Efficient strong integrators for linear stochastic systems

By Gabriel Lord, Simon J.A. Malham and Anke Wiese

School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK (20th February 2006)

This paper is dedicated to the memory of Nairo Aparicio, a friend and collaborator of SJAM who passed away on 20th June 2005

We present numerical schemes for the strong solution of linear stochastic differential equations driven by two Wiener processes and with non-commutative vector fields. These schemes are based on the Neumann and Magnus expansions. We prove that for a sufficiently small stepsize, the half order Magnus and a new modified order one Magnus integrator are globally more accurate than classical stochastic numerical schemes or Neumann integrators of the corresponding order. These Magnus methods will therefore always be preferable provided the cost of computing the matrix exponential is not significant. Further, for small stepsizes the accurate representation of the Lévy area between the two driving processes dominates the computational cost for all methods of order one and higher. As a consequence, we show that the accuracy of *all* stochastic integrators asymptotically scales like the square-root of the computational cost. This has profound implications on the effectiveness of higher order integrators. In particular in terms of efficiency, there are generic scenarios where order one Magnus methods compete with and even outperform higher order methods. We consider the consequences in applications such as linear-quadratic optimal control, filtering problems and the pricing of path-dependent financial derivatives.

Keywords: linear stochastic differential equations, numerical methods, strong convergence, stochastic linear-quadratic control

1. Introduction

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

$$S(t) = I + \sum_{i=0}^{d} \int_0^t a_i(\tau) S(\tau) \, \mathrm{d}W_i(\tau), \tag{1.1}$$

or more succinctly,

$$S = I + \mathsf{K} \circ S, \tag{1.2}$$

where $W_0(t) \equiv t$ and $W_i(t)$, for $i = 1, \dots, d$, are independent scalar Wiener processes and $a_0(t)$ and $a_i(t)$ are given $n \times n$ coefficient matrices. In the abbreviated form (1.2), we set

$$\mathsf{K} \equiv \mathsf{K}_0 + \mathsf{K}_1 + \cdots + \mathsf{K}_d,$$

where the K_i, $i = 0, \ldots, d$, are the linear integral operators

$$(\mathsf{K}_i \circ S)(t) \equiv \int_0^t a_i(\tau) S(\tau) \, \mathrm{d}W_i(\tau) \,.$$

We can think of $S(t)$ as the fundamental matrix solution or flow-map associated with a linear stochastic differential equation of exactly the same form as (1.1), except for an n-vector of unknowns $Y(t)$ starting with initial data Y_0 at time $t = 0$ (rather than the identity matrix I) so that $Y(t) = S(t) Y_0$. In this paper we are specifically interested in the case of more than one Wiener process; later on for ease of exposition, we take $d = 2$.

The solution of the integral equation for S is known as the *Peano–Baker series*, *Feynman–Dyson path ordered exponential*, *Chen-Fleiss series* or *Neumann series*

$$S(t) = (\mathsf{I} - \mathsf{K})^{-1} \circ I \equiv (\mathsf{I} + \mathsf{K} + \mathsf{K}^2 + \mathsf{K}^3 + \mathsf{K}^4 + \cdots) \circ I \,.$$

The logarithm of the Neumann expansion is known as the *Magnus expansion* (Magnus 1954), i.e. we can write

$$S(t) = \exp(\sigma(t)) \,,$$

where

$$\sigma(t) = \mathsf{K} \circ I + \mathsf{K}^2 \circ I - \tfrac{1}{2} (\mathsf{K} \circ I)^2 + \cdots , \tag{1.3}$$

See Kunita (1980), Ben Arous (1989), Castel (1993) and Burrage (1999) for the derivation and convergence of the stochastic Magnus expansion; Iserles *et al.* (2000) for a deterministic review; Lyons (1998) and Sipiläinen (1993) for extensions to rough signals; Lyons & Victoir (2004) for a recent application to probabilistic methods for solving PDEs; and Sussmann (1988) for a related product expansion.

In the case when the coefficient matrices $a_i(t) = a_i$, $i = 0, \ldots, d$ are constant and non-commutative, the solution to the linear problem (1.1) is non-trivial and given by the Neumann series (we adopt the standard notation for multiple Stratonovich integrals $J_{\alpha_1 \cdots \alpha_m}(t)$—see Kloeden & Platen 1999)

$$S^{\mathrm{neu}}(t) = \sum_{m=0}^{\infty} \sum_{\alpha \in \mathbb{P}_m} a_{\alpha_m} \cdots a_{\alpha_1} J_{\alpha_1 \cdots \alpha_m}(t) \,. \tag{1.4}$$

Here $\mathbb{P}_m$ is the set of all combinations of multi-indices $\alpha = (\alpha_1, \ldots, \alpha_m)$ of length m with $\alpha_i \in \{0, 1, \ldots, d\}$. There are some special non-commutative cases when we can write down an explicit analytical solution. For example when there are only two independent scalar Wiener processes, the stochastic differential equation

$$S(t) = I + a_1 \cdot \int_0^t S(\tau) \, \mathrm{d}W_1(\tau) + \int_0^t S(\tau) \, \mathrm{d}W_2(\tau) \cdot a_2 \,,$$

has the explicit analytical solution $S(t) = \exp(a_1 W_1(t)) \cdot \exp(a_2 W_2(t))$. Here the underlying vector fields a_1 and a_2 are respectively, separately integrable left and right actions with respect to W_1 and W_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. Numerical schemes based on deterministic Runge–Kutta methods have also been derived—see Kloeden & Platen (1999) and Talay (1995). 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 & Gaines (1995), Burrage (1999) and Misawa (2001).

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 Gaines & Lyons 1997). In this context, our goal in this paper is to prove that for linear stochastic differential equations driven by two Wiener processes:

1. Superior accuracy is provided by order 1/2 and a modified class of order 1 Magnus integrators, over the corresponding Neumann integrators.

2. Accuracy of all stochastic integrators asymptotically scales like the square-root of the computational cost for small stepsizes.

Statement 1 reflects that the exponential of the Magnus series is a natural solution ansatz for linear stochastic differential equations. As a result the remainder for a Magnus integrator contains relatively fewer terms compared to the corresponding Neumann remainder. Statement 2 naturally arises in the time-ordered integration of information generated at infinitesimally small scales by the two driving Wiener signals. In particular, for small stepsizes the accurate representation of the Lévy area (or chordal area process) $\frac{1}{2}(J_{12} - J_{21})$ between the two Wiener processes dominates the computational cost for all methods of order one and higher. Coincidentally, half-order methods, which do not require the Lévy area, also naturally obey this square-root scaling.

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.

The evaluation effort is mostly much smaller for the Magnus integrators than for the Neumann integrators. This is because there are usually fewer terms in the Magnus expansion compared to the corresponding Neumann expansion to the same order. For Magnus integrators though, there is the additional computational expense associated with computing the matrix exponential. However when the cost of computing the matrix exponential is not significant, we expect Magnus integrators to be preferable to classical stochastic numerical integrators (using Statement 1). This will be the case for systems that are small or sparse or systems with underlying symmetries for which the matrix exponential is simplified and computed cheaply. This is also true 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.

When two or more Wiener processes are present, to obtain a higher order stochastic integrator we need to include the Lévy area or equivalently, the multiple Stratonovich integral J_{12}. In the variable time-step scenario, order $1/2$ integrators do not necessarily converge to the correct solution (see Gaines & Lyons 1997) and a successful integrator must include the Lévy area (see Lyons 1998, who proves that the solution is continuously controlled by the driving processes and the Lévy area).

Unfortunately, for high accuracy pathwise integrators of order 1 or more, the overall computational effort of the integrator is dominated by the quadrature effort associated with evaluating the Lévy area. Indeed evaluating the Lévy area to high orders of accuracy is computationally more expensive than evaluating even higher order multiple Stratonovich integrals. This is the well-known *bottleneck* problem associated with high order stochastic pathwise integration schemes that traditionally limits their application (see Kloeden & Platen 1999 page 367, and Schurz 2002).

While Gaines & Lyons (1994) and more recently Wiktorsson (2001) and Stump & Hill (2005) provide methods for efficiently sampling the Lévy area across a given time-step, we choose to approximate the Lévy area and all 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 variable time-step schemes (Gaines & Lyons 1997) and filtering problems where the driving processes (say W_1 and W_2) are observed signals. In such a scenario, 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 (see Gaines & Lyons 1997).

For pathwise integrators, the accurate representation of the Lévy area implies Statement 2. The consequences are startling. For small stepsizes, numerical methods of differing orders are only distinguished by the multiplicative constants (in the scaling law) which are proportional to the global error coefficients. The global error coefficients depend on linear combinations of products of the underlying vector fields. Hence for some generic scenarios, such as when the global error coefficients for methods of different orders do not differ by an order of magnitude, then in terms of efficiency, order one Magnus methods compete with and even outperform higher order methods. If methods of order $3/2$ or higher are slightly more accurate the simplicity of implementing the order 1 method might still make it preferable.

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 §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 §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 §4. We prove that for a sufficiently small stepsize, the order $1/2$ Magnus integrator and a new modified order 1 Magnus integrator are globally more accurate than their Neumann counterparts. We then turn our attention in §5 to the method of approximating multiple Stratonovich integrals by their conditional expectations (as proposed in

Gaines & Lyons 1997). We show that the accurate representation of the Lévy area dominates the quadrature effort for all methods of order 1 and higher, causing the *bottleneck*. We prove in §6 that this implies the square-root scaling law between the global error and computational effort. In §7 we present numerical experiments that reflect our theoretical results, and in particular illustrate the superior accuracy of Magnus methods (already observed by Sipiläinen 1993 and Burrage 1999). Also in §7, we apply Neumann and Magnus integrators to a stochastic Riccati differential system that can be linearized. Since for the linearized system, expensive matrix-matrix multiplications can be achieved independent of the path, the Magnus and Neumann methods perform better than an explicit Runge–Kutta type method applied directly to the nonlinear Riccati system. Lastly in §8, we outline further 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 $a_i(t) = a_i$, $i = 0, 1, \ldots, d$. 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 for two Wiener processes up to terms with L^2-norm of order 2. Let $\sigma_m(t)$ denote the truncated Magnus series

$$\sigma_m(t) = \sum_{\alpha \in \mathbb{Q}_m} c_\alpha J_\alpha \,, \tag{2.1}$$

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, $m = 1/2, 1, 3/2, \ldots$. The terms c_α are linear combinations of finitely many (more precisely exactly length α) products of the a_i, $i = 0, 1, \ldots, d$. Let $|\mathbb{Q}_m|$ denote the cardinality of $\mathbb{Q}_m$.

Theorem 2.1 (Convergence). *For any $t \leqslant 1$, the exponential of the truncated Magnus series, $\exp\big(\sigma_m(t)\big)$, is square-integrable. Further, if $S(t)$ is the solution of the stochastic differential equation (1.1), there exists a constant $C(m)$ such that*

$$\big\|S(t) - \exp\big(\sigma_m(t)\big)\big\|_{L^2} \leqslant C(m)\, t^{m+1/2} \,. \tag{2.2}$$

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

$$\big\|(\sigma_m(t))^k\big\|_{L^2} \leqslant \Big(\max_{\alpha \in \mathbb{Q}_m} \|c_\alpha\|_{\mathrm{op}}\Big)^k \cdot \sum_{\substack{\alpha_i \in \mathbb{Q}_m \\ i=1,\ldots,k}} \|J_{\alpha_1} J_{\alpha_2} \cdots J_{\alpha_k}\|_{L^2} \,. \tag{2.3}$$

Note that the maximum of the operator norm 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 α_i and α_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 $\ell(\alpha_i) \leqslant 2m$ for $i = 1, \ldots, k$, 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 $k \leqslant \ell(\beta) \leqslant 2mk$.

From equation (5.2.34) in Kloeden & Platen, 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 $\ell(\gamma) \leqslant \ell(\beta)$. Further, from Remark 5.2.8 in Kloeden & Platen, $\ell(\gamma) + n(\gamma) \geqslant \ell(\beta) + n(\beta)$, where $n(\beta)$ and $n(\gamma)$ denote the number of zeros in β and γ, respectively. From Lemma 5.7.3 in Kloeden & Platen,

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

Noting that $\ell(\gamma) \leqslant \ell(\beta) \leqslant 2mk$ and $\ell(\gamma) + n(\gamma) \geqslant k$, it follows that for $t \leqslant 1$, $\|J_\beta\|_{L^2} \leqslant 2^{4mk-1}\, t^{k/2}$. 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,

$$\left\| \left(\sigma_m(t)\right)^k \right\|_{L^2} \leqslant \left(\max_{\alpha \in \mathbb{Q}_m} \|c_\alpha\|_{\mathrm{op}} \cdot |\mathbb{Q}_m| \cdot 2^{6m} \cdot t^{1/2} \right)^k .$$

Hence $\exp\big(\sigma_m(t)\big)$ is square-integrable.

Second we prove (2.2). Let $S_m(t)$ denote Neumann series solution (1.4) truncated to included terms of order up to and including t^m. We have

$$\left\| S(t) - \exp\big(\sigma_m(t)\big) \right\|_{L^2} \leqslant \left\| S(t) - S_m(t) \right\|_{L^2} + \left\| S_m(t) - \exp\big(\sigma_m(t)\big) \right\|_{L^2} . \quad (2.4)$$

We know $S(t) \in L^2$ (see Lemma III.2.1 in Gihman & Skorohod 1979). Furthermore, for any order m, $S_m(t)$ corresponds to the truncated Taylor expansion involving terms of order up to and including t^m. Hence $S_m(t)$ is a strong approximation to $S(t)$ to that order with the remainder consisting of $\mathcal{O}(t^{m+1/2})$ terms (see Proposition 5.9.1 in Kloeden & Platen 1999). It follows from the definition of the Magnus series as the logarithm of the Neumann series, that the terms of order up to and including t^m in $\exp\big(\sigma_m(t)\big)$ correspond with $S_m(t)$; the error consists of $\mathcal{O}(t^{m+1/2})$ terms. $\qquad\square$

Remark. Ben Arous (1989) and Castell (1993) prove the remainder of the exponential of any truncation of the Magnus series is bounded in probability as $t \to 0$ (in the full nonlinear case). Our result holds 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-autonomous case with suitable conditions on the coefficient matrices (see Proposition 5.10.1 in Kloeden & Platen).

3. Global and local error

Suppose $S(t_n, t_{n+1})$ is the exact and $\hat{S}(t_n, t_{n+1})$ is the approximate fundamental solution across the interval $[t_n, t_{n+1}]$ where $t_n = nh$. Let $R(t_n, t_{n+1})$ be the difference between these exact and approximate fundamental solutions so that

$$S(t_n, t_{n+1}) = \hat{S}(t_n, t_{n+1}) + R(t_n, t_{n+1}) . \quad (3.1)$$

Definition (Local truncation error). *We define the* local truncation error *associated with any such approximation as those terms in the remainder $R(t_n, t_{n+1})$ that contribute at leading order in the stepsize h to the global error.*

Remark. The local truncation error is a matrix valued random variable constructed from the leading order terms in $R(t_n, t_{n+1})$ as well as some higher order terms that can, as we will see, contribute to the global truncation error at leading order.

Definition (Local error). *We define the* local error *as*

$$\mathcal{L} \equiv \sup_{\|Y_0\|_2=1} \|R(t_n, t_{n+1}) Y_0\|_{L^2},$$

where $\|\cdot\|_2$ *is the vector 2-norm.*

Definition (Strong global error). *We define the* strong global error *associated with an approximate solution to the stochastic differential equation* (1.1) *over the global interval of integration* $[0, T] = \cup_{n=0}^{N-1}[t_n, t_{n+1}]$ *as*

$$\mathcal{E} \equiv \sup_{\|Y_0\|_2=1} \left\| \left(\prod_{n=N-1}^{0} S(t_n, t_{n+1}) - \prod_{n=N-1}^{0} \hat{S}(t_n, t_{n+1}) \right) Y_0 \right\|_{L^2}.$$

Remark. 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 we substitute our truncation-remainder decomposition (3.1) for the exact solution into our definition for the strong global error, we get for small h,

$$\mathcal{E} = \sup_{\|Y_0\|_2=1} \left\| \left(\sum_{n=0}^{N-1} \hat{S}(t_{n+1}, t_N) R(t_n, t_{n+1}) \hat{S}(t_0, t_n) \right) Y_0 \right\|_{L^2}, \qquad (3.2)$$

up to higher order terms, in fact $\mathcal{O}(\mathcal{L}^{3/2} h^{-3/4})$. By $\hat{S}(t_n, t_m)$ with $m > n$ we mean the approximate solution across the interval $[t_n, t_m]$ constructed by composing $\hat{S}(t_k, t_{k+1})$ across the intervening intervals $[t_k, t_{k+1}]$ with $k = n, \ldots, m-1$.

Definition (Global remainder). *We identify the* global remainder *as the matrix-valued random variable*

$$\mathcal{R} \equiv \sum_{n=0}^{N-1} \hat{S}(t_{n+1}, t_N) R(t_n, t_{n+1}) \hat{S}(t_0, t_n).$$

The square of the global truncation error (3.2) at leading order is therefore

$$\mathcal{E}^2 = \sup_{\|Y_0\|_2=1} Y_0^T \, \mathbb{E}(\mathcal{R}^T \mathcal{R}) \, Y_0.$$

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

$$R(t_n, t_{n+1}) = \sum_{\alpha} A_\alpha J_\alpha(t_n, t_{n+1}). \qquad (3.3)$$

Here α is a multi-index and the terms A_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 Stratonovich integrals, including permutations of α). The global remainder thus has the form

$$\mathcal{R} \equiv \sum_{n=0}^{N-1} \sum_{\alpha} \left(\hat{S}(t_{n+1}, t_N) A_\alpha \hat{S}(t_0, t_n) \right) J_\alpha(t_n, t_{n+1}).$$

To construct the global truncation error we need to compute

$$\mathbb{E}\big(\mathcal{R}^T \mathcal{R}\big)$$
$$= \sum_{n=0}^{N-1} \sum_{\alpha,\beta} \mathbb{E}\Big(\big(\hat{S}(t_{n+1},t_N)A_\alpha \hat{S}(t_0,t_n)\big)^T \big(\hat{S}(t_{n+1},t_N)A_\beta \hat{S}(t_0,t_n)\big)\Big) \cdot$$
$$\cdot \mathbb{E}\big(J_\alpha(t_n,t_{n+1})\, J_\beta(t_n,t_{n+1})\big)$$
$$+ \sum_{n \neq m} \sum_{\alpha,\beta} \mathbb{E}\Big(\big(\hat{S}(t_{n+1},t_N)A_\alpha \hat{S}(t_0,t_n)\big)^T \big(\hat{S}(t_{m+1},t_N)A_\beta \hat{S}(t_0,t_m)\big)\Big) \cdot$$
$$\cdot \mathbb{E}\big(J_\alpha(t_n,t_{n+1})\big) \mathbb{E}\big(J_\beta(t_m,t_{m+1})\big). \qquad (3.4)$$

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 $A_\alpha J_\alpha$ with local L^2-norm up to and including $\mathcal{O}(h^M)$. The leading terms 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 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 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).

4. Uniformly accurate Magnus integrators

We can compare the local accuracy of the Neumann and Magnus integrators through the leading terms of their remainders $R(t_n,t_{n+1})$. Hereafter we only consider the case of two driving Wiener processes $W_1(t)$ and $W_2(t)$ and constant coefficient matrices a_i, $i = 0, 1, 2$. At this juncture, the reader might like to re-acquaint themselves with the explicit Neumann and Magnus integrators presented in Appendix A.

The remainder of an order M Neumann integrator $R(t_n,t_{n+1}) \equiv R_M^{\mathrm{neu}}(t_n,t_{n+1})$ is simply given by the terms not included in the Neumann approximation. For an order M Magnus integrator, suppose $\sigma_M(t_n,t_{n+1})$ is the Magnus expansion on $[t_n,t_{n+1}]$, truncated to include the term s_M, and that $\rho_M(t_n,t_{n+1})$ is the corresponding remainder, i.e.

$$\sigma(t_n,t_{n+1}) = \sigma_M(t_n,t_{n+1}) + \rho_M(t_n,t_{n+1}).$$

Then the remainder $R(t_n,t_{n+1}) = R_M^{\mathrm{mag}}(t_n,t_{n+1})$ associated with the Magnus approximation on expanding the exponentials is

$$R_M^{\mathrm{mag}}(t_n,t_{n+1}) = \exp\big(\sigma(t_n,t_{n+1})\big) - \exp\big(\sigma_M(t_n,t_{n+1})\big)$$
$$= \exp\big(\sigma_M(t_n,t_{n+1}) + \rho_M(t_n,t_{n+1})\big) - \exp\big(\sigma_M(t_n,t_{n+1})\big)$$
$$= \rho_M(t_n,t_{n+1}) + R_M^*(t_n,t_{n+1}) + \mathcal{O}(\sigma_M^2 \rho_M), \qquad (4.1)$$

where

$$R_M^*(t_n,t_{n+1}) = \tfrac{1}{2}\big(\sigma_M(t_n,t_{n+1})\rho_M(t_n,t_{n+1}) + \rho_M(t_n,t_{n+1})\sigma_M(t_n,t_{n+1})\big), \quad (4.2)$$

can contribute to the global error at leading order.

Theorem 4.1 (Local error comparison). *For a sufficiently small stepsize h, the Magnus integrators of global orders $1/2$ and 1 have smaller local error than the corresponding Neumann integrators, i.e.*

$$\mathcal{L}^{\mathrm{mag}} \leqslant \mathcal{L}^{\mathrm{neu}}.$$

Proof. Over one time interval $[t_n, t_{n+1}]$, we set

$$\hat{R}_M \equiv R_M^{\mathrm{neu}} - R_M^{\mathrm{mag}}.$$

For order $M = 1/2$ integrators, a straightforward calculation shows (up to terms of higher order that only contribute $\mathcal{O}(h)$ terms to the global error)

$$R_{1/2}^{\mathrm{mag}} = \tfrac{1}{2}[a_1, a_2](J_{21} - J_{12}) \quad \text{and} \quad \hat{R}_{1/2} = \tfrac{1}{2}(a_1 a_2 + a_2 a_1)(J_{21} + J_{12}).$$

Since these two terms are uncorrelated, we have that

$$\mathbb{E}\big((R_{1/2}^{\mathrm{neu}})^T R_{1/2}^{\mathrm{neu}}\big) = \mathbb{E}\big((R_{1/2}^{\mathrm{mag}})^T R_{1/2}^{\mathrm{mag}}\big) + \mathbb{E}\big((\hat{R}_{1/2})^T \hat{R}_{1/2}\big),$$

establishing the stated result for the order $1/2$ integrators.

For the order $M = 1$ Neumann and Magnus integrators we have (up to terms of higher order that only contribute $\mathcal{O}(h^{3/2})$ terms to the global error)

$$
\begin{aligned}
R_1^{\mathrm{mag}} = {} & \tfrac{1}{2}[a_0, a_1](J_{10} - J_{01}) + \tfrac{1}{2}[a_0, a_2](J_{20} - J_{02}) \\
& + [a_1, [a_1, a_2]]\left(J_{112} - \tfrac{1}{2}J_1 J_{12} + \tfrac{1}{12}J_1^2 J_2\right) \\
& + [a_2, [a_2, a_1]]\left(J_{221} - \tfrac{1}{2}J_2 J_{21} + \tfrac{1}{12}J_2^2 J_1\right) \\
& + \tfrac{1}{8}\left(4a_0^2 + a_1^4 + a_2^2 a_1^2 + a_1^2 a_2^2 + a_2^4 + 2(a_1^2 a_0 + a_0 a_1^2 + a_2^2 a_0 + a_0 a_2^2)\right)h^2, \\
\hat{R}_1 = {} & \tfrac{1}{2}(a_0 a_1 + a_1 a_0)J_1 J_0 + \tfrac{1}{2}(a_0 a_2 + a_2 a_0)J_2 J_0 \\
& - \tfrac{1}{12}\left[a_1, [a_1, a_2]\right]J_1^2 J_2 + \tfrac{1}{2}a_1^2 a_2 J_1 J_{21} + \tfrac{1}{2}a_2 a_1^2 J_1 J_{12} \\
& - \tfrac{1}{12}\left[a_2, [a_2, a_1]\right]J_2^2 J_1 + \tfrac{1}{2}a_2^2 a_1 J_2 J_{12} + \tfrac{1}{2}a_1 a_2^2 J_2 J_{21} \\
& + a_1^3 J_{111} + a_2^3 J_{222} + \tfrac{1}{12}\left([a_1, [a_1, a_0]] + [a_2, [a_2, a_0]]\right)h^2.
\end{aligned}
$$

Here $\hat{R}_1$ and R_1^{mag} are correlated, however a long but straightforward calculation shows that

$$\mathbb{E}\big((R_1^{\mathrm{neu}})^T R_1^{\mathrm{neu}}\big) = \mathbb{E}\big((R_1^{\mathrm{mag}})^T R_1^{\mathrm{mag}}\big) + h^3 X^{\mathrm{T}} B X + \mathcal{O}(h^{7/2}),$$

where X is the $12n \times n$ matrix consisting of the $n \times n$ blocks

$$X_1 = \big[a_1, [a_1, a_2]\big], \ X_3 = a_1^2 a_2, \ X_5 = a_2^2 a_1, \ X_7 = a_1^3, \ X_9 = a_0 a_1, \ X_{11} = a_0 a_2,$$
$$X_2 = \big[a_2, [a_2, a_1]\big], \ X_4 = a_2 a_1^2, \ X_6 = a_1 a_2^2, \ X_8 = a_2^3, \ X_{10} = a_1 a_0, \ X_{12} = a_2 a_0,$$

and B is the $12n \times 12n$ matrix consisting of $n \times n$ diagonal blocks of the form $b_{ij}I_{n \times n}$ for $i,j = 1, \ldots, 12$ and where $b = [b_{ij}]$ is given by

$$
b = \begin{pmatrix}
\frac{1}{144} & 0 & -\frac{1}{48} & -\frac{1}{48} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{144} & 0 & 0 & -\frac{1}{48} & -\frac{1}{48} & 0 & 0 & 0 & 0 & 0 & 0 \\
-\frac{1}{48} & 0 & \frac{7}{24} & \frac{1}{12} & 0 & 0 & 0 & \frac{1}{8} & 0 & 0 & \frac{1}{8} & \frac{1}{8} \\
-\frac{1}{48} & 0 & \frac{1}{12} & \frac{7}{24} & 0 & 0 & 0 & \frac{1}{8} & 0 & 0 & \frac{1}{8} & \frac{1}{8} \\
0 & -\frac{1}{48} & 0 & 0 & \frac{7}{24} & \frac{1}{12} & \frac{1}{8} & 0 & \frac{1}{8} & \frac{1}{8} & 0 & 0 \\
0 & -\frac{1}{48} & 0 & 0 & \frac{1}{12} & \frac{7}{24} & \frac{1}{8} & 0 & \frac{1}{8} & \frac{1}{8} & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{8} & \frac{1}{8} & \frac{5}{12} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\
0 & 0 & \frac{1}{8} & \frac{1}{8} & 0 & 0 & 0 & \frac{5}{12} & 0 & 0 & \frac{1}{4} & \frac{1}{4} \\
0 & 0 & 0 & 0 & \frac{1}{8} & \frac{1}{8} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{8} & \frac{1}{8} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\
0 & 0 & \frac{1}{8} & \frac{1}{8} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{4} \\
0 & 0 & \frac{1}{8} & \frac{1}{8} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{4}
\end{pmatrix}
$$

which has three zero and nine strictly positive eigenvalues, and hence is positive definite. Hence there exists a non-singular matrix c such that $b = c^T c$. Let C be the $12n \times 12n$ matrix consisting of $n \times n$ diagonal blocks of the form $c_{ij}I_{n \times n}$ for $i,j = 1, \ldots, 12$. Then $B = C^T C$ and is therefore also positive definite; thus establishing the stated result for order 1 integrators. $\square$

We now introduce two new modified global order 1 Magnus integrators.

Definition (Uniformly accurate Magnus integrator). *We define the* uniformly accurate Magnus integrator *by*

$$
\sigma^{\mathrm{umag}} = a_1 J_1 + a_2 J_2 + a_0 J_0 + \tfrac{1}{2}[a_1, a_2](J_{21} - J_{12})
$$
$$
+ \tfrac{h^2}{12}\big([a_1, [a_1, a_0]] + [a_2, [a_2, a_0]] + a_1[a_2, [a_2, a_1]] + a_2[a_1, [a_1, a_2]]\big).
$$

Definition (Alternative accurate Magnus integrator). *We define the* alternative accurate Magnus integrator *by*

$$
\sigma^{\mathrm{amag}} = a_1 J_1 + a_2 J_2 + a_0 J_0 + \tfrac{1}{2}[a_1, a_2](J_{21} - J_{12})
$$
$$
+ \tfrac{h^2}{12}\big(a_1 a_2^2 a_1 + a_2 a_1^2 a_2\big) - \tfrac{h^2}{6}\big(a_1 a_0 a_1 + a_2 a_0 a_2 + a_1 a_2 a_1 a_2 + a_2 a_1 a_2 a_1\big).
$$

Remark. For either modified Magnus integrator, the additional terms are extremely cheap to compute as they are constant. Therefore they do not significantly add to the computational burden.

Theorem 4.2 (Global error comparison). *For a sufficiently small stepsize h, the order $1/2$ Magnus integrator is globally more accurate than the order $1/2$ Neumann integrator. In addition, for the order 1 integrators we have,*

$$
\mathcal{E}^{\mathrm{umag}} \leqslant \mathcal{E}^{\mathrm{amag}} \leqslant \mathcal{E}^{\mathrm{neu}},
$$

i.e. the uniformly accurate Magnus integrator is globally more accurate than the alternative accurate Magnus integrator which in turn is globally more accurate than the order 1 Neumann integrator. In addition, the uniformly accurate Magnus integrator is globally more accurate than the order 1 (unmodified) Magnus integrator.

Proof. To compare the global accuracy of the Neumann and Magnus integrators, we need to consider further contributions to the global truncation error at leading order from two possible sources, from terms in:

1. ρ_M with non-zero expectation, a half order higher than the leading terms;

2. R_M^*—the next order term in the Magnus remainder (4.2).

For the Magnus integrator of order $1/2$, the terms that might contribute to the leading order global truncation error from either source have zero expectation, and so contribute to the global truncation error at higher order. Hence we can deduce directly from the corresponding local error result, that the order $1/2$ Magnus integrator is globally more accurate than the corresponding Neumann integrator.

For the order 1 Magnus integrator there are additional terms from both sources that need to be considered. The terms in ρ_1 that need to be included are

$$\tfrac{1}{12}\big([a_1,[a_1,a_0]] + [a_2,[a_2,a_0]]\big)h^2 \,. \tag{4.3}$$

The terms in R_1^* that need to be included are

$$\tfrac{1}{12}\big(a_1[a_2,[a_2,a_1]] + a_2[a_1,[a_1,a_2]]\big)h^2 \,. \tag{4.4}$$

In both cases we replaced the terms by their expectations since they only contribute to the global error at leading order through the off-diagonal sum. Combining (4.3) and (4.4), the additional terms in the Magnus remainder we need to consider are the deterministic terms

$$\mathfrak{R}^{\mathrm{mag}} = \tfrac{1}{12}\big([a_1,[a_1,a_0]] + [a_2,[a_2,a_0]] + a_1[a_2,[a_2,a_1]] + a_2[a_1,[a_1,a_2]]\big)h^2 \,.$$

Further the additional terms in the Neumann remainder we need to consider are

$$\mathfrak{R}^{\mathrm{neu}} = \big(\tfrac{1}{2}a_0^2 + \tfrac{1}{4}(a_1^2 a_0 + a_0 a_1^2 + a_2^2 a_0 + a_0 a_2^2) + \tfrac{1}{8}(a_1^4 + a_2^2 a_1^2 + a_1^2 a_2^2 + a_2^4)\big)h^2 \,.$$

If we set

$$A = a_1^2 a_0 + a_0 a_1^2 + a_2^2 a_0 + a_0 a_2^2 \,,$$
$$B = 2(a_1 a_0 a_1 + a_2 a_0 a_2 + a_1 a_2 a_1 a_2 + a_2 a_1 a_2 a_1) - a_1 a_2^2 a_1 - a_2 a_1^2 a_2 \,,$$
$$C = a_1^2 a_2^2 + a_2^2 a_1^2 \,,$$
$$D = a_1^4 + a_2^4 \,,$$

then

$$\mathfrak{R}^{\mathrm{mag}} = \tfrac{1}{12}(A - B + C)h^2 \quad \text{and} \quad \mathfrak{R}^{\mathrm{neu}} = \big(\tfrac{1}{2}a_0^2 + \tfrac{1}{4}A + \tfrac{1}{8}C + \tfrac{1}{8}D\big)h^2 \,.$$

The difference

$$
\begin{aligned}
(\mathfrak{R}^{\mathrm{neu}})^T & \mathfrak{R}^{\mathrm{neu}} - (\mathfrak{R}^{\mathrm{mag}})^T \mathfrak{R}^{\mathrm{mag}} \\
&= \tfrac{1}{4}(a_0^2)^T a_0^2\, h^4 + \tfrac{1}{16}\big((a_0^2)^T (2A + C + D) + (2A + C + D)^T a_0^2\big)h^4 \\
&\quad + \tfrac{1}{18}A^T A\, h^4 + \tfrac{1}{144}\big(A^T B + B^T A\big)h^4 + \tfrac{7}{288}\big(A^T C + C^T A\big)h^4 \\
&\quad - \tfrac{1}{144}B^T B\, h^4 + \tfrac{1}{144}\big(B^T C + C^T B\big)h^4 + \tfrac{5}{576}C^T C\, h^4 \\
&\quad + \tfrac{1}{32}\big(A^T D + D^T A\big)h^4 + \tfrac{1}{64}\big(C^T D + D^T C\big)h^4 + \tfrac{1}{64}D^T D\, h^4 \,,
\end{aligned}
$$

is not in general positive definite (note we can consider this difference in isolation because all other terms in the remainders have zero expectation). However if we include the terms $\mathfrak{R}^{\mathrm{mag}}$ in our order 1 Magnus expansion—as we do in the *uniformly accurate Magnus integrator*—the terms $\mathfrak{R}^{\mathrm{mag}}$ are consequently not present in the integrator remainder. This establishes that the uniformly accurate Magnus integrator is more accurate than the order 1 Neumann expansion as well as the order 1 (unmodified) Magnus integrator.

We can refine this result further. The term involving $B^T B$ is the only negative definite term. Including this term in the Magnus integrator—as we have done in the *alternative accurate Magnus integrator*—establishes its superiority over the Neumann integrator.

Now comparing the terms left in the remainders of the uniformly accurate Magnus integrator and alternative accurate Magnus integrator, we get for a sufficiently small stepsize h,

$$(\mathcal{E}^{\mathrm{amag}})^2 = (\mathcal{E}^{\mathrm{umag}})^2 + \tfrac{1}{144} \sup_{\|Y_0\|_2=1} \big\|(A+C)Y_0\big\|_2^2 \, h^4 \,,$$

establishing the first inequality stated in the theorem. □

Remark. The alternative accurate Magnus integrator contains the minimum set of additional terms to ensure its superior accuracy over the order 1 Neumann integrator. Note that the corresponding terms we might think of including in the Neumann expansion have zero expectation and so contribute to the global truncation error at a higher order. We could however, improve the order 1 Neumann integrator to include all the terms in $\mathfrak{R}^{\mathrm{neu}}$. But then the uniformly accurate Magnus integrator would still be globally more accurate than such a modified order 1 Neumann integrator. This is because the terms with zero expectation in the remainders R_1^{mag} and R_1^{neu} will be the only terms contributing to the global errors, and the local error comparison result in Theorem 4.1 guarantees the corresponding global result.

5. Quadrature

We start by emphasizing two inherent scales.

1. Quadrature scale Δt—the *smallest scale* on which the discrete Wiener paths $W_1(t)$ and $W_2(t)$ are generated.

2. Time-step scale h—on which the stochastic differential equation is stepped forward.

To evaluate the numerical solution to our stochastic differential equation over the interval $[t_n, t_{n+1}]$ we need to approximate Stratonovich integrals such as J_{12}. The main idea is to approximate these integrals by their corresponding expectations conditioned on the filtration representing intervening knowledge of the Wiener paths (Clark & Cameron 1980; Gaines & Lyons 1997)

$$\mathcal{F}_Q = \{\Delta W_i(t_n + q\,\Delta t)\colon i = 1, 2;\ q = 0, \ldots, Q-1;\ n = 0, \ldots, N-1\}\,,$$

where

$$\Delta W_i(t_n + q\,\Delta t) \equiv W_i(t_n + (q+1)\,\Delta t) - W_i(t_n + q\,\Delta t)\,,$$

and $Q\Delta t \equiv h$, i.e. Q is the number of discrete Wiener path intervals in each time-step interval. We can think of having knowledge of the paths $W_1(t)$ and $W_2(t)$ in the interval $[0, T] = \cup_{n=0}^{N-1}[t_n, t_{n+1}]$ at the points $t_n + q\Delta t$, $q = 0, \ldots, Q - 1$.

For example the conditional expectation of J_{12} is (we set $\tau_q \equiv t_n + q\Delta t$):

$$\hat{J}_{12}(t_n, t_{n+1}) = \tfrac{1}{2} \sum_{q=0}^{Q-1} \Big(\big(W_1(\tau_{q+1}) - W_1(\tau_0)\big) + \big(W_1(\tau_q) - W_1(\tau_0)\big) \Big) \Delta W_2(\tau_q).$$

which can be found in Gaines & Lyons (1997). More generally for higher Stratonovich integrals such as J_{112}, J_{120} etc. used in our analysis, the conditional expectations and the corresponding local quadrature errors are derived by approximating the integrals with their discretizations and using the following formula for the conditional distribution of $W(s) - W(u)$ given $W(t) - W(t_0)$ for $t_0 \le u \le s \le t$ (see Arnold 1974)

$$W(s) - W(u)\big|W(t) - W(t_0) \sim \mathrm{N}\left(\frac{s - u}{t - t_0}\big(W(t) - W(t_0)\big), s - u - \frac{(s-u)^2}{t - t_0} \right).$$

These conditional expectations are intimately linked to the polygonal area/volume approximations, where the paths $W_1(t)$ and $W_2(t)$ are approximated by piecewise linear interpolations of the filtration set $\mathcal{F}_Q$—see Wong & Zakai (1965), Kloeden & Platen (1999) and Gyöngy & Michaletzky (2004). More precisely, if we substitute the approximations $W_i(\tau) \approx (\Delta W_i(\tau_q)/\Delta t)(\tau - \tau_q) + W_i(\tau_q)$ over each subinterval $\tau_q \le \tau \le \tau_{q+1}$ into the Stratonovich integrals, we obtain polygonal area/volume approximations for them. It turns out that the polygonal approximation for J_{12} is precisely $\hat{J}_{12}$, while the polygonal volume approximation for J_{112} is the same as $\hat{J}_{112}$ up to an additive asymptotically small term $\frac{1}{12}\Delta t\big(W_2(t_{n+1}) - W_2(t_n)\big)$.

We now examine the error in these approximations, in particular for $J_{12}(t_n, t_{n+1})$. Note that $\mathbb{E}\big[J_{12}(\tau_q, \tau_{q+1})\big|\mathcal{F}_Q\big] = \frac{1}{2}\Delta W_1(\tau_q)\Delta W_2(\tau_q)$. With this in mind,

$$\big\|J_{12}(t_n, t_{n+1}) - \hat{J}_{12}(t_n, t_{n+1})\big\|_{L^2}^2 = \sum_{q=0}^{Q-1} \mathbb{E}\left(\mathrm{Var}\big[J_{12}(\tau_q, \tau_{q+1})\big|\mathcal{F}_Q\big] \right) = \mathcal{O}\big(h^2/Q\big),$$

where in the last step we used that $Q\Delta t = h$. Clark & Cameron (1980) prove this is the maximum rate of convergence. Analogous estimates derived for the local quadrature errors associated with our approximations $\hat{J}_{112}$, $\hat{J}_{120}$ and $\hat{J}_{1112}$ include terms which involve the conditional error associated with J_{12}, and these are the terms that give the leading order estimates shown in Table 1. Karhunen–Loève approximations have analogous rates of convergence but do incur slightly better multiplicative order one constants (see Kloeden & Platen 1999, page 367). The convergence for $\hat{J}_{10}$ concurs with Hofmann & Müller-Gronbach (2004).

6. Global error vs computational effort

The global error for an integrator of order M on $[0, T]$ with $Nh = T$ is

$$\mathcal{E} = K_T(M)\, h^M + K_Q(M)\, h^{\frac{1}{2}}/\sqrt{Q}, \tag{6.1}$$

Table 1. *Estimates for local quadrature errors and the computational effort associated with our conditional expectation approximations. The top row indicates how the conditional errors scale, while the bottom three rows indicate the number of quadrature points Q required for these approximations to be sufficiently accurate representations to the correct order (locally and therefore globally at a half order less).*

Quadrature		$\hat{J}_{12}$	$\hat{J}_{112}$	$\hat{J}_{10}$	$\hat{J}_{120}$	$\hat{J}_{1112}$
local error		$h/\sqrt{Q}$	$h^{3/2}/\sqrt{Q}$	$h^{3/2}/\sqrt{Q}$	$h^2/\sqrt{Q}$	$h^2/\sqrt{Q}$
$\mathcal{U}$	$\mathcal{O}(h^{3/2})$	h^{-1}	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	$\mathcal{O}(h^2)$	h^{-2}	h^{-1}	h^{-1}	$\cdots$	$\cdots$
	$\mathcal{O}(h^{5/2})$	h^{-3}	h^{-2}	h^{-2}	h^{-1}	h^{-1}

where

$$K_T(M) = \text{coefficient in the truncation error},$$

$$K_Q(M) = \text{coefficient in the quadrature error for } \hat{J}_{12}.$$

Note that the global quadrature error can be deduced by combining the arguments for the local L^2 quadrature error in §5 with those in §3 for computing the global truncation errors. Also we suppose we have used Q points in each subinterval $[t_n, t_{n+1}]$ to construct $\hat{J}_{12}$. For numerical methods of order $3/2$ or higher, we should also include the quadrature error associated with approximating higher order multidimensional stochastic integrals such as $\hat{J}_{112}$. In practice we will construct $\hat{J}_{112}$ by only using the minimum number of quadrature points to obtain the correct order for the global quadrature error. For example the quadrature error associated with $\hat{J}_{112}$, using Table 1 would be $\mathcal{O}(h/\sqrt{q})$. If we choose $q = h\,Q$ we ensure this error is of the same order as that for $\hat{J}_{12}$. To account for the additional terms representing the quadrature error associated with approximating the higher order stochastic integrals we, with a slight abuse of notation, subsume them into the $\hat{J}_{12}$ error term, and to emphasize this we write $K_Q(M)$—the coefficient depending on the order M of the method.

The global computational effort (measured in flops) associated with generating a Magnus approximation of global order M is for small h,

$$\mathcal{U} = \big(4Q + c_M\,n^2 + c_E + \mathcal{O}(h\,Q)\big)N, \qquad (6.2)$$

which we have constructed by simply adding up the local effort contributions over the N subintervals of $[0, T]$ of length h. The quantities shown are the local computational efforts (in flops) associated as follows:

$$4Q = \text{computing the integral approximation } \hat{J}_{12};$$

$$c_M\,n^2 = \text{evaluating the truncated Magnus series};$$

$$c_E = \text{computing an } n \times n \text{ matrix exponential}.$$

Here note that we use n to denote the size of the system and the flop counts $c_M\,n^2 + c_E$ for each Magnus integrator of order M can be found in Table 2. The

correction term $\mathcal{O}(h\,Q)$ represents the effort associated with any higher order multiple stochastic integrals such as $\hat{J}_{112}$.

Theorem 6.1 (Global effort vs error). *The global computational effort $\mathcal{U}$ measured in flops is related to the global error $\mathcal{E}$ for sufficiently small stepsize h by*

$$\mathcal{U} = \underbrace{\left((c_M\,n^2 + c_E)TK_{\mathcal{E}}^{1/M}(M)\right)\mathcal{E}^{-1/M}}_{\mathcal{U}^{\mathrm{eval}}} + \underbrace{\left(4\,T^2K_{\mathcal{E}}^2(M)\right)\mathcal{E}^{-2}}_{\mathcal{U}^{\mathrm{quad}}}, \qquad (6.3)$$

where $\mathcal{U}^{\mathrm{eval}}$ and $\mathcal{U}^{\mathrm{quad}}$ are the computational efforts associated with evaluation and quadrature, and $K_{\mathcal{E}}(M) \equiv K_Q(M) + K_T(M)$.

Proof. To guarantee a numerical method of global order M in (6.1) we must choose

$$Q = h^{1-2M}. \qquad (6.4)$$

This ensures $\hat{J}_{12}$ is a sufficiently accurate representation of J_{12}. With this choice for Q, we have for a sufficiently small stepsize h,

$$\mathcal{E} = K_{\mathcal{E}}(M)\,h^M. \qquad (6.5)$$

Combining the results (6.4), (6.5) and that $N = T/h$ with our expression (6.2) for the global computational effort establishes the theorem. $\square$

There are two asymptotic regimes of interest. Firstly, when the evaluation effort dominates the overall computational effort so that we can ignore the quadrature effort. And secondly, when quadrature effort dominates the overall computational effort so that we can ignore the evaluation effort. Ignoring each of the effects mentioned in turn in (6.3) establishes the following results.

Corollary 6.2 (Standard scaling). *If $\mathcal{U}^{\mathrm{eval}} \gg \mathcal{U}^{\mathrm{quad}}$ then for a sufficiently small stepsize h:*

$$\mathcal{E} \sim \left((c_M\,n^2 + c_E)T\right)^M K_{\mathcal{E}}(M)\mathcal{U}^{-M}.$$

In other words the log-log plot of global error vs effort has a negative slope given by the order M of the method.

Corollary 6.3 (Square-root scaling). *If $\mathcal{U}^{\mathrm{quad}} \gg \mathcal{U}^{\mathrm{eval}}$ then for a sufficiently small stepsize h:*

$$\mathcal{E} \sim 2\,TK_{\mathcal{E}}(M)\mathcal{U}^{-1/2}.$$

In a log-log plot of global error vs effort all methods have the same slope of $-1/2$.

For a given fixed large computational effort, which order Magnus method is most accurate? Note that for the order $1/2$ method there is no quadrature effort.

Corollary 6.4 (Global error vs effort). *For the order $M = 1/2$, 1 and $3/2$ Magnus integrators the global errors $\mathcal{E}_M$ are given in terms of the computational effort $\mathcal{U} \gg 1$ by:*

$$\mathcal{E}_{1/2} = \left((c_{1/2}\,n^2 + c_E)T\right)^{1/2} K_{\mathcal{E}}(1/2)\mathcal{U}^{-1/2};$$
$$\mathcal{E}_1 \sim 2\,TK_{\mathcal{E}}(1)\mathcal{U}^{-1/2};$$
$$\mathcal{E}_{3/2} \sim 2\,TK_{\mathcal{E}}(3/2)\mathcal{U}^{-1/2}.$$

Table 2. *Floating point operation counts per step for different order schemes applied to linear constant coefficient stochastic differential equations driven by two Wiener processes. The Magnus column includes the $6n^3$ flops needed to compute the matrix exponential.*

	For each path		Independent of the path	
Order	Neumann	Magnus	Neumann	Magnus
$\frac{1}{2}$	$9n^2$	$5n^2 + 6n^3$	$4n^3$	n/a
1	$13n^2$	$7n^2 + 6n^3$	$8n^3$	$4n^3$
$1\frac{1}{2}$	$56n^2$	$19n^2 + 6n^3$	$50n^3$	$28n^3$

Further, taking differences of the logarithms of these expressions we have

$$\log \mathcal{E}_{1/2} - \log \mathcal{E}_1 \sim \tfrac{1}{2}\log(c_{1/2}\, n^2 + c_E) - \tfrac{1}{2}\log 4\, T + \log K_{\mathcal{E}}(1/2) - \log K_{\mathcal{E}}(1)\,, \tag{6.6}$$

$$\log \mathcal{E}_1 - \log \mathcal{E}_{3/2} \sim \log K_{\mathcal{E}}(1) - \log K_{\mathcal{E}}(3/2)\,. \tag{6.7}$$

In other words the Magnus methods of order 1 and 3/2 converge for $\mathcal{U} \gg 1$ to within the fixed gap given by (6.7).

7. Numerical simulations

(a) Linear system

Our first numerical simulation is for a homogeneous and autonomous linear problem involving two Wiener processes with coefficient matrices

$$a_0 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{pmatrix}, \qquad a_1 = \begin{pmatrix} 0 & 1 \\ -\frac{1}{2} & -\frac{51}{200} \end{pmatrix} \qquad \text{and} \qquad a_2 = \begin{pmatrix} 1 & 1 \\ 1 & \frac{1}{2} \end{pmatrix}, \tag{7.1}$$

and initial data $(\frac{1}{2}\ 1)^T$ (we found Higham 2001 a very useful starting point for our Matlab simulations). In Figure 1 we show how the error scales with CPU clocktime. We see that the superior accuracy of the Magnus integrators is achieved for the same computational cost. Note that $T = 1$ and $n = 2$. In addition $c_E \approx 6n^3 = 48$ flops—we used a $(6,6)$ Padé approximation with scaling to compute the matrix exponential—see Moler & Van Loan (2003) and also Iserles & Zanna (2002). Then assuming $K_{\mathcal{E}}(1/2)$, $k_{\mathcal{U}}$ and $K_{\mathcal{E}}(1)$ are all strictly order 1, and using that $c_{1/2} = 5$ from Table 2, we see from (6.6) that $\log \mathcal{E}_{1/2} - \log \mathcal{E}_1 \approx \frac{1}{2}\log(20 + c_E) \approx 0.9$, which is in good agreement with the difference shown in Figure 1. We can also see in Figure 1 that there is not too much to choose between the order 1 and 3/2 integrators, they are separated by a fixed small gap theoretically predicted by (6.7). Note that in Figure 1 the uniformly accurate Magnus integrator and order 1 (unmodified) Magnus integrator yield virtually identical performances. A separate check of the contribution of the terms in $\mathfrak{R}^{\mathrm{mag}}$ to the global error for our example corroborates this observation (for small h).

(b) Riccati system

Our second application is for stochastic Riccati differential systems—some classes of which can be linearized (see Freiling 2004 and Schiff & Shnider 1999). Such
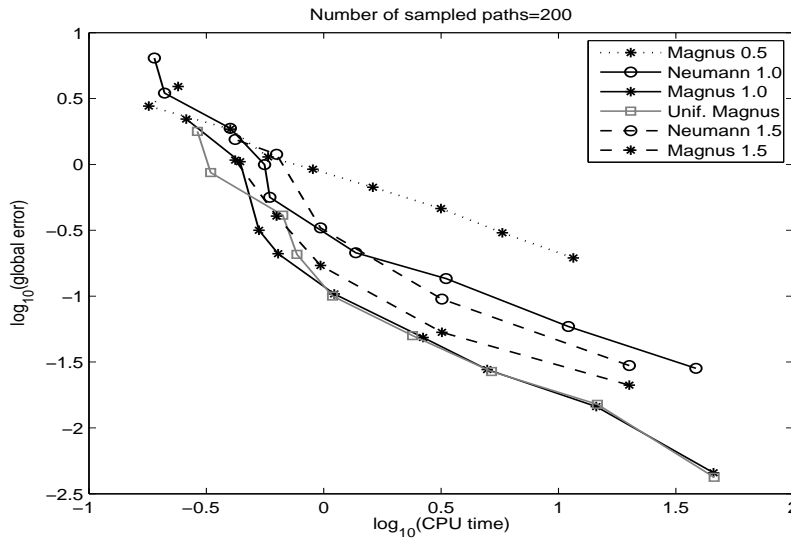
Figure 1. Global error vs CPU clocktime for the model problem at time $t = 1$ with two Wiener processes. The error corresponding to the largest step size takes the shortest time to compute.

systems arise in stochastic linear-quadratic optimal control problems, for example, mean-variance hedging in finance (see Bobrovnytska & Schweizer 2004 and Kohlmann & Tang 2003)—though often these are backward problems (which we intend to investigate in a separate study). Consider for example Riccati equations of the form

$$S(t) = I + \sum_{i=0}^{d} \int_0^t \big(S(\tau)A_i(\tau)S(\tau) + B_i(\tau)S(\tau) + S(\tau)C_i(\tau) + D_i(\tau)\big)\,\mathrm{d}W_i(\tau)\,.$$

If $\mathbb{A}_i(t) \equiv \begin{pmatrix} B_i(t) & D_i(t) \\ -A_i(t) & -C_i(t) \end{pmatrix}$ and $\mathbb{U} = (U\ V)^T$ satisfies the linear system

$$\mathbb{U}(t) = \mathbb{I} + \sum_{i=0}^{d} \int_0^t \mathbb{A}_i(\tau)\mathbb{U}(\tau)\,\mathrm{d}W_i(\tau)\,,$$

then $S = UV^{-1}$ solves the Riccati equation above—note that $\mathbb{I} \equiv (I\ I)^T$.

We consider here a Riccati problem with two additive Wiener processes, W_1 and W_2, and coefficient matrices

$$A_0 = \begin{pmatrix} -1 & 1 \\ -\frac{1}{2} & -1 \end{pmatrix}, \qquad C_0 = \begin{pmatrix} -\frac{1}{2} & 0 \\ -1 & -1 \end{pmatrix} \qquad \text{and} \qquad D_0 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{pmatrix},$$

and we take $D_1 = a_1$ and $D_2 = a_2$, where a_1 and a_2 are given in (7.1). All other coefficient matrices are zero. The initial data is the identity matrix (for S and therefore U and V also).

Note that for this example the coefficient matrices $\mathbb{A}_1$ and $\mathbb{A}_2$ are upper right block triangular and therefore nilpotent of degree 2, and also that $\mathbb{A}_1\mathbb{A}_2$ and $\mathbb{A}_2\mathbb{A}_1$ are identically zero. This means that S_1 and s_1 are identically zero (see Appendix A) and the order 3/2 Neumann and Magnus integrators collapse to the simpler forms:

$$S(t_n,t_{n+1}) = I + \mathbb{A}_1 J_1 + \mathbb{A}_2 J_2 + \mathbb{A}_0 J_0 + \mathbb{A}_0\mathbb{A}_1 J_{10} + \mathbb{A}_1\mathbb{A}_0 J_{01}$$
$$+ \mathbb{A}_0\mathbb{A}_2 J_{20} + \mathbb{A}_2\mathbb{A}_0 J_{02} + \tfrac{1}{2}(\mathbb{A}_0)^2 h^2 \,,$$

$$\sigma(t_n,t_{n+1}) = \mathbb{A}_1 J_1 + \mathbb{A}_2 J_2 + \mathbb{A}_0 J_0 + \tfrac{1}{2}[\mathbb{A}_0,\mathbb{A}_1](J_{10}-J_{01})$$
$$+ \tfrac{1}{2}[\mathbb{A}_0,\mathbb{A}_2](J_{20}-J_{02}) - \tfrac{1}{6}(\mathbb{A}_1\mathbb{A}_0\mathbb{A}_1 + \mathbb{A}_2\mathbb{A}_0\mathbb{A}_2)h^2 \,.$$

For either integrator, if we include only $\mathbb{A}_1 J_1 + \mathbb{A}_2 J_2 + \mathbb{A}_0 J_0$ we obtain order 1 integrators. The number of terms in each order 3/2 integrator is roughly equal, and so for a given stepsize the Magnus integrator should be more expensive to compute due to the cost of computing the 4×4 matrix exponential. Also the order 1 integrators do not involve quadrature effort whilst the order 3/2 integrators involve the quadrature effort associated with approximating J_{10}—see Table 1. Using arguments analogous to those at the end of §6 we can deduce the following expressions for the global errors and efforts for small stepsize h: $\mathcal{E}_1 \sim K_{\mathcal{E}}(1)\,h$, $\mathcal{U}_1 \sim (\tilde{c}_1 n^2 + \tilde{c}_E)\,T\,h^{-1}$ and $\mathcal{E}_{3/2} \sim K_{\mathcal{E}}(3/2)\,h^{3/2}$, $\mathcal{U}_{3/2} \sim (8\,h^{-1} + \tilde{c}_{3/2}n^2 + \tilde{c}_E)\,T\,h^{-1}$. Consequently we see that we expect the slope of a log-log plot of global error vs computational effort to be -1 for the order 1 integrators, and ignoring the evaluation effort for the order 3/2 integrators, we expect to see a slope of $-3/4$. Hence for very small stepsize h the order 3/2 integrators globally perform worse than the order 1 integrators for the same effort. Further, the slope gets progressively smaller in magnitude for higher order methods.

For comparison, we use a nonlinear Runge–Kutta type order 3/2 scheme for the case of two additive noise terms (from page 383 of Kloeden & Platen) applied directly to the original Riccati equation:

$$S(t_n,t_{n+1}) = S(t_n) + f\big(S(t_n)\big)h + D_1 J_1 + D_2 J_2$$
$$+ \tfrac{h}{4}\big(f(Y_1^+) + f(Y_1^-) + f(Y_2^+) + f(Y_2^-) - 4f\big(S(t_n)\big)\big)$$
$$+ \tfrac{1}{2\sqrt{h}}\big(\big(f(Y_1^+) - f(Y_1^-)\big)J_{10} + \big(f(Y_2^+) - f(Y_2^-)\big)J_{20}\big)\,, \qquad (7.2)$$

where $Y_j^{\pm} = S(t_n) + \tfrac{h}{2}f\big(S(t_n)\big) \pm D_j\sqrt{h}$ and $f(S) = SA_0S + B_0S + SC_0 + D_0$.

In Figure 2 we show the global error vs 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 (7.2), we must evaluate $f(S)$ five times per step per path costing $20n^3 + 54n^2$ flops—here and subsequently $n = 2$ refers to the size of the original system. The Neumann and Magnus integrators the evaluation costs are $16(2n \times n) = 32n^2$ and $6(2n)^3 + 11(2n)^2 = 48n^3 + 44n^2$ flops, respectively (directly counting from the schemes outlined above). Hence for the same relatively 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 and the efforts of all the order 3/2 integrators are not much different. The Magnus integrator then outperforms the other two due to its superior accuracy.
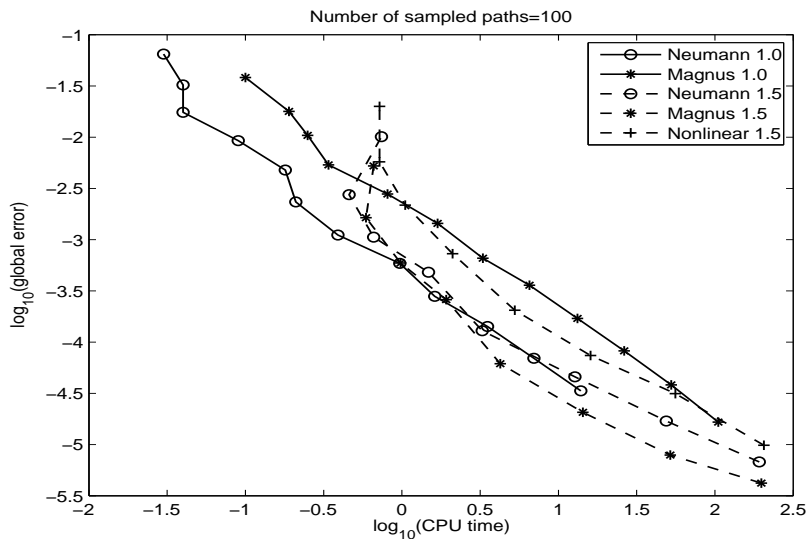
Figure 2. Global error vs CPU clocktime for the Riccati problem at time $t = 1$.

8. Concluding remarks

Our results suggest the uniformly accurate Magnus integrator is the optimal method in dynamic programming or filtering applications such as any linear feedback control system (or some neural or mechanical systems in nature). An important class of schemes we have not mentioned thusfar are the *asymptotically efficient* Runge–Kutta schemes derived by Newton (1991). Such schemes have the optimal minimum leading error coefficient among all schemes which are $\mathcal{F}_Q$-measurable. Castell & Gaines (1995) state that the order $1/2$ Magnus integrator is asymptotically efficient, and in the case of one Wiener process (take a_2 to be the zero matrix) the order 1 uniformly accurate Magnus integrator we present in §4 is asymptotically efficient. In the case of two or more Wiener processes, we expect our order 1 uniformly accurate Magnus integrator to be a prime candidate for the corresponding asymptotically efficient scheme.

Lastly, some extensions of our work that we intend to investigate further are: (1) implementing a variable step scheme following Gaines & Lyons (1997), using analytic expressions for the local truncation errors (see Aparicio *et al.* 2004); (2) to consider the Lie-group structure preserving properties of Magnus methods in the stochastic setting (though see Castell & Gaines 1995; Iserles *et al.* 2000; Kunita 1980; Burrage *et al.* 2004; Misawa 2001; and also Milstein *et al.* 2002 for a possible symplectic application); (3) applications to nonlinear stochastic differential equations (see Ben Arous 1989 and Castell & Gaines 1995, and Casas & Iserles 2005 in the deterministic case) and (4) pricing path-dependent options.

Appendix A.

We present Neumann and Magnus integrators up to global order 2 in the case of two Wiener processes $W_1(t)$ and $W_2(t)$, and with constant coefficient matrices a_i, $i = 0, 1, 2$. The Neumann expansion (1.4) for the solution over an interval $[t_n, t_{n+1}]$, where $t_n = nh$, is

$$S^{\text{neu}}(t_n, t_{n+1}) \approx I + S_{1/2} + S_1 + S_{3/2} + S_2\,, \tag{A 1}$$

where

$$
\begin{aligned}
S_{1/2} &= a_1 J_1 + a_2 J_2 + a_0 J_0 + a_1^2 J_{11} + a_2^2 J_{22}\,, \\
S_1 &= a_2 a_1 J_{12} + a_1 a_2 J_{21}\,, \\
S_{3/2} &= a_0^2 J_{00} + a_0 a_1 J_{10} + a_1 a_0 J_{01} + a_0 a_2 J_{20} + a_2 a_0 J_{02} \\
&\quad + a_1^3 J_{111} + a_2 a_1^2 J_{112} + a_1 a_2 a_1 J_{121} + a_1^2 a_2 J_{211} \\
&\quad + a_2^2 a_1 J_{122} + a_2 a_1 a_2 J_{212} + a_1 a_2^2 J_{221} + a_2^3 J_{222} \\
&\quad + a_1^2 a_0 J_{011} + a_0 a_1^2 J_{110} + a_2^2 a_0 J_{022} + a_0 a_2^2 J_{220} \\
&\quad + a_1^4 J_{1111} + a_2^2 a_1^2 J_{1122} + a_1^2 a_2^2 J_{2211} + a_2^4 J_{2222}\,, \\
S_2 &= a_1 a_0 a_1 J_{101} + a_2 a_0 a_2 J_{202} + a_0 a_1 a_2 J_{210} + a_0 a_2 a_1 J_{120} \\
&\quad + a_1 a_0 a_2 J_{201} + a_1 a_2 a_0 J_{021} + a_2 a_0 a_1 J_{102} + a_2 a_1 a_0 J_{012} \\
&\quad + a_2 a_1^3 J_{1112} + a_1^2 a_2 a_1 J_{1121} + a_1^2 a_2 a_1 J_{1211} + a_2 a_1^3 J_{2111} \\
&\quad + a_1 a_2^3 J_{2221} + a_2 a_1 a_2^2 J_{2212} + a_2^2 a_1 a_2 J_{2122} + a_2^3 a_1 J_{1222} \\
&\quad + a_2 a_1 a_2 a_1 J_{1212} + a_1 a_2 a_1 a_2 J_{2121} + a_1 a_2^2 a_1 J_{1221} + a_2 a_1^2 a_2 J_{2112}\,.
\end{aligned}
$$

The corresponding Magnus expansion with

$$S^{\text{mag}}(t_n, t_{n+1}) = \exp\big(\sigma(t_n, t_{n+1})\big)\,. \tag{A 2}$$

is

$$\sigma(t_n, t_{n+1}) \approx s_{1/2} + s_1 + s_{3/2} + s_2\,, \tag{A 3}$$

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

$$
\begin{aligned}
s_{1/2} &= a_1 J_1 + a_2 J_2 + a_0 J_0\,, \\
s_1 &= \tfrac{1}{2}[a_1, a_2](J_{21} - J_{12})\,, \\
s_{3/2} &= \tfrac{1}{2}[a_0, a_1](J_{10} - J_{01}) + \tfrac{1}{2}[a_0, a_2](J_{20} - J_{02}) \\
&\quad + [a_1, [a_1, a_2]]\left(J_{112} - \tfrac{1}{2}J_1 J_{12} + \tfrac{1}{12}J_1^2 J_2\right) \\
&\quad + [a_2, [a_2, a_1]]\left(J_{221} - \tfrac{1}{2}J_2 J_{21} + \tfrac{1}{12}J_2^2 J_1\right) \\
&\quad + [a_1, [a_1, a_0]]\left(J_{110} - \tfrac{1}{2}J_1 J_{10} + \tfrac{1}{12}J_1^2 J_0\right) \\
&\quad + [a_2, [a_2, a_0]]\left(J_{220} - \tfrac{1}{2}J_2 J_{20} + \tfrac{1}{12}J_2^2 J_0\right)\,, \\
s_2 &= + [a_2, [a_1, a_0]]\left(J_{120} + \tfrac{1}{2}J_1 J_{02} + \tfrac{1}{2}J_0 J_{21} - \tfrac{2}{3}J_0 J_1 J_2\right) \\
&\quad + [a_1, [a_2, a_0]]\left(J_{210} + \tfrac{1}{2}J_2 J_{01} + \tfrac{1}{2}J_0 J_{12} - \tfrac{2}{3}J_0 J_1 J_2\right) \\
&\quad - [a_1, [a_1, [a_1, a_2]]]\left(J_{1112} - \tfrac{1}{2}J_1 J_{112} + \tfrac{1}{12}J_1^2 J_{12}\right) \\
&\quad - [a_2, [a_2, [a_2, a_1]]]\left(J_{2221} - \tfrac{1}{2}J_2 J_{221} + \tfrac{1}{12}J_2^2 J_{21}\right) \\
&\quad + [a_1, [a_2, [a_1, a_2]]]\left(\tfrac{1}{24}J_1^2 J_2^2 - \tfrac{1}{2}J_2 J_{112} + \tfrac{1}{6}J_1 J_2 J_{21} - \tfrac{1}{2}J_1 J_{221} + J_{1122}\right)\,.
\end{aligned}
$$

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. The Magnus expansion, up to and including the term $s_{3/2}$, can be found in Burrage (1999)—using the Jacobi identity we have one less term. Note that we have explicitly used the relationships between multi-dimensional stochastic integrals generated by partial integration (see Gaines 1995 and Kloeden & Platen 1999). This is well known, indeed Gaines 1995 and Kawski 2001 consider the shuffle algebra associated with these relations. Using these relations in the Neumann expansion does not significantly reduce the number of terms. However it is clear that all the higher order multi-dimensional integrals can be directly expressed in terms of only a few specific integrals of that order and so the Magnus or Neumann approximations of order 2 shown above can both be expressed in the computationally favourable basis

$$\{J_0, J_1, J_2, J_{12}, J_{01}, J_{02}, J_{112}, J_{221}, J_{110}, J_{220}, J_{120}, J_{210}, J_{1112}, J_{2221}, J_{1122}\}\ .$$

Though there are many variants of this basis we could use, the basis we have chosen reveals explicitly that we should expect to be able to approximate all the higher order terms by single sums (including J_{1122}).

As an example, to construct a global order 3/2 Magnus integrator, the local remainder is $R_{3/2}^{\mathrm{mag}} = \rho_{3/2} + \cdots = s_2 + \cdots$. All the terms in s_2 have L^2-norm of order h^2. They all have zero expectation as well and so only contribute to the global error through the diagonal sum in (3.4) generating terms of order $h^{3/2}$, consistent with the order of the integrator. Note that in $s_{3/2}$ we included the two terms (the last two) with L^2-norm of order h^2 because they have non-zero expectation though, as explained at the end of §3, we can replace them by their expectations. Similar arguments explain the form of the order 3/2 Neumann integrator and how analogous terms can be replaced by their expectations. Lastly, note that in the case of one Wiener process, $\exp(a_0 t + a_1 J_1)$ generates a global order 1 Magnus scheme—see Castell & Gaines (1995).

Suppose that instead of the linear Stratonovich stochastic differential equation (1.2) we have $S = I + \mathsf{K} \circ S + F$, where $F(t) \equiv \int_0^t A_f(\tau)\,\mathrm{d}W_f(\tau)$ is a non-homogeneous term with a Wiener process W_f independent of W_i, for $i = 1, \ldots, d$, and $A_f(t)$ is a given $n \times n$ coefficient matrix. The solution S can be decomposed into its homogeneous and particular integral components, $S = S_H + S_P$, where the homogeneous component S_H can be solved as outlined in the introduction and the non-homogeneous component is $S_P(t) = (\mathsf{I} + \mathsf{K} + \mathsf{K}^2 + \mathsf{K}^3 + \cdots) \circ F$ or equivalently $S_P(t) = S_H(t) \int_0^t S_H^{-1}(\tau) A_f(\tau)\,\mathrm{d}W_f(\tau)$.

For the non-autonomous case, where the coefficient matrices $a_i(t)$, $i = 0, 1, 2$ are not constant, we assume they have Taylor series expansions on $[t_n, t_{n+1}]$ of the form $a_i(t_n + h) = \tilde{a}_i(t_n) + b_i(t_n)h + \cdots$. Then we need to modify the order 2 autonomous numerical schemes by replacing the a_i by $\tilde{a}_i$ at each step, and adding the terms

$$\tilde{S}(t_n, t_{n+1}) \approx b_1 J_{01} + b_2 J_{02} + \tfrac{1}{2} b_0 J_{00} + \tilde{a}_1 b_1 J_{011} + \tilde{a}_1 b_2 J_{021} + \tilde{a}_2 b_1 J_{012} + \tilde{a}_2 b_2 J_{022}$$

to the Neumann expansion, or the following terms to the Magnus expansion

$$\begin{aligned}
\tilde{\sigma}(t_n, t_{n+1}) \approx\ & b_1 J_{01} + b_2 J_{02} + \tfrac{1}{2} b_0 J_{00} + \tilde{a}_1 b_1 (J_{110} - \tfrac{1}{2} J_1 J_{10}) - \tfrac{1}{2} b_1 \tilde{a}_1 J_1 J_{01} \\
& + \tilde{a}_2 b_1 (J_{012} - \tfrac{1}{2} J_2 J_{01}) - \tfrac{1}{2} b_1 \tilde{a}_2 J_2 J_{01} + \tilde{a}_1 b_2 (J_{021} - \tfrac{1}{2} J_1 J_{02}) \\
& - \tfrac{1}{2} b_2 \tilde{a}_1 J_1 J_{02} + \tilde{a}_2 b_2 (J_{220} - \tfrac{1}{2} J_2 J_{20}) - \tfrac{1}{2} b_2 \tilde{a}_2 J_2 J_{02}\ .
\end{aligned}$$

References

Aparicio, N.D., Malham, S.J.A. & Oliver, M. 2005 Numerical evaluation of the Evans function by Magnus integration, *BIT*, **45**, 219–258.

Arnold, L. 1974 *Stochastic differential equations*. Wiley.

Ben Arous, G. 1989 Flots et series de Taylor stochastiques. *Probab. Th. Rel. Fields* **81**, 29–77.

Bobrovnytska, O. & Schweizer, M. 2004 Mean-variance hedging and stochastic control: beyond the Brownian setting. *IEEE Trans. on Automatic Control* **49**(3), 396–408.

Burrage, P.M. 1999 Runge–Kutta methods for stochastic differential equations. Ph.D. thesis, University of Queensland.

Burrage, K. & Burrage, P.M. 1999 High strong order methods for non-commutative stochastic ordinary differential equation systems and the Magnus formula. *Physica* D **133**, 34–48.

Burrage, K., Burrage, P.M. & Tian, T. 2004 Numerical methods for strong solutions stochastic differential equations: an overview. *Proc. R. Soc. Lond.* A **460**, 373–402.

Casas, F. & Iserles, A. 2005 Explicit Magnus expansions for nonlinear equations. *Cambridge NA reports*.

Castell, F. 1993 Asymptotic expansion of stochastic flows. *Probability Theory and Related Fields* **96**, 225–239.

Castell, F. & Gaines, J. 1995 An efficient approximation method for stochastic differential equations by means of the exponential Lie series. *Mathematics and computers in simulation* **38**, 13–19.

Clark, J.M.C. & Cameron, R.J. 1980 The maximum rate of convergence of discrete approximations for stochastic differential equations. In *Lecture Notes in Control and Information Sciences* **25**, pp. 162–171.

Freiling, G. 2002 A survey of nonsymmetric Riccati equations. *Linear Algebra and its Applications* **351–352**, 243–270.

Gaines, J.G. 1995 A basis for iterated stochastic integrals. *Mathematics and computers in simulation* **38**, 7–11.

Gaines, J.G. & Lyons, T.J. 1994 Random generation of stochastic area integrals. *SIAM J. Appl. Math.* **54**(4), 1132–1146.

Gaines, J.G. & Lyons, T.J. 1997 Variable step size control in the numerical solution of stochastic differential equations. *SIAM J. Appl. Math.* **57**(5), 1455–1484.

Gihman, I.I. & Skorohod, A.V. 1979 *The theory of stochastic processes III*. Springer.

Gyöngy, I. & Michaletzky, G. 2004 On Wong–Zakai approximations with δ-martingales. *Proc. R. Soc. Lond.* A **460**, 309–324.

Higham, D. J. 2001 An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review* **43**(3), 525–546.

Hofmann, N. & Müller-Gronbach, T. 2004 On the global error of Itô–Taylor schemes for strong approximation of scalar stochastic differential equations. *Journal of Complexity* **20**, 732–752.

Iserles, A., Munthe-Kaas, H.Z., Nørsett, S.P. & Zanna, A. 2000 Lie-group methods. *Acta Numerica*, 215–365.

Iserles, A. & Zanna, A. 2002 Efficient computation of the matrix exponential by generalized polar decompositions. Technical report 2002/NA09, DAMTP, University of Cambridge.

Kawski, M. 2001 The combinatorics of nonlinear controllability and noncommuting flows. *Lectures given at the Summer School on Mathematical Control Theory, Trieste, September 3–28*.

Kloeden, P.E. & Platen, E. 1999 *Numerical solution of stochastic differential equations*. Springer.

Kohlmann, M. & Tang, S. 2003 Multidimensional backward stochastic Riccati equations and applications. *SIAM J. Control Optim.* **41**(6), 1696–1721.

Kunita, H. 1980 On the representation of solutions of stochastic differential equations. *Lecture Notes in Math.* **784**, Springer–Verlag, 282–304.

Lyons, T. 1998 Differential equations driven by rough signals. *Rev. Mat. Iberoamericana* **14**(2), 215–310.

Lyons, T. & Victoir, N. 2004 Cubature on Wiener space. *Proc. R. Soc. Lond.* A **460**, 169–198.

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

Milstein, G.N., Repin, Yu. M. & Tretyakov, M.V. 2002 Numerical methods for stochastic systems preserving symplectic structure. *SIAM J. Numer. Anal.* **40**(4), 1583–1604.

Misawa, T. 2001 A Lie algebraic approach to numerical integration of stochastic differential equations. *SIAM J. Sci. Comput.* **23**(3), 866–890.

Moler, C. & Van Loan, C. 2003 Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later, *SIAM Review* **45**, 3–49.

Newton, N.J. 1991 Asymptotically efficient Runge–Kutta methods for a class of Ito and Stratonovich equations. *SIAM J. Appl. Math.* **51**, 542–567.

Schiff, J. & Shnider, S. 1999 A natural approach to the numerical integration of Riccati differential equations. *SIAM J. Numer. Anal.* **36**(5), 1392–1413.

Schurz, H. 2002 A brief introduction to numerical analysis of (ordinairy) stochastic differential equations without tears. In *Handbook of Stochastic Analysis and Applications* (ed. V. Lakshmikantham & D. Kannan). Marcel Dekker, pp. 237–359.

Sipiläinen, E.-M. 1993 A pathwise view of solutions of stochastic differential equations. Ph.D. thesis, University of Edinburgh.

Stump, D.M. & Hill, J.M. 2005 On an infinite integral arising in the numerical integration of stochastic differential equations. *Proc. R. Soc.* A **461**, 397–413.

Sussmann, H.J. 1988 Product expansions of exponential Lie series and the discretization of stochastic differential equations. In *Stochastic Differential Systems, Stochastic Control Theory, and Applications* (ed. W. Fleming & J. Lions). Springer IMA Series, vol. 10, pp. 563–582.

Talay, D. 1995 Simulation and numerical analysis of stochastic differential systems. In *Probabilistic Methods in Applied Physics* (ed. P. Kree and W. Wedig). Lecture Notes in Physics, vol. 451, chap. 3, pp. 63–106.

Wiktorsson, M. 2001 Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions. *The Annals of Applied Probability* **11**(2), 470–487.

Wong, E. & Zakai, M. 1965 On the relation between ordinary differential and stochastic differential equations. *Int. J. Engng Sci.* **3**, 213–229.