Weak error rates for numerical schemes of non-singular Stochastic Volterra equations with application to stochastic volatility models

Pierre Bras*† and Masaaki Fukasawa‡

Abstract

We study the weak error rate for the Euler-Maruyama scheme for Stochastic Volterra equations (SVE) with application to pricing under stochastic volatility models. SVEs are non-Markovian stochastic differential equations with memory kernel. We assume in particular that the kernel is non-singular and $\mathcal{C}^4$. We show that the weak error rate is of order $O(1/N)$ where N is the number of steps of the Euler-Maruyama scheme, thus giving the same weak error rate as for SDEs. Our proof consists in adapting the classic weak error proof for Markov processes to SVEs; to this end we rely on infinite dimensional functionals and on their derivatives.

Keywords– Stochastic Volterra equation, Rough volatility, Euler-Maruyama scheme, Weak error rate.
MSC Classification– 65C30, 60H35

1 Introduction

Stochastic Volterra equations (SVE) have recently attracted much attention in the mathematical finance community in the context of rough volatility modelling, which is more able to reproduce some features of asset prices [4, 21, 12, 14, 6, 28, 16, 17, 18]. SVEs have also been introduced with regular (non-singular) kernel for modelling in population dynamics, biology and physics [31], in order to generalize modelling to non-Markovian stochastic systems with some memory effect. They were also motivated in particular by the physics of heat transfer [24] and have been mathematically studied since [8, 35]. Applications, e.g. pricing options in financial practice, require numerical methods to simulate the solution of the SVE, such as simulation through Euler-Maruyama schemes.

In the present paper, we give bounds for the weak error of the Euler-Maruyama scheme with N steps of an SVE on a finite time interval $[0, T]$ in the case where the kernel is non-singular. We consider two different Euler schemes: one where the kernel is not discretized, thus requiring the simulation of a large Gaussian matrix with covariance, and another one where the kernel is discretized, thus requiring only the simulation of (independent) Brownian increments.

A first bound on the weak error can be obtained from bounds on the strong error, however this is sub-optimal in general. For example, for Stochastic Differential Equations (SDE) and under general regularity assumptions the strong error is of order $O(1/\sqrt{N})$ but the weak error is of order $O(1/N)$. Such bounds get even worse in the case of SVE with fractional kernel, giving a weak error bounded by the strong error which is order $O(N^{-H})$, where $H \in (0, 1/2)$ is the Hurst parameter of the fractional kernel and is small ($H \simeq 0.1$, see for example [21, 18, 7, 9] for empirical numerical estimations of H) in many financial applications. In [36, 27], the authors prove strong error rates for rough and regular SVEs and show in particular that in the regular case the strong error rate is of order $O(1/\sqrt{N})$ as for SDEs, see section 2.2.

Moreover in [36] bounds are given for the weak error for the multi-level Euler-Maruyama scheme, however the authors only assume that the weak error is bounded by the strong error (see [36, Section 2.3]). In [19, 32] it is shown that the strong error is exactly of order H and the authors give the expression of the limit law of the (rescaled) strong error.

Motivated by rough volatility modeling, a series of recent papers [6, 5, 20, 15] focus on the analysis of the discretization error of $X_T = \int_0^T \sigma(\hat{W}_t^H)dB_t$ where $\hat{W}_t^H = \int_0^t (t-s)^{H-1/2}dW_s$, $H \in (0, 1/2)$, $B_t = \rho W_t + \sqrt{1-\rho^2}W_t^{\perp}$, W and $W^{\perp}$ are independent Brownian motions. [6] considers the case $\sigma(x) = x$ – which applies

*Sorbonne Université, Laboratoire de Probabilités, Statistique et Modélisation, UMR 8001, case 158, 4 pl. Jussieu, F-75252 Paris Cedex 5, France. E-mail: pierre.bras@sorbonne-universite.fr.
†Corresponding author.
‡Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka 560-0043, Japan. E-mail: fukasawa.m.es@osaka-u.ac.jp.

in particular to the rough Stein-Stein model [37] – and shows that the weak error rate is at least $H + 1/2$; the approach is based on an infinite dimensional Markov representation of X. In [5], the authors apply the duality method to this context to show that the weak error rate is at least $H + 1/2$ for $\sigma(x) = x$ and at least $2H$ under a general model. The same technique is sharpen in [20], where weak error rates are proved to be of order $(3H + 1/2) \wedge 1$ for $\sigma(x) = x$ or for general σ with cubic test function. More recently, [15] also proved that the weak error rate is of order $(3H + 1/2) \wedge 1$ for a reasonably large class of functions σ and polynomial test functions, including the rough Bergomi and the rough Stein-Stein models [37, 12]. The authors even obtain a rate of order 1 in the uncorrelated case $\rho = 0$. In these papers, the proof generally relies on successive Malliavin integration by parts with respect to the Brownian motion W to obtain a representation of the error as iterated integrals. In [10], the authors show a weak error rate of at least $H+1/2$ for smooth test functions. We recall that a (possibly rough) Volterra stochastic volatility model is a special case of a (possibly singular) two-dimensional SVE, where the first process is an asset price satisfying $dS_t = S_t\sqrt{V_t}dB_t$ for some Brownian motion B and where the second process (V_t) is the volatility satisfying some (possibly rough) Volterra stochastic equation, then giving a matrix kernel K for the joint process being diagonal and constant on its first coordinate. In [3], the authors propose a numerical scheme based on multifactor SDE approximation that applies to both regular and rough SVEs where the kernel is completely monotone, and give strong error bounds that depend on the choice of the kernel approximation.

Our main result states that for SVEs with non-singular kernel, the weak error of the Euler-Maruyama scheme is of order $O(1/N)$, which is the same rate as in the classic SDE case, under regularity assumptions on the coefficients and the kernel: we assume that the kernel is defined on the whole interval $[0, T]$ and is $\mathcal{C}^4$, that the drift and the diffusion coefficients are $\mathcal{C}^5$, bounded with bounded derivatives. Our strategy of proof consists in adapting the domino method from [38] to the SVE case. In the SDE case and for a function $f : \mathbb{R}^d \to \mathbb{R}$, the domino strategy consists in a step-by-step decomposition of the weak error to produce an upper bound as follows:

$$\begin{aligned}|\mathbb{E}f(\bar{X}^x_T) - \mathbb{E}f(X^x_T)| &= |\bar{P}_h \circ \cdots \circ \bar{P}_h f(x) - P_{t_N} f(x)| \\ &\leq \sum_{k=1}^{N} \left|\bar{P}_h \circ \cdots \circ \bar{P}_h \circ (\bar{P}_h - P_h) \circ P_{t_N - t_k} f(x)\right|, \end{aligned} \tag{1.1}$$

where $h = T/N$, $t_k = kh$, X is the solution to the SDE and $\bar{X}$ is the corresponding Euler-Maruyama scheme, P and $\bar{P}$ are the semi-group operators associated to X and $\bar{X}$ respectively. Then showing that with $g := P_{t_n - t_k} f$, the short term weak error $(\bar{P}_h - P_h)g(x)$ is of order $O(1/N^2)$, the sum in (1.1) is then of order $O(1/N)$. An elementary proof in the SDE case using the domino method can be found in [33, Section 7.6]. This strategy fundamentally relies on a Markov semi-group and thus cannot be directly applied to the SVE case, as the future trajectory of X depends on the whole previous trajectory in this last case. Instead, we consider the solution of an SVE as a Markov process on the infinite dimensional state space of trajectories Ω and then we define a infinite-dimensional Markov semi-group type operator, allowing us to use the previously introduced domino method.

We then show that the weak error in small time is of order $O(1/N^2)$ by establishing an Itô type formula for functionals $g : \Omega \to \mathbb{R}$ of SVE processes, involving the Fréchet derivatives of g. Such approach involving the derivatives of functionals on infinite dimensional state space and establishing Itô formula for SVEs was also developed in [13] and [39]. In [13] the author gives an Itô formula for $d_t g((X_s)_{s\in[0,t]})$ where X_t is a (Markov) semi-martingale; in [39] the Itô formula is given for $d_t g((\tilde{X}^t)_{s\in[0,T]})$ where $\tilde{X}^t$ is an SVE or rough process for $s \in [0, t]$ and completed by some given $\mathcal{F}_t$-measurable process on times greater than t. The formula we establish is a new contribution and is valid for $d_t g(\int_0^t Z(\cdot, s)dW_s)$ where $Z(u, \cdot)$ is an adapted semi-martingale for all u. Moreover we could not apply the result from [39] to our strategy of proof.

Using a hybrid approach, combining ideas from both finite and infinite dimensional settings, the Itô formula with a finite dimensional Brownian motion on the one side, and the Fréchet derivatives of path functionals and the Markov property on an infinite dimensional state space on the other side, we establish the weak convergence rate of the Euler-Maruyama for SVEs. Studying the non-singular case by adapting the classic domino method to path-dependent setting could be a first step for studying weak error rates for fractional SVEs. More specifically, the outline of the proof in this last case would be to follow the same path-wise domino strategy and to still rely on the new Itô formula in theorem 3.4. This would be achieved by using regular approximations of kernels or short-time scaling of the kernel. We refer to the proof of theorem 4.1 which is in fact the main proof of the paper, where we give bounds for the difference between the two semi-group type operators $\bar{P}_{h,t_k} - P_{h,t_k}$ in short time. However, we highlight here that the adaptation to the non-singular case is not straightforward as the singular is more difficult to tackle and generally requires more advanced techniques. At this stage we are not sure whether the same technique could be applied and further investigation is needed.

We give numerical evidence of the convergence rate we obtained on a Monte Carlo option pricing problem with a stochastic volatility model where the volatility follows some non-singular SVE. Proving weak error rates allows to design weighted and unweighted multi-level Richardson-Romberg extrapolation estimators [22, 23, 30] that exploit the faster convergence of the weak error in comparison with the strong error; such application is particularly critical for the Monte Carlo simulation of Volterra and rough stochastic equations where the Vanilla Euler-Maruyama scheme has large time complexity (N^2).

The article is organized as follows. In section 2 we give the precise setting and assumptions of the problem we consider, in particular the regularity assumptions on the coefficients and on the kernel of the SVE, we recall the existing results on strong error rates and we state our main theorem. We then give a list of examples of SVEs from the literature to which our main result is applicable, and we show how the computational cost of standard and multi-level estimators can be improved using weak error rates. In section 3, we give general results on random paths $(\varphi_u^t)_{u\geq 0, t\in[0,T]}$ which are adapted with respect to t, in particular we establish an Itô formula for infinite dimensional functionals $g : \Omega \to \mathbb{R}$. The proof of the theorem is given in section 4. Considering (1.1), the proof of our main result is decomposed into three parts: we prove that the short term error is of order $O(1/N^2)$, applying Itô formula for a regular functional $g : \Omega \to \mathbb{R}$ as in the classic proof in the SDE case. Secondly we show that with g being the concatenation of discrete kernels applied to f, then the functional g is indeed differentiable in the Fréchet meaning and with bounded Fréchet derivatives. Lastly, in section 5, we empirically check the weak convergence rate for some SVE model with non-singular kernel.

2 Setting and main results

2.1 Setting

Let us consider the following SVE in $\mathbb{R}^d$, $d \in \mathbb{N}$:

$$X_t = X_0 + \int_0^t K_1(t,s) b(X_s) ds + \int_0^t K_2(t,s)\sigma(X_s) dW_s, \quad t \in [0,T], \tag{2.1}$$

where (W_t) is a standard Brownian motion in $\mathbb{R}^{q_3}$ defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where

$$b : \mathbb{R}^d \to \mathbb{R}^{q_1}, \quad K_1 : [0,T]^2 \to \mathcal{M}_{d,q_1}(\mathbb{R}), \quad \sigma : \mathbb{R}^d \to \mathcal{M}_{q_2,q_3}(\mathbb{R}), \quad K_2 : [0,T]^2 \to \mathcal{M}_{d,q_2}(\mathbb{R}),$$

and where $q_1, q_2, q_3 \in \mathbb{N}$ and for $a, b \in \mathbb{N}$, $\mathcal{M}_{a,b}(\mathbb{R})$ denotes the set of $a \times b$ matrices with coefficients in $\mathbb{R}$. We denote by $(\mathcal{F}_t)_{t\in[0,T]}$ the filtration generated by the Brownian motion.

The starting point $X_0 \in \mathbb{R}^d$ is fixed. Let us make the following assumptions on the coefficients b and σ and on the kernels K_1 and K_2.

For $(\mathcal{A}, d_{\mathcal{A}})$ and $(\mathcal{B}, d_{\mathcal{B}})$ two metric spaces and $k \in \mathbb{N}$, we consider the following sets of functions from $\mathcal{A}$ to $\mathcal{B}$:

- $\mathcal{C}^k(\mathcal{A}, \mathcal{B})$: functions that are k times differentiable with continuous derivatives,
- $\mathcal{C}_b^k(\mathcal{A}, \mathcal{B})$: functions that are bounded, k times differentiable with continuous and bounded derivatives,
- $\widetilde{\mathcal{C}}_b^k(\mathcal{A}, \mathcal{B})$: functions that are k times differentiable with continuous and bounded derivatives.

When there is no ambiguity on the spaces, we also use the notations $\mathcal{C}^k$, $\mathcal{C}_b^k$ and $\widetilde{\mathcal{C}}_b^k$ respectively.

For f some function between metric spaces, if f is Lipschitz-continuous then we denote $[f]_{\text{Lip}}$ its Lipschitz constant.

Assumption 2.1.

(i) $K_1 \in \mathcal{C}^2([0,T]^2, \mathcal{M}_{d,q_1}(\mathbb{R}))$ *and* $K_2 \in \mathcal{C}^4([0,T]^2, \mathcal{M}_{d,q_2}(\mathbb{R}))$, *which guarantees that* K_1 *(resp.* K_2*) is bounded with bounded derivatives up to order 2 (resp. 4).*

(ii) $b \in \widetilde{\mathcal{C}}_b^5(\mathbb{R}^d, \mathbb{R}^{q_1})$ *and* $\sigma \in \mathcal{C}_b^5(\mathbb{R}^d, \mathcal{M}_{q_3,q_2}(\mathbb{R}))$.

Then theorem 2.1 guarantees that the solution of (2.1) is well defined (see for example [1, Lemma 5.29] and [27, Theorem 1.1]).

To simplify the notations and for more readability of the proofs, we assume hereafter that all the objects considered are one-dimensional, i.e. that $d = q_1 = q_2 = q_3 = 1$. However the main results in section 2.3 remain valid for any (finite) dimensions d, q_1, q_2 and q_3, re-writing the proofs by replacing the the one-dimensional products by matrix products and writing them as sums over indices.

Let us define the Euler-Maruyama scheme associated to (2.1). For $N \in \mathbb{N}$, we define the time step and the regular subdivision

$$h := T/N, \quad t_k := kT/N, \quad k \in \{0, \dots, N\} \tag{2.2}$$

and

$$\bar{X}_t = X_0 + \int_0^t K_1(t,s) b(\bar{X}_{\underline{s}}) ds + \int_0^t K_2(t,s) \sigma(\bar{X}_{\underline{s}}) dW_s, \quad t \in [0,T], \tag{2.3}$$

where for $s \in [0,T]$, we define

$$\underline{s} = \lfloor s/h \rfloor h.$$

The solution of (2.3) can be recursively simulated as

$$\bar{X}_t = X_0 + \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}\wedge t} K_1(t,s) b(\bar{X}_{t_j}) ds + \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}\wedge t} K_2(t,s) \sigma(\bar{X}_{t_j}) dW_s, \quad t \in [t_k, t_{k+1}],$$

where the integrals $(\int_{t_j}^{t_{j+1}} K_2(t,s) dW_s)_j$ can be simulated on the discrete grid $(t_k)_{0 \le k \le N}$ by generating the independent sequence of Gaussian vectors

$$\Big(\int_{t_j}^{t_{j+1}} K_2(t_k, s) dW_s \Big)_{k=j,\dots,N}, \quad j = 0, \dots, N-1,$$

using the Cholesky decomposition of the covariance matrix

$$\Big(\int_{t_j}^{t_{j+1}} K_2(t_{k_1}, s) K_2(t_{k_2}, s) ds \Big)_{k_1,k_2=j,\dots,N}.$$

We also define the Euler-Maruyama scheme associated to (2.1) with discretization of the kernels as

$$\overrightarrow{X}_t = X_0 + \int_0^t K_1(t,\underline{s}) b(\overrightarrow{X}_{\underline{s}}) ds + \int_0^t K_2(t,\underline{s}) \sigma(\overrightarrow{X}_{\underline{s}}) dW_s, \quad t \in [0,T]. \tag{2.4}$$

This scheme is more convenient to simulate as it only requires the simulation of the Brownian increments $(W_{t_{k+1}} - W_{t_k})_{0 \le k \le N-1}$.

With no ambiguity, in the proofs we shall use the notation $\bar{\psi}$ for ψ some function defined on $\mathbb{R}^d$, such that for every process Y and every $s \in [0,T]$ we have $\bar{\psi}(Y_s) = \psi(Y_{\underline{s}})$.

We extend the definition of K_1 and K_2 on $\mathbb{R}^+ \times \mathbb{R}^+$ with $\mathbb{R}^+ := [0,\infty)$, such that for $i = 1,2$, $K_i(t,s) = 0$ for $(t,s) \notin [0,2T] \times [0,2T]$ and such that K_i is still bounded with bounded derivatives up to order 2.

In this paper, we use the notation C to denote a positive real constant, which may change from line to line. The constant C depends on the parameters of the problem: the coefficients and the kernels of the SVE, the time horizon T.

Let us first prove a bound on the moments.

Lemma 2.2. *Let X, $\bar{X}$, $\overrightarrow{X}$ be the solution of (2.1), (2.3) and (2.4) respectively; under theorem 2.1 we have*

$$\sup_{t\in[0,T]} \mathbb{E}|X_t|^2 + \sup_{t\in[0,T]} \mathbb{E}|\bar{X}_t|^2 + \sup_{t\in[0,T]} \mathbb{E}|\overrightarrow{X}_t|^2 < +\infty. \tag{2.5}$$

Proof. For every $t \in [0,T]$, from the definition of X and using $(\alpha+\beta+\gamma)^2 \le 3(\alpha^2+\beta^2+\gamma^2)$ and $(\alpha+\beta)^2 \le 2(\alpha^2+\beta^2)$, we get

$$\begin{aligned}
\sup_{s\in[0,t]} \mathbb{E}|X_s|^2 &\le 3|X_0|^2 + 3\|K_2\|_\infty^2 \|\sigma\|_\infty^2 T + 3\|K_1\|_\infty^2 \sup_{s\in[0,t]} \mathbb{E}\Big| \int_0^s |b(X_u)| du \Big|^2 \\
&\le C + C\Big(2T^2 |b(0)|^2 + 2[b]_{\mathrm{Lip}}^2 \mathbb{E}\Big| \int_0^t |X_u| du \Big|^2 \Big) \\
&\le C\big(1 + \int_0^t \mathbb{E}|X_u|^2 du\big) \le C\big(1 + \int_0^t \sup_{u\in[0,s]} \mathbb{E}|X_u|^2 du\big),
\end{aligned}$$

with similar inequalities for $\bar{X}$ and $\overrightarrow{X}$. The result follows from the Grönwall Lemma. □

2.2 Strong error bounds [36]

Let us first recall strong error results for regular SVEs.

Theorem 2.3 (Adapted from [36], Theorem 2.2 and [27], Theorems 1.2 and 1.4)**.** *Assume that b and σ are Lipschitz-continuous and $K_1, K_2 \in \mathcal{C}^1([0,T]^2)$. Then for $p \geq 1$,*

$$\mathbb{E}\Big[\big|\bar{X}_T - X_T\big|^p\Big] + \mathbb{E}\Big[\big|\overrightarrow{X}_T - X_T\big|^p\Big] \leq C_p N^{-p/2}. \tag{2.6}$$

Remark 2.4

- The results from [36, 27] are in fact more general and may also be applied to non regular settings. The latter [27] requires fewer assumptions than [36] and gives results for $\bar{X}$, so-called K-discrete Euler scheme, whereas [36] only gives bounds for $\overrightarrow{X}$. However for concision we simplify the assumptions and we adapt these results to our "regular" setting, assuming in particular that the kernels are differentiable.
- We can deduce a weak error bound from the strong error bound (2.6): if $f : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz-continuous, then

$$\Big|\mathbb{E}[f(\overrightarrow{X}_T)] - \mathbb{E}[f(X_T)]\Big| \leq \mathbb{E}\Big|f(\overrightarrow{X}_T) - f(X_T)\Big| \leq [f]_{\text{Lip}}\mathbb{E}\Big[\big|\overrightarrow{X}_T - X_T\big|\Big] = O(N^{-1/2}), \tag{2.7}$$

  however this result is sub-optimal.

2.3 Main results

Theorem 2.5. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be $\widetilde{\mathcal{C}}_b^5$ and assume theorem 2.1. Then we have*

$$\mathbb{E}[f(\bar{X}_T)] - \mathbb{E}[f(X_T)] = O\left(1/N\right), \tag{2.8}$$

$$\mathbb{E}[f(\overrightarrow{X}_T)] - \mathbb{E}[f(X_T)] = O\left(1/N\right). \tag{2.9}$$

Remark 2.6

- Following theorem 2.2, we have

$$\mathbb{E}|X_T| + \mathbb{E}|\bar{X}_T| + \mathbb{E}|\overrightarrow{X}_T| < +\infty.$$

  Since f is Lipschitz-continuous, we get

$$\mathbb{E}|f(X_T)| + \mathbb{E}|f(\bar{X}_T)| + \mathbb{E}|f(\overrightarrow{X}_T)| < +\infty.$$

- The strategy of proof we develop in section 4 does not allow us to give weak error bounds for path-dependent functionals.
- We prove that the weak order of convergence of the Euler-Maruyama scheme for SVEs with regular kernels is the same as for SDEs, however the computation time for this scheme for SVEs is of order N^2, against order N for SDEs.
- The proofs for (2.8) and (2.9) are largely similar; in the regular setting and provided that K_i and $\partial_2 K_i$ are bounded, $i = 1, 2$, then the numerical error coming from the discretization of the kernels is at most of same order of the error coming discretization of b and σ, see section 4.5.

Remark 2.7 (Discussion on theorem 2.1) The classic proof in the SDE case [38] requires the assumption $b, \sigma \in \widetilde{\mathcal{C}}_b^4$ (also see [33, Section 7.6]), while we require $b \in \widetilde{\mathcal{C}}_b^5$ and $\sigma \in \mathcal{C}_b^5$. This is mainly because of the technical assumptions in theorem 3.2 and the Itô formula in theorem 3.4. In the latter, we assume that the functional $G \in \widetilde{\mathcal{C}}_b^3$ instead of $G \in \widetilde{\mathcal{C}}_b^2$, whereas $\nabla^3 G$ does not appear in the final formula (3.6), thus we require b and σ to have bounded derivatives up to order 5 instead of 4 in order to obtain theorem 4.2. In the former, we require in particular $\sup_{u,s} |F(u,s)| \leq C$ almost surely, for some fixed $C > 0$, and thus we need σ to be bounded in order to get a bound on $\mathbb{E}\|P_{r,t}(\omega)\|_\infty$.

2.4 Applications

2.4.1 Examples of SVEs satisfying theorem 2.1

We give some examples of SVEs satisfying theorem 2.1, hence to which our main result 2.5 is applicable.

- Fractional Ornstein-Uhlenbeck processes [11] are defined by the equation:

$$X_t = X_0 - \lambda \int_0^t (X_s - \theta)ds + \sigma W_t^H,$$

where $\theta,\ \lambda,\ \sigma \in [0,\infty)$ and W^H is a fractional Brownian motion with index $H \in (0,1]$. Such processes admit the following Volterra version with regular kernel:

$$X_t = X_0 - \lambda \int_0^t (X_s - \theta)ds + \sigma \int_0^t K_2(t,s)dW_s$$

for any K_2 satisfying theorem 2.1. This is a particular case of affine Volterra processes [26], where we consider σ being constant.

- The Gaussian Stein-Stein model, introduced in [2] as a generalization of the classic Stein-Stein model [37], reads:

$$\begin{cases} dS_t = S_t V_t dB_t, \quad S_0 > 0, \\ V_t = V_0 + g_0(t) + \kappa \int_0^t K(t-s)V_s ds + \nu \int_0^t K(t-s)dW_s \end{cases}$$

where B and W are correlated Brownian motions with correlation $\rho \in [-1,1]$, $\kappa,\ \nu \in [0,\infty)$, the function $g_0 : [0,T] \to \mathbb{R}$ is deterministic and continuous and K is a measurable kernel. In this model, S describes some asset price and V its square volatility and if K and g_0 are regular i.e. $K \in \mathcal{C}^4$ and $g_0 \in \mathcal{C}^2$, then the above equation on V falls under our setting. We conduct a further numerical analysis of the weak convergence rate for this equation in section 5.

- Neural SVEs, introduced in [34], are used to approximate SVEs with neural networks. More precisely they can be written as

$$X_t = X_0 g_\theta(t) + \int_0^t K_{1,\theta}(t-s)b_\theta(Z_s)ds + \int_0^t K_{2,\theta}(t-s)\sigma_\theta(Z_s)dW_s$$

where g_θ, $K_{i,\theta}$, $i = 1,2$, σ_θ are neural functions parametrized by the multi-dimensional parameter θ. More specifically, let us assume that b and σ are compositions of the sigmoid activation function $\mathrm{sig}(x) := (1 + e^{-x})^{-1}$ and the softplus activation function $\mathrm{sp}(x) := \log(1+e^x)$ – which is used as a smooth approximation of the ReLU function – and affine functions in x parametrized by θ, and furthermore let us assume that for σ the last function in the composition is the sigmoid function. Then b has bounded successive derivatives and σ is bounded with bounded successive derivatives, thus satisfying theorem 2.1. This is because we have $\mathrm{sig}'(x) = \mathrm{sig}(x)(1-\mathrm{sig}(x))$, $\mathrm{sp}'(x) = 1-\mathrm{sig}(x)$, so that successive derivatives of sigmoid and softplus functions can be written as polynomials in the sigmoid function which is bounded with exponentially decaying derivatives, but we do not give an extensive analysis here.

2.4.2 Error analysis, multi-level Monte Carlo

Given $\varepsilon > 0$, we analyze the computational cost of a Monte Carlo estimator to achieve an error of order $O(\varepsilon)$ on the estimation of $\mathbb{E}f(X_T)$. If M is the number of Monte Carlo samples and N is the number of steps, then the cost of $\overrightarrow{X}$ is MN^2. To control the statistical error we need to set $M = O(\varepsilon^{-2})$ and to control the bias discretization error we need to set $N = O(\varepsilon^{-1})$ according to theorem 2.5, which is summarized in the following proposition.

Proposition 2.8. *Let* $\mathrm{Cost}_1(\varepsilon)$ *be the cost of the Monte Carlo estimator given by simulations of* $f(\overrightarrow{X}_T)$ *to achieve an error* ε. *Then*

$$\mathrm{Cost}_1(\varepsilon) = O(\varepsilon^{-4}).$$

We remark that using the weak error bound (2.9) instead of (2.7), we can improve the cost from $O(\varepsilon^{-6})$ in [36, Proposition 2.9] to $O(\varepsilon^{-4})$.

The MLMC method. The multi-level Monte Carlo (MLMC) method [22, 30], first introduced for SDEs, can also be used for SVEs; it consists in correcting the bias with successive layers of refined estimators and relies on strong and weak error bounds. It can be written as

$$\frac{1}{M_0}\sum_{i=1}^{M_0} f(\overrightarrow{X}_T^{N_0,0,i}) + \sum_{\ell=1}^{R}\frac{1}{M_\ell}\sum_{i=1}^{M_\ell}\left(f(\overrightarrow{X}_T^{N_\ell,\ell,i}) - f(\overrightarrow{X}_T^{N_{\ell-1},\ell,i})\right), \tag{2.10}$$

where (M_ℓ) are the sizes of the Monte Carlo simulations, $R \in \mathbb{N}$ is the number of refining layers, $(\overrightarrow{X}^{N_\ell,\ell,i})_{\ell,i}$ are independent simulations of $\overrightarrow{X}_T$ with N_ℓ steps with $N_\ell = 2^{-\ell}h_0$ and $\overrightarrow{X}^{N_{k_1},\ell_1,i}$ and $\overrightarrow{X}^{N_{k_2},\ell_2,j}$ are built on the same Brownian motion if $\ell_1 = \ell_2$ and $k_1 = k_2$, and independent otherwise. Following the proof of [36, Proposition 2.11] while improving the bound on the bias terms $\mathbb{E}f(\overrightarrow{X}^{N_\ell,\ell,i_T}) - \mathbb{E}f(X_T)$ from $N_\ell^{-1/2}$ to N_ℓ^{-1}, we improve the cost of the MLMC estimator (2.10) from ε^{-4} to ε^{-3}. Actually, following the optimization procedure of the parameters R, (M_ℓ), h_0 described in [30, Theorem 3.11], we obtain that the cost of the optimal MLMC estimator for level error ε is

$$\text{Cost}_2(\varepsilon) = O(\varepsilon^{-2}\log(1/\varepsilon)^{-2}).$$

We remark the significant benefits of multi-level methods to simulation settings where the bias is computationally expensive to reduce; in this case, the benefits are greater than when applied to SDEs. Moreover, if we assume or can obtain a high-order expansion of the weak error (2.9), then we could apply the weighted MLMC method [30] – also called multi-level Richardson-Romberg (ML2R) – to the simulation of SVEs with even higher gains.

3 Preliminary results on infinite dimensional paths

3.1 State space and path derivatives

For $T' \in \mathbb{R}^+$, we consider the infinite dimensional state space $\Omega_{T'}$ being the space of $\mathbb{R}$-valued continuous trajectories on $[0,\infty)$ with support included in $[0,T']$, with the topology of the supremum norm. For $\omega \in \Omega_{T'}$ such that ω is $\mathcal{C}^1$, we denote $\dot\omega$ its derivative. If ω is Lipschitz-continuous, we denote $[\omega]_{\text{Lip}}$ its Lipschitz constant. We denote by $\widetilde{0}$ the path on $\mathbb{R}^+$ constant to 0.

For $g : \Omega_{T'} \to \mathbb{R}$ and for $\omega \in \Omega_{T'}$, we define, when it exists, $\nabla g(\omega)$ as the Fréchet derivative of g with respect to ω, which is a linear operator on $\Omega_{T'}$:

$$g(\omega+\eta) = g(\omega) + \langle \nabla g(\omega), \eta\rangle + o(\|\eta\|_\infty), \quad \eta \in \Omega_{T'}.$$

More generally, for $\ell \in \mathbb{N}$ we define, when it exists, the derivative of g of order ℓ recursively as the ℓ-multilinear operator on $\Omega_{T'}^{\otimes \ell}$:

$$\begin{aligned}&\langle \nabla^{\ell-1} g(\omega+\eta^1), \bigotimes_{j=2}^{\ell}\eta^j\rangle = \langle \nabla^{\ell-1} g(\omega), \bigotimes_{j=2}^{\ell}\eta^j\rangle + \langle \nabla^{\ell} g(\omega), \bigotimes_{j=1}^{\ell}\eta^j\rangle + o(\|\eta^1\|_\infty),\\ &\eta^i \in \Omega_{T'},\ i = 1,\dots,\ell.\end{aligned}$$

We use the notation $\otimes$ only to enhance the multilinearity of $\nabla^\ell g$.

Remark 3.1 The path derivative can be made explicit in some simple cases:

- If $g(\omega) = \widetilde{g}(\omega_{u_0})$ for some fixed $u_0 \in \mathbb{R}^+$ and for some $\widetilde{g} : \mathbb{R} \to \mathbb{R}$, then we have

$$\langle \nabla^\ell g(\omega), \bigotimes_{j=1}^{\ell}\eta^j\rangle = \nabla^\ell \widetilde{g}(\omega_{u_0}) \cdot \bigotimes_{j=1}^{\ell}\eta^j_{u_0}. \tag{3.1}$$

- If $g(\omega) = \int_0^{T'} \widetilde{g}(\omega_u)du$ for some $\widetilde{g} : \mathbb{R} \to \mathbb{R}$, then we have

$$\langle \nabla^\ell g(\omega), \bigotimes_{j=1}^{\ell}\eta^j\rangle = \int_0^{T'} \nabla^\ell \widetilde{g}(\omega_u) \cdot \bigotimes_{j=1}^{\ell}\eta^j_u du. \tag{3.2}$$

3.2 Expectation of the supremum of a random path process

Lemma 3.2. *Let $F(u,s)_{u\geq 0, s\in[0,T]}$ be a $\mathbb{R}$-valued random process adapted to the filtration $\mathcal{F}$ with respect to its second variable, such that for every $s \in [0,T]$ and $u > T'$, $F(u,s) = 0$, and such that $\sup_{u,s}|F(u,s)| \leq C$ almost surely, $\partial_1 F$ exists and $\sup_{u,s}|\partial_1 F(u,s)| \leq C_1$ almost surely, for some C, $C_1 \in \mathbb{R}^+$, and let $(M_s)_{0\leq s\leq T}$ be a $\mathbb{R}$-valued martingale adapted to $\mathcal{F}$ with $\mathbb{E}\langle M\rangle_T < \infty$. For $r \in [0,T]$ let us define*

$$\varphi_u := \int_0^r F(u,s)dM_s, \quad u \geq 0.$$

Then there exists a continuous modification $\widetilde{\varphi}$ of φ and a constant $C(T')$ depending on T' such that

$$\mathbb{E}\sup_{u\geq 0}|\widetilde{\varphi}_u|^2 \leq C(T')^2 C_1^2 \mathbb{E}\langle M\rangle_r. \tag{3.3}$$

Proof. For $u_1, u_2 \in [0, T']$ we have

$$\mathbb{E}|\varphi_{u_1} - \varphi_{u_2}|^2 = \mathbb{E}\Big|\int_0^r (F(u_1,s) - F(u_2,s))dM_s\Big|^2 = \mathbb{E}\int_0^r |F(u_1,s) - F(u_2,s)|^2 d\langle M\rangle_s$$
$$\leq |u_1 - u_2|^2 C_1^2 \mathbb{E}\langle M\rangle_r,$$

so that using the Kolmogorov continuity theorem (A.1), there exists a modification $\widetilde{\varphi}$ of φ which is almost surely α-Hölder for every $\alpha \in (0, 1/2)$, and taking for example $\alpha = 1/4$ we have

$$\mathbb{E}\Big[\Big(\sup_{u_1,u_2\in[0,T'],u_1\neq u_2} \frac{|\widetilde{\varphi}_{u_1} - \widetilde{\varphi}_{u_2}|}{|u_1 - u_2|^{1/4}}\Big)^2\Big] \leq C(T')^{3/2} C_1^2 \mathbb{E}\langle M\rangle_r,$$

where C is an universal constant, so that taking $u_1 = u$ and $u_2 = 0$ with $\varphi_0 = 0$ we obtain

$$\mathbb{E}\sup_{u\geq 0}\big|\widetilde{\varphi}_u\big|^2 = \mathbb{E}\sup_{u\in[0,T']}\big|\widetilde{\varphi}_u\big|^2 \leq C(T')^2 C_1^2 \mathbb{E}\langle M\rangle_r.$$

□

Remark 3.3 In the following, we will use theorem 3.2 for families of trajectories $(\varphi^t)_{t\in[0,T]}$ of the form

$$\varphi_u^t = \int_0^t F(u,s)dM_s \quad \text{or} \quad \varphi_u^t = \int_0^t F(t+u,s)dM_s.$$

When the assumptions of theorem 3.2 are checked, the bound (3.3) is true up to some modification of φ, i.e. for a family of trajectories $(\widetilde{\varphi}_u^t)$ such that

$$\forall t \in [0,T],\ \forall u \geq 0,\ \mathbb{P}(\varphi_u^t = \widetilde{\varphi}_u^t) = 1 \quad \text{and} \quad \forall t \in [0,T],\ u \mapsto \widetilde{\varphi}_u^t \text{ is continuous.}$$

Without loss of generality, each time we use theorem 3.2 we do not mention explicitly the modification.

3.3 A general Itô formula for path-dependent functionals

In this section we prove an extension of the classic Itô formula to processes of the form $G(\varphi^t)$, where $G : \Omega_{T'} \to \mathbb{R}$ and where for every $t \in [0,T]$, φ^t is some $\mathcal{F}_t$-measurable random path.

Theorem 3.4. *Let us consider the family of random paths $(\varphi_u^t)_{t\in[0,T],\ u\geq 0}$ such that*

$$\varphi_u^t = \varphi_u^0 + \int_0^t Z_1(u,s)ds + \int_0^t Z_2(u,s)dW_s, \tag{3.4}$$

where for every $u \geq 0$, $s \mapsto Z_i(u,s)$, $i = 1,2$, is an adapted $\mathbb{R}$-valued semi-martingale such that $\partial_1 Z_i$ and $\partial_{11}^2 Z_2$ exist almost surely and there exists some $C \geq 0$ such that

$$\sup_{s\in[0,T]} \mathbb{E}\sup_{u\geq 0}|Z_1(u,s)|^2 + \sup_{s\in[0,T]} \mathbb{E}\sup_{u\geq 0}|\partial_1 Z_1(u,s)| \leq C,$$
$$\|Z_2\|_\infty + \|\partial_1 Z_2\|_\infty + \|\partial_{11}^2 Z_2\|_\infty \leq C \quad \textit{almost surely,} \tag{3.5}$$

and such that $Z_i(u,s) = 0$ for $u > T'$, $T' \in \mathbb{R}^+$. We also assume that $\varphi^0 \in \Omega_{T'} \cap \mathcal{C}^1$ and is Lipschitz-continuous. Moreover, let $G : \Omega_{T'} \to \mathbb{R}$ with bounded pathwise derivatives up to order 3. Then we have almost surely

$$G(\varphi^t) = G(\varphi^0) + \int_0^t \langle \nabla G(\varphi^s), Z_1(\cdot,s)\rangle ds + \int_0^t \langle \nabla G(\varphi^s), Z_2(\cdot,s)\rangle dW_s$$
$$+ \frac{1}{2}\int_0^t \langle \nabla^2 G(\varphi^s), Z_2(\cdot,s)^{\otimes 2}\rangle ds. \tag{3.6}$$

Remark 3.5 We highlight the fact that in (3.4), the values of $Z_i(u,s)$ cannot depend on t. For example, if we consider the SVE

$$\varphi_u^t = \varphi_u^0 + \int_0^t K_2(t+u,s)\sigma(A_s)dW_s$$

for some adapted semi-martingale A, then we need to write φ_u^t as

$$\varphi_u^t = \varphi_u^0 + \int_0^t K_2(s+u,s)\sigma(A_s)dW_s + \int_0^t \Big(\int_0^s \partial_1 K_2(s+u,v)\sigma(A_v)dW_v\Big)ds.$$

Proof. We first remark that if G only depends on a finite number of times $u_1, \dots, u_n \in \mathbb{R}^+$, i.e. if we have

$$\forall \omega \in \Omega_{T'},\ G(\omega) = \widetilde{G}(\omega_{u_1}, \dots, \omega_{u_n}), \quad \widetilde{G} \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}),$$

then we have

$$\begin{aligned}
&\forall \eta \in \Omega,\ \langle \nabla G(\omega), \eta \rangle = \sum_{i=1}^n \partial_i \widetilde{G}(\omega_{u_1}, \dots, \omega_{u_n}) \eta_{u_i}, \\
&\forall \eta^1, \eta^2 \in \Omega,\ \langle \nabla^2 G(\omega), \eta^1 \otimes \eta^2 \rangle = \sum_{1 \le i,j \le n} \partial_{ij} \widetilde{G}(\omega_{u_1}, \dots, \omega_{u_n}) \eta^1_{u_i} \eta^2_{u_j},
\end{aligned}$$

and then (3.6) directly comes from the classic Itô formula.

Then, let us define for $n \in \mathbb{N}$ the regular subdivision of $[0, T']$: $u_i^n = iT'/n$, $0 \le i \le n$ and for $\omega \in \Omega_{T'}$ we define ω^n as the affine interpolation of ω on the subdivision $(u_i^n)_i$ i.e. ω^n is equal to the affine interpolation on $[0, T']$ and then ω^n and ω are both equal to 0 on $[T', \infty)$; we also define G^n as for every $\omega \in \Omega_{T'}$, $G^n(\omega) = G(\omega^n)$. Then for every $\omega \in \Omega_{2T}$ we have $G^n(\omega) = \widetilde{G}^n(\omega_{u_1}, \dots, \omega_{u_n})$ where $\widetilde{G}^n : \mathbb{R}^n \to \mathbb{R}$ is the composition of the affine interpolation $\mathscr{L}^n : \mathbb{R}^n \to \Omega_{2T}$, which is a bounded linear operator, and of G, so is $\mathcal{C}^2$ and then (3.6) is true for G^n. Moreover for every $\omega, \eta \in \Omega_{T'}$ such that ω is Lipschitz-continuous we have

$$|G(\omega) - G^n(\omega)| = |G(\omega) - G(\omega^n)| \le \|\nabla G\|_\infty \|\omega - \omega^n\|_\infty \le C \|\nabla G\|_\infty [\omega]_{\mathrm{Lip}}/n.$$

Moreover, remarking that the affine interpolation is a bounded linear operator, we get that G^n is also differentiable with

$$\langle \nabla G^n(\omega), \eta \rangle = \langle \nabla G(\omega^n), \eta^n \rangle$$

so that

$$\big| \langle \nabla G(\omega), \eta \rangle - \langle \nabla G^n(\omega), \eta \rangle \big| \le \|\nabla^2 G\|_\infty \|\omega - \omega^n\|_\infty \|\eta\|_\infty + \|\nabla G\|_\infty \|\omega\|_\infty \|\eta - \eta^n\|_\infty. \tag{3.7}$$

Moreover for $\eta^1, \eta^2 \in \Omega_{T'}$ we have

$$\langle \nabla^2 G^n(\omega), \eta^1 \otimes \eta^2 \rangle = \langle \nabla^2 G(\omega^n), (\eta^1)^n \otimes (\eta^2)^n \rangle$$

so that

$$\begin{aligned}
\big| \langle \nabla^2 G(\omega), \eta^1 \otimes \eta^2 \rangle - \langle \nabla^2 G^n(\omega), \eta^1 \otimes \eta^2 \rangle \big| &\le \|\nabla^3 G\|_\infty \|\omega - \omega^n\|_\infty \|\eta^1\|_\infty \|\eta^2\|_\infty \\
&\quad + \|\nabla^2 G\|_\infty \big(\|\omega\|_\infty \|\eta^1 - (\eta^1)^n\|_\infty \|\eta^2\|_\infty + \|\omega\|_\infty \|\eta^1\|_\infty \|\eta^2 - (\eta^2)^n\|_\infty \big).
\end{aligned} \tag{3.8}$$

Writing (3.6) with G^n gives

$$\begin{aligned}
G^n(\varphi^t) = G^n(\varphi^0) &+ \int_0^t \langle \nabla G((\varphi^s)^n), Z_1^n(\cdot, s) \rangle ds + \int_0^t \langle \nabla G((\varphi^s)^n), Z_2^n(\cdot, s) \rangle dW_s \\
&+ \frac{1}{2} \int_0^t \langle \nabla^2 G((\varphi^s)^n), Z_2^n(\cdot, s)^{\otimes 2} \rangle ds,
\end{aligned} \tag{3.9}$$

and we have

$$\mathbb{E}|G^n(\varphi^t) - G(\varphi^t)|^2 \le 2\|G\|_\infty \mathbb{E}|G^n(\varphi^t) - G(\varphi^t)| \le C \|G\|_\infty \|\nabla G\|_\infty \mathbb{E}[\varphi^t]_{\mathrm{Lip}} n^{-1}$$

with φ^t being $\mathcal{C}^1$ with

$$\dot{\varphi}_u^t = \dot{\varphi}_0^t + \int_0^t \partial_1 Z_1(u, s) ds + \int_0^t \partial_1 Z_2(u, s) dW_s,$$

where the interchange is ensured by the stochastic Fubini theorem. But following theorem 3.2 with assumption (3.5), $\mathbb{E}[\varphi^t]_{\mathrm{Lip}} < \infty$, so that $G^n(\varphi^t)$ converges to $G(\varphi^t)$ in L^2. We proceed the same way for $G^n(\varphi^0)$.

Moreover, we have

$$\begin{aligned}
&\mathbb{E}\Big| \int_0^t \langle \nabla G((\varphi^s)^n), Z_1^n(\cdot, s) \rangle ds - \int_0^t \langle \nabla G(\varphi^s), Z_1(\cdot, s) \rangle ds \Big| \\
&\quad \le \int_0^t \mathbb{E}|\langle \nabla G(\varphi^s), Z_1^n(\cdot, s) - Z_1(\cdot, s) \rangle| ds + \int_0^t \mathbb{E}|\langle \nabla G((\varphi^s)^n) - \nabla G(\varphi^s), Z_1^n(\cdot, s) \rangle| ds
\end{aligned}$$

$$
\begin{aligned}
&\le \|\nabla G\|_\infty \int_0^t \mathbb{E}\sup_{u\ge 0}|\partial_1 Z_1(u,s)|ds\ n^{-1} + \|\nabla^2 G\|_\infty \int_0^t \mathbb{E}\big[[\varphi^s]_{\mathrm{Lip}}\sup_{u\ge 0}|Z_1^n(u,s)|\big]ds\ n^{-1}\\
&\le T\|\nabla G\|_\infty \sup_{s\in[0,T]}\mathbb{E}\sup_{u\ge 0}|\partial_1 Z_1(u,s)|n^{-1}\\
&\quad + \|\nabla^2 G\|_\infty \Big(\int_0^t \mathbb{E}[\varphi^s]_{\mathrm{Lip}}^2 ds\Big)^{1/2}\Big(\int_0^t \mathbb{E}\sup_{u\ge 0}|Z_1(u,s)|^2 ds\Big)^{1/2} n^{-1}\\
&\le T\|\nabla G\|_\infty \sup_{s\in[0,T]}\mathbb{E}\sup_{u\ge 0}|\partial_1 Z_1(u,s)|n^{-1}\\
&\quad + T^{1/2}\|\nabla^2 G\|_\infty \Big(\int_0^t \mathbb{E}[\varphi^s]_{\mathrm{Lip}}^2 ds\Big)^{1/2}\big(\sup_{s\in[0,T]}\mathbb{E}\sup_{u\ge 0}|Z_1(u,s)|^2\big)^{1/2} \underset{n\to\infty}{\longrightarrow} 0,
\end{aligned}
$$

where we used (3.7) and that $\mathbb{E}[\varphi^s]_{\mathrm{Lip}}^2 \le C$ with theorem 3.2, where C does not depend on s.

Furthermore, following (3.7) we have

$$
\begin{aligned}
&\mathbb{E}\Big|\int_0^t \langle \nabla G((\varphi^s)^n), Z_2^n(\cdot,s)\rangle dW_s - \int_0^t \langle \nabla G(\varphi^s), Z_2(\cdot,s)\rangle dW_s\Big|^2\\
&= \int_0^t \mathbb{E}\big|\langle \nabla G((\varphi^s)^n), Z_2^n(\cdot,s)\rangle - \langle \nabla G(\varphi^s), Z_2(\cdot,s)\rangle\big|^2 ds\\
&\le 2\int_0^t \mathbb{E}\big|\|\nabla G\|_\infty\|Z_2\|_\infty\big(\langle \nabla G((\varphi^s)^n), Z_2^n(\cdot,s)\rangle - \langle \nabla G(\varphi^s), Z_2(\cdot,s)\rangle\big)\big|ds\\
&\le 2C\int_0^t \mathbb{E}\Big[\|\nabla G\|_\infty\|Z_2\|_\infty\Big(\|\nabla^2 G\|_\infty n^{-1}\|Z_2\|_\infty[\varphi^s]_{\mathrm{Lip}} + \|\nabla G\|_\infty n^{-1}\|\partial_1 Z_2\|_\infty\|\varphi^s\|_\infty\Big)\Big]ds\\
&\le Cn^{-1}\int_0^t \big(\mathbb{E}[\varphi^s]_{\mathrm{Lip}} + \mathbb{E}\|\varphi^s\|_\infty\big)ds
\end{aligned}
$$

and using theorem 3.2, we have $\mathbb{E}\|\varphi^s\|_\infty \le C$ and $\mathbb{E}[\varphi^s]_{\mathrm{Lip}} \le C$ where C does not depend on s, so that the above quantity converges to 0 as $n\to\infty$.

Last, using (3.8) we get

$$
\begin{aligned}
&\mathbb{E}\Big|\int_0^t \langle \nabla^2 G((\varphi^s)^n), Z_2^n(\cdot,s)^{\otimes 2}\rangle ds - \int_0^t \langle \nabla^2 G(\varphi^s), Z_2(\cdot,s)^{\otimes 2}\rangle ds\Big|\\
&\le \int_0^t \mathbb{E}\big(C\|\nabla^3 G\|_\infty n^{-1}\|Z_2\|_\infty^2[\varphi^s]_{\mathrm{Lip}} ds + C\|\nabla^2 G\|_\infty\|Z_2\|_\infty\|\partial_1 Z_2\|_\infty n^{-1}\|\varphi^s\|_\infty\big)ds\\
&\le Cn^{-1}\int_0^t \mathbb{E}[\varphi^s]_{\mathrm{Lip}} ds + Cn^{-1}\int_0^t \mathbb{E}\|\varphi^s\|_\infty ds \underset{n\to\infty}{\longrightarrow} 0.
\end{aligned}
$$

□

4 Proof of theorem 2.5

In this section, we only give the full proof for (2.8). The proof of (2.9) is similar, as explained in section 4.5.

4.1 Definition of the infinite dimensional semi-group and domino strategy

We define the infinite dimensional semi-group type operator that we use for the domino strategy. We do not apply the domino strategy on X directly; instead we define an auxiliary process Y such that X can be induced from Y, as follows. Let us consider the following family of processes:

$$
Y_t(u) = \int_0^t K_1(t+u,s)b(X_s)ds + \int_0^t K_2(t+u,s)\sigma(X_s)dW_s, \quad u\ge 0,\ t\in[0,T]. \tag{4.1}
$$

Following theorem 3.2, for every $u\ge 0$ and $t\in[0,T]$, $Y_t(u)$ is well defined and $Y_t: u\mapsto Y_t(u)$ is continuous. Moreover, for every $t\in[0,T]$, Y_t is $\mathcal{F}_t$-measurable (but the process $(Y_t(u))_{u\ge 0}$ is not adapted w.r.t. u) and writing

$$
Y_t(u) = \int_0^t \Big(\int_0^u \partial_1 K_1(t+v,s)dv + K_1(t,s)\Big)b(X_s)ds
$$

$$+\int_0^t \Big(\int_0^u \partial_1 K_2(t+v,s)dv + K_2(t,s)\Big)\sigma(X_s)dW_s$$
$$= \int_0^u \Big(\int_0^t \partial_1 K_1(t+v,s)b(X_s)ds\Big)dv + \int_0^t K_1(t,s)b(X_s)ds$$
$$+\int_0^u \Big(\int_0^t \partial_1 K_2(t+v,s)\sigma(X_s)dW_s\Big)dv + \int_0^t K_2(t,s)\sigma(X_s)dW_s,$$

where the interchange is ensured by the stochastic Fubini theorem, we obtain that Y_t is almost surely $\mathcal{C}^1$ and Lipschitz-continuous (using theorem 3.2 again) with

$$\dot{Y}_t(u) = \int_0^t \partial_1 K_1(t+u,s)b(X_s)ds + \int_0^t \partial_1 K_2(t+u,s)\sigma(X_s)dW_s. \tag{4.2}$$

We also note that

$$Y_t(0) = X_t - X_0, \quad t \in [0,T]. \tag{4.3}$$

Then we have

$$X_v = X_0 + Y_t(v-t) + \int_t^v K_1(v,s)b(X_s)ds + \int_t^v K_2(v,s)\sigma(X_s)dW_s, \tag{4.4}$$
$$Y_v(u) = Y_t(v-t+u) + \int_t^v K_1(v+u,s)b(X_s)ds + \int_t^v K_2(v+u,s)\sigma(X_s)dW_s, \tag{4.5}$$
$$0 \le t \le v \le T,\ u \ge 0.$$

This leads us to define the following non-homogeneous semi-group type operator $P_{r,t}$ for $t \in [0,T]$ and $r \in [0, T-t]$ on Ω_{2T}:

$$P_{r,t}(\omega)_u = \omega_{r+u} + \int_t^{t+r} K_1(t+r+u,s)b(\widetilde{X}_s)ds + \int_t^{t+r} K_2(t+r+u,s)\sigma(\widetilde{X}_s)dW_s, \quad u \ge 0, \tag{4.6}$$

where $(\widetilde{X}_s)_{s\in[t,t+r]}$ is the solution of the following SVE:

$$\widetilde{X}_v = X_0 + \omega_{v-t} + \int_t^v K_1(v,s)b(\widetilde{X}_s)ds + \int_t^v K_2(v,s)\sigma(\widetilde{X}_s)dW_s \tag{4.7}$$

and where we omit here the dependency of $\widetilde{X}$ in t and in ω. Since $K_i(u,s) = 0$ for $u \ge 2T$, we have indeed that for every $\omega \in \Omega_{2T}$, $P_{r,t}(\omega) \in \Omega_{2T}$ almost surely and then $P_{r,t}$ maps Ω_{2T} to a set of $\mathcal{F}_{t+r}$-measurable trajectories in Ω_{2T}. Then following (4.4) and (4.5) we have

$$P_{r,t}(Y_t) = Y_{t+r}, \tag{4.8}$$

hence the designation "semi-group type operator", although $P_{r,t}$ is not a true semi-group on Ω_{2T}.

Likewise, we define

$$\bar{Y}_t(u) = \int_0^t K_1(t+u,s)b(\bar{X}_{\underline{s}})ds + \int_0^t K_2(t+u,s)\sigma(\bar{X}_{\underline{s}})dW_s, \quad u \ge 0,\ t \in [0,T] \tag{4.9}$$

as well as the semi-group corresponding to the Euler-Maruyama scheme (2.3) for $k \in \{0,\dots,N-1\}$ and $r \in [0, T-t_k]$:

$$\bar{P}_{r,t_k}(\omega)_u = \omega_{r+u} + \int_{t_k}^{t_k+r} K_1(t_k+r+u,s)b(X_0+\omega_0)ds + \int_{t_k}^{t_k+r} K_2(t_k+r+u,s)\sigma(X_0+\omega_0)dW_s, \tag{4.10}$$

for $u \ge 0$, so that we have

$$\bar{Y}_t(0) = \bar{X}_t - X_0, \quad t \in [0,T] \tag{4.11}$$

and for $r \in [0,h]$:

$$\bar{P}_{r,t_k}(\bar{Y}_{t_k})_u = \bar{Y}_{t_k}(r+u) + \int_{t_k}^{t_k+r} K_1(t_k+r+u,s)b(X_0+\bar{Y}_{t_k}(0))ds$$
$$+\int_{t_k}^{t_k+r} K_2(t_k+r+u,s)\sigma(X_0+\bar{Y}_{t_k}(0))dW_s$$

$$= \bar{Y}_{t_k+r}(u).$$

By a slight abuse of notation, we use the notations P and $\bar{P}$ also to denote the semi-group type operators such that for every $g : \Omega_{2T} \to \mathbb{R}$, $\omega \in \Omega_{2T}$, $t \in [0,T]$ and $r \in [0, T-t]$ we have

$$P_{r,t}g(\omega) := \mathbb{E}g(P_{r,t}(\omega)), \quad \bar{P}_{r,t}g(\omega) := \mathbb{E}g(\bar{P}_{r,t}(\omega)).$$

More precisely, we consider $P_{r,t}g$ and $\bar{P}_{r,t}g$ for $g \in \widetilde{\mathcal{C}}_b^1(\Omega_{2T}, \mathbb{R})$ so that for $\omega \in \Omega_{2T}$,

$$\mathbb{E}\big|g(P_{r,t}(\omega))\big| \leq |g(\widetilde{0})| + \|\nabla g\|_\infty \mathbb{E}\|P_{r,t}(\omega)\|_\infty < \infty,$$

as $P_{r,t}(\omega) \in \Omega_{2T}$ almost surely with $\mathbb{E}\|P_{r,t}(\omega)\|_\infty < \infty$ according to theorem 3.2 and theorem 2.2. Then $P_{r,t}$ and $\bar{P}_{r,t}$ map $\widetilde{\mathcal{C}}_b^1(\Omega_{2T}, \mathbb{R})$ to a set of measurable functions from Ω_{2T} to $\mathbb{R}$.

For general semi-groups $Q_1, \ldots, Q_r$ we denote their composition as

$$\prod_{k=1}^{r} Q_k := Q_1 \circ \ldots \circ Q_r.$$

Then we obtain $X_T = Y_T(0) + X_0 = P_{T,0}(\widetilde{0})_0 + X_0$ and $\bar{X}_T = \big(\prod_{k=0}^{N-1} \bar{P}_{h,t_{N-1-k}}(\widetilde{0})\big)_0 + X_0$.

Now for $f : \mathbb{R} \to \mathbb{R}$ being $\widetilde{\mathcal{C}}_b^5$ we define

$$\widetilde{f} : \omega \in \Omega_{2T} \mapsto f(\omega_0 + X_0). \tag{4.12}$$

Following theorem 3.1, we also have $\widetilde{f} \in \widetilde{\mathcal{C}}_b^5$. Moreover we can write

$$\mathbb{E}f(\bar{X}_T) = \mathbb{E}f\big(\big(\prod_{k=0}^{N-1} \bar{P}_{h,t_{N-1-k}}(\widetilde{0})\big)_0 + X_0\big) = \mathbb{E}\widetilde{f}\big(\prod_{k=0}^{N-1} \bar{P}_{h,t_{N-1-k}}(\widetilde{0})\big) = \left(\prod_{k=0}^{N-1} \bar{P}_{h,t_k}\widetilde{f}\right)(\widetilde{0}).$$

We highlight that in our notations the order of the operators is reversed whether $\prod_k \bar{P}_{h,t_k}$ is applied to some $\omega \in \Omega_{2T}$ or to some $g : \Omega_{2T} \to \mathbb{R}$. We then rewrite the weak error as

$$\begin{aligned}
\mathbb{E}f(\bar{X}_T) - \mathbb{E}f(X_T) &= \left(\prod_{k=0}^{N-1} \bar{P}_{h,t_k}\widetilde{f}\right)(\widetilde{0}) - \left(\prod_{k=0}^{N-1} P_{h,t_k}\widetilde{f}\right)(\widetilde{0}) \\
&= \sum_{k=0}^{N-1} \left(P_{kh,0} \circ (\bar{P}_{h,t_k} - P_{h,t_k}) \circ \prod_{j=k+1}^{N-1} \bar{P}_{h,t_j}\widetilde{f}\right)(\widetilde{0}),
\end{aligned} \tag{4.13}$$

where we used a telescopic sum.

4.2 Weak error in small time

In this section, we give a bound on the weak error in small time for the one-step Euler-Maruyama scheme $(\bar{P}_{h,t_k} - P_{h,t_k})g(\omega)$, where $g : \Omega_{2T} \to \mathbb{R}$ is some smooth functional and $\omega \in \Omega_{2T}$ and $\omega \in \mathcal{C}^2$.

Proposition 4.1. *Let $\omega \in \Omega_{2T}$ and be $\mathcal{C}^2$, $g : \Omega_{2T} \to \mathbb{R}$ with bounded (pathwise) derivatives up to order 5 and $k \in \{0, \ldots, N-1\}$. Then we have*

$$\big|(\bar{P}_{h,t_k} - P_{h,t_k})g(\omega)\big| \leq C(1 + [\omega]_{\text{Lip}})h^2, \tag{4.14}$$

where the constant C does not depend on k nor ω nor h.

Proof. We can assume that ω is Lipschitz-continuous without loss of generality.

- Let us consider $(\widetilde{X}_s)_{s\in[t_k,t_k+h]}$ as defined in (4.7). Then for $v \in [t_k, t_{k+1}]$ and $\varepsilon \in [0, t_{k+1} - v]$ we have

$$\begin{aligned}
\widetilde{X}_{v+\varepsilon} - \widetilde{X}_v = \omega_{v+\varepsilon-t_k} - \omega_{v-t_k} &+ \int_{t_k}^{v} (K_1(v+\varepsilon, s) - K_1(v,s))b(\widetilde{X}_s)ds \\
&+ \int_{v}^{v+\varepsilon} K_1(v+\varepsilon, s)b(\widetilde{X}_s)ds \\
&+ \int_{t_k}^{v} (K_2(v+\varepsilon, s) - K_2(v,s))\sigma(\widetilde{X}_s)dW_s + \int_{v}^{v+\varepsilon} K_2(v+\varepsilon, s)\sigma(\widetilde{X}_s)dW_s
\end{aligned}$$

so that

$$d\widetilde{X}_v = \dot{\omega}_{v-t_k}dv + K_1(v,v)b(\widetilde{X}_v)dv + K_2(v,v)\sigma(\widetilde{X}_v)dW_v$$
$$+ \left(\int_{t_k}^v \partial_1 K_1(v,s)b(\widetilde{X}_s)ds + \int_{t_k}^v \partial_1 K_2(v,s)\sigma(\widetilde{X}_s)dW_s\right)dv.$$

It follows from the classic Itô formula that for $\psi : \mathbb{R} \to \mathbb{R}$ and $v \geq t_k$ we have

$$\begin{aligned} d\psi(\widetilde{X}_v) &= \nabla\psi(\widetilde{X}_v)(K_1(v,v)b(\widetilde{X}_v)dv + \dot{\omega}_{v-t_k}dv + K_2(v,v)\sigma(\widetilde{X}_v)dW_v) \\ &+ \frac{1}{2}\nabla^2\psi(\widetilde{X}_v)K_2^2(v,v)\sigma^2(\widetilde{X}_v)dv \\ &+ \nabla\psi(\widetilde{X}_v)\left(\int_{t_k}^v \partial_1 K_1(v,s)b(\widetilde{X}_s)ds + \int_{t_k}^v \partial_1 K_2(v,s)\sigma(\widetilde{X}_s)dW_s\right)dv \end{aligned} \tag{4.15}$$

In particular we remark that for $r \in [0,h]$ and if ψ is $\mathcal{C}^2$ with bounded derivatives,

$$\begin{aligned} &\left|\mathbb{E}\psi(\widetilde{X}_{t_k+r}) - \psi(X_0+\omega_0)\right| \qquad (4.16) \\ &\leq \|\nabla\psi\|_\infty(\|K_1\|_\infty(|b(0)| + [b]_{\text{Lip}}\sup_v \mathbb{E}|\widetilde{X}_v|)r + [\omega]_{\text{Lip}}) + \frac{1}{2}\|\nabla^2\psi\|_\infty\|K_2\|_\infty^2\|\sigma\|_\infty^2 r \\ &\quad + \|\nabla\psi\|_\infty\|\partial_1 K_1\|_\infty(|b(0)| + [b]_{\text{Lip}}\sup_v \mathbb{E}|\widetilde{X}_v|)\frac{r^2}{2} + \|\nabla\psi\|_\infty\|\partial_1 K_2\|_\infty\|\sigma\|_\infty\frac{2r^{3/2}}{3} \\ &\leq C(1+[\omega]_{\text{Lip}})r, \qquad (4.17) \end{aligned}$$

where we use theorem 2.2 on $\widetilde{X}$ and where we bound the last term as follows:

$$\begin{aligned} &\left|\mathbb{E}\int_{t_k}^{t_k+r}\nabla\psi(\widetilde{X}_v)\left(\int_{t_k}^v \partial_1 K_2(v,s)\sigma(\widetilde{X}_s)dW_s\right)dv\right| \\ &\leq \|\nabla\psi\|_\infty\int_{t_k}^{t_k+r}\mathbb{E}\left|\int_{t_k}^v \partial_1 K_2(v,s)\sigma(\widetilde{X}_s)dW_s\right|dv \\ &\leq \|\nabla\psi\|_\infty\|\partial_1 K_2\|_\infty\|\sigma\|_\infty\int_{t_k}^{t_k+r}(v-t_k)^{1/2}dv. \end{aligned}$$

- On the other side for $r \geq 0$ we have

$$\begin{aligned} P_{r+\varepsilon,t_k}(\omega)_u - P_{r,t_k}(\omega)_u &= \omega_{r+\varepsilon+u} - \omega_{r+u} + \int_{t_k+r}^{t_k+r+\varepsilon} K_1(t_k+r+\varepsilon+u,s)b(\widetilde{X}_s)ds \\ &+ \int_{t_k+r}^{t_k+r+\varepsilon} K_2(t_k+r+\varepsilon+u,s)\sigma(\widetilde{X}_s)dW_s \\ &+ \int_{t_k}^{t_k+r}(K_1(t_k+r+\varepsilon+u,s) - K_1(t_k+r+u,s))b(\widetilde{X}_s)ds \\ &+ \int_{t_k}^{t_k+r}(K_2(t_k+r+\varepsilon+u,s) - K_2(t_k+r+u,s))\sigma(\widetilde{X}_s)dW_s \end{aligned}$$

so that we can write

$$\begin{aligned} dP_{r,t_k}(\omega)_u &= K_1(t_k+r+u,t_k+r)b(\widetilde{X}_{t_k+r})dr + K_2(t_k+r+u,t_k+r)\sigma(\widetilde{X}_{t_k+r})dW_{t_k+r} \\ &+ \dot{\omega}_{r+u}dr + \left(\int_{t_k}^{t_k+r}\partial_1 K_1(t_k+r+u,s)b(\widetilde{X}_s)ds + \int_{t_k}^{t_k+r}\partial_1 K_2(t_k+r+u,s)\sigma(\widetilde{X}_s)dW_s\right)dr \end{aligned}$$

so that for $G : \Omega_{2T} \to \mathbb{R}$ being $\widetilde{\mathcal{C}}_b^3$ and using theorem 3.4 – checking the assumption (3.5) with theorem 2.2 for the drift part and theorem 3.2 for the diffusion part – we obtain

$$\begin{aligned} dG(P_{r,t_k}(\omega)) &= \langle\nabla G(P_{r,t_k}(\omega)), \dot{\omega}_{r+\cdot}\rangle dr \\ &+ \langle\nabla G(P_{r,t_k}(\omega)), (K_1(t_k+r+u,t_k+r))_{u\geq 0}\rangle b(\widetilde{X}_{t_k+r})dr \\ &+ \langle\nabla G(P_{r,t_k}(\omega)), (K_2(t_k+r+u,t_k+r))_{u\geq 0}\rangle\sigma(\widetilde{X}_{t_k+r})dW_{t_k+r} \end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2}\langle\nabla^2 G(P_{r,t_k}(\omega)),(K_2(t_k+r+u,t_k+r))_{u\geq 0}^{\otimes 2}\rangle\sigma^2(\widetilde{X}_{t_k+r})dr\\
&+\Big\langle\nabla G(P_{r,t_k}(\omega)),\Big(\int_{t_k}^{t_k+r}\partial_1 K_1(t_k+r+u,s)b(\widetilde{X}_s)ds\\
&\quad+\int_{t_k}^{t_k+r}\partial_1 K_2(t_k+r+u,s)\sigma(\widetilde{X}_s)dW_s\Big)_{u\geq 0}\Big\rangle dr.
\end{aligned}\tag{4.18}
$$

In particular, we remark that

$$
\begin{aligned}
&\big|\mathbb{E}G(P_{r,t_k}(\omega))-G(\omega)\big|\\
&\leq\|\nabla G\|_\infty[\omega]_{\text{Lip}}r+\|\nabla G\|_\infty\|K_1\|_\infty\big(|b(0)|+[b]_{\text{Lip}}\sup_v\mathbb{E}|\widetilde{X}|_v\big)r+\frac{1}{2}\|\nabla^2 G\|_\infty\|K_2\|_\infty^2\|\sigma\|_\infty^2 r\\
&\quad+\|\nabla G\|_\infty\|\partial_1 K_1\|_\infty\big(|b(0)|+[b]_{\text{Lip}}\sup_v\mathbb{E}|\widetilde{X}_v|\big)\frac{r^2}{2}+C\|\nabla G\|_\infty\|\sigma\|_\infty\|\partial_{11}^2 K_2\|_\infty^2\frac{2r^{3/2}}{3},\\
&\leq C(1+[\omega]_{\text{Lip}})r.
\end{aligned}\tag{4.19}
$$

where we used theorem 3.2 and theorem 2.2 to bound the last term.

Thus for $g:\Omega_{2T}\to\mathbb{R}$ being $\mathcal{C}_b^5$ we have

$$
\begin{aligned}
\mathbb{E}g(P_{h,t_k}(\omega))-g(\omega)&=\mathbb{E}\Big[\int_0^h\langle\nabla g(P_{r,t_k}(\omega)),\dot{\omega}_{r+\cdot}\rangle dr\\
&\quad+\frac{1}{2}\int_0^h\langle\nabla^2 g(P_{r,t_k}(\omega)),(K_2(t_k+r+u,t_k+r))_{u\geq 0}^{\otimes 2}\rangle\sigma^2(\widetilde{X}_{t_k+r})dr\\
&\quad+\int_0^h\langle\nabla g(P_{r,t_k}(\omega)),(K_1(t_k+r+u,t_k+r))_{u\geq 0}\rangle b(\widetilde{X}_{t_k+r})dr\\
&\quad+\int_0^h\Big\langle\nabla g(P_{r,t_k}(\omega)),\Big(\int_{t_k}^{t_k+r}\partial_1 K_1(t_k+r+u,s)b(\widetilde{X}_s)ds\\
&\qquad+\int_{t_k}^{t_k+r}\partial_1 K_2(t_k+r+u,s)\sigma(\widetilde{X}_s)dW_s\Big)_{u\geq 0}\Big\rangle dr\Big]\\
&=:\sum_{i=1}^5 I_i.
\end{aligned}
$$

Likewise, we obtain a similar formula on $\mathbb{E}g(\bar{P}_{h,t_k}(\omega))-g(\omega)$, replacing b by $\bar{b}$ and σ by $\bar{\sigma}$, and we write

$$
\mathbb{E}g(\bar{P}_{h,t_k}(\omega))-g(\omega)=:\sum_{i=1}^5\bar{I}_i.
$$

We shall now inspect the quantity $I_i-\bar{I}_i$ for every $i=1,\ldots,5$.

- For fixed $r\in[0,h]$, let $G_r:\eta\mapsto\langle\nabla g(\eta),\dot{\omega}_{r+\cdot}\rangle$. Then $G_r\in\mathcal{C}_b^3$ and we have

$$
\langle\nabla G_r(\eta),\tau\rangle=\langle\nabla^2 g(\eta),\dot{\omega}_{r+\cdot}\otimes\tau\rangle,\quad\langle\nabla^2 G_r(\eta),\tau^1\otimes\tau^2\rangle=\langle\nabla^3 g(\eta),\dot{\omega}_{r+\cdot}\otimes\tau^1\otimes\tau^2\rangle.
$$

Applying the Itô formula (4.18) again to $\alpha\mapsto\mathbb{E}G_r(P_{\alpha,t_k}(\omega))$ for $\alpha\in[0,r]$ and with the estimate (4.19), we obtain

$$
|\mathbb{E}G_r(P_{r,t_k}(\omega))-G_r(\omega)|\leq Cr(1+[\omega]_{\text{Lip}}).
$$

Similarly, we have

$$
\big|\mathbb{E}G_r(\bar{P}_{r,t_k}(\omega))-G_r(\omega)\big|\leq Cr(1+[\omega]_{\text{Lip}}),
$$

and then

$$
\begin{aligned}
|I_1-\bar{I}_1|=\Big|\mathbb{E}\Big[\int_0^h G_r(P_{r,t_k}(\omega))dr-\int_0^h G_r(\bar{P}_{h,t_k}(\omega))dr\Big]\Big|&\leq C(1+[\omega]_{\text{Lip}})\int_0^h rdr\\
&\leq C(1+[\omega]_{\text{Lip}})h^2.
\end{aligned}
$$

- For fixed $r\in[0,h]$, let

$$
G_r:\eta\mapsto\frac{1}{2}\langle\nabla^2 g(\eta),(K_2(t_k+r+u,t_k+r))_{u\geq 0}^{\otimes 2}\rangle.
$$

Then $G_r \in \mathcal{C}_b^3$ and we have

$$\langle \nabla G_r(\eta), \tau \rangle = \frac{1}{2}\langle \nabla^3 g(\eta), (K_2(t_k + r + u, t_k + r))_{u\geq 0}^{\otimes 2} \otimes \tau \rangle,$$
$$\langle \nabla^2 G_r(\eta), \tau^1 \otimes \tau^2 \rangle = \frac{1}{2}\langle \nabla^4 g(\eta), (K_2(t_k + r + u, t_k + r))_{u\geq 0}^{\otimes 2} \otimes \tau^1 \otimes \tau^2 \rangle.$$

Applying the Itô formulae we obtained in (4.18) and in (4.15) and the classic Itô formula for a product, we get

$$\begin{aligned}
&\mathbb{E}[G_r(P_{r,t_k}(\omega))\sigma^2(\widetilde{X}_{t_k+r})] - G_r(\omega)\sigma^2(X_0 + \omega_0) \\
&= \mathbb{E}\Big[\int_0^r d\big(G_r(P_{\alpha,t_k}(\omega))\big)\sigma^2(\widetilde{X}_{t_k+\alpha}) + G_r(P_{\alpha,t_k}(\omega))d\big(\sigma^2(\widetilde{X}_{t_k+\alpha})\big) \\
&\qquad + d\langle G_r(P_{\cdot,t_k}(\omega)), \sigma^2(\widetilde{X}_{t_k+\cdot})\rangle_\alpha\Big] \\
&=: A_1 + A_2 + A_3,
\end{aligned}$$

but σ^2 is bounded and following (4.19), we obtain that $A_1 \leq C(1 + [\omega]_{\text{Lip}})r$; the same way and since G_r is bounded (independently on r) and following (4.17) we have $A_2 \leq C(1 + [\omega]_{\text{Lip}})r$. Moreover we have

$$\begin{aligned}
d\langle G_r(P_{\cdot,t_k}(\omega)), \sigma^2(\widetilde{X}_{t_k+\cdot})\rangle_\alpha &= \nabla\sigma^2(\widetilde{X}_{t_k+\alpha})K_2(t_k + \alpha, t_k + \alpha)\sigma^2(\widetilde{X}_{t_k+\alpha}) \\
&\quad \cdot \langle \nabla G_r(P_{\alpha,t_k}(\omega)), (K_2(t_k + \alpha + u, t_k + \alpha))_{u\geq 0}\rangle d\alpha
\end{aligned}$$

so that same way we get $A_3 \leq Cr$. Thus we finally obtain

$$\big|\mathbb{E}G_r(P_{r,t_k}(\omega))\sigma^2(\widetilde{X}_{t_k+r}) - G_r(\omega)\sigma^2(X_0 + \omega_0)\big| \leq C(1 + [\omega]_{\text{Lip}})r. \tag{4.20}$$

The same way we have

$$\big|\mathbb{E}G_r(\bar{P}_{r,t_k}(\omega))\sigma^2(X_0 + \omega_0) - G_r(\omega)\sigma^2(X_0 + \omega_0)\big| \leq C(1 + [\omega]_{\text{Lip}})r,$$

so that

$$\big|I_2 - \bar{I}_2\big| = \Big|\mathbb{E}\Big[\int_0^h (G_r(P_{r,t_k}(\omega))\sigma^2(\widetilde{X}_{t_k+r}) - G_r(\bar{P}_{r,t_k}(\omega))\sigma^2(X_0 + \omega_0))dr\Big]\Big| \leq C(1 + [\omega]_{\text{Lip}})h^2.$$

- For $r \in [0, h]$ and $u \geq 0$ let us define

$$\varphi_u^r := \int_{t_k}^{t_k+r} \partial_1 K_2(t_k + r + u, s)\sigma(\widetilde{X}_s)dW_s$$

and let us write

$$d\varphi_u^r = \partial_1 K_2(t_k + r + u, t_k + r)\sigma(\widetilde{X}_{t_k+r})dW_{t_k+r} + \left(\int_{t_k}^{t_k+r} \partial_{11}^2 K_2(t_k + r + u, s)\sigma(\widetilde{X}_s)dW_s\right) dr. \tag{4.21}$$

Moreover, using theorem 3.2, we have $\mathbb{E}\|\varphi^r\|_\infty^2 \leq C$. We also define

$$G : (\eta^1, \eta^2) \in \Omega_{2T}^2 \mapsto \langle \nabla g(\eta^1), \eta^2 \rangle.$$

Then $G \in \mathcal{C}_b^3$ with

$$\langle \nabla G(\eta^1, \eta^2), (\tau^1, \tau^2)\rangle = \langle \nabla^2 g(\eta^1), \tau^1 \otimes \eta^2 \rangle + \langle \nabla g(\eta^1), \tau^2 \rangle,$$
$$\langle \nabla^2 G(\eta^1, \eta^2), (\tau^1, \tau^2)^{\otimes 2}\rangle = \langle \nabla^3 g(\eta^1), (\tau^1)^{\otimes 2} \otimes \eta^2 \rangle + 2\langle \nabla^2 g(\eta^1), \tau^1 \otimes \tau^2 \rangle.$$

Using (4.18) and (4.21), for every $r \in [0, h]$ we have

$$\begin{aligned}
&\langle \nabla g(P_{r,t_k}(\omega)), \left(\int_{t_k}^{t_k+r} \partial_1 K_2(t_k + r + u, s)\sigma(\widetilde{X}_s)dW_s\right)_{u\geq 0}\rangle \\
&= \int_0^r \Big[\langle \nabla^2 g(P_{\alpha,t_k}(\omega)), \dot{\omega}_{r+\cdot} \otimes \varphi^\alpha \rangle + \langle \nabla^2 g(P_{\alpha,t_k}(\omega)), (K_1(t_k + \alpha + u, t_k + \alpha))_{u\geq 0} \otimes \varphi^\alpha \rangle \\
&\cdot b(\widetilde{X}_{t_k+\alpha})
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2}\langle\nabla^3 g(P_{\alpha,t_k}(\omega)),(K_2(t_k+\alpha+u,t_k+u))_{u\geq 0}^{\otimes 2}\otimes\varphi^\alpha\rangle\sigma^2(\widetilde{X}_{t_k+\alpha})\\
&+\langle\nabla^2 g(P_{\alpha,t_k}(\omega)),\Big(\int_{t_k}^{t_k+\alpha}\partial_1 K_1(t_k+\alpha+u,s)b(\widetilde{X}_s)ds\\
&\quad+\int_{t_k}^{t_k+\alpha}\partial_1 K_2(t_k+\alpha+u,s)\sigma(\widetilde{X}_s)dW_s\Big)_{u\geq 0}\otimes\varphi^\alpha\rangle\Big]d\alpha\\
&+\int_0^r\langle\nabla^2 g(P_{\alpha,t_k}(\omega)),(K_2(t_k+\alpha+u,t_k+u))_{u\geq 0}\otimes\varphi^\alpha\rangle\sigma(\widetilde{X}_{t_k+\alpha})dW_{t_k+\alpha}\\
&+\int_0^r\Big[\langle\nabla g(P_{\alpha,t_k}(\omega)),\Big(\int_{t_k}^{t_k+\alpha}\partial_{11}^2 K_2(t_k+\alpha+u,s)\sigma(\widetilde{X}_s)dW_s\Big)_{u\geq 0}\rangle d\alpha\\
&+\int_0^r\langle\nabla g(P_{\alpha,t_k}(\omega)),(\partial_1 K_2(t_k+\alpha+u,t_k+\alpha))_{u\geq 0}\rangle\sigma(\widetilde{X}_{t_k+\alpha})dW_{t_k+\alpha}\\
&+\int_0^r\langle\nabla^2 g(P_{\alpha,t_k}(\omega)),(K_2(t_k+\alpha+u,t_k+u))_{u\geq 0}\otimes(\partial_1 K_2(t_k+\alpha+u,t_k+\alpha))_{u\geq 0}\rangle\\
&\quad\cdot\sigma^2(\widetilde{X}_{t_k+\alpha})d\alpha
\end{aligned}
$$

so that

$$
\Big|\mathbb{E}\langle\nabla g(P_{r,t_k}(\omega)),\Big(\int_{t_k}^{t_k+r}\partial_1 K_2(t_k+r+u,s)\sigma(\widetilde{X}_s)dW_s\Big)_{u\geq 0}\rangle\Big|\leq C(1+[\omega]_{\mathrm{Lip}})r.
$$

The same way, we obtain

$$
\Big|\mathbb{E}\langle\nabla g(\bar{P}_{r,t_k}(\omega)),\Big(\int_{t_k}^{t_k+r}\partial_1 K_2(t_k+r+u,s)\bar{\sigma}(X_0+\omega(0))dW_s\Big)_{u\geq 0}\rangle\Big|\leq C(1+[\omega]_{\mathrm{Lip}})r
$$

and then

$$
|I_5-\bar{I}_5|\leq C(1+[\omega]_{\mathrm{Lip}})h^2.
$$

- The arguments to prove that

$$
|I_3-\bar{I}_3|+|I_4-\bar{I}_4|\leq C(1+[\omega]_{\mathrm{Lip}})h^2
$$

are the same or simpler.

□

4.3 Proof that the derivatives of g are bounded

In this section, we prove that if we choose $g:\Omega\to\mathbb{R}$ as in (4.13), then g has bounded derivatives up to order 5 so that we can apply theorem 4.1 to g.

Lemma 4.2. *Let $k\in\{0,\dots,N-1\}$, let us define*

$$
g(\omega):=\prod_{j=k}^{N-1}\bar{P}_{h,t_j}\widetilde{f}(\omega)=\mathbb{E}\left[f\big((\prod_{j=k}^{N-1}\bar{P}_{h,t_{N-1+k-j}}\cdot\omega)_0\big)\right],
$$

where $\widetilde{f}$ is defined in (4.12). Then g is five times differentiable with

$$
\|\nabla^\ell g\|_\infty\leq C,\quad \ell\in\{1,\dots,5\}.
$$

Proof. We can rewrite

$$
g(\omega)=\mathbb{E}\left[f(\omega_{T-t_k}+\int_{t_k}^T K_1(T,s)\bar{b}(\hat{X}_s^\omega)ds+\int_{t_k}^T K_2(T,s)\bar{\sigma}(\hat{X}_s^\omega)dW_s)\right]
$$

where $\hat{X}^\omega$ follows the piecewise SVE:

$$
\hat{X}_v^\omega=X_0+\omega_{v-t_k}+\sum_{\ell=k}^j b(\hat{X}_{t_\ell}^\omega)\int_{t_\ell}^{t_{\ell+1}\wedge v}K_1(v,s)ds+\sum_{\ell=k}^j\sigma(\hat{X}_{t_\ell}^\omega)\int_{t_\ell}^{t_{\ell+1}\wedge v}K_2(v,s)dW_s,
$$

$$v \in [t_j, t_{j+1}],\ j \in \{k, \dots, N-1\}.$$

The process $\hat{X}^\omega$ depends on k, but we omit this dependency in the notation without ambiguity. We define the tangent process $(Z_v^\omega)_{v\in[T-t_k,T]}$ as the process of the linear operators on Ω_{2T} by induction as:

$$\langle Z_v^\omega, \eta\rangle = \eta_{v-t_k} + \sum_{\ell=k}^{j} \nabla b(\hat{X}_{t_\ell}^\omega) \cdot \langle Z_{t_\ell}^\omega, \eta\rangle \int_{t_\ell}^{t_{\ell+1}\wedge v} K_1(v,s)ds$$
$$+ \sum_{\ell=k}^{j} \nabla \sigma(\hat{X}_{t_\ell}^\omega) \cdot \langle Z_{t_\ell}^\omega, \eta\rangle \int_{t_\ell}^{t_{\ell+1}\wedge v} K_2(v,s)dW_s, \quad v \in [t_j, t_{j+1}],\ j \in \{k, \dots, N-1\}, \tag{4.22}$$

so that for every fixed v, we have $Z_v^\omega = \nabla \hat{X}_v^\omega$, where ∇ is taken with respect to ω. We now give a bound on $\|Z_v^\omega\|$. For every $\eta \in \Omega_{2T}$ and $v \in [T_k, T]$ we have

$$\langle Z_v^\omega, \eta\rangle = \eta_{v-t_k} + \int_{t_k}^{v} K_1(v,s)\nabla \bar{b}(\hat{X}_s^\omega)\langle Z_{\underline{s}}^\omega, \eta\rangle ds + \int_{t_k}^{v} K_2(v,s)\nabla \bar{\sigma}(\hat{X}_s^\omega)\langle Z_{\underline{s}}^\omega, \eta\rangle dW_s.$$

Let us denote

$$\varphi_v := \sup_{s\in[t_k,v]} \mathbb{E}\|Z_s^\omega\|^2, \quad v \in [t_k, T]$$

and we have (using the inequality $(a+b+c)^2 \le 3(a^2+b^2+c^2)$):

$$\varphi_v \le 3 + 3\int_{t_k}^{v} (T\|K_1\|_\infty^2 \|\nabla b\|_\infty^2 + \|K_2\|_\infty^2 \|\nabla\sigma\|_\infty^2)\varphi_s ds$$

so that using the Gronwall inequality:

$$\varphi_v \le 3\exp\left((v-t_k)(T\|K_1\|_\infty^2\|\nabla b\|_\infty^2 + \|K_2\|_\infty^2\|\nabla\sigma\|_\infty^2)\right) \le C.$$

Then we have

$$\langle \nabla g(\omega), \eta\rangle = \mathbb{E}\left[\nabla f\big((\prod_{j=k}^{N-1} \bar{P}_{h,t_{N-1+k-j}} \cdot \omega)_0\big) \cdot \Big(\eta_{T-t_k} + \int_{t_k}^{T} K_1(T,s)\nabla\bar{b}(\hat{X}_s^\omega)\langle Z_{\underline{s}}^\omega, \eta\rangle ds \right.$$
$$\left. + \int_{t_k}^{T} K_2(T,s)\nabla\bar{\sigma}(\hat{X}_s^\omega)\langle Z_{\underline{s}}^\omega, \eta\rangle dW_s\Big)\right]$$

implying

$$\|\nabla g\|_\infty \le \|\nabla f\|_\infty \Big(1 + \int_{t_k}^{T} \|K_1\|_\infty \|\nabla b\|_\infty \varphi_s^{1/2} ds + (\int_{t_k}^{T} \|K_2\|_\infty^2 \|\nabla\sigma\|_\infty^2 \varphi_s ds)^{1/2}\Big)$$

thus implying that ∇g is bounded (independently of k and N).

We prove that the derivatives of g are bounded up to order 5 by following the same method.

□

4.4 Conclusion: proof of theorem 2.5

Proof. Let us consider (4.13) again and for $k \in \{0, \dots, N-1\}$ we set $\omega^k := P_{kh,0}(\widetilde{0})$ and

$$g_{k+1} := \prod_{j=k+1}^{N-1} \bar{P}_{h,t_j} \widetilde{f}.$$

Then following theorem 4.2, we have that $g_{k+1} \in \widetilde{\mathcal{C}}_b^5$. On the other side, we have that ω^k is $\mathcal{C}^2$ with

$$\dot{\omega}_u^k = \int_0^{t_k} \partial_1 K_1(t_k+u,s)b(\widetilde{X}_s)ds + \int_0^{t_k} \partial_1 K_2(t_k+u,s)\sigma(\widetilde{X}_s)dW_s,$$
$$\ddot{\omega}_u^k = \int_0^{t_k} \partial_{11}^2 K_1(t_k+u,s)b(\widetilde{X}_s)ds + \int_0^{t_k} \partial_{11}^2 K_2(t_k+u,s)\sigma(\widetilde{X}_s)dW_s$$

where the interchange is ensured by the stochastic Fubini theorem in the same way as in (4.2), and following theorem 3.2, we obtain that $\mathbb{E}[\omega^k]_{\text{Lip}} \le C$. Then applying theorem 4.1 with $g = g_{k+1}$ and $\omega = \omega^k$ we get

$$\left| \left(P_{kh,0} \circ (\bar{P}_{h,t_k} - P_{h,t_k}) \circ \prod_{j=k+1}^{n-1} \bar{P}_{h,t_j} \widetilde{f} \right) (\widetilde{0}) \right| = \left| \left(P_{kh,0} \circ (\bar{P}_{h,t_k} - P_{h,t_k}) g \right) (\widetilde{0}) \right|$$
$$\le \mathbb{E} \left| \left((\bar{P}_{h,t_k} - P_{h,t_k}) g \right) (\omega^k) \right| \le C(1 + \mathbb{E}[\omega^k]_{\text{Lip}}) h^2 \le C h^2.$$

Summing over $k \in \{0, \dots, N-1\}$ yields

$$\mathbb{E}[f(\bar{X}_T)] - \mathbb{E}[f(X_T)] = O\left(\frac{1}{N}\right).$$

□

4.5 Proof of weak error for the scheme with discretization of the kernels

The proof of (2.9) for the scheme $\overrightarrow{X}$ defined in (2.4) is similar to the proof for (2.8). We define the associated semi-group type operator as

$$\overrightarrow{P}_{r,t_k}(\omega)_u = \omega_{r+u} + \int_{t_k}^{t_k+r} K_1(t_k + r + u, t_k) b(X_0 + \omega_0) ds + \int_{t_k}^{t_k+r} K_2(t_k + r + u, t_k) \sigma(X_0 + \omega_0) dW_s,$$

for $u \ge 0$. The estimate for the weak error in small time (4.14) also holds for $|(\overrightarrow{P}_{h,t_k} - P_{h,t_k}) g(\omega)|$; the only necessary adaptation in the proof is for the estimate for I_2 and I_3. Indeed, instead of (4.20) we need to prove

$$\left| \mathbb{E} G_r(P_{r,t_k}(\omega)) \sigma^2(\widetilde{X}_{t_k+r}) - \frac{1}{2} \langle \nabla^2 g(\omega), K_2(t_k + r + \cdot, t_k)^{\otimes 2} \rangle \sigma^2(X_0 + \omega_0) \right| \le C(1 + [\omega]_{\text{Lip}}) r.$$

But we have

$$\alpha \mapsto \langle \nabla^2 g(\omega), K_2(t_k + r + \cdot, t_k + \alpha)^{\otimes 2} \rangle$$

is $\mathcal{C}^1$ with derivative

$$\alpha \mapsto 2 \langle \nabla^2 g(\omega), K_2(t_k + r + \cdot, t_k) \otimes \partial_2 K_2(t_k + r + \cdot, t_k) \rangle$$

and since g, K_2, $\partial_2 K_2$ and σ are bounded we have

$$\left| \mathbb{E} G_r(P_{r,t_k}(\omega)) \sigma^2(\widetilde{X}_{t_k+r}) - \frac{1}{2} \langle \nabla^2 g(\omega), K_2(t_k + r + \cdot, t_k)^{\otimes 2} \rangle \sigma^2(X_0 + \omega_0) \right|$$
$$\le \left| \mathbb{E} G_r(P_{r,t_k}(\omega)) \sigma^2(\widetilde{X}_{t_k+r}) - G_r(\omega) \sigma^2(X_0 + \omega_0) \right|$$
$$+ \left| \mathbb{E} G_r(\omega) \sigma^2(X_0 + \omega_0) - \frac{1}{2} \langle \nabla^2 g(\omega), K_2(t_k + r + \cdot, t_k)^{\otimes 2} \rangle \sigma^2(X_0 + \omega_0) \right|$$
$$\le C(1 + [\omega]_{\text{Lip}}) r.$$

The argument for the estimate of I_3 is similar.

Having proved the estimate for the weak error in small time, the conclusion of the proof is the same as for $\bar{X}$.

5 Simulations

In order to numerically check the convergence rate obtained in theorem 2.5, we empirically measure the weak convergence rate in the case of a stochastic volatility model where the volatility follows some Volterra equation. We consider the following Volterra version of the Stein-Stein model [37] where the analogous rough version was introduced in [2]:

$$\begin{cases} dS_t = S_t V_t dB_t, \quad S_0 > 0, \\ V_t = V_0 + g_0(t) + \kappa \int_0^t K(t-s) V_s ds + \nu \int_0^t K(t-s) dW_s \end{cases} \tag{5.1}$$

where the asset price process S and the square volatility process V take their values in $\mathbb{R}$, the function $g_0 : \mathbb{R}^+ \to \mathbb{R}$ is deterministic and continuous, the processes B and W are standard Brownian motions with correlation $\rho \in [-1, 1]$ and the non-singular kernel K is given by the shifted power-law kernel [25]:

$$K(t) = A_1 (A_2 + t)^{-1/4},$$

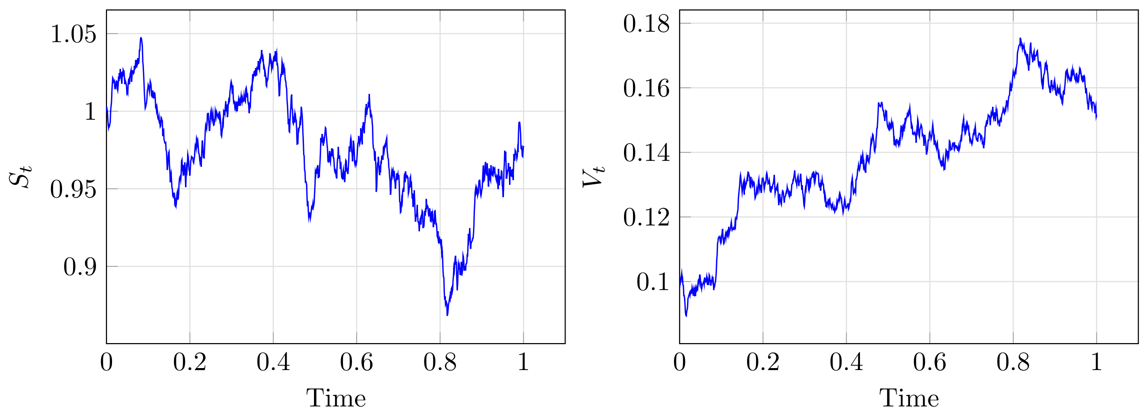

Figure 1: Example of trajectory of the asset price and the square volatility processes following the SVE (5.1).

with $A_1,\ A_2 > 0$. The process $(S_t, V_t)^\top$ in (5.1) is a special case of the Volterra equation (2.1) with 2×2 matrix kernels

$$K_1(t,s) = K_2(t,s) = \begin{pmatrix} 1 & 0 \\ 0 & K(t-s) \end{pmatrix}.$$

We consider the payoff function given by the Call option

$$f(x) = (x - \mathcal{K})_+$$

with strike $\mathcal{K} \geq 0$.

We simulate $(\overrightarrow{S}_T, \overrightarrow{V}_T)$ by discretizing the kernel K using weights matching the second moment, see [20, Section 3] and we plot $\mathbb{E}f(\overrightarrow{S}_T^{\lfloor \beta N \rfloor}) - \mathbb{E}f(\overrightarrow{S}_T^N)$ for some $\beta \in (1,2]$ and for different values of N, where N is the number of steps in the Euler-Maruyama scheme of the SVE. If $\mathbb{E}f(\overrightarrow{S}_T^N) = \mathbb{E}f(S_T) + O(1/N)$, then we should also have

$$\mathbb{E}f(\overrightarrow{S}_T^{\lfloor \beta N \rfloor}) - \mathbb{E}f(\overrightarrow{S}_T^N) = O(1/N).$$

An example of trajectory is given in fig. 1 and the results are given in fig. 2 with the following parameters:

$$T = 1,\ S_0 = 1,\ \mathcal{K} = 1,\ \kappa = 0.01,\ X_0 = 0.1,\ \rho = -0.7,\ \nu = 0.05,\ A_1 = 0.3,\ A_2 = 0.02,$$
$$\beta = 1.5,\ g_0 : t \mapsto (4\theta)/(3A_1)t^{3/4},\ \theta = 0.01.$$

We empirically obtain a convergence rate for the weak error which is approximatively -1, thus confirming the results in theorem 2.5.

Acknowledgements

The authors thank Professor Gilles Pagès, Sorbonne University, Paris, and Professor Toshihiro Yamada, Hitotsubashi University, Tokyo, for helpful discussions.

This work was conducted as the first author was doing an internship at Graduate School of Engineering Science, Osaka University, as a JSPS International Research Fellow. The first author thanks the Japanese Society for Promotion of Science JSPS for the financial support under the JSPS Summer Program.

References

[1] E. Abi Jaber, *Stochastic Invariance and Stochastic Volterra Equations*, theses, Université Paris sciences et lettres, Oct. 2018, https://tel.archives-ouvertes.fr/tel-01956320.

[2] E. Abi Jaber, *The characteristic function of Gaussian stochastic volatility models: an analytic expression*, Finance Stoch., 26 (2022), pp. 733–769, https://doi.org/10.1007/s00780-022-00489-4, https://doi.org/10.1007/s00780-022-00489-4.

[3] A. Alfonsi and A. Kebaier, *Approximation of Stochastic Volterra Equations with kernels of completely monotone type*, arXiv e-prints, (2021), arXiv:2102.13505, p. arXiv:2102.13505, https://doi.org/10.48550/arXiv.2102.13505, https://arxiv.org/abs/2102.13505.

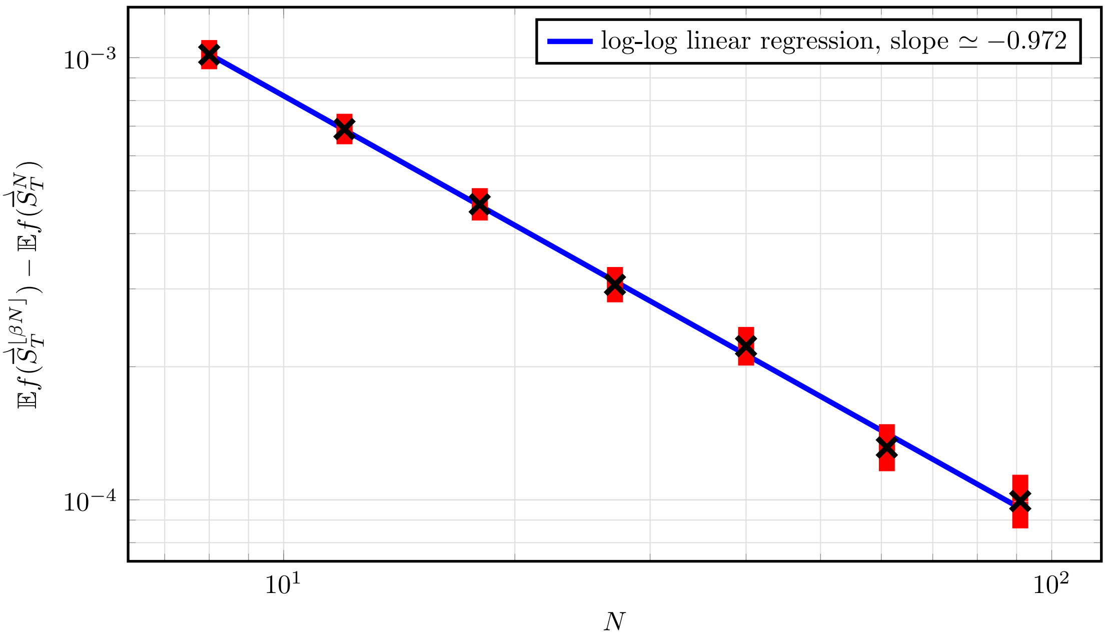

Figure 2: Simulation of (5.1) and weak error in log-log scale. We give in red the 95% confidence intervals where the number of trajectories is $512 \times 1024 \times 2000$ for each value of N.

[4] E. Alòs, J. A. León, and J. Vives, *On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility*, Finance Stoch., 11 (2007), pp. 571–589, https://doi.org/10.1007/s00780-007-0049-1, https://doi.org/10.1007/s00780-007-0049-1.

[5] C. Bayer, M. Fukasawa, and S. Nakahara, *Short communication: On the weak convergence rate in the discretization of rough volatility models*, SIAM Journal on Financial Mathematics, 13 (2022), pp. SC66–SC73, https://doi.org/10.1137/22M1482871, https://doi.org/10.1137/22M1482871, https://arxiv.org/abs/https://doi.org/10.1137/22M1482871.

[6] C. Bayer, E. J. Hall, and R. Tempone, *Weak error rates for option pricing under linear rough volatility*, International Journal of Theoretical and Applied Finance, 25 (2022), p. 2250029, https://doi.org/10.1142/S0219024922500297, https://doi.org/10.1142/S0219024922500297, https://arxiv.org/abs/https://doi.org/10.1142/S0219024922500297.

[7] M. Bennedsen, A. Lunde, and M. S. Pakkanen, *Decoupling the Short- and Long-Term Behavior of Stochastic Volatility*, Journal of Financial Econometrics, 20 (2021), pp. 961–1006, https://doi.org/10.1093/jjfinec/nbaa049, https://doi.org/10.1093/jjfinec/nbaa049, https://arxiv.org/abs/https://academic.oup.com/jfec/article-pdf/20/5/961/46754291/nbaa049.pdf.

[8] M. A. Berger and V. J. Mizel, *Volterra equations with Itô integrals. II*, J. Integral Equations, 2 (1980), pp. 319–337.

[9] A. E. Bolko, K. Christensen, M. S. Pakkanen, and B. Veliyev, *A GMM approach to estimate the roughness of stochastic volatility*, Journal of Econometrics, 235 (2023), pp. 745–778, https://doi.org/https://doi.org/10.1016/j.jeconom.2022.06.009, https://www.sciencedirect.com/science/article/pii/S0304407622001476.

[10] O. Bonesini, A. Jacquier, and A. Pannier, *Rough volatility, path-dependent PDEs and weak rates of convergence*, arXiv e-prints, (2023), arXiv:2304.03042, p. arXiv:2304.03042, https://doi.org/10.48550/arXiv.2304.03042, https://arxiv.org/abs/2304.03042.

[11] P. Cheridito, H. Kawaguchi, and M. Maejima, *Fractional Ornstein-Uhlenbeck processes*, Electronic Journal of Probability, 8 (2003), pp. 1 – 14, https://doi.org/10.1214/EJP.v8-125, https://doi.org/10.1214/EJP.v8-125.

[12] P. F. Christian Bayer and J. Gatheral, *Pricing under rough volatility*, Quantitative Finance, 16 (2016), pp. 887–904, https://doi.org/10.1080/14697688.2015.1099717, https://doi.org/10.

1080/14697688.2015.1099717, https://arxiv.org/abs/https://doi.org/10.1080/14697688.2015.1099717.

[13] B. Dupire, *Functional Itô calculus*, Quant. Finance, 19 (2019), pp. 721–729, https://doi.org/10.1080/14697688.2019.1575974, https://doi.org/10.1080/14697688.2019.1575974.

[14] O. El Euch, M. Fukasawa, and M. Rosenbaum, *The microstructural foundations of leverage effect and rough volatility*, Finance Stoch., 22 (2018), pp. 241–280, https://doi.org/10.1007/s00780-018-0360-z, https://doi.org/10.1007/s00780-018-0360-z.

[15] P. K. Friz, W. Salkeld, and T. Wagenhofer, *Weak error estimates for rough volatility models*, arXiv e-prints, (2022), arXiv:2212.01591, p. arXiv:2212.01591, https://doi.org/10.48550/arXiv.2212.01591, https://arxiv.org/abs/2212.01591.

[16] M. Fukasawa, *Short-time at-the-money skew and rough fractional volatility*, Quant. Finance, 17 (2017), pp. 189–198, https://doi.org/10.1080/14697688.2016.1197410, https://doi.org/10.1080/14697688.2016.1197410.

[17] M. Fukasawa, *Volatility has to be rough*, Quant. Finance, 21 (2021), pp. 1–8, https://doi.org/10.1080/14697688.2020.1825781, https://doi.org/10.1080/14697688.2020.1825781.

[18] M. Fukasawa, T. Takabatake, and R. Westphal, *Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics*, Mathematical Finance, 32 (2022), pp. 1086–1132, https://doi.org/https://doi.org/10.1111/mafi.12354, https://onlinelibrary.wiley.com/doi/abs/10.1111/mafi.12354, https://arxiv.org/abs/https://onlinelibrary.wiley.com/doi/pdf/10.1111/mafi.12354.

[19] M. Fukasawa and T. Ugai, *Limit distributions for the discretization error of stochastic Volterra equations*, arXiv e-prints, to appear in Annals of Applied Probability, (2021), arXiv:2112.06471, p. arXiv:2112.06471, https://doi.org/10.48550/arXiv.2112.06471, https://arxiv.org/abs/2112.06471.

[20] P. Gassiat, *Weak error rates of numerical schemes for rough volatility*, SIAM Journal on Financial Mathematics, 14 (2023), pp. 475–496, https://doi.org/10.1137/22M1485760, https://doi.org/10.1137/22M1485760, https://arxiv.org/abs/https://doi.org/10.1137/22M1485760.

[21] J. Gatheral, T. Jaisson, and M. Rosenbaum, *Volatility is rough*, Quant. Finance, 18 (2018), pp. 933–949, https://doi.org/10.1080/14697688.2017.1393551, https://doi.org/10.1080/14697688.2017.1393551.

[22] M. B. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617, https://doi.org/10.1287/opre.1070.0496, https://doi.org/10.1287/opre.1070.0496.

[23] M. B. Giles and L. Szpruch, *Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation*, The Annals of Applied Probability, 24 (2014), pp. 1585 – 1620, https://doi.org/10.1214/13-AAP957, https://doi.org/10.1214/13-AAP957.

[24] G. Gripenberg, S. O. Londen, and O. Staffans, *Volterra Integral and Functional Equations*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 1990, https://doi.org/10.1017/CBO9780511662805.

[25] J. Guyon and M. El Amrani, *Does the Term-Structure of Equity At-the-Money Skew Really Follow a Power Law?*, Available at SSRN, (2022), https://doi.org/10.2139/ssrn.4174538.

[26] E. A. Jaber, M. Larsson, and S. Pulido, *Affine Volterra processes*, The Annals of Applied Probability, 29 (2017), pp. 3155 – 3200, https://doi.org/10.1214/EJP.v8-125.

[27] B. Jourdain and G. Pagès, *Convex ordering for stochastic Volterra equations and their Euler schemes*, arXiv e-prints, (2022), arXiv:2211.10186, p. arXiv:2211.10186, https://doi.org/10.48550/arXiv.2211.10186, https://arxiv.org/abs/2211.10186.

[28] P. Jusselin and M. Rosenbaum, *No-arbitrage implies power-law market impact and rough volatility*, Math. Finance, 30 (2020), pp. 1309–1336, https://doi.org/10.1111/mafi.12254, https://doi.org/10.1111/mafi.12254.

[29] J.-F. Le Gall, *Brownian motion, martingales, and stochastic calculus*, vol. 274 of Graduate Texts in Mathematics, Springer, [Cham], french ed., 2016, https://doi.org/10.1007/978-3-319-31089-3, https://doi.org/10.1007/978-3-319-31089-3.

[30] V. Lemaire and G. Pagès, *Multilevel Richardson-Romberg extrapolation*, Bernoulli, 23 (2017), pp. 2643–2692, https://doi.org/10.3150/16-BEJ822, https://doi.org/10.3150/16-BEJ822.

[31] S.-E. A. Mohammed, *Stochastic differential systems with memory: Theory, examples and applications*, in Stochastic Analysis and Related Topics VI, L. Decreusefond, B. Øksendal, J. Gjerde, and A. S. Üstünel, eds., Boston, MA, 1998, Birkhäuser Boston, pp. 1–77.

[32] D. Nualart and B. Saikia, *Error distribution of the Euler approximation scheme for stochastic Volterra equations*, Journal of Theoretical Probability, 36 (2023), pp. 1829–1876, https://doi.org/https://doi.org/10.1007/s10959-022-01222-9.

[33] G. Pagès, *Numerical probability*, Universitext, Springer, Cham, 2018, https://doi.org/10.1007/978-3-319-90276-0, https://doi.org/10.1007/978-3-319-90276-0. An introduction with applications to finance.

[34] D. J. Prömel and D. Scheffels, *Neural stochastic Volterra equations: learning path-dependent dynamics*, arXiv e-prints, (2024), arXiv:2407.19557, p. arXiv:2407.19557, https://doi.org/10.48550/arXiv.2407.19557, https://arxiv.org/abs/2407.19557.

[35] P. Protter, *Volterra equations driven by semimartingales*, Ann. Probab., 13 (1985), pp. 519–530, http://links.jstor.org/sici?sici=0091-1798(198505)13:2<519:VEDBS>2.0.CO;2-3&origin=MSN.

[36] A. Richard, X. Tan, and F. Yang, *Discrete-time simulation of stochastic Volterra equations*, Stochastic Process. Appl., 141 (2021), pp. 109–138, https://doi.org/10.1016/j.spa.2021.07.003, https://doi.org/10.1016/j.spa.2021.07.003.

[37] E. M. Stein and J. C. Stein, *Stock price distributions with stochastic volatility: An analytic approach*, The Review of Financial Studies, 4 (1991), pp. 727–752, http://www.jstor.org/stable/2962156 (accessed 2023-02-10).

[38] D. Talay and L. Tubaro, *Expansion of the global error for numerical schemes solving stochastic differential equations*, Stochastic Anal. Appl., 8 (1990), pp. 483–509 (1991), https://doi.org/10.1080/07362999008809220, https://doi.org/10.1080/07362999008809220.

[39] F. Viens and J. Zhang, *A martingale approach for fractional Brownian motions and related path dependent PDEs*, Ann. Appl. Probab., 29 (2019), pp. 3489–3540, https://doi.org/10.1214/19-AAP1486, https://doi.org/10.1214/19-AAP1486.

A Appendix

We use the following version of the Kolmogorov continuity theorem, giving the precise upper bound constant.

Theorem A.1 (Kolmogorov continuity theorem). *Let $(X_t)_{t\in[0,T]}$ be a $\mathbb{R}^d$-valued random process and assume that for some $p, \epsilon > 0$,*

$$\mathbb{E}[|X_t - X_s|^p] \leq C_0 |t-s|^{1+\varepsilon}, \quad t, s \in [0,T].$$

Then there exists a modification $\widetilde{X}$ of X which is α-Hölder continuous for every $\alpha \in (0, \varepsilon/p)$ and with

$$\mathbb{E}\left[\left(\sup_{t,s\in[0,T],\ t\neq s} \frac{|\widetilde{X}_t - \widetilde{X}_s|}{|t-s|^{\alpha}}\right)^p\right] \leq C_0 \left(\frac{2^{1+\alpha}}{1-2^{-\alpha}}\right)^p \frac{T^{1+\varepsilon-\alpha p}}{1-2^{-(\varepsilon-\alpha p)}}.$$

Proof. We refer to the proof of [29, Theorem 2.9, Lemma 2.10], with an immediate adaptation if we do not assume $T = 1$. □