LECTURE NOTES ON STOCHASTIC DIFFERENTIAL EQUATION

ANANTA KUMAR MAJEE

1. Introduction

Consider the problem of estimating the integral of a function f over the unit interval. Let $\alpha = \int_0^1 f(x)\,dx$. We may represent α as an expectation $\mathbb{E}\big[f(U)\big]$, where U is uniformly distributed random variable between 0 and 1.

Definition 1.1 (Uniform distribution)**.** A random variable X has uniform distribution, denoted as $X \sim \text{Unif}\,[a,b]$ if it has a probability density function $\bar{f}$ given by

$$\bar{f}(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0, & \text{otherwise.} \end{cases}$$

Definition 1.2 (Normal distribution)**.** A random variable X is normally distributed with mean μ and variance σ^2, denoted as $X \sim \mathcal{N}(\mu, \sigma^2)$, if it has a density function $\bar{f}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. We say that a d-dimensional random variable Y is normally distributed with mean $\mu \in \mathbb{R}^d$ and covariance matrix Σ, if it has a density given by

$$\bar{f}(x) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp\big(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\big), \quad x \in \mathbb{R}^d$$

with $|\Sigma|$ the determinant of Σ.

If $Z \sim \mathcal{N}(0,1)$, then $\mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$. If $Y \sim \mathcal{N}(\mu, \Sigma)$, then its ith component Y_i has distribution $\mathcal{N}(\mu_i, \sigma_i^2)$ with $\sigma_i^2 = \Sigma_{ii}$. If $X \sim \mu, \pm$, then for any linear transformation A, $AX \sim \mathcal{N}(A\mu, A\Sigma A^\top)$.

Definition 1.3 (Central limit theorem)**.** Let $X_1, X_2, \cdots$ are iid random variables with mean μ and variance σ^2. Then

$$\sqrt{N}\Big(\frac{1}{N}\sum_{i=1}^N X_i - \mu\Big) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Definition 1.4 (Law of Large number)**.** Let $X_1, X_2, \cdots$ be iid random variable with mean $\mu = \mathbb{E}[X_i]$. Then, **Strong Law of Large number** says that

$$\bar{X}_N = \frac{1}{N}\sum_{i=1}^N X_i \xrightarrow{\text{a.s.}} \mu \quad \text{as } N \to \infty.$$

Moreover, **Weak Law of Large number** gives

$$\bar{X}_N = \frac{1}{N}\sum_{i=1}^N X_i \xrightarrow{\mathbb{P}} \mu \quad \text{as } N \to \infty.$$

Let $\hat{\alpha}_N = \sum_{i=1}^N \frac{1}{N} f(U_i)$, where $U_1, U_2, \cdots$ are independent and uniformly distributed over $[0,1]$. Suppose f is integrable over $[0,1]$. Then by Strong Law of Large number, we see that $\hat{\alpha}_N \longrightarrow \alpha$ with probability 1 as $N \to \infty$. Define the error E_N via

$$E_N := \hat{\alpha}_N - \alpha = \frac{1}{N}\sum_{i=1}^N f(U_i) - \int_0^1 f(x)\,dx = \sum_{i=1}^N \frac{f(U_i) - \mathbb{E}[f(U)]}{N}.$$

By Central limit theorem, $\sqrt{N}E_N \rightharpoonup \sigma Y$, where $Y \sim \mathcal{N}(0,1)$ and

$$\sigma^2 = \int_0^1 f^2(x)\,dx - \Big(\int_0^1 f(x)\,dx\Big)^2 = \int_0^1 \Big(f(x) - \int_0^1 f(x)\,dx\Big)^2\,dx.$$

In practice, σ^2 is approximated by $\sigma_N^2 := \frac{1}{N-1}\sum_{i=1}^N \Big(f(U_i) - \sum_{j=1}^N \frac{f(U_j)}{N}\Big)^2$.

1.1. **Pseudo random numbers:** Our goal is to generate approximate random numbers so called Pseudo random numbers. By a generator of genuinely random numbers, we mean a mechanism for producing a sequence of random variables $U_1, U_2, \cdots$ with the property that

i) each U_i is uniformly distributed between 0 and 1.
ii) The U_i' are mutually independent.

Linear Congruential Generators: The general linear congruential generator, first proposed by Lehmer, takes the form

$$\begin{aligned} x_{i+1} &= (ax_i + c) \bmod m \\ u_{i+1} &= \frac{x_{i+1}}{m}. \end{aligned}$$

We say that the generator has **full period** if the number of distinct values generated from any seed x_0 is $m-1$.

Conditions for full period: If $c \neq 0$, then the conditions are

i) c and m are relatively prime i.e., $\gcd(c,m) = 1$.
ii) every prime number that divides m divides $a-1$.
iii) $a-1$ is divisible by 4 if m is.

If $c = 0$ and m is prime, then for any $(x_0 \neq 0)$ the conditions are

i) $a^{m-1} - 1$ is a multiple of m.
ii) $a^j - 1$ is not a multiple of m for $j = 1, 2, \cdots m-2$.

Observe that, if m is a power of 2 and $c \neq 0$, the generator has full period if c is odd and $a = 4n+1$ for some n. In practice, we consider the algorithm $x_{i+1} = ax_i \bmod m$ where a and m are relatively prime with the combinations $m = 2^{31}$, $a = 2^{16}+3$ or $m = 2^{31}-1$, $a = 7^5$. In Monte Carlo computations, we use the Pseudo random numbers $\{u_i\}_{i=1}^N$, which for $N << 2^{31}$ behaves approximately as independent uniformly distributed variables.

Multiple recursive generator: It is an extension of a linear congruential generator which uses a higher-order recursion of the form

$$x_i = (a_1 x_{i-1} + a_2 x_{i-2} + \cdots + a_k x_{i-k}) \bmod m$$

followed by $u_i = \frac{x_i}{m}$. A seed for this generator consists of initial values $x_{k-1}, x_{k-2}, \cdots, x_0$.

1.2. **Inverse transform method:** We want to generate a random variable X with the property that $\mathbb{P}(X \le x) = F(x)$, where F is a given cumulative distribution function. The inverse transform method sets

$$X = F^{-1}(U), \quad U \sim \text{Unif}\,[0,1]$$

where $F^{-1}(u) = \inf\{x : F(x) \ge u\}$. Let U_i be the uniformly distributed random variables on [0,1] which can be generated using linear congruential generators algorithm. Therefore

$$\mathbb{P}(U_i \le u) = \begin{cases} 0, & \text{if} \quad u < 0 \\ u, & \text{if} \quad 0 \le u \le 1 \\ 1, & \text{if} \quad u > 1. \end{cases}$$

We will now show that $\mathbb{P}(X \le x) = F(x)$ for this algorithm.

$$\mathbb{P}(X \le x) = \mathbb{P}\big(F^{-1}(U) \le x\big) = \mathbb{P}\big(U \le F(x)\big) = F(x).$$

Last equality holds true as F is cumulative distribution function and $U \sim \text{Unif}\,[0,1]$.

Example 1.1 (Exponential distribution with mean $\theta > 0$)**.** *We want to generate a random variable X such that $\mathbb{P}(X \le x) = F(x)$, where*

$$F(x) = 1 - e^{-\frac{x}{\theta}}, \quad x \ge 0.$$

By Inverse transform method, X is given by the formula $X = F^{-1}(U), \quad U \sim Unif[0,1]$. Let us calculate

$$F^{-1}(u) = \inf\{x : F(x) \ge u\} = \inf\{x : 1 - e^{-\frac{x}{\theta}} \ge u\} = \inf\{x : 1 - u \ge e^{-\frac{x}{\theta}}\} = \inf\{x : x \ge -\theta \log(1-u$$
$$= -\theta \log(1-u).$$

Thus, since U and $1-U$ has same distribution, we see that

$$X = -\theta \log(1-U), \quad or \quad X = -\theta \log(U) \quad U \sim Unif[0,1].$$

Example 1.2 (Discrete random variable)**.** *We want to find a discrete random variable X such that $\mathbf{p}(k) = \mathbb{P}(X = k)$, $k \ge 0$ where $\mathbf{p}$ is the probability mass function. Inverse transform method gives*

$$X = \begin{cases} 0, & if \quad U \le \mathbf{p}(0) \\ k, & if \quad \sum_{i=0}^{k-1} \mathbf{p}(i) < U \le \sum_{i=0}^{k} \mathbf{p}(i), \quad k \ge 1. \end{cases}$$

This algorithm is easily verified directly by recalling that $\mathbb{P}\big(a < U \le b\big) = b - a$ for $0 \le a < b \le 1$. Here we use $a = \sum_{i=0}^{k-1} \mathbf{p}(i) < b = \sum_{i=0}^{k} \mathbf{p}(i)$ and so $b - a = \mathbf{p}(k)$.

Example 1.3 (Bernoulli$\,(p)$ and Binomial$\,(n,p)$ distributions)**.** *We want to generate a Bernoulli$\,(p)$ random variable X in which*

$$\mathbb{P}(X = 0) = 1 - p, \quad \mathbb{P}(X = 1) = p \quad \textit{for some } p \in (0,1)$$

To do so, we have the following algorithm

i) Generate $U \sim Unif[0,1]$
ii) Set $X = 0$ if $U \le 1-p$, and $X = 1$ if $U > 1-p$.

For Binomial(n,p) distribution, the probability mass function is given by

$$\mathbf{p}(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 \le k \le n.$$

Let $X = \sum_{i=1}^n Y_i$ where Y_i are iid Bernoulli(p) distributed random variables. So the algorithm reads as

> *i) Generate n iid random variables $U_1, U_2, \cdots U_n \sim$ Unif$[0,1]$*
> *ii) For each $1 \le i \le n$, set $Y_i = 0$ if $U_i \le 1-p$, and $Y_i = 1$ if $U_i > 1-p$*
> *iii) Set $X = \sum_{i=1}^n Y_i$. Then $X \sim$ Binomial(n,p).*

1.3. **Acceptance-Rejection Method:** We wish to sample a random variable X from a distribution F in which the probability density function f is known.
Idea: Find an alternative probability distribution G with density function $g(x)$, which we could be sampled using an efficient algorithm, such that the function $g(x)$ is 'close' to $f(x)$ satisfying $\sup_x \frac{f(x)}{g(x)} \le c$ for some constant $c > 0$. In practice, we would want c as close to 1 as possible.

Acceptance-Rejection Method is described by the following algorithm.

> i) Generate a random variable Y distributed as G.
> ii) Generate $U \sim \text{Unif}\,[0,1]$, independent of Y.
> ii) If $U \le \frac{f(Y)}{cg(Y)}$, set $X = Y$ (accept); otherwise go back to i) (reject).

The algorithm says that $\frac{f(Y)}{cg(Y)}$ is independent of U and $0 < \frac{f(Y)}{cg(Y)} \le 1$, and in the end we obtain our X as having the conditional distribution of Y given that the event $\{U \le \frac{f(Y)}{cg(Y)}\}$ occurs.
Consequence: The number of iterations N needed to successfully generate X is a random variable having geometric distribution with 'success' probability $p = \mathbb{P}\big(U \le \frac{f(Y)}{cg(Y)}\big)$, and hence $\mathbb{E}\big[N\big] = \frac{1}{p}$.

Exercise 1.1. *Show that $p := \mathbb{P}\big(U \le \frac{f(Y)}{cg(Y)}\big) = \frac{1}{c}$.*

Solution: Since U is uniformly distributed over $[0,1]$, we see that

$$\mathbb{P}\Big(U \le \frac{f(Y)}{cg(Y)} \big| Y \le y\Big) = \frac{f(y)}{cg(y)}$$
$$\implies p = \mathbb{P}\big(U \le \frac{f(Y)}{cg(Y)}\big) = \int_{-\infty}^{\infty} \mathbb{P}\Big(U \le \frac{f(Y)}{cg(Y)} \big| Y \le y\Big) g(y)\, dy = \int_{-\infty}^{\infty} \frac{f(y)}{cg(y)} g(y)\, dy = \frac{1}{c}.$$

Last equality holds true as f is a probability density function. Therefore, $p = \frac{1}{c}$ and hence $\mathbb{E}[N] = c$, the bounding constant.

Remark 1.1. In view of the above exercise, we see that it is desirable to choose out alternative density g so as to minimize this constant $c = \sup_x \big\{\frac{f(x)}{g(x)}\big\}$. Thus, the expected number of iterations of the algorithm required until an X is successfully generated is exactly the bounding constant c.

Proof of the algorithm: We need to show that $F(y) = \mathbb{P}\Big(Y \leq y \big| U \leq \frac{f(Y)}{cg(Y)}\Big)$. Let $B = \{U \leq \frac{f(Y)}{cg(Y)}\}$ and $A = \{Y \leq y\}$. Now by Bayes' theorem, one has

$$\mathbb{P}(A\big|B) = \frac{\mathbb{P}(B\big|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B\big|A)G(y)}{p} = cG(y)\mathbb{P}(B\big|A).$$

Let us calculate $\mathbb{P}(B\big|A)$.

$$\mathbb{P}(B\big|A) = \frac{\mathbb{P}\Big(U \leq \frac{f(Y)}{cg(Y)}, Y \leq y\Big)}{G(y)} = \int_{-\infty}^{y} \frac{\mathbb{P}\Big(U \leq \frac{f(Y)}{cg(Y)}\Big|Y = s \leq y\Big)}{G(y)} g(s)\,ds = \frac{1}{G(y)}\int_{-\infty}^{y} \frac{f(s)}{cg(s)} g(s)\,ds$$
$$= \frac{1}{cG(y)}\int_{-\infty}^{y} f(s)\,ds = \frac{F(y)}{cG(y)}.$$

Therefore, $\mathbb{P}(A\big|B) = cG(y).\frac{F(y)}{cG(y)} = F(y)$.

Example 1.4 (Beta distribution)**.** *The beta density on* $[0,1]$ *with parameter* α, β *is given by*

$$f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad 0 \leq x \leq 1, \quad \textit{with } B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

where Γ *is the gamma function. The case* $\alpha = \beta = \frac{1}{2}$ *is the 'arcsine' distribution*

$$F(x) = \frac{2}{\pi} arcSin(\sqrt{x}),\ 0 \leq x \leq 1.$$

If $\alpha, \beta \geq 1$ *and at least one of the parameters exceeds* 1*, the beta density is unimodal and achieves its maximum at* $x = \frac{\alpha-1}{\alpha+\beta-2}$*. Let* $c = f\Big(\frac{\alpha-1}{\alpha+\beta-2}\Big)$*. Then* $f(x) \leq c$ *for all* $x \in [0,1]$*. So, we may choose* $g(x) \equiv 1$ *for* $0 \leq x \leq 1$*, which is in fact beta density with parameters* $\alpha = \beta = 1$*. In this case the acceptance-rejection method becomes*

1) *Generates* U_1, U_2 *from uniform distribution over* $[0,1]$*.*
2) *If* $cU_2 \leq f(U_1)$*, set* $X = U_1$*, otherwise go back to* $1)$*.*

Generating Univariate Normal: We want to generate two independent standard normal random variables via **Box-Müler method**. The algorithm reads as

i) Generate two independent random variables U, $V \sim \mathrm{Unif}\,[0,1]$.
ii) Set $R = -2\log(U)$ and $Z = 2\pi V$
ii) Set $X = \sqrt{R}\cos(Z)$, and $Y = \sqrt{R}\sin(Z)$. Then X and Y are independent $\mathcal{N}(0,1)$-distributed random variables.

<u>**Verification:**</u> The variables x and y are independent standard normal variables if and only if their joint density function is $\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}$. Let $x = r\cos(\theta)$ and $y = r\sin(\theta)$ where $0 \leq \theta < 2\pi$ and $0 \leq r < +\infty$. Then

$$e^{-\frac{x^2+y^2}{2}}\,dx\,dy = re^{-\frac{r^2}{2}}\,dr\,d\theta = d\big(e^{-\frac{r^2}{2}}\big)\,d\theta.$$

The random variables θ and r could be sampled by taking θ to be uniformly distributed in the interval $[0,2\pi)$ and $e^{-\frac{r^2}{2}}$ to be uniformly distributed in $(0,1]$. Thus, θ and r can be sampled as $\theta = 2\pi V$ and $r = \sqrt{-2\log(U)}$, where U and V are independent random variables and $U, V \sim \text{Unif}\,[0,1]$. This ensures that x and y are independent standard normal random variables.

In the above **Box-Müler** method, one needs to compute sine and cosine functions. One can also use acceptance-rejection method to generate a standard normal random variable $X \sim \mathcal{N}(0,1)$ which avoids the computations of sine and cosine functions. Moreover, if we can generate from the absolute value, $|X|$, then by symmetry we can obtain X by independently generating a random variable S (for sign) that is ± 1 with probability $\frac{1}{2}$ and setting $X = S|X|$. In other words, we generate a random variable $U \sim \text{Unif}\,[0,1]$ and set

$$X = \begin{cases} |X|, & \text{if } \; U \le 0.5 \\ -|X|, & \text{if } \; U > 0.5. \end{cases}$$

Let $g(x) = e^{-x}$, $\;g \ge 0$, the exponential density with rate 1. Then $c = \max_x \left\{\frac{f(x)}{g(x)}\right\} = \sqrt{\frac{2e}{\pi}}$ and hence $\frac{f(y)}{cg(y)} = e^{-\frac{(y-1)^2}{2}}$, where $f(y) = \frac{2}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$, density for $|X|$. Thus, the acceptance-rejection algorithm is then

i) Generate Y with an exponential distribution at rate 1 i.e., generate $U \sim \text{Unif}\,[0,1]$ and set $Y = -\log(U)$
ii) Generate $V \sim \text{Unif}\,[0,1]$, independent from Y.
iii) If $V \le e^{-\frac{(Y-1)^2}{2}}$, set $|X| = Y$; otherwise go back to i)
iv) Generate $Z \sim \text{Unif}\,[0,1]$. Set $X = |X|$ if $Z \le 0.5$, set $X = -|X|$ if $Z > 0.5$.

Note that $V \le e^{-\frac{(Y-1)^2}{2}}$ if and only if $-\log(V) \ge \frac{(Y-1)^2}{2}$ and since $-\log(V)$ is exponential rate 1, we can simplify the algorithm:

i) Generate two independent exponentials at rate 1; $Y_1 = -\log(U_1)$ and $Y_2 = -\log(U_2)$ where $U_1, U_2 \sim \text{Unif}\,[0,1]$.
ii) If $Y_2 \ge \frac{(Y_1-1)^2}{2}$, set $|X| = Y_1$; otherwise go back to i)
iii) Generate $U \sim \text{Unif}\,[0,1]$. Set $X = |X|$ if $U \le 0.5$, set $X = -|X|$ if $U > 0.5$.

Generating Multivariate Normals: If $Z \sim \mathcal{N}(0, \text{Id})$, then $X = \mu + AZ \sim \mathcal{N}(\mu, AA^\top)$. Thus, the problem of sampling X from the multivariate normal $\mathcal{N}(\mu, \Sigma)$ reduces to finding a matrix A for which $AA^\top = \Sigma$. For positive definite covariance matrix Σ, one can use **Cholesky Factorization** to find a lower triangular matrix A with $AA^\top = \Sigma$. In this case, entries of A are given by

$$A_{i,j} = \Big(\Sigma_{i,j} - \sum_{k=1}^{j-1} A_{ik}A_{jk}\Big)\Big/ A_{jj} \qquad j < i, \quad \text{and} \quad A_{ii} = \sqrt{\Sigma_{i,i} - \sum_{k=1}^{i-1} A_{ik}^2}.$$

For positive semi-definite covariance matrix Σ, one can use **eigenvector factorization** method:

$$\Sigma = BDB^\top$$

where B is the orthogonal matrix with columns $v_1, v_2, \cdots, v_d$ such that $\Sigma v_i = \lambda_i v_i$ and D is the diagonal matrix with diagonal entries $\lambda_1, \lambda_2, \cdots, \lambda_d$. Set $A = BD^{\frac{1}{2}}$. Then

$AA^\top = \Sigma$. Methods for calculating B and D are included in many mathematical software libraries.

1.4. **Brownian motion and its simulation.**

Definition 1.5. (Brownian motion) A scalar standard Brownian motion or standard Wiener process over $[0, T]$ is a random variable $W(t)$ that depends continuously on $t \in [0, T]$ and satisfies the following conditions.

i) $W(0) = 0$ $\mathbb{P}$-a.s.
ii) The increment $W(t) - W(s)$ is normally distributed with mean 0 and variance $t - s$ for all $0 \le s < t \le T$. Thus, $W(t) - W(s) \sim \sqrt{t-s}\mathcal{N}(0, 1)$ where $\mathcal{N}(0, 1)$ denotes a normally distributed random variable with mean 0 and variance 1.
iii) It has a independent increment i.e., for all $0 \le s < t < u < v \le T$, the increments $W(t) - W(s)$ and $W(v) - W(u)$ are independent.

For computational purpose, it is useful to consider discretized Brownian motion. Set $dt = \frac{T}{N}$ for some $N \in \mathbb{N}^+$ and let W_j denote $W(t_j)$ with $t_j = jdt$. Since $W(0) = 0$ $\mathbb{P}$-a.s., we have

$$W_j = W_{j-1} + dW_j, \;\; j = 1, 2, \cdots, N, \tag{1.1}$$

where each dW_j is an independent random variable of the form $\sqrt{dt}\mathcal{N}(0, 1)$. Here, the random number generator **randn** is used -each call to **randn** produces an independent "pseudorandom" number from the $\mathcal{N}(0, 1)$ distribution. In order to make experiments repeatable, the initial state of the random number generator to be set. We set the state, arbitrary, to be 100 with the command **randn**('state',100). Here is the MATLAB code for the simulation of Brownian motion.

```
 % Brownian path simulation
clc
clear all
rng( default );
T=1;
N=5000;
dt=T/N;
dW=zeros(1,N); % preallocate arrays
W=zeros(1,N); % for efficiency
dW(1)=sqrt(dt)*randn;
W(1)= dW(1); % since W(0)=0 is not allowed
for j=2:N
    dW(j)= sqrt(dt)* randn;
    W(j)=W(j-1) + dW(j);
end
plot([0:dt:T],[0,W],  g- )  % plot W against t
xlabel( t , FontSize , 16)
ylabel( W(t) , FontSize , 16,  Rotation , 0)
```

To perform the simulation more elegantly and efficiently, we replace the *for loop* with higher level "vectorized" commands. By **randn**(1,N) we creates a 1-by-N array of independent $\mathcal{N}(0, 1)$ samples. The function *cumsum* computes the cumulative sum of its argument, so the jth element of the 1-by-N array W is $dW(1) + dW(2) + \cdots + dW(j)$.

```
 % Brownian path simulation:vectorized
clc
clear all
randn( state ,100); % set the state of randn
T=1;
N=5000;
dt=T/N;
dW=sqrt(dt)*randn(1,N); % increments
W=cumsum(dW); % cumulative sum
plot([0:dt:T],[0,W],  r- )  % plot W against t
xlabel( t , FontSize , 16)
ylabel( W(t) , FontSize , 16,  Rotation , 0)
```

Next we evaluate the function $u(W(t)) = \exp(t+\frac{1}{2}W(t))$ along 1000 discretized Brownian paths. dW is an M-by-N array such that $dW(i,j)$ gives the increment dW_j in (1.1) for the ith path. We use **cumsum**$(dW, 2)$ to form cumulative sums across the second(column) dimension. We use **repmat**$(t, [M1])$ to produce an M-by-N array whose rows are all copies of t. By $U\text{mean} = \text{mean}(U)$, it computes column-wise averages, so Umean is a 1-by-N array whose jth entry is the sample average of $u(W(t_j))$. One can apply Itô's formula to the function $f(t,x) = \exp(t+\frac{1}{2}x)$ and obtain $\mathbb{E}[u(W(t))] = \exp(\frac{9t}{8})$.

```
 % Function along a Brownian path simulation
clc
clear all
randn( state ,100); % set the state of randn
T=1;
N=5000;
dt=T/N; t=[dt:dt:1];
M=1000; % number of sample paths
dW=sqrt(dt)*randn(M,N); % increments
W=cumsum(dW,2); % cumulative sum
U=exp(repmat(t,[M 1])+ 0.5*W);
Umean=mean(U);
plot([0 t],[1, Umean],  b- ), hold on  % plot mean over M paths
plot([0,t],[ones(5,1),U(1:5;:)],  r-- ), hold on % plot 5 individual paths
xlabel( t , FontSize , 16)
ylabel( U(t) , FontSize , 16,  Rotation , 0, HorizontalAlignment ,  right )
legend( mean of 1000 paths ,  5 individual paths , 2)
averr=norm(Umean-exp(9*t/8)), inf ) % sample error
```

One can notice that $u(W(t))$ is non smooth along individual paths but its sample average is smooth. In the above code, we use **averr** to compute the maximum deviation between the sample average and the exact expected value over all points t_j. **averr** will decrease if one increase the number of sample paths.

2. Euler approximation and SDE

We will prove the convergence of the forward Euler approximation of stochastic differential equation, following the convergence proof for Itô integrals. Consider the following stochastic differential equation (SDE)

$$\begin{aligned} dX(t) &= a(t, X(t))\, dt + b(t, X(t))\, dW(t) \quad t < T \\ X(0) &= X_0 \end{aligned}, \tag{2.1}$$

where $a : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are given functions, and X_0 is square integrable $\mathbb{R}^d$- valued random variable.

2.1. **Euler approximation of (2.1).** Let $\{\bar{t}_n\}_{n=0}^N$ with $\bar{t}_0 = 0$ and $\bar{t}_N = T$ be a set of nodal points of a grid in the time interval $[0, T]$. We define the discrete stochastic process $\bar{X}$ by the forward Euler method

$$\bar{X}(\bar{t}_{n+1}) - \bar{X}(\bar{t}_n) = a(\bar{t}_n, \bar{X}(\bar{t}_n))(\bar{t}_{n+1} - \bar{t}_n) + b(\bar{t}_n, \bar{X}(\bar{t}_n))\big(W(\bar{t}_{n+1}) - W(\bar{t}_n)\big), \tag{2.2}$$

for $n = 0, 1, \cdots, N-1$. We extend it continuously to all values at t by

$$\bar{X}(t) = \bar{X}(\bar{t}_n) + \int_{\bar{t}_n}^{t} a(\bar{t}_n, \bar{X}(\bar{t}_n))\, ds + \int_{\bar{t}_n}^{t} b(\bar{t}_n, \bar{X}(\bar{t}_n))\, dW(s), \quad \bar{t}_n \le t < \bar{t}_{n+1}$$

i.e., $\bar{X} : [0, T] \times \Omega \to \mathbb{R}^d$ satisfying

$$d\bar{X}(s) = \bar{a}(s, \bar{X})\, ds + \bar{b}(s, \bar{X}) dW(s), \quad \bar{t}_n \le s < \bar{t}_{n+1}, \tag{2.3}$$

where $\bar{a}(s, \bar{X}) = a(\bar{t}_n, \bar{X}(\bar{t}_n))$, $\bar{b}(s, \bar{X}) = b(\bar{t}_n, \bar{X}(\bar{t}_n))$ for $\bar{t}_n \le s < \bar{t}_{n+1}$, and the nodal values of the process $\bar{X}$ is defined by the Euler method (2.2).

Theorem 2.1. *Let $\bar{X}$ and $\tilde{X}$ be forward Euler approximation of the stochastic process $X : [0, T] \times \Omega \to \mathbb{R}^d$ satisfying the SDE (2.1) with time steps $\{\bar{t}_n\}_{n=0}^N$ and $\{\tilde{t}_m\}_{m=0}^M$ respectively. Let*

$$\Delta t = \max\Big\{ \max_{0\le n\le N-1} \bar{t}_{n+1} - \bar{t}_n, \max_{0\le m\le M-1} \tilde{t}_{m+1} - \tilde{t}_m \Big\},$$

and there exists a constant $C > 0$ such that

$$\begin{aligned} \big|a(t, x) - a(t, y)\big|_{\mathbb{R}^d} + \big|b(t, x) - b(t, y)\big|_{\mathcal{L}(\mathbb{R}^d,\mathbb{R}^m)} &\le C\big|x - y\big|_{\mathbb{R}^d}, \\ \big|a(t, x) - a(s, x)\big|_{\mathbb{R}^d} + \big|b(t, x) - b(s, x)\big|_{\mathcal{L}(\mathbb{R}^d,\mathbb{R}^m)} &\le C\big(1 + |x|_{\mathbb{R}^d}\big)\sqrt{|t - s|}. \end{aligned} \tag{2.4}$$

Then, there exists a constant $K > 0$ such that

$$\max\Big\{\mathbb{E}[|\bar{X}(t)|^2_{\mathbb{R}^d}], \mathbb{E}[|\tilde{X}(t)|^2_{\mathbb{R}^d}]\Big\} \le KT, \quad t < T \tag{2.5}$$

$$\mathbb{E}\Big[\big|\bar{X}(t) - \tilde{X}(t)\big|^2_{\mathbb{R}^d}\Big] \le K\Delta t, \quad t < T. \tag{2.6}$$

Proof. To prove (2.5), it is enough to prove $\max\limits_{0\le n\le N} \mathbb{E}\big[|Y_n|^2_{\mathbb{R}^d}\big] \le C_T$ where $Y_n = \bar{X}(\bar{t}_n)$. Let us denote by $|\cdot|$ the Euclidean norm in $\mathbb{R}^d$. Let $\Delta\bar{t} = \max\limits_{0\le n\le N-1} \bar{t}_{n+1} - \bar{t}_n$. Note that, thanks to the assumption (2.4)

$$\big|a(t, x)\big|_{\mathbb{R}^d} + \big|b(t, x)\big|_{\mathcal{L}(\mathbb{R}^d,\mathbb{R}^m)} \le C\Big(1 + \big|x\big|_{\mathbb{R}^d}\Big). \tag{2.7}$$

By using the formula $\big(a - b, a\big) = \frac{1}{2}\Big(|a|^2 - |b|^2 + |a - b|^2\Big)$, we have, in view of (2.2), and (2.7)

$$\frac{1}{2}\Big(|Y_{n+1}|^2 - |Y_n|^2 + |Y_{n+1} - Y_n|^2\Big)$$

$$= \Big(a(\bar{t}_n, Y_n)(\bar{t}_{n+1} - \bar{t}_n), Y_{n+1}\Big) + \Big(b(\bar{t}_n, Y_n)\Delta\bar{W}_n, Y_{n+1} - Y_n\Big) + \Big(b(\bar{t}_n, Y_n)\Delta\bar{W}_n, Y_n\Big)$$
$$\le C\Delta\bar{t}\big(1 + |Y_n|\big)|Y_{n+1}| + C\big(1 + |Y_n|\big)|Y_{n+1} - Y_n||\Delta\bar{W}_n| + \Big(b(\bar{t}_n, Y_n)\Delta\bar{W}_n, Y_n\Big),$$

where $\Delta\bar{W}_n = \big(W(\bar{t}_{n+1}) - W(\bar{t}_n)\big)$. Now we sum over $0 \le n \le j < N$ and use Cauchy Schwartz inequality to obtain

$$|Y_{j+1}|^2 + \sum_{n=0}^{j} |Y_{n+1} - Y_n|^2 \le C\Delta\bar{t}\sum_{n=0}^{j} |Y_n|^2 + C\sum_{n=0}^{j} (1 + |Y_n|^2)||\Delta\bar{W}_n|^2 + \sum_{n=0}^{j} \Big(\int_{\bar{t}_n}^{\bar{t}_{n+1}} b(\bar{t}_n, Y_n)dW(s), Y_n\Big)$$

By using the properties of conditional expectation and Brownian increment, we see that

$$\mathbb{E}\Big[\Big(\int_{\bar{t}_n}^{\bar{t}_{n+1}} b(\bar{t}_n, Y_n)dW(s), Y_n\Big)\Big] = 0,$$

$$\mathbb{E}\Big[\big(1 + |Y_n|^2\big)|\Delta\bar{W}_n|^2\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\big(1 + |Y_n|^2\big)||\Delta\bar{W}_n|^2\Big|\mathcal{F}_{\bar{t}_n}\Big]\Big] = \mathbb{E}\Big[\big(1 + |Y_n|^2\big)\mathbb{E}\Big[|\Delta\bar{W}_n|^2\Big|\mathcal{F}_{\bar{t}_n}\Big]\Big]$$
$$\le \Delta\bar{t}\,\mathbb{E}\big[\big(1 + |Y_n|^2\big)\big].$$

Combining these estimates, we have

$$\mathbb{E}\big[|Y_{j+1}|^2\big] \le C\Big(1 + \mathbb{E}\big[|X_0|^2\big]\Big) + C\Delta\bar{t}\sum_{n=0}^{j}\mathbb{E}\big[|Y_n|^2\big], \quad 0 \le j \le N.$$

One can use discrete version of Gronwall's lemma to conclude (2.5).

Next we will prove (2.6). Since $\tilde{X}$ is an Euler approximation associated on a grid in $[0, T]$ with nodes $\{\tilde{t}_m\}_{m=0}^{M}$, we see that

$$\bar{X}(s) - \tilde{X}(s) = \int_0^s (\bar{a} - \tilde{a})(t)\,dt + \int_0^s (\bar{b} - \tilde{b})(t)\,dW(t) \equiv \int_0^s \Delta a(t)\,dt + \int_0^s \Delta b(t)\,dW(t). \tag{2.8}$$

Now the definition of the discretized solutions implies that

$$\Delta a(t) = a(\bar{t}_n, \bar{X}(\bar{t}_n)) - a(\tilde{t}_m, \tilde{X}(\tilde{t}_m))$$
$$= \Big(a(\bar{t}_n, \bar{X}(\bar{t}_n)) - a(t, \bar{X}(t))\Big) + \Big(a(t, \bar{X}(t)) - a(t, \tilde{X}(t))\Big) + \Big(a(t, \tilde{X}(t)) - a(\tilde{t}_m, \tilde{X}(\tilde{t}_m))\Big)$$
$$\equiv a_1 + a_2 + a_3$$

where $t \in [\tilde{t}_m, \tilde{t}_{m+1}) \cap [\bar{t}_n, \bar{t}_{n+1})$. In view of the assumption (2.4), we have

$$|a_2| \le C\big|\bar{X}(t) - \tilde{X}(t)|,$$
$$|a_1| \le \big|a(\bar{t}_n, \bar{X}(\bar{t}_n)) - a(t, \bar{X}(\bar{t}_n))\big| + \big|a(t, \bar{X}(\bar{t}_n)) - a(t, \bar{X}(t))\big|$$
$$\le C\big|\bar{X}(\bar{t}_n) - \bar{X}(t)\big| + C\big(1 + |\bar{X}(\bar{t}_n)\big)\sqrt{|t - \bar{t}_n|}. \tag{2.9}$$

Again, thanks to (2.7), we see that

$$\big|\bar{X}(\bar{t}_n) - \bar{X}(t)\big| = \Big|a(\bar{t}_n, \bar{X}(\bar{t}_n)(t - \bar{t}_n) + b(\bar{t}_n, \bar{X}(\bar{t}_n)\big(W(t) - W(\bar{t}_n)\big)\Big|$$
$$\le C\big(1 + |\bar{X}(\bar{t}_n)\big)\Big\{(t - \bar{t}_n) + \big|W(t) - W(\bar{t}_n)\big|\Big\}. \tag{2.10}$$

The combination of (2.9) and (2.10) gives

$$|a_1| \le C\big(1 + |\bar{X}(\bar{t}_n)\big)\Big\{\sqrt{|t - \bar{t}_n|} + \big|W(t) - W(\bar{t}_n)\big|\Big\}.$$

In a similar way, one can show that

$$|a_3| \le C\big(1 + |\tilde{X}(\tilde{t}_m)|\big)\Big\{\sqrt{|t-\tilde{t}_m|} + \big|W(t) - W(\tilde{t}_m)\big|\Big\}.$$

Thus, we get

$$\begin{aligned}|\Delta a(t)|^2 \le C\Bigg\{ & |\bar{X}(t) - \tilde{X}(t)|^2 + \big(1 + |\bar{X}(\bar{t}_n)|^2\big)\Big(|t-\bar{t}_n| + \big|W(t) - W(\bar{t}_n)\big|^2\Big) \\ & + \big(1 + |\tilde{X}(\tilde{t}_m)|^2\big)\Big(|t-\tilde{t}_m| + \big|W(t) - W(\tilde{t}_m)\big|^2\Big)\Bigg\}. \end{aligned} \tag{2.11}$$

Note that, by using conditional expectation and (2.5)

$$\mathbb{E}\Big[\big(1 + |\bar{X}(\bar{t}_n)|^2\big)\big|W(t) - W(\bar{t}_n)\big|^2\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\big(1 + |\bar{X}(\bar{t}_n)|^2\big)\big|W(t) - W(\bar{t}_n)\big|^2\Big|\mathcal{F}_{\bar{t}_n}\Big]\Big]$$

$$= \mathbb{E}\Big[\big(1 + |\bar{X}(\bar{t}_n)|^2\big)\mathbb{E}\Big[\big|W(t) - W(\bar{t}_n)\big|^2\Big|\mathcal{F}_{\bar{t}_n}\Big]\Big] = \mathbb{E}\Big[\big(1 + |\bar{X}(\bar{t}_n)|^2\big)|t-\bar{t}_n|\Big] \le \Delta t\mathbb{E}\Big[\big(1 + |\bar{X}(\bar{t}_n)|^2\big)\Big],$$

$$\mathbb{E}\Big[\big(1 + |\tilde{X}(\tilde{t}_m)|^2\big)\big|W(t) - W(\tilde{t}_m)\big|^2\Big] \le \Delta t\mathbb{E}\Big[\big(1 + |\tilde{X}(\tilde{t}_m)|^2\big)\Big].$$

Taking expectation in (2.11) and using above estimates along with (2.5), we arrive at

$$\mathbb{E}\big[|\Delta a(t)|^2\big] \le C\Big(\mathbb{E}\big[|\bar{X}(t) - \tilde{X}(t)|^2\big] + \Delta t\Big). \tag{2.12}$$

Similarly, one can show that

$$\mathbb{E}\big[|\Delta b(t)|^2\big] \le C\Big(\mathbb{E}\big[|\bar{X}(t) - \tilde{X}(t)|^2\big] + \Delta t\Big). \tag{2.13}$$

We define a refined grid $\{t_h\}_{h=0}^{P}$ by the union $\{t_h\} \equiv \{\bar{t}_n\} \cup \{\tilde{t}_m\}$. Observe that both the functions $\Delta a(t)$ and $\Delta b(t)$ are adapted and piecewise constant on the refined grid. In view of the error representation (2.8), the estimates (2.12) and (2.13), and Itô isometry, we have

$$\begin{aligned}\mathbb{E}\big[|\bar{X}(s) - \tilde{X}(s)|^2\big] &\le 2\mathbb{E}\Big[\Big(\int_0^s \Delta a(t)\,dt\Big)^2 + \Big(\int_0^s \Delta b(t) dW(t)\Big)^2\Big] \\ &\le 2\mathbb{E}\Big[s\int_0^s |\Delta a(t)|^2\,dt\Big] + 2\int_0^s \mathbb{E}\Big[|\Delta b(t)|^2\Big]\,dt \\ &\le C\Delta t + C\int_0^s \mathbb{E}\big[|\bar{X}(t) - \tilde{X}(t)|^2\big]\,dt.\end{aligned}$$

An application of Gronwall's lemma then implies that

$$\mathbb{E}\big[|\bar{X}(s) - \tilde{X}(s)|^2\big] \le \Delta t C\exp(CT), \quad s < T$$

which finishes the proof. □

Theorem 2.1 implies that the Euler iterates $\{\bar{X}\}_{\Delta t_n}$ is a Cauchy sequence in $L^2(0,T;L^2(\Omega))$. Therefore it has a limit, say X. By Hölder's inequality, we get that

$$\int_0^t a(s, \bar{X}^{\Delta t_n}(s))\,ds \to \int_0^t a(s, X(s))\,ds \quad \text{in} \quad L^2(\Omega)$$

and by the Itô isometry it follows that

$$\int_0^t b(s, \bar{X}^{\Delta t_n}(s))\,dW(s) \to \int_0^t b(s, X(s))\,dW(s) \quad \text{in} \quad L^2(\Omega).$$

Hence we conclude that for all $t \in [0,T]$,

$$X(t) = X_0 + \int_0^t a(s, X(s))\, ds + \int_0^t b(s, X(s))\, dW(s) \quad \mathbb{P}\text{-a.s.}$$

In other words, $X(t)$ is a strong solution of the SDE (2.1).

Here is the MATLAB code for Euler-Maruyama approximation for linear SDE

$$dX(t) = \lambda X(t)\, dt + \mu X(t)\, dW(t); \quad X(0) = X_0 \tag{2.14}$$

whose exact solution is given by $X(t) = X_0 \exp((\lambda - \frac{1}{2}\mu^2)t + \mu W(t))$.

Exercise 2.1. *Show that $X(t)$ satisfies the following*

$$X(t) = X_0 \exp\Big(\mu W(t) + (\lambda - \frac{1}{2}\mu^2)t\Big).$$

Moreover, the expected value of the stock price $X(t)$ consequently agrees with the deterministic solution of (2.14) corresponding to $\mu = 0$.

Solution: Let $X(t)$ be the price of stock at time t satisfying the SDE (2.14). We apply Itô's formula to the functional $x \mapsto \log x$ to have

$$d(\log X) = \frac{dX}{X} - \frac{1}{2}\frac{\mu^2 X^2}{X^2}\, dt = \frac{dX}{X} - \frac{1}{2}\mu^2\, dt = \big[\lambda - \frac{1}{2}\mu^2\big]\, dt + \mu dW(t),$$

and therefore

$$\log X = \log X_0 + \big[\lambda - \frac{1}{2}\mu^2\big]t + \mu W(t) \implies X = X_0 \exp\Big(\mu W(t) + \big(\lambda - \frac{1}{2}\mu^2\big)t\Big).$$

Observe that,

$$\mathbb{E}[X(t)] = X_0 + \mu \int_0^t \mathbb{E}[X(s)]\, ds \implies \frac{d}{dt}\mathbb{E}[X(t)] = \lambda \mathbb{E}[X(t)]; \quad X(0) = X_0$$
$$\implies \log \frac{\mathbb{E}[X(t)]}{X_0} = \lambda t \implies \mathbb{E}[X(t)] = X_0 \exp(\lambda t).$$

This implies that the expected value of the stock price $X(t)$ consequently agrees with the deterministic solution of 2.14 corresponding to $\mu = 0$.

For numerical simulation, we consider $\lambda = 2$, $\mu = 1$ and $X_0 = 1$. We compute a discretized Brownian path over $[0,1]$ with $dt = 2^{-8}$. For convenience, we always choose the step size for numerical method to be an integer multiple of the increment dt for the Brownian path. This ensures that the set of points $\{t_j\}$ on which the discretized Brownian path is based contains the points $\{\tau_j\}$ at which the EM method is computed. Here $\tau_j = jRdt$ for some integer $R > 0$.

```
%Euler-Maruyama method on linear SDE
% SDE: dX(t)= lambda*X(t)dt + mu*X(t) dW(t); X(0)= Xzero
% Computations of exact solution
clc;
clear all;
randn( state , 100000);
lambda=2; mu=1; Xzero=1; % problem parameters
T=1; N=2^8; dt= 1/N;
dW= sqrt(dt)* randn(1,N); % creating Brownian increments
W=cumsum(dW); % randn(1,N) creates a 1-by-N array of independent N(0,1) samples
```

```
% and cumsum computes the cumulative sum of its argument
% For example, jth element of the 1-by-N array W is dW(1)+  + dW(j)
Xext=Xzero*exp((lambda-0.5*mu^2)*(dt:dt:T) + mu*W);
plot([0:dt:T],[Xzero, Xext],  g ), hold on
% Euler-Maruyama uses the time step R*dt
R=4; Dt=R*dt; L=N/R;  % L EM steps of size Dt
Xem=zeros(1,L); % preallocate for efficiency
% EM methods requires the increment W(t_j)-W(t_{j-1}) which is given by
% W(j*R*dt)-W((j-1)*R*dt)= \sum_{k=(j-1)*R +1}^{j*R} dW_k
Xtemp=Xzero;
for j=1:L
     Winc=sum(dW(R*(j-1)+1: j*R));
     Xtemp=Xtemp+ lambda* Dt* Xtemp + mu*Xtemp*Winc;
     Xem(j)=Xtemp;
end
Dtvals=dt*(2.^([0,9]));
plot([0:Dt:T],[Xzero, Xem],  r ), hold off
legend( Exact solution ,  Euler-Maruyama method )
xlabel( t ,  FontSize , 12)
ylabel( X , FontSize , 16,  Rotation , 0)
Emerr=abs(Xem(end)-Xext(end))
```

Definition 2.1. Let $\Delta t > 0$ be sufficiently small and $\tau = n\Delta t \in [0, T]$ is fixed .

- We say that a method has a *strong order of convergence* equal to γ, if there exists a constant $C > 0$ such that
$$\mathbb{E}\Big[\big|X_n - X(\tau)\big|\Big] \le C\Delta t^{\gamma}$$
- We say that a method a *weak order of convergence* equal to γ, if there exists a constant $C > 0$ such that for all smooth functions g having polynomial growth
$$\Big|\mathbb{E}\big[g(X_n)\big] - \mathbb{E}\big[g(X(\tau))\big]\Big| \le C\Delta t^{\gamma}$$

In view of the proof of Theorem 2.1, we see that Euler-Maruyama method has strong order of convergence $\gamma = \frac{1}{2}$. We will test it numerically for the linear SDE. We use 10 different time steps $\Delta t = 2^{p-1}dt$ for $1 \le p \le 10$ to evolute the error for 10000 different sample paths and then use the function *mean* to average over all the sample paths.

```
% Strong error of Euler-Maruyama method on linear SDE
% SDE: dX(t)= lambda*X(t)dt + mu*X(t) dW(t); X(0)= Xzero

clc;
clear all;
rng( default );
lambda=2; mu=1; Xzero=1; % problem parameters
T=1; N=2^8; dt= 1/N;
M=10000 % number of sample paths
Xerr=zeros(M, 10) % allocation for M-by-10 array
for s=1:M
    dW= sqrt(dt)* randn(1,N); % creating Brownian increments
    W=cumsum(dW); % discrete Brownian paths
    Xext=Xzero*exp((lambda-0.5*mu^2) + mu*W(end)); % evoluating exact solution at t=1
```

```
    for p=1:10
        % Euler−Maruyama uses the time step R*dt
        R= 2ˆ(p−1); Dt=R*dt; L=N/R;  % L EM steps of size Dt= Rdt
        Xtemp=Xzero;
        for j=1:L
            Winc=sum(dW(R*(j−1)+1: j*R));
            Xtemp=Xtemp+ lambda* Dt* Xtemp + mu*Xtemp*Winc;
            Xem(j)=Xtemp;
        end
    Xerr(s,p)=abs(Xtem−Xext) % strore the error at t=1
    end
end
Dtvals=dt*(2.ˆ([0:9]));
subplot(221)  % top LH picture
loglog(Dtvals, mean(Xerr) , b*− ), hold on % plots our approximation against
%Delta t on a log−log scale.
loglog(Dtvals,(Dtvals.ˆ(.5)), r−− ), hold off    % reference slope of 1/2
xlabel( \Delta t ) , ylabel( Sample average of |X(T)−X_L| )
title( emstrong.m ,  FontSize , 10)
```

2.2. **Weak error for Euler approximation.** We want to simulate the expected value $\mathbb{E}\big[g(X(T))\big]$ for a solution X of a given SDE (2.1) with a given function g. Let $\bar{X}$ be a forward Euler approximation of X given by (2.2)-(2.3). Then the error $\mathbb{E}\Big[g(X(T)) - g(\bar{X}(T))\Big]$ is called weak error or time discretization error. We want to discuss about this error. For this, we need some preparation. Consider the differential operator

$$
\begin{aligned}
&L : C^2([0,T]\times\mathbb{R}^d) \to C(\mathbb{R}) \\
&Lf(t,x) := \frac{1}{2}\sum_{i,j=1}^{d} \sigma_{ij}(t,x)\partial_{ij} f(t,x) + \sum_{i=1}^{d} a_i(t,x)\partial_i f(t,x)
\end{aligned}
$$

where $\sigma = bb^\top$ with strict ellipticity conditions: for all $x\in\mathbb{R}^d$, $\sum_{i,j=1}^d \sigma_{ij}(t,x)\xi_i\xi_j > 0$ for all $\xi = (\xi_1,\xi_2,\cdots,\xi_d)\in\mathbb{R}^d$ uniformly in $t\in[0,T]$, and a, b are smooth given functions related to the SDE (2.1).

Theorem 2.2. *Let a, b and g are smooth and bounded functions and X be the solution of the SDE (2.1). Define $u(t,x) = \mathbb{E}\Big[g(X(T))\Big|X(t)=x\Big]$. Then u is the solution of the Kolmogorov backward equation*

$$
\begin{aligned}
u_t(t,x) + Lu(t,x) &= 0, \quad t<T \\
u(T,x) &= g(x).
\end{aligned} \tag{2.15}
$$

Proof. Since a, b and g are smooth bounded functions and the operator L is uniformly elliptic, the equation (2.15) has an unique smooth solution, say $\bar{u}$. To prove the theorem, it is required to show that $\bar{u} = \mathbb{E}\Big[g(X(T))\Big|X(t)=x\Big]$. We apply Itô formula to the functional $\bar{u}(t,X(t))$ to have

$$\bar{u}(T,X(T)) - \bar{u}(t,X(t))$$

$$= \int_t^T \Big(\bar{u}_t(s, X(s)) + \frac{1}{2} \sum_{i,j=1}^d \sigma_{ij}(s, X(s)) \partial_{ij} \bar{u}(s, X(s)) + \sum_{i=1}^d a_i(s, X(s)) \partial_i \bar{u}(s, X(s)) \Big)\, ds$$
$$+ \int_t^T \nabla_x \bar{u}(s, X(s)) b(s, X(s))\, dW(s)$$
$$= \int_t^T \nabla_x \bar{u}(s, X(s)) b(s, X(s))\, dW(s) \quad (\text{as } \bar{u}_t(s, X(s)) + L\bar{u}(s, X(s)) = 0)$$

Taking expectation and noting that $\bar{u}(T, X(T)) = g(X(T))$, we obtain

$$\mathbb{E}\Big[g(X(T)) \Big| X(t) = x \Big] - \bar{u}(t, x) = \mathbb{E}\Big[\int_t^T \nabla_x \bar{u}(s, X(s)) b(s, X(s))\, dW(s) \Big| X(t) = x \Big] = 0.$$

Therefore, $\bar{u}(t, x) = \mathbb{E}\Big[g(X(T)) \Big| X(t) = x \Big]$ which finishes the proof. □

Exercise 2.2. *Let $d = m = 1$, and a, b, g, h and V are smooth and bounded functions. Let X be the solution of the SDE (2.1) for $d = m = 1$. Then show that*

$$u(t, x) = \mathbb{E}\Big[g(X(T)) e^{\int_t^T V(s, X(s))\, ds} \big| X(t) = x \Big] + \mathbb{E}\Big[- \int_t^T h(s, X(s)) e^{\int_t^s V(\tau, X(\tau))\, d\tau}\, ds \big| X(t) = x \Big]$$

solves the PDE

$$\begin{aligned} u_t(t, x) + a(t, x) u_x(t, x) + \frac{1}{2} b^2(t, x) u_{xx}(t, x) + V(t, x) u(t, x) = h(t, x), \quad t < T \\ u(T, x) = g(x). \end{aligned} \tag{2.16}$$

Solution: Since the functions a, b, g, h and V are smooth and bounded, the PDE (2.16) is a linear second order (uniformly elliptic) parabolic PDE, and hence it has a unique smooth solution, say $\bar{u}$. Let $G(s) = e^{\int_t^s V(\tau, X(\tau))\, d\tau}$. Apply Itô product rule, we have

$$d(\bar{u}(s, X(s)) G(s)) = G(s) \Big(L\bar{u}(s, X(s)) + \bar{u}(s, X(s)) V(s, X(s)) \Big)\, ds + G(s) b(s, X(s)) \bar{u}_x(s, X(s)) dW(s),$$

where L is given in Theorem 2.2. Integrating both sides from t to T and noting the fact that $\bar{u}$ is the solution of (2.16), we obtain

$$\begin{aligned} &\bar{u}(T, X(T)) e^{\int_t^T V(s, X(s))\, ds} - \bar{u}(t, x) \\ &= \int_t^T h(s, X(s)) e^{\int_t^s V(\tau, X(\tau))\, d\tau}\, ds + \int_t^T e^{\int_t^s V(\tau, X(\tau))\, d\tau} a(s, X(s)) \bar{u}(s, X(s))\, dW(s) \end{aligned}$$

Taking conditional expectation given $X(t) = x$, we arrive at

$$\bar{u}(t, x) = \mathbb{E}\Big[g(X(T)) e^{\int_t^T V(s, X(s))\, ds} \big| X(t) = x \Big] + \mathbb{E}\Big[- \int_t^T h(s, X(s)) e^{\int_t^s V(\tau, X(\tau))\, d\tau}\, ds \big| X(t) = x \Big].$$

Since the PDE has an unique solution, we conclude that $u(t, x)$ solves the PDE (2.16).

Let us now focus on the time discretization error $\mathbb{E}\Big[g(X(T)) - g(\bar{X}(T)) \Big]$. Regarding this, we have the following theorem

Theorem 2.3. *Assume that a, b and g are smooth and decay sufficiently fast as $|x| \to \infty$, and X be the solution of the SDE (2.1). Let $\bar{X}$ be a forward Euler approximation given by (2.2)-(2.3). Then there holds*

$$\mathbb{E}\Big[g(X(T)) - g(\bar{X}(T))\Big] \le \mathcal{O}(\max \Delta t).$$

Proof. In view of Theorem 2.2, we see that $u(t,x) = \mathbb{E}\Big[g(X(T))\big|X(t) = x\Big]$ solves the PDE (2.15). In particular

$$u(0, X(0)) = \mathbb{E}\big[g(X(T))\big]. \tag{2.17}$$

Now, an application of Itô formula applied to the function $u(t, \bar{X}(t))$ gives

$$\begin{aligned} du(t,\bar{X}(t)) &= \Big\{u_t(t,\bar{X}(t)) + \frac{1}{2}\sum_{i,j=1}^d \bar{\sigma}_{ij}(t,\bar{X})\partial_{ij}u(t,\bar{X}(t)) + \sum_{i=1}^d \bar{a}_i(t,\bar{X})\partial_i u(t,\bar{X}(t))\Big\}\, dt \\ &\qquad + \nabla_x u(t,\bar{X}(t))\bar{b}(t,\bar{X})\, dW(t) \\ &= \Big\{\frac{1}{2}\sum_{i,j=1}^d \big(\bar{\sigma}_{ij}(t,\bar{X}) - \sigma_{ij}(t,\bar{X}(t))\big)\partial_{ij}u(t,\bar{X}(t)) + \sum_{i=1}^d \big(\bar{a}_i(t,\bar{X}) - a_i(t,\bar{X}(t))\big)\partial_i u(t,\bar{X}(t))\Big\}\, d \\ &\qquad + \nabla_x u(t,\bar{X}(t))\bar{b}(t,\bar{X})\, dW(t) \end{aligned}$$

Integrating from 0 to T and then taking expectation, we have, in view of (2.17)

$$\begin{aligned} \mathbb{E}\Big[g(\bar{X}(T)) - g(X(T))\Big] &= \int_0^T \mathbb{E}\Big[\frac{1}{2}\sum_{i,j=1}^d \big(\bar{\sigma}_{ij}(t,\bar{X}) - \sigma_{ij}(t,\bar{X}(t))\big)\partial_{ij}u(t,\bar{X}(t))\Big]\, dt \\ &\qquad + \int_0^T \mathbb{E}\Big[\sum_{i=1}^d \big(\bar{a}_i(t,\bar{X}) - a_i(t,\bar{X}(t))\big)\partial_i u(t,\bar{X}(t))\Big]\, dt. \end{aligned} \tag{2.18}$$

Let $\Delta t_n = t_{n+1} - t_n$ where $\{t_n\}_{n=0}^N$ are nodal points of a grid of the interval $[0,T]$. We first show that

$$f_1(t) := \mathbb{E}\Big[\sum_{i=1}^d \big(\bar{a}_i(t,\bar{X}) - a_i(t,\bar{X}(t))\big)\partial_i u(t,\bar{X}(t))\Big] = \mathcal{O}(\Delta t_n), \quad t_n \le t < t_{n+1}. \tag{2.19}$$

Since $\bar{a}(t,\bar{X}) = a(t_n, \bar{X}(t_n))$, we see that $f_1(t_n) = 0$. Therefore, (2.19) will hold if we show that $|f_1'(t)| \le C$ for some constant $C > 0$. To do so, let $\alpha(t,x) = \sum_{i=1}^d \big(a_i(t_n,\bar{X}(t_n)) - a_i(t,x)\big)\partial_i u(t,x)$ so that $f(t) = \mathbb{E}\Big[\alpha(t,\bar{X}(t))\Big]$. By Itô formula we have

$$f'(t) = \frac{d}{dt}\mathbb{E}\Big[\alpha(t,\bar{X}(t))\Big] = \Big[\alpha_t(t,\bar{X}(t)) + \bar{a}(t,\bar{X}(t))\nabla_x\alpha(t,\bar{X}) + \frac{1}{2}\sum_{i,j=1}^d \bar{\sigma}_{ij}(t,\bar{X})\partial_{ij}\alpha(t,\bar{X}(t))\Big].$$

Since a, b and g are smooth functions and decay sufficiently fast as $|x| \to \infty$, we see that $f'(t) = \mathcal{O}(1)$. In other words, there exists a constant $C > 0$ such that $|f'(t)| \le C$ for

$t_n < t < t_{n+1}$. Consequently (2.19) holds true. Similarly, one can show that

$$f_2(t) := \mathbb{E}\Big[\frac{1}{2}\sum_{i,j=1}^{d}\big(\bar{\sigma}_{ij}(t,\bar{X}) - \sigma_{ij}(t,\bar{X}(t))\big)\partial_{ij}u(t,\bar{X}(t))\Big] = \mathcal{O}(\Delta t_n), \quad t_n \le t < t_{n+1}. \tag{2.20}$$

We combine (2.19)-(2.20) in (2.18) to conclude

$$\mathbb{E}\Big[g(X(T)) - g(\bar{X}(T))\Big] \le \mathcal{O}(\max \Delta t).$$

□

We will test the weak rate of convergence numerically for linear SDE. It is worth emphasizing that for the computations of weak rate of convergence, we use different paths for each step size Δt as weak convergence concerns only the mean of the solution, and so we are free to use any $\sqrt{\Delta t}\mathcal{N}(0,1)$ sample for the increment $W(\tau_j) - W(\tau_{j-1})$ on any step.

Exercise 2.3. *Show that* $\mathbb{E}[X(t)] = X_0 \exp(\lambda t)$, *where* X *is a strong solution of the Black-Scholes equation* (2.14).

```
% Weak error of Euler—Maruyama method on linear SDE
% SDE: dX(t)= lambda*X(t)dt + mu*X(t) dW(t); X(0)= Xzero

clc;
clear all;
randn( state , 100);
lambda=2; mu=0.1; Xzero=1; % problem parameters
T=1;
M=10000 % number of sample paths

Xem=zeros(5, 1);
for p=1:5
    Dt=2^(p—10);
     L=T/Dt; % number of Euler steps of size dt
     Xtemp=Xzero*ones(M,1)
     for j=1:L
          Winc=sqrt(Dt)*randn(M,1);
          Xtemp=Xtemp + Dt*lambda*Xtemp + mu*Xtemp.*Winc;
     end
     Xem(p)=mean(Xtemp);
end
Xerr=abs(Xem—exp(lambda));

Dtvals=2.^([1:5]—10);
 % subplot(222)  % top RH picture
loglog(Dtvals, Xerr, b*— ), hold on
loglog(Dtvals, Dtvals, r— ), hold off    % reference slope of 1
xlabel( \Delta t ) , ylabel(  |E(X(T))— Sample average of X_L| )
title( emweak.m ,  FontSize , 10)
```

3. Markov process

An useful property of a process is Markov property. To define the Markov property, one needs to define so called *transition function* and *semi- group*.

Definition 3.1. (Transition function) A transition funtion $\{\pi_t : t \ge 0\}$ on $\mathbb{R}^n$ is a family of kernels $\pi_t : \mathbb{R}^n \times \mathcal{B}(\mathbb{R}^n) \to [0,1]$ such that

i) For $t \ge 0$ and $x \in \mathbb{R}^n$), $\pi_t(x,\cdot)$ is a probability measure on $\mathbb{R}^n$.
ii) For $t \ge 0$ and $A \in \mathcal{B}(\mathbb{R}^n)$, the mapping $x \mapsto \pi_t(x,A)$ is measurable.
iii) For $s,t \ge 0$, $x \in \mathbb{R}^n$, and $A \in \mathcal{B}(\mathbb{R}^n)$

$$\pi_{t+s}(x,A) = \int_{\mathbb{R}^n} \pi_t(y,A)\pi_s(x,dy). \tag{3.1}$$

The equation (3.1) is called the **Chapman-Kolmogorov** equation.

Given a transition functions $\{\pi_t : t \ge 0\}$ on $\mathbb{R}^n$, we can define a one-parameter family of linear operators $\big(\Pi_t\big)_{t\ge 0}$ from the space of bounded Borel functions into itself as follows:

$$(\Pi_t f)(x) := \int_{\mathbb{R}^n} f(y)\pi_t(x,dy). \tag{3.2}$$

Exercise 3.1. *Let $\{\pi_t : t \ge 0\}$ be a transition function on $\mathbb{R}^n$ and Π_t be its associated one-parameter family of linear operators defined by (3.2). Then the following properties are satisfied:*

a) $\Pi_t \boldsymbol{1} = \boldsymbol{1}$; $\quad \Pi_t f \ge 0$ *if* $f \ge 0$; $\quad \|\Pi_t\|_{\mathcal{L}(\mathcal{B}_b(\mathbb{R}^n))} \le 1$ *for all* $t \ge 0$.
b) The semi-group property holds: for every $s,t \ge 0$,

$$\Pi_{t+s} = \Pi_t \Pi_s.$$

Definition 3.2. (Markov process) A stochastic process $(X_t)_{t\ge 0}$ defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ is called a Markov process if there exists a transition function $\{\pi_t : t \ge 0\}$ on $\mathbb{R}^n$ such that for every bounded Borel function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}\Big(f(X_{t+s})|\mathcal{F}_s\Big) = (\Pi_t f)(X_s).$$

Proposition 3.1. *Let $(X_t)_{t\ge 0}$ be a one dimensional stochastic process defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ such that $X_0 = 0$ a.s. Then $(X_t)_{t\ge 0}$ is a standard Brownian motion if and only if it is a Markov process with semi group:*

$$\Pi_0 = \mathrm{Id}; \quad (\Pi_t f)(x) = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} f(y) \exp(-\frac{(x-y)^2}{2t})\,dy, \quad t > 0,\ x \in \mathbb{R}.$$

Proof. Let $(X_t)_{t\ge 0}$ be a Markov process defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ with given semi group. Let $(\mathcal{F}_t)_{t\ge 0}$ be the natural filtration of $(X_t)_{t\ge 0}$. Since $(X_t)_{t\ge 0}$ is a Markov process, we have

$$\begin{aligned}
&\mathbb{E}\Big[\exp(i\lambda X_{t+s})|\mathcal{F}_s\Big] = \Pi_t \exp(i\lambda X_s) = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} \exp(i\lambda(X_s + y)) \exp(-\frac{y^2}{2t})\,dy \\
\implies &\mathbb{E}\Big[\exp(i\lambda(X_{t+s} - X_s))|\mathcal{F}_s\Big] = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} \exp(i\lambda y) \exp(-\frac{y^2}{2t})\,dy = \exp(-\frac{\lambda^2 t}{2}).
\end{aligned}$$

This shows in particular that the increments of $(X_t)_{t\geq 0}$ are stationary and independent. Now for $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$, $) < t_1 < \cdots < t_n$

$$\mathbb{E}\Big[\exp\big(i\sum_{k=1}^n \lambda_k(X_{t_{k+1}} - X_{t_k})\big)\Big] = \prod_{k=1}^n \mathbb{E}\big[\exp\big(i\lambda_k(X_{t_{k+1}} - X_{t_k})\big)\big]$$

$$= \prod_{k=1}^n \mathbb{E}\big[\exp\big(i\lambda_k X_{t_{k+1}-t_k}\big)\big] = \exp\big(-\frac{1}{2}\sum_{k=1}^n (t_{k+1}-t_k)\lambda_k^2\big).$$

Hence $(X_t)_{t\geq 0}$ is a standard Brownian motion.

Conversely, let $(X_t)_{t\geq 0}$ be a standard Brownian motion with natural filtration $\mathcal{F}_t$. If f is a bounded Borel function and $s, t \geq 0$, then we have

$$\mathbb{E}\Big(f(X_{t+s})|\mathcal{F}_s\Big) = \mathbb{E}\Big(f(X_{t+s} - X_s + X_s)|\mathcal{F}_s\Big) = \mathbb{E}\Big(f(X_{t+s} - X_s + X_s)|X_s\Big)$$

where the last equality follows from the fact that $X_{t+s} - X_s$ is $\mathcal{F}_s$ independent and X_s is $\mathcal{F}_s$-measurable. In other words, we obtain $\mathbb{E}\Big(f(X_{t+s})|\mathcal{F}_s\Big) = \mathbb{E}\Big(f(X_{t+s})|X_s\Big)$. For $x \in \mathbb{R}$, we have

$$\mathbb{E}\Big[f(X_{t+s})\big|X_s = x\Big] = \mathbb{E}\Big[f(X_{t+s} - X_s + X_s)\big|X_s = x\Big] = \mathbb{E}\Big[f(Y_t + x)\Big] \tag{3.3}$$

where Y_t is a random variable independent from X_s and normally distributed with mean 0 and variance t. Therefore we have

$$\mathbb{E}\Big[f(X_{t+s})\big|X_s = x\Big] = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} f(x+y)\exp(-\frac{y^2}{2t})\,dy$$

$$\mathbb{E}\Big[f(X_{t+s})\big|\mathcal{F}_s\Big] = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} f(y+X_s)\exp(-\frac{y^2}{2t})\,dy = \Pi_t f(X_s).$$

i.e., (X_t) is a Markov process.

□

The following theorem shows how to construct martingales from solutions of the backward heat equation.

Theorem 3.2. *Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. Let $f : [0,\infty) \times \mathbb{R} \to \mathbb{C}$ be a function such that $f \in C^{1,2}([0,\infty) \times \mathbb{R}; \mathbb{C})$, and for $t \geq 0$, there exist $K > 0$ and $\alpha > 0$ such that*

$$\sup_{0\leq s\leq t} |f(s,x)| \leq K e^{\alpha|x|}.$$

Then the process $(f(t, B_t))_{t\geq 0}$ is a martingale if and only if

$$\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} = 0. \tag{3.4}$$

Proof. Let $t > 0$. In what follows, we denote by $\mathbb{F} = (\mathcal{F}_t)_{t>0}$ the natural filtration of the Brownian motion. Since the density of the normally distributed random variable with mean 0 and variance t is $g(x) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} \exp(-\frac{(x-y)^2}{2t})\,dy$, by using independent of the Brownian increments , we see that for $s < t$

$$\mathbb{E}\Big[f(t, B_t)\big|\mathcal{F}_s\Big] = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi(t-s)}} f(t,y)\exp\Big(-\frac{(y-B_s)^2}{2(t-s)}\Big)\,dy.$$

Therefore, the process $(f(t,B_t))_{t\ge 0}$ is a martingale if and only if for $0<s<t$ and $x\in\mathbb{R}$

$$f(s,x)=\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi(t-s)}}f(t,y)\exp\Big(-\frac{(y-x)^2}{2(t-s)}\Big)\,dy. \tag{3.5}$$

We are thus led to characterize the functions f that satisfy the above functional equation (3.5) and the given growth conditions. So, let f be a function that satisfy the given conditions and the functional equation (3.5). For a fixed $t>0$, the function

$$g:(s,x)\mapsto\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi(t-s)}}f(t,y)\exp\Big(-\frac{(y-x)^2}{2(t-s)}\Big)\,dy$$

which is defined on $[0,t)\times\mathbb{R}$ satisfies the equation $\frac{\partial g}{\partial s}+\frac{1}{2}\frac{\partial^2 g}{\partial x^2}=0$. Since $f(s,x)=g(s,x)$, we see that f is a solution of the backward heat equation (3.4).

Conversely, let f satisfy the required growth conditions and the equation (3.4). Let $t>0$ is fixed. Then the function $h:=f-g$ where g is defined as above satisfies $\frac{\partial h}{\partial s}+\frac{1}{2}\frac{\partial^2 h}{\partial x^2}=0$. Moreover, for all $x\in\mathbb{R}$,

$$\lim_{s\to t}h(s,x)=\lim_{s\to t}\big(f(s,x)-g(s,x)\big)=0.$$

From the classical uniqueness of solution results for the heat equation, we deduce that $h\equiv 0$. In other words, (3.5) is satisfied. This completes the proof. □

Exercise 3.2. *Let $(X_t)_{t\ge0}$ be a Markov process with semi group Π_t. Show that if $T>0$, the process $(\Pi_{T-t}f(X_t))_{t\ge0}$ is a martingale. By using Doob's stopping theorem, deduce that if S is a stopping time such that $S\le T$ a.s., then $\mathbb{E}\Big[f(X_T)\big|\mathcal{F}_S\Big]=\Pi_{T-S}f(X_S)$, where f is a bounded Borel function.*

Solution: We have seen that for any $t\ge0$, Π_t is a contraction mapping from the space of bounded Borel functions into itself. Therefore, $\Pi_t f$ is a bounded Borel function. Since $(X_t)_{t\ge0}$ is a Markov process, we have

$$\mathbb{E}\Big[\Pi_{T-t}f(X_t)\big|\mathcal{F}_s\Big]=\mathbb{E}\Big[\Pi_{T-t}f(X_{t-s+s})\big|\mathcal{F}_s\Big]=\Pi_{t-s}\Pi_{T-t}f(X_s)=\Pi_{T-s}f(X_s).$$

In other words, the process $(\Pi_{T-t}f(X_t))_{t\ge0}$ is a martingale.

As a Corollary of Doob's martingale theorem, we have the following: let T_1 and T_2 be two bounded stopping times with $T_1\le T_2$ and $(M_t)_{t\ge0}$ be a continuous martingale with $\mathbb{E}\big[M_{T_1}\big],\,\mathbb{E}\big[M_{T_2}\big]<\infty$, then $\mathbb{E}\Big[M_{T_2}\big|\mathcal{F}_{T_1}\Big]=M_{T_1}$. In our case, $M_t=\Pi_{T-t}f(X_t)$ and $T_1=S$ and $T_2=T$. Thus,

$$\mathbb{E}\Big[f(X_T)\big|\mathcal{F}_S\Big]=\mathbb{E}\Big[\Pi_{T-T}f(X_T)\big|\mathcal{F}_S\Big]=\mathbb{E}\Big[M_T\big|\mathcal{F}_S\Big]=M_S=\Pi_{T-S}f(X_S).$$

It is remarkable that given any transition function, it is always possible to find a corresponding Markov process.

Theorem 3.3. *Let $\big\{\pi_t:t\ge0\big\}$ be a transition function on $\mathbb{R}^n$. Let ν be a probability measure on $\mathbb{R}^n$. Then there exists a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a stochastic process $(X_t)_{t\ge0}$ such that the following hold:*

i) $\mathscr{L}(X_0)=\nu$

ii) If $f : \mathbb{R}^n \to \mathbb{R}$ is a bounded Borel function then, for all $s, t \geq 0$, there holds $\mathbb{E}\Big[f(X_{t+s})\big|\mathcal{F}_s^X\Big] = \Pi_t f(X_s)$, where $\mathcal{F}^X$ is the natural filtration of X.

Proof. Let A be a Borel set in $\mathbb{R}^n$ and B a Borel set in $(\mathbb{R}^n)^{\otimes m}$. For $0 = t_0 < t_1 < \cdots < t_m$, we define

$$\mu_{t_0,t_1,\cdots,t_m}(A \times B) := \int_A \int_B \pi_{t_1}(z, dx_1)\pi_{t_2-t_1}(x_1, dx_2) \cdots \pi_{t_m-t_{m-1}}(x_{m-1}, dx_m)\,\nu(dz).$$

The measure $\mu_{t_0,t_1,\cdots,t_m}$ is therefore a probability measure on $\mathbb{R}^n \times (\mathbb{R}^n)^{\otimes m}$. Since for a Borel set C in $\mathbb{R}^n$ and $x \in \mathbb{R}^n$

$$\pi_{t+s}(x, C) = \int_{\mathbb{R}^n} \pi_t(y, C)\pi_s(x, dy)$$

we deduce that this family of probability measures satisfy the assumptions of the Daniell-Kolmogorov theorem. Therefore we can find a process $(X_t)_{t\geq 0}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ whose finite dimensional distributions are given by $\mu_{t_0,t_1,\cdots,t_m}$. Now,

$$\mu_0(A) = \int_A \nu(dz) = \nu(A) \quad \forall A \in \mathcal{B}(\mathbb{R}^n).$$

In this case, $\Omega =$ set of all functions $[0, \infty) \mapsto \mathbb{R}^n$, $\mathcal{F} = \sigma$-algebra generated by cylindrical sets, and $\mathbb{P}(X_{t_1} \in A_1, \cdots X_{t_m} \in A_m) = \mu_{t_1,t_2,\cdots,t_m}(A_1 \times A_2 \times \cdots \times A_m)$ with $X_t(f) = f(t)$, the canonical process. In other words, i) holds true.

To prove ii) we have to show that if $f : \mathbb{R}^n \to \mathbb{R}$ and $F : (\mathbb{R}^n)^{\otimes m} \to \mathbb{R}$ are bounded Borel functions and if $0 = t_0 < t_1 < \cdots < t_m$, then

$$\mathbb{E}\Big[f(X_{t_m})F(X_{t_0}, \cdots X_{t_{m-1}})\Big] = \mathbb{E}\Big[\Pi_{t_m-t_{m-1}} f(X_{t_{m-1}})F(X_{t_0}, \cdots X_{t_{m-1}})\Big]. \tag{3.6}$$

Thanks to Fubini's theorem we have

$$\begin{aligned}
&\mathbb{E}\Big[f(X_{t_m})F(X_{t_0}, \cdots X_{t_{m-1}})\Big] \\
&= \int_{(\mathbb{R}^n)^{\otimes(m+1)}} f(x_m)F(z, x_1, \cdots, x_{m-1})\pi_{t_1}(z, dx_1)\pi_{t_2-t_1}(x_1, dx_2) \\
&\qquad\qquad \cdots \pi_{t_m-t_{m-1}}(x_{m-1}, dx_m)\,\nu(dz) \\
&= \int_{\mathbb{R}^n} \int_{(\mathbb{R}^n)^{\otimes m}} \Pi_{t_m-t_{m-1}} f(x_{m-1})F(z, x_1, \cdots, x_{m-1})\pi_{t_1}(z, dx_1)\pi_{t_2-t_1}(x_1, dx_2) \\
&\qquad\qquad \cdots \pi_{t_{m-1}-t_{m-2}}(x_{m-2}, dx_{m-1})\,\nu(dz) \\
&= \mathbb{E}\Big[\Pi_{t_m-t_{m-1}} f(X_{t_{m-1}})F(X_{t_0}, \cdots X_{t_{m-1}})\Big].
\end{aligned}$$

This completes the proof of the theorem. □

Definition 3.3. (Strong Markov property) Let $(X_t)_{t\geq 0}$ be a Markov process with transition function $\big\{\pi_t : t \geq 0\big\}$. We say that $(X_t)_{t\geq 0}$ is a *strong Markov process* if for any bounded Borel function $F : \mathbb{R}^n \to \mathbb{R}$ and any finite stopping time S of the filtration $(\mathcal{F}_t^X)_{t\geq 0}$ one has

$$\mathbb{E}\Big[f(X_{S+t})\big|\mathcal{F}_S^X\Big] = \Pi_t f(X_S).$$

Proposition 3.4. *Let $(B_t)_{t\geq 0}$ be a standard Brownian motion and T be a finite stopping time. The process $(B_{T+t} - B_T)_{t\geq 0}$ is a standard Brownian motion, independent from $\mathcal{F}_T^B$.*

Proof. Let T be a finite stopping time of the filtration $\{\mathcal{F}_t^B\}_{t\ge 0}$. We first assume that T bounded. Define

$$\widetilde{B}_t := B_{T+t} - B_T, \quad t \ge 0.$$

Let $\lambda \in \mathbb{R}$. Define $M_t := \exp(i\lambda B_t + \frac{\lambda^2}{2}t)$. Them M_t is a martingale. Let $0 \le s \le t$. Then $t+T$ and $s+T$ are two stopping times. By Doob's stopping theorem we have

$$\mathbb{E}\Big[M_{t+T}\big|\mathcal{F}_{s+T}\Big] = M_{s+T}.$$

This show that

$$\mathbb{E}\Big[\exp\Big(i\lambda(B_{t+T} - B_{s+T})\Big)\big|\mathcal{F}_{s+T}\Big] = \exp\big(-\frac{\lambda^2}{2}(t-s)\big).$$

In other words, the increments of $(\widetilde{B}_t)_{t\ge 0}$ are independent and stationary. Moreover, for $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$, $0 < t_1 < \cdots < t_n$

$$\begin{aligned}\mathbb{E}\Big[\exp\big(i\sum_{k=1}^n \lambda_k(\widetilde{B}_{t_{k+1}} - \widetilde{B}_{t_k})\big)\Big] &= \prod_{k=1}^n \mathbb{E}\big[\exp\big(i\lambda_k(\widetilde{B}_{t_{k+1}} - \widetilde{B}_{t_k})\big)\big] \\ &= \exp\big(-\frac{1}{2}\sum_{k=1}^n (t_{k+1}-t_k)\lambda_k^2\big).\end{aligned}$$

Hence $(\widetilde{B}_t)_{t\ge 0}$ is a standard Brownian motion, independent from $\mathcal{F}_T^B$. If T is not bounded almost surely, then we can consider the stopping time $T \wedge N$ and from the previous step, the finite-dimensional distributions $\Big(B_{t_1+(T\wedge N)} - B_{T\wedge N}, \cdots, B_{t_n+(T\wedge N)} - B_{t_{n-1}+(T\wedge N)}\Big)$ do not depend on N and are the same as a Brownian motion. We can then let $N \to \infty$ to conclude the result. $\square$

Exercise 3.3. *$(B_t)_{t\ge 0}$ is a strong Markov process.*

Solution: Let f is a bounded Borel function and $t \ge 0$. Let S be a finite stopping time. From Proposition 3.4, we have

$$\mathbb{E}\Big(f(B_{t+S})|\mathcal{F}_S\Big) = \mathbb{E}\Big(f(B_{t+S} - B_S + B_S)|\mathcal{F}_S\Big).$$

Since $B_{t+S} - B_S$ is independent from $\mathcal{F}_S$, we first deduce that

$$\mathbb{E}\Big(f(B_{t+S})|\mathcal{F}_S\Big) = \mathbb{E}\Big(f(B_{t+S})|B_S\Big).$$

For $x \in \mathbb{R}$, we have

$$\mathbb{E}\Big[f(B_{t+S})\big|B_S = x\Big] = \mathbb{E}\Big[f(B_{t+S} - B_S + B_S)\big|B_S = x\Big] = \mathbb{E}\Big[f(Y_t + x)\Big] \tag{3.7}$$

where Y_t is a random variable independent from B_S and normally distributed with mean 0 and variance t. Therefore we have

$$\begin{aligned}\mathbb{E}\Big[f(B_{t+S})\big|B_S = x\Big] &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} f(x+y)\exp(-\frac{y^2}{2t})\,dy \\ \mathbb{E}\Big[f(B_{t+S})\big|\mathcal{F}_S\Big] &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} f(y+B_S)\exp(-\frac{y^2}{2t})\,dy = \Pi_t f(B_S).\end{aligned}$$

i.e., (B_t) is a strong Markov process.

Exercise 3.4. *Let $(B_t)_{t\geq 0}$ be a standard Brownian motion on $\mathbb{R}$. For any $a > 0$, define the hitting time that Brownian motion hits a i.e.,*

$$T_a := \inf\{t \geq 0 : B_t = a\}$$

i) Show that T_a is almost surely finite stopping time.
ii) Show that the process $(\widetilde{B}_t)_{t\geq 0}$ given by

$$\widetilde{B}_t = B_t \chi_{t\leq T_a} + (2a - B_t)\chi_{t>T_a}$$

is again a standard Brownian motion.

Solution: Notice that

$$\{T_a \leq t\} = \{\exists\, s \in [0,t] : W_s = a\} \in \mathcal{F}_t.$$

Thus, T_a is a stopping time. Since $\mathbb{P}\{B_t = a\} = 0$ we obtain

$$\mathbb{P}\{T_a \leq t\} = \mathbb{P}\{T_a \leq t, B_t > a\} + \mathbb{P}\{T_a \leq t, B_t < a\}.$$

Now any path that starts at 0 and ends above $a > 0$ must cross a, and so $\mathbb{P}\{T_a \leq t, B_t > a\} = \mathbb{P}\{B_t > a\}$. Since T_a is a stopping time, the strong Markov property of the Brownian motion implies that $X_t := B_{T_a+t} - B_{T_a} = B_{T_a+t} - a$ is a Brownian motion, independent of B_r, $r \leq T_a$. Again, since the normal distribution is symmetric around zero, we have $\mathbb{P}\{T_a \leq t, B_t < a\} = \mathbb{P}\{T_a \leq t, B_t > a\} = \mathbb{P}\{B_t > a\}$. Therefore, we have

$$\mathbb{P}\{T_a \leq t\} = 2\mathbb{P}\{B_t > a\} = 2\int_a^\infty \exp\big(-\frac{x^2}{2t}\big)\frac{dx}{\sqrt{2\pi t}} = \int_0^t (2\pi s^3)^{-\frac{1}{2}} a \exp(-\frac{a^2}{2s})\, ds$$

where the last equality follows by substituting $x = a\sqrt{\frac{t}{s}}$ in the second equality. In other words, the distribution and the density functions of T_a are given by $F_{T_a}(t) := 2\int_a^\infty \exp\big(-\frac{x^2}{2t}\big)\frac{dx}{\sqrt{2\pi t}}\, dx$ and $f_{T_a}(t) := (2\pi t^3)^{-\frac{1}{2}} a \exp(-\frac{a^2}{2t})$ respectively. In particular, $\mathbb{P}\{T_a < \infty\} = 1$. This completes the proof of i).

To prove ii), we consider the processes $X := (B_t)_{t\leq T_a}$, $Y = (B_{t+T_a} - a)_{t\geq 0}$, and $Z = -Y$. By the strong Markov property, Y is a Brownian motion and independent of X. Hence X is independent of Z. Thus $(X, Y) \overset{d}{=} (X, Z)$. Concatenating X with Y gives B while Concatenating X with Z gives $\widetilde{B}$. Hence $\widetilde{B} \overset{d}{=} B$. In other words, $\widetilde{B}_t = B_t\chi_{t\leq T_a} + (2a - B_t)\chi_{t>T_a}$ is a standard Brownian motion. This completes the proof.

3.1. **Expansion of the global weak error for Euler scheme.** We have seen that for sufficiently smooth f such that it decays sufficiently fast as $|x| \to \infty$,

$$\big|E_{\rm rr}(T,k)\big| = \Big|\mathbb{E}\Big[f(X_T) - f(Y_N^k)\Big]\Big| \leq C_T k$$

where $X = \{X_t : 0 \leq t \leq T\}$ is a strong solution to the SDE (2.1), while $Y^k = \{Y_j^k : 1 \leq j \leq N\}$ is a Euler approximation with size $k > 0$ associated to the SDE (2.1). We want to study $E_{\rm rr}(T,k)$ as a function of k.

Let us consider the class F_T of functions $\phi : [0,T] \times \mathbb{R}^n \to \mathbb{R}$ with the following properties: ϕ is of class C^∞, and for some positive integer s and positive C_T such that

$$\big|\phi(\theta,x)\big| \leq C_T\big(1 + |x|^s\big) \qquad \forall\, \theta \in [0,T] \text{ and } x \in \mathbb{R}^n.$$

We say that $\phi \in F_T$ is *homogeneous* if it does not depend on the time variable, i.e., $\phi(t,x)=\phi(x)$ for all $t\in[0,T]$. We consider the differential operator $\mathcal{L}$ associated to the SDE (2.1)

$$\mathcal{L} := \frac{1}{2}\sum_{i,j=1}^{d}\sigma_{ij}(t,x)\partial_{ij} + \sum_{i=1}^{d} a_i(t,x)\partial_i$$

where $\sigma = bb^\top$ with strict ellipticity conditions. Consider the functions a and b are C^∞ functions whose derivatives of any order are bounded. Consider a homogeneous function f of F_T, and define $u(t,x)=\mathbb{E}\big[f(X_T^{t,x})\big]$ where $X_s^{t,x}$ is a strong solution of the SDE (2.1) started at the point x at time t. Then, thanks to Theorem 2.2, u satisfies the following Kolmogorov backward equation

$$\begin{cases}\partial_t u + \mathcal{L}u = 0\\ u(T,x) = f(x).\end{cases}$$

Moreover, since a, b are smooth functions with bounded derivatives, and $f\in F_T$, there exist positive constants $m_\alpha(T)$ and $C_\alpha(T)$ such that

$$\big|\partial_\alpha u(t,x)\big| \le C_\alpha(T)\big(1+|x|^{m_\alpha(T)}\big) \tag{3.8}$$

where ∂_α denotes the mixed partial derivatives of order $|\alpha|$

$$\frac{\partial^{|\alpha|}}{\partial_1^{\alpha_1}\cdots\partial_r^{\alpha_r}} \quad \text{with } \alpha=(\alpha_1,\cdots,\alpha_r), \text{ and } |\alpha| = \sum_{i=1}^{r}\alpha_i.$$

To study the expansion of the weak error in terms to discretized parameter k, we need the higher order moments for the Euler approximation. We have shown (in the proof of existence and uniqueness theorem for SDE via Euler method) that there exists a constant $C>0$, independent of $k>0$ such that $\sup\limits_{1\le i\le N}\mathbb{E}\big[\|Y_i^k\|_{\mathbb{R}^n}^2\big]\le C$. Moreover, by using continuified Euler approximation along with BDG inequality, one can show that for any positive integer $m\ge 2$, there exists a positive constant C_m, independent of k such that

$$\sup_{1\le i\le N}\mathbb{E}\Big[\|Y_i^k\|_{\mathbb{R}^n}^m\Big] \le \exp(C_m T). \tag{3.9}$$

Note that $u(0,X_0)=u(0,Y_0^k)=\mathbb{E}\big[f(X_T)\big]$. Thus the weak error $E_{\rm rr}(T,k)$ takes the form

$$E_{\rm rr}(T,k) = \mathbb{E}\big[u(T,Y_N^k)\big] - \mathbb{E}\big[u(0,Y_0^n)\big] = \sum_{i=1}^{N}\mathbb{E}\Big[u(t_i,Y_i^k)-u(t_{i-1},Y_{i-1}^k)\Big].$$

We compute $\mathbb{E}\Big[u(t_i,Y_i^k)-u(t_{i-1},Y_{i-1}^k)\Big]$ by performing a Taylor expansion at the point (t_{i-1},Y_{i-1}^k). Since u is a smooth solution of the Kolmogorov equation, by Taylor expansion we have

$$\begin{aligned}
&u(t+\Delta t, x+\Delta x) - u(t,x)\\
&= \Delta t\frac{\partial u(t,x)}{\partial t} + \frac{1}{2}(\Delta t)^2\frac{\partial^2 u(t,x)}{\partial t^2} + \Delta t\sum_{|\alpha|=1}\Delta x^\alpha\frac{\partial u(t,x)}{\partial t}\partial_\alpha u(t,x)\\
&+\frac{1}{2}\sum_{|\alpha|=2}\Delta x^\alpha\frac{\partial u(t,x)}{\partial t}\partial_\alpha u(t,x) + \frac{1}{6}\Delta t\sum_{|\alpha|=3}\Delta x^\alpha\frac{\partial u(t,x)}{\partial t}\partial_\alpha u(t,x)
\end{aligned}$$

$$+\frac{1}{|\alpha|!}\sum_{|\alpha|=1}^{5}\Delta x^{\alpha}\partial_{\alpha}u(t,x)+\cdots ...$$

where $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_r)$ and Δx^{α} means $\Delta x^{\alpha}=(\Delta x_1)^{\alpha_1}..(\Delta x_r)^{\alpha_r}$. Note that $t_i=t_{i-1}+k$. Also, we can estimate $Y_i^k-Y_{i-1}^k$ in terms of k by using conditions on a and b along with Euler scheme. Moreover, one case use (3.8) and (3.9) together with the above Taylor expansion for u (after easy but tedious computations) to have

$$\mathbb{E}\Big[u(t_i,Y_i^k)-u(t_{i-1},Y_{i-1}^k)\Big]=k^2\mathbb{E}\Big[\psi_e(t_{i-1},Y_{i-1}^k)\Big]+k^3R_{i,k} \tag{3.10}$$

where

$$\begin{aligned}\psi_e(t,x)=&\frac{1}{2}\sum_{i,j=1}^{n}a_i(t,x)a_j(t,x)\partial_{ij}u(t,x)+\frac{1}{2}\sum_{i,j,l=1}^{n}a_i(t,x)\sigma_{jl}(t,x)\partial_{ijl}u(t,x)\\&+\frac{1}{8}\sum_{i,j,l,m=1}^{n}\sigma_{ij}(t,x)\sigma_{lm}(t,x)\partial_{ijlm}u(t,x)+\frac{1}{2}\frac{\partial^2u(t,x)}{\partial t^2}\\&\quad+\frac{1}{2}\sum_{i,j=1}^{n}\sigma_{ij}(t,x)\frac{\partial}{\partial t}\partial_{ij}u(t,x)+\sum_{i=1}^{n}a_i(t,x)\frac{\partial}{\partial t}\partial_iu(t,x),\end{aligned} \tag{3.11}$$

and there exists a constant C_T independent of $k>0$ such that $|R_{i,k}|\le C_T$. In view of (3.10), weak error $E_{\rm rr}(T,h)$ takes of the form

$$E_{\rm rr}(T,h)=k^2\sum_{j=0}^{N-1}\mathbb{E}\Big[\psi_e(t_j,Y_j^k)\Big]+k^2\mathcal{R}_{N,k};\quad|\mathcal{R}_{N,k}|\le C_T. \tag{3.12}$$

Lemma 3.5. *There exists a real number C_T, independent of k such that*

$$k\sum_{j=0}^{N-1}\mathbb{E}\Big[\big|\psi_e(t_j,Y_j^k)\big|\Big]\le C_T.$$

Proof. The result follows in view of the definition of $\psi_e(t,x)$ (cf. (3.11)) and the estimates in (3.8) and (3.9):

$$k\sum_{j=0}^{N-1}\mathbb{E}\Big[\big|\psi_e(t_j,Y_j^k)\big|\Big]\le k\sum_{j=0}^{N-1}C_\alpha(T)\big(1+\|Y_j^k\|_{\mathbb{R}^n}^{m_\alpha(T)}\big)\Big]\le CkN\big(1+\exp(CT)\big)\le C_T.$$

□

Note that, if $\phi\in F_T$, then the function $v(\theta;t,x)$ defined by $v(\theta;t,x)=\mathbb{E}\big[\phi(\theta;X_T^{t,x})\big]$ verifies the following equation

$$\begin{cases}\frac{\partial v}{\partial t}+\mathcal{L}v=0\\ v(\theta;T,x)=\phi(\theta,x).\end{cases}$$

Moreover, for all $\theta\in[0,T]$, $v(\theta;t,x)$ satisfies the estimate (3.8).

Lemma 3.6. *For any function $\phi\in F_T$, there exists a real number C_T, independent of k such that*

$$\mathbb{E}\Big[\phi(\theta,Y_N^k)-\phi(\theta,X_T)\Big]=R_T(h)\quad\textit{with}\ |R_T(h)|\le C_Th. \tag{3.13}$$

Proof. In view of (3.12), we see that for any $\phi \in F_T$,

$$\mathbb{E}\Big[\phi(\theta, Y_N^k) - \phi(\theta, X_T)\Big] = k^2 \sum_{j=0}^{N-1} \mathbb{E}\Big[\psi_e(t_j, Y_j^k)\Big] + k^2 R_{N,k} := R_N(k)$$

Since $|R_{N,k}| \le C_T$ and $k \sum_{j=0}^{N-1} \mathbb{E}\Big[\big|\psi_e(t_j, Y_j^k)\big|\Big] \le C_T$ (see Lemma 3.5), we observe that $|R_T(k)| \le C_T k$. This completes the proof. □

Lemma 3.7.

$$\Big|k \sum_{j=1}^{N-1} \mathbb{E}\big[\psi_e(t_j, Y_j^k)\big] - \int_0^T \mathbb{E}\big[\psi_e(s, X_s)\big]\, ds\Big| = \mathcal{O}(k).$$

Proof. By the triangle inequality, we obtain

$$\begin{aligned}
&\Big|k \sum_{j=1}^{N-1} \mathbb{E}\big[\psi_e(t_j, Y_j^k)\big] - \int_0^T \mathbb{E}\big[\psi_e(s, X_s)\big]\, ds\Big| \\
&\le k \sum_{j=1}^{N-1} \Big|\mathbb{E}\big[\psi_e(t_j, Y_j^k) - \psi_e(t_j, X_{t_j})\big]\Big| + \Big|h \sum_{j=1}^{N-1} \mathbb{E}\big[\psi_e(t_j, X_{t_j})\big] - \int_0^T \mathbb{E}\big[\psi_e(s, X_s)\big]\, ds\Big|
\end{aligned}$$

Note that $\psi_e(t,x)$ belongs to F_T and therefore by Lemma 3.6, $k \sum_{j=1}^{N-1} \Big|\mathbb{E}\big[\psi_e(t_j, Y_j^k) - \psi_e(t_j, X_{t_j})\big]\Big| = \mathcal{O}(k)$. Again, since $s \mapsto \mathbb{E}\big[\psi_e(s, X_s)\big]$ has a continuous first derivative, we see that $\Big|h \sum_{j=1}^{N-1} \mathbb{E}\big[\psi_e(t_j, X_{t_j})\big] - \int_0^T \mathbb{E}\big[\psi_e(s, X_s)\big]\, ds\Big| \le Ck$. Hence the result follows by combining these estimates. □

Now we are in a position to state and prove the following theorem regarding the global weak error expansion of Euler scheme.

Theorem 3.8. *Let a, $b \in C^\infty$ and have all bounded derivatives and $f \in F_T$. Then the weak error has the following expression:*

$$E_{\rm rr}(T,k) = k \int_0^T \mathbb{E}\big[\psi_e(s, X_s)\big]\, ds + \mathcal{O}(k^2), \tag{3.14}$$

where $\psi_e(\cdot,\cdot)$ is defined by (3.11).

Proof. In view of (3.12), we see that

$$\begin{aligned}
E_{\rm rr}(T,k) :&= \mathbb{E}\Big[u(T, Y_N^k) - u(0, Y)\Big] = k^2 \sum_{j=0}^{N-1} \mathbb{E}\big[\psi_e(t_j, Y_j^k)\big] + k^2 R_{N,k} \\
&= k \int_0^T \mathbb{E}\big[\psi_e(s, X_s)\big]\, ds + k\Big\{k \sum_{j=0}^{N-1} \mathbb{E}\big[\psi_e(t_j, Y_j^k)\big] - \int_0^T \mathbb{E}\big[\psi_e(s, X_s)\big]\, ds\Big\} \\
&\qquad + k^2 R_{N,k}.
\end{aligned}$$

Thanks to Lemma 3.7 and the fact that $|R_{N,k}| \le C_T$, we observe that

$$k\Big\{k \sum_{j=0}^{N-1} \mathbb{E}\big[\psi_e(t_j, Y_j^k)\big] - \int_0^T \mathbb{E}\big[\psi_e(s, X_s)\big]\, ds\Big\} + k^2 R_{N,k} = \mathcal{O}k^2.$$

In other words, we conclude that (3.14) holds true. This finishes the proof. $\square$

Some applications:

1) Just as in the case of ordinary differential equations, it is possible to carry on to the stochastic case some usual applications of the expansion of the error. In view of Theorem 3.8, we have
$$E_{\rm rr}(T,h) = e_1(T)k + \mathcal{O}(k^2). \tag{3.15}$$
We can control the discretization step size k in order to get the error less than a given tolerance. For example, we first perform an approximation with the step size k, and then a second one with step size $\frac{k}{2}$. Then, thanks to (3.15), we have
$$\mathbb{E}\big[f(Y_N^k)\big] - \mathbb{E}\big[f(Y_N^{\frac{k}{2}})\big] = e_1(T)\frac{k}{2} + \mathcal{O}(k^2)$$
from which we can estimate $e_1(T)$. Again, in view of (3.15), we can choose a new k in order to get the error less than a given tolerance.
2) The second application is the use of polynomial extrapolation methods to improve the approximation to the true solution. Consider the following new approximation (the Romberg extrapolation)
$$Z_T^k := 2\mathbb{E}\big[f(Y_N^{\frac{k}{2}})\big] - \mathbb{E}\big[f(Y_N^k)\big].$$
Since $\mathbb{E}\big[f(Y_N^k)\big] = \mathbb{E}\big[f(X_T)\big] = -e_1(T)k + \mathcal{O}(k^2)$, it is easy to check that
$$\mathbb{E}\big[f(X_T)\big] - Z_T^k = \mathcal{O}(k^2).$$
i.e., it is possible to get a result of precision of order k^2 from the results given by a first-order scheme.

3.2. **Higher order scheme.** The key to the construction of most higher-order numerical approximations is usually obtained from the truncated expansion of the variables of interest over small increments. The well-known Taylor formula provides the basis for the derivation of most deterministic numerical algorithms. In the stochastic case, a stochastic Taylor expansion for Itô SDE was first described by Wagner and Platen (1978). The Wagner-Platen formula is obtained by iterated applications of the Itô formula to the integrands in the integral version of the SDE (2.1) for $d = 1 = m$:
$$f(t,X_t) = f(t_0,X_{t_0}) + \int_{t_0}^t Lf(s,X_s)\,ds + \int_{t_0}^t \Lambda f(s,X_s)\,dW_s, \tag{3.16}$$
where
$$L \equiv \partial_t + a\partial_x + \frac{1}{2}b^2\partial_x^2, \qquad \Lambda \equiv b\partial_x.$$
Again we can apply Itô formula to have
$$\begin{cases} Lf(s,X_s) = Lf(t_0,X_{t_0}) + \int_{t_0}^s L^2 f(s_1,X_{s_1})\,ds_1 + \int_{t_0}^s \Lambda L f(s_1,X_{s_1})\,dW_{s_1} \\ \Lambda f(s,X_s) = \Lambda f(t_0,X_{t_0}) + \int_{t_0}^s L\Lambda f(s_1,X_{s_1})\,ds_1 + \int_{t_0}^s \Lambda^2 f(s_1,X_{s_1})\,dW_{s_1}. \end{cases} \tag{3.17}$$
We substitute (3.17) in (3.16) to have
$$f(t,X_t) = f(t_0,X_{t_0}) + Lf(t_0,X_{t_0})(t-t_0) + \Lambda f(t_0,X_{t_0})(W_t - W_{t_0})$$

$$+\int_{t_0}^{t}\int_{t_0}^{s} L^2 f(s_1, X_{s_1})\, ds_1\, ds + \int_{t_0}^{t}\int_{t_0}^{s} \Lambda L f(s_1, X_{s_1})\, dW_{s_1}\, ds$$
$$+\int_{t_0}^{t}\int_{t_0}^{s} L\Lambda f(s_1, X_{s_1})\, ds_1\, dW_s + \int_{t_0}^{t}\int_{t_0}^{s} \Lambda^2 f(s_1, X_{s_1})\, dW_{s_1}\, dW_s. \quad (3.18)$$

Now consider the simple case: namely a and b appearing in the SDE (2.1) are independent of the time variable. Then, by choosing $f(t,x) \equiv x$ in (3.18), we obtain the expansion

$$X_t = X_{t_0} + a(X_{t_0})(t-t_0) + b(X_{t_0})\int_{t_0}^{t} dW_s + b(X_{t_0})b'(X_{t_0})\int_{t_0}^{t}\int_{t_0}^{s} dW_{s_1}\, dW_s + R_{t_0,t} \quad (3.19)$$

where $R_{t_0,t}$ represents some remainder term consisting of higher order multiple stochastic integrals. Multiple Itô integrals of the form

$$I_{(1)} := \int_{t_n}^{t_{n+1}} dW_s;\;\; I_{(1,1)} := \int_{t_n}^{t_{n+1}}\int_{t_n}^{s} dW_{s_1}\, dW_s;\;\; I_{(0,1)} := \int_{t_n}^{t_{n+1}}\int_{t_n}^{s} ds_1\, dW_s;$$
$$I_{(1,0)} := \int_{t_n}^{t_{n+1}}\int_{t_n}^{s} dW_{s_1}\, ds;\;\; I_{(1,1,1)} := \int_{t_n}^{t_{n+1}}\int_{t_n}^{s_3}\int_{t_n}^{s_2} dW_{s_1}\, dW_{s_s}\, dW_{s_3}$$

on the interval $[t_n, t_{n+1}]$ form the random building blocks in the Wagner-Platen expansions.

Exercise 3.5. *Show that*

$$I_{(1,1)} := \int_{t_n}^{t_{n+1}}\int_{t_n}^{s} dW_{s_1}\, dW_s = \frac{1}{2}\Big((I_{(1)})^2 - (t_{n+1}-t_n)\Big).$$

Solution: In view of the Itô formula $x \mapsto |x|^2$, we see that

$$\int_{t_n}^{t} W_s dW_s = \frac{1}{2}(W_t^2 - W_{t_n}^2) - \frac{1}{t}(t-t_n).$$

Now,

$$\begin{aligned} I_{(1,1)} &= \int_{t_n}^{t_{n+1}} (W_s - W_{t_n})\, dW_s = \int_{t_n}^{t_{n+1}} W_s dW_s - W_{t_n}(W_{t_{n+1}} - W_{t_n}) \\ &= \frac{1}{2}(W_t^2 - W_{t_n}^2) - \frac{1}{t}(t-t_n) - W_{t_n}(W_{t_{n+1}} - W_{t_n}) = \frac{1}{2}(W_{t_{n+1}} - W_{t_n})^2 - \frac{1}{2}(t_{n+1}-t_n) \\ &= \frac{1}{2}\Big((I_{(1)})^2 - (t_{n+1}-t_n)\Big). \end{aligned}$$

Exercise 3.6. *Express $I_{(1,1,1)}$ as a polynomial in $I_{(1)}$.*

Truncating the Itô-Taylor expansion at an appropriate point (setting $R_{t_0,t} = 0$ in (3.19)) we have the following **Milstein scheme**:

$$Y_{j+1}^k = Y_j^k + k a(Y_j^k) + b(Y_j^k)\Delta_j W + b(Y_j^k)b'(Y_j^k) I_{(1,1)} \quad \forall j = 1, 2, \cdots N-1 \quad (3.20)$$

where $I_{(1,1)} := \int_{t_j}^{t_{j+1}}\int_{t_j}^{s} dW_{s_1}\, dW_s$. In general this scheme has strong order $\gamma = 1.0$. Thus, adding one more term from the Wagner-Platen formula to the Euler scheme already provides an improvement in efficiency. Moreover, one can consider the following higher-order scheme whose strong order $\gamma = 1.5$

$$Y_{n+1}^k = Y_n^k + a(Y_n^k)k + b(Y_n^k)\Delta_n W + b(Y_n^k)b'(Y_n^k) I_{(1,1)} + b(Y_n^k)a'(Y_n^k) I_{(1,0)}$$

$$+\left(a(Y_n^k)a'(Y_n^k)+\frac{1}{2}b^2(Y_n^k)a''(Y_n^k)\right)\frac{k^2}{2}+\left(a(Y_n^k)b'(Y_n^k)+\frac{1}{2}b^2(Y_n^k)b''(Y_n^k)\right)I_{(0,1)}$$
$$+b(Y_n^k)\left\{b(Y_n^k)b''(Y_n^k)+(b'(Y_n^k))^2\right\}I_{(1,1,1)}. \tag{3.21}$$

Disadvantage of higher-order scheme: A disadvantage of higher-order strong Taylor approximations is the fact that derivatives of the drift and diffusion coefficients have to be computed at each step. This can be avoided by considering derivative-free approximations, such as the Runge-kutta type method. We will discuss this scheme later (if time permits).

Now we text the strong convergence order of the Milstein's scheme for the nonlinear SDE of the form

$$dX(t)=rX(t)(K-X(t))dt+\beta X(t)dW(t); \quad X_0 = Xzero = 1$$

where r, K and β are constants. This SDE arises in population dynamics.

```
% Strong convergence of Milstein scheme
%%%%%%%%% Scheme %%%%%%%%%%%%%%%%%%
% Y_{j+1}= Y_j + \Delta t a(Y_j) + b(Y_j)\Delta_jW +
% 1/2 b(Y_j)b^{\prime}(Y_j) ( (\Delta_W)^2 - \Delta t)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% We consider consider the following SDE
% dX(t)= r X(t)(K-X(t))dt + \beta X(t) dW(t);  which arises in propulation
% dynamics, where r, K and \beta are constants.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clc
rng( default );
T=1;    % Final time
N=2^9;
r=1; beta=0.25; K=1; Xzero=1; % parameters for the SDE
M=5000; % number of sample paths
dt=T/N; % time steps
R=[1; 16; 32; 64; 128; 256; 512;1024]; % Milstein stepsizes are R*dt
dW=sqrt(dt)*randn(M,N); % increments for Brownian motion
Xmil=zeros(M,8);
for p=1:8
    Dt=R(p)*dt;
    L= N/R(p); % L time steps of size Dt=R*dt\
    Xtemp=Xzero*ones(M,1); % initial discretization for each paths
    for j=1:L
        Winc=sum(dW(:, R(p)*(j-1)+1: R(p)*j),2);
        Xtemp= Xtemp + Dt*r*Xtemp.*(K-Xtemp) + beta*Xtemp.*Winc...
            + 0.5*beta^2*Xtemp.*(Winc.^2-Dt);
    end
    Xmil(:,p)= Xtemp; % Milstein solution at final time for eact time step
end
%%%%%%%%%%%%%%
% We take the Milstein solution with step Dt=dt to be a good approximation
% of the exact solution and called it Xref
Xref=Xmil(:,1);
```

```
Xerr= abs(Xmil(:,2:8)-repmat(Xref,1,7)); % error in each path
mean(Xerr);  % strong error for Milstein scheme
Dtvals=dt*R(2:8);  % Milstein time steps

loglog(Dtvals, mean(Xerr),  b*- ), hold on
loglog(Dtvals, Dtvals,  r-- ), hold off  % reference slope of 1
xlabel( Delta t )
ylabel( Sample average of |X(T)-X_L| )
title( Strong error for Milstein scheme ,  FontSize ,10)
```

4. Stability analysis of numerical methods for linear SDE

The concept of strong and weak convergence concern the accuracy of a numerical method over a finite interval $[0, T]$ for small step sizes Δt. However in many applications, the long-term behavior ($t \to \infty$) of an SDE is of interest.

4.1. **Linear SDE: multiplicative noise.** Consider the linear SDE of the form

$$\begin{cases} dX(t) = \lambda X(t)dt + \mu X(t)dW(t) & t > 0, \\ X(0) = 1, \end{cases} \tag{4.1}$$

where $\lambda, \mu \in \mathbb{C}$ with $\mathcal{R}(\lambda) < 0$. The exact solution is given by

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

If $\mu = 0$, then the equation (4.1) reduces to a deterministic linear test equation, and $\lim\limits_{t\to\infty} X(t) = 0$. In order to generalize this idea to the SDE case, we need to precise the meaning of "$\lim\limits_{t\to\infty} X(t) = 0$" as random variables are infinite-dimensional objects and hence norms are not equivalent in general. We will consider the most common measures of stability:

i) **Mean square stability (MS-stability):** $\lim\limits_{t\to\infty} \mathbb{E}[|X(t)|^2] = 0$.
ii) **Asymptotic stability:** $\lim\limits_{t\to\infty} |X(t)| = 0$, with probability 1.

Lemma 4.1. *The solution of the text equation (4.1) is MS-stable if and only if*

$$2\mathcal{R}(\lambda) + |\mu|^2 < 0. \tag{4.2}$$

Proof. Assume that $Y(t) = \mathbb{E}[|X(t)|^2]$. Then application of Itô's formula $y \mapsto |y|^2$ yields that $Y(t)$ satisfies the following ODE:

$$\begin{cases} \frac{dY(t)}{dt} = (2\mathcal{R}(\lambda) + |\mu|^2)Y(t) \\ Y(0) = 1 \end{cases} \tag{4.3}$$

which has the solution $Y(t) = \exp((2\mathcal{R}(\lambda) + |\mu|^2)t)$. Thus, we see that $X(t)$ is MS-stable if and only if the condition (4.2) holds. This completes the proof. $\square$

Remark 4.1. It can be shown that the solution of the test equation (4.1) is **asymptotic**-stable if and only if $\mathcal{R}(2\lambda - \mu^2) < 0$. Therefore, MS-stability of the solution of (4.1) implies that it is **asymptotic**-stable as $\mathcal{R}(2\lambda - \mu^2) \le 2\mathcal{R}(\lambda) + |\mu|^2 < 0$.

Now suppose that the parameters λ and μ are chosen so that SDE (4.1) is stable in the mean-square or asymptotic sense. Let X_n be the associated numerical scheme for the SDE (4.1). A natural question is then ask the values of mesh size of the time discretization in which the numerical solution X_n is stable in an analogous sense. Now we define Y_n corresponding to $Y(t)$ by $Y_n = \mathbb{E}[|X_n|^2]$. Then Y_n satisfies the following one-step difference equation:

$$Y_{n+1} = R(h,l)Y_n, \quad \text{where } \; h = k\lambda, \; l = -\frac{\mu^2}{\lambda}.$$

We shall call $R(h,l)$ the stability function of the scheme. Clearly $Y_n \to 0$ as $n \to \infty$ if and only if $|R(h,l)| < 1$.

Definition 4.1. (MS-stability of the scheme) Let the parameters λ and μ satisfies (4.2). The numerical scheme associated to SDE (4.1) is said to be MS-stable for those values of h and l such that $|R(h,l)| < 1$. The set $\mathscr{R}$ given by $\mathscr{R} := \big\{(h,l) : |R(h,l)| < 1\big\}$ is analogously called the domain of MS-stability of the scheme.

Stability functions of schemes: We now consider various numerical scheme to (4.1), and derive its stability function $R(h,l)$.

a) Euler-Maruyama scheme: Euler-Maruyama scheme for the problem (4.1) reads as

$$X_{n+1} = X_n + k\lambda X_n + \mu X_n \Delta_n W = (1+\lambda k)X_n + +\mu X_n \Delta_n W.$$

Thus

$$\mathbb{E}\Big[|X_{n+1}|^2\Big] = |1+\lambda k|^2 \mathbb{E}[|X_n|^2] + k|\mu|^2 \mathbb{E}[|X_n|^2] = \Big(|1+\lambda k|^2 + |\mu|^2 k\Big)\mathbb{E}[|X_n|^2],$$

and therefore we see that

$$R(h,l) = |1+h|^2 + |lk|.$$

b) Semi-implicit Euler scheme:

$$X_{n+1} = X_n + k\Big(\alpha\lambda X_{n+1} + (1-\alpha)\lambda X_n\Big) + \mu X_n \Delta_n W \quad \alpha \in [0,1].$$

Here the scheme is called the trapezoidal Euler scheme if $\alpha = \frac{1}{2}$, and the backward Euler scheme if $\alpha = 1$. A simple algebra reveals that

$$(1-\alpha h)X_{n+1} = \big(1 + (1-\alpha)h\big)X_n + \mu X_n \Delta_n W.$$

Thus $Y_n = \mathbb{E}[|X_n|^2]$ satisfies

$$Y_{n+1} = \frac{\big|1+(1-\alpha)h\big|^2 + |\mu|^2 k}{|1-\alpha h|^2} Y_n.$$

Hence

$$R(h,l;\alpha) := \frac{\big|1+(1-\alpha)h\big|^2 + |hl|}{|1-\alpha h|^2}.$$

In particular, for trapezoidal Euler scheme, $R(h,l) = \frac{|1+\frac{1}{2}h|^2 + |hl|}{|1-\frac{1}{2}h|^2}$, and for backward Euler scheme $R(h,l) = \frac{1+|hl|}{|1-h|^2}$.

c) Milstein's scheme:

$$X_{n+1} := X_n + \lambda k X_n + \mu X_n \Delta_n W + \frac{\mu^2}{2} X_n\big((\Delta_n W)^2 - k\big)$$
$$= \big(1 + \lambda k - \frac{k\mu^2}{2}\big) X_n + \mu X_n \Delta_n W + \frac{\mu^2}{2} X_n (\Delta_n W)^2.$$

By using binomial formula, one has

$$|X_{n+1}|^2 = \big(1 + \lambda k - \frac{k\mu^2}{2}\big)^2 X_n^2 + 2\mu X_n^2 \big(1 + \lambda k - \frac{k\mu^2}{2}\big)\Delta_n W + \mu^3 X_n^2 (\Delta_n W)^3$$
$$+ \Big(1 + \big(1 + \lambda k - \frac{k\mu^2}{2}\big)\Big)|\mu|^2 X_n^2 |\Delta_n W|^2 + \frac{\mu^4}{4} X_n^2 |\Delta_n W|^4$$

Taking expectation, we obtain

$$Y_{n+1} = \Big\{\big(1 + \lambda k - \frac{k\mu^2}{2}\big)^2 + k\mu^2\Big(1 + \big(1 + \lambda k - \frac{k\mu^2}{2}\big)\Big) + \frac{3}{4}k^2|\mu|^4\Big\} Y_n$$
$$= \Big\{|1 + \lambda k|^2 + k|\mu|^2 (1 + \frac{1}{2}k|\mu|^2)\Big\} Y_n.$$

Thus, associated stability function $R(h,l)$ is given by

$$R(h,l) := |1 + h|^2 + |hl| + \frac{1}{2}|h^2 l^2|.$$

d) Semi-implicit Milstein's scheme:

$$X_{n+1} := X_n + \lambda\Big(\alpha X_{n+1} + (1-\alpha) X_n\Big) k + \mu X_n \Delta_n W + \frac{\mu^2}{2} X_n\big((\Delta_n W)^2 - k\big) \quad \alpha \in [0,1].$$

Here the scheme is called the trapezoidal Milstein's scheme if $\alpha = \frac{1}{2}$, and the backward Milstein's scheme if $\alpha = 1$. One can derive the stability function $R(h,l;\alpha)$ as

$$R(h,l;\alpha) := \frac{|1 + (1-\alpha)h|^2 + |hl| + \frac{1}{2}|hl|^2}{|1 - \alpha h|^2}.$$

Lemma 4.2. *The semi-implicit Euler scheme with $\frac{1}{2} \le \alpha \le 1$ is unconditionally MS-stable for $0 \le l \le 2$.*

Proof. The stability function of the semi-implicit Euler scheme is given by the following expression:

$$R(h,l;\alpha) := \frac{\big|1 + (1-\alpha)h\big|^2 + |hl|}{|1 - \alpha h|^2}.$$

If $0 \le \alpha \le \frac{1}{2}$, the boundary of the region i.e., $|R(h,l;\alpha)| = 1$ intersects the h-axis at the point $(0, \frac{2}{2\alpha - 1})$. On the other hand, if $\frac{1}{2} \le \alpha \le 1$, then the boundary of the region does not intersect the h-axis and has the asymptote $l = 2$. This finishes the proof. $\square$

The asymptotic version of the Euler-Maruyama scheme for (4.1) can be studied via the strong law of large number and the law of the iterated logarithm, leading to

$$\lim_{n\to\infty} |X_n| = 0, \mathbb{P}\text{-a.s.} \iff \mathbb{E}\Big[\log\big|1 + k\lambda + \sqrt{k}\mu\mathcal{N}(0,1)\big|\Big] < 0. \tag{4.4}$$

Here is the numerical evidence for MS and asymptotic stability of the Euler-Maruyama solution for the SDE (4.1).

```
% Test for MS-stability of the linear SDE
% dX(t)= \lambda X(t) dt + \mu X(t) dW(t); X(0)= Xzero=1
% where \lambda and \mu are complex valued.

%%%%%%%%%%%%%%%%%%%  MS-conditions  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Re(\lambda) + 1/2 |\mu|^2 < 0 ; |1+ k \lambda|^2 + k|\mu|^2 < 1
% where k is a mesh-size of the time interval [0,T]

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clc
rng( default );
T=30;    % Final time
M= 100000; % number of sample path
lambda=-3; mu=sqrt(3); Xzero=1; % problem parametres
ltype={ b- ,  r-- ,  m- , g-- }; % linetypes for plot

subplot(211) %%%   Mean-square stability  %%%%%%%%%%%%%%%%%%%%%
for m=1:4
    dt=2^(1-m);
    N= T/dt; % number of steps for step size dt
    Xms=zeros(1,N);
    Xtemp=Xzero*ones(M,1);
    for j=1:N
        Winc =sqrt(dt)*randn(M,1);
        Xtemp= Xtemp + dt*lambda*Xtemp + mu*Xtemp.*Winc;
        Xms(j)=mean(Xtemp.^2); % mean-square estimate
    end
    semilogy([0:dt:T], [Xzero, Xms],ltype{m},  Linewidth , 2), hold on
end
legend( dt=1 , dt=1/2 ,  dt=1/4 ,  dt=1/8 )
title( Mean-square:\lambda=-3,\mu=\surd 3 ,  FontSize , 16)
ylabel ( E[|X|^2] ,  FontSize , 12), axis([0,T,1e-20, 1e+20]), hold off

subplot(212)  %%% Asymptotic-stability: a single path %%%%%%
T=300;
%%%%%%%%%%  Asymptotic stable conditions %%%%%%%%%%%%%%%%%%%%%%%%%%%

% Re(\lambda - 1/2 \mu^2) < 0, and
% E{log|1+ k\lambda + \sqrt{k}\mu N(0,1)|}<0

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

lambda=0.5; mu= sqrt(6); % problem parameters satisfying 1st condition
for m=1:4
    dt= 2^(1-m);
    N=T/dt;
    Xemabs=zeros(1,N); Xtemp=Xzero;
    for j=1:N
        Winc=sqrt(dt)* randn;
        Xtemp= Xtemp + dt*lambda*Xtemp + mu*Xtemp*Winc;
        Xemabs(j)=abs(Xtemp);
    end
    semilogy([0:dt:T], [Xzero,Xemabs], ltype{m},  Linewidth , 2), hold on
```

```
end
legend( dt=1 , dt=1/2 ,  dt=1/4 ,  dt=1/8 )
title( Asymptotic—single path: \lambda=1/2, \mu=\surd 6 ,  FontSize , 16)
ylabel ( |X| ,  FontSize , 12), axis([0,T,1e—50, 1e+100]), hold off
```

4.2. **Linear SDE: additive noise.** Let us consider the linear SDE with additive noise of the form

$$\begin{cases} dX(t) = \lambda X(t)dt + \mu dW(t) \quad t > 0, \\ X(0) = x_0, \end{cases} \tag{4.5}$$

where $\lambda, \mu \in \mathbb{C}$ with $\mathcal{R}(\lambda) < 0$. Exact solution of (4.5) is given by

$$X(t) = \exp(\lambda t)x_0 + \mu \int_0^t \exp(\lambda(t-s))\, dW(s).$$

Observe that

$$\begin{aligned} \mathbb{E}\Big[|X(t)|^2\Big] &= \exp(2\mathcal{R}(\lambda)t)\mathbb{E}\big[|x_0|^2\big] - \frac{|\mu|^2}{2\mathcal{R}(\lambda)}\big(1-\exp(2\mathcal{R}(\lambda)t)\big) \\ &\longrightarrow -\frac{|\mu|^2}{2\mathcal{R}(\lambda)} \quad (t \uparrow \infty) \quad (\because \mathcal{R}(\lambda) < 0). \end{aligned}$$

Let us consider the Euler-Maruyama scheme:

$$X_{n+1} = (1+\lambda k)X_n + \mu \Delta_n W.$$

This implies that

$$\mathbb{E}\big[|X_{n+1}|^2\big] = |1+\lambda k|^2 \mathbb{E}\big[|X_n|^2\big] + |\mu|^2 k.$$

Continuing the iteration, one has

$$\begin{aligned} \mathbb{E}\big[|X_{n+1}|^2\big] &= |1+\lambda k|^2\Big\{|1+\lambda k|^2\mathbb{E}\big[|X_{n-1}|^2\big] + |\mu|^2 k\Big\} + |\mu|^2 k \\ &= |1+\lambda k|^4\mathbb{E}\big[|X_{n-1}|^2\big] + \Big(1+|1+\lambda k|^2\Big)|\mu|^2 k \\ &= |1+\lambda k|^{2(n+1)}\mathbb{E}\big[|x_0|^2\big] + \Big\{|1+\lambda k|^{2n} + \cdots + |1+\lambda k|^2 + 1\Big\}|\mu|^2 k \\ &= |1+\lambda k|^{2(n+1)}\mathbb{E}\big[|x_0|^2\big] + \frac{|1+\lambda k|^{2(n+1)} - 1}{|1+\lambda k|^2 - 1}|\mu|^2 k \\ &= |1+\lambda k|^{2(n+1)}\mathbb{E}\big[|x_0|^2\big] + \frac{|1+\lambda k|^{2(n+1)} - 1}{2\mathcal{R}(\lambda) + |\lambda|^2 k}|\mu|^2 \end{aligned}$$

Thus, if $|1+\lambda k| < 1$, Then

$$\mathbb{E}\big[|X_{n+1}|^2\big] \longrightarrow -\frac{|\mu|^2}{2\mathcal{R}(\lambda) + |\lambda|^2 k} \quad (n \to \infty)$$

and hence

$$\lim_{k\to 0}\Big(\lim_{n\to\infty}\mathbb{E}\big[|X_n|^2\big]\Big) = -\frac{|\mu|^2}{2\mathcal{R}(\lambda)}.$$

Definition 4.2. (Asymptotically consistent in mean square) A numerical method is said to be asymptotically consistent in mean square if the numerical solution $\{X_n\}$ for the test equation (4.5) satisfies

$$\lim_{k\to 0}\Big(\lim_{n\to\infty}\mathbb{E}\big[|X_n|^2\big]\Big) = -\frac{|\mu|^2}{2\mathcal{R}(\lambda)}.$$

In view of the above discussions, we arrive at the following theorem:

Theorem 4.3. *The Euler-Maruyama scheme for (4.5) is asymptotically consistent in mean square if* $|1+\lambda k|<1$.

Definition 4.3. (Numerical stable in mean) The numerical method is said to be numerical stable in mean if the numerical solution X_n for the test equation (4.5) is satisfied $\mathbb{E}[X_n]\to 0$ as $n\to\infty$.

Exercise 4.1. *Show that*

i) The Euler-Maruyama scheme for (4.5) is numerical stable in mean if $|1+\lambda k|<1$.
ii) The semi-implicit Euler scheme for (4.5) is numerical stable in mean if

$$\big|1+\frac{h}{1-\alpha h}\big|<1. \tag{4.6}$$

Moreover, under the condition (4.6), The semi-implicit Euler scheme for (4.5) is asymptotically consistent in mean square.

References

[1] P. Glasserman Monte Carlo Methods in Financial Engineering *Applications of Mathematics (New York), 53.* Stochastic Modelling and Applied Probability. Springer-Verlag, New York, 2004.

[2] P. E. Kloeden, E. Platen. Numerical Solution of Stochastic Differential Equations. *Applications of Mathematics (New York), 23.* Stochastic Modelling and Applied Probability. Springer-Verlag, New York, 1992.

[3] D. J. Higham. An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations. *SIAM Review.* Vol. 43, No. 3, pp. 525-546

[4] Y. Saito, T. Mitsui. Stability Analysis of Numerical Schemes for Stochastic Differential Equations. *SIAM. J. Numer. Anal.* Vol. 33, N0. 6, pp. 2254-2267 (1996).

[5] D. Talay, L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.* 8 (1990), no. 4, 483-509 (1991).

[6] E. Gobet. Weak approximation of killed diffusion using Euler schemes. *Stochastic Process. Appl.* 87 (2000), no. 2, 167-197.