LECTURE 4 STOCHASTIC DIFFERENTIAL EQUATIONS AND SOLUTIONS

Let us consider the following simple stochastic ordinary equation:

$$dX(t) = -\lambda X(t)\,dt + dW(t), \quad \lambda > 0. \tag{0.1}$$

It can be readily verified by Ito's formula that the following process

$$X(t) = e^{-\lambda t}x_0 + \int_0^t e^{-\lambda(t-s)}\,dW(s) \tag{0.2}$$

satisfies Equation (0.1). By the Kolmogrov continuity theorem, the solution is Hölder continuous of order less than 1/2 in time since

$$\mathbb{E}[|X(t) - X(s)|^2] \le (t-s)^2(\frac{2}{\lambda} + x_0^2) + |t-s|\,. \tag{0.3}$$

This simple model shows that the solution to a stochastic differential equation is Hölder continuous of order less than 1/2 and thus does not have derivatives in time. This low regularity of solutions leads to different concerns in SODEs (and their numerical methods) from ODEs.

1. Existence and uniqueness of strong solutions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(W(t), \mathcal{F}_t^W) = ((W_1(t), \dots, W_m(t))^\top, \mathcal{F}_t^W)$ be an m-dimensional standard Wiener process, where $\mathcal{F}_t^W$, $0 \le t \le T$, is an increasing family of σ-subalgebras of $\mathcal{F}$ induced by $W(t)$. Consider the system of Ito SODEs

$$dX = a(t,X)dt + \sum_{r=1}^m \sigma_r(t,X)dW_r(t), \quad t \in (t_0, T], \; X(t_0) = x_0, \tag{1.1}$$

where X, a, σ_r are m-dimensional column-vectors and x_0 is independent of w. We assume that $a(t,x)$ and $\sigma(t,x)$ are sufficiently smooth and globally Lipschitz.

Remark 1.1. *The SODEs* (1.1) *can be rewritten in Stratonovich sense under mild conditions. The equation* (1.1) *can be written as*

$$dX = [a(t,X) - c(t,X)]dt + \sum_{r=1}^m \sigma_r(t,X)dW_r(t), \quad t \in (t_0, T], \; X(t_0) = x_0, \tag{1.2}$$

where

$$c(t,X) = \frac{1}{2}\sum_{r=1}^m \frac{\partial \sigma_r(t,X)}{\partial x}\sigma_r(t,X),$$

and $\frac{\partial \sigma_r}{\partial x}$ *is the Jacobi matrix of the column-vector* σ_r*:*

$$\frac{\partial \sigma_r}{\partial x} = \begin{bmatrix} \frac{\partial \sigma_r}{\partial x_1} & \cdots & \frac{\partial \sigma_r}{\partial x_m} \end{bmatrix} = \begin{bmatrix} \frac{\partial \sigma_{1,r}}{\partial x_1} & \cdots & \frac{\partial \sigma_{1,r}}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial \sigma_{m,r}}{\partial x_1} & \cdots & \frac{\partial \sigma_{m,r}}{\partial x_m} \end{bmatrix}.$$

We denote $f \in \mathbb{L}_{\rm ad}(\Omega; L^2([a,b]))$ if $f(t)$ is adapted to $\mathcal{F}_t$ and $f(t,\omega) \in L^2([a,b])$, i.e.,

$$f \in \mathbb{L}_{\rm ad}(\Omega; L^2([a,b])) = \left\{ f(t,\omega) | f(t,\omega) \text{ is } \mathcal{F}_t\text{-measurable and } \mathbb{P}(\int_a^b f_s^2\,ds < \infty) = 1 \right\}.$$

Here $\{\mathcal{F}_t;\, a \le t \le b\}$ is a filtration such that

Date: November 3, 2019.

- for each t, $f(t)$ and $W(t)$ are $\mathcal{F}_t$-measurable, i.e., $f(t)$ and $W(t)$ are adapted to the filtration $\mathcal{F}_t$.
- for any $s \leq t$, $W(t) - W(s)$ is independent of the σ-filed $\mathcal{F}_s$.

Definition 1.2 (A strong solution to a SODE)**.** *We say that $X(t)$ is a (strong) solution to SDE (1.1) if*

- $a(t, X(t)) \in \mathbb{L}_{\text{ad}}(\Omega, L^1([c,d]))$,
- $\sigma(t, X(t)) \in \mathbb{L}_{\text{ad}}(\Omega, L^2([c,d]))$,
- *and $X(t)$ satisfies the following integral equation a.s.*

$$X(t) = x + \int_0^t a(s, X(s))\, ds + \int_0^t \sigma(s, X(s))\, dW(s). \tag{1.3}$$

In general, it is difficult to give a necessary and sufficient condition for the existence and uniqueness of strong solutions. Usually, we can give sufficient conditions.

Theorem 1.3 (Existence and uniqueness)**.** *If X_0 is $\mathcal{F}_0$-measurable and $\mathbb{E}[X_0^2] < \infty$. The coefficients a, σ satisfy the following conditions.*

- *(Lipschitz condition) a and σ are Lipschitz continuous, i.e., there is a constant $K > 0$ such that*

$$|a(x) - a(y)| + \sum_{r=1}^{m} |\sigma_r(x) - \sigma_r(y)| \leq K|x - y|.$$

- *(Linear growth) a and σ grow at most linearly i.e., there is a $C > 0$ such that*

$$|a(x)| + |\sigma(x)| \leq C(1 + |x|),$$

then the SDE above has a unique strong solution and the solution has the following properties

- *$X(t)$ is adapted to the filtration generated by X_0 and $W(s)$ $(s \leq t)$.*
- $\mathbb{E}[\int_0^t X^2(s)\, ds] < \infty$.

See [Øksendal, 2003, Chapter 5] for a proof.

Here are some examples where the conditions in the theorem are satisfied.

- (Geometric Brownian motion) For μ, $\sigma \in \mathbb{R}$,

$$dX(t) = \mu X(t)\, dt + \sigma X(t)\, dW(t), \quad X_0 = x.$$

- (Sine process) For $\sigma \in \mathbb{R}$,

$$dX(t) = \sin(X(t))\, dt + \sigma\, dW(t), \quad X_0 = x.$$

- (modified Cox-Ingersoll-Ross process) For θ_1, $\theta_2 \in \mathbb{R}$,

$$dX(t) = -\theta_1 X(t)\, dt + \theta_2 \sqrt{1 + X(t)^2}\, dW(t), \quad X_0 = x. \quad \theta_1 + \frac{\theta_2^2}{2} > 0.$$

Remark 1.4. *The condition in the theorem is also known as global Lipschitz condition. A straightforward generalization is one-sided Lipschitz condition (global monotone condition)*

$$(x - y)^\top (a(x) - a(y)) + p_0 \sum_{r=1}^{m} |\sigma_r(x) - \sigma_r(y)|^2 \leq K|x - y|^2, \quad p_0 > 0,$$

and the growth condition can also be generalized as

$$x^\top a(x) + p_1 \sum_{r=1}^{m} |\sigma_r(x)|^2 \leq C(1 + |x|^2).$$

Theorem 1.5 (Regularity of the solution)**.** *Under the conditions of Theorem 1.3, the solution is continuous and there exists a constant $C > 0$ depending only on t that*

$$\mathbb{E}[|X(t) - X(s)|^{2p}] \leq C\, |t - s|^p\,, \quad p \geq 1.$$

The proof of this theorem rely on the Burkholder-Davis-Gundy inequality. Then by the Kolmogorov continuity theorem, we can conclude that the solution is only Hölder continuous with exponent less than 1/2, which is the same as Brownian motion.

2. Solution methods

This process (0.2) here is a special case of the Ornstein-Uhlenbeck process, which satisfies the equation

$$dX(t) = \kappa(\theta - X(t))\,dt + \sigma\,dW(t). \tag{2.1}$$

where κ, $\sigma > 0$, $\theta \in \mathbb{R}$. The solution to (2.1) can be obtained by *the method of change-of-variable*: $Y(t) = \theta - X(t)$. Then by Ito's formula we have

$$dY(t) = -\kappa Y(t)\,dt + \sigma\,d(-W(t)).$$

Similar to (0.2), the solution is

$$Y(t) = e^{-\kappa t}Y_0 + \sigma\int_0^t e^{-\kappa(t-s)}\,d(-W(s)). \tag{2.2}$$

Then by $Y(t) = \theta - X(t)$, we have

$$X(t) = X_0e^{-\kappa t} + \theta(1 - e^{-\kappa t}) + \sigma\int_0^t e^{-\kappa(t-s)}\,dW(s).$$

In a more general case, we can use similar ideas to find explicit solutions to SODEs.

2.1. The integrating factor method.

We apply the integrating factor method to solve nonlinear SDEs of the form

$$dX(t) = f(t, X(t))\,dt + \sigma(t)X(t)\,dW(t), \quad X_0 = x. \tag{2.3}$$

where f is a continuous deterministic function defined from $\mathbb{R}^+ \times \mathbb{R}$ to $\mathbb{R}$.

- Step 1. Solve the equation

  $$dG(t) = \sigma(t)G(t)\,dW(t).$$

  Then we have

  $$G(t) = \exp(\int_0^t \sigma(s)\,dW(s) - \frac{1}{2}\int_0^t \sigma^2(s)\,ds).$$

  The integrating factor function is defined by $F(t) = G^{-1}(t)$. It can be readily verified that $F(t)$ satisfies

  $$dF(t) = -\sigma(t)F(t)\,dW(t) + \sigma^2(t)F(t)\,dt.$$

- Step 2. Let $X(t) = G(t)C(t)$ and then $C(t) = F(t)X(t)$. Then by the product rule, (2.3) can be written as

  $$d(F(t)X(t)) = F(t)f(t, X(t))\,dt.$$

  Then C_t satisfies the following "deterministic" ODE

  $$dC(t) = F(t)f(t, G(t)C(t)). \tag{2.4}$$

- Step 3. Once we obtain $C(t)$, we get $X(t)$ from $X(t) = G(t)C(t)$.

Remark 2.1. *When* (2.4) *cannot be explicitly solved, we may use some numerical methods to obtain* $C(t)$*.*

Example 2.2. *Use the integrating factor method to solve the SDE*

$$dX(t) = (X(t))^{-1}\,dt + \alpha X(t)\,dW(t), \quad X_0 = x > 0,$$

where α *is a constant.*

Solution. Here $f(t,x) = x^{-1}$ and $F(t) = \exp(-\alpha W(t) + \frac{\alpha^2}{2}t)$. We only need to solve

$$dC(t) = F(t)[G^{-1}(t)C(t)]^{-1} = F^2(t)/C(t).$$

This gives $d(C(t))^2 = 2F^2(t)\,dt$ and thus

$$(C(t))^2 = 2\int_0^t \exp(-2\alpha W(s) + \alpha^2 s)\,ds + x^2.$$

Since the initial condition is $x > 0$, we take $Y(t) > 0$ such that

$$X(t) = G(t)Y(t) = \exp(\alpha W(t) - \frac{\alpha^2}{2}t)\sqrt{2\int_0^t \exp(-2\alpha W(s) + \alpha^2 s)\,ds + x^2} > 0.$$

2.2. **Moment equations of solutions.** For a more complicated SODE, we cannot obtain a solution that can be written explicitly in terms of $W(t)$. For example, the modified Cox-Ingersoll-Ross model (2.5) does not have an explicit solution:

$$dX(t) = \kappa(\theta - X(t))dt + \sigma\sqrt{X(t)}dW(t), \quad X_0 = x, \tag{2.5}$$

However, we can say a bit more about **the moments of the process** $X(t)$. Write (2.5) in its integral form:

$$X(t) = x + \kappa\int_0^t (\theta - X(s))ds + \sigma\int_0^t \sqrt{X(s)}\,dW(s) \tag{2.6}$$

and using Ito's formula gives

$$X^2(t) = x^2 + (2\kappa\theta + \sigma^2)\int_0^t X(s)\,ds - 2\kappa\int_0^t X(s)^2\,ds + 2\sigma\int_0^t (X(s))^{3/2}\,dW(s). \tag{2.7}$$

From this equation and the properties of Ito's integral, we can obtain the moments of the solution. *The first moment* can be obtained by taking expectation over both sides of (2.6):

$$m_t := \mathbb{E}[X(t)] = x + \kappa\left(\theta t - \int_0^t \mathbb{E}[X(s)]\,ds\right),$$

because the expectation of the stochastic integral part is zero[1]. We can then solve the following ODE:

$$dm_t = \kappa(\theta - m_t)dt.$$

The solution is given by:

$$m_t = \theta + (x - \theta)e^{-\kappa t}.$$

For *the second moment*, we get from (2.7) that

$$\mathbb{E}[X^2(t)] = x^2 + (2\kappa\theta + \sigma^2)\int_0^t \mathbb{E}[X(s)]ds - 2\kappa\int_0^t \mathbb{E}[X^2(s)]ds.$$

This is again an ODE similar to the one for m_t to solve:

$$\mathbb{E}[X^2(t)] = x^2 + (2\kappa\theta + \sigma^2)\Big(\theta t + (x-\theta)\frac{(1-e^{-\kappa t})}{\kappa}\Big) - 2\kappa\int_0^t \mathbb{E}[X^2(s)]ds.$$

Here we also assume that we have $\int_0^t \mathbb{E}[|X(s)|^3]\,ds < \infty$ so that $\int_0^t (X(s))^{3/2}\,dW(s)$ is an Ito integral with a square-integrable integrand and thus $\mathbb{E}[\int_0^t (X(s))^{3/2}\,dW(s)] = 0$.

Remark 2.3. *It can be shown using Feller's test* [Karatzas and Shreve, 1991, Theorem 5.29] *that the solution to* (2.7) *exists and is unique when* $2\kappa\theta > \sigma^2$ *and* $X_0 \geq 0$. *Moreover, the solution is strictly positive when* $X_0 > 0$. *If* $\mathbb{E}[|X_0|^3] < \infty$, *then* $\mathbb{E}[|X(t)|^p] < \infty$, $1 \leq p \leq 3$.

[1] Here we need to verify that $\int_0^t \sqrt{X(s)}\,dW(s)$ is indeed Ito's integral with a square-integrable integrand, by showing that $\int_0^t \mathbb{E}[|X(s)|]\,ds < \infty$. See Remark 2.3.

Unfortunately, even the first few moments are difficult to obtain in general. For example, we cannot get a closure for the second-order moment of the following SDE

$$dX(t) = \kappa(\theta - X(t))\,dt + \big(X(t)\big)^{\alpha}\,dW(t), \quad \frac{1}{2} < \alpha < 1.$$

We cannot even obtain the first-order moment of the following SDE

$$dX(t) = \sin(X(t))\,dt + dW(t), \quad X_0 = x.$$

3. Numerical methods for SODEs

As explicit solutions to SODEs are usually hard to find, we seek numerical approximation of solutions.

3.1. **Derivation of numerical methods based on numerical integration.** A starting point for numerical SODEs is numerical integration. Consider the SODE (1.1) over $[t,\, t+h]$:

$$X(t+h) = X(t) + \int_t^{t+h} a(s, X(s))\,ds + \sum_{r=1}^{m}\int_t^{t+h} \sigma_r(s, X(s))\,dW_r.$$

The simplest scheme for (1.1) is the forward Euler scheme. In the forward Euler scheme, we replace (approximate)

$$\int_t^{t+h} a(s, X(s))\,ds \text{ with } \int_t^{t+h} a(t, X(t))\,ds = a(t, X(t))h$$

and

$$\int_t^{t+h} \sigma_r(s, X(s))\,dW_r \text{ with } \int_t^{t+h} \sigma_r(t, X(t))\,dW_r = \sigma_r(t, X(t))(W_r(t+h) - W_r(t)).$$

Then we obtain the forward Euler scheme (also known as Euler-Maruyama scheme):

$$X_{k+1} = X_k + a(t_k, X_k)h + \sum_{r=1}^{m} \sigma_l(t_k, X_k)\Delta_k W_r, \tag{3.1}$$

where $h = (T - t_0)/N,\quad t_k = t_0 + kh,\quad k = 0, \dots, N$. $X_0 = x_0$ and $\Delta_k W_r = W_r(t_{k+1}) - W_r(t_k)$. The Euler scheme can be implemented by replacing the increments $\Delta_k W_r$ with Gaussian random variables:

$$X_{k+1} = X_k + a(t_k, X_k)h + \sum_{r=1}^{m} \sigma_l(t_k, X_k)\sqrt{h}\xi_{l,k+1}, \tag{3.2}$$

where $\xi_{r,k+1}$ are i.i.d. $\mathcal{N}(0,1)$ random variables.

Replacing (approximating) the drift term with its value at $t+h$, we have

$$\int_t^{t+h} a(s, X(s))\,ds \approx \int_t^{t+h} a(t+h, X(t+h))\,ds = a(t+h, X(t+h))h.$$

The resulting scheme is called backward Euler scheme (also known as drift-implicit Euler scheme)

$$X_{k+1} = X_k + a(t_{k+1}, X_{k+1})h + \sum_{r=1}^{m} \sigma_l(t_k, X_k)\Delta_k W_r, \quad k = 0, 1, \dots, N-1. \tag{3.3}$$

The following schemes can be considered as extensions of forward and backward Euler schemes

$$X_{k+1} = X_k + [(1-\lambda)a(t_k, X_k) + \lambda a(t_{k+1}, X_{k+1})]h + \sum_{r=1}^{m} \sigma_r(t_k, X_k)\sqrt{h}\xi_{r,k+1}, \tag{3.4}$$

where $\lambda \in [0,1]$, or similarly

$$X_{k+1} = X_k + a((1-\lambda)t_k + \lambda t_{k+1},\, (1-\lambda)X_k + \lambda X_{k+1})h + \sum_{r=1}^{m} \sigma_r(t_k, X_k)\sqrt{h}\xi_{r,k+1}. \tag{3.5}$$

We can also derive numerical methods for (1.1) in order to get high-order convergence. For example, in (1.1), we can approximate the diffusion terms σ_r using their half-order Ito-Taylor's expansion which leads to the Milstein scheme [Milstein, 1995]. Let us illustrate the derivation of the Milstein scheme for an autonomous SODE (a and σ do not explicitly depend on t) when $m = 1$ and $r = 1$, i.e., for scalar equation with single noise. With the following approximation,

$$\begin{aligned}\int_t^{t+h} a(X(s))\,ds &\approx \int_t^{t+h} a(X_t)\,ds = a(X(t))h \\ \int_t^{t+h} \sigma(X(s))\,dW(s) &\approx \int_t^{t+h} [\sigma(X(t)) + \int_t^s \sigma'(X(t))\sigma(X_t)\,dW(\theta)]\,dW(s) \\ &= \sigma(X(t))(W(t+h)-W(t)) + \sigma'(X(t))\sigma(X(t)) \int_t^{t+h}\int_t^s dW(\theta)\,dW(s),\end{aligned}$$

we can obtain the Milstein scheme

$$X_{k+1} = X_k + a(X_k)h + \sigma(X_k)(W(t_{k+1}) - W(t_k)) + \frac{1}{2}\sigma(X_k)\sigma'(X_k)[(W(t_{k+1}) - W(t_k))^2 - h].$$

One can also derive a drift-implicit Milstein scheme as follows:

$$X_{k+1} = X_k + a(X_{k+1})h + \sigma(X_k)(W(t_{k+1}) - W(t_k)) + \frac{1}{2}\sigma(X_k)\sigma'(X_k)[(W(t_{k+1}) - W(t_k))^2 - h].$$

For (1.1), the Milstein scheme is as follows, see e.g. [Kloeden and Platen, 1992, Milstein and Tretyakov, 2004],

$$X_{k+1} = X_k + a(t_k, X_k)h + \sum_{r=1}^{m} \sigma_r(t_k, X_k)\xi_{rk}\sqrt{h} + \sum_{i,l=1}^{m} \Lambda_i\sigma_l(t, X_k)I_{i,l,t_k}, \tag{3.6}$$

where $I_{i,l,t_k} = \int_{t_k}^{t_{k+1}} \int_{t_k}^{s} dW_i\,dW_l$. To efficiently evaluate this double integral, see Chapter 4 of [Zhang and Karniadakis, 2017]. The scheme (3.6) is of first-order mean-square convergence. For commutative noises, i.e.

$$\Lambda_i\sigma_l = \Lambda_l\sigma_i, \quad \Lambda_l = \sigma_l^\top \frac{\partial}{\partial x}, \tag{3.7}$$

we can use only increments of Brownian motions of the double Ito integral in (3.6) since

$$I_{i,l,t_k} + I_{l,i,t_k} = (\xi_{ik}\xi_{lk} - \delta_{il})h,$$

where δ_{il} is the Kronecker delta function. In this case, we have a simplified version of (3.6):

$$X_{k+1} = X_k + a(t_k, X_k)h + \sum_{r=1}^{m} \sigma_r(t_k, X_k)\xi_{lk}\sqrt{h} + \frac{1}{2}\sum_{i,l=1}^{m} \Lambda_i\sigma_l(t, X_k)(\xi_{ik}\xi_{lk} - \delta_{il})h. \tag{3.8}$$

There has been an extensive literature on numerical methods for SODES. We refer to [Higham, 2001, Schurz, 2002] for introduction to numerical methods for SODEs and to [Kloeden and Platen, 1992, Milstein, 1995] for a systematic construction of numerical methods for SODEs.

For numerical methods for SODEs and SPDEs, the key issues are whether a numerical method converges and in what sense and whether it is stable in some sense, as well as how fast it converges.

3.2. Strong convergence.

Definition 3.1 (Strong convergence in L^p). *A method (scheme) is said to have a strong convergence order γ in L^p if there exists a constant $K > 0$ independent of h such that*

$$\mathbb{E}[|X_k - X(t_k)|^p] \le Kh^{p\gamma}$$

for any $k = 0, 1, \ldots, N$ and $Nh = T$ and sufficiently small h.

In many applications and in this book, a strong convergence refers to convergence in the mean-square sense, i.e., $p = 2$.

If the coefficients of (1.1) satisfy the conditions in Theorem 1.3, the forward Euler scheme (3.1) and the backward Euler scheme (3.3) are convergent with half-order ($\gamma = 1/2$) in the mean-square sense (strong convergence order half), i.e.,

$$\max_{1\leq k\leq N} \mathbb{E}[|X(t_k) - X_k|^2] \leq Kh,$$

where K is positive constant independent of h. When the noise is additive, i.e., the coefficients of noises are functions of time instead of functions of the solutions, these schemes are of first-order convergence.

Under the conditions in Theorem 1.3, the Milstein scheme (3.6) can be shown to have a strong convergence order one, i.e., $\gamma = 1$.

Note that all these schemes are one-step schemes. One can use the following Milstein's fundamental theorem to derive their mean-square convergence order. Introduce the one-step approximation $\bar{X}_{t,x}(t+h)$, $t_0 \leq t < t+h \leq T$, for the solution $X_{t,x}(t+h)$ of (1.1), which depends on the initial point (t,x), a time step h, and $\{W_1(\theta) - W_1(t), \dots, W_m(\theta) - W_m(t),\ t \leq \theta \leq t+h\}$ and which is defined as follows:

$$\bar{X}_{t,x}(t+h) = x + A(t,x,h; W_i(\theta) - W_i(t),\ i = 1,\dots,m,\ t\leq\theta\leq t+h). \tag{3.9}$$

Using the one-step approximation (3.9), we recurrently construct the approximation $(X_k, \mathcal{F}_{t_k})$, $k = 0,\dots,N$, $t_{k+1} - t_k = h_{k+1}$, $T_N = T$:

$$\begin{aligned} X_0 = X(t_0),\ X_{k+1} &= \bar{X}_{t_k,\bar{X}_k}(t_{k+1}) \qquad (3.10)\\ &= X_k + A(t_k, X_k, h_{k+1}; W_i(\theta) - W_i(t_k),\ i = 1,\dots,m,\ t_k\leq\theta\leq t_{k+1}). \end{aligned}$$

For simplicity, we will consider a uniform time step size, i.e., $h_k = h$ for all k. The proof of the following theorem can be found in [Milstein, 1995, Milstein and Tretyakov, 2004, Chapter 1].

Theorem 3.2 (Fundamental convergence theorem of one-step numerical methods). *Suppose that*

(i) *the coefficients of* (1.1) *are Lipschitz continuous;*
(ii) *the one-step approximation* $\bar{X}_{t,x}(t+h)$ *from* (3.9) *has the following orders of accuracy: for some* $p \geq 1$ *there are* $\alpha \geq 1$, $h_0 > 0$, *and* $K > 0$ *such that for arbitrary* $t_0 \leq t \leq T-h$, $x \in \mathbb{R}^d$, *and all* $0 < h \leq h_0$:

$$|\mathbb{E}[X_{t,x}(t+h) - \bar{X}_{t,x}(t+h)]| \leq K(1+|x|^2)^{1/2} h^{q_1}\,,$$

$$\left[\mathbb{E}|X_{t,x}(t+h) - \bar{X}_{t,x}(t+h)|^{2p}\right]^{1/(2p)} \leq K(1+|x|^{2p})^{1/(2p)} h^{q_2} \tag{3.11}$$

with

$$q_2 \geq \frac{1}{2}\,,\ q_1 \geq q_2 + \frac{1}{2}\,;$$

Then for any N *and* $k = 0,1,\dots,N$ *the following inequality holds:*

$$\left[\mathbb{E}|X_{t_0,X_0}(t_k) - \bar{X}_{t_0,X_0}(t_k)|^{2p}\right]^{1/(2p)} \leq K(1+\mathbb{E}|X_0|^{2p})^{1/(2p)} h^{q_2-1/2}, \tag{3.12}$$

where $K > 0$ *do not depend on* h *and* k, *i.e., the order of accuracy of the method* (3.10) *is* $q = q_2 - 1/2$.

Many other schemes of strong convergence based on Ito-Taylor's expansion have been developed, such as Runge-Kutta methods, predictor-corrector methods and splitting (split-step) methods, see e.g. [Kloeden and Platen, 1992, Milstein and Tretyakov, 2004].

3.3. **Weak convergence.** The weak integration of SDEs is computing the expectation

$$\mathbb{E}[f(X(T)], \tag{3.13}$$

where $f(x)$ is a sufficiently smooth function with growth at infinity not faster than a polynomial:

$$|f(x)| \leq K(1+|x|^{\kappa}) \tag{3.14}$$

for some $K > 0$ and $\kappa \geq 1$.

Definition 3.3 (Weak convergence). *A method (scheme) is said to have a weak convergence order γ if there exists a constant $K > 0$ independent of h such that*

$$|\mathbb{E}[f(X_k)] - \mathbb{E}[f(X(t_k))]| \leq Kh^\gamma$$

for any $k = 0, 1, \dots, N$ and $Nh = T$ and sufficiently small h.

Under the conditions of Theorem 1.3, the following error estimate holds for the forward Euler scheme (3.2) (see e.g. [Milstein and Tretyakov, 2004, Chapter 2]):

$$|\mathbb{E}f(X_k) - \mathbb{E}f(X(t_k))| \leq Kh, \tag{3.15}$$

where $K > 0$ is a constant independent of h. The backward Euler scheme (3.3) and the Milstein scheme (3.6), are all of weak convergence order 1.

This first-order weak convergence of the forward Euler scheme can also be achieved by replacing $\xi_{l,k+1}$ with discrete random variables [Milstein and Tretyakov, 2004], e.g., the weak Euler scheme has the form

$$\tilde{X}_{k+1} = \tilde{X}_k + ha(t_k, \tilde{X}_k) + \sqrt{h} \sum_{r=1}^{m} \sigma_r(t_k, \tilde{X}_k)\zeta_{r,k+1}, \qquad k = 0, \dots, N-1, \tag{3.16}$$

where $\tilde{X}_0 = x_0$ and $\zeta_{r,k+1}$ are i.i.d. random variables with the law

$$P(\zeta = \pm 1) = 1/2. \tag{3.17}$$

The following error estimate holds for (3.16)-(3.17) (see e.g. [Milstein and Tretyakov, 2004, Chapter 2]):

$$|\mathbb{E}f(\tilde{X}_N) - \mathbb{E}f(X(T))| \leq Kh, \tag{3.18}$$

where $K > 0$ can be a different constant than that in (3.15).

Introducing the function $\varphi(y)$, $y \in \mathbb{R}^{mN}$, so that

$$\varphi(\xi_{1,1}, \dots, \xi_{r,1}, \dots, \xi_{1,N}, \dots, \xi_{m,N}) = f(X_N), \tag{3.19}$$

we have

$$\begin{aligned} \mathbb{E}[f(X(T)] &\approx \mathbb{E}f(X_N) = \mathbb{E}\varphi(\xi_{1,1}, \dots, \xi_{r,1}, \dots, \xi_{1,N}, \dots, \xi_{m,N}) \qquad (3.20) \\ &= \frac{1}{(2\pi)^{mN/2}} \int_{\mathbb{R}^{rN}} \varphi(y_{1,1}, \dots, y_{m,1}, \dots, y_{1,N}, \dots, y_{m,N}) \exp\left(-\frac{1}{2}\sum_{i=1}^{mN} y_i^2\right) dy. \end{aligned}$$

Further, it is not difficult to see from (3.16)-(3.17) that

$$\begin{aligned} \mathbb{E}[f(X(T)] &\approx \mathbb{E}f(\tilde{X}_N) = \mathbb{E}\varphi(\zeta_{1,1}, \dots, \zeta_{m,1}, \dots, \zeta_{1,N}, \dots, \zeta_{m,N}) \qquad (3.21) \\ &= Q_2^{\otimes mN} \varphi(y_{1,1}, \dots, y_{m,1}, \dots, y_{1,N}, \dots, y_{m,N}), \end{aligned}$$

where Q_2 is the Gauss-Hermite quadrature rule with nodes ± 1 and equal weights 1/2. We note that $\mathbb{E}[f(\tilde{X}_N)]$ can be viewed as an approximation of $\mathbb{E}[f(X_N)]$ and that (cf. (3.15) and (3.18)) $\left|\mathbb{E}[f(X_N)] - \mathbb{E}[f(\tilde{X}_N)]\right| = \mathcal{O}(h)$.

Remark 3.4. *Let $\zeta_{l,k+1}$ in* (3.16) *be i.i.d. random variables with the law*

$$P(\zeta = \mathsf{y}_{n,j}) = \mathsf{w}_{n,j}, \quad j = 1, \dots, n, \tag{3.22}$$

where $\mathsf{y}_{n,j}$ are nodes of the Gauss-Hermite quadrature Q_n and $\mathsf{w}_{n,j}$ are the corresponding quadrature weights. Then

$$\mathbb{E}f(\tilde{X}_N) = \mathbb{E}\varphi(\zeta_{1,N}, \dots, \zeta_{m,N}) = Q_n^{\otimes mN} \varphi(y_{1,1}, \dots, y_{m,N}),$$

which can be a more accurate approximation of $\mathbb{E}[f(X_N)]$ than $\mathbb{E}[f(\tilde{X}_N)]$ from (3.21) *but the weak-sense error for the SDEs approximation $\mathbb{E}f(\tilde{X}_N) - \mathbb{E}f(X(T))$ remains of order $\mathcal{O}(h)$.*

We can also use the second-order weak scheme (3.23) for (1.1) (see, e.g. [Milstein and Tretyakov, 2004, Chapter 2]):

$$\begin{aligned} X_{k+1} &= X_k + ha(t_k, X_k) + \sqrt{h}\sum_{i=1}^{m} \sigma_i(t_k, X_k)\xi_{i,k+1} + \frac{h^2}{2}\mathfrak{L}a(t_k, X_k) \\ &\quad + h\sum_{i=1}^{m}\sum_{j=1}^{r} \Lambda_i \sigma_j(t_k, X_k)\eta_{i,j,k+1} + \frac{h^{3/2}}{2}\sum_{i=1}^{m}(\Lambda_i a(t_k, X_k) + \mathfrak{L}\sigma_i(t_k, X_k))\xi_{i,k+1}, \\ k &= 0, \dots, N-1, \end{aligned} \tag{3.23}$$

where $X_0 = x_0$; $\eta_{i,j} = \frac{1}{2}\xi_i\xi_j - \gamma_{i,j}\zeta_i\zeta_j/2$ with $\gamma_{i,j} = -1$ if $i < j$ and $\gamma_{i,j} = 1$ otherwise;

$$\Lambda_l = \sum_{i=1}^{m} \sigma_l^i \frac{\partial}{\partial x_i}, \quad \mathfrak{L} = \frac{\partial}{\partial t} + \sum_{i=1}^{m} a^i \frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{r=1}^{m}\sum_{i,j=1}^{m} \sigma_l^i \sigma_l^i \frac{\partial^2}{\partial x_i \partial x_j};$$

and $\xi_{i,k+1}$ and $\zeta_{i,k+1}$ are mutually independent random variables with Gaussian distribution or with the laws $P(\xi = 0) = 2/3$, $P(\xi = \pm\sqrt{3}) = 1/6$ and $P(\zeta = \pm 1) = 1/2$. The following error estimate holds for (3.23) (see e.g. [Milstein and Tretyakov, 2004, Chapter 2]):

$$|\mathbb{E}f(X(T)) - Ef(X_N)| \le Kh^2.$$

We again refer to [Kloeden and Platen, 1992, Milstein and Tretyakov, 2004] for more weakly convergent numerical schemes.

3.4. **Linear stability.** To understand the stability of time integrators for SODEs, we consider the following linear model:

$$dX = \lambda X\, dt + \sigma\, dW(t), \quad X_0 = x, \quad \lambda < 0. \tag{3.24}$$

Consider one-step methods of the following form:

$$X_{n+1} = A(z)X_n + B(z)\sqrt{h}\sigma\xi_n, \quad z = \lambda h. \tag{3.25}$$

Here h is the time step size and $A(z)$ and $B(z)$ are analytic functions, $\delta W_n = \sqrt{h}\xi_n$ are i.i.d. Gaussian random variables.

For (3.24), we have $X(t)$ is a Gaussian random variable and

$$\lim_{t\to\infty} \mathbb{E}[X(t)] = 0, \quad \lim_{t\to\infty} \mathbb{E}[X^2(t)] = \frac{\sigma^2}{2\lambda}.$$

It can be readily shown that X_n is also a Gaussian random variable with $\mathbb{E}[X_n] = A^n(z)x$ and

$$\lim_{n\to\infty} \mathbb{E}[|X_n|^2] = \frac{\sigma^2}{2\lambda}R(z), \quad R(z) = -\frac{2zB^2(z)}{1 - A^2(z)}.$$

Here are some examples:

- Euler scheme: $A(z) = 1 + z$, $B(z) = 1$, and $R(z) = \frac{2}{2+z}$.
- Backward Euler scheme: $A(z) = B(z) = \frac{1}{1-z}$ and $R(z) = \frac{2}{2-z}$.
- midpoint scheme, $A(z) = \frac{1+z/2}{1-z/2}$ and $B(z) = \frac{1}{1-z/2}$ and $R(z) = 1$.

For the Euler schemes, $|A(z)| < 1$ implies that $-2 < \lambda h = z < 0$. For the backward Euler scheme and the midpoint scheme, any $h > 0$ will leads to $|A(z)| < 1$.

When $R(z) = 1$ and $|A(z)| < 1$, We call the one-step scheme is mean-square stable. If $R(z) = 1$ and $|A(z)| < 1$ holds for all $h > 0$, we call the scheme A-stable in the mean-squarse sense.

For long-time integration, *L-stability* is also helpful when a stiff problem (e.g. λ is large) is investigated. The L-stability requires A-stability and that $A(-\infty) = 0$ such that $\lim_{z\to-\infty} \mathbb{E}[X_n] = \lim_{z\to-\infty} A^n(z)x = 0$ for any fixed n. When λh is large (e.g., λ is too large to have such a practical h that λh is small), L-stable schemes can still obtain the decay of the solution $\mathbb{E}[X(t)] = 0$ even with moderately small time step sizes while A-stable schemes usually require very small h. For example, the trapezoidal rule is A-stable but not L-stable since $A(-\infty) = 1$. In practice, this means that for an extremely large λ, the trapezoidal rule damps the mean since $\mathbb{E}[X_{n+1}] = \lim_{z\to-\infty} A(z)\mathbb{E}[X_n] = \mathbb{E}[X_n]$ while $\mathbb{E}[X(t_{n+1})] = \lim_{z\to-\infty} \exp(-z)\mathbb{E}[X_{t_n}] = 0$, where $t_{n+1} - t_n = h$.

However, when $A(z)$ and $B(z)$ are rational functions of z, it is impossible to have both A-stability and L-stability since when $R(z) = 1$, it holds that $A(-\infty) = 1$. The claim can be proved by the argument of contradiction.

Remark 3.5. *It is still possible to have a scheme such that it is L-stable and A-stable. Define*

$$\tilde{X}_n = C(z)X_n + D(z)\sigma\sqrt{h}\xi_n, \quad z = \lambda h,$$

where X_n is from (3.25). For example, for the backward Euler scheme, $A(z) = B(z) = \frac{1}{1-z}$ and

$$\lim_{n\to\infty} \mathbb{E}[\left|\tilde{X}_n\right|^2] = \frac{\sigma^2}{2\lambda}(C^2(z)R(z) - 2zD^2(z)).$$

The limit is exactly the same as the variance of $X(\infty)$ when $C(z) = 1$ and $D(z) = (1 - z/2)^{-1}$. Such a scheme with $\tilde{X}$ approximating X is called a post-processing *scheme or a* predictor-corrector *scheme.*

The linear model (3.24) is designed for additive noise. For multiplicative noise, we can consider the following geometric Brownian motion.

$$dX = \lambda X\, dt + \sigma X\, dW(t), \quad X(0) = 1. \tag{3.26}$$

Here we assume that $\lambda, \sigma \in \mathbb{R}$. The solution to (3.26) is

$$X = \exp((\lambda - \frac{1}{2}\sigma^2)t + \sigma W(t)).$$

The solution is mean-square stable if $\lambda + \frac{\sigma^2}{2} < 0$, i.e., $\lim_{t\to\infty} \mathbb{E}[X^2(t)] = 0$. It is asymptotic stable ($\lim_{t\to\infty} |X(t)| = 0$) if $\lambda - \frac{\sigma^2}{2} < 0$. The mean-square stability implies asymptotic stability. Here we consider only mean-square stability.

Applying the forward Euler scheme (3.2) to the linear model (3.26), we have

$$X_{k+1} = (1 + \lambda h + \sigma\sqrt{h}\xi_k)X_k.$$

The second moment is $\mathbb{E}[X_{k+1}^2] = \mathbb{E}[X_k^2]\mathbb{E}[(1 + \lambda h + \sigma\sqrt{h}\xi_k)^2] = \mathbb{E}[X_k^2]\big((1 + \lambda h)^2 + h\sigma^2\big)$. For $\lim_{k\to\infty} \mathbb{E}[X_{k+1}^2] = 0$, we need $(1 + \lambda h)^2 + h\sigma^2 < 1$. Similarly, for the backward Euler scheme (3.3), we need $1 + \sigma^2 h < (1 - \lambda h)^2$.

We call the region of $(\lambda h, \sigma^2 h)$ where a scheme is mean-square stable the mean-square stability region of the scheme. To allow relative large h for stiff problems, e.g., when λ is large, we need a large stability region. Usually, explicit schemes have smaller stability regions than implicit (including semi-implicit and drift-implicit) schemes do.

Both schemes (3.2) and (3.6) are explicit and hence they have small stability regions. To improve the stability region, we can use some semi-implicit (drift-implicit) schemes, e.g. (3.3) and drift-implicit Milstein scheme.

3.5. **Nonlinear and stiff SODEs.** Let $X_{t_0,X_0}(t) = X(t)$, $t_0 \le t \le T$, be a solution of the system (1.1). We will assume the following.

Assumption 3.6. *(i) The initial condition is such that*

$$\mathbb{E}|X_0|^{2p} \le K < \infty, \quad \textit{for all} \;\; p \ge 1. \tag{3.27}$$

(ii) For a sufficiently large $p_0 \ge 1$ there is a constant $c_1 \ge 0$ such that for $t \in [t_0, T]$,

$$(x - y, a(t,x) - a(t,y)) + \frac{2p_0 - 1}{2}\sum_{r=1}^{m} |\sigma_r(t,x) - \sigma_r(t,y)|^2 \le c_1|x - y|^2, \;\; x, y \in \mathbb{R}^d. \tag{3.28}$$

(iii) There exist $c_2 \ge 0$ and $\varkappa \ge 1$ such that for $t \in [t_0, T]$,

$$|a(t,x) - a(t,y)|^2 \le c_2(1 + |x|^{2\varkappa-2} + |y|^{2\varkappa-2})|x - y|^2, \;\; x, y \in \mathbb{R}^d. \tag{3.29}$$

The condition (3.28) implies that

$$(x, a(t,x)) + \frac{2p_0 - 1 - \varepsilon}{2} \sum_{r=1}^{m} |\sigma_r(t,x)|^2 \leq c_0 + c_1'|x|^2, \quad t \in [t_0, T],\ x \in \mathbb{R}^d, \tag{3.30}$$

where $c_0 = |a(t,0)|^2/2 + \frac{(2p_0-1-\varepsilon)(2p_0-1)}{2\varepsilon} \sum_{r=1}^{m} |\sigma_r(t,0)|^2$ and $c_1' = c_1 + 1/2$. The inequality (3.30) together with (3.27) is sufficient to ensure finiteness of moments: there is $K > 0$

$$\mathbb{E}|X_{t_0,X_0}(t)|^{2p} < K(1 + \mathbb{E}|X_0|^{2p}),\ 1 \leq p \leq p_0 - 1, \quad t \in [t_0, T]. \tag{3.31}$$

Also, (3.29) implies that

$$|a(t,x)|^2 \leq c_3 + c_2'|x|^{2\varkappa}, \quad t \in [t_0, T],\ x \in \mathbb{R}^d, \tag{3.32}$$

where $c_3 = 2|a(t,0))|^2 + 2c_2(\varkappa - 1)/\varkappa$ and $c_2' = 2c_2(1 + \varkappa)/\varkappa$.

Example 3.7. *Here is an example for Assumption 3.6 (ii):*

$$dX = -\mu X|X|^{r_1 - 1}dt + \lambda X^{r_2} dW,$$

where $\mu,\ \lambda > 0$, $r_1 \geq 1$, *and* $r_2 \geq 1$. *If* $r_1 + 1 > 2r_2$ *or* $r_1 = r_2 = 1$, *then (3.28) is valid for any* $p_0 \geq 1$. *If* $r_1 + 1 = 2r_2$ *and* $r_1 > 1$ *then (3.28) is valid for* $1 \leq p_0 \leq \mu/\lambda^2 + 1/2$.

We introduce the one-step approximation $\bar{X}_{t,x}(t+h)$, $t_0 \leq t < t+h \leq T$, for the solution $X_{t,x}(t+h)$ of (1.1), as in (3.9) and (3.10).

The following theorem is a generalization of Milstein's fundamental theorem (see [Milstein, 1995, Milstein and Tretyakov, 2004, Chapter 1]) from the global to non-globally Lipschitz case.

For simplicity, we will consider a uniform time step size, i.e., $h_k = h$ for all k.

Theorem 3.8 ([Zhang and Karniadakis, 2017]). *Suppose (i) Assumption 3.6 holds;*

(ii) The one-step approximation $\bar{X}_{t,x}(t+h)$ *from* (3.9) *has the following orders of accuracy: for some* $p \geq 1$ *there are* $\alpha \geq 1$, $h_0 > 0$, *and* $K > 0$ *such that for arbitrary* $t_0 \leq t \leq T - h$, $x \in \mathbb{R}^d$, *and all* $0 < h \leq h_0$:

$$|\mathbb{E}[X_{t,x}(t+h) - \bar{X}_{t,x}(t+h)]| \leq K(1 + |x|^{2\alpha})^{1/2} h^{q_1}, \tag{3.33}$$

$$\left[\mathbb{E}|X_{t,x}(t+h) - \bar{X}_{t,x}(t+h)|^{2p}\right]^{1/(2p)} \leq K(1 + |x|^{2\alpha p})^{1/(2p)} h^{q_2} \tag{3.34}$$

with

$$q_2 \geq \frac{1}{2},\ q_1 \geq q_2 + \frac{1}{2}; \tag{3.35}$$

(iii) The approximation X_k *from* (3.10) *has finite moments, i.e., for some* $p \geq 1$ *there are* $\beta \geq 1$, $h_0 > 0$, *and* $K > 0$ *such that for all* $0 < h \leq h_0$ *and all* $k = 0, \ldots, N$:

$$\mathbb{E}|X_k|^{2p} < K(1 + \mathbb{E}|X_0|^{2p\beta}). \tag{3.36}$$

Then for any N *and* $k = 0, 1, \ldots, N$ *the following inequality holds:*

$$\left[\mathbb{E}|X_{t_0,X_0}(t_k) - \bar{X}_{t_0,X_0}(t_k)|^{2p}\right]^{1/(2p)} \leq K(1 + \mathbb{E}|X_0|^{2\gamma p})^{1/(2p)} h^{q_2 - 1/2}, \tag{3.37}$$

where $K > 0$ *and* $\gamma \geq 1$ *do not depend on* h *and* k, *i.e., the order of accuracy of the method* (**??**) *is* $q = q_2 - 1/2$.

Corollary 3.9. *In the setting of Theorem 3.8 for* $p \geq 1/(2q)$ *in* (3.37), *there is* $0 < \varepsilon < q$ *and an a.s. finite random variable* $C(\omega) > 0$ *such that*

$$|X_{t_0,X_0}(t_k) - X_k| \leq C(\omega) h^{q - \varepsilon},$$

i.e., the method (3.10) *for* (1.1) *converges with order* $q - \varepsilon$ *a.s.*

The corollary is proved using the Borel-Cantelli lemma.

Besided the implicit schemes, some explicit schemes can be also used for nonlinear SODEs, e. g., balanced schemes, truncated/projection schemes.

Balanced schemes

Specifically, the balanced Euler schemes can be written as

$$X_{k+1} = X_k + a(t_k, X_k)h + \sum_{r=1}^{m} \sigma_r(t_k, X_k)(W_r(t_{k+1}) - W_r(t_k)) + P(t_k, t_{k+1}, X_k, X_{k+1}, h, W_r(t_{k+1}) - W_r(t_k)),$$

where the term P is not zero unless $h = 0$. It can be considered as a *penalty method* or a *Lagrange multiplier method* for stiff SDEs.

A special form is the so-called tamed Euler scheme.

$$X_{k+1} = X_k + \mathcal{T}_0(a(t_k, X_k)h) + \sum_{r=1}^{m} \mathcal{T}_r(\sigma_r(t_k, X_k)(W_r(t_{k+1}) - W_r(t_k))),$$

where $\mathcal{T}(y) = \frac{y}{1+|y|}$, or $\mathcal{T}(y) = \tanh(y)$. Tamed Milstein schemes can be also derived.

Remark 3.10. *Apply the tame function only when the drift or diffusion coefficients grow superlinearly. No tame function is needed if the coefficient grows linearly.*

Truncated schemes/Projection schemes

Suppose that $|a(x)| + \sum_{r=1}^{m} |\sigma_r(x)| \leq \mu(R)$ when $|x| \leq R$ and μ is a strictly increasing function with positive values.

Take $\mathcal{T}(h) \leq h^{-\frac{1}{4}}$ and then the following scheme converges

$$X_{k+1} = X_k + ha(\tilde{X}_k) + \sum_{r=1}^{m} \sigma_r(t_k, \tilde{X}_k)\xi_{rk}\sqrt{h}., \tag{3.38}$$

$$\tilde{X_k} = X_k \mathbf{1}_{|X_k| \leq \mu^{-1}(\mathcal{T}(h))} + \operatorname{sgn}(X_k)\mu^{-1}(\mathcal{T}(h))\mathbf{1}_{|X_k| > \mu^{-1}(\mathcal{T}(h))}. \tag{3.39}$$

The above scheme converges even for locally continuous coefficients.

Theorem 3.11. *Assume that (local Lipschitz condition)*

$$|a(x) - a(y)| + \sum_{r=1}^{m} |\sigma_r(x) - \sigma_r(y)| \leq K_R |x - y|, \tag{3.40}$$

and (linear growth condition)

$$x^\top a(x) + \frac{p-1}{2} \sum_{r=1}^{m} |\sigma_r(x)| \leq K(1 + |x|^2), \quad p > 2, K > 0. \tag{3.41}$$

The truncated scheme converges when $q \in [2, p]$*:*

$$\lim_{h \to 0} \mathbb{E}[|X_N - X(t_N)|^q] = 0.$$

3.6. **Wong-Zakai approximation.** Suppose that the trajectories of the Brownian motion are approximated by piece-wise continuously-differentiable functions

$$\lim_{n \to \infty} \sup_{0 < t < T} |W_n(t) - W(t)| = 0 \tag{3.42}$$

with probability one. Consider the ordinary differential equation

$$\frac{dX_n(t)}{dt} = a\,(t, X_n(t)) + \sigma\,(t, X_n(t))\,\frac{dW_n(t)}{dt}, X_n(0) = x. \tag{3.43}$$

Then the limit X is the solution of the stochastic ordinary differential equation in the Stratonovich form:

$$\frac{dX(t)}{dt} = a\,(t, X(t)) + \sigma\,(t, X)t)) \circ \frac{dW(t)}{dt}, X(0) = x \tag{3.44}$$

Appendix A. Burkholder-Davis-Gundy inequality

For any $1 \le p < \infty$, there exist constants $c_p, C_p > 0$ such that for all (local) martingales X with $X_0 = 0$ and stopping times τ, the following inequality holds:

$$c_p \mathbb{E}[[X]_\tau^{p/2}] \le \mathbb{E}[(X_\tau^*)^p] \le C_p \mathbb{E}\left[[X]_\tau^{p/2}\right].$$

Here $X_t^* = \sup_{s \le t} |X_s|$ is the maximum process X_t and $[X]$ is the quadratic variation of X. Furthermore, for continuous (local) martingales, this statement holds for all $0 < p < \infty$.

The proof is based on Doob's maximal inequality and Ito's formula.

Appendix B. Fully implicit schemes

Fully implicit schemes are also used in practice because of their symplecticity-preserving property and effectiveness in long-term integration. The following fully implicit scheme is from [Milstein and Tretyakov, 2004]:

$$\begin{aligned} X_{k+1} &= X_k + a(t_{k+\lambda}, (1-\lambda)X_k + \lambda X_{k+1})h \\ &\quad -\lambda \sum_{r=1}^{m} \sum_{j=1}^{d} \frac{\partial \sigma_r}{\partial x^j}(t_{k+\lambda}, (1-\lambda)X_k + \lambda X_{k+1}) \sigma_r^j(t_{k+\lambda}, (1-\lambda)X_k + \lambda X_{k+1})h \\ &\quad + \sum_{r=1}^{m} \sigma_r(t_{k+\lambda}, (1-\lambda)X_k + \lambda X_{k+1}) \left(\zeta_{rh}\right)_k \sqrt{h}, \end{aligned} \tag{B.1}$$

where $0 < \lambda \le 1$, $t_{k+\lambda} = t_k + \lambda h$ and $(\zeta_{rh})_k$ are i.i.d. random variables so that

$$\zeta_h = \begin{cases} \xi, \ |\xi| \le A_h, \\ A_h, \ \xi > A_h, \\ -A_h, \ \xi < -A_h, \end{cases} \tag{B.2}$$

with $\xi \sim \mathcal{N}(0,1)$ and $A_h = \sqrt{2l|\ln h|}$ with $l \ge 1$. We recall [Milstein and Tretyakov, 2004, Lemma 1.3.4] that

$$\left|\mathbb{E}[(\xi^2 - \zeta_h^2)]\right| \le (1 + 2\sqrt{2l|\ln h|})h^l. \tag{B.3}$$

All these fully implicit schemes are of half-order convergence in the mean-square sense, see e.g. [Milstein and Tretyakov, 2004, Chapter 1]. When the noise is additive, i.e., the coefficients of noises are functions of time instead of functions of the solutions, these schemes are of first-order convergence.

References

[Higham, 2001] Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.*, 43(3):525–546 (electronic).

[Karatzas and Shreve, 1991] Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition.

[Kloeden and Platen, 1992] Kloeden, P. E. and Platen, E. (1992). *Numerical solution of stochastic differential equations*. Springer-Verlag, Berlin.

[Milstein, 1995] Milstein, G. N. (1995). *Numerical integration of stochastic differential equations*. Kluwer Academic Publishers Group, Dordrecht.

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

[Øksendal, 2003] Øksendal, B. (2003). *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth edition. An introduction with applications.

[Schurz, 2002] Schurz, H. (2002). Numerical analysis of stochastic differential equations without tears. In *Handbook of stochastic analysis and applications*, volume 163 of *Statist. Textbooks Monogr.*, pages 237–359. Dekker, New York.

[Zhang and Karniadakis, 2017] Zhang, Z. and Karniadakis, G. E. (2017). *Numerical methods for stochastic partial differential equations with white noise*, volume 196 of *Applied Mathematical Sciences*. Springer, Cham.