

CONVERGENCE RATES FOR ADAPTIVE WEAK APPROXIMATION OF STOCHASTIC DIFFERENTIAL EQUATIONS

KYOUNG-SOOK MOON, ANDERS SZEPESSY, RAÚL TEMPONE, AND GEORGIOS E.
ZOURARIS

ABSTRACT. Convergence rates of adaptive algorithms for weak approximations of Itô stochastic differential equations are proved for the Monte Carlo Euler method. Two algorithms based either on optimal stochastic time steps or optimal deterministic time steps are studied. The analysis of their computational complexity combines the error expansions with a posteriori leading order term introduced in [A. Szepessy, R. Tempone and G. Zouraris, *Comm. Pure and Appl. Math.*, 54, 1169-1214, 2001] and an extension of the convergence results for adaptive algorithms approximating deterministic ordinary differential equations, derived in [K-S. Moon, A. Szepessy, R. Tempone and G. Zouraris, preprint <http://www.nada.kth.se/~szepessy/ode.ps>]. The main step in the extension is the proof of the almost sure convergence of the error density. Both adaptive algorithms are proven to stop with the optimal number of steps up to a problem independent factor defined in the algorithm. Numerical examples illustrate the behavior of the adaptive algorithms, motivating when stochastic and deterministic adaptive time step are superior to constant time steps and when adaptive stochastic steps are superior to adaptive deterministic steps.

1. INTRODUCTION TO ADAPTIVE ALGORITHMS FOR SDE'S

This work derives convergence rates of adaptive algorithms for weak approximation of Itô stochastic differential equations

$$(1.1) \quad dX_k(t) = a_k(t, X(t))dt + \sum_{\ell=1}^{\ell_0} b_k^\ell(t, X(t))dW^\ell(t), \quad t > 0$$

where $k = 1, \dots, d$ and $(X(t; \omega))$ is a stochastic process in \mathbb{R}^d , with randomness generated by the independent one dimensional Wiener processes $W^\ell(t; \omega)$, $\ell = 1, \dots, \ell_0$, on the probability space (Ω, \mathcal{F}, P) , cf. [19], [28]. The functions $a(t, x) \in \mathbb{R}^d$ and $b^\ell(t, x) \in \mathbb{R}^d$, $\ell = 1, \dots, \ell_0$, are given drift and diffusion fluxes.

The goal is to construct approximations to the expected value $E[g(X(T))]$ by a Monte Carlo method, for a given function $g : \mathbb{R}^d \rightarrow \mathbb{R}$. A topical example of such an expected value is to compute option prices in mathematical finance, cf. [18]. Other related models based on stochastic dynamics are used, e.g., for stochastic

1991 *Mathematics Subject Classification.* 65C30, 65Y20, 65L50, 60H35.

Key words and phrases. adaptive mesh refinement algorithm, computational complexity, Monte Carlo method, almost sure convergence.

This work is supported by the Swedish Research Council, UdelaR and UdeM in Uruguay, the Swedish Parallel and Scientific Computing Institute (PSCI), the Swedish National Board for Industrial and Technical Development (NUTEK), the EU-TMR project viscosity solutions and their applications.

climate prediction and for wave propagation in random media, cf. [21],[1]. The Monte Carlo Euler method approximates the unknown process X by the Euler method $\bar{X}(t_n)$ (cf. [20, 22]), which is a time discretization based on the nodes $0 = t_1 < t_2 < \dots < t_{N+1} = T$ where

$$(1.2) \quad \bar{X}(t_{n+1}) - \bar{X}(t_n) = \Delta t_n a(t_n, \bar{X}(t_n)) + \sum_{\ell=1}^{\ell_0} \Delta W_n^\ell b^\ell(t_n, \bar{X}(t_n)),$$

and $\Delta t_n \equiv t_{n+1} - t_n$, $\Delta W_n^\ell \equiv W^\ell(t_{n+1}) - W^\ell(t_n)$, $n = 1, 2, \dots, N + 1$. The aim of the adaptive algorithm is to choose the size of the time steps, Δt_n , and the number of independent identically distributed samples $\bar{X}(\cdot, \omega_j)$, $j = 1, 2, \dots, M$, such that the computational work, $N \cdot M$, is minimal while the approximation error is bounded by a given error tolerance, TOL, i.e. the event

$$(1.3) \quad \left| E[g(X(T))] - \frac{1}{M} \sum_{j=1}^M g(\bar{X}(T; \omega_j)) \right| \leq \text{TOL}$$

has a probability close to one. A priori error estimates of the computational error in (1.3) was first derived by Talay and Tubaro in [33]. The work [31] modified Talay's and Tubaro's error expansion to an expansion with computable leading order term in a posteriori form, based on computable stochastic flows and discrete dual backward problems.

Here we derive convergence rates of two algorithms including dividing and merging of time steps, with either stochastic or deterministic time steps. The difference between the two algorithms is that the stochastic time steps may use different meshes for each realization, while the deterministic time steps use the same mesh for all realizations. The construction and the analysis of the adaptive algorithms are inspired by the related work [25], on adaptive algorithms for deterministic ordinary differential equations, and use the error estimates from [31].

There are numerous adaptive algorithms for ordinary and partial differential equations, cf. [16], [2], [3], [5, 6], [4], [11], [17], [24], but the theoretical understanding of convergence rates of adaptive algorithms is not as well developed; there are however recent important contributions. The work [14], [15] and [26] prove optimal convergence rates for strong approximation of stochastic differential equations. DeVore studies in [9] the efficiency of adaptive approximation of functions, including wavelet expansions, based on smoothness conditions in Besov spaces. Inspired by this approximation results, Cohen, Dahmen and DeVore prove in [7] that a wavelet-based adaptive N -term approximation algorithm produces a solution with optimal error $\mathcal{O}(N^{-s})$ in the energy norm for linear coercive elliptic problems, see also [8]. The work [25] connects DeVore's smoothness conditions to error densities for adaptive approximation of ordinary differential equations. In particular [25] constructs an algorithm and proves that it stops with the optimal number of time steps, up to a problem independent factor defined in the algorithm; for any p -th order accurate method, the optimal number of adaptive steps is proportional to the p -th root of the $L^{\frac{1}{p+1}}$ quasi-norm of the error density, while the number of constant steps, with the same error, is proportional to the p -th root of the larger L^1 -norm of the error density. This work generalizes [25] to weak approximation of stochastic differential equations.

There are two main results on efficiency and accuracy of the adaptive algorithms described in Section 3. In view of accuracy with probability close to one, the approximation errors in (1.3) are asymptotically bounded by the specified error tolerance times a problem independent factor as the tolerance parameter tends to zero. In view of efficiency, both the algorithms with stochastic steps and deterministic steps stop with the optimal expected number of final time steps and optimal number of final time steps respectively, up to a problem independent factor. The number of final time steps is related to the numerical effort needed to compute the approximation. To be more precise, the total work for deterministic steps is roughly $M \cdot N$ where M is the final number of realizations and N is the final number of time steps, since the work to determine the mesh turns out to be negligible. On the other hand, the total work with stochastic steps is on average bounded by $M \cdot E[N_{\text{tot}}]$, where the total number, N_{tot} , of steps including all refinement levels is bounded by $\mathcal{O}(N \log N)$ with N steps in the final refinement; for each realization it is necessary to determine the mesh, which may vary for each realization.

The accuracy and efficiency results are based on the fact that the error density, ρ which measures the approximation error for each interval following (2.1), converges *almost surely* or *a.s.* as the error tolerance tends to zero. This convergence can be understood by the *a.s.* convergence of the approximate solution, \bar{X} , as the maximal step size tends to zero. Although the time steps are not adapted to the standard filtration generated by W for the stochastic time stepping algorithm, the work [31] proved that the corresponding approximate solution converges to the correct adapted solution X . This result makes it possible to prove the martingale property of the approximate error term with respect to a specific filtration, see Lemma 4.2. Therefore Theorem 4.1 and 4.4 use Doob's inequality to prove the *a.s.* convergence of \bar{X} . Similar results of pointwise convergence with constant step sizes, adapted to the standard filtration, are surveyed by Talay in [32].

The outline of the paper is: Section 2 states the a posteriori error expansion, proved in [31] and used in the adaptive algorithms; Section 3 describes and analyzes the adaptive algorithms with stochastic time steps and deterministic time steps; Section 4 proves *a.s.* convergence of the error density; and finally Section 5 presents numerical experiments based on the adaptive algorithms.

For simplicity, we introduce the following notation

$$d_{ij} \equiv \frac{1}{2} b_i^\ell b_j^\ell, \quad \partial_k \equiv \frac{\partial}{\partial x_k}, \quad \partial_{ki} \equiv \frac{\partial^2}{\partial x_k \partial x_i}, \quad \dots$$

with the summation convention, i.e., if the same subscript appears twice in a term, the term denotes the sum over the range of this subscript, e.g.

$$c_{ik} \partial_k b_j \equiv \sum_{k=1}^d c_{ik} \partial_k b_j.$$

For a derivative ∂_α the notation $|\alpha|$ is its order.

2. A POSTERIORI ERROR EXPANSION

The main result of [31] is new expansions of the computational error with computable leading order term in a posteriori form. The result was inspired by a corresponding a priori analysis derived in [33], with the main difference that the

weight for the local error contribution to the global error can be computed efficiently by stochastic flows and discrete dual backward problems in [25]. These a posteriori error expansions can be used in adaptive algorithms, in order to control the approximation error. Although [31] proposed adaptive algorithms, the main focus in that work was on error estimates. Properties regarding the stopping, efficiency and accuracy of the adaptive algorithms, following the ideas in [25] are first studied here. Assume that the process X satisfies (1.1) and its approximation, \bar{X} , is given by (1.2), then these error expansions in Theorems 1.2 and 2.2 of [31] have the form

$$(2.1) \quad E[g(X(T)) - g(\bar{X}(T))] = E \left[\sum_{n=1}^N \tilde{\rho}_n \Delta t_n^2 \right] \\ = E \left[\sum_{n=1}^N \rho_n \Delta t_n^2 \right] + \text{higher order terms,}$$

where $\rho_n \Delta t_n^2$ are computable error indicators, i.e. they provide information for further improvement of the time mesh and ρ_n measures the density of the global error in (2.1). A typical adaptive algorithm does two things iteratively:

- (1) if the error indicators satisfy an accuracy condition then it stops; otherwise
- (2) the algorithm chooses where to refine the mesh and then makes an iterative step to (1).

In addition to estimate the global error $E[g(X(T)) - g(\bar{X}(T))]$ in the sense of (2.1), the indicators $\rho_n \Delta t_n^2$ also give simple information on where to refine to reach an optimal mesh, based on the almost sure convergence of the density ρ_n as we refine the discretization, see Section 4.

In the remaining part of this section we state one error expansion from [31], which can be used with either stochastic or deterministic time steps. The work [31] also proves a second error expansion which requires less computational work per realization, in particular for large d , however this expansion is only valid with deterministic time steps and it has larger statistical error. Although adaptive algorithms based on this second expansion work well in practice, the larger statistical error makes it difficult to analyze precisely. Therefore this work focus on the first error expansion.

The following Lemma 2.1 and Theorem 2.2, derived in [31], describe the error expansion which is used in the adaptive algorithms in Section 3. Assume that for all times $t \in [t_n, t_{n+1})$ and all outcomes ω the time steps $\Delta t(t) = \Delta t_n$ are constructed by the refinement criterion

$$(2.2) \quad \begin{aligned} \Delta t(t) &= T2^{-m}, \text{ for some positive integer } m = m(t, \omega), \\ |\rho(t, \omega)| (\Delta t(t))^2 &< \text{constant,} \end{aligned}$$

with an approximate error density function, $\rho(t, \omega) = \rho_n(\omega)$, satisfying, for $s, t \in [0, T]$ and all outcomes ω , the uniform upper and lower bounds

$$(2.3) \quad \begin{aligned} c(\text{TOL}) \leq |\rho(s, \omega)| &\leq C(\text{TOL}), \\ |\partial_{W(t)} \rho(s, \omega)| &\leq C(\text{TOL}), \end{aligned}$$

for some positive functions c and C , with $\text{TOL}/c(\text{TOL}) \rightarrow 0$ as $\text{TOL} \rightarrow 0$. Here $\partial_{W(t)} Y$ denotes the Malliavin derivative which is the first variation of a process Y

with respect to a perturbation $dW(t)$ at time t of the Wiener process, cf. [27] and [31].

Lemma 2.1. *Suppose there are positive constants k and C and integer m_0 with the bounds*

$$\begin{aligned} g &\in C_{loc}^{m_0}(\mathbb{R}^d), \quad |\partial_\alpha g(x)| \leq C(1 + |x|^k), \quad \text{for all } |\alpha| \leq m_0, \\ E[|X(0)|^{2k+d+1} + |\bar{X}(0)|^{2k+d+1}] &\leq C, \end{aligned}$$

and

$$a \text{ and } b \text{ are bounded in } C^{m_0}([0, T] \times \mathbb{R}^d).$$

Assume that \bar{X} is constructed by the forward Euler method with step sizes Δt_n satisfying (2.2, 2.3) and the corresponding $\Delta W_n \equiv W(t_{n+1}) - W(t_n)$ are generated by Brownian bridges, based on the stochastic time step algorithm in Section 3. Assume also that $\bar{X}(0) = X(0)$. Then there exists a sufficiently large integer m_0 such that

$$(2.4) \quad \sup_{0 \in [0, T]} \sqrt{E[|X(t) - \bar{X}(t)|^2]} = \mathcal{O}\left(\sqrt{\Delta t_{sup}}\right) = \mathcal{O}\left(\sqrt{\frac{\text{TOL}}{c(\text{TOL})}}\right) \rightarrow 0,$$

as $\text{TOL} \rightarrow 0$, where $\Delta t_{sup} \equiv \sup_{n, \omega} \Delta t_n(\omega)$.

Theorem 2.2. *Suppose that a, b, g and X satisfy the assumptions in Lemma 2.1 and $E[|X(0)|^{k_0}] \leq C$ for some $k_0 \geq 16$. Then the time discretization error has the expansion*

$$(2.5) \quad \begin{aligned} E[g(X(T)) - g(\bar{X}(T))] &= E\left[\sum_{n=1}^N \rho(t_n, \bar{X})(\Delta t_n)^2\right] \\ &+ \mathcal{O}\left(\sqrt{\frac{\text{TOL}}{c(\text{TOL})}} \left(\frac{C(\text{TOL})}{c(\text{TOL})}\right)^{8/k_0}\right) E\left[\sum_{n=1}^N (\Delta t_n)^2\right] \end{aligned}$$

with computable leading order terms, where

$$(2.6) \quad \begin{aligned} \rho(t_n, \bar{X}) &\equiv \frac{1}{2} \left(\frac{\partial}{\partial t} a_k + \partial_j a_k a_j + \partial_{ij} a_k d_{ij} \right) \varphi_k(t_{n+1}) \\ &+ \frac{1}{2} \left(\frac{\partial}{\partial t} d_{km} + \partial_j d_{km} a_j + \partial_{ij} d_{km} d_{ij} + 2\partial_j a_k d_{jm} \right) \varphi'_{km}(t_{n+1}) \\ &+ \partial_j d_{km} d_{jr} \varphi''_{kmr}(t_{n+1}), \end{aligned}$$

and the terms in the sum of (2.6) are evaluated at the a posteriori known points $(t_n, \bar{X}(t_n))$, i.e.

$$\partial_\alpha a \equiv \partial_\alpha a(t_n, \bar{X}(t_n)), \quad \partial_\alpha b \equiv \partial_\alpha b(t_n, \bar{X}(t_n)), \quad \partial_\alpha d \equiv \partial_\alpha d(t_n, \bar{X}(t_n)).$$

Here $\varphi \in \mathbb{R}^d$ is the solution of the discrete dual backward problem

$$(2.7) \quad \begin{aligned} \varphi_i(t_n) &= \partial_i c_j(t_n, \bar{X}(t_n)) \varphi_j(t_{n+1}), \quad t_n < T, \\ \varphi_i(T) &= \partial_i g(\bar{X}(T)), \end{aligned}$$

with

$$(2.8) \quad c_i(t_n, x) \equiv x_i + \Delta t_n a_i(t_n, x) + \Delta W_n^\ell b_i^\ell(t_n, x)$$

and its first and second variation

$$(2.9) \quad \varphi'_{ij} \equiv \partial_{x_j(t_n)} \varphi_i(t_n) \equiv \frac{\partial \varphi_i(t_n; \overline{X}(t_n) = x)}{\partial x_j},$$

$$(2.10) \quad \varphi''_{ikm}(t_n) \equiv \partial_{x_m(t_n)} \varphi'_{ik}(t_n) \equiv \frac{\partial \varphi'_{ik}(t_n; \overline{X}(t_n) = x)}{\partial x_m},$$

which satisfy for $t_n < T$

$$(2.11) \quad \begin{aligned} \varphi'_{ik}(t_n) &= \partial_i c_j(t_n, \overline{X}(t_n)) \partial_k c_p(t_n, \overline{X}(t_n)) \varphi'_{jp}(t_{n+1}) \\ &\quad + \partial_{ik} c_j(t_n, \overline{X}(t_n)) \varphi_j(t_{n+1}), \\ \varphi'_{ik}(T) &= \partial_{ik} g(\overline{X}(T)), \end{aligned}$$

and

$$(2.12) \quad \begin{aligned} \varphi''_{ikm}(t_n) &= \partial_i c_j(t_n, \overline{X}(t_n)) \partial_k c_p(t_n, \overline{X}(t_n)) \partial_m c_r(t_n, \overline{X}(t_n)) \varphi''_{jpr}(t_{n+1}) \\ &\quad + \partial_{im} c_j(t_n, \overline{X}(t_n)) \partial_k c_p(t_n, \overline{X}(t_n)) \varphi'_{jp}(t_{n+1}) \\ &\quad + \partial_i c_j(t_n, \overline{X}(t_n)) \partial_{km} c_p(t_n, \overline{X}(t_n)) \varphi'_{jp}(t_{n+1}) \\ &\quad + \partial_{ik} c_j(t_n, \overline{X}(t_n)) \partial_m c_p(t_n, \overline{X}(t_n)) \varphi'_{jp}(t_{n+1}) \\ &\quad + \partial_{ikm} c_j(t_n, \overline{X}(t_n)) \varphi_j(t_{n+1}), \\ \varphi''_{ikm}(T) &= \partial_{ikm} g(\overline{X}(T)), \end{aligned}$$

respectively.

The previous result can be directly applied to the particular case of deterministic time steps. The deterministic time stepping algorithm uses the sample average of $\rho(t_n, \overline{X})$ in (2.6), that is the error density becomes

$$(2.13) \quad \overline{\rho}(t_n, \overline{X}) \equiv \frac{1}{M_T} \sum_{j=1}^{M_T} \rho(t_n, \overline{X})(\omega_j),$$

which has variance $\text{Var}[\overline{\rho}_n] = \mathcal{O}(M_T^{-1})$.

3. ADAPTIVE ALGORITHMS FOR SDE

This section presents two adaptive time stepping algorithms and analyzes the basic properties of the adaptive algorithms for the computational problem (1.1). These adaptive algorithms choose adaptively the number of realizations and the time steps, including dividing and merging, to bound the approximation error by a given error tolerance.

The computational error in (1.3) naturally separates into the time discretization error and the statistical error

$$(3.1) \quad \begin{aligned} E[g(X(T))] - \frac{1}{M} \sum_{j=1}^M g(\overline{X}(T; \omega_j)) \\ = (E[g(X(T)) - g(\overline{X}(T))] + \left(E[g(\overline{X}(T))] - \frac{1}{M} \sum_{j=1}^M g(\overline{X}(T; \omega_j)) \right)) \\ \equiv \mathcal{E}_T + \mathcal{E}_S. \end{aligned}$$

The time steps for the realizations of the approximate solution \overline{X} are determined from statistical approximations of the time discretization error, \mathcal{E}_T , and the number, M , of realizations of \overline{X} is determined from the statistical error, \mathcal{E}_S . The statistical

error and the time discretization error are combined in order to bound the computational error (3.1). Therefore we split a given error tolerance TOL into a statistical tolerance, TOL_S , and a time discretization tolerance, TOL_T . The computational work is roughly $\mathcal{O}(N \cdot N) = \mathcal{O}(\text{TOL}_T^{-1} \text{TOL}_S^{-2})$, therefore we use

$$(3.2) \quad \text{TOL}_T = \frac{1}{3}\text{TOL} \quad \text{and} \quad \text{TOL}_S = \frac{2}{3}\text{TOL},$$

by minimizing $\text{TOL}_T^{-1} \text{TOL}_S^{-2}$ under the constraint $\text{TOL}_T + \text{TOL}_S = \text{TOL}$.

Control of the statistical error. For M independent samples $\{Y(\omega_j)\}_{j=1}^M$ of a random variable Y , with $E[|Y|^6] < \infty$, define the sample average $\mathcal{A}(Y; M)$ and the sample standard deviation $\mathcal{S}(Y; M)$ of Y by

$$(3.3) \quad \mathcal{A}(Y; M) \equiv \frac{1}{M} \sum_{j=1}^M Y(\omega_j) \quad \text{and} \quad \mathcal{S}(Y; M) \equiv [\mathcal{A}(Y^2; M) - (\mathcal{A}(Y; M))^2]^{1/2}.$$

Let $\sigma_Y \equiv (E[|Y - E[Y]|^2])^{1/2}$ and consider the random variable

$$Z_M \equiv \frac{\sqrt{M}}{\sigma_Y} (\mathcal{A}(Y; M) - E[Y])$$

with cumulative distribution function $F_{Z_M}(x) \equiv P(Z_M \leq x)$, $x \in \mathbb{R}$. Let $\lambda \equiv (E[|Y - E[Y]|^3])^{1/3} / \sigma_Y < \infty$, then the Berry-Esseen theorem, cf. [10], gives the following estimate in the central limit theorem

$$\sup_{x \in \mathbb{R}} |F_{Z_M}(x) - \Phi(x)| \leq \frac{3}{\sqrt{M}} \lambda^3$$

for the rate of convergence of F_{Z_M} to the distribution function, Φ , of a normal random variable with mean zero and variance one, i.e.

$$(3.4) \quad \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}s^2\right) ds.$$

Since in the examples below M is sufficiently large, i.e. $M \gg 36\lambda^6$, the statistical error

$$\mathcal{E}_S(Y; M) \equiv E[Y] - \mathcal{A}(Y; M)$$

satisfies, by the Berry-Esseen theorem, the following probability approximation

$$P\left(\left|\mathcal{E}_S(Y; M)\right| \leq c_0 \frac{\sigma_Y}{\sqrt{M}}\right) \simeq 2\Phi(c_0) - 1.$$

In practice choose some constant $c_0 \geq 1.65$, so the normal distribution satisfies $1 > 2\Phi(c_0) - 1 \geq 0.901$ and the event

$$(3.5) \quad |\mathcal{E}_S(Y; M)| \leq \mathbf{E}_S(Y; M) \equiv c_0 \frac{\mathcal{S}(Y; M)}{\sqrt{M}}$$

has probability close to one, which involves the additional step to approximate σ_Y by $\mathcal{S}(Y; M)$, cf. [12]. Thus, in the computations $\mathbf{E}_S(Y; M)$ is a good approximation of the statistical error $\mathcal{E}_S(Y; M)$.

For a given $\text{TOL}_S > 0$, the goal is to find M such that $\mathbf{E}_S(Y; M) \leq \text{TOL}_S$. The following algorithm adaptively finds the number of realizations M to compute the sample average $\mathcal{A}(Y; M)$ as an approximation to $E[Y]$. With probability close to one, depending on c_0 , the statistical error in the approximation is then bounded by

TOL_S , see Theorems 3.3 and 3.9. Technical reasons to prove *a.s.* convergence in Lemma 4.4 motivates our choice $M = 2^n$, $n \in \mathbb{N}$.

```

routine Monte-Carlo( $\text{TOL}_S, M_0; EY$ )
  Set the batch counter  $m = 1$ ,  $M[1] = M_0$  and  $E_S[1] = +\infty$ .
  Do while ( $E_S[m] > \text{TOL}_S$ )
    Compute  $M[m]$  new samples of  $Y$ , along with the sample average
     $EY \equiv \mathcal{A}(Y; M[m])$ , the sample variance  $S[m] \equiv \mathcal{S}(Y; M[m])$  and
    the deviation  $E_S[m+1] \equiv E_S(Y, M[m])$ .
    Compute  $M[m+1]$  by change_M ( $M[m], S[m], \text{TOL}_S; M[m+1]$ ).
    Increase  $m$  by 1.
  end-do
end of Monte-Carlo

```

```

routine change_M ( $M_{in}, S_{in}, \text{TOL}_S; M_{out}$ )
  (3.6)  $M^* = \min\left\{\text{integer part}\left(\frac{c_0 S_{in}}{\text{TOL}_S}\right)^2, \text{MCH} \times M_{in}\right\}$ 
   $n = \text{integer part}(\log_2 M^*) + 1$ 
   $M_{out} = 2^n$ .

```

```

end of change_M

```

Here, M_0 is a given initial value for M , and $\text{MCH} > 1$ is a positive integer parameter introduced to avoid a large new number of realizations in the next batch due to a possibly inaccurate sample standard deviation $S[m]$. Indeed, $M[m+1]$ cannot be greater than $\text{MCH} \times M[m]$.

Control of the time discretization error. For given time nodes $0 = t_1 < \dots < t_{N+1} = T$, let the piecewise constant mesh function Δt be determined by

$$\Delta t(\tau) \equiv \Delta t_n \text{ for } \tau \in [t_n, t_{n+1}) \text{ and } n = 1, \dots, N.$$

Then the number of time steps that corresponds to a mesh Δt , for the interval $[0, T]$, is given by

$$(3.7) \quad N(\Delta t) \equiv \int_0^T \frac{1}{\Delta t(\tau)} d\tau.$$

Consider, for $\tau \in [t_n, t_{n+1})$ and $n = 1, \dots, N$, the piecewise constant function

$$(3.8) \quad \rho(\tau) \equiv \text{sign}(\rho_n) \max(|\rho_n|, \delta)$$

which measures the density of the time discretization error, where $\rho_n = \rho(t_n, \bar{X})$ is defined by (2.6) for the stochastic time stepping algorithm or $\rho_n = \bar{\rho}(t_n, \bar{X})$ from (2.13) for the deterministic time stepping algorithm. Here we have $\text{sign}(x) \equiv x/|x|$, for $x \in \mathbb{R} - \{0\}$, $\text{sign}(0) \equiv 1$, and we use the positive parameter, δ , satisfying $\delta + \text{TOL}^2/\delta \rightarrow 0$, in order to guarantee that $\Delta t_{sup} \rightarrow 0$ and that $\rho(t)$ and ρ_n have the same limit, as $\text{TOL}_T \rightarrow 0$. From now on, with a slight abuse of notation, $\rho(t_n) = \rho_n$ denotes the modified density (3.8).

Following the error expansion in Theorem 2.2, the time discretization error is approximated by

$$(3.9) \quad |\mathcal{E}_T| = |E[g(X(T)) - g(\bar{X}(T))]| \lesssim E \left[\sum_{n=1}^N r_n \right]$$

using the error indicator, r_n , defined by

$$(3.10) \quad r_n \equiv |\rho(t_n)|\Delta t_n^2$$

with the modified error density defined by (3.8). To motivate the adaptivity procedure for the time partition, let us now formulate an optimal choice of the time steps by minimizing the expected computational work subject to the accuracy constraint

$$(3.11) \quad E \left[\sum_{n=1}^N r_n \right] \leq \text{TOL}_T.$$

More precisely, solve

$$(3.12) \quad \min E[N(\widetilde{\Delta t})], \text{ such that } \widetilde{\Delta t} \in \mathcal{K} \text{ and } E \left[\sum_{n=1}^{N(\widetilde{\Delta t})} r_n \right] \leq \text{TOL}_T,$$

where \mathcal{K} is the feasible set for the mesh function $\widetilde{\Delta t}$ and $N(\widetilde{\Delta t})$ is the corresponding number of steps. The optimal choice of time steps in \mathcal{K} is based on the given density $\rho_n[k]$, which is piecewise constant on the mesh $\Delta t[k], k = 1, 2, \dots$. The choice of \mathcal{K} determines either deterministic time steps or stochastic time steps. For example, if we let

$$\mathcal{K} \equiv \left\{ \begin{array}{l} \widetilde{\Delta t} \in L_{dt}^2((0, T)) : \widetilde{\Delta t} \text{ is deterministic, positive and} \\ \text{piecewise constant on } \Delta t[k] \end{array} \right\}$$

then the objective function in (3.12) becomes deterministic and a standard application of a Lagrange multiplier shows that the minimizer of the problem (3.12) satisfies

$$(3.13) \quad E[r_n] = \text{constant}, \text{ for all time steps } n,$$

which sets the basis for the refinement procedure with deterministic time steps. On the other hand, letting

$$\mathcal{K} \equiv \left\{ \begin{array}{l} \widetilde{\Delta t} \in L_{dt \times P}^2((0, T) \times \Omega) : \widetilde{\Delta t} \text{ is stochastic, positive and} \\ \text{piecewise constant on } \Delta t[k](\omega) \end{array} \right\}$$

leads to

$$(3.14) \quad r_n(\omega) = \text{constant}, \text{ for all time steps } n \text{ and for all realizations } \omega,$$

which sets the basis for the refinement procedure with stochastic time steps.

Thus, the adaptive algorithm with stochastic time steps uses the optimal conditions (3.11) and (3.14) to construct the mesh which may be different for each realization. On the other hand, the adaptive algorithm with deterministic time steps uses the optimal conditions (3.11) and (3.13) to construct the mesh which is the same for all realizations.

3.1. Convergence rates for stochastic time steps. The optimal conditions (3.11), (3.14) and the restriction (2.2) motivate that the goal of the adaptive algorithm is to construct a time partition Δt of $[0, T]$ for each realization such that

$$(3.15) \quad s_2 \frac{\text{TOL}_T}{E[N]} \leq |\rho_n|\Delta t_n^2 \leq s_1 \frac{\text{TOL}_T}{E[N]}, \quad \forall n = 1, \dots, N$$

where s_1 and s_2 are given constants satisfying $0 < s_2 < s_1$. Note that in practice the quantity $E[N]$ is not known and we can only estimate it by a sample average

$\mathcal{A}(N; M)$ from the previous batch of realizations. The statistical error $|E[N] - \mathcal{A}(N; M)|$ is then bounded by $E_S(N; M)$, with probability close to one, by the same argument as in (3.5). The remainder of this section analyzes an adaptive algorithm based on (3.15) with respect to stopping, accuracy and efficiency.

Let $\bar{N}[j] \equiv \mathcal{A}(N; M[j])$ be the sample average of the final number of time steps in the j -th batch of $M[j]$ numbers of realizations. To achieve (3.15), start with an initial partition $\Delta t[1]$ and then specify iteratively a new partition $\Delta t[k+1]$, from $\Delta t[k]$, using the following dividing and merging strategy: for each realization in the m -th batch and for each time step $n = 1, 2, \dots, N[k]$,

$$(3.16) \quad \text{if } r_n[k] \geq s_1 \frac{\text{TOL}_T}{\bar{N}[m-1]}, \text{ then divide } \Delta t_n[k] \text{ into } H \text{ substeps}$$

$$(3.17) \quad \text{elseif } \max(r_n[k], r_{n+1}[k]) \leq s_2 \frac{\text{TOL}_T}{\bar{N}[m-1]}, \text{ then merge } \Delta t_n[k] \text{ and}$$

$\Delta t_{n+1}[k]$ into one step, and increase n by 1,

else let the new step be the same as the old one.

endif

Here H is a given integer greater than 1, which bounds the increment of the number of time steps from one iteration to the next. The following analysis, for fixed H , can easily be extended to bounded and varying H .

The dividing and merging strategy (3.16, 3.17) motivates the following stopping criteria: for each realization of the m -th batch

$$(3.18) \quad \text{if } \left(r_n[k] < S_1 \frac{\text{TOL}_T}{\bar{N}[m-1]}, \forall n = 1, \dots, N[k] \right) \text{ and}$$

$$(3.19) \quad \left(\max(r_n[k], r_{n+1}[k]) > S_2 \frac{\text{TOL}_T}{\bar{N}[m-1]}, \forall n = 1, \dots, N[k] - 1 \right)$$

then stop the dividing-merging process.

Here S_1 and S_2 are given constants satisfying $0 < S_2 < s_2 < s_1 < S_1$. The combination of (3.9), (3.11) and the upper bound (3.18) asymptotically guarantee a given level of accuracy,

$$|E[g(X(T)) - g(\bar{X}(T))]| < S_1 \text{TOL}_T,$$

while the lower bound (3.19) implies efficiency by almost optimal time steps. When almost all r_n satisfy $r_n < s_1 \text{TOL}_T / \bar{N}$, the reduction of the error may be slow. Therefore the algorithm stops if $\max_n r_n < S_1 \text{TOL}_T / \bar{N}$ and $\min_n (\max(r_n, r_{n+1})) > S_2 \text{TOL}_T / \bar{N}$. What is the right choice for the constants $S_2 < s_2 < s_1 < S_1$ of the dividing and merging strategy? Before determining sufficient conditions for the constants, we will describe the algorithm with stochastic time steps in detail.

Remark 3.1. *In practice, our numerical tests show that*

$$\frac{|E[g(X(T))] - \mathcal{A}(g(\bar{X}(T)); M)|}{\text{TOL}} \approx \frac{s_1}{2} < \frac{1}{3} S_1,$$

so we choose $s_1 = 2$.

The adaptive algorithm. The adaptive, stochastic time stepping algorithm has a structure similar to the `Monte-Carlo` routine. First we split the specified error tolerance by (3.2) and the outer loop computes the batches of realizations of \bar{X} , until

an estimate for the statistical error (3.5) is below the tolerance, TOL_S . In the inner loop for each realization, we apply our dividing and merging strategy (3.16, 3.17) to a given initial time mesh iteratively until the approximate solution is sufficiently resolved, in other words, until the approximate error density and the time steps satisfy the stopping criteria (3.18, 3.19) with a given time discretization tolerance TOL_T . This procedure needs to sample the Wiener process, W , on finer or coarser partitions, given its values on coarser or finer ones respectively. Therefore, the use of Brownian bridges is natural to preserve the required independence between the Wiener increments.

Now we are ready for the detailed definition of the adaptive algorithm with stochastic steps:

Algorithm S

Initialization Choose:

- (1) an error tolerance, $\text{TOL} \equiv \text{TOL}_S + \text{TOL}_T$,
- (2) a number, $N[1]$, of initial uniform steps $\Delta t[1]$ for $[0, T]$ and set $\bar{N} = N[1]$,
- (3) a number, $M[1]$, of initial realizations,
- (4) an integer $H \geq 2$ for the number of subdivisions of a refined time step, a number, $s_1 = 2$ in (3.16) and a rough estimate of c in (3.20) to compute s_2, S_1, S_2 using (3.21, 3.22, 3.23), and
- (5) a constant $c_0 \geq 1.65$ and an integer $\text{MCH} \geq 2$ to determine the number of realizations in (3.6).

Set the two iteration counters, m for batches and k for time refinement levels, to 1. Set the stochastic error $\mathbf{E}_S[1] = +\infty$.

Do while ($\mathbf{E}_S[m] > \text{TOL}_S$)

For realizations $j = 1, \dots, M[m]$

 Set $k = 1$ and $r[1] = +\infty$.

 Start with the initial partition $\Delta t[1]$ and generate $\Delta W[1]$.

Do while ($r[k]$ violates the stopping (3.18, 3.19))

 Compute the approximation $\bar{X}[k]$ and the error indicator $r[k]$ in (3.10) using the error density (2.6) on $\Delta t[k]$ with the known Wiener increments $\Delta W[k]$.

If ($r[k]$ violates the stopping (3.18, 3.19))

For time steps $i = 1, \dots, N[k]$

 Do the dividing and merging process (3.16, 3.17) to compute $\Delta t[k+1]$ from $\Delta t[k]$ and compute $\Delta W[k+1]$ from $\Delta W[k]$ using Brownian bridges.

end-for

end-if

 Increase k by 1.

end-do

end-for

Compute the sample average $Eg \equiv \mathcal{A}(g(\bar{X}(T)); M[m])$, the sample standard deviation $\mathcal{S}[m] \equiv \mathcal{S}(g(\bar{X}(T)); M[m])$ and the a posteriori bound for the statistical error $\mathbf{E}_S[m] \equiv \mathbf{E}_S(g(\bar{X}(T)), M[m])$ in (3.5).

if ($\mathbf{E}_S[m] > \text{TOL}_S$)

 Compute $M[m+1]$ by `change_M`($M[m], \mathcal{S}[m], \text{TOL}_S; M[m+1]$), cf. (3.6), and update $\bar{N} = \mathcal{A}(N; M[m])$, where the random variable N is

the final number of time steps on each realization.

end-if

Increase m by 1.

end-do

Accept Eg as an approximation of $E[g(X(T))]$, since the estimate of the computational error is bounded by TOL.

Note that the dividing and merging strategy in (3.16, 3.17) may do infinite loops at some time steps if the constant s_2 is too close to s_1 . Clearly, we want to avoid the case where the time step $\Delta t(t)[k]$ is divided but in the next iteration, $\Delta t(t)[k+1]$ is merged. An important ingredient to avoid such instability is the proof in Section 4 that the error density converges *a.s.* as the maximal time step tends to zero. Therefore, for sufficiently small initial meshes, which may depend on the realization and small tolerance, there exists a positive constant c such that for all $t \in [0, T]$ and for each realization

$$(3.20) \quad c \leq \left| \frac{\rho(t)[k+1]}{\rho(t)[k]} \right| \leq c^{-1}.$$

Stopping of the adaptive algorithm. The right choice of the parameters $S_2 < s_2 < s_1 < S_1$ is explained by

Theorem 3.2 (Stopping). *Suppose the assumptions of Theorem 2.2 hold and the adaptive algorithm uses the strategy (3.8), (3.16, 3.17), (3.18, 3.19). Assume that (3.20) holds with $c \geq 4^{-1}$, and that*

$$(3.21) \quad s_2 < \frac{c}{H^2} s_1,$$

$$(3.22) \quad S_2 < cs_2,$$

$$(3.23) \quad S_1 > \frac{1}{c} s_1.$$

Then for each realization the adaptive dividing-merging process stops by the stopping criteria (3.18, 3.19), after a finite number of operations.

Proof. If s_2 is too close to s_1 , the algorithm may be unstable in the sense that a time step is first divided and then in the next iteration the step is merged, or similarly a step is first merged and then divided. To analyze this unstable situation assume that (3.20) holds. Merging two steps, where at least one has just been divided requires that

$$(3.24) \quad r_i[k] \geq s_1 \frac{\text{TOL}_T}{\bar{N}}, \quad \text{and} \quad \max(r_i[k+1], r_{i+1}[k+1]) \leq s_2 \frac{\text{TOL}_T}{\bar{N}}$$

with $\Delta t_i[k+1] = \Delta t_i[k]/H$, so that by (3.10) and (3.20)

$$(3.25) \quad \begin{aligned} s_2 &\geq \frac{\bar{N}}{\text{TOL}_T} \frac{|\rho[k+1]|}{|\rho[k]|} \left(\frac{\Delta t[k+1]}{\Delta t[k]} \right)^2 r[k] \\ &\geq \frac{\bar{N}}{\text{TOL}_T} c \frac{1}{H^2} \left(s_1 \frac{\text{TOL}_T}{\bar{N}} \right) = \frac{c}{H^2} s_1. \end{aligned}$$

Similarly, dividing following merging requires $s_2 \geq cs_1/4$, which is less demanding than the smaller (3.25). Therefore the condition (3.21) avoids the dividing-merging instability.

The next step is to understand the evolution of

$$r_{max}[k] \equiv \max_i r_i[k], \text{ and } r_{min}[k] \equiv \min_i \max(r_i[k], r_{i+1}[k])$$

for iterations $k = 1, 2, \dots$. It is advantageous if r_{max} decreases and r_{min} increases quickly to levels close to the bounds $s_1 \text{TOL}_T / \bar{N}$ and $s_2 \text{TOL}_T / \bar{N}$, respectively. Indeed, there holds

$$(3.26) \quad r_{max}[k+1] < \frac{c^{-1}}{H^2} r_{max}[k],$$

provided $r_{max}[k] > H^2 s_1 \text{TOL}_T / \bar{N}$, and

$$(3.27) \quad r_{min}[k+1] > 4c r_{min}[k],$$

provided $r_{min}[k] < s_2 \text{TOL}_T / (4\bar{N})$. In other words, the error indicators r_{max} and r_{min} decrease and increase, respectively, with a constant factor away from the bounds

$$r_{max} > H^2 s_1 \text{TOL}_T / \bar{N} \text{ and } r_{min} < s_2 \text{TOL}_T / (4\bar{N}).$$

If r_{max} is close to the dividing bound, i.e. if

$$(3.28) \quad r_{max}[k] \leq H^2 s_1 \text{TOL}_T / \bar{N},$$

then in the next iteration $r_{max}[k+1] \leq c^{-1} s_1 \text{TOL}_T / \bar{N}$, since a divided step cannot be merged by (3.21). Therefore $r_{max}[k+1]$ enters the stopping region, $r_{max} < S_1 \text{TOL}_T / \bar{N}$, provided $S_1 > s_1/c$ and the condition (3.28) remains satisfied for the later iterations provided $c^{-1} \leq H^2$.

Similarly, if r_{min} is close to the merging bound, i.e. if

$$(3.29) \quad r_{min}[k] \geq s_2 \text{TOL}_T / (4\bar{N}),$$

then $r_{min}[k+1] \geq c s_2 \text{TOL}_T / \bar{N}$. Therefore r_{min} belongs to the stopping region, $r_{min}[k+1] > S_2 \text{TOL}_T / \bar{N}$, provided $S_2 < c s_2$ and the condition (3.29) remains satisfied for the later iterations provided $c \geq 4^{-1}$. When both r_{max} and r_{min} have entered their stopping regions, the dividing-merging process stops, and the algorithm continues with the next realization. \square

Accuracy of the adaptive algorithm. The adaptive algorithm guarantees that the estimated error is bounded by $S_1 \text{TOL}_T + \text{TOL}_S = (\frac{S_1}{3} + \frac{2}{3}) \text{TOL}$. The next question is whether the true error is bounded by $(\frac{S_1}{3} + \frac{2}{3}) \text{TOL}$ asymptotically. Using the upper bound (3.18) of the error indicators and the *a.s.* convergence of ρ , proved in Section 4, the approximate error has the estimate

Theorem 3.3 (Accuracy). *Suppose that the assumptions of Theorem 2.2 hold. Then the adaptive algorithm (3.8), (3.16, 3.17), (3.18, 3.19) satisfies, for any constant $c_0 > 0$ defined in (3.6),*

$$(3.30) \quad \liminf_{\text{TOL} \rightarrow 0} P \left(\frac{|E[g(X(T))] - \mathcal{A}(g(\bar{X}(T)); M)|}{\text{TOL}} \leq \frac{S_1}{3} + \frac{2}{3} \right) \geq \int_{-c_0}^{c_0} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

Proof. To simplify the proof, we split the fraction on the left hand side of (3.30) into a statistical error and a time discretization error, for all $s \in \mathbb{R}_+$,

$$\begin{aligned}
(3.31) \quad & \liminf_{\text{TOL} \rightarrow 0} P \left(\frac{|E[g(X(T)) - \mathcal{A}(g(\bar{X}(T)); M)]|}{\text{TOL}} \leq \frac{S_1}{3} + \frac{2}{3} \right) \\
& \geq \liminf P \left(\frac{|E[g(X(T)) - g(\bar{X}(T))]|}{\text{TOL}} + \frac{|E[g(\bar{X}(T)) - \mathcal{A}(g(\bar{X}(T)); M)]|}{\text{TOL}} \leq \frac{S_1}{3} + \frac{2}{3} \right) \\
& \geq \liminf P \left(\frac{|E[g(X(T)) - g(\bar{X}(T))]|}{\text{TOL}} \leq s \right. \\
& \quad \left. \text{and } \frac{|E[g(\bar{X}(T)) - \mathcal{A}(g(\bar{X}(T)); M)]|}{\text{TOL}} \leq \frac{S_1}{3} + \frac{2}{3} - s \right) \\
& = \liminf \left(P \left(\frac{|E[g(X(T)) - g(\bar{X}(T))]|}{\text{TOL}} \leq s \right) \times \right. \\
& \quad \left. P \left(\frac{|E[g(\bar{X}(T)) - \mathcal{A}(g(\bar{X}(T)); M)]|}{\text{TOL}} \leq \frac{S_1}{3} + \frac{2}{3} - s \right) \right).
\end{aligned}$$

The time discretization error. When the adaptive algorithm stops, the error estimate (2.1) and the stopping bound for Δt (3.18) imply

$$\begin{aligned}
|E[g(X(T)) - g(\bar{X}(T))]| & \leq E \left[\sum_{i=1}^N \Delta t_i \int_{t_{i-1}}^{t_i} |\tilde{\rho}(\tau)| d\tau \right] \\
& < \sqrt{S_1 \text{TOL}_T} E \left[\frac{1}{\sqrt{N}} \sum_{i=1}^N \int_{t_{i-1}}^{t_i} \frac{|\tilde{\rho}(\tau)|}{\sqrt{|\rho(\tau)|}} d\tau \right].
\end{aligned}$$

Apply the dominated convergence theorem to the right hand side of the above equation to get

$$\begin{aligned}
\limsup_{\text{TOL} \rightarrow 0} \frac{|E[g(X(T)) - g(\bar{X}(T))]|}{\text{TOL}_T} & \leq \sqrt{S_1} E \left[\limsup_{\text{TOL} \rightarrow 0} \left(\frac{\int_0^T |\tilde{\rho}(\tau)| / \sqrt{|\rho(\tau)|} d\tau}{\sqrt{\text{TOL}_T N}} \right) \right] \\
& \leq \sqrt{S_1} E \left[\frac{1}{\sqrt{\liminf_{\text{TOL} \rightarrow 0} \text{TOL}_T N}} \limsup_{\text{TOL} \rightarrow 0} \int_0^T \frac{|\tilde{\rho}(\tau)|}{\sqrt{|\rho(\tau)|}} d\tau \right].
\end{aligned}$$

Corollary 4.3 in Section 4 shows that both $\tilde{\rho}$ and ρ converge *a.s.* to the limit $\hat{\rho}$, so that dominated convergence, based on (3.8), establishes

$$(3.32) \quad \lim_{\text{TOL} \rightarrow 0} \int_0^T \frac{|\tilde{\rho}(\tau)|}{\sqrt{|\rho(\tau)|}} d\tau = \int_0^T \sqrt{|\hat{\rho}(\tau)|} d\tau \text{ a.s.}$$

since $|\tilde{\rho}|/\sqrt{|\rho|}$ is uniformly bounded in L^1 using the facts (2.5) and (3.8).

Rewrite the inequality (3.18) as

$$\sqrt{|\rho|} < \sqrt{S_1 \frac{\text{TOL}_T}{N[m-1]} \frac{1}{\Delta t_i}},$$

integrate both sides, use the definition (3.7) and take the sample average of $M[m]$ independent samples to obtain

$$(3.33) \quad \bar{N}[m] \geq \frac{\sqrt{\bar{N}[m-1]}}{\sqrt{S_1 \text{TOL}_T}} \mathcal{A} \left(\int_0^T \sqrt{|\rho(\tau)|} d\tau; M[m] \right).$$

Take the limit in (3.33) and use that Corollary 4.3 implies

$$\lim_{\text{TOL} \rightarrow 0} \mathcal{A} \left(\int_0^T \sqrt{|\rho(\tau)|} d\tau; M[m] \right) = E \left[\int_0^T \sqrt{|\hat{\rho}(\tau)|} d\tau \right]$$

to get

$$\sqrt{\liminf_{\text{TOL} \rightarrow 0} \text{TOL}_T \bar{N}} \geq \frac{1}{\sqrt{S_1}} E \left[\int_0^T \sqrt{|\hat{\rho}(\tau)|} d\tau \right],$$

and consequently

$$\limsup_{\text{TOL} \rightarrow 0} \frac{|E[g(X(T)) - g(\bar{X}(T))]|}{\text{TOL}_T} \leq S_1.$$

Since $\text{TOL}_T = \frac{1}{3} \text{TOL}$, this deterministic limit implies that for all $s > S_1/3$

$$(3.34) \quad \liminf_{\text{TOL} \rightarrow 0} P \left(\frac{|E[g(X(T)) - g(\bar{X}(T))]|}{\text{TOL}} \leq s \right) = 1.$$

The statistical error. Use that in (3.6) the number of realizations is

$$M \geq \left(\frac{c_0 \mathcal{S}(g(\bar{X}(T)); M)}{\text{TOL}_s} \right)^2$$

to rewrite $1/\text{TOL} \leq (2/3c_0) \sqrt{M}/\mathcal{S}(g(\bar{X}(T)); M)$ by (3.2), so that

$$\begin{aligned} & P \left(\frac{|E[g(\bar{X}(T))] - \mathcal{A}(g(\bar{X}(T)); M)|}{\text{TOL}} \leq \frac{S_1}{3} + \frac{2}{3} - s \right) \\ & \geq P \left(\frac{|E[g(\bar{X}(T))] - \mathcal{A}(g(\bar{X}(T)); M)|}{\mathcal{S}(g(\bar{X}(T)); M)} \sqrt{M} \leq \frac{3c_0}{2} \left(\frac{S_1}{3} + \frac{2}{3} - s \right) \right). \end{aligned}$$

Let $\chi \equiv (E[g(X(T))] - \mathcal{A}(g(X(T)); M))/\mathcal{S}(g(X(T)); M)$ and $\bar{\chi} \equiv (E[g(\bar{X}(T))] - \mathcal{A}(g(\bar{X}(T)); M))/\mathcal{S}(g(\bar{X}(T)); M)$. Then, write

$$\bar{\chi} = (\bar{\chi} - \chi) + \chi$$

and note that $\bar{\chi} - \chi \rightarrow 0$, weakly as $\Delta t \rightarrow 0$, and that $\mathcal{S}(g(X(T)); M) \rightarrow \sigma_{g(X(T))}$, strongly as $M \rightarrow \infty$. Therefore the central limit theorem shows that $\lim_{\text{TOL} \rightarrow 0} \bar{\chi}$ is normally distributed with mean zero and variance one and consequently

$$(3.35) \quad \begin{aligned} & \sup_{s > \frac{S_1}{3}} \lim_{\text{TOL} \rightarrow 0} P \left(\frac{|E[g(\bar{X}(T))] - \mathcal{A}(g(\bar{X}(T)); M)|}{\mathcal{S}(g(\bar{X}(T)); M)} \sqrt{M} \leq \frac{3c_0}{2} \left(\frac{S_1}{3} + \frac{2}{3} - s \right) \right) \\ & = \sup_{s > \frac{S_1}{3}} \int_{-\frac{3c_0}{2} \left(\frac{S_1}{3} + \frac{2}{3} - s \right)}^{\frac{3c_0}{2} \left(\frac{S_1}{3} + \frac{2}{3} - s \right)} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \int_{-c_0}^{c_0} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx. \end{aligned}$$

Finally the combination of (3.31), (3.34) and (3.35) proves the theorem. \square

Efficiency of the adaptive algorithm. The minimal expected number of time steps in the class of stochastic time steps $\Delta t(t, \omega)$ depends on the stochastic data and the individual realizations \overline{X} through the constraint $E[g(X(T)) - g(\overline{X}(T))] = \text{TOL}_T$. The conditions (3.11) and (3.14) imply that the optimal expected number of time steps, $E[N_S]$, satisfies (cf. [31])

$$(3.36) \quad E[N_S] = \frac{1}{\text{TOL}_T} \left(E \left[\int_0^T \sqrt{|\rho(\tau)|} d\tau \right] \right)^2.$$

On the other hand, for the constant time steps $\Delta t = \text{constant}$, the number of steps, N_C , to achieve $\sum_{i=1}^N |\rho_i| \Delta t_i^2 = \text{TOL}_T$, becomes

$$(3.37) \quad N_C = \frac{T}{\text{TOL}_T} \left(\int_0^T E[|\rho(\tau)|] d\tau \right).$$

Therefore, Jensen's inequality shows that the adaptive method with stochastic time steps uses fewer time steps than the method with a constant time step, i.e.,

$$(3.38) \quad E[N_S] \leq N_C.$$

The following theorem uses the lower bound (3.19) of the error indicators to show that the algorithm (3.16 - 3.19) generates a mesh which is optimal, up to a multiplicative constant $C = 4/S_2$.

Theorem 3.4 (Efficiency). *Suppose that the assumptions of Theorem 2.2 hold. Then there exists a constant $C > 0$, asymptotically bounded by $4/S_2$, such that, using a final batch of $M[m]$ realizations, the final sample average*

$$\overline{N}[m] \equiv \mathcal{A}(N; M[m])$$

of the number of adaptive steps of the algorithm (3.16, 3.17), (3.18, 3.19) satisfies

$$(3.39) \quad \text{TOL}_T \frac{\overline{N}[m]^2}{\overline{N}[m-1]} < C \left(\mathcal{A} \left(\int_0^T \sqrt{|\rho(\tau)|} d\tau; M[m] \right) \right)^2,$$

and asymptotically

$$(3.40) \quad \limsup_{\text{TOL}_T \rightarrow 0} \text{TOL}_T E[N] \leq \frac{4}{S_2} \left(E \left[\int_0^T \sqrt{|\hat{\rho}(\tau)|} d\tau \right] \right)^2.$$

Proof. For each realization, Theorem 3.3 shows that the adaptive dividing-merging stops, by the stopping criteria (3.18, 3.19), and the set of time steps satisfies,

$$\{\Delta t_i : i = 1, \dots, N\} = \mathcal{D} \cup \mathcal{M} \text{ and } \mathcal{D} \cap \mathcal{M} = \emptyset$$

where

$$(3.41) \quad \mathcal{D} \equiv \left\{ \Delta t_i : S_2 \frac{\text{TOL}_T}{\overline{N}[m-1]} < |\rho_i| \Delta t_i^2 < S_1 \frac{\text{TOL}_T}{\overline{N}[m-1]}, i = 1, \dots, N \right\}$$

$$(3.42) \quad \mathcal{M} \equiv \left\{ \Delta t_i : |\rho_i| \Delta t_i^2 \leq S_2 \frac{\text{TOL}_T}{\overline{N}[m-1]}, i = 1, \dots, N \right\}.$$

Condition (3.19) shows that there is no successive pair of time steps which belongs to the set \mathcal{M} . This means that the number of steps $N_{\mathcal{M}}$ in \mathcal{M} is *at most* the same as the half of the number of steps, N , i.e. $N_{\mathcal{M}} \leq \lceil \frac{N}{2} \rceil$ where $\lceil a \rceil$ rounds the number

a to the nearest integer greater than or equal to a . Then for each realization, the number of steps satisfies

$$(3.43) \quad N[m] = N_{\mathcal{D}} + N_{\mathcal{M}} \leq N_{\mathcal{D}} + \left\lceil \frac{N[m]}{2} \right\rceil \leq 2N_{\mathcal{D}}[m] + 1 \leq 3N_{\mathcal{D}}[m].$$

Take the sample average of both sides of $M[m]$ independent samples to get

$$(3.44) \quad \frac{1}{3}\bar{N}[m] \leq \bar{N}_{\mathcal{D}}[m].$$

The time steps $\Delta t_i \in \mathcal{D}$ satisfy

$$(3.45) \quad \sqrt{S_2 \frac{\text{TOL}_T}{\bar{N}[m-1]} \frac{1}{\Delta t_i}} < \sqrt{|\rho|},$$

and by integrating both sides with the definition (3.7) and taking the sample average, we get

$$(3.46) \quad \sqrt{S_2 \frac{\text{TOL}_T}{\bar{N}[m-1]} \bar{N}_{\mathcal{D}}[m]} < \mathcal{A} \left(\int_0^T \sqrt{|\rho(\tau)|} d\tau; M \right).$$

The estimate (3.44) shows

$$\sqrt{S_2 \frac{\text{TOL}_T}{\bar{N}[m-1]} \frac{\bar{N}[m]}{3}} \leq \sqrt{S_2 \frac{\text{TOL}_T}{\bar{N}[m-1]} \bar{N}_{\mathcal{D}}[m]},$$

which together with (3.46) proves (3.39) where $C = 9/S_2$. A more precise use of the last inequality in (3.43) shows $C \rightarrow 4/S_2$, as $N \rightarrow \infty$.

To prove (3.40), apply dominated convergence theorem to (3.39) and use Corollary 4.3 to get

$$(3.47) \quad \limsup_{\text{TOL}_T \rightarrow 0} \text{TOL}_T \bar{N} \leq \frac{4}{S_2} \left(E \left[\int_0^T \sqrt{|\hat{\rho}(\tau)|} d\tau \right] \right)^2.$$

Integrate (3.45) and take the expected value to obtain

$$(3.48) \quad \begin{aligned} \limsup_{\text{TOL}_T \rightarrow 0} \text{TOL}_T E[N[m]] \\ \leq \sqrt{\frac{4}{S_2}} \limsup_{\text{TOL}_T \rightarrow 0} E \left[\sqrt{\text{TOL}_T \bar{N}[m-1]} \int_0^T \sqrt{|\rho(\tau)|} d\tau \right]. \end{aligned}$$

Dominated convergence, the *a.s.* convergence of error density, $\rho \rightarrow \hat{\rho}$, and the bound (3.47) imply

$$\begin{aligned} \limsup_{\text{TOL}_T \rightarrow 0} \text{TOL}_T E[N] &\leq \sqrt{\frac{4}{S_2}} E \left[\limsup_{\text{TOL}_T \rightarrow 0} \left(\sqrt{\text{TOL}_T \bar{N}[m-1]} \int_0^T \sqrt{|\rho(\tau)|} d\tau \right) \right] \\ &\leq \sqrt{\frac{4}{S_2}} E \left[\limsup_{\text{TOL}_T \rightarrow 0} \sqrt{\text{TOL}_T \bar{N}[m-1]} \int_0^T \sqrt{|\hat{\rho}(\tau)|} d\tau \right] \\ &\leq \frac{4}{S_2} \left(E \left[\int_0^T \sqrt{|\hat{\rho}(\tau)|} d\tau \right] \right)^2, \end{aligned}$$

which proves (3.40). \square

Remark 3.5. *The error density condition (3.20) also implies constraints on the optimal mesh, for instance $H = 2$ and the assumption $\frac{1}{2}(\rho_i[k] + \rho_{i+1}[k]) = \rho(t_i)[k-1]$ show that*

$$(3.49) \quad 2c - 1 \leq \left| \frac{\rho_{i+1}[k]}{\rho_i[k]} \right| \leq 2c^{-1} - 1.$$

Remark 3.6. *If the number of elements in each refinement iteration increase only very slowly, the total work including all refinement levels becomes proportional to the product of the number of steps in the finest mesh times the number of refinement levels, J , which satisfies $\min \Delta t = H^{-J}T/N[1] = \mathcal{O}(\text{TOL})$, so that*

$$J = \mathcal{O}(\log(\text{TOL}^{-1})) = \mathcal{O}(\log N).$$

Therefore the average of the total number of time steps is essentially bounded by $E[N \log N]$.

3.2. Convergence rates for deterministic time steps. The main difference between the stochastic and the deterministic time step algorithm is that the additional work to find the optimal deterministic steps requires a much smaller number, M_T , of realizations than total number of realizations M . The approximation of the time discretization error in the right hand side of (3.9) can be separated into two parts

$$(3.50) \quad E \left[\sum_{n=1}^N r_n \right] \leq \mathcal{A} \left(\sum_{n=1}^N r_n; M_T \right) + \left| E \left[\sum_{n=1}^N r_n \right] - \mathcal{A} \left(\sum_{n=1}^N r_n; M_T \right) \right|,$$

where the second error term in the right hand side of (3.50) is approximated by

$$(3.51) \quad \left| E \left[\sum_{n=1}^N r_n \right] - \mathcal{A} \left(\sum_{n=1}^N r_n; M_T \right) \right| \lesssim E_{TS} \equiv c_0 \frac{\mathcal{S} \left(\sum_{n=1}^N r_n; M_T \right)}{\sqrt{M_T}}$$

and the first term defines $E_{TT} \equiv \mathcal{A} \left(\sum_{n=1}^N r_n; M_T \right)$. Then for a given $\text{TOL}_T > 0$, the goal is to construct a partition Δt of $[0, T]$, with as few time steps and realizations M_T as possible, such that

$$E_{TT} + E_{TS} \leq \text{TOL}_T.$$

To this end, first split the time discretization tolerance TOL_T in two positive parts TOL_{TT} and TOL_{TS} for E_{TT} and E_{TS} respectively. The statistical error of the time discretization using the density (2.13) is $\mathcal{O}(\Delta t_{max}/\sqrt{M_T})$. Therefore the percentage of the tolerance, TOL , devoted to the control of the statistical time discretization error can be arbitrary small as $\text{TOL} \rightarrow 0$. In practice we choose

$$(3.52) \quad \text{TOL}_{TT} = \frac{2}{3}\text{TOL}_T = \frac{2}{9}\text{TOL} \text{ and } \text{TOL}_{TS} = \frac{1}{3}\text{TOL}_T = \frac{1}{9}\text{TOL}.$$

The control of the statistical time discretization error determines the number of realizations M_T to ensure a reliable choice of the time discretization in the deterministic time stepping algorithm.

Take into account (3.8) and define the error density, $\bar{\rho}$, by

$$E_{TT} \equiv \mathcal{A} \left(\sum_{n=1}^N r_n; M_T \right) = \sum_{n=1}^N \bar{\rho}_n \Delta t_n^2.$$

Following the optimal conditions (3.11) and (3.13), the goal of the adaptive algorithm described below is to construct a time partition Δt of $[0, T]$ such that

$$(3.53) \quad d_2 \frac{\text{TOL}_{TT}}{N} \leq |\bar{\rho}_n| \Delta t_n^2 \leq d_1 \frac{\text{TOL}_{TT}}{N}, \quad \forall n = 1, \dots, N$$

where d_1 and d_2 are given constants satisfying $0 < d_2 < d_1$. This section analyzes an adaptive algorithm based on (3.53), with respect to stopping, accuracy and efficiency.

To achieve (3.53), start as in Section 3.1 with an initial partition $\Delta t[1]$ and then specify iteratively a new partition $\Delta t[k+1]$, from $\Delta t[k]$, using the following dividing and merging strategy: for $n = 1, 2, \dots, N[k]$ let $\bar{r}_n \equiv |\bar{\rho}_n| \Delta t_n^2$ and

$$(3.54) \quad \text{if } \bar{r}_n[k] \geq d_1 \frac{\text{TOL}_{TT}}{N[k]}, \text{ then divide } \Delta t_n[k] \text{ into } H \text{ substeps}$$

$$(3.55) \quad \begin{aligned} &\text{elseif } \max(\bar{r}_n[k], \bar{r}_{n+1}[k]) \leq d_2 \frac{\text{TOL}_{TT}}{N[k]}, \text{ then merge } \Delta t_n[k] \text{ and} \\ &\quad \Delta t_{n+1}[k] \text{ into one step, and increment } n \text{ by 1,} \\ &\text{else let the new step be the same as the old one.} \\ &\text{endif} \end{aligned}$$

until the following stopping criteria is satisfied:

$$(3.56) \quad \text{if } \left(\bar{r}_n[k] < D_1 \frac{\text{TOL}_{TT}}{N[k]}, \quad \forall n = 1, \dots, N[k] \right) \text{ and}$$

$$(3.57) \quad \left(\max(\bar{r}_n[k], \bar{r}_{n+1}[k]) > D_2 \frac{\text{TOL}_{TT}}{N[k]}, \quad \forall n = 1, \dots, N[k] - 1 \right)$$

then the dividing-merging process stops.

Here D_1 and D_2 are given constants satisfying $0 < D_2 < d_2 < d_1 < D_1$. The combination of (3.50) and (3.56) asymptotically guarantees a given level of accuracy, $\mathbf{E}_{TT} < D_1 \text{TOL}_{TT}$ and the lower bound (3.57) implies efficiency by almost optimal time steps. The positive numbers D_1 and D_2 are motivated to avoid slow convergence in case almost all \bar{r}_n satisfy (3.56, 3.57), as in Section 3.1. Before determining the sufficient conditions for the constants $D_2 < d_2 < d_1 < D_1$, we will describe the algorithm with deterministic time steps in detail.

Remark 3.7. *In practice, our numerical tests show that*

$$\frac{|E[g(X(T))] - \mathcal{A}(g(\bar{X}(T)); M)|}{\text{TOL}} \approx \frac{d_1}{2} < \frac{2}{9} D_1,$$

so we choose $d_1 = 2$.

The adaptive algorithm. First we split the specified error tolerance into three parts, TOL_S , TOL_{TT} and TOL_{TS} by (3.2) and (3.52). The first loop below determines the mesh with M_T realizations by changing iteratively the time steps using our dividing and merging strategy (3.54, 3.55) until $|\bar{\rho}|$ and Δt satisfy the stopping criteria (3.56, 3.57) and the statistical error estimate $\mathbf{E}_{TS} \leq \text{TOL}_{TS}$. Then, the second loop with fixed mesh chooses the number M of realizations, using (3.6) until $\mathbf{E}_S \leq \text{TOL}_S$ holds.

Now we are ready for the detailed definition of the adaptive algorithm with deterministic steps:

Algorithm D**Initialization** Choose:

- (1) an error tolerance, $\text{TOL} \equiv \text{TOL}_S + \text{TOL}_{TT} + \text{TOL}_{TS}$,
- (2) a number, $N[1]$, of initial uniform steps $\Delta t[1]$ for $[0, T]$,
- (3) a number, $M[1]$, of initial realizations and set $M_T[1] = M[1]$,
- (4) an integer $H \geq 2$ for the number of subdivisions of a refined time step, a number, $d_1 = 2$ in (3.54) and a rough estimate of c in (3.58) to compute d_2, D_1, D_2 using (3.59, 3.60, 3.61), and
- (5) a constant $c_0 \geq 1.65$ and an integer $\text{MCH} \geq 2$ to determine the number of realizations in (3.6).

Set the iteration counter, k , for time refinement levels, to 1 and set the statistical error, $\text{E}_{TS} = +\infty$ and $\bar{r}[1] = +\infty$.

Do while ($\bar{r}[k]$ violates the stopping (3.56, 3.57) or $\text{E}_{TS} > \text{TOL}_{TS}$)

 Compute, with the same mesh $\Delta t[k]$, sample averages and error estimates by calling **Euler**.

If ($\bar{r}[k]$ violates the stopping (3.56, 3.57))

 For all time steps $i = 1, \dots, N[k]$, do the dividing and merging process (3.54, 3.55) to compute $\Delta t[k+1]$ from $\Delta t[k]$.

elseif ($\text{E}_{TS} > \text{TOL}_{TS}$)

 Compute $M_T[k+1]$ by **change_M** ($M_T[k], \mathcal{V}_{TS}[k], \text{TOL}_{TS}; M_T[k+1]$).

 Set $\Delta t[k+1] = \Delta t[k]$.

end-if

 Increase k by 1.

end-do

Compute an approximation, Eg , for $E[g(\bar{X}(T))]$ with fixed time mesh $\Delta t = \Delta t[k]$ by **Monte-Carlo**($M_T[k], \text{TOL}_S; Eg$).

Accept Eg as an approximation of $E[g(X(T))]$, since the estimate of the computational error is bounded by TOL .

routine Euler

 Compute $M_T[k]$ **new** realizations of the Euler method with the same partition $\Delta t[k]$ and update the approximations of the time discretization error indicators $\bar{r}[k]$ and the statistical time discretization error $\text{E}_{TS}[k]$.

end-of-Euler

Stopping of the adaptive algorithm. The right choice of the parameters $D_2 < d_2 < d_1 < D_1$ is explained by

Theorem 3.8 (Stopping). *Suppose the assumptions of Theorem 2.2 hold and the adaptive algorithm uses the strategy (3.8), (3.54, 3.55), (3.56, 3.57). Assume that there exists a positive constant $c \geq 2^{-1}$ such that for all $t \in [0, T]$ and for each realization*

$$(3.58) \quad c \leq \left| \frac{\bar{\rho}(t)[k+1]}{\bar{\rho}(t)[k]} \right| \leq c^{-1}$$

and that

$$(3.59) \quad d_2 < \frac{c}{2H^2}d_1,$$

$$(3.60) \quad D_2 < \frac{c}{2}d_2,$$

$$(3.61) \quad D_1 > \frac{H}{c}d_1.$$

Then the adaptive dividing-merging process stops by the stopping criteria (3.56, 3.57), after a finite number of operations.

Proof. The proof is the same as for Theorem 3.2 in [25], which is also similar to Theorem 3.2. \square

Accuracy of the adaptive algorithm. The adaptive algorithm guarantees that the estimated error is bounded by a given error tolerance, $D_1 \text{TOL}_T + \text{TOL}_S = (\frac{2}{9}D_1 + \frac{2}{3})\text{TOL}$. By applying the same arguments as in the proof of the Theorem 3.3, the true error for the deterministic time stepping is also bounded by $(\frac{2}{9}D_1 + \frac{2}{3})\text{TOL}$ asymptotically, using the upper bound (3.56) of the error indicators and the *a.s.* convergence of $|\bar{\rho}|$ in Section 4.

Theorem 3.9 (Accuracy). *Suppose that the assumptions of Theorem 2.2 hold. Then the adaptive algorithm (3.8), (3.54, 3.55), (3.56, 3.57) satisfies, for any constant $c_0 > 0$ defined in (3.6)*

$$(3.62) \quad \liminf_{\text{TOL} \rightarrow 0} P \left(\frac{|E[g(X(T))] - \mathcal{A}(g(\bar{X}(T))); M|}{\text{TOL}} \leq \frac{2}{9}D_1 + \frac{2}{3} \right) \geq \int_{-c_0}^{c_0} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

Proof. By applying the same arguments as in the proof of Theorem 3.3, we obtain, for the time discretization error,

$$(3.63) \quad \frac{|E[g(X(T)) - g(\bar{X}(T))]|}{\text{TOL}_{TT}} < D_1 E \left[\frac{\int_0^T |\tilde{\rho}(\tau)| / \sqrt{|\bar{\rho}(\tau)|} d\tau}{\int_0^T \sqrt{|\bar{\rho}(\tau)|} d\tau} \right].$$

Corollary 4.3 and Theorem 2.2 show that $\tilde{\rho}$ converges *a.s.* to the limit $\hat{\rho}$ and Theorem 4.5 proves that the sample average, $\bar{\rho}$, converges to $E[\hat{\rho}]$. Therefore dominated convergence shows that the fraction in the right hand side of (3.63) converges to 1, i.e.,

$$\begin{aligned} \lim_{\text{TOL} \rightarrow 0} E \left[\frac{\int_0^T |\tilde{\rho}(\tau)| / \sqrt{|\bar{\rho}(\tau)|} d\tau}{\int_0^T \sqrt{|\bar{\rho}(\tau)|} d\tau} \right] &= E \left[\lim_{\text{TOL} \rightarrow 0} \frac{\int_0^T |\tilde{\rho}(\tau)| / \sqrt{|\bar{\rho}(\tau)|} d\tau}{\int_0^T \sqrt{|\bar{\rho}(\tau)|} d\tau} \right] \\ &= E \left[\frac{\int_0^T |\hat{\rho}(\tau)| / \sqrt{|E[\hat{\rho}(\tau)]|} d\tau}{\int_0^T \sqrt{|E[\hat{\rho}(\tau)]|} d\tau} \right] = 1. \end{aligned}$$

This deterministic limit implies that for all $s > \frac{2}{9}D_1$

$$(3.64) \quad \liminf_{\text{TOL} \rightarrow 0} P \left(\frac{|E[g(X(T)) - g(\bar{X}(T))]|}{\text{TOL}} \leq s \right) = 1,$$

and consequently the combination (3.64) and (3.35) proves (3.62). \square

Efficiency of the adaptive algorithm. Within the class of deterministic time steps, the conditions (3.11) and (3.13) give the minimal number of steps

$$(3.65) \quad N_D = \frac{1}{\text{TOL}_T} \left(\int_0^T \sqrt{E[|\bar{\rho}(\tau)|]} d\tau \right)^2.$$

Using the definitions (3.36) and (3.37) of the minimal expected number, $E[N_S]$, of stochastic time steps and the number, N_C , of constant time steps respectively, we obtain

$$(3.66) \quad E[N_S] \leq N_D \leq N_C$$

by applying Jensen's inequality.

Similar to Theorem 3.4 with N instead of \bar{N} , we prove the following theorem using the lower bound (3.57) of the error indicators to show that the algorithm (3.54 - 3.57) generates a mesh which is optimal, up to a multiplicative constant $C = 6/D_2$.

Theorem 3.10 (Efficiency). *Suppose that the assumptions of Theorem 2.2 hold. Then there exists a constant $C > 0$, asymptotically bounded by $6/D_2$, such that the number, N , of adaptive steps of the algorithm (3.8) (3.54, 3.55), (3.56, 3.57) satisfies*

$$(3.67) \quad \text{TOL}_T N < C \left(\int_0^T \sqrt{|\bar{\rho}(\tau)|} d\tau \right)^2.$$

4. ALMOST SURE CONVERGENCE OF THE ERROR DENSITY

This subsection proves pathwise *a.s.* convergence of the error density using *a.s.* convergence of the approximate solution of the Monte Carlo Euler method based on the adaptive algorithms presented in Section 3. Before presenting the main result, we extend the Euler method, for theoretical purposes only, to $t \in [0, T]$ by

$$(4.1) \quad \bar{X}(t) - \bar{X}(0) = \int_0^t \bar{a}(s; \bar{X}) ds + \sum_{\ell=1}^{\ell_0} \int_0^t \bar{b}^\ell(s; \bar{X}) dW^\ell(s)$$

where \bar{a} and \bar{b}^ℓ are the piecewise constant approximations, see Figure 1,

$$\bar{a}(s; \bar{X}) \equiv a(t_n, \bar{X}(t_n)) \quad \text{and} \quad \bar{b}^\ell(s; \bar{X}) \equiv b^\ell(t_n, \bar{X}(t_n)) \quad \text{for } s \in [t_n, t_{n+1}).$$

4.1. Stochastic time steps. Let us now for $H = 2$ define the maximum step size for a fixed realization by

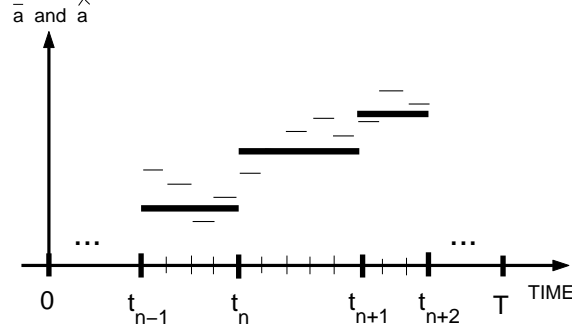
$$(4.2) \quad \Delta t_{sup}(\eta) \equiv \sup_{1 \leq n \leq N} \Delta t_n(\eta) = T2^{-\eta},$$

for a positive integer η . Theorem 4.1 and Lemma 4.2 below show that the approximate solution converges *a.s.* to the correct limit.

Theorem 4.1. *Suppose that a, b, g and X satisfy the assumptions in Lemma 2.1. with the maximum time step size $\Delta t_{sup}(\eta)$ in (4.2). Then for any $\alpha < 1/2$*

$$(4.3) \quad \lim_{\eta \rightarrow \infty} \Delta t_{sup}^{-\alpha}(\eta) \sup_{t \in [0, T]} |X(t) - \bar{X}(t)| = 0 \quad \text{a.s.},$$

where $\Delta t_{sup}^\beta(\eta) \equiv (\Delta t_{sup}(\eta))^\beta$ for $\beta \in \mathbb{R}$.


 FIGURE 1. The piecewise constant approximations \bar{a} and \hat{a} .

Proof. To simplify the proof, let us introduce the forward Euler approximation \hat{X} of X with uniform time steps, $\hat{\Delta}t$, on a much finer grid than Δt , so that $\{i\hat{\Delta}t : i = 0, \dots, \hat{N}\}$ includes all time steps for \bar{X} . We can extend \hat{X} as in (4.1) to $t \in [0, T]$ by

$$(4.4) \quad \hat{X}(t) - \hat{X}(0) = \int_0^t \hat{a}(s; \hat{X}) ds + \sum_{\ell=1}^{\ell_0} \int_0^t \hat{b}^\ell(s; \hat{X}) dW^\ell(s)$$

where \hat{a} and \hat{b}^ℓ are the piecewise constant approximations, see Figure 1, $\hat{a}(s; \hat{X}) = a(i\hat{\Delta}t, \hat{X}(i\hat{\Delta}t))$ and $\hat{b}^\ell(s; \hat{X}) = b^\ell(i\hat{\Delta}t, \hat{X}(i\hat{\Delta}t))$ for $s \in [i\hat{\Delta}t, (i+1)\hat{\Delta}t)$. Then, the error splits into

$$(4.5) \quad \begin{aligned} & \Delta t_{sup}^{-\alpha}(\eta) \sup_t |X(t) - \bar{X}(t)| \\ & \leq \Delta t_{sup}^{-\alpha}(\eta) \sup_t |X(t) - \hat{X}(t)| + \Delta t_{sup}^{-\alpha}(\eta) \sup_t |\hat{X}(t) - \bar{X}(t)|. \end{aligned}$$

Let us first study the *a.s.* convergence of the second term in (4.5). The convergence of the standard Euler method with uniform time steps in the first term follows then by a similar derivation or by [32]. To show *a.s.* convergence of the second term we observe that

$$(4.6) \quad \hat{X}(t) - \bar{X}(t) = \int_0^t \Delta a(s) ds + \int_0^t \Delta b^\ell(s) dW^\ell(s),$$

where for $t \in [t_n, t_{n+1})$,

$$(4.7) \quad \begin{aligned} \Delta a(t) &= \hat{a}(t; \hat{X}) - \bar{a}(t; \bar{X}), \\ \Delta b^\ell(t) &= \hat{b}^\ell(t; \hat{X}) - \bar{b}^\ell(t; \bar{X}). \end{aligned}$$

Now let us define the second term of the right hand side in (4.6) by

$$(4.8) \quad Y(t) \equiv \int_0^t \Delta b^\ell(s) dW^\ell(s).$$

We have $\sup_t \sqrt{E[|Y(t)|^2]} = \mathcal{O}(\Delta t_{sup}^{\frac{1}{2}})$ by using the mean square strong convergence of the Euler approximation in Lemma 2.1. Lemma 4.2 below verifies that Y is a continuous martingale with respect to a filtration generated by $\{W(s), \Delta t(s) :$

$s \leq t$ }, so that $|Y|$ is a submartingale. Therefore Doob's inequality and Jensen's inequality give, for any $\alpha' \in \mathbb{R}$,

$$\begin{aligned} P\left(\sup_{0 \leq t \leq T} |Y(t)| \geq \Delta t_{sup}^{\alpha'}(\eta)\right) &\leq \frac{1}{\Delta t_{sup}^{\alpha'}(\eta)} E[|Y(t)|] \\ &\leq \frac{1}{\Delta t_{sup}^{\alpha'}(\eta)} \sqrt{E[|Y(t)|^2]} \\ &= \mathcal{O}(\Delta t_{sup}^{\frac{1}{2}-\alpha'}). \end{aligned}$$

The definition (4.2) then implies, for a positive constant C ,

$$\begin{aligned} \sum_{\eta=1}^{\infty} P\left(\sup_{0 \leq t \leq T} |Y(t)| \geq \Delta t_{sup}^{\alpha'}(\eta)\right) &\leq \sum_{\eta=1}^{\infty} \mathcal{O}(\Delta t_{sup}^{\frac{1}{2}-\alpha'}) \\ (4.9) \qquad \qquad \qquad &\leq C \sum_{\eta=1}^{\infty} 2^{-\eta(\frac{1}{2}-\alpha')} < \infty, \end{aligned}$$

provided $\frac{1}{2} - \alpha' > 0$. Therefore the Borel-Cantelli lemma implies

$$P\left(\sup_{0 \leq t \leq T} |Y(t)| \geq \Delta t_{sup}^{\alpha'}(\eta) \text{ infinitely often}\right) = 0,$$

i.e., with $\alpha < \frac{1}{2}$

$$(4.10) \qquad \Delta t_{sup}^{-\alpha}(\eta) \sup_{0 \leq t \leq T} |Y(t)| \rightarrow 0 \text{ a.s.}, \text{ as } \eta \rightarrow \infty.$$

The first term in the right hand side of (4.6) satisfies

$$\sup_t \left| \int_0^t \Delta a(s) ds \right| \leq \int_0^T |\Delta a(s)| ds.$$

Therefore Chebyshev's inequality yields

$$\begin{aligned} P\left(\sup_t \left| \int_0^t \Delta a(s) ds \right| \geq \Delta t_{sup}^{\alpha'}(\eta)\right) &\leq P\left(\int_0^T |\Delta a(s)| ds \geq \Delta t_{sup}^{\alpha'}(\eta)\right) \\ &\leq \frac{1}{\Delta t_{sup}^{2\alpha'}(\eta)} E\left[\left(\int_0^T |\Delta a(s)| ds\right)^2\right] \\ &\leq \frac{T}{\Delta t_{sup}^{2\alpha'}(\eta)} E\left[\int_0^T |\Delta a(s)|^2 ds\right] \\ &= \mathcal{O}\left(\Delta t_{sup}^{1-2\alpha'}(\eta)\right), \end{aligned}$$

where the last equality follows from Lemma 2.1. The argument in (4.9) and the Borel-Cantelli lemma give similarly $\Delta t_{sup}^{-\alpha}(\eta) \sup_t \left| \int_0^t \Delta a(s) ds \right| \rightarrow 0$ a.s. for $\alpha < \frac{1}{2}$, as $\eta \rightarrow \infty$, and consequently for $\alpha < \frac{1}{2}$

$$(4.11) \qquad \Delta t_{sup}^{-\alpha}(\eta) \sup_{t \in [0, T]} |\hat{X}(t) - \bar{X}(t)| \rightarrow 0 \text{ a.s. as } \eta \rightarrow \infty.$$

The same arguments applied to the first term in the right hand side of (4.5) yield for $\alpha < \frac{1}{2}$,

$$(4.12) \quad \Delta t_{sup}^{-\alpha}(\eta) \sup_{t \in [0, T]} |X(t) - \hat{X}(t)| \rightarrow 0 \text{ a.s. as } \eta \rightarrow \infty.$$

Here Lemma 4.2 holds directly for \hat{X} , since \hat{X} is adapted to the standard σ -algebra generated by $W(s)$. The combination (4.12) and (4.11) proves (4.3). \square

The key to our proof of Theorem 4.1 is that $Y_t \equiv Y(t)$ in (4.8) is a martingale with respect to a filtration \mathcal{F}_t :

Lemma 4.2. *Suppose that a, b, g and X satisfy the assumptions in Lemma 2.1. Let the process Y_t be defined by (4.8) and the filtration \mathcal{F}_t be the σ -algebra generated by $\{W(s), \Delta t(s) : s \leq t\}$. Then Y_t satisfies*

- (i): $E[|Y_t|] < \infty$, for $0 \leq t \leq T$,
- (ii): Y_t is adapted to \mathcal{F}_t ,
- (iii): $E[Y_t | \mathcal{F}_s] = Y_s$ for $t \geq s$,

i.e., Y_t is a martingale with respect to \mathcal{F}_t .

Proof. To prove the lemma, we observe by Lemma 2.1 that $E[|Y_t|] \leq \sqrt{E[|Y_t|^2]} < \infty$ and the construction of \bar{X} implies $Y_t \in \mathcal{F}_t$ for all t , so that (i) and (ii) holds. Note that Y_t is not adapted to the standard filtration generated by W only. Using conditional expectation and the notations in Theorem 4.1

$$(4.13) \quad \begin{aligned} E[Y_t | \mathcal{F}_s] &= E[Y_s | \mathcal{F}_s] + E \left[\int_s^t \Delta b^\ell dW^\ell \mid \mathcal{F}_s \right] \\ &= Y_s + E \left[\int_s^t \Delta b^\ell dW^\ell \mid \mathcal{F}_s \right]. \end{aligned}$$

It remains to verify $E \left[\int_s^t \Delta b^\ell dW^\ell \mid \mathcal{F}_s \right] = 0$. Since Δb^ℓ is piecewise constant, rewrite $\int_s^t \Delta b^\ell dW^\ell = \sum_{i=n_1}^{n_2} \Delta b^\ell(i\hat{\Delta}t) \hat{\Delta}W_i^\ell$, where $\hat{\Delta}W_i^\ell = W^\ell((i+1)\hat{\Delta}t) - W^\ell(i\hat{\Delta}t)$ with the uniform time steps, $\hat{\Delta}t$, which is finer than Δt ; at the end point t of the interval $[s, t]$ we let $\hat{\Delta}W_{n_2}^\ell = W^\ell(t) - W^\ell(n_2\hat{\Delta}t)$ and similarly for the end point s .

Let us now study one term $E \left[\Delta b^\ell(\hat{n}\hat{\Delta}t) \hat{\Delta}W_{\hat{n}}^\ell \right]$. Divide the interval, $[\hat{n}\hat{\Delta}t, (\hat{n}+1)\hat{\Delta}t]$, into the union of disjoint intervals $[\hat{n}\hat{\Delta}t + \sum_{j=1}^{m-1} \hat{\Delta}t_j, \hat{n}\hat{\Delta}t + \sum_{j=1}^m \hat{\Delta}t_j]$, so that $\hat{\Delta}t = \sum_{m=1}^{\hat{N}} \hat{\Delta}t_m$ with the corresponding Wiener increments

$$\begin{aligned} \hat{\Delta}W_m^\ell &= W^\ell(\hat{n}\hat{\Delta}t + \sum_{j=1}^m \hat{\Delta}t_j) - W^\ell(\hat{n}\hat{\Delta}t + \sum_{j=1}^{m-1} \hat{\Delta}t_j), \\ \check{\Delta}W_m^\ell &= W^\ell((\hat{n}+1)\hat{\Delta}t) - W^\ell(\hat{n}\hat{\Delta}t + \sum_{j=1}^m \hat{\Delta}t_j). \end{aligned}$$

The end intervals $(n_2\hat{\Delta}t, t)$ and $(s, n_1\hat{\Delta}t)$ are treated similarly.

We claim that $\bar{X}(t_n)$ is essentially independent of the increments $W(\hat{t} + \hat{\Delta}t) - W(\hat{t})$, provided $\hat{\Delta}t(\hat{t})$ is sufficiently small and $t_n < \hat{t}$. The stopping criteria (3.18, 3.19), for each final acceptable time discretization imply that $\min_n (S_1 \frac{\text{TOL}_T}{N} - r_n)$

is positive for all realizations. The approximate solution $\overline{X}(\tau)$ depends on $dW(\hat{t})$, for $\tau < \hat{t} \equiv \hat{n}\hat{\Delta t}$, only through changes in the mesh. We shall show that, provided $\hat{\Delta t}$ is sufficiently small and conditioned on the σ -algebra $\mathcal{M}(\hat{t}, \hat{\Delta t})$ generated by $\{dW(\tau) : \tau < \hat{t}, \text{ or } \tau > \hat{t} + \hat{\Delta t}\}$, the probability to change the mesh by a change in only $\hat{\Delta W}(\hat{t})$ is arbitrary small, thus $\overline{X}(\tau)$ will be essentially independent of $\hat{\Delta W}(\hat{t})$ for $\tau < \hat{t}$. Conditioned on $\mathcal{M}(\hat{t}, \hat{\Delta t})$, let $r_n(\hat{\Delta W}) \equiv E[r_n | \mathcal{M}(\hat{t}, \hat{\Delta t})]$ denote the dependence of the error indicator r_n on the noise $\hat{\Delta W}$. The Malliavin derivative, $\partial_{W^\ell(t)} r_n$, and Taylor's formula imply

$$r_n(\hat{\Delta W}) = r_n(0) + \hat{\Delta W}^\ell \int_0^1 \partial_{W^\ell(t)} r_n(\tau \hat{\Delta W}) d\tau.$$

The mesh generated by $r_n(\hat{\Delta W})$ and $r_n(0)$ is the same provided

$$\begin{aligned} (4.14) \quad 0 &< S_1 \frac{\text{TOL}_T}{N} - r_n(\hat{\Delta W}) \\ &= S_1 \frac{\text{TOL}_T}{N} - \left(r_n(0) + \hat{\Delta W}^\ell \int_0^1 \partial_{W^\ell(t)} r_n(\tau \hat{\Delta W}) d\tau \right) \\ &\equiv S_1 \frac{\text{TOL}_T}{N} - (r_n(0) + \hat{\Delta W} \cdot r'_n), \text{ for all } n, \end{aligned}$$

and

$$(4.15) \quad \min_n \left(S_1 \frac{\text{TOL}_T}{N} - r_n(0) \right) \equiv \epsilon(\omega) > 0.$$

Therefore, (4.14, 4.15) hold if $\hat{\Delta W} \cdot r'_n < \epsilon(\omega)$ for all n . Let

$$\check{\epsilon} \equiv \epsilon / ((\sup_n \Delta t_n)^2 C(\text{TOL})),$$

then $|\hat{\Delta W}| < \check{\epsilon}$ implies $\hat{\Delta W} \cdot r'_n < \epsilon$, since

$$\hat{\Delta W} \cdot r'_n \leq |\hat{\Delta W}| |r'_n| < \check{\epsilon} |r'_n| = \epsilon (\Delta t_n)^2 \frac{|\int_0^1 \partial_{W(t)} \rho(t_n, \tau \hat{\Delta W}) d\tau|}{(\sup \Delta t_n)^2 C(\text{TOL})} \leq \epsilon,$$

provided $\rho(t_n, \hat{\Delta W}) \equiv E[\rho(t_n) | \mathcal{M}(\hat{t}, \hat{\Delta t})]$ and ρ is an approximate error density function satisfying (2.3).

Consequently, the following independence claim holds:

$$(4.16) \quad \begin{aligned} &\overline{X}(t_n) \text{ is independent of } \hat{\Delta W}(\hat{t}) \text{ conditioned on} \\ &|\hat{\Delta W}(\hat{t})| < \check{\epsilon} \text{ and } \mathcal{M}(\hat{t}, \hat{\Delta t}), \text{ for } t_n < \hat{t}, \end{aligned}$$

and the conditional probability to have different meshes with $r_n(0)$ or $r_n(\hat{\Delta W}(\hat{t}))$ is, for sufficiently small $\hat{\Delta t}$, bounded by

$$(4.17) \quad P \left(|\hat{\Delta W}| \geq \check{\epsilon} \mid \mathcal{M}(\hat{t}, \hat{\Delta t}), \mathcal{F}_s \right) = 1 - P \left(|\hat{\Delta W}| < \check{\epsilon} \mid \mathcal{M}(\hat{t}, \hat{\Delta t}), \mathcal{F}_s \right) \leq (\hat{\Delta t})^3$$

since

$$\begin{aligned} P\left(|\hat{\Delta}W| < \check{\epsilon} \mid \mathcal{M}(\hat{t}, \hat{\Delta}t), \mathcal{F}_s\right) &= P\left(|\hat{\Delta}W| < \check{\epsilon} \mid \mathcal{M}(\hat{t}, \hat{\Delta}t)\right) \\ &\geq 1 - C_{\ell_0} \exp\left(-\frac{(\check{\epsilon})^2}{4\hat{\Delta}t}\right) > 1 - (\hat{\Delta}t)^3. \end{aligned}$$

Let us define the σ -algebra \mathcal{G} generated by $\{dW(\tau) : \tau < \hat{n}\hat{\Delta}t \text{ or } \tau > \hat{n}\hat{\Delta}t + \hat{\Delta}t\}$ and $\{\Delta t(\tau) : \tau < \hat{n}\hat{\Delta}t\}$. Then using the results (4.17) and (4.16), we get

$$\begin{aligned} &E\left[\Delta b^\ell \hat{\Delta}W_{\hat{n}}^\ell \mid \mathcal{F}_s\right] = E\left[E\left[\Delta b^\ell \hat{\Delta}W_{\hat{n}}^\ell \mid \mathcal{G}\right] \mid \mathcal{F}_s\right] \\ &= E\left[E\left[\Delta b^\ell \left(\check{\Delta}W_1^\ell + \mathbf{1}_{\{|\hat{\Delta}W_1| < \check{\epsilon}\}} + \mathbf{1}_{\{|\hat{\Delta}W_1| \geq \check{\epsilon}\}}\right) \hat{\Delta}W_1^\ell \mid \mathcal{G}\right] \mid \mathcal{F}_s\right] \\ (4.18) \quad &= E\left[E\left[\Delta b^\ell \check{\Delta}W_1^\ell \mid \mathcal{G}\right] \mid \mathcal{F}_s\right] + E\left[E\left[\Delta b^\ell \mathbf{1}_{\{|\hat{\Delta}W_1| < \check{\epsilon}\}} \hat{\Delta}W_1^\ell \mid \mathcal{G}\right] \mid \mathcal{F}_s\right] \\ &\quad + E\left[E\left[\Delta b^\ell \mathbf{1}_{\{|\hat{\Delta}W_1| \geq \check{\epsilon}\}} \hat{\Delta}W_1^\ell \mid \mathcal{G}\right] \mid \mathcal{F}_s\right]. \end{aligned}$$

By (4.17), the last term of (4.18) becomes

$$(4.19) \quad E\left[E\left[\Delta b^\ell \mathbf{1}_{\{|\hat{\Delta}W_1| \geq \check{\epsilon}\}} \hat{\Delta}W_1^\ell \mid \mathcal{G}\right] \mid \mathcal{F}_s\right] = E\left[\mathcal{O}((\hat{\Delta}t_1)^3)\right]$$

and from (4.16) the second term of (4.18) becomes

$$\begin{aligned} (4.20) \quad &E\left[E\left[\Delta b^\ell \mathbf{1}_{\{|\hat{\Delta}W_1| < \check{\epsilon}\}} \hat{\Delta}W_1^\ell \mid \mathcal{G}\right] \mid \mathcal{F}_s\right] \\ &= E\left[E\left[\Delta b^\ell \mid \mathcal{G}, \{|\hat{\Delta}W_1| < \check{\epsilon}\}, \mathcal{F}_s\right]\right. \\ &\quad \left.\times \underbrace{E\left[\mathbf{1}_{\{|\hat{\Delta}W_1| < \check{\epsilon}\}} \hat{\Delta}W_1^\ell \mid \mathcal{G}, \{|\hat{\Delta}W_1| < \check{\epsilon}\}, \mathcal{F}_s\right] P(|\hat{\Delta}W_1| < \check{\epsilon} \mid \mathcal{G}, \mathcal{F}_s)}_{=0}\right] = 0. \end{aligned}$$

Therefore we obtain from (4.19) and (4.20)

$$(4.21) \quad E\left[\Delta b^\ell \hat{\Delta}W_{\hat{n}}^\ell \mid \mathcal{F}_s\right] = E\left[E\left[\Delta b^\ell \check{\Delta}W_1^\ell \mid \mathcal{G}\right] \mid \mathcal{F}_s\right] + E\left[\mathcal{O}((\hat{\Delta}t_1)^3)\right].$$

Apply this $\hat{\Delta}W, \check{\Delta}W$ argument recursively to $E[\Delta b^\ell \check{\Delta}W_m^\ell]$ conditioned on $\mathcal{M}(i\Delta t + \sum_{j=1}^m \hat{\Delta}t_j, \hat{\Delta}t_{m+1})$ to get

$$(4.22) \quad E\left[\Delta b^\ell \hat{\Delta}W_{\hat{n}}^\ell \mid \mathcal{F}_s\right] = E\left[\sum_m \mathcal{O}((\hat{\Delta}t_m)^3)\right] = E\left[\int_{(\hat{n}-1)\hat{\Delta}t}^{\hat{n}\hat{\Delta}t} \mathcal{O}((\hat{\Delta}t)^2) dt\right].$$

Letting $\hat{\Delta}t \rightarrow 0+$ in (4.22) proves $E[\Delta b^\ell \hat{\Delta}W_{\hat{n}}^\ell \mid \mathcal{F}_s] = 0$ and apply the same arguments to all intervals $[\hat{n}\hat{\Delta}t, (\hat{n}+1)\hat{\Delta}t]$, so that by (4.13) $E[Y_t \mid \mathcal{F}_s] = Y_s$ for $t \geq s$. \square

To verify *a.s.* convergence of the error density, let us recall the definition of the variation of a process Y : the first variation of a function $F(Y(T))$ with respect to

a perturbation in the initial location of the path Y , at time s , is denoted by

$$(4.23) \quad \begin{aligned} F'(T; s) &= \partial_{y(s)} F(Y(T)) \\ &\equiv \left(\frac{\partial}{\partial y_1} F(Y(T); Y(s) = y), \dots, \frac{\partial}{\partial y_d} F(Y(T); Y(s) = y) \right). \end{aligned}$$

The definition (4.23) implies that the first variation, X' , of the solution X with respect to a perturbation in the initial location at t satisfies

$$(4.24) \quad \begin{aligned} dX'_{ik}(s) &= \partial_j a_i(s, X(s)) X'_{jk}(s) + \partial_j b_i^\ell X'_{jk}(s) dW^\ell(s), \quad s > t, \\ X'_{ik}(t) &= \delta_{ik} = \begin{cases} 0, & i \neq k, \\ 1, & i = k, \end{cases} \end{aligned}$$

and similarly one can derive the equations for the second and the third variation of X , cf. [31].

The definition of c_i in (2.8) shows that the forward Euler approximation, \bar{X}' , in (4.24) can be written

$$(4.25) \quad \begin{aligned} \bar{X}'_{ik}(t_{n+1}; t_m) &= \partial_j c_i(t_n, \bar{X}(t_n)) \bar{X}'_{jk}(t_n; t_m) \\ \bar{X}'_{ik}(t_m) &= \delta_{ik}. \end{aligned}$$

Then the equations for φ in (2.7) and for \bar{X}' in (4.25) imply

$$\begin{aligned} 0 &= \sum_{n=m}^{N-1} (\varphi_i(t_n) - \partial_i c_j(t_n, \bar{X}(t_n)) \varphi_j(t_{n+1})) \bar{X}'_{ik}(t_n; t_m) \\ &= \sum_{n=m}^{N-1} \varphi_i(t_{n+1}) \left(\bar{X}'_{ik}(t_{n+1}; t_m) - \partial_j c_i(t_n, \bar{X}(t_n)) \bar{X}'_{jk}(t_n; t_m) \right) \\ &\quad + \varphi_i(t_m) \bar{X}'_{ik}(t_m; t_m) - \varphi_i(T) \bar{X}'_{ik}(T; t_m) \\ &= \varphi_i(t_m) \bar{X}'_{ik}(t_m; t_m) - \varphi_i(T) \bar{X}'_{ik}(T; t_m), \end{aligned}$$

i.e.,

$$(4.26) \quad \varphi_k(t_m) = \partial_i g(\bar{X}(T)) \bar{X}'_{ik}(T; t_m),$$

by using the initial conditions of φ and \bar{X}' . The definitions of the first and second variations of φ in (2.9) and (2.10) also yield

$$(4.27) \quad \varphi'_{kn}(t_m) = \partial_i g(\bar{X}(T)) \bar{X}''_{ikn}(T; t_m) + \partial_{ir} g(\bar{X}(T)) \bar{X}'_{rn}(T; t_m) \bar{X}'_{ik}(T; t_m)$$

and

$$(4.28) \quad \begin{aligned} \varphi''_{knm}(t_m) &= \partial_i g(\bar{X}(T)) \bar{X}'''_{iknm}(T; t_m) + \partial_{ir} g(\bar{X}(T)) \bar{X}'_{rm}(T; t_m) \bar{X}''_{ikn}(T; t_m) \\ &\quad + \partial_{ir} g(\bar{X}(T)) \bar{X}'_{ik}(T; t_m) \bar{X}''_{rnm}(T; t_m) + \partial_{ir} g(\bar{X}(T)) \bar{X}''_{ikm}(T; t_m) \bar{X}'_{rn}(T; t_m) \\ &\quad + \partial_{irv} g(\bar{X}(T)) \bar{X}'_{vm}(T; t_m) \bar{X}'_{ik}(T; t_m) \bar{X}'_{rn}(T; t_m). \end{aligned}$$

Define the function $\hat{\rho}$ following the error density ρ in (2.6) by substituting \bar{X} by the limit X and replacing $\varphi, \varphi', \varphi''$ by the corresponding limits in (4.26-4.28) where \bar{X} is substituted by X .

Now we are ready to prove the *a.s.* convergence of the error density :

Corollary 4.3. *Suppose the assumptions of Theorem 4.1. Then the error density ρ in (2.6) converges a.s. to the limit $\hat{\rho}$, defined above, as the specified error tolerance tends to zero.*

Proof. To prove the corollary, it is necessary to understand the convergence of the discrete dual solutions, φ , in (2.7) and its first and second variation, φ', φ'' . Using the definitions of the variations of X , let us consider an augmented system for $Z = (X, X', X'', X''')^T$ and let I denote the $d \times d$ identity matrix. Then (4.24) and the equations for X'' and X''' can be written

$$(4.29) \quad \begin{aligned} dZ &= A(t, Z)dt + B^\ell(t, Z)dW^\ell(t), \quad t > t_0 \\ Z(t_0) &= (x, I, 0, 0, 0)^T. \end{aligned}$$

The Euler approximation $\bar{Z} = (\bar{X}, \bar{X}', \bar{X}'', \bar{X}''')^T$ of Z in (4.29) with piecewise constant drift and diffusion fluxes satisfies

$$d\bar{Z} = \bar{A}(t; \bar{Z})dt + \bar{B}^\ell(t; \bar{Z})dW^\ell$$

defined as in (4.1).

Applying Theorem 4.1 to \bar{Z} and the augmented system (4.29) shows that the approximation \bar{Z} converges a.s. to the process Z . Therefore, the representations (4.26- 4.28) and the regularity assumptions on the functions g, a and b imply that φ, φ' and φ'' converge a.s.. Consequently, the error density ρ converges a.s. to the true error density $\hat{\rho}$ as the specified error tolerance, TOL, tends to zero, i.e., as the maximum step size tends to zero. \square

4.2. Deterministic time steps.

Lemma 4.4. *Suppose that a, b, g and X satisfy the assumptions in Lemma 2.1. Then for any $\alpha \in (0, 1/2)$*

$$(4.30) \quad \lim_{M \rightarrow \infty} M^\alpha \sup_{t \in [0, T]} |E[\bar{X}(t)] - \mathcal{A}(\bar{X}(t); M)| = 0 \text{ a.s.}$$

Proof. The idea is to split the difference in (4.30) into a sum of a martingale and a bounded integral and use Doob's inequality for the martingale. Thus, representation (4.1) shows

$$(4.31) \quad \begin{aligned} E[\bar{X}(t)] - \mathcal{A}(\bar{X}(t); M) &= \underbrace{E\left[\int_0^t \bar{b}^\ell(s; \bar{X}(s))dW\right]}_{=0} - \mathcal{A}\left(\int_0^t \bar{b}^\ell(s; \bar{X}(s))dW^\ell; M\right) \\ &+ \int_0^t (E[\bar{a}(s; \bar{X}(s))] - \mathcal{A}(\bar{a}(s; \bar{X}(s)); M)) ds. \end{aligned}$$

Denote by $\{\mathcal{F}_{M,t}\}_{t \geq 0}$ the filtration generated by the M independent Wiener processes that define the sample average $\mathcal{A}(\bar{X}(t); M)$ and let the σ -algebra generated by the step sizes be $\sigma(\Delta t)$. The construction of Algorithm D implies that $\sigma(\Delta t)$ is independent of $\mathcal{F}_{M,t}$. The conditional expectation process

$$Y_M(t) \equiv E\left[\mathcal{A}\left(\int_0^t \bar{b}^\ell(s; \bar{X}(s))dW^\ell; M\right) \mid \sigma(\Delta t)\right]$$

is then a $\mathcal{F}_{M,t}$ martingale and Doob's inequality yields

$$\begin{aligned}
P(M^\alpha \sup_{t \in [0, T]} |Y_M(t)| \geq \epsilon) &\leq M^{2\alpha} \frac{\text{Var}[Y_M(T)]}{\epsilon^2} \\
&= M^{2\alpha} \frac{\text{Var}[E[\mathcal{A}(\int_0^T \bar{b}^\ell(s, \bar{X}(s)) dW^\ell(s); M) | \sigma(\Delta t)]]}{\epsilon^2} \\
&= \frac{\text{Var}[E[\int_0^T \bar{b}^\ell(s, \bar{X}(s)) dW^\ell(s) | \sigma(\Delta t)]]}{M^{1-2\alpha} \epsilon^2} \\
&\leq \frac{C_T}{M^{1-2\alpha} \epsilon^2}.
\end{aligned}$$

Here C_T is a constant that, due to the smoothness assumptions in the drift and diffusion coefficients, bounds uniformly the variance of $\int_0^T \bar{b}^\ell(s, \bar{X}(s)) dW^\ell(s)$. The construction (3.6) shows that all M belong to the set $\{2^k | k \in \mathbb{N}\}$. Let $M_k \equiv 2^k$, then

(4.32)

$$\sum_{k=1}^{\infty} P\left(M_k^\alpha \sup_{t \in [0, T]} |Y_{M_k}(t)| \geq \epsilon\right) \leq \frac{C_T}{\epsilon^2} \sum_{k=1}^{\infty} \frac{1}{(M_k)^{1-2\alpha}} < \infty$$

and Borel-Cantelli's lemma implies

$$(4.33) \quad P\left((M_k)^\alpha \sup_{t \in [0, T]} |Y_{M_k}(t)| \geq \epsilon \text{ infinitely often}\right) = 0,$$

i.e. $(M_k)^\alpha \sup_{t \in [0, T]} |Y_{M_k}(t)| \rightarrow 0$ a.s., as $k \rightarrow \infty$ for $0 < \alpha < 1/2$. Now consider the second term in (4.31), namely $\int_0^t (E[\bar{a}(s; \bar{X}(s))] - \mathcal{A}(\bar{a}(s; \bar{X}(s)); M_k)) ds$. First note that we have

$$(4.34) \quad \sup_{t \in [0, T]} \left| \int_0^t E[\bar{a}(s; \bar{X}(s))] - \mathcal{A}(\bar{a}(s; \bar{X}(s)); M_k) ds \right| \leq \int_0^T |E[\bar{a}(s; \bar{X}(s))] - \mathcal{A}(\bar{a}(s; \bar{X}(s)); M_k)| ds$$

so that a combination of Chebychev's inequality and Cauchy-Schwartz inequality yield

$$\begin{aligned}
&P\left((M_k)^\alpha \sup_{t \in [0, T]} \left| \int_0^t (E[\bar{a}(s; \bar{X}(s))] - \mathcal{A}(\bar{a}(s; \bar{X}(s)); M_k)) ds \right| \geq \epsilon\right) \\
&\leq \frac{M_k^{2\alpha}}{\epsilon^2} E\left[\left(\int_0^T |E[\bar{a}(s; \bar{X}(s))] - \mathcal{A}(\bar{a}(s; \bar{X}(s)); M_k)| ds\right)^2\right] \\
&\leq \frac{M_k^{2\alpha}}{\epsilon^2} T \int_0^T E\left[(E[\bar{a}(s; \bar{X}(s))] - \mathcal{A}(\bar{a}(s; \bar{X}(s)); M_k))^2\right] ds \\
&= \frac{T}{\epsilon^2 M_k^{1-2\alpha}} \int_0^T \text{Var}[\bar{a}(s; \bar{X}(s))] ds \leq \frac{C_T}{\epsilon^2 M_k^{1-2\alpha}}.
\end{aligned}$$

The analogous Borel-Cantelli arguments (4.32-4.33) therefore imply for $\alpha < 1/2$

$$(4.35) \quad (M_k)^\alpha \sup_{t \in [0, T]} \left| \int_0^t E[\bar{a}(s; \bar{X}(s))] - \mathcal{A}(\bar{a}(s; \bar{X}(s)); M_k) ds \right| \rightarrow 0 \text{ a.s.}$$

which combined with (4.33) proves the lemma. \square

Theorem 4.5 (Convergence). *Suppose the assumptions of Lemma 4.4 hold. Then the error density $\bar{\rho}$ in (2.13) converges a.s. to $E[\hat{\rho}]$, defined in Corollary 4.3, as the specified error tolerance tends to zero.*

Proof. First split $E[\hat{\rho}] - \mathcal{A}(\bar{\rho}; M) = E[\hat{\rho} - \bar{\rho}] + (E[\bar{\rho}] - \mathcal{A}(\bar{\rho}; M))$. The first term is bounded by $\mathcal{O}(\Delta t_{sup})$ using, for the augmented system in (4.29), the weak convergence of order one, proved for both stochastic and deterministic time steps in Theorem 2.2. The a.s. convergence of the second term follows from applying Lemma 4.4 to the augmented system in (4.29). \square

5. NUMERICAL EXPERIMENTS

This section presents numerical results from the implementation of the adaptive algorithms described in Section 3, namely the adaptive algorithms with deterministic time steps (Algorithm D) and with stochastic time steps (Algorithm S).

Algorithm D uses a Matlab version 6 implementation and simulates the ℓ_0 independent Wiener processes with a pseudo-random number generator, based on either a linear congruential recursion,

$$(5.1) \quad u_{k+1} = (au_k + c) \pmod{m}$$

with $a = 16807$, $c = 0$ and $m = 2^{31} - 1$, or on a more advanced random number generator that uses several seeds proposed by Marsaglia, see [23].

Algorithm S uses a double precision FORTRAN 77 implementation, with the same linear congruential pseudo-random number generator (5.1). In particular, the program applies a double precision modification of the functions `ran1` and `gasdev` proposed in [30], provided with an initial seed.

In all computations the number of subdivisions of a refined step is $H = 2$, and the constants to determine the number of realizations in (3.6) are $c_0 = 1.65$ and $MCH = 16$.

We prescribe the error tolerance $TOL = 0.025$ with the splitting $TOL_S = \frac{2}{3}TOL$ and $TOL_T = \frac{1}{3}TOL$ and Algorithm D uses $TOL_T = TOL_{TS} + TOL_{TT}$, where $TOL_{TS} = \frac{1}{9}TOL$ and $TOL_{TT} = \frac{2}{9}TOL$. Observe that the choice $TOL_{TT} = \frac{2}{9}TOL$ instead of $TOL_{TT} = \frac{1}{3}TOL$ will in general give a final number of time steps in Algorithm D that is around 50% larger than the expected value of the number of time steps in Algorithm S. However, this apparent disadvantage is largely overcome by the fact that in Algorithm S we have to compute the time error approximations for all realizations, while in Algorithm D this is done just in the first loop, which usually has a smaller number of realizations.

Unless otherwise stated, the time adaptivity procedures in Algorithm D and Algorithm S are applied with initial uniform partitions of $[0, T]$ having $N[1] = 20$ subintervals and the initial number of realizations is $M[1] = 64$.

The constants for the merging and dividing procedure are taken in accordance with Remarks 3.1 and 3.7 and the corresponding stopping results in Theorems 3.2 and 3.8. Indeed, we use $d_1 = s_1 = 2$ and $d_2 = s_2 = 1/40$, while the corresponding constants for the stopping criteria are $D_1 = S_1 = 8$ and $D_2 = S_2 = 1/160$, respectively.

In order to reduce the computational effort, we use antithetic variates, see [13], to reduce the variance in some computations.

To see the practical effect of the merging procedure, both algorithms are tested with and without it. The results with merging correspond to the table values Merge = 1 and similarly, the results with the antithetic variance reduction correspond to Anti = 1. Due to the construction of our algorithms, the number of realizations, M , is a power of 2, see (3.6). In fact, the final number of realizations is either $M = 2^{15} = 32768$ or $M = 2^{17} = 131072$, depending on the use of variance reduction and the accuracy level, $\text{TOL} = 0.025$.

In **Algorithm S** the total amount of computational work is proportional to the total number of time steps performed in all realizations. The latter is shown for all the computations under the table entry $\sum MN$. Similarly, in **Algorithm D** we show for each computation the table entry $1^{st} \sum MN$, which corresponds to the total number of time steps in all realizations of the first loop that determines the mesh, and $2^{nd} \sum MN$, which corresponds to the total number of time steps in all realizations of the second loop with fixed mesh. We make a distinction between the amount of work in the first and the second loop since only the first entails the computation of time discretization error approximations, and therefore the first loop uses much more computational effort per time step.

For each computation with **Algorithm D** we show the final number of time steps, N , as well as an estimate of the optimal number of time steps, N^* , that would be required to achieve the desired accuracy. The optimal number of time steps is based on the approximate error density defined in (2.13) and (3.8), and it is given by $N^* = (\int_0^T \sqrt{|\bar{\rho}(\tau)|} d\tau)^2 / \text{TOL}_{TT}$.

Similarly, for each of the computations with **Algorithm S** we show the sample average of the final number of time steps, $\mathcal{A}(N; M)$ and its sample standard deviation, $\mathcal{S}(N; M)$. We use the estimate $(\mathcal{A}(\int_0^T \sqrt{|\bar{\rho}(\tau)|} d\tau; M))^2 / \text{TOL}_{TT}$ for the optimal expected number of time steps that would be required to achieve the desired accuracy, based on the approximate error density defined in (2.6) and (3.8).

Example 5.1.

Let us consider a real constant $\alpha \in (0, T)$ and the linear stochastic differential equation

$$(5.2) \quad \begin{aligned} dX(t) &= \begin{cases} X(t)dW(t), & t \in [0, \alpha] \\ \frac{X(t)dt}{2\sqrt{t-\alpha}} + X(t)dW(t), & t \in (\alpha, T] \end{cases} \\ X(0) &= 1 \end{aligned}$$

with the unique solution

$$X(t) = \begin{cases} \exp(W(t) - \frac{1}{2}t), & t \in [0, \alpha] \\ \exp(W(t) - \frac{1}{2}t) \exp(\sqrt{t-\alpha}), & t \in [\alpha, T]. \end{cases}$$

Here we choose $T = 1$ and $\alpha = T/3$. Our goal is to approximate $E[X(T)] = \exp(\sqrt{T-\alpha})$. To avoid evaluating arbitrarily large values of the drift in (5.2) we modify the drift to be

$$a(t, x) = \begin{cases} 0, & t \in [0, \alpha] \\ \frac{x}{2\sqrt{t-\alpha} + \text{TOL}^4}, & t \in (\alpha, T] \end{cases}$$

yielding a higher order perturbation $\mathcal{O}(\text{TOL}^2)$ in the computed result and in the size of the optimal time steps. Due to the time discontinuity of the drift function and to ensure optimal convergence of the adaptive algorithms, we modify the Euler method by

$$\bar{X}_{n+1} - \bar{X}_n = a(\hat{t}, \bar{X}_n)\Delta t_n + \bar{X}_n\Delta W_n, \quad n = 0, \dots$$

where we choose the stochastic evaluation time $\hat{t} \in \{t_n, t_{n+1}\}$ by

$$|a(\hat{t}, \bar{X}_n)| = \max(|a(t_n, \bar{X}_n)|, |a(t_{n+1}, \bar{X}_n)|).$$

Observe that the use of \hat{t} does not change the adapted nature of the Euler method.

We compare the results of **Algorithm D** and **Algorithm S** in Tables 1, 2, 3 and 4. First, Tables 1 and 2 show results from **Algorithm D**, the only difference being the choice of the pseudo random number generator. Results in Table 1 use [23] and those in Table 2 are based on the linear congruential random number generator (5.1). As expected, the level of difference between these results is comparable to the effect of taking different initial seeds in the computations.

Next, Tables 3 and 4 show results from **Algorithm S**. Table 3 shows results when the time adaptivity procedure is started with the same mesh for each realization, namely a uniform mesh with 20 time steps, whereas the results of Table 4 use that the time adaptivity procedure is started with the optimal mesh from the previous realization, yielding less computational work since the drift singularity occurs at a deterministic time. Table 4 shows a situation where merging is needed, namely when using the stochastic time stepping algorithm and the optimal mesh from the previous realization as a starting guess for the next realization. In this case the merging procedure avoids an excessive increment of the average number of time steps.

In this example **Algorithm D** and **Algorithm S** select the final number of time steps in a similar way, since there is no remarkable influence from the stochastic term $X(t)dW(t)$ in the dynamics of (5.2). Besides this, Figure 3 shows that both **Algorithm D** and **Algorithm S** have a remarkable advantage over a computation with uniform time steps which needs 1.7×10^5 time steps to achieve the given level of accuracy $\text{TOL}_{TT} \approx 5.6 \times 10^{-3}$, compared to on average less than 600 final adaptive steps. However, there is a noticeable difference in the total number of operations in all levels, and the deterministic time stepping algorithm has a clear advantage. Table 1 shows that variance reduction is useful in the second loop of the deterministic time stepping algorithm where the statistical error to control has an associated large variance. Finally, there is no significant difference between the results with the merging procedure or without merging and in addition, the final number of time steps is close to the estimated optimal number of time steps.

Algorithm D

Merge	Anti	M	N	$1^{st} \sum MN$	$2^{nd} \sum MN$	N^*
1	1	2^{15}	4.8×10^2	0.6×10^6	48×10^6	4.2×10^2
1	0	2^{17}	5.2×10^2	0.3×10^6	77×10^6	4.7×10^2
0	1	2^{15}	4.6×10^2	0.6×10^6	46×10^6	4.4×10^2
0	0	2^{17}	5.0×10^2	0.3×10^6	74×10^6	4.4×10^2

TABLE 1. (Example 5.1) Numerical results using the Matlab 6 pseudo random number generator, see [23].

Algorithm D

Merge	Anti	M	N	$1^{st} \sum MN$	$2^{nd} \sum MN$	N^*
1	1	2^{15}	5.0×10^2	0.6×10^6	50×10^6	4.1×10^2
1	0	2^{17}	6.4×10^2	0.35×10^6	95×10^6	5.8×10^2
0	1	2^{15}	5.0×10^2	0.6×10^6	50×10^6	4.9×10^2
0	0	2^{17}	5.7×10^2	0.35×10^6	85×10^6	4.5×10^2

TABLE 2. (Example 5.1) Similar to Table 1, but using the linear congruential pseudo random number generator from (5.1) to simulate the Wiener increments.

Algorithm S

Merge	M	$\mathcal{A}(N; M)$	$\mathcal{S}(N; M)$	$\sum MN$
1	2^{15}	2.2×10^2	4×10	170×10^6
0	2^{15}	2.3×10^2	4×10	175×10^6

TABLE 3. (Example 5.1) For each realization the time adaptivity procedure is started with the same uniform mesh that has 20 time steps. This computation uses antithetic variates and the estimate for the expected value of the optimal number of time steps is in this case 2.1×10^2 .

Algorithm S

M	$\mathcal{A}(N; M)$	$\mathcal{S}(N; M)$	$\sum MN$
2^{15}	5.8×10^2	1.5×10^2	56×10^6

TABLE 4. (Example 5.1) For each realization the time adaptivity procedure is started with the optimal mesh from the previous realization. This computation uses both antithetic variates and the merging procedure. The estimate for the expected value of the optimal number of time steps is in this case 2.1×10^2 .

Example 5.2.

Now we change (5.2) taking α no longer a constant but a uniformly distributed random variable independent of the Wiener process, i.e. $\alpha \sim U(\underline{\alpha}, \bar{\alpha})$ with $\underline{\alpha} = 1/22$ and $\bar{\alpha} = 1 - 1/22$, for $T = 1$. The conditional expectation $E[X(T)|\alpha] = \exp(\sqrt{T - \alpha})$ can be used to compute

$$E[X(T)] = \frac{1}{\bar{\alpha} - \underline{\alpha}} \int_{\underline{\alpha}}^{\bar{\alpha}} \exp(\sqrt{T - \alpha}) d\alpha,$$

which is now the functional to approximate. Since the position of the drift singularity is stochastic, the deterministic time stepping **Algorithm D** gives the approximation error $\mathcal{O}(\sqrt{\Delta t_{\text{sup}}})$, so that $N_D \sim \text{TOL}_T^{-2}$; and **Algorithm D** behaves like a uniform time stepping algorithm which needs 1.3×10^5 time steps to achieve the given level of accuracy $\text{TOL}_{TT} \approx 5.6 \times 10^{-3}$. Thus, **Algorithm S** has a clear advantage in this example, with asymptotic accuracy $\mathcal{O}(\Delta t_{\text{sup}})$, so that $E[N_S] \sim \text{TOL}_T^{-1}$ using on average less than 200 final time steps, as Table 5 and Figure 4 show.

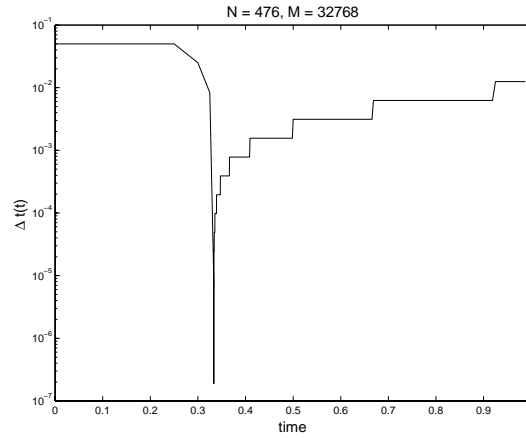


FIGURE 2. (Example 5.1). Computation with variance reduction and the merging procedure. Final mesh function for Algorithm D and $TOL = 0.025$.

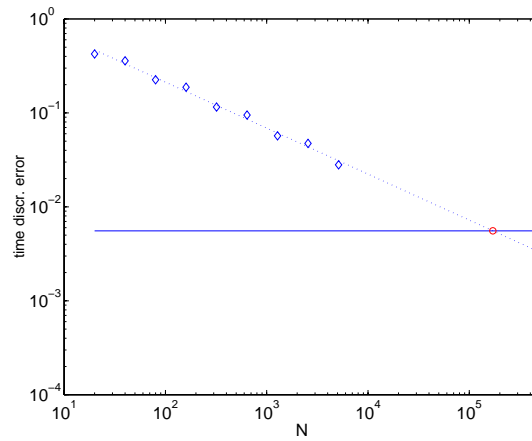


FIGURE 3. (Example 5.1). Computations with uniform time steps and variance reduction. The computational order of convergence for the time discretization is 0.49 and the estimated number of time steps to achieve the accuracy $TOL_{TT} \approx 5.6 \times 10^{-3}$ becomes 1.7×10^5 . The computational time discretization error is given by $A(g(X(T)) - g(\bar{X}(T)); 2^{15})$.

REFERENCES

1. V.Kh. Abdullaev, J.C. Bronski and G. Papanicolaou, *Soliton perturbations and the random Kepler problem*, Phys. D., **135**, 369-386, 2000
2. M. Ainsworth and J. T. Oden, *A posteriori error estimation in finite element analysis*, Comput. Methods Appl. Mech. Engrg., **142** (1997), 1-88.
3. I. Babuška and W. C. Rheinboldt, *Error estimates for adaptive finite element computations*, SIAM J. Numer. Anal., **15** (1978), 736-754.

Algorithm S

Merge	M	$\mathcal{A}(N; M)$	$\mathcal{S}(N; M)$	$\sum MN$
1	2^{15}	1.8×10^2	5×10	140×10^6
0	2^{15}	1.9×10^2	5×10	146×10^6

TABLE 5. (Example 5.2) Both computations use antithetic variates and in this case there is no significant advantage in using the merging procedure. The estimate for the expected value of the optimal number of time steps is in this case 1.7×10^2 .

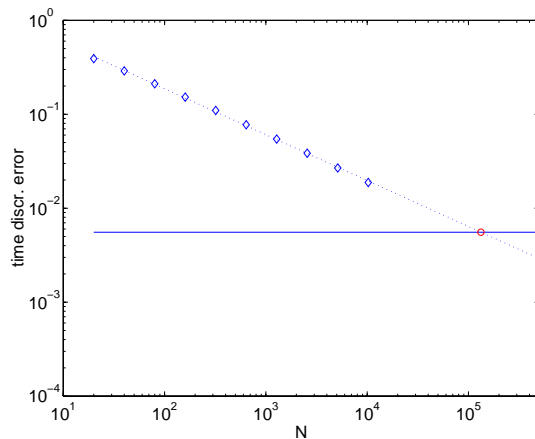


FIGURE 4. (Example 5.2). Computations with uniform time steps and variance reduction. The order of convergence for the time discretization is 0.49 and the estimated number of time steps to achieve the accuracy $\text{TOL}_{TT} \approx 5.6 \times 10^{-3}$ becomes 1.3×10^5 . The computational time discretization error is given by $\mathcal{A}(g(X(T)) - g(\bar{X}(T)); 2^{15})$.

4. I. Babuška and M. Vogelius, *Feedback and adaptive finite element solution of one-dimensional boundary value problems*, Numer. Math. **44** (1984), no. 1, 75–102.
5. R. Becker and R. Rannacher, *A feed-back approach to error control in finite element methods: basic analysis and examples*, East-West J. Numer. Math., **4** (1996), no. 4, 237–264.
6. R. Becker and R. Rannacher, *An optimal control approach to a posteriori error estimation in finite element methods*, Acta Numerica, 2001, 1–102.
7. A. Cohen, W. Dahmen and R. DeVore, *Adaptive wavelet method for elliptic operator equations: convergence rates*, Math. Comp. , **70** No. 233 (2000), 27–75.
8. W. Dahmen, *Wavelet and multiscale methods for operator equations*, Acta Numerica, **6** (1997), 55–228.
9. R. A. DeVore, *Nonlinear approximation*, Acta Numerica, (1998), 51–150.
10. R. Durrett, *Probability: theory and examples*, Duxbury Press, Belmont CA, 1964.
11. K. Eriksson, D. Estep, P. Hansbo and C. Johnson, *Introduction to adaptive methods for differential equations*, Acta Numerica, (1995), 105–158.
12. G. S. Fishman, *Monte Carlo: Concepts, Algorithms, Applications*, Springer-Verlag, New York, 1996.
13. J. M. Hammersley and K. W. Morton *A new Monte Carlo technique: antithetic variates* Proc. Cambridge Philos. Soc. **52**, 449–475, 1956.
14. N. Hofmann, T. Müller-Gronbach and K. Ritter, *Optimal approximation of stochastic differential equations by adaptive step-size control*, Math. Comp. **69**, (2000), 1017–1034.

15. N. Hofmann, T. Müller-Gronbach and K. Ritter, *The optimal discretization of stochastic differential equations*, J. Complexity **17**, (2001), 117–153.
16. E. Harrier, S.P. Norsett and G. Wanner, *Solving Ordinary Differential Equations I*, Springer-Verlag, 1993.
17. C. Johnson and A. Szepessy, *Adaptive finite element methods for conservation laws based on a posteriori error estimates*, Comm. Pure Appl. Math., **48**, (1995), 199–234.
18. E. Jouini, J. Cvitanic and M. Musiela, *Option pricing, interest rates and risk management* Cambridge University Press, 2001.
19. I. Karatzas and S.E. Shreve, *Brownian motion and stochastic calculus*, Graduate Texts in Mathematics 113, Springer-Verlag, New York, 1988.
20. P.E. Kloeden and E. Platen *Numerical solution of stochastic differential equations*, Applications of Mathematics, **23**, Springer-Verlag, New York, 1992.
21. A. Majda, I. Timofeyev and E. Vanden Eijnden, *A mathematical framework for stochastic climate models*, Comm. Pure Appl. Math., **54**, 891-974, 2001.
22. G. N. Milstein, *Numerical integration of stochastic differential equations*, Mathematics and its Applications, **313**, Kluwer Academic Publishers, Dordrecht, 1995.
23. C. Moler, *Random thoughts*, <http://www.mathworks.com/company/newsletter/pdf/Cleve.pdf>
24. K.-S. Moon, A. Szepessy, R. Tempone and G.E. Zouraris, Hyperbolic differential equations and adaptive numerics, in *Theory and numerics of differential equations, Durham 2000* (Eds. J.F. Blowey, J.P. Coleman and A.W. Craig), Springer Verlag, in press.
25. K.-S. Moon, A. Szepessy, R. Tempone and G.E. Zouraris, *Convergence rates for adaptive approximation of ordinary differential equations based on global and local errors*, preprint, 2001.
26. T. Müller-Gronbach, *The optimal uniform approximation of systems of stochastic differential equations*, Nr. A-8-2000, Freie Universität Berlin, Preprint 2000.
27. D. Nualart, *The Malliavin calculus and related topics*, Probability and its Applications, Springer-Verlag, New York, 1995.
28. B. Øksendal, *Stochastic differential equations*, Fifth edition, Springer-Verlag, Berlin, 1998.
29. S. K. Park and K. W. Miller, *Random number generators: Good ones are hard to find*, Comm ACM, vol. 32, 1988.
30. W. H. Press *Numerical recipes in FORTRAN: the art of scientific computing* 2nd edition, Cambridge University Press, Cambridge, United Kingdom, 1992
31. A. Szepessy, R. Tempone and G.E. Zouraris, *Adaptive weak approximation of stochastic differential equations*, Comm. Pure Appl. Math. **54** (2001), no. 10, 1169–1214.
32. D. Talay, *Simulation of Stochastic Differential Systems*, Probabilistic Method in Applied Physics, (P. Krée ; W. Wedig Eds.), 54–96, Lecture notes in physics 451, Springer, 1995.
33. D. Talay and L. Tubaro, *Expansion of the global error for numerical schemes solving stochastic differential equations*, Stochastic Anal. Appl., **8**, 483–509, 1990.

INSTITUTIONEN FÖR NUMERISK ANALYS OCH DATALOGI, KUNGL. TEKNISKA HÖGSKOLAN, S-100 44 STOCKHOLM, SWEDEN.
E-mail address: moon@nada.kth.se

MATEMATISKA INSTITUTIONEN, KUNGL. TEKNISKA HÖGSKOLAN, S-100 44 STOCKHOLM, SWEDEN.
E-mail address: szepessy@nada.kth.se

INSTITUTIONEN FÖR NUMERISK ANALYS OCH DATALOGI, KUNGL. TEKNISKA HÖGSKOLAN, S-100 44 STOCKHOLM, SWEDEN.
E-mail address: rtempone@nada.kth.se

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF THE AEGEAN, GR-83200 KARLOVASSI, SAMOS, GREECE.
E-mail address: zouraris@nada.kth.se