SIAM J. Numer. Anal.
Vol. 47, No. 1, pp. 181–203

B–SERIES ANALYSIS OF STOCHASTIC RUNGE–KUTTA METHODS THAT USE AN ITERATIVE SCHEME TO COMPUTE THEIR INTERNAL STAGE VALUES*

KRISTIAN DEBRABANT† AND ANNE KVÆRNØ‡

Abstract. In recent years, implicit stochastic Runge–Kutta (SRK) methods have been developed both for strong and weak approximations. For these methods, the stage values are only given implicitly. However, in practice these implicit equations are solved by iterative schemes such as simple iteration, modified Newton iteration or full Newton iteration. We employ a unifying approach for the construction of stochastic B-series which is valid both for Itô- and Stratonovich-stochastic differential equations (SDEs) and applicable both for weak and strong convergence to analyze the order of the iterated Runge–Kutta method. Moreover, the analytical techniques applied in this paper can be of use in many other similar contexts.

Key words. stochastic Runge–Kutta method, composite method, stochastic differential equation, iterative scheme, order, internal stage values, Newton's method, weak approximation, strong approximation, growth functions, stochastic B-series

AMS subject classifications. 65C30, 60H35, 65C20, 68U20

DOI. 10.1137/070704307

1. Introduction. In many applications, e.g., in epidemiology and financial mathematics, taking stochastic effects into account when modeling dynamical systems often leads to stochastic differential equations (SDEs). Therefore, in recent years, the development of numerical methods for the approximation of SDEs has become a field of increasing interest; see, e.g., [16, 22] and references therein. Many stochastic schemes fall into the class of stochastic Runge–Kutta (SRK) methods. SRK methods have been studied both for strong approximation [1, 10, 11, 16], where one is interested in obtaining good pathwise solutions, and for weak approximation [8, 9, 16, 19, 21, 32], which focuses on the expectation of functionals of solutions. Order conditions for these methods are found by comparing series of the exact and the numerical solutions. In this paper, we will concentrate on the use of B-series and rooted trees. Such series are surprisingly general; as formal series they are independent of the choice of the stochastic integral, Itô or Stratonovich, or whether weak or strong convergence is considered. This is demonstrated in section 2. For solving SDEs which are stiff, implicit SRK methods have to be considered, which also has been done both for strong [4, 11, 12] and weak [7, 12, 17] approximation. For these methods, the stage values are only given implicitly. However, in practice these implicit equations are solved by iterative schemes like simple iteration or some kind of Newton iterations. For the numerical solution of ODEs such iterative schemes have been studied [13, 14], and it was shown that certain growth functions defined on trees give a precise description of the development of the iterations. Exactly the same growth functions apply to SRKs, as we prove in section 3. Only when these results

*Received by the editors October 2, 2007; accepted for publication (in revised form) June 9, 2008; published electronically October 29, 2008.
http://www.siam.org/journals/sinum/47-1/70430.html

†Fachbereich Mathematik, Technische Universität Darmstadt, Schloßgartenstr.7, D-64289 Darmstadt, Germany (debrabant@mathematik.tu-darmstadt.de).
‡Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway (anne@math.ntnu.no).

are interpreted in terms of the order of the overall scheme, we distinguish Itô from Stratonovich SDEs, weak from strong convergence. This is discussed in sections 4 and 5.

When considering strong convergence, it is difficult to implement fully implicit SRK methods in combination with Newton iterations due to the possible singularity of the numerical procedure. Therefore, various techniques have been developed to circumvent this problem [4]. One possibility is the use of so-called truncated random variables, which have finite distribution and can approximate the increment of Wiener processes to a chosen order [4, 23]. As the concrete choice of random variables in the numerical methods is not specified in this paper, all considerations are without any change also valid for SRK methods with such modified random variables.

Another possibility is to use composite methods [31], which are combinations of a semi-implicit SRK and an implicit SRK. Based on the results for conventional SRK methods, convergence results for iterated composite methods are given in section 6. Finally, in section 7 we present two numerical examples.

Let $(\Omega, \mathcal{A}, \mathcal{P})$ be a probability space. We denote by $(X(t))_{t\in I}$ the stochastic process which is the solution of either a d-dimensional Itô or Stratonovich SDE defined by

$$X(t) = x_0 + \int_{t_0}^{t} g_0(X(s))ds + \sum_{l=1}^{m} \int_{t_0}^{t} g_l(X(s)) \star dW_l(s) \tag{1.1}$$

with an m-dimensional Wiener process $(W(t))_{t\geq 0}$ and $I = [t_0, T]$. The integral w.r.t. the Wiener process has to be interpreted either as Itô integral with $\star dW_l(s) = dW_l(s)$ or as Stratonovich integral with $\star dW_l(s) = \circ dW_l(s)$. We assume that the Borel-measurable coefficients $g_l : \mathbb{R}^d \to \mathbb{R}^d$ are sufficiently differentiable and satisfy a Lipschitz and a linear growth condition. For Stratonovich SDEs, we require in addition that the g_l are differentiable and that also the vectors $g_l' g_l$ satisfy a Lipschitz and a linear growth condition. Then the existence and uniqueness theorem [15] applies.

To simplify the presentation, we define $W_0(s) = s$, so that (1.1) can be written as

$$X(t) = x_0 + \sum_{l=0}^{m} \int_{t_0}^{t} g_l(X(s)) \star dW_l(s). \tag{1.2}$$

In the following we denote by Ξ a set of families of measurable mappings,

$$\Xi := \big\{\{\varphi(h)\}_{h\geq 0} : \ \varphi(h) : \ \Omega \to \mathbb{R} \text{ is } \mathcal{A}-\mathcal{B}\text{-measurable } \forall h \geq 0\big\},$$

and by Ξ_0 the subset of Ξ defined by

$$\Xi_0 := \big\{\{\varphi(h)\}_{h\geq 0} \in \Xi : \ \varphi(0) \equiv 0\big\}.$$

Let a discretization $I^h = \{t_0, t_1, \ldots, t_N\}$ with $t_0 < t_1 < \ldots < t_N = T$ of the time interval I with step sizes $h_n = t_{n+1} - t_n$ for $n = 0, 1, \ldots, N-1$ be given. Now, we consider the general class of s-stage SRK methods given by $Y_0 = x_0$ and

$$Y_{n+1} = Y_n + \sum_{l=0}^{m} \sum_{\nu=0}^{M} \left(z^{(l,\nu)^\top} \otimes I_d \right) g_l \left(H^{(l,\nu)} \right) \tag{1.3a}$$

for $n = 0, 1, \ldots, N-1$ with $Y_n = Y(t_n)$, $t_n \in I^h$, $I_d \in \mathbb{R}^{d,d}$ the identity matrix, and

$$H^{(l,\nu)} = \mathbb{1}_s \otimes Y_n + \sum_{r=0}^{m} \sum_{\mu=0}^{M} \left(Z^{(l,\nu)(r,\mu)} \otimes I_d \right) g_r \left(H^{(r,\mu)} \right) \tag{1.3b}$$

for $l = 0, \ldots, m$ and $\nu = 0, \ldots, M$ with $\mathbb{1}_s = (1, \ldots, 1)^\top \in \mathbb{R}^s$,

$$g_l\left(H^{(l,\nu)}\right) = \left(g_l\left(H_1^{(l,\nu)}\right)^\top, \ldots, g_l\left(H_s^{(l,\nu)}\right)^\top\right)^\top$$

and

$$z^{(l,\nu)} \in \Xi_0^s, \qquad Z^{(l,\nu)(r,\mu)} \in \Xi_0^{s,s}$$

for $l, r = 0, \ldots, m$, $\mu, \nu = 0, \ldots, M$.

The formulation (1.3) is a slight generalization of the class considered in [27] and includes the classes of SRK methods considered in [4, 11, 18, 20, 28, 29, 30] as well as the SRK methods considered in [12, 16, 25]. It defines a d-dimensional approximation process Y with $Y_n = Y(t_n)$.

Application of an iterative scheme yields

$$\begin{aligned} H_{k+1}^{(l,\nu)} &= \mathbb{1}_s \otimes Y_n + \sum_{r=0}^{m}\sum_{\mu=0}^{M}\left(Z^{(l,\nu)(r,\mu)} \otimes I_d\right) g_r\left(H_k^{(r,\mu)}\right) \\ &\quad + \sum_{r=0}^{m}\sum_{\mu=0}^{M}\left(Z^{(l,\nu)(r,\mu)} \otimes I_d\right) J_k^{(r,\mu)}\left(H_{k+1}^{(r,\mu)} - H_k^{(r,\mu)}\right), \end{aligned} \tag{1.4a}$$

$$Y_{n+1,k} = Y_n + \sum_{l=0}^{m}\sum_{\nu=0}^{M}\left({z^{(l,\nu)}}^\top \otimes I_d\right) g_l\left(H_k^{(l,\nu)}\right) \tag{1.4b}$$

with some approximation $J_k^{(r,\mu)}$ to the Jacobian of $g_r(H_k^{(r,\mu)})$ and a predictor $H_0^{(l,\nu)}$. In the following we assume that (1.4a) can be solved uniquely at least for small enough h_n. We consider simple iterations with $J_k^{(r,\mu)} = 0$ (i.e., predictor-corrector methods), modified Newton iterations with $J_k^{(r,\mu)} = I_s \otimes g_r'(x_0)$, and full Newton iterations.

2. Some notation, definitions, and preliminary results. In this section we introduce some necessary notation and provide a few definitions and preliminary results that will be used later.

2.1. Convergence and consistency. Here we will give the definitions for both weak and strong convergence and results which relate convergence to consistency.

Let $C_P^l(\mathbb{R}^d, \mathbb{R}^{\hat{d}})$ denote the space of all $g \in C^l(\mathbb{R}^d, \mathbb{R}^{\hat{d}})$ fulfilling a polynomial growth condition [16].

DEFINITION 1. *A time discrete approximation $Y = (Y(t))_{t \in I^h}$ converges weakly with order p to X as $h \to 0$ at time $t \in I^h$ if for each $f \in C_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R})$ there exist a constant C_f and a finite $\delta_0 > 0$ such that*

$$|\,\mathrm{E}(f(Y(t))) - \mathrm{E}(f(X(t)))| \le C_f\, h^p$$

holds for each $h \in \,]0, \delta_0[\,$.

Now, let $le_f(h; t, x)$ be the weak local error of the method starting at the point (t, x) with respect to the functional f and step size h, i.e.,

$$le_f(h; t, x) = \mathrm{E}\left(f(Y(t+h)) - f(X(t+h)) | Y(t) = X(t) = x\right).$$

The following theorem due to Milstein [22], which holds also in the case of general one-step methods, shows that, as in the deterministic case, consistency implies convergence.

THEOREM 1. *Suppose the following conditions hold:*

- *The coefficients g_l are continuous, satisfy a Lipschitz condition, and belong to $C_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R}^d)$ for $l = 0, \dots, m$. For Stratonovich SDEs, we require in addition that the g_l are differentiable and that also the vectors $g_l' g_l$ satisfy a Lipschitz condition and belong to $C_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R}^d)$ for $l = 0, \dots, m$.*
- *For sufficiently large r (see, e.g., [22] for details) the moments $\mathrm{E}(\|Y(t_n)\|^{2r})$ exist for $t_n \in I^h$ and are uniformly bounded with respect to N and $n = 0, 1, \dots, N$.*
- *Assume that for all $f \in C_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R})$ there exists a $K \in C_P^0(\mathbb{R}^d, \mathbb{R})$ such that*

$$|le_f(h; t, x)| \leq K(x)\, h^{p+1}$$

is valid for $x \in \mathbb{R}^d$ and $t,\, t+h \in I^h$, i.e., the approximation is weak consistent of order p.

Then the method (1.3) *is convergent of order p in the sense of weak approximation.*

Whereas weak approximation methods are used to estimate the expectation of functionals of the solution, strong approximation methods approach the solution pathwise.

DEFINITION 2. *A time discrete approximation $Y = (Y(t))_{t \in I^h}$ converges strongly, respectively, in the mean square with order p to X as $h \to 0$ at time $t \in I^h$ if there exists a constant C and a finite $\delta_0 > 0$ such that*

$$\mathrm{E}\,\|Y(t) - X(t)\| \leq C\, h^p, \qquad \textit{respectively,} \qquad \sqrt{\mathrm{E}(\|Y(t) - X(t)\|^2)} \leq C\, h^p$$

holds for each $h \in\,]0, \delta_0[\,$.

In this article we will consider convergence in the mean square sense. But as by Jensen's inequality we have

$$(\mathrm{E}\,\|Y(t) - X(t)\|)^2 \leq \mathrm{E}(\|Y(t) - X(t)\|^2),$$

mean square convergence implies strong convergence of the same order.

Now, let $le^m(h; t, x)$ and $le^{ms}(h; t, x)$, respectively, be the mean and mean square local error, respectively, of the method starting at the point (t, x) with respect to the step size h; i.e.,

$$\begin{aligned} le^m(h; t, x) &= \mathrm{E}\left(Y(t+h) - X(t+h) | Y(t) = X(t) = x\right), \\ le^{ms}(h; t, x) &= \sqrt{\mathrm{E}\left((Y(t+h) - X(t+h))^2 | Y(t) = X(t) = x\right)}. \end{aligned}$$

The following theorem due to Milstein [22] which holds also in the case of general one step methods shows that in the mean square convergence case we obtain order p if the mean local error is consistent of order p and the mean square local error is consistent of order $p - \frac{1}{2}$.

THEOREM 2. *Suppose the following conditions hold:*

- *The coefficients g_l are continuous and satisfy a Lipschitz condition for $l = 0, \dots, m$, and $\mathrm{E}(\|X(t_0)\|^2) < \infty$. For Stratonovich SDEs, we require in addition that the g_l are differentiable and that also the vectors $g_l' g_l$ satisfy a Lipschitz condition.*
- *There exists a constant K independent of h such that*

$$\|le^m(h; t, x)\| \leq K \sqrt{1 + \|x\|^2}\, h^{p_1}, \qquad le^{ms}(h; t, x) \leq K \sqrt{1 + \|x\|^2}\, h^{p + \frac{1}{2}}$$

with $p \geq 0$, $p_1 \geq p+1$ *is valid for* $x \in \mathbb{R}^d$, *and* $t, t+h \in I^h$; *i.e., the approximation is consistent in the mean of order* $p_1 - 1 \geq p$ *and in the mean square of order* $p - \frac{1}{2}$.

Then the SRK method (1.3) *is convergent of order* p *in the sense of mean square approximation.*

For Stratonovich SDEs, this result is also obtained by Burrage and Burrage [2].

2.2. Stochastic B-series. In this section we will develop stochastic B-series for the solution of (1.2) as well as for the numerical solution given by (1.3). B-series for deterministic ODEs were introduced by Butcher [6]. Today such series appear as a fundamental tool to do local error analysis on a wide range of problems. B-series for SDEs have been developed by Burrage and Burrage [1, 2] to study strong convergence in the Stratonovich case, by Komori, Mitsui, and Sugiura [20] and Komori [18] to study weak convergence in the Stratonovich case, and by Rößler [26, 27] to study weak convergence in both the Itô and the Stratonovich case. However, the distinction between the Itô and the Stratonovich integrals depends only on the definition of the integrals, not on how the B-series are constructed. Similarly, the distinction between weak and strong convergence depends only on the definition of the local error. Thus, we find it convenient to present a uniform and self-contained theory for the construction of stochastic B-series. We will present results and proofs in a certain detail, since some intermediate results will be used in later sections.

Following the idea of Burrage and Burrage, we introduce the set of colored, rooted trees related to the SDE (1.1), as well as the elementary differentials associated with each of these trees.

Definition 3 (trees). *The set of* $m+1$*-colored, rooted trees*

$$T = \{\emptyset\} \cup T_0 \cup T_1 \cup \cdots \cup T_m$$

is recursively defined as follows:

(a) *The graph* $\bullet_l = [\emptyset]_l$ *with only one vertex of color* l *belongs to* T_l.

Let $\tau = [\tau_1, \tau_2, \ldots, \tau_\kappa]_l$ *be the tree formed by joining the subtrees* $\tau_1, \tau_2, \ldots, \tau_\kappa$ *each by a single branch to a common root of color* l.

(b) *If* $\tau_1, \tau_2, \ldots, \tau_\kappa \in T$, *then* $\tau = [\tau_1, \tau_2, \ldots, \tau_\kappa]_l \in T_l$.

Thus, T_l is the set of trees with an l-colored root, and T is the union of these sets.

Definition 4 (elementary differentials). *For a tree* $\tau \in T$ *the elementary differential is a mapping* $F(\tau) : \mathbb{R}^d \to \mathbb{R}^d$ *defined recursively by*

(a) $F(\emptyset)(x_0) = x_0$,

(b) $F(\bullet_l)(x_0) = g_l(x_0)$,

(c) *If* $\tau = [\tau_1, \tau_2, \ldots, \tau_\kappa]_l \in T_l$, *then*

$$F(\tau)(x_0) = g_l^{(\kappa)}(x_0)\left(F(\tau_1)(x_0), F(\tau_2)(x_0), \ldots, F(\tau_\kappa)(x_0)\right).$$

As will be shown in the following, both the exact and the numerical solutions, including the iterated solutions as we will see later, can formally be written in terms of B-series.

Definition 5 (B-series). *Given a mapping* $\phi : T \to \Xi$ *satisfying*

$$\phi(\emptyset)(h) \equiv 1 \quad \text{and} \quad \phi(\tau)(0) \equiv 0, \quad \forall \tau \in T \backslash \{\emptyset\}.$$

A (stochastic) B-series is then a formal series of the form

$$B(\phi, x_0; h) = \sum_{\tau \in T} \alpha(\tau) \cdot \phi(\tau)(h) \cdot F(\tau)(x_0),$$

where $\alpha : T \to \mathbb{Q}$ *is given by*

$$\alpha(\emptyset) = 1, \qquad \alpha(\bullet_l) = 1, \qquad \alpha(\tau = [\tau_1, \ldots, \tau_\kappa]_l) = \frac{1}{r_1! r_2! \cdots r_q!} \prod_{j=1}^{\kappa} \alpha(\tau_j),$$

where $r_1, r_2, \ldots, r_q$ *count equal trees among* $\tau_1, \tau_2, \ldots, \tau_\kappa$*.*

If $\phi : T \to \Xi^s$, then $B(\phi, x_0; h) = [B(\phi_1, x_0; h), \ldots, B(\phi_s, x_0; h)]^\top$.

The next lemma proves that if $Y(h)$ can be written as a B-series, then $f(Y(h))$ can be written as a similar series, where the sum is taken over trees with a root of color f and subtrees in T. The lemma is fundamental for deriving B-series for the exact and the numerical solution. It will also be used for deriving weak convergence results.

LEMMA 3. *If* $Y(h) = B(\phi, x_0; h)$ *is some B-series and* $f \in C^\infty(\mathbb{R}^d, \mathbb{R}^{\hat{d}})$*, then* $f(Y(h))$ *can be written as a formal series of the form*

$$f(Y(h)) = \sum_{u \in U_f} \beta(u) \cdot \psi_\phi(u)(h) \cdot G(u)(x_0), \tag{2.1}$$

where U_f *is a set of trees derived from* T*, by*

(a) $[\emptyset]_f \in U_f$*, and if* $\tau_1, \tau_2, \ldots, \tau_\kappa \in T$*, then* $[\tau_1, \tau_2, \ldots, \tau_\kappa]_f \in U_f$*.*

(b) $G([\emptyset]_f)(x_0) = f(x_0)$ *and*
$G(u = [\tau_1, \ldots, \tau_\kappa]_f)(x_0) = f^{(\kappa)}(x_0)\big(F(\tau_1)(x_0), \ldots, F(\tau_\kappa)(x_0)\big)$*.*

(c) $\beta([\emptyset]_f) = 1$ *and* $\beta(u = [\tau_1, \ldots, \tau_\kappa]_f) = \frac{1}{r_1! r_2! \ldots r_q!} \prod_{j=1}^{\kappa} \alpha(\tau_j)$*, where* $r_1, r_2, \ldots, r_q$ *count equal trees among* $\tau_1, \tau_2, \ldots, \tau_\kappa$*.*

(d) $\psi_\phi([\emptyset]_f)(h) \equiv 1$ *and* $\psi_\phi(u = [\tau_1, \ldots, \tau_\kappa]_f)(h) = \prod_{j=1}^{\kappa} \phi(\tau_j)(h)$*.*

Proof. Writing $Y(h)$ as a B-series, we have

$$\begin{aligned} f(Y(h)) =& f\left(\sum_{\tau \in T} \alpha(\tau) \cdot \phi(\tau)(h) \cdot F(\tau)(x_0)\right) \\ =& \sum_{\kappa=0}^{\infty} \frac{1}{\kappa!} f^{(\kappa)}(x_0) \left(\sum_{\tau \in T \setminus \{\emptyset\}} \alpha(\tau) \cdot \phi(\tau)(h) \cdot F(\tau)(x_0)\right)^{\kappa} \\ =& f(x_0) + \sum_{\kappa=1}^{\infty} \frac{1}{\kappa!} \sum_{\{\tau_1, \tau_2, \ldots, \tau_\kappa\} \in T \setminus \{\emptyset\}} \frac{\kappa!}{r_1! r_2! \cdots r_q!} \\ & \cdot \left(\prod_{j=1}^{\kappa} \alpha(\tau_j) \cdot \phi(\tau_j)(h)\right) f^{(\kappa)}(x_0)\big(F(\tau_1)(x_0), \ldots, F(\tau_\kappa)(x_0)\big), \end{aligned}$$

where the last sum is taken over all possible unordered combinations of κ trees in T. For each set of trees $\tau_1, \tau_2, \ldots, \tau_\kappa \in T$ we assign a $u = [\tau_1, \tau_2, \ldots, \tau_\kappa]_f \in U_f$. The theorem is now proved by comparing term by term with (2.1). ☐

Remark 1. For example, in the definition of weak convergence, just $f \in C_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R})$ is required. Thus $f(Y(h))$ can be written only as a truncated B-series, with a remainder term. However, to simplify the presentation in the following we assume that all derivatives of $f, g_0, \ldots, g_l$ exist.

We will also need the following result.

LEMMA 4. *If* $Y(h) = B(\phi_Y, x_0; h)$ *and* $Z(h) = B(\phi_Z, x_0; h)$ *and* $f \in C^\infty(\mathbb{R}^d, \mathbb{R}^{\hat{d}})$*, then*

$$f'(Y(h))Z(h) = \sum_{u \in U_f} \beta(u) \cdot \Upsilon(u)(h) \cdot G(u)(x_0)$$

with

$$\Upsilon([\emptyset]_f)(h) \equiv 0, \qquad \Upsilon([u = [\tau_1, \dots, \tau_\kappa]_f)(h) = \sum_{i=1}^{\kappa} \left(\prod_{\substack{j=1 \\ j \neq i}}^{\kappa} \phi_Y(\tau_j)(h) \right) \phi_Z(\tau_i)(h)$$

with $\beta(u)$ *and* $G(u)(x_0)$ *given by Lemma* 3. The proof is similar to the deterministic case (see [24]).

When Lemma 3 is applied to the functions g_l on the right-hand side of (1.2) we get the following result: If $Y(h) = B(\phi, x_0; h)$, then

$$g_l(Y(h)) = \sum_{\tau \in T_l} \alpha(\tau) \cdot \phi_l'(\tau)(h) \cdot F(\tau)(x_0) \tag{2.2}$$

in which

$$\phi_l'(\tau)(h) = \begin{cases} 1 & \text{if } \tau = \bullet_l, \\ \prod_{j=1}^{\kappa} \phi(\tau_j)(h) & \text{if } \tau = [\tau_1, \dots, \tau_\kappa]_l \in T_l. \end{cases}$$

THEOREM 5. *The solution* $X(t_0 + h)$ *of* (1.2) *can be written as a B-series* $B(\varphi, x_0; h)$ *with*

$$\varphi(\emptyset)(h) \equiv 1, \quad \varphi(\bullet_l)(h) = W_l(h), \quad \varphi(\tau = [\tau_1, \dots, \tau_\kappa]_l)(h) = \int_0^h \prod_{j=1}^{\kappa} \varphi(\tau_j)(s) \star dW_l(s).$$

Proof. Write the exact solution as some B-series $X(t_0 + h) = B(\varphi, x_0; h)$. By (2.2) the SDE (1.2) can be written as

$$\sum_{\tau \in T} \alpha(\tau) \cdot \varphi(\tau)(h) \cdot F(\tau)(x_0) = x_0 + \sum_{l=0}^{m} \sum_{\tau \in T_l} \alpha(\tau) \cdot \int_0^h \varphi_l'(\tau)(s) \star dW_l(s) \cdot F(\tau)(x_0).$$

Comparing term by term we get

$$\varphi(\emptyset)(h) \equiv 1, \quad \text{and} \quad \varphi(\tau)(h) = \int_0^h \varphi'(\tau)(s) \star dW_l(s) \quad \text{for} \quad \tau \in T_l, \quad l = 0, 1, \dots, m.$$

The proof is completed by induction on τ. ▯

The same result is given for the Stratonovich case in [2, 18], but it clearly also applies to the Itô case.

The definition of the order of the tree, $\rho(\tau)$, is motivated by the fact that $\mathrm{E}\, W_l(h)^2 = h$ for $l \geq 1$ and $W_0(h) = h$.

DEFINITION 6 (order). *The order of a tree* $\tau \in T$ *is defined by*

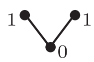
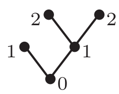
$$\rho(\emptyset) = 0, \quad \rho([\tau_1, \dots, \tau_\kappa]_f) = \sum_{i=1}^{\kappa} \rho(\tau_i)$$

and

$$\rho(\tau = [\tau_1, \dots, \tau_\kappa]_l) = \sum_{i=1}^{\kappa} \rho(\tau_i) + \begin{cases} 1 & \textit{for } l = 0, \\ \frac{1}{2} & \textit{otherwise.} \end{cases}$$

TABLE 2.1

Examples of trees and corresponding functions $\rho(\tau)$, $\alpha(\tau)$, and $\varphi(\tau)$. The integrals $\varphi(\tau)$ are also expressed in terms of multiple integrals $J_{(\dots)}$ for the Stratonovich (S) *and $I_{(\dots)}$ for the Itô* (I) *cases; see* [16] *for their definition. In bracket notation, the trees will be written as $\bullet_l$, $[[\bullet_2]_0]_1$, $[\bullet_1, \bullet_1]_0$, and $[\bullet_1, [\bullet_2, \bullet_2]_1]_0$, respectively.*

τ	$\rho(\tau)$	$\alpha(\tau)$	$\varphi(\tau)(h)$
$\bullet_l$	$\begin{cases} 1 & \text{if } l = 0 \\ \frac{1}{2} & \text{if } l \neq 0 \end{cases}$	1	$W_l(h) = \begin{cases} h & \text{if } l = 0 \\ J_{(l)} & \text{(S)} \\ I_{(l)} & \text{(I)} \end{cases}$
	2	1	$\int_0^h \int_0^{s_1} W_2(s_2) \star ds_2 \star dW_1(s_1) = \begin{cases} J_{(2,0,1)} & \text{(S)} \\ I_{(2,0,1)} & \text{(I)} \end{cases}$
	2	$\frac{1}{2}$	$\int_0^h W_1(s)^2 \star ds = \begin{cases} 2J_{(1,1,0)} & \text{(S)} \\ 2I_{(1,1,0)} + I_{(0,0)} & \text{(I)} \end{cases}$
	3	$\frac{1}{2}$	$\int_0^h W_1(s_1) \left(\int_0^{s_1} W_2(s_2)^2 \star dW_1(s_2) \right) \star ds_1 = \begin{cases} 4J_{(2,2,1,1,0)} + 2J_{(2,1,2,1,0)} + 2J_{(1,2,2,1,0)} & \text{(S)} \\ 4I_{(2,2,1,1,0)} + 2I_{(2,1,2,1,0)} + 2I_{(1,2,2,1,0)} & \\ + 2I_{(0,1,1,0)} + 2I_{(2,2,0,0)} + I_{(1,0,1,0)} + I_{(0,0,0)} & \text{(I)} \end{cases}$

In Table 2.1 some trees and the corresponding values for the functions ρ, α, and φ are presented.

The following result is similar to results given in [1].

THEOREM 6. *If the coefficients $Z^{(l,\nu),(r,\mu)} \in \Xi_0^{s,s}$ and $z^{(l,\nu)} \in \Xi_0^s$, then the numerical solution Y_1 as well as the stage values can be written in terms of B-series*

$$H^{(l,\nu)} = B\left(\Phi^{(l,\nu)}, x_0; h\right), \qquad Y_1 = B(\Phi, x_0; h)$$

for all l, ν, with

$$\Phi^{(l,\nu)}(\emptyset)(h) \equiv \mathbb{1}_s, \quad \Phi^{(l,\nu)}(\bullet_r)(h) = \sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \mathbb{1}_s, \tag{2.3a}$$

$$\Phi^{(l,\nu)}(\tau = [\tau_1, \dots, \tau_\kappa]_r)(h) = \sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \prod_{j=1}^{\kappa} \Phi^{(r,\mu)}(\tau_j)(h) \tag{2.3b}$$

and

$$\Phi(\emptyset)(h) \equiv 1, \quad \Phi(\bullet_l)(h) = \sum_{\nu=0}^{M} z^{(l,\nu)} \mathbb{1}_s, \tag{2.4a}$$

$$\Phi(\tau = [\tau_1, \dots, \tau_\kappa]_l)(h) = \sum_{\nu=0}^{M} z^{(l,\nu)} \prod_{j=1}^{\kappa} \Phi^{(l,\nu)}(\tau_j)(h). \tag{2.4b}$$

Proof. Write $H^{(l,\nu)}$ as a B-series, that is

$$H^{(l,\nu)} = \sum_{\tau \in T} \alpha(\tau) \left(\Phi^{(l,\nu)}(h) \otimes I_d\right) \left(\mathbb{1}_s \otimes F(\tau)(x_0)\right).$$

Use the definition of the method (1.3) together with (2.2) to obtain

$$H^{(l,\nu)} = \mathbb{1}_s \otimes x_0 + \sum_{r=0}^{m} \sum_{\mu=0}^{M} \sum_{\tau \in T_r} \alpha(\tau) \left(\left(Z^{(l,\nu)(r,\mu)} \cdot \left(\Phi^{(r,\mu)} \right)'_r (\tau)(h) \right) \otimes I_d \right) (\mathbb{1}_s \otimes F(\tau)(x_0))$$

with $(\Phi^{(r,\mu)})'_r(\tau)(h) = ((\Phi_1^{(r,\mu)})'_r(\tau)(h), \ldots, (\Phi_s^{(r,\mu)})'_r(\tau)(h))^\top$. Comparing term-by-term gives the relations (2.3). The proof of (2.4) is similar. □

To decide the weak order we will also need the B-series of the function f, evaluated at the exact and the numerical solution. From Theorem 5, Theorem 6, and Lemma 3 we obtain

$$f(X(t_0 + h)) = \sum_{u \in U_f} \beta(u) \cdot \psi_\varphi(u)(h) \cdot G(u)(x_0),$$

$$f(Y_1) = \sum_{u \in U_f} \beta(u) \cdot \psi_\Phi(u)(h) \cdot G(u)(x_0),$$

with

$$\psi_\varphi([\emptyset]_f)(h) \equiv 1, \quad \psi_\varphi(u = [\tau_1, \ldots, \tau_\kappa]_f)(h) = \prod_{j=1}^{\kappa} \varphi(\tau_j)(h)$$

and

$$\psi_\Phi([\emptyset]_f)(h) \equiv 1, \quad \psi_\Phi(u = [\tau_1, \ldots, \tau_\kappa]_f)(h) = \prod_{j=1}^{\kappa} \Phi(\tau_j)(h).$$

So, for the weak local error it follows

$$le_f(h; t, x) = \sum_{u \in U_f} \beta(u) \cdot \mathrm{E}\left[\psi_\Phi(u)(h) - \psi_\varphi(u)(h)\right] \cdot G(u)(x).$$

For the mean and mean square local error we obtain from Theorem 5 and Theorem 6,

$$le^{ms}(h; t, x) = \sqrt{\mathrm{E}\left(\sum_{\tau \in T} \alpha(\tau) \cdot (\Phi(\tau)(h) - \varphi(\tau)(h)) \cdot F(\tau)(x) \right)^2},$$

$$le^m(h; t, x) = \sum_{\tau \in T} \alpha(\tau) \cdot \mathrm{E}\left(\Phi(\tau)(h) - \varphi(\tau)(h)\right) \cdot F(\tau)(x).$$

With all the B-series in place, we can now present the order conditions for the weak and strong convergence for both the Itô and the Stratonovich case.[1] We have weak

[1] As usual we assume that method (1.3) is constructed such that $E\psi_\varphi(u)(h) = \mathcal{O}(h^{\rho(u)})$ $\forall u \in U_f$ and $\varphi(\tau)(h) = \mathcal{O}(h^{\rho(\tau)})$ $\forall \tau \in T$, respectively, where especially in the latter expression the $\mathcal{O}(\cdot)$-notation refers to the $L^2(\Omega)$-norm and $h \to 0$. These conditions are fulfilled if for $i, j = 1, \ldots, s$, $k \in \mathbb{N} = \{0, 1, \ldots\}$ it holds that

$$(z_i^{(l,\nu)})^{2^k} = \begin{cases} \mathcal{O}(h^{(2^k)}) & l = 0 \\ \mathcal{O}(h^{(2^{k-1})}) & l > 0 \end{cases}, \quad (Z_{ij}^{(l,\nu)(r,\mu)})^{2^k} = \begin{cases} \mathcal{O}(h^{(2^k)}) & l = 0 \\ \mathcal{O}(h^{(2^{k-1})}) & l > 0 \end{cases}.$$

consistency of order p if and only if

$$\mathrm{E}\,\psi_\Phi(u)(h) = \mathrm{E}\,\psi_\varphi(u)(h) + \mathcal{O}(h^{p+1}) \quad \forall u \in U_f \text{ with } \rho(u) \le p + \frac{1}{2}, \tag{2.5}$$

where $\rho(u = [\tau_1, \dots, \tau_\kappa]_f) = \sum_{j=1}^{\kappa} \rho(\tau_j)$ ((2.5) slightly weakens conditions given in [27]), and mean square global order p if [4]

$$\Phi(\tau)(h) = \varphi(\tau)(h) + \mathcal{O}\left(h^{p+\frac{1}{2}}\right) \quad \forall \tau \in T \text{ with } \rho(\tau) \le p, \tag{2.6}$$

$$\mathrm{E}\,\Phi(\tau)(h) = \mathrm{E}\,\varphi(\tau)(h) + \mathcal{O}\left(h^{p+1}\right) \quad \forall \tau \in T \text{ with } \rho(\tau) \le p + \frac{1}{2} \tag{2.7}$$

and all elementary differentials $F(\tau)$ fulfill a linear growth condition. Instead of the last requirement it is also enough to claim that there exists a constant C such that $\|g_j'(y)\| \le C \quad \forall y \in \mathbb{R}^m$, $j = 0, \dots, M$ (which implies the global Lipschitz condition) and all necessary partial derivatives exist [2].

3. B-series of the iterated solution and growth functions. In this section we will discuss how the iterated solution defined in (1.4) can be written in terms of B-series, that is,

$$H_k^{(l,\nu)} = B\left(\Phi_k^{(l,\nu)}, x_0; h\right) \quad \text{and} \quad Y_{1,k} = B(\Phi_k, x_0; h).$$

For notational convenience, in the following the h-dependency of the weight functions will be suppressed, so $\Phi(\tau)(h)$ will be written as $\Phi(\tau)$. Further, all results are valid for all $l = 0, \dots, m$ and $\nu = 0, \dots, M$.

Assume that the predictor can be written as a B-series,

$$H_0^{(l,\nu)} = B\left(\Phi_0^{(l,\nu)}, x_0; h\right),$$

satisfying $\Phi_0^{(l,\nu)}(\emptyset) = \mathbb{1}_s$ and $\Phi_0^{(l,\nu)}(\tau) = \mathcal{O}(h^{\rho(\tau)})\ \forall \tau \in T$. The most common situation is the use of the trivial predictor $H^{(l,\nu)} = \mathbb{1}_s \otimes x_0$, for which $\Phi_0^{(l,\nu)}(\emptyset) = \mathbb{1}_s$ and $\Phi_0^{(l,\nu)}(\tau) = 0$ otherwise.

The iteration schemes we discuss here differ only in the choice of $J_k^{(r,\mu)}$ in (1.4). For all schemes, the following lemma applies. The proof follows directly from Lemma 3.

LEMMA 7. *If $H_k^{(l,\nu)} = B(\Phi_k^{(l,\nu)}, x_0; h)$, then $Y_{1,k} = B(\Phi_k, x_0; h)$ with*

$$\Phi_k(\emptyset) \equiv 1, \quad \Phi_k(\bullet_l) = \sum_{\nu=0}^{M} z^{(l,\nu)} \mathbb{1}_s, \quad \Phi_k(\tau = [\tau_1, \dots, \tau_\kappa]_l) = \sum_{\nu=0}^{M} z^{(l,\nu)} \prod_{j=1}^{\kappa} \Phi_k^{(l,\nu)}(\tau_j).$$

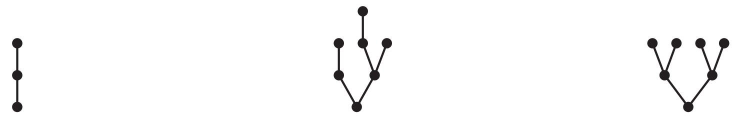
Further,

$$f(Y_{1,k}) = \sum_{u \in U_f} \beta(u) \cdot \psi_{\Phi_k}(u) \cdot G(u)(x_0)$$

with

$$\psi_{\Phi_k}([\emptyset]_f) = 1, \quad \psi_{\Phi_k}(u = [\tau_1, \dots, \tau_\kappa]_f) = \prod_{j=1}^{\kappa} \Phi_k(\tau_j),$$

where $\beta(u)$ and $G(u)(x_0)$ are given in Lemma 3.



$\mathfrak{h}(\tau) = 3,\ \mathfrak{r}(\tau) = \mathfrak{d}(\tau) = 1;\quad \mathfrak{h}(\tau) = 4,\ \mathfrak{r}(\tau) = 3,\ \mathfrak{d}(\tau) = 2;\quad \mathfrak{h}(\tau) = \mathfrak{r}(\tau) = \mathfrak{d}(\tau) = 3$

FIG. 3.1. *Examples of trees and their growth functions for simple* ($\mathfrak{h}$)*, modified Newton* ($\mathfrak{r}$)*, and full Newton* ($\mathfrak{d}$) *iterations.*

We are now ready to study each of the iteration schemes. In each case, we will first find the recurrence formula for $\Phi_k^{(l,\nu)}(\tau)$. From this we define a growth function $\mathfrak{g}(\tau)$.

DEFINITION 7 (growth function). *A growth function* $\mathfrak{g} : T \to \mathbb{N}$ *is a function satisfying*

$$\begin{aligned} &\Phi_k^{(l,\nu)}(\tau) = \Phi^{(l,\nu)}(\tau), \quad \forall \tau \in T, \quad \mathfrak{g}(\tau) \le k \\ &\qquad \Rightarrow \quad \Phi_{k+1}^{(l,\nu)}(\tau) = \Phi^{(l,\nu)}(\tau), \quad \forall \tau \in T, \quad \mathfrak{g}(\tau) \le k+1, \end{aligned} \tag{3.1}$$

for all $k \ge 0$.

This result should be sharp in the sense that in general there exists $\tau \ne \emptyset$ with $\Phi_0^{(l,\nu)}(\tau) \ne \Phi^{(l,\nu)}(\tau)$ and $\Phi_k^{(l,\nu)}(\tau) \ne \Phi^{(l,\nu)}(\tau)$ when $k < \mathfrak{g}(\tau)$. From Lemma 7 we also have

$$\begin{aligned} \Phi_k(\tau) = \Phi(\tau) \qquad & \forall \tau = [\tau_1, \ldots, \tau_\kappa]_l \in T, \quad \mathfrak{g}'(\tau) = \max_{j=1}^{\kappa} \mathfrak{g}(\tau_i) \le k, \\ \psi_{\Phi_k}(u) = \psi_{\Phi(\tau)} \qquad & \forall u = [\tau_1, \ldots, \tau_\kappa]_f \in U_f, \quad \mathfrak{g}'(u) = \max_{j=1}^{\kappa} \mathfrak{g}'(\tau_i) \le k. \end{aligned} \tag{3.2}$$

The growth functions give a precise description of the development of the iterations. As we will see, the growth functions are exactly the same as in the deterministic case (see [13, 14]). However, to get applicable results, we will at the end need the relation between the growth functions and the order. Further, we will also take advantage of the fact that $\mathrm{E}\,\Phi(\tau) = 0$ and $\mathrm{E}\,\psi_\Phi(u) = 0$ for some trees. These aspects are discussed in the next sections. Examples of trees and the values of the growth functions for the three iteration schemes are given in Figure 3.1.

The simple iteration. Simple iterations are described by (1.4a) with $J_k^{(r,\mu)} = 0$, that is,

$$H_{k+1}^{(l,\nu)} = \mathbb{1}_s \otimes x_0 + \sum_{r=0}^{m} \sum_{\mu=0}^{M} \left(Z^{(l,\nu)(r,\mu)} \otimes I_d \right) g_r \left(H_k^{(r,\mu)} \right). \tag{3.3}$$

By (2.2) and Theorem 6 we easily get the next two results.

LEMMA 8. *If* $H_0^{(l,\nu)} = B(\Phi_0^{(l,\nu)}, x_0; h)$*, then* $H_k^{(l,\nu)} = B(\Phi_k^{(l,\nu)}, x_0; h)$*, where*

$$\Phi_{k+1}^{(l,\nu)}(\emptyset) \equiv \mathbb{1}_s, \quad \Phi_{k+1}^{(l,\nu)}(\tau = [\tau_1, \ldots, \tau_\kappa]_r) = \sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \prod_{j=1}^{\kappa} \Phi_k^{(r,\mu)}(\tau_j).$$

The corresponding growth function is given by

$$\mathfrak{h}(\emptyset) = 0, \quad \mathfrak{h}([\tau_1, \ldots, \tau_\kappa]_l) = 1 + \max_{j=1}^{\kappa} \mathfrak{h}(\tau_j).$$

The function $\mathfrak{h}(\tau)$ is the height of τ, that is, the maximum number of nodes along one branch. The functions $\mathfrak{h}'(\tau)$ and $\mathfrak{h}'(u)$ are defined by (3.2), with $\mathfrak{g}$ replaced by $\mathfrak{h}$.

The modified Newton iteration. In this subsection, we consider the modified Newton iteration (1.4a) with $J_k^{(r,\mu)} = I_s \otimes g_r'(x_0)$. The B-series for $H_k^{(l,\nu)}$ and the corresponding growth function can now be described by the following lemma.

LEMMA 9. *If* $H_0^{(l,\nu)} = B(\Phi_0^{(l,\nu)}, x_0; h)$, *then* $H_k^{(l,\nu)} = B(\Phi_k^{(l,\nu)}, x_0; h)$ *with*

$$\Phi_{k+1}^{(l,\nu)}(\emptyset) \equiv \mathbb{1}_s,$$

$$\Phi_{k+1}^{(l,\nu)}(\tau) = \begin{cases} \sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \prod_{j=1}^{\kappa} \Phi_k^{(r,\mu)}(\tau_j) & \text{if } \tau = [\tau_1, \dots, \tau_\kappa]_r \in T \text{ and } \kappa \geq 2, \\ \sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \Phi_{k+1}^{(r,\mu)}(\tau_1) & \text{if } \tau = [\tau_1]_r \in T. \end{cases} \tag{3.4}$$

The corresponding growth function is given by

$$\mathfrak{r}(\emptyset) = 0, \qquad \mathfrak{r}(\bullet_l) = 1, \qquad \mathfrak{r}(\tau = [\tau_1, \dots, \tau_\kappa]_l) = \begin{cases} \mathfrak{r}(\tau_1) & \text{if} \quad \kappa = 1, \\ 1 + \max_{j=1}^{\kappa} \mathfrak{r}(\tau_j) & \text{if} \quad \kappa \geq 2. \end{cases}$$

The function $\mathfrak{r}(\tau)$ is one plus the maximum number of ramifications along any branch of the tree.

Proof. The iteration scheme (1.4a) can be rewritten in B-series notation as

$$\begin{aligned} &\sum_{\tau \in T} \alpha(\tau) \cdot \Phi_{k+1}^{(l,\nu)}(\tau) \otimes F(\tau)(x_0) = \mathbb{1} \otimes x_0 \\ &\quad + \sum_{r=0}^{m} \sum_{\tau \in T_r} \alpha(\tau) \cdot \left(\sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \left(\Phi_k^{(r,\mu)} \right)_r' (\tau) \right) \otimes F(\tau)(x_0) \\ &\quad + \sum_{r=0}^{m} \sum_{\tau_1 \in T_r} \alpha(\tau_1) \cdot \left(\sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \left(\Phi_{k+1}^{(r,\mu)}(\tau_1) - \Phi_k^{(r,\mu)}(\tau_1) \right) \right) \otimes (g_r'(x_0) F(\tau_1)(x_0)), \end{aligned} \tag{3.5}$$

where we have used (2.2). Clearly, $\Phi_{k+1}^{(l,\nu)}(\emptyset) \equiv \mathbb{1}_s$ for all $k \geq 0$ and

$$\Phi_{k+1}^{(l,\nu)}(\bullet_r) = \sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \mathbb{1}_s,$$

proving the lemma for $\tau = \bullet_r = [\emptyset]_r$. Next, let $\tau = [\tau_1]_r$, where $\tau_1 \neq \emptyset$. Then $F(\tau)(x_0) = g_r'(x_0) F(\tau_1)$. Comparing equal terms on both sides of the equation, using $\alpha(\tau) = \alpha(\tau_1)$, we get

$$\Phi_{k+1}^{(l,\nu)}(\tau) = \sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \left(\left(\Phi_k^{(r,\mu)} \right)_r' (\tau) + \Phi_{k+1}^{(r,\mu)}(\tau_1) - \Phi_k^{(r,\mu)}(\tau_1) \right).$$

Since $(\Phi_k^{(r,\mu)})_r'(\tau) = \Phi_k^{(r,\mu)}(\tau_1)$ the lemma is proved for all $\tau = [\tau_1]_r$. For $\tau = [\tau_1, \dots, \tau_\kappa]_r$ with $\kappa \geq 2$ the last sum of (3.5) contributes nothing, thus

$$\Phi_{k+1}^{(l,\nu)}(\tau) = \sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \left(\Phi_k^{(r,\mu)} \right)_r' (\tau),$$

which concludes the proof of (3.4).

The second statement of the lemma is obviously true for $\tau = \emptyset$. Let τ be any tree satisfying $\mathfrak{r}(\tau) \le k+1$. Then either $\tau = [\tau_1]_l$ with $\mathfrak{r}(\tau_1) \le k+1$ or $\tau = [\tau_1, \dots, \tau_\kappa]_l$ with $\kappa \ge 2$ and $\mathfrak{r}(\tau_i) \le k$. In the latter case, we have by the hypothesis, by (3.4) and Theorem 6, that

$$\Phi_{k+1}^{(l,\nu)}(\tau) = \sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \prod_{j=1}^{\kappa} \Phi^{(r,\mu)}(\tau_j) = \Phi^{(l,\nu)}(\tau).$$

In the first case, it follows easily by induction on τ that $\Phi_{k+1}^{(l,\nu)}(\tau) = \Phi^{(l,\nu)}(\tau)$ since $\Phi_{k+1}^{(l,\nu)}(\tau) = \sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \Phi_{k+1}^{(r,\mu)}(\tau_1)$. □

The full Newton iteration. In this subsection, we consider the full Newton iteration (1.4a) with

$$J_k^{(r,\mu)} = g_r'\left(H_k^{(r,\mu)}\right).$$

It follows that the B-series for $H_k^{(l,\nu)}$ and the corresponding growth function satisfy.

LEMMA 10. *If $H_0^{(l,\nu)} = B(\Phi_0^{(l,\nu)}, x_0; h)$, then $H_k^{(l,\nu)} = B(\Phi_k^{(l,\nu)}, x_0; h)$ with*

$$\begin{aligned} &\Phi_{k+1}^{(l,\nu)}(\emptyset) \equiv \mathbb{1}_s, \\ &\Phi_{k+1}^{(l,\nu)}(\tau) = \sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \prod_{j=1}^{\kappa} \Phi_k^{(r,\mu)}(\tau_j) \\ &\qquad + \sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \sum_{i=1}^{\kappa} \left(\prod_{\substack{j=1 \\ j\neq i}}^{\kappa} \Phi_k^{(r,\mu)}(\tau_j) \right) \left(\Phi_{k+1}^{(r,\mu)}(\tau_i) - \Phi_k^{(r,\mu)}(\tau_i) \right), \end{aligned} \tag{3.6}$$

where $\tau = [\tau_1, \dots, \tau_\kappa]_r$ and the rightmost $\prod$ is taken to be $\mathbb{1}_s$ if $\kappa = 1$. The corresponding growth function is given by

$$\mathfrak{d}(\emptyset) = 0, \qquad \mathfrak{d}(\bullet_l) = 1,$$

$$\mathfrak{d}(\tau = [\tau_1, \dots, \tau_\kappa]_l) = \begin{cases} \max_{j=1}^{\kappa} \mathfrak{d}(\tau_j) & \text{if} \quad \gamma = 1, \\ \max_{j=1}^{\kappa} \mathfrak{d}(\tau_j) + 1 & \text{if} \quad \gamma \ge 2, \end{cases}$$

where γ is the number of subtrees in τ satisfying $\mathfrak{d}(\tau_i) = \max_{j=1}^{\kappa} \mathfrak{d}(\tau_j)$.

The function $\mathfrak{d}$ is called the doubling index of τ.

Proof. Using (2.2) and Lemma 4 the scheme (1.4a) can be written as

$$\begin{aligned} &\sum_{\tau \in T} \alpha(\tau) \cdot \Phi_{k+1}^{(l,\nu)}(\tau) \otimes F(\tau)(x_0) = \mathbb{1} \otimes x_0 \\ &\qquad + \sum_{r=0}^{m} \sum_{\tau \in T_r} \alpha(\tau) \cdot \left(\sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \left(\Phi_k^{(r,\mu)}\right)_r' (\tau) \right) \otimes F(\tau)(x_0) \\ &\qquad + \sum_{r=0}^{m} \sum_{u \in U_{g_r}} \beta(u) \cdot \left(\sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \Upsilon_k^{(r,\mu)}(u) \right) \otimes G(u)(x_0), \end{aligned} \tag{3.7}$$

where

$$\Upsilon_k^{(r,\mu)}(u = [\tau_1, \dots, \tau_\kappa]_{g_r}) = \sum_{i=1}^{\kappa} \left(\prod_{\substack{j=1 \\ j\neq i}}^{\kappa} \Phi_k^{(r,\mu)}(\tau_j) \right) \left(\Phi_{k+1}^{(r,\mu)}(\tau_i) - \Phi_k^{(r,\mu)}(\tau_i) \right).$$

From the definition of $F(\tau)$, $G(u = [\tau_1, \ldots, \tau_\kappa]_{g_r})(x_0) = F(\tau = [\tau_1, \ldots, \tau_\kappa]_r)(x_0)$. The sum over all $u \in U_{g_r}$ can be replaced by the sum over all $\tau \in T_r$, and the result is proved. Next, we will prove that $\mathfrak{d}(\tau)$ satisfies the implication (3.1) given in Definition 7. We will do so by induction on $n(\tau)$, the number of nodes in τ. Since $\emptyset$ is the only tree satisfying $n(\tau) = 0$, and $\Phi_{k+1}^{(r,\mu)}(\emptyset) = \Phi^{(r,\mu)}(\emptyset) \equiv \mathbb{1}_s$, the conclusion of (3.1) is true for all $\tau \in T$ with $n(\tau) = 0$. Let $\bar{n} \geq 1$ and assume by the induction hypothesis that the conclusion of (3.1) holds for any tree satisfying $\mathfrak{d}(\tau) \leq k+1$ and $n(\tau) < \bar{n}$. We will show that $\Phi_{k+1}^{(r,\mu)}(\bar{\tau}) = \Phi^{(r,\mu)}(\bar{\tau})$ for all $\bar{\tau}$ satisfying $\mathfrak{d}(\bar{\tau}) \leq k+1$ and $n(\bar{\tau}) \leq \bar{n}$. Let $\bar{\tau} = [\tau_1, \ldots, \tau_\kappa]_l$ where $n(\tau_j) < \bar{n}$ for $j = 1, \ldots, \kappa$. Since $\mathfrak{d}(\bar{\tau}) \leq k+1$ there is at most one subtree τ_j satisfying $\mathfrak{d}(\tau_j) = k+1$, let us for simplicity assume this to be the last one. Thus $\mathfrak{d}(\tau_j) \leq k$ for $j = 1, \ldots, \kappa-1$ and $\mathfrak{d}(\tau_\kappa) \leq k+1$. Consequently, $\Phi_k^{(r,\mu)}(\tau_j) = \Phi^{(r,\mu)}(\tau_j)$, $j = 1, \ldots, \kappa-1$ by the hypothesis of (3.1), and $\Phi_{k+1}^{(r,\mu)}(\tau_j) = \Phi^{(r,\mu)}(\tau_j)$, $j = 1, \ldots, \kappa$ by the induction hypothesis. By (3.6) and Theorem 6,

$$\begin{aligned}
\Phi_{k+1}^{(l,\nu)}(\bar{\tau}) = &\sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \prod_{j=1}^{\kappa} \Phi_k^{(r,\mu)}(\tau_j) \\
&+ \sum_{\mu=0}^{M} \sum_{i=1}^{\kappa} Z^{(l,\nu)(r,\mu)} \left(\prod_{\substack{j=1 \\ j\neq i}}^{\kappa} \Phi_k^{(r,\mu)}(\tau_j) \right) \left(\Phi_{k+1}^{(r,\mu)}(\tau_i) - \Phi_k^{(r,\mu)}(\tau_i) \right) \\
= &\sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \left(\prod_{j=1}^{\kappa-1} \Phi_k^{(r,\mu)}(\tau_j) \right) \Phi_k^{(r,\mu)}(\tau_\kappa) \\
&+ \sum_{\mu=0}^{M} Z^{(l,\nu)(r,\mu)} \left(\prod_{j=1}^{\kappa-1} \Phi^{(r,\mu)}(\tau_j) \right) \left(\Phi^{(r,\mu)}(\tau_\kappa) - \Phi_k^{(r,\mu)}(\tau_\kappa) \right) \\
= &\Phi^{(l,\nu)}(\bar{\tau}),
\end{aligned}$$

completing the induction proof. ▯

4. General convergence results for iterated methods. Now we will relate the results of the previous section to the order of the overall scheme. In the following, we assume that the predictors satisfy the conditions

$$\begin{aligned}
&\Phi_0^{(l,\nu)}(\tau) = \Phi^{(l,\nu)}(\tau) && \forall \tau \in T \text{ with } \mathfrak{g}(\tau) \leq \mathcal{G}_0, \\
&\Phi_0^{(l,\nu)}(\tau) \in \left\{ \Phi^{(l,\nu)}(\tau), 0 \right\} && \forall \tau \in T \text{ with } \mathfrak{g}(\tau) \leq \hat{\mathcal{G}}_0,
\end{aligned} \tag{4.1}$$

where $\mathcal{G}_0$ and $\hat{\mathcal{G}}_0$ are chosen as large as possible. In particular, the trivial predictor satisfies $\mathcal{G}_0 = 0$ while $\hat{\mathcal{G}}_0 = \infty$. We assume further that in analogy to (3.1) we have

$$\begin{aligned}
&\Phi_k^{(l,\nu)}(\tau) \in \left\{ \Phi^{(l,\nu)}(\tau), 0 \right\}, \quad \forall \tau \in T, \quad \mathfrak{g}(\tau) \leq k \\
&\quad \Rightarrow \quad \Phi_{k+1}^{(l,\nu)}(\tau) \in \left\{ \Phi^{(l,\nu)}(\tau), 0 \right\}, \quad \forall \tau \in T, \quad \mathfrak{g}(\tau) \leq k+1,
\end{aligned} \tag{4.2}$$

for all $k \geq 0$. By Lemmas 8, 9, and 10 this is guaranteed for the iteration schemes considered here.

It follows from (3.1), (3.2), and (4.2) that

$$(4.3)\qquad \begin{aligned} &\Phi_k(\tau) = \Phi(\tau) && \forall \tau \in T \text{ with } \mathfrak{g}'(\tau) \le \mathcal{G}_0 + k,\\ &\Phi_k(\tau) \in \{\Phi(\tau), 0\} && \forall \tau \in T \text{ with } \mathfrak{g}'(\tau) \le \hat{\mathcal{G}}_0 + k \end{aligned}$$

as well as

$$(4.4)\qquad \begin{aligned} &\psi_{\Phi_k}(u) = \psi_\Phi(u) && \forall u \in U_f \text{ with } \mathfrak{g}'(u) \le \mathcal{G}_0 + k,\\ &\psi_{\Phi_k}(\tau) \in \{\psi_\Phi(u), 0\} && \forall u \in U_f \text{ with } \mathfrak{g}'(u) \le \hat{\mathcal{G}}_0 + k. \end{aligned}$$

The next step is to establish the relation between the order and the growth function of a tree. We have chosen to do so by some maximum height functions, given by

(4.5)
$$\mathcal{G}_T(q) = \max_{\tau\in T}\{\mathfrak{g}'(\tau) : \rho(\tau) \le q\}, \qquad \mathcal{G}_{T,\varphi}(q) = \max_{\tau\in T}\{\mathfrak{g}'(\tau) : \mathrm{E}\,\varphi(\tau) \ne 0, \rho(\tau) \le q\},$$
$$\mathcal{G}_{U_f}(q) = \max_{u\in U_f}\{\mathfrak{g}'(u) : \rho(u) \le q\}, \quad \mathcal{G}_{U_f,\psi_\varphi}(q) = \max_{u\in U_f}\{\mathfrak{g}'(u) : \mathrm{E}\,\psi_\varphi(u) \ne 0,\ \rho(u) \le q\}.$$

Note that the definition relates to the weights of the exact, not the numerical, solution. We are now ready to establish results on weak and strong convergence for the iterated solution.

Weak convergence. Let p be the weak order of the underlying scheme. The weak order of the iterated solution after k iterations is $\min(q_k, p)$ if

$$\mathrm{E}\,\psi_{\Phi_k}(u) = \mathrm{E}\,\psi_\Phi(u) \qquad \forall u \in U_f, \quad \rho(u) \le q_k + \frac{1}{2}.$$

If $q_k \le p$ we can take advantage of the fact that $0 = \mathrm{E}\,\psi_\varphi(u) = \mathrm{E}\,\psi_\Phi(u) + \mathcal{O}(h^{p+1})$ for some u, and thereby relax the conditions to

$$(4.6)\qquad \begin{aligned} &\psi_{\Phi_k}(u) = \psi_\Phi(u) && \forall u \in U_f \text{ with } \mathrm{E}\,\psi_\varphi(u) \ne 0,\\ &\psi_{\Phi_k}(u) \in \{\psi_\Phi(u), 0\} && \forall u \in U_f \text{ with } \mathrm{E}\,\psi_\varphi(u) = 0. \end{aligned}$$

By (4.4), this is fulfilled for all u of order $\rho(u) \le \min(q_k, p)$ if

$$\mathcal{G}_{U_f,\Psi_\varphi}\left(q_k + \frac{1}{2}\right) \le \mathcal{G}_0 + k \qquad \text{and} \qquad \mathcal{G}_{U_f}\left(q_k + \frac{1}{2}\right) \le \hat{\mathcal{G}}_0 + k.$$

The results can then be summarized in the following theorem.

THEOREM 11. *The iterated method is of weak order $q_k \le p$ after*

$$\max\left\{\mathcal{G}_{U_f,\psi_\varphi}\left(q_k + \frac{1}{2}\right) - \mathcal{G}_0, \mathcal{G}_{U_f}\left(q_k + \frac{1}{2}\right) - \hat{\mathcal{G}}_0\right\}$$

iterations.

Strong convergence. The strong convergence case can be treated similarly. Let p now be the mean square order of the underlying method. The iterated solution is of order $\min(p, q_k)$ if

$$(4.7)\qquad \begin{aligned} &\Phi_k(\tau) = \Phi(\tau) \quad \forall \tau \in T \text{ with } \rho(\tau) \le q_k,\\ &\Phi_k(\tau) = \Phi(\tau) \quad \forall \tau \in T \text{ with } \rho(\tau) = q_k + \frac{1}{2}, \quad \mathrm{E}\,\phi(\tau) \ne 0,\\ &\Phi_k(\tau) \in \{\Phi(\tau), 0\} \quad \forall \tau \in T \text{ with } \rho(\tau) = q_k + \frac{1}{2}, \quad \mathrm{E}\,\phi(\tau) = 0. \end{aligned}$$

According to (4.3) these are satisfied if all the conditions

$$\mathcal{G}_T(q_k) \le \mathcal{G}_0 + k, \qquad \mathcal{G}_T\left(q_k + \frac{1}{2}\right) \le \hat{\mathcal{G}}_0 + k, \qquad \text{and} \qquad \mathcal{G}_{T,\varphi}\left(q_k + \frac{1}{2}\right) \le \mathcal{G}_0 + k$$

are satisfied. We can summarize this by the following theorem.

THEOREM 12. *The iterated method is of mean square order $q_k \le p$ after*

$$\max\left\{\max\left\{\mathcal{G}_T(q_k), \mathcal{G}_{T,\varphi}\left(q_k + \frac{1}{2}\right)\right\} - \mathcal{G}_0, \mathcal{G}_T\left(q_k + \frac{1}{2}\right) - \hat{\mathcal{G}}_0\right\}$$

iterations.

5. Growth functions and order. In this section we will discuss the relation between the order of trees and the growth functions defined in section 3. Let us start with the following lemma.

LEMMA 13. *For $k \ge 1$,*

$$\begin{aligned} \mathfrak{h}'(\tau) = k \quad &\Rightarrow \quad \rho(\tau) \ge \frac{k}{2} + \frac{1}{2}, \\ \mathfrak{r}'(\tau) = k \quad &\Rightarrow \quad \rho(\tau) \ge k, \\ \mathfrak{d}'(\tau) = k \quad &\Rightarrow \quad \rho(\tau) \ge 2^{k-1}. \end{aligned}$$

The same result is valid for $\mathfrak{h}'(u)$, $\mathfrak{r}'(u)$, and $\mathfrak{g}'(u)$.

Proof. Let $\mathcal{T}_{\mathfrak{h},k}$, $\mathcal{T}_{\mathfrak{r},k}$, and $\mathcal{T}_{\mathfrak{d},k}$ be sets of trees of minimal order satisfying $\mathfrak{h}(\tau) = k$ $\forall \tau \in \mathcal{T}_{\mathfrak{h},k}$, $\mathfrak{r}(\tau) = k$ $\forall \tau \in \mathcal{T}_{\mathfrak{r},k}$, and $\mathfrak{d}(\tau) = k$ $\forall \tau \in \mathcal{T}_{\mathfrak{d},k}$ (see Figure 5.1), and denote this minimal order by $\rho_{\mathfrak{h},k}$, $\rho_{\mathfrak{r},k}$ and $\rho_{\mathfrak{d},k}$. Minimal order trees are built up only by stochastic nodes. It follows immediately that $\mathcal{T}_{\mathfrak{h},1} = \mathcal{T}_{\mathfrak{r},1} = \mathcal{T}_{\mathfrak{d},1} = \{\bullet_l : \ l \ge 1\}$. Since $\rho(\bullet_l) = 1/2$ for $l \ge 1$, the results are proved for $k = 1$. It is easy to show by induction on k that

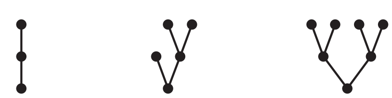
$$\begin{aligned} &\mathcal{T}_{\mathfrak{h},k} = \{[\tau]_l : \ \tau \in \mathcal{T}_{\mathfrak{h},k-1}, \ l \ge 1\}, && \rho_{\mathfrak{h},k} = \rho_{\mathfrak{h},k-1} + \frac{1}{2} = \frac{k}{2}, \\ &\mathcal{T}_{\mathfrak{r},k} = \{[\bullet_{l_1}, \tau]_{l_2} : \ \tau \in \mathcal{T}_{\mathfrak{r},k-1}, \ l_1, l_2 \ge 1\}, && \rho_{\mathfrak{r},k} = \rho_{\mathfrak{r},k-1} + 1 = k - \frac{1}{2}, \\ &\mathcal{T}_{\mathfrak{d},k} = \{[\tau_1, \tau_2]_l : \ \tau_1, \tau_2 \in \mathcal{T}_{\mathfrak{d},k-1}, \ l \ge 1\}, && \rho_{\mathfrak{d},k} = 2\rho_{\mathfrak{d},k-1} + \frac{1}{2} = 2^{k-1} - \frac{1}{2}. \end{aligned} \tag{5.1}$$

For each $\mathfrak{g}$ being either $\mathfrak{h}$, $\mathfrak{r}$, or $\mathfrak{d}$, the minimal order trees satisfying $\mathfrak{g}'(\tau'_{\mathfrak{g},k}) = k$, $\mathfrak{g}'(u_{\mathfrak{g},k}) = k$ are $\tau'_{\mathfrak{g},k} = [\tau_{\mathfrak{g},k}]_l$ with $\tau_{\mathfrak{g},k} \in \mathcal{T}_{\mathfrak{g},k}$ and $l \ge 1$, and $u_{\mathfrak{g},k} = [\tau'_{\mathfrak{g},k}]_f$. Both are of order $\rho(\tau_{\mathfrak{g},k}) + 1/2$. ∎

Let $\mathcal{G}_T(q)$ and $\mathcal{G}_{U_f}(q)$ be defined by (4.5). Then the following corollary holds.

COROLLARY 14. *For $q \ge \frac{1}{2}$ we have*

$$\mathcal{G}_T(q) = \mathcal{G}_{U_f}(q) = \begin{cases} 2q - 1 & \textit{for simple iterations,} \\ \lfloor q \rfloor & \textit{for modified Newton iterations,} \\ \lfloor \log_2(q) \rfloor + 1 & \textit{for full Newton iterations.} \end{cases}$$



$\mathcal{T}_{\mathfrak{h},3}$ $\mathcal{T}_{\mathfrak{r},3}$ $\mathcal{T}_{\mathfrak{d},3}$

FIG. 5.1. *Minimal order trees with $\mathfrak{g}(\tau) = 3$. The sets $\mathcal{T}_{\mathfrak{g},3}$ consist of all such trees with only stochastic nodes.*

Proof. The minimal order trees are also the maximum height/ramification number/doubling index trees, in the sense that as long as $\rho(\tau'_{\mathfrak{g},k}) \le q < \rho(\tau'_{\mathfrak{g},k+1})$ there are no trees of order q for which the growth function can exceed k. □

Let $T^S \subset T$ and $U_f^S \subset U_f$ be the set of trees with an even number of each kind of stochastic nodes. Further, let $T^I \subset T_0$ and $U_f^I \subset U_f$ be the set of trees constructed from the root ($\bullet_0$ or $\bullet_f$), by a finite number of steps of the form:

(i) add one deterministic node, or

(ii) add two equal stochastic nodes, neither of them being a father of the other.

Clearly $T^I \subset T^S$ and $U_f^I \subset U_f^S$. From [5, 26] we have

$$(5.2)\qquad \begin{aligned} \mathrm{E}\,\varphi(\tau) = 0 \quad \text{if} \quad \tau \notin &\begin{cases} T^S & \text{in the Stratonovich case,} \\ T^I & \text{in the Itô case,} \end{cases} \\ \mathrm{E}\,\psi_\varphi(u) = 0 \quad \text{if} \quad u \notin &\begin{cases} U_f^S & \text{in the Stratonovich case,} \\ U_f^I & \text{in the Itô case.} \end{cases} \end{aligned}$$

Considering only trees for which $\mathrm{E}\,\varphi$ or $\mathrm{E}\,\psi_\varphi$ are different from zero, we get the following lemma.

LEMMA 15. *For $k \ge 1$,*

$$\mathfrak{h}'(\tau) = k \quad \Rightarrow \quad \rho(\tau) \ge \begin{cases} \lceil \frac{k+1}{2} \rceil & \textit{if} \quad \tau \in T^S, \\ k+1 & \textit{if} \quad \tau \in T^I, \end{cases}$$

$$\mathfrak{r}'(\tau) = k \quad \Rightarrow \quad \rho(\tau) \ge \begin{cases} k & \textit{if} \quad \tau \in T^S, \\ k+1 & \textit{if} \quad \tau \in T^I, \end{cases}$$

$$\mathfrak{d}'(\tau) = k \quad \Rightarrow \quad \rho(\tau) \ge \begin{cases} 2^{k-1} & \textit{if} \quad \tau \in T^S, \\ 2^{k-1}+1 & \textit{if} \quad \tau \in T^I. \end{cases}$$

This result is also valid for $\mathfrak{h}'(u)$, $\mathfrak{r}'(u)$, and $\mathfrak{g}'(u)$, with $T^{\cdot}$ replaced by $U_f^{\cdot}$.

Proof. In the Stratonovich case, we consider only trees of integer order, which immediately gives the results. In the Itô case, let $\tau_{\mathfrak{g},k}$, $\tau'_{\mathfrak{g},k}$ be the minimal order trees used in the proof of Lemma 13. Unfortunately $\tau'_{\mathfrak{g},k}$ has a stochastic root, so $\tau'_{\mathfrak{g},k} \notin T^I$, and there are no trees $\tau \in T^I$ of order $\rho(\tau_{\mathfrak{g},k}) + 1/2$ satisfying $\mathfrak{g}'(\tau) = k$. When $\mathfrak{g}$ is either $\mathfrak{r}$ or $\mathfrak{d}$ then the tree $[\tau_{\mathfrak{g}}, \bullet_l]_0 \in T^I$ if all the stochastic nodes are of color $l \ge 1$. The order of this tree is $\rho(\tau_{\mathfrak{g}}) + 3/2$, proving the result for $\mathfrak{r}'(\tau)$ and $\mathfrak{d}'(\tau)$. Let $\hat{\tau}'_{\mathfrak{h},k} \in T^I$ be a tree of minimal order satisfying $\mathfrak{h}'(\hat{\tau}'_{\mathfrak{h},k}) = k$. Clearly, $\hat{\tau}'_{\mathfrak{h},1}$ can be either $[\bullet_0]_0$ or $[\bullet_l, \bullet_l]_0$ with $l \ge 1$, both of order 2. From the construction of trees in T^I it is clear that the height of the tree can be increased only by one for each order, thus $\rho(\hat{\tau}'_{\mathfrak{h},k}) = k+1$. The result for U_f^I follows immediately. □

Let $\mathcal{G}_{T,\varphi}(q)$ and $\mathcal{G}_{U_f,\psi_\varphi}(q)$ be given by (4.5). Then the analogue of Corollary 14 is as follows.

COROLLARY 16. *For $q \ge \frac{1}{2}$ we have in the Stratonovich case*

$$\mathcal{G}_{T,\varphi}(q) = \mathcal{G}_{U_f,\psi_\varphi}(q) = \begin{cases} \max\{0, 2\lfloor q \rfloor - 1\} & \textit{for simple iterations,} \\ \lfloor q \rfloor & \textit{for modified Newton iterations,} \\ \lfloor \log_2(q) \rfloor + 1 & \textit{for full Newton iterations,} \end{cases}$$

Table 5.1
Number of iterations needed to achieve order p when using the simple, modified, or full Newton iteration scheme in the Itô or Stratonovich case for strong or weak approximation.

	Stratonovich			Itô					
	Strong/weak appr.			Weak appr.			Strong appr.		
p	simple	mod.	full	simple	mod.	full	simple	mod.	full
$\frac{1}{2}$	1	1	1	0	0	0	0	0	0
1	1	1	1	0	0	0	1	1	1
$1\frac{1}{2}$	3	2	2	1	1	1	2	1	1
2	3	2	2	1	1	1	3	2	2
$2\frac{1}{2}$	5	3	2	2	2	1	4	2	2
3	5	3	2	2	2	1	5	3	2

and in the Itô case

$$\mathcal{G}_{T,\varphi}(q) = \mathcal{G}_{U_f,\psi_\varphi}(q) = \begin{cases} \max\{0, \lfloor q \rfloor - 1\} & \textit{for simple iterations,} \\ \max\{0, \lfloor q \rfloor - 1\} & \textit{for modified Newton iterations,} \\ \max\{0, \lfloor \log_2(q) \rfloor\} & \textit{for full Newton iterations.} \end{cases}$$

For the trivial predictor, Table 5.1 gives the number of iterations needed to achieve a certain order of convergence. The results concerning the Stratonovich case when considering strong approximation and using the simple iteration scheme where already obtained by Burrage and Tian [3] analyzing predictor corrector methods.

6. Convergence results for composite methods. Composite methods have been introduced by Tian and Burrage [31]. At each step either a semi-implicit Runge–Kutta method or an implicit Runge–Kutta method is used in order to obtain better stability properties, which results in the method

$$\begin{aligned} Y_{n+1} = Y_n &+ \lambda_n \sum_{l=0}^{m} \sum_{\nu=0}^{M} \left({z^{(1,l,\nu)}}^\top \otimes I_d \right) g_l \left(H^{(1,l,\nu)} \right) \\ &+ (1-\lambda_n) \sum_{l=0}^{m} \sum_{\nu=0}^{M} \left({z^{(2,l,\nu)}}^\top \otimes I_d \right) g_l \left(H^{(2,l,\nu)} \right) \end{aligned} \tag{6.1a}$$

for $n = 0, 1, \ldots, N-1$, $t_n \in I^h$, $\lambda_n \in \{0, 1\}$ and

$$H^{(j,l,\nu)} = \mathbb{1}_s \otimes Y_n + \sum_{r=0}^{m} \sum_{\mu=0}^{M} \left(Z^{(j,l,\nu)(r,\mu)} \otimes I_d \right) g_r \left(H^{(j,r,\mu)} \right), \qquad j = 1, 2. \tag{6.1b}$$

Here, the method coefficients with superscripts 1 are those of the implicit SRK and the method coefficients with superscripts 2 are those of the semi-implicit SRK. Let $\Phi^{(1)}$, $\Phi^{(2)}$ be the corresponding weight-functions and $\Phi_k^{(1)}$, $\Phi_k^{(2)}$ be the corresponding weight-functions of the iterated methods. Then the weight-function Φ of the composite method is given by $\Phi = \lambda_1 \Phi^{(1)} + (1-\lambda_1)\Phi^{(2)}$, and similarly we have for the iterated method $\Phi_k = \lambda_1 \Phi_k^{(1)} + (1-\lambda_1)\Phi_k^{(2)}$. It follows that the convergence conditions (4.6) and (4.7), respectively, are satisfied if and only if they are satisfied as well for the underlying implicit SRK as for the semi-implicit SRK. Thus, an iterated composite method has the same order as the original composite method, if in each step the number of iterations is chosen according to Theorem 11 and Theorem 12, respectively. For the trivial predictor, the number of iterations needed to achieve a certain order of convergence is again given by Table 5.1.

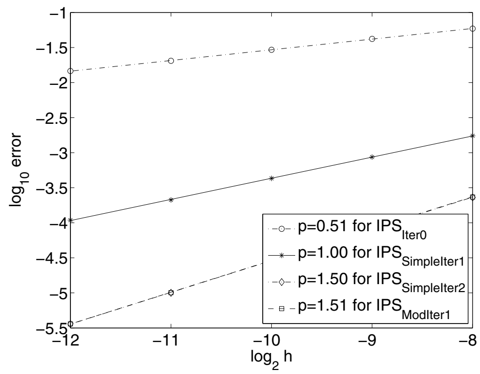


FIG. 7.1. *Error of IPS applied to* (7.1) *without iteration, with one or two simple iterations, and with one modified Newton iteration (the last two results nearly coincide).*

7. Numerical examples. In the following, we analyze numerically the order of convergence of three SRK methods in dependence on the kind and number of iterations. In each example, the solution is approximated with step sizes $2^{-8}, \dots, 2^{-12}$ and the sample average of $M = 20{,}000$ independent simulated realizations of the absolute error is calculated in order to estimate the expectation.

As a first example, we apply the drift implicit strong order 1.5 scheme due to Kloeden and Platen [16], implemented as a stiffly accurate SRK scheme with six stages and denoted by IPS; i.e., for one-dimensional Wiener processes

$$Y_{n+1} = Y_n + \sum_{l=0}^{1} \left({z^{(l)}}^\top \otimes I_d \right) g_l(H), \qquad H = \mathbb{1}_6 \otimes Y_n + \sum_{l=0}^{1} \left(Z^{(l)} \otimes I_d \right) g_l(H),$$

$$Z^{(0)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ h & 0 & 0 & 0 & 0 & 0 \\ h & 0 & 0 & 0 & 0 & 0 \\ h & 0 & 0 & 0 & 0 & 0 \\ h & 0 & 0 & 0 & 0 & 0 \\ \frac{h}{2} & a & -a & 0 & 0 & \frac{h}{2} \end{pmatrix}, \qquad Z^{(1)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \sqrt{h} & 0 & 0 & 0 & 0 & 0 \\ -\sqrt{h} & 0 & 0 & 0 & 0 & 0 \\ \sqrt{h} & \sqrt{h} & 0 & 0 & 0 & 0 \\ \sqrt{h} & -\sqrt{h} & 0 & 0 & 0 & 0 \\ I_{(1)} & b+c & b-c & d & -d & 0 \end{pmatrix},$$

$$z^{(0)} = \left(\frac{h}{2}, a, -a, 0, 0, \frac{h}{2} \right)^\top, \qquad z^{(1)} = \left(I_{(1)}, b+c, b-c, d, -d, 0 \right)^\top,$$

$$a = \frac{I_{(1,0)} - \frac{1}{2} I_{(1)} h}{2\sqrt{h}}, \quad b = \frac{I_{(0,1)}}{2h}, \quad c = \frac{I_{(1,1)}}{2\sqrt{h}} - d, \quad d = \frac{I_{(1,1,1)}}{2h},$$

to the nonlinear SDE [16, 21]

$$dX(t) = \left(\tfrac{1}{2} X(t) + \sqrt{X(t)^2 + 1} \right) dt + \sqrt{X(t)^2 + 1}\, dW(t), \qquad X(0) = 0, \tag{7.1}$$

on the time interval $I = [0, 1]$ with the solution $X(t) = \sinh(t + W(t))$.

The results at time $t = 1$ are presented in Figure 7.1, where the orders of convergence correspond to the slope of the regression lines. As predicted by Table 5.1 we observe strong order 0.5 without iteration, strong order 1.0 for one simple iteration, and strong order 1.5 for two simple or one modified Newton iteration.

As second example, we apply the diagonal implicit strong order 1.5 method DIRK4 which for one-dimensional Wiener processes is given by

$$Y_{n+1} = Y_n + \sum_{l=0}^{1} \left({z^{(l)}}^\top \otimes I_d \right) g_l(H), \qquad H = \mathbb{1}_3 \otimes Y_n + \sum_{l=0}^{1} \left(Z^{(l)} \otimes I_d \right) g_l(H)$$

with coefficients[2]

$$z^0 = h\alpha, \quad z^1 = J_{(1)}\gamma^{(1)} + \frac{J_{(1,0)}}{h}\gamma^{(2)},$$

$$Z^{(0)} = hA, \quad Z^{(1)} = J_{(1)}B^{(1)} + \frac{J_{(1,0)}}{h}B^{(2)} + \sqrt{h}B^{(3)},$$

$$\alpha^\top = (0.169775184, 0.297820839, 0.042159965, 0.490244012),$$

$${\gamma^{(1)}}^\top = (-1.008751744, 0.285118644, 0.760818846, 0.962814254),$$

$${\gamma^{(2)}}^\top = (1.507774621, 1.085932735, -1.458091242, -1.135616114),$$

$$A = \begin{pmatrix} 0.240968725 & 0 & 0 & 0 \\ 0.167810317 & 0.160243373 & 0 & 0 \\ -0.002766912 & 0.473332751 & 0.178081733 & 0 \\ 0.415057712 & 0.115126049 & 0.020652745 & 0.130541130 \end{pmatrix},$$

$$B_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -0.476890860 & 0 & 0 & 0 \\ 0.514160282 & 0.012424879 & 0 & 0 \\ -0.879966702 & 0.412866280 & 0.711524058 & 0 \end{pmatrix},$$

$$B_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1.287951512 & 0 & 0 & 0 \\ 0.665416412 & -0.686930244 & 0 & 0 \\ 0.703868780 & 0.876627859 & -0.321270197 & 0 \end{pmatrix},$$

$$B_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0.568300129 & -0.568300129 & 0 & 0 \\ 1.614193125 & -0.618659748 & -0.995533377 & 0 \\ 0.660721631 & -0.714401673 & -0.896487337 & 0.950167380 \end{pmatrix}$$

to the corresponding Stratonovich version of (7.1). This method is constructed such that the regularity of the linear system which has to be solved in each modified Newton iteration step does not depend directly on $J_{(1)}$ and $J_{(1,0)}$.

The results at time $t = 1$ are presented in Figure 7.2. As predicted by Table 5.1 we observe no convergence without iteration, strong order 1.0 for one or two simple iterations or one modified Newton iteration, and strong order 1.5 in the case of three simple iterations or two modified Newton iterations.

[2]For typographical reasons, we restrict ourselves to an accuracy of $5 \cdot 10^{-10}$. A 16-digits version of the coefficients can be obtained on request from the authors.

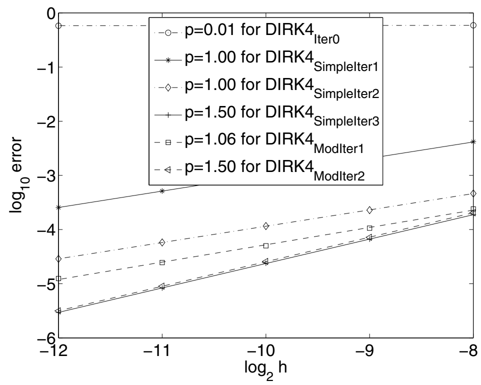


FIG. 7.2. *Error of DIRK4 applied to the Stratonovich version of* (7.1) *without iteration, with one, two, or three simple iterations, and with one or two modified Newton iterations (the results for three simple iterations and two modified Newton iterations nearly coincide).*

Finally, we apply the drift implicit strong order 1.0 scheme due to Kloeden and Platen [16], implemented as a stiffly accurate SRK scheme in the form

$$Y_{n+1} = Y_n + \sum_{l=0}^{m}\sum_{\nu=0}^{m}\left(z^{(l,\nu)^\top} \otimes I_d\right) g_l\left(H^{(\nu)}\right),$$

$$H^{(\nu)} = \mathbb{1}_2 \otimes Y_n + \sum_{l=0}^{m}\sum_{\mu=0}^{m}\left(Z^{(\nu)(l,\mu)} \otimes I_d\right) g_l\left(H^{(\mu)}\right), \quad \nu = 0,\dots,m,$$

$$Z^{(0)(0,0)} = \begin{pmatrix} 0 & 0 \\ \frac{h}{2} & \frac{h}{2}\end{pmatrix}, \qquad Z^{(0)(0,\mu)} = \begin{pmatrix} 0 & 0 \\ 0 & 0\end{pmatrix}, \qquad Z^{(0)(l,0)} = \begin{pmatrix} 0 & 0 \\ I_{(l)} & 0\end{pmatrix},$$

$$Z^{(0)(l,\mu)} = \begin{pmatrix} 0 & 0 \\ -\frac{I_{(\mu,l)}}{\sqrt{h}} & \frac{I_{(\mu,l)}}{\sqrt{h}}\end{pmatrix}, \quad Z^{(j)(0,0)} = \begin{pmatrix} 0 & 0 \\ h & 0\end{pmatrix}, \qquad Z^{(j)(l,l)} = \begin{pmatrix} 0 & 0 \\ \sqrt{h} & 0\end{pmatrix},$$

$$Z^{(j)(l,\mu)} = \begin{pmatrix} 0 & 0 \\ 0 & 0\end{pmatrix} \text{ for } l \neq \mu, \qquad z^{(0,0)} = \left(\tfrac{h}{2}, \tfrac{h}{2}\right)^\top, \qquad z^{(0,\mu)} = (0,0)^\top,$$

$$z^{(l,0)} = \left(I_{(l)}, 0\right)^\top, \qquad z^{(l,\mu)} = \left(-\tfrac{I_{(\mu,l)}}{\sqrt{h}}, \tfrac{I_{(\mu,l)}}{\sqrt{h}}\right)^\top, \qquad j,l,\mu = 1,\dots,m,$$

and denoted by IPS10 to the following nonlinear problem of dimension two driven by two Wiener processes in which there is no commutativity between the driving terms,

$$dX_1(t) = \left(\tfrac{1}{2}X_1(t) + \sqrt{X_1(t)^2 + X_2(t)^2 + 1}\right) dt + \sqrt{X_2(t)^2+1}\, dW_1(t)$$

$$+ \cos X_1(t)\, dW_2(t), \tag{7.2a}$$

$$dX_2(t) = \left(\tfrac{1}{2}X_1(t) + \sqrt{X_2(t)^2 + 1}\right) dt + \sqrt{X_1(t)^2+1}\, dW_1(t) + \sin X_2(t)\, dW_2(t), \tag{7.2b}$$

$$X(0) = 0. \tag{7.2c}$$

As here we don't know the exact solution, to approximate it we use IPS10 with two modified Newton iterations and a step size ten times smaller than the actual step size. The multiple Itô integrals are approximated as described in [16]. The results at time

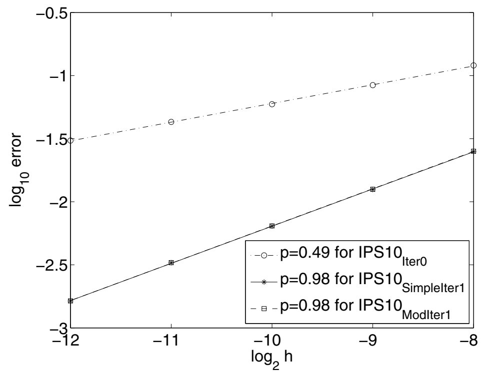


FIG. 7.3. *Error of IPS*10 *applied to* (7.2) *without iteration, with one simple iteration and with one modified Newton iteration (the last two results nearly coincide).*

$t = 1$ are presented in Figure 7.3. As predicted by Table 5.1 we observe strong order 0.5 without iteration and strong order 1.0 for one simple iteration or one modified Newton iteration.

8. Conclusion. For stochastic Runge–Kutta methods that use an iterative scheme to compute their internal stage values, we derived convergence results based on the order of the underlying Runge–Kutta method, the choice of the iteration method, the predictor, and the number of iterations. This was done by developing a unifying approach for the construction of stochastic B-series, which is valid both for Itô- and Stratonovich-SDEs and can be used both for weak and strong convergence. We expect this to be useful also for the further development and analysis of stochastic Runge–Kutta type methods.

Acknowledgement. We thank the anonymous referees for their valuable comments and suggestions, which helped to improve this paper.

REFERENCES

[1] K. Burrage and P. M. Burrage, *High strong order explicit Runge–Kutta methods for stochastic ordinary differential equations*, Appl. Numer. Math., 22 (1996), pp. 81–101. Special issue celebrating the centenary of Runge–Kutta methods.

[2] K. Burrage and P. M. Burrage, *Order conditions of stochastic Runge–Kutta methods by B-series*, SIAM J. Numer. Anal., 38 (2000), pp. 1626–1646.

[3] K. Burrage and T. H. Tian, *Predictor-corrector methods of Runge–Kutta type for stochastic differential equations*, SIAM J. Numer. Anal., 40 (2002), pp. 1516–1537.

[4] K. Burrage and T. H. Tian, *Implicit stochastic Runge–Kutta methods for stochastic differential equations*, BIT, 44 (2004), pp. 21–39.

[5] P. M. Burrage, *Runge–Kutta methods for stochastic differential equations*, Ph.D. thesis, The University of Queensland, Brisbane, 1999.

[6] J. C. Butcher, *Coefficients for the study of Runge–Kutta integration processes*, J. Austral. Math. Soc., 3 (1963), pp. 185–201.

[7] K. Debrabant and A. Rößler, *Diagonally drift-implicit Runge–Kutta methods of weak order one and two for Itô SDEs and stability analysis*, Appl. Numer. Math., 2008, in press, doi:10.1016/j.apnum.2008.03.011.

[8] K. Debrabant and A. Rößler, *Families of efficient second order Runge–Kutta methods for the weak approximation of Itô stochastic differential equations*, Appl. Numer. Math., 2008, in press, doi:10.1016/j.apnum.2008.03.012.

[9] K. Debrabant and A. Rößler, *Classification of stochastic Runge–Kutta methods for the weak approximation of stochastic differential equations*, Math. Comput. Simulation, 77 (2008), pp. 408–420.

[10] T. C. Gard, *Introduction to stochastic differential equations*, Monographs and Textbooks in Pure Appl. Math. 114, Marcel Dekker Inc., New York, 1988.
[11] D. B. Hernández and R. Spigler, *Convergence and stability of implicit Runge–Kutta methods for systems with multiplicative noise*, BIT, 33 (1993), pp. 654–669.
[12] D. J. Higham, *Mean-square and asymptotic stability of the stochastic theta method*, SIAM J. Numer. Anal., 38 (2000), pp. 753–769.
[13] K. R. Jackson, A. Kværnø, and S. P. Nørsett, *The use of Butcher series in the analysis of Newton-like iterations in Runge–Kutta formulas*, Appl. Numer. Math., 15 (1994), pp. 341–356. International Conference on Scientific Computation and Differential Equations (Auckland, 1993).
[14] K. R. Jackson, A. Kværnø, and S. P. Nørsett, *An analysis of the order of Runge–Kutta methods that use an iterative scheme to compute their internal stage values*, BIT, 36 (1996), pp. 713–765.
[15] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, Graduate Texts in Mathematics 113, 2nd ed., Springer-Verlag, New York, 1991.
[16] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, Applications of Mathematics 21, 2nd ed., Springer-Verlag, Berlin, 1999.
[17] Y. Komori, *Weak first- or second-order implicit Runge–Kutta methods for stochastic differential equations with a scalar Wiener process*, J. Comput. Appl. Math., 217 (2008), pp. 166–179.
[18] Y. Komori, *Multi-colored rooted tree analysis of the weak order conditions of a stochastic Runge–Kutta family*, Appl. Numer. Math., 57 (2007), pp. 147–165.
[19] Y. Komori, *Weak second-order stochastic Runge–Kutta methods for non-commutative stochastic differential equations*, J. Comput. Appl. Math., 206 (2007), pp. 158–173.
[20] Y. Komori, T. Mitsui, and H. Sugiura, *Rooted tree analysis of the order conditions of ROW-type scheme for stochastic differential equations*, BIT, 37 (1997), pp. 43–66.
[21] V. Mackevičius and J. Navikas, *Second order weak Runge–Kutta type methods of Itô equations*, Math. Comput. Simulation, 57 (2001), pp. 29–34.
[22] G. N. Milstein, *Numerical integration of stochastic differential equations*, Mathematics and its Applications 313, Kluwer Academic Publishers Group, Dordrecht, 1995. Translated and revised from the 1988 Russian original.
[23] G. N. Milstein, Y. M. Repin, and M. V. Tretyakov, *Numerical methods for stochastic systems preserving symplectic structure*, SIAM J. Numer. Anal., 40 (2003), pp. 1583–1604.
[24] S. P. Nørsett and A. Wolfbrandt, *Order conditions for Rosenbrock type methods*, Numer. Math., 32 (1979), pp. 1–15.
[25] W. P. Petersen, *A general implicit splitting for stabilizing numerical simulations of Itô stochastic differential equations*, SIAM J. Numer. Anal., 35 (1998), pp. 1439–1451.
[26] A. Rößler, *Stochastic Taylor expansions for the expectation of functionals of diffusion processes*, Stoch. Anal. Appl., 22 (2004), pp. 1553–1576.
[27] A. Rößler, *Rooted tree analysis for order conditions of stochastic Runge–Kutta methods for the weak approximation of stochastic differential equations*, Stoch. Anal. Appl., 24 (2006), pp. 97–134.
[28] A. Rößler, *Runge–Kutta methods for Itô stochastic differential equations with scalar noise*, BIT, 46 (2006), pp. 97–110.
[29] A. Rößler, *Second order Runge–Kutta methods for Itô stochastic differential equations*, Technical report Preprint 2479, TU Darmstadt, 2006.
[30] A. Rößler, *Second order Runge–Kutta methods for Stratonovich stochastic differential equations*, BIT, 47 (2007), pp. 657–680.
[31] T. H. Tian and K. Burrage, *Two-stage stochastic Runge–Kutta methods for stochastic differential equations*, BIT, 42 (2002), pp. 625–643.
[32] A. Tocino and J. Vigo-Aguiar, *Weak second order conditions for stochastic Runge–Kutta methods*, SIAM J. Sci. Comput., 24 (2002), pp. 507–523.