AN INTRODUCTION TO SDES

MARIA CAMERON

Contents

1. Brownian Motion

Various processes in nature are often modeled by stochastic differential equations of the form

$$dx = b(x,t)dt + \sigma(x,t)dw,$$

where the function $b(x,t)$ is called *the drift field*, the matrix function $\sigma(x,t)$ is called the diffusion matrix, and the factor dw is the increment of the stochastic process called *the Brownian motion*. The goal of this section is to understand what the Brownian motion is.

1.1. Definition of Brownian Motion.

Definition 1. *A stochastic process (in the strict sense) is a function $v(\omega,t)$ of two arguments, where $\omega \in \Omega$, $(\Omega, \mathcal{B}, P)$ is a probability space, and $t \in \mathbb{R}$, such that*

- *for each ω, $v(\omega,t)$ is a function of t, and*
- *for each t, $v(\omega,t)$ is a random variable.*

Definition 2. *Brownian motion (in mathematical terminology) is a stochastic process $w(\omega,t)$, $\omega \in \Omega$, $0 \le t < \infty$, that satisfies the following four axioms:*

(1) *$w(\omega,0) = 0$ for all ω.*
(2) *For almost all ω, $w(\omega,t)$ is a continuous function of t.*
(3) *For each $0 \le s \le t$, $w(\omega,t) - w(\omega,s)$ is a Gaussian random variable with mean 0 and variance $t - s$.*
(4) *$w(\omega,t)$ has independent increments, i.e., if*

$$0 \le t_1 < t_2 < \ldots < t_n$$

then

$$w(\omega,t_i) - w(\omega,t_{i-1}) \quad \text{for} \quad i = 2,\ldots,n \quad \text{are independent.}$$

Remark What is called the Brownian motion in mathematics is called the Wiener process in physics. What is called the Brownian motion in physics is called the Ornstein-Uhlenbeck process in mathematics.

Here is an equivalent definition of Brownian motion.

Definition 3. *A process $w(\omega,t)$ on a probability space $(\Omega, \mathcal{B}, P)$ is called a Brownian motion if*

(1) *Sample paths $w(\omega,t)$ are continuous functions of t for almost all $\omega \in \Omega$.*
(2) *For any $k > 1$ and $0 \le t_1 \le \ldots \le t_k$, the random vector $(w(\omega,t_1),\ldots,w(\omega,t_k))$ is Gaussian with mean 0 and covariance matrix*

$$B(t_i,t_j) = E[w(t_i), w(t_j)] = \min\{t_i,t_j\} \equiv t_i \wedge t_j, \quad 1 \le i,j \le k.$$

Definition 4. *A d-dimensional Brownian motion is defined as the vector process*

$$w(t) = (w_1(t), \ldots w_d(t)),$$

where $w_k(t)$, $1 \le k \le d$ are independent Brownian motions.

1.2. **Existence of Brownian motion.** The question about the existence of the Brownian motion is not trivial. For example, if we upgrade axiom 2 in Definition 2 to require differentiability, such a process simply would not exist.

The original construction of Brownian motion (the Wiener process) was done by Norbert Wiener (1894 – 1964). He has shown that the Fourier series

$$w(t) = \frac{a_0}{\sqrt{\pi}}t + \sqrt{\frac{2}{\pi}}\sum_{k=1}^{\infty}\frac{a_k}{k}\sin(kt), \tag{1}$$

where a_k, $k = 0, 1, 2, \ldots$, are independent Gaussian random variables with mean 0 and variance 1, converges, and its sum satisfies Definition 2 for $0 \le t \le 1$.

In [2], the existence of Brownian motion follows from Kolmogorov's theorem about the existence of stochastic processes with covariance satisfying certain conditions.

Exercise A simple construction of the Brownian motion can be done as follows. Consider a random walk on a mesh in the (x, t)-space, $t \ge 0$, $x \in \mathbb{R}$ in which the time-step k and the space step h are related to the time step k via $k = h^2$. Start at the origin at time 0. At any discrete moment of time $t = k,\ 2k,\ 3k, \ldots$, take a step to left or to the right with probability 1/2. Let k and h tend to zero in such a manner that the relationship $k = h^2$ is maintained. Apply the Central Limit Theorem and obtain the Brownian motion.

Two constructions of Brownian Motion are discussed in Appendix A.

1.3. **Elementary properties of Brownian motion.**

- **The covariance function** of the Brownian motion is

$$E[w(t_1)w(t_2)] = \min\{t_1, t_2\} \equiv t_1 \wedge t_2. \tag{2}$$

  Indeed, suppose $t_2 > t_1$. Then

$$\begin{aligned} E[w(t_1)w(t_2)] &= E[w(t_1)w(t_1) + w(t_1)(w(t_2) - w(t_1)] \\ &= E[w(t_1)w(t_1)] + E[w(t_1)(w(t_2) - w(t_1))] = t_1. \end{aligned}$$

- **Nowhere differentiability with probability 1.** Consider the random

$$\frac{w(\omega, t + \Delta t) - w(\omega, t)}{\Delta t}.$$

  It is Gaussian with mean 0 and variance $(\Delta t)^{-1}$, which tends to infinity as $\Delta t \to 0$. Hence $w(\omega, t)$ is differentiable nowhere with probability 1.
- **White noise.** Despite the regular derivative of a Brownian motion does not exist, one can consider its derivative in the sense of distributions. This derivative $\eta(\omega, t)$ is called *white noise* and is defined by the property

$$\int_{t_1}^{t_2} \eta(\omega, t)dt = w(\omega, t_2) - w(\omega, t_1).$$

- **Scaling and Symmetry.** If $w(t)$ is a Brownian motion then so are the processes defined by

$$x(t) := \frac{1}{\sqrt{c}} w(ct) \quad \text{for any positive constant} \quad c,$$

$$y(t) = -w(t).$$

- **Time inversion.** Let $w(t)$ be a Brownian motion. Then so is the process defined by

$$x(t) = \begin{cases} tw\left(1/t\right), & 0 < t < \infty, \\ 0, & t = 0. \end{cases} \tag{3}$$

- **Invariance under rotations and reflections (orthogonal transformations).** Let $w(t)$ be a d-dimensional Brownian motion, and T be a $d \times d$ orthogonal matrix (i.e., $T^\top = T^{-1}$). Then the process

$$x(t) = Tw(t)$$

is also a d-dimensional Brownian motion.

1.4. **A brief introduction into the Wiener measure.** When we need to average some function of Brownian motion, the measure to be taken is the Wiener measure. Therefore, we need to specify the measurable space where the Wiener measure will be defined, i.e., the set of outcomes Ω and the σ-algebra on Ω.

The set of outcomes will be the set of continuous functions $y(t)$ satisfying $y(0) = 0$. The σ-algebra will contain all cylinder sets of the form

$$C = \{f(t) \text{ is continuous} \mid a \le f(s) < b\}, \tag{4}$$

where a, b, and s are parameters defining C.

The Wiener measure of each cylinder set (4) is defined so that it is equal to the probability that a Brownian motion $w(\omega, t)$ passes through the window $[a, b)$ at time s:

$$P(C) = \int_a^b \frac{e^{-x^2/2s}}{\sqrt{2\pi s}} dx.$$

The Wiener measure of the intersection of two cylinders

$$C_1 = \{f(t) \text{ is continuous} \mid a_1 \le f(s_1) < b_1\} \quad \text{and} \quad C_2 = \{f(t) \text{ is continuous} \mid a_2 \le f(s_2) < b_2\}$$

is defined so that it is equal to the probability that a Brownian motion $w(\omega, t)$ passes through both windows:

$$a_1 \le w(\omega, s_1) < b_1, \quad \text{and} \quad a_2 \le w(\omega, s_2) < b_2.$$

Assume $s_1 < s_2$. Taking into account that the increments of a Brownian motion are independent Gaussian random variables, we calculate

$$P(C_1 \cap C_2) = \int_{a_1}^{b_1} \frac{e^{-x^2/2s_1}}{\sqrt{2\pi s_1}} dx \int_{a_2}^{b_2} \frac{e^{-(y-x)^2/2(s_2-s_1)}}{\sqrt{2\pi(s_2 - s_1)}} dy.$$

The notation for the Wiener measure is dW. Thus, the solution of the heat equation (**??**) can be written as

$$u(x, t) = \int \phi(x + w(\omega, t)) dW.$$

Example 1 Compute $\int F dW$ where $F(w) = \int_0^1 w^4(\omega, s) ds$.

$$\int F dW = \int dW \int_0^1 w^4(\omega, s) ds = \int_0^1 ds \int dW w^4(\omega, s)$$
$$= \int_0^1 ds \int_{-\infty}^{\infty} x^4 \frac{e^{-x^2/2s}}{\sqrt{2\pi s}} dx = \int_0^1 3s^2 ds = 1.$$

1.5. **Markov property of Brownian motion.**

Definition 5. *A stochastic process $\zeta(t)$ on $[0,T]$ is called a Markov process if for any sequences $0 \le t_0 < \ldots < t_n \le T$ and x_0, x_1, ..., x_n, its transition probability distribution function has the property*

$$\mathbb{P}(\zeta(t_n) < x_n \mid \zeta(t_{n-1}) < x_{n-1}, \ldots, \zeta(t_0) < x_0) = \mathbb{P}(\zeta(t_n) < x_n \mid \zeta(t_{n-1}) < x_{n-1}).$$

The transition probability density function, defined by

$$p(x_n, t_n \mid x_{n-1}, t_{n-1}; \ldots; x_0, t_0) := \frac{\partial}{\partial x_n}\mathbb{P}(\zeta(t_n) < x_n \mid \zeta(t_{n-1}) < x_{n-1}, \ldots, \zeta(t_0) < x_0)$$

then satisfies

$$p(x_n, t_n \mid x_{n-1}, t_{n-1}; \ldots; x_0, t_0) = p(x_n, t_n \mid x_{n-1}, t_{n-1}). \tag{5}$$

For any three times $t > \tau > s$ and any three points x, y, z we can write the identities

$$p(y, t; z, \tau \mid x, s) = p(y, t \mid z, \tau; x, s)p(z, \tau \mid x, s) = p(y, t \mid z, \tau)p(z, \tau \mid x, s).$$

The last equality is a consequence of the Markov property. This identity implies the Chapman-Kolmogorov equation:

$$p(y, t \mid x, s) = \int p(y, t; z, \tau \mid x, s)dz = \int p(y, t \mid z, \tau)p(z, \tau \mid x, s)dz. \tag{6}$$

Theorem 1. *The Brownian motion is a Markov process.*

Proof. Take any sequences $0 = t_0 < \ldots < t_n \le T$ and $x_0 = 0$, x_1, ..., x_n and consider the joint pdf of the vector

$$w = (w(t_1), w(t_2), \ldots, w(t_n)).$$

It is given by

$$p(x_1, t_1; x_2, t_2; \ldots; x_n, t_n) = \prod_{k=1}^{n}\left[\frac{\exp\left\{-\frac{(x_k - x_{k-1})^2}{2(t_k - t_{k-1})}\right\}}{\sqrt{2\pi(t_k - t_{k-1})}}\right]. \tag{7}$$

Recall that

$$p(x_n, t_n \mid x_{n-1}, t_{n-1}; \ldots; x_1, t_1) = \frac{p(x_1, t_1; x_2, t_2; \ldots; x_n, t_n)}{p(x_1, t_1; x_2, t_2; \ldots; x_{n-1}, t_{n-1})}.$$

Using Eq. (7) we get

$$p(x_n, t_n \mid x_{n-1}, t_{n-1}; \ldots; x_1, t_1) = \frac{\prod_{k=1}^{n}\left[\frac{\exp\left\{-\frac{(x_k - x_{k-1})^2}{2(t_k - t_{k-1})}\right\}}{\sqrt{2\pi(t_k - t_{k-1})}}\right]}{\prod_{k=1}^{n-1}\left[\frac{\exp\left\{-\frac{(x_k - x_{k-1})^2}{2(t_k - t_{k-1})}\right\}}{\sqrt{2\pi(t_k - t_{k-1})}}\right]} = \frac{\exp\left\{-\frac{(x_n - x_{n-1})^2}{2(t_n - t_{n-1})}\right\}}{\sqrt{2\pi(t_n - t_{n-1})}}$$
$$= p(x_n, t_n \mid x_{n-1}, t_{n-1}).$$

Hence, the Brownian motion is a Markov process. □

2. An Introduction to SDEs.

Here we follow the discussion found in [1, 4]. Consider a stochastic process $x(\omega, t) \equiv x(t)$ obeying the following evolution law:

$$dx(t) = b(x(t), t)dt + \sigma(x(t), t)dw, \tag{8}$$

where w is the standard Brownian motion and the functions b and σ are smooth.

This evolution law is called a *stochastic differential equation* (SDE). If $\sigma(x(t), t) = 0$, Eq. (8) becomes an ordinary differential equation (ODE)

$$\frac{dx}{dt} = b(x, t). \tag{9}$$

Suppose $x(0) = x_0$. Eq. (9) is equivalent to the following integral equation

$$x(t) = x_0 + \int_0^t b(x(s), s)ds. \tag{10}$$

A solution of an ODE is a function satisfying the ODE. If the ODE is complemented with an initial condition, then the solution to the corresponding initial-value problem is a function satisfying the given initial condition.

Now we will discuss the meaning of SDE (8). Similarly to Eq. (11) we can write

$$x(t) = x_0 + \int_0^t b(x(s), s)ds + \int_0^t \sigma(x(s), s)dw(s) \tag{11}$$

2.1. **Ito's and Stratonovich's stochastic integrals.** First assume that the function σ is independent of x, i.e., $\sigma(x, t) \equiv \sigma(t)$. We partition the interval $[0, t]$ into $t_0 = 0 < t_1 < t_2 < \ldots < t_n = t$ and denote the fineness of the partition by Δ:

$$\Delta := \max_{1 \le i \le n} |t_i - t_{i-1}|.$$

Then we define

$$\int_0^t \sigma(s)dw(s) = \lim_{\substack{n\to\infty \\ \Delta\to 0}} \sum_{i=0}^{n-1} \sigma_i (w(t_{i+1}) - w(t_i))$$

where σ_i is chosen so that it approximates $\sigma(t)$ on the subinterval $[t_i, t_{i+1}]$.

The case where σ depends on x is much more difficult. We can proceed as before and write

$$\int_0^t \sigma(x(s), s)dw(s) = \lim_{\substack{n\to\infty \\ \tau\to 0}} \sum_{i=1}^{n} \sigma_i (w(t_{i+1}) - w(t_i)),$$

however, the value of this limit will depend on how we choose σ_i approximating $\sigma(x(t), t)$ on the interval $[t_i, t_{i+1}]$. There are two common choices. We will give their general definition. Let $f(w(t), t)$ be a smooth function depending on time and a Brownian Motion $w(t)$. In particular, $f(w(t), t)$ can be chosen to coincide with $\sigma(x(t), t)$ where $x(t)$ is a solution of SDE (8), i.e., a stochastic process depending on $w(t)$.

- **The Ito stochastic integral** is defined by the choice $f_i = f(t_i)$, i.e., f is evaluated at the left end of each subinterval:

$$\int_0^t f(w(s), s)dw = \lim_{\substack{n\to\infty \\ \Delta\to 0}} \sum_{i=0}^{n-1} f(w(t_i), t_i)(w(t_{i+1}) - w(t_i)).$$

- **The Stratonovich stochastic integral** is defined by the choice $f_i = f(t_{1+1/2})$, where $t_{i+1/2} \equiv \frac{1}{2}(t_i + t_{i+1})$. i.e., f is evaluated at the midpoint of each subinterval. The Stratonovich stochastic integral is marked by $\circ$:

$$\int_0^t f(w(s), s) \circ dw = \lim_{\substack{n\to\infty \\ \Delta\to 0}} \sum_{i=0}^{n-1} f(w(t_{i+1/2}), t_{i+1/2})(w(t_{i+1}) - w(t_i)).$$

An example in Appendix B demonstrates that Ito's and Stratonovich's stochastic integrals are different. However, one always can convert Stratonovich's stochastic integral to Ito's stochastic integral – see Wiki. From now on, we will work only with Ito's stochastic integrals and Ito's stochastic differential equations.

2.2. **Elementary properties of stochastic integral.** Here we follow the discussions in [4, 5]. We will use a shorter notation denoting $f(w(t), t)$ by $f(t, \omega)$ where ω is the stochastic argument of the Brownian Motion.

We will a consider stochastic process $f(t, \omega)$ on $0 \le t \le T$ satisfying the following conditions:

Condition (1): $f(t, \omega)$ is independent of the increments of the Brownian motion $w(t, \omega)$ in the future, i.e., $f(t, \omega)$ is independent of $w(t + s, \omega) - w(t, \omega)$ for all $s > 0$. Such processes are called *adapted to the Brownian filtration* $\mathcal{F}_t$.

Condition (2):

$$\int_0^T E[f^2(s, \omega)]ds < \infty.$$

Now we list some useful elementary properties. The first two properties are similar to those of the Riemann integral. The other ones are specific for the Ito integral. Let $f(t, \omega)$ and $g(t, \omega)$ be any functions satisfying conditions (1) and (2) above.

(1) Linearity:

$$\int_0^t (af(s, \omega) + bg(s, \omega))dw(s, \omega) = a\int_0^t f(s, \omega)dw(s, \omega) + b\int_0^t g(s, \omega)dw(s, \omega).$$

(2) Additivity. Let $0 < T_1 < T$. Then

$$\int_0^T f(s, \omega)dw(s, \omega) = \int_0^{T_1} f(s, \omega)dw(s, \omega) + \int_{T_1}^T f(s, \omega)dw(s, \omega).$$

(3) If f is a deterministic function, i.e., $f(s, \omega) \equiv f(s)$, then

$$\int_0^t f(s)dw(s, \omega) \sim N\left(0, \int_0^t f^2(s)ds\right).$$

(4) For any $0 \le \tau \le t \le T$,

$$E\left[\int_0^t f(s, \omega)dw(s, \omega)\right] = 0;$$
$$E\left[\int_0^t f(s, \omega)dw(s, \omega) \,\middle|\, \int_0^\tau f(s, \omega)dw(s, \omega) = x\right] = x;$$
$$E\left[\left(\int_0^T f(s, \omega)dw(s, \omega)\right)^2\right] = \int_0^T E[f^2(s, \omega)]ds.$$

(5)

$$E\left[\int_0^T f(s, \omega)dw(s, \omega)\int_0^T g(s, \omega)dw(s, \omega)\right] = \int_0^T E[f(s, \omega)g(s, \omega)]ds.$$

2.3. **Construction of the Ito integral.** First note that for any constant random function $f(\omega)$

$$\int_a^b f(\omega)dw(s) = f(\omega)(w(b) - w(a)).$$

Therefore, it is easy to construct the Ito integral for any simple random function $h(t, \omega)$ that assumes a finite number of values. The integrals for simple functions can be extended to integrals for any functions satisfying conditions 1 and 2.

Theorem 2. *For every function $f(t, \omega)$ satisfying Conditions 1 and 2, there is a sequence of step functions $f_n(t, \omega)$ satisfying Conditions 1 and 2 such that*

$$\lim_{n\to\infty}\int_0^T |f(t, \omega) - f_n(t, \omega)|^2 dt = 0 \tag{12}$$

for almost all $\omega \in \Omega$, and the limit

$$I(t, \omega) := \lim_{n\to\infty}\int_0^t f_n(s, \omega)dw(s, \omega) \tag{13}$$

is uniform in T for almost all $\omega \in \Omega$ and is independent of the sequence $f_n(t, \omega)$ satisfying conditions 1 and 2.

2.4. **Existence and uniqueness of solutions of the Ito SDEs.** Consider the SDE

$$dx(t) = b(x(t),t)dt + \sigma(x(t),t)dw, \quad x(0) = x_0 \in \mathbb{R}^d, \quad t \in [0,T], \tag{14}$$

where w is the standard Brownian motion. We assume that the functions b and σ satisfy the following conditions. There exists a constant C such that

$$\|b(x,t)\| + \|\sigma(x,t)\|_F \le C(1+\|x\|), \quad \text{for all} \quad x \in \mathbb{R}^d, \quad t \in [0,T], \tag{15}$$

where $\|\cdot\|_F$ denotes the Frobenius matrix norm:

$$\|A\|_F := \sqrt{\mathrm{tr}(A^\top A)},$$

and

$$\|b(x,t) - b(y,t)\| + \|\sigma(x,t) - \sigma(y,t)\|_F \le C\|x-y\|, \quad \text{for all} \quad x,y \in \mathbb{R}^d, \quad t \in [0,T]. \tag{16}$$

The first condition says that b and σ do not grow faster than linearly in x, and the second condition is an analogue of the Lipschitz condition. In this case, Eq. (14) with $E[\|x_0\|^2] < \infty$ has a unique solution such that

$$E\left[\int_0^t \|x(s)\|^2 ds\right] < \infty \quad \text{for all} \quad t \in [0,T].$$

From now on, we will autonomize SDEs to save some writing. If b and/or σ explicitly depend on t, we introduce a new independent variable s, declare that t is a new dependent variable, and add the equation $dt = ds$. This is called "autonomization". Therefore, the assumption that the SDE is autonomous does not lead to the loss of generality.

2.5. **Notation common in probability books.** It is common in the probability community to denote stochastic processes by capital letters with subscripts specifying their time arguments:

$$dX_t = b(X_t)dt + \sigma(X_t)dw, \quad X_0 = x, \quad t \in [0,T], \ x \in \mathbb{R}^d. \tag{17}$$

Terminology:

- X_t satisfying Eq. (17) is called a *diffusion process*;
- $b(x)$ is called a *drift*;
- the matrix $\Sigma(x) = \sigma(x)\sigma^\top(x)$ is called a *diffusion matrix*.

Exercise Show that b and Σ satisfy:

$$\lim_{t\to s} E\left[\frac{X_t - X_s}{t-s} \mid X_s = x\right] = b(x,s) \tag{18}$$

$$\lim_{t\to s} E\left[\frac{[X_t - X_s][X_t - X_s]^\top}{t-s} \mid X_s = x\right] = \Sigma(x,s). \tag{19}$$

2.6. **The Ornstein-Uhlenbeck process.** The Ornstein-Uhlenbeck process models the velocity of a heavy particle pushed around by light particles. The variable X_t is the velocity of the particle. For simplicity, we consider it in 1D:

$$dX_t = -\gamma X_t dt + \sqrt{2D}dw, \quad X_0 = x \in R \tag{20}$$

where x is a fixed number, γ is the friction coefficient, and D is the diffusion coefficient.

The exact solution of Eq. (20) can be written in the closed form involving a stochastic integral. We proceed as we do when we solve a linear non-homogeneous first order ODE. Switch the term $-\gamma X_t$ to the left-hand side and multiply the equation by the integrating factor $\exp(\gamma t)$. Then we get

$$d\left(e^{\gamma t}X_t\right) = \sqrt{2D}e^{\gamma t}dw.$$

Suppose $X_0 = x$. Integrating from 0 to t we obtain:

$$e^{\gamma t}X_t - x = \sqrt{2D}\int_0^t e^{\gamma s}dw_s.$$

Hence the solution of Eq. (20) is

$$X_t = xe^{-\gamma t} + \sqrt{2D}\int_0^t e^{-\gamma(t-s)}dw_s.$$

The solution X_t is a Gaussian random variable with mean $xe^{-\gamma t}$ and variance $\frac{D}{\gamma}(1-e^{-2\gamma t})$. The variance is found as follows. We partition the interval $[0,t]$ into n equal subintervals of length $h = t/n$ and let $n \to \infty$:

$$\begin{aligned}
&\text{Var}\left(\sqrt{2D}\int_0^t e^{-\gamma(t-s)}dw_s\right) = 2De^{-2\gamma t}E\left[(\int_0^t e^{\gamma s}dw_s)^2\right] \\
&= 2De^{-2\gamma t}\lim_{n\to\infty} E\left[\sum_{j=0}^{n-1} e^{\gamma 2jh}[w((j+1)h) - w(jh)]^2\right] \\
&= 2De^{-2\gamma t}\int_0^t e^{2\gamma s}ds = 2De^{-2\gamma t}\frac{1}{2\gamma}\left(e^{2\gamma t} - 1\right) \\
&= \frac{D}{\gamma}\left(1 - e^{-2\gamma t}\right).
\end{aligned}$$

As $t \to \infty$, the velocity X_t of the particle becomes a Gaussian random variable with mean 0 and variance $\frac{D}{\gamma}$.

Now let us calculate the covariance function of the Ornstein-Uhlenbeck process. Suppose $t \geq s$.

$$\begin{aligned}
\text{Cov}(X_s, X_t) &= E\left[\left(\sqrt{2D}\int_0^s e^{-\gamma(s-\tau)}dw_\tau\right)\left(\sqrt{2D}\int_0^t e^{-\gamma(t-\tau)}dw_\tau\right)\right] \\
&= 2De^{-\gamma(t+s)}E\left[\left(\int_0^s e^{\gamma\tau}dw_\tau\right)\left(\int_0^s e^{\gamma\tau}dw_\tau + \int_s^t e^{\gamma\tau}dw_\tau\right)\right] \\
&= \frac{D}{\gamma}e^{-\gamma(t+s)}\left(e^{2\gamma s} - 1\right) \\
&= \left[-\frac{D}{\gamma}\right]e^{-\gamma(t+s)} + \frac{D}{\gamma}e^{-\gamma(t-s)}.
\end{aligned}$$

Note that as t and s tend to infinity, the first term decays to zero. Then the covariance function depends only on the difference $t - s$ and is given by

$$R(t-s) = \frac{D}{\gamma}e^{-\gamma|t-s|}.$$

The Ornstein-Uhlenbeck process (20) with the initial condition $x \sim N(0, \frac{D}{\gamma})$ can serve as a model for a colored noise.

3. The Ito calculus

SDEs can be solved analytically only in special cases. Even if so, we often need not the formula for the solution but the expected value of some function defined on the random trajectories. This function can be the first passage time to a given region of the phase space or the probability to reach first one given region rather than the other given region. To answer such kind of questions, it is handy to be able to calculate the time evolution of functions defined on trajectories. The Ito formula provides us with a tool to do it.

3.1. A derivation of the Ito formula.

The most important result in the Ito calculus is the Ito formula.

3.1.1. *1D case.* For simplicity, we will start with the 1D case. Let $x(t)$ be trajectory obeying the ODE

$$\frac{dx}{dt} = b(x).$$

Then any function $f(x(t), t)$ evolves in time according to

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\frac{dx}{dt} = \frac{\partial f}{\partial t} + b(x)\frac{\partial f}{\partial x}. \tag{21}$$

Now let $x(t, \omega)$ be a trajectory of the SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dw.$$

One could naively write

$$df = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dX_t = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}(b(X_t)dt + \sigma(X_t)dw)$$

but this would be **WRONG.** This is because $dw = O(\sqrt{dt})$ and $(dw)^2 = O(dt)$. The correct differential of f is given by Eq. (22) below.

Let us derive it. We want to find the differential, i.e., the part of the increment of the order of dt or larger for a function $f(X_t, t)$, where $dX_t = b(X_t)dt + \sigma(X_t)dw$. We will write a formal Taylor expansion of $f(X_t, t)$ and keep all terms of the order of dt or larger. To save some space, the arguments (X_t, t) in all derivatives will be omitted.

$$f(X_t+dX_t, t+dt) = f(X_t, t)+\frac{\partial f}{\partial t}dt+\frac{\partial f}{\partial x}dX_t+\frac{1}{2}\frac{\partial^2 f}{\partial t^2}(dt)^2+\frac{\partial^2 f}{\partial x\partial t}dX_t dt+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(dX_t)^2+\dots.$$

The term dw is $O(\sqrt{dt})$. The term $dX_t dt$ is $O((dt)^{3/2})$. The term $(dw)^2$ contained in $(dX_t)^2$ is $O(dt)$. Hence we need to keep only the term $\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(dX_t)^2$ out of the second order terms. Therefore,

$$df(X_t, t) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(dX_t)^2.$$

Writing dX_t explicitly we get:

$$\begin{aligned}(dX_t)^2 &= (b(X_t)dt + \sigma(X_t)dw)(b(X_t)dt + \sigma(X_t)dw) \\ &= b^2(X_t)(dt)^2 + 2b(X_t)\sigma(X_t)dtdw + \sigma^2(dw)^2 = \sigma^2(X_t)dt + o(dt).\end{aligned}$$

Hence,

(22) $$\boxed{df(X_t,t) = \left[\frac{\partial f}{\partial t} + b(X_t)\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2(X_t)\frac{\partial^2 f}{\partial x^2}\right]dt + \sigma(X_t)\frac{\partial f}{\partial x}dw.}$$

Note that Eq. (22) can be rewritten as

(23) $$df(X_t,t) = \left[\frac{\partial f}{\partial t} + Lf\right]dt + \sigma(x)\frac{\partial f}{\partial x}dw, \quad \text{where} \quad L = b(X_t)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(X_t)\frac{\partial^2}{\partial x^2}.$$

L is called the infinitesimal generator of the process.

3.1.2. *Multidimensional case.* Now we turn to the multidimensional case is stated below. We define the generator by

(24) $$L = b(X_t)\cdot\nabla + \frac{1}{2}\Sigma(X_t) : \nabla\nabla,$$

where the symbol ":" means

$$A : B := \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}B_{ij} \equiv \mathsf{tr}\left(A^\top B\right)$$

and

$$(\nabla\nabla)_{ij} = \frac{\partial^2}{\partial x_i \partial x_j}.$$

In the coordinate form, L can be rewritten as

(25) $$L = \sum_{j=1}^{d} b_j(X_t)\frac{\partial}{\partial x_j} + \frac{1}{2}\sum_{i,j=1}^{d}\Sigma_{ij}(X_t)\frac{\partial^2}{\partial x_i \partial x_j}.$$

Theorem 3. (Ito's formula) *Assume that b and σ satisfy Eqs.* (15) *and* (16) *and* $E[\|X_0\|^2] < \infty$. *Let $f(x,t)$ be twice continuously differentiable in x in $\mathbb{R}^d$ and continuously differentiable in t on $[0,T]$. Then the process $f(X_t,t)$ satisfies:*

(26)
$$\boxed{f(X_t,t) = f(X_0,0) + \int_0^t \frac{\partial f(X_s,s)}{\partial s}ds + \int_0^t Lf(X_s,s)ds + \int_0^t \nabla f(X_s,s)^\top \sigma(X_s)dw_s.}$$

Eq. (26) is equivalent to

(27) $$\frac{df}{dt} = \frac{\partial f}{\partial t} + \nabla f \cdot \frac{dX_t}{dt} + \frac{1}{2}\frac{dX_t}{dt}\cdot\nabla\nabla f\frac{dX_t}{dt},$$

or

(28) $$df(X_t,t) = \frac{\partial f}{\partial t}dt + \sum_{i=1}^{d}\frac{\partial f}{\partial x_i}dX_i + \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2 f}{\partial x_i \partial x_j}dX_i dX_j,$$

Ito's formula is proven e.g. in [4].

3.2. **The geometric Brownian motion.** The geometric Brownian motion is a stochastic process satisfying the following SDE

$$dX_t = \mu X_t dt + \sigma X_t dw, \tag{29}$$

where w is the standard Brownian motion, μ (the percentage drift) and σ (the percentage volatility) are constants. The SDE (29) is used in mathematical finance to model the stock prices in the Black-Scholes model.

Eq. (29) has an analytic solution that can be found as follows. Introduce the new dependent variable $Y_t = \log X_t$. Using Ito's formula (22) and taking into account that $\frac{dY}{dX} = \frac{1}{X}$ and $\frac{d^2Y}{dX^2} = -\frac{1}{X^2}$ we write the differential of Y_t:

$$dY_t = \left(\mu X_t \frac{dY}{dX} + \frac{1}{2}\sigma^2 X_t^2 \frac{d^2Y}{dX^2}\right) dt + \sigma X_t \frac{dY}{dX} dw = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dw.$$

The right-hand side of the SDE for Y_t is independent of Y_t and hence Y_t is found just by integration of the right-hand side:

$$Y_t = Y_0 + \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma w_t.$$

Returning to $X_t = \exp(Y_t)$ we get the exact solution of Eq. (29):

$$X_t = X_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma w_t}. \tag{30}$$

We say that a random variable X is *lognormal* if its logarithm is Gaussian, i.e. if $\log X \sim N(m, s^2)$.

Exercise Check that if $Y \sim N(m, s^2)$ then

$$E[e^Y] = e^{m + \frac{s^2}{2}}, \quad \text{Var}(e^Y) = e^{2m+s^2}(e^{s^2} - 1),$$

and the pdf of $X = e^Y$ is given by

$$f_{e^Y}(x) = \frac{1}{x\sqrt{2\pi s^2}} e^{-\frac{(\log x - m)^2}{2\sigma^2}}.$$

Note that

$$Y_t = \log X_t \sim N\left(Y_0 + \left(\mu - \frac{\sigma^2}{2}\right) t, \sigma^2 t\right),$$

hence the geometric Brownian motion X_t has a log-normal distribution. The mean and the variance of X_t are

$$E[X_t] = X_0 e^{\mu t}, \quad \text{Var}(X_t) = X_0^2 e^{2\mu t}\left(e^{\sigma^2 t} - 1\right).$$

3.3. **Backward Kolmogorov equation.** Imagine that we are interested in some quantity f that depends on X_t evolving according to the SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dw.$$

Suppose that we want to find the expected value of f at a future time T given that at the present time t, $X_t = x$. For example, you can think of f being an option price that depends on the stock price X. We denote the expected value of f at time T conditioned on $X_t = x$ by $u(x,t)$. Let us find the time evolution of

$$u(x,t) := E[f(X_T) \mid X_t = x] = \int_{\mathbb{R}^d} f(y)p(y,T|x,t)dy,$$

In words, imagine that we start a stochastic process X_s at time t at the point x. We stop it at a fixed time T. We want to find the expected value of $f(X_T)$. This expected value $u(x,t)$ depends on the initial time t and the initial point x. Obviously, for the terminal time $t = T$ we have: $u(x,T) = f(x)$. Using Ito's formula (26) we calculate

$$u(X_T,T) - u(X_t,t) = \int_t^T \frac{\partial}{\partial s}u(X_s,s)ds + \int_t^T Lu(X_s,s)ds + \int_t^T \nabla u(X_s,s)^\top \sigma(X_s)dw_s.$$

Now we will take expected values of both parts of this equation conditioned on $X_t = x$. Note that

$$E[u(X_T,T) \mid X_t = x] = E\left[f(X_T) \mid X_t = x\right] = u(x,t)$$

and

$$E[u(X_t,t) \mid X_t = x] = u(x,t).$$

Hence the conditional expectation of the left-hand side is 0. Also note that by property (4) in Section 2.2,

$$E\left[\int_t^T \nabla u(X_s,s)^\top \sigma(X_s)dw_s\right] = 0.$$

Therefore, for all x, t and T we have

$$\int_t^T \left[\frac{\partial}{\partial s}u(x,s) + Lu(x,s)\right] ds = 0.$$

Hence for all $t \le T$, $u(x,t)$ satisfies the PDE with the final condition:

$$\frac{\partial}{\partial t}u(x,t) + Lu(x,t) = 0 \quad u(x,T) = f(x).$$

Re-defining u as

$$u(x,t) = E[f(X_t) \mid X_0 = x], \quad \text{i.e.,} \quad u(x,t) = u_{\text{old}}(x,T-t),$$

we obtain that for $0 \le t \le T$

$$\boxed{\frac{\partial}{\partial t}u(x,t) = Lu(x,t), \quad u(x,0) = f(x).} \tag{31}$$

Eq. (31) is called the *Backward Kolmogorov Equation*. It describes the time evolution of expected values. Note that in the re-definition of u we used the fact that X_t evolves according to an autonomous SDE which is invariant with respect to a time shift.

3.4. **The expected first passage time.** Let $A \subset \mathbb{R}^d$ be some region. The first passage time to A is defined as

$$\tau_A = \inf\{t \geq 0 \mid X_t \in A\}.$$

Let $u(x,t)$ be the expected first passage time to A for the process X_t starting at x, i.e.,

$$u(x,t) = E\left[\tau_A \mid X_t = x\right].$$

We calculate:

$$u(X_{\tau_A}, \tau_A) - u(X_t, t) = \int_t^{\tau_A} \frac{\partial}{\partial s} u(X_s, s) ds + \int_t^{\tau_A} Lu(X_s, s) ds + \int_t^{\tau_A} \nabla u(X_s, s)^\top \sigma(X_s) dw_s.$$

Next, we take the expected values of the left- and right-hand side of this equation conditioned on $X_t = x$. Taking into account that

$$E[u(X_{\tau_A}, \tau_A) \mid X_t = x] = E\left[\tau_A \mid X_t = x\right] = u(x,t) \quad \text{and} \quad E[u(X_t, t) \mid X_t = x] = u(x,t)$$

we get for all t and x

$$\int_t^{\tau_A} \frac{\partial}{\partial s} u(x,s) ds + \int_t^{\tau_A} Lu(x,s) ds = 0.$$

We also note that

$$\frac{\partial}{\partial t} u(x,t) = 1.$$

Finally, if $x \in A$ we have $\tau_A = 0$. Hence the mean first passage time $u(x)$ satisfies the following boundary value problem

$$\boxed{Lu = -1, \quad x \notin A, \quad x(\partial A) = 0.} \tag{32}$$

Example 2 Let $dX_t = \sqrt{2\beta^{-1}} dw$, $X_0 = x$, i.e., a scaled one-dimensional Brownian motion starting at the point x. Let us find the expected exit time from the interval $[-1, 1]$. As we have shown, the expected exit time $u(x)$ satisfies Eq. (32), which in our case becomes

$$\beta^{-1} u'' = -1, \quad -1 \leq x \leq 1, \quad u(1) = u(-1) = 0.$$

Solving this equation we obtain

$$u(x) = \frac{\beta}{2}(1 - x^2).$$

3.5. **The committor equation.** Let $A \subset \mathbb{R}^d$ and $B \subset \mathbb{R}^d$ be some regions. The committor function $q(x)$ is defined as the probability that the process starting at the point x first reaches B rather than A [6, 7]. Let us derive a boundary-value problem for the commitor. It is clear that $q(\partial A) = 0$ and $q(\partial B) = 1$. For $x \in (A \cup B)^c$ let us define the first passage time to $A \cup B$, i.e.,

$$\tau_{AB} = \inf\{t \geq 0 \mid X_t \in A \cup B\}.$$

We calculate:

$$q(X_{\tau_{AB}}) - q(X_0) = \int_0^{\tau_{AB}} Lq(X_s)ds + \int_0^{\tau_{AB}} \nabla q(X_s)^\top \sigma(X_s)dw_s.$$

Take the expected values of the left- and right-hand side of this equation conditioned on $X_0 = x$. We get that for all x

$$q(x) - q(x) = \int_0^{\tau_{AB}} Lq(x)ds = 0$$

Therefore, the solution of the boundary-value problem

$$Lq = 0, \quad x \in (A \cup B)^c, \quad q(\partial A) = 0, \quad q(\partial B) = 1$$

is the committor function.

Example 3 Let $dX_t = -V'(x)dt + \sqrt{2\beta^{-1}}dw$, $X_0 = x$, i.e., a particle moving according to the overdamped Langevin dynamics in the potential force field $V(x)$. In the 1D case, the committor equation can be solved exactly. We have:

$$-V'(x)q'(x) + \beta^{-1}q''(x) = 0, \quad a \leq x \leq b, \quad q(a) = 0, \quad q(b) = 1.$$

Multiply this equation by $\beta e^{-\beta V(x)}$. Then its left-hand side becomes a complete differential:

$$\left(e^{-\beta V(x)}q'(x)\right)' = 0.$$

Integrating this equation and taking the boundary conditions into account we get

$$q(x) = \frac{\int_a^x e^{\beta V(y)}dy}{\int_a^b e^{\beta V(y)}dy}.$$

3.6. **The generator of a Markov process.** In Section 3.3, we fixed a function f and we considered the expectation of f at time t as a function of the initial point x and time t:

$$u(x,t) := E[f(X_t) \mid X_0 = x].$$

Now we fix the time t and consider the same expectation as a map applied to the set of continuous and bounded functions $f(x)$, $x \in \mathbb{R}^d$. Therefore, we define the family of operators indexed by t:

$$(P_t f)(x) := E[f(X_t) \mid X_0 = x] = \int_{\mathbb{R}^d} f(y)p(y,t \mid x,0)dy.$$

We will call the operator P_t the *transfer operator*. It is analogous to the stochastic matrix P in the discrete-time Markov chains.

The operator P_t possesses the semigroup properties:

$$P_0 = I, \quad P_{t+s} = P_t \circ P_s \quad \text{for all} \quad t, s \geq 0.$$

Indeed,

$$(P_0 f)(x) := E[f(X_0) \mid X_0 = x] = f(x),$$

Hence P_0 is the identity. Recall the Chapman-Kolmogorov equation (Eq. (6)) expressing the Markov property. Using it, we write:

$$\begin{aligned}(P_{t+s}f)(x) &= \int_{\mathbb{R}^d} f(y)p(y, t+s \mid x, 0)dy = \int_{\mathbb{R}^d} f(y)dy \int_{\mathbb{R}^d} p(y, t+s \mid z, t)p(z, t \mid x, 0)dz \\ &= \int_{\mathbb{R}^d} dz p(z, t \mid x, 0) \int_{\mathbb{R}^d} p(y, t+s \mid z, t)f(y)dy = (P_t P_s f)(x).\end{aligned}$$

Due to this, the operator P_t is often referred to as the *Markov semigroup*.

Now consider the limit as $t \downarrow 0$:

$$(Lf)(x) := \lim_{t \to 0+} \frac{(P_t f)(x) - f(x)}{t}. \tag{33}$$

Assume that this limit exists. This limit is called the *infinitesimal generator* of the transfer operator P_t or the *generator* of the Markov process X_t.

Recall that $(P_t f)(x) = u(x, t)$. Eq. (33) implies that

$$\frac{\partial u}{\partial t} = \lim_{s \to 0} \frac{P_{t+s}f - P_t f}{s} = LP_t f = Lu, \quad u(x, 0) = f(x).$$

This is the Backward Kolmogorov equation that we have obtained in Section 3.3. Matching it with Eq. (31) we see that the operator L must be given by Eq. (24).

3.7. **The adjoint semigroup and the forward Kolmogorov equation.** In Section 3.6 we considered the expected value of $f(X_t)$ conditioned on $X_0 = x$. Now we assume that X_0 does not start at x with probability 1 but the starting point is distributed according to a pdf $\mu_0(x)$. Then the expected value of $f(X_t)$ is

$$E[f(X_t)] = \int_{\mathbb{R}^d} P_t f(x)\mu_0(x)dx.$$

Writing $P_t f$ explicitly and switching the order of integration we obtain

$$\begin{aligned}E[f(X_t)] &= \int_{\mathbb{R}^d} P_t f(x)\mu_0(x)dx = \int_{\mathbb{R}^d} \mu_0(x)dx \int_{\mathbb{R}^d} dy f(y)p(y, t \mid x, 0) \\ &= \int_{\mathbb{R}^d} dy f(y) \int_{\mathbb{R}^d} \mu_0(x)p(y, t \mid x, 0)dx \\ &=: \int_{\mathbb{R}^d} dy f(y) P_t^* \mu_0. \qquad (34)\end{aligned}$$

In the original order of integration, we froze the pdf $\mu_0(x)$ while evolved $f(X_t)$ in time. After switching the order of integration, we froze f and evolved the pdf μ in time. Finally, we have defined the evolution operator for the pdf:

$$\mu_t(x) := (P_t^* \mu_0)(x) := \int_{\mathbb{R}^d} \mu_0(x)p(y, t \mid x, 0)dx. \tag{35}$$

The operator P_t^* is adjoint to the transfer operator P_t. Indeed, consider the inner product

$$(f,g) := \int_{\mathbb{R}^d} f(x)g(x)dx.$$

Then Eq. (34) shows that

$$E[f(X_t)] = (P_t f, \mu_0) = (f, P_t^* \mu_0).$$

The infinitesimal generator for the adjoint semigroup P_t^* is defined by

$$(L^*\mu_0)(x) := \lim_{t\to 0} \frac{(P_t^*\mu_0)(x) - \mu_0(x)}{t}. \tag{36}$$

It is easy to check that L and L^* are adjoint, i.e., for all admissible f and g,

$$(Lf,g) = (f, L^* g). \tag{37}$$

Eqs. (35) and (36) show that the time evolution of the probability density function is given by

$$\frac{\partial \mu(x,t)}{\partial t} = L^*\mu(x,t), \quad \mu(x,0) = \mu_0(x). \tag{38}$$

Eq. (38) is called the *forward Kolmogorov equation* or the *Fokker-Planck equation*.

Eq. (24) allows us to find the adjoint generator L^* explicitly. Consider the process governed by the SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dw, \quad X(0) = x, \quad t \ge 0,\ x \in \mathbb{R}^d.$$

For it, the generator is the differential operator given by

$$L = b(x)\cdot\nabla + \frac{1}{2}\Sigma(x) : \nabla\nabla.$$

To find the adjoint generator L^*, we consider the identity Eq. (37):

$$\begin{aligned}(Lf,g) &= \int_{\mathbb{R}^d} (b\cdot\nabla f + \frac{1}{2}\Sigma : \nabla\nabla f) g dx \\ &= \int_{\mathbb{R}^d} f\left(-\nabla\cdot(gb) + \frac{1}{2}\nabla\nabla : (\Sigma g)\right) dx = (f, L^* g).\end{aligned}$$

Here we have integrated by parts the first term once and the second term twice. Hence,

$$L^* g = -\nabla\cdot(gb) + \frac{1}{2}\nabla\nabla : (\Sigma g) \tag{39}$$

Example 4 We will elaborate the procedure of obtaining L^* in 2D. The extension to higher dimensions in straightforward. Consider a 2D stochastic process of the form

$$\begin{aligned}dX_t &= b_1(X_t,Y_t)dt + \sigma_{11}(X_t,Y_t)dw_1 + \sigma_{12}(X_t,Y_t)dw_2 \\ dY_t &= b_2(X_t,Y_t)dt + \sigma_{21}(X_t,Y_t)dw_1 + \sigma_{22}(X_t,Y_t)dw_2\end{aligned}$$

In the vector notations it looks as

$$d\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix}.$$

The matrix $\Sigma = \sigma\sigma^\top$ is

$$\Sigma \equiv \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} \equiv \begin{pmatrix} \sigma_{11}^2 + \sigma_{12}^2 & \sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22} \\ \sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22} & \sigma_{21}^2 + \sigma_{22}^2 \end{pmatrix}.$$

The generator L applied to a function f is:

$$Lf = b_1\partial_x f + b_2\partial_y f + \frac{1}{2}\left[\Sigma_{11}\partial_{xx} f + 2\Sigma_{12}\partial_{xy} f + \Sigma_{22}\partial_{yy} f\right].$$

The adjoint generator L^* is found from the identity $(Lf, g) = (f, L^*g)$:

$$\begin{aligned}(Lf,g) &= \int_{\mathbb{R}^2}\left[gb_1\partial_x f + gb_2\partial_y f + \frac{1}{2}g\left[\Sigma_{11}\partial_{xx} f + 2\Sigma_{12}\partial_{xy} f + \Sigma_{22}\partial_{yy} f\right]\right] dxdy \\ &= \int_{\mathbb{R}^2} f\left[-\partial_x(b_1 g) - \partial_y(b_2 g) + \frac{1}{2}\left[\partial_{xx}(\Sigma_{11} g) + 2\partial_{xy}(\Sigma_{12} g) + \partial_{yy}(\Sigma_{22} g)\right]\right] dxdy \\ &= (f, L^* g).\end{aligned}$$

Hence

$$L^* g = -\partial_x(b_1 g) - \partial_y(b_2 g) + \frac{1}{2}\left[\partial_{xx}(\Sigma_{11} g) + 2\partial_{xy}(\Sigma_{12} g) + \partial_{yy}(\Sigma_{22} g)\right].$$

3.8. **The invariant pdf.** Recall irreducible continuous-time Markov chains. The invariant measure is the solution of $\pi L = 0$ or, equivalently, $L^\top \pi^\top = 0$. π and $\pi^\top$are row and column vectors respectively. If the invariant measure is normalizable, we normalize it so that $\sum_i \pi_i = 1$ and call it the invariant distribution. For irreducible Markov chains with a finite number of states, any probability distribution converges over time to the unique invariant distribution π, i.e., for any initial distribution p_0, the solution of

$$\frac{dp}{dt} = pL, \quad p(0) = p_0$$

converges to π. Such Markov chains are called *ergodic*. The Markov Chain Monte Carlo methods employ this property. Also recall that irreducibility of a continuous-time Markov chain with a finite number of states implies that the eigenvalue 0 of L has multiplicity one.

Suppose that the equation $L^* f = 0$ for a Markov process X_t has a unique positive solution μ up to a multiplicative constant, and this solution is normalizable so that $\int \mu(x)dx = 1$, then μ is the unique invariant pdf. In this case, for any initial pdf μ_0, the pdf μ_t converges to μ as $t \to \infty$:

$$\lim_{t\to\infty} P_t^* \mu_0 = \mu.$$

Such Markov processes are also called *ergodic*.

The unique invariant pdf $\mu(x)$ satisfies the stationary forward Kolmogorov equation (stationary Fokker-Planck equation):

$$L^*\mu = 0, \quad \int_{\mathbb{R}^d} \mu(x)dx = 1. \tag{40}$$

Exercise (1) Show that the generator of the 1D Ornstein-Uhlenbeck process (20) is given by

$$L = -\gamma x \frac{d}{dx} + D\frac{d^2}{dx^2}.$$

(2) Integrating by parts, derive the expression for the adjoint generator

$$L^*g = \frac{d}{dx}(\gamma x g) + D\frac{d^2 g}{dx^2}.$$

(3) Solve the equation

$$L^*\mu = 0, \qquad \int_{-\infty}^{\infty} \mu dx = 1$$

and find that the invariant pdf for the 1D Ornstein-Uhlenbeck process is

$$\mu(x) = \sqrt{\frac{\gamma}{2\pi D}} e^{-\frac{\gamma x^2}{2D}}. \tag{41}$$

4. The Langevin Dynamics

4.1. The full Langevin dynamics.

The Langevin equation models the dynamics of heavy particles in the potential force field pushed around by light particles:

$$\begin{aligned} dq &= \frac{p}{m}dt \\ dp &= (-\nabla V(q) - \gamma p)dt + \sqrt{2\gamma m\beta^{-1}}dw, \end{aligned} \tag{42}$$

where $(q,p) \in \mathbb{R}^{2d}$ are the positions and momenta of the heavy particles, γ is the friction coefficient, m is the mass of the heavy particles, and $-\nabla V(q)$ is the potential force acting on the heavy particles. Eq. (42) can be written in the form (17) by introducing

$$X_t = \begin{bmatrix} q \\ p \end{bmatrix}, \quad b(X_t) = \begin{bmatrix} p/m \\ -\nabla V(q) - \gamma p \end{bmatrix}, \quad \sigma = \sqrt{2\gamma m\beta^{-1}} \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix},$$

where I is the $d \times d$ identity matrix.

Exercise (1) Show that the infinitesimal generator for Eq. (42) is given by

$$L = \frac{p}{m}\cdot\nabla_q - \nabla_q V \cdot \nabla_p + \gamma\left(-p\nabla_p + m\beta^{-1}\Delta_p\right).$$

(2) Derive the expression for the adjoint generator

$$L^*g = -\frac{p}{m}\cdot\nabla_q g + \nabla_q V\cdot\nabla_p g + \gamma\left(\nabla_p\cdot(pg) + m\beta^{-1}\Delta_p g\right).$$

(3) Solve the stationary Fokker-Planck equation and show that the invariant pdf is given by

$$\mu(q,p) = \frac{1}{Z}e^{-\beta H(q,p)}, \quad \text{where} \quad H(q,p) = \frac{|p|^2}{2m} + V(q).$$

4.2. **The overdamped Langevin dynamics.** Suppose the friction coefficient γ in Eq. (42) is large and/or the mass m is small, i.e., $m\gamma^{-1}$ is small. We divide the equation for p by γ:

$$\gamma^{-1}dp = \gamma^{-1}mdv = (-\gamma^{-1}\nabla V(q) - p)dt + \gamma^{-1}\sqrt{2m\gamma\beta^{-1}}dw.$$

We use that assumption that $m\gamma^{-1}$ is small and set the left-hand side of the SDE above to 0. Then we replace p with $m\frac{dq}{dt}$ in the right-hand side and multiply both sides by γ:

$$m\gamma\frac{dq}{dt} = -\nabla V(q) + \sqrt{2m\gamma\beta^{-1}}\frac{dw}{dt}$$

or

$$m\gamma dq = -\nabla V(q)dt + m\gamma\sqrt{\frac{2\beta^{-1}}{m\gamma}}dw(t)$$

Now we want to cancel out $m\gamma$. To do so, we rescale the time by introducing $\tau = (m\gamma)^{-1}t$. Then $dt = m\gamma d\tau$. Recall that if $w(t)$ is a Brownian motion, then for any $c > 0$, $c^{-1/2}w(ct)$ is also a brownian motion, i.e., $c^{1/2}w(t) = w(ct)$. Hence

$$dw(t) = dw(m\gamma\tau) = (m\gamma)^{1/2}dw(\tau).$$

Therefore, choosing the new time τ and canceling $m\gamma$ we get

$$dq = -\nabla V(q)d\tau + \sqrt{2\beta^{-1}}dw(\tau).$$

For the overdamped Langevin dynamics, the generator L is given by

$$L = -\nabla V\cdot\nabla + \beta^{-1}\Delta.$$

The adjoint generator L^* is

$$L^*g = -\nabla\cdot(g\nabla V) + \beta^{-1}\Delta g.$$

The invariant pdf is

$$\mu = \frac{1}{Z}e^{-\beta V(x)}.$$

Appendix A. Construction of the Brownian motion

A.1. **Construction of the Brownian motion via a refinement procedure.** The Brownian motion on $0 \le t \le 1$ can be constructed by a recursive refinement procedure [3, 2]. Consider the following collection of sets

$$\mathcal{D}_n = \left\{\frac{k}{2^n} \mid 0 \le k \le 2^n\right\}$$

of dyadic points. At each refinement step, we will define a Gaussian random walk satisfying the axioms of the Brownian Motion at the dyadic points (see Fig. 1).

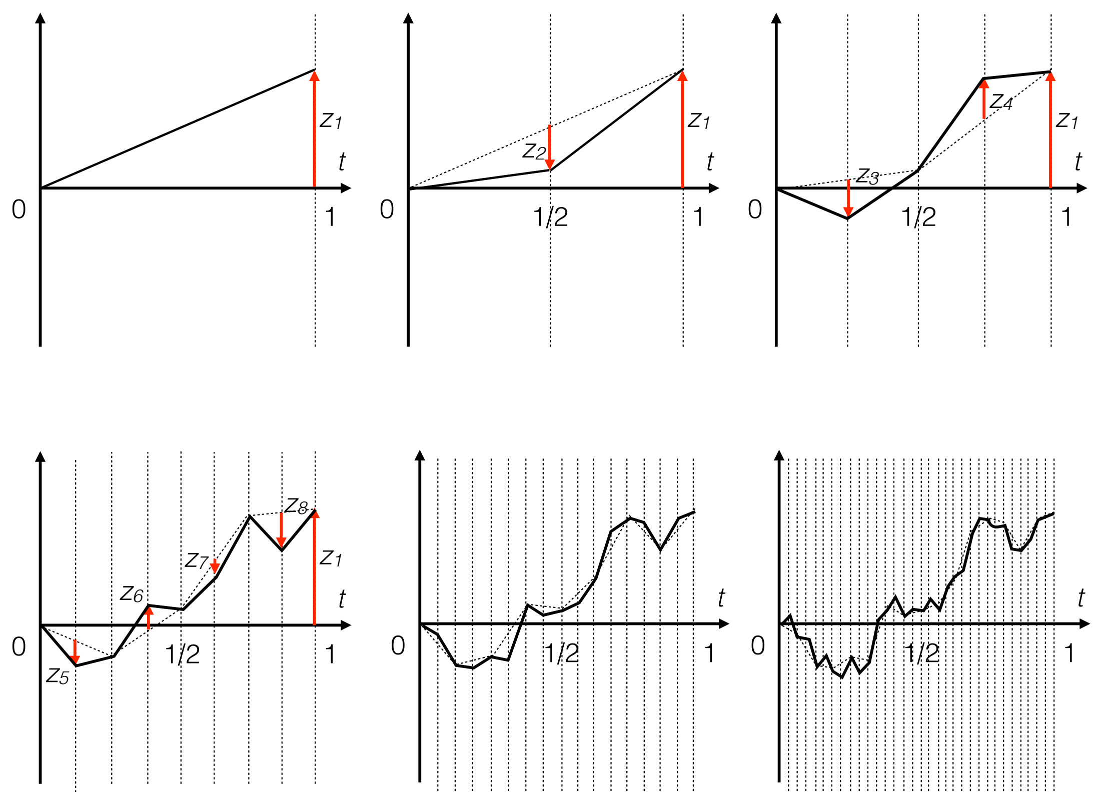

FIGURE 1. First six steps of construction a Brownian motion from Brownian random walks by a refinement procedure. Here $z_1, \ldots, z_8$ are independent Gaussian random variables such that $z_1 \in N(0,1)$, $z_2 \in N(0,1/4)$, $z_3, z_4 \in N(0, \frac{1}{8})$, $z_5, z_6, z_7, z_8 \in N(0, \frac{1}{16})$.

Let $\{z_0, z_{n,j}\}$ where $n = 1, 2, 3, \ldots$, $j = 1, 2, \ldots, 2^{n-1}$, be a collection of independent Gaussian random variables with mean 0 and variance 1. Recall that

$$\text{if} \quad z \in N(0,1) \quad \text{then} \quad \frac{z}{2^{\alpha}} \in N(0, 2^{-2\alpha}).$$

We start by constructing a Gaussian random walk on $\mathcal{D}_0$ by setting

$$B_0(0) = 0 \quad \text{and} \quad B_0(1) = z_0.$$

Then we refine it to a Gaussian random walk B_1 on D_1 by setting

$$B_1(\mathcal{D}_0) = B_0(\mathcal{D}_0), \quad B_1\left({}^{1}\!/\!{}_{2}\right) = \frac{B_0(1) + B_0(0)}{2} + \frac{z_{1,1}}{2}.$$

Note that

$$B_1\left(1/2\right) = \frac{z_0}{2} + \frac{z_{1,1}}{2}.$$

Hence

$$\mathrm{Var}\left(B_1\left(1/2\right)\right) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

as desired. Let us show that the increments of B_1 restricted to the dyadic set $\mathcal{D}_1 = \{0, 1/2, 1\}$ are independent, i.e., that $B_1(1/2) - B_1(0)$ and $B(1) - B(1/2)$ are independent. Indeed, a linear combination of independent Gaussian random variables $x_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ and $x_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ is Gaussian: $ax_1 + bx_2 \sim \mathcal{N}(am_1 + bm_2, a^2\sigma_1^2 + b^2\sigma_2^2)$ (check it!). Then we calculate:

$$\begin{aligned}
B_1\left(1/2\right) - B_1(0) &= \frac{z_0}{2} + \frac{z_{1,1}}{2} \sim \mathcal{N}\left(0, 1/2\right), \\
B_1(1) - B_1\left(1/2\right) &= z_0 - \frac{z_0}{2} - \frac{z_{1,1}}{2} = \frac{z_0}{2} - \frac{z_{1,1}}{2} \sim \mathcal{N}\left(0, 1/2\right), \\
\mathbb{E}\left[\left(\frac{z_0}{2} + \frac{z_{1,1}}{2}\right)\left(\frac{z_0}{2} - \frac{z_{1,1}}{2}\right)\right] &= \mathbb{E}\left[\left(\frac{z_0}{2}\right)^2 - \left(\frac{z_{1,1}}{2}\right)^2\right] = \frac{1}{2} - \frac{1}{2} = 0.
\end{aligned}$$

Since uncorrelated Gaussian random variables are independent, we conclude that $B_1(1/2) - B_1(0)$ and $B(1) - B(1/2)$ are independent.

Next, we refine B_1 to a Gaussian random walk B_2 on $\mathcal{D}_2$ by setting $B_2(\mathcal{D}_1) = B_1(\mathcal{D}_1)$ and

$$\begin{aligned}
B_2\left(1/4\right) &= \frac{1}{2}\left[B_1\left(1/2\right) + B_1(0)\right] + \frac{z_{2,1}}{2\sqrt{2}}, \\
B_2\left(3/4\right) &= \frac{1}{2}\left[B_1(1) + B_1\left(1/2\right)\right] + \frac{z_{2,2}}{2\sqrt{2}}.
\end{aligned}$$

Then,

$$\begin{aligned}
\mathrm{Var}\left(B_2\left(1/4\right)\right) &= \mathrm{Var}\left(1/2\left[B_1\left(1/2\right) + B_1(0)\right] + \frac{z_{2,1}}{2\sqrt{2}}\right) = \frac{1}{4}\cdot\frac{1}{2} + \frac{1}{8} = \frac{1}{4}, \\
\mathrm{Var}\left(B_2\left(3/4\right)\right) &= \mathrm{Var}\left(\frac{1}{2}\left[B_1(1) + B_1\left(1/2\right)\right] + \frac{z_{2,2}}{2\sqrt{2}}\right) \\
&= \mathrm{Var}\left(\frac{1}{2}\left[B_1(1) - B_1\left(1/2\right) + 2B_1\left(1/2\right)\right] + \frac{z_{2,2}}{2\sqrt{2}}\right) \\
&= \frac{1}{4}\left[\frac{1}{2} + 4\frac{1}{2}\right] + \frac{1}{8} = \frac{3}{4}.
\end{aligned}$$

as desired. Show that the increments of B_2 restricted to the dyadic set $\mathcal{D}_2 = \{0, 1/4, 1/2, 3/4, 1\}$ are independent. First show that $B_2(2^{-2}(k+1)) - B_2(2^{-2}k)$ and $B_2(2^{-2}(l+1)) - B_2(2^{-2}l)$

for $0 \le k < l \le 3$ are independent. If k is even, i.e., $k = 2p$, we have:

$$B_2\left(\frac{p}{2}+\frac{1}{4}\right) - B_2\left(\frac{p}{2}\right) = \frac{1}{2}\left[B_1\left(\frac{p}{2}\right) + B_1\left(\frac{p+1}{2}\right)\right] + \frac{z_{2,p+1}}{2\sqrt{2}} - B_2\left(\frac{p}{2}\right) =$$
$$\frac{1}{2}\left[B_1\left(\frac{p+1}{2}\right) - B_1\left(\frac{p}{2}\right)\right] + \frac{z_{2,p+1}}{2\sqrt{2}} \sim \mathcal{N}\left(0, \frac{1}{4}\cdot\frac{1}{2}+\frac{1}{8}\right) = \mathcal{N}\left(0,\frac{1}{4}\right).$$

If k is odd, i.e., $k = 2p+1$, we have:

$$B_2\left(\frac{p+1}{2}\right) - B_2\left(\frac{p}{2}+\frac{1}{4}\right) = B_2\left(\frac{p+1}{2}\right) - \frac{1}{2}\left[B_1\left(\frac{p}{2}\right) + B_1\left(\frac{p+1}{2}\right)\right] - \frac{z_{2,p+1}}{2\sqrt{2}} =$$
$$\frac{1}{2}\left[B_1\left(\frac{p+1}{2}\right) - B_1\left(\frac{p}{2}\right)\right] - \frac{z_{2,p+1}}{2\sqrt{2}} \sim \mathcal{N}\left(0, \frac{1}{4}\cdot\frac{1}{2}+\frac{1}{8}\right) = \mathcal{N}\left(0,\frac{1}{4}\right).$$

Note that both $\frac{1}{2}\left[B_1\left(\frac{p+1}{2}\right) - B_1\left(\frac{p}{2}\right)\right]$ and $\frac{z_{2,p+1}}{2\sqrt{2}}$ are independent Gaussian random variables with mean 0 and variance 1/8. If $k = 2p$ and $l = 2p+1$, using an argument similar to the one used for showing that that $B_1(1/2) - B_1(0)$ and $B(1) - B(1/2)$ are independent, we show that

$$B_2\left(\frac{p}{2}+\frac{1}{4}\right) - B_2\left(\frac{p}{2}\right) \quad \text{and} \quad B_2\left(\frac{p+1}{2}\right) - B_2\left(\frac{p}{2}+\frac{1}{4}\right)$$

are independent. If k and l are such that $floor(k/2) < floor(l/2)$, then the argument above implies that $B_2(2^{-2}(k+1)) - B_2(2^{-2}k) = x_1 + x_2$ and $B_2(2^{-2}(l+1)) - B_2(2^{-2}l) = y_1 + y_2$, where x_1, x_2, y_1, y_2 are independent Gaussian random variables with mean 0 and variance 1/8. Therefore, $B_2(2^{-2}(k+1)) - B_2(2^{-2}k)$ and $B_2(2^{-2}(l+1)) - B_2(2^{-2}l)$ for $0 \le k < l \le 3$ are independent. Finally, the increments over non-overlapping (no common interior points) of B_2 restricted to $\mathcal{D}_2$ are independent as they are sums of mutually independent increments.

Continuing inductively, we define $(i): B_n(\mathcal{D}_{n-1}) = B_{n-1}(\mathcal{D}_{n-1})$; (ii) for $\mathcal{D}_n \backslash \mathcal{D}_{n-1}$

$$B_n\left(\frac{k}{2^{n-1}}+\frac{1}{2^n}\right) = \frac{1}{2}\left(B_{n-1}\left(\frac{k}{2^{n-1}}\right) + B_{n-1}\left(\frac{k+1}{2^{n-1}}\right)\right) + \frac{z_{n,k+1}}{2^{(n+1)/2}},$$

where $z_{n,k+1} \sim \mathcal{N}(0,1)$. It is shown in [3] that if one continues this refinement procedure up to infinity, the resulting process satisfies the definition of Brownian motion. The first six steps of this procedure are illustrated in Fig. 1.

A.2. **Construction of the Brownian motion by a random walk.** This construction is left as an exercise.

Exercise Consider the mesh

$$\{t_j \mid t_j = jh, \quad h = \tfrac{1}{N}, \quad 0 \le j \le N\}.$$

Let $\{z_j\}_{j=1}^N$ be independent Gaussian random variables with mean 0 and variance 1. Consider the Gaussian random walk $B_h(t)$ defined by

$$\begin{aligned} B_h(0) &= 0, \\ B_h(t_j) &= B_h(t_{j-1}) + z_j\sqrt{h}, \quad j = 1, \dots, N, \\ B_h(t) &= \frac{1}{h}\left[B_h(t_{j-1})(t_j - t) + B_h(t_j)(t - t_{j-1})\right], \quad t_{j-1} < t < t_j, \quad j = 1, \dots, N. \end{aligned}$$

Prove that this random walk satisfies axioms (1)-(4) of Brownian motion at the points t_j, $j = 0, 1, \dots, N$.

Exercise A simple construction of the Brownian motion can be done as follows. Consider a random walk on a mesh in the (x, t)-space, $t \geq 0$, $x \in \mathbb{R}$ in which the time-step k and the space step h are related to the time step k via $k = h^2$. Start at the origin at time 0. At any discrete moment of time $t = k, 2k, 3k, \dots$, take a step to left or to the right with probability $1/2$. Let k and h tend to zero in such a manner that the relationship $k = h^2$ is maintained. Apply the Central Limit Theorem and obtain the Brownian motion.

The last exercise will be solved in Section **??** ahead.

Appendix B. An example comparing Ito's and Stratonovich's stochastic integrals

Calculate $\int_a^b w dw$ (the Ito stochastic integral) and $\int_a^b w \circ dw$ (the Stratonovich stochastic integral). Let the partition be uniform, i.e.,

$$\Delta t = \frac{b-a}{n}.$$

We start with the Ito stochastic integral.

$$\begin{aligned} \int_a^b w dw &= \lim_{\Delta \to 0} \sum_i w(t_i)\left[w(t_{i+1}) - w(t_i)\right] \\ &= \frac{1}{2}\lim_{\Delta \to 0} \sum_i \left[w^2(t_{i+1}) - w^2(t_i) - (w(t_{i+1}) - w(t_i))^2\right] \end{aligned}$$

Note that

$$\begin{aligned} \sum_i (w^2(t_{i+1}) - w^2(t_i)) &= w^2(t_n) - w^2(t_{n-1}) + w^2(t_{n-1}) - \ldots + w^2(t_1) - w^2(t_0) \\ &= w^2(b) - w^2(a). \end{aligned}$$

We also compute

$$E\left[\sum_i (w(t_{i+1}) - w(t_i))^2\right] = \sum_i (t_{i+1} - t_i) = b - a.$$

$$\begin{aligned}\text{Var}\left(\sum_i (w(t_{i+1}) - w(t_i))^2\right) &= \sum_i \left[E[(w(t_{i+1}) - w(t_i))^4] - \left(E[(w(t_{i+1}) - w(t_i))^2]\right)^2\right. \\ &\le \sum_i E\left[(w(t_{i+1}) - w(t_i))^4\right] = \frac{n}{\sqrt{2\pi\Delta t}} \int_{-\infty}^{\infty} x^4 e^{-x^2/2\Delta t} dx \\ &= 3n(\Delta t)^2 = \frac{3n(b-a)^2}{n^2} = \frac{3(b-a)^2}{n} \to 0 \quad \text{as} \quad n \to \infty.\end{aligned}$$

Here we have used the fact that the fourth central moment of $\mathcal{N}(\mu, \sigma^2)$ is $3\sigma^4$. Hence

$$\int_a^b w dw = \frac{w^2(b) - w^2(a)}{2} - \frac{b-a}{2}.$$

The expected value of this integral is zero. Indeed, $w(t_i)\left[w(t_{i+1}) - w(t_i)\right]$ is a product of two independent Gaussian random variables with mean 0, hence $E\left[w(t_i)\left[w(t_{i+1}) - w(t_i)\right]\right] = 0$ for all i. Therefore,

$$E\left[\int_a^b w dw\right] = E\left[\lim_{\Delta \to 0} \sum_i w(t_i)\left[w(t_{i+1}) - w(t_i)\right]\right] = 0.$$

Now we will calculate the Stratonovich stochastic integral:

$$\begin{aligned}\int_a^b w \circ dw &= \lim_{\Delta \to 0} \sum_i w(t_{i+1/2})(w(t_{i+1}) - w(t_i)) \\ &= \lim_{\Delta \to 0} \sum_i [w(t_{i+1/2}) - w(t_i) + w(t_i)][w(t_{i+1}) - w(t_{i+1/2}) + w(t_{i+1/2}) - w(t_i)] \\ &= \lim_{\Delta \to 0} \sum_i (w(t_{i+1/2}) - w(t_i))^2 \\ &+ \lim_{\Delta \to 0} \sum_i (w(t_{i+1/2}) - w(t_i))(w(t_{i+1}) - w(t_{i+1/2})) \\ &+ \lim_{\Delta \to 0} \sum_i w(t_i)\left[w(t_{i+1}) - w(t_i)\right].\end{aligned}$$

The first limit can be evaluated by finding its mean $(b-a)/2$ and showing that its variance tends to zero as $n \to \infty$. The second limit is zero as it is the sum of products of independent Gaussian random variables with mean zero and vanishing variance as as $n \to \infty$. The third limit is the Ito stochastic integral that we have just evaluated. Hence,

$$\int_a^b w \circ dw = \frac{b-a}{2} + \frac{w^2(b) - w^2(a)}{2} - \frac{b-a}{2} = \frac{w^2(b) - w^2(a)}{2}.$$

As you see,

$$\int_a^b w \circ dw \neq \int_a^b w dw.$$

References

[1] A. Chorin and O. Hald, Stochastic Tools in Mathematics and Science, 3rd edition, Springer 2013
[2] L. Koralov and Ya. Sinai, Theory of probability and stochastic processes, 2nd edition, Springer, 2007
[3] Peter Moertens and Yuval Peres, Brownian Motion, electronic book, 2008
[4] Zeev Schuss, Theory and Applications of Stochastic Processes, An analytical approach, Springer, 2010
[5] Grigorios Pavliotis, Stochastic processes and Applications, Diffusion Processes, the Fokker-Planck, and Langevin Equations, Springer, 2014
[6] W. E, E. Vanden-Eijnden, Towards a theory of transition paths, Journal of statistical physics 123 (3) (2006) 503.
[7] W. E, E. Vanden-Eijnden, Transition-path theory and path-finding algorithms for the study of rare events, Annu. Rev. Phys. Chem. 61 (2010) 391–420.
[8] Wikipedia