Introduction to the geometry of stochastic differential equations and diffusion semigroups
SNSB Bucharest lectures

Fabrice Baudoin
Department of Mathematics
Purdue University
fbaudoin@math.purdue.edu

Contents

Chapter 1

Brownian motion and Stochastic processes

The first stochastic process that has been extensively studied is the so-called Brownian motion, named in honor of the botanist Robert Brown (1773-1858), who observed and described in 1828 the random movement of particles suspended in a liquid or gas. One of the first mathematical studies of this process goes back to the mathematician Louis Bachelier (1870-1946), in 1900, who presented a stochastic modelling of the stock and option markets. But, mainly due to the lack of rigorous foundations of probability theory at that time, the seminal work of Bachelier was partly *heuristic* and has been ignored for a long time by mathematicians. Let us observe that independently of Bachelier, in his 1905 paper, Albert Einstein (1879-1955) brought this stochastic process to the attention of physicists by presenting it as a way to indirectly confirm the existence of atoms and molecules.

1.1 Introduction to the Brownian motion: The point of view of Bachelier

The purpose of Bachelier was to provide a mathematical model that could describe and predict the evolution of a stock price. Let X_t denote the price at time t of a given asset. By convention, we will assume that $X_0 = 0$ (otherwise, we work with $X_t - X_0$). The first observation is that the price X_t can not be predicted with absolute certainty. It seems therefore reasonable to assume that X_t is a random variable defined on some probability space. One of the initial problems of Bachelier was to understand the distribution of prices at given times, that is the law of the random variable $(X_{t_1}, ..., X_{t_n})$, where $t_1, ..., t_n$ are fixed. In what follows, we briefly discuss how he tackled this problem from scratch.

The two following fundamental observations of Bachelier were based on empirical observations:

- If τ is very small then, in absolute value, the price variation $X_{t+\tau} - X_t$ is of order $\sigma\sqrt{\tau}$, where σ is a positive parameter (nowadays called the volatility of the asset);
- The expectation of a speculator is always zero[1] (nowadays, a generalization of this principle is called the absence of arbitrage).

Next, Bachelier assumes that for every $t > 0$, X_t has a density with respect to the Lebesgue measure, let us say $p(t, x)$. The two above observations imply that for τ small,

$$p(t+\tau, x) = \frac{1}{2}p(t, x-\sigma\sqrt{\tau}) + \frac{1}{2}p(t, x+\sigma\sqrt{\tau}).$$

Indeed, due to the first observation, if the price is x at time $t+\tau$, it means that at time t the price was equal to $x-\sigma\sqrt{\tau}$ or to $x+\sigma\sqrt{\tau}$. According to the second observation, each of this case produces with probability $\frac{1}{2}$.
Now Bachelier assumes that $p(t, x)$ is regular enough and uses the following approximations coming from a Taylor expansion:

$$p(t+\tau, x) \simeq p(t, x) + \tau\frac{\partial p}{\partial t}(t, x)$$

$$p(t, x-\sigma\sqrt{\tau}) \simeq p(t, x) - \sigma\sqrt{\tau}\frac{\partial p}{\partial x}(t, x) + \frac{1}{2}\sigma^2\tau\frac{\partial^2 p}{\partial x^2}(t, x)$$

$$p(t, x+\sigma\sqrt{\tau}) \simeq p(t, x) + \sigma\sqrt{\tau}\frac{\partial p}{\partial x}(t, x) + \frac{1}{2}\sigma^2\tau\frac{\partial^2 p}{\partial x^2}(t, x).$$

This gives

$$\frac{\partial p}{\partial t} = \frac{1}{2}\sigma^2\frac{\partial^2 p}{\partial x^2}(t, x).$$

This is the so-called heat equation, which is the primary example of a diffusion equation. Explicit solutions to this equation are known, and by using the fact that at time 0, p is the Dirac distribution at 0, it is obtained that:

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

[1] Quoted and translated from Bachelier: *It seems that the market, the aggregate of speculators, can believe in neither a market rise nor a market fall, since, for each quoted price, there are as many buyers as sellers..*

It means that X_t has a Gaussian distribution with mean 0 and variance σ^2. Let now $0 < t_1 < ... < t_n$ be fixed times and $I_1, ..., I_n$ be fixed intervals. In order to compute

$$\mathbb{P}(X_{t_1} \in I_1, ..., X_{t_n} \in I_n)$$

the next step is to assume that the above analysis did not depend on the origin of time, or more precisely that the best information available at time t is given by the price X_t. That leads first to the following computation

$$\begin{aligned}\mathbb{P}(X_{t_1} \in I_1, X_{t_2} \in I_2) &= \int_{I_1} \mathbb{P}(X_{t_2} \in I_2 \mid X_{t_1} = x_1) p(t_1, x_1) dx_1 \\ &= \int_{I_1} \mathbb{P}(X_{t_2 - t_1} + x_1 \in I_2 \mid X_{t_1} = x_1) p(t_1, x_1) dx_1 \\ &= \int_{I_1 \times I_2} p(t_2 - t_1, x_2 - x_1) p(t_1, x_1) dx_1 dx_2,\end{aligned}$$

which is easily generalized to

$$\mathbb{P}(X_{t_1} \in I_1, ..., X_{t_n} \in I_n) = \int_{I_1 \times \cdots \times I_n} p(t_n - t_{n-1}, x_n - x_{n-1}) \cdots p(t_2 - t_1, x_2 - x_1) p(t_1, x_1) dx_1 dx_2 \cdots dx_n.$$

In many regards, the previous computations were not rigorous but heuristic. One of the main issues here, is that the sequence of random variables X_t is not well defined from a mathematical point of view. In the next section, we will provide a rigorous construction of this object $X_t, t \in \mathbb{R}$ on which worked Bachelier and that is called the Brownian motion.
Let us observe that after the work of Bachelier and the work by Black and Scholes and Merton, Brownian motion and its generalizations are at the basis of the modern mathematical modelling of financial markets.

1.2 Stochastic processes

1.2.1 Measure theory in function spaces

Stochastic processes can be seen as random variables taking their values in a function space. It is therefore important to understand the natural associated σ-algebras.
Let $\mathcal{A}(\mathbb{R}_+, \mathbb{R}^d)$, $d \geq 1$, be the set of functions $\mathbb{R}_{\geq 0} \to \mathbb{R}^d$. We denote by $\mathcal{T}(\mathbb{R}_{\geq 0}, \mathbb{R}^d)$ the σ-algebra generated by the so-called cylindrical sets

$$\{f \in \mathcal{A}(\mathbb{R}_{\geq 0}, \mathbb{R}^d), f(t_1) \in I_1, ..., f(t_n) \in I_n\}$$

where

$$t_1, ..., t_n \in \mathbb{R}_{\geq 0}$$

and where $I_1, ..., I_n$ are products of intervals $\Pi_{k=1}^d (a_i^k, b_i^k]$.

Remark 1.2.1 *As a σ-algebra $\mathcal{T}(\mathbb{R}_{\geq 0}, \mathbb{R}^d)$ is also generated by the following families:*

- $$\{f \in \mathcal{A}(\mathbb{R}_{\geq 0}, \mathbb{R}^d), f(t_1) \in B_1, ..., f(t_n) \in B_n\}$$
  where
  $$t_1, ..., t_n \in \mathbb{R}_{\geq 0}$$
  and where $B_1, ..., B_n$ are Borel sets in $\mathbb{R}^d$.

- $$\{f \in \mathcal{A}(\mathbb{R}_{\geq 0}, \mathbb{R}^d), (f(t_1), ..., f(t_n)) \in B\}$$
  where
  $$t_1, ..., t_n \in \mathbb{R}_{\geq 0}$$
  and where B is a Borel set in $(\mathbb{R}^d)^{\otimes n}$.

Exercise 1.2.2 *Show that the following sets are not in $\mathcal{T}([0,1], \mathbb{R})$:*

1. $$\{f \in \mathcal{A}([0,1], \mathbb{R}), \sup_{t \in [0,1]} f(t) < 1\}$$

2. $$\{f \in \mathcal{A}([0,1], \mathbb{R}), \exists t \in [0,1] f(t) = 0\}$$

The above exercise shows that the σ-algebra $\mathcal{T}(\mathbb{R}_{\geq 0}, \mathbb{R}^d)$ is not rich enough to include *natural* events; this is due to the fact that the space $\mathcal{A}(\mathbb{R}_{\geq 0}, \mathbb{R}^d)$ is by far too big. In this course, we shall be interested in processes with continuous paths. In that case, we use the space of continuous functions $\mathcal{C}(\mathbb{R}_{\geq 0}, \mathbb{R}^d)$ endowed with the σ-algebra $\mathcal{B}(\mathbb{R}_{\geq 0}, \mathbb{R}^d)$ generated by

$$\{f \in \mathcal{C}(\mathbb{R}_{\geq 0}, \mathbb{R}^d), f(t_1) \in I_1, ..., f(t_n) \in I_n\}$$

where

$$t_1, ..., t_n \in \mathbb{R}_{\geq 0}$$

and where $I_1, ..., I_n$ are products of intervals $\Pi_{k=1}^d (a_i^k, b_i^k]$.. This σ-algebra enjoys nice properties. It is for instance generated by the open sets of the topology of uniform convergence on compact sets.

Exercise 1.2.3 *Show that the following sets are in $\mathcal{B}([0,1], \mathbb{R})$:*

1. $$\{f \in \mathcal{C}([0,1], \mathbb{R}), \sup_{t \in [0,1]} f(t) < 1\}$$

2. $$\{f \in \mathcal{C}([0,1], \mathbb{R}), \exists t \in [0,1] f(t) = 0\}$$

1.2.2 Stochastic processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Definition 1.2.4 *On $(\Omega, \mathcal{F}, \mathbb{P})$, a (d-dimensional) stochastic process is a sequence $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ of $\mathbb{R}^d$-valued random variables that are measurable with respect to $\mathcal{F}$.*

A process $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ can also be seen as an application

$$X(\omega) \in \mathcal{A}(\mathbb{R}_{\geq 0}, \mathbb{R}^d), t \to X_t(\omega).$$

These applications are called the paths of the process. The application $X : (\Omega, \mathcal{F}) \to (\mathcal{A}(\mathbb{R}_{\geq 0}, \mathbb{R}^d), \mathcal{T}(\mathbb{R}_{\geq 0}, \mathbb{R}^d))$ is measurable. The probability measure defined by

$$\mu(A) = \mathbb{P}(X^{-1}(A)), A \in \mathcal{T}(\mathbb{R}_+, \mathbb{R}^d)$$

is called the law (or distribution) of $(X_t)_{t\in\mathbb{R}_+}$.

For $t \geq 0$, we denote by π_t the application that sends $f \in \mathcal{A}(\mathbb{R}_{\geq 0}, \mathbb{R}^d)$ to $f(t)$. The stochastic process $(\pi_t)_{t\in\ \mathbb{R}_{\geq 0}}$ which is defined on the probability space $(\mathcal{A}(\mathbb{R}_{\geq 0}, \mathbb{R}), \mathcal{T}(\mathbb{R}_{\geq 0}, \mathbb{R}^d), \mu)$ is called the canonical process associated to X. It is a process with law μ.

Definition 1.2.5 *A process $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ is said to be measurable if the application*

$$(t, \omega) \to X_t(\omega)$$

is measurable with respect to the σ-algebra $\mathcal{B}(\mathbb{R}_{\geq 0}) \otimes \mathcal{F}$, that is if

$$\forall A \in \mathcal{B}(\mathbb{R}^d), \{(t, \omega), X_t(\omega) \in A\} \in \mathcal{B}(\mathbb{R}_{\geq 0}) \otimes \mathcal{F}.$$

The paths of a measurable process are, of course, measurable functions $\mathbb{R}_{\geq 0} \to \mathbb{R}^d$.

Definition 1.2.6 *If a process X takes its values in $\mathcal{C}(\mathbb{R}_{\geq 0}, \mathbb{R}^d)$, that is if the paths of X are continuous functions, then we say that $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ is a continuous process.*

If $(X_t)_{t\in\mathbb{R}_+}$ is a continuous process then the application $X : (\Omega, \mathcal{F}) \to (\mathcal{C}(\mathbb{R}_{\geq 0}, \mathbb{R}^d), \mathcal{B}(\mathbb{R}_{\geq 0}, \mathbb{R}))$ is measurable and the law of X is a probability measure on $(\mathcal{C}(\mathbb{R}_{\geq 0}, \mathbb{R}^d), \mathcal{B}(\mathbb{R}_{\geq 0}, \mathbb{R}^d))$. Moreover, it can be shown that a continuous process is measurable in the sense of the definition 1.2.5.

1.3 The Daniell-Kolmogorov extension theorem

The Daniell-Kolmogorov extension theorem is one of the first deep theorems of the theory of stochastic processes. It provides existence results for nice probability measures on path (function) spaces. It is however non constructive and relies on the axiom of choice. In what follows, in order to avoid heavy notations we restrict to the one dimensional case $d = 1$. The multidimensional extension is straightforward.

Definition 1.3.1 *Let X be a stochastic process. For $t_1, ..., t_n \in \mathbb{R}_{\geq 0}$ we denote by $\mu_{t_1,...,t_n}$ the probability distribution of the random variable*

$$(X_{t_1}, ..., X_{t_n}).$$

It is therefore a probability measure on $\mathbb{R}^n$. This probability measure is called a finite dimensional distribution of the process X.

If two processes have the same finite dimensional distributions, then it is clear that the two processes induce the same distribution on the path space $\mathcal{A}(\mathbb{R}_{\geq 0}, \mathbb{R})$ because cylinders generate the σ-algebra $\mathcal{T}(\mathbb{R}_{\geq 0}, \mathbb{R})$.
The finite dimensional distributions of a given process satisfy the two following properties: If $t_1, ..., t_n \in \mathbb{R}_{\geq 0}$ and if τ is a permutation of the set $\{1, ..., n\}$, then:

1.

$$\mu_{t_1,...,t_n}(A_1 \times ... \times A_n) = \mu_{t_{\tau(1)},...,t_{\tau(n)}}(A_{\tau(1)} \times ... \times A_{\tau(n)}), \quad A_i \in \mathcal{B}(\mathbb{R}).$$

2.

$$\mu_{t_1,...,t_n}(A_1 \times ... \times A_{n-1} \times \mathbb{R}) = \mu_{t_1,...,t_{n-1}}(A_1 \times ... \times A_{n-1}), \quad A_i \in \mathcal{B}(\mathbb{R}).$$

Conversely,

Theorem 1.3.2 *(Daniell 1918, Kolmogorov 1933)*
Assume given for every $t_1, ..., t_n \in \mathbb{R}_{\geq 0}$ a probability measure $\mu_{t_1,...,t_n}$ on $\mathbb{R}^n$. Let us assume that these probability measures satisfy:

1.

$$\mu_{t_1,...,t_n}(A_1 \times ... \times A_n) = \mu_{t_{\tau(1)},...,t_{\tau(n)}}(A_{\tau(1)} \times ... \times A_{\tau(n)}), \quad A_i \in \mathcal{B}(\mathbb{R}).$$

2.

$$\mu_{t_1,...,t_n}(A_1 \times ... \times A_{n-1} \times \mathbb{R}) = \mu_{t_1,...,t_{n-1}}(A_1 \times ... \times A_{n-1}), \quad A_i \in \mathcal{B}(\mathbb{R}).$$

Then, there is a unique probability measure μ on $(\mathcal{A}(\mathbb{R}_+, \mathbb{R}), \mathcal{T}(\mathbb{R}_+, \mathbb{R}))$ such that for $t_1, ..., t_n \in \mathbb{R}_{\geq 0}$, $A_1, ..., A_n \in \mathcal{B}(\mathbb{R})$:

$$\mu(\pi_{t_1} \in A_1, ..., \pi_{t_n} \in A_n) = \mu_{t_1,...,t_n}(A_1 \times ... \times A_n).$$

The Daniell-Kolmogorov theorem is often used to construct processes thanks to the following corollary:

Corollary 1.3.3 *Assume given for every $t_1, ..., t_n \in \mathbb{R}_{\geq 0}$ a probability measure $\mu_{t_1,...,t_n}$ on $\mathbb{R}^n$. Let us further assume that these measures satisfy the assumptions of the Daniell-Kolmogorov theorem. Then, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as well as a process $(X_t)_{t\geq 0}$ defined on this space such that the finite dimensional distributions of $(X_t)_{t\geq 0}$ are given by the $\mu_{t_1,...,t_n}$'s.*

Proof. As a probability space we chose

$$(\Omega, \mathcal{F}, \mathbb{P}) = (\mathcal{A}(\mathbb{R}_{\geq 0}, \mathbb{R}), \mathcal{T}(\mathbb{R}_{\geq 0}, \mathbb{R}), \mu)$$

where μ is the probability measure given by the Daniell-Kolmogorov theorem. The canonical process $(\pi_t)_{t\geq 0}$ defined on $\mathcal{A}(\mathbb{R}_{\geq 0}, \mathbb{R})$ by $\pi_t(f) = f(t)$ does the job. □

We now turn to the proof of the Daniell-Kolmogorov theorem. This proof proceeds in several steps, is difficult and can be skipped by the student in a first reading.

As a first step, let us recall the Caratheodory extension theorem that is often useful for the effective construction of measures (cf. the construction of the Lebesgue measure on $\mathbb{R}$):

Theorem 1.3.4 *(Caratheodory) Let Ω be a non empty set and let $\mathcal{A}$ be a family of subsets that satisfy:*

1. *$\Omega \in \mathcal{A}$;*
2. *Si $A, B \in \mathcal{A}$, $A \cup B \in \mathcal{A}$;*
3. *Si $A \in \mathcal{A}$, $\Omega \backslash A \in \mathcal{A}$.*

Let $\sigma(\mathcal{A})$ be the σ-algebra generated by $\mathcal{A}$. If μ_0 is σ-additive measure on $(\Omega, \mathcal{A})$, then there exists a unique measure[2] μ on $(\Omega, \sigma(\mathcal{A}))$ such that for $A \in \mathcal{A}$,

$$\mu_0(A) = \mu(A).$$

[2] The σ-additivity is not provided by the theorem

As a first step, we prove the following fact:

Lemma 1.3.5 *Let $B_n \subset \mathbb{R}^n$, $n \in \mathbb{N}$ be a sequence of Borel sets that satisfy*

$$B_{n+1} \subset B_n \times \mathbb{R}.$$

Let us assume that for every $n \in \mathbb{N}$ a probability measure μ_n is given on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ and that these probability measures satisfy:

$$\mu_n(B_n) > \varepsilon,$$

where $0 < \varepsilon < 1$. There exists a sequence of compact sets $K_n \subset \mathbb{R}^n$, $n \in \mathbb{N}$, such that:

- $K_n \subset B_n$
- $K_{n+1} \subset K_n \times \mathbb{R}$.
- $\mu_n(K_n) \geq \frac{\varepsilon}{2}$.

Proof.
For every n, we can find a compact set $K_n^* \subset \mathbb{R}^n$ such that

$$K_n^* \subset B_n$$

and

$$\mu_n(B_n \backslash K_n^*) \leq \frac{\varepsilon}{2^{n+1}}.$$

Let

$$K_n = (K_1^* \times \mathbb{R}^{n-1}) \cap ... \cap (K_{n-1}^* \times \mathbb{R}) \cap K_n^*.$$

It is easily checked that:

- $K_n \subset B_n$
- $K_{n+1} \subset K_n \times \mathbb{R}$.

Moreover,

$$\begin{aligned}
\mu_n(K_n) &= \mu_n(B_n) - \mu_n(B_n \backslash K_n) \\
&= \mu_n(B_n) - \mu_n\left(B_n \backslash \left((K_1^* \times \mathbb{R}^{n-1}) \cap ... \cap (K_{n-1}^* \times \mathbb{R}) \cap K_n^*\right)\right) \\
&\geq \mu_n(B_n) - \mu_n\left(B_n \backslash \left((K_1^* \times \mathbb{R}^{n-1})\right)\right) - ... - \mu_n\left(B_n \backslash \left(K_{n-1}^* \times \mathbb{R}\right)\right) - \mu_n(B_n \backslash K_n^*) \\
&\geq \mu_n(B_n) - \mu_n(B_1 \backslash K_1^*) - ... - \mu_n(B_n \backslash K_n^*) \\
&\geq \varepsilon - \frac{\varepsilon}{4} - ... - \frac{\varepsilon}{2^{n+1}} \\
&\geq \frac{\varepsilon}{2}.
\end{aligned}$$

□

With this in hands, we can now turn to the proof of the Daniell-Kolmogorov theorem.

Proof. For the cylinder

$$\mathcal{C}_{t_1,...,t_n}(B) = \{f \in \mathcal{A}(\mathbb{R}_+, \mathbb{R}), (f(t_1), ..., f(t_n)) \in B\}$$

where

$$t_1, ..., t_n \in \mathbb{R}_{\geq 0}$$

and where B is a Borel subset of $\mathbb{R}^n$, we define

$$\mu\left(\mathcal{C}_{t_1,...,t_n}(B)\right) = \mu_{t_1,...,t_n}(B).$$

Thanks to the assumptions on the $\mu_{t_1,...,t_n}$'s, it is seen that such a μ is well defined and satisfies:

$$\mu\left(\mathcal{A}(\mathbb{R}_{\geq 0}, \mathbb{R})\right) = 1.$$

The set $\mathcal{A}$ of all the possible cylinders $\mathcal{C}_{t_1,...,t_n}(B)$ satisfies the assumption of Caratheodory theorem. Therefore, in order to conclude, we have to show that μ is σ-additive, that is, if $(C_n)_{n\in\mathbb{N}}$ is a sequence of pairwise disjoint cylinders and if $C = \cup_{n\in\mathbb{N}} C_n$ is a cylinder then

$$\mu(C) = \sum_{n=0}^{+\infty} \mu(C_n).$$

This is the difficult part of the theorem.
Since for $N \in \mathbb{N}$,

$$\mu(C) = \mu\left(C \backslash \cup_{n=0}^{N} C_n\right) + \mu\left(\cup_{n=0}^{N} C_n\right),$$

we just have to show that

$$\lim_{N\to+\infty} \mu(D_N) = 0.$$

where $D_N = C \backslash \cup_{n=0}^{N} C_n$.
The sequence $(\mu(D_N))_{N\in\mathbb{N}}$ is positive decreasing and therefore converges. Let assume that it converges toward $\varepsilon > 0$. We shall prove that in that case

$$\cap_{N\in\mathbb{N}} D_N \neq \emptyset,$$

which is clearly absurd.
Since D_N is a cylinder, the event $\cup_{N\in\mathbb{N}} D_N$ only involves a coutable sequence of times $t_1 < ... < t_n < ...$ and we may assume (otherwise we can add convenient other sets in the sequence of the D_N's) that every D_N can be described as follows

$$D_N = \{f \in \mathcal{A}(\mathbb{R}_{\geq 0}, \mathbb{R}), (f(t_1), ..., f(t_N)) \in B_N\}$$

where $B_n \subset \mathbb{R}^n$, $n \in \mathbb{N}$, is a sequence of Borel sets such that

$$B_{n+1} \subset B_n \times \mathbb{R}.$$

Since we assumed $\mu(D_N) \geq \varepsilon$, we can use the previous lemma to construct a sequence of compact sets $K_n \subset \mathbb{R}^n$, $n \in \mathbb{N}$, such that:

- $K_n \subset B_n$
- $K_{n+1} \subset K_n \times \mathbb{R}$.
- $\mu_{t_1,...,t_n}(K_n) \geq \frac{\varepsilon}{2}$.

Since K_n is non-empty, we pick

$$(x_1^n, ..., x_n^n) \in K_n.$$

The sequence $(x_1^n)_{n\in\mathbb{N}}$ has a convergent subsequence $(x_1^{j_1(n)})_{n\in\mathbb{N}}$ that converges toward $x_1 \in K_1$. The sequence $((x_1^{j_1(n)}, x_2^{j_1(n)})_{n\in\mathbb{N}}$ has a convergent subsequence that converges toward $(x_1, x_2) \in K_2$. By pursuing this process we obtain a sequence $(x_n)_{n\in\mathbb{N}}$ such that for every n,

$$(x_1, ..., x_n) \in K_n.$$

The event

$$\{f \in \mathcal{A}(\mathbb{R}_+, \mathbb{R}), (f(t_1), ..., f(t_N)) = (x_1, ..., x_N)\}$$

is in D_N, this leads to the expected contradiction. Therefore, the sequence $(\mu(D_N))_{N\in\mathbb{N}}$ converges toward 0, which implies the σ-additivity of μ. $\square$

1.4 The Kolmogorov continuity theorem

The Daniell-Kolmogorov theorem is a very useful tool since it provides existence results for stochastic processes. Nevertheless, this theorem does not say anything about the paths of this process. The following theorem, due to Kolmogorov, precises that, under mild conditions, we can work with processes whose paths are quite regular.

Definition 1.4.1 *A function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}^d$ is said to be Hölder with exponent $\alpha > 0$ if there exists a constant $C > 0$ such that for $s, t \in \mathbb{R}_{\geq 0}$,*

$$\|f(t) - f(s)\| \leq C \mid t - s \mid^\alpha .$$

Hölder functions are in particular continuous.

Definition 1.4.2 *A stochastic process* $(\tilde{X}_t)_{t\geq 0}$ *is called a modification of the process* $(X_t)_{t\geq 0}$ *if for* $t \geq 0$,

$$\mathbb{P}\left(X_t = \tilde{X}_t\right) = 1.$$

Remark 1.4.3 *We can observe that if* $(\tilde{X}_t)_{t\geq 0}$ *is a modification of* $(X_t)_{t\geq 0}$ *then* $(\tilde{X}_t)_{t\geq 0}$ *has the same distribution as* $(X_t)_{t\geq 0}$.

Theorem 1.4.4 *(Kolmogorov 1956) Let* $\alpha, \varepsilon, c > 0$. *If a d-dimensional process* $(X_t)_{t\in[0,1]}$ *defined* $(\Omega, \mathcal{F}, \mathbb{P})$ *satisfies for* $s, t \in [0,1]$,

$$\mathbb{E}\left(\|X_t - X_s\|^\alpha\right) \leq c \mid t-s \mid^{1+\varepsilon},$$

then there exists a modification of the process $(X_t)_{t\in[0,1]}$ *that is a continuous process and whose paths are* γ *Hölder for every* $\gamma \in [0, \frac{\varepsilon}{\alpha})$.

Proof.
For $n \in \mathbb{N}$, we denote

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

and

$$\mathcal{D} = \cup_{n\in\mathbb{N}} \mathcal{D}_n.$$

Let $\gamma \in [0, \frac{\varepsilon}{\alpha})$.
From Chebychev inequality:

$$\begin{aligned}
\mathbb{P}\left(\max_{1\leq k\leq 2^n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-\gamma n}\right) &= \mathbb{P}\left(\cup_{1\leq k\leq 2^n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-\gamma n}\right) \\
&\leq \sum_{k=1}^{2^n} \mathbb{P}\left(\|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-\gamma n}\right) \\
&\leq \sum_{k=1}^{2^n} \frac{\mathbb{E}\left(\|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\|^\alpha\right)}{2^{-\gamma\alpha n}} \\
&\leq c 2^{-n(\varepsilon-\gamma\alpha)}
\end{aligned}$$

Therefore, since $\gamma\alpha > \varepsilon$,

$$\sum_{n=1}^{+\infty} \mathbb{P}\left(\max_{1\leq k\leq 2^n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-\gamma n}\right) < +\infty,$$

From the Borel-Cantelli lemma, we can thus find a set $\Omega^* \in \mathcal{F}$ such that $\mathbb{P}(\Omega^*) = 1$ and such that for $\omega \in \Omega^*$, there exists $N(\omega)$ such that for $n \geq N(\omega)$,

$$\max_{1 \leq k \leq 2^n} \|X_{\frac{k}{2^n}}(\omega) - X_{\frac{k-1}{2^n}}(\omega)\| < 2^{-\gamma n}.$$

The paths of $X_{/\Omega^*}$ are consequently γ-Hölder on $\mathcal{D}$. Indeed, let $\omega \in \Omega^*$ and $s, t \in \mathcal{D}$ such that

$$| s - t | \leq \frac{1}{2^n}$$

with $n \geq N(\omega)$.
We can find an increasing and stationary sequence $(s_n)_{n \in \mathbb{N}}$ converging toward s, such that $s_n \in \mathcal{D}_n$ and

$$| s_{n+1} - s_n | = 2^{-(n+1)} \quad \text{ou} \quad 0.$$

In the same way, we can find an analogue sequence $(t_n)_{n \in \mathbb{N}}$ that converges toward t. We have then On a alors:

$$X_t - X_s = \sum_{i=n}^{+\infty} (X_{s_{i+1}} - X_{s_i}) + (X_{s_n} - X_{t_n}) + \sum_{i=n}^{+\infty} (X_{t_i} - X_{t_{i+1}}),$$

where the above sums are actually finite.
Therefore,

$$\begin{aligned} \|X_t - X_s\| &\leq 2 \sum_{i=n}^{+\infty} \max_{1 \leq k \leq 2^i} \|X_{\frac{k}{2^i}}(\omega) - X_{\frac{k-1}{2^i}}(\omega)\| \\ &\leq 2 \sum_{i=n}^{+\infty} 2^{-\gamma i} \\ &\leq \frac{2}{1 - 2^{-\gamma}} 2^{-\gamma n}. \end{aligned}$$

Hence the paths of $X_{/\Omega^*}$ are γ-Hölder on the set $\mathcal{D}$.
For $\omega \in \Omega^*$, let $t \to \tilde{X}_t(\omega)$ be the unique continuous function that agrees with $t \to X_t(\omega)$ on $\mathcal{D}$. For $\omega \notin \Omega^*$, we set $\tilde{X}_t(\omega) = 0$. The process $(\tilde{X}_t)_{t \in [0,1]}$ is the desired modification of $(X_t)_{t \in [0,1]}$. □

1.5 Gaussian processes and Brownian motion

1.5.1 Gaussian processes

Definition 1.5.1 *A real-valued stochastic process $(X_t)_{t \geq 0}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a Gaussian process if all the finite dimensional distributions of X are Gaussian random variables.*

The distribution of a Gaussian process is characterized by its mean function

$$m(t) = \mathbb{E}(X_t)$$

and its covariance function

$$R(s,t) = \mathbb{E}\left((X_t - m(t))(X_s - m(s))\right).$$

We can observe that the covariance function $R(s,t)$ is symmetric ($R(s,t) = R(t,s)$) and positive, that is for $a_1, ..., a_n \in \mathbb{R}$ and $t_1, ..., t_n \in \mathbb{R}_{\geq 0}$,

$$\begin{aligned}\sum_{1\leq i,j\leq n} a_i a_j R(t_i,t_j) &= \sum_{1\leq i,j\leq n} a_i a_j \mathbb{E}\left((X_{t_i} - m(t_i))(X_{t_j} - m(t_j))\right) \\ &= \mathbb{E}\left(\left(\sum_{i=1}^n (X_{t_i} - m(t_i))\right)^2\right) \\ &\geq 0.\end{aligned}$$

Conversely, as an easy application of the Daniell-Kolmogorov theorem,

Proposition 1.5.2 *Let $m : \mathbb{R}_{\geq 0} \to \mathbb{R}$ and let $R : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ be a symmetric and positive function. There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Gaussian process $(X_t)_{t\geq 0}$ defined on it, whose mean function is m and whose covariance function is R.*

1.5.2 Brownian motion

Definition 1.5.3 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A continuous real-valued process $(B_t)_{t\geq 0}$ is called a standard Brownian motion if it is a Gaussian process with mean function*

$$m(t) = 0$$

and covariance function

$$R(s,t) = \min(s,t).$$

Remark 1.5.4 *It is seen that $R(s,t) = \min(s,t)$ is a covariance function, because it is symmetric and for $a_1, ..., a_n \in \mathbb{R}$ and $t_1, ..., t_n \in \mathbb{R}_{\geq 0}$,*

$$\begin{aligned}\sum_{1\leq i,j\leq n} a_i a_j \min(t_i,t_j) &= \sum_{1\leq i,j\leq n} a_i a_j \int_0^{+\infty} \mathbf{1}_{[0,t_i]}(s)\mathbf{1}_{[0,t_j]}(s)ds \\ &= \int_0^{+\infty}\left(\sum_{i=1}^n a_i \mathbf{1}_{[0,t_i]}(s)\right)^2 ds \geq 0.\end{aligned}$$

Remark 1.5.5 *The distribution of a standard Brownian motion is called the Wiener measure.*

Remark 1.5.6 *A d-dimensional stochastic process $(B_t)_{t\geq 0}$ is called a standard Brownian motion if*

$$(B_t)_{t\geq 0} = (B_t^1, \cdots, B_t^d)_{t\geq 0}$$

where the processes $(B_t^i)_{t\geq 0}$ are independent standard Brownian motions.

Exercise 1.5.7 *Let $(B_t)_{t\geq 0}$ be a standard one-dimensional Brownian motion. Show the following properties:*

1. *$B_0 = 0$ a.s.;*
2. *For any $t > s \geq 0$, $B_t - B_s$ has the distribution as B_{t-s};*
3. *For any $t > s \geq 0$, the random variable $B_t - B_s$ is independent of the random variable $(B_{t_1}, ..., B_{t_n})$ whenever $0 < t_1, ..., t_n < s$.*
4. *If $0 < t_1 < \cdots < t_n$ are given times and if $I_1, \cdots, I_n$ are intervals then*

   $$\mathbb{P}(B_{t_1} \in I_1, ..., B_{t_n} \in I_n) = \int_{I_1\times\cdots\times I_n} p(t_n - t_{n-1}, x_n - x_{n-1}) \cdots p(t_2 - t_1, x_2 - x_1) p(t_1, x_1) dx_1 dx_2 \cdots dx_n,$$

   where $p(t,x) = \frac{e^{-\frac{x^2}{2t}}}{\sigma\sqrt{2\pi t}}$.
5. *For every $c > 0$, the process $(B_{ct})_{t\geq 0}$ has the same law as the process $(\sqrt{c}B_t)_{t\geq 0}$;*
6. *The process $(tB_{\frac{1}{t}})_{t\geq 0}$ has the same law as the process $(B_t)_{t\geq 0}$;*

Theorem 1.5.8 *There exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process on it that is a standard Brownian motion.*

Proof. From the Proposition 1.5.2, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Gaussian process $(X_t)_{t\geq 0}$ on it, whose mean function is 0 and covariance function is

$$\mathbb{E}(X_s X_t) = \min(s,t).$$

We have for $n \geq 0$ and $0 \leq s \leq t$:

$$\mathbb{E}\left((X_t - X_s)^{2n}\right) = \frac{(2n)!}{2^n n!}(t-s)^n.$$

Therefore, by using the Kolmogorov continuity theorem, there exists a modification $(B_t)_{t\geq 0}$ of $(X_t)_{t\geq 0}$ whose paths are locally γ-Hölder if $\gamma \in [0, \frac{n-1}{2n})$. $\square$

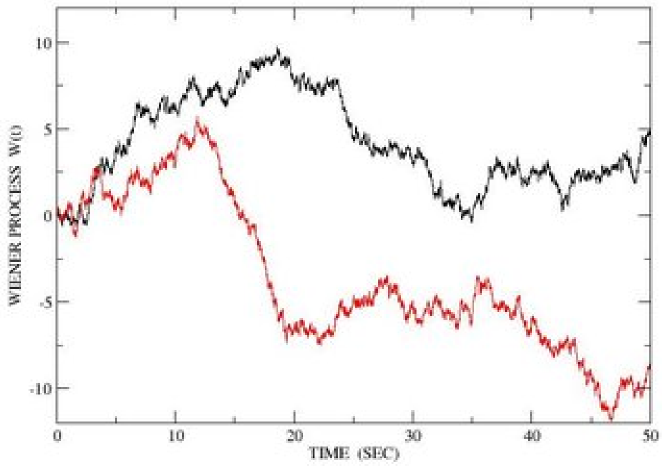

Figure 1.1: Sample paths of a Brownian motion

From the previous proof, we can also deduce that the paths of a standard Brownian motion are locally γ-Hölder for every $\gamma < \frac{1}{2}$. It can moreover be shown that they are not $\frac{1}{2}$-Hölder and almost nowhere differentiable.

For further reading on the Brownian motion we refer the interested student to [17].

Chapter 2

Diffusion processes

2.1 Markov process associated to a diffusion operator

2.1.1 Diffusion operators

The primary example of a diffusion equation is the heat equation. The problem is to find a real valued function $\phi(t,x)$ that is smooth on $(0,+\infty)\times\mathbb{R}^n$, continuous on $[0,+\infty)\times\mathbb{R}^n$ and that satisfies:

$$\frac{\partial\phi}{\partial t}(t,x)=\frac{1}{2}\Delta\phi(t,x),\quad \phi(0,x)=f(x), \tag{2.1.1}$$

where $\Delta=\sum_{i=1}^{d}\frac{\partial^2}{\partial x_i^2}$ is the Laplace operator on $\mathbb{R}^n$ and f is a given function on $\mathbb{R}^n$.

Remark 2.1.1 *This equation describes how temperature evolves. Assume that at time $t=0$ the temperature at a point $x\in\mathbb{R}^n$ is $f(x)$, then the temperature at time t will be given by $\phi(t,x)$.*

Remark 2.1.2 *The coefficient $\frac{1}{2}$ in front of the Laplace operator is here for a normalization reason that will later be apparent (it can be removed by replacing t by $2t$ in the following computations).*

This equation is easily solved by using Fourier transforms.

Exercise 2.1.3 *Assume that the function $f\in\mathbf{L}^2(\mathbb{R}^n,\mathbb{R})$. Show that the unique solution to the equation (2.1.1) is given by*

$$\phi(t,x)=\frac{1}{(2\pi t)^{\frac{n}{2}}}\int_{\mathbb{R}^n}e^{-\frac{\|x-y\|^2}{2t}}f(y)dy.$$

Beside the fact that it is possible to obtain a closed solution, it is remarkable that positivity is preserved by the equation. Namely, if f is non negative then $\phi(t,x)$ is also non negative. This is a fundamental feature shared by the so-called diffusion equations.

Example 2.1.4 *For instance, the equation*

$$\frac{\partial \phi}{\partial t}(t,x) = -\frac{1}{2}\frac{\partial^2 \phi}{\partial x^2}(t,x), \quad \phi(0,x) = f(x),$$

does not preserve positivity. Indeed, if $f(x) = x^2$*, then* $\phi(t,x) = x^2 - t$ *is a solution which is not non negative.*

Definition 2.1.5 *A differential operator* L *on* $\mathbb{R}^n$*, is called a diffusion operator if it can be written*

$$L = \sum_{i,j=1}^{n} \sigma_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial x_i},$$

where b_i *and* σ_{ij} *are continuous functions on* $\mathbb{R}^n$ *and for every* $x \in \mathbb{R}^n$*, the matrix* $(\sigma_{ij}(x))_{1 \le i,j \le n}$ *is a symmetric and positive matrix.*

Diffusion operators satisfy a maximum principle. Before we state this principle let us recall a simple result from linear algebra.

Lemma 2.1.6 *Let* A *and* B *be two symmetric positive matrices, then*

$$\mathbf{tr}(AB) \ge 0.$$

Proof. Since A is symmetric positive, there exists a symmetric and positive matrix S such that $S^2 = A$. We have then

$$\mathbf{tr}(AB) = \mathbf{tr}(S^2 B) = \mathbf{tr}(SBS) = \mathbf{tr}({}^t SBS).$$

The matrix ${}^t SBS$ is seen to be symmetric and positive and thus has a positive trace. □

Proposition 2.1.7 *(Maximum principle for diffusion operators) Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a smooth function that attains a local minimum at* x*. If* L *is a diffusion operator then* $Lf(x) \ge 0$*.*

Proof. Let

$$L = \sum_{i,j=1}^{n} \sigma_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(x) \frac{\partial}{\partial x_i},$$

and let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function that attains a local minimum at x. We have

$$\begin{aligned} Lf(x) &= \sum_{i,j=1}^{n} \sigma_{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \\ &= \mathbf{tr} \left(\sigma(x) \mathbf{Hess} f(x) \right). \end{aligned}$$

Where $\sigma(x)$ is the symmetric and positive matrix with coefficient $\sigma_{ij}(x)$ and $\mathbf{Hess} f(x)$ is the Hessian matrix of f, that is the symmetric matrix with coefficient $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$. Since x is a local minimum of f, $\mathbf{Hess} f(x)$ is a positive matrix. We can now use the previous lemma to get the expected result. □

Actually, together with the linearity and the local property, this maximum principle is characteristic to the diffusion operators.

Theorem 2.1.8 *Let $\mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R})$ be the set of smooth functions $\mathbb{R}^n \to \mathbb{R}$ and $\mathcal{C}(\mathbb{R}^n, \mathbb{R})$ be the set of continuous functions $\mathbb{R}^n \to \mathbb{R}$. Let now*

$$L : \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R}) \to \mathcal{C}(\mathbb{R}^n, \mathbb{R})$$

be an operator such that:

1. *L is linear;*
2. *L is a local operator; That is, if $f, g \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R})$ coincide on a neighnorhood of x, then $Lf(x) = Lg(x)$;*
3. *If $f \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R})$ has a local minimum at x, $Lf(x) \geq 0$.*

Then L is a diffusion operator.

2.1.2 Diffusion equations

In the sequel, we consider a diffusion operator L as well as a Borel measure μ on $\mathbb{R}^n$. We will assume that if $f \in \mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$, $1 \leq p \leq +\infty$, then there is a unique real valued function $\phi(t, x)$ that is smooth on $(0, +\infty) \times \mathbb{R}^n$, continuous on $[0, +\infty) \times \mathbb{R}^n$ and that satisfies:

$$\frac{\partial \phi}{\partial t}(t, x) = L\phi(t, x), \quad \phi(0, x) = f(x). \tag{2.1.2}$$

This equation is called the diffusion equation associated with the diffusion operator L. In general, unlike the heat equation this is not possible to explicitly solve equation (7.1.1). To understand solutions, it is however extremely useful to introduce a one-parameter family of operators defined in the following way: If $f \in \mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$, $1 \le p \le +\infty$, then , for $t \ge 0$, we set

$$(\mathbf{P}_t f)(x) = \phi(t, x),$$

where ϕ is the unique solution of equation (7.1.1). We shall assume that $\mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$ is left stable by $\mathbf{P}_t$ for any $t \ge 0$ and $1 \le p \le +\infty$, and that for $f \in \mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$, $g \in \mathbf{L}^q_\mu(\mathbb{R}^n, \mathbb{R})$, with $\frac{1}{p} + \frac{1}{q} = 1$,

$$\int_{\mathbb{R}^n} f \mathbf{P}_t g d\mu = \int_{\mathbb{R}^n} g \mathbf{P}_t f d\mu. \tag{2.1.3}$$

In order to simplify the following discussion, we shall also assume that the Schwartz space $\mathcal{S}$ of smooth rapidly decreasing functions on $\mathbb{R}^n$ is a dense subset of $\mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$, $1 \le p < +\infty$ and is left globally stable by L and by $\mathbf{P}_t$ for any $t \ge 0$[1].

Remark 2.1.9

- *Observe that, due to the linearity of equation (7.1.1),* $\mathbf{P}_t$ *is for any* $t \ge 0$*, a linear operator* $\mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R}) \to \mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$.

- *Observe that if we apply (7.1.2) with* $f \in \mathcal{S}$ *and* $g = 1$*, then we get*

  $$\int_{\mathbb{R}^n} f \mathbf{P}_t 1 d\mu = \int_{\mathbb{R}^n} \mathbf{P}_t f d\mu.$$

  But $\mathbf{P}_t 1 = 1$*, therefore for* $f \in \mathcal{S}$*,*

  $$\int_{\mathbb{R}^n} \mathbf{P}_t f d\mu = \int_{\mathbb{R}^n} f d\mu.$$

  By differentiating with respect to t *this last equality, we also obtain that for* $f \in \mathcal{S}$

  $$\int_{\mathbb{R}^n} L f d\mu = 0.$$

[1] In general we work with an algebra of functions that depends on L and that satisfies these properties

- *We can also observe that for $f \in \mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$ and $s, t \geq 0$, $\mathbf{P}_{t+s} f = \mathbf{P}_t(\mathbf{P}_s f)$. This property $\mathbf{P}_{t+s} = \mathbf{P}_t \circ \mathbf{P}_s$ is called the semigroup property for the one-parameter family of operators $(\mathbf{P}_t)_{t \geq 0}$. Due to the fact that we moreover have at $t = 0$,*
$$\frac{d}{dt} \mathbf{P}_t f = Lf,$$
we often use the notation $\mathbf{P}_t = e^{tL}$.

We then have the following theorem that asserts that equation (7.1.1) preserves the positivity.

Theorem 2.1.10 *(Preservation of positivity) Let $f \in \mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$, $1 \leq p \leq +\infty$, be a μ-almost surely non negative function. Then for every $t \geq 0$, the function $\mathbf{P}_t f$ is μ-almost surely non negative.*

Proof. The first step is to observe that if ψ is a smooth convex function and if $f \in \mathcal{S}$, then we have

$$\begin{aligned} L(\psi \circ f)(x) &= \sum_{i,j=1}^n \sigma_{ij}(x) \frac{\partial^2 (\psi \circ f)}{\partial x_i \partial x_j}(x) + \sum_{i=1}^n b_i(x) \frac{\partial (\psi \circ f)}{\partial x_i}(x) \\ &= (\psi' \circ f)(x) Lf(x) + \sum_{i,j=1}^n \sigma_{ij}(x) \frac{\partial f}{\partial x_i}(x) \frac{\partial f}{\partial x_j}(x) \\ &\geq (\psi' \circ f)(x) Lf(x). \end{aligned}$$

Now,

$$\begin{aligned} \frac{\partial}{\partial t} \int_{\mathbb{R}^n} \psi(\mathbf{P}_t f) d\mu &= \int_{\mathbb{R}^n} \psi'(\mathbf{P}_t f) \frac{\partial}{\partial t} \mathbf{P}_t f d\mu \\ &= \int_{\mathbb{R}^n} \psi'(\mathbf{P}_t f) L \mathbf{P}_t f d\mu \\ &\leq \int_{\mathbb{R}^n} L \psi(\mathbf{P}_t f) d\mu = 0. \end{aligned}$$

The functional $\int_{\mathbb{R}^n} \psi(\mathbf{P}_t f) d\mu$ is therefore non increasing. We deduce that

$$\int_{\mathbb{R}^n} \psi(\mathbf{P}_t f) d\mu \leq \int_{\mathbb{R}^n} \psi(f) d\mu.$$

In particular, by approximating the function $x \to |x|$ by smooth convex functions, for $f \in \mathcal{S}$,

$$\int_{\mathbb{R}^n} |\mathbf{P}_t f| \, d\mu \leq \int_{\mathbb{R}^n} |f| d\mu. \tag{2.1.4}$$

Thus if $f \in \mathcal{S}$ is non negative,

$$\int_{\mathbb{R}^n} |\mathbf{P}_t f| \, d\mu \leq \int_{\mathbb{R}^n} |f| d\mu = \int_{\mathbb{R}^n} f d\mu = \int_{\mathbb{R}^n} \mathbf{P}_t f d\mu,$$

which implies that almost surely $\mathbf{P}_t f$ is non negative.

Let now $f \in \mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$, $1 \leq p < +\infty$, be a μ-almost surely non negative function. There exists a sequence of non negative functions $f_n \in \mathcal{S}$ that converges toward f in $\mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$. We are going to show that the sequence $\mathbf{P}_t f_n$ converges toward $\mathbf{P}_t f$, which implies that $\mathbf{P}_t f$ is μ-almost surely non negative. According to (2.1.4), for $m, n \in \mathbb{N}$,

$$\int_{\mathbb{R}^n} |\mathbf{P}_t f_n - \mathbf{P}_t f_m| \, d\mu \leq \int_{\mathbb{R}^n} |f_n - f_m| d\mu.$$

The sequence $\mathbf{P}_t f_n$ is therefore a Cauchy sequence and thus converges in $\mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$. For every $g \in \mathbf{L}^q_\mu(\mathbb{R}^n, \mathbb{R})$, with $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\int_{\mathbb{R}^n} f_m \mathbf{P}_t g d\mu = \int_{\mathbb{R}^n} g \mathbf{P}_t f_m d\mu,$$

and therefore

$$\int_{\mathbb{R}^n} f \mathbf{P}_t g d\mu = \int_{\mathbb{R}^n} g \lim_{m \to +\infty} \mathbf{P}_t f_m d\mu.$$

This implies that $\lim_{m \to +\infty} \mathbf{P}_t f_m = \mathbf{P}_t f$.

If $f \in \mathbf{L}^\infty_\mu(\mathbb{R}^n, \mathbb{R})$ is non negative, there exists a bounded an increasing sequence of non negative $f_n \in \mathbf{L}^1_\mu(\mathbb{R}^n, \mathbb{R})$ that converges to f, μ-almost surely. The sequence $\mathbf{P}_t f_n$ is increasing and bounded and therefore converges μ-almost surely. Finally, it is shown as above that this limit has to be equal to $\mathbf{P}_t f$. □

2.1.3 Associated Markov process

The preservation of positivity by diffusion equations makes the bridge between probability theory and the theory of diffusion equations. More precisely:

Exercise 2.1.11 *If A is a Borel set in $\mathbb{R}^n$, $t \geq 0$, and $x \in \mathbb{R}^n$ we define*

$$P_t(x, A) = (\mathbf{P}_t 1_A)(x),$$

where 1_A is the indicator function of A. Show that the following properties are satisfied:

1. *For $t \geq 0$ and $x \in \mathbb{R}^n$, $P_t(x, \cdot)$ is a probability measure on $\mathbb{R}^n$;*

2. *For $t \geq 0$ and A Borel set in $\mathbb{R}^n$ the application $x \to P_t(x, A)$ is measurable;*

3. *For $s, t \geq 0$, $x \in \mathbb{R}^n$ and A Borel set in $\mathbb{R}^n$,*

$$P_{t+s}(x, A) = \int_{\mathbb{R}^n} P_t(y, A) P_s(x, dy).$$

With this in hands, we can now show that there is a natural family of stochastic process associated to the equation (7.1.1).

Theorem 2.1.12 *Let ν be a probability measure on $\mathbb{R}^n$. There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process $(X_t)_{t \geq 0}$ such that:*

1. *The distribution of X_0 is ν ;*

2. *If $f : \mathbb{R}^n \to \mathbb{R}$ is a bounded and Borel function*

$$\mathbb{E}\left(f(X_{t+s}) \mid \mathcal{F}_s^X\right) = (\mathbf{P}_t f)(X_s), \quad s, t \geq 0$$

where $\mathcal{F}^X$ is the natural filtration [2] of X.

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

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

where we use the notation $P_t(x, \cdot)$ introduced in Exercise 2.1.11. $\mu_{t_0, t_1, \dots, t_n}$ is therefore a probability measure on $\mathbb{R}^n \times (\mathbb{R}^n)^{\otimes m}$. Since for a Borel set C in $\mathbb{R}^n$ and $x \in \mathbb{R}^n$ we have

$$P_{t+s}(x, C) = \int_{\mathbb{R}^n} P_t(y, C) P_s(x, dy),$$

we deduce that this family of probability satisfies the assumptions of the Daniell-Kolmogorov theorem. Therefore, we can find a process $(X_t)_{t \geq 0}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ whose finite dimensional distributions are given by the $\mu_{t_0, t_1, \dots, t_n}$'s. Let us now prove that this process satisfies the property stated in the theorem. First, the distribution of X_0 is ν because

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

[2] Recall that $\mathcal{F}_s^X$ is the smallest σ-algebra that makes measurable all the random variables $(X_{t_1}, \cdots, X_{t_m})$, $0 \leq t_1 \leq \cdots \leq t_m \leq s$.

We now have to prove that if $f : \mathbb{R}^n \to \mathbb{R}$ is a bounded and Borel function and if $0 < s, t$, then

$$\mathbb{E}\left(f(X_{s+t}) \mid \mathcal{F}_s^X\right) = (\mathbf{P}_t f)(X_s).$$

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

$$\mathbb{E}\left(f(X_{t_m})F(X_{t_0}, ..., X_{t_{m-1}})\right) = \mathbb{E}\left((\mathbf{P}_{t_m - t_{m-1}} f)(X_{t_{m-1}})F(X_{t_0}, ..., X_{t_{m-1}})\right).$$

But according to Fubini's theorem

$$\begin{aligned}
&\mathbb{E}\left(f(X_{t_m})F(X_{t_0}, ..., X_{t_{m-1}})\right) \\
=&\int_{(\mathbb{R}^n)^{\otimes(m+1)}} f(x_m)F(z, x_1, ..., x_{m-1})P_{t_1}(z, dx_1)P_{t_2-t_1}(x_1, dx_2)...P_{t_m - t_{m-1}}(x_{m-1}, dx_m)\nu(dz) \\
=&\int_{\mathbb{R}^n}\int_{(\mathbb{R}^n)^{\otimes m}} (\mathbf{P}_{t_m - t_{m-1}} f)(x_{m-1})F(z, x_1, ..., x_{m-1})P_{t_1}(z, dx_1)P_{t_2-t_1}(x_1, dx_2)...P_{t_{m-1} - t_{m-2}}(x_{m-2}, dx_{m-1})\nu(dz) \\
=&\mathbb{E}\left((\mathbf{P}_{t_m - t_{m-1}} f)(X_{t_{m-1}})F(X_{t_0}, ..., X_{t_{m-1}})\right).
\end{aligned}$$

□

Remark 2.1.13 *Observe the degree of freedom we have on the distribution of X_0.*

Remark 2.1.14 *It is possible to show (see the Theorem 2.2.10 below) that we can always work with a continuous version of $(X_t)_{t \geq 0}$.*

We are finally led to the following definition.

Definition 2.1.15 *A continuous stochastic process $(X_t)_{t \geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is called a diffusion process with semigroup $\mathbf{P}_t$ and infinitesimal generator L if for every bounded and Borel function $f : \mathbb{R}^n \to \mathbb{R}$*

$$\mathbb{E}\left(f(X_{t+s}) \mid \mathcal{F}_s^X\right) = (\mathbf{P}_t f)(X_s), \quad s, t \geq 0.$$

Remark 2.1.16

- *Diffusion processes are Markov processes, that is*

$$\mathbb{E}\left(f(X_{t+s}) \mid \mathcal{F}_s^X\right) = \mathbb{E}\left(f(X_{t+s}) \mid X_s\right), \quad s, t \geq 0.$$

- *The operator L is called the infinitesimal generator of $(X_t)_{t \geq 0}$ because it characterizes how the process evolves in small time intervals. If τ is small,*

$$\mathbb{E}\left(f(X_{t+\tau}) \mid \mathcal{F}_t^X\right) = (\mathbf{P}_\tau f)(X_t) = f(X_t) + \tau L f(X_t) + o(\tau).$$

One of the purpose of stochastic differential equations will be to explicitly construct diffusion processes.

2.2 Diffusion operator associated to a Markov process

We have seen in the previous section, that due to the positivity preserving property, it was possible to a associate a Markov process with a diffusion operator. In this section, we shall show that under mild conditions the converse is also true. Namely, we can associate a diffusion operator to most of the (continuous) Markov processes. Intuitively, a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a Markov process if for every bounded and Borel function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}\left(f(X_{t+s}) \mid \mathcal{F}_s^X\right) = \mathbb{E}\left(f(X_{t+s}) \mid X_s\right), \quad s, t \geq 0.$$

Let us now turn to a more precise definition. The notion of transition function for Markov process is the analogue in continuous time of the a transition matrix associated to a Markov chain.

Definition 2.2.1 *A transition function $\{P_t, t \geq 0\}$ on $\mathbb{R}^n$ is a family of kernels*

$$P_t : \mathbb{R} \times \mathcal{B}(\mathbb{R}^n) \to [0, 1]$$

such that:

1. *For $t \geq 0$ and $x \in \mathbb{R}^n$, $P_t(x, \cdot)$ is a probability measure on $\mathbb{R}^n$;*
2. *For $t \geq 0$ and A Borel set in $\mathbb{R}^n$ the application $x \to P_t(x, A)$ is measurable;*
3. *For $s, t \geq 0$, $x \in \mathbb{R}^n$ and A Borel set in $\mathbb{R}^n$,*

$$P_{t+s}(x, A) = \int_{\mathbb{R}^n} P_t(y, A) P_s(x, dy). \tag{2.2.5}$$

The relation (2.2.5) is called the Chapman-Kolmogorov relation.
A transition function can also be seen as a one parameter family of linear operators $(\mathbf{P}_t)_{t \geq 0}$ from the space of bounded Borel functions into itself:

$$(\mathbf{P}_t f)(x) = \int_{\mathbb{R}} f(y) P_t(x, dy).$$

With this in hands, we can now provide the definition of a (homogeneous) Markov process:

Definition 2.2.2 *A stochastic process $(X_t)_{t \geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a Markov process if there exists a transition function $\{P_t, t \geq 0\}$ on $\mathbb{R}^n$ such that for every bounded and Borel function $f : \mathbb{R}^n \to \mathbb{R}$,*

$$\mathbb{E}\left(f(X_{t+s}) \mid \mathcal{F}_s^X\right) = \mathbf{P}_t f(X_s), \quad s, t \geq 0.$$

Remark 2.2.3 *We may also speak of the Markov property with respect to a given filtration. A stochastic process $(X_t)_{t\geq 0}$ defined on a filtered probability space*[3] $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ *is called a Markov process with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ if there exists a transition function $\{P_t, t \geq 0\}$ on $\mathbb{R}^n$ such that for every bounded and Borel function $f : \mathbb{R}^n \to \mathbb{R}$,*

$$\mathbb{E}\left(f(X_{t+s}) \mid \mathcal{F}_s\right) = \mathbf{P}_t f(X_s), \quad s, t \geq 0.$$

If $\{P_t, t \geq 0\}$ is a transition function, then the following properties are satisfied:

- $\mathbf{P}_t 1 = 1$;
- For every $t \geq 0$, $\mathbf{P}_t$ is a positive operator, in the sense that if f is non negative, so is $\mathbf{P}_t f$;
- For every $t \geq 0$, $\mathbf{P}_t$ is a contraction from the space of bounded Borel functions into itself (that is, it is a continuous operator with a norm smaller than 1);
- The semigroup property holds: For every $s, t \geq 0$,

$$\mathbf{P}_{t+s} = \mathbf{P}_t \mathbf{P}_s, s, t \geq 0.$$

We have already met such one-parameter families of operators, when we solved diffusion equations. Our goal, here, will now be to identify sufficient conditions ensuring that $(\mathbf{P}_t)_{t\geq 0}$ is actually the semigroup associated with a diffusion equation.

Exercise 2.2.4 *Let $f : \mathbb{R}_{\geq 0} \to \mathcal{M}_n(\mathbb{R})$ (set of $n \times n$ matrices) be a function such that f is continuous at 0, $f(0) = \mathbf{I}_n$ and*

$$f(t+s) = f(t)f(s), s, t \geq 0.$$

Show that there exists $M \in \mathcal{M}_n(\mathbb{R})$ such that for $t \geq 0$,

$$f(t) = e^{tM}.$$

This exercises suggests that with a continuity assumption at 0 we should be able to write

$$\mathbf{P}_t = e^{t\mathcal{L}},$$

where L is an operator is an operator defined on some function space.
In what follows, we denote by $\mathcal{C}_0(\mathbb{R}^n, \mathbb{R})$ the set of continuous functions $f : \mathbb{R}^n \to \mathbb{R}$ such that $\lim_{\|x\|\to+\infty} f(x) = 0$.

[3] Recall that a filtration $(\mathcal{F}_t)_{t\geq 0}$ is an increasing sequence of sub σ-fields of $\mathcal{F}$

Definition 2.2.5 *Let $\{P_t, t \geq 0\}$ be a transition function. We say that $\{P_t, t \geq 0\}$ is a Feller-Dynkin transition function if:*

1. $\mathbf{P}_t : \mathcal{C}_0(\mathbb{R}^n, \mathbb{R}) \to \mathcal{C}_0(\mathbb{R}^n, \mathbb{R})$;
2. $\mathbf{P}_0 = \mathbf{Id}$;
3. $\forall f \in \mathcal{C}_0(\mathbb{R}^n, \mathbb{R}), \forall x \in \mathbb{R}^n, \lim_{t\to 0}(\mathbf{P}_t f)(x) = f(x)$.

Remark 2.2.6 *It is possible to show that if $\{P_t, t \geq 0\}$ is a Feller-Dynkin transition function, then actually $\forall f \in \mathcal{C}_0(\mathbb{R}^n, \mathbb{R})$*

$$\lim_{t\to 0} \|\mathbf{P}_t f - f\|_\infty = 0.$$

Remark 2.2.7 *If the transition function of a Markov process is a Feller-Dynkin transition function, then we say that the process is a Feller-Dynkin process.*

We then have the following result:

Proposition 2.2.8 *There exists a dense subset $\mathcal{D} \subset \mathcal{C}_0(\mathbb{R}^n, \mathbb{R})$ such that for $f \in \mathcal{D}$, there exists $g \in \mathcal{C}_0(\mathbb{R}, \mathbb{R})$ such that:*

$$\lim_{t\to 0} \left\| \frac{\mathbf{P}_t f - f}{t} - g \right\|_\infty .$$

Proof. Let us consider the following bounded operators on $\mathcal{C}_0(\mathbb{R}^n, \mathbb{R})$:

$$A_t = \frac{1}{t}\int_0^t \mathbf{P}_s ds.$$

For $f \in \mathcal{C}_0(\mathbb{R}^n, \mathbb{R})$ and $h > 0$,

$$\begin{aligned} \frac{1}{t}\left(\mathbf{P}_t A_h f - A_h f\right) &= \frac{1}{ht}\int_0^h \mathbf{P}_{s+t} f - \mathbf{P}_s f ds \\ &= \frac{1}{ht}\int_0^t \mathbf{P}_{s+h} f - \mathbf{P}_s f ds. \end{aligned}$$

Therefore,

$$\lim_{t\to 0} \frac{1}{t}\left(\mathbf{P}_t A_h f - A_h f\right) = \frac{1}{h}\left(\mathbf{P}_h f - f\right).$$

Since $\lim_{h\to 0} A_h f = f$, we get the expected result. $\square$

If $f \in \mathcal{D}$, the g corresponding to f is of course unique and shall be denoted by L. The linear operator $L : \mathcal{D} \to \mathcal{C}_0(\mathbb{R}, \mathbb{R})$ is called the infinitesimal operator of the Feller-Dynkin transition function $\{P_t, t \geq 0\}$ and the space $\mathcal{D}$ is called the domain of L. We would like to show that, under suitable assumptions, L is a diffusion operator.

Theorem 2.2.9 *Let $\{P_t, t \geq 0\}$ be a Feller-Dynkin transition function on $\mathbb{R}^n$ such that for every $\varepsilon > 0$, and every compact set $K \subset \mathbb{R}^n$,*

$$\lim_{t \to 0} \sup_{x \in K} \frac{1}{t} P_t(x, \mathbf{B}(x, \varepsilon)^c) = 0,$$

where

$$\mathbf{B}(x, \varepsilon)^c = \{y \in \mathbb{R}^n, \|y - x\| \geq \varepsilon\}.$$

Let us assume that the domain $\mathcal{D}$ of the the infinitesimal operator L of $\{P_t, t \geq 0\}$ contains the space of smooth and compactly supported functions. Then, L is a diffusion operator, that is

$$L = \sum_{i,j=1}^{n} \sigma_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(x) \frac{\partial}{\partial x_i},$$

where b_i and σ_{ij} are continuous functions on $\mathbb{R}^n$ and for every $x \in \mathbb{R}^n$, the matrix $(\sigma_{ij}(x))_{1 \leq i,j \leq n}$ is a symmetric and positive matrix.

Proof. We are going to use Theorem 2.1.8, to show that L has to be a diffusion operator. We have to show that under our assumptions:

1. L is linear;
2. L is a local operator;
3. If f is smooth, compactly supported, and has a maximum at x, then $Lf(x) \leq 0$.

The first point is obvious. Let now f be a smooth and compactly supported function that vanishes in a neighborhood of $x \in \mathbb{R}^n$. Since for every $\varepsilon > 0$,

$$\lim_{t \to 0} \frac{1}{t} P_t(x, \mathbf{B}(x, \varepsilon)^c) = 0,$$

we have

$$\lim_{t \to 0} \frac{\mathbf{P}_t f(x)}{t} = 0,$$

and therefore $Lf(x) = 0$.
Finally, let f be a smooth, compactly supported function that attains its maximum at x. We have $f \leq f(x)$. Therefore $\mathbf{P}_t f - f(x) \leq 0$, which implies $Lf(x) \leq 0$. □

Finally, for Feller-Dynkin transition functions, we have the following existence theorem.

Theorem 2.2.10 *Let $\{P_t, t \geq 0\}$ be a Feller-Dynkin transition function on $\mathbb{R}^n$ such that for every $\varepsilon > 0$, and every compact set $K \subset \mathbb{R}^n$,*

$$\lim_{t \to 0} \sup_{x \in K} \frac{1}{t} P_t(x, \mathbf{B}(x, \varepsilon)^c) = 0.$$

Let now ν be a probability measure on $\mathbb{R}^n$. There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a continuous stochastic process $(X_t)_{t \geq 0}$ such that:

- *The distribution of X_0 is ν;*
- *$(X_t)_{t \geq 0}$ is a Markov process with transition function $\{P_t, t \geq 0\}$.*

2.3 Examples of diffusion processes

2.3.1 Brownian motion

Let us recall that a n-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a process $(B_t)_{t \geq 0}$ that can be written

$$B_t = (B_t^1, \cdots, B_t^n)$$

where $B^1, \cdots, B^n$ are independent one-dimensional Brownian motions.

Proposition 2.3.1 *$(B_t)_{t \geq 0}$ is a diffusion process with semigroup*

$$\mathbf{P}_t f(x) = \frac{1}{(2\pi t)^{\frac{n}{2}}} \int_{\mathbb{R}^n} e^{-\frac{\|x-y\|^2}{2t}} f(y) dy$$

and infinitesimal generator $L = \frac{1}{2}\Delta$.

Proof. We already know that the unique solution of

$$\frac{\partial \phi}{\partial t}(t, x) = \frac{1}{2} \Delta \phi(t, x), \quad \phi(0, x) = f(x),$$

is given by

$$\phi(t, x) = \mathbf{P}_t f(x) = \frac{1}{(2\pi t)^{\frac{n}{2}}} \int_{\mathbb{R}^n} e^{-\frac{\|x-y\|^2}{2t}} f(y) dy.$$

Let now $f : \mathbb{R}^n \to \mathbb{R}$ be a bounded and Borel function and let $s, t \geq 0$. We have:

$$\mathbb{E}(f(B_{t+s}) \mid \mathcal{F}_s) = \mathbb{E}(f(B_{t+s} - B_s + B_s) \mid \mathcal{F}_s^B).$$

Since $B_{t+s} - B_s$ is independent from $\mathcal{F}_s^B$, we deduce that

$$\mathbb{E}(f(B_{t+s}) \mid \mathcal{F}_s^B) = \mathbb{E}(f(B_{t+s}) \mid B_s).$$

For $x \in \mathbb{R}^n$,

$$\mathbb{E}(f(B_{t+s}) \mid B_s = x) = \mathbb{E}(f(B_{t+s} - B_s + B_s) \mid B_s = x) = \mathbb{E}(f(X_t + x)),$$

where X_t is a random variable independent from B_s and such that X_t is normally distributed with mean 0 and variance $t\mathbf{I}_n$. Therefore,

$$\mathbb{E}(f(B_{t+s}) \mid B_s = x) = \int_{\mathbb{R}^n} f(x+y) \frac{e^{-\frac{\|y\|^2}{2t}}}{(\sqrt{2\pi t})^n} dy$$

and

$$\mathbb{E}(f(B_{t+s}) \mid \mathcal{F}_s^B) = \frac{1}{(2\pi t)^{\frac{n}{2}}} \int_{\mathbb{R}^n} e^{-\frac{\|B_s - y\|^2}{2t}} f(y) dy.$$

□

2.3.2 Ornstein-Uhlenbeck process

Let $\theta \in \mathbb{R} \backslash \{0\}$. The diffusion operator on $\mathbb{R}$

$$\mathcal{L} = \theta x \frac{d}{dx} + \frac{1}{2} \frac{d^2}{dx^2}$$

is called the Ornstein-Uhlenbeck operator. Let $(B_t)_{t \geq 0}$ be a one-dimensional Brownian motion and let $\theta \in \mathbb{R} \backslash \{0\}$. We consider the process

$$X_t = e^{\theta t} B_{\frac{1 - e^{-2\theta t}}{2\theta}}.$$

Proposition 2.3.2 $(X_t)_{t \geq 0}$ *is a diffusion process with semigroup*

$$(\mathbf{P}_t f)(x) = \int_{\mathbb{R}} f\left(e^{\theta t} x + \sqrt{\frac{e^{2\theta t} - 1}{2\theta}} y \right) \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy.$$

and infinitesimal generator $\mathcal{L}$.

Proof. It is easily seen by a straightforward computation that if ϕ solves the heat equation

$$\frac{\partial \phi}{\partial t}(t,x) = \frac{1}{2} \frac{\partial^2 \phi}{\partial x^2} \phi(t,x), \quad \phi(0,x) = f(x),$$

then the function

$$\psi(t,x) = \phi\left(\frac{e^{2\theta t}-1}{2\theta}, e^{\theta t}x\right)$$

solves the diffusion equation

$$\frac{\partial\psi}{\partial t}(t,x) = \theta x\frac{\partial\psi}{\partial x}(t,x) + \frac{1}{2}\frac{\partial^2\psi}{\partial x^2}(t,x), \quad \phi(0,x) = f(x). \tag{2.3.6}$$

Therefore, the unique solution to equation (2.3.6) is given by

$$\psi(t,x) = (\mathbf{P}_t f)(x) = \int_{\mathbb{R}} f\left(e^{\theta t}x + \sqrt{\frac{e^{2\theta t}-1}{2\theta}}y\right)\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}dy.$$

We now compute

$$\begin{aligned}
\mathbb{E}\left(f(X_{t+s}) \mid \mathcal{F}_s^X\right) &= \mathbb{E}\left(f\left(e^{\theta(t+s)}B_{\frac{1-e^{-2\theta(t+s)}}{2\theta}}\right) \mid \mathcal{F}_s^X\right) \\
&= \frac{1}{\left(2\pi\frac{e^{-2\theta s}-e^{-2\theta(t+s)}}{2\theta}\right)^{\frac{n}{2}}}\int_{\mathbb{R}^n}\exp\left(-\frac{\theta\|y\|^2}{e^{-2\theta s}-e^{-2\theta(t+s)}}\right)f\left(e^{\theta(t+s)}(y+e^{-\theta s}x)\right)dy \\
&= (\mathbf{P}_t f)(X_s)
\end{aligned}$$

□

2.3.3 Black-Scholes diffusion

On $\mathbb{R}$, the diffusion operator

$$\mathcal{L} = \mu x\frac{d}{dx} + \frac{1}{2}\sigma^2 x^2\frac{d^2}{dx^2}.$$

is called the Black-Scholes operator. Let $(B_t)_{t\geq 0}$ be a one-dimensional Brownian motion and let $\mu \in \mathbb{R}$, $\sigma > 0$. We consider the process

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

We let as an exercise the following result:

Proposition 2.3.3 *$(X_t)_{t\geq 0}$ is a diffusion process with semigroup*

$$(\mathbf{P}_t f)(x) = \int_{\mathbb{R}} f\left(xe^{\left(\mu-\frac{\sigma^2}{2}\right)t+\sigma y}\right)\frac{e^{-\frac{y^2}{2t}}}{\sqrt{2\pi t}}dy.$$

and infinitesimal generator $\mathcal{L}$.

2.3.4 Bessel process

On $\mathbb{R}_{\geq 0}$, the operator

$$\mathcal{L} = \frac{n-1}{2}\frac{d}{dx} + \frac{1}{2}\frac{d^2}{dx^2},$$

where $n \geq 1$ is an integer, is called the Bessel diffusion operator. Let $(B_t)_{t\geq 0}$ be a n-dimensional Brownian motion. We consider

$$X_t = \sqrt{\sum_{i=1}^{n}(B_t^i)^2}.$$

We let as an exercise

Proposition 2.3.4 *$(X_t)_{t\geq 0}$ is a diffusion process with semigroup*

$$\begin{aligned}(\mathbf{P}_t f)(x) &= \frac{1}{t}\int_{\mathbb{R}_{\geq 0}} f(y)\left(\frac{y}{x}\right)^{\frac{n}{2}-1} I_{\frac{n}{2}-1}\left(\frac{xy}{t}\right) e^{-\frac{x^2+y^2}{2}}\, dy, && x>0\\ &= \frac{2^{1-\frac{n}{2}}\Gamma(n/2)}{t^{n/2}}\int_{\mathbb{R}_{\geq 0}} f(y)\, y^{n-1} e^{-\frac{y^2}{2t}}\, dy && x=0\end{aligned}$$

and infinitesimal generator $\mathcal{L}$. The function $I_{\frac{n}{2}-1}$ is the modified Bessel function of the first kind with index $\frac{n}{2}-1$.

Chapter 3

Stochastic differential equations

Stochastic differential equations provide a pathwise construction of diffusions.

3.1 Stochastic calculus

In this section, we review briefly (mainly without proof) the theory of stochastic calculus. The stochastic integration is a natural, easy and fruitful integration theory which is due to K. Itô in 1944. For further details on this theory we refer to the excellent books by Revuz-Yor [17] or by Protter [16].

3.1.1 Martingales

In what follows, we work on a filtered probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathcal{F}, \mathbb{P})$ which satisfies the usual conditions, that is:

1. $(\mathcal{F}_t)_{t\geq 0}$ is a filtration, i.e. an increasing family of sub-σ-fields of $\mathcal{F}$;

2. for any $t \geq 0$, $\mathcal{F}_t$ is complete with respect to $\mathbb{P}$, i.e. every subset of a set of measure zero is contained in $\mathcal{F}_t$;

3. $(\mathcal{F}_t)_{t\geq 0}$ is right continuous, i.e. for any $t \geq 0$,

$$\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s.$$

Consider an adapted and continuous process $(M_t)_{t\geq 0}$ defined on $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathcal{F}, \mathbb{P})$.

Definition 3.1.1 *The process $(M_t)_{t\geq 0}$ is said to be a martingale (with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$) if:*

1. *For* $t \geq 0$, $\mathbb{E}\left(\mid M_t \mid\right) < +\infty$*;*

2. *For* $0 \leq s \leq t$, $\mathbb{E}\left(M_t \mid \mathcal{F}_s\right) = M_s$*.*

For instance a standard Brownian motion is a martingale.
In the theory of stochastic processes, it is often useful to deal with random times. A real valued positive random variable T is said to be a stopping time with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ if for any $t \geq 0$,

$$\{T \leq t\} \in \mathcal{F}_t.$$

If T is a stopping time the smallest σ-field which contains all the events $\{T \leq t\}$, $t \geq 0$, is denoted $\mathcal{F}_T$. For martingales, we have the following proposition, which is known as the stopping theorem.

Proposition 3.1.2 *The following properties are equivalent:*

1. *The process* $(M_t)_{t\geq 0}$ *is a martingale;*

2. *For any bounded stopping time* T, $\mathbb{E}(M_T) = \mathbb{E}(M_0)$*;*

3. *For any pair of bounded stopping times* S *and* T*, with* $S \leq T$, $\mathbb{E}\left(M_T \mid \mathcal{F}_S\right) = M_S$*.*

Actually, if T is a bounded stopping time, then the process $(M_{t\wedge T})_{t\geq 0}$ is also a martingale.
A martingale $(M_t)_{t\geq 0}$ is said to be square integrable if for $t \geq 0$, $\mathbb{E}\left(M_t^2\right) < +\infty$. In that case, from Jensen's inequality the function $t \to \mathbb{E}\left(M_t^2\right)$ is increasing. We also have the so-called Doob's inequality

$$\mathbb{E}\left(\sup_{t\geq 0} M_t^2\right) \leq 4 \sup_{t\geq 0} \mathbb{E}\left(M_t^2\right),$$

which is one of the cornerstone of the stochastic integration. Observe therefore that if $\sup_{t\geq 0} \mathbb{E}\left(M_t^2\right) < +\infty$, the martingale $(M_t)_{t\geq 0}$ is uniformly integrable and converges in L^2 to a square integrable random variable M_∞ which satisfies

$$\mathbb{E}\left(M_\infty \mid \mathcal{F}_t\right) = M_t,\ t \geq 0.$$

If $(M_t)_{t\geq 0}$ is a square integrable martingale, there exists a unique increasing process denoted $(\langle M\rangle_t)_{t\geq 0}$ which satisfies:

1. $\langle M\rangle_0 = 0$;

2. The process $(M_t^2 - \langle M\rangle_t)_{t\geq 0}$ is a martingale.

This increasing process $(\langle M\rangle_t)_{t\geq 0}$ is called the quadratic variation of the martingale $(M_t)_{t\geq 0}$. This terminology comes from the following property. If $0 = t_0 \leq t_1 \leq \cdots \leq t_n = t$ is a subdivision of the time interval $[0,t]$ whose mesh tends to 0, then in probability

$$\lim_{n\to+\infty}\sum_{i=0}^{n-1}\left(M_{t_{i+1}}-M_{t_i}\right)^2 = \langle M\rangle_t.$$

For technical reasons (localization procedures), we often have to consider a wider class than martingales.

Definition 3.1.3 *The process $(M_t)_{t\geq 0}$ is said to be a local martingale (with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$) if there exists a sequence $(T_n)_{n\geq 0}$ of stopping times such that:*

1. *The sequence $(T_n)_{n\geq 0}$ is increasing and $\lim_{n\to+\infty} T_n = +\infty$ almost surely;*
2. *For every $n \geq 1$, the process $(M_{t\wedge T_n})_{t\geq 0}$ is a uniformly integrable martingale with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$.*

A martingale is always a local martingale but the converse is not true. Nevertheless, a local martingale $(M_t)_{t\geq 0}$ such that for every $t \geq 0$,

$$\mathbb{E}\left(\sup_{s\leq t} \mid M_s \mid\right) < +\infty,$$

is a martingale. If $(M_t)_{t\geq 0}$ is a local martingale, there still exists a unique increasing process denoted $(\langle M\rangle_t)_{t\geq 0}$ which satisfies:

1. $\langle M\rangle_0 = 0$;
2. The process $(M_t^2 - \langle M\rangle_t)_{t\geq 0}$ is a local martingale.

This increasing process $(\langle M\rangle_t)_{t\geq 0}$ is called the quadratic variation of the local martingale $(M_t)_{t\geq 0}$. By polarization, it is easily seen that, more generally, if $(M_t)_{t\geq 0}$ and $(N_t)_{t\geq 0}$ are two continuous local martingales, then there exists a unique continuous process denoted $(\langle M,N\rangle_t)_{t\geq 0}$ and called the quadratic covariation of $(M_t)_{t\geq 0}$ and $(N_t)_{t\geq 0}$ which satisfies:

1. $\langle M,N\rangle_0 = 0$;
2. The process $(M_tN_t - \langle M,N\rangle_t)_{t\geq 0}$ is a local martingale.

Before we turn to the theory of stochastic integration, we conclude this section with Lévy's characterization of Brownian motion.

Proposition 3.1.4 *Let $(M_t)_{t\geq 0}$ be a d-dimensional continuous local martingale such that $M_0 = 0$ and*

$$\langle M^i, M^i \rangle_t = t, \ \langle M^i, M^j \rangle_t = 0 \text{ if } i \neq j.$$

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

3.1.2 Stochastic integration

Let $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathcal{F}, \mathbb{P})$ be a filtered probability space which satisfies the usual conditions specified before. We aim now at defining an integral $\int_0^t H_s dM_s$ where $(M_t)_{t\geq 0}$ is an adapted continuous square integrable martingale such that $\sup_{t\geq 0} \mathbb{E}\left(M_t^2\right) < +\infty$ and $(H_t)_{t\geq 0}$ an adapted process which shall be in a *good* class. Observe that such an integral could not be defined trivially since the Young's integration theory does not cover the integration against paths which are less than $\frac{1}{2}$-Hölder continuous. First, we define the class of integrands. The predictable σ-field $\mathcal{P}$ associated with the filtration $(\mathcal{F}_t)_{t\geq 0}$ is the σ-field generated on $\mathbb{R}_{\geq 0} \times \Omega$ by the space of indicator functions $\mathbf{1}_{]S,T]}$, where S and T are two stopping times such that $S \leq T$. An adapted stochastic process $(H_t)_{t\geq 0}$ is said to be predictable if the application $(t, \omega) \rightarrow H_t(\omega)$ is measurable with respect to $\mathcal{P}$. Observe that a continuous adapted process is predictable.

Let us first assume that $(H_t)_{t\geq 0}$ is a predictable elementary process that can be written

$$H_t = \sum_{i=1}^{n-1} H_i \mathbf{1}_{]T_i, T_{i+1}]}(t)$$

where H_i is a $\mathcal{F}_{T_i}$ measurable bounded random variable, and where $(T_i)_{1\leq i\leq n}$ is a finite increasing sequence of stopping times. In that case, a natural definition for $\int_0^t H_s dM_s$ is

$$\int_0^t H_s dM_s = \sum_{i=1}^{n-1} H_i \left(M_{t\wedge T_{i+1}} - M_{t\wedge T_i}\right).$$

Then, we observe that the process $\left(\int_0^t H_s dM_s\right)_{t\geq 0}$ is a bounded martingale which satisfies furthermore from Doob's inequality

$$\mathbb{E}\left(\sup_{t\geq 0}\left(\int_0^t H_s dM_s\right)^2\right) \leq 4 \parallel H \parallel_\infty^2 \mathbb{E}(M_\infty^2).$$

We also note that

$$\mathbb{E}\left(\left(\int_0^t H_s dM_s\right)^2\right) = \mathbb{E}\left(\sum_{i=1}^{n-1} H_i \left(M_{t\wedge T_{i+1}} - M_{t\wedge T_i}\right)^2\right).$$

Assume now that $(H_t)_{t\geq 0}$ is a bounded continuous adapted process. To define $\int_0^t H_s dM_s$, the idea is of course to approximate $(H_t)_{t\geq 0}$ with elementary processes $(H_t^p)_{t\geq 0}$ and to check the convergence of $\left(\int_0^t H_s^p dM_s\right)_{t\geq 0}$ with respect to a suitable norm. Precisely, let us define for any $p \in \mathbb{N}^*$, the following stopping times:

$$T_0^p = 0$$

$$T_1^p = \inf \left\{ t > 0, \mid H_t \mid \geq \frac{1}{p} \right\},$$

and by iteration

$$T_{n+1}^p = \inf \left\{ t > T_n^p, \mid H_t - H_{T_n^p} \mid \geq \frac{1}{p} \right\}.$$

We now define

$$H_t^p = \sum_{i=1}^{n-1} H_{T_i^p} \mathbf{1}_{]T_i^p, T_{i+1}^p]}(t).$$

For this sequence of processes $(H_t^p)_{t\geq 0}$, it easy to show that

$$\mathbb{E}\left(\sup_{t\geq 0} \left(\int_0^t (H_s^p - H_s^q) dM_s \right)^2 \right) \to_{p,q\to+\infty} 0.$$

From this, we can deduce that there exists a continuous martingale denoted $\left(\int_0^t H_s dM_s\right)_{t\geq 0}$ such that

$$\int_0^t H_s dM_s = \lim_{p\to+\infty} \int_0^t H_s^p dM_s$$

uniformly for t on compact sets. We furthermore have

$$\mathbb{E}\left(\sup_{t\geq 0} \left(\int_0^t H_s dM_s \right)^2 \right) \leq 4 \parallel H \parallel_\infty^2 \mathbb{E}(M_\infty^2).$$

and

$$\mathbb{E}\left(\left(\int_0^t H_s dM_s \right)^2 \right) = \mathbb{E}\left(\int_0^t H_s^2 d\langle M \rangle_s \right).$$

Now, by localization, it is not difficult to extend naturally the definition of $\int_0^t H_s dM_s$ in the general case where:

1. $(M_t)_{t\geq 0}$ is a semimartingale, that is, $(M_t)_{t\geq 0}$ can be written

   $$M_t = A_t + N_t,$$

   where $(A_t)_{t\geq 0}$ is a bounded variation process and $(N_t)_{t\geq 0}$ is a local martingale;

2. $(H_t)_{t\geq 0}$ is a locally bounded predictable process.

In this setting, we have:

$$\int_0^t H_s dM_s = \int_0^t H_s dA_s + \int_0^t H_s dN_s.$$

Observe that, since $(A_t)_{t\geq 0}$ is a bounded variation process, the integral $\int_0^t H_s dA_s$ is simply a Riemann-Stieltjes integral. The process $\left(\int_0^t H_s dN_s\right)_{t\geq 0}$ is a local martingale.

The class of semimartingales appears then as a good class of integrators in the theory of stochastic integration. It can be shown that this is actually the widest possible class if we wish to obtain a *natural* integration theory. The decomposition of a semimartingale $(M_t)_{t\geq 0}$ under the form

$$M_t = A_t + N_t,$$

is essentially unique under the condition $N_0 = 0$. The process $(A_t)_{t\geq 0}$ is called the bounded variation part of $(M_t)_{t\geq 0}$. The process $(N_t)_{t\geq 0}$ is called the local martingale part of $(M_t)_{t\geq 0}$. If $(M_t^1)_{t\geq 0}$ and $(M_t^2)_{t\geq 0}$ are two semimartingales, then we define the quadratic covariation $(\langle M^1, M^2\rangle_t)_{t\geq 0}$ of $(M_t^1)_{t\geq 0}$ and $(M_t^2)_{t\geq 0}$ as being $(\langle N^1, N^2\rangle_t)_{t\geq 0}$ where $(N_t^1)_{t\geq 0}$ and $(N_t^2)_{t\geq 0}$ are the local martingale parts.

Throughout this course we shall only deal with continuous processes so that in the sequel, we shall often omit to precise the continuity of the processes which will be considered. Moreover, we shall preferably use Stratonovitch's integration rather than Itô's integration. If $(N_t)_{0\leq t\leq T}$, $T > 0$, is an $\mathcal{F}$-adapted real valued local martingale and if $(\Theta_t)_{0\leq t\leq T}$ is an $\mathcal{F}$-adapted continuous process satisfying $\mathbb{E}\left(\int_0^T \Theta_t^2 d\langle N\rangle_t\right)$, then by definition,

$$\int_0^T \Theta_t \circ dN_t = \int_0^T \Theta_t \cdot dN_t + \frac{1}{2}\langle \Theta, N\rangle_T,$$

where:

1. $\int_0^T \Theta_t \circ dN_t$ is the Stratonovitch integral of $(\Theta_t)_{0\leq t\leq T}$ against $(N_t)_{0\leq t\leq T}$;

2. $\int_0^T \Theta_t \cdot dN_t$ is the Itô integral of $(\Theta_t)_{0\leq t\leq T}$ against $(N_t)_{0\leq t\leq T}$;

3. $\langle \Theta, N\rangle_T$ is the quadratic covariation at time T between $(\Theta_t)_{0\leq t\leq T}$ and $(N_t)_{0\leq t\leq T}$.

3.1.3 Itô's formula

The Itô's formula is certainly the most important formula of stochastic calculus.

Theorem 3.1.5 *Let $(X_t)_{t\geq 0} = (X_t^1, \cdots, X_t^n)_{t\geq 0}$ be a n- dimensional continuous semimartingale. Let now $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2 function. We have*

$$\begin{aligned} f(X_t) &= f(X_0) + \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x_i}(X_s)dX_s^i + \frac{1}{2}\sum_{i,j=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)d\langle X^i, X^j\rangle_s \\ &= f(X_0) + \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x_i}(X_s) \circ dX_s^i. \end{aligned}$$

3.2 Stochastic differential equations

We now turn to the theory of stochastic differential equations. Stochastic differential equations are the differential equations corresponding to the theory of the stochastic integration. They provide a pathwise construction of diffusion processes.

3.2.1 Existence and uniqueness for solutions

Let us consider a diffusion operator L on $\mathbb{R}^n$,

$$L = \frac{1}{2}\sum_{i,j=1}^n a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x)\frac{\partial}{\partial x_i},$$

where b_i and a_{ij} are continuous functions on $\mathbb{R}^n$ such that for every $x \in \mathbb{R}^n$, the matrix $(a_{ij}(x))_{1\leq i,j\leq n}$ is a symmetric and positive matrix. Since the matrix a is symmetric and positive, it admits a square root, that is, there exists a symmetric and positive matrix σ such that

$$\sigma^2 = a.$$

Let us now introduce a filtered probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathcal{F}, \mathbb{P})$ which satisfies the usual conditions and on which is defined a n-dimensional Brownian motion $(B_t)_{t\geq 0}$. The main theorem is the following:

Theorem 3.2.1 *Let us assume that b and σ are smooth, and that their derivatives of any order are bounded.*
Then, for every $x_0 \in \mathbb{R}^n$, there exists a unique and adapted process $(X_t^{x_0})_{t\geq 0}$ such that for $t \geq 0$

$$X_t^{x_0} = x_0 + \int_0^t b(X_s^{x_0})ds + \int_0^t \sigma(X_s^{x_0})dB_s. \tag{3.2.1}$$

Moreover, if we denote for a bounded and Borel function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbf{P}_t f(x) = \mathbb{E}\left(f(X_t^x)\right), \quad t \geq 0, x \in \mathbb{R}^n,$$

then, for every $x_0 \in \mathbb{R}^n$, $(X_t^{x_0})_{t\geq 0}$ is a Markov process with semigroup $(\mathbf{P}_t)_{t\geq 0}$.

Finally, if $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth and compactly supported function, then the function

$$\phi(t,x) = \mathbb{E}\left(f(X_t^x)\right) = \mathbf{P}_t f(x)$$

is the unique bounded solution of the diffusion equation

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

Proof.
For the first part of the theorem, as for the proof of the Cauchy-Lipschitz theorem that asserts existence and uniqueness of solutions for ordinary differential equations, the idea is to apply a fixed point theorem in a convenient Banach space. Let us first observe that from our assumptions, there exists $K > 0$ such that

1. $\|b(x) - b(y)\| + \|\sigma(x) - \sigma(y)\| \leq K\|x - y\|$, $x, y \in \mathbb{R}^n$;
2. $\|b(x)\| + \|\sigma(x)\| \leq K(1 + \|x\|)$, $x \in \mathbb{R}^n$.

For $T > 0$, let us consider the space $\mathcal{E}_T$ of continuous and adapted processes such that

$$\mathbb{E}\left(\sup_{0\leq s\leq T} \mid X_s \mid^2\right) < +\infty$$

endowed with the norm

$$\| X \|^2 = \mathbb{E}\left(\sup_{0\leq s\leq T} \mid X_s \mid^2\right).$$

It is easily seen that $(\mathcal{E}_T, \|\cdot\|)$ is a Banach space.

Step one: We first prove that if a continuous and adapted process $(X_t^{x_0})_{t\geq 0}$ is a solution of the equation (3.2.1) then, for every $T > 0$, $(X_t^{x_0})_{0\leq t\leq T} \in \mathcal{E}_T$.
Let us fix $T > 0$ and consider for $n \in \mathbb{N}$ the stopping times

$$T_n = \inf\{t \geq 0, \|X_t^{x_0}\| > n\},$$

For $t \le T$,

$$X^{x_0}_{t\wedge T_n} = x_0 + \int_0^{t\wedge T_n} b(X^{x_0}_s)ds + \int_0^{t\wedge T_n} \sigma(X^{x_0}_s)dB_s.$$

Therefore, by using the inequality

$$\|a+b+c\|^2 \le 3(\|a\|^2 + \|b\|^2 + \|c\|^2),$$

we get

$$\left\|X^{x_0}_{t\wedge T_n}\right\|^2 \le 3\left(\|x_0\|^2 + \left\|\int_0^{t\wedge T_n} b(X^{x_0}_s)ds\right\|^2 + \left\|\int_0^{t\wedge T_n} \sigma(X^{x_0}_s)dB_s\right\|^2\right).$$

Thus,

$$\mathbb{E}\left(\sup_{0\le u\le t\wedge T_n} \|X^{x_0}_u\|^2\right)$$

$$\le 3\left(\|x_0\|^2 + \mathbb{E}\left(\sup_{0\le u\le t\wedge T_n}\left\|\int_0^{u\wedge T_n} b(X^{x_0}_s)ds\right\|^2\right) + \mathbb{E}\left(\sup_{0\le u\le t\wedge T_n}\left\|\int_0^{u\wedge T_n} \sigma(X^{x_0}_s)dB_s\right\|^2\right)\right)$$

By using our assumptions, we first estimate

$$\mathbb{E}\left(\sup_{0\le u\le t\wedge T_n}\left\|\int_0^{u\wedge T_n} b(X^{x_0}_s)ds\right\|^2\right) \le K^2\mathbb{E}\left(\left(\int_0^{t\wedge T_n}(1+\|X^{x_0}_s\|)ds\right)^2\right)$$

By using our assumptions and Doob's inequality, we now estimate

$$\mathbb{E}\left(\sup_{0\le u\le t\wedge T_n}\left\|\int_0^{u\wedge T_n} \sigma(X^{x_0}_s)dB_s\right\|^2\right) \le 4K^2\mathbb{E}\left(\int_0^{t\wedge T_n}(1+\|X_s\|)^2ds\right).$$

Therefore, from the inequality $\|a+b\|^2 \le 2(\|a\|^2+\|b\|^2)$, we get

$$\mathbb{E}\left(\sup_{0\le u\le t\wedge T_n}\|X^{x_0}_u\|^2\right) \le 3\left(\|x_0\|^2 + 2(K^2T+4K^2)\int_0^t\left(1+\mathbb{E}\left(\sup_{0\le u\le s\wedge T_n}\left\|X^{x_0}_{u\wedge T_n}\right\|^2\right)ds\right)\right).$$

We may now apply Gronwall's lemma to the function

$$t \to \mathbb{E}\left(\sup_{0\le u\le t\wedge T_n}\|X^{x_0}_u\|^2\right)$$

and deduce

$$\mathbb{E}\left(\sup_{0\le u\le T\wedge T_n}\|X^{x_0}_u\|^2\right) \le C$$

where C is a constant that does not depend on n. Fatou's lemma implies by passing to the limit when $n \to +\infty$ that

$$\mathbb{E}\left(\sup_{0 \le u \le T} \|X_u^{x_0}\|^2\right) \le C.$$

We conclude, as expected, that

$$(X_t^{x_0})_{0 \le t \le T} \in \mathcal{E}_T.$$

More generally, by using the same arguments we can observe that if a continuous and adapted process satisfies

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s,$$

with $\mathbb{E}(X_0^2) < +\infty$, then $(X_t)_{0 \le t \le T} \in \mathcal{E}_T$.

Step 2: We now show existence and uniqueness of solutions for the equation (3.2.1) on a time interval $[0, T]$ where T is small enough.
Let us consider the application Φ that sends a continuous and adapted process $(X_t)_{0 \le t \le T}$ to the process

$$\Phi(X)_t = x_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s.$$

By using successively the inequalities $(a+b)^2 \le 2(a^2+b^2)$, Schwarz inequality and Doob's inequality, we get

$$\| \Phi(X) - \Phi(Y) \|^2 \le 2(K^2T^2 + 4K^2T) \| X - Y \|^2 .$$

Moreover, arguing the same way as above, we can prove

$$\| \Phi(0) \|^2 \le 3(x_0^2 + K^2T^2 + 4K^2T).$$

Therefore, if T is small enough Φ is a Lipschitz map $\mathcal{E}_T \to \mathcal{E}_T$ whose Lipshitz constant is strictly less than 1. Consequently, it has a unique fixed point. This fixed point is, of course the unique solution of (3.2.1) on the time interval $[0, T]$. Here again, we can observe that the same reasoning applies if x_0 is replaced by a random variable X_0 that satisfies $\mathbb{E}(X_0^2) < +\infty$.

Step 3:
In order to get a solution of (3.2.1) on $[0, +\infty)$, we may apply the previous step to get a solution on intervals $[T_n, T_{n+1}]$, where $T_{n+1} - T_n$ is small enough and $T_n \to +\infty$.

This will provide a solution of (3.2.1) on $[0,+\infty)$. This solution is unique, from the uniqueness on each interval $[T_n, T_{n+1}]$.

Step 4:
We now turn to the proof of the second part of the theorem, namely the Markov property of solutions. The key point, here, is to observe that solutions are actually adapted to the natural filtration of the Brownian motion $(B_t)_{t\geq 0}$. More precisely, there exists on the space of continuous functions $[0,+\infty)\to\mathbb{R}^n$ a predictable functional such that for $t\geq 0$:

$$X_t^{x_0} = F(x_0, (B_u)_{0\leq u\leq t}).$$

Indeed, let us first work on $[0,T]$ where T is small enough. In that case, as seen in the Step 2, the process $(X_t^{x_0})_{0\leq t\leq T}$ is the unique fixed point of the application Φ that was above defined. Alternatively, one can interpret this by observing that $(X_t^{x_0})_{0\leq t\leq T}$ is the limit in $\mathcal{E}_T$ of the sequence of processes $(X_t^n)_{0\leq t\leq T}$ inductively defined by

$$X^{n+1} = \Phi(X^n), \quad X^0 = x_0.$$

It is easily checked that for each X^n there is a predictable functional F_n such that

$$X_t^n = F_n(x_0, (B_u)_{0\leq u\leq t}),$$

which proves the above claim when T is small enough. To get the existence of F for any T, we proceed as in Step 3.
With this hands, we can now prove the Markov property. Let $s\geq 0$. For $t\geq 0$, we have

$$\begin{aligned} X_{s+t}^{x_0} &= X_s + \int_s^{s+t} b(X_u^{x_0})du + \int_s^{s+t}\sigma(X_u^{x_0})dB_u \\ &= X_s + \int_0^t b(X_{u+s}^{x_0})du + \int_0^t \sigma(X_s^{x_0})d(B_{u+s}-B_s). \end{aligned}$$

Consequently, from uniqueness of solutions,

$$X_{s+t}^{x_0} = F(X_s^{x_0}, (B_{u+s}-B_s)_{0\leq u\leq t}).$$

We deduce that for a bounded and Borel function $f:\mathbb{R}^n\to\mathbb{R}$,

$$\mathbb{E}\left(f(X_{s+t}^{x_0}) \mid \mathcal{F}_s\right) = \mathbb{E}\left(f(F(X_s^{x_0}, (B_{u+s}-B_s)_{0\leq u\leq t})) \mid \mathcal{F}_s\right) = \mathbf{P}_t f(X_s^{x_0}),$$

because $(B_{u+s}-B_s)_{0\leq u\leq t}$ is a Brownian motion independent of $\mathcal{F}_s$.

Step 5:
We finally prove the last part of the theorem. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth and compactly supported function. From Itô's formula, we deduce that the process

$$f\left(X_t^x\right) - \int_0^t Lf(X_s^x)ds$$

is a martingale. Therefore, for $t \geq 0$,

$$\mathbb{E}\left(f\left(X_t^x\right)\right) = f(x) + \int_0^t \mathbb{E}\left(Lf(X_s^x)\right)ds.$$

In other words,

$$\mathbf{P}_t f(x) = f(x) + \int_0^t \mathbf{P}_s Lf(x)ds.$$

We can deduce from that

$$\frac{d}{dt}\mathbf{P}_t f = \mathbf{P}_t Lf. \tag{3.2.2}$$

In particular, at $t = 0$

$$\frac{d}{dt}\mathbf{P}_t f = Lf.$$

At that point, we will admit the fact that

$$\mathbf{P}_t f(x) = \mathbb{E}\left(f(X_t^x)\right)$$

is smooth with respect to the variable x (This actually stems from the differentiablity of $x \to X_t^x$). Therefore, on one hand for any $s \geq 0$, at $t = 0$

$$\frac{d}{dt}\mathbf{P}_t\mathbf{P}_s f = L\mathbf{P}_s f.$$

But on the other hand at $t = 0$

$$\frac{d}{dt}\mathbf{P}_t\mathbf{P}_s f = \mathbf{P}_s\frac{d}{dt}\mathbf{P}_t f = \mathbf{P}_s Lf.$$

We conclude that $\mathbf{P}_t Lf = L\mathbf{P}_t f$ and deduce from (3.2.2) that the function

$$\phi(t,x) = \mathbb{E}\left(f(X_t^x)\right) = \mathbf{P}_t f(x)$$

is a bounded solution of the diffusion equation

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

We finally prove uniqueness. So, let $\phi(t,x)$ be a bounded solution of the diffusion equation

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

By using Itô's formula, it is readily checked that for $T \geq 0$, the process $(\phi(T-t, X_t^x))_{0 \leq t \leq T}$ is a bounded local martingale and thus a martingale. It has therefore a constant expectation. This gives

$$\phi(T,x) = \mathbb{E}\left(f(X_T^x)\right).$$

□

Definition 3.2.2 *An equation like (3.2.1) is called a stochastic differential equation.*

3.2.2 The language of vector fields

For geometric purposes, it is often very useful to use Stratonovitch integrals and the language of vector fields in the study of stochastic differential equations.
Let $\mathcal{O} \subset \mathbb{R}^n$ be a non empty open set. A smooth vector field V on $\mathcal{O}$ is a smooth map

$$\begin{array}{rccl} V: & \mathcal{O} & \rightarrow & \mathbb{R}^n \\ & x & \rightarrow & (v_1(x), ..., v_n(x)). \end{array}$$

A vector field V defines a differential operator acting on the smooth functions $f : \mathcal{O} \to \mathbb{R}$ as follows:

$$(Vf)(x) = \sum_{i=1}^n v_i(x) \frac{\partial f}{\partial x_i}.$$

We note that V is a derivation, that is a map on $\mathcal{C}^\infty(\mathcal{O}, \mathbb{R})$, linear over $\mathbb{R}$, satisfying for $f, g \in \mathcal{C}^\infty(\mathcal{O}, \mathbb{R})$,

$$V(fg) = (Vf)g + f(Vg).$$

An interesting result is that, conversely, any derivation on $\mathcal{C}^\infty(\mathcal{O}, \mathbb{R})$ is a vector field.
With these notations, it is readily checked that if $V_0, V_1, \cdots, V_d$ are smooth vector fields on $\mathbb{R}^n$, then the second order differential operator

$$V_0 + \frac{1}{2} \sum_{i=1}^d V_i^2$$

is a diffusion operator.
We have the following translation of Theorem 4.1.1 whose proof is let as a strongly recommended exercise:

Theorem 3.2.3 *Let $(B_t)_{t\geq 0}$ be a d-dimensional Brownian motion. Let us assume that $V_0, V_1, \cdots, V_d$ are smooth vector fields on $\mathbb{R}^n$, and that their derivatives of any order are bounded.*
Then, for every $x_0 \in \mathbb{R}^n$, there exists a unique and adapted process $(X_t^{x_0})_{t\geq 0}$ such that for $t \geq 0$

$$X_t^{x_0} = x_0 + \int_0^t V_0(X_s^{x_0})ds + \sum_{i=1}^d \int_0^t V_i(X_s^{x_0}) \circ dB_s^i. \tag{3.2.3}$$

Moreover, if $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function, the following Itô's formula holds

$$f(X_t^{x_0}) = f(x_0) + \int_0^t V_0 f(X_s^{x_0})ds + \sum_{i=1}^d \int_0^t V_i f(X_s^{x_0}) \circ dB_s^i.$$

Finally, if $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth and compactly supported function, then the function

$$\phi(t,x) = \mathbb{E}\left(f(X_t^x)\right) = \mathbf{P}_t f(x)$$

is the unique bounded solution of the diffusion equation

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

where

$$L = V_0 + \frac{1}{2}\sum_{i=1}^d V_i^2.$$

Chapter 4

Malliavin calculus and elliptic stochastic differential equations

4.1 Hypoellipticity and stochastic differential equations

Consider a stochastic differential equation

$$X_t^{x_0} = x_0 + \int_0^t V_0(X_s^{x_0})ds + \sum_{i=1}^n \int_0^t V_i(X_s^{x_0}) \circ dB_s^i, \; t \geq 0, \tag{4.1.1}$$

where $x_0 \in \mathbb{R}^n$, $V_0, V_1, ..., V_n$ are C^∞ bounded vector fields on $\mathbb{R}^n$ and $(B_t)_{t\geq 0}$ is a n-dimensional standard Brownian motion.
Let us first recall (see Chapter 3) that for every $x_0 \in \mathbb{R}^n$ and every smooth function $f : \mathbb{R}^n \to \mathbb{R}$ which is compactly supported,

$$\mathbb{E}\left(f(X_t^{x_0})\right) = \left(P_t f\right)(x_0), \tag{4.1.2}$$

where P_t is the diffusion semigroup associated to the diffusion operator

$$\mathcal{L} = V_0 + \frac{1}{2}\sum_{i=1}^n V_i^2.$$

We are interested in the existence of smooth densities for the random variables $X_t^{x_0}$, $t > 0$, $x_0 \in \mathbb{R}^n$. According to formula (4.1.2), this question is therefore equivalent to the question of the existence of a smooth transition kernel with respect to the Lebesgue measure for the operators P_t. Let us recall the following definition which comes from functional analysis.

Definition 4.1.1 *A differential operator $\mathcal{G}$ defined on an open set $\mathcal{O} \subset \mathbb{R}^n$ is called hypoelliptic if, whenever u is a distribution on $\mathcal{O}$, u is a smooth function on any open set $\mathcal{O}' \subset \mathcal{O}$ on which $\mathcal{G}u$ is smooth.*

It is possible to show that the existence of a smooth transition kernel with respect to the Lebesgue measure for P_t is equivalent to the hypoellipticity of $\mathcal{L}$. Therefore, our initial question about the existence of smooth densities for the random variables $X_t^{x_0}$, $t > 0$, $x_0 \in \mathbb{R}^n$ is equivalent to the hypoellipticity of $\mathcal{L}$. One of the main results of this chapter is to prove a weak version of the celebrated Hörmander's theorem:

Theorem 4.1.2 *Assume that for every $x_0 \in \mathbb{R}^n$,*

$$\mathbf{span}(V_1(x_0), ..., V_n(x_0)) = \mathbb{R}^n,$$

then the operator $\mathcal{L}$ is hypoelliptic.

Remark 4.1.3 *The strong version of Hörmander's theorem involves a condition on the Lie algebra generated by the vector fields V_0,V_1,...,V_n.*

Remark 4.1.4 *Observe that in the above theorem the sufficient condition for hypoellipticity does not involve the drift term V_0.*

Remark 4.1.5 *If for every $x_0 \in \mathbb{R}^n$,*

$$\mathbf{span}(V_1(x_0), ..., V_n(x_0)) = \mathbb{R}^n,$$

we shall say that the stochastic differential equation 4.1.1 is elliptic.

Remark 4.1.6 *It is also possible to obtain hypoellipticity results for second order differential operators which can not be written as a sum of squares and stochastic differential written in Itô's form. More precisely, If we consider the following stochastic differential equation written in Itô's form:*

$$X_t^{x_0} = x_0 + \int_0^t b(X_s^{x_0})ds + \int_0^t \sigma(X_s^{x_0})dB_s,$$

where σ is a C^∞ bounded field of $n \times n$ matrices, then a sufficient condition for the existence and smoothness of density of $X_t^{x_0}$, $t > 0$ is that the matrix σ is always invertible.

The original proof of Hörmander was rather complicated and has been considerably simplified since by using the theory of pseudo-differential operators. The probabilistic counterpart of the theorem, which is the existence of a smooth density for the random variable $X_t^{x_0}$, $t > 0$, has first been pointed out in Malliavin where, in order the reprove the theorem under weaker assumptions, the author has developed a stochastic calculus of variations which is now known as the Malliavin calculus. This is this approach that we are going to follow.

4.2 Malliavin calculus

In this section, we introduce the basic tools of Malliavin calculus which are used in the proof of Hörmander's theorem. For further details, we refer to the excellent book:

David Nualart: The Malliavin calculus and related topics, Springer 2006.

Let us consider the Wiener space of continuous paths:

$$\mathbb{W}^{\otimes n} = (\mathcal{C}([0,1],\mathbb{R}^n), (\mathcal{B}_t)_{0\le t\le 1}, \mathcal{B}_1, \mathbb{P})$$

where:

1. $\mathcal{C}([0,1],\mathbb{R}^n)$ is the space of continuous functions $[0,1] \to \mathbb{R}^n$;
2. $(B_t)_{t\ge 0}$ is the coordinate process defined by $B_t(f) = f(t)$, $f \in \mathcal{C}([0,1],\mathbb{R}^n)$;
3. $\mathbb{P}$ is the Wiener measure on $[0,1]$, that is the law of a n-dimensional standard Brownian motion indexed by the time interval $[0,1]$;
4. $(\mathcal{B}_t)_{0\le t\le 1}$ is the ($\mathbb{P}$-completed) natural filtration of $(B_t)_{0\le t\le 1}$.

A $\mathcal{B}_1$ measurable real valued random variable F is said to be cylindric if it can be written

$$F = f\left(\int_0^1 h_s^1 dB_s, ..., \int_0^1 h_s^m dB_s\right)$$

where $h^i \in \mathbf{L}^2([0,1],\mathbb{R}^n)$ and $f : \mathbb{R}^m \to \mathbb{R}$ is a C^∞ bounded function. The set of cylindric random variables is denoted $\mathcal{S}$.
The derivative of $F \in \mathcal{S}$ is the $\mathbb{R}^n$ valued stochastic process $(\mathbf{D}_t F)_{0\le t\le 1}$ given by

$$\mathbf{D}_t F = \sum_{i=1}^m h^i(t) \frac{\partial f}{\partial x_i}\left(\int_0^1 h_s^1 dB_s, ..., \int_0^1 h_s^m dB_s.\right).$$

More generally, we can introduce iterated derivatives. If $F \in \mathcal{S}$, we set

$$\mathbf{D}^k_{t_1,...,t_k} F = \mathbf{D}_{t_1} ... \mathbf{D}_{t_k} F.$$

We may consider $\mathbf{D}^k F$ as a square integrable random process indexed by $[0,1]^k$ and valued in $\mathbb{R}^d$. For any $p \ge 1$, the operator $\mathbf{D}^k$ is closable on $\mathcal{S}$. We denote $\mathbb{D}^{k,p}$ the closure of the class of cylindric random variables with respect to the norm

$$\|F\|_{k,p} = \left(\mathbb{E}\left(F^p\right) + \sum_{j=1}^k \mathbb{E}\left(\left\|\mathbf{D}^j F\right\|^p_{\mathbf{L}^2([0,1]^j,\mathbb{R}^n)}\right)\right)^{\frac{1}{p}},$$

and

$$\mathbb{D}^\infty = \bigcap_{p \geq 1} \bigcap_{k \geq 1} \mathbb{D}^{k,p}.$$

We have the following key result which makes Malliavin calculus so useful when one want to study the existence of densities for random variables.

Theorem 4.2.1 *Let $F = (F_1, ..., F_m)$ be a $\mathcal{B}_1$ measurable random vector such that:*

1. *for every $i = 1, ..., m$, $F_i \in \mathbb{D}^\infty$;*
2. *the matrix*

$$\Gamma = \left(\int_0^1 \langle \mathbf{D}_s F^i, \mathbf{D}_s F^j \rangle_{\mathbb{R}^n} ds \right)_{1 \leq i,j \leq m}$$

   is invertible.

Then F has a density with respect to the Lebesgue measure. If moreover, for every $p > 1$,

$$\mathbb{E} \left(\frac{1}{\mid \det \Gamma \mid^p} \right) < +\infty,$$

then this density is smooth.

Remark 4.2.2 *The matrix Γ is often called the Malliavin matrix of the random vector F.*

This theorem relies on the following lemma of Fourier analysis for which we shall use the following notation: If $\phi : \mathbb{R}^n \to \mathbb{R}$ is a smooth function then for $\alpha = (i_1, ..., i_k) \in \{1, ..., n\}^k$, we denote

$$\partial_\alpha \phi = \frac{\partial^k}{\partial x_{i_1} \cdots \partial x_{i_k}} \phi.$$

Lemma 4.2.3 *Let μ be a probability measure on $\mathbb{R}^n$ such that for every smooth and compactly supported function $\phi : \mathbb{R}^n \to \mathbb{R}$,*

$$\left| \int_{\mathbb{R}^n} \partial_\alpha \phi d\mu \right| \leq C_\alpha \|\phi\|_\infty,$$

where $\alpha \in \{1, ..., n\}^k$, $k \geq 1$, $C_\alpha > 0$. Then μ is absolutely continuous with respect to the Lebesgue measure with a smooth density.

Proof. The idea is to show that we may assume that μ is compactly supported and then use Fourier transforms techniques. Let $x_0 \in \mathbb{R}^n$, $R > 0$ and $R' > R$. Let Ψ be a smooth function on $\mathbb{R}^n$ such that $\Psi = 1$ on the ball $\mathbf{B}(x_0, R)$ and $\Psi = 0$ outside the ball $\mathbf{B}(x_0, R')$. Let ν be the measure on $\mathbb{R}^n$ that has a density Ψ with respect to μ. It is easily seen, by induction and integrating by parts that for every smooth and compactly supported function $\phi : \mathbb{R}^n \to \mathbb{R}$,

$$\left| \int_{\mathbb{R}^n} \partial_\alpha \phi d\nu \right| \le C'_\alpha \|\phi\|_\infty,$$

where $\alpha \in \{1, ..., n\}^k$, $k \ge 1$, $C'_\alpha > 0$. Now, if we can prove that under the above assumption ν has a smooth density, then we will able to conclude that ϕ has a smooth density because $x_0 \in \mathbb{R}^n$ and R, R' are arbitrary. Let

$$\hat{\nu}(y) = \int_{\mathbb{R}^n} e^{i\langle y,x\rangle} \nu(dx)$$

be the Fourier transform of the measure μ. The assumption implies that $\hat{\nu}$ is rapidly decreasing (apply the inequality with $\phi(x) = e^{i\langle y,x\rangle}$). We conclude that ν has a smooth density with respect to the Lebesgue measure and that this density f is given by the inverse Fourier transform formula:

$$f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-i\langle y,x\rangle} \hat{\nu}(y) dy.$$

□

We may now turn to the proof of Theorem 4.2.1.

Proof. The proof relies on an integration by parts formula. Let ϕ be a smooth and compactly supported function on $\mathbb{R}^n$. Since $F_i \in \mathbb{D}^\infty$, we easily deduce that $\phi(F) \in \mathbb{D}^\infty$ and that

$$\mathbf{D}\phi(F) = \sum_{i=1}^n \partial_i \phi(F) \mathbf{D} F_i.$$

Therefore

$$\int_0^1 \langle \mathbf{D}_t \phi(F), \mathbf{D}_t F_j \rangle dt = \sum_{i=1}^n \partial_i \phi(F) \int_0^1 \langle \mathbf{D}_t F_i, \mathbf{D}_t F_j \rangle dt.$$

We conclude that

$$\partial_i \phi(F) = \sum_{j=1}^n (\Gamma^{-1})_{i,j} \int_0^1 \langle \mathbf{D}_t \phi(F), \mathbf{D}_t F_j \rangle dt.$$

We now admit, and refer to the book of Nualart for further details, that there exists an operator $\delta : \mathbb{D}^\infty \to \mathbb{D}^\infty$ called the divergence operator (or Skorohod integral) such that for $H \in \mathbb{D}^\infty$,

$$\mathbb{E}\left(H\delta u_t\right) = \mathbb{E}\left(\int_0^1 u_t \mathbf{D}_t H dt\right),$$

whenever u is a measurable process such that $u_t \in \mathbb{D}^\infty$. By using inductively this integration by parts formula in the above equality, we easily deduce that the assumptions of the previous lemma are satisfied. $\square$

4.3 Proof of Hörmander's theorem

Theorem 4.3.1 *Let $(X_t^x)_{t\geq 0}$ denote the solution of the stochastic differential equation*

$$X_t^x = x + \sum_{i=1}^d \int_0^t V_i(X_s^{x_0}) \circ dB_s^i. \tag{4.3.3}$$

Then, for every $i = 1, ..., n$, $X_1^i \in \mathbb{D}^\infty$. Moreover,

$$\mathbf{D}_t^j X_1 = \mathbf{J}_{0\to 1}\mathbf{J}_{0\to t}^{-1} V_j(X_t), \quad j = 1, ..., d, \quad 0 \leq t \leq 1,$$

where $(\mathbf{J}_{0\to t})_{t\geq 0}$ is the first variation process defined by

$$\mathbf{J}_{0\to t} = \frac{\partial X_t^x}{\partial x},$$

and where $\mathbf{D}_t^j X_1^i$ is the j-th component of $\mathbf{D}_t X_1^i$.

Chapter 5

Stochastic Taylor expansions

5.1 Motivation

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^∞ bounded function and denote by $(X_t^{x_0})_{t\geq 0}$ the solution of (3.2.3) with initial condition $x \in \mathbb{R}^n$. First, by Itô's formula, we have

$$f(X_t^x) = f(x) + \sum_{i=0}^{d} \int_0^t (V_i f)(X_s^x) \circ dB_s^i, \ t \geq 0,$$

where we use the notation $B_t^0 = t$. Now, a new application of Itô's formula to $V_i f(X_s^x)$ leads to

$$f(X_t^x) = f(x) + \sum_{i=1}^{d} (V_i f)(x) B_t^i + \sum_{i,j=1}^{d} \int_0^t \int_0^s (V_j V_i f)(X_u^x) \circ dB_u^j \circ dB_s^i.$$

We may iterate this process. For this, let us introduce the following notations:

1. $$\Delta^k[0,t] = \{(t_1, ..., t_k) \in [0,t]^k, t_1 \leq ... \leq t_k\};$$

2. If $I = (i_1, ... i_k) \in \{0, ..., d\}^k$ is a word with length k,

   $$\int_{\Delta^k[0,t]} \circ dB^I = \int_{0 \leq t_1 \leq ... \leq t_k \leq t} \circ dB_{t_1}^{i_1} \circ ... \circ dB_{t_k}^{i_k},$$

   and $n(I)$ is the number of 0's in I.

We can then continue the above procedure and get that for every $N \geq 1$

$$f(X_t^x) = f(x) + \sum_{k=1}^{N} \sum_{I \in \{0,...,d\}^k, k+n(I) \leq N} (V_{i_1} ... V_{i_k} f)(x) \int_{\Delta^k[0,t]} \circ dB^I + \mathbf{R}_N(t, f, x),$$

for some remainder term $\mathbf{R}_N(t,f,x)$ which is easily computed, and shown to satisfy

$$\sup_{x\in\mathbb{R}^n}\sqrt{\mathbb{E}\left(\mathbf{R}_N(t,f,x)^2\right)}\le C_N t^{\frac{N+1}{2}}\sup_{(i_1,...,i_k),k+n(I)=N+1 \text{ or } N+2}\|V_{i_1}\cdots V_{i_k}f\|_\infty.$$

This shows that, in *small times*, the sum

$$f(x)+\sum_{k=1}^{N}\sum_{I\in\{0,...,d\}^k,k+n(I)\le N}(V_{i_1}...V_{i_k}f)(x)\int_{\Delta^k[0,t]}\circ dB^I$$

is a more and more accurate approximation of $f(X_t^x)$ when $N\to+\infty$.

Remark 5.1.1 *For further details on the above discussion, we refer to Ben Arous [4] and Kloeden-Platen [11]. Related discussions for the Taylor expansion of solutions of equations driven by rough paths may be found in Inahama [10] and Friz-Victoir [8].*

5.2 Chen series

Let $\mathbb{R}[[X_0,...,X_d]]$ be the non commutative algebra over $\mathbb{R}$ of the formal series with $d+1$ indeterminates, that is the set of series

$$Y=y_0+\sum_{k=1}^{+\infty}\sum_{I\in\{0,1,...,d\}^k}a_{i_1,...,i_k}X_{i_1}...X_{i_k}.$$

Definition 5.2.1 *If $x:\mathbb{R}_{\ge0}\to\mathbb{R}^d$ is an absolutely continuous path, the Chen series of x is the formal series:*

$$\mathfrak{S}(x)_t=1+\sum_{k=1}^{+\infty}\sum_{I\in\{0,1,...,d\}^k}\left(\int_{0\le t_1\le...\le t_k\le t}dx_{t_1}^{i_1}\cdots dx_{t_k}^{i_k}\right)X_{i_1}\cdots X_{i_k},\quad t\ge0,$$

with the convention $x_t^0=t$.

The exponential of $Y\in\mathbb{R}[[X_0,...,X_d]]$ is defined by

$$\exp(Y)=\sum_{k=0}^{+\infty}\frac{Y^k}{k!},$$

and the logarithm of Y by

$$\ln(Y) = \sum_{k=1}^{+\infty} \frac{(-1)^k}{k}(Y-1)^k.$$

The Chen-Strichartz formula that we will prove in this subsection, is an explicit formula for $\ln \mathfrak{S}(x)_t$.

Remark 5.2.2 *As a preliminary, let us first try to understand a simple case: the commutative case.*
We denote $\mathcal{S}_k$ the group of the permutations of the index set $\{1, ..., k\}$ and if $\sigma \in \mathcal{S}_k$, we denote for a word $I = (i_1, ..., i_k)$, $\sigma \cdot I$ the word $(i_{\sigma(1)}, ..., i_{\sigma(k)})$. If $X_0, X_1, ..., X_d$ were commuting[1]*, we would have*

$$\mathfrak{S}(x)_t = \mathbf{1} + \sum_{k=1}^{+\infty} \sum_{I=(i_1,...,i_k)} X_{i_1}...X_{i_k} \left(\frac{1}{k!} \sum_{\sigma \in \mathcal{S}_k} \int_{\Delta^k[0,t]} dx^{\sigma \cdot I} \right).$$

Since

$$\sum_{\sigma \in \mathcal{S}_k} \int_{\Delta^k[0,t]} dx^{\sigma \cdot I} = x_t^{i_1}...x_t^{i_k},$$

we get,

$$\mathfrak{S}(x)_t = \mathbf{1} + \sum_{k=1}^{+\infty} \frac{1}{k!} \sum_{I=(i_1,...,i_k)} X_{i_1}...X_{i_k} x_t^{i_1}...x_t^{i_k} = \exp\left(\sum_{i=0}^{d} X_i x_t^i \right).$$

We define the Lie bracket between two elements U and V of $\mathbb{R}[[X_0, ..., X_d]]$ by

$$[U, V] = UV - VU.$$

Moreover, if $I = (i_1, ..., i_k) \in \{0, ..., d\}^k$ is a word, we denote by X_I the commutator defined by

$$X_I = [X_{i_1}, [X_{i_2}, ..., [X_{i_{k-1}}, X_{i_k}]...].$$

The universal Chen's theorem asserts that the Chen series of a path is the exponential of a Lie series.

[1] Rigorously, this means that we work in $\mathbb{R}[[X_0, X_1, \cdots, X_d]]/\mathcal{J}$ where $\mathcal{J}$ is the two-sided ideal generated by the relations $X_iX_j - X_jX_i = 0$

Theorem 5.2.3 *[Chen-Strichartz expansion theorem] If $x:\mathbb{R}_{\geq 0}\to\mathbb{R}^d$ is an absolutely continuous path, then*

$$\mathfrak{S}(x)_t=\exp\left(\sum_{k\geq 1}\sum_{I\in\{0,1,\dots,d\}^k}\Lambda_I(x)_t X_I\right),\ t\geq 0,$$

where for $k\geq 1,\ I\in\{0,1,...,d\}^k$:

- *$\mathcal{S}_k$ is the set of the permutations of $\{0,...,k\}$;*
- *If $\sigma\in\mathcal{S}_k$, $e(\sigma)$ is the cardinality of the set*

$$\{j\in\{0,...,k-1\},\sigma(j)>\sigma(j+1)\},$$

- $$\Lambda_I(x)_t=\sum_{\sigma\in\mathcal{S}_k}\frac{(-1)^{e(\sigma)}}{k^2\begin{pmatrix}k-1\\ e(\sigma)\end{pmatrix}}\int_{0\leq t_1\leq\dots\leq t_k\leq t}dx_{t_1}^{\sigma^{-1}(i_1)}\cdots dx_{t_k}^{\sigma^{-1}(i_k)},\quad t\geq 0.$$

Remark 5.2.4 *The first terms in the Chen-Strichartz formula are:*

1. $$\sum_{I=(i_1)}\Lambda_I(x)_t X_I=\sum_{k=0}^{d}x_t^i X_i;$$

2. $$\sum_{I=(i_1,i_2)}\Lambda_I(x)_t X_I=\frac{1}{2}\sum_{0\leq i<j\leq d}[X_i,X_j]\int_0^t x_s^i dx_s^j-x_s^j dx_s^i.$$

We shall give the proof of this theorem in the case where the path x_t is piecewise affine that is

$$dx_t=a_i dt$$

on the interval $[t_i,t_{i+1})$ where $0=t_0\leq t_1\leq\cdots\leq t_N=T$. We may then conclude by a limiting argument. The proof relies on several lemmas.

Lemma 5.2.5 (Chen's relations) *Let x_t be an absolutely continuous path. For any word $(i_1,...,i_n)\in\{0,1,...,d\}^n$ and any $0<s<t$,*

$$\int_{\Delta^n[0,t]}dx^{(i_1,\dots,i_n)}=\sum_{k=0}^{n}\int_{\Delta^k[0,s]}dx^{(i_1,\dots,i_k)}\int_{\Delta^{n-k}[s,t]}dx^{(i_{k+1},\dots,i_n)},$$

where we used the following notations:

1.

$$\int_{\Delta^k[s,t]} dx^{(i_1,\dots,i_k)} = \int_{s\le t_1\le\dots\le t_k\le t} \circ dx^{i_1}_{t_1}\dots dx^{i_k}_{t_k};$$

2. if I is a word with length 0, then $\int_{\Delta^0[0,t]} \circ dx^I = 1$.

Proof. It follows readily by induction on n by noticing that

$$\int_{\Delta^n[0,t]} dx^{(i_1,\dots,i_n)} = \int_0^t \left(\int_{\Delta^{n-1}[0,t_n]} dx^{(i_1,\dots,i_{n-1})} \right) dx^{i_n}_{t_n}.$$

□

The previous lemma implies the following flow property for the Chen series:

Lemma 5.2.6 *Let x_t be an absolutely continuous path. For $0 < s < t$,*

$$\mathfrak{S}(x)_t = \mathfrak{S}(x)_s \left(\mathbf{1} + \sum_{k=1}^{+\infty} \sum_{I=(i_1,\dots i_k)} X_{i_1}\dots X_{i_k} \int_{\Delta^k[s,t]} dx^I \right).$$

Proof. We have, thanks to the previous lemma,

$$\begin{aligned} &\mathfrak{S}(x)_s \left(\mathbf{1} + \sum_{k=1}^{+\infty} \sum_{I} X_{i_1}\dots X_{i_k} \int_{\Delta^k[s,t]} dx^I \right) \\ =&\mathbf{1} + \sum_{k,k'=1}^{+\infty} \sum_{I,I'} X_{i_1}\dots X_{i_k} X_{i'_1}\dots X_{i'_{k'}} \int_{\Delta^k[s,t]} dx^I \int_{\Delta^{k'}[0,s]} dx^{I'} \\ =&\mathbf{1} + \sum_{k=1}^{+\infty} \sum_{I} X_{i_1}\dots X_{i_k} \int_{\Delta^k[0,t]} dx^I \\ =&\mathfrak{S}(x)_t. \end{aligned}$$

□

With this in hands, we may now come back to the proof of the Chen-Strichartz expansion theorem in the case where x_t is piecewise affine. By using inductively the previous proposition, we obtain

$$\mathfrak{S}(x)_T = \prod_{n=0}^{N-1} \left(\mathbf{1} + \sum_{k=1}^{+\infty} \sum_{I=(i_1,\dots i_k)} X_{i_1}\dots X_{i_k} \int_{\Delta^k[t_n,t_{n+1}]} dx^I \right)$$

Since, on $[t_n, t_{n+1})$,

$$dx_t = a_n dt,$$

we have

$$\int_{\Delta^k[t_n,t_{n+1}]} dx^I = a_n^{i_1}\cdots a_n^{i_k}\int_{\Delta^k[t_n,t_{n+1}]} dt_{i_1}\cdots dt_{i_k} = a_n^{i_1}\cdots a_n^{i_k}\frac{(t_{n+1}-t_n)^k}{k!}.$$

Therefore

$$\begin{aligned}\mathfrak{S}(x)_T &= \prod_{n=0}^{N-1}\left(\mathbf{1}+\sum_{k=1}^{+\infty}\sum_{I=(i_1,\ldots i_k)} X_{i_1}\ldots X_{i_k} a_n^{i_1}\cdots a_n^{i_k}\frac{(t_{n+1}-t_n)^k}{k!}\right)\\ &= \prod_{n=0}^{N-1}\exp\left((t_{n+1}-t_n)\sum_{i=0}^{d} a_n^i X_i\right)\end{aligned}$$

We now use the Baker-Campbell-Hausdorff-Dynkin formula:

Proposition 5.2.7 (**Baker-Campbell-Hausdorff-Dynkin formula**) *If* $y_1,\cdots,y_N\in\mathbb{R}^{d+1}$ *then,*

$$\prod_{n=1}^{N}\exp\left(\sum_{i=0}^{d} y_n^i X_i\right) = \exp\left(\sum_{k\geq 1}\sum_{I\in\{0,1,\ldots,d\}^k}\beta_I(y_1,\cdots,y_N)X_I\right),$$

where for $k\geq 1,\ I\in\{0,1,...,d\}^k$ *:*

$$\beta_I(y_1,\cdots,y_N) = \sum_{\sigma\in\mathcal{S}_k}\sum_{0=j_0\leq j_1\leq\cdots\leq j_{N-1}\leq k}\frac{(-1)^{e(\sigma)}}{j_1!\cdots j_{N-1}!k^2\begin{pmatrix}k-1\\ e(\sigma)\end{pmatrix}}\prod_{\nu=1}^{N} y_\nu^{\sigma^{-1}(i_{j_{\nu-1}+1})}\cdots y_\nu^{\sigma^{-1}(i_{j_\nu})}.$$

We get therefore:

$$\mathfrak{S}(x)_T = \exp\left(\sum_{k\geq 1}\sum_{I\in\{0,1,\ldots,d\}^k}\beta_I(t_1a_0,\cdots,(t_N-t_{N-1})a_{N-1})X_I\right).$$

It is finally an easy exercise to check, by using the Chen relations, that:

$$\beta_I(t_1a_0,\cdots,(t_N-t_{N-1})a_{N-1}) = \sum_{\sigma\in\mathcal{S}_k}\frac{(-1)^{e(\sigma)}}{k^2\begin{pmatrix}k-1\\ e(\sigma)\end{pmatrix}}\int_{0\leq t_1\leq\ldots\leq t_k\leq t} dx_{t_1}^{\sigma^{-1}(i_1)}\cdots dx_{t_k}^{\sigma^{-1}(i_k)},\quad t\geq 0.$$

Remark 5.2.8 *The seminal result of Chen [6] asserted that* $\ln\mathfrak{S}(x)_T$ *was a Lie series. The coefficients of this expansion were computed by Ben Arous [4] , Castell [5], and Strichartz [18].*

5.3 Brownian Chen series

Chen's theorem can actually be extended to Brownian paths (see Baudoin [3], Ben Arous [4], Castell [5], Fliess [7]) and even to rough paths (see Lyons [12], Friz-Victoir [9]).

Definition 5.3.1 *If $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion, the Chen series of B is the formal series:*

$$\mathfrak{S}(B)_t = 1 + \sum_{k=1}^{+\infty} \sum_{I \in \{0,1,\ldots,d\}^k} \left(\int_{0 \leq t_1 \leq \ldots \leq t_k \leq t} \circ dB_{t_1}^{i_1} \cdots \circ dB_{t_k}^{i_k} \right) X_{i_1} \cdots X_{i_k}, \quad t \geq 0,$$

with the convention $B_t^0 = t$, and $\circ$ denotes Stratonovitch integral.

Theorem 5.3.2 *If $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion, then*

$$\mathfrak{S}(B)_t = \exp\left(\sum_{k \geq 1} \sum_{I \in \{0,1,\ldots,d\}^k} \Lambda_I(B)_t X_I \right), \ t \geq 0,$$

where for $k \geq 1$, $I \in \{0,1,...,d\}^k$,

$$\Lambda_I(B)_t = \sum_{\sigma \in \mathcal{S}_k} \frac{(-1)^{e(\sigma)}}{k^2 \begin{pmatrix} k-1 \\ e(\sigma) \end{pmatrix}} \int_{0 \leq t_1 \leq \ldots \leq t_k \leq t} \circ dB_{t_1}^{\sigma^{-1}(i_1)} \cdots \circ dB_{t_k}^{\sigma^{-1}(i_k)}, \quad t \geq 0.$$

If

$$Y = y_0 + \sum_{k=1}^{+\infty} \sum_{I \in \{0,1,\ldots,d\}^k} a_{i_1,\ldots,i_k} X_{i_1} \ldots X_{i_k}.$$

is a random series, that is if the coefficients are real random variables defined on a probability space, we will denote

$$\mathbb{E}(Y) = \mathbb{E}(y_0) + \sum_{k=1}^{+\infty} \sum_{I \in \{0,1,\ldots,d\}^k} \mathbb{E}(a_{i_1,\ldots,i_k}) X_{i_1} \ldots X_{i_k}.$$

as soon as the coefficients of Y are integrable, where $\mathbb{E}$ stands for the expectation. The following theorem gives the expectation (see Baudoin [3], Lyons-Victoir [13]) of the Brownian Chen series:

Theorem 5.3.3 *For $t \geq 0$,*

$$\mathbb{E}\left(\mathfrak{S}(B)_t\right) = \exp\left(t\left(X_0 + \frac{1}{2}\sum_{i=1}^{d} X_i^2\right)\right).$$

Proof. An easy computation shows that if $\mathcal{I}_n$ is the set of words with length n obtained by all the possible concatenations of the words

$$\{0\}, \{(i,i)\}, \quad i \in \{1, ..., d\},$$

1. If $I \notin \mathcal{I}_n$ then
$$\mathbb{E}\left(\int_{\Delta^n[0,t]} \circ dB^I\right) = 0;$$

2. If $I \in \mathcal{I}_n$ then
$$\mathbb{E}\left(\int_{\Delta^n[0,t]} \circ dB^I\right) = \frac{t^{\frac{n+n(I)}{2}}}{2^{\frac{n-n(I)}{2}}\left(\frac{n+n(I)}{2}\right)!},$$
where $n(I)$ is the number of 0 in I (observe that since $I \in \mathcal{I}_n$, n and $n(I)$ necessarily have the same parity).

Therefore,

$$\mathbb{E}\left(\mathfrak{S}(B)_t\right) = 1 + \sum_{k=1}^{+\infty} \sum_{I \in \mathcal{I}_k} \frac{t^{\frac{k+n(I)}{2}}}{2^{\frac{k-n(I)}{2}}\left(\frac{k+n(I)}{2}\right)!} X_{i_1} ... X_{i_k}$$

$\square$

Chapter 6

Approximation of solutions of stochastic differential equations

6.1 Exponential of a vector field

Let $\mathcal{O} \subset \mathbb{R}^n$ be a non empty open set and V be a smooth vector field on $\mathcal{O}$. It is a basic result in the theory of ordinary differential equations that if $K \subset \mathcal{O}$ is compact, there exist $\varepsilon > 0$ and a smooth mapping

$$\Phi : (-\varepsilon, \varepsilon) \times K \to \mathcal{O},$$

such that for $x \in K$ and $-\varepsilon < t < \varepsilon$,

$$\frac{\partial \Phi}{\partial t}(t,x) = X(\Phi(t,x)),\ \Phi(0,x) = x.$$

Furthermore, if $y : (-\eta, \eta) \to \mathbb{R}^n$ is a C^1 path such that for $-\eta < t < \eta$, $y'(t) = X(y(t))$, then $y(t) = \Phi(t, y(0))$ for $-\min(\eta, \varepsilon) < t < \min(\eta, \varepsilon)$. From this characterization of Φ it is easily seen that for $x \in K$ and $t_1, t_2 \in \mathbb{R}$ such that $\mid t_1 \mid + \mid t_2 \mid < \varepsilon$,

$$\Phi(t_1, \Phi(t_2, x)) = \Phi(t_1 + t_2, x).$$

Because of this last property, the solution mapping $t \to \Phi(t,x)$ is called the exponential mapping, and we denote $\Phi(t,x) = e^{tV}(x)$. It always exists if $\mid t \mid$ is sufficiently small. If e^{tV} can be defined for any $t \in \mathbb{R}$, then the vector field is said to be complete. For instance if $\mathcal{O} = \mathbb{R}^n$ and if V is C^∞-bounded then the vector field V is complete.

6.2 Lie bracket of vector fields

We have already mentioned that a vector field V may be seen derivation, that is a map on $\mathcal{C}^\infty(\mathcal{O},\mathbb{R})$, linear over $\mathbb{R}$, satisfying for $f,g \in \mathcal{C}^\infty(\mathcal{O},\mathbb{R})$,

$$V(fg) = (Vf)g + f(Vg).$$

Also, conversely, any derivation on $\mathcal{C}^\infty(\mathcal{O},\mathbb{R})$ is a vector field. If V' is another smooth vector field on $\mathcal{O}$, then it is easily seen that the operator $VV' - V'V$ is a derivation. It therefore defines a smooth vector field on $\mathcal{O}$ which is called the Lie bracket of V and V' and denoted $[V,V']$. A straightforward computation shows that for $x \in \mathcal{O}$,

$$[V,V'](x) = \sum_{i=1}^n \left(\sum_{j=1}^n v_j(x) \frac{\partial v'_i}{\partial x_j}(x) - v'_j(x) \frac{\partial v_i}{\partial x_j}(x) \right) \frac{\partial}{\partial x_i}.$$

Observe that the Lie bracket satisfies obviously $[V,V'] = -[V',V]$ and the so-called Jacobi identity, that is:

$$[V,[V',V'']] + [V',[V'',V]] + [V'',[V,V']] = 0.$$

6.3 Castell's approximation theorem

Combining the stochastic Taylor expansion with the Chen-Strichartz formula leads to the following result due to Castell [5]:

Theorem 6.3.1 (Castell approximation theorem) *Let $(B_t)_{t\geq 0}$ be a d-dimensional Brownian motion. Let us assume that $V_0, V_1, \cdots, V_d$ are C^∞ bounded vector fields on $\mathbb{R}^n$, Then, for the solution $(X_t^{x_0})_{t\geq 0}$ of the following stochastic differential equation*

$$X_t^{x_0} = x_0 + \int_0^t V_0(X_s^{x_0})ds + \sum_{i=1}^d \int_0^t V_i(X_s^{x_0}) \circ dB_s^i, \qquad (6.3.1)$$

we have for every $N \geq 1$,

$$X_t^{x_0} = \exp\left(\sum_{k=1}^N \sum_{I \in \{0,1,\dots,d\}^k, k+n(I)\leq N} \Lambda_I(B)_t X_I \right)(x_0) + t^{\frac{N+1}{2}} \mathbf{R}_N(t),$$

where $n(I)$ denotes the number of 0's in the word I and where the remainder term $\mathbf{R}_N(t)$ is bounded in probability when $t \to 0$. More precisely, $\exists\ \alpha, c > 0$ such that $\forall A > c$,

$$\lim_{t\to 0} \mathbb{P}\left(\sup_{0\leq s\leq t} s^{\frac{N+1}{2}} \mid \mathbf{R}_N(s) \mid \geq A t^{\frac{N+1}{2}} \right) \leq \exp\left(-\frac{A^\alpha}{c} \right).$$

6.4 Approximation of solutions of diffusion equations

We consider the following linear partial differential equation

$$\frac{\partial \Phi}{\partial t} = \mathcal{L}\Phi, \quad \Phi(0,x) = f(x), \tag{6.4.2}$$

where $\mathcal{L}$ is an operator on $\mathcal{E}$ that can be written

$$\mathcal{L} = V_0 + \frac{1}{2}\sum_{i=1}^{d} V_i^2,$$

the V_i's being smooth and compactly supported[1] vector fields on $\mathbb{R}^n$. It is known that the solution of (6.4.2) can be written

$$\Phi(t,x) = (e^{t\mathcal{L}} f)(x) = \mathbf{P}_t f(x).$$

If $I \in \{0,1,...,d\}^k$ is a word, we denote as before

$$V_I = [V_{i_1}, [V_{i_2}, ..., [V_{i_{k-1}}, V_{i_k}]...].$$

and

$$d(I) = k + n(I),$$

where $n(I)$ is the number of 0's in the word I.
For $N \geq 1$, let us consider

$$\mathbf{P}_t^N = \mathbb{E}\left(\exp\left(\sum_{I, d(I) \leq N} \Lambda_I(B)_t V_I\right)\right).$$

For instance

$$\mathbf{P}_t^1 = \mathbb{E}\left(\exp\left(\sum_{i=1}^{d} B_t^i V_i\right)\right),$$

and

$$\mathbf{P}_t^2 = \mathbb{E}\left(\exp\left(\sum_{i=0}^{d} B_t^i \nabla_i + \frac{1}{2}\sum_{1\leq i<j\leq d}\int_0^t B_s^i dB_s^j - B_s^j dB_s^i [V_i, V_j]\right)\right).$$

[1]This assumption will not be restrictive for us because we shall eventually be interested in local results

The meaning of this last notation is the following. If f is a smooth and bounded function, then $(\mathbf{P}_t^N f)(x) = \mathbb{E}(\Psi(1,x))$, where $\Psi(\tau,x)$ is the solution of the first order partial differential equation with random coefficients:

$$\frac{\partial \Psi}{\partial \tau}(\tau,x) = \sum_{I,d(I)\le N} \Lambda_I(B)_t (V_I \Psi)(\tau,x), \quad \Psi(0,x)=f(x).$$

Finally, let us consider the following family of norms: If f is a C^∞ bounded function, then for $k \ge 0$,

$$\parallel f \parallel_k = \sup_{0\le l\le k} \sup_{0\le i_1,\cdots,i_l \le d} \sup_{x\in\mathbb{R}^n} \parallel V_{i_1}\cdots V_{i_l} f(x) \parallel .$$

Theorem 6.4.1 *Let $N\ge 1$ and $k \ge 0$. If f is a C^∞ bounded function, then*

$$\parallel \mathbf{P}_t f - \mathbf{P}_t^N f \parallel_k = O\left(t^{\frac{N+1}{2}}\right), \quad t\to 0.$$

Proof. First, by using the scaling property of Brownian motion and expanding out the exponential with Taylor formula we obtain

$$\exp\left(\sum_{I,d(I)\le N} \Lambda_I(B)_t V_I\right) f = \left(\sum_{k=0}^N \frac{1}{k!}\left(\sum_{I,d(I)\le N} \Lambda_I(B)_t V_I\right)^k\right) f + t^{\frac{N+1}{2}} \mathbf{R}_N^1(t),$$

where the remainder term $\mathbf{R}_N^1(t)$ is such that $\mathbb{E}\left(\parallel \mathbf{R}_N^1(t) \parallel_k\right)$ is bounded when $t\to 0$. We now observe that, due to Theorem 5.3.2, the rearrangement of terms in the previous formula gives

$$\left(\sum_{k=0}^N \frac{1}{k!}\left(\sum_{I,d(I)\le N} \Lambda_I(B)_t V_I\right)^k\right) f = f + \sum_{I,d(I)\le N} \int_{\Delta^{|I|}[0,t]} \circ dB^I V_{i_1}...V_{i_{|I|}} f + t^{\frac{N+1}{2}} \mathbf{R}_N^2(t),$$

where $\mathbb{E}\left(\parallel \mathbf{R}_N^2(t) \parallel_k\right)$ is bounded when $t\to 0$. Therefore

$$\exp\left(\sum_{I,d(I)\le N} \Lambda_I(B)_t V_I\right) f = f + \sum_{I,d(I)\le N} \int_{\Delta^{|I|}[0,t]} \circ dB^I V_{i_1}...V_{i_{|I|}} f + t^{\frac{N+1}{2}} \mathbf{R}_N^3(t),$$

and

$$\mathbf{P}_t^N f = f + \sum_{I,d(I)\le N} \mathbb{E}\left(\int_{\Delta^{|I|}[0,t]} \circ dB^I\right) V_{i_1}...V_{i_{|I|}} f + t^{\frac{N+1}{2}} \mathbb{E}\left(\mathbf{R}_N^3(t)\right),$$

where $\mathbb{E}(\| \mathbf{R}_N^3(t) \|_k)$ is bounded when $t \to 0$. We now recall (see the proof of Theorem 5.3.3) that if $\mathcal{I}_n$ is the set of words with length n obtained by all the possible concatenations of the words

$$\{0\}, \{(i,i)\}, \quad i \in \{1, ..., d\},$$

1. If $I \notin \mathcal{I}_n$ then
$$\mathbb{E}\left(\int_{\Delta^n[0,t]} \circ dB^I \right) = 0;$$

2. If $I \in \mathcal{I}_n$ then
$$\mathbb{E}\left(\int_{\Delta^n[0,t]} \circ dB^I \right) = \frac{t^{\frac{n+n(I)}{2}}}{2^{\frac{n-n(I)}{2}} \left(\frac{n+n(I)}{2} \right)!},$$

   where $n(I)$ is the number of 0 in I (observe that since $I \in \mathcal{I}_n$, n and $n(I)$ necessarily have the same parity).

We conclude therefore

$$\| \mathbf{P}_t^N f - \sum_{k \le \frac{N+1}{2}} \frac{t^k}{k!} \mathcal{L}^k f \|_k = O\left(t^{\frac{N+1}{2}} \right).$$

Since it is known that

$$\| \mathbf{P}_t f - \sum_{k \le \frac{N+1}{2}} \frac{t^k}{k!} \mathcal{L}^k f \|_k = O\left(t^{\frac{N+1}{2}} \right),$$

the theorem is proved. $\square$

6.5 Approximation on the diagonal of elliptic heat kernels

In this subsection, we keep the framework of the previous subsection and we assume furthermore that at given point $x_0 \in \mathbb{R}^n$ the vector fields $V_1(x_0), ..., V_d(x_0)$ form a basis of $\mathbb{R}^n$ which of course implies $n = d$. In that case, it is known that the random variable $X_t^{x_0}$ admits a smooth density $p_t(x_0, \cdot)$ with respect to the Lebesgue measure of $\mathbb{R}^n$. In other words,

$$\mathbb{P}(X_t^{x_0} \in dy) = p_t(x_0, y) dy,$$

for some smooth function $p(x_0, \cdot) : (0, +\infty) \times \mathbb{R}^n \to \mathbb{R}_{\ge 0}$.

We are interested in $p_t(x_0, x_0)$ in small times. It is known (see for instance Azencott, [1]), that the following asymptotic expansion holds when $t \to 0$,

$$p_t(x_0, x_0) = \frac{1}{t^{\frac{d}{2}}} \left(\sum_{k=0}^{N} a_k(x_0) t^k \right) + O\left(t^{N+1-\frac{d}{2}} \right), \quad N \geq 0,$$

for some constants $a_0(x_0), \cdots, a_N(x_0)$. The following proposition provides an effective way to compute these constants.

Proposition 6.5.1 *For $N \geq 1$, when $t \to 0$,*

$$p_t(x_0, x_0) = d_t^N(x_0) + O\left(t^{\frac{N+1-d}{2}} \right),$$

where $d_t^N(x_0)$ is the density at 0 of the random variable $\sum_{I, d(I) \leq N} \Lambda_I(B)_t V_I(x_0)$

For instance, by applying the previous proposition with $N = 1$, we get

$$a_0(x) = \frac{1}{(2\pi)^{\frac{d}{2}}} \frac{1}{|\det(V_1(x_0), \cdots, V_d(x_0))|}$$

Chapter 7

An introduction to the theory of log-Sobolev inequalities

This last Chapter is only an introduction to the theory of functional inequalities for diffusion operators. For a detailed account on this we refer to the following very clear and complete survey written by Michel Ledoux: An Introduction to the geometry of Markov processes

7.1 Framework

We use here the framework of Chapter 2 that we remind below.
In the sequel, we consider a diffusion operator L as well as a Borel measure μ on $\mathbb{R}^n$. We will assume that if $f \in \mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$, $1 \le p \le +\infty$, then there is a unique real valued function $\phi(t,x)$ that is smooth on $(0,+\infty) \times \mathbb{R}^n$, continuous on $[0,+\infty) \times \mathbb{R}^n$ and that satisfies:

$$\frac{\partial \phi}{\partial t}(t,x) = L\phi(t,x), \quad \phi(0,x) = f(x). \tag{7.1.1}$$

As usual, if $f \in \mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$, $1 \le p \le +\infty$, then , for $t \ge 0$, we set

$$(\mathbf{P}_t f)(x) = \phi(t,x),$$

where ϕ is the unique solution of equation (7.1.1). We shall assume that $\mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$ is left stable by $\mathbf{P}_t$ for any $t \ge 0$ and $1 \le p \le +\infty$, and that for $f \in \mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$, $g \in \mathbf{L}^q_\mu(\mathbb{R}^n, \mathbb{R})$, with $\frac{1}{p} + \frac{1}{q} = 1$,

$$\int_{\mathbb{R}^n} f \mathbf{P}_t g d\mu = \int_{\mathbb{R}^n} g \mathbf{P}_t f d\mu. \tag{7.1.2}$$

As in Chapter 2, we shall also assume that the Schwartz space $\mathcal{S}$ of smooth rapidly decreasing functions on $\mathbb{R}^n$ is a dense subset of $\mathbf{L}^p_\mu(\mathbb{R}^n, \mathbb{R})$, $1 \leq p < +\infty$ and is left globally stable by L and by $\mathbf{P}_t$ for any $t \geq 0$.

7.2 The curvature inequality

We associate with the diffusion operator L two canonical differential bilinear forms: For $f, g \in \mathcal{S}$, we set

$$\Gamma(f,g) = \frac{1}{2}(L(fg) - fLg - gLf),$$

and

$$\Gamma_2(f,g) = \frac{1}{2}\big[L\Gamma(f,g) - \Gamma(f, Lg) - \Gamma(g, Lf)\big].$$

Example 7.2.1 *If $L = \Delta - \langle \nabla U, \nabla \cdot \rangle$, where U is a smooth function, one can check*

$$\Gamma(f,g) = \langle \nabla f, \nabla g \rangle,$$

$$\Gamma_2(f,f) = \|\mathbf{Hess} f\|_2^2 + \langle \mathbf{Hess} U(\nabla f), \nabla f \rangle.$$

It is easily seen that since L is a diffusion operator, the bilinear form Γ is positive, more precisely a straightforward computation shows that if

$$L = \sum_{i,j=1}^{n} \sigma_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(x) \frac{\partial}{\partial x_i},$$

then

$$\Gamma(f,f) = \sum_{i,j=1}^{n} \sigma_{ij}(x) \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \geq 0.$$

The bilinear form Γ_2 is, in general, not positive but we set the following definition:

Definition 7.2.2 (Curvature inequality) *We say that the operator L satisfies the curvature inequality with parameter $\rho \in \mathbb{R}$, if for every $f \in \mathcal{S}$,*

$$\Gamma_2(f,f) \geq \rho \Gamma(f,f).$$

Remark 7.2.3 *The name curvature inequality stems from the fact that the Laplace-Beltrami operator of a Riemannian manifold satisfies the curavture inequality with parameter ρ, if and only if the Ricci curvature (as a $(0,2)$ tensor) is bounded from below by ρ.*

Example 7.2.4 *The operator $L = \Delta - \langle \nabla U, \nabla \cdot \rangle$ satisfies the curvature inequality with parameter ρ if and only if*

$$\mathbf{Hess} U \geq \rho.$$

The curvature inequality is equivalent to the following inequality:

Proposition 7.2.5 *The curvature inequality with parameter $\rho \in \mathbb{R}$ is satisfied if and only if for every $f \in \mathcal{S}$, and every $t \geq 0$,*

$$\sqrt{\Gamma(\mathbf{P}_t f, \mathbf{P}_t f)} \leq e^{-\rho t} \mathbf{P}_t \sqrt{\Gamma(f, f)}.$$

Proof.
We refer to Lemma 1.3. in Ledoux.

□

Remark 7.2.6 *By Cauchy-Schwarz inequality, the above inequality implies*

$$\Gamma(\mathbf{P}_t f, \mathbf{P}_t f) \leq e^{-2\rho t} \mathbf{P}_t \Gamma(f, f).$$

Exercise 7.2.7 *Let us assume that the operator L satisfies the curvature inequality with parameter $\rho > 0$. Show that*

$$\mu(\mathbb{R}^n) < +\infty.$$

In this case, we will assume that $\mu(\mathbb{R}^n) = 1$.

7.3 Log-Sobolev inequality

A first basic consequence of the curvature inequality is the so-called Poincaré inequality:

Proposition 7.3.1 *Let us assume that the operator L satisfies the curvature inequality with parameter $\rho > 0$. Then the following Poincaré inequality is satisfied:*

$$\int_{\mathbb{R}^n} f^2 d\mu \leq \left(\int_{\mathbb{R}^n} f d\mu \right)^2 + \frac{1}{\rho} \int_{\mathbb{R}^n} \Gamma(f, f) d\mu,$$

where f is a function such that $f, Lf \in \mathbf{L}^2_\mu(\mathbb{R}^n, \mathbb{R})$.

Proof. Let $f \in \mathcal{S}$. We have by assumption

$$\Gamma_2(f,f) \geq \rho\Gamma(f,f).$$

Therefore

$$\int_{\mathbb{R}^n} \Gamma_2(f,f)d\mu \geq \rho \int_{\mathbb{R}^n} \Gamma(f,f)d\mu.$$

But

$$\int_{\mathbb{R}^n} \Gamma_2(f,f)d\mu = -\int_{\mathbb{R}^n} \Gamma(f,Lf)d\mu = \int_{\mathbb{R}^n} (Lf)^2 d\mu.$$

Therefore

$$\int_{\mathbb{R}^n} (Lf)^2 d\mu \geq \rho \int_{\mathbb{R}^n} \Gamma(f,f)d\mu = -\rho \int_{\mathbb{R}^n} fLf d\mu.$$

By density, this last inequality is seen to hold for every function f such that $f, Lf \in \mathbf{L}^2_\mu(\mathbb{R}^n,\mathbb{R})$. It means that the L^2 spectrum of $-L$ lies in $\{0\} \cup [\rho, +\infty)$, which directly implies the Poincaré inequality. □

As a consequence of the previous Poincaré inequality, we deduce the following convergence to equilibrium for $\mathbf{P}_t$,

Proposition 7.3.2 *Let us assume that the operator L satisfies the curvature inequality with parameter $\rho > 0$. Let $f, Lf \in \mathbf{L}^2_\mu(\mathbb{R}^n,\mathbb{R})$, then when $t \to +\infty$, in $\mathbf{L}^2_\mu(\mathbb{R}^n,\mathbb{R})$,*

$$\mathbf{P}_t f \to \int_{\mathbb{R}^n} f d\mu.$$

More precisely,

$$\int_{\mathbb{R}^n} \left(\mathbf{P}_t f - \int_{\mathbb{R}^n} f d\mu \right)^2 d\mu \leq e^{-2\rho t} \int_{\mathbb{R}^n} \Gamma(f,f)d\mu.$$

Proof. First, observe that we may assume $\int_{\mathbb{R}^n} f d\mu = 0$. Then, the idea is to consider the functional

$$\Psi(t) = \int_{\mathbb{R}^n} (\mathbf{P}_t f)^2 d\mu.$$

We have

$$\Psi'(t) = 2\int_{\mathbb{R}^n} \mathbf{P}_t f L \mathbf{P}_t f d\mu = -2\int_{\mathbb{R}^n} \Gamma(\mathbf{P}_t f, \mathbf{P}_t f)d\mu.$$

If we now use Poincaré inequality with $\mathbf{P}_t f$, we end up with the following differential inequality

$$\Psi'(t) \leq -2\Psi(t).$$

This implies

$$\Psi(t) \leq e^{-2\rho t}\Psi(0),$$

which is the required inequality. $\square$

We now turn to the so-called log-Sobolev inequality

Theorem 7.3.3 *Let us assume that the operator L satisfies the curvature inequality with parameter $\rho > 0$. Then, for $f \in \mathcal{S}$,*

$$\int_{\mathbb{R}^n} f^2 \ln f^2 d\mu \leq \int_{\mathbb{R}^n} f^2 d\mu \ln \left(\int_{\mathbb{R}^n} f^2 d\mu \right) + \frac{2}{\rho} \int_{\mathbb{R}^n} \Gamma(f,f) d\mu.$$

Proof. By considering $\sqrt{f}$ instead of f, it is enough to show that if f is positive,

$$\int_{\mathbb{R}^n} f \ln f d\mu \leq \int_{\mathbb{R}^n} f d\mu \ln \left(\int_{\mathbb{R}^n} f d\mu \right) + \frac{1}{2\rho} \int_{\mathbb{R}^n} \frac{\Gamma(f,f)}{f} d\mu.$$

We now have

$$\begin{aligned}
\int_{\mathbb{R}^n} f \ln f d\mu - \int_{\mathbb{R}^n} f d\mu \ln \left(\int_{\mathbb{R}^n} f d\mu \right) &= - \int_0^{+\infty} \frac{d}{dt} \int_{\mathbb{R}^n} \mathbf{P}_t f \ln \mathbf{P}_t f d\mu dt \\
&= - \int_0^{+\infty} \int_{\mathbb{R}^n} L \mathbf{P}_t f \ln \mathbf{P}_t f d\mu dt \\
&= \int_0^{+\infty} \int_{\mathbb{R}^n} \Gamma(\mathbf{P}_t f, \ln \mathbf{P}_t f) d\mu dt \\
&= \int_0^{+\infty} \int_{\mathbb{R}^n} \frac{\Gamma(\mathbf{P}_t f, \mathbf{P}_t f)}{\mathbf{P}_t f} d\mu dt
\end{aligned}$$

Now, we know that

$$\Gamma(\mathbf{P}_t f, \mathbf{P}_t f) \leq e^{-2\rho t} \left(\mathbf{P}_t \sqrt{\Gamma(f,f)} \right)^2.$$

And, from Cauchy-Schwarz inequality,

$$(\mathbf{P}_t \sqrt{\Gamma(f,f)})^2 \leq \mathbf{P}_t \frac{\Gamma(f,f)}{f} \mathbf{P}_t f.$$

Therefore,

$$\int_{\mathbb{R}^n} f \ln f d\mu - \int_{\mathbb{R}^n} f d\mu \ln \left(\int_{\mathbb{R}^n} f d\mu \right) \leq \int_0^{+\infty} e^{-2\rho t} dt \int_{\mathbb{R}^n} \frac{\Gamma(f,f)}{f} d\mu,$$

which is the required inequality. $\square$

Bibliography

[1] R. Azencott: Densité des diffusions en temps petit: développements asymptotiques (part I), Sem. Prob. , tome 18, (1984), p.402-498.

[2] R. Azencott: Formule de Taylor stochastique et développements asymptotiques d'intégrales de Feynman. In Azema, Yor (Eds.), Séminaire de probabilités **XVI**, LNM **921**, 237-284, Springer, (1982).

[3] F. Baudoin: An Introduction to the Geometry of Stochastic Flows, Imperial College Press, (2004).

[4] G. Ben Arous: Flots et séries de Taylor stochastiques, Journal of Probability Theory and Related Fields, **81**, 29-77, (1989).

[5] F. Castell: Asymptotic expansion of stochastic flows, Prob. Rel. Fields, **96**, 225-239, (1993).

[6] K.T. Chen: Integration of paths, Geometric invariants and a Generalized Baker-Hausdorff formula, Annals of Mathematics, **65**, n1, (1957).

[7] M. Fliess, D. Normand-Cyrot: Algèbres de Lie nilpotentes, formule de Baker-Campbell-Hausdorff et intégrales itérées de K.T. Chen, in Séminaire de Probabilités, LNM **920**, Springer-Verlag, (1982).

[8] P. Friz, N. Victoir: Euler Estimates for Rough Differential Equations, Journal of Differential Equations, Vol 244, Issue 2 (2008) 388-412

[9] P. Friz, N. Victoir: Multidimensional stochastic processes as rough paths. Theory and Applications. Cambridge Studies in Advanced Mathematics (CUP, 2009)

[10] Y. Inahama: A stochastic Taylor-like expansion in the rough path theory, Preprint from Tokyo Institute of Technology, 2007

[11] Kloeden, P. E. , Platen, E. : Numerical solution of stochastic differential equations. Applications of Mathematics, vol. 23. Springer, (1992).

[12] T. Lyons: Differential Equations Driven by Rough Signals.Revista Mathemàtica Iberio Americana, Vol **14**, No **2**, 215 - 310, (1998).

[13] T. Lyons, N. Victoir: Cubature on Wiener space, Proc. R. Soc. Lond. A (2004) 460, 169-198.

[14] P. Malliavin: Stochastic calculus of variations and hypoelliptic operators. In: *Proc. Inter, Symp. on Stoch. Diff. Equations, Kyoto 1976*, Wiley, 195-263, (1978).

[15] P. Malliavin: Stochastic Analysis, Grundlehren der mathematischen Wissenschaften, Vol. **313**, Springer, (1997).

[16] Protter P. (2005): Stochastic Integration and Differential Equations, Second Edition, Version 2.1 Springer-Verlag, Heidelberg.

[17] Revuz D., Yor M. (1999): Continuous Martingales and Brownian Motion, third edition, Springer-Verlag, Berlin.

[18] R.S. Strichartz: The Campbell-Baker-Hausdorff-Dynkin formula and solutions of differential equations, Jour. Func. Anal., **72**, 320-345, (1987).

[19] M.E. Taylor: Partial Differential Equations, Qualitative Studies of Linear Equations, Springer, Applied Mathematical Sciences **116**, (1996).