

EFFICIENT STRONG INTEGRATORS FOR LINEAR STOCHASTIC SYSTEMS

GABRIEL LORD*, SIMON J.A. MALHAM*†, AND ANKE WIESE*

Abstract. We present numerical schemes for the strong solution of linear stochastic differential equations driven by an arbitrary number of Wiener processes. These schemes are based on the Neumann (stochastic Taylor) and Magnus expansions. Firstly, we consider the case when the governing linear diffusion vector fields commute with each other, but not with the linear drift vector field. We prove that numerical methods based on the Magnus expansion are more accurate in the mean-square sense than corresponding stochastic Taylor integration schemes. Secondly, we derive the maximal rate of convergence for arbitrary multi-dimensional stochastic integrals approximated by their conditional expectations. Consequently, for general nonlinear stochastic differential equations with non-commuting vector fields, we deduce explicit formulae for the relation between error and computational costs for methods of arbitrary order. Thirdly, we consider the consequences in two numerical studies, one of which is an application arising in stochastic linear-quadratic optimal control.

Key words. linear stochastic differential equations, strong numerical methods, Magnus expansion, stochastic linear-quadratic control

AMS subject classifications. 60H10, 60H35, 93E20

1. Introduction. We are interested in designing efficient numerical schemes for the strong approximation of linear Stratonovich stochastic differential equations of the form

$$y_t = y_0 + \sum_{i=0}^d \int_0^t a_i(\tau) y_\tau dW_\tau^i, \quad (1.1)$$

where $y \in \mathbb{R}^p$, $W^0 \equiv t$, (W^1, \dots, W^d) is a d -dimensional Wiener process and $a_0(t)$ and $a_i(t)$ are given $p \times p$ coefficient matrices. We call ‘ $a_0(t)y$ ’ the linear *drift* vector field and ‘ $a_i(t)y$ ’ for $i = 1, \dots, d$ the linear *diffusion* vector fields. We can express the stochastic differential equation (1.1) more succinctly in the form

$$y = y_0 + K \circ y, \quad (1.2)$$

where $K \equiv K_0 + K_1 + \dots + K_d$ and $(K_i \circ y)_t \equiv \int_0^t a_i(\tau) y_\tau dW_\tau^i$. The solution of the integral equation for y is known as the *Neumann series*, *Peano–Baker series*, *Feynman–Dyson path ordered exponential* or *Chen–Fleiss series*

$$y_t = (I - K)^{-1} \circ y_0 \equiv (I + K + K^2 + K^3 + \dots) \circ y_0.$$

The flow-map or fundamental solution matrix S_t maps the initial data y_0 to the solution $y_t = S_t y_0$ at time $t > 0$. It satisfies an analogous matrix valued stochastic differential equation to (1.2) with the $p \times p$ identity matrix as initial data. The logarithm of the Neumann expansion for the flow-map is the *Magnus expansion*. We can thus write the solution to the stochastic differential equation (1.1) in the form

$$y_t = (\exp \sigma_t) y_0,$$

*Maxwell Institute for Mathematical Sciences and School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK (G.J.Lord@hw.ac.uk, S.J.Malham@ma.hw.ac.uk, A.Wiese@hw.ac.uk). (21/8/2007)

†SJAM would like to dedicate this paper to the memory of Nairo Aparicio, a friend and collaborator who passed away on 20th June 2005.

where

$$\sigma_t = \ln((I - K)^{-1} \circ I) \equiv K \circ I + K^2 \circ I - \frac{1}{2}(K \circ I)^2 + \dots \quad (1.3)$$

See Magnus [41], Kunita [35], Azencott [3], Ben Arous [4], Strichartz [54], Castell [12], Burrage [7], Burrage and Burrage [8] and Baudoin [6] for the derivation and convergence of the original and also stochastic Magnus expansion; Iserles, Munthe-Kaas, Nørsett and Zanna [30] for a deterministic review; Lyons [38] and Sipiläinen [53] for extensions to rough signals; Lyons and Victoir [40] for a recent application to probabilistic methods for solving partial differential equations; and Sussmann [56] for a related product expansion.

In the case when the coefficient matrices $a_i(t) = a_i$, $i = 0, \dots, d$ are constant and non-commutative, the solution to the linear problem (1.1) is non-trivial and given by the Neumann series or stochastic Taylor expansion (see Kloeden and Platen [33])

$$y_t^{\text{neu}} = \sum_{\ell=0}^{\infty} \sum_{\alpha \in \mathbb{P}_\ell} J_{\alpha_\ell \dots \alpha_1}(t) a_{\alpha_1} \dots a_{\alpha_\ell} y_0, \quad (1.4)$$

where

$$J_{\alpha_\ell \dots \alpha_1}(t) \equiv \int_0^t \int_0^{\xi_1} \dots \int_0^{\xi_{\ell-1}} dW_{\xi_\ell}^{\alpha_\ell} \dots dW_{\xi_2}^{\alpha_2} dW_{\xi_1}^{\alpha_1}.$$

Here \mathbb{P}_ℓ is the set of all combinations of multi-indices $\alpha = \{\alpha_1, \dots, \alpha_\ell\}$ of length ℓ with $\alpha_k \in \{0, 1, \dots, d\}$ for $k = 1, \dots, \ell$. There are some special non-commutative cases when we can write down an explicit analytical solution. For example the stochastic differential equation $dy_t = a_1 y_t dW_t^1 + y_t a_2 dW_t^2$ with the identity matrix as initial data has the explicit analytical solution $y_t = \exp(a_1 W_t^1) \cdot \exp(a_2 W_t^2)$. However in general we cannot express the Neumann solution series (1.4) in such a closed form.

Classical numerical schemes such as the Euler-Maruyama and Milstein methods correspond to truncating the stochastic Taylor expansion to generate global strong order 1/2 and order 1 schemes, respectively. Stochastic Runge-Kutta numerical methods have also been derived—see Kloeden and Platen [33] and Talay [57]. At the linear level, the Neumann, stochastic Taylor and Runge-Kutta type methods are equivalent. In the stochastic context, Magnus integrators have been considered by Castell and Gaines [13], Burrage [7], Burrage and Burrage [8] and Misawa [46].

We present numerical schemes based on truncated Neumann and Magnus expansions. Higher order multiple Stratonovich integrals are approximated across each time-step by their expectations conditioned on the increments of the Wiener processes on suitable subdivisions (see Newton [49] and Gaines and Lyons [22]). What is new in this paper is that we:

1. Prove the strong convergence of the truncated stochastic Magnus expansion for small stepsize;
2. Derive uniformly accurate higher order stochastic integrators based on the Magnus expansion in the case of commuting linear diffusion vector fields;
3. Prove the maximal rate of convergence for arbitrary multi-dimensional stochastic integrals approximated by their conditional expectations;
4. Derive explicit formulae for the relation between error and computational costs for methods of arbitrary order in the case of general nonlinear, non-commuting governing vector fields.

Our results can be extended to nonlinear stochastic differential equations with analogous conditions on the governing nonlinear vector fields, where the exponential Lie series (replacing the Magnus expansion) can be evaluated using the Castell–Gaines approach.

In the first half of this paper, sections 2–5, we focus on proving the convergence of the truncated Magnus expansion and establishing Magnus integrators that are more accurate than Neumann (stochastic Taylor) schemes of the same order. The numerical schemes we present belong to the important class of *asymptotically efficient* schemes introduced by Newton [49]. Such schemes have the optimal minimum leading error coefficient among all schemes that depend on increments of the underlying Wiener process only. Castell and Gaines [13, 14] prove that the order 1/2 Magnus integrator driven by a d -dimensional Wiener process and a modified order 1 Magnus integrator driven by a 1-dimensional Wiener process are asymptotically efficient. We extend this result of Castell and Gaines to an arbitrary number of driving Wiener processes. We prove that if we assume the linear diffusion vector fields commute, then an analogously modified order 1 Magnus integrator and a new order 3/2 Magnus integrator are globally more accurate than their corresponding Neumann integrators.

There are several potential sources of cost contributing to the overall computational effort of a stochastic numerical integration scheme. The main ones are the efforts associated with:

- Evaluation: computing (and combining) the individual terms and special functions such as the matrix exponential;
- Quadrature: the accurate representation of multiple Stratonovich integrals.

There are usually fewer terms in the Magnus expansion compared to the Neumann expansion to the same order, but there is the additional computational expense of computing the matrix exponential. When the cost of computing the matrix exponential is not significant, due to their superior accuracy we expect Magnus integrators to be preferable to classical stochastic numerical integrators. This will be the case for systems that are small (see Moler and Van Loan [47] and Iserles and Zanna [31]) or for large systems when we only have to compute the exponential of a large sparse matrix times given vector data for which we can use Krylov subspace methods (see Moler and Van Loan [47] and Sidje [52]). Magnus integrators are also preferable when using higher order integrators (applied to non-sparse systems of any size) when high accuracies are required. This is because in this scenario, quadrature computational cost dominates integrator effort.

In the second half of this paper, sections 6–8, we focus on the quadrature cost associated with approximating multiple Stratonovich integrals to a degree of accuracy commensurate with the order of the numerical method implemented. Our conclusions apply generally to the case of nonlinear, non-commuting governing vector fields. The governing set of vector fields and driving path process (W^1, \dots, W^d) generate the unique solution process $y \in \mathbb{R}^p$ to the stochastic differential equation (1.1). For a scalar driving Wiener process W the Itô map $W \mapsto y$ is continuous in the topology of uniform convergence. For a d -dimensional driving processes with $d \geq 2$ the Universal Limit Theorem implies that the Itô map $(W^1, \dots, W^d) \mapsto y$ is continuous in the p -variation topology, in particular for $2 \leq p < 3$ (see Lyons [38], Lyons and Qian [39] and Malliavin [43]). Since Wiener paths with $d \geq 2$ have finite p -variation for $p > 2$, approximations to y constructed using successively refined approximations to the driving path will only converge to the correct solution y if we include information about the Lévy chordal areas of the driving path (the L^2 -norm of the 2-variation of

a Wiener process is finite though). Hence if we want to implement a scheme using adaptive stepsize we should consider order 1 or higher pathwise stochastic numerical methods (see Gaines and Lyons [22]).

However simulating multiple Stratonovich integrals accurately is costly! For classical accounts of this limitation on applying higher order pathwise stochastic numerical schemes see Kloeden and Platen [33, p. 367], Milstein [44, p. 92] and Schurz [51, p. 58] and for more recent results see Gaines and Lyons [21, 22], Wiktorsson [58], Cruzeiro, Malliavin and Thalmaier [17], Stump and Hill [55] and Giles [24, 25].

Taking a leaf from Gaines and Lyons [22] we consider whether it is computationally cheaper to collect a set of sample data over a given time interval and then evaluate the solution (conditioned on that sample data), than it is to evaluate the solution frequently, say at every sample time. The resounding result here is that of Clark and Cameron [16] who prove that when the multiple Stratonovich integral J_{12} is approximated by its expectation conditioned on intervening sample points, the maximal rate of L^2 -convergence is of order $h/Q^{1/2}$ where h is the integration steplength and Q is the sampling rate. We extend this result to multiple Stratonovich integrals $J_{\alpha_1, \dots, \alpha_\ell}$ of arbitrary order approximated by their expectation conditioned on intervening information sampled at the rate Q . Indeed we prove that the maximal rate of convergence is $h^{\ell/2}/Q^{1/2}$ when $\alpha_1, \dots, \alpha_\ell$ are non-zero indices (and an improved rate of convergence if some of them are zero). In practice the key information is how the accuracy achieved scales with the effort required to produce it on the global interval of integration say $[0, T]$ where $T = Nh$. We derive an explicit formula for the relation between the global error and the computational effort required to achieve it for a multiple Stratonovich integral of arbitrary order when the indices $\alpha_1, \dots, \alpha_\ell$ are distinct. This allows us to infer the effectiveness of strong methods of arbitrary order for systems with non-commuting vector fields. For a given computational effort which method delivers the best accuracy? The answer not only relies on methods that are more accurate at a given order. It also is influenced by three regimes for the stepsize that are distinguished as follows. In the first large stepsize regime the evaluation effort is greater than the quadrature effort; higher order methods produce superior performance for given effort. Quadrature effort exceeds evaluation effort in the second smaller stepsize regime. We show that in this regime when $d = 2$, or when $d \geq 3$ and the order of the method $M \leq 3/2$, then the global error scales with the computational effort with an exponent of $-1/2$. Here more accurate higher order methods still produce superior performance for given effort; but not at an increasing rate as the stepsize is decreased. However when $d \geq 3$ for strong methods with $M \geq 2$ the global error versus computational effort exponent is worse than $-1/2$ and this distinguishes the third very small stepsize regime. The greater exponent means that eventually lower order methods will deliver greater accuracy for a given effort.

We have chosen to approximate higher order integrals over a given time step by their expectations conditioned on the increments of the Wiener processes on suitable subdivisions. This is important for adaptive time-step schemes (Gaines and Lyons [22]) and filtering problems where the driving processes (say W^1 and W^2) are observed signals. However it should be noted that Wiktorsson [58] has provided a practical method for efficiently sampling the set of multiple Stratonovich multiple integrals $\{J_{ij} : i, j = 1, \dots, d\}$ across a given time-step associated with a d -dimensional driving process (see Gilling and Shardlow [23] for a practical implementation). Wiktorsson simulates the tail distribution in a truncated Karhunen–Loeve Fourier series approximation of these integrals which produces a convergence rate of order $hd^{3/2}/Q$

where Q is analogously the number of required independent normally distributed samples.

Other potential sources of computational effort might be path generation and memory access. Path generation effort depends on the application context. This cost is at worst proportional to the quadrature effort where we could subsume it. Memory access efforts depend on the processing and access memory environment. To reveal higher order methods (which typically require more path information) in the best light possible, we have ignored this effect.

Our paper is outlined as follows. We start in Section 2 by proving that the exponential of every truncation of the Magnus series converges to the solution of our linear stochastic differential equation (1.1). In Section 3 we define the strong error measures we use and how to compute them. Using these, we explicitly compare the local and then global errors for the Magnus and Neumann integrators in Section 4 and thus establish our stated results for uniformly accurate Magnus integrators. In Section 5 we show that when the linear diffusion vector fields do not commute we cannot expect the corresponding order 1 Magnus integrator to in general be globally more accurate than the order 1 Neumann integrator. We then turn our attention in Section 6 to the method of approximating multiple Stratonovich integrals by their conditional expectations. We prove the maximal rate of convergence for an arbitrary multiple Stratonovich integral in Section 6. We then use this result in Section 7 to show how the global error scales with the computational effort for numerical schemes of arbitrary order. The shuffle algebra of multiple Stratonovich integrals generated by integration by parts allows for different representations and therefore bases for the solution of a stochastic differential equation. Some choices of basis representation are more efficiently approximated than others and we investigate in Section 8 the impact of this choice. In Section 9 we present numerical experiments that reflect our theoretical results. To illustrate the superior accuracy of the uniformly accurate Magnus methods we apply them to a stochastic Riccati differential system that can be reformulated as a linear system which has commuting diffusion vector fields. Since for the linear system expensive matrix-matrix multiplications can be achieved independent of the path, the Neumann method performs better than an explicit Runge–Kutta type method applied directly to the nonlinear Riccati system. We also numerically solve an explicit linear system with governing linear vector fields that do not commute for two and also three driving Wiener processes—Magnus integrators also exhibit superior accuracy in practice in these cases also. Lastly in Section 10, we outline how to extend our results to nonlinear stochastic differential equations and propose further extensions and applications.

2. Strong convergence of truncated Magnus series. We consider here the case when the stochastic differential equation (1.1) is driven by d Wiener processes with constant coefficient matrices $a_i(t) = a_i$, $i = 0, 1, \dots, d$. The Neumann expansion has the form shown in (1.4). We construct the Magnus expansion by taking the logarithm of this Neumann series as in (1.3). In Appendix A we explicitly give the Neumann and Magnus expansions up to terms with L^2 -norm of order $3/2$. Let $\hat{\sigma}_t$ denote the truncated Magnus series

$$\hat{\sigma}_t = \sum_{\alpha \in \mathbb{Q}_m} J_\alpha(t) c_\alpha, \quad (2.1)$$

where \mathbb{Q}_m denotes the finite set of multi-indices α for which $\|J_\alpha\|_{L^2}$ is of order up to and including t^m . Note that here m is a half-integer index, $m = 1/2, 1, 3/2, \dots$. The

terms c_α are linear combinations of finitely many (more precisely exactly length α) products of the a_i , $i = 0, 1, \dots, d$. Let $|\mathbb{Q}_m|$ denote the cardinality of \mathbb{Q}_m .

THEOREM 2.1 (Convergence). *For any $t \leq 1$, the exponential of the truncated Magnus series, $\exp \hat{\sigma}_t$, is square-integrable. Further, if y_t is the solution of the stochastic differential equation (1.1), there exists a constant $C(m)$ such that*

$$\|y_t - \exp \hat{\sigma}_t \cdot y_0\|_{L^2} \leq C(m) t^{m+1/2}. \quad (2.2)$$

Proof. First we show that $\exp \hat{\sigma}_t \in L^2$. Using the expression (2.1) for $\hat{\sigma}_t$, we see that for any number k , $(\hat{\sigma}_t)^k$ is a sum of $|\mathbb{Q}_m|^k$ terms, each of which is a k -multiple product of terms $J_\alpha c_\alpha$. It follows that

$$\|(\hat{\sigma}_t)^k\|_{L^2} \leq \left(\max_{\alpha \in \mathbb{Q}_m} \|c_\alpha\|_{\text{op}} \right)^k \cdot \sum_{\substack{\alpha_i \in \mathbb{Q}_m \\ i=1, \dots, k}} \|J_{\alpha(1)} J_{\alpha(2)} \cdots J_{\alpha(k)}\|_{L^2}. \quad (2.3)$$

Note that the maximum of the operator norm $\|\cdot\|_{\text{op}}$ of the coefficient matrices is taken over a finite set. Repeated application of the product rule reveals that the product $J_{\alpha(i)} J_{\alpha(j)}$, where $\alpha(i)$ and $\alpha(j)$ are multi-indices of length $\ell(\alpha(i))$ and $\ell(\alpha(j))$, is a linear combination of $2^{\ell(\alpha(i)) + \ell(\alpha(j)) - 1}$ multiple Stratonovich integrals. Since $\ell(\alpha(i)) \leq 2m$ for $i = 1, \dots, k$, each term ' $J_{\alpha(1)} J_{\alpha(2)} \cdots J_{\alpha(k)}$ ' in (2.3) is thus the sum of at most 2^{2mk-1} Stratonovich integrals J_β . We also note that $k \leq \ell(\beta) \leq 2mk$.

From equation (5.2.34) in Kloeden and Platen [33], every multiple Stratonovich integral J_β can be expressed as a finite sum of at most $2^{\ell(\beta)-1}$ multiple Itô integrals I_γ with $\ell(\gamma) \leq \ell(\beta)$. Further, from Remark 5.2.8 in Kloeden and Platen [33], $\ell(\gamma) + n(\gamma) \geq \ell(\beta) + n(\beta)$, where $n(\beta)$ and $n(\gamma)$ denote the number of zeros in β and γ , respectively. From Lemma 5.7.3 in Kloeden and Platen [33],

$$\|I_\gamma\|_{L^2} \leq 2^{\ell(\gamma) - n(\gamma)} t^{(\ell(\gamma) + n(\gamma))/2}.$$

Noting that $\ell(\gamma) \leq \ell(\beta) \leq 2mk$ and $\ell(\gamma) + n(\gamma) \geq k$, it follows that for $t \leq 1$, we have $\|J_\beta\|_{L^2} \leq 2^{4mk-1} t^{k/2}$. Since the right hand side of equation (2.3) consists of $|\mathbb{Q}_m|^k 2^{2mk-1}$ Stratonovich integrals J_β , we conclude that,

$$\|(\hat{\sigma}_t)^k\|_{L^2} \leq \left(\max_{\alpha \in \mathbb{Q}_m} \|c_\alpha\|_{\text{op}} \cdot |\mathbb{Q}_m| \cdot 2^{6m} \cdot t^{1/2} \right)^k.$$

Hence $\exp \hat{\sigma}_t$ is square-integrable.

Second we prove (2.2). Let \hat{y}_t denote Neumann series solution (1.4) truncated to included terms of order up to and including t^m . We have

$$\|y_t - \exp \hat{\sigma}_t \cdot y_0\|_{L^2} \leq \|y_t - \hat{y}_t\|_{L^2} + \|\hat{y}_t - \exp \hat{\sigma}_t \cdot y_0\|_{L^2}. \quad (2.4)$$

We know $y_t \in L^2$ (see Gihman and Skorohod [23] or Arnold [2]). Furthermore, for any order m , \hat{y}_t corresponds to the truncated Taylor expansion involving terms of order up to and including t^m . Hence \hat{y}_t is a strong approximation to y_t to that order with the remainder consisting of $\mathcal{O}(t^{m+1/2})$ terms (see Proposition 5.9.1 in Kloeden and Platen [33]). It follows from the definition of the Magnus series as the logarithm of the flow-map Neumann series, that the terms of order up to and including t^m in $\exp \hat{\sigma}_t \cdot y_0$ correspond with \hat{y}_t ; the error consists of $\mathcal{O}(t^{m+1/2})$ terms. \square

Convergence of approximations based on truncations of the stochastic Taylor expansion has been studied in Kloeden and Platen [33], see Propositions 5.10.1, 5.10.2, and 10.6.3. Ben Arous [4] and Castell [12] prove the remainder of the exponential of any truncation of the Magnus series is bounded in probability as $t \rightarrow 0$ (in the full nonlinear case). Burrage [7] shows that the first terms up to and including order 3/2 Magnus expansion coincide with the terms in the Taylor expansion of the same order. Our result holds for any order in L^2 for sufficiently small t . A more detailed analysis is needed to establish results concerning the convergence radius. Similar arguments can be used to study the non-constant coefficient case with suitable conditions on the coefficient matrices (see Proposition 5.10.1 in Kloeden and Platen [33] for the corresponding result for the Taylor expansion).

Note that above and in subsequent sections, one may equally consider a stochastic differential equation starting at time $t_0 > 0$ with square-integrable \mathcal{F}_{t_0} -measurable initial data y_0 . Here $(\mathcal{F}_t)_{t \geq 0}$ denotes the underlying filtration.

3. Global and local error. Suppose $S_{t_n, t_{n+1}}$ and $\hat{S}_{t_n, t_{n+1}}$ are the exact and approximate flow-maps across the interval $[t_n, t_{n+1}]$, respectively; both satisfying the usual flow-map semi-group property: composition of flow-maps across successive intervals generates the flow-map across the union of those intervals. We call the difference between the exact and approximate flow-maps

$$R_{t_n, t_{n+1}} \equiv S_{t_n, t_{n+1}} - \hat{S}_{t_n, t_{n+1}} \quad (3.1)$$

the *local flow remainder*. For an approximation $\hat{y}_{t_{n+1}}$ across the interval $[t_n, t_{n+1}]$ the local remainder is thus $R_{t_n, t_{n+1}} y_{t_n}$. Our goal here is to see how the leading order terms in the local remainders accumulate, contributing to the global error.

DEFINITION 3.1 (Strong global error). *We define the strong global error associated with an approximate solution \hat{y}_T to the stochastic differential equation (1.1) over the global interval of integration $[0, T]$ as $\mathcal{E} \equiv \|y_T - \hat{y}_T\|_{L^2}$.*

The global error can be decomposed additively into two components, the global truncation error due to truncation of higher order terms, and the global quadrature error due to the approximation of multiple Stratonovich integrals retained in the approximation. If $[0, T] = \cup_{n=0}^{N-1} [t_n, t_{n+1}]$ where $t_n = nh$ then for small stepsize $h = T/N$ we have

$$\begin{aligned} \mathcal{E} &= \left\| \left(\prod_{n=N-1}^0 S_{t_n, t_{n+1}} - \prod_{n=N-1}^0 \hat{S}_{t_n, t_{n+1}} \right) y_0 \right\|_{L^2} \\ &= \left\| \left(\sum_{n=0}^{N-1} \hat{S}_{t_{n+1}, t_N} R_{t_n, t_{n+1}} \hat{S}_{t_0, t_n} \right) y_0 \right\|_{L^2} + \mathcal{O}(\max_n \|R_{t_n, t_{n+1}}\|_{L^2}^{3/2} h^{-3/2}). \end{aligned} \quad (3.2)$$

The local flow remainder has the following form in the case of constant coefficients a_i , $i = 1, \dots, d$, (see for example the integrators in Appendix A):

$$R_{t_n, t_{n+1}} = \sum_{\alpha} J_{\alpha}(t_n, t_{n+1}) c_{\alpha}.$$

Here α is a multi-index and the terms c_{α} represent products or commutations of the constant matrices a_i . The J_{α} represent Stratonovich integrals (or linear combinations, of the same order, of products of integrals including permutations of α). The global

error \mathcal{E}^2 at leading order in the stepsize is thus

$$y_0^T \sum_{\alpha, \beta} \left(\sum_n \mathbb{E}(J_\alpha(t_n, t_{n+1}) J_\beta(t_n, t_{n+1})) \mathbb{E} \left((\hat{S}_{t_{n+1}, t_N} c_\alpha \hat{S}_{t_0, t_n})^T (\hat{S}_{t_{n+1}, t_N} c_\beta \hat{S}_{t_0, t_n}) \right) \right. \\ \left. + \sum_{n \neq m} \mathbb{E}(J_\alpha(t_n, t_{n+1})) \mathbb{E}(J_\beta(t_m, t_{m+1})) \mathbb{E} \left((\hat{S}_{t_{n+1}, t_N} c_\alpha \hat{S}_{t_0, t_n})^T (\hat{S}_{t_{m+1}, t_N} c_\beta \hat{S}_{t_0, t_m}) \right) \right) y_0.$$

Hence in the global truncation error we distinguish between the *diagonal sum* consisting of the the first sum on the right-hand side above, and the *off-diagonal sum* consisting of the second sum above with $n \neq m$.

Suppose we include in our integrator all terms with local L^2 -norm up to and including $\mathcal{O}(h^M)$. The leading terms $J_\alpha c_\alpha$ in $R_{t_n, t_{n+1}}$ thus have L^2 -norm $\mathcal{O}(h^{M+1/2})$. Those with zero expectation will contribute to the diagonal sum, generating $\mathcal{O}(h^M)$ terms in the global error, consistent with a global order M integrator. However those with non-zero expectation contribute to the off-diagonal double sum. They will generate $\mathcal{O}(h^{M-1/2})$ terms in the global error. We must thus either include them in the integrator, or more cheaply, only include their expectations—the corresponding terms $(J_\alpha - \mathbb{E}(J_\alpha))c_\alpha$ of order $h^{M+1/2}$ in $R_{t_n, t_{n+1}}$ will then have zero expectation and only contribute through the diagonal sum—see for example Milstein [44, p. 12]. This also guarantees the next order term in the global error estimate (3.2), whose largest term has the upper bound $\max_n \|R_{t_n, t_{n+1}}\|_{L^2}^{3/2} h^{-3/2}$, only involves higher order contributions to the leading $\mathcal{O}(h^M)$ error.

Note that high order integrators may include multiple Stratonovich integral terms. We approximate these multiple integrals by their conditional expectations to the local order of approximation $h^{M+1/2}$ of the numerical method. Hence terms in the *integrator* of the form $J_\alpha c_\alpha$ are in fact approximated by $\mathbb{E}(J_\alpha | \mathcal{F}_Q) c_\alpha$, their expectation conditioned on intervening path information \mathcal{F}_Q (see Section 6 for more details). This generates terms of the form $(J_\alpha - \mathbb{E}(J_\alpha | \mathcal{F}_Q))c_\alpha$ in the local flow remainder, which have zero expectation and hence contribute to the global error through the diagonal sum generating $\mathcal{O}(h^M)$ terms.

4. Uniformly accurate Magnus integrators. Our goal is to identify a class Magnus integrators that are more accurate than Neumann (stochastic Taylor) integrators of the same order for any governing set of linear vector fields (for the integrators of order 1 and 3/2 we assume the diffusion vector fields commute). We thus compare the local accuracy of the Neumann and Magnus integrators through the leading terms of their remainders. We consider the case of constant coefficient matrices a_i , $i = 0, 1, \dots, d$.

The local flow remainder of a Neumann integrator R^{neu} is simply given by the terms not included in the flow-map Neumann approximation. Suppose $\hat{\sigma}$ is the truncated Magnus expansion and that ρ is the corresponding remainder, i.e. $\sigma = \hat{\sigma} + \rho$. Then the local flow remainder R^{mag} associated with the Magnus approximation is

$$\begin{aligned} R^{\text{mag}} &= \exp \sigma - \exp \hat{\sigma} \\ &= \exp(\hat{\sigma} + \rho) - \exp \hat{\sigma} \\ &= \rho + \frac{1}{2}(\hat{\sigma}\rho + \rho\hat{\sigma}) + \mathcal{O}(\hat{\sigma}^2\rho). \end{aligned} \tag{4.1}$$

Hence the local flow remainder of a Magnus integrator R^{mag} is the truncated Magnus expansion remainder ρ , and higher order terms $\frac{1}{2}(\hat{\sigma}\rho + \rho\hat{\sigma})$ that can contribute to the

global error at leading order through their expectations. For the integrators considered in this section these higher order terms do not contribute in this way, however for the order 1 integrator we consider in the next section they do.

DEFINITION 4.1 (Uniformly accurate Magnus integrators). *When the linear diffusion vector fields commute so that $[a_i, a_j] = 0$ for all $i, j \neq 0$, we define the order 1 and order 3/2 uniformly accurate Magnus integrators by*

$$\hat{\sigma}_{t_n, t_{n+1}}^{(1)} = J_0 a_0 + \sum_{i=1}^d (J_i a_i + \frac{h^2}{12} [a_i, [a_i, a_0]]),$$

and

$$\hat{\sigma}_{t_n, t_{n+1}}^{(3/2)} = J_0 a_0 + \sum_{i=1}^d (J_i a_i + \frac{1}{2} (J_{i0} - J_{0i}) [a_0, a_i] + \frac{h^2}{12} [a_i, [a_i, a_0]]).$$

By *uniformly* we mean for any given set of governing linear vector fields (or equivalently coefficient matrices a_i , $i = 0, 1, \dots, d$) for which the diffusion vector fields commute, and for any initial data y_0 .

THEOREM 4.2 (Global error comparison). *For any initial condition y_0 and sufficiently small fixed stepsize $h = t_{n+1} - t_n$, the order 1/2 Magnus integrator is globally more accurate in L^2 than the order 1/2 Neumann integrator. If in addition we assume the linear diffusion vector fields commute so that $[a_i, a_j] = 0$ for all $i, j \neq 0$, then the order 1 and 3/2 uniformly accurate Magnus integrators are globally more accurate in L^2 than the corresponding Neumann integrators. In other words, if \mathcal{E}^{mag} denotes the global error of the order 1/2 Magnus integrator or the uniformly accurate Magnus integrators of order 1 or order 3/2, respectively, and \mathcal{E}^{neu} is the global error of the Neumann integrators of the corresponding order, then at each of those orders,*

$$\mathcal{E}^{\text{mag}} \leq \mathcal{E}^{\text{neu}}. \quad (4.2)$$

Proof. Let R^{mag} and R^{neu} denote the local flow remainders corresponding to the Magnus and Neumann approximations across the interval $[t_n, t_{n+1}]$ with $t_n = nh$. A direct calculation reveals that

$$\mathbb{E}((R^{\text{neu}})^T R^{\text{neu}}) = \mathbb{E}((R^{\text{mag}})^T R^{\text{mag}}) + D_M + \mathcal{O}(h^{2M+1/2}),$$

where if we set $\hat{R} \equiv R^{\text{neu}} - R^{\text{mag}}$ then

$$D_M \equiv \mathbb{E}(\hat{R}^T R^{\text{mag}}) + \mathbb{E}((R^{\text{mag}})^T \hat{R}) + \mathbb{E}(\hat{R}^T \hat{R}). \quad (4.3)$$

We now show explicitly in each of the three cases of the theorem that at leading order: R^{mag} and \hat{R} are uncorrelated and hence D_M is positive semi-definite. This implies that the local remainder for the Neumann expansion is larger than that of the Magnus expansion. Hereafter assume the indices $i, j, k, l \in \{1, \dots, d\}$.

For the *order 1/2 integrators* we have to leading order (see the form of the expansions in Appendix A)

$$R^{\text{mag}} = \sum_{i < j} \frac{1}{2} (J_{ij} - J_{ji}) [a_j, a_i],$$

and

$$\hat{R} = \sum_i (J_{ii} - \frac{1}{2}h) a_i^2 + \sum_{i < j} \frac{1}{2} (J_{ij} + J_{ji}) (a_j a_i + a_i a_j),$$

which are uncorrelated by direct inspection.

We henceforth assume $[a_i, a_j] = 0$ for all i, j . For the *uniformly accurate order 1 integrator* we have to leading order (again see Appendix A)

$$R^{\text{mag}} = \sum_i \frac{1}{2} (J_{i0} - J_{0i}) [a_0, a_i],$$

and

$$\begin{aligned} \hat{R} &= \frac{1}{2} h^2 a_0^2 + \sum_i \frac{1}{2} (J_{i0} + J_{0i}) (a_0 a_i + a_i a_0) + \frac{1}{4} h^2 (a_i^2 a_0 + a_0 a_i^2) \\ &\quad + \sum_{i,j,k} (J_{ijk} - \mathbb{E}(J_{ijk})) a_k a_j a_i, \end{aligned}$$

which again, by direct inspection are uncorrelated.

For the *uniformly accurate order 3/2 integrator* the local flow remainders are

$$\begin{aligned} R^{\text{neu}} &= \sum_i (J_{i0} - \frac{1}{4} h^2) a_0 a_i^2 + J_{i0i} a_i a_0 a_i + (J_{0ii} - \frac{1}{4} h^2) a_i^2 a_0 \\ &\quad + \sum_{i < j} \left((J_{0ji} + J_{0ij}) a_i a_j a_0 + J_{j0i} a_i a_0 a_j \right. \\ &\quad \quad \left. + J_{i0j} a_j a_0 a_i + (J_{j0i} + J_{i0j}) a_0 a_i a_j \right) \\ &\quad + \sum_{i,j,k,l} (J_{ijkl} - \mathbb{E}(J_{ijkl})) a_l a_k a_j a_i, \end{aligned}$$

and

$$\begin{aligned} R^{\text{mag}} &= \sum_i \frac{1}{12} (J_i^2 J_0 - h^2 - 6J_{i0i}) [a_i, [a_i, a_0]] \\ &\quad + \sum_{i < j} \frac{1}{6} (J_0 J_i J_j - 3(J_{i0j} + J_{j0i})) [a_j, [a_i, a_0]]. \end{aligned}$$

Consequently we have

$$\begin{aligned} \hat{R} &= \frac{1}{12} \sum_i \left((6J_{i0} J_i - J_i^2 J_0 - 2h^2) a_0 a_i^2 + 2(J_i^2 J_0 - h^2) a_i a_0 a_i \right. \\ &\quad \left. + (6J_{0i} J_i - J_i^2 J_0 - 2h^2) a_i^2 a_0 \right) \\ &\quad + \frac{1}{6} \sum_{i < j} \left((3(J_i J_{0j} + J_j J_{0i}) - J_0 J_i J_j) a_i a_j a_0 \right. \\ &\quad \quad + (J_0 J_i J_j + 3(J_{i0j} - J_{j0i})) a_j a_0 a_i \\ &\quad \quad + (J_0 J_i J_j + 3(J_{j0i} - J_{i0j})) a_i a_0 a_j \\ &\quad \quad \left. + (3(J_i J_{j0} + J_j J_{i0}) - J_0 J_i J_j) a_0 a_i a_j \right) \\ &\quad + \sum_{i,j,k,l} (J_{ijkl} - \mathbb{E}(J_{ijkl})) a_l a_k a_j a_i. \end{aligned}$$

First we note that the terms in \hat{R} of the form $J_{ijkl}a_l a_k a_j a_i$, for which at least three of the indices distinct, are uncorrelated with any terms in R^{mag} . We thus focus on the terms of this form with at most two distinct indices, namely

$$\sum_{i < j} (J_{ii}J_{jj} - \frac{1}{4}h^2)a_j^2 a_i^2 + \sum_{i \neq j} J_{iii}J_j a_j a_i^3 + \sum_i (J_{iiii} - \frac{1}{8}h^2)a_i^4$$

and other remaining terms in \hat{R} . Since $\mathbb{E}[J_{i0i}|J_i] \equiv h(J_i^2 - h)/6$ and $\mathbb{E}[J_{i0j}|J_i, J_j] \equiv hJ_i J_j/6$ for $i \neq j$ the following conditional expectations are immediate

$$\begin{aligned} \mathbb{E}[(J_i^2 J_0 - h^2)(J_i^2 J_0 - h^2 - 6J_{i0i})|J_i] &= 0, \\ \mathbb{E}[(J_{iiii} - \frac{1}{8}h^2)(J_i^2 J_0 - h^2 - 6J_{i0i})|J_i] &= 0, \\ \mathbb{E}[(J_{ii}J_{jj} - \frac{1}{4}h^2)(J_i^2 J_0 - h^2 - 6J_{i0i})|J_i, J_j] &= 0, \\ \mathbb{E}[(J_0 J_i J_j + 3(J_{i0j} - J_{j0i}))(J_0 J_i J_j - 3(J_{i0j} + J_{j0i}))|J_i, J_j] &= 0, \\ \mathbb{E}[(J_{ii}J_{jj} - \frac{1}{4}h^2)(J_0 J_i J_j - 3(J_{i0j} + J_{j0i}))|J_i, J_j] &= 0, \\ \mathbb{E}[J_{iii}J_j(J_0 J_i J_j - 3(J_{i0j} + J_{j0i}))|J_i, J_j] &= 0. \end{aligned}$$

Hence the expectations of the terms shown are also zero. Secondly, direct computation of the following expectations reveals

$$\begin{aligned} \mathbb{E}((6J_{0i}J_i - J_i^2 J_0 - 2h^2)(J_i^2 J_0 - h^2 - 6J_{i0i})) &= 0, \\ \mathbb{E}\left(\left(3(J_i J_{0j} + J_j J_{0i}) - J_0 J_i J_j\right)(J_0 J_i J_j - 3(J_{i0j} + J_{j0i}))\right) &= 0. \end{aligned}$$

Hence R^{mag} and \hat{R} are uncorrelated.

The corresponding global comparison results (4.2) now follow directly in each of the three cases above using that the local errors accumulate and contribute to the global error as diagonal terms in the standard manner described in detail at the end of Section 3. Note that we include the terms $\frac{1}{12}h^2[a_i, [a_i, a_0]]$ in the order 1 uniformly accurate Magnus integrator. These terms appear at leading order in the global remainder where they would otherwise generate a non-positive definite contribution to D_M in (4.3). \square

Note that the Magnus integrator $\hat{\sigma}^{(1)}$ is an order 1 integrator without the terms $\frac{1}{12}h^2[a_i, [a_i, a_0]]$. However they are cheap to compute and including them in the integrator guarantees global superior accuracy independent of the set of governing coefficient matrices.

5. Non-commuting linear diffusion vector fields. What happens when the linear diffusion vector fields do not commute, i.e. we have $[a_i, a_j] \neq 0$ for non-zero indices $i \neq j$? Hereafter assume $i, j, k \in \{1, \dots, d\}$. Consider the case of order 1 integrators. The local flow remainders are

$$R^{\text{neu}} = \sum_i (J_{i0}a_0 a_i + J_{0i}a_i a_0) + \sum_{i,j,k} J_{kji}a_i a_j a_k$$

and

$$\begin{aligned} R^{\text{mag}} &= \sum_i \frac{1}{2}(J_{i0} - J_{0i})[a_0, a_i] + \sum_{i \neq j} (J_{iij} - \frac{1}{2}J_i J_{ij} + \frac{1}{12}J_i^2 J_j)[a_i, [a_i, a_j]] \\ &+ \sum_{i < j < k} ((J_{ijk} + \frac{1}{2}J_j J_{ki} + \frac{1}{2}J_k J_{ij} - \frac{2}{3}J_i J_j J_k)[a_i, [a_j, a_k]] \\ &+ (J_{jik} + \frac{1}{2}J_i J_{kj} + \frac{1}{2}J_k J_{ji} - \frac{2}{3}J_i J_j J_k)[a_j, [a_i, a_k]]). \end{aligned}$$

Computing D_M in (4.3) gives

$$D_M = h^2 \sum_{i \neq j} U_{ij}^T B U_{ij} + h^2 \sum_{i \neq j \neq k} U_{ijk}^T C U_{ijk}$$

where $U_{ij} = (a_j a_i^2, a_i a_j a_i, a_i^2 a_j, a_j^3, a_0 a_j, a_j a_0)^T \in \mathbb{R}^{6p \times p}$ and in addition we have that $U_{ijk} = (a_k a_j a_i, a_j a_k a_i, a_k a_i a_j, a_i a_k a_j, a_j a_i a_k, a_i a_j a_k)^T \in \mathbb{R}^{6p \times p}$. Here $B, C \in \mathbb{R}^{6p \times 6p}$ consist of $p \times p$ diagonal blocks of the form $b_{k\ell} I_p$ and $c_{k\ell} I_p$ where

$$b = \frac{1}{144} \begin{pmatrix} 31 & 10 & 1 & 18 & 12 & 24 \\ 10 & 4 & 10 & 0 & 0 & 0 \\ 1 & 10 & 31 & 18 & 24 & 12 \\ 18 & 0 & 18 & 60 & 36 & 36 \\ 12 & 0 & 24 & 36 & 36 & 36 \\ 24 & 0 & 12 & 36 & 36 & 36 \end{pmatrix},$$

and

$$c = \frac{1}{36} \begin{pmatrix} 4 & 1 & 1 & 1 & 1 & -2 \\ 1 & 4 & 1 & -2 & 1 & 1 \\ 1 & 1 & 4 & 1 & -2 & 1 \\ 1 & -2 & 1 & 4 & 1 & 1 \\ 1 & 1 & -2 & 1 & 4 & 1 \\ -2 & 1 & 1 & 1 & 1 & 4 \end{pmatrix}.$$

Again, c is positive semi-definite with eigenvalues $\{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, 0, 0\}$. However b has eigenvalues $\{\frac{1}{6}, \frac{1}{48}(5 + \sqrt{41}), \frac{1}{48}(5 - \sqrt{41}), 0.94465, 0.0943205, -0.03897\}$, where the final three values are approximations to the roots of $288x^3 - 2888x^2 + 14x + 1$.

The eigenvalues of b and c , respectively, are multiple eigenvalues for the matrices B and C , respectively. This implies that there are certain matrix combinations and initial conditions, for which the order 1 Taylor approximation is more accurate in the mean-square sense than the Magnus approximation. However, the two negative values are small in absolute value compared to the positive eigenvalues. For the majority of systems, one can thus expect the Magnus approximation to be more accurate (as already observed by Sipiläinen [53], Burrage [7] and Burrage and Burrage [8]). For any given linear system of stochastic differential equations, the scheme that is more accurate can be identified using the results above.

In this case there are terms from $\frac{1}{2}(\hat{\sigma}\rho + \rho\hat{\sigma})$ in Magnus remainder expression (4.1) that appear at leading order in the local flow remainder. These terms are of the form $\frac{1}{12}(a_i[a_j, [a_j, a_i]] + a_j[a_i, [a_i, a_j]])h^2$. They make a negative definite contribution to global error, though this can be negated by including them as cheaply computable terms in the Magnus integrator (indeed we recommend doing so).

6. Quadrature. We start by emphasizing that there are two inherent scales:

- *Quadrature scale* Δt on which the discrete Wiener paths are generated;
- *Evaluation scale* h on which the stochastic differential equation is advanced.

The idea is to approximate multiple Stratonovich integrals by their corresponding expectations conditioned on the σ -algebra representing intervening knowledge of the Wiener paths (Clark and Cameron [16]; Newton [49]; Gaines and Lyons [22]). Hence we approximate $J_\alpha(t_n, t_{n+1})$ by $\mathbb{E}(J_\alpha(t_n, t_{n+1}) | \mathcal{F}_Q)$ where

$$\mathcal{F}_Q = \{\Delta W_{t_n+q\Delta t}^i : i = 1, \dots, d; q = 0, \dots, Q-1; n = 0, \dots, N-1\},$$

with $\Delta W_{t_n+q\Delta t}^i \equiv W_{t_n+(q+1)\Delta t}^i - W_{t_n+q\Delta t}^i$, and $Q\Delta t \equiv h$, i.e. Q is the number of Wiener increments. We extend the result of Clark and Cameron [16] on the maximum rate of convergence to arbitrary order multiple Stratonovich integrals.

LEMMA 6.1 (Quadrature error). *Suppose at least two of the indices in the multi-index $\alpha = \{\alpha_1, \dots, \alpha_\ell\}$ are distinct. Define*

$$j^* \equiv \min\{j \leq \ell - 1 : \alpha_i = \alpha_l, \forall i, l > j\}.$$

Let $n(\alpha)$ denote the number of zeros in α , and $n^*(\alpha) = n(\{\alpha_{j^*}, \dots, \alpha_\ell\})$. Then the L^2 error in the multiple Stratonovich integral approximation $\mathbb{E}(J_\alpha(t_n, t_{n+1}) | \mathcal{F}_Q)$ is

$$\|J_\alpha(t_n, t_{n+1}) - \mathbb{E}(J_\alpha(t_n, t_{n+1}) | \mathcal{F}_Q)\|_{L^2} = \mathcal{O}\left(\frac{h^{(\ell+n(\alpha))/2}}{Q^{(n^*(\alpha)+1)/2}}\right). \quad (6.1)$$

Proof. For any s, t , we write for brevity

$$\widehat{J}_\alpha(s, t) = \mathbb{E}(J_\alpha(s, t) | \mathcal{F}_Q).$$

We define α^{-k} as the multi-index obtained by deleting the last k indices, that is $\alpha^{-k} = \{\alpha_1, \dots, \alpha_{\ell-k}\}$. We set $\tau_q \equiv t_n + q\Delta t$, where $q = 0, \dots, Q-1$. Then

$$\begin{aligned} J_\alpha(t_n, t_{n+1}) &= \sum_{q=0}^{Q-1} \int_{\tau_q}^{\tau_{q+1}} J_{\alpha^{-1}}(t_n, \tau) dW_\tau^{\alpha_\ell} \\ &= \sum_{q=0}^{Q-1} \left(\sum_{k=1}^{\ell-1} J_{\alpha^{-k}}(t_n, \tau_q) J_{\alpha_{\ell-k+1}, \dots, \alpha_\ell}(\tau_q, \tau_{q+1}) + J_\alpha(\tau_q, \tau_{q+1}) \right). \end{aligned}$$

Thus we have

$$\begin{aligned} J_\alpha(t_n, t_{n+1}) - \widehat{J}_\alpha(t_n, t_{n+1}) &= \sum_{q=0}^{Q-1} \left(\sum_{k=1}^{\ell-1} (J_{\alpha^{-k}}(t_n, \tau_q) - \widehat{J}_{\alpha^{-k}}(t_n, \tau_q)) J_{\alpha_{\ell-k+1}, \dots, \alpha_\ell}(\tau_q, \tau_{q+1}) \right. \\ &\quad + \widehat{J}_{\alpha^{-k}}(t_n, \tau_q) (J_{\alpha_{\ell-k+1}, \dots, \alpha_\ell}(\tau_q, \tau_{q+1}) - \widehat{J}_{\alpha_{\ell-k+1}, \dots, \alpha_\ell}(\tau_q, \tau_{q+1})) \\ &\quad \left. + J_\alpha(\tau_q, \tau_{q+1}) - \widehat{J}_\alpha(\tau_q, \tau_{q+1}) \right). \quad (6.2) \end{aligned}$$

We prove the assertion by induction over ℓ . For $\ell = 2$, the first two terms in the sum (6.2) are zero and

$$\begin{aligned} \left\| \sum_{q=0}^{Q-1} (J_\alpha(\tau_q, \tau_{q+1}) - \widehat{J}_\alpha(\tau_q, \tau_{q+1})) \right\|_2^2 &= \mathcal{O}(Q(\Delta t)^{\ell+n(\alpha)}) \\ &= \mathcal{O}\left(\frac{h^{\ell+n(\alpha)}}{Q^{\ell+n(\alpha)-1}}\right) \\ &= \mathcal{O}\left(\frac{h^{\ell+n(\alpha)}}{Q^{n^*(\alpha)+1}}\right). \end{aligned}$$

Here we have used that fact that $\ell = 2$, $j^* = 1$ and thus $n(\alpha) = n^*(\alpha)$.

Assume now that $\ell > 2$. We will investigate the order of each of the three types of terms in (6.2) separately. If at least two indices in α^{-k} are distinct, then by induction hypothesis

$$\|J_{\alpha^{-k}}(t_n, \tau_q) - \widehat{J}_{\alpha^{-k}}(t_n, \tau_q)\|_2^2 = \mathcal{O}\left(\frac{(\tau_q - t_n)^{\ell-k+n(\alpha^{-k})}}{q^{n^*(\alpha^{-k})+1}}\right),$$

and since $\tau_q - t_n = q\Delta t$ and $Q\Delta t = h$, we have for each $k = 1, \dots, \ell - 1$,

$$\begin{aligned} \sum_{q=0}^{Q-1} \|(J_{\alpha^{-k}}(t_n, \tau_q) - \widehat{J}_{\alpha^{-k}}(t_n, \tau_q)) J_{\alpha_{\ell-k+1}, \dots, \alpha_\ell}(\tau_q, \tau_{q+1})\|_2^2 \\ = \mathcal{O}\left(\frac{h^{\ell+n(\alpha)}}{Q^{n^*(\alpha^{-k})+n(\alpha)-n(\alpha^{-k})+k}}\right). \end{aligned}$$

Note that we have

$$n^*(\alpha^{-k}) + n(\alpha) - n(\alpha^{-k}) + k \geq n^*(\alpha) + k \geq n^*(\alpha) + 1. \quad (6.3)$$

If all indices in α^{-k} are equal, then $\|J_{\alpha^{-k}}(t_n, \tau_q) - \widehat{J}_{\alpha^{-k}}(t_n, \tau_q)\|_2^2 = 0$.

For the second term in (6.2) we have for each $k = 1, \dots, \ell - 1$,

$$\begin{aligned} \sum_{q=0}^{Q-1} \|\widehat{J}_{\alpha^{-k}}(t_n, \tau_q) (J_{\alpha_{\ell-k+1}, \dots, \alpha_\ell}(\tau_q, \tau_{q+1}) - \widehat{J}_{\alpha_{\ell-k+1}, \dots, \alpha_\ell}(\tau_q, \tau_{q+1}))\|_2^2 \\ = \begin{cases} 0, & \text{if } \alpha_\ell = \dots = \alpha_{\ell-k+1}, \\ \mathcal{O}\left(\frac{h^{\ell+n(\alpha)}}{Q^{n(\alpha)-n(\alpha^{-k})+k-1}}\right), & \text{otherwise.} \end{cases} \end{aligned}$$

Note that in the second of these cases

$$n(\alpha) - n(\alpha^{-k}) + k - 1 \geq n^*(\alpha) + 1. \quad (6.4)$$

For the third term we have

$$\sum_{q=0}^{Q-1} \|J_\alpha(\tau_q, \tau_{q+1}) - \widehat{J}_\alpha(\tau_q, \tau_{q+1})\|_2^2 = \mathcal{O}\left(\frac{h^{\ell+n(\alpha)}}{Q^{\ell+n(\alpha)-1}}\right).$$

Again, note that we have

$$\ell + n(\alpha) - 1 \geq n^*(\alpha) + 1. \quad (6.5)$$

Equality holds in at least one of (6.3) to (6.5). To see this we distinguish the case when the last two indices are equal, $\alpha_{\ell-1} = \alpha_\ell$, and the case when the last two indices are distinct, $\alpha_{\ell-1} \neq \alpha_\ell$. If $\alpha_{\ell-1} = \alpha_\ell$, then

$$n^*(\alpha^{-1}) + n(\alpha) - n(\alpha^{-1}) = n^*(\alpha).$$

Since in this case at least two indices in α^{-1} are distinct, equality holds for $k = 1$ in (6.3). If $\alpha_{\ell-1} \neq \alpha_\ell$, then

$$n^*(\alpha) = n(\alpha) - n(\alpha^{-2}),$$

and thus equality holds for $k = 2$ in (6.4). Hence the lemma follows. \square

Note that each multiple Stratonovich integral $J_\alpha(t_n, t_{n+1})$ can be thought of as an ℓ -dimensional topologically conical volume in (W^1, \dots, W^ℓ) -space. The surface of the conical volume is *panelled* with each panel distinguished by a double Stratonovich integral term involving two consecutive indices from α . The edges between the panels are distinguished by a triple integral and so forth. The conditional expectation approximation $\mathbb{E}(J_\alpha(t_n, t_{n+1}) | \mathcal{F}_Q)$ can also be decomposed in this way. In the L^2 -error estimate for this approximation, the leading terms are given by sums over the panels which also confirm the estimate (6.1).

Approximations of multiple Stratonovich integrals constructed using their conditional expectations are intimately linked to those constructed using paths W_t^i that are approximated by piecewise linear interpolations of the intervening sample points. The difference between the two approaches are asymptotically smaller terms. For more details see Wong and Zakai [59], Kloeden and Platen [33], Hofmann and Müller-Gronbach [29] and Gyöngy and Michaletzky [27].

7. Global error vs computational effort. We examine in detail the step-size/accuracy regimes for which higher order stochastic integrators are feasible and also when they become less efficient than lower order schemes. In any strong simulation there are two principle sources of computational effort. Firstly there is evaluation effort $\mathcal{U}^{\text{eval}}$ associated with evaluating the vector fields, their compositions and any functions such as the matrix exponential. Secondly there is the quadrature effort $\mathcal{U}^{\text{quad}}$ associated with approximating multiple stochastic integrals to an accuracy commensurate with the order of the method. For a numerical approximation of order M the computational evaluation effort measured in flops over $N = Th^{-1}$ evaluation steps is

$$\mathcal{U}^{\text{eval}} = (c_M p^2 + c_E) Th^{-1}.$$

Here p is the size of the system, c_M represents the number of scalar-matrix multiplications and matrix-matrix additions for the order M truncated Magnus expansion, and c_E is the effort required to compute the matrix exponential. Note that if we implement an order 1/2 method there is no quadrature effort. Hence since $\mathcal{E} = \mathcal{O}(h^{1/2})$ we have $\mathcal{E} = \mathcal{O}((\mathcal{U}^{\text{eval}})^{-1/2})$.

Suppose we are required to simulate $J_{\alpha_1 \dots \alpha_\ell}(t_n, t_{n+1})$ with all the indices distinct with a global error of order h^M ; naturally $\ell \geq 2$ and $M \geq \ell/2$.

LEMMA 7.1. *The quadrature effort \mathcal{U} measured in flops required to approximate $J_{\alpha_1 \dots \alpha_\ell}(t_n, t_{n+1})$ with a global error of order h^M when all the indices are distinct and non-zero, is to leading order in h :*

$$\mathcal{U} = \mathcal{O}(h^{-\beta(M, \ell)}),$$

where

$$\beta(M, \ell) = (\ell - 1)(2M + 1 - \ell) + 1.$$

Since we stipulate the global error associated with the multiple integral approximation to be $\mathcal{E} = \mathcal{O}(h^M)$, we have

$$\mathcal{E} = \mathcal{O}(\mathcal{U}^{-M/\beta(M, \ell)}).$$

Proof. The quadrature effort required to construct $\mathbb{E}(J_{\alpha_1 \dots \alpha_\ell}(t_n, t_{n+1}) | \mathcal{F}_Q)$, which is a $(\ell - 1)$ -multiple sum, over $[0, T]$ is $\mathcal{U} = \mathcal{O}(Q^{\ell-1} N)$ with $N = Th^{-1}$. Using

TABLE 7.1

Slopes of the logarithm of the global error \mathcal{E} versus the logarithm of the quadrature effort \mathcal{U} , i.e. the exponent $-M/\beta(M, \ell)$, for different values of ℓ and M when $J_{\alpha_1 \dots \alpha_\ell}$ has distinct non-zero indices.

	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$	$\ell = 6$
no zero index					
$M = 1$	$-1/2$	\dots	\dots	\dots	\dots
$M = 3/2$	$-1/2$	$-1/2$	\dots	\dots	\dots
$M = 2$	$-1/2$	$-2/5$	$-1/2$	\dots	\dots
$M = 5/2$	$-1/2$	$-5/14$	$-5/14$	$-1/2$	\dots
$M = 3$	$-1/2$	$-1/3$	$-3/10$	$-1/3$	$-1/2$

Lemma 6.1 to achieve a global accuracy of order h^M and therefore local L^2 -norm of order $h^{M+1/2}$ for this integral, requires that $Q = h^{\ell-1-2M}$. \square

In Table 7.1 we quote values for the exponent $-M/\beta(M, \ell)$ for different values of M and ℓ in the case when all the distinct indices $\alpha_1, \dots, \alpha_\ell$ are non-zero.

Suppose we are given a stochastic differential equation driven by a d -dimensional Wiener process with non-commuting governing vector fields. To successfully implement a strong numerical method of order M we must guarantee that the global error associated with each multiple integral present in the integrator that is approximated by its conditional expectation is also of order M . If we implement a numerical method of order $M \leq d/2$, we will *in general* be required to simulate multiple Stratonovich integrals with distinct indices of length ℓ with $2 \leq \ell \leq 2M \leq d$. We will also have to simulate multiple integrals with repeated indices of length $\ell_r \leq 2M$. These integrals will require the same or fewer quadrature points than those with distinct indices as we can take advantage of the repeated indices—see Section 8 for more details. Such integrals therefore represent lower order corrections to the quadrature effort. Similarly multiple integrals with distinct indices that involve a zero index have index length that is one less than similar order multiple integrals with distinct non-zero indices. Hence they will also represent lower order corrections to the quadrature effort.

To examine the scaling exponent in the relation between the global error \mathcal{E} and the quadrature effort $\mathcal{U}^{\text{quad}}$, which is the sum total of the efforts required to approximate all the required multiple integrals to order h^M , we use Table 7.1 as a guide for the dominant scalings. For methods of order $M \leq d/2$, if $d = 2$ and we implement a method of order $M = 1$, then the dominant exponent is $-1/2$. Similarly if $d = 3$ then order 1 and $3/2$ methods also invoke a dominant scaling exponent of $-1/2$ for the integrals of length $\ell = 2$ and $\ell = 3$. If $d = 4$ then methods of order 1 and $3/2$ have the same scaling exponent of $-1/2$, however the method of order 2 involves multiple integrals with three indices which are all distinct and the dominant scaling exponent for them is $-2/5$.

If we implement a method of order $M > d/2$ then we will be required to simulate multiple Stratonovich integrals with distinct indices of length ℓ with $2 \leq \ell \leq d$. We must also simulate higher order multiple integrals with indices of length ℓ_r involving repetitions with $d \leq \ell_r \leq 2M$; these may be cheaper to simulate than multiple integrals of the same length with distinct indices (again see Section 8). When $d = 2$ the dominant scaling exponent is $-1/2$ for all orders. For $d \geq 3$ the dominant scaling exponent is *at best* $-1/2$, and so forth.

Lastly, we give an estimate for the critical stepsize h_{cr} below which the quadrature effort dominates the evaluation effort. Since $\ell = M + 1$ minimizes $\beta(M, \ell)$ we have

the following estimate.

COROLLARY 7.2. *For the case of general non-commuting governing vector fields and a numerical approximation of order M , we have $\mathcal{U}^{\text{eval}} \geq \mathcal{U}^{\text{quad}}$ if and only if $h \geq h_{\text{cr}}$ where the critical stepsize*

$$h_{\text{cr}} = \mathcal{O}\left(\left(T(c_M P^2 + c_E)\right)^{-1/(1-\beta(M, \ell_{\text{max}}))}\right),$$

where $\ell_{\text{max}} = \max\{d, M + 1\}$.

In practice when we implement numerical methods for stochastic differential equations driven by a d -dimensional Wiener process we expect that for $h \geq h_{\text{cr}}$ the evaluation effort dominates the computational cost. In this scenario integrators of order M scale like their deterministic counterparts. Consider what we might expect to see in a log-log plot of global error verses computational cost. As a function of increasing computational cost we expect the global error for each method to fan out with slope $-M$, with higher order methods providing superior accuracy for a given effort. However once the quadrature effort starts to dominate, the scaling exponents described above take over. When $d = 2$ for example and all methods dress themselves with the scaling exponent $-1/2$, then we expect to see parallel graphs with higher order methods still providing superior accuracy for a given cost. However higher order methods that assume a scaling exponent worse than $-1/2$ will eventually re-intersect the graphs of their lower order counterparts and past that regime should not be used.

Note that in the case when all the diffusion vector fields commute, methods of order 1 do not involve any quadrature effort and hence $\mathcal{E} = \mathcal{O}((\mathcal{U}^{\text{eval}})^{-1})$. Using Lemma 6.1, we can by analogy with the arguments in the proof of Lemma 7.1, determine the dominant scaling exponents for methods of order $M \geq 3/2$. For example the L^2 -error associated with approximating J_{0i} by its expectation conditioned on intervening information is of order $h^{3/2}/Q$. Hence we need only choose $Q = h^{-1/2}$ to achieve to achieve a global error of order $3/2$. In this case the dominant scaling exponent is -1 . However the L^2 -error associated with approximating J_{0ij} for $i \neq j$ is of order $h^2/Q^{1/2}$ whereas for J_{i0j} and J_{ij0} it is of order h^2/Q . For the case of diffusing vector fields we do not need to simulate J_{0ij} , and so for a method of order 2 the dominant scaling exponent is still -1 . However more generally the effort associated with approximating J_{0ij} dominates the effort associated with the other two integrals.

8. Efficient quadrature bases. When multiple Stratonovich integrals contain repeated indices, are they as cheap to compute as the corresponding lower dimensional integrals with an equal number of distinct indices (none of them repeated)?

Let $i \cdots i_p$ denote the multi-index with p copies of the index i . Repeated integration by parts yields the formulae

$$J_{i \cdots i_p j i \cdots i_q} = \sum_{k=1}^q (-1)^{k+1} J_{i \cdots i_k} J_{i \cdots i_p j i \cdots i_{q-k}} + (-1)^{q+2} \int J_{i \cdots i_p} J_{i \cdots i_q} dJ_j, \quad (8.1a)$$

$$J_{i \cdots i_p j \cdots j_q} = \sum_{k=1}^{q-1} (-1)^{k+1} J_{j \cdots j_k} J_{i \cdots i_p j \cdots j_{q-k}} + (-1)^{q+1} \int J_{i \cdots i_p} dJ_{j \cdots j_q}. \quad (8.1b)$$

The first relation (8.1a) suggests that any integral of the form $J_{i \cdots i_p j i \cdots i_q}$ can always be approximated by a single sum. This last statement is true for $q = 1$. If we assume it is true for $q - 1$ and apply the relation (8.1a) we establish by induction that $J_{i \cdots i_p j i \cdots i_q}$ can be approximated by a single sum. A similar induction argument using (8.1b) then

also establishes that any integral of the form $J_{i \dots i_p j \dots j_q}$ can also be approximated by a single sum. Hence in both cases the quadrature effort is proportional to QN .

Implicit in the relations (8.1) is the natural underlying shuffle algebra created by integration by parts (see Gaines [20, 19], Kawksi [32] and Munthe-Kaas and Wright [48]). Two further results are of interest. Firstly we remark that by integration by parts we have the following two shuffle product results:

$$J_{i_1 i_2 i_3} J_{i_4} = J_{i_1 i_2 i_3 i_4} + J_{i_1 i_2 i_4 i_3} + J_{i_1 i_4 i_2 i_3} + J_{i_4 i_1 i_2 i_3}, \quad (8.2a)$$

$$J_{i_1 i_2} J_{i_3 i_4} = J_{i_1 i_2 i_3 i_4} + J_{i_1 i_3 i_2 i_4} + J_{i_3 i_1 i_2 i_4} + J_{i_3 i_4 i_1 i_2} + J_{i_3 i_1 i_4 i_2} + J_{i_1 i_3 i_4 i_2}. \quad (8.2b)$$

If we replace $\{i_1, i_2, i_3, i_4\}$ by $\{i, i, j, j\}$ in (8.2b) and (8.2a) and then by $\{i, j, i, j\}$ in (8.2a) and (8.2b), respectively, we obtain the linear system of equations

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} J_{jijj} \\ J_{ijji} \\ J_{jii}j \\ J_{ijij} \end{pmatrix} = \begin{pmatrix} J_{ii}J_{jj} - J_{iijj} - J_{jjii} \\ J_{iji}J_j - J_{jjii} \\ J_{iij}J_j - 2J_{iijj} \\ J_{ij}J_{ij} - 4J_{iijj} \end{pmatrix}. \quad (8.3)$$

By direct inspection the coefficient matrix on the left-hand side has rank 4 and so all the multiple Stratonovich integrals J_{jijj} , J_{ijji} , $J_{jii}j$ and J_{ijij} can be expressed in terms of J_{iijj} and J_{jjii} and products of lower order integrals, all of which can be approximated by single sums (note that $J_{iji} \equiv J_{ij}J_i - 2J_{iij}$).

Now consider the set of multiple Stratonovich integrals

$$\mathfrak{J} = \{J_{i_1 i_2 i_3 i_4 i_5} : \{i_1, i_2, i_3, i_4, i_5\} \in \text{perms}\{i, i, i, j, j\}\} \setminus \{J_{iijj}, J_{jjii}\},$$

where we exclude the elements J_{iijj} and J_{jjii} which we know can be approximated by single sums from (8.1b). By considering the shuffle relations generated by products of the form: $J_{i_1 i_2 i_3 i_4} J_{i_5}$, $J_{i_1 i_2 i_3} J_{i_4 i_5}$, $J_{i_1 i_2} J_{i_3 i_4 i_5}$ and $J_{i_1} J_{i_2} J_{i_3} J_{i_4} J_{i_5}$ and substituting in the 10 elements with indices from ‘perms $\{i, i, i, j, j\}$ ’ we obtain a linear system of equations analogous to (8.3) with 50 equations for the 8 unknowns in \mathfrak{J} . However direct calculation shows that the corresponding coefficient matrix has rank 7. In particular, all of the multiple integrals in \mathfrak{J} can be expressed in terms of J_{iijj} , J_{jjii} and J_{jijji} . Hence the set of multiple integrals with indices from ‘perms $\{i, i, i, j, j\}$ ’ cannot all be approximated by single sums, but in fact require a double sum to approximate J_{jijji} .

For simplicity assume $d = 1$. Consider numerical schemes of increasing order M . If $3/2 \leq M \leq 3$ all the necessary multiple integrals can be approximated by single sums—at the highest order in this range indices involving permutations of $\{1, 1, 1, 1, 0\}$ and $\{1, 1, 0, 0\}$ are included for which the corresponding integrals can be approximated by single sums. For methods of order $M \geq 7/2$ we require at least double sums to approximate the necessary multiple integrals.

When $d = 2$, for methods of order $M = 1, 3/2$ the integrals involved can be approximated by single sums, but for $M = 2$ integrals involving indices with permutations of $\{2, 1, 0\}$ are included which can only be approximated by double sums. If there were no drift vector field then for $1 \leq M \leq 2$ the necessary multiple integrals can all be approximated by single sums, but for $M = 5/2$ we need to include multiple integrals involving indices with permutations of $\{2, 2, 1, 1, 1\}$ which require approximation by double sums. We can in principle extend these results to higher values M , however methods of order $M \geq d/2$ for $d \geq 3$ are not commonly implemented!

9. Numerical simulations.

9.1. Riccati system. Our first application is for stochastic Riccati differential systems—some classes of which can be reformulated as linear systems (see Freiling [18] and Schiff and Shnider [50]). Such systems arise in stochastic linear-quadratic optimal control problems, for example, mean-variance hedging in finance (see Bobrovnytska and Schweizer [5] and Kohlmann and Tang [34])—though often these are backward problems (which we intend to investigate in a separate study). Consider for example Riccati equations of the form

$$u_t = u_0 + \sum_{i=0}^d \int_0^t (u_\tau A_i(\tau) u_\tau + B_i(\tau) u_\tau + u_\tau C_i(\tau) + D_i(\tau)) dW_\tau^i.$$

If $y = (U \ V)^T$ satisfies the linear stochastic differential system (1.1), with

$$a_i(t) \equiv \begin{pmatrix} B_i(t) & D_i(t) \\ -A_i(t) & -C_i(t) \end{pmatrix},$$

then $u = UV^{-1}$ solves the Riccati equation above.

We consider here a Riccati problem with two additive Wiener processes, W^1 and W^2 , and coefficient matrices

$$D_0 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & 1 \\ -\frac{1}{2} & -\frac{51}{200} \end{pmatrix} \quad \text{and} \quad D_2 = \begin{pmatrix} 1 & 1 \\ 1 & \frac{1}{2} \end{pmatrix}, \quad (9.1)$$

and

$$A_0 = \begin{pmatrix} -1 & 1 \\ -\frac{1}{2} & -1 \end{pmatrix}, \quad \text{and} \quad C_0 = \begin{pmatrix} -\frac{1}{2} & 0 \\ -1 & -1 \end{pmatrix}.$$

All other coefficient matrices are zero. The initial data is the 2×2 identity matrix, i.e. $u_0 = I_2$ and therefore $U_0 = I_2$ and $V_0 = I_2$ also. We found Higham [28] a very useful starting point for our Matlab simulations.

Note that for this example the coefficient matrices a_1 and a_2 are upper right block triangular and therefore nilpotent of degree 2, and also that $a_1 a_2$ and $a_2 a_1$ are identically zero so that in particular $[a_1, a_2] = 0$. The number of terms in each integrator at either order 1 or 3/2 is roughly equal, and so for a given stepsize the uniformly accurate Magnus integrators should be more expensive to compute due to the cost of computing the 4×4 matrix exponential—we used a (6, 6) Padé approximation with scaling to compute the matrix exponential. See Moler and Van Loan [47] and also Iserles and Zanna [31], the computational cost is roughly 6 times the system size cubed. Also note the order 1 integrators do not involve quadrature effort whilst the order 3/2 integrators involve the quadrature effort associated with approximating J_{10} and J_{20} . For comparison, we use a nonlinear Runge–Kutta type order 3/2 scheme for the case of two additive noise terms (from Kloeden and Platen [33, p. 383]) applied directly to the original Riccati equation:

$$\begin{aligned} S_{t_n, t_{n+1}} &= S_{t_n} + f(S_{t_n})h + D_1 J_1 + D_2 J_2 \\ &\quad + \frac{h}{4} (f(Y_1^+) + f(Y_1^-) + f(Y_2^+) + f(Y_2^-) - 4f(S_{t_n})) \\ &\quad + \frac{1}{2\sqrt{h}} ((f(Y_1^+) - f(Y_1^-))J_{10} + (f(Y_2^+) - f(Y_2^-))J_{20}), \end{aligned} \quad (9.2)$$

where $Y_j^\pm = S_{t_n} + \frac{h}{2} f(S_{t_n}) \pm D_j \sqrt{h}$ and $f(S) = SA_0 S + B_0 S + SC_0 + D_0$.

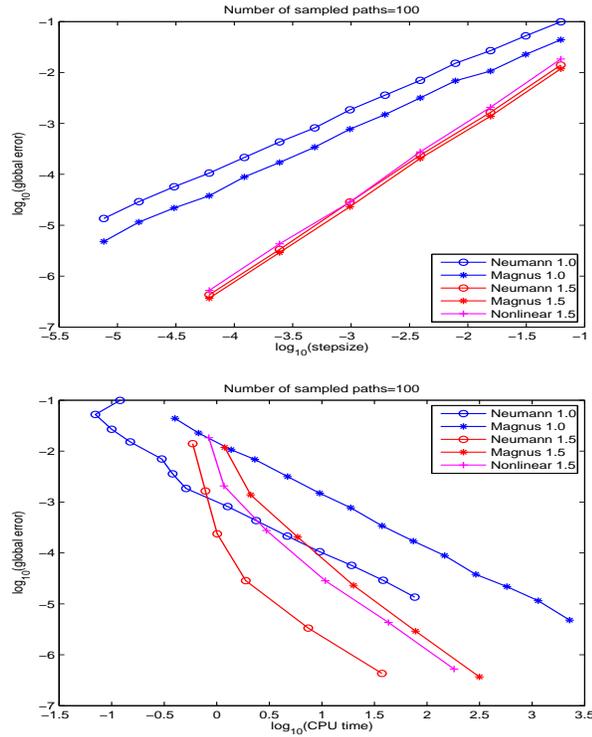


FIG. 9.1. Global error vs stepsize (top) and vs CPU clocktime (bottom) for the Riccati problem at time $t = 1$. The Magnus integrators of order 1 and 3/2 shown are the uniformly accurate Magnus integrators from Section 4.

In Figure 9.1 we show how the global error scales with stepsize and also CPU clocktime for this Riccati problem. Note that as anticipated, for the same step size (compare respective plot points starting from the left), the order 1 Magnus integrator is more expensive to compute and more accurate than the order 1 Neumann integrator. Now compare the order 3/2 integrators. For the nonlinear scheme (9.2), we must evaluate $f(S)$ five times per step per path costing $20p^3 + 54p^2$ flops—here $p = 2$ refers to the size of the original Riccati system. For the Neumann and Magnus integrators the evaluation costs are $16(2p \times p) = 32p^2$ and $6(2p)^3 + 11(2p)^2 = 48p^3 + 44p^2$ flops, respectively (directly counting from the schemes). Hence for large stepsize we expect the Neumann integrator to be cheapest and the Magnus and nonlinear Runge–Kutta integrators to be more expensive. However for much smaller stepsizes the quadrature effort should start to dominate. The efforts of all the order 3/2 integrators will not be much different and the Magnus integrator then outperforms the other two due to its superior accuracy.

9.2. Linear system. Our second numerical example is for a homogeneous and constant coefficient linear problem involving two Wiener processes with coefficient matrices $a_i \equiv D_i$ where the D_i , $i = 0, 1, 2$ are given in (9.1) and do *not commute*, and initial data $y_0 = (\frac{1}{2} \ 1)^T$. In Figure 9.2 we show how the error scales with stepsize and CPU clocktime. We see that the superior accuracy of the Magnus integrators is achieved for the same computational cost. Note that in the case $M = 1$ we have

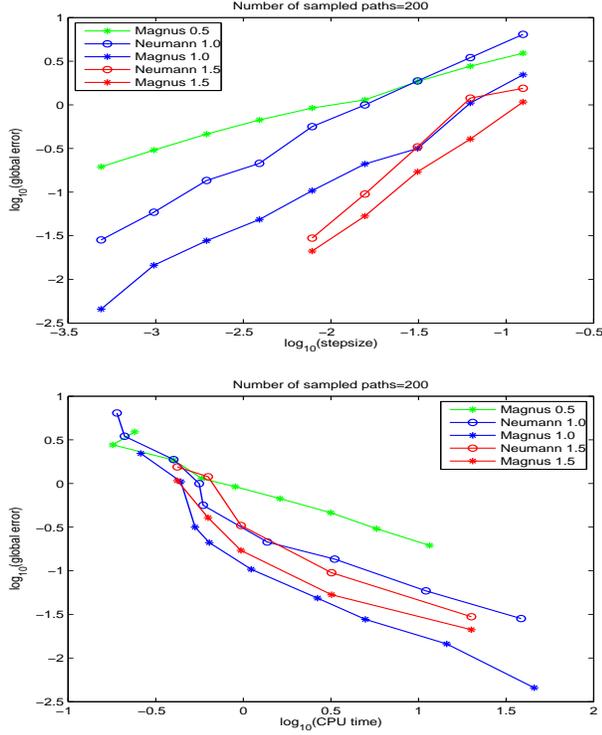


FIG. 9.2. Global error vs stepsize (top) and vs CPU clocktime (bottom) for the model problem at time $t = 1$ with 2 driving Wiener processes. The error corresponding to the largest step size takes the shortest time to compute.

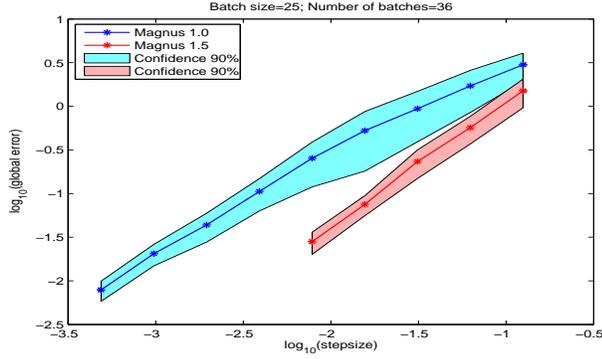


FIG. 9.3. Confidence intervals for the global errors of the uniformly accurate Magnus integrators for the model problem at time $t = 1$ with 2 driving Wiener processes.

$d = \ell = 2$. For the case when $h \leq h_{cr}$, when computational cost is dominated by quadrature effort, the relation between the global error \mathcal{E}_1 and computational cost \mathcal{U} is, ignoring $\mathcal{U}^{\text{eval}}$ and taking the logarithm,

$$\log \mathcal{E}_1 \approx \log K_1 - \frac{1}{2} \log \mathcal{U}.$$

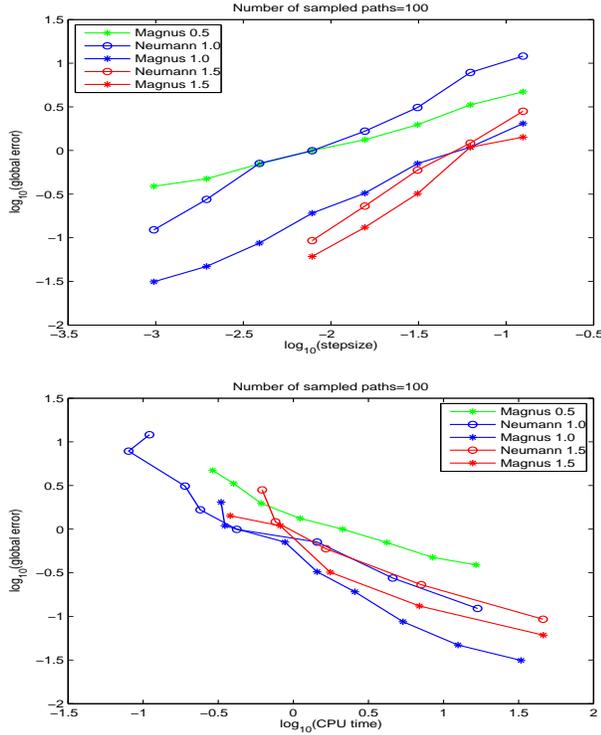


FIG. 9.4. Global error vs stepsize (top) and vs CPU clocktime (bottom) for the model problem at time $t = 1$ with 3 driving Wiener processes. The error corresponding to the largest step size takes the shortest time to compute.

For the order 1/2 Magnus method the computational cost \mathcal{U} is given solely by the evaluation effort and therefore we have

$$\log \mathcal{E}_{1/2} = \log K_{1/2} + \frac{1}{2} \log T(c_{1/2}p^2 + c_E) - \frac{1}{2} \log \mathcal{U}.$$

Further note that $K_{1/2}$ and K_1 are strictly order one, and that $T = 1$, $p = 2$, $c_{1/2} = 5$ and $c_1 = 7$ (counting directly from the integration scheme). In addition $c_E \approx 6p^3 = 48$ flops using the (6, 6) Padé approximation with scaling. Substituting these values into the difference of these last two estimates reveals that

$$\log \mathcal{E}_1 - \log \mathcal{E}_{1/2} \approx -\frac{1}{2} \log 68 \approx -0.9,$$

which is in good agreement with the difference shown in Figure 9.2.

Also for this example, we have shown 90% confidence intervals in Figure 9.3 for the global errors of the uniformly accurate Magnus integrators of orders 1 and 3/2. The confidence intervals become narrower as the stepsize decreases and order of the method increases, as expected—see Kloeden and Platen [33, pp. 312–316].

In Figure 9.4 we consider a linear stochastic differential system driven by *three* independent scalar Wiener processes, with the same vector fields as in the linear system with two Wiener processes just considered but with an additional linear diffusion vector field characterized by the coefficient matrix

$$a_3 = \begin{pmatrix} \frac{1}{4} & \frac{2}{5} \\ \frac{1}{6} & \frac{1}{7} \end{pmatrix},$$

which does not commute with a_0 , a_1 or a_2 . Using Table 7.1 we would expect to see for small stepsizes for the order 1 and 3/2 Magnus and Neumann methods, that the global error scales with the computational effort with exponent $-1/2$. This can be seen in Figure 9.4.

10. Concluding remarks. Our results can be extended to nonlinear stochastic differential equations. Consider a stochastic differential system governed by $(d + 1)$ nonlinear autonomous vector fields $V_i(y)$ —instead of the linear vector fields ‘ $a_i(t)y$ ’ in (1.1)—and driven by a d -dimensional Wiener process (W^1, \dots, W^d) . If we take the logarithm of the stochastic Taylor series for the flow-map we obtain the exponential Lie series (see Chen [15] and Strichartz [54])

$$\sigma_t = \sum_{i=0}^d J_i(t)V_i + \sum_{j>i=0}^d \frac{1}{2}(J_{ij} - J_{ji})(t)[V_i, V_j] + \dots$$

Here $[\cdot, \cdot]$ is the Lie–Jacobi bracket on the Lie algebra of vectors fields defined on \mathbb{R}^p . The solution y_t of the system at time $t > 0$ is given by $y_t = \exp \sigma_t \circ y_0$ (see for example Ben Arous [4] or Castell and Gaines [13, 14]). Across the interval $[t_n, t_{n+1}]$ let $\hat{\sigma}_{t_n, t_{n+1}}$ denote the exponential Lie series truncated to a given order, with multiple integrals approximated by their expectations conditioned on intervening sample points. Then

$$\hat{y}_{t_{n+1}} = \exp(\hat{\sigma}_{t_n, t_{n+1}}) \circ y_{t_n}$$

is an approximation to the exact solution $y_{t_{n+1}}$ at the end point of the interval. The truncated and conditioned exponential Lie series $\hat{\sigma}_{t_n, t_{n+1}}$ is itself an ordinary vector field. Hence to compute $\hat{y}_{t_{n+1}}$ we solve the ordinary differential system

$$u'(\tau) = \hat{\sigma}_{t_n, t_{n+1}} \circ u(\tau)$$

with $u(0) = y_{t_n}$ across the interval $\tau \in [0, 1]$ (see Castell and Gaines [13, 14]). If we use a sufficiently accurate ordinary differential integrator commensurate with the order of the truncation of the exponential Lie series then $u(1) \approx \hat{y}_{t_{n+1}}$ to that order.

Hence the order 1/2 exponential Lie series integrator is more accurate than the Euler–Maruyama method. Further in the case of commuting diffusion vector fields, the uniformly accurate exponential Lie series integrators of order 1 and 3/2 are more accurate than the stochastic Taylor approximations of the corresponding order. This generalization is discussed in Malham and Wiese [42]. An important future investigation to justify the viability of these schemes would be the relation between global error and computational cost, which must take into account the additional computational effort associated with the ordinary differential solver.

An important application for our results that we have in mind for the future are large order problems driven by a large number of Wiener processes. High-dimensional problems occur in many financial applications, for example in portfolio optimization or risk management and in the context of option pricing when high-dimensional models are used, for example for the pricing of interest rate options or rainbow options. Large order problems also arise when numerically solving stochastic parabolic partial differential equation driven by a multiplicative noise term which is white noise in time and spatially smooth (see for example Lord and Shardlow [37]). Here we think of projecting the system onto a finite spatial basis set which results in a large system of coupled ordinary stochastic differential equations each driven by a multiplicative noise term. The high dimension d of the driving Wiener process will now be an

important factor in the computational cost for order $M \geq 1$ as for example we will need to simulate $\frac{1}{2}d(d-1)$ multiple integrals J_{ij} ; though the results of Wiktorsson [58] suggest this can be improved upon. Krylov subspace methods for computing large matrix exponentials would be important for efficient implementation of our methods for this case (see Moler and Van Loan [47] and Sidje [52]).

Lastly, extensions of our work that we also intend to investigate further are: (1) implementing a variable step scheme following Gaines and Lyons [22], Lamba, Mattingly and Stuart [36], Burrage and Burrage [10] and Burrage, Burrage and Tian [9]—by using analytic expressions for the local truncation errors (see Aparicio, Malham and Oliver [1]); (2) pricing path-dependent options; (3) deriving strong symplectic numerical methods (see Milstein, Repin and Tretyakov [45]); and (4) constructing numerical methods based on the nonlinear Magnus expansions of Casas and Iserles [11].

Acknowledgments. We thank Kevin and Pamela Burrage, Sandy Davie, Terry Lyons, Per-Christian Moan, Nigel Newton, Tony Shardlow, Josef Teichmann and Michael Tretyakov for stimulating discussions. The research in this paper was supported by the EPSRC first grant number GR/S22134/01. A significant portion of this paper was completed while SJAM was visiting the Isaac Newton Institute in the Spring of 2007. We would also like to thank the anonymous referees for their very helpful comments and suggestions.

Appendix A. Neumann and Magnus integrators. We present Neumann and Magnus integrators up to global order 3/2 in the case of a d -dimensional Wiener process (W^1, \dots, W^d) , and with constant coefficient matrices a_0 and a_i , $i = 1, \dots, d$. The Neumann expansion for the solution of the stochastic differential equation (1.1) over an interval $[t_n, t_{n+1}]$, where $t_n = nh$, is

$$y_{t_n, t_{n+1}}^{\text{neu}} \approx (I + S_{1/2} + S_1 + S_{3/2}) y_{t_n},$$

where (the indices i, j, k run over the values $1, \dots, d$)

$$\begin{aligned} S_{1/2} &= J_0 a_0 + \sum_i J_i a_i + \sum_i J_{ii} a_i^2, \\ S_1 &= \sum_{i \neq j} J_{ij} a_i a_j, \\ S_{3/2} &= \sum_i (J_{i0} a_0 a_i + J_{0i} a_i a_0) + \sum_{i, j, k} J_{kji} a_i a_j a_k \\ &\quad + \sum_i (J_{0ii} a_i^2 a_0 + J_{i0i} a_0 a_i^2) + \sum_{i, j} J_{iij} a_i^2 a_j^2 + J_{00} a_0^2. \end{aligned}$$

The corresponding approximation using the Magnus expansion is

$$y_{t_n, t_{n+1}}^{\text{mag}} \approx \exp(s_{1/2} + s_1 + s_{3/2}) y_{t_n},$$

where, with $[\cdot, \cdot]$ as the matrix commutator,

$$\begin{aligned}
s_{1/2} &= J_0 a_0 + \sum_i J_i a_i, \\
s_1 &= \sum_{i < j} \frac{1}{2} (J_{ji} - J_{ij}) [a_i, a_j], \\
s_{3/2} &= \sum_i \frac{1}{2} (J_{i0} - J_{0i}) [a_0, a_i] + \sum_{i \neq j} (J_{ij} - \frac{1}{2} J_i J_{ij} + \frac{1}{12} J_i^2 J_j) [a_i, [a_i, a_j]] \\
&\quad + \sum_{i < j < k} ((J_{ijk} + \frac{1}{2} J_j J_{ki} + \frac{1}{2} J_k J_{ji} - \frac{2}{3} J_i J_j J_k) [a_i, [a_j, a_k]] \\
&\quad\quad + (J_{jik} + \frac{1}{2} J_i J_{kj} + \frac{1}{2} J_k J_{ji} - \frac{2}{3} J_i J_j J_k) [a_j, [a_i, a_k]]) \\
&\quad + \sum_i (J_{ii0} - \frac{1}{2} J_i J_{i0} + \frac{1}{12} J_i^2 J_0) [a_i, [a_i, a_0]].
\end{aligned}$$

To obtain a numerical scheme of global order M using the Neumann or Magnus expansion, we must use all the terms up to and including S_M or s_M , respectively. Leading order terms of order $M + 1/2$ with non-zero expectation can be replaced by their expectations (as detailed at the end of Section 3). Further, extending these solution series to the non-homogeneous and/or the non-constant coefficient case is straightforward.

REFERENCES

- [1] N. D. APARICIO, S. J. A. MALHAM, AND M. OLIVER, *Numerical evaluation of the Evans function by Magnus integration*, BIT, 45 (2005), pp. 219–258.
- [2] L. ARNOLD, *Stochastic differential equations*, Wiley, 1974.
- [3] R. AZENCOTT, *Formule de Taylor stochastique et développement asymptotique d'intégrales de Feynman*, Seminar on Probability XVI, Lecture Notes in Math., 921 (1982), Springer, pp. 237–285.
- [4] G. BEN AROUS, *Flots et series de Taylor stochastiques*, Probab. Theory Related Fields, 81 (1989), pp. 29–77.
- [5] O. BOBROVNYTSKA AND M. SCHWEIZER, *Mean-variance hedging and stochastic control: beyond the Brownian setting*, IEEE Trans. on Automatic Control, 49(3) (2004), pp. 396–408.
- [6] F. BAUDOIN, *An introduction to the geometry of stochastic flows*, Imperial College Press, 2004.
- [7] P. M. BURRAGE, *Runge–Kutta methods for stochastic differential equations*, Ph.D. thesis, University of Queensland, 1999.
- [8] K. BURRAGE AND P. M. BURRAGE, *High strong order methods for non-commutative stochastic ordinary differential equation systems and the Magnus formula*, Phys. D, 133 (1999), pp. 34–48.
- [9] K. BURRAGE, P. M. BURRAGE, AND T. TIAN, *Numerical methods for strong solutions of stochastic differential equations: an overview*, Proc. R. Soc. Lond. A, 460 (2004), pp. 373–402.
- [10] P. M. BURRAGE AND K. BURRAGE, *A variable stepsize implementation for stochastic differential equations*, SIAM J. Sci. Comput., 24(3) (2002), pp. 848–864.
- [11] F. CASAS AND A. ISERLES, *Explicit Magnus expansions for nonlinear equations*, Technical report NA2005/05, DAMTP, University of Cambridge, 2005.
- [12] F. CASTELL, *Asymptotic expansion of stochastic flows*, Probab. Theory Related Fields, 96 (1993), pp. 225–239.
- [13] F. CASTELL AND J. GAINES, *An efficient approximation method for stochastic differential equations by means of the exponential Lie series*, Math. Comput. Simulation, 38 (1995), pp. 13–19.
- [14] ———, *The ordinary differential equation approach to asymptotically efficient schemes for solution of stochastic differential equations*, Ann. Inst. H. Poincaré Probab. Statist., 32(2) (1996), pp. 231–250.
- [15] K. T. CHEN, *Integration of paths, geometric invariants and a generalized Baker–Hausdorff formula*, Annals of Mathematics, 65(1) (1957), pp. 163–178.

- [16] J. M. C. CLARK AND R. J. CAMERON, *The maximum rate of convergence of discrete approximations for stochastic differential equations*, in Lecture Notes in Control and Information Sciences, Vol. 25, 1980, pp. 162–171.
- [17] A.B. CRUZEIRO, P. MALLIAVIN AND A. THALMAIER, *Geometrization of Monte-Carlo numerical analysis of an elliptic operator: strong approximation*, C. R. Acad. Sci. Paris, Ser. I, 338 (2004), pp. 481–486.
- [18] G. FREILING, *A survey of nonsymmetric Riccati equations*, Linear Algebra Appl., 351–352 (2002), pp. 243–270.
- [19] J. G. GAINES, *The algebra of iterated stochastic integrals*, Stochastics and stochastics reports, 49(3–4) (1994), 169–179.
- [20] ———, *A basis for iterated stochastic integrals*, Math. Comput. Simulation, 38 (1995), pp. 7–11.
- [21] J. G. GAINES AND T. J. LYONS, *Random generation of stochastic area integrals*, SIAM J. Appl. Math., 54(4) (1994), pp. 1132–1146.
- [22] ———, *Variable step size control in the numerical solution of stochastic differential equations*, SIAM J. Appl. Math., 57(5) (1997), pp. 1455–1484.
- [23] I. I. GIHMAN, AND A. V. SKOROHOD, *The theory of stochastic processes III*, Springer, 1979.
- [24] M.B. GILES, *Multi-level Monte Carlo path simulation*, OUCL, Report no. 06/03.
- [25] ———, *Improved multilevel Monte Carlo convergence using the Milstein scheme*, OUCL, Report no. 06/22.
- [26] H. GILSING AND T. SHARDLOW, *SDELab: stochastic differential equations with MATLAB*, MIMS EPrint: 2006.1, ISSN 1749-9097.
- [27] I. GYÖNGY, AND G. MICHALETZKY, *On Wong–Zakai approximations with δ -martingales*, Proc. R. Soc. Lond. A, 460 (2004), pp. 309–324.
- [28] D. J. HIGHAM, *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Rev., 43(3) (2001), pp. 525–546.
- [29] N. HOFMANN, AND T. MÜLLER-GRONBACH, *On the global error of Itô–Taylor schemes for strong approximation of scalar stochastic differential equations*, J. Complexity, 20 (2004), pp. 732–752.
- [30] A. ISERLES, H. Z. MUNTHE-KAAS, S. P. NØRSETT, AND A. ZANNA, *Lie-group methods*, Acta Numer., (2000), pp. 215–365.
- [31] A. ISERLES AND A. ZANNA, *Efficient computation of the matrix exponential by generalized polar decompositions*, SIAM J. Numer. Anal., 42(5) (2005), pp. 2218–2256.
- [32] M. KAWSKI, *The combinatorics of nonlinear controllability and noncommuting flows*, Lectures given at the Summer School on Mathematical Control Theory, Trieste, September 3–28, 2001.
- [33] P. E. KLOEDEN AND E. PLATEN, *Numerical solution of stochastic differential equations*, Springer, 1999.
- [34] M. KOHLMANN AND S. TANG, *Multidimensional backward stochastic Riccati equations and applications*, SIAM J. Control Optim., 41(6) (2003), pp. 1696–1721.
- [35] H. KUNITA, *On the representation of solutions of stochastic differential equations*, Lecture Notes in Math. 784, Springer-Verlag, 1980, pp. 282–304.
- [36] H. LAMBA, J.C. MATTINGLY AND A.M. STUART, *An adaptive Euler-Maruyama scheme for SDEs: convergence and stability*, IMA Journal of Numerical Analysis, 27 (2007), pp. 479–506.
- [37] G.J. LORD AND T. SHARDLOW, *Post processing for stochastic parabolic partial differential equations*, SIAM J. Numer. Anal. 45(2) (2007), pp. 870–889.
- [38] T. LYONS, *Differential equations driven by rough signals*, Rev. Mat. Iberoamericana, 14(2) (1998), pp. 215–310.
- [39] T. LYONS AND Z. QIAN, *System control and rough paths*, Oxford University Press, 2002.
- [40] T. LYONS AND N. VICTOIR, *Cubature on Wiener space*, Proc. R. Soc. Lond. A, 460 (2004), pp. 169–198.
- [41] W. MAGNUS, *On the exponential solution of differential equations for a linear operator*, Comm. Pure Appl. Math., 7 (1954), pp. 649–673.
- [42] S. J.A. MALHAM AND A. WIESE, *Stochastic Lie group integrators*, SISC, accepted.
- [43] P. MALLIAVIN, *Stochastic analysis*, Grundlehren der mathematischen Wissenschaften 313, Springer, 1997.
- [44] G. N. MILSTEIN, *Numerical integration of stochastic differential equations*, Mathematics and its applications, Kluwer Academic Publishers, 1994.
- [45] G. N. MILSTEIN, YU. M. REPIN, AND M. V. TRETYAKOV, *Numerical methods for stochastic systems preserving symplectic structure*, SIAM J. Numer. Anal., 40(4) (2002), pp. 1583–1604.
- [46] T. MISAWA, *A Lie algebraic approach to numerical integration of stochastic differential equa-*

- tions, SIAM J. Sci. Comput., 23(3) (2001), pp. 866–890.
- [47] C. MOLER AND C. VAN LOAN, *Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later*, SIAM Rev., 45 (2003), pp. 3–49.
- [48] H.Z. MUNTHE-KAAS AND W.M. WRIGHT, *On the Hopf algebraic structure of Lie group integrators*, preprint, 2006.
- [49] N. J. NEWTON, *Asymptotically efficient Runge–Kutta methods for a class of Itô and Stratonovich equations*, SIAM J. Appl. Math., 51 (1991), pp. 542–567.
- [50] J. SCHIFF AND S. SHNIDER, *A natural approach to the numerical integration of Riccati differential equations*, SIAM J. Numer. Anal., 36(5) (1999), pp. 1392–1413.
- [51] H. SCHURZ, *A brief introduction to numerical analysis of (ordinary) stochastic differential equations without tears*, in Handbook of Stochastic Analysis and Applications, V. Lakshmikantham and D. Kannan, eds., Marcel Dekker, 2002, pp. 237–359.
- [52] R. B. SIDJE, *Expokit: a software package for computing matrix exponentials*, ACM Transactions on Mathematical Software, 24(1) (1998), pp. 130–156.
- [53] E.-M. SIPILÄINEN, *A pathwise view of solutions of stochastic differential equations*, Ph.D. thesis, University of Edinburgh, 1993.
- [54] R. S. STRICHARTZ, *The Campbell–Baker–Hausdorff–Dynkin formula and solutions of differential equations*, J. Funct. Anal., 72 (1987), pp. 320–345.
- [55] D. M. STUMP AND J. M. HILL, *On an infinite integral arising in the numerical integration of stochastic differential equations*, Proc. R. Soc. A, 461 (2005), pp. 397–413.
- [56] H. J. SUSSMANN, *Product expansions of exponential Lie series and the discretization of stochastic differential equations*, in Stochastic Differential Systems, Stochastic Control Theory, and Applications, W. Fleming and J. Lions, eds., Springer IMA Series, Vol. 10, 1988, pp. 563–582.
- [57] D. TALAY, *Simulation and numerical analysis of stochastic differential systems*, in Probabilistic Methods in Applied Physics, P. Kree and W. Wedig, eds., Lecture Notes in Physics, Vol. 451, 1995, chap. 3, pp. 63–106.
- [58] M. WIKTORSSON, *Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions*, Ann. Appl. Probab., 11(2) (2001), pp. 470–487.
- [59] E. WONG AND M. ZAKAI, *On the relation between ordinary differential and stochastic differential equations*, Internat. J. Engrg Sci., 3 (1965), pp. 213–229.
- [60] Y. YAMATO, *Stochastic differential equations and nilpotent Lie algebras*, Z. Wahrsch. Verw. Gebiete, 47(2) (1979), pp 213–229.