# MULTI-STEP METHODS FOR SDES AND THEIR APPLICATION TO PROBLEMS WITH SMALL NOISE *

EVELYN BUCKWAR AND RENATE WINKLER†

**Abstract.** In this paper the numerical approximation of solutions of Itô stochastic differential equations is considered, in particular for equations with a small parameter $\epsilon$ in the noise coefficient. We construct stochastic linear multi-step methods and develop the fundamental numerical analysis concerning their mean-square consistency, numerical stability in the mean-square sense and mean-square convergence. For the special case of two-step Maruyama schemes we derive conditions guaranteeing their mean-square consistency. Further, for the small noise case we obtain expansions of the local error in terms of the stepsize and the small parameter $\epsilon$. Simulation results using several explicit and implicit stochastic linear $k$-step schemes, $k = 1,\ 2$, illustrate the theoretical findings.

**AMS subject classifications.** 60H35, 65C30, 65L06, 60H10

**1. Introduction.** We consider stochastic differential equations (SDEs) of the form

$$X(s)\Big|_{t_0}^{t} = \int_{t_0}^{t} f(X(s), s)\,\mathrm{d}s + \int_{t_0}^{t} G(X(s), s)\,\mathrm{d}W(s), \quad X(t_0) = X_0, \tag{1.1}$$

for $t \in \mathcal{J}$, where $\mathcal{J} = [t_0, T]$. The drift and diffusion functions are given as $f : \mathbb{R}^n \times \mathcal{J} \to \mathbb{R}^n$, $G = (g_1, \ldots, g_m) : \mathbb{R}^n \times \mathcal{J} \to \mathbb{R}^{n\times m}$, respectively. The process $W$ is an $m$-dimensional Wiener process on the given probability space $(\Omega, \mathcal{F}, P)$ with a filtration $(\mathcal{F}_t)_{t\in\mathcal{J}}$, and $X_0$ is a given $\mathcal{F}_{t_0}$-measurable initial value, independent of the Wiener process and with finite second moment. We assume that there exists a path-wise unique strong solution $X(\cdot)$ of (1.1).

The aim of this article is to analyse the mean-square convergence properties of, in general, drift-implicit linear multi-step methods (LMMs) for the approximation of the solution of (1.1). An advantage LMMs have in deterministic numerics is that they require a low number of evaluations of the right-hand side. Thus for problems with an expensive right-hand side they are often preferable to Runge-Kutta schemes with the same order of convergence. We recall that in the deterministic case a high order of convergence is always based on sufficient smoothness of the solution of the differential equation. In contrast, the solution of an SDE is not smooth in the ordinary sense and a high order of convergence is achievable only by including more information on the driving Wiener process, i.e. a sufficient number of multiple stochastic integrals. Our motivation for considering stochastic multi-step methods (SLMMs) is firstly understanding when and why these schemes converge in the mean-square sense. Secondly, we want to recapture some of the properties of deterministic LMMs for SDEs with small noise, i.e. SDEs that can be written in the form

$$X(s)\Big|_{t_0}^{t} = \int_{t_0}^{t} f(X(s), s)ds + \int_{t_0}^{t} \epsilon\hat{G}(X(s), s)dW(s), \quad t \in \mathcal{J}, \quad X(t_0) = X_0, \tag{1.2}$$

*The work of the first author was supported by the DFG-Project BU 1073/2, the SFB 373 "Quantification and Simulation of Economic Processes" and the EU Network "Statistical Methods for Dynamical Stochastic Systems". The work of the second author was supported by the BMBF under the grant 03ROM3B3 and the DFG research center "Mathematics for key technologies" (FZT 86) in Berlin.

†Humboldt-Universität zu Berlin, Institut für Mathematik, (buckwar/winkler@mathematik.hu-berlin.de)

where $\epsilon \ll 1$ is a small parameter, and $\hat{G} = (\hat{g}_1, \dots, \hat{g}_m) : \mathbb{R}^n \times \mathcal{J} \to \mathbb{R}^{n\times m}$ and its derivatives are assumed to have moderate values. For this we will derive expansions of the error in terms of the small parameter $\epsilon$ and the step-size, as has been done for one-step methods in [17]. Then the small diffusion term makes it unnecessary to use high order multiple stochastic integrals and Wiener increments will be sufficient if the step-size is not too small.

Stochastic two-step methods already appear in [16] (for additive noise) and in [12] (see also the references there). In [2, 3] two-step methods for Itô SDEs are analysed. Stochastic versions of Adams methods for order up to five have been implemented and tested for SDEs with additive noise in [7]. Consistency of SLMMs for Stratonovich SDEs has been considered in [5], in addition stochastic Adams methods have been implemented as predictor-corrector schemes and tested.

In Section 2 we introduce the class of SLMMs considered and provide necessary definitions and useful facts. In Section 3 we consider the solvability of the discrete system and the boundedness of the iterates. We then establish two fundamental results, the first one concerns the relation between mean-square numerical stability of the SLMM and Dahlquist's root condition, and the second one concerns the relation between mean-square convergence, mean-square numerical stability and mean-square consistency of the SLMM. This mirrors the results and the essential role of zero-stability in the deterministic analysis of discretisation schemes, see e.g. [6, 9, 11, 13]. In Section 4 we consider two-step-Maruyama methods and give conditions for their mean-square consistency. These conditions allow to determine the parameters for the stochastic part from the parameters of the deterministic scheme and reduce to those of the underlying deterministic schemes when there is no noise. We then apply the two-step-Maruyama methods to the SDE with small noise (1.2) and derive expansions of the local error in a manner similar to that in [17]. In Section 5 we provide illustrative examples. The appendix contains the proof of Theorem 3.2.

**2. Definitions and preliminary results.** We denote by $|\cdot|$ the Euclidian norm in $\mathbb{R}^n$ and by $\|\cdot\|$ the corresponding induced matrix norm. The mean-square norm of a vector-valued square-integrable random variable $Z \in L_2(\Omega, \mathbb{R}^n)$ will be denoted by

$$\|Z\|_{L_2} := (\mathbb{E}|Z|^2)^{1/2} .$$

We define a deterministic grid on $\mathcal{J}$ as $t_0 < t_1 < \ldots < t_N = T$ with (for simplicity) a constant step-size $h := T/N$ and $t_\ell = \ell \cdot h,\ \ell = 0, \dots, N$. We consider a stochastic linear $k$-step method, which for $\ell = k, \dots, N$, takes the form

$$\sum_{j=0}^{k} \alpha_j\, X_{\ell-j} = h \sum_{j=0}^{k} \beta_j\, f(X_{\ell-j}, t_{\ell-j}) + \sum_{j=1}^{k} \Gamma_j(X_{\ell-j}, t_{\ell-j})\, I^{t_{\ell-j}, t_{\ell-j+1}}. \tag{2.1}$$

We set $\alpha_0 = 1$. We require given initial values $X_0, \dots, X_{k-1} \in L_2(\Omega, \mathbb{R}^n)$ such that $X_\ell$ is $\mathcal{F}_{t_\ell}$-measurable for $\ell = 0 \dots, k-1$. Every diffusion term $\Gamma_j(x,t)\, I^{t,t+h}$ is a finite sum of terms each containing an appropriate function $h$ of $x$ and $t$ multiplied by a multiple Wiener integral over $[t, t+h]$, i.e. it takes the general form

$$\Gamma_j(x,t)\, I^{t,t+h} = \sum_{r=1}^{m} h_j^r(x,t)\, I_r^{t,t+h} + \sum_{\substack{r_1,r_2=0 \\ r_1+r_2>0}}^{m} h_j^{r_1,r_2}(x,t)\, I_{r_1,r_2}^{t,t+h} + \dots\ .$$

A general multiple Wiener integral is given by

$$I^{t,t+h}_{r_1,r_2,\dots,r_j}(y) = \int_t^{t+h} \int_t^{s_1} \dots \int_{t+h}^{s_{j-1}} y(X(s_j),s_j)\;\mathrm{d}W_{r_1}(s_j)\dots\mathrm{d}W_{r_j}(s_1), \tag{2.2}$$

where $r_i \in \{0,1,\dots,m\}$ and $\mathrm{d}W_0(s) = \mathrm{d}s$. If $y \equiv 1$ we write $I^{t,t+h}_{r_1,r_2,\dots,r_j}$. Note that the integral $I^{t,t+h}_r$ is simply the increment $W_r(t+h) - W_r(t)$ of the scalar Wiener process $W_r$. The term $I^{t,t+h}$ denotes the collection of multiple Wiener integrals associated with the interval $[t,t+h]$.

We emphasize that an explicit discretization is used for the diffusion term. For $\beta_0 = 0$, the SLMM (2.1) is explicit, otherwise it is drift-implicit.

We give two examples of two-step methods.

Example 2.1. The first is a stochastic variant of the implicit two-step BDF method, which we term BDF2-Maruyama method. For $\ell = 2,\dots,N$, it takes the form

$$\begin{aligned} X_\ell - \frac{4}{3}X_{\ell-1} + \frac{1}{3}X_{\ell-2} = h\;\frac{2}{3}f(X_\ell,t_\ell) \\ + \sum_{r=1}^m g_r(X_{\ell-1},t_{\ell-1})\;I_r^{t_{\ell-1},t_\ell} - \frac{1}{3}\sum_{r=1}^m g_r(X_{\ell-2},t_{\ell-2})\;I_r^{t_{\ell-2},t_{\ell-1}}. \end{aligned}$$

Here one has $\alpha_0 = 1,\ \alpha_1 = -\frac{4}{3},\ \alpha_2 = \frac{1}{3},\ \beta_0 = \frac{2}{3},\ \beta_1 = \beta_2 = 0$, and

$$\Gamma_1(x,t)\;I^{t,t+h} = \sum_{r=1}^m g_r(x,t)\;I_r^{t,t+h}, \qquad \Gamma_2(x,t)\;I^{t,t+h} = -\frac{1}{3}\sum_{r=1}^m g_r(x,t)\;I_r^{t,t+h}.$$

Example 2.2. The second example is a Milstein variant of the two-step Adams-Bashforth-method:

$$\begin{aligned} X_\ell - X_{\ell-1} = h\;\left(\frac{3}{2}f(X_{\ell-1},t_{\ell-1}) - \frac{1}{2}f(X_{\ell-2},t_{\ell-2})\right) + \sum_{r=1}^m g_r(X_{\ell-1},t_{\ell-1})\;I_r^{t_{\ell-1},t_\ell} \\ + \sum_{r_1,r_2=1}^m (g_{r_1})'_x g_{r_2}(X(t_{\ell-1}),t_{\ell-1})\;I_{r_1,r_2}^{t_{\ell-1},t_\ell}. \end{aligned}$$

For this method one has

$$\alpha_0 = 1,\ \alpha_1 = -1,\ \alpha_2 = 0,\ \beta_0 = 0,\ \beta_1 = \frac{3}{2},\ \beta_2 = -\frac{1}{2}, \quad \Gamma_2(x,t)\;I^{t,t+h} \equiv 0$$

and

$$\Gamma_1(x,t)\;I^{t,t+h} = \sum_{r=1}^m g_r(x,t)\;I_r^{t,t+h} + \sum_{r_1,r_2=1}^m (g_{r_1})'_x g_{r_2}(x,t)\;I_{r_1,r_2}^{t,t+h}.$$

We will consider mean-square convergence of SLMMs in the sense discussed in Milstein and others [1, 16, 20]. Note that in the literature the term *strong* convergence is sometimes used synonymously for our expression *mean-square* convergence.

Definition 2.3. *We call the SLMM (2.1) for the approximation of the solution of the SDE (1.1)* **mean-square convergent** *if the global error $X(t_\ell) - X_\ell$ satisfies*

$$\max_{\ell=1,\dots,N} \|X(t_\ell) - X_\ell\|_{L_2} \to 0 \textit{ as } h \to 0,$$

*we say it is* **mean-square convergent with order** $\gamma$ *($\gamma > 0$) if the global error satisfies*

$$\max_{\ell=1,\dots,N} \|X(t_\ell) - X_\ell\|_{L_2} \le C \cdot h^\gamma,$$

*with a grid-independent constant* $C > 0$ *.*

In the following we will define what we understand by local errors. 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 in the fundamental work of Milstein in [15, 16], where for the first time the relation between local and global errors of one-step methods for SDEs has been clarified. However, only the first definition extends easily to multi-step methods, hence we will use it here. For comparison of these principles in the deterministic setting see [10, Chapters II.3,III.4].

Definition 2.4. *We define the* **local error** *of the SLMM (2.1) for the approximation of the solution of the SDE (1.1), for* $\ell = k, \dots, N$*, as*

$$L_\ell := \sum_{j=0}^{k} \alpha_j X(t_{\ell-j}) - h \sum_{j=0}^{k} \beta_j f(X(t_{\ell-j}), t_{\ell-j}) - \sum_{j=1}^{k} \Gamma_j(X(t_{\ell-j}), t_{\ell-j})\, I^{t_{\ell-j}, t_{\ell-j+1}}. \tag{2.3}$$

We aim to conclude mean square convergence from local properties of the SLMM by means of numerical stability in the mean-square sense. Numerical stability concerns the influence of perturbations of the right-hand side of the discrete scheme on the global solution of that discrete scheme. Sources of perturbations may be the local error, round-off errors or defects in the approximate solution of implicit schemes. The mean-square stability estimate of the global error is based on the mean square norm **and** on the conditional mean of the perturbations. In the case of one-step schemes this appears e.g. in [1, 20], we refer in particular to the discussion in [16, Chapter 1.4]. We remark that in the case of $k$-step schemes the conditional mean has to be taken with respect to the $\sigma$-algebra $\mathcal{F}_{t_{\ell-k}}$.

In our analysis we thus consider the following discrete system, the perturbed form of (2.1), for $\ell = k, \dots, N$

$$\sum_{j=0}^{k} \alpha_j \tilde{X}_{\ell-j} = h \sum_{j=0}^{k} \beta_j f(\tilde{X}_{\ell-j}, t_{\ell-j}) + \sum_{j=1}^{k} \Gamma_j(\tilde{X}_{\ell-j}, t_{\ell-j})\, I^{t_{\ell-j}, t_{\ell-j+1}} + D_\ell, \tag{2.4}$$

with initial values $\tilde{X}_\ell = X_\ell + D_\ell,\ \ell = 0, \dots, k-1$. We suppose that the perturbations $D_\ell$ are $\mathcal{F}_{t_\ell}$-measurable and that $D_\ell \in L_2(\Omega, \mathbb{R}^n)$.

Remark 2.5. *It is useful to represent the perturbations in the form*

$$\begin{array}{c} D_\ell = R_\ell + S_\ell =: R_\ell + \sum\limits_{j=1}^{k} S_{j,\ell-j+1}, \quad \ell = k, \dots, N, \\ \text{where each } S_{j,\ell} \text{ is } \mathcal{F}_{t_\ell} \text{ measurable with } \mathbb{E}(S_{j,\ell}|\mathcal{F}_{t_\ell}) = 0\ . \end{array} \tag{2.5}$$

*The representation (2.5) is not unique. One extreme possibility is* $R_\ell = D_\ell,\ S_\ell = 0$, *another, more useful one, is given by*

$$\begin{array}{rcl} R^*_\ell &=& \mathbb{E}(D_\ell|\mathcal{F}_{t_{\ell-k}}), \qquad S^*_\ell = D_\ell - R^*_\ell, \\ S^*_{j,\ell-j+1} &=& \mathbb{E}(D_\ell - R^*_\ell - \sum_{i=j+1}^k S^*_{i,\ell-i+1}|\mathcal{F}_{t_{\ell-j+1}}),\ j = k, k-1, \dots, 1\ . \end{array} \tag{2.6}$$

*This construction guarantees the required measurability conditions in (2.5). We also note that this decomposition is orthogonal in* $L_2(\Omega)$, *i.e.*

$$\|D_\ell\|^2_{L_2} = \|R^*_\ell\|^2_{L_2} + \sum_{j=1}^k \|S^*_{j,\ell-j+1}\|^2_{L_2}. \tag{2.7}$$

*As an example one obtains for* $k = 3$

$$\begin{aligned} R^*_\ell &= \mathbb{E}(D_\ell|\mathcal{F}_{t_{\ell-3}}) \\ S^*_{3,\ell-2} &= \mathbb{E}(D_\ell - R^*_\ell|\mathcal{F}_{t_{\ell-2}}) \\ S^*_{2,\ell-1} &= \mathbb{E}(D_\ell - R^*_\ell - S^*_{3,\ell-2}|\mathcal{F}_{t_{\ell-1}}) \\ S^*_{1,\ell} &= D_\ell - R^*_\ell - S^*_{3,\ell-2} - S^*_{2,\ell-1}. \end{aligned}$$

Now we give the precise definition of mean square stability and consistency that we consider in this paper.

Definition 2.6. *We call the SLMM (2.1)* ***numerically stable in the mean-square sense*** *if there exist constants* $h_0 > 0$ *and* $S > 0$ *such that for all stepsizes* $h < h_0$ *and for all* $\mathcal{F}_{t_\ell}$ *measurable perturbations* $D_\ell \in L_2(\Omega, \mathbb{R}^n)$ $(\ell = 0, \dots, N)$ *and all their representations (2.5), the following inequality holds*

$$\max_{\ell=0,\dots,N} \|X_\ell - \tilde{X}_\ell\|_{L_2} \le S\Big\{ \max_{\ell=0,\dots,k-1} \|D_\ell\|_{L_2} + \max_{\ell=k,\dots,N} \big( \frac{\|R_\ell\|_{L_2}}{h} + \frac{\|S_\ell\|_{L_2}}{h^{1/2}} \big) \Big\}, \tag{2.8}$$

*where* $(X_\ell)_{\ell=1}^N$ *and* $(\tilde{X}_\ell)_{\ell=1}^N$ *are the solutions of the SLMM (2.1) and the perturbed discrete system (2.4), respectively.*

We refer to $S$ as the stability constant and to (2.8) as the stability inequality.

Definition 2.7. *We call the SLMM (2.1) for the approximation of the solution of the SDE (1.1)* **mean-square consistent** *if the local error* $L_\ell$ *satisfies*

$$h^{-1}\ \|\mathbb{E}(L_\ell|\mathcal{F}_{t_{\ell-k}})\|_{L_2} \to 0 \text{ for } h \to 0, \text{ and } \ h^{-1/2}\ \|L_\ell\|_{L_2} \to 0 \text{ for } h \to 0.$$

*We call the SLMM (2.1) for the approximation of the solution of the SDE (1.1)* **mean-square consistent** *of order* $\gamma$ $(\gamma > 0)$, *if the local error* $L_\ell$ *satisfies*

$$\|\mathbb{E}(L_\ell|\mathcal{F}_{t_{\ell-k}})\|_{L_2} \le \bar{c} \cdot h^{\gamma+1}\ , \text{ and } \quad \|L_\ell\|_{L_2} \le c \cdot h^{\gamma+\frac{1}{2}}\ , \quad \ell = 1, \dots, N\ ,$$

*with constants* $c\,, \bar{c} > 0$ *only depending on the SDE and its solution.*

We remind the reader that consistency is only concerned with the local error. In the case that we disregard other sources of errors in (2.4) we only have to deal with perturbations $D_\ell = L_\ell$.

Lemma 2.8. *The SLMM (2.1) is mean-square consistent of order* $\gamma$ , *if*

$$\|R_\ell\|_{L_2} \le \bar{c} \cdot h^{\gamma+1}\ , \text{ and } \quad \|S_\ell\|_{L_2} \le c \cdot h^{\gamma+\frac{1}{2}}\ , \quad \ell = 1, \dots, N\ ,$$

*for any representation (2.5) of the local error $D_\ell = L_\ell$. The SLMM (2.1) is mean-square consistent of order $\gamma$, if and only if*

$$\|R^*_\ell\|_{L_2} \le \bar{c}\cdot h^{\gamma+1}\ , \ \text{and} \quad \|S^*_\ell\|_{L_2} \le c\cdot h^{\gamma+\frac{1}{2}}\ , \quad \ell = 1,\dots,N\ ,$$

*where the representation (2.6) is chosen for the local error $D_\ell = L_\ell$.*

*Proof.* First, let $L_\ell = R_\ell + S_\ell =: R_\ell + \sum_{j=1}^{k} S_{j,\ell-j+1}$ be a representation (2.5) with $\|R_\ell\|_{L_2} \le \bar{c}\cdot h^{\gamma+1}$ , and $\|S_\ell\|_{L_2} \le c\cdot h^{\gamma+\frac{1}{2}}$ , $\ell = k,\dots,N$ . By the conditions $\mathbb{E}(S_{j,\ell}|\mathcal{F}_{t_\ell}) = 0$ we conclude

$$\|\mathbb{E}(L_\ell|\mathcal{F}_{t_{\ell-k}})\|_{L_2} = \|\mathbb{E}(R_\ell|\mathcal{F}_{t_{\ell-k}})\|_{L_2} \le \|R_\ell\|_{L_2} \le \bar{c}\cdot h^{\gamma+1}.$$

Further, we have, for $h \le 1$,

$$\|L_\ell\|_{L_2} = \|R_\ell + S_\ell\|_{L_2} \le \|R_\ell\|_{L_2} + \|S_\ell\|_{L_2} \le \bar{c}\cdot h^{\gamma+1} + c\cdot h^{\gamma+\frac{1}{2}} \le (\bar{c}+c)\cdot h^{\gamma+\frac{1}{2}}.$$

Second, let the SLMM (2.1) be mean-square consistent of order $\gamma$, i.e. $\|\mathbb{E}(L_\ell|\mathcal{F}_{t_{\ell-k}})\|_{L_2} \le \bar{c}\cdot h^{\gamma+1}$ , and $\|L_\ell\|_{L_2} \le c\cdot h^{\gamma+\frac{1}{2}}$ , $\ell = k,\dots,N$. Because of $R^*_\ell = \mathbb{E}(L_\ell|\mathcal{F}_{t_{\ell-k}})$ we then, obviously, have $\|R^*_\ell\|_{L_2} \le \bar{c}\cdot h^{\gamma+1}$, and, further

$$\|S^*_\ell\|_{L_2} = \|L_\ell\|_{L_2} - \|R^*_\ell\|_{L_2} \le \|L_\ell\|_{L_2} \le c\cdot h^{\gamma+\frac{1}{2}}.$$

□

For further reference we state the following definitions and results.

Definition 2.9. *A function $f : \mathbb{R}^n \times \mathcal{J} \to \mathbb{R}^n$ satisfies a* **uniform Lipschitz condition** *with respect to $x$, if there exists a positive constant $L_f$, such that*

$$|f(x,t) - f(y,t)| \le L_f|x-y|, \quad \forall x,y \in \mathbb{R}^n, t \in \mathcal{J}. \tag{2.9}$$

*A function $\Gamma : \mathbb{R}^n \times \mathcal{J} \to \mathbb{R}^{n\times m_\Gamma}$ satisfies a* **uniform Lipschitz condition** *with respect to $x$, if there exists a positive constant $L_\Gamma$, such that*

$$\|\Gamma(x,t) - \Gamma(y,t)\| \le L_\Gamma|x-y| \quad \forall x,y \in \mathbb{R}^n, t \in \mathcal{J}. \tag{2.10}$$

*Let $C^{s,s-1}$ denote the class of all functions from $\mathbb{R}^n \times \mathcal{J}$ to $\mathbb{R}^n$ having continuous partial derivatives up to order $s-1$ and, in addition, continuous partial derivatives of order $s$ with respect to the first variable.*

*Let $C^K$ denote the class of functions $y$ from $\mathbb{R}^n \times \mathcal{J}$ to $\mathbb{R}^n$ that satisfy a* **linear growth condition** *in the form*

$$|y(x,t)| \le K(1+|x|^2)^{\frac{1}{2}}, \qquad \forall y \in \mathbb{R}^n, t \in \mathcal{J}. \tag{2.11}$$

Definition 2.10. *The* **characteristic polynomial** *of (2.1) is given by*

$$\rho(\zeta) = \alpha_k\zeta^k + \alpha_{k-1}\zeta^{k-1} + \ldots + \alpha_0. \tag{2.12}$$

*The SLMM (2.1) is said to fulfil* **Dahlquist's root condition**, *if*

*i) The roots of $\rho(\zeta)$ lie on or within the unit circle;*
*ii) The roots on the unit circle are simple.*

Lemma 2.11. *(A discrete version of Gronwall's lemma) Let $a_\ell$, $\ell = 1,\dots,N$, and $C_1$, $C_2$ be nonnegative real numbers and assume that the inequalities*

$$a_\ell \le C_1 + C_2\frac{1}{N}\sum_{i=1}^{\ell-1} a_i, \quad \ell = 1,\dots,N,$$

*are valid. Then we have* $\max_{\ell=1,\ldots,N} a_\ell \le C_1 \exp(C_2)$.

To estimate the multiple integrals (2.2) we will use the following lemma (cf. Lemma 2.1 and 2.2 in [16]).

LEMMA 2.12. *For any function $y$ belonging to the class $C^K$, and any $t \in \mathcal{J}$, $h > 0$, such that $t+h \in \mathcal{J}$, we have that*

$$\mathbb{E}(I^{t,t+h}_{r_1 \ldots r_j}(y)|\mathcal{F}_t) = 0 \quad \textit{if } r_i \neq 0 \textit{ for some } i \in \{1,\ldots,j\}, \tag{2.13}$$

$$\|\mathbb{E}(I^{t,t+h}_{r_1,\ldots,r_j}(y)|\mathcal{F}_t)\|_{L_2} \le \|I^{t,t+h}_{r_1,\ldots,r_j}(y)\|_{L_2} = \mathcal{O}(h^{l_1+l_2/2}), \tag{2.14}$$

*where $l_1$ is the number of zero indices $r_i$ and $l_2$ the number of non-zero indices $r_i$.*

**3. Global properties of stochastic LMMs.** In this section we will first establish the solvability of the recurrence equations (2.4) (and thus of (2.1)), then we will discuss numerical stability and mean-square convergence of the SLMM (2.1). The former characterizes the robustness of a numerical scheme with respect to small perturbations, such as rounding errors. As a property of the numerical scheme alone, it is not a priori giving evidence on the approximation power of the scheme (which may very well approximate a different problem than intended). However, numerical stability and consistency together yield convergence of the numerical solution to the exact solution. In order to distinguish this stability concept from others, it is sometimes called zero-stability or in honour of Dahlquist, also D-stability. It should not be mistaken for properties like asymptotic stability, which guarantee that for fixed step-sizes (and long or unbounded time-intervals) qualitative properties of the exact solutions like damping behaviour in dissipative systems are preserved by the discrete approximations. For further discussions we refer the reader to the deterministic literature (see e.g. [6, 13, 14]). In the stochastic literature mean-square numerical stability for one-step-schemes has been considered in [1, 3, 4, 8, 19, 20]. In the first four of these works only perturbations in the initial data have been treated.

We now turn to the solvability of the recurrence equations. If in (2.1) and (2.4) the parameter $\beta_0 = 0$, the discrete systems are explicit and every iterate $\tilde{X}_\ell$, $\ell \ge k$, can be obtained explicitly, i.e. the recurrence equations (2.1) and (2.4) obviously have unique solutions. In the case of implicit systems we need to consider the solvability of the systems of nonlinear equations (2.1) and (2.4). 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.)

THEOREM 3.1. *Suppose that $\beta_0 \neq 0$ and the drift-coefficient $f$ satisfies (2.9) and assume that $2\, h\, \beta_0\, L_f < 1$. Then the perturbed discrete scheme (2.4) and, in consequence, the SLMM (2.1) have a unique solution. If, in addition the coefficients $\Gamma_j$ satisfy (2.10), then the mean-square norm of the iterates exists.*

*Proof.* The proof of the existence of unique solutions of the perturbed discrete system (2.4) (and thus of (2.1)) follows the line of proofs used in the deterministic analysis of multi-step schemes. The idea is to express (2.4) as

$$\tilde{X}_\ell = h\beta_0 f(\tilde{X}_\ell, t_\ell) + \tilde{B}_\ell\ , \tag{3.1}$$

where

$$\tilde{B}_\ell := -\sum_{j=1}^{k} \alpha_j \tilde{X}_{\ell-j} + h\sum_{j=1}^{k} \beta_j f(\tilde{X}_{\ell-j}, t_{\ell-j}) + \sum_{j=1}^{k} \Gamma_j(X_{\ell-j}, t_{\ell-j})\, I^{t_{\ell-j},t_{\ell-j+1}} + \tilde{D}_\ell$$

is a known $\mathcal{F}_{t_\ell}$ -measurable random variable, when we suppose that $\tilde{X}_{\ell-j}$ are known $\mathcal{F}_{t_{\ell-j}}$-measurable random variables for $j = 1, \ldots, k$. We can then view (2.4) as a fixed-point equation in $x$

$$x = h\beta_0 f(x, t_\ell) + b_\ell$$

and apply the contraction mapping principle. We refer e.g. to [9, Thm.6.1.1] for more details.
It remains to show that the second moments of the iterates exist. We start from the assumption $\mathbb{E}|\tilde{X}_{\ell-j}|^2 < \infty$ , $j = 1, \ldots, k$ on the initial values. Recursively, we conclude that $\mathbb{E}|\tilde{X}_\ell|^2 < \infty$ , $\ell = k, \ldots, N$ by comparing $\tilde{X}_\ell$ with the solution of the fixed-point equation (3.1) for $\tilde{B}_\ell = 0$, i.e., with the solution of the deterministic implicit equation $x = h\beta_0 f(x, t_\ell)$, and applying Lipschitz continuity arguments. For more details we refer to [20, Thm.5]. ◻

We now formulate our main theorem on numerical stability. The proof is given in the appendix.
THEOREM 3.2. *The stochastic linear multi-step method (2.1) is numerically stable in the mean-square sense for every continuous $f$ and $\Gamma_j$ satisfying (2.9) and (2.10), respectively, if and only if its characteristic polynomial $\rho(\zeta)$ (2.12) satisfies Dahlquist's root condition given in Definition 2.10.*

With the powerful notion of numerical stability in the mean-square sense, together with mean-square consistency the mean-square convergence follows almost immediately.
THEOREM 3.3. *A mean-square consistent SLMM (2.1) for the approximation of the solution of SDE (1.1) is mean-square convergent for all continuous $f$ and $\Gamma_j$ satisfying (2.9) and (2.10), respectively, if and only if it is numerically stable in the mean-square sense. If, in addition, it is mean-square consistent with order $\gamma > 0$, then the SLMM (2.1) is mean-square convergent with order $\gamma$.*
*Proof.* First, let us assume that the mean-square consistent numerical method (2.1) is mean-square convergent. Then the necessity of stability can essentially be proved as in the deterministic case. Set $f \equiv 0$, $\Gamma_j \equiv 0$, $X_0 = 0$. Then (2.1) reduces to $\sum_{j=1}^{k} \alpha_j X_{\ell-j} = 0$, $l = k, k+1, \ldots$, a deterministic homogeneous difference equation, and stability follows by standard arguments, see e.g. [9].
Second, let us assume that the numerical method (2.1) is mean-square stable and consistent with order $\gamma > 0$. Then mean square convergence with order $\gamma$ follows by applying the stability estimate (2.8) to $\{\tilde{X}_\ell := X(t_\ell)\}$ related to the perturbations $\{D_\ell := L_\ell = R^*_\ell + S^*_\ell\}$. ◻

**4. Two-step-Maruyama schemes.** In this section we consider linear two-step-Maruyama schemes, thus we have for $\ell = 2, \ldots, N$

$$\sum_{j=0}^{2} \alpha_j X_{\ell-j} = h \sum_{j=0}^{2} \beta_j f(X_{\ell-j}, t_{\ell-j}) + \sum_{j=1}^{2} \gamma_j \sum_{r=1}^{m} g_r(X_{\ell-j}, t_{\ell-j})\; I_r^{t_{\ell-j}, t_{\ell-j+1}}. \quad (4.1)$$

For sufficiently smooth drift and diffusion coefficients $f,\ g_1, \ldots, g_m$ Theorem 3.2 applies and the two-step scheme (4.1) is mean square stable if the coefficients $\alpha_0,\ \alpha_1,\ \alpha_2$ satisfy Dahlquist's root condition. Then the scheme (4.1) is mean-square convergent of some order $\gamma$, if it is mean-square consistent of that order. Thus we will be concerned with mean-square consistency of the above scheme and derive order conditions in terms of the coefficients $\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \gamma_1, \gamma_2$. In general, the mean-square order of convergence will be not higher than $\frac{1}{2}$, since the only information about the driving noise process that the Maruyama-type schemes include are the Wiener increments. We note that the simple Euler-Maruyama method would suffice to obtain the same order of convergence. However, convergence is an asymptotic property, i.e. it holds for $h \to 0$ and a result concerning the order of convergence may not provide sufficient information about the size of the actual error that arise for reasonable choices of the step-size. In particular when one considers equations with a small noise term as in Equation (1.2), one may find that the influence of the noise is not dominant and properties of the methods in the deterministic setting are recovered to a certain extent.

From the deterministic theory we know that for linear $k$-step methods

$$\sum_{j=0}^{k} \alpha_j x_{\ell+j} = h \sum_{j=0}^{k} \beta_j f(x_{\ell+j}, t_{\ell+j}), \quad \text{applied to } x'(t) = f(x(t), t),$$

the local error is of order $p+1$ for sufficiently smooth functions $f$ if

$$\sum_{j=0}^{k} \alpha_j = 0 \qquad \text{and} \qquad \sum_{j=0}^{k} \alpha_j j^q = q \sum_{j=0}^{k} \beta_j j^{q-1} \quad \text{for } q = 1, \ldots, p.$$

In the first part of this section we derive consistency conditions for the two-step scheme (4.1) applied to the general SDE (1.1). We establish a representation of the local error $L_\ell$ in terms of certain multiple stochastic integrals obtained by the Itô-Taylor expansion. It turns out that consistency is guaranteed under the above conditions for deterministic order 1 and additional conditions that determine the method parameters $\gamma_1$ and $\gamma_2$.

In the second part of this section we consider the application of the scheme (4.1) to SDEs with small noise, i.e. SDEs that can be written in the form (1.2) with a small parameter $\epsilon$ in the diffusion coefficient $G(x,t) = \epsilon \hat{G}(x,t)$. We follow ideas of [17] and develop the local error in powers of the stepsize $h$ and the small parameter $\epsilon$. The expansion yields the deterministic conditions for order two and we discuss for which choices of $\epsilon$ and $h$ the stochastic component in the error estimates becomes small compared to these order 2 terms.

**4.1. Two-step schemes for general SDEs.** To analyse the local error $L_\ell$ of the scheme (4.1) for the SDE (1.1) and to achieve a suitable representation (2.5) we want to derive appropriate Itô-Taylor expansions, where we take special care to separate the multiple stochastic integrals over the different subintervals of integration. We introduce operators $\Lambda_0$ and $\Lambda_r$, $r = 1, \ldots, m$, defined on $C^{2,1}$ and $C^{1,0}$, respectively, by

$$\Lambda_0 y = y'_t + y'_x f + \frac{1}{2} \sum_{r=1}^{m} \sum_{i,j=1}^{n} y''_{x_i x_j} g_{ri} g_{rj} \ , \quad \Lambda_r y = y'_x g_r \ , \ r = 1, \ldots, m, \tag{4.2}$$

and remind the reader of the notation for multiple Wiener integrals (2.2). Using these operators the Itô formula for a function $y$ in $C^{2,1}$ and the solution $X$ of (1.1) reads

$$y(X(t), t) = y(X(t_0), t_0) + I_0^{t_0,t}(\Lambda_0 y) + \sum_{r=1}^{m} I_r^{t_0,t}(\Lambda_r y), \quad t \in \mathcal{J}. \tag{4.3}$$

Applying the Itô-formula (4.3) on the corresponding intervals to the drift coefficient $f$ as well as to the diffusion coefficients $g_r$ yields for $s \in [t_{\ell-j}, t_{\ell-j+1}]$, $j = 1, 2$

$$f(X(s), s) = f(X(t_{\ell-j}), t_{\ell-j}) + I_0^{t_{\ell-j},s}(\Lambda_0 f) + \sum_{r=1}^{m} I_r^{t_{\ell-j},s}(\Lambda_r f), \tag{4.4}$$

$$g_r(X(s), s) = g_r(X(t_{\ell-j}), t_{\ell-j}) + I_0^{t_{\ell-j},s}(\Lambda_0 g_r) + \sum_{q=1}^{m} I_q^{t_{\ell-j},s}(\Lambda_q g_r). \tag{4.5}$$

We trace back the values of the drift coefficient to the point $t_{\ell-2}$ to obtain

$$f(X(t_{\ell-1}), t_{\ell-1}) = f(X(t_{\ell-2}), t_{\ell-2}) + I_0^{t_{\ell-2},t_{\ell-1}}(\Lambda_0 f) + \sum_{r=1}^{m} I_r^{t_{\ell-2},t_{\ell-1}}(\Lambda_r f), \tag{4.6}$$

$$\begin{aligned} f(X(t_\ell), t_\ell) &= f(X(t_{\ell-2}), t_{\ell-2}) + I_0^{t_{\ell-2},t_{\ell-1}}(\Lambda_0 f) + I_0^{t_{\ell-1},t_\ell}(\Lambda_0 f) \\ &\quad + \sum_{r=1}^{m} I_r^{t_{\ell-2},t_{\ell-1}}(\Lambda_r f) + \sum_{r=1}^{m} I_r^{t_{\ell-1},t_\ell}(\Lambda_r f). \end{aligned} \tag{4.7}$$

For the general SDE (1.1) we have the following result.

LEMMA 4.1. *Assume that the coefficients $f$, $g_r, r = 1, \ldots, m$ of the SDE (1.1) belong to the class $C^{2,1}$ with $\Lambda_0 f, \Lambda_0 g_r, \Lambda_r f, \Lambda_q g_r \in C^K$ for $r, q = 1, \ldots, m$. Then the local error (2.3) of the stochastic 2-step scheme (4.1) allows the representation*

$$L_\ell = R_\ell^\circ + S_{1,\ell}^\circ + S_{2,\ell-1}^\circ, \quad \ell = 2, \ldots, N, \tag{4.8}$$

*where $R_\ell^\circ, S_{j,\ell}^\circ, j = 1, 2$ are $\mathcal{F}_{t_\ell}$-measurable with $\mathbb{E}(S_{j,\ell}^\circ | \mathcal{F}_{t_{\ell-1}}) = 0$ and*

$$\begin{aligned} R_\ell^\circ &= \Big[\sum_{j=0}^{2} \alpha_j\Big] X(t_{\ell-2}) + \Big[2\alpha_0 + \alpha_1 - \sum_{j=0}^{2} \beta_j\Big] h f(X(t_{\ell-2}), t_{\ell-2}) + \tilde{R}_\ell^\circ, \\ S_{1,\ell}^\circ &= \Big[\alpha_0 - \gamma_1\Big] \sum_{r=1}^{m} g_r(X(t_{\ell-1}), t_{\ell-1}) I_r^{t_{\ell-1},t_\ell} + \tilde{S}_{1,\ell}^\circ, \\ S_{2,\ell-1}^\circ &= \Big[(\alpha_0 + \alpha_1) - \gamma_2\Big] \sum_{r=1}^{m} g_r(X(t_{\ell-2}), t_{\ell-2}) I_r^{t_{\ell-2},t_{\ell-1}} + \tilde{S}_{2,\ell-1}^\circ \end{aligned}$$

*with*

$$\|\tilde{R}_\ell^\circ\|_{L_2} = O(h^2), \quad \|\tilde{S}_{1,\ell}^\circ\|_{L_2} = O(h), \quad \|\tilde{S}_{2,\ell-1}^\circ\|_{L_2} = O(h). \tag{4.9}$$

COROLLARY 4.2. *Let the coefficients $f$, $g_r$, $r = 1, \ldots, m$, of the SDE (1.1) satisfy the assumptions of Lemma 4.1 and suppose they are Lipschitz continuous with respect to their first variable. Let the coefficients of the stochastic linear two-step scheme (4.1) satisfy Dahlquist's root condition and the consistency conditions*

$$\sum_{j=0}^{2} \alpha_j = 0, \quad 2\alpha_0 + \alpha_1 = \sum_{j=0}^{2} \beta_j, \quad \alpha_0 = \gamma_1, \quad \alpha_0 + \alpha_1 = \gamma_2. \tag{4.10}$$

*Then the global error of the scheme (4.1) applied to (1.1) allows the expansion*

$$\max_{\ell=2,N} \|X(t_\ell) - X_\ell\|_{L_2} = \mathcal{O}(h^{1/2}) + \mathcal{O}(\max_{\ell=0,1} \|X(t_\ell) - X_\ell\|_{L_2}) \,.$$

*Proof. (of Corollary 4.2)* By Lemma 4.1 we have the representation (4.8) for the local error. Applying the consistency conditions (4.10) yields

$$R^\circ_\ell = \tilde{R}^\circ_\ell, \quad S^\circ_{1,\ell} = \tilde{S}^\circ_{1,\ell}, \quad S^\circ_{2,\ell-1} = \tilde{S}^\circ_{2,\ell-1}, \quad \ell = 2, \ldots, N.$$

As the scheme (4.1) satisfies Dahlquist's root condition, it is numerically stable in the mean-square sense. Now the assertion follows from the estimates (4.9) by means of the stability inequality. ◻

*Proof. (of Lemma 4.1)* To derive a representation of the local error in the form (4.8) we evaluate and resume the deterministic parts at the point $(X(t_{\ell-2}), t_{\ell-2})$ and separate the stochastic terms carefully over the different subintervals $[t_{\ell-2}, t_{\ell-1}]$ and $[t_{\ell-1}, t_\ell]$. This ensures the independence of the random variables. It does make the calculations more messy, though. By rewriting

$$\sum_{j=0}^{2} \alpha_j X(t_{\ell-j}) = \alpha_0\big(X(t_\ell) - X(t_{\ell-1})\big) + (\alpha_0 + \alpha_1)\big(X(t_{\ell-1}) - X(t_{\ell-2})\big) + \big(\sum_{j=0}^{2} \alpha_j\big) X(t_{\ell-2}),$$

we can express the local error (2.3) as

$$\begin{aligned} L_\ell = \alpha_0\big(X(t_\ell) - X(t_{\ell-1})\big) + (\alpha_0 + \alpha_1)\big(X(t_{\ell-1}) - X(t_{\ell-2})\big) + \sum_{j=0}^{2} \alpha_j X(t_{\ell-2}) \\ -h \sum_{j=0}^{2} \beta_j f(X(t_{\ell-j}), t_{\ell-j}) - \sum_{j=1}^{2} \gamma_j G(X(t_{\ell-j}), t_{\ell-j}) \Delta W_{\ell-j+1}. \end{aligned}$$

The SDE (1.1) implies the identities

$$\begin{aligned} &X(t_{\ell-1}) - X(t_{\ell-2}) = \int_{t_{\ell-2}}^{t_{\ell-1}} f(X(s), s) \mathrm{d}s + \sum_{r=1}^{m} \int_{t_{\ell-2}}^{t_{\ell-1}} g_r(X(s), s) \mathrm{d}W_r(s) \\ &= h f(X(t_{\ell-2}), t_{\ell-2}) + I_{00}^{t_{\ell-2}, t_{\ell-1}}(\Lambda_0 f) + \sum_{r=1}^{m} I_{r0}^{t_{\ell-2} t_{\ell-1}}(\Lambda_r f) \\ &\quad + \sum_{r=1}^{m} g_r(X(t_{\ell-2}), t_{\ell-2}) I_r^{t_{\ell-2}, t_{\ell-1}} + \sum_{r=1}^{m} I_{0r}^{t_{\ell-2}, t_{\ell-1}}(\Lambda_0 g_r) + \sum_{r,q=1}^{m} I_{qr}^{t_{\ell-2}, t_{\ell-1}}(\Lambda_q g_r), \end{aligned}$$

and, additionally using (4.6),

$$\begin{aligned} X(t_\ell) - X(t_{\ell-1}) &= \int_{t_{\ell-1}}^{t_\ell} f(X(s), s) \mathrm{d}s + \sum_{r=1}^{m} \int_{t_{\ell-1}}^{t_\ell} g_r(X(s), s) \mathrm{d}W_r(s) \\ &= h\big\{f(X(t_{\ell-2}), t_{\ell-2}) + I_0^{t_{\ell-2}, t_{\ell-1}}(\Lambda_0 f) + \sum_{r=1}^{m} I_r^{t_{\ell-2}, t_{\ell-1}}(\Lambda_r f)\big\} \\ &\qquad + I_{00}^{t_{\ell-1}, t_\ell}(\Lambda_0 f) + \sum_{r=1}^{m} I_{r0}^{t_{\ell-1} t_\ell}(\Lambda_r f) \\ &\quad + \sum_{r=1}^{m} g_r(X(t_{\ell-1}), t_{\ell-1}) I_r^{t_{\ell-1}, t_\ell} + \sum_{r=1}^{m} I_{0r}^{t_{\ell-1}, t_\ell}(\Lambda_0 g_r) + \sum_{r,q=1}^{m} I_{qr}^{t_{\ell-1}, t_\ell}(\Lambda_q g_r). \end{aligned}$$

Inserting this and the expansions (4.6, 4.7) into the local error formula and reordering the terms, yields

$$\begin{aligned}L_\ell = &\Big[\sum_{j=0}^{2}\alpha_j\Big]X(t_{\ell-2}) + \Big[2\alpha_0+\alpha_1-\sum_{j=0}^{2}\beta_j\Big]hf(X(t_{\ell-2}),t_{\ell-2}) + \tilde{R}^\circ_\ell\\ &+\Big[\alpha_0-\gamma_1\Big]\sum_{r=1}^{m} g_r(X(t_{\ell-1}),t_{\ell-1})I_r^{t_{\ell-1},t_\ell} + \tilde{S}^\circ_{1,\ell}\\ &+\Big[(\alpha_0+\alpha_1)-\gamma_2\Big]\sum_{r=1}^{m} g_r(X(t_{\ell-2}),t_{\ell-2})I_r^{t_{\ell-2},t_{\ell-1}} + \tilde{S}^\circ_{2,\ell-1},\end{aligned}$$

where

$$\begin{aligned}\tilde{R}^\circ_\ell = &\alpha_0\big\{hI_0^{t_{\ell-2},t_{\ell-1}}(\Lambda_0 f) + I_{00}^{t_{\ell-1},t_\ell}(\Lambda_0 f)\big\} + (\alpha_0+\alpha_1)I_{00}^{t_{\ell-2},t_{\ell-1}}(\Lambda_0 f)\\ &-h\beta_0\big\{I_0^{t_{\ell-2},t_{\ell-1}}(\Lambda_0 f) + I_0^{t_{\ell-1},t_\ell}(\Lambda_0 f)\big\} - h\beta_1 I_0^{t_{\ell-2},t_{\ell-1}}(\Lambda_0 f),\end{aligned} \tag{4.11}$$

$$\begin{aligned}\tilde{S}^\circ_{1,\ell} = &\sum_{r=1}^{m}\Big(\alpha_0 I_{r0}^{t_{\ell-1},t_\ell}(\Lambda_r f) - h\beta_0 I_r^{t_{\ell-1},t_\ell}(\Lambda_r f)\Big) + \alpha_0\sum_{r=1}^{m} I_{0r}^{t_{\ell-1},t_\ell}(\Lambda_0 g_r)\\ &+\alpha_0\sum_{r,q=1}^{m} I_{qr}^{t_{\ell-1},t_\ell}(\Lambda_q g_r)),\end{aligned} \tag{4.12}$$

$$\begin{aligned}\tilde{S}^\circ_{2,\ell-1} = &h(\alpha_0-\beta_0-\beta_1)\sum_{r=1}^{m} I_r^{t_{\ell-2},t_{\ell-1}}(\Lambda_r f) + (\alpha_0+\alpha_1)\sum_{r=1}^{m} I_{r0}^{t_{\ell-2},t_{\ell-1}}(\Lambda_r f)\\ &+(\alpha_0+\alpha_1)\sum_{r=1}^{m} I_{0r}^{t_{\ell-2},t_{\ell-1}}(\Lambda_0 g_r) + (\alpha_0+\alpha_1)\sum_{r,q=1}^{m} I_{qr}^{t_{\ell-2},t_{\ell-1}}(\Lambda_q g_r).\end{aligned} \tag{4.13}$$

Finally, the estimates (4.9) are derived by means of Lemma 2.12, where the last terms in (4.12) and (4.13) determine the order $\mathcal{O}(h)$. ∎

**4.2. Two-step schemes for small noise SDEs.** For the numerical integration of ODEs two-step schemes of order 2 or higher are particularly interesting. They offer a high order of convergence for low computational cost per step. In this section we consider the special case of SDEs of the form (1.2) where we have a small parameter $\epsilon$ in the diffusion coefficients $g_r = \epsilon\hat{g}_r,\ r = 1,\dots,m$. Lemma 4.1 provides a representation for the local error (2.3) of the stochastic linear two step scheme (4.1) applied to (1.2). Starting from this expression we will further analyse the local error by expanding the term $\Lambda_0 f$ appearing in $\tilde{R}^\circ_\ell$ (4.11). Naturally, this requires more smoothness of the coefficients. A sufficient condition would be $\Lambda_0 f \in C^{2,1}$, for which, in general, one needs the existence of fourth order derivatives of $f$ with respect to $x$. However, for small noise SDES, the term $f'_x f + f'_t$ dominates $\Lambda_0 f$. This allows to weaken the smoothness assumptions again. The expansion of $\Lambda_0 f$ also yields additional multiple Itô-integrals whose conditional expectation vanishes. By moving these terms from $\tilde{R}^\circ_\ell$ into the stochastic parts of the representation of the local error we achieve better estimates. With this analysis we are able to prove that some of the potential of deterministic two-step schemes can be recovered in the special case of small noise SDEs.

To be able to exploit the effect of the small parameter $\epsilon$ in the expansions of the local error we introduce operators $\Lambda_0^f, \hat{\Lambda}_0$ and $\hat{\Lambda}_r,\ r = 1, \ldots, m$ defined on $C^{2,1}$ and $C^{1,0}$, respectively, by

$$\Lambda_0^f y := y_t' + y_x' f, \quad \hat{\Lambda}_0 y := \frac{1}{2}\sum_{r=1}^{m}\sum_{i,j=1}^{n} y''_{x_i x_j}\hat{g}_{ri}\hat{g}_{rj}, \quad \hat{\Lambda}_r y := y_x' \hat{g}_r \,. \tag{4.14}$$

In terms of the original definition (4.2) we have

$$\Lambda_0 y = \Lambda_0^f y + \epsilon^2 \hat{\Lambda}_0 y \quad \text{ and } \quad \Lambda_r y = \epsilon \hat{\Lambda}_r y. \tag{4.15}$$

LEMMA 4.3. *Assume that the coefficients $f, \hat{g}_r, r = 1, \ldots, m$ of the small noise SDE (1.2), as well as $\Lambda_0^f f = f_x' f + f_t'$ belong to the class $C^{2,1}$ with $\Lambda_0 f, \Lambda_0 \hat{g}_r, \hat{\Lambda}_r f$, $\hat{\Lambda}_q \hat{g}_r, \Lambda_0 \Lambda_0^f f$, $\hat{\Lambda}_r \Lambda_0^f f \in C^K$ for $r, q = 1, \ldots, m$. Let the stochastic 2-step scheme (4.1) satisfy the consistency conditions (4.10). Then the local error (2.3) of the method (4.1) for the small noise SDE (1.2) allows the representation*

$$L_\ell = R_\ell^\diamond + S_{1,\ell}^\diamond + S_{2,\ell-1}^\diamond, \quad \ell = 2, \ldots, N, \tag{4.16}$$

*where $R_\ell^\diamond, S_{j,\ell}^\diamond, j = 1, 2$ are $\mathcal{F}_{t_\ell}$-measurable with $\mathbb{E}(S_{j,\ell}^\diamond | \mathcal{F}_{t_{\ell-1}}) = 0$, and*

$$\begin{aligned}
R_\ell^\diamond &= \Big[(4\alpha_0 + \alpha_1) - (4\beta_0 + 2\beta_1)\Big]\frac{h^2}{2}(f_t' + f_x' f)(X(t_{\ell-2}), t_{\ell-2}) + \tilde{R}_\ell^\diamond, \\
S_{1,\ell}^\diamond &= \tilde{S}_{1,\ell}^\circ + \tilde{S}_{1,\ell}^\diamond, \\
S_{2,\ell-1}^\diamond &= \tilde{S}_{2,\ell-1}^\circ + \tilde{S}_{2,\ell-1}^\diamond,
\end{aligned}$$

*where*

$$\|\tilde{R}_\ell^\diamond\|_{L_2} = O(h^3 + \epsilon^2 h^2), \quad \|\tilde{S}_{1,\ell}^\diamond\|_{L_2} = O(\epsilon h^{5/2}), \quad \|\tilde{S}_{2,\ell-1}^\diamond\|_{L_2} = O(\epsilon h^{5/2}). \tag{4.17}$$

*The terms $\tilde{S}_{1,\ell}^\circ, \tilde{S}_{2,\ell-1}^\circ$ are given by (4.12, 4.13) in the proof of Lemma 4.1 and satisfy here*

$$\|\tilde{S}_{1,\ell}^\circ\|_{L_2} = \mathcal{O}(\epsilon^2 h + \epsilon h^{3/2}), \quad \|\tilde{S}_{2,\ell}^\circ\|_{L_2} = \mathcal{O}(\epsilon^2 h + \epsilon h^{3/2}). \tag{4.18}$$

*Proof.* We have from Lemma 4.1, if the consistency conditions (4.10) are satisfied, the representation

$$L_\ell = \tilde{R}_\ell^\circ + \tilde{S}_{1,\ell}^\circ + \tilde{S}_{2,\ell-1}^\circ,\ \ell = 2, \ldots, N,$$

where $\tilde{R}_\ell^\circ$, $\tilde{S}_{1,\ell}^\circ$, $\tilde{S}_{2,\ell-1}^\circ$ are given by (4.11, 4.12, 4.13). Splitting $\Lambda_0 f = \Lambda_0^f f + \epsilon^2 \hat{\Lambda}_0 f$ immediately yields $\tilde{R}_\ell^\circ = \tilde{R}_\ell^{\circ f} + \epsilon^2 \hat{R}_\ell^\circ$ with

$$\begin{aligned}
\tilde{R}_\ell^{\circ f} := {}& (\alpha_0 - \beta_0 - \beta_1) h I_0^{t_{\ell-2}, t_{\ell-1}}(\Lambda_0^f f) + (\alpha_0 + \alpha_1) I_{00}^{t_{\ell-2}, t_{\ell-1}}(\Lambda_0^f f) \\
& + \alpha_0 I_{00}^{t_{\ell-1}, t_\ell}(\Lambda_0^f f) - h\beta_0 I_0^{t_{\ell-1}, t_\ell}(\Lambda_0^f f)
\end{aligned} \tag{4.19}$$

$$\begin{aligned}
\hat{R}_\ell^\circ := {}& (\alpha_0 - \beta_0 - \beta_1) h I_0^{t_{\ell-2}, t_{\ell-1}}(\hat{\Lambda}_0 f) + (\alpha_0 + \alpha_1) I_{00}^{t_{\ell-2}, t_{\ell-1}}(\hat{\Lambda}_0 f) \\
& + \alpha_0 I_{00}^{t_{\ell-1}, t_\ell}(\hat{\Lambda}_0 f) - h\beta_0 I_0^{t_{\ell-1}, t_\ell}(\hat{\Lambda}_0 f).
\end{aligned} \tag{4.20}$$

We note that (4.20) appears with the factor $\epsilon^2$ in the local error representation, thus yielding the $\mathcal{O}(\epsilon^2 h^2)$ term in the estimate of $\|\tilde{R}_\ell^\diamond\|_{L_2}$ in (4.17) . We concentrate on developing $\tilde{R}_\ell^{\circ f}$ in more detail. Applying the Itô-formula (4.3) to $\Lambda_0^f f(X(s),s)$ for $s\in[t_{\ell-2},t_{\ell-1}]$ and integrating yields

$$I_0^{t_{\ell-2},s}(\Lambda_0^f f) = (s-t_{\ell-2})\Lambda_0^f f(X(t_{\ell-2}),t_{\ell-2}) + I_{00}^{t_{\ell-2},s}(\Lambda_0\Lambda_0^f f) + \epsilon\sum_{r=1}^m I_{r0}^{t_{\ell-2},s}(\hat\Lambda_r\Lambda_0^f f).$$

For $s=t_{\ell-1}$ we obtain an expression for the first integral in (4.19) . Integrating again we obtain for the second integral in (4.19)

$$I_{00}^{t_{\ell-2},t_{\ell-1}}(\Lambda_0^f f) = \tfrac{h^2}{2}\Lambda_0^f f(X(t_{\ell-2}),t_{\ell-2}) + I_{000}^{t_{\ell-2},t_{\ell-1}}(\Lambda_0\Lambda_0^f f) + \epsilon\sum_{r=1}^m I_{r00}^{t_{\ell-2},t_{\ell-1}}(\hat\Lambda_r\Lambda_0^f f).$$

Both the other integrals are over the interval $[t_{\ell-1},t_\ell]$. In the analogous expressions for these the term $\Lambda_0^f f(X(t_{\ell-1}),t_{\ell-1})$ has to be substituted by

$$\Lambda_0^f f(X(t_{\ell-1}),t_{\ell-1}) = \Lambda_0^f f(X(t_{\ell-2}),t_{\ell-2}) + I_0^{t_{\ell-2},t_{\ell-1}}(\Lambda_0\Lambda_0^f f) + \epsilon\sum_{r=1}^m I_r^{t_{\ell-2},t_{\ell-1}}(\Lambda_r\Lambda_0^f f).$$

Then we obtain from (4.19)

$$\tilde{R}_\ell^{\circ f} = \Big[(\alpha_0+4\alpha_1)-(4\beta_0+2\beta_1)\Big]\frac{h^2}{2}\Lambda_0^f f(X(t_{\ell-2}),t_{\ell-2}) + \tilde{R}_\ell^{\diamond f} + \tilde{S}_{1,\ell}^{\diamond} + \tilde{S}_{2,\ell}^{\diamond},$$

where

$$\begin{aligned}\tilde{R}_\ell^{\diamond f} &= (\alpha_0-2\beta_0)\frac{h^2}{2}I_0^{t_{\ell-2},t_{\ell-1}}(\Lambda_0\Lambda_0^f f)\\ &\qquad + (\alpha_0-\beta_0-\beta_1)hI_{00}^{t_{\ell-2},t_{\ell-1}}(\Lambda_0\Lambda_0^f f) - \beta_0 h I_{00}^{t_{\ell-1},t_\ell}(\Lambda_0\Lambda_0^f f)\\ &\qquad + (\alpha_0+\alpha_1)I_{000}^{t_{\ell-2},t_{\ell-1}}(\Lambda_0\Lambda_0^f f) + \alpha_0 I_{000}^{t_{\ell-1},t_\ell}(\Lambda_0\Lambda_0^f f),\end{aligned}$$

$$\tilde{S}_{1,\ell}^{\diamond} = \alpha_0\epsilon\sum_{r=1}^m I_{r00}^{t_{\ell-1},t_\ell}(\hat\Lambda_r\Lambda_0^f f) - h\beta_0\epsilon\sum_{r=1}^m I_{r0}^{t_{\ell-1},t_\ell}(\hat\Lambda_r\Lambda_0^f f),$$

$$\begin{aligned}\tilde{S}_{2,\ell}^{\diamond} &= (\alpha_0-2\beta_0)\frac{h^2}{2}\epsilon\sum_{r=1}^m I_r^{t_{\ell-2},t_{\ell-1}}(\hat\Lambda_r\Lambda_0^f f) + (\alpha_0-\beta_0-\beta_1)h\epsilon\sum_{r=1}^m I_{r0}^{t_{\ell-2},t_{\ell-1}}(\hat\Lambda_r\Lambda_0^f f)\\ &\qquad + (\alpha_0+\alpha_1)\epsilon\sum_{r=1}^m I_{r00}^{t_{\ell-2},t_{\ell-1}}(\hat\Lambda_r\Lambda_0^f f).\end{aligned}$$

We arrive at $\tilde{R}_\ell^{\diamond} = \tilde{R}_\ell^{\diamond f} + \epsilon^2\hat{R}_\ell^{\circ}$. Finally, the estimates (4.17) are derived by means of Lemma 2.12. ∎

Corollary 4.4. *Let the coefficients $f$, $\hat g_r$, $r=1,\dots,m$, of the SDE (1.2) satisfy the assumptions of Lemma 4.3 and suppose they are Lipschitz continuous with respect to their first variable. Let the coefficients of the stochastic linear two-step scheme (4.1) satisfy Dahlquist's root condition and the consistency conditions (4.10) and*

$$(\alpha_0+4\alpha_1)-(4\beta_0+2\beta_1)=0. \tag{4.21}$$

*Then the global error of the scheme (4.1) applied to (1.2) allows the expansion*

$$\max_{\ell=2,\dots,N}\|X(t_\ell)-X_\ell\|_{L_2} = \mathcal{O}(h^2+\epsilon h+\epsilon^2 h^{1/2}) + \mathcal{O}(\max_{\ell=0,1}\|X(t_\ell)-X_\ell\|_{L_2})\,.$$

*Proof.* Lemma 4.3 stated the representation (4.16) for the local error. Applying the consistency condition (4.21) yields $R_\ell^{\diamond} = \tilde{R}_\ell^{\diamond}$ and by (4.17) we have $\|R_\ell^{\diamond}\|_{L_2} = \mathcal{O}(h^3+\epsilon^2h^2)$. The stochastic terms $S_{1,\ell}^{\diamond}$, $S_{2,\ell-1}^{\diamond}$ are dominated by $\tilde{S}_{1,\ell}^{\circ}$, $\tilde{S}_{2,\ell-1}^{\circ}$ and thus are of order of magnitude $\mathcal{O}(\epsilon^2 h + \epsilon h^{3/2})$. As the scheme (4.1) satisfies Dahlquist's root condition, it is numerically stable in the mean-square sense. Applying the stability inequality (2.8) to the representation (4.16) of the local error yields the assertion. □

REMARK 4.5. *We note that $R_\ell^\diamond$ is responsible for the $\mathcal{O}(h^2)$ term in the expansion of the global error. In the limit $\epsilon \to 0$ this is the only remaining term. It reflects the convergence properties of the scheme in the deterministic setting. On the other hand, for asymptotically small stepsizes $h \to 0$ and fixed parameter $\epsilon$, even if it is small, the term of order $\mathcal{O}(\epsilon^2 h^{1/2})$ causes the low order $\frac{1}{2}$ of convergence. The question arises for which choices of $\epsilon$ and $h$ the schemes still show the order 2 behaviour. Thus we are interested in when the term $\mathcal{O}(h^2)$ dominates the term $\mathcal{O}(\epsilon h + \epsilon^2 h^{1/2})$. Clearly both these terms depend on the actual coefficients $f, \hat{g}_r, r = 1, \ldots, m$, of the SDE (1.2) and their derivatives. Assuming moderate function values, the term $\mathcal{O}(h^2)$ dominates $\mathcal{O}(\epsilon h + \epsilon^2 h^{1/2})$, if $h^2 \gg \epsilon^2 h^{1/2}$, i.e. $h \gg \epsilon^{4/3}$, and $h^2 \gg \epsilon h$, i.e. $h \gg \epsilon$. Obviously, in general the second condition is stronger. Summarizing, we can expect order 2 behaviour if $h \gg \epsilon$.*
*However, even if the chosen stepsize does not satisfy this condition for a given $\epsilon$, schemes satisfying the consistency condition (4.21) often show a better performance than other schemes. The reason is that their error is dominated by $\mathcal{O}(\epsilon h)$ instead of $\mathcal{O}(\epsilon^2 h^{1/2})$, resulting in an order 1 behaviour with the small parameter $\epsilon$ in the error constant. Again assuming moderate function values, one may expect this for $\epsilon h \gg \epsilon^2 h^{1/2}$, i.e. $h \gg \epsilon^2$.*

**5. Test results.** We report results for several explicit and implicit stochastic linear $k$-step schemes for $k = 1, 2$, applied to the simple bilinear scalar SDE

$$X(t) = 1 + \int_0^t \alpha X(s) ds + \int_0^t \beta X(s) dW(s), \; t \in [0, 1], \tag{5.1}$$

with coefficients $f(x, t) := \alpha x$, $G(x, t) = (g(x, t)) = (\beta x)$, parameters $\alpha, \beta \in \mathbb{R}$ and a scalar Wiener process $W$. Its solution is given by the geometric Brownian motion

$$X(t) = \exp\big((\alpha - \frac{1}{2}\beta^2)t + \beta W(t)\big).$$

Table 5.1 summarizes the methods we have implemented and tested. All methods, including the one-step schemes, satisfy the consistency conditions (4.10) and all methods, excluding the Euler schemes, satisfy the consistency condition (4.21). In the following we present results for $\alpha = -1$ and different values of the parameter $\beta$, which takes the role of the parameter $\epsilon$. We plot the achieved accuracy versus the stepsizes in logarithmic scale. The accuracy is measured as the maximum approximate $L_2$-norm of the global errors in the time-interval $[0, 1]$:

$$\max_{\ell=1,\ldots,N} \Big(\frac{1}{M} \sum_{j=1}^{M} |X(t_\ell, \omega_j) - X_\ell(\omega_j)|^2\Big)^{1/2} \;\approx\; \max_{\ell=1,\ldots,N} \|X(t_\ell) - X_\ell\|_{L_2}$$

where N denotes the number of steps and M the number of computed paths. In our computations we used $M = 100$. Lines with slopes $-1$ and $-0.5$ are provided in some figures to enable comparisons with convergence of order 1 or 1/2.
Our first test concerned the effect of applying a method which is not numerically stable in the mean-square sense, i.e. the coefficients of the method do not satisfy Dahlquist's

| Method | $\alpha_0$ | $\alpha_1$ | $\alpha_2$ | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\gamma_1$ | $\gamma_2$ |
|---|---|---|---|---|---|---|---|---|
| unstable method | 1 | $-3$ | 2 | 0 | $\frac{1}{2}$ | $-\frac{3}{2}$ | 1 | $-2$ |
| BDF 2 | 1 | $-\frac{4}{3}$ | $\frac{1}{3}$ | $\frac{2}{3}$ | 0 | 0 | 1 | $-\frac{1}{3}$ |
| explicit Euler | 1 | $-1$ | 0 | 0 | 1 | 0 | 1 | 0 |
| implicit Euler | 1 | $-1$ | 0 | 1 | 0 | 0 | 1 | 0 |
| trapezoidal rule | 1 | $-1$ | 0 | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 | 1 | 0 |
| Adams-Bashforth 2 | 1 | $-1$ | 0 | 0 | $\frac{3}{2}$ | $-\frac{1}{2}$ | 1 | 0 |
| Adams-Moulton 2 | 1 | $-1$ | 0 | $\frac{5}{12}$ | $\frac{8}{12}$ | $-\frac{1}{12}$ | 1 | 0 |
| Milne-Simpson | 1 | 0 | $-1$ | $\frac{1}{3}$ | $\frac{4}{3}$ | $\frac{1}{3}$ | 1 | 1 |

TABLE 5.1
*Table 1: coefficients of two-step schemes*

root condition. In this case, one of the roots of the characteristic polynomial $\rho$ (2.12) is 2. Figure 1 shows the behaviour of the error, when the stepsize decreases. We note that the scale is logarithmic!

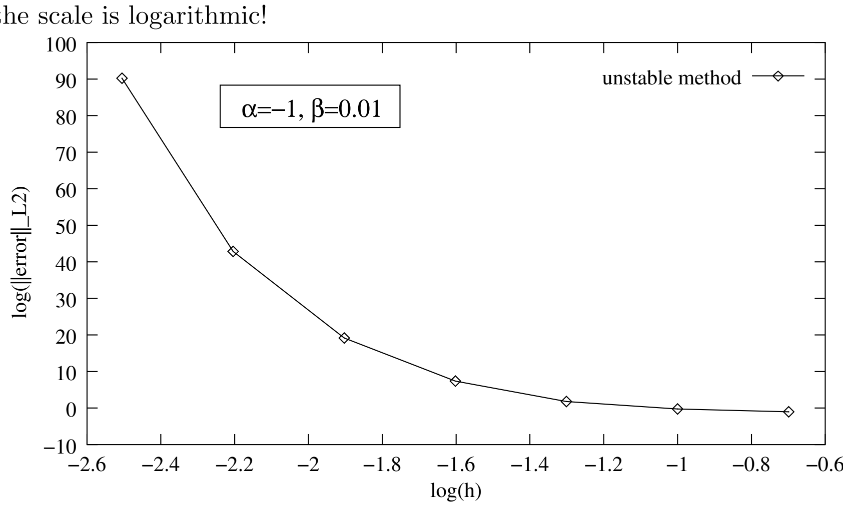


Figure 1: Simulation results for the SDE (5.1) with the unstable scheme

In Figure 2 we show the effect of applying the stable methods to deterministic ODES, i.e. with $\beta = 0$. Clearly the orders of these methods are reflected by the slope of the lines in the diagram. They reach from 1 for the Euler schemes over 2 for the trapezoidal rule, the two-step BDF and Adams-Bashforth methods, 3 for the two-step Adams-Moulton method to 4 for the Milne-Simpson method. We note that the achievable accuracy is restricted by the machine precision. Additionally, especially the Milne-Simpson method is affected by a phenomenon called *weak instability*, see e.g. [10].

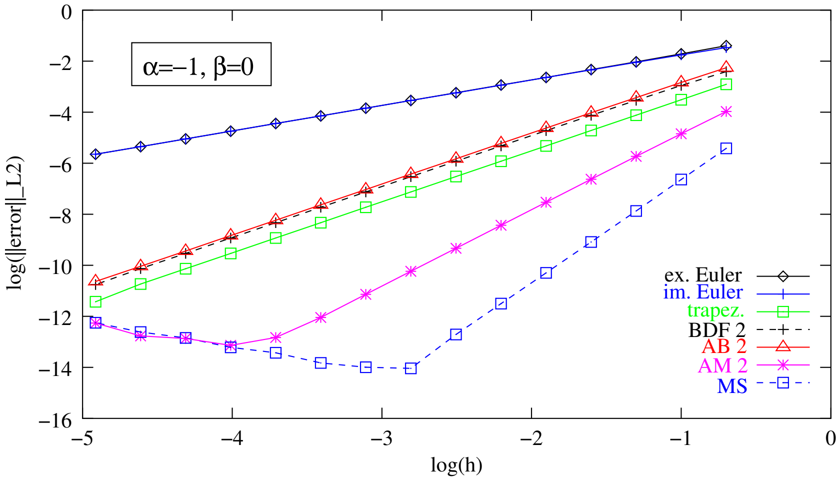


Figure 2: Simulation results for the SDE (5.1) for $\alpha = -1$ and $\beta = 0$ (deterministic case)

Figures 3 to 6 show simulation results for an ensemble of 100 computed paths using the parameter $\alpha = -1$ and several values for $\beta$. The results illustrate the observations made in Remark 4.5. We see in all figures except Figure 6 that the error of the other schemes is smaller than that of the Euler schemes. The error appears to be of the size $\max(\beta h, ch^2)$, $c$ being an error constant particular to the scheme. Figure 6 relates to the case where there is no small parameter in the diffusion coefficient. Here the order $h^{1/2}$ term dominates the error for all chosen stepsizes. Furthermore, due to the larger noise and the small number of paths, the statistical error in the approximation is visible.

For the Euler methods one clearly observes the dominating deterministic order 1 behaviour in Figures 3 and 4. The asymptotic mean-square order 1/2 becomes visible for larger values of the noise parameter $\beta$ in Figures 5 and 6, in the former only for small values of the stepsize $h$.

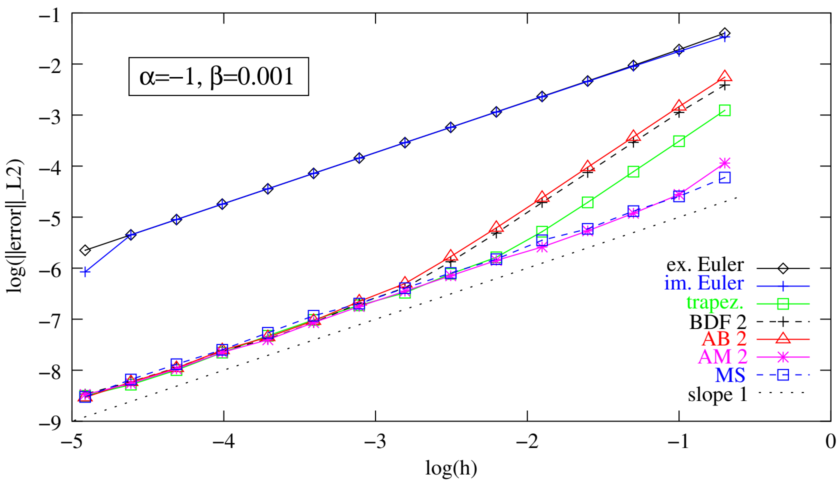


Figure 3

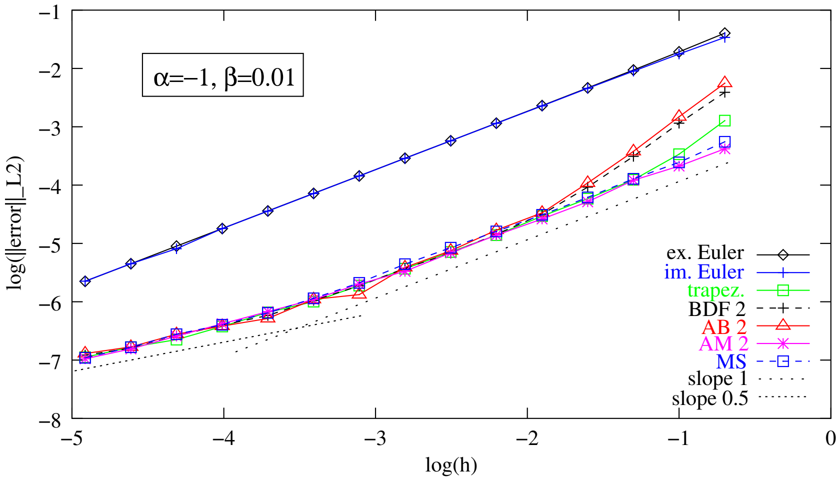


Figure 4

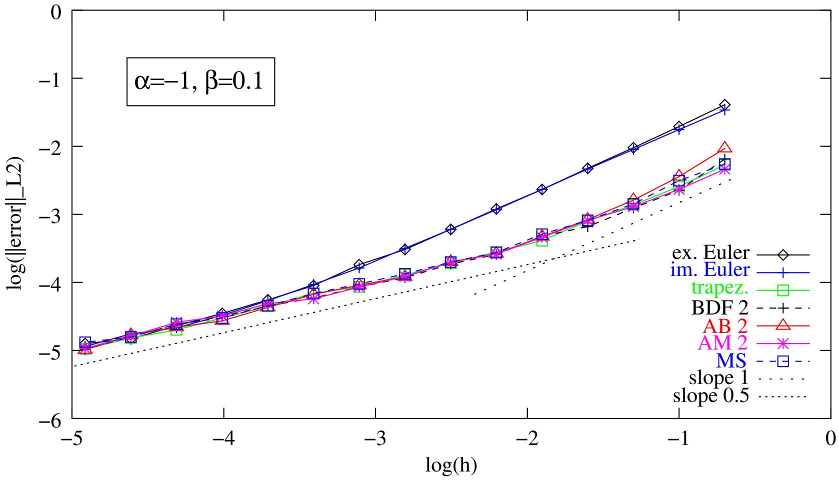


Figure 5

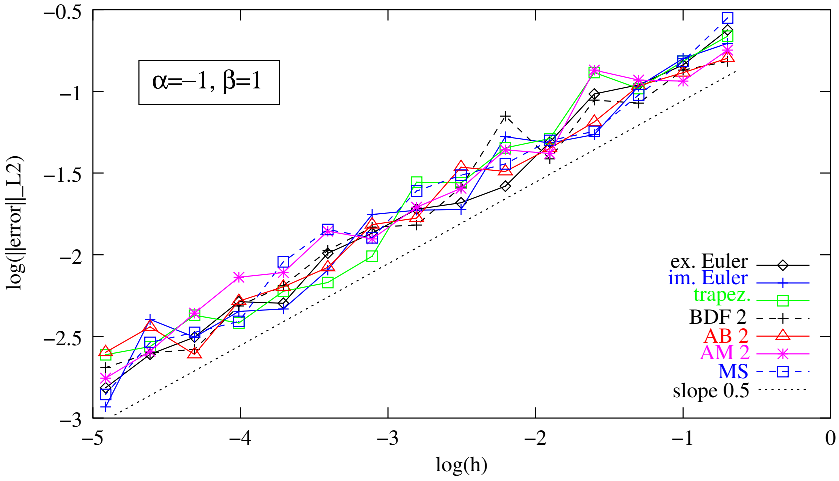


Figure 6

**Appendix A. Proof of Theorem 3.2.**

*Proof.* **Necessity:** This part can be proved as in the deterministic case, i. e. we take the equation $X'(t) = 0$, then $f$ and $\Gamma_j$ satisfy obviously (2.9) and (2.10). We then follow in principle the proof of [9, Thm.6.3.3].

**Sufficiency:** Since the SLMM (2.1) contains the stochastic part related to the $\Gamma_j$, we can not rely on the theory of difference equations and the representations of their solutions. Instead, we will follow the route of rewriting the $k$-step recurrence equation as a one-step recurrence equation in a higher dimensional space (see e.g. [10, Chap.III.4][18, Chap.8.2.1]).

For $X_\ell$ and $\tilde{X}_\ell$ being the solutions of (2.1) and (2.4), respectively, let the $n$-dimensional vector $E_\ell$ be defined as the difference $X_\ell - \tilde{X}_\ell$. We have with $E_0, \ldots, E_{k-1} \in L_2(\Omega, \mathbb{R}^n)$ for $\ell = k, \ldots, N$, the recursion

$$E_\ell = -\sum_{j=1}^{k} \alpha_j \; E_{\ell-j} + \underbrace{h \; \sum_{j=0}^{k} \beta_j \; \Delta f_{\ell-j}}_{=:\Delta\phi^\ell} + \underbrace{\sum_{j=1}^{k} \Delta\Gamma_{j,\ell-j} \; I^{t_{\ell-j}, t_{\ell-j+1}}}_{=:\Delta\psi^\ell} - D_\ell, \tag{A.1}$$

where

$$\begin{aligned} \Delta f_{\ell-j} &:= f(X_{\ell-j}, t_{\ell-j}) - f(\tilde{X}_{\ell-j}, t_{\ell-j}) \\ \Delta\Gamma_{j,\ell-j} &:= \Gamma_j(X_{\ell-j}, t_{\ell-j}) - \Gamma_j(\tilde{X}_{\ell-j}, t_{\ell-j}). \end{aligned}$$

We rearrange this $k$-step recursion in the space $L_2(\Omega, \mathbb{R}^n)$ to a one-step recursion in $L_2(\Omega, \mathbb{R}^{k\times n})$. Together with the trivial identities $E_{\ell-1} = E_{\ell-1}, \; \ldots \; E_{\ell-k+1} = E_{\ell-k+1}$ we obtain

$$\underbrace{\begin{pmatrix} E_\ell \\ E_{\ell-1} \\ \vdots \\ E_{\ell-k+1} \end{pmatrix}}_{=:\,\mathcal{E}_\ell} = \underbrace{\begin{pmatrix} -\alpha_1 I & \cdots & \cdots & -\alpha_k I \\ I & 0 & & \\ & \ddots & \ddots & \\ & & I & 0 \end{pmatrix}}_{=:\,\mathcal{A}} \underbrace{\begin{pmatrix} E_{\ell-1} \\ E_{\ell-2} \\ \vdots \\ E_{\ell-k} \end{pmatrix}}_{=:\,\mathcal{E}_{\ell-1}} + \underbrace{\begin{pmatrix} \Delta\phi^\ell \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{=:\,\Delta\Phi_\ell} + \underbrace{\begin{pmatrix} \Delta\psi^\ell \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{=:\,\Delta\Psi_\ell} + \underbrace{\begin{pmatrix} -D_\ell \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{=:\,\mathcal{D}_\ell}$$

or, in compact form

$$\mathcal{E}_\ell = \mathcal{A}\mathcal{E}_{\ell-1} + \Delta\Phi_\ell + \Delta\Psi_\ell + \mathcal{D}_\ell \;,\; \ell = k, \ldots, N \quad \text{and} \quad \mathcal{E}_{k-1} = (-D_{k-1}, -D_{k-2}, \ldots, -D_0)^T,$$

where $\mathcal{E}_\ell \in L_2(\Omega, \mathbb{R}^{k\times n}),\; \ell = k-1, k, \ldots, N$. The vector $\mathcal{E}_{k-1}$ consists of the perturbations to the initial values. We now trace back the recursion in $\mathcal{E}_\ell$ to the initial vector $\mathcal{E}_{k-1}$. For $\ell = k, \ldots, N$ we have

$$\begin{aligned} \mathcal{E}_\ell &= \mathcal{A}\mathcal{E}_{\ell-1} + \Delta\Phi_\ell + \Delta\Psi_\ell + \mathcal{D}_\ell \\ &= \mathcal{A}(\mathcal{A}\mathcal{E}_{\ell-2} + \Delta\Phi_{\ell-1} + \Delta\Psi_{\ell-1} + \mathcal{D}_{\ell-1}) + \Delta\Phi_\ell + \Delta\Psi_\ell + \mathcal{D}_\ell \\ &= \mathcal{A}^2\mathcal{E}_{\ell-2} + (\Delta\Phi_\ell + \mathcal{A}\Delta\Phi_{\ell-1}) + (\Delta\Psi_\ell + \mathcal{A}\Delta\Psi_{\ell-1}) + (\mathcal{D}_\ell + \mathcal{A}\mathcal{D}_{\ell-1}) \\ &\;\;\vdots \\ &= \mathcal{A}^{\ell-k+1}\mathcal{E}_{k-1} + \sum_{i=0}^{\ell-k} \mathcal{A}^i \Delta\Phi_{\ell-i} + \sum_{i=0}^{\ell-k} \mathcal{A}^i \Delta\Psi_{\ell-i} + \sum_{i=0}^{\ell-k} \mathcal{A}^i \mathcal{D}_{\ell-i} \\ &= \mathcal{A}^{\ell-k+1}\mathcal{E}_{k-1} + \sum_{i=k}^{\ell} \mathcal{A}^{\ell-i} \Delta\Phi_i + \sum_{i=k}^{\ell} \mathcal{A}^{\ell-i} \Delta\Psi_i + \sum_{i=k}^{\ell} \mathcal{A}^{\ell-i} \mathcal{D}_i \;. \end{aligned}$$

A crucial point for the subsequent calculations is to find a scalar product inducing a matrix norm such that this norm of the matrix $\mathcal{A}$ is less than or equal to 1 (see e.g. [10, Chap.III.4,Lemma 4.4]. This is possible if the eigenvalues of the Frobenius matrix $\mathcal{A}$ lie inside the unit circle of the complex plane and are simple if their modulus is equal to 1. The eigenvalues of $\mathcal{A}$ are the roots of the characteristic polynomial $\rho$ (2.12) and due to the assumption that Dahlquist's root condition is satisfied they have the required property. Then there exists a non-singular matrix $\mathcal{C}$ with a block-structure like $\mathcal{A}$ such that $\|\mathcal{C}^{-1}\mathcal{A}\mathcal{C}\|_2 \leq 1$, where $\|\cdot\|_2$ denotes the spectral matrix norm that is induced by the Euclidian vector norm in $\mathbb{R}^{k\times n}$. We can thus choose a scalar product for $\mathcal{X},\mathcal{Y} \in \mathbb{R}^{k\times n}$ as

$$\langle \mathcal{X},\mathcal{Y}\rangle_* := \langle \mathcal{C}^{-1}\mathcal{X},\mathcal{C}^{-1}\mathcal{Y}\rangle_2$$

and then have $|.|_*$ as the induced vector norm on $\mathbb{R}^{k\times n}$ and $\|\cdot\|_*$ as the induced matrix norm with $\|\mathcal{A}\|_* = \|\mathcal{C}^{-1}\mathcal{A}\mathcal{C}\|_2 \leq 1$. We also have

$$\langle \mathcal{X},\mathcal{Y}\rangle_* = \mathcal{X}^T\mathcal{C}^{-T}\mathcal{C}^{-1}\mathcal{Y} = \mathcal{X}^T\mathcal{C}^*\mathcal{Y} \quad \text{with} \quad \mathcal{C}^* = \mathcal{C}^{-T}\mathcal{C}^{-1} = (c^*_{ij}I_n)_{i,j=1,\ldots,k}.$$

Due to the norm equivalence there are constants $c^*, c_* > 0$ such that

$$|\mathcal{X}|_2^2 \leq c^*|\mathcal{X}|_*^2 \quad \text{and} \quad |\mathcal{X}|_*^2 \leq c_*|\mathcal{X}|_\infty^2 \quad \forall \mathcal{X} \in \mathbb{R}^{k\times n}\ ,$$

where $|\mathcal{X}|_2^2 = \sum_{j=1,\ldots,k}|x_j|^2$, $|\mathcal{X}|_\infty = \max_{j=1,\ldots,k}|x_j|$ for $\mathcal{X} = (x_1^T,\ldots,x_k^T)^T$. For the special vectors $\mathcal{X} = (x^T,0,\ldots,0)^T$ and $\mathcal{Y} = (y^T,0,\ldots,0)^T$ with $\mathcal{X},\mathcal{Y} \in \mathbb{R}^{k\times n}$ and $x,y \in \mathbb{R}^n$, one has $\langle \mathcal{X},\mathcal{Y}\rangle_* = c^*_{11}\langle x,y\rangle_2 = c^*_{11}x^Ty$, where $c^*_{11}$ is given by the matrix $\mathcal{C}^*$.

We now apply $|.|_*^2$ to estimate $|\mathcal{E}_\ell|_*^2$ and, later, $\mathbb{E}|\mathcal{E}_\ell|_*^2$. We start with

$$|\mathcal{E}_\ell|_*^2 \leq 4\Big\{\underbrace{|\mathcal{A}^{\ell-k+1}\mathcal{E}_{k-1}|_*^2}_{1)} + \underbrace{|\sum_{i=k}^{\ell}\mathcal{A}^{\ell-i}\Delta\Phi_i|_*^2}_{2)} + \underbrace{|\sum_{i=k}^{\ell}\mathcal{A}^{\ell-i}\Delta\Psi_i|_*^2}_{3)} + \underbrace{|\sum_{i=k}^{\ell}\mathcal{A}^{\ell-i}\mathcal{D}_i|_*^2}_{4)}\Big\}.$$

For the term labelled 1) we have $|\mathcal{A}^{\ell-k+1}\mathcal{E}_{k-1}|_*^2 \leq |\mathcal{E}_{k-1}|_*^2$, and thus

$$\mathbb{E}|\mathcal{A}^{\ell-k+1}\mathcal{E}_{k-1}|_*^2 \leq \mathbb{E}|\mathcal{E}_{k-1}|_*^2\ . \tag{A.2}$$

For the term labelled 2) we have

$$\begin{aligned}
&|\sum_{i=k}^{\ell}\mathcal{A}^{\ell-i}\Delta\Phi_i|_*^2 \leq (\ell-k+1)\sum_{i=k}^{\ell}|\mathcal{A}^{\ell-i}\Delta\Phi_i|_*^2 \ \leq\ N\sum_{i=k}^{\ell}|\Delta\Phi_i|_*^2 = \frac{T}{h}c^*_{11}\sum_{i=k}^{\ell}|\Delta\phi^i|^2\\
&\leq h\ T\ c^*_{11}\ (k+1)\sum_{i=k}^{\ell}\sum_{j=0}^{k}|\beta_j\ \Delta f_{i-j}|^2 \leq h\ T\ c^*_{11}\ (k+1)\ L_f^2\sum_{i=k}^{\ell}\sum_{j=0}^{k}\beta_j^2\ |E_{i-j}|^2\\
&\leq h\ T\ c^*_{11}\ (k+1)\ L_f^2\Big\{\beta_0^2\ |E_\ell|^2 + \sum_{i=k}^{\ell}\{\beta_0^2\ |E_{i-1}|^2 + \sum_{j=1}^{k}\beta_j^2\ |E_{i-j}|^2\}\Big\}\\
&\leq h\ T\ c^*_{11}\ (k+1)\ L_f^2\Big\{c^*\ \beta_0^2\ |\mathcal{E}_\ell|_*^2 + C_\beta\ c^*\sum_{i=k-1}^{\ell-1}|\mathcal{E}_i|_*^2\Big\},
\end{aligned}$$

where $C_\beta = 2\ \max_{j=0,\ldots,k}\beta_j$. Hence,

$$\mathbb{E}|\sum_{i=k}^{\ell}\mathcal{A}^{\ell-i}\Delta\Phi_i|_*^2 \leq h\ T\ c^*_{11}\ (k+1)\ L_f^2\ \Big\{c^*\ \beta_0^2\ \mathbb{E}|\mathcal{E}_\ell|_*^2 + C_\beta\ c^*\sum_{i=k-1}^{\ell-1}\mathbb{E}|\mathcal{E}_i|_*^2\Big\}. \tag{A.3}$$

We will now treat the term labelled 3). For that purpose we introduce the notation $\Delta\Psi_{j,i-j} := ((\Delta\Gamma_{j,i-j} I^{t_{i-j},t_{i-j+1}})^T, 0, \ldots, 0)^T$. Using this we can write

$$\Delta\Psi_i = ((\Delta\psi^i)^T, 0, \ldots, 0)^T = ((\sum_{j=1}^{k} \Delta\Gamma_{j,i-j}\ I^{t_{i-j},t_{i-j+1}})^T, 0, \ldots, 0)^T = \sum_{j=1}^{k} \Delta\Psi_{j,i-j}$$

and $\quad |\sum_{i=k}^{\ell} \mathcal{A}^{\ell-i}\Delta\Psi_i|_*^2 \;=\; |\sum_{i=k}^{\ell} \mathcal{A}^{\ell-i} \sum_{j=1}^{k} \Delta\Psi_{j,i-j}|_*^2.$

Every $\Delta\Psi_{j,i-j}$ is $\mathcal{F}_{t_{i-j+1}}$-measurable and $\mathbb{E}(\Delta\Psi_{j,i-j}|\mathcal{F}_{t_{i-j}}) = 0$. We can now reorder the last term above such that we have a sum of terms where each term contains all multiple Wiener integrals over just one subinterval. The expectation of products of terms from different subintervals vanishes, hence we obtain

$$\begin{aligned}
&\mathbb{E}|\sum_{i=k}^{\ell} \mathcal{A}^{\ell-i}\Delta\Psi_i|_*^2\\
&= \mathbb{E}|\mathcal{A}^{\ell-k}\Delta\Psi_{k,0}|_*^2\\
&\quad + \mathbb{E}|\mathcal{A}^{\ell-k-1}\Delta\Psi_{k,1} + \mathcal{A}^{\ell-k}\Delta\Psi_{k-1,1}|_*^2\\
&\quad \vdots\\
&\qquad + \mathbb{E}|\mathcal{A}^{\ell-2k+1}\Delta\Psi_{k,k-1} + \mathcal{A}^{\ell-2k+2}\Delta\Psi_{k-1,k-1} + \ldots + \mathcal{A}^{\ell-k}\Delta\Psi_{1,k-1}|_*^2\\
&\qquad \vdots\\
&\qquad + \mathbb{E}|\mathcal{A}^{0}\Delta\Psi_{k,\ell-k} + \mathcal{A}^{1}\Delta\Psi_{k-1,\ell-k} + \ldots + \mathcal{A}^{k-1}\Delta\Psi_{1,\ell-k}|_*^2\\
&\quad \vdots\\
&\quad + \mathbb{E}|\mathcal{A}^{0}\Delta\Psi_{2,\ell-2} + \mathcal{A}^{1}\Delta\Psi_{1,\ell-2}|_*^2\\
&\quad + \mathbb{E}|\mathcal{A}^{0}\Delta\Psi_{1,\ell-1}|_*^2
\end{aligned}$$

$$\begin{aligned}
&\leq k\ \sum_{i=k}^{\ell}\sum_{j=1}^{k} \mathbb{E}|\Delta\Psi_{j,i-j}|_*^2 \ \leq k\ c_{11}^*\ \sum_{i=k}^{\ell}\sum_{j=1}^{k} \mathbb{E}\|\Delta\Gamma_{j,i-j}\|^2\ \mathbb{E}|I^{t_{i-j},t_{i-j+1}}|^2\\
&\leq h\ k\ c_{11}^*\ L_\Gamma^2\ \sum_{i=k}^{\ell}\sum_{j=1}^{k} \mathbb{E}|E_{i-j}|^2 \ \leq\ h\ k\ c_{11}^*\ L_\Gamma^2\ c^*\ \sum_{i=k}^{\ell} |\mathcal{E}_{i-1}|_*^2.
\end{aligned}$$

Thus, for the term labelled 3), we obtain

$$\mathbb{E}|\sum_{i=k}^{\ell} \mathcal{A}^{\ell-i}\Delta\Psi_i|_*^2 \leq\ h\ k\ c_{11}^*\ L_\Gamma^2\ c^*\ \sum_{i=k-1}^{\ell-1} |\mathcal{E}_i|_*^2. \tag{A.4}$$

We will, for a shorter notation, deal with the term labelled 4), i.e. the perturbations $D_i$ in $\mathcal{D}_i$, after obtaining an intermediate result. Using (A.2), (A.3) and (A.4) and setting $L_0 := L_f^2\ (k+1)\ c_{11}^*\ T\ c^*\ \beta_0^2$ and $L := L_f^2\ (k+1)\ c_{11}^*\ T\ c_\beta^* \ +\ L_\Gamma^2\ k\ c_{11}^*\ c^*$, we have now arrived at

$$\mathbb{E}|\mathcal{E}_\ell|_*^2 \leq 4\Big\{\mathbb{E}|\mathcal{E}_{k-1}|_*^2 + hL_0\mathbb{E}|\mathcal{E}_\ell|^2 + hL\sum_{i=k-1}^{\ell-1}\mathbb{E}|\mathcal{E}_i|_*^2 + \mathbb{E}|\sum_{i=k}^{\ell}\mathcal{A}^{\ell-i}\mathcal{D}_i|_*^2\Big\},\ \ell = k, \ldots, N.$$

If necessary we choose a bound $h^0$ on the stepsize such that $4\,h\,L_0 < \frac{1}{2}$ holds for all $h < h^0$ and conclude that

$$\mathbb{E}|\mathcal{E}_\ell|_*^2 \le 8\,\mathbb{E}|\mathcal{E}_{k-1}|_*^2 + 8\,\mathbb{E}|\sum_{i=k}^{\ell}\mathcal{A}^{\ell-i}\mathcal{D}_i|_*^2 + 8\,L\,T\,\frac{1}{N}\sum_{i=k-1}^{\ell-1}\mathbb{E}|\mathcal{E}_i|_*^2.$$

We now apply Lemma 2.11 with $a_\ell := 0,\ \ell = 1,\dots,k-2$ and $a_\ell := \mathbb{E}|\mathcal{E}_\ell|_*^2,\ \ell = k-1,\dots,N$, and obtain the intermediate result

$$\max_{\ell=k-1,\dots,N}\mathbb{E}|\mathcal{E}_\ell|_*^2 \le \hat{S}\Big\{\mathbb{E}|\mathcal{E}_{k-1}|_*^2 + \max_{\ell=k,\dots,N}\mathbb{E}|\sum_{i=k}^{\ell}\mathcal{A}^{\ell-i}\mathcal{D}_i|_*^2\Big\},\ \hat{S} := 8\exp(8LT)\ . \quad \text{(A.5)}$$

It remains to deal with the term labelled 4), i.e. the perturbations $D_i$ in $\mathcal{D}_i$. We decompose $D_i$, and, analogously, $\mathcal{D}_i$ into

$$D_i = R_i + S_i = R_i + \sum_{j=1}^{k} S_{j,i-j+1}, \qquad \mathcal{D}_i = \mathcal{R}_i + \mathcal{S}_i = \mathcal{R}_i + \sum_{j=1}^{k}\mathcal{S}_{j,i-j+1},$$

where $S_{j,i-j+1}$ is $\mathcal{F}_{t_{i-j+1}}$-measurable with $\mathbb{E}(S_{j,i-j+1}|\mathcal{F}_{t_{i-j}}) = 0$ for $i = k,\dots,N$ and $j = 1,\dots,k$.
Then $\mathbb{E}\langle\mathcal{A}^{\ell_1}\mathcal{S}_{j_1,i_1},\mathcal{A}^{\ell_2}\mathcal{S}_{j_2,i_2}\rangle_* = 0$ for $i_1 \ne i_2$, and by similar computations as above we obtain

$$\begin{aligned}\mathbb{E}|\sum_{i=k}^{\ell}\mathcal{A}^{\ell-i}\mathcal{D}_i|_*^2 &\le 2\,(\ell-k+1)\sum_{i=k}^{\ell}\mathbb{E}|\mathcal{A}^{\ell-i}\mathcal{R}_i|_*^2 + 2\,k\,\sum_{i=k}^{\ell}\sum_{j=1}^{k}\mathbb{E}|\mathcal{A}^{\ell-i}\mathcal{S}_{j,i-j+1}|_*^2\\ &\le 2\sum_{i=k}^{\ell}\Big(\frac{T}{h}\,\mathbb{E}|\mathcal{R}_i|_*^2 + k\sum_{j=1}^{k}\mathbb{E}|\mathcal{S}_{j,i-j+1}|_*^2\Big).\end{aligned}$$

Inserting this into the intermediate result (A.5) we obtain

$$\max_{\ell=k-1,\dots,N}\mathbb{E}|\mathcal{E}_\ell|_*^2 \le \hat{S}\Big\{\mathbb{E}|\mathcal{E}_{k-1}|_*^2 + 2\sum_{i=k}^{\ell}\Big(\frac{T}{h}\,\mathbb{E}|\mathcal{R}_i|_*^2 + k\sum_{j=1}^{k}\mathbb{E}|\mathcal{S}_{j,i-j+1}|_*^2\Big)\Big\},$$

and thus $\max\limits_{\ell=k-1,\dots,N}\mathbb{E}|E_\ell|^2$

$$\le c^*\hat{S}\Big\{c_*\max_{\ell=0,\dots,k-1}\mathbb{E}|E_\ell|^2 + 2\,c_{11}^*\max_{\ell=k,\dots,N}\Big(\frac{T^2}{h^2}\mathbb{E}|R_\ell|^2 + \frac{kT}{h}\sum_{j=1}^{k}\mathbb{E}|S_{j,\ell-j+1}|^2\Big)\Big\}.$$

Taking the square root yields the final estimate

$$\begin{aligned}&\max_{\ell=k-1,\dots,N}\|E_\ell\|_{L_2}\\ &\le \sqrt{c^*\hat{S}}\left\{\sqrt{c_*}\max_{\ell=0,\dots,k-1}\|E_\ell\|_{L_2} + \sqrt{2c_{11}^*}\max_{\ell=1,\dots,N}\left(\frac{T}{h}\|R_\ell\|_{L_2} + \sqrt{\frac{kT}{h}\sum_{j=1}^{k}\|S_{j,\ell-j+1}\|_{L_2}^2}\right)\right\}\\ &\le S\left\{\max_{\ell=0,\dots,k-1}\|E_\ell\|_{L_2} + \max_{\ell=1,\dots,N}\left(\frac{\|R_\ell\|_{L_2}}{h} + \frac{\sqrt{\sum_{j=1}^{k}\|S_{j,\ell-j+1}\|_{L_2}^2}}{\sqrt{h}}\right)\right\},\end{aligned}$$

which completes the proof. $\square$

REFERENCES

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

[2] R. H. Bokor, *On two-step methods for stochastic differential equations*, Acta Cybernetica, 13 (1997), pp. 197–207.

[3] ———, *Convergence and stability properties for numerical approximations of stochastic ordinary differential equations*, PhD Thesis, Universität Zagreb, 2000.

[4] ———, *Stochastically stable one-step approximations of solutions of stochastic ordinary differential equations*, J. Appl. Numer. Math., 44 (2003), pp. 299–312.

[5] L. Brugnano, K. Burrage, and P. Burrage, *Adams-type methods for the numerical solution of stochastic ordinary differential equations*, BIT, 40 (2000), pp. 451–470.

[6] G. Dahlquist, *33 years of numerical instability, Part I*, BIT, 25 (1985), pp. 188–204.

[7] G. Denk and S. Schäffler, *Adams methods for the efficient solution of stochastic differential equations with additive noise*, Computing, 59 (1996), pp. 153–161.

[8] T. C. Gard, *Introduction to Stochastic Differential Equations*, Marcel-Dekker, New York and Basel, 1988.

[9] W. Gautschi, *Numerical analysis. An introduction*, Birkhäuser, Boston, 1997.

[10] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving ordinary differential equations. I: Nonstiff problems. 2. rev. ed.*, vol. 8 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1993.

[11] E. Isaacson and H. Keller, *Analysis of numerical methods*, John Wiley and Sons, New York, 1966.

[12] P. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, Springer, Berlin, 1992.

[13] J. Lambert, *Numerical methods for ordinary differential systems: the initial value problem*, John Wiley & Sons , Chichester, 1991.

[14] P. Lax and R. Richtmyer, *Survey of the stability of linear finite difference equations*, Comm. Pure Appl. Math., 9 (1956), pp. 267–293.

[15] G. Milstein, *Theorem on the order of convergence for mean-square approximations of solutions of stochastic differential equations*, Theory Probab. Appl., 32 (1987), pp. 738–741. translation from Teor. Veroyatn. Primen. 32, No.4, 809-811 (1987).

[16] G. Milstein, *Numerical integration of stochastic differential equations*, Kluwer, 1995. Translation from the Russion original of 1988.

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

[18] R. Plato, *Numerische Mathematik kompakt. Grundlagenwissen für Studium und Praxis*, Vieweg, Braunschweig, 2000.

[19] W. Römisch and R. Winkler, *Stochastic DAEs in circuit simulation*, in Modeling, Simulation and Optimization of Integrated Circuits, A. G. K. Antreich, R. Bulirsch and P. Rentrop, eds., Birkhäuser, 2003, pp. 303–318.

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