KMT theory applied to approximations of SDE

A. M. Davie

September 18, 2014

Abstract

The dyadic method of Komlós, Major and Tusnády is a powerful way of constructing simultaneous normal approximations to a sequence of partial sums of i.i.d. random variables. We use a version of this KMT method to obtain order 1 approximation in a Vaserstein metric to solutions of vector SDEs under a mild non-degeneracy condition using an easily implemented numerical scheme.

1 Introduction

The pathwise simulation of solutions of vector stochastic differential equations is challenging because, using standard methods, to obtain approximations to order greater than $\frac{1}{2}$ requires simulation of iterated integrals of the Brownian path, which is difficult. One approach is to seek approximations in a Vaserstein metric, meaning that there is a coupling between the approximate and exact solutions with respect to which the error is of the desired order. [2] describes an easily generated scheme, based on the standard order 1 Milstein scheme, which is order 1 in a Vaserstein metric, provided the SDE has a nondegenerate diffusion term. Here we describe a modified version of the scheme from [2] which gives order 1 under a weaker nondegeneracy condition. The proof uses a construction of a coupling based on the KMT method.

Section 2 reviews the basics of SDE approximation and states the main result. Section 3 briefly reviews the KMT theorem and presents some required material from coupling and optimal transport theory. The rest of the paper is devoted to the proof of the theorem and a relevant example.

Some other work on SDE approximation using coupling is described in the final chapter of volume 2 of [6]. We also mention [1] which obtains an order $\frac{2}{3} - \epsilon$ bound in a Vaserstein metric for the Euler method in one dimension.

The author is grateful to the referee for several suggestions which improved the presentation of the paper.

2 Approximation of SDEs

Here we briefly review the Milstein scheme and formulate our new version.

Consider an Itô SDE

$$dx_i(t) = \sum_{k=1}^{d} b_{ik}(t, x(t))dW_k(t), \qquad x_i(0) = x_i^{(0)}, \qquad i = 1, \cdots, q \tag{1}$$

on an interval $[0, T]$, for a q-dimensional vector $x(t)$, with a d-dimensional driving Brownian path $W(t)$. If the coefficients $b_{ik}(t, x)$ satisfy a global Lipschitz condition

$$|a_i(t, x) - a_i(t, y)| \leq C|x - y|, \qquad |b_{ik}(t, x) - b_{ik}(t, y)| \leq C|x - y| \tag{2}$$

for all $x, y \in \mathbb{R}^q$, $t \in [0, T]$ and all i, k, where C is a constant, and if a_i and b_i are continuous in t for each x, then (1) has a unique solution $x(t)$ which is a process adapted to the filtration induced by the Brownian motion. This solution satisfies satisfies $\mathbb{E}|x(t)|^p < \infty$ for each $p \in [1, \infty)$ and $t \in [0, T]$.

The standard approach to the strong or pathwise approximation of the solution of (1), as described for example in [4], is to divide $[0, T]$ into a finite number N of subintervals, which we shall usually assume to be of equal length $h = T/N$, and to approximate the equation on each subinterval using a stochastic Taylor expansion. Such expansions are described in detail in chapter 5 of [4]. The simplest such approximation, using only the linear term in the expansion, gives the Euler (also known as Euler-Maruyama) scheme

$$x_i^{(j+1)} = x_i^{(j)} + \sum_{k=1}^{d} b_{ik}(t_j, x^{(j)}) V_k^{(j)} \tag{3}$$

while adding the quadratic terms gives the Milstein scheme

$$x_i^{(j+1)} = x_i^{(j)} + \sum_{k=1}^{d} b_{ik}(t_j, x^{(j)}) V_k^{(j)} + \sum_{k,l=1}^{d} \rho_{ikl}(t_j, x^{(j)}) I_{kl}^{(j)} \tag{4}$$

where $V_k^{(j)} = W_k((j+1)h) - W_k(jh)$, $I_{kl}^{(j)} = \int_{jh}^{(j+1)h} \{W_k(t) - W_k(jh)\} dW_l(t)$ and $\rho_{ikl}(t, x) = \sum_{m=1}^{q} b_{mk}(t, x) \frac{\partial b_{il}}{\partial x_m}(t, x)$.

Assuming (2) the Euler scheme has order $\frac{1}{2}$, in the sense that $\mathbb{E}(\max_{j=1}^{N} |x^{(j)} - x(jh)|^2) = O(h)$ and under a stronger smoothness condition on the b_{ik} the Milstein scheme has order 1, indeed

$$\mathbb{E}(\max_{j=1}^{N} |x^{(j)} - x(jh)|^2) = O(h^2) \tag{5}$$

(see Kloeden and Platen [2], Section 10.3). These L^2 bounds can be extended to L^p for any $p \geq 1$.

The Euler scheme is straightforward to implement, as the only random variables one has to generate are the normally-distributed $V_k^{(j)}$, but for Milstein one has also to generate the 'area integrals' $I_{kl}^{(j)}$ which is non-trivial if $d \geq 2$. Order $\frac{1}{2}$ is the best one can do in general when the only random variables generated are the $V_k^{(j)}$.

We remark here that we can write $I_{kl}^{(j)} = \frac{1}{2}(V_k^{(j)} V_l^{(j)} - h\delta_{kl}) + \zeta_k^{(j)} V_l^{(j)} - \zeta_l^{(j)} V_k^{(j)} + K_{kl}^{(j)}$ with random variables $\zeta_k^{(j)}$, $K_{kl}^{(j)}$ for $1 \leq k, l \leq d$ all having zero mean, variance $\frac{h}{12}$, satisfying $K_{kl}^{(j)} = -K_{lk}^{(j)}$, and such that the $d(d+1)/2$ random variables consisting of $\zeta_k^{(j)} : 1 \leq k \leq d$ and $K_{kl}^{(j)} : 1 \leq k < l \leq d$ are mutually uncorrelated (though not independent).

Motivated by this remark we consider the following modification of Milstein, which requires the generation of normal random variables only.

$$\tilde{x}_i^{(j+1)} = \tilde{x}_i^{(j)} + \sum_{k=1}^{d} b_{ik}(t_j, \tilde{x}^{(j)}) \tilde{V}_k^{(j)} + \sum_{k,l=1}^{d} \rho_{ikl}(t_j, \tilde{x}^{(j)}) J_{kl}^{(j)} \tag{6}$$

where again the $\tilde{V}_k^{(j)}$ are independent $N(0, h)$ and $J_{kl}^{(j)} = \frac{1}{2}(\tilde{V}_k^{(j)} \tilde{V}_l^{(j)} - h\delta_{kl}) + z_k^{(j)} \tilde{V}_l^{(j)} - z_l^{(j)} \tilde{V}_k^{(j)} + \lambda_{kl}^{(j)}$, where the $z_k^{(j)}$ for $1 \leq k \leq d$ and $\lambda_{kl}^{(j)}$ for $1 \leq k < l \leq d$ are independent $N(0, \frac{h}{12})$, and then we set $\lambda_{lk}^{(j)} = -\lambda_{kl}^{(j)}$ for $k < l$ and $\lambda_{kk}^{(j)} = 0$.

Our main result is that, under suitable regularity conditions and a fairly mild nondegeneracy condition, the scheme (6) has order 1 under a suitable coupling. To formulate the nondegeneracy condition, we define for each $(t, x) \in [0, T] \times \mathbb{R}^d$ a linear mapping $L_{t,x} : \mathbb{R}^d \oplus S_d \to \mathbb{R}^q$ by $L_{t,x}(r, s)_i = \sum_{k=1}^{d} b_{ik}(t, x) r_k + \sum_{k,l=1}^{d} \rho_{ikl}(t, x) s_{kl}$ for $r \in \mathbb{R}^d$ and $s \in S_d$, where S_d is the space of skew-symmetric $d \times d$ matrices. We will require that L_x be surjective for each $(t,x) \in [0,T] \times \mathbb{R}^d$; this is equivalent to requiring that for each (t,x) the vectors $b^k(x)$ and $[b^k, b^l](x)$, for $1 \le k, l \le d$, will span $\mathbb{R}^q$ (here b^k is the vector whose ith component is b_{ik}, and $[b^k, b^l]$ denotes the Lie bracket, regarding b^k and b^l as vector fields on $\mathbb{R}^q$ for each t). This can be thought of as a strengthened Hörmander condition. In fact we need a version with some uniformity in (t,x), which we state precisely in the main theorem:

Theorem 1. *Suppose that the first and second derivatives of b_{ik} are bounded on $[0,T] \times \mathbb{R}^q$, and that there constants $\delta > 0$ and $K > 0$ such that for each $(t,x) \in [0,T] \times \mathbb{R}^q$ the image under $L_{t,x}$ of the unit ball in $\mathbb{R}^r \oplus S_d$ contains the ball of radius $\delta(1+|x|)^{-K}$ in $\mathbb{R}^q$.*

Then there is a constant $C > 0$ such that if $N \in \mathbb{N}$ is given we can find independent $N(0, \frac{h}{12})$ random variables $z_k^{(j)}$ for $1 \le k \le d$ and $\lambda_{kl}^{(j)}$ for $1 \le k < l \le d$ and $0 \le j < N$, defined on the same probability space as the Brownian path $W(t)$, such that, if $\tilde{x}^{(j)}$ is as given by the scheme (6), we have $\mathbb{E}|\tilde{x}^{(j)} - x(jh)|^2 \le Ch^2$ for $j = 1, \cdots, N$.

A similar result is proved in [2] for the scheme

$$x_i^{(j+1)} = x_i^{(j)} + \sum_{k=1}^{d} b_{ik}(t_j, x^{(j)}) V_k^{(j)} + \frac{1}{2} \sum_{k,l=1}^{d} \rho_{ikl}(t_j, x^{(j)})(V_k^{(j)} V_l^{(j)} - h\delta_{kl})$$

under a stronger nondegeneracy condition that the matrix (b_{ik}) has rank q.

The proof of theorem 1 occupies much of the remainder of the paper. We note here some properties of the joint characteristic function χ of the random variables (z_k), (λ_{kl}), regarded as a function on $\mathbb{R}^{d(d+1)/2}$. An explicit expression for χ can be found in [9]. What we require are the following (taking the case $h = 1$, from which the general case can be deduced by scaling): χ extends to be analytic on a 'strip' $\{x + iy : x, y \in \mathbb{R}^{d(d+1)/2}$ in $\mathbb{C}^{d(d+1)/2}$ and $|y| < \delta\}$ for some $\delta > 0$, and $|\chi(x+iy)| < C(1+|x|)^{-1}$ on this strip.

3 Coupling and KMT theory

If we have two probability spaces $(\mathcal{X}, \mathcal{F}, \mathbb{P})$ and $(\mathcal{Y}, \mathcal{G}, \mathbb{Q})$ then a *coupling* between $\mathbb{P}$ and $\mathbb{Q}$ is a measure on $\mathcal{X} \times \mathcal{Y}$ which has $\mathbb{P}$ and $\mathbb{Q}$ as its marginal distributions. Theorem 1 asserts the existence of a coupling between the probability space of the Brownian path and that of the random variables used in the approximation (6). We collect here some results on couplings which we shall need.

First we mention the Vaserstein metrics on probability measures on $\mathbb{R}^n$. If $\mathbb{P}_1$ and $\mathbb{P}_2$ are such measures, we define $\mathbb{W}_p(\mathbb{P}_1, \mathbb{P}_2)$ to be the infimum of $\mathbb{E}|X - Y|^p$, taken over all couplings between $\mathbb{P}_1$ and $\mathbb{P}_2$ where X and Y have distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ respectively. For $p \ge 1$ one can then show that $\mathbb{W}_p$ is a metric on the set of all probability measures $\mathbb{P}$ on $\mathbb{R}^n$ having finite pth moment (i.e. satisfying $\int_{\mathbb{R}^n} |x|^p d\mathbb{P}(x) < \infty$). $\mathbb{W}_p$ is known as the p-Vaserstein metric after [7]. (Note: we use the transliteration 'Vaserstein' from the Cyrillic as that is the one used by Vaserstein himself; 'Wasserstein' is also used in the literature).

We also note the elementary result (see e.g. proposition 7.10 in [8]) that

$$\mathbb{W}_p(\mu, \nu) \le 2^{(p-1)/p} \left\{ \int |x|^p d|\mu - \nu|(x) \right\}^{1/p} \tag{7}$$

for any two probability measures μ, ν on $\mathbb{R}^n$ and for any $p \ge 1$.

This is quite a good bound if $p = 1$ but is less good for $p > 1$; we shall however use it for bounding some small remainder terms.

The KMT theorem [5] is a form of simultaneous Central Limit Theorem using coupling. A variant of this result (modified from the original to be closer to the type of result we will use) states that if $\mathbb{P}$ is a suitably well-behaved probability distribution on $\mathbb{R}$, with zero mean, variance 1 and zero 3rd moment, then there is a constant $C > 0$ such that the following holds: if $n \in \mathbb{N}$ and $X_1, \cdots, X_n$ are independent with distribution $\mathbb{P}$, and if $Y_1, \cdots, Y_n$ are independent $N(0,1)$, then there is a coupling between the random vectors $(X_1, \cdots, X_n)$ and $(Y_1, \cdots, Y_n)$ such that

$$\mathbb{E}\left\{\sum_{i=1}^{k}(X_i - Y_i)\right\}^2 \leq C$$

for $k = 1, \cdots, n$.

There are various generalisations in the literature. Einmahl [3] extended the result to vector random variables and Zaitsev [10] further extended it to non-identical distributions which are uniformly non-degenerate. What we require is a variant of this latter result where the distributions are themselves random. It is not clear that this can be easily deduced from results in the literature so we prefer to give a self-contained argument in the context we need. This argument will use the lemma and corollary below, on polynomial perturbations of normal distributions. We denote by ϕ the standard normal $N(0, I)$ distribution on $\mathbb{R}^q$.

Lemma 2. *Let X be an $\mathbb{R}^q$-valued random variable with $N(0, I)$ distribution, let $p : \mathbb{R}^q \to \mathbb{R}^q$ be a polynomial function of degree 3, and define $\rho : \mathbb{R}^q \to \mathbb{R}^q$ by $\rho(x) = x + p(x)$. Let $\mathbb{P}$ be the probability distribution of $\rho(X)$ and let ν be the signed measure on $\mathbb{R}^q$ with density $\phi(y)(1 + y.p(y) - \nabla.p(y))$. Then for any $M \geq 1$ we have a bound*

$$\int_{\mathbb{R}^q} (1 + |y|)^M d|\mathbb{P} - \nu|(y) \leq C\epsilon^2 \tag{8}$$

where C is a positive constant depending only on q and M, and ϵ is an upper bound for the absolute values of the coefficients of p.

Proof. We use C_1, C_2 etc to denote positive constants which depend only on q and M. First we can find $C_1 \geq 1$ such that

$$\max(|\rho(x) - x|,\ \|D\rho(x) - I\|) \leq C_1\epsilon(1 + |x|)^3 \tag{9}$$

and

$$\max(|r(x)|,\ \|Dr(x)\|) \leq C_1\epsilon^2(1 + |x|)^9 \tag{10}$$

for all $x \in \mathbb{R}^q$, where $r(x) = p(x) - p(x + p(x))$. Then let $R = (2C_1\epsilon)^{-1/6} - 1$ and let $B_R = \{x \in \mathbb{R}^q : |x| < R\}$ (which will of course be empty if $R \leq 0$, which can happen if ϵ is not very small).

Now define a measure μ as the image under ρ of the restriction to B_R of the $N(0, I)$ distribution on $\mathbb{R}^q$. We also define $\tilde{\nu} = \nu|\rho(B_R)$. Then we have

$$\int_{\mathbb{R}^q} (1 + |y|)^M d|\mathbb{P} - \nu|(y) \leq \Omega_1 + \Omega_2 + \Omega_3$$

where $\Omega_1 = \int_{\mathbb{R}^q}(1 + |y|)^M d|\mu - \tilde{\nu}|(y)$, $\Omega_2 = \int_{\mathbb{R}^q}(1 + |y|)^M d(\mathbb{P} - \mu)(y)$ and $\Omega_3 = \int_{\mathbb{R}^q}(1 + |y|)^M d|\nu - \tilde{\nu}|(y)$.

We first bound Ω_1. To this end we note that, by the definition of R, for $x \in B_R$ the RHS of (9) is bounded by $\frac{1}{2}|\epsilon|^{1/2}$. It then follows from (9) that for $x \in B_R$ we have $\|D\rho(x) - I\| \leq \frac{1}{2}$ and so ρ is bijective on B_R. Then the density f of $\tilde{\nu}$ on $\rho(B_R)$ is given by $f(y) = \det D\rho^{-1}(y)\phi(\rho^{-1}(y))$ and so we have

$$\Omega_1 = \int_{\rho(B_R)} (1 + |y|)^M |\det D\rho^{-1}(y)\phi(\rho^{-1}(y)) - (1 + y.p(y) - \nabla.p(y))\phi(y)|dy$$

To bound the RHS, we fix $x \in B_R$ and set $y = \rho(x)$, noting that $|x-y| \le \min(1, |y|^{-1})$ by (9). Noting that $x = y - p(y) + r(x)$ and using the bound (10) we readily find that

$$|\phi(x) - (1 + y.p(y) - \nabla.p(y))\phi(y)| \le C_2\epsilon^2(1+|y|^{C_3})\phi(y)$$

and

$$|\det D\rho^{-1}(y) - (1 - \nabla.p(y))| \le C_2\epsilon^2(1+|y|^{C_3})$$

. From this we easily deduce that $\Omega_1 \le C_4\epsilon^2$.

Similar bounds for Ω_2 and Ω_3 are also easily proved, using the exponential decay of ϕ, and the result follows. □

We require a corollary of this lemma, for which we first introduce some notation.

Let P denote the space of all real-valued polynomials on $\mathbb{R}^q$, and P^q the space of $\mathbb{R}^q$-valued functions $p = (p_1, \cdots, p_q)$ on $\mathbb{R}^q$ such that each p_i is a polynomial. Let P_0 denote the subspace of $S \in P$ such that $\int_{\mathbb{R}^q} S(y)\phi(y)dy = 0$. We can characterise P_0 as follows. Let $\mathcal{L} : P^q \to P$ be the linear mapping defined by $\mathcal{L}p(x) = \nabla.p(x) - x.p(x)$. Then $\nabla.(\phi p)(x) = \mathcal{L}p(x)\phi(x)$ and it follows from the divergence theorem that $\mathcal{L}p \in P_0$ for every $p \in P^q$. In the converse direction, we note that if $u \in P$ has degree $n \ge 1$ then $\mathcal{L}\nabla u = -nu + r$ where $r \in P$ has degree less than n. If this u is in P_0 then we have $r \in P_0$ and by induction on n we can deduce that u is in the range of $\mathcal{L}$. So P_0 is precisely the range of $\mathcal{L}$.

Corollary 3. *Let $g \in P$ have degree 4, let μ be the measure with density ϕ (i.e. the standard normal probability measure) and let λ be a probability measure on $\mathbb{R}^q$ such that*

$$\int_{\mathbb{R}^q} (1+|y|)^M d|(1+g)\mu - \lambda|(y) \le \alpha \tag{11}$$

Then $\mathbb{W}_M(\mu, \lambda) \le C(\epsilon + (\epsilon^2 + \alpha)^{1/M})$, where C is a positive constant depending only on q and M, and ϵ is an upper bound for the absolute values of the coefficients of g.

Proof. Let $\beta = \int g d\mu$. Then (11) gives $|\beta| \le \alpha$, and $g - \beta \in P_0$. So by replacing g by $g - \beta$ we can assume $g \in P_0$.

Then as described above we can find $p \in P^q$ with $\mathcal{L}p = g$, and from the construction of p it is clear that p has degree 3 and its coefficients are bounded by $C_1\epsilon$. Let X be an $N(0, I)$ random variable and let $Y = X + p(X)$. By lemma 2 we have

$$\int_{\mathbb{R}^q} (1+|y|)^M d|g\mu - \nu|(y) \le C_2\epsilon^2$$

and so

$$\int_{\mathbb{R}^q} (1+|y|)^M d|\nu - \lambda|(y) \le C_2\epsilon^2 + \alpha$$

Hence by (7), $\mathbb{W}_M(\nu, \lambda) \le C_3(\epsilon^2 + \alpha)^{1/M}$. Finally $\mathbb{E}|Y - X|^M = \mathbb{E}|p(X)|^M \le C_4\epsilon^M$ so $\mathbb{W}_M(\mu, \nu) \le C_4^{1/M}\epsilon$ and the result follows by the triangle inequality. □

4 First reduction

For the Milstein approximation $(x_i^{(j)})$ given above, we know that $\mathbb{E}|x^{(j)} - x(jh)|^2 \le Ch^2$ holds under the assumptions of the theorem. So to prove the theorem it suffices to obtain a bound $\mathbb{E}|\tilde{x}^{(j)} - x^{(j)}|^2 \le Ch^2$. We will construct a coupling between the set of random variables $V_k^{(j)}$, $I_{kl}^{(j)}$ used for Milstein and the set of random variables used by (6), such that this bound holds.

We first split each of the random variables $V_k^{(j)}$ as the sum of two parts: $V_k^{(j)} = Q_k^{(j)} + R_k^{(j)}$ where $Q_k^{(j)} \sim N(0, h - h^2)$ and $R_k^{(j)} \sim N(0, h^2)$ are independent. (See the remarks following the proof of Theorem 1 for discussion of this splitting). Now let $(u_i^{(j)})$ be the modified Euler approximation defined by the recurrence relation

$$u_i^{(j+1)} = u_i^{(j)} + \sum_{k=1}^{d} b_{ik}(u^{(j)}) Q_k^{(j)} \tag{12}$$

with $u^{(0)} = x^{(0)}$. Then define the $q \times q$ matrix $A^{(j)}$ by $A_{il}^{(j)} = \sum_{k=1}^{d} \frac{\partial b_{ik}}{\partial x_l}(u^{(j)}) Q_k^{(j)}$, and a modified matrix by $\hat{A}^{(j)} = A^{(j)}$ if $\|A^{(j)}\| \le \frac{1}{2}$, and $\hat{A}^{(j)} = 0$ otherwise. We also define a modified version of $I_{kl}^{(j)}$ by replacing V by Q, namely $\overline{I}_{kl}^{(j)} = \frac{1}{2}(Q_k^{(j)} Q_l^{(j)} - (h - h^2)\delta_{kl}) + \zeta_k^{(j)} Q_l^{(j)} - \zeta_l^{(j)} Q_k^{(j)} + K_{kl}^{(j)}$. Then define $\alpha^{(j)}$ by the recurrence relation

$$\alpha_i^{(j+1)} = \{(I + \hat{A}^{(j)})\alpha^{(j)}\}_i + \sum_{k=1}^{d} b_{ik}(u^{(j)}) R_k^{(j)} + \sum_{k,l=1}^{d} \rho_{ikl}(u^{(j)}) \overline{I}_{kl}^{(j)} \tag{13}$$

with $\alpha^{(0)} = 0$. Next define $\beta^{(j)} = x^{(j)} - u^{(j)} - \alpha^{(j)} \in \mathbb{R}^q$ and note that then $\beta^{(0)} = 0$. We have $\beta_i^{(j+1)} - \beta^{(j)} = x_i^{(j+1)} - x_i^{(j)} - (u_i^{(j+1)} - u_i^{(j)}) - (\alpha_i^{(j+1)} - \alpha_i^{(j)})$ and using (4), (12) and (13) we find after some rearrangement that

$$\begin{aligned}
\beta_i^{(j+1)} - \beta_i^{(j)} = &\sum_{k,l=1}^{d} \frac{\partial b_{ik}}{\partial x_l}(u^{(j)}) \beta_l^{(j)} Q_k^{(j)} + \sum_{k=1}^{d} \left\{ b_{ik}(x^{(j)}) - b_{ik}(u^{(j)}) - \sum_{l=1}^{q} \frac{\partial b_{ik}}{\partial x_l}(u^{(j)})(x_l^{(j)} - u_l^{(j)}) \right\} Q_k^{(j)} \\
&+ \sum_{k=1}^{d} \{b_{ik}(x^{(j)}) - b_{ik}(u^{(j)})\} R_k^{(j)} + \sum_{k,l=1}^{d} (\rho_{ikl}(x^{(j)}) - \rho_{ikl}(u^{(j)})) I_{kl} \\
&+ \{(A^{(j)} - \hat{A}^{(j)})\alpha^{(j)}\}_i + \sum_{k,l=1}^{d} \rho_{ikl}(u^{(j)})(I_{kl}^{(j)} - \overline{I}_{kl}^{(j)})
\end{aligned} \tag{14}$$

We now bound the RHS of (14). First note that, conditional on the random variables $Q^{(i)}, R^{(i)}, \zeta^{(i)}, K^{(i)}$ for $i < j$, each of the 6 terms on the RHS has expectation 0. Also the first term has variance bounded by $C_1 \mathbb{E}|\beta^{(j)}|^2 h$. Next, we see that the scheme (12) has order $\frac{1}{2}$, being an Euler scheme with the random term scaled by $\sqrt{1-h} = 1 + O(h)$, so that $\mathbb{E}|x^{(j)} - u^{(j)}|^2 \le C_2 h$. Then we see that each of the other 3 terms on the RHS has variance bounded by $C_3 h^3$. Then we conclude from (14) that $\mathbb{E}|\beta^{(j+1)}|^2 \le (1 + C_1 h)\mathbb{E}|\beta^{(j)}|^2 + C_3 h^3$ and hence that

$$\mathbb{E}|x^{(j)} - u^{(j)} - \alpha^{(j)}|^2 = \mathbb{E}|\beta^{(j)}|^2 \le C_4 h^2 \tag{15}$$

for $j = 1, \cdots, N$.

We can do a similar analysis for $(x_i^{(j)})$ as defined by (6) using random variables $\tilde{V}_k^{(j)}$, $z_k^{(j)}$ and $\lambda_{kl}^{(j)}$ as above. We again write $\tilde{V}_k^{(j)} = \tilde{Q}_k^{(j)} + \tilde{R}_k^{(j)}$ with $\tilde{Q}_k^{(j)} \sim N(0, h - h^2)$ and $\tilde{R}_k^{(j)} \sim N(0, h^2)$. Our intention is to construct a coupling between the two sets of random variables so that they all all defined on the same probability space, on which we can compare the two approximations. Our coupling will satisfy $\tilde{Q}^{(j)} = Q^{(j)}$, so we will assume this from now on.

Then, using the same $\hat{A}^{(j)}$ and $u^{(j)}$ as above, we define $\tilde{\alpha}^{(j)}$ by the recurrence relation

$$\tilde{\alpha}_i^{(j+1)} = \{(I + \hat{A}^{(j)})\tilde{\alpha}^{(j)}\}_i + \sum_{k=1}^{d} b_{ik}(u^{(j)}) \tilde{R}_k^{(j)} + \sum_{k,l=1}^{d} \rho_{ikl}(u^{(j)}) \overline{J}_{kl}^{(j)} \tag{16}$$

with $\tilde{\alpha}^{(0)} = 0$, where $\overline{J}^{(j)}_{kl} = \frac{1}{2}(Q^{(j)}_k Q^{(j)}_l - (h-h^2)\delta_{kl}) + z^{(j)}_k Q^{(j)}_l - z^{(j)}_l Q^{(j)}_k + \lambda^{(j)}_{kl}$ and just as before we obtain a bound

$$\mathbb{E}|\tilde{x}^{(j)} - u^{(j)} - \tilde{\alpha}^{(j)}|^2 = \mathbb{E}|\tilde{\beta}^{(j)}|^2 \leq C_5 h^2 \tag{17}$$

From the bounds (15) and (17) we see that to prove the theorem it is enough to obtain a bound

$$\mathbb{E}|\alpha^{(j)} - \tilde{\alpha}^{(j)}|^2 \leq Ch^2 \tag{18}$$

We prove this in the next section.

As preparation we note some properties of the process $(u^{(j)})$. We let $\mathcal{G}$ denote the σ-algebra generated by $Q^{(0)}, \cdots, Q^{(N-1)}$, so that the $u^{(j)}$ and $\hat{A}^{(j)}$ are $\mathcal{G}$-measurable. As $u(j)$ is an Euler approximation to (1), with the random term scaled by $\sqrt{1-h}$, standard bounds apply and we have $\mathbb{E}|u^{(j)}|^p \leq C(p)$ for any $p \geq 1$. We also define $B^{(r)} = (I + \hat{A}^{(1)})^{-1} \cdots (I + \hat{A}^{(r)})^{-1}$ and we readily obtain $\mathbb{E}\|B^{(r)}\|^p \leq C(p)$ and $\mathbb{E}\|(B^{(r)})^{-1}\|^p \leq C(p)$.

5 Proof of theorem

Throughout we use C to denote a constant which may depend on the SDE but is independent of N; each occurrence may be different.

With $B^{(r)}$ as defined above we set

$$\gamma^{(r)} = B^{(r)} \left\{ \sum_{k=1}^{d} R^{(r)}_k b_k(u^{(r)}) + \sum_{k,l=1}^{d} \sigma_{kl}(u^{(r)}) Q^{(r)}_l \zeta^{(r)}_k + \sum_{1 \leq k < l \leq d} \sigma_{kl}(u^{(r)}) K^{(r)}_{kl} \right\},$$

$$\tilde{\gamma}^{(r)} = B^{(r)} \left\{ \sum_{k=1}^{d} \tilde{R}^{(r)}_k b_k(u^{(r)}) + \sum_{k,l=1}^{d} \sigma_{kl}(u^{(r)}) Q^{(r)}_l z^{(r)}_k + \sum_{1 \leq k < l \leq d} \sigma_{kl}(u^{(r)}) \lambda^{(r)}_{kl} \right\},$$

where $\sigma_{kl}(x)$ is the vector in $\mathbb{R}^q$ whose ith component is $\rho_{ikl}(x) - \rho_{ilk}(x)$, and we see that

$$\alpha^{(j)} - \tilde{\alpha}^{(j)} = (B^{(j)})^{-1} \sum_{r=0}^{j-1} (\gamma^{(r)} - \tilde{\gamma}^{(r)}) \tag{19}$$

It is convenient to reformulate the above expressions for $\gamma^{(r)}$ and $\tilde{\gamma}^{(r)}$ using random variables scaled to have variance 1. We let $m = d(d+3)/2$ and define random vectors $X^{(r)} = (X^{(r)}_1, \cdots, X^{(r)}_m)$ by $X^{(r)}_k = h^{-1} R^{(r)}_k$ for $k = 1, \cdots, d$; $X^{(r)}_k = (12/h)^{1/2} \zeta^{(r)}_{k-d}$ for $k = d+1, \cdots 2d$; $X_{(k+1)(d-k/2)+l} = 12^{1/2} h^{-1} K^{(r)}_{kl}$. Then (conditional on $\mathcal{G}$), $X^{(r)}$ has mean 0 and covariance matrix I. We can then write $h^{-1}\gamma^{(r)} = G_r X^{(r)}$ where G_r is a $q \times m$ matrix defined in terms of $B^{(r)}, b_k(u^{(r)}), \sigma_{kl}(u^{(r)}), Q^{(r)}$. In the same way we have $h^{-1}\tilde{\gamma}^{(r)} = G_r \tilde{X}^{(r)}$ where $\tilde{X}^{(r)}$ is $N(0, I)$.

We have inequalities

$$\|G_r\| \leq \|B^{(r)}\| \left(\sum_{k=1}^{d} |b^{(r)}_k (u^{(r)}| + \sum_{k,l=1}^{d} |\sigma_{kl}(u^{(r)}|(h^{-1/2}|Q^{(r)}_l| + 1) \right)$$

and $G_r G_r^t \geq B^{(r)} F(u^{(r)}) B^{(r)t}$ where $F(x) = \sum_{k=1}^{d} b_k(x) b_k(x)^t + \frac{1}{12} \sum_{k<l} \sigma_{kl}(x) \sigma_{kl}(x)^t$. We note that the nondegeneracy hypothesis in the theorem implies that $\|(F(x)F(x)^t)^{-1}\| \leq C(1+|x|)^{-2K}$. From these bounds and those at the end of the last section we deduce that $\mathbb{E}\|G_r\|^p \leq C(p)$ and $\mathbb{E}\|(G_r G_r^t)^{-1}\|^p \leq C(p)$ for all $p \geq 1$. We remark that, conditional on $\mathcal{G}$, $\gamma^{(r)}$ and $\tilde{\gamma}^{(r)}$ have the same covariance matrix $h^2 G_r G_r^t$.

From now on we assume for convenience that N is a power of 2, $N = 2^\kappa$ (this can always be arranged by extending the SDE to the interval $[0, 2^\kappa h]$ where κ is the smallest integer such that $2^\kappa \geq N$).

Let $E_0 = \{0, 1, \cdots, 2^\kappa - 1\}$. We call a subset E of E_0 dyadic if it is of the form $E = \{m2^n, m2^n + 1, \cdots, (m+1)2^n - 1\}$ for some $n \in \{0, 1, \cdots, \kappa\}$ and $m \in \{0, 1, \cdots, 2^{\kappa - n} - 1\}$. We see then that, for each n, the dyadic sets of size 2^n form a partition of E_0, and each dyadic set of size 2^{n+1} is the union of two dyadic sets of size 2^n. For each dyadic set E of size 2^n we then define $\gamma_E = \sum_{r \in E} \gamma^{(r)}$, $\tilde{\gamma}_E = \sum_{r \in E} \tilde{\gamma}^{(r)}$ and $H_E = 2^{-n} \sum_{r \in E} G_r G_r^t$. Note that since, conditional on $\mathcal{G}$, the random variables $\gamma^{(0)}, \cdots, \gamma^{(N-1)}$ are independent, H_E is the (conditional) covariance matrix of $Y_E := 2^{-n/2} h^{-1} \gamma_E$. The same applies to $\tilde{\gamma}_E$.

The idea is to construct couplings between $\tilde{\gamma}_E$ and γ_E recursively, starting with E_0 and proceeding by successive bisection. For this purpose we use the following lemma, which is a version of the Central Limit Theorem saying that the density of γ_E is close to the (Gaussian) density of $\tilde{\gamma}_E$.

Lemma 4. *Let E be a dyadic set of size 2^n, and let f_E be the density function of Y_E, conditional on $\mathcal{G}$. Fix η with $0 < \eta < \frac{1}{12}$. Then, provided $\|G_r\| < 2^{\eta n}$ and $\|(G_r G_r^t)^{-1}\| < 2^{2\eta n}$ for each $r \in E$, we have for $|v| < 2^{\eta n}$ that*

$$\left| \frac{f_E}{\tilde{f}_E}(v) - 1 - p_E(v) \right| < C 2^{(16\eta - 2)n}$$

where $\tilde{f}_E(v) = (2\pi)^{-q/2} (\det H_E)^{-1/2} \exp(-\frac{1}{2} v^t H_E^{-1} v)$ is the density of $\tilde{Y}_E$ and $p_E(v)$ is a 4th degree polynomial whose coefficients are bounded by $C 2^{(4\eta - 1)n}$.

Proof. Note first that the bounds on G_r imply $\|H_E\| \leq 2^{2n\eta}$ and $\|H_E^{-1}\| \leq 2^{2n\eta}$.

Let ψ be the characteristic function of the random variable $X^{(r)}$ (which is independent of r). ψ is real-valued and even on $\mathbb{R}^m$, and extends to a complex-analytic function on a 'strip' $\{x + iy : x, y \in \mathbb{R}^m, |y| < a\}$ for some $a > 0$. In a neighbourhood of 0 in $\mathbb{C}$, $\log \psi$ has a convergent expansion $\log \psi(z) = -\frac{1}{2}|z|^2 + c_4(z) + c_6(z) + \cdots$ where $c_k(z)$ is a homogeneous polynomial of degree k, and $|c_k(z)| \leq (C|z|)^k$ for even $k \geq 4$. Then $\psi(z) = \exp -\frac{z^t z}{2} + \chi(z)$ where $\chi(z) = c_4(z) + c_6(z) + \cdots$. From this it follows that there exists $\delta > 0$ such that

$$\text{if } x, y \in \mathbb{R} \text{ with } 2|y| \leq |x| < \delta \text{ then } |\psi(x + iy)| \leq e^{-|x|^2/6} \tag{20}$$

Then using the decay of ψ as $x \to \infty$ and the fact that $|\psi(x)| < 1$ for $x \in \mathbb{R}$ with $x \neq 0$, we can find γ with $0 < \gamma < 1$ and $\delta' > 0$ so that

$$\text{if } x, y \in \mathbb{R} \text{ with } |x| \geq \delta \text{ and } |y| \leq \delta' \text{ then } |\psi(x + iy)| \leq \min(\gamma, C|x|^{-1}) \tag{21}$$

Now let Ψ be the characteristic function of Y_E; then $\Psi(u) = \prod_{r \in E} \psi(2^{-n/2} G_r^t u)$ and $f_E(v) = (2\pi)^{-q/2} \int_{\mathbb{R}^q} e^{-iu^t v} \Psi(u) du$, which by translating the subspace of integration in $\mathbb{C}^q$ by $-iH^{-1}v$ we can write as

$$f_E(v) = (2\pi)^{-q} e^{-v^t H_E^{-1} v} \int_{\mathbb{R}^q} e^{-iu^t v} \Psi(u - iH_E^{-1} v) du \tag{22}$$

If $|u| \geq 2^{4\eta n + 1}$ we can write $\Psi(u - iH_E^{-1}v) = \prod_{r \in E} \psi(2^{-n/2} G_r^t u - i 2^{-n/2} G_r^t H_E^{-1} v)$. Now using (20) and (21) we see that each term in the product is bounded by either $\min(\gamma, (C 2^{n(\eta + 1/2)} |u|^{-1})$ or $\exp(-2^{-(1+2\eta)n} |u|^2/6)$, and we deduce that $|\Psi(u - iH_E^{-1}v)| \leq \min(\gamma, (C 2^{n(\eta+1/2)} |u|^{-1})^{2^n} + \exp(-2^{-2\eta n} |u|^2/6)$ for $|u| \geq 2^{4\eta n + 1}$. It then follows easily that

$$\int_{|u| \geq 2^{4\eta n + 1}} |\Psi(u - iH_E^{-1} v)| du \leq C \left\{ 2^{nm} \gamma^{2^n} + \exp(-2^{6\eta n - 1}) \right\} \tag{23}$$

To get a bound for $|u| \le 2^{4\eta n+1}$ we write $w = u - iH_E^{-1}v$ and note that $e^{-iu^tv}\Psi(w) = \exp(\frac{1}{2}v^tH_E^{-1}v - \frac{1}{2}u^tH_Eu + \Lambda(w)$ where $\Lambda(w) = \sum_{r\in E}\chi(2^{-n/2}G_r^tw) = \sum_{k=2}^{\infty} S_{2k}(w)$ where $S_{2k}(w) = 2^{-kn}\sum_{r\in E} c_{2k}(G_r^tw)$. We see that S_{2k} is a homogeneous polynomial of degree $2k$ and satisfies $|S_{2k}(w)| \le C2^{(1-k+2k\eta)n}|w|^{2k}$. We find that $|e^{\Lambda(w)} - 1 - S_4(w)| \le C2^{(8\eta-2)n}|w|^6$ and hence that

$$e^{-\frac{1}{2}v^tH_E^{-1}v}\int_{|u|\le 2^{4\eta n+1}} |e^{-iu^tv}\Psi(u-iH_E^{-1}v)-(1+S_4(u-iH_E^{-1}v))e^{-u^2H_Eu}|du \le C2^{(16\eta-2)n} \quad (24)$$

We also have $\int_{|u|\ge 2^{4\eta n+1}}|1+S_4(u-iH_E^{-1}v)|e^{-\frac{1}{2}u^tH_Eu}du \le Ce^{-2^{\eta n}}$ and combining these bounds the lemma follows, with $p_E(v) = \int_{\mathbb{R}^q} S_4(u - iH_E^{-1}v)e^{-\frac{1}{2}u^2H_Eu}du$ which is a polynomial of degree 4 whose coefficients are bounded by $C2^{(4\eta-1)n}$. □

Initial step. We start the construction by finding a coupling between $\tilde{Y}_{E_0}$ and Y_{E_0}. Let $\mathcal{E}_0$ be the event that condition (27) below holds with $E = E_0$. Then provided $\mathcal{E}_0$ holds, lemma 4 gives $|f_{E_0}(y)/\tilde{f}_{E_0}(y) - 1 - p_{E_0}(y)| < C2^{(16-2\eta)n}$ for $|y| \le 2^{n\eta/3}$. To apply Corollary 3 we write $y = H_{E_0}^{1/2}u$ and $g(u) = (\det H_{E_0})^{1/2}f_{E_0}(H_{E_0}^{1/2}U)$, and deduce that

$$\int_{\mathcal{A}}(1+|u|)^3\left|\left\{1+p_{E_0}\left(H_{E_0}^{1/2}u\right)\right\}\phi(u) - g(u)\right|du < C2^{(16\eta-2)n} \quad (25)$$

where $\mathcal{A} = \{u \in \mathbb{R}^q : |H_{E_0}^{1/2}u| < 2^{n\eta/3}\}$. One can easily see that the integral over $\mathcal{A}^c$ is bounded by $C2^{-2n}$ so that (25) holds with the integral over $\mathbb{R}^q$. And the polynomial $p_{E_0}(H^{1/2}u)$ has coefficients bounded by $C2^{(4\eta-1)n}$ so from Corollary 3 we have $\mathbb{W}_3(g,\phi) \le C2^{(16\eta-2)n/3}$. Then $\mathbb{W}_3(f_{E_0}, \tilde{f}_{E_0}) \le \|H_{E_0}^{1/2}\|\mathbb{W}_3(g,\phi) \le C2^{(17\eta-2)n/3}$. So can can find a coupling between Y_{E_0} and Y_E so that

$$\mathbb{E}|Y_{E_0} - \tilde{Y}_{E_0}|^3 \le C2^{(17\eta-2)n} \quad (26)$$

Recursive step. Let E be a dyadic set of size 2^n where $n \ge 1$. We can write E in a unique way as the union of two disjoint dyadic sets F and G of size 2^{n-1} and note that $Y_F + Y_G = 2^{1/2}Y_E$ and $\tilde{Y}_F + \tilde{Y}_G = 2^{1/2}\tilde{Y}_E$. We suppose a coupling between $\tilde{Y}_E$ and Y_E has been defined, conditional on $\mathcal{G}$. In other words, for each choice of $Q^{(0)}, \cdots, Q^{(N-1)}$, we have a joint distribution of Y_E and $\tilde{Y}_E$ with the correct conditional marginal distributions. We wish to extend this coupling to a coupling between (Y_F, Y_G) and $(\tilde{Y}_F, \tilde{Y}_G)$.

For each $x \in \mathbb{R}^q$, let f_x be the density of Y_F conditional on $Y_E = x$ and on $\mathcal{G}$, and let $\tilde{f}_x$ be the density of $\tilde{Y}_F$ conditional on $\tilde{Y}_E = x$ and on $\mathcal{G}$. We note that the conditional distribution of $\tilde{Y}_F$, given $\tilde{Y}_E = x$ and $\mathcal{G}$, is $N(Jx, H)$ where $J = H_FH_E^{-1}$ and $H = \frac{1}{2}H_FH_E^{-1}H_G$. So $\tilde{f}_x$ is the density function of $N(Jx, H)$.

We need to find a coupling between Y_F and $\tilde{Y}_F$, conditional on $Y_E = x$ and $\tilde{Y}_E = \tilde{x}$. To do this we need a coupling between the distributions with densities f_x and $\tilde{f}_{\tilde{x}}$. We shall in fact construct a coupling between f_x and $\tilde{f}_x$, then use the fact that $\tilde{f}_{\tilde{x}}$ is just $\tilde{f}_x$ translated by $J(\tilde{x} - x)$.

First we note that $f_x(y) = \frac{2^{1/2}f_F(y)f_G(2^{1/2}x-y)}{f_E(x)}$. Then the provided the condition

$$\|G_r\| < 2^{\eta n/6} \quad \text{and} \quad \|(G_rG_r^t)^{-1}\| < 2^{\eta n/3} \quad \text{for each } r \in E \quad (27)$$

holds, applying lemma 4 to each of E, F, G gives

$$\left|\frac{f_x(y)}{\tilde{f}_x(y)} - 1 - p_x(y)\right| < C2^{(16\eta-2)n} \quad (28)$$

for $|x|, |y| \le 2^{n\eta/3}$, where $p_x(y) = p_F(y) + p_G(2^{1/2}x - y) - p_E(x)$.

Let $\Omega = \{x \in \mathbb{R}^q : \mathbb{E}(|Y_F|^3 \chi_{|Y_F| \geq 2^{\eta n/3}} | Y_E = x \ \& \ \mathcal{G}) > 2^{-2n}\}$, and let $p = 60/\eta$. Then we see that, provided (27) holds,

$$\mathbb{P}(Y_E \in \Omega | \mathcal{G}) \leq 2^{2n} \mathbb{E}(|Y_F|^3 \chi_{|Y_F| \geq 2^{\eta n/3}} | \mathcal{G}) \leq 2^{-18n} \mathbb{E}(|Y_F|^{p+3} | \mathcal{G}) \leq C 2^{-18n} \|H_E\|^{(p+3)/2} \leq C 2^{-2n} \tag{29}$$

Let $\mathcal{E}$ denote the event that (27) holds, $|Y_E| \leq 2^{n\eta/6-1}$ and $Y_E \notin \Omega$. Write $x = Y_E$. In order to apply Corollary 3 to the conditional distribution of Y_F, we make the change of variable $y = Jx + H^{1/2}u$, noting that then $\tilde{f}_x(y) = (\det H)^{-1/2}\phi(u)$. We define $g_x(u) = (\det H)^{1/2} f_x(Jx + H^{1/2}u)$. Then, provided $\mathcal{E}$ holds, (28) gives

$$\int_{\mathcal{A}} (1 + |u|)^3 \left| \left\{ 1 + p_x(Jx + H^{1/2}u) \right\} \phi(u) - g_x(u) \right| du < C 2^{(16\eta - 2)n} \tag{30}$$

where $\mathcal{A} = \{u \in \mathbb{R}^q : |Jx + H^{1/2}u| < 2^{n\eta/3}\}$. Also, writing $y = Jx + H^{1/2}u$, if $|y| \geq 2^{n\eta/3}$ we have $|H^{1/2}u| \leq 2|y|$ so $|u| \leq 2^{1+n\eta/6}|y|$ and then, using $x \notin \Omega$, we find $\int_{\mathcal{A}^c} (1+|u|)^3 g_x(u) \leq 2^{(\eta-2)n}$. We also easily get $\int_{\mathcal{A}} (1+|u|)^3 |1 + p_x(Jx + H^{1/2}u)\phi(u)| du < C 2^{-2n}$. Putting these bounds together, we obtain

$$\int_{\mathbb{R}^q} (1 + |u|)^3 \left| \left\{ 1 + p_x(Jx + H^{1/2}u) \right\} \phi(u) - g_x(u) \right| du < C 2^{(16\eta - 2)n} \tag{31}$$

The polynomial $p_x(Jx + H^{1/2}u)$ has coefficients bounded by $C 2^{(5\eta-1)n}$ and then applying Corollary 3 we deduce that $\mathbb{W}_3(g_x, \phi) \leq C 2^{(16\eta-2)n/3}$. Then $\mathbb{W}_3(f_x, \tilde{f}_x) \leq \|H^{1/2}\| \mathbb{W}_3(g_x, \phi) \leq C 2^{(17\eta-2)n/3}$. In other words, conditional on $Y_E = x$ and assuming $\mathcal{E}$, we can find a random variable Y_F^* with density $\tilde{f}_x$ such that $\mathbb{E}|Y_F^* - Y_F|^3 \leq C 2^{(17\eta-2)n}$. If $\mathcal{E}$ fails then we find an arbitrary variable Y^* with density $\tilde{f}_x$. One easily finds that $\mathbb{P}(\mathcal{E}) \leq C 2^{-6n}$ and then taking expectation over $\mathcal{G}$ and Y_E we find that unconditionally

$$\mathbb{E}|Y_F^* - Y_F|^3 \leq C 2^{(17\eta-2)n} \tag{32}$$

We can now complete the recursive step by defining

$$\tilde{Y}_F = Y_F^* + H_F H_E^{-1}(\tilde{Y}_E - Y_E) \tag{33}$$

which has the correct conditional density $\tilde{f}_{\tilde{x}}$ with $\tilde{x} = \tilde{Y}_E$. Then we must have $\tilde{Y}_G = 2^{1/2}\tilde{Y}_E - \tilde{Y}_F$. It is natural to define $Y_G^* = 2^{1/2} Y_E - Y_F^*$; then one sees that (32) and (33) both hold with F replaced by G.

Conclusion of proof. We apply the recursive procedure as described above, starting with E_0 (initial step), then using the recursive step to proceed from dyadic sets of size 2^n to dyadic sets of size 2^{n-1}, to generate a coupling for every dyadic set. The result can be summarised as follows: conditional on $\mathcal{G}$ we have constructed a coupling between the sets of random variables (Y_E) and $(\tilde{Y}_E)$, each ranging over the dyadic sets E, such that (32) and (33) hold whenever F is a dyadic set of size 2^{n-1} contained in a dyadic set E of size 2^n.

Now consider a given dyadic set E of size 2^n. We can write in a unique way $E = E_k \subseteq E_{k-1} \subseteq \cdots \subseteq E_0$ where $k = \kappa - n$ and, for each j, E_j is a dyadic set of size $2^{\kappa - j}$. Then from (33) we obtain $\tilde{Y}_E - Y_E = \sum_{j=1}^k H_{E_k} H_{E_j}^{-1}(Y_{E_j}^* - Y_{E_j}) + H_{E_k} H_{E_0}^{-1}(\tilde{Y}_{E_0}) - Y_{E_0})$. From this, using (26) and (32) along with the L^p bounds for $H_{E_j}^{-1}$ and H_{E_k}, and using Hölder's inequality, we obtain $\|\tilde{Y}_E - Y_E\|_{5/2} \leq C 2^{3(17\eta-2)n/10}$. Thus

$$\|\tilde{\gamma}_E - \gamma_E\|_{5/2} \leq C 2^{(51\eta-1)n/10} h \tag{34}$$

holds whenever E is a dyadic set of size 2^n. We now apply this to (19). If $1 \leq j \leq N$ we can write $\{0, 1, \cdots, j-1\}$ as a union of dyadic sets $E_1 \cup \cdots \cup E_k$ where $E_1, \cdots, E_k$ have different sizes. Then (19) gives $\alpha^{(j)} - \tilde{\alpha}^{(j)} = (B^{(j)})^{-1} \sum_{i=1}^{k} (\tilde{\gamma}_{E_i} - \gamma_{E_i})$. Provided $\eta < \frac{1}{51}$, (34) then gives (18) using Hölder's inequality and an L^{10} bound for $(B^{(j)})^{-1}$. This completes the proof of the theorem.

Remarks. A natural question is whether the theorem is true without the nondegeneracy condition. Without this condition the KMT-type argument faces considerable technical difficulties, but I would conjecture that the theorem is still true.

The splitting $V^{(j)} = Q^{(j)} + R^{(j)}$ is introduced in order to allow the vectors b^k as well as the Lie brackets be included in the nondegeneracy condition. If we have the nondegeneracy condition with the brackets only (i.e. the term in r_k is omitted from the definition of $L_{t,x}$) then we can prove the theorem without this splitting - but this condition is considerable stronger (e.g. it cannot hold if $q = d = 2$).

Note that our result is slightly weaker than than the bound (5) for Milstein which has a max over j. (5) is deduced from the bound for individual j using Doob's martingale inequality; however we cannot apply this to our scheme because our coupling does not preserve the filtrations, so the error $\tilde{x}^{(j)} - x(jh)$ is not a martingale. The following example shows that the analogue of (5) fails for scheme (6), whatever coupling is used.

Example. We consider the SDE with $q = 3$ and $d = 2$ given by

$$dx_1 = dW_1, \quad dx_2 = dW_2, \quad dx_3 = x_1 dW_2 - x_2 dW_1$$

on the time interval $[0, 1]$, with initial condition $x_i(0) = 0$.

This SDE has solution $x_1 = W_1$, $x_2 = W_2$, $x_3(t) = \int_0^t (W_1(s)dW_2(s) - W_2(s)dW_1(s))$. We find that $\rho_{312}(x) = 1$, $\rho_{321}(x) = -1$ and all other ρ_{ikl} are zero. It is then easy to check that the hypotheses of Theorem 1 are satisfied. We also note that the Milstein approximation is exact, in that $x^{(j)} = x(jh)$ for each j.

We claim that there is a constant $c > 0$ such that, for any $N \in \mathbb{N}$ the approximation using scheme (6) with $h = \frac{1}{N}$ and any coupling between the random variables $\tilde{V}_k^{(j)}, z_k^{(j)}, \lambda_{12}^{(j)}$ used by (6) and the Brownian path W, we have

$$\mathbb{P}(\max_{0 \le j < N} |\tilde{x}^{(j)} - x(jh)| \ge cN^{-1} \log N) > \frac{1}{2} \tag{35}$$

To prove this claim we first note that

$$x_3^{(j+1)} - x_3^{(j)} = x_1^{(j)} V_2^{(j)} - x_2^{(j)} V_1^{(j)} + I_{12}^{(j)} - I_{21}^{(j)} \tag{36}$$

and

$$\tilde{x}_3^{(j+1)} - \tilde{x}_3^{(j)} = \tilde{x}_1^{(j)} \tilde{V}_2^{(j)} - \tilde{x}_2^{(j)} \tilde{V}_1^{(j)} + 2(z_1^{(j)} \tilde{V}_2^{(j)} - z_2^{(j)} \tilde{V}_1^{(j)} + \lambda_{12}^{(j)}) \tag{37}$$

We also define random variables $M = \max_{0 \le j < N} |\tilde{x}^{(j)} - x^{(j)}|$, $K = \max_{1 \le j \le N} |W(jh)|$ and $\tilde{K} = \max_{1 \le j \le N} |(\tilde{x}_1^{(j)}, \tilde{x}_2^{(j)})|$. And we set $X^j = h^{-1}(I_{12}^{(j)} - I_{21}^{(j)})$, $Y^{(j)} = \frac{2}{h}(z_1^{(j)} \tilde{V}_2^{(j)} - z_2^{(j)} \tilde{V}_1^{(j)})$ and $Z^{(j)} = \frac{2}{h} \lambda_{12}^{(j)}$. Then subtracting (36) from (37) and using the above definitions we find that

$$h|X^{(j)} - Y^{(j)} - Z^{(j)}| \le 2M(1 + 2K + 2\tilde{K}) \tag{38}$$

The idea is to use (38) to get a lower bound for M. For this we need the distributions of the random variables on the LHS of (38). First note that, from the known distribution of the Lévy area, $X^{(j)}$ has density $\frac{1}{2}\text{sech}(\pi x/2)$ so $\mathbb{P}(|X^{(j)}| \ge \lambda) \ge C_1 e^{-\pi\lambda/2}$ for $\lambda > 0$. And $Y^{(j)}$ can be expressed as $\frac{1}{2\sqrt{3}}(P^2 - Q^2 + R^2 - S^2)$ where P, Q, R, S are independent $N(0,1)$, so that $P^2 + R^2$ and $Q^2 + S^2$ have exponential distributions, and then a simple calculation shows that $Y^{(j)}$ has a symmetric exponential distribution with $\mathbb{P}(|Y^{(j)}| > \lambda) = e^{-\sqrt{3}\lambda}$. Moreover $Z^{(j)}$ has $N(0, \frac{1}{3})$ distribution, from which one finds easily that $\mathbb{P}(|Y^{(j)} + Z^{(j)}| > \lambda) \le C_2 e^{-5\lambda/3}$ (using $\frac{5}{3} < \sqrt{3}$).

Now fix α and β with $\frac{3}{5} < \beta < \alpha < \frac{2}{\pi}$. Then we have

$$\mathbb{P}(\max_{0 \le j < N} |X^{(j)}| \le \alpha \log N) \le (1 - C_1 e^{-\pi\alpha \log N/2})^N \le \exp(C_1 N^{1-\frac{\pi\alpha}{2}})$$

and

$$\mathbb{P}(\max_{0 \le j < N} |Y^{(j)} + Z^{(j)}| \ge \beta \log N) \le C_2 N^{1-\frac{3\beta}{5}}$$

So if N is large enough we have $\mathbb{P}(\max_{0 \le j < N} |X^{(j)}| \le \alpha \log N) \le \frac{1}{8}$ and $\mathbb{P}(\max_{0 \le j < N} |Y^{(j)} + Z^{(j)}| \ge \beta \log N) \le \frac{1}{8}$. Moreover we can find a constant C_3 so that $\mathbb{P}(K \ge C_3) \le \frac{1}{8}$ and $\mathbb{P}(\tilde{K} \ge C_3) \le \frac{1}{8}$. Then, with probability at least $\frac{1}{2}$, we have

$$\max_{0 \le j < N} |X^{(j)}| \ge \alpha \log N, \quad \max_{0 \le j < N} |Y^{(j)} + Z^{(j)}| \le \beta \log N, \quad K \le C_3, \text{ and } \tilde{K} \le C_3 \tag{39}$$

Finally, using (38), (39) implies $2M(1 + 4C_3) \ge (\alpha - \beta) h \log N$, giving the required result.

References

[1] A. Alfonsi, B. Jourdain and A. Kohatsu-Higa, Pathwise optimal transport bounds between a one-dimensional diffusion and its Euler scheme, *Ann. Appl. Probab.* **24** (2014), 1049-1080.

[2] A. M. Davie, Pathwise approximation of stochastic differential equations using coupling, preprint.

[3] U. Einmahl, Extensions of results of Komlós, Major and Tusnády to the multivariate case, *J. Multivariate Anal.* **28** (1989), 20-68.

[4] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag 1995.

[5] J. Komlós, P. Major and G. Tusnády, An approximation of partial sums of independent RV's and the sample DF. I, *Z. Wahr. und Wer. Gebiete* **32** (1975), 111-131.

[6] S. T. Rachev and L. Ruschendorff, *Mass Transportation Problems*, Volume 1, Theory; Volume 2, Applications. Springer-Verlag 1998.

[7] L. N. Vaserstein, Markov processes over denumerable products of spaces describing large system of automata, *Problemy Peredaci Informacii* **5** (1969), 64-72 (Russian). Translation in *Problems of Information Transmission* **5** (1969), 47-52.

[8] C. Villani, *Topics in Optimal Transportation*, AMS 2003.

[9] M. Wiktorsson, Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions, *Ann. Appl. Probab.* **11** (2001), 470-487.

[10] A. Zaitsev, Multidimensional version of a result of Sakhanenko on the invariance principle for vectors with finite exponential moments. I,II,III *Teor. Veroyatnost. i Primenen*, **45** (2000), 718-738; **46** (2001), 535-561 and 744-769. Translations in *Theory Probab. Appl.* **45** (2002), 624-641; **46** (2003), 490-514 and 676-698.

School of Mathematics
University of Edinburgh