# Lecture notes for Numerik IVc - Numerics for Stochastic Processes, Wintersemester 2012/2013. Instructor: Prof. Carsten Hartmann

Scribe: H. Lie

February 2, 2013

## Contents

Figure 1.1: Simulation of a butane molecule and its approximation by a 3-state Markov chain (states in blue, green and yellow; solvent molecules not shown).

# 1 Day 1, 16.10.2012: Modelling issues

## 1.1 Time-discrete Markov chains

Time index set $I$ is discrete, e.g. $I \subseteq \mathbb{N}$ and state space $S$ is countable or finite, e.g. $S = \{s_1, s_2, s_3\}$ (see Figure 1.1). Key objects are transition probabilities. For a state space $S = \{1, \ldots, n\}$, the transition probabilities $p_{ij}$ satisfy

$$p_{ij} = \mathbb{P}\left(X_{t+1} = j \mid X_t = i\right)$$

and yield a row-stochastic matrix $P = (p_{ij})_{i,j \in S}$.

## 1.2 Markov jump processes

These are time-continuous, discrete state-space Markov chains. Time index set $I \subseteq \mathbb{R}_+$, $S$ discrete. For a fixed time step $h > 0$, the transition probabilities are given by (see Figure 1.2)

$$\mathbb{P}\left(X_{t+h} = s_j \mid X_t = s_i\right) = h\ell_{ij} + o(h)$$

where $L = (\ell_{ij})_{i,j \in S}$ and $P_h$ are matrices satisfying $P_h = \exp(hL)$.

**Note:** the matrix $L$ is row sum zero, i.e. $\sum_j \ell_{ij} = 0$. The waiting times for the Markov chain in any state $s_i$ are exponentially distributed in the sense that

$$\mathbb{P}\left(X_{t+s} = s_i, \ s \in [0, \tau) \mid X_t = s_i\right) = \exp(\ell_{ii}\tau)$$

3
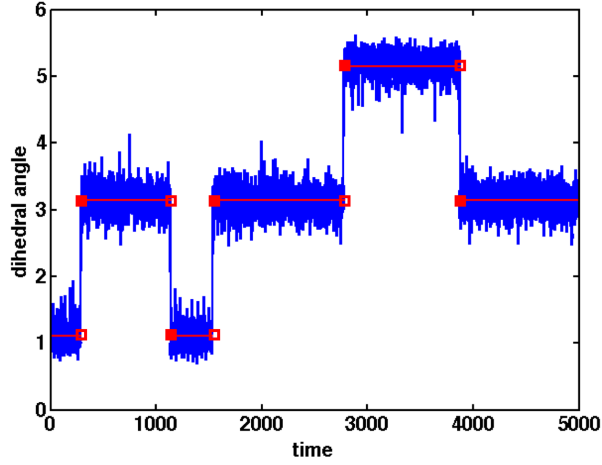
Figure 1.2: Simulation of butane: typical time series of the central dihedral angle (blue: metastable diffusion process, red: Markov jump process)

and the 'average waiting time' is $-\ell_{ii}$ (by definition of the exponential distribution).

Note: the spectrum of the matrix $P_h$ is contained within the unit disk, i.e. for every eigenvalue $\lambda$ of $P_h$, $|\lambda| \le 1$. This property is a consequence of $P_h$ being row-stochastic, i.e. that $\sum_j P_{h,ij} = 1$. Since $P_h = \exp(hL)$ it follows that

$$\sigma(P_h) \subset D := \left\{ x \in \mathbb{R}^2 \mid |x| \le 1 \right\} \Leftrightarrow \sigma(L) \subset \mathbb{C}^- = \left\{ y \in \mathbb{C} \mid \mathrm{Re}(y) \le 0 \right\}$$

**Example 1.1.** *Suppose one has a reversible reaction in which one has a large collection of $N$ molecules of the same substance. The molecules can be either in state $A$ or state $B$ and the molecules can change between the two states. Let $k^+$ denote the rate of the reaction in which molecules change from state $A$ to $B$ and let $k^-$ denote the rate at which molecules change from state $B$ to $A$.*

*For $t > 0$, consider the quantity*

$$\mu_i^A(t) := \mathbb{P}\left(\text{number of molecules in state } A \text{ at time } t \text{ is } i\right)$$

*where $i = \{0, \dots, N\}$. One can define quantities $\mu_i^B(t)$ in a similar way, and one can construct balance laws for these quantities, e.g.*

$$\frac{d\mu_i^A(t)}{dt} = k^+ \mu_{i+1}^A(t) + k^- \mu_{i-1}^A(t) - (k^+ + k^-)\mu_i^A(t).$$

*The above balance law can be written in vector notation using a tridiagonal matrix $L$. By adding an initial condition one can obtain an initial value problem*

$$\frac{d\mu^A(t)}{dt} = L^\top \mu^A(t), \quad \mu^A(0) = \mu_0.$$

*The solution of the initial value problem above is*

$$\mu^A(t) = \mu_0 \exp\left(tL^\top\right).$$

4

## 1.3  Stochastic differential equations

These are time-continuous, continuous state space Markov chains. Stochastic differential equations (SDEs) may be considered to be ordinary differential equations (ODEs) with an additional noise term (cf. Figure 1.2). Let $b : \mathbb{R}^n \to \mathbb{R}^n$ be a smooth vector field and let $x(t)$ be a deterministic dynamical system governed by the vector field $b(\cdot)$. Then $x(t)$ evolves according to

$$\frac{dx}{dt} = b(x), \quad x(0) = x_0. \tag{1.1}$$

Now let $(B_t)_{t>0}$ be Brownian motion in $\mathbb{R}^d$, and let $(X_t)_{t>0}$ be a dynamical system in $\mathbb{R}^d$ which evolves according to the equation

$$\frac{dX_t}{dt} = b(X_t) + \frac{dB_t}{dt}. \tag{1.2}$$

The additional term $\frac{dB_t}{dt}$ represents 'noise', or random perturbations from the environment, but is not well-defined because the paths of Brownian motion are nowhere differentiable. Therefore, one sometimes writes

$$dX_t = b(X_t)dt + dB_t,$$

which is shorthand for

$$X_t = X_0 + \int_0^t b(X_t)dt + \int_0^t dB_t.$$

The most common numerical integration method for SDEs is the forward Euler method. If $x$ is a $C^1$ function of time $t$, then

$$\left.\frac{dx}{dt}\right|_{t=s} = \lim_{h \to 0} \frac{x(s+h) - x(s)}{h}.$$

The forward Euler method for ODEs of the form (1.1) is given by

$$X_{t+h} = X_t + hb(X_t)$$

and for SDEs of the form (1.2) it is given by

$$X_{t+h} = X_t + hb(X_t) + \xi_h$$

where $0 < h \ll 1$ is the integration time step and the noise term $\xi$ in the Euler method for SDEs is modeled by a mean-zero Gaussian random variable.

For stochastic dynamical systems which evolve according to SDEs as in (1.2), one can consider the probability that a system at some point $x \in \mathbb{R}^d$ will be in a set $A \subset \mathbb{R}^d$ after a short time $h > 0$:

$$\mathbb{P}\left(X_{t+h} \in A \mid X_t = x\right).$$

The associated transition probability density functions of these stochastic dynamical systems are Gaussian because the noise term in (1.2) is Gaussian.

What has been the generator matrix $L$ in case of a Markov jump process is an infinite-dimensional operator acting on a suitable Banach space. Specifically,

$$Lf(x_0) = \lim_{t \to 0} \frac{\mathbb{E}_{x_0}[f(X_t)] - f(x_0)}{t},$$

provided that the limit exists. Here $f : \mathbb{R}^n \to \mathbb{R}$ is any measurable function and $\mathbb{E}_{x_0}[\cdot]$ denotes the expectation over all random paths of $X_t$ satisfying $X_0 = x_0$. $L$ is a second-order differential operator if $f$ is twice differentiable.

# 2 Day 2, 23.10.2012: Probability theory basics

**Suggested references:** [8, 12]

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space, where $\Omega$ is a set and $\mathcal{E} \subseteq 2^\Omega$ is a $\sigma$-field or $\sigma$-algebra on $\Omega$, and $\mathbb{P}$ is a probability measure (i.e., $\mathbb{P}$ is a nonnegative, countably additive measure on $(\Omega, \mathcal{E})$ with the property $\mathbb{P}(\Omega) = 1$).

## 2.1 Conditioning of random variables

Let $A \in \mathcal{E}$ be a set of nonzero measure, i.e. $\mathbb{P}(A) > 0$ and define $\mathcal{E}_A$ to be the set of all subsets of $A$ which are elements of $\mathcal{E}$, i.e.

$$\mathcal{E}_A := \{E \subset A \mid E \in \mathcal{E}\}.$$

**Definition 2.1** (Conditional probability, part I). *For an event $A$ and an event $E \in \mathcal{E}_A$, the* conditional probability *of $E$ given $A$ is*

$$\mathbb{P}(E|A) := \frac{\mathbb{P}(E \cap A)}{\mathbb{P}(A)}.$$

**Remark 2.2.** *Think of $\mathbb{P}_A := \mathbb{P}(\cdot \,|A)$ as a probability measure on the measurable space $(A, \mathcal{E}_A)$.*

Given a set $B \in \mathcal{E}$, the *characteristic* or *indicator* function $\chi_B : \Omega \to \{0, 1\}$ satisfies

$$\chi_B(x) = \begin{cases} 1 & x \in B \\ 0 & x \notin B. \end{cases}$$

**Definition 2.3** (Conditional expectation, part I). *Let $X : \Omega \to \mathbb{R}$ be a random variable with finite expectation with respect to $\mathbb{P}$. The* conditional expectation *of $X$ given an event $A$ is*

$$\mathbb{E}(X|A) = \frac{\mathbb{E}[X\chi_A]}{\mathbb{P}(A)}.$$

**Remark 2.4.** *We have*

$$\mathbb{E}(X|A) = \frac{1}{\mathbb{P}(A)} \int_A X d\mathbb{P} = \int X d\mathbb{P}_A.$$

**Remark 2.5.** *Observe that $\mathbb{P}(E|A) = \mathbb{E}[\chi_E|A]$.*

Up to this point we have only considered the case where $A$ satisfies $\mathbb{P}(A) > 0$. We now consider the general case.

**Definition 2.6** (Conditional expectation, part II). *Let $X : \Omega \to \mathbb{R}$ be an integrable random variable with respect to $\mathbb{P}$ and let $\mathcal{F} \subset \mathcal{E}$ be any sub-sigma algebra of $\mathcal{E}$. The* conditional expectation *of $X$ given $\mathcal{F}$ is a random variable $Y := \mathbb{E}[X|\mathcal{F}]$ with the following properties:*

- *$Y$ is measurable with respect to $\mathcal{F}$: $\forall B \in \mathcal{B}(\mathbb{R})$, $Y^{-1}(B) \in \mathcal{F}$.*

- *We have*

$$\int_F X d\mathbb{P} = \int_F Y d\mathbb{P} \quad \forall F \in \mathcal{F}.$$

**Remark 2.7.** *The second condition in the last definition amounts to the projection property as can be seen by noting that*

$$\mathbb{E}\left[X\chi_F\right] = \int_F X d\mathbb{P} = \int_F Y d\mathbb{P} = \mathbb{E}\left[Y\chi_F\right] = \mathbb{E}\left[\mathbb{E}\left[X|\mathcal{F}\right]\chi_F\right].$$

*By the Radon-Nikodym theorem [12], the conditional expectation exists and is unique up to $\mathbb{P}$-null sets.*

**Definition 2.8** (Conditional probability, part II). *Define the conditional probability of an event $E \in \mathcal{E}$ given $A$ by $\mathbb{P}(E|A) := \mathbb{E}\left[\chi_E|A\right]$*

**Exercise 2.9.** *Let $X$, $Y : \Omega \to \mathbb{R}$ and scalars $a, b \in \mathbb{R}$. Prove the following properties of the conditional expectation:*

- *(Linearity):*
$$\mathbb{E}\left[aX + bY|A\right] = a\mathbb{E}\left[X|A\right] + b\mathbb{E}\left[Y|A\right].$$

- *(Law of total expectation):*
$$\mathbb{E}\left[X\right] = \mathbb{E}\left[X|A\right] + \mathbb{P}(A) + \mathbb{E}\left[X|A^c\right]\mathbb{P}(A^c)$$

- *(Law of total probability):*
$$\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c).$$

**Example 2.10.** *The following is a collection of standard examples.*

- *Gaussian random variables: Let $X_1$, $X_2$ be jointly Gaussian with distribution $N(\mu, \Sigma)$, where*
$$\mu = \begin{pmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \end{pmatrix}, \quad \Sigma = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

*such that $\Sigma$ is positive definite. The density of the distribution is*
$$\rho(x) = \frac{1}{\sqrt{det(2\pi\Sigma)}} \exp\left[-\frac{1}{2}(x-\mu)^\top \Sigma (x-\mu)\right]$$

*(Ex.: Compute the distribution of $X_1$ given that $X_2 = a$ for some $a \in \mathbb{R}$.)*

- *(Conditioning as coarse-graining): Let $Z = \{Z_i\}_{i=1}^M$ be a partition of $\Omega$, i.e. $\Omega = \cup_{i=1}^M Z_i$ with $Z_i \cap Z_j = \emptyset$ and define*
$$Y(\omega) = \sum_{i=1}^M \mathbb{E}\left[X \mid Z_i\right]\chi_{Z_i}(\omega).$$

*Then $Y = \mathbb{E}\left[X|Z\right]$ is a conditional expectation (cf. Figure 2.1)*

- *(Exponential waiting times): exponential waiting times are random variables $T : \Omega \to [0, \infty)$ with the memoryless property:*
$$\mathbb{P}\left(T > s + t \mid T > s\right) = \mathbb{P}\left(T > t\right).$$

*This property is equivalent to the statement that $T$ has an exponential distribution, i.e. that $\mathbb{P}(T > t) = \exp(-\lambda t)$ for a parameter value $\lambda > 0$.*
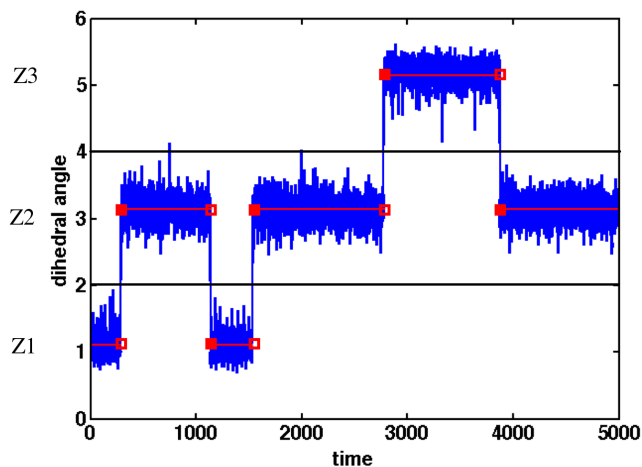
Figure 2.1: Simulation of butane, coarse-grained into three states $Z_1$, $Z_2$, $Z_3$.

## 2.2 Stochastic processes

**Definition 2.11** (Stochastic process). *A stochastic process $X = \{X_t\}_{t \in I}$ is a collection of random variables on a probability space $(\Omega, \mathcal{E}, \mathbb{P})$ indexed by a parameter $t \in I \subseteq [0, \infty)$. We call $X$*

- discrete *in time if $I \subseteq \mathbb{N}_0$*

- continuous *in time if $I = [0, T]$ for any $T < \infty$.*

How does one define probabilities for $X$? We provide a basic argument to illustrate the possible difficulties in defining the probability of a stochastic process in an unambiguous way. By definition of a stochastic process, $X_t = X_t(\omega)$ is measurable for every fixed $t \in I$, but if one has an event of the form

$$E = \{\omega \in \Omega \mid X_t(\omega) \in [a, b] \ \forall t \in I\}$$

how does one define the probability of this event? If $t$ is discrete, the $\sigma$-additivity of $\mathbb{P}$ saves us, together with the measurability of $X_t$ for every $t$. If, however, the process is time-continuous, $X_t$ is defined only almost surely (a.s.) and we are free to change $X_t$ on a set $A_t$ with $\mathbb{P}(A_t) = 0$. By this method we can change $X_t$ on $A = \cup_{t \in I} A_t$, The problem now is that $\mathbb{P}(A)$ need not be equal to zero even though $\mathbb{P}(A_t) = 0 \ \forall t \in I$. Furthermore, $\mathbb{P}(E)$ may not be uniquely defined. So what can we do? The solution to the question of how to define probabilities for stochastic processes is to use finite-dimensional distributions or marginals.

**Definition 2.12.** *(Finite dimensional distributions): Fix $d \in \mathbb{N}$, $t_1, \ldots, t_d \in I$. The* finite-dimensional distributions *of the stochastic process $X$ for $(t_1, \ldots, t_d)$ are defined as*

$$\mu_{t_1, \ldots, t_d}(B) := \mathbb{P}_{(X_{t_k})_{k=1,\ldots,d}}(B) = \mathbb{P}\left(\{\omega \in \Omega \mid (X_{t_1}(\omega), \ldots, X_{t_d}(\omega)) \in B\}\right)$$

*for $B \in \mathcal{B}(\mathbb{R}^d)$.*

Here and in the following we use the shorthand notation $\mathbb{P}_Y := \mathbb{P} \circ Y^{-1}$ to denote the *push forward* of $\mathbb{P}$ by the random variable $Y$.

**Theorem 2.13.** *(Kolmogorov Extension Theorem): Fix $d \in \mathbb{N}$, $t_1, \ldots, t_d \in I$, and let $\mu_{t_1,\ldots,t_d}$ be a consistent family of finite-dimensional distributions, i.e.*

- *for any permutation $\pi$ of $(1, \ldots, d)$,*

$$\mu_{t_1,\ldots,t_d}(B_1 \times \ldots B_d) = \mu_{(t_{\pi(1)},\ldots,t_{\pi(d)})}(B_{\pi(1)} \times \ldots \times B_{\pi(d)})$$

- *For $t_1, \ldots, t_{d+1} \in I$, we have that*

$$\mu_{t_1,\ldots,t_{d+1}}(B_1 \times \ldots B_d \times \mathbb{R}) = \mu_{t_1,\ldots,t_d}(B_1 \times \ldots \times B_d).$$

*Then there exists a stochastic process $X = (X_t)_{t \in I}$ with $\mu_{t_1,\ldots t_d}$ as its finite-dimensional distribution.*

**Remark 2.14.** *The Kolmogorov Extension Theorem does not guarantee uniqueness, not even $\mathbb{P}$-a.s. uniqueness, and, as we will see later on, such a kind of uniqueness would not be a desirable property of a stochastic process.*

**Definition 2.15.** *(Filtration generated by a stochastic process $X$): Let $\mathcal{F} = \{\mathcal{F}_t\}_{t \in I}$ with $\mathcal{F}_s \subset \mathcal{F}_t$ for $s < t$ be a filtration generated by $\mathcal{F}_t = \sigma(\{X_s \mid s \leq t\})$ is called the filtration generated by $X$.*

## 2.3 Markov processes

**Definition 2.16.** *A stochastic process $X$ is a* Markov process *if*

$$\mathbb{P}(X_{t+s} \in A \mid \mathcal{F}_s) = \mathbb{P}(X_{t+s} \in A \mid X_s) \tag{2.1}$$

*where*

$$\mathbb{P}(\cdot \mid X_s) := \mathbb{P}(\cdot \mid \sigma(X_s)),$$
$$\mathbb{P}(E \mid \sigma(X_s)) := \mathbb{E}[\chi_E \mid \sigma(X_s)]$$

*for some event $E$.*

**Remark 2.17.** *If $I$ is discrete, then $X$ is a Markov process if*

$$\mathbb{P}(X_{n+1} \in A \mid X_0 = x_0, \ldots, X_n = x_n) = \mathbb{P}(X_{n+1} \in A \mid X_n = x_n)$$

**Example 2.18.** *Consider a Markov Chain $(X_t)_{t \in \mathbb{N}_0}$ on a continuous state space $S \subset \mathbb{R}$ and let $\mathcal{S}$ be a $\sigma$-algebra on $S$. Let the evolution of $(X_t)_{t \in \mathbb{N}_0}$ be described by the* transition kernel *$p(\cdot, \cdot) : S \times \mathcal{S} \to [0, 1]$ which gives the single-step transition probabilities:*

$$p(x, A) := \mathbb{P}(X_{t+1} \in A \mid X_t = x)$$
$$= \int_A q(x, y) dy.$$

*In the above, $A \in \mathcal{B}(S)$ and $q = \frac{d\mathbb{P}}{d\lambda}$ is the density of the transition kernel with respect to Lebesgue measure. The transition kernel has the property that*

$\forall x \in S$, $p(x, \cdot)$ is a probability measure on $S$, while for every $A \in S$, $p(\cdot, A)$ is a measurable function on $S$.

For a concrete example, consider the Euler-Maruyama discretization of an SDE for a fixed time step $\Delta t$,

$$X_{n+1} = X_n + \sqrt{\Delta t}\xi_{n+1}, \quad X_0 = 0,$$

where $(\xi_i)_{i \in \mathbb{N}}$ are independent, identically distributed (i.i.d) Gaussian $\mathcal{N}(0,1)$ random variables. The process $(X_i)_{i \in \mathbb{N}}$ is a Markov Chain on $\mathbb{R}$. The transition kernel $p(x, A)$ has the Gaussian transition density

$$q(x, y) = \frac{1}{\sqrt{2\pi\Delta t}} \exp\left[-\frac{1}{2}\frac{|y-x|^2}{\Delta t}\right].$$

Thus, if $X_n = x$, then the probability that $X_{n+1} \in A \subset \mathbb{R}$ is given by

$$\mathbb{P}\left(X_{n+1} \in A | X_n = x\right) = \int_A q(x, y) dy.$$

# 3 Day 3, 30.10.2012: Brownian motion

**Suggested references**: [1, 12, 15]

## 3.1 Scaled random walks

Recapitulation:

- A stochastic process $X = (X_t)_{t \in I}$ is a collection of random variables $X_t : \Omega \to \mathbb{R}$ indexed by $t \in I$ (e.g. $I = [0, \infty)$) on some probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

- A filtration $\mathcal{F} := (\mathcal{F}_t)_{t \in I}$ is a collection of increasing sigma-algebras satisfying $\mathcal{F}_t \subset \mathcal{F}_s$ for $t < s$. A stochastic process $X$ is said to be *adapted* to $\mathcal{F}$ if $(X_s)_{s \leq t}$ is $\mathcal{F}_t$-measurable. For example, if we define $\mathcal{F}_t := \sigma(X_s : s \leq t)$, then $X$ is adapted to $\mathcal{F}$.

- The probability distribution of a random variable $X$ is given in terms of its finite dimensional distributions.

**Example 3.1** (Continued from last week). *Let $I = \mathbb{N}_0$ and consider a sequence $(X_n)_{n \in \mathbb{N}_0}$ of random variables $X_n = X_n^{\Delta t}$ governed by the relation*

$$X_{n+1}^{\Delta t} = X_n^{\Delta t} + \sqrt{\Delta t}\xi_{n+1}, \quad X_0^{\Delta t} = 0 \, a.s. \tag{3.1}$$

*where $\Delta t > 0$, and $(\xi_k)_{k \in \mathbb{N}_0}$ are i.i.d. random variables with $\mathbb{E}[\xi_k] = 0$ and $\mathbb{E}[\xi_k^2] = 1$ (not necessarily Gaussian). To obtain a continuous-time stochastic process, the values of the stochastic process on non-integer time values may be obtained by linear interpolation (cf. Figure 3.1 below). We want to consider the limiting behaviour of the stochastic process in the limit as $\Delta t$ goes to zero. Set $\Delta t = t/N$ for a fixed terminal time $t < \infty$ and let $N \to \infty$ ($\Delta t \to 0$). Then, by the central limit theorem,*

$$X_N^{\Delta t} = \sqrt{\frac{t}{N}} \sum_{k=1}^{N} \xi_k \rightharpoonup \sqrt{t}Z \tag{3.2}$$
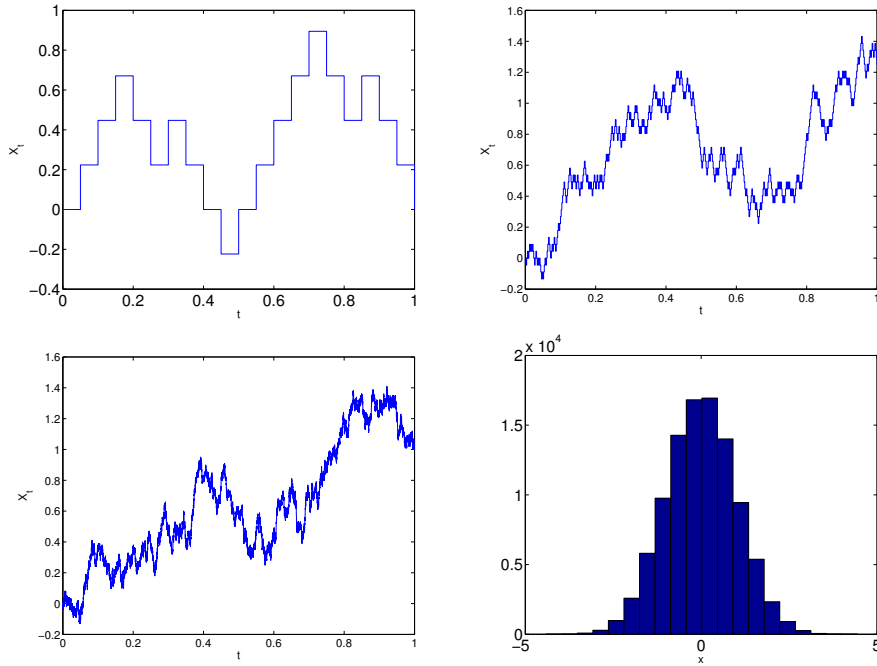
Figure 3.1: Sample paths of $(X_n^{\Delta t})_n$ for $\Delta t = 0.05, 0.002, 0.001$ over the unit time interval $[0, 1]$, with piecewise constant interpolation. The lower right plot shows the histogram (i.e., the unnormalized empirical distribution) of $(X_{1000}^{\Delta t})$ at time $t = 1$, averaged over $10\,000$ independent realizations.

*where $Z \sim \mathcal{N}(0,1)$, and "$\rightharpoonup$" means "convergence in distribution", i.e., weak convergence of the induced probability measure; equivalently, the limiting random variable is distributed according to $\mathcal{N}(0,t)$. In other words the limiting distribution of the random variable $X_N^{\Delta t}$ for fixed $t = N\Delta t$ is the same as the distribution of a centered Gaussian random variable with variance $t$. As this is true for any $t > 0$, we can think of the limiting process as a continuous-time Markov process $B = (B_t)_{t>0}$ with Gaussian transition probabilities,*

$$\mathbb{P}\left(B_{t+s} \in A \mid B_s = x\right) = \int_A q_{s,t}(x,y)dy$$
$$= \frac{1}{\sqrt{2\pi|t-s|}} \int_A \exp\left(-\frac{|y-x|^2}{2|t-s|}\right) dy.$$

*The stochastic process $B$ is homogeneous or time-homogeneous because the transition probability density $q_{s,t}(\cdot,\cdot)$ does not depend on the actual values of $t$ and $s$, but only on their difference, i.e.,*

$$q_{s,t}(\cdot,\cdot) = \tilde{q}_{|s-t|}(\cdot,\cdot) \tag{3.3}$$

**Remark 3.2.** *The choice of exponent $1/2$ in $\sqrt{\Delta t} = (\Delta t)^{1/2}$ in (3.2) is unique. For $(\Delta t)^\alpha$ with $\alpha \in (0, \frac{1}{2})$, the limit of $X_n^{\Delta t}$ "explodes" in the sense that the*

*variance of the process blows up, i.e., $\mathbb{E}[(X_N^{\Delta t})^2] \to \infty$ as $N \to \infty$. On the other hand, for $(\Delta t)^\alpha$ with $\alpha > 1/2$, $X_N^{\Delta t} \to 0$ in probability as $N \to \infty$.*

## 3.2 Constructing Brownian motion

Brownian motion is named after the British botanist, Robert Brown (1773-1858), who first observed the random motion of pollen particles suspended in water. Einstein called the Brownian process "Zitterbewegung" in his 1905 paper, *Über die von der molekularkinetischen Theorie der Wärme gefordete Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen.* The Brownian motion is a continuous-time stochastic process which is nowhere differentiable. It is also a *martingale* in the sense that on average, the particle stays in the same location at which it was first observed. In other words, the best estimate of where the particle will be after a time $t > 0$ is its initial location.

**Definition 3.3.** *(Brownian motion) The stochastic process $B = (B_t)_{t>0}$ with $B_t \in \mathbb{R}$ is called the 1-dimensional* Brownian motion *or the 1-dimensional* Wiener process *if it has the following properties:*

*(i) $B_0 = 0$ $\mathbb{P}$-a.s.*

*(ii) $B$ has* independent increments, *i.e., for all $s < t$, $(B_t - B_s)$ is a random variable which is independent of $B_r$ for $0 \leq r \leq s$.*

*(iii) $B$ has* stationary, Gaussian increments, *i.e., for $t > s$ we have[1]*

$$B_t - B_s \overset{D}{=} B_{t-s} \tag{3.4a}$$

$$\overset{D}{=} \mathcal{N}(0, t - s). \tag{3.4b}$$

*(iv) Trajectories of Brownian motion are continuous functions of time.*

We now make precise some important notions:

**Definition 3.4.** *(Filtered probability space) A filtered probability space is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t \geq 0}$ such that $\forall t \geq 0$,*

$$\mathcal{F}_t \subset \mathcal{F}.$$

**Remark 3.5.** *One may write $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ to refer to a filtered probability space. However, if one is working with a particular stochastic process $X$, one may consider the sigma-algebra $\mathcal{F}$ on $\Omega$ to simply be the smallest sigma-algebra which contains the union of the $\mathcal{F}_t^X$, where $\mathcal{F}_t^X := \sigma(X_s : s \leq t)$. In symbols, we define the sigma-algebra in the probability space to be*

$$\mathcal{F} := \vee_{t \geq 0} \mathcal{F}_t := \sigma\left(\cup_{t \geq 0} \mathcal{F}_t\right).$$

**Definition 3.6** (Martingale)**.** *A stochastic process $X = (X_t)_{t>0}$ is a* martingale *with respect to a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ if $X$ satisfies the following properties:*

*(i) $X$ is adapted to $\mathcal{F}$, i.e. $X_t$ is measurable with respect to $\mathcal{F}_t$ for every $t \geq 0$*

---

[1] The notation "$X \overset{D}{=} Y$" means "$X$ has the same distribution as $Y$".

*(ii) X is integrable: $X \in L^1(\Omega, \mathbb{P})$, i.e.*

$$\mathbb{E}\left[|X|\right] = \int_\Omega |X(\omega)| d\mathbb{P}(\omega) < \infty$$

*(iii) X has the* martingale property*: $\forall t > s \geq 0$*

$$\mathbb{E}\left[X_t | \mathcal{F}_s\right] = X_s.$$

**Definition 3.7.** *(Gaussian process) A 1-dimensional process $G = (G_t)_{t>0}$ is called a* Gaussian process *if for any collection $(t_1, \ldots, t_m) \subset I$ for arbitrary $m \in \mathbb{N}_0$, the random variable $(G_{t_1}, \ldots, G_{t_m})$ has a Gaussian distribution, i.e. it has a density*

$$f(g) = \frac{1}{\sqrt{det(2\pi\Sigma)}} \exp\left[-\frac{1}{2}(g-\mu)^\top \Sigma^{-1}(g-\mu)\right] \tag{3.5}$$

*where $g = (g_1, \ldots, g_m)$, $\mu \in \mathbb{R}^m$ is a constant vector of means and $\Sigma = \Sigma^\top \in \mathbb{R}^{m \times m}$ is a symmetric positive semi-definite matrix.*

**Remark 3.8.** *The Brownian motion process is a Gaussian process with the vector of means $\mu = 0$ and covariance matrix*

$$\Sigma = \begin{pmatrix} t_1 & 0 & \ldots & 0 \\ 0 & t_2 - t_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & t_m - t_{m-1} \end{pmatrix} \tag{3.6}$$

*The covariance matrix is diagonal due to the independence of the increments of Brownian motion.*

**Remark 3.9.** *Further remarks are in order.*

*(a) Conditions (i)-(iii) define a consistent family of finite-dimensional distributions. Hence, the existence of the process B is guaranteed by the Kolmogorov Extension Theorem.*

*(b) Conditions (i)-(iii) imply that $\mathbb{E}\left[B_t\right] = 0$ and $\mathbb{E}\left[B_t B_s\right] = \min(t, s)$ $\forall s, t \in \mathbb{R}$. The proof is left as an exercise.*

*(c) The discrete process $(X_n^{\Delta t})_{n \in \mathbb{N}_0}$ converges in distribution to a Brownian motion $(B_t)_{t \geq 0}$ if the time discrete is linearly interpolated between two successive points. In other words, if we consider the continuous-time stochastic processes $(X_t^{\Delta t})_{t>0}$ (which is obtained by linear interpolation between the $X_N^{\Delta t}$) and B as random variables on the space of continuous trajectories $(C(\mathbb{R}_+)$ and $\mathcal{B}(C(\mathbb{R}_+)))$, then the process $(X_t^{\Delta t})_{t>0}$ converges in distribution to B.*

*(d) We have that*

$$\mathbb{E}\left[(B_t - B_s)^2\right] = \mathbb{E}\left[(B_{t-s})^2\right] \text{ by (3.4a) in Definition 3.3}$$
$$= |t - s| \text{ by (3.4b) in Definition 3.3.}$$

*(e) Brownian motion enjoys the following scaling invariance, also known as self-similarity of Brownian motion: for every $t > 0$ and $\alpha > 0$,*

$$B_t \stackrel{D}{=} \alpha^{-1/2} B_{\alpha t}.$$

## 3.3 Karhunen-Loève expansion

Observe that we have constructed Brownian motion by starting with the scaled random walk process and using the Kolmogorov Extension Theorem. Now we present an alternative method for constructing Brownian motion that is useful for numerics, called the Karhunen-Loève expansion of Brownian motion. We will consider this expansion for Brownian motion on the unit time interval $[0, 1]$.

Let $\{\eta_k\}_{k \in \mathbb{N}}$ be a collection of independent, identically distributed (i.i.d) Gaussian random variables distributed according to $\mathcal{N}(0, 1)$, and let $\{\phi_k(t)\}_{k \in \mathbb{N}}$ be an orthonormal basis of

$$L^2([0, 1]) = \left\{ u : [0, 1] \to \mathbb{R} : \int_0^1 |u(t)|^2 dt < \infty \right\}. \tag{3.7}$$

By construction, the basis functions satisfy

$$\langle \phi_i, \phi_j \rangle = \int_0^1 \phi_i(t) \phi_j(t) dt = \delta_{ij},$$

and we can represent any function $\forall f \in L^2([0, 1])$ by

$$f(t) = \sum_{k \in \mathbb{N}} \alpha_k \phi_k(t)$$

for $\alpha_k = \langle f, \phi_k \rangle$. We have the following result.

**Theorem 3.10.** *(Karhunen-Loève): The process $(W_t)_{0 \leq t \leq 1}$ defined by*

$$W_t = \sum_{k \in \mathbb{N}} \eta_k \int_0^t \phi_k(s) ds \tag{3.8}$$

*is a Brownian motion.*

*Proof.* We give only a sketch of the proof. For details, see the Appendix in [12], or [7]). The key components of the proof are to show the following:

(i) The infinite sum which defines the Karhunen-Loève expansion is absolutely convergent, uniformly on $[0, 1]$.

(ii) It holds that $\mathbb{E}[W_t] = 0$ and $\mathbb{E}[W_t W_s] = \min(s, t)$.

$\square$

# 4 Day 4, 06.11.2012: Brownian motion, cont'd

**Suggested references:** [7, 18]

## 4.1 More properties of Brownian motion

From last week, we saw that the Brownian motion $(B_t)_{t \geq 0}$ is a continuous-time stochastic process on $\mathbb{R}$ with

- stationary, independent, Gaussian increments

- a.s. continuous paths. That is, for fixed $\omega$, each $(B_t)_{t \geq 0}(\omega)$ is a continuous trajectory in $\mathbb{R}$.

Moreover the scaled random walk defined by

$$X_{n+1}^{\Delta t} = X_n^{\Delta t} + \sqrt{\Delta t} \xi_{n+1}$$

with linear interpolation converges weakly (i.e. converges in distribution) to the Brownian motion process. Above, the $(\xi_n)_{n \in \mathbb{N}}$ are independent, identically distributed (i.i.d) normalized Gaussian random variables (i.e. $\xi_n$ is Gaussian with mean zero and variance 1).

**Remark 4.1.** *Two remarks are in order.*

- *Continuity can be understood using the Lévy construction of Brownian motion on the set of dyadic rationals,*

$$D := \bigcup_{n \in \mathbb{N}} D_n, \quad D_n := \left\{ \frac{k}{2^n} \ : \ k = 0 \ldots, 2^n \right\}.$$

*The construction of Brownian motion on the unit time interval is as follows. Let $\{Z_t\}_{t \in D}$ be a collection of independent, normalized random variables defined on a probability space. Define the collection of functions $(F_n)_{n \in \mathbb{N}}$, where $F_n : [0,1] \to \mathbb{R}$ are given by*

$$F_n(t) := \begin{cases} 0 & t \in D_{n-1} \\ 2^{-(j+1)/2} Z_t & t \in D_j \setminus D_{j-1} \\ lin.\ interp. & in\ between. \end{cases}$$

*Then the process*

$$B(t) = \sum_{n=1}^{\infty} F_n(t).$$

*is indeed a Brownian motion on $[0,1]$. The Gaussianity of the $\{Z_t\}_{t \in D}$ leads to the stationary, independent Gaussian increments of the process $(B_t)_{t \in [0,1]}$. The continuity of the process follows from an application of the Borel-Cantelli Lemma, which states that there exists a random and almost surely finite number $N \in \mathbb{N}$ such that for all $n \geq N$ and $d \in D_n$, $|Z_d| < c\sqrt{n}$ holds. This boundedness condition implies that $\forall n \geq N$ we have a decay condition for the $F_n$:*

$$\|F_n\|_\infty < c\sqrt{n} 2^{-n/2}.$$

*Therefore the sum $\sum_j F_j(\cdot)$ converges uniformly on $[0,1]$. As each $F_j$ is continuous and the uniform limit of continuous functions is continuous, the process $(B_t)_{t \in [0,1]}$ is continuous. For more details, see [14].*

- *The Hausdorff dimension $\dim_{\mathcal{H}}$ of Brownian motion paths depends on the dimension of the space $\mathbb{R}^d$ in which the Brownian motion paths live.[2] Let $B_{[0,1]} = \{B_t \in \mathbb{R}^d : t \in [0,1]\}$ be the graph of $B_t$ over $I = [0,1]$. Then*

$$\dim_{\mathcal{H}} B_{[0,1]} = \begin{cases} 3/2 & d = 1 \\ 2 & d \geq 2. \end{cases}$$

---

[2]If you do not know what this is, just think of the box counting dimension that is an upper limit of the Hausdorff dimension.
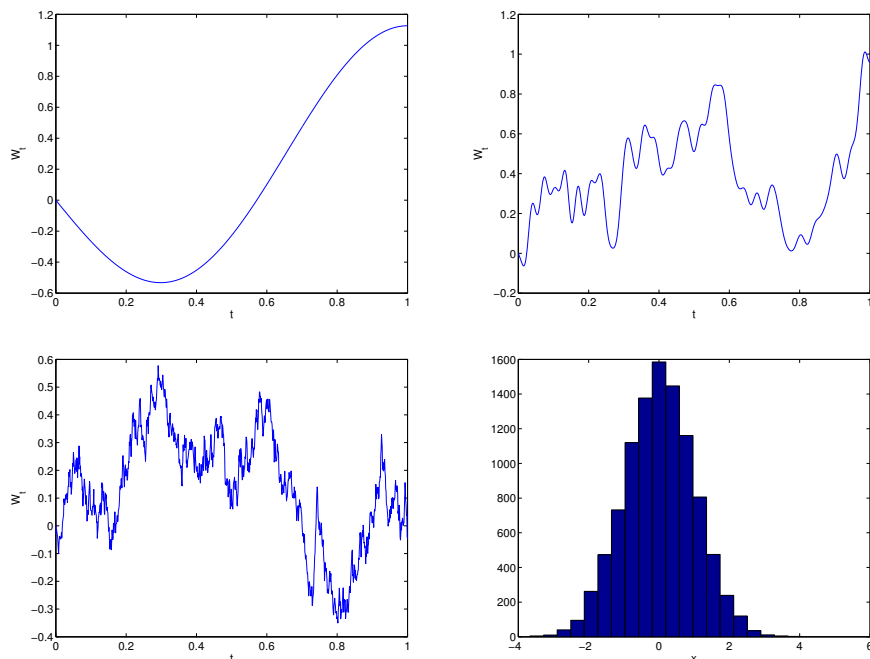
Figure 4.1: Sample paths of the Karhunen-Loève expansion of ($W_t$ for $M =$ 2, 64, 2048 basis functions (you can guess which one is which). The lower right plot shows the unnormalized histogram of $W_t$ at time $t = 1$, using $M = 64$ basis functions and averaged over 10 000 independent realizations.

*The significance of this is as follows: if you consider Brownian motion paths confined to a smooth and compact two-dimensional domain and impose reflecting boundary conditions, then the Brownian motion paths will fill the domain in the limit as $t \to \infty$.*

## 4.2   Brownian bridge

Recall the Karhunen-Loève expansion of Brownian motion:

**Theorem 4.2.** *Let $\{\eta_k\}_{k \in \mathbb{N}}$ be i.i.d. normalized random variables and $\{\phi_k\}_{k \in \mathbb{N}}$ form a real orthonormal basis of $L^2([0,1])$. Then*

$$W_t = \sum_{k \in \mathbb{N}} \eta_k \int_0^t \phi_k(s) ds$$

*is a Brownian motion on the interval $I = [0, 1]$.*

**Exercise 4.3.** *Show that, for the definition of $(W_t)_{t \in [0,1]}$ above, it holds that $\mathbb{E}[W_t W_s] = \min(s, t)$.*

**Remark 4.4.** *Unlike the scaled random walk construction of Brownian motion, no forward iterations are required here. This helps for the consideration of round-off errors in the construction of $(W_t)_{t \in [0,1]}$. Furthermore:*

16

- *Standard choices for the orthonormal basis $\{\phi_k\}_{k \in \mathbb{N}}$ are Haar wavelets or trigonometric functions. Hence the numerical error can be controlled by truncating the series and by the choice of the basis.*

- *To obtain a Brownian motion on any general time interval $[0, T]$, it suffices to use the scaling property, e.g.*

$$W_{[0,T]} \overset{D}{=} \sqrt{T} W_{[0,1]/T}$$
$$= \sqrt{T} \sum_{k \in \mathbb{N}} \eta_k \int_0^{t/T} \phi_k(s) ds.$$

**Application: filtering of Brownian motion**

Suppose we know that $W_0 = 0$ and $W_1$ is equal to some constant $\omega$. Without loss of generality, let $\omega = 0$. Suppose we wanted to generate a Brownian motion path which interpolated between the values $W_0 = 0$ and $W_1 = 0$.

**Definition 4.5.** *A continuous, mean-zero Gaussian process $(BB_t)_{t \geq 0}$ is called a Brownian bridge to $\omega$ if it has the same distribution as $(W_t)_{t \in [0,1]}$ conditional on the terminal value $W_1 = \omega$. Equivalently, $(BB_t)_{t \geq 0}$ is a Brownian bridge if*

$$\mathrm{Cov}\left[BB_t BB_s\right] = \min(s, t) - st.$$

**Lemma 4.6.** *If $(W_t)_{t \in [0,1]}$ is a Brownian motion, then $BB_t = W_t - tW_1$ is a Brownian bridge.*

*Proof.* Observe that

$$\mathbb{E}\left[BB_t\right] = \mathbb{E}\left[W_t - tW_1\right] = 0 - t \cdot 0 = 0,$$

so that $(BB_t)_{t \in [0,1]}$ is indeed mean-zero. The process $(BB_t)_{t \in [0,1]}$ inherits continuity from the process $(W_t)_{t \in [0,1]}$. The covariance process is given by

$$\mathrm{Cov}(BB_t BB_s) = \mathbb{E}\left[BB_t BB_s\right] = \mathbb{E}\left[(W_t - tW_1)(W_s - sW_1)\right]$$
$$= \mathbb{E}\left[W_t W_s\right] - t \underbrace{\mathbb{E}\left[W_1 W_s\right]}_{=\min(s,1)} - s \underbrace{\mathbb{E}\left[W_1 W_t\right]}_{=\min(t,1)} + ts\mathbb{E}\left[W_1 W_1\right]$$
$$= \min(t, s) - ts - st + ts.$$

$\square$

## 4.3 Simulating a Brownian bridge

**First approach**: forward iteration, using Euler's method. The time interval is $[0, 1]$ and we have a time step of $\Delta t := 1/N$, so we have $(N+1)$ discretized time nodes $(t_n = n\Delta t)_{n=0,\ldots,N}$ and $(N+1)$ values $(Y_n^{\Delta t})_{n=0,\ldots,N}$. Let $\{\xi_n\}_{n=0,\ldots,N-1}$ be a collection of i.i.d. normalized random variables. Forward iteration gives

$$Y_{n+1}^{\Delta t} = Y_n^{\Delta t}\left(1 - \frac{\Delta t}{1 - t_n}\right) + \sqrt{\Delta t}\xi_{n+1}.$$

It holds that $1 - t_{N-1} = \Delta t$ by definition of $\Delta t = 1/N$. Therefore from the formula above we have

$$Y_N^{\Delta t} = \sqrt{\Delta t}\xi_{N+1}.$$

Therefore $Y_N^{\Delta t}$ is a mean zero Gaussian random variable with variance $\Delta t$. While this implies that $Y_N^{\Delta t}$ should converge in probability to the value 0 as the step size $\Delta t \to 0$, the forward iteration approach is not optimal because the random variable $\xi_{N+1}$ is continuous, so

$$\mathbb{P}\left(Y_N^{\Delta t} = 0\right) = 0.$$

Therefore this construction of the Brownian bridge to the value $\omega = 0$ will in general not yield processes which are at 0 at time $t = 1$. As a matter of fact, $Y_N^{\Delta t}$ is unbounded and can be arbitrarily far away from zero.

**Second approach**: Recall the Karhunen-Loéve construction of Brownian motion and choose trigonometric functions as an orthonormal basis. Then the process $(W_t)_{t \in [0,1]}$ given by

$$W_t(\omega) = \sqrt{2} \sum_{k=1}^{M} \eta_k(\omega) \frac{\sin((k - \frac{1}{2})\pi t)}{(k - \frac{1}{2})\pi}$$

is a Brownian motion and we can define the Brownian bridge to $\omega$ at $t = 1$ by

$$BB_t = W_t - t(W_1 - \omega).$$

**Remark 4.7.** *It holds that*

$$BB_t = \sqrt{2} \sum_{k \in \mathbb{N}} \eta_k \frac{\sin(k\pi t)}{k\pi}$$
$$= \sum_{k \in \mathbb{N}} \eta_k \sqrt{\lambda_k} \psi_k(t),$$

*where* $\{\lambda_k, \psi_k\}_{k \in \mathbb{N}} = \{\sqrt{2}/k\pi, \sin(k\pi t)\}_{k \in \mathbb{N}}$ *is the eigensystem of the* covariance *operator* $T : L^1([0,1]) \to L^1([0,1])$ *of the process* $(BB_t)_{t \in [0,1]}$, *defined by*

$$(Tu)(t) = \int_0^1 \underbrace{\mathrm{Cov}(BB_t BB_s)}_{= \min(t,s) - st} u(s) ds,$$

*i.e.,*

$$T\psi_k(\cdot) = \lambda_k \psi_k(\cdot).$$

*The second approach works for* any *stochastic process which has finite variance over a finite time interval. For details, see [18].*

# 5 Day 5, 13.11.2012: Stochastic integration

**Suggested references:** [1, 12, 15]

## 5.1 Integrating against Brownian motion

Recall that Brownian motion $(B_t)_{t > 0}$ is a stochastic process with the following properties:

- $B_0 = 0$ $\mathbb{P}$-a.s.
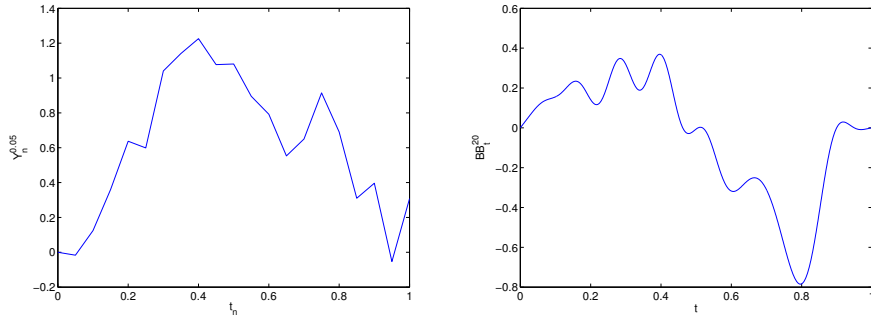
Figure 4.2: Sample paths of the Brownian bridge approximation, using the Euler scheme with $\Delta t = 0.05$ (left panel) and Karhunen-Loève expansion with $M = 20$ basis functions (right panel).

- $\forall 0 \leq t_0 < t_1 < t_2 < \ldots < t_n$, the increments $B_{t_i} - B_{t_{i-1}}$ are independent for $i = 1, \ldots, n$ and Gaussian with mean 0 and variance $t_i - t_{i-1}$.

- $t \mapsto B_t(\omega)$ is continuous $\mathbb{P}$-a.s. but is $\mathbb{P}$-a.s. nowhere differentiable.

One of the motivations for the development of the stochastic integral lies in financial mathematics, where one wishes to determine the price of an asset that evolves randomly. The French mathematician Louis Bachelier is generally considered one of the first people to model random asset prices. In his PhD thesis, Bachelier considered the following problem. Let the value $S_t$ of an asset at time $t > 0$ be modeled by

$$S_t = \sigma B_t$$

where $\sigma > 0$ is a scalar that describes the volatility of the stock price. Let $f(t)$ be the amount of money an individual invests in the asset in some infinitesimal time interval $[t, t + dt]$. Then the wealth of the individual at the end of a time interval $[0, T]$ is given by

$$\int_0^T f(t)dS_t = \sigma \int_0^T f(t)dB_t.$$

However, it is not clear what the expression '$dB_t$' means. In this section, we will consider what an integral with respect to $dB_t$ means, and we will also consider the case when the function $f$ depends not only on time but on the random element $\omega$.

## 5.2 The Itô integral for simple functions

The first idea is to rewrite

$$\int f(t)dB_t = \int f(t)\frac{dB_t}{dt}dt$$

but as Brownian motion is almost surely nowhere differentiable, we cannot write $\frac{dB_t}{dt}$.

19

The second idea is to proceed as in the definition of the Lebesgue integral: start with simple step functions and later extend the definition to more general functions by the Itô Isometry.

Step 1: Consider simple functions

$$f(t) = \sum_{i=1}^{n} a_i \chi_{(t_i, t_{i+1}]}(t)$$

where $\chi_A$ is the indicator function of a set $A$ satisfying

$$\chi_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

Observe that $f$ takes a finite number $n$ of values. By the theory of Lebesgue integration, we know that the set of these simple functions is dense in $L^2([0, \infty))$. We also know that the usual Riemann integral of such a function $f$ corresponds to the area under the graph of $f$, with

$$\int_0^\infty f(t)dt = \sum_i a_i(t_{i+1} - t_i)$$

Step 2: We now extend the method above to stochastic integral with respect to Brownian motion:

$$\int f(t)dB_t = \sum a_i(B_{t_{i+1}} - B_{t_i}).$$

**Remark 5.1.** *By the equation above, it follows that the integral $\int f(t)dB_t$ is a random variable, since the $B_{t_i}$ are random variables. Since increments of Brownian motion are independent and Gaussian, the integral $\int f(t)dB_t$ is normally distributed with zero mean. What about its variance?*

**Lemma 5.2.** *(Itô Isometry for simple functions) For a simple function $f(t) = \sum_i a_i \chi_{(t_i, t_{i+1}]}(t)$, it holds that*

$$\mathbb{E}\left[\left(\int_0^\infty f(t)dB_t\right)^2\right] = \int_0^\infty (f(t))^2 \, dt.$$

*Proof.*

$$\mathrm{var}\left(\int_0^\infty f(t)dB_t\right) = \mathrm{var}\left(\sum_i a_i\left(B_{t_{i+1}} - B_{t_i}\right)\right)$$

$$= \sum_{i=1}^n a_i^2 \mathrm{var}\left(B_{t_{i+1}} - B_{t_i}\right)$$

$$= \sum_{i=1}^n a_i^2 \left(t_{i+1} - t_i\right)$$

$$= \sum_{i=1}^n a_i^2 \int_0^\infty \chi_{(t_i,t_{i+1}]}dt$$

$$= \int \sum_{i=1}^n a_i^2 \chi_{(t_i,t_{i+1}]}dt$$

$$= \int \left(f(t)\right)^2 dt.$$

Therefore

$$\mathrm{var}\left(\int f(t)dB_t\right) = \mathbb{E}\left[\left(\int f(t)dB_t\right)^2\right] - \left(\underbrace{\mathbb{E}\left[\int f(t)dB_t\right]}_{=0}\right)^2$$

$$= \mathbb{E}\left[\left(\int f(t)dB_t\right)^2\right].$$

$\square$

Step 3: Now we extend the definition of the integral to $L^2([0,\infty))$. The main result is the following

**Theorem 5.3.** (Itô integral for $L^2([0,\infty))$ functions) *The definition of the Itô integral can be extended to elements $f \in L^2([0,\infty))$ by setting*

$$\int_0^\infty f(t)dB_t := \lim_{n\to\infty} \int_0^\infty f_n(t)dB_t$$

*where the sequence $(f_n)_{n\in\mathbb{N}}$ is a sequence of a simple functions satisfying $f_n \to f$ in $L^2([0,\infty))$, i.e.*

$$\|f_n - f\|_{L^2([0,\infty))} = \left(\int_0^\infty (f_n - f)^2(t)dt\right)^{1/2} \xrightarrow[n\to\infty]{} 0.$$

*By the Itô isometry, we can show that $(\int f_n(t)dB_t)_{n\in\mathbb{N}}$ is a Cauchy sequence in the weighted $L^2$ space*

$$L^2(\Omega,\mathbb{P}) := \left\{F: \Omega \to \mathbb{R} :\ \|F\|_{L^2(\Omega,\mathbb{P})} < \infty\right\}$$

*of measurable functions $F$ where*

$$\|F\|_{L^2(\Omega,\mathbb{P})}^2 = \int |F|^2(\omega)\, d\mathbb{P}(\omega).$$

*To show that the sequence $(\int f_n(t)dB_t)_{n\in\mathbb{N}}$ is a Cauchy sequence, let $(f_l)_{l\in\mathbb{N}}$ be a sequence of functions converging to $f$ in $L^2([0,\infty))$ and consider for $m, n \in \mathbb{N}$*

$$\left\|\int f_n(t)dB_t - \int f_m(t)dB_t\right\|_{L^2(\Omega,\mathbb{P})}$$

$$= \left(\mathbb{E}\left[\left(\int f_n(t)dB_t - \int f_m(t)dB_t\right)^2\right]\right)^{1/2}$$

$$= \left(\mathbb{E}\left[\left(\int f_n(t) - f_m(t)dB_t\right)^2\right]\right)^{1/2}$$

$$= \left(\int (f_n(t) - f_m(t))^2\, dt\right)^{1/2} \quad \text{(Itô isometry)}$$

$$= \|f_n - f_m\|_{L^2([0,\infty))}$$

$$\leq \|f_n - f\|_{L^2([0,\infty))} + \|f_m - f\|_{L^2([0,\infty))}$$

*and using that $\|f_n - f\|_{L^2([0,\infty))}$ and $\|f_m - f\|_{L^2([0,\infty))} \to 0$ as $m, n \to \infty$, the result follows.*

Since $L^2(\Omega,\mathbb{P})$ is complete, the limit exists and is in the same space. Moreover, by the Itô isometry, the limit is independent of the sequence $(f_n)_{n\in\mathbb{N}}$ used to approximate $f$ in $L^2([0,\infty))$ (see [4] for an example). $\qquad\square$

**Example 5.4.** *Consider the random variable $\int_0^\infty \exp(-t)dB_t$. How is it distributed? Using the Itô Isometry, the random variable is Gaussian with mean zero and variance $\frac{1}{2} = \int_0^\infty \exp(-2t)dt$.*

**Corollary 5.5.** *The Itô Isometry holds as well for $f \in L^2([0,\infty))$, not just simple functions.*

Step 4: Now we consider functions $f$ which depend both on the random element $\omega$ as well as time $t$. That is, we consider stochastic integrals of stochastic processes $f : [0,\infty) \times \Omega \to \mathbb{R}$ with the following properties:

(i) $f$ is $\mathcal{B} \times \mathcal{F}$-measurable, where $\mathcal{B}$ is the Borel sigma-algebra on $[0,\infty)$ and $\mathcal{F}$ is a given sigma-algebra on $\Omega$.

(ii) $f(t,\omega)$ is adapted with respect to $\mathcal{F}_t$, where $\mathcal{F}_t := \sigma\left(B_s \;:\; s \leq t\right)$

(iii) $\mathbb{E}\left[\int |f(t,\omega)|^2 dt\right] < \infty$.

Consider simple stochastic processes of the form

$$f(t,\omega) = \sum_{i=1}^n a_i(\omega)\chi_{(t_i,t_{i+1}]}(t).$$

Then

$$\int f(t,\omega)dB_t = \sum_{i=1}^n a_i(\omega)\left(B_{t_{i+1}} - B_{t_i}\right).$$

## 5.3 Ambiguities in defining stochastic integrals

**Example 5.6.** *Fix $n \in \mathbb{N}$, fix a time step $\Delta t := 2^{-n}$ and define the time nodes $t_i := i\Delta t$ for $i = 0, 1, 2, \ldots$.. Let $(B_t)_{t>0}$ be the standard Brownian motion. Define the following processes on $[0, \infty)$:*

$$f_1(t, \omega) = \sum_{i \in \mathbb{N}} B_{t_i}(\omega) \chi_{[t_i, t_{i+1})}(t)$$

$$f_2(t, \omega) = \sum_{i \in \mathbb{N}} B_{t_{i+1}}(\omega) \chi_{[t_i, t_{i+1})}(t).$$

*Now fix $T > 0$ and $N$ such that $T = t_N = N\Delta t = N2^{-n}$ and compute the expected values of the integrals of $f_1$ and $f_2$ over $[0, T]$. By the independent increments property of Brownian motion (or the martingale property of Brownian motion), we have*

$$\mathbb{E}\left[\int_0^T f_1(t, \omega) dB_t\right] = \sum_{i=0}^{N-1} \mathbb{E}\left[B_{t_i}\left(B_{t_{i+1}} - B_{t_i}\right)\right] = 0.$$

*Using the fact above with linearity of expectation, we also have*

$$
\begin{aligned}
\mathbb{E}\left[\int_0^T f_2(t, \omega) dB_t\right] &= \sum_{i=0}^{N-1} \mathbb{E}\left[B_{t_{i+1}}\left(B_{t_{i+1}} - B_{t_i}\right)\right] - 0 \\
&= \sum_{i=0}^{N-1} \left(\mathbb{E}\left[B_{t_{i+1}}\left(B_{t_{i+1}} - B_{t_i}\right)\right] - \mathbb{E}\left[B_{t_i}\left(B_{t_{i+1}} - B_{t_i}\right)\right]\right) \\
&= \sum_{i=0}^{N-1} \mathbb{E}\left[\left(B_{t_{i+1}} - B_{t_i}\right)^2\right] \\
&= \sum_{i=0}^{N-1} t_{i+1} - t_i = T.
\end{aligned}
$$

*In the case of Riemann integration of deterministic integrals, letting $n \to \infty$ would lead to the result that both integrals above are equal. We see that for stochastic integration, this is not the case; even if we let $n \to \infty$, the expectations of the Itô integrals would not be equal. This is because the choice of endpoint of the interval matters in stochastic integration. Choosing the left endpoint (i.e. choosing $B_{t_i}$) for $f_1$ and the right endpoint (i.e. $B_{t_{i+1}}$ for $f_2$ leads to different expectations. Note also that taking the right endpoint in $f_2$ leads to $f_2$ not being adapted, since $B_{t_{i+1}}$ is not measurable with respect to $\mathcal{F}_t$ for $t < t_{i+1}$. Therefore, by property (ii) above, we may not integrate $f_2$ with respect to $dB_t$ in the way we have just described.*

# 6  Day 6, 20.11.2012: Stochastic integration, cont'd

## 6.1  Itô isometry

We extend the Itô integral to the case

$$I[f](\omega) = \int_0^t f(s, \omega) dB_s(\omega),$$

where $B_t$ is one-dimensional Brownian motion. One aim is to understand

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \ X_0 = x$$

that is SDE shorthand for

$$X_t = x + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s\,.$$

A second objective later on will be to analyze discretizations of SDEs, such as

$$X_{n+1} - X_n = b(t_n, X_n)\Delta t + \sigma(t_n, X_n)\Delta B_n.$$

We begin with a couple of definitions.

**Definition 6.1.** *We call $\|\cdot\|_{\mathcal{V}}$ the norm defined by*

$$\|f\|_{\mathcal{V}}^2 = \mathbb{E}\left[\int_s^t |f(u, \cdot)|^2 du\right] = \int_\Omega \int_s^t |f(u, \omega)|^2 du \ d\mathbb{P}(\omega)\,.$$

**Definition 6.2** (Cf. the considerations at the bottom of p. 22)**.** *Let $\mathcal{V} = \mathcal{V}(s, t)$ be the class of functions $f : [0, \infty) \times \Omega \to \mathbb{R}$ with*

(i) $(t, \omega) \mapsto f(t, \omega)$ *is $\mathcal{B} \times \mathcal{F}$-measurable*[3]

(ii) $f(t, \cdot)$ *is $\mathcal{F}_t$-adapted*

(iii) $\|f\|_{\mathcal{V}} < \infty$.

**Definition 6.3.** *A simple function $\varphi : [0, \infty) \times \Omega \to \mathbb{R}$ is a function of the form*

$$\varphi(t, \omega) = \sum_j e_j(\omega)\chi_{[t_j, t_{j+1})}(t)$$

*where each $e_j$ is $\mathcal{F}_{t_j}$-measurable and $\{\mathcal{F}_t\}_{t \geq 0}$ with $\mathcal{F}_t = \sigma(B_s \colon s \leq t)$ is the filtration generated by Brownian motion.*

**Definition 6.4.** *The Itô integral for a simple function $\varphi$ is defined by*

$$I[\varphi](\omega) = \int_0^t \varphi(s, \omega)dB_s(\omega) = \sum_j e_j(\omega)(B_{t_{j+1}} - B_{t_j}).$$

**Lemma 6.5.** Itô Isometry*: If $\varphi(t, \omega)$ is a bounded, simple function, then*

$$\mathbb{E}\left[\left(\int_s^t \varphi(u, \omega)dB_u(\omega)\right)^2\right] = \mathbb{E}\left[\int_s^t |\varphi(u, \omega)|^2 du\right].$$

*Proof.* Define $\Delta B_j := B_{t_{j+1}} - B_{t_j}$. Then $I[\varphi] = \sum_j e_j \Delta B_j$. By independence of $e_i e_j \Delta B_i$ from $\Delta B_j$ when $i \neq j$, we have that

$$\mathbb{E}\left[e_i e_j \Delta B_i \Delta B_j\right] = \begin{cases} 0 & i \neq j \\ \mathbb{E}\left[e_j^2\right](t_{j+1} - t_j) & i = j. \end{cases}$$

---

[3]Here again: $\mathcal{B} = \mathcal{B}([0, \infty))$ is the $\sigma$-algebra od Borel sets over $[0, \infty)$.

Therefore,

$$\mathbb{E}\left[\left(\int_s^t \varphi(u,\cdot)dB_u\right)^2\right] = \sum_{i,j} \mathbb{E}\left[e_i e_j \Delta B_i \Delta B_j\right]$$

$$= \sum_j \mathbb{E}\left[e_j^2\right](t_{j+1} - t_j) = \mathbb{E}\left[\int_s^t \varphi(u,\cdot)du\right].$$

$\square$

## 6.2  Itô Integral for random integrands

Now we will extend the Itô integral to $\mathcal{V} = \mathcal{V}(s,t)$, by extending the Itô integral to progressively larger classes of functions.

Step 1: Let $g \in \mathcal{V}$ be a uniformly bounded function which is continuous for each $\omega$. Then there exists a sequence of simple functions $(\varphi_n)_{n\in\mathbb{N}}$ such that

$$\|\varphi_n - g\|_{\mathcal{V}} \to 0$$

as $n \to \infty$.

*Proof.* Choose $\varphi_n(t,\omega) = \sum_j g(t_j,\omega)\chi_{[t_j,t_{j+1})}(t)$. Then $\varphi_n \to g$ in $L^2([s,t])$ for each $\omega \in \Omega$, and hence $\|\varphi_n - g\|_{\mathcal{V}}^2 \to 0$, i.e.,

$$\mathbb{E}\left[\int_s^t (\varphi_n - g)^2\, du\right] = \int_\Omega \left(\int_s^t (\varphi_n - g)^2\, du\right) d\mathbb{P} \to 0$$

as $n \to \infty$. $\square$

Step 2: Let $h \in \mathcal{V}$ be bounded. Then there exists a bounded sequence of functions $(g_n)_{n\in\mathbb{N}} \subset \mathcal{V}$ such that each $g_n$ is continuous in $t$ for each $\omega$ and for each $n \in \mathbb{N}$, such that $\|g_n - h\|_{\mathcal{V}} \to 0$.

*Proof.* Suppose that $|h(t,\omega)| \le M < \infty$. For each $n$, let $\psi_n$ be defined by

(i)  $\psi_n(x) = 0$ for $x \in (-\infty, -\frac{1}{n}] \cup [0,\infty)$

(ii)  $\int_{\mathbb{R}} \psi_n(x)dx = 1$.

Now define the functions $g_n : [0,\infty) \times \Omega \to \mathbb{R}$ by

$$g_n(t,\omega) = \int_0^t \psi_n(s-t)h(s,\omega)ds.$$

Then it holds that $g_n \to h$ in $L^2([s,t])$, i.e., $\|g_n - h\|_{L^2([s,t])} \to 0$ as $n \to \infty$. As $h$ is bounded, we can apply the bounded convergence theorem to obtain

$$\mathbb{E}\left[\int_s^t (g_n - h)^2\, du\right] \to 0$$

as $n \to \infty$. $\square$

**Remark 6.6.** *In the limit as $n \to \infty$, the $\psi_n(x)$ become more sharply peaked at $x = 0$—in other words, they approach a Dirac delta distribution:*

$$h(t, \omega) = \int_0^\infty \delta(s - t) h(s, \omega) ds \,.$$

Step 3: Let $f \in \mathcal{V}$. Then there exists a sequence of functions $(h_n)_{n \in \mathbb{N}} \subset \mathcal{V}$ such that $h_n$ is bounded for each $n \in \mathbb{N}$ and $\|h_n - f\|_\mathcal{V} \to 0$ as $n \to \infty$.

*Proof.* Define

$$h_n(t, \omega) = \begin{cases} -n & f(t, \omega) < -n \\ f(t, \omega) & -n \leq f(t, \omega) \leq n \\ n & f(t, \omega) > n. \end{cases}$$

Then the assertion follows by the dominated convergence theorem. $\square$

By Steps 1-3, $f \in \mathcal{V}$ can be approximated by sequences of simply functions $\varphi_n$ in the sense that

$$\|f - \varphi_n\|_\mathcal{V} \to 0 \,.$$

Therefore we can define

$$I[f](\omega) = \int_s^t f(u, \omega) dB_u(\omega) = \lim_{n \to \infty} \int_s^t \varphi_n(u, \omega) dB_u(\omega) \,,$$

where by the Itô isometry, the limit exists in $L^2(\Omega, \mathbb{P})$ because

$$\left( \int \varphi_n(u, \omega) dB_u(\omega) \right)_{n \in \mathbb{N}}$$

is a Cauchy sequence in $L^2(\Omega, \mathbb{P})$; see p. 22.

**Definition 6.7.** *Let $f \in \mathcal{V} = \mathcal{V}(s, t)$. Then the* Itô integral *of $f$ is defined by*

$$\lim_{n \to \infty} \int_s^t \varphi_n(u, \omega) dB_u(\omega)$$

*where $(\varphi_n)_{n \in \mathbb{N}}$ is a sequence of simple functions with $\varphi_n \to f$ in $\mathcal{V}$, i.e.,*

$$\mathbb{E}\left[ \int_s^t (f(u, \omega) - \varphi(u, \omega))^2 \, du \right] \to 0 \text{ as } n \to \infty.$$

**Corollary 6.8.** (Itô isometry) *For all $f \in \mathcal{V} = \mathcal{V}(s, t)$, we have*

$$\mathbb{E}\left[ \left( \int_s^t f(u, \omega) dB_u(\omega) \right)^2 \right] = \mathbb{E}\left[ \int_s^t |f(u, \omega)|^2 \, du \right].$$

**Theorem 6.9.** *Let $f, g \in \mathcal{V}(0, t)$ and $0 \leq s \leq u \leq t$. Then (a.s.):*

*(i)*

$$\int_s^t f(\tau, \omega) dB_\tau = \int_s^u f(\tau, \omega) dB_\tau + \int_u^t f(\tau, \omega) dB_\tau \,.$$

*(ii)*

$$\int_s^t (\alpha f + \beta g dB_u = \alpha \int_s^t f dB_u + \beta \int_s^t g dB_u \quad \forall \alpha, \beta \in \mathbb{R}$$

*(iii)*

$$\mathbb{E}\left[\int_0^t f(s,\omega) dB_s\right] = 0$$

*(iv)*

$$\int_s^t f(u,\omega) dB_u$$

is $\mathcal{F}_t$-measurable.

*Proof.* Exercise. $\qquad\square$

## 6.3 Ornstein-Uhlenbeck process

**Example 6.10** (OU process)**.** *Consider the linear SDE for constants $A, B \in \mathbb{R}$*

$$dX_t = AX_t dt + B dW_t, \quad X_0 = x,$$

*which means*

$$X_t = x + A \int_0^t X_s ds + B \int_0^t dW_s.$$

*One can show that the solution to the linear SDE can be expressed using the variation-of-constants-formula*

$$X_t = e^{At} x + \int_0^t e^{A(t-s)} B dW_s.$$

*The solution $(X_t)_{t>0}$ is a Gaussian process, so it is completely specified by its mean and variance*

$$\mathbb{E}[X_t] = e^{At} x \text{ by property (iv) above}$$

$$\mathbb{E}\left[(X_t - \mathbb{E}[X_t])^2\right] = \mathbb{E}\left[\left(\int_0^t e^{A(t-s)} B dW s\right)^2\right]$$

$$= \mathbb{E}\left[\int_0^t \left(e^{A(t-s)} B\right)^2 ds\right] \text{ by Itô isometry}$$

$$= \int_0^t e^{2A(t-s)} B^2 \, ds$$

$$= \frac{B^2}{2A}\left(e^{2At} - 1\right)$$

**Remark 6.11.** *The main things to remember are that the approximation procedure for defining the Itô integral reduces to the Itô isometry for elementary functions $\varphi_n \to f$ (convergence in $\mathcal{V}$) and that the limiting integral $I[f]$ is in $L^2(\Omega, \mathbb{P})$. Specifically, we have proved that $I[\varphi_n] \to I[f]$ in $L^2(\Omega, \mathbb{P})$, i.e.,*

$$\mathbb{E}\left[(I[\varphi_n] - I[f])^2\right] = \int_\Omega (I[\varphi_n](\omega) - I[f](\omega))^2 \, d\mathbb{P}(\omega) \to 0$$

*as $n \to \infty$.*

# 7 Day 7, 27.11.2012: Itô calculus

**Recapitulation:** The Itô integral for functions $f \in \mathcal{V} \cong L^2(\Omega \times [0,T], \mathbb{P} \otimes \lambda)$ is defined by

$$I[f](\omega) = \int_0^T f(\omega, s) dB_s(\omega) = \lim_{n \to \infty} \sum_j e_j^n \left( B_{t_{j+1}^n} - B_{t_j^n} \right),$$

with convergence in $L^2(\Omega, \mathbb{P})$. Here the $(e_j^n)_{n,j \in \mathbb{N}}$ is a sequence of random variables that are measurable with respect to $\sigma(B_s \colon s \le t_j^n)$, and

$$(\varphi_n)_{n \in \mathbb{N}}, \quad \varphi_n(\omega, t) = \sum_j e_j^n(\omega) \chi_{[t_j^n, t_{j+1}^n)}(t)$$

is a sequence of simple functions such that $\|\varphi_n - f\|_{\mathcal{V}} \to 0$.

The Itô integral provides the solution to the stochastic differential equation

$$dX_t(\omega) = b(X_t(\omega), t) dt + \sigma(X_t(\omega), t) dB_t(\omega), \quad X_0 = x. \tag{7.1}$$

Specifically, assuming that $(X_t)_{t \ge 0}$ is adapted to the filtration generated by $B_t$, i.e., $X_t$ is measurable with respect to $\sigma(B_s \colon s \le t)$, we have

$$X_t = x + \int_0^t b(X_s, s)\, ds + \int_0^t \sigma(X_s, s)\, dB_s.$$

## 7.1 Functions of bounded and quadratic variation

What we now need is a theory of differentiation that is useful in solving equations such as (7.1) and which can explain properties of the Itô integral, such as

$$\int_0^t B_t dB_t = \frac{1}{2} B_t^2 - \frac{1}{2} t.$$

**Exercise 7.1.** *Prove the above equation.*

**Definition 7.2.** *Let $T > 0$. A sequence $(\Delta_n)_{n \in \mathbb{N}}$ of partitions of $[0,T]$, with*

$$\Delta_n = \{t_0^n, \ldots, t_{k^n}^n\} \subset [0,T], \quad 0 = t_0^n < t_1^n < \cdots < t_{k^n}^n = T$$

*is called a* refinement of partitions *of $[0,T]$ if the sequence satisfies*

$$\Delta_{n+1} \supset \Delta_n \quad \& \quad |\Delta_n| := \max_i |t_i^n - t_{i-1}^n| \to 0 \quad as\ n \to 0.$$

**Example 7.3.** *An example of a refinement of partitions is the sequence of dyadic partitions. Let $T = 1$, and define*

$$\Delta_n = \left\{ \frac{j}{2^n} \colon j = 0, 1, \ldots, 2^n - 1, 2^n \right\}.$$

**Definition 7.4.** *A function $f : [0,T] \to \mathbb{R}$ is of* bounded variation *(BV) if*

$$\sup_{n \in \mathbb{N}} |f(t_i^n) - f(t_{i-1}^n)| < \infty$$

*for all refinements of partitions $(\Delta_n)_{n \in \mathbb{N}}$.*

**Definition 7.5.** *A function $f : [0, T] \to \mathbb{R}$ is of* quadratic variation *(QV) if its quadratic variation*

$$\langle f \rangle_t := \sup_n \sum_{t_i^n \leq t} |f(t_i^n) - f(t_{i-1}^n)|^2 < \infty.$$

*is finite for every $t \in [0, T]$ and over all refinements of partitions.*

**Remark 7.6.** *We make some remarks which we will not prove, with the exception of (iii).*

(i) *Continuously differentiable functions are BV functions.*

(ii) *If one integrates against a BV function, the resulting Riemann-Stieltjes integral is independent of the refinement of partitions.*

(iii) *Continuous BV functions have zero QV:*

$$\begin{aligned}
0 \leq \langle f \rangle_t &= \sup_{n \in \mathbb{N}} \sum_{t_i^n \leq t} |f(t_i^n) - f(t_{i-1}^n)|^2 \\
&= \lim_{n \to \infty} \sum_{t_i^n \leq t} |f(t_i^n) - f(t_{i-1}^n)|^2 \\
&\leq \max_i |f(t_i^n) - f(t_{i-1}^n)| \lim_{n \to \infty} \sum_{t_i^n \leq t} |f(t_i^n) - f(t_{i-1}^n)| \\
&\leq C \lim_{n \to \infty} |f(t_i^n) - f(t_{i-1}^n)| \text{ (since } f \text{ is a BV function)} \\
&= 0 \quad \text{(by continuity of } f \text{) } .
\end{aligned}$$

(iv) *The quadratic variation of Brownian motion at time $t$ is equal to $t$:*

$$\langle B \rangle_t = t.$$

*Brownian motion is not of bounded variation.*

(v) *Given an interval $[0, T]$ for $T > 0$ and a function $f : [0, T] \to \mathbb{R}$ of QV, the quadratic variation $\langle f \rangle_t$ is a BV function of time, which follows from the fact that $\langle \cdot \rangle$ is monotonic as a function of time.*

## 7.2 Itô's formula

**Theorem 7.7** (Itô's formula I)**.** *Let $F \in C^{2,1}(\mathbb{R}, [0, T])$ and let $X = B \in C([0, T])$ be Brownian motion. Then*

$$F(X_t, t) = F(0, 0) + \int_0^t \frac{\partial F}{\partial x}(X_s, s) \, dX_s + \int_0^t \left( \frac{1}{2} \frac{\partial^2}{\partial x^2} + \frac{\partial}{\partial s} \right) F(X_s, s) \, ds \,.$$

*Proof.* For convenience, we will drop the time-dependence of $F$, so that $F(x, s) = F(x)$. By Taylor's theorem,

$$F(X_{t_i^n}) - F(X_{t_{i-1}^n}) = F'(X_{t_{i-1}^n})(X_{t_i^n} - X_{t_{i-1}^n}) + \frac{1}{2} F''(\xi_i^n)(X_{t_i^n} - X_{t_{i-1}^n})^2$$

for a number $\xi_i^n \in (X_{t_i^n}, X_{t_{i-1}^n})$. Then

$$F(X_t) - F(X_0) = \underbrace{\sum_{t_i^n \leq t} F'(X_{t_{i-1}^n})(X_{t_i^n} - X_{t_{i-1}^n})}_{I_n} + \underbrace{\frac{1}{2} \sum_{t_i^n \leq t} F''(\xi_i^n)(X_{t_i^n} - X_{t_{i-1}^n})^2}_{Q_n}.$$

We consider the two sums separately. As for the first sum, we observe that $I_n$ is a discrete version of the Itô integral, and therefore

$$\left\| I_n - \int_0^t \frac{\partial F}{\partial x}(X_s, s)\, dX_s \right\|_{L^2(\Omega, \mathbb{P})}^2 \to 0$$

as $n \to \infty$. Using (v) from the preceding remark, we know that the quadratic variation of $(X_t)_{t \in [0,T]}$ is itself a BV function of time. Therefore $Q_n$ converges to the standard Riemann-Stieltjes integral,

$$\frac{1}{2} \int_0^t F''(X_s)\, d\langle X \rangle_s = \frac{1}{2} \int_0^t F''(X_s)\, ds\,,$$

where, again, convergence is in $L^2(\Omega, \mathbb{P})$. $\qquad\square$

**Corollary 7.8** (Itô's Formula II). *Let $B = (B_t)_{t \geq 0}$ be d-dimensional Brownian motion and let $X = (X_t)_{t \geq 0}$ be the n-dimensional solution to the Itô stochastic differential equation*

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dB_t\,,$$

*where $b \colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$, $\sigma \colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^{n \times d}$. Let $F \in C^{2,1}(\mathbb{R}^n, [0, \infty))$. Then $Y_t := F(X_t, t)$ solves the Itô equation*

$$dY_t = \underbrace{\nabla_x F(X_t, t) \cdot dX_t + \frac{\partial F}{\partial t}(X_t, t)dt}_{BV\ part\ (by\ chain\ rule)} + \underbrace{\frac{1}{2} dX_t \cdot \nabla_x^2 F(X_t, t) dX_t}_{QV\ part}$$

$$= \left( \frac{\partial F}{\partial t} + \sum_{i=1}^n \frac{\partial F}{\partial x_i} b_i + \frac{1}{2}\left(\sigma\sigma^\top : \nabla_x^2 F\right) \right)(X_t, t)\, dt + \left(\sigma^\top \nabla_x F\right)(X_t, t) \cdot dB_t$$

*where $A \colon B = (A^T B)$ denotes the matrix inner product, and we have obtained the last equation by substituting $dX_t = b(X_t, t)dt + \sigma(X_t, t)dB_t$, using the rules*

$$dtdt = dtdB_t^i = dB_t^j dt = 0 \quad and \quad dB_t^i dB_t^j = \delta_{ij} dt \quad (i, j = 1, \ldots, d)\,,$$

*where $B_t^i$ denotes the i-th component of $B_t$.*

**Remark 7.9.** *The matrix family $a(\cdot, \cdot) := \sigma\sigma^\top \colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^{n \times n}$ is sometimes called the diffusion matrix.*

**Remark 7.10.** *Note that, for functions that have a quadratic variation, Itô's formula is what is chain rule for functions of bounded variation. Dropping the dependence on $(X_t, t)$ for the moment, one may rewrite the last equation as*

$$dY_t = \left( \frac{\partial F}{\partial t} + \sum_{i=1}^n \frac{\partial F}{\partial x_i} b_i + \frac{1}{2} \sum_{i,j=1}^n a_{ij} \frac{\partial^2 F}{\partial x_i \partial x_j} \right) dt + \sum_{i=1}^d \left( \sum_{j=1}^n \sigma_{ij} \frac{\partial F}{\partial x_j} \right) dB_t^i\,.$$

**Historical remarks**

Itô's original work was published in 1951.[4] However it was recently revealed that in 1940 Wolfgang Döblin, brother of novelist Alfred Döblin, French-German mathematician and student of Maurice Fréchet and Paul Lévy, sent a sealed letter to the Académie Française, while he was on the German front with the French army (as a telephone operator). Döblin committed suicide before he was captured by the German troops and burned all his mathematical notes. According to Döblin's last will, the letter to the Académie Française was opened in the year 2000 and found to contain a proof of Itô's lemma.[5]

## 7.3 Geometric Brownian motion

Consider the *Geometric Brownian motion* $S = (S_t)_{t \geq 0}$ that is the solution of the Itô stochastic differential equation (SDE)

$$dS_t = \mu S_t dt + \sigma S_t dB_t, \quad S_0 > 0.$$

We claim that the solution to the SDE is

$$S_t = S_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma B_t\right].$$

This can be seen as follows: using Itô's formula for $F(x) = \log x$, we find

$$Y_t = \log S_t$$
$$\Rightarrow dY_t = \frac{dS_t}{S_t} - \frac{\sigma^2}{2} \frac{S_t^2}{S_t^2} dt$$
$$= \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dB_t$$

and therefore

$$Y_t = Y_0 + \int_0^t \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma \int_0^t dB_t$$
$$= Y_0 + \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma B_t$$

which proves that $S_t$ follows the log-normal distribution with mean $\left(\mu - \frac{\sigma^2}{2}\right) t$ and variance $\sigma^2$. Moreover,

$$\mathbb{E}[S_t] = \exp(\mu t).$$

**Remark 7.11.** *The geometric Brownian motion is sometimes used to model the growth of one's wealth subject to some positive interest rate $\mu > 0$ and random fluctuations due to market conditions, represented by the volatility-modified Brownian motion term $\sigma B_t$. Good to know that $\mathbb{E}[S_t] = \exp(\mu t)$.*

---

[4]Kiyoshi Itô, 1915–2008, Japanese Mathematician; the famous lemma appeared in [K. Itô, On stochastic differential equations, *Memoirs AMS* **4**, 1–51, 1951].

[5]For a summary of Döblin's work, see [B. Bru and M. Yor, Comments on the life and mathematical legacy of Wolfgang Doeblin, Finance Stochast. **6**, 4–47, 2002]
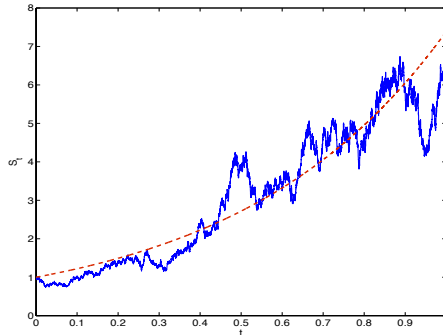
Figure 7.1: Typical realization of Geometric Brownian Motion $(S_t)_{t\in[0,1]}$ for $\mu = 2$ and $\sigma = 1$. The red dashed line shows the mean $\mathbb{E}[S_t]$.

*It is also known, however, that the Brownian motion satisfies the Law of the Iterated Logarithm (see, e.g., [15, Thm. 5.1.2])*

$$\limsup_{t\to\infty} \frac{B_t}{\sqrt{2t\log\log t}} = +1$$

$$\liminf_{t\to\infty} \frac{B_t}{\sqrt{2t\log\log t}} = -1\,,$$

*which states that Brownian grows sublinearly. Since*

$$S_t = S_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right]$$

*it follows that depending on different values of $\mu$ and $\sigma$, the wealth process $(S_t)_{t\geq 0}$ is dominated by the linear drift term. Indeed S, can exhibit rather different behaviours in the limit as $t \to \infty$, depending on the values of $\mu$ and $\sigma$:*

*(a) If $\mu < \frac{\sigma^2}{2}$, then $S_t \to 0$ as $t \to \infty$.*

*(b) If $\mu = \frac{\sigma^2}{2}$, then $\limsup_{t\to\infty} S_t = \infty$, $\liminf_{t\to\infty} S_t = 0$.*

*(c) If $\mu > \frac{\sigma^2}{2}$, then $S_t \nearrow \infty$ as $t \to \infty$.*

*(All the statements above hold $\mathbb{P}$-almost surely.) The mind-blowing aspect of geometric Brownian motion is the seemingly contradictory property that, even though the expected value grows exponentially with time for every volatility value $\sigma$, the process will hit zero with probability 1 whenever the volatility is sufficiently large, i.e. when $\sigma > \sqrt{2\mu}$. That is, even though the expected wealth grows exponentially (and thus never hits zero), for $\mathbb{P}$-almost all $\omega$, all the wealth will vanish due to fluctuations in the long time limit, i.e., every single market player goes broke with probability one. Think about it!*

# 8 Day 8, 04.12.2012: Itô stochastic differential equations

## 8.1 Martingale representation theorem

Let us repeat Itô's Lemma: given $F \in C^2(\mathbb{R}^n)$ and an Itô SDE

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dB_t,$$

the process $(Y_t)_{t \geq 0}$, $Y_t := F(X_t)$ solves the Itô SDE

$$dY_t = \nabla F(X_t) \cdot dX_t + \frac{1}{2}dX_t \cdot \nabla^2 F(X_t)dX_t$$

$$= \left( \nabla F \cdot b + \frac{1}{2}\sigma\sigma^\top : \nabla^2 F \right) dt + \left( \sigma^\top \nabla F \right) \cdot dB_t$$

where $u \cdot v = u^\top v$ denotes the usual inner product between vectors and $A : B = \mathrm{tr}(A^\top B)$ is the inner product between matrices. Furthermore, in the step from the first to the second line, we have used the rule that

$$dtdt = dtdB_t^i = dB_t^i dt = 0 \quad dB_t^i dB_t^j = \delta_{ij} dt, \quad i, j = 1, \ldots, n.$$

If we define the second-order differential operator (the infinitesimal generator of the stochastic "flow" $X_t$ that will be introduced later on)

$$L\varphi = \frac{1}{2}\sigma\sigma^\top : \nabla^2 \varphi + b \cdot \nabla \varphi,$$

Itô's formula may be rewritten as

$$dY_t = (LF)(X_t, t)dt + \left( \sigma^\top \nabla F \right)(X_t, t) \cdot dB_t.$$

**Remark 8.1.** *The fact that the usual chain does not apply for Itô processes has to do with the definition of the corresponding stochastic integral. We shall briefly comment on what makes Itô integral special.*

*(i) The* Martingale Representation Theorem *states that every $\mathcal{F}_t$-martingale $(X_t)_{t \geq 0}$ (i.e., $X_t$ is adapted to the filtration generated by $B_t$ and satisfies $X_s = \mathbb{E}[X_t | \mathcal{F}_s]$ ) can be written as the integral*

$$X_t = X_0 + \int_0^t \phi(\omega, s)dB_s$$

*for a function $\phi \in \mathcal{V}(0, t)$, that is uniquely determined. Conversely, every Itô integral of the form*

$$\int_0^t \phi_s dB_s$$

*is a martingale with respect to $(\mathcal{F}_t)_{t \geq 0}$*

*(ii) The* Stratonovich integral *is another stochastic integral, distinct from the Itô integral, that is based on the midpoint rule, i.e.,*

$$\int \psi(\omega, s) \circ dB_s = \lim_{n \to \infty} \sum_j \psi\left( \omega, t_{j+\frac{1}{2}} \right) \left( B_{t_{j+1}} - B_{t_j} \right)$$

*where we emphasize the different notation using the "∘" symbol and where*

$$t_{j+\frac{1}{2}} := \frac{t_j + t_{j+1}}{2}.$$

*The Stratonovich integral has the property that*

$$\mathbb{E}\left[\int \psi_s \circ dB_s\right] \neq 0\,,$$

*hence the Stratonovich integral is not a martingale, unlike the Itô integral. Furthermore, the thus defined integral does not satisfy the Itô Isometry. On the other hand, the usual chain rule applies.*

*(iii) The Stratonovich integral is used for integrating* Stratonovich SDEs

$$dX_t = b(X_t, t)dt + \sigma(X_t, t) \circ dB_t\,,$$

*where, again, the "∘" indicates that the SDE has to be interpreted in the Stratonovich sense (i.e., integrated using the Stratonovich integral). One can also convert Stratonovich SDEs to Itô SDEs using the conversion rule*

$$dX_t = b(X_t, t)dt + \sigma(X_t, t) \circ dB_t = \left(b + \frac{1}{2}\sigma\nabla\sigma\right)(X_t, t)dt + \sigma(X_t, t)dB_t.$$

## 8.2 Stochastic differential equations: Existence and uniqueness

We now want to find possible solutions $(X_t)_{t\geq 0} \subset \mathbb{R}^n$ for Itô SDEs of the form

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dB_t. \tag{8.1}$$

where $b: \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ and $\sigma: \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^{n\times d}$ are measurable functions.

**Definition 8.2** (Strong solution). *Let $T > 0$. A process $(X_t)_{t\in[0,T]}$ is called a* strong solution *of (8.1) if the map $t \mapsto X_t$ is almost surely continuous and adapted to the filtration generated by the Brownian motion $(B_t)_{t\in[0,T]}$ and if it holds for $\mathbb{P}$-almost all $\omega$ (i.e., if it holds $\mathbb{P}$-almost surely) that*

$$X_t(\omega) = X_0(\omega) + \int_0^t b(X_s(\omega), s)ds + \int_0^t \sigma(X_s, \omega)dB_s(\omega) \quad (\omega \text{ fixed})\,.$$

**Definition 8.3** (Uniqueness). *The solution of (8.1) is called* unique *or* pathwise unique *if*
$$\mathbb{P}(X_0 = \tilde{X}_0) \quad \Rightarrow \quad \mathbb{P}(X_t = \tilde{X}_t) \; \forall t \in [0, T]$$
*for any two solutions $(X_t)_{t\in[0,T]}$ and $(\tilde{X}_t)_{t\in[0,T]}$ of (8.1).*

**Theorem 8.4** (Existence and uniqueness). *Let $T > 0$ and $b, \sigma$ in (8.1) satisfy*

*(i) (Global Lipschitz condition): $\forall x, y \in \mathbb{R}^n$, $\forall t \in [0, T]$,*

$$|b(x, t) - b(y, t)| + |\sigma(x, t) - \sigma(y, t)| \leq L|x - y|$$

*for some constant $0 < L < \infty$.[6]*

---

[6] Here the norm on the matrix terms $\sigma(\cdot, \cdot)$ is arbitrary and may be taken to be, e.g., the Frobenius norm $|\sigma| = (\sum_{i,j}|\sigma_{ij}|^2)^{1/2}$.

(ii) (Sublinear growth condition): $\forall x, \in \mathbb{R}^n$ and $\forall t \in [0, T]$,

$$|b(x, t)| + |\sigma(x, t)| \leq G(1 + |x|)$$

for some $0 < G < \infty$.

Given that the above conditions hold, if we have $\mathbb{E}\left[X_0^2\right] < \infty$, then (8.1) has a pathwise unique, strong solution for any $T > 0$.

*Proof.* See [15, Thm 5.2.1]. The main elements of the proof of the above theorem are the Itô Isometry and a Picard-Lindelöf-like fixed-point iteration, just as in case of ordinary differential equations. $\qquad \square$

## 8.3 Applications from physics, biology and finance

**Example 8.5.** *We now consider some examples which use Itô's formula.*

(i) (Geometric Brownian motion): *the Itô SDE in this example is*

$$dS_t = \mu S_t dt + \sigma S_t dB_t, \quad S_0 > 0$$

*and the solution to this SDE is*

$$S_t = S_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma B_t\right]$$

(ii) (Ornstein-Uhlenbeck process): *the OU process is a Gaussian process whose evolution is given by*

$$dX_t = AX_t dt + B dW_t, \quad X_0 = x_0$$

*for $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times d}$ (i.e., $W_t$ is d-dimensional Brownian motion). The solution to the above SDE is*

$$X_t = e^{At} x_0 + \int_0^t e^{A(t-s)} B \, dW_s$$

(iii) (Brownian Bridge): *the Brownian bridge $(BB_t)_{t \in [0,1]}$ is a Gaussian process with the property $BB_0 = BB_1 = 0$. The associated SDE is*

$$dBB_t = \frac{1 - BB_t}{1 - t} dt + dB_t, \quad BB_0 = 0.$$

*The solution to the above SDE is*

$$BB_t = (1 - t) \int_0^t \frac{dB_s}{1 - s}.$$

*The proof is left as an exercise.*

(iv) (Logistic growth model): *Consider the ODE*
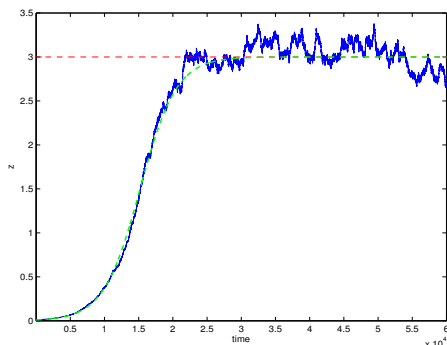
$$\frac{dz}{dt} = rz(C - z), \quad z(0) > 0$$

Figure 8.1: Typical realization of the logistic growth model with $r = 0.5$, $C = 3$ and $\sigma = 0.1$; for comparison, the dashed green curve shows the deterministic model with $\sigma = 0$; the red straight line shows the capacity bound $C$.

*that describes logistic growth of a population where $r > 0$ is the (initial) growth rate and $C > 0$ is the capacity bound. If we add random perturbations we obtain the Itô SDE*

$$dZ_t = rZ_t(C - Z_t)dt + \sigma Z_t dB_t, \quad Z_0 > 0.$$

*which describes logistic growth in a random environment (see Figure 8.1). This is an interesting example because the drift coefficient is not globally Lipschitz—even worse, the drift term can become unbounded. Despite this, one can obtain an analytic solution to the SDE above, given by*

$$\hat{Z}_t = \frac{\exp\left[\left(rC - \frac{\sigma^2}{2}\right)t + \sigma B_t\right]}{\hat{Z}_0^{-1} + \int_0^t \exp\left[\left(rC - \frac{\sigma^2}{2}\right)s + \sigma B_s\right]ds}.$$

**Central issues for numerical methods for SDEs**

We now consider some key properties which we shall use in evaluating the quality of a numerical method for solving Itô SDEs of the form

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\epsilon}dB_t \tag{8.2}$$

with a smooth function $V \colon \mathbb{R} \to \mathbb{R}$ (8.1) that will not meet the requirements of the existence and uniqueness theorem in general.

(i) There are various choices for stable numerical schemes for solving equations like (8.2). As we will see below, one such choice is

$$\tilde{X}_{n+1} - \tilde{X}_n = -\Delta t \nabla(\tilde{X}_n) + \sqrt{2\epsilon\Delta t}\,\xi_{n+1},$$

where $\Delta t > 0$ and the $\xi_n$ are suitable i.i.d. random variables, e.g., standard normal or uniform on the set $\{-1, 1\}$, such that $\tilde{X}_n \approx X(t_n)$ on a sufficiently fine grid $0 = t_0 < t_1 < t_3 < \dots$ with $\Delta t = t_{n+1} - t_n$.
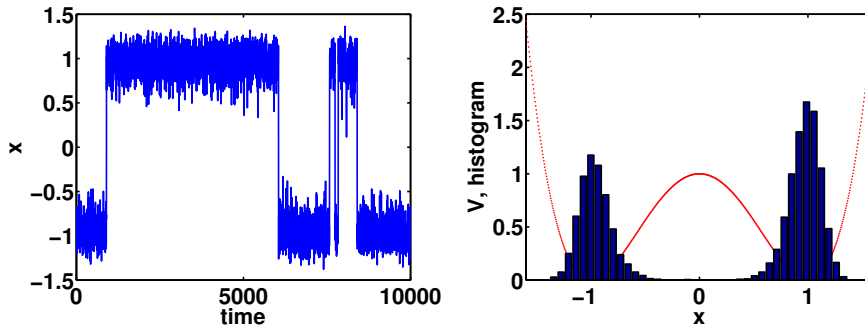
Figure 8.2: Typical realization of (8.2) with a bistable potential $V$; the solution has been computed using the Euler method (see the next section).

(ii) The numerical scheme under (i) can be shown to yield an approximation to the continuous SDE on any *finite* time interval, but diverges when $n \to \infty$. On the other hand, we may be interested in the limiting behaviour, specifically in the stationary distribution of the process (if it exists). In our case, (an under certain technical assumption on $V$), the process satisfies

$$(\mathbb{P} \circ X_t^{-1})(A) \to \int_A e^{-V/\epsilon} dx$$

as $t \to \infty$ and for all Borel sets $A \subset \mathbb{R}$ (assuming that the integral on the right hand side is properly normalized). For the continuous process convergence of the distribution can be shown to hold in $L^1$, but it may be very slow if $\epsilon$ in (8.2) is small. For the discrete approximation this question of convergence does not have an easy answer, for the standard numerical schemes are not asymptotically stable and the numerical discretization introduces a bias in the stationary distribution (see Figure 8.2).

(iii) Can we compute functionals of paths of $X_t$? For example, can we compute quantities, such as

$$\mathbb{E}\left[\phi(X_T)\right], \quad \mathbb{E}\left[\int_0^T \psi(X_t, t) dt\right]$$

for bounded continuous functions $\phi$, $\psi$ and $T > 0$, or can we compute

$$\mathbb{E}[\tau | X_0 = x]$$

with $\tau$ being some random stopping time (e.g., a first hitting time of a set $E \subset \mathbb{R}$). Questions dealing with such functionals, but also the long-term stability issue under (ii) will lead us to Markov Chain Monte-Carlo (MCMC) methods for PDEs and the celebrated Feynman-Kac formula.

# 9 Day 9, 11.12.2012: Stochastic Euler method

**Suggested reference:** [9]

We motivate the ideas in this method by considering the deterministic initial value problem

$$\frac{dx}{dt} = b(x,t), \quad x(0) = x_0,$$

for $t \in [0, T]$. The initial value problem has the solution

$$x(t) = x_0 + \int_0^t b(x(s), s)ds, \quad t \in [0, T]$$

and we can approximate the true solution, employing a suitable quadrature rule for the integral, e.g., the "rectangle rule":

$$x(t_{n+1}) = x(t_n) + \int_{t_n}^{t_{n+1}} b(x(s), s)ds$$

$$\approx x(t_n) + \int_{t_n}^{t_{n+1}} b(x(t_n), t_n)ds$$

$$= x(t_n) + b(x(t_n), t_n)(t_{n+1} - t_n).$$

Given a sufficiently fine grid of time nodes $0 = t_0 < t_1 < \ldots < t_N = T$ with fixed time step $\Delta t = t_{n+1} - t_n$, we recognize the forward Euler method

$$x_{n+1} = x_n + \Delta t b(x_n, t_n).$$

The discretization error induced by the Euler scheme can be shown to satisfy

$$\sup_{n=1,\ldots,N} |x(t_n) - x_n| \le C\Delta t.$$

for a $0 < C < \infty$ that is independent of $\Delta t$.

## 9.1 The Euler-Maruyama scheme

We construct a simple quadrature rule for the Itô integral. To this end let $(X_t)_{t \in [0,T]}$ solve the Itô SDE

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dB_t, \quad X_0 = x$$

with $b : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^{n \times m}$ for $(B_t)_{t>0}$ a $m$-dimensional Brownian motion. We wish to approximate $(X_t)_{t>0}$ on the uniform grid

$$\{0 = t_0 < t_1 < \ldots < t_N = T\}, \quad \Delta t = t_{n+1} - t_n.$$

The "rectangle rule" for the solution between $[t_n, t_{n+1}] \subset [0, T]$ is

$$X_{t_{n+1}} = X_{t_n} + \int_{t_n}^{t_{n+1}} b(X_s, s)ds + \int_{t_n}^{t_{n+1}} \sigma(X_s, s)dB_s$$

$$\approx X_{t_n} + \int_{t_n}^{t_{n+1}} b(X_{t_n}, t_n)ds + \int_{t_n}^{t_{n+1}} \sigma(X_{t_n}, t_n)dB_s$$

$$= X_{t_n} + \Delta t b(X_{t_n}, t_n) + \underbrace{(B_{t_{n+1}} - B_{t_n})}_{=:\Delta B_n \sim N(0, \Delta t)} \sigma(X_{t_n}, t_n).$$

**Definition 9.1.** (Euler-Maruyama scheme)*: For $n = 0, \dots, N - 1$, the* Euler-Maruyama scheme *or* Euler's method *gives the n-th iterate as*

$$\tilde{X}_{n+1} = \tilde{X}_n + \Delta t b(\tilde{X}_n, t_n) + \sigma(\tilde{X}_n, t_n) \Delta B_n, \quad \tilde{X}_0 = x$$

**Remark 9.2.** *A few remarks are in order.*

(i) *Euler's method is consistent with the definition of the Itô integral, in that is evaluates the integrand at the left endpoint of the interval.*

(ii) *The Euler method gives the values of the numerical path at the time nodes. A numerical path is obtained by linear interpolation: for $t \in [t_n, t_{n+1}]$,*

$$\tilde{X}_t(\omega) = \tilde{X}_n(\omega) + \frac{(t - t_n)}{\Delta t} \left( \tilde{X}_{n+1}(\omega) - \tilde{X}_n(\omega) \right)$$
$$= \tilde{X}_n + (t - t_n) b(\tilde{X}_n, t_n) + \frac{t - t_n}{\Delta t} \sigma(\tilde{X}_n, t_n) \left( B_{t_{n+1}} - B_{t_n} \right).$$

*Note that $\tilde{X}_t$, $t \leq t_{n+1}$ depends on $B_{t_{n+1}}$, i.e., the interpolant $\tilde{X}_t$ is not non-anticipating.*

(iii) *Sometimes one wishes to refine the partition for a specific realization, i.e. using the same path $\omega$. For example, by halving the time step from from $\Delta t$ to $\Delta t/2$, one obtains new grid points*

$$t_{n+\frac{1}{2}} = t_n + \frac{\Delta t}{2}$$

*The values of the refined Brownian motion can be computed by the rule*

$$B_{t_{n+\frac{1}{2}}}(\omega) = \frac{1}{2} \left[ B_{t_n}(\omega) + B_{t_{n+1}}(\omega) \right] + \frac{1}{2} \sqrt{\Delta t} \, \xi_n(\omega)$$

*where the $\xi_n \sim N(0, 1)$ are i.i.d. standard normal random variables.*

## 9.2 Convergence of Euler's method

We can guess that $\tilde{X} \approx X(t_n)$, but in which sense?

**Example 9.3.** *Let $X_t = B_t$ and $Y_t = -B_t$; then $X_t \sim Y_t$ (that is, $(X_t)_{t>0}$ and $(Y_t)_{t>0}$ have the same distribution) but $|X_t - Y_t| = 2|B_t|$ is unbounded for all $t$. Hence pathwise comparisons may not be very informative.*

**Definition 9.4.** (Strong convergence)*: Let $X_{t_n}$ denotes the value of the true solution $(X_t)_{t \in [0,T]}$ of our SDE at the time $t_n$. A numerical scheme $(\tilde{X}_n)_n = (\tilde{X}_n^{\Delta t})_{n=0,\dots,N-1}$ is called* strongly convergent *of order $\gamma > 0$ if*

$$\max_{n=0,\dots,N-1} \mathbb{E}[|\tilde{X}_n - X_{t_n}|] \leq C \Delta t^\gamma$$

*where $0 < C < \infty$ is independent of $\Delta t$ but can depend on the length $T = N \Delta t$ of the time interval.*

**Definition 9.5.** (Weak convergence): *A numerical scheme* $(\tilde{X}_n)_n$ *is called weakly convergent of order* $\delta > 0$ *if*

$$\max_{n=0,\ldots,N-1} \left| \mathbb{E}[f(\tilde{X}_n)] - \mathbb{E}[f(X_{t_n})] \right| \leq D \Delta t^\delta$$

*for all functions* $f$ *in a suitably chosen class of functions (e.g. the space of continuous, bounded functions* $C_b(\mathbb{R}^n$, *or the space of polynomials of degree* $k$*). The constant* $D$ *is independent of* $\Delta t$, *but may depend function being considered.*

**Remark 9.6.** *A mnemonic for the difference between strong and weak convergence is that strong convergence is about the mean of the error, while weak convergence is about the error of the mean.*

**Lemma 9.7.** *Let* $f$ *be globally Lipschitz. Then strong convergence implies weak convergence. The converse does not hold*

*Proof.* Since $f$ is globally Lipschitz, $\exists\, 0 < L < \infty$ such that

$$|f(x) - f(y)| \leq L|x - y| \quad \forall x, y \in \mathbb{R}^n.$$

Then

$$
\begin{aligned}
\left| \mathbb{E}[f(\tilde{X}_n)] - \mathbb{E}[f(X_{t_n})] \right| &= \left| \mathbb{E}[f(\tilde{X}_n) - f(X_{t_n})] \right| \\
&\leq \mathbb{E}[|f(\tilde{X}_n) - f(X_{t_n})|] \\
&\leq L \mathbb{E}[|\tilde{X}_n - X_{t_n}|].
\end{aligned}
$$

Now consider the converse statement. Let $\tilde{X}_n = -X_{t_n}$, with $\mathbb{E}[\tilde{X}_n] = 0$ and $\mathbb{E}[|\tilde{X}_n|] \neq 0$. Then

$$|\mathbb{E}[\tilde{X}_n - \mathbb{E}[X_{t_n}]| = |0 - 0| = 0\,,$$

but

$$\mathbb{E}[|\tilde{X}_n - X_{t_n}|] = 2\mathbb{E}[|\tilde{X}_n|] \neq 0\,,$$

which concludes the proof. $\qquad\square$

The next theorem states that Euler's method is strongly (and hence weakly) convergent.

**Theorem 9.8.** *Let* $T > 0$ *and* $b, \sigma$ *in (8.1) satisfy*

(i) *(Global Lipschitz condition):* $\forall x, y \in \mathbb{R}^n$, $\forall t \in [0, T]$,

$$|b(x,t) - b(y,t)| + |\sigma(x,t) - \sigma(y,t)| \leq L|x - y|$$

*for some constant* $0 < L < \infty$.

(ii) *(Growth condition):* $\forall x, \in \mathbb{R}^n$ *and* $\forall t \in [0, T]$,

$$|b(x,t)|^2 + |\sigma(x,t)|^2 \leq G(1 + |x|^2)$$

*for some* $0 < G < \infty$.

*Then Euler's method is strongly convergent of order* $\gamma = 1/2$ *and weakly convergent of order* $\delta = 1$.

The proof is essentially based on the integral version of Gronwall's lemma (as is common when Lipschitz conditions are involved):

**Lemma 9.9** (Gronwall Lemma). *Let $y\colon [0,T] \to \mathbb{R}$ be non-negative and integrable such that*

$$y(t) \le A + B \int_0^t y(s)\,ds\,, \quad 0 \le t \le T\,,$$

*for some constants $A, B > 0$. Then*

$$y(t) \le Ae^{Bt}\,, \quad 0 \le t \le T\,.$$

Now we can prove that the Euler-Maruyama scheme convergences.

*Proof of Theorem 9.8.* We will show only strong convergence and leave the weak convergence part as an exercise. Without loss of generality, we assume that $b(x,t) = b(x)$ and $\sigma(x,t) = \sigma(x)$ are independent of $t$ and that $x \in \mathbb{R}$ is scalar—this will greatly simplify the notation. Since $L^2(\Omega, P) \subset L^1(\Omega, P)$, i.e.,

$$\mathbb{E}[|\tilde{X}_n - X_{t_n}|] \le \sqrt{\mathbb{E}[|\tilde{X}_n - X_{t_n}|^2]}\,,$$

it suffices to prove that

$$\mathbb{E}[|\tilde{X}_n - X_{t_n}|^2] \le C^2 \Delta t$$

for sufficiently small $\Delta t$.

Now let $\tau \in [0,T)$ and define $n_\tau \in \mathbb{N}$ by $\tau \in [t_{n_\tau}, t_{n_\tau+1})$ with $t_k = k\Delta t$. Further let $\bar{X}_\tau = \tilde{X}_{n_\tau}$ be the piecewise constant interpolant of the time-discrete Markov chain $\tilde{X}_0, \tilde{X}_1, \tilde{X}_2, \ldots$. Then

$$
\begin{aligned}
\bar{X}_\tau - X_\tau &= \tilde{X}_{n_\tau} - \left( x + \int_0^\tau b(X_s)\,ds + \int_0^\tau \sigma(X_s)\,dB_s \right) \\
&= \sum_{i=0}^{n_\tau - 1} (\tilde{X}_{i+1} - \tilde{X}_i) - \int_0^\tau b(X_s)\,ds - \int_0^\tau \sigma(X_s)\,dB_s \\
&= \sum_{i=0}^{n_\tau - 1} (b(\tilde{X}_i)\Delta t + \sigma(\tilde{X}_i)\Delta B_{i+1}) - \int_0^\tau b(X_s)\,ds - \int_0^\tau \sigma(X_s)\,dB_s \\
&= \int_0^{t_{n_\tau}} b(\bar{X}_s)\,ds + \int_0^{t_{n_\tau}} \sigma(\bar{X}_s)\,dB_s - \int_0^\tau b(X_s)\,ds - \int_0^\tau \sigma(X_s)\,dB_s \\
&= \underbrace{\int_0^{t_{n_\tau}} (b(\bar{X}_s) - b(X_s))\,ds + \int_0^{t_{n_\tau}} (\sigma(\bar{X}_s) - \sigma(X_s))\,dB_s}_{\text{discretization error}} \\
&\quad - \underbrace{\left( \int_{t_{n_\tau}}^\tau b(X_s)\,ds + \int_{t_{n_\tau}}^\tau \sigma(X_s)\,dB_s \right)}_{\text{interpolation error}}
\end{aligned}
$$

Squaring both sides of the equality and taking the expectation, it follows with the inequality $(a + b + c + d)^2 \le 4(a^2 + b^2 + c^2 + d^2)$,

$$\mathbb{E}[|\bar{X}_\tau - X_\tau|^2] \le 4\mathbb{E}\left[\left(\int_0^{t_{n_\tau}} (b(\bar{X}_s) - b(X_s))\, ds\right)^2\right]$$

$$+ 4\mathbb{E}\left[\left(\int_0^{t_{n_\tau}} (\sigma(\bar{X}_s) - \sigma(X_s))\, dB_s\right)^2\right]$$

$$+ 4\mathbb{E}\left[\left(\int_{t_{n_\tau}}^\tau b(X_s)\, ds\right)^2\right] + 4\mathbb{E}\left[\left(\int_{t_{n_\tau}}^\tau \sigma(X_s)\, dB_s\right)^2\right]$$

We will now estimate the right hand side of the inequality term by term, using Lipschitz and growth conditions:

1. Noting the inner Riemann integral can be interpreted as a scalar product between the functions $g(s) = 1$ and $f(s) = b(\bar{X}_s) - b(X_s)$, we find

$$\mathbb{E}\left[\left(\int_0^{t_{n_\tau}} (b(\bar{X}_s) - b(X_s))\, ds\right)^2\right] \underbrace{\le t_{n_\tau} \int_0^{t_{n_\tau}} \mathbb{E}[|b(\bar{X}_s) - b(X_s)|^2]\, ds}_{\text{Cauchy-Schwarz \& Fubini}}$$

$$\underbrace{\le T L^2 \int_0^{t_{n_\tau}} \mathbb{E}[|\bar{X}_s - X_s|^2]\, ds}_{\text{Lipschitz bound}} \,.$$

2. For the stochastic integral, the Itô isometry implies

$$\mathbb{E}\left[\left(\int_0^{t_{n_\tau}} (\sigma(\bar{X}_s) - \sigma(X_s))\, dB_s\right)^2\right] \underbrace{= \int_0^{t_{n_\tau}} \mathbb{E}[|\sigma(\bar{X}_s) - \sigma(X_s)|^2]\, ds}_{\text{Itô isometry}}$$

$$\underbrace{\le L^2 \int_0^{t_{n_\tau}} \mathbb{E}[|\bar{X}_s - X_s|^2]\, ds}_{\text{Lipschitz bound}} \,.$$

3. For the interpolation error coming from the drift, we have

$$\mathbb{E}\left[\left(\int_{t_{n_\tau}}^\tau b(X_s)\, ds\right)^2\right] \underbrace{\le (\tau - t_{n_\tau}) \int_{t_{n_\tau}}^\tau \mathbb{E}[|b(X_s)|^2]\, ds}_{\text{Cauchy-Schwarz \& Fubini}}$$

$$\underbrace{\le \Delta t\, G \int_{t_{n_\tau}}^\tau (1 + \mathbb{E}[|X_s|^2])\, ds}_{\text{sublinear growth}}$$

$$\le M_1 (\Delta t)^2 \,,$$

for a constant $0 < M_1 < \infty$. In the last step we have used that $\mathbb{E}[|X_s|^2]$ is finite by the assumptions on the coefficients $b$ and $\sigma$.

4. Finally, using the Itô isometry again, we can bound the remaining stochastic integral by

$$\mathbb{E}\left[\left(\int_{t_{n_\tau}}^{\tau} \sigma(X_s)\, dB_s\right)^2\right] = \underbrace{\int_{t_{n_\tau}}^{\tau} \mathbb{E}[|\sigma(X_s)|^2]\, ds}_{\text{Itô isometry}}$$

$$\underbrace{\leq G \int_{t_{n_\tau}}^{\tau} (1 + \mathbb{E}[|X_s|^2])\, ds}_{\text{sublinear growth}}$$

$$\underbrace{\leq M_2 \Delta t}_{\mathbb{E}[|X_s|^2] < \infty},$$

for a constant $0 < M_2 < \infty$.

Setting $y(t) = \mathbb{E}[|\bar{X}_t - X_t|^2]$, the assertion follows from Gronwall's lemma with $A = M\Delta t$ for a $M > (M_1\Delta t + M_2)$ and $B = L^2(1 + T)$. $\qquad\square$

## 9.3 Some practical issues

**Remark 9.10.** *Note that the (strong) error bound for Euler's method grows exponentially with $T$, hence becomes essentially of order one if $T = \mathcal{O}(-\log \Delta t)$.*

**Remark 9.11.** *We shall briefly comment on some implementation issues.*

*(i) The standard implementation of Euler's method is*

$$\tilde{X}_{n+1} = \tilde{X}_n + \Delta t b(\tilde{X}_n, t_n) + \sqrt{\Delta t}\sigma(\tilde{X}_n, t_n)\xi_{n+1}$$

*where the $\xi_n$ are standard normal, i.i.d. random variables.*

*(ii) The simplified Euler method uses i.i.d. $\xi'_n \sim U(\{\pm 1\})$ random variables, i.e. $\mathbb{P}(\xi'_n = 1) = \mathbb{P}(\xi'_n = -1) = 1/2$. The advantage of the simplified Euler scheme is that it is much faster to generate $U(\{\pm 1\})$ random variables than $N(0, 1)$ random variables. The disadvantage of the simplified Euler scheme is that it gives only weak convergence.*

**Exercise 9.12.** *Show that the simplified Euler method is weakly convergent.*

# 10 Day 10, 18.12.2012: Kolmogorov backward equation

**Suggested references:** $[1, 9, 15]$

## 10.1 Further remarks on the Euler-Maruyama scheme

For an inhomogeneous Itô SDE of the form

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dB_t, \quad X_0 = x, \quad t \in [0, T],$$

with $b\colon \mathbb{R}^n \times [0,T] \to \mathbb{R}^n$ and $\sigma\colon \mathbb{R}^n \times [0,T] \to \mathbb{R}^{n\times m}$, Euler's method reads

$$\tilde{X}_{n+1} = \tilde{X}_n + \Delta t\, b(\tilde{X}_n, t_n) + \sqrt{2\Delta t}\,\sigma(\tilde{X}_n, t_n)\,\xi_{n+1} \quad (n = 0, \dots, N-1)$$

where $0 = t_0 < t_1 < t_2 \dots < t_N = T$ with $t_{n+1} - t_n = \Delta t$, and $\xi_n \sim \mathcal{N}(0, I)$ i.i.d. Recall that Euler's method is strongly convergent of order $1/2$ and weakly convergent of order 1, i.e.,

$$\sup_n \mathbb{E}[|\tilde{X}_n - X_{t_n}|] \leq C\Delta t^{1/2}$$

$$\sup_n |\mathbb{E}[f(\tilde{X}_n)] - \mathbb{E}[f(X_{t_n})]| \leq D\Delta t$$

for $C, D$, independent of $\Delta t$. In the last inequality, $f\colon \mathbb{R}^n \to \mathbb{R}$ is chosen from a suitable class of functions, e.g., $\mathcal{C}_b = \{f\colon \text{continuous and bounded}\}$ or $\mathcal{P}_k = \{f\colon \text{polynomials of degree at most } k\}$.[7] Note that $D$ may depend on $f$.

**Remark 10.1.** *A few remarks are in order.*

(i) *One cannot construct a fully implicit numerical method for SDEs analogous to the backward Euler method for ODEs, due to a lack of measurability of the resulting discretized solution (see Example 5.6 on p. 23).*

(ii) *For functionals of $(X_t)_{t>0}$ of the form $\mathbb{E}[f(X_T)]$ we can use a Monte-Carlo method to get unbiased estimates, e.g.,*

$$\mathbb{E}\left[f(X_T)\right] = \frac{1}{M}\sum_{i=1}^{M} f(\tilde{X}_N(\omega_i)) + \mathcal{O}\left(\Delta t + M^{-1/2}\right)$$

*Here the error term $\mathcal{O}\left(\Delta t + M^{-1/2}\right)$ represents the width of the confidence interval for estimating the mean $\mathbb{E}[\cdot]$. The numerical complexity, if we want to achieve an accuracy $\epsilon > 0$ with a certain confidence level, goes like $O(\epsilon^{-3})$.*

## 10.2  Strongly continuous semigroups

Starting from

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t\,, \quad X_0 = x$$

we want to compute

$$\mathbb{E}[f(X_t)] = \underbrace{\int_\Omega f(X_t(\omega))d\mathbb{P}(\omega)}_{\text{avg. over realizations of } X_t}$$

$$= \underbrace{\int_{\mathbb{R}^n} f(y)d\mathbb{P}_{X_t}(y)}_{\text{avg. over distribution of } X_t}$$

where $f$ is integrable and $\mathbb{P}_{X_t} = \mathbb{P}\circ X_t^{-1}$ denotes the distribution of $X_t$ at time $t > 0$ (i.e., the push forward of the base measure $\mathbb{P}$ by $X_t$).

**Remark 10.2.** *Note that, for a fixed initial value $X_0 = x$, the expectation $\mathbb{E}[f(X_t)] = \mathbb{E}[f(X_t)|X_0 = x]$ is always a function of $x$ and $t$.*

---

[7]The space $\mathcal{C}_b(\mathbb{R}^n)$ is dual to the space of all finite Borel measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

**Example 10.3** (Heat equation). *Let $X_t = B_t + x$, where $(B_t)_{t>0}$ is Brownian motion in $\mathbb{R}$. Then $X_t \sim N(x, t)$ with density*

$$\rho(y, t) := \rho(y, t; x) = \frac{1}{(2\pi t)^{1/2}} \exp\left(-\frac{|y - x|^2}{2t}\right).$$

In the following we will assume that $\mathbb{P}_{X_t}$ has a smooth density $\rho$ with respect to Lebesgue measure, i.e., $d\mathbb{P}_{X_t}(y) = \rho(y, t)dy$.

We seek an evolution equation for the conditional expectation

$$u(x, t) := \int_{\mathbb{R}^n} f(y)\rho(y, t; x)dy = \mathbb{E}[f(X_t)|X_0 = x].$$

in terms of the functions $u$ or $\rho$. We will first need a couple of definitions.

**Definition 10.4** (Transition kernel). *Let $(X_t)_{t>0} \subset \mathbb{R}^n$ be a homogeneous Markov process. The* transition kernel *(also called* transition function*) is a function $p\colon [0, \infty) \times \mathbb{R}^n \times \mathcal{B}(\mathbb{R}^n) \to [0, 1]$ defined by*

$$p(t, x, A) = \mathbb{P}\left(X_{t+s} \in A | X_s = x\right) \quad s, t \geq 0, \quad x \in \mathbb{R}^n, \quad A \in \mathcal{B}(\mathbb{R}^n)$$

*that satisfies the Chapman-Kolmogorov equation*

$$p(t + s, x, A) = \int_{\mathbb{R}^n} p(t, y, A)p(s, x, dy) = \int_A \left(\int_{\mathbb{R}^n} q(t, y, z)q(s, x, y)dy\right) dz.$$

*Here $q(t, x, \cdot)$ is the kernel density with respect to the Lebesgue measure (that we tacitly assume to exist if not stated otherwise).*

**Definition 10.5** (Strongly continuous contraction semigroup). *Let $(K, \|\cdot\|)$ be a Banach space with norm $\|\cdot\|$. A strongly continuous contraction semigroup is a one-parameter family of operators $S_t\colon K \to K$ such that*

(i) *$S_0 = \mathrm{Id}$ and $S_{t+u} = S_t \circ S_u$ for all $t, u > 0$ (semigroup property)*

(ii) *$\lim_{t\to 0} S_t f = f$ where the limit is understood with respect to the strong operator topology, i.e., $\|S_t f - f\| \to 0$ as $t \to 0$ (continuity)*

(iii) *$\|S_t\| \leq 1$ for all $t > 0$ (contraction property)*

**Definition 10.6.** (Feller process)*: A Markov process $(X_t)_{t>0}$ is called a* Feller *process if the operators defined by*

$$(S_t f)(x) := \mathbb{E}\left[f(X_t)|X_0 = x\right] = \int f(y)q(t, x, y)dy$$

*form a strongly continuous contraction semigroup in the Banach space $(\mathcal{C}_0, \|\cdot\|_\infty)$ where $\mathcal{C}_0$ is the space of bounded functions which vanish at infinity, i.e. the space of bounded functions with compact support, and $\|\cdot\|_\infty$ is the supremum norm.*

**Example 10.7** (Heat equation, cont'd). *Let $(B_t)_{t>0}$ be a Brownian motion. The Markov process $(X_t)_{t>0}$ defined by $X_t = B_t + x$ is a Feller process with transition density*

$$q(t, x, y) = \frac{1}{(2\pi t)^{1/2}} \exp\left(-\frac{|y - x|^2}{2t}\right).$$

## 10.3 Feynman-Kac theorem (backward equation)

For all practical purposes the solutions to the SDEs of the above form are of Feller type. Let us introduce the shorthand

$$\mathbb{E}_x[f(X_t)] = \mathbb{E}[f(X_t)|X_0 = x].$$

We will now state one version of the famous Feynman-Kac theorem[8] that allows for expressing the semigroup $S_t$ in terms of the solution to a partial differential equation (PDE); for further information the reader is referred to [15].

**Theorem 10.8** (Feynman-Kac formula). *Fix $T > 0$. Let $u(x,t)$ be the solution of the linear, parabolic PDE*

$$\left(\frac{\partial}{\partial t} + L\right) u(x,t) = 0, \quad u(x,T) = f(x), \tag{10.1}$$

*where*

$$L\phi = \frac{1}{2}\sigma\sigma^\top : \nabla^2\phi + b \cdot \nabla\phi.$$

*Then*

$$u(x, t=0) = \mathbb{E}_x\left[f(X_T)\right]$$

*Proof.* The proof uses Itô's formula for $Y_t = u(X_t, t)$:

$$dY_t = \nabla u \cdot dX_t + \frac{\partial u}{\partial t}dt + \frac{1}{2}\left(\sigma\sigma^\top : \nabla^2 u\right)dt$$

$$= \left(\frac{\partial}{\partial t} + L\right) u(X_t, t)dt + \left(\sigma^\top \nabla u\right)(X_t, t) \cdot dB_t.$$

Now integrate both sides of the equation from $0$ to $T$ and use (10.1) to obtain

$$\underbrace{u(X_T, T)}_{=f(X_T)} - u(X_0, 0) = \underbrace{\int_0^T \left(\frac{\partial}{\partial t} + L\right) u(X_t, t)dt}_{=0} + \int_0^T \left(\sigma^\top \nabla u\right)(X_t, t) \cdot dB_t$$

Note that the rightmost integral is a martingale, given suitable conditions on the coefficients of (10.1) that guarantee that $u$ is continuously differentiable. Taking expectations then yields the assertion:

$$\mathbb{E}_x\left[f(X_T)\right] = u(x, 0).$$

$\square$

**Remark 10.9.** *Equation (10.1) is a so-called backward evolution equation, a PDE with a terminal condition, and is known as the* Kolmogorov backward equation.

By altering the proof of Theorem 10.8, it is now straightforward to derive variants of the Feynman-Kac formula and even a forward evolution equation for the transition density.

---

[8]Richard P. Feynman (1918–1988), American Physicist and Nobel Prize Winner; Mark (Marek) Kac (1914–1984), Polish-American mathematician.

(i) Let $v(x, t)$ solve

$$\left(\frac{\partial}{\partial t} - L\right) v(x, t) = 0, \quad v(x, 0) = f(x). \tag{10.2}$$

Then $v(x, t) = \mathbb{E}_x[f(X_t)]$.

(ii) Let

$$\rho(y, t; x) dy = \mathbb{P}\left(X_t \in [y, y + dy] \mid X_0 = x\right).$$

We will also use the notation $\rho(y, t) := \rho(y, t; x)$ when the starting point $x$ does not change; note that $\rho(y, t; x) = q(t, x, y)$ is the same as the transition density for fixed starting value $X_0 = x$.

Using integration by parts it can be shown that $\rho$ solves the *Kolmogorov forward equation* (also known as the *Fokker-Planck Equation*),

$$\left(\frac{\partial}{\partial t} - A\right) \rho(y, t) = 0, \quad \lim_{t \to 0} \rho(\cdot, t) = \delta_x \tag{10.3}$$

where $A = L^*$ is the formal $L^2$-adjoint of the operator $L$, i.e.

$$\langle f, Lg \rangle = \int f(x) \left(Lg\right)(x) dx = \int \left(Af\right)(x) g(x) dy = \langle Af, g \rangle.$$

We leave the proofs as an exercise to the reader.

**Example 10.10** (Heat equation, cont'd). *The* heat kernel

$$\rho(y, t) = \frac{1}{(2\pi t)^{1/2}} \exp\left(-\frac{|y - x|^2}{2t}\right)$$

*solves the PDE*

$$\frac{\partial \rho}{\partial t} = \frac{1}{2} \frac{\partial^2 \rho}{\partial y^2}, \quad \lim_{t \to 0} \rho(\cdot, t) = \delta_x$$

**Semigroup interpretation**

The Feynman-Kac theorem provides a connection between certain Itô SDEs and partial differential equations (most notably Kolmogorov forward and backward equations). In terms of the solution to (10.2) our semigroup reads

$$(S_t f)(x) = \mathbb{E}_x\left[f(X_t)\right] = \left(e^{tL} f\right)(x),$$

where $v(x, t) = (e^{tL} f)(x)$ is the formal solution of the PDE

$$\frac{\partial v}{\partial t} = Lv, \quad v(x, 0) = f(x)$$

In other words, the semigroup has the suggestive representation

$$S_t = e^{tL}.$$

We say that $S_t$ *is generated by* $L$.

# 11 Day 11, 08.01.2013: Exit times

## 11.1 Some more semigroup theory

Let $(X_t)_{t \geq 0} \subset \mathbb{R}^n$ be the solution of the time-homogenous SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t\,, \quad X_0 = x\,,$$

with, e.g.,

$$b(x) = -\nabla V(x)\,, \quad \sigma(x) = \sqrt{2\epsilon}\,,$$

with $\epsilon > 0$ and a smooth function $V\colon \mathbb{R}^n \to \mathbb{R}$ that is bounded below and sufficiently growing at infinity (see Fig. 8.2 on p. 37). Further let $(S_t)_{t \geq 0}$ be the corresponding semigroup, defined by

$$(S_t f)(x) = \mathbb{E}_x[f(X_t)]$$

where $f \in \mathcal{C}_0(\mathbb{R}^n)$. Here $\mathcal{C}_0(\mathbb{R}^n)$ denotes the space of continous functions that vanish at infinity, and it comes with a natural norm $\|f\|_\infty = \sup_{x \in \mathbb{R}^n} |f(x)|$; we have already seen that $S_t$ defines a strongly continuous contraction semigroup on $(\mathcal{C}_0(\mathbb{R}^n), \|\cdot\|_\infty)$, and that it can be represented by the exponential $S_t = \exp(tL)$ of the the second-order differential operator

$$L\phi = \frac{1}{2}\sigma\sigma^\top : \nabla^2 \phi + b \cdot \nabla \phi\,.$$

There is more to say about the relation between the semigroup and the diffusion operator $L$ that is expressed in the following definition.

**Definition 11.1** (Infinitesimal generator). *Let $X = (X_t)_{t \geq 0}$ be a continuous-time Markov process with semigroup $(S_t)_{t \geq 0}$ on a Banach space $(K, \|\cdot\|)$. The infinitesimal generator of $X$ is defined as*

$$L\phi(x) = \lim_{t \to 0} \frac{(S_t \phi)(x) - \phi(x)}{t}\,, \tag{11.1}$$

*provided that the limit exists in $(K, \|\cdot\|)$.*

**Remark 11.2.** *If $X = (X_t)_{t \geq 0}$ is time-discrete on a, say, countable state space (i.e., $X$ is a Markov chain with transition matrix $P = (p_{xy})_{x,y}$), then*

$$(S_n f)(x) = \sum_y f(y) p_{xy}^{(n)} = \left(P^n f\right)_x\,,$$

*which suggests to define the* discrete generator *by $Lf = Pf - f$, i.e., $L = P - \mathrm{Id}$.*

**Theorem 11.3.** *If $(X_t)_t$ is the solution of*

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t\,, \quad X_0 = x$$

*and $\varphi \in C^2(\mathbb{R}^n)$ then the infinitesimal generator from (11.1) is given by*

$$L = \frac{1}{2}\sigma\sigma^\top : \nabla^2 + b \cdot \nabla.$$

*Proof.* The theorem largely follows from Itô's formula; see [15, Thm. 7.3.3]. $\qquad \square$

**Example 11.4** (Heat equation, cont'd). *The operator in the n-dimensional heat equation*

$$\frac{\partial \rho}{\partial t} = \frac{1}{2} \sum_{i=1}^{n} \frac{\partial^2 \rho}{\partial x_i^2}$$

*is the $L^2$-adjoint of the generator associated of the n-dimensional Brownian motion $B_t$, i.e., the generator of $B_t$ is one half the Laplacian,*

$$L = \frac{1}{2} \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}$$

*(Note that the Laplacian is essentially self-adjoint.)*

**Exercise 11.5.** *Prove that $LS_t = S_t L$, i.e., $L\left(\mathbb{E}_x\left[f(X_t)\right]\right) = \mathbb{E}_x\left[Lf(X_t)\right]$.*

## 11.2 Exit times: more Feynman-Kac formulae

Let $\Omega \subset \mathbb{R}^n$ be an open and bounded domain with smooth boundary, $\partial \Omega \in \mathcal{C}^\infty$. In this section, we investigate connections between stopped SDEs and linear boundary value problems (BVPs) of the form

$$Au = f \quad \text{in } \Omega$$
$$u = g \quad \text{on } \partial \Omega$$

where $A$ is a second-order differential operator, $f \in \mathcal{C}(\Omega)$, $g \in \mathcal{C}(\partial \Omega)$ and the solution $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$. The function $g$ prescribes the values of $u$ on the boundary $\partial \Omega$. The next theorem is a stopping-time version of Itô's formula.

**Theorem 11.6.** (Dynkin's formula): *Let $\tau$ be a stopping time with $\mathbb{E}_x\left[\tau\right] < \infty$. Then for $f \in \mathcal{C}_0^2(\mathbb{R}^n)$, we have*

$$\mathbb{E}_x\left[f(X_\tau)\right] = f(x) + \mathbb{E}_x\left[\int_0^\tau Lf(X_s)ds\right].$$

*Proof.* The assertion follows from Itô's formula and the Martingale property of stopped stochastic integrals; see [15, Lem. 7.3.2]. □

**Example 11.7.** *We want to compute the mean first exit time (MFET) of n-dimensional Brownian Motion from the open ball $D = \{x \in \mathbb{R}^n \colon |x| < R\}$ of radius $R > 0$ (see Fig. 11.1). To this end let $X_t = B_t + x$ where we assume that $X_0 = x \in D$. We choose $k \in \mathbb{N}$ and define the stopping time $\tau_k = k \wedge \tau_D$, with*

$$\tau_D = \inf\{t > 0 \colon X_t \notin D\}.$$

*By definition our stopping time $\tau_k$ is a.s. finite, so that Dynkin's formula for the function $f(x) = |x|^2$ for $x \in D$ yields*

$$\mathbb{E}_x\left[f\left(X_{\tau_k}\right)\right] = f(x) + \mathbb{E}_x\left[\int_0^{\tau_k} \frac{1}{2}\Delta f(X_s)ds\right]$$
$$= |x|^2 + n\mathbb{E}_x\left[\tau_k\right].$$

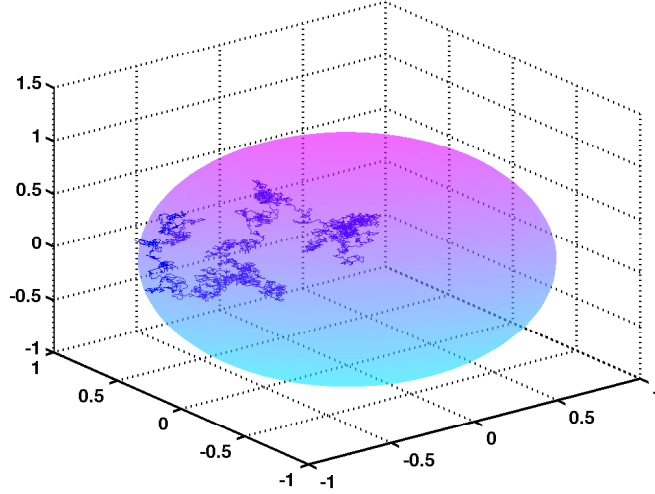*In particular, since $\mathbb{E}_x[f(X_{\tau_k})] \leq \mathbb{E}_x[f(X_{\tau_D})] = R^2$, it follows that*

Figure 11.1: How long does 3-dimensional Brownian motion need on average to leave a sphere of radius $R = 1$? See Example 11.7 below.

$$\mathbb{E}_x\left[\tau_k\right] \leq \frac{R^2 - |x|^2}{n}.$$

*Now $\tau_k \to \tau_D$ as $k \to \infty$ and hence, by monotone convergence,*

$$\mathbb{E}_x\left[\tau_D\right] = \frac{R^2 - |x|^2}{n}.$$

*Observe that the MFET is inversely proportional to the dimension of the ball, which can be understood as follows. If the initial point $x$ is $\epsilon$ away from the set boundary, i.e., if $R^2 - |x|^2 = \epsilon^2$ then every dimension adds another direction in which the particle can exit. Therefore $\mathbb{E}_x[\tau_D] \propto 1/n$.*

**Remark 11.8.** *If $\tau$ is the first exit time of an open, bounded set $\Omega \subset \mathbb{R}^n$,*

$$\tau_\Omega := \inf\left\{t > 0 \colon X_t \notin \Omega\right\}$$

*then $\mathbb{E}_x[\tau] < \infty$.*

Our next result establishes the connection between the mean first exit time from a set with a boundary value problem specified on that set.

**Lemma 11.9** (Mean first exit time). *Let $\Omega \subset \mathbb{R}^n$ be an open, bounded domain with sufficiently nice boundary $\partial\Omega$.*

$$\tau = \inf\left\{t > 0 \colon X_t \notin \Omega\right\}$$

*Let $\varphi \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$ solve the boundary value problem*

$$\begin{aligned} L\varphi(x) &= -1\,, \quad x \in \Omega \\ \varphi(x) &= 0\,, \quad x \in \partial\Omega. \end{aligned} \tag{11.2}$$
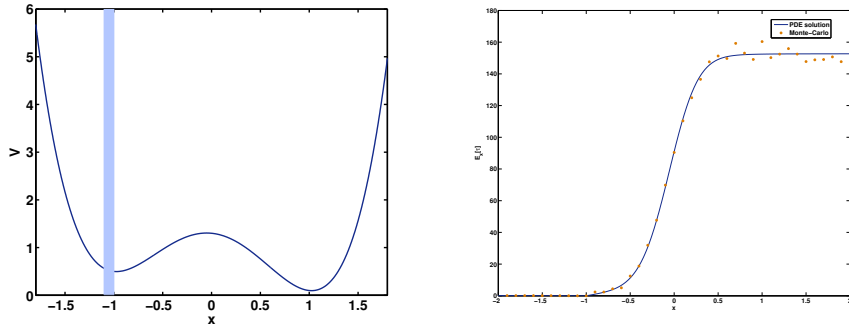
Figure 11.2: Tilted double well potential (left panel) and mean first passage time of the light blue region (right panel). The solid line in the right panel shows the finite difference solution of the MFPT using 1500 grid points, the dots are the Monte-Carlo estimates for 1000 independent realizations, starting from each initial value (from $x = -2$ to $x = 2$ in 0.1 steps).

*Then $\varphi(x) = \mathbb{E}_x[\tau]$.*

*Proof.* We can use Itô's formula (or Dynkin's formula) for $\varphi$:

$$\varphi(X_\tau) = \varphi(X_0) + \int_0^\tau L\varphi(X_s)ds + \int_0^\tau \left(\sigma^\top \nabla \varphi\right)(X_s)dB_s.$$

By the assumption that $\varphi$ is $\mathcal{C}^2$ inside the domain $\Omega$, it follows that the rightmost term is a Martingale; taking expectations and using the PDE for $\varphi$, we find

$$\mathbb{E}_x[\underbrace{\varphi(X_\tau)}_{=0}] = \varphi(x) + \mathbb{E}_x\left[\int_0^\tau \underbrace{L\varphi(X_s)}_{=-1} ds\right] + \mathbb{E}_x\left[\int_0^\tau \left(\sigma^\top \nabla \varphi\right)(X_s)dB_s\right],$$

where the leftmost term follows from the boundary condition $\varphi|_{\partial\Omega} = 0$. Hence

$$0 = \varphi(x) - \mathbb{E}_x[\tau]$$

which proves the lemma. $\qquad\square$

## 11.3 Applications of mean first exit time formulae

The boundary value problem (11.2) for the MFET can be used to accurately compute the exit time rather when the system is not too high-dimensional (say, at most three-dimensional) and Monte-Carlo sampling would be inefficient (e.g, if the exit times are extremely large). As an example let us consider our one-dimensional paradigm, diffusion in a double-well potential:

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\epsilon}\, dB_t, \quad X_0 = x\,, \tag{11.3}$$

with $V$ as shown in the left panel of Figure 11.2 below and $\epsilon \ll 1$. Further let $A = [-1.1, -1]$. We want to compute the mean first passage time (MFPT) of $A$, i.e., the mean first exit time of the set $\mathbb{R} \setminus A \subset \mathbb{R}$. The right panel of Figure 11.2
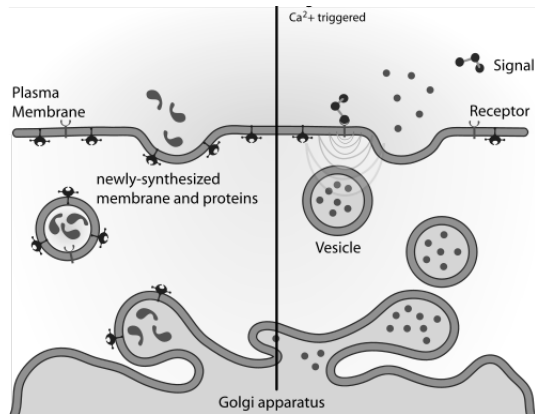
Figure 11.3: Process of exocytosis (source: Wikipedia).

shows the comparison of a brute force Monte-Carlo and a PDE solution of the MFPT. Here the Monte-Carlo solution was based on an the unbiased estimator

$$\hat{\varphi}(x) = \frac{1}{N} \sum_{i=1}^{N} \tilde{\tau}_i^x$$

where $\tilde{\tau}^x$ denotes the numerical approximation of the first passage time, using an Euler discretization of (11.3) with initial value $x \in \{-2, -1.9, \dots, 1.9, 2\}$ and $N = 1000$ independent realizations. The PDE solution was obtained using a finite difference discretization of the mixed BVP

$$\begin{aligned} L\varphi(x) &= -1\,, \quad x \in (-3, 3) \setminus [-1.1, 1] \\ \varphi(x) &= 0\,, \quad x \in [-1.1, 1] \\ \varphi'(x) &= 0\,, \quad x \in \{-3, 3\} \end{aligned}$$

on a fine grid with grid size $\Delta x = 0.004$ (i.e., 1500 grid points). In terms of runnig time, the PDE solution clearly rules out the Monte-Carlo method (1 sec. vs. 10 hrs.), moreover the PDE solution is much more accurate as can be seen from the figure. However Monte-Carlo becomes competitive if the dimension grows, for the computational costs are independent of the dimension (for a single initial condition), whereas they grow exponentially with the dimension in case of a grid-based PDE solution—the number of grid points grows like $\mathcal{O}(2^n)$.

We briefly mention some applications in which MFETs are sought.

**Exocytosis** Exocytosis is a process by which a cell ejects the content of secretory vesicles into the extracellular domain (see Fig. 11.3). Here the cell can be modeled by an open bounded domain in $\mathbb{R}^3$, and the elastic plasma membrane by a nice boundary $\partial \Omega$. The whole process is diffusion-dominated but strongly affected by the cell geometry and crowding. A typical question that biologists ask is, e.g., how long it takes on average for a secretory vesicle (that can be well modelled as a point particle) that is produced in the Golgi apparatus to reach the cell membrane.
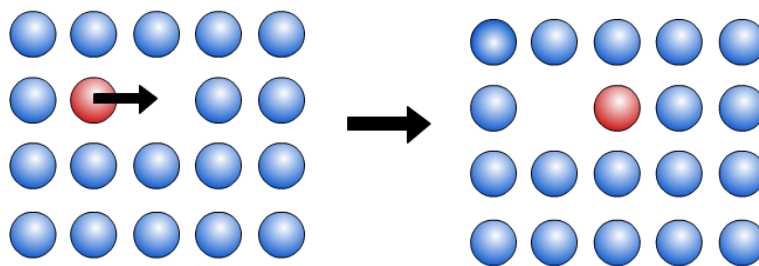
Figure 11.4: Migration of atoms into the vacancy in a crystal lattice that can be interpreted as the diffusion of the vacancy (source: University of Cambridge).

**Vacancy diffusion in crystals**  A vacancy is a "missing atom" in an atomic lattice. If a vacancy is present, one of the adjacent atoms can move into the vacancy, creating a new vacancy on the former position of the atom. By the symmetry of the crystal lattice there is an equal probability that any of the adjacent atoms will move into the vacancy, including the atom that has just created the vacancy (see Fig. 11.4). It is possible to think of this mechanism as a moving vacancy, rather than the motion of atoms into the vacancy. In fact the vacancy undergoes some kind of random walk or diffusion between lattice sites and. Physicists are interested in measuring the diffusivity of the vacancy, i.e., how long on average it takes for the vacancy to propagate through the crystal, which can be expressed in terms of the mean exit time of a test particle from the periodic unit cell of the crystal.

**Transition state theory**  Consider the reaction of a molecule $A$ to a different molecule $B$ by formation of a molecular complex $AB$, just in the way that is depicted in Figure 11.5. If the reaction is temperature-activated the situation can be conveniently modelled by a one-dimensional diffusion in a bistable potential as is described by (11.3), with $V = G$ being the thermodynamic free energy associated with the reaction and $0 < \varepsilon \ll 1$ being the temperature in the system. If we let $\Omega = (-\infty, b)$ denote the product state and $\partial\Omega = b$, the exit time problem (11.2) can be shown to have the explicit solution

$$\mathbb{E}_x[\tau] = \frac{1}{\varepsilon} \int_b^x \int_x^\infty \exp\left(\frac{G(z) - G(x)}{\varepsilon}\right) dz\, dy\,.$$

Now suppose that $a$ is the unique minimum of $G$ in $\Omega$ and define $\Delta G^\ddagger = G(b) - G(a)$. Then it follows by Laplace's method (see [3]) that

$$\lim_{\varepsilon \searrow 0} \varepsilon \log \mathbb{E}_x[\tau] = \Delta G^\ddagger\,. \tag{11.4}$$

Observe that in the limit $\epsilon \to 0$ the MEFT $\mathbb{E}_x[\tau]$ becomes essentially $\mathbb{E}_a[\tau]$ and hence independent of the actual initial condition. In other words, at small temperature the system is always close to the minimum at $x = a$. If we define the rate $k_{AB}$ of transitions between reactant and product state as $1/\mathbb{E}_x[\tau]$—the transition time between the transition state and the product state is negligble—
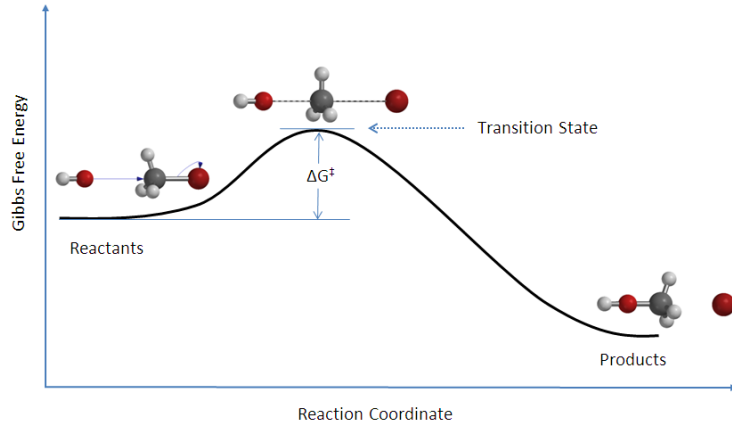
Figure 11.5: Temperature-activated reaction of molecule $A$ to a molecule $B$ by formation of a transition state $AB$ (source: Wikipedia).

and drop the limit, our asymptotic formula turns into

$$k_{AB} \asymp \exp\left(-\frac{\Delta G_{\ddagger}^{\ddagger}}{k_B T}\right),$$

with $\epsilon = k_B T$ where $T$ is the physical temperature and $k_B$ is Boltzmann's constant. The last equation is known as *Kramers-Eyring law* in physics or *Arrhenius formula* in chemistry.

**Yet another Feynman-Kac formula**

One possible generalization of Lemma 11.9 is here: Let $\psi \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\bar{\Omega})$ solve the linear boundary value problem

$$\begin{aligned} L\psi(x) &= f, & x \in \Omega \\ \psi(x) &= g, & x \in \partial\Omega. \end{aligned} \tag{11.5}$$

on the open bounded set $\Omega \subset \mathbb{R}^n$ for bounded continuous functions $f$ and $g$. Letting $\tau$ denote the first exit time of $X_t$ (generated by $L$) from $\Omega$, we have

$$\psi(x) = \mathbb{E}_x\left[g(X_\tau) - \int_0^\tau f(X_s)ds\right].$$

Again, the formula follows by using Dynkin's formula:

$$\mathbb{E}_x\left[\psi(X_\tau)\right] = \psi(x) + \mathbb{E}_x\left[\int_0^\tau L\psi(X_s)ds\right]$$

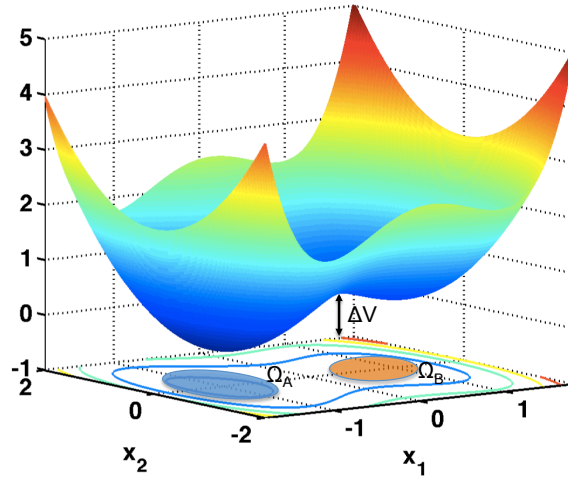$$\Rightarrow \psi(x) = \mathbb{E}_x\left[g(X_\tau) - \int_0^\tau f(X_s)ds\right].$$

54

Figure 12.1: Bistable potential $V : \mathbb{R}^2 \to \mathbb{R}$ with basins of attraction $\Omega_A$ and $\Omega_B$; $\Delta V > 0$ is the potential barrier between the left well and the saddle point.

## 12 Day 12, 15.01.2013: Fokker-Planck equation

### 12.1 Propagation of densities

We begin our discussion by considering again the SDE

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\varepsilon}dB_t \qquad (12.1)$$

with a bistable potential $V : \mathbb{R}^n \to \mathbb{R}$. Let $\tau_A$ be the stopping time

$$\tau_A = \inf \{ t > 0 \ : \ X_t \notin \Omega_A \} \,,$$

of an open bounded set $\Omega_A \subset \mathbb{R}^n$ (see Fig. 12.1 below). Recall the small noise asymptotics for $\tau_A$, which should be read as in (11.4):

$$\mathbb{E}_x [\tau_A] \asymp e^{\Delta V/\varepsilon} \quad \text{as } \varepsilon \to 0.$$

Let $(X_t)_{t \geq 0}$ be an infinitely long realization of (12.1) and consider the situation depicted in Figure 12.1 where the wells $\Omega_A$, $\Omega_B$ may represent, e.g.,

- biologically relevant conformations of a molecule,
- ice age and interglacial periods of our climate system, or
- reactant and product of a catalytic reaction.

An interesting question now is how much time does $X_t$ spend in either $\Omega_A$ or $\Omega_B$? To this end recall that the distribution of $X_t$ at time $t$ is given by

$$p(t, x, A) = \mathbb{P}(X_t \in A \mid X_0 = x) \,, \quad A \subset \mathbb{R}^n \text{ Borel set}$$

with density $\rho(y, t; x)$, i.e.

$$p(t, x, \Omega_B) = \int_{\Omega_B} \rho(y, t; x)dy \,.$$

## 12.2 Adjoint semigroup: transfer operator

We seek an evolution equation for the density $\rho$ and study its properties for the long time limit $t \to \infty$ (and for $\varepsilon$ being small). When we do not need to consider the initial value $x$, we will use $\rho(y, t) := \rho(y, t; x)$ to refer to the density.

**Theorem 12.1.** *Let $(X_t)_{t \geq 0}$ solve the SDE*

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad X_0 = x$$

*with the initial value $x$ being randomly distributed according to the density $\rho_0$. Then the density $\rho(\cdot, t; x) = \rho(\cdot, t)$ solves the Fokker-Planck equation*

$$\left( \frac{\partial}{\partial t} - L^* \right) \rho(y, t) = 0$$

*where $\rho(y, 0) = \rho_0(y)$ and $L^*$ is the $L^2$-adjoint of the infinitesimal generator*

$$L = \frac{1}{2} \sigma \sigma^\top : \nabla^2 + b \cdot \nabla.$$

*Proof.* Let $u(x, t) = \mathbb{E}_x[f(X_t)] = (S_t f)(x)$ where we recall that $(S_t)_{t > 0}$ is the one-parameter family of semigroups $S_t = e^{tL}$. Then

$$\mathbb{E}[f(X_t)] = \int_{\mathbb{R}^n} \mathbb{E}_x[f(X_t)] \rho_0(x)dx$$

$$= \int_{\mathbb{R}^n} u(x, t)\rho_0(x)dx$$

$$= \int_{\mathbb{R}^n} (S_t f)(x)\rho_0(x)dx$$

$$= \int_{\mathbb{R}^n} f(x)(S_t^* \rho_0)(x)dx.$$

Note that

$$\frac{\partial}{\partial t} \int_{\mathbb{R}^n} u(x, t)\rho_0(x)dx = \frac{\partial}{\partial t} \int_{\mathbb{R}^n} (Lu)(x, t)\rho_0(x)dx$$

$$= \int_{\mathbb{R}^n} (S_t L f)(x)\rho_0(x)dx$$

$$= \int_{\mathbb{R}^n} f(x)(L^* S_t^* \rho_0)(x)dx,$$

which implies that $(\exp(tL))^* = \exp(tL^*)$, i.e., $L^*$ generates the evolution semi-group $S_t^* \colon (\mathcal{C}_0(\mathbb{R}^n))^* \to (\mathcal{C}_0(\mathbb{R}^n))^*$ on the dual space $(\mathcal{C}_0(\mathbb{R}^n))^*$—this the space of (nonnegative) finite Borel measures.[9] By definition of $\rho$, we also have

$$\mathbb{E}[f(X_t)] = \int f(y)\, d\mathbb{P}_{X_t}(y) = \int f(y)\rho(y, t)dy.$$

Therefore,

$$\int f(y)(e^{tL^*} \rho_0)(y)dy = \int f(y)\rho(y, t)dy$$

---

[9]Note that the adjoint is defined with respect to the inner product in the space $L^2(\mathbb{R}^n)$, but using a density argument it can be shown that it is sufficient to consider $f \in \mathcal{C}_0(\mathbb{R}^n)$, since $\mathcal{C}_0(\mathbb{R}^n)$ is a dense subspace of $L^2(\mathbb{R}^n)$.

and, since $f \in \mathcal{C}_0(\mathbb{R}^n)$ was arbitrary, it follows that

$$\left(e^{tL^*}\rho_0\right)(y) = \rho(y,t)$$

with $\rho(y,0) = \rho_0(y)$. This yields

$$\frac{\partial \rho}{\partial t} = L^*\left(e^{L^*t}\rho_0\right) = L^*\rho$$

so $\rho$ solves the Fokker-Planck equation. $\qquad\square$

**Example 12.2** (Ornstein-Uhlenbeck Process). *The Ornstein-Uhlenbeck (OU) process is described by the SDE*

$$dX_t = AX_t dt + B dW_t\,, \quad X_0 = x\,,$$

*where we assume that $X_0 = x$ is fixed, $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$. The infinitesimal generator of the process $X_t$ satisfies (for all scalar-valued twice continuously differentiable functions $h$ in its domain)*

$$Lh(x) = \frac{1}{2}BB^\top : \nabla^2 h(x) + (Ax) \cdot \nabla h(x)$$

*and the formal $L^2$-adjoint of $L$, $L^*$ satisfies*

$$\langle L^*u, v\rangle_{L^2} = \int (L^*u)\,v dx = \int u\,(Lv)\,dx = \langle u, Lv\rangle_{L^2}\,.$$

*We calculate $L^*$ using the above and integration by parts:*

$$\begin{aligned}
\int u\,(Lv)\,dx &= \int u\left(\frac{1}{2}BB^\top : \nabla^2 v + (Ax) \cdot \nabla v\right) dx \\
&= -\frac{1}{2}\int \nabla u \cdot \left(\frac{1}{2}BB^\top \nabla v\right) dx - \int v \nabla \cdot (uAx) dx \\
&= \frac{1}{2}\int v\left(\frac{1}{2}BB^\top : \nabla^2 u\right) dx - \int v\left(\nabla u \cdot (Ax) + u\nabla \cdot (Ax)\right) dx \\
&= \int v\left(\frac{1}{2}BB^\top : \nabla^2 u - (Ax) \cdot \nabla u - \mathrm{tr}(A)u\right) dx
\end{aligned}$$

*which implies that (here $\mathrm{tr}(\cdot)$ denotes the trace operator)*

$$L^*h(x) = \frac{1}{2}BB^\top : \nabla^2 h(x) - (Ax) \cdot \nabla h(x) - \mathrm{tr}(A)h(x).$$

*This gives us the Fokker-Planck equation,*

$$\frac{\partial \rho}{\partial t} - \frac{1}{2}BB^\top : \nabla^2 \rho + \nabla \cdot (\rho Ax) = 0\,,$$

*the solution of which we know already (cf. Example 8.5), namely,*

$$X_t \sim \mathcal{N}\left(e^{At}x, \int_0^t e^{As}BB^\top e^{A^\top s}ds\right)$$

In other words, the OU process at time $t > 0$ is Gaussian with mean

$$\mathbb{E}[X_t] = e^{At}x$$

and covariance

$$\mathbb{E}\big[\,(X_t - \mathbb{E}[X_t])\,(X_t - \mathbb{E}[X_t])^\top\,\big] = \int_0^t e^{At}BB^\top e^{A^\top t}\,dt,$$

In particular we observe that

$$\lim_{t \searrow 0} \rho(y,t) = \delta_x(y) \quad \text{(weakly)}.$$

If the eigenvalue of $A$ lie entirely in the left complex half-plane $\mathbb{C}^-$ (i.e., have strictly negative real part), then both mean and covariance converge as $t \to \infty$:

$$\mathbb{E}[X_t] \to 0$$

and

$$\mathbb{E}\big[\,(X_t - \mathbb{E}[X_t])\,(X_t - \mathbb{E}[X_t])^\top\,\big] \to \int_0^\infty e^{At}BB^\top e^{A^\top t}\,dt\,,$$

i.e., for almost all initial conditions, the solution $\rho(x,t)$ of the Fokker-Planck equation converges to the Gaussian density

$$\rho_\infty = \mathcal{N}\left(0, \int_0^\infty e^{As}BB^\top e^{A^\top s}\,ds\right)$$

**Exercise 12.3.** *Compute the solution of the Fokker-Planck equation*

$$\frac{\partial \rho}{\partial t} = \frac{\partial^2}{\partial x^2}\left(\frac{\sigma^2 x^2}{2}\rho\right) - \frac{\partial}{\partial x}\left(\mu x \rho\right), \quad \lim_{t \searrow 0} \rho(x,t) = \delta_a(x)$$

*for some $a > 0$. (Hint: write down the corresponding SDE.)*

## 12.3 Invariant measures

We now ask whether the solution $(X_t)_{t \geq 0}$ associated with a given SDE has an invariant measure, and if so, whether it is approached independently of the initial distribution of $X_0$. Let us make these questions precise.

**Definition 12.4** (Invariant measure). *A finite, nonnegative Borel measure $\mu_\infty$ is called an* invariant measure *if $X_0 \sim \mu_\infty$ implies that $X_t \sim \mu_\infty$ for all $t > 0$.*

It readily follows from the properties of the Fokker-Planck equation, that if $\rho_\infty$ is the Lebesgue density associated with the invariant measure $\mu_\infty$, then

$$L^*\rho_\infty = 0\,.$$

In other words, stationary distributions are steady-state solutions of the Fokker-Planck equation.

**Example 12.5** (Ornstein-Uhlenbeck process, cont'd). *If $\lambda(A) \subset \mathbb{C}^-$, i.e., all eigenvalues of $A$ have strictly negative real part, then*

$$\rho_\infty(x) = (\det(2\pi C_\infty))^{-1/2} \exp\left(-\frac{1}{2}x^\top C_\infty^{-1} x\right),$$

*where*

$$C_\infty = \int_0^\infty e^{As}BB^\top e^{A^\top s}\,ds\,.$$

*(If $C_\infty$ is not invertible, then $C_\infty^{-1}$ is the Moore-Penrose pseudo-inverse.)*

**Boltzmann-Gibbs measures**

An interesting special class of SDEs with unique known invariant measure are gradient systems of the form

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\varepsilon}dB_t, \quad X_0 = x.$$

For this class of SDEs, the density of the invariant measure is given by the

$$\rho_\infty = \frac{1}{Z}e^{-V/\varepsilon}, \quad Z = \int_{\mathbb{R}^n} e^{-V(x)/\varepsilon}dx,$$

provided that the integral exists. The corresponding invariant measure is called *Boltzmann distribution* or *Gibbs measure*.

Recall that our initial question was to ask how much time does the process $(X_t)_{t>0}$ spend in $\Omega_B$. We now answer that question:

$$\mu_\infty(\Omega_B) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \chi_{\Omega_B}(X_t)dt \quad \text{a.s.}$$

for $\mu_\infty$-almost all initial conditions $X_0 = x$. (We are anticipating results from the forthcoming lectures.) The above expression means that the invariant measure of $\Omega_B$ is given by the average time the process $(X_t)_{t>0}$ spends in $\Omega_B$, where the average is taken over an infinitely long realization of time. Note that we can also express the invariant measure of $\Omega_B$ by a *spatial* average using the associated density of the invariant measure,

$$\mu_\infty(\Omega_B) = \frac{1}{Z} \int_{\Omega_B} e^{-V(x)/\varepsilon}dx.$$

The spatial average is usually difficult to compute in dimensions greater than two or three, but the time average can be computed in a Markov Chain Monte-Carlo (MCMC) fashion by running a long realization of the SDE.

# 13 Day 13, 22.01.2013: Long-term behaviour of SDEs

**Suggested references:** [5, 17]

## 13.1 Invariant measures, cont'd

We confine our attention to gradient-type systems of the form

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\varepsilon}dB_t, \tag{13.1}$$

with a smooth potential $V: \mathbb{R}^n \to \mathbb{R}$ that is bounded below and sufficiently growing at infinity (e.g., coercive). For sufficiently smooth functions (e.g., $\mathcal{C}_0^2(\mathbb{R}^n)$), the infinitesimal generator of $X_t$ and its formal $L^2$-adjoint are

$$L = \epsilon\Delta = \nabla V \cdot \nabla, \quad L^* = \epsilon\Delta + \nabla V \cdot \nabla + \Delta V.$$

The corresponding Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = \epsilon\Delta\rho + \nabla \cdot (\rho\nabla V), \quad \rho(x,0) = \rho_0(x) \tag{13.2}$$

with $\rho_0 \geq 0$ has the unique stationary solution

$$\rho_\infty = \frac{1}{Z}e^{-V/\varepsilon}\,, \quad Z = \int_{\mathbb{R}^n} e^{-V(x)/\varepsilon}dx\,,$$

where $Z < \infty$ by the assumptions on $V$.

**Example 13.1** (OU process with symmetric stiffness matrix)**.** *The SDE*

$$dX_t = -KX_t dt + \sqrt{2\epsilon}dW_t\,, \quad X_0 = x\,,$$

*with symmetric positive definite (s.p.d.) stiffness matrix $K \in \mathbb{R}^{n \times n}$ is a gradient system with quadratic potential*

$$V(x) = \frac{1}{2}x^\top Kx\,.$$

*Its unique invariant measure is Gaussian with mean zero and covariance $C_\infty = \epsilon K^{-1}$ and has full topological support (i.e., has a strictly positive density $\rho_\infty$).*

**Remark 13.2.** *A few remarks are in order.*

1. *The Fokker-Planck equation can be written in divergence form. In our case it reads*

$$\frac{\partial\rho}{\partial t} = \nabla \cdot J\,, \quad J = \epsilon\nabla\rho + \rho\nabla V\,.$$

*Integrating $\rho(x,t)$ over $x \in \mathbb{R}^n$ and using Gauss' theorem, it follows that*

$$\frac{d}{dt}\int_{\mathbb{R}^n}\rho(x,t)dx = 0\,,$$

*which implies that*

$$\|\rho(\cdot,t)\|_{L^1(\mathbb{R}^n)} = \|\rho_0\|_{L^1(\mathbb{R}^n)} = 1\,.$$

*That is, normalized densities stay properly normalized throughout time.*

2. *It follows by PDE arguments (strong maximum principle for elliptic equations) that $\rho_\infty$ is the only stationary solution of (13.2).*

3. *It is often helpful to think of $\rho_\infty$ as the solution to the eigenvalue problem*

$$L^*v = \lambda v$$

*for the simple eigenvalue $\lambda = 0$ (it must be simple, because $\rho_\infty$ is unique). Equivalently, we may think of $\rho_\infty$ as the solution to the fixed-point equation*

$$S_t^*\rho_\infty = \rho_\infty\,,$$

*where $(S_t^*)_{t \geq 0}$ is the adjoint semigroup or transfer operator (propagator). In other words, $\rho_\infty$ is the eigenfunction of $S_t^* = \exp(tL^*)$ to the eigenvalue $\sigma = 1$. Note that $S_t^*$ and $L^*$ have the same eigenvectors; their eigenvalues are related by $\sigma(t) = \exp(\lambda t)$, i.e., the eigenvalues of the transfer operator depend on time, with the sole exception of the simple eigenvalue $\sigma = 1$.*
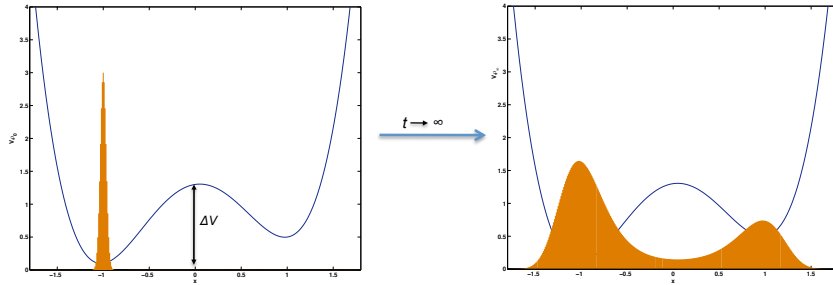
Figure 13.1: Bistable potential $V \colon \mathbb{R}^2 \to \mathbb{R}$ with basins of attraction $\Omega_A$ and $\Omega_B$; $\Delta V > 0$ is the potential barrier between the left well and the saddle point.

## 13.2 Convergence to equilibrium

We want to study the long-term behaviour of solutions to (13.2). In particular, we want to know at which speed the stationary distribution is approached (if at all). Let us start with some preliminary considerations:

**Contraction property.** Recall that the semigroup $(S_t)_{t \geq 0} \colon \mathcal{C}_0(\mathbb{R}^n) \to \mathcal{C}_0(\mathbb{R}^n)$ associated with $(X_t)_{t \geq 0}$ is a contraction, i.e., $\|S_t\| \leq 1$ for all $t$ in the operator norm that is induced by the (supremum) norm on $\mathcal{C}_0(\mathbb{R}^n)$. We further know that $\|S_t\|$ is bounded from below by the spectral radius $r_\sigma(S_t)$. Since the eigenvalue $\sigma = 1$ is simple and the generator $L$ has real spectrum (as we will see below), it follows that $r_\sigma(S_t) = 1$ and that all other eigenvalues are strictly less than 1 in modulus. The same goes for $S_t^*$ and $L^*$. Hence we can expect that all solutions of (13.2) converge to the stationary solution $\rho_\infty$, i.e.,

$$\rho = S_t^* \rho_0 \to \rho_\infty$$

in some suitable norm (e.g., $L^1$). This is the good news.

**Relaxation time scale.** Now comes the bad news. Imagine a situation as in Figure 13.1 with small noise and recall the asymptotic expression (11.4) for the mean first exit time from the left well of the potential $V$: if $\tau = \inf\{t > 0 \colon X_t \notin \text{left well}\}$ denotes the first exit time from the left energy well of $V$ as shown in Figure 13.1, then the mean first exit time satisfies

$$\lim_{\varepsilon \searrow 0} \varepsilon \log \mathbb{E}_x[\tau] = \Delta V \,.$$

This implies that for the initial density in the left well to propagate to the right well and relax to the stationary distribution, we have to wait at least

$$T = \mathcal{O}\!\left(e^{\Delta V / \epsilon}\right).$$

Hence when $\Delta V \gg \epsilon$, as is the case in most interesting applications, the typical time scale on which the convergence $\rho \to \rho_\infty$ takes place is exponentially large in the dominant barrier height of the potential $V$.

**Theorem 13.3** (Bakry & Emery, 1983)**.** *Let $V$ satisfy the convexity condition*

$$\nabla^2 V \geq \alpha I$$

*for some $\alpha > 0$. Further let $\rho_t = \rho(\cdot, t)$ denote the solution to the Fokker-Planck equation (13.2) with initial density $\rho_0 \in L^1(\mathbb{R}^n)$. Then*

$$\|\rho_t - \rho_\infty\|_{L^1(\mathbb{R}^n)} \leq e^{-\alpha \epsilon t} \|\rho_0 - \rho_\infty\|_{L^1(\mathbb{R}^n)}\,.$$

The proof of the theorem is quite involved (e.g., see [11]), and we will prove a slightly simpler version instead. This will suffice to get the idea of the general proof that is based on entropy estimates. We will need the following definition.

**Definition 13.4** (Relative entropy)**.** *Let $\mu$ be a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ that is absolutely continuous with respect to $d\mu_\infty = \rho_\infty dx$. The quantity*

$$H(\mu\|\mu_\infty) = \int_{\mathbb{R}^n} \log\left(\frac{d\mu}{d\mu_\infty}\right) d\mu$$

*is called* relative entropy *or* Kullback-Leibler divergence *of $\mu$ and $\mu_\infty$. We declare that $H = \infty$ if $\mu$ is not absolutely continuous with respect to $\mu_\infty$.*

We use the agreement that $0\log 0 = 0$. It can be shown, using Jensen's inequality, that $H$ is nonnegative, with $H(\mu\|\mu_\infty) = 0$ if and only if $\mu = \mu_\infty$ a.e. Even though $H$ is not symmetric in its arguments, it can be used to bound the variation distance between $\mu$ and $\mu_\infty$: if $\rho$ and $\rho_\infty$ denote the Lebesgue densities with respect to $\mu$ and $\mu_\infty$, then

$$\|\rho - \rho_\infty\|_{L^1(\mathbb{R}^n)} \leq \sqrt{8H(\mu\|\mu_\infty)}\,. \tag{13.3}$$

The inequality is known as *Csiszár-Kullback inequality* [2, 10].

**Lemma 13.5.** *Let $\rho_t = \rho(x, t)$ be the solution of the Fokker-Planck equation for a quadratic potential*

$$V(x) = \frac{\alpha}{2}|x|^2\,, \ \alpha > 0\,.$$

*Then*

$$\|\rho_t - \rho_\infty\|_{L^1(\mathbb{R}^n)} \leq e^{-\alpha \epsilon t} \|\rho_0 - \rho_\infty\|_{L^1(\mathbb{R}^n)}\,.$$

*Proof.* The sketch of proof is based on a Gronwall estimate of the entropy production rate $dH/dt$. Specifically, let (in a slight abuse of notation)

$$H(\rho_t\|\rho_\infty) = \int_{\mathbb{R}^n} \log\left(\frac{\rho_t}{\rho_\infty}\right) \rho_t \, dx\,.$$

Further let $\|\rho_0\|_{L^1(\mathbb{R}^n)} = 1$. Using integration by parts and the norm conservation property of the Fokker-Planck equation, it follows that

$$
\begin{aligned}
\frac{dH}{dt} &= \int_{\mathbb{R}^n} \frac{\partial \rho_t}{\partial t} \log \rho_t \, dx + \frac{\alpha}{2\epsilon} \int_{\mathbb{R}^n} |x|^2 \frac{\partial \rho_t}{\partial t} \, dx \\
&= -\int_{\mathbb{R}^n} (\nabla \log \rho_t) \cdot (\epsilon \nabla \rho_t + \alpha x \rho_t) \, dx - \frac{\alpha}{\epsilon} \int_{\mathbb{R}^n} x \cdot (\epsilon \nabla \rho_t + \alpha x \rho_t) \, dx \\
&= -\int_{\mathbb{R}^n} \left( \epsilon \frac{|\nabla \rho_t|^2}{\rho_t} + 2\alpha x \cdot \nabla \rho_t + \frac{\alpha^2}{\epsilon}|x|^2 \rho_t \right) dx \\
&= -\int_{\mathbb{R}^n} \left| \sqrt{\epsilon}\nabla \log \rho_t + \frac{\alpha}{\sqrt{\epsilon}}x \right|^2 \rho_t \, dx\,,
\end{aligned}
$$

which shows that the entropy $H$ is nonincreasing along solutions of the Fokker-Planck equation, i.e., $dH/dt \leq 0$. (This is to say that $H$ is a Lyapunov function for $\rho_t$.) Now comes a bit of magic: by using a functional inequality for the entropy, called *logarithmic Sobolev inequality* (e.g., see [11]), it can be shown that the last integral can be bounded from above by $-2\alpha\epsilon H$, i.e.,

$$\frac{d}{dt}H(\rho_t\|\rho_\infty) \leq -2\alpha\epsilon H(\rho_t\|\rho_\infty)\,.$$

The inequality is a (functional) differential inequality, and it follows by the Gronwall Lemma on page 41 that

$$H(\rho_t\|\rho_\infty) \leq e^{-2\alpha\epsilon t}H(\rho_0\|\rho_\infty)\,.$$

The Csiszár-Kullback inequality (13.3) then yields the desired result. $\qquad\square$

Most of the rigorous results that prove exponential convergence of $\rho_t$ towards the stationary distribution are for strictly convex potentials. From the properties of the semigroup $S_t$ or its adjoint $S_t^*$ we may nonetheless expect that

$$\|\rho_t - \rho_\infty\| \leq Ce^{-\lambda t}$$

holds true in some suitable norm and for some constants $C, \lambda > 0$. Let us briefly explain why.

## 13.3 Spectral gaps and the space $L^2(\mathbb{R}^n, \mu_\infty)$

We begin with a short excurse on the spectral properties of the generator and its semigroup. To this end, let $L^2(\mathbb{R}^n, \mu_\infty)$ denote the space of functions that are square-integrable with respect to the invariant measure $\mu_\infty$, i.e.,

$$L^2(\mathbb{R}^n, \mu_\infty) = \left\{u\colon \mathbb{R}^n \to \mathbb{R} : \int_{\mathbb{R}^n} |u(x)|^2 \rho_\infty(x)\,dx < \infty\right\}.$$

This is a Hilbert space with scalar product

$$\langle u, v\rangle_{\mu_\infty} = \int_{\mathbb{R}^n} u(x)v(x)\,\rho_\infty(x)\,dx$$

and induced norm $\|\cdot\|_{\mu_\infty}$. It can be readily seen that $L$ is symmetric as an operator on $\mathcal{C}_0^2(\mathbb{R}^n) \cap L^2(\mathbb{R}^n, \mu_\infty)$:

$$\langle Lu, v\rangle_{\mu_\infty} = -\epsilon \int_{\mathbb{R}^n} \nabla u(x)\cdot\nabla v(x)\rho_\infty(x)\,dx = \langle u, Lv\rangle_{\mu_\infty}\,. \qquad (13.4)$$

Technical details aside, any (bounded) symmetric operator on a Hilbert space has real eigenvalues and orthogonal eigenvectors.[10] Moreover, by the above,

$$\langle Lf, f\rangle_{\mu_\infty} = -\epsilon\|\nabla f\|_{\mu_\infty}\,,$$

---

[10] Clearly $L$ here is an unbounded operator, but it can be shown that it is essentially self-adjoint on a suitable subspace of $\mathcal{C}_0^2(\mathbb{R}^n) \cap L^2(\mathbb{R}^n, \mu_\infty)$ which implies that its spectrum is discrete and real, with eigenfunctions that form an orthonormal basis of $L^2(\mathbb{R}^n, \mu_\infty)$.

which proves that $L$ is negative definite on $L^2(\mathbb{R}^n, \mu_\infty)$. The corresponding eigenvalue problem

$$L\phi_n = \lambda_n \phi_n\,, \quad n = 0, 1, 2, \dots$$

is well posed with discrete eigenvalues

$$\lambda_0 > \lambda_1 \geq \lambda_2 \geq \dots\,,$$

where $\lambda_0 = 0$ is simple, and pairwise orthogonal eigenvectors $\{\phi_n\}_{n=0}^\infty$ that (if properly normalized) form an orthonormal basis of $L^2(\mathbb{R}^n, \mu_\infty)$, i.e.,

$$\langle \phi_i, \phi_j \rangle_{\mu_\infty} = \delta_{ij}\,. \tag{13.5}$$

We can now represent any function $f(\cdot, t) \in L^2(\mathbb{R}^n, \mu_\infty)$ in terms of the eigenfunctions by the expansion

$$f(x, t) = \sum_{j=0}^\infty c_j(t)\phi_j(x)\,, \quad c_j(t) = \langle f(\cdot, t), \phi_j \rangle_{\mu_\infty}\,.$$

**Fokker-Planck equation in $L^2(\mathbb{R}^n, \mu_\infty)$**

For our purposes it is convenient to normalize the distribution by $\rho_\infty$; instead of $\rho_t$, we consider the evolution of a function $\psi_t = \psi(x, t)$ defined by

$$\rho(x, t) = \psi(x, t)\rho_\infty(x)\,.$$

Inserting the last expression into the Fokker-Planck equation (13.2) it is easy to show that $\psi_t$ solves the backward Kolmogorov equation

$$\frac{\partial \psi}{\partial t} = L\psi\,, \quad \psi_0 = \rho_0/\rho_\infty\,.$$

We assume that $\psi_0 \in L^2(\mathbb{R}^n, \mu_\infty)$, which is the case whenever $\rho_0 \in L^1(\mathbb{R}^n)$ is continuous and, e.g., compactly supported, such that $\psi_0$ admits the expansion

$$\psi_0(x) = \sum_{j=0}^\infty \langle \psi_0, \phi_j \rangle_{\mu_\infty} \phi_j(x)\,.$$

Also $\psi_t = \exp(tL)\psi_0$ can be expanded into eigenfunctions. This yields

$$\psi_t = \sum_{j=0}^\infty \langle \psi_0, \phi_j \rangle_{\mu_\infty} e^{tL}\phi_j = \sum_{j=0}^\infty \langle \psi_0, \phi_j \rangle_{\mu_\infty} e^{t\lambda_j}\phi_j$$

where we have tacitly assumed that the sum is uniformly convergent and used that $L$ and $\exp(tL)$ have the same eigenvectors $\phi_j$. Now recall that $\phi_0 = 1$ is the unique eigenvector to the eigenvalue $\lambda_0 = 0$, corresponding to the stationary distribution $\rho_\infty$, and observe that

$$\langle \psi_0, \phi_0 \rangle_{\mu_\infty} = \int_{\mathbb{R}^n} \frac{\rho_0}{\rho_\infty} \rho_\infty \, dx = \|\rho_0\|_{L^1(\mathbb{R}^n)} = 1\,.$$

Therefore, with (13.5) and $\psi_\infty = 1$,

$$\|\psi_t - \psi_\infty\|_{L^2(\mathbb{R}^n, \mu_\infty)}^2 = \sum_{j=1}^\infty e^{2\lambda_j t} \langle \psi_0, \phi_j \rangle_{\mu_\infty}^2 \leq e^{2\lambda_1 t}\|\psi_0\|_{L^2(\mathbb{R}^n, \mu_\infty)}^2\,,$$

which shows that $\psi_t \to \psi_\infty$ in $L^2(\mathbb{R}^n, \mu_\infty)$, exponentially fast with rate given by the first eigenvalue or *spectral gap* $\lambda_1 < 0$. In other words,

$$\|\psi_t - \psi_\infty\|_{L^2(\mathbb{R}^n, \mu_\infty)} \le Ce^{\lambda_1 t}, \tag{13.6}$$

with $C = \|\psi_0\|_{L^2(\mathbb{R}^n, \mu_\infty)}$.

**Remark 13.6.** *If we remove the scaling by the invariant density and translate (13.6) into probability densities again, we see that*

$$\int_{\mathbb{R}^n} |\rho_t - \rho_\infty|^2 \rho_\infty^{-1} dx \le Ce^{\lambda_1 t}.$$

*This is not quite what we would like to have as the assumption $C < \infty$ requires that $\rho_0$ is square integrable with respect to $\rho_\infty \propto \exp(V/\epsilon)$; a more natural norm would be $L^1$, but at least our short calculation should have made clear that convergence to equilibrium is related to the first nonzero eigenvalue(s) of the generator L. Unfortunately, one can show that*

$$\lambda_1 \asymp e^{-\Delta V/\epsilon} \quad as \ \epsilon \to 0,$$

*that is, in the presence of large energy barriers, the convergence is rather slow as $\lambda_1 \approx 0$ (even though it is still exponential).*

# 14 Day 14, 29.01.2013: Markov chain approximations

Recall that the probability distribution associated to an SDE of the form

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\varepsilon}dB_t, \quad X_0 \sim \rho_0 \tag{14.1}$$

satisfies for all Borel sets $A \subset \mathbb{R}^n$

$$\lim_{t \to \infty} \mathbb{P}(X_t \in A) = \mu_\infty(A) \quad \text{a.s.}$$

where

$$\mu_\infty(A) = \int_A \rho_\infty(x)dx.$$

As we have seen, the convergence above is exponentially fast with a rate $\lambda_1 < 0$ given by the first nonzero eigenvalue of the generator $L = \varepsilon\Delta - \nabla V \cdot \nabla$. Unfortunately, for small $\epsilon > 0$,

$$\lambda_1 \approx -(\mathbb{E}_x[\tau])^{-1},$$

with $\tau$ being the first exit time from the deepest well in the potential energy landscape, in other words, $\lambda_1$ may be arbitrarily close to zero.

## 14.1 Ergodicity

We can evaluate $\rho_\infty(x) \propto e^{-V(x)/\varepsilon}$ but we cannot sample from it. There are two central issues related to sampling the stationary from simulations of (14.1).

How do we compute expectation value with respect to $\rho_\infty$? Put differently, given a sufficiently long realization $(X_t)_{t \in [0,T]}$ of (14.1), can we assume that

$$\int_{\mathbb{R}^n} f d\mu_\infty \approx \frac{1}{T} \int_0^T f(X_t) \, dt \,,$$

i.e., does $X_t$ sample $\rho_\infty$? And if so, how large must $T$ be? This brings us to the second question: Can we say anything about the small-noise asymptotics for the distribution of $X_t$ and time $t$? If yes, it might tell us something about a reasonable choice of $T$, as the typical relaxation time is exponentially in the eigenvalue that depends on $\epsilon$.[11]

**Remark 14.1.** *In the case of a two-well potential with $V(x) = 0$ at the two minima $x = \pm 1$, it can be shown that*

$$\mu_\infty \rightharpoonup \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1} \,,$$

*i.e., that the invariant density converges (in distribution, or in the weak-$*$ sense) to the weighted sum of two Dirac measures concentrated at the minima.*

**Theorem 14.2** (Strong Law of Large Numbers). *Let $f \in \mathcal{C}_b(\mathbb{R}^n)$. Then (given the usual growth, smoothness and boundedness conditions on $V$),*

$$\mathbb{E}\,[f] = \lim_{T \to \infty} \frac{1}{T} \int_0^T f(X_t) dt \quad a.s.$$

*for almost all initial conditions $X_0 = x$.*

*Proof.* See [6, Thm. 2.5]. □

The interpretation of the above statement is that if the SDE has a unique invariant measure, then almost all realizations converge to the same invariant distribution, in the sense that expectations agree, a property that is known as *ergodicity.*

**Remark 14.3.** *The Law of Large Numbers does not hold for general SDEs. For example, the infinitesimal generator $L$ may have pure imaginary eigenvalues or the invariant density $\rho_\infty$ may not be everywhere positive. In these cases, the system may not be ergodic or there may depend on the initial conditions.*

## 14.2 Markov Chain Monte Carlo (MCMC)

Now we consider two possible methods for sampling $\rho_\infty$, i.e., for computing expectation values of an integrable function $f$.

**First attempt:** We use the Euler-Maruyama scheme

$$\tilde{X}_{n+1} = \tilde{X}_n - \Delta t \nabla V(\tilde{X}_n) + \sqrt{2\varepsilon}\Delta B_{n+1} \,,$$

and take the sample mean

$$\mathbb{E}\,[f] \approx \frac{1}{N} \sum_{i=1}^N f(\tilde{X}_i).$$

We make the following observations:

---

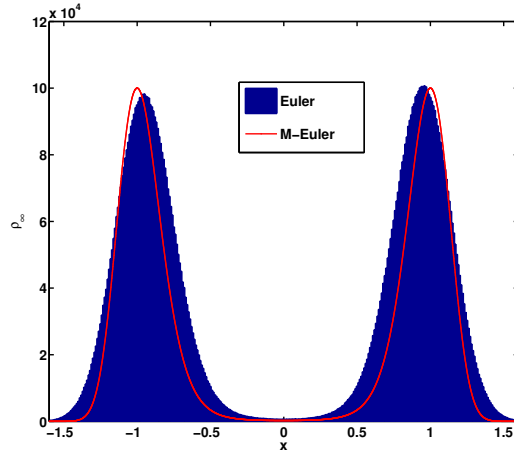[11] We may expect that this choice will have to do with Kramer's formula, equation (11.4).

Figure 14.1: Empirical distribution and exact Boltzmann density (red curve) for the symmetric double-well potential.

(i) We know that on any finite time interval $[0, T]$ the weak error satisfies

$$\max_n |\mathbb{E}[f(X_{t_n})] - \mathbb{E}[f(\tilde{X}_n)]| \leq C\Delta t$$

where $C = Ae^{BT}$ for suitable constants $A, B > 0$ (see Section 9.2). This implies that the error in this attempt is essentially of order 1 as $T \to \infty$.

(ii) We may think of $(\tilde{X}_n)_{n \in \mathbb{N}_0}$ as a homogeneous Markov chain with the transition probability kernel

$$Q(x, A) = \mathbb{P}(\tilde{X}_{n+1} \in A \mid \tilde{X}_n = x)$$

for a Borel set $A \subset \mathbb{R}^n$, with the Gaussian transition density

$$q(x, y) = (4\pi\varepsilon\Delta t)^{-n/2} \exp\left(-\frac{|y - x + \Delta t \nabla V(x)|^2}{4\varepsilon\Delta t}\right).$$

The associated invariant distribution satisfies

$$\tilde{\mu}_\infty(A) = \int_{\mathbb{R}^n} Q(x, A) d\tilde{\mu}_\infty(x) \tag{14.2}$$

or equivalently

$$\tilde{\rho}_\infty(y) = \int_{\mathbb{R}^n} q(x, y)\tilde{\rho}_\infty(x) dx.$$

(Recall that in the discrete-state space setting the invariant distribution is the left eigenvector of the transition matrix to the eigenvalue $\lambda = 1$, i.e., we have $\tilde{\mu}_\infty(x) = \sum_y \tilde{\mu}_\infty(y)q_{xy}$; equation (14.2) is simply the continuous analogue of the last identity.) It is easy to see that (see Fig. 14.1)

$$\mu_\infty(A) \neq \int_{\mathbb{R}^n} Q(x, A) d\mu_\infty(x), \tag{14.3}$$

i.e., $\mu_\infty$ is not invariant under the Euler scheme.[12] In fact for an arbitrary

---

[12] Try $V(x) = x^2$.

67

potential $V$, the Markov chain with transition probability $Q$ must not have an invariant measure at all (i.e., the chain may be transient).

**Remark 14.4.** *It is not advisable to let $\Delta t \to 0$ and $N \to \infty$ in the attempt above to reduce the numerical error, because even with large time steps, simulations take a long time. Decreasing the time steps would make simulations too long to be practical.*

**Second attempt:** By slightly altering the transition rule of the Euler scheme we can construct a Markov chain $(Y_n)_n \geq 0$ that is similar to $(\tilde{X}_n)_{n\geq 0}$ with transition probability $P(x, A)$ satisfying

$$\mu_\infty(A) = \int_{\mathbb{R}^n} P(x, A) d\mu_\infty(x).$$

We will do this by "Metropolizing" the euler scheme. To this end let us recall some facts from the theory for Markov chains:

(i) A sufficient condition for invariance is *detailed balance*, i.e.

$$\rho_\infty(x)p(x, y) = \rho_\infty(y)p(y, x) \quad \forall x, y \in \mathbb{R}^n$$

where $p(x, y)$ is the Lebesgue density of the transition kernel $P(x, \cdot)$. Then

$$\int_{\mathbb{R}^n} \rho_\infty(x)p(x, y)dy = \rho_\infty(x)P(x, \mathbb{R}^n) = \rho_\infty(x) = \int_{\mathbb{R}^n} \rho_\infty(y)p(y, x)dy,$$

where the last equality is exactly the condition that $\rho_\infty$ must fulfill in order to be the invariant density under $P$.

(ii) If the Markov chain is irreducible and aperiodic with strictly positive invariant measure $\mu_\infty > 0$, then the Markov chain is ergodic, i.e. for almost all initial conditions $x$ it follows that

$$\lim_{n\to\infty} \sup_{A\subset\mathbb{R}^n} |P^n(x, A) - \mu_\infty(A)| = 0$$

where $P^n(x, A) := \mathbb{P}(X_n \in A \mid X_0 = x)$ and the supremum runs over all measurable sets $A \subset \mathbb{R}^n$ (total variation norm); e.g., see [13].

## 14.3 Metropolis-Adjusted Langevin Algorithm (MALA)

We combine the Euler scheme for equation (14.1), also known as *Langevin algorithm*, with the Metropolis Hastings scheme.

**Algorithm 14.5.** *MALA proceeds by iterating the following two steps (proposal and acceptance-rejection step):*

1. *Let $Y_n = x$ and generate proposal according to*

$$y = x - \Delta t\nabla V(x) + \sqrt{2\varepsilon}\,\Delta B_{n+1}.$$

*(The proposal step $x \mapsto y$ has transition probability density $q(x, y)$.)*
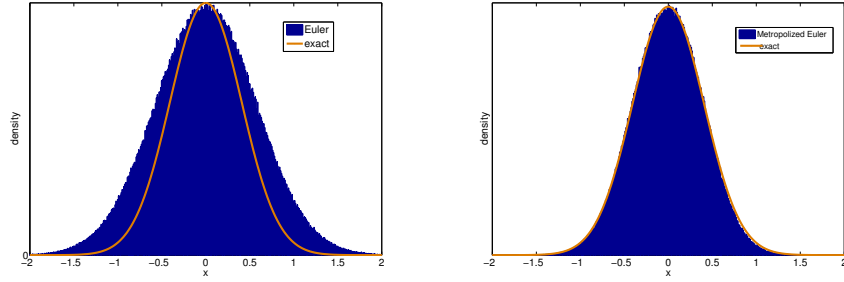
Figure 14.2: Empirical distributions and exact Boltzmann density (orange curve) for the symmetric double-well potential: The left panel shows the systematic bias of Euler's method, the right panel shows the Metropolized simulation.

2. *Accept $Y_{n+1} = y$ with probability*

$$a(x,y) := \min\left\{1, \frac{\rho_\infty(y)q(y,x)}{\rho_\infty(x)q(x,y)}\right\};$$

*otherwise, set $Y_{n+1} = x$.*

The above acceptance criterion yields a Markov chain with transition probability density $p(x,y) = q(x,y)a(x,y)$ for $x \neq y$. The total rejection probability, i.e., the probability that the chain remains at $x$ then is

$$r(x) = 1 - \int_{\mathbb{R}^n} q(x,y)a(x,y)dy$$

The transition kernel of MALA can be expressed as

$$P(x,dy) = q(x,y)a(x,y)dy + \left(1 - \int_{\mathbb{R}^n} q(x,y)a(x,y)dy\right)\delta_x(dy).$$

The Markov chain described by the transition density above satisfies detailed balance with respect to the invariant density $\rho_\infty$ and inherits all the nice properies from the Euler scheme (such as irreducibility and aperiodicity), so the Markov chain is ergodic. In particular, for any stable time step $\Delta t > 0$,

$$\sup_{A \subset \mathbb{R}^n} |P^n(x,A) - \mu_\infty(A)\| \to 0 \quad \text{as } n \to \infty$$

and

$$\mathbb{E}\left[f\right] = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} f(Y_i) \quad a.s.$$

**Example 14.6.** *Figure 14.2 shows a comparison between the ordinary and the Metropolis-adjusted Euler scheme for a quadratic potential. We observe that for a reasonable time step, the Euler scheme shows a systematic bias that is corrected by the acceptance-rejection procedure of MALA.*

In some cases the convergence can be shown to be geometric, which is the analogue of the exponential convergence in the time-continuous case.

**Theorem 14.7** (Geometric ergodicity of MALA). *If $V(x) \sim |x|^q$ for $|x| \to \infty$ with $1 \le q < 2$, then*

$$\sup_{A \subset \mathbb{R}^n} |P^n(x, A) - \mu_\infty(A)| \le C\varrho^n$$

*for some $0 < \varrho < 1$.*

*Proof.* See [16] $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

**Remark 14.8.** *We conclude with some comments on MALA.*

1. *Convergence of MALA holds for all stable step sizes $\Delta t$ and under fairly mild conditions on the potential $V$, basically polynomial growth is enough.*

2. *The speed of convergence does depend on the details of the potential landscape; for potentials that do not fall under the category of Theorem 14.7, the convergence can be shown to be slower than geometric (see [16]).*

3. *Even though MALA converges independently of $\Delta t$, the acceptance probability decreases when $\Delta t$ is increased. Hence there is a trade-off between many integration steps and small rejection rate ($\Delta t$ small) and fewer integration steps with a possibly large rejection rate ($\Delta t$ large).*

# References

[1] Ludwig Arnold. *Stochastische Differentialgleichungen: Theorie und Anwendung.* Wiley & Sons, 1973.

[2] Imre Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Stud. Sci. Math. Hung.*, 2:299–318, 1967.

[3] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications.* Springer, Heidelberg, 1998.

[4] Richard Durrett. *Stochastic Calculus: A Practical Introduction.* CRC Press, 1996.

[5] Wilhelm Huisinga and Eike Meerbach. *Markov chains for everybody.* Lecture Notes, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 2005.

[6] Jonathan C. Mattingly, Andrew Stuart and Desmond J. Higham. Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise. *Stoch. Proc. Appl.*, 101:185–232, 2002.

[7] Ioannis Karatzas and Steven Shreve. *Brownian Motion and Stochastic Calculus.* Springer, New York, 1991.

[8] Achim Klenke. *Wahrscheinlichkeitstheorie.* Springer, Berlin, 2006.

[9] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin, 1992.

[10] Solomon Kullback. A lower bound for discrimination information in terms of variation. *IEEE Trans. Inform. Theory*, 4:126–127, 1967.

[11] Peter A. Markowich and Cédric Villani. On the trend to equilibrium for the fokker-planck equation: an interplay between physics and functional analysis. *Matematica Contemporanea*, 19:1–31, 2000.

[12] David Meintrup and Stefan Schäffler. *Stochastik*. Springer, Berlin, 2005.

[13] Sean P. Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 2005.

[14] Peter Mörters and Yuval Peres. *Brownian Motion (Cambridge Series in Statistical and Probabilistic Mechanics)*. Cambridge University Press, 2010.

[15] Bernt K. Øksendal. *Stochastic Differential Equations: An Introduction With Applications*. Springer, Berlin, 2003.

[16] Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2:341–363, 1996.

[17] Marco Sarich. *Multiscale Stochastic Processes*. Lecture Notes, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 2011.

[18] Dongbin Xiu. *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press, 2010.