined component-wise by

$$B_t^{0;j;n} := \int_0^t (s - \underline{s}) \, dW_s^j; \qquad A_t^{0;n} := 2 \int_0^t (s - \underline{s}) \, ds; \tag{42}$$

$$B_t^{2;l,j;n} := \int_0^t \sqrt{2}(W_s^l - W_{\underline{s}}^l) \, dW_s^j; \qquad A_t^{1;j;n} := \int_0^t (W_s^j - W_{\underline{s}}^j) \, ds; \tag{43}$$

$$B_t^{3;i,j;n} := \int_0^t \nu_{\underline{s}}^{i;n} \, dW_s^j; \qquad B_t^{1;i;n} := \int_0^t \nu_{\underline{s}}^{i;n} \, ds. \tag{44}$$

Then we have the following limit results:

$$(X, \tilde{X}^n, W, nA^{0;n}, \sqrt{n}B^{1;n}) \quad \Rightarrow \quad (X, X, W, A^0, B^1) \tag{45}$$

$$\begin{aligned}&(X, \tilde{X}^n, W, n^{1/2}A^{0;n}, n^{1/2}B^{0;n}, n^{1/2}A^{1;n}, n^{1/2}B_s^{2;n}, B^{1;n}, B^{3;n})\\ &\qquad\Rightarrow \quad (X, X, W, 0, 0, 0, B^2, 0, B^3),\end{aligned} \tag{46}$$

where $A_t^0 = t$; B^0, B^1, B^2 *and* B^3 *are standard Brownian motions defined on an extension of the space* W, *with dimensions* d_x, d_x^2, $|\mathcal{I}| \times d_x$ *and* $|\mathcal{I}|$, *respectively.*

Moreover, we have $\{B^0, B^2, B^3, W\}$ *are independent*; $\{B^0, B^2, B^1, W\}$ *are independent*, *and* $B^{1;n}$, $\sqrt{n}B^{2;n}$ *and* $B^{3;n}$ *are "good" in the sense of Theorem* 4.8.

The proof of Proposition 4.10 will be given in Section A.1.

PROPOSITION 4.11. *Under the assumptions and notation of Proposition* 4.1, *assume that* $r^y > \varsigma + 3$ *in* $(\mathcal{H}_{s.s.})$, *and that there is* $\beta_2 \geq 0$ *such that the asymptotic expansion*

$$\frac{\Gamma_M^{[\varsigma/2+1]}}{\Gamma_M^{[\varsigma/2]}} = C_2 n^{-\beta_2} + o(n^{-\beta_2}), \tag{47}$$

where ς *is defined in* (29), *holds*. *Let*

$$\rho = \mathbf{1}_{\{\beta_0 > \beta_1\}}(\beta_2 \wedge (\beta_0 - \beta_1)) + \mathbf{1}_{\{\beta_0 < \beta_1\}}(\beta_0 \wedge (\beta_1 - \beta_0)). \tag{48}$$

(i) *Let* Φ_F *be the* $d_x \times d_x$ *matrix defined component-wise as*

$$\Phi_F^{i,j}(x) := C_0^{-1} \int \langle \sigma^* D_y \phi_{F^i}(x, y), \sigma^* D_y \phi_{F^j}(x, y)\rangle \mu^x(dy),$$

where ϕ_{F^i} *is the solutions of the Poisson equation* (13) *with source* F^i. *Let*

$$\varphi_F(x) := \mathbf{1}_{\{\beta_1 \geq \beta_0\}}\sqrt{\Phi_F(x)}; \qquad R_F^i(x) := \mathbf{1}_{\{\beta_0 \geq \beta_1\}} C_1 \int \bar{v}_{F^i}^{\varsigma, r^y}(x, y)\mu^x(dy),$$

with the square root meaning the Cholesky root. Then there exists a family of independent standard Gaussian variables $\{\nu_k^{i;n}\}_{n \in \mathbb{N}^*; 0 \leq k \leq n; 1 \leq i \leq d_x}$, *such that each* $\nu_k^{i;n}$ *is* $\bar{\mathcal{F}}_{t_k}$ *measurable and*

$$\mathbb{E}\left[\left| n^\beta (F^i(x) - \tilde{F}^{i;n}(x, t_k)) - \sum_{j=1}^{d_x} \varphi_F^{i,j}(x)\nu_k^{j;n} - R_F^i(x)\right|^2\right] = O(n^{-2\rho}),$$

for all $x \in \mathbb{R}^{d_x}$.

(ii) *Under the full SDE case, define in a similar way a d_x^2 dimensional random function R_H and a $d_x^2 \times d_x^2$ dimensional random function Φ_H, with*

$$\Phi_H^{i,j,i',j'}(x) := C_0^{-1} \int \langle \sigma^* D_y \phi_{H^{i,j}}(x,y), \sigma^* D_y \phi_{H^{i',j'}}(x,y) \rangle \mu^x(dy);$$

$$\varphi_H(x) := \mathbf{1}_{\{\beta_1 \geq \beta_0\}} \sqrt{\Phi_H(x)};$$

$$R_H^{i,j}(x) := \mathbf{1}_{\{\beta_0 \geq \beta_1\}} C_1 \int \bar{v}^{\varsigma, r^y}_{H^{i,j}}(x,y) \mu^x(dy).$$

Then there exists a family of independent standard Gaussian variables $\{\nu_k^{i,j;n}\}_{n \in \mathbb{N}^*; 0 \leq k \leq n; 0 \leq i,j \leq d_x}$, such that each $\nu_k^{i,j;n}$ is $\bar{\mathcal{F}}_{t_k}$ measurable and

$$\mathbb{E}\left[\left| n^\beta \bigl(H^{i,j}(x) - \tilde{H}^{i,j;n}(x,t_k)\bigr) - \sum_{i',j'=1}^{d_x} \varphi_H^{i,j,i',j'}(x) \nu_k^{i',j';n} - R_H^{i,j}(x) \right|^2\right]$$
$$= O\bigl(n^{-2\rho}\bigr),$$

for all $x \in \mathbb{R}^{d_x}$. Moreover, letting R_G, φ_G be defined component-wise for $0 \leq i', j' \leq d_x$ as

$$R_G^{i,i} = \frac{R_H^{i,i} - 2\sum_{k=1}^{i-1} R_G^{i,k} G^{i,k}}{2G^{i,i}},$$

$$\varphi_G^{i,i,i',j'} = \frac{\varphi_H^{i,i,i',j'} - 2\sum_{j=1}^{i-1} \varphi_G^{i,j,i',j'} G^{i,j}}{2G^{i,i}},$$

and for $i > j$,

$$R_G^{i,j} = \frac{R_H^{i,j} - R_G^{j,j} G^{i,j} - \sum_{l=1}^{j-1} (R_G^{j,l} G^{i,l} + R_G^{i,l} G^{j,l})}{G^{j,j}},$$

$$\varphi_G^{i,j,i',j'} = \frac{\varphi_H^{i,j,i',j'} - \varphi_G^{j,j,i',j'} G^{i,j} - \sum_{l=1}^{j-1} [\varphi_G^{j,l,i',j'} G^{i,l} + \varphi_G^{i,l,i',j'} G^{j,l}]}{G^{j,j}}.$$

Then

$$\mathbb{E}\left[\left| n^\beta \bigl(G^{i,j}(x) - \tilde{G}^{i,j;n}(x,t_k)\bigr) - \sum_{i',j'=1}^{d_x} \varphi_G^{i,j,i',j'}(x) \nu_k^{i',j';n} - R_G^{i,j}(x) \right|^2\right]$$
$$= O\bigl(n^{-2\rho}\bigr).$$

PROOF. (i) We prove the first claim. We use the expansion of Proposition 3.8 up to order ς as in Proposition 4.1, and estimates (24)–(26), (32)–(34) to get for any x that

$$\begin{aligned} &\mathbb{E}\left[\left| \bigl(F^i(x) - \tilde{F}^{i;n}(x,q)\bigr) - \frac{1}{\Gamma_M}\bigl(N_{F^i,M}(x,q) + A^{(\varsigma)}_{F^i,M}(x,q)\bigr) \right|^2\right] \\ &\quad = O\bigl((\Gamma_M)^{-2}[1 + (\Gamma_M^{[\varsigma/2+1]})^2]\bigr). \end{aligned} \tag{49}$$

Let us examine separately three cases depending on the relation between β_0 and β_1:

- If $\beta_0 > \beta_1$: In this case $\beta = \beta_1$, and by definition of β_1 it follows that

$$\begin{aligned}
&\mathbb{E}\big[\big|n^{\beta}\big(F^i(x) - \tilde{F}^{i;n}(x,q)\big) - R_F^i(x)\big|^2\big] \\
&\quad \le K\mathbb{E}\big[\big|\big(\Gamma^{[\varsigma/2]}\big)^{-1}\Gamma_M\big(F^i(x) - \tilde{F}^{i;n}(x,q)\big) - R_F^i(x)\big|^2\big] \\
&\quad \le K'\mathbb{E}\big[\big|\Gamma_M^{[\varsigma/2]}\big|^{-2}\big|\Gamma_M\big(F^i(x) - \tilde{F}^{i;n}(x,q)\big) \\
&\qquad\qquad - N_{F^i,M}(x,q) - A^{\varsigma}_{F^i,M}(x,q)\big|^2\big] \\
&\qquad + K'\mathbb{E}\big[\big|\big(\Gamma_M^{|[\varsigma/2]}\big)^{-1}\big(A^{\varsigma}_{F^i,M}(x,q) - R_F^i(x)\big)\big|^2\big] \\
&\qquad + K'\mathbb{E}\big[\big|\big(\Gamma_M^{|[\varsigma/2]}\big)^{-1}N_{F^i,M}(x,q)\big|^2\big].
\end{aligned} \tag{50}$$

The first term in the right-hand side of (50) can be controlled by rescaling (49) to get

$$\begin{aligned}
&\mathbb{E}\big[\big|\Gamma_M^{[\varsigma/2]}\big|^{-2}\big|\Gamma_M\big(F^i(x) - \tilde{F}^{i;n}(x,q)\big) - N_{F^i,M}(x,q) - A^{\varsigma}_{F^i,M}(x,q)\big|^2\big] \\
&\quad = O\big(\big(\Gamma_M^{[\varsigma/2]}\big)^{-2}\big[\big(\Gamma_M^{[\varsigma/2+1]}\big)^2 + 1\big]\big).
\end{aligned} \tag{51}$$

From (25) we control the third term in the right-hand side of (50)

$$\mathbb{E}\big[\big|\big(\Gamma_M^{[\varsigma/2]}\big)^{-1}N_{F^i,M}(x,q)\big|^2\big] = O\big(\big(\Gamma_M^{[\varsigma/2]}\big)^{-2}\Gamma_M\big). \tag{52}$$

To control the second term of (50), let us define

$$\bar{A}^{\varsigma}_{F^i,M}(x,q) := \sum_{k=1}^{M} \gamma_k^{\varsigma/2}\bar{v}^{\varsigma,r^y}_{F^i}\big(x,\bar{Y}^{x,q}_{k-1}\big) \tag{53}$$

for $\bar{v}^{\varsigma,r^y}_{F^i}$ defined in (19). We can compare $(\Gamma_M^{[\varsigma/2]})^{-1}A^{\varsigma}_{F^i,M}$ and $(\Gamma_M^{[\varsigma/2]})^{-1} \times \bar{A}^{\varsigma}_{F^i,M}$ in L_2 by (23). Indeed, thanks to controls (32) and (33), and the fact that for some $K \in \mathbb{R}^+$, $\Gamma_M^{[\varsigma]} \le K\Gamma_M$, we have

$$\begin{aligned}
&\mathbb{E}\big[\big|\big(\Gamma_M^{[\varsigma/2]}\big)^{-1}\big(A^{\varsigma}_{F^i,M}(x,q) - \bar{A}^{\varsigma}_{F^i,M}(x,q)\big)\big|^2\big] \\
&\quad = \mathbb{E}\left[\left|\big(\Gamma_M^{[\varsigma/2]}\big)^{-1}\sum_{k=1}^{M}\gamma_k^{\varsigma/2}\big(v^{\varsigma,r^y}_{F^i}\big(x,\bar{Y}^{x,q}_{k-1},U_k^q\big) - \bar{v}^{\varsigma,r^y}_{F^i}\big(x,\bar{Y}^{x,q}_{k-1}\big)\big)\right|^2\right] \\
&\quad = O\big(\big(\Gamma^{[\varsigma/2]}\big)^{-2}\Gamma_M\big).
\end{aligned} \tag{54}$$

It remains to show that

$$\mathbb{E}\big[\big|\big(\Gamma_M^{[\varsigma/2]}\big)^{-1}\bar{A}^{\varsigma}_{F^i,M}(x,q) + R_F^i(x)\big|^2\big] = O\big(n^{-2\rho}\big). \tag{55}$$

Indeed, from the definition of β_0 and β_1, $\Gamma_M^{[\varsigma/2]} = O(n^{\beta_0-\beta_1})$ so that it diverges. Moreover, from the assumed regularity hypothesis, $\bar{v}_{F^i}^{\varsigma,r^y}(x,y)$ is $C_{p,b}^{2,r^y-\varsigma}$. Therefore, Proposition 3.14 guarantees the existence of a solution to the centered Poisson equation with source $\bar{A}_{F^i}^{\varsigma,r^y}(x,y)$ of the same regularity, and thus Proposition 3.10 shows that $\bar{A}_{F^i,M}^{\varsigma}(x,q)$ converges uniformly with respect to x in L^2 to $-R_F^i(x)$ with rate $(\beta_0-\beta_1)\wedge\beta_2 \geq \rho$ since

$$\begin{aligned}
&(\Gamma_M^{\varsigma/2})^{-2}(1+\Gamma_M^{[\varsigma]}+\Gamma_M^{[\varsigma-1]}+(\Gamma_M^{[\varsigma/2+1]})^2) \\
&\quad \leq K(\Gamma_M^{\varsigma/2})^{-2}(\Gamma_M+(\Gamma_M^{[\varsigma/2+1]})^2) = O(n^{-2((\beta_0-\beta_1)\wedge\beta_2)}).
\end{aligned}$$

The claim follows from replacing (51), (52), (54) and (55) in (50).

- If $\beta_1 > \beta_0$, we follow a similar approach. We expand the rescaled error term to find

$$\begin{aligned}
&\mathbb{E}\big[\big|n^\beta\big(F^i(x)-\tilde{F}^{i;n}(x,q)\big)-R_F^i(x)\big|^2\big] \\
&\quad \leq K'\mathbb{E}\big[|\Gamma_M|^{-1/2}\big|\Gamma_M\big(F^i(x)-\tilde{F}^{i;n}(x,q)\big) \\
&\qquad\qquad - N_{F^i,M}(x,q)-A_{F^i,M}^{\varsigma}(x,q)\big|^2\big] \\
&\qquad + K'\mathbb{E}\big[\big|(\Gamma_M)^{-1/2}A_{F^i,M}^{\varsigma}(x,q)\big|^2\big] \\
&\qquad + K'\mathbb{E}\big[\big|(\Gamma_M)^{-1/2}\big(N_{F^i,M}(x,q)-\Phi_F\big)\big|^2\big].
\end{aligned} \tag{56}$$

By rescaling (49) we get

$$\begin{aligned}
&\mathbb{E}\big[|\Gamma_M|^{-1/2}\big|\Gamma_M\big(F^i(x)-\tilde{F}^{i;n}(x,q)\big)-N_{F^i,M}(x,q)-A_{F^i,M}^{\varsigma}(x,q)\big|^2\big] \\
&\quad = O\big((\Gamma_M)^{-1}\big[1+\big(\Gamma_M^{[\varsigma/2+1]}\big)^2\big]\big),
\end{aligned}$$

and from (33),

$$\mathbb{E}\big[\big|(\Gamma_M)^{-1/2}A_{F^i,M}^{\varsigma}(x,q)\big|^2\big] = O\big((\Gamma_M)^{-1}\big(\Gamma_M^{[\varsigma/2]}\big)^2\big).$$

So it remains to consider the N_M term. Note that since the U_k^q are independent standard Gaussian vectors, $(C_0\sqrt{\Gamma_M})^{-1}N_{F^i,M}(x,q)$ when i ranges $1,\dots,d_x$ is a Gaussian vector.

Let us study its covariance matrix Φ_F^n. Using (22) we get for $i,j=1,\dots,n$

$$\begin{aligned}
\Phi_F^{i,j;n}(x,q) &:= \mathbb{E}\bigg[\frac{1}{\Gamma_M}N_{F^i,M}(x,q)N_{F^j,M}(x,q')\bigg] \\
&= \mathbf{1}_{\{q=q'\}}\sum_{k,k'=1}^{M}\gamma_k\big\langle\sigma^*(\cdot)D_y\phi_{F^i}(\cdot),\sigma^*(\cdot)D_y\phi_{F^j}(\cdot)\big\rangle\big(x,\tilde{Y}_{k-1}^{x,q}\big).
\end{aligned}$$

Define $\varphi_F^n = \sqrt{\Phi_F^n}$ (the Cholesky decomposition). Then, there exists a family of independent Gaussian variables $\nu_{t_k}^{i,j;n}$, $\bar{\mathcal{F}}_{t_k}$-measurable such that

$$(\Gamma_M)^{-1} N_{F^i,M}(x,q) = \sum_{j=1}^{d_x} \varphi^{i,j;n} \nu_{t_k}^{i,j;n}.$$

Moreover, from Proposition 3.6 and Proposition 4.1, we have that $\Phi_F^n(x,q)$ converges uniformly in x in L^2 to $\Phi_F(x)$ as defined in the claim with rate $O(n^{-\beta})$. By Theorem 3.3 we get the same uniform convergence for φ_F^n. The claim follows in this case.

- The case $\beta_0 = \beta_1$ is straightforward from what has been proven in the previous cases.

(ii) Since H, $\tilde{H}^n$ satisfy the same properties as F, $\tilde{F}^n$, we get the claim for R_H, φ_H and $\nu_k^{i,j;n}$ by analogous arguments. Replacing this result in the sensitivity of the Cholesky procedure given in Lemma 3.4, and taking into account the independence of the Gaussian entries, we get the claim for R_G and φ_G. $\square$

Let $\{\upsilon_n\}$ be a sequence of increasing positive numbers, and let us consider the sequence of rescaled error processes ζ^n, defined by

$$\zeta_t^n := \upsilon_n (X_t - \tilde{X}_t^n).$$

We can show that this sequence of processes converges in distribution in the uniform convergence topology to a process ζ defined as the solution to a certain stochastic differential equation. We divide the analysis in two main cases: a first one in which $G(x) \equiv 0$, that is, when X is the solution to an ordinary differential equation, and the case when $G(x)$ is nondegenerate. Just as in the asymptotic error obtained for the usual stochastic Euler method given in Jacod and Protter (1998), we will obtain different rates and different components in the equation for both cases.

THEOREM 4.12 (Limit distribution). *Under the assumptions and notation of Proposition 4.11, let $\rho, R_F, \varphi_F, R_G, \varphi_G$ be defined as in Proposition 4.11 and β defined in (31).*

(i) *[ODE case-$G(x) \equiv 0$.] Let B^1 be the Brownian process given in Proposition 4.10. Let $r = 1 \wedge (1/2 + \beta)$, and suppose $\rho \geq r - \beta$. Let*

$$\zeta_t^n := n^r (X_t - \tilde{X}_t^n).$$

Then $\zeta^n \Rightarrow \zeta^\infty$ in the uniform convergence sense, where ζ^∞ is solution of the system

$$\zeta_t^{\infty,i} = \sum_{j=1}^{d_x}\left(\int_0^t \partial_{x^j} F^i(X_s)\zeta_s^{\infty,j}\,ds + \mathbf{1}_{\{\beta\ge 1/2\}}\frac{1}{2}\int_0^t \partial_{x^j} F^i(X_s)F^j(X_s)\,ds\right) + \mathbf{1}_{\{\beta\le 1/2\}}\left(\int_0^t R_F^i(X_s)\,ds + \sum_{l=1}^{d_x}\int_0^t \varphi_F^{i,l}(X_s)\,dB_s^{1;l}\right). \tag{57}$$

(ii) [*SDE case-$G(x)\neq 0$.*] *Let B^2 and B^3 be the independent Brownian processes given in Proposition 4.10. Let $r = (1/2\wedge\beta)$ and*

$$\zeta_t^n := n^r\left(X_t - \tilde{X}_t^n\right).$$

Then $\zeta^n \Rightarrow \zeta^\infty$, where ζ^∞ is solution of the system for $i = 1,\dots,d_x$ of

$$\begin{aligned}\zeta_t^{\infty,i} &= \sum_j\left(\int_0^t \partial_{x^j} F^i(X_s)\zeta_s^{\infty,j}\,ds + \int_0^t R_G^{i,j}(X_s)\,dW_s^j\right)\\ &\quad + \mathbf{1}_{\{\beta\le 1/2\}}\sum_{j,k,l=1}^{d_x}\int_0^t \varphi_G^{i,j,l,k}(X_s)\,dB_s^{3;l,k,j}\\ &\quad + \mathbf{1}_{\{\beta\le 1/2\}}\sum_{j,l=1}^{d_x}\int_0^t \partial_{x^j} G^{i,l}(X_s)\zeta_s^{\infty,j}\,dW_s^l\\ &\quad + \mathbf{1}_{\{\beta\ge 1/2\}}\frac{1}{\sqrt{2}}\sum_{j,k,l=1}^{d_x}\int_0^t \partial_{x^j} G^{i,l}(X_s)G^{j,k}(X_s)\,dB_s^{2;k,l}.\end{aligned} \tag{58}$$

Let us remark that if $\beta > 1/2$ in Theorem 4.12, the error of the Euler scheme dominates: we recover the limit distribution error for an Euler scheme with exact coefficients given in Kurtz and Protter (1991b) or Jacod and Protter (1998). By contrast, if $\beta < 1/2$, it is the decreasing Euler estimate error that becomes dominant. Since a higher β is generally only achieved by paying a higher price in the required number of steps for the decreasing Euler step, the optimal choice implies fixing $\beta = 1/2$.

Before proving Theorem 4.12, let us show how it implies Theorem 2.1.

PROOF OF THEOREM 2.1. The result is obtained, from Theorems 4.6 and 4.12, since $(\mathcal{H}_{s.s.})$ and $(\mathcal{H}_{f.s.})$ are directly assumed and as the sequence defined as $\gamma_k = \gamma_1 k^{-\theta}$ for $0 < \theta < 1$ satisfies Hypothesis $(\mathcal{H}_\gamma)$. Moreover, recall that we fixed $M(n) = \lceil M_1 n^{1/(1-\theta)} \rceil$, and we have for n large enough,

$$\Gamma_M \approx \frac{\gamma_0 M_1^{1-\theta} n}{1-\theta}, \qquad \frac{\Gamma_M^{[\varsigma/2]}}{\Gamma_M} \approx \frac{(1-\theta) M_1^{-(\varsigma/2-1)\theta} n^{-(\varsigma/2-1)\theta/(1-\theta)}}{1-\varsigma\theta/2},$$
$$\frac{\Gamma_M^{[\varsigma/2+1]}}{\Gamma_M^{[\varsigma/2]}} \approx \frac{(1-\varsigma\theta/2) M_1^{-\theta} n^{-\theta/(1-\theta)}}{1-(\varsigma/2+1)\theta},$$

so that we get from Proposition 4.11, that $\beta_0 = 1/2$ and

$$\beta_1 = \frac{(\varsigma/2-1)\theta}{1-\theta}, \qquad \beta_2 = \frac{\theta}{1-\theta}, \qquad C_0 \approx \frac{\gamma_0 M_1^{1-\theta}}{1-\theta}, \qquad C_1 \approx \frac{(1-\theta) M_1^{-\theta}}{1-2\theta}.$$

Recall that ς is defined in (29) and stands for the first nonzero term in the error expansion of the decreasing Euler estimator. Let us assume we are in the worst case when it attains its minimal value $\varsigma = 4$. Hence

$$\beta_1 = \frac{\theta}{1-\theta}, \qquad C_1 = \frac{(1-\theta) M_1^{-\theta}}{1-2\theta}.$$

Let us now deduce the conditions on θ are then deduced from the conditions in Theorem 4.12 for each of our study cases:

- ODE with random coefficients: From the conditions of Theorem 4.12 we have
$$r = 1 \wedge \left(\tfrac{1}{2} + \beta\right) = \tfrac{1}{2} + \left(\tfrac{1}{2} \wedge \beta\right) = \tfrac{1}{2} + \left(\tfrac{1}{2} \wedge (\beta_0 \wedge \beta_1)\right) = \tfrac{1}{2} + \beta$$
since we should verify $\rho \geq r - \beta = 1/2$, this implies
$$|\beta_0 - \beta_1| = \left| \frac{1}{2} - \frac{\theta}{1-\theta} \right| \geq \rho \geq \frac{1}{2},$$
which is the case if $\theta \in [1/2, 1)$. Moreover, since in this case $\beta_1 \geq 1 > \beta_0 = 1/2$, we get $r = 1/2$, and the R_F term disappears.
- Full SDE case: We have $r = \beta = 1/2 \wedge (\theta/(1-\theta))$ the only restriction comes from imposing $\beta = 1/2$. This is obtained for $1/3 \leq \theta < 1$. Note that the R_G term is different from zero only if $\theta = 1/3$.

Finally, note that if $\varsigma > 4$, we get from the constraints $\theta \in [1/2, 1)$ in the ODE with random coefficients case that $\beta_1 > \beta_0 + 1/2$ and from fixing $\theta \in [1/3, 1)$ in the full SDE case that $\beta_1 > \beta_0 = 1/2$, $\beta_1 > \beta_2$. In both those cases the R_G term is zero. □

REMARK 4.13. It should be noted from the proof of Theorem 2.1 that knowing a priori that $\varsigma > 4$ makes it possible to obtain a lower inferior bound for θ in the theorem. Since in general we do not know ς, we have stated our results with the sometimes sub-optimal limits.

PROOF OF THEOREM 4.12. (a) Let us deal first with the full SDE case. We have from the definition of ζ^n that

$$\zeta_t^n = \int_0^t n^r \big(F(X_s) - \tilde{F}^n(\tilde{X}_{\underline{s}}^n, \underline{s})\big)\, ds + \int_0^t n^r \big(G(X_s) - \tilde{G}^n(\tilde{X}_{\underline{s}}^n, \underline{s})\big)\, dW_s. \tag{59}$$

Let us examine each one of these terms separately. Denoting by x^i the ith component of x, let $x, y \in \mathbb{R}^{d_x}$. We define the set of vectors $\Delta^j(x, y)$

$$\Delta^j(x, y) := \begin{cases} x, & \text{for } j = 0, \\ (y_1, y_2, \dots, y_j, x_{j+1}, x_{j+2}, \dots, x_{d_x})^*, & \text{for } 1 \le j \le d_x, \end{cases}$$

and

$$\Delta^j F^i(x, y) := \mathbf{1}_{\{x^j \neq y^j\}} \left(\frac{F^i(\Delta^{j-1}(x, y)) - F^i(\Delta^j(x, y))}{x^j - y^j} \right) + \mathbf{1}_{\{x^j = y^j\}} \partial_{x^j} F^i(x),$$

and recalling that

$$\tilde{X}_s^{j,n} - \tilde{X}_{\underline{s}}^{j,n} = F^j(\tilde{X}_{\underline{s}}^n)(s - \underline{s}) + \sum_{l=1}^{d_x} G^{j,l}(\tilde{X}_{\underline{s}}^n)(W_s^l - W_{\underline{s}}^l),$$

we have

$$\begin{aligned} &\int_0^t n^r \big[F^i(X_s) - \tilde{F}^{i;n}(\tilde{X}_{\underline{s}}^n, \underline{s})\big]\, ds \\ &\quad = \int_0^t n^r \big(F^i(X_s) - F^i(\tilde{X}_s^n)\big)\, ds + \int_0^t n^r \big(F^i(\tilde{X}_s^n) - F^i(\tilde{X}_{\underline{s}}^n)\big)\, ds \\ &\qquad + \int_0^t n^r \big(F^i(\tilde{X}_{\underline{s}}^n) - \tilde{F}^{i;n}(\tilde{X}_{\underline{s}}^n, \underline{s})\big)\, ds \end{aligned}$$

so that

$$\begin{aligned} &\int_0^t n^r \big[F^i(X_s) - \tilde{F}^{i;n}(\tilde{X}_{\underline{s}}^n, \underline{s})\big]\, ds \\ &\quad = \int_0^t \sum_j \Big[n^r \Delta^j F^i(X_s, \tilde{X}_s^n)(X_s^j - \tilde{X}_s^{j;n}) \\ &\qquad\qquad + n^r \Delta^j F^i(\tilde{X}_s^n, \tilde{X}_{\underline{s}}^n) F^j(\tilde{X}_{\underline{s}}^n)(s - \underline{s}) \\ &\qquad\qquad + \sum_{l=1}^{d_x} n^r \Delta^j F^i(\tilde{X}_s^n, \tilde{X}_{\underline{s}}^n) G^{j,l}(\tilde{X}_{\underline{s}}^n)(W_s^l - W_{\underline{s}}^l) \Big]\, ds \\ &\qquad + n^{r-\beta} \int_0^t n^\beta \big(F^i(\tilde{X}_{\underline{s}}^n) - \tilde{F}^{i;n}(\tilde{X}_{\underline{s}}^n, \underline{s})\big)\, ds. \end{aligned}$$

Following the same approach we obtain for each $l = 1, \dots, d_x$,

$$\begin{aligned}
&\int_0^t n^r \big[G^{i,l}(X_s) - \tilde{G}^{i,l;n}(\tilde{X}^n_{\underline{s}}, \underline{s})\big]\, dW^l_s \\
&\quad = \int_0^t \sum_j \Bigg[n^r \Delta^j G^{i,l}(X_s, \tilde{X}^n_s)(X^j_s - \tilde{X}^{j;n}_s) \\
&\qquad\qquad + n^r \Delta^j G^{i,l}(\tilde{X}^n_s, \tilde{X}^n_{\underline{s}}) F^j(\tilde{X}^n_{\underline{s}})(s - \underline{s}) \\
&\qquad\qquad + \sum_{k=1}^{d_x} n^r \Delta^j G^{i,l}(\tilde{X}^n_s, \tilde{X}^n_{\underline{s}}) G^{j,k}(\tilde{X}^n_{\underline{s}})(W^k_s - W^k_{\underline{s}}) \Bigg] dW^l_s \\
&\qquad + n^{r-\beta} \int_0^t n^\beta \big(G^{i,l}(\tilde{X}^n_{\underline{s}}) - \tilde{G}^{i,l;n}(\tilde{X}^n_{\underline{s}}, \underline{s})\big)\, dW^l_s.
\end{aligned}$$

By identifying terms in the obvious way, we write

$$\zeta^{i,n}_t = \big(P^{i,n}_1(t) + P^{i,n}_2(t)\big) + \int_0^t \langle Q^{i;n}_1(s), \zeta^n_s\rangle\, ds + \sum_{l=1}^{d_x} \int_0^t \langle Q^{i,l;n}_2(s), \zeta^n_s\rangle\, dW^l_s,$$

where Q^i_1, $Q^{i,l}_2$ are d_x dimensional random processes with components

$$Q^{j,i;n}_1(s) = \Delta^j F^i(X_s, \tilde{X}^n_s), \qquad Q^{j,i,l;n}_2(s) = \Delta^j G^{i,l}(X_s, \tilde{X}^n_s)$$

and

$$\begin{aligned}
P^{i;n}_2(s) &= n^{r-\beta} \int_0^t n^\beta \big(F^i(\tilde{X}^n_{\underline{s}}) - \tilde{F}^{i;n}(\tilde{X}^n_{\underline{s}}, \underline{s})\big)\, ds \\
&\qquad + n^{r-\beta} \int_0^t n^\beta \big(G^{i,l}(\tilde{X}^n_{\underline{s}}) - \tilde{G}^{i,l;n}(\tilde{X}^n_{\underline{s}}, \underline{s})\big)\, dW^l_s.
\end{aligned}$$

$$\begin{aligned}
P^{i;n}_1(s) &= \int_0^t \sum_j \Bigg[n^r \Delta^j F^i(\tilde{X}^n_s, \tilde{X}^n_{\underline{s}}) F^j(\tilde{X}^n_{\underline{s}})(s - \underline{s}) \\
&\qquad\qquad + \sum_{l=1}^{d_x} n^r \Delta^j F^i(\tilde{X}^n_s, \tilde{X}^n_{\underline{s}}) G^{j,l}(\tilde{X}^n_{\underline{s}})(W^l_s - W^l_{\underline{s}}) \Bigg] ds \\
&\quad + \int_0^t \sum_{j,l=1}^{d_x} \Bigg[n^r \Delta^j G^{i,l}(\tilde{X}^n_s, \tilde{X}^n_{\underline{s}}) F^j(\tilde{X}^n_{\underline{s}})(s - \underline{s}) \\
&\qquad\qquad \times \sum_{k=1}^{d_x} n^r \Delta^j G^{i,l}(\tilde{X}^n_s, \tilde{X}^n_{\underline{s}}) G^{j,k}(\tilde{X}^n_{\underline{s}})(W^k_s - W^k_{\underline{s}}) \Bigg] dW^l_s.
\end{aligned}$$

(b) In this step, we introduce a nicer diffusion and study its convergence, and prove it shares the limit distribution of the previous SDE. Let

$$\check{\zeta}_t^{i,n} = \left(\check{P}_1^{i,n}(t) + \check{P}_2^{i,n}(t)\right) + \int_0^t \langle \check{Q}_1^{i;n}(s), \check{\zeta}_s^n \rangle \, ds + \sum_{l=1}^{d_x} \int_0^t \langle \check{Q}_2^{i,l;n}(s), \check{\zeta}_s^n \rangle \, dW_s^l,$$

where

$$\check{Q}_1^{i;n}(s) = \nabla F^i(X_s); \qquad \check{Q}_2^{i,l;n}(s) = \nabla G^{i,l}(X_s);$$

$$\begin{aligned}\check{P}_1^{i;n}(s) = {} & \frac{1}{2} \int_0^t n^r \langle \nabla F^i(X_s), F(X_s) \rangle \, dA^{0;n} \\ & + \sum_{l=1}^{d_x} \int_0^t n^r \langle \nabla F^i(X_s), G^{\cdot,l}(X_s) \rangle \, dA_s^{1;l;n} \\ & + \int_0^t n^r \langle \nabla G^{i,l}(X_s), F(X_s) \rangle \, dB_s^{0;l,n} \\ & + \sum_{k,l=1}^{d_x} \frac{1}{\sqrt{2}} \int_0^t n^r \langle \nabla G^{i,l}(X_s) G^{\cdot,k}(X_s) \rangle \, dB_s^{2;k,l,n},\end{aligned}$$

$$\begin{aligned}\check{P}_2^{i;n}(s) = {} & n^{r-\beta} \int_0^t \sum_{j,k,l=1}^{d_x} \varphi_G^{i,j,l,k}(X_s) \, dB_s^{3;l,k,j,n} + n^{r-\beta} \int_0^t \sum_{j=1}^{d_x} R_G^{i,j}(X_s) \, dW_s^j \\ & + n^{r-\beta} \int_0^t \sum_{j=1}^{d_x} \varphi_F^{i,j}(X_s) \, dB_s^{1;j,n} + n^{r-\beta} \int_0^t R_F^i(X_s) \, ds,\end{aligned}$$

for $R_F, R_G, \varphi_F, \varphi_G$ defined in Proposition 4.11. By $(\mathcal{H}_{s.s.})$, F, G are bounded; by Lemma 3.13, ∇F and ∇G are well defined and bounded and have bounded derivatives; and from the definition of $R_F, R_G, \varphi_F, \varphi_G$ are C_b^1.

Note that (46) in Proposition 4.10 gives us goodness and convergence of the tuple $(n^r A^{0,n}, n^r A^{1,n}, n^r B^{0,n}, B^{1,n}, n^r B^{2,n}, B^{3,n})$. Hence, by virtue of Theorem 5.4 in Kurtz and Protter (1991a) $\check{\zeta}^n(\cdot \wedge \tau_K^n)$ is tight and any limit point will satisfy (58) on the interval $[0, \tau_K]$ where $\tau_K = (\inf\{t : |\zeta(t)| > K\} \wedge T)$. Moreover

$$\sup_{0 \le s \le \tau_K} \|\check{P}_s^n\|, \qquad \sup_{0 \le s \le \tau_K} \|\check{Q}_1^n(s)\|, \qquad \sup_{0 \le s \le \tau_K} \|\check{Q}_2^n(s)\|$$

are tight.

(c) We prove now that both ζ^n and $\check{\zeta}^n$ have the same limit on the interval $[0, \tau_K]$. By Theorem 4.9, it suffices to prove that sup norm of the difference of the coefficients converge in probability. By Theorem 4.6 the regularity properties of F and the mean value theorem we have

$$\mathbb{E}\Big[\sup_{0 \le t \le \tau_K} \big|Q_1^{i,n}(t) - \check{Q}_1^{i,n}(t)\big|\Big] \le \mathbb{E}\Big[\sup_x \big|D^2 F(x)\big| \sup_{0 \le t \le \tau_K} \big|X_t - \tilde{X}_t^n\big|\Big] \to 0.$$

The terms of Q_2^n, P_1^n are treated in the same way. On the other hand, we get from Corollary 4.4, Proposition 3.2, and Burckholder–Davis–Gundy inequality that

$$\sup_{0\le t\le T}\left|n^{r-\beta}\int_0^t\left[n^\beta\big(F^i(\cdot)-\tilde{F}^{i;n}(\cdot,\underline{s})\big)-\sum_{j=1}^{d_x}\varphi_F^{i,j}(\cdot)\nu_{\underline{s}}^{j;n}-R_F^i(\cdot)\right](\tilde{X}_{\underline{s}}^n)\,ds\right|,$$

$$\begin{aligned}\sup_{0\le t\le T}\Bigg|n^{r-\beta}\int_0^t\Bigg[&n^\beta\big(G^{i,j}(\cdot)-\tilde{G}^{i,j;n}(\cdot,\underline{s})\big)\\&-\sum_{j,k=1}^{d_x}\varphi_G^{i,j,l,k}(\cdot)\nu_{\underline{s}}^{l,k;n}-R_G^{i,j}(\cdot)\Bigg](\tilde{X}_{\underline{s}}^n)\,dW_s^l\Bigg|\end{aligned}$$

are tight and converge to zero.

Thus, by Theorem 4.9 we will have that $\zeta^{i;n}$ and $\check{\zeta}^{i;n}$ will converge to the same limit.

(d) Finally, note that $\tau_K^n\to\infty$ and $\tau_K\to\infty$, proving our claim in the full SDE case.

(e) To prove (i) it suffices to follow the same approach. We obtain an equivalent development for the ODE with random coefficients case (replacing by zero all the "g-terms"). The rest of the proof proceeds as before, this time using (45) for the weak convergence of the tuple. □

5. The EMsDS algorithm. Given the error expansion for the decreasing step algorithm presented in Proposition 3.8, it seems natural to explore if a Richardson–Romberg extrapolation may be used to obtain the approximation with the same convergence properties we have proven. The idea of such a procedure is to decrease the complexity by performing a linear combination of two (or more) realizations of the algorithm with carefully chosen parameters. We borrow here the procedure as defined in Lemaire (2005).

Let λ be a positive real. If $\{\gamma_k\}$ is a sequence of steps satisfying $(\mathcal{H}_\gamma)$, the sequence $\gamma_k^\lambda:=\frac{\gamma_k}{\lambda}$ will also satisfy $(\mathcal{H}_\gamma)$. We will denote Γ_M^λ and $\Gamma_M^{\lambda,[r]}$ the sum of the γ_k^λ and its power as before.

Let us denote by $\tilde{F}^{\lambda,M}(x,q)$ the approximation as defined in (8) when the coefficients $\{\gamma_k^\lambda\}_{k\in\mathbb{N}^*}$ are used.

With ς given as in (29), let us define the *extrapolated approximation estimator* as

$$\hat{F}^{\lambda;M(n)}(x,q)=\frac{1}{\lambda^{\varsigma/2-1}-1}\big(\lambda^{\varsigma/2-1}\tilde{F}^{\lambda,M(n)}(x,q)-\tilde{F}^{M(n)}(x,q)\big). \tag{60}$$

The first question we might ask is if estimator (60) does converge to the actual ergodic average, and what type of properties it inherits. To clarify the situation consider an extension of (2). Let $\vec{Y}^x = (Y^{1;x}, Y^{2;x})^*$ with

$$
\begin{aligned}
Y_t^{1;x} &= y_0^1 + \int_0^t \frac{b(x, Y_s^{1;x})}{\lambda}\, ds + \int_0^t \frac{\sigma(x, Y_s^{1;x})}{\sqrt{\lambda}}\, d\hat{W}_s^1, \\
Y_t^{2;x} &= y_0^2 + \int_0^t b(x, Y_s^{2;x})\, ds + \int_0^t \sigma(x, Y_s^{2;x})\, d\hat{W}_s^2.
\end{aligned} \tag{61}
$$

If $\hat{W}^1$ and $\hat{W}^2$ are independent, then this system satisfies $(\mathcal{H}_{f.s.})$ with a unique invariant measure defined by $\vec{\mu}^x(d\vec{y}) = \mu^x(dy^1)\mu^x(dy^2)$. If we define

$$
\vec{f}(x, \vec{y}) := \frac{1}{\lambda^{\varsigma/2-1} - 1}\bigl(\lambda^{\varsigma/2-1} f(x, y^1) - f(x, y^2)\bigr), \tag{62}
$$

and defining in an analogous way $\vec{h}$, then it can be seen that $\vec{f}, \vec{g}, \vec{h} := \vec{g}\vec{g}^*$ satisfy $(\mathcal{H}_{s.s.})$. Moreover if we apply the decreasing step algorithm to $\vec{f}$ (resp., $\vec{h}$) in the extended framework, we obtain the expression (60). *Hence, we conclude that the EMsDS algorithm is equivalent to the MsDS algorithm applied to an extended system.*

Let us denote by $\hat{X}^n$ the approximation of the diffusion X using the extrapolated version of the algorithm. In view of the discussion we presented before, the following result is mainly a corollary of Theorems 4.6 and 4.12, and extends the main Theorem to the extrapolation algorithm. It shows the advantage of using the EMsDS algorithm: assuming higher regularity, all the properties of the MsDS algorithm are conserved, but the extrapolated version *allows a lower value for θ in the definition of the sequence $\gamma_k = \gamma_0 k^{-\theta}$*. More precisely we pass from $1/2$ to $1/3$ in the ODE case and from $1/3$ to $1/5$ in the SDE case as minimal θ values. As a consequence of this reduction, the complexity of the modified version is in general asymptotically lower than that of the nonextrapolated version (refer to the efficiency analysis on Section 6.1).

THEOREM 5.1. *Let $0 < \theta < 1$, $\gamma_1 \in \mathbb{R}^+$ and $\gamma_k = \gamma_1 k^{-\theta}$. Assume $(\mathcal{H}_{f.s.})$ and $(\mathcal{H}_{s.s.})$, $M(n)$ defined as in Theorem 2.1, and assume in addition that $r_y > 5$. Let $\hat{X}^n$ be the approximated diffusion where we replace the ergodic estimator* (8) *by* (60).

(i) (*Strong convergence*). *There exists a constant K such that*

- *Case $g \equiv 0$* (*ODE with random coefficients*):

$$
\mathbb{E}\Bigl[\sup_{0\le t\le T} |X_t - \hat{X}_t^n|^2\Bigr] \le K n^{-2[(1-\theta)\wedge 2\theta]/(1-\theta)}.
$$

- (*Full SDE case*):

$$
\mathbb{E}\Bigl[\sup_{0\le t\le T} |X_t - \hat{X}_t^n|^2\Bigr] \le K n^{-[(1-\theta)\wedge 4\theta]/(1-\theta)}.
$$

(ii) (*Limit distribution*). *Assume in addition that* $r^y \geq 8$, *and define*

$$\hat{C}_\varphi := \frac{(\lambda^3+1)^{1/2}}{\lambda-1}; \qquad \hat{C}_1 := \frac{\gamma_0^2(1-\theta)M_1^{-2\theta}}{1-3\theta};$$

$$\hat{\varphi}_F(x) := \mathbf{1}_{\{\theta=1/5\}}\hat{C}_\varphi\sqrt{\hat{\Phi}_F(x)}; \qquad \hat{\varphi}_G(x) := \mathbf{1}_{\{\theta=1/5\}}\hat{C}_\varphi\sqrt{\hat{\Phi}_G(x)};$$

$$\hat{R}_F^i(x) := \mathbf{1}_{\{\theta\geq 1/5\}}\hat{C}_1(1-\lambda^{-1})\int \bar{v}_{F^i}^{\varsigma+2,r^y}(x,y)\mu^x(dy);$$

$$\hat{R}_H^{i,j}(x) := \mathbf{1}_{\{\theta\geq 1/5\}}\hat{C}_1(1-\lambda^{-1})\int \bar{v}_{H^{i,j}}^{\varsigma+2,r^y}(x,y)\mu^x(dy).$$

- [*ODE case*: $G(x)\equiv 0$]. *If* $\theta \geq 1/3$, *then* $\hat{\zeta}^n := n(X_t - \hat{X}^n)$ *satisfies the limit distribution result given in Theorem* 2.1(a) *with new coefficients* $\hat{\varphi}^F$ *instead of* φ^F.
- (*SDE case*). *If* $\theta \geq 1/5$, *then* $\hat{\zeta}^n := n^{1/2}(X_t - \hat{X}^n)$ *satisfies the limit distribution result given in Theorem* 2.1(b) *with the coefficients* $\hat{R}^F$, $\hat{R}^G$, $\hat{\varphi}^F$ *and* $\hat{\varphi}^G$ *instead of* R^F, R^G, φ_F *and* φ_g, *respectively.*

PROOF OF THEOREM 5.1. We will deduce the proof only for the full SDE case the other case being analogous. We assume that $\varsigma = 4$, which is the most common case.

(a) As in the proof of Theorem 2.1, the sequence of coefficients satisfies $(\mathcal{H}_\gamma)$. Moreover, the EMsDS algorithm is the MsDS algorithm applied to an extended system, and hence the strong convergence and limit distribution properties are a consequence from Theorems 4.6 and 4.12: it remains just to express the values of the functions and constants appearing in Propositions 4.1 and 4.11 in terms of the original system.

Indeed, recall that

$$\vec{b}(x,\vec{y}) = \begin{pmatrix} \lambda^{-1}b(x,y^1) \\ b(x,y^2) \end{pmatrix}; \qquad \vec{\sigma}(x,\vec{y}) = \begin{pmatrix} \lambda^{-1/2}\sigma(x,y^1) & 0 \\ 0 & \sigma(x,y^2) \end{pmatrix}. \tag{63}$$

By (i) in Proposition 4.1 applied to the extended problem [i.e., for the system (61) and $\vec{f}$ defined in (62)], we have a solution for the extended centered Poisson equation given by

$$\vec{\phi}_{F^i}(x,\vec{y}) = (\lambda-1)^{-1}(\lambda^2\phi_{F^i}(x,y^1) - \phi_{F^i}(x,y^2)),$$

that is, the solution of equation (13) with function F^i under the extended set-up is a linear combination of the solution in the original set-up. Thus, for any $j > 0$,

$$D_y^j\vec{\phi}_{F^i}(x,\vec{y}) = \frac{1}{\lambda-1}\begin{pmatrix} \lambda^2 D_y^j\phi_{F^i}(x,y^1) \\ -D_y^j\phi_{F^i}(x,y^2) \end{pmatrix}. \tag{64}$$

It follows that

$$
\begin{aligned}
D_y^j \vec{\phi}_{F^i}(x,\vec{y})\mathbb{E}[\langle \vec{b}(x,\vec{y})^{\otimes(l-j)}, (\vec{\sigma}(x,\vec{y})U_1^0)^{\otimes(2j-l)}\rangle] \\
= \frac{\lambda^2}{\lambda-1} D_y^j \phi_{F^i}(x,y^1)\mathbb{E}\biggl[\biggl\langle\biggl(\frac{b(x,y^1)}{\lambda}\biggr)^{\otimes(l-j)}, \biggl(\frac{\sigma(x,y^1)}{\sqrt{\lambda}}U_1^0\biggr)^{\otimes(2j-l)}\biggr\rangle\biggr] \\
- \frac{1}{\lambda-1} D_y^j \phi_{F^i}(x,y^2)\mathbb{E}[\langle \vec{b}(x,y^2)^{\otimes(l-j)}, (\sigma(x,y^2)U_1^0)^{\otimes(2j-l)}\rangle].
\end{aligned}
$$

Therefore

$$
\vec{\bar{v}}^{l,r^y}_{F^i} = \biggl(\frac{\lambda^{(4-l)/2}-1}{\lambda-1}\biggr)\bar{v}^{l,r^y}_{F^i}, \tag{65}
$$

and we deduce that the terms of the error expansion will be zero for $l \le 5$.

(b) Let $\vec{\varsigma}$ be defined by (29) under the extended setup. From (65) we conclude that $\vec{\varsigma} \ge 6$, being $\vec{\varsigma} = 6$ the worst case. Hence, we deduce that defining

$$
\beta_0 = \frac{1}{2}, \qquad \hat{\beta}_1 = \frac{2\theta}{1-\theta}, \qquad \hat{\beta}_2 = \frac{\theta}{1-\theta},
$$

then

$$
\begin{aligned}
&\Gamma_M \approx \frac{\gamma_0 M_1^{1-\theta} n}{1-\theta}, \qquad \frac{\Gamma_M^{[3]}}{\Gamma_M} \approx \frac{\gamma_0^2(1-\theta)M_1^{-2\theta}n^{-2\theta/(1-\theta)}}{1-3\theta}, \\
&\frac{\Gamma_M^{[4]}}{\Gamma_M^{[3]}} \approx \frac{\gamma_0(1-3\theta)M_1^{-\theta}n^{-\theta/(1-\theta)}}{1-4\theta},
\end{aligned}
$$

and so, $\beta_1, \hat{\beta}_2, \hat{\beta}_3$ are the coefficients appearing in Proposition 4.1 applied to this setup. We conclude as well that $\hat{R}^i_F$ is the function appearing in Proposition 4.11. Similar developments for H allow us to extend the conclusion to $\hat{R}^{i,j}_H$.

(c) Finally, looking at the definition of φ_F and Φ_F from Proposition 4.11 and (64) we get that

$$
\begin{aligned}
\hat{\Phi}^{i,j}_F(x) = \frac{C_0^{-1}}{(\lambda-1)^2}\biggl(\lambda^2 \int \langle \sigma^* D_y\phi_{F^i}, \sigma^* D_y \phi_{F^j}\rangle(x,y^1)\mu^x(dy^1) \\
+ \int \langle \sigma^* D_y\phi_{F^i}, \sigma^* D_y \phi_{F^j}\rangle(x,y^2)\mu^x(dy^2)\biggr);
\end{aligned}
$$

that is, $\hat{\Phi}_F(x) = (\lambda^2+1)(\lambda-1)^{-2}\Phi_F(x)$. We get a similar result for $\hat{\Phi}_G$. We obtain the value $\hat{C}_\varphi$ given in the statement. The claim follows. □

REMARK 5.2. $\hat{C}_\varphi$ is a constant multiplying the uncertainty coming from the decreasing step estimator. Since we would like this quantity as small as possible, having an explicit value for $\hat{C}_\varphi$ is very useful from a numerical point of view: we can choose λ to minimize $\hat{C}_\varphi$. We get

$$\lambda_* = 1 + (\sqrt{3}+1)^{1/3} + (\sqrt{3}+1)^{-1/3} \approx 3.196$$

inducing $\hat{C}_\varphi \approx 2.64$. This is the initial additional cost that has to be paid for the extrapolation, making the EMsDS algorithm useful for large n, where the reduction in complexity of the EMsDS is enough to compensate for the higher error.

6. Numerical results.

6.1. *Efficiency analysis.* We can approximate the execution time of both algorithms, the original and extrapolated versions of the algorithm, by estimating the total number of operations needed to perform one path approximation of the effective equation (3). Note that since both algorithms share the same structure, a similar analysis is valid for both of them: the total cost $\kappa(n)$ of the algorithm with n steps may be written as

$$\kappa(n) = \left[\kappa_1(n, d_x, d_y) + \kappa_2(d_x)\right]n,$$

where κ_1 stands for the cost coefficient estimation at each step of the decreasing Euler, and κ_2 for the cost of calculating the Euler iteration. The latter will be of order $O(d_x)$ in the ODE case and $O(d_x^2)$ for the SDE case.

Let us focus now on κ_1. Both algorithms perform $M_1 n^{1/(1-\theta)}$ iterations for approximating the diffusion $\tilde{Y}$ and the calculation of estimators $\tilde{F}, \tilde{G}$. For the MsDS algorithm, each one of these iterations has a cost of $O(d_y d_x)$ in the *ODE* case, or $O(d_y d_x^2)$ in the SDE case. In the latter, we need also to perform a Cholesky decomposition with a cost of $O(d_x^3)$ operations. Hence

$$\kappa_1^{\mathrm{MsDS}}(n, d_x, d_y) = \begin{cases} O\left(d_y d_x n^{1/(1-\theta)}\right), & \text{in ODE case,} \\ O\left(\left[d_y d_x^2 + d_x^3\right] n^{1/(1-\theta)}\right), & \text{in SDE case.} \end{cases}$$

On the other hand, from the definition of the EMsDs algorithm, we get $\kappa_1^{\mathrm{EMsDS}} \le \lambda \kappa_1^{\mathrm{MsDS}}$, and thus both share the same order of complexity, with the only difference that θ *is allowed to be smaller in the extrapolated algorithm.*

It may be more interesting to compare the *efficiency* of both algorithms, that is, the time spent to obtain a given error tolerance Δ. We have from Theorems 2.1 and 5.1 that $\Delta(n) := O(n^{-1})$ for the ODE, and $\Delta(n) := O(n^{-1/2})$ for the SDE case. Replacing the minimum possible θ values we obtain the complexity figures given in Table 1.

How do these figures compare with a straightforward Euler scheme applied to the original system? For the ODE case, an Euler scheme implemented for the original system (1) would require a total of $(dx + dy)\varepsilon^{-1}\Delta^{-2}$ operations. Then the MsDS algorithm is more efficient if $\varepsilon < \Delta(d_x \vee d_y)^{-1}$, and the EMsDS

TABLE 1
Minimal efficiency (operations for fixed error) of the basic and extrapolated algorithm for ODE and full SDE cases

	ODE	**ODE (extrapol.)**	**SDE**	**SDE (extrapol.)**
$\theta_{\min}$	$1/2$	$1/3$	$1/3$	$1/5$
$\tau_{\min}(\Delta)$	$O(d_y d_x \Delta^{-3})$	$O(d_x d_y \Delta^{-2.5})$	$O([d_x^2 d_y + d_x^3]\Delta^{-5})$	$O([d_x^2 d_y + d_x^3]\Delta^{-4.5})$

if $\varepsilon < \Delta^{1/2}(d_x \vee d_y)^{-1}$. With respect to the algorithm presented in E, Liu and Vanden-Eijnden (2005), the efficiency is equivalent to the one obtained when using a weak scheme of order one for approximating the ergodic averages. The advantage of our method is that we have in addition to the rate of convergence an expression for a C.L.T. type result.

In the SDE case, on the other hand, the proposed algorithm will be advantageous in the case in which $\varepsilon < \Delta^3(d_x \vee d_y)^{-1}$ for the MsDS version, and $\varepsilon < \Delta^{2.5}(d_x \vee d_y)^{-1}$ for the EMsDS. In other words, our proposed algorithms will be more efficient in our regime of interest of a strong scale separation (i.e., when $\varepsilon \to 0$). It should be remarked that the SDE case is not explicitly studied for the algorithm in E, Liu and Vanden-Eijnden (2005).

6.2. *Numerical tests.*

6.2.1. *A toy problem.* Let us illustrate the main features of the algorithm by evaluating its behavior when used for solving a toy system for which we are able to obtain an exact solution. Consider

$$dY_t^x = \left(\left(|x|^2+1\right)^{-1/2} - Y_t^x\right) + \sqrt{2}\, d\tilde{W}_t,$$

which is an Ornstein–Uhlenbeck system having a unique invariant measure with normal distribution with mean $(|x|^2+1)^{-1/2}$ and variance 1, and define the SDE system

$$dX_t = F(X_t)\,dt + G(X_t)\,dW_t,$$

with

$$f(x,y) := \begin{pmatrix} 1 + y - \left(|x|^2+1\right)^{-1/2} \\ 1 \end{pmatrix};$$

$$g(x,y) := \sqrt{\frac{|x|^2+1}{2|x|^2+3}(y^2+1)}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

with F, G defined as before and where $\tilde{W}$ is a real Brownian motion independent of the planar Brownian motion W. The form of the assumed coefficients is chosen to satisfy the regularity and uniform bound hypothesis in $(\mathcal{H}_{s.s.})$ and $(\mathcal{H}_{f.s.})$ and

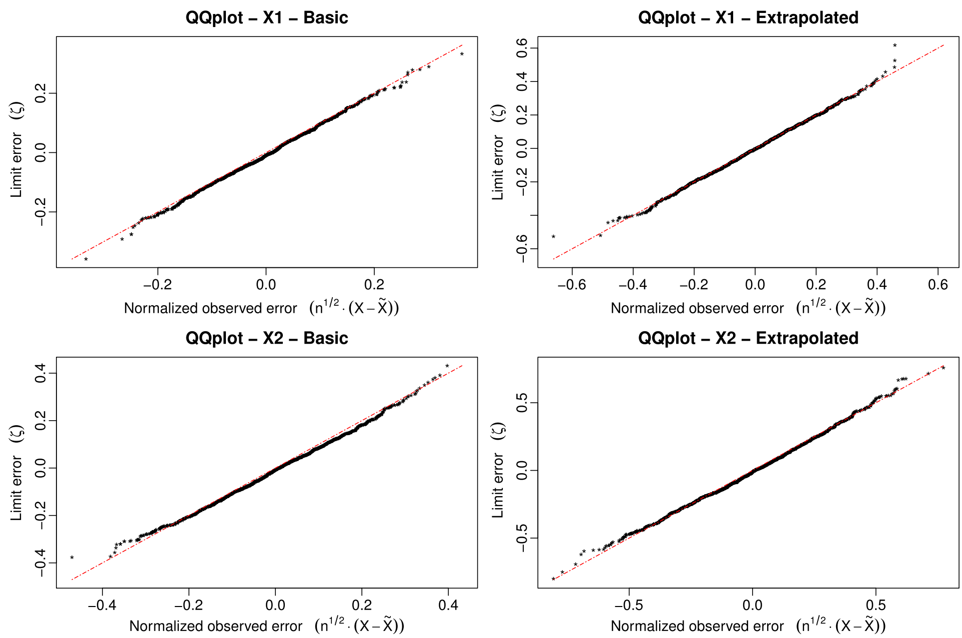

FIG. 1. *Q–Q plot comparing the rescaled errors in the simulation with $n = 510$ and the theoretical limit distribution (the reference line represents a perfect match). Left: SDE decreasing step. Right: SDE interpolated.*

to give a simple effective equation expression. In fact, it is easily verified that the exact effective equation is

$$X_s = \begin{pmatrix} x_0^1 + s + W_s^1 \\ x_0^2 + s + W_s^1 + W_s^2 \end{pmatrix}.$$

We will look at the numerical results of applying the decreasing step with sequence $\gamma_k = k^{-1/3}$ and the EMsDS version with sequence $\gamma_k = k^{-1/5}$ and $\lambda = 3$. Let us examine the distribution of the error at a fixed time $T = 1$ (i.e., $\zeta = \tilde{X}_1 - X_1$). Figure 1 shows a Q–Q plot of the rescaled simulated errors $\sqrt{n}\zeta$ and the limit distribution error in the studied cases. As shown, the empirical distributions obtained after 1600 simulations with $n = 510$ verify the expected limit behavior.

Figure 2. Left plots in a log–log scale the evolution of the L_2 error

$$\zeta_{L_2} = \left(\mathbb{E}\Big[\sup_{0\le t\le T} |\tilde{X}_t - X_t|^2\Big]\right)^{1/2}$$

in function of the number of steps n, comparing both versions of the algorithm. The empirically obtained slope (close to -0.5 in both cases) represents the power of the approximation and is the one expected from the convergence theorems.

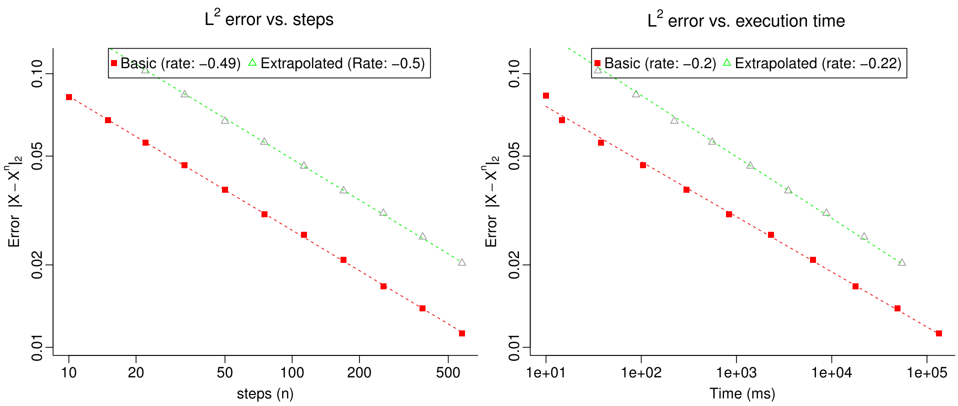

FIG. 2. *Left*: *L_2 error as a function of steps for the SDE case (log–log scale). Note that the estimated values for the slopes verify the rate of convergence for the algorithm in both implementations.* *Right*: *L_2 error as a function of execution time for the SDE case (log–log scale). Although a higher price must be payed for a small step number, the slope difference signals a change in the asymptotic order of convergence.*

We show as well in Figure 2 (right) a comparison in the efficiency of both methods (measured as the error in terms of the execution time) of each one of the algorithms. The effect of the extrapolation in the cost of the algorithm is evidenced in the difference in slope of the empirical plot for both algorithms. Note that solving for Δ in Table 1 we get $\Delta_{\text{MsDS}} = O(\tau^{-0.2})$ and $\Delta_{\text{EMsDS}} \approx O(\tau^{-0.222})$, values that are retrieved in the numerical experiment. It is worth observing the difference in the intercept of both lines, showing that the higher slope comes with a cost in the initial error. The conclusion drawn from the toy example may well be generalized: the user should consider implementing the extrapolated version only when requiring a very high precision on the approximation results.

6.2.2. *Pricing in finance*. We apply now the algorithm to a pricing problem in finance. Consider the mean-reverting corrected Heston's stochastic volatility model presented in Fouque and Lorig (2011) and given by

$$
\begin{aligned}
dX_t &= rX_t\,dt + \Sigma_t X_t\,dW_t^x,\\
dY_t &= \varepsilon^{-1} Z_t(m - Y_t)\,dt + \nu\sqrt{2Z_t\varepsilon^{-1}}\,dW_t^y,\\
dZ_t &= \kappa(\theta - Z_t)\,dt + \sigma\sqrt{Z_t}\,dW_t^z,\\
\Sigma_t &= \sqrt{Z_t}(1 + Y_t^2),
\end{aligned}
$$

where we assume W_t^x, W_t^y, W_t^z are one-dimensional Brownian motions with correlations ρ_{xy}, ρ_{xz} and ρ_{yz}. We suppose the model is already written in terms of the risk neutral probability measure with known parameters and initial conditions given in Table 2. We are interested in pricing several types of options depending

TABLE 2
Initial condition and parameters of the model

x_0	z_0	y_0	m	ν	κ	r	θ	σ	ρ_{xy}	ρ_{yz}	ρ_{xz}
100	0.24	0.06	0.06	1.0	1.0	0.05	1.0	0.39	0	0	−0.33

on the whole trajectory on this model. For this test, we price a floating strike Asian call (the payoff being $AC_{\text{float}} = S_T - T^{-1} \int S_t \, dt$) and a lookback call with floating strike (with payoff $LC_{\text{float}} = S_T - S_{\min}$).

In this test, we compare the algorithm with a simple Euler scheme with different values for ε. We carry out 6000 Monte Carlo simulations. The results are presented in Table 3.

Note that the system does not satisfy all the hypothesis $(\mathcal{H}_{f.s.})$ and $(\mathcal{H}_{s.s.})$, particularly it fails to satisfy the boundedness of the coefficients with respect to the slow variables, and the uniform ellipticity hypothesis. Nevertheless, the MsDS algorithm seems to work even under these relaxed conditions, and, in addition, appears to be more stable than the algorithm using small values of ε. Note as well that for similar values of total operations [represented by the column $n \times M(n)$], the MsDS algorithm gives better results.

APPENDIX A: TECHNICAL RESULTS

A.1. Weak convergence of tuples.

PROOF OF PROPOSITION 4.10. (a) Let us start by proving (45). Note that the approximations defined by (9) are defined in the same sample space of the effective

TABLE 3
Simulation values

Method	ε	n	$M(n)$	$n \times M(n)$	**Asian**	**Lookback**
Euler	10^{-3}	5×10^6	1	5×10^4	40,988	81,591
Euler	10^{-3}	10^7	1	1×10^5	40,503	81,256
Euler	10^{-3}	2×10^7	1	2×10^5	40,086	80,769
Euler	10^{-4}	5×10^6	1	5×10^5	22,091	54,119
Euler	10^{-4}	10^7	1	1×10^6	21,897	53,806
Euler	10^{-4}	2×10^7	1	2×10^6	20,908	52,095
Euler	10^{-5}	5×10^6	1	5×10^6	18,203	45,947
Euler	10^{-5}	10^7	1	1×10^7	15,164	39,123
Euler	10^{-5}	2×10^7	1	2×10^7	20,659	51,240
MsDS	–	50	3540	1.77×10^5	20,738	47,920
MsDS	–	100	10,010	1×10^6	20,681	48,841
MsDS	–	200	28,290	5.66×10^6	20,669	49,557

equation (3) and that we have, thanks to Theorem 4.6, that

$$\sup_{0\le s\le t} |\tilde{X}^n_s - X_s| \xrightarrow{P} 0.$$

Hence,

$$(X, \tilde{X}^n, W) \Rightarrow (X, X, W). \tag{66}$$

Now, $nA^{0;n}$ is deterministic, continuous and

$$\lim_{n\to\infty} nA^{0;n}_t = \lim_{n\to\infty} 2n\left(\frac{t^2}{2} - \frac{\lfloor nt\rfloor(\lfloor nt\rfloor - 1)}{2n^2}\right) = t,$$

and the convergence is uniform in t.

On the other hand, we can easily verify that for any t, and $1 \le i \le d_x$,

$$\sqrt{n}B^{1;i,n}_t = \sqrt{n}\int_0^t \nu^{i;n}_{\underline{s}}\, ds = \sum_{i=0}^{\lfloor nt\rfloor} \frac{1}{\sqrt{n}}\nu^{i;n}_{\underline{s}} + \frac{nt-\lfloor nt\rfloor}{n^{3/2}}\nu^{i;n}_{\underline{t}},$$

but by the Cauchy–Schwarz inequality we have

$$\mathbb{E}\left[\left|\sup_{0\le t\le T}\left(\frac{nt-\lfloor nt\rfloor}{n^{3/2}}\nu^{i;n}_{\underline{t}}\right)\right|^2\right] \le \mathbb{E}\left[\left|\sum_{k=1}^{n} \frac{1}{n^{3/2}}|\nu^{i;n}_{t_k}|\right|^2\right] \le \mathbb{E}\left[\frac{n}{n^3}\sum |\nu^{i;n}_{t_k}|^2\right] \to 0.$$

Then it suffices to study the convergence of the Gaussian martingale $\sum_{i=0}^{\lfloor nt\rfloor} n^{-1/2} \times \nu^{i;n}_{\underline{s}}$. Let $0 \le j \le d_x$. Then the independence properties and an application of a multi-dimensional C.L.T. gives us that

$$\left\langle \sum_{i=0}^{\lfloor nt\rfloor} \frac{1}{\sqrt{n}}\nu^{i;n}_{\underline{s}}, \sum_{i=0}^{\lfloor nt\rfloor} \frac{1}{\sqrt{n}}\nu^{j;n}_{\underline{s}} \right\rangle = \frac{1}{n}\sum_{i=0}^{\lfloor nt\rfloor} \nu^{i;n}_{\underline{s}}\nu^{j;n}_{\underline{s}} \xrightarrow{P} \delta_{i=j}.$$

We conclude that B^1 is (up to a modification) a Brownian motion independent from W and X, by remarking its Gaussian nature with independent increments property and covariance matrix as the one of the standard Brownian. Thus (45) follows. Note that we have shown property $(*)$ as well, and consequently goodness of the sequence.

(b) To prove (46) note first that $\sqrt{n}B^{2;i,j;n}$ is a continuous martingale. In view of the results in Jacod (1997), we examine the component-wise quadratic variation. By standard techniques we find

$$\begin{aligned}\langle \sqrt{n}B^{2;i,j;n}, \sqrt{n}B^{2;i',j;n}\rangle_t &= 2n\int_0^t (W^i_s - W^i_{\underline{s}})(W^{i'}_s - W^{i'}_{\underline{s}})\, ds \\ &\xrightarrow{P} \mathbf{1}_{\{i=i'\}}t,\end{aligned}$$

and due to independence we find as well that, taking, $j \neq j'$,

$$\langle \sqrt{n} B^{2;i,j;n}, W^j \rangle_t \xrightarrow{P} 0,$$
$$\langle \sqrt{n} B^{2;i,j;n}, \sqrt{n} B^{2;i,j';n} \rangle_t = 0; \qquad \langle \sqrt{n} B^{2;k;n}, W^j \rangle_t = 0.$$

By Theorem 4-1 in Jacod (1997), $B^{2;n}$ convergences stably in law toward B^2 a standard Brownian Motion independent from W; for the definition of this type of convergence see Aldous and Eagleson (1978) or Jacod (1997). Since all the processes are continuous, stable convergence in law implies joint convergence. Therefore considering (66), we have

$$(X, \tilde{X}^n, W, B^{2;n}) \Rightarrow (X, X, W, B^2).$$

Note that we proved tightness of the quadratic variation of the martingale $\sqrt{n} B^{2;n}$, so that it has property $(*)$, and therefore it is good.

Now, $B^{3;n}$ is also a continuous Gaussian martingale, and we can make use again of Theorem 4-1 in Jacod (1997). Let us check the convergence in probability of its quadratic variation toward tI and of its quadratic covariation with respect to the other martingales. Indeed, it is straightforward that if $j \neq j'$, $\langle B^{3;i,j;n}, B^{3;i',j';n} \rangle_t = 0$, while we deduce from the multidimensional C.L.T. $\langle B^{3;i,j;n}, B^{3;i',j;n} \rangle_t \xrightarrow{P} \mathbf{1}_{\{i=i'\}} t$. As before this also shows goodness of $B^{3;n}$. Using the same techniques we prove for any i, j, l that $\langle B^{3;l,j;n}, \sqrt{n} B^{2;i;n} \rangle_t = 0$, and $\langle B^{3;l,j;n}, W \rangle = 0$. Hence

$$(X, \tilde{X}^{(n)}, W, B^{2;n}, B^{3;n}) \Rightarrow (X, X, W, B^2, B^3).$$

We prove now the convergence in probability toward zero of the remaining terms in the left side tuple in (46).

Since $n^{-1/2} \to 0$ and $nA^{0,n} \Rightarrow A^0$, we have $n^{1/2} A^{0,n} \Rightarrow 0$ and thus $n^{1/2} \times A^{0,n} \xrightarrow{P} 0$.

On the other hand, for any $t \geq 0$ and k,

$$\begin{aligned}
\mathbb{E}[\langle \sqrt{n} B^{0;k;n}, \sqrt{n} B^{0;k;n} \rangle_t] &= n \int_0^t (s - \underline{s})^2 \, ds \\
&= \sum_{i=0}^{\lfloor nt \rfloor} n \int_0^{1/n} r^2 \, dr + n \int_0^{t - \lfloor nt \rfloor / n} r^2 \, dr \\
&\leq \sum_{i=0}^{\lfloor nt \rfloor + 1} \frac{1}{3n^2} = O(n^{-1}).
\end{aligned}$$

So that by the Burckholder–Davis–Gundy inequality, $\mathbb{E}[\sup_{0 \leq t \leq T} |\sqrt{n} B^{0;n}|^2]$ tends to zero as $n \to \infty$, implying $\sqrt{n} B^{0;n} \xrightarrow{P} 0$. In addition, it can be readily seen that

$$\mathbb{E}[|nA^{1;k;n} A^{1;j;n}|] = 0$$

for $j \neq k$, so that we have by using convex and Cauchy–Schwarz inequalities,

$$\mathbb{E}\Big[\sup_{0\le t\le T}|\sqrt{n}A_t^{1;n}|^2\Big] \le nT\sum_{j=1}^{d_x}\int_0^T \mathbb{E}[(W_s^j - W_{\underline{t}s}^j)] \le \frac{d_x T}{2},$$

and hence, by the law of large numbers, $\sqrt{n}A^{1;n} \xrightarrow{P} \mathbb{E}[\sqrt{n}A^{1;n}] = 0$.

Finally, as $\sqrt{n}B^{1;n}$ converges in law to a Brownian, $B^{3;n} \xrightarrow{P} 0$. Therefore (46) is proved. $\square$

APPENDIX B: CHOLESKY DECOMPOSITION

PROOF OF LEMMA 3.4. Since $G + \Delta G$ is the lower triangular factor of $H + \Delta H$, we have

$$(G_{i,i} + \Delta G_{i,i})^2 = H_{i,i} + \Delta H_{i,i} - \sum_{k=1}^{i-1}(G_{i,k} + \Delta G_{i,k})^2.$$

By algebraic manipulation and the fact that G is the Cholesky decomposition of H, we get

$$\Delta G_{i,i} = \frac{\Delta H_{i,i} - 2\sum_{k=1}^{i-1}\Delta G_{i,k}G_{i,k}}{2G_{i,i}} - \left((\Delta G_{i,i})^2 + \sum_{k=1}^{i-1}(\Delta G_{i,k})^2\right).$$

The first claim follows by controlling the last term by induction in i, Theorem 3.3 and norm equivalence given by (11). The case $i > j$ is proved in the same way. $\square$

Acknowledgments. The author would like to thank François Delarue for his help and support during the preparation of this work and the anonymous referee for his suggestions that greatly improved the paper.

REFERENCES

ALDOUS, D. J. and EAGLESON, G. K. (1978). On mixing and stability of limit theorems. *Ann. Probab.* **6** 325–331. MR0517416

BENSOUSSAN, A., LIONS, J.-L. and PAPANICOLAOU, G. (1978). *Asymptotic Analysis for Periodic Structures*. *Studies in Mathematics and Its Applications* **5**. North-Holland, Amsterdam. MR0503330

E, W., LIU, D. and VANDEN-EIJNDEN, E. (2005). Analysis of multiscale methods for stochastic differential equations. *Comm. Pure Appl. Math.* **58** 1544–1585. MR2165382

E, W., ENGQUIST, B., LI, X., REN, W. and VANDEN-EIJNDEN, E. (2007). Heterogeneous multiscale methods: A review. *Commun. Comput. Phys.* **2** 367–450. MR2314852

FOUQUE, J.-P. and LORIG, M. J. (2011). A fast mean-reverting correction to Heston's stochastic volatility model. *SIAM J. Financial Math.* **2** 221–254. MR2775413

FOUQUE, J.-P., PAPANICOLAOU, G. and SIRCAR, K. R. (2000). *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge Univ. Press, Cambridge. MR1768877

FOUQUE, J.-P., PAPANICOLAOU, G., SIRCAR, R. and SOLNA, K. (2003). Singular perturbations in option pricing. *SIAM J. Appl. Math.* **63** 1648–1665. MR2001213

JACOD, J. (1997). On continuous conditional Gaussian martingales and stable convergence in law. In *Séminaire de Probabilités, XXXI. Lecture Notes in Math.* **1655** 232–246. Springer, Berlin. MR1478732

JACOD, J. and PROTTER, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** 267–307. MR1617049

JAKUBOWSKI, A., MÉMIN, J. and PAGÈS, G. (1989). Convergence en loi des suites d'intégrales stochastiques sur l'espace $\mathbf{D}^1$ de Skorokhod. *Probab. Theory Related Fields* **81** 111–137. MR0981569

KURTZ, T. G. and PROTTER, P. (1991a). Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19** 1035–1070. MR1112406

KURTZ, T. G. and PROTTER, P. (1991b). Wong–Zakai corrections, random evolutions, and simulation schemes for SDEs. In *Stochastic Analysis* 331–346. Academic Press, Boston, MA. MR1119837

KURTZ, T. G. and PROTTER, P. E. (1996). Weak convergence of stochastic integrals and differential equations. In *Probabilistic Models for Nonlinear Partial Differential Equations* (*Montecatini Terme*, 1995). *Lecture Notes in Math.* **1627** 1–41. Springer, Berlin. MR1431298

LAMBERTON, D. and PAGÈS, G. (2002). Recursive computation of the invariant distribution of a diffusion. *Bernoulli* **8** 367–405. MR1913112

LEMAIRE, V. (2005). Estimation récursive de la mesure invariante d'un processus de diffusion. Ph.D. thesis, Univ. de Marne-la-Vallée.

MAJDA, A. J., TIMOFEYEV, I. and VANDEN-EIJNDEN, E. (2001). A mathematical framework for stochastic climate models. *Comm. Pure Appl. Math.* **54** 891–974. MR1829529

MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes. III. Foster–Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25** 518–548. MR1234295

PARDOUX, E. and VERETENNIKOV, A. Y. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29** 1061–1085. MR1872736

PARDOUX, È. and VERETENNIKOV, A. Y. (2003). On Poisson equation and diffusion approximation. II. *Ann. Probab.* **31** 1166–1192. MR1988467

SUN, J. G. (1991). Perturbation bounds for the Cholesky and QR factorizations. *BIT* **31** 341–352. MR1112229

TALAY, D. (1990). Second-order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics and Stochastic Reports* **29** 13–36.

VERETENNIKOV, A. Y. (1997). On polynomial mixing bounds for stochastic differential equations. *Stochastic Process. Appl.* **70** 115–127. MR1472961

VERETENNIKOV, A. Y. (2011). On Sobolev solutions of Poisson equations in $\mathbb{R}^d$ with a parameter. *J. Math. Sci.* (*N.Y.*) **179** 48–79. MR3014098

UNIVERSITÉ NICE SOPHIA ANTIPOLIS
CNRS, LJAD, UMR 7351
PARC VALROSE
06108 NICE CEDEX 02
FRANCE
E-MAIL: camilo@unice.fr