UNIVERSITY OF BERGEN Department of Mathematics

Introduction to Stochastic Differential Equations - Applications to numerical analysis and stochastic geometry

Adrien LAURENT and Gianmarco VEGA-MOLINO

Spring 2022

UNIVERSITY OF BERGEN
Faculty of Mathematics and Natural Sciences

Abstract

These notes go with a course on stochastic differential equations, that was given at the university of Bergen in Spring 2022. While we skip some of the details, we hope to give a practical understanding of how SDEs work, and how to manipulate them, as well as a few applications that appear in the recent literature.

We begin by giving an overview of the tools from measure theory and probability theory that we need. We define the Brownian motion, and martingales in general, and use them to define the stochastic integral with respect to a martingale. We give the definition of SDEs, and prove the existence and uniqueness of a solution in the globally Lipschitz context. We play with the Ito formula, the Stratonovich integral, and we say a word about rough paths and SPDEs.

For the applications, we present the necessary tools for studying strong and weak approximations of SDEs, and we give the proofs of convergence of a few numerical schemes. In stochastic geometry, we give the ideas to define stochastic processes on manifolds.

We use the following textbooks and articles for this course:

- measure theory: [3]
- probability theory: [12]
- stochastic analysis: [11, 6, 2]
- numerical analysis: [5, 10, 4, 9]

Acknowledgements. The authors would like to thank all the participants for their helpful comments and their interest in this course. The computations for the high weak order approximations were performed at the University of Geneva on the Baobab cluster using the Julia programming language. The code is available on MittUiB, or on demand.

Contents

Chapter 1

Introduction

An ordinary differential equation (ODEs) is an equation that contains a function $x\colon \mathbb{R} \to \mathbb{R}^d$ and its derivatives. We typically study ODEs of the form

$$x'(t) = f(x(t)), \quad x(0) = x_0, \tag{1.0.1}$$

where f is a given smooth vector field. If f is globally Lipschitz, then there exists a unique solution to this problem. An alternative way of writing (1.0.1) is the *integral formulation*:

$$x(t) = x_0 + \int_0^t f(x(s))ds.$$

A stochastic differential equation (SDE) is a differential equation that includes a random perturbation that we call noise. We could write it as

$$X'(t) = f(X(t)) + g(X(t))\xi(t), \quad X(0) = X_0,$$

where ξ is the noise. The definition of ξ is not straightforward, and one prefers to use a Brownian motion W (with $\xi = dW/dt$). We will write

$$dX(t) = f(X(t))dt + g(X(t))dW_t, \quad X(0) = X_0,$$

which exactly stands for the integral formulation

$$X(t) = X_0 + \int_0^t f(X(s))ds + \int_0^t g(X(s))dW_s.$$

The solution of this equation (if it exists) is a random variable. We have many things to understand. First, what is the Brownian motion W? Second, how can we define the stochastic integral $\int_0^t g(X(s))dW_s$? Finally, under which conditions do we have existence and uniqueness to a SDE.

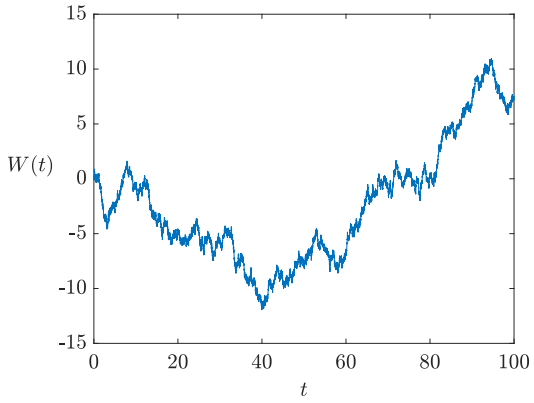
The Brownian motion is a stochastic process. We observe that is has "size $\sqrt{t}$". It is definitely not differentiable in general, but it is continuous.

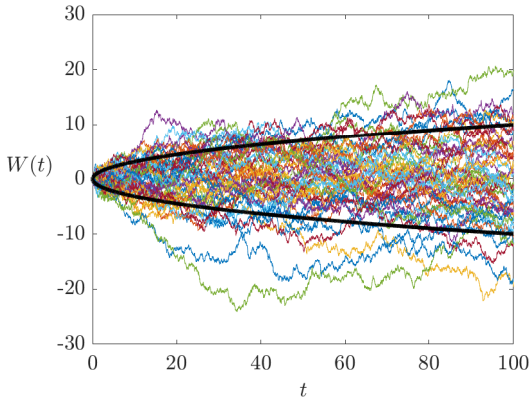
We observe surprising results. One would think that the solution of

$$dX(t) = aX(t)dt + bX(t)dW_t, \quad X(0) = 1,$$

is

$$X(t) = \exp(at + bW_t).$$





This is not true in general. It depends on the definition we take for the stochastic integral. With the Rienmann integral, we know that defining it with left, right or any rectangles yields the same result. This is not the case with the stochastic integral. For the definition we will take of the integral, the solution of the linear SDE is

$$X(t) = \exp(\left(a - \frac{b^2}{2}\right) t + bW_t).$$

At the end of the reading group, we will see how to approximate numerically the solution of a SDE, and we will study stochastic processes on manifolds.

Chapter 2

A quick review of measure theory and probability theory

2.1 Measure theory

2.1.1 The Riemann integral

Recall, given a real interval $[a, b]$ a partition of length n is a sequence $a = x_0 < x_1 < \cdots < x_n = b$, and a tagged partition is a partition along with another sequence $t_1, \ldots, t_n$ such that $t_i \in [x_{i-1}, x_i]$.

For a function $f : \mathbb{R} \to \mathbb{R}$ and a tagged partition P of length n we define the Riemann sum

$$R(f, P) := \sum_{i=1}^{n} f(t_i)(x_i - x_{i-1}).$$

In principle, then, the Riemann integral is the limit of Riemann sums as the mesh

$$\|P\| = \max_i |x_{i+1} - x_i|$$

vanishes. More precisely,

Definition 2.1.1. *Fix a function f and a real number R, and suppose for any $\epsilon > 0$ there exists a $\delta > 0$ such that for any tagged partition P with mesh $\|P\| < \delta$ it holds that*

$$|R(f, P) - s| < \epsilon.$$

Then we say that f is Riemann integrable, s is the Riemann integral of f over the interval $[a, b]$, and write

$$\int_a^b f(x)\, dx = s.$$

Of course, this can be a bit cumbersome to use, and so we consider the equivalent notion of Darboux integral. Fixing a partition P of length n, we define the upper and lower Darboux sums over P

$$U(f, P) := \sum_{i=1}^{n} \sup_{t_i \in [x_{i-1}, x_i]} f(t_i)(x_i - x_{i-1})$$

$$L(f,P) := \sum_{i=1}^{n} \inf_{t_i \in [x_{i-1},x_i]} f(t_i)(x_i - x_{i-1})$$

that is, the upper (lower) Darboux sum over a partition P maximizes (minimizes) the value over each subinterval. Then the upper and lower Darboux sums of f are defined as

$$U(f) := \inf_{P \in \mathcal{P}_{[a,b]}} U(f,P)$$
$$L(f) := \sup_{P \in \mathcal{P}_{[a,b]}} L(f,P)$$

where $\mathcal{P}_{[a,b]}$ denotes the set of all partitions of $[a,b]$. Finally, if $U(f)$ and $L(f)$ are both finite and equal then we define the Darboux integral

$$\int_a^b f(x)\ dx := U(f) = L(f)$$

and say that the function f is Darboux integrable.

Theorem 2.1.2. *A function f is Riemann integrable over an interval $[a,b]$ if and only if it is Darboux integrable over the same interval, and the integrals are always equal.*

Theorem 2.1.3. *Suppose f is Riemann integrable over the interval $[a,b]$. Let P_n be any sequence of tagged partitions of $[a,b]$ with the property that $\|P_n\| \xrightarrow{n\to+\infty} 0$. Then*

$$\lim_{n\to+\infty} R(f,P_n) \to \int_a^b f(x)\ dx$$

The key point to take from the preceding theorem is the following: if f is Riemann integrable, the choice of tags does not matter to the computation of the integral so long as the mesh of the partition vanishes.

Riemann-Stieltjes Integral

We want to generalize the notion of integration by putting different "weight" at different points on the domain. That is, we want to know when it's reasonable to define

$$\int_a^b f(x)\ dg(x) = \lim_{\|P_n\|\to 0} \sum_{i=1}^{n} f(t_i)(g(x_i) - g(x_{i-1}))$$

for some tagged partition P_n.

In the case that g is continuous and monotonic there is no issue; however it is possible for this expression to be ill-defined when g oscillates too much.

Definition 2.1.4. *Let $g\colon \mathbb{R} \to \mathbb{R}$. Then the <u>total variation</u> of g is given by*

$$V_a^b(g) := \sup_{P \in \mathcal{P}_{[a,b]}} \sum_{i=1}^{n} |g(x_i) - g(x_{i-1})|$$

where $\mathcal{P}_{[a,b]}$ denotes all partitions of $[a,b]$.

We say g is of <u>bounded variation</u> if $V_a^b(g) < \infty$.

The formal definition of the Riemann-Stieltjes integral is analogous to that of the Riemann integral, and we have the following standard existence result.

Theorem 2.1.5. *If f is continuous and g is of bounded variation, then the Riemann-Stieltjes integral*

$$\int_a^b f(x)\ dg(x)$$

is defined.

Remark 2.1.6. *The proceeding can be strengthened; so long as the set of discontinuities of f is discrete and disjoint from the set of discontinuities of g then the integral will be defined.*

Quadrature formulas

In numerical analysis, we use quadrature formulas to approximate integrals

$$\int_0^1 f(x)dx \simeq \sum_{k=0}^{K} \alpha_k f(x_k).$$

We say that the quadrature is of order p if it is exact for f polynomial of degree at most $p-1$. For instance, the left rectangle and right rectangle quadratures have order one, and the midpoint rule has order two. We divide the interval $[a,b]$ in N subintervals of equal size. We apply the quadrature formula on each subinterval, and we obtain, under regularity assumptions on f, that

$$\lim_{N\to\infty} \sum_{n=0}^{N-1} \sum_{k=0}^{K} \alpha_k f(a+(b-a)/N(n+x_k)) = \int_0^1 f(x)dx.$$

It means that whatever the quadrature I take, it always converges to the integral of f. This is not the case with the stochastic integral.

2.1.2 Measure Spaces

We can consider the integration of a function f over the interval $[a,b]$ to be a way of "measuring" the interval. In order to generalize that, we recall the fundamentals of measure theory.

First, given a space X, we recall that a σ-algebra $\mathcal{A}$ of X is a collection of subsets of X containing the empty set and closed under complements and countable unions. We note that the following stronger statements follow from the definition for a σ-algebra:

- $\mathcal{A}$ contains both X and the empty set.
- $\mathcal{A}$ is closed under countable intersections and countable unions.

We call a space X equipped with a σ-algebra $\mathcal{A}$ a measurable space; a map $\mu\colon \mathcal{A} \to [0,+\infty]$ such that

- $\mu(\varnothing) = 0$
- (σ-additive) $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \sum_{i=1}^{\infty} \mu(A_i)$ for any $A_i \in \mathcal{A}$

is called a (σ-additive) measure, and the triple $(X,\mathcal{A},\mu)$ is called a measure space.

Remark 2.1.7. *Given a measure space $(X, \mathcal{A}, \mu)$ one can define its completion $(X, \bar{\mathcal{A}}, \bar{\mu})$ such that*

- *$\mathcal{A} \subseteq \bar{\mathcal{A}}$ with $\bar{\mu}(A) = \mu(A)$ for all $A \in \mathcal{A}$,*
- *If $\mu(A) = 0$ for some $A \in \mathcal{A}$, then every set $B \subseteq A$ is in $\bar{\mathcal{A}}$ and $\bar{\mu}(B) = 0$.*

A measure space such that $(X, \mathcal{A}, \mu) = (X, \bar{\mathcal{A}}, \bar{\mu})$ is said to be complete, and we will always assume that any measure space we work with is complete.

Note: there are many generalizations of these notions, but we will work exclusively with σ-additive measures, and frequently on probability spaces. That is, a measure space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space if $\mathbb{P}(\Omega) = 1$, in which case we say

- Ω is a sample space,
- $\mathcal{F}$ is a set of events,
- $\mathbb{P}$ is a probability measure.

Measurable functions

Suppose $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ are measurable spaces, and let $f\colon X \to Y$. We say that f is measurable if for all $B \in \mathcal{B}$ it holds that $f^{-1}(B) \in \mathcal{A}$.

In particular, we consider the case $Y = \mathbb{R}$ and $\mathcal{B} = \mathcal{B}(\mathbb{R})$ the Borel σ-algebra. That is, $\mathcal{B}(\mathbb{R})$ is the smallest σ-algebra containing the open sets of $\mathbb{R}$.

Theorem 2.1.8. *The Borel σ-algebra is generated by any of the following sets:*

- $\{(-\infty, b), b \in \mathbb{R}\}$
- $\{(-\infty, b], b \in \mathbb{R}\}$
- $\{(a, \infty), a \in \mathbb{R}\}$
- $\{[a, \infty), a \in \mathbb{R}\}$

Lebesgue measure

We consider the specific case on $\mathbb{R}$. For an interval $[a, b]$, the reasonable choice of measure is clearly $l([a, b]) := b - a$. We define the Lebesgue outer measure as

$$\mathcal{L}^*(E) = \inf\left\{\sum_i l(I_i)\colon E \subseteq \bigcup_i I_i, I_i \text{ is a countable collection}\right\}$$

There is the Carathéodory condition for a set E that for any set A

$$\mathcal{L}^*(A) = \mathcal{L}^*(A \cap E) + \mathcal{L}^*(A \cap E^c)$$

Theorem 2.1.9. *The Lebesgue measurable collection $\mathcal{A}$ of subsets of $\mathbb{R}$ that satisfy the Carathéodory condition is a σ-algebra, and the triple $(\mathbb{R}, \mathcal{A}, \mathcal{L})$ is a measure space. We call $\mathcal{L} = \mathcal{L}^*|_{\mathcal{A}}$ the Lebesgue measure.*

2.1.3 Construction of the Lebesgue integral

Given a measure space $(X, \mathcal{A}, \mu)$ we can define a notion of integration determined by the measure.

1. For the indicator function
$$\chi_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$
we want to set
$$\int_X \chi_A \, d\mu := \mu(A)$$
which we extend to simple functions of the form
$$f = \sum_i \alpha_i \chi_{A_i}$$
as
$$\int_X f \, d\mu := \sum_i \alpha_i \int_X \chi_{A_i} \, d\mu = \sum_i \alpha_i \mu(A_i)$$
for countable families of measurable sets $A_i \in \mathcal{A}$.

2. We extend to non-negative, measurable functions f by
$$\int_X f \, d\mu = \sup_{\tilde{f} \in \mathbf{F}} \int_X \tilde{f} \, d\mu$$
where
$$\mathbf{F} = \{\tilde{f} \colon \tilde{f}(x) \leqslant f(x), \tilde{f} \text{ simple}\}$$
denotes the set of simple functions bounded above by f.

3. For a general measurable function f, write $f = f^+ - f^-$ where f^+ and f^- are non-negative measurable functions, then define
$$\int_X f \, d\mu = \int_X f^+ \, d\mu - \int_X f^- \, d\mu.$$

The nontrivial part of this is to show that the second step is well-defined for any measurable function.

Lebesgue integration vs Riemann integration

We have the important result showing that the Lebesgue integral is strictly a strengthening of the Riemann integral.

Theorem 2.1.10. *Let $f \colon \mathbb{R} \to \mathbb{R}$ be Riemann integrable. Then it is also Lebesgue integrable using the Lebesgue measure on $\mathbb{R}$, and the integrals agree. That is, for $a < b$,*
$$\int_a^b f(x) \, dx = \int_{[a,b]} f(x) \, d\mathcal{L}(x)$$

2.2 Probability theory

2.2.1 Probability space, random variables

Definition 2.2.1. *A probability space is a measure space* $(\Omega, \mathcal{F}, \mathbb{P})$ *with* $\mathbb{P}(\Omega) = 1$.

The sets A in $\mathcal{F}$ are called events, and $\mathbb{P}(A)$ is the probability of the event A. If $\mathbb{P}(A) = 1$, we say that A holds almost surely (or a.s.).

We write $(E, \mathcal{E})$ a measurable space (typically $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$).

Definition 2.2.2. *A function* $X : \Omega \to E$ *is a random variable is it measurable, that is,*

$$\forall A \in \mathcal{E}, X^{-1}(A) \in \mathcal{F}.$$

Definition 2.2.3. *Let* $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ *be a random variable. The law of* X *is the probability measure given by*

$$\mu_X(A) = \mathbb{P}(X \in A) = \mathbb{P} \circ X^{-1}(A).$$

Then, $(E, \mathcal{E}, \mu_X)$ *is a probability space.*

Example: Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and X be the result of a D6. Then, with $E = \Omega$ and $\mathcal{E} = \mathcal{F}$, X is a random variable. It satifies $\mathbb{P}(X = i) = 1/6$, and it defines μ_X completely:

$$\mu_X = \frac{1}{6}(\delta_1 + \cdots + \delta_6),$$

where δ_x is the probability measure satisfying $\delta_x(A) = \mathbb{1}_{\{x \in A\}}$.

When the law of a random variable can be written with a countable number of δ_x, it is called a discrete random variable. Let us now see the continuous random variables.

Definition 2.2.4. *A random variable* $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *has a density w.r.t the Lebesgue measure* λ *if there exists* $\rho : \Omega \to \mathbb{R}^d$ *s.t.*

$$\mu_X(A) = \int_A \rho(x) d\lambda(x), A \in \mathcal{B}(\mathbb{R}^d).$$

We also write it $d\mu_X = \rho d\lambda$.

Example: A few densities (+drawings):

- the uniform law $\mathcal{U}([a, b])$, $\rho(x) = \frac{1}{b-a}\mathbb{1}_{\{x \in [a,b]\}}$,
- the exponential distribution $\mathcal{E}(\lambda)$, $\rho(x) = \lambda e^{-\lambda x}\mathbb{1}_{\{x \geqslant 0\}}$,
- the Gaussian distribution $\mathcal{N}(m, \sigma^2)$, $\rho(x) = \frac{e^{-(x-m)^2/2\sigma^2}}{\sqrt{2\pi}\sigma}$.

2.2.2 Independence

Definition 2.2.5. *Two events A, $B \in \mathcal{F}$ are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. For a family $(A_i)_{i\in I}$, it generalizes in $\mathbb{P}(\cap_{i\in J} A_i) = \prod_{i\in J} \mathbb{P}(A_i)$, for all $J \subset I$ finite. The σ-algebras $(\mathcal{F}_i)$ are independent if for all $A_i \in \mathcal{F}_i$, the events $(A_i)_{i\in I}$ are independent. The random variables (X_i) are independent if the σ-algebras $\sigma(X_i)$ are independent.*

Note that if we take multiple random variables X_i defined on the same measure space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $(E_i, \mathcal{E}_i)$, then $X = (X_1, \dots, X_n)$ takes values in $(E_1 \times \cdots \times E_n, \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n)$. The X_i are called the marginals of X.

Proposition 2.2.6. *The random variables $X_1, \dots, X_n$ are independent if and only if*

$$\mu_(X_1, \dots, X_n) = \mu_{X_1} \dots \mu_{X_n}.$$

Equivalently, if $X = (X_1, \dots, X_n)$ has a density $rho_X(x_1, \dots, x_n) = \rho_1(x_1) \dots \rho_n(x_n)$, then the random variables $X_1, \dots, X_n$ are independent.

Example: If $\rho(x, y) = \frac{1}{3\pi} \exp(-(x^2 + 2xy + 5y^2)/6)$, the density of X is given by $\rho_X(x) = \int \rho(x, y)dy = \frac{1}{\sqrt{15\pi/2}} e^{-4x^2/30}$. Similarly one obtains for Y, $\rho_Y(y) = \frac{1}{\sqrt{3\pi/2}} e^{-4y^2/6}$. As $\rho(x, y) \neq \rho_X(x)\rho_Y(y)$, the random variables are not independent.

2.2.3 Expectation, variance, characteristic function

Definition 2.2.7. *Given a random variable X and a function ϕ, the expectation of $\phi(X)$ is the Lebesgue integral of $\phi(X)$ with respect to the probability measure $\mathbb{P}$. If $\mathbb{E}[|\phi(X)|] < \infty$, the integral exists and we have*

$$\mathbb{E}[\phi(X)] = \int_\Omega \phi(X(\omega))d\mathbb{P}(\omega) = \int_E \phi(x)d\mu_X(x).$$

The variance of X is the quantity $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Exercise: If $X \sim b(p)$, $\mathbb{E}[X] = p$ and $\mathrm{Var}(X) = p(1-p)$. If $X \sim \mathcal{N}(m, \sigma^2)$, its expectation is m and its variance is σ^2.

Proposition 2.2.8. *If $X_1, \dots, X_n$ are independent r.v. such that $\mathbb{E}[|\phi_i(X_i)|] < \infty$ (resp. $\mathrm{Var}(X_i) < \infty$), we have*

$$\mathbb{E}[\phi_1(X_1) \dots \phi_n(X_n)] = \mathbb{E}[\phi_1(X_1)] \dots \mathbb{E}[\phi_n(X_n)]$$

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

Lemma 2.2.9 (Chebyshev's inequality)**.** *If X is a positive random variable and $p \geqslant 1$, then*

$$\mathbb{P}(X \geqslant \lambda) \leqslant \frac{1}{\lambda^p}\mathbb{E}[X^p]$$

Exercise: prove it!

Definition 2.2.10. *The characteristic function of a random variable X taking values in $\mathbb{R}^d$ is*

$$\varphi_X(\lambda) = \mathbb{E}[e^{i\lambda\cdot X}].$$

The characteristic function is the equivalent of the Fourier transform in probability.

Proposition 2.2.11.

- *If X and Y share the same characteristic function, they have the same law.*
- *If $X_1,\dots, X_n$ are independent, then*

$$\varphi_{X_1+\cdots+X_n}(\lambda) = \varphi_{X_1}(\lambda)\dots\varphi_{X_n}(\lambda).$$

- *If X is real valued, then $\varphi_X^{(k)}(0) = i^k\mathbb{E}[X^k]$.*

Exercise: prove the second and third points.

2.2.4 Modes of convergence

Definition 2.2.12. *Here are the different modes of convergence :*

- *(X_n) converges to X almost surely if there exists $A \in \mathcal{F}$ with $\mathbb{P}(A) = 1$ such that for all $\omega \in A$, $X_n(\omega) \to X(\omega)$. We write $X_n \to X$ a.s..*
- *(X_n) converges to X in L^p if $\mathbb{E}[|X_n - X|^p] \to 0$. We write $X_n \xrightarrow{L^p} X$.*
- *(X_n) converges to X in probability if for all $\delta > 0$, $\mathbb{P}(|X_n - X| > \delta) \to 0$. We write $X_n \xrightarrow{\mathbb{P}} X$.*
- *(X_n) converges to X in law if $\mu_{X_n}(A) \to \mu_X(A)$ for all $A \in \mathcal{E}$. We write $X_n \Rightarrow X$.*

Proposition 2.2.13. *almost sure $\Rightarrow \mathbb{P} \Rightarrow \mathcal{L}$ and $L^\infty \Rightarrow \cdots \Rightarrow L^1 \Rightarrow \mathbb{P}$.*

Proposition 2.2.14.

- *If $X_n \to X$ a.s., and f is continuous, then $f(X_n) \to f(X)$ a.s. (resp. in $\mathbb{P}$, in L^p).*
- *There exists a metric on the convergence in probability.*
- *If $X_n \xrightarrow{\mathbb{P}} X$, there exists a subsequence that converges a.s. to X.*
- *If $X_n \xrightarrow{\mathbb{P}} X$ and (X_n) is uniformly integrable, then $X_n \xrightarrow{L^1} X$.*
- *If $X_n \Rightarrow X$ and X is a constant, then $X_n \xrightarrow{\mathbb{P}} X$.*

The Borel-Cantelli lemma is a useful tool to go from a convergence in probability to an almost sure convergence.

Lemma 2.2.15 (Borel-Cantelli)**.** *Let (A_i) be a sequence of events in $\mathcal{F}$, and*

$$A = \{\omega \in \Omega,\ \text{there exists an infinity of } i \text{ s.t. } \omega \in A_i\}.$$

If $\sum \mathbb{P}(A_i) < \infty$, then $\mathbb{P}(A) = 0$.

Theorem 2.2.16. *The following assertions are equivalent:*

- $X_n \Rightarrow X$
- *for all* $g\colon \mathbb{R} \to \mathbb{R}$ *continuous and bounded,* $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$.
- $\varphi_{X_n}(\lambda) \to \varphi_X(\lambda)$

Example: if $\mathbb{P}(X_n = 1 + 1/n) = 1/2$ and $\mathbb{P}(X_n = 0) = 1/2$, prove in different ways that $X_n \Rightarrow b(1/2)$.

2.2.5 Gaussian random variables, Gaussian vectors and limit theorems

Definition 2.2.17. *A real random variable* X *is a standard Gaussian if it has the following density with respect to the Lebesgue measure:*

$$f_X(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

We write $X \sim \mathcal{N}(0,1)$ *in this case, and* $X \sim \mathcal{N}(\mu, \sigma^2)$ *if* $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

Exercise: Using the characteristic function $\varphi_X(\lambda) = \mathbb{E}[\exp(i\lambda X)]$, show that the sum of independent Gaussians is a Gaussian. We recall that φ describes completely random variables, that is, if $\varphi_X = \varphi_Y$, then X and Y follow the same law.

Definition 2.2.18. *A random vector* $X \in \mathbb{R}^d$ *is a Gaussian vector if for all* $a \in \mathbb{R}^d$, $a \cdot X$ *is Gaussian.*

Proposition 2.2.19. *A Gaussian vector* X *is uniquely determined by its expectation* $\mu = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])$ *and its covariance matrix* $\Gamma = (\mathrm{Cov}(X_i, X_j))_{ij}$. *If* Γ *is positive definite, the associated density is*

$$f_X(x) = \frac{e^{-(x-\mu)^T \Gamma^{-1}(x-\mu)/2}}{(2\pi)^{d/2}\sqrt{\det(\Gamma)}}.$$

Example 2.2.20. *If* $X_1, \dots, X_d$ *are independent standard Gaussians,* X *is a random vector and* $X \sim \mathcal{N}(0, I_d)$.

Proposition 2.2.21. *Let* (X_n) *be a sequence of Gaussians* $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ *with* $\sigma_n > 0$, *such that* $X_n \Rightarrow X$. *Then,* X *is Gaussian and* $X \sim \mathcal{N}(\lim \mu_n, \lim \sigma_n^2)$. *If in addition,* $X_n \xrightarrow{\mathbb{P}} X$, *then* $X_n \xrightarrow{L^p} X$.

Exercise: Admitting that $\lim \mu_n$ and $\lim \sigma_n^2$ exist, prove the convergence in law with the help of the characteristic function.

Let us now look at the law of large numbers.

Theorem 2.2.22 (Law of Large Numbers)**.** *Let* (X_n) *be a sequence of iid random variables. If* $\mathbb{E}[|X_1|] < \infty$, *then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}[X_1] \quad a.s.$$

Application: (Monte-Carlo estimators) Let (X_n) be a sequence of iid uniform random variables on B, a measurable bounded domain in $\mathbb{R}^d$. Let f be a map such that $f(X_1) \in L^1$, then

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to \frac{1}{\lambda(B)}\int_B f d\lambda \quad a.s.$$

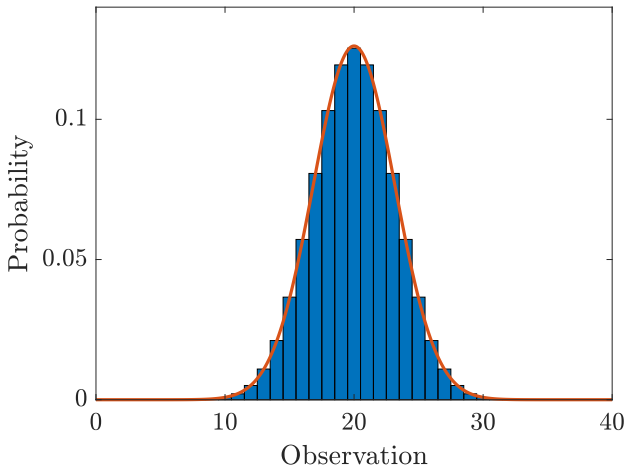
We can approximate integrals with this kind of estimators. The speed of convergence is usually much slower than for quadrature formulas in low dimension. If we work in high dimension, it is the class of methods we have. Moreover, Monte-Carlo estimators do not assume much regularity assumptions on the function f.

Theorem 2.2.23 (Central Limit Theorem)**.** *Let (X_n) be real iid L^2 random variables with $\mu = \mathbb{E}[X_n]$ and $\sigma^2 = \mathrm{Var}(X_n)$. Let $S_N = \frac{1}{N}\sum_{n=1}^{N} X_n$, then we have*

$$\frac{\sqrt{N}(S_N - \mu)}{\sigma} \Rightarrow \mathcal{N}(0,1).$$

Proof. Use the characteristic function, apply a Taylor expansion up to order two, and take the limit. □

Exercise: (Moivre-Laplace) Take $X_n \sim b(1/2)$ and prove the CLT in this context, with the help of the characteristic functions (Hint: Find the law of S_N).



Chapter 3

Stochastic processes, martingales and Brownian motion

3.1 Stochastic processes

Definition 3.1.1. *Let $T \in \mathbb{R}$ be an index set (typically $\mathbb{R}^+$, $[0,T]$ or $\mathbb{N}$) and $(X(t))_{t\in T}$ be a collection of random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in a measurable space $(E, \mathcal{E})$ (typically $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$). Then, X is a stochastic process.*

The map X is actually a map $(t,\omega) \in T \times \Omega \mapsto X(t,\omega)$. $X(.,\omega)$ is called a trajectory, realization or a path. If $T \in \mathbb{Z}$, X is a discrete stochastic process.

Definition 3.1.2. *Given a stochastic process $X(t)$, we denote $\mathcal{F}_t = \sigma(X(s), 0 \leqslant s \leqslant t)$. If $\mathcal{F}_s \subset \mathcal{F}_t$ for $s \leqslant t$, then $\{\mathcal{F}_t, t \in T\}$ is called a filtration associated to X.*

The filtration represents the information we have up to time t. The more the time moves on, the more information we have.

Definition 3.1.3. *A stochastic process $X(t)$ is adapted to the filtration $(\mathcal{F}_t)$ if $X(t)$ is $\mathcal{F}_t$-measurable.*

It makes sense to say that 2 random variables are the same if they are equal almost everywhere. For the equivalence of stochastic processes, we have multiple choices.

Definition 3.1.4. *Let $(X(t))$ and $(Y(t))$ be two stochastic processes indexed by T, defined respectively on $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\Omega', \mathcal{F}', \mathbb{P}')$, and taking values in a measurable space $(E, \mathcal{E})$.*

- *X and Y are equivalent if for all $t_1, \ldots, t_n \in T$, the marginals $(X(t_1), \ldots, X(t_n))$ and $(Y(t_1), \ldots, Y(t_n))$ have the same law.*
- *Y is a modification of X if for all $t \in T$, $\mathbb{P}(X(t) \neq Y(t)) = 0$.*
- *X and Y are indistinguishable if $\mathbb{P}(\sup_{t\in T} |X(t) - Y(t)|_E = 0) = 1$.*

We claim that $iii) \Rightarrow ii) \Rightarrow i)$. If T is a countable set, then $iii) \Leftrightarrow ii)$.

Example 3.1.5. *Let $Y = 0$ and $X = \mathbb{1}_{\omega=t}$ with $T = \Omega = [0,1]$ and the Lebesgue measure. Then, for t fixed, $\mathbb{P}(X(t) \neq Y(t)) = \mathbb{P}(\{\omega, \omega = t\}) = \mathbb{P}(\{t\}) = 0$, and Y is a modification of X. However, we have*

$$\mathbb{P}(\sup_{t\in T} |X(t) - Y(t)| = 0) = 0,$$

that is X and Y are not indistinguishable.

Theorem 3.1.6 (Criterion of Kolomogorov)**.** *Let $(X(t))_{t\in[0,T]}$ be a stochastic process taking values in $(E,\mathcal{E})$ such that there exists C, $\varepsilon>0$, $\delta>1$, s.t. for all s, t,*

$$\mathbb{E}[|X(t)-X(s)|_E^{\delta}] \leqslant C\,|t-s|^{1+\varepsilon}\,.$$

Then, for all $\alpha\in[0,\varepsilon/\delta[$, there exists a continuous modification Y of X whose trajectories $Y(.,\omega)$ are α-Hölderian on $[0,T]$ almost surely.

3.2 Conditional expectation and martingales

Definition 3.2.1. *Let X be a random variable on a probability space $(\Omega,\mathcal{F},\mathcal{P})$, and let $\mathcal{G}\in\mathcal{F}$ be a σ-algebra. The conditional expectation $\mathbb{E}[X|\mathcal{G}]$ of X given $\mathcal{G}$ is a random variable Z such that*

- *Z is $\mathcal{G}$-measurable,*
- *$\mathbb{E}[X\mathbb{1}_A]=\mathbb{E}[Z\mathbb{1}_A]$ ($\int_A X d\mathbb{P}=\int_A Z d\mathbb{P}$) for all $A\in\mathcal{G}$.*

If Y is a random variable on the same probability space, we define the conditional expectation $\mathbb{E}[X|Y]$ of X given Y as $\mathbb{E}[X|\sigma(Y)]$.

Remark 3.2.2. *If $X\in L^2(\Omega,\mathcal{F})$, the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ can be understood as the projection of X on the vector space $L^2(\Omega,\mathcal{G})$. It solves the least squares problem:*

$$\|\mathbb{E}[X|\mathcal{G}]-X\|^2=\min_{Y\in L^2(\Omega,\mathcal{G})}\|Y-X\|^2\,.$$

Example 3.2.3. *For discrete random variables, the conditional expectation is given by*

$$\mathbb{E}[X|Y=y]=\sum x_i\mathbb{P}(X=x_i|Y=y)=\sum x_i\frac{\mathbb{P}(X=x_i,Y=y)}{\mathbb{P}(Y=y)}.$$

For instance, let us play a game with two dice. X and Y are respectively the values of the first two dice, and $Z=X+Y$. The law of $X|Y=2$ is $\mathcal{U}(\{1,\dots,6\})$. The law of $Z|Y=2$ is $\mathcal{U}(\{3,\dots,8\})$. The law of $X|Z=5$ is $\mathcal{U}(\{1,\dots,4\})$. The conditional expectation $-\mathbb{E}[X|Z=5]=2.5$.

For continuous random variables, the density of $X|Y$ is

$$\rho_{X|Y=y}(x)=\frac{\rho_{(X,Y)}(x,y)}{\rho_Y(y)}.$$

It is indeed a density as

$$\int\rho_{X|Y=y}(x)dx=1.$$

Proposition 3.2.4. *If $X\in L^1(\Omega,\mathcal{F})$, the conditional expectation exists and is unique up to sets of measure zero. In addition, the conditional expectation satisfies the following properties:*

- *$X\mapsto\mathbb{E}[X|\mathcal{G}]$ is linear a.s.,*
- *if X is $\mathcal{G}$-measurable and $XY\in L^1$, $\mathbb{E}[X|\mathcal{G}]=X$ and $\mathbb{E}[XY|\mathcal{G}]=X\mathbb{E}[Y|\mathcal{G}]$ a.s.,*
- *$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]]=\mathbb{E}[X]$,*

- *If X and Y are independent, $\mathbb{E}[X|Y] = \mathbb{E}[X]$ a.s.,*
- *If $X \leqslant Y$ a.s., $\mathbb{E}[X|\mathcal{G}] \leqslant \mathbb{E}[Y|\mathcal{G}]$ a.s..*

Definition 3.2.5. *Let X be an adapted stochastic process on $T \subset \mathbb{R}$ such that $\mathbb{E}[|X(t)|] \leqslant \infty$ for all $t \in T$. Then, $X(t)$ is a martingale if*

$$X(s) = \mathbb{E}[X(t)|\mathcal{F}(s)] \quad a.s., \quad s \leqslant t.$$

$X(t)$ is a submartingale if

$$X(s) \leqslant \mathbb{E}[X(t)|\mathcal{F}(s)] \quad a.s., \quad s \leqslant t.$$

$X(t)$ is a supermartingale if

$$X(s) \geqslant \mathbb{E}[X(t)|\mathcal{F}(s)] \quad a.s., \quad s \leqslant t.$$

The term martingale appeared in casinos when creating gambling strategies. Indeed it typically represents a fair game.

Example 3.2.6. *We toss a coin multiple times. If the result is head, we gain 1 NOK, and we lose 1 NOK if its tail. We denote X_n the result of the n-th throw, i.e. $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = 0.5$. Let us write $S_n = X_1 + \cdots + X_n$ the total sum we got or lost at the n-th step. Then, S_n is integrable, and*

$$\begin{aligned}\mathbb{E}[S_{n+p}|S_1, \ldots, S_n] &= \mathbb{E}[S_{n+p}|X_1, \ldots, X_n] \\ &= \mathbb{E}[X_1 + \cdots + X_n|X_1, \ldots, X_n] + \mathbb{E}[X_{n+1} + \cdots + X_{n+p}|X_1, \ldots, X_n] \\ &= X_1 + \cdots + X_n + \mathbb{E}[X_{n+1} + \cdots + X_{n+p}] \\ &= S_n.\end{aligned}$$

Thus (S_n) is a martingale.

We should use local martingales for the construction of the integral, but we will use martingales for the sake of simplicity (any martingale is a local martingale).

Definition 3.2.7. *An adapted stochastic process $(M_t)_t$ is a local martingale if there exists a sequence of stopping time τ_n such that $\tau_n \to \infty$ and $(M_{\min(t,\tau_n)})_t$ is a uniformly integrable martingale.*

Exercise (Gambler's ruin): Let $X_1, \ldots, X_n$ be iid variables such that $\mathbb{P}(X_1 = 1) = p$ and $\mathbb{P}(X_1 = -1) = 1 - p = q$. The random walk (S_n) on $\{0, ..., N\}$ is defined as follows. For $i \geqslant 1$, as long as S_{i-1} is different from 0 and N, its i-jump is X_i. Values 0 and N are called absorbing for the walk, meaning that S_n is stationary if it reaches 0 or N. This example was first studied by de Moivre, who was modeling a gambler who would bet 1 NOK at every hand until either he is ruined or he reaches a total fortune of N. De Moivre's martingale is defined by $Y_n = (q/p)^{S_n}$.

- Check that Y_n is indeed a martingale. What is the associated σ-algebra?
- Compute $\mathbb{P}(\text{walk is absorbed at } 0|S_0 = k)$.

Definition 3.2.8. *A stopping time $T\colon \Omega \to [0,\infty[$ relative to the filtration $\mathcal{F}_t$ is a map such that for every t, $\{T \leqslant t\} \in \mathcal{F}_t$.*

Theorem 3.2.9 (Doob's stopping time)**.** *Let X be a martingale, and T a stopping time, both with respect to the filtration $\mathcal{F}_t$. If one of the three conditions is satisfied:*

- *T is bounded,*
- *X is bounded and $T < \infty$ a.s.*
- *$T \in L^1$ and there exists a real K such that for all n,*

$$|X_n - X_{n-1}| \leqslant K \quad a.s.$$

Then, we have

$$\mathbb{E}[X_T] = \mathbb{E}[X_0].$$

3.3 The Brownian motion

In 1827, Robert Brown observes stochastic dynamics in the motion of pollen in water. In 1900, Bachelier develops a first theory of stochastic analysis and applies it to describe fluctuations in stock prices. In 1905, Einstein defines the Brownian motion.

We want to model the motion of particles in 1D. We denote $f(x,t)$ the density of particles at x and time t, $\rho_x(y,\tau)$ the probability that a particle in x moves to $x+y$ in a time τ. We thus have

$$f(x,t+\tau) = \int_{-\infty}^{+\infty} f(x-y,t)\rho_x(y,\tau)dy.$$

A Taylor expansion gives

$$\begin{aligned} f(x,t+\tau) = f(x,t)\overbrace{\int_{-\infty}^{+\infty}\rho_x(y,\tau)dy}^{=1} + \frac{\partial f}{\partial x}(x,t)\overbrace{\int_{-\infty}^{+\infty} y\rho_x(y,\tau)dy}^{=0} \\ + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x,t)\int_{-\infty}^{+\infty} y^2\rho_x(y,\tau)dy + \dots \end{aligned}$$

We write

$$D = \lim_{\tau\to 0}\int_{-\infty}^{+\infty} y^2\frac{\rho_x(y,\tau)}{\tau}dy.$$

Then f satisfies the PDE

$$\frac{\partial f}{\partial t}(x,t) = \frac{D}{2}\frac{\partial^2 f}{\partial x^2}(x,t),$$

whose solution is $f(x,t) = \frac{1}{\sqrt{2\pi Dt}}e^{-\frac{x^2}{2Dt}}$, i.e. the probability distribution of a Gaussian random variable $\mathcal{N}(0,Dt)$.

Definition 3.3.1. *A real-valued stochastic process W defined on $\mathbb{R}^+$ is a Brownian motion if*

- *$W_0 = 0$,*
- *$W_t - W_s \sim \mathcal{N}(0,t-s)$,*
- *for all times $0 < t_1 < \dots < t_n$, the increments $W(t_1)$, $W(t_2)-W(t_1)$, …, $W(t_n)-W(t_{n-1})$ are independent.*

Exercise: find the density of $X = (W(t_1), W(t_2), \dots, W(t_n))$. For $n = 2$, we find

$$\rho_X(x_1, x_2) = \frac{1}{2\pi\sqrt{t_1(t_2 - t_1)}} \exp(-\frac{x_1^2}{2t_1}) \exp(-\frac{(x_2 - x_1)^2}{2(t_2 - t_1)}).$$

Proposition 3.3.2. *A Brownian motion satisfies the following properties:*

- $W_t \sim \mathcal{N}(0, t)$, W *is a martingale,*
- $\mathbb{E}[W_t] = 0$, $\mathbb{E}[W_t^2] = t$, $\mathbb{E}[W_t W_s] = \min(t, s)$,
- W *has a* α*-Hölderian modification for* $0 < \alpha < 1/2$. *It is in particular continuous.*
- *For almost every* $\omega \in \Omega$, $W(., \omega)$ *is nowhere differentiable.*

The derivative of W can be understood as a distribution, and is often called white noise.

Proof. 1 and 2 are straightforward. We admit 4. 3 uses the Kolmogorov criterion. Indeed, we have $\mathbb{E}[|W(t) - W(s)|^2] = |t - s|$. Similarly, one finds

$$\mathbb{E}[|W(t) - W(s)|^{2k}] = \frac{(2k)!}{2^k k!} |t - s|^k .$$

Using the Kolmogorov criterion with $\delta = 2k$ and $\varepsilon = k-1$ gives that there exists a modification Y of W that is α-Hölderian with $\alpha < \frac{k-1}{2k} = \frac{1}{2} - \frac{1}{2k}$, and this for every k. Hence the result. □

Definition 3.3.3. *A Brownian motion in* $\mathbb{R}^d$ *is a vector whose components are independent Brownian motions in* $\mathbb{R}$.

Remark 3.3.4 (Numerical simulation of BM)**.** *We use a discretization* $t_n = nh$ *of* $[0, T]$ *with* $Nh = T$. *Let* $\xi_1, \dots, \xi_N$ *be independent Gaussian vectors of law* $\mathcal{N}(0, I_d)$. *We define*

$$W_{n+1} = W_n + \sqrt{h}\xi_n, \quad W_0 = 0.$$

Then, W *is a numerical approximation of the Brownian motion. Note that the increments are indeed independent, that* $W_n \sim \mathcal{N}(0, t_n I_d)$, *and that* (W_n) *is a martingale.*

Proposition 3.3.5. *Let* W *be a Brownian motion, then*

- *(homogenization)* $W_{t+h} - W_h$ *is a BM,*
- *(symmetry)* $-W_t$ *is a BM,*
- *(scaling/autosimilarity)* $\lambda W_{t/\lambda^2}$ *is a BM for* $\lambda > 0$,
- *(inverting time)* $tW_{1/t}$ *is a BM.*

Let W *and* M *be 2 stochastic processes such that* $M(t) = W_t^2 - t$, *then* W *is a BM iff* M *is a martingale. (Theorem of Lévy)*

Proof. The first points are left as an exercise (you can also plot it on a computer to observe these properties numerically). Let us prove one implication of the last point. $M(t)$ is adapted and L^1 as $W_t^2 \sim \chi^2$. Then

$$\mathbb{E}[M(t) - M(s)|\mathcal{F}_s] = \mathbb{E}[W_t^2 - W_s^2|\mathcal{F}_s] + s - t$$

$$\begin{aligned}&= \mathbb{E}[(W_t - W_s)^2 + 2W_s(W_t - W_s)|\mathcal{F}_s] + s - t\\ &= \mathbb{E}[(W_t - W_s)^2] + 2W_s\mathbb{E}[(W_t - W_s)] + s - t\\ &= t - s + 0 + s - t = 0.\end{aligned}$$

Hence the result. □

We define the random variables $T_a = \inf\{t \geqslant 0, W_t = a\}$.

Proposition 3.3.6. *The T_a are stopping times, they are finite a.s. (recurrence of the BM), and for $a < 0 < b$, they satisfy*

$$\mathbb{P}(T_a < T_b) = \frac{b}{b-a}, \quad \mathbb{P}(T_b < T_a) = \frac{-a}{b-a}.$$

Proof. By recurrence of the BM, we have

$$\mathbb{P}(T_a < T_b) + \mathbb{P}(T_b < T_a) = 1.$$

Then, we apply the Doob theorem to the martingale $W_{t\wedge T_a \wedge T_b}$, and we obtain in the limit $t \to \infty$:

$$\mathbb{E}[W_{T_a \wedge T_b}] = a\mathbb{P}(T_a < T_b) + b\mathbb{P}(T_b < T_a) = \mathbb{E}[W_0] = 0.$$

We admit the recurrence of the BM (see law of Blumenthal). □

3.4 Construction of the Brownian motion

Construction with random walks/numerical approximation: The Brownian motion can be seen as a limit of random walks. We work on $[0, T]$, with a discretization $t_n = n\Delta t$. We define the sequence of iid random variables (Y_k) by $\mathbb{P}(Y_k = \pm\sqrt{\Delta t}) = 0.5$. Then, the expectation is 0 and the variance Δt. The sum of these random variables $X_n = Y_1 + \cdots + Y_n$ satisfies $\mathbb{E}[X_n] = 0$ and $\mathbb{E}[(X_n)^2] = n\Delta t = t_n$. We interpolate linearly the points X_n in a function $X(t)$, that satisfies

$$\mathbb{E}[X(t)] = 0, \quad \mathbb{E}[(X(t))^2] = t + o(\Delta t).$$

One can prove that, as $\Delta t \to 0$, $X(t)$ converges in law to a Brownian motion. Indeed, by the CLT,

$$\frac{X_n}{t_n} = \frac{Y_1/\sqrt{\Delta t} + \cdots + Y_n/\sqrt{\Delta t}}{\sqrt{n}} \Rightarrow \mathcal{N}(0, 1),$$

that one can extends into $X(t) \Rightarrow \mathcal{N}(0, t)$. We refer to [5] for more details on the simulation of Brownian motions. Note that one could replace the Y_k by any r.v. with expectation 0 and variance 1.

Wiener's construction: Another possibility for constructing the Brownian motion on $[0, 1]$ is to sum standard Gaussian random variables against an orthonormal basis. In $L^2([0, 1])$, we have

$$(\mathbb{1}_{[0,t]}, \mathbb{1}_{[0,s]}) = \min(s, t).$$

In $L^2(\Omega)$, we have

$$\mathbb{E}[W_t W_s] = \min(s, t).$$

The idea is that the map $\mathbb{1}_{[0,t]} \to W_t$ can be extended as an isometry, called the Wiener isometry I_W. We take (e_n) an orthonormal basis of $L^2([0,1])$, then $\xi_n = (I_W(e_n))$ is an orthonormal basis of $L^2(\Omega)$, that is, $\mathbb{E}[\xi_i \xi_j] = \delta_{ij}$. Choosing ξ_n to be standard independent Gaussian random variables is natural. Then, we have

$$W_t = \sum_{n=1}^{\infty} \xi_n (\mathbb{1}_{[0,t]}, e_n).$$

This series converges in L^2. For the following particular choice of orthonormal basis, the series converges uniformly:

$$W_t = \xi_0 t + \sum_{n=1}^{\infty} \xi_n \frac{\sqrt{2}\sin(\pi n t)}{\pi n}.$$

Note that the series of the differentials does not converge in general.

3.5 Regularity of trajectories

Definition 3.5.1. *Let $f\colon [a,b] \to \mathbb{R}$ and $\Pi = (t_1 = a, t_2, \dots, t_p = b)$ a partition of $[a,b]$. We write*

$$V_\Pi f = \sum_{t_i \in \Pi} |f(t_{i+1}) - f(t_i)|, \quad V_\Pi^2 f = \sum_{t_i \in \Pi} |f(t_{i+1}) - f(t_i)|^2.$$

The total variation of f on $[a,b]$ is $V(f) = \sup_\Pi V_\Pi f$, while the quadratic variation of f on $[a,b]$ is $V^2(f) = \lim_{|\Pi|\to 0} V_\Pi^2 f$, where $|\Pi| = \sup |t_{i+1} - t_i|$. If the total variation of f is finite (resp. quadratic variation), we say that f has bounded variation (resp. bounded quadratic variation).

Proposition 3.5.2. *If f is continuous and $V(f) < \infty$, then $V^2(f) = 0$.*

Proof. We have the estimate

$$V_\Pi^2 f \leqslant \max_{|u-v|\leqslant|\Pi|} |f(u) - f(v)|\, V(f).$$

Taking the limit $|\Pi| \to 0$ yields the result. □

Exercise: If f is $\mathcal{C}^1([a,b])$, then $V_\Pi f = \int_a^b |f'(x)|\, dx$.

Proposition 3.5.3. *If W is a BM, then its quadratic variation on $[0,T]$ is $V^2(W) = T$ almost surely (and $V(W) = \infty$ almost surely).*

Proof. We prove the preliminary results that $\mathbb{E}[V^2(W) - T] = 0$ and $\mathbb{E}[(V^2(W) - T)^2] = 0$. The rest of the proof uses the Borel-Cantelli lemma and the Markov inequality.

First, we have by a simple calculation

$$V_\Pi^2(W) - T = \sum_{i=1}^{p-1} (W(t_{i+1}) - W(t_i))^2 - (t_{i+1} - t_i).$$

As the random variables in the sum have mean zero, we get $\mathbb{E}[V^2(W) - T] = 0$.

For the second moment, we have

$$\mathbb{E}[(V^2(W) - T)^2] = \sum_{i=1}^{p-1} \mathbb{E}[(W(t_{i+1}) - W(t_i))^4] - 2\sum_{i=1}^{p-1} (t_{i+1} - t_i)\mathbb{E}[(W(t_{i+1}) - W(t_i))^2]$$

$$
\begin{aligned}
&+ \sum_{i=1}^{p-1} (t_{i+1} - t_i)^2 \\
&= \sum_{i=1}^{p-1} 3(t_{i+1} - t_i)^2 - 2(t_{i+1} - t_i)^2 + (t_{i+1} - t_i)^2 \\
&= 2 \sum_{i=1}^{p-1} (t_{i+1} - t_i)^2 = 0,
\end{aligned}
$$

where we used that the quadratic variation of the differentiable function $f(t) = t$ vanishes. □

Definition 3.5.4. *Let M be a martingale, then we write $\langle M, M \rangle_t$ the quadratic variation of M in probability on $[0, t]$. For M, N two martingales, we write (polarization)*

$$\langle M, N \rangle_t = \frac{1}{2}(\langle M + N, M + N \rangle_t - \langle M, M \rangle_t - \langle N, N \rangle_t).$$

Theorem 3.5.5. *We have*

$$\langle M, N \rangle_t = \lim_{|\Pi| \to 0}^{\mathbb{P}} \sum_i (M_{t_{i+1}} - M_{t_i})(N_{t_{i+1}} - N_{t_i}).$$

$\langle ., . \rangle_t$ is bilinear symmetric. Moreover $M_t^2 - \langle M, M \rangle_t$ is a martingale.

Indeed for the Brownian motion, we had that $W_t^2 - t$ is a martingale.

We now realise that martingales and BM have a lot in common. The following result links the two concepts.

Theorem 3.5.6. *If M_t is a martingale bounded in L^2 with $M_0 = 0$, then there exists a BM W_t such that $M_t = W(\langle M, M \rangle_t)$.*

3.6 Brownian motion and PDEs

Let $u(x, t) = \mathbb{E}[\phi(x + W(t))]$, we will see in the following that u satisfies the following PDE:

$$\frac{\partial u}{\partial t} = \frac{1}{2} \Delta u.$$

We give the Dirichlet problem, for D a bounded open set of $\mathbb{R}^d$, and $f : \partial D \to \mathbb{R}$ continuous, there exists a unique function u that is $\mathcal{C}^2(D)$, continuous on $\overline{D}$, and such that

$$\Delta u = 0, \qquad u = f \text{ on } \partial D.$$

The Walk On Sphere (WOS) algorithm is the following. First, for $x \in D$ and $r \in]0, 1]$, and (U_n) a sequence of iid $\mathcal{U}(S^d)$ random variables, we define

$$X_x(n + 1) = X_x(n) + r d(X_x(n), \partial D) U_n, \qquad X_x(1) = x.$$

One proves that the sequence $(X_x(n))$ converges a.s. to $X_x(\infty) \in \partial D$.

Theorem 3.6.1. *Let $u(x) = \mathbb{E}[f(X_x(\infty))]$, then u is the unique solution of the Dirichlet problem.*

There exists a handful of numerical methods to solve the Dirichlet problem such as finite differences or finite elements. The WOS algorithm has the advantage that it is easy to implement, and its convergence does not require some regularity assumption on D and its boundary. In high dimension, it outperforms the deterministic algorithms.

Project: see Villa-Morales, 2011.

Chapter 4

Construction of the stochastic integral

References for this chapter - [6, 2].

4.1 Construction of the Itô integral

4.2 The Ito formula

4.3 A word on the Stratonovich integral

Chapter 5

Stochastic differential equations

5.1 Definition, examples

Let $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t, W(t))$ be the probability space, equipped with a filtration such that W is adapted for this filtration. We consider the stochastic differential equation

$$dX(t) = f(t, X(t))dt + g(t, X(t))dW(t), X(0) = X_0,$$

with $f(t,x) \in \mathbb{R}^d$, $g(t,x) \in \mathbb{R}^{d\times k}$ and W a k-dimensional Brownian motion. We understand this equation as its integral formulation. If g is a constant, we say that it is a SDE with additive noise. If it is $g(x)$, we say multiplicative noise. The function f is called the drift, while we call gdW the martingale part/diffusion of the SDE.

We recall the Stratonovich conversion

$$dX = f(X)dt + g(X) \circ dW$$

mean

$$dX = f(X)dt + \frac{1}{2}g'(x)g(x) + g(X)dW.$$

We first mention a few important examples. One of the central models is the Black-Scholes model

$$dS(t) = S(t)(\mu dt + \sigma dW(t)).$$

$S(t)$ is the price of an asset, μ the drift and σ the volatility. The solution is given by

$$S(t) = S_0 \exp((\mu - \sigma^2/2)t + \sigma W(t)).$$

If $\sigma = 0$, we typically find the rate of a savings bank account (and μ is usually small). If σ is large, the market is driven by randomness. *See numerical experiments.*

The underdamped Langevin equation models the motion of particles in a fluid

$$\begin{aligned} dq &= pdt \\ dp &= -\gamma p dt - \nabla V(q)dt + \sigma dW. \end{aligned}$$

If one lets the friction γ go to infinity, we obtain the overdamped Langvin dynamics

$$dX = -\nabla V(X)dt + \sigma dW,$$

which we will study closely in the numerical analysis part of the course. If V is quadratic, we obtain an Ornstein–Uhlenbeck process

$$dX = -\lambda X dt + \sigma dW.$$

Exercise: compute the expectation and covariance $\mathrm{Cov}(X(s), X(t))$ of X.

If the SDE has the form

$$dX(t) = \Pi_{\mathcal{M}}(X(t))f(X(t))dt + \Pi_{\mathcal{M}}(X(t))g(X(t)) \circ dW(t), X(0) = X_0,$$

where $\Pi_{\mathcal{M}}(x) = I_d - \nabla\zeta \left|\nabla\zeta\right|^{-2} \nabla\zeta^T$ and ζ is a smooth function taking values in $\mathbb{R}$, then $X(t)$ lies on the manifold $\mathcal{M}$ given by the constraint $\zeta(X(t)) = \zeta(X_0)$ (Exercise: check that indeed $d\zeta(X(t)) = 0$). For instance, for the sphere, $\Pi_{\mathcal{S}^{d-1}}(x) = I_d - xx^T/(x^Tx)$. The Brownian motion on the sphere satisfies:

$$dX = (I_d - xx^T/(x^Tx)) \circ dW.$$

In 2D, it simplifies into

$$dX = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} X \circ dW = -\frac{1}{2} X dt + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} X dW.$$

One can check that $(\cos(W(t)), \sin(W(t)))$ is the solution.

We mention the stochastic oscillator $q'' = -\omega^2 q dt + dW$ that is written rigorously as

$$\begin{aligned} dq &= p dt \\ dp &= -\omega^2 q dt + dW. \end{aligned}$$

Its solution is the following if $p(0) = 0$:

$$q(t) = q_0 \cos(\omega t) + \frac{1}{\omega} \int_0^t \sin(\lambda(t-s)) dW(s).$$

See numerical experiments.

Exercise: compute the variance of $q(t)$. Find the exact formula if $p(0) \neq 0$.

In general, we do not have an explicit expression of the solution of a given SDE. We are usually interested in its behaviour, or in the average of a function of the solution $\mathbb{E}[\phi(X(t))]$. For Langevin dynamics, $\mathbb{E}[|p(t)|^2]$ is proportional to the temperature of the system, which is a key quantity in thermodynamics. (Exercise: what is the equation satisfied by $\mathbb{E}[S(t)]$ in the Black-Scholes model?)

5.2 Existence and uniqueness of solutions

A more general theorem is the following.

Theorem 5.2.1. *Let $f(t,x)$ and $g(t,x$ be measurable functions, that are Lipschitz in x uniformly in $t \in [0,T]$ and satisfy*

$$|f(t,x)| \leqslant C(1+|x|), \quad |g(t,x)| \leqslant C(1+|x|).$$

If $X_0 \in L^2(\Omega)$, then the SDE

$$dX(t) = f(t,X(t))dt + g(t,X(t))dW(t), X(0) = X_0,$$

has a unique solution in

Proof. Do the proof together with the class. See Kuo (go fast on the Borel-Cantelli part for simplicity). □

5.3 SDEs and PDEs

5.3.1 Generator and the Kolmogorov equation

Often in physical applications, one is interested in the average of a function of the solution of the SDE, that is, $u(x,t) = \mathbb{E}[\phi(X(t))|X(0) = x]$. This function u is a major link between SDEs and PDEs.

We introduce the space $\mathcal{C}^p_P(\mathbb{R}^d, \mathbb{R})$ of test functions that are in $\mathcal{C}^p$ such that ϕ and its first p derivatives have at most polynomial growth. For simplicity, we work with autonomous SDEs from now on. Our notation for derivatives is the following

$$\phi^{(k)}(a^1, \dots, a^k) = \sum_{i_1,\dots,i_k=1}^{d} \frac{\partial^k \phi}{\partial_{x_{i_1}} \dots \partial_{x_{i_k}}} a^1_{i_1} \dots a^k_{i_k}.$$

We also denote by (e_i) the canonical basis of $\mathbb{R}^q$ (depending on the context).

The function u satisfies the following PDE.

Theorem 5.3.1 (Talay)**.** *Let $f(x)$ and $g(x) \in \mathcal{C}^{2p+2}$ with bounded partial derivatives, let $\phi \in \mathcal{C}^{2p+2}_P$, then $u(x,t) \in \mathcal{C}^{2p+2}_P$ and it satisfies*

$$\frac{\partial u}{\partial t}(x,t) = \mathcal{L}u(x,t), \quad u(x,0) = \phi(x).$$

The linear operator $\mathcal{L}$ is called the generator of the SDE and is given by

$$\mathcal{L}\phi = \phi' f + \frac{1}{2}\sum_{i=1}^{k} \phi''(ge_i, ge_i).$$

This PDE is called the forward Kolmogorov equation. There exists an alternate way to write the last term of $\mathcal{L}$ in the literature, but the one we use here is more convenient for the numerical analysis part.

Example: The generator of $dX = dW$ is $\mathcal{L} = \frac{1}{2}\Delta$, and the Kolmogorov equation associated to simple Brownian dynamics is the heat equation (hence the name diffusion part). For the overdamped Langevin dynamics, it is $\mathcal{L} = -\nabla V \cdot \nabla + \frac{\sigma^2}{2}\Delta$. Note that for additive noise, the generator is always elliptic. For the constrained overdamped Langevin, we find

$$\mathcal{L}\phi = \phi'(\Pi_{\mathcal{M}} f) + \frac{\sigma^2}{2}\sum_{i=1}^{d} \phi''(\Pi_{\mathcal{M}} e_i, \Pi_{\mathcal{M}} e_i) + \frac{\sigma^2}{2}\sum_{i=1}^{d} \phi'(\Pi'_{\mathcal{M}}(\Pi_{\mathcal{M}} e_i)e_i).$$

This operator is hypoelliptic, but is elliptic on the manifold $\mathcal{M}$.

Proof. We apply the Itô formula to $\phi(X(t))$. It gives

$$d\phi(X) = \phi'(f)dt + \phi'(g)dW + \frac{1}{2}\sum_{i=1}^{k} \phi''(ge_i, ge_i)dt.$$

We write the integral formulation and take the expectation

$$\mathbb{E}[\phi(X(t))] = \phi(x) + \int_0^t \mathbb{E}[\phi'(X(s))(f(X(s)))]ds + \frac{1}{2}\sum_{i=1}^{k}\int_0^t \mathbb{E}[\phi''(X(s))(g(X(s))e_i, g(X(s))e_i)]ds.$$

We deduce

$$u(x,t) = \phi(x) + \int_0^t \mathbb{E}[\mathcal{L}\phi(X(s))|X(0) = x]ds.$$

We admit that $\mathbb{E}$ and $\mathcal{L}$ commute (for simplicity), and we get the result. □

A question now is: given a generator $\mathcal{L}$, does there exist an SDE associated to it? We refer to [6] for an answer. (give a few insights without details)

Application of the methodology: solving the Dirichlet problem with the WOS.

5.3.2 A word on SPDEs

Stochastic partial differential equations are beyond the scope of this course. We just give a few examples, that are important in numerics.

Stochastic Schrödinger equations with a stochastic perturbation

$$du = \Delta u dt + |u|^{2\sigma} u dt + g(u)dW,$$

or with a white noise dispersion (fiber optics)

$$idu = \Delta u \circ d\beta + |u|^{2\sigma} u dt.$$

The noise can also depend on the space variable x, and not only in t. In practice, all sorts of noises are useful, and it is important to understand what kind of noise is associated to the equation.

Stochastic heat equations with a stochastic perturbation

$$du = \Delta u dt + g(u)dW = 0,$$

or with a white noise dispersion

$$du = \Delta u \circ d\beta.$$

To simplify these equations, we consider spatial discretizations of them. For instance, for the Schrödinger equation, we get an equation of the form

$$dX = AXdt + F(X)dt + G(X)dW = 0,$$

where X is a vector that contains the coefficients associated to the different frequencies.

Chapter 6

Hints of numerical analysis of SDEs

6.1 Different modes of convergence

We want to solve the SDE $dX = f(X)dt + g(X)dW$ (recall dimensions). The question now is: what do we want to approximate? Approximating the exact trajectory is called a strong approximation. As discussed before, we can approximate the average of a function of the solution. This is a weak approximation. For certain systems, called ergodic, the system reaches an equilibrium. It then makes sense to approximate the average of a function of the solution at equilibrium. We keep in mind that for underdamped Langevin dynamics, this quantity of interest is typically the temperature of the system.

6.1.1 Strong approximation of SDEs

We present the Euler-Maruyama method

$$X_{n+1} = X_n + hf(X_n) + g(X_n)\Delta W_n.$$

We use the notation $\Delta W_n = W(t_{n+1}) - W(t_n)$ for simplicity. Note that $\Delta W_n \sim \mathcal{N}(0, hI_d)$.

We also introduce the Milstein method

$$X_{n+1} = X_n + hf(X_n) + g(X_n)\Delta W_n + \frac{1}{2}g'(X_n)g(X_n)(\Delta W_n^2 - h).$$

We find again this correction term that we saw in the conversion Itô-Stratonovich.

Definition 6.1.1. *A method (X_n) is said to have strong order $\gamma > 0$ if for T fixed and $h \leqslant h_0$ small enough such that $Nh = T$, there exists a constant independent of h and N such that*

$$\mathbb{E}[|X(t_n) - X_n|^2] \leqslant Ch^\gamma, \quad n = 0, \dots, N.$$

Proposition 6.1.2. *The Euler method has order $1/2$ in general and the Milstein method has order 1. If we have additive noise (g is constant), the two methods coincide and have order 1.*

Proof. Hard and technical. □

It is considered hard to build high order methods. One needs to simulate iterated stochastic integrals, and these are expensive and difficult to simulate.

6.1.2 Weak approximation of SDEs

A possible solution is to approximate $u(x,t) = \mathbb{E}[\phi(X(t))|X(0) = x]$ instead. Then we care only about the *law of the solution*. We rewrite the Euler-Maruyama scheme as

$$X_{n+1} = X_n + hf(X_n) + \sqrt{h}g(X_n)\xi_n.$$

The random variable (ξ_n) are independent identically distributed and follow *approximately* the same law as $\Delta W_n/\sqrt{h} \overset{\mathcal{L}}{=} \mathcal{N}(0, I_d)$. We will see that if we want weak order p, we ask that

$$\mathbb{E}[\xi_i^{2q}] = \frac{(2q)!}{2^q q!}, \quad q = 0, \dots, p.$$

An obvious choice would be $\xi_i \sim \mathcal{N}(0,1)$, but we often prefer to use bounded random variables for stability reasons. For instance, for order 2, one can take

$$\mathbb{P}(\xi_i = 0) = \frac{2}{3}, \quad \mathbb{P}(\xi_i = \pm\sqrt{3}) = \frac{1}{6}.$$

Definition 6.1.3. *A method (X_n) is said to have weak order $\gamma > 0$ if for T fixed and $h \leqslant h_0$ small enough such that $Nh = T$, for all test functions $\phi \in \mathcal{C}_P^\infty$, there exists a constant independent of h and N such that*

$$|\mathbb{E}[\phi(X(t_n))] - \mathbb{E}[\phi(X_n)]| \leqslant Ch^\gamma, \quad n = 0, \dots, N.$$

The intuition from this definition is that if you approximate correctly the moments of a bounded random variable, you approximate its law correctly (see characteristic function/Laplace transform).

In practice, we have to approximate the expectation with a statistical estimator. We apply the integrator (X_n) for different randomness a high number of times M. Then, our approximation of $u(X_0, T)$ is

$$\bar{U} = \frac{1}{M}\sum_{i=1}^{M} \phi(X_N^{(m)}).$$

The new error estimate that we get is of the form

$$\left|\bar{U} - u(X_0, T)\right| \leqslant Ch^\gamma + \frac{C}{\sqrt{M}},$$

where the error bound for the estimator is given by the CLT. The term in $\mathcal{O}(M^{-1/2})$ is hard to improve. There are variance-reduction techniques, for instance with antithetic coupling or MLMC methods, but this term is often the biggest in the expansions. This is why order two is already high order in the stochastic setting. The main term of the error is not the discretization term. We emphasize though that combining variance-reduction techniques and high-order discretizations is not straightforward.

Proposition 6.1.4. *The Euler-Maruyama method has weak order 1 in general.*

In particular, the weak order is an integer for the class of methods we will consider.

6.1.3 Ergodicity and approximation of the invariant measure

We say that the process $X(t)$ is ergodic if it has a unique invariant measure μ satisfying for each test function ϕ and for any deterministic initial condition $X_0 = x$,

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \phi(X(s))ds = \int_{\mathbb{R}^d} \phi(x)d\mu(x).$$

It means that, whatever the initial condition, the trajectory will visit in long time the entire space. It will spend more time at some places than in others, and this is described by the measure μ.

Thanks to the uniform ellipticity of the Langevin generator, the overdamped Langevin dynamics are ergodic (with growth assumptions on the potential). The constrained Langevin dynamics are also ergodic. Note that μ is absolutely continuous wrt the Lebesgue measure in the $\mathbb{R}^d$ case, and wrt to the measure on $\mathcal{M}$ induced by the Euclidean metric in the manifold case.

6.2 Weak convergence analysis

6.2.1 Euler-Maruyama method

Do the proof of global order 1 of [10]. Skip the details on the regularity of the modified test function.

6.2.2 High order weak integration of SDEs

Present from [7] the Talay-Tubaro expansion, the trees, the application on Runge-Kutta methods for solving Langevin dynamics. Show the order 2 calculations and talk about the different formalisms of stochastic trees.
Open questions:

- Runge-Kutta methods of any order for multiplicative noise ([1])
- Associated tree formalism
- Simpler formalism in the context of the invariant measure?

6.2.3 High order integration with constraints

Present from [8, 9] the equation, constraints, examples from molecular dynamics, the generator, the projection methods, the constrained Euler method, the tree system, algebraic properties of the tree system, order conditions in simple cases.
Open questions:

- Simpler formalism for writing the generator and the Talay-Tubaro expansion?
- Different class of methods than projection methods, Lie-group methods?
- Rewrite everything directly on a manifold, without the need of the embedding in $\mathbb{R}^d$.

Chapter 7

Hints of stochastic differential geometry

Reference for this chapter - *SDE and PDE: Solving PDE by running a Brownian Motion*, A. Thalmaier.

Bibliography

[1] K. Debrabant and A. Rößler. Families of efficient second order Runge-Kutta methods for the weak approximation of Itô stochastic differential equations. *Appl. Numer. Math.*, 59(3-4):582–594, 2009.

[2] L. C. Evans. *An introduction to stochastic differential equations*. American Mathematical Society, Providence, RI, 2013.

[3] G. B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[4] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006. Structure-preserving algorithms for ordinary differential equations.

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

[6] H.-H. Kuo. *Introduction to stochastic integration*. Universitext. Springer, New York, 2006.

[7] A. Laurent and G. Vilmart. Exotic aromatic B-series for the study of long time integrators for a class of ergodic SDEs. *Math. Comp.*, 89(321):169–202, 2020.

[8] A. Laurent and G. Vilmart. Order conditions for sampling the invariant measure of ergodic stochastic differential equations on manifolds. *To appear in Found. Comput. Math.*, 2021.

[9] T. Lelièvre, M. Rousset, and G. Stoltz. *Free energy computations*. Imperial College Press, London, 2010. A mathematical perspective.

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

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

[12] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 1994.