On the stochastic Magnus expansion and its application to SPDEs

Kevin Kamm[*]   Stefano Pagliarani[†]   Andrea Pascucci[‡]

This version: May 14, 2021

Abstract

We derive the stochastic version of the Magnus expansion for linear systems of stochastic differential equations (SDEs). The main novelty with respect to the related literature is that we consider SDEs in the Itô sense, with progressively measurable coefficients, for which an explicit Itô-Stratonovich conversion is not available. We prove convergence of the Magnus expansion up to a stopping time τ and provide a novel asymptotic estimate of the cumulative distribution function of τ. As an application, we propose a new method for the numerical solution of stochastic partial differential equations (SPDEs) based on spatial discretization and application of the stochastic Magnus expansion. A notable feature of the method is that it is fully parallelizable. We also present numerical tests in order to asses the accuracy of the numerical schemes.

1 Introduction

The Magnus expansion (hereafter referred to as ME) is a classical tool to solve non-autonomous linear differential equations. Generalizations of the ME to *Stratonovich SDEs* are well-known and were proposed by several authors (see for instance [2], [3], [5], [32] and the references given in Section 1.2). In this paper we derive, for the first time to the best of our knowledge, the ME for *Itô SDEs* under general assumptions which do not guarantee an explicit Itô-Stratonovich conversion, namely progressively measurable stochastic coefficients. Our main results are the convergence of the stochastic ME up to a stopping time τ and a novel asymptotic estimate of the cumulative distribution function of τ. The latter improves some previous estimates obtained in purely Markovian settings and is based on an application

---

[*]Dipartimento di Matematica, Università di Bologna, Bologna, Italy. **e-mail**: kevin.kamm@unibo.it
[†]Dipartimento di Matematica, Università di Bologna, Bologna, Italy. **e-mail**: stefano.pagliarani9@unibo.it
[‡]Dipartimento di Matematica, Università di Bologna, Bologna, Italy. **e-mail**: andrea.pascucci@unibo.it

of Morrey's inequality. We also explore possible applications to the numerical solution of stochastic partial differential equations (SPDEs).

Let $d, q \in \mathbb{N}$ and consider the linear matrix-valued Itô SDE

$$\begin{cases} dX_t = B_t X_t dt + A_t^{(j)} X_t dW_t^j, \\ X_0 = I_d, \end{cases}$$

with $A^{(1)}, \ldots, A^{(q)}, B$ being real $(d \times d)$-matrix-valued bounded stochastic processes, I_d the identity $(d \times d)$-matrix and $W = (W^1, \ldots, W^q)$ a q-dimensional standard Brownian motion. In (1), as well as anywhere throughout the paper, we use Einstein summation convention to imply summation of terms containing W^j, over the index j from 1 to q.

In the deterministic case, i.e. $A^{(j)} \equiv 0$, $j = 1, \ldots, q$, (1) reduces to the matrix-valued ODE

$$\begin{cases} \frac{d}{dt} X_t = B_t X_t, \\ X_0 = I_d, \end{cases}$$

which admits an explicit solution, in terms of matrix exponential, in the time-homoge-neous case. Namely, if $B_t \equiv B$, the unique solution to (1) reads as

$$X_t = e^{tB}, \qquad t \geq 0.$$

However, in the non-autonomous case, the ODE (1) does not admit an explicit solution. In particular, if B_t is not constant, the solution X_t typically differs from $e^{\int_0^t B_s ds}$. This is due to the fact that, in general, B_t and B_s do not commute for $t \neq s$. As it turns out, a representation of the solution in terms of a matrix exponential is still possible, at least for short times, i.e.

$$X_t = e^{Y_t},$$

for $t \geq 0$ suitably small and Y_t real valued $(d \times d)$-matrix. Moreover, Y admits a semi-explicit expansion as a series of iterated integrals involving nested Lie commutators of the function B at different times. Such representation is known as *Magnus expansion* ([23]) and its first terms read as

$$\begin{aligned} Y_t = \int_0^t B_s ds &+ \frac{1}{2} \int_0^t ds \int_0^s [B_s, B_u] du \\ &+ \frac{1}{6} \int_0^t ds \int_0^s du \int_0^u \left([B_s, [B_u, B_r]] + [B_r, [B_u, B_s]] \right) dr + \cdots, \end{aligned}$$

where $[A, B] := AB - BA$ denotes the Lie commutator. The ME has a wide range of physical applications and the related literature has grown increasingly over the last decades (see, for instance, the excellent survey paper [3] and the references given therein).

In the stochastic case, when $j = 1$, $B_t \equiv 0$ and A is constant, i.e. $A_t(\omega) \equiv A$, the Itô

equation (1) reduces to

$$\begin{cases} dX_t = AX_t dW_t, \\ X_0 = I_d, \end{cases}$$

whose explicit solution can be easily proved to be of the form (1), with

$$Y_t = -\frac{1}{2}A^2 t + AW_t, \qquad t \geq 0.$$

In general, when the matrices $A_t^{(j)}, A_s^{(j)}, B_t, B_s$ with $t \neq s$ do not commute, an explicit solution to (1) is not known. For instance, in the non-commutative case, neither the equation

$$\begin{cases} dX_t = Bdt + AX_t dW_t, \\ X_0 = I_d, \end{cases}$$

nor the equation

$$\begin{cases} dX_t = A_t X_t dW_t, \\ X_0 = I_d, \end{cases}$$

admit an explicit solution, save some particular cases (see for instance the example in Section 3.3). Among the approximation tools that were developed in the literature to solve stochastic differential equations, including (1), some Magnus-type expansions that extend (1)-(1) were derived in different contexts. We now go on to describe our contribution to this stream of literature, and then to firm our results within the existing ones. In particular, a detailed comparison with existing stochastic MEs previously derived by other authors will be provided below, in the last subsection.

1.1 Description of the main results.

In this paper we derive a Magnus-type representation formula for the solution to the Itô SDE (1), which is (1) together with

$$Y_t = Y_t^{(1)} + Y_t^{(2)} + Y_t^{(3)} + \cdots \qquad t \in [0, \tau],$$

for τ suitably small, strictly positive stopping time. In analogy with the deterministic ME, the general term $Y^{(n)}$ can be expressed recursively, and contains iterated stochastic integrals of nested Lie commutators of the processes $B, A^{(j)}$ at different times.

In the case $j = 1$, the first two terms of the expansion read as

$$Y_t^{(1)} = \int_0^t B_s ds + \int_0^t A_s dW_s,$$

$$Y_t^{(2)} = \frac{1}{2}\int_0^t \left(\left[B_s, \int_0^s B_u du + \int_0^s A_u dW_u \right] - A_s^2 \right) ds$$

$$+ \frac{1}{2}\int_0^t \left[A_s, \int_0^s B_u du + \int_0^s A_u dW_u \right] dW_s.$$

For example, in the case of SDE (1) the latter can be reduced to

$$Y_t^{(1)} = Bt + AW_t,$$
$$Y_t^{(2)} = [A, B] \left(\frac{1}{2} tW_t - \int_0^t W_s ds \right) - \frac{1}{2} A^2 t.$$

Notice that the last expressions do not contain stochastic integrals. In fact, in the general autonomous case, and if $j = 1$, all the iterated stochastic integrals in $Y^{(n)}$ can be solved for any n (see Corollary 5.2.4 in [16]). Therefore, in this case the expansion becomes numerically computable by only approximating Lebesgue integrals, as opposed to stochastic Runge-Kutta schemes, which typically require the numerical approximation of stochastic integrals. As we shall see in the numerical tests in Section 3, this feature allows us to choose a sparser time-grid in order to save computation time. This feature is also preserved in some non-autonomous cases as illustrated in Section 3.

A notable feature of the expansion is the possibility of parallelizing the computation of its terms. In contrast to standard iterative methods, which require the solution at a given time-step in order to go through the next step in the iteration, the discretization of the integrals in the terms $Y^{(n)}$ can be done simultaneously for all the time steps. Conclusively, this entails the possibility of parallelizing *over all times* in the time-grid and makes the numerical implementation of the stochastic ME perfectly GPU applicable.

As it often happens when deriving convergent (either asymptotically or absolutely) expansions, a formal derivation precedes the rigorous result: that is what we do for (1)-(1.1) in Section 2.2. Just like the derivation of the deterministic ME relies on the possibility of writing the logarithm Y as the solution to an ODE, in the stochastic case the first step consists in representing Y as the solution to an SDE. Such representation of Y will be more involved compared to the deterministic case because of the presence of the second order derivatives of the exponential map coming from the application of Itô's formula. This is a distinctive feature of our derivation with respect to other analogous results in the Stratonovich setting where the standard chain-rule applies. With the SDE representation for Y at hand, the expansion (1.1) stems, like in the deterministic case, from applying a Dyson-type perturbation procedure to the SDE solved by Y.

In the deterministic case, the convergence of the ME (1) to the exact logarithm of the solution to (1) was studied by several authors, who proved progressively sharper lower bounds on the maximum $\bar{t}$ such that the convergence to the exact solution is assured for any $t \in [0, \bar{t}]$. At the best of our knowledge, the sharpest estimate was given in [26], namely

$$\bar{t} \geq \sup \left\{ t \geq 0 \mid \int_0^t \|B_s\| \, ds < \pi \right\},$$

where $\|B_s\|$ denotes the spectral norm. Note that the existence of a real logarithm of X_t is an issue that underlies the study of the convergence of the ME. We state here our main result, proved in Section 2.3, which deals with these matters in the stochastic case, when the coefficients $B, A^{(j)}$ in (1) are progressively measurable processes. We defer a comparison with

previous convergence results for stochastic Magnus-type expansions to the next subsection. We denote by $\mathcal{M}^{d \times d}$ the space of the $(d \times d)$-matrices with real entries. Also, for an $\mathcal{M}^{d \times d}$-valued stochastic process $M = (M_t)_{t \in [0,T]}$, we set $\|M\|_T := \||M\|_F\|_{L^\infty([0,T] \times \Omega)}$, where $\|\cdot\|_F$ denotes the Frobenius (Euclidean entry-wise) norm.

Theorem 1.1. *Let $A^{(1)}, \ldots, A^{(q)}$ and B be bounded, progressively measurable, $\mathcal{M}^{d \times d}$-valued processes defined on a filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t \geq 0})$ equipped with a standard q-dimensional Brownian motion $W = (W^1, \cdots, W^q)$. For $T > 0$ let also $X = (X_t)_{t \in [0,T]}$ be the unique strong solution to* (1) *(see Lemma 2.10). There exists a strictly positive stopping time $\tau \leq T$ such that:*

(i)   *X_t has a real logarithm $Y_t \in \mathcal{M}^{d \times d}$ up to time τ, i.e.*

$$X_t = e^{Y_t}, \qquad 0 \leq t < \tau;$$

(ii)   *the following representation holds P-almost surely:*

$$Y_t = \sum_{n=0}^{\infty} Y_t^{(n)}, \qquad 0 \leq t < \tau,$$

     where $Y^{(n)}$ is the n-th term in the stochastic ME as defined in (2.2) *and* (2.2)–(2.2)*;*

(iii)   *there exists a positive constant C, only dependent on $\|A^{(1)}\|_T, \ldots, \|A^{(q)}\|_T$, $\|B\|_T$, T and d, such that*

$$P(\tau \leq t) \leq Ct, \qquad t \in [0,T].$$

The proof of (i) relies on the continuity of X together with a standard representation for the matrix logarithm. The key point in the proof of (ii) consists in showing that $X_t^{\varepsilon, \delta}$ and its logarithm $Y_t^{\varepsilon, \delta}$ are holomorphic as functions of (ε, δ), where $X_t^{\varepsilon, \delta}$ represents the solution of (1) when $A^{(j)}$ and B are replaced by $\varepsilon A^{(j)}$ and δB, respectively. Once this is established, the representation (1.1) follows from observing that, by construction, the series in (1.1) is exactly the formal power series of $Y_t^{\varepsilon, \delta}$ at $(\varepsilon, \delta) = (1, 1)$. To prove the holomorphicity of $X_t^{\varepsilon, \delta}$ we follow the same approach typically adopted to prove regularity properties of stochastic flows. Namely, in Lemma 2.10 we state some maximal L^p and Hölder estimates (with respect to the parameters) for solutions to SDEs with random coefficients and combine them with the Kolmogorov continuity theorem. Finally, the proof of (iii) owes one more time to the L^p estimates in Lemma 2.10 and to a Sobolev embedding theorem to obtain pointwise estimates w.r.t. the parameters (ε, δ) above.

Theorem 1.1 has been used in the recent paper [34] (cf. Lemma 1) where a semi-linear non-commutative Itô-SDEs is studied and Euler, Milstein and derivative-free numerical schemes are developed, with a convergence analysis for those schemes.

In the last part of the paper we perform numerical tests with the Magnus expansion. In particular, Section 3.2 is devoted to the application of the stochastic ME to the numerical solution of parabolic stochastic partial differential equations (SPDEs). The idea is to discretize the SPDE *only in space* and then approximate the resulting linear matrix-valued SDE

by truncating the series in (1.1)-(1.1). The goal here is to propose the application of stochastic MEs as novel approximation tools for SPDEs; we study the error of this approximating procedure only numerically, in a case where an explicit benchmark is available, and we defer the theoretical error analysis to further studies.

1.2 Review of the literature and comparison.

Stochastic generalizations of the MEs were proposed by several authors. To the best of our knowledge, we recognize mainly two streams of research.

The beginning of the first one can be traced back to the work [2], where the author derived *exponential* stochastic Taylor expansions (see also [1], [16] for general stochastic Taylor series) of the solution of a system of Stratonovich SDEs with values on a manifold $\mathcal{M}$, i.e.

$$\begin{cases} dX_t = B(X_t)dt + A^{(j)}(X_t) \circ dW_t^j, \\ X_0 = x_0, \end{cases}$$

with $B, A^{(j)}$ being smooth, deterministic and autonomous vector fields on $\mathcal{M}$. The stochastic flow of (1.2) is represented in terms of the exponential map of a stochastic vector field Y, i.e.

$$X_t(x_0) = \exp Y_t(x_0),$$

the vector field Y being expressed by an infinite series of iterated stochastic integrals multiplying nested commutators of the vector fields $B, A^{(j)}$. This representation is proved up to a strictly positive stopping time and extends some previous results in [11], [31] for the commutative case and in [33], [19], [13] for the nilpotent case. Refinements of [2] were proved in [6] making the expansion of Y more explicit. Later, numerical methods based on these representations were proposed in [7] and [8]. Such techniques, known as Castell-Gaines methods, require the approximation of the solution to a time-dependent ODE besides the approximation of iterated stochastic integrals. Truncating the expansion of Y at a specified order, these schemes turn out to be asymptotically efficient in the sense of Newton [28].

If $\mathcal{M} = \mathcal{M}^{d \times d}$ and the vector fields are linear, then (1.2) reduces to the Stratono-vich version of (1) with $B, A^{(j)}$ constant matrices, and the representation of X given in [2] can be seen as a stochastic ME, in that the exponential map of Y reduces to the multiplication by a matrix exponential. In fact, in this case the expansion in [2] becomes explicit in terms of iterated stochastic integrals, and can be shown to coincide with the expansion in this paper by applying Itô-Stratonovich conversion formula. In the very interesting paper [22], the authors study several computational aspects of numerical schemes stemming from the truncated ME, in which the iterated stochastic integrals are approximated by their conditional expectation. Besides showing that asymptotic efficiency holds for an arbitrary number of Brownian components, they compare the theoretical accuracy with the one of analogous schemes based on Dyson (or Neumann) series, which are obtained by applying stochastic Taylor expansion directly on the equation. They find that, although the theoretical accuracy of Magnus schemes is not superior, Magnus-based approximations seem more accurate than

their Dyson counterparts in practice. They also discuss the computational cost deriving from approximating the iterated stochastic integrals and the matrix exponentiation, in relation to different features of the problem such as the dimension and the number of Brownian motions, as well as to the order of the numerical scheme.

The second stream of literature is explicitly aimed at extending the original Magnus results to stochastic settings and can be traced back to [5] where the ME is derived via formal arguments for a linear system of Stratonovich SDEs with deterministic coefficients. Clearly, in the autonomous case such expansion coincides with the one obtained by Ben Arous ([2]), whereas in the non-autonomous case, $B \equiv 0$ and $j = 1$, it is formally equivalent to the deterministic ME (1) with all the Lebesgue integrals replaced by Stratonovich ones. The authors of [5] do not address the convergence of the ME, but rather study computational aspects of the resulting approximation, in particular in comparison with Runge-Kutta stochastic schemes. The authors of [24] consider the Ito SDE (1) with constant coefficients, and propose to resolve via Euler method the SDE (2.2) for the logarithm of the solution. In [32] the ME for the Stratonovich version of (1) with deterministic coefficients is applied to solve non-linear SDEs; however, the error analysis of the truncated expansion seems flawed, since the fact that the Magnus series converges only up to a positive stopping time is overlooked. In [27], a general procedure for designing higher strong order methods for Itô SDEs on matrix Lie groups is outlined.

We now go on to discuss the contribution of this paper with respect to the existing literature. In the first place, Theorem 1.1 on the convergence of the ME requires very weak conditions on the coefficients, which are stochastic processes satisfying the sole assumption of progressive measurability. This is a novel aspect compared to the results in [2], [6], which surely cover a wider class of SDEs, but under the assumption of time-independent deterministic coefficients. We point out that this feature is also relevant in light of the fact that our result is stated for Itô SDEs as opposed to Stratonovich ones. Indeed, while this difference might appear as minor in the Markovian case, where a simple conversion formula exists (cf. [10] and [21]), it becomes substantial in the case of progressively measurable coefficients. We also point out that, even in the Markovian non-autonomous case, convergence issues were not discussed in [5] and [22].

Another novel aspect of our result concerns the estimate (1.1) for the cumulative distribution function of the stopping time τ up to which the Magnus series converges to the real logarithm of the solution: this kind of estimate was unknown even in the autonomous case. Theorem 11 in [2] (see also [6]) provides an asymptotic estimate for the truncation error of the logarithm, which in the linear case studied in this paper would read as

$$Y_t = \sum_{n=1}^{N} Y_t^{(n)} + t^{\frac{N+1}{2}} R_t, \qquad t < T,$$

with R bounded in probability. Although this type of result holds for the general SDE (1.2), it is weaker than Theorem 1.1 in the linear case. In fact, it can be obtained by (1.1) together with the standard estimate $\left\| \sup_{0 \le s \le t} \|Y_s^{(n)}\|_F \right\|_{L^2(\Omega)} \le C t^{\frac{N+1}{2}}$, but not the other way around.

A rigorous error analysis of the ME is left for future research, as well as applications to non-linear SDEs (see [32] for a recent attempt in this direction).

The rest of the paper is structured as follows. In Section 2 we derive the ME and prove Theorem 1.1. In particular, Section 2.1 contains the key Lemma 2.1 with a representations for the first and second order differentials through which the terms $Y^{(n)}$ in (1.1)-(1.1) will be defined, and some preliminary results that will be used to derive the expansion. Section 2.2 contains a formal derivation of (1.1)-(1.1). Section 2.3 is entirely devoted to the proof of Theorem 1.1.

In Section 3 we first introduce a numerical test for an SDE with constant, non-commuting coefficients. The formulas for the first three orders of the ME in this test will also be used in in Subsection 3.2, where we present the application of the ME to the numerical solution of parabolic SPDEs. In particular, in Subsection 3.2.1 we recall some general facts about stochastic Cauchy problems, in Subsection 3.2.2 we introduce the finite-difference–Magnus approximation scheme and we check the effectiveness of the proposed approach through numerical tests. Finally, we provide an additional numerical test to assess the accuracy of the ME in the case of time-dependent coefficients.

2 Itô-Stochastic ME

In this section we define the terms in the expansion (1.1) and prove Theorem 1.1.

2.1 Preliminaries

Let $\mathcal{M}^{d \times d}$ be the vector space of $(d \times d)$ real-valued matrices. For the readers' convenience we recall the following notations. Throughout the paper we denote by $[\cdot, \cdot]$ the standard Lie brackets operation, i.e.

$$[M, N] \coloneqq MN - NM, \qquad M, N \in \mathcal{M}^{d \times d},$$

and by $\|\cdot\|$ the spectral norm on $\mathcal{M}^{d \times d}$. Also, we denote by β_k, $k \in \mathbb{N}_0$, the Bernoulli numbers defined as the derivatives of the function $x \mapsto x/(e^x - 1)$ computed at $x = 0$. For sake of convenience we report the first three Bernoulli numbers: $\beta_0 = 1$, $\beta_1 = -\frac{1}{2}$, $\beta_2 = \frac{1}{6}$. Note also that $\beta_{2m+1} = 0$ for any $m \in \mathbb{N}$.

We now define the operators that we will use in the sequel. For a fixed $\Sigma \in \mathcal{M}^{d \times d}$, we let:

- $\mathrm{ad}_\Sigma^j : \mathcal{M}^{d \times d} \to \mathcal{M}^{d \times d}$, for $j \in \mathbb{N}_0$, be the linear operators defined as

$$\mathrm{ad}_\Sigma^0(M) \coloneqq M,$$
$$\mathrm{ad}_\Sigma^j(M) \coloneqq \left[\Sigma, \mathrm{ad}_\Sigma^{j-1}(M)\right], \qquad j \in \mathbb{N}.$$

To ease notation we also set $\mathrm{ad}_\Sigma \coloneqq \mathrm{ad}_\Sigma^1$;

- $e^{\mathrm{ad}_\Sigma} : \mathcal{M}^{d\times d} \to \mathcal{M}^{d\times d}$ be the linear operator defined as

$$e^{\mathrm{ad}_\Sigma}(M) := \sum_{n=0}^{\infty} \frac{1}{n!} \mathrm{ad}_\Sigma^n(M) = e^\Sigma M e^{-\Sigma},$$

where $e^\Sigma := \sum_{j=0}^{\infty} \frac{\Sigma^j}{j!}$ is the standard matrix exponential;

- $\mathcal{L}_\Sigma : \mathcal{M}^{d\times d} \to \mathcal{M}^{d\times d}$ be the linear operator defined as

$$\mathcal{L}_\Sigma(M) := \int_0^1 e^{\mathrm{ad}_{\tau\Sigma}}(M) d\tau = \sum_{n=0}^{\infty} \frac{1}{(n+1)!} \mathrm{ad}_\Sigma^n(M);$$

- $\mathcal{Q}_\Sigma : \mathcal{M}^{d\times d} \times \mathcal{M}^{d\times d} \to \mathcal{M}^{d\times d}$ be the bi-linear operator defined as

$$\mathcal{Q}_\Sigma(M, N) := \mathcal{L}_\Sigma(M)\,\mathcal{L}_\Sigma(N) + \int_0^1 \tau \left[\mathcal{L}_{\tau\Sigma}(N), e^{\mathrm{ad}_{\tau\Sigma}}(M) \right] d\tau$$

$$= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{\mathrm{ad}_\Sigma^n(M)}{(n+1)!} \frac{\mathrm{ad}_\Sigma^m(N)}{(m+1)!} + \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{[\mathrm{ad}_\Sigma^n(N), \mathrm{ad}_\Sigma^m(M)]}{(n+m+2)(n+1)!m!}.$$

In the next lemma we provide explicit expressions for the first and second order differentials of the exponential map $\mathcal{M}^{d\times d} \ni M \mapsto e^M$. We recall that this map is smooth and in particular, it is continuously twice differentiable.

Lemma 2.1. *For any $\Sigma \in \mathcal{M}^{d\times d}$, the first and the second order differentials at Σ of the exponential map $\mathcal{M}^{d\times d} \ni M \mapsto e^M$ are given by*

$$M \mapsto \mathcal{L}_\Sigma(M)\, e^\Sigma = e^\Sigma\, \mathcal{L}_{-\Sigma}(M), \qquad\qquad M \in \mathcal{M}^{d\times d},$$

$$(M, N) \mapsto \mathcal{Q}_\Sigma(M, N)\, e^\Sigma, \qquad\qquad M, N \in \mathcal{M}^{d\times d},$$

where $\mathcal{L}_\Sigma$ and $\mathcal{Q}_\Sigma$ are the linear and (symmetric) bi-linear operators as defined in (2.1)-(2.1).

We point out that this result, though very basic, is novel and of independent interest (for instance it was recently employed in [14]).

Proof. The first part of the statement, concerning the first order differential, is a classical result; its proof can be found in [3, Lemma 2] among other references.

We prove the second part. Fix $M \in \mathcal{M}^{d\times d}$ and denote by $\partial_M e^\Sigma$ the first order directional derivative of e^Σ w.r.t. M, i.e.

$$\partial_M e^\Sigma := \frac{d}{dt} e^{\Sigma+tM} \Big|_{t=0}.$$

By the first part, we have

$$\partial_M e^\Sigma = \mathcal{L}_\Sigma(M)\, e^\Sigma, \qquad \Sigma \in \mathcal{M}^{d\times d}.$$

We now show that, for any $M, N \in \mathcal{M}^{d\times d}$, the second order directional derivative

$$\partial_{N,M} e^\Sigma := \frac{d}{dt} \partial_M e^{\Sigma+tN} \Big|_{t=0}$$

is given by

$$\partial_{N,M}e^{\Sigma} = \mathcal{Q}_{\Sigma}(N, M)\, e^{\Sigma}, \qquad \Sigma \in \mathcal{M}^{d \times d}.$$

We have

$$\frac{d}{dt}\partial_M e^{\Sigma+tN} = \frac{d}{dt}\Big(\mathcal{L}_{\Sigma+tN}(M)\, e^{\Sigma+tN}\Big)$$

$$= \mathcal{L}_{\Sigma+tN}(M)\, \mathcal{L}_{\Sigma+tN}(N)\, e^{\Sigma+tN} + \Big(\frac{d}{dt}\mathcal{L}_{\Sigma+tN}(M)\Big)\, e^{\Sigma+tN}.$$

We use the definition (2.1) and exchange the differentiation and integration signs to obtain

$$\frac{d}{dt}\mathcal{L}_{\Sigma+tN}(M) = \int_0^1 \frac{d}{dt}e^{\mathrm{ad}_{\tau(\Sigma+tN)}}(M)d\tau$$

(by (2.1))

$$= \int_0^1 \frac{d}{dt}\Big(e^{\tau(\Sigma+tN)}Me^{-\tau(\Sigma+tN)}\Big)d\tau$$

$$= \int_0^1 \Big(\frac{d}{dt}e^{\tau(\Sigma+tN)}\Big)\, M\, e^{-\tau(\Sigma+tN)}d\tau + \int_0^1 e^{\tau(\Sigma+tN)}\, M\, \frac{d}{dt}e^{-\tau(\Sigma+tN)}d\tau$$

(by employing the two expressions in (2.1) for the first-order differential)

$$= \int_0^1 \tau \mathcal{L}_{\tau(\Sigma+tN)}(N)\, e^{\tau(\Sigma+tN)}\, M\, e^{-\tau(\Sigma+tN)}d\tau$$

$$- \int_0^1 \tau e^{\tau(\Sigma+tN)}\, M\, e^{-\tau(\Sigma+tN)}\, \mathcal{L}_{\tau(\Sigma+tN)}(N)d\tau$$

$$= \int_0^1 \tau\Big[\mathcal{L}_{\tau(\Sigma+tN)}(N), e^{\mathrm{ad}_{\tau(\Sigma+tN)}}(M)\Big]d\tau.$$

This, together with (2.1), proves (2.1).

To conclude, we prove equality (2.1). It is enough to observe that

$$\int_0^1 \tau\, \mathcal{L}_{\tau\Sigma}(N)\, e^{\mathrm{ad}_{\tau\Sigma}}(M)d\tau = \int_0^1 \tau \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{\mathrm{ad}_{\tau\Sigma}^n(N)}{(n+1)!}\frac{\mathrm{ad}_{\tau\Sigma}^m(M)}{m!}d\tau$$

$$= \int_0^1 \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{\mathrm{ad}_{\Sigma}^n(N)}{(n+1)!}\frac{\mathrm{ad}_{\Sigma}^m(M)}{m!}\tau^{n+m+1}d\tau$$

$$= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{\mathrm{ad}_{\Sigma}^n(N)\mathrm{ad}_{\Sigma}^m(M)}{(n+m+2)(n+1)!m!}.$$

$\square$

Proposition 2.2 (Itô formula). *Let Y be an $\mathcal{M}^{d \times d}$-valued Itô process of the form*

$$dY_t = \mu_t dt + \sigma_t^j dW_t^j.$$

Then we have

$$de^{Y_t} = \left(\mathcal{L}_{Y_t}(\mu_t) + \frac{1}{2} \sum_{j=1}^{q} \mathcal{Q}_{Y_t}(\sigma_t^j, \sigma_t^j) \right) e^{Y_t} dt + \mathcal{L}_{Y_t}(\sigma_t^j) \, e^{Y_t} dW_t^j.$$

Proof. The statement follows from the multi-dimensional Itô formula (see, for instance, [29]) combined with Lemma 2.1 and applied to the exponential process e^{Y_t}. $\qquad\qquad\square$

We also have the following inversion formula for the operator $\mathcal{L}_\Sigma$.

Lemma 2.3 (**Baker, 1905**). *Let $\Sigma \in \mathcal{M}^{d \times d}$. The operator $\mathcal{L}_\Sigma$ is invertible if and only if the eigenvalues of the linear operator ad_Σ are different from $2m\pi$, $m \in \mathbb{Z} \setminus \{0\}$. Furthermore, if $\|\Sigma\| < \pi$, then*

$$\mathcal{L}_\Sigma^{-1}(M) = \sum_{k=0}^{\infty} \frac{\beta_k}{k!} ad_\Sigma^k(M), \qquad M \in \mathcal{M}^{d \times d}.$$

For a proof to Lemma 2.3 we refer the reader to [3].

2.2 Formal derivation

In this section we perform formal computations to derive the terms $Y^{(n)}$ appearing in the ME (1.1). Although such computations are heuristic at this stage, they are meant to provide the reader with an intuitive understanding of the principles that underlie the expansion procedure. Their validity will be proved a fortiori, in Section 2.3, in order to prove Theorem 1.1.

Let $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t \geq 0})$ be a filtered probability space. Assume that, for any $\varepsilon, \delta \in \mathbb{R}$, the process $X^{\varepsilon, \delta} = (X_t^{\varepsilon, \delta})_{t \geq 0}$ solves the Itô SDE

$$\begin{cases} dX_t^{\varepsilon, \delta} = \delta B_t X_t^{\varepsilon, \delta} dt + \varepsilon A_t^{(j)} X_t^{\varepsilon, \delta} dW_t^j, \\ X_0^{\varepsilon, \delta} = I_d, \end{cases}$$

and that it admits the exponential representation

$$X_t^{\varepsilon, \delta} = e^{Y_t^{\varepsilon, \delta}}$$

with $Y^{\varepsilon, \delta}$ being an $\mathcal{M}^{d \times d}$-valued Itô process. Clearly, if $(\varepsilon, \delta) = (1, 1)$, then (2.2)-(2.2) reduce to (1)-(1). Assume now that $Y^{\varepsilon, \delta}$ is of the form (2.2). Then, Proposition 2.2 yields

$$\varepsilon A_t^{(j)} = \mathcal{L}_{Y_t^{\varepsilon, \delta}}(\sigma_t^j), \qquad j = 1, \dots, q,$$

$$\delta B_t = \mathcal{L}_{Y_t^{\varepsilon, \delta}}(\mu_t) + \frac{1}{2} \sum_{j=1}^{q} \mathcal{Q}_{Y_t^{\varepsilon, \delta}}(\sigma_t^j, \sigma_t^j).$$

Inverting now (2.2)-(2.2), in accord with (2.3), one obtains

$$\sigma_t^j = \mathcal{L}_{Y_t^{\varepsilon,\delta}}^{-1}\big(\varepsilon A_t^{(j)}\big) = \varepsilon \sum_{k=0}^{\infty} \frac{\beta_k}{k!} \mathrm{ad}_{Y_t^{\varepsilon,\delta}}^k\big(A_t^{(j)}\big), \qquad j = 1,\ldots,q,$$

$$\mu_t = \mathcal{L}_{Y_t^{\varepsilon,\delta}}^{-1}\Big(\delta B_t - \frac{1}{2}\sum_{j=1}^{q} \mathcal{Q}_{Y_t^{\varepsilon,\delta}}\big(\sigma_t^j, \sigma_t^j\big)\Big)$$

$$= \sum_{k=0}^{\infty} \frac{\beta_k}{k!} \mathrm{ad}_{Y_t^{\varepsilon,\delta}}^k\Big(\delta B_t - \frac{1}{2}\sum_{j=1}^{q}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\Big(\frac{\mathrm{ad}_{Y_t^{\varepsilon,\delta}}^n\big(\sigma_t^j\big)}{(n+1)!}\frac{\mathrm{ad}_{Y_t^{\varepsilon,\delta}}^m\big(\sigma_t^j\big)}{(m+1)!}$$

$$+ \frac{\big[\mathrm{ad}_{Y_t^{\varepsilon,\delta}}^n\big(\sigma_t^j\big), \mathrm{ad}_{Y_t^{\varepsilon,\delta}}^m\big(\sigma_t^j\big)\big]}{(n+m+2)(n+1)!m!}\Big)\Big)\Big).$$

Equivalently, $Y^{\varepsilon,\delta}$ solves the Itô SDE

$$\begin{cases} dY_t^{\varepsilon,\delta} = \mu^{\varepsilon,\delta}(t, Y_t^{\varepsilon,\delta})dt + \sigma_j^{\varepsilon}(t, Y_t^{\varepsilon,\delta})dW_t^j, \\ Y_0^{\varepsilon,\delta} = 0, \end{cases}$$

with

$$\sigma_j^{\varepsilon}(t, \cdot) = \varepsilon \sum_{n=0}^{\infty} \frac{\beta_n}{n!} \mathrm{ad}_{\cdot}^n\big(A_t^{(j)}\big), \qquad j = 1,\ldots,d,$$

$$\mu^{\varepsilon,\delta}(t, \cdot) = \sum_{n=0}^{\infty} \frac{\beta_n}{n!} \mathrm{ad}_{\cdot}^n\Big(\delta B_t - \frac{1}{2}\sum_{j=1}^{q} \mathcal{Q}_{\cdot}\big(\sigma_j^{\varepsilon}(t, \cdot), \sigma_j^{\varepsilon}(t, \cdot)\big)\Big).$$

We now assume that $Y^{\varepsilon,\delta}$ admits the representation

$$Y_t^{\varepsilon,\delta} = \sum_{n=0}^{\infty}\sum_{r=0}^{n} Y_t^{(r,n-r)} \varepsilon^r \delta^{n-r},$$

for a certain family $(Y^{(r,n-r)})_{n,r \in \mathbb{N}_0}$ of stochastic processes. In particular, setting $(\varepsilon, \delta) = (1,1)$, (2.2) would yield

$$Y_t = \sum_{n=0}^{\infty} Y_t^{(n)} \quad \text{with} \quad Y_t^{(n)} := \sum_{r=0}^{n} Y_t^{(r,n-r)}.$$

Remark 2.4. Note that it is possible to re-order the double series $\sum_{n=0}^{\infty}\sum_{r=0}^{n} Y_t^{(r,n-r)}$ according to any arbitrary choice, for the latter will be proved to be absolutely convergent. The above choice for $Y^{(n)}$ contains all the terms of equal order by weighing ε and δ in the same way. A different choice, which respects the probabilistic relation $\sqrt{\Delta t} \approx \Delta W_t$, corresponds to weighing δ as ε^2. This would lead to setting

$$Y_t^{(n)} := \sum_{r=0}^{\lfloor \frac{n}{2} \rfloor} Y_t^{(n-2r,r)}$$

in (2.2).

Remark 2.5. Observe that, if the function $(\varepsilon, \delta) \mapsto Y_0^{\varepsilon,\delta}$ is assumed to be continuous P-almost surely, then the initial condition in (2.2) implies

$$Y_0^{(i,j)} = 0 \quad P\text{-a.s.}, \qquad i, j \in \mathbb{N}_0,$$

and thus

$$Y_0^{(n)} = 0 \quad P\text{-a.s.}, \qquad n \in \mathbb{N}_0.$$

We now plug (2.2) into (2.2) and collect all terms of equal order in ε and δ. Up to order 2 we obtain

$$
\begin{aligned}
\varepsilon^0 \delta^0 : \quad & Y_t^{(0,0)} = 0, \\
\varepsilon^1 \delta^0 : \quad & Y_t^{(1,0)} = \int_0^t A_s^{(j)} dW_s^j, \\
\varepsilon^0 \delta^1 : \quad & Y_t^{(0,1)} = \int_0^t B_s ds, \\
\varepsilon^2 \delta^0 : \quad & Y_t^{(2,0)} = -\frac{1}{2} \int_0^t (A_s^{(j)})^2 ds + \frac{1}{2} \int_0^t \Big[A_s^{(j)}, \int_0^s A_r^{(i)} dW_r^i \Big] dW_s^j, \\
\varepsilon^1 \delta^1 : \quad & Y_t^{(1,1)} = \frac{1}{2} \int_0^t \Big[B_s, \int_0^s A_r^{(j)} dW_r^j \Big] ds + \frac{1}{2} \int_0^t \Big[A_s^{(j)}, \int_0^s B_r dr \Big] dW_s^j, \\
\varepsilon^0 \delta^2 : \quad & Y_t^{(0,2)} = \frac{1}{2} \int_0^t \Big[B_s, \int_0^s B_r dr \Big] ds,
\end{aligned}
$$

for any $t \geq 0$, where we used, one more time, Einstein summation convention to imply summation over the indexes i, j and Remark 2.5 to set all the initial conditions equal to zero. Proceeding by induction, one can obtain a recursive representation for the general term $Y^{(r,n-r)}$ in (2.2), namely:

$$Y_t^{(r,n-r)} = \int_0^t \mu_s^{r,n-r} ds + \int_0^t \sigma_s^{r,n-r,j} dW_s^j, \qquad n \in \mathbb{N}_0, \ r = 0, \ldots, n,$$

where the terms $\sigma^{r,n-r,j}, \mu^{r,n-r}$ are defined recursively as

$$
\begin{aligned}
\sigma_s^{r,n-r,j} &:= \sum_{i=0}^{n-1} \frac{\beta_i}{i!} S_s^{r-1,n-r,i}(A^{(j)}), \\
\mu_s^{r,n-r} &:= \sum_{i=0}^{n-1} \frac{\beta_i}{i!} S_s^{r,n-r-1,i}(B) - \frac{1}{2} \sum_{j=1}^{q} \sum_{i=0}^{n-2} \frac{\beta_i}{i!} \sum_{q_1=2}^{r} \sum_{q_2=0}^{n-r} S^{r-q_1,n-r-q_2,i}(Q^{q_1,q_2,j}),
\end{aligned}
$$

with

$$Q_s^{q_1,q_2,j} := \sum_{i_1=2}^{q_1} \sum_{i_2=0}^{q_2} \sum_{h_1=1}^{i_1-1} \sum_{h_2=0}^{i_2} \sum_{p_1=0}^{q_1-i_1} \sum_{p_2=0}^{q_2-i_2} \sum_{m_1=0}^{p_1+p_2} \sum_{m_2=0}^{q_1-i_1-p_1+q_2-i_2-p_2}$$

$$\left(\frac{S_s^{p_1,p_2,m_1}\big(\sigma_s^{h_1,h_2,j}\big)}{(m_1+1)!} \frac{S_s^{q_1-i_1-p_1,q_2-i_2-p_2,m_2}\big(\sigma_s^{i_1-h_1,i_2-h_2,j}\big)}{(m_2+1)!} \right.$$

$$\left. + \frac{\big[S_s^{p_1,p_2,m_1}\big(\sigma_s^{i_1-h_1,i_2-h_2,j}\big), S_s^{q_1-i_1-p_1,q_2-i_2-p_2,m_2}\big(\sigma_s^{h_1,h_2,j}\big)\big]}{(m_1+m_2+2)(m_1+1)!m_2!} \right),$$

and with the operators S being defined as

$$S_s^{r-1,n-r,0}(A) := \begin{cases} A & \text{if } r = n = 1, \\ 0 & \text{otherwise,} \end{cases}$$

$$S_s^{r-1,n-r,i}(A) := \sum_{\substack{(j_1,k_1),\dots,(j_i,k_i)\in\mathbb{N}_0^2 \\ j_1+\cdots+j_i=r-1 \\ k_1+\cdots+k_i=n-r}} \big[Y_s^{(j_1,k_1)}, \big[\dots, \big[Y_s^{(j_i,k_i)}, A_s\big] \dots \big]\big]$$

$$= \sum_{\substack{(j_1,k_1),\dots,(j_i,k_i)\in\mathbb{N}_0^2 \\ j_1+\cdots+j_i=r-1 \\ k_1+\cdots k_i=n-r}} \mathrm{ad}_{Y_s^{(j_1,k_1)}} \circ \cdots \circ \mathrm{ad}_{Y_s^{(j_i,k_i)}}(A_s), \qquad i \in \mathbb{N}.$$

Remark 2.6. All the processes $Y^{(r,n-r)}$, with $n \in \mathbb{N}$ and $r = 0, \dots, n$, are well defined according to the recursion (2.2)-(2.2)-(2.2), as long as B and $A^{(1)}, \dots, A^{(q)}$ are bounded and progressively measurable stochastic processes.

Example 2.7. As we already pointed out in the introduction, in the case $j = 1$ and $B \equiv 0$, the SDE (1) admits an explicit solution given by

$$Y_t = -\frac{1}{2}A^2 t + AW_t, \qquad t \geq 0,$$

and the terms in the ME (2.2) read as

$$Y_t^{(1)} = AW_t, \qquad Y_t^{(2)} = -\frac{1}{2}A^2 t, \qquad Y_t^{(n)} = 0, \ n \geq 3.$$

In particular, the ME coincides with the exact solution with the first two terms.

2.3 Convergence analysis

In this section we prove Theorem 1.1. To avoid ambiguity, only in this section, we denote by $\mathcal{M}_{\mathbb{R}}^{d\times d}$ and $\mathcal{M}_{\mathbb{C}}^{d\times d}$ the spaces of $(d \times d)$-matrices with real and complex entries, respectively; on these spaces we shall make use of the Frobenius norm denoted by $\|\cdot\|_F$. We say that a matrix-valued function is holomorphic if all its entries are holomorphic functions. We recall that $W = (W^1, \dots, W^q)$ is a q-dimensional standard Brownian motion and $A^{(1)}, \dots, A^{(q)}, B$ are bounded $\mathcal{M}_{\mathbb{R}}^{d\times d}$-valued progressively measurable stochastic processes defined on a filtered

probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t \geq 0})$. Also recall that, for any $\mathcal{M}_{\mathbb{R}}^{d \times d}$-valued process $M = (M_t)_{t \in [0,T]}$, we set $\|M\|_T := \| \|M\|_F \|_{L^\infty([0,T] \times \Omega)}$.

We start with two preliminary lemmas.

Lemma 2.8. *Assume that* $Y = (Y_t^{\varepsilon, \delta})_{\varepsilon, \delta \in \mathbb{R}, t \in \mathbb{R}_{\geq 0}}$ *is a* $\mathcal{M}_{\mathbb{R}}^{d \times d}$-*valued stochastic process that can be represented as a convergent series of the form* (2.2). *If* Y *solves the SDE* (2.2) *up to a positive stopping time* τ, *then* $Y^{(r, n-r)}$ *in* (2.2) *are Itô processes and satisfy* (2.2)-(2.2)-(2.2) *for any* $t < \tau$.

Proof. We prove (2.2)-(2.2)-(2.2) only for $n = 0, 1$. Namely, we show that (2.2), (2.2) and (2.2) hold up to time τ, P-a.s. The representation for the general term $Y^{(r, n-r)}$ can be proved by induction; we omit the details for brevity.

Since Y is of the form (2.2) then $Y_t^{(0,0)} = Y_t^{0,0}$ for any $t < \tau$. Moreover, since Y solves the SDE (2.2) then $Y^{0,0} \equiv 0$ on $[0, \tau[$, P-a.s. Thus (2.2) holds up to time τ, P-a.s.

Now, (2.2) yields

$$Y_t^{\varepsilon, 0} = \varepsilon \int_0^t A_s^{(j)} dW_s^j + \varepsilon R_t^\varepsilon, \qquad t \in [0, \tau[, \quad P\text{-a.s.},$$

where

$$R_t^\varepsilon = \int_0^t \left(\sum_{k=1}^\infty \frac{\beta_k}{k!} \text{ad}_{Y_s^{\varepsilon,0}}^k (A_s^{(j)}) \right) dW_s^j$$
$$- \frac{\varepsilon}{2} \int_0^t \mathcal{L}_{Y_s^{\varepsilon,0}}^{-1} \left(\mathcal{Q}_{Y_s^{\varepsilon,0}} \left(\sum_{k=0}^\infty \frac{\beta_k}{k!} \text{ad}_{Y_s^{\varepsilon,0}}^k (A_s^{(j)}), \sum_{k=0}^\infty \frac{\beta_k}{k!} \text{ad}_{Y_s^{\varepsilon,0}}^k (A_s^{(j)}) \right) \right) ds.$$

Note that, again by (2.2), $R^0 \equiv 0$ P-a.s. Moreover, representation (2.2) implies continuity of $\varepsilon \mapsto Y_t^{\varepsilon, 0}$ near $\varepsilon = 0$, which in turn implies the continuity of $\varepsilon \mapsto R_t^\varepsilon$. Thus we have $\lim_{\varepsilon \to 0} R_t^\varepsilon = R_t^0$ P-a.s. This, together with (2.3) and (2.2) implies that (2.2) necessarily holds, up to time τ, P-a.s.

Similarly, (2.2) yields

$$Y_t^{0, \delta} = \delta \int_0^t B_s ds + \delta Q_t^\delta, \qquad t \in [0, \tau[, \quad P\text{-a.s.},$$

with

$$Q_t^\delta = \int_0^t \left(\sum_{k=1}^\infty \frac{\beta_k}{k!} \text{ad}_{Y_s^{0,\delta}}^k (B_s) \right) ds$$

and the same argument employed above yields (2.2) up to time τ, P-almost surely. $\qquad \square$

Lemma 2.9. *Let* $M \in \mathcal{M}_{\mathbb{C}}^{d \times d}$ *be nonsingular and such that* $\|M - I_d\| < 1$ *where* $\|\cdot\|$ *is the spectral norm. Then* M *has a unique logarithm, which is*

$$\log M = \sum_{n=1}^\infty (-1)^{n+1} \frac{(M - I_d)^n}{n}$$
$$= (M - I_d) \int_0^\infty \frac{1}{1 + \mu} (\mu I_d + M)^{-1} d\mu.$$

In particular, we have

$$\|\log M\| \leq -\log\left(1 - \|M - I_d\|\right).$$

Proof. The first representation is a standard result. The second representation stems from the factorization $M = VJV^{-1}$ with J in Jordan form, under the assumption that M has no non-positive real eigenvalues, i.e. $\lambda \in \mathbb{C} \backslash] - \infty, 0]$ for any λ eigenvalue of M. This last property, however, is ensured by the assumption $\|M - I_d\| < 1$. Indeed, the latter implies

$$\|Mv - v\|_F < 1, \qquad v \in \mathbb{R}^d, \ |v| = 1,$$

which in turn implies that, if λ is a real eigenvalue of M and v is one of its normalized eigenvectors, then

$$1 > \|Mv - v\|_F = |\lambda v - v| = |\lambda - 1|.$$

$\square$

We have one last preliminary lemma, containing some technical results concerning the solutions to (2.2). These are semi-standard, in that they can be inferred by combining and adapting existing results in the literature.

Lemma 2.10. *For any $T > 0$ and $\varepsilon, \delta \in \mathbb{C}$, the SDE (2.2) has a unique strong solution $(X_t^{\varepsilon,\delta})_{t \in [0,T]}$. For any $p \geq 1$ and $h > 0$ there exists a positive constant κ, only dependent on $\|A^{(1)}\|_T, \ldots, \|A^{(q)}\|_T, \|B\|_T, d, T, h$ and p, such that*

$$E\left[\|X_t^{\varepsilon,\delta} - X_s^{\varepsilon',\delta'}\|_F^{2p}\right] \leq \kappa\left(|t - s|^p + (|\varepsilon - \varepsilon'| + |\delta - \delta'|)^{2p}\right),$$

$$E\left[\sup_{0 \leq u \leq t} \|X_u^{\varepsilon,\delta} - X_0^{\varepsilon,\delta}\|_F^{2p}\right] \leq \kappa t^p(|\varepsilon| + |\delta|)^{2p},$$

for any $0 \leq t, s \leq T$ and $\varepsilon, \delta, \varepsilon', \delta' \in \mathbb{C}$ with $|\varepsilon|, |\delta|, |\varepsilon'|, |\delta'| \leq h$.

Up to modifications, $(X_t^{\varepsilon,\delta})_{\varepsilon,\delta \in \mathbb{C}, \, t \in [0,T]}$ is a continuous process such that:

i)   *for any $t \in [0, T]$, the function $(\varepsilon, \delta) \mapsto X_t^{\varepsilon,\delta}$ is holomorphic;*

ii) *the functions $(t, \varepsilon, \delta) \mapsto \partial_\varepsilon X_t^{\varepsilon,\delta}$ and $(t, \varepsilon, \delta) \mapsto \partial_\delta X_t^{\varepsilon,\delta}$ are continuous;*

iii) *for any $p \geq 1$ and $h > 0$ there exists a positive constant κ only dependent on $\|A^{(1)}\|_T, \ldots, \|A^{(q)}\|_T, \|B\|_T, d, T, h$ and p, such that*

$$E\left[\sup_{0 \leq s \leq t} \left\{\|\partial_\varepsilon X_s^{\varepsilon,\delta}\|_F^{2p} + \|\partial_\delta X_s^{\varepsilon,\delta}\|_F^{2p}\right\}\right] \leq \kappa t^p(|\varepsilon| + |\delta|)^p,$$

for any $t \in [0, T]$ and $|\varepsilon|, |\delta| \leq h$.

Proof. Existence of the solution and estimates (2.10)-(2.10) of the moments follow from the results in Section 5, Chapter 2 in [17] (in particular, see Corollary 5 on page 80 and Theorem 7 on page 82).

The second part of the statement is a refined version of the Kolmogorov continuity theorem in the form that can be found for instance in Section 2.3 in [20]: a detailed proof is provided in [15]. $\square$

Remark 2.11. The existence and uniqueness for the solution to (1) is a particular case of the previous result.

We are now in the position to prove Theorem 1.1.

of Theorem 1.1. We fix $h > 1$, $T > 0$, and let $(X_t^{\varepsilon,\delta})_{\varepsilon,\delta \in \mathbb{C}, t \in [0,T]}$ be the solution of the SDE (2.2) as defined in Lemma 2.10. Moreover, for $t \in]0, T]$, we set $Q_{t,h} :=]0, t[\times B_h(0)$ where $B_h(0) = \{(\varepsilon, \delta) \in \mathbb{C}^2 \mid |(\varepsilon, \delta)| < h\}$.

<u>Part (i)</u>: as $X_0^{\varepsilon,\delta} = I_d$, by continuity the random time defined as

$$\tau := \sup \left\{ t \in [0,T] \, \Big| \, \|X_s^{\varepsilon,\delta} - I_d\|_F < 1 - e^{-\pi} \text{ for any } (s, \varepsilon, \delta) \in Q_{t,h} \right\}$$

is strictly positive. Furthermore, again by continuity,

$$(\tau \leq t) = \bigcup_{(s,\varepsilon,\delta) \in \widetilde{Q}_{t,h}} \left(\|X_s^{\varepsilon,\delta} - I_d\|_F \geq 1 - e^{-\pi} \right), \qquad t \in [0, T],$$

where $\widetilde{Q}_{t,h}$ is a countable, dense subset of $Q_{t,h}$, which implies that τ is a stopping time.

Let $(t, \varepsilon, \delta) \in Q_{\tau,h}$: by Lemma 2.9 applied to $M = X_t^{\varepsilon,\delta}$ we have

$$Y_t^{\varepsilon,\delta} := \log X_t^{\varepsilon,\delta} = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{(X_t^{\varepsilon,\delta} - I_d)^n}{n}$$
$$= (X_t^{\varepsilon,\delta} - I_d) \int_0^{\infty} \frac{1}{1+\mu} (\mu I_d + X_t^{\varepsilon,\delta})^{-1} d\mu.$$

Notice that $X_t^{\varepsilon,\delta}$ (and therefore also $Y_t^{\varepsilon,\delta}$) is real for $\varepsilon, \delta \in \mathbb{R}$: in particular, $Y_t = Y_t^{1,1}$ is real and this proves Part (i).

<u>Part (ii)</u>: since $(\varepsilon, \delta) \mapsto X_t^{\varepsilon,\delta}$ is holomorphic, we can differentiate (2.3) to infer that $(\varepsilon, \delta) \mapsto Y_t^{\varepsilon,\delta}$ is holomorphic as well: indeed, we have for $(t, \varepsilon, \delta) \in Q_{\tau,h}$

$$\partial_\varepsilon Y_t^{\varepsilon,\delta} = \partial_\varepsilon X_t^{\varepsilon,\delta} \int_0^{\infty} \frac{1}{1+\mu} (\mu I_d + X_t^{\varepsilon,\delta})^{-1} d\mu$$
$$+ (X_t^{\varepsilon,\delta} - I_d) \int_0^{\infty} \frac{1}{1+\mu} (\mu I_d + X_t^{\varepsilon,\delta})^{-1} (\partial_\varepsilon X_t^{\varepsilon,\delta}) (\mu I_d + X_t^{\varepsilon,\delta})^{-1} d\mu,$$

and similarly by differentiating w.r.t. to δ. Then the expansion of $Y_t^{\varepsilon,\delta}$ in power series at $(\varepsilon, \delta) = (0, 0)$ is absolutely convergent on $B_h(0)$ and the representation (2.2) holds on $Q_{\tau,h}$ for some random coefficients $Y_t^{(r,n-r)}$. To conclude we need to show that the latter are as given by (2.2)-(2.2)-(2.2). Then (1.1) will stem from (2.2) by setting $(\varepsilon, \delta) = (1, 1)$.

In light of Lemma 2.9, the logarithmic map is continuously twice differentiable on the open subset of $\mathcal{M}_{\mathbb{C}}^{d \times d}$ of the matrices M such that $\|M - I_d\| < 1$: thus $Y_t^{\varepsilon,\delta}$ admits an Itô representation (2.2) for $(t, \varepsilon, \delta) \in Q_{\tau,h}$. Then Proposition 2.2 together with (2.2) yield (2.2)-(2.2) P-a.s. up to τ for any $(\varepsilon, \delta) \in B_h(0) \cap \mathbb{R}^2$. Furthermore, by estimate (2.9) of Lemma 2.9 we also have $\left\| Y_t^{\varepsilon,\delta} \right\| < \pi$ for $t < \tau$. Therefore, we can apply Baker's Lemma 2.3 to invert $\mathcal{L}_{Y_t^{\varepsilon,\delta}}$ in (2.2)-(2.2) and obtain that $Y^{\varepsilon,\delta}$ solves (2.2) up to τ for any $(\varepsilon, \delta) \in B_h(0) \cap \mathbb{R}^2$. Part

(ii) then follows from Lemma 2.8.

<u>Part (iii)</u>: for $t \leq T$ let

$$f_t(\varepsilon, \delta) := \max_{s \in [0,t]} \|X_s^{\varepsilon,\delta} - I_d\|_F, \qquad M_t := \sup_{(\varepsilon,\delta) \in B_h(0)} f_t(\varepsilon, \delta).$$

By definition (2.3), we have

$$P(\tau \leq t) \leq P\left(M_t \geq 1 - e^{-\pi}\right) \leq \frac{1}{(1 - e^{-\pi})^2} E\left[M_t^2\right],$$

and therefore (1.1) follows by suitably estimating $E\left[M_t^2\right]$. To prove such an estimate we will show in the last part of the proof that f_t belongs to the Sobolev space $W^{1,2p}(B_h(0))$ for any $p \geq 1$ and we have

$$E\left[\|f_t\|_{W^{1,2p}(B_h(0))}^{2p}\right] \leq Ct^p, \qquad t \in [0, T],$$

where the positive constant C depends only on $\|A^{(1)}\|_T, \ldots, \|A^{(q)}\|_T$, $\|B\|_T$, d, T, h and p. Since $f_t \in W^{1,2p}(B_h(0))$ and $B_h(0) \subseteq \mathbb{R}^4$, by Morrey's inequality (cf., for instance, Corollary 9.14 in [4]) for any $p > 2$ we have

$$M_t \leq c_0 \|f_t\|_{W^{1,2p}(B_h(0))},$$

where c_0 is a a positive constant, dependent only on p and h (in particular, c_0 is independent of ω). Combining (2.3) with (2.3), for a fixed $p > 2$ we have

$$E\left[M_t^2\right] \leq c_0^2 E\left[\|f_t\|_{W^{1,2p}(B_h(0))}^2\right] \leq$$

(by Hölder inequality)

$$\leq c_0^2 Ct, \qquad t \in [0, T].$$

This last estimate, combined with (2.3), proves (1.1).

To conclude, we are left with the proof of (2.3). First we have

$$E\left[\int_{B_h(0)} |f_t(\varepsilon, \delta)|^{2p} d\varepsilon \, d\delta\right] = \int_{B_h(0)} E\left[|f_t(\varepsilon, \delta)|^{2p}\right] d\varepsilon \, d\delta \leq Ct^p,$$

where we used the estimate (2.10) of Lemma 2.10 in the last inequality. Fix now $t \in \,]0, T]$, $(\varepsilon, \delta), (\varepsilon', \delta') \in B_h(0)$ such that $f_t(\varepsilon', \delta') \leq f_t(\varepsilon, \delta)$ and set

$$\bar{t} \in \underset{0 \leq s \leq t}{\arg\max} \|X_s^{\varepsilon,\delta} - I_d\|_F, \qquad \widetilde{t} \in \underset{0 \leq s \leq t}{\arg\max} \|X_s^{\varepsilon',\delta'} - I_d\|_F.$$

Note that the arg max above do exist in that the process $g_s(\varepsilon, \delta) := X_s^{\varepsilon,\delta} - I_d$ is continuous

in s and we have

$$
\begin{aligned}
\left|f_t(\varepsilon,\delta) - f_t(\varepsilon',\delta')\right| &= \left|\|g_{\bar{t}}(\varepsilon,\delta)\|_F - \|g_{\bar{t}}(\varepsilon',\delta')\|_F\right| \leq \left|\|g_{\bar{t}}(\varepsilon,\delta)\|_F - \|g_{\bar{t}}(\varepsilon',\delta')\|_F\right| \\
&\leq \|g_{\bar{t}}(\varepsilon,\delta) - g_{\bar{t}}(\varepsilon',\delta')\|_F \leq \sup_{0\leq s\leq t} \|g_s(\varepsilon,\delta) - g_s(\varepsilon',\delta')\|_F \\
&\leq \left|(\varepsilon,\delta)-(\varepsilon',\delta')\right| \sup_{0\leq s\leq t} \sup_{\substack{|\bar{\varepsilon}-\varepsilon|\leq|\varepsilon'-\varepsilon| \\ |\bar{\delta}-\delta|\leq|\delta'-\delta|}} \|\nabla g_s(\bar{\varepsilon},\bar{\delta})\|_F,
\end{aligned}
$$

where $\nabla = \nabla_{\varepsilon,\delta}$. This, as $(s,\varepsilon,\delta) \mapsto \nabla g_s(\varepsilon,\delta)$ is continuous on $Q_{t,h}$, implies $f_t \in W^{1,2p}(B_h(0))$ and yields the key inequality

$$
|\nabla f_t(\varepsilon,\delta)| \leq \sup_{0\leq s\leq t} \|\nabla X_s^{\varepsilon,\delta}\|_F, \qquad (\varepsilon,\delta) \in B_h(0).
$$

Therefore, we have

$$
\begin{aligned}
E\left[\int_{B_h(0)} |\nabla f_t(\varepsilon,\delta)|^{2p}\,d\varepsilon\,d\delta\right] &= \int_{B_h(0)} E\left[|\nabla f_t(\varepsilon,\delta)|^{2p}\right] d\varepsilon\,d\delta \\
&\leq \int_{B_h(0)} E\left[\sup_{0\leq s\leq t} \|\nabla X_s^{\varepsilon,\delta}\|_F^{2p}\right] d\varepsilon\,d\delta \leq Ct^p,
\end{aligned}
$$

where we used the estimate (2.10) of Lemma 2.10 in the last inequality. This, together with (2.3), proves (2.3) and conclude the proof. $\qquad\square$

3 Numerical tests and applications to SPDEs

We present here some numerical tests in order to confirm the accuracy of the approximate solutions to (1) stemming from the truncation of the series (1.1). We also show how this approximation can be applied to approximate the solutions to stochastic partial differential equations (SPDEs) of parabolic type.

   We consider two examples of SDEs (one in Section 3.1 and one in Section 3.3), for which we compute the first three terms of the ME given by (2.2)-(2.2) and present numerical experiments to test the accuracy of the approximate solutions to (1) stemming from it. In both cases we consider $j=1$ in (1) and replace $A^{(1)}$ with A to shorten notation. The first example will be for constant matrices A and B. In the second one we will consider $B \equiv 0$ and a deterministic upper diagonal A_t. For each numerical test we will implement the exponential of the truncated ME up to order $n=1,2$ and 3, i.e.

$$
X^{(n)} := e^{\sum_{i=1}^n Y^{(i)}}, \qquad n=1,2,3,
$$

and compare it with a benchmark solution to (1). In Section 3.2 we turn our attention to the application of the ME to the numerical resolution of SPDEs. In particular, in the numerical tests we will make use of the ME for constant matrices discussed in Section 3.1.

Error and notations.   Throughout this section we will employ the following tags:

-     `euler` for the solution obtained with Euler-Maruyama scheme, which was implemented with Matlab's pagefun for the matrix multiplication on a single GPU and vectorized over all samples;

-     `exact` to denote the time-discretization of an explicit solution, if available;

-     `m1`, `m2` and `m3` for the time-discretization of the Magnus approximations in (3), up to order 1,2 and 3, respectively.

For the numerical error analysis in the SDE examples we will make use of the following norms. Denoting by X^{ref} and by X^{app} a benchmark and an approximate solution, respectively, to (1) and by $(t_k)_{k=0,\dots,N}$ a homogeneous discretization of $[0,t]$, we consider the random variable

$$\text{Err}_t := \frac{\Delta}{t} \sum_{k=1}^{N} \frac{\|X_{t_k}^{\text{ref}} - X_{t_k}^{\text{app}}\|_F}{\|X_{t_k}^{\text{ref}}\|_F} \approx \frac{1}{t} \int_0^t \frac{\|X_s^{\text{ref}} - X_s^{\text{app}}\|_F}{\|X_s^{\text{ref}}\|_F} ds \qquad \text{with } \Delta = \frac{t}{N},$$

namely a discretization of the time-averaged relative error on the interval $[0,t]$. This is a way to measure the error on the whole trajectory as opposed to the error at a specific given time. Then we use Monte Carlo simulation, with M independent realizations of the discretized Brownian trajectories, to approximate the distribution of Err_t.

The matrix norm above is the Frobenius norm. In the following tests, `m1`, `m2` and `m3` will always play the role of X^{app}, `exact` always the role of X^{ref}, whereas `euler` will be either X^{app} or X^{ref} depending on whether `exact` is available or not.

We used for the calculations `Matlab R2021a` with Parallel Computing Toolbox running on Windows 10 Pro, on a machine with the following specifications: processor Intel(R) Core(TM) i7-8750H @ 2.20 GHz, 2x32 GB (Dual Channel) Samsung SODIMM RAM @ 2667 MHz, and a NVIDIA GeForce RTX 2070 with Max-Q Design (8 GB GDDR6 RAM). Also, we will make use of the Matlab built-in routine `expm` for the computation of the matrix exponential. As it turns out, this represents the most expensive step in the implementation of the Magnus approximation. However important, the pursue of optimized method for the matrix exponentiation is an extended topic of separate interest, which goes beyond the goals of this paper. Therefore, here we will limit ourselves to pointing out, separately, the computational times for the approximations of the logarithm and of the matrix exponential.

In the implementation we simulate the Brownian motion first and use it as an input for each scheme to be able to compare the trajectories of each scheme amongst each other.

3.1 Example: constant A and B.

With a slight abuse of notation, we consider $A_t \equiv A$ and $B_t \equiv B$. Recall that, if A and B do not commute, there is in general no closed-form solution to (1). The first three terms of the

ME read as

$$Y_t^{(1)} = Bt + AW_t, \qquad Y_t^{(2)} = [A, B] \left(\frac{1}{2} tW_t - \int_0^t W_s ds \right) - \frac{1}{2} A^2 t,$$

$$Y_t^{(3)} = [[B, A], A] \left(\frac{1}{2} \int_0^t W_s^2 ds - \frac{1}{2} W_t \int_0^t W_s ds + \frac{1}{12} tW_t^2 \right)$$

$$+ [[B, A], B] \left(\int_0^t sW_s ds - \frac{1}{2} t \int_0^t W_s ds - \frac{1}{12} t^2 W_t \right).$$

We point out that, in this case, all the stochastic integrals appearing in the ME can be solved in terms of Lebesgue integrals by using Itô's formula. Therefore, in order to discretize $Y^{(n)}$ it is not necessary to approximate stochastic integrals. This allows to use a sparser time grid compared to the Euler method, for which the discretization of stochastic integrals is necessary. In particular, the theoretical speed of convergence with respect to the time-step is of order $\sqrt{\Delta}$ for Euler-Maruyama scheme and of order Δ for deterministic Euler, which is the scheme used to discretize the Lebesgue integrals in the Magnus expansion above. In the following numerical tests, we discretize in time with mesh Δ equal to 10^{-4} for `euler` and equal to $\sqrt{\Delta} = 10^{-2}$ for `m1`, `m2` and `m3`. Note that, as it is confirmed by the results in Table 1, choosing a finer time-discretization for `euler` (our reference method here) is essential in order to make it comparable with `m3`. Furthermore, in the example of Section 3.3, where an explicit solution is available, we show (see Tables 7 and 9) that choosing a sparser time-grid (say $\Delta = 10^{-3}$) the Euler-Maruyama method incurs a sensitive loss of precision.

It is also clear that the implementation is totally parallelizable, in that $Y^{(1)}$, $Y^{(2)}$ and $Y^{(3)}$ do not depend on each other and thus they can be computed in parallel. More importantly, the discretization of the integrals in each $Y^{(n)}$ can be parallelized as the latter are explicit and not implicitly defined through a differential equation.

We choose A and B at random and normalize them by their spectral norms. In particular, the results below refer to

$$A = \begin{pmatrix} 0.335302 & -0.645492 \\ -0.264419 & 0.634641 \end{pmatrix}, \qquad B = \begin{pmatrix} -0.0572262 & 0.0493763 \\ -0.665366 & 0.742744 \end{pmatrix}.$$

In Figure 1 we plot one realization of the trajectories of the top-left component $(X_t)_{11}$, computed with the methods above, up to time $t = 0.75$. In Table 1 we show the expectations $E[\mathrm{Err}_t]$ for different values of t, with `euler` as benchmark solution, computed via Monte Carlo simulation with 10^3 samples. The same samples are used in Figure 2 to plot the empirical CDF of Err_t. The computational times for the 10^3 sampled trajectories of X, up to time $t = 1$, computed with `m1`, `m2`, `m3` and `euler` are reported in Table 2. For the Magnus methods we separate the time to compute the approximate logarithm from the one to compute the matrix exponential.

Remark 3.1. We can see from Table 2 that the Magnus methods `m1`, `m2` and `m3` are significanty faster than `euler`. The reason for this is two-fold: on the one hand, we have the possibility of parallelizing the Magnus methods over both time and samples, while `euler` is
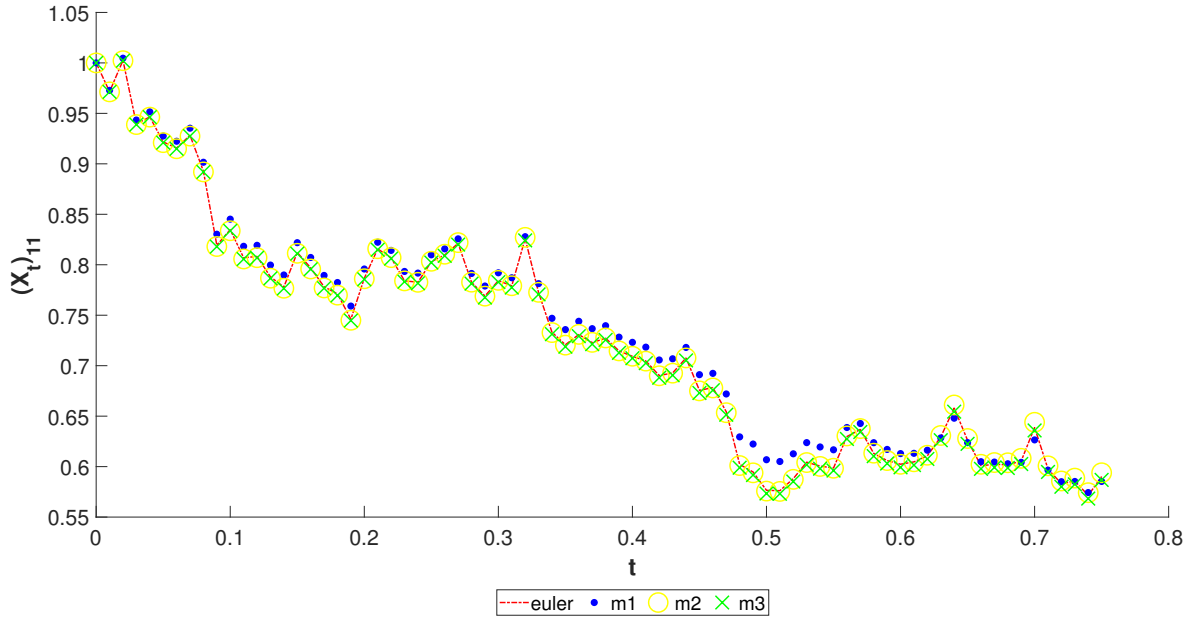
Figure 1: A and B constant. One realization of the trajectories of the top-left component $(X_t)_{11}$, computed with `euler`, `m1`, `m2`, `m3`.

Table 1: A and B constant. Values of $E[\text{Err}_t]$ (in percentage) for `m1`, `m2`, `m3`, with `euler` as benchmark solution, obtained with 10^3 samples.

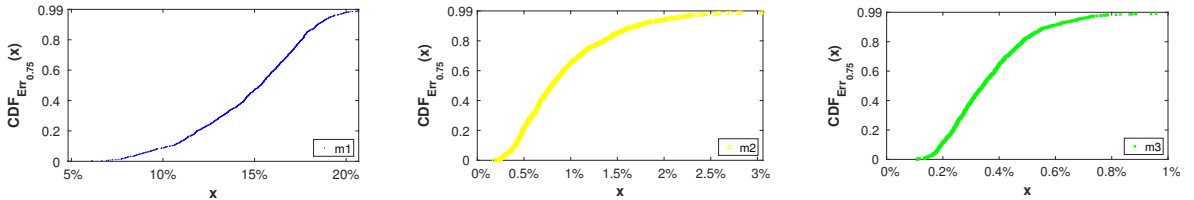| Method	$t = 0.25$	$t = 0.5$	$t = 0.75$	$t = 1$	$t = 2$	$t = 3$
Euler $\Delta = 10^{-4}$, Magnus $\Delta = 10^{-2}$						
`m1`	4.59 %	9.6 %	14.8 %	21.1 %	49.5 %	86.1 %
`m2`	0.217 %	0.503 %	0.951 %	1.55 %	5.29 %	10.9 %
`m3`	0.176 %	0.256 %	0.371 %	0.543 %	2.03 %	5.25 %

Figure 2: A and B constant. Empirical CDF of Err_t, at $t = 0.75$, for `m1`, `m2`, `m3`, with `euler` as benchmark solution, obtained with 10^3 samples.

22

Table 2: A and B constant. Computational times for 10^3 sampled trajectories of X, up to time $t = 1$, computed with m1, m2, m3 and euler.

Method	Log	Matrix Exp	Total
euler $\Delta = 10^{-4}$	0	0	6.7784
m1 $\Delta = 10^{-2}$	0.0093466	0.535331	0.544678
m2 $\Delta = 10^{-2}$	0.0221759	0.569511	0.591687
m3 $\Delta = 10^{-2}$	0.0475184	0.584098	0.631616

only parallelizable over all samples, and on the other hand, we can discretize the Magnus expansion with a time-step that is the square root of the one used for Euler-Maruyama, due to the different rates of convergence.

In our numerical experiments we already use 6 CPU cores to parallelize the computation of the matrix exponential on the CPU, while we use one GPU to compute the Magnus logarithm. For euler we speed up in each iteration the matrix multiplications by using pagefun on a GPU to parallelize over all samples. As for m1, m2 and m3, if we were to increase the number of CPU cores to, say, 12, we could see an approximate reduction in the computation time of matrix exponentiation by half (plus overhead), making it about 24 times as fast as the Euler method.

Now, the very nature of euler (see Example 3.2 together with Table 3) as an iterative scheme yields another advantage of the Magnus methods; namely, that the computation of the logarithm is very fast and if one needs only the solution of the SDE at the terminal time then one has to compute the matrix exponential only at a single time. Let us consider Table 2 for the moment. In this particular experiment it would mean that we can divide the computational time of the matrix exponentiation by approximately $\Delta^{-1} = 10^2$ without increasing the CPU core count. Hence, the Magnus methods would require approximately only 0.04 seconds plus effects from distributing the memory to the different processors. The euler method, in contrast, does not benefit from this because, as an iterative method, it must fully evaluate the trajectories.

Such situations are not uncommon; for example, in mathematical finance pricing a European call option depends only on the terminal time of the underlying process, giving the Magnus methods a tremendous advantage even without increasing CPUs or GPUs. We will illustrate such a situation in Example 3.2 together with Table 3. In calibration procedures, such as fitting a model to data at few points in time, the Magnus method also excels for the same reason.

Example 3.2. In this example we want to demonstrate the benefit, explained in Remark 3.1, of using the Magnus methods compared to iterative schemes, such as the Euler method, when calculating the first, second and third element-wise moments of the terminal value of a matrix-valued SDE. Precisely, we evaluate $E\big[((X_t)_{ij})^k\big]$ for $i, j = 1, \ldots, d$ and $k = 1, 2, 3$. We will keep the same parameters as in Table 1.

Table 3: A and B constant. Computational times and values of first, second and third moment at the terminal time $t = 1$ for `m1`, `m2`, `m3`, using $\Delta = 10^{-2}$, and `euler`, using $\Delta = 10^{-4}$, obtained with 10^3 samples.

Method	$E\left[((X_t)_{11})^k\right]$	$E\left[((X_t)_{12})^k\right]$	$E\left[((X_t)_{21})^k\right]$	$E\left[((X_t)_{22})^k\right]$	Total time
		First moment, $k = 1$			
euler	0.884995	0.136974	-0.913738	1.99784	6.83223
m1	1.23538	-0.510346	-1.38672	2.88552	0.11308
m2	0.92461	0.0488442	-0.889341	2.00131	0.16783
m3	0.886685	0.132748	-0.910886	1.9915	0.185966
		Second moment, $k = 2$			
euler	1.20982	1.09315	1.78757	7.06842	6.83263
m1	2.49141	3.21348	3.99804	15.4156	0.113674
m2	1.31038	1.18291	1.7166	7.07746	0.168869
m3	1.21421	1.09186	1.77727	7.00593	0.186955
		Third moment, $k = 3$			
euler	2.62519	-3.20706	-5.74199	40.8729	6.83229
m1	8.21939	-20.9025	-18.9891	136.058	0.1131
m2	2.95392	-4.05804	-5.44452	40.8636	0.168065
m3	2.6546	-3.27689	-5.70576	40.2687	0.185915

The results of this example are summarized in Table 3. In this table columns 2–4 contain the values of the element-wise moments at the terminal time of the solution to the SDE with constant coefficients starting with the upper left corner of the solution matrix, then the upper right, lower left and lower right, respectively. In the last column we present the computational times in seconds.

The values of the moments do not differ significantly between `euler` and `m3`, and remarkably the Magnus methods are roughly 35 times as fast in this particular example. We stress again at this point that a coarser time-grid for `euler` would not be comparable to the accuracy of `m3`.

In the interesting paper [12] a non-linear extension in the case of commuting A and B can be found and applications to SPDEs via space discretizations are discussed, which is the same approach we take in the next subsection with the ME.

3.2 Applications to SPDEs

The aim of this subsection is to apply the previously derived ME for the numerical solution of parabolic stochastic partial differential equations (SPDEs). We derive an approximation scheme for the general case of variable coefficients, which we only test in the case of the stochastic heat-equation (Example 3.4), for which an exact solution is available.

3.2.1 Stochastic Cauchy problem and fundamental solution

Let $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t \geq 0})$ be a filtered probability space endowed with a real Brownian motion W. We consider the stochastic Cauchy problem

$$\begin{cases} du_t(x) = \mathbf{L}_t u_t(x) dt + \mathbf{G}_t u_t(x) dW_t, & t > 0, \ x \in \mathbb{R}, \\ u_0 = \varphi, \end{cases}$$

where $\mathbf{L}_t$ is the elliptic linear operator acting as

$$\mathbf{L}_t u_t(x) = \frac{1}{2} \mathbf{a}_t(x) \partial_{xx} u_t(x) + \mathbf{b}_t(x) \partial_x u_t(x) + \mathbf{c}_t(x) u_t(x),$$

and $\mathbf{G}_t$ is the first-order linear operator acting as

$$\mathbf{G}_t u_t(x) = \sigma_t(x) \partial_x u_t(x) + \mathbf{g}_t(x) u_t(x).$$

The coefficients $(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{g}, \sigma)$ are random fields indexed by $(t, x) \in [0, \infty[\times \mathbb{R}$ and the initial datum φ is a random field on $\mathbb{R}$. A classical solution to (3.2.1) is understood here as a predictable and almost-surely continuous random field $u = u_t(x)$ over $[0, \infty[\times \mathbb{R}$, such that $u_t \in C^2(\mathbb{R})$ a.s. for any $t > 0$ and

$$u_t(x) = \varphi(x) + \int_0^t \mathbf{L}_\tau u_\tau(x) d\tau + \int_0^t \mathbf{G}_\tau u_\tau(x) dW_\tau, \qquad t \geq 0, \ x \in \mathbb{R}.$$

There is a vast literature on stochastic SPDEs and problems of the form (3.2.1), under suitable measurability, regularity and boundedness assumptions on the coefficients and on the initial datum: see, for instance, [18], [25], [9], [30] and the references therein.

Note that, in analogy with deterministic PDEs, the solution of the Cauchy problem (3.2.1) can be written, in some cases, as a convolution of the initial datum with a *stochastic fundamental solution* $p(t, x; 0, \xi)$, i.e.

$$u_t(x) = \int_{\mathbb{R}} p(t, x; 0, \xi) \varphi(\xi) d\xi, \qquad (t, x) \in]0, \infty[\times \mathbb{R},$$

with $p(t, x; 0, \xi)$ being a random field that solves the SPDE in (3.2.1) with respect to the variables (t, x) and which approximates a Dirac delta centered at ξ as t approaches 0.

3.2.2 Finite-difference Magnus scheme

We employ the stochastic ME to develop an approximation scheme for the Cauchy problem (3.2.1). Our goal here is only to hint at the possibility that the stochastic ME is a useful tool for the numerical solution of SPDEs. Therefore, we keep the exposition at a heuristic level and postpone the rigorous study of the problem for further research.

The idea is to apply finite-difference space-discretization for the operators $\mathbf{L}$ and $\mathbf{G}$, and then ME to solve the resulting linear (matrix-valued) Itô SDE. We fix a bounded interval $[a, b]$ and use the following notation: for a given $d \in \mathbb{N}$, we denote by ς_d a mesh of $d + 2$ equidistant points in $[a, b]$, i.e.

$$\varsigma_d = \{x_i^d \mid x_i^d = a + ih, \ i = 0, \dots, d + 1\}, \qquad h := \frac{b - a}{d + 1},$$

and for any random field $\mathbf{f}(x)$, $x \in \mathbb{R}$, we denote by $\mathbf{f}^d = (\mathbf{f}_0^d, \dots, \mathbf{f}_{d+1}^d)$ the random vector whose components correspond to $\mathbf{f}$ evaluated at the points of the mesh, namely

$$\mathbf{f}_i^d = \mathbf{f}(x_i^d), \qquad i = 0, \dots, d + 1.$$

Following the classical centered finite-difference discretization, we approximate the spatial derivatives in each point as

$$\partial_x u_t(x_i^d) \approx \frac{u_{t,i}^d - u_{t,i-1}^d}{h}, \quad \partial_{xx} u_t(x_i^d) \approx \frac{u_{t,i+1}^d - 2u_{t,i}^d + u_{t,i-1}^d}{h^2}, \qquad i = 1, \dots, d,$$

to obtain the system of Itô SDEs

$$\begin{cases} du_{t,i}^d = (\mathbf{L}_t^d u_t^d)_i dt + (\mathbf{G}_t^d u_t^d)_i dW_t, \\ u_{0,i}^d = \varphi_i^d, \end{cases}$$

for $i = 1, \cdots, d$, where $\mathbf{L}_t^d$ and $\mathbf{G}_t^d$ are now the operators acting as

$$(\mathbf{L}_t^d u_t^d)_i = \frac{1}{2}\mathbf{a}_{t,i}^d \frac{u_{i+1}^d - 2u_i^d + u_{i-1}^d}{h^2} + \mathbf{b}_{t,i}^d \frac{u_i^d - u_{i-1}^d}{h} + \mathbf{c}_{t,i}^d u_i^d,$$

$$(\mathbf{G}_t^d u_t^d)_i = \sigma_{t,i}^d \frac{u_i^d - u_{i-1}^d}{h} + \mathbf{g}_{t,i}^d u_i^d.$$

By imposing some boundary conditions, for instance

$$u_{t,0}^d = u_{t,d+1}^d = 0, \qquad t > 0,$$

the system of SDEs (3.2.2) can be cast in the framework of the previous section. More precisely, under condition (3.2.2), system (3.2.2) is equivalent to

$$\begin{cases} d\bar{u}_t^d = A_t^d \bar{u}_t^d dt + B_t^d \bar{u}_t^d dW_t, \\ \bar{u}_0^d = \bar{\varphi}^d, \end{cases}$$

where we set

$$\bar{u}_t^d = (u_{t,1}^d, \ldots, u_{t,d}^d), \qquad \bar{\varphi}^d = (\varphi_1^d, \ldots, \varphi_d^d),$$

and A_t^d, B_t^d are the random tridiagonal $(d \times d)$-matrices given by

$$A_t^d = \begin{pmatrix} -\frac{\mathbf{a}_{t,1}^d}{h^2} + \frac{\mathbf{b}_{t,1}^d}{h} + \mathbf{c}_{t,1}^d & \frac{1}{2}\frac{\mathbf{a}_{t,1}^d}{h^2} & \cdots & & 0 \\ \frac{1}{2}\frac{\mathbf{a}_{t,2}^d}{h^2} - \frac{\mathbf{b}_{t,2}^d}{h} & -\frac{\mathbf{a}_{t,2}^d}{h^2} + \frac{\mathbf{b}_{t,2}^d}{h} + \mathbf{c}_{t,2}^d & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \frac{1}{2}\frac{\mathbf{a}_{t,d-1}^d}{h^2} \\ 0 & \cdots & \frac{1}{2}\frac{\mathbf{a}_{t,d}^d}{h^2} - \frac{\mathbf{b}_{t,d}^d}{h} & -\frac{\mathbf{a}_{t,d}^d}{h^2} + \frac{\mathbf{b}_{t,d}^d}{h} + \mathbf{c}_{t,d}^d \end{pmatrix},$$

$$B_t^d = \begin{pmatrix} \frac{\sigma_{t,1}^d}{h} + \mathbf{g}_{t,1}^d & 0 & \cdots & & 0 \\ -\frac{\sigma_{t,2}^d}{h} & +\frac{\sigma_{t,2}^d}{h} + \mathbf{g}_{t,2}^d & \ddots & & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & -\frac{\sigma_{t,d}^d}{h} & +\frac{\sigma_{t,d}^d}{h} + \mathbf{g}_{t,d}^d \end{pmatrix}.$$

Now, the solution to (3.2.2) can be written as

$$\bar{u}_t^d = X_t^d \bar{u}_0^d, \qquad t \geq 0,$$

where X^d is in turn the solution to the $\mathcal{M}^{d \times d}$-valued Itô SDE

$$\begin{cases} dX_t^d = A_t^d X_t^d dt + B_t^d X_t^d dW_t, \\ X_0^d = I_d. \end{cases}$$

Remark 3.3. The components of X^d can be regarded as approximations of the integrals of the fundamental solution of the SPDE in (3.2.1), when it exists, on each sub-interval $[\frac{1}{2}(x_{j-1}^d + x_j^d), \frac{1}{2}(x_j^d + x_{j+1}^d)]$, namely

$$(X_t^d)_{i,j} \approx \int_{\frac{1}{2}(x_{j-1}^d + x_j^d)}^{\frac{1}{2}(x_j^d + x_{j+1}^d)} p(t, x_i^d; 0, \xi) d\xi =: (\mathcal{I}_t^d)_{i,j}, \qquad i, j = 1, \ldots, d.$$

Example 3.4. We consider a special case of (3.2.1) with $\mathbf{a}_t \equiv \mathbf{a} > 0$, $\mathbf{b}, \mathbf{c}, \mathbf{g} \equiv 0$ and $\sigma_t \equiv \sigma > 0$. Hence, we consider the stochastic heat equation

$$du_t = \frac{\mathbf{a}}{2} \partial_{xx} u_t(x) dt + \sigma \partial_x u_t(x) dW_t, \qquad t > 0, \ x \in \mathbb{R},$$

with $\mathbf{a} > \sigma^2$, whose stochastic fundamental solution is given explicitly by

$$p(t, x; 0, \xi) := \frac{1}{\sqrt{2\pi(\mathbf{a} - \sigma^2)t}} \exp\left(-\frac{(x + \sigma W_t - \xi)^2}{2(\mathbf{a} - \sigma^2)t}\right), \qquad t > 0, \ x, \xi \in \mathbb{R}.$$

The matrices A_t^d and B_t^d in (3.2.2) now read as

$$A_t^d \equiv \frac{\mathbf{a}}{h^2} \begin{pmatrix} -1 & \frac{1}{2} & \cdots & 0 \\ \frac{1}{2} & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{1}{2} \\ 0 & \cdots & \frac{1}{2} & -1 \end{pmatrix}, \qquad B_t^d \equiv \frac{\sigma}{h} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & -1 & 1 \end{pmatrix}.$$

In particular, they do not commute and are constant for fixed d.

Table 4: SPDE. Values of $\mathbb{E}[\mathrm{Err}_t^d]$ (in percentage) for `m1` and `m3`, with $d = 50$, obtained with 50 independent samples. Parameters as in (3.4).

Method	$t = 0.1$	$t = 0.2$	$t = 0.3$	$t = 0.4$	$t = 0.5$
`euler`	9.1364 %	5.5467 %	5.3231 %	4.8377 %	4.5829 %
`m1`	8.0746 %	5.3243 %	4.9617 %	4.7273 %	5.3065 %
`m3`	9.1337 %	5.5296 %	5.3310 %	4.8314 %	4.5704 %

Table 5: SPDE. Values of $\mathbb{E}[\mathrm{Err}_t^d]$ (in percentage) for `m1` and `m3`, with $d = 100$, obtained with 50 independent samples. Parameters as in (3.4).

Method	$t = 0.1$	$t = 0.2$	$t = 0.3$	$t = 0.4$	$t = 0.5$
euler	$4.2053\,\%$	$3.4600\,\%$	$2.8214\,\%$	$2.3524\,\%$	$2.0370\,\%$
m1	$4.4452\,\%$	$5.1232\,\%$	$4.7061\,\%$	$4.8807\,\%$	$4.8397\,\%$
m3	$4.2576\,\%$	$3.4543\,\%$	$2.8172\,\%$	$2.3598\,\%$	$2.0467\,\%$

Table 6: SPDE. Values of $\mathbb{E}[\mathrm{Err}_t^d]$ (in percentage) for `m1` and `m3`, with $d = 200$, obtained with 50 independent samples. Parameters as in (3.4).

Method	$t = 0.1$	$t = 0.2$	$t = 0.3$	$t = 0.4$	$t = 0.5$
euler	$2.1832\,\%$	$1.4403\,\%$	$1.4190\,\%$	$1.2174\,\%$	$1.1532\,\%$
m1	$4.9891\,\%$	$5.0444\,\%$	$4.8249\,\%$	$5.0042\,\%$	$5.2603\,\%$
m3	$2.1690\,\%$	$1.4364\,\%$	$1.4467\,\%$	$1.2140\,\%$	$1.1420\,\%$

In the next numerical test we compare the approximate solutions to (3.2.2), obtained with the stochastic ME in the special case of constant coefficients (3.1), with the $\mathcal{M}^{d \times d}$-valued stochastic process $\mathcal{I}^d$, whose components are given by the integral in (3.3) with p as in (3.4). In doing this, we shall keep in mind that the difference between the latter quantities can be decomposed into two errors, namely: the one between X^d and its approximation, and the one between X^d and $\mathcal{I}^d$. In turn, the latter is the result of both space-discretization and the error that stems by imposing null boundary conditions (see (3.2.2)). In particular, this last error cannot be reduced by refining the space-grid. Therefore, the analysis should be restricted to the "central" components of $\mathcal{I}^d$, namely those which do not depend on the values of the fundamental solution in the vicinity of the boundary $\{a, b\}$. This motivates the definition that follows. For a given $\kappa \in \mathbb{N}$ with $\kappa < d$, and a given approximation $X^{d,\mathrm{app}}$ of X^d, we define the process

$$\mathrm{Err}_t^d := \frac{\|\widetilde{\mathcal{I}}_t^{d,\kappa} - \widetilde{X}_t^{d,\kappa,\mathrm{app}}\|_F}{\|\widetilde{\mathcal{I}}_t^{d,\kappa}\|_F}, \qquad t > 0,$$

where $\widetilde{\mathcal{I}}_t^{d,\kappa}$ and $\widetilde{X}_t^{d,\kappa,\mathrm{app}}$ are the projections on $\mathcal{M}^{\kappa \times d}$ obtained by selecting the central κ rows of $\mathcal{I}_t^d$ and $X_t^{d,\mathrm{app}}$, respectively. The matrix norm above is the Frobenius norm. The role of $X^{d,\mathrm{app}}$ will be played by the time-discretization of the truncated ME (1.1)-(1.1). In particular, we will denote by `m1` and `m3` the discretized first and third-order MEs of X^d, respectively. We will not consider the second-order Magnus approximation `m2` as it appears less stable than the others. Note that, being A_t^d and B_t^d constant matrices, the first three terms of the ME are given explicitly by (3.1).

In the numerical experiments we set

$$a = -2, \ b = 2, \quad \mathbf{a} = 0.2, \quad \sigma = 0.15.$$

Setting the parameter κ in (3.4), which determines the number of rows that are taken into account to asses the error, as $\kappa = \lfloor d/2 \rfloor$, we study the expectation of Err_t^d up to $t = 0.5$. Such choice for κ and t allows us to study the error in a region that is suitably away from the boundary. Indeed, choosing κ as above implies x_i^d in (3.3) ranging roughly from -1 to 1. On the other hand, the standard-deviation parameter associated to the Gaussian density (3.4) at $t = 0.5$ is roughly 0.30, while the mean parameter is $0.15 \times W_{0.5}$, whose standard deviation is in turn roughly 0.10. Therefore, both $(\widetilde{\mathcal{I}}_t^{d,\kappa})_{i,1}$ and $(\widetilde{\mathcal{I}}_t^{d,\kappa})_{i,d}$ are likely to be very close to zero, thus meeting the null boundary condition implied by (3.2.2).

In Tables 4, 5, 6, we report the approximate values of $\mathbb{E}[\text{Err}_t^d]$ for $d = 50, 100$ and 200, respectively. These were obtained via simulation of 50 trajectories of W with time step-size $\Delta = 10^{-4}$. Now, let us inspect the Tables 4–6 in more detail. As a reminder, these results were obtained by using the `exact` solution as a reference, which is available in this particular example. In Table 4 it is noticeable that `euler` and `m3` can exhibit worse results for small times compared to `m1`. This is due to the coarse space approximation with only 52 space grid points. Increasing the number d of grid points improves the error of `euler` and `m3` for all displayed times, which can be seen in Tables 5 and 6 by comparing each column for the same final time. Finally, notice that the third-order Magnus expansion has the same magnitude of error as the `euler` scheme for all final times.

3.3 Example: $j = 1$, $B = 0$ and A_t upper triangular.

We now test the ME on an SDE with time-dependent coefficients and with known explicit solution. Set

$$A_t = \begin{pmatrix} 2 & t \\ 0 & -1 \end{pmatrix}, \qquad B_t \equiv 0.$$

In this case (1) admits an explicit solution, which can be obtained by using Itô's formula, given by

$$X_t = \begin{pmatrix} e^{2(W_t - t)} & e^{2(W_t - t)} \left(\int_0^t s\, e^{-3W_s + \frac{3}{2}s}\, dW_s - 2 \int_0^t s\, e^{-3W_s + \frac{3}{2}s}\, ds \right) \\ 0 & e^{-(W_t + \frac{1}{2}t)} \end{pmatrix}.$$

Table 7: B_t and A_t as in (3.3). Values of $E[\mathrm{Err}_t]$ (in percentage) for `euler`, `m1`, `m2`, `m3`, with `exact` as benchmark solution, obtained with 10^3 samples.

Method	$t = 0.25$	$t = 0.5$	$t = 0.75$	$t = 1$	$t = 2$	$t = 3$
`euler` $\Delta = 10^{-4}$	0.503 %	0.716 %	0.883 %	1.08 %	2.07 %	3.3 %
`euler` $\Delta = 10^{-3}$	1.56 %	2.12 %	2.56 %	2.92 %	4.16 %	5.6 %
`m1` $\Delta = 10^{-2}$	19.6 %	42.9 %	70.8 %	108 %	409 %	1490 %
`m2` $\Delta = 10^{-2}$	0.147 %	0.659 %	1.71 %	3.36 %	11.3 %	20.6 %
`m3` $\Delta = 10^{-2}$	0.117 %	0.309 %	0.77 %	1.61 %	6.44 %	13.9 %

The first three terms of the ME read as

$$Y_t^{(1)} = \begin{pmatrix} 2W_t & tW_t - \int_0^t W_s ds \\ 0 & -W_t \end{pmatrix},$$

$$Y_t^{(2)} = -\frac{1}{2}\begin{pmatrix} 4t & \frac{1}{2}t^2 \\ 0 & t \end{pmatrix} - \frac{3}{2}\begin{pmatrix} 0 & W_t \int_0^t W_u du - \int_0^t W_u^2 du \\ 0 & 0 \end{pmatrix},$$

$$Y_t^{(3)} = \begin{pmatrix} 0 & \frac{3}{4}(t - W_t^2)\int_0^t W_u du - \frac{3}{2}\int_0^t W_u u du \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & \frac{9}{4}W_t \int_0^t W_u^2 du - \frac{3}{2}\int_0^t W_u^3 du + \frac{3}{8}t^2 W_t \\ 0 & 0 \end{pmatrix}.$$

Again, all the stochastic integrals appearing in the ME can be solved in terms of Lebesgue integrals by using Itô's formula, which allows us to use one more time a sparser time grid compared to the Euler method and the discretized exact solution. In the following numerical tests, we discretize in time with mesh Δ equal to 10^{-4} for `exact` and equal to 10^{-2} for `m1`, `m2` and `m3`. For `euler` we run two experiments with mesh equal to 10^{-4} and 10^{-3}. Note that `euler` serves here as an alternative approximation and that choosing a finer time-discretization for `euler` and `exact` (our reference method here) is again essential in order to make them comparable with `m3`.

In Table 7 we show the expectations $E[\mathrm{Err}_t]$ for different values of t, with `exact` as benchmark solution, computed via Monte Carlo simulation with 10^3 samples. The same samples are used in Figure 3 to plot the empirical CDF of Err_t. It is clear from the results that the time-step size $\Delta = 10^{-3}$ is not small enough in order for `euler` to yield accurate results. Also note that `m3` outperforms `euler` with $\Delta = 10^{-4}$ up to $t = 0.75$.

The computational time for 10^3 sampled trajectories, up to time $t = 1$, which is given in Table 8, is approximately 0.7 seconds for `exact`, 4.6 seconds for `euler` and 0.6 seconds for either `m1`, `m2` or `m3`. The latter, however, is divided as follows: nearly 0.05 seconds to compute the ME and nearly 0.55 seconds to compute the matrix exponential with the Matlab function `expm`. Let us recall Remark 3.1 and note that the computation of the logarithm via

Table 8: Computational times (seconds) for 10^3 sampled trajectories of X, up to time $t = 1$, choosing $\Delta = 10^{-2}$ for `m1`, `m2` and `m3`, and $\Delta = 10^{-4}$ for `euler` and `exact`.

	B_t and A_t as in (3.3)			B_t and A_t as in (3.3) normalized by its spectral norm		
Method	Log	Matrix Exp	Total	Log	Matrix Exp	Total
`exact`	0	0	0.70544	0	0	0.807434
`euler`	0	0	4.59658	0	0	4.53311
`m1`	0.0186238	0.51252	0.531143	0.0128825	0.522425	0.535308
`m2`	0.0245188	0.517689	0.542207	0.0557674	0.527644	0.583412
`m3`	0.0441915	0.530973	0.575165	0.212351	0.471555	0.683906

ME is very fast thanks to the possibility of parallelizing the computation of the integrals in $Y^{(1)}, Y^{(2)}$ and $Y^{(3)}$.

As it appears in the results above, the accuracy of the ME quickly deteriorates as the time increases. This is largely due the fact that the spectral norm

$$\|A_t\| = \sqrt{\frac{1}{2}\left(t^2 + \sqrt{t^4 + 10t^2 + 9} + 5\right)}$$

is an increasing function of t. This behavior shall not come as a surprise, since the proof of Theorem 1.1 already uncovered the relation between the convergence time τ and the spectral norms of A_t and B_t. Such relation is also consistent with the convergence condition (1.1) that holds in the deterministic case. In order to asses numerically the impact of the spectral norm of A_t on the quality of the Magnus approximation, we now repeat the experiments on the equation obtained by normalizing A_t as in (3.3) with respect to $\|A_t\|$. As it turns out, the accuracy of `m1`, `m2` and `m3` improves considerably with this normalization. Note that, in this case, (1) no longer admits a closed-form solution, while the representation for the terms $Y^{(1)}, Y^{(2)}$ and $Y^{(3)}$ in the ME is omitted for it becomes rather tedious to write. In Figure 4 we plot one realization of the trajectories of the top-right component $(X_t)_{12}$, computed with all the methods above, up to time $t = 10$. In this case we did not plot a diagonal component of the solution because the latter are exact for `m2` and `m3`, up to discretization errors of Lebesgue integrals. Table 9 and Figure 5 are analogous to Table 7 and Figure 3 and are obtained again with 10^3 independent samples. The computational times, reported in Table 8, are comparable with those of the non normalized case. The same can said about the accuracy results reported in Table 9, which are comparable with those in Table 7.
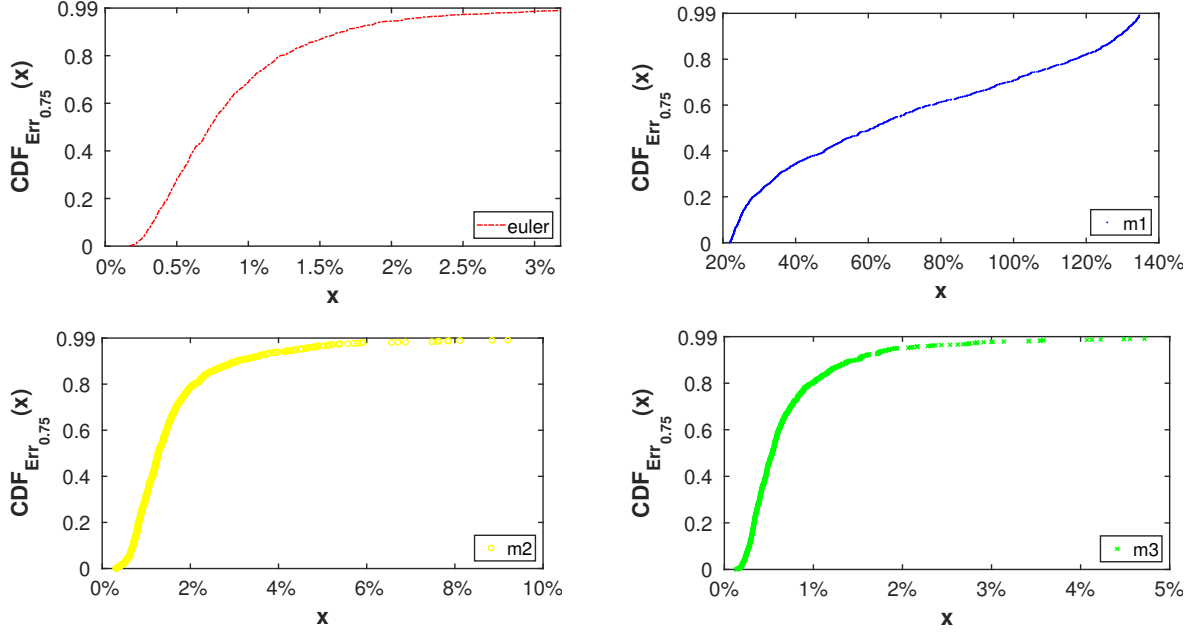
Figure 3: $B \equiv 0$ and A_t as in (3.3). Empirical CDF of Err_t, at $t = 0.75$, for `euler`, `m1`, `m2`, `m3`, with `exact` as benchmark solution, obtained with 10^3 samples.

Table 9: B_t and A_t as in (3.3) normalized by its spectral norm. Values of $E[\mathrm{Err}_t]$ (in percentage) for `euler`, `m1`, `m2`, `m3`, with `exact`, using $\Delta = 10^{-4}$, as benchmark solution, obtained with 10^3 samples.

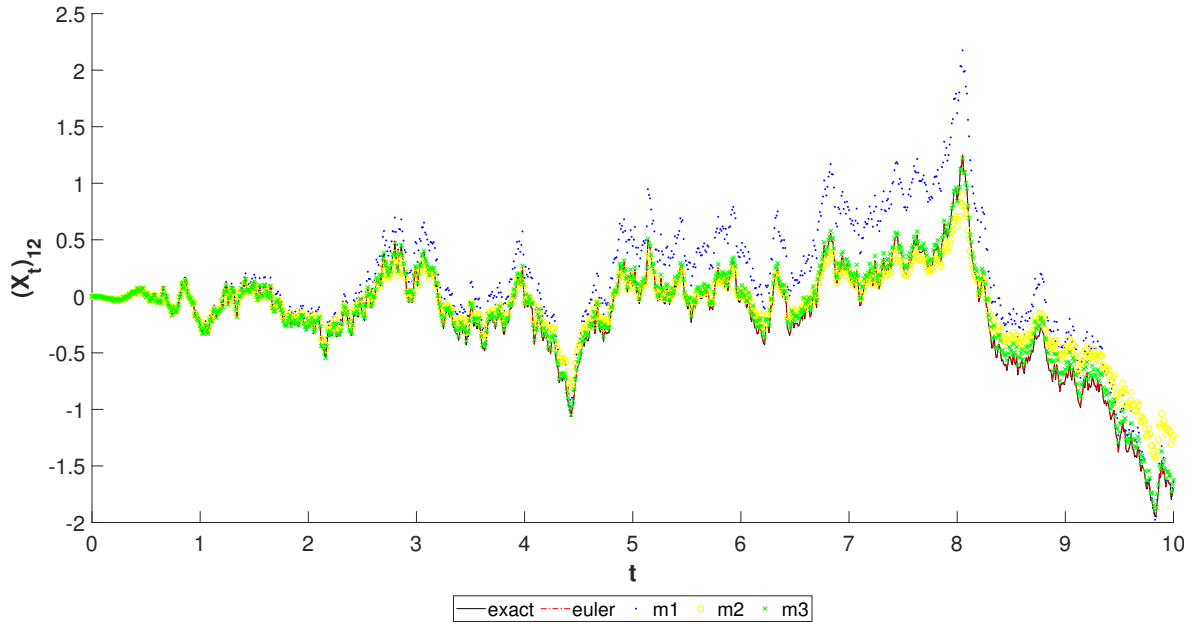| Method	$t = 0.25$	$t = 0.5$	$t = 0.75$	$t = 1$	$t = 2$	$t = 3$	$t = 10$
`euler` $\Delta = 10^{-4}$	0.134%	0.191%	0.227%	0.263%	0.382%	0.489%	0.715%
`euler` $\Delta = 10^{-3}$	0.416%	0.572%	0.691%	0.777%	0.965%	1.12%	1.35%
`m1` $\Delta = 10^{-2}$	4.69%	9.25%	13.6%	18.1%	34%	47%	90.9%
`m2` $\Delta = 10^{-2}$	0.047%	0.109%	0.263%	0.529%	2.01%	3.92%	15.2%
`m3` $\Delta = 10^{-2}$	0.0537%	0.0821%	0.126%	0.201%	0.642%	1.28%	5.31%

Figure 4: B_t and A_t as in (3.3) normalized by its spectral norm. One realization of the trajectories of the top-right component $(X_t)_{12}$, computed with `exact`, `euler`, `m1`, `m2`, `m3`.

Declarations

Funding

Conflicts of interests

The authors have no relevant financial or non-financial interests to disclose.

Data availability

All data generated or analysed during this study are included in this published article. In particular the code to produce the numerical experiments is available at
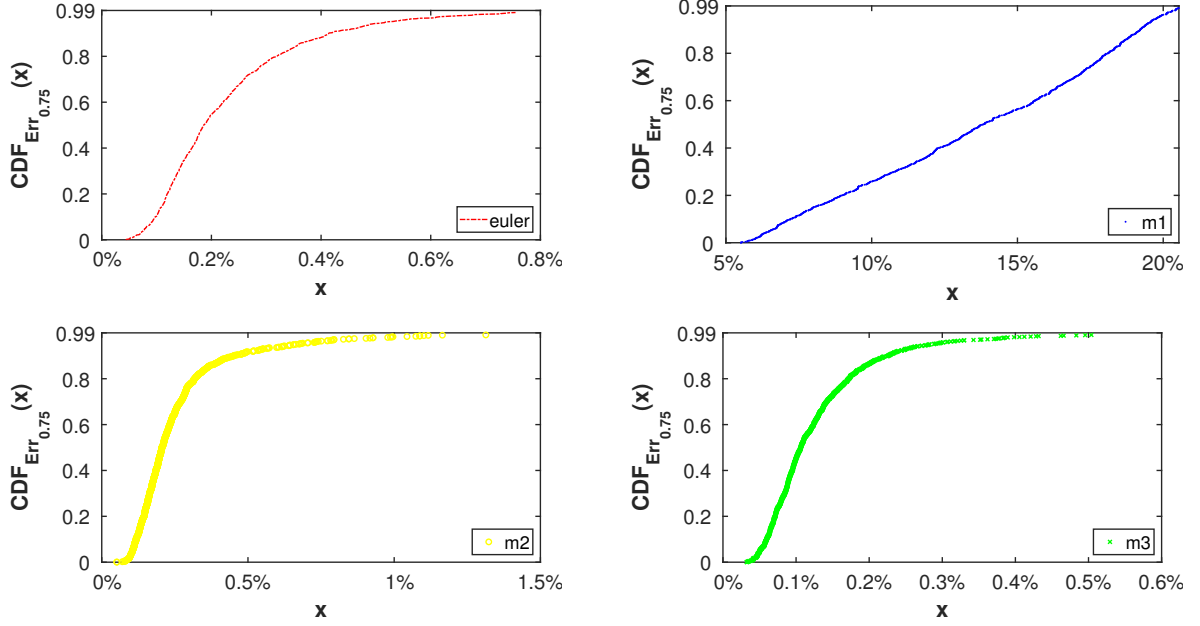https://github.com/kevinkamm/StochasticMagnusExpansion.

Figure 5: B_t and A_t as in (3.3) normalized by its spectral norm. Empirical CDF of Err_t, at $t = 0.75$, for `euler`, `m1`, `m2`, `m3`, with `exact` as benchmark solution, obtained with 10^3 samples.

References

[1] Azencott, R. Formule de Taylor stochastique et développement asymptotique d'intégrales de Feynman. In *Seminar on Probability, XVI, Supplement*, vol. 921 of *Lecture Notes in Math.* Springer, Berlin-New York, 1982, pp. 237–285.

[2] Ben Arous, G. Flots et séries de Taylor stochastiques. *Probab. Theory Related Fields 81*, 1 (1989), 29–77.

[3] Blanes, S., Casas, F., Oteo, J. A., and Ros, J. The Magnus expansion and some of its applications. *Phys. Rep. 470*, 5-6 (2009), 151–238.

[4] Brezis, H. *Functional analysis, Sobolev spaces and partial differential equations.* Universitext. Springer, New York, 2011.

[5] Burrage, K., and Burrage, P. High strong order methods for non-commutative stochastic ordinary differential equation systems and the Magnus formula. *Physica D: Nonlinear Phenomena 133*, 1-4 (1999), 34–48.

[6] Castell, F. Asymptotic expansion of stochastic flows. *Probab. Theory Related Fields 96*, 2 (1993), 225–239.

[7] Castell, F., and Gaines, J. An efficient approximation method for stochastic differential equations by means of the exponential Lie series. *Math. Comput. Simulation 38*, 1-3 (1995), 13–19. Probabilités numériques (Paris, 1992).

[8] Castell, F., and Gaines, J. The ordinary differential equation approach to asymptotically efficient schemes for solution of stochastic differential equations. *Ann. Inst. H. Poincaré Probab. Statist. 32*, 2 (1996), 231–250.

[9] Chow, P.-L. *Stochastic partial differential equations*, second ed. Advances in Applied Mathematics. CRC Press, Boca Raton, FL, 2015.

[10] Correales, A., and Escudero, C. Ito vs Stratonovich in the presence of absorbing states, 2018.

[11] Doss, H. Liens entre équations différentielles stochastiques et ordinaires. *Ann. Inst. H. Poincaré Sect. B (N.S.) 13*, 2 (1977), 99–125.

[12] Erdogan, U., and Lord, G. J. A new class of exponential integrators for SDEs with multiplicative noise. *IMA Journal of Numerical Analysis 39*, 2 (03 2018), 820–846.

[13] Fliess, M., and Normand-Cyrot, D. Algèbres de Lie nilpotentes, formule de Baker-Campbell-Hausdorff et intégrales itérées de K. T. Chen. In *Seminar on Probability, XVI*, vol. 920 of *Lecture Notes in Math.* Springer, Berlin-New York, 1982, pp. 257–267.

[14] Friz, P. K., Hager, P., and Tapia, N. Unified signature cumulants and generalized Magnus Expansions. arXiv:2102.03345.

[15] Kamm, K. *PhD thesis - Doctorate in Mathematics - University of Bologna*. In preparation.

[16] Kloeden, P. E., and Platen, E. *Numerical solution of stochastic differential equations*, vol. 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[17] Krylov, N. V. *Controlled diffusion processes*, vol. 14. Springer Science & Business Media, 2008.

[18] Krylov, N. V., and Rozovskii, B. L. The Cauchy problem for linear stochastic partial differential equations. *Izv. Akad. Nauk SSSR Ser. Mat. 41*, 6 (1977), 1329–1347, 1448.

[19] Kunita, H. On the representation of solutions of stochastic differential equations. In *Seminar on Probability, XIV (Paris, 1978/1979) (French)*, vol. 784 of *Lecture Notes in Math.* Springer, Berlin, 1980, pp. 282–304.

[20] Kunita, H. *Stochastic flows and jump-diffusions*, vol. 92 of *Probability Theory and Stochastic Modelling*. Springer, Singapore, 2019.

[21] Kuo, H.-H. *Introduction to stochastic integration*. Universitext. Springer, New York, 2006.

[22] Lord, G., Malham, S. J. A., and Wiese, A. Efficient strong integrators for linear stochastic systems. *SIAM J. Numer. Anal. 46*, 6 (2008), 2892–2919.

[23] Magnus, W. On the exponential solution of differential equations for a linear operator. *Comm. Pure Appl. Math. 7* (1954), 649–673.

[24] Marjanovic, G., and Solo, V. Numerical methods for stochastic differential equations in matrix Lie groups made simple. *IEEE Trans. Automat. Control 63*, 12 (2018), 4035–4050.

[25] Mikulevicius, R. On the Cauchy problem for parabolic SPDEs in Hölder classes. *Ann. Probab. 28*, 1 (2000), 74–103.

[26] Moan, P. C., and Niesen, J. Convergence of the Magnus series. *Found. Comput. Math. 8*, 3 (2008), 291–301.

[27] Muniz, M., Ehrhardt, M., Günther, M., and Winkler, R. Higher Strong Order Methods for Itô SDEs on Matrix Lie Groups. arXiv:2102.04131.

[28] Newton, N. J. Asymptotically efficient Runge-Kutta methods for a class of Itô and Stratonovich equations. *SIAM J. Appl. Math. 51*, 2 (1991), 542–567.

[29] Pascucci, A. *PDE and martingale methods in option pricing*, vol. 2 of *Bocconi & Springer Series*. Springer, Milan; Bocconi University Press, Milan, 2011.

[30] Pascucci, A., and Pesce, A. The parametrix method for parabolic spdes. *Stochastic Processes and their Applications 130*, 10 (2020), 6226–6245.

[31] Sussmann, H. J. Product expansions of exponential Lie series and the discretization of stochastic differential equations. In *Stochastic differential systems, stochastic control theory and applications (Minneapolis, Minn., 1986)*, vol. 10 of *IMA Vol. Math. Appl.* Springer, New York, 1988, pp. 563–582.

[32] Wang, Z., Ma, Q., Yao, Z., and Ding, X. The Magnus Expansion for Stochastic Differential Equations. *J. Nonlinear Sci. 30*, 1 (2020), 419–447.

[33] Yamato, Y. Stochastic differential equations and nilpotent Lie algebras. *Z. Wahrsch. Verw. Gebiete 47*, 2 (1979), 213–229.

[34] YANG, G., BURRAGE, K., KOMORI, Y., BURRAGE, P., AND DING, X. A class of new magnus-type methods for semi-linear non-commutative itô stochastic differential equations. *Numerical Algorithms* (Mar 2021).