SIAM J. SCI. COMPUT.
Vol. 34, No. 1, pp. A137–A160

THE DIRECT RICHARDSON pTH ORDER (DRp) SCHEMES: A NEW CLASS OF TIME INTEGRATION SCHEMES FOR STOCHASTIC DIFFERENTIAL EQUATIONS*

PAVEL P. POPOV† AND STEPHEN B. POPE†

Abstract. We describe a new family of weak pth order accurate SDE time integration schemes, called the direct Richardson pth order accurate (DRp) schemes. The DRp schemes use the idea of Richardson extrapolation on Euler time steps, performed by way of an acceptance-rejection algorithm. Previous applications of Richardson extrapolation to the Euler scheme are applicable only when the objective is to estimate a functional of the final distribution of the process. In contrast, provided that the diffusion matrix is strictly positive definite, the DRp class of schemes can be used in all applications which require a weak SDE time integration scheme. Numerical results have been obtained, and a comparison is made between the second- and third-order accurate DRp schemes and other modern SDE time integration schemes, indicating that the DRp schemes incur less error than standard algorithms based on Ito–Taylor expansions, and have similar computational efficiency. Finally, we provide a proof of the convergence properties of the DRp schemes.

Key words. stochastic differential equation, weak approximation, Richardson extrapolation, rejection sampling

AMS subject classifications. 60H35, 65C30

DOI. 10.1137/100801445

1. Introduction.

In this work, we introduce a new class of weak pth order accurate numerical schemes for the integration of the n-dimensional nonhomogeneous and anisotropic Ito stochastic differential equation (SDE):

$$\mathrm{d}\mathbf{X} = \mathbf{D}\left(\mathbf{X}, t\right) \mathrm{d}t + \sigma\left(\mathbf{X}, t\right) \mathrm{d}\mathbf{W}, \tag{1}$$

where $\mathbf{X}(t)$ is the random process governed by the Ito SDE, $\mathbf{D}\left(\mathbf{x}, t\right)$ denotes the drift, which is an n-dimensional vector field, $\sigma\left(\mathbf{x}, t\right)$ is a strictly positive definite $n \times n$ matrix field, and $\mathrm{d}\mathbf{W}$ as usual indicates that the SDE is driven by the standard n-dimensional Wiener process. In this form and generality, the Ito SDE has numerous applications in science and engineering [4]: for example, its solution is an essential component of particle-based numerical schemes for turbulent combustion [3], which is the authors' particular interest.

Given the wide range of applicability of the Ito SDE, it is not surprising that over the years numerous methods have been developed for its numerical approximation, in both the strong and weak senses. Strong SDE integration schemes aim to accurately reconstruct the trajectory $\mathbf{X}\left(t\right)$ as a function of the Wiener sample path $\mathbf{W}\left(t\right)$. In contrast, weak schemes only need to satisfy the condition that the distribution of the numerical solution approximates that of the actual SDE solution. Here, we concentrate on numerical schemes which exhibit weak convergence.

One of the most widespread numerical schemes for the solution of nonstiff SDEs in Ito form is the forward Euler scheme. Using Δt to denote the length of a time step,

*Submitted to the journal's Methods and Algorithms for Scientific Computing section July 8, 2010; accepted for publication (in revised form) October 7, 2011; published electronically January 31, 2012. This work was supported by the Air Force Office of Scientific Research through grant FA 9550-09-1-0047.

http://www.siam.org/journals/sisc/34-1/80144.html

†Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14853 (ppp7@cornell.edu, s.b.pope@cornell.edu).

and η to denote a sample vector from the n-dimensional standard normal distribution $\mathcal{N}(\mathbf{0}, I)$, a single time step of the Euler scheme has the form

$$\mathbf{Z}^E(t+\Delta t) = \mathbf{Z}^E(t) + \mathbf{D}\left(\mathbf{Z}^E(t), t\right)\Delta t + \sigma\left(\mathbf{Z}^E(t), t\right)\eta\sqrt{\Delta t}, \tag{2}$$

and it is well known [4] that the Euler scheme is strong 0.5th order accurate and weak first-order accurate.

There are also numerous higher-order weak SDE integration schemes in existence [7, 4, 3, 12]. To the authors' knowledge, all of these are based on approximating higher-order terms in the Ito–Taylor expansion of $\mathbf{X}(t)$. One way to perform this approximation is through the explicit evaluation of derivatives of the fields $\mathbf{D}(\mathbf{x}, t)$ and $\sigma(\mathbf{x}, t)$ at the initial point (for an example, see the second-order weak Taylor scheme presented in [4]). Alternatively, the fields can be sampled at additional points, thereby implicitly obtaining the required derivatives through finite differences (for an example, see the two families of weak second-order derivative-free schemes developed in [12]).

Here, we present a different approach—to approximate each time step increment through Euler steps of varying length, and then employ Richardson extrapolation via an acceptance-rejection procedure in order to obtain a higher-order scheme. This may sound somewhat familiar to the reader, inasmuch as Richardson extrapolation on Euler solutions is a well-known [9] method for obtaining higher-order estimates of functionals $E(g(\mathbf{X}(T)))$ of the solution at the end time T. Additionally, the recently developed multilevel Monte Carlo method [5, 6] uses an approximation based on Euler solutions with varying time steps to minimize the computational cost for approximating functionals of $\mathbf{X}(t)$ over the entire interval $t \in [0, T]$. Furthermore, the acceptance-rejection methodology has previously been used in SDE time-stepping algorithms, for example, the method introduced in [2], for the purposes of obtaining an integrator which is ergodic with respect to the SDE equilibrium distribution.

The significant difference in the class of schemes that we present is that the Richardson extrapolation is applied at each time step directly to the PDF of the random variable that approximates the SDE solution. This yields a family of SDE integration schemes, called the direct Richardson pth order (DRp) schemes, which are pth order accurate at each time step, and not just at the end of the simulation. This high-order accuracy at each time step is essential when the solution to the SDE is just one part of a more complex simulation, such as, for example, in a Lagrangian Monte Carlo solution for turbulent reactive flows.

To elaborate upon this issue, we note that in such a solution [13], the chemical composition in the reaction domain is represented by an ensemble of particles whose temperature and density affects the velocity and diffusivity of the flow. The solution algorithm alternates between an SDE time step, which advances the particle locations using the current velocity and diffusivity fields, and a finite-volume time step, which solves the Navier–Stokes and turbulence modeling equations, using density and temperature fields based on the new particle locations.

Thus, while the methods described in [9, 5, 6] are computationally more efficient for the purpose of evaluating expectations of functionals of an a priori known SDE, they cannot be employed in an application of mathematical physics such as the one outlined in the above paragraph, because in such an application the coefficients of (1) at a given time t are not known until the particles' positions have been advanced to time t. For this reason, we compare the performance of the DRp class of schemes with that of SDE time integrators such as those of Kloeden and Platen [4] and the

Cao and Pope [3], which can also be used in an application where the SDE coefficients are known only locally in time, i.e., only for the duration of the current finite volume time step.

In this work, we present the general form of the DRp family of SDE integration schemes and provide a proof of the weak pth order accuracy of its members, subject to certain smoothness criteria on the SDE drift and diffusion terms. We also provide results from numerical test cases which compare the performance of the two simplest DRp schemes—DR2 and DR3—with that of other modern weak second-order accurate SDE integration schemes.

The organization of the rest of this paper is as follows: In section 2, we present the simplest member of the DRp family, DR2, and comment on its implementation in a computational code. In section 3, we present results from numerical test cases which compare the performance of DR2 and DR3 with that of two modern weak second-order schemes. The numerical results indicate that, for the purpose of approximating Ito SDEs with strictly positive definite diffusion, the DRp schemes are at least as efficient as current SDE integration schemes.

Section 4 introduces Richardson extrapolation and the way in which it is employed in the DRp schemes. Section 5 introduces the regularity and boundedness criteria which need to be satisfied by the fields $\mathbf{D}(\mathbf{x},t)$ and $\sigma(\mathbf{x},t)$ in order to achieve weak pth order accuracy. Section 6 introduces the framework and notation for the general form of the DRp schemes, and section 7 gives a pseudocode description of the general DRp scheme. Section 8 contains a proof of the weak pth order accuracy of the DRp schemes, and section 9 provides a summary. Finally, Appendices A and B contain the proofs of two theorems which are used in section 8.

The mathematical developments in the second half of this paper are essential, as they prove the convergence of the schemes proposed by the authors. Nevertheless, it is appreciated that some readers are primarily interested in employing a DRp scheme for a particular application, and may wish to skip the more theoretical aspects of this work. Such readers are advised to read sections 2, 3, 5, as well as section 4 up to and including (20), and section 7 up to and including (41).

2. Description of the second-order member of the DRp family: DR2.

Below, we give a pseudocode description of a single time step of length Δt of the weak second-order accurate DR2 scheme. Without loss of generality, we assume that the initial position is $\mathbf{X}(t=0)=\mathbf{0}$. Also, for the sake of compactness of notation, we shall use $B(\mathbf{x},t)$ to denote $\sigma(\mathbf{x},t)\,\sigma^T(\mathbf{x},t)$.

1. Obtain two independent samples $\eta_1,\eta_2 \sim \mathcal{N}(\mathbf{0},I)$ from the standard normal distribution.
2. Set

$$\mathbf{U}_1^{(2)} \equiv \mathbf{0} + \mathbf{D}(\mathbf{0},0)\,\frac{\Delta t}{2} + \sigma(\mathbf{0},0)\,\eta_1\sqrt{\frac{\Delta t}{2}},$$

$$\mathbf{U}_2^{(2)} \equiv \mathbf{U}_1^{(2)} + \mathbf{D}\left(\mathbf{U}_1^{(2)},\frac{\Delta t}{2}\right)\frac{\Delta t}{2} + \sigma\left(\mathbf{U}_1^{(2)},\frac{\Delta t}{2}\right)\eta_2\sqrt{\frac{\Delta t}{2}}. \tag{3}$$

3. Compute

$$\mathbf{v}_1 \equiv \mathbf{U}_2^{(2)} - \mathbf{U}_1^{(2)} - \mathbf{D}(\mathbf{0},0)\,\frac{\Delta t}{2},$$

$$\mathbf{v}_2 \equiv \mathbf{U}_2^{(2)} - \mathbf{U}_1^{(2)} - \mathbf{D}\left(\mathbf{U}_1^{(2)},\frac{\Delta t}{2}\right)\frac{\Delta t}{2},$$

$$f_1 \equiv \frac{1}{\left|B\left(\mathbf{0},0\right)\right|^{1/2}} \exp\left(-\frac{1}{\Delta t}\mathbf{v}_1^T \left(B\left(\mathbf{0},0\right)\right)^{-1}\mathbf{v}_1\right),$$

$$f_2 \equiv \frac{1}{\left|B\left(\mathbf{U}_1^{(2)},\frac{\Delta t}{2}\right)\right|^{1/2}} \exp\left(-\frac{1}{\Delta t}\mathbf{v}_2^T \left(B\left(\mathbf{U}_1^{(2)},\frac{\Delta t}{2}\right)\right)^{-1}\mathbf{v}_2\right),$$

$$H \equiv 1 - \frac{f_1}{2f_2}. \tag{4}$$

4. Sample a one-dimensional (1D) random variable ξ from the standard uniform distribution: $\xi \sim U(0,1)$.
5. If the acceptance criterion

$$\xi < \max\left(H, 0.1\right) \tag{5}$$

is met, then set $\mathbf{Z}_{\Delta t}^{DR2} = \mathbf{U}_2^{(2)}$. Else, go back to step 1.

The above algorithm is easily implemented in a computational code: each acceptance-rejection step involves two evaluations of the drift and diffusion fields, along with the computations involved in calculating H, which are trivial for the case of isotropic diffusion, and can be optimized to involve just one determinant evaluation and one solution of an $n \times n$ linear system for the anisotropic case.

Taking a more abstract view on the DR2 algorithm, we can see that it is basically a rejection sampling algorithm which takes two Euler time steps (3) at each sampling step, and accepts with probability $\max\left(H, 0.1\right)$. The reader is referred to [1] for an accessible introduction to rejection sampling methods, also known as acceptance-rejection methods. The essential idea of acceptance-rejection methods is that we can generate a random variable with a desired distribution (the target distribution) by obtaining a sample from a simpler distribution on the same sample space (the instrumental distribution) and accepting that sample with a probability based on the ratio of the PDFs associated with the target and instrumental distributions.

One might wonder, then, where exactly is the Richardson extrapolation after which this scheme is named. To answer this question, let us define $f_{\Delta t}^2\left(\mathbf{x},\mathbf{y}\right)$ to be the PDF of the random variable $\left(\mathbf{U}_1^{(2)},\mathbf{U}_2^{(2)}\right)$, in other words the PDF associated with the event $\left\{\mathbf{U}_1^{(2)}=\mathbf{x},\mathbf{U}_2^{(2)}=\mathbf{y}\right\}$. Also, let $\eta_3 \sim \mathcal{N}\left(\mathbf{0},I\right)$, define $\overline{\mathbf{U}_2^{(2)}}$ by

$$\overline{\mathbf{U}_2^{(2)}} \equiv \mathbf{U}_1^{(2)} + \mathbf{D}\left(\mathbf{0},0\right)\frac{\Delta t}{2} + \sigma\left(\mathbf{0},0\right)\eta_3\sqrt{\frac{\Delta t}{2}}, \tag{6}$$

and let us define $f_{\Delta t}^1\left(\mathbf{x},\mathbf{y}\right)$ to be the PDF of the random variable $\left(\mathbf{U}_1^{(2)},\overline{\mathbf{U}_2^{(2)}}\right)$, in other words, the PDF associated with the event $\left\{\mathbf{U}_1^{(2)}=\mathbf{x},\overline{\mathbf{U}_2^{(2)}}=\mathbf{y}\right\}$. It can then be shown that

$$\frac{f_{\Delta t}^1\left(\mathbf{U}_1^{(2)},\mathbf{U}_2^{(2)}\right)}{f_{\Delta t}^2\left(\mathbf{U}_1^{(2)},\mathbf{U}_2^{(2)}\right)} = \frac{f_1}{f_2}, \tag{7}$$

and so, since $f_{\Delta t}^2\left(\mathbf{x},\mathbf{y}\right)$ is the instrumental distribution of the rejection sampling, we arrive at the conclusion that the PDF of the pair $\left(\mathbf{U}_1^{(2)},\mathbf{U}_2^{(2)}\right)$ which gets accepted by the DR2 algorithm approximates $2f_{\Delta t}^2\left(\mathbf{x},\mathbf{y}\right)\left(1-\frac{f_{\Delta t}^1(\mathbf{x},\mathbf{y})}{2f_{\Delta t}^2(\mathbf{x},\mathbf{y})}\right) = 2f_{\Delta t}^2\left(\mathbf{x},\mathbf{y}\right) - f_{\Delta t}^1\left(\mathbf{x},\mathbf{y}\right)$.

For the above result to hold, however, we need to show that the probability of H deviating significantly from $\frac{1}{2}$ is negligible; we shall elaborate on this claim, and demonstrate its correctness, later on.

Note, however, that (6) implies that $\overline{\mathbf{U}_2^{(2)}}$ has the same distribution as the end value after a single Euler step of length Δt, whereas by its definition $\mathbf{U}_2^{(2)}$ is the end value after two Euler steps of length $\frac{\Delta t}{2}$ each. Therefore, the PDF of $\mathbf{Z}_{\Delta t}^{DR2}$ approximates the second-order Richardson extrapolation of the PDFs of the final value at $t = \Delta t$, as given by one and two Euler steps.

And so we see that the DR2 scheme applies Richardson extrapolation directly to the PDF of the process's value after a single time step. This makes it weakly second-order accurate, as demonstrated by the numerical results presented in the next section.

3. Numerical test cases: Comparing the accuracy and efficiency of DR2 and DR3 with that of other weakly convergent SDE integration schemes. In this section, we provide results from numerical test cases which compare the accuracy and computational cost of the DR2 and DR3 schemes with those of two other modern SDE time integration schemes. The first of these is the weak second-order midpoint scheme, developed by Cao and Pope [3], which we refer to as the Cao and Pope scheme, or CP for short. In the authors' experience [8], the Cao and Pope SDE integration scheme is both accurate and computationally efficient. However, having been designed specifically for its application in Lagrangian Monte Carlo methods for turbulent combustion, the CP scheme can only treat cases with isotropic diffusion. We compare its performance with that of DR3 on a 1D, and hence isotropic, test case.

We also make a comparison with the multidimensional explicit second-order weak scheme described by Kloeden and Platen [4, pp. 486–487] and generalized to a family of derivative-free weak second-order schemes by Tocino and Vigo-Aguiar [12] (the particular member of the family of schemes used here is referred to as SIE-A in [12]). As the Kloeden and Platen (KP) scheme allows for anisotropic diffusion, we compare its performance with that of DR2 on a two-dimensional (2D) anisotropic test case.

Two dimensional, anisotropic test case: Comparison between the DR2 and KP schemes. We perform a simulation on the domain $\mathbf{x} = (x, y) \in [0, 2\pi) \times [0, 2\pi)$ with periodic boundary conditions, from $t = 0$ to $t = 1$. We specify an analytic solution with the functional form

$$f(x,y,t) = \sum_{k,l,m=0}^{3} R_{klm}^{1,f} \sin(xk + yl + \pi tm) + R_{klm}^{2,f} \cos(xk + yl + \pi tm) \tag{8}$$

for the PDF of the process $\mathbf{X}(t)$. The same functional form is used for the coefficients $(\sigma_{ij})_{i,j=1}^{2}$ of the diffusion matrix σ:

$$\sigma_{ij}(x,y,t) = \sum_{k,l,m=1}^{3} R_{klm}^{1,\sigma_{ij}} \sin(xk + yl + \pi tm) + R_{klm}^{2,\sigma_{ij}} \cos(xk + yl + \pi tm). \tag{9}$$

At the beginning of the simulation, the coefficients $R_{klm}^{1,f}, R_{klm}^{2,f}, R_{klm}^{1,\sigma_{ij}}, R_{klm}^{2,\sigma_{ij}}$ are assigned randomly from a standard normal distribution. Then all of the coefficients are rescaled and a constant offset is added, in order to enforce

$$\min_{x,y,t}(f(x,y,t)) = \frac{1}{8\pi^2}, \qquad \max_{x,y,t}(f(x,y,t)) = \frac{3}{8\pi^2},$$

$$\min_{x,y,t}(\sigma_{ij}(x,y,t)) = 0.6, \qquad \max_{x,y,t}(\sigma_{ij}(x,y,t)) = 1.4 \text{ for } i = j,$$
$$\min_{x,y,t}(\sigma_{ij}(x,y,t)) = -0.3, \qquad \max_{x,y,t}(\sigma_{ij}(x,y,t)) = 0.3 \text{ for } i \neq j, \tag{10}$$

thus ensuring that $f(x,y,t)$ is positive and integrates to 1, and that $\sigma(x,y,t)$ is positive definite. We then specify $\mathbf{D}(x,y,t)$ such that the Fokker–Planck equation is satisfied. Denoting by $\tilde{f}(x,y,t)$ the PDF of a given numerical approximation $\mathbf{Z}(t)$ to $\mathbf{X}(t)$, and denoting by $\tilde{A}^{1,f}_{kl}, \tilde{A}^{2,f}_{kl}$ the coefficients of the Fourier expansion of $\tilde{f}(x,y,t=1)$,

$$\tilde{f}(x,y,t=1) = \sum_{k,l\in(-\infty,+\infty)} \tilde{A}^{1,f}_{kl} \sin(xk+yl) + \tilde{A}^{2,f}_{kl} \cos(xk+yl), \tag{11}$$

with a similar definition for the Fourier coefficients $A^{1,f}_{kl}, A^{2,f}_{kl}$ of $f(x,y,t=1)$ we estimate the following measure of error between $f(x,y,t=1)$ and $\tilde{f}(x,y,t=1)$:

$$\epsilon_f = \sqrt{\sum_{k,l=0}^{4} \left(\tilde{A}^{1,f}_{kl} - A^{1,f}_{kl}\right)^2 + \left(\tilde{A}^{2,f}_{kl} - A^{2,f}_{kl}\right)^2}. \tag{12}$$

Note that ϵ_f can be estimated only stochastically from the sample PDF of $\tilde{f}(x,y,t = 1)$. We use a sufficient number of samples of $\mathbf{Z}(t=1)$ to ensure that the 95% confidence interval for ϵ_f has a width smaller than the sample mean for ϵ_f, and we employ a jackknife estimator to reduce bias. In addition to estimating the error, we also measure the computational cost of each numerical scheme, in terms of microseconds per sample computation on a single processor. The machine used was a medium-sized cluster of 35 3.0GHz quad-core Xeon processors, and the numerical test cases were implemented in Fortran 90, using the Intel 10.032 compiler.

The results for this test case are given in Table 1, and in graphical form in Figures 1 and 2. In Figure 1, which is a log-log plot of error vs. time step, we see that both the DR2 and KP data points fall close to a line of slope two, which confirms the second-order accuracy of both schemes. In Table 1, it can be seen that DR2 has a slightly higher computational cost—it takes about 20% more time for the same time step size—than the KP scheme. This, however, is offset by the lower error produced by the DR2 scheme—only a third of the KP error—to yield a numerical method which is overall computationally more efficient for this test case. This can be seen in Figure 2, which is a log-log plot of error vs. computational cost. On this plot, the DR2 data points are closer to the lower left corner of the plot, which indicates that a given level of error can be achieved at lower computational cost by the DR2 scheme.

Table 1
Summary of accuracy and computational cost of the DR2 and KP schemes for the 2D anisotropic test case.

DR2	Δt	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$
	ϵ_f	1.04e-3	2.91e-4	9.06e-5	2.75e-5	7.67e-6
	95% CI half-width	2.31e-4	5.65e-5	1.34e-5	3.10e-6	7.80e-7
	μs/sample	14.56	27.16	50.65	100.38	195.92
KP	Δt	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	
	ϵ_f	2.58e-3	1.01e-3	3.09e-4	8.49e-5	
	95% CI half-width	2.90e-4	8.55e-5	1.60e-5	4.80e-6	
	μs/sample	13.04	22.98	43.01	84.92	

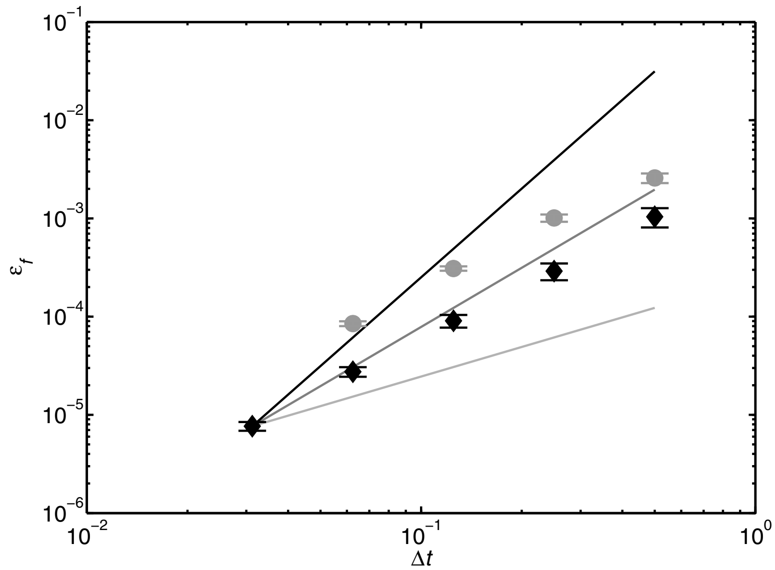

FIG. 1. *Numerical results comparing the accuracy of the DR2 scheme (black diamonds, horizontal bars denote 95% confidence intervals) with that of the KP scheme (gray circles), for the 2D, anisotropic test case. The light gray, dark gray, and black sloped reference lines illustrate, respectively, first-, second-, and third-order convergence.*

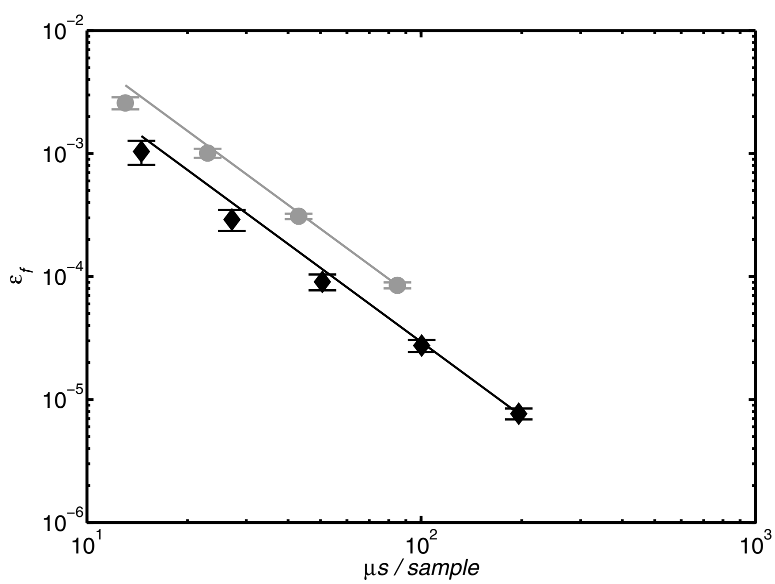

FIG. 2. *Numerical results comparing the computational efficiency of the DR2 scheme (black diamonds, horizontal bars denote 95% confidence intervals) with that of the KP scheme (gray circles), for the 2D, anisotropic test case. The black sloped reference line illustrates the second-order convergence behavior of DR2. The gray sloped reference line illustrates the second-order convergence behavior of KP.*

Table 2
Summary of accuracy and computational cost of the DR3 and CP schemes for the 1D isotropic test case.

DR3	Δt	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	
	ϵ_f	2.08e-3	5.62e-4	1.10e-4	1.53e-5	
	95% CI half-width	6.65e-4	8.85e-5	1.05e-5	1.19e-5	
	μs/sample	7.81	15.60	30.69	60.85	
CP	Δt	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$
	ϵ_f	9.07e-3	2.37e-3	7.32e-4	1.69e-4	4.28e-5
	95% CI half-width	1.40e-3	3.35e-4	8.77e-5	2.13e-5	5.34e-6
	μs/sample	3.22	6.27	11.82	23.61	47.03

One-dimensional, isotropic test case: Comparison between the DR3 and CP schemes. We have chosen to test DR3 (which is defined in section 7) on a 1D test case due to the prohibitive computational cost associated with the stochastic verification of the convergence properties of an SDE integration scheme with a high order of accuracy. For example, for a third-order scheme, halving the time step requires 2^7 times more computational effort in order to obtain a reasonable confidence interval.

The methodology is similar to that of the 2D test case described above. In particular, we set

$$f(x,t) = 1 + 0.3\sin(x - \pi t),$$
$$B(x,t) = 1 + 0.25\sin(x + \pi/3), \tag{13}$$

with $D(x,t)$ such that the Fokker–Planck equation is satisfied. We simulate on the periodic domain $x \in [0, 2\pi)$, from $t = 0$ to $t = 1$, and we define the error as

$$\epsilon_f = \sqrt{\sum_{k=0}^{6} \left(\tilde{A}_k^{1,f} - A_k^{1,f}\right)^2 + \left(\tilde{A}_k^{2,f} - A_k^{2,f}\right)^2}, \tag{14}$$

where again $\tilde{A}_k^{1,f}, \tilde{A}_k^{2,f}$ and $A_k^{1,f}, A_k^{2,f}$ are the Fourier coefficients of $\tilde{f}(x, t = 1)$ and $f(x, t = 1)$, respectively.

The results are presented in Table 2, and in graphical form in Figures 3 and 4. In Figure 3, which is a log-log plot of error vs. time step, it can be seen that the CP data points fall close to a line of slope two, confirming the second-order convergence of this scheme, whereas the DR3 scheme achieves third-order convergence for time steps lower than $\Delta t = 1/8$. Also, as can be seen in Table 2, for all time steps the error produced by DR3 is at least 4 times smaller than that produced by CP. On the other hand, the computational cost of DR3 is about 2.5 times higher for a given time step than that of Cao and Pope's scheme. In Figure 4, which is a log-log plot of error vs. computational cost, it can be seen that at high error levels, CP is more computationally efficient (its data points lie to the left of those of DR3), whereas at low error levels DR3 is more efficient, as it has attained its higher order of convergence.

Based on these numerical test cases, we establish the practical significance of the DR2 and DR3 schemes, which (for fixed Δt) produce less error than other modern SDE integration schemes, have comparable computational efficiency, and can be implemented with ease in a computational code, as we saw from the pseudocode description of DR2.

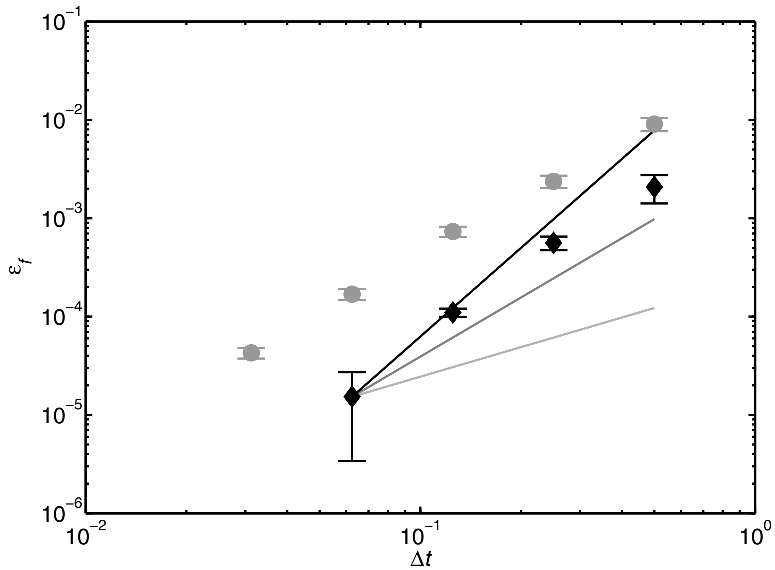

FIG. 3. *Numerical results comparing the accuracy of the DR3 scheme (black diamonds, horizontal bars denote 95% confidence intervals) with that of the CP scheme (gray circles), for the 1D, isotropic test case. The light gray, dark gray, and black sloped reference lines illustrate, respectively, first-, second-, and third-order convergence.*

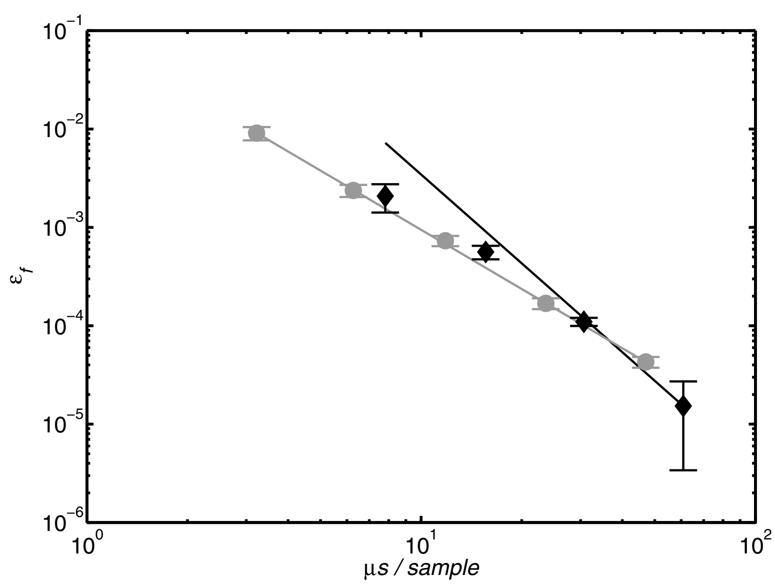

FIG. 4. *Numerical results comparing the computational efficiency of the DR3 scheme (black diamonds, horizontal bars denote 95% confidence intervals) with that of the CP scheme (gray circles), for the 1D, isotropic test case. The black sloped reference line illustrates the third-order convergence behavior of DR3 for small time steps. The gray sloped reference line illustrates the second-order convergence behavior of CP.*

Finally, we note that as the DRp class of SDE integration schemes are based on an acceptance-rejection procedure on a fixed number of explicit Euler time steps, its stability properties are the same as that of the explicit Euler scheme. While this may prevent DRp from being an appropriate SDE integrator for certain applications, it does not pose a problem in the context of turbulent reactive flow simulations and related applications in mathematical physics, in which the time steps used are small, for the sake of time resolution of the turbulent flow, and for stability of the finite volume solver. Indeed, explicit schemes such as those developed in [12, 3] and even explicit Euler are the SDE integrators most often used in this field.

In the subsequent sections, we present the mathematical theory of the general DRp scheme and prove its properties. We start this with a brief description of Richardson extrapolation and the manner in which it is used by the DRp family of schemes in order to achieve weak pth order accuracy.

4. Richardson extrapolation and its use by the DRp scheme.

Richardson extrapolation, as introduced in [10], introduces the elegant idea that, if we have a first-order accurate numerical approximation $A^1_{\Delta t}$ to an exact solution A, and the error with respect to some linear functional $g(\cdot)$ varies smoothly,

$$g(A^1_{\Delta t} - A) = \sum_{i=1}^{\infty} K_i \Delta t^i, \tag{15}$$

then we can construct from $A^1_{\Delta t}$ a second-order accurate approximation $A^2_{\Delta t}$ to A by setting $A^2_{\Delta t} = 2A^1_{\Delta t/2} - A^1_{\Delta t}$. Then (15) gives us that

$$\begin{aligned} g(A^2_{\Delta t} - A) &= 2g(A^1_{\Delta t/2} - A) - g(A^1_{\Delta t} - A) \\ &= \sum_{i=1}^{\infty} K_i (2(\Delta t/2)^i - \Delta t^i) \\ &= \sum_{i=2}^{\infty} K_i (2(\Delta t/2)^i - \Delta t^i), \end{aligned} \tag{16}$$

and so we see that the first-order component of the error has vanished. Applying this method inductively, we can obtain a scheme with an arbitrarily high order of accuracy from the first-order scheme $A^1_{\Delta t}$ by the linear combination

$$A^p_{\Delta t} = \sum_{k=1}^{p} l^p_k A^1_{\frac{\Delta t}{2^{k-1}}}, \tag{17}$$

where the coefficients l^p_k satisfy the following recursive relation:

$$\begin{aligned} \left(l^2_1, l^2_2\right) &= (-1, 2), \\ \left(l^p_1, \ldots, l^p_p\right) &= \frac{\left[2^{p-1}\left(0, l^{p-1}_1, \ldots, l^{p-1}_{p-1}\right) - \left(l^{p-1}_1, \ldots, l^{p-1}_{p-1}, 0\right)\right]}{2^{p-1} - 1}. \end{aligned} \tag{18}$$

In the context of solutions to SDEs [4, 9], Richardson extrapolation has been used to obtain pth order accurate estimates for the expected value of a function of the SDE solution at the end time, $E\left(g\left(\mathbf{X}(T)\right)\right)$. This can done by computing, for each time step $\Delta t, \Delta t/2, \ldots$, Euler solutions with that time step, which we shall denote by $\mathbf{Z}^E_{\Delta t}(T), \mathbf{Z}^E_{\Delta t/2}(T), \ldots$, and approximating $E(g(\mathbf{X}(T)))$ by the expected value of the Richardson extrapolate of $g\left(\mathbf{Z}^E_{\Delta t}(T)\right)$:

$$E(g(\mathbf{X}(T))) \approx E\left(\sum_{k=1}^{p} l_k^p g\left(\mathbf{Z}^E_{\frac{\Delta t}{2^{k-1}}}(T)\right)\right). \tag{19}$$

The approximation given by (19), while elegant and effective, is applicable only if we are interested in a functional of the solution at the end time, T. On the other hand, in many applications it is necessary to use an SDE integration procedure which gives an accurate result at each intermediate time step, due to the fact that the SDE is coupled to another process. As an example, in the implementation of a Monte Carlo method for turbulent combustion [13], an overall time step may consist of a transport substep (in which an SDE of the form of (1) is solved), followed by a reaction substep and a diffusion substep. As the last of these substeps uses the values provided by the first, it is easily seen that the transport substep needs to employ an SDE integration scheme which is accurate at intermediate times as well.

With this in mind, we adopt an alternative way of performing Richardson extrapolation on the Euler SDE solutions. Without loss of generality, let $\mathbf{X}(t=0) = \mathbf{0}$, let, for $k = 0, \ldots, p-1$, $f^k_{\Delta t}(\mathbf{x})$ be the probability density functions (PDFs) of the random variables $\mathbf{Z}^E_{\frac{\Delta t}{2^k}}(\Delta t)$, respectively, and let $f_{\mathbf{X}(\Delta t)}(\mathbf{x})$ be the PDF of the random variable $\mathbf{X}(\Delta t)$ which we are approximating numerically. Furthermore, let us denote by $\mathbf{Z}^{DRp}_{\Delta t}$ the random variable which is the DRp solution after a single time step of length Δt, and let $f^{DRp}_{\Delta t}(\mathbf{x})$ be its PDF.

Following [12], we note that a sufficient condition for the weak pth order accuracy of the DRp scheme is that it satisfies, for any multi-index of nonnegative integers $(i_1, i_2, \ldots, i_n)$ with $\sum_{m=1}^n i_m \leq 2p+2$, the inequality

$$\left| E\left(\prod_{m=1}^{n}\left(Z^{DRp}_{\Delta t,m}\right)^{i_m} - \prod_{m=1}^{n}\left(X_m(\Delta t)\right)^{i_m}\right)\right| \leq \tilde{C}\Delta t^{p+1}, \tag{20}$$

where $\tilde{C}$ is a positive constant which depends only on n, p and the fields $\mathbf{D}(\mathbf{x},t), \sigma(\mathbf{x},t)$. Previously [9, 11] it has also been demonstrated that for fields $\mathbf{D}(\mathbf{x},t), \sigma(\mathbf{x},t)$ which are sufficiently smooth, there exist constants $C^E_1, C^E_2, \ldots, C^E_{p+1}$ such that

$$\left| E\left(\prod_{m=1}^{n}\left(Z^E_{\Delta t,m}\right)^{i_m} - \prod_{m=1}^{n}\left(X_m(\Delta t)\right)^{i_m}\right) - \sum_{m=1}^{p} C^E_m \Delta t^m\right| \leq \tilde{C}^E_{p+1}\Delta t^{p+1}, \tag{21}$$

and hence

$$\left| E\left(\sum_{k=1}^{p} l_k^p \prod_{m=1}^{n}\left(Z^E_{\frac{\Delta t}{2^{k-1}},m}\right)^{i_m} - \prod_{m=1}^{n}\left(X_m(\Delta t)\right)^{i_m}\right)\right| \leq \tilde{C}^E_{p+1}\Delta t^{p+1}. \tag{22}$$

We have designed the DRp solution so that $f^{DRp}_{\Delta t}(\mathbf{x})$ satisfies

$$\int\left| f^{DRp}_{\Delta t}(\mathbf{x}) - \sum_{k=0}^{p-1} l^p_{k+1} f^k_{\Delta t}(\mathbf{x})\right| |\mathrm{d}\mathbf{x}| \leq C\Delta t^{p+1} \tag{23}$$

for a given finite constant C, independent of Δt. It is proved in Theorem 2 in Appendix B that (23) implies that

$$\left| E\left(\prod_{m=1}^{n} \left(Z_{\Delta t,m}^{DRp} \right)^{i_m} - \sum_{k=1}^{p} l_k^p \prod_{m=1}^{n} \left(Z_{\frac{\Delta t}{2^{k-1}},m}^{E} \right)^{i_m} \right) \right| \leq C' \Delta t^{p+1}, \tag{24}$$

where C' is another constant which depends only on n, p and the fields $\mathbf{D}(\mathbf{x}, t), \sigma(\mathbf{x}, t)$. Since (20) follows directly from (22), (24), we see that (23) and Theorem 2 imply that the DRp scheme is weak pth order accurate.

In the remainder of this paper, we give a general description of the DRp schemes and a proof of (23). First, however, we need to specify smoothness and boundedness criteria on the SDE drift and diffusion fields, which are necessary for the correct operation of the DRp schemes.

5. Smoothness requirements of the DRp scheme. As previously mentioned, we are computing, in $\mathbb{R}^n \times [0, T]$, weak solutions to the SDE problem as given by (1). We require that the fields $\mathbf{D}, \sigma$ be smooth (all derivatives exist and are bounded), so that the result of (21), derived in [11], holds true. Furthermore, for the correct operation of the DRp family of numerical schemes, we require that the drift vector field, $\mathbf{D}(\mathbf{x}, t) \in R^n$, be bounded,

$$\| \mathbf{D}(\mathbf{x}, t) \| \leq C_1^D, \tag{25}$$

and globally Lipschitz continuous in both space and time,

$$\| \mathbf{D}(\mathbf{x_1}, t_1) - \mathbf{D}(\mathbf{x_2}, t_2) \| \leq C_2^{D,\mathbf{x}} \| \mathbf{x}_1 - \mathbf{x}_2 \| + C_2^{D,t} |t_1 - t_2|, \tag{26}$$

and that the diffusion field of matrices, $\sigma(\mathbf{x}, t) \in \mathrm{Mat}(n, n)$, be globally bounded,

$$\| \sigma(\mathbf{x}, t) \mathbf{v} \| \leq C_1^\sigma \| \mathbf{v} \| \quad \text{for any } \mathbf{v} \in R^n, \tag{27}$$

as well as being globally coercive,

$$\| \sigma(\mathbf{x}, t) \mathbf{v} \| \geq C_2^\sigma \| \mathbf{v} \| \quad \text{for any } \mathbf{v} \in R^n, \tag{28}$$

and globally Lipschitz continuous with respect to the matrix norm,

$$\| \sigma(\mathbf{x_1}, t_1) - \sigma(\mathbf{x_2}, t_2) \| \leq C_3^{\sigma,\mathbf{x}} \| \mathbf{x}_1 - \mathbf{x}_2 \| + C_3^{\sigma,t} |t_1 - t_2|. \tag{29}$$

Here it is understood that constants $C_1^D, C_2^{D,\mathbf{x}}, C_2^{D,t}, C_1^\sigma, C_2^\sigma, C_3^{\sigma,\mathbf{x}}, C_3^{\sigma,t}$ are finite and strictly positive. Note also that (27)–(29) imply similar regularity conditions on σ^{-1}, the matrix inverse of σ:

$$\left\| \sigma^{-1}(\mathbf{x}, t) \mathbf{v} \right\| \leq C_1^{\sigma^{-1}} \| \mathbf{v} \| \quad \text{for any } \mathbf{v} \in R^n, \tag{30}$$

$$\left\| \sigma^{-1}(\mathbf{x}, t) \mathbf{v} \right\| \geq C_2^{\sigma^{-1}} \| \mathbf{v} \| \quad \text{for any } \mathbf{v} \in R^n, \tag{31}$$

and

$$\| \sigma^{-1}(\mathbf{x_1}, t_1) - \sigma^{-1}(\mathbf{x_2}, t_2) \| \leq C_3^{\sigma^{-1},\mathbf{x}} \| \mathbf{x}_1 - \mathbf{x}_2 \| + C_3^{\sigma^{-1},t} |t_1 - t_2|, \tag{32}$$

where $C_1^{\sigma^{-1}} = \frac{1}{C_2^\sigma}$, $C_2^{\sigma^{-1}} = \frac{1}{C_1^\sigma}$, $C_3^{\sigma^{-1},\mathbf{x}} = \frac{C_3^{\sigma,\mathbf{x}}}{(C_2^\sigma)^2}$, and $C_3^{\sigma^{-1},t} = \frac{C_3^{\sigma,t}}{(C_2^\sigma)^2}$ are again finite and strictly positive constants. For the sake of compactness of notation, we shall call C_1^D, $C_2^{D,\mathbf{x}}$, $C_2^{D,t}$, C_1^σ, C_2^σ, $C_3^{\sigma,\mathbf{x}}$, $C_3^{\sigma,t}$, $C_1^{\sigma^{-1}}$, $C_2^{\sigma^{-1}}$, $C_3^{\sigma^{-1},\mathbf{x}}$, $C_3^{\sigma^{-1},t}$ the Richardson regularity constants, and we shall denote them collectively as $\{C\}$.

We note that in the application of Lagrangian Monte Carlo turbulent reactive flow simulations, the above regularity conditions hold for simulations on well-resolved finite-volume grids and in the absence of compressible shocks (i.e., a subsonic simulation).

6. Framework of the DRp scheme. Here we describe the random variable $\mathbf{Z}_{\Delta t}^{DRP}$ and demonstrate why its PDF satisfies (23). Due to the fact that the Richardson extrapolation vectors l_k^p have negative components, (23) requires that the PDF of $\mathbf{Z}_{\Delta t}^{DRP}$ approximate a nonconvex linear combination of other PDFs, which are in themselves easily sampled from. To achieve this goal, we use an acceptance-rejection approach, with $f_{\Delta t}^{p-1}(\mathbf{x})$ as the instrumental distribution.

An additional concept which we need to achieve this is that of the 2^{p-1}-step sample path which corresponds to the random variables $\mathbf{Z}^E_{\frac{\Delta t}{2^k}}$. This concept embodies the idea that if we do not update the coefficients for the second step, two Euler steps of length $\Delta t/2$ produce exactly the same result as a single Euler step of length Δt, and hence we can use time steps of length $\frac{\Delta t}{2^{p-1}}$ to sample from each of $\mathbf{Z}^E_{\frac{\Delta t}{2^k}}$ by updating the values of $\mathbf{D}(\mathbf{x},t), \sigma(\mathbf{x},t)$ only when we reach a time which is an integer multiple of $\frac{\Delta t}{2^k}$, instead of at each time step.

More concretely, let us denote $\lfloor i \rfloor_{p,k} \equiv \left\lfloor \frac{i}{2^{p-1-k}} \right\rfloor 2^{p-1-k}$, let $\eta_i \sim \mathcal{N}(\mathbf{0}, I)$ be independent samples from the standard normal distribution, and define the random variable $\mathbf{U}^{(k)}$, for each $k = 0, 1, \ldots, p-1$ as $\mathbf{U}^{(k)} = \left(\mathbf{U}_i^{(k)}\right)_{i=1}^{2^{p-1}}$, where $\mathbf{U}_i^{(k)}$ are defined by

$$
\begin{aligned}
\mathbf{U}_0^{(k)} &= \mathbf{0},\\
\mathbf{U}_i^{(k)} &= \mathbf{U}_{i-1}^{(k)} + \mathbf{D}\left(\mathbf{U}^{(k)}_{\lfloor i-1 \rfloor_{p,k}}, \lfloor i-1 \rfloor_{p,k} \frac{\Delta t}{2^{p-1}}\right) \frac{\Delta t}{2^{p-1}}\\
&\quad + \sigma\left(\mathbf{U}^{(k)}_{\lfloor i-1 \rfloor_{p,k}}, \lfloor i-1 \rfloor_{p,k} \frac{\Delta t}{2^{p-1}}\right) \eta_i \sqrt{\frac{\Delta t}{2^{p-1}}}.
\end{aligned}
\tag{33}
$$

Finally, let $f_{\Delta t}^{U,k}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}})$ be the probability density function of $\mathbf{U}^{(k)}$, i.e., the PDF of the event $\left\{\mathbf{U}_i^{(k)} = \mathbf{x}_i\right\}$. This variable, $\mathbf{U}^{(k)}$ is what we will call the 2^{p-1}-step sample path generated by the 2^k-step Euler scheme. Note that it is an $n2^{p-1}$-dimensional random variable, and, correspondingly, its PDF is defined on an $n2^{p-1}$-dimensional sample space.

Noting that in the definition of $\mathbf{U}^{(k)}$, the coefficients $\mathbf{D}, \sigma$ are updated only when the number of the time step is an integer multiple of 2^{p-1-k}, we have that

$$
\begin{aligned}
\mathbf{U}^{(k)}_{i2^{p-1-k}} &= \mathbf{U}^{(k)}_{(i-1)2^{p-1-k}} + \mathbf{D}\left(\mathbf{U}^{(k)}_{(i-1)2^{p-1-k}}, (i-1)\frac{\Delta t}{2^k}\right)\frac{\Delta t}{2^k}\\
&\quad + \sigma\left(\mathbf{U}^{(k)}_{(i-1)2^{p-1-k}}, (i-1)\frac{\Delta t}{2^k}\right)\left[\sum_{j=(i-1)2^{p-1-k}+1}^{i2^{p-1-k}} \eta_j \sqrt{\frac{\Delta t}{2^{p-1}}}\right],
\end{aligned}
\tag{34}
$$

and so, since the vectors η_j are independent standard normal random variables, we have that the term in the square brackets in (34) is distributed as $N\left(0, \frac{\Delta t}{2^k}\right)$, which implies that the random variables $\left\{\mathbf{U}^{(k)}_{i2^{p-1-k}} | \mathbf{U}^{(k)}_{(i-1)2^{p-1-k}} = \mathbf{x}\right\}$ and $\left\{\mathbf{Z}_i^k | \mathbf{Z}_{i-1}^k = \mathbf{x}\right\}$

are identically distributed; in other words, the 2^{p-1-k} steps corresponding to (34) are identical to a singe Euler step of length $\frac{\Delta t}{2^k}$.

The above paragraph implies that, for each $k = 0, 1, \ldots, p-1$, the PDF $f^k_{\Delta t}(\mathbf{x})$ is the marginal PDF of $f^{U,k}_{\Delta t}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}})$ in the last variable, and so we can achieve the goal of (23) by approximating a linear combination of the PDFs of the 2^{p-1}-step sample paths.

The reason for this approach is the rather unexpected result that if we compute a realization of $\mathbf{U}^{(p-1)}$, we can also compute exactly the ratio $\frac{f^{U,k}_{\Delta t}(\mathbf{U}^{(p-1)})}{f^{U,p-1}_{\Delta t}(\mathbf{U}^{(p-1)})}$. To see how, define $f^{U,k}_{\Delta t, U_j}(\mathbf{x}_j; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{j-1})$ to be the PDF of $\mathbf{U}^{(k)}_j$, conditional upon $\{\mathbf{U}^{(k)}_1 = \mathbf{x}_1, \mathbf{U}^{(k)}_2 = \mathbf{x}_2, \ldots, \mathbf{U}^{(k)}_{j-1} = \mathbf{x}_{j-1}\}$, and note that

$$f^{U,k}_{\Delta t}\left(\mathbf{U}^{(p-1)}\right) = \prod_{j=1}^{2^{p-1}} f^{U,k}_{\Delta t, U_j}\left(\mathbf{U}^{(p-1)}_j; \mathbf{U}^{(p-1)}_1, \mathbf{U}^{(p-1)}_2, \ldots, \mathbf{U}^{(p-1)}_{j-1}\right), \tag{35}$$

which gives us that

$$\frac{f^{U,k}_{\Delta t}\left(\mathbf{U}^{(p-1)}\right)}{f^{U,p-1}_{\Delta t}\left(\mathbf{U}^{(p-1)}\right)} = \prod_{j=1}^{2^{p-1}} \frac{f^{U,k}_{\Delta t, U_j}\left(\mathbf{U}^{(p-1)}_j; \mathbf{U}^{(p-1)}_1, \mathbf{U}^{(p-1)}_2, \ldots, \mathbf{U}^{(p-1)}_{j-1}\right)}{f^{U,p-1}_{\Delta t, U_j}\left(\mathbf{U}^{(p-1)}_j; \mathbf{U}^{(p-1)}_1, \mathbf{U}^{(p-1)}_2, \ldots, \mathbf{U}^{(p-1)}_{j-1}\right)}, \tag{36}$$

where the factors of the above expression are easily evaluated, as they are just evaluations of the joint normal distributions which correspond to the Euler steps in (33):

$$\begin{aligned} & f^{U,k}_{\Delta t, U_j}(\mathbf{x}_j; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{j-1}) \\ & = \frac{1}{(2\pi)^{n/2} \left| B(\mathbf{x}^*, t^*) \frac{\Delta t}{2^{p-1}} \right|^{1/2}} \exp\left(-\mathbf{v}^T \frac{1}{2} \left(B(\mathbf{x}^*, t^*) \frac{\Delta t}{2^{p-1}} \right)^{-1} \mathbf{v} \right), \end{aligned} \tag{37}$$

where, for the sake of brevity in the above equation, we use the notation

$$\mathbf{x}^* = \mathbf{x}_{\lfloor j-1 \rfloor_{p,k}}; \qquad t^* = \lfloor j-1 \rfloor_{p,k} \frac{\Delta t}{2^{p-1}} \tag{38}$$

and

$$\mathbf{v} = \left[\mathbf{x}_j - \mathbf{x}_{j-1} - \mathbf{D}(\mathbf{x}^*, t^*) \frac{\Delta t}{2^{p-1}} \right]. \tag{39}$$

Note that (36)–(39) require values of $\mathbf{D}(\mathbf{x}), \sigma(\mathbf{x})$ at $\mathbf{x} = \mathbf{U}^* = \mathbf{U}^{(p-1)}_{\lfloor j-1 \rfloor_{p,k}}$ only, and in the process of computing a realization of $\mathbf{U}^{(p-1)}$ (i.e., taking 2^{p-1} Euler steps of length $\frac{\Delta t}{2^{p-1}}$) we have already computed these values, so no further sampling of the diffusion and drift fields is needed in order to evaluate the products in (36).

And so, we are ready to proceed to the pseudocode description of the DRp scheme.

7. Pseudocode description of the DRp scheme. Here, we give a pseudocode description of the pth order accurate direct Richardson scheme. As in the previous sections we assume, without loss of generality, that the initial location is $\mathbf{Z}(t=0) = \mathbf{0}$, and we describe the algorithm by which we calculate $\mathbf{Z}^{DRp}_{\Delta t}$. First, we choose a parameter $c \in (0, 0.5)$ which will serve as a lower bound for the acceptance probability; in the computational results presented in this work, the value $c = 0.1$ is used. Later on, we shall demonstrate that the unconditional acceptance probability of the DRp algorithm converges to 0.5 in the limit $\Delta t \downarrow 0$.

1. Obtain a sample of the random variable $\mathbf{U}^{(p-1)}$, according to (33) (i.e., take 2^{p-1} Euler time steps).
2. For each $k = 0, 1, \ldots, p-2$, evaluate $\frac{f_{\Delta t}^{U,k}\left(\mathbf{U}^{(p-1)}\right)}{f_{\Delta t}^{U,p-1}\left(\mathbf{U}^{(p-1)}\right)}$ according to (36)–(39), and calculate

$$H\left(\mathbf{x}_1, \ldots, \mathbf{x}_{2^{p-1}}\right) \equiv \frac{1}{2} \sum_{k=0}^{p-1} l_{k+1}^p \frac{f_{\Delta t}^{U,k}\left(\mathbf{x}_1, \ldots, \mathbf{x}_{2^{p-1}}\right)}{f_{\Delta t}^{U,p-1}\left(\mathbf{x}_1, \ldots, \mathbf{x}_{2^{p-1}}\right)} \tag{40}$$

   at $\left(\mathbf{x}_1, \ldots, \mathbf{x}_{2^{p-1}}\right) = \mathbf{U}^{(p-1)}$.
3. Sample a random variable ξ with a standard uniform distribution: $\xi \sim U(0,1)$.
4. If the acceptance criterion

$$\xi < \max\left(H\left(\mathbf{U}^{(p-1)}\right), c\right) \tag{41}$$

   is met, then set $\mathbf{Z}_{\Delta t}^{DRp} = \mathbf{U}_{2^{p-1}}^{(p-1)}$. Else, go back to step 1.

As can be seen, the DR2 algorithm described earlier is the particular case of the above algorithm when $p = 2$ and $c = 0.1$. Examining (41) it is easily seen that the probability of acceptance at each iteration of steps 1–4 is at least equal to c. In fact, in Appendix A we shall prove the following theorem, which implies that as the time step Δt decreases, the probability of acceptance converges to $\frac{1}{2}$.

THEOREM 1. *For any integer $m, p \geq 1$ and any real number $d > 0$, if the fields $\mathbf{D}(\mathbf{x}, t)$ and $\sigma(\mathbf{x}, t)$ satisfy the regularity conditions stated in* (25)–(32)*, then there exists a constant $\bar{C} \in (0, \infty)$, dependent only on $p, m, d, \{C\}$, such that if the set $E \subset R^{n2^{p-1}}$ is defined by*

$$E = \left\{ \left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}\right) \in R^{n2^{p-1}} \;\middle|\; \left| \frac{f_{\Delta t}^{U,k}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}\right)}{f_{\Delta t}^{U,p-1}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}\right)} - 1 \right| > d \right\}, \tag{42}$$

then

$$\int_E f_{\Delta t}^{U,k}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}\right) |\mathrm{d}\mathbf{x}_1| \, |\mathrm{d}\mathbf{x}_2| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}| \leq \bar{C} \Delta t^m \tag{43}$$

and

$$\int_E f_{\Delta t}^{U,p-1}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}\right) |\mathrm{d}\mathbf{x}_1| \, |\mathrm{d}\mathbf{x}_2| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}| \leq \bar{C} \Delta t^m. \tag{44}$$

Theorem 1 is a very powerful result, as it implies that the probability of the quantity $H\left(\mathbf{U}^{(p-1)}\right)$ differing considerably from $\frac{1}{2}$ on a given acceptance-rejection step decreases faster than any power of Δt, as $\Delta t \downarrow 0$. This implies that the unconditional acceptance probability of the DRp algorithm converges to $\frac{1}{2}$ as $\Delta t \downarrow 0$.

To see why this is so, note that by the definition of the Richardson extrapolation coefficients, (18), we have that $\sum_{k=0}^{p-1} l_{k+1}^p = 1$; hence $H\left(\mathbf{U}^{(p-1)}\right)$ differing considerably from $\frac{1}{2}$ implies that at for least one $k \in \{0, 1, 2, \ldots, p-1\}$, $\frac{f_{\Delta t}^{U,k}\left(\mathbf{U}^{(p-1)}\right)}{f_{\Delta t}^{U,p-1}\left(\mathbf{U}^{(p-1)}\right)}$ differs considerably from 1, and by Theorem 1 the probability of this is negligible.

More concretely, choosing an arbitrary $c' \in \left(c, \frac{1}{2}\right)$ and setting $d = \frac{1-2c'}{\sum_{k=0}^{p-1}\left|l_{k+1}^p\right|}$, we get that

$$H\left(\mathbf{U}^{(p-1)}\right) \notin [c', 1-c'] \tag{45}$$

implies that for at least one $k \in \{0, 1, \ldots, p-1\}$ we have that $\left|\frac{f_{\Delta t}^{U,k}(\mathbf{U}^{(p-1)})}{f_{\Delta t}^{U,p-1}(\mathbf{U}^{(p-1)})} - 1\right| > d$. However, denoting by $P_{\mathbf{U^{(p-1)}}}\{\cdot\}$ the probability of a given event (dependent on $\mathbf{U}^{(p-1)}$) for one sampling of $\mathbf{U}^{(p-1)}$, i.e., one acceptance-rejection step, we have that

$$P_{\mathbf{U^{(p-1)}}}\left\{\left|\frac{f_{\Delta t}^{U,k}\left(\mathbf{U}^{(p-1)}\right)}{f_{\Delta t}^{U,p-1}\left(\mathbf{U}^{(p-1)}\right)} - 1\right| > d\right\} = \int_E f_{\Delta t}^{U,p-1}\left(\mathbf{x}_1, \ldots, \mathbf{x}_{2^{p-1}}\right)|\mathrm{d}\mathbf{x}_1| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}|, \tag{46}$$

where E is as defined in (42). Hence, applying Theorem 1 for this value of d, an arbitrary $m \geq p+1$, and all $k = 0, 1, \ldots, p-1$, we get that there exists a constant $\bar{C}$ such that

$$\begin{aligned} &P_{\mathbf{U^{(p-1)}}}\left\{H\left(\mathbf{U}^{(p-1)}\right) \notin [c', 1-c']\right\} \\ &\leq \sum_{k=0}^{p-1} P_{\mathbf{U^{(p-1)}}}\left\{\left|\frac{f_{\Delta t}^{U,k}\left(\mathbf{U}^{(p-1)}\right)}{f_{\Delta t}^{U,p-1}\left(\mathbf{U}^{(p-1)}\right)} - 1\right| > d\right\} \leq \bar{C}\Delta t^m. \end{aligned} \tag{47}$$

Since $c' \in \left(c, \frac{1}{2}\right)$ and $m \geq p+1$ are arbitrary, (47) implies that for an arbitrarily narrow interval $[c', 1-c']$, centered on $\frac{1}{2}$, as $\Delta t \downarrow 0$, the probability that $H\left(\mathbf{U}^{(p-1)}\right)$ falls outside of this interval decreases faster than any power of Δt. This, combined with (41) (which implies that the probability of acceptance is at least c, on each acceptance-rejection step of the DRp algorithm), implies that the DRp acceptance-rejection algorithm is well behaved, and as $\Delta t \downarrow 0$, the expected number of steps to acceptance converges to 2.

In the next section, we use the result of Theorem 1 to prove that, with $\mathbf{Z}_{\Delta t}^{DRp}$ being generated by the above algorithm, $f_{\Delta t}^{DRp}$ satisfies (23).

8. Proving that the DRp scheme satisfies its objective. Before we proceed, we need some additional definitions. Let us define $V_{c'}$ as follows:

$$V_{c'} = \left\{\left(\mathbf{x}_1, \ldots, \mathbf{x}_{2^{p-1}}\right) \in R^{n2^{p-1}} \middle| H\left(\mathbf{x}_1, \ldots, \mathbf{x}_{2^{p-1}}\right) \in [c', 1-c']\right\}. \tag{48}$$

It is important to note that, by Theorem 1, the complement of $V_{c'}$, $(V_{c'})^c$ contains a negligible part of the mass of the PDFs $f_{\Delta t}^{U,k}\left(\mathbf{x}_1, \ldots, \mathbf{x}_{2^{p-1}}\right)$, in the sense that

$$\int_{(V_{c'})^c} f_{\Delta t}^{U,k}\left(\mathbf{x}_1, \ldots, \mathbf{x}_{2^{p-1}}\right)|\mathrm{d}\mathbf{x}_1| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}| \leq \bar{C}\Delta t^m. \tag{49}$$

Next, let $\mathbf{W}^{(p-1)} = \left(\mathbf{W}_i^{(p-1)}\right)_{i=1}^{2^{p-1}}$ be the value of $\mathbf{U}^{(p-1)}$ at the last step of the acceptance-rejection loop, that is, the value which leads to an acceptance, and let $f_{\Delta t}^{W,p-1}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}\right)$ be its PDF on the sample space $R^{n2^{p-1}}$. Finally, let $P_{\mathbf{W^{(p-1)}}}\{\cdot\}$ denote the probability of a given event dependent on $\mathbf{W}^{(p-1)}$. Note that the difference between $P_{\mathbf{U^{(p-1)}}}\{\cdot\}$ and $P_{\mathbf{W^{(p-1)}}}\{\cdot\}$ is that the former is the probability for one acceptance-rejection step, whereas the latter is the probability for the entire time step, i.e., until the algorithm results in acceptance.

We then have that

$$\mathbf{Z}_{\Delta t}^{DRp} = \mathbf{W}_{2^{p-1}}^{(p-1)} \tag{50}$$

and

$$f_{\Delta t}^{DRp}(\mathbf{x}) = \int_{R^{n(2^{p-1}-1)}} f_{\Delta t}^{W,p-1}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}-1},\mathbf{x})\,|\mathrm{d}\mathbf{x}_1|\,|\mathrm{d}\mathbf{x}_2|\ldots|\mathrm{d}\mathbf{x}_{2^{p-1}-1}|\,. \tag{51}$$

Now, by (48), we have that, on $V_{c'}$, $\max(H(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}}),c) = H(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}}) \in [c', 1-c']$, and so for $(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}}) \in V_{c'}$

$$\begin{aligned}
&\max(H(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}}),c) \times f_{\Delta t}^{U,p-1}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}})\\
&= \frac{1}{2}\sum_{k=0}^{p-1} l_{k+1}^{p} \frac{f_{\Delta t}^{U,k}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}})}{f_{\Delta t}^{U,p-1}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}})} \times f_{\Delta t}^{U,p-1}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}})\\
&= \frac{1}{2}\sum_{k=0}^{p-1} l_{k+1}^{p} f_{\Delta t}^{U,k}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}})\,. \qquad (52)
\end{aligned}$$

Therefore, since $f_{\Delta t}^{U,p-1}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}})$ is the instrumental distribution in the DRp acceptance-rejection scheme, and the acceptance probability is as defined in (41), we have that, on $V_{c'}$,

$$\frac{f_{\Delta t}^{W,p-1}(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{2^{p-1}})}{\sum_{k=0}^{p-1} l_{k+1}^{p} f_{\Delta t}^{U,k}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}})} = Q_1, \tag{53}$$

where Q_1 is a constant that does not vary with $(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{2^{p-1}})$, and hence, because $f_{\Delta t}^{W,p-1}$ is strictly positive, we have that

$$\begin{aligned}
&\int_{V_{c'}} \left| f_{\Delta t}^{W,p-1}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}}) - \sum_{k=0}^{p-1} l_{k+1}^{p} f_{\Delta t}^{U,k}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}}) \right| |\mathrm{d}\mathbf{x}_1|\ldots|\mathrm{d}\mathbf{x}_{2^{p-1}}|\\
&= \left| \int_{V_{c'}} f_{\Delta t}^{W,p-1}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}}) - \sum_{k=0}^{p-1} l_{k+1}^{p} f_{\Delta t}^{U,k}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}})\,|\mathrm{d}\mathbf{x}_1|\ldots|\mathrm{d}\mathbf{x}_{2^{p-1}}| \right|\\
&\le \int_{(V_{c'})^c} f_{\Delta t}^{W,p-1}(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{2^{p-1}})\,|\mathrm{d}\mathbf{x}_1|\ldots|\mathrm{d}\mathbf{x}_{2^{p-1}}|\\
&\quad + \int_{(V_{c'})^c} \left| \sum_{k=0}^{p-1} l_{k+1}^{p} f_{\Delta t}^{U,k}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}}) \right| |\mathrm{d}\mathbf{x}_1|\ldots|\mathrm{d}\mathbf{x}_{2^{p-1}}|, \qquad (54)
\end{aligned}$$

where, in order to get the inequality in the above (54), we used the triangle inequality and the fact that $f_{\Delta t}^{W,p-1}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}})$ and $\sum_{k=0}^{p-1} l_{k+1}^{p} f_{\Delta t}^{U,k}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}})$ both integrate to 1 over the entire space $R^{n2^{p-1}}$. The above result, and (51), imply that

$$\begin{aligned}
&\int \left| f_{\Delta t}^{DRP}(\mathbf{x}) - \sum_{k=0}^{p-1} l_{k+1}^{p} f_{\Delta t}^{k}(\mathbf{x}) \right| |\mathrm{d}\mathbf{x}|\\
&= \int_{R^{n2^{p-1}}} \left| f_{\Delta t}^{W,p-1}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}}) - \sum_{k=0}^{p-1} l_{k+1}^{p} f_{\Delta t}^{U,k}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}}) \right| |\mathrm{d}\mathbf{x}_1|\ldots|\mathrm{d}\mathbf{x}_{2^{p-1}}|\\
&\le 2\int_{(V_{c'})^c} f_{\Delta t}^{W,p-1}(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{2^{p-1}})\,|\mathrm{d}\mathbf{x}_1|\ldots|\mathrm{d}\mathbf{x}_{2^{p-1}}|\\
&\quad + 2\int_{(V_{c'})^c} \left| \sum_{k=0}^{p-1} l_{k+1}^{p} f_{\Delta t}^{U,k}(\mathbf{x}_1,\ldots,\mathbf{x}_{2^{p-1}}) \right| |\mathrm{d}\mathbf{x}_1|\ldots|\mathrm{d}\mathbf{x}_{2^{p-1}}|\,. \qquad (55)
\end{aligned}$$

Now, by (49), we have that

$$\int_{(V_{c'})^c} \left| \sum_{k=0}^{p-1} l_{k+1}^p f_{\Delta t}^{U,k}\left(\mathbf{x}_1, \ldots, \mathbf{x}_{2^{p-1}}\right) \right| |\mathrm{d}\mathbf{x}_1| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}| \leq \bar{C}\Delta t^m \sum_{k=0}^{p-1} \left| l_{k+1}^p \right|, \tag{56}$$

and by the definition of $f_{\Delta t}^{W,p-1}$ we have that

$$\int_{(V_{c'})^c} f_{\Delta t}^{W,p-1}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}\right) |\mathrm{d}\mathbf{x}_1| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}| = P_{\mathbf{W}^{(p-1)}} \left\{ \mathbf{W}^{(p-1)} \in (V_{c'})^c \right\}. \tag{57}$$

Also, by (47) and the fact that the acceptance probability is at least c, we have that

$$\begin{aligned} & P_{\mathbf{W}^{(p-1)}} \left\{ \mathbf{W}^{(p-1)} \in (V_{c'})^c \right\} \\ & \leq \frac{1}{c} P_{\mathbf{U}^{(p-1)}} \left\{ \mathbf{U}^{(p-1)} \in (V_{c'})^c \right\} \leq \frac{1}{c} \bar{C}\Delta t^m. \end{aligned} \tag{58}$$

Substituting the result of (56)–(58) into (55), we get that

$$\int \left| f_{\Delta t}^{DRp}(\mathbf{x}) - \sum_{k=0}^{p-1} l_k^p f_{\Delta t}^k(\mathbf{x}) \right| |\mathrm{d}\mathbf{x}| \leq 2\bar{C}\Delta t^m \left(\frac{1}{c} + \sum_{k=0}^{p-1} \left| l_{k+1}^p \right| \right), \tag{59}$$

which proves that the DRp scheme satisfies the required condition, (23). In fact, as $m \geq p+1$ can be arbitrarily large, the result of (59) is much stronger than the required condition of (23): we have just shown that the L^1 difference between $f_{\Delta t}^{DRp}(\mathbf{x})$ and the Richardson extrapolate, $\sum_{k=0}^{p-1} l_k^p f_{\Delta t}^k(\mathbf{x})$, decreases faster than any power of Δt, as $\Delta t \downarrow 0$. This concludes the proof that the DRp scheme is weak pth order accurate.

9. Summary and conclusions. We have developed a new class of weak pth order accurate SDE integration schemes, for the solution of nonhomogeneous, anisotropic Ito SDEs with strictly positive definite diffusion matrices. These schemes, called the direct Richardson pth order accurate (DRp) schemes, perform Richardson extrapolation on the Euler algorithm in a conceptually new way, by means of an acceptance-rejection algorithm, after each time step.

Unlike previous applications of Richardson extrapolation to an Euler SDE solution, which are only applicable to the problem of estimating functionals of the distribution of the SDE process at the end time, the DRp solution is weak pth order accurate at each time step of the simulation, and can therefore be applied to any problem which requires a weakly convergent SDE integration scheme.

A simplified description of a particular member of the DRp class, DR2, has been provided. This description illustrates the elegance of the direct Richardson schemes and their ease of implementation in a computational code. Numerical results have been provided for both 2D anisotropic and 1D isotropic test cases, which compare the performance of DR2 and DR3 with that of other modern SDE integration schemes, in particular those developed by Kloeden and Platen [4] and by Cao and Pope [3]. The numerical results indicate that the error of the DR2, DR3 schemes is smaller than that of existing schemes based on Ito–Taylor expansions, whereas the computational cost of the DR schemes is somewhat higher, so that the overall computational efficiency is comparable. This suggests that the DRp family of SDE integration schemes are a practical alternative to existing SDE integration schemes, with the benefit of being easier to implement, especially in cases where the SDE diffusion is isotropic, or its matrix decomposition is known.

Appendix A. Proof of Theorem 1. Here, we prove Theorem 1.

THEOREM 1. *For any integer* $m, p \geq 1$ *and any real number* $d > 0$*, if the fields* $\mathbf{D}(\mathbf{x}, t)$ *and* $\sigma(\mathbf{x}, t)$ *satisfy the regularity conditions stated in* (25)–(32)*, then there exists a constant* $\bar{C} \in (0, \infty)$*, dependent only on* $p, m, d, \{C\}$*, such that if the set* $E \subset R^{n2^{p-1}}$ *is defined by*

$$E = \left\{ (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}) \in R^{n2^{p-1}} \;\middle|\; \left| \frac{f_{\Delta t}^{U,k}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}})}{f_{\Delta t}^{U,p-1}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}})} - 1 \right| > d \right\}, \tag{60}$$

then

$$\int_E f_{\Delta t}^{U,k}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}) \, |\mathrm{d}\mathbf{x}_1| \, |\mathrm{d}\mathbf{x}_2| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}| \leq \bar{C} \Delta t^m \tag{61}$$

and

$$\int_E f_{\Delta t}^{U,p-1}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}) \, |\mathrm{d}\mathbf{x}_1| \, |\mathrm{d}\mathbf{x}_2| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}| \leq \bar{C} \Delta t^m. \tag{62}$$

Proof. We use the convention $\mathbf{x}_0 = \mathbf{0}$. For a given value of $\epsilon \in (0, \infty)$, consider the set $G_{\Delta t}^{\epsilon} \in R^{n2^{p-1}}$ defined by

$$G_{\Delta t}^{\epsilon} = \left\{ (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}) \in R^{n2^{p-1}} \;\middle|\; \|\mathbf{x}_j - \mathbf{x}_{j-1}\| \leq \Delta t^{1/2-\epsilon} \right\}. \tag{63}$$

Our proof consists of two parts. In the first, we demonstrate that for a suitable ϵ and a small enough Δt, a point $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}})$ is in E only if it is also in the complement of $G_{\Delta t}^{\epsilon}$ (in other words $E \subseteq (G_{\Delta t}^{\epsilon})^c$). In the second part, we prove that the integral of either $f_{\Delta t}^{U,k}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}})$ or $f_{\Delta t}^{U,p-1}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}})$ over the set $(G_{\Delta t}^{\epsilon})^c$ is small; the reader can get some intuition as to why this is so by noting that, by their definition, the functions $f_{\Delta t}^{U,k}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}})$ and $f_{\Delta t}^{U,p-1}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}})$ consist of products of Gaussian PDFs with characteristic width $\Delta t^{1/2}$, whereas $G_{\Delta t}^{\epsilon}$ is a region of characteristic width $\Delta t^{1/2-\epsilon}$, and for small Δt we have that $\Delta t^{1/2-\epsilon} > \Delta t^{1/2}$, so that $G_{\Delta t}^{\epsilon}$ contains most of the mass of $f_{\Delta t}^{U,k}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}})$ and $f_{\Delta t}^{U,p-1}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}})$.

Consider a given point $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}) \in G_{\Delta t}^{\epsilon}$. We have then, from (35)–(39), that

$$\begin{aligned} &\frac{f_{\Delta t, U_j}^{U,k}(\mathbf{x}_j; \mathbf{x}_1, \ldots, \mathbf{x}_{j-1})}{f_{\Delta t, U_j}^{U,p-1}(\mathbf{x}_j; \mathbf{x}_2, \ldots, \mathbf{x}_{j-1})} \\ &= \frac{\left| B(\mathbf{x}_{j-1}, t_{j-1}) \frac{\Delta t}{2^{p-1}} \right|^{1/2}}{\left| B(\mathbf{x}^*, t^*) \frac{\Delta t}{2^{p-1}} \right|^{1/2}} \times \frac{\exp\left(-\mathbf{v}^T \frac{1}{2} \left(B(\mathbf{x}^*, t^*) \frac{\Delta t}{2^{p-1}} \right)^{-1} \mathbf{v} \right)}{\exp\left(-\tilde{\mathbf{v}}^T \frac{1}{2} \left(B(\mathbf{x}_{j-1}, t_{j-1}) \frac{\Delta t}{2^{p-1}} \right)^{-1} \tilde{\mathbf{v}} \right)}, \end{aligned} \tag{64}$$

where $\mathbf{x}^*, \mathbf{v}, t^*$ are as defined in (38), (39), and we introduce, for the sake of compactness, the notation, $t_{j-1} = (j-1)\frac{\Delta t}{2^{p-1}}$, and we shall also denote

$$\tilde{\mathbf{v}} = \left[\mathbf{x}_j - \mathbf{x}_{j-1} - \mathbf{D}(\mathbf{x}_{j-1}, t_{j-1}) \frac{\Delta t}{2^{p-1}} \right].$$

Considering the term $\frac{\left|B(\mathbf{x}_{j-1},t_{j-1})\frac{\Delta t}{2^{p-1}}\right|^{1/2}}{\left|B(\mathbf{x}^*,t^*)\frac{\Delta t}{2^{p-1}}\right|^{1/2}}$ first, we have that since $\|\mathbf{x}_{j-1}-\mathbf{x}^*\| \leq \Delta t^{1/2-\epsilon}$ and $|t_{j-1}-t^*| \leq \Delta t$, then from the regularity equations, (25)–(32),

$$\left|\frac{\left|B\left(\mathbf{x}_{j-1},t_{j-1}\right)\frac{\Delta t}{2^{p-1}}\right|^{1/2}}{\left|B\left(\mathbf{x}^*,t^*\right)\frac{\Delta t}{2^{p-1}}\right|^{1/2}}-1\right| \leq \left(\frac{C_3^{\sigma,\mathbf{x}}\Delta t^{1/2-\epsilon}+C_3^{\sigma,t}\Delta t}{C_2^{\sigma}}\right)^n. \tag{65}$$

Next, considering the term

$$\frac{\exp\left(-\mathbf{v}^T\frac{1}{2}\left(B\left(\mathbf{x}^*,t^*\right)\frac{\Delta t}{2^{p-1}}\right)^{-1}\mathbf{v}\right)}{\exp\left(-\tilde{\mathbf{v}}^T\frac{1}{2}\left(B\left(\mathbf{x}_{j-1},t_{j-1}\right)\frac{\Delta t}{2^{p-1}}\right)^{-1}\tilde{\mathbf{v}}\right)},$$

we have that

$$\frac{\exp\left(-\mathbf{v}^T\frac{1}{2}\left(B\left(\mathbf{x}^*,t^*\right)\frac{\Delta t}{2^{p-1}}\right)^{-1}\mathbf{v}\right)}{\exp\left(-\tilde{\mathbf{v}}^T\frac{1}{2}\left(B\left(\mathbf{x}_{j-1},t_{j-1}\right)\frac{\Delta t}{2^{p-1}}\right)^{-1}\tilde{\mathbf{v}}\right)} = \exp\left(-\frac{2^{p-2}}{\Delta t}\left(\mathbf{w}^T\mathbf{w}-\tilde{\mathbf{w}}^T\tilde{\mathbf{w}}\right)\right), \tag{66}$$

where $\mathbf{w} = \sigma\left(\mathbf{x}^*,t^*\right)^{-1}\mathbf{v}$ and $\tilde{\mathbf{w}} = \sigma\left(\mathbf{x}_{j-1},t_{j-1}\right)^{-1}\tilde{\mathbf{v}}$. Now, we have that $\mathbf{w}^T\mathbf{w} - \tilde{\mathbf{w}}^T\tilde{\mathbf{w}} = \left(\mathbf{w}-\tilde{\mathbf{w}}\right)^T\left(\mathbf{w}+\tilde{\mathbf{w}}\right)$, and using again (25)–(32), we have that

$$\|\mathbf{w}+\tilde{\mathbf{w}}\| \leq 2C_1^{\sigma^{-1}}\left(\Delta t^{1/2-\epsilon}+C_1^D\frac{\Delta t}{2^{p-1}}\right) \tag{67}$$

and

$$\begin{aligned}\mathbf{w}-\tilde{\mathbf{w}} &= \sigma\left(\mathbf{x}^*,t^*\right)^{-1}\mathbf{v}-\sigma\left(\mathbf{x}_{j-1},t_{j-1}\right)^{-1}\tilde{\mathbf{v}} \\ &= \left(\sigma\left(\mathbf{x}^*,t^*\right)^{-1}-\sigma\left(\mathbf{x}_{j-1},t_{j-1}\right)^{-1}\right)\mathbf{v}+\sigma\left(\mathbf{x}_{j-1},t_{j-1}\right)^{-1}\left(\mathbf{v}-\tilde{\mathbf{v}}\right) \\ &= \left(\sigma\left(\mathbf{x}^*,t^*\right)^{-1}-\sigma\left(\mathbf{x}_{j-1},t_{j-1}\right)^{-1}\right)\mathbf{v} \\ &\quad +\sigma\left(\mathbf{x}_{j-1},t_{j-1}\right)^{-1}\left(\mathbf{D}\left(\mathbf{x}_{j-1},t_{j-1}\right)-\mathbf{D}\left(\mathbf{x}^*,t^*\right)\right)\frac{\Delta t}{2^{p-1}}.\end{aligned} \tag{68}$$

Again invoking (25)–(32) to bound the magnitude of the two terms on the rightmost side of (68), we have that

$$\begin{aligned}\|\mathbf{w}-\tilde{\mathbf{w}}\| &\leq \left(\Delta t^{1/2-\epsilon}+C_1^D\frac{\Delta t}{2^{p-1}}\right)\left(C_3^{\sigma^{-1},\mathbf{x}}2^{p-1}\Delta t^{1/2-\epsilon}+C_3^{\sigma^{-1},t}\Delta t\right) \\ &\quad +\frac{\Delta t}{2^{p-1}}C_1^{\sigma^{-1}}\left(C_2^{D,\mathbf{x}}2^{p-1}\Delta t^{1/2-\epsilon}+C_2^{D,t}\Delta t\right).\end{aligned} \tag{69}$$

Combining the results of (67), (69), we get that for $\Delta t < 1, \epsilon \in (0, 0.5)$ there exists a constant $A_1 \in (0,\infty)$ such that

$$\left\|\mathbf{w}^T\mathbf{w}-\tilde{\mathbf{w}}^T\tilde{\mathbf{w}}\right\| \leq A_1\Delta t^{3/2-3\epsilon}. \tag{70}$$

Choosing $\epsilon = 1/12$, substituting the result of (70) into (66), and combining with the result of (65), we get that for any $h \in (0,\infty)$, we can choose Δt small enough such that if $(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{2^{p-1}}) \in G_{\Delta t}^{1/12}$, then we have that $\left|\frac{f_{\Delta t,U_j}^{U,k}(\mathbf{x}_j;\mathbf{x}_1,\ldots,\mathbf{x}_{j-1})}{f_{\Delta t,U_j}^{U,p-1}(\mathbf{x}_j;\mathbf{x}_2,\ldots,\mathbf{x}_{j-1})}-1\right| \leq h$.

Since this result holds for any $j \in \{1, 2, 3, \ldots, 2^{p-1}\}$, by (36) we can choose h small enough that $\left|\frac{f_{\Delta t}^{U,k}(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{2^{p-1}})}{f_{\Delta t}^{U,p-1}(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{2^{p-1}})} - 1\right| \leq d$ for any $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}) \in G_{\Delta t}^{1/12}$.

Therefore, for Δt small enough, we have that $E \subseteq \left(G_{\Delta t}^{1/12}\right)^c$, which implies that

$$\int_E f_{\Delta t}^{U,k}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}\right) |\mathrm{d}\mathbf{x}_1| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}|$$
$$\leq \int_{\left(G_{\Delta t}^{1/12}\right)^c} f_{\Delta t}^{U,k}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}\right) |\mathrm{d}\mathbf{x}_1| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}| . \tag{71}$$

Note that if $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}) \in \left(G_{\Delta t}^{1/12}\right)^c$, then for some i between 1 and 2^{p-1} we have that $\|\mathbf{x}_i - \mathbf{x}_{i-1}\| \geq \Delta t^{5/12}$, and so we have that

$$\int_{\left(G_{\Delta t}^{1/12}\right)^c} f_{\Delta t}^{U,k}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}\right) |\mathrm{d}\mathbf{x}_1| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}|$$
$$\leq \sum_{i=1}^{2^{p-1}} \int_{\left\{\|\mathbf{x}_i - \mathbf{x}_{i-1}\| \geq \Delta t^{5/12}\right\}} f_{\Delta t}^{U,k}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}\right) |\mathrm{d}\mathbf{x}_1| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}| , \tag{72}$$

where $\left\{\|\mathbf{x}_i - \mathbf{x}_{i-1}\| \geq \Delta t^{5/12}\right\}$ denotes the set of all points $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}) \in R^{n2^{p-1}}$ such that $\|\mathbf{x}_i - \mathbf{x}_{i-1}\| \geq \Delta t^{5/12}$. Let's consider a single term in the sum in (72). We have, by (35), that

$$\int_{\left\{\|\mathbf{x}_i - \mathbf{x}_{i-1}\| \geq \Delta t^{5/12}\right\}} f_{\Delta t}^{U,k}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}\right) |\mathrm{d}\mathbf{x}_1| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}|$$
$$= \int_{\left\{\|\mathbf{x}_i - \mathbf{x}_{i-1}\| \geq \Delta t^{5/12}\right\}} \prod_{j=1}^{2^{p-1}} f_{\Delta t, U_j}^{U,k}\left(\mathbf{x}_j; \mathbf{x}_1, \ldots, \mathbf{x}_{j-1}\right) |\mathrm{d}\mathbf{x}_1| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}|$$
$$= \int_{\left\{\|\mathbf{x}_i - \mathbf{x}_{i-1}\| \geq \Delta t^{5/12}\right\}} \prod_{j=1}^{i} f_{\Delta t, U_j}^{U,k}\left(\mathbf{x}_j; \mathbf{x}_1, \ldots, \mathbf{x}_{j-1}\right) |\mathrm{d}\mathbf{x}_1| \ldots |\mathrm{d}\mathbf{x}_i| , \tag{73}$$

where the second inequality follows from Fubini's theorem and the fact that, since $f_{\Delta t, U_j}^{U,k}\left(\mathbf{x}_j; \mathbf{x}_1, \ldots, \mathbf{x}_{j-1}\right)$ are conditional PDFs, as defined in (37), then for any $\mathbf{x}_1, \ldots,$ $\mathbf{x}_{j-1}$ we have that

$$\int_{R^n} f_{\Delta t, U_j}^{U,k}\left(\mathbf{x}_j; \mathbf{x}_1, \ldots, \mathbf{x}_{j-1}\right) |\mathrm{d}\mathbf{x}_j| = 1. \tag{74}$$

By inductively applying (74), we also get that

$$\int_{R^{n(i-1)}} \prod_{j=1}^{i-1} f_{\Delta t, U_j}^{U,k}\left(\mathbf{x}_j; \mathbf{x}_1, \ldots, \mathbf{x}_{j-1}\right) |\mathrm{d}\mathbf{x}_1| |\mathrm{d}\mathbf{x}_2| \ldots |\mathrm{d}\mathbf{x}_{i-1}| = 1. \tag{75}$$

Applying (75) and Fubini's theorem to (73), we get that

$$\int_{\left\{\|\mathbf{x}_i - \mathbf{x}_{i-1}\| \geq \Delta t^{5/12}\right\}} \prod_{j=1}^{i} f_{\Delta t, U_j}^{U,k}\left(\mathbf{x}_j; \mathbf{x}_1, \ldots, \mathbf{x}_{j-1}\right) |\mathrm{d}\mathbf{x}_1| \ldots |\mathrm{d}\mathbf{x}_i|$$
$$\leq \sup_{(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1})} \left(\int_{\left\{\|\mathbf{x}_i - \mathbf{x}_{i-1}\| \geq \Delta t^{5/12}\right\}} f_{\Delta t, U_i}^{U,k}\left(\mathbf{x}_i; \mathbf{x}_1, \ldots, \mathbf{x}_{i-1}\right) |\mathrm{d}\mathbf{x}_i| \right), \tag{76}$$

where the supremum in the above inequality is taken over all $(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}) \in R^{n(i-1)}$. Note, however, that by its definition, (37), we have that the function $f^{U,k}_{\Delta t, U_i}(\mathbf{x}_i; \mathbf{x}_1, \ldots, \mathbf{x}_{i-1})$ is a multivariate Gaussian distribution whose covariance matrix roughly scales as $\frac{\Delta t}{2^{p-1}}$, and so it is intuitively easy to see that integrating that distribution over the region $\left\{\|\mathbf{x}_i - \mathbf{x}_{i-1}\| \geq \Delta t^{5/12}\right\}$ gives a quantity which decreases faster than any power of Δt as $\Delta t \downarrow 0$. More concretely, we have that, for fixed $\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}$, using the definition of $f^{U,k}_{\Delta t, U_i}$ and the bounds given by (25)–(32), the inequality

$$
\begin{aligned}
(77)\quad & f^{U,k}_{\Delta t, U_i}(\mathbf{x}_i; \mathbf{x}_1, \ldots, \mathbf{x}_{i-1}) \\
& \leq \frac{1}{\left(2\pi \frac{\Delta t}{2^{p-1}}\right)^{n/2} (C^{\sigma}_2)^n} \exp\left(- \frac{\left(C^{\sigma^{-1}}_2\right)^2 2^{p-2}}{\Delta t} \left\| \mathbf{x}_i - \mathbf{x}_{i-1} - \mathbf{D}(\mathbf{x}^*, t^*) \frac{\Delta t}{2^{p-1}} \right\|^2 \right),
\end{aligned}
$$

holds, and for $\Delta t \leq \left(\frac{2^{p-2}}{C^D_1}\right)^{12/7}$, we have that $\|\mathbf{x}_i - \mathbf{x}_{i-1}\| \geq \Delta t^{5/12}$ implies that $\left\|\mathbf{x}_i - \mathbf{x}_{i-1} - \mathbf{D}(\mathbf{x}^*, t^*) \frac{\Delta t}{2^{p-1}}\right\| \geq \frac{1}{2}\|\mathbf{x}_i - \mathbf{x}_{i-1}\|$; hence for $\Delta t \leq \left(\frac{2^{p-2}}{C^D_1}\right)^{12/7}$ and $\|\mathbf{x}_i - \mathbf{x}_{i-1}\| \geq \Delta t^{5/12}$, we have that

$$
\begin{aligned}
(78)\quad & f^{U,k}_{\Delta t, U_i}(\mathbf{x}_i; \mathbf{x}_1, \ldots, \mathbf{x}_{i-1}) \\
& \leq \frac{1}{\left(2\pi \frac{\Delta t}{2^{p-1}}\right)^{n/2} (C^{\sigma}_2)^n} \exp\left(- \frac{\left(C^{\sigma^{-1}}_2\right)^2 2^{p-4}}{\Delta t} \|\mathbf{x}_i - \mathbf{x}_{i-1}\|^2 \right),
\end{aligned}
$$

which implies that

$$
\begin{aligned}
& \sup_{(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1})} \left(\int_{\left\{\|\mathbf{x}_i - \mathbf{x}_{i-1}\| \geq \Delta t^{5/12}\right\}} f^{U,k}_{\Delta t, U_i}(\mathbf{x}_i; \mathbf{x}_1, \ldots, \mathbf{x}_{i-1}) |\mathrm{d}\mathbf{x}_i| \right) \\
& \leq \int_{\left\{\|\mathbf{y}\| \geq \Delta t^{5/12}\right\}} \frac{1}{\left(2\pi \frac{\Delta t}{2^{p-1}}\right)^{n/2} (C^{\sigma}_2)^n} \exp\left(- \frac{\left(C^{\sigma^{-1}}_2\right)^2 2^{p-4}}{\Delta t} \|\mathbf{y}\|^2 \right) |\mathrm{d}\mathbf{y}| \\
(79)\quad & = \int_{\left\{\|\mathbf{z}\| \geq \Delta t^{-1/12}\right\}} \frac{1}{\left(\frac{2\pi}{2^{p-1}}\right)^{n/2} (C^{\sigma}_2)^n} \exp\left(- \left(C^{\sigma^{-1}}_2\right)^2 2^{p-4} \|\mathbf{z}\|^2 \right) |\mathrm{d}\mathbf{z}|,
\end{aligned}
$$

where the third line follows from the second by the simple change of variables $\mathbf{z} = \mathbf{y}\Delta t^{-1/2}$. Using this result in conjunction with (76), (73), (72), we get that for $\Delta t \leq \left(\frac{2^{p-2}}{C^D_1}\right)^{12/7}$

$$
\begin{aligned}
& \int_{\left(G^{1/12}_{\Delta t}\right)^c} f^{U,k}_{\Delta t}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^{p-1}}) |\mathrm{d}\mathbf{x}_1| \ldots |\mathrm{d}\mathbf{x}_{2^{p-1}}| \\
(80)\quad & \leq 2^{p-1} \int_{\left\{\|\mathbf{z}\| \geq \Delta t^{-1/12}\right\}} \frac{1}{\left(\frac{2\pi}{2^{p-1}}\right)^{n/2} (C^{\sigma}_2)^n} \exp\left(- \left(C^{\sigma^{-1}}_2\right)^2 2^{p-4} \|\mathbf{z}\|^2 \right) |\mathrm{d}\mathbf{z}|,
\end{aligned}
$$

and it is a simple, albeit lengthy, calculus exercise to demonstrate that for any integer m there exists a constant $\bar{C}$ such that

$$
2^{p-1} \int_{\left\{\|\mathbf{z}\| \geq \Delta t^{-1/12}\right\}} \frac{1}{\left(\frac{2\pi}{2^{p-1}}\right)^{n/2} (C^{\sigma}_2)^n} \exp\left(- \left(C^{\sigma^{-1}}_2\right)^2 2^{p-4} \|\mathbf{z}\|^2 \right) |\mathrm{d}\mathbf{z}| \leq \bar{C}\Delta t^m,
$$
(81)

which proves the first part of Theorem 1, (43). The proof of the second part is completely identical, with the only difference being that we have $(\mathbf{x}_{i-1}, t_{i-1})$ in place of $(\mathbf{x}^*, t^*)$ in (77), which has no impact on the argument of (77)–(81). And so, we have proven Theorem 1.

Appendix B. Proof of Theorem 2. Here, we state and prove Theorem 2, which was used to demonstrate that the DRp schemes satisfy the criterion for weak pth order accuracy.

THEOREM 2. *For the random variable* $\mathbf{Z}_{\Delta t}^{DRp}$ *defined in section* 7*, if* (23) *holds, then so does* (24)*.*

Proof. We shall divide the domain, $\mathbb{R}^n$, into two parts: a ball of radius 1 centered on the origin, $\mathcal{B}(\mathbf{0},1)$, and its complement, $\mathcal{B}(\mathbf{0},1)^c$. We have that

$$\begin{aligned} &\left| E\left(\prod_{m=1}^{n}\left(Z_{\Delta t,m}^{DRp}\right)^{i_m} - \sum_{k=1}^{p} l_k^p \prod_{m=1}^{n}\left(Z^{E}_{\frac{\Delta t}{2^{k-1}},m}\right)^{i_m}\right)\right| \\ &= \left|\int \prod_{m=1}^{n} x_m^{i_m}\left(f_{\Delta t}^{DRp}(\mathbf{x}) - \sum_{k=0}^{p-1} l_{k+1}^p f_{\Delta t}^k(\mathbf{x})\right)|\mathrm{d}\mathbf{x}|\right| \\ &\le \int_{\mathcal{B}(\mathbf{0},1)} \|\mathbf{x}\|^{\sum_{m=1}^n i_m}\left|f_{\Delta t}^{DRp}(\mathbf{x}) - \sum_{k=0}^{p-1} l_{k+1}^p f_{\Delta t}^k(\mathbf{x})\right||\mathrm{d}\mathbf{x}| \\ &\quad + \int_{\mathcal{B}(\mathbf{0},1)^c} \|\mathbf{x}\|^{\sum_{m=1}^n i_m}\left|f_{\Delta t}^{DRp}(\mathbf{x}) - \sum_{k=0}^{p-1} l_{k+1}^p f_{\Delta t}^k(\mathbf{x})\right||\mathrm{d}\mathbf{x}|, \end{aligned} \tag{82}$$

and since $\|\mathbf{x}\|^{\sum_{m=1}^n i_m}$ is bounded on $\mathcal{B}(\mathbf{0},1)$, we have that (23) implies that there exists a constant, C'', such that

$$\int_{\mathcal{B}(\mathbf{0},1)} \|\mathbf{x}\|^{\sum_{m=1}^n i_m}\left|f_{\Delta t}^{DRp}(\mathbf{x}) - \sum_{k=0}^{p-1} l_{k+1}^p f_{\Delta t}^k(\mathbf{x})\right||\mathrm{d}\mathbf{x}| \le C''\Delta t^{p+1}. \tag{83}$$

Now, let us consider the second term in (82) and obtain a result analogous to (83) for it. From the boundedness of σ and $\mathbf{D}$, there exists a parameter Δt_0, dependent on $\{C\}$ and n only, such that for all $\Delta t \le \Delta t_0$ and any k between 0 and $p-1$, we have that $\|\mathbf{x}\| \ge 1$ implies that

$$f_{\Delta t}^k(\mathbf{x}) \le \frac{1}{(2\pi\Delta t)^{n/2}(2C_1^\sigma)^n}\exp\left(-\frac{\|\mathbf{x}\|^2}{2\Delta t\,(2C_1^\sigma)^2}\right). \tag{84}$$

Also, since $c > 0$, as defined in section 7, gives a lower bound on the acceptance probability, we have that

$$\frac{f_{\Delta t}^{W,p-1}}{f_{\Delta t}^{U,p-1}} \le \frac{1}{c}, \tag{85}$$

which implies that

$$\frac{f_{\Delta t}^{DRp}}{f_{\Delta t}^{p-1}} \le \frac{1}{c}, \tag{86}$$

and so (84), (86) jointly imply that for $\Delta t \leq \Delta t_0$ and $\mathbf{x} \in \mathcal{B}(\mathbf{0},1)^c$, we have that

$$\begin{aligned} &\left| f_{\Delta t}^{DRp}(\mathbf{x}) - \sum_{k=0}^{p-1} l_{k+1}^p f_{\Delta t}^k(\mathbf{x}) \right| \\ &\leq \left(\frac{1}{c} + \sum_{k=0}^{p-1} \left| l_{k+1}^p \right| \right) \frac{1}{(2\pi\Delta t)^{n/2} (2C_1^\sigma)^n} \exp\left(-\frac{\|\mathbf{x}\|^2}{2\Delta t \,(2C_1^\sigma)^2} \right), \end{aligned} \tag{87}$$

which implies that for $\Delta t \leq \Delta t_0$ we have

$$\begin{aligned} &\int_{\mathcal{B}(\mathbf{0},1)^c} \|\mathbf{x}\|^{\sum_{m=1}^n i_m} \left| f_{\Delta t}^{DRp}(\mathbf{x}) - \sum_{k=0}^{p-1} l_{k+1}^p f_{\Delta t}^k(\mathbf{x}) \right| |\mathrm{d}\mathbf{x}| \\ &\leq \int_{\mathcal{B}(\mathbf{0},1)^c} \|\mathbf{x}\|^{\sum_{m=1}^n i_m} \frac{\frac{1}{c} + \sum_{k=0}^{p-1} \left| l_{k+1}^p \right|}{(2\pi\Delta t)^{n/2} (2C_1^\sigma)^n} \exp\left(-\frac{\|\mathbf{x}\|^2}{2\Delta t \,(2C_1^\sigma)^2} \right) |\mathrm{d}\mathbf{x}| \,. \end{aligned} \tag{88}$$

As the Gaussian distribution on the second line of (88) has characteristic width $2C_1^\sigma\sqrt{\Delta t}$, whereas the domain of integration is over all $\mathbf{x}$ with $\|\mathbf{x}\| \geq 1$, it is a simple, albeit lengthy, calculus exercise to demonstrate that there exists a constant C'' such that

$$\int_{\mathcal{B}(\mathbf{0},1)^c} \|\mathbf{x}\|^{\sum_{m=1}^n i_m} \frac{\frac{1}{c} + \sum_{k=0}^{p-1} \left| l_{k+1}^p \right|}{(2\pi\Delta t)^{n/2} (2C_1^\sigma)^n} \exp\left(-\frac{\|\mathbf{x}\|^2}{2\Delta t \,(2C_1^\sigma)^2} \right) |\mathrm{d}\mathbf{x}| \leq C'' \Delta t^{p+1}. \tag{89}$$

Combining the results of (83), (89) into (82), (24) follows immediately, which concludes the proof of Theorem 2.

REFERENCES

[1] S. Asmussen and P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Springer, New York, 2007.

[2] N. Bou-Rabee and E. Vanden-Eijnden, *Pathwise accuracy and ergodicity of metropolized integrators for SDEs*, Comm. Pure Appl. Math., 63 (2010), pp. 655–696.

[3] R. Cao and S. B. Pope, *Numerical integration of stochastic differential equations: Weak second-order mid-point scheme for application in the composition PDF method*, J. Comput. Phys., 185 (2003), pp. 194–212.

[4] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, 1992.

[5] M. B. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.

[6] M. B. Giles, D. J. Higham, and X. Mao, *Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff*, Finance Stoch., 13 (2009), pp. 403–413.

[7] G. N. Milstein and M. V. Tretyakov, *Stochastic Numerics for Mathematical Physics*, Springer, Berlin, 2004.

[8] P. P. Popov and S. B. Pope, *Stochastic particle advection in hybrid large eddy simulation/filtered density function methods*, Bull. Amer. Phys. Soc., 53 (15) (2008).

[9] P. Protter and D. Talay, *The Euler scheme for Levy driven stochastic differential equations*, Ann. Probab., 25 (1997), pp. 393–423.

[10] L. F. Richardson and J. A. Gaunt, *The deferred approach to the limit. Part* I. *Single lattice. Part* II. *Interpenetrating lattices*, Philos. Trans. Roy. Soc. A, 226 (1927), pp. 299–361.

[11] D. Talay and L. Tubaro, *Expansion of the global error for numerical schemes solving stochastic differential equations*, Stoch. Anal. Appl., 8 (1990), pp. 94–120.

[12] A. Tocino and J. Vigo-Aguiar, *Weak second order conditions for stochastic Runge–Kutta methods*, SIAM J. Sci. Comput., 24 (2002), pp. 507–523.

[13] H. Wang, P. P. Popov, and S. B. Pope, *Weak second order splitting schemes for Lagrangian Monte Carlo particle methods for the composition PDF/FDF transport equations*, J. Comput. Phys., 229 (2010), pp. 1852–1878.