Approximation of SDE solutions using local asymptotic expansions

A. M. Davie¹

This article is dedicated to István Gyöngy on the occasion of his 70th birthday.

Received: 20 September 2021 / Revised: 9 November 2021 / Accepted: 15 November 2021 /

Published online: 10 December 2021

© The Author(s) 2021

Abstract

We develop an asymptotic expansion for small time of the density of the solution of a non-degenerate system of stochastic differential equations with smooth coefficients, and apply this to the stepwise approximation of solutions. The asymptotic expansion, which takes the form of a multivariate Edgeworth-type expansion, is obtained from the Kolmogorov forward equation using some standard PDE results. To generate one step of the approximation to the solution, we use a Cornish–Fisher type expansion derived from the Edgeworth expansion. To interpret the approximation generated in this way as a strong approximation we use couplings between the (normal) random variables used and the Brownian path driving the SDE. These couplings are constructed using techniques from optimal transport and Vaserstein metrics. The type of approximation so obtained may be regarded as intermediate between a conventional strong approximation and a weak approximation. In principle approximations of any order can be obtained, though for higher orders the algebra becomes very heavy. In order $1/2$ the method gives the usual Euler approximation; in order 1 it gives a variant of the Milstein method, but which needs only normal variables to be generated. However the method is somewhat limited by the non-degeneracy requirement.

Keywords SDE · Approximate solution · Coupling · Vaserstein metric

Mathematics Subject Classification Primary 60H10 · Secondary 49Q20

✉ A. M. Davie
a.davie@ed.ac.uk

¹ School of Mathematics, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh EH9 3FD, Scotland

1 Introduction

In this paper we describe an approach to the stepwise approximation of solutions of stochastic differential equations using an asymptotic expansion for each step. Consider an SDE

$$dx(t) = a(t, x(t))dt + B(t, x(t))dW(t) \quad (1)$$

where $x(t) \in \mathbb{R}^q$, $a(t, x) \in \mathbb{R}^2$, $B(t, x)$ is a $q \times d$ matrix and W is a d -dimensional standard Brownian motion, with an initial condition $x(0) = x^{(0)}$. The first step of a stepwise approximation with step h requires the simulation of an approximation to $x(h)$. The standard approach, as described in [9], is based on stochastic Taylor expansions, giving rise to the Euler approximation $x(h) \approx x^{(0)} + ha(0, x^{(0)}) + B(0, x^{(0)})W(h)$ and the Milstein approximation

$$\begin{aligned} x_i(h) \approx & x_i^{(0)} + ha_i(0, x^{(0)}) + \sum_{j=1}^d b_{ij}(0, x^{(0)})W_j(h) \\ & + \sum_{j,k=1}^d \rho_{ijk}(0, x^{(0)}) \int_0^h W_j(t)dW_k(t) \end{aligned} \quad (2)$$

where $B = (b_{ij})$ and $\rho_{ijk} = \sum_{m=1}^q b_{mj} \frac{\partial b_{ik}}{\partial x_m}$, as the lowest-order approximations. The difficulty with this method is that the stochastic integrals involved are hard to generate, already in the case of the double integrals appearing in (2). In [3] an approach to resolving this difficulty is described, under the assumption that $B(0, x^{(0)})$ has rank q , using a perturbation method to generate approximations to the stochastic integrals. This gives an analogue of (2)

$$\begin{aligned} x_i(h) \approx & x_i^{(0)} + ha_i(0, x^{(0)}) + \sum_{j=1}^d b_{ij}(0, x^{(0)})X_j \\ & + \frac{1}{2} \sum_{j,k=1}^d \rho_{ijk}(0, x^{(0)})(X_j X_k - h\delta_{jk}) \end{aligned} \quad (3)$$

where $X_1, \dots, X_d$ are independent $N(0, h)$ random variables.

Here we use a different approach, which involves first deriving a asymptotic expansion to the distribution of $x(h)$. For small h a first approximation to this distribution is the normal distribution $N(x_0, h\Sigma^{(0)})$ where $\Sigma^{(0)} = B(0, x_0)B(0, x_0)^t$, which suggests an Edgeworth-type asymptotic expansion of the form

$$f_h(y) \sim \phi_{\Sigma^{(0)}}(y) \left(1 + \sum_{j=1}^{\infty} h^{j/2} S_j(y) \right) \quad (4)$$

for the density function f_h of $h^{-1/2}(x(h) - x_0)$ where ϕ_Σ is the density of $N(0, \Sigma)$ and the S_j are polynomials on $\mathbb{R}^q$. Provided $B(0, x_0)$ has rank q , so that Σ is nonsingular, and under smoothness conditions on a and B , we derive such an expansion from the Kolmogorov forward equation for the density, and make precise in what sense it is an asymptotic expansion of f_h .

Given the expansion (4), one can then look to approximate $x(h)$ in distribution by a random variable of the form $x_0 + h^{1/2}(X + \sum_{j=1}^k h^{j/2} p_j(X))$ where X is an $\mathbb{R}^q$ -valued random variable with $N(0, \Sigma^{(0)})$ distribution, and the p_j are $\mathbb{R}^q$ -valued polynomials in q variables. It is shown in [4] that any random variable of this form has a density of the form (4), and the relation between the polynomials $p_1, \dots, p_k$ and the sequence (S_j) in (4) is described there in algebraic terms - it is in effect a multivariate version of the relation between Edgeworth and Cornish-Fisher expansions in one variable.

We show that, in principle at least, this programme can be carried out to give approximations to arbitrary order provided the coefficients a and B are smooth and as long as the essential condition that B has rank q continues to hold. The error in the approximation to $x(h)$ is assessed by a suitable coupling between $x(h)$ and the approximation, using methods from the theory of optimal transport.

Sections 2 and 3 give background material on optimal transport and on Edgeworth and Cornish–Fisher-type expansions. Section 4 describes the construction of the expansion (4) and the sense in which it is an effective asymptotic expansion. The application to stepwise approximation schemes is described in Sect. 5 for a single step, and then applied to multistep approximation in Sect. 6, where the main approximation bounds are stated and proved. Section 7 concludes with a comparison of this approximation scheme with standard strong and weak approximation methods, indicating that it is stronger than weak approximation but somewhat weaker than standard strong approximation, and mentions some possible directions for further work.

For general background on numerical approximation of SDE solutions see [9]. In a rather different direction we mention [6] which shows how discrete approximations can be used to prove existence of strong solutions of SDE.

I am very pleased to be able to contribute to the Special Issue in honour of Professor Gyöngy, who has contributed so much to the field of Stochastic Analysis.

2 Optimal transport background

We will use some notation and results from optimal transport theory. If $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are probability measures on $\mathbb{R}^q$, then for $p \geq 1$ the Wasserstein distance $\mathbb{W}_M(\mathbb{P}, \tilde{\mathbb{P}})$ is defined as $\inf(\mathbb{E}|X - \tilde{X}|^M)^{1/M}$ where the inf is over all joint distributions on $\mathbb{R}^q \times \mathbb{R}^q$ for $\mathbb{R}^q$ -valued random variables X and $\tilde{X}$ which have marginals $\mathbb{P}$ and $\tilde{\mathbb{P}}$. We sometimes abuse notation and write this as $\mathbb{W}_M(X, \tilde{X})$ or $\mathbb{W}_M(X, \tilde{\mathbb{P}})$; we will mainly consider the case when $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are absolutely continuous w.r.t. Lebesgue measure, with densities f and $\tilde{f}$, and then we may use the notation $\mathbb{W}_M(f, \tilde{f})$. One result we need is that if f and $\tilde{f}$ are probability densities on $\mathbb{R}^q$ then

$$\mathbb{W}_M(f, \tilde{f}) \leq 2 \left(\int_{\mathbb{R}^q} |x|^M |f(x) - \tilde{f}(x)| dx \right)^{1/M} \quad (5)$$

for any $M \geq 1$. This follows from Proposition 7.10 in [15]. We will in fact need a specific coupling (i.e. joint distribution) between two random variables X and $\tilde{X}$, with densities f and $\tilde{f}$ respectively, which realises (5), and one such is as follows: let $a = \int (f - \tilde{f})_+ = \int (\tilde{f} - f)_+$, where g_+ means $\max(g, 0)$, then with probability $1 - 2a$ generate X with density $(1 - 2a)^{-1} \min(f(x), \tilde{f}(x))$ and set $\tilde{X} = X$, otherwise generate X and $\tilde{X}$ independently with densities $a^{-1}(f(x) - \tilde{f}(x))_+$ and $a^{-1}(\tilde{f}(x) - f(x))_+$ respectively. Then one readily checks that X and $\tilde{X}$ do have the specified densities, and a straightforward calculation gives

$$\mathbb{E}|X - \tilde{X}|^M \leq 2^{M-1} \int |x|^M |f(x) - \tilde{f}(x)| dx \quad (6)$$

from which (5) follows. Details can be found in [15].

Much more on this topic can be found in the books of Rachev and Ruschendorf [12] and Villani [15, 16]. And a note on our spelling of ‘Vaserstein’: his original paper [14] is in Russian, we have used the transliteration ‘Vaserstein’ (rather than the alternative ‘Wasserstein’) from the Cyrillic alphabet as that is the one he himself has used in his English-language publications.

3 Polynomial perturbations of normal distributions

In this section we outline some results from [4] which we will use. Fix $q \in \mathbb{N}$ and let P denote the space of all real-valued polynomials on $\mathbb{R}^q$, and P^q the space of $\mathbb{R}^q$ -valued polynomial functions on $\mathbb{R}^q$. We also fix a positive-definite $q \times q$ matrix Σ . Let $p_1, \dots, p_k \in P^q$. For $\epsilon \in \mathbb{R}$ we define $\rho_\epsilon : \mathbb{R}^q \rightarrow \mathbb{R}^q$ by $\rho_\epsilon(x) = x + \sum_{j=1}^k \epsilon^j p_j(x)$. We are interested in the distribution of $\rho_\epsilon(X)$ where X is an $\mathbb{R}^q$ -valued random variable with $N(0, \Sigma)$ distribution and ϵ is close to 0. Pretending that ρ_ϵ is bijective, then this distribution has a density given by

$$f_\epsilon(y) = \det(D\rho_\epsilon^{-1}(y)) \phi_\Sigma(\rho_\epsilon^{-1}(y)) \quad (7)$$

By expanding $\rho_\epsilon^{-1}(y)$ formally in powers of ϵ one can obtain a formal expansion $f_\epsilon(y) \sim \phi_\Sigma(y)(1 + \sum_{k=1}^\infty \epsilon^k S_k(y))$ where $S_k \in P$.

We remark that, if $p \in P^q$ and $\det(Dp)$ is not identically 0, then the zero set of $\det(Dp)$ is a closed set of measure zero, and p is locally invertible on its complement, from which one can deduce that $p(X)$ has a density. Applying this to ρ_ϵ , one sees that $\det(D\rho_\epsilon(x)) = 1 + \epsilon h(\epsilon, x)$ where h is a polynomial, so that if ϵ is small enough it cannot vanish identically. Hence $\rho_\epsilon(X)$ has a density f_ϵ for ϵ small enough.

The construction of the S_j from the p_j is described in [4]. One sees that for $j < n$, S_j depends only on $\{p_m : m \leq j\}$. It follows that, given a sequence $p_1, p_2, \dots$ with $p_j \in P^q$ we obtain a well-defined sequence $S_1, S_2, \dots$. Now we introduce the notation $\mathcal{P}$ for the set of all sequences $(u_1, u_2, \dots)$ with $u_j \in P$, and similarly $\mathcal{P}^q$.

We write $\mathcal{P}_\Sigma$ for the set of sequences satisfying $\int u_k(y)\phi_\Sigma(y)dy = 0$ for all k . Then we define the linear mapping $\mathcal{S}_\Sigma : \mathcal{P}^q \rightarrow \mathcal{P}_\Sigma$ by $\mathcal{S}_\Sigma(p_1, p_2, \dots) = (S_1, S_2, \dots)$ as just described. It is shown in [4] that this mapping is onto. In one dimension it is also (1-1); it gives the connection between Cornish-Fisher expansions and Edgeworth expansions. But when $q > 1$ it is not (1-1), meaning that there are many Cornish-Fisher expansions for a given Edgeworth expansion.

The calculation of S_n from $p_1, \dots, p_n$ is straightforward in principle, though the algebra becomes heavy as n increases. As indicated in [4], we find $S_1(y) = y^t \Sigma^{-1} p_1(y) - \nabla \cdot p_1(y)$ and

$$\begin{aligned} S_2(y) = & \frac{1}{2}(S_1(y))^2 - \frac{1}{2}p_1(y)^t \Sigma^{-1} p_1(y) - y^t \Sigma^{-1} Dp_1(y)p_1(y) \\ & + \frac{1}{2}\text{tr}(Dp_1(y)^2) + p_1(y)^t \nabla(\nabla \cdot p_1)(y) + y^t \Sigma^{-1} p_2(y) - \nabla \cdot p_2(y) \end{aligned}$$

The following result from [4] demonstrates the sense in which the series $(1 + \sum_{j=1}^\infty \epsilon^j S_j)\phi_\Sigma$ can be regarded as an asymptotic expansion for the density f_ϵ of $\rho_\epsilon(X)$.

Proposition 1 *With notation as above, for any n and $M \geq 1$ there exists $C > 0$ so that*

$$\int_{\mathbb{R}^q} (1 + |y|)^M \left| f_\epsilon(y) - \left(1 + \sum_{j=1}^n \epsilon^j S_j(y) \right) \phi_\Sigma(y) \right| \leq C |\epsilon|^{n+1} \quad (8)$$

for all sufficiently small ϵ .

Here we will be concerned with the reverse direction, where we have a sequence $(S_1, S_2, \dots)$ giving an asymptotic expansion of a family (f_ϵ) of probability densities, and wish to find a corresponding sequence $(p_1, p_2, \dots)$ for the purpose of generating an approximation to $\mathbb{P}_\epsilon$. Generally, if $(f_\epsilon)_{\epsilon \in (0,1)}$ is a family of probability densities on $\mathbb{R}^q$ and $(S_n)_{n \in \mathbb{N}}$ is a sequence in $\mathcal{P}$, then we say that (S_n) is an α_Σ -sequence for $(\mathbb{P}_\epsilon)$ if, for any $M \geq 1$, there exists $C > 0$ so that (8) holds for all sufficiently small $\epsilon > 0$. Note that an α_Σ -sequence is automatically an $\mathcal{A}_\Sigma$ -sequence as defined in [4] - the notion of $\mathcal{A}_\Sigma$ -sequence is slightly more general and allows for probability measures without densities, but the above notion of α_Σ sequence is somewhat simpler and suffices for our purpose here.

Then Theorem 4 of [4] gives the following:

Proposition 2 *Suppose $(f_\epsilon : \epsilon \in (0, 1))$ and $(\tilde{f}_\epsilon : \epsilon \in (0, 1))$ are families of probability densities on $\mathbb{R}^q$ having respectively an α_Σ -sequence $(S_1, S_2, \dots)$ and an α_Σ -sequence $(\tilde{S}_1, \tilde{S}_2, \dots)$. Suppose also that, for some $n \in \mathbb{N}$, we have $S_j = \tilde{S}_j$ for $1 \leq j \leq n$. Let $M \geq 1$ be given. Then we can find $C > 0$ such that $\mathbb{W}_M(f_\epsilon, \tilde{f}_\epsilon) \leq C |\epsilon|^{n+1}$ for all $\epsilon \in (0, 1)$.*

Note that Proposition 1 says that $(S_1, S_2, \dots)$ is an α_Σ -sequence for $\rho_\epsilon(X)$. The idea now is that, if we have an α_Σ -sequence $(S_1, S_2, \dots)$ for some family f_ϵ , we can

find a sequence $(p_1, p_2, \dots)$ giving the same $(S_1, S_2, \dots)$, and then use the resulting $\rho_\epsilon(X)$ as an approximation to f_ϵ ; Proposition 2 then gives a bound for the accuracy of this approximation. In this context we now describe, for later use, a specific coupling between f_ϵ and $\rho_\epsilon(X)$ which attains the bound in Proposition 2.

To do this, first fix $M \geq 1$ and let $r \in \mathbb{N}$ with $r + 1 \geq M(n + 1)$. Then let $Z_\epsilon = \rho_\epsilon(X) = X + \sum_{k=1}^n \epsilon^k p_k(X)$ and $V_\epsilon = X + \sum_{k=1}^r \epsilon^k \tilde{p}_k(X)$, where X has $N(0, \Sigma)$ distribution, noting that then $V_\epsilon - Z_\epsilon = \sum_{k=n+1}^r \epsilon^k \tilde{p}_k(X)$, so that $\mathbb{E}|V_\epsilon - Z_\epsilon|^M \leq C\epsilon^{M(n+1)}$. Now if Y_ϵ has density f_ϵ then, using (8), the construction described at the end of Section 1 gives an explicit coupling between Y_ϵ and V_ϵ with $\mathbb{E}|Y_\epsilon - V_\epsilon|^M \leq C\epsilon^{M(n+1)}$. Using this coupling to generate Y_ϵ conditional on V_ϵ we then get a joint distribution of $Y_\epsilon, V_\epsilon, Z_\epsilon$ for which $\mathbb{E}|Y_\epsilon - Z_\epsilon|^M \leq C\epsilon^{M(n+1)}$ as required. We remark that this deduction of a coupling between Y_ϵ and Z_ϵ , given a coupling between V_ϵ and Z_ϵ , and one and one between Y_ϵ and V_ϵ , is an example of *gluing* of couplings, as discussed on page 11 of [16].

In the case $M = 2$ we can give a more precise estimate of the $\mathbb{W}_2$ distance in Proposition 2, using Theorem 11 and equation (22) from [4] (with $\tilde{\Sigma} = \Sigma$). We find that, under the hypotheses of Proposition 2, we have $\mathbb{W}_2(\mathbb{P}_\epsilon, \tilde{\mathbb{P}}_\epsilon) = B\epsilon^{n+1} + O(\epsilon^{n+2})$ where $B = (\int_{\mathbb{R}^q} |p(x)|^2 \phi_\Sigma(x) dx)^{1/2}$, where p is the unique polynomial in P^q which is a gradient and which satisfies $\mathcal{L}_\Sigma p = S_{n+1} - \tilde{S}_{n+1}$.

General references for Edgeworth-type expansions in several variables are [1] and [11]. For the classical theory of Edgeworth and Cornish-Fisher expansions in one variable see [7] and [8] for example. One method of constructing Cornish-Fisher expansions in several variables is given in [13].

4 Construction of asymptotic expansions

We return to the SDE $dx(t) = a(t, x(t))dt + B(t, x(t))dW(t)$ with $x(0) = x^{(0)}$ with initial condition $x(0) = x^{(0)}$ of (1), and consider the distribution of the random variable $Y_t = t^{-1/2}(x(t) - x^{(0)})$. For the present we make the following assumption: (*) a and B are C^∞ on $(0, T) \times \mathbb{R}^q$, for each multi-index α the derivatives $D^\alpha a$ and $D^\alpha B$ are bounded, and also B has rank q everywhere and $(BB^t)^{-1}$ is bounded.

Then Y_t will have a density f_t and we look for an asymptotic expansion

$$f_t(y) \sim \phi_{\Sigma^{(0)}}(y) \left(1 + \sum_{n=1}^{\infty} t^{n/2} S_n(y) \right) \quad (9)$$

valid for small $t > 0$, to which we can apply the methods from section 3, with $\epsilon = t^{1/2}$. To do this we relate f_t to the density of the solution of the SDE: let $u(t, x)$ denote the density of $x(t)$. Then u satisfies the forward equation $Lu = 0$ where

$$Lu(t, x) = \partial_t u(t, x) + \sum_i \partial_i (a_i(t, x)u(t, x)) - \frac{1}{2} \sum_{i,j} \partial_{ij} (\Sigma_{ij}(t, x)u(t, x)) \quad (10)$$

where $\Sigma(t, x) = B(t, x)B(t, x)^t$ and ∂_{ij} means $\frac{\partial^2}{\partial x_i \partial x_j}$, and we have $f_t(y) = t^{q/2}u(t, t^{1/2}y)$. We also have $u(t, x) = t^{-q/2}f_t(t^{-1/2}(x - x^{(0)}))$ and then, if we have an expansion (9) we obtain

$$u(t, x) \sim t^{-q/2}\phi_{\Sigma^{(0)}}(t^{-1/2}(x - x^{(0)})) \left(1 + \sum_{n=1}^{\infty} t^{n/2} S_n(t^{-1/2}(x - x^{(0)})) \right) \quad (11)$$

The idea is to substitute the RHS of (11) for $u(t, x)$ in (10), then find polynomials $S_1, S_2, \dots$ so that $Lu = 0$ holds in a formal power series sense, and finally verify that with these $(S_1, S_2, \dots)$ the expansion (9) is a valid asymptotic expansion of (f_t) .

To carry out the first step, we first note that by an orthogonal change of variables we may arrange that $\Sigma^{(0)}$ is diagonal, with positive entries $\lambda_1, \dots, \lambda_q$. We also impose the requirement that $a(t, x)$ and $B(t, x)$ are infinitely differentiable in a neighbourhood of $(0, x^{(0)})$. Then we have Taylor expansions $a_i(t, x) \sim \sum_{k, \alpha} a_{ik\alpha} t^k (x - x^{(0)})^\alpha = \sum a_{ik\alpha} t^{k+|\alpha|/2} y^\alpha$, where $y = t^{-1/2}(x - x^{(0)})$, and $\Sigma_{ij}(t, x) \sim \sum_{k, \alpha} \sigma_{ijk\alpha} t^k (x - x^{(0)})^\alpha = \sum \sigma_{ijk\alpha} t^{k+|\alpha|/2} y^\alpha$, where the sums are over all nonnegative integers k and nonnegative multi-indices α of length q . We have $\sigma_{ijk\alpha} = \sigma_{jik\alpha}$ and $\sigma_{ij00} = \lambda_i \delta_{ij}$.

Then the sought expansion (11) can be written as $u(t, x) \sim \sum_{n=0}^{\infty} t^{\frac{n-q}{2}} T_n(t^{-1/2}(x - x^{(0)}))$ where $T_n(y) = S_n(y)\phi_{\Sigma^{(0)}}(y)$. Then

$$\partial_t u(t, x) \sim \frac{1}{2} \sum_{n=0}^{\infty} t^{\frac{n-q-2}{2}} ((n-q)T_n(y) - y \cdot \nabla T_n(y)) \quad (12)$$

where $y = t^{-1/2}(x - x^{(0)})$. Likewise we have

$$\sum_{i,j=1}^q \partial_{ij}(\Sigma_{ij}(t, x)u(t, x)) \sim \sum_{n=0}^{\infty} t^{\frac{n-q-2}{2}} \left(\sum_{i=1}^q \lambda_i \partial_{ii} T_n(y) + Q_n(y)\phi_{\Sigma^{(0)}}(y) \right) \quad (13)$$

and

$$\sum_i \partial_i(a_i(t, x)u(t, x)) \sim \sum_{n=1}^{\infty} t^{\frac{n-q-2}{2}} R_n(y)\phi_{\Sigma^{(0)}}(y) \quad (14)$$

where Q_n and R_n are polynomials depending only on S_k for $k < n$. Then substituting (12), (14) and (13) into $Lu = 0$ using (10), and comparing coefficients of $t^{\frac{n-q-2}{2}}$, we obtain $(n-q)T_n(y) - y \cdot \nabla T_n(y) = \sum_{i=1}^q \lambda_i \partial_{ii} T_n(y) + Q_n(y)\phi_{\Sigma^{(0)}}(y)$. Now we have $\partial_i T_n(y) = (\partial_i S_n(y) - \lambda_i^{-1} y_i S_n(y))\phi_{\Sigma^{(0)}}(y)$ and $\partial_{ii} T_n(y) = (\partial_{ii} S_n(y) - 2\lambda_i^{-1} y_i \partial_i S_n(y) + \lambda_i^{-2} (y_i^2 - \lambda_i) S_n(y))\phi_{\Sigma^{(0)}}(y)$, and we deduce finally that

$$nS_n(y) + y \cdot \nabla S_n(y) - \sum_{i=1}^q \lambda_i \partial_{ii} S_n(y) = \tilde{Q}_n(y) \quad (15)$$

is the requirement for (10) to hold in a formal power series sense, where $\tilde{Q}_n = Q_n - 2R_n$.

Now for $n \in \mathbb{N}$ we define an operator Λ_n , acting on polynomials in q variables, by $\Lambda_n p(y) = np(y) + y \cdot \nabla p(y) - \sum_{i=1}^q \lambda_i \partial_{ii} p(y)$, then we see that for any non-negative multi-index α we have $\Lambda_n(y^\alpha) = (n + |\alpha|)y^\alpha + g$ where g has degree $|\alpha| - 2$. It then follows by induction on degree that the equation $\Lambda_n p = Q$ has a unique polynomial solution p for any given polynomial Q . Since (15) can be written as $\Lambda_n S_n = \tilde{Q}_n$, we conclude by induction on n that there is a unique sequence of polynomials (S_n) satisfying (15).

Recursive calculation of S_n

For notational simplicity, we assume here that $\lambda_i = 1$ for each i (which we can always arrange by taking a non-orthogonal initial coordinate change), and then we find after some calculations that (15) can be written as

$$\Lambda_n S_n = \tilde{Q}_n = 2 \sum^* a_{ik\alpha} A_{i\alpha} S_{n-1-2k-|\alpha|} + \sum' \sigma_{ijk\alpha} B_{ij\alpha} S_{n-2k-\alpha} \quad (16)$$

where $A_{i\alpha}$ and $B_{ij\alpha}$ are operators on polynomials defined by

$$A_{i\alpha} S(y) = (y^{\alpha+e_i} - \alpha_i y^{\alpha-e_i}) S(y) + y^\alpha \partial_i S(y) \quad (17)$$

and

$$\begin{aligned} B_{ij\alpha} S(y) = & \{(\alpha_i \alpha_j - \delta_{ij} \alpha_i) y^{\alpha-e_i-e_j} - 2\alpha_i y^{\alpha-e_i+e_j} + (y_i y_j - \delta_{ij}) y^\alpha\} S(y) \\ & + 2(\alpha_i y^{\alpha-e_i} - y^{\alpha+e_i}) \partial_j S(y) + y^\alpha \partial_{ij} S(y) \end{aligned} \quad (18)$$

and where $\sum^*$ denotes a sum over all i, k, α with $i \in \{1, \dots, q\}$, $k \geq 0$ and multi-indices α such that $2k + |\alpha| < n$, while $\sum'$ is a similar sum over i, j, k, α such that $0 < 2k + |\alpha| \leq n$; here e_i denotes the multi-index with i 'th entry 1 and all other entries 0.

In order to apply the recurrence relation (16) for S_n it may be helpful to express the RHS in terms of Hermite polynomials and note the easily verified fact that

$$\Lambda_n H_\alpha = (n + |\alpha|) H_\alpha \quad (19)$$

where $H_\alpha(y) = H_{\alpha_1}(y_1) \cdots H_{\alpha_q}(y_q)$. The algebra involved in this inductive construction is in general quite heavy, as is often the case with Edgeworth-type expansions. The case $n = 1$ is relatively simple - we find that $\Lambda_1 S_1(y) = 2 \sum_i a_{i00} y_i + \Omega$ where $\Omega = \sum_{i,j,s} \lambda_{ijs} y_s (y_i y_j - \delta_{is} - \delta_{js} - \delta_{ij})$, with $\lambda_{ijs} = \sigma_{ij0e_s} = \partial_s \Sigma_{ij}(0, x^{(0)})$. We can write $\Omega = \sum_s \lambda_{sss} H_3(y_s) + \sum_{i \neq s} (2\lambda_{isi} + \lambda_{iis}) H_2(y_i) y_s + \sum^* \lambda_{ijs} y_i y_j y_s$ where $\sum^*$ denotes a sum over all cases where i, j, s are distinct. Then we see that Ω is a linear combination of H_α terms with $|\alpha| = 3$, and deduce using (19) that

$$\begin{aligned}
S_1(y) &= \sum_{i=1}^q a_{i00} y_i + \frac{1}{2} \Omega = \sum_i a_{i00} y_i + \frac{1}{4} \sum_{i,j,s=1}^q \lambda_{ijs} y_i y_j y_s \\
&\quad - \frac{1}{4} \sum_{i,s=1}^q (2\lambda_{isi} + \lambda_{iis}) y_s
\end{aligned} \tag{20}$$

The algebra is simpler in the case $q = 1$. In this case we have $\Lambda_n p(y) = np(y) + yp'_n(y) - p''_n(y)$ and (16) becomes

$$\begin{aligned}
\Lambda_n S_n &= -2 \sum_{k,r}^* a_{kr} y^{r-1} ((r - y^2) S_{n-1-2k-r}(y) + y S_{n-1-2k-r}(y)) \\
&\quad + \sum_{k,r}^{\prime} \sigma_{kr} y^{r-2} \{ (r(r-1) - (2r+1)y^2 + y^4) S_{n-2k-r} \\
&\quad + 2y(r - y^2) s'_{n-2k-r} + y^2 s''_{n-2k-r} \}
\end{aligned}$$

where $\sum^*$ is a sum over non-negative integers (k, r) such that $2k+r < n$, while $\sum'$ is a sum over pairs such that $0 < 2k+r \leq n$. From this we obtain $S_1(y) = a_{00}y + \frac{\sigma_{01}}{4} H_3(y)$ (which also follows from (20)) and

$$\begin{aligned}
S_2 &= \frac{1}{2} a_{0,0}^2 H_2 + \frac{1}{4} a_{0,0} \sigma_{0,1} (H_4 + H_2) + \frac{1}{2} a_{0,1} H_2 + \sigma_{0,1}^2 \left(\frac{1}{32} H_6 + \frac{1}{4} H_4 \right) \\
&\quad + \sigma_{0,2} \left(\frac{1}{6} H_4 + \frac{1}{4} H_2 \right) + \frac{1}{4} \sigma_{1,0} H_2
\end{aligned}$$

We also note two properties of S_n which follow easily by induction on n using (16): S_n is odd or even according as n is odd or even, and S_n has degree at most $3n$.

The following theorem makes precise the sense in which the expansion (9) is an asymptotic expansion of the distribution of Y_t .

Theorem 1 Suppose $x(t)$ is a solution of an SDE of the form (1) on $(0, T)$ with initial condition $x(0) = x^{(0)}$, and that $\mathbb{E}|x(t)|^K < \infty$ for all $K \geq 1$. Suppose also that the coefficients a and B are C^∞ on a neighbourhood of $(0, x^{(0)})$, and that $B(0, x^{(0)})$ has rank q . Let S_n be the unique sequence of polynomials satisfying (15). Then (S_n) is an $\alpha_{\Omega(0)}$ -sequence for the distribution of Y_t ; in other words, given $n \in \mathbb{N}$ and $M \geq 1$ there exist $C > 0$ such that the bound

$$\int_{\mathbb{R}^q} (1 + |y|)^M d|\mathbb{P}_t - \nu_{t,n}|(y) \leq Ct^{(n+1)/2} \tag{21}$$

holds for all $t \in (0, T)$, where $\mathbb{P}_t$ is the distribution measure of Y_t and $\nu_{t,n}$ is the measure with density $f_{t,n}(y) := \phi_{\Sigma(0)}(y) (1 + \sum_{k=1}^n t^{k/2} S_k(y))$

Proof We fix n and M . and first assume (*); we shall use the notation $C_1, C_2, \dots$ for constants which depend only on n, M and bounds for $(BB^t)^{-1}$ and for $D^\alpha a$ and $D^\alpha B$ for finitely many α . Let $N = \max(q + d, 2n + 6)$ define u_N by the truncated

expansion $u_N(t, x) = \sum_{n=0}^N t^{\frac{n-q}{2}} T_n(t^{-1/2}(x - x^{(0)}))$. Then the fact that (10) holds in a formal power series sense implies that $Lu_N = \psi_N$ for $t > 0$, where $\psi_N(t, x) = t^{\frac{N-q-1}{2}} \phi_{\Sigma^{(0)}}(y) V_N(t^{1/2}, y)$ where V_N is a polynomial in $q + 1$ variables. We extend ψ_N and u_N to $W := (-\infty, T) \times \mathbb{R}^q$ by setting then to 0 for $t \leq 0$. Then using $N > q + 3$ one can check that ψ_N is C^1 , and bounded with bounded first derivatives on W . And, as a distribution on W , we have $Lu_N = \psi_N + \delta_0$, where δ_0 is the unit point mass at $(0, 0)$. Moreover, if $0 < \tau < T$, then ψ_N and its first derivatives are bounded by $C_1 \tau^{\frac{N-q-4}{2}}$ on $W_\tau = (-\infty, \tau] \times \mathbb{R}^q$. Also $Lu = \delta_0$ so $L(u - u_N) = \psi_N$. Then by standard PDE results (for example using Theorem 8.10.1 in [10], with any $\delta \in (0, 1)$, and noting the comment after equation (8.0.2) that the condition $c \leq 0$ can be avoided) we conclude that $|u(t, x) - u_N(t, x)| \leq C_2 \tau^{\frac{N-q-4}{2}}$ on W_τ . This means that $|f_t(y) - f_{t,N}(y)| \leq C_2 t^{\frac{N-4}{2}}$ for $0 < t < T$, $y \in \mathbb{R}^q$, and hence

$$\int_{\mathbb{R}^q} (1 + |y|)^{-1-q} |f_{t,N}(y) - f_t(y)| \leq C_3 t^{(N-4)/2} \quad (22)$$

Let $R = 2M + q + 1$, then we have $\int (1 + |y|)^R |f_t(y)| dy \leq C_4$, and clearly $\int (1 + |y|)^R |f_{t,N}(y)| dy \leq C_5$, so we have

$$\int (1 + |y|)^R |f_{t,N} - f_t(y)| dy \leq C_6 \quad (23)$$

Applying Cauchy–Schwartz to (22) and (23), we obtain $\int (1 + |y|)^M |f_{t,N}(y) - f_t(y)| dy \leq C_7 t^{(n+1)/2}$. We also have $\int (1 + |y|)^M |f_{t,N}(y) - f_{t,n}(y)| dy \leq C_7 t^{(n+1)/2}$ and hence $\int (1 + |y|)^M |f_{t,n}(y) - f_t(y)| dy \leq C_9 t^{(n+1)/2}$ which is (21) in this case.

To treat the general case, given a and B satisfying the hypotheses of the theorem, we find $\tilde{a}$ and $\tilde{B}$ satisfying (*), and such that they agree with a and B respectively for $|x - x_0| < \delta$, for some $\delta > 0$. Let $\tilde{\mathbb{P}}_t$ be the distribution of $\tilde{Y}_t$. Then, for given M , we have C such that

$$\int_{\mathbb{R}^q} (1 + |y|)^M d|\tilde{\mathbb{P}}_t - \nu_{t,n}|(y) \leq C t^{(n+1)/2} \quad (24)$$

for each t . Now $\int (1 + |y|)^{2M} d|\mathbb{P}_t - \tilde{\mathbb{P}}_t|(y) \leq K$, while

$$\int d|\mathbb{P}_t - \tilde{\mathbb{P}}_t| \leq 2\mathbb{P}(|x(s) - x^{(0)}| \geq \delta \text{ for some } s \in [0, t]) \leq K t^{n+1}$$

for some constant K . Then Cauchy–Schwartz gives $\int (1 + |y|)^M d|\mathbb{P}_t - \tilde{\mathbb{P}}_t|(y) \leq K^{(n+1)/2}$, which together with (24) gives (21). $\square$

We remark that the hypotheses of this theorem allow cases where the distribution $\mathbb{P}_t$ does not have a density. For example the one-dimensional SDE $dx = |x|^{1/2} dW$ with initial condition $x(0) = -1$ has a solution which stays at 0 after first reaching 0, and for this solution $P(x(t) = 0) > 0$ for all $t > 0$ so $\mathbb{P}_t$ has a point mass at 0.

In fact the application of theorem 1 in the following sections only uses the case (*), but we have included the general case in case the asymptotic expansion may have some independent interest.

5 One-step approximation

Now we apply the results described in the previous sections to the construction of an approximation scheme. Given an SDE of the form (1), we find the polynomials S_n in the expansion (9) of the density of Y_t , then use the methods of [4] as outlined in Sect. 3 above to construct a Cornish-Fisher type expansion $X + t^{1/2}p_1(X) + tp_2(X) + \dots$, where X is an $N(0, \Sigma^{(0)})$ random variable, which expansion can be truncated to give a random variable whose distribution approximates that of Y_t . To illustrate the procedure we consider the relatively simple case of p_1 . We assume for simplicity that $d = q$, and then we assume as usual that $B(0, x^{(0)})$ is invertible. As before, by suitable choice of coordinates we can assume that $B(0, x^{(0)}) = I$, and the $\Sigma^{(0)} = I$. We recall the expression (20) for S_1 , and note that the requirement for p_1 is that $y^t p_1(y) - \nabla \cdot p_1(y) = S_1(y)$. One choice for p_1 given by the approximation scheme (3), as follows: the scheme (3) with $h = t$ gives the approximation to $Y_t = t^{-1/2}(x(t) - x^{(0)})$

$$(Y_t)_i \approx X_i + t^{1/2} \left(a_i + \frac{1}{2} \sum_{j,k=1}^d \rho_{ijk} (X_j X_k - \delta_{jk}) \right) \quad (25)$$

which corresponds to $p_1(y)_i = a_i + \frac{1}{2} \sum_{j,k=1}^d \rho_{ijk} (y_j y_k - \delta_{jk})$. Then we obtain

$$\begin{aligned} S_1(y) &= y \cdot p_1(y) - \nabla \cdot p_1(y) = \frac{1}{2} \sum_{i,j,k=1}^d \rho_{ijk} y_i y_j y_k \\ &\quad + \sum_{i=1}^d y_i \left(a_i - \frac{1}{2} \sum_{k=1}^d (\rho_{ikk} + \rho_{kik} + \rho_{kki}) \right) \end{aligned} \quad (26)$$

Recalling that $B = I$ we see that $\lambda_{ijs} = \rho_{isj} + \rho_{jsi}$ and then the above agrees with (20) as claimed.

We note in passing that the scheme (3) also occurs in a different setting, in the antithetic multilevel method for weak approximation in [5].

As mentioned above, the choice of p_1 for a given S_1 is far from unique if $q > 1$; one can add to p_1 any $p \in P^q$ such that $y^t p(y) - \nabla \cdot p(y) = 0$, without changing the corresponding S_1 . An example is the scheme derived by geometric methods in [2], equation (3), which corresponds to the choice $p_1(y)_i = a_i + \frac{1}{2} \sum_{j,k=1}^d (\rho_{ijk} + \rho_{jki} - \rho_{jik})(y_j y_k - \delta_{jk})$: it is easily checked that this gives the same S_1 as above.

In general, given the Edgeworth-type expansion with $S_1, S_2, \dots$ as in Sect. 4, for any choice of $(p_1, p_2, \dots)$ such that $\Sigma_{\Sigma^{(0)}}(p_1, p_2, \dots) = (S_1, S_2, \dots)$, we can construct an approximation for any given n by

$$x(t) = x^{(0)} + t^{1/2} Y_t \approx x^{(0)} + t^{1/2} X + \sum_{r=1}^n t^{(r+1)/2} p_r(X) + t^{(n+2)/2} \mathbb{E} p_{n+1}(X) \quad (27)$$

where X is an $N(0, \Sigma^{(0)})$ random variable. Then, writing $Z(t)$ for the RHS of (27), from Propositions 1 and 2, together with Theorem 1, we have $\mathbb{W}_M(x(t), Z(t)) = O(t^{(n+2)/2})$ for any $M \geq 1$. And the discussion following Proposition 2 gives (for any given $M \geq 1$) a specific coupling between $x(t)$ and $Z(t)$ which attains this bound. Note that the last term in (27) does not give any improvement in this W_M bound, but it does give the bound $|\mathbb{E}x(t) - \mathbb{E}Z(t)| = O(t^{(n+3)/2})$ which is needed to get an optimal bound for the multi-step approximation in the next section.

We remark that if n is odd then $\mathbb{E} p_{n+1}(X) = \int S_{n+1}(y) y dy = 0$ since S_{n+1} is even, so the last term on the right of (27) appears only if n is even.

6 Convergence of stepwise approximations

Now we consider the approximation of a solution of (1) by successive application of the approximation (27) with a fixed time-step h . More precisely, we consider a sequence $x^{(k)}$ defined by a recurrence

$$x^{(k+1)} = x^{(k)} + h^{1/2} X^{(k)} + \sum_{r=1}^n h^{(r+1)/2} p_r^{(k)}(X^{(k)}) h^{(n+2)/2} \mathbb{E} p_{n+1}^{(k)}(X^{(k)}) \quad (28)$$

where $X^{(k)}$ is an $N(0, \Sigma^{(k)})$ random variable. Here $\Sigma^{(k)} = \Sigma(kh, x^{(k)})$ and the sequence $(p_1^{(k)}, p_2^{(k)}, \dots)$ satisfies $\mathcal{S}_{\Sigma^{(k)}}(p_1^{(k)}, p_2^{(k)}, \dots) = (S_1^{(k)}, S_2^{(k)}, \dots)$ where the sequence $(S_1^{(k)}, S_2^{(k)}, \dots)$ is obtained by the construction in Sect. 4 applied to (1) with initial value $x^{(k)}$.

We assume first that the coefficients a and B satisfy condition (*) from the proof of Theorem 1. Let $u^{(k)}(t)$ be the solution of (1) for $t > kh$ satisfying $u^{(k)}(kh) = x^{(k)}$, and let M be fixed. Then the procedure described at the end of Sect. 6 gives for each k , conditional on $x^{(k)}$, a coupling between $x^{(k+1)}$ and $u^{(k)}(kh)$ such that $\mathbb{E}|x^{(k+1)} - u^{(k)}(kh)|^M \leq Ch^{M(n+2)/2}$ and $|\mathbb{E}(x^{(k+1)} - u^{(k)}(kh))| \leq Ch^{M(n+3)/2}$. Using these couplings we can enlarge the probability space, and the associated filtration $\mathcal{F}_t$, on which the solution $x(t)$ is defined, so that there is a sequence $(\tilde{x}_k)$ of random variables so that $\tilde{x}_k$ is $\mathcal{F}_{kh}$ measurable, $(\tilde{x}^{(1)}, \tilde{x}^{(2)}, \dots)$ have the same joint distribution as $(x^{(1)}, x^{(2)}, \dots)$, and we have the bounds

$$\mathbb{E}|\tilde{x}^{(k+1)} - u^{(k)}(kh)|^M \leq Ch^{M(n+2)/2} \text{ and } |m_k| \leq Ch^{M(n+3)/2} \text{ a.s.} \quad (29)$$

where $m_k = \mathbb{E}(\tilde{x}^{(k+1)} - u^{(k)}(kh) | \mathcal{F}_{kh})$. We can now state the main error bound:

Theorem 2 *With the above notation, and assuming that a and B satisfy (*), for given $T > 0$ and $M \geq 1$ there is a constant C such that for any $h > 0$ the following bound holds:*

$$\mathbb{E} \max_{k \leq T/h} |\tilde{x}^{(k)} - x(kh)|^M \leq Ch^{M(n+1)/2} \quad (30)$$

Proof We have $x((k+1)h) - \tilde{x}^{(k+1)} = x((k+1)h) - u^{(k)}((k+1)h) + u^{(k)}((k+1)h) - \tilde{x}^{(k+1)} = x(kh) - \tilde{x}^{(k)} + \int_{kh}^{(k+1)h} (B(t, x(t)) - B(t, u^{(k)}(t))) dW_t + \int_{kh}^{(k+1)h} (a(t, x(t)) - a(t, u^{(k)}(t))) dt - \tilde{x}^{(k+1)} + u^{(k)}(kh)$ from which, using the bounds (29), it is straightforward to deduce that

$$\mathbb{E}|x((k+1)h) - \tilde{x}^{(k+1)}|^M \leq (1 + Kh)\mathbb{E}|x(kh) - \tilde{x}^k|^M$$

and from this that the bound (30) holds for individual k . To get the maximal bound, note that $x(kh) - u^{(k)}$ can be written as $M_k + V_k$ where

$$M_k = \sum_{j=0}^{k-1} \left(\int_{jh}^{(j+1)h} (B(t, x(t)) - B(t, u^{(j)}(t))) dW_t + u^{(j)}(jh) - \tilde{x}^{(j+1)} - m_j \right)$$

and $V_k = \sum_{j=0}^{k-1} \left(\int_{jh}^{(j+1)h} (a(t, x(t)) - a(t, u^{(j)}(t))) dW_t + m_j \right)$. Then (M_k) is a martingale and (30) follows from the maximal martingale inequality and the uniform bound for m_j . $\square$

The global condition (*) is rather restrictive and we now give an application of Theorem 2 which covers a rather more general situation. We suppose a and B are C^∞ and B has rank q on $[0, T] \times U$ where U is an open set in $\mathbb{R}^q$, and consider (1) with $x^{(0)} \in U$. We fix compact sets K and K^* such that $x^{(0)} \in \text{int}(K)$ and $K \subset \text{int}(K^*) \subset U$. Then for a given time-step $h > 0$ and $n \in \mathbb{N}$ we define a modified version of the recursion (28) as follows: as long as the RHS of (28) is in K^* , we define $x^{(k+1)}$ by (28), but as soon as the RHS is outside K^* we set $x^{(j)} = x^{(k)}$ for all $j > k$.

Then we assert the following:

Theorem 3 *With the above assumptions and notation, given $M \geq 1$ and $n \in \mathbb{N}$ there exists $C > 0$ such that, for any $h > 0$, there is a probability space, equipped with a filtration $\mathcal{F}_t$, on which the solution $x(t)$ to (1) is defined, along with a sequence of random vectors $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \dots$, having the same joint distribution as $(x^{(1)}, x^{(2)}, \dots)$ as defined above, and such that $\tilde{x}^{(k)}$ is $\mathcal{F}_{kh}$ -measurable, and*

$$\mathbb{E} \max_{k \leq \tau/h} |\tilde{x}^{(k)} - x(kh)|^M \leq Ch^{M(n+1)/2} \quad (31)$$

where τ is the escape time of $x(t)$ from K .

Proof We can find modified coefficients $\tilde{a}(t, x)$ and $\tilde{B}(t, x)$ which satisfy (*) and agree with a and B on K^* . Then we apply Theorem 2 to the modified system. The vectors $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \dots$ from the modified system will have the required distribution, except that any $\tilde{x}^{(k)}$ which is outside K^* has to be modified. Since $x(kh) \in K$ for $k \leq \tau/h$, it follows from (30) that the probability of such a modification being needed is $\leq C\delta^{-M}h^{M(n+1)/2}$ where δ is the distance of K from the complement of K^* . When no modification is needed the bound (30) applies, and (31) follows. $\square$

The idea is that one can generate random variables $x^{(1)}, x^{(2)}, \dots$ using (28), and identify them with $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \dots$ since the distributions are the same. In this sense the bounds (30) and (31) may be regarded as strong error bounds. In the concluding section we compare this method with standard strong and weak approximation methods and discuss the extent to which it qualifies as a strong approximation.

7 Concluding remarks

Here we compare the approximation scheme described above to standard strong and weak convergence schemes, and also mention some open questions.

Standard strong approximation schemes, such as described in [9], typically involve approximating the solution $x(t)$ on $[0, T]$ of an SDE by approximations $x^{(k)}$ for $k = 1, \dots, N$, where the $x^{(k)}$ are calculated from simulated increments of the Brownian path, and possibly simulated iterated integrals of the path. One then seeks a bound of the form $\mathbb{E} \max_{k=1}^N |x^{(k)} - x(kh)|^M \leq Ch^{M\gamma}$ where γ is the order of the method. For weak approximation of order γ one is concerned with the estimation of $\mathbb{E}f(x(T))$ for suitable f and only requires bounds of the form $|\mathbb{E}f(z^{(N)}) - \mathbb{E}f(x(T))| \leq Ch^\gamma$. The bounds given in Theorems 2 and 3 then look more like standard strong bounds than weak bounds, but differ in that the random variables $\tilde{x}^{(k)}$ are not measurable w.r.t. the probability space generated by the Brownian path, and their connection to the Brownian path is not explicit. One might describe our method as strong approximation to a weak solution, the ‘weak’ referring to the fact that the solution is on a larger probability space than the Brownian path. It is natural to ask if one can get the same bounds but with $x^{(k)}$ measurable w.r.t. the σ -field generated by the path W . This seems quite possible but would require a different type of coupling.

To illustrate that our method is stronger than standard weak approximation we note that if f is a Lipschitz function on $\mathbb{R}^q$ then (30) with $M = 1$ implies that $|\mathbb{E}f(\tilde{x}^{(N)}) - \mathbb{E}f(x(T))| = O(h^{(n+1)/2})$, so that one can obtain weak approximation of arbitrary order for Lipschitz f , whereas standard weak approximation requires increasing differentiability of f for higher order.

Some general discussion of the interpretation of error bounds based on couplings and Wasserstein distances as a substitute for strong bounds can be found in Section 12 of [3].

Another question on our coupling construction is whether the same $\tilde{x}^{(k)}$ could be chosen for all M - the construction given above is essentially dependent on M .

A different aspect of our approximation scheme which may be worth further investigation is the choice of $(p_1, p_2, \dots)$ corresponding to given $(S_2, S_2, \dots)$, i.e. the choice of a Cornish-Fisher expansion for a given Edgeworth expansion, which in dimension > 1 is not unique. As mentioned at the end of Section 3, $\mathbb{W}_2$ bounds are easier to estimate precisely than $\mathbb{W}_M$ in general. For example, in the context of (27), one could ask for a choice of $p_1, \dots, p_n$ which minimises the leading order term in the expansion of $\mathbb{W}_2(x(t), Z(t))$ in powers of t . This is an essentially algebraic problem which merits exploration.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bhattacharya, R.N., Rao, R.R.: Normal approximations and asymptotic expansions, SIAM (2010)
2. Cruzeiro, A.B., Malliavin, P., Thalmeyer, A.: Geometrization of Monte-Carlo numerical analysis of an elliptic operator: strong approximation, C. R. Acad. Sci. Paris, Ser. I. **338**: 481–486
3. Davie, A.M.: Pathwise approximation of stochastic differential equations using coupling, preprint: www.maths.ed.ac.uk/~adavie/coum.pdf
4. Davie, A.M.: Polynomial perturbations of normal distributions, preprint: www.maths.ed.ac.uk/~adavie/polg.pdf
5. Giles, M.B., Szpruch, L.: Antithetic multilevel Monte Carlo estimation for multidimensional SDEs without Lévy area simulation, [arXiv:1202.6283](https://arxiv.org/abs/1202.6283)
6. Gyöngy, I., Krylov, N.: Existence of strong solutions for Itô's stochastic equations via approximations. Probab. Theory Relat. Fields **105**, 143–158 (1996)
7. Hall, P.: The Bootstrap and Edgeworth Expansion. Springer-Verlag, Berlin (1992)
8. Stuart, A., Ord, K.: Kendall's Advanced Theory of Statistics, Volume 1, 6th edition, Edward Arnold (1994)
9. Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer-Verlag, Berlin (1999)
10. Krylov, N.V.: Lectures on Elliptic and Parabolic Equations in Hölder Spaces, AMS Graduate Studies in Mathematics Volume 12, (1996)
11. Petrov, V.V.: Sums of Independent Random Variables. Springer-Verlag, Berlin (1975)
12. Rachev, S.T., Ruschendorf, L.: Mass Transportation Problems, Volume 1, Theory; Volume 2, Applications. Springer-Verlag, Berlin (1998)
13. Takeuchi, K.: A multivariate generalisation of Cornish-Fisher expansion and its applications. Keizaigaku Ronshu **44**(2), 1–12 (1978). (in Japanese)
14. Vaserstein, L.N.: Markov processes over denumerable products of spaces describing large system of automata. Problemy Peredaci Informacii **5**, 64–72 (1969). (in Russian)
15. Villani, C.: Topics in Optimal Transportation, AMS (2003)
16. Villani, C.: Optimal Transport Old and New. Springer-Verlag, Berlin (2009)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.