Rooted Tree Analysis for Order Conditions of Stochastic Runge-Kutta Methods for the Weak Approximation of Stochastic Differential Equations

Andreas Rößler

Darmstadt University of Technology, Fachbereich Mathematik, Schlossgartenstr.7, D-64289 Darmstadt, Germany

Abstract

A general class of stochastic Runge-Kutta methods for the weak approximation of Itô and Stratonovich stochastic differential equations with a multi-dimensional Wiener process is introduced. Colored rooted trees are used to derive an expansion of the solution process and of the approximation process calculated with the stochastic Runge-Kutta method. A theorem on general order conditions for the coefficients and the random variables of the stochastic Runge-Kutta method is proved by rooted tree analysis. This theorem can be applied for the derivation of stochastic Runge-Kutta methods converging with an arbitrarily high order.

Key words: stochastic Runge-Kutta method, stochastic differential equation, weak approximation, rooted tree analysis, order condition
MSC 2000: 65C30, 60H35, 65L05, 60H10, 34F05

1 Introduction

In recent years many numerical methods have been proposed for the approximation of stochastic differential equations (SDEs), see e.g. [7], [9], [10], [13], [18] and [19]. Substantially, numerical methods for strong and for weak approximations can be distinguished. While strong approximations focus on a good approximation of the path of a solution, weak approximations are applied if a good distributional approximation is needed. In the present paper, a

⋆ 18.10.2004
Email address: roessler@mathematik.tu-darmstadt.de (Andreas Rößler).

class of stochastic Runge-Kutta (SRK) methods for weak approximation of Itô and Stratonovich SDEs is introduced in Section 2. Similar to the deterministic setting, order conditions for SRK methods are calculated by comparing the numerical solution with the exact solution over one step assuming exact initial values. Therefore, the actual solution of the SDE and the numerical approximation process have to be expanded by a stochastic Taylor series. However, even for low orders such expansions become much more complex than in the deterministic setting where it is already a lengthy task. In order to handle this task in an easy way, a rooted tree theory based on three different kinds of colored nodes is established in Section 3, which is a generalization of the rooted tree theory due to Butcher [3]. Thus, colored trees are applied in Section 4 and 5 to give a representation of the solution and the approximation process calculated with the SRK method in order to allow a rooted tree analysis of order conditions. A similar approach with two different kinds of nodes has been introduced by Burrage & Burrage [1], [2] for a SRK method converging in the strong sense as well as in Komori et al. [8] for ROW-type schemes for Stratonovich SDEs. Finally, the main Theorem 6.4 presented in Section 6 immediately yields all order conditions for the coefficients and the random variables of the introduced SRK method such that it converges with an arbitrarily given order in the weak sense. As a result of this theorem, the lengthy calculation and comparison of Taylor expansions can be avoided.

Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathcal{F}_t)_{t \geq 0}$ and let $I = [t_0, T]$ for some $0 \leq t_0 < T < \infty$. We consider the solution $(X_t)_{t \in I}$ of either a d-dimensional Itô stochastic differential equation system

$$dX_t = a(s, X_s)\, ds + b(s, X_s)\, dW_s \tag{1}$$

or a d-dimensional Stratonovich stochastic differential equation system

$$dX_t = a(s, X_s)\, ds + b(s, X_s) \circ dW_s \tag{2}$$

with an $\mathcal{F}_{t_0}$-measurable initial condition $X_{t_0} = x_0 \in \mathbb{R}^d$ such that for some $l \in \mathbb{N}$ holds $E(\|X_{t_0}\|^{2l}) < \infty$. Here, $W = ((W_t^1, \dots, W_t^m))_{t \geq 0}$ is an m-dimensional Wiener process w.r.t. $(\mathcal{F}_t)_{t \geq 0}$. SDE (1) and (2) can be written in integral form

$$X_t = x_0 + \int_{t_0}^t a(s, X_s)\, ds + \sum_{j=1}^m \int_{t_0}^t b^j(s, X_s) * dW_s^j \tag{3}$$

for $t \in I$, where the jth column of the $d \times m$-matrix function $b = (b^{i,j})$ is denoted by b^j for $j = 1, \dots, m$. Here, the second integral w.r.t. the Wiener process has to be interpreted either as an Itô integral in case of SDE (1) or as a Stratonovich integral in case of SDE (2), which is indicated by the asterisk.

The solution $(X_t)_{t \in I}$ of a Stratonovich SDE with drift a and diffusion b is also solution of an Itô SDE as in (1) and therefore also a diffusion process, however with the modified drift

$$\tilde{a}^i(t,x) = a^i(t,x) + \frac{1}{2} \sum_{j=1}^{d} \sum_{k=1}^{m} b^{j,k}(t,x) \frac{\partial b^{i,k}}{\partial x^j}(t,x) \tag{4}$$

for $i = 1, \ldots, d$ and the same diffusion b, i.e.

$$\begin{aligned} X_t &= X_{t_0} + \int_{t_0}^{t} a(s, X_s)\, ds + \int_{t_0}^{t} b(s, X_s) \circ dW_s \\ &= X_{t_0} + \int_{t_0}^{t} \tilde{a}(s, X_s)\, ds + \int_{t_0}^{t} b(s, X_s)\, dW_s. \end{aligned} \tag{5}$$

The solution of the stochastic differential equation (3) is sometimes denoted by $X^{t_0, X_{t_0}}$ in order to emphasize the initial condition. We suppose that the drift $a : I \times \mathbb{R}^d \to \mathbb{R}^d$ and the diffusion $b : I \times \mathbb{R}^d \to \mathbb{R}^{d \times m}$ are measurable functions satisfying a linear growth and a Lipschitz condition

$$\|a(t,x)\| + \|b(t,x)\| \leq C\,(1 + \|x\|) \tag{6}$$

$$\|a(t,x) - a(t,y)\| + \|b(t,x) - b(t,y)\| \leq C\,\|x - y\| \tag{7}$$

for all $x, y \in \mathbb{R}^d$ and all $t \in I$ with some constant $C > 0$. Then the conditions of the Existence and Uniqueness Theorem are fulfilled for the Itô SDE (1) (see, e.g., [6]). If the conditions also hold with a replaced by the modified drift $\tilde{a}$ in the Itô SDE, then the Existence and Uniqueness Theorem also applies to the Stratonovich SDE (2).

In the following, let $C_P^l(\mathbb{R}^d, \mathbb{R})$ denote the space of l times continuously differentiable functions $g \in C^l(\mathbb{R}^d, \mathbb{R})$ for which all partial derivatives up to order l have polynomial growth. That is, for which there exist constants $K > 0$ and $r \in \mathbb{N}$ depending on g, such that $|\partial_x^i g(x)| \leq K\,(1 + \|x\|^{2r})$ holds for all $x \in \mathbb{R}^d$ and any partial derivative $\partial_x^i g$ of order $i \leq l$.

Let $I_h = \{t_0, t_1, \ldots, t_N\}$ be a discretization of the time interval $I = [t_0, T]$ such that

$$0 \leq t_0 < t_1 < \ldots < t_N = T \tag{8}$$

and define $h_n = t_{n+1} - t_n$ for $n = 0, 1, \ldots, N-1$ with the maximum step size

$$h = \max_{0 \leq n \leq N-1} h_n.$$

In the following, we consider a class of approximation processes of the type $Y^{t,x}(t+h) = A(t, x, h; \xi)$ where ξ is a random variable or in general a vector of random variables, with moments of sufficiently high order, and A is a vector valued function of dimension d. We write $Y_n = Y^{t_0, X_{t_0}}(t_n)$ and we construct the sequence

$$\begin{aligned} Y_0 &= X_{t_0} \\ Y_{n+1} &= A(t_n, Y_n, h; \xi_n), \qquad n = 0, 1, \ldots, N-1, \end{aligned} \tag{9}$$

where ξ_0 is independent of Y_0, while ξ_n for $n \geq 1$ is independent of $Y_0, \ldots, Y_n$ and $\xi_0, \ldots, \xi_{n-1}$. Then we can define weak convergence with some order p of an approximation process.

Definition 1.1 *A time discrete approximation process $Y = (Y_t)_{t \in I_h}$ converges weakly with order p to X as $h \to 0$ at time T if for each $f \in C_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R})$ exists a constant C_f, which does not depend on h, and a finite $h_0 > 0$ such that*

$$|E(f(X_T)) - E(f(Y_T))| \leq C_f \, h^p \tag{10}$$

holds for each $h \in]0, h_0[$.

Since we are interested in calculating a global approximation converging in the weak sense with some desired order p, we make use of the following slightly modified theorem due to Milstein (1986) [12].

Theorem 1.2 *Suppose the following conditions hold:*

(i) *the coefficients a^i in the case of SDE (1), $\tilde{a}^i$ in the case of SDE (2) and $b^{i,j}$ are continuous, satisfy a Lipschitz condition (7) and belong to $C_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R})$ with respect to x for $i = 1, \ldots, d$, $j = 1, \ldots, m$,*

(ii) *for sufficiently large r (specified below) the moments $E(\|Y_n\|^{2r})$ exist and are uniformly bounded with respect to N and $n = 0, 1, \ldots, N$,*

(iii) *assume that for all $f \in C_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R})$ the following* local error estimation

$$|E(f(X^{t,x}(t+h))) - E(f(Y^{t,x}(t+h)))| \leq K(x) \, h^{p+1} \tag{11}$$

is valid for $x \in \mathbb{R}^d$, $t, t+h \in I$ and $K \in C_P^0(\mathbb{R}^d, \mathbb{R})$.

Then for all N and all $n = 0, 1, \ldots, N$ the following global error estimation

$$|E(f(X^{t_0, X_{t_0}}(t_n))) - E(f(Y^{t_0, X_{t_0}}(t_n)))| \leq C \, h^p \tag{12}$$

holds for all $f \in C_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R})$, where C is a constant, i.e. the method (9) has order of accuracy p in the sense of weak approximation.

A proof of Theorem 1.2 can be found in [11], [12] and [16]. Lemma 1.3 gives sufficient conditions for assumption (ii) of Theorem 1.2 (see also [11], [12]).

Lemma 1.3 *Suppose that for $h < 1$ the conditions*

$$\|E(A(t_n, x, h; \xi_n) - x)\| \leq C_1 (1 + \|x\|) \, h, \tag{13}$$

$$\|A(t_n, x, h; \xi_n) - x\| \leq M(\xi_n)(1 + \|x\|) \, h^{1/2} \tag{14}$$

hold where $M(\xi_n)$ has moments of all orders, i.e. $E((M(\xi_n))^i) \leq C_2$, $i \in \mathbb{N}$, with constants C_1 and C_2 independent of h. Then for every even number $2r$ the expectations $E(\|Y_n\|^{2r})$ exist and are uniformly bounded with respect to N and $n = 1, \ldots, N$, if only $E(\|Y_0\|^{2r})$ exists.

2 A Class of Stochastic Runge-Kutta Methods

In the following a class of *stochastic Runge-Kutta methods* is introduced for the approximation of both, Itô and Stratonovich stochastic differential equation systems w.r.t. an m-dimensional Wiener process. In order to preserve the most possible generality, the considered class of stochastic Runge-Kutta methods is of type (9) and has the following structure

$$\begin{aligned} Y_0 &= x_0 \\ Y_{n+1} &= A(t_n, Y_n, h_n; \theta_\nu(h_n) : \nu \in \mathcal{M}) \end{aligned} \tag{15}$$

where $\mathcal{M}$ is an arbitrary finite set of multi-indices with $\kappa = |\mathcal{M}|$ elements and $\theta_\nu(h)$, $\nu \in \mathcal{M}$, are some suitable random variables. For the weak approximation of the solution $(X_t)_{t \in I}$ of the d-dimensional SDE system (3), considered either with respect to Itô or Stratonovich calculus, the general class of s-stage stochastic Runge-Kutta methods is given by $Y_0 = x_0$ and

$$\begin{aligned} Y_{n+1} = Y_n &+ \sum_{i=1}^{s} z_i^{(0,0)} \; a\left(t_n + c_i^{(0,0)} h_n, H_i^{(0,0)}\right) \\ &+ \sum_{i=1}^{s} \sum_{k=1}^{m} \sum_{\nu \in \mathcal{M}} z_i^{(k,\nu)} \; b^k\left(t_n + c_i^{(k,\nu)} h_n, H_i^{(k,\nu)}\right) \end{aligned} \tag{16}$$

with

$$\begin{aligned} H_i^{(0,0)} &= Y_n + \sum_{j=1}^{s} Z_{ij}^{(0,0),(0,0)} \; a\left(t_n + c_j^{(0,0)} h_n, H_j^{(0,0)}\right) \\ &\quad + \sum_{j=1}^{s} \sum_{r=1}^{m} \sum_{\mu \in \mathcal{M}} Z_{ij}^{(0,0),(r,\mu)} \; b^r\left(t_n + c_j^{(r,\mu)} h_n, H_j^{(r,\mu)}\right) \\ H_i^{(k,\nu)} &= Y_n + \sum_{j=1}^{s} Z_{ij}^{(k,\nu),(0,0)} \; a\left(t_n + c_j^{(0,0)} h_n, H_j^{(0,0)}\right) \\ &\quad + \sum_{j=1}^{s} \sum_{r=1}^{m} \sum_{\mu \in \mathcal{M}} Z_{ij}^{(k,\nu),(r,\mu)} \; b^r\left(t_n + c_j^{(r,\mu)} h_n, H_j^{(r,\mu)}\right) \end{aligned}$$

for $i = 1, \ldots, s$, $k = 1, \ldots, m$, $\nu \in \mathcal{M}$ and $n = 0, 1, \ldots, N-1$ where

$$\begin{aligned} z_i^{(0,0)} &= \alpha_i \, h_n & z_i^{(k,\nu)} &= \sum_{\iota \in \mathcal{M}} \gamma_i^{(\iota)\,(k,\nu)} \; \theta_\iota(h_n) \\ Z_{ij}^{(0,0),(0,0)} &= A_{ij}^{(0,0),(0,0)} \, h_n & Z_{ij}^{(0,0),(r,\mu)} &= \sum_{\iota \in \mathcal{M}} B_{ij}^{(\iota)\,(0,0),(r,\mu)} \; \theta_\iota(h_n) \\ Z_{ij}^{(k,\nu),(0,0)} &= A_{ij}^{(k,\nu),(0,0)} \, h_n & Z_{ij}^{(k,\nu),(r,\mu)} &= \sum_{\iota \in \mathcal{M}} B_{ij}^{(\iota)\,(k,\nu),(r,\mu)} \; \theta_\iota(h_n) \end{aligned}$$

for $i,j = 1,\dots,s$. Here $\alpha_i, \gamma_i^{(\iota)(k,\nu)}, A_{ij}^{(k,\nu),(0,0)}, B_{ij}^{(\iota)(k,\nu),(r,\mu)} \in \mathbb{R}$ are the coefficients of the SRK method and as usual the weights can be defined by

$$c^{(0,0)} = A^{(0,0),(0,0)} e \qquad\qquad c^{(k,\nu)} = A^{(k,\nu),(0,0)} e \tag{17}$$

with $e = (1,\dots,1)^T$. If $A_{ij}^{(k,\nu),(0,0)} = B_{ij}^{(\iota)(k,\nu),(r,\mu)} = 0$ for $j \geq i$ then (16) is called an explicit SRK method, otherwise it is called implicit. The introduced class of SRK methods can be characterized by an extended Butcher array

$$\begin{array}{c|c|c|c|c} c^{(0,0)} & A^{(0,0),(0,0)} & B^{(\iota_1)(0,0),(r,\mu)} & \dots\dots\dots & B^{(\iota_\kappa)(0,0),(r,\mu)} \\ \hline c^{(k,\nu)} & A^{(k,\nu),(0,0)} & B^{(\iota_1)(k,\nu),(r,\mu)} & \dots\dots\dots & B^{(\iota_\kappa)(k,\nu),(r,\mu)} \\ \hline & \alpha^T & \gamma^{(\iota_1)(k,\nu)^T} & \dots\dots\dots & \gamma^{(\iota_\kappa)(k,\nu)^T} \end{array} \tag{18}$$

for $k,r = 1,\dots,m$ and $\iota_i, \nu, \mu \in \mathcal{M}$ for $1 \leq i \leq \kappa$. We assume that the random variables $\theta_\nu(h_n)$ satisfy the moment condition

$$E\left(\theta_{\nu_1}^{p_1}(h_n) \cdot \ldots \cdot \theta_{\nu_\kappa}^{p_\kappa}(h_n)\right) = O\left(h_n^{(p_1+\ldots+p_\kappa)/2}\right) \tag{19}$$

for all $p_i \in \mathbb{N}_0$ and $\nu_i \in \mathcal{M}$, $1 \leq i \leq \kappa$. This moment condition ensures a contribution of each random variable having an order of magnitude $O(\sqrt{h})$, i.e. having mean-square order $\frac{1}{2}$.

Some SRK schemes which belong to the introduced general class of SRK methods can be found in [14], [15] and [16]. Further, many schemes proposed in recent literature like in [7] or [20] are also covered. Usually, the set $\mathcal{M}$ may consist of some multi-indices $(j_1,\dots,j_l)$ with $0 \leq j_i \leq m$ for $i = 1,\dots,l$ and the random variables may be chosen as multiple Itô or Stratonovich integrals of type $I_{(j_1,\dots,j_l)}/h^q$ or $J_{(j_1,\dots,j_l)}/h^q$, depending on the calculus that is used.

3 Stochastic Rooted Tree Theory

The SDE system (3) can be represented by an autonomous SDE system

$$X_t = x_0 + \int_{t_0}^t a(X_s)\, ds + \sum_{j=1}^m \int_{t_0}^t b^j(X_s) \; * dW_s^j \tag{20}$$

with one additional equation representing time. Hence without loss of generality, it is sufficient to treat autonomous SDE systems in the following. First of all a definition of colored graphs which will be suitable in the rooted tree theory for SDEs w.r.t. a multi-dimensional Wiener process is given (see [17]).

Definition 3.1 *Let l be a positive integer.*

(1) A monotonically labelled S-tree (stochastic tree) $\boldsymbol{t}$ *with $l = l(\boldsymbol{t})$ nodes is a pair of maps $\boldsymbol{t} = (\boldsymbol{t}', \boldsymbol{t}'')$ with*

$$\boldsymbol{t}' : \{2, \ldots, l\} \to \{1, \ldots, l-1\}$$
$$\boldsymbol{t}'' : \{1, \ldots, l\} \to A$$

so that $\boldsymbol{t}'(i) < i$ for $i = 2, \ldots, l$. Unless otherwise noted, we choose the set $A = \{\gamma, \tau, \sigma_{j_k}, k \in \mathbb{N}\}$ where j_k is a variable index with $j_k \in \{1, \ldots, m\}$.

(2) LTS denotes the set of all monotonically labelled S-trees. Here two trees $\boldsymbol{t} = (\boldsymbol{t}', \boldsymbol{t}'')$ and $\boldsymbol{u} = (\boldsymbol{u}', \boldsymbol{u}'')$ just differing by their colors $\boldsymbol{t}''$ and $\boldsymbol{u}''$ are considered to be identical if there exists a bijective map $\pi : A \to A$ with $\pi(\gamma) = \gamma$ and $\pi(\tau) = \tau$ so that $\boldsymbol{t}''(i) = \pi(\boldsymbol{u}''(i))$ holds for $i = 1, \ldots, l$.

As a result of this $\mathbf{t}'$ defines a father son relation between the nodes, i.e. $\mathbf{t}'(i)$ is the father of the son i. Furthermore the color $\mathbf{t}''(i)$, which consists of one element of the set A, is added to the node i for $i = 1, \ldots, l(\mathbf{t})$. The node of type $\gamma = \otimes$ is denoted as the *root* and always sketched as the lowest node of the graph, $\tau = \bullet$ is a deterministic node and $\sigma_{j_k} = \bigcirc_{j_k}$ is a stochastic node with a variable index $j_k \in \{1, \ldots, m\}$. So the variable index j_k is associated with the j_kth component of the corresponding m-dimensional Wiener process of the considered SDE. In case of a one-dimensional Wiener process one can omit the variable indices since we have $j_k = 1$ for all $k \in \mathbb{N}$ (see also [16]). As an example Figure 1 presents two elements of LTS.

Fig. 1. Two elements of LTS with $j_1, j_2 \in \{1, \ldots, m\}$.

For the labelled S-tree $\mathbf{t}_I$ in Figure 1 we have $\mathbf{t}_I'(2) = \mathbf{t}_I'(3) = 1$ and $\mathbf{t}_I'(4) = 2$. The color of the nodes is given by $\mathbf{t}_I''(1) = \gamma$, $\mathbf{t}_I''(2) = \tau$, $\mathbf{t}_I''(3) = \sigma_{j_1}$ and $\mathbf{t}_I''(4) = \sigma_{j_2}$.

Definition 3.2 *Let $\boldsymbol{t} = (\boldsymbol{t}', \boldsymbol{t}'') \in LTS$. We denote by $d(\boldsymbol{t}) = \sharp\{i : \boldsymbol{t}''(i) = \tau\}$ the number of deterministic nodes, by $s(\boldsymbol{t}) = \sharp\{i : \boldsymbol{t}''(i) = \sigma_{j_k}, k \in \mathbb{N}\}$ the number of stochastic nodes and by $n(\boldsymbol{t}) = \sharp\{i : \boldsymbol{t}''(i) = \boldsymbol{t}''(i+1) = \sigma_{j_k}, k \in \mathbb{N}\}$ the number of pairs of stochastic nodes with the same variable index. The* order *$\rho(\boldsymbol{t})$ of the tree $\boldsymbol{t}$ is defined as $\rho(\boldsymbol{t}) = d(\boldsymbol{t}) + \frac{1}{2}s(\boldsymbol{t})$ with $\rho(\gamma) = 0$.*

The order of the trees $\mathbf{t}_I$ and $\mathbf{t}_{II}$ presented in Figure 1 can be calculated as $\rho(\mathbf{t}_I) = \rho(\mathbf{t}_{II}) = 2$. Every labelled S-tree can be written as a combination of three different brackets defined as follows.

Definition 3.3 *If $\mathbf{t}_1, \ldots, \mathbf{t}_k$ are colored trees then we denote by*

$$(\mathbf{t}_1, \ldots, \mathbf{t}_k), \quad [\mathbf{t}_1, \ldots, \mathbf{t}_k] \quad \textit{and} \quad \{\mathbf{t}_1, \ldots, \mathbf{t}_k\}_j$$

the tree in which $\mathbf{t}_1, \ldots, \mathbf{t}_k$ are each joined by a single branch to $\otimes$, $\bullet$ and $\bigcirc_j$, respectively (see Figure 2).

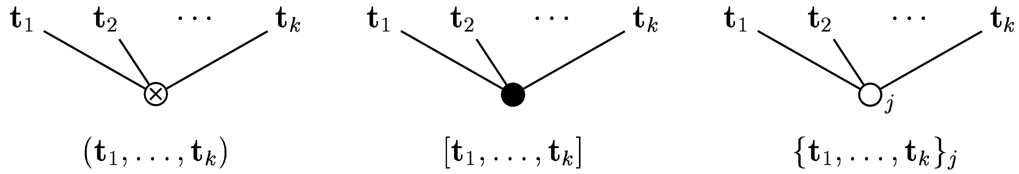


Fig. 2. Writing a colored S-tree with brackets.

Therefore proceeding recursively, for the two examples $\mathbf{t}_I$ and $\mathbf{t}_{II}$ in Figure 1 we obtain $\mathbf{t}_I = ([\bigcirc_{j_2}], \bigcirc_{j_1}) = ([\sigma_{j_2}], \sigma_{j_1})$ and $\mathbf{t}_{II} = (\{\bullet, \bigcirc_{j_2}\}_{j_1}) = (\{\tau, \sigma_{j_2}\}_{j_1})$.

Due to the fact that we are interested in calculating weak approximations, it will turn out that we can concentrate our considerations to one representative tree of each equivalence class.

Definition 3.4 *Let $\mathbf{t} = (\mathbf{t}', \mathbf{t}'')$ and $\mathbf{u} = (\mathbf{u}', \mathbf{u}'')$ be elements of LTS. Then the trees $\mathbf{t}$ and $\mathbf{u}$ are equivalent, i.e. $\mathbf{t} \sim \mathbf{u}$, if the following hold:*

(i) $l(\mathbf{t}) = l(\mathbf{u})$

(ii) There exist two bijective maps

$$\begin{array}{l} \psi : \{1, \ldots, l(\mathbf{t})\} \to \{1, \ldots, l(\mathbf{t})\} \quad \textit{with} \quad \psi(1) = 1, \\ \pi : A \to A \quad \textit{with} \quad \pi(\gamma) = \gamma \quad \textit{and} \quad \pi(\tau) = \tau, \end{array}$$

so that the following diagram commutes

$$\begin{array}{ccccc} \{2, \ldots, l(\mathbf{t})\} & \xrightarrow{\mathbf{t}'} & \{1, \ldots, l(\mathbf{t})\} & \xrightarrow{\mathbf{t}''} & \\ \downarrow \psi & & \downarrow \psi & & A \\ \{2, \ldots, l(\mathbf{t})\} & \xrightarrow{\mathbf{u}'} & \{1, \ldots, l(\mathbf{t})\} & \xrightarrow{\pi(\mathbf{u}'')} & \end{array}$$

The set of all equivalence classes under the relation $\sim$ is denoted by $TS = LTS/\sim$. We denote by $\alpha(\mathbf{t})$ the cardinality of $\mathbf{t}$, i.e. the number of possibilities of monotonically labelling the nodes of $\mathbf{t}$ with numbers $1, \ldots, l(\mathbf{t})$.

Thus, a monotonically labelled S-tree $\mathbf{u}$ is equivalent to $\mathbf{t}$, if each label i is replaced by $\psi(i)$ and if each stochastic node σ_{j_k} with variable index j_k is replaced by an other stochastic node $\pi(\sigma_{j_k})$. Thus, all trees in Figure 3 belong to the same equivalence class as $\mathbf{t}_I$ in the example above, since the indices j_1 and j_2 are just renamed either by j_2 and j_1 or j_8 and j_3, respectively. Finally the graphs differ only in the labelling of their number indices.

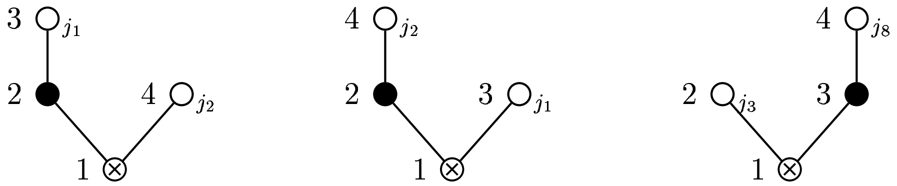


Fig. 3. Trees of the same equivalence class.

For every rooted tree $\mathbf{t} \in LTS$, there exists a corresponding *elementary differential* which is a direct generalization of the differential in the deterministic case (see, e.g., [3]). For $j \in \{1, \ldots, m\}$, the elementary differential is defined recursively by

$$F(\gamma)(x) = f(x), \qquad F(\tau)(x) = a(x), \qquad F(\sigma_j)(x) = b^j(x),$$

for single nodes and by

$$F(\mathbf{t})(x) = \begin{cases} f^{(k)}(x) \cdot (F(\mathbf{t}_1)(x), \ldots, F(\mathbf{t}_k)(x)) & \text{for } \mathbf{t} = (\mathbf{t}_1, \ldots, \mathbf{t}_k) \\ a^{(k)}(x) \cdot (F(\mathbf{t}_1)(x), \ldots, F(\mathbf{t}_k)(x)) & \text{for } \mathbf{t} = [\mathbf{t}_1, \ldots, \mathbf{t}_k] \\ {b^j}^{(k)}(x) \cdot (F(\mathbf{t}_1)(x), \ldots, F(\mathbf{t}_k)(x)) & \text{for } \mathbf{t} = \{\mathbf{t}_1, \ldots, \mathbf{t}_k\}_j \end{cases} \tag{21}$$

for a tree $\mathbf{t}$ with more than one node. Here $f^{(k)}$, $a^{(k)}$ and ${b^j}^{(k)}$ define a symmetric k-linear differential operator, and one can choose the sequence of labelled S-trees $\mathbf{t}_1, \ldots, \mathbf{t}_k$ in an arbitrary order. For example, the Ith component of $a^{(k)} \cdot (F(\mathbf{t}_1), \ldots, F(\mathbf{t}_k))$ can be written as

$$(a^{(k)} \cdot (F(\mathbf{t}_1), \ldots, F(\mathbf{t}_k)))^I = \sum_{J_1, \ldots, J_k = 1}^{d} \frac{\partial^k a^I}{\partial x^{J_1} \ldots \partial x^{J_k}} (F^{J_1}(\mathbf{t}_1), \ldots, F^{J_k}(\mathbf{t}_k))$$

where the components of vectors are denoted by superscript indices, which are chosen as capitals. As a result of this we get for $\mathbf{t}_I$ and $\mathbf{t}_{II}$ the elementary differentials

$$F(\mathbf{t}_I) = f''(a'(b^{j_2}), b^{j_1}) = \sum_{J_1, J_2 = 1}^{d} \frac{\partial^2 f}{\partial x^{J_1} \partial x^{J_2}} \Big(\sum_{K_1 = 1}^{d} \frac{\partial a^{J_1}}{\partial x^{K_1}} b^{K_1, j_2} \cdot b^{J_2, j_1} \Big)$$

$$F(\mathbf{t}_{II}) = f'({b^{j_1}}''(a, b^{j_2})) = \sum_{J_1 = 1}^{d} \frac{\partial f}{\partial x^{J_1}} \Big(\sum_{K_1, K_2 = 1}^{d} \frac{\partial^2 b^{J_1, j_1}}{\partial x^{K_1} \partial x^{K_2}} a^{K_1} \cdot b^{K_2, j_2} \Big)$$

It has to be pointed out that the elementary differentials for the trees presented in Figure 3 coincide with $F(\mathbf{t}_I)$ if the variable indices j_i are simply renamed by a suitable bijective mapping π.

4 Taylor Expansion for Itô and Stratonovich SDEs

For the expansion of the expectation of some functional applied to the solution $(X_t)_{t\in I}$ of the d-dimensional SDE (20) considered either w.r.t. Itô or Stratonovich calculus, some subsets $LTS(I)$ and $LTS(S)$ of LTS have to be introduced, respectively.

Definition 4.1 *For $* \in \{I, S\}$ let $LTS(*)$ denote the set of trees $\boldsymbol{t} \in LTS$ having a root $\gamma = \otimes$ and which can be constructed by a finite number of steps of the form*

a) adding a deterministic node $\tau = \bullet$, or
b) adding two stochastic nodes $\sigma_{j_k} = \bigcirc_{j_k}$, where both nodes get the same new variable index j_k for some $k \in \mathbb{N}$. Additionally, in the case of $* = I$ neither of the two nodes is allowed to be the father of the other.

The nodes have to be labelled in the same order as they have been added by the construction of the tree. Further $TS(*) = LTS(*)/\sim$ denotes the equivalence class under the relation of Definition 3.4 restricted to $LTS(*)$ and $\alpha_*(\boldsymbol{t})$ denotes the cardinality of $\boldsymbol{t}$ in $LTS(*)$ for $* \in \{I, S\}$, respectively.

Since the number of stochastic nodes is always even with $n(\mathbf{t}) = s(\mathbf{t})/2$, the order $\rho(\mathbf{t})$ has to be an integer and $\mathbf{t}$ owns the variable indices $j_1, \ldots, j_{n(\mathbf{t})}$. As the construction of the trees in $LTS(I)$ is more restrictive than of the ones in $LTS(S)$, it holds $LTS(I) \subset LTS(S)$.

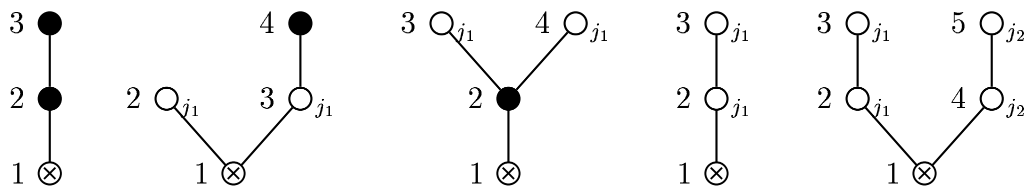


Fig. 4. Some trees which belong to $LTS(I)$ or $LTS(S)$.

All trees of Figure 4 belong to $LTS(S)$, however only the first three trees belong to $LTS(I)$. For the last tree, there is similar tree $(\{\sigma_{j_2}\}_{j_1}, \{\sigma_{j_2}\}_{j_1})$ which belongs to $LTS(I)$. The only difference is the sequence of the construction, i.e. the correct father-son relationship for the stochastic nodes. Clearly, a tree like $(\{\tau\}_{j_1})$ or $(\{[\sigma_{j_1}]\}_{j_1})$ neither belongs to $LTS(I)$ nor to $LTS(S)$.

The following result gives an expansion for the solution process of an Itô and a Stratonovich SDE, respectively, by the use of colored rooted trees.

Theorem 4.2 *Let $(X_t)_{t\in I}$ be the solution of the stochastic differential equation system (20) with initial value $X_{t_0} = x_0 \in \mathbb{R}^d$. Then for $p \in \mathbb{N}_0$ and*

$f, a^i, \tilde{a}^i, b^{i,j} \in C_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R})$ for $i = 1, \ldots, d$, $j = 1, \ldots, m$ and for $t \in [t_0, T]$ the following truncated expansion holds:

$$\begin{aligned} E^{t_0,x_0}(f(X_t)) &= \sum_{\substack{\boldsymbol{t} \in LTS(*) \\ \rho(\boldsymbol{t}) \le p}} \sum_{j_1,\ldots,j_{s(\boldsymbol{t})/2}=1}^{m} \frac{F(\boldsymbol{t})(x_0)}{2^{s(\boldsymbol{t})/2}\,\rho(\boldsymbol{t})!} (t-t_0)^{\rho(\boldsymbol{t})} + O((t-t_0)^{p+1}) \\ &= \sum_{\substack{\boldsymbol{t} \in TS(*) \\ \rho(\boldsymbol{t}) \le p}} \sum_{j_1,\ldots,j_{s(\boldsymbol{t})/2}=1}^{m} \frac{\alpha_*(\boldsymbol{t})\,F(\boldsymbol{t})(x_0)}{2^{s(\boldsymbol{t})/2}\,\rho(\boldsymbol{t})!} (t-t_0)^{\rho(\boldsymbol{t})} + O((t-t_0)^{p+1}) \end{aligned} \tag{22}$$

Here, $* = I$ for the Itô version of SDE (20), and $* = S$ for the Stratonovich version of SDE (20).

Proof. For a proof we refer to Theorem 3.2 and Theorem 4.2 together with Proposition 5.1 in [17]. □

5 Taylor Expansion for the SRK method

In order to derive conditions such that the stochastic Runge-Kutta method (16) converges in the weak sense with some specified order, a Taylor expansion of the numerical solution based on colored rooted trees has to be developed. We follow the approach of Butcher [3] in a similar way as in Burrage and Burrage [1], [2], Hairer [4], Hairer, Nørsett and Wanner [5] and Rößler [16].

For notational convenience, in this section we define $\overline{\mathcal{M}} = \mathcal{M} \cup \{0\}$ and we denote by $\theta_0(h) = h$ and by $\theta(h) = (\theta_0(h), \theta_{\nu_1}(h), \ldots, \theta_{\nu_\kappa}(h))^T$, $\nu_i \in \mathcal{M}$, the corresponding $\kappa+1$-dimensional vector[1] with $\kappa = |\mathcal{M}|$. Further, it is assumed that $\theta_\nu(0) = 0$ for all $\nu \in \overline{\mathcal{M}}$. Due to condition (17), it is sufficient to consider autonomous SRK methods (16) in the following. We denote t_n by t_0 and for a given $t = t_0 + h$ the approximations Y_n and Y_{n+1} are denoted by $Y(t_0)$ and $Y(t)$, respectively. Further, the values $H_i^{(k,\nu)}$ are denoted by $H_i^{(k,\nu)}(t)$ in order to stress the dependency on t of the random variables $\theta_0(t-t_0), \theta_{\nu_1}(t-t_0), \ldots, \theta_{\nu_\kappa}(t-t_0)$ appearing in $H_i^{(k,\nu)}$. For the Taylor expansion of the SRK method $Y(t) = A(Y(t_0), \theta(t-t_0))$ as a function of $\theta_0, \theta_{\nu_1}, \ldots, \theta_{\nu_\kappa}$, the differential operator $\mathcal{D}^k$ for $k \in \mathbb{N}$ is introduced as

$$\mathcal{D}^k = \sum_{\nu_1,\ldots,\nu_k \in \overline{\mathcal{M}}} \Delta\theta_{\nu_1} \cdot \Delta\theta_{\nu_2} \cdot \ldots \cdot \Delta\theta_{\nu_k} \cdot \frac{\partial^k}{\partial\theta_{\nu_1}\partial\theta_{\nu_2}\ldots\partial\theta_{\nu_k}} \tag{23}$$

[1] Then $Y(t) = A(t_0, Y(t_0), \theta(t-t_0))$ and $Y(t_0) = A(t_0, Y(t_0), 0, \ldots, 0)$.

with $\Delta\theta_\nu = \theta_\nu(h) - \theta_\nu(0)$ and we denote by $\mathcal{D}^0 \equiv \mathrm{Id}$. Under the assumption that f, a and b^j, $1 \leq j \leq m$, are sufficiently differentiable, we apply the Theorem of Taylor and get for $n \in \mathbb{N}$

$$f(Y(t)) = \sum_{k=0}^{n} \frac{\mathcal{D}^k f(Y(t_0))}{k!} + \mathcal{R}_n(t, t_0) \tag{24}$$

with a remainder term $\mathcal{R}_n$ which can be written in Lagrange form as

$$\mathcal{R}_n(t, t_0) = \frac{\mathcal{D}^{n+1} f(Y(t_0 + \xi\, h))}{(n+1)!} \tag{25}$$

with some $\xi \in]0, 1[$ and $h = t - t_0$.

The next step is the computation of $\mathcal{D}^k f(Y(t_0))$ for $k \in \mathbb{N}_0$, i.e. the kth derivative of the numerical solution $f(Y(t))$. Therefore, generalized versions of the *Leibniz formula* and of *Faà di Bruno's formula* (see, e.g., [5]) are helpful.

To begin with, a multi-dimensional version of the Leibniz formula fitted to the expansion of the SRK method is given. Let $q \in \mathbb{N}$, $k \in \{0, 1, \ldots, m\}$ and $\nu \in \overline{\mathcal{M}}$. Then the formula

$$\begin{aligned}
\sum_{\nu_1,\ldots,\nu_q \in \overline{\mathcal{M}}} \frac{\partial^q H_i^{(k,\nu)}(t)^J}{\partial\theta_{\nu_1} \ldots \partial\theta_{\nu_q}} = {} & q \cdot \sum_{j=1}^{s} A_{ij}^{(k,\nu),(0,0)} \sum_{\nu_1,\ldots,\nu_{q-1} \in \overline{\mathcal{M}}} \frac{\partial^{q-1} a(H_j^{(0,0)}(t))^J}{\partial\theta_{\nu_1} \ldots \partial\theta_{\nu_{q-1}}} \\
& + q \cdot \sum_{j=1}^{s} \sum_{r=1}^{m} \sum_{\mu,\nu_q \in \mathcal{M}} B_{ij}^{(\nu_q)(k,\nu),(r,\mu)} \sum_{\nu_1,\ldots,\nu_{q-1} \in \overline{\mathcal{M}}} \frac{\partial^{q-1} b^r(H_j^{(r,\mu)}(t))^J}{\partial\theta_{\nu_1} \ldots \partial\theta_{\nu_{q-1}}} \\
& + \sum_{j=1}^{s} A_{ij}^{(k,\nu),(0,0)}\, \theta_0(t - t_0) \sum_{\nu_1,\ldots,\nu_q \in \overline{\mathcal{M}}} \frac{\partial^q a(H_j^{(0,0)}(t))^J}{\partial\theta_{\nu_1} \ldots \theta_{\nu_q}} \\
& + \sum_{j=1}^{s} \sum_{r=1}^{m} \sum_{\mu,\iota \in \mathcal{M}} B_{ij}^{(\iota)(k,\nu),(r,\mu)}\, \theta_\iota(t - t_0) \sum_{\nu_1,\ldots,\nu_q \in \overline{\mathcal{M}}} \frac{\partial^q b^r(H_j^{(r,\mu)}(t))^J}{\partial\theta_{\nu_1} \ldots \theta_{\nu_q}}
\end{aligned} \tag{26}$$

can be easily calculated (see also [16], Lemma 2.5.3). In order to state a generalized version of Faà di Bruno's formula [5], we introduce a special set of trees corresponding to the derivatives of the composition of two functions. For example, if we consider $g \circ h$, we get for the Jth component of the third

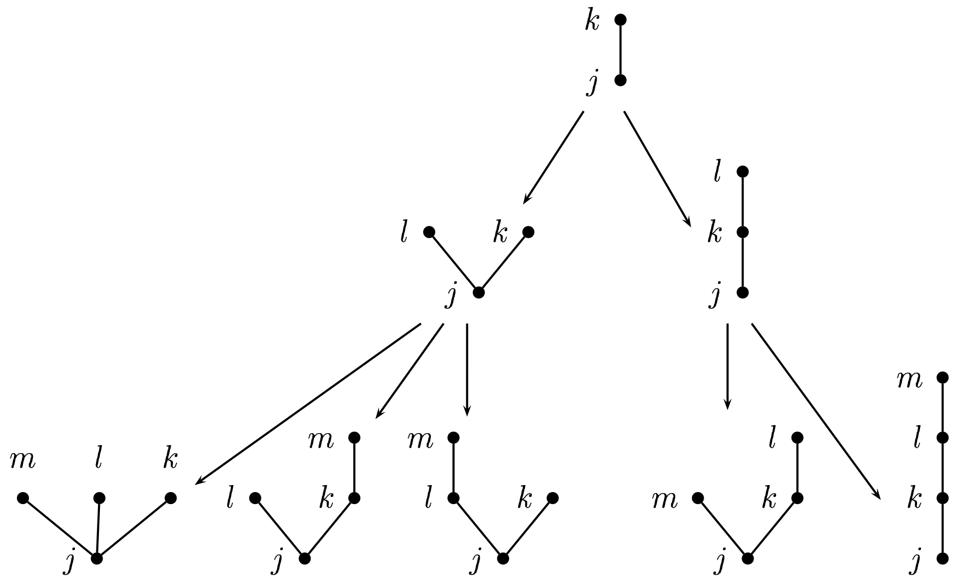


Fig. 5. Some special trees representing the derivatives of $g(h)^J$.

derivative

$$\frac{\partial^3 g(h)^J}{\partial x^K \partial x^L \partial x^M} = \sum_{K_1,K_2,K_3} g^J_{K_1 K_2 K_3}(h) \left(\frac{\partial h^{K_1}}{\partial x^M} \cdot \frac{\partial h^{K_2}}{\partial x^L} \cdot \frac{\partial h^{K_3}}{\partial x^K} \right)$$
$$+ \sum_{K_1,K_2} g^J_{K_1 K_2}(h) \left(\frac{\partial h^{K_1}}{\partial x^L} \cdot \frac{\partial^2 h^{K_2}}{\partial x^K \partial x^M} \right) + \sum_{K_1,K_2} g^J_{K_1 K_2}(h) \left(\frac{\partial^2 h^{K_1}}{\partial x^L \partial x^M} \cdot \frac{\partial h^{K_2}}{\partial x^K} \right)$$
$$+ \sum_{K_1,K_2} g^J_{K_1 K_2}(h) \left(\frac{\partial h^{K_1}}{\partial x^M} \cdot \frac{\partial^2 h^{K_2}}{\partial x^K \partial x^L} \right) + \sum_{K_1} g^J_{K_1}(h) \left(\frac{\partial^3 h^{K_1}}{\partial x^K \partial x^L \partial x^M} \right). \tag{27}$$

The corresponding special trees are presented in the last line of Figure 5. Here the number m of indices $K_1, \ldots, K_m$ depends on the number of ramifications of the root. Each time $g(h)^J$ is differentiated, one has to

(i) differentiate the first factor $g^J_{K_1 \ldots}$, i.e., add a new branch to the root j,

(ii) increase the number of derivatives of each of the h functions by 1, which is presented by lengthening the corresponding branch.

So each time we differentiate, we have to add a new label. *All* trees which are obtained in this way are those *special trees* which have no ramifications except at the root.

In order to take into account colored stochastic trees with their meaning for the expansion of the SRK method in the following, special trees having either a root of type γ, τ or σ_j have to be considered. This is due to the analysis of the composed functions $f(Y(t))$, $a(H_i^{(0,0)}(t))$ and $b^j(H_i^{(j,\nu)}(t))$.

Definition 5.1 *The set of* special labelled trees *with q nodes having no ramifications except at the root is denoted by $SLTS_q$. For $\boldsymbol{u} \in SLTS_q$ we denote by $m = m(\boldsymbol{u})$ the number of ramifications of the root of $\boldsymbol{u}$. Further we denote by $SLTS_q^{(M)} \subset SLTS_q$ with $M \subset A = \{\gamma, \tau, \sigma_{j_k} : k \in \mathbb{N}\}$ the set of special labelled trees in $SLTS_q$ having a root of type π with $\pi \in M$.*

Now a formula similar to Faà di Bruno's formula fitted to the stochastic setting can be stated.

Lemma 5.2 *For $q \in \mathbb{N}$, $\pi \in A$ and functions $g : \mathbb{R}^d \to \mathbb{R}^r$ and $h : \mathbb{R}^{\kappa+1} \to \mathbb{R}^d$, the multi-dimensional chain rule*

$$\sum_{\nu_1,\ldots,\nu_q \in \overline{\mathcal{M}}} \frac{\partial^q g(h)^J}{\partial\theta_{\nu_1}\ldots\partial\theta_{\nu_q}} = \sum_{\boldsymbol{u}\in SLTS_{q+1}^{(\pi)}} \sum_{K_1,\ldots,K_{m(\boldsymbol{u})}=1}^{d} g^J_{K_1\ldots K_{m(\boldsymbol{u})}}(h) \times$$
$$\times \left(\left(\sum_{\nu_1,\ldots,\nu_{\delta_1}\in\overline{\mathcal{M}}} \frac{\partial^{\delta_1} h^{K_1}}{\partial\theta_{\nu_1}\ldots\partial\theta_{\nu_{\delta_1}}} \right) \cdot \ldots \cdot \left(\sum_{\nu_1,\ldots,\nu_{\delta_{m(\boldsymbol{u})}}\in\overline{\mathcal{M}}} \frac{\partial^{\delta_{m(\boldsymbol{u})}} h^{K_{m(\boldsymbol{u})}}}{\partial\theta_{\nu_1}\ldots\partial\theta_{\nu_{\delta_{m(\boldsymbol{u})}}}} \right) \right) \tag{28}$$

holds. Here $m = m(\boldsymbol{u})$ denotes the number of ramifications of the root of the special tree $\boldsymbol{u} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m)_\pi$ with a root of type $\pi \in A$ and $\delta_i = l(\boldsymbol{u}_i)$ describes the number of nodes of the subtree $\boldsymbol{u}_i$ for $1 \leq i \leq m$ with $\delta_1 + \ldots + \delta_{m(\boldsymbol{u})} = q$ for all $\boldsymbol{u} \in SLTS_{q+1}^{(\pi)}$.

Proof. We prove Lemma 5.2 by induction on q. For $q = 1$ and $\pi \in A$ we have

$$\sum_{\nu_1\in\overline{\mathcal{M}}} \frac{\partial g(h)^J}{\partial\theta_{\nu_1}} = \sum_{\nu_1\in\overline{\mathcal{M}}} \sum_{K_1=1}^{d} \frac{\partial g(h)^J}{\partial x^{K_1}} \cdot \frac{\partial h^{K_1}}{\partial\theta_{\nu_1}} = \sum_{\mathbf{u}\in SLTS_2^{(\pi)}} \sum_{K_1=1}^{d} g^J_{K_1}(h) \left(\sum_{\nu_1\in\overline{\mathcal{M}}} \frac{\partial h^{K_1}}{\partial\theta_{\nu_1}} \right)$$

with the set $SLTS_2^{(\pi)} = \{\, (\tau)_\pi \,\}$, $m(\mathbf{u}) = 1$ and $\delta_1 = 1$. Assuming now that the hypothesis (28) holds for q, we prove it for $q + 1$. Therefore we write shortly

$$(h^K)^{(\delta)} = \sum_{\nu_1,\ldots,\nu_\delta\in\overline{\mathcal{M}}} \frac{\partial^\delta h^K}{\partial\theta_{\nu_1}\ldots\partial\theta_{\nu_\delta}} \tag{29}$$

and we thus get

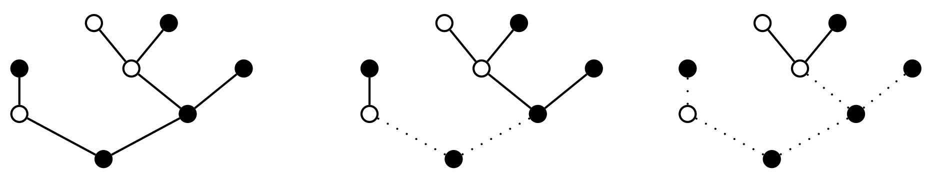
$$
\begin{aligned}
&\sum_{\nu_1,\dots,\nu_{q+1}\in\overline{\mathcal{M}}} \frac{\partial^{q+1} g(h)^J}{\partial\theta_{\nu_1}\dots\partial\theta_{\nu_{q+1}}} = \sum_{\nu_{q+1}\in\overline{\mathcal{M}}} \frac{\partial}{\partial\theta_{\nu_{q+1}}} \Big(\sum_{\nu_1,\dots,\nu_q\in\overline{\mathcal{M}}} \frac{\partial^q g(h)^J}{\partial\theta_{\nu_1}\dots\partial\theta_{\nu_q}} \Big) \\
&= \sum_{\nu_{q+1}\in\overline{\mathcal{M}}} \frac{\partial}{\partial\theta_{\nu_{q+1}}} \Big(\sum_{\mathbf{u}\in SLTS_{q+1}^{(\pi)}} \sum_{K_1,\dots,K_m=1}^{d} g^J_{K_1\dots K_m}(h)\cdot(h^{K_1})^{(\delta_1)}\cdot\ldots\cdot(h^{K_m})^{(\delta_m)} \Big) \\
&= \sum_{\mathbf{u}\in SLTS_{q+1}^{(\pi)}} \sum_{K_1,\dots,K_m,K=1}^{d} g^J_{K_1\dots K_m K}(h)\cdot(h^{K})^{(1)}\cdot(h^{K_1})^{(\delta_1)}\cdot\ldots\cdot(h^{K_m})^{(\delta_m)} \\
&+ \sum_{\mathbf{u}\in SLTS_{q+1}^{(\pi)}} \sum_{K_1,\dots,K_m=1}^{d} g^J_{K_1\dots K_m}(h)\cdot(h^{K_1})^{(\delta_1+1)}\cdot(h^{K_2})^{(\delta_2)}\cdot\ldots\cdot(h^{K_m})^{(\delta_m)} \\
&+ \dots \\
&+ \sum_{\mathbf{u}\in SLTS_{q+1}^{(\pi)}} \sum_{K_1,\dots,K_m=1}^{d} g^J_{K_1\dots K_m}(h)\cdot(h^{K_1})^{(\delta_1)}\cdot\ldots\cdot(h^{K_{m-1}})^{(\delta_{m-1})}\cdot(h^{K_m})^{(\delta_m+1)} \\
&= \sum_{\mathbf{u}\in SLTS_{q+2}^{(\pi)}} \sum_{K_1,\dots,K_m=1}^{d} g^J_{K_1\dots K_m}(h)\cdot(h^{K_1})^{(\delta_1)}\cdot\ldots\cdot(h^{K_m})^{(\delta_m)}.
\end{aligned}
$$

□

As in the deterministic setting, the density $\gamma(\mathbf{t})$ of a tree is a measure of its non-bushiness and can be similarly defined for stochastic colored trees.

Definition 5.3 *For $\boldsymbol{t} = (\boldsymbol{t}', \boldsymbol{t}'') \in LTS$ let $\gamma(\boldsymbol{t})$ be defined recursively by*

$$
\begin{aligned}
\gamma(\boldsymbol{t}) &= 1 && \text{if } l(\boldsymbol{t}) = 1, \\
\gamma(\boldsymbol{t}) &= \prod_{i=1}^{m} \gamma(\boldsymbol{t}_i) && \text{if } \boldsymbol{t} = (\boldsymbol{t}_1, \dots, \boldsymbol{t}_m), \\
\gamma(\boldsymbol{t}) &= l(\boldsymbol{t}) \prod_{i=1}^{m} \gamma(\boldsymbol{t}_i) && \text{if } \boldsymbol{t} = [\boldsymbol{t}_1, \dots, \boldsymbol{t}_m] \text{ or } \boldsymbol{t} = \{\boldsymbol{t}_1, \dots, \boldsymbol{t}_m\}_j.
\end{aligned}
$$



$\gamma(\mathbf{t}) = 8 \qquad \cdot 2 \cdot 5 \qquad \cdot 3 = 240$

Fig. 6. Example for the definition of $\gamma(\mathbf{t})$ for a tree $\mathbf{t} \in LTS$.

In order to have a more suitable notation for the proof of the main theorem of this section, i.e. the theorem about the expansion of the approximation calculated with the stochastic Runge-Kutta method by rooted trees, we introduce the following denomination:

Definition 5.4 *Let $\boldsymbol{t} = (\boldsymbol{t}', \boldsymbol{t}'') \in LTS$ be a tree with $l = l(\boldsymbol{t}) > 1$ nodes which are denoted by $i_1 < i_2 < \ldots < i_l$, consisting of $s = s(\boldsymbol{t}) \leq l$ stochastic nodes $\sigma_{j_1}, \sigma_{j_2}, \ldots, \sigma_{j_s}$. Then we denote for $i \in \{i_2, \ldots, i_l\}$ by*

$$Z_{\boldsymbol{t}'(i),i} = \begin{cases} z_i^{(0,0)} & \text{if } \boldsymbol{t}''(i) = \tau \text{ and } \boldsymbol{t}''(\boldsymbol{t}'(i)) = \gamma \\ \sum_{j_k=1}^m \sum_{\nu_k \in \mathcal{M}} z_i^{(j_k,\nu_k)} & \text{if } \boldsymbol{t}''(i) = \sigma_{j_k} \text{ and } \boldsymbol{t}''(\boldsymbol{t}'(i)) = \gamma \\ Z_{\boldsymbol{t}'(i),i}^{(0,0),(0,0)} & \text{if } \boldsymbol{t}''(i) = \tau \text{ and } \boldsymbol{t}''(\boldsymbol{t}'(i)) = \tau \\ Z_{\boldsymbol{t}'(i),i}^{(j_k,\nu_k),(0,0)} & \text{if } \boldsymbol{t}''(i) = \tau \text{ and } \boldsymbol{t}''(\boldsymbol{t}'(i)) = \sigma_{j_k} \\ \sum_{j_k=1}^m \sum_{\nu_k \in \mathcal{M}} Z_{\boldsymbol{t}'(i),i}^{(0,0),(j_k,\nu_k)} & \text{if } \boldsymbol{t}''(i) = \sigma_{j_k} \text{ and } \boldsymbol{t}''(\boldsymbol{t}'(i)) = \tau \\ \sum_{j_r=1}^m \sum_{\nu_r \in \mathcal{M}} Z_{\boldsymbol{t}'(i),i}^{(j_k,\nu_k),(j_r,\nu_r)} & \text{if } \boldsymbol{t}''(i) = \sigma_{j_r} \text{ and } \boldsymbol{t}''(\boldsymbol{t}'(i)) = \sigma_{j_k} \end{cases} \tag{30}$$

Further, we denote by

$$\Phi_{i_1}(\boldsymbol{t}) = \sum_{i_2,\ldots,i_l=1}^{s} Z_{\boldsymbol{t}'(i_2),i_2} \cdot \ldots \cdot Z_{\boldsymbol{t}'(i_l),i_l} \tag{31}$$

the corresponding coefficient function and define $\Phi_{i_1}(\boldsymbol{t}) = 1$ if $l(\boldsymbol{t}) = 1$.

We will now state a proposition which allows a representation of the derivatives of the stochastic Runge-Kutta method w.r.t. rooted trees.

Proposition 5.5 *Let $q \in \mathbb{N}$, $k \in \{0, 1, \ldots, m\}$, $\nu \in \overline{\mathcal{M}}$ and $A = \{\tau, \sigma_{j_r} : r \in \mathbb{N}\}$. We denote by*

$$\begin{aligned} z_{i_1} &= \begin{cases} z_{i_1}^{(0,0)} & \text{if } \boldsymbol{t}''(i_1) = \tau \\ \sum_{j_r=1}^m \sum_{\nu_r \in \mathcal{M}} z_{i_1}^{(j_r,\nu_r)} & \text{if } \boldsymbol{t}''(i_1) = \sigma_{j_r} \end{cases} \\ Z_{i,i_1}^{(k,\nu)} &= \begin{cases} Z_{i,i_1}^{(k,\nu),(0,0)} & \text{if } \boldsymbol{t}''(i_1) = \tau \\ \sum_{j_r=1}^m \sum_{\mu_r \in \mathcal{M}} Z_{i,i_1}^{(k,\nu),(j_r,\nu_r)} & \text{if } \boldsymbol{t}''(i_1) = \sigma_{j_r} \end{cases} \end{aligned} \tag{32}$$

Then the derivatives of the Jth component of $H_i^{(k,\nu)}(t_0)$ satisfy

$$\mathcal{D}^q H_i^{(k,\nu)}(t_0)^J = \sum_{\substack{\boldsymbol{t} \in LTS \\ l(\boldsymbol{t})=q}} \gamma(\boldsymbol{t}) \sum_{i_1=1}^{s} Z_{i,i_1}^{(k,\nu)} \cdot \Phi_{i_1}(\boldsymbol{t}) \cdot F(\boldsymbol{t})(Y(t_0))^J. \tag{33}$$

The Jth component of the numerical solution $Y(t_0)$ satisfies

$$\mathcal{D}^q Y(t_0)^J = \sum_{\substack{\boldsymbol{t} \in LTS \\ l(\boldsymbol{t})=q}} \gamma(\boldsymbol{t}) \sum_{i_1=1}^{s} z_{i_1} \cdot \Phi_{i_1}(\boldsymbol{t}) \cdot F(\boldsymbol{t})(Y(t_0))^J. \tag{34}$$

Proof. Because of the similarity of $Y(t)$ and $H_i^{(k,\nu)}(t)$, it is satisfactory to prove the first equation (33) only. Then the second equation (34) follows substituting $Z_{ij}^{(k,\nu)(0,0)}$ and $Z_{ij}^{(k,\nu)(r,\mu)}$ by $z_j^{(0,0)}$ and $z_j^{(r,\mu)}$, respectively, in $H_i^{(k,\nu)}(t)$, which equals it to $Y(t)$, and by the definition of $\mathcal{D}^q$ in (23).

We prove equation (33) by induction on q. For $q=1$ and $A=\{\tau,\sigma_{j_k}:k\in\mathbb{N}\}$ there are two trees $\mathbf{t}_1=\tau$ and $\mathbf{t}_2=\sigma_{j_1}$ with $l(\mathbf{t}_1)=l(\mathbf{t}_2)=1$ in LTS and

$$\begin{aligned}\sum_{\nu_1\in\overline{\mathcal{M}}}\Delta\theta_{\nu_1}\cdot\frac{\partial H_i^{(k,\nu)}(t_0)^J}{\partial\theta_{\nu_1}} &= \sum_{i_1=1}^{s} Z_{i,i_1}^{(k,\nu),(0,0)}\cdot a(Y(t_0))^J \\ &\quad+\sum_{i_1=1}^{s}\sum_{j_1=1}^{m}\sum_{\nu_1\in\mathcal{M}} Z_{i,i_1}^{(k,\nu),(j_1,\nu_1)}\cdot b^{j_1}(Y(t_0))^J \\ &= \sum_{\substack{\mathbf{t}\in LTS\\ l(\mathbf{t})=1}}\gamma(\mathbf{t})\sum_{i_1=1}^{s} Z_{i,i_1}^{(k,\nu)}\cdot F(\mathbf{t})(Y(t_0))^J.\end{aligned} \tag{35}$$

For a better understanding, we also consider the case $q=2$. Here we have to consider the trees $\mathbf{t}_3=[\tau]$, $\mathbf{t}_4=[\sigma_{j_1}]$, $\mathbf{t}_5=\{\tau\}_{j_1}$ and $\mathbf{t}_6=\{\sigma_{j_2}\}_{j_1}$ with $l(\mathbf{t})=2$ nodes in LTS. Then we get

$$\begin{aligned}&\sum_{\nu_1,\nu_2\in\overline{\mathcal{M}}}\Delta\theta_{\nu_1}\cdot\Delta\theta_{\nu_2}\cdot\frac{\partial H_i^{(k,\nu)}(t_0)^J}{\partial\theta_{\nu_1}\partial\theta_{\nu_2}}\\ &= 2\sum_{i_1,i_2=1}^{s} Z_{i,i_1}^{(k,\nu),(0,0)}\,Z_{i_1,i_2}^{(0,0),(0,0)}\sum_{K_1=1}^{d}\frac{\partial a(Y(t_0))^J}{\partial x^{K_1}}\,a(Y(t_0))^{K_1}\\ &+2\sum_{i_1,i_2=1}^{s}\sum_{j_1=1}^{m}\sum_{\nu_1\in\mathcal{M}} Z_{i,i_1}^{(k,\nu),(0,0)}\,Z_{i_1,i_2}^{(0,0),(j_1,\nu_1)}\sum_{K_1=1}^{d}\frac{\partial a(Y(t_0))^J}{\partial x^{K_1}}\,b^{j_1}(Y(t_0))^{K_1}\\ &+2\sum_{i_1,i_2=1}^{s}\sum_{j_1=1}^{m}\sum_{\nu_1\in\mathcal{M}} Z_{i,i_1}^{(k,\nu),(j_1,\nu_1)}\,Z_{i_1,i_2}^{(j_1,\nu_1),(0,0)}\sum_{K_1=1}^{d}\frac{\partial b^{j_1}(Y(t_0))^J}{\partial x^{K_1}}\,a(Y(t_0))^{K_1}\\ &+2\sum_{i_1,i_2=1}^{s}\sum_{j_1,j_2=1}^{m}\sum_{\nu_1,\nu_2\in\mathcal{M}} Z_{i,i_1}^{(k,\nu),(j_1,\nu_1)}\,Z_{i_1,i_2}^{(j_1,\nu_1),(j_2,\nu_2)}\sum_{K_1=1}^{d}\frac{\partial b^{j_1}(Y(t_0))^J}{\partial x^{K_1}}b^{j_2}(Y(t_0))^{K_1}\\ &=\sum_{\substack{\mathbf{t}\in LTS\\ l(\mathbf{t})=2}}\gamma(\mathbf{t})\sum_{i_1=1}^{s} Z_{i,i_1}^{(k,\nu)}\cdot\Phi_{i_1}(\mathbf{t})\cdot F(\mathbf{t})(Y(t_0))^J.\end{aligned} \tag{36}$$

Now, we assume that equation (33) holds for some $q-1$ and prove the case q. The first step is the application of formula (26) in order to obtain

$$\begin{aligned}
\mathcal{D}^q H_i^{(k,\nu)}(t_0)^J &= \sum_{\nu_1,\ldots,\nu_q \in \overline{\mathcal{M}}} \Delta\theta_{\nu_1} \cdot \ldots \cdot \Delta\theta_{\nu_q} \cdot \frac{\partial^q H_i^{(k,\nu)}(t_0)^J}{\partial\theta_{\nu_1} \ldots \partial\theta_{\nu_q}} \\
&= q \sum_{i_1=1}^{s} Z_{i,i_1}^{(k,\nu),(0,0)} \sum_{\nu_1,\ldots,\nu_{q-1} \in \overline{\mathcal{M}}} \Delta\theta_{\nu_1} \cdot \ldots \cdot \Delta\theta_{\nu_{q-1}} \frac{\partial^{q-1} a(H_{i_1}^{(0,0)}(t_0))^J}{\partial\theta_{\nu_1} \ldots \theta_{\nu_{q-1}}} \\
&+ q \sum_{i_1=1}^{s} \sum_{j_1=1}^{m} \sum_{\nu_1 \in \mathcal{M}} Z_{i,i_1}^{(k,\nu)(j_1,\nu_1)} \times \\
&\times \sum_{\nu_1,\ldots,\nu_{q-1} \in \overline{\mathcal{M}}} \Delta\theta_{\nu_1} \cdot \ldots \cdot \Delta\theta_{\nu_{q-1}} \frac{\partial^{q-1} b^{j_1}(H_{i_1}^{(j_1,\nu_1)}(t_0))^J}{\partial\theta_{\nu_1} \ldots \partial\theta_{\nu_{q-1}}}.
\end{aligned} \tag{37}$$

As the second step, we make use of Lemma 5.2 twice. Firstly, equation (28) is applied to trees $\mathbf{u} \in SLTS_q^{(\tau)}$ (i.e., trees having a root of type τ) and secondly, to trees $\mathbf{u} \in SLTS_q^{(\sigma_{j_1})}$ (i.e., trees having a root of type σ_{j_1}). Thus with $\delta_1 + \ldots + \delta_{m(\mathbf{u})} = q-1$ we obtain

$$\begin{aligned}
&\sum_{\nu_1,\ldots,\nu_{q-1} \in \overline{\mathcal{M}}} \Delta\theta_{\nu_1} \cdot \ldots \cdot \Delta\theta_{\nu_{q-1}} \frac{\partial^{q-1} a(H_{i_1}^{(0,0)}(t_0))^J}{\partial\theta_{\nu_1} \ldots \partial\theta_{\nu_{q-1}}} \\
&\quad = \sum_{\mathbf{u} \in SLTS_q^{(\tau)}} \sum_{K_1,\ldots,K_m=1}^{d} a_{K_1 \ldots K_m}^J (H_{i_1}^{(0,0)}(t_0)) \times \\
&\quad \times \Bigg(\Bigg(\sum_{\nu_1,\ldots,\nu_{\delta_1} \in \overline{\mathcal{M}}} \Delta\theta_{\nu_1} \cdot \ldots \cdot \Delta\theta_{\nu_{\delta_1}} \frac{\partial^{\delta_1} H_{i_1}^{(0,0)}(t_0)^{K_1}}{\partial\theta_{\nu_1} \ldots \partial\theta_{\nu_{\delta_1}}} \Bigg) \cdot \ldots \\
&\quad \ldots \cdot \Bigg(\sum_{\nu_1,\ldots,\nu_{\delta_m} \in \overline{\mathcal{M}}} \Delta\theta_{\nu_1} \cdot \ldots \cdot \Delta\theta_{\nu_{\delta_m}} \frac{\partial^{\delta_m} H_{i_1}^{(0,0)}(t_0)^{K_m}}{\partial\theta_{\nu_1} \ldots \partial\theta_{\nu_{\delta_m}}} \Bigg) \Bigg)
\end{aligned} \tag{38}$$

and analogously

$$\sum_{\nu_1,\dots,\nu_{q-1}\in\overline{\mathcal{M}}} \Delta\theta_{\nu_1}\cdot\ldots\cdot\Delta\theta_{\nu_{q-1}} \frac{\partial^{q-1} b^{j_1}(H_{i_1}^{(j_1,\nu_1)}(t_0))^J}{\partial\theta_{\nu_1}\dots\partial\theta_{\nu_{q-1}}}$$
$$= \sum_{\mathbf{u}\in SLTS_q^{(\sigma_{j_1})}} \sum_{K_1,\dots,K_m=1}^{d} {b^{j_1}}^J_{K_1\dots K_m}(H_{i_1}^{(j_1,\nu_1)}(t_0)) \times$$
$$\times \left(\left(\sum_{\nu_1,\dots,\nu_{\delta_1}\in\overline{\mathcal{M}}} \Delta\theta_{\nu_1}\cdot\ldots\cdot\Delta\theta_{\nu_{\delta_1}} \frac{\partial^{\delta_1} H_{i_1}^{(j_1,\nu_1)}(t_0)^{K_1}}{\partial\theta_{\nu_1}\dots\partial\theta_{\nu_{\delta_1}}}\right)\cdot\ldots\right.$$
$$\left.\ldots\cdot\left(\sum_{\nu_1,\dots,\nu_{\delta_m}\in\overline{\mathcal{M}}} \Delta\theta_{\nu_1}\cdot\ldots\cdot\Delta\theta_{\nu_{\delta_m}} \frac{\partial^{\delta_m} H_{i_1}^{(j_1,\nu_1)}(t_0)^{K_m}}{\partial\theta_{\nu_1}\dots\partial\theta_{\nu_{\delta_m}}}\right)\right). \tag{39}$$

Finally, we replace the derivatives of $H_{i_1}^{(0,0)}$ and $H_{i_1}^{(j_1,\nu_1)}$, which appear in (38) and (39) with $\delta_i \le q-1$, $1\le i\le m=m(\mathbf{u})$, by the induction hypothesis (33) and rearrange the sums. Then we get for (37):

$$\sum_{\nu_1,\dots,\nu_q\in\overline{\mathcal{M}}} \Delta\theta_{\nu_1}\cdot\ldots\cdot\Delta\theta_{\nu_q} \frac{\partial^q H_i^{(k,\nu)}(t_0)^J}{\partial\theta_{\nu_1}\dots\partial\theta_{\nu_q}}$$
$$= q \sum_{\mathbf{u}\in SLTS_q^{(\tau)}} \sum_{\substack{\mathbf{t}_1\in LTS\\ l(\mathbf{t}_1)=\delta_1}} \cdots \sum_{\substack{\mathbf{t}_m\in LTS\\ l(\mathbf{t}_m)=\delta_m}} \gamma(\mathbf{t}_1)\cdot\ldots\cdot\gamma(\mathbf{t}_m)\times$$
$$\times \sum_{i_1=1}^{s} Z_{i,i_1}^{(k,\nu)(0,0)} \left(\sum_{k_1=1}^{s} Z_{i_1,k_1}^{(0,0)}\Phi_{k_1}(\mathbf{t}_1)\cdot\ldots\cdot\sum_{k_m=1}^{s} Z_{i_1,k_m}^{(0,0)}\Phi_{k_m}(\mathbf{t}_m)\right)\times$$
$$\times \sum_{K_1,\dots,K_m=1}^{d} a^J_{K_1\dots K_m}(H_{i_1}^{(0,0)}(t_0))\cdot\Big(F(\mathbf{t}_1)(Y(t_0))^{K_1}\cdot\ldots\cdot F(\mathbf{t}_m)(Y(t_0))^{K_m}\Big)$$
$$+ q \sum_{\mathbf{u}\in SLTS_q^{(\sigma_{j_1})}} \sum_{\substack{\mathbf{t}_1\in LTS\\ l(\mathbf{t}_1)=\delta_1}} \cdots \sum_{\substack{\mathbf{t}_m\in LTS\\ l(\mathbf{t}_m)=\delta_m}} \gamma(\mathbf{t}_1)\cdot\ldots\cdot\gamma(\mathbf{t}_m)\times$$
$$\times \sum_{j_1=1}^{m}\sum_{\nu_1\in\mathcal{M}}\sum_{i_1=1}^{s} Z_{i,i_1}^{(k,\nu)(j_1,\nu_1)} \left(\sum_{k_1=1}^{s} Z_{i_1,k_1}^{(j_1,\nu_1)}\Phi_{k_1}(\mathbf{t}_1)\cdot\ldots\cdot\sum_{k_m=1}^{s} Z_{i_1,k_m}^{(j_1,\nu_1)}\Phi_{k_m}(\mathbf{t}_m)\right)\times$$
$$\times \sum_{K_1,\dots,K_m=1}^{d} {b^{j_1}}^J_{K_1\dots K_m}(H_{i_1}^{(j_1,\nu_1)}(t_0))\cdot\Big(F(\mathbf{t}_1)(Y(t_0))^{K_1}\cdot\ldots\cdot F(\mathbf{t}_m)(Y(t_0))^{K_m}\Big) \tag{40}$$

where i_1 denotes the root of $\mathbf{u}$ and $k_1,\dots,k_m$ denote the roots of the trees $\mathbf{t}_1,\dots,\mathbf{t}_m$, respectively.

The main difficulty is now to understand that to each tuple of trees

$$(\mathbf{u},\mathbf{t}_1,\dots,\mathbf{t}_m) \quad \text{with} \quad \mathbf{u}\in SLTS_q^{(\pi)}, \quad \mathbf{t}_i\in LTS, \quad l(\mathbf{t}_i)=\delta_i$$

with $\pi \in A$ and $\sum_{i=1}^m \delta_i = q-1$, there corresponds exactly one labelled tree $\mathbf{t} = (\mathbf{t}', \mathbf{t}'') \in LTS$ with $l(\mathbf{t}) = q$ such that the root i_1 of $\mathbf{t}$ is of type π and such that

$$\gamma(\mathbf{t}) = q \cdot \gamma(\mathbf{t}_1) \cdot \ldots \cdot \gamma(\mathbf{t}_m) \tag{41}$$

and for $\pi = \tau$

$$\begin{aligned} F(\mathbf{t})(Y(t_0))^J &= \sum_{K_1,\ldots,K_m=1}^{d} a^J_{K_1\ldots K_m}(Y(t_0))\, F(\mathbf{t}_1)(Y(t_0))^{K_1} \ldots F(\mathbf{t}_m)(Y(t_0))^{K_m} \\ \Phi_{i_1}(\mathbf{t}) &= \sum_{k_1,\ldots,k_m=1}^{s} Z^{(0,0)}_{i_1,k_1} \ldots Z^{(0,0)}_{i_1,k_m} \Phi_{k_1}(\mathbf{t}_1) \ldots \Phi_{k_m}(\mathbf{t}_m) \end{aligned} \tag{42}$$

or for $\pi = \sigma_{j_1}$

$$\begin{aligned} F(\mathbf{t})(Y(t_0))^J &= \sum_{K_1,\ldots,K_m=1}^{d} {b^{j_1}}^J_{K_1\ldots K_m}(Y(t_0))\, F(\mathbf{t}_1)(Y(t_0))^{K_1} \ldots F(\mathbf{t}_m)(Y(t_0))^{K_m} \\ \Phi_{i_1}(\mathbf{t}) &= \sum_{k_1,\ldots,k_m=1}^{s} Z^{(j_1,\nu_1)}_{i_1,k_1} \ldots Z^{(j_1,\nu_1)}_{i_1,k_m} \Phi_{k_1}(\mathbf{t}_1) \ldots \Phi_{k_m}(\mathbf{t}_m) \end{aligned} \tag{43}$$

holds, respectively. This labelled tree $\mathbf{t}$ is obtained if the branches of $\mathbf{u}$ are replaced by the trees $\mathbf{t}_1, \ldots, \mathbf{t}_m$ and the corresponding labels are taken over in a natural way, i.e., in the same order (see Figure 7).

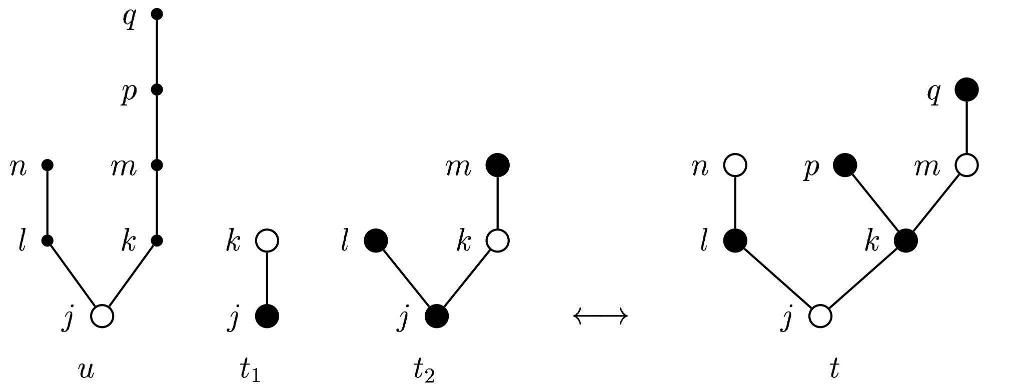


Fig. 7. Example for the bijection of $(\mathbf{u}, \mathbf{t}_1, \ldots, \mathbf{t}_m) \leftrightarrow \mathbf{t}$ with $\pi = \sigma$.

In this way, for $\pi = \tau$ and $\pi = \sigma_{j_1}$ *all* trees $\mathbf{t} = (\mathbf{t}', \mathbf{t}'') \in LTS$ with $l(\mathbf{t}) = q$ appear exactly *once*. Thus (40) becomes (33) after inserting (41), (42) and (43), respectively. □

Since the Taylor expansion contains the coefficients of the SRK method, we define a coefficient function Φ_S which assigns to every tree an *elementary weight*. So for every $\mathbf{t} \in TS$ or $\mathbf{t} \in LTS$ the function Φ_S is defined recursively by

$$\Phi_S(\mathbf{t}) = \begin{cases} \prod_{i=1}^{\lambda} \Phi_S(\mathbf{t}_i) & \text{if } \mathbf{t} = (\mathbf{t}_1, \dots, \mathbf{t}_\lambda) \\ {z^{(0,0)}}^T \prod_{i=1}^{\lambda} \Psi^{(0,0)}(\mathbf{t}_i) & \text{if } \mathbf{t} = [\mathbf{t}_1, \dots, \mathbf{t}_\lambda] \\ \sum_{\nu \in \mathcal{M}} {z^{(k,\nu)}}^T \prod_{i=1}^{\lambda} \Psi^{(k,\nu)}(\mathbf{t}_i) & \text{if } \mathbf{t} = \{\mathbf{t}_1, \dots, \mathbf{t}_\lambda\}_k \end{cases} \tag{44}$$

where $\Psi^{(0,0)}(\emptyset) = \Psi^{(k,\nu)}(\emptyset) = e$ with $\gamma = (\emptyset)$, $\tau = [\emptyset]$, $\sigma_k = \{\emptyset\}_k$ and

$$\Psi^{(0,0)}(\mathbf{t}) = \begin{cases} Z^{(0,0),(0,0)} \prod_{i=1}^{\lambda} \Psi^{(0,0)}(\mathbf{t}_i) & \text{if } \mathbf{t} = [\mathbf{t}_1, \dots, \mathbf{t}_\lambda] \\ \sum_{\mu \in \mathcal{M}} Z^{(0,0),(r,\mu)} \prod_{i=1}^{\lambda} \Psi^{(r,\mu)}(\mathbf{t}_i) & \text{if } \mathbf{t} = \{\mathbf{t}_1, \dots, \mathbf{t}_\lambda\}_r \end{cases} \tag{45}$$

$$\Psi^{(k,\nu)}(\mathbf{t}) = \begin{cases} Z^{(k,\nu),(0,0)} \prod_{i=1}^{\lambda} \Psi^{(0,0)}(\mathbf{t}_i) & \text{if } \mathbf{t} = [\mathbf{t}_1, \dots, \mathbf{t}_\lambda] \\ \sum_{\mu \in \mathcal{M}} Z^{(k,\nu),(r,\mu)} \prod_{i=1}^{\lambda} \Psi^{(r,\mu)}(\mathbf{t}_i) & \text{if } \mathbf{t} = \{\mathbf{t}_1, \dots, \mathbf{t}_\lambda\}_r \end{cases} \tag{46}$$

Here $e = (1, \dots, 1)^T$ and the product of vectors in the definition of $\Psi^{(0,0)}$ and $\Psi^{(k,\nu)}$ is defined by component-wise multiplication, i.e. with $(a_1, \dots, a_n) * (b_1, \dots, b_n) = (a_1 b_1, \dots, a_n b_n)$. Now we get immediately the following representation of the stochastic Runge-Kutta approximation w.r.t. rooted trees.

Corollary 5.6 *Assume that the drift a and the diffusion b^j, $1 \leq j \leq m$, are sufficiently differentiable. Then, the one-step approximation $Y(t) = Y(t_0 + h)$ with $h \in]0, \infty[$, given by the stochastic Runge-Kutta method (16), can be represented as*

$$\begin{aligned} Y(t)^J &= Y(t_0)^J + \sum_{\substack{\boldsymbol{t} \in LTS \\ l(\boldsymbol{t}) \leq n}} \frac{\gamma(\boldsymbol{t}) \sum_{i_1=1}^{s} z_{i_1} \Phi_{i_1}(\boldsymbol{t}) F(\boldsymbol{t})(Y(t_0))^J}{l(\boldsymbol{t})!} + \mathcal{R}_n(t, t_0) \\ &= Y(t_0)^J + \sum_{\substack{\boldsymbol{t} \in TS \\ l(\boldsymbol{t}) \leq n}} \frac{\alpha(\boldsymbol{t}) \gamma(\boldsymbol{t}) \sum_{i_1=1}^{s} z_{i_1} \Phi_{i_1}(\boldsymbol{t}) F(\boldsymbol{t})(Y(t_0))^J}{l(\boldsymbol{t})!} + \mathcal{R}_n(t, t_0) \end{aligned} \tag{47}$$

for $n \in \mathbb{N}$ and with $\alpha(\boldsymbol{t})$ denoting the cardinality of the tree $\boldsymbol{t} \in LTS$ with $A = \{\tau, \sigma\}$. Using the coefficient function Φ_S, we get analogously

$$\begin{aligned} Y(t)^J &= Y(t_0)^J + \sum_{\substack{\boldsymbol{t} \in LTS \\ l(\boldsymbol{t}) \leq n}} \sum_{j_1, \dots, j_{s(\boldsymbol{t})}=1}^{m} \frac{\gamma(\boldsymbol{t}) \Phi_S(\boldsymbol{t}) F(\boldsymbol{t})(Y(t_0))^J}{l(\boldsymbol{t})!} + \mathcal{R}_n(t, t_0) \\ &= Y(t_0)^J + \sum_{\substack{\boldsymbol{t} \in TS \\ l(\boldsymbol{t}) \leq n}} \sum_{j_1, \dots, j_{s(\boldsymbol{t})}=1}^{m} \frac{\alpha(\boldsymbol{t}) \gamma(\boldsymbol{t}) \Phi_S(\boldsymbol{t}) F(\boldsymbol{t})(Y(t_0))^J}{l(\boldsymbol{t})!} + \mathcal{R}_n(t, t_0). \end{aligned} \tag{48}$$

Proof. This follows directly from the Theorem of Taylor (see (24)) and Proposition 5.5. □

As a final step, we extend this representation of the approximation $Y(t)$ to our primarily problem of a representation for $f(Y(t))$. Therefore we consider a suitable subset $LTS(\Delta)$ of LTS w.r.t. the set $A = \{\gamma, \tau, \sigma_{j_k} : k \in \mathbb{N}\}$, where γ represents the function f.

Definition 5.7 *Let $LTS(\Delta)$ denote the set of trees $\boldsymbol{t} = (\boldsymbol{t}', \boldsymbol{t}'') \in LTS$ w.r.t. $A = \{\gamma, \tau, \sigma_{j_k} : k \in \mathbb{N}\}$ such that*

a) the root is of type $\boldsymbol{t}''(1) = \gamma$ and all other nodes are either deterministic or stochastic nodes, i.e. $\boldsymbol{t}''(i) \in \{\tau, \sigma_{j_k} : k \in \mathbb{N}\}$ for $2 \leq i \leq l(\boldsymbol{t})$,
b) all stochastic nodes own a different variable index j_k, $1 \leq k \leq s(\boldsymbol{t})$, i.e. for two different stochastic nodes $i \neq l$ holds $\boldsymbol{t}''(i) \neq \boldsymbol{t}''(l)$.

Further $TS(\Delta) = LTS(\Delta)/\sim$ denotes the equivalence class under the relation of Definition 3.4 restricted to $LTS(\Delta)$ and $\alpha_\Delta(\boldsymbol{t})$ denotes the cardinality of $\boldsymbol{t}$ in LTS.

Here it has to be pointed out that $LTS(I) \subset LTS(S) \subset LTS(\Delta)$ since the rules of construction for the trees $\mathbf{t}$ in $LTS(I)$ and in $LTS(S)$ are more restrictive than for the trees $\mathbf{t} \in LTS(\Delta)$. However in contrast to $LTS(I)$ and $LTS(S)$, a tree $\mathbf{t} \in LTS(\Delta)$ has $s(\mathbf{t})$ different variable indices $j_1, \ldots, j_{s(t)}$ while a tree $\mathbf{u}$ in $LTS(I)$ or $LTS(S)$ has only $n(\mathbf{u}) = s(\mathbf{u})/2$ different variable indices. For example, the tree $(\{[\sigma_{j_2}]\}_{j_1})$ is an element of $LTS(\Delta)$ while it is neither an element of $LTS(I)$ nor of $LTS(S)$. With the definition of the set $LTS(\Delta)$, we can now formulate our main result for the expansion of the stochastic Runge-Kutta method. It gives an expansion of $f(Y(t))$ which is required for the calculation of order conditions for the SRK method.

Theorem 5.8 *For the one-step approximation $Y(t) = Y(t_0 + h)$, $h \in]0, \infty[$, defined by the stochastic Runge-Kutta method (16), a function $f : \mathbb{R}^d \to \mathbb{R}$ and for $n \in \mathbb{N}$ the expansion*

$$\begin{aligned} f(Y(t)) &= \sum_{\substack{\boldsymbol{t} \in LTS(\Delta) \\ l(\boldsymbol{t})-1 \leq n}} \sum_{j_1, \ldots, j_{s(\boldsymbol{t})}=1}^{m} \frac{\gamma(\boldsymbol{t}) \cdot \Phi_S(\boldsymbol{t}) \cdot F(\boldsymbol{t})(Y(t_0))}{(l(\boldsymbol{t})-1)!} + \mathcal{R}_n(t, t_0) \\ &= \sum_{\substack{\boldsymbol{t} \in TS(\Delta) \\ l(\boldsymbol{t})-1 \leq n}} \sum_{j_1, \ldots, j_{s(\boldsymbol{t})}=1}^{m} \frac{\alpha_\Delta(\boldsymbol{t}) \cdot \gamma(\boldsymbol{t}) \cdot \Phi_S(\boldsymbol{t}) \cdot F(\boldsymbol{t})(Y(t_0))}{(l(\boldsymbol{t})-1)!} + \mathcal{R}_n(t, t_0) \end{aligned} \tag{49}$$

holds provided all necessary derivatives of f, a and b^j, $1 \leq j \leq m$, exist.

Proof. Let $A = \{\gamma, \tau, \sigma_{j_k} : k \in \mathbb{N}\}$. We apply Lemma 5.2 with $\pi = \gamma$ and conclude that

$$\begin{aligned} \mathcal{D}^q f(Y(t_0)) &= \sum_{\nu_1,\ldots,\nu_q \in \overline{\mathcal{M}}} \Delta\theta_{\nu_1} \cdot \ldots \cdot \Delta\theta_{\nu_q} \cdot \frac{\partial^q f(Y(t_0))}{\partial\theta_{\nu_1} \ldots \partial\theta_{\nu_q}} \\ &= \sum_{\mathbf{u} \in SLTS^{(\gamma)}_{q+1}} \sum_{K_1,\ldots,K_m=1}^{d} f_{K_1 \ldots K_m}(Y(t_0)) \cdot \left(\mathcal{D}^{\delta_1} Y(t_0)^{K_1} \ldots \mathcal{D}^{\delta_m} Y(t_0)^{K_m}\right) \end{aligned} \tag{50}$$

where $m = m(\mathbf{u})$ and $\delta_1 + \ldots + \delta_m = q$. Now Proposition 5.5 yields

$$\begin{aligned} \mathcal{D}^q f(Y(t_0)) = & \sum_{\mathbf{u} \in SLTS^{(\gamma)}_{q+1}} \sum_{K_1,\ldots,K_m=1}^{d} f_{K_1 \ldots K_m}(Y(t_0)) \times \\ & \times \Bigg(\Big(\sum_{\substack{\mathbf{t}_1 \in LTS \\ l(\mathbf{t}_1)=\delta_1}} \gamma(\mathbf{t}_1) \sum_{k_1=1}^{s} z_{k_1} \Phi_{k_1}(\mathbf{t}_1) \cdot F(\mathbf{t}_1)(Y(t_0))^{K_1}\Big) \cdot \ldots \times \\ & \times \ldots \cdot \Big(\sum_{\substack{\mathbf{t}_m \in LTS \\ l(\mathbf{t}_m)=\delta_m}} \gamma(\mathbf{t}_m) \sum_{k_m=1}^{s} z_{k_m} \Phi_{k_m}(\mathbf{t}_m) \cdot F(\mathbf{t}_m)(Y(t_0))^{K_m}\Big)\Bigg) \end{aligned} \tag{51}$$

where $\mathbf{t}_1, \ldots, \mathbf{t}_m \in LTS$ are considered w.r.t. $A = \{\tau, \sigma_{j_k} : k \in \mathbb{N}\}$ and $k_1, \ldots, k_m$ denote the roots of the trees $\mathbf{t}_1, \ldots, \mathbf{t}_m$, respectively. Now nearly the same considerations as in the proof of Proposition 5.5 apply: To each tuple of trees $(\mathbf{u}, \mathbf{t}_1, \ldots, \mathbf{t}_m)$ with $\mathbf{u} \in SLTS^{(\gamma)}_{q+1}$, $\mathbf{t}_i \in LTS$, $l(\mathbf{t}_i) = \delta_i$ and with $\sum_{i=1}^m \delta_i = q$, there corresponds exactly one labelled tree $\mathbf{t} = (\mathbf{t}', \mathbf{t}'') \in LTS(\Delta)$ with $l(\mathbf{t}) = q + 1$ such that the root i_1 of $\mathbf{t}$ is of type γ and

$$\begin{aligned} \gamma(\mathbf{t}) &= \gamma(\mathbf{t}_1) \cdot \ldots \cdot \gamma(\mathbf{t}_m) \\ F(\mathbf{t})(Y(t_0)) &= \sum_{K_1,\ldots,K_m=1}^{d} f_{K_1 \ldots K_m}(Y(t_0)) \cdot F(\mathbf{t}_1)(Y(t_0))^{K_1} \ldots F(\mathbf{t}_m)(Y(t_0))^{K_m} \\ \tilde{\Phi}(\mathbf{t}) :&= \prod_{k \in \mathbf{t}'^{-1}(i_1)} \sum_{k=1}^{s} z_k \Phi_k(\mathbf{t}_k) = \sum_{k_1,\ldots,k_m=1}^{s} z_{k_1} \Phi_{k_1}(\mathbf{t}_1) \cdot \ldots \cdot z_{k_m} \Phi_{k_m}(\mathbf{t}_m) \end{aligned} \tag{52}$$

where $\mathbf{t}_k$ denotes the subtree of $\mathbf{t}$ having the node k as a root.

The labelled tree $\mathbf{t}$ is obtained if the branches of $\mathbf{u}$ are replaced by the trees $\mathbf{t}_1, \ldots, \mathbf{t}_m$ and the corresponding labels are taken over in a natural way, i.e. in the same order (see Figure 7). In this way *all* trees $\mathbf{t} = (\mathbf{t}', \mathbf{t}'') \in LTS(\Delta)$ with $l(\mathbf{t}) = q+1$ appear exactly *once*. Applying the usual tensor notation and substituting $\tilde{\Phi}(\mathbf{t})$ by $\Phi_S(\mathbf{t})$, we get

$$\begin{aligned}
\mathcal{D}^q f(Y(t_0)) &= \sum_{\substack{\mathbf{t}\in LTS(\Delta)\\ l(\mathbf{t})=q+1}} \gamma(\mathbf{t}) \cdot \tilde{\Phi}(\mathbf{t}) \cdot F(\mathbf{t})(Y(t_0)) \\
&= \sum_{\substack{\mathbf{t}\in LTS(\Delta)\\ l(\mathbf{t})=q+1}} \sum_{j_1,\dots,j_{s(\mathbf{t})}=1}^{m} \gamma(\mathbf{t}) \cdot \Phi_S(\mathbf{t}) \cdot F(\mathbf{t})(Y(t_0))
\end{aligned} \tag{53}$$

With $\Phi_S(\gamma) = 1$, $F(\gamma)(Y(t_0)) = f(Y(t_0))$ and the Theorem of Taylor (24) we finally arrive at (49) which completes the proof. $\square$

6 Order Conditions for SRK Methods

In this section, conditions such that the stochastic Runge-Kutta method (16) converges in the weak sense with order p to the solution of the stochastic differential equation (20) are considered. Therefore, we give a suitable representation of the approximation due to the SRK method.

Proposition 6.1 *Let $Y(t) = Y(t_0+h)$ with $h \in]0, h_0[$ and $Y(t_0) = x_0$ denote the one-step approximation defined by the stochastic Runge-Kutta method (16). Assume that for the random variables holds $\theta_\iota(h) = \sqrt{h} \cdot \vartheta_\iota$ for $\iota \in \mathcal{M}$ with a bounded random variable ϑ_ι. Then for $f : \mathbb{R}^d \to \mathbb{R}$ and $p \in \mathbb{N}$ the expansion*

$$E^{t_0,x_0}\left(f\left(Y(t)\right)\right) = \sum_{\substack{\boldsymbol{t}\in TS(\Delta)\\ \rho(\boldsymbol{t})\leq p+\frac{1}{2}}} \sum_{j_1,\dots,j_{s(\boldsymbol{t})}=1}^{m} \frac{\alpha_\Delta(\boldsymbol{t})\,\gamma(\boldsymbol{t})\,F(\boldsymbol{t})(x_0)\,E\left(\Phi_S(\boldsymbol{t})\right)}{(l(\boldsymbol{t})-1)!} + O\left(h^{p+1}\right) \tag{54}$$

holds for sufficient small $h_0 > 0$, provided $f, a^i, b^{i,j} \in C_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R})$ for all $i = 1, \dots, d$ and $j = 1, \dots, m$.

Proof. Apply Theorem 5.8 with $n = 2(p+\frac{1}{2})$ and simply take the expectation of equation (49). By the definition of Φ_S and due to (19), for all $\mathbf{t} \in TS(\Delta)$ the expectation becomes

$$E(\Phi_S(\mathbf{t})) = O(h^{d(\mathbf{t})+\frac{1}{2}s(\mathbf{t})}) = O(h^{\rho(\mathbf{t})}).$$

Now, for all trees $\mathbf{t} \in TS(\Delta)$ appearing in the sum of equation (49) and which do not appear in the sum of (54), i.e. trees with $l(\mathbf{t}) \leq 2p+2$ and $\rho(\mathbf{t}) \geq p+1$, we have $E(\Phi_S(\mathbf{t})) = O(h^{p+1})$. As a result of this, we finally have to prove that $E^{t_0,x_0}(\mathcal{R}_{2p+1}(t,t_0)) = O(h^{p+1})$ holds. In the following, let $h < 1$. The autonomous version of the SRK method (16) can be written as

$$\begin{aligned} H^{(k,\nu)} &= (e \otimes I)\, Y_n + \sum_{r=0}^{m} \sum_{\mu,\iota \in \overline{\mathcal{M}}} \theta_\iota(h) \left(B^{(\iota)\,(k,\nu),(r,\mu)} \otimes I \right) G_r \left(H^{(r,\mu)} \right) \\ Y_{n+1} &= Y_n + \sum_{k=0}^{m} \sum_{\nu,\iota \in \overline{\mathcal{M}}} \theta_\iota(h) \left(\gamma^{(\iota)\,(k,\nu)} \otimes I \right) G_k \left(H^{(k,\nu)} \right) \end{aligned} \tag{55}$$

Here, denote $\theta_0(h) = h$ and $\gamma^{(0)\,(0,0)} = \alpha$, $\gamma^{(0)\,(k,\nu)} = 0$ for $k \neq 0$ or $\nu \neq 0$, $B^{(0)\,(k,\nu),(0,0)} = A^{(k,\nu),(0,0)}$ and $B^{(0)\,(k,\nu),(r,\mu)} = 0$ for $r \neq 0$ or $\mu \neq 0$. Further we denote $b^0 = a$, $G_k(H^{(k,\nu)}) = (b^k(H_1^{(k,\nu)})^T, \dots, b^k(H_s^{(k,\nu)})^T)^T \in \mathbb{R}^{d \cdot s}$, $H^{(k,\nu)} = (H_1^{(k,\nu)^T}, \dots, H_s^{(k,\nu)^T})^T \in \mathbb{R}^{d \cdot s}$, $I \in \mathbb{R}^{d \times d}$ and $e = (1, \dots, 1)^T \in \mathbb{R}^s$. In the following the norm $\|G_k(H^{(k,\nu)})\| = \max_{1 \leq i \leq s} \|b^k(H_i^{(k,\nu)})\|$ is used. Then, with the linear growth condition $\|G_k(H^{(k,\nu)})\| \leq C_1(1 + \|H^{(k,\nu)}\|)$ and

$$C_2 = \max_{\iota,k,\nu,r,\mu} \left\{ \left\| B^{(\iota)\,(k,\nu),(r,\mu)} \otimes I \right\|, \left\| \gamma^{(\iota)\,(k,\nu)} \otimes I \right\|, \|e \otimes I\| \right\}$$

the following inequality holds:

$$\begin{aligned} \max_{(k,\nu)} \left\| H^{(k,\nu)} \right\| &\leq C_2 \|Y_n\| + \sum_{r=0}^{m} \sum_{\mu,\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)|\, C_2\, C_1 \left(1 + \left\| H^{(r,\mu)} \right\| \right) \\ &\leq C_2 \|Y_n\| + (m+1) |\overline{\mathcal{M}}|^2 \max_{\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)|\; C_1 C_2 \left(1 + \max_{(k,\nu)} \left\| H^{(k,\nu)} \right\| \right) \end{aligned} \tag{56}$$

Let $C_3 = (m+1)\, |\overline{\mathcal{M}}|^2\, C_1 C_2$. Then for $\max_{\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)| \leq \frac{1}{2\, C_3}$ holds

$$\begin{aligned} \max_{(k,\nu)} \left\| H^{(k,\nu)} \right\| &\leq \left(C_2 \|Y_n\| + C_3 \max_{\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)| \right) \frac{1}{1 - C_3 \max_{\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)|} \\ &\leq 2\, C_2 \|Y_n\| + 2\, C_3 \max_{\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)| \\ &\leq C_4\, (1 + \|Y_n\|) \end{aligned} \tag{57}$$

Next, consider the qth derivative. By (26) and similar considerations, we get with the application of Lemma 5.2 using the notation (29) that

$$
\begin{aligned}
&\left\| \sum_{\nu_1,\ldots,\nu_q \in \overline{\mathcal{M}}} \frac{\partial^q H^{(k,\nu)\,J}}{\partial\theta_{\nu_1}\ldots\partial\theta_{\nu_q}} \right\| \\
&\leq q \sum_{r=0}^{m} \sum_{\mu,\nu_q \in \overline{\mathcal{M}}} \left\| B^{(\nu_q)\,(k,\nu),(r,\mu)} \otimes I \right\| \left\| \sum_{\nu_1,\ldots,\nu_{q-1} \in \overline{\mathcal{M}}} \frac{\partial^{q-1} G_r\left(H^{(r,\mu)}\right)^J}{\partial\theta_{\nu_1}\ldots\partial\theta_{\nu_{q-1}}} \right\| \\
&+ \sum_{r=0}^{m} \sum_{\mu,\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)| \left\| B^{(\iota)\,(k,\nu),(r,\mu)} \otimes I \right\| \left\| \sum_{\nu_1,\ldots,\nu_q \in \overline{\mathcal{M}}} \frac{\partial^q G_r\left(H^{(r,\mu)}\right)^J}{\partial\theta_{\nu_1}\ldots\partial\theta_{\nu_q}} \right\| \\
&\leq q\,|\overline{\mathcal{M}}| C_2 \sum_{r=0}^{m} \sum_{\mu \in \overline{\mathcal{M}}} \sum_{\mathbf{u} \in SLTS_q^{(\sigma_r)}} \sum_{K_1,\ldots,K_{m(\mathbf{u})}=1}^{d\cdot s} \left\| G^J_{r\,K_1\ldots K_{m(\mathbf{u})}}\left(H^{(r,\mu)}\right) \right\| \times \\
&\times \left\| \left(H^{(r,\mu)\,K_1}\right)^{(\delta_1)} \right\| \cdot \ldots \cdot \left\| \left(H^{(r,\mu)\,K_{m(\mathbf{u})}}\right)^{(\delta_{m(\mathbf{u})})} \right\| \\
&+ |\overline{\mathcal{M}}| C_2 \max_{\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)| \sum_{r=0}^{m} \sum_{\mu \in \overline{\mathcal{M}}} \sum_{\substack{\mathbf{u} \in SLTS_{q+1}^{(\sigma_r)} \\ m(\mathbf{u})>1}} \sum_{K_1,\ldots,K_{m(\mathbf{u})}=1}^{d\cdot s} \left\| G^J_{r\,K_1\ldots K_{m(\mathbf{u})}}\left(H^{(r,\mu)}\right) \right\| \times \\
&\times \left\| \left(H^{(r,\mu)\,K_1}\right)^{(\delta_1)} \right\| \cdot \ldots \cdot \left\| \left(H^{(r,\mu)\,K_{m(\mathbf{u})}}\right)^{(\delta_{m(\mathbf{u})})} \right\| \\
&+ |\overline{\mathcal{M}}| C_2 \max_{\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)| \sum_{r=0}^{m} \sum_{\mu \in \overline{\mathcal{M}}} \sum_{K_1=1}^{d\cdot s} \left\| G^J_{r\,K_1}\left(H^{(r,\mu)}\right) \right\| \cdot \left\| \left(H^{(r,\mu)\,K_1}\right)^{(q)} \right\|.
\end{aligned}
\tag{58}
$$

Due to the Lipschitz condition and the polynomial growth condition, we have $\|G^J_{r\,K_1}(H^{(r,\mu)})\| \leq L$ and $\|G^J_{r\,K_1\ldots K_{m(\mathbf{u})}}(H^{(r,\mu)})\| \leq C_5(1 + (\max_{(k,\nu)} \|H^{(k,\nu)}\|)^{2l})$ which is bounded by some constant C_6 only depending on $\|Y_n\|$ due to (57). Therefore, we get with $C_7 = C_2\,|\overline{\mathcal{M}}|^2\,(m+1)$

$$
\begin{aligned}
\max_{J,(k,\nu)} \left\| \left(H^{(k,\nu)\,J}\right)^{(q)} \right\| &\leq q\,C_7 \sum_{\mathbf{u} \in SLTS_q^{(\sigma)}} (d\cdot s)^{m(\mathbf{u})}\, C_6 \prod_{i=1}^{m(\mathbf{u})} \max_{J,(k,\nu)} \left\| \left(H^{(k,\nu)\,J}\right)^{(\delta_i)} \right\| \\
&+ C_7 \max_{\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)| \sum_{\substack{\mathbf{u} \in SLTS_{q+1}^{(\sigma)} \\ m(\mathbf{u})>1}} (d\cdot s)^{m(\mathbf{u})}\, C_6 \prod_{i=1}^{m(\mathbf{u})} \max_{J,(k,\nu)} \left\| \left(H^{(k,\nu)\,J}\right)^{(\delta_i)} \right\| \\
&+ C_7 \max_{\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)|\, (d\cdot s)\, L \max_{J,(k,\nu)} \left\| \left(H^{(k,\nu)\,J}\right)^{(q)} \right\|.
\end{aligned}
\tag{59}
$$

Let $C_8 = C_7\, d\, s\, L$. Then we get with $\max_{\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)| \leq \frac{1}{2\,C_8}$ that

$$\max_{J,(k,\nu)} \left\| \left(H^{(k,\nu)^J} \right)^{(q)} \right\| \leq 2\, q\, C_7 \sum_{\mathbf{u} \in SLTS_q^{(\sigma)}} (d \cdot s)^{m(\mathbf{u})}\, C_6 \prod_{i=1}^{m(\mathbf{u})} \max_{J,(k,\nu)} \left\| \left(H^{(k,\nu)^J} \right)^{(\delta_i)} \right\|$$
$$+ 2\, C_7 \max_{\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)| \sum_{\substack{\mathbf{u} \in SLTS_{q+1}^{(\sigma)} \\ m(\mathbf{u})>1}} (d \cdot s)^{m(\mathbf{u})}\, C_6 \prod_{i=1}^{m(\mathbf{u})} \max_{J,(k,\nu)} \left\| \left(H^{(k,\nu)^J} \right)^{(\delta_i)} \right\| \tag{60}$$

holds with $\delta_i = \delta_i(\mathbf{u}) \leq q-1$ because $m(\mathbf{u}) > 1$. Especially for $q = 1$ where due to the linear growth condition $C_6 = C_9\,(1 + \|Y_n\|)$, we get

$$\max_{J,(k,\nu)} \left\| \sum_{\nu_1 \in \overline{\mathcal{M}}} \frac{\partial H^{(k,\nu)^J}}{\partial \theta_{\nu_1}(h)} \right\| \leq C_{10}\,(1 + \|Y_n\|)\,. \tag{61}$$

Applying formula (60) recursively and using finally (57) yields an upper bound $C_q(Y_n)$ of the qth derivative of $H^{(k,\nu)}$ only depending on $\|Y_n\|$ for all $q \in \mathbb{N}$. Due to the definition of C_2 and the same structure of $Y_{n+1} = A(Y_n, \theta(h))$ as $H^{(k,\nu)}$, the same upper bound holds also for the qth derivative of $A(Y_n, \theta(h))$. Since $f \in C_P^{2p+2}(\mathbb{R}^d, \mathbb{R})$, we get for $\xi \in]0,1[$ and $|\theta_\iota(h)| \leq \sqrt{h}\, C_\vartheta$ with the Jensen inequality

$$\begin{aligned}
&\left\| E^{t_0,x_0} \left(\sum_{\nu_1,\dots,\nu_{2p+2} \in \overline{\mathcal{M}}} \Delta\theta_{\nu_1} \cdot \ldots \cdot \Delta\theta_{\nu_{2p+2}} \frac{\partial^{2p+2} f(A(Y(t_0), \xi\, \theta(h)))}{\partial \theta_{\nu_1} \ldots \partial \theta_{\nu_{2p+2}}} \right) \right\| \\
&\leq E^{t_0,x_0} \Bigg(\left(\max_{\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)| \right)^{2p+2} \times \\
&\times \sum_{\mathbf{u} \in SLTS_{2p+3}^{(\gamma)}} \sum_{K_1,\dots,K_m=1}^{d} \|f_{K_1 \dots K_m}(A(Y(t_0), \xi\, \theta(h)))\| \prod_{i=1}^{m(\mathbf{u})} C_{\delta_i}(Y(t_0)) \Bigg) \\
&\leq h^{p+1} C_\vartheta^{2p+2} \sum_{\mathbf{u} \in SLTS_{2p+3}^{(\gamma)}} d^{m(\mathbf{u})} C_f \left(1 + (C_4(1 + \|x_0\|))^{2r(\mathbf{u})} \right) \prod_{i=1}^{m(\mathbf{u})} C_{\delta_i}(x_0)
\end{aligned} \tag{62}$$

and it follows $E^{t_0,x_0}(\mathcal{R}_{2p+1}(t,t_0)) = O(h^{p+1})$. □

The result of Proposition 6.1 can also be proved for general unbounded random variables in the case of explicit SRK methods (see [16], Proposition 2.6.1). However, especially for weak approximations it is usual to use bounded random variables which are often easier to generate (see, e.g., [7], [11], [19]).

The approximation Y has to be uniformly bounded with respect to the number N of steps in order to guaranty convergence. Therefore, sufficient conditions for the random variables and for some coefficients of the stochastic Runge-Kutta method are calculated.

Proposition 6.2 *Let $a^i, b^{i,j} \in C^1(\mathbb{R}^d, \mathbb{R})$ satisfy a Lipschitz and a linear growth condition and let for all $1 \leq k \leq m$ and $\nu \in \mathcal{M}$*

$$E\left(\sum_{i=1}^{s} \sum_{\iota \in \mathcal{M}} {\gamma_i^{(\iota)}}^{(k,\nu)} \theta_\iota(h)\right) = 0. \tag{63}$$

Further assume that each random variable can be expressed as $\theta_\iota(h) = \sqrt{h} \cdot \vartheta_\iota$ for $\iota \in \mathcal{M}$ with a bounded random variable ϑ_ι. Then the approximation Y by the stochastic Runge-Kutta method (16) has uniformly bounded moments, i.e. for $r \in \mathbb{N}$ the expectation $E(\|Y_n\|^{2r})$ is uniformly bounded w.r.t. the number of steps N for all $n = 0, 1, \ldots, N$.

Proof. Let $h < 1$. Using the notation (55) we get with the linear growth condition and with (57)

$$\begin{aligned} \|A(Y_n, \theta(h)) - Y_n\| &\leq \sum_{k=0}^{m} \sum_{\nu,\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)| \, C_2 \left\| G_k\left(H^{(k,\nu)}\right) \right\| \\ &\leq (m+1) \, |\overline{\mathcal{M}}|^2 \max_{\iota \in \overline{\mathcal{M}}} |\theta_\iota(h)| \, C_2 \, C_9 \left(1 + \max_{(k,\nu)} \left\| H^{(k,\nu)} \right\| \right) \\ &\leq C_{11} \left(1 + \|Y_n\|\right) \sqrt{h}. \end{aligned} \tag{64}$$

Next, we get with one step of the Taylor-expansion of G_k for $\xi \in]0,1[$ that

$$\begin{aligned} \|E\left(A(Y_n, \theta(h)) - Y_n\right)\| \leq & \left\| \sum_{k=1}^{m} \sum_{\nu,\iota \in \mathcal{M}} E\left(\theta_\iota(h)\right) \left({\gamma^{(\iota)}}^{(k,\nu)} \otimes I\right) G_k\left((e \otimes I)\, Y_n\right) \right\| \\ + h \left\| {\gamma^{(0)}}^{(0,0)} \otimes I \right\| \, \|G_0\left((e \otimes I)\, Y_n\right)\| + & \left\| E\left(\sum_{k=0}^{m} \sum_{\nu,\iota \in \overline{\mathcal{M}}} \theta_\iota(h) \left({\gamma^{(\iota)}}^{(k,\nu)} \otimes I\right) \times \right.\right. \\ \times \sum_{\mu \in \overline{\mathcal{M}}} \sum_{J=1}^{d \cdot s} \frac{\partial G_k\left(H^{(k,\nu)}\left(\xi \theta(h)\right)\right)}{\partial x^J} & \left.\left. \frac{\partial H^{(k,\nu)}\left(\xi \theta(h)\right)^J}{\partial \theta_\mu} \Delta\theta_\mu(h) \right) \right\|. \end{aligned} \tag{65}$$

The first summand on the right hand side vanishes due to (63). With a Lipschitz constant L for G and the linear growth condition, we get with the Jensen inequality

$$
\begin{aligned}
&\|E\left(A(Y_n,\theta(h))-Y_n\right)\| \leq h\, C_2\, C_9\, (1+\|(e\otimes I)Y_n\|)\\
&+E\left(\sum_{k=0}^{m}\sum_{\nu,\iota\in\overline{\mathcal{M}}}\left\|{\gamma^{(\iota)}}^{(k,\nu)}\otimes I\right\|\sum_{J=1}^{d\cdot s}L\left\|\sum_{\mu\in\overline{\mathcal{M}}}\frac{\partial H^{(k,\nu)}\left(\xi\theta(h)\right)^J}{\partial\theta_\mu}\right\|\left(\max_{\iota\in\overline{\mathcal{M}}}|\theta_\iota(h)|\right)^2\right).
\end{aligned}
\tag{66}
$$

Finally, applying (61) and the condition $|\theta_\iota(h)| \leq \sqrt{h}\, C_\vartheta$, we get

$$
\|E\left(A(Y_n,\theta(h))-Y_n\right)\| \leq C_{12}\,(1+\|Y_n\|)\; h \tag{67}
$$

Now, Lemma 1.3 can be applied because (64) and (67) are fulfilled. This yields the existence of $E(\|Y_n\|^{2r})$ for all $r \in \mathbb{N}$ and provides that the moments are uniformly bounded with respect to N and $n = 1,\ldots,N$. □

The next step is to compare the representations of the solution of the stochastic differential equation in Theorem 4.2 with the representation of the approximation in Proposition 6.1. Due to Theorem 1.2 these representations have to coincide up to order $p+1$ locally. This leads to different conditions w.r.t. trees in $TS(I)$ and $TS(S)$ on the one hand and trees in $TS(\Delta)\setminus TS(I)$ and $TS(\Delta)\setminus TS(S)$ on the other hand, respectively. Having in mind that for $\mathbf{t} \in TS(I)$ or $\mathbf{t} \in TS(S)$ we have $s(\mathbf{t})/2$ different variable indices while for the same tree $\mathbf{t} \in TS(\Delta)$ we have twice as much, i.e. $s(\mathbf{t})$ different variable indices, we use the following helpful definition.

Definition 6.3 *Let $|\mathbf{t}|$ denote the tree which is obtained if the nodes σ_{j_i} of $\mathbf{t}$ are replaced by σ, i.e. by omitting all variable indices. Let a tree $\mathbf{t} \in TS(*)$ for $* \in \{I,S\}$ with variable indices $j_1,\ldots,j_{s(\mathbf{t})/2}$ be given and let $\mathbf{u} \in TS(\Delta)$ with variable indices $\hat{j}_1,\ldots,\hat{j}_{s(\mathbf{u})}$ denote the tree which is equivalent to $\mathbf{t}$ except for the variable indices, i.e. $|\mathbf{t}| \sim |\mathbf{u}|$ with $s(\mathbf{t}) = s(\mathbf{u})$. For a fixed choice of correlations of type $j_k = j_l$ or $j_k \neq j_l$, $1 \leq k < l \leq s(\mathbf{t})/2$, between the indices $j_1,\ldots,j_{s(\mathbf{t})/2}$, let $\beta(\mathbf{t})$ denote the number of all possible correlations between the indices $\hat{j}_1,\ldots,\hat{j}_{s(\mathbf{u})}$ of tree $\mathbf{u}$ such that $\mathbf{t} \sim \mathbf{u}$ holds. In the case of $s(\mathbf{t}) = 0$ or $\mathbf{t} \in TS(\Delta)\setminus TS(*)$, $* \in \{I,S\}$, define $\beta(\mathbf{t}) = 1$.*

Remark that in case of $m = 1$ we have $\beta(\mathbf{t}) = 1$ for all $\mathbf{t} \in TS(*)$, $* \in \{I,S\}$. As an example consider the trees $\mathbf{t} = (\sigma_{j_1},\sigma_{j_1},\sigma_{j_2},\sigma_{j_2}) \in TS(I)$ and $\mathbf{u} = (\sigma_{\hat{j}_1},\sigma_{\hat{j}_2},\sigma_{\hat{j}_3},\sigma_{\hat{j}_4}) \in TS(\Delta)$. For the correlation $j_1 = j_2$ of $\mathbf{t}$ we have exactly one possibility for the choice of a correlation of $\mathbf{u}$: We have to choose $\hat{j}_1 = \hat{j}_2 = \hat{j}_3 = \hat{j}_4$, i.e. in this case we have $\beta(\mathbf{t}) = 1$. However, in case of the correlation $j_1 \neq j_2$ for $\mathbf{t}$, there are three different possible correlations for $\mathbf{u}$: We can choose $\hat{j}_1 = \hat{j}_2 \neq \hat{j}_3 = \hat{j}_4$, $\hat{j}_1 = \hat{j}_3 \neq \hat{j}_2 = \hat{j}_4$ or $\hat{j}_1 = \hat{j}_4 \neq \hat{j}_2 = \hat{j}_3$, thus we have $\beta(\mathbf{t}) = 3$. As a second example, for the trees $\mathbf{t} = (\sigma_{j_1},\sigma_{j_2},\{\sigma_{j_2}\}_{j_1}) \in TS(I)$ and $\mathbf{u} = (\sigma_{\hat{j}_1},\sigma_{\hat{j}_2},\{\sigma_{\hat{j}_4}\}_{\hat{j}_3}) \in TS(\Delta)$, two different correlations are distinguished. On the one hand we have the correlation $j_1 = j_2$ for $\mathbf{t}$ where we get the only possible correlation $\hat{j}_1 = \hat{j}_2 = \hat{j}_3 = \hat{j}_4$ for $\mathbf{u}$, i.e. $\beta(\mathbf{t}) = 1$. On the other hand we have $j_1 \neq j_2$ as a correlation for $\mathbf{t}$ allowing us two different correlations $\hat{j}_1 = \hat{j}_3 \neq \hat{j}_2 = \hat{j}_4$ and $\hat{j}_2 = \hat{j}_3 \neq \hat{j}_1 = \hat{j}_4$ for $\mathbf{u}$. Thus we get $\beta(\mathbf{t}) = 2$ in the latter case.

The main theorem for stochastic Runge-Kutta methods of type (16) yields general conditions for the coefficients and the random variables of the method such that convergence with some order p in the weak sense is assured. Remark that for every tree $\mathbf{t} \in TS(*)$ with variable indices $j_1, \dots, j_{s(\mathbf{t})/2}$ there exists a tree $\mathbf{u} \in TS(\Delta)$ with $|\mathbf{u}| \sim |\mathbf{t}|$ and variable indices $\hat{j}_1, \dots, \hat{j}_{s(\mathbf{u})}$ such that for some suitable correlation of type $\hat{j}_k = \hat{j}_l$ or $\hat{j}_k \neq \hat{j}_l$, $1 \leq k < l \leq s(\mathbf{u})$, we have $\mathbf{t} \sim \mathbf{u}$ and thus $\mathbf{u} \in TS(*)$ with $\alpha_*(\mathbf{u}) = \alpha_*(\mathbf{t})$ for $* \in \{I, S\}$. However, we have $\alpha_*(\mathbf{u}) = 0$ for all $\mathbf{u} \in TS(\Delta) \setminus TS(*)$ for $* \in \{I, S\}$.

Theorem 6.4 *Let X be the solution of either an Itô or Stratonovich stochastic differential equation (3) considered w.r.t. an m-dimensional Wiener process with $f, a^i, \tilde{a}^i, b^{i,j} \in C_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R})$ for $i = 1, \dots, d$ and $j = 1, \dots, m$. Then the approximation Y by the stochastic Runge-Kutta method (16) with maximum step size h is of weak order p, if for all $\boldsymbol{t} \in TS(\Delta)$ with $\rho(\boldsymbol{t}) \leq p + \frac{1}{2}$ and all correlations of type $j_k = j_l$ or $j_k \neq j_l$, $1 \leq k < l \leq s(\boldsymbol{t})$, between the indices $j_1, \dots, j_{s(\boldsymbol{t})} \in \{1, \dots, m\}$ the equations*

$$\frac{\alpha_*(\boldsymbol{t}) \cdot h^{\rho(\boldsymbol{t})}}{2^{s(\boldsymbol{t})/2} \cdot \rho(\boldsymbol{t})!} = \frac{\alpha_\Delta(\boldsymbol{t}) \cdot \beta(\boldsymbol{t}) \cdot \gamma(\boldsymbol{t}) \cdot E(\Phi_S(\boldsymbol{t}))}{(l(\boldsymbol{t}) - 1)!} \tag{68}$$

hold for $* = I$ in case of Itô SDEs and $* = S$ in case of Stratonovich SDEs, provided that (17) and (19) hold and that the approximation Y has uniformly bounded moments w.r.t. the number N of steps.

Proof. Apply Theorem 1.2 and compare the coefficients from the representations of the solution in Theorem 4.2 with the coefficients of the SRK method in Proposition 6.1, where $TS(*) \subseteq TS(\Delta)$, $* \in \{I, S\}$. Finally, we take into account the summation w.r.t. variabel indices. Therefore, the correlation index $\beta(\mathbf{t})$ has to be added and we yield the conditions (68). □

Remark 6.5 *Theorem 6.4 provides uniform weak convergence with order p on the interval $I = [t_0, T]$ for the stochastic Runge-Kutta method in the case of a non-random time discretization I_h. That is for each $f \in C_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R})$ there exists a finite constant C_f not depending on the maximum step size h such that*

$$\max_{0 \leq k \leq N} \left| E(f(X_{t_k})) - E(f(Y_k)) \right| \leq C_f \, h^p \tag{69}$$

holds. This is a consequence of Theorem 1.2 (see, e.g., [7], [11]).

Table A.1 contains all S-trees of $TS(I)$ and $TS(S)$ up to order two with the corresponding cardinalities α_I and α_S. Table A.3 contains all S-trees of $TS(\Delta)$ up to order 2.5 with the values of α_Δ. The cardinalities can be determined very easily as the number of possibilities to build up the considered tree due to the corresponding rules of growth. Together with Table A.2 containing the values of β, we can consider the following example:

Example 6.6 *Assume that $m \geq 1$.*

a) As a first example, let us have a look at tree $\mathbf{t}_{2.5} = (\sigma_{j_1}, [\sigma_{j_2}]) \in TS(\Delta)$ with parameters $l(\mathbf{t}_{2.5}) = 4$, $\gamma(\mathbf{t}_{2.5}) = 2$, $s(\mathbf{t}_{2.5}) = 2$, $\alpha_\Delta(\mathbf{t}_{2.5}) = 3$ and $\rho(\mathbf{t}_{2.5}) = 2$. Then the following correlations have to be distinguished: For $j_1 = j_2$ follows that $\mathbf{t}_{2.5} \in TS(*)$ with $\alpha_I(\mathbf{t}_{2.5}) = \alpha_S(\mathbf{t}_{2.5}) = 2$ and $\beta(\mathbf{t}_{2.5}) = 1$. Then for $j_1 \in \{1, \ldots, m\}$ Theorem 6.4 yields the conditions

$$E\left(\left(\sum_{\nu\in\mathcal{M}} z^{(j_1,\nu)^T} e\right)\left(z^{(0,0)^T}\left(\sum_{\mu\in\mathcal{M}} Z^{(0,0),(j_1,\mu)} e\right)\right)\right) = \frac{2\cdot 3!\cdot h^2}{2^1\cdot 2!\cdot 3\cdot 1\cdot 2}.$$

Here, the conditions for Itô and Stratonovich calculus coincide. However, for $j_1 \neq j_2$ follows $\mathbf{t}_{2.5} \notin TS(*)$, i.e. $\alpha_I(\mathbf{t}_{2.5}) = \alpha_S(\mathbf{t}_{2.5}) = 0$, and thus one gets for $j_1, j_2 \in \{1, \ldots, m\}$ with $j_1 \neq j_2$ the additional conditions

$$E\left(\left(\sum_{\nu\in\mathcal{M}} z^{(j_1,\nu)^T} e\right)\left(z^{(0,0)^T}\left(\sum_{\mu\in\mathcal{M}} Z^{(0,0),(j_2,\mu)} e\right)\right)\right) = 0.$$

b) Consider $\mathbf{t}_{2.11} = (\sigma_{j_1}, \sigma_{j_2}, \sigma_{j_3}, \sigma_{j_4}) \in TS(\Delta)$ with $l(\mathbf{t}_{2.11}) = 5$, $\gamma(\mathbf{t}_{2.11}) = 1$, $s(\mathbf{t}_{2.11}) = 4$, $\alpha_\Delta(\mathbf{t}_{2.11}) = 1$ and $\rho(\mathbf{t}_{2.11}) = 2$. The following correlations have to be analyzed: For $j_1 = j_3 \neq j_2 = j_4$ we have $\mathbf{t}_{2.11} \in TS(*)$ with $\alpha_I(\mathbf{t}_{2.11}) = \alpha_S(\mathbf{t}_{2.11}) = 1$ and $\beta(\mathbf{t}_{2.11}) = 3$. Then Theorem 6.4 yields the condition

$$E\left(\left(\sum_{\nu\in\mathcal{M}} z^{(j_1,\nu)^T} e\right)^2\left(\sum_{\nu\in\mathcal{M}} z^{(j_2,\nu)^T} e\right)^2\right) = \frac{4!\cdot h^2}{2^2\cdot 2!\cdot 3}$$

with $j_1, j_2 \in \{1, \ldots, m\}$, $j_1 \neq j_2$, for both, Itô and Stratonovich calculus. For $j_1 = j_2 = j_3 = j_4$ we have $\mathbf{t}_{2.11} \in TS(*)$ with $\alpha_I(\mathbf{t}_{2.11}) = \alpha_S(\mathbf{t}_{2.11}) = 1$ and $\beta(\mathbf{t}_{2.11}) = 1$. Again, Theorem 6.4 yields the condition

$$E\left(\left(\sum_{\nu\in\mathcal{M}} z^{(j_1,\nu)^T} e\right)^4\right) = \frac{4!\cdot h^2}{2^2\cdot 2!\cdot 1}$$

with $j_1 \in \{1, \ldots, m\}$ for both, Itô and Stratonovich calculus. For all remaining correlations of the indices follows that $\mathbf{t}_{2.11} \notin TS(*)$ and thus $\alpha_I(\mathbf{t}_{2.11}) = \alpha_S(\mathbf{t}_{2.11}) = 0$. Therefore, the condition $E(\Phi_S(\mathbf{t}_{2.11})) = 0$ has to be fulfilled for the remaining correlations.

c) For $\boldsymbol{t}_{2.12} = (\sigma_{j_1}, \sigma_{j_2}, \{\sigma_{j_4}\}_{j_3})$ with $l(\boldsymbol{t}_{2.12}) = 5$, $\gamma(\boldsymbol{t}_{2.12}) = 2$, $s(\boldsymbol{t}_{2.12}) = 4$, $\alpha_\Delta(\boldsymbol{t}_{2.12}) = 6$ and $\rho(\boldsymbol{t}_{2.12}) = 2$, consider the following correlations: For $j_1 = j_2 \neq j_3 = j_4$ we have $\boldsymbol{t}_{2.12} = \boldsymbol{t}_{2.12a} \in TS(S)$ with $\alpha_S(\boldsymbol{t}_{2.12a}) = 2$ and $\beta(\boldsymbol{t}_{2.12a}) = 1$. Therefore, we get the condition

$$E\left(\left(\sum_{\nu\in\mathcal{M}} z^{(j_1,\nu)^T} e\right)^2 \left(\sum_{\nu\in\mathcal{M}} z^{(j_3,\nu)^T} \left(\sum_{\mu\in\mathcal{M}} Z^{(j_3,\nu)(j_3,\mu)} e\right)\right)\right) = \frac{2\cdot 4!\cdot h^2}{2^2\cdot 2!\cdot 6\cdot 2}$$

with $j_1, j_3 \in \{1,\dots,m\}$, $j_1 \neq j_3$, for Stratonovich calculus. However, since $\boldsymbol{t}_{2.12a} \notin TS(I)$ we get for Itô calculus the condition $E(\Phi_S(\boldsymbol{t}_{2.12a})) = 0$ since $\alpha_I(\boldsymbol{t}_{2.12a}) = 0$. For $j_1 = j_3 \neq j_2 = j_4$ or $j_2 = j_3 \neq j_1 = j_4$ we have $\boldsymbol{t}_{2.12} = \boldsymbol{t}_{2.12b} \in TS(*)$ with $\alpha_I(t_{2.12b}) = \alpha_S(\boldsymbol{t}_{2.12b}) = 4$ and $\beta(t_{2.12b}) = 2$. Here we get the condition

$$E\left(\left(\sum_{\nu\in\mathcal{M}} z^{(j_1,\nu)^T} e\right)\left(\sum_{\nu\in\mathcal{M}} z^{(j_2,\nu)^T} e\right)\left(\sum_{\nu\in\mathcal{M}} z^{(j_1,\nu)^T} \left(\sum_{\mu\in\mathcal{M}} Z^{(j_1,\nu)(j_2,\mu)} e\right)\right)\right) = \frac{4\cdot 4!\cdot h^2}{2^2\cdot 2!\cdot 6\cdot 2\cdot 2} \tag{70}$$

with $j_1, j_2 \in \{1,\dots,m\}$, $j_1 \neq j_2$, for Itô and Stratonovich calculus. Further, for $j_1 = j_2 = j_3 = j_4$ we have $\boldsymbol{t}_{2.12} \in TS(*)$ with $\alpha_I(\boldsymbol{t}_{2.12}) = 0 + 4$, $\alpha_S(\boldsymbol{t}_{2.12}) = 2 + 4$ and $\beta(\boldsymbol{t}_{2.12}) = 1$. Therefore, we get the conditions

$$E\left(\left(\sum_{\nu\in\mathcal{M}} z^{(j_1,\nu)^T} e\right)^2 \left(\sum_{\nu\in\mathcal{M}} z^{(j_1,\nu)^T} \left(\sum_{\mu\in\mathcal{M}} Z^{(j_1,\nu)(j_1,\mu)} e\right)\right)\right) = \frac{\alpha_*(\boldsymbol{t}_{2.12})\, 4!\, h^2}{2^2\cdot 2!\cdot 6\cdot 2}$$

with $j_1 \in \{1,\dots,m\}$. Thus we have different conditions for Itô and Stratonovich calculus. Finally, for all remaining correlations the conditions $E(\Phi_S(\boldsymbol{t}_{2.12})) = 0$ have to hold due to $\boldsymbol{t}_{2.12} \notin TS(*)$ in these cases.

7 Conclusions

The present paper introduces a very general class of stochastic Runge-Kutta methods for the approximation of stochastic differential equations. Explicit as well as implicit SRK methods for non-autonomous SDE systems w.r.t. to a multi-dimensional Wiener process are considered. A rigorous analysis of the weak convergence for the SRK method is given. Therefore, colored rooted trees are introduced and an expansion of the solution and of the approximation process is given. Finally, a theorem giving directly the order conditions for arbitrary high order of convergence is proved. The main advantages of the rooted tree analysis are as follows: The required colored rooted trees can be easily determined. So in contrast to the usual direct comparison of the Taylor expansions, one needs not to calculate the derivatives of f, a and b. It has to be pointed out that the calculated order conditions dependent on the coefficients and the random variables of the SRK method. Therefore, the order conditions can also be used for the determination of suitable random variables for the SRK method. In order to get a closed theory, the presented results cover SRK methods for the approximation of both, Itô and Stratonovich SDE systems. Finally, the presented colored rooted tree theory and the introduced SRK methods generalize the well known theory for deterministic Runge-Kutta methods due to Butcher [3]. In the case of $b \equiv 0$ and $f(x) = x$, i.e. an ordinary differential equation, the SRK method coincides with a deterministic Runge-Kutta method and also the order conditions coincide with the deterministic order conditions. For some examples of SRK methods and the corresponding analysis of order conditions with colored rooted trees, we refer to Rößler [16].

A Tables

Table A.1: Trees $\mathbf{t} \in TS(*)$, $* \in \{I, S\}$, of order $\rho(\mathbf{t}) \leq 2.5$ with variable indices $j_1, j_2 \in \{1, \ldots, m\}$.

$\mathbf{t}$	tree	α_I	α_S	ρ	$\mathbf{t}$	tree	α_I	α_S	ρ
$\mathbf{t}_{0.1}$	γ	1	1	0	$\mathbf{t}_{2.11}$	$(\sigma_{j_1}, \sigma_{j_1}, \sigma_{j_2}, \sigma_{j_2})$	1	1	2
$\mathbf{t}_{1.1}$	(τ)	1	1	1	$\mathbf{t}_{2.12a}$	$(\sigma_{j_1}, \sigma_{j_1}, \{\sigma_{j_2}\}_{j_2})$	0	2	2
$\mathbf{t}_{1.2}$	$(\sigma_{j_1}, \sigma_{j_1})$	1	1	1	$\mathbf{t}_{2.12b}$	$(\sigma_{j_1}, \sigma_{j_2}, \{\sigma_{j_2}\}_{j_1})$	4	4	2
$\mathbf{t}_{1.3}$	$(\{\sigma_{j_1}\}_{j_1})$	0	1	1	$\mathbf{t}_{2.13a}$	$(\sigma_{j_1}, \{\sigma_{j_2}, \sigma_{j_2}\}_{j_1})$	2	2	2
$\mathbf{t}_{2.1}$	$([\tau])$	1	1	2	$\mathbf{t}_{2.13b}$	$(\sigma_{j_2}, \{\sigma_{j_2}, \sigma_{j_1}\}_{j_1})$	0	2	2
$\mathbf{t}_{2.2}$	(τ, τ)	1	1	2	$\mathbf{t}_{2.14a}$	$(\sigma_{j_1}, \{\{\sigma_{j_2}\}_{j_2}\}_{j_1})$	0	2	2
$\mathbf{t}_{2.3}$	$([\{\sigma_{j_1}\}_{j_1}])$	0	1	2	$\mathbf{t}_{2.14b}$	$(\sigma_{j_2}, \{\{\sigma_{j_2}\}_{j_1}\}_{j_1})$	0	2	2
$\mathbf{t}_{2.4}$	$([\sigma_{j_1}, \sigma_{j_1}])$	1	1	2	$\mathbf{t}_{2.15a}$	$(\{\sigma_{j_1}\}_{j_1}, \{\sigma_{j_2}\}_{j_2})$	0	1	2
$\mathbf{t}_{2.5}$	$(\sigma_{j_1}, [\sigma_{j_1}])$	2	2	2	$\mathbf{t}_{2.15b}$	$(\{\sigma_{j_2}\}_{j_1}, \{\sigma_{j_2}\}_{j_1})$	2	2	2
$\mathbf{t}_{2.6}$	$(\{\sigma_{j_1}\}_{j_1}, \tau)$	0	2	2	$\mathbf{t}_{2.16}$	$(\{\sigma_{j_1}, \sigma_{j_2}, \sigma_{j_2}\}_{j_1})$	0	1	2
$\mathbf{t}_{2.7}$	$(\sigma_{j_1}, \sigma_{j_1}, \tau)$	2	2	2	$\mathbf{t}_{2.17a}$	$(\{\sigma_{j_1}, \{\sigma_{j_2}\}_{j_2}\}_{j_1})$	0	1	2
$\mathbf{t}_{2.8}$	$(\sigma_{j_1}, \{\tau\}_{j_1})$	2	2	2	$\mathbf{t}_{2.17b}$	$(\{\sigma_{j_2}, \{\sigma_{j_2}\}_{j_1}\}_{j_1})$	0	2	2
$\mathbf{t}_{2.9}$	$(\{\{\tau\}_{j_1}\}_{j_1})$	0	1	2	$\mathbf{t}_{2.18}$	$(\{\{\sigma_{j_2}, \sigma_{j_2}\}_{j_1}\}_{j_1})$	0	1	2
$\mathbf{t}_{2.10}$	$(\{\sigma_{j_1}, \tau\}_{j_1})$	0	1	2	$\mathbf{t}_{2.19}$	$(\{\{\{\sigma_{j_2}\}_{j_2}\}_{j_1}\}_{j_1})$	0	1	2

Remark A.1 *If we choose $j_1 = j_2$ then some trees of Table A.1 may coincide. In this case α_* has to be taken as the sum of the values α_* from the coinciding trees. As an example, for $j_1 = j_2$ we get $\alpha_I(\mathbf{t}_{2.15}) = 0+2$ and $\alpha_S(\mathbf{t}_{2.15}) = 1+2$.*

Table A.2: The correlation coefficient $\beta(\mathbf{t})$ for some trees $\mathbf{t} \in TS(*)$, $* \in \{I, S\}$, and $j_1, j_2 \in \{1, \ldots, m\}$. For trees with $\rho(\mathbf{t}) \leq 2.5$ which are not listed holds $\beta(\mathbf{t}) = 1$.

$\mathbf{t}$	correlation	α_I	α_S	β	$\mathbf{t}$	correlation	α_I	α_S	β
$\mathbf{t}_{2.11}$	$j_1 \neq j_2$	1	1	3	$\mathbf{t}_{2.12b}$	$j_1 \neq j_2$	4	4	2
$\mathbf{t}_{2.11}$	$j_1 = j_2$	1	1	1	$\mathbf{t}_{2.12}$	$j_1 = j_2$	4	6	1
$\mathbf{t}_{2.13b}$	$j_1 \neq j_2$	0	2	2	$\mathbf{t}_{2.16}$	$j_1 \neq j_2$	0	1	3
$\mathbf{t}_{2.13}$	$j_1 = j_2$	2	4	1	$\mathbf{t}_{2.16}$	$j_1 = j_2$	0	1	1

Table A.3: Trees $\mathbf{t} \in TS(\Delta)$ of order $\rho(\mathbf{t}) \leq 2.5$ with arbitrary choice of $j_1, j_2, j_3, j_4, j_5 \in \{1, \ldots, m\}$.

$\mathbf{t}$	tree	α_Δ	$\mathbf{t}$	tree	α_Δ
$\mathbf{t}_{0.1}$	γ	1	$\mathbf{t}_{0.5.1}$	(σ_{j_1})	1
$\mathbf{t}_{1.1}$	(τ)	1	$\mathbf{t}_{1.2}$	$(\sigma_{j_1}, \sigma_{j_2})$	1
$\mathbf{t}_{1.3}$	$(\{\sigma_{j_2}\}_{j_1})$	1			
$\mathbf{t}_{1.5.1}$	$([\sigma_{j_1}])$	1	$\mathbf{t}_{1.5.2}$	$(\{\tau\}_{j_1})$	1
$\mathbf{t}_{1.5.3}$	(τ, σ_{j_1})	2	$\mathbf{t}_{1.5.4}$	$(\sigma_{j_1}, \sigma_{j_2}, \sigma_{j_3})$	1
$\mathbf{t}_{1.5.5}$	$(\{\sigma_{j_2}\}_{j_1}, \sigma_{j_3})$	3	$\mathbf{t}_{1.5.6}$	$(\{\sigma_{j_2}, \sigma_{j_3}\}_{j_1})$	1
$\mathbf{t}_{1.5.7}$	$(\{\{\sigma_{j_3}\}_{j_2}\}_{j_1})$	1			
$\mathbf{t}_{2.1}$	$([\tau])$	1	$\mathbf{t}_{2.2}$	(τ, τ)	1
$\mathbf{t}_{2.3}$	$([\{\sigma_{j_2}\}_{j_1}])$	1	$\mathbf{t}_{2.4}$	$([\sigma_{j_1}, \sigma_{j_2}])$	1
$\mathbf{t}_{2.5}$	$(\sigma_{j_1}, [\sigma_{j_2}])$	3	$\mathbf{t}_{2.6}$	$(\{\sigma_{j_2}\}_{j_1}, \tau)$	3
$\mathbf{t}_{2.7}$	$(\sigma_{j_1}, \sigma_{j_2}, \tau)$	3	$\mathbf{t}_{2.8}$	$(\sigma_{j_1}, \{\tau\}_{j_2})$	3
$\mathbf{t}_{2.9}$	$(\{\{\tau\}_{j_2}\}_{j_1})$	1	$\mathbf{t}_{2.10}$	$(\{\sigma_{j_2}, \tau\}_{j_1})$	2
$\mathbf{t}_{2.11}$	$(\sigma_{j_1}, \sigma_{j_2}, \sigma_{j_3}, \sigma_{j_4})$	1	$\mathbf{t}_{2.12}$	$(\sigma_{j_1}, \sigma_{j_2}, \{\sigma_{j_4}\}_{j_3})$	6
$\mathbf{t}_{2.13}$	$(\sigma_{j_1}, \{\sigma_{j_3}, \sigma_{j_4}\}_{j_2})$	4	$\mathbf{t}_{2.14}$	$(\sigma_{j_1}, \{\{\sigma_{j_4}\}_{j_3}\}_{j_2})$	4

$\mathbf{t}$	tree	α_Δ	$\mathbf{t}$	tree	α_Δ
$\mathbf{t}_{2.15}$	$(\{\sigma_{j_2}\}_{j_1},\{\sigma_{j_4}\}_{j_3})$	3	$\mathbf{t}_{2.16}$	$(\{\sigma_{j_2},\sigma_{j_3},\sigma_{j_4}\}_{j_1})$	1
$\mathbf{t}_{2.17}$	$(\{\sigma_{j_2},\{\sigma_{j_4}\}_{j_3}\}_{j_1})$	3	$\mathbf{t}_{2.18}$	$(\{\{\sigma_{j_3},\sigma_{j_4}\}_{j_2}\}_{j_1})$	1
$\mathbf{t}_{2.19}$	$(\{\{\{\sigma_{j_4}\}_{j_3}\}_{j_2}\}_{j_1})$	1	$\mathbf{t}_{2.20}$	$(\{[\sigma_{j_2}]\}_{j_1})$	1
$\mathbf{t}_{2.5.1}$	(τ,τ,σ_{j_1})	3	$\mathbf{t}_{2.5.2}$	$([\sigma_{j_1}],\tau)$	3
$\mathbf{t}_{2.5.3}$	$([\tau],\sigma_{j_1})$	3	$\mathbf{t}_{2.5.4}$	$([\tau,\sigma_{j_1}])$	2
$\mathbf{t}_{2.5.5}$	$([[\sigma_{j_1}]])$	1	$\mathbf{t}_{2.5.6}$	$(\{\tau\}_{j_1},\tau)$	3
$\mathbf{t}_{2.5.7}$	$(\{\tau,\tau\}_{j_1})$	1	$\mathbf{t}_{2.5.8}$	$(\{[\tau]\}_{j_1})$	1
$\mathbf{t}_{2.5.9}$	$([\{\tau\}_{j_1}])$	1	$\mathbf{t}_{2.5.10}$	$(\tau,\sigma_{j_1},\sigma_{j_2},\sigma_{j_3})$	4
$\mathbf{t}_{2.5.11}$	$([\sigma_{j_1}],\sigma_{j_2},\sigma_{j_3})$	6	$\mathbf{t}_{2.5.12}$	$(\tau,\{\sigma_{j_2}\}_{j_1},\sigma_{j_3})$	12
$\mathbf{t}_{2.5.13}$	$([\sigma_{j_1},\sigma_{j_2}],\sigma_{j_3})$	4	$\mathbf{t}_{2.5.14}$	$([\{\sigma_{j_2}\}_{j_1}],\sigma_{j_3})$	4
$\mathbf{t}_{2.5.15}$	$([\sigma_{j_1}],\{\sigma_{j_3}\}_{j_2})$	6	$\mathbf{t}_{2.5.16}$	$(\tau,\{\sigma_{j_2},\sigma_{j_3}\}_{j_1})$	4
$\mathbf{t}_{2.5.17}$	$(\tau,\{\{\sigma_{j_3}\}_{j_2}\}_{j_1})$	4	$\mathbf{t}_{2.5.18}$	$([\sigma_{j_1},\sigma_{j_2},\sigma_{j_3}])$	3
$\mathbf{t}_{2.5.19}$	$([\sigma_{j_1},\{\sigma_{j_3}\}_{j_2}])$	3	$\mathbf{t}_{2.5.20}$	$([\{\sigma_{j_2},\sigma_{j_3}\}_{j_1}])$	1
$\mathbf{t}_{2.5.21}$	$([\{\{\sigma_{j_3}\}_{j_2}\}_{j_1}])$	1	$\mathbf{t}_{2.5.22}$	$(\{\tau\}_{j_1},\sigma_{j_2},\sigma_{j_3})$	6
$\mathbf{t}_{2.5.23}$	$(\{\tau\}_{j_1},\{\sigma_{j_3}\}_{j_2})$	6	$\mathbf{t}_{2.5.24}$	$(\{\tau,\sigma_{j_2}\}_{j_1},\sigma_{j_3})$	8
$\mathbf{t}_{2.5.25}$	$(\{[\sigma_{j_2}]\}_{j_1},\sigma_{j_3})$	4	$\mathbf{t}_{2.5.26}$	$(\{\tau,\sigma_{j_2},\sigma_{j_3}\}_{j_1})$	3
$\mathbf{t}_{2.5.27}$	$(\{\tau,\{\sigma_{j_3}\}_{j_2}\}_{j_1})$	3	$\mathbf{t}_{2.5.28}$	$(\{[\sigma_{j_2}],\sigma_{j_3}\}_{j_1})$	3
$\mathbf{t}_{2.5.29}$	$(\{[\sigma_{j_2},\sigma_{j_3}]\}_{j_1})$	1	$\mathbf{t}_{2.5.30}$	$(\{[\{\sigma_{j_3}\}_{j_2}]\}_{j_1})$	1
$\mathbf{t}_{2.5.31}$	$(\{\{\tau\}_{j_2}\}_{j_1},\sigma_{j_3})$	4	$\mathbf{t}_{2.5.32}$	$(\{\{\tau\}_{j_2},\sigma_{j_3}\}_{j_1})$	3
$\mathbf{t}_{2.5.33}$	$(\{\{\tau,\sigma_{j_3}\}_{j_2}\}_{j_1})$	2	$\mathbf{t}_{2.5.34}$	$(\{\{[\sigma_{j_3}]\}_{j_2}\}_{j_1})$	1
$\mathbf{t}_{2.5.35}$	$(\{\{\{\tau\}_{j_3}\}_{j_2}\}_{j_1})$	1	$\mathbf{t}_{2.5.36}$	$(\sigma_{j_1},\sigma_{j_2},\sigma_{j_3},\sigma_{j_4},\sigma_{j_5})$	1
$\mathbf{t}_{2.5.37}$	$(\sigma_{j_1},\sigma_{j_2},\sigma_{j_3},\{\sigma_{j_5}\}_{j_4})$	10	$\mathbf{t}_{2.5.38}$	$(\sigma_{j_1},\sigma_{j_2},\{\sigma_{j_4},\sigma_{j_5}\}_{j_3})$	10
$\mathbf{t}_{2.5.39}$	$(\sigma_{j_1},\sigma_{j_2},\{\{\sigma_{j_5}\}_{j_4}\}_{j_3})$	10	$\mathbf{t}_{2.5.40}$	$(\sigma_{j_1},\{\sigma_{j_3}\}_{j_2},\{\sigma_{j_5}\}_{j_4})$	15
$\mathbf{t}_{2.5.41}$	$(\{\sigma_{j_2}\}_{j_1},\{\sigma_{j_4},\sigma_{j_5}\}_{j_3})$	10	$\mathbf{t}_{2.5.42}$	$(\{\sigma_{j_2}\}_{j_1},\{\{\sigma_{j_5}\}_{j_4}\}_{j_3})$	10
$\mathbf{t}_{2.5.43}$	$(\{\sigma_{j_2},\sigma_{j_3},\sigma_{j_4}\}_{j_1},\sigma_{j_5})$	5	$\mathbf{t}_{2.5.44}$	$(\{\sigma_{j_2},\{\sigma_{j_4}\}_{j_3}\}_{j_1},\sigma_{j_5})$	15
$\mathbf{t}_{2.5.45}$	$(\{\{\sigma_{j_3},\sigma_{j_4}\}_{j_2}\}_{j_1},\sigma_{j_5})$	5	$\mathbf{t}_{2.5.46}$	$(\{\{\{\sigma_{j_4}\}_{j_3}\}_{j_2}\}_{j_1},\sigma_{j_5})$	5
$\mathbf{t}_{2.5.47}$	$(\{\sigma_{j_2},\sigma_{j_3},\sigma_{j_4},\sigma_{j_5}\}_{j_1})$	1	$\mathbf{t}_{2.5.48}$	$(\{\sigma_{j_2},\sigma_{j_3},\{\sigma_{j_5}\}_{j_4}\}_{j_1})$	6
$\mathbf{t}_{2.5.49}$	$(\{\sigma_{j_2},\{\sigma_{j_4},\sigma_{j_5}\}_{j_3}\}_{j_1})$	4	$\mathbf{t}_{2.5.50}$	$(\{\sigma_{j_2},\{\{\sigma_{j_5}\}_{j_4}\}_{j_3}\}_{j_1})$	4

$\mathbf{t}$	tree	α_Δ	$\mathbf{t}$	tree	α_Δ
$\mathbf{t}_{2.5.51}$	$(\{\{\sigma_{j_3}\}_{j_2}, \{\sigma_{j_5}\}_{j_4}\}_{j_1})$	3	$\mathbf{t}_{2.5.52}$	$(\{\{\sigma_{j_3}, \sigma_{j_4}, \sigma_{j_5}\}_{j_2}\}_{j_1})$	1
$\mathbf{t}_{2.5.53}$	$(\{\{\sigma_{j_3}, \{\sigma_{j_5}\}_{j_4}\}_{j_2}\}_{j_1})$	3	$\mathbf{t}_{2.5.54}$	$(\{\{\{\sigma_{j_4}, \sigma_{j_5}\}_{j_3}\}_{j_2}\}_{j_1})$	1
$\mathbf{t}_{2.5.55}$	$(\{\{\{\{\sigma_{j_5}\}_{j_4}\}_{j_3}\}_{j_2}\}_{j_1})$	1			

References

[1] K. Burrage and P. M. Burrage, High strong order explicit Runge-Kutta methods for stochastic ordinary differential equations, *Appl. Numer. Math.*, **22**, No. 1-3, (1996) 81–101.

[2] K. Burrage and P. M. Burrage, Order conditions of stochastic Runge-Kutta methods by B-series, *SIAM J. Numer. Anal.*, **38**, No. 5, (2000) 1626–1646.

[3] J. C. Butcher, *Numerical methods for ordinary differential equations* (John Wiley & Sons, Chichester, 2003).

[4] E. Hairer, Order Conditions for Numerical Methods for Partitioned Ordinary Differential Equations *Numer. Math.* **36**, (1981) 431–445.

[5] E. Hairer and S. P. Nørsett and G. Wanner, Solving Ordinary Differential Equations I (Springer-Verlag, Berlin, 1993).

[6] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus* (Springer New-York, Berlin, 1999).

[7] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Applications of Mathematics 23, Springer-Verlag, Berlin, 1999).

[8] Y. Komori and T. Mitsui and H. Sugiura, Rooted tree analysis of the order conditions of ROW-type scheme for stochastic differential equations, BIT, Vol. 37, No. 1, (1997) 43-66.

[9] J. Lehn, A. Rößler and O. Schein, Adaptive schemes for the numerical solution of SDEs - a comparison, *J. Comput. Appl. Math.*, **138**, No. 2, (2002) 297–308.

[10] V. Mackevicius and J. Navikas, Second-order weak Runge-Kutta type methods for Itô equations, Math. Comput. Simul., Vol. 57, (1–2), 29–34, 2001.

[11] G. N. Milstein, *Numerical Integration of Stochastic Differential Equations* (Kluwer Academic Publishers, Dordrecht, 1995).

[12] G. N. Milstein, Weak Approximation of Solutions of Systems of Stochastic Differential Equations, *Theory Probab. Appl.* **30** (1986) 750–766.

[13] N. J. Newton, Asymptotically efficient Runge-Kutta methods for a class of Itô and Stratonovich equations, SIAM J. Appl. Math., Vol. 51, No. 2, 542–567, 1991.

[14] A. Rößler, Embedded Stochastic Runge-Kutta Methods, *Proc. Appl. Math. Mech.*, Vol. 2, Iss. 1, 461-462, 2003.

[15] A. Rößler, Runge-Kutta methods for Stratonovich stochastic differential equation systems with commutative noise, *J. Comput. Appl. Math.*, **164-165** (2004) 613–627.

[16] A. Rößler, Runge-Kutta Methods for the Numerical Solution of Stochastic Differential Equations, Ph.D. thesis, Darmstadt University of Technology. (Shaker Verlag, Aachen, 2003).

[17] A. Rößler, Stochastic Taylor Expansions for the Expectation of Functionals of Diffusion Processes, *To appear in: Stochastic Anal. Appl.*

[18] W. Rümelin, Numerical treatment of stochastic differential equations, SIAM J. Numer. Anal., Vol. 19, No. 3, 604–613, 1982.

[19] D. Talay, Second-order discretization schemes of stochastic differential systems for the computation of the invariant law, Stochastics and Stochastic Rep., Vol. 29, No. 1, 13–36, 1990.

[20] A. Tocino and J. Vigo-Aguiar, Weak second order conditions for stochastic Runge-Kutta methods, SIAM J. Sci. Comput., Vol. 24, No. 2, 507–523, 2002.