Accepted Manuscript

High-order numerical schemes for jump-SDEs

Alexander Grigo

PII: S0377-0427(18)30749-0
DOI: https://doi.org/10.1016/j.cam.2018.12.018
Reference: CAM 12061

To appear in: *Journal of Computational and Applied Mathematics*

Received date : 13 February 2018
Revised date : 10 December 2018

Please cite this article as: A. Grigo, High-order numerical schemes for jump-SDEs, *Journal of Computational and Applied Mathematics* (2018), https://doi.org/10.1016/j.cam.2018.12.018

High-Order Numerical Schemes for Jump SDEs

Alexander Grigo[a]

[a] *Department of Mathematics, University of Oklahoma, 72019 Norman OK, USA*

Abstract

In this paper we propose an algorithm to numerically simulate Markov processes of jump type. While these processes can naturally be generated as solutions to jump-SDEs the algorithm we propose is instead based on the interlacing construction of the process. We show that one can construct in the sense of strong convergence a high order numerical scheme based on high order ODE solvers. This result is in sharp contrast to the well known difficulty of constructing high-order numerical schemes for diffusion processes.

Keywords: discretization scheme, jump process, strong order of convergence

1. Introduction

In this paper we describe an efficient numerical algorithm to simulate the continuous time Markov process X in $\mathbb{R}^d$ determined by its infinitesimal generator $\mathcal{G}$ given by

$$\mathcal{G}f(x) = \nabla_{v(x)} f(x) + v(x) + \int \Lambda(x)\,\Gamma(x, dx')\,[f(x') - f(x)] \tag{1}$$

for all $f \in C_c^\infty$. If the vector field v is globally Lipschitz continuous (and therefore satisfies a linear growth condition), the transition (probability) kernel $\Gamma(x, dy)$ has finite first and second moments, and the rate function $\Lambda(x)$ is bounded from above, then it is known, e.g. Kolokoltsov [1, 2], that the martingale problem corresponding to $\mathcal{G}$ has a solution for any initial distribution. In order to formulate our main result we will impose more stringent regularity conditions, so the existence of X_t is always guaranteed.

One method of simulating trajectories of X_t is based on what is commonly referred to as thinning, where the times at which the jumps occur are oversampled by a Poisson process with intensity at least as large as $\sup_x \Lambda(x)$. And at every instance of those predetermined times of a potential jump a jump actually occurs with an appropriately chosen probability and distribution. Details

Email address: grigo@math.ou.edu (Alexander Grigo)

[1] The author gratefully acknowledges the support by the NSF grant DMS-1413428.

of this approach can be found in section 4.2 of Ethier and Kurtz [3], and how this method can be used to construct the process X_t, in the first place.

Another method to simulate the process X_t is as follows. Using the forward Euler method to define the meaning of a stochastic differential equation, as is customarily done e.g. Ikeda and Watanabe [4], Protter [5], Bichteler [6], one can define the process X_t as the solution to the stochastic differential equation (SDE)

$$d\mathsf{X}_t = v(\mathsf{X}_{t-})\,dt + d\mathsf{Y}_t(\mathsf{X}_{t-}) \tag{2}$$

where $\mathsf{Y}_t(x)$ denotes the pure-jump Levy process induced by the measure $\nu(x,.)$, where the measure $\nu(x,.)$ is determined by the requirement $\int \nu(x,dz)\,f(x+z) = \int \Lambda(x)\,\Gamma(x,dx')\,f(x')$ for all $f \in C_c^\infty$. Details to this approach can be found in Kolokoltsov [1, 2]. As with the thinning method, the SDE representation of the process X_t provides a coupling of all trajectories of X_t corresponding to different initial data, and hence allows for a path-wise analysis of X_t. The very definition of the meaning of the SDE (2) as limits of the forward Euler method readily lends itself to a straightforward numerical scheme for simulating trajectories X_t, just as the thinning method. In fact, the Euler method is a well-known and frequently used method in simulating SDEs with or without jumps, e.g. Milstein [7], Milstein and Tretyakov [8], Kloeden and Platen [9], Bichteler [6].

The approach we present in this paper is different from both of the above mentioned methods. It is based on the following observation. Let $\varphi_t(x)$ denote the solution of the initial value problem $\dot z = v(z)$, $z(0) = x$, which generates a globally defined flow in $\mathbb{R}^d$. Then the identity

$$\begin{aligned}\mathsf{E}[f(\mathsf{X}_t)\,|\,\mathsf{X}_0 = x] &= \exp\Big\{-\int_0^t \Lambda(\varphi_s(x))\,ds\Big\}\,f(\varphi_t(x)) + \\ &+ \int_0^t \exp\Big\{-\int_0^r \Lambda(\varphi_s(x))\,ds\Big\}\,\Lambda(\varphi_r(x))\times \\ &\qquad\times \int \Gamma(\varphi_r(x),dx')\;\mathsf{E}[f(\mathsf{X}_t)\,|\,\mathsf{X}_r = x']\,dr\;,\end{aligned}$$

which holds for any $f \in C_c^\infty$, allows for the following construction of X_t. It is well known, e.g. Kallenberg [10], that the transition kernel $\Gamma(x,dx')$ can always be represented as $\int \Gamma(x,dx')\,f(x') = \int \mu(d\xi)\,f(\gamma(\xi,x))$ for any $f \in C_c^\infty$ and $x \in \mathbb{R}^d$, where μ is a suitably chosen distribution on some auxiliary space Ξ, and suitable map $\gamma\colon \Xi\times\mathbb{R}^d \to \mathbb{R}^d$. In complete generality one can choose μ to be the uniform distribution on $\Xi = [0,1]$, but in practice other choices are sometimes favorable, and we will make use of this below. Then for any sequence $(\theta_n)_{n\in\mathbb{N}}$ of independent standard exponential random variables, and for any sequence $(\xi_n)_{n\in\mathbb{N}}$ of independent Ξ–valued random variables with common distribution μ, and for any initial value $x \in \mathbb{R}^d$ we construct the corresponding path of X_t by the interlacing procedure: Set the initial values

$$\tau_0 = 0\;, \quad \mathsf{X}_0 = x \tag{3a}$$

and supposing that X_t is already defined for $t \le \tau_n$ we continue defining X_t by

$$\begin{aligned} \tau_{n+1} &= \inf\Big\{ t > \tau_n : \int_0^t \Lambda(\varphi_s(\mathsf{X}_{\tau_n}))\, ds > \theta_{n+1} \Big\} \\ \mathsf{X}_t &= \varphi_{t-\tau_n}(\mathsf{X}_{\tau_n}) \quad \text{for all} \quad \tau_n < t < \tau_{n+1} \\ \mathsf{X}_{\tau_{n+1}} &= \gamma(\xi_n, \varphi_{\tau_{n+1}-\tau_n}(\mathsf{X}_{\tau_n})) \quad \text{for} \quad t = \tau_{n+1} \end{aligned} \tag{3b}$$

and then repeating this process recursively. This process of interlacing not only provides an explicit coupling for all trajectories of X_t corresponding to all the different initial data X_0. It also provides a completely transparent description of X_t, which readily lends itself to modeling various problems where a deterministic evolution is intertwined with randomly occurring sudden changes. For instance, in Leite et al. [11], Broda et al. [12] this model was used to model the effect of random (sudden) events in population dynamics, and Xue and Othmer [13] for models of chemotaxis.

Furthermore, the explicit coupling construction provided by the interlacing procedure (3) readily indicates a numerical scheme to simulate the process X_t that (a) avoids unnecessary oversampling of the jump times as in the thinning method, and (b) allows for employing, as will be shown in the subsequent part of this paper, higher-order methods unlike the Euler–scheme used in the definition of the SDE representation of X_t. Namely, we simulate X_t using an efficient high-order numerical scheme, e.g. Hairer et al. [14], Hairer and Wanner [15], Stoer and Bulirsch [16], for the ordinary differential equation (ODE) $\dot{z} = v(z)$ between the jump times $(\tau_n)_{n\in\mathbb{N}}$, and perform the jumps at the jump times. The fact that despite the presence of the jumps this approach indeed yields efficient numerical schemes of arbitrary order to path-wise simulate X_t is the main result of this paper. This is in sharp contrast to corresponding results for simulating diffusions or jump-diffusions, where it is known to be very difficult to construct higher-order numerical schemes, e.g. Milstein [7], Milstein and Tretyakov [8], Kloeden and Platen [9], Platen [6].

In Section 2 we describe the ODE solver we admit in our algorithm. The actual algorithm and the main convergence results are stated in Section 3. Final remarks are contained in Section 4.

2. Description of the ODE solver

The interlacing procedure (3) shows that except for the jumps the process X_t is the solution to the ODE $\dot{z} = v(z)$. Therefore, we describe a few key aspects of what kind of algorithm we consider to solve ODEs. It will, however, turn out that the ODE we are interested in is not simply $\dot{z} = v(z)$, which is the reason why in this section we consider a generic initial value problem

$$\dot{y}(t) = g(y(t)) \,, \quad y(0) = y_0 \in \mathbb{R}^l \,, \tag{4}$$

where the specific assumptions we impose on the function g are given below. In principle we could allow for an explicit time-dependence, but since we restrict ourself to time-homogeneous processes we have no need for this slightly generalized setup.

There is a vast literature of how to numerically approximate the solution to (4) efficiently, e.g. Hairer et al. [14], Butcher [17], Stoer and Bulirsch [16]. In order to explain the structure of the numerical schemes we will consider in this paper we start with an analysis of Runge-Kutta methods. These methods are well-known and extensively used in practice, and we refer the reader to Hairer et al. [14], Butcher [17], Stoer and Bulirsch [16] for details.

Runge-Kutta methods are so-called one-step methods which partition starts with a sequence $0 = t_0 < t_1 < t_2 < \dots$ of points in time at which a numerical approximation $\hat{y}$ to $y(t)$ is produced according to the one-step recurrence relation

$$\hat{y}(t_{k+1}) = \hat{y}(t_k) + h_k \, \Phi(\hat{y}(t_k), h_k) \quad \text{with} \quad h_k = t_{k+1} - t_k \tag{5}$$

for $k = 0, 1, \dots$ for a given initial value $\hat{y}(0)$ which may or may not be equal to $y(0)$. The key ingredient in this kind of method is, of course, the choice of the function $\Phi(y, h)$. For Runge-Kutta methods there is a specific way to construct $\Phi(y, h)$, which will not be of concern to us. What will matter to us, however, is that it is possible, e.g. Hairer et al. [14], Butcher [17], Stoer and Bulirsch [16], to construct Φ such that for some $\bar{L} > 0$, $\bar{C} > 0$, $\bar{h} > 0$, $p = 1, 2, \dots$, $q = 0, 1, 2, \dots$ we have

$$\begin{aligned} \| \, \Phi(y_1, h) - \Phi(y_2, h) \, \| &\leq \bar{L} \, \| \, y_1 - y_2 \, \| \\ \| \, y(h) - y(0) - h \, \Phi(y(0), h) \, \| &< C \, \| \, g(y(0)) \, \| \, (1 \vee \| \, g(y(0)) \, \|^q) \, h^{p+1} \end{aligned} \tag{6}$$

for all $0 < h \leq \bar{h}$ and $y(0), y_1, y_2 \in \mathbb{R}^l$. The left-hand-side is usually referred to as local truncation error and p is called the order of the method. A sufficient condition for this type of local truncation error estimate is the following result, e.g. Hairer et al. [14], Butcher [17], Stoer and Bulirsch [16].

Lemma 2.1 (Local truncation error for Runge-Kutta methods, e.g. Hairer et al. [14], Butcher [17], Stoer and Bulirsch [16])**.** *If all derivatives of g up to and including order $p - 1$ are uniformly bounded, then p–th order Runge–Kutta methods satisfy the local truncation error estimate* (6) *with $q = p - 1$.*

By a minor variation of the proofs of the error estimate found in Hairer et al. [14], Butcher [17], Stoer and Bulirsch [16] we obtain the following global error estimate

Lemma 2.2 (Global truncation error)**.** *Suppose the first derivative of g is uniformly bounded by L, i.e. $\| \, Dg(y) \, \| \leq L$ for all $y \in \mathbb{R}^l$, and suppose further that Φ satisfies the local truncation error estimate* (6)*. Then*

$$\begin{aligned} \| \, y(t_k) - \hat{y}(t_k) \, \| \leq e^{t_k \, \bar{L}} \, \| \, y(0) - \hat{y}(0) \, \| + \bar{C} \, e^{t_k \, (\bar{L} \vee L)} \, \frac{e^{q \, L \, t_k} - 1}{q \, L} \times \\ \times \, \| \, g(y(0)) \, \| \, (1 \vee \| \, g(y(0)) \, \|^q) \left[\max_{0 \leq l \leq k-1} h_l \right]^p \end{aligned} \tag{7}$$

for all $k = 1, 2, \dots$.

Proof. Let for any $y \in \mathbb{R}^l$ denote by $Y(t,y)$ the solution of the initial value problem $\dot{Y} = g(Y)$, $Y(0) = y$. By definition of $y(t)$ and $\hat{y}(t_k)$ we have

$$\begin{aligned} y(t_{k+1}) - \hat{y}(t_{k+1}) = y(t_k) - \hat{y}(t_k) + h_k \left[\Phi(y(t_k), h_k) - \Phi(\hat{y}(t_k), h_k)\right] \\ + y(t_{k+1}) - y(t_k) - h_k \, \Phi(y(t_k), h_k) \end{aligned}$$

so that the assumed local truncation error estimate (6) implies

$$\begin{aligned} \| \, y(t_{n+1}) - \hat{y}(t_{n+1}) \, \| \le (1 + h_k \, \bar{L}) \, \| \, y(t_k) - \hat{y}(t_k) \, \| \\ + \bar{C} \, \| \, g(y(t_k)) \, \| \, (1 \vee \| \, g(y(t_k)) \, \|^q) \, h_k^{p+1} \end{aligned}$$

for all $k = 0, 1, \ldots$. With Gronwall's inequality it follows that

$$\begin{aligned} \| \, y(t_k) - \hat{y}(t_k) \, \| \le e^{t_k \, \bar{L}} \, \| \, y(0) - \hat{y}(0) \, \| \\ + \bar{C} \sum_{l=0}^{k-1} e^{(t_k - t_{l+1}) \, \bar{L}} \, \| \, g(y(t_l)) \, \| \, (1 \vee \| \, g(y(t_l)) \, \|^q) \, h_l^{p+1} \end{aligned}$$

for all $k = 0, 1, \ldots$.

The final observation is that the above estimate involves only the true solution $y(t)$, sampled at $t = t_k$. And since

$$\begin{aligned} \frac{d}{dt} \| \, g(y(t)) \, \|^2 = 2 \, g(y(t))^T \, Dg(y(t)) \, g(y(t)) \le 2 \, \| \, Dg(y(t)) \, \| \, \| \, g(y(t)) \, \|^2 \\ \le 2 \, L \, \| \, g(y(t)) \, \|^2 \end{aligned}$$

it follows again from Gronwall's inequality that

$$\| \, g(y(t)) \, \|^2 \le \| \, g(y(0)) \, \|^2 \, e^{2 \, L \, t} \quad \text{i.e.} \quad \| \, g(y(t)) \, \| \le \| \, g(y(0)) \, \| \, e^{L \, t}$$

for all $t \ge 0$. Upon substituting this estimate for $\| \, g(y(t)) \, \|$ in the above estimate for $\| \, y(t_k) - \hat{y}(t_k) \, \|$ yields

$$\begin{aligned} \| \, y(t_k) - \hat{y}(t_k) \, \| &\le e^{t_k \, \bar{L}} \, \| \, y(0) - \hat{y}(0) \, \| \\ &\quad + \bar{C} \sum_{l=0}^{k-1} e^{(t_k - t_{l+1}) \, \bar{L}} \, \| \, g(y(0)) \, \| \, e^{L \, t_l} \, (1 \vee \| \, g(y(t_l)) \, \|^q) \, h_l^{p+1} \\ &\le e^{t_k \, \bar{L}} \, \| \, y(0) - \hat{y}(0) \, \| \\ &\quad + \bar{C} \, e^{t_k \, (\bar{L} \vee L)} \, \| \, g(y(0)) \, \| \sum_{l=0}^{k-1} (1 \vee \| \, g(y(t_l)) \, \|^q) \, h_l^{p+1} \\ &\le e^{t_k \, \bar{L}} \, \| \, y(0) - \hat{y}(0) \, \| \\ &\quad + \bar{C} \, e^{t_k \, (\bar{L} \vee L)} \, \| \, g(y(0)) \, \| \, (1 \vee \| \, g(y(0)) \, \|^q) \sum_{l=0}^{k-1} e^{q \, L \, t_l} \, h_l^{p+1} \end{aligned}$$

for all $k = 1, 2, \ldots$. With

$$\sum_{l=0}^{k-1} e^{q\,L\,t_l}\, h_l^{p+1} \le \Big[\max_{0\le l\le k-1} h_l\Big]^p \sum_{l=0}^{k-1} e^{q\,L\,t_l}\, h_l \le \Big[\max_{0\le l\le k-1} h_l\Big]^p \int_0^{t_k} e^{q\,L\,s}\, ds$$

we obtain the claimed estimate on $\| y(t_k) - \hat{y}(t_k) \|$. □

The Runge-Kutta methods, as any other one-step method, generates numerical approximations to $y(t)$ only at the prescribed (or preselected) times $0 = t_0 < t_1 < t_2 < \ldots$, for which the local truncation error estimate (6), provides a global truncation estimate (7). In the construction of Runge-Kutta methods, i.e. the construction of Φ, one can Hairer et al. [14] also efficiently produce an interpolation $\hat{y}(t)$ of $\hat{y}(t_k)$, $\hat{y}(t_{k+1})$ for any $t_k \le t \le t_{k+1}$ that satisfies the same local truncation error estimate for $\| y(t) - \hat{y}(t) \|$ as in (6). This is usually referred to as dense output, and an efficient Runge-Kutta implementation of this is called a continuous Runge-Kutta method, see Hairer et al. [14] for details.

Of course, there are other numerical methods to solve the initial value problem (4), for example multi-step methods or adaptive methods. We refer the reader to Hairer et al. [14], Butcher [17], Stoer and Bulirsch [16] and references therein for an analysis of various methods. What matters to us in this paper is simply the fact that one can construct efficient algorithms to produce an approximation $\hat{y}(t)$ that agrees with $y(t)$ for all t up to an explicit error bound as in (7). Therefore we will consider numerical methods that satisfy the following properties:

$$\begin{aligned} 0 = t_0 &< t_1 < \ldots \\ h_k = t_{k+1} - t_k &\le h \quad \text{for all } n = 0, 1, \ldots \\ \| y(t) - \hat{y}(t) \| &\le e^{t\,A_1} \,\| y(0) - \hat{y}(0) \| \\ &\quad + A_2\, t\, e^{t\,A_3} \,\| g(y(0)) \|\, (1 \vee \| g(y(0)) \|^q)\, h^p \end{aligned} \tag{8}$$

where h is a given parameter controlling the precision of the algorithm (the maximal step size to be used), $p, q \in \mathbb{N}$ are describing the order property of the algorithm, and $A_1, A_2, A_3 \ge 0$ are some fixed constants.

As mentioned, continuous Runge-Kutta methods are of this type, including those that adaptively choose the step size.

3 Coupling of X_t and its numerical approximation $\tilde{\mathsf{X}}_t$

Let us fix the notation $\tilde{\mathsf{X}}_t$ for any (numerical or not) approximation of X_t. In order to obtain path-wise estimates on the difference between X_t and $\tilde{\mathsf{X}}_t$ two principal steps need to be carried out. Firstly, both processes X_t, $\tilde{\mathsf{X}}_t$ must be constructed on the same probability space, for otherwise they cannot be compared at all. Secondly, to quantify the comparison of the paths of X_t, $\tilde{\mathsf{X}}_t$ for some range of t a suitable metric has to be agreed upon. The first issue will

be addressed by an explicit simultaneous construction of X_t and $\tilde{\mathsf{X}}_t$, and will be made clear in the description of Algorithm 3.1 below. So we start this section explaining the metric we use to compare X and $\tilde{\mathsf{X}}$.

3.1. Introduction of notation and basic definitions

Let $T > 0$ denote some fixed time that describes the time interval $[0, T]$ on which we attempt to (construct and) compare the paths of X and $\tilde{\mathsf{X}}$. Since X and $\tilde{\mathsf{X}}$ are constructed below to have paths that have left limits and are continuous from the right (cadlag paths) the natural choice for the space of paths is the Skorohod space $D([0, T], \mathbb{R}^d)$ consisting of all cadlag paths. Details can be found in many standard textbooks, e.g. Ethier and Kurtz [5], Jacod and Shiryaev [18], and we will simply provide here a description of the usual Prohorov metric ϱ_P on $D([0, T], \mathbb{R}^d)$, which makes it into a complete separable metric space whose Borel sigma algebra is the one generated by the evaluation maps. Namely, for any two paths $x, \tilde{x} \in D([0, T], \mathbb{R}^d)$ their distance $\varrho_P(x, \tilde{x})$ is defined as

$$\varrho_P(x, \tilde{x}) = \inf_{\lambda} \left\{ \sup_{0 \le s < t \le T} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right| + 1 \wedge \sup_{0 \le t \le T} \| x(\lambda(t)) - \tilde{x}(t) \| \right\} , \quad (9)$$

where the infimum is taken over all time changes λ, i.e. $\lambda \colon [0, T] \to [0, T]$ increasing continuous bijections. A more tractable metric is the so-called Skorohod metric

$$\varrho_S(x, \tilde{x}) = \inf_{\lambda} \left\{ \sup_{0 \le s < t \le T} |\lambda(t) - t| + 1 \wedge \| x(\lambda(t)) - \tilde{x}(t) \| \right\} . \quad (10)$$

The metric ϱ_S generates the same topology as ϱ_P, but lacks the completeness property. This, however, will not be of concern to us.

It is standard practice to interpret the process X_t restricted to $t \in [0, T]$ as a random variable with values in the complete metric space $D([0, T], \mathbb{R}^d)$. This point of view is useful when defining the sense in which X and its approximation $\tilde{\mathsf{X}}$ are to be compared path-wise.

3.2. The numerical algorithm

In this section we describe the algorithm for constructing $\tilde{\mathsf{X}}$. A key result will be that this construction also provides a coupling with X, which is crucial for the comparison. Recall that the interlacing procedure (3) provides an explicit construction of X as solutions to the ODE $\dot{x} = v(x)$ on random time intervals $[\tau_n, \tau_{n+1})$ which are connected by random jumps. However, since the distribution of $\tau_{n+1} - \tau_n$ depends on the path of X_t we will not only solve the ODE $\dot{x} = v(x)$, but rather do we consider the ODE

$$z = (x, y) \quad w(x, y) = (v(x), \Lambda(x)) , \quad \dot{z} = w(z) , \quad (11)$$

whose solution we denote by $\psi_t(z)$. Furthermore, let

$$\tilde{z}(t) = \tilde{\psi}_t(z) , \quad 0 = \tilde{t}_0(z) < \tilde{t}_1(z) < \dots \quad (12)$$

denote the numerical approximation (not necessarily a flow) of $\psi_t(\tilde{z})$ with values algorithmically generated at times $(\tilde{t}_k(z))_{k\in\mathbb{N}}$ using an ODE solver satisfying our assumption (8). With this notation we can describe an extended interlacing algorithm, recall (3), to simultaneously construct X and $\tilde{\mathsf{X}}$:

Algorithm 3.1 (Extended interlacing procedure to construct X, $\tilde{\mathsf{X}}$)**.**

1. *Fix the random input* $(\theta_n)_{n\in\mathbb{N}}$, $(\xi_n)_{n\in\mathbb{N}}$, *and fix the initial values* $\mathsf{X}_0 = \tilde{\mathsf{X}}_0 \in \mathbb{R}^d$, $\tilde{y}_0 = 0$, $\tau_0 = \tilde{\tau}_0 = 0$. *Furthermore, fix* $h_\star > 0$ *and* $\varepsilon_\star > 0$.
2. *Suppose that for some* $n \in \mathbb{N}$ *the jump times* $(\tau_i)_{i=0,\dots,n}$, $(\tilde{\tau}_i)_{i=0,\dots,n}$ *and* $(\mathsf{X}_t)_{0\le t\le\tau_n}$, $(\tilde{\mathsf{X}}_t)_{0\le t\le\tilde{\tau}_n}$ *have been constructed.*
   (a) *Define* τ_{n+1} *through*

$$\tau_{n+1} = \inf\Big\{t > \tau_n : \int_{\tau_n}^{t} \lambda(\varphi_s(\mathsf{X}_{\tau_n}))\, ds > \theta_n\Big\}$$

   and define X_t *via*

$$\begin{aligned}\mathsf{X}_t &= \varphi_{t-\tau_n}(\mathsf{X}_{\tau_n}) \quad \text{for all} \quad \tau_n < t < \tau_{n+1}\\ \mathsf{X}_{\tau_{n+1}} &= \gamma(\xi_n, \varphi_{\tau_{n+1}-\tau_n}(\mathsf{X}_{\tau_n}))\end{aligned}$$

   for all $\tau_n < t \le \tau_{n+1}$.
   (b) *Use an ODE solver satisfying our assumption* (8) *to construct for* $h = h_\star$

$$(\tilde{x}(t), \tilde{y}(t)) = \tilde{\psi}_t(\tilde{z}_n)\,, \quad 0 = \tilde{t}_0(\tilde{z}_n) < \tilde{t}_1(\tilde{z}_n) < \dots$$

   for the initial value $\tilde{z}_n = (\tilde{\mathsf{X}}_{\tau_n}, 0)$, *recall* (12) *for the notation used. Let* $N \in \mathbb{N}$ *be*

$$N = \inf\{k \in \mathbb{N} : \tilde{y}(\tilde{t}_{k+1}) > \theta_n\}$$

   and use the dense output of the ODE solver to determine a value $\tilde{t}_N \le \tilde{\Delta}_n < \tilde{t}_{N+1}$ *such that*

$$|\tilde{y}(\tilde{\Delta}_n) - \theta_n| \le \varepsilon_\star\,.$$

   Then set

$$\tilde{\tau}_{n+1} = \tilde{\tau}_n + \tilde{\Delta}_n$$

   and define $\tilde{\mathsf{X}}_t$ *via*

$$\begin{aligned}\tilde{\mathsf{X}}_t &= \tilde{x}(t - \tilde{\tau}_n) \quad \text{for all} \quad \tilde{\tau}_n < t < \tilde{\tau}_{n+1}\\ \tilde{\mathsf{X}}_{\tilde{\tau}_{n+1}} &= \gamma(\xi_n, \tilde{x}(\tilde{\Delta}_n))\end{aligned}$$

   for all $\tilde{\tau}_n < t \le \tilde{\tau}_{n+1}$.
3. *Repeatedly applying step 2 the processes* X_t *and* $\tilde{\mathsf{X}}_t$ *can be defined recursively until both are defined on an interval containing* $[0,T]$. *Then restrict* t *to* $[0,T]$ *to obtain two paths* $(\mathsf{X}_t)_{t\in[0,T]}$, $(\tilde{\mathsf{X}}_t)_{t\in[0,T]}$ *in* $D([0,T],\mathbb{R}^d)$.

3.3. A priori estimates

For any random input $(\theta_n)_{n\in\mathbb{N}}$, $(\xi_n)_{n\in\mathbb{N}}$, and any initial value $\mathsf{X}_0 = \tilde{\mathsf{X}}_0 \in \mathbb{R}^d$ Algorithm 3.1 produces a pair of paths $(\mathsf{X}_t, \tilde{\mathsf{X}}_t)$ and a sequence of jump times $(\tau_n, \tilde{\tau}_n)_{n\in\mathbb{N}}$. Our standing assumption on the ODE solver used to produce the numerical approximation is that it satisfies the truncation error estimate given in (8).

Using the notation as in Algorithm 3.1 we define the time change map

$$\lambda(\tilde{t}) = \tau_{n+1}\,\frac{\tilde{t}-\tilde{\tau}_n}{\tilde{\tau}_{n+1}-\tilde{\tau}_n} + \tau_n\,\frac{\tilde{\tau}_{n+1}-\tilde{t}}{\tilde{\tau}_{n+1}-\tilde{\tau}_n} \quad \text{for all } \tilde{\tau}_n \le \tilde{t} \le \tilde{\tau}_{n+1}\,, \tag{13}$$

which maps $[\tilde{\tau}_n, \tilde{\tau}_{n+1}]$ bijectively onto $[\tau_n, \tau_{n+1}]$ for all $n \in \mathbb{N}$ (or rather for all n for which Algorithm 3.1 produces τ_n and $\tilde{\tau}_n$).

Lemma 3.2 (Single step a priori error estimate) *Suppose that $0 < \Lambda_- \le \Lambda(x)$ for all x, and suppose that Λ is globally Lipschitz with Lipschitz constant L_Λ. Further, suppose that v is globally Lipschitz with Lipschitz constant L_v, and suppose also that $\| \gamma(\xi,x) - \gamma(\xi,x') \| \le L_\gamma \| x - x' \|$ for all x, x' and all ξ. Given the values of τ_n, $\tilde{\tau}_n$, X_{τ_n}, $\tilde{\mathsf{X}}_{\tilde{\tau}_n}$, $h_\star$, $\varepsilon_\star$ it follows that*

$$\begin{aligned}|\tau_{n+1} - \tau_n - \tilde{\tau}_{n+1} + \tilde{\tau}_n| \le{}& \frac{1}{\Lambda_-}\, e^{(\tilde{\tau}_{n+1}-\tilde{\tau}_n)\,A_1} \,\left\| \mathsf{X}_{\tau_n} - \tilde{\mathsf{X}}_{\tilde{\tau}_n} \right\| \\ &+ \frac{A_2}{\Lambda_-}\,(\tilde{\tau}_{n+1}-\tilde{\tau}_n)\, e^{(\tilde{\tau}_{n+1}-\tilde{\tau}_n)\,A_3} \times \\ &\times \| w(\mathsf{X}_{\tau_n},0) \|\, (1 \vee \| w(\mathsf{X}_{\tau_n},0) \|^q)\, h_\star^p\end{aligned}$$

and

$$\begin{aligned}\frac{1}{1\vee L_\gamma} \sup_{\tilde{\tau}_n \le \tilde{t} \le \tilde{\tau}_{n+1}} \left\| \mathsf{X}_{\lambda(\tilde{t})} - \tilde{\mathsf{X}}_{\tilde{t}} \right\| \le{}& \frac{|e^{L_v\,(\tau_{n+1}-\tau_n)} - e^{L_v\,(\tilde{\tau}_{n+1}-\tilde{\tau}_n)}|}{L_v}\, \| v(\mathsf{X}_{\tau_n}) \| \\ &+ e^{(\tilde{\tau}_{n+1}-\tilde{\tau}_n)\,A_1} \,\left\| \mathsf{X}_{\tau_n} - \tilde{\mathsf{X}}_{\tilde{\tau}_n} \right\| \\ &+ A_2\,(\tilde{\tau}_{n+1}-\tilde{\tau}_n)\, e^{(\tilde{\tau}_{n+1}-\tilde{\tau}_n)\,A_3} \times \\ &\times \| w(\mathsf{X}_{\tau_n},0) \|\, (1 \vee \| w(\mathsf{X}_{\tau_n},0) \|^q)\, h_\star^p\,.\end{aligned}$$

Proof. Recall that $\tilde{\mathsf{X}}_t$ is constructed by numerically approximating the initial value problem $\dot{z} = w(z)$ initial value $z(\tilde{\tau}_n) = (\tilde{\mathsf{X}}_{\tilde{\tau}_n}, 0)$, call the resulting numerical approximation $\tilde{z}(t)$ for all $t \ge \tilde{\tau}_n$ (and not just for $t \le T$), and then restricting t to some appropriate time interval. The process X_t is chosen as the exact solution to the same differential equation, except with initial data $(\mathsf{X}_{\tau_n}, 0)$, restricted to some appropriately chosen time interval. Denote the solution valid for all t by simply $z(t)$. By assumption on the ODE solver used to produce the numerical approximation satisfies the truncation error estimate given in (8)

$$\begin{aligned}\| z(t+\tau_n) - \tilde{z}(t+\tilde{\tau}_n) \| \le{}& e^{t\,A_1}\, \| z(\tau_n) - \tilde{z}(\tilde{\tau}_n) \| \\ &+ A_2\, t\, e^{t\,A_3}\, \| w(z(\tau_n)) \|\, (1 \vee \| w(z(\tau_n)) \|^q)\, h_\star^p\end{aligned}$$

for all $t \geq 0$. Since $\| z(\tau_n) - \tilde{z}(\tilde{\tau}_n) \| = \left\| \mathsf{X}_{\tau_n} - \tilde{\mathsf{X}}_{\tilde{\tau}_n} \right\|$ it follows that

$$\begin{aligned}
|y(\tau_{n+1}) - y(\tilde{\tau}_{n+1} - \tilde{\tau}_n + \tau_n)| &\leq |\theta_n - \tilde{y}(\tilde{\tau}_{n+1})| + |\tilde{y}(\tilde{\tau}_{n+1}) - y(\tilde{\tau}_{n+1} - \tilde{\tau}_n + \tau_n)| \\
&\leq \varepsilon_\star + e^{(\tilde{\tau}_{n+1} - \tilde{\tau}_n) A_1} \left\| \mathsf{X}_{\tau_n} - \tilde{\mathsf{X}}_{\tilde{\tau}_n} \right\| \\
&\quad + A_2 \, (\tilde{\tau}_{n+1} - \tilde{\tau}_n) \, e^{(\tilde{\tau}_{n+1} - \tilde{\tau}_n) A_3} \times \\
&\quad \times \| w(\mathsf{X}_{\tau_n}, 0) \| \, (1 \vee \| w(\mathsf{X}_{\tau_n}, 0) \|^q) \, h_\star^p \, .
\end{aligned}$$

From the uniform lower bound on Λ it follows that

$$|\tau_{n+1} - \tau_n - \tilde{\tau}_{n+1} + \tilde{\tau}_n| \leq \frac{1}{\Lambda_-} \, |y(\tau_{n+1}) - y(\tilde{\tau}_{n+1} - \tilde{\tau}_n + \tau_n)| \, ,$$

which thus proves the claimed bound on $|\tau_{n+1} - \tau_n - \tilde{\tau}_{n+1} + \tilde{\tau}_n|$.

Similarly, by assumed bound on the approximation error

$$\begin{aligned}
\| x(t + \tau_n) - \tilde{x}(t + \tilde{\tau}_n) \| &\leq e^{t \, A_1} \left\| \mathsf{X}_{\tau_n} - \tilde{\mathsf{X}}_{\tilde{\tau}_n} \right\| \\
&\quad + \varepsilon_\star \, e^{t \, A_3} \| w(\mathsf{X}_{\tau_n}, 0) \| \, (1 \vee \| w(\mathsf{X}_{\tau_n}, 0) \|^q) \, h_\star^p
\end{aligned}$$

for all $t \geq 0$. Hence

$$\begin{aligned}
\left\| x(\lambda(\tilde{t})) - \tilde{x}(\tilde{t}) \right\| &\leq \left\| x(\lambda(\tilde{t})) - x(\tilde{t} - \tilde{\tau}_n + \tau_n) \right\| \\
&\quad + \left\| x(\tilde{t} - \tilde{\tau}_n + \tau_n) - \tilde{x}(\tilde{t} - \tilde{\tau}_n + \tilde{\tau}_n) \right\| \\
&\leq \left\| x(\lambda(\tilde{t})) - x(\tilde{t} - \tilde{\tau}_n + \tau_n) \right\| + e^{(\tilde{\tau}_{n+1} - \tilde{\tau}_n) A_1} \left\| \mathsf{X}_{\tau_n} - \tilde{\mathsf{X}}_{\tilde{\tau}_n} \right\| \\
&\quad + A_2 \, (\tilde{\tau}_{n+1} - \tilde{\tau}_n) \, e^{(\tilde{\tau}_{n+1} - \tilde{\tau}_n) A_3} \times \\
&\quad \times \| w(\mathsf{X}_{\tau_n}, 0) \| \, (1 \vee \| w(\mathsf{X}_{\tau_n}, 0) \|^q) \, h_\star^p
\end{aligned}$$

for all $\tilde{\tau}_n \leq \tilde{t} \leq \tilde{\tau}_{n+1}$, where

$$\lambda(\tilde{t}) = \tau_{n+1} \, \frac{\tilde{t} - \tilde{\tau}_n}{\tilde{\tau}_{n+1} - \tilde{\tau}_n} + \tau_n \, \frac{\tilde{\tau}_{n+1} - \tilde{t}}{\tilde{\tau}_{n+1} - \tilde{\tau}_n} \, .$$

By a standard Gronwall estimate, similar to what was done in the proof of Lemma 2.2, it follows that

$$\begin{aligned}
\left\| x(\lambda(\tilde{t})) - x(\tilde{t} - \tilde{\tau}_n + \tau_n) \right\| &= \left\| x(\lambda(\tilde{t}) - \tau_n + \tau_n) - x(\tilde{t} - \tilde{\tau}_n + \tau_n) \right\| \\
&\leq \frac{|e^{L_v \, (\lambda(\tilde{t}) - \tau_n)} - e^{L_v \, (\tilde{t} - \tilde{\tau}_n)}|}{L_v} \, \| v(x(\tau_n)) \| \\
&= \frac{|e^{L_v \, \frac{\tau_{n+1} - \tau_n}{\tilde{\tau}_{n+1} - \tilde{\tau}_n} \, (\tilde{t} - \tilde{\tau}_n)} - e^{L_v \, (\tilde{t} - \tilde{\tau}_n)}|}{L_v} \, \| v(\mathsf{X}_{\tau_n}) \| \\
&\leq \frac{|e^{L_v \, (\tau_{n+1} - \tau_n)} - e^{L_v \, (\tilde{\tau}_{n+1} - \tilde{\tau}_n)}|}{L_v} \, \| v(\mathsf{X}_{\tau_n}) \|
\end{aligned}$$

for all $\tilde{\tau}_n \le \tilde{t} \le \tilde{\tau}_{n+1}$, and hence

$$\begin{aligned} \| x(\lambda(\tilde{t})) - \tilde{x}(\tilde{t}) \| \le{}& \frac{|e^{L_v(\tau_{n+1}-\tau_n)} - e^{L_v(\tilde{\tau}_{n+1}-\tilde{\tau}_n)}|}{L_v} \| v(\mathsf{X}_{\tau_n}) \| \\ &+ e^{(\tilde{\tau}_{n+1}-\tilde{\tau}_n) A_1} \left\| \mathsf{X}_{\tau_n} - \tilde{\mathsf{X}}_{\tilde{\tau}_n} \right\| \\ &+ A_2 (\tilde{\tau}_{n+1} - \tilde{\tau}_n) e^{(\tilde{\tau}_{n+1}-\tilde{\tau}_n) A_3} \times \\ &\quad \times \| w(\mathsf{X}_{\tau_n}, 0) \| (1 \vee \| w(\mathsf{X}_{\tau_n}, 0) \|^q) h_\star^p \end{aligned}$$

for all $\tilde{\tau}_n \le \tilde{t} \le \tilde{\tau}_{n+1}$. In particular, this estimate applies to

$$\left\| \mathsf{X}_{\tau_{n+1}^-} - \tilde{\mathsf{X}}_{\tilde{\tau}_{n+1}^-} \right\| = \| x(\lambda(\tilde{\tau}_{n+1}^-)) - \tilde{x}(\tilde{\tau}_{n+1}) \| \ .$$

The assumed uniform Lipschitz continuity of φ implies

$$\left\| \mathsf{X}_{\tau_{n+1}} - \tilde{\mathsf{X}}_{\tilde{\tau}_{n+1}} \right\| \le L_\varphi \left\| \mathsf{X}_{\tau_{n+1}^-} - \tilde{\mathsf{X}}_{\tilde{\tau}_{n+1}^-} \right\|$$

which can be combined with the previous estimate to obtain the claimed estimate on $\left\| \mathsf{X}_{\lambda(\tilde{t})} - \tilde{\mathsf{X}}_{\tilde{t}} \right\|$. □

Assuming a uniform upper bound on Λ, implies that the number of jumps of X_t for $0 \le t \le T$ has a tail distribution that can be estimated by an exponentially small (in the number of jumps) term, which is uniform in the choice of the initial value X_0. In particular, the number of jumps is almost surely finite. Therefore iterating the estimate provided in Lemma 3.2 for as many times as there are jumps we readily obtain the following

Theorem 3.3 (Almost sure convergence of $\mathcal{O}(H^p)$)**.** *Suppose the same assumptions as in Lemma 3.2 and suppose also that* $\sup_x \| w(x,0) \| \le M$ *for some* $M \ge 1$*. If* $\varepsilon_\star = \mathcal{O}(h_\star^p)$*, then Algorithm 3.1 constructs a coupling* $(\mathsf{X}_t, \tilde{\mathsf{X}}_t)$ *such that* $\varrho_P(\mathsf{X}_t, \tilde{\mathsf{X}}_t) = \mathcal{O}(H^p)$ *holds almost surely and for all initial values* $\mathsf{X}_0 = \tilde{\mathsf{X}}_0 \in \mathbb{R}^d$*. The order of convergence is uniform in the number of jumps of* X_t *within* $[0, T]$*.*

A slightly more careful analysis of the preceding argument provides an estimate of the average of ϱ_S.

Theorem 3.4 (Convergence in average)**.** *Under the same assumptions as in Theorem 3.3 it follows that for every* $j \in \mathbb{N}$

$$\mathsf{E}\, \varrho_S(\mathsf{X}, \tilde{\mathsf{X}}) \le (1+T) \frac{(T\,M)^{j+1}}{(j+1)!} + \mathcal{O}(H^p) \ .$$

Proof. Fix the initial condition $\mathsf{X}_0 = \tilde{\mathsf{X}}_0 \in \mathbb{R}^d$. Notice that the estimates provided by Lemma 3.2 are naturally set up to estimate $|\tau_{n+1} - \tilde{\tau}_{n+1}|$ and

$\sup_{\tilde\tau_n \le \tilde t \le \tilde\tau_{n+1}} \left\| \mathsf{X}_{\lambda(\tilde t)} - \tilde{\mathsf{X}}_{\tilde t} \right\|$ in terms of the corresponding quantities with n replaced by $n-1$. Let N denote the number of jumps of X in $[0,T]$. Then

$$\mathsf{Pr}\,[N > j] \le \mathsf{Pr}\left[\hat N > j\right]$$

where $\hat N$ is the number of jumps within $[0,T]$ of a Poisson process with rate M. And since $\hat N$ has a Poisson distribution with parameter $T\,M$ it follows that

$$\mathsf{Pr}\,[N > j] \le e^{-T\,M} \sum_{k=j+1}^{\infty} \frac{(T\,M)^k}{k!} \le \frac{(T\,M)^{j+1}}{(j+1)!}\,.$$

Since the worst case of the right-hand-side in the definition of ϱ_S as in (10) is given by $T+1$ we obtain the following estimate

$$\begin{aligned}
\mathsf{E}\,\varrho_S(\mathsf{X},\tilde{\mathsf{X}}) &= \mathsf{E}\,\varrho_S(\mathsf{X},\tilde{\mathsf{X}})\,\mathbf{1}\{N > j\} + \mathsf{E}\,\varrho_S(\mathsf{X},\tilde{\mathsf{X}})\,\mathbf{1}\{N \le j\} \\
&\le (1+T)\;\mathsf{Pr}\,[E_1 + \ldots + E_{j+1} \le M\,T] + \mathsf{E}\,\varrho_S(\mathsf{X},\tilde{\mathsf{X}})\,\mathbf{1}\{N \le j\} \\
&\le (1+T)\,\frac{(T\,M)^{j+1}}{(j+1)!} + \mathsf{E}\,\varrho_S(\mathsf{X},\tilde{\mathsf{X}})\,\mathbf{1}\{N \le j\}\;.
\end{aligned}$$

By Theorem 3.3 we have $\varrho_S(\mathsf{X},\tilde{\mathsf{X}}) = \mathcal{O}(H^p)$ uniformly as long as $N \le j$, which completes the proof. □

4. Conclusions

As should be clear from the strategy of proof for Theorem 3.3 and Theorem 3.4 the uniform upper bounds for Λ and v are just made to simplify the arguments. One can allow for linear growth bounds (which are already implied by the global Lipschitz assumption imposed on w) under bounds on the q–th moment of the transition kernel Γ in order to control the terms involving the q–th power of $\| w(\mathsf{X}_\cdot, 0) \|$, which is necessary when iterating the a priori estimate provided by Lemma 3.2.

Similarly, instead of fixing (i.e. conditioning on) the initial condition $\mathsf{X}_0 = \tilde{\mathsf{X}}_0 \in \mathbb{R}^d$ in Theorem 3.3 and Theorem 3.4 one can allow for a distribution of initial data, which would naturally lead to an estimate of the Wasserstein-Kantorovich distance of X and $\tilde{\mathsf{X}}$ when using the cost function ϱ_S.

Finally, as is clear from the interlacing construction of X, our Algorithm 3.1 combines an efficient numerical ODE solver with the least number of random variables necessary to construct a coupling of X_t and $\tilde{\mathsf{X}}_t$. In this sense the proposed algorithm is optimal.

[1] V. N. Kolokoltsov, On Markov processes with decomposable pseudo-differential generators, Stoch. Stoch. Rep. 76 (1) (2004) 1–44, ISSN 1045-1129, URL https://doi.org/10.1080/10451120410001661250.

[2] V. N. Kolokoltsov, The Lévy-Khintchine type operators with variable Lipschitz continuous coefficients generate linear or nonlinear Markov processes and semigroups, Probab. Theory Related Fields 151 (1-2) (2011) 95–123, ISSN 0178-8051, URL https://doi.org/10.1007/s00440-010-0293-8.

[3] S. N. Ethier, T. G. Kurtz, Markov processes, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, ISBN 0-471-08186-8, URL https://doi.org/10.1002/9780470316658, characterization and convergence, 1986.

[4] N. Ikeda, S. Watanabe, Stochastic differential equations and diffusion processes, vol. 24 of *North-Holland Mathematical Library*, North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, second edn., ISBN 0-444-87378-3, 1989.

[5] P. E. Protter, Stochastic integration and differential equations, vol. 21 of *Stochastic Modelling and Applied Probability*, Springer-Verlag, Berlin, ISBN 3-540-00313-4, URL https://doi.org/10.1007/978-3-662-10061-5, second edition. Version 2.1, Corrected third printing, 2005.

[6] K. Bichteler, Stochastic integration with jumps, vol. 89 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, Cambridge, ISBN 0-521-81129-5, URL https://doi.org/10.1017/CBO9780511549878, 2002.

[7] G. N. Milstein, Numerical integration of stochastic differential equations, vol. 313 of *Mathematics and its Applications*, Kluwer Academic Publishers Group, Dordrecht, ISBN 0-7923-3213-X, URL https://doi.org/10.1007/978-94-015-8455-5, translated and revised from the 1988 Russian original, 1995.

[8] G. N. Milstein, M. V. Tretyakov, Stochastic numerics for mathematical physics, Scientific Computation, Springer-Verlag, Berlin, ISBN 3-540-21110-1, URL https://doi.org/10.1007/978-3-662-10063-9, 2004.

[9] P. E. Kloeden, E. Platen, Numerical solution of stochastic differential equations, vol. 23 of *Applications of Mathematics (New York)*, Springer-Verlag, Berlin, ISBN 3-540-54062-8, URL https://doi.org/10.1007/978-3-662-12616-5, 1992.

[10] O. Kallenberg, Foundations of modern probability, Probability and its Applications (New York), Springer-Verlag, New York, second edn., ISBN 0-387-95313-2, URL https://doi.org/10.1007/978-1-4757-4015-8, 2002.

[11] M. Leite, N. Petrov, E. Weng, Stationary distributions of semistochastic processes with disturbances at random times and with random severity, Nonlinear Anal. Real World Appl. 13 (2) (2012) 497–512, ISSN 1468-1218, URL https://doi.org/10.1016/j.nonrwa.2011.02.025.

[12] J. Broda, A. Grigo, N. P. Petrov, Convergence rates for semistochastic processes, Discrete & Continuous Dynamical Systems - B 24 (2019) 109–125, ISSN 1531-3492, URL https://doi.org/10.3934/dcdsb.2019001.

[13] C. Xue, H. G. Othmer, Multiscale models of taxis-driven patterning in bacterial populations, SIAM J. Appl. Math. 70 (1) (2009) 133–167, ISSN 0036-1399, URL https://doi.org/10.1137/070711505.

[14] E. Hairer, S. P. Nørsett, G. Wanner, Solving ordinary differential equations. I, vol. 8 of *Springer Series in Computational Mathematics*, Springer-Verlag, Berlin, second edn., ISBN 3-540-56670-8, nonstiff problems, 1993.

[15] E. Hairer, G. Wanner, Solving ordinary differential equations. II, vol. 14 of *Springer Series in Computational Mathematics*, Springer-Verlag, Berlin, second edn., ISBN 3-540-60452-9, URL https://doi.org/10.1007/978-3-642-05221-7, stiff and differential-algebraic problems, 1996.

[16] J. Stoer, R. Bulirsch, Introduction to numerical analysis, vol. 12 of *Texts in Applied Mathematics*, Springer-Verlag, New York, third edn., ISBN 0-387-95452-X, URL https://doi.org/10.1007/978-0-387-21738-3, translated from the German by R. Bartels, W. Gautschi and C. Witzgall, 2002.

[17] J. C. Butcher, Numerical methods for ordinary differential equations, John Wiley & Sons, Ltd., Chichester, third edn., ISBN 978-1-119-12150-3, URL https://doi.org/10.1002/9781119121534, with a foreword by J. M. Sanz-Serna, 2016.

[18] J. Jacod, A. N. Shiryaev, Limit theorems for stochastic processes, vol. 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, Springer-Verlag, Berlin, second edn., ISBN 3-540-43932-3, URL https://doi.org/10.1007/978-3-662-05265-5, 2003.