TWO-SIDED ERROR ESTIMATES FOR STOCHASTIC ONESTEP AND MULTISTEP METHODS*

RAPHAEL KRUSE†

Abstract. This paper presents a unifying theory for the numerical analysis of stochastic onestep and multistep methods. In addition to well-known results on the error of strong convergence we prove a two-sided error estimate. This is characterized by Dahlquist's strong root condition and is used to determine the maximum order of convergence. In particular, we apply our theory to the stochastic theta method, BDF2-Maruyama and higher order Itô-Taylor schemes. The main ingredient of the stability analysis is a stochastic version of Spijker's norm.

Key words. SODE, Itô-Taylor schemes, BDF2-Maruyama, stochastic multistep method, two-sided error estimate, consistency, bistability, stochastic Spijker norm

AMS subject classifications. Primary 65C20, 65C30, 65L20 ; Secondary 65L06, 65L70

1. Introduction. In this paper we present a unifying theory to analyse the strong error of convergence for numerical methods applied to stochastic ordinary differential equations (SODEs). Motivated by the theory of discrete approximation (see [7, 8, 9, 10], [25], [26, 27, 28, 29]) we develop a notion of consistency, stability and convergence which allows to derive sharper versions of well-known results concerning the convergence of onestep schemes [16, 19, 20] and multistep methods [4].

The improvements are concerned with three aspects of the theory: A special choice of norms allows us to prove bistability in the sense of [29]. Our first result is a characterization of bistability in terms of a strong version of Dahlquist's root condition. Secondly, we derive a two-sided estimate of the strong error of convergence. This can be achieved by a suitable stochastic version of the deterministic Spijker norm (see [23, 24], [25, Ch.2.2], [11, Ch.III.8]). Finally, we use these two-sided error estimates to prove the maximum order of convergence and extend a known result [6] for Euler-Maruyama type methods to higher order schemes.

Our analysis applies to a wide range of stochastic onestep and multistep methods. In this paper we are only concerned with the stochastic theta method, higher order Itô-Taylor schemes and the BDF2-Maruyama method. But our results immediately carry over to all stochastic linear multistep methods mentioned in [4].

For the stochastic theta method a slightly weaker version of the two-sided error estimate was shown in [3]. The present article uses the same conceptual idea, but follows a different path in the proof of bistability. We also note that further numerical stability concepts have been developed for multistep methods [4], for stochastic differential algebraic equations [31] and stochastic delay equations [2, 5]. In the following we give a more technical outline of the paper.

We are interested in the numerical approximation of $\mathbb{R}^d$-valued stochastic processes X, which satisfy an ordinary Itô stochastic differential equation [1, 18, 21] of

*This work was supported by CRC 701 'Spectral Analysis and Topological Structures in Mathematics'.

†Department of Mathematics, Bielefeld University, P.O. Box 100131, 33501 Bielefeld, Germany (rkruse@math.uni-bielefeld.de).

the form

$$\begin{aligned} dX(t) &= b^0(t,X(t))dt + \sum_{r=1}^{m} b^r(t,X(t))dW^r(t), \quad t \in [0,T], \\ X(0) &= X_0. \end{aligned} \tag{1.1}$$

The drift and diffusion coefficient functions $b^r : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $r = 0,\dots,m$, are assumed to be measurable. The processes W^r, $r = 1,\dots,m$, are real and independent standard Brownian motions on a given complete probability space $(\Omega, \mathcal{F}, P)$, adapted to the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ which fulfills the usual conditions (i.e. the filtration is right-continuous and each $\mathcal{F}_t$ contains all P-null sets).

In addition, we assume that the following usual assumptions [1, 18, 21] hold:

(A1) The initial value X_0 is an $\mathcal{F}_0$-measurable and $\mathbb{R}^d$-valued random variable satisfying

$$\mathbb{E}(|X_0|^2) < \infty.$$

(A2) There exists a constant $K > 0$ such that

$$|b^r(t,x)| \le K(1 + |x|) \text{ and } |b^r(t,x) - b^r(t,y)| \le K|x - y|$$

for all $x, y \in \mathbb{R}^d$, $t \in [0,T]$ and $r = 0,\dots,m$.

Here we denote by $\mathbb{E}$ the expectation with respect to P and by $|\cdot|$ the Euclidean norm in $\mathbb{R}^d$. Assumptions (A1) and (A2) are sufficient to assure the existence and uniqueness of a strong Itô solution to (1.1) (see [1, 18, 21]), i.e. there exists a unique, P-a.s. continuous and $(\mathcal{F}_t)_{t\in[0,T]}$-adapted process X which satisfies

$$X(t) = X_0 + \int_0^t b^0(s,X(s))ds + \sum_{r=1}^{m} \int_0^t b^r(s,X(s))dW^r(s) \tag{1.2}$$

for all $t \in [0,T]$ and

$$\mathbb{E}\left(\int_0^T |X(s)|^2 ds\right) < \infty.$$

Next, we introduce a general form of a stochastic k-step method which we use for the characterization of bistability. For simplicity we consider equidistant step size $h = \frac{T}{N}$ for $N \in \mathbb{N}$ and the time grid

$$\tau_h = \{t_i = ih \mid i = 0,\dots,N\}.$$

Note that our analysis for onestep methods is not restricted to equidistant time grids (c.f. [3] for the stochastic theta method).

We are concerned with stochastic k-step methods written as

$$\begin{aligned} Y_i &= \tilde{X}_i, \text{ for } i = 0,\dots,k-1, \\ \sum_{j=0}^{k} a_j Y_{i+j-k} &= \Phi_h(t_i, Y_{i-k},\dots,Y_i, (I_\alpha^{t_{i+j-k}})_{\alpha\in\mathcal{A}, j=1,\dots,k}), \text{ for } i = k,\dots,N, \end{aligned} \tag{1.3}$$

where $a_j \in \mathbb{R}$, $a_k \neq 0$ and the initial values $\tilde{X}_i$, $i = 0,\dots,k-1$, are $\mathcal{F}_{t_i}$-measurable, square integrable random variables. In order to compute the approximation Y_i of the solution $X(t_i)$ the increment function Φ_h depends on the time $t_i \in \tau_h$, a family of stochastic increments $(I_\alpha^{t_{i+j-k}})_{\alpha\in\mathcal{A},j=1,\dots,k}$, the k predecessors of Y_i and, in the case of an implicit multistep method, it also depends on Y_i itself. In the next section we give more details on Φ_h.

A special case of a k-step method is the Euler-Maruyama scheme: $k = 1$,

$$\begin{aligned} Y_0 &= X_0, \\ Y_i - Y_{i-1} &= hb^0(t_{i-1}, Y_{i-1}) + \sum_{r=1}^{m} b^r(t_{i-1}, Y_{i-1}) I_{(r)}^{t_i}, \text{ for } i = 1, \dots, N, \end{aligned}$$

with the stochastic increments $I_{(r)}^{t_i} = W^r(t_i) - W^r(t_{i-1})$. In [16] it is shown that the Euler-Maruyama scheme converges at least with order $\gamma = \frac{1}{2}$ in the strong sense, i.e. there exists a constant $C > 0$ such that

$$\left(\mathbb{E}\left(\max_{0\le i\le N} |X(t_i) - Y_i|^2\right)\right)^{\frac{1}{2}} \le Ch^\gamma,$$

where X is the unique solution to (1.1). In [6] J.M.C. Clark and R.J. Cameron have shown that, in general, $\gamma = \frac{1}{2}$ is also the maximum rate of convergence for the Euler-Maruyama scheme (and for any method which only uses the Brownian motion at grid points).

In order to derive similar results for k-step methods we write (1.3) as an operator equation $A_h X_h = 0$, where the operator A_h acts on the set of adapted grid functions. This is done for the general form (1.3) and for the stochastic theta method, the Itô-Taylor schemes and BDF2-Maruyama in Section 2. Now, the strong convergence is written in terms of the norm

$$\|Y_h\|_{0,h} = \left(\mathbb{E}\left(\max_{0\le i\le N} |Y_h(t_i)|^2\right)\right)^{\frac{1}{2}}. \tag{1.4}$$

On the other side, the local truncation error is measured by the following stochastic version of Spijker's norm

$$\|Y_h\|_{-1,h} = \sum_{j=0}^{k-1} \|Y_h(t_j)\|_{L^2(\Omega)} + \left(\mathbb{E}\left(\max_{k\le i\le N} \left|\sum_{j=k}^{i} Y_h(t_j)\right|^2\right)\right)^{\frac{1}{2}}. \tag{1.5}$$

In the analysis of deterministic multistep methods it is well-known that Spijker's norm leads to optimal stability properties [10]. In this paper we will show, that under some conditions the following bistability inequality

$$C_1\|A_hY_h - A_hZ_h\|_{-1,h} \le \|Y_h - Z_h\|_{0,h} \le C_2\|A_hY_h - A_hZ_h\|_{-1,h} \tag{1.6}$$

is equivalent to Dahlquist's strong root condition. We refer to Section 3 for a precise formulation of our results and to Section 4 for the proof of the bistability inequality.

If we apply the bistability inequality to the restriction $r_h^E X$ of the unique solution X to the time grid τ_h and the grid function X_h, which is generated by the k-step method (1.3), i.e. $A_h X_h = 0$, we obtain the two-sided error estimate

$$C_1\|A_h r_h^E X\|_{-1,h} \le \|r_h^E X - X_h\|_{0,h} \le C_2\|A_h r_h^E X\|_{-1,h}.$$

In Section 5 we derive upper bounds for the local truncation error $\|A_h r_h^E X\|_{-1,h}$ of the stochastic theta method, the higher order Itô-Taylor schemes and the BDF2-Maruyama scheme in terms of the step size h. In Section 6 we use the left-hand side of the two-sided error estimate to discuss the maximum order of convergence for these k-step methods.

2. Numerical schemes. In this section we rewrite the general k-step method (1.3) as an operator equation $A_h X_h = 0$ and introduce the corresponding spaces and norms. The operator formulation is motivated by the discrete approximation theory [29]. Our notion of consistency, stability and convergence will be formulated in terms of the operator A_h. At the end of this section we present some well-known numerical schemes, which will be analysed in more detail in the sequel of this paper.

Given a time grid τ_h we define the set $\mathcal{G}_h := \mathcal{G}(\tau_h, L^2(\Omega, \mathcal{F}, \mathbb{R}^d))$ to be the space of all adapted and $L^2(\Omega) := L^2(\Omega, \mathcal{F}, P; \mathbb{R}^d)$-valued grid functions. That is, for $Y_h \in \mathcal{G}_h$, the random variables $Y_h(t_i)$ are square-integrable and $\mathcal{F}_{t_i}$-measurable for all $t_i \in \tau_h$. Next, we endow $\mathcal{G}_h$ with the norms (1.4) and (1.5) and we denote the Banach spaces $(\mathcal{G}_h, \| \cdot \|_{0,h})$ and $(\mathcal{G}_h, \| \cdot \|_{-1,h})$ by E_h and F_h, respectively.

Now the operator $A_h : E_h \to F_h$ representing the k-step method (1.3) is given by

$$[A_h Y_h](t_i) = Y_h(t_i) - \tilde{X}_i \tag{2.1}$$

for $0 \leq i \leq k-1$ and by

$$\begin{aligned} [A_h Y_h](t_i) = &\sum_{j=0}^{k} a_j Y_h(t_{i+j-k}) \\ &- \Phi_h(t_i, Y_h(t_{i-k}), \ldots, Y_h(t_i), (I_\alpha^{t_{i+j-k}})_{\alpha \in \mathcal{A}, j=1,\ldots,k}) \end{aligned} \tag{2.2}$$

for $k \leq i \leq N$ and $Y_h \in E_h$. Please note, that the initial values $\tilde{X}_i \in L^2(\Omega, \mathcal{F}_{t_i}, P; \mathbb{R}^d)$ of the k-step method are incorporated into the definition of A_h. Clearly, if a grid function X_h is generated by the k-step method (1.3) then $A_h X_h = 0$.

Next, we turn to the increment function Φ_h and to the stochastic increments $(I_\alpha^{t_{i+j-k}})_{\alpha \in \mathcal{A}, j=1,\ldots,k}$. Let $\mathcal{A}$ be a nonempty, finite set of multi-indices $\alpha = (j_1, \ldots, j_\ell)$, where $j_i \in \{0, \ldots, m\}$ for $i = 1, \ldots, \ell$. By $\ell = \ell(\alpha) \in \mathbb{N}$ we denote the length of α. For $\alpha = (j_1, \ldots, j_\ell) \in \mathcal{A}$ the stochastic increment $I_\alpha^{t_i}$ is given by the ℓ-fold iterated stochastic Itô-integral

$$I_\alpha^{t_i} = \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s_1} \cdots \int_{t_{i-1}}^{s_{\ell-1}} dW^{j_1}(s_\ell) \cdots dW^{j_\ell}(s_1),$$

with $dW^0(s) = ds$. For example, we have $I_{(0)}^{t_i} = t_i - t_{i-1} = h$ and $I_{(r)}^{t_i} = W^r(t_i) - W^r(t_{i-1}) \in \mathcal{F}_{t_i}$ for $r > 0$.

For A_h to be well-defined the increment function Φ_h needs to satisfy

$$\Phi_h(t_i, Y_h(t_{i-k}), \ldots, Y_h(t_i), (I_\alpha^{t_{i+j-k}})_{\alpha \in \mathcal{A}, j=1,\ldots,k}) \in L^2(\Omega, \mathcal{F}_{t_i}, P; \mathbb{R}^d) \tag{2.3}$$

for all $Y_h \in E_h$ and $t_i \in \tau_h$. In the following we will introduce three different numerical schemes and show that (2.3) is fulfilled in each case.

Example (Stochastic theta method). Let $\theta \in [0,1]$. For a time grid τ_h the stochastic theta method (STM) is given by the recursion

$$\begin{aligned} Y_0 &= \tilde{X}_0, \\ Y_i - Y_{i-1} &= h\left((1-\theta) b^0(t_{i-1}, Y_{i-1}) + \theta b^0(t_i, Y_i)\right) + \sum_{r=1}^{m} b^r(t_{i-1}, Y_{i-1}) I_{(r)}^{t_i}, \end{aligned} \tag{2.4}$$

for $1 \leq i \leq N$.

Obviously, the STM is a onestep method ($k = 1$) and one can choose $\mathcal{A} := \{(r) \,|\, r = 0, \dots, m\}$. For a given grid function $Y_h \in E_h$ the corresponding increment function Φ_h^{STM} is defined by

$$\begin{aligned}\Phi_h^{STM}&(t_i, Y_h(t_{i-1}), Y_h(t_i), (I_\alpha^{t_i})_{\alpha\in\mathcal{A}})\\ &= h\left((1-\theta)b^0(t_{i-1}, Y_h(t_{i-1})) + \theta b^0(t_i, Y_h(t_i))\right) + \sum_{r=1}^m b^r(t_{i-1}, Y_h(t_{i-1}))I_{(r)}^{t_i}.\end{aligned}$$

By assumption (A2) the random variable $\Phi_h^{STM}(t_i, Y_h(t_{i-1}), Y_h(t_i), (I_\alpha^{t_i})_{\alpha\in\mathcal{A}})$ is square-integrable and $\mathcal{F}_{t_i}$-measurable.

For the choice $\theta = 0$ one gets the classic Euler-Maruyama scheme. Unlike the deterministic case, the STM converges in general for every choice of θ with the order $\gamma = \frac{1}{2}$ (see the next section). An important application of the STM is the approximation of stiff stochastic differential equations (see [12]). □

Example (BDF2-Maruyama). As a prototype for drift-linear k-step methods we consider the BDF2-Maruyama scheme [4] which is given by

$$\begin{aligned}Y_0 &= \tilde{X}_0, \qquad Y_1 = \tilde{X}_1,\\ Y_i - \frac{4}{3}Y_{i-1} + \frac{1}{3}Y_{i-2} &= h\frac{2}{3}b^0(t_i, Y_i) + \sum_{r=1}^m b^r(t_{i-1}, Y_{i-1})I_{(r)}^{t_i}\\ &\quad - \frac{1}{3}\sum_{r=1}^m b^r(t_{i-2}, Y_{i-2})I_{(r)}^{t_{i-1}}, \qquad 2 \le i \le N.\end{aligned} \tag{2.5}$$

As before, one can choose $\mathcal{A} := \{(r) \,|\, r = 0, \dots, m\}$. The increment function Φ_h^{BDF} of the 2-step method takes the form

$$\begin{aligned}\Phi_h^{BDF}&(t_i, Y_h(t_{i-2}), Y_h(t_{i-1}), Y_h(t_i), (I_\alpha^{t_{i+j-2}})_{\alpha\in\mathcal{A}, j=1,2})\\ &= h\frac{2}{3}b^0(t_i, Y_h(t_i)) + \sum_{r=1}^m b^r(t_{i-1}, Y_h(t_{i-1}))I_{(r)}^{t_i} - \frac{1}{3}\sum_{r=1}^m b^r(t_{i-2}, Y_h(t_{i-2}))I_{(r)}^{t_{i-1}}\end{aligned}$$

for grid functions $Y_h \in E_h$ and all $t_i \in \tau_h$. Again, by the linear growth condition (A2), the random variable $\Phi_h^{BDF}(t_i, Y_h(t_{i-2}), Y_h(t_{i-1}), Y_h(t_i), (I_\alpha^{t_{i+j-2}})_{\alpha\in\mathcal{A}, j=1,2})$ is square-integrable and $\mathcal{F}_{t_i}$-measurable. Hence, the associated operator $A_h^{BDF} : E_h \to F_h$ is well-defined.

It turns out that the BDF2-Maruyama scheme also converges with the strong order $\gamma = \frac{1}{2}$. In the deterministic case, linear multistep methods usually are of higher order than the Euler method. Therefore, one expects a better approximation of the dominating drift term in systems with small noise and the approximation error is significantly smaller than the error of the Euler-Maruyama scheme. We refer to [4] for a detailed discussion. □

Now we turn to the higher order Itô-Taylor schemes which are based on an iterated application of Itô's formula to the integrands of (1.2), provided that all appearing integrals and derivatives exist. We refer to the books [16, 19, 20] for a rigorous derivation.

Example (Itô-Taylor scheme). As in [16], for $\gamma \in \{\frac{n}{2} \,|\, n \in \mathbb{N}\}$, we consider the finite set of multi-indices

$$\mathcal{A}_\gamma = \left\{\alpha = (j_1, \dots, j_\ell) \,|\, 1 \le \ell(\alpha) + n(\alpha) \le 2\gamma \text{ or } \ell(\alpha) = n(\alpha) = \gamma + \frac{1}{2}\right\},$$

where we write $n(\alpha) \in \mathbb{N}$ for the number of components of α which are equal to 0. The Itô-Taylor scheme of order γ is given by

$$\begin{aligned} Y_0 &= \tilde{X}_0, \\ Y_i - Y_{i-1} &= \textstyle\sum_{\alpha \in \mathcal{A}_\gamma} f_\alpha(t_{i-1}, Y_{i-1}) I_\alpha^{t_i}, \quad 1 \leq i \leq N. \end{aligned} \tag{2.6}$$

Here, for $\alpha = (j_1, \dots, j_\ell)$, the coefficient functions $f_\alpha : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ are defined by

$$f_\alpha(t,x) = (L^{j_1} \cdots L^{j_\ell} f)(t,x),$$

where $f : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ is the projection with respect to the second coordinate, i.e. $f(t,x) = x$. The L^r are differential operators of the form

$$\begin{aligned} L^0 &= \frac{\partial}{\partial t} + \sum_{i=1}^d b^{0,i} \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d \sum_{r=1}^m b^{r,i} b^{r,j} \frac{\partial^2}{\partial x_i \partial x_j}, \\ L^r &= \sum_{i=1}^d b^{r,i} \frac{\partial}{\partial x_i}, \quad r = 1, \dots, m, \end{aligned}$$

where $b^{r,i}$ denotes the i-th component of the coefficient function b^r for $i = 1, \dots, d$ and $r = 0, \dots, m$.

If we choose $\gamma = \frac{1}{2}$ then the set $\mathcal{A}_{\frac{1}{2}}$ consists of all multi-indices of length 1, i.e. $\mathcal{A}_{\frac{1}{2}} = \{(0), (1), \dots, (m)\}$, and the coefficient functions f_α simplify to the drift and diffusion coefficient functions of the SODE (1.1). Thus the Itô-Taylor scheme of order $\frac{1}{2}$ is the well-known Euler-Maruyama scheme. One also easily checks that the choice $\gamma = 1$ leads to the Milstein method.

The associated increment function Φ_h^{ITS} is given by

$$\Phi_h^{ITS}(t_i, Y_h(t_{i-1}), Y_h(t_i), (I_\alpha^{t_i})_{\alpha \in \mathcal{A}_\gamma}) = \textstyle\sum_{\alpha \in \mathcal{A}_\gamma} f_\alpha(t_{i-1}, Y_h(t_{i-1})) I_\alpha^{t_i}$$

where $Y_h \in E_h$. Under the following additional assumption the increment function is well-defined:

(A3) The assumptions of Theorem 5.5.1 in [16], (i.e. the coefficient functions b^r of the SODE (1.1) are sufficiently smooth such that the functions f_α and the Itô-Taylor expansion exists up to the order γ) are satisfied and for all $\alpha \in \mathcal{A}_\gamma$ there exists a constant $L_\alpha > 0$ such that

$$|f_\alpha(t,x) - f_\alpha(t,y)| \leq L_\alpha |x - y| \text{ and } |f_\alpha(t,x)| \leq L_\alpha (1 + |x|)$$

for all $x, y \in \mathbb{R}^d$ and $t \in [0,T]$.

In practice it can be costly to simulate the iterated stochastic increments $I_\alpha^{t_i}$. This may outweight the advantage of the higher order of convergence. However, in many important applications the diffusion coefficients have some special properties which allow to simplify the Itô-Taylor schemes in a way that the use of iterated stochastic integrals can be avoided. We refer to the corresponding discussions in [16] and [22, 30].
□

3. Definitions and main results. In this section we introduce our notions of consistency and (numerical) bistability of a multistep method which are motivated by the work of Stummel [29]. For a comparison to related notions in the literature and for a more detailed embedding into the abstract theory of discrete approximations we refer to [3] and [17], respectively. In the second part of this section we give a precise formulation of our assumptions, the characterization of the bistability of a multistep method and the two-sided error estimates. We start with the definition of a consistent multistep method.

DEFINITION 3.1. *The multistep method $(A_h)_{h>0}$ is called consistent of order $\gamma > 0$, if there exist a constant $C > 0$ and an upper step size bound $\overline{h} > 0$, such that the estimate*

$$\|A_h r_h^E X\|_{-1,h} \leq Ch^\gamma \tag{3.1}$$

holds for all grids τ_h with $h \leq \overline{h}$, where $r_h^E X$ denotes the restriction of the analytic solution X of (1.1) to the time grid τ_h.

The left-hand side of (3.1) is called *local truncation error* or *consistency error* and uses our stochastic version of Spijker's norm (1.5). The standard procedure of eliminating convergence errors by successive triangle inequalities from local errors (see "Lady Windemere's fan" diagram in [11]) is not sharp enough to produce two-sided error estimates. Next, we come to the definition of bistability.

DEFINITION 3.2. *The multistep method $(A_h)_{h>0}$ is called bistable with respect to the norms $\|\cdot\|_{0,h}$, $\|\cdot\|_{-1,h}$, if there exist constants $C_1, C_2 > 0$ and an upper step size bound $\overline{h} > 0$, such that the operators $A_h : E_h \to F_h$ are bijective and the estimate*

$$C_1\|A_h Y_h - A_h Z_h\|_{-1,h} \leq \|Y_h - Z_h\|_{0,h} \leq C_2\|A_h Y_h - A_h Z_h\|_{-1,h} \tag{3.2}$$

holds for all $Y_h, Z_h \in E_h$ and all grids τ_h with $h \leq \overline{h}$.

If only the right-hand side inequality in (3.2) is true we say that the multistep method is *stable*. As we will see below, consistency and stability are sufficient for the convergence of the multistep method $(A_h)_{h>0}$.

DEFINITION 3.3. *The multistep method $(A_h)_{h>0}$ is called convergent of order $\gamma > 0$, if there exist a constant $C > 0$ and an upper step size bound $\overline{h} > 0$, such that the operators $A_h : E_h \to F_h$ are bijective and the estimate*

$$\|X_h - r_h^E X\|_{0,h} \leq Ch^\gamma$$

holds for all time grids τ_h with $h \leq \overline{h}$. Here X_h and $r_h^E X$ denote the solution to $A_h X_h = 0$ and the restriction of the analytic solution X to the time grid τ_h, respectively.

A bistable multistep method can be characterized by Dahlquist's strong root condition. The characteristic polynomial ρ of the k-step method (1.3) is given by

$$\rho(z) = \sum_{j=0}^{k} a_j z^j, \quad z \in \mathbb{C}.$$

The strong root condition reads as follows:

Strong root condition If $z \in \mathbb{C}$ with $\rho(z) = 0$, then either $|z| < 1$ or $z = 1$ is a simple root of ρ.

In [4] the authors showed for a different pair of norms that the usual root condition (all roots of ρ lie within the unit circle and all roots with modulus 1 are of multiplicity 1) is necessary and sufficient for the stability of a stochastic multistep method. But, as we will see in the next section, the usual root condition is not sharp enough to characterize bistability.

For our stability theorem we also need the following Lipschitz-type assumptions on the increment function Φ_h.

(S1) There exists $L > 0$ such that for all $j = k, \dots, N$, $Y_h \in \mathcal{G}_h$ and $Z \in L^2(\Omega, \mathcal{F}_{t_j}, P; \mathbb{R}^d)$

$$\begin{aligned}&\Big\| \Phi_h(t_j, Y_h(t_{j-k}), \dots, Y_h(t_{j-1}), Y_h(t_j), (I_\alpha^{t_{j+i-k}})_{\alpha \in \mathcal{A}, i=1,\dots,k}) \\ &\qquad - \Phi_h(t_j, Y_h(t_{j-k}), \dots, Y_h(t_{j-1}), Y_h(t_j) + Z, (I_\alpha^{t_{j+i-k}})_{\alpha \in \mathcal{A}, i=1,\dots,k}) \Big\|_{L^2(\Omega)} \\ &\le Lh \|Z\|_{L^2(\Omega)}.\end{aligned}$$

(S2) There exists $L > 0$ such that for all $j = k, \dots, N$, $Y_h, Z_h \in \mathcal{G}_h$

$$\begin{aligned}&\mathbb{E}\Bigg(\max_{k \le i \le j} \Big| \textstyle\sum_{\eta=k}^{i} \Big[\Phi_h(t_\eta, Y_h(t_{\eta-k}), \dots, Y_h(t_\eta), (I_\alpha^{t_{\eta+l-k}})_{\alpha \in \mathcal{A}, l=1,\dots,k}) \\ &\qquad\qquad - \Phi_h(t_\eta, Z_h(t_{\eta-k}), \dots, Z_h(t_\eta), (I_\alpha^{t_{\eta+l-k}})_{\alpha \in \mathcal{A}, l=1,\dots,k}) \Big] \Big|^2 \Bigg) \\ &\le Lh \sum_{\eta=0}^{j} \mathbb{E}\left(\max_{0 \le i \le \eta} |Y_h(t_i) - Z_h(t_i)|^2 \right).\end{aligned}$$

Now we are in the position to formulate our first main result.

Theorem 3.4 (Characterization of bistability). *Assume that the multistep method $(A_h)_{h>0}$ satisfies $\rho(1) = 0$, $a_k \neq 0$ and the Lipschitz assumptions (S1), (S2). Then*

$$(A_h)_{h>0} \text{ is bistable}$$

if and only if

$$(A_h)_{h>0} \text{ satisfies the strong root condition.}$$

The proof of theorem 3.4 is deferred to Section 4. The next theorem makes use of the bistability inequality (3.2).

Theorem 3.5. *Assume that the multistep method $(A_h)_{h>0}$ is bistable. Then for $\gamma > 0$*

$$(A_h)_{h>0} \text{ is consistent of order } \gamma$$

if and only if

$$(A_h)_{h>0} \text{ is convergent of order } \gamma.$$

Moreover, there exist constants $C_1, C_2 > 0$ and an upper step size bound $\overline{h} > 0$ such that the two-sided error estimate

$$C_1 \|A_h r_h^E X\|_{-1,h} \le \|X_h - r_h^E X\|_{0,h} \le C_2 \|A_h r_h^E X\|_{-1,h} \tag{3.3}$$

holds for all $h < \overline{h}$, where $X_h \in E_h$ solves $A_h X_h = 0$ and $r_h^E X$ denotes the restriction of the analytic solution X to the time grid τ_h.

Proof. Since $(A_h)_{h>0}$ is bistable there exist an upper step size bound $\overline{h} > 0$ such that the operators $A_h : E_h \to F_h$ are bijective for all $h < \overline{h}$. Thus, there exists a unique grid function $X_h \in E_h$ such that $A_h X_h = 0$. Applying the bistability inequality (3.2) to X_h and the restriction $r_h^E X$ yields the two-sided error estimate (3.3). The first statement of the theorem is now evident. □

The rest of this section is devoted to the three approximation schemes which were introduced in Section 2. The first theorem is concerned with the bistability of these methods and will also be proved in the next section.

THEOREM 3.6.

(i) *Under the assumptions (A1) and (A2) the stochastic theta method and the BDF2-Maruyama scheme are bistable.*

(ii) *Under the assumptions (A1), (A2) and (A3) the Itô-Taylor schemes are bistable.*

The next theorem deals with the consistency of the approximation schemes and is based on the following additional assumptions. Here we use the notation of the remainder set $\mathcal{B}(\mathcal{A}_\gamma)$ of the Itô-Taylor expansion (c.f. [16]) which is given by

$$\mathcal{B}(\mathcal{A}_\gamma) = \{\alpha = (j_1, j_2, \dots, j_\ell) \,|\, j_1 = 0, \dots, m, \, \alpha \notin \mathcal{A}_\gamma, \, (j_2, \dots, j_\ell) \in \mathcal{A}_\gamma\}.$$

(C1) The initial values are consistent of order γ, i.e. there exist a constant $C > 0$ and $\overline{h} > 0$ such that for all $h \le \overline{h}$

$$\max_{0 \le i \le k-1} \|X(t_i) - \tilde{X}_i\|_{L^2(\Omega)} \le C h^\gamma.$$

(C2) There exists a constant $K > 0$ such that

$$|b^r(t,x) - b^r(s,x)| \le K(1 + |x|)\sqrt{|t-s|}$$

for all $x \in \mathbb{R}^d$, $t, s \in [0,T]$.

(C3) For all $\alpha \in \mathcal{B}(\mathcal{A}_\gamma)$ we have

$$\int_0^T \mathbb{E}\left(|f_\alpha(s, X(s))|^2\right) ds < \infty.$$

The assumption (C2) is already used in [16] to prove convergence of the Euler-Maruyama scheme. The assumption (C3) is fulfilled if all coefficient functions f_α, $\alpha \in \mathcal{B}(A_\gamma)$, satisfy a linear growth condition. Now we formulate the consistency theorem.

THEOREM 3.7.

(i) *Under the assumptions (A1), (A2), (C1) and (C2) the stochastic theta method and BDF2-Maruyama are consistent of order* $\gamma = \frac{1}{2}$.

(ii) *Under the assumptions (A1), (A2), (A3), (C1) and (C3) the Itô-Taylor scheme of order* γ *is consistent of order* γ.

The proof is deferred to Section 5. From theorems 3.5, 3.6 and 3.7 one immediately obtains the following result:

COROLLARY 3.8.

(i) *Under the assumptions (A1), (A2), (C1) and (C2) the stochastic theta method and BDF2-Maruyama are convergent of order* $\gamma = \frac{1}{2}$.

(ii) *Under the assumptions (A1), (A2), (A3), (C1) and (C3) the Itô-Taylor scheme of order* γ *is convergent of order* γ.

Moreover, in both cases, the two-sided error estimate (3.3) *is valid.*

We close this section with the following *remarks.*

1. We remark that by our choice of the norm (1.4) the convergence in definition 3.3 and corollary 3.8 is understood in the L^2-sense. In particular, the numerical solution X_h of the equation $A_h X_h = 0$ converges uniformly at each grid point to the restriction of the analytic solution X. The L^2-convergence implies a good pathwise approximation for each sample path $\omega \in \Omega$. In addition to this notion of strong convergence we mention the concepts of (numerical) weak convergence (see [16, 19, 20]) and of pathwise convergence (see [13, 14, 15]) which. however, are not considered in this paper.

2. One can use a slightly different pair of norms where the maximum occurs outside the expectation, i.e. $|||V_h|||_{0,h} := \max_{t_i \in \tau_h} \|X_h(t_i) - X(t_i)\|_{L^2(\Omega)}$, and obtains similar results. For the stochastic theta method a proof is given in [3].

3. In our approach we work with grid functions only. According to [16, Ch.10.6] one can interpolate the numerical approximation to an adapted, right-continuous stochastic procress $X_h : [0, T] \to \mathbb{R}^d$ with existing left limits (cadlag) such that $X_h(t)$ converges uniformly in t to the analytic solution $X(t)$ with the same order that holds at the grids points. □

4. Characterization of bistability. In this section we prove the Theorems 3.4 and 3.6. The proofs are done in several steps, each in an own subsection. First we show that the numerical schemes form Section 2 fulfill the strong root condition and the stability assumptions (S1), (S2). Hence Theorem 3.6 directly follows from Theorem 3.4.

Next we show that the operator A_h of the general k-step method (1.3) is invertible under assumption (S1). In the third subsection we write the k-step method as a sum of a linear operator and the increment function. We show that the k-step method is bistable if and only if the linear operator is bistable. Finally, in the last subsection, we show that the linear operator is bistable if and only if Dahlquist's strong root condition is satisfied.

In the last two subsections we apply techniques used by R.D. Grigorieff [10] for a similar analysis of deterministic multistep methods.

4.1. Proof of Theorem 3.6. In this subsection we show that under the given assumptions the stochastic theta method, the BDF2-Maruyama scheme and the Itô-Taylor scheme of order γ satisfy the stability assumptions (S1), (S2) and the strong root condition. Thus Theorem 3.6 follows from Theorem 3.4.

By the definitions of the operators A_h^{STM}, A_h^{BDF} and A_h^{ITS} the conditions $\rho(1) = 0$, $a_k \neq 0$ and the strong root condition are satisfied in each case (the roots of the characteristic polynomial of the BDF2-Maruyama scheme are $z_1 = 1$, $z_2 = \frac{1}{3}$). Thus it remains to prove the stability assumptions (S1),(S2).

First we do this for the stochastic theta method (2.4). Let $Y_h \in \mathcal{G}_h$, $j = 1, \dots, N$ and $Z \in L^2(\Omega, \mathcal{F}_{t_j}, P; \mathbb{R}^d)$, then by the Lipschitz-assumption (A2)

$$\begin{aligned}
&\Big\| \Phi_h^{STM}(t_j, Y_h(t_{j-1}), Y_h(t_j), (I_{(r)}^{t_j})_{r=0,\dots,m}) \\
&\qquad\qquad - \Phi_h^{STM}(t_j, Y_h(t_{j-1}), Y_h(t_j) + Z, (I_{(r)}^{t_j})_{r=0,\dots,m}) \Big\|_{L^2(\Omega)} \\
&= \left\| h\theta \left(b^0(t_j, Y_h(t_j)) - b^0(t_j, Y_h(t_j) + Z) \right) \right\|_{L^2(\Omega)} \\
&\leq Lh \|Z\|_{L^2(\Omega)},
\end{aligned}$$

where $L = \theta K$. This proves (S1) for the stochastic theta method. By the inequality $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$ we obtain for (S2)

$$
\begin{aligned}
&\mathbb{E}\left(\max_{1\leq i\leq j}\left|\sum_{\eta=1}^{i}\Big(\Phi_h^{STM}(t_\eta, Y_h(t_{\eta-1}), Y_h(t_\eta), (I_{(r)}^{t_\eta})_{r=0,\dots,m})\right.\right.\\
&\qquad\qquad\left.\left. -\Phi_h^{STM}(t_\eta, Z_h(t_{\eta-1}), Z_h(t_\eta), (I_{(r)}^{t_\eta})_{r=0,\dots,m})\Big)\right|^2\right)\\
&\leq 3\mathbb{E}\left(\max_{1\leq i\leq j}\left|\sum_{\eta=1}^{i} h(1-\theta)\left(b^0(t_{\eta-1}, Y_h(t_{\eta-1})) - b^0(t_{\eta-1}, Z_h(t_{\eta-1}))\right)\right|^2\right)\\
&+ 3\mathbb{E}\left(\max_{1\leq i\leq j}\left|\sum_{\eta=1}^{i} h\theta\left(b^0(t_\eta, Y_h(t_\eta)) - b^0(t_\eta, Z_h(t_\eta))\right)\right|^2\right)\\
&+ 3\mathbb{E}\left(\max_{1\leq i\leq j}\left|\sum_{\eta=1}^{i}\sum_{r=1}^{m}\left(b^r(t_{\eta-1}, Y_h(t_{\eta-1})) - b^r(t_{\eta-1}, Z_h(t_{\eta-1}))\right) I_{(r)}^{t_\eta}\right|^2\right)\\
&=: T_1 + T_2 + T_3.
\end{aligned}
$$

We estimate the three summands separately. For T_1 Jensen's inequality and the Lipschitz-assumption (A2) yield

$$
\begin{aligned}
T_1 &\leq 3\mathbb{E}\left(\max_{1\leq i\leq j} ih^2(1-\theta)^2\sum_{\eta=1}^{i}\left|b^0(t_{\eta-1}, Y_h(t_{\eta-1})) - b^0(t_{\eta-1}, Z_h(t_{\eta-1}))\right|^2\right)\\
&\leq 3(1-\theta)^2 T\sum_{\eta=1}^{j} h\mathbb{E}\left(\left|b^0(t_{\eta-1}, Y_h(t_{\eta-1})) - b^0(t_{\eta-1}, Z_h(t_{\eta-1}))\right|^2\right)\\
&\leq 3(1-\theta)^2 TK^2 h\sum_{\eta=1}^{j}\mathbb{E}\left(|Y_h(t_{\eta-1}) - Z_h(t_{\eta-1})|^2\right)\\
&\leq Lh\sum_{\eta=0}^{j}\mathbb{E}\left(\max_{0\leq i\leq\eta}|Y_h(t_i) - Z_h(t_i)|^2\right),
\end{aligned}
$$

where the constant $L > 0$ only depends on θ, T and K. The term T_2 is estimated analogously. By the martingale property of the stochastic Itô-integrals we are allowed to apply Doob's martingale inequality to term T_3. Then we use $\mathbb{E}\left(\left|I_{(r)}^{t_\eta}\right|^2\right) = h$ and finish the estimate by

$$
\begin{aligned}
T_3 &\leq 12\mathbb{E}\left(\left|\sum_{\eta=1}^{j}\sum_{r=1}^{m}\left(b^r(t_{\eta-1}, Y_h(t_{\eta-1})) - b^r(t_{\eta-1}, Z_h(t_{\eta-1}))\right) I_{(r)}^{t_\eta}\right|^2\right)\\
&\leq 12\sum_{\eta=1}^{j}\sum_{r=1}^{m} hK^2\mathbb{E}\left(|Y_h(t_{\eta-1}) - Z_h(t_{\eta-1})|^2\right)\\
&\leq Lh\sum_{\eta=0}^{j}\mathbb{E}\left(\max_{0\leq i\leq\eta}|Y_h(t_i) - Z_h(t_i)|^2\right).
\end{aligned}
$$

Here the constant $L > 0$ depends on m and K. Altogether we have shown that the stochastic theta method satisfies assumption (S2).

Since the BDF2-Maruyama scheme can be written as a sum of two stochastic theta methods with different parameter values θ the proof of (S1), (S2) is basically the same as above and therefore skipped.

The Itô-Taylor schemes are explicit onestep methods and (S1) is clearly satisfied. It remains to prove (S2) for the Itô-Taylor scheme of order γ. For $Y_h, Z_h \in \mathcal{G}_h$ and $j = 1, \ldots, N$ we compute

$$\begin{aligned}
&\mathbb{E}\Bigg(\max_{1\le i\le j}\Bigg|\sum_{\eta=1}^{i}\Big(\Phi_h^{ITS}(t_\eta, Y_h(t_{\eta-1}), Y_h(t_\eta), (I_\alpha^{t_\eta})_{\alpha\in\mathcal{A}_\gamma})\\
&\qquad\qquad -\Phi_h^{ITS}(t_\eta, Z_h(t_{\eta-1}), Z_h(t_\eta), (I_\alpha^{t_\eta})_{\alpha\in\mathcal{A}_\gamma})\Big)\Bigg|^2\Bigg)\\
&= \mathbb{E}\left(\max_{1\le i\le j}\left|\sum_{\eta=1}^{i}\sum_{\alpha\in\mathcal{A}_\gamma}\left[f_\alpha(t_{\eta-1}, Y_h(t_{\eta-1})) - f_\alpha(t_{\eta-1}, Z_h(t_{\eta-1}))\right] I_\alpha^{t_\eta}\right|^2\right)\\
&\le |\mathcal{A}_\gamma| \sum_{\alpha\in\mathcal{A}_\gamma}\mathbb{E}\left(\max_{1\le i\le j}\left|\sum_{\eta=1}^{i}\left[f_\alpha(t_{\eta-1}, Y_h(t_{\eta-1})) - f_\alpha(t_{\eta-1}, Z_h(t_{\eta-1}))\right] I_\alpha^{t_\eta}\right|^2\right).
\end{aligned}$$

Since $|A_\gamma| < \infty$ it is sufficient to estimate each summand separately. For all multi-indices $\alpha \in \mathcal{A}_\gamma$ of the form $\alpha = (0, \ldots, 0)$, i.e. $\ell(\alpha) = n(\alpha)$, we have $I_\alpha^{t_\eta} = \frac{1}{\ell(\alpha)!} h^{\ell(\alpha)}$. In this case we apply Jensen's inequality and the Lipschitz-assumption (A3) and estimate the summand by

$$\begin{aligned}
&\mathbb{E}\left(\max_{1\le i\le j}\left|\sum_{\eta=1}^{i}\left[f_\alpha(t_{\eta-1}, Y_h(t_{\eta-1})) - f_\alpha(t_{\eta-1}, Z_h(t_{\eta-1}))\right] I_\alpha^{t_\eta}\right|^2\right)\\
&\le \frac{T}{(\ell(\alpha)!)^2} h^{2\ell(\alpha)-1}\sum_{\eta=1}^{j}\mathbb{E}\left(\left|f_\alpha(t_{\eta-1}, Y_h(t_{\eta-1})) - f_\alpha(t_{\eta-1}, Z_h(t_{\eta-1}))\right|^2\right)\\
&\le \frac{T}{(\ell(\alpha)!)^2} L_\alpha h^{2\ell(\alpha)-1}\sum_{\eta=1}^{j}\mathbb{E}\left(\left|Y_h(t_{\eta-1}) - Z_h(t_{\eta-1})\right|^2\right)\\
&\le \frac{T}{(\ell(\alpha)!)^2} L_\alpha h^{2\ell(\alpha)-1}\sum_{\eta=0}^{j}\mathbb{E}\left(\max_{0\le i\le\eta}\left|Y_h(t_i) - Z_h(t_i)\right|^2\right).
\end{aligned}$$

For multi-indices $\alpha \in \mathcal{A}_\gamma$ with $\ell(\alpha) \neq n(\alpha)$ we have $\mathbb{E}(I_\alpha^{t_\eta} | \mathcal{F}_{t_{\eta-1}}) = 0$ with probability 1 (c.f. Lemma 5.7.1 in [16]) and there exists a constant C such that $\mathbb{E}(|I_\alpha^{t_\eta}|^2) \le Ch^{\ell(\alpha)+n(\alpha)}$ (c.f. Lemma 5.7.2 in [16] or Lemma 5.3 below). Hence, under the given assumptions, the stochastic process $(S_i)_{i=0,\ldots,N}$ with

$$S_i := \sum_{\eta=1}^{i}\left(f_\alpha(t_{\eta-1}, Y_h(t_{\eta-1})) - f_\alpha(t_{\eta-1}, Z_h(t_{\eta-1}))\right) I_\alpha^{t_\eta}$$

is a discrete, square-integrable martingale. Once again we apply Doob's martingale inequality and obtain

$$\begin{aligned}
\mathbb{E}\left(\max_{1\le i\le j}|S_i|^2\right) &\le 4\mathbb{E}\left(|S_j|^2\right)\\
&= 4\sum_{\eta=1}^{j}\mathbb{E}\left(\left|\left[f_\alpha(t_{\eta-1},Y_h(t_{\eta-1}))-f_\alpha(t_{\eta-1},Z_h(t_{\eta-1}))\right]I_\alpha^{t_\eta}\right|^2\right)\\
&\le 4CL_\alpha\sum_{\eta=1}^{j}h^{\ell(\alpha)+n(\alpha)}\mathbb{E}\left(|Y_h(t_{\eta-1})-Z_h(t_{\eta-1})|^2\right)\\
&\le 4CL_\alpha h^{\ell(\alpha)+n(\alpha)}\sum_{\eta=0}^{j}\mathbb{E}\left(\max_{0\le i\le \eta}|Y_h(t_i)-Z_h(t_i)|^2\right).
\end{aligned}$$

Since $\ell(\alpha)+n(\alpha)\ge 1$ we have shown (S2) for the Itô-Taylor scheme of order γ.

4.2. Invertibility of A_h. In this subsection we begin the proof of Theorem 3.4 by discussing the invertibility of the operator $A_h : E_h \to F_h$ of the general k-step method (1.3). The following lemma summarizes our result.

LEMMA 4.1. *Under the assumptions (S1) and $a_k \neq 0$ there exists an upper step size bound $\overline{h} > 0$ such that the operators $A_h : E_h \to F_h$ are bijective for all $h < \overline{h}$.*

Proof. Let $Y_h \in F_h$. The equation $A_h X_h = Y_h$ is written in terms of grid functions, hence we have to solve a system of equations of the form

$$[A_h X_h](t_i) = Y_h(t_i) \tag{4.1}$$

for all $t_i \in \tau_h$. We show that this equation is uniquely solvable for $t_i \in \tau_h$ if the solution is already uniquely determined for all $t_j \in \tau_h$ with $j < i$.

For $0 \le i \le k-1$ we have $[A_h X_h](t_i) = X_h(t_i) - \tilde{X}_i$, where $\tilde{X}_i$ denotes the i-th initial value of the multistep method. Hence $X_h(t_i) := \tilde{X}_i + Y_h(t_i) \in L^2(\Omega,\mathcal{F}_{t_i},P;\mathbb{R}^d)$ is the unique solution of (4.1) for $0\le i\le k-1$.

Next assume that for $j \ge k$ a unique and adapted grid function $(X_h(t_i))_{0\le i\le j-1}$ is known such that equation (4.1) holds for all $0 \le i < j$. Now the equation $[A_hX_h](t_j) = Y_h(t_j)$ is equivalently written in fixed point form as

$$X_h(t_j) = F_h(t_j, X_h(t_j)),$$

where $F_h(t_j,\cdot) : L^2(\Omega,\mathcal{F}_{t_j},P;\mathbb{R}^d) \to L^2(\Omega,\mathcal{F}_{t_j},P;\mathbb{R}^d)$ is given by

$$\begin{aligned}
F_h(t_j,Z) = \frac{1}{a_k}\Bigg(&Y_h(t_j) - \sum_{i=0}^{k-1} a_i X_h(t_{j+i-k})\\
&+ \Phi_h(t_j, X_h(t_{j-k}),\dots,X_h(t_{j-1}), Z, (I_\alpha^{t_{j+\eta-k}})_{\alpha\in\mathcal{A},\eta=1,\dots,k})\Bigg)
\end{aligned}$$

for $Z \in L^2(\Omega,\mathcal{F}_{t_j},P;\mathbb{R}^d)$. By assumption (S1) we get

$$\left\|F_h(t_j,Z)-F_h(t_j,\tilde{Z})\right\|_{L^2(\Omega)} \le Lh\frac{1}{a_k}\left\|Z-\tilde{Z}\right\|_{L^2(\Omega)}.$$

Hence, for h small enough, $F_h(t_j,\cdot)$ is a contraction in $L^2(\Omega,\mathcal{F}_{t_j},P;\mathbb{R}^d)$ and there exists a unique fixed point, which we denote by $X_h(t_j)$.

By induction we obtain a unique and adapted grid function X_h on the whole time grid τ_h which solves $A_hX_h = Y_h$. Therefore the operator A_h is invertible under assumption (S1). □

4.3. Reduction to the linear part. An important step for the characterization of a bistable multistep method is to realize that the bistability only depends on the linear part of the operator A_h as long as the remainder part satisfies a Lipschitz condition. By the linear part we mean the operator $L_h : E_h \to F_h$ which is given by

$$[L_hY_h](t_i) = \begin{cases} Y_h(t_i), & \text{for } 0 \le i \le k-1, \\ \sum_{j=0}^k a_jY_h(t_{i+j-k}), & \text{for } k \le i \le N. \end{cases} \tag{4.2}$$

The residual operator is denoted by $T_h := A_h - L_h$. The goal of this subsection is to prove the following lemma which is a generalization of a corresponding result for deterministic multistep methods [10].

Lemma 4.2. *Under the assumptions (S1), (S2) and $a_k \neq 0$ the multistep method $(A_h)_{h>0}$ is bistable if and only if the sequence of operators $(L_h)_{h>0}$ is bistable.*

For the proof we need the following discrete Gronwall-lemma.

Lemma 4.3. *Consider constants $\gamma_1,\gamma_2 \ge 0$ and a real sequence $(x_j)_{j=0,\dots,N}$, $N \in \mathbb{N}$, with*

$$x_j \le \gamma_1 + \gamma_2 \sum_{\eta=0}^{j-1} x_\eta$$

for all $j = 0,\dots,N$. Then $x_j \le \gamma_1 e^{j\gamma_2}$ for all $j = 0,\dots,N$.

Proof of Lemma 4.2. Note that by assumption $a_k \neq 0$ and Lemma 4.1 it is clear that there exists an upper step size bound $\overline{h} > 0$ such that the operators A_h and L_h are both bijective for all $h < \overline{h}$. Hence we only have to show that (3.2) holds for one operator if and only if it holds for the other one.

First we assume that the bistability inequality (3.2) holds for the operator L_h. As a start we prove that the estimate

$$\begin{aligned} \left(\mathbb{E}\left(\max_{0\le i\le j}|Y_h(t_i)-Z_h(t_i)|^2\right)\right)^{\frac{1}{2}} \le C_2 \Bigg[&\sum_{i=0}^{k-1} \|Y_h(t_i)-Z_h(t_i)\|_{L^2(\Omega)} \\ &+ \left(\mathbb{E}\left(\max_{k\le i\le j}\left|\textstyle\sum_{\eta=k}^i (L_hY_h(t_\eta) - L_hZ_h(t_\eta))\right|^2\right)\right)^{\frac{1}{2}}\Bigg] \end{aligned} \tag{4.3}$$

is valid for all $h < \overline{h}$, $Y_h, Z_h \in E_h$ and all $0 \le j \le N$. For the proof we fix a step size $h < \overline{h}$, a grid function $Y_h \in E_h$ and $0 \le j \le N$ arbitrary. For every $Z_h \in E_h$ there exists a unique solution $X_h \in E_h$ to the difference equation

$$[L_hX_h](t_i) = \begin{cases} [L_hZ_h](t_i), & \text{for } 0 \le i \le j, \\ [L_hY_h](t_i), & \text{for } j+1 \le i \le N, \end{cases}$$

since L_h is bijective for all $h < \overline{h}$. As in Subsection 4.2 one shows that $X_h(t_i) = Z_h(t_i)$

for all $i \le j$. By (3.2) we obtain

$$
\begin{aligned}
&\left(\mathbb{E}\left(\max_{0\le i\le j}|Y_h(t_i)-Z_h(t_i)|^2\right)\right)^{\frac{1}{2}}\\
&=\left(\mathbb{E}\left(\max_{0\le i\le j}|Y_h(t_i)-X_h(t_i)|^2\right)\right)^{\frac{1}{2}}\\
&\le \|Y_h-X_h\|_{0,h}\\
&\le C_2\,\|L_hY_h-L_hX_h\|_{-1,h}\\
&=C_2\left[\sum_{i=0}^{k-1}\|Y_h(t_i)-Z_h(t_i)\|_{L^2}+\left(\mathbb{E}\left(\max_{k\le i\le j}\left|\textstyle\sum_{\eta=k}^{i}\left(L_hY_h(t_\eta)-L_hZ_h(t_\eta)\right)\right|^2\right)\right)^{\frac{1}{2}}\right]
\end{aligned}
$$

which proves the estimate (4.3). By inserting $L_h = A_h - T_h$ into (4.3) we get

$$
\begin{aligned}
&\left(\mathbb{E}\left(\max_{0\le i\le j}|Y_h(t_i)-Z_h(t_i)|^2\right)\right)^{\frac{1}{2}}\\
&\le C_2\left[\sum_{i=0}^{k-1}\|Y_h(t_i)-Z_h(t_i)\|_{L^2}\right.\\
&\quad\left.+\left(\mathbb{E}\left(\max_{k\le i\le j}\left|\textstyle\sum_{\eta=k}^{i}\left(A_hY_h(t_\eta)-T_hY_h(t_\eta)-A_hZ_h(t_\eta)+T_hZ_h(t_\eta)\right)\right|^2\right)\right)^{\frac{1}{2}}\right]\\
&\le C_2\left[\|A_hY_h-A_hZ_h\|_{-1,h}+\left(\mathbb{E}\left(\max_{k\le i\le j}\left|\textstyle\sum_{\eta=k}^{i}\left(T_hY_h(t_\eta)-T_hZ_h(t_\eta)\right)\right|^2\right)\right)^{\frac{1}{2}}\right].
\end{aligned}
$$

For the second summand assumption (S2) yields

$$
\begin{aligned}
&\mathbb{E}\left(\max_{k\le i\le j}\left|\textstyle\sum_{\eta=k}^{i}\left(T_hY_h(t_\eta)-T_hZ_h(t_\eta)\right)\right|^2\right)\\
&=\mathbb{E}\left(\max_{k\le i\le j}\left|\sum_{\eta=k}^{i}\Big(\Phi_h(t_\eta,Y_h(t_{\eta-k}),\ldots,Y_h(t_\eta),(I_\alpha^{t_{\eta+l-k}})_{\alpha\in\mathcal{A},l=1,\ldots,k})\right.\right.\\
&\qquad\left.\left.-\Phi_h(t_\eta,Z_h(t_{\eta-k}),\ldots,Z_h(t_\eta),(I_\alpha^{t_{\eta+l-k}})_{\alpha\in\mathcal{A},l=1,\ldots,k})\Big)\right|^2\right)\\
&\le Lh\sum_{\eta=0}^{j}\mathbb{E}\left(\max_{0\le i\le\eta}|Y_h(t_i)-Z_h(t_i)|^2\right).
\end{aligned}
\tag{4.4}
$$

Thus

$$
\begin{aligned}
&(1-2C_2Lh)\mathbb{E}\left(\max_{0\le i\le j}|Y_h(t_i)-Z_h(t_i)|^2\right)\\
&\le 2C_2\left[\|A_hY_h-A_hZ_h\|_{-1,h}^2+Lh\sum_{\eta=0}^{j-1}\mathbb{E}\left(\max_{0\le i\le\eta}|Y_h(t_i)-Z_h(t_i)|^2\right)\right].
\end{aligned}
$$

From Lemma 4.3 and for all $h < \min(\overline{h}, \frac{1}{4C_2L})$ we derive the estimate

$$\mathbb{E}\left(\max_{0\leq i\leq j} |Y_h(t_i) - Z_h(t_i)|^2\right) \leq \frac{2C_2}{1-2C_2Lh} \|A_hY_h - A_hZ_h\|_{-1,h}^2 e^{\frac{j2C_2Lh}{1-2C_2Lh}}$$
$$\leq 4C_2 \|A_hY_h - A_hZ_h\|_{-1,h}^2 e^{4TC_2L}.$$

Since $Y_h \in E_h$ and $0 \leq j \leq N$ were chosen arbitrary the operator A_h is stable for all $h < \min(\overline{h}, \frac{1}{4C_2L})$, i.e. there exists a constant $\tilde{C}_2$ independent of h such that

$$\|Y_h - Z_h\|_{0,h} \leq \tilde{C}_2 \|A_hY_h - A_hZ_h\|_{-1,h}$$

holds for all $Y_h, Z_h \in E_h$. Further we compute

$$\|A_hY_h - A_hZ_h\|_{-1,h}$$
$$\leq \|L_hY_h - L_hZ_h\|_{-1,h} + \left(\mathbb{E}\left(\max_{k\leq j\leq N}\left|\sum_{i=k}^{j}\left[T_hY_h(t_i) - T_hZ_h(t_i)\right]\right|^2\right)\right)^{\frac{1}{2}}$$
$$\leq \left(\frac{1}{C_1} + \sqrt{L(T+1)}\right)\|Y_h - Z_h\|_{0,h},$$

where we used $A_h = L_h + T_h$, the left-hand side of the bistablity inequality (3.2) for L_h and the estimate (4.4). Altogether we have shown the bistability of the operators $(A_h)_{h>0}$.

By interchanging the role of the operators $(A_h)_{h>0}$ and $(L_h)_{h>0}$ appropriately one proves the bistability of $(L_h)_{h>0}$ analogously. □

4.4. Bistability of the linear part. In this subsection we deal with the missing link between Lemma 4.2 and Theorem 3.4. Thus we have to show the following result:

LEMMA 4.4. *Under the assumptions $\rho(1) = 0$, $a_k \neq 0$ the sequence of operators $(L_h)_{h>0}$ is bistable if and only if Dahlquist's strong root condition is satisfied.*

By the assumption $\rho(1) = 0$ we can write

$$\rho(z) = \rho^*(z)(z-1),$$

where $\rho^*(z) = \sum_{j=0}^{k-1} a_j^* z^j$ is a polynomial of degree $k-1$ with $a_{k-1}^* \neq 0$. We introduce the operator $L_h^* : E_h \to F_h$ defined by

$$[L_h^*Y_h](t_i) = \begin{cases} Y_h(t_i), & \text{for } 0 \leq i \leq k-2, \\ \sum_{j=0}^{k-1} a_j^* Y_h(t_{i+j-k+1}), & \text{for } k-1 \leq i \leq N. \end{cases} \tag{4.5}$$

Note that ρ^* is the characteristic polynomial of the multistep method $(L_h^*)_{h<0}$. Moreover, we have

$$L_hY_h(t_i) = L_h^*Y_h(t_i) - L_h^*Y_h(t_{i-1}) \tag{4.6}$$

for all $i = k, \dots, N$. The following result will be useful for the proof of Lemma 4.4:

LEMMA 4.5. *Under the assumptions $\rho(1) = 0$, $a_k \neq 0$ the sequence of linear operators $(L_h)_{h>0}$ is bistable if and only if there exist constants $\lambda_1, \lambda_2 > 0$ such that the inequalities*

$$\lambda_1\|Y_h\|_{0,h} \leq \sum_{j=0}^{k-1}\|Y_h(t_j)\|_{L^2} + 2\left(\mathbb{E}\left(\max_{k-1\leq j\leq N}|L_h^*Y_h(t_j)|^2\right)\right)^{\frac{1}{2}} \leq \lambda_2\|Y_h\|_{0,h} \tag{4.7}$$

hold for all $h > 0$ *and* $Y_h \in E_h$.

Proof. By the linearity of the operators $(L_h)_{h>0}$ the bistability inequality (3.2) is written as

$$C_1\|Y_h\|_{0,h} \leq \|L_h Y_h\|_{-1,h} \leq C_2\|Y_h\|_{0,h}$$

for $Y_h \in \mathbb{E}_h$. The relationship (4.6) gives

$$\begin{aligned}
\|L_h Y_h\|_{-1,h} &= \sum_{j=0}^{k-1} \|Y_h(t_j)\|_{L^2} + \left(\mathbb{E}\left(\max_{k\leq j\leq N}\left|\sum_{i=k}^{j} L_h Y_h(t_i)\right|^2\right)\right)^{\frac{1}{2}} \\
&= \sum_{j=0}^{k-1} \|Y_h(t_j)\|_{L^2} + \left(\mathbb{E}\left(\max_{k\leq j\leq N} |L_h^* Y_h(t_j) - L_h^* Y_h(t_{k-1})|^2\right)\right)^{\frac{1}{2}} \\
&\leq \sum_{j=0}^{k-1} \|Y_h(t_j)\|_{L^2} + 2\left(\mathbb{E}\left(\max_{k-1\leq j\leq N} |L_h^* Y_h(t_j)|^2\right)\right)^{\frac{1}{2}}.
\end{aligned}$$

Conversely, we have

$$L_h^* Y_h(t_j) = \sum_{i=k}^{j} L_h Y_h(t_i) + L_h^* Y_h(t_{k-1}),$$

which we use to obtain

$$\begin{aligned}
&\sum_{j=0}^{k-1} \|Y_h(t_j)\|_{L^2} + 2\left(\mathbb{E}\left(\max_{k-1\leq j\leq N} |L_h^* Y_h(t_j)|^2\right)\right)^{\frac{1}{2}} \\
&= \sum_{j=0}^{k-1} \|Y_h(t_j)\|_{L^2} + 2\left(\mathbb{E}\left(\max_{k-1\leq j\leq N} \left|\textstyle\sum_{i=k}^{j} L_h Y_h(t_i) + L_h^* Y_h(t_{k-1})\right|^2\right)\right)^{\frac{1}{2}} \\
&\leq 2\left(\mathbb{E}\left(\max_{k\leq j\leq N} \left|\textstyle\sum_{i=k}^{j} L_h Y_h(t_i)\right|^2\right)\right)^{\frac{1}{2}} + \sum_{j=0}^{k-1}(1 + 2|a_j^*|)\|Y_h(t_j)\|_{L^2} \\
&\leq 2\left(1 + \textstyle\sum_{j=0}^{k-1} |a_j^*|\right) \|L_h Y_h\|_{-1,h}. \qquad \square
\end{aligned}$$

In the next step we collect results on difference equations written in terms of L^2-valued grid functions. For $Z_h \in \mathcal{G}_h$ the unique solution $Y_h \in \mathcal{G}_h$ to the equation $L_h^* Y_h = Z_h$ is given by

$$Y_h(t_i) = \sum_{\eta=0}^{k-2} v_i^\eta Z_h(t_\eta) + \sum_{\eta=k-1}^{N} w_i^\eta Z_h(t_\eta), \tag{4.8}$$

where for $\eta = 0, \ldots, k-2$ the real sequence $(v_i^\eta)_{i=0,\ldots,N}$ solves the homogeneous difference equations

$$\begin{aligned}
\sum_{j=0}^{k-1} a_j^* v_{i-k+1+j}^\eta &= 0, \qquad i = k-1, \ldots, N, \\
v_i^\eta &= \delta_{i,\eta}, \qquad i = 0, \ldots, k-2,
\end{aligned} \tag{4.9}$$

for $\eta = 0, \dots, k-2$ and the real sequence $(w_i^\eta)_{i=0,\dots,N}$ solves the inhomogeneous difference equations

$$\begin{aligned} \sum_{j=0}^{k-1} a_j^* w_{i-k+1+j}^\eta &= \delta_{i,\eta}, \qquad i = k-1, \dots, N, \\ w_i^\eta &= 0, \qquad\quad i = 0, \dots, k-2, \end{aligned} \tag{4.10}$$

for $\eta = k-1, \dots, N$ with $\delta_{i,j} = 0$ for $i \neq j$ and $\delta_{i,i} = 1$. It is well-known how the solutions to the linear difference equations (4.9), (4.10) can be expressed by the roots of the characteristic polynomial $\rho^*(z) = \sum_{j=0}^{k-1} a_j^* z^j$:

Let $\zeta_i \in \mathbb{C}$, $i = 1, \dots, s$, be the pairwise distinct roots of ρ^* with multiplicity $k_i \geq 1$ ($k_1 + \dots + k_s = k-1$). A fundamental system of solutions to the homogeneous difference equation (4.9) is given by

$$u_j^{i,\kappa} = \left(\prod_{\nu=j-\kappa+1}^{j} \nu \right) \zeta_i^{j-\kappa}, \quad i = 1, \dots, s, \ \kappa = 1, \dots, k_i, \ j = 0, \dots, N,$$

where $\prod_\emptyset = 1$. All solutions $(v_j^\eta)_{j=0,\dots,N}$ to (4.9) can be written as

$$v_j^\eta = \sum_{i=1}^{s} \sum_{\kappa=1}^{k_i} c_{i,\kappa}^\eta u_j^{i,\kappa},$$

where the coefficients $c_{i,\kappa}^\eta \in \mathbb{C}$ are uniquely determined by the initial values (in particular, they are independent of N).

Now consider the real-valued solution $(x_i)_{i=0,\dots,N}$ to the homogeneous difference equation

$$\begin{aligned} \sum_{j=0}^{k-1} a_j^* x_{i-k+1+j} &= 0, \qquad\qquad i = k-1, \dots, N, \\ x_i = 0, \quad x_{k-2} &= \frac{1}{a_{k-1}^*}, \qquad i = 0, \dots, k-3. \end{aligned}$$

For $i < 0$ we define $x_i := 0$. Then we have

$$w_i^\eta = x_{i-\eta+k-2} \tag{4.11}$$

for the solution to (4.10), since

$$\sum_{j=0}^{k-1} a_j^* w_{i-k+1+j}^\eta = \sum_{j=0}^{k-1} a_j^* x_{i+j-\eta-1} = \delta_{i,\eta}.$$

Note that $(x_i)_{i=0,\dots,N}$ solves a homogeneous difference equation. Hence it also has a representation as a linear combination of the fundamental solutions $(u_j^{i,\kappa})_{j=1,\dots,N}$.

Remark. Under the usual root condition one can prove that the fundamental solutions to the homogeneous difference equation (4.9) are uniformly bounded for all $N \in \mathbb{N}$. This is sufficient to show that the solution $Y_h \in \mathcal{G}_h$ to $L_h Y_h = Z_h$ satisfies

$$\|Y_h\|_{0,h} \leq C \|Z_h\|_{0,h}$$

for a constant $C > 0$ which is independent of h. From this result one derives the stability of the operators $(L_h)_{h>0}$ but for a different pair of norms (c.f. [10] for deterministic multistep methods). $\square$

Proof of Lemma 4.4. By Lemma 4.5 it remains to show that the inequalities (4.7) hold if and only if the strong version of Dahlquist's root condition holds.

We first prove that the strong root condition is sufficient for the inequalities (4.7) to be true. Let $Y_h \in E_h$ denote the solution to $L_h^* Y_h = Z_h \in F_h$. Using the representation (4.8) gives

$$\begin{aligned}
\|Y_h\|_{0,h} &\leq \sum_{\eta=0}^{k-2} \left(\mathbb{E} \left(\max_{0\leq i\leq N} |v_i^\eta Z_h(t_\eta)|^2 \right) \right)^{\frac{1}{2}} + \left(\mathbb{E} \left(\max_{0\leq i\leq N} \left| \textstyle\sum_{\eta=k-1}^N w_i^\eta Z_h(t_\eta) \right|^2 \right) \right)^{\frac{1}{2}} \\
&\leq \sum_{\eta=0}^{k-2} \max_{0\leq i\leq N} |v_i^\eta| \|Z_h(t_\eta)\|_{L^2} + \left(\mathbb{E} \left(\max_{0\leq i\leq N} \left[\textstyle\sum_{\eta=k-1}^N |w_i^\eta| \, |Z_h(t_\eta)| \right]^2 \right) \right)^{\frac{1}{2}} \\
&\leq \left(\max_{0\leq \eta\leq k-2} \max_{0\leq i\leq N} |v_i^\eta| \right) \sum_{j=0}^{k-2} \|Z_h(t_j)\|_{L^2} \\
&\qquad + \max_{0\leq i\leq N} \left(\textstyle\sum_{\eta=k-1}^N |w_i^\eta| \right) \left(\mathbb{E} \left(\max_{k-1\leq \eta\leq N} |Z_h(t_\eta)|^2 \right) \right)^{\frac{1}{2}} \\
&\leq C_N \left[\sum_{j=0}^{k-2} \|Y_h(t_j)\|_{L^2} + 2 \left(\mathbb{E} \left(\max_{k-1\leq \eta\leq N} |L_h^* Y_h(t_\eta)|^2 \right) \right)^{\frac{1}{2}} \right],
\end{aligned}$$

where the constant C_N is given by

$$C_N := \max_{0\leq \eta\leq k-2} \max_{0\leq i\leq N} |v_i^\eta| + \frac{1}{2} \max_{0\leq i\leq N} \sum_{\eta=k-1}^N |w_i^\eta|.$$

The first part of the inequalities (4.7) is proved if we can show

$$\sup_{N\in\mathbb{N}} C_N < \infty. \tag{4.12}$$

But under the strong root condition all roots of ρ^* satisfy $|\zeta_i| \leq r_0 < 1$ for $i = 1, \ldots, s$. Hence there exists a constant $C > 0$ such that

$$|u_j^{i,\kappa}| \leq C r_0^j, \quad j = 0, \ldots, N, \tag{4.13}$$

for all $i = 1, \ldots, s$ and $\kappa = 1, \ldots, k_i$. Since $(v_j^\eta)_{j=0,\ldots,N}$ and $(x_j)_{j=0,\ldots,N}$ are finite linear combinations of the fundamental system $(u_j^{i,\kappa})_{j=0,\ldots,N}$ the estimate (4.13) is also valid for these sequences. By the relation (4.11) we compute

$$\sum_{\eta=k-1}^N |w_i^\eta| = \sum_{\eta=k-1}^N |x_{i-\eta+k-2}| = \sum_{\eta=k-1}^{i+k-1} |x_{i-\eta+k-1}| \leq C \sum_{\eta=0}^{i} r_0^\eta < C \frac{1}{1-r_0} < \infty,$$

where we used $x_i = 0$ for $i < 0$. Altogether this proves (4.12).

The right hand side of (4.7) follows directly from

$$\sum_{j=0}^{k-2} \|Y_h(t_j)\|_{L^2} + 2\left(\mathbb{E}\left(\max_{k-1\le\eta\le N} |L_h^* Y_h(t_\eta)|^2\right)\right)^{\frac{1}{2}}$$
$$\le (k-1)\|Y_h\|_{0,h} + 2\left(\mathbb{E}\left(\max_{k-1\le\eta\le N} \left[\sum_{j=0}^{k-1} |a_j^*|\,|Y_h(t_{\eta-k+1+j})|\right]^2\right)\right)^{\frac{1}{2}}$$
$$\le \left(k-1+2\sum_{j=0}^{k-1}|a_j^*|\right)\|Y_h\|_{0,h}\,.$$

Consequently, the strong root condition is sufficient for the inequalities (4.7) and for the bistability of the operators L_h.

Conversely, assume that the inequalities (4.7) hold for all $h > 0$ and $Y_h \in E_h$ and that ρ does not satisfy the strong root condition, i.e. there exists $\zeta \in \mathbb{C}$ with $\rho^*(\zeta) = 0$ and $|\zeta| \ge 1$.

First, we focus on the case $|\zeta| = 1$. Define $z_j = j(\zeta^j + \overline{\zeta}^j) \in \mathbb{R}$ and let $Y_h(t_j) := z_j Y$ for $Y \in L^2(\Omega, \mathcal{F}_{t_0}, P; \mathbb{R}^d)$. Then $Y_h \in E_h$ and if we apply L_h^* to Y_h we get

$$L_h^* Y_h(t_j) = \sum_{\eta=0}^{k-1} a_\eta^* Y_h(t_{\eta+j-k+1}) = \sum_{\eta=0}^{k-1} a_\eta^* (\eta+j-k+1)\left(\zeta^{\eta+j-k+1} + \overline{\zeta}^{\eta+j-k+1}\right)$$
$$= -\zeta^{j-k+1}\rho^*(\zeta) + \zeta^{j-k+2}\frac{d}{dz}\rho^*(\zeta) - \overline{\zeta}^{j-k+1}\rho^*(\overline{\zeta}) + \overline{\zeta}^{j-k+2}\frac{d}{dz}\rho^*(\overline{\zeta}).$$

Since ρ^* is a real polynomial we also have $\rho^*(\overline{\zeta}) = 0$ and thus

$$\max_{k-1\le j\le N} |L_h^* Y_h(t_j)| \le \left|\frac{d}{dz}\rho^*(\zeta)\right| + \left|\frac{d}{dz}\rho^*(\overline{\zeta})\right| < \infty.$$

Combining this with (4.7) gives us

$$\lambda_1 \|Y_h\|_{0,h} \le \sum_{j=0}^{k-1} \|Y_h(t_j)\|_{L^2} + 2\left(\mathbb{E}\left(\max_{k-1\le j\le N} |L_h^* Y_h(t_j)|^2\right)\right)^{\frac{1}{2}} < \infty$$

for $\lambda_1 > 0$. On the other hand we have

$$\lim_{h\to 0} \|Y_h\|_{0,h} = \infty$$

which contradicts (4.7).

The case $|\zeta| > 1$ also contradicts (4.7) by using $z_j = \zeta^j + \overline{\zeta}^j \in \mathbb{R}$ for $j \in \mathbb{N}_0$. ◻

5. Consistency. The aim of this section is to prove Theorem 3.7. We deal with each numerical scheme in a separate subsection.

5.1. Consistency of the stochastic theta method. Before we start with the estimate of the local truncation error we quote the following useful result from [1, 18]:

THEOREM 5.1. *Under the assumptions (A1) and (A2) the solution X to* (1.1) *satisfies*

$$\mathbb{E}\left(|X(t)|^2\right) \le \left(1 + \mathbb{E}(|X_0|^2)\right) e^{Ct}$$

and

$$\mathbb{E}\left(|X(t)-X_0|^2\right) \le C\left(1+\mathbb{E}(|X_0|^2)\right) te^{Dt}$$

for all $0 \le t \le T$ *and some constants* $C, D > 0$ *depending only on* K *and* T.

In particular, using the semigroup property of X, *one can prove the estimate*

$$\mathbb{E}\left(|X(t)-X(s)|^2\right) \le C|t-s|$$

for all $t, s \in [0,T]$ *and some constant* $C > 0$ *depending only on* K, T *and* $\mathbb{E}(|X_0|^2)$.

By the definition of our stochastic version of Spijker's norm the local truncation error of the stochastic theta method can be written as

$$\|A_h^{STM} r_h^E X\|_{-1,h} = \|X(0)-\tilde{X}_0\|_{L^2(\Omega)} + \left(\mathbb{E}\left(\max_{1\le i\le N}\left|\sum_{j=1}^{i}[A_h^{STM} r_h^E X](t_j)\right|^2\right)\right)^{\frac{1}{2}}.$$

The assumption (C1) assures the consistency of the initial value $\tilde{X}_0$ with the order $\gamma = \frac{1}{2}$. Thus it remains to estimate the second summand

$$\begin{aligned}
&\left(\mathbb{E}\left(\max\nolimits_{1\le i\le N} |\textstyle\sum_{j=1}^{i}[A_h^{STM} r_h^E X](t_j)|^2\right)\right)^{\frac{1}{2}}\\
&= \Big(\mathbb{E}\Big(\max\nolimits_{1\le i\le N}\Big|\textstyle\sum_{j=1}^{i}\Big[X(t_j)-X(t_{j-1})-h(1-\theta)b^0(t_{j-1},X(t_{j-1}))\\
&\qquad -h\theta b^0(t_j,X(t_j)) - \textstyle\sum_{r=1}^{m} b^r(t_{j-1},X(t_{j-1}))I_{(r)}^{t_j}\Big]\Big|^2\Big)\Big)^{\frac{1}{2}}.
\end{aligned}$$

From the representation (1.2) and the triangle inequality we obtain the estimate

$$\le (1-\theta)\left(\mathbb{E}\left(\max_{1\le i\le N}\left|\sum_{j=1}^{i}\int_{t_{j-1}}^{t_j} b^0(s,X(s))-b^0(t_{j-1},X(t_{j-1}))ds\right|^2\right)\right)^{\frac{1}{2}} \tag{5.1}$$

$$+\,\theta\left(\mathbb{E}\left(\max_{1\le i\le N}\left|\sum_{j=1}^{i}\int_{t_{j-1}}^{t_j} b^0(s,X(s))-b^0(t_j,X(t_j))ds\right|^2\right)\right)^{\frac{1}{2}} \tag{5.2}$$

$$+\left(\mathbb{E}\left(\max_{1\le i\le N}\left|\sum_{j=1}^{i}\sum_{r=1}^{m}\int_{t_{j-1}}^{t_j} b^r(s,X(s))-b^r(t_{j-1},X(t_{j-1}))dW^r(s)\right|^2\right)\right)^{\frac{1}{2}}. \tag{5.3}$$

We estimate the terms separately. Jensen's inequality yields for the square of the first term (5.1)

$$\begin{aligned}
T_1 &:= \mathbb{E}\left(\max_{1\le i\le N}\left|i\sum_{j=1}^{i}\frac{1}{i}h\int_{t_{j-1}}^{t_j}\frac{1}{h}(b^0(s,X(s))-b^0(t_{j-1},X(t_{j-1})))ds\right|^2\right)\\
&\le \mathbb{E}\left(\max_{1\le i\le N}\left[ih\sum_{j=1}^{i}\int_{t_{j-1}}^{t_j}\left|b^0(s,X(s))-b^0(t_{j-1},X(t_{j-1}))\right|^2 ds\right]\right)\\
&= T\sum_{j=1}^{N}\int_{t_{j-1}}^{t_j}\mathbb{E}\left(\left|b^0(s,X(s))-b^0(t_{j-1},X(t_{j-1}))\right|^2\right)ds.
\end{aligned}$$

Using the assumptions (A2) and (C2) we find for $r = 0, \dots, m$

$$\begin{aligned}
&|b^r(s, X(s)) - b^r(t_{j-1}, X(t_{j-1}))|^2 \\
&\leq (|b^r(s, X(s)) - b^r(s, X(t_{j-1}))| + |b^r(s, X(t_{j-1})) - b^r(t_{j-1}, X(t_{j-1}))|)^2 \\
&\leq \left(K|X(s) - X(t_{j-1})| + K(1 + |X(t_{j-1})|)\sqrt{|s - t_{j-1}} \right)^2 \\
&\leq 2K^2 |X(s) - X(t_{j-1})|^2 + 2K^2 \left(1 + |X(t_{j-1})|\right)^2 |s - t_{j-1}|.
\end{aligned}$$

Applying Theorem 5.1 leads to the estimate

$$\begin{aligned}
&\mathbb{E}\left(|b^r(s, X(s)) - b^r(t_{j-1}, X(t_{j-1}))|^2 \right) \\
&\leq 2K^2 \mathbb{E}\left(|X(s) - X(t_{j-1})|^2 \right) + 4K^2 \left(1 + \mathbb{E}\left(|X(t_{j-1})|^2 \right) \right) |s - t_{j-1}| \\
&\leq C|s - t_{j-1}|,
\end{aligned} \tag{5.4}$$

where the constant C only depends on K, T and $\mathbb{E}\left(|X_0|^2\right)$. Hence we complete our estimate of the term T_1 as follows

$$T_1 \leq T \sum_{j=1}^{N} \int_{t_{j-1}}^{t_j} C|s - t_{j-1}| ds = \frac{1}{2} C T^2 h. \tag{5.5}$$

Replacing t_{j-1} by t_j in (5.4) leads to the analogous result for the term in (5.2), i.e.

$$T_2 := \mathbb{E}\left(\max_{1 \leq i \leq N} \left| \sum_{j=1}^{i} \int_{t_{j-1}}^{t_j} b^0(s, X(s)) - b^0(t_j, X(t_j)) ds \right|^2 \right) \leq \frac{1}{2} C T^2 h. \tag{5.6}$$

Thus, it remains to estimate the third term (5.3)

$$T_3 := \mathbb{E}\left(\max_{1 \leq i \leq N} \left| \sum_{j=1}^{i} \sum_{r=1}^{m} \int_{t_{j-1}}^{t_j} b^r(s, X(s)) - b^r(t_{j-1}, X(t_{j-1})) dW^r(s) \right|^2 \right).$$

By the martingale property of the stochastic Itô-integral (c.f. Theorem 1.5.12 in [18]) we are able to apply Doob's martingale inequality. Then we use the martingale property and the independence of the Wiener processes to interchange the sums and the expectation (c.f. Theorem 1.5.21 and Lemma 1.5.22 in [18]). This yields

$$\begin{aligned}
T_3 &\leq 4\mathbb{E}\left(\left| \sum_{j=1}^{N} \sum_{r=1}^{m} \int_{t_{j-1}}^{t_j} b^r(s, X(s)) - b^r(t_{j-1}, X(t_{j-1})) dW^r(s) \right|^2 \right) \\
&= 4 \sum_{j=1}^{N} \sum_{r=1}^{m} \mathbb{E}\left(\left| \int_{t_{j-1}}^{t_j} b^r(s, X(s)) - b^r(t_{j-1}, X(t_{j-1})) dW^r(s) \right|^2 \right) \\
&= 4 \sum_{j=1}^{N} \sum_{r=1}^{m} \int_{t_{j-1}}^{t_j} \mathbb{E}\left(|b^r(s, X(s)) - b^r(t_{j-1}, X(t_{j-1}))|^2 \right) ds,
\end{aligned}$$

where we used the Itô-isometry in the last step. Again we apply (5.4) and obtain

$$T_3 \leq 4 \sum_{j=1}^{N} \sum_{r=1}^{m} \int_{t_{j-1}}^{t_j} C|s - t_{j-1}| ds = 2CTmh. \tag{5.7}$$

Combining (C1) and the estimates (5.5), (5.6) and (5.7) we arrive at the final estimate

$$\|A_h r_h^E X\|_{-1,h} \leq Ch^{\frac{1}{2}} + (1-\theta)\sqrt{\frac{1}{2}CT^2 h} + \theta\sqrt{\frac{1}{2}CT^2 h} + \sqrt{2CTmh} = \tilde{C}h^{\frac{1}{2}},$$

where the constant $\tilde{C} > 0$ only depends on K, T, $\mathbb{E}\left(|X_0|^2\right)$ and m. Thus, the stochastic theta method is consistent of order $\gamma = \frac{1}{2}$.

5.2. Consistency of the BDF2-Maruyama method. In this subsection we prove the consistency of the BDF2-Maruyama method under the assumptions (A1), (A2), (C1) and (C2). The idea is to write BDF2-Maruyama as a sum of two stochastic theta methods and then to use the result from the previous subsection. The local truncation error is given by

$$\|A_h^{BDF} r_h^E X\|_{-1,h} = \sum_{j=0}^{1} \|X(t_j) - \tilde{X}_j\|_{L^2} + \left(\mathbb{E}\left(\max_{2\leq i\leq N}\left|\sum_{j=2}^{i}[A_h^{BDF} r_h^E X](t_j)\right|^2\right)\right)^{\frac{1}{2}}$$

Again, by assumption (C1), the initial values are assumed to be consistent of order $\gamma = \frac{1}{2}$. Thus we only have to estimate the second term. But this one can be written by the definition (2.5) of A_h^{BDF} and the triangle inequality as follows

$$\begin{aligned}
&\left(\mathbb{E}\left(\max_{2\leq i\leq N}\left|\sum_{j=2}^{i}[A_h^{BDF} r_h^E X](t_j)\right|^2\right)\right)^{\frac{1}{2}} \\
&= \left(\mathbb{E}\left(\max_{2\leq i\leq N}\left|\sum_{j=2}^{i}\left[X(t_j) - \frac{4}{3}X(t_{j-1}) + \frac{1}{3}X(t_{j-2}) - \frac{2}{3}hb^0(t_j, X(t_j))\right.\right.\right.\right. \\
&\qquad\qquad \left.\left.\left.\left. - \sum_{r=1}^{m} b^r(t_{j-1}, X(t_{j-1})) I_{(r)}^{t_j} + \frac{1}{3}\sum_{r=1}^{m} b^r(t_{j-2}, X(t_{j-2})) I_{(r)}^{t_{j-1}}\right]\right|^2\right)\right)^{\frac{1}{2}} \\
&\leq \left(\mathbb{E}\left(\max_{2\leq i\leq N}\left|\sum_{j=2}^{i}\left[X(t_j) - X(t_{j-1}) - \frac{2}{3}hb^0(t_j, X(t_j)) - \frac{1}{3}hb^0(t_{j-1}, X(t_{j-1}))\right.\right.\right.\right. \\
&\qquad\qquad \left.\left.\left.\left. - \sum_{r=1}^{m} b^r(t_{j-1}, X(t_{j-1})) I_{(r)}^{t_j}\right]\right|^2\right)\right)^{\frac{1}{2}} \\
&+ \left(\mathbb{E}\left(\max_{2\leq i\leq N}\left|\frac{1}{3}\sum_{j=2}^{i}\left[X(t_{j-1}) - X(t_{j-2}) - hb^0(t_{j-1}, X(t_{j-1}))\right.\right.\right.\right. \\
&\qquad\qquad \left.\left.\left.\left. - \sum_{r=1}^{m} b^r(t_{j-2}, X(t_{j-2})) I_{(r)}^{t_{j-1}}\right]\right|^2\right)\right)^{\frac{1}{2}}.
\end{aligned}$$

Now the two summands can be estimated as in the previous subsection. Thus the BDF2-Maruyama scheme is also consistent of order $\gamma = \frac{1}{2}$.

5.3. Consistency of higher order Itô-Taylor schemes. In this subsection we prove the consistency of the Itô-Taylor schemes. Choose $\gamma \in \{\frac{n}{2} \,|\, n \in \mathbb{N}\}$ such that assumptions (A1), (A2), (A3), (C1) and (C3) are satisfied. For the estimate we need the following result on Itô-Taylor expansions from [16].

THEOREM 5.2. *Under the assumptions (A1), (A2), (A3) the Itô-Taylor expansion*

$$X(t_i) = X(t_{i-1}) + \sum_{\alpha \in \mathcal{A}_\gamma} f_\alpha(t_{i-1}, X(t_{i-1})) I_\alpha^{t_i} + \sum_{\alpha \in \mathcal{B}(\mathcal{A}_\gamma)} I_\alpha[f_\alpha(\cdot, X(\cdot))]_{t_{i-1}}^{t_i},$$

holds for all $i = 1, \ldots, N$, where for $\alpha = (j_1, \ldots, j_\ell) \in \mathcal{B}(\mathcal{A}_\gamma)$

$$I_\alpha[f_\alpha(\cdot, X(\cdot))]_{t_{i-1}}^{t_i} = \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s_1} \cdots \int_{t_{i-1}}^{s_{\ell-1}} f_\alpha(s_\ell, X(s_\ell)) dW^{j_1}(s_\ell) \cdots dW^{j_\ell}(s_1).$$

For the proof we refer to Theorem 5.5.1 in [16]. Now the local truncation error of the Itô-Taylor scheme of order γ takes the form

$$\|A_h^{ITS} r_h^E X\|_{-1,h} = \|X(0) - \tilde{X}_0\|_{L^2} + \left(\mathbb{E} \left(\max_{1 \le i \le N} \left| \sum_{j=1}^{i} [A_h^{ITS} r_h^E X](t_j) \right|^2 \right) \right)^{\frac{1}{2}}.$$

Again, by assumption (C1), the initial value $\tilde{X}_0$ is assumed to be sufficiently consistent. Thus we are only concerned with the second summand

$$\begin{aligned}
&\left(\mathbb{E} \left(\max_{1 \le i \le N} \left| \sum_{j=1}^{i} [A_h^{ITS} r_h^E X](t_j)] \right|^2 \right) \right)^{\frac{1}{2}} \\
&= \left(\mathbb{E} \left(\max_{1 \le i \le N} \left| \sum_{j=1}^{i} \left[X(t_i) - \sum_{\alpha \in \mathcal{A}_\gamma} f_\alpha(t_{j-1}, X(t_{j-1})) I_\alpha^{t_j} \right] \right|^2 \right) \right)^{\frac{1}{2}} \\
&= \left(\mathbb{E} \left(\max_{1 \le i \le N} \left| \sum_{j=1}^{i} \left[\sum_{\alpha \in \mathcal{B}(\mathcal{A}_\gamma)} I_\alpha[f_\alpha(\cdot, X(\cdot))]_{t_{j-1}}^{t_j} \right] \right|^2 \right) \right)^{\frac{1}{2}} \\
&\le \sum_{\alpha \in \mathcal{B}(\mathcal{A}_\gamma)} \left(\mathbb{E} \left(\max_{1 \le i \le N} \left| \sum_{j=1}^{i} I_\alpha[f_\alpha(\cdot, X(\cdot))]_{t_{j-1}}^{t_j} \right|^2 \right) \right)^{\frac{1}{2}},
\end{aligned}$$

where we applied Theorem 5.2 and the triangle inequality. Since the remainder set $\mathcal{B}(\mathcal{A}_\gamma)$ is finite (c.f. [16]) it is enough to estimate each summand separately. First we consider all multi-indices $\alpha \in \mathcal{B}(A_\gamma)$ with $\ell = \ell(\alpha) = n(\alpha)$, i.e. $\alpha = (0, \ldots, 0)$. For these multi-indices one computes

$$
\begin{aligned}
&\mathbb{E}\left(\max_{1\le i\le N}\left|\sum_{j=1}^{i} I_\alpha\left[f_\alpha(\cdot,X(\cdot))\right]_{t_{j-1}}^{t_j}\right|^2\right)\\
&= \mathbb{E}\left(\max_{1\le i\le N}\left|\sum_{j=1}^{i}\int_{t_{j-1}}^{t_j}\int_{t_{j-1}}^{s_1}\cdots\int_{t_{j-1}}^{s_{\ell-1}} f_\alpha(s_\ell,X(s_\ell))ds_\ell\cdots ds_1\right|^2\right)\\
&= \mathbb{E}\left(\max_{1\le i\le N}\left|\sum_{j=1}^{i}\frac{1}{(\ell-1)!}\int_{t_{j-1}}^{t_j} f_\alpha(s,X(s))(t_j-s)^{\ell-1}ds\right|^2\right)\\
&\le \left(\frac{1}{(\ell-1)!}\right)^2 \mathbb{E}\left(\max_{1\le i\le N}\left[ih\sum_{j=1}^{i}\int_{t_{j-1}}^{t_j}|f_\alpha(s,X(s))|^2|t_j-s|^{2(\ell-1)}ds\right]\right),
\end{aligned}
$$

where we used Jensen's inequality in the last step. We complete the estimate by

$$
\begin{aligned}
&\le \left(\frac{1}{(\ell-1)!}\right)^2 T\sum_{j=1}^{N}\int_{t_{j-1}}^{t_j}\mathbb{E}\left(|f_\alpha(s,X(s))|^2\right)ds\, h^{2\ell-2}\\
&= \left(\frac{1}{(\ell-1)!}\right)^2 T\int_0^T\mathbb{E}\left(|f_\alpha(s,X(s))|^2\right)ds\, h^{2\ell-2}.
\end{aligned}
$$

By assumption (C3) the integral is finite and by the definitions of $\mathcal{A}_\gamma$ and $\mathcal{B}(\mathcal{A}_\gamma)$ we have $\alpha\in\mathcal{B}(\mathcal{A}_\gamma)$ with $\ell(\alpha)=n(\alpha)$ only if $\ell=\ell(\alpha)=\gamma+1$ or $\ell=\ell(\alpha)=\gamma+\frac{3}{2}$. Hence $h^{2\ell-2}=\mathcal{O}(h^\gamma)$, which is also the order of the complete term.

Thus it remains to estimate the summands with all indices $\alpha\in\mathcal{B}(\mathcal{A}_\gamma)$ such that $n(\alpha)<\ell(\alpha)$. In this case note that $\mathbb{E}\left(I_\alpha\left[f_\alpha(\cdot,X(\cdot))\right]_{t_{j-1}}^{t_j}\mid\mathcal{F}_{t_i}\right)=0$ for all $i<j$ (c.f. Lemma 5.7.1 in [16]). Therefore $(S_i)_{i=1,\dots,N}$ with $S_i=\sum_{j=1}^{i}I_\alpha[f_\alpha(\cdot,X(\cdot))]_{t_{j-1}}^{t_j}$ is a discrete martingale. Furthermore, by Lemma 5.3 below, we have the following estimate of the second moment:

$$
\mathbb{E}\left(\left|I_\alpha\left[f_\alpha(\cdot,X(\cdot))\right]_{t_{j-1}}^{t_j}\right|^2\right)\le h^{\ell(\alpha)+n(\alpha)-1}\int_{t_{j-1}}^{t_j}\mathbb{E}\left(|f_\alpha(u,X(u))|^2\right)du
$$

for all $j=1,\dots,N$. Thus we are allowed to apply Doob's martingale inequality and obtain

$$
\begin{aligned}
\mathbb{E}\left(\max_{1\le i\le N}\left|\sum_{j=1}^{i}I_\alpha\left[f_\alpha(\cdot,X(\cdot))\right]_{t_{j-1}}^{t_j}\right|^2\right) &\le 4\mathbb{E}\left(\left|\sum_{j=1}^{N}I_\alpha\left[f_\alpha(\cdot,X(\cdot))\right]_{t_{j-1}}^{t_j}\right|^2\right)\\
&= 4\sum_{j=1}^{N}\mathbb{E}\left(\left|I_\alpha\left[f_\alpha(\cdot,X(\cdot))\right]_{t_{j-1}}^{t_j}\right|^2\right)\\
&\le 4\int_0^T\mathbb{E}\left(|f_\alpha(u,X(u))|^2\right)du\; h^{\ell(\alpha)+n(\alpha)-1},
\end{aligned}
$$

where we used the martingale property of the stochastic integrals and Lemma 5.3. Also in this case we have $\ell(\alpha)+n(\alpha)-1\ge 2\gamma$ by the definitions of $\mathcal{A}_\gamma$ and $\mathcal{B}(\mathcal{A}_\gamma)$.

Hence under the given assumptions the Itô-Taylor scheme of order γ is consistent of order γ.

LEMMA 5.3. *Assume that the stochastic process $f : [0,T] \to \mathbb{R}^d$ is stochastically integrable with respect to the iterated Itô-integral I_α. If*

$$\int_s^t \mathbb{E}\left(|f(u)|^2\right) du < \infty$$

for all $0 \le s < t \le T$ then

$$\mathbb{E}\left(\left|I_\alpha\left[f(\cdot)\right]_s^t\right|^2\right) \le \int_s^t \mathbb{E}\left(\left|f(u)\right|^2\right) du \; (t-s)^{\ell(\alpha)+n(\alpha)-1}$$

for all $0 \le s < t \le T$ and all multi-indices α.

Proof. The proof is similar to the proofs of Lemmas 2.1 and 2.2 in [19] and done by an inductive argument. If $\ell(\alpha) = 1$ and hence $\alpha = (j_1)$, then the estimate holds with equality in the case $j_1 \neq 0$ by the Itô-isometry. If $j_1 = 0$, then the estimate is just Jensen's inequality.

Let $\ell(\alpha) > 1$ with $\alpha = (j_1, \ldots, j_\ell)$. First consider the case $j_\ell = 0$. Then by Jensen's inequality

$$\begin{aligned}
&\mathbb{E}\left(\left|I_\alpha\left[f(\cdot)\right]_s^t\right|^2\right) \\
&= \mathbb{E}\left(\left|\int_s^t \int_s^{s_1} \cdots \int_s^{s_{\ell-1}} f(s_\ell) dW^{j_1}(s_\ell) \cdots dW^{j_\ell}(s_1)\right|^2\right) \\
&\le (t-s) \int_s^t \mathbb{E}\left(\left|\int_s^{s_1} \cdots \int_s^{s_{\ell-1}} f(s_\ell) dW^{j_1}(s_\ell) \cdots dW^{j_{\ell-1}}(s_2)\right|^2\right) ds_1 \\
&= (t-s) \int_s^t \mathbb{E}\left(\left|I_{\tilde\alpha}\left[f(\cdot)\right]_s^{s_1}\right|^2\right) ds_1,
\end{aligned}$$

where $\tilde\alpha = (j_1, \ldots, j_{\ell-1})$ with $\ell(\tilde\alpha) = \ell(\alpha) - 1$ and $n(\tilde\alpha) = n(\alpha) - 1$. By the induction hypothesis we get

$$\begin{aligned}
\mathbb{E}\left(\left|I_{\tilde\alpha}\left[f(\cdot)\right]_s^{s_1}\right|^2\right) &\le \int_s^{s_1} \mathbb{E}\left(\left|f(u)\right|^2\right) du \; (s_1 - s)^{\ell(\tilde\alpha)+n(\tilde\alpha)-1} \\
&\le \int_s^t \mathbb{E}\left(|f(u)|^2\right) du \; (t-s)^{\ell(\alpha)+n(\alpha)-3}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathbb{E}\left(\left|I_\alpha\left[f(\cdot)\right]_s^t\right|^2\right) &\le (t-s) \int_s^t \int_s^t \mathbb{E}\left(\left|f(u)\right|^2\right) du \; (t-s)^{\ell(\alpha)+n(\alpha)-3} ds_1 \\
&= \int_s^t \mathbb{E}\left(\left|f(u)\right|^2\right) du \; (t-s)^{\ell(\alpha)+n(\alpha)-1}.
\end{aligned}$$

If $j_\ell \neq 0$ the Itô-isometry gives

$$\mathbb{E}\left(\left|I_\alpha\left[f(\cdot)\right]_s^t\right|^2\right) = \int_s^t \mathbb{E}\left(\left|I_{\tilde\alpha}\left[f(\cdot)\right]_s^{s_1}\right|^2\right) ds_1.$$

After applying the induction hypothesis one uses $n(\tilde\alpha) = n(\alpha)$ to obtain the same order. □

6. Maximum order of convergence. In this section we extend a well-known result from J.M.C. Clark and R.J. Cameron [6]: They constructed an example to show that, in general, a numerical scheme has the maximum order of convergence $\frac{1}{2}$ if it only uses the increments $W^r(t_i) - W^r(t_{i-1})$ of the driving Wiener processes. We show that this result follows in a natural way for the BDF2-Maruyama method from the two-sided error estimate (3.3). Moreover, we present a generalization of Clark and Cameron's example to treat the higher order Itô-Taylor schemes. We refer to [3] for a discussion of the stochastic theta method.

THEOREM 6.1. *In general, the maximum order of convergence*

(i) *of the BDF2-Maruyama method is equal to* $\frac{1}{2}$,

(ii) *of the Itô-Taylor scheme of order* γ *is equal to* γ.

Proof. For the BDF2-Maruyama method we consider the SODE

$$dX(t) = \begin{pmatrix} 1 & 0 \\ 0 & X_1(t) \end{pmatrix} d \begin{pmatrix} W^1(t) \\ W^2(t) \end{pmatrix}, \quad X(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{6.1}$$

with the analytic solution

$$X(t) = \begin{pmatrix} W^1(t) \\ \int_0^t W^1(s) dW^2(s) \end{pmatrix}, \quad \text{for } t \in [0,T]. \tag{6.2}$$

Note that (6.1) satisfies the assumptions (A1), (A2) and (C2). As initial values of the BDF2-Maruyama method we choose the corresponding analytic solution, i.e. $\tilde{X}_i = X(t_i)$, $i = 0,1$. By the two-sided error estimate (3.3) it is enough to estimate the local truncation error from below. It follows

$$\begin{aligned}
&\left\| A_h^{BDF} r_h^E X \right\|_{-1,h}^2 \\
&= \mathbb{E}\left(\max_{2\le i\le N} \left| \sum_{j=2}^{i} \left[X(t_j) - \frac{4}{3} X(t_{j-1}) + \frac{1}{3} X(t_{j-2}) - \begin{pmatrix} I_{(1)}^{t_j} \\ X_1(t_{j-1}) I_{(2)}^{t_j} \end{pmatrix} \right.\right.\right. \\
&\qquad\qquad \left.\left.\left. + \frac{1}{3} \begin{pmatrix} I_{(1)}^{t_{j-1}} \\ X_1(t_{j-2}) I_{(2)}^{t_{j-1}} \end{pmatrix} \right] \right|^2 \right) \\
&= \mathbb{E}\left(\max_{2\le i\le N} \left| \textstyle\sum_{j=2}^{i} \left[\int_{t_{j-1}}^{t_j} W^1(s) dW^2(s) - \frac{1}{3} \int_{t_{j-2}}^{t_{j-1}} W^1(s) dW^2(s) \right.\right.\right. \\
&\qquad\qquad \left.\left.\left. - \textstyle\int_{t_{j-1}}^{t_j} W^1(t_{j-1}) dW^2(s) + \frac{1}{3} \int_{t_{j-2}}^{t_{j-1}} W^1(t_{j-2}) dW^2(s) \right] \right|^2 \right) \\
&\ge \max_{2\le i\le N} \mathbb{E}\left(\left| \textstyle\int_{t_{i-1}}^{t_i} W^1(s) - W^1(t_{i-1}) dW^2(s) - \frac{1}{3} \int_{t_0}^{t_1} W^1(s) - W^1(t_0) dW^2(s) \right.\right. \\
&\qquad\qquad \left.\left. + \textstyle\frac{2}{3} \sum_{j=2}^{i-1} \int_{t_{j-1}}^{t_j} W^1(s) - W^1(t_{j-1}) dW^2(s) \right|^2 \right).
\end{aligned}$$

As in the previous sections we use the martingale property of the stochastic Itô-integrals and the Itô-isometry. This yields

$$\begin{aligned}
&= \max_{2\le i\le N} \left[\textstyle\int_{t_{i-1}}^{t_i} s - t_{i-1}\, ds + \frac{1}{9} \int_{t_0}^{t_1} s - t_0\, ds + \frac{4}{9} \sum_{j=2}^{i-1} \int_{t_{j-1}}^{t_j} s - t_{j-1}\, ds \right] \\
&= \left[1 + \frac{1}{9} + \frac{4}{9}(N-2) \right] \frac{1}{2} h^2 \ge \frac{2}{9} T h,
\end{aligned}$$

which proves the assertion for the BDF2-Maruyama method.

Next, we turn to the Itô-Taylor scheme of order $\gamma = \frac{n}{2}$. We consider the $(n+1)$-dimensional SODE

$$\begin{aligned} &dX_1(t) = dW^1(t), dX_2(t) = X_1(t)dW^2(t), \ldots, dX_{n+1}(t) = X_n(t)dW^{n+1}(t),\\ &\quad X(0) = 0 \in \mathbb{R}^{n+1} \end{aligned} \tag{6.3}$$

which has the solution

$$X_1(t) = W^1(t) = I_{(1)}[1]_0^t, X_2(t) = I_{(1,2)}[1]_0^t, \ldots, X_{n+1}(t) = I_{(1,\ldots,n+1)}[1]_0^t, \tag{6.4}$$

where we used the notation from Theorem 5.2. First, as for the BDF2-Maruyama method, one checks that the first n components are exactly approximated by the Itô-Taylor scheme of order $\frac{n}{2}$. In order to keep the notation simple we only do this for the Milstein scheme $(n = 2)$ which is written as

$$\begin{aligned} X_h(0) &= 0 \in \mathbb{R}^3,\\ X_h(t_i) &= X_h(t_{i-1}) + \begin{pmatrix} I_{(1)}^{t_i} \\ X_{h,1}(t_{i-1})I_{(2)}^{t_i} + I_{(1,2)}^{t_i} \\ X_{h,2}(t_{i-1})I_{(3)}^{t_i} + X_{h,1}(t_{i-1})I_{(2,3)}^{t_i} \end{pmatrix}. \end{aligned}$$

For the first component we have

$$X_1(t_j) = I_{(1)}[1]_0^{t_j} = W^1(t_j) = \sum_{i=1}^{j} W^1(t_i) - W^1(t_{i-1}) = \sum_{i=1}^{j} I_{(1)}^{t_i} = X_{h,1}(t_j).$$

For the second component we have

$$\begin{aligned} X_2(t_j) = I_{(1,2)}[1]_0^{t_j} &= I_{(1,2)}[1]_0^{t_{j-1}} + \int_{t_{j-1}}^{t_j} W^1(s)dW^2(s)\\ &= I_{(1,2)}[1]_0^{t_{j-1}} + W^1(t_{j-1})I_{(2)}^{t_j} + \int_{t_{j-1}}^{t_j} W^1(s) - W^1(t_{j-1})dW^2(s)\\ &= I_{(1,2)}[1]_0^{t_{j-1}} + X_{h,1}(t_{j-1})I_{(2)}^{t_j} + I_{(1,2)}^{t_j}. \end{aligned}$$

Now, an inductive arguments yields $X_2(t_j) = X_{h,2}(t_j)$ for all $j = 0, \ldots, N$. For the last component we compute

$$X_3(t_j) = X_3(t_{j-1}) + X_2(t_{j-1})I_{(3)}^{t_j} + X_1(t_{j-1})I_{(2,3)}^{t_j} + I_{(1,2,3)}^{t_j}$$

which shows that the local truncation error of the Milstein method takes the form

$$\|A_h^{ITS} r_h^E X\|_{-1,h}^2 = \mathbb{E}\left(\max_{1\le i\le N} \left|\sum_{j=1}^{i} I_{(1,2,3)}^{t_j}\right|^2\right)$$

For the general Itô-Taylor scheme of order $\gamma = \frac{n}{2}$, one can prove analogously

$$\|A_h^{ITS} r_h^E X\|_{-1,h}^2 = \mathbb{E}\left(\max_{1\le i\le N} \left|\sum_{j=1}^{i} I_{(1,\ldots,n+1)}^{t_j}\right|^2\right).$$

By using the martingale property and the Itô-isometry we arrive at the lower bound

$$\|A_h^{ITS} r_h^E X\|_{-1,h} \ge \left(\frac{T}{(n+1)!}\right)^{\frac{1}{2}} h^\gamma.$$

□

7. Conclusions. In this paper we presented a unifying theory for the convergence analysis of stochastic onestep and multistep methods. We derived a two-sided error estimate which we used to discuss the maximum order of convergence for the stochastic theta method, BDF2-Maruyama and the higher order Itô-Taylor schemes.

One important root of this theory is our notion of consistency and bistability which stems from the abstract framework of discrete approximations as it is formulated by F. Stummel. Then, the proof of bistability relied on our choice of the function spaces and norms, in particular on the stochastic version of Spijker's norm. The usefulness of this norm originates from the fact that we are allowed to interchange summation and expectation in the L^2-norm by the martingale property of the stochastic integrals. That is why we do not need an additional consistency condition on the mean of the numerical methods, in contrast to a different approach in the literature by G.N. Milstein [19]. Note that, in a sense, the Spijker norm is of $W^{-1,\infty}$ Sobolev type.

The characterization of the bistability uses well-known techniques for deterministic multistep methods. After the reduction to the linear operator L_h the same arguments hold for every multistep method with values in an abstract Banach space.

Therefore, we hope to carry over some elements of our theory to infinite dimensional problems. In particular, we are optimistic to prove similar results for stochastic delay equations. But in Subsection 4.1 it was shown that the constant $L > 0$ in assumption (S1) may grow with the number m of Wiener processes. It is not yet clear how to solve this problem if our notion is applied to a numerical scheme for stochastic partial differential equations, e.g. the stochastic heat equation on the real line with white noise.

REFERENCES

[1] Ludwig Arnold, *Stochastic differential equations: theory and applications*, Wiley-Interscience [John Wiley & Sons], New York, 1974. Translated from the German.

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

[3] Wolf-Jürgen Beyn and Raphael Kruse, *Two-sided error estimates for the stochastic theta method*, tech. report, CRC 701, Bielefeld University, 2009. (to appear in DCDS B).

[4] Evelyn Buckwar and Renate Winkler, *Multistep methods for SDEs and their application to problems with small noise*, SIAM J. Numer. Anal., 44 (2006), pp. 779–803 (electronic).

[5] ———, *Multi-step Maruyama methods for stochastic delay differential equations*, Stoch. Anal. Appl., 25 (2007), pp. 933–959.

[6] J. M. C. Clark and R. J. Cameron, *The maximum rate of convergence of discrete approximations for stochastic differential equations*, in Stochastic differential systems (Proc. IFIP-WG 7/1 Working Conf., Vilnius, 1978), vol. 25 of Lecture Notes in Control and Information Sci., Springer, Berlin, 1980, pp. 162–171.

[7] Rolf Dieter Grigorieff, *Diskrete Approximation von Eigenwertproblemen. I. Qualitative Konvergenz*, Numer. Math., 24 (1975), pp. 355–374.

[8] ———, *Diskrete Approximation von Eigenwertproblemen. II. Konvergenzordnung*, Numer. Math., 24 (1975), pp. 415–433.

[9] ———, *Diskrete Approximation von Eigenwertproblemen. III. Asymptotische Entwicklungen*, Numer. Math., 25 (1975/76), pp. 79–97.

[10] ———, *Numerik gewöhnlicher Differentialgleichungen. Band 2*, B. G. Teubner, Stuttgart, 1977. Mehrschrittverfahren, Unter Mitwirkung von Hans Joachim Pfeiffer, Teubner Studienbücher: Mathematik.

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

[12] Desmond J. Higham, *Mean-square and asymptotic stability of the stochastic theta method*, SIAM J. Numer. Anal., 38 (2000), pp. 753–769 (electronic).

[13] A. Jentzen, P. E. Kloeden, and A. Neuenkirch, *Pathwise convergence of numerical schemes for random and stochastic differential equations*, in Foundation of Computational Mathematics, F. Cucker, A. Pinkus, and M. Todd, eds., Cambridge University Press, 2009, Hong Kong, 2008, pp. 140–161.

[14] A. Jentzen, P. E. Kloeden, and A. Neuenkirch, *Pathwise approximation of stochastic differential equations on domains: higher order convergence rates without global Lipschitz coefficients*, Numer. Math., 112 (2009), pp. 41–64.

[15] P. E. Kloeden and A. Neuenkirch, *The pathwise convergence of approximation schemes for stochastic differential equations*, LMS J. Comput. Math., 10 (2007), pp. 235–253 (electronic).

[16] Peter E. Kloeden and Eckhard Platen, *Numerical solution of stochastic differential equations*, vol. 23 of Applications of Mathematics (New York), Springer-Verlag, Berlin, 1992.

[17] Raphael Kruse, *Discrete approximation of stochastic differential equations*, in Bol. Soc. Esp. Mat. Apl., Proc. Conf., NSDS 09, Sevilla, 2009. (to appear).

[18] Xuerong Mao, *Stochastic differential equations and their applications*, Horwood Publishing Series in Mathematics & Applications, Horwood Publishing Limited, Chichester, 1997.

[19] G. N. Milstein, *Numerical integration of stochastic differential equations*, vol. 313 of Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1995. Translated and revised from the 1988 Russian original.

[20] G. N. Milstein and M. V. Tretyakov, *Stochastic numerics for mathematical physics*, Scientific Computation, Springer-Verlag, Berlin, 2004.

[21] Bernt Øksendal, *Stochastic differential equations*, Universitext, Springer-Verlag, Berlin, sixth ed., 2003. An introduction with applications.

[22] Tobias Rydén and Magnus Wiktorsson, *On the simulation of iterated Itô integrals*, Stochastic Process. Appl., 91 (2001), pp. 151–168.

[23] Marc Nico Spijker, *Stability and convergence of finite-difference methods*, vol. 1968 of Doctoral dissertation, University of Leiden, Rijksuniversiteit te Leiden, Leiden, 1968.

[24] ———, *On the structure of error estimates for finite-difference methods*, Numer. Math., 18 (1971/72), pp. 73–100.

[25] Hans J. Stetter, *Analysis of discretization methods for ordinary differential equations*, Springer-Verlag, New York, 1973. Springer Tracts in Natural Philosophy, Vol. 23.

[26] Friedrich Stummel, *Diskrete Konvergenz linearer Operatoren. I*, Math. Ann., 190 (1970/71), pp. 45–92.

[27] ———, *Diskrete Konvergenz linearer Operatoren. II*, Math. Z., 120 (1971), pp. 231–264.

[28] ———, *Diskrete Konvergenz linearer Operatoren. III*, in Linear operators and approximation (Proc. Conf., Oberwolfach, 1971), Birkhäuser, Basel, 1972, pp. 196–216. Internat. Ser. Numer. Math., Vol. 20.

[29] ———, *Approximation methods in analysis*, Matematisk Institut, Aarhus Universitet, Aarhus, 1973. Lectures delivered during the Spring Semester, 1973, Lecture Notes Series, No. 35.

[30] Magnus Wiktorsson, *Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions*, Ann. Appl. Probab., 11 (2001), pp. 470–487.

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