Weak second order multi-revolution composition methods for highly oscillatory stochastic differential equations with additive or multiplicative noise

Gilles Vilmart*

April 15, 2014†

Abstract

We introduce a class of numerical methods for highly oscillatory systems of stochastic differential equations with general noncommutative noise. We prove global weak error bounds of order two uniformly with respect to the stiffness of the oscillations, which permits to use large time steps. The approach is based on the micro-macro framework of multi-revolution composition methods recently introduced for deterministic problems and inherits its geometric features, in particular to design integrators preserving exactly quadratic first integral. Numerical experiments, including the stochastic nonlinear Schrödinger equation with space-time multiplicative noise, illustrate the performance and versatility of the approach.

Keywords: highly-oscillatory stochastic differential equation, composition method, quadratic first integral conservation, multiplicative noise, time-dependent stochastic Schrödinger equation.

MSC numbers: 60H10, 35Q55.

1 Introduction

Consider a nonlinear system of (Itô[1]) stochastic differential equations (SDEs) of the form

$$dX(t) = (\varepsilon^{-1}AX(t) + f(X(t)))dt + \sum_{r=1}^{m} g^r(X(t))dW_r(t), \qquad 0 \le t \le T, \tag{1}$$

where $\varepsilon > 0$ is fixed, $A \in \mathbb{R}^{d\times d}$ is a constant matrix, $f, g^r : \mathbb{R}^d \to \mathbb{R}^d$ are smooth and Lipschitz vector fields, $X(0) = X_0$ is the initial condition, and $W_r, r = 1, \dots, m$ are independent standard Wiener processes. On the highly oscillatory part of the problem, we assume that

$$\exp(A) = I. \tag{2}$$

*Université de Genève, Section de mathématiques, 2-4 rue du Lièvre, CP 64, 1211 Genève 4, Switzerland. On leave from École Normale Supérieure de Rennes, INRIA Rennes, IRMAR, CNRS, UEB, av. Robert Schuman, F-35170 Bruz, France, Gilles.Vilmart@ens-rennes.fr

†A typographical error was corrected in Theorem 3.3 (Dec. 2014).

[1]We focus on Itô SDEs to simplify the presentation, but our analysis applies also to Stratonovich SDEs using the conversion formula.

This means that the flow of the stiff oscillatory part $\frac{dX}{dt} = \varepsilon^{-1}AX$, given by $x \mapsto \exp(\varepsilon^{-1}tA)x$, is periodic with respect to t, with period ε. Thus, multiple oscillatory frequencies in $\varepsilon^{-1}A$ are allowed if they are integer multiples of 2π. Such class of problems includes second order SDEs of the form

$$\ddot{X}(t) = -\varepsilon^{-2}K^2X(t) + a(X(t)) + \sum_{r=1}^{m} b^r(X(t))\dot{W}_r(t) \tag{3}$$

where $K \in \mathbb{R}^{d\times d}$ is a constant positive symmetric matrix with eigenvalues in $2\pi\mathbb{N}$ and $a, b^r : \mathbb{R}^d \to \mathbb{R}^d$ are smooth vector fields. An interesting situation is the case where $a = -\nabla U$ derives from a potential $U : \mathbb{R}^d \to \mathbb{R}$ and where additive noise is considered, i.e. the functions g^r are constant and there exists a constant matrix $B \in \mathbb{R}^{d\times m}$ such that

$$(g^1(q), \dots, g^m(q)) = B. \tag{4}$$

Consider the Hamiltonian, which represents the energy of the system (3),

$$H(p, q) = \frac{1}{2}\left(p^T p + \varepsilon^{-2}q^T K^2 q + U(q)\right).$$

A standard application of the Itô formula to (3) yields that the average of Hamiltonian grows linearly with time due to the additive noise perturbation. Precisely, setting $Q(t) = X(t)$ and $P(t) = \dot{X}(t)$, we have

$$\mathbb{E}(H(P(t), Q(t))) = H(P(0), Q(0)) + \frac{t}{2}\mathrm{trace}(BB^T) \tag{5}$$

and this energy is exactly conserved along time only in the deterministic case ($B = 0$). The linear growth (5) is not recovered in general by standard explicit integrators, e.g. the Euler-Maruyama method where a super linear growth can be observed [10, 11].

The class of problems (1) also includes spectral spatial discretizations of the nonlinear Schrödinger equation with a stochastic space-time noise. Consider for instance

$$i\frac{\partial u(x,t)}{\partial t} = -\Delta u(x,t) + \varepsilon V(x)|u(x,t)|^{2q}u(x,t) + \sqrt{\varepsilon}g(u(x,t))\dot{W}(x,t), \quad t \geq 0, x \in (0,1), \tag{6}$$

where we consider periodic boundary conditions in dimension one of space for simplicity. Here, $W(x,t)$ denotes a real-valued white noise which is white in time and correlated in space.[2] We refer for more details to [15] where numerical simulations of the stochastic Schrödinger equation are presented to investigate the influence of space-time noise over the stochastic blow up time of the solutions, and to [13] where the strong and weak convergence rates of the Euler-Maryama method applied to a class of stochastic nonlinear Schrödinger equations are shown to be the same as for SDEs in finite dimension (i.e. weak order 1) in contrast to the parabolic case. In the case of multiplicative noise in the Stratonovitch sense where $g(u) = \sigma u$ (with σ a constant), we have that the L^2 norm $\int_0^1 |u(x,t)|^2 dx$ is a first integral similarly to the deterministic: it is exactly conserved along time almost surely on the existence interval of the solution. The Hamiltonian energy is however not conserved in general in the presence of noise. Considering a pseudo-spectral spatial representation of the form $u(t,x) \approx \sum_{k=-\ell+1}^{\ell} \xi_k(t)e^{ikx}$ with a fixed number ℓ of Fourier modes, we then

[2] Here, the correlation in space arises from the fixed number ℓ of Fourier modes considered in the noise.

arrive at the following system of SDEs with Stratonovitch noise for the Fourier modes $\xi_k(t) \in \mathbb{C},\ -\ell < k \le \ell$,

$$d\xi_k = \big(-ik^2\xi_k + \varepsilon\, f_k(\ldots,\xi_{-1},\xi_0,\xi_1,\ldots)\big)dt + i\sigma\sqrt{\varepsilon}\sum_{\ell<j\le\ell} \xi_{k-j}\circ dW_j, \tag{7}$$

where f_k corresponds to the nonlinear term $V|u|^{2q}u$, W_j $(-\ell < j \le \ell)$ are independent standard Wiener processes, and we set $\xi_{j\pm 2\ell} = \xi_j$ for all j in the above sum (the convolution product of the solution frequencies and the Wiener processes). The Stratonovitch system of SDEs (7) can be recast into the format (1) by rescaling time (that is, by rewriting the system in terms of the new time variable $\hat{t} = \frac{2\pi}{\varepsilon}t$) and using the Stratonovitch to Itô conversion formula.

For the numerical approximation of highly oscillatory problems of the form (1), standard SDE integrators require in general a time stepsize of the same magnitude as ε to achieve stability and accuracy, which can be prohibitively expansive for small values of ε, already in the deterministic setting ($g^r = 0, r = 1,\ldots,m$ in (1)), and thus appropriate integrators are needed. For the numerical study of highly oscillatory stochastic integrators and not assuming (2), we mention the recent work [34], where splitting methods for the Langevin equations are analyzed, and the works [10, 12] where stochastic trigonometric methods are proposed and analyzed, and further extended to the linear stochastic wave equation [11]. The trigonometric methods are a modification of the classical leap-frog method for (3) where filter functions are introduced to avoid numerical resonances (see [21, Chap. XIII] in the context of deterministic problems). A common aspect of the aforementioned works is the study of the strong convergence (with orders 1/2 or 1), i.e. the error $\mathbb{E}(|X(t_k)-X_k|)$ (where $t_k = kh$ and h is the stepsize) for approximating individual trajectories $X(t)$ themselves in the case of additive noise (i.e. (4) holds). In contrast, we consider here the general case of a general nonlinear noncommutative noise, and we focus on the weak error of convergence, i.e. the error $|\mathbb{E}(\phi(X(t_k))) - \mathbb{E}(\phi(X_k))|$ where ϕ is a smooth test function.

In [7], the class of deterministic multi-revolution composition methods was recently introduced for systems of the form (1) with $g^r = 0, r = 1,\ldots,m$. Observe that the exact flow of this problem after one period ε is a smooth perturbation of the identity. The multi-revolution approach, as first proposed in [23, 27], permits the approximation of the Nth iterate of a smooth deterministic near identity map $\varphi_\varepsilon(y) = y+\mathcal{O}(\varepsilon)$ from $\mathbb{R}^d$ into itself, but calculating only a few compositions (hence the name multi-revolution). Setting $H := N\varepsilon$, a example of such composition method is the order two approximation of the Nth iterate of the map φ_ε,

$$\varphi_{(H+\varepsilon)/2}\circ\varphi^{-1}_{-(H-\varepsilon)/2}(y) = \varphi^N_\varepsilon(y) + \mathcal{O}(H^3).$$

It is shown in [7] that the constant symbolized by $\mathcal{O}$ is independent of $N \ge 1$ and $\varepsilon \le \varepsilon_0$ if $\varphi_\varepsilon(y)$ is a C^3 function of (y,ε). Next, considering for φ_ε the flow of an highly oscillatory system after one highly oscillatory period permits to design large time step geometric integrators. The aim of this paper is to extend and analyze this approach to the stochastic context with general nonlinear multi-dimensional drift and diffusion functions.

It is known (see [29] and the recent work [9]) in a deterministic context ($g^r = 0$) that the solution $X(t)$ of (1) is asymptotically close to an effective nonstiff problem of the form

$$\frac{d\overline{X}(t)}{dt} = \big(a_1 + \varepsilon a_2 + +\varepsilon^2 a_3 + \ldots\big)\big(\overline{X}(t)\big), \qquad \overline{X}(0) = X_0 \tag{8}$$

where

$$
\begin{aligned}
a_1(x) &= \int_0^1 e^{sA} f(e^{-sA}x)ds \\
a_2(x) &= -\frac{1}{2}\int_0^1\int_0^{s_2}\Big(e^{s_2A} f(e^{(s_1-s_2)A} f(e^{-s_1A}x)) - e^{s_1A} f(e^{(s_2-s_1)A} f(e^{-s_2A}x))\Big)ds_1ds_2, \\
&\dots
\end{aligned} \tag{9}
$$

at times which are integer multiples of the oscillatory period. Precisely, truncating the above series (8) (which diverges for general nonlinear problems) after the ε^{p-1} term yields $X(t) = \overline{X}(t) + \mathcal{O}(\varepsilon^p)$ for all $t = k\varepsilon \le T$ with $k \in \mathbb{N}$, and this remainder can be made exponentially small for analytic data. We observe that the vector field of the effective problem (8) involving multiple integrals can be difficult to simulate in general. Although an analogous effective problem can be constructed in some cases (see Remark 3.5 for additive noise), we highlight that the proposed methods do not involve the pre-calculation of such an effective problem but applies directly to the original problem (1), with micro and macro stepsizes in the spirit of the so-called Heterogeneous Multiscale Method [2, 16, 17].

This paper is organized as follows. In Section 2, we introduce the stochastic multi-revolution composition methods and state our main weak convergence results. In Section 3, we perform the error analysis of the proposed methods for general nonlinear systems of SDEs. In Section 4, we present several numerical examples and comparisons with other oscillatory integrators that corroborate the theoretical orders of convergence and illustrate the good qualitative behavior of the proposed methods for various problems over long times, including the stochastic nonlinear Schrödinger equation with multiplicative noise.

2 Multi-revolution composition methods for oscillatory SDEs

We first introduce semi-discrete multi-revolution methods involving a macro time stepsize H which can be possibly much larger than the oscillatory period ε. We next introduce the fully-discrete methods by coupling the multi-revolution approach with a micro integrator involving in addition a micro stepsize h. We present weak error estimates with respect to H, h with error constants independent of ε. The proofs are postponed to Section 3.

2.1 Semi-discrete multi-revolution composition methods

We introduce the following integrator which permits to integrate (1) with oscillatory period ε on a time interval of size $\mathcal{O}(1)$ at a computational cost independent of ε by considering appropriate auxiliary non-stiff SDE problems.

Algorithm 2.1 (Semi-discrete S-MRCM)**.** *For the approximation of the flow after time $H = N\varepsilon$ of (1) where $N \in \mathbb{N}$, we consider the scheme $X_k \mapsto X_{k+1}$ defined by*

$$
dK_{k,1} = (-AK_{k,1} + \alpha_N H f(K_{k,1}))dt + \sqrt{\alpha_N H}\sum_{r=1}^m g^r(K_{k,1})dW_{k,r,1},\ K_{k,1}(0) = X_k \tag{10}
$$

$$
dK_{k,2} = (AK_{k,2} + \beta_N H f(K_{k,2}))dt + \sqrt{\beta_N H}\sum_{r=1}^m g^r(K_{k,2})dW_{k,r,2},\ K_{k,2}(0) = K_{k,1}(1) \tag{11}
$$

$$
X_{k+1} = K_{k,2}(1)
$$

where $W_{k,r,1}, W_{k,r,2}, r = 1, \dots, m$ are independent Wiener processes, and we define

$$\alpha_N := \frac{1}{2} - \frac{1}{2N}, \qquad \beta_N := \frac{1}{2} + \frac{1}{2N}. \tag{12}$$

For notational brevity, we shall sometimes drop the dependence on k of $K_{k,j}$ and $W_{k,r,j}$. The advantage of the above Algorithm 2.1 is that the stiff problem (1) on a time interval $H = N\varepsilon$ including N fast oscillations (due to the term $\varepsilon^{-1}AX$ in (1)) is approached by the resolution of two non-stiff problems posed on the time interval $(0, 1)$ involving each only one oscillation, and thus more convenient to solve. In other words, in this multi-revolution approach, N fast oscillations of (1) are approached by only two oscillations of analogous problems with appropriate coefficients. This is advantageous compared to standard integrators for large values of N, i.e. for small values of ε.

Remark 2.2. *In the context of deterministic oscillatory problems, order conditions for MRCMs up to arbitrary high order are derived in [7] using the algebraic framework of labelled rooted trees, and integrators up to order 4 are exhibited. However, it is known [30, 31] that a composition method with real coefficients of order strictly larger than 2 necessarily involves negative coefficients, see [4] for an elegant geometric proof. Since stochastic problems such as (1) are not reversible in time, such high order methods cannot straightforwardly be applied. A possibility to circumvent this order barrier for non-reversible problems would be to use complex coefficients, as proposed in [5, 22] in the context of deterministic diffusion problems. The use of complex coefficients in this context is however beyond the scope of the present paper.*

The next task is to prove accuracy estimates between the exact solution $X(t)$ of (1) and the approximation X_k from Algorithm 2.1 where $t = kH$ and $H = N\varepsilon$ with error constants independent of $H \leq H_0$ and $\varepsilon \leq \varepsilon_0$. We make the following smoothness assumption of the data.

(H) The functions $f, g^r, r = 1, \dots, m$ are C^6-functions with all partial derivatives bounded.

This implies that f, g^r are Lipschitz continuous (but not necessarily bounded), and thus the solution of (1) exists and is unique. In order to state our weak error estimates, we denote $C^p_P(\mathbb{R}^d, \mathbb{R})$ the set of functions of class C^p where all the partial derivatives have a polynomial growth, i.e. for each partial derivative ϕ up to order p, there exist $C > 0$ and $r \in \mathbb{N}$ such that

$$|\phi(x)| \leq C(1 + |x|^r). \tag{13}$$

We shall prove in Section 3 the following second order weak error estimate for the semi-discrete methods.

Theorem 2.3. *Let $T > 0$. Assume (2) and **(H)**. Consider X_k the numerical solution of Algorithm 2.1 and $X(t)$ the exact solution of (1). Then, for all $\phi \in C^6_P(\mathbb{R}^d, \mathbb{R})$, and all $H = N\varepsilon$ with $N \in \mathbb{N}$, and $k \in \mathbb{N}$ with $kH \leq T$,*

$$|\mathbb{E}(\phi(X_k)) - \mathbb{E}(\phi(X(kH)))| \leq CH^2,$$

where the constant C is independent of ε, H, k, N.

We highlight that the weak accuracy estimate of Theorem 2.3 holds uniformly with respect to ε where both $\varepsilon \ll 1$ and $\varepsilon \simeq 1$ are allowed. Observe in addition that Algorithm 2.1 is exact for $N = 1$ (i.e. the error is zero) because $K_{k,1}(1) = X_k$ in (10) and (11) reduces to a time transformation $\hat{t} = t/\varepsilon$ of (1).

Remark 2.4. *Notice that setting $\alpha_N = 0, \beta_N = 1$ in (12) in the definition of Algorithm 2.1 would yield the following multi-revolution method $X_k \mapsto X_{k+1}$ of weak order 1,*

$$dK_k = (AK_k + Hf(K_k))dt + \sqrt{H}\sum_{r=1}^{m} g^r(K_k)dW_{k,r},\ K_k(0) = X_k,$$
$$X_{k+1} = K_k(1).$$

Indeed, the error estimate of Theorem 2.3 with H^2 replaced by H remains valid for the above scheme for all $\phi \in C_P^4(\mathbb{R}^d, \mathbb{R})$ (observe in this case $K_{k,2}(0) = K_{k,1}(1) = X_k$). In this paper, we shall however focus on the more accurate weak second order methods. The error analysis of the first order method can be performed analogously.

2.2 Fully-discrete multi-revolution composition methods

Algorithm 2.1 is called semi-discrete because each step requires the resolution of two systems of SDEs, whose solution is not known in general and needs to be approximated numerically. It was noted in the deterministic context [7] that in principle, any nonstiff integrator could be used. However, a natural choice for such approximation is to use a splitting method where the oscillatory and non oscillatory parts of the problem are solved separately in an efficient way (sometimes exactly), as proposed and studied recently in [34] in the stochastic context of the Langevin equation.

We now formulate the fully-discrete multi-revolution stochastic methods for the class of problems (1). It involves a micro stepsize h and a macro stepsize H where $H \gg \varepsilon$ is allowed. We highlight again that both H and h are of moderate size, independently of the smallness of the stiff parameter ε. The approach involves a weak integrator $Y_{k+1} = \Phi_h(Y_k)$ with stepsize h for the nonstiff SDE problem

$$dY(t) = f(Y(t))dt + \sum_{r=1}^{m} g^r(Y(t))dW_r(t) \tag{14}$$

where compared to (1), the stiff term $\varepsilon^{-1}A$ has been removed. The integrator Φ_h can be the exact solution if computationally available, or an efficient weak approximation.

Algorithm 2.5 (Fully-discrete S-MRCM)**.** *Consider a macro stepsize $H = N\varepsilon$ and a micro stepsize $h = 1/n$ with $N, n \in \mathbb{N}^*$. For the approximation of the flow after time H of (1), we consider the scheme $X_{k+1} \mapsto X_k$ defined by the composition*

$$X_{k+1} = (e^{hA/2} \circ \Phi_{\beta_N Hh} \circ e^{hA/2})^n \circ (e^{-hA/2} \circ \Phi_{\alpha_N Hh} \circ e^{-hA/2})^n (X_k)$$

where the micro integrator Φ_h is a weak integrator for (14) with stepsize h, and α_N, β_N are defined in (12). The exponents n indicate that the maps are composed n times with independent random variables.

Notice that Algorithm 2.5 requires for each time step $2n$ applications of the nonstiff integrator Φ_h, applied with independent random variables, and $2n$ evaluations of exponentials (using $e^{hA/2} \circ e^{hA/2} = e^{hA}$). The following theorem with proof postponed to Section 3 states that it has weak second order of accuracy with respect to the micro and macro stepsizes H, h, uniformly with respect to ε. To this aim, we assume that the integrator Φ_h satisfies for all $x \in \mathbb{R}^d$, and all $h \leq h_0$,

$$|\mathbb{E}(Y_{k+1} - Y_k | Y_k = x)| \leq C(1 + |x|)h, \qquad |Y_{k+1} - Y_k| \leq M_k(1 + |Y_k|)\sqrt{h}, \tag{15}$$

where C, M_n are independent of h, x and M_k is a random variable with finite moments of all orders. We further assume the following local weak order two estimate, for all $\phi \in C_P^6(\mathbb{R}^d, \mathbb{R})$,

$$|\mathbb{E}(\phi(\Phi_h(Y_0))) - \mathbb{E}(\phi(Y(h)))| \leq C(x)h^3, \tag{16}$$

where $C(x)$ is independent h and has a polynomial growth (13).

Theorem 2.6. *Let $T > 0$. Assume* (2) *and* ***(H)****. Consider X_k the numerical solution of Algorithm 2.5 and $X(t)$ the exact solution of* (1)*. Assume that the integrator Φ_h for* (14) *satisfies* (15) *and* (16)*. Then, for all $\phi \in C_P^6(\mathbb{R}^d, \mathbb{R})$, and all $h = 1/n$ and $H = N\varepsilon$ small enough with $n, N, k \in \mathbb{N}$ with $kH \leq T$,*

$$|\mathbb{E}(\phi(X_k)) - \mathbb{E}(\phi(X(kH)))| \leq C(H^2 + h^2), \tag{17}$$

where C is independent of $\varepsilon, H, n, k, N, h$.

The proofs of the above weak convergence estimates (Theorem 2.3 and Theorem 2.6) are provided in Section 3.

Remark 2.7. *We highlight that* (15) *and* (16) *are natural assumptions for a weak order two integrator. Indeed, a classical theorem of Milstein [24] (see [25, Chap. 2.2]) permits to deduce from the local error* (16) *the global error $|\mathbb{E}(\phi(Y_k)) - \mathbb{E}(\phi(Y(hk)))| \leq Ch^2$ for all $hk \leq T$. Crucial is the assumption* (15)*, which is easily satisfied by any reasonable integrator for* (14)*, and that automatically yield that Y_k has bounded moments of any order for all $kh \leq T$ (see [25, Lemma 2.2, p. 102]). Examples of such integrators with additional favorable geometric properties are presented in Section 4.*

We end this section with the following remark, which shown that the multi-revolution approach has links with weak methods for systems of SDEs with small noise of the form

$$dZ(s) = a(Z(s))dt + \varepsilon b(Z(s))dt + \sqrt{|\varepsilon|}\sum_{r=1}^{m} c^r(Z(s))dW_r(s), \quad Z(0) = Z_0 \tag{18}$$

as proposed and analyzed in [26] (see also [25, Chap. 3]). Such schemes applied to (18) have global weak order $\mathcal{O}(h^p + \varepsilon h^q)$ where $p < q$ on bounded time intervals, using appropriate smoothness assumptions on the vector fields a, b, c^r.

Remark 2.8. *Setting $a(Z) = AZ, b(Z) = f(Z), \sigma^r = g^r, Z_0 = X_0$, the system* (18) *is equivalent to* (1) *via the time transformation $s = t/\varepsilon$. In this case, the multi-revolution approach permits to approximate $Z(s)$ on longer time intervals with $s = \mathcal{O}(\varepsilon^{-1})$ with a computational cost and an accuracy both independent of ε. Indeed, following the lines of the proof of Theorem 2.6, if $\Psi_{h,\varepsilon}$ is an integrator for* (18) *with weak order $\mathcal{O}(h^p +$*

εh^q), then, provided the numerical moments remain uniformly bounded, the multi-revolution composition scheme

$$X_{k+1} = (\Psi_{h,\beta_N H})^n \circ (\Psi_{-h,-\alpha_N H})^n (X_k),$$

with α_N, β_N defined in (12) *and $h = 1/n$, can be shown to approximate the solution $X(t)$ of* (1) *at time $t = kH$ with $H = N\varepsilon$ with global weak error $\mathcal{O}(H^2 + H^{-1}h^p + h^q)$ for all $kH \leq T$ (equivalently to approximate* (18) *at time $s = kN \leq \varepsilon^{-1}T$), where the constant in $\mathcal{O}$ is independent of H, h, ε.*

3 Weak convergence analysis

A crucial ingredient for the analysis is the variation of constant formula: the solution $X(t)$ of (1) satisfies

$$X(t) = e^{t\varepsilon^{-1}A}X_0 + \int_0^t e^{(t-s)\varepsilon^{-1}A} f(X(s))ds + \int_0^t e^{(t-s)\varepsilon^{-1}A} \sum_{r=1}^m g^r(X(s))dW_r(s). \quad (19)$$

In other words, considering the change of variables $Y(t) = e^{-\varepsilon^{-1}tA}X(t)$, we have that $Y(t)$ is the solution of the non-autonomous SDE problem

$$dY(t) = e^{-\varepsilon^{-1}tA} f(e^{\varepsilon^{-1}tA}Y(t))dt + \sum_{r=1}^m e^{-\varepsilon^{-1}tA} g^r(e^{\varepsilon^{-1}tA}Y(t))dW_r(t). \quad (20)$$

A feature of the semi-discrete Algorithm 2.1 is that it is exact in the case of additive noise (4) and in the absence of the nonlinearity ($f = 0$), as stated in the following proposition. This is a consequence of the homogeneity and scaling in time properties of Wiener processes. Notice in contrast that the stochastic trigonometric methods [10, 12] are not exact in this case.

Proposition 3.1. *Consider the numerical solution of* (1) *with $f = 0$ and additive noise* (4)*. Then, Algorithm 2.1 is exact in the sense that $X(t)$, $t = kN\varepsilon$ and X_k have the same law of probability for all ε and all $N, k, \in \mathbb{N}$.*

Proof. The solutions of (10) and (11) can be expressed using the variation of constant formula (19), which yields, using $e^A = e^{-A} = I$,

$$X_1 = X_0 + \sqrt{H} \int_0^1 e^{-sA} B d\tilde{W}(s)$$

where we notice that $\tilde{W}(s) := \sqrt{\alpha_N}W_1(s) + \sqrt{\beta_N}W_2(s)$ is a standard m-dimensional Wiener process because $\alpha_N + \beta_N = 1$ (recall that W_1 and W_2 are assumed independent). In comparison, using (19) and the change of variable $\hat{s} = \varepsilon^{-1}s$, and the periodicity of $t \mapsto e^{tA}$ with period 1, the exact solution of (1) with $f = 0$ and (4) satisfies

$$X(H) = X_0 + \int_0^N e^{-sA} B dW(\varepsilon s) = X_0 + \int_0^1 e^{-sA} B \sum_{k=0}^{N-1} dW(\varepsilon s + k).$$

Using standard homogeneity and scaling in time properties of the Wiener process, we have that $H^{-1}\sum_{k=0}^{N-1}(W(\varepsilon s + k) - W(k))$ is also a standard Wiener process. Using the independence of $W_{k,1}, W_{k,2}, k = 0, 1, 2, \ldots$ in Algorithm 2.1, we obtain that $X(H)$ has the same law of probability as X_1, and we conclude the proof that $X(t_k)$ and X_k have the same law of probability by induction on k. □

Using (2), we observe that for integer multiples of the oscillatory period, the solutions of (1) and (20) coincide: $X(N\varepsilon) = Y(N\varepsilon)$ for all $N \in \mathbb{N}$ and all ε. The advantage of considering the form (20) compared to (1) is that the drift and diffusion functions $(x,t) \mapsto e^{-\varepsilon^{-1}tA} f(e^{\varepsilon^{-1}tA} Y(t))$ and $(x,t) \mapsto e^{-\varepsilon^{-1}tA} g^r(e^{\varepsilon^{-1}tA} Y(t))$ are C^6-functions with all partial derivatives with respect to the spatial variable x bounded uniformly with respect to ε. Using these regularity, we may recall the following two classical results taken from [33].

Lemma 3.2. *[33, Lemma 1]. Assume **(H)** and consider the solution $Y(t)$ of (20). Then, there exist a constant $C > 0$ and $r \in \mathbb{N}$ such that for all $p \in \mathbb{N}, t \in [0,T], X_0 \in \mathbb{R}^d$,*

$$\mathbb{E}(|X(t)|^p) \leq C(1 + |X_0|^r).$$

In addition, there exists a version of the process $X(t)$ which is almost surely 6 times continuously differentiable with respect to the initial condition $X(0) = X_0$, and the derivatives with respect to the initial condition, $(\partial^k X(t))/(\partial X_0^k), k = 1, \ldots, 6$, have bounded moments uniformly with respect to $\varepsilon \in \mathbb{R}$, $t \in [0,T], X_0 \in \mathbb{R}^d$.

Theorem 3.3. *[33, Theorem 2.2]. Assume **(H)** and consider the solution $Y(t)$ of (20). For $T > 0$ and all $\phi \in C_P^6(\mathbb{R}^d, \mathbb{R})$, the function $v(x,t) := \mathbb{E}(\phi(Y(t))|Y(T) = x)$, $0 \leq t \leq T$, is solution of the Backward Kolmogorov partial differential equation*

$$\frac{\partial v}{\partial t} + \mathcal{L}_\varepsilon(-t)v = 0, \qquad v(x,T) = \phi(x),$$

where the generator of (20) is defined by[3]

$$\begin{aligned} \mathcal{L}_\varepsilon(-t)\phi(x) &= e^{-\varepsilon^{-1}tA} f(e^{\varepsilon^{-1}tA}x) \cdot \nabla\phi(x) \\ &+ \frac{1}{2}\sum_{r=1}^{m} \phi''(x)\left(e^{-\varepsilon^{-1}tA} g^r(e^{\varepsilon^{-1}tA}x), e^{-\varepsilon^{-1}tA} g^r(e^{\varepsilon^{-1}tA}x)\right) \end{aligned} \tag{21}$$

and the partial derivatives $\partial_t^{i_0}\partial_{x_{i_1}} \cdots \partial_{x_{i_k}} v(x,t)$ with $2i_0 + k \leq 6$, $1 \leq i_j \leq d$, $i_0, k \geq 0$, are continuous with polynomial growth (13) where r, C are independent of ε, $x \in \mathbb{R}^d$, and $t \in [0,T]$. In particular,

$$u(x,t) := \mathbb{E}(\phi(Y(t))|Y(0) = x) \tag{22}$$

is equal to $v(x, T-t)$, and assuming $T/\varepsilon \in \mathbb{N}$ and (2), we obtain that $u(x,t)$ in (22) is the solution of

$$\frac{\partial u}{\partial t} = \mathcal{L}_\varepsilon(t)u, \qquad u(x,0) = \phi(x), \tag{23}$$

for all $t \in [0,T]$.

We deduce from Theorem 3.3 a weak Taylor expansion for $u(x,H) = \mathbb{E}(\phi(X(H)))$ where $X(t)$ is the solution of (1).

Proposition 3.4. *Assume the hypotheses of Theorem 3.3. Then, $u(x,t)$ defined in (22) satisfies for all $H = N\varepsilon$, with $N \in \mathbb{N}$,*

$$u(x,H) = \phi(x) + H\int_0^1 \mathcal{L}_1(s_1)\phi(x)ds_1 + \frac{H^2 - H\varepsilon}{2}\int_0^1\int_0^1 \mathcal{L}_1(s_1)\mathcal{L}_1(s_2)\phi(x)ds_2ds_1$$

[3]We denote $\nabla\phi(x)$ the gradient with respect to x of ϕ and $\phi''(x)(\cdot,\cdot)$ the second derivative of ϕ, which is a symmetric bilinear form.

$$+ \quad H\varepsilon \int_0^1 \int_0^{s_1} \mathcal{L}_1(s_1)\mathcal{L}_1(s_2)\phi(x)ds_2ds_1 + \mathcal{O}(H^3)$$

where the constant in $\mathcal{O}(H^3)$ is independent of N, ε with a polynomial growth (13) *with respect to x, and $\mathcal{L}_1$ is defined in* (21) *with $\varepsilon = 1$.*

Proof. Iterating "á la Picard" the integral relation $u(x,t) = \phi(x) + \int_0^t \mathcal{L}_\varepsilon(s)u(x,s)ds$, we obtain

$$\begin{aligned} u(x,t) &= \phi(x) + \int_0^t \mathcal{L}_\varepsilon(s_1)\phi(x)ds_1 + \int_0^t \int_0^{s_1} \mathcal{L}_\varepsilon(s_1)\mathcal{L}_\varepsilon(s_2)\phi(x)ds_2ds_1 \\ &+ \int_0^t \int_0^{s_1} \int_0^{s_2} \mathcal{L}_\varepsilon(s_1)\mathcal{L}_\varepsilon(s_2)\mathcal{L}_\varepsilon(s_3)u(x,s_3)ds_3ds_2ds_1 \end{aligned}$$

which yields using Theorem 3.3,

$$u(x,H) = \phi(x) + \int_0^H \mathcal{L}_\varepsilon(s_1)\phi(x)ds_1 + \int_0^H \int_0^{s_1} \mathcal{L}_\varepsilon(s_1)\mathcal{L}_\varepsilon(s_2)\phi(x)ds_2ds_1 + \mathcal{O}(H^3)$$

where the constant in $\mathcal{O}(H^3)$ satisfies the polynomial growth (13) uniformly with respect to N, ε. Using the change of variable $\hat{s}_i = \varepsilon^{-1}s_i$, we deduce

$$u(x,H) = \phi(x) + \varepsilon \int_0^N \mathcal{L}_1(s_1)\phi(x)ds_1 + \varepsilon^2 \int_0^N \int_0^{s_1} \mathcal{L}_1(s_1)\mathcal{L}_1(s_2)\phi(x)ds_2ds_1 + \mathcal{O}(H^3).$$

Finally, using the periodicity assumption (2) yields the identities

$$\begin{aligned} \int_0^N \mathcal{L}_1(s_1)\phi(x)ds_1 &= N\int_0^1 \mathcal{L}_1(s_1)\phi(x)ds_1, \\ \int_0^N \int_0^{s_1} \mathcal{L}_1(s_1)\mathcal{L}_1(s_2)\phi(x)ds_2ds_1 &= \frac{N(N-1)}{2}\int_0^1 \int_0^1 \mathcal{L}_1(s_1)\mathcal{L}_1(s_2)\phi(x)ds_2ds_1 \\ &+ N\int_0^1 \int_0^{s_1} \mathcal{L}_1(s_1)\mathcal{L}_1(s_2)\phi(x)ds_2ds_1, \end{aligned}$$

which permits to conclude the proof. □

A consequence of Proposition 3.4 is the following remark, which gives some insight on a stochastic effective problem in the case of additive noise, analogously to (8).

Remark 3.5. *Consider the additive case* (4) *with dimensions $d = m$ and $B = I$. It can be checked that for all $t_k = k\varepsilon \leq T$ with $k \in \mathbb{N}$, and all $\phi \in C_P^6(\mathbb{R}^d, \mathbb{R})$,*

$$|\mathbb{E}(\phi(X(t_k))) - \mathbb{E}(\phi(\overline{X}(t_k)))| \leq C\varepsilon^2,$$

where C is independent of N, ε, where $\overline{X}(t)$ solves the effective SDE

$$d\overline{X} = \Big(a_1 + \varepsilon(a_2 + \sigma^2\Delta b_2)\Big)(\overline{X})dt + \sigma(I + \varepsilon b_2'(\overline{X}))dW(t), \qquad \overline{X}(0) = X_0 \tag{24}$$

where a_1, a_2 are defined in (9) *and*

$$b_2(x) = \frac{1}{4}\int_0^1 \Big(s_1 e^{s_1 A} f(e^{-s_1 A}x) - \int_0^{s_1} e^{s_2 A} f(e^{-s_2 A}x)d_{s_2}\Big)ds_1.$$

The proof of (24) *is deduced showing first the local estimate* $|\mathbb{E}(\phi(X(\varepsilon))) - \mathbb{E}(\phi(\overline{X}(\varepsilon)))| \leq C\varepsilon^3$ *(using Proposition 3.4 with* $N = 1$ *for* $X(\varepsilon)$ *and analogously a weak Taylor expansion for* $\overline{X}(\varepsilon)$*). Interestingly, the effective SDE* (24) *has multiplicative noise in general, although the original SDE* (1) *with* $B = I$ *in* (4) *has additive noise.*

We may now derive a local weak error estimate for the semi-discrete S-MRCM.

Lemma 3.6. *Assume* (2) *and* ***(H)****. Consider* X_1 *the numerical solution of Algorithm 2.1 after one step and* $X(t)$ *the exact solution of* (1)*. Then, for all initial condition* $X_0 = x$*, all* $\phi \in C_P^6(\mathbb{R}^d, \mathbb{R})$*, and all* $H = N\varepsilon$ *with* $N \in \mathbb{N}$*,*

$$|\mathbb{E}(\phi(X_1)) - \mathbb{E}(\phi(X(H)))| \leq C(x)H^3,$$

where the constant $C(x)$ *is independent of* ε, H *and has a polynomial growth* (13)*.*

Proof. We observe that the SDE (10) has the form (1) by setting $\varepsilon = -1$ and replacing f by $\alpha_N H f$ and g^r by $\sqrt{\alpha_N H} g^r$. Applying Proposition 3.4 with $N = 1$, we deduce

$$\begin{aligned}\mathbb{E}(\phi(K_1(1))|K_1(0) = x) &= \phi(x) + \alpha_N H \int_0^1 \mathcal{L}_{-1}(s_1)\phi(x)ds_1 \\ &+ (\alpha_N H)^2 \int_0^1 \int_0^{s_1} \mathcal{L}_{-1}(s_1)\mathcal{L}_{-1}(s_2)\phi(x)ds_2ds_1 + \mathcal{O}(H^3). \quad (25)\end{aligned}$$

Analogously for the solution $K_2(t)$ of (11), we have

$$\begin{aligned}\mathbb{E}(\phi(K_2(1))|K_2(0) = x) &= \phi(x) + \beta_N H \int_0^1 \mathcal{L}_1(s_1)\phi(x)ds_1 \\ &+ (\beta_N H)^2 \int_0^1 \int_0^{s_1} \mathcal{L}_1(s_1)\mathcal{L}_1(s_2)\phi(x)ds_2ds_1 + \mathcal{O}(H^3), \quad (26)\end{aligned}$$

where the constants in the above remainders $\mathcal{O}(H^3)$ are independent of N, ε and have a polynomial growth (13) with respect to x.

We next have by the Fubini theorem and using properties of conditional expectancies (recall that W_1, W_2 in (10),(11) are independent Wiener processes),

$$\mathbb{E}(\phi(X_1)|X_0 = x) = \mathbb{E}^{W_1}\Big(\mathbb{E}^{W_2}\Big(\phi(K_2(1))|K_2(0) = K_1(1)\Big)|K_1(0) = x\Big) \quad (27)$$

where the notations $\mathbb{E}^{W_1}, \mathbb{E}^{W_2}$ refers to the expectation with respect to the Wiener process W_1, W_2, respectively. Applying (25) with ϕ replaced by $x = K_2(0) \mapsto \mathbb{E}^{W_2}(\phi(K_2(1)))$, which is, by Theorem 3.3, almost surely of class C^6 with derivatives of polynomial growth, we deduce for all $X_0 = x$ the estimate[4]

$$\begin{aligned}\mathbb{E}(\phi(X_1)) &= \Big(I + \alpha_N H \int_0^1 \mathcal{L}_{-1}(s_1)ds_1 + (\alpha_N H)^2 \int_0^1 \int_0^{s_1} \mathcal{L}_{-1}(s_1)\mathcal{L}_{-1}(s_2)ds_2ds_1\Big) \\ &\circ\Big(I + \beta_N H \int_0^1 \mathcal{L}_1(s_1)ds_1 + (\beta_N H)^2 \int_0^1 \int_0^{s_1} \mathcal{L}_1(s_1)\mathcal{L}_1(s_2)ds_2ds_1\Big)\phi(x) + \mathcal{O}(H^3)\end{aligned}$$

[4]Observe that the ordering of composition is the opposite compared to the ordering (10),(11) in Algorithm 2.5. This effect is known as the "Vertauschungssatz" of Gröbner in the deterministic literature, see e.g. [21, Sect. III.5.1].

$$
\begin{aligned}
&= \Big(I + \alpha_N H \int_0^1 \mathcal{L}_{-1}(s_1)ds_1 + \beta_N H \int_0^1 \mathcal{L}_1(s_1)ds_1 \\
&+ \alpha_N\beta_N H^2 \int_0^1 \mathcal{L}_{-1}(s_1)ds_1 \int_0^1 \mathcal{L}_1(s_1)ds_1 \\
&+ (\alpha_N H)^2 \int_0^1 \int_0^{s_1} \mathcal{L}_{-1}(s_1)\mathcal{L}_{-1}(s_2)ds_2ds_1 \\
&+ (\beta_N H)^2 \int_0^1 \int_0^{s_1} \mathcal{L}_1(s_1)\mathcal{L}_1(s_2)ds_2ds_1\Big)\phi(x) + \mathcal{O}(H^3).
\end{aligned}
$$

A consequence of (2) is $\int_0^1 \mathcal{L}_{-1}(s_1)ds_1 = \int_0^1 \mathcal{L}_1(s_1)ds_1$ and

$$
\int_0^1\int_0^{s_1}\mathcal{L}_{-1}(s_1)\mathcal{L}_{-1}(s_2)ds_2ds_1 = \int_0^1\int_0^1\mathcal{L}_1(s_1)\mathcal{L}_1(s_2)ds_2ds_1 - \int_0^1\int_0^{s_1}\mathcal{L}_1(s_1)\mathcal{L}_1(s_2)ds_2ds_1.
$$

We deduce

$$
\begin{aligned}
\mathbb{E}(\phi(X_1)) &= \Big(I + (\alpha_N + \beta_N)H \int_0^1 \mathcal{L}_1(s_1)ds_1 \\
&+ (\alpha_N\beta_N + \alpha_N^2)H^2 \int_0^1 \int_0^1 \mathcal{L}_1(s_1)\mathcal{L}_1(s_2)ds_2ds_1 \\
&+ (\beta_N^2 - \alpha_N^2)H^2 \int_0^1 \int_0^{s_1} \mathcal{L}_1(s_1)\mathcal{L}_1(s_2)ds_2ds_1\Big)\phi(X_0) + \mathcal{O}(H^3).
\end{aligned}
$$

Using (12), we observe that $\alpha_N + \beta_N = 1$, $(\alpha_N\beta_N + \alpha_N^2)H^2 = (H^2 - H\varepsilon)/2$ and $(\beta_N^2 - \alpha_N^2)H^2 = H\varepsilon$. Comparing the above Taylor expansion with the one in Proposition 3.4 permits to conclude the proof. □

Based on the local error estimate of Lemma 3.6, we may now give the proof of Theorem 2.3 for the global order two of convergence in the weak sense. To this aim, the following lemma, which is shown in the proof of [25, Lemma 2.2, p. 102] is a crucial ingredient.

Lemma 3.7. *[25] Consider a discrete process $\{Y_k\}$ ($k \in \mathbb{N}$) satisfying* (15)*. Assume further that Y_0 has finite moments of all orders. Then, for all integer p, there exists C_p such that*

$$
\mathbb{E}(|Y_k|^p) \leq e^{C_p kh}(\mathbb{E}(|Y_0|^p) + 1)
$$

for all $k \in \mathbb{N}$ and $h \leq h_0$.

Proof of Theorem 2.3. We use a well known result of Milstein [24] (see [25, Chap. 2.2]) which permits to deduce automatically the global error estimate of Theorem 2.3 from the local error estimate of Lemma 3.6. To this aim, we only have to check is that the moments $\mathbb{E}(|X_k|^{2r})$ of the numerical solution are bounded for all k, H with $0 \leq kH \leq T$ uniformly with respect to k and H sufficiently small. This is a consequence of Lemma 3.7 applied with $h = H$ to the discrete process $\{Y_k\}$ defined by $Y_{2k} = X_k$ and $Y_{2k+1} = K_{k,1}(1)$ where $K_{k,1}$ is given in (10). Considering the SDEs (10) and (11), the estimates (15) with $h = H$ for the discrete flows $K_{k,1}(0) \mapsto K_{k,1}(1)$ and $K_{k,2}(0) \mapsto K_{k,2}(1)$ are a straightforward consequence of the Lischitzness of $f, g^r, r = 1, \ldots, m$. This permits to conclude the proof. □

In order to prove Theorem 2.6 for the the global weak accuracy of the fully-discrete method, we need to check that the numerical solution has uniformly bounded moments of any order. This is the purpose of the following lemma which shows that the numerical moments are uniformly bounded in spite of the oscillatory terms $e^{\pm hA/2}$ involved in the scheme.

Lemma 3.8. *Assume the hypotheses of Theorem 2.6 and consider the numerical solution X_k of Algorithm 2.5. Then, for all $p \in \mathbb{N}$, there exists a constant C_p such that for all $H = N\varepsilon$, $k \in \mathbb{N}$ with $kH \leq T$,*

$$\mathbb{E}(|X_k|^p) \leq C_p.$$

Proof. We observe that the numerical solution X_{k+1} is calculated inductively from X_k using a composition of $2n$ stochastic mappings of the form

$$e^{\delta hA/2} \circ \Phi_{\gamma hH} \circ e^{\delta hA/2}$$

where $\delta = -1, \gamma = \alpha_N$ or $\delta = 1, \gamma = \beta_N$. Let $\{Y_\kappa\}, \kappa = 0, 1, 2, \ldots$ denote the discrete process arising from this composition (note that $X_k = Y_{2nk}$). We next define the discrete process $\widetilde{Y}_\kappa$ from Y_κ by

$$\widetilde{Y}_{2nk+j} = e^{jhA} Y_{2nk+j}, \quad \widetilde{Y}_{2nk+n+j} = e^{-jhA} Y_{2nk+n+j}, \quad , j = 1, \ldots, n.$$

For $1 \leq j \leq n$, we obtain $\widetilde{Y}_{2nk+j+1} = e^{h(j+1/2)A} \circ \Phi_{\alpha_N hH} \circ e^{-h(j+1/2)A}(\widetilde{Y}_{2nk+j})$, which yields

$$\widetilde{Y}_{2nk+j+1} - \widetilde{Y}_{2nk+j} = e^{h(j+1/2)A}(\Phi_{\alpha_N hH}(e^{-h(j+1/2)A}\widetilde{Y}_{2nk+j}) - e^{-h(j+1/2)A}\widetilde{Y}_{2nk+j}).$$

Analogously, for $n+1 \leq j \leq 2n$, the above identity holds with A replaced by $-A$ and α_N replaced by β_N. Using the estimates (15) on Φ_h, the inequality $|e^{\lambda A}\xi| \leq c|\xi|$ for all $\lambda \in \mathbb{R}, \xi \in \mathbb{R}^d$, and the bounds $\alpha_N \leq 1, \beta_N \leq 1$ we deduce for all κ,

$$|\mathbb{E}(\widetilde{Y}_{\kappa+1} - \widetilde{Y}_\kappa | \widetilde{Y}_\kappa = x)| \leq C(1 + |x|)Hh, \qquad |\widetilde{Y}_{\kappa+1} - \widetilde{Y}_\kappa| \leq M_\kappa(1 + |\widetilde{Y}_\kappa|)\sqrt{Hh},$$

where C is independent of $N, n, H, h, \varepsilon, \kappa, k$. Applying Lemma 3.7 then yields that $\widetilde{Y}_\kappa$ has uniformly bounded moments of any order for $\kappa Hh \leq 2T$. We conclude the proof using $h = 1/n$ and $X_k = \widetilde{Y}_{2nk}$ for all k. □

We shall also need the following result, whose proof follows standard arguments. For the sake of completeness, a proof is provided in Appendix.

Lemma 3.9. *Assume the hypotheses of Theorem 2.6. Let $\delta = -1, \gamma = \alpha_N$ (resp. $\delta = 1, \gamma = \beta_N$). Then, the integrator*

$$Y_{k+1} = e^{\delta hA/2} \circ \Phi_{\gamma Hh} \circ e^{\delta hA/2}(Y_k),$$

applied the SDE (10) *(respectively* (11)*) satisfies for all $\phi \in C_P^6(\mathbb{R}^d, \mathbb{R})$ and all $H, h = 1/n$ small enough,*

$$|\mathbb{E}(\phi(Y_n)) - \mathbb{E}(\phi(K_j(1)))| \leq C(x)h^2 H,$$

where $j = 1$ (resp. $j = 2$) and $C(x)$ is independent of $h, H, N, n, k, \varepsilon$ and satisfies (13).

Proof of Theorem 2.6. Consider the numerical solution denoted $\hat{X}_k$ of the semi-discrete Algorithm 2.1. Applying Lemma 3.9 for (10) and (11), we deduce

$$|\mathbb{E}(\phi(\hat{X}_1)) - \mathbb{E}(\phi(X_1))| \leq C(x)h^2 H.$$

Using Lemma 3.6, we deduce the local error estimate

$$|\mathbb{E}(\phi(X_1)) - \mathbb{E}(\phi(X(H)))| \leq C(x)(H^3 + h^2 H). \tag{28}$$

In addition, by Lemma 3.8, the numerical solution X_k as uniformly bounded moments of any order, and the Milstein theorem [24] (see [25, Chap. 2.2]) then yields the global error estimate of Theorem 2.6. □

4 Numerical experiments

In this final section, we consider four problems to illustrate numerically the weak convergence rates of the proposed methods and the versatility of the approach. We investigate not only the accuracy of the methods compared to other oscillatory integrators but also their long time qualitative behavior. We also compare our approach with other known highly oscillatory integrators.

4.1 An illustrative example: Kubo oscillators

We first focus on the so-called Kubo oscillators. A nonlinear version with multiplicative Stratonovitch noise, as considered recently in [10], is

$$\begin{aligned} dQ(t) &= (-\varepsilon^{-1}P(t) + Pf(P(t),Q(t)))dt + \sigma P(t) \circ dW(t) \\ dP(t) &= (\varepsilon^{-1}Q(t) - Qf(P(t),Q(t)))dt - \sigma Q(t) \circ dW(t) \end{aligned} \tag{29}$$

where $P(t), Q(t) \in \mathbb{R}$ and the same one-dimensional Wiener process is considered in (29) for both components. Notice that (29) can be put in the form (1) using the time transformation $\hat{t} = t/2\pi$ and the Stratonovitch to Itô conversion formula. For the nonstiff part of the system (29), we choose as micro integrator a Strang splitting method

$$\Phi_h = \Phi^V_{h/2} \circ \Phi^W_h \circ \Phi^V_{h/2}, \tag{30}$$

where, on the one hand, the noise part $dQ = \sigma P \circ dW,\ dP = -\sigma Q \circ dW$ is integrated with

$$\Phi^V_h : \begin{pmatrix} q \\ p \end{pmatrix} \mapsto \begin{pmatrix} Q \\ P \end{pmatrix} = \begin{pmatrix} \cos(\sigma\sqrt{h}\xi) & \sin(\sigma\sqrt{h}\xi) \\ -\sin(\sigma\sqrt{h}\xi) & \cos(\sigma\sqrt{h}\xi) \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix}$$

where ξ are independent random variables with $\mathbb{P}(\xi = \pm\sqrt{3}) = 1/6$, $\mathbb{P}(\xi = 0) = 2/3$. Notice that replacing above $\sqrt{h}\xi$ by $\Delta W_k = W((n+1)k) - W(nk)$ would make Φ^V_h exact. The above choice of discrete random variable makes Φ^V_h a weak second order integrator (using $\mathbb{E}(\xi^2) = 1, \mathbb{E}(\xi^4) = 3$). On the other hand, the nonlinear part $dQ = \sigma P f(P,Q)dt,\ dP = -\sigma Q f(P,Q)dt$ is approximated by the implicit midpoint rule defined as

$$\Phi^W_h : \begin{pmatrix} q \\ p \end{pmatrix} \mapsto \begin{pmatrix} Q \\ P \end{pmatrix} = \begin{pmatrix} q + hP_M f(P_M, Q_M) \\ p - hQ_M f(P_M, Q_M) \end{pmatrix}$$

where $P_M = (p+P)/2, Q_M = (q+Q)/2$. In the implementation of Φ^W_h, we use fixed point iterations until convergence up to round-off errors. Notice that the splitting scheme (30) has weak second order of accuracy (see Proposition 6.1 in Appendix with $\varepsilon = 1$).

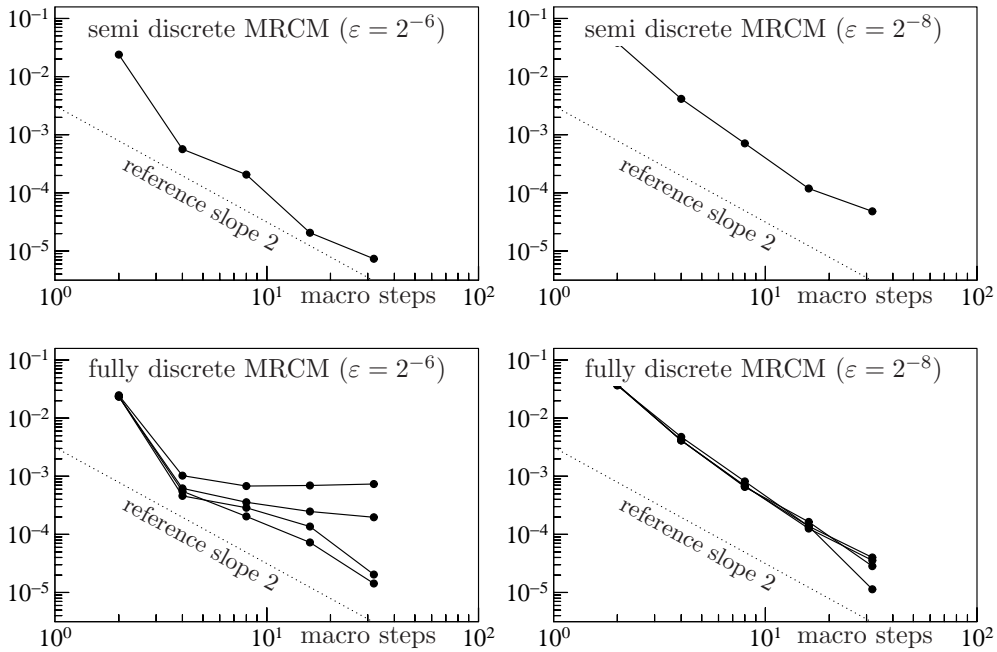

Figure 1: Multi-revolution methods for the Kubo oscillator (29) with nonlinearity $f(p,q) = p^3 + q^5$. Error in $\mathbb{E}(Q(T)^2)$ versus number of macro steps (final time $T = 2\pi$). Top pictures: semi-discrete S-MRCM. Bottom pictures: fully-discrete S-MRCM with $n = 8, 16, 32, 64$ micro steps per macro step (respectively from top to bottom).

Conservation of quadratic first integrals An interesting feature of (29) is that the quantity $C(p,q) = p^2 + q^2$ is exactly conserved along time for all trajectories. Precisely, observing that $d\big(C(P(t),Q(t))\big) = 0$ (as shown for instance in [10, Prop. 3.2]), we have almost surely

$$P(t)^2 + Q(t)^2 = P(0)^2 + Q(0)^2, \qquad \text{for all } t > 0. \tag{31}$$

Since both integrators Φ_h^V and Φ_h^W exactly conserve the quadratic first integral, we have that Φ_h in (30) exactly conserves $P(t)^2 + Q(t)^2$, and thus also the corresponding fully-discrete S-MRCM of Algorithm 2.5: $P_k^2 + Q_k^2 = P_0^2 + Q_0^2$ for all $k \in \mathbb{N}$ and all numerical trajectories of the method. Note in contrast that the strong method proposed in [10] does not conserve exactly this quadratic first integral in general.

Weak convergence rates Since the scheme exactly conserves the quadratic first integral P^2+Q^2, we have that the weak error bound (17) holds provided that $f : \mathbb{R}^2 \to \mathbb{R}$ in (29) is of class C^6 in a neighbourhood of $\{(p,q)\ ;\ p^2+q^2 = P_0+Q_0\}$ where we consider a deterministic initial condition (here $Q_0 = 1, P_0 = 0$). We consider the nonlinearity $f(p,q) = p^3 + q^5$ (similarly to [10]). For $\varepsilon = 2^{-6} \simeq 1.6 \cdot 10^{-2}$ (left pictures) and $\varepsilon = 2^{-8} \simeq 3.9 \cdot 10^{-3}$ (right pictures), we plot for many different macro steps H the error for the second moment of the first component $\mathbb{E}(|Q(T)|^2)$ at the final time $T = 2\pi$ as a function of the number of macro steps T/H. The expectation is approximated using the average over $M = 10^7$ trajectories to make the Monte-Carlo error sufficiently small compared to the weak accuracy of the methods. In the top pictures, we take $n = 1024$ micro steps in each macro step, so that the micro discretization can be considered as nearly exact (see semi-discrete Algorithm

2.1). In the bottom pictures, the four lines correspond respectively to $n = 8, 16, 32, 64$ micro steps per macro step (from top to bottom lines). In the top pictures, we observe the expected lines of slope 2, as proved in the semi-discrete error analysis of Theorem 2.3. In the bottom pictures, we observe the expected lines of slope 2 only for a sufficiently fine micro stepsize, as shown in Theorem 2.6. As a reference solution, we consider here the standard Strang splitting $e^{\varepsilon^{-1}hA/2} \circ \Phi_h \circ e^{\varepsilon^{-1}hA/2}$ with Φ_h defined in (30) with small stepsize $h = 2^{-15}T \simeq 1.9 \cdot 10^{-4}$.

4.2 A test problem with non-commutative noise

In some situations (e.g. the stochastic nonlinear Schrödinger model (7) studied in Section 4.4), the exact solution of the non oscillatory system (14) is available and easy to compute. Otherwise a weak second order approximation is needed in general. Notice that already for nonstiff stochastic problems and without structure assumptions (i.e. for a non-commutative noise), methods with high strong order are generally costly to simulate because they involve multiple stochastic integrals, such as $\int_{t_0}^{h} \int_0^{s_2} dW_q(s_1) dW_r(s_2)$, which are costly to approximate strongly for $q \neq r$. This is however not true for weak methods where such multiple stochastic integrals can be approximated efficiently in a weak sense using appropriate discrete random variables. The aim of this section is to illustrate that this is again not a difficulty in our highly oscillatory context. We consider the following nonlinear oscillatory problem in dimension $d = 2$, which is a modification of a scalar test SDE from [14], with a non-commutative Itô noise with $m = 10$ independent driving Wiener processes,

$$\begin{aligned} dQ(t) &= -\varepsilon^{-1} P(t) dt, \quad Q(0) = 1, \\ dP(t) &= \varepsilon^{-1} Q(t) dt + \sum_{j=1}^{10} a_j^{-1} \sqrt{P(t)^2 + Q(t)^2 + b_j^{-1}(1 - Q(t))}\, dW_j(t), \quad P(0) = 0. \end{aligned} \tag{32}$$

where $P(t), Q(t) \in \mathbb{R}$. The values of the constants $a_j, j = 1, \ldots, 10$ are respectively $5, 5, 10, 15, 30, 15, 10, 5, 10, 15$, and the values of $b_j, j = 1, \ldots, 10$ are respectively $4, 3, 5, 2, 1, 2, 4, 5, 10, 10$. Considering the averaged oscillatory energy $E = \mathbb{E}\left(P^2 + Q^2\right)$, an application of the Itô formula yields $\frac{dE(t)}{dt} = \sum_{j=1}^{10} a_j^{-2}(E(t) + b_j^{-1}(1 - \cos(t/\varepsilon)))$, which can be solved analytically as

$$\mathbb{E}\left(P(t)^2 + Q(t)^2\right) = e^{at} + b(a + a^3\varepsilon^2)^{-1}\left(e^{at} + a^2\varepsilon^2 \cos(t/\varepsilon) - a\varepsilon \sin(t/\varepsilon) - a^2\varepsilon^2 - 1\right),$$

where $a = \sum_{j=1}^{10} a_j^{-2} = 37/225$, $b = \sum_{j=1}^{10} a_j^{-2} b_j^{-1} = 257/6000$. We shall use this formula to check the accuracy of the fully-discrete S-MRCM (Algorithm 2.5).

For the integrator $X_1 = \Phi_h(X_0)$ needed to integrate the non-stiff part of problem (32), which has the form (14) with $f = 0$, we compare two different schemes which fulfil the assumptions of Theorem 2.6. On the one hand we simply use the Euler-Maruyama method of weak order 1,

$$X_1 = X_0 + \sqrt{h} \sum_{r=1}^{m} g^r(X_0) \xi_r,$$

where ξ_r are independent discrete random variables satisfying $\mathbb{P}(\xi_r = \pm\sqrt{3}) = 1/6$, $\mathbb{P}(\xi_r = 0) = 2/3$. On the other hand, we consider the following Runge-Kutta type scheme of weak order two, which is a derivative free version of the Milstein-Talay method [32] derived in

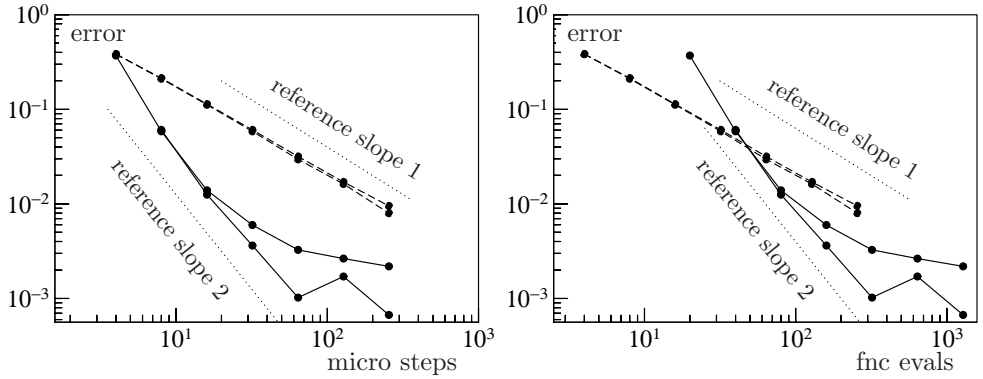

Figure 2: Multi-revolution methods for the test problem (32) with non-commutative noise with $\varepsilon = 2^{-8}$. Error in the averaged energy $\mathbb{E}(P(T)^2 + Q(T)^2)$ at $T = 2\pi$ versus number of micro steps (left picture) or the number of diffusion function evaluations (right picture). Solid lines: method (33) with weak order 2 for the nonstiff part. Dashed lines: Euler-Maruyama method with weak order 1 for the nonstiff part. We use in both cases the multi-revolution parameter $N = 128$ and $N = 256$ (from top to bottom).

[3, Lemma 3.1] using an idea in [28] to make the number of evaluations of each diffusion function $g^r, r = 1, \ldots, m$ independent of the dimensions d, m,

$$\begin{aligned} X_1 &= X_0 + \frac{1}{2}\sum_{r=1}^{m}\Big(g^r\big(X_0 + \sum_{q=1}^{m} g^q(X_0)J_{q,r}\big) - g^r\big(X_0 - \sum_{q=1}^{m} g^q(X_0)J_{q,r}\big)\Big) \\ &+ \frac{\sqrt{h}}{2}\sum_{r=1}^{m}\Big(g^r(X_0 + \sqrt{\frac{h}{2}}\sum_{q=1}^{m} g^q(X_0)\chi_q) + g(X_0 - \sqrt{\frac{h}{2}}\sum_{q=1}^{m} g^q(X_0)\chi_q)\Big)\xi_r, \end{aligned} \tag{33}$$

where the quantity $J_{q,r}$ is a weak approximation of the multiple stochastic integral $\int_0^h \int_0^{s_2} dW_q(s_1)dW_r(s_2)$ given by [25, p. 96, eq. (1.25)]

$$J_{q,r} = \begin{cases} h(\xi_r\xi_r - 1)/2, & \text{if } q = r, \\ h(\xi_q\xi_r - \chi_q)/2, & \text{if } r < q, \\ h(\xi_q\xi_r + \chi_r)/2, & \text{if } r > q, \end{cases}$$

and $\chi_l, \xi_l,\ l = 1 \ldots m$ are independent discrete random variables satisfying $\mathbb{P}(\chi_l = \pm 1) = 1/2, \mathbb{P}(\xi_l = \pm\sqrt{3}) = 1/6, \mathbb{P}(\xi_l = 0) = 2/3$.

In Figure 2, we consider Algorithm 2.5 applied to problem (32) with final time $T = 2\pi$, and oscillatory frequency $\varepsilon = 2^{-8}$. We plot the error in the averaged energy $\mathbb{E}(P(T)^2 + Q(T)^2) \simeq 3.2816345$ as a function of the total number of micro steps (left picture) or as a function of the number of evaluations of the diffusion functions $g^r, r = 1, \ldots, m$ (right picture), taking into account that method (33) requires 5 diffusion function evaluations per step in contrast to the Euler-Maruyama method with a single evaluation of the diffusion functions per step. Since these evaluations dominate the cost of the method for a large dimension m, this is a natural measure of the computational cost. The averages are obtained using 10^8 independent trajectories. We use successively the multi-revolution parameters $N = 128$ and $N = 256$, and obtain in Figure 2 the expected lines of slope 2 and 1, except for large numbers of micro steps where the macro error (and also the Monte-Carlo error) is not negligible compared to the micro error. This confirms the error analysis in Theorem 2.6.

We observe in the right picture of Figure 2 that the S-MRCM using a weak second order nonstiff method is more efficient than the S-MRCM using the Euler-Maruyama method in the case where high accuracy is needed (finer than one percent in this example).

4.3 A modification of the Fermi-Pasta-Ulam problem with additive noise

We next consider a problem inspired from [21], which is a single-frequency modification of the Fermi-Pasta-Ulam problem often used to test methods for highly-oscillatory problems. It corresponds to a second order Hamiltonian system of the form (3) with $d = 6$ degrees of freedom and additive noise (4),

$$dQ = \nabla_P H(P,Q)dt, \qquad dP = -\nabla_Q H(P,Q) + \sigma B dW(t). \tag{34}$$

Its Hamiltonian function is defined by

$$E(p,q) = \frac{1}{2}\sum_{i=1}^{6} p_i^2 + \frac{1}{2\varepsilon^2}\sum_{i=4}^{6} q_i^2 + V(q), \tag{35}$$

with the quartic interaction potential

$$V(q) = \frac{1}{4}((q_1-q_4)^4 + (q_2-q_5-q_1-q_4)^4 + (q_3-q_6-q_2-q_5)^4 + (q_3+q_6)^4),$$

where $\varepsilon > 0$ is a small parameter. Again, following [7] in the deterministic setting, this problem can be put in the form (1) up to a time transformation $\hat{t} = t/2\pi$, where $A \in \mathbb{R}^{12\times 12}$ is given by

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\pi\eta I \\ 0 & 0 & 0 & 0 \\ 0 & -2\pi\eta^{-1} I & 0 & 0 \end{pmatrix},$$

where $\eta = \varepsilon$ is fixed, and 0, I denote respectively zero and identity matrices of size 3×3. This model describes the motion of a chain composed with three soft nonlinear springs and three stiff linear springs (see [21] in the deterministic context). The components $i = 1,2,3$ are associated to a slow motion, while the components $i = 4,5,6$ oscillate rapidly at the frequency ε^{-1}. We consider the initial condition $Q(0) = (1,0,0,\varepsilon,0,0)^T$, $P(0) = (1,0,0,1,0,0)^T$, and similarly to [12], we consider an additive noise perturbation (4) in dimension $m = 1$ with

$$B = \sigma(0,1,1/2,5,0,1/100)^T. \tag{36}$$

Remark 4.1. *Our stochastic error analysis is performed for simplicity for a constant matrix A independent of ε. However, analogously to the deterministic case in [7], it can be extended to the present situation where A depends on $\eta = \varepsilon$. Indeed, since a, b^r in (3) are independent of the highly oscillatory velocities, we have that the unbounded factor η^{-1} involved in A never appears in $e^{tA}f(e^{-tA}y)$ and $e^{tA}g^r(e^{-tA}y)$ where f, g^r are the vector fields in the formulation (20) of the problem.*

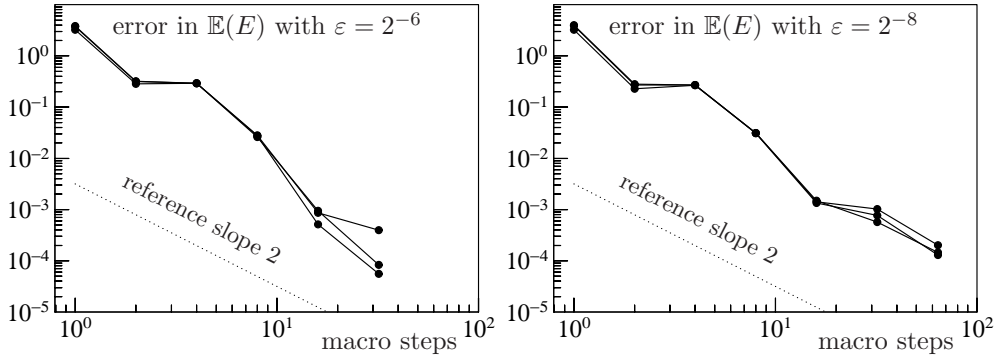

Figure 3: Multi-revolution methods for the FPU like problem (34) with Hamiltonian (35). Error in the Hamiltonian $\mathbb{E}(E(P(T), Q(T)))$ (from top to bottom pictures, respectively) versus the number of macro steps (final time $T = 2\pi$). The lines corresponds to the fully-discrete S-MRCM with $n = 8, 16, 32$ micro steps per macro step, respectively.

Weak convergence rates We emphasize that our convergence analysis applies only to Lipschitz vector fields, which is not the case of problem (34)-(35). However, numerical experiments still exhibit the high weak order two of convergence of the method predicted in the Lipschitz case. We consider the fully-discrete Algorithm 2.5 where the micro integrator is defined as a Strang Splitting (30) where Φ_h^V integrates exactly the deterministic problem with Hamiltonian $\frac{1}{2}\sum_{i=1}^{3} p_i^2$ and Φ_h^W integrates the position components (see [7] for details on the implementation). In Figure 3, we plot the convergence curves of the S-MRCM for average Hamiltonian energy $\mathbb{E}(E)$ at final time $T = 2\pi$ for $\varepsilon = 2^{-6}$ (left picture) and $\varepsilon = 2^{-8}$ (right picture), respectively. We plot the errors at time $T = 2\pi$ as a function of the number of macro steps T/H where the macro stepsize is $H = 2\pi N\varepsilon$ (taking into account the time transformation $\hat{t} = t/(2\pi)$). The 3 curves in each plot correspond respectively to $n = 8, 16, 32$ micro steps per macro steps (from top to bottom). We observe convergence curves with slope 2, which corroborate the fully-discrete error analysis of Theorem 2.6. The expectations are obtained as the averages over 10^8 trajectories of the methods, and the reference solution is computed using a standard Strang splitting with stepsize $h = 2^{-16}T \simeq 10^{-4}$.

Long time behavior Although our error analysis applies only to bounded time intervals with respect to ε, we investigate in Figure 4 the long time behavior of the S-MRCM on long time intervals of size $T = 2\pi\varepsilon^{-1}$ with $\varepsilon = 1/200$. We consider the evolution along time of the three stiff spring energies

$$I_j = \frac{1}{2}p_{3+j}^2 + \frac{1}{2\varepsilon^2}q_{3+j}^2, \qquad j = 1, 2, 3$$

the total oscillatory energy $I = I_1 + I_2 + I_3$ and the Hamiltonian energy in (35). In the deterministic case, it is classical (see e.g. [21]) that the Hamiltonian of the exact solution is exactly conserved, while $I(t) = I(0) + \mathcal{O}(\varepsilon)$ is nearly conserved along time and called an adiabatic invariant. In addition energy exchanges at the time scale ε^{-1} can be observed between the stiff spring energies I_1, I_2, I_3. This can be observed in Figure 4 (left pictures) for a reference solution (Strang splitting method with small stepsize $h = 10^{-3}$) and the MRCM of deterministic order 2 as first proposed in [7].

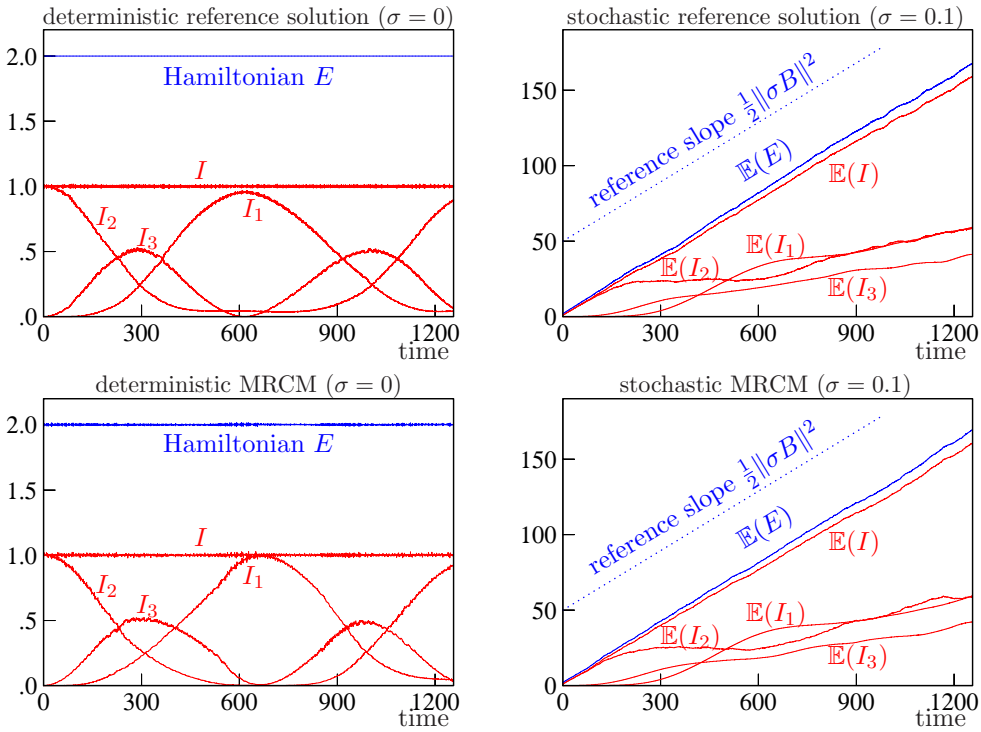

Figure 4: Energy exchanges in the FPU-like problem (34) with $\varepsilon = 1/200$ on the time interval $(0, 2\pi\varepsilon^{-1})$. Comparison of the S-MRCM ($N = 10$,$n = 8$) (bottom pictures) and a reference solution (Strang splitting with $h = 10^{-3}$, top pictures). Stochastic case ($\sigma = 0.1$): averages over 10^3 numerical trajectories.

In the additive noise case (where we take $\sigma = 0.1$ in (36)), there is still an average energy exchange between the stiff springs, as observed in [12], and a linear drift (5) of the average Hamiltonian energy with slope $\frac{1}{2}\|\sigma B\|^2$ is expected for the exact solution. We observe an excellent behavior of the proposed S-MRCM, where the expected energy linear drift with the expected slope (5) can be observed (compare the S-MRCM and reference solutions in right pictures). We took the macro parameter $N = 10$ and $n = 8$ micro steps per macro time steps. On the time interval $(0, 2\pi\varepsilon^{-1})$, this corresponds to $nN^{-1}\varepsilon^{-2} = 3.2 \cdot 10^4$ forces evaluations of the potential force ∇V for each numerical trajectory. In comparison, for the reference solution with $h \simeq 10^{-3}$ (Strang splitting), we make $2\pi\varepsilon^{-1}h^{-1} \simeq 1.3 \cdot 10^6$ force evaluations to achieve a satisfactory solution. In this experiment, all expectations are obtained using the averages over 10^3 trajectories. Notice that for a standard explicit integrator, such as the Euler-Maruyama method (not represented here), a super linear growth of the numerical Hamiltonian would be observed (see [12]), yielding a completely wrong qualitative behavior.

4.4 The stochastic nonlinear Schrödinger equation

Although our analysis applies only to finite dimensional systems of SDEs, we finally consider the SPDE model (6) of the nonlinear Schrödinger equation with a multiplicative space-time noise,

$$i\frac{\partial u(x,t)}{\partial t} = -\Delta u(x,t) + \varepsilon V(x)|u(x,t)|^2 u(x,t) + \sigma\sqrt{\varepsilon}u(x,t) \circ \dot{W}(x,t), \quad t > 0, \qquad (37)$$

$$u(x,0) \quad = \quad \cos x + \sin x,$$

where $V(x) = 2\cos(2x)$, $\sigma = 10^{-2}$ is a constant, and $W(x,t)$ is a space-time noise which is white in time and meant in the Stratonovitch sense and spatially correlated in space. Problem (37) is posed for simplicity in dimension one of space on the interval $(0,1)$ with periodic boundary conditions. In the deterministic setting ($\sigma = 0$) this problem is analyzed theoretically in [19] and considered in [6, 7] to illustrate highly oscillatory integrators, and it has a unique global solution in all Sobolev spaces H^s for arbitrary $s \geq 0$.

We consider the S-MRCM (Algorithm 2.5) applied to the spectral formulation (7), which yields to a system of SDEs with a multiplicative noise of Stratonovitch type in dimension 2ℓ. We observe that $y(t) = \exp\big(-it\alpha|y_0|^2 - i\beta W(t)\big)y_0$ is the the exact solution of the complex scalar Stratonovitch SDE $idy = \alpha|y|^2 y + \beta u \circ dW, y(0) = y_0$ for all real parameters α, β. It is thus natural to define the micro integrator Φ_h of the S-MRCM algorithm by $\Phi_h = \mathcal{F}_\ell \circ \psi_h \circ \mathcal{F}_\ell^{-1}$ where $\mathcal{F}_\ell$ denotes the Fast Fourier Transform and ψ_h is given by

$$\psi_h : (u_j)_{-\ell<j\leq\ell} \in \mathbb{C}^{2\ell} \mapsto \left(\exp\Big(-ihV(x_j)|u_j|^2 - i\sigma\frac{\sqrt{h}}{\sqrt{\Delta x}}\chi_j\Big)u_j \right)_{-\ell<j\leq\ell}$$

where $x_j = j\Delta x$, $\Delta x = 1/(2\ell)$ and χ_j are independent random variables with $\mathbb{P}(\chi_j = \pm\sqrt{3}) = 1/6$, $\mathbb{P}(\chi_j = 0) = 2/3$).

Features of the considered scheme are that it is explicit and the complexity with respect to the space dimension parameter ℓ reduces to $\mathcal{O}(\ell \log \ell)$ thanks to the Fast Fourier Transform algorithm. In addition, the mass $\sum_{\ell<j\leq\ell} |\xi_j(t)|^2$, which is a quadratic first integral of (7), is exactly conserved by the numerical scheme (up to round-off errors) because the method is a composition of mass preserving flows.

Convergence rates For $\varepsilon = 10^{-4}$ and $\ell = 128$ spatial Fourier modes we plot in Figure 5 the error in $\mathbb{E}(|\xi_1(T)|^2)$ and $\mathbb{E}(|\xi_3(T)|^2)$ (second moment of the first and third modes) at time $T = 2\pi\varepsilon^{-1}$ as a function of the number of macro steps, here $1, 2, 4, 8, 16, 20, 40$, respectively. The number of micro steps per macro step is $n = 100$, and we take the averages over 10^5 trajectories. Again we observe the expected lines of slope 2. As reference solution, we take the numerical solution with $N = 10$, which corresponds to 10^3 macro steps on the time interval $(0, T)$.

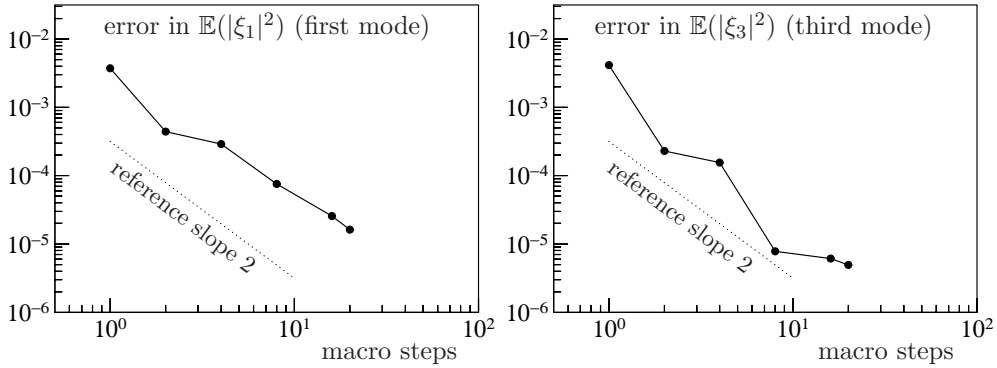

Figure 5: Stochastic nonlinear Schrödinger problem (37) with $\varepsilon = 10^{-4}$. Error of S-MRCM at time $T = 2\pi\varepsilon^{-1}$ for $\mathbb{E}(|\xi_1|^2)$ (first mode) and $\mathbb{E}(|\xi_3|^2)$ (third mode) as a function of the number of macro steps. Number of micro steps per macro step: $n = 100$.

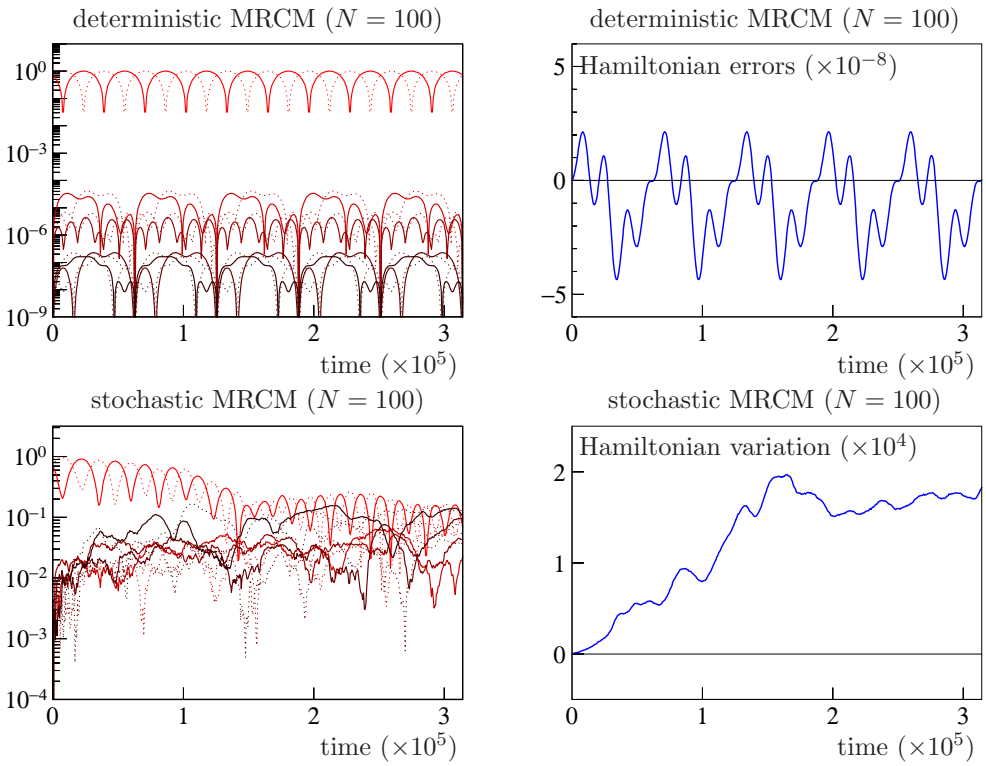

Figure 6: Nonlinear Schrödinger problem (37) with $\varepsilon = 10^{-4}$ on the time interval $(0, 2\pi\varepsilon^{-1})$. Top pictures: deterministic case ($\sigma = 0$). Bottom pictures: one sample trajectory in the stochastic case ($\sigma = 1$). Left pictures: Plot of the actions $|\xi_j(t)|$ along time t, for $j = 1, 3, 5, \ldots, 9$ (solid lines) and for $j = -1, -3, -5, \ldots, -9$ (dotted lines) with colors from red ($|j| = 1$) to black ($|j| = 9$). Right pictures: Hamiltonian energy along time (in blue).

Long time behavior Again for $\varepsilon = 10^{-4}$, we plot in Figure 6 the solution of S-MRCMs with parameter $N = 100$ for the first Fourier modes $|\xi_j(t)|$ for $|j| = 1, 3, 5, \ldots, 9$ (notice that $\xi_j(t) = 0$ for even indices j if $\sigma = 0$) along time in the interval $(0, 10\pi\varepsilon^{-1})$. In the top pictures, we consider the deterministic case ($\sigma = 0$) for which the Hamiltonian is exactly conserved by the exact solution (see the top right picture where the Hamiltonian error of the S-MRCM remains small, of size 10^{-8}). It was shown in [19] that a nonlinear beating effect occurs for the first mode of the exact solution

$$|\xi_1(t)|^2 = \frac{1 + \sin(2\varepsilon t)}{2} + \mathcal{O}(\varepsilon^{1/8}), \quad |\xi_{-1}(t)|^2 = \frac{1 - \sin(2\varepsilon t)}{2} + \mathcal{O}(\varepsilon^{1/8}), \quad t \leq \varepsilon^{-9/8}.$$

This behavior of the exact solution is well reproduced by the deterministic MRCM as reported in [7] (see top left picture). In the stochastic case of a multiplicative space-time noise ($\sigma = 0.1$), an analogous beating effect with period π/ε can still be observed in the first mode (see a sample trajectory in bottom pictures). Notice that in this stochastic context, the Hamiltonian energy is no longer a conserved quantity for the exact solution (see a sample trajectory in bottom right picture).

4.5 Comparison with other oscillatory integrators

In this section, we compare the efficiency of the proposed multi-revolution composition methods with other existing methods for the Kubo oscillator (Section 4.1) and the FPU type problem (Section 4.3). Precisely, we consider the following five methods:

- S-MRCM2: the multi-revolution methods of weak order two (Algorithm 2.5).
- S-MRCM1: the multi-revolution modification of weak order one discussed in Remark 2.4.
- SPLT2: the standard Strang splitting method between the highly oscillatory part and the nonstiff part, of weak order two,

$$X_{k+1} = e^{\varepsilon^{-1}hA/2} \circ \Phi^W_{h/2} \circ \Phi^V_h \circ \Phi^W_{h/2} \circ e^{\varepsilon^{-1}hA/2}(X_k) \tag{38}$$

  where Φ^V_h integrates the deterministic nonlinear part and Φ^W_h integrates the noise, as described in Sections 4.1 and 4.3 for problems (29), (34), respectively.

- TRIG: the implicit trigonometric method proposed and studied numerically in [10, eq. (19)] for the Kubo oscillator (29),

$$\begin{aligned} \begin{pmatrix} Q_{k+1} \\ P_{k+1} \end{pmatrix} &= e^{\varepsilon^{-1}hA} \left(\begin{pmatrix} Q_k \\ P_k \end{pmatrix} + \frac{h}{2} \begin{pmatrix} P_k f(P_k, Q_k) \\ -Q_k f(P_k, Q_k) \end{pmatrix} + \frac{\sigma}{2} \begin{pmatrix} P_k \\ -Q_k \end{pmatrix} \Delta W_k \right) \\ &+ \frac{h}{2} \begin{pmatrix} P_{k+1} f(P_{k+1}, Q_{k+1}) \\ -Q_{k+1} f(P_{k+1}, Q_{k+1}) \end{pmatrix} + \frac{\sigma}{2} \begin{pmatrix} P_{k+1} \\ -Q_{k+1} \end{pmatrix} \Delta W_k \end{aligned} \tag{39}$$

  where we have the identity $e^{\varepsilon^{-1}hA} = \begin{pmatrix} \cos(\varepsilon^{-1}h) & -\sin(\varepsilon^{-1}h) \\ \sin(\varepsilon^{-1}h) & \cos(\varepsilon^{-1}h) \end{pmatrix}$ and $\Delta W_k = W(t_{k+1}) - W(t_k)$ are independent centred Gaussian random vectors with variance hI.

- TRIG1: the explicit trigonometric method proposed in [10, 11] for second order SDEs with additive noise (3)-(4),

$$\begin{aligned} \begin{pmatrix} Q_{k+1} \\ P_{k+1} \end{pmatrix} &= \begin{pmatrix} \cos(\varepsilon^{-1}hK) & h\,\mathrm{sinc}(\varepsilon^{-1}hK) \\ -\varepsilon^{-1}K\sin(\varepsilon^{-1}hK) & \cos(\varepsilon^{-1}hK) \end{pmatrix} \begin{pmatrix} Q_k \\ P_k \end{pmatrix} \\ &+ \frac{h}{2} \begin{pmatrix} h\Psi a(\Phi Q_k) \\ \Psi_0 a(\Phi Q_k) + \Psi_1 a(\Phi Q_{k+1}) \end{pmatrix} + \begin{pmatrix} h\,\mathrm{sinc}(\varepsilon^{-1}hK) B \Delta W_k \\ \cos(\varepsilon^{-1}hK) B \Delta W_k \end{pmatrix} \end{aligned} \tag{40}$$

  where we denote the position and velocity vectors $X = Q$ and $\dot{X} = P$. Here, $\Phi, \Psi, \Psi_0, \Psi_1$ are filter matrices introduced to avoid numerical resonances arising already in the deterministic setting [21, Chap. XIII] and defined in [10, 11] following [20],

$$\Phi = \mathrm{sinc}(\xi), \quad \Psi = \mathrm{sinc}^3(\xi), \quad \Psi_0 = \cos(\xi)\,\mathrm{sinc}^2(\xi), \quad \Psi_1 = \mathrm{sinc}^2(\xi),$$

  with $\xi = \varepsilon^{-1}hK$ where we note that $\mathrm{sinc}(\xi) = \xi^{-1}\sin(\xi)$ is defined for possibly singular matrices ξ. We recall that the scheme (40) does not require the eigenvalues of the highly oscillatory matrix K to be integer multiples in contrast to the multi-revolution approach where (2) is assumed. In the analysis [11] of the stiff scheme (40), it is proved a strong convergence of order one for the position, $\mathbb{E}(\|Q_k - Q(t_k)\|^2)^{1/2} \leq Ch$ for all $kh \leq T$ (which implies weak order one) and $h \geq c\varepsilon$, with constants C, c that are independent of ε, h. A bound $\mathbb{E}(\|P_k - P(t_k)\|^2)^{1/2} \leq C$ is proved on the velocities.

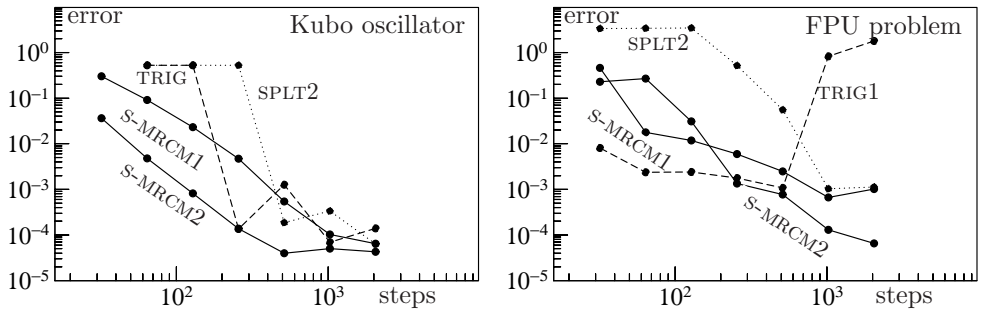

Figure 7: Precision/work diagrams. Weak errors versus the total number of steps for multi-revolution methods or orders 1 and 2 (solid lines), standard Strang splitting methods (38) (dotted lines), and trigonometric methods (dashed lines). Highly oscillatory frequency $\varepsilon = 2^{-8}$. Left picture: error in $\mathbb{E}(Q(T)^2)$ (second moment of the first component) for the Kubo oscillator (29). Right picture: error in $\mathbb{E}(E(T))$ (average energy) for the FPU problem (34). Fully-discrete S-MRCM with $n = 8$ micro steps per macro step.

In Figure 7, we consider the Kubo oscillator (29) and the FPU problem (34) as described in Sections 4.1 and 4.3 with highly oscillatory frequency $\varepsilon = 2^{-8}$. We plot the weak errors (in $\mathbb{E}(Q(T)^2)$ and $\mathbb{E}(E(T))$, respectively) versus the total number of steps for multi-revolution methods S-MRCM1 and S-MRCM2, the standard Strang splitting method SPLT2, and trigonometric methods from [10, 11] (method TRIG in (39) for the Kubo oscillator and method TRIG1 in (40) for the FPU problem). For the multi-revolution methods, the considered number of steps corresponds to the total number of micro-steps, namely $2nT/H$ for S-MRCM2 and nT/H for S-MRCM1 where T/H is the number of macro steps (we choose here $n = 8$ micro steps). Since the most expansive part of all integrators is the integration of the nonlinearity of the problems, this is a fair measure of the computational cost of the methods. The expectancies are computed as the averages over 10^7 independent trajectories.[5]

For both problems, we observe that the multi-revolution composition methods S-MRCM1 and S-MRCM2 exhibit lines of slopes 1 and 2 which corroborates the uniformly accurate weak convergence rates of Theorem 2.6 and Remark 2.4. Thus, any prescribed accuracy at a fixed time T can be achieved with a computation cost which is independent of the smallness of ε. In contrast, the standard splitting methods for both problems and the trigonometric method TRIG in (39) for the Kubo oscillator require small stepsizes $h \leq C\varepsilon$ for the methods to converge. This means that these standard schemes have a computation cost which grows as $\mathcal{O}(\varepsilon^{-1})$ as $\varepsilon \to 0$ which makes these nonstiff integrators inefficient for highly oscillatory problems. The stiff trigonometric method TRIG1 in (40) for the FPU problem has a line with slope 1 which corroborates the strong convergence analysis in [11] and is efficient for large time steps. Notice however that the scheme converges only in the regime $h \geq c\varepsilon$. This is not surprising because this stepsize restriction is needed in the analysis in [11]. We did not apply the stiff trigonometric method TRIG1 to the Kubo oscillator because this scheme is designed for second order SDEs with additive noise and does not generalizes straightforwardly to the multiplicative noise case.

[5] In the results for the Kubo oscillator (left picture in Figure 7), notice that the errors appear bounded from below for the finer number of steps $1024, 2048$. This is due to the Monte-Carlo error, here with size $\simeq 5 \cdot 10^{-4}$ for 10^7 trajectories.

5 Conclusion

We have presented and analyzed a class of large time step integrators for highly oscillatory SDEs based on the idea of multi-revolution composition methods originally introduced for deterministic problems [7]. The versatile micro-macro approach can in principle be coupled with any micro integrator in the spirit of the Heterogeneous Multiscale Method [2, 16, 17]. This permits to construct methods with favourable geometric properties such as quadratic first integral preserving methods, as illustrated in the numerical experiments. Since the multi-revolution methods allow a (macro) stepsize $H \gg \varepsilon$ and have uniform accuracy with respect to the oscillatory period ε, the proposed approach could be coupled with the multilevel Monte-Carlo method for SDEs [18], which permits to significantly speed up the standard Monte-Carlo method by reducing the variance of the scheme. Indeed, applied to stiff problems, the multilevel approach requires a scheme allowing both coarse and fine time steps, as shown recently in [1] in the context of stiff stochastic diffusion problems. Finally, we mention that the analysis of the multi-revolution approach is studied in [8] for the linear Schrödinger equation in the deterministic case. The extension to the stochastic case will be investigated in future works.

Acknowledgements The author is grateful to Arnaud Debussche for helpful discussions on the topic of SPDEs, and to Assyr Abdulle for hospitality and parallel computing resources at the ANMC chair, École Polytechnique Fédérale de Lausanne. This work was partially supported by the Fonds National Suisse, project No. 200020_144313/1.

6 Appendix

The proof of Lemma 3.9 used in the proof of the main Theorem 2.6 is an immediate consequence of the following more general proposition applied with $f_1 = \delta A$, $g_1^r = 0$, $f_2 = f$, $g_2^r = g^r$, and $\varepsilon = \gamma H$. Although this result on the Strang splitting method is well known in deterministic contexts (see for instance [21]), it seems not available in the literature in a stochastic context, and it can be useful by itself also for $\varepsilon = 1$.

Proposition 6.1. *For a fixed ε, consider the system of SDEs*

$$dX(t) = (f_1 + \varepsilon f_2)(X(t))dt + \sum_{r=1}^{m} (g_1^r + \sqrt{\varepsilon} g_2^r)(X(t))dW_r(t), \qquad X(0) = X_0 \tag{41}$$

*where $f_1, f_2, g_1^r, g_2^r : \mathbb{R}^d \to \mathbb{R}^d$, $r = 1, \dots, m$ are smooth vector fields satisfying **(H)**. Consider the stochastic flows Φ_h^1 and Φ_h^2 with time h of the auxiliary SDE problems $dX = f_1 dt + \sum_{r=1}^m g_1^r dW_r$ and $dX = f_2 dt + \sum_{r=1}^m g_2^r dW_r$, respectively. Then, the integrator*

$$X_{k+1} = \Phi_{h/2}^1 \circ \Phi_{\varepsilon h}^2 \circ \Phi_{h/2}^1 (X_k),$$

where $\Phi_{h/2}^1$, $\Phi_{\varepsilon h}^2$, and $\Phi_{h/2}^1$ are applied with independent Wiener processes, satisfies for all $\phi \in C_P^6(\mathbb{R}^d, \mathbb{R})$ and all $hk \le T$,

$$|\mathbb{E}(\phi(X_k)) - \mathbb{E}(\phi(X(kh)))| \le C\varepsilon h^2,$$

where C is independent of h, ε and $X(t)$ is the solution of (41). In addition, considering for Φ_h^2 a weak second order approximation satisfying (15) and (16), the above error estimates remains valid.

Proof. We denote by $e^{t(\mathcal{L}_1+\varepsilon\mathcal{L}_2)}\phi(x)$ and $e^{t\mathcal{L}_j}\phi(x), j=1,2$ the solutions at time t of the backward Kolmogorov equation variant (23) with $\mathcal{L}_\varepsilon$ replaced by $\mathcal{L}_1+\varepsilon\mathcal{L}_2$ and $\mathcal{L}_j$, $j=1,2$, respectively. We have by [33, Thm. 2.2] (see Theorem 3.3 with $A=0$),

$$\mathbb{E}(\phi(X(h))|X(0)=x) = e^{h(\mathcal{L}_1+\varepsilon\mathcal{L}_2)}\phi(x)$$

and analogously $\mathbb{E}(\phi(\Phi_h^j(x))) = e^{h\mathcal{L}_j}\phi(x)$, for $j=1,2$, where the generator $\mathcal{L}_j$ is defined by $\mathcal{L}_j\phi(x) := f_j(x)\cdot\nabla\phi(x) + \frac{1}{2}\sum_{r=1}^m \phi''\left(g_j^r(x), g_j^r(x)\right)$ and $e^{h(\mathcal{L}_1+\varepsilon\mathcal{L}_2)}\phi(x)$ and $e^{h\mathcal{L}_j}\phi(x)$ are C^3 functions with respect to h with derivatives with polynomial growth (13). Analogously to (27) in the proof of Lemma 3.6, we deduce

$$\mathbb{E}(\phi(X_1)|X_0=x) = e^{\frac{h}{2}\mathcal{L}_1}\circ e^{h\varepsilon\mathcal{L}_2}\circ e^{\frac{h}{2}\mathcal{L}_1}\phi(x).$$

We obtain for all initial condition $X_0=x$,

$$e(h) := \mathbb{E}(\phi(X_1)) - \mathbb{E}(\phi(X(h))) = e^{\frac{h}{2}\mathcal{L}_1}\circ e^{\varepsilon h\mathcal{L}_2}\circ e^{\frac{h}{2}\mathcal{L}_1}\phi(x) - e^{h\mathcal{L}_1+\varepsilon h\mathcal{L}_2}\phi(x).$$

Next, a straightforward Taylor series calculation yields $\frac{d^p e(h)}{dh^p}\Big|_{h=0} = 0$ for $p=0,1,2$ (see e.g. [21, Chap. III.4]). Using [33, Thm. 2.2] (see Theorem 3.3 with $A=0$) applied repeatedly with ϕ replaced by $e^{\frac{h}{2}\mathcal{L}_1}\phi$ and $e^{\varepsilon h\mathcal{L}_2}\circ e^{\frac{h}{2}\mathcal{L}_1}\phi$, the quantity $e(h)$ is of class C^3 with respect to h and the third derivative has the polynomial growth (13) with respect to x. In addition, by the variation of constant formula,

$$e(h) = \varepsilon\int_0^h e^{s\mathcal{L}_1+\varepsilon s\mathcal{L}_2}\mathcal{L}_2 e^{(h-s)\mathcal{L}_1}\phi ds - \varepsilon\int_0^h e^{h\mathcal{L}_1/2}e^{\varepsilon s\mathcal{L}_2}\mathcal{L}_2 e^{h\mathcal{L}_1/2}\phi ds,$$

which yields that the third derivative of $e(h)$ has size $\mathcal{O}(\varepsilon)$ for all h. By the Taylor formula, we deduce for all initial condition $X_0=x$ the local error bound

$$|e(h)| \le C(x)\varepsilon h^3$$

where $C(x)$ is independent of h,ε and satisfies the polynomial growth (13) with respect to x. We conclude the proof of the global weak error estimate using the Milstein theorem in [24] (see Remark 2.7). In the case where Φ_h^2 a weak second order approximation instead of the exact flow, the above local error estimate remains valid for all h small enough, and using Remark 2.7 permits to conclude the proof of Proposition 6.1. □

References

[1] A. Abdulle and A. Blumenthal. Stabilized multilevel Monte Carlo method for stiff stochastic differential equations. *J. Comput. Phys.*, 251:445–460, 2013.

[2] A. Abdulle, W. E, B. Engquist, and E. Vanden-Eijnden. The heterogeneous multiscale method. *Acta Numer.*, 21:1–87, 2012.

[3] A. Abdulle, G. Vilmart, and K. Zygalakis. Weak second order explicit stabilized methods for stiff stochastic differential equations. *SIAM J. Sci. Comput.*, 35(4):1792–1814, 2013.

[4] S. Blanes and F. Casas. On the necessity of negative coefficients for operator splitting schemes of order higher than two. *Appl. Num. Math.*, 54:23–37, 2005.

[5] F. Castella, P. Chartier, S. Descombes, and G. Vilmart. Splitting methods with complex times for parabolic equations. *BIT*, 49(3):487–508, 2009.

[6] F. Castella, P. Chartier, F. Méhats, and A. Murua. Stroboscopic averaging for the nonlinear Schrödinger equation. Technical report, Institut de Recherche Mathématique de Rennes, Konputazio Zientziak eta A.A. Saila, 2012.

[7] P. Chartier, J. Makazaga, A. Murua, and G. Vilmart. Multi-revolution composition methods for highly oscillatory differential equations. *to appear in Numerische Mathematik (published online)*, 2014.

[8] P. Chartier, F. Mehats, and M. Thalhammer. Convergence analysis of multi-revolution composition time-splitting pseudo-spectral methods for linear Schrödinger equations. *in preparation*, 2013.

[9] P. Chartier, A. Murua, and J. M. Sanz-Serna. Higher-order averaging, formal series and numerical integration I: B-series. *Found. Comput. Math.*, 10(6):695–727, 2010.

[10] D. Cohen. On the numerical discretisation of stochastic oscillators. *Math. Comput. Simulation*, 82(8):1478–1495, 2012.

[11] D. Cohen, S. Larsson, and M. Sigg. A trigonometric method for the linear stochastic wave equation. *SIAM J. Numer. Anal.*, 51(1):204–222, 2013.

[12] D. Cohen and M. Sigg. Convergence analysis of trigonometric methods for stiff second-order stochastic differential equations. *Numer. Math.*, 121(1):1–29, 2012.

[13] A. de Bouard and A. Debussche. Weak and strong order of convergence of a semidiscrete scheme for the stochastic nonlinear Schrödinger equation. *Appl. Math. Optim.*, 54(3):369–399, 2006.

[14] K. Debrabant and A. Rößler. Diagonally drift-implicit runge–kutta methods of weak order one and two for itô sdes and stability analysis. *Appl. Num. Math.*, 59(3–4):595 – 607, 2009.

[15] A. Debussche and L. Di Menza. Numerical simulation of focusing stochastic nonlinear Schrödinger equations. *Phys. D*, 162(3-4):131–154, 2002.

[16] W. E and B. Engquist. The heterogeneous multiscale methods. *Commun. Math. Sci.*, 1(1):87–132, 2003.

[17] W. E, B. Engquist, X. Li, W. Ren, and E. Vanden-Eijnden. Heterogeneous multiscale methods: a review. *Commun. Comput. Phys.*, 2(3):367–450, 2007.

[18] M. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[19] B. Grébert and C. Villegas-Blas. On the energy exchange between resonant modes in nonlinear Schrödinger equations. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 28(1):127–134, 2011.

[20] V. Grimm and M. Hochbruck. Error analysis of exponential integrators for oscillatory second-order differential equations. *J. Phys. A*, 39(19):5495–5507, 2006.

[21] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer Series in Computational Mathematics 31. Springer-Verlag, Berlin, second edition, 2006.

[22] E. Hansen and A. Ostermann. High order splitting methods for analytic semigroups exist. *BIT*, 49(3):527–542, 2009.

[23] B. Melendo and M. Palacios. A new approach to the construction of multirevolution methods and their implementation. *Appl. Numer. Math.*, 23(2):259–274, 1997.

[24] G. Milstein. Weak approximation of solutions of systems of stochastic differential equations. *Theory Probab. Appl.*, 30(4):750–766, 1986.

[25] G. Milstein and M. Tretyakov. *Stochastic numerics for mathematical physics*. Scientific Computing. Springer-Verlag, Berlin and New York, 2004.

[26] G. N. Milstein and M. V. Tretyakov. Numerical methods in the weak sense for stochastic differential equations with small noise. *SIAM J. Numer. Anal.*, 34(6):2142–2167, 1997.

[27] L. R. Petzold, L. O. Jay, and J. Yen. Numerical solution of highly oscillatory ordinary differential equations. In *Acta Numerica*, volume 6, pages 437–483. Cambridge Univ. Press, Cambridge, 1997.

[28] A. Rößler. Second order Runge-Kutta methods for Itô stochastic differential equations. *SIAM J. Numer. Anal.*, 47(3):1713–1738, 2009.

[29] J. Sanders, F. Verhulst, and J. Murdock. *Averaging methods in nonlinear dynamical systems*. Springer, New York, second edition, 2007.

[30] Q. Sheng. Solving linear partial differential equations by exponential splitting. *IMA J. Numer. Anal.*, 9:199–212, 1989.

[31] G. J. Sussman and J. Wisdom. Chaotic evolution of the solar system. *Science*, 257:56–62, 1992.

[32] D. Talay. Efficient numerical schemes for the approximation of expectations of functionals of the solution of a SDE and applications. *Lecture Notes in Control and Inform. Sci., Springer*, 61:294–313, 1984.

[33] D. Talay. Discrétisation d'une équation différentielle stochastique et calcul approché d'espérances de fonctionnelles de la solution. *RAIRO Modél. Math. Anal. Numér.*, 20(1):141–179, 1986.

[34] M. Tao, H. Owhadi, and J. E. Marsden. Structure preserving stochastic impulse methods for stiff Langevin systems with a uniform global error of order 1 or 1/2 on position. *arXiv:1006.4657v1*, 2010.