2

Brownian motion and stochastic calculus

In this chapter, we review some basic concepts for stochastic processes and stochastic calculus as well as numerical integration methods in random space for obtaining statistics of stochastic processes.

We start from Gaussian processes and their representations in Chapter 2.1 and then introduce Brownian motion and its properties and approximations in Chapter 2.2. We discuss basic concepts in stochastic calculus: Ito integral in Chapter 2.3 and Ito formula in Chapter 2.4. We then focus on numerical integration methods in random space such as Monte Carlo methods, quasi-Monte Carlo methods, Wiener chaos method, and stochastic collocation method (sparse grid collocation method) in Chapter 2.5. Examples of applying these methods to a simple equation are provided in Chapter 2.5.5 with Matlab code. In Chapter 2.6 of bibliographic notes, we present a review on different types of approximation of Brownian motion and a brief review on *pros* and *cons* of numerical integration methods in random space. Various exercises are provided for readers to familiarize themselves with basic concepts presented in this chapter.

2.1 Gaussian processes and their representations

On a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ $(\Omega = \mathbb{R})$, if a cumulative distribution function of a random variable X is normal, i.e.,

$$\mathbb{P}(X < x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}\, dy, \quad \sigma > 0.$$

Z. Zhang, G.E. Karniadakis, *Numerical Methods for Stochastic Partial Differential Equations with White Noise*,
Applied Mathematical Sciences 196, DOI 10.1007/978-3-319-57511-7_2

then the random variable X is called a Gaussian (normal) random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Here X is completely characterized by its mean μ and its standard deviation σ. We denote $X \sim \mathcal{N}(\mu, \sigma^2)$. The probability density function of X is

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

When $\mu = 0$ and $\sigma = 1$, we call X a standard Gaussian (normal) random variable. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$, i.e., Z is a standard Gaussian (normal) random variable.

Example 2.1.1 *If X_i are mutually independent Gaussian random variables, then $\sum_{j=1}^{N} a_i X_i$ is a Gaussian random variable for any $a_i \in \mathbb{R}$. In particular, if $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent, then*

$$\alpha X_1 + \beta X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \alpha^2\sigma_1^2 + \beta^2\sigma_2^2).$$

Definition 2.1.2 (Gaussian random vector) *A $\mathbb{R}^n$-valued random vector $X = (X_1, X_2, \ldots, X_n)^\top$ has an n-variate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ if $X = \boldsymbol{\mu} + AZ$ where the matrix A is of size $n \times n$, $\Sigma = AA^\top$, and $Z = (Z_1, Z_2, \ldots, Z_n)^\top$ is a vector with independent standard Gaussian (normal) components.*

When $n = 1$, X is a (univariate) Gaussian random variable. The probability density of X is

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n}\,|\Sigma|^{1/2}} e^{-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2}}.$$

Example 2.1.3 *A set of random variables $\{X_i\}_{i=1}^n$ are called* jointly Gaussian *if $\sum_{i=1}^n a_i X_i$ is a Gaussian random variable for any $a_i \in \mathbb{R}$. Then $X = (X_1, X_2, \ldots, X_n)^\top$ is a Gaussian random vector.*

The correlation of two random variables (vectors) is a normalized version of the covariance, with values ranging from -1 to 1:

$$\mathrm{Corr}(X, Y) = \frac{\mathbb{C}\mathrm{ov}[(X, Y)]}{\sqrt{\mathbb{V}\mathrm{ar}[X]\mathbb{V}\mathrm{ar}[Y]}}, \quad \mathbb{C}\mathrm{ov}[(X, Y)] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^\top].$$

When $\mathrm{Corr}(X, Y) = 0$, we say X and Y are uncorrelated.

Definition 2.1.4 (Gaussian process) *A collection of random variables is called a Gaussian process, if the joint distribution of any finite number of its members is Gaussian. In other words, a Gaussian process is a $\mathbb{R}^d$-valued stochastic process with continuous time (or with index) t such that $(X(t_0), X(t_1), \cdots, X(t_n))^\top$ is a $n+1$-dimensional Gaussian random vector for any $0 \le t_0 < t_1 < \cdots < t_n$.*

The Gaussian process is denoted as $X = \{X(t)\}_{t\in T}$ where T is a set of indexes. Here $T = [0, \infty)$.

The consistency theorem of Kolmogorov [255, Theorem 2.2] implies that the finite dimensional distribution of a Gaussian stochastic process $X(t)$ is uniquely characterized by two functions: the mean function $\mu_t = \mathbb{E}[X(t)]$ and the covariance function $C(t, s) = \mathbb{C}\text{ov}[X(t), X(s)]$.

A Gaussian process $\{X(t)\}_{t\in T}$ is called a centered Gaussian process if the mean function $\mu(t) = \mathbb{E}[X(t)] = 0$ for all $t \in T$.

Given a function $\mu(t) : T \to \mathbb{R}$ and a nonnegative definite function $C(t, s) : T\times T \to \mathbb{R}$, there exists a Gaussian process $\{X(t)\}_{t\in T}$ with the mean function $\mu(t)$ and the covariance function $C(t, s)$.

To find such a Gaussian process, we can use the following expansion.

Theorem 2.1.5 (Karhunen-Loève expansion) *Let $X(t)$ be a Gaussian stochastic process defined over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $t \in [a, b]$, $-\infty < a < b < \infty$. Suppose that $X(t)$ has a continuous covariance function $C(t, s) = \mathbb{C}ov[(X(t), X(s))] = \mathbb{E}[(X(t) - \mathbb{E}[X(t)])(X(s) - \mathbb{E}[X(s)])]$. Then $X(t)$ admits the following representation*

$$X(t) = \mathbb{E}[X(t)] + \sum_{k=1}^{\infty} Z_k e_k(t),$$

where the convergence is in L^2, uniform in t (i.e., $\lim_{n\to\infty} \max_{t\in[a,b]} \mathbb{E}[(X(t) - \mathbb{E}[X(t)] - \sum_{k=1}^{n} Z_k e_k(t))^2] = 0$) and

$$Z_k = \int_a^b \big(X(t) - \mathbb{E}[X(t)]\big) e_k(t)\, dt.$$

Here the eigenfunctions e_k's of C_X with respective eigenvalues λ_k's form an orthonormal basis of $L^2([a, b])$ and

$$\int_a^b C(t, s) e_k(t)\, dt = \lambda_k e_k(s), \quad k \geq 1.$$

Furthermore, the random variables Z_k's have zero-mean, are uncorrelated, and have variance λ_k

$$\mathbb{E}[Z_k] = 0, \quad \textit{for all } k \geq 1 \qquad \textit{and} \qquad \mathbb{E}[Z_i Z_j] = \delta_{ij}\lambda_j, \quad \textit{for all } i, j \geq 1.$$

This is a direct application of Mercer's theorem [497] on a representation of a symmetric positive-definite function as a sum of a convergent sequence of product functions. The stochastic process $X(t)$ can be non-Gaussian.

The covariance function $C(t, s)$ can be represented as $C(t, s) = \sum_{k=1}^{\infty} \lambda_k e_k(t) e_k(s)$. The variance of $X(t)$ is the sum of the variances of the individual components of the sum:

$$\mathbb{V}\mathrm{ar}[X(t)] = \mathbb{E}[(X(t) - \mathbb{E}[X(t)])^2] = \sum_{k=0}^{\infty} e_k^2(t)\mathbb{V}\mathrm{ar}[Z_k] = \sum_{k=1}^{\infty} \lambda_k e_k^2(t).$$

Here Z_k are uncorrelated random variables.

The domain where the process is defined can be extended to domains in $\mathbb{R}^d$. In Table 2.1 we present a list of covariance functions commonly used in practice. Here the constant l is called correlation length, K_ν is the modified

Table 2.1. A list of covariance functions.

Wiener process	$\min(x, y),\ x, y \geq 0$
White Noise	$\sigma^2\delta(x-y), \quad x, y \in \mathbb{R}^d$
Gaussian	$\exp\left(-\frac{\lvert x-y\rvert^2}{2l^2}\right), \quad x, y \in \mathbb{R}^d$
Exponential	$\exp\left(-\frac{\lvert x-y\rvert}{l}\right), \quad x, y \in \mathbb{R}^d$
Matérn kernel	$\frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}\,\lvert x-y\rvert}{l}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}\,\lvert x-y\rvert}{l}\right), \quad x, y \in \mathbb{R}^d$
Rational quadratic	$(1+\lvert x-y\rvert^2)^{-\alpha}, \quad x, y \in \mathbb{R}^d, \quad \alpha \geq 0$

Bessel function of order ν, and $\Gamma(\cdot)$ is the gamma function:

$$\Gamma(t) = \int_0^{\infty} x^{t-1}e^{-x}\,dx,\ t > 0.$$

Here are some examples of Karhunen-Loève expansion for Gaussian processes.

Example 2.1.6 (Brownian motion) *When $C(t,s) = \min(t,s)$, $t \in [0,1]$, then the Gaussian process $X(t)$ can be written as*

$$X(t) = \sqrt{2}\sum_{k=1}^{\infty} \xi_k \frac{\sin\left(\left(k-\frac{1}{2}\right)\pi t\right)}{\left(k-\frac{1}{2}\right)\pi}.$$

Here ξ_k's are mutually independent standard Gaussian random variables. One can show that for $t, s \in [0,1]$, the eigenvectors of the covariance function $\min(t,s)$ are

$$e_k(t) = \sqrt{2}\sin\left(\left(k-\tfrac{1}{2}\right)\pi t\right),$$

and the corresponding eigenvalues are

$$\lambda_k = \frac{1}{(k-\frac{1}{2})^2\pi^2}.$$

In the next section, we know that the process in Example 2.1.6 is actually a *Brownian motion.*

Example 2.1.7 (Brownian Bridge) *Let $X(t)$, $0 \le t \le 1$, be the Gaussian process in Example 2.1.6. Then $Y(t) = X(t) - tX(1)$, $0 \le t \le 1$, is also a Gaussian process and admits the following Karhunen-Loève expansion:*

$$Y(t) = \sum_{k=1}^{\infty} \eta_k \frac{\sqrt{2}\sin(k\pi t)}{k\pi}.$$

Here η_k's are mutually independent standard Gaussian random variables.

Example 2.1.8 (Ornstein-Uhlenbeck process) *Consider a centered one-dimensional Gaussian process with an exponential covariance function $\exp(-\frac{|t-s|}{l})$. The Karhunen-Loève expansion of such a Gaussian process over $[-a, a]$ is*

$$O(t) = \sum_{k=1}^{\infty} \xi_k \sqrt{\lambda_k} e_k(t),$$

where $\lambda_k = \frac{2l}{l^2\theta_k^2+1}$ and the corresponding eigenvalues are

$$e_{2i}(t) = \frac{\cos(\theta_{2i}t)}{\sqrt{2 + \frac{\sin(2\theta_{2i}a)}{2\theta_{2i}}}}, e_{2i-1}(t) = \frac{\sin(\theta_{2i-1}t)}{\sqrt{2 - \frac{\sin(2\theta_{2i-1}a)}{2\theta_{2i-1}}}}, \text{for all } i \ge 1,\ t \in [-a, a].$$

The θ_k's are solutions to the following transcendental equation

$$1 - l\theta\tan(a\theta) = 0 = l\theta + \tan(a\theta).$$

See p. 23, Section 2.3 of [155] or [245] for a derivation of such an expansion.

For more general forms of covariance functions $C(t, s)$, it may not be possible to find explicitly the eigenvectors and eigenvalues. The Karhunen-Loève expansion can be found numerically, and in practice only a finite number of terms in the expansion are required. Specifically, we usually perform a principal component analysis by truncating the sum at some N such that

$$\frac{\sum_{i=1}^{N} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i} = \frac{\sum_{i=1}^{N} \lambda_i}{\int_a^b \mathbb{V}\mathrm{ar}[X(t)]\, dt} \ge \alpha.$$

Here α is typically taken as 0.9, 0.95, and 0.99. The eigenvalues and eigenfunctions are found by solving numerically the following eigenproblem (integral equation):

$$\int_a^b C(t, s)e_k(t)\, ds = \lambda_k e_k(s), \quad s \in [a, b] \text{ and } k = 1, 2, \ldots, N.$$

See, e.g., Section 2.3 of [155] or [419] for a Galerkin method for this problem. We can also apply the Nyström method or the quadrature method, where the integral is replaced with a representative weighted sum. Several numerical methods for representing a stochastic process with a given covariance kernel are presented in [308, Chapter 7], where numerical methods are not based on Karhunen-Loève expansion, but based on Fourier analysis and other methods.

The decay of eigenvalues in the Karhunen-Loève expansion depends on the smoothness of covariance functions.

Definition 2.1.9 ([419]) *A covariance function $C : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ is said to be piecewise analytic/smooth/$H^{p,q}$ on $\mathcal{D} \times \mathcal{D}$, $0 \le p, q \le \infty$ if there exists a partition $\mathfrak{D} = \{D_j\}_{j=1}^J$ into a finite sequence of simplexes D_j and a finite family $\mathfrak{G} = \{G_j\}_{j=1}^J$ of open sets in $\mathbb{R}^d$ such that*

$$\bar{D} = \cup_{j=1}^J \bar{D}_j,\ \bar{D}_j \subseteq G_j, \quad 1 \le j \le J$$

and such that $C|_{D_j \times D_j'}$ has an extension to $G_j \times G_{j'}$ which is analytic in $D_j \times D_{j'}$/is smooth in $D_j \times D_{j'}$/is in $H^p(D_j) \otimes H^q(D_{j'})$ for any pair (j, j').

The following conclusions on the eigenvalues in the Karhunen-Loève expansion are from [419].

Theorem 2.1.10 *Assume that $C \in L^2(\mathcal{D} \times \mathcal{D})$ be a symmetric covariance function which leads to a compact and nonnegative operator from $L^2(\mathcal{D})$ defined by $\mathfrak{C}u(x) = \int_{\mathcal{D}} C(x, y)u(y)\, dy$. If C is piecewise analytic on $\mathcal{D} \times \mathcal{D}$, then the eigenvalues λ_k in the Karhunen-Loève expansion satisfy that*

$$0 \le \lambda_k \le K_1 e^{-K_2 k^{1/d}}, \quad k \ge 1.$$

The constants K_1 and K_2 depend only on the covariance function C and the domain $\mathcal{D}$. If C is piecewise $H^p(\mathcal{D}) \otimes L^2(\mathcal{D})$ with $p \ge 1$, then the eigenvalues λ_k in the Karhunen-Loève expansion decay algebraically fast

$$0 \le \lambda_k \le K_3 k^{-p/d}, \quad k \ge 1.$$

For the Gaussian covariance function, $C(x, y) = \sigma^2 \exp(-(|x - y|^2)/\gamma^2/\mathrm{diam}(\mathcal{D}))$. Then the eigenvalues λ_k in the Karhunen-Loève expansion decay exponentially fast:

$$0 \le \lambda_k \le K_4 \gamma^{-k^{1/d}-2}/\Gamma(0.5k^{1/d}), \quad k \ge 1.$$

An different approach to show the decay of the eigenvalues is presented in [308, Chapter 7] using Fourier analysis for isotropic covariance kernels (the two-point covariance kernel depends only on distances of two points).

Theorem 2.1.11 ([419]) *Assume that the process $a(x, \omega)$ has a covariance function C, which is piecewise analytic/in $H^{p,q}$ on $\mathcal{D} \times \mathcal{D}$. Then the eigenfunctions are analytic/in H^p in each $\bar{D}_j \in \mathfrak{D}$.*

With further conditions on the domain D_j in $\mathfrak{D}$, it can be shown that the derivatives of eigenfunctions $e_k(x)$'s decay at the speed of $|\lambda_k|^{-s}$ when C is piecewise smooth where $s > 0$ is an arbitrary number.

2.2 Brownian motion and white noise

Definition 2.2.1 (One-dimensional Brownian motion) *A one-dimensional continuous time stochastic process $W(t)$ is called a standard Brownian motion if*

- *$W(t)$ is almost surely continuous in t,*
- *$W(t)$ has independent increments,*
- *$W(t) - W(s)$ obeys the normal distribution with mean zero and variance $t - s$.*
- *$W(0) = 0$.*

It can be readily shown that $W(t)$ is Gaussian process. We then call $\dot{W}(t) = \frac{d}{dt}W$, formally the first-order derivative of $W(t)$ in time, *white noise*.

By Example 2.1.6 and Exercise 2.7.7, then the Brownian motion $W(t)$, $t \in [0,1]$ can be represented by

$$W(t) = \sqrt{2}\sum_{i=1}^{\infty} \xi_i \frac{\sin\left(\left(i-\frac{1}{2}\right)\pi t\right)}{\left(i-\frac{1}{2}\right)\pi}, \quad t \in [0,1],$$

where ξ_i's are mutually independent standard Gaussian random variables. The Brownian motion and white noise can also be defined in terms of orthogonal expansions. Suppose that $\{m_k(t)\}_{k\geq 1}$ is a complete orthonormal system (CONS) in $L^2([0,T])$. The Brownian motion $W(t)$, $t \in [0,T]$ can be defined by (see, e.g., [315])

$$W(t) = \sum_{i=1}^{\infty} \xi_i \int_0^t m_i(s)\, ds, \ t \in [0,T], \tag{2.2.1}$$

where ξ_i's are mutually independent standard Gaussian random variables. It can be checked that the Gaussian process defined by (2.2.1) is indeed a standard Brownian motion by Definition 2.2.1. Correspondingly, the white noise is defined by

$$\dot{W}(t) = \sum_{i=1}^{\infty} \xi_i m_i(t), \ t \in [0,T]. \tag{2.2.2}$$

When $m_i(t) = \sqrt{2/T}\cos((i-1/2)\pi t/T)$, $i \geq 1$, then the representation (2.2.1) coincides with the *Karhunen-Loève expansion* of Brownian motion in Example 2.1.6 when $T = 1$.

Definition 2.2.2 (Multidimensional Brownian motion) *A continuous stochastic process $W_t = (W_1(t), \ldots, W_m(t))^\top$ is called an m-dimensional Brownian motion on $\mathbb{R}^m$ when $W_i(t)$ are mutually independent standard Brownian motions on $\mathbb{R}$.*

Definition 2.2.3 (Multidimensional Brownian motion, alternative definitions) *An $\mathbb{R}^d$-valued continuous Gaussian process $X(t)$ with mean function $\mu(t) = \mathbb{E}[X(t)]$ and the covariance function $C(t,s) = \mathbb{C}ov[(X(t), X(s))] = \mathbb{E}[(X(s) - \mu(s))(X(t) - \mu(t))^\top]$ is called a d-dimensional Brownian motion if for any $0 \leq t_0 < t_1 < \cdots < t_n$,*

- *$X(t_i)$ and $X(t_{i+1}) - X(t_i)$ are independent;*
- *the covariance function (a matrix) is a diagonal matrix with entries $\min(t_i, t_j)$, $0 \leq i, j \leq n$.*

When $\mu(t) = 0$ for all t and $C(t,s) = \min(t,s)$, the Gaussian process is called a standard Brownian motion.

2.2.1 Some properties of Brownian motion

Theorem 2.2.4 *The covariance $\mathbb{C}ov[(W(t), W(s))] = \mathbb{E}[W(t)W(s)] = \min(t,s)$.*

- *Time-homogeneity: For any $s > 0$, $\tilde{W}(t) = W(t+s) - W(s)$ is a Brownian motion, independent of $\sigma(W(u),\, u \leq s)$.*
- *Brownian scaling: For every $c > 0$, $cW(t/c^2)$ is a Brownian motion.*
- *Time inversion: Let $\tilde{W}(0) = 0$ and $\tilde{W}(t) = tW(1/t)$, $t > 0$. Then $\tilde{W}(t)$ is a Brownian motion.*

Corollary 2.2.5 (Strong law of large numbers for Brownian motion) *If $W(t)$ is a Brownian motion, then it holds almost surely that*

$$\lim_{t\to\infty} \frac{W(t)}{t} = 0.$$

Theorem 2.2.6 (Law of the iterated logarithm) *Let W_t be a standard Brownian motion. Then*

$$\mathbb{P}(\limsup_{t\to 0} \frac{W_t}{\sqrt{2t\,|\log\log(t)|}} = 1) = 1, \quad \mathbb{P}(\liminf_{t\to 0} \frac{W_t}{\sqrt{2t\,|\log\log(t)|}} = -1) = 1.$$

$$\mathbb{P}(\limsup_{t\to \infty} \frac{W_t}{\sqrt{2t\log\log(t)}} = 1) = 1, \quad \mathbb{P}(\liminf_{t\to \infty} \frac{W_t}{\sqrt{2t\log\log(t)}} = -1) = 1.$$

Example 2.2.7 (Ornstein-Uhlenbeck process) *Consider a centered one-dimensional Gaussian process with exponential covariance function $\exp(-\frac{|t-s|}{\sigma})$. The Gaussian process is usually called a Ornstein-Uhlenbeck process. Suppose that $W(t)$ is a standard Brownian motion. For $t \geq 0$, the Ornstein-Uhlenbeck process can be written as*

$$O(t) = e^{-\frac{t}{\sigma}} W(e^{\frac{2t}{\sigma}}).$$

Example 2.2.8 *The Brownian bridge $X(t)$ is a one-dimensional Gaussian process with time $t \in [0,1]$ and covariance $\mathbb{C}ov[(X(t), X(s))] = \min(t,s) - ts = \begin{cases} s(1-t),\ 0 \le s \le t \le 1. \\ t(1-s),\ 0 \le t \le s \le 1. \end{cases}$*

Suppose that $W(t)$ is a standard Brownian motion. Then $X(t)$ can be represented by

$$X(t) = W(t) - tW(1) = t\big(W(t) - W(1)\big) + (1-t)\big(W(t) - W(0)\big), \quad 0 \le t \le 1.$$

The process $X(t)$ bridges $W(t) - W(1)$ and $W(t) - W(0)$. It can be readily verified that $\mathbb{C}\text{ov}[(X(t),\, X(s))] = \min(t,s) - ts$ and $X(t)$ is continuous and starting from 0. Moreover,

$$W(t) = (t+1)X(\frac{t}{t+1}), \quad X(t) = (1-t)W(\frac{t}{1-t}).$$

Regularity of Brownian motion

For deterministic functions, $f(x)$, $x \in \mathbb{R}$ is Hölder continuous of order α if and only if there exists a constant C such that

$$|f(x+h) - f(x)| \le Ch^{\alpha}, \quad \text{for all } h > 0 \text{ and all } x.$$

When $\alpha = 1$, we call it Lipschitz continuous. When C depends on x, then we call it locally Hölder continuous of order α

$$|f(x+h) - f(x)| \le C(x)h^{\alpha}, \quad \text{for all small enough } h > 0.$$

Definition 2.2.9 *Consider two stochastic processes, $X(t)$ and $Y(t)$, defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We call $Y(t)$ a modification (or version) of $X(t)$ if for every $t \ge 0$, we have*

$$\mathbb{P}(X(t) = Y(t)) = 1.$$

Theorem 2.2.10 (Kolmogorov and Centsov continuity theorem, [255, Section 2.2.B]) *Given a stochastic process $X(t)$ with $t \in [a,b]$, if there exist constants $p > r$, $K > 0$ such that*

$$\mathbb{E}[|X(t) - X(s)|^p] \le K\,|t-s|^{1+r}\,, \text{ for } t, s \in [a,b],$$

then $X(t)$ has a modification $Y(t)$ which is almost everywhere (in ω) continuous: for all $t, s \in [a,b]$,

$$|Y(t,\omega) - Y(s,\omega)| \le C(\omega)\,|t-s|^{\alpha}\,,\ 0 < \alpha < \frac{r}{p}.$$

For $X(\boldsymbol{t})$, $\boldsymbol{t} \in T \subseteq \mathbb{R}^d$, if there exist constants $p > r$, K such that

$$\mathbb{E}[|X(\boldsymbol{t}) - X(\boldsymbol{s})|^p] \leq K\,|\boldsymbol{t} - \boldsymbol{s}|^{d+r}\,, \text{ for } \boldsymbol{t}, \boldsymbol{s} \in T,$$

then $X(\boldsymbol{t})$ has a modification $Y(\boldsymbol{t})$ which is almost everywhere in ω continuous: for all $\boldsymbol{t}$, $\boldsymbol{s} \in T$,

$$\mathbb{E}[\big(\sup_{\boldsymbol{s}\neq\boldsymbol{t}} \frac{|Y(\boldsymbol{t},\omega) - Y(\boldsymbol{s},\omega)|^{\alpha}}{|\boldsymbol{t} - \boldsymbol{s}|}\big)^p] < \infty,\ 0 < \alpha < \frac{r}{p}.$$

Theorem 2.2.11 *For $\alpha < \frac{1}{2}$, the Brownian motion has a modification which is of locally Hölder continuous of order α.*

Proof. For integer $n \geq 1$, by Kolmogorov and Centsov continuity theorem, it only requires to show that

$$\mathbb{E}[|W(t) - W(s)|^{2n}] \leq C_n\,|t - s|^n\,.$$

Then recalling the conclusion from Exercise 2.7.1 leads to the conclusion.

Theorem 2.2.12 ([255, Section 2.9.D]) *The Brownian motion is nowhere differentiable: for almost all ω, the sample path (realization, trajectory) $W(t,\omega)$ is nowhere differentiable as function of t. Moreover, for almost all ω, the path (realization, trajectory) $W(t,\omega)$ is nowhere Hölder continuous with exponent $\alpha > \frac{1}{2}$.*

Definition 2.2.13 (p-variation) *The p-variation of a real-valued function f, defined on an interval $[a,b] \subset \mathbb{R}$, is the quantity*

$$|f|_{p,\mathrm{TV}} = \sup_{\Pi_n,\,|\pi_n|\to 0} \sum_{i=0}^{n-1} |f(x_{i+1}) - f(x_i)|^p,$$

where the supremum runs over the set of all partitions Π_n of the given interval.

Theorem 2.2.14 (Unbounded total variation of Brownian motion) *The paths (realizations, trajectories) of Brownian motion are of infinite total variation almost surely (a.s., with probability one).*

Proof. Without loss of generality, let's consider the interval $[0,1]$.

$$|W|_{1,\mathrm{TV}} = \sup_{\Pi_n} \sum_{i=0}^{n-1} |W(t_{i+1}) - W(t_i)| \geq \sum_{i=0}^{n-1} |W(\frac{i+1}{n}) - W(\frac{i}{n})| =: V_n.$$

Denote by $W(\frac{i+1}{n}) - W(\frac{i}{n}) = \frac{\xi_i}{\sqrt{n}}$. Then ξ_i's are i.i.d. $\mathcal{N}(0,1)$ random variables. Observe that $\mathbb{E}[V_n] = \sqrt{n}\mathbb{E}[|\xi_1|]$ and $\mathbb{V}\mathrm{ar}[V_n] = 1 - (\mathbb{E}[|\xi_1|])^2$.

Then it follows from the Chebyshev inequality (see Appendix D), we have

$$\begin{aligned}\mathbb{P}(V_n \geq \frac{1}{2}\mathbb{E}[|\xi_1|]\sqrt{n}) &= \mathbb{P}(V_n - \mathbb{E}[|\xi_1|]\sqrt{n} \geq -\frac{1}{2}\mathbb{E}[|\xi_1|]\sqrt{n}) \\ &\geq 1 - \mathbb{P}(\big|V_n - \mathbb{E}[|\xi_1|]\sqrt{n}\big| \geq \frac{1}{2}\mathbb{E}[|\xi_1|]\sqrt{n}) \\ &\geq 1 - \frac{\mathbb{V}\mathrm{ar}[V_n]}{\big(\frac{1}{2}\mathbb{E}[|\xi_1|]\sqrt{n}\big)^2} = 1 - 4\frac{1-(\mathbb{E}[|\xi_1|])^2}{n(\mathbb{E}[|\xi_1|])^2}.\end{aligned}$$

Thus we have

$$\mathbb{P}(|W|_{1,\mathrm{TV}} \geq \frac{\mathbb{E}[|\xi_1|]}{2}\sqrt{n}) \geq \mathbb{P}(V_n \geq \frac{\mathbb{E}[|\xi_1|]}{2}\sqrt{n}) = 1 - 4\frac{1-(\mathbb{E}[|\xi_1|])^2}{n(\mathbb{E}[|\xi_1|])^2}.$$

Letting $n \to \infty$, we obtain

$$\mathbb{P}(|W|_{1,\mathrm{TV}} = \infty) = 1.$$

2.2.2 Approximation of Brownian motion

According to the representation in Chapter 2.2, we have at least three approximate representations for Brownian motion by a finite number of random variables.

By Definition 2.2.1, the Brownian motion at time t_{n+1} can be approximated by

$$\sum_{i=0}^{n} \Delta W_i = \sum_{i=0}^{n} \sqrt{\Delta \mathsf{t}_i}\xi_i, \text{ where } \Delta W_i = W(\mathsf{t}_{i+1}) - W(\mathsf{t}_i), \text{ and } \Delta \mathsf{t}_i = \mathsf{t}_{i+1} - \mathsf{t}_i, \tag{2.2.3}$$

where ξ_i's are i.i.d. standard Gaussian random variables. A sample path (realization, trajectory) of Brownian motion is illustrated in Figure 2.1. Here is Matlab code for generating Figure 2.1.

Code 2.1. A sample path of Brownian motion.

```
% One realization of W(t) at time grids k*dt
clc, clear all
t =2.5;
n= 1000;
dt = t / n;
%  Increments of Brownian motion
Winc  = zeros ( n + 1,1);
% Declare the status of random number generator -- Mersenne
% Twister
rng(100, 'twister');
Winc(1:n) = sqrt ( dt ) * randn ( n, 1);
%  Brownian motion - cumulative sum of all previous
%  increments
```

```
W(2:n+1,1) = cumsum ( Winc(1:n) );
figure(10)
plot((1:n+1).'*dt,W,'b-','Linewidth',2);
xlabel('t')
ylabel('W(t)')
axis tight
```

One popular approximation of Brownian motion in continuous time is piecewise linear approximation (also known as polygonal approximation, see, e.g., [457, 481, 482] or [241, p. 396]), i.e.,

$$W^{(n)}(t) = W(\mathtt{t}_i) + (W(\mathtt{t}_{i+1}) - W(\mathtt{t}_i))\frac{t - \mathtt{t}_i}{\mathtt{t}_{i+1} - \mathtt{t}_i}, \quad t \in [\mathtt{t}_i, \mathtt{t}_{i+1}). \tag{2.2.4}$$

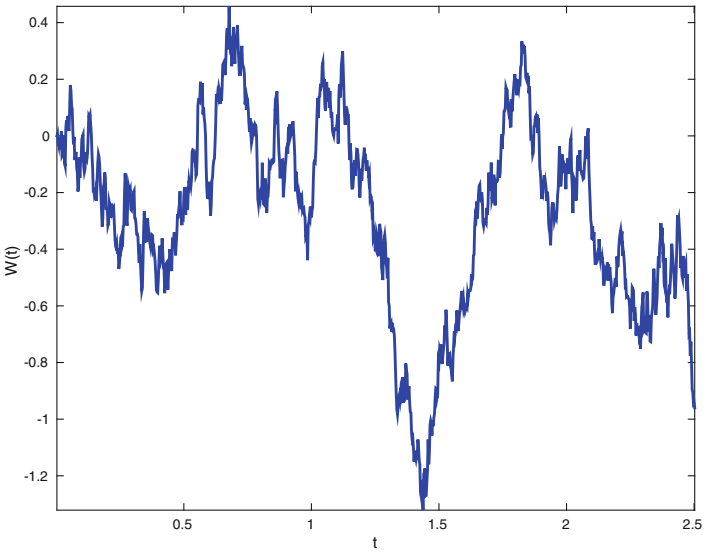
Another way to approximate Brownian motion is by a truncated *orthogonal expansion*:

$$W^{(\mathtt{n})}(t) = \sum_{i=1}^{\mathtt{n}} \xi_i \int_0^t m_i(s)\, ds, \quad \xi_j = \int_0^T m_j(t)\, dW, \quad t \in [0, T], \tag{2.2.5}$$

where $\{m_i(t)\}$ is a CONS in $L^2([0,T])$ and ξ_j are mutually independent standard Gaussian random variables.

In this book, we mostly use the cosine basis $\{m_l(s)\}_{l\geq 1}$ given by

Fig. 2.1. An illustration of a sample path of Brownian motion using cumulative summation of increments.



$$m_1(s) = \frac{1}{\sqrt{T}}, \quad m_l(s) = \sqrt{\frac{2}{T}} \cos(\frac{\pi(l-1)s}{T}), \quad l \geq 2, \quad 0 \leq s \leq T, \quad (2.2.6)$$

or a piecewise version of spectral expansion (2.2.5) by taking a partition of $[0, T]$, e.g., $0 = \mathsf{t}_0 < \mathsf{t}_1 < \cdots \mathsf{t}_{\mathsf{K}-1} < \mathsf{t}_{\mathsf{K}} = T$. We then have

$$W^{(\mathsf{n},\mathsf{K})}(t) = \sum_{k=1}^{\mathsf{K}} \sum_{l=1}^{\mathsf{n}} M_{k,l}(t)\xi_{k,l}, \quad \xi_{k,l} = \int_{I_k} m_{k,l}(s)\, dW(s) \qquad (2.2.7)$$

where $M_{k,l}(t) = \int_{\mathsf{t}_{k-1}}^{\mathsf{t}_k \wedge t} m_{k,l}(s)\, ds$ ($\mathsf{t}_k \wedge t = \min(\mathsf{t}_k, t)$), $\{m_{k,l}\}_{l=1}^{\infty}$ is a CONS in $L^2(I_k)$ and $I_k = [\mathsf{t}_{k-1}, \mathsf{t}_k)$. The random variables $\xi_{k,l}$ are i.i.d. standard Gaussian random variables. Sometimes (2.2.7) is written as

$$W^{(\mathsf{n},\mathsf{K})}(t) = \sum_{k=1}^{\mathsf{K}} \sum_{l=1}^{\mathsf{n}} \int_0^t \mathbf{1}_{I_k}(s) m_{k,l}(s)\, ds \xi_{k,l}, \qquad (2.2.8)$$

where $\mathbf{1}_{\cdot}$ is the indicator function.

Here different choices of CONS lead to different representations. The orthonormal piecewise constant basis over time interval $I_k = [\mathbf{t}_{k-1}, \mathbf{t}_k)$, with $h_k = (\mathbf{t}_k - \mathbf{t}_{k-1})/\sqrt{\mathsf{n}}$, is

$$m_{k,l}(t) = \frac{1}{\sqrt{h}} \chi_{[\mathbf{t}_{k-1}+(l-1)h_k, \mathbf{t}_{k-1}+lh_k)}, \quad l = 1, 2, \cdots, \mathsf{n}. \qquad (2.2.9)$$

When $\mathsf{n} = 1$, this basis gives the classical piecewise linear interpolation (2.2.4). The orthonormal Fourier basis gives Wiener's representation (see, e.g., [259, 358, 391]):

$$m_{k,1} = \frac{1}{\sqrt{\mathbf{t}_k - \mathbf{t}_{k-1}}}, \quad m_{k,2l} = \sqrt{\frac{2}{\mathbf{t}_k - \mathbf{t}_{k-1}}} \sin\big(2l\pi \frac{t - \mathbf{t}_{k-1}}{\mathbf{t}_k - \mathbf{t}_{k-1}}\big),$$
$$m_{k,2l+1}(t) = \sqrt{\frac{2}{\mathbf{t}_k - \mathbf{t}_{k-1}}} \cos\big(2l\pi \frac{t - \mathbf{t}_{k-1}}{\mathbf{t}_k - \mathbf{t}_{k-1}}\big), \; t \in [\mathbf{t}_{k-1}, \mathbf{t}_k). \quad (2.2.10)$$

Note that taking only $m_{k,1}$ ($\mathsf{n} = 1$) in (2.2.10) again leads to the piecewise linear interpolation (2.2.4). Besides, we can also use the Haar wavelet basis, which gives the Levy-Ciesielsky representation [255].

Remark 2.2.15 *Once we have a formal representation (approximation) of Brownian motion, we then can readily obtain a formal representation (approximation) of white noise and thus we skip the formulas for white noise.*

Lemma 2.2.16 *Consider a uniform partition of $[0, T]$, i.e., $\mathsf{t}_k = k\Delta$, $\mathsf{K}\Delta = T$. For $t \in [0, T]$, there exists a constant $C > 0$ such that*

$$\mathbb{E}[\big(W(t) - W^{(\mathsf{n},\mathsf{K})}(t)\big)^2] \leq C\frac{\Delta}{n},$$

and for sufficient small $\epsilon > 0$

$$\left|W(t) - W^{(\mathsf{n},\mathsf{K})}(t)\right| \leq \mathcal{O}\left(\left(\frac{\Delta}{n}\right)^{1/2-\epsilon}\right).^{1} \tag{2.2.11}$$

For $t = \mathsf{t}_k$, *we have*

$$W(\mathsf{t}_k) - W^{(\mathsf{n},\mathsf{K})}(\mathsf{t}_k) = 0, \tag{2.2.12}$$

if the CONS $\{m_{k,l}\}_{l=1}^{\infty}$ *contains* $\frac{1}{\sqrt{\mathsf{t}_k - \mathsf{t}_{k-1}}}$ *as its elements, i.e.,* $\int_{\mathsf{t}_{k-1}}^{\mathsf{t}_k} m_{k,l}(s) \frac{1}{\sqrt{\mathsf{t}_k - \mathsf{t}_{k-1}}}\, ds = \delta_{l,1}$.

Proof. By the spectral approximation of $W(t)$ (2.2.7) and the fact that $\xi_{k,l,i}$ are i.i.d., we have

$$\begin{aligned}
\mathbb{E}[(W(t) - W^{(\mathsf{n},\mathsf{K})}(t))^2] &= \sum_{k=1}^{\mathsf{K}} \sum_{l=\mathsf{n}+1}^{\infty} \left(\int_{\mathsf{t}_{k-1}}^{t\wedge\mathsf{t}_k} m_{k,l}(s)\, ds\right)^2 \\
&= \sum_{k=1}^{\mathsf{K}} \sum_{l=\mathsf{n}+1}^{\infty} \left(\int_{\mathsf{t}_{k-1}}^{\mathsf{t}_k} \chi_{[0,t]}(s) m_{k,l}(s)\, ds\right)^2 \\
&\leq C\frac{\Delta}{n},
\end{aligned}$$

where $\mathsf{t}_k \wedge t = \min(\mathsf{t}_k, t)$ and we have applied the standard estimate for L^2-projection using piecewise orthonormal basis $m_{k,l}(s)$, see, e.g., [417] and have used the fact that the indicator function $\chi_{[0,t]}(s)$ belongs to the Sobolev space $H^{1/2}((0,T))$ for any $T > t$.

Once we have the L^2-estimate, we can apply the Borel-Cantelli lemma (see Appendix D) to obtain the almost sure (a.s.) convergence (2.2.11).

If $t = \mathsf{t}_k$, we have $\int_{\mathsf{t}_{k-1}}^{\mathsf{t}_k} m_{k,l}(s)\, ds = 0$ for any $l \geq 2$ if $m_{k,1} = \frac{1}{\sqrt{\mathsf{t}_k - \mathsf{t}_{k-1}}}$. and thus (2.2.12) holds.

Though any CONS in $L^2([0,T])$ can be used in the spectral approximation (2.2.5), we use a CONS containing a constant in the basis. Consequently, we have the following relation

$$\int_{t_n}^{t_{n+1}} dW^{(\mathsf{n})}(t) = \Delta W_n, \quad \Delta W_n = W(t_{n+1}) - W(t_n). \tag{2.2.13}$$

We will use these approximations in most of the chapters in the book for Wong-Zakai approximation.

[1] The big "$\mathcal{O}$" implies that the error is bounded by a positive constant times the term in the parenthesis.

2.3 Brownian motion and stochastic calculus

As Brownian motion $W(t)$ is not a process of bounded variation, the integral $\int_0^t f(t)\,dW(t)$ cannot be interpreted using Riemann-Stieltjes integration or Lebesgue-Stieltjes integration, even for very smooth stochastic process f. However, it can be understood as Ito integral or Stratonovich integral. For an adapted process $f(t)$ (with respect to the natural filtration of Brownian motion), the Ito integral is defined as, see, e.g., [388], for all partitions of the interval $[0, T]$,

$$\lim_{|\Pi_n|\to 0} \mathbb{E}[\Big(\int_0^T f(t)\cdot dW - \sum_{i=1}^{n-1} f(t_{i-1})\Delta W_i\Big)^2] = 0,$$

where $\Pi_n = \{0 = t_0 < t_1 < t_2 < \cdots < t_n = T\}$ is a partition of the interval $[0, T]$ and $|\Pi_n| = \max_{0\le i\le n-1} |t_{i+1} - t_i|$. The finite sum in this definition is defined at the *left-hand points* in each subinterval of the partition. For Stratonovich calculus, the finite sum is defined at the *midpoints* in each subinterval of the partition. i.e.,

$$\int_0^T f(t)\circ dW = \lim_{|\Pi_n|\to 0} \sum_{i=1}^{n-1} f(\frac{t_{i-1}+t_i}{2})\Delta W_i.$$

Again, the limit is understood in the mean-square sense, see, e.g., [388, 431].

Example 2.3.1 *It can be readily checked that*

$$\int_0^T W(t)\,dW(t) = \frac{W^2(T)-T}{2}, \quad \int_0^T W(t)\circ dW(t) = \frac{W^2(T)}{2}. \quad (2.3.1)$$

Let us show that the first formula holds. By simple calculation and the properties of increments of Brownian motion, for $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$, we have

$$\begin{aligned}
&\mathbb{E}[\Big(\frac{W^2(T)-T}{2} - \sum_{i=0}^{n-1} W(t_i)\Delta W_i\Big)^2] \\
&= \mathbb{E}[\Big(\frac{W^2(T)-T}{2} - \sum_{i=0}^{n-1} \frac{W(t_i)+W(t_{i+1})}{2}\Delta W_i + \frac{1}{2}\sum_{i=1}^{n}(\Delta W_i)^2\Big)^2] \\
&= \mathbb{E}[\Big(\frac{W^2(T)-T}{2} - \sum_{i=0}^{n-1} \frac{W^2(t_{i+1})-W^2(t_i)}{2} + \frac{1}{2}\sum_{i=1}^{n}(\Delta W_i)^2\Big)^2] \\
&= \mathbb{E}[\Big(\frac{-T}{2} + \frac{1}{2}\sum_{i=1}^{n}(\Delta W_i)^2\Big)^2] \to 0, \quad n\to\infty.
\end{aligned}$$

The second formula can be derived similarly.

Moreover, we have the following conversion formula.

Theorem 2.3.2 (Conversion of a Stratonovich integral to an Ito integral) *A Stratonovich integral can be computed via the Ito integral:*

$$\int_0^T f(t, W(t) \circ dW(t) = \int_0^T f(t, W(t))\, dW(t) + \frac{1}{2}\int_0^T \partial_x f(t, W(t))\, dt.$$

Here $f(t, W(t))$ is a scalar function and $\partial_x f$ is the derivative with respect to the second argument of f. When $f \in \mathbb{R}^{m\times n}$ is a matrix function, then

$$[\int_0^T f(t, W(t) \circ dW(t)]_i = [\int_0^T f(t, W(t))\, dW(t)]_i + \frac{1}{2}\int_0^T \sum_{j=1}^n \partial_{x_j} f_{i,j}(t, W(t))\, dt, \quad i = 1, 2, \cdots, m.$$

Here v_i means the i-th component of a vector v.

The proof can be done using the definition of two integrals and mean value theorem. We leave the proof to interested readers.

Definition 2.3.3 (Quadratic covariation) *The quadratic covariation of two processes X and Y is*

$$[X, Y]_t = \lim_{|\Pi_n|\to 0} \sum_{k=1}^n (X(t_k) - X(t_{k-1}))\,(Y(t_k) - Y(t_{k-1}))\,.$$

Here $\Pi_n = \{0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = t\}$ is an arbitrary partition of the interval $[0, t]$.

When $X = Y$, the quadratic covariation becomes the quadratic variation:

$$[X]_t = [X, X]_t = \lim_{|\Pi_n|\to 0} \sum_{k=1}^n X(t_k) - X(t_{k-1})^2.$$

The quadratic covariation can be computed by the polarization identity:

$$[X, Y]_t = \tfrac{1}{4}([X + Y]_t - [X - Y]_t).$$

With the definition of quadratic covariation, we have

$$\int_0^T f(t, W(t) \circ dW(t) = \int_0^T f(t, W(t))\, dW(t) + \frac{1}{2}\int_0^T \partial_x f(t, W(t))\, d[W, W]_t.$$

More generally, we have the following conversion rule

$$\int_0^t Y(s) \circ dX(s) = \int_0^t Y(s)\, dX(s) + \frac{1}{2}[X, Y]_t. \tag{2.3.2}$$

For Ito integral, we have the following properties. Define

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

Theorem 2.3.4 *For $f, g \in \mathbb{L}^2_{ad}(\Omega; L^2([0,T]))$, we have*

- *(linear property)* $\int_0^t (af(s) + bg(s))\, dW(s) = a\int_0^t f(s)\, dW(s) + b\int_0^t g(s)\, dW(s), \quad a, b \in \mathbb{R},$
- *(Ito isometry)* $\mathbb{E}[\big(\int_0^t f(s)\, dW(s)\big)^2] = \int_0^t \mathbb{E}[f^2(s)]\, ds,$
- *(Generalized Ito isometry)* $\mathbb{E}[\int_0^t f(s)\, dW(s) \int_0^t g(s)\, dW(s)] = \int_0^t \mathbb{E}[f(s) g(s)]\, dt,$
- $M_t = \int_0^t f(s)\, dW(s)$ *is a continuous martingale. Moreover, the quadratic variation of* M_t *is* $|M|_{2,\mathrm{TV}} = \int_0^t f^2(s)\, ds$, $0 \le t \le T$ *and*

$$\mathbb{E}[\sup_{0\le t\le T} \big(\int_0^t f(s)\, dW(s)\big)^2] \le 4\mathbb{E}[\int_0^t \mathbb{E}[f^2(s)]\, ds].$$

Example 2.3.5 (Multiple Ito integral) *Assuming that $W(t)$ is a standard Brownian Motion, show that*

$$\frac{1}{n!}\int_0^t \int_0^{t_n} \cdots \int_0^{t_2} dW(t_1) \cdots dW(t_n) = t^{n/2} H_n(\frac{W(t)}{\sqrt{t}}).$$

Here H_n is the n-th Hermite polynomial:

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}. \tag{2.3.3}$$

When $n = 0$, $H_n(x) = 1$ and we use the convention that when $n < 1$ the integral is defined as 1. When $n = 1$, $\int_0^t dW(t_1) = W(t) = t^{1/2} H_1(\frac{W(t)}{\sqrt{t}})$ as $H_1(x) = x$. Then it can be shown by induction that the integrand in the left-hand side is in $\mathbb{L}^2_{\mathrm{ad}}(\Omega; L^2([0,t]))$ and thus the multiple integral is indeed an Ito integral and is equal to the right-hand side.

When we want to define the integral $\int_0^t f(t)\, dW(t)$ via a spectral representation of Brownian motion instead of using increments of Brownian motion, we have to use the so-called Ito-Wick product (Wick product) and Stratonovich product.

The use of Ito-Wick product relies on two facts: the integrand f can be expressed as Hermite polynomial expansion of some random variables ("random basis") and the product is well defined for these basis. The basis and the Ito-Wick product will be shown and defined shortly. Specifically, let $\{\xi_k\}_{k=1}^{\infty}$ be a sequence of mutually independent standard Gaussian random variables from a spectral representation of Brownian motion and let $\mathcal{F} = \sigma(\xi_k)_{k\geq 1}$. The following Cameron-Martin theorem states that any element in $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ can be represented by a linear combination of elements in the Cameron-Martin basis (2.3.4).

Theorem 2.3.6 (Cameron-Martin [57]) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. The collection $\Xi = \{\xi_\alpha\}_{\alpha\in\mathcal{J}}$ is an orthonormal basis in $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, where ξ_α's are defined as*

$$\xi_\alpha := \prod_{\alpha} \left(\frac{H_{\alpha_l}(\xi_l)}{\sqrt{\alpha_l!}} \right), \ \xi_l = \int_0^t m_l(s)\, dW(s), \quad \alpha \in \mathcal{J}, \tag{2.3.4}$$

where $\{m_l\}$ is a complete orthonormal basis in $L^2([0,t])$ and $\mathcal{J}$ is the collection of of multi-indices $\alpha = (\alpha_l)_{l\geq 1}$ of finite length, i.e.,

$$\mathcal{J} = \left\{ \alpha = (\alpha_l)_{l\geq 1}, \quad \alpha_l \in \{0, 1, 2, \ldots\}, \ |\alpha| := \sum_l \alpha_l < \infty \right\}.$$

Any $\eta \in \mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ can be represented as the following Wiener chaos expansion

$$\eta = \sum_{\alpha\in\mathcal{J}} \eta_\alpha \xi_\alpha, \quad \eta_\alpha = \mathbb{E}[\eta\xi_\alpha], \quad \text{and} \quad \mathbb{E}[\eta^2] = \sum_{\alpha\in\mathcal{J}} \eta_\alpha^2. \tag{2.3.5}$$

The collection Ξ of random variables $\{\xi_\alpha, \alpha \in \mathcal{J}\}$ is called the Cameron-Martin basis. It can be readily shown that $\mathbb{E}[\xi_\alpha \xi_\beta] = 1$ if $\alpha = \beta$ and 0 otherwise. See some specific examples of the Cameron-Martin basis in Chapter 2.5.3.

Following [223, Section 2.5] and [316], under certain conditions on $f(t)$ (continuous semi-martingale with respect to the natural filtration of Brownian motion), we have

$$\int_0^t f(t)\, dW(t) = \int_0^t f(t) \diamond \dot{W}(t)\, dt, \tag{2.3.6}$$

where the definition of Ito-Wick product "$\diamond$" is based on the product for elements of the Cameron-Martin basis: $\xi_\alpha \diamond \xi_\beta = \sqrt{\frac{(\alpha+\beta)!}{\alpha!\beta!}}\xi_{\alpha+\beta}$.

With the approximation (2.2.5), Ogawa [386] defined the following so-called Ogawa integral,

$$\int_0^t f(t) * dW(t) = \sum_{i=1}^{\infty} \int_0^t f(s) m_i(s)\, ds \int_0^t m_i(s)\, dW(s), \quad t \in [0,T] \quad (2.3.7)$$

where $\{m_i(t)\}$ is the CONS on $L^2([0,T])$. (Note that Ogawa's original integral is only defined when $t = T$.) Nualart and Zakai [385] proved that the Ogawa integral is equivalent to the Stratonovich integral if the Ogawa integral exists, with the Stratonovich integral defined as

$$\int_0^t f(t) \circ dW(t) = \lim_{|\Pi_n| \to 0} \sum_{i=1}^{n} \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} f(s)\, ds (W(t_{i+1}) - W(t_i)). \quad (2.3.8)$$

Moreover, if the integrand $f(t)$ is a continuous semi-martingale (e.g., Ito process in Definition 2.4.2) on the natural filtration of $W(t)$, then the Ogawa integral coincides with the Stratonovich integral defined at the mid-points.

$$\int_0^t f(t) \circ dW(t) = \lim_{|\Pi_n| \to 0} \sum_{i=1}^{n} f(\frac{t_i + t_{i+1}}{2})(W(t_{i+1}) - W(t_i)). \quad (2.3.9)$$

As application of stochastic integrals, the fractional Brownian motion $B_H(t)$, $t \geq 0$, can be introduced. It is a centered Gaussian process with the following covariance function

$$\mathbb{E}[B_H(t) B_H(s)] = \tfrac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}), \quad 0 < H < 1.$$

The constant H is called the Hurst index or Hurst parameter. The fractional Brownian motion can be represented by

$$B_H(t) = B_H(0) + \frac{1}{\Gamma(H+1/2)} \left\{ \int_{-\infty}^{0} \left[(t-s)^{H-1/2} - (-s)^{H-1/2} \right] dW(s) + \int_0^t (t-s)^{H-1/2}\, dW(s) \right\}.$$

2.4 Stochastic chain rule: Ito formula

One motivation of Ito formula is to evaluate Ito integral with a complicated integrand.

Theorem 2.4.1 (Ito formula in the simplest form) *If f and its first two derivatives are continuous on $\mathbb{R}$, then it holds with probability one (almost surely, a.s.) that*

$$f(W(t)) = f(W(t_0)) + \int_{t_0}^{t} f'(W(s))\,dW(s) + \frac{1}{2}\int_{t_0}^{t} f''(W(s))\,ds.$$

From the theorem, we can compute the Ito integral $\int_{t_0}^{t} f'(W(s))\,dW(s)$ by

$$\int_{t_0}^{t} f'(W(s))\,dW(s) = f(W(t)) - f(W(t_0)) - \frac{1}{2}\int_{t_0}^{t} f''(W(s))\,ds.$$

Definition 2.4.2 (Ito process) *An Ito process is a stochastic process of the form*

$$X_t = X(t_0) + \int_{t_0}^{t} a(s)\,ds + \int_{t_0}^{t} \sigma(s)\,dW(s),$$

where $X(t_0)$ is $\mathcal{F}_{t_0}$-measurable, a_s and σ_s are adapted w.r.t. $\mathcal{F}_s$ and

$$\int_{t_0}^{t} |a(s)|\,ds < \infty, \quad \int_{t_0}^{t} \|\sigma(s)\|^2\,ds < \infty \ a.s..$$

The filtration $\{\mathcal{F}_s,\, t_0 \le s \le t\}$ is defined such that

- for any s, B_s, $a(s)$ and $\sigma(s)$ are $\mathcal{F}_s$-measurable;
- for any $t_1 \le t_2$, $B_{t_2} - B_{t_1}$ is independent of $\mathcal{F}_{t_1}$.

Suppose that $X(t)$ exists a.s. such that

$$X(t) = X(t_0) + \int_{t_0}^{t} a(s, X(s))\,ds + \sum_{r=1}^{m}\int_{t_0}^{t} \sigma_r(s, X(s))\,dW_r(s).$$

Here $X(t)$, $X(t_0), a$, $\sigma_r \in \mathbb{R}^d$ and $\sigma \in \mathbb{R}^{d\times m}$. Also, $W_r(s)$'s are mutually independent Brownian motions, $a(s)$ and $\sigma(s)$ are adapted w.r.t. $\mathcal{F}_s$, and

$$\int_{t_0}^{t} |a(s, X(s))|\,ds < \infty, \quad \sum_{r=1}^{m}\int_{t_0}^{t} |\sigma_r(s, X(s))|^2\,ds < \infty \text{ a.s..}$$

The filtration $\{\mathcal{F}_s,\, t_0 \le s \le t\}$ is defined such that for any s, $W_r(s)$ is $\mathcal{F}_s$-measurable and for any $t_1 \le t_2$, $W_r(t_2) - W_r(t_1)$ is independent of $\mathcal{F}_{t_1}$. Ito formula for a $C^1([0,T]; C^2(\mathbb{R}^d))$ function $f(t,\cdot)$ is

$$f(t, X(t)) = f(t_0, X(t_0)) + \sum_{r=1}^{m}\int_{t_0}^{t} \Lambda_r f(s, X(s))\,dW_r(s) + \int_{t_0}^{t} \mathcal{L} f(s, X(s))\,ds,$$

where

$$\Lambda_r = \sigma_r^{\top}\nabla = \sum_{i=1}^{d} \sigma_{i,r}\frac{\partial}{\partial x_i}, \quad \nabla = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_d}\right),$$

$$\mathcal{L} = \frac{\partial}{\partial t} + a^{\top}\nabla + \frac{1}{2}\sum_{r=1}^{m}\sum_{i,j=1}^{d} \sigma_{i,r}\sigma_{j,r}\frac{\partial^2}{\partial x_i \partial x_j}.$$

Remark 2.4.3 *For the multi-dimensional Ito formula, we can use the following table to memorize the formula when $W_j(t)$ are mutually independent Brownian motions.*

$\times$	$dW_j(t)$	dt
$dW_i(t)$	$\mathbf{1}_{\{i=j\}}dt$	0
dt	0	0

Theorem 2.4.4 (Integration by parts formula) *Let $X(t)$ and Y_t be Ito processes defined in Definition 2.4.2. Then the following integration by parts formula holds*

$$X(t)Y(t) = X(t_0)Y(t_0) + \int_{t_0}^{t} X(s)\,dY(s) + \int_{t_0}^{t} Y(s)\,dX(s) + \int_{t_0}^{t} dX(s)\,dY(s).$$

Here $dX(s)\,dY(s)$ can be formally computed using the table in Remark 2.4.3.

This integration by parts formula is a corollary of multi-dimensional Ito formula for Ito processes.

Consider two Ito processes, X and Y: $dX = a^X(t)\,dt + \sigma^X(t)\,dW(t)$ and $dY = a^Y(t)\,dt + \sigma^Y(t)\,dW(t)$. Then we have from Remark 2.4.3 that

$$dXdY = (a^X(s)\,dt + \sigma^X(t)\,dW(t)(a^Y(s)\,dt + \sigma^Y(t)\,dW(t))) = \sigma^X(t)\sigma^Y(t)\,dt.$$

2.5 Integration methods in random space

Numerical SODEs and SPDEs are usually dependent on the Monte Carlo method and its variants to obtain the desired statistics of the solutions. The fundamental quantity of interest is of the following form

$$\frac{1}{(\sqrt{2\pi})^d}\int_{\mathbb{R}^d} f(x)e^{-\frac{|x|^2}{2}}\,dx.$$

This is a standard numerical integration problem. Let us consider $d = 1$ in a general setting. An integral can be treated as an expectation of random variable

$$\int_{\mathbb{R}} f(x)p(x)\,dx = \mathbb{E}[f(X)], \tag{2.5.1}$$

where X has a probability density $p(x) \geq 0$ and $\int_{\mathbb{R}} p(x)\,dx = 1$ (Here $p(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$).

2.5.1 Monte Carlo method and its variants

A simple (standard, brute force) Monte Carlo estimator of this integration is

$$\mu = \mathbb{E}[f(X)] \approx \frac{1}{n}\sum_{i=1}^{n} f(X_i) =: \hat{\mu} \tag{2.5.2}$$

where X_i are copies of X (i.i.d. with X). The convergence of (2.5.2) is guaranteed by the law of large numbers. Observe that $\hat{\mu}$ is a random variable and the mean of this estimator is $\mathbb{E}[\hat{\mu}] = \mu$, the variance of the estimator is

$$\mathbb{V}\mathrm{ar}[\hat{\mu}] = \frac{\mathbb{V}\mathrm{ar}[f(X)]}{n}. \tag{2.5.3}$$

The error of the Monte Carlo estimator is measured by the following *confidence interval* (95%)

$$(\mu - 1.96\sqrt{\frac{\mathbb{V}\mathrm{ar}[f(X)]}{n}},\ \mu + 1.96\sqrt{\frac{\mathbb{V}\mathrm{ar}[f(X)]}{n}}),$$
$$\sigma^2(f(X)) = \mathbb{E}[f^2(X)] - (\mathbb{E}[f(X)])^2.$$

In practice, we use the following since we don't know μ:

$$(\hat{\mu} - 1.96\sqrt{\frac{\mathbb{V}\mathrm{ar}[f(X)]}{n}},\ \hat{\mu} + 1.96\sqrt{\frac{\mathbb{V}\mathrm{ar}[f(X)]}{n}}),$$
$$\sigma^2(f(X)) = \mathbb{E}[f^2(X)] - (\mathbb{E}[f(X)])^2.$$

The error estimate of the Monte Carlo estimator is based on the central limit theorem. Specifically, when n is large, the Monte Carlo estimator $\hat{\mu}$ is treated as a Gaussian random variable and

$$Z =: \frac{\hat{\mu} - \mu}{\sqrt{\mathbb{V}\mathrm{ar}[f(X)]/n}} = \frac{\hat{\mu} - \mu}{\sqrt{\mathbb{V}\mathrm{ar}[\hat{\mu}]}}$$

is a standard Gaussian random variable. Here the number 1.96[2] is the approximate z value such that $\mathbb{P}(Z < z) = 0.95$ and the number $1.96\frac{\mathbb{V}\mathrm{ar}[f(X)]}{\sqrt{n}}$ is called the statistical error.

The convergence rate of the Monte Carlo estimator can be also shown by the Chebyshev inequality, the fact that $\mathbb{E}[\hat{\mu}] = \mu$ and (2.5.3):

$$\mathbb{P}(|\mathbb{E}[f(X)] - \hat{\mu}| \geq \sqrt{\frac{\mathbb{V}\mathrm{ar}[\hat{\mu}]}{\varepsilon}}) = \mathbb{P}(|\mathbb{E}[f(X)] - \hat{\mu}| \geq n^{-1/2}\sqrt{\frac{\mathbb{V}\mathrm{ar}[f(X)]}{\varepsilon}}) \leq \varepsilon.$$

For any fixed ε, the error of the Monte Carlo estimator decreases in the rate $\mathcal{O}(n^{-1/2})$ when we increase the number of samples.

In practice, the random numbers X_i are replaced with pseudorandom numbers using pseudorandom number generators, see, e.g., [358, 376]. Also, the variance $\mathbb{V}\mathrm{ar}[f(X)]$ is replaced by its numerical value (called empirical variance):

$$\sqrt{\frac{1}{n-1}\Big(\sum_{i=0}^{n} f^2(X_i) - \hat{\mu}^2\Big)} \text{ or } \sqrt{\frac{1}{n-1}\Big(\sum_{i=0}^{n} \big(f(X_i) - \hat{\mu}\big)^2}. \tag{2.5.4}$$

In Matlab (Matlab 2011b and later version), the estimator of $\mathbb{E}[X^p]$ ($p \geq 1$, $X \sim \mathcal{N}(0,1)$) and the error can be implemented as follows.

[2]The value 2 is used sometimes in this book.

Code 2.2. Estimation of the second moment of a standard Gaussian random variable.

```
% Declare the status of random number generators -- Mersenne
% Twister
 rng(100, 'twister');
 n= 1000;  % the number of sample points
 X =  randn ( n, 1);
 p=2;
 mu_hat =  mean (X.^p);
 mu =1;   % X is a standard Gaussian random variable
 mc_int_err = mu_hat - mu;   %  integration error
 stat_err = 1.96*sqrt(var (X.^2)/n);  %   statistical error
```

When the variance $\mathbb{V}\mathrm{ar}[f(X)]$ is large, the Monte Carlo estimator $\hat{\mu}$ is not accurate: the empirical variance in (2.5.4) may be large and the statistical error in the confidence interval can be too large that the interval is not trusted; moreover, the empirical variance may not be reliable either.

To have an accurate Monte Carlo estimator, i.e., a small confidence interval, e.g., when we want $\sqrt{\frac{\mathbb{V}\mathrm{ar}[f(X)]}{n}} = 10^{-2}$ but $\mathbb{V}\mathrm{ar}[f(X)] = 10$, we then need $n = 10^4\mathbb{V}\mathrm{ar}[f(X)]) = 10^5$ Monte Carlo sample points. To reduce this number of Monte Carlo sample points, we need "the variance $\mathbb{V}\mathrm{ar}[f(X)]$ small." To "reduce" the variance, there are several methods available, such as importance sampling (change of measure), control variates, stratified sampling (decomposing the sampling domain into smaller sub-domains). For such variance reduction methods, one can refer to many books on this topic, such as [259, 264, 376].

For one-dimensional integration, the convergence rate of Monte Carlo method is simply $O(n^{-1/2})$ and too slow to compete with deterministic numerical integration methods. Actually, the advantage of Monte Carlo methods is their efficiency for high dimensional integration (large d). The statistical error of a Monte Carlo estimator does not depend on the dimensionality.

Multilevel Monte Carlo method

One of the recently developed variance reduction methods, the so-called multilevel Monte Carlo method, see, e.g., [156] has attracted a lot of attention for numerical SODEs and SPDEs. The idea of the multilevel Monte Carlo method is to write the desired statistics in a telescoping sum and then to sample the difference terms (between terms defined on two different mesh sizes) in the telescoping sum with a small number of sampling paths, where the corresponding "variance" is small. The multilevel Monte Carlo method for (2.5.1) is based on the following formulation.[3]

[3] As a method of control variate, it can be written in the following form

$$\int_{\mathbb{R}} f(x)p(x)\,dx = \int_{\mathbb{R}} [f(x) - \lambda\big(g(x) - \mathbb{E}[g]\big)]p(x)\,dx,$$

$$\int_{\mathbb{R}} f(x)p(x)\,dx = \int_{\mathbb{R}} [f(x) - f_0(x)]p(x)\,dx + \int_{\mathbb{R}} f_L(x)p(x)\,dx$$
$$+ \sum_{l=0}^{L-1} \int_{\mathbb{R}} [f_l(x) - f_{l+1}(x)]p(x)\,dx,\ L \geq 0.^{4}$$

The control variate f_0 is chosen such that $f_0 \approx f$ and it is much cheaper to simulate f_L (and $\mathbb{E}[f_L(x)]$), and $f_l - f_{l+1}$ has smaller variances than f's. Consider $L = 0$. The Monte Carlo estimate is then

$$\frac{1}{N_0}\sum_{k=1}^{N_0} f_0(X_k) + \frac{1}{N_1}\sum_{k=1}^{N_1}[f(X_{N_0+k}) - f_0(X_{N_0+k})].$$

Let C_0 be the cost for the Monte Carlo estimate of $\mathbb{E}[f_0]$ and C_1 the cost for the Monte Carlo estimate of $\mathbb{E}[f - f_0]$. The total cost of the estimate is $N_0C_0 + N_1C_1$ while the variance is $N_0^{-1}\mathbb{V}\mathrm{ar}[f_0] + N_1^{-1}\mathbb{V}\mathrm{ar}[f - f_0]$. For a fixed cost, we can minimize the variance by choosing

$$\frac{C_0}{C_1} = \frac{V_0/N_0^2}{V_1/N_1^2},\ \text{i.e.,} \frac{N_1}{N_0} = \frac{\sqrt{V_1/C_1}}{\sqrt{V_0/C_0}}.$$

For $L > 0$, the idea is similar and one can take N_l proportional to $\sqrt{V_l/C_l}$. Applying to simulation of statistics of solutions to SDEs, the difference terms can be defined on finer meshes (smaller time step sizes) which would admit smaller variances and thus require a smaller number of sampling paths. The computational cost is thus reduced.

2.5.2 Quasi-Monte Carlo methods

Quasi-Monte Carlo methods (QMC) were originally designed as deterministic integration methods in random space and allowed only moderately high dimensional integrations, see, e.g., [376, 423]. QMC has a similar estimator as the Monte Carlo estimator

$$\mathrm{QMC}_n(f) = \frac{1}{n}\sum_{k=1}^{n} f(\mathbf{x}_k).$$

However, one significant difference is that the sequence x_k is deterministic instead of a random or pseudo-random sequence. The sequence is designed to provide better uniformity (measured in *discrepancy*) than a random sequence and to satisfy the basic (worse-case) bound

---

where the control variate $g(x)$ has known expectation $\mathbb{E}[g]$ (w.r.t. $p(x)$) and g is well correlated with f and optimal value for λ can be estimated by a few samples.

[4] We use the convention that when $L = 0$, the summation is zero.

$$|\mathbb{E}[f] - \mathrm{QMC}_n(f)| \le C(\log n)^k n^{-1} \left| \frac{\partial^d f}{\partial x_1 \cdots \partial x_d} \right|_{1,\mathrm{TV}},$$

where the constants $C > 0$ and $k \ge 0$ do not depend on N but may depend on the dimension d. Here it is required that the mixed derivative of f has a bounded total variation while MC requires only a bounded variance. Compared to Monte Carlo methods, the convergence rate is approximately proportional to n^{-1} instead of $n^{-1/2}$ and there is no statistical error since a deterministic sequence is used. However, the convergence rate is smaller if the dimension d is large (usually less than 40). To overcome the dependence on dimension, one can use randomized quasi-Monte Carlo methods.

Some commonly used quasi-Monte Carlo sequences are Halton and Sobol sequences, Niederreiter sequences, etc. In an example at Chapter 2.5.5 and in Chapter 9.4, we will use randomized quasi-Monte Carlo sequences, Halton sequence, and Sobol sequence, where the Matlab code for these sequences is provided.

We refer to [56] for an introduction to the Monte Carlo method and quasi-Monte Carlo methods and [115] for recent development in deterministic and randomized quasi-Monte Carlo methods.

Compared to the Monte Carlo type method, the following two methods have no statistical errors and allow efficient short-time integration of SPDEs.

2.5.3 Wiener chaos expansion method

The Cameron-Martin theorem (Theorem 2.3.6) provides a spectral representation of square-integrable stochastic processes defined on the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This representation is also known as Wiener chaos expansion, see, e.g., [155, 479, 485].

Let us review what we have in Chapter 2.3. The Cameron-Martin basis is listed in Table 2.2 where there is only one Wiener process in (2.3.4).

Table 2.2. Some elements of the Cameron-Martin basis ξ_α in (2.3.4).

$\lvert\alpha\rvert$	α	ξ_α
0	$\alpha = (0, 0, \ldots)$	1
1	$\alpha = (0, \ldots, 0, 1, 0, \ldots)$	$H_1(\xi_i) = \xi_i$
2	$\alpha = (0, \ldots, 0, 2, 0, \ldots)$	$H_2(\xi_i)/\sqrt{2} = \frac{1}{\sqrt{2}}\left(\xi_i^2 - 1\right)$
2	$\alpha = (0, \ldots, 0, 1, 0, \ldots, 0, 1, 0, \ldots)$	$H_1(\xi_i)H_1(\xi_j) = \xi_i\xi_j$

We need in practice to truncate the number of random variables, i.e., let the elements of α be zero with large indexes. To be precise, we introduce the following notation: the order of multi-index α:

$$d(\alpha) = \max\{l \ge 1 : \alpha_{k,l} > 0 \text{ for some } k \ge 1\}.$$

Also, we need to limit the number of $|\alpha|$. We actually define truncated set of multi-indices

$$\mathcal{J}_{\mathsf{N},\mathsf{n}} = \{\alpha \in \mathcal{J} : \ |\alpha| \le \mathsf{N}, \quad d(\alpha) \le \mathsf{n}\}.$$

In this set, there is a finite number of n dimensional random variables and the number is

$$\sum_{i=0}^{\mathsf{N}} \binom{\mathsf{n}+i-1}{i} = \binom{\mathsf{n}+\mathsf{N}}{\mathsf{N}} = \binom{\mathsf{n}+\mathsf{N}}{\mathsf{n}}.$$

In Table 2.3, we list the elements in a truncated Cameron-Martin basis. More

Table 2.3. Elements of a truncated Cameron-Martin basis ξ_α for a finite dimensional random space where $\alpha \in \mathcal{J}_{\mathsf{N},\mathsf{n}}$, $\mathsf{N} = 2$ and $\mathsf{n} = 2$.

$\lvert\alpha\rvert$	α	ξ_α
0	$\alpha = (0,0)$	1
1	$\alpha = (1,0)$	$H_1(\xi_1) = \xi_1$
1	$\alpha = (0,1)$	$H_1(\xi_2) = \xi_2$
2	$\alpha = (2,0)$	$H_2(\xi_1)/\sqrt{2} = \frac{1}{\sqrt{2}}\left(\xi_1^2 - 1\right)$
2	$\alpha = (0,2)$	$H_2(\xi_2)/\sqrt{2} = \frac{1}{\sqrt{2}}\left(\xi_2^2 - 1\right)$
2	$\alpha = (1,1)$	$H_1(\xi_1)H_1(\xi_2) = \xi_1\xi_2$

examples of the basis can be generated using the representation of the Hermite polynomial (2.3.3). Here are the first seven Hermite polynomials:

$$\begin{aligned}
H_0(x) &= 1, \quad H_1(x) = x,\\
H_2(x) &= x^2 - 1,\\
H_3(x) &= x^3 - 3x,\\
H_4(x) &= x^4 - 6x^2 + 3,\\
H_5(x) &= x^5 - 10x^3 + 15x,\\
H_6(x) &= x^6 - 15x^4 + 45x^2 - 15.
\end{aligned}$$

The Hermite polynomials can be represented (computed) by the three-term recurrence relation

$$H_{n+1}(x) = xH_n(x) - nH_{n-1}(x),\ n \ge 1, \quad H_0(x) = 1,\ H_1(x) = x. \quad (2.5.5)$$

Let's now consider the Wiener chaos expansion for computing the integration (2.5.1). The method is based on the Wiener chaos expansion of $f(X)$, where X is a standard Gaussian random variable. Suppose that

$$f(X) = \sum_{j=0}^{\infty} f_j \xi_j, \ \text{where } \xi_j' s \text{ are from the Cameron-Martin basis.}$$

Once we find f_j, we compute the integration (2.5.1) by

$$\mathbb{E}[f(X)] = \mathbb{E}[\sum_{j=0}^{\infty} f_j \xi_j] = f_0.$$

In practice, the coefficients of the Wiener chaos expansion, f_j's can be numerically computed using the governing stochastic equation. Note that ξ_j are orthonormal and thus $\mathbb{E}[\xi_j] = 0$ for $j \geq 1$ and $f_j = \mathbb{E}[f(X)\xi_j]$. Moreover, we can numerically compute $\mathbb{E}[g(f(X))]$ by

$$\mathbb{E}[g(f(X))] \approx \mathbb{E}[g(\sum_{j=0}^{\mathsf{N}} f_j \xi_j)].$$

We will illustrate the idea of Wiener chaos method as a numerical method in Chapter 2.5.5. We will see that the Wiener chaos method is essentially a spectral Galerkin method in the random space.

2.5.4 Stochastic collocation method

In the framework of deterministic integration methods for SPDEs in random space, another solution for nonlinear SPDEs or linear SPDEs with random coefficient is to employ collocation techniques in random space. Here by stochastic collocation methods, we mean the sampling strategies using high dimensional deterministic quadratures (with certain polynomial exactness) to evaluate desired expectations of solutions to SPDEs.

Let's now consider the stochastic collocation method (SCM) for computing the integration (2.5.1). Note that the deterministic integral can be computed by any quadrature rule. Here we consider a sequence of one-dimensional Gauss–Hermite quadrature rules Q_n with number of nodes $n \in \mathbb{N}$ for univariate functions $\psi(\mathsf{y})$, $\mathsf{y} \in \mathbb{R}$:

$$Q_n \psi(\mathsf{y}) = \sum_{\alpha=1}^{n} \psi(\mathsf{y}_{n,\alpha}) \mathsf{w}_{n,\alpha}, \tag{2.5.6}$$

where $\mathsf{y}_{n,1} < \mathsf{y}_{n,2} < \cdots < \mathsf{y}_{n,n}$ are the roots of the Hermite polynomial $H_n(\mathsf{y}) = (-1)^n e^{\mathsf{y}^2/2} \frac{d^n}{d\mathsf{y}^n} e^{-\mathsf{y}^2/2}$ and $\mathsf{w}_{n,\alpha} = n!/(n^2[H_{n-1}(\mathsf{y}_{n,\alpha})]^2)$ are the associated weights. It is known that $Q_n\psi$ is exactly equal to the integral $I_1\psi$ when ψ is a polynomial of degree less than or equal to $2n-1$, i.e., the polynomial degree of exactness of Gauss–Hermite quadrature rules Q_n is equal to $2n-1$. The integration (2.5.1) can be computed by

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)p(x)\,dx \approx \sum_{k=0}^{\mathsf{N}} f(\mathsf{y}_{\mathsf{N},k})\mathsf{w}_k.$$

The Gauss-Hermite quadrature points $\mathsf{y}_{\mathsf{N},k}$ are the zeros (roots) of the $\mathsf{N}+1$-th order Hermite polynomial $H_{\mathsf{N}+1}$. The weights w_i's are the corresponding quadrature weights. The statistics $\mathbb{E}[g(f(X))]$ can be approximated by

$$\mathbb{E}[g(f(X))] \approx \sum_{k=0}^{\mathsf{N}} g(f(\mathsf{y}_{\mathsf{N},k}))\mathsf{w}_k.$$

The quadrature rule here is a starting point of stochastic collocation methods.

As we are usually looking for numerical approximation of some statistics like $\mathbb{E}[f(X)]$, we can simply look for "function values" on some deterministic quadrature points corresponding to certain quadrature rules. The stochastic collocation methods are then collocating a stochastic equation at these deterministic quadrature points using the classical collocation method in random space and seeking the "function values" at the quadrature points by solving the resulting equations they satisfy.

We will illustrate the idea of stochastic collocation method as a numerical method in Chapter 2.5.5. The stochastic collocation method is a spectral collocation method in the random space.

Smolyak's sparse grid

Sparse grid quadrature is a certain reduction of product quadrature rules, which decreases the number of quadrature nodes and allows effective integration in moderately high dimensions [425] (see also [154, 381, 477]).

Consider a d-dimensional integral of a function $\varphi(y)$, $y \in \mathbb{R}^d$, with respect to a Gaussian measure:

$$I_d\varphi := \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \varphi(y) \exp\left(-\frac{1}{2}\sum_{i=1}^{d} y_i^2\right) dy_1 \cdots dy_d. \tag{2.5.7}$$

We can approximate the multidimensional integral $I_d\varphi$ by a tensor product quadrature rule

$$\begin{aligned} I_d\varphi \approx \bar{I}_d\varphi &:= Q_n \otimes Q_n \cdots \otimes Q_n \varphi(y_1, y_2, \cdots, y_d) \\ &= Q_n^{\otimes d}\varphi(y_1, y_2, \cdots, y_d) \\ &= \sum_{\alpha_1=1}^{n} \cdots \sum_{\alpha_d=1}^{n} \varphi(\mathsf{y}_{n,\alpha_1}, \ldots, \mathsf{y}_{n,\alpha_d})\mathsf{w}_{n,\alpha_1} \cdots \mathsf{w}_{n,\alpha_d}, \end{aligned} \tag{2.5.8}$$

where for simplicity we use the same amount on nodes in all the directions. The quadrature $\bar{I}_d\varphi$ is exact for all polynomials from the space $\mathcal{P}_{k_1} \otimes \cdots \otimes \mathcal{P}_{k_d}$ with $\max_{1\le i\le d} k_i = 2n-1$, where $\mathcal{P}_k$ is the space of one-dimensional polynomials of degree less than or equal to k (we note in passing that this fact is easy to prove using probabilistic representations of $I_d\varphi$ and $\bar{I}_d\varphi$). Computational costs of quadrature rules are measured in terms of a number of function evaluations, which is equal to n^d in the case of the tensor product (2.5.8), i.e., the computational cost of (2.5.8) grows exponentially fast with dimension.

The sparse grid of Smolyak [425] reduces the computational complexity of the tensor product rule (2.5.8) by exploiting the difference quadrature formulas:

$$A(\mathsf{L},d)\varphi := \sum_{d\le|\mathbf{i}|\le\mathsf{L}+d-1} (Q_{i_1}-Q_{i_1-1})\otimes\cdots\otimes(Q_{i_d}-Q_{i_d-1})\varphi,$$

where $Q_0 = 0$ and $\mathbf{i} = (i_1, i_2, \ldots, i_d)$ is a multi-index with $i_k \ge 1$ and $|\mathbf{i}| = i_1 + i_2 + \cdots + i_d$. The number L is usually referred to as the level of the sparse grid. The sparse grid rule (2.5.9) can also be written in the following form [477]:

$$A(\mathsf{L},d)\varphi = \sum_{\mathsf{L}\le|\mathbf{i}|\le\mathsf{L}+d-1} (-1)^{\mathsf{L}+d-1-|\mathbf{i}|}\binom{d-1}{|\mathbf{i}|-\mathsf{L}} Q_{i_1}\otimes\cdots\otimes Q_{i_d}\varphi. \tag{2.5.9}$$

The quadrature $A(\mathsf{L},d)\varphi$ is exact for polynomials from the space $\mathcal{P}_{k_1}\otimes\cdots\otimes\mathcal{P}_{k_d}$ with $|\mathbf{k}| = 2\mathsf{L}-1$, i.e., for polynomials of total degree up to $2\mathsf{L}-1$ [381, Corollary 1].

Denote the set of sparse grid points $\mathsf{x}_\kappa = (\mathsf{x}_\kappa^1, \cdots, \mathsf{x}_\kappa^d)$ by $\mathcal{H}_{\mathsf{L}}^{nq}$, where x_κ^j $(1 \le j \le d)$ belongs to the set of points used by the quadrature rule Q_{i_j}. According to (2.5.9), we only need to know the function values at the sparse grid $\mathcal{H}_{\mathsf{L}}^{nq}$:

$$A(\mathsf{L},d)\varphi = \sum_{\kappa=1}^{\eta(\mathsf{L},nq)} \varphi(\mathsf{x}_\kappa)\mathsf{W}_\kappa, \quad \mathsf{x}_\kappa = (\mathsf{x}_\kappa^1, \cdots, \mathsf{x}_\kappa^d) \in \mathcal{H}_{\mathsf{L}}^{nq}, \tag{2.5.10}$$

where W_κ are determined by (2.5.9) and the choice of the quadrature rules Q_{i_j} and they are called the sparse grid quadrature weights. Due to (2.5.9), the total number of nodes used by this sparse grid rule is estimated by

$$\#\mathcal{H}_{\mathsf{L}}^{nq} = \eta(\mathsf{L},d) \le \sum_{\mathsf{L}\le|\mathbf{i}|\le\mathsf{L}+d-1} i_1\times\cdots\times i_d.$$

Table 2.4 lists the number of sparse grid points up to level 5 when the level is not greater than d.

Table 2.4. The number of sparse grid points for the sparse grid quadrature (2.5.9) using the one-dimensional Gauss-Hermite quadrature rule (2.5.6), when the sparse grid level $\mathsf{L} \le d$.

	$\mathsf{L}=1$	$\mathsf{L}=2$	$\mathsf{L}=3$	$\mathsf{L}=4$	$\mathsf{L}=5$
$\eta(\mathsf{L},d)$	1	$2d+1$	$2d^2+2d+1$	$\frac{4}{3}d^3+2d^2+\frac{14}{3}d+1$	$\frac{2}{3}d^4+\frac{4}{3}d^3+\frac{22}{3}d^2+\frac{8}{3}d+1$

The quadrature $\bar{I}_d\varphi$ from (2.5.8) is exact for polynomials of total degree $2\mathsf{L}-1$ when $n=\mathsf{L}$. It is not difficult to see that if the required polynomial exactness (in terms of total degree of polynomials) is relatively small, then the sparse grid rule (2.5.9) substantially reduces the number of function evaluations compared with the tensor-product rule (2.5.8). For instance, suppose that the dimension $d=40$ and the required polynomial exactness is equal to 3. Then the cost of the tensor product rule (2.5.8) is $3^{40} \doteq 1.215\,8\times 10^{19}$ while the cost of the sparse grid rule (2.5.9) based on the one-dimensional rule (2.5.6) is 3281.

2.5.5 Application to SODEs

Let's consider the stochastic differential equation (1.1.2) where

$$dX = W(t)X\,dt, \quad 0<t\le 1 \quad X(0)=X_0=1, \tag{2.5.11}$$

where $W(t)$ is a standard Brownian motion.

We employ the simplest time discretization – forward Euler scheme. For a uniform partition of $[0,1]$, $t_k=kh$, $1\le k\le N$, and $Nh=1$. The forward Euler scheme is

$$X_{k+1}=X_k+W(t_k)X_kh, \quad 0\le k\le N-1. \tag{2.5.12}$$

The goal here is to numerically compute the mean and variance of the solution or simply $\mathbb{E}[X_N^p]$, $p=1,2$.

Here we notice that $W(t_k)$ need further discretization. We recall that in Chapter 2.2.2, there are two ways to approximate $W(t_k)$. The first one is to use increments of Brownian motions (2.2.3) and the forward Euler scheme becomes

$$X_{k+1}=X_k+\sqrt{h}\sum_{i=0}^{k}\xi_iX_kh, \quad 0\le k\le N-1. \tag{2.5.13}$$

Here $\xi_0=0$. Then we have that the solution can be represented by $X_N(\xi_1,\xi_2,\ldots,\xi_{N-1})$, and the desired statistics are of the form (2.5.7). We can then use the methods described in the last section.

For Monte Carlo methods, we can use pseudo-random number generators. In Matlab, the Monte Carlo method for (2.5.13) can be implemented as follows:

Code 2.3. Monte Carlo method with the forward Euler scheme for Equation (2.5.11) using (2.5.13).

```
clc, clear all
rng(100, 'twister');    % for repeatable pseudo random
% sequences.
t_final = 1;   x_ini= 1;
```

```
N= 1000;  h =t_final/N;
num_sample_path = 1e4;
% time marching, Euler scheme
W_k = 0;
X_k = x_ini* ones(num_sample_path,1);
for  k= 1: N-1
     W_k = W_k + sqrt(h)* randn(num_sample_path,1);
     X_k = X_k + W_k.*X_k*h;
end
X_mean = mean(X_k);
X_second_moment = mean(X_k.^2);
X_mean_stat_error = 1.96 * sqrt(var(X_k)/ num_sample_path);
X_second_moment_stat_error = 1.96 * sqrt (var (X_k.^2)/
                              num_sample_path);
```

For quasi-Monte Carlo method, we use the scrambled Halton sequence [262],

Code 2.4. Quasi-Monte Carlo method with the forward Euler scheme for Equation (2.5.11) using (2.5.13).

```
clc, clear all
t_final = 1;   x_ini= 1;
N= 1000;  h =t_final/N;
num_sample_path = 1e4;
rng(100, 'twister');    % for repeatable  randomized
% quasi-random sequences.
qmc_sequence= haltonset(N-1 ,'Skip',1e3,'Leap',20);
% Halton sequence
qmc_sequence= scramble( qmc_sequence, 'RR2');   % scramble,
% randomizing
qmc_sequence = net( qmc_sequence, num_sample_path);
qmc_sequence=erfinv(2*qmc_sequence-1)*sqrt(2);  % an inverse
% transformation for Gaussian
% time marching, Euler scheme
W_k = 0;
X_k = x_ini* ones(num_sample_path,1);
for  k= 1: N-1
    W_k = W_k + sqrt(h)* qmc_sequence(:,k);
    X_k = X_k + W_k.*X_k*h;
end
X_mean = mean(X_k);
X_second_moment = mean(X_k.^2);
X_mean_stat_error = 1.96 * sqrt(var(X_k)/ num_sample_path);
X_second_moment_stat_error = 1.96 * sqrt( var (X_k.^2)/
                             num_sample_path);
```

In this code, the scrambled Halton sequence is used, see [262]. We can also use the scrambled Sobol sequence [340] instead of a scrambled Halton sequence.

Code 2.5. A scrambled Sobol sequence of quasi-Monte Carlo method.

```
qmc_sequence = sobolset(N-1,'Skip',1e3,'Leap',20);
% Sobol sequence
qmc_sequence = scramble(qmc_sequence,'MatousekAffineOwen');
qmc_sequence = erfinv(2*qmc_sequence-1)*sqrt(2);
% inverse transformation
```

A stochastic collocation method requires the values of X_k at quadrature points (in random space) according to (2.5.8). To find these values, we apply the collocation method in random space: for $\kappa = 1, \ldots, \eta(\mathsf{L}, N-1)$,

$$\begin{aligned}&\mathbb{E}[X_{k+1}(\xi_1,\ldots,\xi_k)\delta((\xi_1,\ldots,\xi_k)-(\mathsf{x}_\kappa^1,\cdots,\mathsf{x}_\kappa^k))]\\ =\ &\mathbb{E}[(1+\sqrt{h}\sum_{i=0}^{k}\xi_i h))X_k(\xi_1,\ldots,\xi_{k-1})\delta((\xi_1,\ldots,\xi_k)-(\mathsf{x}_\kappa^1,\cdots,\mathsf{x}_\kappa^k))].\end{aligned}$$

By the property of the delta function, we have a system of deterministic and decoupled equations:

$$X_{k+1}(\mathsf{x}_\kappa^1,\cdots,\mathsf{x}_\kappa^k) = (1+\sqrt{h}\sum_{i=0}^{k}\mathsf{x}_\kappa^i h)X_k(\mathsf{x}_\kappa^1,\cdots,\mathsf{x}_\kappa^{k-1}),\quad \kappa = 1,\ldots,\eta(\mathsf{L},N-1).$$

Here we use the sparse grid code 'nwspgr.m' at http://www.sparse-grids.de/. We now list the code for sparse grid collocation methods.

Code 2.6. Sparse grid collocation with the forward Euler scheme for Equation (2.5.11) using (2.5.13).

```
clc, clear all
t_final = 1;   X_ini= 1;
N= 40;  h =t_final/N;
sparse_grid_dim = N-1;
sparse_grid_level = 2;
[sparse_grid_nodes, sparse_grid_weights]=nwspgr('GQN', ...
    sparse_grid_dim,  sparse_grid_level);
num_sample_path=size(sparse_grid_weights,1);
% time marching, Euler scheme
W_k = 0;
X_k = X_ini* ones(num_sample_path,1);
for  k= 1: N-1
    W_k = W_k + sqrt(h)* sparse_grid_nodes(:,k);
    X_k = X_k + W_k.*X_k*h;
```

```
end
X_mean = sum(X_k.*sparse_grid_weights);
X_second_moment = sum(X_k.^2.*sparse_grid_weights);
```

Consider now the Wiener chaos method for (2.5.13). Suppose that $X_{k+1} = \sum_{\alpha\in\mathcal{J}_{\mathsf{N},k}} x_{\alpha,k}\xi_\alpha$ for $1 \le k \le N-1$. We first apply a Galerkin method in random space – multiplying by the Cameron-Martin basis ξ_β, $\beta \in\in \mathcal{J}_{\mathsf{N},k}$ over both sides of (2.5.13) and taking expectation (integration over the random space). We then have

$$\mathbb{E}[\xi_\beta \sum_{\alpha\in\mathcal{J}_{\mathsf{N},k}} x_{\alpha,k}\xi_\alpha] = \mathbb{E}[\xi_\beta(1+\sqrt{h}\sum_{i=1}^{k}\xi_i h))\sum_{\alpha\in\mathcal{J}_{\mathsf{N},k-1}} x_{\alpha,k-1}\xi_\alpha].$$

By the orthonormality of the Cameron-Martin basis, we have

$$x_{\beta,k} = x_{\beta,k-1}\mathbf{1}_{\beta_k=0} + h^{3/2}\sum_{i=1}^{k}\sum_{\alpha\in\mathcal{J}_{\mathsf{N},k-1}} x_{\alpha,k-1}\mathbb{E}[\xi_i\xi_\alpha\xi_\beta].$$

This turns the discrete stochastic equation from the forward Euler scheme into a system of deterministic equations of the Wiener chaos expansion coefficients $x_{\beta,k+1}$. This system of deterministic equation of the coefficients is called *propagator* of the discrete stochastic equation (2.5.13).

To solve for $\xi_{\beta,k+1}$, one needs to find the expectations of the triples $\mathbb{E}[\xi_i\xi_\alpha\xi_\beta]$. Recalling the recurrence relation (2.5.5) and orthogonality of Hermite polynomials, the triples are zero unless $\alpha = \beta \pm \varepsilon_i$, where ε_i is a multiindex with $|\varepsilon_i| = 1$ and its only nonzero element is the i-th one. Recalling the (2.5.5) and when $\beta_i = \alpha_i + 1$, we have

$$\begin{aligned}\mathbb{E}[\xi_i\xi_{\alpha_i}\xi_{\beta_i}] &= \mathbb{E}[\xi_i H_{\alpha_i}(\xi_i)H_{\beta_i}(\xi_i)]/\sqrt{\alpha_i!}/\sqrt{\beta_i!}\\ &= \mathbb{E}[H_{\alpha_i+1}(\xi_i)H_{\beta_i}(\xi_i)]/\sqrt{\alpha_i!}\sqrt{\beta_i!} = \sqrt{\alpha_i+1}.\end{aligned}$$

Then, the triples can be computed as

$$\mathbb{E}[\xi_i\xi_\alpha\xi_\beta] = \sqrt{\alpha_i+1}\mathbf{1}_{\alpha+\varepsilon_i=\beta} + \sqrt{\beta_i+1}\mathbf{1}_{\beta+\varepsilon_i=\alpha}. \tag{2.5.14}$$

We have a *propagator* ready for implementation: for $\beta \in \mathcal{J}_{\mathsf{N},k}$,

$$\begin{aligned}x_{\beta,k} = \big(x_{\beta,k-1} + h^{3/2}\sum_{i=1}^{k-1}\big(\sqrt{\beta_i}x_{\beta-\varepsilon_i,k-1} + \sqrt{\beta_i+1}x_{\beta+\varepsilon_i,k-1}\mathbf{1}_{|\beta|\le\mathsf{N}-1}\big)\big)\\ \mathbf{1}_{\beta_k=0} + h^{3/2}x_{\alpha,k-1}\mathbf{1}_{\beta_k=1}\mathbf{1}_{\alpha=(\beta_1,\ldots,\beta_{k-1})}.\end{aligned}$$

Now let's consider the spectral approximation of Brownian motion (2.2.5). Applying the spectral approximation (2.2.5), the problem (2.5.11) can be discretized as follows:

$$d\tilde{X} = W^{(\mathsf{n})}(t)\tilde{X}\,dt, \quad \tilde{X}_0 = 1. \tag{2.5.15}$$

Different from (2.5.13), Equation (2.5.15) is still continuous in time. We need a further discretization in time. Let us again use the forward Euler scheme:

$$\tilde{X}_{k+1} = \tilde{X}_k + W^{(\mathsf{n})}(t_k)\tilde{X}_k h, \quad \tilde{X}_0 = 1. \tag{2.5.16}$$

With the approximation (2.2.5) and the cosine basis (2.2.6), we have

$$\begin{aligned} W^{(\mathsf{n})}(t_k) &= \sum_{i=1}^{\mathsf{n}} \xi_i M_i(t_k), \quad M_1(t) = t,\ M_i(t) = \int_0^t m_i(s)\, ds \\ &= \frac{\sqrt{2}}{(i-1)\pi} \sin(\pi(i-1)t), i \geq 2. \end{aligned}$$

The Monte Carlo method is similar to what we have before. The only difference between Code 2.3 and the code here is in computing the approximated Brownian motion.

Code 2.7. Monte Carlo method with the forward Euler scheme for Equation (2.5.11) using (2.5.16) (WZ).

```
clc, clear all
rng(100, 'twister');    % for repeatable pseudo random
% sequences.
t_final = 1;   x_ini= 1;
N= 1000;  h =t_final/N;
num_sample_path = 1e4;
% time marching, Euler scheme
W_k = 0;
X_k = x_ini* ones(num_sample_path,1);
n=40;  % truncation of specral approximation
xi=randn(num_sample_path,n);
for  k= 1: N-1
    t_k=k*h;
    W_k = t_k*xi(:,1);
    for i=2:n
       W_k = W_k+ sqrt(2)*xi(:,i)*sin(pi*(i-1)*t_k)/(i-1)/pi;
    end
    X_k = X_k + W_k.*X_k*h;
end
X_mean = mean(X_k);
X_second_moment = mean(X_k.^2);
X_mean_stat_error = 1.96 * sqrt( var(X_k)/num_sample_path);
X_second_moment_stat_error = 1.96 * sqrt( var (X_k.^2)/
                             num_sample_path);
```

Similarly, the sparse grid collocation method here differs from Code 2.6 in computing the approximated Brownian motion.

Code 2.8. Sparse grid collocation with the forward Euler scheme for Equation (2.5.11) using (2.5.16) (WZ).

```
clc, clear all
t_final = 1;   X_ini= 1;
N= 10000;  h =t_final/N;
n=40;  % truncation of specral approximation
sparse_grid_dim = n;     % n
sparse_grid_level = 2;  %  less than 5
[sparse_grid_nodes, sparse_grid_weights]=nwspgr('GQN', ...
    sparse_grid_dim,  sparse_grid_level);
num_sample_path=size(sparse_grid_weights,1);
% time marching, Euler scheme
W_k = 0;
X_k = X_ini* ones(num_sample_path,1);
for  k= 1: N-1
    t_k=k*h;
    W_k = t_k*sparse_grid_nodes(:,1);
    for i=2:n
        W_k = W_k+ sqrt(2)* sparse_grid_nodes(:,i)
              *sin(pi*(i-1)*t_k)/(i-1)/pi;
    end
    X_k = X_k + W_k.*X_k*h;
end
X_mean = sum(X_k.*sparse_grid_weights);
X_second_moment = sum(X_k.^2.*sparse_grid_weights);
```

Let us derive the Wiener chaos method for (2.5.16). Suppose that $\tilde{X}_k = \sum_{\alpha\in\mathcal{J}_{\mathsf{N},\mathsf{n}}} x_{\alpha,k}\xi_\alpha$ for $k \geq 1$. We first apply a Galerkin method in random space – multiplying the Cameron-Martin basis ξ_β, $\beta \in \mathcal{J}_{\mathsf{N},,\mathsf{n}}$ over both side of (2.5.13) and taking expectation (integration over the random space). We then have

$$\mathbb{E}[\xi_\beta \sum_{\alpha\in\mathcal{J}_{\mathsf{N},,\mathsf{n}}} x_{\alpha,k+1}\xi_\alpha] = \mathbb{E}[\xi_\beta(1+\sum_{i=1}^{\mathsf{n}}\xi_i M_i(t_k)) \sum_{\alpha\in\mathcal{J}_{\mathsf{N},\mathsf{n}}} x_{\alpha,k}\xi_\alpha].$$

By the orthonormality of the Cameron-Martin basis, we have the following *propagator* of the discrete stochastic equation (2.5.16):

$$x_{\beta,k+1} = x_{\beta,k} + h\sum_{i=1}^{\mathsf{n}} M_i(t_k) \sum_{\alpha\in\mathcal{J}_{\mathsf{N},\mathsf{n}}} x_{\alpha,k}\mathbb{E}[\xi_i\xi_\alpha\xi_\beta].$$

Recall the fact (2.5.14) and we have a *propagator* ready for implementation

$$x_{\beta,k+1} = x_{\beta,k} + h\sum_{i=1}^{\mathsf{n}} M_i(t_k)\big(\sqrt{\beta_i}x_{\beta-\varepsilon_i,k} + \sqrt{\beta_i+1}x_{\beta+\varepsilon_i,k}\mathbf{1}_{|\beta|\leq\mathsf{N}-1}\big),$$
$$\beta \in \mathcal{J}_{\mathsf{N},\mathsf{n}}.$$

In Table 2.5, we present first two moments (mean and second moment) obtained by the Monte Carlo method and quasi-Monte Carlo method and stochastic collocation methods using different approximation of Brownian motion for Equation (2.5.11). We recall from Chapter 1.1 that the first moment at $t = 1$ is $\mathbb{E}[X(1)] = X(0)\exp(1/2) \approx 1.6487$ and the second moment is $\mathbb{E}[X^2(1)] = X^2(0)\exp(2 \approx 7.3891$. From the table, we can see that quasi-Monte Carlo methods give the most accurate values for both the mean and the second moment while the stochastic collocation method is the least accurate. One possible reason for the failure of the stochastic collocation method is the high dimensionality in random space. It is believed that beyond 40 dimension the stochastic collocation method is empirically inefficient.

2.6 Bibliographic notes

Other approximations of Brownian motion. Besides the piecewise linear approximation of Brownian motion (2.2.4), there are several other approximations, e.g., the mollifier approximation (see, e.g., [241, p. 397] and [387]), or the random walk by Donsker's theorem (see, e.g., [255]). However, we omit the discussion of construction of Brownian motion here as only the first two approaches are used in practice and in this book.

Table 2.5. The first two moments using different methods in random space and Euler scheme in time for Equation (2.5.11) at $t = 1$. For Wong-Zakai (WZ) approximation, we use $\mathsf{n} = 40$. The time step size is 10^{-4}. The exact moments are given by (1.1.4): the mean is $\exp(1/3) \approx 1.3956$ and the second moment is $\exp(3/2) \approx 1.9477$.

Methods	Mean	Second moment	Approximation of Brownian motion
Monte Carlo (MC)	1.1856 ± 0.0148	1.9771 ± 0.0717	Increments
Quasi-Monte Carlo (QMC)	1.2237 ± 0.0152	2.0969 ± 0.0673	Increments
Stochastic collocation method (SCM)	1.1545	1.6325	Increments
Monte Carlo (MC)	1.1777 ± 0.0145	1.9353 ± 0.0609	WZ
Quasi-Monte Carlo (QMC)	1.5662 ± 0.0173	3.2312 ± 0.0903	WZ
Stochastic collocation method (SCM)	1.1695	1.7141	WZ

Mollification of Brownian motion is also known as the mollifier approximation and is used in the approximation of stochastic integrals and SODEs, see, e.g., [117, 177, 239, 323, 325, 387, 397, 510, 511]

$$\tilde{W}(t) = \int_{t_n}^{t} \int_{\mathbb{R}} K(\theta, s)\, dW(s)\, d\theta, \quad t \in [\mathtt{t}_n, \mathtt{t}_{n+1}), \tag{2.6.1}$$

where K is symmetric. This type of approximation was proposed for a method of lines for SODEs in [387], where no numerical results were presented. When this approximation is applied to SODEs, consistency (convergence without order) has been proved in [117, 177, 179, 326], etc. In [241], the approaches of piecewise linear approximation and mollification have been unified with proved convergence order, known as Ikeda-Nakao-Yamato-type approximations, see also [195].

Spectral approximation of Brownian motion. With a piecewise constant basis in (2.2.5), the use of multiple Ito integrals (Wiener chaos expansion) and multiple Stratonovich integrals was addressed in [53, 54]. When the spectral basis is chosen as Haar wavelets, the approximation is also known as Levy-Ciesielsky approximation, see, e.g., [255]). The expansion (2.2.1) is an extension of Fourier expansion of Brownian motion proposed by Wiener [391, Chapter IX]. It is also known as Levy-Ciesielski approximation [81, 255, 278], Ito-Niso approximation [243], or Fourier approximation [391]; see [240] for a historical review of this approximation.

The approximation with trigonometric orthonormal basis has been used in Wiener chaos methods (see, e.g., [55, 225, 315, 316, 318, 505]) and will be the approximation for our Wong-Zakai approximation throughout this book. See Chapter 4 for the Wong-Zakai approximation using (2.2.5) for SODEs with time delay.

For the multilevel Monte Carlo method for numerical SPDEs, see, e.g., [1, 25, 75, 82, 440, 441] for elliptic equations with random coefficients, [363–365] for stochastic hyperbolic equations, [24, 158] for stochastic parabolic equations, and [160] for a stochastic Brinkman problem. Ref. [1] proposed time discretization schemes with large stability regions to further reduce the cost of the multilevel Monte Carlo method. However, it has been shown that the performance of multilevel Monte Carlo methods is *not robust*, see, e.g., [264, Chapter 4] when the variances of the desired stochastic processes are large.

Quasi-Monte Carlo methods have also been investigated for numerical SPDEs. Some randomized quasi-Monte Carlo methods have been successfully applied to solve stochastic elliptic equations with random coefficients, see, e.g., [164, 165, 281, 282] where the solution is analytic in random space (parameter space). For a good review on quasi-Monte Carlo methods, see [115].

Wiener chaos expansion methods. As numerical methods, they have been summarized in [155, 439, 487]. This idea of representing a random variable (process) with orthogonal polynomial of the random variable (process) with respect to its corresponding probability density function is not limited to Gaussian random variables (processes) and has been extended to more general cases, see, e.g., [415, 488]. Xiu and Karniadakis [488] developed the Wiener-Askey polynomial chaos using a broad class of Askey's orthogonal polynomials [10]. Soize and Ghanem [427] discuss chaos expansions with re-

spect to an arbitrary probability measure, see also Wan and Karniadakis [468]. These methods are known as generalized polynomial chaos, or gPC, see reviews in, e.g., [484, 485].

For linear problems driven by white noise in space or in time, the Wiener chaos expansion method has been investigated both theoretically (see, e.g., [315–318]) and numerically (see, e.g., [469, 505]). The advantage of Wiener chaos expansion method is that the resulting system of PDEs is linear, lower triangular, and deterministic. Also, the Wiener chaos expansion method can be of high accuracy. However, there are two main difficulties for the Wiener chaos expansion as a numerical method. The first is the *efficiency of long-time integration*. Usually, this method is only efficient for short-time integration, see, e.g., [53, 315]. This limitation can be somewhat removed when a recursive procedure is adopted for computing certain statistics, e.g., first two moments of the solution, see, e.g., [505, 507]. The second is *nonlinearity*. When SPDEs are nonlinear, Wiener chaos expansion methods result in fully coupled systems of deterministic PDEs while the interactions between different Wiener chaos expansion terms necessitate exhaustive computation. This effect has been shown numerically through the stochastic Burgers equation and the Navier-Stokes equations [225].

One remedy for nonlinear problems is to introduce the *Wick-Malliavin approximation* for nonlinear terms. Wick-Malliavin approximations can be seen as a perturbation of a Wick product formulation by adding high-order Malliavin derivatives of the nonlinear terms to the Wick product formulation, see [346, 462] and Chapter 11 for details. Basically, lower level Wick-Malliavin approximation (with lower-order Malliavin derivatives) allows weaker nonlinear interactions between the Wiener chaos expansion terms. Let us consider the Burgers equation with additive noise, for example. When only the Wick product is used (zeroth-order Malliavin derivatives only), the resulting system is lower triangular and contains only one nonlinear equation. When Malliavin derivatives of up to first-order are used, the resulting system of PDEs is only weakly coupled and contains only two nonlinear equations. This approach has been shown to be very efficient for short-time integration of equations with quadratic nonlinearity and small noise, see, e.g., [462].

The Wick product had been formulated in [223] for various SPDEs before the Wick-Malliavin approximation was introduced. The Wick product formulation has been explored with finite element methods in physical space, see, e.g., [254, 327–333, 443–445, 498] and also [469] for a brief review on SPDEs equipped with Wick product. In Chapter 11, we will discuss the Wick-Malliavin approximation for linear elliptic equations with multiplicative noise and some nonlinear equations with quadratic nonlinearity and additive noise.

A stochastic collocation method (SCM) was first introduced in [439] and later on by other authors, see, e.g., [11] and [486]. While WCE (Wiener Chaos Expansion) is a spectral Galerkin method in random space, see, e.g., [155, 488]), SCM is a stochastic version of deterministic collocation

methodology. As collocation methods for deterministic problems, see, e.g., [163]), the stochastic collocation methods exhibit high accuracy comparable to the WCE performance, see, e.g., [123] for elliptic equations with random coefficients.

For stochastic differential equations with color noise, it has been demonstrated in a number of works (see, e.g., [11, 12, 35, 370, 378, 486, 500] and references therein) that stochastic collocation methods (Smolyak's sparse grid collocation, SGC) can be a competitive alternative to the Monte Carlo technique and its variants in the case of differential equations. The success of these methods relies on the solution smoothness in the random space and can usually be achieved when it is sufficient to consider only a limited number of random variables (i.e., in the case of a low dimensional random space). As Wiener chaos methods, stochastic collocation methods are limited in practice as they can be used for a small number of random variables. *Based on empirical evidence (see, e.g., [393]), the use of SGC is limited to problems with random space dimensionality of less than* 40.

More efficient algorithms might be built using *anisotropic SGC methods* [172, 379] or goal-oriented quadrature rules, which employ more quadrature points along the "most important" direction, e.g., [373, 389, 390]. Here we consider only isotropic SGC with predetermined quadrature rules. In fact, the effectiveness of adaptive sparse grids relies heavily on the order of importance in random dimension of numerical solutions to stochastic differential equations, which is not always easy to reach. Furthermore, all these sparse grids as integration methods in random space grow quickly with random dimensions and thus cannot be used for longer time integration (usually with large random dimensions).

For SODEs driven by white noise in time, the stochastic collocation method has been known as cubature on Wiener space (e.g., [202, 300, 322, 373, 377]), optimal quantization (e.g., [389, 390]) to solve SODEs in random space, sparse grid of Smolyak type (e.g., [153, 154, 172, 398]), or particle simulation (e.g., [132]). For stochastic collocation methods for equations with color noise, see, e.g., [11, 486].

The stochastic collocation methods result in decoupled systems of equations as Monte Carlo method and its variants do, which can be of great advantage in parallel computation. High accuracy and fast convergence can be also observed for stochastic evolution equations, e.g., [153, 154, 172, 398] where the sparse grid of Smolyak type was used.

However, the fundamental limitation of these collocation methods is the exponential growth of sampling points with an increasing number of random parameters, see, e.g., [172], and thus a failure for longer time integration, see error estimates for cubature on Wiener space (e.g., [28, 70, 116]) and conclusions for optimal quantization (e.g., [389, 390]).

2.7 Suggested practice

Exercise 2.7.1 *Show that for any centered Gaussian random variable,* $\xi \sim \mathcal{N}(0,\sigma^2)$, $\mathbb{E}[\xi^{2n}] = \sigma^{2n}(2n-1)!!$ *and* $\mathbb{E}[\xi^{2n-1}] = 0$, *for any* $n \geq 1$. *Here* $(2n-1)!! = 1 \times 3 \times \cdots \times (2n-1)$.

Exercise 2.7.2 *Consider a Gaussian vector* (X,Y) *with* X, Y *satisfying* $\mathbb{C}ov[(X,Y)] = 0$ *(uncorrelated). Then* X *and* Y *are independent.*

Exercise 2.7.3 *For a Gaussian random variable* $X \sim \mathcal{N}(0,\sigma^2)$ *and a Bernoulli* Z *with* $\mathbb{P}(Z = \pm 1) = \frac{1}{2}$, X *and* Z *are independent. Show that*

a) the product ZX *is a Gaussian random variable;*
b) and X *and* ZX *are uncorrelated;*
c) but X *and* ZX *are not independent.*

Check whether $(X, ZX)^\top$ *is a Gaussian random vector or not.*

Exercise 2.7.4 *Assume that* X *and* Y *are Gaussian random variables, then* $X+Y$ *independent of* $X-Y$ *implies* X *independent of* Y *if and only if* (X,Y) *is a Gaussian random vector.*

Exercise 2.7.5 *Show that the covariance function* C *is nonnegative definite, i.e., for all* $t_1, \ldots t_k \in T$ *and all* $z_1, z_2, \ldots, z_k \in \mathbb{R}$

$$\sum_{i=1}^{k}\sum_{j=1}^{k} C(t_i,t_j) z_i z_j \geq 0.$$

Hint: assume that $\mathbb{E}[X(t)] = 0$ and write the formula as $\mathbb{E}[\big(\sum_{i=1}^k z_i X(t_i)\big)^2]$.

Exercise 2.7.6 *Assume that* $W(t)$ *is a standard Brownian motion. Show that the covariance of* $W(t)$ *is* $\mathbb{C}ov[(W(t), W(s))] = \min(t,s)$.

Exercise 2.7.7 *If* $X(t)$ *is a one-dimensional Gaussian process with covariance* $\mathbb{C}ov[X(t), X(s)] = \min(t,s)$, *then it is a one-dimensional Brownian motion.*

Exercise 2.7.8 *Use the definition of Brownian motion to show that the process in* (2.2.1) *is indeed a Brownian motion on* $[0,T]$.

Exercise 2.7.9 *Compute the Hölder exponent* α *for the following functions:* $f(t) = \sqrt{t}$, $f(t) = t^\beta$ $(\beta > 0)$ *for* $t \in [0,1]$ *and for* $t \in [\varepsilon, 1]$ $(0 < \varepsilon < 1)$.

Exercise 2.7.10 *Show that if there exists a constant* C *such that for any* x,

$$|f(x+h) - f(x)| \leq Ch^\beta, \ \beta > 1,$$

then $f(x)$ *is a constant.*

Exercise 2.7.11 *For integer $n \geq 1$, verify that*

$$\mathbb{E}[|B_t - B_s|^{2n}] \leq C_n |t-s|^n .$$

Exercise 2.7.12 *If f is differentiable and its derivative is Riemann-integrable, its total variation over the interval $[a,b]$ is*

$$|f|_{1,\mathrm{TV},[a,b]} = \int_a^b |f'(t)|\, dt.$$

Exercise 2.7.13 *Show that a Brownian motion $W(t)$ has a bounded quadratic variation and any p-variation of $W(t)$ is zero when $p \geq 3$.*

Exercise 2.7.14 *Assume that V_n is from Theorem 2.2.14. Show that $\mathbb{E}[V_n] = \sqrt{n}\mathbb{E}[|\xi_1|]$ and $\mathbb{V}ar[V_n] = 1 - (\mathbb{E}[|\xi_1|])^2$.*

Exercise 2.7.15 *Compute $\mathbb{E}[|\xi_1|]$ in Theorem 2.2.14.*

Exercise 2.7.16 *Show that $\xi_l = \int_0^T m_l(s)\, dW(s)$'s are i.i.d. standard Gaussian random variables, where $\{m_l\}$ is a complete orthonormal basis in $L^2([0,T])$.*

Exercise 2.7.17 *Show by definition the second formula in Example 2.3.1.*

Exercise 2.7.18 *Using Taylor's expansion and bounded quadratic variation of Brownian motion, prove Theorem 2.4.1 when f has bounded first two derivatives.*

http://www.springer.com/978-3-319-57510-0

Numerical Methods for Stochastic Partial Differential Equations with White Noise

Zhang, Z.; Karniadakis, G.E.

2017, XV, 394 p. 36 illus., 34 illus. in color. With online files/update., Hardcover
ISBN: 978-3-319-57510-0