Stefano Bonaccorsi & Enrico Priola

From Brownian Motion to Stochastic Differential Equations

10th Internet Seminar

October 23, 2006

Contents

Part I Brownian Motion

Part I

Brownian Motion

1

Basic concepts of probability theory

This lecture contains basic concepts from probability theory and fix notation that will be used throughout the course. The first part of the lecture introduces some tools from measure theory – with particular regard to Fourier transform in $\mathbb{R}^n$ – of special interest in probability. Later, we consider *random variables* X and introduce some probabilistic concepts. Several textbooks exist on the subject, and our introduction by no means is intended to be complete; for further details, we refer (among others) to `Billingsley` [Bi95], `Dudley` [Du02] or `Kallenberg` [Ka02].

Let Ω be a set. A collection $\mathcal{G}$ of subsets of Ω is called *field* (or *algebra*) if it satisfies: $\Omega \in \mathcal{G}$; $A \in \mathcal{G}$ implies $A^c = \Omega \setminus A$ (the complement of A in Ω) belongs to $\mathcal{G}$; $\mathcal{G}$ is stable under finite intersections of elements of $\mathcal{G}$. The collection $\mathcal{G}$ is called *σ-field* (or *σ-algebra*) if is a field and in addition it is stable under countable intersections of elements. A *pre-measure* μ on a field $\mathcal{G}$ is a function $\mu : \mathcal{G} \to [0, +\infty]$ which is additive (i.e., for any A, $B \in \mathcal{G}$ with $A \cap B = \emptyset$ we have $\mu(A \cup B) = \mu(A) + \mu(B)$). A pre-measure μ is σ-additive if for any sequence $\{A_n\} \subset \mathcal{G}$ such that $A = \bigcup_{n \geq 0} A_n \in \mathcal{G}$ and $A_i \cap A_j = \emptyset$, for $i \neq j$, one has $\mu(A) = \sum_{n \geq 0} \mu(A_n)$. A pre-measure μ is said to be finite if $\mu(\Omega) < +\infty$.

Let $\mathcal{F}$ be a σ-field of subsets of Ω. A σ-additive pre-measure on $\mathcal{F}$ is called a *measure* on $\mathcal{F}$ (or on $(\Omega, \mathcal{F})$). A *probability measure* $\mathbb{P}$ on the space $(\Omega, \mathcal{F})$ is a measure on $(\Omega, \mathcal{F})$ with total mass $\mathbb{P}(\Omega) = 1$. We shall denote by $(\Omega, \mathcal{F}, \mathbb{P})$ the reference *probability space*. Each $A \in \mathcal{F}$ will be called an *event*.

There is no loss of generality in assuming, since now, that the reference probability space is *complete*. By definition, a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is *complete* if $\mathcal{F}$ contains all $\mathbb{P}$-*null sets*. Recall that $N \subset \Omega$ is a $\mathbb{P}$-null set if there exists $B \in \mathcal{F}$ such that $N \subset B$ and $\mathbb{P}(B) = 0$. Actually, it is always possible to enlarge the σ-field $\mathcal{F}$ and extend the probability measure in order to get a complete probability space.

An *Euclidean space* is a subset of $\mathbb{R}^n$, $n \in \mathbb{N}$, endowed with the topology induced by the euclidean metric of $\mathbb{R}^n$; it will be denoted by E. The Borel σ-algebra of E (i.e., the smallest σ-algebra which contains all open sets of E) will be indicated by $\mathcal{E}$ or $\mathcal{B}(E)$. We denote by $|\cdot|$ and $\langle \cdot, \cdot \rangle$ respectively the euclidean norm and the inner product in $\mathbb{R}^n$.

1.1 Measure theory in Euclidean spaces

Let $\mathcal{M}(E)$ be the set of all *Borel finite measures* on $(E, \mathcal{E})$. Every $\mu \in \mathcal{M}(E)$ is necessarily regular, i.e., for any $B \in \mathcal{E}$ and for every $\varepsilon > 0$, there exists a compact set $K \subset B$ such that $\mu(B \setminus K) < \varepsilon$. For a given $\mu \in \mathcal{M}(E)$ we define:

$$\mu(f) = \int_E f(y)\,\mu(\mathrm{d}y), \qquad f \in C_b(E);$$

here $C_b(E)$ denotes the Banach space of all real continuous and bounded functions on E endowed with the supremum norm.

The *weak convergence of measures* in $\mathcal{M}(E)$ is defined as follows:

$$\text{weak convergence:} \quad \mu_n \to \mu \quad \text{if} \quad \mu_n(f) \to \mu(f) \quad \forall\, f \in C_b(E).$$

Uniqueness of weak limit follows from the fact that $C_b(E)$ is rich enough to distinguish two measures. Namely, if for two measures μ and ν in $\mathcal{M}(E)$ it holds

$$\int_E \phi(x)\,\mu(\mathrm{d}x) = \int_E \phi(x)\,\nu(\mathrm{d}x)$$

for all $\phi \in C_b(E)$ then $\mu = \nu$.

Finally, note that if the sequence $\{\mu_n\}$ which weakly converge to μ consists of probability measures then μ is a probability measure as well.

1.1.1 Fourier transform

Let $\mu \in \mathcal{M}(\mathbb{R}^n)$; its *Fourier transform* or *characteristic function*, denoted by $\hat{\mu}$, is a mapping from $\mathbb{R}^n$ into $\mathbb{C}$ defined by

$$\hat{\mu}(u) = \int \mathrm{e}^{\mathrm{i}\langle u,x\rangle}\mu(\mathrm{d}x), \qquad u \in \mathbb{R}^n. \tag{1.1}$$

Lemma 1.1. *Let μ be a Borel probability measure on $\mathbb{R}^n$; then its Fourier transform is a uniformly continuous and bounded mapping, with $\hat{\mu}(0) = 1$.*

Proof. Since $|\mathrm{e}^{\mathrm{i}\langle u,x\rangle}| = 1$ for any choice of (real) vectors u and x, we get

$$|\hat{\mu}(u)| \le \int |\mathrm{e}^{\mathrm{i}\langle u,x\rangle}|\mu(\mathrm{d}x) = \int 1\mu(\mathrm{d}x) = 1, \quad u \in \mathbb{R}^n,$$

and, in particular, the equality holds for $\hat{\mu}(0)$.

Let us show that $\hat{u}$ is uniformly continuous. Fix $\varepsilon > 0$. Since μ is regular there exists a compact set K such that $\mu(\mathbb{R}^n \setminus K) < \varepsilon$ (K is contained in a ball $B(0,\alpha)$ for some $\alpha > 0$). Using that $|\langle x,u\rangle| \le |x|\alpha$, for $u \in K$, we get that there exists $\delta > 0$ such that $|\mathrm{e}^{\mathrm{i}\langle x_1,u\rangle} - \mathrm{e}^{\mathrm{i}\langle x_2,u\rangle}| = |1 - \mathrm{e}^{\mathrm{i}\langle (x_2-x_1),u\rangle}| < \varepsilon$ for $|x_2 - x_1| < \delta$ and for any $u \in K$. Therefore

$$\begin{aligned}|\hat{\mu}(x_1) - \hat{\mu}(x_2)| \le &\int_K |\mathrm{e}^{\mathrm{i}\langle x_1,u\rangle} - \mathrm{e}^{\mathrm{i}\langle x_2,u\rangle}|\,\mu(\mathrm{d}u) + \int_{\mathbb{R}^n\setminus K} |\mathrm{e}^{\mathrm{i}\langle x_1,u\rangle} - \mathrm{e}^{\mathrm{i}\langle x_2,u\rangle}|\,\mu(\mathrm{d}u)\\ \le &\varepsilon\mu(K) + 2\mu(\mathbb{R}^n \setminus K) < 2\varepsilon + \varepsilon\mu(\mathbb{R}^n), \quad \text{for} \quad |x_1 - x_2| < \delta.\end{aligned}$$

□

Uniqueness and continuity theorem

A basic result in the theory of characteristic functions concerns the continuity of the mapping $\mu \mapsto \hat{\mu}$ with respect to the weak convergence of measures. This theorem is placed in the very center of probability theory and statistics. As a corollary, we obtain that a characteristic function (or Fourier transform) uniquely determines the measure it comes from.

Theorem 1.2 (Lévy continuity theorem). *Given a sequence $\{\mu_n\}$ of Borel probability measures on $\mathbb{R}^n$, the following hold:*

(a) if μ_n weakly converges to $\mu \in \mathcal{M}(\mathbb{R}^n)$, then $\hat{\mu}_n \to \hat{\mu}$ uniformly on compact sets in $\mathbb{R}^n$;
(b) if the sequence of Fourier transform $\hat{\mu}_n$ converges (pointwise) to a function ϕ in $\mathbb{R}^n$, and moreover ϕ is continuous in 0*, then there exists a unique Borel probability measure μ such that $\phi = \hat{\mu}$ and $\mu_n \to \mu$ in weak sense.*

In particular, if we take $\mu_n \equiv \nu$, we conclude that a Borel probability measure μ on $\mathbb{R}^n$ is uniquely determined by its characteristic function $\hat{\mu}$.

1.1.2 Dynkin's π-λ theorem

We present in this section the π-λ theorem of Dynkin. It is an important tool, which allows to extend a result proved in a class $\mathcal{C}$ of event to the generated σ-field $\sigma(\mathcal{C})$. An application of this result is called *monotone-class argument*. We first introduce some notation.

Let $\mathcal{O}$ be a class of subsets of Ω. We denote by $\sigma(\mathcal{O})$ the smallest σ-algebra which contains $\mathcal{O}$. This is given by the intersection of all σ-algebras in Ω containing $\mathcal{O}$ (note that the collection of all subsets in Ω is a σ-algebra which contains $\mathcal{O}$).

A class $\mathcal{A}$ of subsets of Ω is called a *π-system* if it is closed under the action of finite intersections: for any $A, B \in \mathcal{A}$, $A \cap B \in \mathcal{A}$. Further, a class $\mathcal{L}$ is called a *λ-system* if it verifies the following properties

1. $\Omega \in \mathcal{L}$;
2. $\mathcal{L}$ is closed under the action of complement: $A \in \mathcal{L}$ implies $A^c \in \mathcal{A}$;
3. $\mathcal{L}$ is closed under countable union of *disjoint* sets: if $\{A_j \in \mathcal{L},\ j \in \mathbb{N}\}$ and $A_i \cap A_j = \emptyset$ for $i \neq j$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{L}$.

Problem 1.1.
1. Prove that a class that is both a π-system and a λ-system is a σ-field.
2. Prove that if a λ-system $\mathcal{L}$ contains both A and $A \cap B$, it contains also $A \cap B^c$.
3. Given two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ on $(\Omega, \mathcal{F})$, prove that $\mathcal{A} = \{A \in \mathcal{F} : \mathbb{P}_1(A) = \mathbb{P}_2(A)\}$ is a λ-system.

We can now state the Dynkin's π-λ theorem.

Lemma 1.3. *If $\mathcal{A}$ is a π-system and $\mathcal{L}$ is a λ-system, then $\mathcal{A} \subset \mathcal{L}$ implies that $\sigma(\mathcal{A}) \subset \mathcal{L}$.*

Proof. We denote $\ell(\mathcal{A})$ the minimal λ-system over $\mathcal{A}$, i.e., the intesection of all λ-systems that contains $\mathcal{A}$. Further, for any set A, let $\mathcal{G}_A$ denote the class of sets B such that $A \cap B \in \ell(\mathcal{A})$.

To get the result, we prove that $\ell(\mathcal{A})$ is a π-system; due to exercise 1.1(1), this means that it is a σ-field, hence it contains $\sigma(\mathcal{A})$ (due to the minimality of the σ-algebra $\sigma(\mathcal{A})$); on the other hand, if $\mathcal{L}$ is a λ-system containing $\mathcal{A}$, it also contains $\ell(\mathcal{A})$, again for the minimality of the λ-system $\ell(\mathcal{A})$. Therefore, we have the following sequence of inclusions:

$$\mathcal{A} \subset \sigma(\mathcal{A}) \subset \ell(\mathcal{A}) \subset \mathcal{L},$$

which shows the result.

We proceed with the first step in our construction. If $A \in \ell(\mathcal{A})$, $\mathcal{G}_A$ is a λ-system. In fact we have: 1. $A \cap \Omega = A \in \ell(\mathcal{A})$, hence $\Omega \in \mathcal{G}_A$; 2. $B \in \mathcal{G}_A$ means $A \cap B \in \ell(\mathcal{A})$ then, by exercise 1.1(2), $A \cap B^c \in \ell(\mathcal{A})$ and $B^c \in \ell(\mathcal{A})$; 3. if $\{B_n,\ n \in \mathbb{N}\} \subset \mathcal{G}_A$ is a sequence of disjoint sets, then $\{A \cap B_n,\ n \in \mathbb{N}\} \subset \ell(\mathcal{A})$, so that $\bigcup(A \cap B_n) = A \bigcap(\cup B_n) \in \ell(A)$ which means $(\cup B_n) \in \mathcal{G}_A$.

Now take $A \in \mathcal{A}$; then for any $B \in \mathcal{A}$ we have $B \in \mathcal{G}_A$, hence $\mathcal{A} \subset \mathcal{G}_A$ and $\ell(A) \subset \mathcal{G}_A$, due to the minimality of the λ-system $\ell(\mathcal{A})$. In turn, this implies that for $B \in \ell(\mathcal{A}) \subset \mathcal{G}_A$, $A \cap B \in \ell(\mathcal{A})$. This implies that $A \in \mathcal{A}$ implies $A \in \mathcal{G}_B$, hence $\mathcal{A} \subset \mathcal{G}_B$ and $\ell(\mathcal{A}) \subset \mathcal{G}_B$.

We have then: if $A \in \ell(\mathcal{A})$ and $B \in \ell(\mathcal{A})$, then $A \in \mathcal{G}_B$, i.e., $A \cap B \in \ell(\mathcal{A})$, which means that $\ell(\mathcal{A})$ is closed under finite intersection, i.e., it is a π-system.

□

Problem 1.2. Using (3) in Exercise 1.1 and Lemma 1.3 prove that two Borel probability measures on $\mathbb{R}$ which coincide on bounded intervals are equal.

1.1.3 Caratheodory's extension theorem

In this section, we recall the celebrated *Caratheodory's extension theorem*; we shall let some details from measure theory as exercises. This result will be used in constructing gaussian measures on the space $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$, as well as in the proof of Kolmogorov's extension theorem.

We define $\mathbb{R}^\infty$ as the vector space of all real valued sequences, endowed with the product topology; it becomes a complete separable metric space by considering the distance

$$d(x,y) = \sum_{k=1}^{\infty} 2^{-k} \frac{|x_k - y_k|}{1 + |x_k - y_k|}, \qquad x = \{x_k\},\ y = \{y_k\} \in \mathbb{R}^\infty,$$

which induces exactly the product topology.

On the space $\mathbb{R}^\infty$ we consider the family $\mathcal{R}$ of all *cylindrical sets* $I_{k_1,\dots,k_n,A}$, where $n \in \mathbb{N}$, $k_1, \dots, k_n \in \mathbb{N}$ and $A \in \mathcal{B}(\mathbb{R}^n)$,

$$I_{k_1,\dots,k_n,A} := \{x \in \mathbb{R}^\infty \,:\, (x_{k_1}, \dots, x_{k_n}) \in A\}.$$

It is clear that the family $\mathcal{R}$ of cylindrical sets is an algebra. For instance,

$$(I_{k_1,\dots,k_n,A})^c = I_{k_1,\dots,k_n,A^c}.$$

Note that the σ-algebra generated by the algebra $\mathcal{R}$ of cylindrical sets coincides with the Borel σ-algebra generated by all open sets of $\mathbb{R}^\infty$. Indeed, any closed ball with respect to the metric d is a countable intersection of cylindrical sets.

The following definition concernes the continuity of a measure.

Definition 1.4. *Let $\mathcal{R}$ be an algebra of subsets of Ω. We say that a collection $\{A_n\} \subset \mathcal{R}$* decreases to 0 *if $A_{n+1} \subset A_n$ for each n and $\bigcap_{n=1}^{\infty} A_n = \emptyset$.*

A pre-measure $\mu : \mathcal{R} \to [0,1]$ is continuous in 0 *if for every sequence $\{A_n\} \subset \mathcal{R}$ decreasing to 0, it follows that $\mu(A_n) \to 0$ as $n \to 0$.*

Problem 1.3. We are given an algebra $\mathcal{R}$; every pre-measure μ on $\mathcal{R}$ which is continuous in 0, is also σ-additive. On the other hand, every finite σ-additive pre-measure μ is continuous in 0.

The main result in this section deals with the following problem: given a probability pre-measure μ on a space Ω, that is σ-additive on an algebra $\mathcal{R}$, is it possible to extend it univocally to a σ-additive measure defined on a σ-algebra which contains $\mathcal{R}$? The answer is positive, if we take, as σ-algebra containing $\mathcal{R}$, the σ-algebra $\sigma(\mathcal{R})$; this result is stated in the next *Caratheodory's extension theorem.*

Theorem 1.5. *Let μ be a pre-measure on $(\Omega, \mathcal{R})$, where $\mathcal{R}$ is an algebra; assume that μ is continuous in 0 and that $\mu(\Omega) = 1$. Then, there exists a unique probability measure (that we will still denote by μ) which extends μ to the σ-algebra $\sigma(\mathcal{R})$.*

1.2 Random variables

A function $X : \Omega \to E$ which is measurable from $(\Omega, \mathcal{F})$ into $(E, \mathcal{E})$ is called a *random variable* on $(\Omega, \mathcal{F})$ with values in E. The *σ-field generated by X,* denoted by $\sigma(X)$, is the smallest σ-field which makes X measurable. It is clear that $\sigma(X)$ coincides with the set $\{X^{-1}(B),\ B \in \mathcal{E}\}$; moreover the assertion that X is a random variable is equivalent to say that $\sigma(X) \subset \mathcal{F}$.

In general, given a set $\{X_t,\ t \in T\}$ of random variables, where T is a set of indices, the smallest σ-algebra with respect to which all the random variables X_t are measurable is called the *σ-algebra generated by* $\{X_t\}$ and is denoted with $\sigma(X_t, t \in T)$.

For any set $A \in \mathcal{E}$, $(X \in A)$ or $\{X \in A\}$ will denote the event $X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$. The *law* of X is the probability measure μ_X on $(E, \mathcal{E})$ defined as

$$\mu_X(A) = \mathbb{P}(X \in A), \qquad A \in \mathcal{E}.$$

A random variable X is said to have a property $\mathcal{T}$ $\mathbb{P}$-*almost surely* (abbreviated as a.s.) or with probability 1, if

$$A = \{\omega \in \Omega : X(\omega) \text{ has the property } \mathcal{T}\} \in \mathcal{F}$$

and further $\mathbb{P}(A) = 1$.

For a random variable X taking values in $\mathbb{R}^d$, we define the *distribution function* $F : \mathbb{R}^d \to [0,1]$

$$F(x_1, \dots, x_d) = \mathbb{P}(X_1 \le x_1, \dots, X_d \le x_d).$$

F determines the law of the random variable.

An important class of random variables is given by those X whose law is absolutely continuous with respect to the Lebesgue measure. In that case, there exists a *density function* $f : \mathbb{R}^d \to \mathbb{R}$, $f \ge 0$ and $\int_{\mathbb{R}^d} f(x) \mathrm{d}x = 1$, with the property

$$\mu(A) = \int_A f(x)\,\mathrm{d}x, \qquad A \in \mathcal{B}(\mathbb{R}^d).$$

1.2.1 Integration

The *expected value* or mean of a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is the Lebesgue integral

$$\mathbb{E}[X] = \int_\Omega X(\omega)\,\mathbb{P}(\mathrm{d}\omega).$$

All the definitions and results from integration theory apply. If X is real valued, non-negative, its expected value always exists (eventually equals $+\infty$); if X is real valued, we say that X has (finite) mean if at least one (resp., both) of $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ is finite, in which case we set $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$. In the same way, given an E-valued random variable X, we shall say that $|X|$ has finite mean if $|X|$ is integrable, i.e., if $\mathbb{E}[|X|] < \infty$. In a natural way, we can say that X has *finite p-th moment* if $X \in L^p(\Omega, E)$, $1 \le p < +\infty$ (the norms of $L^p(\Omega, E)$ is given by $\|X\|_p = \mathbb{E}[|X|^p]^{1/p}$, $1 \le p < +\infty$); X is bounded if $X \in L^\infty(\Omega, E)$.

We shall frequently refer to the following change-of-variables formula for integrals, which allows to evaluate the expected value of a random variable X using an integral in X.

Proposition 1.6. *Given $X : \Omega \to E$ a random variable having law μ, let $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a measurable function. Then f is μ-integrable if and only if $f \circ X$ is $\mathbb{P}$-integrable, and in this case it holds*

$$\int_E f(x)\,\mu(\mathrm{d}x) = \int_\Omega f(X(\omega))\,\mathbb{P}(\mathrm{d}\omega) = \mathbb{E}[f(X)].$$

For a one-dimensional random variable X, the k-th (absolute) moment of X is determined by

$$\mathbb{E}[|X|^k] = \int_\mathbb{R} |x|^k\,\mu(\mathrm{d}x), \quad k \in \mathbb{N}.$$

If the first two moments of X exist finite, we define the *variance* of X

$$\mathrm{Var}(X) = \mathbb{E}[|X - \mathbb{E}[X]|^2] = \mathbb{E}[|X|^2] - |\mathbb{E}[X]|^2.$$

In case X has a density function f, it holds

$$\mathbb{E}[g(X)] = \int_E g(x) f(x)\,\mathrm{d}x,$$

for any Borel measurable function g provided that $g \cdot f$ is integrable.

Example 1.7. There is an interesting relationship between moments and tail probability. Assume that X is a real valued random variable with $X \ge 0$ a.s., having a density function f; an application of Fubini's theorem implies

$$\begin{aligned}\mathbb{E}[X^p] &= \int_0^\infty x^p f(x)\,\mathrm{d}x = \int_0^\infty \Big(\int_0^x p t^{p-1}\,\mathrm{d}t\Big) f(x)\,\mathrm{d}x \\ &= \int_0^\infty p t^{p-1}\,\mathrm{d}t \int_t^\infty f(x)\,\mathrm{d}x = \int_0^\infty p t^{p-1} \mathbb{P}(X > t)\,\mathrm{d}t, \quad p > 0.\end{aligned}$$

Moreover, if X is a random variable with distribution μ, then its characteristic function φ_X is given by

$$\varphi_X(u) = \mathbb{E}[e^{iuX}] = \int e^{i\langle u,x\rangle}\mu(dx) = \hat{\mu}(u), \quad u \in \mathbb{R}^n.$$

Remark that the existence of finite moments for X implies also regularity of the characteristic function φ_X. For instance, assume that X has finite mean; then φ_X is differentiable and

$$\frac{\partial}{\partial u_j}\varphi_X(u) = i\int x_j e^{i\langle u,x\rangle}\,\mu(dx).$$

In particular, $\nabla_u \varphi_X(0) = i\mathbb{E}[X]$.

This result extends to higher order moments as well. If X has finite second moment, then the Hessian matrix of $\varphi_X(u)$ computed in $u = 0$ is equal to minus the covariance matrix[1] of X.

Among all the results about integration, there stand two specific inequalities of probability measures.

Proposition 1.8 (Jensen's inequality). *Let X be an integrable $\mathbb{R}^d$-valued random variable and let $\Phi : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then $Y = \Phi(X)$ is a real valued random variable, lower semi-integrable, and*

$$\mathbb{E}(\Phi(X)) \geq \Phi(\mathbb{E}(X)). \tag{1.2}$$

Proposition 1.9 (Chebyshev's inequality). *Let X be a real valued, square integrable random variable. Then for every $\alpha > 0$:*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \alpha) \leq \frac{1}{\alpha^2}V(X). \tag{1.3}$$

If further X has finite p-th moment, then it holds

$$\mathbb{P}(|X| \geq \alpha) \leq \frac{\mathbb{E}(|X|^p)}{\alpha^p} \tag{1.4}$$

for every $\alpha > 0$.

1.2.2 Convergence in law

Let $\{X_n\}$ be a sequence of random variables taking values in some Euclidean space $(E, \mathcal{E})$ (a priori, there is a different reference probability space for each X_n). The sequence $\{X_n,\ n \in \mathbb{N}\}$ *converges in law* to some random variable X if the relative sequence of distribution laws $\{\mu_n\}$ is *weakly convergent* to the law μ of X, *i.e.*, for every $\phi : E \to \mathbb{R}$ continuous and bounded,

$$\int_E \phi(x)\,\mu_n(dx) \longrightarrow \int_E \phi(x)\,\mu(dx) \qquad \text{as } n \to \infty.$$

It is natural to compare the convergence in law with other type of convergence for random variables. Assuming that all the random variable X_n are defined on the same probability space, we say that

[1] the *covariance matrix* of a random vector $X = (X_1, \dots, X_n)$ is the symmetric matrix $Q = \big(q_{ij}\big)$ with entries

$$q_{ij} = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])], \qquad i, j = 1, \dots, n.$$

(1) $\{X_n,\ n \in \mathbb{N}\}$ converges a.s. to X if $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$;
(2) $\{X_n,\ n \in \mathbb{N}\}$ converges in probability to X if for any $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$, as $n \to \infty$;
(3) $\{X_n,\ n \in \mathbb{N}\}$ converges in L^p, $p \geq 1$, to X (assuming in addition that X_n and X are in $L^p(\Omega, E)$) if $\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0$.

It turns out that (1) $\Rightarrow$ (2) $\Rightarrow$ (convergence in law) and moreover (3) $\Rightarrow$ (2). In general, these implications can not be reversed.

1.2.3 Independence

In general, the idea of *independence* can be formulated for events $A \in \mathcal{F}$, for random variables and for σ-algebras.

1. Given two events A and B (elements of $\mathcal{F}$), we say that they are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
2. Given two random variables X and $Y : \Omega \to E$, they are independent if, for every $F, G \in \mathcal{E}$:
$$\mathbb{P}(X \in F, Y \in G) = \mathbb{P}(X \in F)\mathbb{P}(Y \in G),$$
where as usual $(X \in F, Y \in G) = \{\omega \in \Omega \,:\, X(\omega) \in F,\ \ Y(\omega) \in G\}$.
3. We shall say that two sub-σ-algebrae $\mathcal{F}'$ and $\mathcal{G}'$ of $\mathcal{F}$ are independent if for any $F \in \mathcal{F}'$ and $G \in \mathcal{G}'$, F e G are independent.

Previous definition of independence easily extends from two to any finite number of objects. For instance, 3. becomes

3'. given $\mathcal{F}_1, \ldots, \mathcal{F}_m$ sub-σ-fields in $\mathcal{F}$, they are (globally) independent if for any choice of $F_1 \in \mathcal{F}_1, \ldots, F_m \in \mathcal{F}_m$, $F_1, \ldots, F_m$ are independent (in particular, thanks to the arbitrariety of our choice of $F_1, \ldots, F_m$, it is enough to show that $\mathbb{P}(\bigcap_{i=1}^m F_i) = \prod_{i=1}^m \mathbb{P}(F_i)$).

It will be also necessary to talk about independence for an infinite number of objects; then, we shall require that any finite subset of objects consists of independent objects.

Problem 1.4. Show that two events A and B are independent if and only if the random variables $\mathbb{1}_A$ and $\mathbb{1}_B$ are independent (hence 1. is a special case of 2.).

Further, it is easy to show that also 2. is a special case of 3., i.e., two random variables X and Y are independent if and only if $\sigma(X)$ and $\sigma(Y)$ are such.

Finally, we say that a random variable X is independent from a sub-σ-field $\mathcal{G}$ if the σ-field generated by X, i.e., $\sigma(X)$, is independent from $\mathcal{G}$.

As an application of Dynkin's π-λ theorem, we can prove the following property of random variables.

Problem 1.5. Let $X_1, \ldots, X_n$ be random variables with values in $\mathbb{R}^k$ and laws $\mu_1, \ldots, \mu_n$ respectively. Set $X = (X_1, \ldots, X_n)$ that is a random vector with values in $\mathbb{R}^{n\,k}$.

(a) Prove that $X_1, \dots, X_n$ are independent if and only if $\mu_X = \mu_1 \otimes \dots \otimes \mu_n$ on $\mathcal{B}(\mathbb{R}^{n\,k})$ (i.e., μ_X is the product measure of $\mu_1, \dots, \mu_n$).
(b) Assume that $k = 1$; then the random variables $X_1, \dots, X_n$ are independent if and only if for any choice of intervals $(a_1, b_1), \dots, (a_n, b_n)$ it holds

$$\mathbb{P}(a_1 \le X_1 < b_1, \dots, a_n \le X_n < b_n) = \prod_{i=1}^n \mathbb{P}(a_i \le X_i < b_i).$$

(c) Assume again that $k = 1$; then the random variables $X_1, \dots, X_n$ are independent if and only if for any choice of functions $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$, Borel measurable and bounded, it holds

$$\mathbb{E}[f_1(X_1) \dots f_n(X_n)] = \mathbb{E}[f_1(X_1)] \dots \mathbb{E}[f_n(X_n)].$$

(d) Let φ_{X_i} be the characteristic function of X_i and φ_X the characteristic function of the vector X. Then $X_1, \dots, X_n$ are independent if and only if

$$\varphi_X(u) = \varphi_{X_1}(u_1) \cdot \ldots \cdot \varphi_{X_n}(u_1), \qquad u \in \mathbb{R}^n.$$

Problem 1.6.
(a) If real random variables $X_1, \dots, X_n$ are independent and belongs to L^2, then

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

(b) Let X and Y be real valued random variables, independent and such that XY is integrable; then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Construct two random variables which show that the converse implication does not hold in $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega = [-1/2, 1/2]^2$, $\mathcal{F} = \mathcal{B}([-1/2, 1/2]^2)$ and $\mathbb{P}$ is the two-dimensional Lebesgue measure.

1.3 Conditional expectation

It is customary to link the idea of conditional probability with the concept of "partial information"; in this section we consider what happens to a random variable in case we modify the information, i.e., we condition with respect to a sub-σ-algebra.

Let $\mathcal{G}$ be a given sub-σ-field in $\mathcal{F}$ and let X be an integrable random variable; note that X in general is not measurable with respect to $\mathcal{G}$. Assume first that $X \ge 0$ a.s. Then we may define $\mathbb{Q}(A) = \mathbb{E}[X \mathbb{1}_A]$ for any $A \in \mathcal{G}$; since Q is a positive measure on $(\Omega, \mathcal{G})$, absolutely continuous with respect to the restriction of $\mathbb{P}$ to $\mathcal{G}$ (still denoted by $\mathbb{P}$), we know by the Radon-Nykodim theorem that there exists an $\mathcal{G}$-measurable integrable random variable $Z = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$. For any $A \in \mathcal{G}$, we get

$$\mathbb{E}[\mathbb{1}_A Z] = \mathbb{Q}(A) = \mathbb{E}[\mathbb{1}_A X]. \tag{1.5}$$

Moreover Z is unique a.s. In the general case, we write $X = X_+ - X_-$, where $X_+ = X \vee 0$ and $X_- = (-X) \vee 0$, and consider as before Z_+ and Z_-. Define $Z = Z_+ - Z_- \in L^1(\Omega, \mathcal{G}, \mathbb{P})$. It is clear that Z verifies (1.5) and Z is unique a.s. We have just proved the following result.

Theorem 1.10 (Existence and uniqueness of conditional expectation). *Let X be a real-valued, integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G}$ be a sub-σ-field in $\mathcal{F}$; then there exists a random variable $Z = \mathbb{E}(X \mid \mathcal{G})$ which is integrable, $\mathcal{G}$-measurable, and verifies (1.5). Further, Z is unique: if Z_1 and Z_2 are integrable, $\mathcal{G}$-measurable, random variables which satisfy (1.5), then $Z_1 = Z_2$ almost surely.*

Definition 1.11. *Let X be an integrable random variable. The integrable $\mathcal{G}$-measurable random variable Z which verifies* (1.5) *is called the* conditional expectation *of X given $\mathcal{G}$. We set $Z = \mathbb{E}(X \mid \mathcal{G})$.*

Notice that the conditional expectation is an "equivalent class" of random variables; hence all the statement below shall be meant to hold almost surely (or, "there exists a version of $\mathbb{E}(X \mid \mathcal{G})$ such that...").

The following result may be useful in some applications. To prove it, use properties of π-systems from section 1.2.3.

Theorem 1.12. *Assume that*

1. *Z is $\mathcal{G}$-measurable;*
2. *$\mathbb{E}(Z\mathbb{1}_D) = \mathbb{E}(X\mathbb{1}_D)$ as D varies in a π-system $\mathcal{D} \subset \mathcal{G}$ which contains Ω and generates $\mathcal{G}$, i.e., $\sigma(\mathcal{D}) = \mathcal{G}$.*

Then $Z = \mathbb{E}(X \mid \mathcal{G})$.

Example 1.13. Let us show how elementary conditional probability $\mathbb{P}(A \mid B)$, $A, B \in \mathcal{F}$, is related to our new terminology. To this purpose, let X be an integrable random variable on $(\Omega, \mathcal{F})$ – that, in particular, we can think as the indicator function of the event $A \in \mathcal{F}$ – and let $\mathcal{G}$ be a sub-σ-algebra in $\mathcal{F}$. Assume that $\mathcal{G}$ is generated by a partition $\{B_i,\ i \in I\}$ of Ω, where I is a numerable set of indices (in particular, we can consider $\mathcal{G} = \mathcal{G}_0 = \{\Omega, B, B^c, \emptyset\}$); then (almost surely) it holds

$$\mathbb{E}(X \mid \mathcal{G}) = \sum_{i \in I} \mathbb{E}_{B_i}(X)\mathbb{1}_{B_i}, \tag{1.6}$$

where $\mathbb{E}_{B_i}(X)$ is the mean of X on B_i:

$$\mathbb{E}_{B_i}(X) = \frac{1}{\mathbb{P}(B_i)} \int_{B_i} X\,\mathbb{P}(d\omega), \qquad \text{if } \ \mathbb{P}(B_i) > 0$$

while we set $\mathbb{E}_{B_i}(X) = 0$ if $\mathbb{P}(B_i) = 0$. Notice that $\mathbb{E}(X \mid \mathcal{G})$ is constant on each B_i.

Thus, as a special case of (1.6), we have, for any $\omega \in B$,

$$\mathbb{P}(A \mid B) := \mathbb{E}(\mathbb{1}_A \mid \mathcal{G}_0)(\omega) = \frac{1}{\mathbb{P}(B)} \int_B \mathbb{1}_A\,\mathbb{P}(\mathrm{d}\omega) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

The next result provides some general properties of the conditional expectation. The proof is immediate and we leave it to the reader.

Proposition 1.14. *Let X and Y be real-valued, integrable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G}$ be sub-σ-field in $\mathcal{F}$; we have, a.s.,*

(1) $\mathbb{E}(\alpha X + \beta Y \mid \mathcal{G}) = \alpha\mathbb{E}(X \mid \mathcal{G}) + \beta\mathbb{E}(Y \mid \mathcal{G}), \quad \alpha, \beta \in \mathbb{R}$*;*
(2) $\mathbb{E}(X \mid \mathcal{G}) = X$ *if* X *is* $\mathcal{G}$*-measurable;*
(3) $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(X)$ *if* X *is independent from* $\mathcal{G}$*;*
(4) $\mathbb{E}(XY \mid \mathcal{G}) = X\mathbb{E}(Y \mid \mathcal{G})$ *if* X *is* $\mathcal{G}$*-measurable and bounded.*

Proposition 1.15. *If* X *is integrable and the sub-*σ*-fields* $\mathcal{G}$ *and* $\mathcal{G}'$ *of* $\mathcal{F}$ *satisfy* $\mathcal{G} \subset \mathcal{G}'$*, then* $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}') \mid \mathcal{G}) = \mathbb{E}(X \mid \mathcal{G})$ *with probability 1.*

This result is known as "absorbing property" of conditional expectation. Heuristically, it says that averaging X with respect to a σ-field $\mathcal{G}'$, then taking another averaging with respect to a smaller σ-field $\mathcal{G}$ is the same that taking the average directly with respect to $\mathcal{G}$.

Proof. Let us denote $Y = \mathbb{E}(X \mid \mathcal{G})$ and $Y' = \mathbb{E}(X \mid \mathcal{G}')$. For any $B \in \mathcal{G}$ we have $\mathbb{E}[X\mathbb{1}_B] = \mathbb{E}[Y\mathbb{1}_B]$; since obviously $B \in \mathcal{G}'$, it holds $\mathbb{E}[X\mathbb{1}_B] = \mathbb{E}[Y'\mathbb{1}_B]$, so that

$$\mathbb{E}[Y'\mathbb{1}_B] = \mathbb{E}[Y\mathbb{1}_B] \quad \text{for any } B \in \mathcal{G},$$

hence $Y = \mathbb{E}(Y' \mid \mathcal{G}) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{G}') \mid \mathcal{G})$.
□

The following result is a generalization of Jensen's inequality for the conditional expectation.

Proposition 1.16. *Let* X *be a real-valued, integrable random variable on* $(\Omega, \mathcal{F}, \mathbb{P})$ *and let* $\mathcal{G}$ *be a sub-*σ*-field in* $\mathcal{F}$*. If* Φ *is a convex real function such that* $\Phi(X)$ *is an integrable random variable, then*

$$\mathbb{E}(\Phi(X) \mid \mathcal{E}) \geq \Phi(\mathbb{E}(X \mid \mathcal{E})).$$

The above property is formally written as the usual Jensen's inequality for the expectation. However, we shall not forget that the conditional expectation is a random variable, while mathematical expectation is a number.

Notice that for every $p \geq 1$ the mapping $x \mapsto |x|^p$ is a convex function on $\mathbb{R}$. As a consequence of Jensen's inequality we obtain, for every $X \in L^p$:

$$|\mathbb{E}(X \mid \mathcal{E})|^p \leq \mathbb{E}(|X|^p \mid \mathcal{E})$$

and also

$$\mathbb{E}|\mathbb{E}(X \mid \mathcal{E})|^p \leq \mathbb{E}(|X|^p),$$

which shows that

$$X \in L^p(\Omega) \quad \Longrightarrow \mathbb{E}(X \mid \mathcal{E}) \in L^p(\Omega).$$

Moreover the conditional expectation is a linear contraction in each $L^p(\Omega)$, $p \geq 1$.

Problem 1.7.

1. Prove Theorem 1.12.
2. Prove the following converse of property (3) in Proposition 1.14: let X be a d-dimensional random variable and $\mathcal{D} \subset \mathcal{F}$ a σ-algebra such that

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\langle\lambda, X\rangle} \mid \mathcal{D}] = \mathbb{E}[\mathrm{e}^{\mathrm{i}\langle\lambda, X\rangle}] \quad \text{a.s.}$$

   for any $\lambda \in \mathbb{R}^d$. Then X and $\mathcal{D}$ are independent.
3. Prove the following form of property (4) in Proposition 1.14: if X is $\mathcal{G}$-measurable, and Y and XY are integrable, then $\mathbb{E}[XY \mid \mathcal{G}] = X\mathbb{E}[Y \mid \mathcal{G}]$.

1.4 Stochastic processes

Given an Euclidean space E and a set of indices T, we consider a mapping $X : (\Omega, \mathcal{F}) \to (E^T, \mathcal{E}^T)$, where $\mathcal{E}^T$ is the minimal *sigma*-field which makes measurable the coordinate variables $\pi_t : E^T \to E$, $\pi_t f = f(t)$. X is measurable if and only if $X_t = \pi_t \circ X : \Omega \to E$ is measurable for every $t \in T$.

Further, there exists on Ω a minimal σ-field which makes measurable all the coordinate random variables $X_t : \Omega \to E$. We can see X as a collection $X = \{X_t,\ t \in T\}$ of random variables, taking values in E, but also as a function $X = X(t, \omega) : T \times \Omega \to E$. We shall call X a *stochastic process* if T is a subset of the real line, and $t \in T$ is meant to represent *time*.

The space E is called the *state space*. Intuitively, using the identification $\omega \longrightarrow \{t \to X_t(\omega)\}$ we can consider $\Omega \subset E^T$; thus, Ω can be thought of as a *path space*. Sometimes, we will proceed with the identification $\Omega = E^T$. The examples of interest for us will be $T = \mathbb{N}$, $\mathbb{Z}$, $[0, \infty)$, $\mathbb{R}$ while E will be a discrete set (Poisson process), or the Euclidean space $\mathbb{R}^n$ $(n \geq 1)$.

Poisson process

Let us consider a telephone exchange; *calls* arrive in sequence, one after the other, and we may count both the number of calls N_t arrived before (or at) time t, or the (random) time of awaiting X_n between call $n-1$ and call n. Fix an initial time $t = 0$, and let $S_n = X_1 + \dots + X_n$ the time of arrival of n-th call. In general, any possible choice of the random variables X_n is allowed, with no restriction on being identically distributed or independent; we shall only assume the following

$$X_j(\omega) > 0, \qquad \sum_{n=1}^{\infty} X_n(\omega) = \infty, \tag{1.7}$$

for (almost) every $\omega \in \Omega$. The same can be expressed in terms of the S_n, asking that

$$S_1(\omega) > 0, \qquad S_n(\omega) > S_{n-1}(\omega), \qquad \sup_n S_n(\omega) = \infty.$$

The number of calls occurring in the interval $[0, t]$ is denoted by N_t; $\{N_t,\ t \geq 0\}$ is a stochastic process, said the *counting process*, having continuous time and countable state space. Hence $\{N_t,\ t \geq 0\}$ is given by

$$N_t(\omega) = \sum_{n \geq 1} \mathbb{1}_{\{S_n \leq t\}}(\omega) = \sum_{n \geq 1} n \mathbb{1}_{\{S_n \leq t < S_{n+1}\}}(\omega), \quad \omega \in \Omega,$$

where X_n are given random variables defined on the reference probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that (1.7) holds and $S_n = X_1 + \dots + X_n$, $n \geq 1$. Note that N_t takes values in $\mathbb{N}$ and moreover $N_0 = 0$ a.s.

By definition, a *Poisson process* is a counting process related to a sequence of *waiting times* X_n which are independent, identically distributed and have *exponential distribution of parameter* $\lambda > 0$, i.e., $X_n : \Omega \to (0, +\infty)$, a.s., and

$$\mathbb{P}(X_n > t) = \mathrm{e}^{-\lambda t}, \qquad t > 0$$

(the exponential distribution of parameter $\lambda > 0$, has density function $\lambda e^{-\lambda t}$ on $(0, \infty)$).

As a first step, we compute the distribution of S_n. Notice that we have $\{S_n > t\}$ equals $\{N_t < n\}$. Using the independence of $\{X_k\}$, we know that the density function $f_n(x)$ of S_n is the n-time convolution of $f_1(x) = \lambda \mathrm{e}^{-\lambda x}$, $x > 0$. We claim that

$$f_n(t) = \lambda \mathrm{e}^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, \qquad t > 0, \;\; n = 1, 2, \dots$$

To prove this it is sufficient to see that

$$f_n * f_1(t) = \int_0^t f_n(t-s) f_1(s) \, \mathrm{d}s = f_{n+1}(t).$$

Using the density $f_n(t)$, we compute

$$\mathbb{P}(S_n > t) = \int_t^\infty f_n(s) \, \mathrm{d}s = \sum_{k=0}^{n-1} \mathrm{e}^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

Now, for any $t \geq 0$, the event $\{N_t = 0\}$ means that nothing happens in the interval $[0, t]$, that is, $\{X_1 > t\}$ and so $\mathbb{P}(N_t = 0) = \mathbb{P}(X_1 > t) = \mathrm{e}^{-\lambda t}$. We can easily compute the distribution of N_t:

$$\mathbb{P}(N_t = n) = \mathbb{P}(S_n \leq t < S_{n+1}) = \mathbb{P}(S_n \leq t, \, S_{n+1} > t) = \mathbb{P}(S_{n+1} > t) - \mathbb{P}(S_n > t) = \mathrm{e}^{-\lambda t} \frac{(\lambda t)^n}{n!}.$$

Lemma 1.17. *A Poisson process verifies the following properties:*

(1) for $0 < t_1 < \cdots < t_n$, N_{t_1}, $N_{t_2} - N_{t_1}$, $\dots$, $N_{t_n} - N_{t_{n-1}}$ *are independent, and*
(2) $\mathbb{P}(N_{t_k} - N_{t_{k-1}} = n) = \mathrm{e}^{-\lambda(t_k - t_{k-1})} \frac{(\lambda(t_k - t_{k-1}))^n}{n!}$, *for* $n \in \mathbb{N}$.

Assertion (1) says that the Poisson process has independent increments; assertion (2) says that the increment $N_t - N_s$ *has a Poisson distribution of parameter* $\lambda(t-s)$, *for any* $t > s > 0$.

Proof. We can fix $t > 0$ and see what happens after time t. By definition, $S_{N_t} \leq t < S_{N_t+1}$, where $S_{N_t}(\omega) := S_{N_t(\omega)}(\omega)$ for $\omega \in \Omega$; the waiting time before the first call after t is $S_{N_t+1} - t$, the waiting times between the first and the second call after t is $X_{N_t+2} = (S_{N_t+2} - t) - (S_{N_t+1} - t)$, and so on. Therefore, we can define a new sequence of waiting times

$$X_1^t = S_{N_t+1} - t, \; X_2^t = X_{N_t+2}, \; \dots$$

and a new counting process

$$N_s^t = N_{t+s} - N_t$$

such that $N_s^t \geq n$ if and only if $X_1^t + \cdots + X_n^t \leq s$, or

$$N_s^t = \max\{n \, : \, X_1^t + \cdots + X_n^t \leq s\}.$$

Notice the similarity between the definitions of N_s and N_s^t, which differs only for the use of different ("shifted") random waiting times.

We show next that $\{N^t_s,\ s \geq 0\}$ is again a Poisson process, proving that X^t_j are independent, identically distributed and with exponential distribution of parameter λ. This will immediately imply the assertion (2).

We first verify that X^t_1 is exponentially distributed; to this purpose we fix n and t, and consider the probability

$$\begin{aligned}\mathbb{P}(S_n \leq t < S_{n+1},\ S_{n+1} - t > s) &= \mathbb{P}(S_n \leq t, X_{n+1} + S_n > s + t) \\ &= \mathbb{E}[\mathbb{1}_{(0,t)}(S_n)\, \mathbb{1}_{(s+t,+\infty)}(S_n + X_{n+1})].\end{aligned}$$

In order to compute the last expectation, we use the independence of S_n and X_{n+1}. This gives that the law of the random variable (S_n, X_{n+1}) with values in $(0, +\infty)^2$ is the product of the laws of S_n and X_{n+1}. Hence we have

$$\begin{aligned}\mathbb{P}(S_n \leq t < S_{n+1}, S_{n+1} - t > s) &= \mathbb{E}[\mathbb{1}_{(0,t)}(S_n)\, \mathbb{1}_{(s+t,+\infty)}(S_n + X_{n+1})] \\ &= \int_0^t \lambda e^{-\lambda u} \frac{(\lambda u)^{n-1}}{(n-1)!}\, du \int_0^\infty \mathbb{1}_{(s+t,\infty)}(u+v) \lambda e^{-\lambda v}\, dv \\ &= \int_0^t \lambda e^{-\lambda u} \frac{(\lambda u)^{n-1}}{(n-1)!}\, du \int_{s+t-u}^\infty \lambda e^{-\lambda v}\, dv = \int_0^t \lambda e^{-\lambda u} \frac{(\lambda u)^{n-1}}{(n-1)!} e^{-\lambda(s+t-u)}\, du \\ &= e^{-\lambda s} e^{-\lambda t} \frac{(\lambda t)^n}{n!} = e^{-\lambda s}\, \mathbb{P}(S_n \leq t < S_{n+1}). \qquad (1.8)\end{aligned}$$

Now, setting $S_0 = 0$,

$$\begin{aligned}\mathbb{P}(X^t_1 > R) &= \sum_{n \geq 0} \mathbb{P}(S_{N_t+1} - t > R,\ N_t = n) = \sum_{n \geq 0} \mathbb{P}(S_{n+1} - t > R, S_n \leq t < S_{n+1}) \\ &= \sum_{n \geq 0} e^{-\lambda R} e^{-\lambda t} \frac{(t\lambda)^n}{n!} = e^{-\lambda R}, \quad R > 0.\end{aligned}$$

This shows that X^t_1 is exponentially distributed. It is easy to check that also X^t_j, $j \geq 2$, are exponentially distributed with parameter λ.

Concerning the independence of X^t_j, we only verify that X^t_1 and X^t_2 are independent. The proof of the independence of any finite subset of $\{X^t_j\}$ follows in the same way. For any $R, M > 0$, we get

$$\begin{aligned}\mathbb{P}(X^t_1 > R, X^t_2 > M) &= \sum_{n \geq 0} \mathbb{P}(S_{N_t+1} - t > R,\ X_{N_t+2} > M, N_t = n) \\ &= \sum_{n \geq 0} \mathbb{P}(S_{n+1} - t > R, X_{n+2} > M, S_n \leq t < S_{n+1}) \\ &= \mathbb{P}(X_{n+2} > M) \sum_{n \geq 0} \mathbb{P}(S_{n+1} - t > R, S_n \leq t < S_{n+1}) = e^{-\lambda M} \sum_{n \geq 0} e^{-\lambda R} e^{-\lambda t} \frac{(t\lambda)^n}{n!} \\ &= e^{-\lambda M} e^{-\lambda R} = \mathbb{P}(X^t_1 > R)\, \mathbb{P}(X^t_2 > M).\end{aligned}$$

Let us come back to formula (1.8). This formula can be generalized as follows.

$$\mathbb{P}(S_n \le t < S_{n+1},\ S_{n+1} - t > s_1,\ X_{n+2} > s_2,\ \dots, X_{n+j} > s_j) \\ = \mathbb{P}(S_n \le t < S_{n+1}) \mathrm{e}^{-\lambda s_1} \cdot \dots \cdot \mathrm{e}^{-\lambda s_j},\ \ j \ge 1.$$

Introducing the set $H = (s_1, \infty) \times \cdots \times (s_j, \infty)$, previous formula can be equivalently written as

$$\mathbb{P}(S_n \le t < S_{n+1},\ (X_1^t, \dots, X_j^t) \in H) = \mathbb{P}(S_n \le t < S_{n+1}) \mathbb{P}((X_1^t, \dots, X_j^t) \in H)$$

(we have also used that $X_1^t, \dots, X_j^t$ are independent and identically exponentially distributed) or

$$\mathbb{P}(N_t = n,\ (X_1^t, \dots, X_j^t) \in H) = \mathbb{P}(N_t = n) \mathbb{P}((X_1^t, \dots, X_j^t) \in H). \tag{1.9}$$

Since the events $\{(X_1^t, \dots, X_j^t) \in H\}$ form a π-system, the above equality can be extended to the case of general $H \in \mathcal{B}(\mathbb{R}^j)$.

Consider next the event $\{N_s^t = m\} = \{N_t - N_s = m\}$: we can express it as $\{(X_1^t, \dots, X_m^t) \in H\}$, where H is the set in $\mathbb{R}^{m+1}$ of points $x_1, \dots, x_{m+1}$ such that

$$x_1 + \cdots + x_m \le s < x_1 + \cdots + x_{m+1}.$$

Hence, for any n, $m \in \mathbb{N}$, (1.9) gives

$$\mathbb{P}(N_t = n,\ N_{t+s} - N_t = m) = \mathbb{P}(N_t = n) \mathbb{P}(N_{t+s} - N_t = m), \tag{1.10}$$

which implies at once that N_t and $N_{t+s} - N_t$ are independent. The extension to the case of 3 or more increments is quite similar.

□

Corollary 1.18. *N_t is a process with stationary increments, i.e., for any $s, t, u, v \in R_+$, $0 \le s < t < +\infty$, $0 \le u < v < +\infty$, such that $t - s = v - u$, the law of $N_t - N_s$ is the same as the law of $N_v - N_u$.*

We leave as an exercise to prove the following equivalent characterization of a Poisson process.

Problem 1.8.

1. The Poisson process with rate $\lambda > 0$ can also be characterized as an integer-valued process, starting from 0, with non-decreasing paths, with independent increments, and such that, as $h \to 0$, uniformly in t,

$$\mathbb{P}(N_{t+h} - N_t = 0) = 1 - \lambda h + o(h), \\ \mathbb{P}(N_{t+h} - N_t = 1) = \lambda h + o(h).$$

2. The following properties show that even if you know the number of arrivals in an interval, nothing can be said to the distribution of their times.
   Let $\{N_t,\ t \ge 0\}$ be a Poisson process. Assume that exactly one jump happens in the interval $[0, t]$, i.e., $N_t = 1$. Show that conditioned on this, X_1 is uniformly distributed in $[0, t]$: $\mathbb{P}(X_1 \le s \mid N_t = 1) = s/t$ for $0 \le s \le t$. This can be eneralized as follows: if we know that exactly n jumps happen in the interval $[s, s+t]$; the times at which jumps occur are uniformly and independently distributed on $[s, s+t]$.

2

Brownian motion

This second lecture is devoted to introduce two main subjects in probability theory: Gaussian measures and Brownian motion. In the first section, we shall introduce the class of Gaussian probability measures on $E = \mathbb{R}^n$, while the second section will be devoted to study Gaussian processes. Both these classes of objects are fundamental in modern probability theory and shall return frequently in the following.

There are several technical points from abstract measure theory involved in the definition of Browian motion. In this lecture, we shall see Brownian motion as an special case of Gaussian process. This characterization will provide useful in the proof of existence of this process, which is the object of third section. We shall appeal to some fundamental theorems due to Kolmogorov in order to prove the result. First, Kolmogorov's existence theorem 2.13 proves the existence of a stochastic process with the same finite dimensional distributions as Brownian motion; this does not suffice, since our definition requires continuity of sample paths: but finally we can appeal to Kolmogorov's theorem of continuity, Theorem 2.16, and conclude the proof of existence.

2.1 Gaussian measures

We start with the simple one dimensional case $E = \mathbb{R}$; we shall say that a Borel probability measure μ on $\mathbb{R}$ is a *Gaussian law* (or a *Gaussian distribution*) $\mathcal{N}(a, \sigma^2)$, for $a \in \mathbb{R}$, $\sigma^2 > 0$, if its density with respect to the Lebesgue measure exists and is given by the function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-a)^2\right), \quad x \in \mathbb{R}.$$

If $\sigma^2 = 0$ we set

$$\mathcal{N}(a, 0) = \delta_a,$$

where δ_a is the Dirac measure (concentrated) in a. If $a = 0$, then $\mathcal{N}(0, \sigma^2)$ is called a centered Gaussian measure.

Let $X : \Omega \to \mathbb{R}$ be a random variable; we say that X is a *Gaussian random variable* if its law μ is a Gaussian measure; in particular, if $\mu = \mathcal{N}(a, \sigma^2)$ we write $X \sim \mathcal{N}(a, \sigma^2)$. In this case, X has mean $\mathbb{E}[X] = a$ and variance $\operatorname{Var}(X) = \sigma^2$. The characteristic function of X (i.e., the Fourier transform of μ) is given by

$$\hat{\mu}(\theta) = \mathbb{E}[e^{iX\theta}] = \int_{\mathbb{R}} e^{i\theta x} \mu(dx) = e^{i\theta a} e^{-\frac{1}{2}\sigma^2\theta^2}.$$

Problem 2.1. Given a sequence $\{\sigma_n,\ n \in \mathbb{N}\}$ of real numbers, converging to 0, let $\{X_n\}$ be a sequence of centered Gaussian random variables, each with variance σ_n^2. Show that the sequence $\{X_n\}$ converges in law to Dirac's measure in 0.

Recall that for given Borel probability measures on $\mathbb{R}$ μ and ν, the convolution measure $\mu * \nu$ is defined by

$$(\mu * \nu)(B) = \int_{\mathbb{R}} \mu(B - x)\nu(dx), \qquad B \in \mathcal{B}(\mathbb{R}).$$

It is known that Fourier transform maps convolution of measures into pointwise product of characteristic functions; using this fact, and the above expression of characteristic function for a Gaussian law, we obtain easily the following result.

Proposition 2.1. *Given two Gaussian laws $\mu = \mathcal{N}(a, \sigma^2)$ and $\nu = \mathcal{N}(b, \tau^2)$, it holds*

$$\widehat{(\mu * \nu)}(\theta) = \hat{\mu}(\theta)\hat{\nu}(\theta) = e^{i\theta(a+b)} \exp\left(-\frac{1}{2}(\sigma^2 + \tau^2)\theta^2\right).$$

Hence, $\mu * \nu$ is the Gaussian law $\mathcal{N}(a + b, \sigma^2 + \tau^2)$. In particular, if X and Y are real independent Gaussian random variables, also $X + Y$ is a Gaussian random variable.

We consider next the multidimensional case. We are given n random variables $X_1, \dots, X_n$, independent, with Gaussian law $\mathcal{N}(0, 1)$; setting $X = (X_1, \dots, X_n)$, the law of the random vector X has density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \dots \frac{1}{\sqrt{2\pi}} e^{-x_n^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}|x|^2} \tag{2.1}$$

and characteristic function

$$e^{-\frac{1}{2}\theta_1^2} \cdots e^{-\frac{1}{2}\theta_n^2} = e^{-\frac{1}{2}|\theta|^2}.$$

If Q is a matrix in $L^+(\mathbb{R}^n)$ (the space of symmetric non-negative defined $n \times n$ matrices), it can be expressed as $Q = AA^*$, where A is a $n \times n$ matrix and A^* denotes the adjoint matrix of A; moreover A is uniquely determined in the class of symmetric matrices: in this case, A is also denoted as $Q^{1/2}$. For given $a \in \mathbb{R}^n$ and $Q \in L^+(\mathbb{R}^n)$, then $Q^{1/2}X + a$ has characteristic function

$$\phi(\theta) = e^{i\langle\theta,a\rangle} e^{-\frac{1}{2}\langle Q^{1/2}\theta, Q^{1/2}\theta\rangle} = e^{i\langle\theta,a\rangle} e^{-\frac{1}{2}\langle Q\theta,\theta\rangle}, \quad \theta \in \mathbb{R}^n. \tag{2.2}$$

A Borel probability measure μ on $\mathbb{R}^n$ is a *(multidimensional) Gaussian law* $\mathcal{N}(a, Q)$ if its characteristic function is given by (2.2).

Remark that if the matrix Q is invertible then also $Q^{1/2}$ is invertible, and so there exists a density g for $Z = Q^{1/2}X + a$; this density is given from (2.1) by a change-of-variables formula

$$g(y) = \frac{1}{(2\pi)^{n/2}(\mathrm{Det}(Q))^{1/2}} e^{-\frac{1}{2}\langle Q^{-1}(y-a),(y-a)\rangle}.$$

Conversely, if the covariance matrix Q is singular, $Q^{1/2}$ shall be singular as well, and the law of $Z = Q^{1/2}X + a$ is supported by some proper affine subspace of $\mathbb{R}^n$, so it cannot have a density. As a special case, we mention the law $\mathcal{N}(a, 0)$, Dirac's delta measure concentrated in a, having characteristic function $\theta \to \mathrm{e}^{\mathrm{i}\langle\theta,a\rangle}$.

The following result can be easily proved by using the Fourier transform.

Proposition 2.2. *Let X be a Gaussian random variable in $\mathbb{R}^n$ with distribution $\mathcal{N}(a, Q)$; consider the random variable $Y = AX + b$, where $A \in L(\mathbb{R}^n, \mathbb{R}^k)$ and $b \in \mathbb{R}^k$. Then Y has Gaussian distribution $\mathcal{N}(Aa + b, AQA^*)$.*

Applying the previous result, we get that, for each Gaussian random variable Z with values in $\mathbb{R}^n$, $Z \sim \mathcal{N}(a, Q)$, there exists a random variable X such that $Z = a + Q^{1/2}X$ and $X \sim \mathcal{N}(0, I)$ (I denotes the identity matrix in $\mathbb{R}^n$). Computing the derivatives in 0 of the characteristic function, we get that a is the mean of Z and Q is the covariance matrix.

In general, two random variables $X_1, X_2 : \Omega \to \mathbb{R}$ are said *uncorrelated* if their covariance matrix is diagonal. Of course, if they are independent then they are also uncorrelated, but in general the converse does not hold. However, for Gaussian families the following holds:

$$\begin{aligned}&\text{if } (X_1, X_2) \text{ is a Gaussian family, then}\\ &X_1 \text{ and } X_2 \text{ Gaussian are independent} \iff \text{they are uncorrelated.}\end{aligned} \tag{2.3}$$

In fact, if X_1 and X_2 are uncorrelated, then the characteristic function of the vector $X = (X_1, X_2)$ has the form $\phi_X(h) = \exp(\mathrm{i}\langle h, m\rangle - \frac{1}{2}\langle Qh, h\rangle)$ where Q is the diagonal covariance matrix. Hence ϕ_X is equal to the product $\phi_{X_1} \cdot \phi_{X_2}$, so that X_1 and X_2 are independent, see Problem 1.5.

More generally, if X_1 and X_2 are respectively $\mathbb{R}^d$-valued and $\mathbb{R}^n$-valued random variables with joint Gaussian distribution such that

$$q_{ij} = \mathbb{E}[(X_1^i - \mathbb{E}[X_1^i])(X_2^j - \mathbb{E}[X_2^j])] = 0, \qquad i = 1, \dots, d \; j = 1, \dots, n,$$

then they are independent.

A similar result holds for n random variables as well; for instance, if $(X_1, \dots, X_n)$ is a Gaussian family of real random variables, then they are independent if and only if their covariance matrix is diagonal.

2.2 Stochastic processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the reference probability space; in this section we mainly consider stochastic processes, taking values in the Euclidean space $E = \mathbb{R}^d$, $d \geq 1$, endowed with the Borel σ-field $\mathcal{E} = \mathcal{B}(E)$.

Recall that a stochastic process $X = \{X_t,\ t \in T\}$ with values in E is a collection of random variables $X_t : \Omega \to E$, $t \in T$. It can be equivalently defined as a measurable mapping $X : (\Omega, \mathcal{F}) \to (E^T, \mathcal{E}^T)$, where $\mathcal{E}^T$ is the minimal σ-field which makes measurable all the projections $\pi_t : E^T \to E$, $\pi_t(f) = f(t)$, $f \in E^T$. We can also see X as a collection of random variables taking values in the space of paths E^T, since $\{X_t(\omega),\ t \in T\}$ is just an ordinary function of time (for every ω).

A stochastic process $\{X_t,\ t \in T\}$ is called *continuous* or *almost surely continuous* if its trajectories (or paths) $t \mapsto X_t(\omega)$, $T \to E$, are continuous (respectively, if almost every trajectory is continuous).

Similar definitions hold for *right continuous*, *almost surely right continuous* stochastic processes, and so on.

Let $\{\mathcal{F}_t, t \in T\}$ be a *filtration*, i.e., a family of sub-σ-fields of $\mathcal{F}$ increasing in time

$$\text{if } s < t \text{ then } \mathcal{F}_s \subset \mathcal{F}_t.$$

Intuitively, "$\mathcal{F}_t$ contains all the information which are available up to time t", i.e., all the events whose occourence can be established up to time t.

We add to the filtration $\{\mathcal{F}_t,\ t \in T\}$ the σ-algebra

$$\mathcal{F}_\infty = \bigvee_{t \in T} \mathcal{F}_t.$$

Recall that the right hand side stands for the minimal σ-field which contains all $\mathcal{F}_t$; this is not the same as $\underset{t \in T}{\cup} \mathcal{F}_t$. Define, for every $t \in T$

$$\mathcal{F}_{t+} = \bigwedge_{t<s\in T} \mathcal{F}_s = \bigcap_{t<s\in T} \mathcal{F}_s.$$

We have clearly, for each $t \in T$, $\mathcal{F}_t \subset \mathcal{F}_{t+}$.

Definition 2.3. *The filtration* $\{\mathcal{F}_t,\ t \in T\}$ *is* right-continuous *if for every* $t \in T$*,* $\mathcal{F}_t = \mathcal{F}_{t+}$.

Assume we are given a stochastic process $X = \{X_t,\ t \in T\}$ and a filtration $\{\mathcal{F}_t,\ t \in T\}$ on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that the process X is *adapted* to the filtration $\{\mathcal{F}_t, t \in T\}$ if, for any $t \in T$ fixed, the random variable X_t is $\mathcal{F}_t$-measurable on Ω; equivalently, we say that X is adapted to $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$.

Notice that it is always possible to construct a filtration with respect to which the process is adapted, by setting $\mathcal{F}_t^X = \sigma(X_s, s \le t)$; $\mathcal{F}_t^X$ is called the *natural filtration* of X. A stochastic process X is adapted to a filtration $\{\mathcal{F}_t,\ t \in T\}$ if and only if one has $\mathcal{F}_t^X \subset \mathcal{F}_t$, $t \in T$.

Let $X = \{X_t,\ t \in T\}$ and $Y = \{Y_t,\ t \in T\}$ be two stochastic processes on $(\Omega, \mathcal{F})$. If the random variables X_t coincide almost surely with Y_t, i.e.,

$$\text{for each } t \in T: \qquad \mathbb{P}\{X_t \neq Y_t\} = 0,$$

X is called a *version* (or a *modification*) of the process Y.

In general, this does not imply that if X is adapted to a filtration $\mathcal{F}_t,\ t \in T$, then also Y is adapted to the same filtration. It may happen, in fact, that the set of zero measure $\{X_t \neq Y_t\}$ does not belong to $\mathcal{F}_t$, so that Y_t is not $\mathcal{F}_t$-measurable.

To avoid this and other technical difficulties, we shall always require that the filtration satisfies in addition the so called *standard assumptions* , i.e.,

(1) the filtration is right continuous, that is, for any $t \in T$: $\mathcal{F}_t = \mathcal{F}_{t+}$;
(2) the filtration is complete, that is, for any $t \in T$, $\mathcal{F}_t$ contains all $\mathbb{P}$-*null sets*.

Remark that each filtration $\{\mathcal{F}_t,\ t \in T\}$ can be completed in order to satisfies hypothesis (2). Indeed, it is enough to enlarge each $\mathcal{F}_t$ by considering the smallest σ-algebra $\mathcal{F}_t \vee \mathcal{N}$ which contains $\mathcal{F}_t$ and the family $\mathcal{N}$ of all $\mathbb{P}$-null sets.

Given a process X, the completion of the natural filtration $\{\mathcal{F}_t^X,\ t \in T\}$ will be called the *completed natural filtration* of X and still denoted by $\{\mathcal{F}_t^X,\ t \in T\}$.

The completed natural filtration of any Brownian motion satisfies the technical hypothesis (1); we will return on this later on.

Definition 2.4. *A* filtered probability space *or a* stochastic basis *is the quadruplet* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$*, where the filtration* $\{\mathcal{F}_t\}$ *verifies the standard assumptions.*

When we say that a stochastic process $X = \{X_t,\ t \in T\}$ *is defined on a stochastic basis* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$*, (implicitly) we also require that it is adapted to* $\{\mathcal{F}_t\}$*.*

Brownian motion

In the following, we introduce the main object of our investigations, namely Brownian motion.

Definition 2.5. *A d-dimensional* Brownian motion $B = \{B_t,\ t \geq 0\}$ *defined on a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P})$ *is a continuous stochastic process such that*

1. $B_0 = 0$ *a.s.;*
2. *for every* $0 \leq s < t$ *the random variable* $B_t - B_s$ *is independent from* $\mathcal{F}_s$*;*
3. *for every* $0 \leq s < t$ *the random variable* $B_t - B_s$ *has Gaussian law* $N(0, (t-s)I_d)$*, where* I_d *is the (d-dimensional) identity matrix.*

We shall again underline that, as opposite to Poisson process, we have just given a formal definition of Brownian motion without providing a construction of it, and we shall return later to the proof of its existence.

Remark 2.6. In the following, if the filtration $\{\mathcal{F}_t\}$ is not explicitly mentioned, we can assume that we are using the completed natural filtration $\{\mathcal{F}_t^B\}$ of B. Remark that in `Protter` [Pr04] it is shown that the completed natural filtration of any Brownian motion is right continuous and so it satisfies the standard assumptions. A Brownian motion B defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t^B,\ t \geq 0\}, \mathbb{P})$ is also called a *natural Brownian motion.*

Problem 2.2. (a) Shows that if in Definition 2.5 we consider the completed natural filtration $\{\mathcal{F}_t^B\}$, then condition 2. is equivalent to the following one:

2'. for any choice of times $0 < t_1 < \cdots < t_m$,

$$B_{t_1},\ B_{t_2} - B_{t_1}, \ldots,\ B_{t_m} - B_{t_{m-1}}$$

are independent random variables.

This condition says the Brownian motion is a process with independent increments (compare with (1) in Lemma 1.17 on the Poisson process).

(b) Consider a Brownian motion $B = \{B_t,\ t \geq 0\}$ defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P})$; let $\mathcal{G}$ be a σ-field independent from $\mathcal{F}_\infty$. Setting $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{G}$, show that B is a $\{\mathcal{G}_t\}$-Browian motion.

Remark 2.7. To conclude this introduction, we compute the moments of a real Brownian motion B_t for fixed t. Since B_t has a Gaussian distribution $\mathcal{N}(0,t)$, for every $n \in \mathbb{N}$ there exists a constant $C_n = \frac{(2n)!}{2^n n!}$ such that

$$\mathbb{E}|Z|^{2n}] = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} x^{2n} \mathrm{e}^{-x^2/2t}\, dx = t^n \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} y^{2n} \mathrm{e}^{-y^2/2}\, dy = C_n t^n.$$

In particular, for $n = 1$ and $n = 2$, we have

$$\mathbb{E}[|B_t|^2] = t, \qquad \mathbb{E}[|B_t|^4] = 3t^2.$$

2.2.1 Gaussian processes

Definition 2.8. *A stochastic process $X = \{X_t,\, t \in T\}$ is called a* Gaussian process *if for every choice of times $t_1, \dots, t_m \in T$ and scalars $\gamma_1, \dots, \gamma_m \in \mathbb{R}$, the random variable $\gamma_1 X_{t_1} + \dots + \gamma_m X_{t_m}$ has Gaussian distribution.*

The following proposition expresses a different characterization of Gaussian processes.

Proposition 2.9. *$X = \{X_t,\, t \in T\}$ is a Gaussian process if and only if for every choice of times $t_1, \dots, t_m \in T$ the random vector $(X_{t_1}, \dots, X_{t_m})$ has a Gaussian distribution in $\mathbb{R}^{d \cdot m}$.*

Proof. We will use Proposition 2.1. First, assume that X is a Gaussian process and fix times $t_1, \dots, t_m \in T$. We introduce $\xi = (X_{t_1}, \dots, X_{t_m})$ and prove that ξ has a Gaussian distribution. Take $d = 1$ for simplicity. For every vector $\gamma = (\gamma_1, \dots, \gamma_m) \in \mathbb{R}^m$, the scalar product $\langle \gamma, \xi \rangle$ is a real Gaussian random variable with mean a and variance σ^2, so that its characteristic function verifies

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\langle \gamma, \xi \rangle \theta}] = \mathrm{e}^{\mathrm{i}a\theta - \sigma^2\theta^2/2}, \quad \theta \in \mathbb{R}. \tag{2.4}$$

Note that in particular each X_{t_i} is a Gaussian random variable. We can compute a and σ^2 in terms of the mean μ and the covariance matrix Γ of the random vector ξ:

$$a = \mathbb{E}[\langle \gamma, \xi \rangle] = \mathbb{E}[\sum_{j=1}^m \gamma_j X_{t_j}] = \sum_{j=1}^m \gamma_j\, \mathbb{E}[X_{t_j}] = \sum_{j=1}^m \gamma_j \mu_{t_j} = \langle \gamma, \mu \rangle,$$

and similarly

$$\sigma^2 = \mathbb{E}[(\langle \gamma, \xi \rangle - \langle \gamma, \mu \rangle)^2] = \mathbb{E}[\langle \gamma, \xi - \mu \rangle^2] = \mathbb{E}[\sum_{i,j=1}^m \gamma_i \gamma_j (X_{t_i} - \mathbb{E}[X_{t_i}])(X_{t_j} - \mathbb{E}[X_{t_j}]) = \langle \Gamma\gamma, \gamma \rangle.$$

By (2.4) with $\theta = 1$, we find that the characteristic function of ξ is

$$\hat{\mu}(\gamma) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\langle \gamma, \xi \rangle}] = \mathrm{e}^{\mathrm{i}\langle \gamma, \mu \rangle - \frac{1}{2}\langle \Gamma\gamma, \gamma \rangle},$$

and we recognize the characteristic function of a Gaussian random variable.

Conversely, if $\xi = (X_{t_1}, \dots, X_{t_m})$ is a random vector, then for every vector $\gamma \in \mathbb{R}^m$ the mapping $\xi \mapsto \langle \gamma, \xi \rangle$ is a linear transformation hence, by Proposition 2.2, maps Gaussian distributions into Gaussian distributions. It follows that X is a Gaussian process.

□

Given two stochastic processes $X = \{X_t,\, t \in T\}$ and $Y = \{Y_t,\, t \in T\}$ on $(\Omega, \mathcal{F}, \mathbb{P})$, we can express their independence in terms of the generated filtrations: the process $\{X_t,\ t \in T\}$ is independent from $\{Y_t,\ t \in T\}$ if $\sigma(X_t,\, t \in T)$ is independent from $\sigma(Y_t,\, t \in T)$.

Proposition 2.10. *Consider two families $X = \{X_t,\, t \in T\}$ and $Y = \{Y_t,\, t \in T\}$ such that $X \cup Y$ is Gaussian. Then X is independent from Y if and only if*

$$\mathbb{E}[X_t Y_s] = \mathbb{E}[X_t]\,\mathbb{E}[Y_s], \tag{2.5}$$

for every $t, s \in T$.

Proof. From (2.5) it follows that for any $t, s \in T$, the Gaussian random variables X_t and Y_s are independent, see (2.3). By a similar criterium, thanks to the arbitrariness of t and s, we have that any two random vectors $(X_{t_1}, \dots, X_{t_n})$ and $(Y_{s_1}, \dots, Y_{s_m})$ are independent.

Now the thesis follows thanks to Dynkin's theorem 1.3. One uses that $\sigma(X) = \sigma(X_t,\, t \in T)$ and $\sigma(Y) = \sigma(Y_t,\, t \in T)$ are generated by the π-systems of events $\{X_{t_1} \in A_1, \dots, X_{t_n} \in A_n\}$ and $\{Y_{s_1} \in B_1, \dots, Y_{s_m} \in B_m\}$ respectively, and that these events are independent due to the independence of the random vectors $(X_{t_1}, \dots, X_{t_n})$ and $(Y_{s_1}, \dots, Y_{s_m})$.

□

Problem 2.3. Give the details of the proof of Proposition 2.10.

Hint: first fix $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R}^d)$ and set $B = \{X_{t_1} \in A_1, \dots, X_{t_n} \in A_n\}$. Define

$$\Lambda = \{A \in \sigma(Y) \text{ such that } A \text{ is independent from } B\}.$$

Show that Λ is a λ-system. Since Λ contains the events $\{Y_{s_1} \in B_1, \dots, Y_{s_m} \in B_m\}$ one deduces that $\Lambda = \sigma(Y)$. How can we finish the proof?

Using the above definition, it is possible to prove that a Brownian motion is a Gaussian process. Choose real numbers $\alpha_1, \dots, \alpha_m$ and times $0 < t_1 < \dots < t_m$: the goal is to prove that $\alpha_1 B_{t_1} + \dots + \alpha_m B_{t_m}$ is a random variable with Gaussian law. For $m = 1$ the claim obviously holds. By induction, we suppose that the claim holds for $m - 1$; thus, we can write

$$\alpha_1 B_{t_1} + \dots + \alpha_m B_{t_m} = [\alpha_1 B_{t_1} + \dots + (\alpha_{m-1} + \alpha_m) B_{t_{m-1}}] + \alpha_m (B_{t_m} - B_{t_{m-1}})$$

and the conclusion follows since the sum of two Gaussian, independent random variables again has a Gaussian distribution.

We may characterize a Brownian motion as a Gaussian process in the following way.

Proposition 2.11. *Given a real Brownian motion $B = \{B_t\}$, then*

a) $B_0 = 0$ *a.s.;*
b) *for any choice of times $0 \le t_1 < t_2 < \dots < t_m$ the m-dimensional random variable $(B_{t_1}, \dots, B_{t_m})$ has a centered Gaussian law;*
c) $\mathbb{E}(B_t B_s) = s \wedge t$.

Conversely, if a continuous process $\{B_t\}$ verifies the above conditions a), b) and c), then it is a natural Brownian motion (i.e., it is a Brownian motion with respect to its completed natural filtration).

Proof. If B is a Brownian motion, condition a) is obvious, condition b) follows by the characterization of Gaussian processes stated in Proposition 2.9 and condition c) holds since, for $s \le t$, we have

$$\mathbb{E}[B_t B_s] = \mathbb{E}[(B_t - B_s)B_s] + \mathbb{E}(B_s^2) = s = s \wedge t.$$

Conversely, if a Gaussian process B satisfies conditions a), b) and c), then property 1. of Definition 2.5 holds; also, the law of $B_t - B_s$ is Gaussian thanks to condition b), it is centered since B_t and B_s are, and its variance is equal to

$$\mathbb{E}[(B_t - B_s)^2] = \mathbb{E}[B_t^2] - 2\mathbb{E}[B_t B_s] + \mathbb{E}[B_s^2] = t - 2t \wedge s + s = t - s,$$

hence $B_t - B_s$ has Gaussian law $N(0, t-s)$. Finally, to show that $B_t - B_s$ is independent from $\mathcal{F}_s^B$, thanks to Proposition 2.10 it is enough to prove that, for each $\tau \le s$, $\mathbb{E}[(B_t - B_s)B_\tau] = 0$:

$$\mathbb{E}[(B_t - B_s)B_\tau] = \mathbb{E}[B_t B_\tau] - \mathbb{E}[B_s B_\tau] = t \wedge \tau - s \wedge \tau = \tau - \tau = 0.$$

□

2.3 Kolmogorov's existence theorem

Often, a stochastic process $X = \{X_t,\ t \in T\}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is described by means of its *finite dimensional distributions*. These are defined as the probability measures $\mu_{t_1,\dots,t_m}$ that X induces on $\mathcal{B}(E^m)$, for each $m \in \mathbb{N}$ and any choice of times $t_1 < t_2 < \cdots < t_m$ in T:

$$\mu_{t_1,\dots,t_m}(A_1 \times \cdots \times A_m) = \mathbb{P}(X_{t_1} \in A_1, \dots, X_{t_m} \in A_m), \qquad A_1, \dots, A_m \in \mathcal{E}.$$

We say that two stochastic processes with values in E (even defined on different probability spaces) are *equivalent* or that they are *equal in law* if they have the same finite dimensional distributions.

Remark 2.12. Let us give a heuristic justification of the term *equality in law.* A stochastic process $X = \{X_t,\ t \in T\}$ can be seen as a random variable with values in E^T. Hence, we may define the law μ_X of X on E^T, endowed with the product σ-algebra $\mathcal{E}^T$, i.e., $\mu_X(F) = \mathbb{P}(X \in F)$, $F \in \mathcal{E}^T$. Now, since μ_X is completely determined by its values on the sets of the form

$$\{f \in E^T \mid f(t_1) \in A_{t_1}, \dots, f(t_m) \in A_{t_m}\}, \qquad A_{t_i} \in \mathcal{E}, i = 1, \dots, m,$$

we have that the finite dimensional distributions of X characterize the law μ_X.

If $\{X_t,\ t \in T\}$ is a real Gaussian process, then its (finite dimensional) distributions are determined by the *mean function*

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

and the *covariance function*

$$\rho(t, s) = \mathbb{E}[(X_t - m(t))(X_s - m(s))].$$

Indeed, the joint density of $X_{t_1}, \dots, X_{t_n}$ is just

$$\mu_{t_1,\dots,t_n}(A_1 \times \dots \times A_n) = \mathcal{N}(m(\boldsymbol{t}), \Sigma)(A_1 \times \dots \times A_n) \tag{2.6}$$

where Σ is the symmetric $n \times n$ matrix

$$\Sigma = \Sigma(t_1, \dots, t_n) = \begin{pmatrix} \rho(t_1,t_1) & \rho(t_1,t_2) & \dots & \rho(t_1,t_n) \\ \rho(t_2,t_1) & \rho(t_2,t_2) & \dots & \rho(t_2,t_n) \\ \vdots & \vdots & \vdots & \vdots \\ \rho(t_n,t_1) & \rho(t_n,t_2) & \dots & \rho(t_n,t_n) \end{pmatrix},$$

$\boldsymbol{x}$ is the row vector $(x_1, \dots, x_n)$ and $m(\boldsymbol{t})$ is the row vector $(m(t_1), \dots, m(t_n))$.

Problem 2.4. Compute the finite dimensional distributions of the Poisson process N_t.

Hint: first, see Exercise 1.8, it holds

$$\mathbb{P}(N_t = n) = \mathrm{e}^{-\lambda t} \frac{(\lambda t)^n}{n!}, \qquad n \in \mathbb{N}.$$

Next, for any finite sequence of times $t_0 = 0 < t_1 < \dots < t_k$, using the fact that N_t has independent increments, show that

$$\mu_{t_1,\dots,t_k}(n_1, \dots, n_k) = \mathbb{P}(N_{t_1} = n_1, \dots, N_{t_k} = n_k) = \prod_{j=1}^{k} \mathrm{e}^{-\lambda(t_j - t_{j-1})} \frac{(\lambda(t_j - t_{j-1}))^{n_j - n_{j-1}}}{(n_j - n_{j-1})!},$$

for any integers $n_0 = 0 \le n_1 \le \dots \le n_k$.

Notice that the finite-dimensional distributions of a stochastic process $\{X_t,\ t \in T\}$ always satisfy the following *consistency condition*:

$$\mu_{t_1,\dots,t_m}(A_1 \times \dots \times A_m) = \mu_{t_1,\dots,t_m,t_{m+1}}(A_1 \times \dots \times A_m \times E), \tag{2.7}$$

with $t_1 < t_2 < \dots < t_m$ in T.

More interesting is the converse problem: given a family of probability measures $\mu_{t_1,\dots,t_m}$ on $\mathcal{B}(E^m)$, for any choice of m and $t_1 < t_2 < \dots < t_m$ in T, which satisfies (2.7), does there exist a stochastic process (defined on some $(\Omega, \mathcal{F}, \mathbb{P})$) having these as finite dimensional distributions? The answer is affirmative and is given by the Kolmogorov existence theorem. To formulate this we introduce some notation.

Let T be an arbitrary set of indices and E^T be the collection of all functions from T into E; E^T can be thought of as a product space of many copies of E, indexed by T. On this space we consider the *projections* π_t, $\pi_t(x(\cdot)) = x(t)$, $x \in E^T$, $t \in T$. Recall that there exists a natural extension to E^T of the Borel σ-field for vector spaces E^k, that we denote by $\mathcal{E}^T$; this is the smallest σ-field which makes each $\pi_t : E^T \to E$ measurable.

Let us define on $(E^T, \mathcal{E}^T)$ the *coordinate process* $Z = \{Z_t,\ t \in T\}$,

$$Z_t : E^T \to E, \quad Z_t(x) = \pi_t(x) = x(t), \qquad x \in E^T.$$

If we provide the space $(E^T, \mathcal{E}^T)$ with a probability measure $\mathbb{P}$, we obtain that $\{Z_t,\ t \in T\}$ is a stochastic process.

Theorem 2.13. *Let $\{\mu_{t_1,\dots,t_m}\}$ be a family of Borel probability distributions satisfying the consistency condition (2.7). Then there exists a unique probability measure $\mathbb{P}$ on $(E^T, \mathcal{E}^T)$ such that the coordinate process $\{Z_t,\ t \in T\}$ on $(E^T, \mathcal{E}^T, \mathbb{P})$ has finite dimensional distributions which coincide with $\mu_{t_1,\dots,t_m}$, for any choice of $t_1 < t_2 < \cdots < t_m$ in T.*

For our pourposes, it is not essential to prove this theorem (the interested reader is referred, for instance, to `Billingsley` [Bi95]). Hence, we only make a remark on the proof in the case $T = \mathbb{N}$.

First one defines $\mathbb{P}$ on the algebra $\mathcal{A}$ of all cylindrical sets

$$I_{t_1,\dots,t_n,A} = \{x \in E^{\mathbb{N}} \,:\, (x_{t_1}, \dots, x_{t_n}) \in A\},$$

where $n \in \mathbb{N}$, $t_1 < \ldots < t_n \in \mathbb{N}$ and $A \in \mathcal{B}(E^n)$, setting

$$\mathbb{P}(I_{t_1,\dots,t_n,A}) := \mu_{t_1,\dots,t_n}(A). \tag{2.8}$$

This definition is meaningful thanks to the consistency condition. It is easy to check that $\mathbb{P}$ is a pre-measure on $\mathcal{A}$. The crucial step of the proof is now to show that $\mathbb{P}$ is continuous (to this purpose, one also uses that Borel measures on E are regular). After the main step is achieved, an application of the Caratheodory extension theorem allows to get that there exists a unique probability measure $\mathbb{P}$ on the σ-algebra generated by $\mathcal{A}$, which coincides with $\mathcal{E}^T$, such that (2.8) holds. Finally, it is clear that finite dimensional distributions of the coordinate process Z_t coincide with the measures $\mu_{t_1,\dots,t_m}$.

Next, we investigate a question arising from the previous result.

Remark 2.14. It must be noticed that in the above construction the choice of E^T is unsatisfactory under several points of view. Let $T = [0,\infty)$ and $E = \mathbb{R}$. It turns out that on the space $\mathbb{R}^{[0,\infty)}$ finite dimensional distributions do not characterize important properties of the trajectories of a process, as for instance continuity.

Indeed, the set $\mathcal{C}$ of continuous functions from $[0,\infty)$ into $\mathbb{R}$ does not belong to $\mathcal{E}^T$. This fact can be explained as follows: given a real function on $[0,\infty)$ it is not possible to establish if it is continuous by looking only to its values on a countable set of points in $[0,\infty)$.

In the next exercise we propose to show a characterization of the σ-algebra $\mathcal{E}^T$, when $T = [0,\infty)$ and $E = \mathbb{R}^n$.

Problem 2.5. Let us consider $\mathcal{E}^T$, when $T = [0,\infty)$ and $E = \mathbb{R}^n$. Show that the family $\mathcal{H}$ of all sets of the form

$$\{x \in E^T \,:\, (x(t_1), x(t_2), \dots, x(t_n), \dots) \in B\},$$

for any choice of sequences of times $t_1 < \ldots < t_n < \ldots$ in T and of sets $B \in \prod_{i\in\mathbb{N}}(\mathcal{E})_i$ (here $\prod_{i\in\mathbb{N}}(\mathcal{E})_i$ denotes the product σ-algebra, where $(\mathcal{E})_i = \mathcal{E}$, $i \in \mathbb{N}$) is a σ-field in E^T.

Deduce that $A \in \mathcal{E}^T$ if and only if there is a sequence of times $t_1 < \ldots < t_n < \ldots$ in T and a set $B \in \prod_{i\in\mathbb{N}}(\mathcal{E})_i$ so that $A = \{x \in E^T \,:\, (x(t_1), x(t_2), \dots, x(t_n), \dots) \in B\}$.

Example 2.15. Let us show that it is possible to construct a process with the same finite-dimensional distributions of the Poisson process $\{N_t,\ t \geq 0\}$, but which exhibits completely different properties of the paths (for instance, they are neither right-continuous, nor monotone non-decreasing).

Take the function $f(t) = t$ if t is rational, $f(t) = 0$ otherwise. Let X_1 be the first waiting time in the definition of the Poisson process $\{N_t\}$ and consider the process $M_t = N_t + f(t + X_1)$. We have

$\mathbb{P}(f(t+X_1)=0)=1$, since the complementary event is $X_1 \in \mathbb{Q}-t$ ($\mathbb{Q}$ is the set of rational point), and the probability that a continuous random variable takes value in a countable set is zero. Thus

$$\mathbb{P}(M_t = N_t) = 1 \qquad \text{for all } t \geq 0,$$

and the process M_t has the same finite-dimensional distributions of N_t, but its paths are everywhere discontinuous, are not integer valued nor monotone.

2.3.1 A construction of Brownian motion using Kolmogorov's theorem

Since a real Brownian motion is a Gaussian process, its finite dimensional distributions are specified by the mean function $m(t)=0$ and covariance function $\rho(t,s)=\min\{s,t\}$ as in (2.6): for $0<t_1<t_2<\dots<t_n$

$$\mu_{t_1,\dots,t_n}(A_1\times\dots\times A_n) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \int\limits_{A_1\times\dots\times A_n} \exp\Big(-\frac{1}{2}\langle \boldsymbol{x}, \Sigma^{-1}\boldsymbol{x}\rangle\Big)\,\mathrm{d}x_1\dots\mathrm{d}x_n, \quad (2.9)$$

where Σ is the $n\times n$ matrix

$$\Sigma = \begin{pmatrix} t_1 & t_1 & \dots & t_1 \\ t_1 & t_2 & \dots & t_2 \\ \vdots & \vdots & \vdots & \vdots \\ t_1 & t_2 & \dots & t_n \end{pmatrix}$$

$\det\Sigma = t_1\cdot(t_2-t_1)\dots(t_n-t_{n-1})>0$ and $\boldsymbol{x}$ is the row vector $(x_1,\dots,x_n)$. Consistency in this case is obvious. Thus Kolmogorov's theorem 2.13 applies: there exists a stochastic process $\{B_t,\ t\geq 0\}$ corresponding to the finite dimensional distributions $\mu_{t_1,\dots,t_n}$ (see (2.9)) on the probability space $(\mathbb{R}^{[0,+\infty)}, \mathcal{E}^{[0,+\infty)}, \mathbb{P})$ with a suitable probability measure $\mathbb{P}$.

Note that a similar construction can be applied, more generally, for proving the existence of a Brownian motion with values in $\mathbb{R}^d$, $d\geq 1$.

However, the Kolmogorov construction does not provide continuity of trajectories of Brownian motion. Actually, we have already observed that the set of continuous functions is not contained in $\mathcal{E}^T$. A second theorem, again due to Kolmogorov, allows to overcome this difficulty.

2.4 Continuity of Brownian motion paths

Recall that a function $f: E\to\mathbb{R}$ is called α-Hölder continuous on $D\subset E$, $\alpha\in(0,1)$, if there exists a constant $c>0$ such that

$$|f(x)-f(y)| \leq c|x-y|^\alpha \qquad \text{for all } x,y\in D.$$

Let $\{X_t,\ t\in T\}$ be a stochastic process (but the result holds, more generally, for a random field indexed by $T\subset\mathbb{R}^d$, $d\geq 2$); the following result, due to Kolmogorov, asserts that a certain degree of regularity in the mean implies some regularity (Hölder continuity) of the trajectories.

Theorem 2.16. *Let $X = \{X_t, t \in T\}$ be a stochastic process in an Euclidean space $(E, \mathcal{E})$ and assume there exist positive constants α, β and c such that*

$$\mathbb{E}\left(|X_t - X_s|^\beta\right) \le c|t-s|^{1+\alpha}. \tag{2.10}$$

Then there exists a version Y of X with continuous sample paths. Further, for any $\gamma < \frac{\alpha}{\beta}$ trajectories $t \mapsto Y_t(\omega)$ are γ-Hölder continuous on every bounded interval, for any $\omega \in \Omega$.

We give the proof in the particular case of $T = (0,1)$ for simplicity of notation. The proof uses the following purely deterministic lemma.

Lemma 2.17. *Let D be the set of all dyadic numbers in (0,1), i.e., $D = \{\frac{k}{2^n} \;:\; n \in \mathbb{N},\; k = 1, \dots, 2^n - 1\}$. Let $f : D \to \mathbb{R}$ be a function and let $\alpha \in (0,1]$. Consider the quantity*

$$c_n = \sup\left\{\frac{|f(x) - f(y)|}{1/2^{\alpha n}} \;:\; x,\, y \in D,\; |x-y| = 1/2^n\right\}, \quad n \in \mathbb{N}.$$

Then $\sup_{n\in\mathbb{N}} c_n < +\infty$ if and only if f is α-Hölder on D.

Proof (Theorem 2.16). Assume $T =]0,1[$ and denote by D the set of dyadic numbers in T; also, denote by $D_n \subset D$ the set of points with coordinates $k/2^n$, for $k = 1, \dots, 2^n - 1$. Finally, fix some $\gamma < \frac{\alpha}{\beta}$.
Step 1. Set

$$B_n = \{\omega : |X_i - X_j| > 2^{-n\gamma} \text{ for some } i, j \in D_n \text{ with } |i-j| = 1/2^n\}.$$

Then, using (2.10) and Chebyshev inequality, we have, for $i, j \in D_n$ such that $|i-j| = 1/2^n$,

$$\mathbb{P}(|X_i - X_j| > 2^{-n\gamma}) \le 2^{n\gamma\beta}\, \mathbb{E}[|X_i - X_j|^\beta] \le 2^{n\gamma\beta}\, 2^{-n(\alpha+1)} = 2^{n(\gamma\beta-\alpha-1)}.$$

This implies that

$$\mathbb{P}(B_n) \le 22^n\, 2^{n(\gamma\beta-\alpha-1)} = C2^{-(\alpha-\beta\gamma)n}, \;\; n \in \mathbb{N}, \tag{2.11}$$

for some $C > 0$.
Step 2. Consider $B = \limsup_n B_n = \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty B_k$; by (2.11) we know that $\sum_n \mathbb{P}(B_n)$ converges, hence $\mathbb{P}(B) = 0$ by the first Borel-Cantelli lemma. Hence $\mathbb{P}(\Omega \setminus B) = 1$ and for any $\omega \in \Omega \setminus B$ there exists $\nu = \nu(\omega)$ such that, for any $n \ge \nu$ we have

$$|X_i(\omega) - X_j(\omega)| \le \frac{1}{2^{n\gamma}}, \quad \text{for any } i, j \in D_n,\; |i-j| = 1/2^n.$$

It follows that, for any $\omega \notin B$, there exists $M = M(\omega) > 0$ such that, for any $n \ge 1$,

$$|X_i(\omega) - X_j(\omega)| \le M(\omega)\, \frac{1}{2^{n\gamma}}, \qquad \text{for any } i, j \in D_n,\; |i-j| = 1/2^n.$$

Step 3. By Lemma 2.17, for every $\omega \notin B$ the trajectory $i \mapsto X_i(\omega)$ is $\gamma-$Hölder continuous as i varies in the set D of dyadic points of T, i.e., there exists $C = C(\omega) > 0$ such that, if $i, j \in D$ then

$$|X_i(\omega) - X_j(\omega)| \le C(\omega)|i-j|^\gamma.$$

In particular, the trajectory is uniformly continuous on D.

Step 4. For every $\omega \notin B$ there exists a unique extension of $X_i(\omega)$, $i \in D$, to a trajectory $\tilde{X}_t(\omega)$ defined for $t \in T$. This extension is Hölder continuous of exponent γ.

Step 5. It remains to prove that $\tilde{X}_t = X_t$ with probability 1, $t \in T$. This clearly holds when $t \in D$; now take $t \in T$ and a sequence of dyadic rationals (t_n) such that $t_n \to t$. Using again (2.10), we show that $X_{t_n} \to X_t$ in probability; hence, passing if necessary to a subsequence, $X_{t_n} \to X_t$ almost surely. But since $X_{t_n} = \tilde{X}_{t_n}$, almost surely, it is $\tilde{X}_t = X_t$ almost surely, $t \in T$, which concludes the proof.
□

Let now $B = \{B_t,\ t \geq 0\}$ be a Brownian motion. Recall, from Remark 2.7 that for every $n \geq 1$

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

By applying Kolmogorov's continuity theorem 2.16 we have that trajectories of the Brownian motion are continuous, and Hölder continuous of exponent γ for every $\gamma < \frac{n-1}{2n}$, i.e., due to the arbitrarity of n, for every $\gamma < \frac{1}{2}$.

Exercises

Problem 2.6. Provide a complete proof of Lemma 2.17.

Problem 2.7. Given a real Brownian motion $\{B_t,\ t \geq 0\}$ on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, we can construct several other processes which share the same property.

1. (*Omogeneity*) For any fixed $s \geq 0$, the process $X_t = B_{t+s} - B_s$ is a Brownian motion with respect to the filtration $\{\mathcal{F}_{t+s},\ t \geq 0\}$.
2. (*Simmetry*) The process $Y_t = -B_t$ is a Brownian motion with respect to the initial filtration $\{\mathcal{F}_t\}$.
3. (*Scale change*) Given a real number $c > 0$, the process $W_t = cB_{t/c^2}$ is a Brownian motion with respect to the filtration $\{\mathcal{F}_{t/c^2},\ t \geq 0\}$.

Next proposition is somewhat surprising; we can phrase it by saying that if we let time flow backward, we still have a Brownian motion.

Proposition 2.18. *Given a real Brownian motion $\{B_t,\ t \geq 0\}$ on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. The stochastic process $Y = \{Y_t,\ t \geq 0\}$ defined by*

$$Y_t = \begin{cases} tB_{1/t} & t > 0 \\ 0 & t = 0 \end{cases}$$

is again a Brownian motion with respect to the completed natural filtration $\mathcal{F}^B = \{\mathcal{F}^B_t,\ t \geq 0\}$.

Proof. We verify that Proposition 2.11 applies. Property a) is obvious; property b) follows easily since Y is again a Gaussian family. Indeed, for any choice of times $0 \leq t_1 < t_2 < \cdots < t_m$, we can write $(Y_{t_1}, \ldots, Y_{t_m})$ as a linear transformation of $(B_{t_1}, \ldots, B_{t_m})$ and then apply Proposition 2.2.

Finally, if $s \leq t$:

$$\mathbb{E}[Y_t Y_s] = st\mathbb{E}[B_{1/t} B_{1/s}] = st\frac{1}{s} \wedge \frac{1}{t} = s = s \wedge t$$

and property c) follows.

Thus, it remains to prove that sample paths are almost surely continuous. This clearly holds for $t > 0$, i.e., there exists a $\mathbb{P}$-null (or negligible) set N^* such that, for every $\omega \in \Omega \setminus N^*$, the mapping $t \mapsto Y_t(\omega)$ is continuous on $(0, +\infty)$. We shall now show that there exists a negligible set N such that outside N, the mapping $t \mapsto Y_t(\omega)$ is continuous on $[0, +\infty)$.

Notice that, since $Y_t - Y_s$ has Gaussian law $\mathcal{N}(0, t-s)$,

$$\mathbb{E}[|Y_t - Y_s|^4] = 3(t-s)^2.$$

Thanks to Kolmogorov's theorem 2.16, there exists a modification $\tilde{Y}$ of Y with continuous paths; hence, for every $t \geq 0$, there exists a $\mathbb{P}$-null set N_t such that

$$Y_t(\omega) = \tilde{Y}_t(\omega) \qquad \forall \omega \in \Omega \setminus N_t.$$

Let N be the union of N^* and the sets N_τ, as τ varies in the positive rational numbers; N is a countable union of negligible sets, hence its probability is zero. For every $\omega \in \Omega \setminus N$, trajectories $t \mapsto Y_t(\omega)$ and $t \mapsto \tilde{Y}_t(\omega)$ are continuous functions on $]0, \infty[$ which coincide on the rational numbers, hence they must coincide on the whole half line $[0, \infty[$. But since $\tilde{Y}_t(\omega)$ is continuous up to 0, so it is $Y_t(\omega)$, and the proof is complete.

□

Addendum. The Borel-Cantelli lemmas

For the sake of completeness, we recall here the Borel-Cantelli lemma. Actually, it consists of two parts, sometimes called the first (second) Borel-Cantelli lemma. Both results concern the limit behaviour of a sequence of events, and as such they are often useful instruments in proving almost sure properties.

Lemma 2.19 (Borel-Cantelli).

(a) *Assume that* $\{A_n,\ n \in \mathbb{N}\}$ *is a sequence of events such that*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < +\infty.$$

Then

$$\mathbb{P}(\limsup_{n \in \mathbb{N}} A_n) = \mathbb{P}(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k) = \mathbb{P}(\{\omega\colon \omega \in A_n \textit{ for infinite indices } n\}) = 0.$$

(b) *Assume that* $\{A_n,\ n \in \mathbb{N}\}$ *is a sequence of pairwise independent events such that*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty.$$

Then

$$\mathbb{P}(\limsup_{n \in \mathbb{N}} A_n) = 1.$$

3

Sample paths of the Brownian motion

In this lecture we discuss which properties are shared, with probability 1, by sample paths of the Brownian motion. We have already seen that there are *positive* results (paths are continuous, and Hölder continuous with arbitrary parameter γ less that $\frac{1}{2}$ on bounded intervals, with probability 1). We can ask about differentiability. The answer is surprising: the sample paths are (almost surely) nowhere differentiable. If continuous, nowhere differentiable functions are considered as pathological in mathematical analysis, they seem to be the norm for trajectories of Brownian motion.

We can give a taste of the construction by considering differentiability at 0. Remember that, by Proposition 2.18, the process $Y_t = tB_{1/t}$ is a Brownian motion, with $Y_0 = 0$. Since

$$\frac{Y_t}{t} = B_{1/t}$$

the derivative of Y_t exists for $t = 0$ if and only if B_t has a limit as $t \to \infty$. But, in fact, we have

$$\sup_{n\to\infty} B_n = +\infty, \qquad \inf_{n\to\infty} B_n = -\infty \tag{3.1}$$

with probability 1; hence, almost surely, paths of the Brownian motion Y_t are not differentiable at 0. A similar argument, using Exercise 2.7(1), shows that with probability 1, Brownian paths are nondifferentiable everywhere. Of course, this can be phrased by saying that Brownian particles show at no point a velocity, thus can only be regarded as a model of the physical reality.

Maxima for Brownian paths

The material contained in this lecture can be found in many classical textbooks, as `Doob` [Do53], `Freedman` [Fr71] or `Lévy` [Lé54, Lé65].

Proposition 3.1. *For almost every ω, the function $B(\cdot, \omega)$ is monotone in no interval.*

Proof. We can simplify the situation. First, it sufficies to prove the result for an interval with rational extremes; then, thanks to Exercise 2.7, we can reduce to prove that no paths are monotone nondecreasing on $[0, 1]$.

Define, for every n, the events

$$A_n = \left\{B(\tfrac{i+1}{n}) - B(\tfrac{i}{n}) \geq 0 \quad \text{for } i = 0, \ldots, n-1\right\}$$

such that the set of nondecreasing paths is contained in $\bigcap\limits_n A_n$. But

$$\mathbb{P}(A_n) = \frac{1}{2^n} \to 0 \tag{3.2}$$

and the proof is complete.

□

Problem 3.1. Prove claim (3.2).

Let us recall that a continuous function f has a *local maximum* at t means that there is $\varepsilon > 0$ such that $f(s) \leq f(t)$ for all $s \in (t-\varepsilon, t+\varepsilon)$. The maximum is *strict* if the inequality is strict, i.e., $f(s) < f(t)$ for all $s \in (t-\varepsilon, t+\varepsilon)$.

We shall need the following properties of continuous functions.

Problem 3.2. Let f be a continuous function on $[0,1]$, monotone in no interval. Then f has a local maximum in $[0,1]$. Further, the set of local maxima is dense in $[0,1]$.

We can state the main result of this section.

Theorem 3.2. *For almost every* ω*, all local maxima of* $\{B_t(\omega),\ t \in [0,1]\}$ *are strict and constitute a dense set in* $[0,1]$.

Proof. Let us denote $M[a,b]$ the maximum value taken by the Browian motion B_t as t varies in $[a,b]$. We claim that for consecutive intervals $a < b < c < d$,

$$\mathbb{P}(M[a,b] \neq M[c,d]) = 1. \tag{3.3}$$

This implies the thesis, since every trajectory whose maximum are all strict verifies $M[a,b] \neq M[c,d]$ for every choice of $a < b < c < d$.

Let us prove (3.3). Setting

$$X = B(c) - B(b), \qquad Y = \max\{B(c+t) - B(c),\ t \in [0, d-c]\},$$

we have

$$\{M[a,b] \neq M[c,d]\} = \{X \neq M[a,b] - B(b) - Y\}.$$

Notice that $\mathcal{F}_b$, X and Y are independent, $M[a,b]$ and $B(b)$ are $\mathcal{F}_b$-measurable; therefore

$$\begin{aligned}\mathbb{P}(X \neq M[a,b] - B(b) - Y) &= \int \mathbb{P}(X \neq x, M[a,b] - B(b) - Y = x)\,\mathrm{d}x \\ &= \int \mathbb{P}(X \neq x)\mathbb{P}(M[a,b] - B(b) - Y = x)\,\mathrm{d}x = 1\end{aligned}$$

since X has a gaussian distribution and $\mathbb{P}(X \neq x) = 1$.

□

The zero set of Brownian motion

We have seen the link between behaviour of the Brownian motion for t near zero and the asymptotic behaviour as $t \to \infty$. Before we proceed with the study of the set of zeroes for the Brownian motion, let us exploit some more this link. First, we propose the following exercise.

Problem 3.3.

1. Prove claim (3.1):
$$\sup_{n\to\infty} B_n = +\infty, \qquad \inf_{n\to\infty} B_n = -\infty.$$
2. Prove that law of large numbers applies:
$$\lim_{t\to\infty} \frac{B_t}{t} = 0 \quad \text{a.s.}$$

Hint: write $(\sup_n |B_n| < \infty) = \bigcup_{k=1}^{\infty} (\sup_n |B_n| < k)$ so that the probability of the event on the left is bounded by the series of the probabilities on the right. But

$$\mathbb{P}(\sup_n |B_n| < k) \le \lim_n \mathbb{P}(|B_n| < k) = \lim_n \mathbb{P}(|B_1| < k/n^{1/2}) \dots$$

Let us consider the following consequence of (3.1). Writing $W_t = tB_{1/t}$ as $t \to 0$, for almost every path there exist two sequences $\{t_n\}$ and $\{s_n\}$ such that $t_{n+1} < s_{n+1} < t_n < s_n$ and $W(t_n) > 0$, $W(s_n) < 0$. Since the Brownian paths are almost surely continuous, there exists a sequence $\{\tau_n\}$ with $t_n < \tau_n < s_n$ and $W(\tau_n) = 0$. We have thus proved that

Corollary 3.3. *With probability one, Bronwian motion returns to the origin infinitely often.*

Let us denote
$$Z(\omega) = \{t \ge 0 \,:\, B_t(\omega) = 0\}$$
and remark that Z is a random variable (taking values in $\mathcal{B}(\mathbb{R}_+)$). Notice that $0 \in Z(\omega)$ for every ω, and previous corollary shows that, with probability one, 0 is an accumulation point for $Z(\omega)$.

Theorem 3.4. *With probability one, the set $Z(\omega)$ is closed, of Lebesgue measure 0, perfect, nowhere dense in $\mathbb{R}_+$.*

Proof. For simplicity, we suppress the phrase “with probability one” in the following. Since $Z(\omega) = B(\omega,\cdot)^{-1}(\mathbb{R}_+)$ is the inverse of a closed set through a continuous mapping, it follows that it is closed. Let λ denote Lebesgue measure on $\mathbb{R}_+$; using Fubini's theorem

$$\mathbb{E}[\lambda(Z(\omega))] = (\lambda \times \mathbb{P})\{(t,\omega) \,:\, B_t(\omega) = 0\} = \int_0^\infty \mathbb{P}(B_t(\omega) = 0)\,\mathrm{d}t = 0$$

since B_t takes every given value with probability zero.

Let I be an interval of the positive real line; since $t \mapsto B_t(\omega)$ is continuous on I, if the set $Z(\omega) \cap I$ were dense in I then necessarily $B_t \equiv 0$ on I, which is absurd, since (otherwise B_t would be differentiable). Hence $Z(\omega)$ is not dense in any interval I.

It remains to prove that $Z(\omega)$ is dense in itself, that is, for any point $t \in Z(\omega)$ there exists a sequence $t_n \in Z(\omega)$, $t_n \neq t$, such that $t_n \to t$. We already know that it is true for $t = 0$.

For a general $x > 0$, let us define

$$\tau_x = \inf\{t \geq x \,:\, B_t(\omega) = 0\};$$

this means that τ_x is the first zero after time x. Notice that

$$\{\tau_x \leq s\} \in \sigma(B_u \,:\, u \leq s) = \mathcal{F}^B_s; \tag{3.4}$$

a random variable like τ_x, non negative, and verifying (3.4) is called a stopping time. We shall return on this in next lectures; for the moment we claim that

$$W_s = B_{\tau_x + s} - B_{\tau_x}$$

is again a Brownian motion (by the way, notice that $W_0 = 0$ and W has continuous paths by construction, so it remains to prove that its finite dimensional distributions agree with those of a Brownian motion).

Now, let $t = \tau_x(\omega)$ be the first zero in $Z(\omega)$ greater or equal to x. Since $W(\omega)$ is a trajectory of a Brownian motion, there is a sequence of positive times $s_n \to 0$ such that $W_{s_n}(\omega) = 0$, and so $B_{t+s_n}(\omega) = 0$, which means that $\{t + s_n\}$ is a sequence in $Z(\omega)$ converging to t, as required.

□

3.1 Regolarity of sample paths

Of particular interest is the study of *variation* of sample paths, since if it were the case, it would have been possible to define a (Lebesgue) integral with respect to them. However, we shall see that this is not the case.

In Section 3.2 we provide the beautiful analysis of P. Lévy about regularity of the Brownian trajectories. We will present the *law of iterated logarithm*, and we prove that the Hölder exponent $\gamma < \frac{1}{2}$ cannot be improved.

3.1.1 Quadratic variation

Let $f : \mathbb{R}_+ \to \mathbb{R}$ be a given function; for any interval $[s, t]$ and partition $\pi = t_0 = s < t_1 < \cdots < t_n = t$ we set

$$qV(f, \pi, [s, t]) = \sum_{i=1}^{n} \Big| f(t_i) - f(t_{i-1}) \Big|^2.$$

We denote $\|\pi\| = \max\{t_i - t_{i-1}\}$ the norm of the partition; then we call f a *finite quadratic variation* function on $[s, t]$ if the following limit exists

$$qV(f, [s, t]) = \lim_{\|\pi\| \to 0} qV(f, \pi, [s, t]) < \infty.$$

We begin presenting a link between quadratic variation and Hölder continuity.

Proposition 3.5. *Fix an interval $[s,t]$; then every function $f : [s,t] \to \mathbb{R}$ Hölder continuous of parameter $\alpha > 1/2$ verifies $qV(f,[s,t]) = 0$.*

Proof. Take a partition π of $[s,t]$; then

$$\begin{aligned}\sum_{i=1}^{n} \Big| f(t_i) - f(t_{i-1}) \Big|^2 &= \sum_{i=1}^{n} (t_i - t_{i-1})^{2\alpha} \left| \frac{f(t_i) - f(t_{i-1})}{(t_i - t_{i-1})^{\alpha}} \right|^2 \\ &\le \sum_{i=1}^{n} C^2 (t_i - t_{i-1})^{2\alpha} \\ &\le C^2 \max |t_i - t_{i-1}|^{2\alpha - 1} \underbrace{\sum_{i=1}^{n} |t_i - t_{i-1}|}_{=t-s}\end{aligned}$$

hence

$$qV(f, \pi, [s,t]) \le C^2 (t-s) \|\pi\|^{2\alpha-1}$$

and, passing to the limit as $\|\pi\| \to 0$, the thesis follows.

□

The next step is the study of quadratic variations for the trajectories $t \mapsto B_t(\omega)$ of a Brownian motion on an interval $[s,t]$. If we set $A_{[s,t]}(\omega) = qV(B.(\omega),[s,t])$, it results to be a random variable $\mathcal{F}_t$-measurable. Surprisingly enough, it happens that this random variable is constant with probability 1, and it holds $A_{[s,t]} = t - s$.

In order to simplify the proof, we give a weaker result, taking convergence in L^2, from where convergence in probability (and convergence almost surely for a subsequence) follows.

Proposition 3.6. *Let $B = \{B_\tau,\ \tau \ge 0\}$ be a Brownian motion, $[s,t]$ be a given interval and π a partition of $[s,t]$. Then*

$$\lim_{\|\pi\| \to 0} qV(B.(\omega), \pi, [s,t]) = t - s \qquad \text{in } L^2.$$

Proof. Since $t - s = \sum_k t_k - t_{k-1}$ we have

$$qV(B., \pi, [s,t]) - (t-s) = \sum_k \Big[(B_{t_k} - B_{t_{k-1}})^2 - (t_k - t_{k-1}) \Big]$$

and

$$\begin{aligned}&\mathbb{E}\Big| qV(B., \pi, [s,t]) - (t-s) \Big|^2 \\ &\quad = \sum_{j,k} \mathbb{E} \Big\{ \Big[(B_{t_k} - B_{t_{k-1}})^2 - (t_k - t_{k-1}) \Big] \Big[(B_{t_j} - B_{t_{j-1}})^2 - (t_j - t_{j-1}) \Big] \Big\} \\ &\quad = \sum_k \mathbb{E} \Big[(B_{t_k} - B_{t_{k-1}})^2 - (t_k - t_{k-1}) \Big]^2\end{aligned}$$

since random variables $(B_{t_k} - B_{t_{k-1}})^2 - (t_k - t_{k-1})$ and $(B_{t_j} - B_{t_{j-1}})^2 - (t_j - t_{j-1})$, for $j \neq k$, are independent with mean 0, the expected value of their product is 0;

$$= \sum_k (t_k - t_{k-1})^2 \mathbb{E}\left[\left(\frac{B_{t_k} - B_{t_{k-1}}}{\sqrt{t_k - t_{k-1}}}\right)^2 - 1\right]^2$$
$$= c \sum_k (t_k - t_{k-1})^2$$

where, for each k, $\frac{B_{t_k} - B_{t_{k-1}}}{\sqrt{t_k - t_{k-1}}}$ is a standard Gaussian distribution, hence

$$c = \mathbb{E}\left[\left(\frac{B_{t_k} - B_{t_{k-1}}}{\sqrt{t_k - t_{k-1}}}\right)^2 - 1\right]^2 = \int_{\mathbb{R}} (x^2 - 1)^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 2.$$

Then we have

$$\mathbb{E}\Big|qV(B., \pi, [s,t]) - (t-s)\Big|^2 \leq 2\|\pi\|(t-s) \to 0$$

as $\|\pi\| \to 0$.

□

3.1.2 Bounded variation

Let $f : \mathbb{R}_+ \to \mathbb{R}$ be a given function; for any interval $[s,t]$ and partition $\pi = t_0 = s < t_1 < \cdots < t_n = t$ we set

$$V(f, \pi, [s,t]) = \sum_{i=1}^n \Big| f(t_i) - f(t_{i-1}) \Big|.$$

We call f a function with *bounded variation* on $[s,t]$ if the following limit exists

$$V(f, [s,t]) = \lim_{\|\pi\| \to 0} V(f, \pi, [s,t]) < \infty.$$

We call f a function with *total bounded variation* if there exists $V(f, [0,\infty)) < \infty$.

We also set

$$V^+(f, \pi, [s,t]) = \sum_{i=1}^n \Big[f(t_i) - f(t_{i-1}) \Big]^+, \qquad V^-(f, \pi, [s,t]) = \sum_{i=1}^n \Big[f(t_i) - f(t_{i-1}) \Big]^-,$$

and similarly

$$V^+(f, [s,t]) = \lim_{\|\pi\| \to 0} V^+(f, \pi, [s,t]), \qquad V^-(f, [s,t]) = \lim_{\|\pi\| \to 0} V^-(f, \pi, [s,t]).$$

Take a function f with bounded variation, with $f(0) = 0$; we set $V^+(x) = V^+(f, [0,x])$ and $V^-(x) = V^-(f, [0,x])$, $V(x) = V(f, [0,x])$. Then we can write f as a difference between two increasing functions

$$f(x) = V^+(x) - V^-(x)$$

while the variation of f is given by $V(x) = V^+(x) + V^-(x)$.

Proposition 3.7. *Trajectories of a Brownian motion are, with probability 1, of unbounded variation on every interval.*

Proof. Fix a partition π of the interval $[a,b]$; the quadratic variation of $B.(\omega)$ is

$$qV(B., \pi, [a,b]) = \sum_k (B_{t_k} - B_{t_{k-1}})^2$$
$$\leq \max_k |B_{t_k} - B_{t_{k-1}}| \sum_k |B_{t_k} - B_{t_{k-1}}|$$

Since trajecotries are continuous we have

$$\lim_{\|\pi\|\to 0} \max_k |B_{t_k} - B_{t_{k-1}}| = 0.$$

Assume that there exists an event $\Omega_{[a,b]}$, with positive probability, such that trajectories $B.(\omega)$ have bounded variation for $\omega \in \Omega_{[a,b]}$. Then for such ω, quadratic variation of the trajectory should be 0, having a contradiction with Proposition 3.6.

We consider the sequence of intervals $[a,b]$ with rational extreme points. Define the set

$$\bar{\Omega} = \bigcup_{0\leq a\leq b\in\mathbb{Q}} \Omega_{[a,b]}$$

it holds that $\bar{\Omega} \in \mathcal{F}$ and $\mathbb{P}(\bar{\Omega}) = 1$. Therefore, for each $\omega \notin \bar{\Omega}$, $B_t(\omega)$ has unbounded variation on every nontrivial interval $[s,t]$ with real valued extreme points (each such interval contains an interval $[a,b]$ with rational end points, and the variation increases with the interval). This concludes the proof. □

Corollary 3.8. *The above results, together with Proposition 3.5, imply that there exists a set of probability 1 such that the corresponding trajectories $t \mapsto B_t(\omega)$ are nowhere α-Hölder continuous for arbitrary exponent $\alpha > \frac{1}{2}$.*

3.2 The law of the iterated logarithm

We present in this section the fine results about regularity of Brownian sample paths. In the proceeding of the course we will not make use of them, and our presentation will try to avoid much technicalities and to give the spirits of the results.

The first result is the law of the iterated logarithm, which describes the oscillations of the Brownian sample paths near zero and infinity.

Theorem 3.9. *For ω in a set of probability 1*

$$\max_{t\to 0^+} \lim \frac{B_t(\omega)}{\sqrt{2t \log\log(1/t)}} = 1. \tag{3.5}$$

Before we proceed with the proof, we shall notice the following consequences of the theorem. By simmetry,

$$\min_{t \to 0^+} \lim \frac{B_t(\omega)}{\sqrt{2t \log\log(1/t)}} = -1,$$

and time inversion provides the following

$$\max_{t \to +\infty} \lim \frac{B_t(\omega)}{\sqrt{2t \log\log t}} = 1, \qquad \min_{t \to +\infty} \lim \frac{B_t(\omega)}{\sqrt{2t \log\log t}} = -1.$$

For almost every trajectory, we can construct two sequences $\{t_n\}$ and $\{s_n\}$, increasing and diverging to $+\infty$, with $s_n < t_n < s_{n+1}$, such that

$$B(t_n) \geq \sqrt{t_n \log\log t_n}, \qquad B(s_n) \leq -\sqrt{s_n \log\log s_n}.$$

Then we see that the oscillations of every path increase more and more; further, due to continuity, we see that the trajectories touch each real value infinitely often.

In preparation for the proof, we propose the following exercises.

Problem 3.4. Compute the tail of the normal distribution: for every $x > 0$ we have

$$\frac{x}{1+x^2} \mathrm{e}^{-x^2/2} \leq \int_x^\infty \mathrm{e}^{-y^2/2} \, \mathrm{d}y \leq \frac{1}{x} \mathrm{e}^{-x^2/2}. \tag{3.6}$$

Problem 3.5. A maximal inequality for the Brownian motion. Consider first a sequence of Gaussian random variable, with zero mean, $\{X_n,\ n \in \mathbb{N}\}$ and the partial sums $S_n = X_1 + \cdots + X_k$. Show that for positive x it holds

$$\mathbb{P}(\max_{k \leq n} S_k \geq x) \leq 2\mathbb{P}(S_n \geq x).$$

Show that this implies

$$\mathbb{P}(\sup_{s \leq \tau \leq t} (B_\tau - B_s) > x) \leq 2\mathbb{P}((B_t - B_s) > x). \tag{3.7}$$

Hint: Brownian motion has continuous paths, hence the probability on the left can be studied for τ in the set of dyadic rational numbers; now, the increments are independent and Gaussian distributed...

Proof (Proof of Theorem 3.9). We shall prove first that

$$\max_{t \to 0^+} \lim \frac{B_t(\omega)}{\sqrt{2t \log\log(1/t)}} \leq 1. \tag{3.8}$$

With the notation $h(t) = \sqrt{\log\log(1/t)}$ and fixed $\delta \in (0,1)$, we choose $\theta \in (0,1)$ such that $\lambda = \theta(1+\delta)^2 > 1$. Define a sequence of times $t_n = \theta^n$ decreasing to 0 and consider the events

$$A_n = \left\{ \max_{t \in [t_{n+1}, t_n]} B_t > (1+\delta) h(t) \right\}.$$

If we prove that $\mathbb{P}(A_n)$ is a converging sequence, we have from Borel-Cantelli lemma that for ω outside a set of measure 0, for every $t \in [t_{n+1}, t_n]$

$$\max_{t\to 0^+}\lim \frac{B_t(\omega)}{h(t)} \le (1+\delta)$$

and from the arbitrariety of δ it follows (3.8).

To estimate the probability of A_n we notice the following inclusion

$$A_n \subset \left\{ \max_{t\in[0,t_n]} B_t > (1+\delta)h(t_{n+1}) \right\}.$$

Using (3.7) we have the following estimate on the probability of this event:

$$\mathbb{P}(\{\max_{t\in[0,t_n]} B_t > (1+\delta)h(t_{n+1})\}) \le 2\mathbb{P}(B_{t_n} > (1+\delta)h(t_{n+1}))$$

$$\le 2\mathbb{P}(\frac{B_{t_n}}{\sqrt{t_n}} > \frac{(1+\delta)h(t_{n+1})}{\sqrt{t_n}})$$

the random variable on the left hand side has a Gaussian standard distribution, so we can use (3.6): set for simplicity $x_n = \frac{(1+\delta)h(t_{n+1})}{\sqrt{t_n}}$ and we get

$$\mathbb{P}(\{\max_{t\in[0,t_n]} B_t > (1+\delta)h(t_{n+1})\}) \le \sqrt{2/\pi}\frac{1}{x_n}\mathrm{e}^{-x_n^2/2}.$$

Compute x_n^2: using the above notation we have $x_n = 2\lambda\log[(n+1)\log(1/\theta)]$ hence

$$\mathbb{P}(A_n) \le C\frac{1}{(n+1)^\lambda}.$$

We prove next the converse inequality. The proof is based on an application of the second part of the Borel-Cantelli lemma. With the same notation of the first part, choose first ε and $\theta \in (0,1)$; we define the events

$$A'_n = \{B_{t_n} - B_{t_{n+1}} \ge (1-\varepsilon)h(t_n)\}.$$

Let us prove that $\sum_n Pr(A'_n)$ diverges. The idea is to use the left inequality of (3.6), which yields, for any x, the estimate

$$\mathbb{P}(\frac{B_{t_n} - B_{t_{n+1}}}{t_n - t_{n+1}} > x) \ge \frac{1}{2\sqrt{2\pi}}\frac{1}{x}\mathrm{e}^{-x^2/2}$$

so, taking $x = x_n = (1-\varepsilon)\frac{h(t_n)}{\sqrt{t_n - t_{n+1}}}$, we have

$$\mathbb{P}(\frac{B_{t_n} - B_{t_{n+1}}}{t_n - t_{n+1}} > x) \ge C\frac{1}{n^{(1-\varepsilon)^2/1-\theta}\sqrt{\log(n)}}.$$

If we take $\theta < 1-(1-\varepsilon)^2$ this term is the general term of a diverging series. Then by Borel-Cantelli lemma we have

$$\mathbb{P}(B_{t_n} - B_{t_{n+1}} \ge (1-\varepsilon)h(t_n) \text{ infinitely often}) = 1.$$

But since from (3.8) we have proved that $B_{t_{n+1}} \ge -(1+\varepsilon)h_{t_{n+1}}$ definitely as $n\to\infty$, we get for infinite indices n

$$B_{t_n} \geq (1-\varepsilon)h(t_n) - (1+\varepsilon)h_{t_{n+1}} = h(t_n)\left(1-\varepsilon-(1+\varepsilon)\frac{h_{t_{n+1}}}{h_{t_n}}\right)$$

Using the limit $\lim_{n\to\infty} \frac{h_{\theta^{n+1}}}{h_{\theta^n}} = \sqrt{\theta}$ and choosing $\varepsilon + (1+\varepsilon)\sqrt{\theta} < \delta$ we have

$$B_{t_n} \geq (1-\delta)h(t_n) \quad \text{infinitely often}$$

□

3.2.1 Modulus of continuity

We have seen that Brownian sample paths are α-Hölder continuous for every $\alpha < \frac{1}{2}$ and, by Corollary 3.8, they are nowhere α-Hölder continuous for every $\alpha > \frac{1}{2}$. This section will treat the limit case $\alpha = \frac{1}{2}$.

The *modulus of continuity* of a real function $f : I \to \mathbb{R}$, where $I \subset \mathbb{R}$ is a given interval, is defined as the function

$$w(\delta) = \sup_{x,y\in I,\ |x-y|\leq\delta} |f(x) - f(y)|.$$

Of course, the modulus of continuity of the Brownian sample paths is bounded above by $C\delta^\alpha$ for any $\alpha > \frac{1}{2}$, since the trajectories are nowehere α-Hölder continuous for such α. On the other hand, it shall be larger that $\sqrt{\log\log(1/\delta)}$ by the law of the iterated logarithm. The exact modulus of continuity for the Brownian sample paths is determined from the following theorem. We omit its proof but refer the interested reader, for instance, to [KS88, Theorem 2.9.5].

Theorem 3.10 (Lévy). *For every $T > 0$, on the interval $I = [0, T]$ we have*

$$\mathbb{P}\left\{\lim_{\delta\to 0^+} \sup_{s<t\in I, |t-s|\leq\delta} \frac{|B_t - B_s|}{(2\delta\log(1/\delta))^{1/2}} = 1\right\} = 1. \tag{3.9}$$

The above theorem states that, given $w(\cdot,\omega)$ the modulus of continuity of the Brownian path $B_t(\omega)$ as $t \in [0, T]$, then a.s.

$$\lim_{\delta\to 0^+} \frac{w(\delta,\omega)}{(2\delta\log(1/\delta))^{1/2}} = 1.$$

This result defines completely the regularity of Brownian trajectories, and we state it in next corollary.

Corollary 3.11. *For ω outside a set of zero measure, every Brownian path is never α-Hölder continuous with exponent $\alpha \geq 1/2$ in a time interval $I \subset \mathbb{R}_+$ with nonempty internal part.*

Proof. Take an interval $J = [q, r]$, with rationals $q < r \in \mathbb{Q}_+$,

$$\lim_{\delta\to 0^+} \sup_{s<t\in J, |t-s|\leq\delta} \frac{|B_t - B_s|}{(2\delta\log(1/\delta))^{1/2}} = 1 \quad \text{a.s.} \tag{3.10}$$

Let us denote $N_{q,r}$ the set of trajectories where (3.10) is not verified, and we know that $N_{q,r}$ is a negligible set. Letting $N = \bigcup_{q<r\in\mathbb{Q}_+} N_{q,r}$, N is again a negligible set.

Now, assume I is an interval with nonempty internal part such that the Brownian path $B(\cdot,\omega)$ is α-Hölder continuous on I with $\alpha \geq 1/2$. Necessarily, there exists an interval $J = [q, r]$ with rational extremes such that $J \subset I$, hence it must be $\omega \in N$, which proves the claim.

□

3.3 The canonical space

Let us consider a real, standard Brownian motion $\{B_t,\ t \in [0,1]\}$, defined on the space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. Having chosen a finite interval helps to simplify some technical details and may be removed if necessary. Without loss of generality, we can assume that $\{B_t\}$ has continuous paths (by taking a modification of the process, if necessary). We consider the mapping $\xi : \omega \to \{t \mapsto B_t(\omega)\}$ which associates to each ω the trajectory, as an element in the space $\mathcal{C} = C_0([0,1])$ of real valued, continuous functions vanishing at 0. $\mathcal{C}$, endowed with the distance

$$d(\omega, \eta) = \sup_{t \in [0,1]} |\omega(t) - \eta(t)|$$

is a complete, separable metric space. Let $\mathcal{G}'$ the Borel σ-field generated by the open sets of $\mathcal{C}$.

Define on the space $\mathcal{C}$, the coordinate mapping functions $X(t, \cdot)(\omega) = \omega(t)$ for every $t \in [0,1]$ and $\omega \in \mathcal{C}$. We consider the σ-algebra $\mathcal{G}$ generated by the finite-dimensional cylindrical sets

$$\{\omega \in \mathcal{C} \,:\, \omega(t_1) \in A_1, \dots, \omega(t_n) \in A_n\},$$

with $t_1, \dots, t_n \in [0,1]$. It has the property that all the coordinate mapping functions $X(t)$ are $\mathcal{G}$-measurable, and $\mathcal{G}$ is the smallset σ-field with this property.

Lemma 3.12. $\mathcal{G} = \mathcal{G}'$.

Proof. The inclusion $\mathcal{G} \subset \mathcal{G}'$ is obvious; let us check the other one. The space $\mathcal{C}$ is separable, hence

$$d(\omega, \eta) = \sup_{t \in \mathbb{Q} \cap [0,1]} |\omega(t) - \eta(t)|;$$

this implies that the spheres $U = \{\omega \in \mathcal{C} \,:\, |\omega_t - \gamma_t| \le \varepsilon\ \forall t \in [0,1]\}$ are in $\mathcal{G}$. But every open set $A \in \mathcal{G}'$ can be written as a countable union of spheres, hence it is in $\mathcal{G}$.
□

As a corollary we have the following result, whose prove we let to the reader.

Problem 3.6. The mapping $\xi : (\Omega, \mathcal{F}) \to (\mathcal{C}, \mathcal{G})$ is measurable, i.e., ξ is a random variable taking values in $\mathcal{C}$.

Thanks to the above result, on the space $(\mathcal{C}, \mathcal{G})$ we can define the image measure $\mathbb{P} = \mu \circ \mathcal{B}^{-1}$, that is called the *Wiener measure*. Define, on the probability space $(\mathcal{C}, \mathcal{G}, \mathbb{P})$, the coordinate mapping process $X(t, \omega) = \omega(t)$ for every $t \in [0,1]$ and $\omega \in \mathcal{C}$. X is equivalent to B and is often called the *canonical* Brownian motion.

Addendum. A second construction of Brownian motion

Here we propose an explicit way to construct a real Brownian motion. This construction is due to Lévy and Ciesielski (see, for instance, `Karatzas and Shreve` [KS88] or `Zabczyk` [Za04]). It does not require to appeal to the Kolmogorov existence theorem.

We proceed in some steps and ask to complete the details as exercise.

I step. We first define the Haar functions $h_k : [0,1] \to \mathbb{R}$, $k \in \mathbb{N}$,

$$h_0(t) = 1, \quad t \in [0,1],$$
$$h_1(t) = \begin{cases} 1 & \text{for } t \in [0,1/2], \\ -1 & \text{for } t \in (1/2,1] \end{cases}$$

and, if $n \in \mathbb{N}$ and $2^n \le k < 2^{n+1}$,

$$h_k(t) = \begin{cases} 2^{n/2} & \text{for } t \in \left[\frac{k-2^n}{2^n}, \frac{k-2^n+1/2}{2^n}\right], \\ -2^{n/2} & \text{for } t \in \left(\frac{k-2^n+1/2}{2^n}, \frac{k-2^n+1}{2^n}\right], \\ 0 & \text{otherwise.} \end{cases}$$

The sequence (h_k) forms an orthonormal and complete system in $L^2(0,1)$.

II step. Define the Schauder functions $s_k : [0,1] \to \mathbb{R}$, $k \in \mathbb{N}$,

$$s_k(t) = \int_0^t h_k(s)ds, \qquad k \in \mathbb{N}.$$

Note that all s_k are non-negative; moreover we have

$$\|s_k\|_\infty = \sup_{t \in [0,1]} |s_k(t)| = 2^{n/2} \frac{1}{2^{n+1}} = 2^{-n/2-1}, \quad 2^n \le k < 2^{n+1}. \tag{3.11}$$

III step. Let us consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which it is well defined a sequence $\{X_n,\ n \in \mathbb{N}\}$ of independent real Gaussian random variables such that the law of each X_n is $\mathcal{N}(0,1)$.

For instance, we can take $\Omega = \prod_{i \in \mathbb{N}} (\mathbb{R})_i$ the product of infinite copies of the real line $\mathbb{R}_i = \mathbb{R}$, $i \in \mathbb{N}$, endowed with the product σ-algebra $\mathcal{F} = \mathcal{B}(\mathbb{R}) \times \ldots \times \mathcal{B}(\mathbb{R}) \times \ldots$. Moreover, we take as measure $\mathbb{P}$ the infinite product of Gaussian laws $\mathcal{N}(0,1)$. We define $X_n(\omega) = \omega_n$, for any $\omega \in \Omega$, $n \in \mathbb{N}$. It is clear that $(\Omega, \mathcal{F}, \mathbb{P}, \{X_n\})$ satisfies our assertion.

IV step. Consider the previous sequence of independent Gaussian random variables (X_n). One has that

$$|X_k(\omega)| \text{ is } \mathrm{O}(\sqrt{\log \mathrm{k}}) \text{ for } k \to \infty, \text{ a.s.}$$

Indeed, we have $\mathbb{P}(|X_k| > 4\sqrt{\log k}) \le c\mathrm{e}^{-4\log k} = \frac{c}{k^4}$, $k \in \mathbb{N}$, and, applying the Borel-Cantelli lemma, for any ω a.s., there exists $k_0 = k_0(\omega)$ such that, if $k \ge k_0$ then $|X_k(\omega)| \le 4\sqrt{\log k}$.

V step. Define

$$B_t(\omega) = \sum_{k \ge 0} X_k(\omega) s_k(t), \qquad \omega \in \Omega,\ t \in [0,1].$$

One verifies that $\{B_t,\ t \in [0,1]\}$ is a Brownian motion (note that $B_0(\omega) = 0$ for any ω).

VI step. We extend the previous definition of B_t when $t \in [0, \infty)$.

Consider a sequence of infinite copies of the same probability space:

$$(\Omega_k, \mathcal{F}_k, \mathbb{P}_k) = (\Omega, \mathcal{F}, \mathbb{P}), \qquad k \geq 0$$

where $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability space on which is defined B_t, $t \in [0,1]$. Let us introduce the product probability space

$$\begin{cases} \dot{\Omega} = \prod_{k \geq 0} \Omega_k, \\ \dot{\mathcal{F}} = \prod_{k \geq 0} \mathcal{F}_k, \\ \dot{\mathbb{P}} = \prod_{k \geq 0} \mathbb{P}_k. \end{cases}$$

Let us denote by B_t^k the Brownian motion on $(\Omega_k, \mathcal{F}_k, \mathbb{P}_k)$, $t \in [0,1]$. Define in a recursive way:

$$\begin{cases} \dot{B}_t = B_t^1, & t \in [0,1], \\ \dot{B}_t = \dot{B}_n + B_{t-n}^{n+1}, & t \in [n, n+1],\ n \geq 1. \end{cases}$$

Hence, for $\omega = (\omega_k) \in \dot{\Omega}$,

$$\begin{aligned} \dot{B}_t(\omega) &= B_t^1(\omega_1), \qquad t \in [0,1], \\ \dot{B}_t(\omega) &= B_1^1(\omega_1) + B_{t-1}^2(\omega_2), \qquad t \in [1,2], \\ &\dots \\ \dot{B}_t(\omega) &= \sum_{k=1}^{n} B_1^k(\omega_k) \ + B_{t-n}^{n+1}(\omega_{n+1}), \qquad t \in [n, n+1]. \end{aligned}$$

One checks that $\{\dot{B}_t,\ t \geq 0\}$ is a real Brownian motion on $(\dot{\Omega}, \dot{\mathcal{F}}, \dot{\mathbb{P}})$.

4

Martingales

In this lecture we introduce in Section 4.1 a new class of random variables, namely the *stopping times*, and in Section 4.2 the class of stochastic processes named *martingales*. Stopping times and martingales have a clear interpretation in gambling and we will adopt sometimes this terminology; this will allow us to explain the ideas which motivate the mathematical definitions. Notice that the french *martingale* denotes the strategy of doubling the bet after a loss.

Both these objects will play a fundamental rôle in the following, starting from the end of this lecture where we show that a Brownian motion is a continuous time martingale and we prove some properties of Brownian motion related to its filtration.

Stopping times enter in a natural way in the theory of Markovian processes, where they define the important class of *strong Markov* processes. It is worth mentioning since now that definition and properties of a stopping time are independent from the probability measure and are rather set-theoretical concepts.

The ideas in this lecture have a central position in all textbooks on the stochastic processes and stochastic integration; for a comprehensive introduction, which contains also the important *limit theorems* for martingales, we refer to `Ethier and Kurtz` [EK86] or the other books listed in the bibliography.

4.1 Stopping times

Suppose that a gambler decide to continue playing roulette against the house until a certain value happens. Let τ be the time at which he will stop playing: the event $\{\tau \le t\}$ will depend only on the observations of the gambler up to time t. This is a good example of what we call a *stopping time*, that is a random variable such that "its happening does not depend on the future".

Definition 4.1. *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in T\}, \mathbb{P})$ be a filtered probability space. A random variable $\tau : \Omega \to T \cup \{+\infty\}$ is called a* stopping time *if for each $t \in T$ the event*

$$\{\tau \le t\} = \{\omega \in \Omega \;:\; \tau(\omega) \le t\}$$

belongs to $\mathcal{F}_t$. We further set

$$\mathcal{F}_\tau = \{A \in \mathcal{F} \;:\; A \cap \{\tau \le t\} \in \mathcal{F}_t \;\text{ for each } t \in T\}.$$

We list below some consequences of the general definition. Recall that T denotes a general subset of $\mathbb{R}_+$ and, in particular T could be $\mathbb{N}$.

1. A stopping time can be deterministic if $\tau(\omega) = t$, for $\omega \in \Omega$ (but also $\tau(\omega) = +\infty$ is a stopping time).
2. Actually, the measure $\mathbb{P}$ plays no rôle here; the only important thing is the filtration $\{\mathcal{F}_t,\ t \in T\}$.
3. It is not difficult to check that $\mathcal{F}_\tau$ is a σ-field (for instance, if $A \in \mathcal{F}_\tau$ then $A^c \in \mathcal{F}_\tau$ since
$$A^c \cap \{\tau \le t\} = \{\tau \le t\} \setminus (A \cap \{\tau \le t\}) \in \mathcal{F}_t).$$
Heuristically, $\mathcal{F}_\tau$ stands for the collection of events for which, at the random time τ, we can decide whether they have occurred or not.
4. Note that also $\{\tau < t\} \in \mathcal{F}_t$, $t \in T$, since
$$\{\tau < t\} = \bigcup_{n \ge 1} \{\tau \le t - \frac{1}{n}\}.$$

Remark 4.2. In case the filtration $\{\mathcal{F}_t,\ t \in T\}$ does not satisfies the standard assumptions (in particular if it is not right-continuous), it is natural to introduce the following definition of optional time, see for instance `Karatzas and Shreve` [KS88].

A random variable $\tau : \Omega \to T \cup \{+\infty\}$ is an *optional time* if $\{\tau < t\} \in \mathcal{F}_t$, $t \in T$. Clearly any stopping time is an optional time. Viceversa, in general an optional time τ is a stopping time only with respect to the filtration $\{\mathcal{F}_{t+}\}$, i.e., $\{\tau \le t\} = \bigcap_{\epsilon > 0} \{\tau < t + \epsilon\} \in \mathcal{F}_{t+}$.

We proceed by considering examples and properties of stopping times.

Problem 4.1. We are given two stopping times σ and τ, defined on the same filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. Then

1. τ is $\mathcal{F}_\tau$-measurable.
2. $\sigma \vee \tau$ (i.e., the maximum between σ and τ), $\sigma \wedge \tau$ (i.e., the minimum between σ and τ) are two stopping times.
3. If $\sigma \le \tau$, then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.

Hint: We only note that 1. holds since $\{\tau \le s\} \cap \{\tau \le t\} \in \mathcal{F}_{t \wedge s}$, for $t, s \in T$, and so $\{\tau \le s\} \in \mathcal{F}_\tau$.

We have seen that the minimum of two stopping times is again a stopping time. This can be extended to the infimum of a sequence of stopping times $\{\tau_n\}$, i.e.,
$$\tau = \inf_{n \in \mathbb{N}} \{\tau_n\}.$$
Indeed, we have
$$\{\tau < t\} = \bigcup_n \{\tau_n < t\} \in \mathcal{F}_t$$
and now $\{\tau \le t\} = \bigcap_{\epsilon > 0} \{\tau < t + \epsilon\} \in \mathcal{F}_{t+} = \mathcal{F}_t$. Thus, if the filstration satisfies the standard assumptions, τ is a stopping time.

We assume next that τ is a stopping time, and ask wether it is possible to approximate it in a shrewd way. Then we set

$$\tau_n = 2^{-n}\lfloor 2^n \tau + 1 \rfloor$$

where $\lfloor x \rfloor$ stands for the integer part of x. For instance, taking $\tau(\omega) = 3/7$, then

$$\tau_1(\omega) = \tau_2(\omega) = \tau_3(\omega) = 1/2, \quad \tau_4(\omega) = \tau_5(\omega) = \tau_6(\omega) = 7/16, \quad \ldots$$

$\{\tau_n\}$ is a sequence of random times taking countable many values (actually, taking values in $2^{-n}\mathbb{N}$); moreover $\tau_n > \tau$, $\tau_n \downarrow \tau$ and, since

$$\{\tau_n \le \frac{k}{2^n}\} = \{\tau < \frac{k}{2^n}\} \in \mathcal{F}_{k/2^n},$$

τ_n is a stopping time. We summarize the above construction in the following statement.

Corollary 4.3. *For any stopping time τ, there exists a sequence of stopping times $\{\tau_n\}$ with respect to the same filtration, such that $\tau_n \downarrow \tau$.*

Also the sum of two stopping times $\sigma + \tau$ is a stopping time. Consider for each t

$$\{\sigma + \tau \le t\} = \bigcup_{r \in \mathbb{Q} \cap [0,t]} \{\sigma \le r,\ \tau \le t - r\}$$

since each member of the union if in $\mathcal{F}_t$ so is the union.

Example 4.4. We have met the Poisson process $\{N_t,\ t \ge 0\}$ in the first lecture. Let

$$\tau = \inf\{t \ge 0 \ :\ N_t = 1\}$$

be the first jump time (recall that $X_1 = \tau$). Then τ is a stopping time with respect to the completed natural filtration $\{\mathcal{F}_t^N,\ t \ge 0\}$ since

$$(\tau > t) = \cap_{s \le t} \{N_s = 0\}$$

belongs to $\mathcal{F}_t^N$, $t \ge 0$. Notice that $\{\mathcal{F}_t^N,\ t \ge 0\}$ satisfies the standard assumptions (the reader should verify this, however see `Protter` [Pr04, page 22]). The stopping time τ may be also called the first passage time from level 1, or the hitting time of the set $B = [1, \infty)$.

Let $\{B_t,\ t \ge 0\}$ be a real Brownian motion on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$; for given $\lambda > 0$, let

$$\tau_\lambda = \inf\{t \ge 0 \ :\ B_t \ge \lambda\}$$

be the first passage time from level λ. This is again a stopping time (with respect to the filtration $\{\mathcal{F}_t, t \ge 0\}$).

Problem 4.2. Show that τ_λ is a stopping time.

4.1.1 Progressively measurable processes

In the last two examples, we have started to relate stopping times and stochastic processes. We already mentioned that for stopping times it is the filtration that matter, so we shall investigate properties of stochastic processes related to the filtration.

Recall that a stochastic process $\{X_t,\ t \in T\}$ with values in $(E, \mathcal{E})$ defined on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \in T\}, \mathbb{P})$, in general, does not show any regularity with respect to time t (since we impose a condition on ω for each fixed t). To overcome this point, we introduce the concept of *progressive measurability.* To simplify notation, we give the definition for $T = \mathbb{R}_+$.

Definition 4.5. *A stochastic process $X = \{X_t, t \geq 0\}$ is called* progressively measurable *if, for every $s > 0$, the mapping $(t, \omega) \longmapsto X_t(\omega)$ is measurable from $[0, s] \times \Omega$, endowed with the product σ-algebra $\mathcal{B}([0, s]) \otimes \mathcal{F}_s$, into $(E, \mathcal{E})$.*

A progressively measurable process is always adapted, but the converse does not hold in general. The following proposition shows a useful sufficient condition for this to happen.

Proposition 4.6. *An adapted stochastic process with right continuous paths is progressively measurable.*

Proof. For given $u \in T$ and $s \leq u$ we set

$$X^{(n)}(s) = \begin{cases} X_{\frac{k+1}{2^n}} & \text{for} \quad s \in \left[\frac{k}{2^n}u, \frac{k+1}{2^n}u\right[, \quad k = 0, \dots, 2^n - 1, \\ X_u & \text{for} \quad s = u. \end{cases}$$

We have $X^{(n)}(s) \to X(s)$, as $n \to \infty$, for every $s \leq u$, $\omega \in \Omega$. Moreover, $X^{(n)}$ is progressively measurable, since for any $\Gamma \in \mathcal{E}$ the set

$$\left\{(s, \omega)\ :\ s \leq u, X_s^{(n)}(\omega) \in \Gamma\right\} = \bigcup_{k<2^n} \left[\frac{k}{2^n}u, \frac{k+1}{2^n}u\right[\times \left\{X_{\frac{k+1}{2^n}} \in \Gamma\right\} \bigcup \{u\} \times \{X_u \in \Gamma\}$$

belongs to $\mathcal{B}([0, u]) \otimes \mathcal{F}_u$. Therefore, the mapping $(s, \omega) \to X_s(\omega)$ is $(\mathcal{B}([0, u]) \otimes \mathcal{F}_u)$-measurable, being pointwise limit of $(\mathcal{B}([0, u]) \otimes \mathcal{F}_u)$-measurable functions.

□

Therefore, both the Brownian motion and the Poisson process are progressively measurable processes. In the following result, we link progressive processes and stopping times.

Proposition 4.7. *Let $\{X_t, t \geq 0\}$ be a progressively measurable process with values in E and τ be a stopping time; define $X_\tau : \Omega \to E$, $X_\tau(\omega) := X_{\tau(\omega)}(\omega)$, for $\omega \in \Omega$. Then X_τ is a $\mathcal{F}_\tau$-measurable random variable.*

Proof. We have that X_τ is a random variable, being the composition of the measurable mappings $\omega \to (\omega, \tau(\omega))$ and $(\omega, t) \to X_t(\omega)$ (here the assumption of X being progressively measurable is not necessary, it is sufficient that X is measurable from $\mathbb{R}_+ \times \Omega$ into E).

Define $\Omega_t = \{\tau \leq t\}$. Then X_τ is measurable as a mapping from $(\Omega_t, \mathcal{F}_t)$ in $(E, \mathcal{E})$ being the composition of the measurable mappings

$$\omega \to (\omega, \tau(\omega)) \text{ from } (\Omega_t, \mathcal{F}_t) \text{ into } (\Omega_t \times [0, t], \mathcal{F}_t \otimes \mathcal{B}([0, t]))$$

and

$$(\omega, s) \to X_s(\omega) \text{ from } (\Omega_t \times [0,t], \mathcal{F}_t \otimes \mathcal{B}([0,t])) \text{ into } (E, \mathcal{E}).$$

Hence, if $A \in \mathcal{E}$:

$$\{\tau \leq t\} \cap \{X_\tau \in A\} = \{\omega \in \Omega_t \, : \, X_\tau(\omega) \in A\} \in \mathcal{F}_t$$

for each t which implies $\{X_\tau \in A\} \in \mathcal{F}_\tau$.

□

Corollary 4.8. *Assume that* $\{X_n,\ n \in \mathbb{N}\}$ *is a stochastic process adapted with respect to the filtration* $\{F_n,\ n \in \mathbb{N}\}$*. Then* X_τ *is* $\mathcal{F}_\tau$*-measurable for any stopping time* τ *with respect to the filtration* $\{F_n\}$*.*

4.2 Martingale

To motivate the introduction of martingales, we discuss a very simple example in terms of a gambler playing at a casino. The successive outcomes of the chosen game are independent each other, and results in a sequence of independent random variables $\{\Delta_n\}$ which represent the amount of win (and loss) of the n-th stake. Let X_0 be the initial fortune of the gambler; at each play he stakes a quantity $W_n \geq 0$ and wins $W_n \Delta_n$: for instance, playing a number on a roulette table, $\Delta_n = 35$ if he gets the number, and $\mathbb{P}(\Delta_n = 35) = \frac{1}{37}$, and $\Delta_n = -1$ if he looses, and $\mathbb{P}(\Delta_n = -1) = 1 - \frac{1}{37}$. Playing red (or black) on the same table leads to $\Delta_n = 1$ with probability $p = \frac{18}{37}$ and $\Delta_n = -1$ with probability $q = 1 - p = \frac{19}{37}$.

After $n-1$ plays, the gambler's capital is $X_{n-1} = X_0 + W_1 \Delta_1 + \cdots + W_{n-1} \Delta_{n-1}$. He next decides the amount W_n to stake on n-play: its decision shall depend only on $X_0, X_1, \dots, X_{n-1}$ and $\Delta_1, \dots, \Delta_{n-1}$, so $W_n \in \mathcal{F}_{n-1}$ is independent from Δ_n.

Example 4.9. Suppose that the gambler choose to bet one stack each time, and he decides to stop after the first loss; let τ be the first time for which $\Delta_n < 0$. Then τ is a stopping time; further, the betting system is $W_n = 1$ if $n \leq \tau$, and $W_n = 0$ otherwise.

In the simplest example with $\Delta_n = \pm 1$ with equal probability $p = q = \frac{1}{2}$, the martingale betting system is defined by $W_n = W_{n-1}$ if $\Delta_{n-1} = +1$ and $W_n = 2W_{n-1}$ otherwise, that is, if we loose, then we double the stack (W_1 is left to be chosen arbitrarily). It is easily seen that $\mathbb{E}[X_n] = X_0$ for every n. The sequence $\{X_n,\ n \geq 1\}$ of the gambler's capital is an example of the class of stochastic processes that we call *martingale*.

The interest for martingale in stochastic analysis, however, goes far beyond this example. In particular, martingale theory gives rise to several powerful inequalities that we shall exploit in connection with Brownian motion, Markov processes and in the construction of stochastic integrals. Also, much of the theory can first be proved for discrete times martingale and then extended to the continuous time case, due to some standard approximation procedure.

Definition 4.10. *Let* $\{X_t,\ t \in T\}$ *a stochastic process with values in* E*, adapted to a filtration* $\{\mathcal{F}_t,\ t \in T\}$*. We say that* X_t *is a* martingale *if*

1. *for each* $t \in T$*,* X_t *is integrable, and*
2. *for each* $s < t \in T$*,* $\mathbb{E}(X_t \mid \mathcal{F}_s) = X_s$ *a.s.* (**martingale property**)*.*

If the filtration is not explicitly stated, we shall assume that it is the natural filtration. If we want to stress the dependence on the filtration $\{\mathcal{F}_t,\ t \in T\}$, we write that X is an $\mathcal{F}_t$-martingale.

Definition 4.11. *Let $\{X_t,\ t \in T\}$ a real valued stochastic process, adapted to a filtration $\{\mathcal{F}_t,\ t \in T\}$. We say that X_t is a* submartingale *if*

1. *for each $t \in T$, X_t is integrable, and*
2. *for each $s < t \in T$, $X_s \le \mathbb{E}(X_t \mid \mathcal{F}_s)$ a.s. (**submartingale property**).*

The process $\{X_t,\ t \in T\}$ is a surmartingale *if $-X_t$ is a submartingale (i.e., the inequality in 2. is reversed).*

Example 4.12. If $T = \mathbb{N}$ and $\{\xi_i,\ i \in \mathbb{N}\}$ is a sequence of independent, zero mean random variables, setting

$$X_n = \xi_1 + \cdots + \xi_n$$

defines a martingale. Actually, $X_n = X_{n-1} + \xi_n$ where ξ_n is independent from $X_1, \ldots, X_{n-1}$ and also from $\mathcal{F}_{n-1} = \sigma(X_1, \ldots, X_{n-1})$, so one gets

$$\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] = \mathbb{E}[X_{n-1} \mid \mathcal{F}_{n-1}] + \mathbb{E}[\xi_n \mid \mathcal{F}_{n-1}] = X_{n-1} + \mathbb{E}[\xi_n] = X_{n-1}.$$

Example 4.13. If X_n is the capital of a gambler after the n-th play and $\mathcal{F}_n$ represents his information about the game, then $X_n = \mathbb{E}(X_{n+1} \mid \mathcal{F}_n)$ says that his expected fortune after next play is the same as his present one. This models for a *fair game*. Also, if we consider any betting strategy $\{W_n,\ n \in \mathbb{N}\}$ that is predictable, i.e., W_n is $\mathcal{F}_{n-1}$ measurable, and $\xi_n = W_n \Delta_n$ has zero mean (which implies –since $W_n \ge 0$ is independent from Δ_n– that $\mathbb{E}[\Delta_n] = 0$, another justification for the name "fair game") then previous example shows that

$$X_n = X_0 + W_1 \Delta_1 + \cdots + W_n \Delta_n$$

is again a martingale. If the game is unfair to the gambler, that is, if $E[\Delta_n] < 0$, then X_n is a submartingale.

In general, linear combination of martingales is again a martingale and linear combination (with positive coefficients) of sub- (sur-)martingale remains a submartingale (resp., a surmartingale).

Assume further that X is a martingale (resp. a submartingale) and $\Phi : \mathbb{R} \to \mathbb{R}$ a convex function (resp. an increasing convex function) such that $\Phi(X_t))$ is still integrable; then $\{\Phi(X_t),\ t \in T\}$ is a submartingale (it is enough to apply Jensen's inequality). In particular, if $\{X_t,\ t \in T\}$ is a martingale, $\{|X_t|,\ t \in T\}$ is a submartingale and if $X_t \in L^2$ for each t, then $\{X_t^2,\ t \in T\}$ is a submartingale (with respect to the same filtration).

Remark 4.14. Let $\{M_t,\ t \in T\}$ be a martingale; then $\mathbb{E}[M_t] = \mathbb{E}[\mathbb{E}(M_t \mid \mathcal{F}_s)] = \mathbb{E}[M_s]$ so that $\mathbb{E}[M_t]$ is constant for each $t \in T$.

In particular, if X_t is a positive martingale, i.e. $X_t \ge 0$ a.s., then $\{X_t,\ t \in T\}$ is bounded as a subset of L^1: $\mathbb{E}[|X_t|] = \mathbb{E}[X_t] = const.$

4.2.1 Inequalities for martingales

In this section we consider a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n,\ n \in \mathbb{N}\}, \mathbb{P})$ on which it is defined a discrete time submartingale $\{X_n,\ n \in \mathbb{N}\}$. We also deal with $\mathcal{F}_n$-stopping times (with values in $\mathbb{N}$).

With start with a useful result.

Proposition 4.15. *Let $X = \{X_n,\ n \in \mathbb{N}\}$ be a martingale and τ be a stopping time for the filtration $\{\mathcal{F}_n,\ n \in \mathbb{N}\}$. Then the stopped process $Y = \{X_{n\wedge\tau},\ n \in \mathbb{N}\}$ is again a martingale with respect to the filtration $\{\mathcal{F}_n,\ n \in \mathbb{N}\}$.*

Proof. Note that $\{\tau \geq n+1\} = \{\tau \leq n\}^c \in \mathcal{F}_n$; moreover $X_{n\wedge\tau}$ and $X_{(n+1)\wedge\tau}$ coincide on $\{\tau \leq n\}$. Hence we find, a.s.,

$$\begin{aligned}&\mathbb{E}(X_{(n+1)\wedge\tau} - X_{n\wedge\tau} \mid \mathcal{F}_n)\\ &\qquad = \mathbb{E}([X_{(n+1)\wedge\tau} - X_{n\wedge\tau}]\, \mathbb{1}_{\{\tau\geq n+1\}} \mid \mathcal{F}_n) = \mathbb{1}_{\{\tau\geq n+1\}}\mathbb{E}(X_{n+1} - X_n \mid \mathcal{F}_n) = 0.\end{aligned}$$

□

We leave as an exercise to prove the following simple fact that will be useful later on.

Problem 4.3. Let $X = \{X_n,\ n \in \mathbb{N}\}$ be a submartingale with respect to the filtration $\{\mathcal{F}_n,\ n \in \mathbb{N}\}$. Let τ be a stopping time with respect to $\{\mathcal{F}_n,\ n \in \mathbb{N}\}$, bounded by a constant $N \in \mathbb{N}$. Prove that

$$X_\tau \leq \mathbb{E}[X_N \mid \mathcal{F}_\tau] \qquad a.s.$$

If we integrate both sides of last formula, we obtain in particular that

$$\mathbb{E}[X_\tau] \leq \mathbb{E}[X_N].$$

Thanks to next theorem, we see that martingale property still holds if we consider bounded stopping times instead of fixed times; this is an important result with several consequences.

Theorem 4.16 (Optional Sampling Theorem). *Let $X = \{X_n,\ n \in \mathbb{N}\}$ be a submartingale and τ, σ be two stopping times for the filtration $\{\mathcal{F}_n,\ n \in \mathbb{N}\}$ bounded by a constant $N \in \mathbb{N}$ and such that $\tau \leq \sigma$ a.s. Then*

$$X_\tau \leq \mathbb{E}[X_\sigma \mid \mathcal{F}_\tau] \qquad a.s. \tag{4.1}$$

Proof. First, note that X_τ and X_σ are both integrable. Indeed, $|X_\tau| \leq \sum_{j=1}^N |X_j|$, a.s., and similarly for X_σ.

We have to prove that, for any $A \in \mathcal{F}_\tau$

$$\int_A X_\tau\, \mathbb{P}(d\omega) \leq \int_A X_\sigma\, \mathbb{P}(d\omega).$$

Letting $A_j = A \cap \{\tau = j\} \in \mathcal{F}_j$, one has thet A is given as a union of disjoint sets $A = \bigcup_{j=1}^{N} A_j$. Thus it is sufficient to show that, for any $j = 1, \ldots, N$, it holds

$$\int_{A_j} X_\sigma\, \mathbb{P}(d\omega) \geq \int_{A_j} X_\tau\, \mathbb{P}(d\omega) = \int_{A_j} X_j\, \mathbb{P}(d\omega).$$

Notice that in A_j it holds $\sigma \geq j$, therefore

$$\int_{A_j} X_{\sigma\wedge j}\,\mathbb{P}(d\omega) = \int_{A_j} X_j\,\mathbb{P}(d\omega).$$

We can show that, for $j \leq m < N$, we have

$$\begin{aligned}\int_{A_j} X_{\sigma\wedge m}\,\mathbb{P}(d\omega) &= \int_{A_j\cap\{\sigma\leq m\}} X_\sigma\,\mathbb{P}(d\omega) + \int_{A_j\cap\{\sigma>m\}} X_m\,\mathbb{P}(d\omega)\\ &\leq \int_{A_j\cap\{\sigma\leq m\}} X_\sigma\,\mathbb{P}(d\omega) + \int_{A_j\cap\{\sigma>m\}} X_{m+1}\,\mathbb{P}(d\omega)\\ &= \int_{A_j} X_{\sigma\wedge(m+1)}\,\mathbb{P}(d\omega)\end{aligned}$$

since $A_j \cap \{\sigma > m\} \in \mathcal{F}_m$. Using a recursive argument, from previous inequality follows

$$\int_{A_j} X_j\,\mathbb{P}(d\omega) = \int_{A_j} X_{\sigma\wedge j}\,\mathbb{P}(d\omega) \leq \int_{A_j} X_{\sigma\wedge N}\,\mathbb{P}(d\omega) = \int_{A_j} X_\sigma\,\mathbb{P}(d\omega)$$

as required.

□

Remark 4.17. Obviously, if $\{X_n\}$ is a martingale then (4.1) becomes an equality; for a surmartingale the inequality is reversed. Further, taking expectation in both sides of (4.1) we have

$$\mathbb{E}[X_\tau] \leq \mathbb{E}[X_\sigma], \qquad \text{for bounded stopping times } \tau \leq \sigma \text{ a.s.}$$

We provide now an interesting consequence of Theorem 4.16. Observe that this result provides an equivalent definition of martingales that does not require conditional expectation.

Corollary 4.18. *Consider a sequence $\{X_n,\ n \in \mathbb{N}\}$ of real, integrable random variables, adapted to a filtration $\{F_n, n \in \mathbb{N}\}$. Then $\{X_n\}$ is a martingale if and only if for each bounded stopping time τ it holds $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$.*

Proof. If $\{X_n\}$ is a martingale, the thesis follows from the Optional Sampling Theorem. Let us consider the converse inequality. Fix $m < n$ and $A \in \mathcal{F}_m$: then we aim to prove that $\mathbb{E}[X_m \mathbb{1}_A] = \mathbb{E}[X_n \mathbb{1}_A]$. Let us choose stopping times τ and σ such that

$$\tau(\omega) = \begin{cases} m & \text{for } \omega \in A\\ n & \text{for } \omega \in A^c\end{cases}$$

and $\sigma(\omega) \equiv n$, so that $\tau \leq \sigma$ a.s. and

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_0] = \mathbb{E}[X_\sigma] = \mathbb{E}[X_n]. \tag{4.2}$$

Now, since

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_\tau \mathbb{1}_A] + \mathbb{E}[X_\tau \mathbb{1}_{A^c}] = \mathbb{E}[X_m \mathbb{1}_A] + \mathbb{E}[X_n \mathbb{1}_{A^c}]$$

and $\mathbb{E}[X_n] = \mathbb{E}[X_n \mathbb{1}_A] + \mathbb{E}[X_n \mathbb{1}_{A^c}]$, equality (4.2) implies

$$\mathbb{E}[X_m \mathbb{1}_A] = \mathbb{E}[X_n \mathbb{1}_A]$$

and the proof is complete.

□

The following inequalities play a fundamental rôle in the theory of martingales. We are given a sequence $X = \{X_n,\ n \in \mathbb{N}\}$ of integrable random variables, adapted to a filtration $\{F_n,\ n \in \mathbb{N}\}$; define $X_n^* = \sup_{j \le n} |X_j|$; then X_n^* is integrable and Chebychev's inequality implies

$$\mathbb{P}(X_n^* \ge \alpha) \le \frac{1}{\alpha}\mathbb{E}[X_n^*], \qquad \alpha > 0.$$

In case X is a positive submartingale, we can substitute the mean of X_n^* with that of X_n.

Theorem 4.19 (Maximal inequality). *Let $\{X_n, n \in \mathbb{N}\}$ be a positive submartingale. Then for every $\lambda > 0$ we have*

$$\mathbb{P}(X_n^* \ge \lambda) \le \frac{1}{\lambda}\mathbb{E}[X_n]. \tag{4.3}$$

Proof. Let $A = (\max_{i=1,\dots,n} X_i \ge \lambda)$; we define a partition $\{A_i\}$ for A setting

$$A_1 = \{\omega \mid X_1(\omega) \ge \lambda\}$$
$$A_2 = \{\omega \mid X_1(\omega) < \lambda \le X_2(\omega)\}$$

and, in general

$$A_i = \{\omega \mid \max_{k=1,\dots,i-1} X_k(\omega) < \lambda \le X_i(\omega)\}$$

so that

$$A = \bigcup_{i=1}^{n} A_i = \{\omega \mid \max_i X_i(\omega) \ge \lambda\}$$

(note that some A_i could be empty). For every i it holds $X_i \ge \lambda$, i.e., $\frac{X_i}{\lambda} \ge 1$ on A_i. Then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A_i) = \sum_{i=1}^{n} \int_{A_i} \mathbb{P}(d\omega) \le \frac{1}{\lambda} \sum_{i=1}^{n} \int_{A_i} X_i \, \mathbb{P}(d\omega).$$

If we apply submartingale property we get

$$\int_{A_i} X_n \, \mathbb{P}(d\omega) \ge \int_{A_i} X_i \, \mathbb{P}(d\omega);$$

then

$$\mathbb{P}(A) \le \frac{1}{\lambda} \sum_{i=1}^{n} \int_{A_i} X_i \, \mathbb{P}(d\omega) \le \frac{1}{\lambda} \sum_{i=1}^{n} \int_{A_i} X_n \mathbb{P}(d\omega) = \frac{1}{\lambda} \int_{A} X_n \mathbb{P}(d\omega) \le \frac{1}{\lambda}\mathbb{E}[X_n],$$

which gives (4.3).

□

Lemma 4.20. *Let* $\{X_n, n \in \mathbb{N}\}$ *be a positive submartingale, p-integrable for some* $p > 1$. *Then, for every* $n \in \mathbb{N}$,

$$\mathbb{E}\left[|X_n^*|^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[X_n^p], \quad \text{where } X_n^* = \max_{1 \le k \le n} X_k.$$

Proof. Let us define, for $\lambda > 0$

$$\tau_\lambda(\omega) = \begin{cases} \inf\{1 \le k \le n \;:\; X_k(\omega) > \lambda\} \\ n+1 \qquad \text{if } \{\quad\} = \varnothing. \end{cases}$$

With the notation introduced in the above proof, it is $A_k = A_k(\lambda) = \mathbb{1}_{\{\tau_\lambda = k\}}$. It is

$$\sum_{k=1}^n \mathbb{1}_{\{\tau_\lambda = k\}} = \mathbb{1}_{\{X_n^* > \lambda\}}$$

which implies, for every $p > 1$,[1]

$$|X_n^*|^p = p \int_0^\infty \lambda^{p-1} \sum_{k=1}^n \mathbb{1}_{\{\tau_\lambda = k\}} \,\mathrm{d}\lambda. \tag{4.4}$$

Since it holds that $\lambda \le X_k$ on A_k, we have

$$\lambda^{p-1} \sum_{k=1}^n \mathbb{1}_{\{\tau_\lambda = k\}} \le \lambda^{p-2} \sum_{k=1}^n X_k \mathbb{1}_{\{\tau_\lambda = k\}}. \tag{4.5}$$

Now we know that X is a submartingale while $\mathbb{1}_{\{\tau_\lambda = k\}}$ is $\mathcal{F}_k$-measurable, hence

$$X_k \mathbb{1}_{\{\tau_\lambda = k\}} \le \mathbb{E}(X_n \mid \mathcal{F}_k) \mathbb{1}_{\{\tau_\lambda = k\}} = \mathbb{E}(X_n \mathbb{1}_{\{\tau_\lambda = k\}} \mid \mathcal{F}_k).$$

Using the above inequality and taking expectation in (4.5) we obtain (recall the property of the mean of a conditional expectation)

$$\mathbb{E}\left[\lambda^{p-1} \sum_{k=1}^n \mathbb{1}_{\{\tau_\lambda = k\}}\right] \le \mathbb{E}\left[\lambda^{p-2} \sum_{k=1}^n X_n \mathbb{1}_{\{\tau_\lambda = k\}}\right]$$

integrating in λ in the last inequality we get

$$\begin{aligned} \mathbb{E}[|X_n^*|^p] &\le p\mathbb{E}\left[\int_0^\infty \lambda^{p-2} \sum_{k=1}^n X_n \mathbb{1}_{\{\tau_\lambda = k\}} \,\mathrm{d}\lambda\right] \\ &= p\mathbb{E}\left[X_n \int_0^\infty \lambda^{p-2} \mathbb{1}_{\{X_n^* > \lambda\}} \,\mathrm{d}\lambda\right] = \frac{p}{p-1} \mathbb{E}[|X_n^*|^{p-1} X_n]; \end{aligned}$$

finally, using Hölder's inequality we have

[1] we use the identity $\mu^p = \int_0^\mu p\lambda^{p-1} \,\mathrm{d}\lambda = \int_0^\infty p\lambda^{p-1} \mathbb{1}_{\{\mu > \lambda\}} \,\mathrm{d}\lambda$

$$\mathbb{E}[|X_n^*|^p] \le \frac{p}{p-1}\mathbb{E}[|X_n^*|^p]^{\frac{p-1}{p}}\mathbb{E}[X_n^p]^{\frac{1}{p}}$$

which proves the thesis.

□

As a direct consequence of the lemma, we get the following result

Theorem 4.21 (Doob's inequality). *Let $X = \{X_n,\ n \in \mathbb{N}\}$ be a martingale such that $\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|^p] < \infty$, for some $1 < p < \infty$. Then*

$$\mathbb{E}\left[\sup_{n\in\mathbb{N}} |X_n|^p\right] \le \left(\frac{p}{p-1}\right)^p \sup_{n\in\mathbb{N}} \mathbb{E}\left[|X_n|^p\right]. \tag{4.6}$$

4.2.2 Continuous time martingales.

Now we extend the main results proved in previous sections to the case of continuous time martingales, where T is an *interval* of $\mathbb{R}_+$ and trajectories are right-continuous. We will see that Brownian motion is a square integrable process with continuous paths on the set of times $T = \mathbb{R}_+$.

Theorem 4.22. *Let $X = \{X_t,\ t \in T\}$ be a right-continuous submartingale on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$.*

(1) If τ and σ are stopping times for the filtration $\{\mathcal{F}_t\}$ with $\tau \le \sigma$ a.s., and σ is bounded a.s., then

$$X_\tau \le \mathbb{E}(X_\sigma \mid \mathcal{F}_\tau) \qquad a.s. \tag{4.7}$$

(2) If $[a,b]$ is an interval in $\mathbb{R}_+$, for every $\lambda > 0$, we have

$$\mathbb{P}(\sup_{s\in[a,b]} X_s \ge \lambda) \le \frac{1}{\lambda}\mathbb{E}[|X_b|]. \tag{4.8}$$

We just provide an outline of the proof of (2). Take any finite sequence $(t_k) \subset [a,b]$, $t_0 = a < t_1 < \ldots < t_n = b$, Theorem 4.19 implies

$$\mathbb{P}(\sup_k X_{t_k} \ge \lambda) \le \frac{1}{\lambda}\mathbb{E}[|X_b|];$$

now the thesis follows since trajectories are right-continuous and so

$$\sup_{\mathbb{Q}\cap[a,b]} X_t = \sup_{[a,b]} X_t.$$

Using Proposition 4.15, we get

Corollary 4.23. *Let $X = \{X_t,\ t \in T\}$ be a right-continuous martingale and τ be a stopping time for the filtration $\{\mathcal{F}_t,\ t \in T\}$. Then the stopped process $Y = \{X_{t\wedge\tau},\ t \in T\}$ is a martingale with respect to the same filtration $\{\mathcal{F}_t,\ t \in T\}$.*

Theorem 4.24. *Let $X = \{X_t,\ t \in T\}$ be a right-continuous martingale on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, such that $\{X_t\}$ is bounded in L^p, $p > 1$. Then, setting $X^* = \sup_{t\in T} |X_t|$, it holds $X^* \in L^p$ and*

$$(\mathbb{E}[|X^*|^p])^{1/p} \le \frac{p}{p-1} \left(\sup_{t\in T} \mathbb{E}[|X_t|^p] \right)^{1/p}. \tag{4.9}$$

A Brownian motion $\{B_t\}$ defined on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ is an $\mathcal{F}_t$-martingale. Indeed we have

$$\mathbb{E}(B_t \mid \mathcal{F}_s) = \mathbb{E}(B_t - B_s \mid \mathcal{F}_s) + \mathbb{E}(B_s \mid \mathcal{F}_s) = B_s,$$

since $B_t - B_s$ is independent from $\mathcal{F}_s$ and with mean 0, while B_s is $\mathcal{F}_s$-measurable. Also, some transformations of Brownian motion are martingales.

Proposition 4.25. *Define the processes $Y = \{Y_t = B_t^2 - t,\ t \ge 0\}$ and $Z = \{Z_t = \mathrm{e}^{\nu B_t - \frac{1}{2}\nu^2 t},\ t \ge 0\}$; Z is called* geometric Brownian motion*. Then both Y and Z are martingale processes with respect to the same filtration $\{F_t,\ t \ge 0\}$ of the Brownian motion $\{B_t,\ t \ge 0\}$.*

Proof. We only prove the assertion concerning Z. The other can be proved similarly.

Take $s < t$; since $B_t - B_s$ is independent from $\mathcal{F}_s$, it follows

$$\begin{aligned}\mathbb{E}(\mathrm{e}^{\nu B_t - \frac{1}{2}\nu^2 t} \mid \mathcal{F}_s) &= \mathbb{E}(\mathrm{e}^{\nu B_s - \frac{1}{2}\nu^2 s} \mathrm{e}^{\nu(B_t - B_s) - \frac{1}{2}\nu^2 (t-s)} \mid \mathcal{F}_s) \\ &= \mathrm{e}^{\nu B_s - \frac{1}{2}\nu^2 s} \mathbb{E}(\mathrm{e}^{\nu(B_t - B_s) - \frac{1}{2}\nu^2 (t-s)}).\end{aligned}$$

It remains to prove that

$$\mathbb{E}(\mathrm{e}^{\nu(B_t - B_s) - \frac{1}{2}\nu^2 (t-s)}) = 1.$$

Setting $h = t - s$, since $B_t - B_s$ has Gaussian distribution $\mathcal{N}(0, h)$, we have

$$\begin{aligned}\mathbb{E}(\mathrm{e}^{\nu(B_t - B_s) - \frac{1}{2}\nu^2 (t-s)}) &= \frac{1}{\sqrt{2\pi h}} \int_{\mathbb{R}} \mathrm{e}^{\nu x - \frac{1}{2}\nu^2 h} \mathrm{e}^{-x^2/2h}\, \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi h}} \int_{\mathbb{R}} \exp\big(- \frac{x^2 - 2\nu h x + \nu^2 h^2}{2h} \big)\, \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi h}} \int_{\mathbb{R}} \mathrm{e}^{-(x-\nu h)^2/2h}\, dx = 1\end{aligned}$$

where the last equality follows since $\frac{1}{\sqrt{2\pi h}} \mathrm{e}^{-(x-\nu h)^2/2h}$ is the density of the Gaussian distribution $\mathcal{N}(\nu h, h)$.

□

Problem 4.4. Prove that $Y = \{Y_t = B_t^2 - t,\ t \ge 0\}$ is an $\mathcal{F}_t$-martingale.

There is also an equivalent characterization of Brownian motion in terms of martingale property: see for instance the book [IW81].

Theorem 4.26. *A continuous d-dimensional process* $\{X_t,\ t \geq 0\}$ *defined on a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P})$ *is a d-dimensional Brownian motion if* $X_0 = 0$*, and, for any* $\lambda \in \mathbb{R}^d$,

$$Y_t^\lambda = \exp(\mathrm{i}\langle \lambda, X_t \rangle + \frac{1}{2}|\lambda|^2 t) \quad \textit{is an } \mathcal{F}_t\textit{-martingale.} \tag{4.10}$$

Proof. Formula (4.10) implies that

$$\mathbb{E}[\exp(\mathrm{i}\lambda \cdot (X_t - X_s)) \mid \mathcal{F}_s] = \mathbb{E}[\exp(-\frac{1}{2}|\lambda|^2(t-s))].$$

Thanks to Exercise 1.7 we know that this implies that $X_t - X_s$ is independent from $\mathcal{F}_s$; further, from the uniqueness of characteristic function, it has a Gaussian distribution with zero mean and covariance matrix $(t-s)I$. Hence properties (a)–(c) of Definition 2.5 holds and X_t is a Brownian motion.

□

5

Markov processes

In this lecture, we are concerned with Markov stochastic processes. Markov property have an appealing intuitive description in terms of conditional expectation, as it is usually expressed by saying that the future after time t, given the present state at time t, is independent from the past before time t, and the process evolves like it is starting afresh at the time t of observation. The Brownian motion, the Poisson process and more generally solutions to stochastic differential equations satisfies such property.

This lecture will provide an introduction to the main results, as we are not in search for the greatest possible generality; important references about this subject are certainly the books by `Ethier and Kurtz` [EK86] or `Blumenthal and Getoor` [BG68], where also hystorical remarks are provided.

5.1 Preliminaries

In this lecture we deal with the positive half-line of continuous times $T = [0, \infty)$; in the same way, however, we could have considered the case of $T = [s, \infty)$, for some $s \in \mathbb{R}$.

Consider a stochastic processes $X = \{X_t,\ t \geq 0\}$ defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P})$, taking values in a euclidean space E endowed with the Borel σ-field $\mathcal{E} = \mathcal{B}(E)$.

We say that X has *initial distribution* μ (where μ is a Borel probability measure on E) if the law of the random variable X_0 is μ, i.e.,

$$\mathbb{P}(X_0 \in \Gamma) = \mu(\Gamma), \qquad \Gamma \in \mathcal{E}.$$

If the initial measure is concentrated in a single point, i.e., $\mu = \delta_x$ is Dirac delta measure at $x \in E$, we say that X *starts from* x *a.s.*

In particular a Brownian motion with values in $\mathbb{R}^d$ and initial distribution μ is a stochastic process satisfying the assumptions of Definition 2.5 except the first one which is replaced by $\mathbb{P}(B_0 \in \Gamma) = \mu(\Gamma)$, $\Gamma \in \mathcal{E}$.

Example 5.1. Given a Brownian motion $B = \{B_t,\ t \geq 0\}$, with values in $\mathbb{R}^d$, defined on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P})$, the family of processes $\{B^x_t := x + B_t,\ t \geq 0, x \in \mathbb{R}^d\}$ is a family of Brownian motions starting from $x \in \mathbb{R}^d$ (each Brownian motion is still defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P})$).

Problem 5.1. Let $B = \{B_t,\ t \geq 0\}$ be a Brownian motion, with values in $\mathbb{R}^d$, defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P})$. Let μ be a Borel probability measure on $\mathbb{R}^d$. On the probability space $(\mathbb{R}^d, \mathcal{K}, \mu)$ we consider the random variable η, $\eta(x) = x$, $x \in \mathbb{R}^d$ (here $\mathcal{K}$ is the completion of $\mathcal{E}$ with respect to μ). Show that the process $\{\bar{B}_t := \eta + B_t,\ t \geq 0\}$ is a Brownian motion with initial distribution μ on $(\Omega \times \mathbb{R}^d, \mathcal{F} \otimes \mathcal{K}, \{\mathcal{F}_t \otimes \mathcal{K},\ t \geq 0\}, \mathbb{P} \otimes \mu)$.

Hint: using Definition 2.5, search for the distribution of $\bar{B}_t - \bar{B}_s$.

In the first lecture, we have introduced the conditional expectation $\mathbb{E}(X \mid \mathcal{D})$ of a real random variable X with respect to a σ-field $\mathcal{D} \subset \mathcal{F}$. Clearly, this conditional expectation can be defined componentwise for random variables X taking values in E.

A particular case of conditional expectation is $\mathbb{E}(f(X) \mid \mathcal{D})$, where $f : E \to \mathbb{R}$ is any Borel measurable, bounded function.

If f is the indicator function of a set $\Gamma \in \mathcal{E}$, we define the conditional probability of the event $\{X \in \Gamma\}$ given $\mathcal{D}$ as

$$\mathbb{P}(X \in \Gamma \mid \mathcal{D}) := \mathbb{E}(\mathbb{1}_\Gamma(X) \mid \mathcal{D}). \tag{5.1}$$

Let $Y : \Omega \to E$ be another random variable; with an abuse of notation, we shall write $\mathbb{E}(X \mid Y) := \mathbb{E}(X \mid \sigma(Y))$ and similarly for $\mathbb{P}(X \in \Gamma \mid Y) = \mathbb{P}(X \in \Gamma \mid \sigma(Y))$ (recall that $\sigma(Y)$ is the smallest σ-algebra which makes Y measurable).

5.2 The Markov property

We begin with a pictorial description of Markov property for a Brownian motion $B = \{B_t,\ t \geq 0\}$ defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P})$. Assume that we can follow its path up to some time $s > 0$ and in particular to observe the value B_s. Conditioning to this observation, what can we say about the probability of finding at a later time t the process $B_t \in \Gamma$ for a given $\Gamma \in \mathcal{E}$? Since $B_t = (B_t - B_s) + B_s$ and the increment $(B_t - B_s)$ is independent from the path history up to time s, we shall be able to express the probability of the event $(B_t \in \Gamma)$ in terms of the initial distribution B_s and the distribution of $B_t - B_s$ (which equals to B_{t-s}). This can be expressed formally by saying that

$$\mathbb{P}(B_t \in \Gamma \mid \mathcal{F}_s) = \mathbb{P}(B_t \in \Gamma \mid B_s), \quad 0 < s < t,\ \Gamma \in \mathcal{E} \tag{5.2}$$

However, for a rigorous proof of (5.2), we need the following technical tool, that is a sort of "freezing lemma", which is often useful.

Lemma 5.2. *We are given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a sub-$\sigma-$field $\mathcal{G}$ such that $X : \Omega \to E$ is $\mathcal{G}$-measurable and $Y : \Omega \to E$ is independent from $\mathcal{G}$. Then, for any Borel and bounded function $\psi : E \times E \to \mathbb{R}$, setting $\Psi(x) = \mathbb{E}[\psi(x, Y)]$, it holds*

$$\mathbb{E}(\psi(X, Y) \mid \mathcal{G}) = \Psi(X). \tag{5.3}$$

Proof. First note that the result is true when $\psi = \mathbb{1}_A \times \mathbb{1}_B$, i.e., $\psi(x, y) = \mathbb{1}_A(x) \cdot \mathbb{1}_B(y)$, for any $x, y \in E$, with $A, B \in \mathcal{E}$. Indeed in this case we have

$$\mathbb{E}(\psi(X, Y) \mid \mathcal{G}) = \mathbb{1}_A(X)\mathbb{E}(\mathbb{1}_B(Y) \mid \mathcal{G}) = \mathbb{1}_A(X)\mathbb{E}[\mathbb{1}_B(Y)] = \Psi(X).$$

The general case follows by a standard approximation argument. Indeed f is a pointwise limit of a sequence f_n of elementary functions, i.e., linear combinations of functions of the type $\mathbb{1}_A \times \mathbb{1}_B$, with $A, B \in \mathcal{E}$, which verify $\sup_{x\in E} |f_n(x)| \leq \sup_{x\in E} |f(x)|$, $n \in \mathbb{N}$.
□

We can now check (5.2). Setting $f(x,y) = \mathbb{1}_\Gamma(x+y)$, it holds

$$\begin{aligned}\mathbb{P}(B_t \in \Gamma \mid \mathcal{F}_s) &= \mathbb{E}(f(B_s,(B_t - B_s)) \mid \mathcal{F}_s) = \mathbb{E}[\mathbb{1}_\Gamma(B_t - B_s + y)]\Big|_{y=B_s} \\ &= \mathbb{E}[\mathbb{1}_\Gamma(B_{t-s} + y)]\Big|_{y=B_s} = \int_{\mathbb{R}^d} \frac{1}{\sqrt{(2\pi)^d (t-s)^d}} \exp\left(-\frac{|z-B_s|^2}{2(t-s)}\right) \mathrm{d}z = \mathbb{P}(B_t \in \Gamma \mid B_s), \quad (5.4)\end{aligned}$$

since $\int_{\mathbb{R}^d} \frac{1}{\sqrt{(2\pi)^d(t-s)^d}} \exp\left(-\frac{|z-B_s|^2}{2(t-s)}\right) \mathrm{d}z \in \sigma(B_s)$.

Equation (5.2) expresses the *Markov property of Brownian motion.*

The definition of Markov property for stochastic processes expresses in a formal way the idea that the present state allows to predict future states as well as the whole history of past and present states does – that is, the process is memoryless.

Definition 5.3. *A stochastic process $X = \{X_t,\ t \geq 0\}$ defined on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P})$, taking values in a euclidean space E, is a* Markov process *(with respect to $\{\mathcal{F}_t,\ t \geq 0\}$) if the following condition holds:*

$$\mathbb{P}(X_{t+s} \in \Gamma \mid \mathcal{F}_s) = \mathbb{P}(X_{t+s} \in \Gamma \mid X_s), \qquad \textit{for any } s,t \geq 0,\ \Gamma \in \mathcal{E}. \quad (5.5)$$

An approximation argument shows that condition (5.5) is equivalent to the following one

$$\mathbb{E}(f(X_{t+s}) \mid \mathcal{F}_s) = \mathbb{E}(f(X_{t+s}) \mid X_s), \qquad \text{for any Borel and bounded function } f: E \to \mathbb{R}. \quad (5.6)$$

In the discussion preceding Definition 5.3 we have proved

Theorem 5.4. *A d-dimensional Brownian motion B with any initial distribution μ is a Markov process.*

Notice that a Brownian motion is both a martingale and a Markov process. This is rather exceptional. For instance, one can prove that Poisson process $\{N_t,\ t \geq 0\}$ verifies the Markov property (with respect to its natural completed filtration), see the following problem, but it is not a martingale (the expected value is strictly increasing in time).

Problem 5.2. Prove that every process $\{X_t,\ t \geq 0\}$ defined on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P})$ with independent increments (i.e., $X_t - X_s$ is independent from $\mathcal{F}_s$, for $0 \leq s < t$) is a Markov process.

Remark 5.5. Let us introduce the family of σ-fields

$$\mathcal{F}^t = \sigma(X_s,\ s \geq t)$$

sometimes called the *post-t filtration*. Notice that the present state X_t belongs both to $\mathcal{F}_t$ and $\mathcal{F}^t$. Markov property (with respect to the natural filtration $\{\mathcal{F}^X_t,\ t \geq 0\}$) can be equivalently written in each of the following forms:

(i) for any $A \in \mathcal{F}_t$, $B \in \mathcal{F}^t$ it holds $\mathbb{P}(A \cap B \mid X_t) = \mathbb{P}(A \mid X_t)\mathbb{P}(B \mid X_t)$.

In this form we clearly realize the (conditional) symmetry between past and future when the present is known (see, for instance, Wentzell [Ve83]).

(ii) for any $B \in \mathcal{F}^t$ it holds $\mathbb{P}(B \mid X_t) = \mathbb{P}(B \mid \mathcal{F}_t)$.

This formulation is a (rather obvious) consequence of (5.5) and a monotone class application.

Problem 5.3. As stated in K.-L. Chung [Ch02], "... big mistakes have been caused by the misunderstanding of the exact meaning of the words "when the present is known". For instance, it is *false* that $\mathbb{P}(B_3 \in (0,1) \mid B_2 \in (0,1),\ B_1 \in (0,1))$ is equal to $\mathbb{P}(B_3 \in (0,1) \mid B_2 \in (0,1))$. Why there is no contradiction with formula (5.2)?

5.3 Markov processes with transition functions

A *time-homogeneous Markov transition probability function* on E (briefly, a transition probability function on E or a t.p.f. on E) is a function $p(t,x,A)$ with $t \geq 0$, $x \in E$ and $A \in \mathcal{E}$, such that

1. for fixed t and A, the map: $x \mapsto p(t,x,A)$ is $\mathcal{E}$-measurable;
2. for fixed t and x, $p(t,x,\cdot)$ is a probability measure on $(E,\mathcal{E})$;
3. the Chapman-Kolmogorov equation holds:

$$p(t+s,x,A) = \int_E p(s,y,A)\, p(t,x,\mathrm{d}y), \qquad \forall\, t,s \geq 0,\ x \in E,\ \ A \in \mathcal{E}. \tag{5.7}$$

4. $p(0,x,\cdot) = \delta_x$, $x \in E$.

Definition 5.6. *A Markov process with transition probability function p is a family of E-valued stochastic processes $X^x = \{X^x_t,\ t \geq 0\}$, possibly depending on $x \in E$, which are defined on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P}^x)$ (note that here only the probability measure $\mathbb{P}^x$ may depend on x) such that*

1. $\mathbb{P}^x(X^x_0 \in \Gamma) = \delta_x(\Gamma)$, for every $x \in E$, $\Gamma \in \mathcal{E}$ (i.e. X^x starts from x, $\mathbb{P}^x$-a.s.);
2. for any $s,t \geq 0$, $x \in E$, and $\Gamma \in \mathcal{E}$,

$$\mathbb{P}^x(X^x_{t+s} \in \Gamma \mid \mathcal{F}_s) = p(t, X^x_s, \Gamma), \qquad \mathbb{P}^x - a.s. \tag{5.8}$$

Note that condition (5.8) implies that $\mathbb{P}^x(X^x_{t+s} \in \Gamma \mid \mathcal{F}_s) = \mathbb{P}^x(X^x_{t+s} \in \Gamma \mid X^x_s)$ so that, in particular, each process X^x satisfies the Markov property (5.5).

The following remarks are important in order to clarify previous definition. We show first that the t.p.f. $p(t,x,\cdot)$ associated to a Markov process X is, summing up, the law of the random variable X^x_t. However, it is not always possible to associate to a Markov process a transition probability function identifying it via the law of the process.

Remark 5.7.

(i) Let us consider equation (5.8). Putting $s = 0$ and taking the expectation $\mathbb{E}^x$ with respect to $\mathbb{P}^x$

$$\mathbb{E}^x[\mathbb{P}^x(X_t^x \in \Gamma \mid \mathcal{F}_0)] = \mathbb{E}^x[\mathbb{1}_\Gamma(X_t^x)] = \mathbb{P}^x(X_t^x \in \Gamma) = p(t,x,\Gamma). \tag{5.9}$$

Hence $p(t, x, \cdot)$ is the law of the random variable X_t^x.

(ii) In our course, an important example of Markov processes with t.p.f. is provided by the solutions of stochastic differential equations; such processes are $X^x = \{X_t^x,\ t \geq 0\}$s defined on a fixed stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. In that case, we are taking $\mathbb{P}^x = \mathbb{P}$, $x \in E$, while the assumption that X^x starts from x means that $X_0^x = x$, $\mathbb{P}$-a.s.; moreover (5.8) becomes

$$\mathbb{P}(X_{t+s}^x \in \Gamma \mid \mathcal{F}_s) = p(t, X_s^x, \Gamma), \qquad \mathbb{P} - a.s.,$$

and (5.9) becomes $\mathbb{P}(X_t^x \in \Gamma) = p(t,x,\Gamma)$, $t \geq 0$, $\Gamma \in \mathcal{E}$.

(iii) Sometimes, it is convenient to fix the process and let the probability measures vary with x in such a way that the mapping:

$$x \mapsto \mathbb{P}^x(A) \quad \text{is a Borel mapping for any } A \in \mathcal{F}. \tag{5.10}$$

Then, we may also consider the random variable $\omega \mapsto \mathbb{P}^{X_t^x(\omega)}(F)$, for any $F \in \mathcal{F}$. Using (5.9), we can write the Markov property in the following "suggestive" form

$$\mathbb{P}^x(X_{s+t} \in A \mid \mathcal{F}_t) = p(s, X_t, A) = \mathbb{P}^{X_t^x}(X_s \in A). \tag{5.11}$$

Example 5.8. The d-dimensional Brownian motion starting from x, i.e., the family of processes $\{B_t^x := x + B_t,\ t \geq 0\}$, $x \in \mathbb{R}^d$, defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P}))$ is rather a special example of Markov process with transition probability function. Indeed, in this case we can explicitly compute the t.p.f. p

$$p(t,x,A) = \mathbb{P}(B_t^x \in A) = \int_A \frac{1}{\sqrt{(2\pi t)^d}} \exp\left(-\frac{|x-y|^2}{2t}\right) \mathrm{d}y, \qquad A \in \mathcal{B}(\mathbb{R}^d),\ \ t > 0,\ x \in \mathbb{R}^d. \tag{5.12}$$

In addition, in this case there exists a density for the t.p.f. $p(t, x, \cdot)$, $t > 0$, $x \in \mathbb{R}^d$, with respect to the Lebesgue measure.

It turns out that for a Markov process associated to a transition probability function p, this function allows to determine all the finite dimensional distributions of the process. For simplicity of notation we shall make the computations in case when $X^x = X$, $x \in E$. We have, for $t > s \geq 0$, $A, B \in \mathcal{E}$,

$$\begin{aligned}\mathbb{P}^x(X_t \in A, X_s \in B) &= \mathbb{E}^x[\mathbb{1}_B(X_s)\,\mathbb{1}_A(X_t)] = \mathbb{E}^x[\mathbb{1}_B(X_s)\,\mathbb{E}^x[\mathbb{1}_A(X_t) \mid \mathcal{F}_s]] \\ &\mathbb{E}^x[\mathbb{1}_B(X_s)\,p(t-s, X_s, A)] = \int_B p(t-s, y, A) p(s, x, \mathrm{d}y)\end{aligned}$$

where in the last equality we are using the fact that $p(s, x, \cdot)$ is the distribution of X_s under $\mathbb{P}^x$.

More generally, using induction, we find, for $0 \leq t_0 < t_1 < \cdots < t_n$, with $A_i \in \mathcal{E}$, $i = 0, \ldots, n$:

$$\begin{aligned}&\mathbb{P}^x(X_{t_0} \in A_0, \ldots, X_{t_n} \in A_n) \\ &\quad = \int_{A_0} p(t_0, x, \mathrm{d}x_0) \int_{A_1} p(t_1 - t_0, x_0, \mathrm{d}x_1) \ldots \int_{A_n} p(t_n - t_{n-1}, x_{n-1}, \mathrm{d}x_n).\end{aligned} \tag{5.13}$$

We conclude the section with a general result about the existence of Markov processes associated with a given t.p.f.. Indeed, the transition function $p(t,x,A)$ uniquely determines a (unique) stochastic Markov process through its finite dimensional distributions. This result follows from Kolmogorov's theorem 2.13.

Theorem 5.9. *Consider a transition probability function $p(t,x,A)$ on E. Then, for any Borel probability distribution μ on E, there exists a process $X = \{X_t,\ t \geq 0\}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P}^\mu)$ taking values E and with initial law μ, such that*

$$\mathbb{P}^\mu(X_{t+s} \in \Gamma \mid \mathcal{F}_s^X) = p(t, X_s, \Gamma), \quad \Gamma \in \mathcal{E} \tag{5.14}$$

(here $\{\mathcal{F}_t^X,\ t \geq 0\}$ denotes the natural completed filtration of X).

Proof. We only sketch the proof which is divided in two steps.

First, define the family of finite dimensional distributions $\mu_{t_0,t_1,\dots,t_n}$, for $0 = t_0 < t_1 < \dots < t_n$, setting

$$\mu_{t_0,t_1,\dots,t_n}(A_0 \times A_1 \times \dots \times A_n) = \int_{A_0} \mu(\mathrm{d}x_0) \int_{A_1} p(t_1 - t_0, x_0, \mathrm{d}x_1) \dots \int_{A_n} p(t_n - t_{n-1}, x_{n-1}, \mathrm{d}x_n)$$

(compare with (5.13)). Chapman-Kolmogorov's equation (5.7) implies that the system of measures $\mu_{t_0,t_1,\dots,t_n}$ verifies the compatibility assumption. Then there exists a process X on a probability space $(\Omega, \mathcal{F}, \mathbb{P}^\mu)$ (we may take $\Omega = E^{[0,+\infty[}$ and $\mathcal{F}$ the completed σ-algebra generated by all cylindrical sets) with the prescribed finite dimensional distributions.

In the second part of the proof one shall verify that this process is actually a Markov process with the prescribed transition probability function and the given initial distribution μ. This last condition is easily checked, but the first one is rather lengthy and we shall omit its verification here.

□

In case when $\mu = \delta_x$, $x \in E$, one simply writes $\mathbb{P}^x$ instead of $\mathbb{P}^{\delta_x}$.

Remark 5.10. The completed natural filtration $\mathcal{F}_t^X$ in Theorem 5.9 is not necessarily right-continuous (hence, in general, it does not satisfy the standard assumptions). Therefore, the theorem does not clarify if the stochastic process X is defined on a stochastic basis.

However, we shall see that for a large class of processes, like Feller processes with right continuous paths (which are rather mild additional assumptions to impose), one can replace in (5.14) the filtration $\{\mathcal{F}_t^X,\ t \geq 0\}$ with the larger filtration $\{\mathcal{F}_{t+}^X,\ t \geq 0\}$ which satisfies the standard assumptions, see for instance `Wentzell` [Ve83]. This way we obtain a Markov process defined on a stochastic basis.

5.4 Strong Markov property

Recall that, by definition, a stopping time τ on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \in T\}, \mathbb{P})$ is a random variable with values in $T \cup \{+\infty\}$ such that $(\tau \leq t) \in \mathcal{F}_t$ for every t. For a stopping time τ, we define $\mathcal{F}_\tau$ the σ-field of sets $A \in \mathcal{F}$ such that $A \cap \{\tau \leq t\} \in \mathcal{F}_t$. If $\tau \equiv t$ is constant, $\mathcal{F}_\tau = \mathcal{F}_{t^+} = \mathcal{F}_t$. Recall that the filtration $\{\mathcal{F}_t,\ t \in T\}$ satisfies the standard assumptions and so in particular is right-continuous.

We give now directly the definition of strong Markov property for a Markov process associated to a transition probability function; the (preliminary) definition of strong Markov process –mimicking Definition 5.3– is left to the reader.

Definition 5.11. *A strong Markov process with transition probability function p is a family of E-valued progressively measurable stochastic processes $X^x = \{X^x_t,\ t \ge 0\}$, possibly depending on $x \in E$, which are defined on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \ge 0\}, \mathbb{P}^x)$ such that*

1. *$\mathbb{P}^x(X^x_0 \in \Gamma) = \delta_x(\Gamma)$, for every $x \in E$, $\Gamma \in \mathcal{E}$;*
2. *for any stopping time τ finite $\mathbb{P}^x$-a.s., for any $t \ge 0$, $x \in E$,*

$$\mathbb{P}^x(X^x_{t+\tau} \in \Gamma \mid \mathcal{F}_\tau) = p(t, X^x_\tau, \Gamma), \qquad \mathbb{P}^x - a.s.,\ \Gamma \in \mathcal{E}. \tag{5.15}$$

Since constant $\tau = s$ is a stopping time, we notice that (5.8) is satisfied: that is, strong Markov processes are, in particular, Markov processes. In general, the converse does not hold, as (5.15) is stronger than the corresponding condition in Definition 5.3.

In the following result we prove that a Brownian motion is a strong Markov process. Indeed, we show more, proving that Brownian motion starts afresh not only at every fixed time, but also starting from a stopping time τ.

Theorem 5.12. *Assume τ is a stopping time, a.s. finite, for the filtration $\{\mathcal{F}_t,\ t \ge 0\}$ associated to a Brownian motion B taking values in $\mathbb{R}^d$ and having a given initial distribution μ. Then, setting*

$$W_t := B_{t+\tau} - B_\tau, \qquad t \ge 0,$$

the stochastic process $W = \{W_t,\ t \ge 0\}$ is a standard Brownian motion on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}^W_t,\ t \ge 0\}, \mathbb{P})$, where $\{\mathcal{F}^W_t,\ t \ge 0\}$ is the natural completed filtration of W; moreover for every choice of times $t_0 \le t_1 \le \dots \le t_n$ the σ-algebras $\mathcal{F}^W_{t_1}, \dots, \mathcal{F}^W_{t_n}$ and $\mathcal{F}_\tau$ are independent.

Proof. For simplicity we consider the case $d = 1$. Assume further that τ is bounded a.s.; the general case can be obtained by considering a suitable sequence of approximating stopping times.

We shall show that the process W_t is gaussian and independent from $\mathcal{F}_\tau$; this will follow if we show that for every choice of times $t_0 \le t_1 \le \dots \le t_n$ and real numbers $\lambda_1, \dots, \lambda_n$ it holds[1]

$$\mathbb{E}\left(\exp\left(\mathrm{i}\sum_{k=1}^n \lambda_k (W(t_k) - W(t_{k-1}))\right) \mid \mathcal{F}_\tau\right) = \prod_{k=1}^n \exp(-\frac{1}{2}\lambda_k^2 (t_k - t_{k-1})). \tag{5.16}$$

Recall that given a Brownian motion B, the process $M_t := \exp(\mathrm{i}\lambda B_t + \frac{1}{2}\lambda^2 t)$ is a (complex valued) martingale with respect to the filtration $\{\mathcal{F}_t\}$. Using Theorem 4.16 we get

$$\begin{aligned}\mathbb{E}(\mathrm{e}^{\mathrm{i}\lambda W(t)} \mid \mathcal{F}_\tau) &= \mathbb{E}(\mathrm{e}^{\mathrm{i}\lambda(B(\tau+t) - B(\tau)} \mid \mathcal{F}_\tau) = \mathrm{e}^{-\mathrm{i}\lambda B(\tau)}\mathbb{E}(\mathrm{e}^{\mathrm{i}\lambda B(\tau+t)} \mid \mathcal{F}_\tau)\\ &= \mathrm{e}^{-\mathrm{i}\lambda B(\tau) - \frac{1}{2}\lambda^2(\tau+t)}\mathbb{E}[\underbrace{\exp(\mathrm{i}\lambda B(\tau+t) + \frac{1}{2}\lambda^2(\tau+t))}_{\mathbb{E}[M_{t+\tau} \mid \mathcal{F}_\tau] = M_\tau} \mid \mathcal{F}_\tau]\\ &= \mathrm{e}^{-\mathrm{i}\lambda B(\tau) - \frac{1}{2}\lambda^2(\tau+t)}\mathrm{e}^{\mathrm{i}\lambda B(\tau) + \frac{1}{2}\lambda^2 \tau} = \mathrm{e}^{\frac{1}{2}\lambda^2 t}\end{aligned}$$

[1] compare with Exercise 1.7(2) and Definition 2.8 of Gaussian processes; further, the process has independent increments since formula (5.16) implies that the covariance matrix is diagonal

and also

$$\mathbb{E}(\mathrm{e}^{\mathrm{i}\lambda(W(t_k)-W(t_{k-1}))} \mid \mathcal{F}_{\tau+t_{k-1}}) = \mathrm{e}^{\frac{1}{2}\lambda^2(t_k-t_{k-1})}.$$

From this identity, formula (5.16) follows easily and the theorem is proved.

□

Arguing as in the proof of Theorem 5.4, using Lemma 5.2, we deduce from the previous theorem

Corollary 5.13. *A d-dimensional Brownian motion starting from x is a strong Markov process with transition probability function p given in* (5.12).

5.5 Applications of Markov property to Brownian motion

We have already seen that *first passage time* for the Brownian motion is a stopping time with respect to the natural filtration, compare Exercise 4.2. In the following result, we apply optional sampling theorem to characterize its properties.

Theorem 5.14. *For fixed $\lambda \neq 0$, let us denote τ_λ the stopping time*

$$\tau_\lambda = \inf\{s > 0 \,:\, B_s = \lambda\},$$

$\tau_\lambda = +\infty$ if $\{\quad\} = \varnothing$. Then τ_λ is a stopping time, called the first passage time in λ*, τ_λ is finite a.s. and its law is characterized by the Laplace transform*

$$\mathbb{E}[\mathrm{e}^{-x\tau_\lambda}] = \mathrm{e}^{-\sqrt{2x\lambda^2}}. \tag{5.17}$$

Proof. Assume for simplicity $\lambda > 0$. Consider the martingale $Z_t = \mathrm{e}^{\nu B_t - \frac{1}{2}\nu^2 t}$: since τ_λ is a stopping time, so it is $\tau_\lambda \wedge t$ for all $t > 0$, and

$$\mathbb{E}[Z_{\tau_\lambda \wedge t}] = \mathbb{E}[Z_0] = 1.$$

We observe the estimate

$$Z_{\tau_\lambda \wedge t} = \exp\left(\nu B_{\tau_\lambda \wedge t} - \frac{1}{2}\nu^2(\tau_\lambda \wedge t)\right) \leq \mathrm{e}^{\nu\lambda}.$$

Moreover, on the set $\{\tau_\lambda < \infty\}$ we have

$$\lim_{t\to\infty} Z_{\tau_\lambda \wedge t} = Z_{\tau_\lambda}$$

while, if $\tau_\lambda = \infty$, we have $\lim\limits_{t\to\infty} Z_{\tau_\lambda \wedge t} = 0$ and an application of dominated convergence theorem leads to

$$\mathbb{E}[\mathbb{1}_{\{\tau_\lambda < \infty\}} Z_{\tau_\lambda}] = 1,$$

that is to say

$$\mathbb{E}\left[\mathbb{1}_{\{\tau_\lambda < \infty\}} \mathrm{e}^{-\frac{1}{2}\nu^2\tau_\lambda}\right] = \mathrm{e}^{-\nu\lambda}.$$

Letting $\nu \downarrow 0$ we get

$$\mathbb{P}(\tau_\lambda < \infty) = 1$$

and also that

$$\mathbb{E}[\mathrm{e}^{-\frac{1}{2}\nu^2\tau_\lambda}] = \mathrm{e}^{-\nu\lambda}.$$

With the change of variable $x = \frac{\nu^2}{2}$ we obtain (5.17). Moreover, the expectation of τ_λ can be obtained by computing the derivative in x of (5.17)

$$\mathbb{E}[\tau_\lambda \mathrm{e}^{-x\tau_\lambda}] = \lambda \frac{\mathrm{e}^{-\sqrt{2x\lambda^2}}}{\sqrt{2x}}$$

and letting $x \downarrow 0$ we get $\mathbb{E}[\tau_\lambda] = +\infty$.

□

Reflection principle

Let $\{B_t,\ t \geq 0\}$ be a standard Brownian motion; for given $\lambda > 0$, let

$$\tau_\lambda = \inf\{t > 0\ :\ B_t \geq \lambda\}$$

be the first passage time in λ. Then we can establish the following relation, known as *reflection principle*:

$$\forall\, \lambda > 0: \qquad \mathbb{P}(\sup_{0\leq t\leq T} B_t \geq \lambda) = \mathbb{P}(\tau_\lambda \leq T) = 2\mathbb{P}(B_T > \lambda). \tag{5.18}$$

Therefore, we also provide the distribution function of the random variable τ_λ:

$$\mathbb{P}(\tau_\lambda \leq t) = \lambda \int_0^t \frac{\mathrm{e}^{-\lambda^2/2s}}{\sqrt{2\pi s^3}}\,\mathrm{d}s.$$

In order to prove (5.18), we introduce some preliminary material. To simplify exposition, we let λ be fixed and suppress the index λ everywhere; recall that $\tau < \infty$ a.s., since $\sup_{n\geq 1} B_n = +\infty$, a.s.

Let $X_t = B_{t\wedge\tau}$ be the Brownian motion absorbed at level λ: on $(\tau \leq t)$ it holds $X_t = \lambda$. Let further $Y_t = B_{t+\tau} - B_\tau$; then by the strong Markov property, $Y = \{Y_t,\, t \geq 0\}$ is a Brownian motion which is independent from $\mathcal{F}_\tau$. A simple application of symmetry implies that also $-Y$ is a Brownian motion. Notice that B_t is equals to

$$\tilde{Z}_t = X_t + Y_{t-\tau}\mathbb{1}_{(t\geq\tau)}, \qquad t \geq 0;$$

define the process

$$Z_t = X_t - Y_{t-\tau}\mathbb{1}_{(t\geq\tau)} = \begin{cases} B_t, & t \leq \tau, \\ 2B_\tau - B_t, & t > \tau \end{cases}, \qquad t \geq 0.$$

Using the properties of the processes Y and $-Y$, it is not difficult to check that

$$\text{the process } \{Z_t,\ t \geq 0\} \text{ is a Brownian motion.} \tag{5.19}$$

This is known as `D. André` reflection principle; it was the first form of strong Markov property for Brownian motion.

We next consider the event $(B_T \le \lambda, \tau \le T)$. Define σ as the first passage time for Z at level λ: it holds that $\sigma = \tau$ (since both depends only on the paths of X_t before reaching λ, and at last both equals the first passage time for the Brownian motion B at level λ). Since the processes B and Z are equal in distribution and $\sigma = \tau$, the probability of $(B_T \le \lambda, \tau \le T)$ coincides with that of $(Z_T \le \lambda, \sigma \le T)$. This implies

$$\mathbb{P}(B_T \le \lambda, \tau \le T) = \mathbb{P}(Z_T \le \lambda, \sigma \le T) = \mathbb{P}(\lambda - Y_{T-\tau} \le \lambda, \sigma \le T) = \mathbb{P}(\lambda \le Y_{T-\tau} + \lambda, \tau \le T)$$
$$= \mathbb{P}(\lambda \le B_T, \tau \le T). \quad (5.20)$$

Problem 5.4. Finish to check relation (5.18). Show that τ_λ has the same distribution as λ^2/N^2, where N is a standard Gaussian distribution.

Hint. The first equality is quite obvious. As far as the second is concerned, notice that

$$(\tau_\lambda \le T) = (B_T > \lambda, \tau_\lambda \le T) \cup (B_T \le \lambda, \tau_\lambda \le T)$$

then apply (5.20).

Notice that this improves estimate (3.7) – a formula which was used in the proof of the Law of the Iterated Logarithm.

Brownian motion on a finite interval

The following construction concerns with a particular class of stopping times defined in terms of a real valued Brownian motion. We consider a family of Brownian motions $B^x = \{B^x_t,\ t \ge 0\}$, starting from $x \in \mathbb{R}$. Let (a, b) be a given finite interval.

We define $\tau^x = \tau^x_{(a,b)}$ be the first exit time from (a, b), i.e.,

$$\tau^x = \inf \{t > 0 \,:\, B^x_t \notin (a, b)\}.$$

It is clear, by continuity of sample paths of the Brownian motion, that $\tau^x = 0$ a.s. for every $x \notin [a, b]$ and $\tau^x > 0$ for every $x \in (a, b)$. It is not clear, a priori, what happens at the interval extremal points.

At the exit time τ^x, B_{τ^x} shall be equal to a or to b, hence we have $\tau^x = \min\{\tau^x_a, \tau^x_b\}$, where τ^x_a is the first passage time from a and τ^x_b is the first passage time from b. Further, if we let $a \to -\infty$ it follows $\lim_{a\to-\infty} \tau_{(a,b)} = \tau_b$ a.s. The following proposition precises the behaviour of the stopping time τ. We shall see that it behaves better than τ_b, compare Theorem 5.14.

Proposition 5.15. τ^x *is a stopping time with finite moments of all orders.*

Proof. We proceed to show that τ^x has moment generating function $M(\lambda) = \mathbb{E}[e^{\lambda\tau^x}]$ that it is finite in a domain $-\varepsilon < \lambda < \varepsilon$, for some $\varepsilon > 0$. This will easily imply that τ^x has finite moments of all orders.

The starting point is the inequality

$$\sup_{x\in(a,b)} \mathbb{P}(\tau^x > 1) \le \sup_{x\in(a,b)} \mathbb{P}(B^x_1 \in (a, b)) := \delta.$$

The number δ can be actually computed, but for our purposes it is sufficient to notice that $0 < \delta < 1$.

The next step uses the Markov property for the Brownian motion; it holds, for $x \in (a, b)$,

$$\begin{aligned}\mathbb{P}(\tau^x > n) &= \mathbb{P}(B_t^x \in (a,b)\,\forall t \in (0,n]) \\ &= \mathbb{P}(B_t^x \in (a,b)\,\forall t \in (0,n-1])\,\mathbb{P}(B_t^x \in (a,b)\,\forall t \in (n-1,n] \mid \mathcal{F}_{n-1})\end{aligned}$$

The first event is equal to $\{\tau^x > n-1\}$ and it is $\mathcal{F}_{n-1}$-measurable, while the second event can be written as $\{\tau^{B^x_{n-1}} > 1\}$[2] and its probability is bounded by δ, hence

$$\mathbb{P}(\tau^x > n) = \mathbb{P}(\tau^x > n-1)\mathbb{P}(\tau^{B^x_{n-1}} > 1) \le \mathbb{P}(\tau^x > n-1)\delta.$$

It follows by induction on n that

$$\mathbb{P}(\tau^x > n) \le \delta^n.$$

By letting $n \to \infty$, we obtain that $\mathbb{P}(\tau^x = \infty) = 0$, i.e., almost surely every path will leave the interval (a, b). Further, for λ small enough such that $e^\lambda \delta < 1$ it is

$$\mathbb{E}[e^{\lambda\tau^x}] \le \sum_{n=1}^\infty e^{\lambda n}\mathbb{P}(\tau^x \in (n-1,n]) \le \sum_{n=1}^\infty e^{\lambda n}\delta^{n-1} < \infty.$$

This means that the moment generating function for τ^x exists and the proof is complete.

□

Problem 5.5. Let us fix $x \in (a, b)$.

(a) We are interested in the exit points from the interval (a, b). Let us denote p_a, respectively p_b, the probability that the exit point is a (resp. b):

$$p_a = \mathbb{P}(B^x_{\tau^x} = a).$$

It is obviously $p_a + p_b = 1$. Use the martingale property of Brownian motion to prove that $ap_a + bp_b = x$, so that we can solve the system for p_a and p_b.

Hint: compute $\mathbb{E}[B^x_{\tau^x}]$.

(b) The next step is to compute $\mathbb{E}[\tau^x]$. Use the martingale property of $(B^x)^2_t - t$ to show that $\mathbb{E}[\tau^x] = a^2 p_a + b^2 p_b - x^2$; this implies $\mathbb{E}[\tau^x] = (x-a)(b-x)$.

Hint: apply Doob's theorem to show that $x^2 = \mathbb{E}[(B^x_{t\wedge\tau})^2 - (t \wedge \tau^x)]$, then let $t \to \infty$.

5.6 Brownian diffusions

In this section, we shall present examples of Markov processes related to Brownian motion, and discuss some of their properties.

In the following, let $\{B_t^x,\ t \ge 0\}$ be a family of real Brownian motions starting from $x \in \mathbb{R}$. Recall the transition probability function $p(t, x, A)$ related to $\{B_t^x,\ t \ge 0\}$

[2] we use Markov property for Brownian motion in the special form given in Remark 5.5(ii)

$$p(t,x,\mathrm{d}y) = \frac{1}{\sqrt{2\pi t}} \exp(-|x-y|^2/2t)\,\mathrm{d}y,$$

The more general brownian kernel $p_\sigma(t,x,y) = \frac{1}{\sqrt{2\sigma^2\pi t}} \exp(-\frac{|x-y|^2}{2\sigma^2 t})$ ($\sigma \neq 0$) is the t.p.f. for $\{\sigma B_t^x,\ t \geq 0\}$, i.e., a Brownian motion with diffusion coefficient σ; note that $p_\sigma(t,x,y)$ is the fundamental solution of the parabolic equation

$$\begin{cases} \frac{\partial}{\partial t} u(t,x) = \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2} u(t,x), & t > 0,\ x \in \mathbb{R} \\ \lim_{t \downarrow 0} p(t,x,\cdot) = \delta_x. \end{cases} \tag{5.21}$$

Hence, we say that $\{\sigma B_t^x,\ t \geq 0\}$ is a *diffusion process* related to the second order operator

$$A = \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2}, \qquad D(A) = C^2(\mathbb{R}).$$

This is a very simple example of diffusion process. The next lecture will provide a precise definition of diffusion processes and a more complete analysis of their properties.

Historical note. Einstein's interpretation of Brownian movement

In this section, we shortly discuss the physical significance of Einstein's result, contained in 1905's paper [Ei05], which had a dramatic impact on modern physics.

At the beginning of 1900s, molecular bombardment was the most likely cause of Brownian movement. The verification was attempted by several researcher, but the results were disappointing: the gap between kinetic energy of the particle and that of the fluid was of the order of a hundred thousand times. The key problem was the measurement of the velocity of a particle undergoing Brownian mouvement, since its trajectory appears highly irregular –like it was a nowhere differentiable curve– how one can compute the tangent at each point? Einstein's approach identified the mean-square displacements of suspended particles rather than their velocities as suitable observable quantities, thus showing a way to circumvent this difficulty.

We start with a suspension of Brownian particles in a fluid, subject to an external force F, in equilibrium. The real form of the force F is unimportant and does not enter in the argument below.

In equilibrium, the force F is balanced by the *osmotic pressure* forces of the suspension

$$F = kT \frac{\nabla \nu}{\nu}, \tag{5.22}$$

where ν is the number of particles per unit volume, T is the absolute temperature and k is Boltzmann's constant (energy per degree).

The Brownian particles in their movement are subject to a resistence due to friction, and the force F gives to each particle a velocity (following *Stokes' law*) of the form $\dfrac{F}{6\pi r \eta}$ where η is the fluid viscosity and r is the *Stokes radius* of the particle. Therefore, $\dfrac{\nu F}{6\pi r\eta}$ is the number of particles passing through a unit area per unit of time due to the presence of the force F. On the other hand, if the

Brownian particles were subject only to diffusion, ν would satisfy the same diffusion equation as (5.21) (notice that the density p is the number density ν divided by the total number of particles)

$$\frac{\partial}{\partial t}\nu = \sigma^2 \frac{\partial^2}{\partial x^2}\nu$$

so that $-\sigma^2 \nabla \nu$ particles pass a unit area per unit of time due to diffusion. In equilibrium, these quantities must sum to zero, so that

$$\frac{\nu F}{6\pi r \eta} = \sigma^2 \nabla \nu. \tag{5.23}$$

Now we can eliminate F and ν between formulae (5.22) and (5.23), thus giving the formula

$$\sigma^2 = \frac{kT}{6\pi r \eta}. \tag{5.24}$$

According to (5.21), the root mean square displacement of a particle is given by $\sqrt{\sigma^2 t}$ (t being the time of observation) and Einstein opened a new promising way for *validating atom's theory*.

`Perrin`'s group started on 1908 to measure the diffusion of the particles, confirming the square root of time law and validating Einstein's kinetic-theory approach. In further experiments over the following five years, Perrin produced a wealth of measurements that could not be contested. His experiments determined Avogadro's number within 19% of currently accepted value (obtained by other means), thus leading Perrin to win the Nobel prize in physics in 1926.

6

Analytic aspects of Markov properties

There are a number of connections between Markov processes and elliptic and parabolic partial differential equations, with particular regard to semigroup theory approach. In this lecture we develop the analytic theory of Markov processes, which was hidden under the probabilistic approach in previous lecture.

We are concerned with a Markov process $X^x = \{X_t^x,\ t \geq 0\}$ with transition function $p(t,x,A)$

$$p(t,x,A) = \mathbb{P}(X_t^x \in A), \tag{6.1}$$

defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, adapted to a filtration $\{\mathcal{F}_t,\ t \geq 0\}$ satisfying the usual assumptions. Our aim is to study a Markov process through its associated transition Markov semigroup $\{T_t,\, t \geq 0\}$ of bounded operators on the space $M_b(E)$ of Borel measurable, bounded functions from E into $\mathbb{R}$

$$T_t : M_b(E) \to M_b(E),\ \ T_t f(x) = \mathbb{E} f(X_t^x).$$

Loosely speaking, if X^x is a Markov process with transition probability function p, then $\{T_t,\, t \geq 0\}$ is a semigroup of linear operators.

Once we have introduced Markov transition semigroups associated to Markov processes, we concentrate on the infinitesimal generator and investigate some of its properties.

Finally, we define the *diffusion processes*, which are – as introduced in last lecture talking about Brownian motion – a class of strong Markov processes with continuous sample paths, whose infinitesimal generator is a second order, possibly degenerate, elliptic differential operator.

Preliminaries

Here we introduce basic functions spaces which will be used in the rest of the lecture; we assume that the reader is familiar with the basis of functional analysis, and as standard reference we use the book of `Rudin`, [Ru91], `Reed and Simon` [ReSi] and `Dudley` [Du02].

We are dealing with real-valued functions defined in a domain E in $\mathbb{R}^n$, endowed with the Borel σ-field. We say that $f : E \to \mathbb{R}$ belongs to $M_b(E)$ if it is Borel measurable and bounded; this space, endowed with the sup-norm $\|f\|_\infty = \sup_{x \in E} |f(x)|$ is a Banach space. Many of the other spaces of functions we shall deal with are Banach spaces; here are some examples

(a) the space $C_b(E)$ of continuous and bounded functions on E, endowed with the norm $\|\cdot\|_\infty$;
(b) the space $C(E)$ of continuous functions on E is a Banach space if and only if E is compact;
(c) the space $C_0(E)$ of continuous functions that vanish at infinity

$$C_0(E) = \{f \in C(E) \,:\, \text{for each } \varepsilon > 0, \text{ the set } \{|f(x)| \geq \varepsilon\} \text{ is compact}\}.$$

6.1 Markov semigroups

¿From the viewpoint of this lecture, transition functions are a convenient way to study Markov processes. In fact, we begin with associating to a transition function a family of bounded linear operators acting on the space $M_b(E)$. As we shall see later, this family forms a semigroup of operators and this fact is connected with the Chapman-Kolmogorov equation for transition functions.

The definition that we introduce below is somewhat more general than the one given in last lecture; in fact, we require only that $p(t, x, E)$ is possibly less than one.

Definition 6.1. *A* time-homogeneous Markov transition function *on E (briefly, a transition function on E or a t.f. on E) is a function $p(t, x, A)$ with $t \geq 0$, $x \in E$ and $A \in \mathcal{E}$, such that*

(a) for fixed t and A, the map: $x \mapsto p(t, x, A)$ is $\mathcal{E}$-measurable;
(b) for fixed t and x, $p(t, x, \cdot)$ is a finite measure on $(E, \mathcal{E})$ such that $p(t, x, E) \leq 1$;
(c) $p(0, x, \cdot) = \delta_x$, $x \in E$;
(d) the Chapman-Kolmogorov equation holds:

$$p(t+s, x, A) = \int_E p(s, y, A)\, p(t, x, \mathrm{d}y), \qquad \forall\, t, s \geq 0,\ x \in E,\ \ A \in \mathcal{E}. \tag{6.2}$$

On the space $M_b(E)$ we define the family of operators

$$T_t f(x) = \int_E f(y) p(t, x, \mathrm{d}y), \qquad f \in M_b(E),\ t \geq 0. \tag{6.3}$$

Condition (a) of Definition 6.1 implies that this operator is well defined from $M_b(E)$ into itself.

Problem 6.1. Show that $T_t : M_b(E) \to M_b(E)$.
Hint: first prove that when $f = 1\!\!1_\Gamma$, $\Gamma \in \mathcal{E}$, then $P_t f \in B_b(E)$ and then use an approximation argument.

Using condition (b) of Definition 6.1, it follows, for each $t \geq 0$, that T_t is a non-negative, contraction linear operator on $M_b(E)$:

$$f \in M_b(E),\ 0 \leq f \leq 1 \text{ on } E \implies 0 \leq T_t f \leq 1 \text{ on } E.$$

Thanks to the Chapman-Kolmogorov condition we have

$$\begin{aligned} T_{t+s} f(x) = \int_E p(t+s, x, \mathrm{d}y)\, f(y) = \int_E \int_E p(t, x, \mathrm{d}z)\, p(s, z, \mathrm{d}y)\, f(y) \\ = \int_E p(t, x, \mathrm{d}z)\, T_s f(z) = T_t(T_s f)(x), \end{aligned}$$

so that the operators $\{T_t,\ t \geq 0\}$ form a semigroup: $T_t T_s = T_{t+s}$, $t, s \geq 0$. We have also, by condition (c) of Definition 6.1, that $T_0 = I$ the identity operator on $M_b(E)$.

Definition 6.2. *The semigroup $T = \{T_t,\ t \geq 0\}$ defined by the transition function p in (6.3) is called a (transition) Markov semigroup on E. If in addition $T_t 1 = 1$, $t \geq 0$, then T is called a conservative Markov semigroup.*

Before proceeding, we remind the fundamental link between t.p.f. and Markov processes. We recall that a t.p.f. (transition probability function) is a t.f. which verifies $p(t, x, E) = 1$, $x \in E$, $t \geq 0$. In literature, a t.f. satisfying this condition is called *conservative*: see also Definition 6.7 below. Clearly, a conservative Markov semigroup is associated to a transition probability function (t.p.f.) p on E.

Theorem 6.3. *To each t.p.f. we can associate a unique Markov process.*

Consider the family of processes X^x associated to p. We have:

$$\mathbb{P}(X_t^x \in \Gamma) = p(t, x, \Gamma).$$

Hence the relation between the semigroup T and X^x is given by

$$T_t f(x) = \mathbb{E}[f(X^x(t))], \qquad f \in M_b(E). \tag{6.4}$$

Note that Markov property can be written, in terms of $\{T_t\}$, as

$$\mathbb{E}(f(X_{t+s}) \mid \mathcal{F}_t) = \mathbb{E}(f(X_{t+s}) \mid X_t) = T_s f(X_t) \tag{6.5}$$

for all $s, t \geq 0$ and $f \in M_b(E)$.

We also say that $\{T_t,\ t \geq 0\}$ is the (conservative) Markov semigroup associated to the Markov process $\{X^x\}$.

6.2 The infinitesimal generator of a Markov semigroup

We consider a Markov semigroup $\{T_t,\ t \geq 0\}$ on the Euclidean space $E \subset \mathbb{R}^d$, $d \geq 1$. The dynamic of this semigroup can be described by means of an operator, which represents the (right)-derivative of the semigroup T_t for $t = 0$. We define the *infinitesimal generator* (briefly, the generator) A for the semigroup $\{T_t,\ t \geq 0\}$ setting

$$Af(x) = \lim_{t \downarrow 0} \frac{T_t f(x) - f(x)}{t} \tag{6.6}$$

for every f in the domain of definition

$$D(A) = \{\, f \in M_b(E) \mid Af(x) \text{ is well defined for each } x \in E \text{ and moreover } \sup_{t>0} \|\frac{T_t f - f}{t}\|_\infty < \infty\}.$$

The operator A is, in general, linear and possibly unbounded. If the semigroup $\{T_t,\ t \geq 0\}$ is associated to the Markov process X^x, $x \in E$, we can formulate (6.6) in an equivalent form

$$Af(x) = \lim_{t \downarrow 0} \frac{\mathbb{E}[f(X_t^x)] - f(x)}{t}. \tag{6.7}$$

Our definition of generator is a variant of the weak generator introduced by Dynkin [Dy65].

Remark 6.4. Given a Markov transition semigroup $\{T_t\}$ on $M_b(E)$, Dynkin introduces the space

$$M_b^0(E) = \{f \in M_b(E) \text{ such that } \lim_{t\to 0} T_t f(x) = f(x), \quad x \in E\}.$$

Moreover he defines the weak generator A of $\{T_t\}$ in the following way. Let

$$D(A) = \big\{f \in M_b(E) \,:\, \text{there exists } g \in M_b^0(E),\ \lim_{t\to 0^+} \tfrac{1}{t}[T_t f(x) - f(x)] = g(x),\ x \in E,$$
$$\text{and } \sup_{t>0} \|\tfrac{1}{t}[T_t f - f]\|_\infty \le M, \text{ for any } t > 0\big\}.$$

Then, for any $f \in D(A)$, $Af(x) := \lim_{t\to 0^+} \frac{1}{t}[T_t f(x) - f(x)]$, $x \in E$.

Example 6.5. As an easy example, let us consider the Markov semigroup associated to the right translation process, and construct the infinitesimal generator of the process.

The (deterministic) right translation process on $\mathbb{R}$ is defined by

$$X^x(t) = x + t, \qquad x \in \mathbb{R}, \quad t \ge 0.$$

This process is associated to the transition semigroup

$$T_t f(x) = f(x+t),$$

The infinitesimal generator is given by

$$Af(x) = \lim_{t\to 0^+} \frac{f(x+t) - f(x)}{t};$$

and we see then that it is identified with the right derivative

$$A = \left(\tfrac{d}{dx}\right)^+$$

for all functions which are Lipschitz continuous.

For every conservative Markov semigroup $\{T_t\}$ on $M_b(E)$ with infinitesimal generator A, the fundamental *maximum principle* holds.

Proposition 6.6. *Let $\{T_t,\, t \ge 0\}$ be a conservative Markov semigroup on E. Let $f \in D(A)$ have an absolute maximum in x (i.e., $f(x) \ge f(y)$ for all $y \in E$, $y \ne x$). Then $Af(x) \le 0$.*

Proof. Since $\{T_t\}$ is conservative we have, for any $t > 0$,

$$\frac{T_t f(x) - f(x)}{t} = \frac{1}{t} \int_E [f(y) - f(x)] p(t, x, \mathrm{d}y) \le 0.$$

□

Problem 6.2. The above result eliminates a number of possible generators. Prove that in $\mathbb{R}$ the operator $Af = f^{(k)}$, the $k-$derivative, can not be a generator of a Markov semigroup if $k \ge 3$.

Problem 6.3. Show that for a Markov semigroup $\{T_t\}$ on $M_b(E)$, the conservative property $T_t 1 = 1$ holds if and only if $1 \in D(A)$ and $A1 = 0$.

There are cases where the transition semigroup $\{T_t\}$ acts naturally on functional spaces other than $M_b(E)$; for instance, the space $C_b(E)$ of bounded continuous functions can be invariant for $\{T_t\}$ and so $\{T_t\}$ acts on $C_b(E)$.

Let L be a closed subspace of $M_b(E)$ which is invariant for $\{T_t\}$, i.e. $T_t(L) \subset L$, $t \geq 0$. In this case, we consider $\{T_t\}$ as a semigroup of linear contractions on L and modify the definition of generator A as follows:

$$\begin{cases} D(A) = \big\{f \in L \,:\, \text{there exists } g \in L \text{ such that } \sup\limits_{t>0} \frac{1}{t}\|T_t f - f\|_\infty < \infty \\ \qquad\qquad \text{and } \lim\limits_{t \to 0^+} \frac{1}{t}(T_t f(x) - f(x)) = g(x),\ x \in E\big\}, \\ Af(x) = \lim\limits_{t \to 0^+} \frac{1}{t}(T_t f(x) - f(x)), \quad f \in D(A),\ x \in E. \end{cases} \tag{6.8}$$

In other words, we are considering the part of A in L.

In the proceeding of this lecture, we are interested in characterizing the infinitesimal generator of a Markov process in terms of its transition function and/or its transition semigroup. Our construction will culminate in a definition of diffusion processes in terms of the infinitesimal generator; this will open the way to the inverse problem of characterize the Markov process associated to a given (second order, differential) operator.

6.2.1 Example: Poisson process

In this section we take as example the real Poisson process $\{N_t,\ t \geq 0\}$ introduced in first lecture. We recall that a Poisson process has indepedent and stationary increments, then $N_{t+s} = (N_{t+s} - N_t) + N_t$ is the sum of independent random variables, with Poisson distributions of parameter λt and λs, respectively; then we obtain (see Exercise 5.2) that $x + N_t$ is Markov with transition probability function

$$p(t, x, A) = \mathbb{P}(x + N_t \in A).$$

For any $t \geq 0$ we consider the operator T_t acting on Borel and bounded functions $u(x)$ by

$$T_t u(x) = \mathbb{E}[u(x + N_t)].$$

As a simple exercise, let us show directly that $\{T_t\}$ satisfies the semigroup law; the idea is to use the freezing lemma as we have done for Brownian motion. Later we will compute its infinitesimal generator.

The semigroup property is

$$\begin{aligned} \mathbb{E}[u(x + N_{t+s})] = \mathbb{E}[\mathbb{E}(u(x + N_{t+s} - N_t) + N_t) \mid \mathcal{F}_t)] &= \mathbb{E}[\mathbb{E}[u(x + y + N_t)]\Big|_{y = N_{t+s} - N_t}] \\ &= \mathbb{E}[\mathbb{E}[u(x + y + N_t)]\Big|_{y = N_s}] = T_s T_t u(x) \end{aligned}$$

where we use respectively independence, Lemma 5.2 and stationarity of increments.

For a Poisson process, the infinitesimal generator can be given explicitly; it holds, for any $u \in M_b(\mathbb{R})$

$$\lim_{\varepsilon\downarrow 0}\frac{T_\varepsilon u(x)-u(x)}{\varepsilon}=\lim_{\varepsilon\downarrow 0}\sum_{k=0}^{\infty}\frac{u(x+k)-u(x)}{\varepsilon}\mathrm{e}^{-\varepsilon\lambda}\frac{(\varepsilon\lambda)^k}{k!}$$
$$=\lim_{\varepsilon\downarrow 0}\sum_{k=1}^{\infty}\frac{u(x+k)-u(x)}{\varepsilon}\mathrm{e}^{-\varepsilon\lambda}\frac{(\varepsilon\lambda)^k}{k!}=\lambda(u(x+1)-u(x)).$$

Therefore A is a difference operator, non local, which maps $u(x)$ into $\lambda(u(x+1)-u(x))$. The operator A is a bounded operator on $M_b(\mathbb{R})$.

6.2.2 Sub-markovian processes

Heuristically, the class of processes that we aim to introduce is given by those stochastic processes $X=\{X^x_t,\ t\geq 0\}$, $x\in E$, which take values in an Euclidean space E only up to some random time τ^x (for instance, one can think to a solution of a differential equation which "explodes" in a finite time). We have

$$\tau^x=\inf\{t\geq 0\,:\,X^x_t\notin E\}$$

with the usual convention that $\tau^x(\omega)=\infty$ if the infimum is taken on an empty set. τ^x is called *lifetime* of the process because it is the instant in which the process ceases to be live in E. Probabilists *kill* the process once its clock has reached τ^x.

The above situation can be rigorously described if we introduce an extra point ∞ attached to E either as an isolated point or as the one-point compactification of E. Thus we denote $E_\infty=E\cup\{\infty\}$, where $\infty\notin E$, endowed with the σ-field $\mathcal{E}_\infty=\sigma(\mathcal{E},\{\infty\})$, and consider the process $\{X_t\}$ taking values in E_∞, compare with `Ikeda and Watanabe` [IW81, Chapter IV].

If E is locally compact, then ∞ is the "point at infinity" and a system of neighborhoods of ∞ is given by the complements of compact sets in E. If E is compact, then ∞ is an isolated point in E_∞. Any real valued function f on E can be extended to $E_\infty=E\cup\{\infty\}$ by setting $f(\infty)=0$.

A family of processes $X=\{X^x\}$, starting from $x\in E$, *but* taking values in E_∞ is called a *sub-markovian process* with transition function p if there exists a transition function $p(t,x,A)$ on E associated with X^x by the formula

$$p(t,x,A)=\mathbb{P}(X^x_t\in A),\qquad A\in\mathcal{E},\quad t\geq 0,\quad x\in E, \tag{6.9}$$

and such that

$$\mathbb{P}(X^x_{t+s}\in A\mid\mathcal{F}_s)=p(t,X^x_s,A),\qquad \mathbb{P}-a.s..$$

It is possible to extend the transition function p, originally given on E, and obtain a t.p.f. on E_∞ as follows:

$$\begin{cases}p'(t,x,A)=p(t,x,A), & t\geq 0,\quad x\in E,\quad A\in\mathcal{E}\\ p'(t,x,\{\infty\})=1-p(t,x,E), & x\neq\infty\\ p'(t,\infty,E)=0, & p'(t,\infty,\{\infty\})=1.\end{cases}$$

The associated Markov process on $(E_\infty,\mathcal{E}_\infty)$ shall be denoted by $\bar X=\{\bar X^x_t,\ t\geq 0\}$, $x\in E_\infty$. The point ∞ is an *absorbing state* for the process $\bar X$, i.e.,

$$\{\bar X^x_s=\infty\}\subset\{\bar X^x_t=\infty\ \forall\, t\geq s\}.$$

Notice that $X^x_t=\bar X^x_t$ on $\{\tau^x>t\}$.

Problem 6.4. Let $B = \{B_t^x,\ t \geq 0\}$ be a family of real Brownian motions starting from $x \in \mathbb{R}$. For any $x > 0$, denote τ_0^x the first passage time from the level 0 of the Brownian motion B^x. Consider the process $\{W_t^x,\ t \geq 0\}$, $x \in (0,\infty)$, defined by

$$W_t^x = B_{t\wedge\tau_0^x}^x = \begin{cases} x + B_t, & t \leq \tau_0 \\ 0, & t \geq \tau_0. \end{cases}$$

This process is called the Brownian motion *absorbed at the origin.*

1. Prove that W is a Markov process with transition function

$$p_-(t,x,\mathrm{d}y) = [p(t,x,y) - p(t,x,-y)]\,\mathrm{d}y,$$

   $x > 0$, $y > 0$.
2. Notice that for fixed $x > 0$, $\{p_-(t,x,\cdot),\ t \geq 0\}$ is a family of measures on $E = (0,\infty)$. We ask to compute the total mass of the measure

$$\int_E p_-(t,x,\mathrm{d}y).$$

Hint: use reflection principle to show that for any $\Gamma \subset (0,\infty)$, $\mathbb{P}(B_t^x \in \Gamma) = \mathbb{P}(B_t^x \in \Gamma,\ B_s^x > 0 \text{ for } 0 \leq s \leq t) + \mathbb{P}(B_t^x \in -\Gamma)$.

The process W on $E = (0,+\infty)$ is a first example of properly sub-markovian processes, i.e., non-conservative stochastic processes. In this case, the *cemetary point* is equal to the origin 0. Indeed we have the following definition.

Definition 6.7. *The sub-Markovian process $X = \{X^x\}$, $x \in E$, will be called* conservative *if and only if the lifetime τ^x is almost surely $+\infty$, for any $x \in E$.*

Hence a conservative process never reaches the attached point ∞ (it is strictly Markovian).

We have already defined a conservative transition semigroup; let us see that these definitions are equivalent, by proving that

$$\mathbb{P}(\tau^x = \infty) = 1 \quad \text{if and only if} \quad T_t\mathbb{1} = \mathbb{1},\ t \geq 0. \tag{6.10}$$

This can be shown easily:

$$1 \geq T_t\mathbb{1}_E(x) = \mathbb{E}[\mathbb{1}_E(X_t^x)] = \mathbb{P}(X_t^x \in E) = \mathbb{P}(\tau^x > t) \geq \mathbb{P}(\tau^x = \infty)$$

and passing to the limit

$$\mathbb{P}(\tau^x = \infty) = \lim_{t\to\infty} \mathbb{P}(\tau^x > t) = \lim_{t\to\infty} T_t\mathbb{1}_E(x).$$

6.3 Feller processes and Feller semigroups

In this section we turn our attention on a special class of Markov processes, named after `W. Feller`. A process X belongs to this class if the transition semigroup on $M_b(E)$ leaves $C_b(E)$ invariant. This also implies some nice continuity properties; it will not be surprising to find that Brownian Motion is a prototype of such processes.

Definition 6.8. *Let $X = \{X_t^x,\ t \ge 0\}$, $x \in E$, be a sub-Markovian process with transition function $p(t,x,\Gamma)$. We say that X (or equivalently p) has the* Feller property *if, for any $f \in C_b(E)$, $t \ge 0$, the function*

$$x \mapsto \mathbb{E}[f(X_t^x)] = \int_E f(y)\, p(t,x,\mathrm{d}y) = T_t f(x)$$

belongs to $C_b(E)$ and moreover, for any $f \in C_b(E)$, the mapping $t \mapsto T_t f(x)$ is continuous on $[0,\infty)$. The Markov semigroup associated to X (or p), i.e. $\{T_t,\ t \ge 0\}$, is called a Feller semigroup.

The right hand side of the above equality is –by definition– the transition semigroup associated to the process X. Feller property implies that given a function $f \in C_b(E)$, then the mapping $T_t f(x)$ shall be continuous with respect to x for each $t \ge 0$; if $x_n \to x_0$ then it must hold

$$\int_E p(t,x_n,\mathrm{d}y) f(y) \longrightarrow \int_E p(t,x_0,dy) f(y), \quad f \in C_b(E).$$

Therefore, an equivalent formulation of the Feller property is that the measure valued mapping $x \mapsto p(t,x,\cdot)$ is weakly continuous. Altough this may seem a natural condition, one can easily convince itself that it does not necessarily hold.

Example 6.9. Construct a Markov process that does not verify Feller condition.

Proof. Consider the case $E = \mathbb{R}$; we define a family of stochastic processes $\{X_t^x,\ t \ge 0,\ x \in \mathbb{R}\}$ setting:

- if $x < 0$, then $X_t^x = x - t$ is the left translation process; also, if $x > 0$, then process defines a right translation: $X_t^x = x + t$;
- if $x = 0$, the process has the same probability to leave to the left or to the right:

$$\mathbb{P}(X_t^0 = t) = \mathbb{P}(X_t^0 = -t) = \frac{1}{2}.$$

Then, we have

$$T_t f(x) = \mathbb{E}[f(X_t^x)] = \begin{cases} f(x+t), & x > 0 \\ \frac{1}{2}[f(t) + f(-t)], & x = 0 \\ f(x-t), & x < 0. \end{cases}$$

Now one can easily verifies that fot arbitrary functions $f \in C_b(\mathbb{R})$ the function $T_t f(x)$ is discontinuous in $x = 0$.

□

Under some additional mild assumptions on regularity of paths, each Feller process is a strong Markov process.

Theorem 6.10. *Let* $\{X^x_t,\ t \geq 0\}$ *be a Feller processes associated to a transition probability function* $p(t,x,A)$*. Assume that the process has right continuous paths. Then, for each stopping time* $\tau < +\infty$*, it holds*

$$\mathbb{E}(f(X^x(t+\tau)) \mid \mathcal{F}_\tau) = T_t f(X^x_\tau). \tag{6.11}$$

Proof. Let us drop the dependence on x. We approximate τ with a sequence of stopping times taking values in the discrete set of dyadic numbers, setting

$$\tau_n = \frac{[2^n \tau] + 1}{2^n}.$$

As $n \to \infty$ $\tau_n \downarrow \tau$. It is

$$\mathcal{F}_\tau = \mathcal{F}_{\tau+} = \bigwedge_{n=1}^{\infty} \mathcal{F}_{\tau_n}.$$

For every time $d \in D$ and for every $A \in \mathcal{F}_\tau$, the event $A_d = A \cap \{\tau_n = d\}$ belongs to $\mathcal{F}_d$, and Markov property implies

$$\mathbb{E}[\mathbb{1}_{A_d} f(X_{d+u})] = \mathbb{E}[\mathbb{1}_{A_d} T_u f(X_d)]$$

Further, since τ_n takes a finite number of values, and $\tau_n = d$ on A_d, we have

$$\mathbb{E}[\mathbb{1}_A f(X_{\tau_n+u})] = \sum_{d \in D} \mathbb{E}[\mathbb{1}_{A_d} f(X_{d+u})] = \sum_{d \in D} \mathbb{E}[\mathbb{1}_{A_d} T_u f(X_d)] = \mathbb{E}[\mathbb{1}_A T_u f(X_{\tau_n})].$$

Since both f and $T_u f$ are continuous functions, and X_t has right continuous paths, passing to the limit as $n \to \infty$ we have

$$\mathbb{E}[\mathbb{1}_A f(X_{\tau+u})] = \mathbb{E}[\mathbb{1}_A T_u f(X_\tau)]$$

Now, using definition of conditional expectation, since $A \in \mathcal{F}_\tau$ is arbitrary, the above equation implies the thesis.

□

We leave as an exercise the proof of the following corollary.

Corollary 6.11. *If the stopping time* τ *takes values in a discrete set (possibly infinite), then property (6.11) holds for every Markov process* $\{X_t,\ t \geq 0\}$*.*

Problem 6.5. Prove Corollary 6.11.

Recall the definition of infinitesimal generator (6.8) (for the present case where $L = C_b(E)$)

$$Af(x) = \lim_{t \downarrow 0} \frac{T_t f(x) - f(x)}{t}$$

for every f in the domain of definition

$$D(A) = \{\, f \in C_b(E) \mid Af(x) \text{ is well defined for each } x \text{ and } \sup_{t>0} \frac{1}{t} \|T_t f - f\| < \infty\}.$$

The generator of a Feller semigroup has some useful properties as it is shown in the next results.

Proposition 6.12. *Let $\{T_t,\, t \geq 0\}$ be a Feller semigroup on E. Then $f \in D(A)$ if and only*

$$T_t f(x) - f(x) = \int_0^t T_s A f(x)\,\mathrm{d}s, \qquad t \geq 0,\ x \in E. \tag{6.12}$$

Moreover if $f \in D(A)$ then $Af = g$ in (6.12).

Proof. It is clear that if (6.12) holds then $f \in D(A)$ and $Af = g$. Let us prove the converse.

Remark that $f \in D(A)$ implies that $T_t f \in D(A)$, $t \geq 0$. Indeed let $\Delta_h = h^{-1}(T_h - I)$, for $h > 0$. There exists $M > 0$ such that, for any $t \geq 0$, $h > 0$,

$$\|\Delta_h(T_t f)\|_\infty = \|T_t(\Delta_h f)\|_\infty \leq \|\Delta_h f\|_\infty \leq M.$$

Moreover

$$\lim_{h\to 0^+} \Delta_h(T_t f)(x) \ = \lim_{h\to 0^+} T_t(\Delta_h f)(x) = \ T_t A f(x),\ \ x \in E.$$

Thus $T_t f \in D(A)$ and moreover $AT_t f = T_t A f$.

Now fix $x \in E$ and consider the mapping: $t \mapsto T_t f(x)$. By the previous argument, there exists the right derivative $\frac{d^+}{dt} T_t f(x) = T_t A f(x)$ at any $t \geq 0$. Moreover the mapping: $t \mapsto T_t A f(x)$ is continuous on $[0, +\infty[$. By a well known lemma of real analysis, we obtain the assertion.
□

Remark 6.13. The previous proof shows also that if $f \in D(A)$, where A is the generator of a Feller semigroup $\{T_t,\, t \geq 0\}$ on $\mathbb{R}^d$, then it holds:

$$T_t f(x) - f(x) = \int_0^t A T_s f(x) ds,\ \ t \geq 0,\ x \in \mathbb{R}^d. \tag{6.13}$$

Hence $u(t,x) := T_t f(x)$ "solves" to the following parabolic Cauchy problem

$$\begin{cases} \partial_t u(t,x) = Au(t,x), & t > 0, \\ u(0,x) = f(x), & x \in \mathbb{R}^d. \end{cases}$$

We will return on this connection between Feller semigroups and parabolic Cauchy problems.

Theorem 6.14. *Let $\{T_t,\, t \geq 0\}$ be a Feller semigroup on E associated to a Feller process $X^x = \{X_t^x,\, t \geq 0\}$, $x \in E$. Then $f \in D(A)$ if and only if there exists $g \in C_b(E)$ such that*

$$M_{t,x}^f = f(X_t^x) - f(x) - \int_0^t g(X_s^x)\mathrm{d}s \qquad \text{is a } \mathbb{P}\text{-martingale} \tag{6.14}$$

(with respect to the same filtration $\{\mathcal{F}_t\}$ of $\{X_t^x\}$) for any $x \in E$. Moreover if $f \in D(A)$ then $Af = g$ in (6.14).

Proof. Assume first that (6.14) holds. Applying the expectation, we have

$$0 = \mathbb{E}\Big[f(X_t^x) - f(x) - \int_0^t g(X_s^x)\mathrm{d}s\Big] = T_t f(x) - f(x) - \int_0^t T_s g(x)\mathrm{d}s,$$

$t \geq 0$, $x \in E$. By Proposition 6.12 we get that $f \in D(A)$ and $Af = g$.

Let now $f \in D(A)$. We have, for $0 \leq s < t$, using the transition function p,

$$\begin{aligned}\mathbb{E}(M_{t,x}^f - M_{s,x}^f \mid \mathcal{F}_s) &= \mathbb{E}\big((f(X_t^x) - f(X_s^x) - \int_s^t g(X_r^x)\,\mathrm{d}r) \mid \mathcal{F}_s\big) \\ &= -f(X_s^x) + \int_E p(t-s, X_s^x, \mathrm{d}y)\, f(y) - \int_s^t \big(\int_E p(r-s, X_s^x, \mathrm{d}y)\, g(y)\big)\,\mathrm{d}r \\ &= T_{t-s} f(X_s^x) - f(X_s^x) - \int_0^{t-s} T_u g(X_s^x)\mathrm{d}u\end{aligned}$$

(to prove that $\mathbb{E}\big((\int_s^t g(X_r^x)\mathrm{d}r) \mid \mathcal{F}_s\big) = \int_s^t \big(\int_E p(r-s, X_s^x, \mathrm{d}y)\, g(y)\big)\,\mathrm{d}r$ we may use an approximation argument by means of Riemann sums). Now, since $f \in D(A)$, by (6.12) we deduce that $\mathbb{E}(M_{t,x}^f - M_{s,x}^f \mid \mathcal{F}_s) = 0$. This completes the proof.
□

Problem 6.6. Prove that the Poisson process $N^x = \{N_t + x,\, t \geq 0\}$ (see Section 6.2.1) starting from $x \in \mathbb{R}$ and a d-dimensional Brownian motion $B^x = \{B_t + x,\, t \geq 0\}$ starting from $x \in \mathbb{R}^d$ are both Feller processes.

Problem 6.7. Let $\{T_t,\ t \geq 0\}$ be the Feller semigroup associated with a Brownian motion B^x, taking values in $\mathbb{R}^d$, $x \in \mathbb{R}^d$, i.e., $\{T_t\}$ is the classical *heat semigroup*. Show that $C_c^2(\mathbb{R}^d) \subset D(A)$ and also that $Af = \frac{1}{2}\Delta f$, $f \in C_c^2(\mathbb{R}^d)$.

6.3.1 Feller semigroups: another approach

Let E_∞ be the one-point compactification of E. We define the space of functions $C_\infty(E)$ as the subspace of $C_b(E)$ which consists of all functions satisfying $\lim_{x\to\infty} f(x) = 0$. It turns out that $C_\infty(E)$ is a closed subspace of $C_b(E)$, and moreover the two spaces can be identified if E is compact. Note also that $C_\infty(E)$ can be identified with $C_0(E)$.

Definition 6.15. *We say that a Feller semigroup $T = \{T_t,\, t \geq 0\}$ on E has the C_∞-property (or that the corresponding transition function p has the C_∞-property) if the space $C_\infty(E)$ is an invariant subspace for T.*

The literature is somehow discording about the precise definition of the family of Feller semigroups. We shall mention, for instance, Ethier and Kurtz [EK86], Schilling [Sc98], Jacob [Ja01] or Durrett [Du96], where Feller property means a transition function having the C_∞-property. Following the literature, we shall call the Feller processes with the C_∞-property *Feller-Dynkin processes*.

Notice further that there is no doubt that a *strong Feller semigroup* should map the bounded measurable functions $M_b(E)$ into $C_b(E)$.

In this section we study some properties of C_∞-transition functions. We begin by proving that C_∞-property implies that the corresponding Markov semigroup $\{T_t,\, t \ge 0\}$ is strongly continuous on the space $C_\infty(E)$ (endowed with the sup norm).

Proposition 6.16. *A Feller semigroup $\{T_t,\ t \ge 0\}$ which leaves $C_\infty(E)$ invariant is strongly continuous when acts on $C_\infty(E)$.*

Proof. In our assumptions, the following pointwise convergence holds:

$$\forall\, f \in C_\infty(E),\ \forall\, x \in E\ :\ T_t f(x) \to f(x) \quad \text{as } t \to 0^+. \tag{6.15}$$

Note that the topological dual of $C_\infty(E)$ is the space of all finite Borel signed measures on E (endowed with the variation norm). Hence (6.15) implies that

$$\int_E T_t f(x)\mu(dx) \to \int_E f(x)\mu(dx) \quad \text{as } t \to 0^+,$$

for any μ finite Borel signed measure on E. We have just proved that weakly $T_t f \to f$ as $t \to 0^+$. By a well known result from semigroup theory (called "weak=strong", see Pazy [Pa83, Theorem 2.1.3]), this implies that $\{T_t\}$ is strongly continuous on $C_\infty(E)$.
□

The following result shows in particular that any *conservative* Feller-Dynkin semigroup is also a Feller semigroup; for a proof, see Schilling [Sc98, Corollary 3.4].

Theorem 6.17. *Every Feller-Dynkin semigroup $\{T_t,\ t \ge 0\}$ on E satisfying $T_t 1 \in C_b(E)$, $t \ge 0$, is ** a Feller semigroup.*

Remark 6.18. In the following lectures, we shall consider Feller processes arising from stochastic differential equations. In such situation the additional C_∞-invariance property is not always true and, even if it holds, it is not easy to prove. A discussion of this problem can be found in Ethier and Kurtz [EK86, page 373] and in the survey paper by Metafune, Pallara and Wacker [MPW02]. This motivates our choice of Feller processes in Definition 6.8.

In the next proposition, we show that Feller processes with the C_∞-property are *uniformly stochastically continuous* , i.e.,

for every compact set $K \subset E$ we have

$$\lim_{t\to 0} \sup_{x \in K} [1 - p(t, x, D_\varepsilon(x))] = 0 \tag{6.16}$$

for every $\varepsilon > 0$, where $D_\varepsilon(x)$ is the open ball of radius ε around x: $D_\varepsilon(x) = \{y \in E\ :\ |x-y| < \varepsilon\}$.

Proposition 6.19. *If p is a Feller transition function with the C_∞-property, then p is uniformly stochastically continuous.*

Proof. Let K be a compact set in E and define the family of mappings

$$f(x) = \begin{cases} 1 - \frac{1}{\varepsilon}|x-y| & \text{if } |x-y| < \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

Then f_x belongs to $C_c(E)$ and

$$\|f_x - f_z\|_\infty \le \frac{1}{\varepsilon}|x-z|. \tag{6.17}$$

Since K is compact, there are a finite number of points $\{x_k\}$ such that

$$K = \bigcup_{k=1}^n B_{\delta\varepsilon/4}(x_k);$$

hence every point in E verifies

$$\min_{k\le n} |x, x_k| < \delta\varepsilon/4.$$

Thus we obtain a similar bound for the family of functions f_x:

$$\min_{k\le n} \|f_x - f_{x_k}\|_\infty \le \delta/4. \tag{6.18}$$

Notice that

$$p(t,x,U_\varepsilon(x)) = \int_E p(t,x,\mathrm{d}y)\, \mathbb{1}_{U_\varepsilon(x)}(y) \ge \int_E p(t,x,\mathrm{d}y)\, f_x(y) = T_t f_x(x)$$

hence

$$1 - p(t,x,U_\varepsilon(x)) \le f_x(x) - T_t f_x(x);$$

it remains to prove that the right hand side converges to 0 as $t \to 0$ uniformly for $x \in K$, that is, there exists t_0 such that

$$|f_x(x) - T_t f_x(x)| \le \delta \qquad \forall\, x \in K,\ t < t_0;$$

take the sup-norm and observe that

$$\|f_x - T_t f_x\| \le \|f_x - f_{x_k}| + |f_{x_k} - T_t f_{x_k}| + |T_t f_{x_k} - T_t f_x|$$

and notice that if we choose witly x_k, the first and the third terms are bounded by the sup-norm, and that T_t is a contraction, hence, using (6.18),

$$\|f_x - T_t f_x\| \le \delta/2 + \|f_{x_k} - T_t f_{x_k}\|$$

but since $\{T_t\}$ is strongly continuous on $C_\infty(E)$, the last term is bounded by $\delta/2$ for sufficiently small t, from where the thesis follows.
□

Remark 6.20. It is possible to prove (see for instance `Ventsel` [Ve83, Theorem 9.3]) that a Markov process with a uniformly stochastically continuous transition probability function possesses a càdlàg version (right continuous sample paths with left limits).

Other results for regularity of paths of Markovian processes associated with t.p.f. will be given later; however, they are not of the upmost importance for us, since in case of solutions of stochastic differential equations we will prove such results directly.

6.4 Further properties of the infinitesimal generator

Let us briefly explain the content of this section, by appealing to the examples discussed so far.

If a process, like the Brownian motion, has continuous paths, it is natural to think that X_t^x is near x for small times t, so that $Af(x) = \lim_{t \to 0} \frac{1}{t}\mathbb{E}[f(X_t^x) - f(x)]$ is a local operator, i.e., it depends only on the values of f in a neighborhood of x. Conversely, if we analyze a jump process like the Poisson process, then the increments depend on the values of X_t^x near x but also near the possible points where the process have moved, which may be far from x: so the operator A is not local. Most important here is the fact that we can provide a condition that implies, at once, the continuity of the sample paths of X – Theorem 6.25 – and that the infinitesimal generator is local – Proposition 6.21.

We say that an operator $A : D(A) \subset M_b(E) \to M_b(E)$ is *local* if the value of $Af(x)$ depends only on the behaviour of f in a neighborhood of x. The next result provides a sufficient condition for the infinitesimal generator A associated to a Markov process X to be local.

Proposition 6.21. *Let p be a transition probability function. Assume that for $x \in E$*

$$\lim_{t \downarrow 0} \frac{1}{t} p(t, x, B_R^c(x)) = 0 \qquad \text{for every } R > 0. \tag{6.19}$$

Then A is a local operator.

Proof. Take $f \in D(A)$; then $Af(x)$ is given by

$$Af(x) = \lim_{t \downarrow 0} \frac{T_t f(x) - f(x)}{t} = \lim_{t \downarrow 0} \frac{1}{t} \int_E [f(y) - f(x)]\, p(t, x, \mathrm{d}y).$$

We split the integral in two parts, so that

$$\frac{1}{t} \int_E [f(y) - f(x)]\, p(t, x, \mathrm{d}y) = \frac{1}{t} \int_{B_R(x)} [f(y) - f(x)]\, p(t, x, \mathrm{d}y) + \frac{1}{t} \int_{B_R^c(x)} [f(y) - f(x)]\, p(t, x, \mathrm{d}y)$$

and use (6.19) in the last integral

$$\frac{1}{t} \int_{B_R^c(x)} [f(y) - f(x)]\, p(t, x, \mathrm{d}y) \le \frac{1}{t} \|f\| p(t, x, B_R^c(x)) \xrightarrow{t \to 0} 0$$

which shows that $Af(x)$ is well defined if and only if the limit

$$\lim_{t \downarrow 0} \frac{1}{t} \int_{B_R(x)} [f(y) - f(x)]\, p(t, x, \mathrm{d}y)$$

exists and, finally, this limit depends only on the behaviour of f in a neighbourhood of x.

□

Under condition (6.19), we may characterize the form of the infinitesimal generator A of the Markovian semigroup P_t. Let us define the space $C_c^2(E)$ as the completition of $C^2(E) \cap C_c(E)$ with respect to the norm

$$\|f\| + \sum_{i=1}^d \|\frac{\partial f}{\partial x_i}\| + \sum_{i,j=1}^d \|\frac{\partial^2 f}{\partial x_i \partial x_j}\|.$$

Proposition 6.22. *Given a Markov process associated to a (conservative) Markov semigroup* $\{T_t\}$ *with infinitesimal generator* A*, assume that condition (6.19) holds, and that* $C^2_c(E) \subset D(A)$*. Then*

$$Af(x) = \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} f(x) + \sum_{k=1}^d b_k(x) \frac{\partial}{\partial x_k} f(x) \tag{6.20}$$

for all $f \in C^2_c(E)$*, and the matrix* $\big(a_{ij}(x)\big)$ *is of positive type.*

A matrix $a = \big(a_{ij}\big)$ is of *positive type* if for any complex vector $\zeta \in \mathbb{C}^d$ it holds

$$\langle a\zeta, \bar\zeta\rangle = \sum_{i,j=1}^d a_{ij} \zeta_i \bar\zeta_j \ge 0.$$

However, the operator A needs not to be elliptic, since the matrix a may be singular.

Proof. The idea is that in our assumptions, A is a local operator such that the maximum principle holds; we aim to prove that the operator A can be at most of second order. Let $f \in C^2_c(\mathbb{R}^d)$ be such that f and its first and second derivatives vanishes in x; we claim that $Af(x) = 0$. This implies that $Af(x)$ is of the form (6.20) for certain real numbers $a_{ij}(x)$, $b_i(x)$ (since if A would involve higher order derivatives, they will give a non zero contribution to $Af(x)$); further, with no loss of generality we can assume that the a_{ij} are symmetric.

Let us prove the claim. Let $g \in C^2_c(E)$ be such that $g(y) \ge 0$ and $g(y) = |y-x|^2$ in a neighbourhood of x; for $\varepsilon > 0$ let $U_\varepsilon(x)$ be a small ball surrounding x such that

$$y \in U_\varepsilon(x) \Longrightarrow |f(y)| \le \varepsilon g(y).$$

Using (6.19) we have

$$\begin{aligned} Af(x) = \lim_{t\downarrow 0} \frac{1}{t} \int f(y) p(t,x,\mathrm{d}y) = \lim_{t\downarrow 0} \frac{1}{t} \int_{U_\varepsilon(x)} f(y) p(t,x,\mathrm{d}y) \\ \le \lim_{t\downarrow 0} \frac{1}{t} \int_{U_\varepsilon(x)} \varepsilon g(y) p(t,x,\mathrm{d}y) = \varepsilon A g(x). \end{aligned}$$

Since ε is arbitrary, the claim $Af(x) = 0$ follows.

Now if $f \in C^2_c(E)$ and $f(x) = 0$ then the positivity of the semigroup implies

$$Af^2(x) = \lim_{t\downarrow 0} \frac{1}{t} \mathbb{E}[T_t f^2(x)] \ge 0.$$

On the other hand, formula (6.20) together with $f(x) = 0$ implies

$$\begin{aligned} Af^2(x) &= \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} f^2(x) + \sum_{k=1}^d b_k(x) \frac{\partial}{\partial x_k} f^2(x) \\ &= 2 \sum_{i,j=1}^d a_{ij}(x) \frac{\partial}{\partial x_i} f(x) \frac{\partial}{\partial x_j} f(x). \end{aligned}$$

Then, since we can choose $\xi_i = \frac{\partial}{\partial x_i} f(x)$ to be arbitrary real numbers, and since $a(x)$ is a real, symmetric matrix, it follows that it is of positive type.

□

Proposition 6.23. *Given a Markov process associated to a (conservative) transition semigroup $\{T_t\}$ with infinitesimal generator A such that $C_c^2(E) \subset D(A)$, assume that condition (6.19) holds. Assume further that, for $x \in E$, the following limits exist for arbitrary $R > 0$:*

$$\lim_{t\downarrow 0} \frac{1}{t} \int_{B_R(x)} (y_i - x_i) p(t, x, \mathrm{d}y) = b_i(x) \quad \textit{(the drift coefficients)} \tag{6.21}$$

$$\lim_{t\downarrow 0} \frac{1}{t} \int_{B_R(x)} (y_i - x_i)(y_j - x_j) p(t, x, \mathrm{d}y) = a_{ij}(x) \quad \textit{(the diffusion coefficients).} \tag{6.22}$$

Then these limits identify the infinitesimal generator A in (6.20) for all $f \in C_c^2(E)$.

Proof. Using (6.19) we get

$$\lim_{t\to 0^+} \frac{1}{t}[T_t f(x) - f(x)] = \lim_{t\to 0^+} \frac{1}{t} \int p(t, x, \mathrm{d}y)(f(y) - f(x))$$
$$= \lim_{t\to 0^+} \frac{1}{t} \int_{B_R(x)} p(t, x, \mathrm{d}y)(f(y) - f(x)).$$

Using second order Taylor's formula for f

$$f(y) = f(x) + \sum_{k=1}^{d} (y_k - x_k) \frac{\partial}{\partial x_k} f(x) + \frac{1}{2} \sum_{i,j=1}^{d} (y_i - x_i)(y_j - x_j) \frac{\partial^2}{\partial x_{ij}^2} f(x) + o(|x - y|^2)$$

we obtain

$$\lim_{t\to 0^+} \frac{1}{t}[T_t f(x) - f(x)] = Af(x) + \lim_{t\to 0^+} \frac{1}{t} \int_{B_R(x)} p(t, x, \mathrm{d}y)\, o(|x - y|^2).$$

This last limit equals zero, since

$$\max_{t\to 0^+}\lim \frac{1}{t} \int_{B_R(x)} p(t, x, \mathrm{d}y)\, o(|x - y|^2)$$
$$\leq \max_{t\to 0^+}\lim \frac{o(R^2)}{R^2} \frac{1}{t} \int_{B_R(x)} p(t, x, \mathrm{d}y)\, |x - y|^2 = \frac{o(R^2)}{R^2} \sum_{i=1}^{d} a_{ii}(x)$$

which concludes the proof, since R is arbitrary.

□

Remark 6.24. The drift and diffusion coefficients can be interpreted heuristically in the following manner: if $d = 1$, then for $f(x) = x$ we obtain from (6.7)

$$\mathbb{E}[X_t - x] = tb(t) + o(t)$$

while for $g(x) = x^2$,

$$\mathbb{E}[(X_t - x)^2] = ta(x) + o(t)$$

as $t \downarrow 0$. In other words, the drift $b(x)$ measures locally the mean velocity of the random movement modeled by X, and $a(x)$ approximates the rate of change in the covariance matrix of the vector $X_t - x$, for small values of $t > 0$.

Needless to say, the assumptions of the above proposition are rather restrictive. It is possible to apply it to Brownian motion (which is the standard example for almost every proposition in the course) but we were not able to give a different example, with a stochastic process of some interest in the applications.

We next state the important Dynkin-Kinney theorem. This result shows that the above condition (6.19) implies continuity of the trajectories for the associated Markov process.

Theorem 6.25. *Assume that $p(t, x, \Gamma)$ is a transition probability function such that condition (6.19) holds. Then there exists a Markov process associated to p with continuous trajectories a.s.*

For reason of space we will not prove the result, but we give a reference to the book of `Ventsel` [Ve83, Theorem 9.1.1].

6.5 Diffusions

The term *diffusion* is associated with the class of continuous Markov processes which can be characterized in terms of their infinitesimal generator.

Definition 6.26. *A* diffusion process *associated to a differential operator A_0*

$$\sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} f(x) + \sum_{k=1}^d b_k(x) \frac{\partial}{\partial x_k} f(x)$$

is a stochastic process $X = \{X_t,\ t \ge 0\}$ such that

1. *is Feller,*
2. *trajectories are continuous, and*
3. *$C_c^2(E) \subset D(A)$ and $Af = A_0 f$, $f \in C_c^2(E)$.*

Recall that we have already proved that a Feller process with (right-) continuous paths has the strong Markov property. In literature it is sometimes required this property at point 1.; also, there are different requirements also the other points: for instance, `Ikeda and Watanabe` states 3. in the martingale form

$$f(X^x(t)) - f(x) - \int_0^t (Af)(X_s^x) \, \mathrm{d}s$$

is a $\mathbb{P}$-martingale for every $f \in D(A)$ and $x \in E$.

7

Harmonic functions and Dirichlet problem

In this lecture we shall go into some model problems as examples of the important links connecting probability theory and functional analysis.

Solutions to several classes of problems in partial differential equations, both in the elliptic as in the parabolic case, can be represented as expected values of specific stochastic processes (which are, usually, solutions to suitable stochastic equations). Such representations allow to study properties of solutions by means of stochastic analysis; however, also the converse holds. As we have seen in Section 6.2, diffusion processes are defined in terms of their infinitesimal generator.

In classical potential theory, a number of interesting problems had their origin in electromagnetism. In this lecture we concentrate on harmonic functions. These are naturally related to Brownian motion. In particular, we will study the Dirichlet problem: find a function u harmonic in a domain D, continuous up to the boundary, with given boundary values $f \in C(\partial D)$.

7.1 Preliminaries

For this lecture we take $E = \mathbb{R}^d$, $d \geq 1$. A *domain* $D \subset E$ is a nonempty, connected open set; ∂D will denote its boundary and $\bar{D}$ its closure. To begin with, we generalize some results from Section 5.5.

Throughout this lecture we consider the *canonical Brownian motion* B_t with values in $\mathbb{R}^d$, defined on the canonical stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \{\mathbb{P}^x,\ x \in \mathbb{R}^d\}$, where $\Omega = C([0, +\infty[; \mathbb{R}^d)$, compare also with Section 3.2.

To define this canonical Brownian motion, we begin with a family of Brownian motions $\{\bar{B}^x_t, t \geq 0\}$, starting from $x \in \mathbb{R}^d$, with values in $\mathbb{R}^d$, defined on some stochastic basis $(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t,\ t \geq 0\}, \bar{\mathbb{P}}\}$.

Then we note that $\Omega = C([0, +\infty[; \mathbb{R}^d)$ is a complete separable metric space endowed with the metric d,

$$d(\omega_1, \omega_2) = \sum_{n \geq 1} \frac{1}{2^n} \frac{\sup_{x \in [0,n]} |\omega_1(x) - \omega_2(x)|}{1 + \sup_{x \in [0,n]} |\omega_1(x) - \omega_2(x)|}, \quad \omega_1,\, \omega_2 \in \Omega.$$

$\mathcal{F}$ denotes the Borel σ-algebra of Ω. Moreover, we introduce the canonical process

$$B_t(\omega) = \omega(t), \quad t \geq 0, \quad \omega \in \Omega. \tag{7.1}$$

We complete the σ-algebra $\sigma(B_s,\ s \in [0,t])$ and we obtain the σ-algebra $\mathcal{F}_t$, $t \geq 0$. The filtration $\{\mathcal{F}_t\}$ satisfies the standard assumptions.

It is not difficult to show that, for each $x \in \mathbb{R}^d$, $\bar{B}^x = \{\bar{B}^x_t,\ t \geq 0\}$ is a measurable mapping from $(\bar{\Omega}, \bar{\mathcal{F}})$ into $(\Omega, \mathcal{F})$, i.e., for each $\omega \in \bar{\Omega}$, we have: $B^x(\omega) = f(t)$, where $f(t) = \bar{B}^x_t(\omega)$, $t \geq 0$.

Define the probability measures $\mathbb{P}^x$, $x \in \mathbb{R}^d$, as the image measures of $\bar{\mathbb{P}}$ under the application B^x, i.e.,

$$\mathbb{P}^x(A) = \mathbb{P}(\omega \in \bar{\Omega} \ :\ B^x(\omega) \in A), \quad A \in \mathcal{F}.$$

We also set $\mathbb{P}^0 = \mathbb{P}$. The measure $\mathbb{P}$ is called the *Wiener measure*. We have $\mathbb{P}^x(\omega \in \Omega \ :\ \omega(0) = x) = 1$, i.e. $\mathbb{P}^x$ is concentrated on the paths leaving from x. Note that on $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P}^x)$ the canonical process given in (7.1) is a Brownian motion starting from x. Finally the expectation with respect to $\mathbb{P}^x$ will be indicated by $\mathbb{E}^x$.

Next we will generalize some results from Section 5.5. For a domain $D \subset E$, we introduce the first exit time

$$\tau_D(\omega) = \inf\{t \geq 0 \ :\ B_t(\omega) = \omega(t) \in D^c\}.$$

If $\{\ \} = \emptyset$, then we set $\tau_D(\omega) = +\infty$.

Problem 7.1.
(i) Show that, for any $A \in \mathcal{F}$, the mapping: $x \mapsto \mathbb{P}^x(A)$ is Borel measurable.
(ii) Let $\{B_t, t \geq 0\}$ a canonical d-dimensional Brownian motion. Let $\Gamma \subset \mathbb{R}^d$ be a closed set and τ be the first hitting time of Γ:

$$\tau(\omega) = \tau_{\Gamma^c}(\omega) = \begin{cases} \inf\{t > 0 \ :\ B_t(\omega) \in \Gamma\} \\ +\infty, \ \text{if } \{\ \cdot\ \} = \varnothing. \end{cases}$$

Show that τ is a stopping time with respect to the filtration $\{\mathcal{F}_t\}$.

The next result shows in particular that, for any bounded domain D, τ_D is finite a.s.

Proposition 7.1. *For a bounded domain $D \subset \mathbb{R}^d$, it holds*

$$\sup_{x \in D} \mathbb{E}^x[\tau_D] < \infty. \tag{7.2}$$

Proof. Let $x \in D$. We first reduce the problem to a simpler one, where we can appeal to the stron markov property of Brownian motion. For every $T \in \mathbb{N}$, it holds

$$\{\tau_D > T\} \subset \bigcap_{n=1}^{T} \{B_n \in D\}.$$

If we compute $\mathbb{P}^z(B_1 \in D)$, $z \in D$, we obtain the upper bound

$$\sup_{z \in D} \mathbb{P}^z(B_1 \in D) = \sup_{z \in D} \int_D \frac{1}{\sqrt{2\pi}} \exp(-|z-y|^2/2)\, \mathrm{d}y := \theta < 1$$

altough it is possible to actually compute θ, it is sufficient to use the bound $\theta < 1$ that is a consequence of the boundedness of D. Next, if we consider the case $T = 2$

$$\begin{aligned}\mathbb{P}^x(B_1 \in D, B_2 \in D) &= \mathbb{E}^x[\mathbb{E}^x(\mathbb{1}_D(B_1)\,\mathbb{1}_D(B_2 - B_1 + B_1)\,|\,\mathcal{F}_1)]\\ &= \mathbb{E}^x[\mathbb{1}_D(B_1)\mathbb{E}^x[\mathbb{1}_D(B_2 - B_1 + y)]_{y=B_1}] = \mathbb{E}^x[\mathbb{1}_D(B_1)\mathbb{E}^{B_1}[\mathbb{1}_D(B_1')]] \le \theta^2\end{aligned}$$

where we use the homogeneity of the Brownian motion: $B_t' = B_t - B_1$ is independent from $\mathcal{F}_1$ and B' is again a standard Brownian motion, and the uniform bound proved above. We obtain by induction that

$$\mathbb{P}^x\left(\bigcap_{n=1}^{T}\{B_n \in D\}\right) \le \theta^T.$$

We use then a little trick for the mean value of random variables: if X is nonnegative, then $\mathbb{E}[X] \le \sum_{j=0}^{\infty} \mathbb{P}(X > j)$[1]; it follows

$$\mathbb{E}^x[\tau_D] \le \sum_{T=0}^{\infty} \mathbb{P}^x\{\tau_D > T\} \le \sum_{T=0}^{\infty} \theta^T = \frac{1}{1-\theta} < \infty.$$

Since θ is independent from x, also the above estimate remains true taking the supremum as x varies in D.

□

Since in the above definition of τ_D we take the infimum on all times $t > 0$, it is not clear what happens when $x \in \partial D$. Actually it could well be $\tau_D > 0$ if the trajectory enters immediately D and leaves only after some times.

Definition 7.2. *We say that a point $a \in \partial D$ is* regular *if $\mathbb{P}^a(\tau_D = 0) = 1$.*

Then, we call *irregular* every point $a \in \partial D$ that is not regular, i.e., if $\mathbb{P}^a(\tau_D = 0) < 1$; for the sake of completeness, we remark that by Blumenthal's zero-or-one law[2]if a point a is irregular then necessarily it holds $\mathbb{P}^a(\tau_D = 0) = 0$.

The following result concerns regularity of points in the one dimensional case.

[1] exercise: convince yourself of this estimate, by taking a random variable with density $f(x)$, $x \ge 0$, and actually writing the expectation of X:

$$\mathbb{E}[X] = \int_0^\infty x f(x)\,\mathrm{d}x = \int_0^\infty \int_y^\infty f(x)\,\mathrm{d}x\,\mathrm{d}y \le \sum_{n=0}^{\infty} \int_n^{n+1} \mathbb{P}(X > y)\,\mathrm{d}y$$

[2] Suppose that B is a Brownian motion defined on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \ge 0\}, \mathbb{P})$, satisfying the standard assumptions. The σ-field $\mathcal{F}_0$ coincides with $\mathcal{F}_{0+} = \bigcap_{t>0} \mathcal{F}_t$ contains the events which happen instantaneously, i.e., those events whose occouring can be stated ad arbitrary time $t > 0$.

Then this σ-field is trivial, i.e., for any $A \in \mathcal{F}_0$ either one has $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. This result is known as *Blumenthal 0-1 law*; for a proof, we refer to `Chung` [Ch82]. In fact, more generally, we have:

Theorem 7.3. *Assume that $\{B_t\}$ is a Brownian motion. Setting $\mathcal{F}_t^s = \sigma(B_u,\ s \le u < t)$ and $\mathcal{F}_{s+}^s = \bigwedge_{t>s} \mathcal{F}_t^s$, then $\mathcal{F}_{s+}^s$ is trivial.*

Proposition 7.4. *With probability 1, a real Brownian motion* $\{B_t,\ t \ge 0\}$ *(starting from* 0*) changes sign infinitely often on every interval* $[0, \varepsilon]$*,* $\varepsilon > 0$.

Proof. Let τ be the first hitting time for $(0, \infty)$ and, similarly, τ' be the first hitting time for $(-\infty, 0)$. Both τ and τ' are stopping times (we use the assumption that the filtration $\{\mathcal{F}_t\}$ satisfies the standard assumptions); further, the symmetry of Brownian motion implies that $\mathbb{P}^0(\tau = 0) = \mathbb{P}^0(\tau' = 0)$ and this value shall be equal to 0 or 1 by Blumenthal's 0-1 law (Theorem 7.3), since the events $\{\tau = 0\}$ and $\{\tau' = 0\}$ both belong to $\mathcal{F}_0$. If it were 0, it should be $\mathbb{P}^0(B_t = 0, \forall t \in [0, \varepsilon]) = 1$, hence almost surely trajectories of B_t have a derivative in $t = 0$, which is not. But then it must hold $\mathbb{P}^0(\tau = 0) = \mathbb{P}^0(\tau' = 0) = 1$, hence for every ω there exist two sequences of times $s_n, t_n \to 0$ such that $B_{s_n} < 0$, $B_{t_n} > 0$ for every $n > 0$.
□

Let $D = (a, b)$ be an interval of the real line; then almost surely trajectories starting from a (and b) enters instantaneously in D^c. We have thus proved the following

Corollary 7.5. *In the one dimensional case, every point* $a \in \mathbb{R}$ *is regular.*

For every $r > 0$ we denote D_r the open ball of $\mathbb{R}^d$, $d \ge 2$, centered in the origin with radius r and $D_r(x) = x + D_r$. On the boundary $\partial D_r = \{x \in \mathbb{R}^d \,:\, |x| = r\}$ we consider the *surface measure* $\lambda_r := \lambda_{\partial D_r}$; the surface area of ∂D_r is $S_r = s_d r^{d-1}$ (for a constant $s_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}$ depending only on the dimension d) and the volume of D_r is $V(D_r) = c_d r^d$.

We recall below an identity concerning the Lebesgue integral of a function f on a ball $D_r(x)$ and the integral of f on the surfaces $x + \partial D_r$:

$$\int_{D_r(x)} f(y)\,\mathrm{d}y = \int_0^r \int_{x+\partial D_\rho} f(y)\,\lambda_\rho(\mathrm{d}y)\,\mathrm{d}\rho. \tag{7.3}$$

The exit time from the ball D_r of radius r will be denoted by $\tau_r = \tau_{D_r}$.

Proposition 7.6. *For a canonical Brownian motion* B *starting from the origin, the random variables* τ_r *and* $B(\tau_r)$ *are independent. Furthermore,* $B(\tau_r)$ *is uniformly distributed on* ∂D_r.

Proof. Let θ be a rotation in $\mathbb{R}^d$ and Θ be the corresponding mapping on Ω, given by $(\Theta\omega)(t) = \theta(\omega(t))$. Since θ preserves distances, it is clear that

$$\tau_r(\omega) = \tau_r(\Theta\omega).$$

Now we claim that continuity of paths implies

$$B(\tau_r) \in \partial D_r \text{ on } \{\tau_r < \infty\}.$$

Let $A \subset \partial D_r$ and define the event

$$\Lambda = \{\omega \,:\, \tau_r(\omega) \le t,\ B(\tau_r(\omega))(\omega) = \omega(\tau_r(\omega)) \in A\} \tag{7.4}$$

and the "rotation" of Λ

$$\Theta^{-1}\Lambda = \{\omega \,:\, \tau_r(\Theta\omega) \le t,\ (\Theta\omega)(\tau_r(\Theta\omega)) = \theta(\omega(\tau_r(\Theta\omega)))\ \in A\}$$

which is equal to

$$\Theta^{-1}\Lambda = \{\omega \,:\, \tau_r(\omega) \le t, \quad \omega(\tau_r(\omega)) \in \theta^{-1}A\}. \tag{7.5}$$

We claim the identity

$$\mathbb{P}(\Lambda) = \mathbb{P}(\Theta^{-1}\Lambda)$$

which is a consequence of the following fact

$$\text{the process } B_t(\Theta\omega) \text{ is again a Brownian motion starting from } 0 \tag{7.6}$$

(this follows by the symmetry of the Gaussian measure in $\mathbb{R}^d$). Using formulae (7.4) and (7.5) we obtain

$$\mathbb{P}(\tau_r \le t,\ B(\tau_r) \in A) = \mathbb{P}(\tau_r \le t,\ B(\tau_r) \in \theta^{-1}A) \tag{7.7}$$

For fixed t, the left hand side defines a measure $\mu(A)$ on the sphere ∂D_r that is rotation invariant, thanks to the arbitrariety of θ on the right hand side. Thus $\mu(A)$ must coincide, up to a scalar factor, with the surface measure on ∂D_r

$$\frac{1}{s_d r^{d-1}} \lambda_r(A)$$

and the identity $\mu(\partial D_r) = \mathbb{P}(\tau_r \le t)$ implies that this number is the required scalar factor, hence

$$\mathbb{P}(\tau_r \le t,\ B(\tau_r) \in A) = \frac{1}{s_d r^{d-1}} \lambda_r(A) \mathbb{P}(\tau_r \le t).$$

We obtain from this formula, at once, that τ_r and $B(\tau_r)$ are independent and that

$$\mathbb{P}(B(\tau_r) \in A) = \frac{1}{s_d r^{d-1}} \lambda_r(A)$$

which concludes the proof.

□

The above construction implies that the measure

$$\mu_r(\mathrm{d}x) = \mathbb{P}(B(\tau) \in \mathrm{d}x) = \frac{1}{s_d r^{d-1}} \lambda_r(\mathrm{d}x)$$

is a probability measure on ∂D_r. In terms of this measure, we obtain the following *average-on-sphere* property for the Brownian motion

$$\int_{\partial D_r} f(y)\, \mu_r(\mathrm{d}y) = \int_{\partial D_r} f(y)\, \mathbb{P}(B(\tau) \in \mathrm{d}y) = \mathbb{E}[f(B(\tau_{D_r}))]. \tag{7.8}$$

7.2 Harmonic functions

We record the classical definition of harmonic functions in a domain D.

Definition 7.7. *A function u on D is* harmonic *if it belongs to $C^2(D)$ (the class of twice continuously differentiable functions on D) and*

$$\Delta u = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} u(x) = 0 \quad \text{in } D.$$

As we shall see in the following, an harmonic function necessarily belongs to the class $C^\infty(D)$ and it possesses the averaging property. It is this property that allows Brownian motion to enter into play.

Definition 7.8. *We say that a function $u : D \to \mathbb{R}$ possesses the* sphere-averaging property *if for every $x \in D$ and $r > 0$ such that $x + \bar{D}_r \subset D$ it holds*

$$u(x) = \frac{1}{s_d r^{d-1}} \int_{\partial D_r} u(x+y)\, \lambda_r(\mathrm{d}y).$$

The equivalence between harmonic property of Definition 7.7 and sphere-averaging property of Definition 7.8 is well known in potential theory. In the following result we give a short (probabilistic) proof of one implication.

Proposition 7.9. *Assume that $u : D \to \mathbb{R}$ is an harmonic function; then u verifies the averaging property.*

Proof. Let $x \in D$ and $r > 0$ be such that $x + \bar{D}_r \subset D$. Arguing as is Theorem 6.14 we have that for given $f \in C_c^2(\mathbb{R}^d)$ the process

$$h(t) = f(B_t) - f(x) - \frac{1}{2} \int_0^t \Delta f(B_s)\, \mathrm{d}s \tag{7.9}$$

is a martingale with respect to the filtration $\{\mathcal{F}_t\}$ under the measure $\mathbb{P}^x$, for each $x \in \mathbb{R}^d$.

We claim that for $u \in C^2(D)$ there exists an extension $\bar{f} \in C_c^2(\mathbb{R}^d)$ such that $\bar{f} \equiv u$ on $x + \bar{D}_r$. Using such $\bar{f}$ in (7.9), we get that $h(t)$ (defined above) is a martingale. Hence the stopped process $h(t \wedge \tau_r)$ is again a martingale, see Corollary 4.23, and it verifies

$$h(t \wedge \tau_r) = u(B(t \wedge \tau_r)) - u(x)$$

since the function $\bar{f} \equiv u$ is harmonic in $x + \bar{D}_r$. Further, since $B_0 = x$, $\mathbb{P}^x$-a.s., $\mathbb{E}^x[h(t)] = \mathbb{E}^x[h(0)] = 0$ hence

$$\mathbb{E}^x[u(B(t \wedge \tau_r))] = \mathbb{E}^x[u(B_0)] = u(x);$$

passing to the limit $t \to \infty$, since τ_r is finite a.s., formula (7.8) yields the thesis.

□

We do not know a probabilistic verification of the opposite implication. A proof, using classical tools from potential analysis, can be found for instance in Dynkin and Yushkevich [DY69] or Durrett [Du96, page 148].

We show next that an harmonic function is smooth. Let $u : D \to \mathbb{R}$ be harmonic; the above proposition shows that for every $r > 0$ it holds

$$u(x) = \mathbb{E}^x[u(B(\tau_r))] = \frac{1}{s_d r^{d-1}} \int_{\partial D_r} u(x+y)\, \lambda_r(\mathrm{d}y)$$

Let ρ be a test function in $C^\infty(E)$ such that $\rho(x) = \rho(|x|)$, $\rho \equiv 0$ for $|x| > \delta$ and $\int \rho(x)\,\mathrm{d}x = 1$. We have then

$$\int_0^\infty \rho(r) s_d r^{d-1}\,\mathrm{d}r = 1.$$

Now, if x and δ are such that $x + \bar{D}_\delta \subset D$, it follows from the averaging property that

$$u(x) = \int_0^\infty \left[\frac{1}{s_d r^{d-1}} \int_{\partial D_r} u(x+y)\,\lambda_r(\mathrm{d}y)\right] \rho(r) s_d r^{d-1}\,\mathrm{d}r$$

hence

$$u(x) = \int_0^\infty \int_{\partial D_r} u(x+y)\,\lambda_r(\mathrm{d}y)\rho(r)\,\mathrm{d}r = \int_0^\infty \int_{\partial(x+D_r)} u(y)\rho(|y-x|)\,\lambda_r(\mathrm{d}y)\,\mathrm{d}r$$

which is, thanks to (7.3) and the assumed properties of ρ, equal to

$$u(x) = \int_E u(y)\rho(|y-x|)\,\mathrm{d}y. \tag{7.10}$$

Proposition 7.10. *Given an harmonic function $u : D \to \mathbb{R}$, then u is of class $C^\infty(D)$.*

Proof. It is sufficient to differentiate u in (7.10); on the right hand side, the function $\rho(|x-y|)$ has continuous and bounded derivatives of all orders and so the same shall hold for the left hand side. □

A lot of examples of harmonic functions are easily constructed. In one dimension, for instance, harmonic functions are just the linear functions.

Problem 7.2.

1. Show that u verifies the averaging property if and only if for any $x \in \mathbb{R}^d$, $r > 0$ it holds

$$u(x) = \frac{1}{V(D_r)} \int_{D_r} u(x+y)\,\mathrm{d}y.$$

2. Show that if u is harmonic, then the same holds for all its partial derivatives of all orders.
3. Show that in dimension $d \geq 3$, every radial harmonic function $u(x) = u(|x|)$ has the form $c_1|x|^{2-d} + c_2$ for constants $c_1, c_2 \in \mathbb{R}$. These functions are called *fundamental radial solutions* of the Laplace equation $\Delta u = 0$. Find the same for $d = 2$.
   Hint: write (and solve) the Laplace equation in radial coordinates.

7.3 Dirichlet problem

The classical form of *Dirichlet problem* (D, f) is the following: given a bounded domain $D \subset \mathbb{R}^d$ and a bounded continuous function $f : \partial D \to \mathbb{R}$, find a function $u : \bar{D} \to \mathbb{R}$ which is continuous on $\bar{D}$, harmonic on D and equal to f on the boundary ∂D, i.e.,

$$\begin{cases} \Delta u = 0 & \text{in } D, \\ u = f & \text{in } \partial D. \end{cases} \tag{7.11}$$

Several physical problem can be stated with the above equation; for instance, the stationary temperature of a region whose boundary is maintained at the given temperature f.

Previous sections show that a natural candidate for the solution is given by

$$u(x) = \mathbb{E}^x[f(B_{\tau_D})], \qquad x \in \bar{D}. \tag{7.12}$$

Let D_r be a small ball such that $x + \bar{D}_r \in D$ and $\tau_r := \tau_{x+D_r}$ be the first exit time from this neighbourhood of x. Then we have, by the strong Markov property

$$B_t = B_t - B_{\tau_r} + B_{\tau_r} = B'_{t'} + B_{\tau_r}$$

The process B' is again a Brownian motion, independent from $\mathcal{F}_{\tau_r}$, and *its* paths will exit from D at *its* stopping time τ'_D. We have, by temporal homogeneity

$$\mathbb{E}^y[f(B_{\tau_D})] = \mathbb{E}^y[f(B'_{\tau'_D})] = u(y)$$

where the first y stands for $B_0 = y$, the second one stands for $B'_0 = B_{\tau_r} = y$.

Therefore, using the strong Markov property, we have

$$u(x) = \mathbb{E}^x[f(B(\tau_D))] = \int \mathbb{P}^x(B(\tau_r) \in \mathrm{d}y)\, \mathbb{E}^y[f(B'(\tau'_D))] = \int \mathbb{P}^x(B(\tau_r) \in \mathrm{d}y)\, u(y)$$

hence u verifies the sphere-averaging property, compare (7.8), and as a consequence we have that it is harmonic in D; it remains to study the behaviour of $u(x)$ as x approaches the boundary ∂D.

In order to prove continuity up to the boundary, we shall characterize those boundary points $a \in \partial D$ such that

$$\lim_{x \in D,\ x \to a} \mathbb{E}^x[f(B_{\tau_D})] = f(a) \tag{7.13}$$

actually holds for every measurable, bounded function $f : \partial D \to \mathbb{R}$ continuous in a. This characterization will be given in terms of regular points, see Definition 7.2. In general, a domain D will be called *regular* if all its boundary points are regular; it will turn out that this definition is consistent with the classical definition of regular domain as "a domain in which the Dirichlet problem (D, f) is solvable for every $f \in C_b(\partial D)$".

The representation (7.12) yields naturally a uniqueness result for the Dirichlet problem (D, f).

Proposition 7.11. *Assume that*

$$\mathbb{P}^a(\tau_D < \infty) = 1, \qquad \forall\, a \in D. \tag{7.14}$$

Given a continuous bounded function f on ∂D, every bounded solution of problem (D, f) has the representation (7.12).

Proof. Take any bounded solution $u(x)$ of problem (D, f); the idea, as in the proof of Proposition 7.9, is to apply again Theorem 6.14, since u is harmonic. Let D_n a sequence of domains invading D

$$D_n = \{y \in D : \inf_{x \in \partial D} |x - y| > 1/n\}.$$

Then $u(B(t \wedge \tau_{D_n})) - u(x) = h(t \wedge \tau_{D_n})$ is a martingale and taking expectation we obtain $\mathbb{E}^x[u(B(t \wedge \tau_{D_n}))] = u(x)$ for every $x \in D_n$.

We pass to the limit as $t \to \infty$ and $n \to \infty$, so that $B(t \wedge \tau_{D_n})$ converges to $B(\tau_D)$: since on the boundary ∂D the function u verifies $u(B(\tau_D)) = f(B(\tau_D))$ we see that the representation (7.12)

$$u(x) = \mathbb{E}^x[f(B_{\tau_D})]$$

holds.

□

In order to prepare the main existence result for the Dirichlet problem (D, f), we state the following property of regular points.

Problem 7.3. If a is a regular point for D, then for every $\varepsilon > 0$ it holds

$$\lim_{x \in D, x \to a} \mathbb{P}^x(\tau_D > \varepsilon) = 0. \tag{7.15}$$

Hint: If we show that the function $g(x) = \mathbb{P}^x(\tau_D > \varepsilon)$ is upper semicontinuous then we have done (why?). However, the function $g_\delta(x) = \mathbb{P}^x(B_s \in D \,\forall\, \delta < s \le \varepsilon)$ is continuous (use the strong Markov property) and $g_\delta(x) \downarrow g(x)$...

Theorem 7.12. *If a is a regular point and $f : \partial D \to \mathbb{R}$ is continuous in a, then the function $u(x)$ defined in (7.12) satisfies*

$$\lim_{x \in D, x \to a} u(x) = f(a). \tag{7.16}$$

In particular, if $f : \partial D \to \mathbb{R}$ is a continuous and bounded function, there exists a unique solution to Dirichlet problem (D, f), which is expressed by the representation (7.12).

Proof. The starting point is a consequence of the reflection principle for the Brownian motion. For every $r > 0$, the probability

$$\mathbb{P}^x(\max_{0 \le t \le \varepsilon} |B_t - B_0| < r) = \mathbb{P}^0(\max_{0 \le t \le \varepsilon} |B_t| < r)$$

is independent from x and converges to 1 as $\varepsilon \to 0$. Therefore

$$\begin{aligned} \mathbb{P}^x(|B(\tau_D) - B(0)| < r) &\ge \mathbb{P}^x\left(\big(\max_{0 \le t \le \varepsilon} |B_t - B_0| < r\big) \cap (\tau_D \le \varepsilon) \right) \\ &\ge \mathbb{P}^0(\max_{0 \le t \le \varepsilon} |B_t| < r) - \mathbb{P}^x(\tau_D > \varepsilon). \end{aligned}$$

Letting $x \to a$ from the interior of D, and then letting $\varepsilon \to 0$, we obtain from (7.15)

$$\lim_{x \in D, x \to a} \mathbb{P}^x(|B(\tau_D) - B(0)| < r) = 1, \qquad \forall\, r > 0. \tag{7.17}$$

From the assumed continuity of f in a we obtain the thesis.

□

Problem 7.4. Starting from formula (7.17), finish the proof of the theorem.

In the 1-dimensional case Dirichlet problem is always solvable, and the solution is piece-wise linear. This may be proved directly, without mentioning regularity of boundary points. One may wonder if the assumption: a is regular, is strictly necessary. The following result answers affermatively to the question in dimension $d \geq 2$.

Corollary 7.13. *Assume that a is an irregular point; then there exists a function $f : \partial D \to \mathbb{R}$, bounded and continuous, such that, letting u be the harmonic function defined in (7.12), it holds*

$$\min_{x \in D, x \to a} \lim u(x) < f(a).$$

Proof. Let f be a function on ∂D such that $f(a) = 1$ and $f(x) < 1$ for every $x \neq a$: we can take, for instance

$$f(x) = (1 - |a - x|) \vee 0.$$

By assumption a is irregular, while the set $\{a\}$ is *polar* (see the Addendum to this lecture) for the Brownian motion in 2 dimensions; it follows

$$\mathbb{P}^a(B(\tau_D) = a) = 0.$$

Then, since $f < 1$ on $\partial D \setminus \{a\}$, we have

$$1 > \mathbb{E}^a[f(B(\tau_D))]$$

further, if τ_r is given by $\tau_r = \inf\{t > 0 : |B(t) - B(0)| \geq r\}$ we have, from $\lim_{r \to 0} \mathbb{P}^a(\tau_r < \tau_D) = 1$ and the strong Markov property applied at τ_r, the identity

$$\mathbb{E}^a[f(B(\tau_D))] = \lim_{r \to 0} \mathbb{E}^a[\mathbb{1}_{\{\tau_r < \tau_D\}} u(B(\tau_r))]$$

but the last quantity is bounded below, as follows

$$\lim_{r \to 0} \mathbb{E}^a[\mathbb{1}_{\{\tau_r < \tau_D\}} u(B(\tau_r))] \geq \lim_{r \to 0} \mathbb{P}^a(\tau_r < \tau_D) \inf_{x \in B(a,r) \cap D} u(x).$$

Then, collecting the above inequalities, we have the thesis.

□

Example 7.14. Let us consider the punctured unit ball $D = \{x : 0 < |x| < 1\}$. The origin is a *polar set* (see the Addendum to this lecture), hence it is irregular for D. A Brownian path starting from the origin leaves D, almost surely, from the external boundary.

The function u defined as in (7.12) is determined only from the values of f on the external boundary $\{|x| = 1\}$, hence it coincides with the harmonic function

$$\tilde{u}(x) = \mathbb{E}^x[f(B(\tau_{D_1})] = \mathbb{E}^x[f(B(\tau_D)], \qquad x \in D; \tag{7.18}$$

on the other hand, since U is a solution of Dirichlet problem (D, f), it must satisfy $u(0) = f(0)$, hence u is continuous in the origin if and only if $\tilde{u}(0) = f(0)$.

There are different way of stating conditions for regularity of boundary points $x \in \partial D$. We limit ourselves to consider the case of 2-dimensional domains. We have seen in previous example that an "isolated" point is irregular for the domain D; the following construction shows –roughly speaking – that this is a necessary and sufficient condition.

Proposition 7.15. *Let $D \subset \mathbb{R}^2$, and assume that for every point $a \in \partial D$ there exist a different point $b \neq a$ in ∂D and a simple arc contained in D^c joining a and b. Then the domain D is regular.*

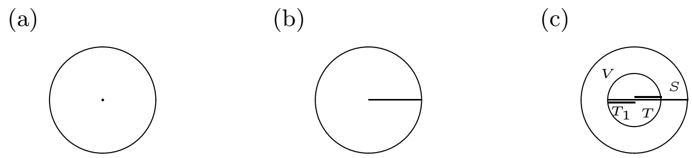

Fig. 7.1. (a): a punctured ball; (b): a ball without a radius; (c): our construction below

For the sake of semplicity, we give the proof in the case when D is the unit ball without a radius $S = \{0\} \times [0, 1)$. We claim that every point in D is regular. This is obvious in case $a \in \partial D$ is in the external boundary; then, the way we prove that the origin is regular will show clearly that all to the points of the segment S are such.

We consider a small ball V centered in the origin, with small radius η; set $T = V \cap S$ and $T_1 = V \cap (\{y = 0\} \setminus S)$ is the reflection of T with respect to the origin; their union is $T_0 = T \cup T_1$. We define the stopping times

$$\sigma_1 = \inf\{t > 0 \,:\, B_t \in T\}, \quad \tau_1 = \inf\{t > 0 \,:\, B_t \in T_1\}, \quad \tau_0 = \inf\{t > 0 \,:\, B_t \in T_0\}.$$

Now the proof works as follows. We know – see for instance Proposition 7.4 – that the second component $B_2(t)$ of the Brownian motion $B(t)$ is zero infinitely often in the time interval $[0, \varepsilon]$, for every $\varepsilon > 0$. Therefore, we can choose a sequence $t_n \to 0$ such that $B_2(t_n) = 0$. It remains to prove that, possibly passing to a subsequence, $B_1(t_n) \in T$ with positive probability, and we have done.

The following points will leave to the thesis:

1. $\{\sigma_1 = 0\} \subset \{\tau_D = 0\}$;
2. $\{\tau_0 = 0\} = \{\tau_1 = 0\} \cup \{\sigma_1 = 0\}$;
3. $\mathbb{P}^0(\sigma_1 = 0) = \mathbb{P}^0(\tau_1 = 0)$;
4. $\mathbb{P}^0(\tau_D = 0) \geq \frac{1}{2}\mathbb{P}^0(\tau_0 = 0)$;
5. $\mathbb{P}^0(\tau_0 = 0) \geq \mathbb{P}^0(\sup_{0 \leq t \leq \varepsilon} |B_1(t)| < \eta) > 0$, for any positive ε.

Problem 7.5. Complete the above proof, by giving the details of the construction in points (1)–(5).

7.3.1 Existence of a density

In this section we consider the Dirichlet problem in the half-plane $D = \{(x, y) : x > 0\}$; this domain is easily seen to be regular, so existence and uniqueness of the solution follows from Theorem 7.12. We can go further and prove the existence of an explicit analytic formula for the solution of the Dirichlet problem.

Let $\{B(t) = (B^1(t), B^2(t)),\ t \geq 0\}$ be a two-dimensional canonical Brownian motion; let $\tau = \tau_D$ be the first exit time for $B(t)$ from the domain D, that is

$$\tau = \inf\{t > 0 : B(t) \notin D\} = \inf\{t > 0 : B^1(t) \leq 0\}$$

and we see that τ only depends on the first component of $B(t)$.

Notice that the exit points from D have the form $(0, \eta)$. Hence formula (7.12) can be written as

$$u(x, y) = \mathbb{E}^{(x,y)}[f(B(\tau))] = \mathbb{E}^{(x,y)}[f(B^2(\tau))].$$

Note that τ is $\mathcal{F}_\tau$-measurable and the process $\{B_2(t),\ t \geq 0\}$ is independent from $\mathcal{F}_\tau$ (since τ depends only on the first component of $\{B(t),\ t \geq 0\}$ and moreover the processes $\{B^1(t),\ t \geq 0\}$ and $\{B^2(t),\ t \geq 0\}$ are independent). Using also a variant of Lemma 5.2, see for instance [EK86], we get

$$\begin{aligned}
u(x, y) = {} & \mathbb{E}^{(x,y)}[f(B^2(\tau(\omega), \omega))] = \mathbb{E}^{(x,y)}[\mathbb{E}^{(x,y)}[f(B^2(\tau(\omega), \omega)) \mid \mathcal{F}_\tau]] \\
= {} & \mathbb{E}^{(x,y)}[\mathbb{E}^{(x,y)}[f(B^2(s, \omega))]_{s=\tau}] = \mathbb{E}^{(x,y)}\Big[\int_{\mathbb{R}} f(t)\, \frac{1}{\sqrt{2\pi\tau}} \exp(-\frac{1}{2\tau}(t - y)^2)\, \mathrm{d}t\Big] \\
= {} & \mathbb{E}^x\Big[\int_{\mathbb{R}} f(t)\, \frac{1}{\sqrt{2\pi\tau_1}} \exp(-\frac{1}{2\tau_1}(t - y)^2)\, \mathrm{d}t\Big],
\end{aligned}$$

where $\mathbb{P}^x$ is a Wiener measure on $C([0, +\infty[, \mathbb{R})$, $x \geq 0$, and

$$\tau_1(\omega) = \inf\{t > 0 : \omega(t) = 0\}, \quad \omega \in C([0, +\infty[, \mathbb{R}).$$

To continue the proof we need to compute the distribution function $s \mapsto \mathbb{P}^x(\tau_1 \leq s)$; notice that the event $\{\tau_1 \leq s\}$ coincides with the event $\{\min_{t \leq s} B^1(t) \leq -x\}$, which is (by the symmetry of the Brownian motion) equivalent to $\{\max_{t \leq s} B^1(s) \geq x\}$; further, the reflection principle implies

$$\mathbb{P}^x(\tau_1 \leq s) = \mathbb{P}(\max_{t \leq s} B^1(t) \geq x) = 2\mathbb{P}(B^1(s) \geq x) = 2\int_x^\infty \frac{1}{\sqrt{2\pi s}} \mathrm{e}^{-y^2/2s}\, \mathrm{d}y$$

and the required density function is obtained by taking the derivative of the above expression

$$\mathbb{P}^x(\tau_1 \in \mathrm{d}s) = \frac{1}{\sqrt{2\pi s}} \frac{x}{s} \mathrm{e}^{-x^2/2s}\, \mathrm{d}s.$$

Hence, we get

$$u(x, y) = \int_{\mathbb{R}} f(t) \Big(\int_0^\infty \frac{x}{2\pi s^2} \exp\left(-\frac{1}{2s}((t - y)^2 + x^2) \right) \mathrm{d}s \Big) \mathrm{d}t$$

using the change of variable $u = -1/s$

$$= \int_{\mathbb{R}} f(t)\Big(\int_{-\infty}^0 \frac{x}{2\pi} \exp\left(-\frac{1}{2}((t-y)^2 + x^2)u \right) \,\mathrm{d}u \Big) \mathrm{d}t$$
$$= \int_{\mathbb{R}} \frac{1}{\pi} \frac{x}{(y-t)^2 + x^2} f(t) \,\mathrm{d}t.$$

7.4 Addendum. Recurrence and transience of the Brownian motion

The following result is concerned with recurrence of Brownian motion. Let $a \in \mathbb{R}^d$ and consider the passage time $\tau_{\{a\}}$ from the singleton $\{a\}$. In the 1-dimensional case, continuity of paths implies that $\mathbb{P}^x(\tau_{\{a\}} < \infty) = 1$. We say that the 1-dimensional Brownian motion is *recurrent*.

From classical results, it is known that a random walk – which is a kind of discrete Brownian motion, in a sense that can be made precise – is recurrent in both the 1-dimensional case and the 2-dimensional case, while it is transient in higher dimensions. Here "recurrence" means that the probability of a return in finite time is 1, hence almost every path returns infinitely often on every point, while "transience" means that there is a positive probability of never returns.

In the case of a Brownian motion, recurrence in the two dimensional case is a delicate matter. The results in literature show that the singletons are transient sets; however, if we enlarge them, considering small disks $D(a, \varepsilon)$ with positive radius $\varepsilon > 0$, then these sets are recurrent.

More easily, in dimensions 3 or higher, Brownian motion is transient. Actually, this follows from the 2-dimensional case, since otherwise we obtain an absurd: if the Brownian motion returns infinitely often to the origin in $\mathbb{R}^d$, $d \geq 3$, then its first 2 components should return infinitely often to 0 in dimension 2, which is not the case.

We begin with the observation that in dimension 2, the function $h(x) = \log|x|$ is harmonic in $\mathbb{R}^2 \setminus \{0\}$, hence the same holds for $c_1 + c_2 \log|x|$. For any domain circular domain D with internal radius r and external radius R, we can choose c_1 and c_2 such that the function f is zero on the outer boundary and equal to 1 on the inner boundary.

Since f is harmonic in D, it verifies $f(x) = \mathbb{E}^x[f(B_{\tau_D})]$, so that

$$f(x) = 1 \cdot \mathbb{P}(|B_{\tau_D}| = r) + 0 \cdot \mathbb{P}(|B_{\tau_D}| = R)$$

is the probability of escaping from D through the inner boundary. Some computation shows that

$$f(x) = \frac{\log(R) - \log(|x|)}{\log(R) - \log(r)}.$$

As an immediate consequence, we obtain the following.

Proposition 7.16. *A Brownian motion on the plane has zero probability of hitting a fixed point a (and never returning to the initial point x).*

We say that a set A is *polar* if for any $x \in \mathbb{R}^d$ it holds $\mathbb{P}^x(\tau_{A^c} < \infty) = 0$. The above proposition states in particular that in dimension 2 every singleton $\{a\}$ (and in particular the origin) is a polar set.

Proof. It is clear that we can start from some point x and examine the probability to ever reach 0. Further, for any $r > 0$, we have that the probability of reaching 0 is less or equal than the probability of entering in the ball centered at the origin with radius r. But, for arbitrary $0 < r < |x| < R$, the probability of entering in the inner circle before reaching the outer one tends to 0 as $r \to 0$ for fixed R.

If the particle starts from x, it will certainly reach the level $|x| + 1$ before reaching 0, then it will reach the level $|x| + 2$ before reaching 0, then it will reach the level $|x| + 3$ and so on. But the motion is continuous, then it takes an infinite time to reach each of the infinite levels $\{|x| + n,\ n \in \mathbb{N}\}$, and the particle shall pass through each of them before reaching 0: so it will never return to the origin.

□

Part II

Stochastic Integration

8

Probability theory in infinite dimensional Hilbert spaces

In this chapter H is a real separable Hilbert space, endowed with a scalar product $\langle\cdot,\cdot\rangle$ and norm $|\cdot|$, $\mathcal{B}(H)$ stands for the σ-algebra of Borel subsets of H, $L(H)$ is the space of linear continuous operators from H into H endowed with the operatorial norm $\|\cdot\|$.

After some preliminary results, we introduce the concept of Gaussian measure on $(H,\mathcal{B}(H))$. Then, we shall consider the *white noise* operator, which is a central element of our future analysis.

Throughout this lecture we also define basic probabilistic concepts in the Hilbert space setting. Many of these concepts were already introduced in the finite dimensional case treated in the initial lectures.

8.1 Preliminaries

Let $\mathcal{M}(H)$ denote the set of all probability measures on $(H,\mathcal{B}(H))$. Recall that μ is a Borel probability measure if μ is a Borel positive measure with $\mu(H)=1$. Our next results show that, also in the infinite dimensional case, any measure $\mu\in\mathcal{M}(H)$ is *regular* and it is possible to use compact sets in the approximation result.

Proposition 8.1. *Let $\mu\in\mathcal{M}(H)$. Then for any Borel set $B\in\mathcal{B}(H)$ we have*

$$\mu(B)=\inf\{\mu(A)\,:\,A\ \textit{open}\ \supset B\}=\sup\{\mu(C)\,:\,C\ \textit{closed}\ \subset B\}. \tag{8.1}$$

Actually, the proof shows that this result still holds, in more generality, assuming only that H is a metric space with distance d and endowed with the Borel σ-field $\mathcal{B}$.

Proof. Let us define the class of sets $\mathcal{K}=\{B\in\mathcal{B}(H)\,:\,(8.1)\text{ holds}\}$. The proof will consist of the following steps

1. $\mathcal{K}$ is a σ-field.
2. $\mathcal{K}$ contains the open sets of H.

Of course, $\mathcal{K}$ contains H itself; further, if $B\in\mathcal{K}$, then B^c belongs to $\mathcal{K}$. Let us now prove that $\mathcal{K}$ is closed under a countable union of sets. Let $\{B_n,\ n\in\mathbb{N}\}\subset\mathcal{K}$ and fix $\varepsilon>0$; we have to show that there exist a closed set C and an open set A such that

$$C \subset \bigcup_n B_n \subset A, \qquad \mu(A \setminus C) \le \varepsilon.$$

It holds, for each $n \in \mathbb{N}$, that there exist C_n closed, A_n open sets such that

$$C_n \subset B_n \subset A_n, \qquad \mu(A_n \setminus C_n) \le \varepsilon \frac{1}{2^{n+1}}.$$

Setting $A = \bigcup_n A_n$ and $S = \bigcup_n C_n$ we obtain that $S \subset \bigcup_n B_n \subset A$ and $\mu(A \setminus S) \le \varepsilon/2$. However, A is open but S is not necessarily closed. We then consider its approximations $S_k = \bigcup_n^k C_n$: S_k is closed as a finite union of closed sets and converges to S, hence $\mu(S_k) \to \mu(S)$ and there is $k = k_\varepsilon$ such that $C := S_k$ verifies $\mu(S \setminus C) \le \varepsilon/2$, so that

$$C \subset S \subset \bigcup_n B_n \subset A, \qquad \mu(A \setminus C) \le \varepsilon.$$

To prove that each open set A belongs to $\mathcal{K}$, we need to construct a sequence of closed sets C_n such that $C_n \subset A$ and $\mu(A \setminus C_n) \to 0$ as $n \to \infty$. It is easily seen that it is sufficient to take the following definition:

$$C_n = \{x \in H \,:\, d(x, A^c) \ge \frac{1}{n}\}$$

where $d(x, A^c)$ is the distance from x to the complementary set A^c of A.

□

Proposition 8.2. *Let $\mu \in \mathcal{M}(H)$. Then for arbitrary $\varepsilon > 0$, there exists a compact set K_ε such that $\mu(H \setminus K_\varepsilon) < \varepsilon$.*

Proof. In the finite dimensional case, the previous proposition implies that we can take a bounded closed set C such that $\mu(H \setminus C) \le \varepsilon$: since closed and bounded sets are compact, the proof is complete.

In the infinite dimensional case, we use that *totally bounded* sets are *relatively compact*; we recall that a set S is totally bounded if for any ε there exists a covering of S given by a finite number of balls of radius ε: see for instance `Dunford and Schwartz` [DS88].

Let us fix a basis $\mathcal{A} = \{a_j,\ j \in \mathbb{N}\}$ in H (more generally, it would be enough to take as $\mathcal{A}$ a countable dense set in H). For any $k \in \mathbb{N}$ there exists $m = m_k$ such that

$$F_k = \bigcup_{j=1}^{m_k} \bar{B}(a_j, \tfrac{1}{k})$$

verifies $\mu(F_k) \ge 1 - 2^{-k}\varepsilon$. Define

$$F = \bigcap_{k=1}^{\infty} F_k.$$

F is totally bounded by construction, hence it is relatively compact, and it is closed (as it is an intersection of closed sets), which means that F is compact. Further,

$$\mu(F^c) = \mu\left(\bigcup_{k=1}^{\infty} F_k^c\right) \le \sum_{k=1}^{\infty} 2^{-k}\varepsilon < \varepsilon.$$

□

We are going to introduce the (finite dimensional) projections in H. Let $\{e_k\}$ be a complete orthonormal basis in H; define $P_n : H \to P_n(H)$, for each $n \in \mathbb{N}$, setting

$$P_n x = \sum_{k=1}^{n} \langle x, e_k \rangle \, e_k, \qquad x \in H. \tag{8.2}$$

Notice that $P_n x \to x$ as $n \to \infty$, for each $x \in H$.

Given a probability measure $\mu \in \mathcal{M}(H)$, we denote by μ_n the image measure of μ on $P_n(H)$ induced by the transformation P_n, i.e., $\mu_n = P_n \circ \mu$,

$$\mu_n(B) = \mu(P_n^{-1}(B)), \qquad B \in \mathcal{B}(P_n(H)).$$

The following result still concerns uniqueness of a probability measure.

Proposition 8.3. *Let $\mu, \nu \in \mathcal{M}(H)$ be such that*

$$\mu_n = \nu_n \textit{ for each } n \in \mathbb{N}.$$

Then $\mu = \nu$.

Problem 8.1. Provide a complete proof of Proposition 8.3.

Trace class operators

Take a linear operator Q from H into H, symmetric and non-negative, i.e., Q verifies

$$\langle Qx, y \rangle = \langle x, Qy \rangle, \quad \langle Qx, x \rangle \ge 0, \quad \forall\ x, y \in H.$$

We shall denote $L^+(H)$ the class of such operators. A special case, of some interest for us is played by the *trace class* operators

$$L_1^+(H) = \{Q \in L^+(H) \,:\, \mathrm{Tr}(Q) < \infty\},$$

where the trace of Q is

$$\mathrm{Tr}(Q) = \sum_{k=1}^{\infty} \langle Q e_k, e_k \rangle$$

for a complete orthonormal system $\{e_k\}$ in H. If $\mathrm{Tr}(Q) < \infty$, then its value is independent from the given system $\{e_k\}$. In particular, trace class operators are compact, and the trace of Q is the sum of all eigenvalues repeated according to their multiplicity.

Random variables

Let H and E be separable Hilbert spaces, $\mu \in \mathcal{M}(H)$ be a probability measure on the space $(H, \mathcal{B}(H))$; set $\mathcal{E} = \mathcal{B}(E)$.

By a *random variable* with values in the space $(E, \mathcal{E})$ we mean a measurable mapping $X : H \to E$ such that

$$A \in \mathcal{E} \implies X^{-1}(A) \in \mathcal{B}(H).$$

The probability measure $\mathcal{L}(X) := X \circ \mu$ induced by X on E is denoted by $X \circ \mu$,

$$(X \circ \mu)(A) = \mu(X^{-1}(A)), \qquad A \in \mathcal{E}.$$

The measure $X \circ \mu$ is said to be the *law* of the random variable X.

We recall the change of variables formula. Let φ be a real bounded Borel mapping $\varphi : E \to \mathbb{R}$; then

$$\int_H \varphi(X(\omega))\, \mu(\mathrm{d}\omega) = \int_E \varphi(x)\, (X \circ \mu)(\mathrm{d}x).$$

Independence of random variables $X_1, \ldots, X_n$ holds if the law of the vector $(X_1, \ldots, X_n)$ coincides with the measure product of the laws $X_1, \ldots, X_n$:

$$\mathcal{L}(X_1, \ldots, X_n) = \mathcal{L}(X_1) \times \cdots \times \mathcal{L}(X_n).$$

In general, a family of random variables $\{X_\alpha,\ \alpha \in T\}$ indexed by some set T, is said to be independent if any finite subset of the family is independent.

Definition 8.4. *Let μ be a probability measure on $(H, \mathcal{B}(H))$; we define the* mean *of μ (if it exists) as the vector $a \in H$ such that*

$$\langle a, h\rangle = \int \langle x, h\rangle\, \mu(\mathrm{d}x), \qquad \forall h \in H. \tag{8.3}$$

If μ has mean a, then we introduce the covariance operator *of μ (if it exists). This is a linear operator Q on H such that*

$$\langle Qh, k\rangle = \int \langle x - a, h\rangle\, \langle x - a, k\rangle\, \mu(\mathrm{d}x), \qquad \forall h, k \in H. \tag{8.4}$$

The Fourier transform (or characteristic function) of μ is defined as follows

$$\hat{\mu}(x) = \int \exp(\mathrm{i}\langle x, y\rangle)\, \mu(\mathrm{d}y), \qquad \forall x \in H. \tag{8.5}$$

In general, mean and covariance operator of a probability measure do not necessarily exist. However, if μ is a probability measure on $(H, \mathcal{B}(H))$ with finite first momentum,

$$\int_H |x|\, \mu(\mathrm{d}x) < +\infty,$$

then there exists the mean of μ. Indeed, the linear functional $F : H \to \mathbb{R}$ defined as

$$F(h) = \int_H \langle x, h \rangle \, \mu(\mathrm{d}x), \qquad h \in H,$$

is countinuous, since

$$|F(h)| \le |h| \int_H |x| \mu(\mathrm{d}x).$$

By the Riesz representation theorem, there exists a unique element $a \in H$ such that

$$\langle a, h \rangle = \int_H \langle x, h \rangle \, \mu(\mathrm{d}x), \qquad h \in H.$$

Therefore, a is the *mean* of μ. Suppose further that the second momentum of μ is finite,

$$\int_H |x|^2 \, \mu(\mathrm{d}x) < +\infty.$$

By using Hölder's inequality it holds that

$$\left(\int_H |x| \, \mu(\mathrm{d}x) \right)^2 \le \int_H |x|^2 \, \mu(\mathrm{d}x) < +\infty$$

and so μ has mean a. Then we can consider the symmetric bilinear form $G : H \times H \to \mathbb{R}$ defined as

$$G(h, k) = \int_H \langle x - a, h \rangle \, \langle x - a, k \rangle \, \mu(\mathrm{d}x), \qquad h, k \in H.$$

G is continuous, since

$$|G(h, k)| \le |h||k| \int_H |x - a|^2 \, \mu(\mathrm{d}x).$$

Again, this implies the existence of a unique linear bounded operator $Q \in L(H)$ such that

$$\langle Qh, k \rangle = \int_H \langle x - a, h \rangle \, \langle x - a, k \rangle \, \mu(\mathrm{d}x), \qquad h, k \in H.$$

Q is the *covariance operator* of μ.

Proposition 8.5. *The covariance operator Q related to a probability measure $\mu \in \mathcal{M}(H)$ is a symmetric, positive and trace class operator.*

Proof. The symmetry and positivity of Q are clear. It remains to show that Q is a trace class operator. Fix a complete orthonormal system $\{e_k, \; k \in \mathbb{N}\}$ in H. Then

$$\langle Qe_k, e_k \rangle = \int_H |\langle x - a, e_k \rangle|^2 \, \mu(\mathrm{d}x), \qquad k \in \mathbb{N}.$$

From the monotone convergence theorem and the Parseval's identity it follows that

$$\mathrm{Tr}(Q) = \sum_{k=1}^{\infty} \int_H |\langle x - a, e_k \rangle|^2 \, \mu(\mathrm{d}x) = \int_H |x - a|^2 \, \mu(\mathrm{d}x) < +\infty.$$

□

Finally, we collect some properties of the Fourier transform of μ.

Proposition 8.6. *The Fourier transform $\hat{\mu}$ of a probability measure μ verifies*

1. $\hat{\mu}(0) = 1$;
2. $\hat{\mu}$ *is uniformly continuous on* H;
3. $\hat{\mu}$ *is a positive defined functional, i.e.,*

$$\sum_{j,k=1}^{n} \hat{\mu}(x_j - x_k)\alpha_j \bar{\alpha}_k \geq 0$$

for every $n \geq 1$, $x_1, \ldots, x_n \in H$, $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$.

Proof. Only the last point deserves some attention. Given $n \geq 1$, $x_1, \ldots, x_n \in H$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$, we have

$$\begin{aligned}
\sum_{j,k=1}^{n} \hat{\mu}(x_j - x_k)\alpha_j \bar{\alpha}_k &= \sum_{j,k=1}^{n} \int e^{\mathrm{i}\langle x_j - x_k, y\rangle}\, \mu(\mathrm{d}y)\, \alpha_j \bar{\alpha}_k \\
&= \sum_{j,k=1}^{n} \int \alpha_j e^{\mathrm{i}\langle x_j, y\rangle} \overline{\left(\alpha_k e^{\mathrm{i}\langle x_k, y\rangle}\right)}\, \mu(\mathrm{d}y) \\
&= \left\langle \sum_{j=1}^{n} \alpha_j e^{\mathrm{i}\langle x_j, \cdot\rangle}, \sum_{j=1}^{n} \alpha_j e^{\mathrm{i}\langle x_j, \cdot\rangle} \right\rangle_{L^2(H,\mu;\mathbb{C})} \\
&= \int \left| \sum_{j=1}^{n} \alpha_j e^{\mathrm{i}\langle x_j, y\rangle} \right|^2 \mu(\mathrm{d}y) \geq 0.
\end{aligned}$$

□

Remark 8.7. In the above proof we have encountered the space $L^2(H, \mu; \mathbb{C})$ of mappings $f : H \to \mathbb{C}$ such that

$$\int |f(x)|^2\, \mu(\mathrm{d}x) < \infty.$$

In probabilistic terms, f is a complex, square integrable random variable on the probability space $(H, \mathcal{B}(H), \mu)$.

Remark 8.8. In the finite dimensional case, the classical Bochner theorem shows that properties 1.–3. of Proposition 8.6 are also sufficient to characterize probability measures using characteristic functions, see Theorem 1.2.

The same does not hold in the infinite dimensional case, as an additional hypothesis is required: we propose in Exercise 8.2 below to construct a counterexample. For a complete characterization of Fourier transforms of probability measures on infinite dimensional Hilbert spaces we refer, for instance, to Theorem 2.13 in `Da Prato and Zabczyk` [DZ92].

We conclude with another uniqueness principle for probability measures.

Proposition 8.9. *The Fourier transform identifies univocally the measure: if μ and ν are probability measures on $(H, \mathcal{B}(H))$ such that $\hat{\mu} = \hat{\nu}$, then $\mu = \nu$.*

Proof. We already know the result for the finite dimensional case, compare with the Lévy Theorem 1.2. In the general case we consider the projections P_n (see (8.2)); then, for each $h \in H$,

$$\hat{\mu}(P_n h) = \int_H \mathrm{e}^{\mathrm{i}\langle x, P_n h\rangle}\, \mu(\mathrm{d}x) = \int_{P_n(H)} \mathrm{e}^{\mathrm{i}\langle P_n x, P_n h\rangle}\, \mu_n(\mathrm{d}x) = \hat{\mu}_n(P_n h)$$

and similarly for ν:

$$\hat{\nu}(P_n h) = \hat{\nu}_n(P_n h).$$

Hence, the measures μ_n and ν_n have the same Fourier transform, so they coincide by teh Lévy Theorem 1.2, and the conclusion follows from Proposition 8.3.

□

8.2 Gaussian measures

In this section we introduce the family of Gaussian measures in an infinite dimensional, separable, Hilbert space H. Recently, this subject has gained some interest from the mathematical analysis point of view; indeed this theory is also linked to infinite dimensional heat equations and even to heat equations perturbed by a linear drift term, see for instance the book by `Da Prato and Zabczyk` [DZ02] and the lecture notes by `Da Prato` [Da01] and `Rhandi` [MR04].

By definition, $\mu \in \mathcal{M}(H)$ is a *Gaussian measure* if, for any $x \in H$, the real valued random variable $X : (H, \mathcal{B}(H), \mu) \to \mathbb{R}$, $X(\omega) = \langle x, \omega\rangle$, $\omega \in H$, possesses a Gaussian law.

To investigate Gaussian measures our main tool will be the Fourier transform.

Given a trace class operator $Q \in L_1^+(H)$, there exists a complete orthonormal system $\{e_k\}$ in H and a sequence $\{\sigma_k^2\}$ of non-negative real numbers such that

$$Q e_k = \sigma_k^2 e_k, \qquad k \in \mathbb{N},$$

and also

$$\mathrm{Tr}(Q) = \sum_{k=1}^{\infty} \sigma_k^2 < +\infty.$$

To simplify our exposition, we will assume that $\mathrm{Ker}(Q) = \{0\}$, i.e., $\sigma_k^2 > 0$ for each $k \in \mathbb{N}$. Therefore, we may and will suppose that all the eigenvalues are ordered in a decreasing order. Set further $x_k = \langle x, e_k\rangle$, so that

$$x = \sum_{k=1}^{\infty} x_k e_k, \qquad |x|^2 = \sum_{k=1}^{\infty} x_k^2.$$

We introduce the natural isomorphism γ between H and ℓ^2 ([1]):

$$x \in H \mapsto \gamma(x) = \{x_k\} \in \ell^2.$$

In the following, we shall identify H with ℓ^2.

Note that l^2 is a Borel subset of $\mathbb{R}^\infty = l^\infty$. Recall that $\mathbb{R}^\infty$ is the vector space of all real valued sequences, endowed with the product topology; it is also a complete separable metric space, endowed with a suitable distance, see the first lecture. We denote by $\mathcal{B}(\mathbb{R}^\infty)$ the Borel σ-algebra of $\mathbb{R}^\infty$.

Theorem 8.10. *We are given a trace class operator $Q \in L_1^+(H)$ and a vector $a \in H$. Then there exists a (unique) probability measure μ on $(H, \mathcal{B}(H))$ such that*

$$\int_H e^{i\langle x,h\rangle}\,\mu(dx) = e^{i\langle a,h\rangle} e^{-\frac{1}{2}\langle Qh,h\rangle}. \tag{8.6}$$

Further, μ is the restriction to ℓ^2 of the product measure

$$\mathcal{N}(a,Q) = \prod_{k=1}^{\infty} \mathcal{N}(a_k, \sigma_k^2)$$

defined on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$.

Proof. Since Fourier transform uniquely determines a probability measure, it remains to show existence of μ (this is not immediate, compare with Remark 8.8.

Consider the sequence of real Gaussian measures $\mu_k = \mathcal{N}(a_k, \sigma_k^2)$, $k \in \mathbb{N}$, and take the product measure $\mu = \prod_{k=1}^{\infty} \mathcal{N}(a_k, \sigma_k^2)$ on the algebra $\mathcal{R} \subset \mathcal{B}(\mathbb{R}^\infty)$ of all ** cylindrical sets $I_{k_1,\dots,k_n,A}$, where $n \in \mathbb{N}$, $k_1, \dots, k_n \in \mathbb{N}$ and $A \in \mathcal{B}(\mathbb{R}^n)$,

$$I_{k_1,\dots,k_n,A} := \{x \in \mathbb{R}^\infty \,:\, (x_{k_1}, \dots, x_{k_n}) \in A\}.$$

Note that

$$\mu(I_{k_1,\dots,k_n,A}) = (\prod_{i=1}^{n} \mu_{k_i})(A)$$

Clearly, μ is additive on $\mathcal{R}$ and it is continuous in 0, since each component is such. Then, by Caratheodory's theorem 1.5, it can be extended to a σ-additive measure on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$.

Next, we show that μ is concentrated on ℓ^2; to this purpose, it is enough to notice that $\int_{\mathbb{R}^\infty} |x|_H^2\,\mu(dx) < \infty$. Indeed, we have easily:

$$\int_{\mathbb{R}^\infty} |x|_H^2\,\mu(dx) = \mathrm{Tr}(Q) + |a|^2. \tag{8.7}$$

Hence $\mu(l^2) = 1$ and so (8.7) holds also when $\mathbb{R}^\infty$ is replaced by H.

Finally, we consider the restriction of μ to ℓ^2, that we continue to denote with μ. Our goal is to show that formula (8.6) holds. Let $P_n : \ell^2 \to \mathbb{R}^n$ be the projection on the first n coordinates, and $\nu_n = \prod_{k=1}^{n} \mu_k$. Then

---

[1] ℓ^2 is the space of real valued, square integrable sequences, endowed with the scalar product $\langle x, y\rangle_2 = \sum_{k=1}^{\infty} x_k y_k$ for $x, y \in \ell^2$.

$$\int_{\ell^2} e^{i\langle x,h\rangle}\,\mu(dx) = \lim_{n\to\infty}\int_{\ell^2} e^{i\langle P_n x,P_n h\rangle}\,\mu(dx)$$
$$= \lim_{n\to\infty}\int_{\mathbb{R}^n} e^{i\langle P_n x,P_n h\rangle}\,\nu_n(dx)$$
$$= \lim_{n\to\infty} e^{i\langle P_n a,P_n h\rangle-\frac{1}{2}\langle QP_n h,P_n h\rangle} = e^{i\langle a,h\rangle-\frac{1}{2}\langle Qh,h\rangle}.$$

□

Problem 8.2. Show that if $\dim H=\infty$ there exists a functional $\phi : H \to \mathbb{C}$ which is continuous, positive defined and such that $\phi(0)=1$, but ϕ is not the Fourier transform of a probability measure μ on $(H,\mathcal{B}(H))$.
Hint: think about the hypothesis we use in the above theorem...

Problem 8.3 (Computing some Gaussian integrals).

1. Conclude the proof of Theorem 8.10 by showing that equality (8.7) holds.
2. Fix $a\in H$, $Q\in L_1^+(H)$ with $\mathrm{Ker}Q=\{0\}$. Setting $\mu\sim\mathcal{N}(a,Q)$, show that a is the mean and Q the covariance operator of μ, i.e.,
$$\int_H x\,\mu(dx)=a,$$
and, for $h,k\in H$,
$$\int_H \langle x,h\rangle\,\langle x,k\rangle\,\mu(dx)=\langle Qh,k\rangle.$$
3. μ has finite variance, in the sense that
$$\int_H |x-a|^2\,\mu(dx)=\mathrm{Tr}(Q). \tag{8.8}$$
4. Assume $T\in L(H)$ is a bounded operator and fix $b\in H$; define the affine mapping $\Gamma x=Tx+b$. Then, the law imagine of μ under Γ is given by the formula
$$\Gamma\circ\mathcal{N}(a,Q)=\mathcal{N}(Ta+b,TQT^*).$$
5. Show that, for any $h\in H$, we have
$$\int_H e^{\langle x,h\rangle}\,\mu(dx)=e^{\langle a,h\rangle}e^{\frac{1}{2}\langle Qh,h\rangle}. \tag{8.9}$$

Take $(\Omega,\mathcal{F},\mu)$ as the reference probability space, where Ω is a separable Hilbert space and μ is a Gaussian measure on Ω. We consider random variables taking values in a separable Hilbert space H; the space $L^2(\Omega,\mu;H)$ is the space of (equivalence classes of) random variables such that

$$\mathbb{E}[|X|^2]=\int_\Omega |X(\omega)|^2\,\mu(d\omega)<\infty.$$

Of course a random variable X on $(\Omega,\mathcal{F},\mu)$ is called Gaussian if its law is a Gaussian measure on H. Gaussian random variables form a closed set in the space of square integrable random variables, as the following proposition shows. This result will be useful often in the sequel.

Proposition 8.11. *Let* $\{X_n,\ n \in \mathbb{N}\}$ *be a sequence of Gaussian random variables with values in an Hilbert space* H*, having distributions* $\mathcal{N}(a_n, Q_n)$*. Assume that* $X_n \to X$ *in* $L^2(\Omega, \mu; H)$*; then* X *is a Gaussian random variable, with mean* $a = \lim a_n$ *and covariance operator*

$$\langle Qh, h\rangle = \lim_{n\to\infty} \langle Q_n h, h\rangle.$$

Proof. We compute the characteristic function of X, i.e., $\mathbb{E}[\mathrm{e}^{i\langle X(\omega),y\rangle_H}]$ and we approximate with X_n, using the assumed convergence, to get

$$\mathbb{E}[\mathrm{e}^{i\langle X(\omega),y\rangle_H}] = \lim_{n\to\infty} \mathbb{E}[\mathrm{e}^{i\langle X_n(\omega),y\rangle_H}]$$

but since X_n has Gaussian distribution, the equality becomes

$$\mathbb{E}[\mathrm{e}^{i\langle X(\omega),y\rangle_H}] = \lim_{n\to\infty} \mathrm{e}^{\mathrm{i}\langle a_n,y\rangle - \frac{1}{2}\langle Q_n y,y\rangle} = \mathrm{e}^{\mathrm{i}\langle a,y\rangle - \frac{1}{2}\langle Qy,y\rangle}.$$

This proves at once all the claims.
□

8.3 White noise

In this section we fix a real separable Hilbert space H and a Gaussian measure

$$\mu = \mathcal{N}(0, Q)$$

on $(H, \mathcal{B}(H))$, where $Q \in L_1^+(H)$ verifies $\mathrm{Ker}(Q) = \{0\}$. We let $\{e_k,\ k \in \mathbb{N}\}$ be a complete orthonormal system in H associated to the eigenvalues σ_k^2 of Q. The idea of “white noise analysis”, due to Hida [Hi80], is to consider white noise, rather than Brownian motion, as the fundamental object to study: a complete treatment of this approach can be found in references [Hi93] or [Ho96].

Also, a related approach to white noise analysis is given by the so-called isonormal Gaussian processes, see `Dudley` [Du02, page 350] and `Nualart` [Nu95, Chapter 1].

Let us denote $Q(H)$ the image of H under Q. Notice that $Q(H)$ is a subset of H, strictly contained in H. Actually, setting $y = Qx$, we get

$$\sum_{k=1}^{\infty} \left(\frac{y_k}{\sigma_k^2}\right)^2 = \sum_{k=1}^{\infty} (x_k)^2 < \infty,$$

hence

$$Q(H) = \left\{ y \in H \,:\, \sum_{k=1}^{\infty} \left(\frac{y_k}{\sigma_k^2}\right)^2 < \infty \right\}.$$

Proposition 8.12. *We have* $\mu(Q(H)) = 0$.

Proof. For any $n, k \in \mathbb{N}$ set

$$U_n = \left\{ y \in H \,:\, \sum_{i=1}^{\infty} \frac{y_i^2}{\sigma_i^2} < n^2 \right\}$$

and

$$U_{n,k} = \left\{ y \in H \,:\, \sum_{i=1}^{2k} \frac{y_i^2}{\sigma_i^2} < n^2 \right\}.$$

Since $U_n \uparrow Q(H)$ as $n \to \infty$ and $U_{n,k} \downarrow U_n$ as $k \to \infty$, it is enough to show that

$$\mu(U_n) = \lim_{k \to \infty} \mu(U_{n,k}) = 0 \tag{8.10}$$

The measure of $U_{n,k}$ is a finite dimensional integral with respect to the Gaussian measure $\mathcal{N}(0, S^2)$, where $S^2 = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_{2k}^2)$

$$\mu(U_{n,k}) = \int_{\{y \in \mathbb{R}^{2k} \,:\, \sum_{i=1}^{2k} \frac{y_i^2}{\sigma_i^2} < n^2\}} \mathcal{N}(0, S^2)(\mathrm{d}y)$$

by the change of variables $x_i = y_i/\sigma_i$ we have

$$\mu(U_{n,k}) = \int_{\{x \in \mathbb{R}^{2k} \,:\, |x|^2 < n^2\}} \mathcal{N}(0, I_{2k})(\mathrm{d}x)$$

where I_{2k} is the identity matrix of size $2k$; passing to polar coordinates we have

$$\mu(U_{n,k}) = \frac{1}{(k-1)!} \int_0^n \mathrm{e}^{-r^2/2} r^{2k-1} \,\mathrm{d}r = \frac{1}{(k-1)!} \int_0^{n^2/2} \mathrm{e}^{-\rho} \rho^{k-1} \,\mathrm{d}\rho \le \frac{1}{k!} (\tfrac{n^2}{2})^k \to 0$$

as $k \to \infty$ and (8.10) follows.

□

Now we define the operator $Q^{1/2}$ by setting

$$Q^{1/2}(x) = \sum_{k=1}^{\infty} \sigma_k \langle x, e_k \rangle e_k, \qquad x \in H.$$

Clearly, $Q^{1/2}(H) \subset H$. This space is the so-called *reproducing kernel* related to the measure $\mathcal{N}(0, Q)$.

Lemma 8.13. *The reproducing kernel space $Q^{1/2}(H)$ is dense in H.*

Proof. We use that $\mathrm{Ker}(Q) = \{0\}$. Assume that $x_0 \in H$ is such that $\langle Q^{1/2} x, x_0 \rangle = 0$ for every $x \in H$: this means that $Q^{1/2}(x_0) = 0$ hence $Q(x_0) = 0$. But since $\mathrm{Ker}(Q) = \{0\}$ this implies $x_0 = 0$.

□

Further, one can show that

$$Q^{1/2}(H) = \left\{ y \in H \,:\, \sum_{k=1}^{\infty} \frac{y_k^2}{\sigma_k} < \infty \right\}.$$

Let us now consider $(H, \mathcal{B}(H), \mu)$ as the reference probability space.

For every $z \in Q^{1/2}(H)$, the mapping $W_z : H \to \mathbb{R}$, $W_z(x) = \langle Q^{-1/2} z, x\rangle$, defines a (real) Gaussian random variable, centered, with variance $|z|_H^2$; this remark can be extended to a any finite choice of elements in the reproducing kernel space $Q^{1/2}(H)$.

Problem 8.4. Let $z_1, \dots, z_n \in Q^{1/2}(H)$ and consider the random variables W_{z_i}, on $(H, \mathcal{B}(H), \mu)$

$$W_{z_i}(x) = \langle Q^{-1/2} z_i, x\rangle.$$

The law of the random variable $(W_{z_1}, \dots, W_{z_n})$ in $\mathbb{R}^n$ is a Gaussian measure with covariance operator $Q = (q_{i,j})$

$$q_{i,j} = \langle z_i, z_j\rangle, \qquad i,j = 1, \dots, n.$$

In particular, the random variables $W_{z_1}, \dots, W_{z_n}$ are independent if and only if the vectors $z_1, \dots, z_n$ forms an orthonormal system.

Hence, the mapping $W : z \mapsto W_z$ is well defined as a mapping from $Q^{1/2}(H)$ into the space of square integrable random variables $L^2(H, \mu) = L^2(H, \mu, \mathbb{R})$. Our aim is to extend this mapping to the whole space H. The resulting application $W : H \to L^2(H, \mu)$ will be called the *white noise* mapping; it will be important in the sequel, for instance in defining the Brownian motion.

Consider again the mapping $W : Q^{1/2}(H) \to L^2(H, \mu)$:

$$W_z(x) = \langle x, Q^{-1/2} z\rangle;$$

since for $z_1, z_2 \in H$ we have

$$\mathbb{E}[W_{z_1} W_{z_2}] = \int W_{z_1}(x) W_{z_2}(x)\, \mu(\mathrm{d}x) = \langle z_1, z_2\rangle$$

it holds that W is an isometry between $Q^{1/2}(H)$ and $L^2(H, \mu)$. Since its domain $Q^{1/2}(H)$ is dense in H, it can be uniquely extended to an isometry $W : H \to L^2(H, \mu)$.

We left as an exercise to verify the properties of the white noise mapping. The proof follows easily using finite-dimensional projections.

Problem 8.5. Fix $n \in \mathbb{N}$, and a sequence $f_1, \dots, f_n \in H$.

1. $(W_{f_1}, \dots, W_{f_n})$ is a Gaussian random variable in $\mathbb{R}^n$, centered, with covariance operator
$$Q_{ij} = \langle f_i, f_j\rangle, \qquad i,j = 1, \dots, n.$$
2. $W_{f_1}, \dots, W_{f_n}$ are independent if and only if $(f_1, \dots, f_n)$ is an orthonormal system in H.

We conclude introducing the exponential mapping $f \mapsto \mathrm{e}^{W_f}$. Choose first $f \in Q^{1/2}(H)$. Then

$$\int \mathrm{e}^{\langle x, Q^{-1/2} f\rangle} \mu(\mathrm{d}x) = \mathrm{e}^{\frac{1}{2}\langle Q Q^{-1/2} f, Q^{-1/2} f\rangle} = \mathrm{e}^{|f|^2/2}.$$

Further, the mapping is continuous on $Q^{1/2}(H)$. Indeed, for $f, g \in Q^{1/2}(H)$, we have

$$\int [\mathrm{e}^{W_f} - \mathrm{e}^{W_g}]^2 \mu(\mathrm{d}x) = \mathrm{e}^{2|f|^2} - 2\mathrm{e}^{|f+g|^2/2} + \mathrm{e}^{2|g|^2} = [\mathrm{e}^{|f|^2} - \mathrm{e}^{|g|^2}]^2 + 2\mathrm{e}^{|f|^2+|g|^2}[1 - \mathrm{e}^{-|f-g|^2/2}]$$

which shows the continuity in f of e^{W_f}. By using a similar computation as above we obtain the following proposition.

Proposition 8.14. *The map $f \mapsto \mathrm{e}^{W_f}$ from $H \to L^2(H, \mu)$ is continuous.*

8.4 A construction of Brownian motion using the white noise mapping

Fix the set of times $T = [0, \infty)$, and let $H = L^2(T)$. Fix an arbitrary operator $Q \in L_1^+(H)$, with $\ker(Q) = \{0\}$, and let $\mu = \mathcal{N}(0, Q)$ be a Gaussian measure on $(H, \mathcal{B}(H))$. We have thus defined the reference probability space

$$(H, \mathcal{B}(H), \mu) = (\Omega, \mathcal{F}, \mathbb{P}).$$

Theorem 8.15. *Set* $B(0) = 0$ *and* $B(t) = W(\mathbb{1}_{[0,t]})$, $t \in T$, *where*

$$\mathbb{1}_{[0,t]}(s) = \begin{cases} 1, & \text{for } s \in [0, t], \\ 0, & \text{otherwise,} \end{cases}$$

and W *is the white noise mapping introduced in Section 8.3. Then* $B = \{B_t,\ t \in T\}$ *is a real Brownian motion on* $(H, \mathcal{B}(H), \mu)$ *(with respect to its natural completed filtration).*

Proof. We verify the properties stated in Definition 2.5. Condition 1. is obvious. Let $0 < s < t$: since

$$B(t) - B(s) = W(\mathbb{1}_{[s,t]}),$$

it follows from Exercise 8.5 that $B(t)$ verifies condition 2. Further, since the functions

$$\mathbb{1}_{[0,t_1]}, \mathbb{1}_{[t_1,t_2]}, \ldots, \mathbb{1}_{[t_{n-1},t_n]}$$

are orthogonal in H, condition 3. follows again from Exercise 8.5. We would have finished if we know that the process B (or a version of B) has continuous sample paths. This part of the proof is the object of next section.

□

8.4.1 The factorization method

In this section we prove continuity of the trajectories of the Brownian motion using the *factorization method* by `L. Schwartz` (in the finite dimensional case) and `Da Prato and Zabczyk` (in the general situation). This is a purely analytic approach which does not require the continuity theorem of Kolmogorov 2.16.

The starting point is provided by the following identity

$$\int_s^t (t-\sigma)^{\alpha-1} (\sigma - s)^{-\alpha} \, d\sigma = \frac{\pi}{\sin(\pi\alpha)}, \qquad 0 \le s < t, \tag{8.11}$$

which holds for each $\alpha \in]0, 1[$. We can also write, for $0 \le s < t$,

$$\mathbb{1}_{[0,t]}(s) = \frac{\sin(\pi\alpha)}{\pi} \int_0^t (t-\sigma)^{\alpha-1} \mathbb{1}_{[0,\sigma](s)} (\sigma - s)^{-\alpha} \, d\sigma.$$

Then, formula (8.11) may be written in the following form

$$\mathbb{1}_{[0,t]} = \frac{\sin(\pi\alpha)}{\pi} \int_0^t (t-\sigma)^{\alpha-1} g_\sigma \, d\sigma,$$

where

$$g_\sigma(s) = \mathbb{1}_{[0,\sigma](s)}(\sigma - s)^{-\alpha}.$$

Choose $\alpha < \frac{1}{2}$, $g_\sigma \in H$ and $\|g_\sigma\|^2 = \frac{\sigma^{1-2\alpha}}{1-2\alpha}$: since the mapping

$$\Omega \to L^2(\Omega, \mu), \qquad f \mapsto W_f$$

is continuous, we get the following representation for the Brownian motion B:

$$B(t) = \frac{\sin(\pi\alpha)}{\pi} \int_0^t (t-\sigma)^{\alpha-1} W(g_\sigma)\,\mathrm{d}\sigma. \tag{8.12}$$

Now it is enough to show that $\sigma \mapsto W_{g_\sigma} \in L^{2m}(0,T)$ $\mathbb{P}$-a.s. for each $T > 0$ and for some $m > \frac{1}{2\alpha}$. Indeed, then the continuity of the trajectories of B shall follow from next Lemma 8.16.

Notice that W_{g_σ} is a Gaussian random variable with normal law $\mathcal{N}(0, \frac{\sigma^{1-2\alpha}}{1-2\alpha})$, hence

$$\int_H |W_{g_\sigma}(\omega)|^{2m}\,\mathbb{P}(\mathrm{d}\omega) = C'_{m,\alpha}\sigma^{m(1-2\alpha)}.$$

Now since $\alpha\, m > \frac{1}{2}$, from Fubini's theorem we obtain

$$\int_0^T \left[\int_H |W_{g_\sigma}|^{2m}\,\mathbb{P}(\mathrm{d}\omega)\right]\mathrm{d}\sigma = \int_H \left[\int_0^T |W_{g_\sigma}|^{2m}\,\mathrm{d}\sigma\right]\mathbb{P}(\mathrm{d}\omega) < +\infty.$$

Hence $\sigma \mapsto W_{g_\sigma} \in L^{2m}(0,T)$ $\mathbb{P}$-a.s. for each $T > 0$, and we get the thesis.
□

Lemma 8.16. *Let H be a real separable Hilbert space; let $T > 0$, and consider a function $f \in L^{2m}(0,T;H)$ where $m > 1$. Set moreover*

$$F(t) = \int_0^t (t-\sigma)^{\alpha-1} f(\sigma)\,\mathrm{d}\sigma, \qquad t \in [0,T],$$

where $\alpha \in]\frac{1}{2m}, 1[$. Then $F \in C([0,T];H)$.

Proof. Using Hölder's inequality (notice that $2m\alpha - 1 > 0$) we have

$$|F(t)| \le \left(\int_0^t (t-\sigma)^{2m(\alpha-1)/(2m-1)}\,\mathrm{d}\sigma\right)^{(2m-1)/2m} \|f\|_{L^{2m}(0,T;H)}. \tag{8.13}$$

Therefore $F \in L^\infty(0,T;H)$. We still have to prove that F is continuous.

Continuity in 0 follows directly from (8.13), hence it remains to show that F is continuous in $[t_0,T]$ for each $t_0 \in (0,T]$. Set, for $\epsilon < t_0$,

$$F_\epsilon(t) = \int_0^{t-\epsilon} (t-\sigma)^{\alpha-1} f(\sigma)\,\mathrm{d}\sigma, \qquad t \in [t_0,T].$$

F_ϵ is obviously continuous in $[t_0,T]$. Moreover, using again Hölder's inequality, we have

$$|F_\epsilon(t) - F(t)| \leq M \left(\frac{2m-1}{2m\alpha - 1} \right)^{(2m-1)/2m} \epsilon^{(2m\alpha-1)/2m} \|f\|_{L^{2m}(0,T;H)}.$$

Therefore $\lim_{\epsilon \to 0} F_\epsilon(t) = F(t)$ uniformly on $[t_0, T]$, so that F is continuous as required.
□

Problem 8.6. Using a suitable modification of the above argument, prove that Brownian sample paths are Hölder continuous.

9

Foundations of Malliavin calculus and stochastic integration

In this chapter we introduce the general concepts that form the basis of the differential calculus on a Gaussian space. The material of this lecture was mainly inspired by the introduction to Malliavin calculus provided by `D. Nualart` [Nu95] as it was applied in the Ph.D. thesis [Bon98]. It is quite difficult to cite some names among the vast literature on the subject, however it was a rich source of inspiration also the book by `Bogachev` [Bo98]. We should of course mention the monograph [Ma97] which presents the original approach by Malliavin.

As in Chapter 8 we start with a real separable Hilbert space H of the form $L^2(T, \mathcal{B}(T), \lambda)$, where $T \subset \mathbb{R}^d$ is the set of times (usual choices are $T = [0,1]$ or $T = \mathbb{R}_+$) and λ is Lebesgue measure on T. We denote the norm and the inner product on H by $|\cdot|$ and $\langle \cdot, \cdot \rangle$.

On the space $(H, \mathcal{B}(H))$ we consider an irreducible, trace class operator Q and we consider a Gaussian probability measure $\mathbb{P} = \mathcal{N}(0, Q)$; we will denote this space with the usual notation $(\Omega, \mathcal{F}, \mathbb{P})$; it is a complete probability space. In that case we can consider the *white noise* $W(\mathbf{1}_A) = W_{1_A}$ based on λ, for $A \in \mathcal{B}(T)$. As we have seen in previous lecture, W is an isometry between H and the space of real Gaussian random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Following the literature, we call the space $(\Omega, \mathcal{F}, \mathbb{P}, H)$ a *Gaussian probability space.*

Throughout this lecture we will always work on

$$(\Omega, \mathcal{F}, \mathbb{P}, H).$$

When $T = \mathbb{R}_+$, we introduce the real Brownian motion $B = \{B_t\}$ (starting from 0) defined by $B_t = W(\mathbb{1}_{[0,t]})$, $t \in T$. Then we focus our interest on the *Wiener integral* of a square integrable function $f \in L2(T, \lambda)$. This is defined via the white noise mapping as $W(f)$. We also write

$$W(f) = \int f(t)\,\mathrm{d}B_t.$$

This integral generates the family of Gaussian variables

$$\mathcal{H}_1 = \{W(f),\ f \in H\}.$$

White noise analysis

For completeness, we mention that in the white noise analysis, the above space $\mathcal{H}_1$ is called the *first Wiener chaos*. The fundamental result in this direction is that the space $L^2(\Omega)$ admits a unique decomposition into orthogonal spaces $\mathcal{H}_n$:

$$L^2(\Omega, \mathcal{F}, \mathbb{P}) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n$$

where $\mathcal{H}_1$ is the previous space of Gaussian random variables.

The theory of multiple Itô-Wiener integrals was initiated by Itô in [It53]; due to its interest, we provide here a very short introduction to it.

The main interest is in the representation of square integrable random variables in terms of (multiple) Wiener integrals. We can define the *multiple Wiener-Itô integral* for a function f_m in $H^{\otimes m} = L^2(T^m)$ by

$$I_m(f_m) = \int_{T^m} f(t_1, \dots, t_m) \, \mathrm{d}B(t_1) \dots \mathrm{d}B(t_m)$$

by a standard procedure: we first define the integral for step functions $a_{i_1,\dots,i_m} \mathbb{1}_{A_{i_1} \times \dots \times A_{i_m}}(t_1, \dots, t_m)$, then we prove that these functions are dense in $L^2(T^m)$ and finally we extend the integral to a linear and continuous operator on $L^2(T^m)$.

As a consequence, it is possible to prove that any square integrable random variable $F \in L^2(\Omega)$ can be expanded into a series of multiple stochastic integrals

$$F = \sum_{m=0}^{\infty} I_m(f_m) \tag{9.1}$$

with $f_0 = \mathbb{E}[F]$, for suitable functions $f_m \in H^{\otimes m}$ which can be taken symmetric in order that the expansion is unique. Since the orthogonality $\mathbb{E}[I_m(f_m) I_n(g_n)] = 0$ holds, each term in the series corresponds to a different Wiener chaos.

Remark 9.1. There is interest in studying white noise also from the view point of the possible applications. In physics, one frequently encounters Gaussian processes and, among them, not only the Brownian motion has a preminent position. Langevin first introduced the *Gaussian white noise* process, that is a process characterized by

$$\mathbb{E}[\xi_t] = 0, \qquad \mathbb{E}[\xi_t \xi_s] = \delta(t - s),$$

where δ is the Dirac measure. White noise process is not a stochastic process in a strict mathematical sense; however, in physics it is often used as a model for very fast oscillations.

The above process enters in the physical construction of the Ornstein-Uhlenbeck process. This process was first introduced as a model for Brownian movement in the 1920's; for an hystorical introduction we refer for instance to Nelson [Ne01].

The last part of the lecture introduces the derivative operator of random variables $F \in L^2(\Omega)$ with respect to ω, which is treated as a parameter. The original presentation is due to P. Malliavin, who first introduced a notion of derivatives of Wiener functionals and applied it to the absolute continuity and the smoothness of density of the probability law induced by the solution of the stochastic differential equation at a fixed time; a basic reference on Malliavin calculus is the monograph [Ma97]. The notation $D_t F$, which we will use for this weak derivative of F, has been introduced in [NZ86]; it well represents the idea of partial derivative with respect to the infinite dimensional vector $\omega = \omega_t$, $t \in T$

9.1 Wiener integral

The aim of this section is to define a stochastic integral for elements $f \in H = L^2(\mathbb{R}_+, \mathrm{d}\lambda)$, i.e. $T = \mathbb{R}_+$,

$$W(f) = \int_T f(s)\,\mathrm{d}B_s. \tag{9.2}$$

By construction, $W(f)$ is a Gaussian random variable with law $\mathcal{N}(0, |f|^2)$.

Remark 9.2. Suppose that f is a step function

$$f(t) = \sum_{j=0}^{n} a_j \mathbb{1}_{A_j}(t),$$

where $a_j \in \mathbb{R}$ and $A_j = [t_j, t_{j+1}) \in \mathcal{B}(\mathbb{R}_+)$. Then one has

$$W(f) = \sum_{j=0}^{n} a_j \big(B(t_{j+1}) - B(t_j)\big).$$

Note that, for given $f \in H$, we have the following adaptedness property: $W(f)$ is $\mathcal{F}_t$-measurable if and only if $f \equiv 0$ on $[t, \infty)$.

We define, as time $t \geq 0$ varies, the family of random variables

$$I_t = \int_0^t f(r)\,\mathrm{d}B_r = \int_T \mathbb{1}_{[0,t]}(r) f(r)\,\mathrm{d}B_r.$$

By construction, it follows that this process is adapted to the filtration $\{\mathcal{F}_t,\ t \geq 0\}$ of Brownian motion, hence the stochastic process $I = \{I_t, t \geq 0\}$ is well defined.

Proposition 9.3. *The stochastic process $\{I_t, t \geq 0\}$ is a Gaussian process with correlation function*

$$\mathbb{E}(I_t I_r) = \int_0^{t \wedge r} f^2(s)\,\mathrm{d}s.$$

Proof. Since $f\mathbb{1}_{[0,t]} \in H$, it holds that I_t is a centered Gaussian random variable with variance $\sigma^2(t) = |f\mathbb{1}_{[0,t]}|^2$.

We show next that it is a Gaussian family. We claim that for every choice of times $t_1, \dots, t_n$ and real numbers $\alpha_1, \dots, \alpha_n$, the random variable $\alpha_1 I(t_1) + \cdots + \alpha_n I(t_n)$ has a Gaussian distribution. However, this follows from the linearity of the white noise mapping.

It remains to compute the correlation $\mathbb{E}(I_t I_r)$. Take $r < t \in T$, then we have

$$\mathbb{E}(I_t I_r) = \mathbb{E}\left[W(f\mathbb{1}_{[0,t]})W(f\mathbb{1}_{[0,r]})\right]$$

and the properties of the white noise mapping implies

$$\mathbb{E}(I_t I_r) = \langle f\mathbb{1}_{[0,t]}, f\mathbb{1}_{[0,r]}\rangle = \int_0^{t\wedge r} f^2(s)\,\mathrm{d}s.$$

□

Problem 9.1. Let $f \in C^1([0,t])$; by using a suitable approximation of f show that it holds, a.s.,

$$\int_0^t f(r)\,\mathrm{d}B_r = f(t)B_t - \int_0^t f'(r)\,B_r\,\mathrm{d}r. \tag{9.3}$$

9.1.1 Multidimensional Brownian motion

Consider the following construction. Suppose that the parameter space is $T = \mathbb{R}_+ \times \{1,\dots,d\}$ and the measure λ is the product of the Lebesgue measure times the uniform measure that gives mass one to each point $1,2,\dots,d$. Then we have $H = L^2(T,\lambda) \equiv L^2(\mathbb{R}_+;\mathbb{R}^d)$. In this situation we have that $B_t = (B_t^1,\dots,B_t^d)$ is a standard d-dimensional Brownian motion, where $B_t^i = W(\mathbb{1}_{[0,t]\times i})$, $t \geq 0$, $1 \leq i \leq d$, are real standard Brownian motions. Furthermore, for any $h \in H$, the random variable $W(h)$ can be obtained as the stochastic integral

$$W(h) = \sum_{i=1}^d \int_{\mathbb{R}_+} h^i(t)\,\mathrm{d}B_t^i.$$

We list some properties of the d-dimensional Brownian motion in Problem 9.4.

Exercises

Problem 9.2. Let us consider the following process, called the *Brownian bridge*

$$X_t = B_t - tB_1, \qquad t \in [0,1].$$

Prove that it is a centered Gaussian process and compute its mean and covariance functions.

Problem 9.3. Given $f \in L^2(0,T)$, compute $\mathbb{E}\left[\exp\left(\int_0^T f(t)\,\mathrm{d}B_t\right)\right]$.

Problem 9.4. Let $\{X_t,\ t \geq 0\}$ be a d-dimensional Gaussian process;

1. show that it is a Brownian motion if and only if
$$\mathbb{E}[X_t^i X_s^j] = \delta_{ij}(t \wedge s).$$
2. Let $A : \mathbb{R}_+ \to M(d,n)$ (the space of all real $d \times n$ matrices) and $z : \mathbb{R}_+ \to \mathbb{R}^n$ be a bounded measurable functions. Let $\{B_t, t \geq 0\}$ be an n-dimensional Brownian motion.
Show that $Y_t = A_t B_t + z_t$ is a Gaussian process and compute its mean and covariance functions.
3. Let $\{B_t,\ t \geq 0\}$ be a d-dimensional standard Brownian motion; show that for any $z \in \mathbb{R}^d$, $\|z\| = 1$, the process $X_t = \langle z, B_t\rangle$ is a real standard Brownian motion respect to the same initial filtration.
4. Let $\{W_t,\ t \geq 0\}$ be a d-dimensional standard Brownian motion; determine the class of matrices $A \in M(d,d)$ such that $Y_t = AW_t$ is again a standard Brownian motion.

9.1.2 Ornstein-Uhlenbeck theory of Brownian movement

The theory of Brownian movement developed by Einstein is, in some sense, unsatisfactory. For instance, he supposed that the deplacement in the time interval (s, t) is independent from the past, thus it shall also be independent from the velocity at time s. Starting from `Langevin`, a new theory was developed in the 1920's, culminating with the `Ornstein-Uhlenbeck` theory of Brownian movement, where a different physical quantity, the *impulse* of the particle, was the central object of interest.

In this model, a faster time scale is considered; therefore, the time interval Δt is small with respect to the time needed in the Einstein model by the particle to lose memory of the velocity, but still great enough such that the number of hits in that interval is large.

Formally, the relevant equation is

$$\frac{\partial^2 X_t}{\partial t^2} = -\alpha \frac{\partial X_t}{\partial t} + \xi_t$$

which means that the force acting on the particle is split in two parts, the first of which depends on the viscosity of the fluid with viscous coefficient $m\alpha$ ([1]), while the second term is a *white noise* process –formally, the derivative of a Brownian motion $\{B_t,\ t \geq 0\}$ (defined on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$) with variance $\sigma^2 > 0$: $\xi_t = \sigma^2 \frac{\partial B_t}{\partial t}$.

We can give a meaning to the above equation by introducing the velocity $V_t = \partial X_t / \partial t$. Then, the velocity field V verifies the Langevin equation

$$\frac{\partial V_t}{\partial t} = -\alpha V_t + \xi_t, \tag{9.4}$$

Given an initial condition $V_0 = v_0$, (9.4) is understood in the integral sense, i.e., we have, a.s.,

$$V_t - v_0 - \alpha \int_0^t V_s \, \mathrm{d}s = \sigma^2 B_t, \quad t \geq 0.$$

One solves equation (9.4) using the variation-of-constants formula, and the solution is given by

$$V_t = \mathrm{e}^{-\alpha t} v_0 + \int_0^t \mathrm{e}^{-\alpha(t-s)} \, \sigma^2 \, \mathrm{d}B_s. \tag{9.5}$$

Indeed, if we introduce $Z_t = V_t - \sigma^2 B_t$ we get that

$$Z_t = v_0 + \alpha \int_0^t (Z_s + \sigma^2 B_s) \, \mathrm{d}s, \qquad t \geq 0.$$

Hence, a.s., Z_t is a C^1-solution to the Cauchy problem

$$\begin{cases} \dot{Z}_t = \alpha(Z_t + \sigma^2 B_t) \\ Z_0 = v_0. \end{cases}$$

which can be explicitly solved. Now, using (9.3), we readly arrive at (9.5).

[1] $\alpha > 0$ is a constant with the same dimension of frequency

By (9.5), we see that $V_t = V_t^{v_0}$, $t \geq 0$, is a Gaussian process with mean $\mathrm{e}^{-\alpha t} v_0$ and covariance

$$\mathbb{E}[V_t V_s] = \mathrm{e}^{-\alpha(t+s)} \sigma^2 \frac{\mathrm{e}^{2\alpha s} - 1}{2\alpha}, \qquad t \geq s \geq 0.$$

Its Fourier transform is given by

$$\mathbb{E}[\mathrm{e}^{\mathrm{i} V_t u}] = \mathrm{e}^{\mathrm{e}^{-\alpha t} v_0 u} \mathrm{e}^{\frac{1}{2}\sigma^2 \frac{1-\mathrm{e}^{2\alpha t}}{2\alpha} u^2}, \qquad u \in \mathbb{R}.$$

Therefore, the law of V_t converges weakly, as $t \to \infty$, to a zero mean Gaussian distribution with variance $\frac{\sigma^2}{2\alpha}$.

Proposition 9.4. *The Ornstein-Uhlenbeck process* $\{V_t^x,\ t \geq 0\}$, $x = v_0 \in \mathbb{R}$ *is Markovian with transition function given by*

$$p(t, x, A) = \mathbb{P}(V_t^x \in A) = \int_A \Big(\frac{\sqrt{2\alpha}}{\sqrt{2\pi\sigma^2(1 - \mathrm{e}^{-2\alpha t})}} \exp\Big(-\frac{\alpha(y - \mathrm{e}^{-\alpha t} x)^2}{(1 - \mathrm{e}^{-2\alpha t})\sigma^2} \Big) \Big) \mathrm{d}y, \quad A \in \mathcal{B}(\mathbb{R}). \tag{9.6}$$

Proof. The formula for the law of V_t^x follows easily by taking the Fourier transform. Let us prove that, for $f = \mathbb{1}_A$, with $A \in \mathcal{B}(\mathbb{R})$ we have

$$\mathbb{E}[f(V_{t+s}^x) \mid \mathcal{F}_s] = p(t, V_s^x, A)$$

We write, using the freezing lemma (see Lemma 5.2)

$$\begin{aligned}
\mathbb{E}(f(V_{t+s}^x) \mid \mathcal{F}_s) &= \mathbb{E}\left(f\Big(\mathrm{e}^{-\alpha(t+s)} x + \int_0^s \mathrm{e}^{-\alpha(t+s-r)}\, \mathrm{d}B_r + \int_s^{t+s} \mathrm{e}^{-\alpha(t+s-r)}\, \mathrm{d}B_r \Big) \mid \mathcal{F}_s \right) \\
&= \mathbb{E}\left(f\Big(y + \mathrm{e}^{-\alpha(t+s)} x + \int_s^{t+s} \mathrm{e}^{-\alpha(t+s-r)}\, \mathrm{d}B_r \Big) \right)\Big|_{y = \mathrm{e}^{-\alpha t} \int_0^s \mathrm{e}^{-\alpha(s-r)}\, \mathrm{d}B_r} \\
&= \mathcal{N}\Big(\mathrm{e}^{-\alpha t}(V_s^x - \mathrm{e}^{-\alpha s} x) + \mathrm{e}^{-\alpha(t+s)} x, \int_s^{t+s} \mathrm{e}^{-2\alpha(t+s-r)} \mathrm{d}r \Big)(A) \\
&= \mathcal{N}\Big(\mathrm{e}^{-\alpha t}(V_s^x - \mathrm{e}^{-\alpha s} x) + \mathrm{e}^{-\alpha(t+s)} x, \int_0^t \mathrm{e}^{-2\alpha u} \mathrm{d}u \Big)(A) \\
&= p(t, V_s^x, A),
\end{aligned}$$

where $\mathcal{N}(a, q)$ denotes the Gaussian measure with mean $a \in \mathbb{R}$ and variance $q > 0$. The assertion is proved.

□

If we fix the initial position $X(0) = x_0$, then since

$$X(t) = x_0 + \int_0^t V_s\, \mathrm{d}s$$

we have that the position of the particle subject to Brownian movement following the Ornstein-Uhlenbeck theory is a Gaussian process $X(t)$ with constant mean x_0 and covariance operator

$$\mathbb{E}[X(t)X(s)] = \frac{\sigma^2}{\alpha^2}\min\{t,s\} + \frac{\sigma^2}{2\alpha^3}\left(-2 + 2\mathrm{e}^{-\alpha t} + 2\mathrm{e}^{-\alpha s} - \mathrm{e}^{-\alpha|t-s|} - \mathrm{e}^{-\alpha(t+s)}\right).$$

It follows that the variance of $X(t)$ is

$$\frac{\sigma^2}{\alpha^2}t + \frac{\sigma^2}{2\alpha^3}\left(-2 + 4\mathrm{e}^{-\alpha t} - \mathrm{e}^{-2\alpha t}\right).$$

The first term in the sum is just the variance of the position following the Einstein's theory; the correction is, in percentage, of order $\frac{1}{\alpha t}$. A typical value for the friction coefficient is $\alpha^{-1} = 10^{-8}s^{-1}$: in the observation we make a small error if we adopt Einstein's value for the variance.

The next result shows in particular that the integral of an Ornstein-Uhlenbeck process is not Markovian.

Problem 9.5. Let $f, g \in H = L^2(0,T)$, and define

$$X(t) = \int_0^t f(s)\,\mathrm{d}B_s, \qquad Y(t) = \int_0^t X(s)g(s)\,\mathrm{d}s.$$

Prove (Y_t) is a Gaussian process in $[0,T]$, with mean function $m(t) = 0$ and covariance function

$$\rho(t,s) = \int_0^{t\wedge s} f^2(u)\left(\int_u^s g(v)\,\mathrm{d}v\right)\left(\int_u^t g(v)\,\mathrm{d}v\right)\mathrm{d}u.$$

Further, show that this process is neither a Markov process, nor a martingale.

Hint: a Gaussian process is Markovian if and only if its correlation function verifies $\rho(t_1,t_3) = \rho(t_1,t_2)\rho(t_2,t_3)$ for $t_1 \le t_2 \le t_3$, see `Feller` [Fe53]. This problem is taken from `Shreve` [Sh, pp. 289–].

9.2 The law of the Brownian motion in space of trajectories

Let $H = L^2(0,1)$ endowed with the Borel σ-field $\mathcal{B}(H)$, $\mu \sim \mathcal{N}(0,Q)$ a Gaussian measure on H and $B = \{B_t,\, t \in [0,1]\}$ be the Brownian motion on the time interval $[0,1]$, defined on the space $(H, \mathcal{B}(H), \mu)$. Our aim is to consider B as a random variable on the space of trajectories, as in Section 3.3; however, in this section, we shall see the trajectories in the space $L^2(0,1)$ of square integrable functions with respect to the Lebesgue measure. We use the notation of Lecture 8.

First, clearly the mapping $B : \Omega \to L^2(0,1)$, $B(\omega)(t) = B_t(\omega)$ is well defined, since

$$\mathbb{E}\left[\int_0^1 |B_t|^2\,\mathrm{d}t\right] = \int_\Omega\int_0^1 |B_t(\omega)|^2\,\mathrm{d}t\,\mu(\mathrm{d}\omega) = \int_0^1\int_\Omega |B_t(\omega)|^2\,\mu(\mathrm{d}\omega)\,\mathrm{d}t = \int_0^1 t\,\mathrm{d}t = \frac{1}{2}$$

so that $B(\omega) \in L^2(0,1)$ for almost all ω. Moreover, we have

$$B \in L^2(\Omega; L^2(0,1)).$$

We define the law of B on $L^2(0,1)$, i.e., the measure $\mathbb{P} = \mu \circ B^{-1}$, the *Wiener measure* in $L^2(0,1)$.

Proposition 9.5. $\mathbb{P}$ *is a Gaussian measure on* $L^2(0,1)$*, with zero mean and covariance operator* $\mathcal{R}$ *given by*

$$(\mathcal{R}h)(t) = \int_0^1 \min\{s,t\} h(s)\,\mathrm{d}s, \qquad h \in L^2(0,1), \quad t \in (0,1). \tag{9.7}$$

Proof. To show that $\mathbb{P}$ is a Gaussian measure, define for $n \in \mathbb{N}$ and $t \in (0,1)$

$$B_n(t;\omega) = W(P_n \mathbb{1}_{[0,t]})(\omega) = \langle \omega, Q^{-1/2} P_n \mathbb{1}_{[0,t]} \rangle = \sum_{k=1}^n \sigma_k^{-1} \langle \omega, e_k \rangle \langle \mathbb{1}_{[0,t]}, e_k \rangle,$$

where P_n is introduced in (8.2) and the mapping $B_n : \Omega \to L^2(0,1)$ given by $B_n(\omega)(t) = B_n(t;\omega)$. B_n is a linear, bounded operator, so it is a $L^2(0,1)$-valued random variable with Gaussian law $\mathcal{L}_n = \mathbb{P} \circ B_n^{-1} \sim \mathcal{N}(0, B_n Q B_n^*)$, where B_n^* is the transpose of B_n. We have in fact

$$\int_{L^2(0,1)} \mathrm{e}^{\mathrm{i}\langle \xi, h \rangle}\, \mathcal{L}_n(\mathrm{d}\xi) = \int_\Omega \mathrm{e}^{\mathrm{i}\langle B_n \omega, h \rangle}\, \mathbb{P}(\mathrm{d}\omega) = \int_\Omega \mathrm{e}^{\mathrm{i}\langle \omega, B_n^* h \rangle}\, \mathbb{P}(\mathrm{d}\omega) = \mathrm{e}^{-\frac{1}{2}\langle Q B_n^* h, B_n^* h \rangle} = \mathrm{e}^{-\frac{1}{2}\langle B_n Q B_n^* h, h \rangle}$$

which shows the claim.

Next step is to prove that

$$\lim_{n \to \infty} B_n = B \qquad \text{in } L^2(\Omega, \mu; L^2(0,1)). \tag{9.8}$$

Then from Proposition 8.11, it will follow that B has a Gaussian distribution.

Let us compute, using Fubini's theorem,

$$\begin{aligned}
\mathbb{E}[|B_n - B|^2_{L^2(0,1)}] &= \int_\Omega \int_0^1 |B(t;\omega) - B_n(t;\omega)|^2 \,\mathrm{d}t\, \mu(\mathrm{d}\omega) \\
&= \int_\Omega \int_0^1 |W((1-P_n)\mathbb{1}_{[0,t]})(\omega)|^2 \,\mathrm{d}t\, \mu(\mathrm{d}\omega) \\
&= \int_0^1 \int_\Omega |W((1-P_n)\mathbb{1}_{[0,t]})(\omega)|^2 \, \mu(\mathrm{d}\omega)\,\mathrm{d}t \\
&= \int_0^1 |(1-P_n)\mathbb{1}_{[0,t]}|^2_{L^2(0,1)} \,\mathrm{d}t
\end{aligned}$$

so (9.8) follows from the dominated convergence theorem.

It remains to compute the covariance operator $\mathcal{R}$. We have

$$\begin{aligned}
\langle \mathcal{R}h, h \rangle &= \mathbb{E}[\langle B, h \rangle^2] \\
&= \int_\Omega \mu(\mathrm{d}\omega) \left\{ \left(\int_0^1 W(\mathbb{1}_{[0,t]})(\omega) h(t)\,\mathrm{d}t \right) \left(\int_0^1 W(\mathbb{1}_{[0,s]})(\omega) h(s)\,\mathrm{d}s \right) \right\} \\
&= \int_0^1 \int_0^1 \mathrm{d}t\,\mathrm{d}s \left\{ h(t)h(s) \int_\Omega W(\mathbb{1}_{[0,t]})(\omega) W(\mathbb{1}_{[0,s]})(\omega)\, \mu(\mathrm{d}\omega) \right\} \\
&= \int_0^1 \int_0^1 \mathrm{d}t\,\mathrm{d}s\, \{h(t)h(s) \min\{s,t\}\}
\end{aligned}$$

and the conclusion follows.

□

Exercises

The next problem deal with the covariance operator of the Wiener measure. It gives a different taste of the relation between Brownian motion and differential operators.

Problem 9.6. Let $\mathcal{R}$ be the covariance operator of the Wiener measure defined in (9.7) and set $A = \mathcal{R}^{-1}$ be an unbounded operator on $L^2(0,1)$. Show that

$$\begin{cases} D(A) = \{h \in H^2(0,1) \,:\, h(0) = h'(1) = 0\} \\ Ah(t) = -h''(t) \quad \forall\, h \in D(A). \end{cases}$$

Now we extend the result to a different process. Let us recall the Brownian bridge, introduced in Exercise 9.2

$$X_t = B_t - tB_1, \qquad t \in [0,1].$$

Problem 9.7. Prove that the law of $\{X_t,\ t \in [0,1]\}$ in $L^2(0,1)$ is a Gaussian measure with zero mean and covariance operator $\mathcal{S}$ defined by

$$\mathcal{S}h(t) = \int_0^1 K(t,s)h(s)\,\mathrm{d}s, \qquad h \in L^2(0,1), \quad t \in [0,1],$$

where the kernel $K(t,s) : [0,1] \times [0,1] \to \mathbb{R}$ is given by

$$K(t,s) = \begin{cases} s(1-t) & 0 \le s \le t \\ t(1-s) & t \le s \le 1. \end{cases}$$

Problem 9.8. Referring to previous exercise, let $B = \mathcal{S}^{-1}$ be an unbounded operator on $L^2(0,1)$. Show that

$$\begin{cases} D(B) = H^2(0,1) \cap H^1_0(0,1) \\ Bh(t) = -h''(t) \quad \forall\, h \in D(B). \end{cases}$$

9.3 The derivative operator

In this section we define the stochastic derivative DF of a square integrable random variable $F : \Omega \to \mathbb{R}$, that means, we want to differentiate F with respect to ω, which we treat as a chance parameter, via a weak notion of derivative, and without assuming any topological structure on the space Ω (see Nualart [Nu95, section 1.2]).

Recall that the white noise W maps H into $\mathcal{H}_1 \subset L^2(\Omega)$; the Malliavin derivative will be an unbounded operator from $L^2(\Omega)$ into H which verifies

$$DW(g) = g, \qquad g \in H.$$

In particular, since $H = L^2(T)$, we can write the above identity pointwise

$$D_t W(g) = g(t), \qquad g \in H, \quad t \in T.$$

We denote by $C_b^\infty(\mathbb{R}^n)$ the set of all infinitely differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ such that f together with all its partial derivatives is bounded on $\mathbb{R}^n$.

Let $\mathcal{S}$ denote the class of smooth random variables, i.e., a random variable $F \in \mathcal{S}$ has the form

$$F = f(W(h_1), \dots, W(h_n)), \tag{9.9}$$

where $f \in C_b^\infty(\mathbb{R}^n)$ and $h_1, \dots, h_n \in H$. Note that the class $\mathcal{S}$ is dense in $L^2(\Omega)$. The derivative of such a random variable F is the stochastic process $\{D_t F,\, t \in T\}$ given by the formula

$$D_t F = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(W(h_1), \dots W(h_n)) h_i(t). \tag{9.10}$$

We consider DF as an element of the space $L^2(T \times \Omega) \equiv L^2(\Omega; H)$. Actually, $DF \in \bigcap_{p \geq 2} L^p(\Omega; H)$.

We fix an element $h \in H$. We introduce the directional derivative D^h on the set $\mathcal{S}$, defined by the formula

$$D^h F = \langle DF, h \rangle.$$

We note the relation

$$D^h F = \lim_{\varepsilon \to 0} \frac{f(W(h_1) + \varepsilon\langle h, h_1\rangle, \dots, W(h_n) + \varepsilon\langle h, h_n\rangle) - f(W(h_1), \dots, W(h_n))}{\varepsilon}.$$

We have the following important integration by parts formula.

Lemma 9.6. *Suppose that $F \in \mathcal{S}$ is a smooth functional and $h \in H$. Then*

$$\mathbb{E}[D^h F] = \mathbb{E}[F\, W(h)].$$

Proof. We may assume that there exists orthogonal elements in H $h = h_1, \dots, h_n$, with $|h_i| = 1$, $i = 1, \dots, n$, and a $f \in C_b^\infty(\mathbb{R}^n)$ such that $F = f(W(h_1), \dots, W(h_n))$. Then

$$\mathbb{E}[\langle DF, h\rangle] = \mathbb{E}[\partial_{x_1} f(W(h_1), \dots, W(h_n))].$$

Let $\mu_n \sim \mathcal{N}(0, I_n)$ be the law of $(W(h_1), \dots, W(h_n))$; then an integration by parts in $\mathbb{R}^n$ leads to

$$\mathbb{E}[\langle DF, h\rangle] = \int_{\mathbb{R}^n} \partial_{x_1} f(x)\, \mu_n(\mathrm{d}x) = \int_{\mathbb{R}^n} f(x) x_1\, \mu_n(\mathrm{d}x) = \mathbb{E}[F W(h_1)]$$

and from this we get the thesis.

□

From this lemma, taking into account that $D(FG) = (DF)G + F(DG)$ we obtain

$$\mathbb{E}[F\langle DG, h\rangle] + \mathbb{E}[G\langle DF, h\rangle] = \mathbb{E}[FGW(h)]. \tag{9.11}$$

Applying the previous result, it is possible to prove

Proposition 9.7. *The derivative operator D (with domain $\mathcal{S}$) is closable as an operator from $L^p(\Omega)$ to $L^p(\Omega; H)$, for any $p \geq 1$.*

Proof. We give the proof in case when $p > 1$. Let $\{F_n,\ n \in \mathbb{N}\} \subset \mathcal{S}$ be such that $F_n \to 0$ in $L^p(\Omega)$ and $DF_n \to G$ in $L^p(\Omega, H)$; our aim is to prove that $G \equiv 0$. Take $h \in H$ and $F = f(W(h_1), \dots, W(h_n))$, where $f \in C_b^\infty(\mathbb{R}^n)$. Then we obtain

$$\mathbb{E}[F\langle G, h\rangle] = \lim_{n\to\infty} \mathbb{E}[F\langle DF_n, h\rangle] = \lim_{n\to\infty} \mathbb{E}[F_n F W(h)] - \mathbb{E}[F_n\langle DF, h\rangle].$$

Now, since $F_n \to 0$ in L^p, and $W(h)$ belongs to each $L^p(\Omega)$, $p \geq 1$, we get that $\mathbb{E}[F\langle G, h\rangle] = 0$, which yields $G \equiv 0$ and the proof is complete.

□

We will denote the domain of D in $L^p(\Omega)$ by $\mathbb{D}^{1,p}$. This means that $\mathbb{D}^{1,p}$ is the closure of the class of smooth random variables $\mathcal{S}$ with respect to the norm

$$\|F\|_{1,p} = \left[\mathbb{E}(|F|^p) + \|DF\|^p_{L^p(\Omega;H)}\right]^{1/p}.$$

For $p = 2$, the space $\mathbb{D}^{1,2}$ is a Hilbert space with the scalar product

$$\langle F, G\rangle_{1,2} = \mathbb{E}(FG) + \mathbb{E}\langle DF, DG\rangle_H.$$

Remark 9.8. Of particular interest is the following result, that is proved in [Nu95]. Take $F \in \mathbb{D}^{1,2}$ and assume it is $\mathcal{F}_A$ measurable, for some $A \in \mathcal{B}(T)$. Then $D_t F$ is zero almost everywhere in $A^c \times \Omega$. In case $T = \mathbb{R}_+$, there above has the following interpretation: if F is $\mathcal{F}_s$-measurable, then its Malliavin derivative $D_t F$ is zero almost surely for $t > s$.

More generally, we can introduce iterated derivatives of (weakly differentiable) random variables. For a smooth random variable F and an integer k, we set

$$D^k_{t_1,\dots,t_k} F = D_{t_1} D_{t_2} \dots D_{t_k} F.$$

Note that the derivative $D^k F$ is a measurable function on the product space $T^k \times \Omega$. For every $p \geq 1$ and natural number $k \geq 1$ we denote by $\mathbb{D}^{k,p}$ the completition of the family $\mathcal{S}$ with respect to the norm

$$\|F\|_{k,p} = \left[\mathbb{E}(|F|^p) + \sum_{j=1}^k \mathbb{E}\|D^j F\|^p_{L^2(T^j)}\right]^{1/p}.$$

We end up this section with two important tools in applications. The first one is a chain rule for the Malliavin derivative operator. The proof of this result follows easily by approximating the random variable F by smooth random variables.

Proposition 9.9. *Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function with bounded partial derivatives and fix $p \geq 1$. Suppose that $F = (F_1, \dots, F_n)$ is a random vector whose components belong to the space $\mathbb{D}^{1,p}$. Then $\phi(F) \in \mathbb{D}^{1,p}$ and*

$$D(\phi(F)) = \sum_{i=1}^n \frac{\partial \phi}{\partial x_i}(F)\, DF_i. \tag{9.12}$$

The last result heuristically states that a random variable is “smooth” if we can approximate it with a sequence of “smooth” random variables.

Proposition 9.10 ([Nu95], Lemma 1.5.3). *Let* $\{F_n,\ n \geq 1\}$ *be a sequence of random variables in* $\mathbb{D}^{k,p}$ *with* $k \geq 1$ *and* $p > 1$. *Assume that* F_n *converges to* F *in* $L^p(\Omega)$ *and*

$$\sup_n \|F_n\|_{k,p} < \infty.$$

Then F *belongs to* $\mathbb{D}^{k,p}$.

10

Itô stochastic calculus

In this lecture we show how it is possible to "integrate" a sufficiently large class of stochastic processes $X = \{X_t,\, t \in T\}$ with respect to a fixed Brownian motion $\{B_t,\, t \geq 0\}$ (defined on a reference stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P})$). The process X will be always defined on the same space $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P})$. In other words, we will construct a stochastic integral with respect to the Brownian motion. This integral, due to `K. Itô`, will also allow to introduce in the sequel the stochastic differential equations.

The discussion of the properties of the Brownian paths shows that, in general, it is not possible to define the integral $\int_0^1 X_s \, \mathrm{d}B_s$ in the Lebesgue-Stjelties sense, arguing "ω by ω". If the paths of process X were of bounded variation on each bounded interval a.s., one may use an integration by parts to define

$$\int_0^1 X_s \, \mathrm{d}B_s = X_1 B_1 - \int_0^1 B_s \, \mathrm{d}X_s.$$

However, this way leaves aside a large number of interesting processes X, like the Brownian motion itself. A way to circumvent this difficulty, and to define a stochastic integral, was introduced in the 1940's by `K. Itô`. His idea was to introduce the Itô stochastic integral not "ω by ω" but instead in a "global way" as a limit in $L^2(\Omega)$ of suitable random variables defined in terms of stochastic integrals involving elementary processes.

Recall that the Wiener integral $f \mapsto W(f) = \int_0^1 f(s) \, \mathrm{d}B_s$ provides an isometry between the space $L^2(0,1)$ and the space of Gaussian random variables; Itô integral provides an isometry between the class $M^2(0,1)$ of square integrable, adapted processes and the space of square integrable and centered random variables.

10.1 The Itô integral

To simplify the notation, in the following we shall consider real valued processes defined on the time interval $[0,T]$. The extensions to other time intervals, like $\mathbb{R}_+$, $[a,b]$ or $t \in \mathbb{R}$ are rather straightforward. In the last part of the lecture we shall descrive the extension to the multidimensional case.

Definition 10.1. *$M^p(0,T)$, $p \geq 1$, is the class of real valued stochastic processes $X = \{X_t,\ t \in [0,T]\}$ which satisfy the following conditions*

i) *X is progressively measurable (with respect to $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \in [0,T]\}, \mathbb{P})$), and*

ii) $\mathbb{E}\Big[\int_0^T |X_s|^p \, \mathrm{d}s\Big] < +\infty$.

Among the elements of M^p we find in particular the *elementary processes*

$$X_t(\omega) = \sum_{i=0}^{n-1} \xi_i(\omega) \mathbb{1}_{[t_i, t_{i+1})}(\omega)$$

where $\xi_i \in L^p(\Omega)$ are random variables, $\mathcal{F}_{t_i}$-measurable, and $0 = t_0 < t_1 < \ldots < t_n = T$.

Recall that any adapted stochastic process with right-continuous trajectories is also progressively measurable.

Lemma 10.2. *Given a process $X \in M^2(0,T)$ there exists a sequence of elementary processes $\{X^{(n)}, n \in \mathbb{N}\}$ in $M^2(0,T)$ which approximate X, i.e.,*

$$\lim_{n \to +\infty} \mathbb{E} \int_0^T |X_t - X_t^{(n)}|^2 \, \mathrm{d}t = 0.$$

We begin with a general result on approximation with step functions, which we shall use often in the following.

Problem 10.1. Let us fix the time interval $[0,T]$, and denote by $[x]$ the integer part of the real number x. For arbitrary $h \in (0,T)$, we introduce the partition $\pi = \{t_0 = 0, t_1, \ldots, t_{n+1} = T\}$ on $[0,T]$, where $n = \left[\frac{T}{h}\right]$ and, for any $k = 1, \ldots, n$, $t_k = kh$.

Given $f \in L^2(0,T)$, we define the step functions

$$f_h(t) = \begin{cases} 0 & \text{if} \quad 0 \leq t < t_1 \\ f_h(t_k) := \frac{1}{h} \int_{(k-1)h}^{kh} f(s) \, \mathrm{d}s & \text{if} \quad t_k \leq t < t_{k+1}. \end{cases} \tag{10.1}$$

Then $f_h \to f$ in $L^2(0,T)$ as $h \to 0$ and moreover

$$\int_0^T |f_h(t)|^2 \, \mathrm{d}t \leq \int_0^T |f(t)|^2 \, \mathrm{d}t.$$

Proof (Lemma 10.2). Let $X \in M^2(0,T)$, define the family of step functions

$$X^{(n)}(t,\omega) = n \int_{(k-1)/n}^{k/n} X(s,\omega) \, \mathrm{d}s \qquad \text{for} \quad k/n \leq t < (k+1)/n,$$

for $\omega \in \Omega$ such that the integral is finite, and $X^{(n)}(t,\omega) = 0$ otherwise; notice however that, by assumption, the second case occurs with probability 0. Using Exercise 10.1 we get

$$\int_0^T |X_t^{(n)}|^2 \,\mathrm{d}t \le \int_0^T |X_t|^2 \,\mathrm{d}t$$

and

$$\int_0^T |X_t^{(n)} - X_t|^2 \,\mathrm{d}t \longrightarrow 0$$

for all trajectories $t \mapsto X(t,\omega)$ which are square integrables on $(0,T)$. However, since $X \in M^2$, the sequence $X^{(n)}$ is uniformly integrable, so that, using the Lebesgue theorem, we get

$$\mathbb{E} \int_0^T |X_t^{(n)} - X_t|^2 \,\mathrm{d}t \longrightarrow 0 \quad \text{as } n \to \infty.$$

□

10.1.1 Stochastic integral for elementary processes

Definition 10.3. *Given an elementary process* $X = \{X_t, t \in [0,T]\} \in M^2$

$$X_t(\omega) = \sum_{i=0}^{n-1} \xi_i(\omega) \mathbf{1}_{[t_i, t_{i+1})}(\omega),$$

we define the stochastic integral *of* X*, denoted by* $\int_0^T X_t \,\mathrm{d}B_t$*, as the random variable*

$$I_T = \sum_{i=0}^{n-1} \xi_i (B_{t_{i+1}} - B_{t_i}).$$

Lemma 10.4. *Given an elementary process* $X = \{X_t, t \in [0,T]\} \in M^2$*, the following properties hold:*

$$\mathbb{E} \int_0^T X_t \,\mathrm{d}B_t = 0 \tag{10.2}$$

$$\mathbb{E} \left[\int_0^T X_t \,\mathrm{d}B_t \right]^2 = \mathbb{E} \int_0^T X_t^2 \,\mathrm{d}t. \tag{10.3}$$

Proof. Since, for each $i = 1, \dots, n$, ξ_i is $\mathcal{F}_{t_i}$-measurable, the centered random variable $B_{t_{i+1}} - B_{t_i}$ is independent from ξ_i; this easily implies (10.2).

Moreover, we have

$$\mathbb{E} \left[\int_0^T X_t \,\mathrm{d}B_t \right]^2 = \mathbb{E} \left[\sum_{i=0}^{n-1} \xi_i (B_{t_{i+1}} - B_{t_i}) \right]^2$$
$$= \sum_{i=0}^{n-1} \mathbb{E}\Big[\xi_i^2 (B_{t_{i+1}} - B_{t_i})^2 \Big] + 2 \sum_{i<j} \mathbb{E}\Big[\xi_i \xi_j (B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j}) \Big].$$

Whenever $i < j$, since $t_i < t_j$, it holds that $(B_{t_{j+1}} - B_{t_j})$ is independent from $\xi_i \xi_j (B_{t_{i+1}} - B_{t_i})$ hence

$$\mathbb{E}[\xi_i \xi_j (B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})] = \mathbb{E}[\xi_i \xi_j (B_{t_{i+1}} - B_{t_i})]\mathbb{E}[B_{t_{j+1}} - B_{t_j}] = 0$$

which implies

$$\begin{aligned}\mathbb{E}\left[\int_0^T X_t \,\mathrm{d}B_t\right]^2 &= \sum_{i=0}^{n-1} \mathbb{E}[\xi_i^2 (B_{t_{i+1}} - B_{t_i})^2] = \sum_{i=0}^{n-1} \mathbb{E}(\xi_i^2)\mathbb{E}(B_{t_{i+1}} - B_{t_i})^2 \\ &= \sum_{i=0}^{n-1} \mathbb{E}(\xi_i^2)(t_{i+1} - t_i) = \mathbb{E}\sum_{i=0}^{n-1} \xi_i^2 (t_{i+1} - t_i) = \mathbb{E}\int_0^T X_t^2 \,\mathrm{d}t.\end{aligned}$$

□

10.1.2 The general case

We have defined the stochastic integral for elementary processes in the class $M^2(0,T)$; our aim is to extend this definition to the whole class. Recall that elementary processes are dense in $M^2(0,T)$, thanks to Lemma 10.2: for each $X \in M^2(0,T)$ there exists a sequence of elementary processes $\{X^{(n)}\}$, contained in $M^2(0,T)$, such that

$$\mathbb{E}\int_0^T |X_t^{(n)} - X_t|^2 \,\mathrm{d}t \longrightarrow 0 \tag{10.4}$$

as $n \to +\infty$.

We consider the sequence of square integrable random variables $I^{(n)} = \int_0^T X_t^{(n)} \,\mathrm{d}B_t$. In order to define the stochastic integral for the process X, it will be sufficient to show that the sequence $I^{(n)}$ is convergent in $L^2(\Omega)$: we shall define *stochastic integral* of X this limit.

Now, it is easy to see that the sequence $\{I^{(n)}\}$ is a Cauchy sequence. Actually, an elementary process remains unchanged if we add points to the original partition; hence, by considering the union of the partitions of $[0,T]$ which define respectively $X^{(n)}$ and $X^{(m)}$, it holds

$$\mathbb{E}[I^{(n)} - I^{(m)}]^2 = \mathbb{E}\left[\int_0^T (X_t^{(n)} - X_t^{(m)}) \,\mathrm{d}B_t\right]^2, \tag{10.5}$$

but the difference of elementary processes is again of such kind, hence we can apply Lemma 10.4 to get

$$\mathbb{E}[I^{(n)} - I^{(m)}]^2 = \mathbb{E}\int_0^T |X_t^{(n)} - X_t^{(m)}|^2 \,\mathrm{d}t.$$

But this implies that there exists the limit (in mean square sense) of $I^{(n)}$ as $n \to \infty$; we shall denote by $I = \int\limits_0^T X_t \,\mathrm{d}B_t$ this limit.

Further, we can prove that this limit is independent from the chosen approximating sequence $\{X^{(n)}\}$. To this aim, let $Y^{(n)}$ be a different sequence of elementary processes which approximate X

in the sense of (10.4). In order to prove that $\int_0^T Y_t^{(n)}\,\mathrm{d}B_t$ has the same limit as $\int_0^T X_t^{(n)}\,\mathrm{d}B_t$, for $n \to \infty$, we introduce one more sequence of elementary processes

$$Z_t^{(n)} = \begin{cases} X_t^{(n)} & \text{if } n \text{ is even} \\ Y_t^{(n)} & \text{if } n \text{ is odd;} \end{cases}$$

the sequence $\{Z^{(n)}\}$ again verifies (10.4), and the above reasoning implies the existence of the limit for the sequence of random variables $\int_0^T Z_t^{(n)}\,\mathrm{d}B_t$, which by construction must be equal to both the limit of $\int_0^T Y_t^{(n)}\,\mathrm{d}B_t$ and to the limit of $\int_0^T X_t^{(n)}\,\mathrm{d}B_t$, as required.

Problem 10.2. Verifies (10.5).

Theorem 10.5. *Let $X \in M^2(0,T)$. It is clear that the stochastic integral defined above is linear on $M^2(0,T)$; moreover we have*

$$\mathbb{E}\int_0^T X_t\,\mathrm{d}B_t = 0 \tag{10.6}$$

$$\mathbb{E}\left[\int_0^T X_t\ \mathrm{d}B_t\right]^2 = \mathbb{E}\int_0^T X_t^2\,\mathrm{d}t. \tag{10.7}$$

Proof. By definition, there exists a sequence of elementary processes $\{X^{(n)}, n \in \mathbb{N}\} \subset M^2$ such that

$$\lim_n \mathbb{E}\int_0^T |X_t^{(n)} - X_t|^2\,\mathrm{d}t = 0$$

and

$$\lim_n \mathbb{E}\left(\int_0^T (X_t^{(n)} - X_t)\,\mathrm{d}B_t\right)^2 = 0.$$

The assertion follows since (10.6) and (10.7) hold for elementary processes and then it is straightforward to pass to the limit.

□

Problem 10.3. Let $X = \{X_t\}$ and $Y = \{Y_t\} \in M^2(0,T)$. Show that

$$\mathbb{E}\left[\int_0^T X_s\,\mathrm{d}B_s \cdot \int_0^T Y_s\,\mathrm{d}B_s\right] = \mathbb{E}\int_0^T X_s Y_s\,\mathrm{d}s. \tag{10.8}$$

10.2 The stochastic integral process

In the following, we consider a stochastic process $X \in M^2(0,T)$ and, for any $t \in [0,T]$, we define the family of random variables

$$I_t = \int_0^t X_r \, \mathrm{d}B_r.$$

By construction, it follows easily that I_t is $\mathcal{F}_t$-measurable, hence $I = \{I_t, t \in [0,T]\}$ defines a stochastic process.

In the following, we shall study properties of the stochastic integral process $\{I_t,\ t \in [0,T]\}$. As in previous section, our study naturally divides in two steps: first we consider integrals of elementary processes, then we proceed to the general case by an approximation argument.

The proof of the following result is elementary and it is left as an exercise to the reader.

Problem 10.4. Let X be an elementary process in $M^2(0,T)$. Then the integral process $I = \{I_t, t \in [0,T]\}$ has continuous trajectories.

Proposition 10.6. *Let X be an elementary process in $M^2(0,T)$. Then the integral process $I = \{I_t, t \in [0,T]\}$ is a square integrable martingale (with respect to $\{\mathcal{F}_t\}$).*

The interest of this proposition (and of Theorem 10.9 below) is that it guarantees the possibility of using –to study integral processes– the powerful inequalities that martingale theory provides.

Proof. Recall that for an elementary process X in $M^2(0,T)$ it holds

$$I_t = \int_0^t X_r \, \mathrm{d}B_r \stackrel{\text{def}}{=} \sum_{i=0}^{n-1} \xi_i (B_{t_{i+1}} - B_{t_i}) + \xi_n (B_t - B_{t_n}).$$

For $s \le t$, we shall verify that $\mathbb{E}(X_t \mid \mathcal{F}_s) = X_s$. Take k be such that $t_k < s < t_{k+1}$: then

$$\int_0^t X_r \, \mathrm{d}B_r = \int_0^s X_r \, \mathrm{d}B_r + \xi_k (B_{t_{k+1}} - B_s) + \dots + \xi_n (B_t - B_{t_n}).$$

We take the conditional expectation with respect to $\mathcal{F}_s$ of the terms on the right-hand side; we recall that $\int\limits_0^s X_r \, \mathrm{d}B_r$ is $\mathcal{F}_s$-measurable so that property $b)$ in Proposition 1.14 implies

$$\mathbb{E}\left(\int_0^s X_r \, \mathrm{d}B_r \mid \mathcal{F}_s \right) = \int_0^s X_r \, \mathrm{d}B_r.$$

Next, we are concerned with the second term. We have that ξ_k is $\mathcal{F}_s$-measurable and the increments of the Brownian motion are independent from the past, hence

$$\mathbb{E}(\xi_k (B_{t_{k+1}} - B_s) \mid \mathcal{F}_s) = \xi_k \mathbb{E}(B_{t_{k+1}} - B_s) \mid \mathcal{F}_s) = \xi_k \mathbb{E}(B_{t_{k+1}} - B_s) = 0.$$

Since $t_i > s$ for $i \ge k+1$, it follows that $\mathcal{F}_s \subset \mathcal{F}_{t_i}$, so that property $d)$ in Proposition 1.14 implies

$$\mathbb{E}(\xi_i(B_{t_{i+1}} - B_{t_i}) \mid \mathcal{F}_s) = \mathbb{E}(\mathbb{E}(\xi_i(B_{t_{i+1}} - B_{t_i}) \mid \mathcal{F}_{t_i}) \mid \mathcal{F}_s) = 0.$$

We have thus proved the equality

$$\mathbb{E}(I_t \mid \mathcal{F}_s) = \mathbb{E}\left(\int_0^t X_r \, \mathrm{d}B_r \mid \mathcal{F}_s\right) = \int_0^s X_r \, \mathrm{d}B_r = I_s$$

which concludes the proof.

□

Remark 10.7. Let $X = \{X_t,\ t \in [0,T]\}$ be an elementary process in $M^2(0,T)$. We consider

$$I_t = \int_0^t X_s \, \mathrm{d}B_s, \qquad t \in [0,T].$$

We have just proved that $\{I_t\}$ is a martingale. Hence $Y_t = |I_t|^2$ is a submartingale and by the maximal inequality we have

$$\mathbb{P}\left(\sup_{t\in[0,T]} Y_t > \lambda\right) \le \frac{1}{\lambda}\mathbb{E}(|Y_T|).$$

Since

$$\mathbb{E}(|Y_T|) = \mathbb{E}\left|\int_0^T X_s \, \mathrm{d}B_s\right|^2 = \mathbb{E}\int_0^T |X_s|^2 \, \mathrm{d}s = \|X\|^2_{M^2(0,T)}$$

we obtain

$$\mathbb{P}\left(\sup_{t\in[0,T]}\left|\int_0^t X_s \, \mathrm{d}B_s\right| > \lambda\right) = \mathbb{P}(\sup_{t\in[0,T]}\left|\int_0^t X_s \, \mathrm{d}B_s\right|^2 > \lambda^2) \le \frac{1}{\lambda^2}\mathbb{E}\int_0^T |X_s|^2 \, \mathrm{d}s.$$

10.2.1 Continuity of trajectories

We shall now extend the regularity properties proved previously for elementary processes to the general case of $X \in M^2(0,T)$. We start with the continuity of trajectories.

We recall that a stochastic process $Y = \{Y_t,\, t \in [0,T]\}$ is called a version of a stochastic process $\{X_t,\, t \in [0,T]\}$, defined on the same stochastic basis of Y, if we have $X_t = Y_t$ a.s., for any $t \in [0,T]$.

Theorem 10.8. *Let X be a stochastic process in $M^2(0,T)$. Then the integral process $I = \{I_t,\, t \in [0,T]\}$, $I_t = \int_0^t X_r \, \mathrm{d}B_r$, has a version $\{J_t,\, t \in [0,T]\}$ with continuous trajectories.*

Proof. Consider a Cauchy sequence $\{X^{(n)}, n \in \mathbb{N}\}$ of elementary processes in $M^2(0,T)$, converging to X in mean square norm:

$$\mathbb{E}\int_0^T |X^{(n+p)} - X^{(n)}|^2 \, \mathrm{d}r \longrightarrow 0$$

as $n \to \infty$. We know that the approximating process $X^{(n)}$ defines a stochastic integral process $I^{(n)}$ with continuous paths, and that integral process I is defined as the L^2-limit of the processes $I^{(n)}$.

If we show that, for almost every ω

$$\sup_{p\in\mathbb{N}} \sup_{t\in[0,T]} \left| \int_0^t [X_r^{(n+p)} - X_r^{(n)}] \, \mathrm{d}B_r \right| \longrightarrow 0 \tag{10.9}$$

as $n \to \infty$, then the functions: $t \mapsto I_t^{(n)}(\omega)$ define a Cauchy sequence which is convergent in $C([0,T])$. Let us denote by $t \mapsto J_t(\omega)$ such limit in $C([0,T])$. Clearly, $\{J_t,\, t \in [0,T]\}$ is an adapted stochastic process and moreover we have $J_t = I_t$ a.s., for $t \in [0,T]$. It follows that $\{J_t\}$ is the version of $\{I_t\}$ we are looking for.

Let us prove (10.9). Thanks to Proposition 10.6, for each $p > 0$ the process $\{(I^{(n+p)} - I^{(n)})_t,\ t \in [0,T]\}$ is a martingale, and since $x \mapsto |x|^2$ is a convex function, then $\{|(I^{(n+p)} - I^{(n)})_t|^2,\ t \in [0,T]\}$ is a submartingale; we apply Doob's inequality (4.8) to this last process and, recalling the isometry (10.3), we obtain

$$\mathbb{P}(\sup_{t\le T} \left| \int_0^t (X_r^{(n+p)} - X_r^{(n)}) \, \mathrm{d}B_r \right| > \lambda) \le \frac{1}{\lambda^2} \sup_{m\in\mathbb{N}} \mathbb{E} \int_0^T (X_r^{(n+m)} - X_r^{(n)})^2 \, \mathrm{d}r \longrightarrow 0,$$

as $n \to \infty$, as shown in Section 10.1.2.

Let us choose two sequences $\{\lambda_k\}$ and $\{\varepsilon_k\}$, positive, converging to 0 and such that $\sum_{k\ge1} (\lambda_k + \varepsilon_k) < \infty$; we fix next a sequence n_k such that for every $p > 0$ it holds

$$\mathbb{P}\left(\sup_{t\le T} \left| \int_0^t (X_r^{(n_k+p)} - X_r^{(n_k)}) \, \mathrm{d}B_r \right| > \lambda_k \right) \le \varepsilon_k.$$

Then we can define the sequence of events

$$A_k = \left\{ \omega \mid \sup_{t\le T} \left| \int_0^t (X_r^{(n_{k+1})} - X_r^{(n_k)}) \, \mathrm{d}B_r \right| > \lambda_k \right\}.$$

Since $\{\varepsilon_k\}$ is a summable sequence, we have $\sum_k \mathbb{P}(A_k) < \sum_k \varepsilon_k < +\infty$; we apply Borel-Cantelli lemma 2.19 to get

$$\mathbb{P}\left(\sup_{t\le T} \left| \int_0^t (X_r^{(n_{k+1})} - X_r^{(n_k)}) \, \mathrm{d}B_r \right| > \lambda_k \quad \text{infinitely often} \right) = 0.$$

Therefore, for every ω in a set of probability 1, there exists a value $\kappa(\omega)$ such that, for every $k \ge \kappa(\omega)$, it holds

$$\sup_{t\le T} \left| \int_0^t (X_r^{(n_{k+1})} - X_r^{(n_k)}) \, \mathrm{d}B_r \right| \le \lambda_k.$$

Since the subsequence $X^{(n_k)}$ converges to X in L^2-norm, noticing that, for $p > 0$

$$X_t^{(n_{k+p})} - X_t^{(n_k)} = \sum_{i=k}^{k+p-1} (X_t^{(n_{i+1})} - X_t^{(n_i)}),$$

we get

$$\sup_{t\le T}\left|\int_0^t (X_r^{(n_{k+p})} - X_r^{(n_k)})\,\mathrm{d}B_r\right|$$
$$\le \sum_{i=k}^{k+p-1} \sup_{t\le T}\left|\int_0^t (X_r^{(n_{i+1})} - X_r^{(n_i)})\,\mathrm{d}B_r\right| \le \sum_{i=k}^{k+p-1}\lambda_i < \sum_{i=k}^{\infty}\lambda_i \longrightarrow 0,$$

as $k \to \infty$, and the thesis follows.

□

According to the previous result, we shall always deal with the continuous version of the integral process $\{I_t\}$.

10.2.2 Martingale property

Theorem 10.9. *Let X be a stochastic process in $M^2(0,T)$. Then the integral process $I = \{I_t, t \in [0,T]\}$ is a martingale.*

Proof. Once more, the proof will follow from an approximation procedure. Assume that $X^{(n)} \in M^2(0,T)$ is a sequence of elementary processes, converging to X in $L^2((0,T)\times\Omega)$; define $I^{(n)}$ the corresponding sequence of integral processes. We know that the process $\{I_t^{(n)},\ t \in (0,T)\}$ is a martingale, hence the identity $\mathbb{E}(I_t^{(n)} \mid \mathcal{F}_s) = I_s^{(n)}$, $t \ge s$, holds a.s.; it remains to prove that, if we take the limit as $n \to \infty$, this identity is maintained.

In particular, it will be sufficient to prove that

$$\mathbb{E}[\mathbb{E}(I_t - I_s \mid \mathcal{F}_s)]^2 = 0;$$

we know that this identity holds for elementary processes, thus, for $t \ge s$,

$$\begin{aligned}\mathbb{E}[\mathbb{E}(I_t - I_s \mid \mathcal{F}_s)]^2 &= \mathbb{E}[\mathbb{E}(I_t - I_s \mid \mathcal{F}_s) - \mathbb{E}(I_t^{(n)} - I_s^{(n)} \mid \mathcal{F}_s)]^2\\ &= \mathbb{E}[\mathbb{E}(I_t - I_t^{(n)} - I_s + I_s^{(n)} \mid \mathcal{F}_s)]^2\\ &= \mathbb{E}[\mathbb{E}(I_t - I_t^{(n)} \mid \mathcal{F}_s) - (I_s - I_s^{(n)})]^2\\ &\le 2\,\mathbb{E}[\mathbb{E}(I_t - I_t^{(n)} \mid \mathcal{F}_s)]^2 + \mathbb{E}[(I_s - I_s^{(n)})]^2.\end{aligned}$$

Now, using Jensen's inequality for conditional expectation ([1]) we obtain $\mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_s)^2) \le \mathbb{E}(X^2)$; therefore, since $I^{(n)}$ converges in L^2-norm to I, we obtain the thesis.

□

Having established the martingale property for the stochastic integral, we can use Doob's inequality (4.9), which implies the following result.

Corollary 10.10. *Let $\{X_t,\ t \in [0,T]\}$ be a stochastic process in $M^2(0,T)$. Then it holds*

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t X_s\,\mathrm{d}B_s\right|^2\right] \le 4\mathbb{E}\left[\int_0^T |X_s|^2\,\mathrm{d}s\right]. \tag{10.10}$$

[1] for every convex mapping Φ it holds $\Phi(\mathbb{E}(X \mid \mathcal{F})) \le \mathbb{E}(\Phi(X) \mid \mathcal{F})$

10.3 Itô processes

In this section, we slightly modify our notation, in order to conform it to the existing literature. Let $\sigma = \{\sigma_t,\ t \in [0,T]\}$ be a stochastic process in $M^2(0,T)$, and define the stochastic process $X = \{X_t,\ t \in [0,T]\}$

$$X_t = \int_0^t \sigma_s \, \mathrm{d}B_s.$$

Intuitively, we can think X_t as the description of a (stochastic) dynamical system, such that the infinitesimal variation $\mathrm{d}X_t$ is of order $\sigma_t \, \mathrm{d}B_t$. In other words, the variance of the process increases of $\sigma_t^2 \mathrm{d}t$, while the expected value remains constant. In literature, we often read that this differential is *of order* $\frac{1}{2}$.

It is natural, then, to add, to the description of the system, a differential *of order 1*. That is, we are given a stochastic process $b = \{b_t,\ t \in [0,T]\} \in M^1(0,T)$ such that the infinitesimal variation of the mean of X_t is $\mathbb{E}[b_t] \, \mathrm{d}t$. However, this choice does not affect the infinitesimal variation of the variance, as it is easily seen.

We are thus lead to the following definition

Definition 10.11. *Let $b = \{b_t,\ t \in [0,T]\} \in M^1$ and $\sigma = \{\sigma_t,\ t \in [0,T]\} \in M^2$ be given stochastic processes; further, let X_0 be a random variable $\mathcal{F}_0$-measurable (hence independent from B_t, for every $t \geq 0$).*

We call Itô process *the stochastic process $X = \{X_t,\ t \in [0,T]\}$ defined by*

$$X_t = X_0 + \int_0^t b_s \, \mathrm{d}s + \int_0^t \sigma_s \, \mathrm{d}B_s. \tag{10.11}$$

We shall also use the following differential form

$$\mathrm{d}X_t = b_t \, \mathrm{d}t + \sigma_t \, \mathrm{d}B_t.$$

which is by definition *equivalent to* (10.11). *We say that the stochastic differential of the Itô process $\{X_t\}$ is $b_t \, \mathrm{d}t + \sigma_t \, \mathrm{d}B_t$.*

Notice that X is a continuous process, since both the integrals in (10.11) are continuous in t.

Exercises

Problem 10.5. Let $\{X_t,\ t \in [0,T]\}$ be a stochastic process in $M^2(0,T)$. Then it holds

$$\mathbb{E}(\int_s^t X_u \, \mathrm{d}B_u \mid \mathcal{F}_s) = 0, \qquad \mathbb{E}(\left(\int_s^t X_u \, \mathrm{d}B_u\right)^2 \mid \mathcal{F}_s) = \mathbb{E}(\int_s^t X_u^2 \, \mathrm{d}u \mid \mathcal{F}_s).$$

Hint: one can start by proving the assertion for elementary processes X...

Before we introduce the Itô formula, it is instructive to compute the following elementary integrals.

Problem 10.6. Let $B = \{B_t,\ t \geq 0\}$ be a real standard Brownian motion. Compute the following integrals

1. $\int_0^T B_t \, \mathrm{d}B_t = \frac{1}{2}(B_T^2 - T)$. In particular, this shows that $\{B_t^2\}$ is an Itô process, with stochastic differential

$$\mathrm{d}(B^2)_t = \mathrm{d}t + 2B_t \, \mathrm{d}B_t. \tag{10.12}$$

2. In connection with the above formula, compute $\int_0^T r \, \mathrm{d}B_r$. This shows that

$$\mathrm{d}(tB_t) = B_t \, \mathrm{d}t + t \, \mathrm{d}B_t.$$

3. Compute $\mathbb{E}\left[B_s \int_s^t B_u \, \mathrm{d}B_u\right]$.
4. Compute $\mathbb{E}\left[\left(B_s \int_s^t B_u \, \mathrm{d}B_u\right)^2\right]$.

10.4 Itô fomula

Theorem 10.12. *Let $X = \{X_t,\, t \in [0,T]\}$ be an Itô process defined by the stochastic differential*

$$X_t = X_0 + \int_0^t b_r \, \mathrm{d}r + \int_0^t \sigma_r \, \mathrm{d}B_r$$

Let $\Phi : [0,T] \times \mathbb{R} \to \mathbb{R}$ be a continuous function in (t,x), differentiable with continuity one time in t e twice in x, having bounded derivatives.

Then $\Phi(t, X_t)$ is again an Itô process with stochastic differential

$$\Phi(t,X_t) - \Phi(0,X_0) = \int_0^t \partial_t \Phi(r,X_r) \, \mathrm{d}r + \int_0^t \partial_x \Phi(r,X_r) b_r \, \mathrm{d}r + \frac{1}{2} \int_s^t \partial_x^2 \Phi(r,X_r) \sigma_r^2 \, \mathrm{d}r + \int_s^t \partial_x \Phi(r,X_r) \sigma_r \, \mathrm{d}B_r. \tag{10.13}$$

Proof. First note that it is enough to show (10.13) when $\{b_t\}$ and $\{\sigma_t\}$ are elementary processes; the general case will then follow by an approximation argument (which is however necessary in order to define the stochastic integral in the formula). Further, notice that we can take b and σ to be constant processes in time, since the formula (10.13) can be proved separately on each interval of the partition defining the processes. Actually, there is a common partition such that b and σ are constant random variables in each subinterval of the partition; with no loss of generality, we can set this time interval equal to $[0,T]$.

Then the initial process has the form

$$X_t = X_0 + bt + \sigma B_t, \qquad t \in [0,T],$$

where X_0, b and σ are $\mathcal{F}_0$-measurable random variables, which means that they are independent from B_t, $t \geq 0$. Further, we shall deal with the process $Y_t = \Phi(t, X_t)$ given by

$$Y_t = \Phi(t, X_0 + bt + \sigma B_t), \qquad t \in [0,T].$$

We fix a time $t \in [0,T]$ and a partition $\pi = \{0 = t_0 < t_1 < \cdots < t_{n+1} = t\}$ with modulus

$$|\pi| = \max_{0 \le k \le n} |t_{k+1} - t_k|.$$

We have

$$Y_t - Y_0 = \sum_{k=0}^{n} \Phi(t_{k+1}, X(t_{k+1})) - \Phi(t_k, X(t_k)). \tag{10.14}$$

Applying Taylor's formula to Φ, we find that there exists numbers $0 \le d_k, d'_k \le 1$ such that

$$\begin{aligned}\Phi(t_{k+1},&X(t_{k+1})) - \Phi(t_k, X(t_k))\\ &= \partial_t \Phi(t_k + d_k(t_{k+1} - t_k), X(t_k))[t_{k+1} - t_k] + \partial_x \Phi(t_k, X(t_k))[X(t_{k+1}) - X(t_k)]\\ &\quad + \frac{1}{2}\partial^2_{xx}\Phi(t_k, X(t_k) + d'_k(X(t_{k+1}) - X(t_k)))[X(t_{k+1}) - X(t_k)]^2.\end{aligned}$$

Thanks to the continuity of the functions $\partial_t \Phi$, $\partial^2_{xx}\Phi$ and the continuity of the trajectories of $\{X_t\}$, we get, ω a.s.,

$$\max_{1 \le k \le n} \max_{0 < d_k < 1} |\partial_t \Phi(t_k + d_k(t_{k+1} - t_k), X(t_k)) - \partial_t \Phi(t_k, X(t_k))| \to 0,$$

as $|\pi| \to 0^+$ and

$$\max_{1 \le k \le n} \left|\partial^2_{xx}\Phi(t_k, X(t_k) + d'_k(X(t_{k+1}) - X(t_k))) - \partial^2_{xx}\Phi(t_k, X(t_k))\right| \to 0,$$

as $|\pi| \to 0^+$. If we set in the previous expressions $d_k = 0$, $d'_k = 0$, we introduce an error which becomes negligible in the limit as $|\pi| \to 0^+$. Write

$$\begin{aligned}Y_t - Y_0 &= \sum_{k=0}^{n} \Phi(t_{k+1}, X(t_{k+1})) - \Phi(t_k, X(t_k))\\ &= \sum_{k=0}^{n} \partial_t \Phi(t_k, X(t_k))[t_{k+1} - t_k] + \sum_{k=0}^{n} \partial_x \Phi(t_k, X(t_k))[X(t_{k+1}) - X(t_k)]\\ &\quad + \sum_{k=0}^{n} \frac{1}{2}\partial^2_{xx}\Phi(t_k, X(t_k))[X(t_{k+1}) - X(t_k)]^2\end{aligned}$$

Passing to the limit as $|\pi| \to 0^+$,

$$\sum_{k=0}^{n} \partial_t \Phi(t_k, X(t_k))[t_{k+1} - t_k] \longrightarrow \int_0^t \partial_s \Phi(s, X_s)\, \mathrm{d}s, \;\; \omega \; a.s.$$

and also

$$\sum_{k=0}^{n} \partial_x \Phi(t_k, X(t_k))[X(t_{k+1}) - X(t_k)]$$

$$\longrightarrow \int_0^t \partial_x \Phi(s, X_s) b \, \mathrm{d}s + \int_0^t \partial_x \Phi(s, X_s) \sigma \, \mathrm{d}B_s \qquad \text{in} \quad L^2(\Omega). \quad (10.15)$$

It remains to treat the last term. Notice that

$$[X(t_{k+1}) - X(t_k)]^2 = b^2(t_{k+1} - t_k)^2 + 2b\sigma(t_{k+1} - t_k)(B(t_{k+1}) - B(t_k)) + \sigma^2(B(t_{k+1}) - B(t_k))^2$$

then we substitute this formula to get

$$\frac{1}{2} \sum_{k=0}^{n} \partial^2_{xx} \Phi(t_k, X(t_k))[X(t_{k+1}) - X(t_k)]^2 = \frac{1}{2} b^2 \sum_{k=0}^{n} \partial^2_{xx} \Phi(t_k, X(t_k))(t_{k+1} - t_k)^2$$
$$+ b\sigma \sum_{k=0}^{n} \partial^2_{xx} \Phi(t_k, X(t_k))(t_{k+1} - t_k)(B(t_{k+1}) - B(t_k))$$
$$+ \frac{1}{2} \sigma^2 \sum_{k=0}^{n} \partial^2_{xx} \Phi(t_k, X(t_k))(B(t_{k+1}) - B(t_k))^2.$$

The continuity of $\partial^2_{xx} \Phi$ and the continuity of trajectories of $\{B_t\}$ imply that the first two sums converge to 0, ω a.s., as $|\pi| \to 0^+$. Finally, we are lead to prove that

$$\sum_{k=0}^{n} \partial^2_{xx} \Phi(t_k, X(t_k))(B(t_{k+1}) - B(t_k))^2 \longrightarrow \int_0^t \partial^2_{xx} \Phi(s, X_s) \, \mathrm{d}s.$$

Recall that we are assuming that $\partial^2_{xx} \Phi(s, x)$ is bounded in $[0, T] \times \mathbb{R}$. It is enough to show that, as $|\pi| \to 0^+$,

$$S_\pi := \sum_{k=0}^{n} \partial^2_{xx} \Phi(t_k, X(t_k))[(B(t_{k+1}) - B(t_k))^2 - (t_{k+1} - t_k)] \longrightarrow 0 \;\; in \; L^2(\Omega).$$

Let, for any $k = 1, \ldots, n$ $\epsilon_k = (B(t_{k+1}) - B(t_k))^2 - (t_{k+1} - t_k)$; they define a Gaussian family of independent random variables, centered and with variance

$$\mathbb{E}[\epsilon_k^2] = 2(t_{k+1} - t_k)^2.$$

Moreover, for each k, ϵ_k is independent from $\partial^2_{xx} \Phi(t_k, X(t_k))$. It follows that for a given sequence of partitions converging to zero: $|\pi| \to 0^+$, the random variables S_π have zero mean and converge to 0 in $L^2(\Omega)$, i.e.,

$$\mathbb{E}(S_\pi^2) \le 2 \|\partial^2_{xx} \Phi\|_\infty \sum_{k=0}^{n} (t_{k+1} - t_k)^2 \le 2t|\pi| \, \|\partial^2_{xx} \Phi\|_\infty \to 0. \quad (10.16)$$

Note that, eventually passing to a subsequence of partitions (π_n) of $[0, t]$ such that $|\pi_n| \to 0$ as $n \to \infty$, we may assume that the limits in (10.15) and (10.16) are limit $a.s.$ as $n \to \infty$. This completes the proof.

□

Remark 10.13. There is a simple way to remember and use Itô formula.
Let $\mathrm{d}X_t = b_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}B_t$; to compute $\mathrm{d}\Phi(t,X_t)$ we write Taylor's formula for $\Phi(t+\mathrm{d}t, X_t+\mathrm{d}X_t)$ in a neighborhood of (t,X_t) to get

$$\mathrm{d}\Phi(t,X_t) = \partial_t\Phi(t,X_t)\,\mathrm{d}t + \partial_x\Phi(t,X_t)\mathrm{d}X_t + \frac{1}{2}\partial^2_{xx}\Phi(t,X_t)\,\mathrm{d}(X^2)_t.$$

It remains to understand the term $\mathrm{d}(X^2)_t$: we have

$$\mathrm{d}(X^2)_t = b_t^2\,(\mathrm{d}t)^2 + \sigma_t^2\,(\mathrm{d}B_t)^2 + 2\sigma_t b_t\,\mathrm{d}t\,\mathrm{d}B_t.$$

Negletting infinitesimal terms of order greater than one: $(\mathrm{d}t)^2$ and $\mathrm{d}t\,\mathrm{d}B_t$, and substituting $(\mathrm{d}B_t)^2$ with $\mathrm{d}t$, we arrive at Itô formula

$$\mathrm{d}\Phi(t,X_t) = \Big[\partial_t\Phi(t,X_t) + b_t\partial_x\Phi(t,X_t) + \frac{1}{2}\sigma_t^2\partial_x^2\Phi(t,X_t)\Big]\,\mathrm{d}t + \sigma_t\partial_x\Phi(t,X_t)\,\mathrm{d}B_t.$$

10.5 Multidimensional Itô integral

Let $\{W_t,\ t\ge 0\}$ be a d-dimensional standard Brownian motion; we assume that W is defined on a probability space $(\Omega,\mathcal{F},\{\mathcal{F}_t\},\mathbb{P})$ equipped with a filtration $\{\mathcal{F}_t\}$ which satisfies the standard assumptions. We put $W_t = (W_t^1,\dots,W_t^d)$. The aim is to define

$$\int_0^t \Phi(s)\,\mathrm{d}W_s, \qquad t\ge 0,$$

where Φ is a stochastic processes taking values in the space $L(\mathbb{R}^d,\mathbb{R}^n)$ of linear operators from $\mathbb{R}^d$ into $\mathbb{R}^n$.

Definition 10.14. *We say that* $\Phi = \big(\phi_{ij}\big)$, $i=1,\dots,n$, $j=1,\dots,d$, *belongs to* $M^p(0,T;L(\mathbb{R}^d,\mathbb{R}^n))$ *if, for any* i *and* j, *the process* $\{\phi_{ij}(s),\ s\in[0,T]\}$ *is progressively measurable and*

$$\mathbb{E}\int_0^T |\phi_{ij}(s)|^p\,\mathrm{d}s < +\infty.$$

The construction can be carried over like in Section 10.1. For $\Phi\in M^2(0,T;L(\mathbb{R}^d,\mathbb{R}^n)$ we have that

$$\int_0^T \Phi(s)\,\mathrm{d}W(s) = \left(\sum_{j=1}^d \int_0^T \phi_{ij}(s)\,\mathrm{d}W_s^j\right)_i, \qquad i=1,\dots,n$$

is a n-dimensional random variable; using the one dimensional results, it is clear that the mapping

$$t\mapsto \int_0^t \Phi(s)\,\mathrm{d}W(s)$$

is a continuous martingale.

Problem 10.7. Prove that for real processes $X_1, X_2 \in M^2(0,T)$ and independent Brownian motions W^1 and W^2, it holds

$$\mathbb{E}\left[\left(\int_0^T X_1(s)\,\mathrm{d}W_s^1\right)\left(\int_0^T X_2(s)\,\mathrm{d}W_s^2\right)\right] = 0. \tag{10.17}$$

Hint: start with elementary processes then use density.

Lemma 10.15. *For $\Phi \in M^2(0,T;L(\mathbb{R}^d,\mathbb{R}^n))$ we have the following* isometry formula

$$\mathbb{E}\left|\int_0^t \Phi(s)\,\mathrm{d}W(s)\right|^2 = \mathbb{E}\int_0^t \|\Phi(s)\|_{HS}^2\,\mathrm{d}s, \tag{10.18}$$

where for any operator $\Phi \in L(\mathbb{R}^d,\mathbb{R}^n)$,

$$\|\Phi\|_{HS}^2 = \sum_{i=1}^m \sum_{j=1}^d |\phi_{ij}|^2.$$

Remark 10.16. We can also notice that the norm $\|\cdot\|_{HS}$ can be equivalently described as $\|\Phi\|_{HS}^2 = \mathrm{Tr}\,\Phi\Phi^*$ or $\|\Phi\|_{HS}^2 = \sum_{j=1}^d |\Phi\, e_j|^2$, where $\{e_1,\dots,e_d\}$ is any orthonormal basis of $\mathbb{R}^d$.

Proof. Let X_t the n-dimensional random variable

$$X_t = \int_0^t \Phi(s)\,\mathrm{d}W(s);$$

the idea is to use the covariance matrix Γ and in particular the relation

$$\mathbb{E}|X_t|^2 = \mathrm{Tr}\ \Gamma. \tag{10.19}$$

Thus the first step is to compute the covariance matrix:

$$\Gamma_{ij} = \mathbb{E}\left[\left(\int_0^t \sum_{k=1}^d \phi_{ik}(s)\,\mathrm{d}W_s^k\right)\left(\int_0^t \sum_{h=1}^d \phi_{jh}(s)\,\mathrm{d}W_s^h\right)\right]$$

but, using (10.17), all the elements outside the diagonal $h = k$ disappears, and it remains

$$\Gamma_{ij} = \sum_{k=1}^d \mathbb{E}\left[\int_0^t \phi_{ik}(s)\phi_{jk}(s)\,\mathrm{d}s\right].$$

Now, summing on the diagonal $i = j$, we get (10.18).
□

Next result provides the actual form of Doob's estimate for the multidimensional stochastic integral.

Lemma 10.17. *For $\Phi \in M^2(0,T;L(\mathbb{R}^d,\mathbb{R}^n)$ we have the following estimate*

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t \Phi(s)\,\mathrm{d}W(s)\right|^2\right] \le 4\mathbb{E}\int_0^T \|\Phi(s)\|_{HS}^2\,\mathrm{d}s. \tag{10.20}$$

Proof. We compute the norm and obtain

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t \Phi(s)\,\mathrm{d}W(s)\right|^2\right] = \mathbb{E}\left[\sup_{t\in[0,T]}\sum_{i=1}^n\left|\int_0^t\sum_{j=1}^d \phi_{ij}(s)\,\mathrm{d}W_s^j\right|^2\right]$$
$$= \sum_{i=1}^n \mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t\sum_{j=1}^d \phi_{ij}(s)\,\mathrm{d}W_s^j\right|^2\right]$$

Each term of the sum is bounded by Doob's theorem 4.24

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t\sum_{j=1}^d \phi_{ij}(s)\,\mathrm{d}W_s^j\right|^2\right] \le 4\mathbb{E}\left|\int_0^T\sum_{j=1}^d \phi_{ij}(s)\,\mathrm{d}W_s^j\right|^2 = 4\mathbb{E}\int_0^T\sum_{j=1}^d |\phi_{ij}(s)|^2\,\mathrm{d}s$$

Putting together this estimates we obtain

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t \Phi(s)\,\mathrm{d}W(s)\right|^2\right] \le 4\mathbb{E}\left[\sum_{i=1}^n\sum_{j=1}^d\int_0^T |\phi_{ij}(s)|^2\,\mathrm{d}s\right].$$

□

10.5.1 Multidimensional Itô's formula

Assume we are given a n-dimensional Itô stochastic process, defined by the Itô differential

$$\mathrm{d}X_t = b(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W_t \tag{10.21}$$

where $\{b(t),\ t \in [0,T]\}$ is a n-dimensional process which belongs to $M^1(0,T;\mathbb{R}^n)$, $\{\sigma(t),\ t \in [0,T]\}$ belongs to $M^2(0,T;L(\mathbb{R}^d,\mathbb{R}^n))$ and W is a d-dimensional Brownian motion. As in the one-dimensional case, identity (10.21) means that

$$X_t = X_0 + \int_0^t b(r)\,\mathrm{d}r + \int_0^t \sigma(r)\,\mathrm{d}W_r, \quad t \in [0,T].$$

The following result extends Itô theorem 10.12 to the case of multidimensional stochastic processes. We shall not give the proof, as it is a direct, lenghty extension of the one-dimensional case.

Theorem 10.18. *Consider a continuous and bounded function $u(t,x)$ defined on $[0,T]\times\mathbb{R}^n$ and taking values in $\mathbb{R}$, such that the partial derivatives $\partial_t u = \frac{\partial}{\partial t}u$, $\partial_{x_i}u = \frac{\partial}{\partial x_i}u$ and $\partial^2_{x_ix_j}u = \frac{\partial^2}{\partial x_i x_j}u$ are continuous and bounded functions su $[0,T]\times\mathbb{R}^n$.*

Then if $\{X_t,\ t\in[0,T]\}$ is a n-dimensional stochastic process, defined by the Itô differential (10.21), the process

$$Y_t = u(t,X_t), \qquad t\in[0,T],$$

Y is a one-dimensional stochastic process with stochastic differential

$$\mathrm{d}Y_t = \Bigg(\partial_t u(t,X_t) + \sum_{i=1}^n b_i(t)\partial_{x_i}u(t,X_t) + \frac{1}{2}\sum_{i,j=1}^n\sum_{l=1}^d \sigma_{il}(t)\sigma_{jl}(t)\partial^2_{x_ix_j}u(t,X_t)\Bigg)\,\mathrm{d}t + \sum_{i=1}^n\sum_{l=1}^d \sigma_{il}u_{x_i}(t,X_t)\,\mathrm{d}W^l_t. \quad (10.22)$$

Remark 10.19. Define the gradient

$$\nabla u = ((\partial_{x_1}u,\dots,\partial_{x_n}u)$$

and the Hessian matrix

$$\nabla^2 u = \left(\partial^2_{x_ix_j}u\right)_{i,j=1,\dots,n}.$$

Then using also the inner product $\langle\cdot,\cdot\rangle$ in $\mathbb{R}^n$, we can write Itô's formula (10.22) in the form

$$\mathrm{d}Y_t = \left(\partial_t u(t,X_t) + \langle b(t),\nabla u(t,X_t)\rangle + \frac{1}{2}\operatorname{Tr}[\sigma(t)\sigma^*(t)(\nabla^2 u)(t,X_t)]\right)\mathrm{d}t + \langle\nabla u(t,X_t),\sigma(t)\,\mathrm{d}W_t\rangle.$$

Remark 10.20. The case $u:[0,T]\times\mathbb{R}^n\to\mathbb{R}^k$, which implies that $Y_t = u(t,X_t)$ is a k-dimensional vector, can be handled easily by the application of formula (10.22) on every component.

An interesting application of the previous theorem concernes the case $u(x,y)=xy$. Suppose that X and Y are real Itô processes

$$\mathrm{d}X_t = b_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}B_t, \qquad \mathrm{d}Y_t = c_t\,\mathrm{d}t + \theta_t\,\mathrm{d}B_t$$

defined on a stochastic basis on which it is defined a real Brownian motion $\{B_t\}$. Assume for simplicity that the real processes $b=\{b_t,\ t\in[0,T]\}$, $c=\{c_t,\ t\in[0,T]\}$, $\sigma=\{\sigma_t,\ t\in[0,T]\}$ and $\theta=\{\theta_t,\ t\in[0,T]\}$ belong to $L^\infty((0,T)\times\Omega)$. Applying Itô formula to u ([2]), we have:

$$\begin{aligned}\mathrm{d}(X_tY_t) &= X_t\,\mathrm{d}Y_t + Y_t\,\mathrm{d}X_t + \sigma_t\theta_t\,\mathrm{d}t\\ &= (X_tc_t + Y_tb_t + \sigma_t\theta_t)\,\mathrm{d}t + (X_t\theta_t + Y_t\sigma_t)\,\mathrm{d}B_t.\end{aligned}$$

This formula is an *integration by parts* rule for stochastic calculus. In integral form it is

[2] to be precise, we should first apply Itô formula to the functions $u_\epsilon(x,y) = xye^{-\epsilon(x^2+y^2)}$, for every $\epsilon>0$, and then pass to the limit as $\epsilon\to 0^+$

$$X(t)Y(t) = X(0)Y(0) + \int_0^t X(s)\,\mathrm{d}Y(s) + \int_0^t Y(s)\,\mathrm{d}X(s) + \int_0^t \sigma_s\theta_s\,\mathrm{d}s.$$

Comparing with the similar formula for deterministic integrals, we notice the additional (Itô) term $\sigma_t\theta_t\,\mathrm{d}t$.

11

Stochastic differential equations with additive noise

Let us start with a *simplified physical model* in which stochastic differential equations appear in a natural way (see also Flandoli [Fl93] and Klebaner [Kl99]).

We consider a microscopic particle moving in a bidimensional fluid. For simplicity, we only deal with one component of the position vector of the particle.

If $X(t)$ denotes the position of the particle at time t, its shift between t and $t+\triangle t$ will be indicated by $\triangle X = X(t+\triangle t) - X(t)$. We can express $\triangle X$ in the following way:

$$X(t+\triangle t) - x = b(t, X(t))\triangle t + \sigma(t, X(t))(B(t+\triangle t) - B(t)),$$

where b and σ are (regular) functions and $B(t)$ denotes a Brownian motion (or more precisely, a trajectory of a Brownian motion).

Recalling the physical interpretation of Brownian motion (as elaborated by Einstein e Smoluchowski at the beginning of 1900) we know that the term $\sigma(t,x)(B(t+\triangle t) - B(t))$ indicates the random movement of the particle in the time interval $(t, t+\triangle t)$ due to the collisions with the molecules of the fluid ($\sigma(t,x)$ measures the effect of temperature of fluid in point x at time t). On the other hand, $b(t,x)\triangle t$ measures the shift of the particle due to the motion of the fluid, whenever we indicate with $b(t,x)$ the velocity of fluid in point x at time t.

We say that b is the *drift coefficient* and σ the *diffusion coefficient*. Passing to the limit as $\triangle t \to 0^+$, we find (*formally*) the equation

$$\mathrm{d}X(t) = b(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}B_t, \qquad t > 0, \tag{11.1}$$

usually endowed with an initial condition $X_0 = z \in \mathbb{R}$. It is not clear how to give a meaning to equation (11.1). Indeed, we can not simply divide by dt and argue "ω by ω" since the "white noise" $\frac{\mathrm{d}B}{\mathrm{d}t}$ is not defined in that sense (remember that trajectories of $B(t)$ are not differentiable, a.s.).

To give a meaning to expression (11.1), Itô proposed to consider it in an integral sense, i.e., to treat

$$X_t = z + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s, \quad t \geq 0. \tag{11.2}$$

As already mentioned, Itô gave a precise meaning to the stochastic integral $\int_0^t \sigma(s, X_s)\,\mathrm{d}B_s$ as a limit in $L^2(\Omega)$ of "elementary stochastic integrals". Nowadays, stochastic differential equations are an important part of probability theory. Examples and applications can be found in many books: for instance, we refer to [KS88], [IW81], [EK86], [SV79], [Kr95].

In this lecture we will treat *stochastic differential equations with additive noise*, i.e., σ does not depend on t and x. Hence we are concerned with equation like

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t.$$

The point here is the choice of σ independent from X_t; the fact that b does not depend on t is not a crucial simplification, but it is done only to simplify the statements and notation.

To prove existence and uniqueness for stochastic equations with additive noise the Itô integral is not needed. However, stochastic equations with additive noise give a useful introduction to the more general equations (11.2), which are called *stochastic equations with multiplicative noise*. As mentioned in Lecture 6, solutions to stochastic differential equations are an important class of diffusion processes; we shall return on this aspect in the following lectures.

11.1 Preliminaries

In this section we shall consider stochastic differential equations with *additive noise*, i.e., equations of the following kind

$$\begin{cases} \mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t, & t \in [s,T], \\ X_s = x \in \mathbb{R}^n, \quad s \geq 0, \end{cases} \tag{11.3}$$

where $\{B_t,\ t \geq 0\}$ is a fixed d-dimensional Brownian motion, defined on a *stochastic basis* $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \geq 0\}, \mathbb{P})$, satisfying the *usual assumptions*. Recall that, in particular, $\{\mathcal{F}_t\}$ can be the natural completed filtration generated by the process $\{B_t\}$.

Here σ is a given real $n \times d$ matrix and $b : \mathbb{R}^n \to \mathbb{R}^n$ is a given Borel measurable function.

Definition 11.1. *A continuous process $\{X_t,\ t \in [s,T]\}$, defined and adapted on $(\Omega, \mathcal{F}, \{\mathcal{F}_t,\ t \in [s,T]\}, \mathbb{P})$ and taking values in $\mathbb{R}^n$, is a* (pathwise) solution *for equation (11.3) if for all $t \geq s \geq 0$ it holds, almost surely,*

$$X_t - x = \int_s^t b(X_r)\,\mathrm{d}r + \sigma\,(B_t - B_s), \qquad t \in [s,T]. \tag{11.4}$$

Sometimes, we write $X(t;s,x)$ or $X_t^{s,x}$ to denote explicitly the dependence of the solution on the initial conditions s and x.

Notice that the trajectories of the solution $\{X_t\}$ have the same regularity of those of Brownian motion. In particular these trajectories are not differentiable almost surely.

We start by mentioning a basic result concerning the case of (globally) Lipschitz continuous coefficients b.

Theorem 11.2. *Let $b : \mathbb{R}^n \to \mathbb{R}^n$ satisfy a Lipschitz condition*

$$|b(x_0) - b(x_1)| \leq L|x_0 - x_1|, \qquad \forall x_0,\ x_1 \in \mathbb{R}^n. \tag{11.5}$$

Then, for each initial condition $x \in \mathbb{R}^n$, there exists a unique solution $\{X_t\}$ of problem (11.3).

We shall not go, for the moment, in the proof of the theorem. In fact, this result follows directly as a corollary from Theorem 11.5 below. Moreover the previous result will be also generalized in the following lectures when we shall consider stochastic differential equations (SDEs for short) with multiplicative noise and Lipschitz continuous coefficients.

11.2 Existence and uniqueness under monotonicity conditions

Since, with probability 1, sample paths of the Brownian motion are continuous functions, we may consider equation (11.3) in a pathwise sense by searching a solution "ω by ω"; thus, equation (11.3) is transformed in a family of ordinary differential equations depending on the parameter ω. As we shall discuss, this approach has a drawback, namely that it will be necessary to prove adaptedness of the solution process. Moreover this approach of arguing "ω by ω" can not be used for general stochastic differential equations with multiplicative noise, where Itô integrals (as opposite to Wiener integrals) are in charge of being used.

We present now a result of existence and uniqueness for the solutions of (11.3) under the following assumptions.

Hypothesis 11.3. The mapping b satisfies the following assumptions

(i) $b \in C^1(\mathbb{R}^n, \mathbb{R}^n)$.
(ii) There exists a constant $K > 0$ such that

$$\langle b(x+y) - b(y), x \rangle \leq K(1 + |x|^2), \qquad x, y \in \mathbb{R}^n, \tag{11.6}$$

where $\langle \cdot, \cdot \rangle$ and $| \cdot |$ denote the scalar product and the euclidean norm in $\mathbb{R}^n$.

Let us discuss the above assumptions.

Remark 11.4. (a) Condition (i) can be generalized, by requiring that the mapping b is locally Lipschitz. Also, condition (ii) is a *monotonicity condition*, which is used in the theory of ordinary differential equations in order to assure the existence of a maximal solution for all times.

(b) If condition (i) holds, and, in addition, b is a Lipschitz continuous mapping on $\mathbb{R}^n$ (hence $L = \sup_{x \in \mathbb{R}^n} |Db(x)| < \infty$, where Db stands for the Jacobian matrix of b) then also (ii) is satisfied, with $K = L$. Actually, it holds

$$\langle b(x+y) - b(y), x \rangle \leq L|x|^2, \qquad x, y \in \mathbb{R}^n.$$

Hence the assumptions of Theorem 11.2 are a special case of Hypothesis 11.3.

(c) In dimension $n = 1$, if b is of class $C^1(\mathbb{R})$ then condition (ii) holds provided that $\sup\limits_{x \in \mathbb{R}} b'(x) = k < \infty$. As an example, consider the equation

$$\mathrm{d}X_t = (aX - X^3)\mathrm{d}t + \mathrm{d}B_t, \qquad X_0 = x \in \mathbb{R}, \quad t \geq 0,$$

for some constant $a \in \mathbb{R}$.

Theorem 11.5. *Under Hypothesis 11.3, for every $x \in \mathbb{R}^n$, there exists a unique solution $\{X_t^{s,x},\ t \ge s\}$ of equation (11.3).*

Further, for every $T > s$, it holds

$$\sup_{t\in[s,T]} \mathbb{E}|X_t^{s,x}|^2 = C_T(x) < \infty. \tag{11.7}$$

To prove the result we shall consider the integral equation (11.4) as a deterministic equation with a continuous pertubration $B_t(\omega)$, for fixed ω; in this way, we can rely on some analytic tools, that we recall below. On the other hand, this approach requires some extra work in order to prove that the solution is $\mathcal{F}_t$-adapted.

The proof of the uniqueness for the solution relies on the following well known result.

Lemma 11.6 (Gronwall's Lemma). *Consider a continuous function $v : [s,T] \to \mathbb{R}$ and constants $c,\ d \ge 0$. If, for every $t \in [s,T]$, it holds*

$$v(t) \le c + d \int_s^t v(s)\mathrm{d}s$$

then

$$v(t) \le ce^{d(T-s)}, \qquad t \in [s,T].$$

Further, next lemma will be useful in the proof of existence of the solution.

Lemma 11.7. *Let $f : [s,\infty) \to \mathbb{R}^n$ be a continuous function, $s \in \mathbb{R}$. Suppose that $b \in C^1(\mathbb{R}^n,\mathbb{R}^n)$ verifies condition* (11.6) *and consider the integral equation below in the space of continuous functions from $[s,\infty)$ into $\mathbb{R}^n$:*

$$u(t) = f(t) + \int_s^t b(u(r))\,\mathrm{d}r, \qquad t \ge s. \tag{11.8}$$

(1) There exists a unique solution u for problem (11.8) *on $[s,\infty)$;*
(2) assume further that b is a Lipschitz function on $\mathbb{R}^n$. Then the following Picard's iteration scheme

$$u_0(t) = f(t),\ \ u_{n+1}(t) = f(t) + \int_s^t b(u_n(r))\,\mathrm{d}r, \qquad t \ge s,\ n \in \mathbb{N},$$

converges to the solution u, uniformly on every bounded interval $[s,T]$, $T > s$.

Proof. We prove (1). Setting $v(t) := u(t) - f(t)$, we search for a C^1-function v which verifies

$$v(t) = \int_s^t b(v(r) + f(r))\,\mathrm{d}r, \qquad t \ge s. \tag{11.9}$$

We write the above equation as a Cauchy problem

$$\begin{cases} \dot v(t) = b(v(t) + f(t)), \\ v(s) = 0. \end{cases}$$

Since $F(t,x) := b(x + f(t))$ is a C^1-function in the x variable, it is known that there exists a unique solution v for the Cauchy problem defined on a maximal interval $[s,T[$. Hence v verifies (11.9) on $[s,T[$. It is also clear that

$$u(t) := v(t) + f(t), \qquad t \in [s,T[,$$

is the unique solution of (11.8) in $[s,T[$. It remains to prove, using (11.6), that $T = \infty$. Taking the scalar product with v we have

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}|v(t)|^2 =& \langle \dot v(t), v(t)\rangle = \langle b(v(t)+f(t)), v(t)\rangle \\ =& \langle b(v(t)+f(t)) - b(f(t)), v(t)\rangle + \langle b(f(t)), v(t)\rangle \\ \le& K(1+|v(t)|^2) + M|v(t)| \le C(1+|v(t)|^2), \qquad t \in [s,T[,\end{aligned}$$

where $M = \sup_{t\in[s,T]} |b(f(t))|$. Integrating in time on the interval $[s,t]$ we get

$$|v(t)|^2 \le 2C(t-s) + 2C\int_s^t |v(r)|^2 \,\mathrm{d}r$$

and Gronwall's Lemma implies that $|v(t)|^2 \le 2C(T-s)\, e^{2C(T-s)}$, for $t \in [s,T[$. Therefore $\sup_{t\in[s,T[} |\dot v(t)| = R_T < \infty$ and $v(t)$ is a Lipschitz function on $[s,T[$. It follows that we can continuously extend v to the closed interval $[s,T]$, and so $T = +\infty$.

Consider next the second part of the thesis. Take $T > s$. It holds, for $t \in [s,T]$,

$$\begin{aligned}|u_1(t) - u_0(t)| &\le \int_s^t |b(f(r))| \,\mathrm{d}r \le B(t-s), \\ |u_2(t) - u_1(t)| &\le \int_s^t |b(u_1(r)) - b(f(r))| \,\mathrm{d}r \le BL\int_s^t (r-s)\,\mathrm{d}r = B\,\frac{L(t-s)^2}{2},\end{aligned}$$

where $B = \sup_{t\in[s,T]} |b(f(t))|$. By induction we obtain

$$|u_{n+1}(t) - u_n(t)| \le B\,\frac{L^n(T-s)^{n+1}}{(n+1)!}, \qquad t \in [s,T], \quad n \in \mathbb{N}.$$

Thus $u_{n+1}(t) = f(t) + \sum_{k=0}^{n} (u_{k+1}(t) - u_k(t))$ converges uniformly on $[s,T]$ to a continuous function u. Taking the limit for $n \to \infty$ in the equation

$$u_{n+1}(t) = f(t) + \int_s^t b(u_n(r))\,\mathrm{d}r$$

we get that u is the solution of problem (11.8) for all $t \ge s$.

□

Proof of Theorem 11.5.

For simplicity of notation, we only give the proof when $s = 0$.

(Uniqueness). Assume that there exists two solutions $\{X_t\}$ and $\{Y_t\}$. For almost every ω, for every $T > 0$, we have

$$|X_t(\omega) - Y_t(\omega)| \le \int_0^t |b(X_s(\omega)) - b(Y_s(\omega))| \, \mathrm{d}s \le c \int_0^t |X_s(\omega) - Y_s(\omega)| \, \mathrm{d}s,$$

for $t \in [0, T]$, where c is the Lipschitz constant of the mapping b on the closed ball $\overline{B(0, R)}$ of radius $R = R(\omega) = \max\limits_{t \in [0,T]} (|X_t(\omega)| + |Y_t(\omega)|)$. From Gronwall's Lemma we deduce that $X_t(\omega) = Y_t(\omega)$, $t \in [0, T]$, and the uniqueness follows from the arbitrariety of T and of ω.

In the same way we can prove that uniqueness holds on any stochastic interval $[0, T(\omega)]$. In other words, assume that $\{X_t\}$ and $\{Y_t\}$ are solutions to (11.4) on $[0, T(\omega)]$, then $X_t(\omega) = Y_t(\omega)$, $t \in [0, T(\omega)]$, almost surely.

(Existence). At first, we shall prove that for almost every ω there exists a solution $\{X_t(\omega),\ t \ge 0\}$ for (11.4); in order to simplify the notation, we shall not state the dependence on the initial condition x.

The existence of a pathwise solution follows by using Lemma 11.7 with

$$f(t) = x + \sigma B_t(\omega)$$

for almost every ω.

We shall prove now that $\{X_t\}$ is adapted to the filtration $\{\mathcal{F}_t\}$, i.e., that $X_t : (\Omega, \mathcal{F}_t) \to \mathbb{R}^n$ is a measurable mapping for each $t > 0$.

This is certainly the case if b is a Lipschitz continuous function on $\mathbb{R}^n$. Indeed, in that case we have that $X_t(\omega)$ is – almost surely in ω – the limit of the Picard approximation processes $\{X_n(t),\ t \ge 0\}$

$$\begin{cases} X_0(t) = x + \sigma B_t, \\ X_{n+1}(t) = x + \sigma B_t + \int_0^t b(X_n(s)) \, \mathrm{d}s \quad t \ge 0,\ n \in \mathbb{N}, \end{cases}$$

the limit being uniform in bounded intervals $[0, T]$: see part (2) of Lemma 11.7. It is easily seen that the approximation processes $\{X_n(t),\ t \ge 0\}$ are $\mathcal{F}_t$-adapted, and so the above stated convergence implies that also the limit process $\{X_t,\ t \ge 0\}$ is $\mathcal{F}_t$-adapted.

Assume next that b is just a locally Lipschitz mapping satisfying the assumptions. We introduce some Lipschitz continuous approximations $\{b_k\}$ for b:

$$b_k(x) = \begin{cases} b(x), & \text{if } |x| \le k, \\ b(k \frac{x}{|x|}), & \text{if } |x| > k. \end{cases}$$

For every k, the mapping b_k is bounded and Lipschitz continuous, hence there exists a unique (global) solution for the stochastic differential equation

$$\begin{cases} \mathrm{d}X_t^k = b_k(X_t^k) \, \mathrm{d}t + \sigma \, \mathrm{d}B_t, \\ X_0 = x \in \mathbb{R}^n,\ t \ge 0,\ k \in \mathbb{N}. \end{cases}$$

From the above discussion we have that the solution $\{X^k(t),\ t \geq 0\}$ is $\mathcal{F}_t$-adapted.

The proof will be complete once we prove that the sequence $X_t^k = X_k(t)$ converges to X_t almost surely, for every $t \geq 0$, i.e.,

$$\lim_{k\to\infty} X_t^k(\omega) = X_t(\omega). \tag{11.10}$$

The above convergence can be stated as follows: for given $t_0 \geq 0$ and for almost every ω, for every $\varepsilon > 0$ there exists $k_0 = k_0(\omega)$ such that for all $k \geq k_0$, $|X_{t_0}^k(\omega) - X_{t_0}(\omega)| < \varepsilon$.

We check (11.10) for every trajectory ω such that $t \mapsto X_t(\omega)$ is a continuous mapping (we already know that this happens with probability 1). Let us consider the number $N = N(\omega) > \sup_{t\in[0,t_0]} |X_t(\omega)|$. By construction, it holds

$$X_t(\omega) - x = \int_0^t b(X_s(\omega))\,\mathrm{d}s + \sigma B_t(\omega) = \int_0^t b_N(X_s(\omega))\,\mathrm{d}s + \sigma B_t(\omega), \quad t \in [0, t_0].$$

Then, the uniqueness of the solution implies that $X_t(\omega) = X_t^N(\omega)$ for all $t \in [0, t_0]$, but also $X_t(\omega) = X_t^k(\omega)$, $t \in [0, t_0]$, for every $k \geq N$. Hence the convergence in (11.10) holds. Actually, we have proved even more, i.e., the convergence is uniform in time on every bounded interval $[0, t_0]$, almost surely:

$$\lim_{k\to\infty} \sup_{t\in[0,t_0]} |X_t^k(\omega) - X_t(\omega)| = 0. \tag{11.11}$$

(Estimate (11.7)*)* The last part of the proof is concerned with the estimate (11.7). The key ingredient in the proof will be Itô formula, applied to the approximating processes $\{X_t^k\}$.

At first, one shall control that these processes belong to $M^2(0, T; \mathbb{R}^n)$ for arbitrary $T > 0$; Since b_k is a bounded Lipschitz continuous mapping, we obtain, for any $k \geq 1$, a.s.,

$$|X_t^k|^2 \leq 3|x|^2 + 3t^2\|b_k\|_\infty^2 + 3\|\sigma\|^2\,|B_t|^2$$

and so

$$\mathbb{E}|X_t^k|^2 \leq C_{T,k}(1 + |x|^2).$$

Hence we find

$$\mathbb{E}\int_0^T |X_t^k|^2\,\mathrm{d}t < \infty,$$

for every $T > 0$ and $k \in \mathbb{N}$ and this gives that $\{X_t^k\} \in M^2(0, T; \mathbb{R}^n)$, for any $T > 0$, $k \in \mathbb{N}$.

We shall apply Itô formula to the function $f(x) = |x|^2$ and the process $\{X_t^k,\ t \geq 0\}$

$$f(X_t^k) - f(x) = \int_0^t \Big[\langle Df(X_s^k), b_k(X_s^k)\rangle + \frac{1}{2}Tr(\sigma\sigma^* D^2 f(X_s^k))\Big]\,\mathrm{d}s + \int_0^t \langle Df(X_s^k), \sigma\,\mathrm{d}B_s\rangle.$$

As already noticed at the end of previous lecture, *a priori* it would not be possible to apply Itô formula to f, which does not belong to $C_b^2(\mathbb{R}^n)$; however, we can use an approximation procedure, working first with $|x|^2\mathrm{e}^{-\varepsilon|x|^2/2}$ and then sending ε to 0, as alrady stated in last lecture. At the end, we get

$$|X_t^k|^2 = |x|^2 + \int_0^t [2\langle X_s^k, b_k(X_s^k)\rangle + \mathrm{Tr}(\sigma\sigma^*)]\,\mathrm{d}s + 2\int_0^t \langle X_s^k, \sigma\mathrm{d}B_s\rangle \tag{11.12}$$

$t \geq 0$, $k \in \mathbb{N}$. We show below two different ways of independent interest to conclude the proof.

First conclusion. We leave as an exercise to show that there exists a constant $L > 0$ (independent from $k \in \mathbb{N}$) such that

$$\langle b_k(x), x\rangle \leq L(1 + |x|^2), \qquad x \in \mathbb{R}^n,\ k \in \mathbb{N}. \tag{11.13}$$

Taking expectation in (11.12) the stochastic integral disappears, hence we obtain

$$\begin{aligned}\mathbb{E}|X_t^k|^2 &= |x|^2 + 2\mathbb{E}\int_0^t \langle X_s^k, b_k(X_s^k)\rangle\,\mathrm{d}s + t\,\mathrm{Tr}(\sigma\sigma^*)\\ &\leq |x|^2 + T\,\mathrm{Tr}(\sigma\sigma^*) + 2L\mathbb{E}\int_0^t (1 + |X_s^k|^2)\,\mathrm{d}s, \qquad t \in [0,T].\end{aligned}$$

By an application of Gronwall's lemma there exists a constant $C_T(x) = (|x|^2 + 2LT + T\,\mathrm{Tr}(\sigma\sigma^*))\mathrm{e}^{2LT}$, independent of $k \in \mathbb{N}$, such that

$$\mathbb{E}|X_t^k|^2 \leq C_T(x), \qquad t \in [0,T].$$

If we apply Fatou's lemma we get

$$\mathbb{E}|X_t|^2 = \mathbb{E}[\liminf_{k\to\infty} |X_t^k|^2] \leq \liminf_{k\to\infty} \mathbb{E}|X_t^k|^2 \leq C_T(x), \qquad t \in [0,T],$$

which implies the thesis.

Second conclusion. In the following, we show an interesting application of stopping times which allows to prove estimate (11.7). This method can be generalized to other kind of stochastic differential equations.

Take a constant $k_0 \in \mathbb{N}$ large enough to satisfy $|x| < k_0$, where x is the initial condition in (11.3); for every $k \geq k_0$ we define the stopping times

$$\tau_k(\omega) = \inf_{t\geq 0}\{\omega \in \Omega\ :\ |X_t^k(\omega)| = k\}$$

with the usual convention that $\tau_k(\omega) = +\infty$ if $\{\cdot\} = \emptyset$. τ_k is the first passage time at level $|x| = k$ for the process $\{X_t^k,\ t \geq 0\}$.

We define the stopped processes $\{X_{t\wedge\tau_k}^k,\ t \geq 0\}$, where $x \wedge y = \min\{x, y\}$. These are well defined, adapted processes on the stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$.

By the theory of stochastic integration we know that the process

$$t \mapsto \int_0^t \langle X_s^k, \sigma\,\mathrm{d}B_s\rangle$$

is a $\mathcal{F}_t$-martingale. Corollary 4.23 states that the same holds for the stopped process

$$t \mapsto \int_0^{t\wedge\tau_k} \langle X_s^k, \sigma\,\mathrm{d}B_s\rangle.$$

Since $|X_{t\wedge\tau_k}^k| \leq k$ it holds that $b_k(X_{t\wedge\tau_k}^k) = b(X_{t\wedge\tau_k}^k)$; we apply Itô's formula to the process $X_{t\wedge\tau_k}^k$ as in (11.12); taking the expectation we get

$$\mathbb{E}|X^k_{t\wedge\tau_k}|^2 = |x|^2 + 2\mathbb{E}\int_0^{t\wedge\tau_k} \langle X^k_s, b(X^k_s)\rangle \, \mathrm{d}s + (t\wedge\tau_k)\operatorname{Tr}(\sigma\sigma^*).$$

Hypothesis (11.6) implies that there exists a constant $L > 0$ such that

$$\langle b(x), x\rangle \le L(1+|x|^2), \qquad x\in\mathbb{R}^n \tag{11.14}$$

see Problem 11.1. Using (11.14) we get, for $t\in[0,T]$

$$\mathbb{E}|X^k_{t\wedge\tau_k}|^2 \le |x|^2 + 2MT + T\operatorname{Tr}(\sigma\sigma^*) + 2M\int_0^t \mathbb{E}|X^k_{s\wedge\tau_k}|^2 \, \mathrm{d}s.$$

By an application of Gronwall's lemma there exists a constant $C_T(x) = (|x|^2 + 2LT + T\operatorname{Tr}(\sigma\sigma^*))\mathrm{e}^{2LT}$, independent from $k\in\mathbb{N}$, such that

$$\mathbb{E}|X^k_{t\wedge\tau_k}|^2 \le C_T(x), \qquad t\in[0,T].$$

We claim that $\lim_{k\to\infty} X^k_{t\wedge\tau_k} = X_t$ almost surely: then, an application of Fatou's Lemma implies

$$\mathbb{E}|X^x_t|^2 = \mathbb{E}[\liminf_{k\to\infty} |X^k_{t\wedge\tau_k}|^2] \le \liminf_{k\to\infty} \mathbb{E}|X^k_{t\wedge\tau_k}|^2 \le C_T(x), \qquad t\in[0,T].$$

and the conclusion follows.

It remains to justify the claim. It is immediate to see that τ_k is a.s. an increasing sequence; if we prove that $\lim_{k\to\infty}\tau_k = +\infty$, a.s., the claim follows. On the contrary, assume that $\lim_{k\to\infty}\tau_k(\omega) = S(\omega) < \infty$ a.s.; then we should have

$$+\infty = \lim_{k\to\infty} |X^k_{\tau_k(\omega)}(\omega)| = X^x_{S(\omega)}(\omega),$$

since $X^k_t(\omega)$ converges to $X_t(\omega)$ uniformly for $t\in[0,S(\omega)]$, compare with (11.11). But this contradicts the fact that X_t is a global solution of (11.3). Then the claim holds and the proof is complete.
□

Along the same lines of the above proof, it is possible to show that the theorem holds assuming a global boundedness of b, without the monotonicity assumption(11.6).

Proposition 11.8. *Assume that $b\in C^1(\mathbb{R}^n,\mathbb{R}^n)$ is bounded on $\mathbb{R}^n$. Then for every initial conditions $x\in\mathbb{R}^n$ and $s\in\mathbb{R}$, there exists a unique global solution $\{X(t;s;x),\ t\ge s\}$ for equation (11.4). Moreover, for every $T>0$, we have*

$$\sup_{t\in[s,T]} \mathbb{E}|X^{s,x}_t|^2 < \infty.$$

In particular,

$$\int_s^T \mathbb{E}|X^{s,x}_r|^2 \, \mathrm{d}s < \infty.$$

Proof. It is sufficient to adapt the proof of Theorem 11.5. Notice the following points:

1. If
$$\sup_{x\in\mathbb{R}^n} |b(x)| = M < \infty,$$
it holds that $|\dot v(t)| = |b(v(t) + f(t))| \le M$, therefore $|v(t) - v(s)| \le M(t-s)$ for $t \in [s, T[$ and it is possible to extend v by continuity to $t = T$, hence $T = +\infty$.
2. From
$$X_t = x + \int_s^t b(X_r)\,\mathrm{d}r + \sigma\,(B_t - B_s),$$
it follows
$$\mathbb{E}|X_t^x|^2 \le 3\mathbb{E}\Big(|x|^2 + M(T-s) + \|\sigma\|^2|B_t - B_s|^2\Big) \le C_{x,T,s}, \qquad t \in [s, T],$$
which shows the thesis.

□

Problem 11.1. Show that (11.13) holds with $L = 2(K + |b(0)|)$, where K is the constant from hypothesis (11.6).
Hint: consider separately the two cases $|x| \le k$ and $|x| > k$. Show that the same estimate holds with b_k replaced by b.

Problem 11.2. Show that Theorem 11.5 can be proved more generally when $x \in \mathbb{R}^n$ is replaced by a random variable $\eta \in L^2(\Omega;\mathbb{R}^n)$ which is $\mathcal{F}_s$-measurable.

Problem 11.3. Notice that equation (11.3) can be solved –again in a pathwise sense– also in cases when the Brownian motion $\{B_t\}$ is replaced by some other stochastic process, suitably regular.

In particular consider the case when $\{B_t\}$ is replaced by a Poisson process $\{N_t\}$, adapted with respect to a filtration $\{\mathcal{F}_t,\, t \ge 0\}$. Show that in this case there exists a unique solution $\{X_t\}$. This is by definition a $\mathcal{F}_t$-adapted process having càdlag trajectories $\mathbb{P}$-a.s. (i.e., the trajectories are in each point continuous from the right and have finite left limit) which solves (11.4) almost surely.

11.3 The Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck process $\{X_t^x,\ t \ge 0\}$, $x \in \mathbb{R}^n$, with values in $\mathbb{R}^n$, is the solution to the following linear stochastic differential equation
$$\begin{cases} \mathrm{d}X_t = AX_t\,\mathrm{d}t + \sigma\,\mathrm{d}B_t, & t \ge 0, \\ X_0 = x \in \mathbb{R}^n, \end{cases} \tag{11.15}$$
where A is a real $n \times n$ matrix. Setting $b(x) := Ax$, $x \in \mathbb{R}^n$, we see that equation (11.15) verifies the assumptions of Theorem 11.5 and there exists a unique global solution to it. Also, following the method of the proof of Theorem 11.5, we can find an explicit formula for $\{X_t^x,\ t \ge 0\}$.

Set $X_t^x = X_t$. We have
$$X_t = x + \int_0^t AX_s + \sigma B_t, \qquad t \ge 0.$$

We introduce the process $v(t) = X_t - \sigma B_t$: we find that, a.s.,

$$\begin{cases} \dot v(t) = Av(t) + A\sigma B_t, \\ v(0) = x. \end{cases}$$

Hence $v(t)$ is given by

$$v(t) = e^{tA}x + \int_0^t e^{(t-s)A} A\sigma B_s \, \mathrm{d}s.$$

Using the integration by parts formula in (9.3) we get

$$v(t) = e^{tA}x - \int_0^t \frac{d}{ds}\big(e^{(t-s)A}\big)\sigma B_s \, \mathrm{d}s = e^{tA}x - \sigma B_t + \int_0^t e^{(t-s)A}\sigma \, \mathrm{d}B_s$$

which implies

$$X_t^x = e^{tA}x + \int_0^t e^{(t-s)A}\sigma \, \mathrm{d}B_s. \tag{11.16}$$

From the above formula, we deduce that the law of X_t^x is the Gaussian measure $N(e^{tA}x, Q_t)$ with mean $e^{tA}x$ and covariance operator Q_t

$$Q_t = \int_0^t e^{sA}\sigma\sigma^* e^{sA^*} \, \mathrm{d}s, \qquad t \ge 0$$

(σ^*, resp. A^*, denote the adjoint of the matrices σ and A). We can therefore compute, for any function $f : \mathbb{R}^n \to \mathbb{R}$ Borel and bounded,

$$\mathbb{E}f(X_t^x) = \int_{\mathbb{R}^n} f(y) N(e^{tA}x, Q_t) \, \mathrm{d}y = \int_{\mathbb{R}^n} f(e^{tA}x + y) N(0, Q_t) \, \mathrm{d}y, \qquad t \ge 0. \tag{11.17}$$

Problem 11.4. Show that the Ornstein-Uhlenbeck process at time t has Gaussian law $N(e^{tA}x, Q_t)$ by using the Fourier transform.

11.3.1 The Ornstein-Uhlenbeck semigroup

Arguing as in Section 9.1.2 one proves that the Ornstein-Uhlenbeck process is a Markov process with transition probability function

$$p(t, x, A) = N(e^{tA}x, Q_t)(A), \quad A \in \mathcal{B}(\mathbb{R}^n).$$

Let us denote by $\{P_t,\ t \ge 0\}$ the (Markovian) Ornstein-Uhlenbeck semigroup acting on $B_b(\mathbb{R}^n)$, defined by

$$P_t f(x) = \mathbb{E}f(X_t^x) = \int_{\mathbb{R}^n} f(e^{tA}x + y) N(0, Q_t) \, \mathrm{d}y, \qquad t \ge 0,$$

for every $f \in B_b(\mathbb{R}^n)$. It is easy to check that

$$P_t(C_b(\mathbb{R}^n)) \subset C_b(\mathbb{R}^n), \ \ t \ge 0,$$

and that $\{P_t,\ t \geq 0\}$ is a Feller semigroup. In addition the function space $C_0(\mathbb{R}^n)$ is invariant for $\{P_t\}$ and so $\{P_t\}$ is a Feller-Dynkin semigroup according to Lecture 6.

To compute its *infinitesimal generator*, we fix $f \in C_0^\infty(\mathbb{R}^n)$ and apply Itô formula to the process $\{f(X_t^x)\}$. Setting $\sigma\sigma^* = q = (q_{ij})$, $A = (A_{ij})$ and $X_t = X_t^x$, we get

$$f(X_t(\omega)) - f(x) = \int_0^t \langle Df(X_u(\omega)), AX_u(\omega)\rangle \,\mathrm{d}u + \frac{1}{2}\int_0^t \Big(\sum_{i,j=1}^n q_{ij}\partial^2_{x_i x_j} f(X_u(\omega))\Big)\,\mathrm{d}u + Y_t(\omega), \qquad t \in [0,T], \quad \text{a.s.,}$$

where Y_t is given by

$$Y_t = \sum_{i,j=1}^n \int_0^t \partial_{x_i} f(X_u)\sigma_{ij} dB_u^j := \int_0^t \langle Df(X_u), \sigma\,\mathrm{d}B_u\rangle.$$

Let us define the possibly degenerate elliptic Ornstein-Uhlenbeck operator

$$\begin{aligned} Lf(x) &= \frac{1}{2}\sum_{i,j=1}^n q_{ij}\partial^2_{x_i x_j} f(x) + \sum_{i,j=1}^n \partial_{x_i} f(x) A_{ij} x_j \\ &= \frac{1}{2}\,\mathrm{Tr}(qD^2 f(x)) + \langle Df(x), Ax\rangle, \qquad x \in \mathbb{R}^n,\ f \in C_0^\infty(\mathbb{R}^n); \end{aligned} \tag{11.18}$$

here $\mathrm{Tr}(B)$ denotes the trace of a matrix B and $D^2 f(x)$ is the Hessian matrix of f in x. Note that the matrix q is symmetric and non-negative definite. The operator L is a particular example of elliptic Kolmogorov operator.

Applying the expectation in the Itô formula we get, using also the operator L,

$$\mathbb{E}(f(X_t^x)) = f(x) + \mathbb{E}\int_0^t Lf(X_s^x)\,\mathrm{d}s, \qquad t \geq 0. \tag{11.19}$$

By Proposition 6.12 we finally obtain the following result.

Proposition 11.9. *Let $\{P_t,\, t \geq 0\}$ be the Ornstein-Uhlenbeck semigroup on $\mathbb{R}^n$ with generator $A : D(A) \subset C_b(\mathbb{R}^n) \to C_b(\mathbb{R}^n)$. Then $C_0^\infty(\mathbb{R}^n) \subset D(A)$ and moreover if $f \in C_0^\infty(\mathbb{R}^n)$ then $Af = Lf$, where L is given in* (11.18).

12

Stochastic differential equations with multiplicative noise

This lecture is dedicated to stochastic differential equations (SDEs for short) with multiplicative noise. In particular, we will prove two theorems on global existence and uniqueness of solutions. The first one is standard and assumes that the coefficients are globally Lipschitz continuous. The second, more advanced result, concerns the case of locally Lipschitz coefficients.

Let $B = \{B_t,\ t \geq 0\}$ be a d-dimensional Brownian motion defined on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$.

Recall the definition of the space $M^p(s,T;E)$, $p \geq 1$, of all progressively measurable processes $\{Y_t,\ t \in [s,T]\}$ with values in E such that

$$\mathbb{E} \int_s^T |Y_r|_E^p \,\mathrm{d}r < \infty, \qquad 0 \leq s < T.$$

Here the space $(E, |\cdot|)$ is either the n-dimensional space $\mathbb{R}^n$ or the space of matrices $L(\mathbb{R}^d, \mathbb{R}^n)$ endowed with the Hilbert-Schmidt norm $\|\sigma\|_{L(\mathbb{R}^d,\mathbb{R}^n)} = \left(\sum_{i=1}^n \sum_{j=1}^d |\sigma_{ij}|^2\right)^{1/2}$.

In this lecture we consider the following class of stochastic differential equations

$$\begin{cases} \mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t, & t \in [s,T], \\ X_s = \eta, & 0 \leq s < T. \end{cases} \tag{12.1}$$

As opposite to the case of equations with additive noise, it will not be possible to study this equation in a pathwise sense; equation (12.1) must be understood in integral sense as

$$X_t = \eta + \int_s^t b(X_r)\,\mathrm{d}r + \int_s^t \sigma(X_r)\,\mathrm{d}B_r, \qquad t \in [s,T]. \tag{12.2}$$

Here the *coefficients* of the equation are $b(x)$, the *drift vector* and $\sigma(x)$, the *dispersion matrix*; we shall always assume that they are Borel measurable functions:

$$b : \mathbb{R}^n \to \mathbb{R}^n, \qquad \sigma : \mathbb{R}^n \to L(\mathbb{R}^d, \mathbb{R}^n);$$

the initial condition η is an $\mathcal{F}_s$-measurable random variable, thus it is independent from the increments $B_{t+h} - B_t$ for all $t \geq s$, $h \geq 0$.

Definition 12.1. *A* (strong) solution $\{X_t,\ t \in [s,T]\}$ *of Equation (12.1) is a continuous stochastic process defined and adapted on* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ *such that*

- *(initial condition)* $\mathbb{P}(X_s = \eta) = 1$*;*
- *(integrability) the process* $\{b(X_t),\ t \in [s,T]\}$ *verifies*

$$\int_s^T |b(X_r)| \, \mathrm{d}r < \infty \qquad a.s.$$

and the process $\{\sigma(X_t),\ t \in [s,T]\}$ *belongs to* $M^2(s,T; L(\mathbb{R}^d, \mathbb{R}^n))$*.*

- *(solution) the integral equation* (12.2) *is verified a.s.*

The solution $\{X_t\}$ depends on η and on $s \geq 0$. When we want to stress such dependence we write

$$X_t^{s,\eta} \quad \text{or} \quad X(t,s,\eta).$$

Problem 12.1 (An approach to diffusions via SDEs). Assume that the coefficients b and σ are bounded and continuous, and suppose that $\{X_t\}$ is a solution to Equation (12.1) with initial condition $X_0 = x \in \mathbb{R}^n$. Define the operator

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

where $a_{ij}(x)$ is the *diffusion matrix*, defined by

$$a_{ij}(x) = \sum_{k=1}^d \sigma_{ik}(x) \sigma_{jk}(x).$$

Using Itô formula, show that for every function $f \in C^2(\mathbb{R}^n)$ which is bounded together with its first and second order derivatives, it holds

$$\lim_{t \downarrow 0} \frac{1}{t} [\mathbb{E} f(X_t) - f(x)] = \mathcal{A}f(x).$$

12.1 Existence of solution: Lipschitz continuous coefficients

If one assume that $\sigma \equiv 0$, then Equation (12.1) reduces to an *ordinary*, nonstochastic differential equation, possibly with a random initial condition. Even in this case, it is well known that conditions on the coefficients shall be imposed in order to have existence and uniqueness of the solution: for instance, a typical requirement imposes locally Lipschitz continuity.

In this section, we prove a basic result assuming a global Lipschitz condition on the coefficients b and σ; later we shall generalize such assumption and concentrate on the case when b and σ are only locally Lipschitz continuous.

To treat the case of globally Lipschitz coefficients, it is useful to introduce a suitable space of stochastic processes in which we will solve the equation by a contraction principle argument.

Definition 12.2. *Given $p \geq 2$, we denote by $L^p(\Omega; C(s,T;\mathbb{R}^n))$ the set of all continuous processes $\{Y_t,\ t \in [s,T]\}$ taking values in $\mathbb{R}^n$, adapted to the filtration $\{\mathcal{F}_t\}$, such that*

$$\mathbb{E}\left[\sup_{t\in[s,T]} |Y_t|^p\right] < +\infty.$$

Identifying two processes $\{Y_t\}$ and $\{Y'_t\}$, such that $\mathbb{P}(Y_t = Y'_t) = 1$, $t \in [s,T]$, it is easy to see that this space is a Banach space with respect to the norm

$$\|Y\|_{L^p} = \left(\mathbb{E}\left[\sup_{t\in[s,T]} |Y_t|^p\right]\right)^{1/p}.$$

A process in the space defined above has a stronger regularity than what is required in $M^p(0,T;\mathbb{R}^n)$. Recall that any adapted continuous process is in particular progressively measurable.

We proceed with our first result.

Theorem 12.3. *Assume that b and σ are Lipschitz continuous mappings on $\mathbb{R}^n$, and take $\eta \in L^2(\Omega, \mathcal{F}_s; \mathbb{R}^n)$. Then there exists a unique solution process $X \in L^2(\Omega; C(s,T;\mathbb{R}^n))$ of (12.2).*

Proof. We give a proof using the contraction principle. By assumption, there exists $K > 0$ such that

$$|b(x) - b(y)| + \|\sigma(x) - \sigma(y)\| \leq K|x - y|, \qquad \forall\, x,\, y \in \mathbb{R}^n,$$

where $\|\sigma(x)\|$ denotes the Hilbert-Schmidt norm of the matrix $\sigma(x) \in L(\mathbb{R}^d;\mathbb{R}^n)$ and $|b(x)|$ is the Euclidean norm of the vector $b(x) \in \mathbb{R}^n$.

We introduce the following operators from $L^2(\Omega; C(s,T;\mathbb{R}^n))$ into itself:

$$F : u \mapsto (Fu)_t = b(u_t), \quad G : u \mapsto (Gu)_t = \sigma(u_t), \qquad t \in [s,T].$$

Let us verify that F and G are Lipschitz continuous mappings. We consider F, the other case being similar: then we aim to compute $\|Fu - Fv\|_{L^2}$, i.e.,

$$\left(\mathbb{E}\left[\sup_{t\in[s,T]} |b(u_t) - b(v_t)|^2\right]\right)^{1/2} \leq K\left(\mathbb{E}\left[\sup_{t\in[s,T]} |u_t - v_t|^2\right]\right)^{1/2} = K\|u - v\|_{L^2}$$

and we obtain

$$\|Fu - Fv\|_{L^2} + \|Gu - Gv\|_{L^2} \leq K\|u - v\|_{L^2}. \tag{12.3}$$

Further, we also introduce the following operators from $L^2(\Omega; C(s,T;\mathbb{R}^n))$ into itself:

$$I : u \mapsto (Iu)_t = \int_s^t u_r\,\mathrm{d}r, \qquad J : u \mapsto (Ju)_t = \int_s^t u_r\,\mathrm{d}B_r, \qquad t \in [s,T].$$

Before proceeding with the proof, let us record an useful result for the proceeding.

Problem 12.2. Show that I is well defined as an operator from $L^2(\Omega; C(s,T;\mathbb{R}^n))$ into itself with operator norm bounded by $(T - s)$, i.e.

$$\|Iu\|_{L^2} \leq (T - s)\|u\|_{L^2}, \qquad u \in L^2(\Omega; C(s,T;\mathbb{R}^n)).$$

Further, Doob's inequality for the Itô integral stated in (10.10) implies that

$$\mathbb{E}\left[\sup_{t\in[s,T]}\left|\int_s^t u_r\,\mathrm{d}B_r\right|^2\right] \le 4\mathbb{E}\left[\int_s^T |u_r|^2\,\mathrm{d}r\right] \le 4\mathbb{E}\left[\int_s^T \sup_{r\in[s,T]}|u_r|^2\,\mathrm{d}r\right]$$

which proves the estimate

$$\|Ju\|_{L^2} \le 2\sqrt{T-s}\,\|u\|_{L^2}.$$

Finally the operator

$$\Phi : L^2(\Omega; C(s,T;\mathbb{R}^n)) \to L^2(\Omega; C(s,T;\mathbb{R}^n)), \qquad \Phi = \eta + I\circ F + J\circ G$$

verifies

$$\|\Phi(u) - \Phi(v)\|_{L^2} \le [(T-s)K + 2\sqrt{T-s}\,K]\;\|u-v\|_{L^2}.$$

If we take $T_0 > s$ such that $[(T_0-s)K + 2\sqrt{T_0-s}\,K] = \alpha < 1$, then there exists a unique solution $\{X_t\} \in L^2(\Omega; C(s,T_0;\mathbb{R}^n))$ defined on $[s,T_0]$.

Notice that the constant α depends only on the length (T_0-s) but it does not depend on s; hence, this constant is not affected if we change the time interval without changing its length. Hence, it is possible to prove the existence of a solution $\{X^1_t\}$, defined on the time interval $[T_0, 2T_0-s]$, of the problem

$$X^1_t = X_{T_0} + \int_{T_0}^t b(X^1_r)\,\mathrm{d}r + \int_{T_0}^t \sigma(X^1_r)\,\mathrm{d}B_r.$$

Clearly, $\{X^1_t\}$ extends $\{X_t\}$ continuously to the time interval $[s, 2T_0-s]$; by the same argument in a finite number of steps it is possible to extend $\{X_t\}$ to the whole interval $[s,T]$.

□

We leave as an exercise to prove the following result, using Doob's inequality.

Problem 12.3. Assume that b and σ are Lipschitz continuous mappings on $\mathbb{R}^n$. Assume that $\{X_t\} \in M^2(s,T;\mathbb{R}^n)$ satisfies the integral equation (12.2). Then $\{X_t\} \in L^2(\Omega; C(s,T;\mathbb{R}^n))$.

12.2 Lipschitz continuity on bounded sets

In this section we are interested to weaken the global Lipschitz condition imposed in the previous section; as in the classical theory of ordinary differential equations, we consider local Lipschitz continuity as our main assumption.

Hypothesis 12.4. We assume that coefficients b and σ are locally Lipschitz continuous, i.e., for any $R > 0$ there exists $K_R > 0$ such that

$$|b(x) - b(y)| + \|\sigma(x) - \sigma(y)\| \le L_R|x-y|, \qquad \forall\, x,\, y \in \mathbb{R}^n,\ |x| \le R,\ |y| \le R. \tag{12.4}$$

Notice that in general such condition ensures existence of a local solution as well as its uniqueness.

In the stochastic case, *local existence* in time means that there exists a stopping time τ, the *explosion time* (which, in some good case, can be also equal to $+\infty$ a.s.), such that the solution $\{X_t\}$ exists for $t < \tau$ and we set $X_\tau = \infty$, the point at infinity of $\mathbb{R}^n$. As we have seen in Lecture 6, this is an absorbing point: once the solution gets at ∞, it remains there forever. We will not develop further this approach; the interested reader is referred, for instance, to `Ikeda and Watanabe` [IW81, Chapter IV]. Instead we shall introduce an additional hypothesis which, together with local Lipschitz continuity, guarantees the global existence and uniqueness of a solution to Equation (12.1).

12.2.1 Lyapunov function

In the theory of dynamical systems, and control theory, Lyapunov functions are a family of functions that can be used to demonstrate the stability or instability of some state points of a system. In our construction, Lyapunov function are a key tool in order to prove global existence for the solution.

As it happens in the above-named fields, we shall face the problem of finding a Lyapunov function for our problem. There is no direct method to obtain a Lyapunov function (in the following we shall propose some examples).

Definition 12.5. *A real function $V \in C^2(\mathbb{R}^n)$ is called a* Lyapunov function *for* (12.1) *if, for some constant $K > 0$, it holds:*

1. $V(x) \geq 0$;
2. $LV(x) \stackrel{def}{=} \langle \nabla V(x), b(x)\rangle + \frac{1}{2}\operatorname{Tr}(\sigma(x)\nabla^2 V(x)\sigma^*(x)) \leq KV(x)$, $x \in \mathbb{R}^n$;
3. $\lim\limits_{|x|\to+\infty} V(x) = +\infty$.

Example 12.6. The basic example of Lyapunov function is of course $V(x) = 1 + |x|^2$. In this case, condition 2. becomes

$$2\langle x, b(x)\rangle + \|\sigma(x)\|^2 \leq K(1 + |x|^2), \qquad x \in \mathbb{R}^n. \tag{12.5}$$

Hence, if b and σ verify condition (12.5), we say that $V(x) = 1+|x|^2$ is a Lyapunov function associated to (12.1).

Example 12.7. We next provide an example of coefficients b and σ for which there exists a Lyapunov function. This contruction is somewhat more general than what is needed for our main result Theorem 12.8, see Remark 12.10.

Assume for simplicity $d = 1$, so the Brownian motion $\{B_t\}$ is a 1-dimensional process and $\sigma(x)$ is a vector in $\mathbb{R}^n$. Let b and σ be locally Lipschitz mappings on $\mathbb{R}^n$; we impose further the condition

$$\begin{cases} \langle b(x), x\rangle \leq K_1(1 + |x|^2) \\ |\langle \sigma(x), x\rangle| \geq K_2|\sigma(x)||x| \end{cases} \tag{12.6}$$

for any $x \in \mathbb{R}^n$, where K_1, K_2 are real constants, with $1/\sqrt{2} < K_2 \leq 1$. Then we can provide explicitly a Lyapunov function $V(x)$ associated to (12.1) as stated in Definition 12.5.

Let φ be an arbitrary function in $C^2(\mathbb{R}_+, \mathbb{R}_+)$ verifying $\varphi(r) = r^\alpha$ in $[r_0, +\infty)$, with $\alpha \in (0, 1 - \frac{1}{2K_2^2})$ and $r_0 > 0$. Then, for the mapping $V(x) = \varphi(|x|^2) + c$ (c is a suitable constant, such that conditions 1. and 2. hold in the ball $B(0, \sqrt{r_0})$) we have, as $|x|^2 > r_0$,

$$
\begin{aligned}
LV(x) &= 2\alpha|x|^{2\alpha-2}\langle b(x), x\rangle + 2\alpha(\alpha-1)|x|^{2\alpha-4}\langle\sigma(x), x\rangle^2 + \alpha|x|^{2\alpha-2}|\sigma(x)|^2 \\
&\le 2\alpha K_1|x|^{2\alpha-2}(1+|x|^2) + \alpha|x|^{2\alpha-2}|\sigma(x)|^2(2K_2^2(\alpha-1)+1) \\
&\le \frac{2\alpha K_1}{r_0^{1-\alpha}} + 2\alpha K_1\varphi(|x|^2) \le 2\alpha K_1(\frac{1}{r_0}+1)V(x).
\end{aligned}
$$

Then $V(x)$ is a Lyapunov function associated to (12.1) and the proof is complete.

Problem 12.4. In particular, a similar construction with $\phi(r) = \sqrt{r}$ shows the existence of a Lyapunov function in case $E = \mathbb{R}$, $b \equiv 0$ and any σ; the non explosion for 1-dimensional stochastic equations of the form $\mathrm{d}X_t = \sigma(X_t)\,\mathrm{d}B_t$ is well known in literature, usually proved with other techniques, see for instance McKean [McK69].

12.2.2 Existence of solution: locally Lipschitz continuous coefficients

We now state the main result of this lecture.

Theorem 12.8. *Assume that coefficients b and σ in* (12.1) *are locally Lipschitz continuous and that there exists a Lyapunov function V for* (12.1) *such that $\mathbb{E}[V(\eta)] < \infty$. Assume also that*

$$\sup_{x\in\mathbb{R}^n} \|\sigma(x)\| < \infty. \tag{12.7}$$

Then there exists a unique global solution $\{X_t,\, t \in [s,T]\}$ to (12.1)*. Moreover the following estimate holds:*

$$\sup_{t\in[s,T]} \mathbb{E}[V(X_t)] < \infty. \tag{12.8}$$

Remark 12.9. Note that if $\eta = x \in \mathbb{R}^n$ a.s., then the assumption $\mathbb{E}[V(\eta)] < \infty$ is automatically satisfied.

Remark 12.10. One may wonder if assumption (12.7) is really necessary to prove the result. The answer is no! We were forced to introduce it because we need to treat the stochastic integral $\int_s^t \sigma(X_r)\,\mathrm{d}B_r$. Our construction in Lecture 10 requires that $\{\sigma(X_t)\}$ belongs to $M^2(s,T;L(\mathbb{R}^d,\mathbb{R}^n))$, which follows from (12.7).

However, there is a way to define the stochastic integral also for the class of of all continuous adapted processes $Y : [s,T]\times\Omega \to L(\mathbb{R}^d,\mathbb{R}^n)$. With this generalization of the stochastic integral, we can avoid hypothesis (12.7) (indeed, clearly $\{\sigma(X_t)\}$ is a continuous adapted process).

Details for this part of stochastic calculus are given in many textbooks, as for instance Karatzas and Shreve [KS88] or Kallenberg [Ka02].

In the special case when the Lyapunov function for (12.1) is $V(x) = 1 + |x|^2$ (see (12.5)) we can weaken (12.7) (yet remaining inside the field of applicability of our definition of stochastic integral). Indeed, in this case, estimate (12.8) shows in particular that the solution $\{X_t\}$ belongs to $M^2(s,T;\mathbb{R}^n)$. Thus if we replace hypothesis (12.7) with the following one

$$\|\sigma(x)\| \le C(1+|x|), \qquad x \in \mathbb{R}^n \tag{12.9}$$

(i.e., σ has a sublinear growth) it follows that $\{\sigma(X_t)\}$ belongs to $M^2(s,T;L(\mathbb{R}^d,\mathbb{R}^n))$ and so the stochastic integral $\int_s^t \sigma(X_r)\,\mathrm{d}B_r$ is well defined.

Before we proceed with the proof of the theorem we need a technical lemma concerning stochastic integrals and stopping times.

Lemma 12.11. *Let $\{G(t)\} \in M^2(s,T;L(\mathbb{R}^d,\mathbb{R}^n))$. Let $\tau \le T$ be a stopping time. Then the following assertions hold:*

(i) $\displaystyle\int_s^\tau G_r \, \mathrm{d}B_r = \int_s^T G(r)\, \mathbb{1}_{\{r\le\tau\}}\, \mathrm{d}B_r$;

(ii) $\displaystyle\mathbb{E}\left[\int_s^\tau G(r)\, \mathrm{d}B_r\right] = 0$;

(iii) $\displaystyle\mathbb{E}\left[\int_s^\tau G(r)\, \mathrm{d}B_r\right]^2 = \mathbb{E}\int_s^\tau \|G(r)\|^2 \mathrm{d}r$.

Proof. We only prove the identity (i), since the other ones follow easily from (i). We first note that the process $\{G(r)\, \mathbb{1}_{\{r\le\tau\}},\, r\in[s,T]\} \in M^2(s,T;L(\mathbb{R}^d,\mathbb{R}^n))$.

In case G is an elementary process and τ takes only a finite number of values, our claims is a direct consequence of the definition of stochastic integral. We only note that, defining the sets $\Omega_k = \{\omega : \tau(\omega) = t_k\}$, $k = 1,\dots,m$, it holds

$$\int_s^\tau G_r \, \mathrm{d}B_r = \sum_{k=1}^m \mathbb{1}_{\Omega_k} \int_s^{t_k} G_r \, \mathrm{d}B_r$$

and moreover $\{G(r)\, \mathbb{1}_{\{r\le\tau\}},\, r\in[s,T]\}$ is an elementary process.

Then we consider the case in which τ takes only a finite number of values and $G \in M^2(s,T;L(\mathbb{R}^d,\mathbb{R}^n))$. We consider a sequence of elementary processes $\{G_n(t)\}$ converging to G:

$$\mathbb{E}\int_s^T \|G(r) - G_n(r)\|^2 \, \mathrm{d}r \longrightarrow 0, \qquad n\to\infty.$$

We have easily that

$$\mathbb{E}\Big|\int_s^\tau [G(r) - G_n(r)]\, \mathrm{d}B_r\Big|^2 + \mathbb{E}\Big|\int_s^T [G(r) - G_n(r)]\, \mathbb{1}_{\{r\le\tau\}}\mathrm{d}B_r\Big|^2 \longrightarrow 0$$

as $n\to\infty$, and so the claim follows also in this case.

Finally, we consider the general case. If τ is a stopping time, we may consider a sequence of stopping times τ_n taking only a finite number of values and approximating τ from the above, i.e., $\tau\le\tau_n$, a.s. and $\tau_n\to\tau$, as $n\to\infty$, a.s. (compare the proof of Theorem 6.10). We can assume that $\tau_n \le T$ (by considering if necessary $\tau_n\wedge T$). Since $I(t) = \int_s^t G(r)\, \mathrm{d}B_r$ is a continuous process and $\tau_n \downarrow \tau$ a.s., it holds $I(\tau_n)\to I(\tau)$ a.s.

It remains to consider the term $\int_s^T G(r)\, \mathbb{1}_{\{r\le\tau_n\}}\, \mathrm{d}B_r$. We find

$$\begin{aligned}\mathbb{E}\Big|\int_s^T G(r)\, \mathbb{1}_{\{r\le\tau_n\}}\, \mathrm{d}B_r - \int_s^T G(r)\, \mathbb{1}_{\{r\le\tau\}}\, \mathrm{d}B_r\Big|^2 &= \mathbb{E}\Big|\int_s^T G(r)\, (\mathbb{1}_{\{r\le\tau_n\}} - \mathbb{1}_{\{r\le\tau\}})\, \mathrm{d}B_r\Big|^2 \\ &= \mathbb{E}\int_s^T \|G(r)\|^2\, \mathbb{1}_{\{\tau<r\le\tau_n\}}\, \mathrm{d}r \;\to\; 0,\end{aligned}$$

as $n\to\infty$, by the dominated convergence theorem. The proof is complete.

□

Proof (Theorem 12.8). We assume $s = 0$ to simplify the notation.

Uniqueness of solution. In this part we only use the hypothesis that coefficients b and σ are locally Lipschitz continuous mappings. Let $\{X_t,\, t \in [0,T]\}$ and $\{Y_t,\, t \in [0,T]\}$ be two continuous processes, defined and adapted on the stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. Define the stopping time $\tau = \tau_R$, $R > 0$,

$$\tau = \begin{cases} \inf\{t \mid 0 \le t \le T: \ |X_t| > R \text{ or } |Y_t| > R\}, \\ T \quad \text{if it is } |X_t| \le R \text{ and } |Y_t| \le R \text{ for } 0 \le t \le T. \end{cases}$$

We assume that $\{X_t,\, t \in [0,T]\}$ and $\{Y_t,\, t \in [0,T]\}$ solve (12.1) in the random interval $[0,\tau]$ and we show that $X_t = Y_t$ a.s. for all $t \in [0,\tau]$.

We know that $\{X_t,\, t \in [0,T]\}$ and $\{Y_t,\, t \in [0,T]\}$ both verify the equation

$$X_t = \eta + \int_0^t b(X_r)\,\mathrm{d}r \ + \int_0^t \sigma(X_r)\,\mathrm{d}B_r, \qquad t \in [0,\tau].$$

Moreover, using Lemma 12.11,

$$\begin{aligned} X_{t\wedge\tau} - Y_{t\wedge\tau} &= \int_0^{t\wedge\tau} [b(X_r) - b(Y_r)]\,\mathrm{d}r + \int_0^{t\wedge\tau} [\sigma(X_r) - \sigma(Y_r)]\,\mathrm{d}B_r \\ &= \int_0^t [b(X_r) - b(Y_r)] \mathbb{1}_{\{r\le\tau\}}\,\mathrm{d}r + \int_0^t [\sigma(X_r) - \sigma(Y_r)] \mathbb{1}_{\{r\le\tau\}}\,\mathrm{d}B_r. \end{aligned}$$

Using the inequality $(a+b)^2 \le 2a^2 + 2b^2$, we get

$$\sup_{t\le s} |X_{t\wedge\tau} - Y_{t\wedge\tau}|^2 \le 2\Big| \int_0^s |b(X_r) - b(Y_r)| \mathbb{1}_{\{r\le\tau\}}\,\mathrm{d}r \Big|^2 + 2 \sup_{t\le s} \Big| \int_0^t [\sigma(X_r) - \sigma(Y_r)] \mathbb{1}_{\{r\le\tau\}}\,\mathrm{d}B_r \Big|^2.$$

Denoting by $L = L_R$ the Lipschitz constant of b and σ on the closed ball centered in 0 with radius $R > 0$, we have, using also Doob's inequality,

$$\begin{aligned} \mathbb{E}\left[\sup_{t\le s} |X_{t\wedge\tau} - Y_{t\wedge\tau}|^2\right] &\le 2TL^2 \mathbb{E} \int_0^s |X_{r\wedge\tau} - Y_{r\wedge\tau}|^2 \mathbb{1}_{\{r\le\tau\}}\,\mathrm{d}r + 2L^2 \mathbb{E} \int_0^s |X_{r\wedge\tau} - Y_{r\wedge\tau}| \mathbb{1}_{\{r\le\tau\}}\,\mathrm{d}r \\ &\le 2TL^2 \mathbb{E} \int_0^s |X_{r\wedge\tau} - Y_{r\wedge\tau}|^2\,\mathrm{d}r + 2L^2 \mathbb{E} \int_0^s |X_{r\wedge\tau} - Y_{r\wedge\tau}|\,\mathrm{d}r \end{aligned}$$

Setting $g(s) = \mathbb{E}\left[\sup_{t\le s} |X_{t\wedge\tau} - Y_{t\wedge\tau}|^2\right]$, we find that $g(s) \le C_T \int_0^s g(r)\,\mathrm{d}r$, $s \in [0,T]$. By the Gronwall lemma we deduce that $g(s) = 0$, $s \in [0,T]$. The assertion follows.

Existence of a global solution. We fix an arbitrary $T > 0$ and we show the existence of a solution $\{X_t\}$ on $[0,T]$. In connection with our assumptions we consider, for every integer $N > 0$, the real mappings

$$\psi_N(x) = \begin{cases} x, & \text{for } |x| \le N \\ Nx/|x|, & \text{for } |x| > N \end{cases}$$

and the functions $b_N = f \circ \psi_N$, $\sigma_N = \sigma \circ \psi_N$. For each $N > 0$ these mappings are (globally) Lipschitz continuous. Therefore, for each N there exists a unique solution $\{X_N(t),\, t \geq 0\}$ to the problem

$$X_N(t) = \eta + \int_0^t b_N(X_N(s))\,\mathrm{d}s + \int_0^t \sigma_N(X_N(s))\,\mathrm{d}B_s, \qquad t \geq 0 \tag{12.10}$$

We call the processes $\{X_N(t)\}$ the approximating solutions. We define also a family of stopping times

$$\tau_N = \begin{cases} \inf\{t \mid 0 \leq t \leq T:\ |X_N(t)| > N\}, \\ T \quad \text{if it is } |X_N(t)| \leq N \ \text{ for } 0 \leq t \leq T. \end{cases} \tag{12.11}$$

Note that $\tau_N \leq T$, a.s., $N \in \mathbb{N}$.
I Step. The approximating solutions have the following properties

1. the solutions $X_N(t) = X_{N+1}(t)$ coincide for any $t \in [0, \tau_N]$;
2. $\tau_N \leq \tau_{N+1}$, a.s..

We start with the solution X_N: (up to time τ_N) it verifies

$$X_N(t \wedge \tau_N) = \eta + \int_0^{t\wedge\tau_N} b_N(X_N(s))\,\mathrm{d}s + \int_0^{t\wedge\tau_N} \sigma_N(X_N(s))\,\mathrm{d}B_s, \qquad t \in [0, T].$$

Since for t in the random interval $[0, \tau_N]$ it holds $|X_N(t)| \leq N$, we have $b_N(X_N(t)) = b_{N+1}(X_N(t)) = b(X_N(t))$ and similarly for σ, so that

$$X_N(t \wedge \tau_N) = \eta + \int_0^{t\wedge\tau_N} b_{N+1}(X_N(s))\,\mathrm{d}s + \int_0^{t\wedge\tau_N} \sigma_{N+1}(X_N(s))\,\mathrm{d}B_s, \qquad t \in [0, T].$$

Therefore, uniqueness of the solution implies that in the random interval $[0, \tau_N]$ it is $X_N(t) = X_{N+1}(t)$; further, since $X_{N+1}(t) = X_N(t)$ is bounded in norm by N in $[0, \tau_N]$, it must be $\tau_{N+1} \geq \tau_N$.

II Step. For any $t \in (0, T]$, we define the measurable sets

$$\Omega_t = \bigcup_{N \geq 1} \{\tau_N \geq t\} \tag{12.12}$$

and show that $\mathbb{P}(\Omega_t) = 1$, $t \in (0, T]$. It is clear that on $\{\tau_N \geq t\}$ it is $X_N(s) = X_{N+1}(s)$ for all $s \in [0, t]$; however, if N is small, it can easily happen that the set $\{\tau_N \geq t\}$ is empty. Since $\tau_N \leq \tau_{N+1}$, it follows that Ω_t is an increasing union and the assertion follows if we verify that

$$\lim_{N\to\infty} \mathbb{P}(\tau_N \geq t) = 1 \text{ or equivalently } \lim_{N\to\infty} \mathbb{P}(\tau_N < t) = 0. \tag{12.13}$$

To prove (12.13) we first take a function $\phi_N \in C_0^\infty(\mathbb{R}^n)$ such that $|\phi_N(x)| \leq 1$ for every $x \in \mathbb{R}^n$, $\phi_N(x) = 1$ for $|x| \leq N$ and $\phi_N(x) = 0$ for $|x| > 2N$. Then we define $V_N(x) = V(x) \cdot \phi_N(x)$ (where V is the Lyapunov function associated to (12.1), which exists by assumption, and K is the bound in Definition 12.5, point 2.) and then we apply Itô formula to the process $\mathrm{e}^{-Kt} V_N(X_N(t))$ (note that $V_N \in C_b^2(\mathbb{R}^n)$). We find, for $t \leq T$,

$$\mathrm{e}^{-K(t\wedge\tau_N)}V_N(X_N(t\wedge\tau_N)) = V_N(\eta) + \int_0^{t\wedge\tau_N} \big(-K\mathrm{e}^{-Kr}V_N(X_N(r)) + \mathrm{e}^{-Kr}LV_N(X_N(r))\big)\,\mathrm{d}r \\ + \int_0^{t\wedge\tau_N} \mathrm{e}^{-Kr}\langle\nabla V_N(X_N(r)), \sigma_N(X_N(r))\,\mathrm{d}B_r\rangle. \quad (12.14)$$

Note that $(\partial_t + L)(\mathrm{e}^{-Kt}V_N) = -K\mathrm{e}^{-Kt}V_N + \mathrm{e}^{-Kt}LV_N$, where

$$LV_N(x) = \langle\nabla V_N(x), b(x)\rangle + \frac{1}{2}\,\mathrm{Tr}(\sigma(x)\nabla^2 V_N(x)\sigma^*(x)).$$

Since, for any $x \in \mathbb{R}^n$ such that $|x| \le N$,

$$LV(x) = LV_N(x) \le KV_N(x),$$

the second term in the right-hand side of (12.14) is nonpositive; next, taking expectation in (12.14) and using Lemma 12.11 to compute the expectation of the stochastic interal, we obtain

$$\mathbb{E}[\mathrm{e}^{-K(t\wedge\tau_N)}V_N(X_N(t\wedge\tau_N))] \le \mathbb{E}[V_N(\eta)] + \mathbb{E}\Big[\int_0^{t\wedge\tau_N} \mathrm{e}^{-Kr}\langle\nabla V_N(X_N(r)), \sigma_N(X_N(r))\,\mathrm{d}B_r\rangle\Big] \\ \le \mathbb{E}[V_N(\eta)].$$

It follows that, for any $t \le T$, since $\mathrm{e}^{-K(t\wedge\tau_N)} \ge \mathrm{e}^{-Kt}$,

$$\mathrm{e}^{Kt}\mathbb{E}[V_N(\eta)] \ge \mathbb{E}[V_N(X_N(t\wedge\tau_N))] \ge \mathbb{E}[V_N(X_N(t\wedge\tau_N))\,\mathbb{1}_{\{\tau_N\le t\}}] = \mathbb{E}[V_N(X_N(\tau_N))\,\mathbb{1}_{\{\tau_N\le t\}}] \\ \ge \Big(\inf_{\{|x|=N\}} V(x)\Big)\,\mathbb{E}[\mathbb{1}_{\{\tau_N\le t\}}] = \Big(\inf_{\{|x|=N\}} V(x)\Big)\,\mathbb{P}(\tau_N \le t).$$

Since $V \to +\infty$ as $x \to \infty$, the quantity $M_N = \inf_{\{|x|=N\}} V(x)$ is arbitrarily large for N large enough; however, $V_N(\eta) \le V(\eta)$, so that for every $t \in [0,T]$ it holds

$$\mathbb{P}(\tau_N \le t) \le \frac{1}{M_N}\mathrm{e}^{Kt}\,\mathbb{E}[V(\eta)] \to 0 \qquad \text{as } N \to \infty$$

which shows (12.13).

III Step. We define the solution $\{X_t\}$. Fix $t \in (0,T]$ and define $X_t : \Omega \to \mathbb{R}^n$ as follows: if $\omega \notin \Omega_t$, we set $X_t(\omega) := 0$; if $\omega \in \Omega_t$, there exists $N \ge 1$ such that $\omega \in \{\tau_N \ge t\}$ and we set

$$X_t(\omega) := X_N(t,\omega).$$

The definition is meaningful thanks to Step 1. Note that $\{X_t\}$ has continuous trajectories: indeed, if $\omega \in \Omega_T$, then there exists $N \ge 1$ such that $\omega \in \{\tau_N = T\}$ and we have that $t \mapsto X_t(\omega) = X_N(t,\omega)$ is continuous on $[0,T]$ (recall that $\mathbb{P}(\Omega_T) = 1$).

Moreover, it is $\mathcal{F}_t$-adapted for any $t \in [0,T]$. To see this we write, for any $A \in B(\mathbb{R}^n)$,

$$\{\omega \,:\, X_t(\omega) \in A\} = \bigcup_{N\ge 1} (\{\omega \,:\, X_t(\omega) \in A\} \cap \{\tau_N \ge t\}) \\ = \bigcup_{N\ge 1} (\{\omega \,:\, X_N(t)(\omega) \in A\} \cap \{\tau_N \ge t\}) \in \mathcal{F}_t.$$

It remains to show that $\{X_t\}$ is actually a solution to (12.1). We have that, for any $t \le T$, $\omega \in \Omega_T$:

$$X_N(t \wedge \tau_N) = \eta + \int_0^{t\wedge\tau_N} b_N(X_N(s))\,\mathrm{d}s + \int_0^{t\wedge\tau_N} \sigma_N(X_N(s))\,\mathrm{d}B_s$$

hence

$$X_{t\wedge\tau_N} = \eta + \int_0^{t\wedge\tau_N} b(X_s)\,\mathrm{d}s + \int_0^{t\wedge\tau_N} \sigma(X_s)\,\mathrm{d}B_s, \qquad N \ge 1. \tag{12.15}$$

Now if $\tilde{\omega} \in \Omega_T$, there exists $\tilde{N}$ such that $\omega \in \{\tau_{\tilde{N}} = T\}$. Thus, computing (12.15) in $\tilde{\omega}$, we get

$$X_t = \eta + \int_0^t b(X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}B_s.$$

This shows that $\{X_t\}$ is a solution to the SDE (12.1).
□

Remark 12.12. We mention an interesting result proved in a paper by Da Prato, Iannelli and Tubaro [Da78]. In this paper it is proved the global existence and uniqueness of a solution to the stochastic differential equation (12.1) without assuming that the coefficients are locally Lipschitz. It is only assumed the following dissipativity condition: there exists $\alpha > 0$ such that

$$2\langle b(x) - b(y), x - y\rangle + \|\sigma(x) - \sigma(y)\|^2 \le \alpha|x - y|^2, \tag{12.16}$$

for every $x, y \in \mathbb{R}^n$. The proof uses suitable approximations of the coefficients by means of Yosida mappings. Note that condition (12.16) is usually considered in the theory of non-linear contraction semigroups in Banach spaces.

Related to condition (12.16), we should mention also a general result obtained by Krylov [Kr95], where the author uses the Euler approximation scheme in order to construct a global solution.

Appendix: an introduction to financial markets

Five years before Einstein's theory of Brownian movement, a similar model was described by the French matematician L. Bachelier in his doctorate thesis, in order to describe the random fluctuations in the stock market of Paris. He was concerned with prices (the macroscopic molecules in Brownian movement theory) whose movements were influenced by agents in the market (which played the rôle of the liquid molecules) and concluded that fluctuations where given by a centered Gaussian distribution with variance $2Dt$, for a positive constant D.

We can not consider such a model as a realistic one, at least due to the fact that the prices can not assume negative values with positive probability; now we propose a different, yet simple model, to describe the stock price process.

Let $\{S_t,\ t \ge 0\}$, be the price of a given stock (but it can also be any other object in the market); in order to describe the variation of the price $\Delta S_t = S_{t+\Delta t} - S_t$, for a small time interval $\Delta t > 0$, we proceed as follows

$$\frac{\Delta S_t}{S_t} = \frac{S_{t+\Delta t} - S_t}{S_t} = \Delta t \cdot \mu + \text{“noise”}$$

where μ is the *drift* of the process (a value which measure the behaviour in the mean) and the noisy term models the typical fluctuations of the market.

We may model the prices with the same terminology of the Brownian movement. The particle is the stock in the market (the fluid); there is a general movement of the fluid (depending on external factors that we may think to be deterministic) and random, local forces, as is, for instance, the balance between the numbers of buyers or sellers acting in a certain moment: any movement of this number cause the price to increase or decrease. However, we do not assume that the action of the noise on the particle is additive, but rather we assume it to be multiplicative; then the relevant proportional variation $\frac{\Delta S_t}{S_t}$ in the interval Δ_t is given by a deterministic part $\mu \triangle t$ and a stochastic part proportional to the variation $\nu \Delta B_t$ of the underlying Brownian motion. Therefore, we say that the process S_t satisfies the following stochastic equation

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}B_t, \tag{12.17}$$

If $\{B_t,\ t \geq 0\}$ is a Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, and $\mu \in \mathbb{R}$, $\nu > 0$ and $S_0 > 0$, the process S_t is given by

$$S_t = S_0 \exp(\mu t - \frac{1}{2}\sigma^2 t + \sigma B_t). \tag{12.18}$$

S_t is the *geometrical Brownian motion*, μ is called the *drift* and σ^2 is the *volatility*. This model plays a key rôle in mathematical finance, since it is at the basis of the renowned Black and Scholes model for pricing derivatives; however, its simplicity implies in particular that it can not be used in times of crisis, when sudden jumps may appear in the general behaviour of the market.

Problem 12.5. Compute Itô differential of the geometric Brownian motion (12.18) and show that it verifies equation (12.17).

Linear stochastic equations

We consider the following linear stochastic differential equation for an $\mathbb{R}^n$-valued process X_t:

$$\begin{cases} \mathrm{d}X_t = [A(t)X_t + f(t)]\,\mathrm{d}t + \sum_{j=1}^{d} \Big[B_j(t)X_t + g_j(t)\Big]\,\mathrm{d}\beta^j(t) \\ X_{t_0} = x_0, \qquad t_0 \leq t \leq T \end{cases} \tag{12.19}$$

where $W_t = (\beta^1(t), \dots, \beta^d(t))$ is a d-dimensional Brownian motion defined on a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. In the above equation, $A(t)$ and $B_j(t)$ are deterministic $n \times n$ matrices and $f(t)$ and $g_j(t)$ are $\mathbb{R}^n$-valued functions defined on $[t_0, T]$.

Theorem 12.13. *The linear stochastic differential equation (12.19) has, for any initial value $x_0 \in \mathbb{R}^n$ a unique continuous solution throughout the interval $[t_0, T]$ provided only that the functions $A(s)$, $f(s)$, $B_j(s)$, $g_j(s)$ are measurable and bounded on that interval.*

The proof of Theorem 12.13 follows easily by a suitable modification of the proof of Theorem 12.3.

Explicit solutions to non-homogeneous linear equations can be obtained, as in the ordinary differential equation case, by the variation of parameters technique. The following result is totally analogous to the ordinary differential equations situation. We introduce the notion of fundamental matrix of the homogeneous equation corresponding to (12.19).

Theorem 12.14. *The linear stochastic differential equation (12.19) with initial value x_0 has, on $[t_0, T]$ the solution*

$$X_t = \Phi_t \left(x_0 + \int_{t_0}^{t} \Phi_s^{-1} \, dY_s \right). \tag{12.20}$$

Here,

$$\mathrm{d}Y_t = [f(t) - \sum_{j=1}^{d} B_j(t) g_j(t)] \, \mathrm{d}t + \sum_{j=1}^{d} g_j(t) \, \mathrm{d}\beta^j(t)$$

and the $n \times n$ matrix Φ_t is the fundamental matrix of the corresponding homogeneous equation, that is, the solution of the stochastic differential equation

$$\begin{cases} \mathrm{d}\Phi_t = A(t)\Phi_t \, \mathrm{d}t + \sum\limits_{j=1}^{d} B_j(t)\Phi_t \, \mathrm{d}\beta^j(t) \\ \Phi(t_0) = I. \end{cases} \tag{12.21}$$

Proof. We only sketch the proof (see `Arnold` [Ar74] for more details). The most interesting part is the proof that the matrix Φ_t is invertible. Indeed, since we already know that a solution to (12.19) exists unique, to prove the theorem is sufficient to verify directly that the given $\{X_t\}$ verifies the equation.

The problem is well known and can be solved with different approaches. Here, we write $\Phi = (\Phi_1, \dots, \Phi_n)$, where Φ_i are column vectors. Hence we have $Y_t := (\det \Phi) = \det(\Phi_1, \dots, \Phi_n)$ and the Itô differential verifies the equation

$$\mathrm{d}Y_t = \left[\sum_{i=1}^{n} \det(\Phi_1, \dots, A\Phi_i, \dots, \Phi_d) + \frac{1}{2} \sum_{i \neq j}^{n} \sum_{k=1}^{m} \det(\Phi_1, \dots, B_k\Phi_i, \dots, B_k\Phi_j, \dots, \Phi_n) \right] \mathrm{d}t$$

$$+ \sum_{k=1}^{d} \left[\sum_{i=1}^{n} \det(\Phi_1, \dots, B_k\Phi_i, \dots, \Phi_d) \right] \mathrm{d}\beta^k(t)$$

Then, we recall the following identities, which hold for any $n \times n$ matrix U:

$$\sum_{i=1}^{n} \det(\Phi_1, \dots, U\Phi_i, \dots, \Phi_n) = \mathrm{Tr} U \cdot \det \Phi \tag{12.22}$$

$$\sum_{i,j=1}^{n} \det(\Phi_1, \dots, U\Phi_i, \dots, U\Phi_j, \dots, \Phi_n) = \frac{1}{2} \left[(\mathrm{Tr} U)^2 - \mathrm{Tr} U^2 \right] \cdot \det \Phi \tag{12.23}$$

and we get the following expression for the differential of Y_t:

$$\begin{cases} \mathrm{d}Y_t = \left[\mathrm{Tr} A + \frac{1}{2} \sum_{k=1}^{d} (\mathrm{Tr} B_k)^2 - \mathrm{Tr} B_k^2 \right] Y_t \, \mathrm{d}t + \sum_{k=1}^{d} \mathrm{Tr} B_k Y_t \, \mathrm{d}\beta^k(t) \\ Y_0 = 1 \end{cases}$$

that is,

$$Y_t = \exp\left(\int_0^t \left[\mathrm{Tr}\Big(A(s) - \sum_{k=1}^{d} B_k^2(s)\Big)\right] \mathrm{d}s + \sum_{k=1}^{d} \int_0^t \mathrm{Tr} B_k(s)\, \mathrm{d}\beta^k(s)\right).$$

Since the determinant of $\Phi(t)$ is strictly positive, the matrix is invertible.

□

13

Diffusion processes, Markov semigroups and Kolmogorov equations

This lecture is dedicated to Markov semigroups associated to diffusion processes which are solutions of stochastic differential equations with globally Lipschitz coefficients. Although the Markov property for solutions of SDEs could be proved under the general hypotheses of our Theorems 11.5 and 12.8, here we only consider the case of globally Lipschitz coefficients for the sake of simplicity. We refer to more advanced books, like `Stroock and Varadhan` [SV79], `Karatzas and Shreve` [KS88], `Ikeda and Watanabe` [IW81], `Ethier and Kurtz` [EK86] for the study of Markov property in full generality.

13.1 Continuous dependence on data

Our basic framework is the same as in the previous lecture. Let $B = \{B_t,\ t \geq 0\}$ be a d-dimensional Brownian motion defined on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. Recall the definition of the space $M^p(s, T; E)$, $p \geq 1$, of all progressively measurable processes $\{Y_t,\ t \in [s, T]\}$ with values in E such that

$$\mathbb{E} \int_s^T |Y_r|_E^p \, \mathrm{d}r < \infty, \qquad 0 \leq s < T.$$

Here the space $(E, |\cdot|)$ is either the n-dimensional space $\mathbb{R}^n$ or the space of matrices $L(\mathbb{R}^d, \mathbb{R}^n)$ endowed with the Hilbert-Schmidt norm $\|\sigma\|_{L(\mathbb{R}^d,\mathbb{R}^n)} = \left(\sum_{i=1}^n \sum_{j=1}^d |\sigma_{ij}|^2 \right)^{1/2}$.

In this lecture we consider the following class of stochastic differential equations

$$X_t = \eta + \int_s^t b(X_r) \, \mathrm{d}r + \int_s^t \sigma(X_r) \, \mathrm{d}B_r, \qquad t \geq s, \tag{13.1}$$

where we assume there exists $K > 0$ such that

$$|b(x) - b(y)| + \|\sigma(x) - \sigma(y)\| \leq K|x - y|, \qquad \forall\, x,\, y \in \mathbb{R}^n, \tag{13.2}$$

and moreover the initial condition $\eta \in L^2(\Omega, \mathbb{R}^n)$ is $\mathcal{F}_s$-measurable.

By the results in the previous lecture, we know that, *for any* $T > 0$, there exists a unique solution $X = \{X_t^{s,\eta},\ t \in [s,T]\}$ which belongs to $L^2(\Omega; C(s,T;\mathbb{R}^n))$, i.e., $\{X_t^{s,\eta},\ t \in [s,T]\}$ is a continuous process with values in $\mathbb{R}^n$, adapted to the filtration $\{\mathcal{F}_t\}$ and such that

$$\mathbb{E}\left[\sup_{t\in[s,T]} |X_t^{s,\eta}|^2\right] < +\infty.$$

Recall that in particular $X \in M^2(s,T;\mathbb{R}^n)$. Note that we can introduce a continuous adapted process $X = \{X_t^{s,\eta},\ t \geq s\}$, global solution of (13.1), simply defining for any $t > s$, $X_t^{s,\eta} = Y_t$, where $\{Y_u^{s,\eta},\ u \in [s,[t]+1]\}$ denotes the solution to (13.1) on the time interval $[s,[t]+1]$ ($[t]$ indicates the integer part of t).

We proceed with our first result.

Proposition 13.1. *Let $T > 0$ be fixed, and $\{X_t,\ t \geq s\}$ be the solution to* (13.1) *under Hypothesis 13.1. Let $p \geq 2$. We have:*

$$\mathbb{E}\left[\sup_{s\leq\theta\leq t}\left|\int_s^\theta \sigma(X_r)\,\mathrm{d}B_r\right|^p\right] \leq c(p,T)\mathbb{E}\left[\int_s^t |\sigma(X_r)|^p\ \mathrm{d}r\right], \qquad t > s, \tag{13.3}$$

for a constant $c = c(p,T)$ which depends only on p and T.

This proposition is a technical, yet important result; we shall sketch the proof in the appendix.

Let us define $\mathcal{S}(X) = \{\mathcal{S}(X)_t,\ t \in [s,T]\}$ the right-hand side of (13.1), for $X \in L^2(\Omega; C(s,T;\mathbb{R}^n))$. The solution of the equation can be obtained as the fixed point $X = \mathcal{S}(X)$ from the Picard approximation scheme $X_{n+1} = \mathcal{S}(X_n)$. Using Proposition 13.1, we estimate the distance between $\mathcal{S}(X)$ and $\mathcal{S}(Y)$ for different X and $Y \in L^2(\Omega; C(s,T;\mathbb{R}^n))$.

Fix any $p \geq 2$ and assume that the initial condition in (13.1) belongs to $L^p(\Omega;\mathbb{R}^n)$. Then

$$\begin{aligned}\mathbb{E}\left[\sup_{s\leq\theta\leq t}|\mathcal{S}(X)_\theta - \mathcal{S}(Y)_\theta|^p\right] \leq 3^{p-1}\,\mathbb{E}|X_s - Y_s|^p + 3^{p-1}\,\mathbb{E}\left[\sup_{s\leq\theta\leq t}\left|\int_s^\theta [b(X_r) - b(Y_r)]\,\mathrm{d}r\right|^p\right] \\ + 3^{p-1}\,\mathbb{E}\left[\sup_{s\leq\theta\leq t}\left|\int_s^\theta [\sigma(X_r) - \sigma(Y_r)]\,\mathrm{d}B_r\right|^p\right]\end{aligned}$$

The stochastic integral is bounded, using (13.3), by

$$c(p,T)\,\mathbb{E}\left[\int_s^t |\sigma(X_r) - \sigma(Y_r)|^p\ \mathrm{d}r\right];$$

we apply Hölder inequality in the deterministic integral and we get

$$\begin{aligned}\mathbb{E}\left[\sup_{s\leq\theta\leq t}\left|\int_s^\theta [b(X_r) - b(Y_r)]\,\mathrm{d}r\right|^p\right] &\leq \mathbb{E}\left[\sup_{s\leq\theta\leq t}\left|(\theta-s)^{p-1}\int_s^\theta |b(X_r) - b(Y_r)|^p\,\mathrm{d}r\right|\right] \\ &\leq (T-s)^{p-1}\,\mathbb{E}\left[\int_s^t |b(X_r) - b(Y_r)|^p\,\mathrm{d}r\right].\end{aligned}$$

Now, recalling that b and σ are Lipschitz continuous functions, with Lipschitz constant K, we obtain the following estimate

$$\mathbb{E}\left[\sup_{s\le\theta\le t}|\mathcal{S}(X)_\theta - \mathcal{S}(Y)_\theta|^p\right] \le 3^{p-1}\,\mathbb{E}|X_s - Y_s|^p + c(p,T,K)\,\mathbb{E}\left[\int_s^t |X_r - Y_r|^p\,\mathrm{d}r\right]. \tag{13.4}$$

Below, we state a useful consequences of (13.4), i.e., we estimate the dependence on the initial conditions. With the usual convention, the constant C may vary from line to line; we write $C = C(a,b,c)$ in order to emphasize the dependence on the parameters listed in brackets.

Proposition 13.2. *Assume Hypothesis 13.1 and take* $\eta,\,\gamma \in L^p(\Omega,\mathcal{F}_s;\mathbb{R}^n)$, $p\ge 2$. *Then, for* $0\le s\le t\le T$, *there exists a constant* $C = C(T,K,p) > 0$ *such that*

$$\mathbb{E}\left[\sup_{s\le\theta\le t}|X_\theta^{s,\eta} - X_\theta^{s,\gamma}|^p\right] \le \mathrm{e}^{C\,(t-s)}\,\mathbb{E}|\eta-\gamma|^p. \tag{13.5}$$

Proof. We set $X_t = X_t^{s,\eta}$ and $Y_t = X_t^{s,\gamma}$; the proof uses Gronwall's lemma and estimate (13.4). We define $\phi(t) = \mathbb{E}\left[\sup_{s\le\theta\le t}|X_\theta - Y_\theta|^p\right]$ and we find from (13.4)

$$\phi(t) \le 3^{p-1}\,\mathbb{E}|\eta-\gamma|^p + C\int_s^t \phi(\theta)\,\mathrm{d}\theta, \qquad T\ge t\ge s$$

so the thesis follows from an application of Gronwall's lemma:

$$\phi(t) \le \mathrm{e}^{C)\,(t-s)}\,\mathbb{E}|\eta-\gamma|^p.$$

□

The following result will be a key step in order to prove Feller property for the Markov semigroup associated to (13.1).

Corollary 13.3. *Assume Hypothesis 13.1 and take* $x, y \in \mathbb{R}^n$. *Then, for any* $p\ge 2$ *and* $0\le s\le t\le T$, *there exists a constant* $C = C(p,T,K)$ *such that*

$$\mathbb{E}\left[\sup_{s\le\theta\le t}|X_\theta^{s,x} - X_\theta^{s,y}|^p\right] \le C|x-y|^p. \tag{13.6}$$

In Lecture 2 we stated a version of Kolmogorov continuity theorem 2.16 for the case of a stochastic process on $T = (0,1)$. Now we need an extension of that result, stated in terms of a random field $\{X^x,\ x\in\mathbb{R}^m\}$. The proof follows exactly the one for the 1-dimensional case, and will not be given.

Theorem 13.4. *Let* $\{X^x,\ x\in\mathbb{R}^m\}$ *be a family of random variables on* $(\Omega,\mathcal{F},\mathbb{P})$ *with values in* $\mathbb{R}^n$. *Suppose that there exist positive constants a, b and c such that*

$$\mathbb{E}[|X^x - X^y|^b] \le c|x-y|^{m+a}, \qquad x,\,y\in\mathbb{R}^m.$$

Then, there exists a random field $\{Y^x,\ x\in\mathbb{R}^m\}$ *such that* $Y^x = X^x$, *a.s.,* $x\in\mathbb{R}^m$, *and the mapping:* $x\mapsto Y^x(\omega)$, *is* γ*-Hölder continuous on each compact set of* $\mathbb{R}^m$, *for any* $\gamma < \frac{a}{b}$, ω*-a.s.*

Now we are able to state our result concerning continuous dependence on the initial condition. With no loss of generality, we fix the initial condition at time $s = 0$ and we denote $\{X_t^x\}$ the solution of (13.1) with initial condition $X_0^x = x$. Applying Corollary 13.3 for $p > n + 1$ and the previous Kolmogorov regularity theorem, we obtain the following.

Theorem 13.5. *Under Hypothesis 13.1, there exists a process $Y = \{Y_t^x,\ t \geq 0\}$, $x \in \mathbb{R}^n$, such that*

1. *the mapping: $t \mapsto Y_t^x$ is continuous from $[0, \infty)$ into $\mathbb{R}^n$, for any $x \in \mathbb{R}^n$, a.s.;*
2. *the mapping: $x \mapsto Y_t^x$ is continuous from $\mathbb{R}^n$ into $\mathbb{R}^n$, for any $t \geq 0$, a.s.;*
3. *$Y_t^x = X_t^x$, a.s. for any $t \geq 0$ and $x \in \mathbb{R}^n$.*

Remark 13.6. With a little more effort it is possible to extend the above result to show continuous dependence of $X_t^{s,x}$ on all three parameters $s \leq t$ and $x \in \mathbb{R}^n$. For this, and more refined results on the regularity of $\{X_t^{s,x}\}$ we refer to the bibliography quoted in the introduction.

Remark 13.7. From now on we will always deal with the continuous version of the solution $\{X_t^{s,x}\}$ whose existence is proved in Theorem 13.5.

Let $X^{s,\eta}$ be a solution of (13.1) with initial condition $\eta \in L^2(\Omega, \mathcal{F}_s; \mathbb{R}^n)$. Next result shows how each given path of the solution depends on the initial condition.

Lemma 13.8. *Assume Hypothesis 13.1 and take the initial condition $\eta \in L^2(\Omega, \mathcal{F}_s; \mathbb{R}^n)$. Then, for $0 \leq s \leq t$, we have, ω a.s.,*

$$X_t^{s,\eta}(\omega) = X_t^{s,\eta(\omega)}(\omega).$$

Proof. Let us first assume that η takes only a finite number of values $y_1, \ldots, y_m \in \mathbb{R}^n$, i.e.,

$$\eta = \sum_{i=1}^m \mathbb{1}_{\{\eta = y_i\}}\, y_i.$$

Then we have

$$X_t^{s,\eta} \mathbb{1}_{\{\eta = y_i\}} = X_t^{s,y_i} \mathbb{1}_{\{\eta = y_i\}}.$$

Indeed, the process on the left hand side is

$$\begin{aligned} X_t^{s,\eta}\, \mathbb{1}_{\{\eta = y_i\}} &= \mathbb{1}_{\{\eta = y_i\}} \left(\eta + \int_s^t b(X_r^{s,\eta})\, \mathrm{d}r + \int_s^t \sigma(X_r^{s,\eta})\, \mathrm{d}B_r \right) \\ &= \mathbb{1}_{\{\eta = y_i\}} \left(y_i + \int_s^t b(X_r^{s,\eta})\, \mathrm{d}r + \int_s^t \sigma(X_r^{s,\eta})\, \mathrm{d}B_r \right) \end{aligned}$$

Let $A_i = \{\eta = y_i\}$. Now if we work on the probability space $(A_i, \mathbb{P}|_{A_i})$ we find easily by uniqueness (using the Lipschitz continuity of coefficients) that $X_t^{s,\eta}\, \mathbb{1}_{\{\eta = y_i\}} = X_t^{s,y_i}\, \mathbb{1}_{\{\eta = y_i\}}$; note in particular that

$$\int_s^t \sigma(X_r^{s,\eta})\, \mathrm{d}B_r\, (\omega) = \int_s^t \sigma(X_r^{s,\eta} \mathbb{1}_{\{\eta = y_i\}})\, \mathrm{d}B_r\, (\omega), \qquad \omega \in A_i,$$

Since $X_t^{s,y_i}(\omega) = X_t^{s,\eta(\omega)}(\omega)$, $\omega \in A_i$, we have the assertion.

To treat the general case, we first approximate η by a sequence of random variables $\{\eta_m\} \subset L^2(\Omega, \mathcal{F}_s; \mathbb{R}^n)$ taking only a finite number of values and such that $\mathbb{E}|\eta_m - \eta|^2 \to 0$, as $m \to \infty$, and $\lim_{m\to\infty} \eta_m(\omega) = \eta(\omega)$, ω a.s.. Then we pass to the limit as $m \to \infty$ in

$$X_t^{s,\eta_m}(\omega) = X_t^{s,\eta_m(\omega)}(\omega).$$

Note that $\lim_{m\to\infty} X_t^{s,\eta_m(\omega)}(\omega) = X_t^{s,\eta(\omega)}(\omega)$, a.s., using that $x \mapsto X_t^{s,x}(\omega)$ is continuous. Moreover, by Proposition 13.2, we know that, possibly passing to a subsequence, $\lim_{m\to\infty} X_t^{s,\eta_m}(\omega) = X_t^{s,\eta}(\omega)$, ω a.s.. The proof is complete.

□

We can now prove the following important theorem which defines the *stochastic flow* property of the solution of (13.1).

Theorem 13.9. *Assume Hypothesis 13.1 and fix an initial condition $x \in \mathbb{R}^n$. Then, for $0 \le s \le r \le t$, we have, ω a.s.,*

$$X_t^{s,x} = X_t^{r,X_r^{s,x}}. \tag{13.7}$$

Proof. We fix $r \ge s$ and introduce $Z_t = X_t^{r,X_r^{s,x}}$, $t \ge r$. The process $\{Z_t,\ t \ge r\}$ solves

$$\begin{aligned} Z_t &= X_r^{s,x} + \int_r^t b(Z_u)\,\mathrm{d}u + \int_r^t \sigma(Z_u)\,\mathrm{d}B_u \\ &= x + \int_s^r b(X_u^{s,x})\,\mathrm{d}u + \int_s^r \sigma(X_u^{s,x})\,\mathrm{d}B_u + \int_r^t b(Z_u)\,\mathrm{d}u + \int_r^t \sigma(Z_u)\mathrm{d}B_u. \end{aligned}$$

If in the previous formula, we replace Z_u with $X_u^{s,x}$, for $u \ge r$, we get

$$Z_t = x + \int_s^t b(X_u^{s,x})\,\mathrm{d}u + \int_s^t \sigma(X_u^{s,x})\,\mathrm{d}B_u, \ \ t \ge r.$$

It follows that $Z_t = X_t^{s,x}$, $t \ge r$, and the assertion is proved.

□

13.2 The Markov property

Let us consider the process $X^x = \{X_t^x,\ t \ge 0\}$, which solves

$$X_t^x = X_t^{0,x} = x + \int_0^t b(X_r^x)\,\mathrm{d}r + \int_0^t \sigma(X_r^x)\,\mathrm{d}B_r, \qquad t \ge 0,\ x \in \mathbb{R}^n.$$

We state the Markov property for X^x: we introduce the candidate to be a transition probability function

$$p(t,x,\Gamma) = \mathbb{P}(X_t^x \in \Gamma)$$

and prove that it verifies the Markov property

$$p(s, X_t^x, \Gamma) = \mathbb{P}(X_{s+t}^x \in \Gamma \mid \mathcal{F}_t).$$

We state the Markov property in the integral form: for any $f \in M_b(\mathbb{R}^n)$, $x \in \mathbb{R}^n$, $t, s \geq 0$, it holds

$$\mathbb{E}[f(X_{t+s}^x) \mid \mathcal{F}_t] = u(s, X_t^x), \quad \text{where} \quad u(s,y) := \int_{\mathbb{R}^n} f(z) p(s, y, \mathrm{d}z), \; y \in \mathbb{R}^n. \tag{13.8}$$

The first remark is that it is enough to prove (13.8) when $f \in C_b(\mathbb{R}^n)$, i.e., f is continuous and bounded on $\mathbb{R}^n$, and the verification of this remark is left as exercise.

Problem 13.1. By using the Dynkin $\pi - \lambda$ lemma, shows that (13.8) holds for any $f \in M_b(\mathbb{R}^n)$ if and only if it holds for any $f \in C_b(\mathbb{R}^n)$.
Hint: define $\mathbb{H} = \{A \in B(\mathbb{R}^n)$ such that (13.8) holds when $f = \mathbb{1}_A\}$; show that $\mathbb{H}$ contains all closed sets and that is a λ-system; deduce that $\mathbb{H} = B(\mathbb{R}^n)$.

Theorem 13.10. *Assume Hypothesis 13.1. Then $X^x = \{X_t^x\}$ is a Markov process on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ with transition function*

$$p(t, x, A) = \mathbb{P}(X_t^x \in A), \quad A \in B(\mathbb{R}^n).$$

Proof. We fix $f \in C_b(\mathbb{R}^n)$, $t, s \geq 0$, $x \in \mathbb{R}^n$, and prove (13.8).

Using the stochastic flow property (13.7), we have $\mathbb{E}[f(X_{t+s}^x) \mid \mathcal{F}_t] = \mathbb{E}[f(X_{t+s}^{t,X_t^x}) \mid \mathcal{F}_t]$. Hence it remains to verify that

$$\mathbb{E}[f(X_{t+s}^{t,X_t^x}) \mid \mathcal{F}_t] = u(s, X_t^x). \tag{13.9}$$

The proof is divided into three steps: first, we prove that the above identity holds when $X_t^x = y$ is a deterministic point, then we proceed to the general case.

Step I. We have to compare the solutions $Y_s^y := X_{t+s}^{t,y}$ and X_s^y; by construction, they are the limits of the following Picard approximations

$$\begin{cases} X_s^{n+1} = \mathcal{S}(X^n)_s = y + \int_0^s b(X_r^n)\,\mathrm{d}r + \int_0^s \sigma(X_r^n)\,\mathrm{d}B_r \\ Y_s^{n+1} = \mathcal{S}'(X^n)_s = y + \int_0^s b(Y_r^n)\,\mathrm{d}r + \int_0^s \sigma(Y_r^n)\,\mathrm{d}B'_r, \quad n \geq 0, \\ X^0 \equiv y, \qquad Y^0 \equiv y, \end{cases} \tag{13.10}$$

where $B' = \{B'_s,\, s \geq 0\}$, $B'_r = B_{t+r} - B_t$, is again a standard d-dimensional Brownian motion which is independent from $\mathcal{F}_t$ (we consider B' as a d-dimensional Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to its natural completed filtration $\{F'_t\}$). It is easily seen by induction that the random variable Y_s^n is independent from $\mathcal{F}_t$, for any $n \geq 0$, and passing to the limit, by the proof of Theorem 12.3 we know that

$$\lim_{m \to \infty} \mathbb{E}[\sup_{h \in [0, T-t]} |X_{t+h}^{t,y} - X_{t+h}^m|^2] = 0, \qquad \forall T > 0,$$

and we obtain that the random variable

$$X_{t+s}^{t,y} \quad \text{is independent from } \mathcal{F}_t, \text{ for any } s \geq 0.$$

Next, consider the two continuous processes $Y^1 = \{Y_s^1,\, s \geq 0\}$, $Y_s^1 := y + b(y)\,s + \sigma(y)\,B'_s$, which is adapted with respect to $\{\mathcal{F}'_t\}$, and and $X^1 = \{X_s^1,\, s \geq 0\}$, $X_s^1 = y + b(y)\,s + \sigma(y)\,B_s$, which is adapted with respect to $\{\mathcal{F}_t\}$. We note that the continuous processes (X^1, B) and (Y^1, B'), both with values in $\mathbb{R}^{n+d}$, have the same law (i.e., they have the same finite-dimensional distributions).

We claim that the same holds for any n, i.e., the continuous processes (X^n, B) and (Y^n, B') have the same law. This can be proved by induction, using the lemma below.

Since mean square convergence implies in particular convergence in law for random variables, passing to the limit as $n \to \infty$, we get in particular that the random variables

$$X^{t,y}_{t+s} \quad \text{and} \quad X^{0,y}_s = X^y_s \quad \text{have the same distribution.} \tag{13.11}$$

Lemma 13.11. *Let b and σ satisfies Hypothesis 13.1. Consider two stochastic bases $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ and $(\Omega, \mathcal{F}, \{\mathcal{F}'_t\}, \mathbb{P})$ on which are defined two d-dimensional Brownian motions, denoted respectively by B and B'. Fix any $T > 0$ and consider two continuous adapted processes $U = \{U_t,\, t \in [0,T]\}$ and $V = \{V_t,\, t \in [0,T]\} \in M^2(0,T;\mathbb{R}^n)$, defined respectively on the stochastic bases $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ and $(\Omega, \mathcal{F}, \{\mathcal{F}'_t\}, \mathbb{P})$. Assume that the processes (U, B) and (V, B') with values in $\mathbb{R}^{n+d}$ have the same law.*

Let X and Y be continuous processes with values in $\mathbb{R}^n$ given by

$$\begin{cases} X_s = \mathcal{S}(U)_s = y + \int_0^s b(U_r)\,\mathrm{d}r + \int_0^s \sigma(U_r)\,\mathrm{d}B_r \\ Y_s = \mathcal{S}(V)_s = y + \int_0^s b(V_r)\,\mathrm{d}r + \int_0^s \sigma(V_r)\,\mathrm{d}B'_r, \quad s \in [0,T], \end{cases} \tag{13.12}$$

Then also the processes (X, B) and (Y, B') have the same law.

Proof. We only briefly sketch the proof and refer to Section 5.4 of Durrett [Du96] for a complete proof.

At first, assume that U and V are elementary processes; then the claim follows in view of the definition of stochastic integral. In the general case, we first approximate U and V with elementary processes and then pass to the limit.

□

We continue with the proof of Theorem 13.10.

Step II. Step I and (13.11) actually show that the following sequence of identities hold:

$$\mathbb{E}[f(X^{t,y}_{t+s}) \mid \mathcal{F}_t] = \mathbb{E}[f(X^{t,y}_{t+s})] = \mathbb{E}[f(X^y_s)] = u(s,y), \qquad y \in \mathbb{R}^n. \tag{13.13}$$

Step III. We show that (13.13) holds also when the initial condition y is replaced by any random variable $Y \in L^2(\Omega, \mathcal{F}_t; \mathbb{R}^n)$, i.e.,

$$\mathbb{E}[f(X^{t,Y}_{t+s}) \mid \mathcal{F}_t] \;=\; u(s,Y) = \mathbb{E}[f(X^y_s)]_{y=Y}. \tag{13.14}$$

It is clear that from this claim we deduce in particular formula (13.11).

First assume that Y takes only a finite number of values $y_1, \ldots, y_m \in \mathbb{R}^n$, i.e., $Y = \sum_{i=1}^m \mathbb{1}_{\{Y=y_i\}}\, y_i$, where the family $\{Y = y_i\}$ is a partition of Ω contained in $\mathcal{F}_t$. We get, using the previous step and the independence of X^{t,y_i}_{t+s} from $\mathcal{F}_t$,

$$\mathbb{E}[f(X^{t,Y}_{t+s})] = \sum_{i=1}^m \mathbb{E}[f(X^{t,y_i}_{t+s})\mathbb{1}_{\{Y=y_i\}}] = \sum_{i=1}^m \mathbb{E}[u(s,y_i)]\,\mathbb{P}(A_i) = \mathbb{E}[u(s,Y)].$$

In the general case, take a sequence $(Y_n) \subset L^2(\Omega, \mathcal{F}_t; \mathbb{R}^n)$ of random variables, taking only a finite number of values, such that

$$\lim_{n\to\infty} \mathbb{E}|Y_n - Y|^2 = 0, \qquad \lim_{n\to\infty} Y_n = Y, \quad a.s.$$

By the continuity properties of $y \mapsto X_{t+s}^{t,y}$, we know that X_{t+s}^{t,Y_n} converges a.s. to $X_{t+s}^{t,Y}$ as $n \to \infty$. Hence, since $f \in C_b(\mathbb{R}^n)$, we know that

$$\lim_{n\to\infty} f(X_{t+s}^{t,Y_n}) = f(X_{t+s}^{t,Y}), \quad a.s.$$

We write

$$\mathbb{E}[f(X_{t+s}^{t,Y_n}) \mid \mathcal{F}_t] \;=\; u(s, Y_n) = \mathbb{E}[f(X_s^y)]_{y=Y_n}. \tag{13.15}$$

Passing to the limit as $n \to \infty$ in (13.15), using also the dominated convergence theorem for conditional expectation, we finally obtain the claim. The proof is complete.

□

13.2.1 Diffusion Markov semigroups

The semigroup of bounded linear operators $\{P_t,\, t \geq 0\}$, $P_t : M_b(\mathbb{R}^n) \to M_b(\mathbb{R}^n)$,

$$P_t f(x) = \mathbb{E}[f(X_t^x)], \;\; f \in M_b(\mathbb{R}^n),\, x \in \mathbb{R}^n \tag{13.16}$$

is called the *diffusion Markov semigroup* associated to $\{X_t^x\}$.

Using the continuity of the mappings $x \mapsto X_t^x(\omega)$ and $t \mapsto X_t^x(\omega)$ (ω-a.s.) together with the dominated convergence theorem, one easily proves

Proposition 13.12. *Assume Hypothesis 13.1. Then the diffusion Markov semigroup $\{P_t\}$ is Feller.*

Problem 13.2. Consider a problem with *additive noise*, i.e., assume that $\sigma(x) \equiv \sigma$ is constant, and assume that b is bounded and Lipschitz continuous on $\mathbb{R}^n$. Show that the diffusion semigroup $\{P_t\}$ is Feller-Dynkin, i.e., $C_0(\mathbb{R}^n)$ is invariant for $\{P_t\}$.

Let us compute the *generator* of $\{P_t\}$. To this purpose, let us introduce the second order (possibly degenerate) elliptic *Kolmogorov operator* L associated to (13.1)

$$Lf(x) = \frac{1}{2}\operatorname{Tr}(\sigma(x)\nabla^2 f(x)\sigma^*(x)) + \langle \nabla f(x), b(x)\rangle,\; f \in C_c^2(\mathbb{R}^n),\; x \in \mathbb{R}^n, \tag{13.17}$$

where $C_c^2(\mathbb{R}^n)$ denotes the space of all functions $f : \mathbb{R}^n \to \mathbb{R}$ having compact support and first and second continuous partial derivatives.

Introducing the symmetric non-negative matrix $q(x) = \sigma(x)\sigma^*(x)$, $x \in \mathbb{R}^n$, L can be written as

$$Lf(x) = \frac{1}{2}\operatorname{Tr}(q(x)\nabla^2 f(x)) + \langle \nabla f(x), b(x)\rangle = \frac{1}{2}\sum_{i,j=1}^n q_{ij}(x)\partial^2_{x_i x_j} f(x) + \sum_{i=1}^n b_i(x)\partial_{x_i} f(x), \tag{13.18}$$

for any $f \in C_c^2(\mathbb{R}^n)$, $x \in \mathbb{R}^n$.

Recall the definition of generator for the Feller semigroup $\{P_t\}$ (see also Lecture 6)

$$\begin{cases} D(A) = \big\{ f \in C_b(\mathbb{R}^n) \,:\, \text{there exists } g \in C_b(\mathbb{R}^n) \text{ such that } \sup\limits_{t>0} \frac{1}{t}\|P_t f - f\|_\infty < \infty \\ \qquad\qquad \text{and } \lim\limits_{t\to 0^+} \frac{1}{t}(P_t f(x) - f(x)) = g(x),\ x \in \mathbb{R}^n \big\}, \\ Af(x) = \lim\limits_{t\to 0^+} \frac{1}{t}(P_t f(x) - f(x)), \quad f \in D(A),\ x \in \mathbb{R}^n. \end{cases} \tag{13.19}$$

Proposition 13.13. *Assume Hypothesis 13.1 and consider the diffusion Markov semigroup $\{P_t\}$, see (13.16). Then $C_c^2(\mathbb{R}^n) \subset D(A)$ and, for any $f \in C_c^2(\mathbb{R}^n)$, $Af(x) = Lf(x)$, $x \in \mathbb{R}^n$.*

Proof. Take $f \in C_c^2(\mathbb{R}^n)$ and set $X_t = X_t^x$. By Itô formula we have

$$f(X_t(\omega)) - f(x) = \int_0^t \langle \nabla f(X_u(\omega)), b(X_u(\omega))\rangle\, \mathrm{d}u + \frac{1}{2}\int_0^t \Big(\sum_{i,j=1}^n q_{ij}(X_u(\omega))\partial^2_{x_i x_j} f(X_u(\omega))\Big)\, \mathrm{d}u$$
$$+ Y_t(\omega) = \int_0^t Lf(X_u(\omega))du + Y_t(\omega), \quad \text{a.s.},\ t \ge 0, \tag{13.20}$$

where $q(x) = \sigma(x)\sigma(x)^*$ and the process $\{Y_t\}$ is given by

$$Y_t = \sum_{i,j=1}^n \int_0^t \partial_{x_i} f(X_u)\sigma_{ij}(X_u) dB_u^j := \int_0^t \langle \nabla f(X_u), \sigma(X_u)\, \mathrm{d}B_u\rangle.$$

Since $\mathbb{E}[Y_t] = 0$, $t \ge 0$, applying expectation in (13.20), we get, for $t \ge 0$,

$$\mathbb{E}[f(X_t^x)] - f(x) = \int_0^t \mathbb{E}[Lf(X_u^x)]\, \mathrm{d}u.$$

Since Lf is bounded on $\mathbb{R}^n$, we get, for any $t > 0$, the upper bound

$$\sup_{x\in\mathbb{R}^n} \left| \frac{\mathbb{E}[f(X_t^x)] - f(x)}{t} \right| = \sup_{x\in\mathbb{R}^n} \left| \frac{P_t f(x) - f(x)}{t} \right| \le \|Lf\|_\infty.$$

Moreover, $\lim\limits_{t\to 0^+} \frac{\mathbb{E}[f(X_t^x)] - f(x)}{t} = Lf(x)$, $x \in \mathbb{R}^n$. The proof is complete.
□

13.3 Kolmogorov equations

In this section we discuss the relation between the diffusion Markov semigroup $\{P_t,\ t \ge 0\}$, see (13.16), and parabolic equations involving the Kolmogorov operator L introduced in (13.18). We do not develop the argument in a complete way; we only present some simple basic facts.

We always assume Hypothesis 13.1 and denote by $\{X_t^x\}$ the solution to (13.1). Hence

$$P_t f(x) = \mathbb{E}[f(X_t^x)], \qquad f \in M_b(\mathbb{R}^n).$$

We will also use the space $C_b^2(\mathbb{R}^n)$ of all functions $f : \mathbb{R}^n \to \mathbb{R}$ which are continuous and bounded on $\mathbb{R}^n$ together with its first and second partial derivatives.

In the following result we do not need to use the Markov property of $\{X_t^x\}$ but only the Itô formula.

Proposition 13.14. *Let $f \in C_b^2(\mathbb{R}^n)$. Then we have:*

$$\mathbb{E}(f(X_t^x)) = f(x) + \mathbb{E} \int_0^t Lf(X_s^x)\, \mathrm{d}s, \qquad t \geq 0, \tag{13.21}$$

and

$$\lim_{h \to 0} \frac{\mathbb{E}[f(X_{t+h}^x)] - \mathbb{E}[f(X_t^x)]}{h} = \mathbb{E}[Lf(X_t^x)], \qquad x \in \mathbb{R}^n,\ t \geq 0. \tag{13.22}$$

Proof. We simply apply the expectation in Itô formula (13.20). We only remark that the process $\{Y_t\}$ in (13.20) is well defined: indeed the estimate

$$|\sigma(X_u^x)^* \nabla f(X_u^x)|^2 \leq C \|\nabla f\|_\infty^2 (1 + |X_u^x|^2),$$

shows that the process $\{\sigma(X_u^x)^* \nabla f(X_u^x),\, u \geq 0\}$, $x \in \mathbb{R}^n$, belongs to $M^2(0, T; \mathbb{R}^n)$, for any $T > 0$.
□

Remark 13.15. The previous result admits the following PDEs-interpretation. Assume that there exists the density $p(t, x, y)$ of the law of the random variable X_t^x. According to Proposition 13.14, $p(t, x, y)$ verifies

$$\int_{\mathbb{R}^n} p(t, x, y) f(y)\, \mathrm{d}y = f(x) + \int_0^t \int_{\mathbb{R}^n} p(s, x, y) L_y f(y)\, \mathrm{d}y\, \mathrm{d}s \tag{13.23}$$

and we say that it is a generalized solution of the so-called *Fokker-Planck parabolic equation*:

$$\begin{cases} \partial_t p(t, x, y) = L_y^* p(t, x, y), & t > 0,\ x \in \mathbb{R}^n, \\ \lim_{t \downarrow 0} p(t, x, y) = \delta_x(y), & x \in \mathbb{R}^n. \end{cases} \tag{13.24}$$

This equation involves the formal adjoint L^* of L:

$$L_y^* \phi(y) = \frac{1}{2} \sum_{i,j=1}^n \partial_{y_i y_j}^2 \big(q_{ij}(y)\phi(y)\big) - \sum_{i=1}^n \partial_{y_i} \big(b_i(y)\phi(y)\big), \qquad \phi \in C_0^2(\mathbb{R}^n),$$

in distributional sense, compare with (13.23). For more details on Fokker-Planck equations, see for instance `Krylov` [Kr95]. The Fokker-Planck equation (13.24) is also called the *Kolmogorov forward equation.*

Let us consider now the following autonomous second order (possibly degenerate) *parabolic Cauchy problem*, involving the Kolmogorov operator L,

$$\begin{cases} \partial_t u(t, x) = Lu(t, x), & t > 0, \\ u(0, x) = f(x), & x \in \mathbb{R}^n, \end{cases} \tag{13.25}$$

where clearly L acts on the x_i-variables and $f \in C_b^2(\mathbb{R}^n)$ is the given initial condition.

We want to show that, under suitable assumptions, $u(t, x) := P_t f(x)$ provides the (unique) classical solution to (13.25).

Definition 13.16. *By a classical solution of* (13.25) *we mean a continuous and bounded function* $u : [0, +\infty) \times \mathbb{R}^n \to \mathbb{R}$ *which has continuous partial derivative* $\partial_t u$ *and continuous and bounded partial derivatives* $\partial_{x_i} u$ *and* $\partial^2_{x_i x_j} u$ *on* $(0, +\infty) \times \mathbb{R}^n$ *and solves* (13.25).

Theorem 13.17. *Assume Hypothesis 13.1 and consider the Cauchy problem* (13.25) *with* $f \in C_b^2(\mathbb{R}^n)$.

(i) (uniqueness) Let u *be a classical solution of* (13.25); *then* $u(t,x) := P_t f(x)$, *where* $\{P_t\}$ *is the diffusion semigroup associated to* (13.1).

(ii) (existence) Assume in addition the following property

$$P_t(C_b^2(\mathbb{R}^n)) \subset C_b^2(\mathbb{R}^n), \;\; t \geq 0. \tag{13.26}$$

Then, defining $v(t,x) := P_t f(x)$, v *is the classical solution to* (13.25).

Proof. (i) Let $X_t = X_t^x$. We fix $t > 0$, $x \in \mathbb{R}^n$, and apply Itô formula to the process $\{u(t-s, X_s),\; s \in [0,t]\}$. We have

$$\begin{aligned} u(0, X_t) &= u(t,x) - \int_0^t \partial_t u(t-r, X_r)\,\mathrm{d}r + \int_0^t Lu(t-r, X_r)\,\mathrm{d}r + \int_0^t \langle \nabla_x u(t-r, X_r), \sigma(X_r)\,\mathrm{d}B_r \rangle \\ &= u(t,x) + \int_0^t \langle \nabla_x u(t-r, X_r), \sigma(X_r)\,\mathrm{d}B_r \rangle, \end{aligned}$$

since u is a solution to (13.25). Applying expectation we infer

$$u(t,x) = \mathbb{E}[u(0, X_t)] = \mathbb{E}[f(X_t)] = P_t f(x), \qquad t \geq 0,\; x \in \mathbb{R}^n.$$

(ii) Set $X_t = X_t^x$. We apply Itô formula to the process $\{f(X_t),\, t \geq 0\}$ to find

$$f(X_t) = f(x) + \int_0^t Lf(X_r)\mathrm{d}r + \int_0^t \langle \nabla f(X_r), \sigma(X_r)\,\mathrm{d}B_r \rangle.$$

Taking expectation, we get

$$\mathbb{E}f(X_{t+h}) - \mathbb{E}f(X_t) = \int_t^{t+h} \mathbb{E}[Lf(X_r)]\,\mathrm{d}r, \qquad h \geq 0.$$

Hence, there exists, for any $t > 0$,

$$\partial_t v(t,x) = \lim_{h \to 0} \frac{P_{t+h} f(x) - P_t f(x)}{h} = \mathbb{E}[Lf(X_t^x)] = P_t(Lf)(x). \tag{13.27}$$

Note that the last formula holds also when $t = 0$ if we consider the limit for $h \to 0^+$.

To compute the previous limit, we have used that the mapping: $s \mapsto \mathbb{E}[Lf(X_s)]$ is continuous on $[0,T]$. Such continuity follows applying the dominated convergence theorem thanks to the estimate

$$\sup_{s \in [0,T]} |Lf(X_s)| \leq C(1 + \sup_{s \in [0,T]} |X_s|), \qquad a.s..$$

Now we use Hypothesis (13.26), i.e., $P_t f \in C_b^2(\mathbb{R}^n)$, and the semigroup property in order to obtain

$$\partial_t v(t,x) = \lim_{h\to 0^+} \frac{P_h\big(P_t f\big)(x) - P_t f(x)}{h} = LP_t f(x), \qquad t>0,\ x\in\mathbb{R}^n. \tag{13.28}$$

We have just showed that

$$\begin{cases} \partial_t (P_t f)(x) = L\big(P_t f\big)(x) = P_t\big(Lf\big)(x), & t>0, \\ P_0 f(x) = f(x), & x\in\mathbb{R}^n. \end{cases}$$

The proof is complete.

□

Remark 13.18. Let us consider a PDEs-interpretation of Theorem 13.17.

Assume that the law $p(t,x,A) = \mathbb{P}(X_t^x \in A)$, $A\in B(\mathbb{R}^n)$, has a density $p(t,x,y)$ with respect to the Lebesgue measure (this is well known if we assume in addition that there exists $\nu>0$ such that $\langle q(x)h,h\rangle \geq \nu|h|^2$, $x,h\in\mathbb{R}^n$). In this case the assertion *(ii)* says that $p(t,x,y)$ is the *fundamental solution* to the parabolic equation (13.25).

The previous results are the starting points for the study of regularity for solutions of parabolic Kolmogorov equations by probabilistic methods (one important reference on this subject is the book by Krylov [Kr95]; see also the books of Nualart [Nu95] and Bass [Bas95].

Note that one can start from the Cauchy problem (13.25), where the operator L is given, for any $f\in C_c^2(\mathbb{R}^n)$, by the formula

$$Lf(x) = \frac{1}{2}\operatorname{Tr}(q(x)\nabla^2 f(x)) + \langle \nabla f(x), b(x)\rangle = \frac{1}{2}\sum_{i,j=1}^n q_{ij}(x)\partial^2_{x_i x_j} f(x) + \sum_{i=1}^n b_i(x)\partial_{x_i} f(x)$$

($q(x)$ is a real $n\times n$ symmetric non-negative matrix). Assume that $b:\mathbb{R}^n\to\mathbb{R}^n$ is Lipschitz continuous and assume also that $q_{ij}\in C^2(\mathbb{R}^n)$ with first and second bounded derivatives on $\mathbb{R}^n$.

Under such assumptions, one can show (see for instance Stroock and Varadhan [SV79]) that there exists a unique real $n\times n$ symmetric non-negative matrix $\sigma(x)$ such that

$$\sigma(x)^2 = q(x),\ \ x\in\mathbb{R}^n, \quad \text{and the mapping: } x\mapsto\sigma(x) \text{ is Lipschitz continuous on } \mathbb{R}^n.$$

Hence, using such $\sigma(x)$, we can associate to L a diffusion process which solves a SDE, involving b and σ, and a related diffusion semigroup $\{P_t\}$. It is clear that regularity properties of $\{P_t\}$ will give regularity properties for the Kolmogorov equation (13.25).

Problem 13.3. Here we require to check the crucial condition (13.26) of Theorem 13.17 in a special case, i.e., when there is *additive noise*. To this purpose, assume that $\sigma(x)\equiv\sigma$ is constant, that $b\in C^2(\mathbb{R}^n,\mathbb{R}^n)$ is Lipschitz continuous on $\mathbb{R}^n$ and show that

$$P_t(C_b^2(\mathbb{R}^n)) \subset C_b^2(\mathbb{R}^n), \ \ t\geq 0. \tag{13.29}$$

Hint: Step I. Denote by $X^x = \{X_t^x,\ t\in[0,T]\}$ the solution of (13.1) on $[0,T]$, with $s=0$. Fix $\omega\in\Omega$, a.s., and introduce the Banach space $H = C([0,T];\mathbb{R}^n)$. Define the mapping $F:\mathbb{R}^n\times H\to H$,

$$F(x,Y)(t) := Y(t) - x - \int_0^t b(Y(r))\,\mathrm{d}r - \sigma B(t), \quad t\in[0,T],\ Y\in H.$$

Prove by the implicit function theorem that, for any ω a.s., the mapping $x \mapsto X^x$ from $\mathbb{R}^n$ into H is twice Frechet-differentiable. Denote by η^i_t and by η^{ij}_t, $t \in [0,T]$, respectively the first derivative at $x \in \mathbb{R}^n$ in the direction e_i and the second derivative at x in the directions e_i and e_j, $i,j = 1,\dots,n$ (we have not indicated the dependence on x for simplicity of notation).

Step II. Show that both η^i_t and η^{ij}_t are solutions of SDEs (which are called the first and second variation equation of (13.1)). Deduce that $\{\eta^i_t,\, t \in [0,T]\}$ and $\{\eta^{ij}_t,\, t \in [0,T]\}$ are stochastic processes defined and adapted on the stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ with values in $\mathbb{R}^n$. Moreover $\{\eta^i_t,\, t \in [0,T]\}$ and $\{\eta^{ij}_t,\, t \in [0,T]\}$ $\in L^\infty([0,T] \times \Omega; \mathbb{R}^n)$ uniformly in $x \in \mathbb{R}^n$.

Step III. Verify that, for any $f \in C^2_b(\mathbb{R}^n)$, one has:

$$\partial_{x_i} P_t f(x) = \mathbb{E}[\langle \nabla f(X^x_t), \eta^i_t\rangle], \quad \partial^2_{x_i x_j} P_t f(x) = \mathbb{E}\big[\langle \nabla^2 f(X^x_t)[\eta^j_t], \eta^i_t\rangle \,+\, \langle \nabla f(X^x_t), \eta^{ij}_t\rangle\big],\ t \in [0,T].$$

Step IV. Deduce (13.29).

Appendix: Proof of Proposition 13.1

We give a proof assuming $s = 0$ and $\mathbb{R}^n = \mathbb{R}$ for simplicity. We are concerned with the stochastic integral of a process $X \in M^2(0,T)$; our aim is to estimate the quantity (13.3)

$$\mathbb{E}\left[\sup_{\theta \le t} \left| \int_0^\theta \sigma(X_r)\,\mathrm{d}B_r \right|^p\right] \le c(p,T)\mathbb{E}\left[\int_0^t |\sigma(X_r)|^p \,\mathrm{d}r\right], \quad t \le T,\ p \ge 2. \tag{13.30}$$

This estimate is a version of the celebrated Burkholder-Davis-Gundy inequality, which can be proved, in larger generality, for martingales: see for instance `Kallenberg` [Ka02, Corollary 17.7].

If we consider Theorem 4.24 and in particular (4.9), we get

$$\mathbb{E}\left[\sup_{\theta \le t} \left| \int_0^\theta \sigma(X_r)\,\mathrm{d}B_r \right|^p\right] \le \left(\frac{p}{p-1}\right)^p \sup_{\theta \le t} \mathbb{E}\left| \int_0^\theta \sigma(X_r)\,\mathrm{d}B_r \right|^p.$$

Let $I_t = \int_0^t \sigma(X_r)\,\mathrm{d}B_r$, $t \in [0,T]$. By a localization procedure via stopping times, possibly replacing I_t with $I_{t\wedge\tau_N}$, where $\tau_N = \inf\{t \ge 0,\colon\ |I_t| > N\}$, $N \in \mathbb{N}$, we may assume, that $|I_t(\omega)| \le K$ for any $t \in [0,T]$, ω-a.s.

Note that $x \mapsto |x|^p$ is a convex function, hence $|I(X)_t|^p$ is a submartingale and so $t \mapsto \mathbb{E}|I(X)_t|^p$ is increasing; it follows that the right-hand side of the last formula is estimated by the quantity

$$\mathbb{E}\left[\left| \int_0^t \sigma(X_r)\,\mathrm{d}B_r \right|^p\right].$$

In case $p = 2$, the Itô isometry formula implies that this quantity is equal to

$$\mathbb{E}\left[\int_0^t |\sigma(X_r)|^2 \,\mathrm{d}r\right],$$

which proves that

$$\mathbb{E}\left[\sup_{\theta\le t}\left|\int_0^\theta \sigma(X_r)\,\mathrm{d}B_r\right|^2\right] \le 4\mathbb{E}\left[\int_0^t |\sigma(X_r)|^2\,\mathrm{d}r\right]. \tag{13.31}$$

In order to prove the result for $p > 2$, we shall use Itô formula. Take $f(x) = |x|^p$; if we apply Itô formula to the process $\{|I_t|^p,\ t \in [0,T]\}$, we get

$$|I_t|^p = p\int_0^t |I_s|^{p-1}\,\mathrm{sign}(I_s)\sigma(X_s)\,\mathrm{d}B_s + \frac{1}{2}p(p-1)\int_0^t |I_s|^{p-2}|\sigma(X_s)|^2\,\mathrm{d}s.$$

In our assumptions, the process $\{|I_s|,\ s \in [0,T]\}$ is bounded and the stochastic integral is well defined and has zero mean, hence

$$\mathbb{E}[|I_t|^p] = c(p)\mathbb{E}\left[\int_0^t |I_s|^{p-2}|\sigma(X_s)|^2\,\mathrm{d}s\right].$$

A first application of Hölder inequality implies

$$\mathbb{E}[|I_t|^p] \le c(p)\,\mathbb{E}\left[\left(\int_0^t |I_s|^p\,\mathrm{d}s\right)^{(p-2)/p}\left(\int_0^t |\sigma(X_s)|^p\,\mathrm{d}s\right)^{2/p}\right]$$

and next, using again the same inequality (this time with respect to the mean),

$$\mathbb{E}[|I_t|^p] \le c(p)\left(\mathbb{E}\left[\int_0^t |I_s|^p\,\mathrm{d}s\right]\right)^{(p-2)/p}\left(\mathbb{E}\left[\int_0^t |\sigma(X_s)|^p\,\mathrm{d}s\right]\right)^{2/p}.$$

We already know that the mapping $t \mapsto \mathbb{E}[|I_t|^p]$ is increasing, so we can estimate the first term in the right hand side with $(T\,\mathbb{E}[|I_t|^p])^{(p-2)/p}$ and, simplifying both sides, we obtain the thesis.

14

Malliavin calculus and stochastic equations with additive noise

This lecture is dedicated to show some applications of Malliavin calculus to the study of regularity of transition probability functions associated to solutions of SDEs. We have confined ourselves to 1-dimensional stochastic differential equations with additive noise. This choice implies a sensible simplification of the exposition; on the other hand, we are missing many interesting aspects of the multidimensional case (like the case of stochastic equations with degenerate noise). We refer to the existing literature on the subject, and in particular to the monograph by `Nualart` [Nu95], for a complete introduction to Malliavin calculus; we also mention that the second part of this lecture follows the notes of `Da Prato` [Da05].

Let us start by mentioning an unexpected relation between the so-called Skorohod integral and Malliavin calculus.

The theory of Skorohod stochastic integral has been developed from the need to give a meaning to (stochastic) integrals of processes that are not necessarily adapted to the natural filtration generated by the Brownian motion with respect to which we integrate. There is a surprising and beautiful relation between this theory and Malliavin calculus, namely the Gaveau-Trauber theorem [GT82]. In this theorem it is proved that the adjoint of the Malliavin derivative is equivalent to the definition of stochastic integral with respect to a generalized Gaussian process introduced by `Skorohod` [Sk75] (based on the chaos expansions in terms of multiple Wiener functionals).

Further, in this lecture we also meet the backward Itô integral, where the stochastic processes to be integrated are adapted to the future filtration of the Brownian motion. For a justification of the interest of this integral, we refer for instance to `Da Prato` [Da98].

14.1 The Skorohod integral

Throughout the lecture we fix the time interval $[0,1]$ (the extension to the case $[0,T]$ is straightforward) and let $H = L^2(0,1)$; we consider, on a Gaussian probability space $(\Omega, \mathcal{F}, \mathbb{P}, H)$, a real valued Brownian motion $B = \{B_t,\ t \in [0,1]\}$. We refer to Lecture 9 for basic definitions and notation on Malliavin calculus.

Let $\{\mathcal{F}_t,\ t \in [0,1]\}$, be the natural completed (right continuous) filtration associated with the Brownian motion B. Recall that $M^2(0,1)$ is the space of all progressively measurable stochastic processes $u : [0,1] \times \Omega \to \mathbb{R}$ such that

$$\mathbb{E}\left[\int_0^1 u^2(s)\,\mathrm{d}s\right] < \infty.$$

Next we define the *adjoint operator* δ of the *Malliavin derivative* D. Recall that the operator D is a closed and unbounded operator with values in $L^2([0,1] \times \Omega)$ defined on the dense subset $\mathbb{D}^{1,2}$ of $L^2(\Omega)$.

Definition 14.1. *Let $u \in L^2([0,1] \times \Omega; \mathbb{R})$; we say that $u \in \mathrm{Dom}\delta$ if there exists a positive constant c (possibly depending on u) such that*

$$\left|\mathbb{E}\int_0^1 D_s F\, u(s)\, ds\right| \leq c\, \|F\|_2 = c\,(\mathbb{E}|F|^2)^{1/2}$$

for any smooth real valued random variable $F \in \mathcal{S}$.

For $u \in \mathrm{Dom}\delta$ we define $\delta(u)$ the unique element in $L^2(\Omega)$ such that, for any $F \in \mathbb{D}^{1,2}$:

$$\mathbb{E}[F\,\delta(u)] = \mathbb{E}\int_0^1 D_s F\, u(s)\,\mathrm{d}s. \tag{14.1}$$

The operator δ is closed since it is an adjoint operator. It is usually called the *Skorohod integral* of the process u and we will use the notation

$$\delta(u) := \int_0^1 u(s)\,\mathrm{d}_s B(s).$$

Lemma 14.2. *Let F be a smooth random variable in $\mathcal{S}$ and $f \in H = L^2(0,1)$. Then the stochastic process $u(t,\omega) = F(\omega) f(t)$ belongs to $\mathrm{Dom}\delta$ and it holds*

$$\delta(u) = -\int_0^1 D_t F\, f(t)\,\mathrm{d}t + W(f)\,F. \tag{14.2}$$

Proof. From the integration-by-parts formula (9.11) we have, for any smooth random variable G:

$$\mathbb{E}[F \int_0^1 D_t G\, f(t)\,\mathrm{d}t] = -\mathbb{E}[G \int_0^1 D_t F\, f(t)\,\mathrm{d}t] + \mathbb{E}[F\,G\,W(f)]$$

and (14.1) leads to the thesis.

□

We have thus a way to introduce a large class of stochastic processes belonging to $\mathrm{Dom}\delta$. We shall define $\mathcal{S}(H)$ the class of smooth stochastic processes of the form

$$\sum_{j=1}^n F_j\, f_j(t), \qquad t \in [0,1],$$

where $F_j \in \mathcal{S}$ (recall the definition of this space in Lecture 9), $f_j \in H = L^2(0,1)$.

There is a commutativity relation between Malliavin derivative and Skorohod integral, which is often useful in applications.

Proposition 14.3. *Suppose that $u \in \mathcal{S}(H)$ is a smooth stochastic process and $h \in H = L^2(0,1)$. Then the Skorohod integral $\delta(u)$ belongs to $\mathbb{D}^{1,2}$ and we have*

$$\langle h, D(\delta(u))\rangle_H = \langle u, h\rangle_H + \int_0^1 \int_0^1 D_t u(s) h(t)\,\mathrm{d}t\,\mathrm{d}_S B(s). \tag{14.3}$$

Proof. Let $u \in \mathcal{S}(H)$ be given by $u(t) = F\,f(t)$, with $F \in \mathcal{S}$ and $f \in H$. Denote by $D^h G = \langle DG, h\rangle_H$ the Malliavin derivative in the direction h. Then it holds

$$\begin{aligned}
\int_0^1 D_t\delta(u)\,h(t)\,\mathrm{d}t &= \int_0^1 D_t[FW(f)]\,h(t)\,\mathrm{d}t - \int_0^1 D_t\Big[\int_0^1 D_s F\,f(s)\,\mathrm{d}s\Big] h(t)\,\mathrm{d}t \\
&= \int_0^1 (D_t F)\,W(f)\,h(t)\,\mathrm{d}t + \int_0^1 F f(t)\,h(t)\,\mathrm{d}t - \int_0^1 D_s\Big[\int_0^1 D_t F\,h(t)\,\mathrm{d}t\Big] f(s)\,\mathrm{d}s \\
&= \int_0^1 u(t)\,h(t)\,\mathrm{d}t + D^h F W(f) - \langle D(D^h F), f\rangle_H \\
&= \langle u, h\rangle_H + \delta(f D^h F)
\end{aligned}$$

and this implies the thesis.

□

Remark 14.4. When we apply formula (14.3) to $h = \mathbb{1}_{[0,t]}$ we obtain the relation

$$\int_0^t D_s\delta(u)\,\mathrm{d}s = \int_0^t u(s)\,\mathrm{d}s + \int_0^1 \int_0^t D_\tau u(s)\,\mathrm{d}\tau\,\mathrm{d}_S B(s);$$

hence, we also obtain, for almost every $t \in [0,1]$,

$$D_t(\delta(u)) = u(t) + \int_0^1 D_t u(\tau)\,\mathrm{d}_S B(\tau).$$

Let us consider two smooth processes $u, v \in \mathcal{S}(H)$; starting with the duality relationship (14.1) we have

$$\mathbb{E}[\delta(u)\,\delta(v)] = \mathbb{E}\int_0^1 u(t)\,D_t\delta(v)\,\mathrm{d}t$$

and the previous remark implies

$$\mathbb{E}[\delta(u)\,\delta(v)] = \mathbb{E}\int_0^1 u(t)\,v(t)\,\mathrm{d}t + \mathbb{E}\int_0^1 u(t)\delta(D_t v)\,\mathrm{d}t.$$

Using again the duality between the operators δ and D we obtain

$$\mathbb{E}[\delta(u)\,\delta(v)] = \mathbb{E}\int_0^1 u(t)\,v(t)\,\mathrm{d}t + \mathbb{E}\int_0^1\int_0^1 D_s u(t)\,D_t v(s)\,\mathrm{d}s\,\mathrm{d}t.$$

As a consequence we get the following estimate for every process in $\mathcal{S}(H)$:

$$\mathbb{E}[\delta(u)^2] \le \mathbb{E}\left[|u|_H^2 + |Du|_{H\otimes H}^2\right]. \tag{14.4}$$

Define $\mathbb{L}^{1,2}$ as the closure of $\mathcal{S}(H)$ in $L^2([0,1]\times\Omega;\mathbb{R})$ with respect to the norm

$$\|u\|_{1,2}^2 = \mathbb{E}\int_0^1 |u(t)|^2\,\mathrm{d}t + \mathbb{E}\int_0^1\int_0^1 (D_s u(t))^2\,\mathrm{d}s\,\mathrm{d}t. \tag{14.5}$$

For any $u \in \mathbb{L}^{1,2}$ there exists a sequence $\{u_n\}$ in $\mathcal{S}(H)$ such that u_n converges to u in $L^2(\Omega;H)$ and Du_n converges to Du in $L^2(\Omega; H\otimes H)$. Therefore, $\delta(u_n)$ is a Cauchy sequence in $L^2(\Omega)$ (use estimate (14.4)) and the closure of the operator δ imply that its limit is $\delta(u)$, hence $\mathbb{L}^{1,2}$ is contained in Domδ. Note that $\mathbb{L}^{1,2}$ is an Hilbert space with respect to the inner product induced by the norm (14.5).

The proposition below is called “isometry formula” and provides an important tool in Malliavin calculus.

Proposition 14.5. *Suppose that u and v are two processes in the space $\mathbb{L}^{1,2}$. Then we have*

$$\mathbb{E}[\delta(u)\,\delta(v)] = \mathbb{E}\int_0^1 u(t)\,v(t)\,\mathrm{d}t + \mathbb{E}\int_0^1\int_0^1 D_s u(t)\,D_t v(s)\,\mathrm{d}s\,\mathrm{d}t. \tag{14.6}$$

Suppose that both processes are adapted to the filtration: by Remark 9.8 we have that $D_s u(t) = 0$ for almost all $s > t$; consequently, the second term in the right hand side of (14.6) is zero, and we recover the usual isometry property of the Itô integral.

Proof. We can assume by a density argument that the processes u and v belong to $\mathcal{S}(H)$; let us write $u_t = Ff_t$ and $v_t = Gg_t$. In this case $\delta(u) = FW(f) - \langle DF, f\rangle_H$.

Using the duality relation (14.1) and formula (14.3) we obtain

$$\mathbb{E}[\delta(u)\delta(v)] = \mathbb{E}\left(\int_0^1 v(t) D_t\delta(u)\,\mathrm{d}t\right) = \mathbb{E}\left(\int_0^1 v(t)\left(u(t) + \int_0^1 D_t u(s)\,\mathrm{d}_s B(s)\right)\mathrm{d}t\right).$$

Finally, we get the thesis applying again the duality relation.

□

Altough it is not true that every adapted process belongs to $\mathbb{L}^{1,2}$, we can show that $M^2(0,1) \subset$ Domδ and that on $M^2(0,1)$, the operator δ is an extension of the Itô integral (with respect to the reference Brownian motion).

Theorem 14.6. *$M^2(0,1)$ is included in* Domδ *and the operator δ restricted to it coincides with the Itô integral, i.e.,*

$$\delta(u) = \int_0^1 u(s)\,\mathrm{d}B(s).$$

Proof. Suppose that $u \in M^2(0,1)$ is an elementary process of the form $u_t = \xi \mathbb{1}_{(a,b)}(t)$, where $(a,b) \subset [0,1]$ and ξ is a square integrable random variable, $\mathcal{F}_a$-measurable. Suppose first that $\xi \in \mathcal{S}$. Then for a smooth random variable η it holds

$$\begin{aligned}\mathbb{E}\int_0^1 (D_t\eta)\xi\mathbb{1}_{(a,b)}(t)\,\mathrm{d}t &= \mathbb{E}\int_0^1 [D_t(\xi\eta) - \eta D_t\xi]\mathbb{1}_{(a,b)}(t)\,\mathrm{d}t \\ &= \mathbb{E}[\eta\xi\delta(\mathbb{1}_{(a,b)}) - \int_0^1 \eta D_t\xi\mathbb{1}_{(a,b)}(t)\,\mathrm{d}t)] = \mathbb{E}[\eta\xi(B(b) - B(a))]\end{aligned}$$

(we have used that $D_t\xi = 0$, $t \in (a,b)$, since ξ is $\mathcal{F}_t$-adapted). This means, using again the duality relationship (14.1),

$$\delta(\xi \mathbb{1}_{(a,b)}) = \xi(B(b) - B(a)). \tag{14.7}$$

Now the general case follows by a limit argument, using the density of smooth random variables in $L^2(\Omega)$ and the closedness of δ.

Further, by linearity, the class of elementary processes

$$X_t(\omega) = \sum_{i=0}^{n-1} \xi_i(\omega) \mathbb{1}_{[t_i,t_{i+1})}(\omega)$$

belongs to Domδ, and again a limit argument, using the density of elementary processes in $M^2(0,1)$, implies the thesis.

□

The following proposition concerns with the product of an integral with a random variable.

Proposition 14.7. *Let $u(t) \in$ Domδ and $Q \in \mathbb{D}^{1,2}$ be such that $Qu(t) \in L^2([0,1] \times \Omega)$ (We don't make any assumption about adapteness). Then $Qu(t)$ belongs to* Domδ *and the following equality holds*

$$\int_0^1 Qu(t)\, \mathrm{d}_S B(t) = Q \int_0^1 u(t)\, \mathrm{d}_S B(t) - \int_0^1 D_t Q\, u(t)\, \mathrm{d}t. \tag{14.8}$$

Proof. Consider a smooth random variable $G \in \mathcal{S}$; then formula (9.11) and the definition of Skorohod integral imply

$$\mathbb{E}[\int_0^1 D_t G\, Qu(t)\, \mathrm{d}t] = \mathbb{E}[\int_0^1 u(t)\, D_t(GQ)\, \mathrm{d}t - \int_0^1 u(t) G D_t Q\, \mathrm{d}t]$$
$$= \mathbb{E}[G \int_0^1 Qu(t)\, \mathrm{d}_S B(t) - G \int_0^1 u(t) D_t Q\, \mathrm{d}t]$$

which proves the result.

□

14.2 One dimensional stochastic equations with additive noise

In this section, we shall be concerned with a stochastic differential equation with additive noise (on the time interval $[0,1]$)

$$\begin{cases} \mathrm{d}X_t = b(X_t)\, \mathrm{d}t + \mathrm{d}B_t, & t \in [s,1], \\ X_s = x \in \mathbb{R}, \quad s \geq 0. \end{cases} \tag{14.9}$$

We consider the following assumption on b,

(H) $b \in C^1(\mathbb{R})$, and, setting $b'(x) = \frac{d}{dx} b(x)$,

$$\|b\|_1 = \sup_{x \in \mathbb{R}} |b'(x)| < \infty. \tag{14.10}$$

In this section we show that the solution $X(t;s,x)$, for any $t \in [0,1]$, is differentiable with respect to the initial condition x and in the Malliavin sense.

The following result is contained as a special case in the last Problem of Lecture 13 and we omit the proof here.

Theorem 14.8. *Assume that Hypothesis* **(H)** *holds. Then $X(t;s,x)$ is differentiable in x, $\mathbb{P}$-a.s., and we have*

$$\partial_x X(t;s,x) = \eta(t;s,x) \tag{14.11}$$

is the solution of the variation equation

$$\begin{cases} \mathrm{d}\eta(t;s,x) = b'(X(t;s,x)) \cdot \eta(t;s,x)\,\mathrm{d}t, & s \le t \in [0,1], \\ \eta(s;s,x) = 1. \end{cases}$$

Remark 14.9. The representation

$$\eta(t;s,x) = \exp\left(\int_s^t b'(X(r;s,x))\,\mathrm{d}r\right)$$

together with Hypothesis **(H)** and the regularity of X implies that $t \mapsto \eta(t;s,x)$ is a continuous and adapted process on $(s,1)$.

We now consider the Malliavin derivative of the flow X. We can show that the solution belongs to the space $\mathbb{D}^{1,\infty} = \bigcap_{p\ge 1} \mathbb{D}^{1,p}$.

Theorem 14.10. *Assume that hypothesis* **(H)** *holds. Let $\{X(t;s,x),\ s \le t \in [0,1]\}$ be the unique continuous solution of equation (14.9). Then $X(t;s,x)$ belongs to $\mathbb{D}^{1,\infty}$ for any $s \le t \in [0,1]$; moreover*

$$\sup_{r\in[s,1]} \mathbb{E}\left[\sup_{t\in[r,1]} |D_r X(t;s,x)|^p\right] < \infty$$

and the Malliavin derivative $D_r X(t;s,x)$ satisfies the following linear equation

$$D_r X(t;s,x) = 1 + \int_r^t b'(X(\tau;s,x)) D_r X(\tau;s,x)\,\mathrm{d}\tau, \qquad s \le r \le t \in [0,1]. \tag{14.12}$$

Proof. With no loss of generality, we may assume that $s = 0$, and for simplicity we suppress the dependence on the initial condition. We will prove by induction that the approximating processes

$$\begin{aligned} X_0(t) &= x \\ X_{n+1}(t) &= x + \int_0^t b(X_n(s))\,\mathrm{d}s + B(t), \qquad n \ge 1, \end{aligned}$$

satisfies the property

$$X_n^i(t) \in \mathbb{D}^{1,\infty} \text{ and } \sup_{r\in[0,1]} \mathbb{E}\left[\sup_{t\in[r,1]} |D_r X_n(t)|^p\right] < \infty. \tag{14.13}$$

In fact, let us consider the Malliavin derivative of $X_{n+1}(t)$; the chain rule for Malliavin derivative implies (recall also that $B(t) = \int_0^1 \mathbb{1}_{[0,t]}(r)\,\mathrm{d}B_r$)

$$D_r X_{n+1}(t) = \int_0^t b'(X_n(s)) D_r X_n(s)\,\mathrm{d}s + \mathbb{1}_{[r,1]}(t)$$

and since the process $X_n(t)$ is adapted to $\mathcal{F}_t$, we obtain

$$D_r X_{n+1}(t) = 1 + \int_r^t b'(X_n(s)) D_r X_n(s)\,\mathrm{d}s, \qquad t \in [r,1].$$

Then for any $r \in [0,t]$,

$$\mathbb{E}\big[\sup_{s\in[r,t]} |D_r X_{n+1}(s)|^p\big] = c_p \left(1 + T^{p-1}\|b\|_1^p \mathbb{E}\int_r^t |D_r X_n(s)|^p\,\mathrm{d}s\right)$$

and (14.13) holds. Moreover, we obtain also that the derivatives $D_r X_n(t)$ are uniformly bounded as $n \to \infty$; therefore, by Proposition 9.10 we obtain that $X_t \in \mathbb{D}^{1,\infty}$ for any $t \in [0,1]$. Finally, applying the Malliavin derivative D to both sides of (14.9) we deduce equation (14.12) for DX_t.

□

We have, then, the following identity between the process η (see (14.11)) and the Malliavin derivative of the flow:

$$D_r X(t;s,x) = \eta(t;r,x), \qquad s < r \le t \in [0,1]. \tag{14.14}$$

14.2.1 Smoothness of the transition function

For fixed time $t = 1$, we denote by $\pi(x,\cdot)$ the law of the random variable $X(1,x)$, so that

$$P_1 f(x) = \int_{\mathbb{R}} f(y)\pi(x,\mathrm{d}y), \quad x \in \mathbb{R},\ f \in C_b(\mathbb{R}).$$

In this section we show that $\pi(x,\cdot)$ has a density with respect to the Lebesgue measure on $\mathbb{R}$. The argument is based on the following elementary analytic result (which cannot be extended to higher dimensions).

Lemma 14.11. *Let m be a probability measure on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ and assume that there is $C > 0$ such that*

$$\left|\int_{\mathbb{R}} f'(x)\,m(\mathrm{d}x)\right| \le C\|f\|_0 \text{ for all } f \in C_b^1(\mathbb{R}). \tag{14.15}$$

Then there exists $\rho \in L^2(\mathbb{R})$ such that

$$m(\mathrm{d}x) = \rho(x)\,\mathrm{d}x. \tag{14.16}$$

Proof. Let $\widehat{m}$ be the Fourier transform of m,

$$\widehat{m}(y) = \int_{\mathbb{R}} \mathrm{e}^{\imath xy}\,\mu(\mathrm{d}x).$$

Let $y \in \mathbb{R}$ and set in (14.15)

$$f(\xi) = \mathrm{e}^{\imath xy}, \quad x \in \mathbb{R}.$$

We obtain

$$|y\widehat{m}(y)| \le C \quad \text{for all } y \in \mathbb{R}.$$

Since $|\widehat{m}(y)| \le 1$ for all $y \in \mathbb{R}$, there is $C_1 > 0$ such that,

$$|\widehat{m}(y)| \le \frac{C_1}{1+|y|}, \qquad \text{for all } y \in \mathbb{R},$$

so that $\widehat{m} \in L^2(\mathbb{R})$. Therefore, setting

$$\rho(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-\imath xy}\widehat{m}(y)\,\mathrm{d}y$$

we can conclude that $\rho \in L^2(\mathbb{R})$ and $m(\mathrm{d}x) = \rho(x)\,\mathrm{d}x$.
□

Theorem 14.12. $\pi(x,\cdot)$ *is absolutely continuous with respect to the Lebesgue measure* λ *on* $\mathbb{R}$*. Moreover, the density* $\frac{\mathrm{d}\pi(x,\cdot)}{\mathrm{d}\lambda}$ *belongs to* $L^2(\mathbb{R})$*.*

Proof. In view of Lemma 14.11 we only have to show that there is $C > 0$ such that

$$\left|\int_{\mathbb{R}} \varphi'(y)\,\pi(x,\mathrm{d}y)\right| = |\mathbb{E}[\varphi'(X(1;x))]| \le C\|\varphi\|_0, \qquad \text{for all } \varphi \in C_b^1(\mathbb{R}). \tag{14.17}$$

Recalling the chain rule for Malliavin derivative (9.12) and relation (14.14) we write

$$\varphi'(X(1,x)) = \frac{D_r\varphi(X(1,x))}{D_rX(1,x)} = D_r\varphi(X(1,x))\Phi(r,1)$$

with

$$\Phi(r,1) = \eta(1;r,x)^{-1} = \exp\left(-\int_r^1 b'(X(s,x))\,\mathrm{d}s\right).$$

Taking the expectation and integrating both sides of the above identity in r between 0 and 1, we obtain

$$\mathbb{E}[\varphi'(X(1,x))] = \mathbb{E}[\int_0^1 D_r\varphi(X(1,x))\Phi(r,1)\,\mathrm{d}r] = \mathbb{E}[\varphi(X(1,x))\delta(\Phi(\cdot,1))]. \tag{14.18}$$

We claim that

$$\{\Phi(r,1),\ r \in [0,1]\} \quad \text{belongs to} \quad \mathrm{Dom}\delta \tag{14.19}$$

and that the isometry formula applies, thus yielding

$$\mathbb{E}|\delta(\Phi(\cdot,1))|^2 = \mathbb{E}\int_0^1 |\Phi(r,1)|^2\,\mathrm{d}r. \tag{14.20}$$

Notice that

$$\mathbb{E}\int_0^1 |\Phi(r,1)|^2\,\mathrm{d}r = \mathbb{E}\int_0^1 \exp\left(-2\int_r^1 b'(X(s,x))\,\mathrm{d}s\right)\,\mathrm{d}r \le \mathbb{E}\int_0^1 \exp\left(2\|b\|_1\,(1-r)\right)\,\mathrm{d}r \le \mathrm{e}^{2\|b\|_1}.$$

Using claim (14.19), we apply Hölder inequality in the right hand side of (14.18), and we find

$$\begin{aligned}|\mathbb{E}[\varphi'(X(1,\xi,x)]| &\le (\mathbb{E}|\varphi(X(1,x))|^2)^{1/2}\,\left(\mathbb{E}|\delta(\Phi(\cdot,1))|^2\right)^{1/2}\\ &\le \|\varphi\|_0\,\mathrm{e}^{\|b\|_1}\end{aligned}$$

and estimate (14.17) is proved.

□

Backward Itô integral

We are left with claim (14.19). The process $\{\Phi(r,1),\ r\in[0,1]\}$ is adapted with respect to the family of σ-fields $\{\mathcal{F}^t,\ t\in[0,1]\}$, where

$$\mathcal{F}^t = \sigma(B_s - B_t,\ s\ge t),\qquad t\in[0,1).$$

Moreover, it has continuous paths and is square integrable on $[0,1]\times\Omega$. We call such kind of processes backward Itô square integrable processes $M^2_b(0,1)$; for these processes it is possible to define a stochastic integral similar to the Itô integral, by means of Riemann sums which are taken in the opposite (time) direction.

Assume that $u\in M^2_b(0,1)$ is an elementary process, i.e.,

$$u(t) = \sum_{k=1}^n X_k\,\mathbb{1}_{(t_k,t_{k+1})}(t),$$

where $X_k\in\mathcal{F}^{t_{k+1}}$; notice that X_k is independent from $B(t_{k+1})-B(t_k)$. Define

$$I_b(u) = \int_0^1 u(t)\,\mathrm{d}_bB(t) = \sum_{k=1}^n X_k\left(B(t_{k+1})-B(t_k)\right).$$

The random variable $I_b(u)$ has zero mean and variance equals to $\|u\|^2_{L^2((0,1)\times\Omega)}$. The construction of the stochastic backward Itô integral now proceeds most likely as in the case treated in Lecture 10.

It is possible to proceed as in the proof of Theorem 14.6 in order to show that $M^2_b(0,1)\subset\mathrm{Dom}\delta$; we leave the details to the interested reader. Further, notice that, for any $u\in M^2_b(0,1)$, it holds $D_r u(t)=0$ for $r\le t$ –since $u(t)$ is adapted to $\mathcal{F}^t$ and it is independent from $\mathcal{F}_t$, hence the second term in the isometry formula for Skorohod integral disappears and we get (14.20).

14.2.2 Regularizing effect of the associated diffusion Markov semigroup

Here we prove that the transition semigroup associated to the solution of (14.9) (where $s=0$) has a regularizing effect. This will be a consequence of a representation formula for the derivative of the semigroup (see (14.22)) which is of independent interest. Recently, extensions of such formula to diffusion semigroups in higher dimensions have been much studied; they are called Bismut-Elworthy-Li formulae (see Elworthy and Li [EL94]).

We again fix time $t = 1$ for simplicity (but the same argument works for any $t > 0$, repeating the construction on the time interval $[0, \lfloor t \rfloor + 1]$). It is clear that for $f \in C_b^1(\mathbb{R})$ we have that $P_1 f$ is differentiable and

$$\frac{d}{d\xi} P_1 f(\xi) = \mathbb{E}[f'(X(1;\xi))\eta(1;0,\xi)]; \tag{14.21}$$

however, this formula contains the derivative of f and it is not useful for our purpose. Our aim is to provide a formula containing f instead of f'. We denote by $UC_b(\mathbb{R})$ the Banach space of all uniformly continuous and bounded functions $f : \mathbb{R} \to \mathbb{R}$ endowed with the sup-norm.

Theorem 14.13. *For all $f \in UC_b(\mathbb{R})$ we have $P_1 f \in C_b^1(\mathbb{R})$ and*

$$\frac{d}{d\xi} P_1 f(\xi) = \mathbb{E}\left[f(X(1;\xi)) \int_0^1 \eta(r;0,\xi)\, \mathrm{d}B_r \right], \quad f \in UC_b(\mathbb{R}). \tag{14.22}$$

Further, there exists a bound for the sup-*norm of the derivative of the semigroup depending only on the* sup-*norm of f and the quantity $\|b\|_1$ from Hypothesis* **(H)***:*

$$\sup_{\xi \in \mathbb{R}} \left| \frac{d}{d\xi} P_1 f(\xi) \right| \le \mathrm{e}^{2\|b\|_1} \|f\|_0, \quad f \in UC_b(\mathbb{R}). \tag{14.23}$$

Proof.
Step I. Assume first that $f \in C_b^1(\mathbb{R})$. The chain rule for the Malliavin derivative implies

$$f'(X(1;\xi)) = \frac{D_r(f(X(1;\xi)))}{D_r X(1;\xi)}, \qquad \forall\, r \in [0,1].$$

If we substitute the above in (14.21) we obtain

$$\frac{d}{d\xi} P_1 f(\xi) = \mathbb{E}\left[\frac{D_r(f(X(1;\xi)))}{\eta(1;r,\xi)} \eta(1;0,\xi) \right]$$

for $r \in [0,1]$: integrating both sides in r between 0 and 1 we obtain

$$\frac{d}{d\xi} P_1 f(\xi) = \int_0^1 \mathbb{E}\left[D_r(f(X(1;\xi)))\, \eta(r;0,\xi) \right] \mathrm{d}r. \tag{14.24}$$

Recall that $\eta(r;0,\xi)$ is a continuous and adapted process hence it belongs to $M^2(0,1)$ and, *a fortiori*, to Domδ. Therefore, on the right hand side of (14.24) we apply (14.1) to get

$$\frac{d}{d\xi} P_1 f(\xi) = \mathbb{E}\left[f(X(1;\xi)) \int_0^1 \eta(r;0,\xi)\, \mathrm{d}B_r \right].$$

Step II. We prove estimate (14.23) for $f \in C_b^1(\mathbb{R})$. Taking the norm in (14.22) and using Hölder inequality on the right hand side we obtain

$$\left|\frac{d}{d\xi}P_1 f(\xi)\right| \le \left(\mathbb{E}\left[f(X(1;\xi))^2\right]\right)^{1/2} \left(\mathbb{E}\left|\int_0^1 \eta(r;0,\xi)\,\mathrm{d}B_r\right|^2\right)^{1/2}.$$

Itô isometry implies

$$\mathbb{E}\left|\int_0^1 \eta(r;0,\xi)\,\mathrm{d}B_r\right|^2 = \mathbb{E}\int_0^1 |\eta(r;0,\xi)|^2\,\mathrm{d}r$$

and since $|\eta(r;0,\xi)|^2 \le \exp(2\|b\|_1 r)$, substituting in the above formula we get the assertion.

Step III. Using the density of $C_b^1(\mathbb{R})$ in $UC_b(\mathbb{R})$ and the fact that the right hand side of (14.22) depends only on f and not on its derivative, it is straightforward to check the first part of the theorem. Also estimate (14.23) follows easily.

□

References

[Ar74] L. Arnold, **Stochastic differential equations: Theory and applications**, (J. Wiley & Sons, 1974).

[Ba00] P. Baldi, **Equazioni differenziali stocastiche e applicazioni**, (Pitagora Editrice, 2000). Seconda edizione.

[Bas95] R. Bass, **Probabilistic techniques in analysis**, (Springer-Verlag, New York, 1995). Probability and its Applications.

[Bau96] H. Bauer, **Probability Theory**, (De Gruyter, 1996). De Gruyter studies in mathematics, 23.

[Bi95] P. Billingsley, **Probability and measure**, (Wiley, 1995). Third edition. Wiley Series in Probability and Mathematical Statistics.

[BG68] R.M. Blumenthal, R.K. Getoor, **Markov processes and potential theory**, (Academic Press, New York-London, 1968).

[Bon98] S. Bonaccorsi, **Some Applications in Malliavin Calculus**, Ph.D. Thesis, University of Trento, 1998.

[Bo98] V. Bogachev, **Gaussian measures**, (American Mathematical Society, Providence, RI, 1998). Mathematical Surveys and Monographs, 62.

[BH91] N. Bouleau and F. Hirsch, **Dirichlet forms and analysis on Wiener space**, (De Gruyter, 1991). De Gruyter Studies in Mathematics, 14.

[Ca96] P. Cattiaux, S. Roelly and H. Zessin, Une approche gibbsienne des diffusions browniennes infinidimensionnelles, *Probab. Theory Related Fields*, 104(2):147–179, 1996.

[Ch82] K.-L. Chung, **Lectures from Markov processes to Brownian motion**, (Springer-Verlag, 1982). Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science], 249.

[Ch02] K.-L. Chung, **Brown, and probability & Brownian motion on the line**, (World Scientific Publishing Co., 2002).

[DD83] D. Dacunha-Castelle and M. Duflo, **Probabilités et statistiques**, Tome 1 & 2, (Masson, 1983).

[Da78] G. Da Prato, M. Iannelli & L. Tubaro, Dissipative functions and finite-dimensional stochastic differential equations, *J. Math. pures et appl.*, 57:173–180, 1978.

[Da98] G. Da Prato, **Introduction to differential stochastic equations**, (Scuola Normale Superiore Pisa, 1998). Collana "Appunti Scuola Normale Superiore".

[Da01] G. Da Prato, **An introduction to infinite dimensional analysis**, (Scuola Normale Superiore Pisa, 2001). Collana "Appunti Scuola Normale Superiore".

[Da03] G. Da Prato, **Introduction to Brownian Motion and Mallivin Calculus**, (Università di Trento, Dipartimento di Matematica, 2003). Lecture Notes Series 17.

[Da05] G. Da Prato, **Introduction to Stochastic Analysis**, (Scuola Normale Superiore, Pisa, 2005).

[DZ92] G. Da Prato and J. Zabczyk, **Stochastic equations in infinite dimensions**, (Cambridge University Press, Cambridge, 1992). Encyclopedia of Mathematics and its Applications, 44.

[DZ02] G. Da Prato and J. Zabczyk, **Second Order Partial Differential Equations in Hilbert Spaces**, (Cambridge University Press, Cambridge, 2002).

[Do53] J. L. Doob, **Stochastic Processes**, (Wiley, 1953).

[Du02] R. M. Dudley, **Real Analysis and Probability**, (Cambridge University Press, Cambridge, 2002). Cambridge Studies in Advanced Mathematics, 74. Revised reprint of the 1989 original.

[DS88] N. Dunford and J.T. Schwartz, **Linear operators. Part I**, (John Wiley & Sons Inc., New York, 1988). Wiley Classics Library, Reprint of the 1958 original.

[Du96] R. Durrett, **Stochastic Calculus. A practical introduction**, (CRC Press, 1996). Probability and Stochastic Series.

[Dy65] E.B. Dynkin, **Markov processes**, Vols. I, II, (Academic Press Inc., Publishers, New York; Springer-Verlag, Berlin-Gttingen-Heidelberg 1965).

[DY69] E. Dynkin and A. Yushkevich, **Markov Processes: Theorems and Problems**, (Plenum Press, New York 1969).

[Ei05] A. Einstein, Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flssigkeiten suspendierten Teilchen, *Ann. Phys.*, 17, 1905.

[EL94] K. D. Elworthy and X. M. Li, Formulae for the derivatives of heat semigroups, *J. Funct. Anal.*, 125, 1994, 252-286.

[EN00] Klaus-Jochen Engel, Rainer Nagel, **One-parameter semigroups for linear evolution equations**, (Springer-Verlag, New York, 2000). With contributions by S. Brendle, M. Campiti, T. Hahn, G. Metafune, G. Nickel, D. Pallara, C. Perazzoli, A. Rhandi, S. Romanelli and R. Schnaubelt. Graduate Texts in Mathematics, 194.

[EK86] S. Ethier and T. Kurtz, **Markov Processes. Characterization and Convergence**, (John Wiley & Sons, Inc., New York, 1986). Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics.

[Fe53] W. Feller, **Introduction to Probability Theory and its Applications** (Wiley, New York, 1953).

[Fl93] F. Flandoli, **Calcolo delle probabilità e applicazioni**, appunti 1993.

[Fr71] David Freedman, **Brownian Motion and Diffusion**, (Springer, 1983).

[Fri75] A. Friedman, **Stochastic differential equations and applications**, (Academic Press, 1975).

[GT82] Bernard Gaveau and Philip Trauber, L'intégrale stochastique comme opérateur de divergence dans l'espace fonctionnel. *J. Funct. Anal.*, 46(2):230-238, 1982.

[GP92] A. Grorud and É. Pardoux, Intégrales hilbertiennes anticipantes par rapport à un processus de Wiener cylindrique et calcul stochastique associé. *Appl. Math. Optim.*, 25(1):31–49, 1992.

[Hi80] T. Hida, **Brownian motion**, (Springer-Verlag, New York, 1980), Applications of Mathematics, 11.

[Hi93] T. Hida, H.-H. Kuo, J. Potthoff, and L. Streit, **White noise**, (Kluwer Academic Publishers Group, Dordrecht, 1993). Mathematics and its Applications, 253.

[Ho96] H. Holden, B. Oksendal, J. Uboe, and T. Zhang, **Stochastic partial differential equations**, (Birkhäuser Boston Inc., Boston, MA, 1996).

[IW81] N. Ikeda and S. Watanabe, **Stochastic differential equations and diffusion processes**, (North-Holland Publishing Co., Amsterdam-New York; Kodansha, Ltd., Tokyo, 1981). North-Holland Mathematical Library, 24.

[It44] K. Itô, Stochastic integral. *Proc. Imp. Acad. Tokyo*, **20**:519–524, 1944.

[It53] K. Itô, Multiple Wiener integral. *J. Math. Soc. Japan* **3**:157–169, 1951.

[IM74] K. Itô and H.P. McKean, Jr., **Diffusion processes and their sample paths**, (Springer-Verlag, Berlin Heidelberg New York, 1974). Second printing, corrected.

[Ja01] N. Jacob, **Pseudo differential operators and Markov processes**, (Imperial College Press, London, 2001). Vol. I. Fourier analysis and semigroups.

[JP00] J. Jacod and P. Protter, **Probability essentials**, (Springer-Verlag, Berlin, 2000). Universitext.

[Ka02] O. Kallenberg, **Foundations of Modern Probability**, (Springer-Verlag, New York, 2002). Second edition.

[KS88] I. Karatzas and S. Shreve, **Brownian motion and stochastic calculus**, (Springer-Verlag, New York, 1988). Graduate Texts in Mathematics, 113.

[Kl99] F. C. Klebaner, **Introduction to stochastic calculus with applications**, (World Scientific Publishing Co., Inc., River Edge, NJ, 1999). Reprint of the 1998 original.

[Kr95] N. V. Krylov, **Introduction to the theory of diffusion processes**, (American Mathematical Society, Providence, RI, 1995). Translated from the Russian manuscript by Valim Khidekel and Gennady Pasechnik. Translations of Mathematical Monographs, 142.

[LL97] D. Lamberton and B. Lapeyre, **Introduction au calcul stochastique appliqueéa la finance**, (Ellipses, Edition Marketing, Paris, 1997). Second edition.

[Let93] G. Letta, **Probabilità elementare**, (Zanichelli Editore, 1993).

[Lé54] P. Lévy, **Le Mouvement Brownien**, (Gauthier Villars, 1954).

[Lé65] P. Lévy, **Processus Stochastiques et Mouvement Brownien**, (Gauthier Villars, 1965).

[Ma97] Paul Malliavin, **Stochastic analysis**, (Springer-Verlag, Berlin, 1997), Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 313.

[McK69] H. P. McKean, Jr., **Stochastic integrals** (Academic Press, New York-London 1969). Probability and Mathematical Statistics, No. 5.

[MPW02] G. Metafune, D. Pallara and M. Wacker, Feller semigroups on $\mathbf{R}^N$, *Semigroup Forum* 65(2):159–205, 2002.

[Ne01] E. Nelson, **Dynamical Theories of Brownian Motion**, second edition, Princeton University Press, August 2001. Posted on the Web at `http://www.math.princeton.edu/ nelson/books.html`

[Nu95] D. Nualart, **The Malliavin calculus and related topics**, (Springer, 1995).

[NP88] D. Nualart and E. Pardoux, Stochastic calculus with anticipating integrands, *Probab. Theory Rel. Fields*, 78:535–581, 1988.

[NZ86] D. Nualart and M. Zakai, Generalized multiple stochastic integrals and the representation of Wiener functionals, *Stochastics*, 23:311–330.

[NZ89] D. Nualart and M. Zakai, The partial Malliavin calculus, in **Seminaire de Probabilités XXIII**, (Springer, 1989). Lecture Notes in Math., 1372.

[MR04] S. Maniglia and A. Rhandi, **Gaussian measures on separable Hilbert spaces and applications**, (Edizioni del Grifo, Quaderno 1/2004).

[Pa83] A. Pazy, **Semigroups of Linear Operators and Applications to Partial Differential Equations**, (Springer-Verlag, New York, 1983).

[Pr04] P. E. Protter, **Stochastic integration and differential equations**, (Springer-Verlag, Berlin, 2004). Second edition.

[ReSi] M. Reed, B. Simon, **Methods of modern mathematical physics**, (Academic Press, New York-London, 1972-1979). I. Functional analysis; II. Fourier analysis, self-adjointness; III. Scattering theory; IV. Analysis of operators.

[Ru91] W. Rudin, **Functional analysis**, (McGraw-Hill, Inc., New York, 1991). Second edition. International Series in Pure and Applied Mathematics.

[Sc98] R. Schilling, Conservativeness and extensions of Feller semigroups, *Positivity*, 2(3):239–256, 1998.

[Sh] S. Shreve, **Stochastic calculus and finance**

[Sk75] A. V. Skorohod, On a generalization of the stochastic integral, *Teor. Verojatnost. i Primenen.*, 20(2):223–238, 1975.

[SV79] D. Stroock, S. R. S. Varadhan, **Multidimensional diffusion processes**, (Springer-Verlag, Berlin-New York, 1979). Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 233.

[Ta88] Kazuaki Taira, **Diffusion processes and partial differential equations**, (Academic Press, 1988).

[Ve83] A. Ventsel, **Teoria dei processi stocastici**, (Editori Riuniti – Edizioni Mir, 1983). English version: **A course in the theory of stochastic processes**, (McGraw-Hill International Book Co., New York, 1981).

[Za04] J. Zabczyk, **Topics in Stochastic Processes**, (Scuola Normale Superiore Pisa, 2004). Collana "Quaderni".

Index