A VARIABLE STEP SIZE RIEMANNIAN SUM FOR AN ITÔ INTEGRAL

E. RAPOO,* *University of South Africa*

Abstract

We investigate the problem of using a Riemannian sum with random subintervals to approximate the iterated Itô integral $\int w \, dw$—or, equivalently, solving the corresponding stochastic differential equation by Euler's method with variable step sizes. In the past this task has been used as a counterexample to illustrate that variable step sizes must be used with extreme caution in stochastic numerical analysis. This article establishes a class of variable step size schemes which do work.

Keywords: Adaptive time discretisation; variable step size; numerical simulation; pathwise approximation; stochastic differential equation; mean-square numerical method

2000 Mathematics Subject Classification: Primary 65C20; 65C30
Secondary 60H05

1. Introduction

It is well known that an Itô integral can be approximated by a suitably chosen Riemannian-type sum. Namely, if we select a series of partitions

$$\tau_n = \{t_i^{(n)}, \ i = 0, \dots, N(n)\},$$

where $0 = t_0^{(n)} < t_1^{(n)} < t_2^{(n)} < \cdots < t_{N(n)}^{(n)} = t$, of the interval $[0, t]$ such that the maximum step lengths in the partitions decrease to 0 as n increases then, for any suitably bounded integrand Y,

$$\int_0^t Y(t) \, dw(t) = \lim_{n \to \infty} \sum_{j=0}^{N(n)-1} Y(t_j^{(n)})(w(t_{j+1}^{(n)}) - w(t_j^{(n)})),$$

where the convergence is almost surely. (This can be proved using Burkholder's inequality. See, for instance, [2, p. 269 and subsequent calculations].) (Here and subsequently, w is a one-dimensional Brownian motion defined in a suitable probabilistic setting; see [10] for the theory of stochastic Itô integrals.) For this approximation to work, the partitions must be deterministic, or they can be random times provided that they depend on the Brownian motion in a nonanticipating way: they have to be stopping times with respect to the underlying Brownian filtration $\mathcal{F}_t = \sigma(w_s, s \le t)$. In other words, the division points must either be chosen in advance, or at least may not look into the future behaviour of the Brownian path $w(\cdot, \omega)$. For partitions not chosen according to these guide lines, the Riemannian sum may not converge, or might converge to an incorrect value.

This limitation in the choice of subdivisions follows from the definition of the Itô integral, which relies on the correct combination of a filtration $\mathcal{F}_t$, a Brownian motion w, and a nonanticipating integrand Y. In particular, the construction of the Itô integral needs the results

$$\mathrm{E}[w_T - w_S \mid \mathcal{F}_S] = 0,$$
$$\mathrm{E}[(w_T - w_S)^2 \mid \mathcal{F}_S] = T - S,$$

for subdivision points $S < T$. These do hold for deterministic S and T, and, according to the optional stopping theorem for martingales, also if S and T are bounded $\mathcal{F}_t$-stopping times (in which case $\mathcal{F}_S$ is interpreted as the filtration of the stopping time S—see, for instance [10]). However, these results are not necessarily true if T is allowed to depend on future values of the Brownian path. Indeed, starting from a particular path of a Brownian motion process, it is possible to create a series of subdivision partitions with maximum step length decreasing to 0, which give a wrong value for the quadratic variation of the Brownian path.

These limitations in the choice of discretisation when approximating Itô integrals by Riemannian sums have their counterpart in the numerical analysis of stochastic differential equations (SDEs), where they appear as restrictions in the use of variable step size methods.

In the numerical analysis of ordinary differential equations, adjusting step sizes has long been used to improve the performance of numerical solution methods. An adjustable-step solution method would usually aim to use, at any part of the solution, the longest possible step length which still guarantees a requested accuracy. Since this means that longer step sizes can be used when the solution seems to go well, and shorter step sizes can be reserved for parts of the solution where they are really needed, a variable step size method should be able to give better accuracy with the same effort (or the same accuracy with less effort) compared to a fixed-step scheme.

The most common way to implement a variable step size method is as follows: first, we attempt a step with the current step length; the step is then either accepted or rejected based on the results of the attempted step. If the step size is rejected then a new attempt is made using a shorter step. The decision of whether to accept or reject a step can be simply based on whether the step taken seems to lead to a too large increment of the solution, or on whether an error estimate based on the step taken (such as the ones provided by some Runge–Kutta methods) is within a required tolerance. Sometimes an error estimate can also be used to calculate an optimal step size for the new attempt if the original step was rejected, or the next step if it was accepted.

Using this type of scheme for SDEs would be expected to run into difficulties since one would in effect need to select the discretisation points after gaining information about the future behaviour of the Brownian path. Traditional stochastic analysis can only guarantee convergence of methods of variable step sizes if the random discretisation times are stopping times. This includes, but is not restricted to, the condition that, for any n and i, $t_{i+1}^{(n)}$ must be $\mathcal{F}(t_i^{(n)})$-adapted (which requires that the length of each interval would need to be fully determined by the starting time of that step), described in [11, p. 321]. Indeed, there are variable step size numerical schemes for SDEs which implement nonrandom step size selection (see, for instance, [18]) or nonanticipating random-step selection (see, for instance, [8] or [9]). More generally, we may use spatial discretisation schemes such as given in [1] or [16], since first hitting times for a continuous diffusion process are stopping times. However, a method which backtracks and tries again if things go badly cannot be implemented using only stopping times or deterministic times.

The theory of rough paths can be used to get around this problem, as explained in [7]. The results in that article (formulated for Stratonovich SDEs) state that the convergence of a variable step size scheme is guaranteed for quite general discretisation steps (deterministic or random, even nonanticipating), but only if one ensures that the one-step method used in the scheme agrees sufficiently well with a second-order Taylor expansion of the solution over each interval; thus, getting right not only the increments of the Brownian motion but also its second-order iterated integrals. In the case of a noncommuting SDE with a driving Brownian motion of dimension two or higher, this means that the numerical scheme must utilise the area integrals of the Brownian motion. In the one-dimensional or commuting cases, the second-order iterated integrals can all be expressed in terms of the Brownian increments but for general random subdivision points; it is still nontrivial to require that the numerical method must get the second-order terms right.

In particular, the ruling in [7] implies that the humble Euler–Maruyama scheme for solving Itô differential equations cannot be expected to work for all possible selections of subdivisions as it is a first-order scheme that pays no attention to the second-order integrals. The example given there as an illustration is an attempt to solve the differential equation

$$\mathrm{d}x_t = w_t \, \mathrm{d}w_t, \qquad w_0 = 0, \tag{1}$$

over the time interval $[0, 1]$ using the Euler method and a variable step size scheme. The variable step size scheme uses only two step sizes, h and $h/2$: normally the step size is h, but two steps of length $h/2$ are taken instead whenever

$$|\triangle w| > \lambda h,$$

that is, when the Brownian increment is considered to be 'too large'. If now $h \to 0$ (a series of dyadic subdivisions is used), it turns out that the scheme converges to the true solution of (1) only when $\lambda = 0$ or $\lambda = 1$, cases which are really fixed-step schemes. For all other values of λ, this particular variable step size scheme gives wrong results: the method does converge, but towards a wrong solution.

Many high-order numerical methods and corresponding variable-step strategies have by now been derived for solving SDEs. (See, for instance, [7], or [4], [5], and [14], which are based on embedded pairs of stochastic Runge–Kutta schemes.) However, the implementation of such schemes is often complicated, especially in the high-dimensional case, since these schemes need to be supplied with various iterated path integrals of the Brownian path—or at least reasonable approximations of these. What makes things particularly difficult for the variable step size schemes is the fact that, for a strong solution, we must ensure that the iterated path integrals and the path increment do come from one and the same Brownian path, and, therefore, we might need to generate, for instance, the conditional area of a Brownian motion over the two halves of an interval, conditional on knowing its area over the whole integral. (See, for instance, [3] and [6] for an account of the difficulties.) For Brownian increments alone, this can of course be done quite easily, using Paul Lévy's construction of Brownian motion.

These practical difficulties in implementing higher-order methods are the reason why the Euler–Maruyama method is still popular: it is the simplest method available, and needs neither derivatives of the coefficient function nor further information about the Brownian path, beyond the path increments. For noncommuting systems of SDEs, it remains the highest-order scheme with easily simulated random variables (see [13], [15], and [20]), and often an Euler method with very small step sizes is used to provide a 'true' solution to compare other numerical

methods to. This popularity is despite its low order of convergence (the L^2-norm of the global error is of the order $h^{0.5}$) and its instability problems. Of course, the low order means that this method in particular could really do with an adjustable-step implementation!

Most existing attempts to implement the Euler method with a variable step size control strategy do not attempt step size adaptation based on the Brownian path; for instance, Lamba *et al.* [12] only adapted the steps to the drift component, and Römisch and Winkler [18] used a nonrandom-step selection. See also the weak adaptive time-step approximation in [19].

The aim of this article is to attempt to answer the following question: if we do wish to use the Euler method to solve the differential equation (1), is there any kind of genuinely random variable step size strategies that *will* work (that is, will improve on the performance of the Euler scheme and also converge to the correct solution)? Since solving (1) using Euler's method is equivalent to approximating the Itô integral $\int w \, \mathrm{d}w$ by a Riemannian-type sum along the subdivision steps used in the Euler scheme, this is equivalent to asking which kinds of random subdivision points could be used in the Riemannian sum to ensure convergence towards the correct Itô integral, while using fewer terms that need to be summed.

Of course, the exact value of the iterated integral $\int w \, \mathrm{d}w$ over any interval is easy to calculate: according to the Itô formula we have $2 \int_0^t w_s \, \mathrm{d}w_s = w_t^2 - t$, which incidentally also makes error analysis in this case very easy. But the case of this simple integral is very illustrative and will give us some valuable information on how to use variable step sizes in conjunction with the Euler–Maruyama method. Also, according to the Itô–Taylor expansion, solutions to general SDEs can be expressed in terms of this and higher-order iterated integrals, and, therefore, the results in this article give an indication of what the minimum requirements would be for a correctly functioning variable step size Euler method for solving more general SDEs in one or higher dimensions.

We will start by describing, in Section 2, the adjusted step size mechanism we will adopt. A brief investigation of the local (one-step) and global errors follows in Section 3. In Section 4 we investigate the bias (L^1-norm) of the global errors, which leads us to introduce several categories of unbiased subdivision criteria whose performance we will compare in the rest of the article. In Section 5 we compare the mean-square error of the unbiased subdivision schemes, and in Section 6 we give a final cost-versus-improvement analysis of the variable step size methods. Finally, in Section 7 we present our conclusions.

2. The variable step size strategy

We wish to use the Euler method, implemented with a simple adjustable step size mechanism which follows the counterexample in [7], to solve the two-dimensional Itô SDE

$$\mathrm{d}x_t = 2y_t \, \mathrm{d}w_t, \qquad \mathrm{d}y_t = \mathrm{d}w_t, \tag{2}$$

with initial values $x_0 = 0$ and $y_0 = 0$ over the interval [0, 1]. Here, w_t is a one-dimensional Brownian motion. Of course, this means that $y_t = w_t$ and x_t is given by

$$x_t = 2 \int_0^t w_s \, \mathrm{d}w_s. \tag{3}$$

In our variable step size numerical scheme the interval [0, 1] is divided into equispaced intervals of length h, but at each interval we have the option of taking two steps of length $h/2$ instead of one step of length h. The decision of whether to take one longer or two shorter steps is made based on the increment of the Brownian motion over the interval in question.

Euler's method applied to solve (2) over a subdivision $0 = t_0 < t_1 < \cdots < t_N = 1$ of the interval $[0, 1]$ uses the one-step iteration formula

$$Y_{i+1} = Y_i + \triangle w_i, \qquad X_{i+1} = X_i + 2Y_i \triangle w_i,$$

where $X_0 = 0$, $Y_0 = 0$, $X_i = X(t_i)$, $Y_i = Y(t_i)$, $w_i = w(t_i)$, and $\triangle w_i = w(t_{i+1}) - w(t_i)$. (Note that throughout this article the upper-case letters X and Y will be used for the numerical solutions, and the lower-case letters x and y will be used for the true solutions.) This will give us the approximate solution X, where

$$X_{i+1} - X_i = 2w_i \triangle w_i$$

and

$$X_k = 2 \sum_{i=0}^{k-1} w(t_i) \triangle w_i, \qquad k = 1, \ldots, N. \tag{4}$$

We see that the Euler discretisation scheme is identical to approximating the Itô integral in (3) by the corresponding Riemannian sum along the given subdivision steps.

The approximate solution function $X(t)$, $t \in [0, 1]$, to the SDE can then be constructed by taking $X(t_k) = X_k$ for $k = 0, \ldots, N$ and using, for instance, linear interpolation between the subdivision points. However, in this article we will only consider the errors at the final point $t = 1$.

In the next two subsections we will discuss the error in the Euler scheme for general subdivision schemes, before returning to our simple full-or-half step scheme.

A subdivision scheme is defined here as a collection $\{\tau_n, \ n = 1, 2, \ldots\}$ of partitions of $[0, 1]$. Namely, we define $\tau_n = \{t_i^{(n)}, \ i = 0, \ldots, N(n)\}$, where $0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_{N(n)}^{(n)} = 1$ gives a partition of the interval $[0, 1]$. The lengths of the subdivision intervals are denoted by $h_i^{(n)} = t_{i+1}^{(n)} - t_i^{(n)}$, and $\delta(\tau_n)$ denotes the maximum step length of partition τ_n:

$$\delta(\tau_n) = \max\{h_i^{(n)}, \ i = 0, \ldots, N(n) - 1\}.$$

For all subdivision schemes, we assume that $\delta(\tau_n) \downarrow 0$ as $n \to \infty$.

The subdivision scheme τ_n can be deterministic or random, in which case it will be determined from the Brownian path, possibly in a nonadaptive way.

2.1. The deterministic subdivision case

If the subdivision scheme τ_n is deterministic, that is, the set of division points $0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_{N(n)}^{(n)} = 1$ of the interval $[0, 1]$ are predetermined (but not necessarily equispaced) for each n, then we know that the Euler scheme converges towards the true solution and, identically, the Riemannian sum converges towards the corresponding Itô integral as n increases: the series of Euler/Riemannian sum approximations $X^{(n)}(1)$ based on the subdivision scheme τ_n, that is,

$$X^{(n)}(1) = 2 \sum_{i=0}^{N(n)-1} w(t_i^{(n)}) \triangle w_i, \tag{5}$$

converges towards the correct solution $x(1)$ almost surely. In fact, we can use the independence of the increments of the Brownian motion over disjoint deterministic intervals to calculate the

error as

$$
\mathrm{E}[x(1) - X^{(n)}(1)]^2 = 4 \sum_{i=0}^{N(n)-1} \mathrm{E}\left[\int_{t_i}^{t_{i+1}} (w(s) - w(t_i^{(n)})) \, dw_s \right]^2
$$

$$
= 2 \sum_{i=0}^{N(n)-1} (h_i^{(n)})^2.
$$

2.2. General subdivisions

For an error estimate valid over random subdivision schemes τ_n, we can use the fact that, for all $t_k \le t_{k+1}$,

$$
2 \int_{t_k}^{t_{k+1}} (w_u - w_{t_k}) \, dw_u = (w_{t_{k+1}} - w_{t_k})^2 - (t_{k+1} - t_k)
$$

$$
= \triangle w_k^2 - h_k. \tag{6}
$$

This gives an expression for the one-step (local) error in the Euler scheme, or, equivalently, the one-term error in the Riemannian sum expression, (5). We see that the total error in the Riemannian sum, or the global error of the Euler method, when a subdivision scheme τ_n is used, is given by

$$
x(1) - X^{(n)}(1) = \sum_{i=0}^{N(n)-1} (\triangle w_k^2 - h_k^{(n)})
$$

$$
= \left(\sum_{i=0}^{N(n)-1} \triangle w_k^2 \right) - 1.
$$

It follows that the Euler method converges towards the correct solution, or, equivalently, the series of Riemannian sums converges towards the correct Itô integral, if and only if the subdivision scheme τ_n is such that

$$
\sum_{i=0}^{N(n)-1} \triangle w_k^2 \to 1
$$

almost surely as n increases—that is, if the subdivision scheme gives the right value for the quadratic variation of the Brownian motion. (This is in agreement with the results in [7]. According to Corollary 4.4 there, the Euler method (4) converges towards the correct solution along a subdivision scheme τ_n if, for all k, h_k and $(\triangle w_k)^2$ agree up to a term which is of order $o(\delta(\tau_n))$.) This does hold for any predetermined (nonrandom) subdivision scheme, as well as for τ_n consisting of stopping times, but not if the subdivision points can be selected using information about the Brownian path.

2.3. The full-or-half step variable step size strategy

We will now return to the simple adjusted step size scheme we will base our calculations on. We will use dyadic subdivisions, so that, for any given value of n, the interval $[0, 1]$ will be divided into $N(n) = 2^n$ intervals of length $h = 2^{-n}$. This gives us the preliminary subdivision points $\tau_n = \{k2^{-n}, k = 0, \ldots, 2^n\}$. Over each of these intervals of length $h = 2^{-n}$, the step length is either accepted as it is, or else it is rejected and two steps of length $h/2 = 2^{-(n+1)}$ are taken instead. The decision of whether to accept or reject a step will be based on the size of

the Brownian increment over the interval. More precisely, a step of length h will be rejected if $\triangle w \in A_h$, where, for each $h > 0$, A_h is a predetermined set. The natural scaling of $\triangle w$ suggests that $A_h = \{\sqrt{h}x : x \in A_1\}$ for all $h > 0$, where A_1 is a fixed set (that is, reject a step if $\triangle w / \sqrt{h} \in A_1$). However, we might also wish to consider other types of scaling.

The final subdivision scheme $\hat{\tau}_n$ will be a random one with $\tau_n \subseteq \hat{\tau}_n \subseteq \tau_{n+1}$—more precisely,

$$\hat{\tau}_n = \tau_n \cup \left(\bigcup_{k: \triangle w_k \in A_{2^{-n}}} (t_k + 2^{-(n+1)}) \right),$$

so the numerical approximation is

$$X^{(n)}(1) = \sum_{t_i \in \hat{\tau}_n} 2w(t_i)\triangle w_i.$$

This can also be written as

$$X^{(n)}(1) = \sum_{t_i \in \tau_n} \triangle \hat{X}_i, \tag{7}$$

where $\triangle \hat{X}_i$, the increment of the solution over the ith step in τ_n, is adjusted according to the variable-step scheme: $\triangle \hat{X}_i = 2w(t_i)\triangle w_i$ if $\triangle w_i \in A_h^C$ (the complement of the set A_h) and $\triangle \hat{X}_i = 2w(t_i)\triangle_1 w_i + 2w(t_i + h/2)\triangle_2 w_i$ if $\triangle w_i \in A_h$. Here, $h = 2^{-n}$, and $\triangle_1 w_i$ and $\triangle_2 w_i$ are the increments of the Brownian motion over the two half-steps of length $h/2$, that is, $\triangle_1 w_i = w(t_i + h/2) - w(t_i)$ and $\triangle_2 w_i = w(t_{i+1}) - w(t_i + h/2)$ (and $\triangle w_i = \triangle_1 w_i + \triangle_2 w_i$). If $A_h = \varnothing$ then this reverts to an ordinary Euler scheme with fixed-step length h; if $A_h = \mathbb{R}$, we get an Euler scheme with the fixed-step length $h/2$. For any other types of sets A_h, we will get different types of variable step size schemes, and the question is which is the optimal choice of the sets A_h. To answer this, first we need to find the errors made in each such scheme, and then we need to compare the improvement in accuracy with the increased cost of the calculations.

3. Error calculations

Since the decision of whether to subdivide an interval $[t_i, t_{i+1}]$ and take two steps instead of one is based only on the Brownian increment over that interval, even after applying our nonadapted variable step size method, it is true that the local errors of method (7) over different intervals in τ_n are independent. Also, we need not worry about error propagation as the global error for our differential equation/Riemannian sum is simply the sum of the local errors. We will now proceed to analyse the local and global errors of our variable step size method.

Using (7), we see that the global error $\hat{E}^{(n)} = x(1) - X^{(n)}(1)$ can be written as

$$\hat{E}^{(n)} = \sum_{t_i \in \tau_n} \hat{e}_i,$$

where $\hat{e}_i$ is the error of the adjustable step size Euler method over the τ_n-interval $[t_i, t_{i+1}]$ of length $h = 2^{-n}$. All the $\hat{e}_i$s are independent, identically distributed random variables, so in the following we will drop the index i and consider a time interval $[t_0, t_1]$ of length h. The time midpoint, $t_{1/2} = (t_0 + t_1)/2$, divides this interval into two subintervals of length $h/2$. The increment of the Brownian motion over $[t_0, t_1]$ is denoted by $\triangle w$, and we have $\triangle w = \triangle_1 w + \triangle_2 w$, where $\triangle_1 w$ and $\triangle_2 w$ are the increments of the Brownian motion over $[t_0, t_{1/2}]$ and $[t_{1/2}, t_1]$, respectively.

Now, if we were always to take a full step of length h (if $A_h = \varnothing$ in our variable step size scheme) then the one-step error over the interval $[t_0, t_1]$ would be, as given in (6),

$$e_{\text{full}} = (\triangle w)^2 - h;$$

similarly, if $A_h = \mathbb{R}$ so that two half-steps are always taken, the one-step error over the interval $[t_0, t_1]$ will be

$$e_{\text{half}} = (\triangle_1 w)^2 - \frac{h}{2} + (\triangle_2 w)^2 - \frac{h}{2} = (\triangle_1 w)^2 + (\triangle_2 w)^2 - h.$$

Clearly, both e_{half} and e_{full} are always bounded below by $-h$. We also have $\mathrm{E}[e_{\text{half}}] = 0 = \mathrm{E}[e_{\text{full}}]$, and $\mathrm{var}(e_{\text{full}}) = 2h^2$ and $\mathrm{var}(e_{\text{half}}) = h^2$.

For general A_h, we have

$$\hat{e} = \mathbf{1}_{A_h^{\mathrm{C}}}(\triangle w) e_{\text{full}} + \mathbf{1}_{A_h}(\triangle w) e_{\text{half}}.$$

The first step in analysing this is to consider the conditional values of e_{full} and e_{half}, given the value of $\triangle w$. Of course, if $\triangle w = x$ is known then $e_{\text{full}} = x^2 - h$ is fully determined. The law of e_{half} given $\triangle w = x$ is easy to find: the conditional distribution of e_{half} is that of $x^2/2 - h + 2y^2$, where y is a normal random variable, independent of $\triangle w$, with mean 0 and variance $h/4$. (A standard Brownian bridge argument is used to find the distribution of $w(t_{1/2}) = w(t_0) + x/2 + y$. See, for instance, [17, p. 35].) In particular,

$$\mathrm{E}[e_{\text{half}} \mid \triangle w = x] = \tfrac{1}{2}(x^2 - h),$$
$$\mathrm{var}(e_{\text{half}} \mid \triangle w = x) = \tfrac{1}{2}h^2,$$
$$\mathrm{E}[e_{\text{half}}^2 \mid \triangle w = x] = \tfrac{1}{2}h^2 + \tfrac{1}{4}(x^2 - h)^2,$$

and the distribution function of e_{half}, given that $\triangle w = x$, over the interval of length h is given by

$$F_{h,x}(x) \equiv \mathrm{P}(e_{\text{half}} \leq x \mid \triangle w = x)$$
$$= 1 - 2\Phi\left(-\sqrt{\frac{2(x + h - x^2/2)}{h}}\right),$$

where Φ is the standard normal distribution function.

Conditioning on the value of $\triangle w$, we can then find the distribution of $\hat{e} = e_i$, the one-step error when our adjustable-step Euler method is used. For our purposes, we need only the first two moments of $\hat{e}_i$. Using the conditional distributions derived above, we obtain

$$\mathrm{E}[\hat{e}_i] = -\int_{A_h} \frac{1}{2}(x^2 - h) p(h, x) \, \mathrm{d}x,$$

and, similarly,

$$\mathrm{E}[\hat{e}_i^2] = 2h^2 + \int_{A_h} \left(\frac{1}{2}h^2 - \frac{3}{4}(x^2 - h)^2\right) p(h, x) \, \mathrm{d}x,$$

where $p(t, x)$ is the transition probability density function of the standard Brownian motion:

$$p(t, x) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right).$$

Theorem 1. *In the scheme 'subdivide if $\triangle w \in A_h$', as defined in (7), for the global error $\hat{E} = x(1) - X^{(n)}(1)$, we have*

$$\mathrm{E}[\hat{E}] = -\int_{A_h} \frac{1}{2} \frac{(x^2 - h)}{h} p(h, x) \, \mathrm{d}x$$

and

$$\mathrm{E}[\hat{E}^2] = 2h + \int_{A_h} \left(\frac{h}{2} - \frac{3}{4} \frac{(x^2 - h)^2}{h} \right) p(h, x) \, \mathrm{d}x - h \, \mathrm{E}[\hat{E}]^2 + \mathrm{E}[\hat{E}]^2.$$

Proof. Since $\hat{E} = \sum_{t_i \in \tau_n} \hat{e}_i$, where the $\hat{e}_i$ are independent, identically distributed random variables, we have

$$\mathrm{E}[\hat{E}] = \sum_{t_i \in \tau_n} \mathrm{E}[\hat{e}_i] = \frac{1}{h} \mathrm{E}[\hat{e}_i] = -\int_{A_h} \frac{1}{2} \frac{(x^2 - h)}{h} p(h, x) \, \mathrm{d}x,$$

$$\mathrm{var}(\hat{E}) = \frac{1}{h} \mathrm{var}(\hat{e}_i) = \frac{1}{h} (\mathrm{E}[\hat{e}_i^2] - \mathrm{E}[\hat{e}_i]^2) = \frac{1}{h} \mathrm{E}[\hat{e}_i^2] - h \, \mathrm{E}[\hat{E}]^2.$$

Using

$$\mathrm{E}[\hat{E}^2] = \mathrm{var}(\hat{E}) + \mathrm{E}[\hat{E}]^2,$$

completes the proof.

The first two moments of the global error $\hat{E}$ in our variable step size Euler scheme were expressed above in terms of the basic step length h and the 'rejection sets' A_h. We are interested in which selections A_h satisfy the following criteria.

- Which selections give unbiased methods, that is, methods for which $\mathrm{E}[\hat{E}] = 0$ for all h?

- Which selections give convergent methods, in that $\mathrm{E}[\hat{E}^2] \downarrow 0$ as $h \to 0$ (or as $n \to \infty$ in the dyadic subdivision scheme)?

- Which methods improve on a fixed-step Euler method by giving better accuracy at the same cost, or the same accuracy at less cost?

4. Bias and how to avoid it

By an unbiased method we mean one for which, for all h, the expected value of the global error vanishes. From Theorem 1 we see that our variable step size scheme is unbiased if, for each h, A_h is such that

$$\mathrm{E}[\hat{E}] = -\int_{A_h} \frac{1}{2} \frac{(x^2 - h)}{h} p(h, x) \, \mathrm{d}x = 0.$$

We will write

$$\mathrm{E}[\hat{E}] = \int_{A_h} G_h(x) \, \mathrm{d}x,$$

where

$$G_h(x) = -\frac{1}{2} \left(\frac{x^2 - h}{h} \right) \frac{1}{\sqrt{2\pi h}} \exp\left(-\frac{x^2}{2h} \right).$$
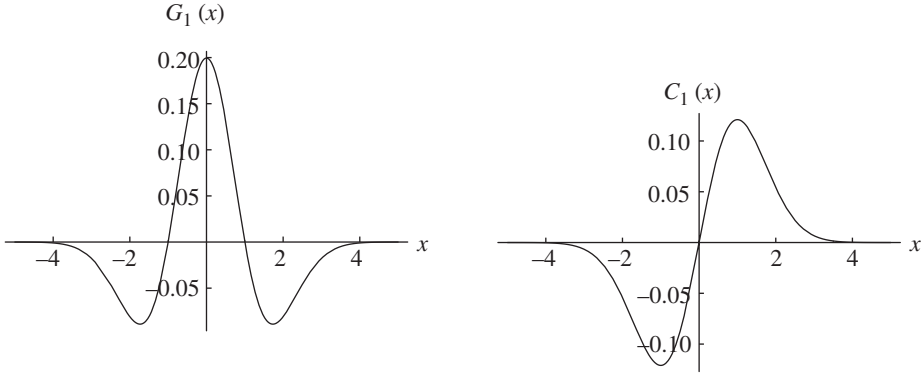
FIGURE 1: Plots of the functions G_1 and C_1.

The function $G_h(x)$ scales with respect to h as

$$G_1(x) = \sqrt{h}\, G_h(x\sqrt{h}).$$

A graph of G_1 is given in Figure 1. Note that $G_h(x) = 0$ at $x = \pm\sqrt{h}$.

The cumulative integral of $G_h(y)$, which we will denote by $C_h(x)$, can be calculated in closed form:

$$C_h(x) = \int_{-\infty}^{x} G_h(y)\,\mathrm{d}y = \frac{1}{2\sqrt{2\pi}}\, \frac{x}{\sqrt{h}}\, \exp\!\left(-\frac{x^2}{2h}\right).$$

This cumulative function scales as follows:

$$C_1(x) = C_h(x\sqrt{h}).$$

A graph of C_1 is given in Figure 1. Note that, for each value of h, the absolute maximum and minimum values (reached at $\pm\sqrt{h}$) have the same value, i.e.

$$\max\{C_h(x),\ x \in \mathbb{R}\} = \frac{1}{2\sqrt{2\pi}}\, \exp\!\left(-\frac{1}{2}\right) \approx 0.1209,$$

$$\min\{C_h(x),\ x \in \mathbb{R}\} = -\frac{1}{2\sqrt{2\pi}}\, \exp\!\left(-\frac{1}{2}\right) \approx -0.1209.$$

The value of the cumulative function equals the bias for a method in which $A_h = (-\infty, x]$ is chosen for each h. The choices $x = -\infty$ or $x = +\infty$ lead to $A_h = \mathbb{R}$ and $A_h = \varnothing$ (fixed step size methods with step lengths $h/2$ and h, respectively), which are therefore seen to be unbiased methods. Another unbiased method is obtained with $x = 0$, that is, if we subdivide whenever the Brownian increment is negative.

This choice of $A_h = (-\infty, x]$ already hints at things to come, since clearly there the bias for a fixed x will not necessarily decrease with h. However, this is not a rational choice for a rejection criterion. Firstly, since the Brownian increments are symmetric about the origin, the set A_h should also be symmetric. Secondly, $\triangle w$ does scale with $\sqrt{h}$ and, therefore, upper and lower limits defining A_h should presumably also have $\sqrt{h}$ as a coefficient. These considerations are also reflected in the symmetry and scaling of the function G_h. Note that when we adopt this kind of scaling for the set A_h (that is, if $A_h = \{\sqrt{h}x : x \in A_1\}$), the resultant bias will not

depend on h but will always be equal to

$$\mathrm{E}[\hat{E}] = \int_{A_1} G_1(x)\,\mathrm{d}x.$$

In [7] the following choice was made: a step is rejected if the Brownian increment is too large, in the sense that $|\triangle w|/\sqrt{h} > \lambda$, where λ is a positive parameter. In our terminology this means taking

$$A_h^C = [-\lambda\sqrt{h}, \lambda\sqrt{h}],$$

which gives a bias, depending only on λ, of

$$\mathrm{bias}(\lambda) = -\frac{1}{\sqrt{2\pi}}\lambda \exp\left(-\frac{\lambda^2}{2}\right).$$

This is equal to 0 for $\lambda = 0$ (always subdivide) or for $\lambda = \infty$ (never subdivide), but, for any other value of λ, it has a strictly negative value, where the largest bias with value

$$-\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-1/2} \approx -0.241\,971$$

is reached at $\lambda = 1$.

It is clear from the graph of G_1 why this choice for A_h is not good: we will always integrate over a section of G_1 which is strictly negative, and, for $\lambda = 1$, we will integrate over all the negative parts of G_1.

Similarly, we will easily see that the strategy of subdividing if

$$\frac{|(\triangle w)^2 - h|}{h} > \lambda$$

(which might seem natural: subdivide if the one-step error is too large) will also have a negative bias.

An investigation of G_1 suggests what kind of choice of A_h would give zero bias. This leads to the following result.

Theorem 2. *We have* $\mathrm{E}[\hat{E}] = 0$ *for the following choices of sets* A_h:

(a) $A_h = \mathbb{R}_-$ *or* $A_h = \mathbb{R}_+$,

(b) $A_h = \{x \colon |x| \leq K\sqrt{h} \text{ or } |x| \geq L\sqrt{h}\}$, *where* $L \geq 1$ *and*

$$K = \sqrt{-W(-L^2\exp(-L^2))},$$

(c) $A_h = \{x \colon K\sqrt{h} < |x| < L\sqrt{h}\}$, *where* L *and* K *are as above.*

Here $W(x)$ is the 'product logarithm' function: $z = W(x)$ is the (real) solution to $z\exp(z) = x$.

Remark 1. Note that if $L > 1$ then we will get $0 < K < 1$. If $L = 1$ then $K = 1$, so Theorem 2(b) also covers the case in which $A_h = \mathbb{R}$ and Theorem 2(c) also covers the case in which $A_h = \varnothing$.

Proof of Theorem 2. Case (a) we have already dealt with.

(b) The bias for $A_1 = \{x : |x| \leq K \text{ or } |x| \geq L\}$ can be evaluated as

$$\mathrm{E}[\hat{E}] = \int_{A_1} G_1(x)\,\mathrm{d}x$$

$$= \frac{1}{\sqrt{2\pi}}\left(K\exp\left(-\frac{K^2}{2}\right) - L\exp\left(-\frac{L^2}{2}\right)\right);$$

thus, for a fixed $L \geq 1$, the values of K for which $\mathrm{E}[\hat{E}] = 0$ are either $K = L$ or $K = \sqrt{-W(-L^2\exp(-L^2))}$. The result for general h in (b) follows by scaling, and the result in (c) from (b), completing the proof.

To summarise, we have identified the following three basic types of unbiased subdivision method.

- *Method A:* Subdivide for positive $\triangle w$ only or for negative $\triangle w$ only. (The decision of whether to subdivide is therefore taken according to the sign of $\triangle w$, but independently of its size.)

- *Method B:* Subdivide when $|\triangle w|/\sqrt{h}$ is either too small (less than or equal to K) or too large (greater than or equal to L), with cutoff points $L \geq 1$ and K, $0 < K \leq 1$, determined from L.

- *Method C:* The complement of method B (accept a step without subdivision when $|\triangle w|/\sqrt{h}$ is too small or too large).

Note that of course there are infinitely many other choices leading to no bias, built from various combinations of the sets above.

Remark 2. In methods B and C, the variable L can be used as a parameter to regulate the frequency of rejected steps. Figure 2 shows the cutoff points which lead to an unbiased method for each choice of L: the upper cutoff point L (dashed line) and the corresponding lower cutoff point K (solid line).
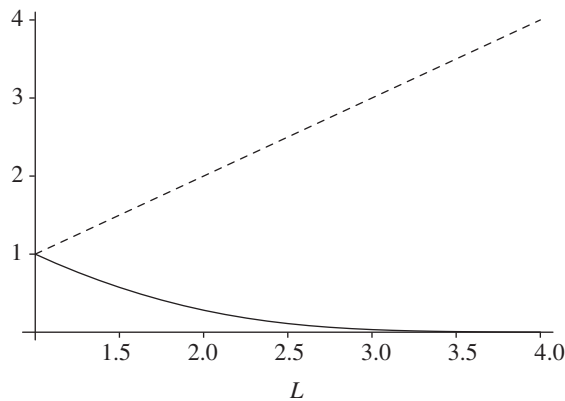

FIGURE 2: The upper cutoff point L (*dashed line*) and the lower cutoff point $K(L)$ (*solid line*) for unbiased methods for varying L.

5. Convergence matters

In the previous section we explained why the subdivision criteria suggested in [7] is biased, and specified three different types of subdivision criteria which do give an unbiased method, that is, for which the expected error is 0 for each and every value of h. But of course we also want the methods to converge.

As a measure of convergence, we will use here the mean square of the global error.

According to Theorem 1, the mean square of the global error $\hat{E}$ for an unbiased method is given by

$$\mathrm{E}[\hat{E}^2] = 2h + \int_{A_h} \left(\frac{h}{2} - \frac{3}{4} \frac{(x^2 - h)^2}{h} \right) p(h, x) \, \mathrm{d}x.$$

If $A_h = \{\sqrt{h}x : x \in A_1\}$ holds then again, for all h,

$$\mathrm{E}[\hat{E}^2] = h \left(2 + \int_{A_1} \left(\frac{1}{2} - \frac{3}{4}(x^2 - 1)^2 \right) p(1, x) \, \mathrm{d}x \right),$$

meaning that, for any choice of A_1, our variable step size modification of the Euler scheme does converge with an error of order h in the mean-square sense. While we therefore cannot expect a better order of error, we can still hope to decrease the size of the error.

For $A_1 = \varnothing$ (never subdivide), we obtain $\mathrm{E}[\hat{E}^2] = 2h$ and, for $A_1 = \mathbb{R}$ (always subdivide), $\mathrm{E}[\hat{E}^2] = h$. (These are the mean-square global errors of the Euler method with fixed step size h and step size $h/2$, respectively.)

Of our three unbiased categories of method, method A (subdivide for positive values only or negative values only) has the mean-square error $\mathrm{E}[\hat{E}^2] = 3h/2$ (in that method we use each of the step sizes h and $h/2$ half of the time on average).

For method B, the mean-square error cannot be calculated in closed form. A plot of the mean-square error as a function of the upper cutoff point L is given in Figure 3(a). The smallest possible mean-square norm, with value $0.8698h$ for method B occurs with $L \approx 1.441\,15$, which means accepting the step for $0.617\,213 < |\triangle w|/\sqrt{h} < 1.441\,15$. Note that this gives better convergence, in the mean-square sense, than always subdividing, despite using fewer steps.

In method C, which is the complement of method B, the mean-square error is $2h$ minus the error for method B with the same value of L.
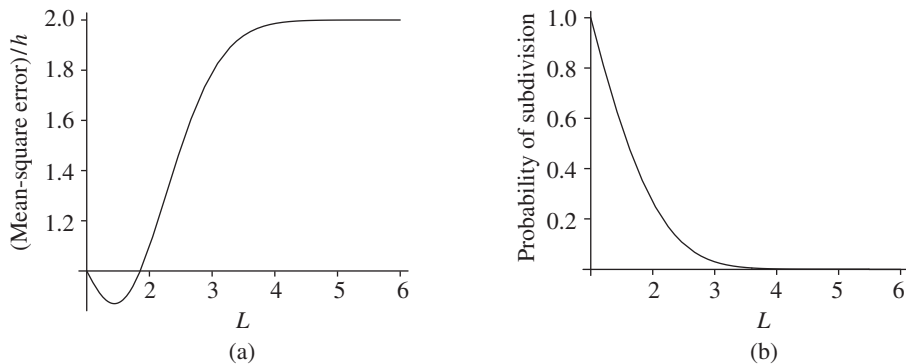

FIGURE 3: (a) Mean-square global error of the variable-step method ($h = 1$); and (b) the probability of subdivision, as a function of the cutoff point L.

6. Efficiency comparisons

We have proved that with due care, it is possible to implement the Euler method for solving (2) with a variable-step algorithm that converges to the correct answer in the mean-square sense. We still need to establish which of these methods actually improve on the fixed-step Euler method, and what is the best choice for the cutoff set A_h.

As always in variable-step methods, the improvement in accuracy will need to be weighted against increased cost in the sense of computational effort. To be preferable, the variable step size method must provide the same accuracy with less cost, or better accuracy at the same cost, compared to the corresponding fixed-step scheme.

In [7] and [14] the variable step size scheme was implemented such that a step with the current step size is attempted, and then accepted or rejected based on the results. The cost of calculations is then taken to be indicated by the number of attempted steps. Here the situation is different in that the decision of whether to accept the step or to subdivide is based simply on the size of the Brownian increment $\triangle w$ over the suggested step interval.

To come up with a cost measure for our case, we note that, for the fixed-step method, the computational effort over one step consists of generating the Brownian increment $\triangle w$ and calculating the Euler increment. In the variable-step method, on the other hand, there is the added cost of determining whether $\triangle w \in A_h$ (i.e. whether the step should be accepted or rejected). If the step is rejected, so that two steps are taken instead of one, then the increments $\triangle_1 w$ and $\triangle_2 w$ must be established. However, the cost of generating the original step and generating the midpoint for the subdivided step can be assumed to be roughly the same (namely, the cost of generating one normal random variable), and far larger than the cost of making the decision of whether to subdivide or not. Based on this, we will measure the cost by the total number of steps taken.

For the fixed-step scheme, we take $1/h$ steps, while, for the variable-step scheme, the number of steps taken will vary from $1/h$ to $2/h$, depending on the choice of the set A_h and on the particular Brownian path. Of course, the average number of steps taken in all cases is easy to find.

As a first comparison of the performance of our variable-step methods, we will simply compare the mean-square errors for a given variable-step method and for the fixed-step method using as its step length the average step length of the variable-step method.

For method A (subdivide for positive $\triangle w$ only or for negative $\triangle w$ only), we subdivide on average half of the original steps and therefore with the basic step length h, the average step length is $3h/4$. The mean-square error $3h/2$ of the variable step size method is identical to that of the fixed-step method with step length $3h/4$.

For method B (subdivide when $|\triangle w|$ is too large or too small; accept steps $K\sqrt{h} \le |\triangle w| \le L\sqrt{h}$), the probability of subdividing a step depends on the product logarithm function and cannot be given in closed form. A graph of the probability of subdivision as a function of the upper cutoff point L is shown in Figure 3(b). From the subdivision probabilities we can find the average step sizes. Figure 4(a) shows, for $h = 1$ and varying values of L, a comparison of the mean-square errors of the variable step size method B and the corresponding fixed-step scheme using the same average number of steps. We see that, for all values of $L > 1$, the variable step size method is more efficient than the corresponding fixed-step method. Figure 4(b) shows the ratio of the mean-square errors, variable/fixed steps, again as a function of L. The best performance is given by selecting $L = 1.702\,04$, for which the mean-square global error of the variable step size is $66.032\,7\%$ of that of the fixed step size. This means accepting steps with $0.440\,622 \le |\triangle w|/\sqrt{h} \le 1.702\,04$; about 43% of steps will be subdivided.
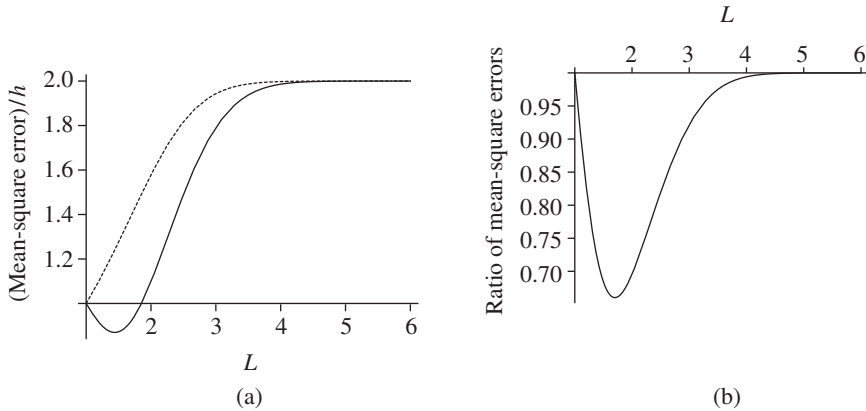
FIGURE 4: (a) Mean-square global errors of the variable-step method (*dashed line*) and the corresponding fixed-step method (*solid line*). (b) Ratio of the mean-square errors of the variable versus the fixed step size methods, where both errors are a function of the cutoff point L.

Remark 3. Note that, more generally, the following holds. Let us fix the number of steps N to be used by the numerical methods. Firstly, apply a fixed-step method using N steps over $[0, 1]$, and, secondly, apply the variable step size method B with parameter L such that the average number of steps over $[0, 1]$ is N. To apply the variable-step method, we need to first calculate the base step size which depends on N and L; applying the method itself will also depend on L. However, the ratio of the mean-square errors of the fixed versus the variable step size methods for any given N do not depend on N but only on L, as shown in Figure 4(b).

For method C, not surprisingly, we find that the mean-square error is always larger than for the corresponding fixed-step method.

Remark 4. The use of the number of steps to measure the cost of the computations above was justified in the case where the Brownian increments are generated as the calculations progress, one step at the time. If the Brownian motion increments are instead retrieved, for instance, from a previously generated Brownian tree, then the reasoning still holds if the h and $h/2$ increments of it are already available. If only h steps of the Brownian motion are available, while $h/2$ steps need to be generated, then it is also necessary to compare the cost of generation versus retrieval.

Above we compared just the average performance of the variable versus fixed-step methods; however, since we are implementing a strong step size adjustment method here, which hopefully intervenes to improve accuracy for particularly 'bad' paths, a strong (pathwise) comparison of the two methods may be more relevant. To this end, $50\,000$ Brownian paths were generated. For each path, firstly, the variable step size method B was applied, and then the fixed-step method using the same number of steps—note that now the number of steps varies for each path. The end results of the two methods were compared, and to measure the performance of the variable-step method in cases where the fixed-size method fares badly, the results were sorted according to the error in the fixed-step method. Figure 5 shows, for different cutoff levels L, the ratio of the absolute errors of the two methods for the $p\%$ worst fixed-step errors. It can be seen, for instance, that with all choices of L,
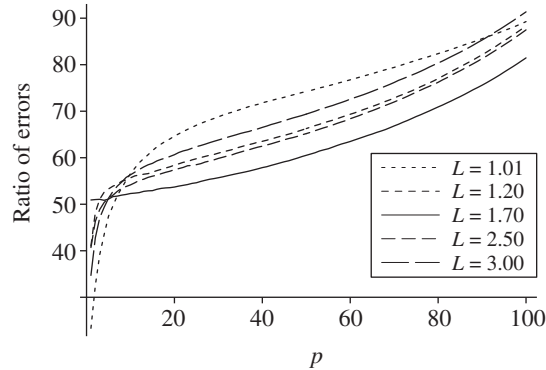
FIGURE 5: Ratio of global errors (variable versus fixed step size methods) in the $p\%$ worst fixed-step cases for different cutoff points L.

the error can be halved in the 10% worst paths. Note that again $L = 1.7$ gives the best performance.

7. Conclusion

We have demonstrated that the reason the variable step size algorithm used as a counter-example in [7] failed was simply because the wrong kind of criteria was used in determining when $\triangle w$, the increment of the Brownian motion on an interval, was 'bad': A more successful method can be based on subdividing if $|\triangle w|$ is too big or too small, so that the step is accepted for $K \leq |\triangle w| \leq L$. The values of L and K have to be chosen suitably, to remove the bias noted in the counterexample of [7]. The resulting variable-step scheme, implemented here in a very simple subdivide-or-not way, can then improve the performance of the Euler method. Further improvement can be obtained, for instance, by repeated subdivision.

Although we considered here only the very special differential equation (2), which in effect amounts to finding which variable step choices do not distort the value of the sum approximation of the Itô integral $\int w \, dw$, or, equivalently which give the correct quadratic variation to the Brownian motion, it is hoped that the result here can be utilised in developing variable step variations of the Euler–Maruyama scheme applied to more general differential equations. The integral $\int w \, dw$ is a key second-order iterated integral in the stochastic Taylor expansion and, therefore, a variable step size method which introduces bias to that integral cannot be expected to converge towards the correct solution.

References

[1] BICHTELER, K. (1981). Stochastic integration and L^p theory of semimartingales. *Ann. Prob.* **9,** 49–89.
[2] BOULEAU, N. AND LÉPINGLE, D. (1994). *Numerical Methods for Stochastic Processes*. John Wiley, New York.
[3] BURRAGE, K., BURRAGE P. M. AND TIAN, T. (2004). Numerical methods for strong solutions of stochastic differential equations: an overview. *Proc. R. Soc. London* **460,** 373–402.
[4] BURRAGE, P. M. AND BURRAGE K. (2002). A variable stepsize implementation for stochastic differential equations. *SIAM J. Sci. Comput.* **24,** 848–864.
[5] BURRAGE P. M., HERDIANA, R. AND BURRAGE, K. (2004). Adaptive stepsize based on control theory for stochastic differential equations. *J. Comput. Appl. Math.* **170,** 317–336.
[6] GAINES, J. G. AND LYONS, T. L. (1994). Random generation of stochastic area integrals. *SIAM J. Appl. Math.* **54,** 1132–1146.

[7] GAINES, J. G. AND LYONS, T. L. (1997). Variable step size control in the numerical solution of stochatic differential equations. *SIAM J. Appl. Math.* **57,** 1455–1484.

[8] HOFMANN, N., MÜLLER-GRONBACH T. AND RITTER, K. (2000). Optimal approximation of stochastic differential equations by adaptive step-size control. *Math. Comput.* **69,** 1017–1034.

[9] HOFMANN, N., MÜLLER-GRONBACH T. AND RITTER, K. (2001). The optimal discretization of stochastic differential equations. *J. Complexity* **17,** 117–153.

[10] IKEDA, N. AND WATANABE, S. (1981). *Stochastic Differential Equations and Diffusion processes*. North-Holland, Amsterdam.

[11] KLOEDEN, P. E. AND PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin.

[12] LAMBA, H., MATTINGLY, J. C. AND STUART, A. M. (2007). An adaptive Euler–Maruyama scheme for SDEs: convergence and stability. *IMA J. Numerical Anal.* **27,** 479–506.

[13] LEHN, J., RÖSSLER, A. AND SCHEIN, O. (2002). Adaptive schemes for the numerical solution of SDEs – a comparison. *J. Comput. Appl. Math.* **138,** 297–308.

[14] MAUTHNER, S. (1998). Step size control in the numerical solution of stochastic differential equations. *J. Comput. Appl. Math.* **100,** 93–109.

[15] MILSTEIN, G. N. AND TRET´YAKOV, M. V. (1997). Mean-square numerical methods for stochastic differential equations with small noises. *SIAM J. Sci. Comput.* **18,** 1067–1087.

[16] NEWTON, N. J. (1990). An efficient approximation for stochastic differential equations on the partition of symmetrical first passage times. *Stoch. Stoch Reports* **29,** 227–258.

[17] REVUZ, D. AND YOR, M. (1994). *Continuous Martingales and Brownian Motion*, 2nd edn. Springer, Berlin.

[18] RÖMISCH, W. AND WINKLER, R. (2006). Stepsize control for mean-square numerical methods for stochastic differential equations with small noise. *SIAM J. Sci. Comput.* **28,** 604–625.

[19] SZEPESSY, A., TEMPONE R. AND ZOURARIS, G. E. (2001). Adaptive weak approximation of stochastic differential equations. *Commun. Pure Appl. Math.* **54,** 1169–1214.

[20] TALAY, D. (1995). Simulation of stochastic differential systems. In *Probabilistic Methods in Applied Physics* (Lecture Notes Physics **451**), eds P. Kree and W. Wedig, Springer, Berlin, pp. 54–96.