One-Step Approximations For Stochastic Functional Differential Equations

Evelyn Buckwar [1]

Humboldt-Universität zu Berlin, Institut für Mathematik, Unter den Linden 6, 10099 Berlin.

Abstract

We consider the problem of strong approximations of the solution of Itô stochastic functional differential equations (SFDEs). We develop a general framework for the convergence of drift-implicit one-step schemes to the solution of SFDEs. We provide examples to illustrate the applicability of the framework.

Key words: Stochastic functional differential equations, Mean-square convergence, Drift-implicit one-step schemes.
PACS: 65C30, 60H35, 34K20, 34K50

1 Introduction

The subject of this note is a general framework for analysing mean-square convergence results for one-step methods approximating the solution of a system of Itô stochastic functional differential equations (SFDEs). We consider SFDEs in the form

$$X(t) = X(0) + \int_0^t F(s, X_s)\ \mathrm{d}s + \sum_{j=1}^m \int_0^t G^j(s, X_s)\ \mathrm{d}W^j(s) \quad \text{for } t \in [0, T], \quad (1)$$

$$X(t) = \Psi(t) \qquad \text{for } t \in J, \qquad \text{where } J := [-\tau, 0]. \quad (2)$$

By $C(J; \mathbb{R}^d)$ we mean the Banach space of all continuous paths from $J \to \mathbb{R}^d$ equipped with the supremum norm $\|\eta\| := \sup_{s \in J} |\eta(s)|$. Throughout the article $|.|$ denotes the Euclidean norm in $\mathbb{R}^d$ and $\langle \cdot, \cdot \rangle$ its induced scalar product. As

Email address: `buckwar@mathematik.hu-berlin.de` (Evelyn Buckwar).
[1] The author was supported by the Deutsche Forschungsgemeinschaft by a grant "Eigene Stelle"and the SFB 373 (Quantifikation und Simulation Ökonomischer Prozesse), Humboldt-Universität zu Berlin.

Preprint submitted to Elsevier Science

usual in the literature on functional differential equations, X_t denotes the history or memory functional of X, which is defined as $X_t(u) = \{X(t+u) : u \in J\}$ with $X_t \in C(J;\mathbb{R}^d)$. Let $(\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t\in[0,T]}, \mathbb{P})$ be a complete probability space with the filtration $\{\mathcal{A}_t\}_{t\in[0,T]}$ satisfying the usual conditions (that is, it is increasing and right-continuous, and each $\{\mathcal{A}_t\}$, $t \in [0,T]$ contains all $\mathbb{P}$-null sets in $\mathcal{A}$) and let $W(t) = (W^1(t), \ldots, W^m(t))^T$ be an m-dimensional standard Brownian motion on that probability space. We denote the mean-square norm of a vector-valued square-integrable random variable $z \in L_2(\Omega;\mathbb{R}^d)$ by $\|z\|_{L_2} := (\mathbb{E}\,|z|^2)^{1/2}$, where $\mathbb{E}$ is expectation with respect to $\mathbb{P}$. The initial path $\Psi(t) : J \to \mathbb{R}^d$ is assumed to be a continuous and $\mathcal{A}_0$-measurable random variable such that $\|\Psi\|_{L_2} < \infty$.

The history functional X_t provides a very general description of the dependence on the past. Of course, the simplest case is given by (deterministic or stochastic) ordinary differential equations (ODEs), that is when there is no dependence on the past. However, a large variety of specific forms of memory appear in the literature and it is common numerical practice to develop appropriate methods according to that specific form. For further reference we give here examples of widespread classes of SFDEs. If we set F and G to be of the form

$$H(s, X(s), X(s-\tau)) \qquad \text{or} \qquad H(s, X(s), X(s-\tau(s))), \tag{3}$$

then we call (1) an SFDE with a discrete or variable lag, respectively. The function $\tau(s)$ with $\tau(s) \le s$ may be random, i.e. a random process independent of the driving Wiener process (see [20, Chap.VI,§3]). If we let F and G represent memory terms of the type

$$H(s, X(s), \int_{-\tau}^{0} K(s,u,X(s+u))\, \mathrm{d}u) \tag{4}$$

then we call (1) an SFDE with a distributed delay term. Note that we do not consider state-dependent delays or vanishing lags here. We refer for example to [5,9,12,14,19,20] for a general background on (deterministic and stochastic) functional differential equations (FDEs).

1.1 A brief review of methods and aims

The representation of the memory term is a central problem in the numerical analysis of FDEs, deterministic and stochastic. In the deterministic literature a large variety of methods has been developed and analysed. The simplest case (apart from the ODE case) of one or more discrete, commensurable delays and a fixed step-size essentially reduces the problem to that of numerical analysis of methods for ODEs and Bellman's method of steps. However, all standard ODE methods have been extended to more complicated types of memory

as well, with for example Hermite interpolation or continuous extensions of Runge-Kutta methods or quadrature dealing with the memory functionals. Several computer packages designed for the numerical solution of FDEs are available on the web. The recent book by Bellen and Zennaro [5] provides a good account of the results in the deterministic case.
For stochastic FDEs the situation is much less satisfactory. A theorem concerning mean-square convergence for explicit one-step schemes applied to SFDEs with discrete delays and global Lipschitz coefficient functions has been presented in [2] (see also the references in that article for previous work on the topic). However, the main method used and investigated is the Euler-Maruyama method, i.e. the stochastic version of the basic Euler scheme. The consistency analysis in [2] was performed for the Euler-Maruyama method. The latter has also been applied to SFDEs with variable delays and local Lipschitz conditions on the coefficient functions, using an interpolation at non-meshpoints by piecewise constants, in [15]. In [16] the Euler-Maruyama method has been applied to SFDEs with the general memory term X_t, where this was linearly interpolated, under local and global Lipschitz conditions on the coefficient functions. In [10] the authors developed and analysed a Milstein scheme for SFDEs with discrete delays. It turns out that the appropriate version of the Itô-formula for the numerical analysis on that class of SFDEs requires the application of Malliavin calculus.
The goal of this article is to provide a systematic approach to the analysis of mean-square convergence for drift-implicit one-step schemes for the approximation of the solution of Eq. (1) under global Lipschitz conditions on the coefficient functions. As in the numerical analysis of deterministic ordinary differential equations (e. g. [8, Chapter II.3], principally for Runge-Kutta methods) and delay differential equations (e. g. [5, Thm. 3.2.8] and [21,22]) or stochastic ordinary differential equations (e. g. [17, Thm. 1.1]), Theorem 1 constitutes the basis for the detailed analysis of particular methods. The theorem says that under certain conditions global error estimates for a method can be inferred from estimates on its local error. An analysis of the local error in methods for FDEs also includes the analysis of the error made in the representation of the memory term. Thus, on the basis of Theorem 1, one can concentrate on the questions concerning quadrature and interpolation methods arising in a stochastic setting. Another important issue is the analysis of methods for equations with small noise (see e.g. [7,18]), i.e. equations of the form

$$X(t) = X(0) + \int_0^t F(s, X_s)\,\mathrm{d}s + \sum_{j=1}^{m} \int_0^t \epsilon\,\hat{G}^j(s, X_s)\,\mathrm{d}W^j(s),$$

where ϵ is a small parameter and the $\hat{G}^j$'s and their derivatives are of moderate size. In this case it would be useful to apply reliable continuous Runge-Kutta methods in the approximation of the drift part with appropriate extensions for the diffusion part. A further example, SFDEs with random lags, also shows that the analysis of numerical methods for functional differential equation needs to take into account the approximation of the specific history term.

This topic will be pursued in future work.
To illustrate the applicability of our general framework we give in Section 4 examples of classes of SFDEs and show how the conditions required in Theorem 1 can be interpreted in these cases.

2 Definitions and Preliminaries

We assume that the drift and diffusion functionals $F : [0,T] \times C(J;\mathbb{R}^d) \to \mathbb{R}^d$ and $G^j : [0,T] \times C(J;\mathbb{R}^d) \to \mathbb{R}^d$, $j = 1,\ldots,m$, are continuous and satisfy uniform Lipschitz conditions and linear growth conditions with respect to their second argument. By [14, Thm 5.2.3] there exists a path-wise unique strong solution to Eq. (1).

We define a family of meshes with a uniform step-size on the interval $[0,T]$ by

$$\mathcal{T}_h^N := \{0 = t_0 < t_1 < t_2 < \cdots < t_N\} \subseteq [0,T], \tag{5}$$

with $t_n = nh,\ n = 0,\ldots,N,\ hN \leq T,\ N \in \mathbb{N}$.
For some types of delay, e.g. for (3) with variable lags, the arguments $t_n - \tau(t_n)$ will not (necessarily) be mesh-points. In this case, after choosing a step-size h and thus a mesh $\mathcal{T}_h^N$, we can define a second non-uniform mesh, which consists of all the $t_n \in \mathcal{T}_h^N$ and, in addition, all the points $t_n - \tau(t_n)$. Thus we may define

$$\mathcal{S}_{\widetilde{N}} := \{t_0 = s_0 < s_1 < \cdots < \cdots < s_{\widetilde{N}} = t_N\} \subseteq [0,T], \tag{6}$$

where (e.g.) for some ℓ, k one has $s_\ell = t_n - \tau(t_n),\ s_k = t_n$ for $t_n \in \mathcal{T}_h^N$ and $t_n - \tau(t_n) > 0$. The number $\widetilde{N}$ is N plus the (finite) number of points which additionally have to be included. We can also represent any point $s_\ell \in \mathcal{S}_{\widetilde{N}}$, with $t_n < s_\ell \leq t_{n+1}$, by

$$s_\ell = t_n + \zeta h, \quad \text{where } t_n, t_{n+1} \in \mathcal{T}_h^N \text{ and } \zeta \equiv \zeta(s_\ell) \in (0,1]. \tag{7}$$

We denote by $\mathcal{J}_h$ the similarly discretized initial interval where $\mathcal{J}_h \subseteq J$.
The approximation method given below incorporates a finite number, say r, of multiple Itô-integrals and it depends linearly on them. We denote multiple Wiener integrals with some function $f(.)$ as an integrand by

$$I_{(j_1,j_2,\ldots,j_l)}^{(t,t+h),\tau_2(t),\ldots,\tau_l(t)}(f) = \int\limits_t^{t+h} \int\limits_{t-\tau_l(t)}^{s_l-\tau_l(t)} \cdots \int\limits_{t-\tau_2(t)}^{s_2-\tau_2(t)} f(s_1)\,\mathrm{d}W^{j_1}(s_1)\ldots\mathrm{d}W^{j_l}(s_l), \tag{8}$$

where $j_i \in \{0,1,\ldots,m\},\ i = 1,\ldots,l$ and $\mathrm{d}W^0(s) = \mathrm{d}s$. If $f \equiv 1$ we omit the argument (f), if $\tau_2(t),\ldots,\tau_l(t)$ are not required and if there is only one Wiener process involved, we omit the corresponding superscripts. For more background on multiple Itô-integrals see [11] or [17], examples of (8) are given by

$$I_{(1)}^{(t,t+h)} = \int_t^{t+h} \mathrm{d}W(s), \qquad I_{(1,2)}^{(t,t+h)} = \int_t^{t+h}\int_t^{s_1} \mathrm{d}W^2(s_2)\,\mathrm{d}W^1(s_1)$$

$$\text{and} \qquad I_{(1,1)}^{(t,t+h),\tau} \int_t^{t+h}\int_{t-\tau}^{s_1-\tau} \mathrm{d}W(s_2)\,\mathrm{d}W(s_1).$$

Note that only the first two integrals appear in the numerical analysis of stochastic ordinary differential equations, the third integral is peculiar to stochastic delay differential equations and is required in the Milstein scheme for SFDEs with constant delays, see [10]. If necessary, then using the fixed step-size h on $\mathcal{T}_h^N$ makes it possible to construct the multiple Itô-integrals from left to right on the fixed, but non-uniform mesh $\mathcal{S}_{\widetilde{N}}$.

In our discussion of numerical methods we will denote by $Y(t_n)$ the approximation of $X(t_n)$ at some point t_n in $\mathcal{T}_h^N$. For simplicity the initial values will be taken as $Y(t_i) := \Psi(t_i)$ for $t_i \in \mathcal{J}_h$. Further, $\{Y_{t_n}^{(F/G)_h}\}$ will denote the approximation of the memory functional X_{t_n} in Eq. (1) at $t_n \in \mathcal{T}_h^N$ and $I_{\phi_i}^{t_n,t_{n+1}}$ represents the collection of multiple Wiener integrals (8) corresponding to the increment function ϕ_i. We define drift-implicit one-step methods for the simulation of the solution X of (1) as

$$Y(t_{n+1}) = Y(t_n) + h\,\phi(t_{n+1}, t_n, \{Y_{t_{n+1}}^{F_h}\}, \{Y_{t_n}^{F_h}\}) + \sum_{i=1}^{r} \phi_i(t_n, \{Y_{t_n}^{G_h}\}) I_{\phi_i}^{t_n,t_{n+1}}. \quad (9)$$

The approximation $\{Y_{t_n}^{(F/G)_h}\}$ is a discrete memory functional from $\mathcal{J}_h$ to $\mathbb{R}^d$. In general it will depend on the particular structure of the dependence on the past in (1). For (3), where (with a certain constraint on the step-size h) the delayed arguments are always part of the mesh $\mathcal{T}_h^N$, the approximation $\{Y_{t_n}^{(F/G)_h}\}$ simply consists of a finite number of already computed values of $Y(t_n)$. If the delayed arguments are not part of $\mathcal{T}_h^N$, but in $\mathcal{S}_{\widetilde{N}}$, we define a continuous interpolant of (9) using the representation (7), i.e. for $t_n \in \mathcal{T}_h^N$ and $\zeta \in (0,1]$ we set

$$\begin{aligned} Y(t_n + \zeta h) &= Y(t_n) + \widetilde{\phi}(\zeta h, t_{n+1}, t_n, \{Y_{t_{n+1}}^{F_h}\}, \{Y_{t_n}^{F_h}\}) \\ &\quad + \sum_{i=1}^{r} \widetilde{\phi}_i(t_n, \{Y_{t_n}^{G_h}\}) I_{\widetilde{\phi}_i}^{t_n, t_n+\zeta h}\,. \end{aligned} \quad (10)$$

The increment functions $\widetilde{\phi}$ and $\widetilde{\phi}_i$ may be given by continuous extensions of the method (9) itself or by some interpolation method. Thus the method (9) together with (10) mirror the *standard approach via continuous ODE methods* presented in Chapters 3 and 4 in Bellen & Zennaro [5]. For (4) the evaluation of the memory functional requires an additional approximation such as the replacement of an integral by a quadrature formula. Thus, the functionals F and

G (and their derivatives) appearing in the increment functions ϕ, ϕ_i, etc. will be replaced by discretized functionals F_h and G_h and we denote this in (9) and (10) by using the superscripts F_h and G_h on $\{Y_{t_n}\}$. We shall require that the discrete functionals F_h and G_h converge in the $\|.\|_{L_2}$ norm to F and G, respectively. Note that it may be necessary to give the corresponding convergence proof when considering e.g. numerical quadrature of stochastic processes, as it may not be part of the existing theory on numerical quadrature.
The method (9), together with the interpolant (10) and the discrete functionals F_h and G_h are required to generate iterates $Y(t_n)$ which are $\mathcal{A}_{t_n}$-measurable. Further important properties required of the increment functions in (9) and consequently in $\{Y_{t_n}^{(F/G)_h}\}$ and (10) are that they are uniformly Lipschitz continuous with the right-hand side of the Lipschitz condition resulting in a finite sum. To achieve this we introduce some rather technical conditions, we give in Section 4 examples of their use. It would be more elegant at this point to write the Lipschitz condition with a supremum over the delay interval (as in [22]), however subsequent estimates with the expectation would be more difficult.

Lipschitz conditions on the increment functions: Assume that there exist positive constants L_ϕ, L_{ϕ_i}, a positive real number M and an integer L, which may both depend on the step-size h, but their product ML is finite. Then we require of ϕ and ϕ_i for all discrete memory functionals $\xi,\ \eta : \mathcal{J}_h \to \mathbb{R}^d$ and all $t, t' \in \mathcal{T}_h^N, t > t'$ and a finite subset $\mathcal{S}_L(t')$ of $\mathcal{S}_{\widetilde{N}}$, $\sharp(\mathcal{S}_L(t')) = L$ ($\sharp(.)$ denoting the number of elements of a set), with $s_\ell < t'$ for all $s_\ell \in \mathcal{S}_L(t')$:

$$|\phi(t,t',\xi_t,\xi_{t'}) - \phi(t,t',\eta_t,\eta_{t'})| \\ \leq L_\phi \Big(|\xi(t)-\eta(t)| + |\xi(t')-\eta(t')| + M \sum_{s_\ell \in \mathcal{S}_L(t')} |\xi(s_\ell)-\eta(s_\ell)| \Big), \quad (11)$$

$$\mathbb{E}\ \Big(\{\phi_i(t,\xi_t)\ -\ \phi_i(t,\eta_t)\}\ I_{\phi_i}^{t,t+h}\ |\mathcal{A}_t \Big)\ =\ 0\ , \quad (12)$$

$$\mathbb{E}\Big(|\{\phi_i(t',\xi_{t'})\ -\ \phi_i(t',\eta_{t'})\}\ I_{\phi_i}^{t',t'+h}|^2 \Big) \\ \leq\ h\ L_{\phi_i}(\,\mathbb{E}|\xi(t')-\eta(t')|^2\ + M \sum_{s_\ell \in \mathcal{S}_L(t')} \mathbb{E}\ |\xi(s_\ell)-\eta(s_\ell)|^2\ ). \quad (13)$$

We shall establish a relationship between convergence, that is, the behaviour of the global error of the approximation to consistency, that is the local error, measured in an appropriate way. We would like to point out that for the analysis of one-step schemes essentially two different but related concepts are used in the literature. In the first one the local error is defined as the defect that is obtained when the exact solution values are inserted into the numerical scheme. In the second one the local error is defined as the difference after one step of the exact and the numerical solution started at an arbitrary deterministic value. These concepts differ in the way the error is transported to the end of the integration interval, in the first via the numerical method, in the second via the exact solution. The second definition has been used by Milstein in the proof of Theorem 1.1 ([17]). For comparison of these principles in the deterministic setting see [8, Chapters II.3,III.4]. In this article we will consider the first approach and we now define what we understand by local errors.

Definition 1 *The local error for the method (9) is defined, with $X(t_n)$ being the exact solution of (1) for $t_n \in \mathcal{T}_h^N$, as the sequence of random vectors in $\mathbb{R}^d$*

$$\delta_h(t_n) = X(t_{n+1}) - X(t_n) \\ - h\, \phi(t_{n+1}, t_n, X^{F_h}_{t_{n+1}}, X^{F_h}_{t_n}) - \sum_{i=1}^{r} \phi_i(t_n, X^{G_h}_{t_n}) I^{t_n,t_{n+1}}_{\phi_i}\,, \tag{14}$$

where $X^{(F/G)_h}_{t_n}$ denotes the evaluation of the discrete functionals F_h and G_h using the exact solution X.

The functionals F and G have to be replaced by discrete approximations such as quadrature formulas when the memory is as (4). For memory of the type (3) $X^{(F/G)_h}_{t_n}$ simply becomes the solution X evaluated at the delayed arguments. We will consider mean-square consistency and convergence of our approximations in the following sense.

Definition 2 *The approximation Y for the solution X of (1) is said to be* mean-square consistent with order p *($p > 0$) if the following estimates hold:*

$$\max_{t_n \in \mathcal{T}_h^N} \|\mathbb{E}(\, \delta_h(t_n)|\mathcal{A}_{t_n})\|_{L_2} \quad \leq \quad C\; h^{p+1} \quad \text{as } h \to 0, \tag{15}$$

and

$$\max_{t_n \in \mathcal{T}_h^N} \|\delta_h(t_n)\|_{L_2} \quad \leq \quad C\; h^{p+\frac{1}{2}} \quad \text{as } h \to 0, \tag{16}$$

where the generic constant C does not depend on h, but may depend on T and the initial data.

Definition 3 *The approximation Y for the solution X of Eq. (1), defined on $\mathcal{T}_h^N$, is said to be* mean-square convergent, with order p, *on the mesh-points, when*

$$\max_{t_n \in \mathcal{T}_h^N} \|X(t_n) \; - \; Y(t_n)\|_{L_2} \quad \leq \quad C\; h^p \text{ as } h \to 0, \tag{17}$$

where $C < \infty$ is independent of h, but may depend on T and on the initial data.

3 Convergence

We first discuss the solvability of the recurrence equation. It is obvious that the approximations $\{Y(t_n)\}_{n\in\mathbb{N}}$ can be computed iteratively, if ϕ in (9) does

not depend on $Y(t_{n+1})$. When the function ϕ in (9) does depend on $Y(t_{n+1})$, the general (and standard) approach to proving existence and uniqueness of a solution is to assume (global) Lipschitz-continuity of the right-hand side of (9) with respect to $Y(t_{n+1})$, require a Lipschitz constant less than 1, and then to apply Banach's contraction mapping principle. In addition, we have to verify that the mean-square norm of the iterates exists. (The straightforward extension to fully implicit systems would serve as an example were the mean-square norm of the iterates does not exist.) The arguments do not differ in an essential way from those in either the deterministic literature (see e.g. [5, Thm.4.2.3]) or for stochastic ordinary differential equations (see e.g. [24, Thm.5]).
We now state the main result of this article.

Theorem 1 *We assume that F and G in Eq. (1) are uniformly Lipschitz-continuous in their second argument. Further we suppose that the increment functions ϕ and ϕ_i in the recurrence (9) satisfy the estimates (11), (12) and (13). Then we obtain for the approximation (9) of the solution of Eq. (1) the estimate*

$$\max_{t_n \in \mathcal{T}_h^N} \|X(t_n) - Y(t_n)\|_{L_2} \leq C \max_{t_n \in \mathcal{T}_h^N} (h^{-1} \parallel \mathbb{E}(\delta_h(t_n)|\mathcal{A}_{t_n})\|_{L_2} + h^{-\frac{1}{2}} \|\delta_h(t_n)\|_{L_2}), \quad (18)$$

where $C < \infty$ is independent of h, but may depend on T and on the initial data. If the method (9) is mean-square consistent of order p, i.e. the method satisfies (15) and (16), then it is convergent in the sense of Definition 3 with order p.

Proof: Define $\mathrm{e}(t_n) := X(t_n) - Y(t_n)$ for $t_n \in \mathcal{T}_h^N$. Note that the error $\mathrm{e}(t_n)$ is $\mathcal{A}_{t_n}$-measurable, since both $X(t_n)$ and $Y(t_n)$ are $\mathcal{A}_{t_n}$-measurable random variables.
By adding and subtracting $h\ \phi(t_{n+1}, t_n, X^{F_h}_{t_{n+1}}, X^{F_h}_{t_n}) + \sum_{i=1}^{r} \phi_i(t_n, X^{G_h}_{t_n}) I^{t_n,t_{n+1}}_{\phi_i}$ and $X(t_n)$ and rearranging we obtain

$$\mathrm{e}(t_{n+1}) = X(t_{n+1}) - Y(t_{n+1}) = \mathrm{e}(t_n) + \delta_h(t_n) + \mathcal{U}(t_n),$$

where $\delta_h(t_n)$ is given by (14) and $\mathcal{U}(t_n)$ is defined as

$$\begin{aligned}\mathcal{U}(t_n) :=& h\ \phi(t_{n+1}, t_n, X^{F_h}_{t_{n+1}}, X^{F_h}_{t_n}) - h\ \phi(t_{n+1}, t_n, \{Y^{F_h}_{t_{n+1}}\}, \{Y^{F_h}_{t_n}\}) \\ &+ \sum_{i=1}^{r} \Big\{\phi_i(t_n, X^{G_h}_{t_n}) - \phi_i(t_n, \{Y^{G_h}_{t_n}\})\Big\} I^{t_n,t_{n+1}}_{\phi_i}\ . \end{aligned} \quad (19)$$

We will frequently use the Cauchy-Schwarz inequality and the inequality $2\ ab \leq a^2 + b^2$. Various properties of conditional expectation, which can be found in, e. g. [23], will also be applied. We obtain

$$
\begin{aligned}
|\,\mathrm{e}(t_{n+1})|^2 &= \langle \mathrm{e}(t_n) + \delta_h(t_n) + \mathcal{U}(t_n), \mathrm{e}(t_n) + \delta_h(t_n) + \mathcal{U}(t_n)\rangle \\
&\leq |\,\mathrm{e}(t_n)|^2 + |\delta_h(t_n)|^2 + |\mathcal{U}(t_n)|^2 + 2\ \langle \mathrm{e}(t_n), \delta_h(t_n)\rangle \\
&\quad +2\ \langle \mathrm{e}(t_n), \mathcal{U}(t_n)\rangle + 2\ |\mathcal{U}(t_n)|\ |\delta_h(t_n)| \\
&\leq |\,\mathrm{e}(t_n)|^2 + 2\ |\delta_h(t_n)|^2 + 2\ \langle \mathrm{e}(t_n), \delta_h(t_n)\rangle \\
&\quad +2\ \langle \mathrm{e}(t_n), \mathcal{U}(t_n)\rangle + 2\ |\mathcal{U}(t_n)|^2\ .
\end{aligned}
$$

Thus, taking expectation and taking the modulus, yields

$$
\begin{aligned}
\mathbb{E}\ |\,\mathrm{e}(t_{n+1})|^2 \leq \mathbb{E}\ |\,\mathrm{e}(t_n)|^2 + \ 2\ \mathbb{E}\ |\delta_h(t_n)|^2 + \underbrace{2\ |\mathbb{E}\ \langle \mathrm{e}(t_n), \delta_h(t_n)\rangle|}_{1)} \\
+ \underbrace{2\ |\ \mathbb{E}\ \langle \mathrm{e}(t_n), \mathcal{U}(t_n)\rangle\ |}_{2)} + \underbrace{2\ \mathbb{E}\ |\mathcal{U}(t_n)|^2}_{3)}\ . \qquad (20)
\end{aligned}
$$

For the term labelled 1) in (20) we obtain immediately

$$
\begin{aligned}
2\ |\mathbb{E}\ \langle \mathrm{e}(t_n), \delta_h(t_n)\rangle| &= 2\ |\ \mathbb{E}\ \Big(\mathbb{E}\ (\langle \mathrm{e}(t_n), \delta_h(t_n)\rangle | \mathcal{A}_{t_n})\ \Big)| \\
&\leq 2\ \ \mathbb{E}\ |\langle \mathrm{e}(t_n), \mathbb{E}\ (\delta_h(t_n)|\mathcal{A}_{t_n})\ \rangle| \\
&\leq 2\ \Big(\ h\ \mathbb{E}\ |\,\mathrm{e}(t_n)|^2\ \Big)^{\frac{1}{2}} \cdot \Big(\ h^{-1}\ \mathbb{E}\ |\ \mathbb{E}(\ \delta_h(t_n)|\mathcal{A}_{t_n})|^2\Big)^{\frac{1}{2}} \\
&\leq h\ \mathbb{E}\ |\,\mathrm{e}(t_n)|^2 + h^{-1}\ \mathbb{E}\ |\ \mathbb{E}(\ \delta_h(t_n)|\mathcal{A}_{t_n})|^2. \qquad (21)
\end{aligned}
$$

In the terms 2) and 3) the Lipschitz conditions on the increment functions will play their part.
For the term labelled 2) in (20) we first estimate $|\ \mathbb{E}\ (\mathcal{U}(t_n)|\mathcal{A}_{t_n})\ |$ using the definition (19) of $\mathcal{U}(t_n)$ and the Lipschitz condition (11) on ϕ as well as condition (12) on $\phi_i,\ i = 1, \ldots, m$. We obtain

$$
\begin{aligned}
&|\ \mathbb{E}\ (\mathcal{U}(t_n)|\mathcal{A}_{t_n})\ | \\
&= |\ \mathbb{E}\ \Big(h\ \phi(t_{n+1}, t_n, X^{F_h}_{t_{n+1}}, X^{F_h}_{t_n}) - h\ \phi(t_{n+1}, t_n, \{Y^{F_h}_{t_{n+1}}\}, \{Y^{F_h}_{t_n}\}\ ) \\
&\quad + \sum_{i=1}^{r} \Big\{\phi_i(t_n, X^{G_h}_{t_n}) - \phi_i(t_n, \{Y^{G_h}_{t_n}\})\Big\}\, I^{t_n, t_{n+1}}_{\phi_i}\ |\mathcal{A}_{t_n}\Big)\ | \\
&\leq h\ L_\phi\ \Big\{\mathbb{E}\ (|X(t_{n+1}) - Y(t_{n+1})|\ |\mathcal{A}_{t_n}) + \mathbb{E}\ (|X(t_n) - Y(t_n)|\ |\mathcal{A}_{t_n}) \\
&\quad + M \sum_{s_\ell \in \mathcal{S}_L(t_n)} \mathbb{E}\ (|X(s_\ell) - Y(s_\ell)|\ |\mathcal{A}_{t_n})\Big\} \\
&= h\ L_\phi\ \mathbb{E}\ \Big(|\,\mathrm{e}(t_{n+1})|\ +\ |\,\mathrm{e}(t_n)|\ +\ M \sum_{s_\ell \in \mathcal{S}_L(t_n)} |\,\mathrm{e}(s_\ell)|\ |\mathcal{A}_{t_n}\Big)\ .
\end{aligned}
$$

Inserting this estimate into the term labelled 2) in (20) yields

$$
\begin{aligned}
2 \mid \mathbb{E}\ \langle \mathrm{e}(t_n), \mathcal{U}(t_n)\rangle \mid\ &=\ 2 \mid \mathbb{E}\ \Big(\mathbb{E}\ (\langle \mathrm{e}(t_n), \mathcal{U}(t_n)\rangle | \mathcal{A}_{t_n})\ \Big)| \\
&\leq 2\ \ \mathbb{E}\ |\langle \mathrm{e}(t_n), |\ \mathbb{E}\ (\mathcal{U}(t_n)|\mathcal{A}_{t_n})\ |\ \rangle| \\
&\leq 2\, h\, L_\phi \Big(\mathbb{E}\, |\langle \mathrm{e}(t_n), \mathrm{e}(t_{n+1})\rangle\ | + \mathbb{E}\, |\langle \mathrm{e}(t_n), \mathrm{e}(t_n)\rangle| + M \sum_{s_\ell \in \mathcal{S}_L(t_n)} \mathbb{E}\, |\langle \mathrm{e}(t_n), \mathrm{e}(s_\ell)\rangle\ |\Big) \\
&\leq h\ L_\phi\ \mathbb{E}\ |\ \mathrm{e}(t_{n+1})|^2\ + h\ C_1\ (\mathbb{E}\ |\ \mathrm{e}(t_n)|^2 +\ M \sum_{s_\ell \in \mathcal{S}_L(t_n)} \mathbb{E}\ |\ \mathrm{e}(s_\ell)|^2)\ . \qquad (22)
\end{aligned}
$$

For the term labelled 4) we have, due to the properties (11) and (13) of the increment functions,

$$
\begin{aligned}
2\, \mathbb{E}\, |\mathcal{U}(t_n)|^2\ &\leq\ 4\, h^2\, \mathbb{E}\, |\phi(t_{n+1}, t_n, X^{F_h}_{t_{n+1}}, X^{F_h}_{t_n}) - \phi(t_{n+1}, t_n, \{Y^{F_h}_{t_{n+1}}\}, \{Y^{F_h}_{t_n}\})|^2 \\
&\quad + 4 \cdot 2^{r-1} \sum_{i=1}^{r} \mathbb{E}\ |\Big\{\phi_i(t_n, X^{G_h}_{t_n}) - \phi_i(t_n, \{Y^{G_h}_{t_n}\})\Big\}\ I^{t_n, t_{n+1}}_{\phi_i}|^2 \\
&\leq 4\, h^2\, L^2_\phi\, \mathbb{E}\, \Big(|X(t_{n+1}) - Y(t_{n+1})| + |X(t_n) - Y(t_n)| + M \sum_{s_\ell \in \mathcal{S}_L(t_n)} |X(s_\ell) - Y(s_\ell)|\Big)^2 \\
&\quad + 4 \cdot 2^{r-1} \sum_{i=1}^{r} h\, L_{\phi_i} \Big(\mathbb{E}\, |X(t_n) - Y(t_n)|^2 + M \sum_{s_\ell \in \mathcal{S}_L(t_n)} \mathbb{E}\, |X(s_\ell) - Y(s_\ell)|^2\Big) \\
&\leq 12\, h^2\, L^2_\phi\, \mathbb{E}\, |\ \mathrm{e}(t_{n+1})|^2\ + h\, C_1\, \mathbb{E}\, |\ \mathrm{e}(t_n)|^2\ +\ h\, C_1\ M \sum_{s_\ell \in \mathcal{S}_L(t_n)} \mathbb{E}\ |\ \mathrm{e}(s_\ell)|^2\ . \qquad (23)
\end{aligned}
$$

Combining these results, we obtain

$$
\begin{aligned}
\mathbb{E}\ |\ \mathrm{e}(t_{n+1})|^2 \leq (1\ +\ h\, C_1\,)\ \mathbb{E}\ |\ \mathrm{e}(t_n)|^2\ &+\ h\ (L_\phi + 12\ h\ L^2_\phi)\ \mathbb{E}\ |\ \mathrm{e}(t_{n+1})|^2 \\
&+\ h\ C_1\ M \sum_{s_\ell \in \mathcal{S}_L(t_n)} \mathbb{E}\ |\ \mathrm{e}(s_\ell)|^2 \\
&+ 2\ \mathbb{E}\ |\delta_h(t_n)|^2 + h^{-1}\ \mathbb{E}\ |\ \mathbb{E}(\ \delta_h(t_n)|\mathcal{A}_{t_n})|^2.
\end{aligned}
$$

Hence

$$
\begin{aligned}
&(1 - h\ (L_\phi + 12\ h\ L^2_\phi))\ \mathbb{E}\ |\ \mathrm{e}(t_{n+1})|^2 \\
&\qquad \leq (1\ +\ h\, C_1\,)\ \mathbb{E}\ |\ \mathrm{e}(t_n)|^2 + h\ C_1\ M\ L \max_{0 \leq i \leq n} \mathbb{E}\ |\ \mathrm{e}(s_i)|^2 \\
&\qquad + C\ h\ \Big(\underbrace{h^{-\frac{1}{2}} (\mathbb{E}\ |\delta_h(t_n)|^2)^{-\frac{1}{2}} + h^{-1}\ (\mathbb{E}\ |\ \mathbb{E}(\ \delta_h(t_n)|\mathcal{A}_{t_n})|^2)^{-\frac{1}{2}}}_{=: \gamma_h(t_n)}\Big)^2. \qquad (24)
\end{aligned}
$$

Set

$$
R_0 = 0, \qquad R_k = \max_{0 \leq i \leq k} \mathbb{E}|\ \mathrm{e}(t_i)|^2 \quad \text{ and } \quad \Gamma_k = \max_{0 \leq i \leq k} \gamma_h(t_i), \qquad (25)
$$

and by using $\dfrac{1 + h\ C_1}{1 - h\ C_2} = 1 + h\ C_3$ with $C_3 = \dfrac{C_1 + C_2}{1 - h\ C_2}$ and requiring $0 < h\ C_2 < 1$ for $C_2 = L_\phi + 12\ h\ L^2_\phi$, we can estimate from (24)

$$
R_{n+1} \leq (1 + h\ C_3)\ R_n + C\, h\, \Gamma^2_n.
$$

By iterating and observing that $h(n+1) = t_{n+1} \leq T$, we obtain

$$
\begin{aligned}
R_{n+1} &\leq (1+C_3 h)^{n+1} R_0 + C\,h\,\Gamma_n^2 \sum_{k=0}^{n} (1+C_3\,h)^k \\
&\leq \frac{C\,h\,\Gamma_n^2}{h}\,\frac{1}{C_3}((1+C_3\,h)^{n+1}-1) \\
&\leq \Gamma_n^2\,\frac{C}{C_3}(e^{C_3\,h(n+1)}-1) \;\leq \Gamma_n^2\,\frac{C\;(e^{C_3\,T}-1)}{C_3}\,.
\end{aligned}
$$

The assertion follows by taking the square-root.

Remark 1 The proof can be easily adapted to cover the error $\mathrm{e}(t_n+\zeta h)$ when one has to include the interpolant (10) at a point s_k in $\mathcal{S}_{\widetilde{N}}$. Further, errors in the starting values can be dealt with in the usual manner by considering a term $R_0 \neq 0$. Obviously, they need to be bounded in the $\|.\|_{L_2}$-norm.

4 Applications

In this section we show how our framework can be applied to several classes of SFDEs covered by Eq. (1), where we have chosen the scalar case for ease of exposition. We always assume that a suitable initial condition (2) is given and that the starting values are obtained from evaluating Ψ on $\mathcal{J}_h$. The Θ-Maruyama method will serve as an archetypical drift-implicit method in all examples.

Example 1 (SODEs)
We begin with the simplest case of an SFDE, which is the instantaneous SODE:

$$
\mathrm{d}X(t) = f(t,X(t))\,\mathrm{d}t + g(t,X(t))\,\mathrm{d}W(t), \quad t\in[0,T],\; X(0)=x_0.
$$

The Θ-Maruyama method reads

$$
\begin{aligned}
Y(t_n) + h\Big(\Theta f(t_{n+1},Y(t_{n+1})) + (1-\Theta) f(t_n,Y(t_n))\Big)& \\
+\, g(t_n,Y(t_n))\; I_{(1)}^{(t_n,t_{n+1})}&\,.
\end{aligned}
$$

Obviously, there is no interpolation of a memory term involved and the functions f and g do not need to be replaced by discrete functionals. The conditions (11) and (13) on the increment functions for some functions ξ, η and all $t, t' \in \mathcal{T}_h^N, t > t'$ reduce to

$$
\begin{aligned}
|\Theta(f(t,\xi(t)) - f(t,\eta(t))) + (1-\Theta)(f(t',\xi(t')) - f(t',\eta(t')))|& \\
\leq L_{\Theta,f}\,\Big(|\xi(t)-\eta(t)| + |\xi(t')-\eta(t')|\Big)&, \\
\text{and } \mathbb{E}|\{g(t',\xi(t')) \;-\; g(t',\eta(t'))\}\; I_{(1)}^{(t',t'+h)}|^2 \leq\; h\; L_g\;\mathbb{E}\,|\xi(t')-\eta(t')|^2&\,,
\end{aligned}
$$

with $M = L = 1$, which are satisfied if f and g are uniformly Lipschitz continuous. We refer e.g., to [1,11,17,24] for results equivalent to Theorem 1 as well as additonal material on the topic of SODEs.

Example 2 (SFDEs with variable delay)
We now consider

$$\begin{aligned} \mathrm{d}X(t) &= f(t, X(t), X(t-\tau_2(t)), \cdots, X(t-\tau_R(t)))\, \mathrm{d}t \\ &\quad + g(t, X(t), X(t-\varsigma_2(t)), \cdots, X(t-\varsigma_Q(t)))\, \mathrm{d}W(t), \qquad t \in [0,T]. \end{aligned}$$

We assume that the functions $\tau_\ell(t), \varsigma_\ell(t)$ are continuous for $t \in [0,T]$ and

$$0 < \tau_\star = \min\{\min_{2\le\ell\le R} \inf_{t\in[0,T]} \tau_\ell(t), \min_{2\le\ell\le Q} \inf_{t\in[0,T]} \varsigma_\ell(t)\}$$

and
$$\tau^\star = \max\{\max_{2\le\ell\le R} \sup_{t\in[0,T]} \tau_\ell(t), \max_{2\le\ell\le Q} \sup_{t\in[0,T]} \varsigma_\ell(t)\} < \infty.$$

One may introduce $\tau_1 = \varsigma_1 = 0$ if desired. The initial condition (2) is given on $[-\tau^\star, 0]$. We include SFDEs with discrete delay here, in particular when there are several non-commensurate delays. Here the Θ-Maruyama method obtains the form

$$\begin{aligned} Y(t_{n+1}) &= Y(t_n) + h(\Theta\, f(t_{n+1}, Y(t_{n+1}), Y(t_{n+1}-\tau_2(t_{n+1})), \cdots, Y(t_{n+1}-\tau_R(t_{n+1}))) \\ &\quad + (1-\Theta)\, f(t_n, Y(t_n), Y(t_n-\tau_2(t_n)), \cdots, Y(t_n-\tau_R(t_n)))) \\ &\qquad + g(t_n, Y(t_n), Y(t_n-\varsigma_2(t_n)), \cdots, Y(t_n-\varsigma_Q(t_n)))\, I_{(1)}^{(t_n,t_{n+1})}. \end{aligned}$$

We do not need a discretized version of f and g, however we may need an interpolant. This may be given by the method, see e.g. the approximation of the stochastic pantograph equation in [3] by continuous Θ-methods or by the continuous extension of the interpolation by piecewise constants in [15]. Then the conditions (11) and (13) on the increment functions can be stated as follows: for some functions ξ, η and all $t, t' \in \mathcal{T}_h^N, t > t'$

$$\begin{aligned} |\Theta\, & (f(t,\xi(t),\xi(t-\tau_2(t)),\cdots,\xi(t-\tau_R(t)) - f(t,\eta(t),\eta(t-\tau_2(t)),\cdots,\eta(t-\tau_R(t))) \\ & + (1-\Theta)\, (f(t',\xi(t'),\xi(t'-\tau_2(t')),\cdots,\xi(t'-\tau_R(t')) \\ & \qquad - f(t',\eta(t'),\eta(t'-\tau_2(t')),\cdots,\eta(t-\tau_R(t')))| \\ & \qquad \le C\, (\, |\xi(t)-\eta(t)| + |\xi(t')-\eta(t')| + \sum_{s_\ell \in S_{L'}(t')} |\xi(s_\ell)-\eta(s_\ell)|\,), \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}|\Big\{ & g(t',\xi(t'),\xi(t'-\varsigma_2(t')),\cdots,\xi(t'-\varsigma_Q(t'))) \\ & - g(t',\eta(t'),\eta(t'-\varsigma_2(t')),\cdots,\eta(t'-\varsigma_Q(t')))\Big\}\, I_{(1)}^{(t',t'+h)}|^2 \\ & \qquad \le C\, h\, (\mathbb{E}|\xi(t')-\eta(t')| + \sum_{s_\ell \in S_{L''}(t')} |\xi(s_\ell)-\eta(s_\ell)|^2), \end{aligned}$$

where $S_{L'}(t') = \{t - \tau_2(t), \dots, t - \tau_R(t), t' - \tau_2(t'), \dots, t' - \tau_R(t')\}$, $L' = 2\,R$ and $S_{L''}(t') = \{t' - \varsigma_2(t'), \dots, t' - \varsigma_Q(t')\}$, $L'' = Q$ (then L can be chosen as $L = \max(L', L'')$) and $M = 1$. In [3], where the continuously extended Θ-Maruyama method has been applied to the stochastic pantograph equation

$$\mathrm{d}X(t) = (aX(t) + bX(qt))\mathrm{d}t + (\sigma_1 + \sigma_2 X(t) + \sigma_3 X(qt))\,\mathrm{d}W(t)\,,\ (0 < q < 1),$$

the above properties were used in the proof of Theorem 3.3, providing mean-square convergence of the method.

Example 3 (SFDEs with distributed memory)
Our third example is given by

$$dX(t) = f(t, X(t), Z(t))\,\mathrm{d}t \ + \ g(t, X(t), Z(t))\,\mathrm{d}W(t), \qquad t \in [0, T],$$

where $Z(t)$ represents a memory term of the type

$$Z(t) \ = \int_{-\tau}^{0} K(t, s, X(t+s))\,\mathrm{d}s \ = \int_{t-\tau}^{t} K(t, s-t, X(s))\,\mathrm{d}s.$$

We choose the mesh $\mathcal{T}_h^N$ with a constant step-size $h = \tau/N_\tau$, $N_\tau \in \mathbb{N}$ and discretize the integral with the same step-size. Thus an interpolant is not required. However, we will need discretized functionals F^h and G^h. We formulate the Θ-Maruyama scheme as

$$\begin{aligned} Y(t_{n+1}) = Y(t_n) \ &+ h\,\Big(\Theta\ f(t_{n+1}, Y(t_{n+1}), \widetilde{Z}(t_{n+1})) \\ &+ (1-\Theta)\ f(t_n, Y(t_n), \widetilde{Z}(t_n))\Big) + g(t_n, Y(t_n), \widehat{Z}(t_n))\ I_{(1)}^{(t_n, t_{n+1})}\,. \end{aligned}$$

The expressions $\widetilde{Z}(t_m)$ and $\widehat{Z}(t_m)$ provide approximations to the integral $Z(t_m)$. Note that they may be different in the drift and diffusion terms, e. g. an implicit one in the drift and an explicit one in the diffusion, as indicated by the notation $\{Y_{t_n}^{F_h}\}$ and $\{Y_{t_n}^{G_h}\}$ in (9) . As in the deterministic case, the choice of a quadrature rule has consequences on the overall performance of the numerical method. We refer to the examples in [4, Section 3], where the second order convergence of the trapezium rule applied to a problem of the form given above with $g \equiv 0$ could only be achieved by also using the (composite) trapezium rule as the quadrature rule for Z. The composite Euler quadrature did not suffice. We thus may choose in the drift the composite trapezium rule, which has the form

$$\widetilde{Z}(t_m) = h \sum_{\ell=m-N_\tau}^{m}{}'' \, K(t_m, t_\ell - t_m, Y(t_\ell)), \qquad m \geq 0.$$

As usual, the notation $\sum''$ means that the first and the last term in the sum are to be halved. For the quadrature in the diffusion we choose an explicit scheme, the composite Euler method

$$\widehat{Z}(t_m) = h \sum_{\ell=m-N_\tau}^{m-1} K(t_m, t_\ell - t_m, Y(t_\ell)), \qquad m \geq 0.$$

These two choices are advantageous in the case of equations with small noise. We will pursue this topic in another article [6]. Note that it may turn out to be useful to consider stochastic quadrature methods given by

$$\bar{Z}(t_m) = \sum_{\ell=m-N_\tau}^{m-1} \Upsilon(t_m, t_\ell - t_m, Y(t_\ell))\, I_\Upsilon^{t_\ell, t_{\ell+1}}, \qquad m \geq 0,$$

where $I_\Upsilon^{t_\ell, t_{\ell+1}}$ represents a collection of appropriate multiple Wiener integrals. Stochastic quadrature methods for integrals with stochastic processes as integrands would arise in the same way as the numerical methods for SODEs have been developed: from the application of the Itô-formula to the integrand. We plan to carry out further investigations along this line.
The functionals F and G in (1) here have the form of f and g above and the discretized functionals F^h and G^h are now given by f and g with $Z(t)$ replaced by the quadrature formulas $\widetilde{Z}(t_m)$ and $\widehat{Z}(t_m)$. The conditions (11) and (13) on the increment functions now turn out to be in practice as follows: for some functions $\xi,\ \eta$ and all $t_m, t_{m-1} \in \mathcal{T}_h^N$

$$\begin{aligned}
|\Theta \Big(& f(t_m, \xi(t_m),\, h \sum_{\ell=m-N_\tau}^{m}{}'' K(t_m, t_\ell - t_m, \xi(t_\ell))) \\
& \qquad - f(t_m, \eta(t_m),\, h \sum_{\ell=m-N_\tau}^{m}{}'' K(t_m, t_\ell - t_m, \eta(t_\ell)))\Big) \\
& + (1-\Theta)\Big(f(t_{m-1}, \xi(t_{m-1}), h \sum_{\ell=m-1-N_\tau}^{m-1}{}'' K(t_{m-1}, t_\ell - t_{m-1}, \xi(t_\ell))) \\
& \qquad - f(t_{m-1}, \eta(t_{m-1}), h \sum_{\ell=m-1-N_\tau}^{m-1}{}'' K(t_{m-1}, t_\ell - t_{m-1}, \eta(t_\ell))\Big)| \\
& \leq C\ (\,|\xi(t_m) - \eta(t_m)| + |\xi(t_{m-1}) - \eta(t_{m-1})| + h \sum_{\ell=m-1-N_\tau}^{m-2} |\xi(t_\ell) - \eta(t_\ell)|\,),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}|\Big\{ & g(t_{m-1}, \xi(t_{m-1}), h \sum_{\ell=m-1-N_\tau}^{m-2} K(t_{m-1}, t_\ell - t_{m-1}, \xi(t_\ell))) \\
& - g(t_{m-1}, \eta(t_{m-1}), h \sum_{\ell=m-1-N_\tau}^{m-2} K(t_{m-1}, t_\ell - t_{m-1}, \eta(t_\ell)))\Big\}\ I_{(1)}^{(t_{m-1}, t_m)}|^2 \\
& \qquad \leq C\ h\ (\mathbb{E}|\xi(t_{m-1}) - \eta(t_{m-1})| + h \sum_{\ell=m-1-N_\tau}^{m-2} \mathbb{E}\,|\xi(s_\ell) - \eta(s_\ell)|^2,
\end{aligned}$$

where $L = N_\tau$ and $M = h$, thus their product is $M\,L = \tau$, which is clearly finite. In [6] the consistency of the Θ-Maruyama method is investigated, using an appropriate Itô-formula. The method has been implemented and some numerical experiments performed.

Acknowledgements
I thank C. T. H. Baker for fruitful discussions that were made possible in part by support from the London Mathematical Society.

References

[1] T. A. Averina and S. S. Artemiev, Numerical analysis of systems of ordinary and stochastic differential equations, VSP, Utrecht, 1997.

[2] C. T. H. Baker and E. Buckwar, *Numerical analysis of explicit one-step methods for stochastic delay differential equations,* LMS J. Comput. Math. 3, pp. 315–335, 2000.

[3] C. T. H. Baker and E. Buckwar, *Continuous* Θ*-methods for the stochastic pantograph equation*, Electron. Trans. Numer. Anal. 11, pp. 131–151, 2000.

[4] C. T. H. Baker and N. Ford, *Convergence of linear multistep methods for a class of delay-integro-differential equations*, in International Series of Numerical Mathematics 86, pp. 47–59, Birkhäuser Verlag Basel, 1988.

[5] A. Bellen and M. Zennaro, Numerical methods for delay differential equations, Oxford University Press, 2003.

[6] E. Buckwar, *The* Θ*-Maruyama scheme for stochastic functional differential equations with distributed memory term*, Preprint: `http://www.mathematik.hu-berlin.de/~buckwar/publications`.

[7] E. Buckwar and R. Winkler, *Multi-step methods for SDEs and their application to problems with small noise*, Preprint 2003-17, Humboldt University Berlin, submitted for publication, 2003.

[8] E. Hairer and S. P. Nørsett and G. Wanner, Solving ordinary differential equations. I: Nonstiff problems. 2. rev. ed., Springer-Verlag Berlin, 1993.

[9] J. Hale and S. M. Verduyn Lunel, Introduction to Functional Differential Equations, Springer Verlag Berlin, 1993.

[10] Y. Hu, S. E. A. Mohammed and F. Yan, *Discrete-time approximations of stochastic delay equations: the Milstein scheme*, to appear in The Annals of Probability.

[11] P. E. Kloeden and E. Platen, Numerical Solution of Stochastic Differential Equations, Springer-Verlag, Berlin, 1992.

[12] V. Kolmanovskii and A. Myshkis, Introduction to the Theory and Applications of Functional Differential Equations, Mathematics and its Applications, Kluwer Academic Publishers, Dordrecht, 1999.

[13] Y. Kuang, Delay Differential Equations with Applications in Population Dynamics, Academic Press, San Diego, 1993.

[14] X. Mao, Stochastic Differential Equations and their Applications, Horwood Publishing Limited, Chichester, 1997.

[15] X. Mao and S. Sabanis, *Numerical solutions of stochastic differential delay equations under local Lipschitz condition*, J. Comput. Appl. Math. 151(1), pp. 215-227, 2003.

[16] X. Mao, *Numerical solutions of stochastic functional differential equations*, LMS J. Comput. Math. 6, pp. 141-161, 2003.

[17] G. N. Milstein, Numerical Integration of Stochastic Differential Equations, Kluwer Academic Publishers Group, Dordrecht, 1995. Translated and revised from the 1988 Russian original.

[18] G.N. Milstein and M.V. Tretyakov, *Mean-square numerical methods for stochastic differential equations with small noise*, SIAM J.Sci. Comput. 18,pp. 1067–1087, 1997.

[19] S. E. A. Mohammed, *Stochastic differential systems with memory. Theory, examples and applications*, in Stochastic Analysis and Related Topics VI. The Geilo Workshop, 1996 , eds. L. Decreusefond, Jon Gjerde, B. Øksendal, A.S.Üstünel, Progress in Probability, Birkhauser ,pp. 1-77, 1998.

[20] S. E. A. Mohammed, Stochastic Functional Differential Equations, Pitman (Advanced Publishing Program), Boston, Mass., 1984.

[21] C. W. Cryer and L. Tavernini, *The numerical solution of Volterra functional differential equations by Euler's method*, SIAM J. numer. Anal. 9, pp. 105-129, 1972.

[22] L. Tavernini, *One-step methods for the numerical solution of Volterra functional differential equations*, SIAM J. numer. Anal. 8, pp. 786-795, 1971.

[23] D. Williams, Probability with martingales, Cambridge University Press, Cambridge, 1991.

[24] R. Winkler, *Stochastic differential algebraic equations of index 1 and applications in circuit simulation*, J. Comput. Appl. Math., 157 (2), pp. 477-505, 2003.