BIT

B-series analysis of iterated Taylor methods

Kristian Debrabant · Anne Kværnø

Abstract For stochastic implicit Taylor methods that use an iterative scheme to compute their numerical solution, stochastic B-series and corresponding growth functions are constructed. From these, convergence results based on the order of the underlying Taylor method, the choice of the iteration method, the predictor, and the number of iterations, for Itô and Stratonovich SDEs, and for weak as well as strong convergence are derived. As special case, also the application of Taylor methods to ODEs is considered. The theory is supported by numerical experiments.

Keywords Stochastic Taylor method · Stochastic differential equation · Iterative scheme · Order · Newton's method · Weak approximation · Strong approximation · Growth function · Stochastic B-series

Mathematics Subject Classification (2000) 65C30 · 60H35 · 65C20 · 68U20

1 Introduction

Besides stochastic Runge–Kutta methods, one important class of schemes to approximate the solution of stochastic differential equations (SDEs) are stochastic Taylor methods. As in the deterministic setting [1], they are especially suitable for problems

Communicated by Anders Szepessy.

K. Debrabant
Fachbereich Mathematik, Technische Universität Darmstadt, Dolivostraße 15, 64293 Darmstadt,
Germany
e-mail: debrabant@mathematik.tu-darmstadt.de

A. Kværnø (✉)
Department of Mathematical Sciences, Norwegian University of Science and Technology,
7491 Trondheim, Norway
e-mail: anne.kvarno@math.ntnu.no

⧍ Springer

with not too high dimension, and here especially in the case of strong approximation, because weak approximation of low-dimensional problems can often be done more efficiently by numerically solving the corresponding deterministic PDE problem obtained by applying the Feynman-Kac formula. For solving stiff SDEs, implicit methods have to be considered, as is illustrated in the following two examples.

Example 1.1 (See [12]) Consider the linear Itô-SDE

$$dX(t) = \mu X(t)\,dt + \sigma X(t)\,dW(t), \qquad X(0) = x_0, \qquad (1.1)$$

with $\mu, \sigma \in \mathbb{C}$. We assume that the exact solution is mean-square stable, i.e.

$$\lim_{t\to\infty} \mathrm{E}(|X(t)|^2) = 0,$$

which is the case if and only if $2\Re\mu + |\sigma|^2 < 0$. To achieve that also the numerical approximation Y_n obtained with the (explicit) Euler-Maruyama scheme with step size h is mean-square stable, i.e. $\lim_{n\to\infty} \mathrm{E}(|Y_n|^2) = 0$, we have to restrict the step size according to $h < h_0 := -\frac{2\Re\mu + |\sigma|^2}{|\mu|^2}$, whereas for $h > h_0$ the numerical approximations explode, $\lim_{n\to\infty} \mathrm{E}(|Y_n|^2) = \infty$. In contrast to this, the semi-implicit Euler scheme is mean-square stable without any step size restriction. For a numerical affirmation, see Fig. 1.

Example 1.2 Consider the following stochastic Van der Pol equation,

$$dX_1(t) = X_2(t)\,dt,$$
$$dX_2(t) = \big(\mu\big(1 - X_1(t)^2\big)X_2(t) - X_1(t)\big)\,dt + \theta\big(1 - X_1(t)^2\big)X_2(t)\,dW(t),$$
$$X_1(0) = x_{0,1}, \qquad X_2(0) = x_{0,2}.$$

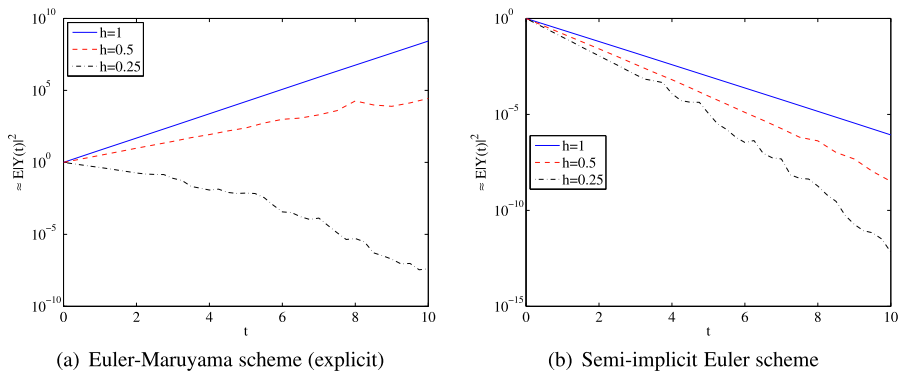(a) Euler-Maruyama scheme (explicit)

(b) Semi-implicit Euler scheme

Fig. 1 Approximation results for the linear test equation (1.1) with $\mu = -3$, $\sigma = \sqrt{3}$, and $x_0 = 1$ by Euler-Maruyama (explicit) and semi-implicit Euler scheme. Here, $\mathrm{E}(|Y(t)|^2)$ is approximated as mean over 10^6 simulations. The explicit scheme is only stable for appropriate step sizes, the semi-implicit scheme is stable for all step sizes

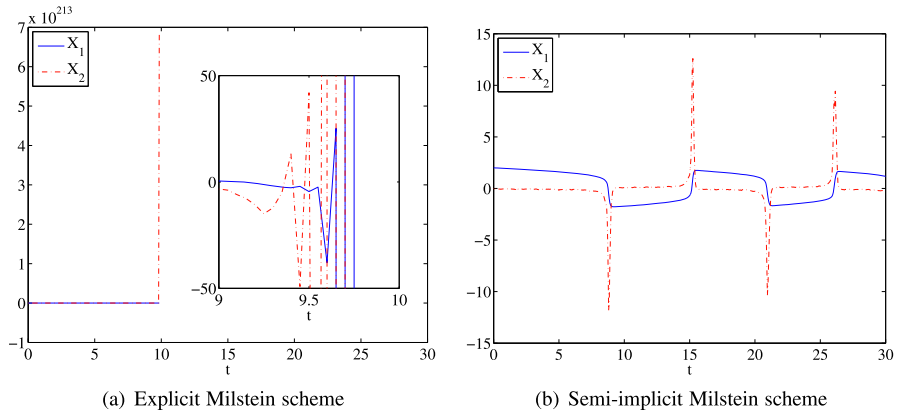(a) Explicit Milstein scheme   (b) Semi-implicit Milstein scheme

Fig. 2 Approximation of Van der Pol equation with $\mu = 10$, $\theta = 1$, $x_{0,1} = 2$, and $x_{0,2} = 0$ by the explicit and semi-implicit Milstein scheme (using the same Brownian path) with step-size $h = 0.05$. The explicit scheme suffers from heavy stability problems and aborts

Application of the explicit Milstein scheme (see Example 1.3 with $\alpha = 0$ and $\beta = 0$) with step-size $h = 0.05$ to approximate a solution path leads to an explosion of the approximation, see Fig. 2(a), whereas application of the semi-implicit Milstein scheme, given by substituting $g_0(Y_n)$ by $g_0(Y_{n+1})$ ($\alpha = 1$ and $\beta = 0$ in Example 1.3), yields (for the same Brownian path) the result of Fig. 2(b).

Implicit stochastic Taylor methods have been considered both for strong [15, 17] and weak [15] approximation. For these methods, the approximation values are only given implicitly. However, in practice these implicit equations are solved by iterative schemes like simple iteration or Newton iteration. The "exact numerical" solution can be written in terms of B-series [8]. As we will prove in this paper, so can the iterated solution. Moreover, for each iteration scheme in question, we will define a growth function. Briefly explained, when the exact numerical and the k times iterated solutions are both written in terms of B-series, then all terms of these series for which the growth function has a value not greater than k coincide. Thus the growth functions give a quite exact description of the development of the iterations. B-series and corresponding growth functions for iterated solutions have been derived for Runge–Kutta methods applied to deterministic ordinary differential equations [14], differential algebraic equations [13], and SDEs [7]. Somewhat surprisingly, the growth functions are exactly the same in all these cases, and, as we will show in this paper, this also holds for implicit Taylor methods.

The outline of the paper is as follows: First, we will give the SDE to be solved and the iterated Taylor methods used for its approximation. In Sect. 2 stochastic B-series are introduced and some useful preliminary results are presented. The main results of the paper can be found in Sect. 3, where the B-series of the iterated solutions are developed and the before mentioned growth functions derived. In Sect. 4, these findings are interpreted in terms of the order of the overall scheme, giving concrete results on the order of the considered methods depending on the kind and number of iterations, both for SDEs and ODEs. Contrary to the results obtained for Runge–Kutta

methods [7], the order of the iteration error is shown to be independent on whether Itô or Stratonovich SDEs, weak or strong convergence is considered. Finally, in Sect. 5 we present several numerical examples to support our theoretical findings.

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 a d-dimensional SDE defined by

$$dX(t) = g_0(X(t)) \, dt + \sum_{l=1}^{m} g_l(X(t)) \star dW_l(t), \qquad X(t_0) = x_0, \qquad (1.2)$$

with an m-dimensional Wiener process $(W(t))_{t \geq 0}$ and $I = [t_0, T]$. As usual, (1.2) is construed as abbreviation of

$$X(t) = x_0 + \int_{t_0}^{t} g_0(X(s)) \, ds + \sum_{l=1}^{m} \int_{t_0}^{t} g_l(X(s)) \star dW_l(s). \qquad (1.3)$$

The integral w.r.t. the Wiener process has to be interpreted e.g. 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 chosen such that SDE (1.3) has a unique solution.

To simplify the presentation, we define $W_0(s) = s$, so that (1.3) can be written as

$$X(t) = x_0 + \sum_{l=0}^{m} \int_{t_0}^{t} g_l(X(s)) \star dW_l(s). \qquad (1.4)$$

Let a discretization $I^h = \{t_0, t_1, \ldots, t_N\}$ with $t_0 < t_1 < \cdots < t_N = T$ 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 stochastic Taylor methods given by $Y_0 = x_0$ and

$$Y_{n+1} = B(\Phi_{ex}, Y_n; h_n) + B(\Phi_{im}, Y_{n+1}; h_n) \qquad (1.5)$$

for $n = 0, 1, \ldots, N - 1$ with $Y_n = Y(t_n)$, $t_n \in I^h$, $\Phi_{ex}(\emptyset) \equiv 1$, $\Phi_{im}(\emptyset) \equiv 0$.

Example 1.3 Consider the family of Milstein schemes applied to an Itô SDE with one-dimensional noise,

$$\begin{aligned} Y_{n+1} = Y_n &+ h_n\big((1 - \alpha)g_0(Y_n) + \alpha g_0(Y_{n+1})\big) \\ &+ I_{(1),h_n}\big((1 - \beta)g_1(Y_n) + \beta g_1(Y_{n+1})\big) \\ &+ (I_{(1,1),h_n} - \beta I_{(1),h_n}^2)[g_1' g_1](Y_n). \end{aligned} \qquad (1.6)$$

Here,

$$I_{(1),h_n} = W(t_{n+1}) - W(t_n) = \Delta W_n,$$

$$I_{(1,1),h_n} = \int_{t_n}^{t_{n+1}} W(s) \, dW(s) = \frac{1}{2}(\Delta W_n^2 - h_n),$$

and the parameters $\alpha, \beta \in [0, 1]$ indicate the degree of implicitness. When $\alpha = \beta = 0$ we have the explicit Milstein scheme, with $\alpha \neq 0$, $\beta = 0$ a semi-implicit scheme. In all cases, the method (1.6) can be written in the form (1.5) with

$$B(\phi_{ex}, Y_n; h_n) = x_0 + h_n(1 - \alpha)g_0(Y_n) + I_{(1),h_n}(1 - \beta)g_1(Y_n)$$
$$+ \left(I_{(1,1),h_n} - \beta I_{(1),h_n}^2\right)(g_1'g_1)(Y_n),$$
$$B(\phi_{im}, Y_{n+1}; h_n) = h_n\alpha g_0(Y_{n+1}) + I_{(1),h_n}\beta g_1(Y_{n+1}).$$

The terms Φ_{ex} and Φ_{im} refer to the time- and method-dependent part of each term: In this case

$$\Phi_{ex}(\bullet_0) = h_n(1 - \alpha), \qquad \Phi_{im}(\bullet_0) = h_n\alpha,$$
$$\Phi_{ex}(\bullet_1) = I_{(1),h_n}(1 - \beta), \qquad \Phi_{im}(\bullet_1) = I_{(1),h_n}\beta,$$
$$\Phi_{ex}([\bullet_1]_1]) = I_{(1,1),h_n} - \beta I_{(1),h_n}^2.$$

The notation will be explained in detail in Sect. 2.

What the general method (1.5) concerns, application of an iterative Newton-type method yields

$$Y_{n+1,k+1} = B(\Phi_{ex}, Y_n; h_n) + B(\Phi_{im}, Y_{n+1,k}; h_n) + J_k(Y_{n+1,k+1} - Y_{n+1,k}) \quad (1.7)$$

with some approximation J_k to the Jacobian of $B(\Phi_{im}, Y_{n+1,k}; h_n)$ and a predictor $Y_{n+1,0}$. In the following we assume that (1.7) can be solved uniquely at least for sufficiently small h_n. To simplify the presentation, we assume further that all step sizes are constant, $h_n = h$.

For the approximation J_k there exist several common choices. If we choose J_k to be the exact Jacobian $\partial_2 B(\Phi_{im}, Y_{n+1,k}; h)$, then we obtain the classical Newton iteration method for solving (1.5), with quadratic convergence. It will be denoted in the following as full Newton iteration. If we choose instead $J_k = \partial_2 B(\Phi_{im}, Y_n; h)$, then we obtain the so called modified Newton iteration method, which is only linearly convergent. Here, J_k is independent of the iteration number k, thus its computation is much cheaper and simpler than in the full Newton iteration case. The third and simplest possibility is to choose J_k equal to zero. In this case we don't even have to solve a linear system for $Y_{n+1,k+1}$. This iteration method is called simple iteration method or predictor corrector method. Its disadvantage is that for stiff systems it requires very small step sizes to converge.

For most stiff problems and problems with additive noise, the use of semi-implicit methods suffices. As will be demonstrated in Sect. 4 these have the advantage that less iterations are required to obtain the correct order of the underlying method.

2 Stochastic B-series

B-series, symbolized by $B(\phi, x_0; h)$, for SDEs were first constructed by Burrage and Burrage [2, 3] to study strong convergence in the Stratonovich case. In the following

years, this approach has been further developed by several authors to study weak and
strong convergence, for the Itô and the Stratonovich case, see e.g. [7] for an overview.
A uniform theory for the construction of stochastic B-series has been presented in [7],
in [8] this approach has been used to construct order conditions for implicit Taylor
methods. Following the exposition of these two papers, we define in this section sto-
chastic B-series and present some preliminary results that will be used later.

2.1 Some useful definitions and preliminary results

Definition 2.1 (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 2.2 (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

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

   With these definitions in place, we can define the stochastic B-series:

Definition 2.3 (B-series) A (stochastic) B-series is 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 $\phi : T \to \Xi := \{\{\varphi(h)\}_{h \geq 0} : \varphi(h) : \Omega \to \mathbb{R}$ is Borel-measurable $\forall h \geq 0\}$ as-
signs to each tree a random variable, and $\alpha : T \to \mathbb{Q}$ 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 $\tau_1, \tau_2, \ldots, \tau_\kappa$.

Note that $\alpha(\tau)$ is the inverse of the order of the automorphism group of τ.

To simplify the presentation, in the following we assume that all elementary differentials exist and all considered B-series converge. Otherwise, one has to consider truncated B-series and discuss the remainder term.

The next lemma is proved in [7].

Lemma 2.1 *If $Y(h) = B(\phi, x_0; h)$ with $\phi(\emptyset) \equiv 1$ 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

(a) *U_f is a set of trees derived from T as follows: $[\emptyset]_f \in U_f$, and if $\tau_1, \tau_2, \ldots, \tau_\kappa \in T \setminus \{\emptyset\}$ then $[\tau_1, \tau_2, \ldots, \tau_\kappa]_f \in U_f$,*

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

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

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

For notational convenience, in the following the h-dependency of the weight functions and the x_0-dependency of the elementary differentials will be suppressed whenever this is unambiguous, so $\Phi(\tau)(h)$ will be written as $\Phi(\tau)$ and $F(\tau)(x_0)$ as $F(\tau)$.

The next step is to present the composition rule for B-series. In the deterministic case, this is e.g. given by [10], using ordered trees. The same rule applies for multicolored trees, as in the stochastic case. But it is also possible to present the result without relying on ordered trees, as done in [8], and this is the approach that will be used in the following.

Consider triples $(\tau, \vartheta, \omega)$ consisting of some $\tau \in T$, a subtree ϑ sharing the root with τ, and a remainder multiset ω of trees left over when ϑ is removed from τ. We also include the empty tree as a possible subtree, in which case the triple becomes $(\tau, \emptyset, \{\tau\})$.

Example 2.1 Two examples of such triples are

and
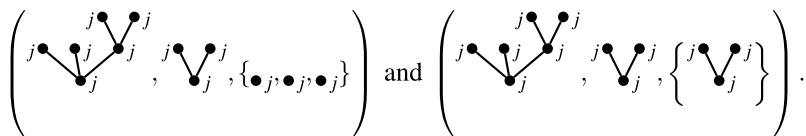
So, for the same τ and ϑ there might be different ω's.

We next define $ST(\tau)$ as the set of all possible subtrees of τ together with the corresponding remainder multiset ω, that is, for each $\tau \in T \setminus \emptyset$ we have

$$ST(\bullet_l) = \big\{ (\emptyset, \{\bullet_l\}), (\bullet_l, \{\emptyset\}) \big\},$$

$$ST(\tau = [\tau_1, \ldots, \tau_\kappa]_l) = \big\{ (\vartheta, \omega) : \vartheta = [\vartheta_1, \ldots, \vartheta_\kappa]_l, \ \omega = \{\omega_1, \ldots, \omega_\kappa\},$$

$$(\vartheta_i, \omega_i) \in ST(\tau_i), \ i = 1, \ldots, \kappa \big\} \cup \big\{ (\emptyset, \{\tau\}) \big\}.$$

We also have to take care of possible equal terms in the formula presented below. This is done as follows: For a given triple $(\tau, \vartheta, \omega)$ write first $\vartheta = [\vartheta_1, \ldots, \vartheta_{\kappa_\vartheta}]_l = [\bar\vartheta_1^{r_1}, \ldots, \bar\vartheta_q^{r_q}]_l$, where the latter only expresses that ϑ is composed by q different nonempty trees, each appearing r_i times, hence $\sum_{i=1}^q r_i = \kappa_\vartheta$. Let $\tau = [\tau_1, \ldots, \tau_\kappa]_l$. For $i = 1, \ldots, q$, each $\bar\vartheta_i$ is a subtree of some of the τ_j's, with corresponding remainder multisets ω_j. Assume that there are exactly p_i different such triples $(\bar\tau_{ik}, \bar\vartheta_i, \bar\omega_{ik})$ each appearing exactly r_{ik} times so that $\sum_{k=1}^{p_i} r_{ik} = r_i$. Finally, let $\bar\delta_k \in \omega$ be the distinct trees with multiplicity s_k, $k = 1, \ldots, p_0$, of the remainder multiset which are directly connected to the root of τ. Then, τ can be written as

$$\tau = [\bar\delta_1^{s_1}, \ldots, \bar\delta_{p_0}^{s_{p_0}}, \bar\tau_{11}^{r_{11}}, \ldots, \bar\tau_{1p_1}^{r_{1p_1}}, \ldots, \bar\tau_{q1}^{r_{q1}}, \ldots, \bar\tau_{qp_q}^{r_{qp_q}}]_l = [\bar\tau_1^{R_1}, \ldots, \bar\tau_Q^{R_Q}]_l, \quad (2.2)$$

where the rightmost expression above indicates that τ is composed by Q different trees each appearing R_i times.

With these definitions, we can state the following theorem, proved in [8]:

Theorem 2.1 (Composition of B-series) *Let* $\phi_x, \phi_y : T \to \Xi$ *and* $\phi_x(\emptyset) \equiv 1$. *Then the B-series* $B(\phi_x, x_0; h)$ *inserted into* $B(\phi_y, \cdot; h)$ *is again a B-series,*

$$B(\phi_y, B(\phi_x, x_0; h); h) = B(\phi_x \circ \phi_y, x_0; h),$$

where

$$\big(\phi_x \circ \phi_y\big)(\tau) = \sum_{(\vartheta, \omega) \in ST(\tau)} \gamma(\tau, \vartheta, \omega) \left(\phi_y(\vartheta) \prod_{\delta \in \omega} \phi_x(\delta) \right)(\tau),$$

$\gamma(\emptyset, \emptyset, \{\emptyset\}) = 1$, *and*

$$\gamma(\tau, \vartheta, \omega) = \frac{R_1! \cdots R_Q!}{s_1! \cdots s_{p_0}! r_{11}! \cdots r_{qp_q}!} \prod_{i=1}^q \prod_{k=1}^{p_i} \gamma(\bar\tau_{ik}, \bar\vartheta_i, \bar\omega_{ik})^{r_{ik}}$$

for τ *given by* (2.2).

The combinatorial term γ gives the number of equal terms that will appear if the composition rule using ordered trees is preferred.

In general, the composition law is not linear, neither is it associative. It is, however, linear in its second operand. Further, if both $\phi_x(\emptyset) = \phi_y(\emptyset) \equiv 1$, then the composition law can be turned into a group operation (Butcher group, see [6, 10, 11] for the deterministic case): The inverse element $\phi^{-1}(\tau)$ can be recursively computed by

$$(\phi \circ \phi^{-1})(\tau) = e(\tau) \equiv \begin{cases} 1 & \text{if } \tau = \emptyset, \\ 0 & \text{otherwise,} \end{cases} \quad (2.3)$$

and associativity is proved by

$$B\big(\phi_z, B\big(\phi_y, B(\phi_x, x_0; h); h\big); h\big) = B\big(\phi_x \circ (\phi_y \circ \phi_z), x_0; h\big)$$
$$= B\big((\phi_x \circ \phi_y) \circ \phi_z, x_0; h\big). \qquad (2.4)$$

This holds even if $\phi_z(\emptyset) \not\equiv 1$.

The next result will be needed for the investigation of modified Newton iterations.

Lemma 2.2 *If $\phi_x(\emptyset) \equiv 0$ we have*

$$\partial_2 B(\phi_y, x_0; h) B(\phi_x, x_0; h) = B(\phi_x * \phi_y, x_0; h),$$

where the bi-linear operator $*$ is given by

$$(\phi_x * \phi_y)(\tau) = \begin{cases} 0 & \text{if } \tau = \emptyset, \\ \sum_{(\vartheta, \{\delta\}) \in SP(\tau)} \gamma(\tau, \vartheta, \{\delta\}) \cdot \phi_y(\vartheta)\phi_x(\delta) & \text{otherwise}, \end{cases} \qquad (2.5)$$

with

$$SP(\tau) = \{(\vartheta, \omega) \in ST(\tau) : \omega \text{ contains exactly one element } \delta\}.$$

Proof Written in full, the statement of the theorem claims that

$$\sum_{\vartheta \in T} \sum_{\delta \in T} \alpha(\vartheta)\alpha(\delta) \cdot \phi_y(\vartheta)\phi_x(\delta) \cdot \partial F(\vartheta) F(\delta)$$

$$= \sum_{\tau \in T \setminus \{\emptyset\}} \alpha(\tau) \left(\sum_{(\vartheta, \{\delta\}) \in SP(\tau)} \gamma(\tau, \vartheta, \{\delta\}) \cdot \phi_y(\vartheta)\phi_x(\delta) \right) \cdot F(\tau). \qquad (2.6)$$

This is true if

$$\partial F(\vartheta) F(\delta) = \sum_{\tau \in \mathcal{A}(\vartheta, \delta)} \beta(\tau, \vartheta, \delta) F(\tau) \qquad (2.7)$$

and

$$\alpha(\vartheta)\alpha(\delta)\beta(\tau, \vartheta, \delta) = \alpha(\tau)\gamma(\tau, \vartheta, \{\delta\}), \qquad (2.8)$$

where $\mathcal{A}(\vartheta, \delta)$ is the set of all τ's constructed by attaching δ to one of the vertices of ϑ. We will prove this by induction.

First, let $\vartheta = \emptyset$. Since $\partial F(\emptyset) F(\delta) = F(\delta)$ we have $\tau = \delta$ and (2.7) and (2.8) are trivially true with $\beta(\tau, \emptyset, \tau) = 1$. Now, let $\vartheta = \bullet_l$. Then $\partial F(\vartheta) F(\delta) = g'_l F(\delta) = F([\delta]_l)$. As $\mathcal{A}(\vartheta, \delta) = \{[\delta]_l\}$ this gives $\beta([\delta]_l, \bullet_l, \delta) = 1$, and again (2.8) is trivially true. Finally, let

$$\vartheta = [\bar{\vartheta}_1^{r_1}, \ldots, \bar{\vartheta}_i^{r_i}, \ldots, \bar{\vartheta}_q^{r_q}]_l$$

with distinct trees $\bar{\vartheta}_1, \ldots, \bar{\vartheta}_q$. Then

$$\partial F(\vartheta) F(\delta)$$

$$= g_l^{(\kappa_\vartheta + 1)} \big(F(\delta), \overbrace{F(\bar{\vartheta}_1), \ldots, F(\bar{\vartheta}_1)}^{r_1 \text{ times}}, \ldots, \overbrace{F(\bar{\vartheta}_q), \ldots, F(\bar{\vartheta}_q)}^{r_q \text{ times}} \big)$$

$$+ \sum_{i=1}^q r_i g_l^{(\kappa_\vartheta)} \big(F(\bar{\vartheta}_1), \ldots, \partial F(\bar{\vartheta}_i) F(\delta), \overbrace{F(\bar{\vartheta}_i), \ldots, F(\bar{\vartheta}_i)}^{r_i - 1 \text{ times}}, \ldots, F(\bar{\vartheta}_q) \big),$$

where $\kappa_\vartheta = \sum_{i=1}^q r_i$, so $\tau \in \mathcal{A}(\vartheta, \delta)$ is either

$$\tau = [\delta, \bar{\vartheta}_1^{r_1}, \ldots, \bar{\vartheta}_i^{r_i}, \ldots, \bar{\vartheta}_q^{r_q}]_l \quad \text{with } \beta(\tau, \vartheta, \delta) = 1$$

or

$$\tau = [\bar{\vartheta}_1^{r_1}, \ldots, \tau_i, \bar{\vartheta}_i^{r_i - 1}, \ldots, \bar{\vartheta}_q^{r_q}]_l \quad \text{with } \tau_i \in \mathcal{A}(\bar{\vartheta}_i, \delta) \quad \text{and}$$

$$\beta(\tau, \vartheta, \delta) = r_i \beta(\tau_i, \bar{\vartheta}_i, \delta).$$

In the first case, if $\delta = \bar{\vartheta}_j$ for some j, then $\alpha(\tau) = \alpha(\vartheta)\alpha(\delta)/M$ and $\gamma(\tau, \vartheta, \{\delta\}) = M$ with $M = r_j + 1$. Otherwise the same is valid with $M = 1$. So (2.8) holds. In the second case, assume that our induction hypothesis (2.8) is true for all $\tau_i \in \mathcal{A}(\bar{\vartheta}_i, \delta)$. We obtain

$$\alpha(\tau) = \frac{r_i}{M} \frac{\alpha(\tau_i)}{\alpha(\bar{\vartheta}_i)} \alpha(\vartheta) \quad \text{and} \quad \gamma(\tau, \vartheta, \{\delta\}) = M \gamma(\tau_i, \bar{\vartheta}_i, \{\delta\})$$

with $M = r_j + 1$ if $\tau_i = \bar{\vartheta}_j$ for some j and $M = 1$ otherwise. It follows that

$$\alpha(\tau)\gamma(\tau, \vartheta, \{\delta\}) = r_i \frac{\alpha(\tau_i)}{\alpha(\bar{\vartheta}_i)} \gamma(\tau_i, \bar{\vartheta}_i, \{\delta\})\alpha(\vartheta) = \alpha(\vartheta)\alpha(\delta) r_i \beta(\tau_i, \bar{\vartheta}_i, \delta)$$

$$= \alpha(\vartheta)\alpha(\delta)\beta(\tau, \vartheta, \delta). \qquad \square$$

2.2 B-series of the exact and the numerical solutions

From the results of the previous subsection, it is possible to find the B-series of the exact and numerical solutions. Here, the proofs are only sketched, for details consult [7, 8].

Theorem 2.2 *The solution $X(t_0 + h)$ of (1.4) can be written as a B-series $B(\varphi, x_0; h)$ with*

$$\varphi(\emptyset) \equiv 1, \qquad \varphi(\bullet_l)(h) = W_l(h),$$

$$\varphi(\tau = [\tau_1, \ldots, \tau_\kappa]_l)(h) = \int_0^h \prod_{j=1}^\kappa \varphi(\tau_j)(s) \star dW_l(s).$$

Proof Write the exact solution as some B-series $X(t_0 + h) = B(\varphi, x_0; h)$. As $\varphi(\emptyset) \equiv 1$, apply Lemma 2.1 to $g_l(X(t_0 + h))$ and Definitions 2.1, 2.2 and 2.3 to obtain

$$g_l(B(\varphi, x_0; h)) = \sum_{\tau \in T_l} \alpha(\tau) \cdot \varphi_l'(\tau)(h) \cdot F(\tau)(x_0) \tag{2.9}$$

in which

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

Insert this into the SDE (1.3) and compare term by term. $\qquad\square$

Theorem 2.3 *The numerical solution Y_1 given by* (1.5) *can be written as a B-series*

$$Y_1 = B(\Phi, x_0; h)$$

with Φ recursively defined by

$$\Phi(\emptyset) \equiv 1, \tag{2.10a}$$

$$\Phi(\tau) = \Phi_{ex}(\tau) + (\Phi \circ \Phi_{im})(\tau). \tag{2.10b}$$

Proof Write $Y_1 = B(\Phi, x_0; h)$ and insert this into (1.5). As $\Phi(\emptyset) = \Phi_{ex}(\emptyset) + \Phi_{im}(\emptyset) \equiv 1$, apply Theorem 2.1, and compare term by term. $\qquad\square$

To study the consistency of the numerical methods, we need to assign to each tree an order:

Definition 2.4 (Tree order) The order of a tree $\tau \in T$ respectively $u \in U_f$ is defined by

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

and

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

In Table 1 some trees and the corresponding values for the functions ρ, α, and φ are presented.

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 Theorems 2.2 and 2.3 and

Table 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 $J_{(\ldots)}$ for the Stratonovich (S) and $I_{(\ldots)}$ for the Itô (I) cases, see [15] 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	$W_l(h) = \begin{cases} h & \text{if } l = 0 \\ J_{(l)} & \text{(S)} \\ I_{(l)} & \text{(I)} \end{cases}$
(tree $[[\bullet_2]_0]_1$)	2	1	$\int_0^h \int_0^{s_1} W_2(s_2) \star ds_2 \star dW_1(s_1) = \begin{cases} J_{(2,0,1)} & \text{(S)} \\ I_{(2,0,1)} & \text{(I)} \end{cases}$
(tree $[\bullet_1, \bullet_1]_0$)	2	$\frac{1}{2}$	$\int_0^h W_1(s)^2 \star ds = \begin{cases} 2J_{(1,1,0)} & \text{(S)} \\ 2I_{(1,1,0)} + I_{(0,0)} & \text{(I)} \end{cases}$
(tree $[\bullet_1, [\bullet_2, \bullet_2]_1]_0$)	3	$\frac{1}{2}$	$\int_0^h W_1(s_1)(\int_0^{s_1} W_2(s_2)^2 \star dW_1(s_2)) \star ds_1$ $= \begin{cases} 4J_{(2,2,1,1,0)} + 2J_{(2,1,2,1,0)} + 2J_{(1,2,2,1,0)} & \text{(S)} \\ 4I_{(2,2,1,1,0)} + 2I_{(2,1,2,1,0)} + 2I_{(1,2,2,1,0)} \\ + 2I_{(0,1,1,0)} + 2I_{(2,2,0,0)} + I_{(1,0,1,0)} + I_{(0,0,0,0)} & \text{(I)} \end{cases}$

Lemma 2.1 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

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

and

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

One can show [5, 9, 15] 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 case the $\mathcal{O}(\cdot)$-notation refers to the $L^2(\Omega)$-norm and $h \to 0$.

In the following we assume that also method (1.5) is consistent with the definition of the tree order, i.e. that it is constructed as usual such that $E\psi_\Phi(u)(h) = \mathcal{O}(h^{\rho(u)})$

$\forall u \in U_f$ and $\Phi(\tau)(h) = \mathcal{O}(h^{\rho(\tau)})$ $\forall \tau \in T$, respectively. These conditions are fulfilled if for $\tau \in T$ and $k \in \mathbb{N} = \{0, 1, \dots\}$ it holds that $(\Phi_{ex}(\tau))^{2^k} = \mathcal{O}(h^{2^k \rho(\tau)})$ and $(\Phi_{im}(\tau))^{2^k} = \mathcal{O}(h^{2^k \rho(\tau)})$.

3 B-series of the iterated solution and growth functions

In this section we will discuss how the iterated solution defined in (1.7) can be written in terms of B-series, that is

$$Y_{1,k} = B(\Phi_k, x_0; h).$$

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

$$Y_{1,0} = B(\Phi_0, x_0; h),$$

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

We are now ready to study each of the iteration schemes, which differ only in the choice of J_k in (1.7). In each case, we will first find the recurrence formula for $\Phi_k(\tau)$. From this we define a growth function $\mathfrak{g}(\tau)$:

Definition 3.1 (Growth function) A growth function $\mathfrak{g} : T \to \mathbb{N}$ is a function satisfying

$$\Phi_k(\tau) = \Phi(\tau) \quad \forall \tau \in T \text{ with } \mathfrak{g}(\tau) \leq k$$

$$\Rightarrow \quad \Phi_{k+1}(\tau) = \Phi(\tau) \quad \forall \tau \in T \text{ with } \mathfrak{g}(\tau) \leq k+1, \tag{3.1}$$

for all $k \geq 0$.

This result should be sharp in the sense that in general there exists $\tau \neq \emptyset$ with $\Phi_k(\tau) \neq \Phi(\tau)$ when $k < \mathfrak{g}(\tau)$. From Lemma 2.1 we also have

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

with

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

where $\beta(u)$ and $G(u)(x_0)$ are given in Lemma 2.1. This implies

$$\psi_{\Phi_k}(u) = \psi_{\Phi(\tau)} \quad \forall u = [\tau_1, \dots, \tau_\kappa]_f \in U_f \text{ with } \mathfrak{g}'(u) = \max_{j=1}^{\kappa} \mathfrak{g}(\tau_i) \leq k. \tag{3.2}$$

As we will see, the growth functions give a precise description of the development of the iterations. However, to get applicable results we will at the end need the relation between the growth functions and the order. These aspects are discussed in the next section. Examples of trees and the values of the growth functions for the three iteration schemes are given in Fig. 3.
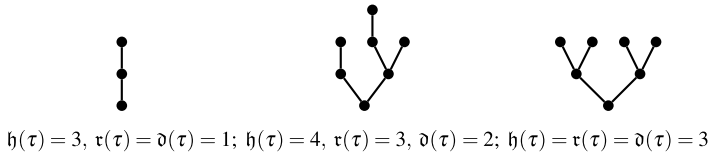
$$\mathfrak{h}(\tau) = 3,\ \mathfrak{r}(\tau) = \mathfrak{d}(\tau) = 1;\ \mathfrak{h}(\tau) = 4,\ \mathfrak{r}(\tau) = 3,\ \mathfrak{d}(\tau) = 2;\ \mathfrak{h}(\tau) = \mathfrak{r}(\tau) = \mathfrak{d}(\tau) = 3$$

Fig. 3 Examples of trees and their growth functions for simple ($\mathfrak{h}$), modified Newton ($\mathfrak{r}$) and full Newton ($\mathfrak{d}$) iterations

3.1 The simple iteration

Simple iterations are described by (1.7) with $J_k = 0$, that is

$$Y_{n+1,k+1} = B(\Phi_{ex}, Y_n; h) + B(\Phi_{im}, Y_{n+1,k}; h). \tag{3.3}$$

By Theorem 2.1 we easily get the following lemma, where, as in the following, all results are valid for all $l = 0, \ldots, m$:

Lemma 3.1 *If $Y_{1,0} = B(\Phi_0, x_0; h)$ then $Y_{1,k} = B(\Phi_k, x_0; h)$, where*

$$\Phi_{k+1}(\emptyset) \equiv 1,$$

$$\Phi_{k+1}(\tau) = \Phi_{ex}(\tau) + (\Phi_k \circ \Phi_{im})(\tau).$$

The corresponding growth function is given by

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

The function $\mathfrak{h}(\tau)$ is the height of τ, that is the maximum number of nodes along one branch.

3.2 The modified Newton iteration

In this subsection we consider the modified Newton iteration

$$Y_{n+1,k+1} = B(\Phi_{ex}, Y_n; h) + B(\Phi_{im}, Y_{n+1,k}; h)$$
$$+ \partial_2 B(\Phi_{im}, Y_n; h)(Y_{n+1,k+1} - Y_{n+1,k}). \tag{3.4}$$

The B-series for $Y_{1,k}$ and the corresponding growth function can now be described by the following lemma:

Lemma 3.2 *If $Y_{1,0} = B(\Phi_0, x_0; h)$ then $Y_{1,k} = B(\Phi_k, x_0; h)$ with*

$$\Phi_{k+1}(\emptyset) \equiv 1,$$
$$\Phi_{k+1}(\tau) = \Phi_{ex}(\tau) + (\Phi_k \circ \Phi_{im})(\tau) + ((\Phi_{k+1} - \Phi_k) * \Phi_{im})(\tau). \tag{3.5}$$

The corresponding growth function is given by

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

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

Proof Theorem 2.1 and Lemma 2.2 imply (3.5). We next prove that $\mathfrak{r}$ is the appropriate growth function. If $\mathfrak{r}(\tau) = 0$ then $\tau = \emptyset$ and $\Phi_0(\tau) = \Phi(\tau)$. Assume now that $\Phi_k(\tau) = \Phi(\tau)$ $\forall\tau$ with $\mathfrak{r}(\tau) \leq k$. If $(\vartheta, \omega) \in ST(\tau) \setminus SP(\tau)$ with $\mathfrak{r}(\tau) \leq k+1$, then $\forall\delta \in \omega$ it holds $\mathfrak{r}(\delta) \leq \mathfrak{r}(\tau)$ and therefore $\Phi_k(\delta) = \Phi(\delta)$. So, by (3.5) we have $\forall\tau$ with $\mathfrak{r}(\tau) \leq k+1$

$$\Phi_{k+1}(\tau) = \Phi_{ex}(\tau) + \sum_{(\vartheta,\omega)\in ST(\tau)\setminus SP(\tau)} \gamma(\tau, \vartheta, \omega) \cdot \Phi_{im}(\vartheta) \prod_{\delta\in\omega} \Phi(\delta)$$

$$+ \sum_{(\vartheta,\{\delta\})\in SP(\tau)} \gamma(\tau, \vartheta, \{\delta\}) \cdot \Phi_{im}(\vartheta)\Phi_{k+1}(\delta),$$

and by induction on the number of nodes of these trees we obtain that $\Phi_{k+1}(\tau) = \Phi(\tau)$ $\forall\tau$ with $\mathfrak{r}(\tau) \leq k+1$. $\qquad\square$

3.3 The full Newton iteration

In this subsection we consider the full Newton iteration (1.7) with $J_k = \partial_2 B(\Phi_{im}, Y_{1,k}; h)$. Extending the $*$-operator to the case when its first operand does not vanish on the empty tree by

$$(\phi_x * \phi_y)(\tau) = \left((\phi_x - \phi_x(\emptyset)e) * \phi_y\right)(\tau) + \phi_x(\emptyset)\phi_y(\tau),$$

it follows that the B-series for $Y_{1,k}$ and the corresponding growth function satisfy:

Lemma 3.3 *If* $Y_{1,0} = B(\Phi_0, x_0; h)$ *then* $Y_{1,k+1} = B(\Phi_{k+1}, x_0; h)$ *with*

$$\Phi_{k+1}(\emptyset) \equiv 1,$$
$$\Phi_{k+1}(\tau) = \Phi_{ex}(\tau) + \left(\Phi_k \circ \left((\Phi_k^{-1} \circ \Phi_{k+1}) * \Phi_{im}\right)\right)(\tau). \tag{3.6}$$

The corresponding growth function is given by

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

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

where γ is the number of trees τ_i in $\tau_1, \ldots, \tau_\kappa$ 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 Writing the iterations in terms of B-series, we get

$$B(\Phi_{k+1}, x_0; h) = B(\Phi_{ex}, x_0; h) + B(\Phi_{im}, Y_{1,k}; h)$$
$$+ \partial_2 B(\Phi_{im}, Y_{1,k}; h) B(\Delta\Phi_k, x_0; h) \tag{3.7}$$

with $\Delta\Phi_k(\tau) = \Phi_{k+1}(\tau) - \Phi_k(\tau)$. Let $x_0 = B(\Phi_k^{-1}, Y_{1,k}; h)$ so that

$$B(\Delta\Phi_k, x_0; h) = B\big(\Delta\Phi_k, B(\Phi_k^{-1}, Y_{1,k}; h); h\big) = B(\Phi_k^{-1} \circ \Delta\Phi_k, Y_{1,k}; h).$$

The use of Lemma 2.2 followed by the use of Theorem 2.1 give the following result:

$$\partial_2 B(\Phi_{im}, Y_{1,k}; h) B(\Phi_k^{-1} \circ \Delta\Phi_k, Y_{1,k}; h) = B\big((\Phi_k^{-1} \circ \Delta\Phi_k) * \Phi_{im}, Y_{1,k}; h\big)$$
$$= B\big(\Phi_k \circ ((\Phi_k^{-1} \circ \Delta\Phi_k) * \Phi_{im}), x_0; h\big).$$

The operator $*$ is bilinear and $\circ$ is linear from the right, thus

$$\Phi_k \circ ((\Phi_k^{-1} \circ \Delta\Phi_k) * \Phi_{im}) = \Phi_k \circ ((\Phi_k^{-1} \circ \Phi_{k+1}) * \Phi_{im}) - \Phi_k \circ \Phi_{im}$$

and the first part of the theorem is proven by (3.7). We will now prove the second part. Assume that $\Phi_k(\tau) = \Phi_{k+1}(\tau) = \Phi(\tau)$ for all τ satisfying $\mathfrak{d}(\tau) \le k$. This is true for $k = 0$ and $\tau = \emptyset$. Let $\Psi_k = \Phi_k^{-1} \circ \Phi_{k+1}$ and notice that by the assumption above, $\Psi_k(\tau)$ equals the unit element $e(\tau)$ if $\mathfrak{d}(\tau) \le k$. Consider a tree τ where $\mathfrak{d}(\tau) = k+1$. For this tree we obtain

$$(\Psi_k \circ \Phi_{im})(\tau) = \sum_{(\vartheta, \{\delta\}) \in SP(\tau)} \gamma(\tau, \vartheta, \{\delta\}) \Phi_{im}(\vartheta) \Psi_k(\delta)$$
$$+ \sum_{(\vartheta, \omega) \in ST(\tau) \setminus SP(\tau)} \gamma(\tau, \vartheta, \omega) \Phi_{im}(\vartheta) \prod_{\delta \in \omega} \Psi_k(\delta)$$
$$= (\Psi_k * \Phi_{im})(\tau) \tag{3.8}$$

since the last sum of (3.8) disappears: For each $(\vartheta, \omega) \in ST(\tau) \setminus SP(\tau)$ (if any) there is at least one $\delta \in \omega$ satisfying $\mathfrak{d}(\delta) \le k$ and thereby $\Psi_k(\delta) = 0$. In this case we obtain

$$\big(\Phi_k \circ ((\Phi_k^{-1} \circ \Phi_{k+1}) * \Phi_{im})\big)(\tau) = (\Phi_k \circ \Phi_k^{-1} \circ \Phi_{k+1} \circ \Phi_{im})(\tau) = (\Phi_{k+1} \circ \Phi_{im})(\tau)$$

by (2.4), so that

$$\Phi_{k+1}(\tau) = \Phi_{ex}(\tau) + (\Phi_{k+1} \circ \Phi_{im})(\tau).$$

The theorem is completed by induction on the number of nodes of τ and on k. $\square$

4 General convergence results for iterated methods

Now we will relate the results of the previous section to the order of the overall scheme. We have weak consistency of order p if and only if

$$\mathrm{E}\,\psi_\Phi(u)(h) = \mathrm{E}\,\psi_\varphi(u)(h) + \mathcal{O}(h^{p+1}) \quad \forall u \in U_f \text{ with } \rho(u) \leq p + \frac{1}{2} \qquad (4.1)$$

((4.1) slightly weakens conditions given in [16]), and mean square global order p if [4]

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

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

and all elementary differentials $F(\tau)$ fulfill a linear growth condition.

Then, the order of the iterated solution after k iterations is q_k if

$$\mathrm{E}\,\psi_{\Phi_k}(u) = \mathrm{E}\,\psi_\varphi(u) \quad \forall u \in U_f \text{ with } \rho(u) \leq q_k + \frac{1}{2} \qquad (4.2)$$

in the weak convergence case respectively

$$\Phi_k(\tau) = \varphi(\tau) \quad \forall \tau \in T \text{ with } \rho(\tau) \leq q_k,$$

$$\mathrm{E}\,\Phi_k(\tau) = \mathrm{E}\,\varphi(\tau) \quad \forall \tau \in T \text{ with } \rho(\tau) = q_k + \frac{1}{2} \qquad (4.3)$$

in the mean square convergence case.

In the following, we assume that the predictors satisfy the condition

$$\Phi_0(\tau) = \Phi(\tau) \quad \forall \tau \in T \text{ with } \mathfrak{g}(\tau) \leq \mathcal{G}_0, \qquad (4.4)$$

where $\mathcal{G}_0$ is chosen as large as possible. In particular, the trivial predictor satisfies $\mathcal{G}_0 = 0$.

It follows from (3.1) and (3.2) that

$$\Phi_k(\tau) = \Phi(\tau) \quad \forall \tau \in T \text{ with } \mathfrak{g}(\tau) \leq \mathcal{G}_0 + k, \qquad (4.5)$$

as well as

$$\psi_{\Phi_k}(u) = \psi_\Phi(u) \quad \forall u \in U_f \text{ with } \mathfrak{g}'(u) \leq \mathcal{G}_0 + k. \qquad (4.6)$$

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 a maximum growth function, given by

$$\mathcal{G}(q) = \max_{\tau \in T} \{\mathfrak{g}(\tau) : \rho(\tau) \leq q\} = \max_{u \in U_f} \{\mathfrak{g}'(u) : \rho(u) \leq q\}. \qquad (4.7)$$

With this definition, by (4.6) respectively (4.5), the conditions (4.2) respectively (4.3) are fulfilled for all u of order $\rho(u) \leq \min(q_k, p)$ respectively all τ of order

$\rho(\tau) \leq \min(q_k, p)$ if

$$\mathcal{G}\left(q_k + \frac{1}{2}\right) \leq \mathcal{G}_0 + k. \tag{4.8}$$

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. E.g. from [9] we have

$$\begin{aligned}
\mathrm{E}\,\varphi(\tau) &= 0 \quad \text{if } \tau \notin T^S, \\
\mathrm{E}\,\psi_\varphi(u) &= 0 \quad \text{if } u \notin U_f^S.
\end{aligned} \tag{4.9}$$

Thus, if the method is as usual constructed such that also $\forall m, n \in \mathbb{N}$ and $\forall \tau_{1,i} \in T$, $i = 1, \ldots, m$, $\forall \tau_{2,j} \in T$, $j = 1, \ldots, n$,

$$\mathrm{E}\left(\prod_{i=1}^{m}\prod_{j=1}^{n} \Phi_{ex}(\tau_{1,i})\Phi_{im}(\tau_{2,j})\right) = 0 \quad \text{if } \sum_{i=1}^{m}\rho(\tau_{1,i}) + \sum_{j=1}^{n}\rho(\tau_{2,i}) \notin \mathbb{N}, \tag{4.10}$$

then in (4.8) $q_k + \frac{1}{2}$ can be replaced by $\lfloor q_k + \frac{1}{2}\rfloor$.

The results can then be summarized in the following theorem:

Theorem 4.1 *If* (4.10) *is fulfilled, then the iterated method is of weak respectively mean square order $q_k \leq p$ after $\mathcal{G}(\lfloor q_k + \frac{1}{2}\rfloor) - \mathcal{G}_0$ iterations, otherwise after $\mathcal{G}(q_k + \frac{1}{2}) - \mathcal{G}_0$ iterations.*

Our next aim is to give explicit formulas for the maximum growth function. Let us start with the following lemma.

Lemma 4.1 *For $k \geq 1$,*

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

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

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 Fig. 4), and denote this minimal order by $\rho_{\mathfrak{h},k}$, $\rho_{\mathfrak{r},k}$, and $\rho_{\mathfrak{d},k}$. Minimal order trees are build up only by stochastic nodes. It follows immediately that $\mathcal{T}_{\mathfrak{h},1} = \mathcal{T}_{\mathfrak{r},1} = \mathcal{T}_{\mathfrak{d},1} = \{\bullet_l : l \geq 1\}$. Since $\rho(\bullet_l) = 1/2$ for $l \geq 1$, the results are proved for $k = 1$. It is easy to show by induction on k that
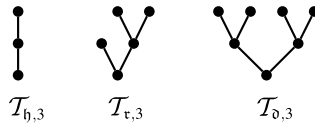
Fig. 4 Minimal order trees with $\mathfrak{g}(\tau) = 3$. The sets $\mathcal{T}_{\mathfrak{g},3}$ consist of all such trees with only stochastic nodes

$$\mathcal{T}_{\mathfrak{h},k} = \{[\tau]_l : \tau \in \mathcal{T}_{\mathfrak{h},k-1}, \; l \geq 1\}, \qquad \rho_{\mathfrak{h},k} = \rho_{\mathfrak{h},k-1} + \frac{1}{2} = \frac{k}{2},$$

$$\mathcal{T}_{\mathfrak{r},k} = \{[\bullet_{l_1}, \tau]_{l_2} : \tau \in \mathcal{T}_{\mathfrak{r},k-1}, \; l_1, l_2 \geq 1\}, \qquad \rho_{\mathfrak{r},k} = \rho_{\mathfrak{r},k-1} + 1 = k - \frac{1}{2}, \quad (4.11)$$

$$\mathcal{T}_{\mathfrak{d},k} = \{[\tau_1, \tau_2]_l : \tau_1, \tau_2 \in \mathcal{T}_{\mathfrak{d},k-1}, \; l \geq 1\}, \qquad \rho_{\mathfrak{d},k} = 2\rho_{\mathfrak{d},k-1} + \frac{1}{2} = 2^{k-1} - \frac{1}{2}.$$

For each $\mathfrak{g}$ being either $\mathfrak{h}$, $\mathfrak{r}$, or $\mathfrak{d}$, the minimal order trees satisfying $\mathfrak{g}'(u_{\mathfrak{g},k}) = k$ are $u_{\mathfrak{g},k} = [\tau_{\mathfrak{g},k}]_f$ with $\tau_{\mathfrak{g},k} \in \mathcal{T}_{\mathfrak{g},k}$, which are of order $\rho(\tau_{\mathfrak{g},k})$. $\qquad\square$

Now we can prove the following corollary.

Corollary 4.1 *For $q \geq \frac{1}{2}$ we have*

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

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

For some methods, these results can be refined. We call a method semi-implicit, if $\Phi_{im}(\tau) \equiv 0 \; \forall \tau \notin T_0$ (remember that T_0 is the set of trees with a deterministic root). Then, by Lemmas 3.1, 3.2, and 3.3 we obtain the following lemma:

Lemma 4.2 *For semi-implicit methods, the corresponding growth functions are given by*

$$\mathfrak{h}_{\mathfrak{s}}(\emptyset) = 0, \qquad \mathfrak{h}_{\mathfrak{s}}([\tau_1, \ldots, \tau_\kappa]_l) = \begin{cases} 1 & \text{if } l > 0, \\ 1 + \max_{j=1}^{\kappa} \mathfrak{h}_{\mathfrak{s}}(\tau_j) & \text{if } l = 0, \end{cases}$$

$$\mathfrak{r}_{\mathfrak{s}}(\emptyset) = 0, \qquad \mathfrak{r}_{\mathfrak{s}}(\bullet_l) = 1,$$

$$\mathfrak{r}_{\mathfrak{s}}(\tau = [\tau_1, \ldots, \tau_\kappa]_l) = \begin{cases} 1 & \text{if } l > 0, \\ \mathfrak{r}_{\mathfrak{s}}(\tau_1) & \text{if } l = 0, \kappa = 1, \\ 1 + \max_{j=1}^{\kappa} \mathfrak{r}_{\mathfrak{s}}(\tau_j) & \text{if } l = 0, \kappa \geq 2, \end{cases}$$

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

$$\mathfrak{d}_\mathfrak{s}(\tau = [\tau_1, \ldots, \tau_\kappa]_l) = \begin{cases} 1 & \text{if } l > 0, \\ \max_{j=1}^\kappa \mathfrak{d}_\mathfrak{s}(\tau_j) & \text{if } l = 0, \gamma = 1, \\ \max_{j=1}^\kappa \mathfrak{d}_\mathfrak{s}(\tau_j) + 1 & \text{if } l = 0, \gamma \geq 2, \end{cases}$$

where γ is the number of trees in $\tau_1, \ldots, \tau_\kappa$ satisfying $\mathfrak{d}_\mathfrak{s}(\tau_i) = \max_{j=1}^\kappa \mathfrak{d}_\mathfrak{s}(\tau_j)$.

This implies immediately:

Lemma 4.3 *For $k \geq 1$,*

$$\mathfrak{h}_\mathfrak{s}(\tau) = k \quad \Rightarrow \quad \rho(\tau) \geq k - \frac{1}{2},$$

$$\mathfrak{r}_\mathfrak{s}(\tau) = k \quad \Rightarrow \quad \rho(\tau) \geq \frac{3}{2}k - 1,$$

$$\mathfrak{d}_\mathfrak{s}(\tau) = k \quad \Rightarrow \quad \rho(\tau) \geq \frac{3}{4}2^k - 1.$$

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

Corollary 4.2 *For semi-implicit methods we have for $q \geq \frac{1}{2}$*

$$\mathcal{G}(q) = \begin{cases} \lfloor q + \frac{1}{2} \rfloor & \text{for simple iterations,} \\ \lfloor \frac{2}{3}(q+1) \rfloor & \text{for modified Newton iterations,} \\ \lfloor \log_2 \frac{q+1}{3} \rfloor + 2 & \text{for full Newton iterations.} \end{cases}$$

For the trivial predictor, Table 2 gives the number of iterations needed to achieve a certain order of convergence, both in the general and in the semi-implicit case.

For the sake of completeness, we also give the corresponding results for (deterministic) Taylor methods applied to deterministic problems. Note that in this case, (4.10) is automatically fulfilled.

Table 2 Number of iterations needed to achieve order p when using the simple, modified or full Newton iteration scheme in the Itô and Stratonovich case for strong or weak approximation, provided (4.10) is fulfilled. In parentheses, the numbers for semi-implicit methods are given

p	Simple iter.	Mod. iter.	Full iter.
$\frac{1}{2}$	2 (1)	1	1
1	2 (1)	1	1
$1\frac{1}{2}$	4 (2)	2	2
2	4 (2)	2	2
$2\frac{1}{2}$	6 (3)	3 (2)	2
3	6 (3)	3 (2)	2

Lemma 4.4 *Suppose that the considered problem is purely deterministic, i.e. $m = 0$ in* (1.3). *Then, for $k \geq 1$,*

$$\mathfrak{h}(\tau) = k \quad \Rightarrow \quad \rho(\tau) \geq k,$$

$$\mathfrak{r}(\tau) = k \quad \Rightarrow \quad \rho(\tau) \geq 2k - 1,$$

$$\mathfrak{d}(\tau) = k \quad \Rightarrow \quad \rho(\tau) \geq 2^k - 1.$$

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

Corollary 4.3 *For deterministic problems, we have for $q \in \mathbb{N}$, $q \geq 1$*

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

5 Numerical examples

In the following, we analyze numerically the order of convergence of several stochastic Taylor methods in dependence on the kind and number of iterations.

As first examples, we apply the semi-implicit Milstein method [15], denoted by SIM and given by (1.6), the implicit Milstein-Taylor method [17], denoted by IM and given by

$$Y_{n+1} = Y_n + h g_0(Y_{n+1}) + I_{(1)} g_1(Y_{n+1}) - (I_{(1,1)} + h)[g_1' g_1](Y_{n+1}),$$

both of strong order 1.0, and the semi-implicit strong order 1.5 Taylor method due to Kloeden and Platen [15, 17], denoted by SIKP and given by

$$Y_{n+1} = Y_n + h g_0(Y_{n+1}) + I_{(1)} g_1(Y_n) + I_{(1,1)}[g_1' g_1](Y_n) - I_{(0,1)}[g_0' g_1](Y_n)$$

$$- \frac{1}{2} h^2 \left[g_0' g_0 + \frac{1}{2} g_0'' g_1^2 \right](Y_{n+1}) + I_{(0,1)} \left[g_1' g_0 + \frac{1}{2} g_1'' g_1^2 \right](Y_n)$$

$$+ I_{(1,1,1)}[g_1'^2 g_1 + g_1'' g_1^2](Y_n),$$

to the non-linear SDE [15]

$$dX(t) = \left(\frac{1}{2} X(t) + \sqrt{X(t)^2 + 1} \right) dt + \sqrt{X(t)^2 + 1} \, dW(t), \qquad X(0) = 0, \quad (5.1)$$

on the time interval $I = [0, 1]$ with the solution $X(t) = \sinh(t + W(t))$. With each method, the solution is approximated with step sizes $2^{-11}, \ldots, 2^{-15}$ and the sample average of $M = 4000$ independent simulated realisations of the absolute error is calculated in order to estimate the expectation.

The results at time $t = 1$ are presented in Fig. 5, where the orders of convergence correspond to the slope of the regression lines. As predicted by Table 2 we observe
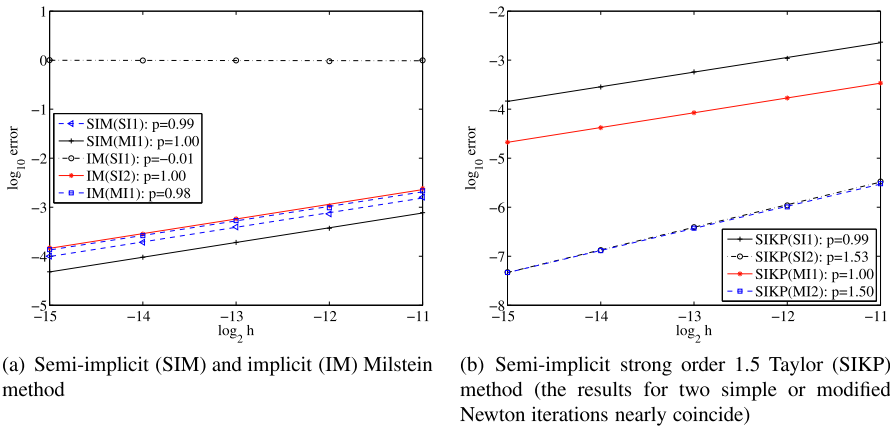
(a) Semi-implicit (SIM) and implicit (IM) Milstein method

(b) Semi-implicit strong order 1.5 Taylor (SIKP) method (the results for two simple or modified Newton iterations nearly coincide)

Fig. 5 Error of several Taylor methods applied to (5.1) with up to two simple (SI) and modified Newton (MI) iterations

strong order 1.0 for one simple or one (modified) Newton iteration of the semi-implicit Milstein method; and no convergence for one simple iteration but strong order 1.0 for two simple or one (modified) Newton iteration of the implicit Milstein-Taylor method. The semi-implicit strong order 1.5 Taylor method yields strong order 1.0 for one and strong order 1.5 for two simple or modified Newton iterations.

Next, we apply the semi-implicit weak order two Taylor scheme due to Platen [15], denoted by SIW and given by

$$Y_{n+1} = Y_n + h g_0(Y_{n+1}) + I_{(1)} g_1(Y_n) + I_{(1,1)}[g_1' g_1](Y_n)$$
$$+ \frac{1}{2} I_{(1)} h \left[-g_0' g_1 + g_1' g_0 + \frac{1}{2} g_1'' g_1^2 \right](Y_n) - \frac{1}{2} h^2 \left[g_0' g_0 + \frac{1}{2} g_0'' g_1^2 \right](Y_{n+1}),$$

to SDE (5.1). Here, we choose as functional $f(x) = p(\text{arsinh}(x))$, where $p(z) = z^3 - 6z^2 + 8z$ is a polynomial. Then the expectation of the solution can be calculated as

$$\mathrm{E}(f(X(t))) = t^3 - 3t^2 + 2t. \tag{5.2}$$

The solution $\mathrm{E}(f(X(t)))$ is approximated with step sizes $2^{-3}, \ldots, 2^{-6}$ and $M = 4 \cdot 10^9$ simulations are performed in order to determine the systematic error of SIW at time $t = 1$. The results with one or two simple or modified Newton iteration steps are presented in Fig. 6. According to Table 2 we expect approximation order one for one iteration and order two for two iterations, which is approved by Fig. 6. Finally, we apply the fully implicit strong order 1.5 Taylor scheme given in [8],

$$Y_{n+1} = Y_n + \frac{1}{2} I_{(1)} g_{1,n+1} + \frac{1}{2} h g_{0,n+1} + \frac{1}{2} (I_{(1,1)} + h) g_{1,n+1}' g_{1,n+1}$$
$$+ \frac{1}{4} h^2 g_{0,n+1}' g_{0,n+1} + \frac{1}{8} h^2 g_{0,n+1}''(g_{1,n+1}, g_{1,n+1}) + \frac{1}{2} I_{(1)} g_1$$

Fig. 6 Error of the weak second order (semi-) implicit Platen method applied to (5.1) with one or two simple (SI) or modified Newton (MI) iterations (the results for one respectively two simple and modified Newton iterations coincide)
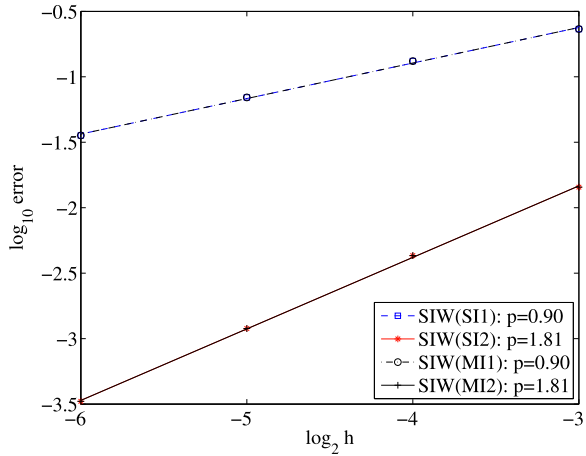
$$
+ \frac{1}{2} h g_0 - \left(h + \frac{1}{2} I_{(1,1)} \right) g_1' g_1 + \frac{1}{2} \left(I_{(0,1)} - I_{(1,0)} \right) g_1' g_0
$$

$$
- \frac{1}{2} (I_{(0,1)} - I_{(1,0)}) g_0' g_1 + \left(\frac{1}{2} I_{(0,1)} - \frac{7}{4} h I_{(1)} - 2 I_{(1,1,1)} \right) g_1''(g_1, g_1)
$$

$$
- \left(\frac{3}{2} h I_{(1)} + 2 I_{(1,1,1)} \right) g_1' g_1' g_1 - \frac{1}{4} h^2 g_0' g_0 - h^2 g_1''(g_0, g_1) - \frac{1}{4} h^2 g_1' g_0' g_1
$$

$$
- \frac{3}{4} h^2 g_1' g_1' g_0 - \frac{1}{8} h^2 g_0''(g_1, g_1) - \frac{1}{4} h^2 g_1' g_1' g_1' g_1 - \frac{5}{8} h^2 g_1' g_1''(g_1, g_1)
$$

$$
- \frac{7}{4} h^2 g_1''(g_1' g_1, g_1) - \frac{3}{4} h^2 g_1'''(g_1, g_1, g_1)
$$

(here we used the abbreviations $g_{l,n+1} = g_l(Y_{n+1})$ and $g_l = g_l(Y_n)$), which is denoted by FIT, and the semi-implicit strong order 1.5 scheme SIKP to the system of non-linear SDEs

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

$$
+ \left(\sin(X_1(t)) + 2 \sin(X_2(t)) \right) dW(t),
$$

$$
dX_2(t) = \left(\frac{1}{2} X_1(t) + \sqrt{X_2(t)^2 + 1} \right) dt + \left(\cos(X_1(t)) + 3 \cos(X_2(t)) \right) dW(t),
$$

$$
X_1(0) = 0, \qquad X_2(0) = 0,
$$

(5.3)

again on the time interval $I = [0, 1]$. The solution is approximated with step sizes $2^{-11}, \ldots, 2^{-15}$ and the sample average of $M = 4000$ independent simulated realisations of the absolute error is calculated in order to estimate the expectation. As here we do not know the exact solution, to approximate it we use SIKP with two simple iterations and a step size ten times smaller than the actual step size.
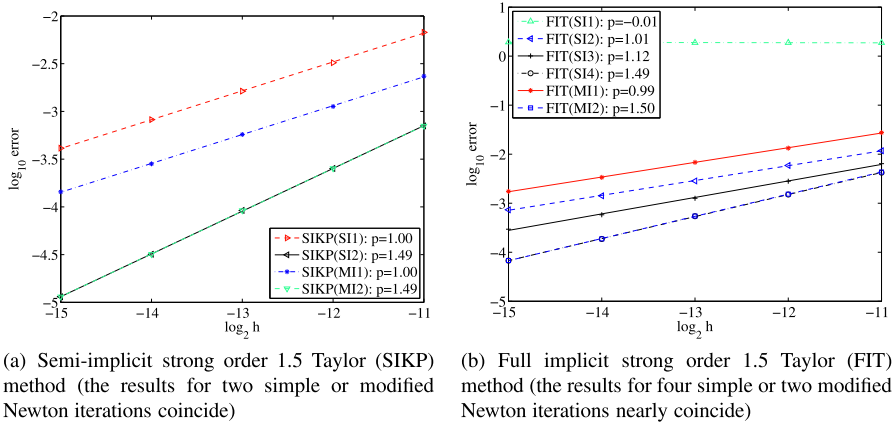
(a) Semi-implicit strong order 1.5 Taylor (SIKP) method (the results for two simple or modified Newton iterations coincide)

(b) Full implicit strong order 1.5 Taylor (FIT) method (the results for four simple or two modified Newton iterations nearly coincide)

Fig. 7 Error of SIKP and FIT applied to (5.3) with different numbers of simple (SI) and modified Newton (MI) iterations

The numerical results at $t = 1$ are presented in Fig. 7. Again, the orders expected according to Table 2 are confirmed.

6 Conclusion

For stochastic implicit Taylor methods that use an iterative scheme to approximate the solution, we derived stochastic B-series and corresponding growth functions. From these, we deduced convergence results based on the order of the underlying Taylor method, the choice of the iteration method, the predictor, and the number of iterations, for Itô and Stratonovich SDEs, and for weak as well as strong convergence. The convergence results are confirmed by numerical experiments. From a practical point of view, this theory might lead to the construction of more efficient numerical schemes for SDEs. But we also like to point out that the similarities of the iteration dependent growth functions $\mathfrak{g}$ for a range of problems (ODEs, DAEs, and SDEs) and underlying methods (Runge–Kutta methods, implicit Taylor methods) indicate an underlying structure that could well be investigated in a more general fashion. In spite of this, the number of iterations needed to obtain the order of the underlying implicit Taylor method does not depend on whether the SDE is of Itô or Stratonovich type. This is in contrast to the results obtained for Runge–Kutta methods for SDEs, for which usually less iterations are needed in the Stratonovich case [7]. The reason for this is that in the latter case certain error terms have vanishing expectation even if they do not vanish themselves. This is not the situation for implicit Taylor methods.

References

1. Barrio, R.: Performance of the Taylor series method for ODEs/DAEs. Appl. Math. Comput. **163**(2), 525–545 (2005). doi:10.1016/j.amc.2004.02.015
2. Burrage, K., Burrage, P.M.: High strong order explicit Runge–Kutta methods for stochastic ordinary differential equations. Appl. Numer. Math. **22**(1–3), 81–101 (1996). doi:10.1016/S0168-9274(96)00027-X. Special issue celebrating the centenary of Runge–Kutta methods
3. Burrage, K., Burrage, P.M.: Order conditions of stochastic Runge–Kutta methods by *B*-series. SIAM J. Numer. Anal. **38**(5), 1626–1646 (2000). doi:10.1137/S0036142999363206 (electronic)
4. Burrage, K., Tian, T.: Implicit stochastic Runge–Kutta methods for stochastic differential equations. BIT **44**(1), 21–39 (2004). doi:10.1023/B:BITN.0000025089.50729.0f
5. Burrage, P.M.: Runge–Kutta methods for stochastic differential equations. Ph.D. thesis, The University of Queensland, Brisbane (1999)
6. Butcher, J.: An algebraic theory of integration methods. Math. Comput. **26**, 79–106 (1972). doi:10.2307/2004720
7. Debrabant, K., Kværnø, A.: B-series analysis of stochastic Runge-Kutta methods that use an iterative scheme to compute their internal stage values. SIAM J. Numer. Anal. **47**(1), 181–203 (2008/2009). doi:10.1137/070704307
8. Debrabant, K., Kværnø, A.: Composition of stochastic B-series with applications to implicit Taylor methods. Appl. Numer. Math. **61**(4), 501–511 (2011). doi:10.1016/j.apnum.2010.11.014
9. Debrabant, K., Kværnø, A.: Stochastic Taylor expansions: weight functions of B-series expressed as multiple integrals. Stoch. Anal. Appl. **28**(2), 293–302 (2010). doi:10.1080/07362990903546504
10. Hairer, E., Lubich, C., Wanner, G.: Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations, 2nd edn. Springer Series in Computational Mathematics, vol. 31. Springer, Berlin (2006)
11. Hairer, E., Wanner, G.: On the Butcher group and general multi-value methods. Computing **13**(1), 1–15 (1974)
12. Higham, D.J.: Mean-square and asymptotic stability of the stochastic theta method. SIAM J. Numer. Anal. **38**(3), 753–769 (2000). doi:10.1137/S003614299834736X (electronic)
13. Jackson, K.R., Kværnø, A., Nørsett, S.P.: The use of Butcher series in the analysis of Newton-like iterations in Runge–Kutta formulas. Appl. Numer. Math. **15**(3), 341–356 (1994). doi:10.1016/0168-9274(94)00031-X. International Conference on Scientific Computation and Differential Equations (Auckland, 1993)
14. Jackson, K.R., Kværnø, A., Nørsett, S.P.: An analysis of the order of Runge–Kutta methods that use an iterative scheme to compute their internal stage values. BIT **36**(4), 713–765 (1996). doi:10.1007/BF01733789
15. Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations. Applications of Mathematics, vol. 21, 2nd edn. Springer, Berlin (1999)
16. Rößler, A.: Rooted tree analysis for order conditions of stochastic Runge–Kutta methods for the weak approximation of stochastic differential equations. Stoch. Anal. Appl. **24**(1), 97–134 (2006). doi:10.1080/07362990500397699
17. Tian, T., Burrage, K.: Implicit Taylor methods for stiff stochastic differential equations. Appl. Numer. Math. **38**(1–2), 167–185 (2001). doi:10.1016/S0168-9274(01)00034-4