Lecture Note for MAT7093: Stochastic Analysis

LI Liying

June 12, 2024

1 Introduction

In this section we will give some motivations to study Brownian motions and stochastic integrals.

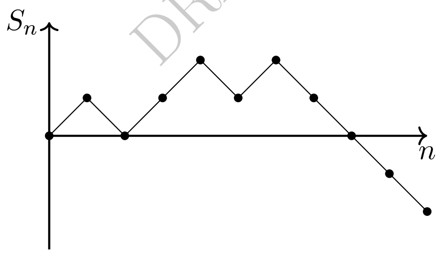
1.1 Stochastic processes

The well-known Central Limit Theorem (CLT) gives the universal behavior of the sum of many small independent variables: for i.i.d. r.v.'s X_i with $\mathsf{E}X_i = 0$, $\mathsf{E}X_i^2 = 1$, one has

$$\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} \Rightarrow_d \mathcal{N}(0,1).$$

Example 1.1 We can take X_i as the results of independent coin flips, so $\mathsf{P}(X_i = \pm 1) = 1/2$.

Write the partial sum as $S_n = X_1 + X_2 + \cdots + X_n$. We can plot the trajectory $n \mapsto S_n$ as below:



The plotted trajectory, which linearly interpolates between (n, S_n), $n \in \mathbb{N}$, can be written as

$$\tilde{S}_t = \begin{cases} S_n, & t = n \in \mathbb{N}, \\ (n+1-t)S_n + (t-n)S_{n+1}, & t \in (n, n+1). \end{cases}$$

Question *What is the limit of $t \mapsto \tilde{S}_t$ as (continuous) trajectories?*

The *Donsker's invariance principle*, a.k.a. the *Functional CLT*, states that in an appropriate sense, the limit is given by the *Brownian motion*, which is a "continuous stochastic process".

Theorem 1.1 (Functional CLT)

$$\Big(\frac{\tilde{S}_{nt}}{\sqrt{n}},\ t \geq 0\Big) \Rightarrow_d \Big(B_t,\ t \geq 0\Big),$$

where $(B_t)_{t\geq 0}$ is the Brownian motion (BM).

Remark 1.2 We will define rigorously what is a "continuous stochastic process" below.

Remark 1.3 The convergence "$\Rightarrow_d$" means convergence in distribution/law. As we are studying random functions rather than random variables, we need to work on probability measures on functional spaces, which are infinite-dimensional and quite different from finite-dimensional spaces like $\mathbb{R}^d$. We will return to this in Section 1.2.

Using the CLT, we can obtain the finite-dimensional distribution (f.d.d.) for Brownian motion. For fixed $t \geq 0$,

$$\mathcal{L}(B_t) = \lim_{n\to\infty} \mathcal{L}\Big(\frac{\tilde{S}_{[nt]}}{\sqrt{n}}\Big) = \lim_{n\to\infty} \mathcal{L}\Big(\frac{\tilde{S}_{[nt]}}{\sqrt{[nt]}} \cdot \sqrt{t}\Big) = \mathcal{N}(0, \sqrt{t}).$$

In general, for $0 = t_1 < t_2 < \cdots < t_m$, it is believable that

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

should have the same distribution as independent $\mathcal{N}(0, t_1)$, $\mathcal{N}(0, t_2 - t_1), \cdots, \mathcal{N}(0, t_m - t_{m-1})$ r.v.'s.

Definition 1.1 *A stochastic process $(X_t)_{t\in T}$ ($T = \mathbb{Z}, \mathbb{R}$, etc) on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is such that for every fixed $t \in T$,*

$$\omega \in \Omega \mapsto X_t(\omega)$$

is a measurable map from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Remark 1.4 As a notation, we may simply write "X_t is $\mathcal{B}(\mathbb{R})/\mathcal{F}$-measurable".

Definition 1.2 *For a stochastic process $(X_t)_{t\in T}$, its finite-dimensional distribution (f.d.d.) is the collection of all the laws*

$$\mathcal{L}(X_{t_1}, X_{t_2}, \ldots, X_{t_m}), \quad t_1, t_2, \ldots, t_m \in T.$$

It follows from Definition 1.1 that all the sets

$$\{(X_{t_1}, X_{t_2}, \ldots, X_{t_m}) \in A\}, \quad A \in \mathcal{B}(\mathbb{R}^m)$$

are measurable, and hence f.d.d. of a stochastic process is well-defined.

Homework (Transformation of BM)

1. Prove the equivalency of the following two conditions: for $0 = t_0 \leq t_1 < \cdots < t_m$,

$$\begin{aligned} & \mathcal{L}(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_m} - B_{t_{m-1}}) = \mathcal{N}(0, \mathrm{diag}\{t_{i+1} - t_i\}_{0\leq i\leq m-1}) \\ \Leftrightarrow \quad & (B_{t_1}, B_{t_2}, \ldots, B_{t_m}) \text{ is a centered Gaussian vector with covariance } \mathsf{E}B_{t_i}B_{t_j} = t_i \wedge t_j. \end{aligned} \tag{1.1}$$

2. Suppose that $(B_t)_{t\geq 0}$ has f.d.d. (1.1). Show that all the following processes have the same f.d.d. (1.1).
   a) $(-B_t)_{t\geq 0}$.
   b) $(B_t^\lambda)_{t\geq 0} := (\frac{1}{\lambda} B_{\lambda^2 t})_{t\geq 0}$. (Fix $\lambda > 0$.)
   c) $(B_t^{(s)})_{t\geq 0} := (B_{t+s} - B_s)_{t\geq 0}$. (Fix $s > 0$.)
   d) $(tB_{1/t})_{t\geq 0}$ (with the convention $0 \cdot B_{1/0} = 0$).

Hint: You can find some basic properties of Gaussian vectors in Section 2.1. This exercise is basically about covariance computation.

It is believable that a stochastic process is more or less determined by all its f.d.d. (which is done by Komolgorov's Extension Theorem, see for example [Shi96, Chap. II.3, Theorem 4]). With the definition of stochastic processes at hand, the next question is what makes a "continuous" stochastic process. To discuss continuity we now take T to be an interval of $\mathbb{R}$ ($T = [a, b]$, $[0, \infty)$, etc). Then, a "continuous" process requires additionally that the map

$$t \mapsto X_t(\omega)$$

is *continuous* for P-a.e. ω.

Remark 1.5 For a generic stochastic process $(X_t)_{t \in \mathbb{R}}$, the sets

$$\mathcal{C} = \{\omega : t \mapsto X_t(\omega) \text{ is continuous.}\}$$

and (for $t_0 \in T$)

$$\mathcal{C}_{t_0} = \{\omega : t \mapsto X_t(\omega) \text{ is continuous at } t = t_0.\}$$

are NOT measurable.

To see this, recall that we can characterize the continuity of a function by sequential convergence, namely,

$$\lim_{t \to t_0} f(t) = f(t_0) \quad \Leftrightarrow \quad \forall t_n \to t_0, \; \lim_{n \to \infty} f(t_n) = f(t_0).$$

Although for any fixed sequence (t_n), the set

$$\{\omega : \lim_{n \to \infty} X_{t_n} = X_{t_0})\} = \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{\omega : |X_{t_n} - X_{t_0}| < \frac{1}{m}\}$$

is in $\mathcal{F}$ (hence measurable), there are uncountably many such sequences (t_n) such that $t_n \to t_0$.

Homework Let $(X_n)_{n \geq 1}$ and X_∞ be r.v.'s on $(\Omega, \mathcal{F}, \mathsf{P})$. Show that

$$\{\omega : \lim_{n \to \infty} X_n(\omega) = X_\infty(\omega)\} = \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{\omega : |X_n(\omega) - X_\infty(\omega)| < \frac{1}{m}\}$$

Conclude that the left hand side belongs to $\mathcal{F}$.

Due to the potential measurability issue, the continuity of a stochastic process is somehow an "independent" property to consider, so additional efforts are always needed for the justification. There are generally two approaches: one is to use Komolgorov's Continuity Test (its usage summarized in Theorem 1.2), the other one is to directly build up probability measures on the desired functional spaces (Section 1.2).

But assuming that this can be done, we are ready to rigorously define what a Brownian motion is. One last thing to do is to specify how we distinguish between different stochastic processes.

Definition 1.3 *Two stochastic processes $X = (X_t)_{t \in T}$, $Y = (Y_t)_{t \in T}$, defined on $(\Omega, \mathcal{F}, \mathsf{P})$, are called* modifications *of each other if*

$$\mathsf{P}(X_t = Y_t) = 1, \quad \forall t \in T.$$

That is, X and Y have the same f.d.d.

Definition 1.4 *Y is called a* version *of X, or* indistinguishable *from X, if for a.e. ω,*

$$X_t = Y_t, \quad \forall t \in T.$$

Clearly, when T is uncountable, the above two definitions are not equivalent.

Remark 1.6 It is tempting to write $\mathsf{P}(X_t = Y_t,\ \forall t \in T) = 1$. However, without additional assumptions on the processes X and Y, it is not clear whether the set $\{X_t = Y_t,\ \forall t \in T\}$ is measurable. If some statement holds for "a.e. ω", what is means is that it is true on an event $\tilde{\Omega}$ with $\mathsf{P}(\tilde{\Omega}) = 1$. It may still be true or not true for some ω in $\tilde{\Omega}^c$, but the point is that at least such exceptional points are contained in a set of zero probability. The issue could be resolved if additionally the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is assumed to be *complete*, in which case all subsets of zero-probability sets are measurable.

Homework Let $X = (X_t)_{t\geq 0}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathsf{P})$ such that $t \mapsto X_t(\omega)$ is continuous for almost every $\omega \in \Omega$. Let τ be a continuous r.v. on $(\Omega, \mathcal{F}, \mathsf{P})$ and $Y = (Y_t)_{t\geq 0}$ be defined as

$$Y_t(\omega) = \begin{cases} X_t(\omega), & t \neq \tau(\omega), \\ X_t(\omega) + 1, & t = \tau(\omega). \end{cases}$$

Show that Y is a stochastic process which is a modification of X, but $t \mapsto Y_t(\omega)$ is NOT continuous for almost every $\omega \in \Omega$.

Definition 1.5 *The (1d, standard) Brownian motion $(B_t)_{t\geq 0}$ is a continuous stochastic process with f.d.d. given by*

$$\mathcal{L}(B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_m} - B_{t_{m-1}}) = \mathcal{N}(0, \operatorname{diag}\{t_{i+1} - t_i\}_{0\leq i\leq m-1}), \quad 0 = t_0 \leq t_1 < \cdots < t_m. \tag{1.2}$$

In particular, $\mathsf{P}(B_0 = 0) = 1$.

The information of f.d.d. of Brownian motion indeed sheds some light on the continuity property. In fact, the continuity condition can be dropped in the above definition, if we allow ourselves to consider stochastic processes up to modifications. The next result is a consequence of the Kolmogorov's Continuity Test.

Theorem 1.2 *If $(X_t)_{t\geq 0}$ has the f.d.d. given in (1.2), then $(X_t)_{t\geq 0}$ has a continuous modification.*

Idea of the proof: We can use the f.d.d. on $\mathbb{Q}_+$ to show that for a.e. ω, $t \mapsto B_t(\omega)$ is uniformly continuous on $\mathbb{Q}_+$, that is, $\forall \varepsilon > 0$, $\exists \delta = \delta(\varepsilon, \omega)$ such that

$$|X_{t_1}(\omega) - X_{t_2}(\omega)| < \delta, \quad \forall |t_1 - t_2| < \varepsilon,\ t_1, t_2 \in \mathbb{Q}_+.$$

Then we can extend the function $t \mapsto X_t(\omega)$ on $\mathbb{Q}_+$ to a continuous function on $\mathbb{R}+$. □

The existence of a stochastic process with any given *consistent* f.d.d. is guaranteed by Kolmogorov's Extension Theorem, although later in this note we will exploit the Gaussian f.d.d. more to give another more explicit construction of Brownian motion (Section 2.2) . Then, using the above theorem we obtain a continuous stochastic process. We will fill in the gaps later in this note.

1.2 Probability measures on metric spaces

Recall that X is a r.v. on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ if $X : \Omega \to \mathbb{R}$ is $\mathcal{B}(\mathbb{R})/\mathcal{F}$-measurable. The distribution of X is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, given by

$$\mathcal{L}(X)(A) = \mathsf{P} \circ X^{-1}(A) = \mathsf{P}(X \in A), \quad A \in \mathcal{B}(\mathbb{R}).$$

The measure $\mathcal{L}(X)$ is determined by $\mathsf{P}(X \leq a)$, $a \in \mathbb{R}$, since $\mathcal{B}(\mathbb{R}) = \sigma\big((-\infty, a],\ a \in \mathbb{R}\big)$.

We want to replace $\mathbb{R}$ by a general metric space (M, d), where M can be as large as the space of all continuous functions. Any stochastic process from a probability measure on the space of continuous functions will automatically be continuous. We start by some basic notions on probability measures on metric spaces.

A metric space (M, d) is a set M equipped with a metric $d : M \times M \to \mathbb{R}_+$ which satisfies

- (symmetry) $d(x,y) = d(y,x)$;
- (positivity) $d(x,y) \geq 0$, and the equality holds only when $x = y$.
- (triangle inequality) $d(x,y) + d(y,z) \geq d(x,z)$.

Example 1.7 1. $M = \mathbb{Z}$, $d(x,y) = |x-y|$.

2. $M = \mathbb{R}^m$, with ℓ_p-distance

$$d_p(x,y) = \begin{cases} \Big[\sum_{i=1}^{m} |x_i - y_i|^p \Big]^{1/p}, & 1 < p < \infty, \\ \max_{1 \leq i \leq m} |x_i - y_i|, & p = \infty. \end{cases}$$

3. $M = \mathcal{C}[0,1]$, $d(x,y) = \sup_{t \in [0,1]} |x(t) - y(t)|$.

For a metric space, its Borel σ-algebra $\mathcal{B}(M)$ is the σ-algebra generated by all the open sets in M, or equivalently, the smallest σ-algebra containing all the open balls

$$B_r(x_0) = \{x : d(x, x_0) < r\}, \quad x_0 \in M, \ r > 0.$$

Definition 1.6 *Let (M,d) be a metric space. An M-value random element (r.e.) on $(\Omega, \mathcal{F}, \mathsf{P})$ is a measurable map from $(\Omega, \mathcal{F})$ to $(M, \mathcal{B}(M))$. The distribution of X is a probability measure on $(M, \mathcal{B}(M))$, given by*

$$(\mathsf{P} \circ X^{-1})(A) = \mathsf{P}(X \in A), \quad A \in \mathcal{B}(M). \tag{1.3}$$

The measure in (1.3) is determined its value on all open balls $B_r(x_0)$.

Example 1.8 Let X be a $\mathcal{C}[0,1]$-valued random element. Then $(X_t)_{t \in [0,1]}$ is a stochastic process.

In fact, for $t \in [0,1]$, we have the composition

$$\omega \mapsto X(\omega) \mapsto X_t(\omega),$$

where the first map is $\mathcal{B}(M)/\mathcal{F}$-measurable by the definition of random elements, and the second map is continuous since it is the evaluation map at given t of continuous functions and hence $\mathcal{B}(\mathbb{R})/\mathcal{B}(M)$-measurable. Therefore, the map $\omega \mapsto X_t(\omega)$ is $\mathcal{B}(\mathbb{R})/\mathcal{F}$-measurable.

Example 1.9 (Coordinate process) Let μ be a measure on $\big(\mathcal{C}(\mathbb{R}_+), \mathcal{B}(\mathcal{C}(\mathbb{R}_+))\big)$. Define

$$(\Omega, \mathcal{F}, \mathsf{P}) = \big(\mathcal{C}(\mathbb{R}_+), \mathcal{B}(\mathcal{C}(\mathbb{R}_+)), \mu\big), \quad X_t(\omega) = \omega_t, \ t \geq 0.$$

Then $(X_t)_{t \geq 0}$ is a continuous stochastic process.

A function $F : M \to \mathbb{R}$ is continuous if $d(x, x_0) \to 0$ implies $|F(x) - F(x_0)| \to 0$.

Definition 1.7 *Let $X^{(n)}$ and X be $\mathcal{C}[0,1]$-valued random elements defined on $(\Omega^{(n)}, \mathcal{F}^{(n)}, \mathsf{P}^{(n)})$ and $(\Omega, \mathcal{F}, \mathsf{P})$. We say that $X^{(n)}$ converge weakly (or converge in distribution/law) to X, denoted by $X^{(n)} \Rightarrow_d X$, if for all bounded and continuous $F : \mathcal{C}[0,1] \to \mathbb{R}$,*

$$\lim_{n \to \infty} \mathsf{E}^{(n)} F(X^{(n)}) = \mathsf{E} F(X).$$

Remark 1.10 It is annoying to work with different probability spaces, but the good news is that the underlying probability spaces are not relevant for the notion of weak convergence. Let $\mu_n = \mathsf{P}^{(n)} \circ [X^{(n)}]^{-1}$ and $\mu = \mathsf{P} \circ X^{-1}$. Then μ_n, μ are all (probability) measures on $(\mathcal{C}[0,1], \mathcal{B}(\mathcal{C}[0,1]))$. By standard functional analysis terminologies, the above definition says that $\mu_n \to \mu$ in the weak-* topology (since measures on metric spaces form the dual space of bounded continuous functions). In probability it is conventional to call it weak convergence.

The Brownian motion gives rise to a measure on $\mathcal{C}[0,1]$, called the *Wiener measure*. It is a probability measure on $\mathcal{C}[0,1]$ whose coordinate process has specific f.d.d.'s. To construct the Wiener measure directly:

- Functional CLT: need to understand (pre-)compact sets in $\mathcal{C}[0,1]$, and use the information of f.d.d. to verify tightness. A good read is [Bil99]).
- Gaussian measures on Banach spaces: more general, but still using the Gaussian information in an essential way. Such construction is needed for the study of stochastic PDEs, where the state space of the Gaussian processes is infinite-dimensional. This is a little beyond the scope of this course, and we will not go into more details other than Definition 2.4. Interesting readers can take a look at [PZ14, Chap. 2] or [Hai, Chap. 2-3].

With the Wiener measure at hand, we can now think of Brownian motion as random continuous functions. We conclude by mentioning the Hölder-continuity property of Brownian motion.

Definition 1.8 *Let $\alpha \in (0,1]$. A continuous function f is called (locally) α-Hölder if every x,*

$$\sup_{y:\ y\neq x} \frac{|f(x)-f(y)|}{|x-y|^\alpha} < \infty.$$

The α-Hölder continuous functions on $[0,T]$ form a complete metric space $\mathcal{C}^\alpha[0,1] \subset \mathcal{C}[0,1]$ under the norm:

$$|f|_{\mathcal{C}^\alpha} = \sup_x |f(x)| + \sup_{x\neq y} \frac{|f(x)-f(y)|}{|x-y|^\alpha}.$$

Theorem 1.3 *For $\alpha \in (0,1/2)$, the Wiener measure P^W is supported on α-Hölder continuous functions, that is,*

$$\forall \alpha \in (0,1/2), \quad \mathsf{P}^W(\omega \in \mathcal{C}^\alpha[0,1]) = 1.$$

Remark 1.11 One can show that for every $\alpha \in (0,1]$, the set of α-Hölder continuous function in $\mathcal{C}[0,1]$ is in $\mathcal{B}(\mathcal{C}[0,1])$, using that fact that a continuous function can be determined by its values on rational points.

1.3 Stochastic integrals and SDEs

Denote by $x(t)$ the position of a particle at time t. The *Langevin dynamics* of the particle is described by the equation

$$m\ddot{x}(t) = -\big(\nabla U\big)\big(x(t)\big) - \gamma\dot{x}(t) + c\eta(t).$$

The equation arises from Newton's second law:

- $m\ddot{x}(t)$ is the mass multiplied by the acceleration. It should be equal to the force, which is the right hand side of the equation.
- U is the potential, and $-\big(\nabla U\big)\big(x(t)\big)$ gives the potential force.
- $-\gamma\dot{x}(t)$ represents the friction which is usually proportional to the velocity $\dot{x}(t)$.
- $c\eta(t)$ is the random forcing, with c controlling its magnitude.

In an ideal physical model, $\eta(t)$ is the so-called *white noise*. As a "stochastic process", it should have at least the following two properties.

- **independence** $\eta(t)$ should be independent over disjoint intervals, namely, if I_1 and I_2 are two disjoint intervals of $\mathbb{R}$, then the two σ-fields

$$\sigma\big(\eta(t),\ t \in I_1\big), \quad \sigma\big(\eta(t),\ t \in I_2\big)$$

are independent.

- **stationarity** the one-dimensional distribution of $\eta(t)$ does not change:

$$\mathcal{L}\big(\eta(t_1)\big) = \mathcal{L}\big(\eta(t_2)\big), \quad \forall t_1 \neq t_2.$$

Brownian motion in fact got its name from the botanist Robert Brown who observed the motion of pollen of plants through a microscope. For things like the pollen, the term $m\ddot{x}(t)$ is negligible compared to other terms since m is so small, the above equation can be approximated by the *overdamped Langevin dynamics*:

$$\dot{x}(t) = -(\nabla u)\big(x(t)\big) + \eta(t) \tag{1.4}$$

For simplicity, we will set all constants (c, γ, etc) to 1 hereafter.

Free motion case. Let us set $U \equiv 0$ in (1.4). This means that no external potential (such as the gravity) is taking effect. We can simply integrate (1.4) to obtain (assuming $x(0) = 0$)

$$x(t) = \int_0^t \eta(s)\, ds.$$

The function $t \mapsto x(t)$ is just the trajectory of a randomly moving light-weighted particle. Based on our assumption on the white noise $\eta(t)$, its antiderivative $x(t)$ will satisfy

- $t \mapsto x(t)$ is continuous; this is really a physical constraint.
- $x(t)$ has independent increments: for all $0 = t_0 \leq t_1 < \cdots < t_m$, $\{x(t_{i+1}) - x(t_i)\}_{1 \leq i \leq m}$ are independent.
- The increments are centered Gaussian: $x(t) - x(s) \sim \mathcal{N}(0, \sigma_{t-s}^2)$. This is because any increment can be written as i.i.d. sums of small r.v.'s:

$$x(t) - x(s) = \sum_{i=0}^{N-1} x(t_{i+1}) - x(t_i), \quad t_i = s + \frac{i(t-s)}{N}.$$

Moreover, due to stationarity, it only makes sense to have σ_{t-s}^2 to be linear: $\sigma_{t-s}^2 = K \cdot (t-s)$ for some constant $K > 0$.

Up to a constant, the only process that satisfies all these conditions is Brownian motion. This means the write noise $\eta(t)$ should be interpreted as the "derivative" of Brownian motion. However, there is one fundamental issue of such interpretation:

Question *The Brownian motion is only α-Hölder continuous for $\alpha < 1/2$. In fact it is nowhere monotone and nowhere differentiable (we will see proofs of these statements later on). Then how should we define $\eta(t) = \dfrac{dB_t}{dt}$?*

The $U \not\equiv 0$ case. Let us consider a more general form

$$\dot{x}(t) = b\big(x(t)\big) + \eta(t), \tag{1.5}$$

where $b : \mathbb{R} \to \mathbb{R}$ is a sufficiently nice function. We are now entering the realm of the *stochastic differential equation (SDE)* . It has a lot of applications in other fields, for example stable diffusion in text-to-image AI models. As we mentioned above, $\eta(t)$ is not a function. At best it could be defined as a generalized function (viewed as a linear functional acting on $\mathcal{C}_0^\infty(\mathbb{R})$). Due to the special structure of (1.5), this issue could be circumvented by considering the equivalent integral equation

$$x(t) = x(0) + \int_0^t b\big(x(s)\big)\, ds + B(t). \tag{1.6}$$

Now the noise enters the equation as a Brownian motion $B(t)$, which is a random continuous function. All terms in (1.6) make sense as long as $x(t)$ is a continuous function. Then standard fixed-point or Picard-iteration techniques can be applied here to construct a unique solution $x(t)$.

First variation of (1.5): the magnitude of the noise is time-dependent. Let us consider

$$\ddot{x}(t) = b\big(x(t)\big) + f(t)\eta(t),$$

where $f(t)$ is a nice (say bounded and smooth) function. Inspired from the integral equation, it suffices to define the so-called *stochastic integral*

$$\int_0^t f(s)\eta(s)\ ds := \int_0^t f(s)\, dB(s) \tag{1.7}$$

The notation on the right hand side is to mimic that of the Riemann–Stieltjes integral. We recall its definition below.

Definition 1.9 *Let g be a function of finite variation (i.e., $g = g^+ - g^-$, where both g^+ and g^- are increasing) and f be a continuous function. Then the* Riemann–Stieltjes integral $\int f dg$ *is defined as*

$$\int_a^b f(s)\, dg(s) := \lim_{|\Delta|\to 0} \sum_{i=1}^N f(\xi_i)\big(g(t_{i+1}) - g(t_i)\big), \tag{1.8}$$

where $\Delta : a = t_0 < t_1 < \cdots < t_N = b$ is a partition, $\xi_i \in (t_i, t_{i+1})$ is arbitrary, and $|\Delta| = \max |t_{i+1} - t_i|$. The limit does not depend on the sequence of partitions or (ξ_i) that are chosen.

Example 1.12 When $g(t) = t$, the Riemann–Stieltjes integral is just the Riemann integral.

A nice thing about the Riemann–Stieltjes integral is that integration by parts holds.

Proposition 1.4 *Let f, g be functions of bounded variation. Then*

$$\int_a^b f(t)\, dg(t) = f(b)g(b) - f(a)g(a) - \int_a^b g(t)\, df(t).$$

Homework Use the Abel transformation (summation by parts)

$$\sum_{k=1}^n u_k(v_{k+1} - v_k) = u_{n+1}v_{n+1} - u_1 v_1 - \sum_{k=1}^n v_{k+1}(u_{k+1} - u_k)$$

to show that integration by parts holds for Riemann–Stieltjes integrals for functions f and g of bounded variation.

Of course, Brownian motion does not have bounded variation; such property is almost requiring differentiability. However, we can still use the idea of integration by parts to define simple stochastic integrals in the form of (1.7) by

$$\int_0^t f(s)\, dB_s := f(t)B_t - \int_0^t B_s\, df(s).$$

It requires only that f has bounded variation.

In fact, the integration-by-part formula suggests a trade-off between the regularities of f and g. A further generalization of Riemann–Stieltjes integral is the *Young's integral*, which says that (1.8) makes sense for $f \in \mathcal{C}^\alpha$, $g \in \mathcal{C}^\beta$ with $\alpha + \beta > 1$. Intuitively, the Riemann–Stieltjes integral corresponds roughly to the case $\alpha = 0$ and $\beta = 1$.

Second variation of (1.5): the magnitude of the noise is both time- and space-dependent.
We are now consider the SDE

$$\ddot{x}(t) = b\big(x(t)\big) + \sigma\big(t, x(t)\big)\eta(t), \tag{1.9}$$

where both b, σ are smooth. Again, with the integral form of the SDE, it all boils down to defining the stochastic integral

$$\int_0^t \sigma\big(s, x(s)\big)\, dB_s. \tag{1.10}$$

We already know that $t \mapsto B_t$ is $\mathcal{C}^\alpha$ with $\alpha < 1/2$. We also note that $x(t)$ cannot be more regular than $B(t)$, and hence no matter how smooth the function σ is, the map $t \mapsto \sigma\big(t, x(t)\big)$ is at most $\mathcal{C}^\beta$ with $\beta < 1/2$. One such simple example is $\displaystyle\int_0^t B_s\, dB_s$. Therefore, it is hopeless to define (1.10) even as a Young's integral, since $\alpha + \beta < 1$. This is as far as classical analysis can take us to. It tells us that the stochastic integral (1.10) cannot be defined for a fixed realization of (B_t). In fact, it could only be defined (or constructed) as a new stochastic process with the help of some new probabilistic tools.

To summarize, two central goals of this course are

1. Define the stochastic integral

   $$\int_0^t Y_s\, dB_s$$

   for very *irregular* stochastic processes $Y = (Y_t)_{t \geq 0}$.

   Again,we emphasize that if $Y \in \mathcal{C}^\beta$, $\beta > 1/2$, then the stochastic integral can be defined for every fixed realization of Brownian motion, but such treatment cannot cover even the simple case where $Y_t = B_t$ itself.

2. Develop a good solution theory for the SDE (1.9).

2 Construction and properties of Brownian motion

2.1 Gaussian r.v.'s and vectors

Gaussianity is crucial in the study of Brownian motion. In many ways, Brownian motion can be seen as a generalization of Gaussian vectors. In this section, we review some basic facts about Gaussian r.v.'s and vectors.

We begin with the definition of a (generalized) Gaussian r.v.

Definition 2.1 *Let $\mu \in \mathbb{R}$ and $\sigma \geq 0$. A Gaussian r.v. X with $\mathcal{N}(\mu, \sigma^2)$ distribution is characterized by any of the following:*

1) Its characteristic function is $\varphi_X(\xi) = \mathbb{E}e^{i\xi X} = e^{i\mu\xi - \frac{\sigma^2}{2}\xi^2}$.

2) $\mathcal{L}(X) = \mathcal{L}(\mu + \sigma \cdot Y)$, where $Y \sim \mathcal{N}(0,1)$ is the standard normal, a r.v. with density $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

3) If $\sigma \neq 0$ (non-degenerate case), then X is a continuous r.v. with density $\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$; if $\sigma = 0$, then $\mathsf{P}(X = 0) = 1$.

Proposition 2.1

1. *If X is a Gaussian r.v. on $(\Omega, \mathcal{F}, \mathsf{P})$, then $X \in L^p(\Omega, \mathcal{F}, \mathsf{P})$, $\forall p \in (0, \infty)$. In particular, for $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mathsf{E}X = 0$ and $\mathrm{Var}(X) = \sigma^2$.*

2. *If $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ and X_i are independent, then $X_1 + X_2 + \cdots + X_n \sim \mathcal{N}(\mu_1 + \cdots + \mu_n, \sigma_1^1 + \cdots + \sigma_n^2)$.*

Proof: The proof is elementary.

1. Direct computation using the Gaussian density.

2. Use the ch.f. of Gaussian r.v.'s.

$\square$

Gaussian r.v.'s have nice properties as elements in $L^2(\Omega, \mathcal{F}, \mathsf{P})$.

Proposition 2.2 *If $X_m \sim \mathcal{N}(\mu_m, \sigma_m^2)$ and $X_m \to X$ in $L^2(\Omega, \mathcal{F}, \mathsf{P})$, then $X \sim \mathcal{N}(\mu, \sigma^2)$ with*

$$\mu = \lim_{m\to\infty} \mu_m, \quad \sigma = \lim_{m\to\infty} \sigma_m. \tag{2.1}$$

Moreover, $X_m \to X$ in $L^p(\Omega, \mathcal{F}, \mathsf{P})$ for any $p > 0$.

Proof: The L^2-convergence of $X_m \to X$ implies the existence of both limits in (2.1). Hence, for each $\xi \in \mathbb{R}$, we have $\varphi_{X_m}(\xi) \to \exp(i\mu\xi - \dfrac{\sigma^2\xi^2}{2})$, which is the ch.f. of $\mathcal{N}(\mu, \sigma^2)$-Gaussian. On the other hand, the L^2-convergence of $X_m \to X$ also implies that $X_m \to X$ in probability, and thus in distribution. so $\varphi_{X_m}(\xi) \to \varphi_X(\xi)$. Therefore, $\varphi_X(\xi) = \exp(i\mu\xi - \dfrac{\sigma^2\xi^2}{2})$, and X indeed has $\mathcal{N}(\mu, \sigma^2)$ distribution, with μ, σ given by (2.1).

For any $q > 0$, it is easy to get a uniform upper bound by direct computation:

$$\sup_m \mathsf{E}|X_m - X|^q \leq C = C(\sup_m \mu_m, \sup_m \sigma_m).$$

By choosing $q > p$, we see that $|X_m - X|^p$ is uniformly integrable. Since $|X_m - X| \to 0$ in probability, this and uniform integrability imply (see [Dur07, Sec. 4.5]) that $\mathsf{E}|X_m - X|^p \to 0$. $\square$

Definition 2.2 *A random vector $X \in \mathbb{R}^d$ is Gaussian if for all $v \in \mathbb{R}^d$, $\langle v, X\rangle$ is a Gaussian r.v.*

Example 2.1

1. $X = (X_1, \ldots, X_d)$ where all X_i's are independent Gaussian random variables.

2. Let $X \in \mathbb{R}^d$ be Gaussian and Q be a $d \times d$ matrix. Then $Y = QX$ is Gaussian, since $\langle v, QX\rangle = \langle Q^T v, X\rangle$ for any vector v.

3. Let $(B_t)_{t\geq 0}$ be Brownian motion. For any $0 \leq t_1 < t_2 < \cdots < t_m$, both random vectors

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

are Gaussian.

Definition 2.3 *A stochastic process $(X_t)_{t\in T}$ is a* Gaussian process *if for any $t_1, t_2, \ldots, t_m \in T$, $(X_{t_1}, \ldots, X_{t_m})$ is a Gaussian vector.*

Example 2.2 The Brownian motion is a (centered) Gaussian process.

Theorem 2.3 *Each of the following is an equivalent definition for a random vector $X \in \mathbb{R}^d$ to be Gaussian.*

1. *There exists $\mu_X \in \mathbb{R}^d$ and a non-negative quadratic form $Q : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ such that the ch.f. of X is*
$$\varphi_X(\xi) = \mathsf{E}e^{i\langle \xi, X\rangle} = e^{i\langle \mu_X, X\rangle - \frac{1}{2}Q(\xi,\xi)}.$$

2. *There exists $\mu_X \in \mathbb{R}^d$, an orthonormal basis (ONB) $\{b_1, \ldots, b_d\}$, and $\varepsilon_1 \geq \varepsilon_2 \geq \cdots \geq \varepsilon_r > 0 = \varepsilon_{r+1} = \cdots = \varepsilon_d$ such that*
$$X \overset{d}{=} Y = \mu_X + \sum_{i=1}^{r} \varepsilon \eta_i \cdot b_i, \quad \eta_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1). \tag{2.2}$$

Proof: **From Definition 2.2 to Item 1.** Since $\langle \xi, X\rangle$ is Gaussian for every $\xi \in \mathbb{R}^d$, we have

$$\varphi_X(\xi) = \mathsf{E}e^{i\langle \xi, X\rangle} = e^{i\mathsf{E}\langle \xi, X\rangle - \frac{1}{2}\operatorname{Var}(\langle \xi, X\rangle)}.$$

We can take $\mu_X = \mathsf{E}X$ (coordinate-wise) so that $\mathsf{E}\langle \xi, X\rangle = \langle \xi, \mu_X\rangle$, and take

$$Q(\xi, \zeta) = \operatorname{Cov}(\langle \xi, X\rangle, \langle \zeta, X\rangle).$$

It is easy to check that $Q(\cdot, \cdot)$ is bilinear, symmetric, and defines a non-negative quadratic form on $\mathbb{R}^d$.

From Item 1 to Item 2. Since Q is a non-negative quadratic form, it can be diagonalized in an ONB $\{b_1, b_2, \ldots, b_d\}$ with eigenvalues $\varepsilon_i^2 \geq 0$:

$$Q(\xi, \zeta) = \sum_{i=1}^{d} (\varepsilon_i)^2 \langle \xi, b_i\rangle\langle \zeta, b_i\rangle.$$

(In matrix form, this is just $Q = B^T \Sigma B$ where $B = \{b_1, \ldots, b_d\}$ and $\Sigma = \operatorname{diag}\{\varepsilon_1^2, \ldots, \varepsilon_d^2\}$.) Without loss of generality we can take $\varepsilon_i \geq 0$ and order them from the largest to the smallest.

Suppose on some probability space we have i.i.d. $\mathcal{N}(0, 1)$ Gaussian r.v.'s η_i and let Y be defined by (2.2). For all $v \in \mathbb{R}^d$,

$$\langle v, Y\rangle = \sum_{i=1}^{r} \varepsilon_i \langle v, b_i\rangle \eta_i$$

is a sum of independent Gaussian r.v.'s, and hence is Gaussian. This verifies that Y is a Gaussian vector. Also, we have

$$\mathsf{E}\langle v, Y\rangle = \langle v, \mu_X\rangle, \quad \operatorname{Var}(\langle v, Y\rangle) = \sum_{i=1}^{r} \varepsilon_i^2 \langle v, b_i\rangle^2 = Q(v, v).$$

So X and Y have the same ch.f., and hence $\mathcal{L}(X) = \mathcal{L}(Y)$ as desired.

From Item 2 to Definition 2.2. It is already done above. $\square$

A Gaussian vector is non-degenerate if the quadratic form Q is non-degenerate, i.e., all eigenvalues are strictly positive. A non-degenerate Gaussian vector has a density, which is more familiar to most people.

Proposition 2.4 *A non-degenerate Gaussian vector $X \in \mathbb{R}^d$ has density*

$$p(x) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det(Q)}} e^{-\frac{1}{2}(x-\mu_X)^T Q^{-1}(x-\mu_X)},$$

where $Q = (Q_{ij}) = \big(\mathrm{Cov}(X_i, X_j)\big)$ is the covariance matrix.

Remark 2.3 Since the distribution of a Gaussian vector is determined by its covariance matrix, the f.d.d. of a centered Gaussian process $X = (X_t)_{t\in T}$ is completely determined by its covariance function

$$\Gamma(s,t) := \mathrm{Cov}(X_s, X_t) = \mathsf{E} X_s X_t, \quad s,t \in T.$$

For Brownian motion, $\Gamma(s,t) = s \wedge t$.

Homework Let X and Y be i.i.d. with $\mathsf{E}X = \mathsf{E}Y = 0$ and $\mathsf{E}X^2 = \mathsf{E}Y^2 = 1$. Suppose that the distribution of (X,Y) is rotational invariant, i.e.,

$$\mathcal{L}(X,Y) = \mathcal{L}(X\cos\theta + Y\sin\theta, -X\sin\theta + Y\cos\theta), \quad \forall \theta \in \mathbb{R}.$$

Show that $\mathcal{L}(X) = \mathcal{L}(Y) = \mathcal{N}(0,1)$.

Hint: rotational invariance implies that the ch.f. takes the form $\varphi_{X,Y}(\xi,\eta) = F(\xi^2 + \eta^2)$.

A Banach space is an infinite-dimensional vector space. The generalization of Gaussian vectors to the infinite dimension is *Gaussian measures on Banach spaces.*

Definition 2.4 (Gaussian measure on Banach spaces) *Let E be a separable Banach space. We say that an E-valued random element X has Gaussian distribution, if $\langle \lambda, X\rangle$ is a Gaussian r.v. for any linear functional $\lambda \in E^*$.*

Example 2.4 For Gaussian vectors in $\mathbb{R}^d$, $E = \mathbb{R}^d = E^*$, that is, any linear functional is the inner product with a fixed vector v. This is exactly Definition 2.2.

Example 2.5 For Brownian motion, $X = (B_t)_{t\in[0,1]}$, $E = \mathcal{C}[0,1]$, and E^* is the space of all finite signed measures on $[0,1]$. Then for $\lambda = \lambda(dt) \in E^*$, $\langle \lambda, X\rangle$ is a centered Gaussian with variance

$$\mathrm{Var}(\langle \lambda, X\rangle) = \mathsf{E}\int_0^1 \int_0^1 B_s\,\lambda(ds) B_t \lambda(dt) = \int_0^1 \Big[\mathsf{E}B_s B_t\Big]\,\lambda(ds)\lambda(dt) \int_0^1\int_0^1 (s\wedge t)\,\lambda(ds)\lambda(dt),$$

where in the last equality the exchange of integration and expectation needs justification.

For the construction of Brownian motion, the variance of $\langle \lambda, X\rangle$, $\lambda \in E^*$, will be given first, and then some general theory will guarantee the existence of a corresponding (centered) Gaussian measure as long as the variance functional induces a positive definite quadratic form, similar to Gaussian vectors.

Homework Let $f(t) = \lambda((t,1])$.

1. Suppose that $\lambda(dt) = \rho(t)\,dt$ for some $\rho \in \mathcal{C}[0,1]$. Show that

$$\int_0^1 \int_0^1 (s\wedge t)\,\lambda(ds)\lambda(dt) = \int_0^1 |f(t)|^2\,dt.$$

   Hint: use integration by parts.

2. (Optional) Prove the same identity for an arbitrary signed measure $\lambda(dt)$.

   Hint: if $\lambda(dt)$ is a signed measure, then f defined as above has bounded variation and $\lambda(dt) = d\big(-f(t)\big)$. Use integration by parts for Riemann–Stieltjes integrals.

2.2 Gaussian white noise

The goal of this section is to construct a centered Gaussian process $(B_t)_{t\in[0,1]}$ with covariance $\mathsf{E}B_tB_s = t \wedge s$. After the construction, the resulting process (called "pre-Brownian motion" in [LeG16]) may not be a.s. continuous; we will discuss how to get continuity in Sections 2.2 and 2.3.

The Kolmogorov's Extension Theorem ([Shi96, Chap. II.3, Theorem 4]) already guarantees the existence of a stochastic process with any prescribed *consistent* f.d.d. However, in the special case of Brownian motion, it is advantageous to have a more explicit construction using the Gaussian white noise.

Surprisingly, it is more convenient to first define a more general stochastic integral $G(f) = \displaystyle\int_0^1 f(t)dB_t$, and then define Brownian motion as a special stochastic integral

$$B_t = \int_0^1 \mathbb{1}_{[0,t]}(s)\, ds.$$

The following discussion shows that the natural class of functions to define $G(f)$ is $L^2[0,1]$, and for such f, $G(f)$ is in fact a Gaussian r.v. This will also motivate the introduction of Gaussian white noise, and the definition of Itô integrals later.

First: f piecewise constant

Suppose that $[0,1]$ is partitioned into $0 = t_0 < t_1 < \cdots < t_m = 1$ and $f(s) = \displaystyle\sum_{i=0}^{m-1} f_i \mathbb{1}_{[t_i,t_{i+1})}(s)$. Then in light of the Riemann–Stieltjes integral, it only makes sense to define $G(f)$ as

$$G(f) := \sum_{i=0}^{m-1} f_i \cdot (B_{t_{i+1}} - B_{t_i}). \tag{2.3}$$

We did not specify $f(1)$, but it does not enter the definition of (2.3) anyway, so it is safe to ignore it. The r.v. in (2.3) is a sum of i.i.d. Gaussian r.v.'s, so it is also Gaussian. It has zero mean, and a variance

$$\operatorname{Var}\big(G(f)\big) = \sum_{i=0}^{m-1} f_i^2(t_{i+1} - t_i) = \int_0^1 |f(t)|^2 dt$$

Second: difference of $G(f_1)$ and $G(f_2)$ for piecewise constant f_i.
Without loss of generality we can assume that f_1 and f_2 has the same partition of $[0,1]$, since otherwise we can enlarge their partitions to a common partition by including all the endpoints. Then, a similar computation yields that $G(f_1) - G(f_2)$ is also a centered Gaussian, with variance

$$\mathsf{E}|G(f_1) - G(f_2)|^2 = |f_1 - f_2|^2_{L^2[0,1]}.$$

Last: general $f \in L^2[0,1]$
Every function $f \in L^2[0,1]$ can be approximated by piecewise functions f_n in $L^2[0,1]$. One way to see is to first approximate any $L^2[0,1]$ function by continuous functions, then to approximate continuous functions by piecewise constant functions. Suppose that $f_n \to f$ in $L^2[0,1]$ and f_n are all piecewise constant. Note that

$$|G(f_n) - G(f_m)|^2_{L^2(\Omega,\mathcal{F},\mathsf{P})} = \mathsf{E}|G(f_n) - G(f_m)|^2 = |f_n - f_m|^2_{L^2[0,1]}$$

Since $f_n \to f$, (f_n) is a Cauchy sequence in $L^2[0,1]$, and hence $\big(G(f_n)\big)$ is a Cauchy sequence in $L^2(\Omega,\mathcal{F},\mathsf{P})$. But $L^2(\Omega,\mathcal{F},\mathsf{P})$ is a complete metric space, which means every Cauchy sequence has a limit; let us denote the limit of $G_N(f)$ by $G(f)$. Note that all $G(f_n)$ are Gaussian, so by Proposition 2.2, the limit $G(f)$ is also Gaussian.

Definition 2.5 (Gaussian white noise) *Let $(E, \mathcal{E})$ be a measurable space, μ be a σ-finite measure on $(E, \mathcal{E})$. Denote by $H = L^2(E, \mathcal{E}, \mu)$. A* Gaussian white noise *(with intensity μ) is an isometry (i.e., preserving the inner product between two inner product spaces) from H to $L^2(\Omega, \mathcal{F}, \mathsf{P})$ with values being (centered) Gaussian r.v.'s. The isometry is given by*

$$G : f \mapsto G(f) \sim \mathcal{N}(0, |f|_H^2).$$

Theorem 2.5 *If the Hilbert space $H = L^2(E, \mathcal{E}, \mu)$ is separable, then there exists a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ such that the Gaussian white noise $G : H \to L^2(\Omega, \mathcal{F}, \mathsf{P})$ exists.*

Remark 2.6 A Hilbert space is an inner product space which is also complete. One can think of a Hilbert space as an infinite-dimensional Euclidean space. All L^2-spaces are Hilbert space by standard real analysis. "Separable" means that there is a dense countable set, which is true when $H = L^2([0, 1])$.

In proving the theorem, the ONLY thing we will use about a separable Hilbert space is the existence of an ONB.

Proposition 2.6 *If H is a separable Hilbert space, then there exist $(e_n)_{n\geq 1} \subset H$, such that*

- $\langle e_n, e_m \rangle = \mathbb{1}_{n=m}$.
- *(basis) for every $f \in H$, it can be written as*

$$f = \sum_{n=1}^{\infty} \langle e_n, f \rangle f_n,$$

where the infinite sum is converging in H.

Such collection $(e_n)_{n\geq 1}$ is called an orthonormal basis *of H.*

Proof of Theorem 2.5: Pick an ONB $(e_n)_{n\geq 1}$ for $H = L^2(E, \mathcal{E}, \mu)$. Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space on which there are i.i.d. $\mathcal{N}(0, 1)$ r.v.'s ξ_n, $n \geq 1$. Let us define

$$G_N(f) = \sum_{n=1}^{N} \xi_n \langle e_n, f \rangle.$$

Then $G_N(f)$, $N \geq 1$, each being a sum of independent Gaussians, are all Gaussian. Also, for $N < N'$,

$$\mathsf{E}|G_N(f) - G_{N'}(f)|^2 = \sum_{N\leq n<N'} |\langle e_n, f \rangle|^2.$$

Since $f \in H = L^2(E, \mathcal{E}, \mu)$ and $|f|_H^2 = \sum_{n=1}^{\infty} |\langle e_n, f \rangle|^2 < \infty$, $\{G_N(f)\}_{N\geq 1}$ is Cauchy in $L^2(\Omega, \mathcal{F}, \mathsf{P})$. Therefore, the following limit in $L^2(\Omega, \mathcal{F}, \mathsf{P})$

$$G(f) = \lim_{N\to\infty} G_N(f) = \sum_{n=1}^{\infty} \xi_n \langle e_n, f \rangle \tag{2.4}$$

exists. Since $G(f)$ is the L^2-limit of Gaussians, it is also Gaussian; moreover, by Proposition 2.1, it has distribution $\mathcal{N}(0, |f|_H^2)$. □

Example 2.7 A Gaussian vector in $\mathbb{R}^d$ is also associated with a Gaussian white noise expansion, with $H = (\mathbb{R}^d, |\cdot|_H)$, and

$$|v|_H^2 = v^T Q v = \sum_{i=1}^{r} \varepsilon_i^2 |\langle v, b_i \rangle|^2.$$

Compare with Item 2 in Theorem 2.3.

Example 2.8 $H = L^2(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}), dt)$. Then $B_t = G(\mathbb{1}_{[0,t]})$ is a centered Gaussian process, with covariance

$$\mathsf{E} B_t B_s = \int_0^\infty \mathbb{1}_{[0,t]}(r)\mathbb{1}_{[0,s]}(r)\, dr = s \wedge t.$$

That is, $(B_t)_{t\geq 0}$ has the same f.d.d. as Brownian motion.

The definition of Gaussian white noise only shows B_t is Gaussian for a fixed t. To see that any f.d.d. is jointly Gaussian, we need to do a little bit more work. This can be also derived from the definition of Gaussian white noise. In fact, any isometry between Hilbert spaces must be linear, so for any $t_1 < \cdots < t_m$ and $v_1, \ldots, v_m$,

$$v_1 B_{t_1} + \cdots + v_m B_{t_m} = G\Big(\sum_{i=1}^{m} v_i \mathbb{1}_{[0,t_i]}\Big)$$

is indeed Gaussian. The covariance computation from variance is a consequence of applying the following *polarization identity* to the inner product spaces $L^2(\Omega, \mathcal{F}, \mathsf{P})$ and $L^2[0,1]$:

$$4\langle f, g\rangle = \langle f+g, f+g\rangle - \langle f-g, f-g\rangle.$$

Remark 2.9 Use the GWN construction of BM, for $f \in L^2[0,\infty)$,

$$\mathsf{E}\Big|\int_0^\infty f(t)\, dB_t\Big|^2 = \mathsf{E}\big|G(f)\big|^2 = \int_0^\infty f^2(t)\, dt. \tag{2.5}$$

This is the simplest form of the celebrated "Itô's Isometry".

2.3 Continuity of Brownian motion via Kolmogorov's Continuity Theorem

A powerful tool to get continuous modification of a stochastic process is the celebrated Komolgorov Continuity Theorem. It extracts information of path regularity from the f.d.d.

Theorem 2.7 *Let $(X_t)_{t\in[0,T]}$ be a stochastic process that satisfies*

$$\mathsf{E}|X_t - X_s|^\alpha \leq K|t-s|^{1+\beta}, \quad \forall 0 \leq s, t \leq T.$$

Then X has a modification $\tilde{X}$ which is γ-Hölder continuous for all $\gamma < \beta/\alpha$.

Example 2.10 Let $(B_t)_{t\in[0,1]}$ be a Gaussian process with $\mathsf{E}B_tB_s = t\wedge s$. Then $B_t - B_s \sim \mathcal{N}(0, t-s)$, and hence $\mathsf{E}|B_t - B_s|^n \leq K_n(t-s)^{n/2}$ for all $n \geq 1$. Since $\dfrac{n/2-1}{n}$ can be arbitrarily close to $1/2$, (B_t) has a modification which is γ-Hölder for all $\gamma < 1/2$.

We first reduce Theorem 2.7 to the case of a fixed γ.

Lemma 2.8 *If X and Y are continuous stochastic processes on $\mathbb{R}$, and Y is a modification of X, then Y is a version of X.*

Proof: By the definition of modifications, $\mathsf{P}(X_t = Y_t) = 1$ for all $t \in \mathbb{R}$. Since the set of rational numbers $\mathbb{Q}$ is countable, we have $\mathsf{P}(X_t = Y_t,\ \forall t \in \mathbb{Q}) = 1$. That is, there is a set $\mathcal{N}$ with probability $\mathsf{P}(\mathcal{N}) = 0$, such that for all $\omega \in \mathcal{N}$,

$$X_t(\omega) = Y_t(\omega), \quad \forall t \in \mathbb{Q}. \tag{2.6}$$

Noting that $t \mapsto X_t(\omega)$ and $t \mapsto Y_t(\omega)$ are always continuous. Hence, if for any ω the condition (2.6) holds, then it follows that

$$X_t(\omega) = Y_t(\omega), \quad \forall t \in \mathbb{R}. \tag{2.7}$$

So (2.7) holds except on a null-set $\mathcal{N}$; this means that Y is a version of X. □

Lemma 2.9 *For Theorem 2.7, it suffices to prove it for any fixed $\gamma < \alpha/\beta$.*

Proof: Suppose that there are modifications $X^{(n)}$ of X which is $\gamma_n = (\alpha/\beta - 1/n)$-Hölder continuous. Then by Lemma 2.8, $X^{(n)}$, $n \geq 1$, are all versions of each other. In particular, there exist null-sets $\mathcal{N}^{(n)}$ such that

$$\forall \omega \in (\mathcal{N}^{(n)})^c : \qquad X_t^{(1)} = X_t^{(n)}, \quad t \in [0, T].$$

Let $\mathcal{N} = \bigcup_{n \geq 2} \mathcal{N}^{(n)}$. Then $\mathcal{N}$ is also a null-set, and for all $\omega \in \mathcal{N}^c$, $X_t^{(1)} = X_t^{(n)}$, $\forall n, t$. Hence, $X^{(1)}$ is γ_n-Hölder for all $n \geq 1$ on the set $\mathcal{N}$. Since γ_n is arbitrarily close to α/β, $X^{(1)}$ is γ-Hölder for any $\gamma < \alpha/\beta$ on $\mathcal{N}$. The proof is complete. □

Proof of Theorem 2.7: Without loss of generality set $T = 1$. Let $\gamma < \beta/\alpha$.

By Markov inequality,

$$\mathsf{P}\Big(|X_{k/2^n} - X_{k-1/2^n}| > 2^{-\gamma n}\Big) \leq K \frac{(1/2^n)^{1+\beta}}{2^{-\gamma n \alpha}} = K 2^{-n(1+\beta-\alpha\gamma)}.$$

By a union bound,

$$\mathsf{P}\big(\sup_{1 \leq k \leq 2^n} |X_{k/2^n} - X_{k-1/2^n}| > 2^{-\gamma n}\big) \leq K \cdot 2^{-(\beta-\alpha\gamma)n}.$$

Since $\sum_{n=1}^{\infty} 2^{-(\beta-\alpha\gamma)n} < \infty$, by Borel–Cantelli, there exists $n_0 = n_0(\omega)$ such that for $n \geq n_0$,

$$|X_{k/2^n} - X_{(k-1)/2^n}| \leq 2^{-\gamma n}, \quad \forall 1 \leq k \leq 2^n. \tag{2.8}$$

Claim: for a.e. ω, X is uniformly γ-Hölder continuous on $D = \cup D_n = \bigcup(\mathbb{Z}/2^n \cap [0,1])$, that is, there exists $M = M(\omega) > 0$ such that

$$|X_s - X_t| < M|t-s|^\gamma, \quad \forall t, s \in D.$$

Assume that the claim is proved. Noting that D is dense in $[0,1]$, we can define

$$\tilde{X}_t = \begin{cases} X_t, & t \in D, \\ \lim\limits_{D \ni t_m \to t} X_{t_m}, & t \notin D. \end{cases}$$

By the uniform γ-Hölder continuity, the limit is independent of (t_m), and the resulting $\tilde{X}_t$ is γ-Hölder continuous with the same constant $C(\omega)$.

Now we turn to the proof of the claim.

Let $t \in [\frac{k}{2^n}, \frac{k+1}{2^n}] \cap D$, $0 \le k \le 2^n - 1$, $n \ge n_0$. Then there exist a sequence $k/2^n = p_n/2^n$, $p_{n+1}/2^{n+1}, \dots, p_N/2^N = t$ such that

$$\left|\frac{p_m}{2^m} - \frac{p_{m+1}}{2^{m+1}}\right| = \frac{1}{2^{m+1}}, \quad n \le m < N.$$

By triangle inequality and (2.8),

$$|X_t - X_{k/2^n}| \le \sum_{m=n}^{N-1} |X_{p_m/2^m} - X_{p_{m+1}/2^{m+1}}| \le \sum_{m=n}^{\infty} 2^{-\gamma m} = \frac{2^{-\gamma n}}{1-2^{-\gamma}}. \tag{2.9}$$

In particular, this and triangle inequality imply that X_t is bounded on $t \in D$. Let $M_0(\omega) = \sup_D X_t$.

For every $s < t$ in D, we can find the biggest n such that

$$\frac{k-1}{2^n} \le s < \frac{k}{2^n} \le t < \frac{k+1}{2^n},$$

and such n necessarily satisfies

$$\frac{1}{2^{n+1}} \le |t-s| \le \frac{1}{2^{n-1}}. \tag{2.10}$$

There are two cases.

Case 1: $n < n_0$**.** Since $|t-s| \ge 2^{-n_0}$, we have

$$\frac{|X_t - X_s|}{|t-s|^\gamma} \le \frac{2M_0}{(2^{-n_0})^\gamma} =: M_1(\omega).$$

Case 2: $n \ge n_0$**.** By triangle inequality, (2.9) and (2.10), we have

$$|X_s - X_t| \le |X_s - X_{k/2^n}| + |X_{k/2^n} - X_t| \le \frac{2^{-\gamma n+1}}{1-2^{-\gamma}} \le \frac{2}{1-2^{-\gamma}}\big(2|t-s|\big)^\gamma =: M_2|t-s|^\gamma.$$

Let $M = \max(M_1(\omega), M_2)$. Then $|X_t - X_s| \le M|t-s|^\gamma$ for all $t, s \in D$. The claim is proved. □

Homework The *Brown sheet* $(\mathbb{B}_{s,t})_{s,t\in[0,1]}$ is a centered Gaussian process with covariance

$$\mathsf{E}\mathbb{B}_{s,t}\mathbb{B}_{s',t'} = (s \wedge s')(t \wedge t'), \quad s,t,s',t' \in [0,1].$$

It can be constructed via GWN with $H = L^2([0,1]^2, \mathcal{B}([0,1]^2), ds \times dt)$ and $\mathbb{B}_{s,t} = G\big(\mathbb{1}_{[0,s]\times[0,t]}\big)$.

1. Show that for each $p \ge 1$, there is some constant $K_p > 0$,

$$\mathsf{E}|\mathbb{B}_{s,t} - \mathbb{B}_{s',t'}|^{2p} \le K_p\big(|s-s'|^p + |t-t'|^p\big), \quad s,t,s',t' \in [0,1].$$

2. Let $0 < \gamma < 1/2$. Show that with probability one, there is a random constant $n_0 = n_0(\omega)$ such that for all $n \ge n_0$,

$$\left|\mathbb{B}_{\frac{k}{2^n},\frac{\ell}{2^n}} - \mathbb{B}_{\frac{k'}{2^n},\frac{\ell'}{2^n}}\right| \le 2^{-\gamma n}, \quad 0 \le k,\ell,k',\ell' \le 2^n,\ |k-k'| + |\ell-\ell'| \le 1.$$

2.4 Lévy's construction of Brownian motion

Using the proof of Theorem 2.5, we can express Brownian motion explicitly in the form of (2.4). In fact, let $\{e_n\}$ be an ONB of $L^2([0,1], dt)$ and $\xi_n \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ on $(\Omega, \mathcal{F}, \mathsf{P})$. Then by Theorem 2.5,

$$B_t(\omega) = \sum_{n=1}^{\infty} \xi_n(\omega) \langle e_n(x), \mathbb{1}_{[0,t]}(x) \rangle \tag{2.11}$$

is a Gaussian process with the f.d.d. of a Brownian motion; moreover, the infinite sum converges in $L^2(\Omega, \mathcal{F}, \mathsf{P})$. But we cannot derive continuity of $t \mapsto B_t(\omega)$ for fixed ω.

Let us take a closer look at the infinite series (2.11). Note that $\beta_n(t) = \langle e_n(x), \mathbb{1}_{[0,t]}(x) \rangle$ is a deterministic, continuous function. Hence, for every fixed N,

$$B_t^N(\omega) = \sum_{n=1}^{N} \xi_n(\omega) \beta_n(t)$$

is also continuous in t for every ω. From classical analysis, for P-a.e. ω, if the Cauchy criterion holds:

$$\sup_{t \in [0,1]} |B_t^N - B_t^{N'}|(\omega) \to 0, \quad N, N' \to \infty, \tag{2.12}$$

then $(B_t^N(\omega))_{t \in [0,1]}$ converges uniformly to some (random) continuous function $\big(\tilde{B}_t(\omega)\big)_{t \in [0,1]}$. The two processes B and $\tilde{B}$ must have the same f.d.d., since for fixed t, $\tilde{B}_t$ is the a.s.-limit of B_t^N, while B_t is the L^2-limit of B_t^N; in other words, $\tilde{B}$ will be a continuous modification of B.

The usual approach to verify the Cauchy criterion is to use *Weierstrass M-test*, which is an estimate for absolute convergence:

$$\sup_{t \in [0,1]} |B_t^N - B_t^{N'}|(\omega) \leq \sum_{N \leq n < N'} |\xi_n| \sup_{t \in [0,1]} |\beta_n(t)|. \tag{2.13}$$

Since $\xi_n \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, it is easy to control the growth of ξ_n: by Borel–Cantelli and the Gaussian tail estimate $\mathsf{P}(|\mathcal{N}(0,1)| \geq a) \leq e^{-a^2/2}$, with probability one, there is a random constant $n_0 = n_0(\omega)$ s.t.

$$|\xi_n| \leq \ln n, \quad \forall n \geq n_0(\omega).$$

Therefore, to apply the M-test, all we need is

$$\sum_{n=1}^{\infty} \ln n \cdot \sup_{t \in [0,1]} |\beta_n(t)| < \infty. \tag{2.14}$$

Can (2.14) be true? Let us look at a common choice for ONB on $L^2[0,1]$ from Fourier series:

$$\{e_n(x)\} = \{1, \sqrt{2}\sin(2\pi n \cdot x), \sqrt{2}\cos(2\pi n \cdot x)\}.$$

For the corresponding $\beta_n(t)$, one has

$$\sup_{t \in [0,1]} |\beta_n(t)| \sim \frac{1}{n}.$$

Since $\displaystyle\sum_{n=1}^{\infty} \frac{\ln n}{n}$ diverges, the M-test cannot apply.

There are two fixes. The first one to choose $\{e_n(x)\}$ more cleverly, so the Cauchy criterion (2.12) holds. See Lévy's construction in the exercise below.

Homework For $n \geq 0$ and $0 \leq k \leq 2^n - 1$, let

$$e_{n,k}(x) = \begin{cases} 2^{\frac{n}{2}}, & \frac{k}{2^n} \leq x < \frac{2k+1}{2^{n+1}}, \\ -2^{\frac{n}{2}}, & \frac{2k+1}{2^{n+1}} \leq x < \frac{k+1}{2^n}, \\ 0, & \text{otherwise}, \end{cases} \qquad \beta_{n,k}(t) = \langle e_{n,k}, \mathbb{1}_{[0,t]} \rangle,$$

and $\xi_{n,k} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. Define $\Delta B_t^n = \sum_{k=0}^{2^n-1} \xi_{n,k}\beta_{n,k}(t)$ and $B_t^N = \sum_{n=0}^{N} \Delta B_t^n$.

1. Show that $\{e_{n,k}\}$ is orthonormal, i.e.,

$$\int_0^1 e_{n,k}(x)e_{n',k'}(x)\,dx = \mathbb{1}_{n=n'}\mathbb{1}_{k=k'}.$$

2. Show that

$$\sup_{t\in[0,1]} |\Delta B_t^n| \leq 2^{-n/2} \cdot \max_{0\leq k\leq 2^n-1} |\xi_{n,k}|.$$

Hint: note that for fixed n, $e_{n,k}$ has disjoint support for different k.

3. Use $\mathsf{P}(|\mathcal{N}(0,1)| \geq a) \leq e^{-a^2/2}$ and Borel–Cantelli Lemma to show that with probability one, there is a random constant $n_0 = n_0(\omega)$ such that

$$|\xi_{n,k}| \leq n, \quad \forall 0 \leq k \leq 2^n - 1,\ n \geq n_0.$$

4. Conclude that with probability 1, $\{B_t^N(\omega), t \in [0,1]\}_{N\geq 1}$ is Cauchy in $\mathcal{C}[0,1]$, that is,

$$\lim_{N,N'\to\infty} \sup_{t\in[0,1]} |B_t^N(\omega) - B_t^{N'}(\omega)| = 0, \quad \text{a.e. } \omega.$$

Another convenient description of Lévy's construction is the following. Let X_k be i.i.d. $\mathcal{N}(0,1)$ and $S_k = X_1 + \cdots + X_k$. Define

$$\tilde{S}_t = \begin{cases} S_k, & t = k \in \mathbb{Z}, \\ (t-k)S_{k+1} + (t+1-k)S_k, & t \in (k, k+1). \end{cases}$$

Then

$$B_t^N \overset{d}{=} \frac{\tilde{S}_{2^N t}}{2^{N/2}}.$$

In this representation, it is easy to verify that B^N has the same f.d.d. as Brownian motion at $t \in \mathbb{Z}/2^N$. By the Functional CLT, B^N converges to Brownian motion in distribution.

Another fix is to utilize the fluctuation of i.i.d. Gaussian and improve the bound on the right hand side of (2.13). As a comparison, recall the Kolmogorov's One-Series Theorem.

Theorem 2.10 *Let X_n be independent with $\mathsf{E}X_n = 0$ and $\sum_{n=1}^{\infty} \mathsf{E}X_n^2 < \infty$. Then $\sum_{n=1}^{\infty} X_n$ converges a.s.*

As a consequence of Theorem 2.10, we can put random ± 1 in front of $1/n$ and get a conditionally converging sum $\sum_{n=1}^{\infty} \frac{\pm 1}{n}$ since $\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$. However, $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$ so absolute convergence bound like (2.13) will fail.

In infinite dimension, the analogue is $\sum_{n=1}^{\infty} \beta_n^2 < \infty$ in the L^2-sense:

$$\sum_{n=1}^{\infty} \int_0^1 \beta_n^2(t)\, dt = \int_0^1 \sum_{n=0}^{\infty} \langle e_n, \mathbb{1}_{[0,t]}\rangle^2\, dt = \int_0^1 |\mathbb{1}_{[0,t]}|^2_{L^2[0,1]}\, dt = \int_0^1 t\, dt < \infty.$$

Some general theory about Gaussian measures is develop to guarantee that (2.11) always converges almost surely, whatever the choice of the ONB $\{e_n\}$, which is a refinement of the construction in Theorem 2.5 (see e.g. [PZ14, Part I, Theorem 2.12]).

3 Filtration and Markov property

3.1 Filtration and stopping times

Definition 3.1 *Let $(X_t)_{t\geq 0}$ be a stochastic process defined on $(\Omega, \mathcal{F}, \mathsf{P})$.*

1. *A filtration $(\mathcal{F}_t)_{t\geq 0}$ is a family of increasing sub–σ-field of $\mathcal{F}_t$, namely,*

$$\mathcal{F}_{t_1} \subset \mathcal{F}_{t_2} \subset \mathcal{F}, \quad \forall 0 \leq t_1 < t_2.$$

2. *X_t is said to be* adapted *to $(\mathcal{F}_t)_{t\geq 0}$, if X_t is measurable w.r.t. $\mathcal{F}_t$ for all $t \geq 0$.*

Example 3.1 (Natural filtration) Let $(X_t)_{t\geq 0}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathsf{P})$. The *natural filtration* is

$$\mathcal{F}_t^X := \sigma\big(X_s : 0 \leq s \leq t\big).$$

Roughly speaking, $\mathcal{F}_t^X$ is the information contained by the process X up to time t. By definition, X_t is $\mathcal{F}_t^X$-measurable, so X is $(\mathcal{F}_t^X)$-adapted.

Definition 3.2 *On the space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathsf{P})$,*

1. *a r.v. T is called a* stopping time *if $\{T \leq t\} \in \mathcal{F}_t$, $\forall t \geq 0$;*
2. *a r.v. T is called an* optional time *if $\{T < t\} \in \mathcal{F}_t$, $\forall t \geq 0$.*

There is a small difference between optional times and stopping times, but under mild assumptions they will be the same. We will see these assumptions by the end of this section. Nevertheless, the next two propositions give some relations between them.

Proposition 3.1 *If T is a stopping time, then T is also optional.*

Proof: We have

$$\{T < t\} = \bigcup_{n=1}^{\infty} \{T \leq t - \frac{1}{n}\} \in \sigma\big(\mathcal{F}_{t-\frac{1}{n}}, n \geq 1\big) \subset \mathcal{F}_t.$$

So T is optional. □

Let $\mathcal{F}_{t+} := \bigcap_{n=1}^{\infty} \mathcal{F}_{t+\frac{1}{n}} = \bigcap_{s>t} \mathcal{F}_s$. The two intersections are equivalent since $\mathcal{F}_t$ is a increasing in t.

Proposition 3.2 *If T is an optional time for $(\mathcal{F}_t)$, then it is a stopping time for $(\mathcal{F}_{t+})$.*

Proof: We have
$$\{T \le t\} = \bigcap_{n=1}^{\infty} \{T < t + \frac{1}{n}\} \in \bigcap_{n=1}^{\infty} \mathcal{F}_{t+\frac{1}{n}} = \mathcal{F}_{t+}.$$
□

Example 3.2 The most common examples of stopping times and optional times are the hitting time of a set. Let $\Gamma \subset \mathbb{R}$ and $(X_t)_{t\ge 0}$ be a $(\mathcal{F}_t)$-adapted process. Then
$$T_\Gamma = \inf\{s \ge 0 : X_s \in \Gamma\}.$$

Proposition 3.3

1. *If Γ is open and X has right-continuous sample paths, then T_Γ is optional.*
2. *If Γ is closed and X has continuous sample paths, then T_Γ is stopping.*

Proof:

1. For $t \ge 0$, we have
$$\{T_\Gamma < t\} = \{\exists s < t : \; X_s \in \Gamma\} = \{\exists q < t, q \in \mathbb{Q} : X_s \in \Gamma\} = \bigcup_{q\in\mathbb{Q}, q<t} \{X_q \in \Gamma\} \in \mathcal{F}_t,$$
where the first equality is due to the definition of infimum,and the second equality due to right-continuity of paths and openness of Γ.
2. For $t \ge 0$, we have
$$\{T_\Gamma > t\} = \{\{X_s\}_{s\in[0,t]} \cap \Gamma = \varnothing\} = \bigcup_{n=1}^{\infty} \{\mathrm{dist}(\{X_s\}_{s\in[0,t]}, \Gamma) \ge \frac{1}{n}\} = \bigcup_{n=1}^{\infty} \bigcap_{q\in[0,t]\cap\mathbb{Q}} \{\mathrm{dist}(X_q, \Gamma) \ge \frac{1}{n}\} \in \mathcal{F}_t.$$
The continuity of X implies that $\{X_s\}_{s\in[0,t]}$ is a compact set, and hence if it does not intersect a closed set Γ, it must have positive distance to Γ; this gives the second equality.

□

Definition 3.3 *A filtration $(\mathcal{F}_t)_{t\ge 0}$ is* right-continuous *if $\mathcal{F}_{t+} = \mathcal{F}_t$ for all $t \ge 0$.*

For a right-continuous filtration, stopping times and optional times are the same. An effortless way to get right-continuous filtration is just to replace $\mathcal{F}_t$ by $\mathcal{F}_{t+}$. Noting that since $\mathcal{F}_t \subset \mathcal{F}_{t+}$, if X_t is $(\mathcal{F}_t)$-adapted, then it is also $(\mathcal{F}_{t+})$-adapted.

Proposition 3.4 *Let $\mathcal{G}_t = \mathcal{F}_{t+}$. Then $(\mathcal{G}_t)_{t\ge 0}$ is right-continuous.*

Proof: We have
$$\mathcal{G}_{t+} = \bigcap_{n=1}^{\infty} \mathcal{G}_{t+\frac{1}{n}} = \bigcap_{n=1}^{\infty} \mathcal{F}_{(t+\frac{1}{n})+} \subset \bigcap_{n=1}^{\infty} \mathcal{F}_{t+\frac{2}{n}} = \mathcal{F}_{t+} = \mathcal{G}_t.$$
□

It is still a valid question to ask how much $\mathcal{F}_t$ is different from $\mathcal{F}_{t+}$. If the filtration is generated by a nice process like the Brownian motion, then the answer is that $\mathcal{F}_t$ and $\mathcal{F}_{t+}$ only differ by null sets. In the case $t = 0$, this can be formulated by the following zero-one law.

Theorem 3.5 (Blumenthal's 0-1 law) *Let $B = (B_t)_{\geq 0}$ be the standard Brownian motion and $\mathcal{F}_t^B$ be its natural filtration. Then $\mathcal{F}_{0+}^B$ is trivial, i.e., $\mathsf{P}(A) = 0$ or 1 for all $A \in \mathcal{F}_{0+}^B$.*

Remark 3.3 Since $B_0 = 0$ for all ω, $\mathcal{F}_0^B = \{\varnothing, \Omega\}$.

Proof: For any $A \in \mathcal{F}_{0+}^B$, $0 < t_1 < \cdots < t_m$ and bounded continuous $g : \mathbb{R}^m \to \mathbb{R}$, we have

$$
\begin{aligned}
\mathsf{E}\mathbb{1}_A g(B_{t_1}, \cdots, B_{t_m}) &= \lim_{n\to\infty} \mathsf{E}\mathbb{1}_A g(B_{t_1} - B_{1/n}, \cdots, B_{t_m} - B_{1/n}) \\
&= \mathsf{E}\mathbb{1}_A \lim_{n\to\infty} \mathsf{E}\mathbb{1}_A g(B_{t_1} - B_{1/n}, \cdots, B_{t_m} - B_{1/n}) \\
&= \mathsf{P}(A) \cdot \mathsf{E} g(B_{t_1}, \cdots, B_{t_m}),
\end{aligned}
$$

where in the first and last equalities, we use the (right-)continuity of $t \mapsto B_t$ at $t = 0$ and the continuity of g, and the Bounded Convergence Theorem, and in the second equality, we use the independence of $B_{t_k} - B_{1/n}$ with $A \in \mathcal{F}_{1/n}$. Then, this implies that $\mathcal{F}_{0+}^B$ is independent of $\sigma(B_t, t > 0)$.

On the other hand, $\mathcal{F}_0^B = \{\varnothing, \Omega\}$, so $\sigma(B_t, t > 0) = \sigma(B_t, t \geq 0)$. Since $\mathcal{F}_{0+}^B \subset \sigma(B_t, t \geq 0)$, we see that $\mathcal{F}_{0+}^B$ is independent of itself. Any such σ-algebra has to be trivial, and this completes the proof. $\square$

Using the zero-one law we can get some surprising results about the sample path of the Brownian motion.

Proposition 3.6 *With probability one,*

$$
\forall \varepsilon > 0, \qquad \sup_{0\leq t\leq\varepsilon} B_t > 0 > \inf_{0\leq t\leq\varepsilon} B_t. \tag{3.1}
$$

Proof: Consider the event

$$
A = \bigcap_{n=1}^{\infty} \{ \sup_{0\leq t\leq 1/n} B_t > 0 \}.
$$

Then since A is the intersection of decreasing events, we have

$$
\mathsf{P}(A) = \lim_{n\to\infty} \mathsf{P}(\sup_{0\leq t\leq 1/n} B_t > 0) \geq \liminf_{n\to\infty} \mathsf{P}(B_{1/n} > 0) = 1/2.
$$

On the other hand, $A \in \mathcal{F}_{0+}^B$, so by Theorem 3.5, $\mathsf{P}(A) = 1$. Hence,

$$
\mathsf{P}\big(\sup_{0\leq t\leq 1/n} B_t > 0 \big) = 1, \quad \forall n \geq 1.
$$

This implies that with probability one, $\sup_{0\leq t\leq\varepsilon} B_t > 0$ for all $\varepsilon > 0$. The other statement for the infimum can be proven similarly. $\square$

We can say something about the zero set of Brownian motion.

Proposition 3.7 *With probability one, there exists a decreasing sequence $t_1(\omega) > t_2(\omega) > \cdots > 0$ such that $B_{t_i} = 0$, i.e., 0 is the limit point of the zero set of B_t.*

Proof: We will construct the sequence (t_i) inductively. By Theorem 3.5, assume (3.1) holds with probability one.

Take $\varepsilon = 1$ in (3.1). Then there exists $s_1, s_1' \in (0, 1]$ such that $B_{s_1} > 0 > B_{s_1'}$. Since $t \mapsto B_t$ is continuous, there exists t_1 between s_1 and s_1' such that $B_{t_1} = 0$.

Now suppose that $t_1, t_2, \ldots, t_n$ have been constructed. Then in (3.1) taking $\varepsilon = t_n$, there exist $s_{n+1}, s_{n+1}' \in (0, t_n]$ such that $B_{s_{n+1}} > 0 > B_{s_{n+1}'}$. Hence there exists t_{n+1} between these two numbers such that $B_{t_{n+1}} = 0$. Clearly $t_{n+1} < t_n$ by this construction. $\square$

Remark 3.4 Suppose that our Brownian motion is constructed on $(\mathcal{C}[0,1], \mathcal{B}(\mathcal{C}[0,1]), \mathsf{P})$. Then clearly, the continuous function f defined by $f(t) = 0$ is not in the set A, so $A \neq \Omega = \mathcal{C}[0,1]$. This means that $\mathcal{F}_0^B \subsetneq \mathcal{F}_{0+}^B$.

Homework For $M > 0$, define $A_M = \bigcap_{n\geq 1} \{ \sup_{0<t\leq 1/n} \frac{B_t}{\sqrt{t}} > M \}$.

1. Show that $\mathsf{P}(A_M) \geq \mathsf{P}(\mathcal{N}(0,1) \geq M)$.
2. Use the zero-one law to deduce that $\mathsf{P}(A_M) = 1$.
3. For every $M > 0$, show that with probability one,
$$\sup_{0<t\leq \frac{1}{n}} \frac{B_t}{\sqrt{t}} > M, \quad \forall n \geq 1.$$
4. Show that with probability one,
$$\sup_{0<t\leq \frac{1}{n}} \frac{B_t}{\sqrt{t}} = +\infty, \quad \forall n \geq 1.$$

3.2 Markov property

We begin with the definition of a Markov process. If the range of t below is restricted to $t = n \in \mathbb{N}$, then one obtains a discrete-time Markov process.

Definition 3.4 *A stochastic process $X = (X_t)_{t\geq 0}$ is Markov if $\forall t, s > 0$,*

$$\mathsf{P}(X_{t+s} \in A \mid \mathcal{F}_t^X) = \mathsf{P}(X_{t+s} \in A \mid X_t), \quad \forall A \in \mathcal{B}(\mathbb{R}), \tag{3.2}$$

or equivalent,

$$\mathsf{E}\Big[F(X_{t+s}) \mid \mathcal{F}_t^X\Big] = \mathsf{E}\Big[F(X_{t+s}) \mid X_t\Big], \quad \forall F \textit{ bounded and measurable.} \tag{3.3}$$

The intuitive meaning of Markov properties is that, conditioned on the past $(\mathcal{F}_t^X)$ is the same as conditioned at the present (X_t), or in other words, knowing the present state X_t, the future $X_{t+s}, s > 0$ is independent of the past $\mathcal{F}_t^X$.

Remark 3.5 With some more efforts, (3.2) or (3.3) are equivalent to their multidimensional versions: for any $t, s_1, \ldots, s_m > 0$,

$$\mathsf{P}((X_{t+s_1}, \cdots, X_{t+s_m}) \in A \mid \mathcal{F}_t^X) = \mathsf{P}((X_{t+s_1}, \cdots, X_{t+s_m}) \in A \mid X_t), \quad \forall A \in \mathcal{B}(\mathbb{R}^m) \tag{3.4}$$

and

$$\mathsf{E}\Big[F(X_{t+s_1}, \cdots, X_{t+s_m}) \mid \mathcal{F}_t^X\Big] = \mathsf{E}\Big[F(X_{t+s_1}, \cdots, X_{t+s_m}) \mid X_t\Big], \quad \forall F \text{ bounded and measurable.} \tag{3.5}$$

Since we will deal with conditional expectation very often, it is useful to collect some basic facts about conditional expection here.

Definition 3.5 *Let $X \in L^1(\Omega, \mathcal{F}, \mathsf{P})$ and $\mathcal{G} \subset \mathcal{F}$ be a sub–σ-field. Then $\mathsf{E}[X \mid \mathcal{G}]$ is the unique $\mathcal{G}$-measurable r.v. (up to modification on a zero-probability set) such that for all $A \in \mathcal{G}$,*

$$\mathsf{E}\Big(\mathsf{E}[X \mid \mathcal{G}]\mathbb{1}_A\Big) = \mathsf{E}X\mathbb{1}_A.$$

Conditional expectation has the following properties. Their proofs can be found in any standard graduate probability textbook, say [Dur07, Shi96], etc.

Proposition 3.8 *The following identities are valid as long as the (conditional) expectations involved make sense.*

1. *If $X \in \mathcal{G}$, then $\mathsf{E}[XY \mid \mathcal{G}] = X\mathsf{E}[Y \mid \mathcal{G}]$.*
2. *If X is independent of $\mathcal{G}$, then $\mathsf{E}[X \mid \mathcal{G}] = \mathsf{E}X$ (that is, an almost sure constant).*
3. *If $\mathcal{G}_1 \subset \mathcal{G}_2$, then $\mathsf{E}\big[\mathsf{E}[X \mid \mathcal{G}_1] \mid \mathcal{G}_2\big] = \mathsf{E}\big[\mathsf{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1\big] = \mathsf{E}[X \mid \mathcal{G}_1]$.*
   In particular, if $\mathsf{E}[X \mid \mathcal{G}_2]$ is $\mathcal{G}_1$-measurable, then $\mathsf{E}[X \mid \mathcal{G}_1] = \mathsf{E}[X \mid \mathcal{G}_2]$.

Besides, all the well-known limit theorems (Fatou, Monotone/Dominated/Bounded Convergence Theorems, etc) and inequalities (Jensen's equality) also a version for conditional expectation.

A key lemma we will use a lot in the context of Markov processes is the following.

Lemma 3.9 *If $X \in \mathcal{G}$ and Y is independent of $\mathcal{G}$, then for any bounded measurable function $F : \mathbb{R}^2 \to \mathbb{R}$, we have*

$$\mathsf{E}[F(X,Y) \mid \mathcal{G}] = \varphi(X),$$

where φ is a deterministic (Borel measurable) function given by

$$\varphi(x) = \mathsf{E}F(x,Y).$$

The above can also be written in short as

$$\mathsf{E}[F(X,Y) \mid \mathcal{G}] = \Big(\mathsf{E}[F(x,Y) \mid \mathcal{G}]\Big)\Big|_{x=X}. \tag{3.6}$$

Remark 3.6 We stress that the substitution of $x = X$ into a deterministic function φ makes the right-hand side of (3.6) $\sigma(X)$-measurable and hence $\mathcal{G}$-measurable.

Proof: Consider the class of functions

$$\mathcal{S} = \{F \text{ bounded measurable } : \mathbb{R}^2 \to \mathbb{R} \text{ such that (3.6) holds}\}.$$

Then $\mathcal{S}$ forms a monotone class, that is, if $F_n \in \mathcal{S}$ and $F_n \wedge F$, then $F \in \mathcal{S}$ as well. Therefore, to show that $\mathcal{S}$ contains all the bounded measurable functions, by standard measure-theoretical argument, it suffices to show that $F(x,y) = \mathbb{1}_A(x)\mathbb{1}_B(y) \in \mathcal{S}$ for all $A, B \in \mathcal{B}(\mathbb{R})$.

Indeed, since $\mathbb{1}_A(X) \in \mathcal{G}$ and $\mathbb{1}_B(Y)$ is independent of $\mathcal{G}$, we have

$$\mathsf{E}\big[\mathbb{1}_A(X)\mathbb{1}_B(Y) \mid \mathcal{G}\big] = \mathbb{1}_A(X)\mathsf{E}\big[\mathbb{1}_B(Y) \mid \mathcal{G}\big] = \mathbb{1}_A(X)\mathsf{P}(Y \in B) = \varphi(X)$$

where

$$\varphi(x) = \mathsf{E}\mathbb{1}_A(x)\mathbb{1}_B(Y) = \mathbb{1}_A(x)\mathsf{P}(Y \in B).$$

This proves the proposition. □

Example 3.7 The Brownian motion is a Markov process.

In fact, $B_{t+s} - B_t$ is independent of $(B_{t_1}, \cdots, B_{t_m})$ for all $t_1, \cdots, t_m \in [0,t]$, so $B_{t+s} - B_t$ is independent of $\mathcal{F}_t^X$. Hence, for all F bounded measurable, applying Lemma 3.9 to $G(x,y) = F(x+y)$, we have

$$\mathsf{E}\Big[F(B_{t+s}) \mid \mathcal{F}_t^X\Big] = \mathsf{E}\Big[G(B_{t+s} - B_t, B_t) \mid \mathcal{F}_t^X\Big] = \Big[\mathsf{E}G(B_{t+s} - B_t, y)\Big]\Big|_{y=B_t},$$

which is a function of B_t and hence $\sigma(B_t)$-measurable. Then Markov property follows from Item 3 in Proposition 3.8.

Example 3.8 Let $f \in L^2_{\mathrm{loc}}[0,\infty) = \{g : g\mathbb{1}_{[0,t]} \in L^2[0,t],\ \forall t > 0\}$. Consider the stochastic integral define via the Gaussian white noise:

$$X_t = \int_0^t f(s)\, dB_s =: G\Big(f\mathbb{1}_{[0,t]}\Big).$$

Then $(X_t)_{t\ge 0}$ is a Markov process.

In fact, the previous analysis for Brownian motion only uses the fact "independent increment" property. To see that such property also holds for X_t, we have from the definition of Gaussian white noise isometry, if $[t_1,t_2]\cap[t_3,t_4] = \varnothing$, then

$$\mathsf{E}(X_{t_4} - X_{t_3})(X_{t_2} - X_{t_1}) = \mathsf{E}G(f\mathbb{1}_{[t_3,t_4]})G(f\mathbb{1}_{[t_1,t_2]}) = \int_0^\infty f^2(s)\mathbb{1}_{[t_1,t_2]}(s)\mathbb{1}_{[t_3,t_4]}(s)\, ds = 0.$$

Since the increments are centered Gaussian, if their covariance is zero, then they are independent.

Homework Let $(B_t)_{t\in[0,1]}$ be the Brownian motion and define $X_t = B_t - tB_1$, $t \in [0,1]$. The process $X = (X_t)_{t\in[0,1]}$ is called the "Brownian Bridge".

1. Show that $(X_t)_{t\ge 0}$ is a centered Gaussian process with covariance

$$\mathsf{E}X_tX_s = s(1-t), \quad \forall 0 \le s < t \le 1.$$

2. Let $t > s > s_1 > s_2 > \cdots > s_n \ge 0$. Show that

$$\mathsf{E}\Big(X_t - \frac{1-t}{1-s}X_s\Big)X_{s_i} = 0, \quad 1 \le i \le n.$$

   Deduce that $X_t - \dfrac{1-t}{1-s}X_s$ is independent of $(X_{s_1},\ldots,X_{s_n})$.

3. Let $t > s$. Show that $X_t - \dfrac{1-t}{1-s}X_s$ is independent of $\mathcal{F}^X_s$.
4. Show that $(X_t)_{t\in[0,1]}$ is Markov.

Next we will introduce the strong Markov property. While the usual Markov property states that future and past are conditionally independent if knowing the present, the strong Markov property allows the "present" to occur at a random stopping time. But first we need to understand how to condition on the information before a stopping time. Recall that a stopping time is a r.v. $T \in [0,\infty]$ such that $\{T \le t\} \in \mathcal{F}^X_t$, $\forall t \ge 0$. In what follows, unless otherwise stated, $\mathcal{F}_t = \mathcal{F}^X_t$ and $\mathcal{F}_\infty = \sigma(\mathcal{F}_t, t \ge 0)$.

Definition 3.6 *The stopping σ-algebra is*

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : \forall t \ge 0,\ A \cap \{T \le t\} \in \mathcal{F}_t\}.$$

Intuitively, $\mathcal{F}_T$ contains the information before a stopping time T.

Example 3.9 Let $a \ge 0$ and consider $T = a$ (a constant r.v.). Then T is a stopping time since

$$\{T \le t\} = \begin{cases} \Omega, & a \le t, \\ \varnothing, & a > t \end{cases} \in \mathcal{F}_t, \quad \forall t \ge 0.$$

Moreover, $\mathcal{F}_T = \mathcal{F}_a$.

We can compare the stopping σ-algebras for different stopping time, or extract information from the stopping σ-algebra.

Proposition 3.10 *If $S \le T$ are two stopping times, then $\mathcal{F}_S \subset \mathcal{F}_T$.*

Remark 3.10 Since $S \le T$, "information before S" is less than "information before T".

Proof: If $A \subset \mathcal{F}_S$, then for every $t \ge 0$,

$$A \cap \{T \le t\} = \Big(A \cap \{S \le t\}\Big) \cap \{T \le t\} \in \mathcal{F}_t.$$

So $A \subset \mathcal{F}_T$. This completes the proof. □

Proposition 3.11 *If T is a stopping time and $S \ge T$ is random time such that S is $\mathcal{F}_T$-measurable, then S is also a stopping time.*

Proof: For each $t \ge 0$, since $\{S \le t\} \in \mathcal{F}_T$,

$$\{S \le t\} = \{S \le t\} \cap \{T \le t\} \in \mathcal{F}_t.$$

This completes the proof. □

Remark 3.11 The stopping time S will take the form $S = f(T)$ for some measurable function f with $f(x) \ge x$.

We also need to impose more measurability constraint on our process $X = (X_t)_{t\ge 0}$.

Definition 3.7 *Let $X = (X_t)_{t\ge0}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathsf{P})$. We say that X is* measurable *if the map*

$$(t,\omega) \mapsto X_t(\omega) : \Big([0,\infty) \times \Omega, \mathcal{B}\big([0,\infty)\big) \otimes \mathcal{F}\Big) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

is measurable.

Proposition 3.12 *Let $X = (X_t)_{t\ge0}$ be measurable and T be a (finite) r.v., then $X_T(\omega) := X_{T(\omega)}(\omega)$ is a r.v.*

Proof: The map $\omega \mapsto X_{T(\omega)}(\omega)$ is the composition of the following two measurable maps:

$$\omega \mapsto (t', \omega') = (T(\omega), \omega'), \quad (t', \omega') \mapsto X_{t'}(\omega').$$

The first map is measurable since T is a r.v., and the second map is measurable since X is measurable. This proves the proposition. □

For adapted process, we introduce the notion of progressive measurability.

Definition 3.8 *Let $X = (X_t)_{t\ge0}$ be an adapted process on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge0}, \mathsf{P})$. We say that X is* progressively measurable *if for every fixed $t \ge 0$, the map*

$$(t,\omega) \mapsto X_t(\omega) : \Big([0,t] \times \Omega, \mathcal{B}\big([0,t]\big) \otimes \mathcal{F}_t\Big) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

is measurable.

Proposition 3.13 *Let $X = (X_t)_{t\ge0}$ be an adapted process on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge0}, \mathsf{P})$ which is progressively measurable and let T be a (finite) stopping time. Then $X_T := X_{T(\omega)}(\omega)$ is a $\mathcal{F}_T$-measurable r.v.*

Proof: Let $A \in \mathcal{B}(\mathbb{R})$. We have

$$\{X_T \in A\} \cap \{T \le t\} = \{X_{T\wedge t} \in A\} \cap \{T \le t\}.$$

It suffices to check that $\{X_{T\wedge t} \in A\} \in \mathcal{F}_t$.

In fact, the map $\omega \mapsto X_{T(\omega)\wedge t}(\omega)$ can be written as the composition of the two maps:

$$\omega \mapsto (t', \omega') = (T(\omega) \wedge t, \omega), \quad (t', \omega') \mapsto X_{t'}(\omega').$$

The first map is measurable from $(\Omega, \mathcal{F}_t)$ to $([0,t] \times \Omega, \mathcal{B}([0,t]) \times \mathcal{F}_t)$ by the definition of stopping times, while the second is measurable since X is progressively measurable. Hence, their composition is also measurable. This proves the proposition. □

Proposition 3.14 *If X is $(\mathcal{F}_t)$-adapted and has right-continuous path, then X is also progressively measurable w.r.t. $(\mathcal{F}_t)$.*

Proof: Fix $t > 0$. For $n \geq 1$ and $0 \leq k \leq 2^n - 1$, define

$$X_s^{(n)}(\omega) = X_{(k+1)/2^n}(\omega), \quad s \in \Big(\frac{kt}{2^n}, \frac{(k+1)t}{2^n}\Big].$$

and $X_0^{(n)}(\omega) = X_0(\omega)$. Then for each n, since X is $(\mathcal{F}_t)$-adapted, it is easy to check that $(s,\omega) \mapsto X_s^{(n)}(\omega)$ is $\mathcal{B}([0,t]) \times \mathcal{F}_t$-measurable. Since for every ω, the sample path $s \mapsto X_s(\omega)$ is right-continuous, we have $\lim_{n\to\infty} X_s^{(n)}(\omega) = X_s(\omega)$ for any $(s,\omega) \in [0,t] \times \Omega$. Therefore, the limit map $(s,\omega) \mapsto X_s(\omega)$ is also $\mathcal{B}([0,t]) \times \mathcal{F}_t$-measurable. This proves the proposition. $\square$

We are ready to state the strong Markov property.

Definition 3.9 *A progressively measurable Markov process $X = (X_t)_{t\geq 0}$ has the strong Markov property if for each a.s. finite stopping time S,*

$$\mathsf{P}(X_{T+t} \in A \mid \mathcal{F}_T) = \mathsf{P}(X_{T+t} \mid X_T). \tag{3.7}$$

Remark 3.12 The strong Markov property can be stated including stopping time T with $\mathsf{P}(T = \infty) > 0$. In that case X_{T+t} makes no sense when $\{T = \infty\}$, so (3.7) only needs to hold on the set $\{T < \infty\}$. For simplicity, we always assume $T < \infty$ a.s. in the sequel.

The Brownian motion has the strong Markov property. We know more about the conditioned process after the any stopping time.

Theorem 3.15 *Let T be a stopping time and define $B_t^{(T)} = B_{T+t} - B_T$. Then $(B_t^{(T)})_{t\geq 0}$ is a standard Brownian motion independent of $\mathcal{F}_T$.*

In particular, Brownian motion has the strong Markov property.

We now use the theorem to check that $(B_t)_{t\geq 0}$ is strongly Markov. The proof of Theorem 3.15 will be postpone to the end of this section.

Derivation of the strong Markov property for $(B_t)_{t\geq 0}$ from Theorem 3.15: Since B is progressively measurable, B_T is $\mathcal{F}_T$-measurable. By Lemma 3.9 and the assumption that $(B_t^{(T)})_{t\geq 0}$ is independent of $\mathcal{F}_T$, for any bounded measurable function F,

$$\mathsf{E}\Big(F(B_{T+t}) \mid \mathcal{F}_T\Big) = \mathsf{E}\Big(F(B_T + B_t^{(T)}) \mid \mathcal{F}_T\Big) = \mathsf{E}\Big(F(B_t^{(T)} + x)\Big)\big|_{x=B_T} \in \sigma(B_T).$$

So by Item 3 of Proposition 3.8, the strong Markov property holds. $\square$

An important consequence of the strong Markov property is the reflection principle. Consider the maximal process $B_t^* = \sup_{0\leq s\leq t} B_t$ and the hitting time $T_a = \inf\{t \geq 0 : B_t = a\}$ for $a > 0$.

Theorem 3.16 (Reflection Principle) *For $a \geq b$,*

$$\mathsf{P}(B_t^* \geq a,\ B_t < b) = \mathsf{P}(B_t > 2a - b).$$

Proof: Clearly, $\{B_t^* \geq a\} = \{T_a \leq t\} \in \mathcal{F}_{T_a}$ and we have

$$\{B_t^* \geq a,\ B_t < b\} = \{T_a \leq t,\ B_{t-T_a}^{(T_a)} < b - a\}.$$

By Theorem Theorem 3.15, $(B_s^{(T_a)})_{s\geq 0}$ is independent of $\mathcal{F}_T$. Since $T_a \in \mathcal{F}_T$ and Brownian motion is symmetric, we see that in distribution,

$$(T_a, (B_s^{(T_a)})_{s\geq 0}) \stackrel{d}{=} (T_a, (-B_s^{(T_a)})_{s\geq 0}).$$

Therefore,

$$\mathsf{P}(T_a \leq t,\ B_{t-T_a}^{(T_a)} < b-a) = \mathsf{P}(T_a \leq t,\ -B_{t-T_a}^{(T_a)} < b-a) = \mathsf{P}(T_a \leq t, B_{t-T_a}^{(T_a)} > a-b).$$

But on the event on the right-hand side, $B_t = B_{T_a} + B_{t-T_a}^{(T_a)} > 2a-b \geq a$, and by continuity, $B_t \geq a$ implies that $T_a \leq t$. So we have

$$\mathsf{P}(T_a \leq t, B_{t-T_a}^{(T_a)} > a-b) = \mathsf{P}(B_{t-T_a}^{(T_a)} > a-b) = \mathsf{P}(X_t > 2a-b),$$

where we use strong Markov property in the last equality. This proves the theorem. □

As a corollary, we have the distribution of the hitting time.

Proposition 3.17 *For $a > 0$,*

$$\mathsf{P}(T_a \leq t) = \mathsf{P}(B_t^* \geq a) = 2\mathsf{P}(B_t \geq a).$$

Proof: Using Theorem 3.16 for $b = a$, we have

$$\mathsf{P}(B_t^* \geq a) = \mathsf{P}(B_t^* \geq a,\ B_t < a) + \mathsf{P}(B_t^* \geq a, B_t \geq a) = \mathsf{P}(B_t > 2a-a) + \mathsf{P}(B_t \geq a) = 2\mathsf{P}(B_t \geq a).$$

□

Proof of Theorem 3.15: Denote by $W = (W_t)_{t\geq 0}$ be a Brownian motion independent of $B = (B_t)_{t\geq 0}$. By the definition of conditional probability, it suffices to show that for all $0 \leq t_1 < t_2 < \cdots < t_m$, all $A \in \mathcal{F}_T$ and all bounded continuous function F on $\mathbb{R}^m$, we have

$$\mathsf{E}F\big(B_{t_1}^{(T)}, B_{t_2}^{(T)}, \cdots, B_{t_m}^{(T)}\big)\mathbb{1}_A = \Big[\mathsf{E}F(W_{t_1}, W_{t_2}, \cdots, W_{t_m})\Big]\mathsf{P}(A). \tag{3.8}$$

Suppose T takes countably many values. Let $T \in \{s_1, s_2, \cdots\}$. Then the LHS of (3.8) is equal to

$$\begin{aligned}
&\sum_{n=1}^{\infty} \mathsf{E}F\big(B_{t_1}^{(T)}, B_{t_2}^{(T)}, \cdots, B_{t_m}^{(T)}\big)\mathbb{1}_A\mathbb{1}_{\{T=s_n\}} \\
&= \sum_{n=1}^{\infty} \mathsf{E}F\big(B_{s_n+t_1} - B_{s_n}, \cdots, B_{s_n+t_m} - B_{s_n}\big)\mathbb{1}_{A\cap\{T=s_n\}} \\
&= \sum_{n=1}^{\infty} \mathsf{E}\Big[\mathsf{E}\Big[F\big(B_{s_n+t_1} - B_{s_n}, \cdots, B_{s_n+t_m} - B_{s_n}\big)\mathbb{1}_{A\cap\{T=s_n\}} \mid \mathcal{F}_{s_n}\Big]\Big] \\
&= \sum_{n=1}^{\infty} \mathsf{E}\Big[\mathbb{1}_{A\cap\{T=s_n\}}\mathsf{E}\Big[F\big(B_{s_n+t_1} - B_{s_n}, \cdots, B_{s_n+t_m} - B_{s_n}\big) \mid \mathcal{F}_{s_n}\Big]\Big] \\
&= \sum_{n=1}^{\infty} \Big(\mathsf{E}\mathbb{1}_{A\cap\{T=s_n\}}\Big)\mathsf{E}F(W_{t_1}, \cdots, W_{t_m}) \\
&= \mathsf{P}(A) \cdot \mathsf{E}F(W_{t_1}, \cdots, W_{t_m}).
\end{aligned}$$

There are two crucial steps: in the third equality we use that $A \cap \{T = s_n\} \in \mathcal{F}_{s_n}$, which holds since T is a stopping time; in the fourth equality we use the simple Markov property for B.

General case. We approximate T be a sequence discrete stopping times:

$$T_k(\omega) = \frac{[2^k T] + 1}{2^k} = \sum_{n=0}^{\infty} \frac{n+1}{2^k} \mathbb{1}_{[\frac{n}{2^k}, \frac{n+1}{2^k})}(T(\omega)). \tag{3.9}$$

Indeed, T_k is stopping since for $t \in [n_0 2^{-k}, (n_0+1)2^{-k})$,

$$\{T_k(\omega) \le t\} = \{T \le \frac{n_0}{2^k}\} \in \mathcal{F}_{\frac{n_0}{2^k}} \subset \mathcal{F}_t,$$

or by Proposition 3.11. Also, since $|T_k - T| \le 2^{-k}$ and $T_k \ge T$, we have $T_k(\omega) \downarrow T(\omega)$ for every ω. Then by the right continuity of $t \mapsto B_t$, $B_t^{(T_k)} \to B_t^{(T)}$ as $k \to \infty$.

Now the left-hand side of (3.8) is equal to

$$\begin{aligned}
&\lim_{k\to\infty} \mathsf{E}\mathbb{1}_A F\big(B_{t_1+T_k} - B_{T_k}, \cdots, B_{t_m+T_k} - B_{T_k}\big) \\
&= \lim_{k\to\infty} \sum_{n=0}^{\infty} \mathsf{E}\mathbb{1}_{A\cap\{T\in[n2^{-k},(n+1)2^{-k})\}} F\big(B_{t_1}^{(T_k)}, \cdots, B_{t_m}^{(T_k)}\big) \\
&= \lim_{k\to\infty} \sum_{n=0}^{\infty} \mathsf{E}\Big[\mathsf{E}\Big[\mathbb{1}_{A\cap\{T_k=(n+1)2^k\}} F\big(B_{t_1}^{(T_k)}, \cdots, B_{t_m}^{(T_k)}\big) \mid \mathcal{F}_{T_k}\Big]\Big] \\
&= \lim_{k\to\infty} \sum_{n=0}^{\infty} \mathsf{E}\Big[\mathbb{1}_{A\cap\{T_k=(n+1)2^k\}} \mathsf{E}\Big[F\big(B_{t_1}^{(T_k)}, \cdots, B_{t_m}^{(T_k)}\big) \mid \mathcal{F}_{T_k}\Big]\Big] \\
&= \lim_{k\to\infty} \sum_{n=0}^{\infty} \Big(\mathsf{E}\mathbb{1}_{A\cap\{T_k=(n+1)2^k\}}\Big) \cdot \mathsf{E}F(W_{t_1}, \cdots, W_{t_m}) \\
&= \mathsf{P}(A)\mathsf{E}F(W_{t_1}, \cdots, W_{t_m}).
\end{aligned}$$

In the third equality we use $A \in \mathcal{F}_T \subset \mathcal{F}_{T_k}$ (Proposition 3.10), and in the fourth equality we use the strong Markov property for T_k. □

Remark 3.13 The proof only relies on the simple Markov property (which guarantees strong Markov property for discrete stopping times) and the right-continuity of sample path (which is used for approximation argument).

Homework Let $B = (B_t)_{t\ge0}$ and $\tilde{B} = (\tilde{B}_t)_{t\ge0}$ be two independent $(\mathcal{F}_t)$-adapted Brownian motion on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge0}, \mathsf{P})$. Let T be an a.s. finite stopping time. Define

$$W_t(\omega) = \begin{cases} B_t(\omega), & t \le T(\omega), \\ B_{T(\omega)} + \Big(\tilde{B}_t(\omega) - \tilde{B}_{T(\omega)}(\omega)\Big), & t > T(\omega). \end{cases}$$

1. Show that $(W_t)_{t\ge0}$ is a continuous, $(\mathcal{F}_t)$-adapted stochastic process.
2. Prove that $W = (W_t)_{t\ge0}$ is a standard Brownian motion by showing that W and B have the same finite dimensional distribution, namely, for all $0 \le t_1 < t_2 < \cdots < t_m$ and all Borel sets $A_1, A_2, \dots, A_m$,

$$\mathsf{P}(W_{t_1} \in A_1, \cdots, W_{t_m} \in A_m) = \mathsf{P}(B_{t_1} \in A_1, \cdots, B_{t_m} \in A_m).$$

Homework Let $B = (B_t)_{t\ge0}$ be the standard Brownian motion. For $a > 0$, let $T_a = \inf\{t \ge 0 : B_t = a\}$ be the first hitting time of a. For $\lambda > 0$, define the Laplace transform of T_a: $e^{-\varphi(\lambda,a)} = \mathsf{E}e^{-\lambda T_a}$. It is not hard to show that φ is a continuous function and we will assume that.

1. Use the strong Markov property to show that $T_a, T_{2a} - T_a, T_{3a} - T_{2a}, \ldots$ are i.i.d. random variables.
2. Show that
$$\varphi(\lambda, na) = n\varphi(\lambda, a), \quad n \geq 1,$$
and use continuity to conclude that $\varphi(\lambda, a) = \varphi(\lambda, 1)a$, $a > 0$.
3. Use the fact that $(\lambda B_{\lambda^{-2}t})_{t\geq 0}$ is also a standard Brownian motion for every $\lambda > 0$ to show that $T_{a\lambda}$ and $\lambda^2 T_a$ have the same distribution.
4. Show that $\varphi(\lambda^2, a) = \varphi(1, \lambda a)$ and conclude that there is a constant $c > 0$ such that
$$\mathsf{E}e^{-\lambda T_a} = e^{-c\sqrt{\lambda}a}.$$

3.3 Augmentation and usual condition

On a probability space $(\Omega, \mathcal{F}, \mathsf{P})$, A is a P-null/negligible set if there exists $N \in \mathcal{F}$ such that $A \subset N$ and $\mathsf{P}(N) = 0$. We recall the definition of a complete σ-field.

Definition 3.10 *We say that $\mathcal{G}$ is complete under the probability measure P if $N_1 \subset N_2$ where $N_2 \in \mathcal{G}$ and $\mathsf{P}(N_2) = 0$, then $N_1 \in \mathcal{G}$.*

Definition 3.11 *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$. we say that a filtration $(\mathcal{F}_t)$ satisfies the "usual condition" if*

1. *$\mathcal{F}_t = \mathcal{F}_{t+}$, i.e., it is right-continuous,*
2. *$\mathcal{F}_t$ contains all the P-null sets, i.e., if $A \subset N \in \mathcal{F}$ and $\mathsf{P}(N) = 0$, then $A \in \mathcal{F}_t$.*

We have seen that if a filtration is right-continuous, then optional times and stopping times are the same. In general, it is just simpler to work with complete probability space. We can always complete a σ-field by adding all the subsets of null sets. The completion of $\mathcal{G}$ under the probability measure P is

$$\begin{aligned}\bar{\mathcal{G}} &= \{G : \exists F \subset \mathcal{G} \text{ and } \mathsf{P}\text{-null set } N \in \mathcal{G} \text{ s.t. } F\Delta G \subset N\} \\ &= \{G : \exists F_1, F_2 \in \mathcal{G}, F_1 \subset F_2, \mathsf{P}(F_1) = \mathsf{P}(F_2) \text{ s.t. } F_1 \subset G \subset F_2\}.\end{aligned}$$

The completed measure on $\bar{\mathcal{G}}$ is defined by $\mathsf{P}(G) = \mathsf{P}(F)$.

With a $(\mathcal{F}_t)$-adapted process X, define the following collections of P-null sets

$$\begin{aligned}\mathcal{N}_t &= \{N : \exists F \subset \mathcal{F}_t^X : \mathsf{P}(F) = 0,\ N \subset F\} \\ \mathcal{N}_\infty &= \{N : \exists F \subset \mathcal{F}_\infty^X : \mathsf{P}(F) = 0,\ N \subset F\}.\end{aligned}$$

There are two ways to complete a filtration.

- **Completion**
$$\overline{\mathcal{F}}_t = \sigma(\mathcal{F}_t^X \cup \mathcal{N}_t) = \{G : \exists F \in \mathcal{F}_t^X \text{ s.t. } F\Delta G \in \mathcal{N}_t\}.$$
- **Augmentation**
$$\widetilde{\mathcal{F}}_t = \sigma(\mathcal{F}_t^X \cup \mathcal{N}_\infty) = \{G : \exists F \in \mathcal{F}_t^X \text{ s.t. } F\Delta G \in \mathcal{N}_\infty\}. \tag{3.10}$$

As we seen in Section 3.1, $\overline{\mathcal{F}}_t$ may not be right continuous: using the set A in the proof of Proposition 3.6, we see

$$\{\varnothing, \Omega\} = \mathcal{F}_0 = \overline{\mathcal{F}}_0 \subsetneq \mathcal{F}_{0+} \subset \overline{\mathcal{F}}_{0+}.$$

Indeed, from the zero-one law Theorem 3.5, even though $\mathcal{F}_{0+}$ is trivial, it still contains information strictly after time $t = 0$. This tells us just doing completion by adding null sets up to time t cannot lead to right-continuous filtration. However, if we do the augmentation, then the resulting filtration will be right-continuous, and thus satisfies the "usual condition".

Theorem 3.18 *Let X be the standard Brownian motion. Then the augmented filtration $(\tilde{\mathcal{F}}_t)_{t\geq 0}$ is right-continuous.*

Proof: The first step is to show that for every bounded $\mathcal{F}^B_\infty$-measurable r.v. Y and $t \geq 0$,

$$\mathsf{E}\Big[Y \mid \mathcal{F}^B_{t+}\Big] = \mathsf{E}\Big[Y \mid B_t\Big]. \tag{3.11}$$

To prove (3.11), it suffices to consider Y taking the form

$$Y = f(B_{t_1}, \cdots, B_{t_n}), \quad 0 \leq t_1 < \ldots < t_{m-1} < t \leq t_m < \cdots < t_n,$$

where f is a bounded continuous function. For $t = 0$, this is the main step in the proof of Theorem 3.5. For $t > 0$, the proof is similar; in order to get $\mathcal{F}^B_{t+}$ instead of $\mathcal{F}^B_t$, one needs to use the right-continuity of $t \mapsto B_t$: for $A \in \mathcal{F}^B_{t+}$,

$$\mathsf{E}\mathbb{1}_A f(B_{t_1}, \cdots, B_{t_n}) = \lim_{\varepsilon \downarrow 0} \mathsf{E}\mathbb{1}_A f(B_{t_1}, \cdots, B_{t_{m-1}}, B_{t_m+\varepsilon}, \cdots, B_{t_n+\varepsilon}).$$

Suppose that (3.11) is proved. Let $F \in \mathcal{F}^B_{t+} \subset \mathcal{F}^B_\infty$. Then by (3.11), $\mathsf{E}[\mathbb{1}_F \mid \mathcal{F}^B_{t+}]$ has a $\sigma(B_t)$-measurable version Z. On the other hand, $\mathsf{E}[\mathbb{1}_F \mid \mathcal{F}^B_{t+}] = \mathbb{1}_F$ a.s. Hence, for $A = \{Z = 1\} \in \mathcal{F}^B_t$, we have $F \Delta A \in \mathcal{N}_\infty$. This implies $F \in \widetilde{\mathcal{F}}_t$. Since F is arbitrary, $\mathcal{F}^B_{t+} \subset \widetilde{\mathcal{F}}_t$.

Next, let $F \subset \widetilde{\mathcal{F}}_{t+} = \bigcap_{n\geq 1} \widetilde{\mathcal{F}}_{t+\frac{1}{n}}$. Then by definition, there exist $G_n \in \mathcal{F}^B_{t+\frac{1}{n}}$ such that $F \Delta G_n \in \mathcal{N}_\infty$. We have

$$\begin{aligned}
& F \Delta G_n \in \mathcal{N}_\infty \\
\Leftrightarrow\quad & \mathbb{1}_F + \mathbb{1}_{G_n} = 0 \mod 2 \text{ a.s.}, \quad \forall n \geq 1, \\
\Leftrightarrow\quad & \mathbb{1}_F + \limsup_{n\to\infty} \mathbb{1}_{G_n} = 0 \mod 2 \text{ a.s.} \\
\Leftrightarrow\quad & F \Delta (\limsup_{n\to\infty} G_n) \in \mathcal{N}_\infty,
\end{aligned}$$

where

$$\limsup_{n\to\infty} G_n = \bigcap_{k=1}^\infty \bigcup_{m=k}^\infty G_m \in \bigcap_{k=1}^\infty \mathcal{F}^B_{t+\frac{1}{k}} \subset \mathcal{F}^B_{t+}.$$

Since $\mathcal{F}^B_{t+} \subset \widetilde{\mathcal{F}}_t$ and $\widetilde{\mathcal{F}}_t$ is complete, we have $F \subset \widetilde{\mathcal{F}}_t$. This shows $\widetilde{\mathcal{F}}_{t+} \subset \widetilde{\mathcal{F}}_t$ and hence the right-continuity of $(\widetilde{\mathcal{F}}_t)_{t\geq 0}$. □

Remark 3.14 We only use the simply Markov property and the right-continuity of the Brownian motion.

3.4 Sample path properties of Brownian motion

In this section we mention some interesting sample path properties of Brownian motion.

Proposition 3.19 (Nowhere monotone) *With probability one, there is no interval $[a, b]$ such that*

$$B_{t_1} \leq B_{t_2} \leq B_{t_3}, \quad \forall a \leq t_1 < t_2 < t_3 \leq b,$$

or

$$B_{t_1} \geq B_{t_2} \geq B_{t_3}, \quad \forall a \leq t_1 < t_2 < t_3 \leq b,$$

Proof: For any $q_1 < q_2$, by Proposition 3.6, with probability one,

$$\sup_{q_1 \leq s \leq q_2} B_s > B_{q_1} > \inf_{q_1 \leq s \leq q_2} B_s. \tag{3.12}$$

Hence, with probability one, Brownian motion is non-monotone in any given interval. By a union bound, Brownian motion is non-monotone simultaneously in all intervals $[q_1, q_2]$, $q_1, q_2 \in \mathbb{Q}$. Since any monotone interval $[a, b]$, if existing, will contain a monotone sub-interval with rational endpoints, the desired conclusion follows. $\square$

Proposition 3.20 (Nowhere differentiable) *With probability one, for every $t \geq 0$, either*

$$D^+ B_t = \limsup_{h \to 0+} \frac{B_{t+h} - B_t}{h} = \infty,$$

or

$$D_+ B_t = \liminf_{h \to 0+} \frac{B_{t+h} - B_t}{h} = -\infty,$$

Proof: See [KS98, pp. 110, Chap. 2, Theorem 9.18]. $\square$

Proposition 3.21 *With probability one, all local maxima of $t \mapsto B_t$ is strict.*

Proof: For $t_1 < t_2 < t_3 < t_4$, let

$$A_{t_1,t_2,t_3,t_4} = \{\omega : \sup_{s \in [t_3,t_4]} B_s - \sup_{s \in [t_1,t_2]} B_s \neq 0\}.$$

Then on $\bigcap_{t_i \in \mathbb{Q}} A_{t_1,\ldots,t_4}$, all local maxima are strict. It suffices to show that $\mathsf{P}(A_{t_1,\ldots,t_4}) = 1$ for all t_i. Indeed, we have

$$\sup_{s \in [t_3,t_4]} B_s - \sup_{s \in [t_1,t_2]} B_s = (B_{t_3} - B_{t_2}) + \inf_{s \in [t_1,t_2]} (B_{t_2} - B_s) + \sup_{s \in [t_3,t_4]} (B_s - B_{t_3})$$

which are sum of three independent, continuous random variables. Hence $\mathsf{P}(A_{t_1,\ldots,t_4}) = 1$. $\square$

Proposition 3.22 *With probability one, the zero set*

$$N(\omega) = \{t \geq 0 : B_t = 0\}$$

is a perfect set (a closed, measure-zero set with no isolated point, like the Cantor set).

Proof: We have

$$\{\omega : N(\omega) \text{ has an isolated point}\} = \bigcup_{a,b\in\mathbb{Q}} \{\omega : \text{ there is exactly one } s \in (a,b) \text{ such that } B_s(\omega) = 0\}.$$

For $t \geq 0$, let $\beta_t = \inf\{s > t : B_s = 0\}$. Then β_t are stopping times. By Proposition 3.7, $\beta_0 = 0$. By the strong Markov properties, $B_{\beta(t)+h} - B_{\beta(t)}$ is a standard Brownian motion, so $\beta_{\beta(t)} = \beta_t$ almost surely. Hence,

$$\{\omega : \text{ there is exactly one } s \in (a,b) \text{ such that } B_s(\omega) = 0\} \subset \{\omega : \beta_a(\omega) < b \text{ and } \beta_{\beta_a(\omega)}(\omega) > b\}$$

has zero probability. This completes the proof. □

4 Martingales

4.1 Basic martingale theory

Definition 4.1 *An adapted stochastic process $(M_t)_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathsf{P})$ is called a martingale if $M_t \in L^1(\Omega, \mathcal{F}, \mathsf{P})$ for all $t \geq 0$, and for all $s, t \geq 0$,*

$$\mathsf{E}\Big[M_{t+s} \mid \mathcal{F}_t\Big] = M_t.$$

If t only takes discrete values (like $\mathbb{Z}$), then we call (M_t) a discrete martingale.

Remark 4.1 If the filtration is not specified, we take the natural filtration $\mathcal{F}_t = \mathcal{F}_t^X$.

Example 4.2 Let X_i be independent random variables with $\mathsf{E}X_i = 0$. Then the partial sum $S_n = X_1 + \cdots + X_n$ forms a martingale, since by independence,

$$\mathsf{E}\Big[S_{n+m} \mid X_1, \cdots, X_n\Big] = X_1 + \cdots + X_n + \mathsf{E}[X_{n+1} + \cdots + X_m] = S_n.$$

Proposition 4.1 *Let $(X_t)_{t\geq 0}$ be a stochastic process with mean zero independent increments. Then*

1. *$(X_t)_{t\geq 0}$ is a martingale.*
2. *If $X_t \in L^2$ for all $t \geq 0$, then $(X_t^2 - \mathsf{E}X_t^2)_{t\geq 0}$ is a martingale.*
3. *If for some $\lambda \in \mathbb{R}$, $\mathsf{E}e^{\lambda X_t} < \infty$ for all $t \geq 0$, then $\left(\dfrac{e^{\lambda X_t}}{\mathsf{E}e^{\lambda X - t}}\right)_{t\geq 0}$ is a martingale.*

Proof:

1. This is obvious.
2. We have for $t > s$,

$$\begin{aligned}
&\mathsf{E}\Big[X_t^2 - X_s^2 \mid \mathcal{F}_s\Big] \\
&= \mathsf{E}\Big[\big(X_t - X_s + X_s\big)^2 - X_s^2 \mid \mathcal{F}_s\Big] \\
&= \mathsf{E}\Big[(X_t - X_s)^2 \mid \mathcal{F}_s\Big] + 2X_s\mathsf{E}\big[X_t - X_s \mid \mathcal{F}_s\big] \\
&= \mathsf{E}(X_t - X_s)^2 = \mathsf{E}(X_t - X_s)(X_t + X_s) - 2\mathsf{E}X_s(X_t - X_s) \\
&= \mathsf{E}X_t^2 - \mathsf{E}X_s^2.
\end{aligned}$$

3. We have for $t > s$,

$$\begin{aligned}\mathsf{E}\Big[e^{\lambda X_t} \mid \mathcal{F}_s\Big] &= e^{\lambda X_s}\mathsf{E}\Big[e^{\lambda(X_t-X_s)} \mid \mathcal{F}_s\Big] \\ &= e^{\lambda X_s}\mathsf{E}e^{\lambda(X_t-X_s)} \\ &= e^{\lambda X_s}\frac{\mathsf{E}e^{\lambda X_t}}{\mathsf{E}e^{\lambda X_s}}.\end{aligned}$$

□

Example 4.3 Let $(B_t)_{t\geq 0}$ be Brownian motion. Then $(B_t)_{t\geq 0}$, $(B_t^2 - t)_{t\geq 0}$, $(e^{\lambda B_t - \frac{1}{2}\lambda^2 t^2})_{t\geq 0}$ are all martingales.

Example 4.4 Let $f \in L^2_{\text{loc}}[0,\infty)$ and consider the stochastic integral defined via Gaussian white noise

$$M_t = \int_0^\infty \mathbb{1}_{[0,t]}(s)f(s)\,dB_s = G\Big(\mathbb{1}_{[0,t]}f\Big).$$

Then (M_t) has mean zero independent increments, and the processes

$$(M_t)_{t\geq 0}, \quad \Big(M_t^2 - \int_0^t f^2(s)\,ds\Big)_{t\geq 0}, \quad \Big(e^{\lambda M_t - \frac{1}{2}\lambda^2\int_0^t f^2(s)\,ds}\Big)_{t\geq 0}$$

are all martingales.

Example 4.5 Let $(N_t)_{t\geq 0}$ be a Poisson process with intensity λ, i.e.,

$$N_t = \max\{n \geq 0 : \xi_1 + \cdots + \xi_n \leq t\}$$

where $(\xi_i)_{i\geq 1}$ are i.i.d. $\text{Exp}(\lambda)$ random variables. Then $(N_t - \lambda t)_{t\geq 0}$ has mean zero independent increments.

Definition 4.2 *Let $(M_t)_{t\geq 0}$ be an adapted process and assume that $M_t \in L^1$ for all $t \geq 0$. We say that $(M_t)_{t\geq 0}$ is a* super-martingale *if*

$$\mathsf{E}[X_t \mid \mathcal{F}_s] \leq X_s, \quad \forall 0 \leq s < t,$$

and say that $(M_t)_{t\geq 0}$ is a sub-martingale *if*

$$\mathsf{E}[X_t \mid \mathcal{F}_s] \geq X_s, \quad \forall 0 \leq s < t.$$

One can use convex/concave functions to generate super- or sub-martingale from martingales.

Proposition 4.2 *If $(M_t)_{t\geq 0}$ is a martingale, and $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function, then $\big(\varphi(M_t)\big)_{t\geq 0}$ is a sub-martingale.*

Proof: Using Jensen's inequality for conditional expectation, we have for all $s < t$,

$$\mathsf{E}\big[\varphi(M_t) \mid \mathcal{F}_s\big] \geq \varphi\Big(\mathsf{E}\big[X_t \mid \mathcal{F}_s\big]\Big) = \varphi(X_s). \tag{4.1}$$

□

Corollary 4.3 *If $(M_t)_{t\geq 0}$ is a sub-martingale and $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and increasing, then $\big(\varphi(M_t)\big)_{t\geq 0}$ is also a sub-martingale.*

Proof: Since φ is increasing and $(M_t)_{t\geq 0}$ is a sub-martingale, the last equality in (4.1) will become

$$\varphi\Big(\mathsf{E}\big[X_t \mid \mathcal{F}_s\big]\Big) \geq \varphi(X_s),$$

and this completes the proof. □

Example 4.6 The function $|x|^p$ $(p \geq 1)$ is convex. The functions $x \vee a$ $(a \in \mathbb{R})$, $x^+ = x \vee 0$ are convex and increasing.

4.2 Convergence of martingales

In this section we discuss the a.s.-limit and L^1-limit of martingales. The main tools are *Doob's Up-crossing Theorem* and uniform integrability.

Let (X_t) be an adapted process (continuous-time or discrete-time) and $a < b$. Consider the following stopping times: $T_b^{(0)} = -\infty$,

$$T_a^{(\ell)} = \inf\{t \geq T_b^{(\ell-1)} : X_t \leq a\}, \quad T_b^{(\ell)} = \inf\{t \geq T_a^{(\ell)} : X_t \geq b\}, \quad \ell \geq 1.$$

In every interval $[T_a^{(\ell)}, T_b^{(\ell)}]$, the process (X_t) completes an up-crossing. The total number of up-crossing in a given interval $[0, n]$ is defined by

$$U_{ab}^X[0, n] = \max\{k : T_b^{(k)} \leq n\}.$$

Theorem 4.4 *Let $(X_n)_{n\geq 1}$ be a sub-martingale, then*

$$\mathsf{E}U_{ab}^X[0, n] \leq \frac{1}{b-a}\mathsf{E}(X_n - a)^+.$$

We have the following corollary about a.s. convergence of martingales.

Proposition 4.5 *If $(X_n)_{n\geq 1}$ is a sub-martingale, and $\sup_n \mathsf{E}X_n^+ < \infty$. Then there exists X such that $X_n \to X$ a.s.*

Proof: The up-crossing number is increasing in n, and hence by assumption,

$$\mathsf{E}U_{ab}^X[0, \infty) = \lim_{n\to\infty} \mathsf{E}U_{ab}^X[0, n] \leq \frac{\sup_n \mathsf{E}X_n^+ + |a|}{b-a} < \infty.$$

This implies $U_{ab}^X[0, \infty) < \infty$ a.s., that is, with probability one, any interval $[a, b]$ is being up-crossed by at most finitely many times. As a consequence, for fixed

$$\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n$$

cannot happen. Taking a union bound over all $[a, b]$ with $a, b \in \mathbb{Q}$, we see that with probability one,

$$\limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n.$$

This proves the statement. □

Example 4.7 If a martingale $(X_n)_{n\geq 0}$ is non-negative, then $\mathsf{E}X_n^+ = \mathsf{E}X_n = \mathsf{E}X_0$, and hence $\lim_{n\to\infty} X_n$ exists.

Next we will discuss the L^1-convergence. Recall the definition of uniform integrability for a family of random variables $\{X_n\}$.

Definition 4.3 *A family of random variables (X_n) is* uniformly integrable, *if*

$$\lim_{M\to\infty} \sup_n \mathsf{E}\mathbb{1}_{\{|X_n|\geq M\}}|X_n| = 0.$$

Uniform integrability is the necessary and sufficient condition for L^1-convergence.

Theorem 4.6 *If $X_n \to X$ a.s., then $X_n \to X$ if and only if (X_n) is uniformly integrable.*

Example 4.8 The following conditions will imply uniform integrability.

1. If there exists $Z \in L^1$ such that $|X_n| \le Z$ for all n, then (X_n) is uniformly integrable. (This is Dominated Convergence Theorem.)
2. If $\sup_n \mathsf{E}|X_n|^p < \infty$ for some $p > 1$, then (X_n) is uniformly integrable.
3. Let $Z \in L^1$. Then the collection of r.v.'s $\{\mathsf{E}[Z \mid \mathcal{G}] : \mathcal{G} \subset \mathcal{F}\}$ is uniformly integrable.

We will prove the last point.

Proposition 4.7 *Let $Z \in L^1(\Omega, \mathcal{F}, \mathsf{P})$. Then the collection of r.v.'s*

$$\{\mathsf{E}[Z \mid \mathcal{G}] : \mathcal{G} \text{ is a sub–}\sigma\text{-field of } \mathcal{F}\}$$

is uniformly integrable.

Proof: Since $Z \in L^1(\Omega, \mathcal{F}, \mathsf{P})$, for every $\varepsilon > 0$, there is $\delta > 0$ such that whenever $\mathsf{P}(A) < \delta$, $\mathsf{E}|Z|\mathbb{1}_A < \varepsilon$.

By Jensen's inequality, for $A = \{|\mathsf{E}[Z \mid \mathcal{G}]| \ge M\} \in \mathcal{G}$, we have

$$\mathsf{E}\mathbb{1}_A|\mathsf{E}[Z \mid \mathcal{G}]| \le \mathsf{E}\mathbb{1}_A\mathsf{E}[|Z| \mid \mathcal{G}] = \mathsf{E}\mathsf{E}[|Z|\mathbb{1}_A \mid \mathcal{G}] = \mathsf{E}|Z|\mathbb{1}_A.$$

When $A = \Omega$, the above inequality gives $\mathsf{E}|\mathsf{E}[Z \mid \mathcal{G}]| \le \mathsf{E}|Z|$. Then by Markov inequality,

$$\mathsf{P}(A) \le \frac{\mathsf{E}|Z|}{M},$$

uniformly for all sub–σ-field $\mathcal{G}$. Combining all these together we prove the statement. □

Proposition 4.8 *A martingale (X_n) is uniformly integrable, if and only if there exists $X_\infty \in L^1$ such that $X_n = \mathsf{E}[X_\infty \mid \mathcal{F}_n]$.*

Proof: **The "$\Rightarrow$" direction.** Uniform integrability implies that $\sup_n \mathsf{E}|X_n| < \infty$, hence Proposition 4.5 implies that there exists X_∞ such that $X_n \to X_\infty$ a.s. But (X_n) is also uniformly integrable, so the limit is also in L^1. Then,

$$\mathsf{E}[X_\infty \mid \mathcal{F}_n] = \lim_{m\to\infty} \mathsf{E}[X_{n+m} \mid \mathcal{F}_n] = X_n$$

as desired.

The "$\Leftarrow$" direction. It follows from Proposition 4.7. □

Adaption to the continuous-time.

The L^1-convergence relies on the uniform integrability, which holds true for continuous-time. The a.s. convergence relies on the up-crossing inequality, and we need extra continuity assumption to take the limit of approximation.

Theorem 4.9 (Continuous-time Doob's Up-crossing Inequality) *Let $(X_t)_{t\ge0}$ be a right-continuous sub-martingale, then for all $T > 0$*

$$\mathsf{E}U_{ab}^X[0, T] \le \frac{1}{b-a}\mathsf{E}(X_T - a)^+.$$

Proof: We can restrict the definition of up-crossings to $D_n = \mathbb{Z}/2^n$. We denote the number of up-crossing observed on D_n by $U_{ab}^X[0,T] \cap D_n$. Since D_n has few points, the number of up-crossing is smaller (an up-crossing can occur on an interval $(k/2^n, (k+1)/2^n)$ and not "seen" by the set D_n). But since X has right-continuous path,

$$U_{ab}^X[0,T] \cap D_n \uparrow U_{ab}^X[0,T], n \to \infty,$$

almost surely. Now on D_n, $(X_t)_{t \in D_n}$ is just a discrete martingale, and we have

$$\mathsf{E} U_{ab}^X[0,T] \cap D_n \leq \frac{1}{b-a} \mathsf{E}(X_T - a)^+.$$

Taking the limit $n \to \infty$ and using the Monotone Convergence Theorem prove the statement. $\square$

4.3 Optional Sampling Theorem

Theorem 4.10 *Let $(X_t)_{t \geq 0}$ be a right-continuous martingale, and $S \leq T$ be two stopping times. Suppose that either*

1. *S, T are bounded, i.e., there is a constant $N > 0$ such that $S, T \leq N$, or*
2. *$(X_t)_{t \geq 0}$ is uniformly integrable.*

Then

$$X_S = \mathsf{E}[X_T \mid \mathcal{F}_T].$$

In particular, $\mathsf{E}X_S = \mathsf{E}X_T = \mathsf{E}X_0$.

Remark 4.9 The first condition implies that $X_t = \mathsf{E}[X_N \mid \mathcal{F}_t]$, and the second condition by Proposition 4.8 implies that $X_t = \mathsf{E}[X_\infty \mid \mathcal{F}_t]$. So both conditions implies that there is a r.v. $Z \in L^1$ such that $X_t = \mathsf{E}[Z \mid \mathcal{F}_t]$ for all t that we care about.

Proof: Let $Z = X_N$ if the first condition holds and $Z = X_\infty$ if the second condition holds. It suffices to show

$$X_T = \mathsf{E}[Z \mid \mathcal{F}_T]. \tag{4.2}$$

Indeed, if (4.2) holds, since $\mathcal{F}_S \leq \mathcal{F}_T$, we have

$$\mathsf{E}[X_T \mid \mathcal{F}_S] = \mathsf{E}\big[\mathsf{E}[Z \mid \mathcal{F}_T] \mid \mathcal{F}_S\big] = \mathsf{E}[Z \mid \mathcal{F}_S] = X_S.$$

The proof of (4.2) will be done in two steps. First we prove it for discrete stopping times, then we use approximation.

Suppose that the range of T is countable, i.e., $T \in \{t_1, t_2, \ldots\}$. Then for all $A \in \mathcal{F}_S$,

$$\begin{aligned}
\mathsf{E}\big(\mathsf{E}[Z \mid \mathcal{F}_S]\mathbb{1}_A\big) = \mathsf{E}Z\mathbb{1}_A &= \sum_{n=1}^\infty \mathsf{E}Z\mathbb{1}_{A \cap \{T = t_n\}} \\
&= \sum_{n=1}^\infty \mathsf{E}\big(\mathbb{1}_{A \cap \{T = t_n\}} \mathsf{E}[Z \mid \mathcal{F}_{t_n}]\big) \\
&= \sum_{n=1}^\infty \mathsf{E}\mathbb{1}_{A \cap \{T = t_n\}} X_{t_n} = \mathsf{E}X_T \mathbb{1}_A,
\end{aligned}$$

where in the second line we use that $A \cap \{T = t_n\} \in \mathcal{F}_{t_n}$ since T is a stopping time.

General case $T \geq 0$. As before, we can approximate T by discrete stopping times

$$T_k = \frac{[2^k T] + 1}{2^k} \downarrow T.$$

Let $A \in \mathcal{F}_T \subset \mathcal{F}_{T_k}$. Then by the first step,

$$\mathsf{E}\mathbb{1}_A X_{T_k} = \mathsf{E}\mathbb{1}_A Z$$

for all T_k. The right continuity of X and $T_k \downarrow T$ imply that $X_{T_k} \to X_T$ a.s., and $X_{T_k} = \mathsf{E}[Z \mid \mathcal{F}_{T_k}]$ and Proposition 4.7 imply that X_{T_k} are uniformly integrable. Therefore,

$$\mathsf{E}\mathbb{1}_A X_T = \lim_{k\to\infty} \mathsf{E}\mathbb{1}_A X_{T_k} = \mathsf{E}\mathbb{1}_A Z.$$

$\square$

Example 4.10 If T is a stopping time, $(M_t)_{t\geq 0}$ is a martingale, then $(M_{t\wedge T})_{t\geq 0}$ is also a martingale. We only need to verify for all $s < t$,

$$\mathsf{E}[M_{t\wedge T} \mid \mathcal{F}_{s\wedge T}] = M_{s\wedge T}.$$

This follows from Theorem 4.10 and the boundedness of the stopping time $s \wedge T$, $t \wedge T$.

Example 4.11 Let $(B_t)_{t\geq 0}$ be Brownian motion, and T_a, T_b be the first hitting time of $a > 0 > b$. Applying Theorem 4.10 to the bounded stopping time $T_a \wedge T_b \wedge n$ gives

$$\mathsf{E}B_{T_a\wedge T_b\wedge n} = \mathsf{E}B_0 = 0. \tag{4.3}$$

Since $|B_{T_a\wedge T_b\wedge n}| \leq |a| \vee |b|$ and $\mathsf{P}(T_a \wedge T_b < \infty) = 1$ (one can easily show for some $\rho < 1$, $\mathsf{P}(T_a \wedge T_b \geq k) \leq \rho^k$), we can take $n \to \infty$ in (4.3) and get

$$0 = \mathsf{E}B_{T_a\wedge T_b} = a\mathsf{P}(T_a < T_b) + b\mathsf{P}(T_a > T_b).$$

Also $\mathsf{P}(T_a < T_b) + \mathsf{P}(T_a > T_b) = 1$. Hence, we have

$$\mathsf{P}(T_a < T_b) = \frac{-b}{a-b}, \quad \mathsf{P}(T_a > T_b) = \frac{a}{a-b}. \tag{4.4}$$

In particular, letting $b \downarrow -\infty$ and $T_b \uparrow \infty$, we obtain $\mathsf{P}(T_a < \infty) = 1$.

Example 4.12 Apply 4.10 to the martingale $(B_t - t^2)_{t\geq 0}$ and the stopping time $T_a \wedge T_b \wedge n$, we have

$$\mathsf{E}B^2_{T_a\wedge T_b\wedge n} - (T_a \wedge T_b \wedge n) = 0.$$

In the limit $n \to \infty$, the first term is bounded by $|a|^2 \vee |b|^2$, the second term is increasing in n, so by Bounded Convergence Theorem and Monotone Convergence Theorem, we have

$$\mathsf{E}B_{T_a\wedge T_b} - (T_a \wedge T_b).$$

Combining with (4.4) we have $\mathsf{E}T_a \wedge T_b = |ab|$. Letting $b \downarrow -\infty$ and obtain $\mathsf{E}T_a = \infty$.

Homework Recall that $(Z_t = e^{\lambda B_t - \frac{\lambda^2}{2}t})_{t\geq 0}$ is a martingale for every $\lambda \in \mathbb{R}$. It is a non-negative super-martingale so it has an almost sure limit Z_∞.

1. Let T_a be the hitting time of $a > 0$. Use the Optional Sampling Theorem to show that

$$\mathsf{E}e^{-cT_a} = e^{-a\sqrt{2c}}, \quad c > 0.$$

2. Show that $Z_\infty = 0$.
Hint: One possible proof is to use Borel–Cantelli to show that for any $\varepsilon, M > 0$, $B_n \le \varepsilon n + M$ for large enough n.

3. Let $S_a = \inf\{t \ge 0 : B_t \ge at + 1\}$, $a > 0$. Use the Optional Sampling Theorem to show that

$$\mathsf{E} e^{2a} \mathbb{1}_{\{S_a \le t\}} + \mathsf{E} e^{2aB_t - 2a^2 t} \mathbb{1}_{\{S_a > t\}} = 1.$$

Take the limit $t \to \infty$ and show that $\mathsf{P}(S_a < \infty) = e^{-2a}$.

We will also mention the Optional Sampling Theorem for sub-/super-martingales.

Definition 4.4 *A (sub-/super-)martingale $(X_t)_{t\ge 0}$ has a last element/is closed by X_∞, if $\exists X_\infty \in L^1$ such that $(X_t)_{0\le t\le\infty}$ forms a (sub-/super-)martingale.*

Example 4.13 If $(M_t)_{t\ge 0}$ is a martingale, then by 4.8, it has a last element if and only if it is uniformly integrable. Moreover, M_∞ is the a.s. and L^1 limit of M_t.

Example 4.14 If $(X_t)_{t\ge 0}$ is a non-negative super-martingale, then it always has a last element $X_\infty = 0$, since it is trivially true that

$$X_t \ge 0 = \mathsf{E}[X_\infty \mid \mathcal{F}_t], \quad \forall t \ge 0.$$

But having a last element is weaker than uniform integrability. Consider $X_t = 1 + B_{t\wedge T_{-1}}$ which is a martingale and hence super-martingale. It is non-negative. It is easy to see that

$$X_\infty = \lim_{t\to\infty} X_t = 1 + B_{T_{-1}} = 0,$$

but $1 = \lim_{t\to\infty} \mathsf{E} X_t \ne \mathsf{E} X_\infty = 0$, so it cannot be uniformly integrable.

Theorem 4.11 *Let $(X_t)_{t\ge 0}$ is a right-continuous sub-martingale and $S \le T$ be two stopping times. If either*

1. *S, T are bounded, or*

2. *$(X_t)_{t\ge 0}$ has a last element $X_\infty \in L^1$,*

then

$$\mathsf{E}[X_T \mid \mathcal{F}_S] \ge X_S.$$

A similar statement also holds for super-martingale.

Sketch of proof: The first step is to prove the theorem for discrete sub-martingales. This is more delicate than the martingale case since it cannot be derived from $\mathsf{E}[X_\infty \mid \mathcal{F}_T] \ge X_T$. For a proof, see [Chu74, Chap. 9] (which is also a good read on discrete martingale theory). Here, discreteness is really essential, while previously we only use the ranges of stopping times are countable.

The second step is to approximate the stopping times S and T by discrete stopping times by above. From $\mathsf{E}\mathbb{1}_A X_{S_n} \le \mathsf{E}\mathbb{1}_A X_{T_n}$, $A \in \mathcal{F}_S$, pass the limit $n \to \infty$ by establishing the uniform continuity of $(X_{S_n})_{n\ge 1}$ and $(X_{T_n})_{n\ge 1}$. $\square$

4.4 Doob's Maximal inequality

We will state the maximal inequality for sub-martingales. Similar statements also hold for super-martingales.

Theorem 4.12 *Let $(X_t)_{t\geq 0}$ be a continuous sub-martingale and $\lambda > 0$. Then*

$$\lambda \mathsf{P}\Big(\sup_{0\leq s\leq t} X_s > \lambda\Big) \leq \mathsf{E}X_t^+, \tag{4.5}$$

$$\lambda \mathsf{P}\Big(\inf_{0\leq s\leq t} X_s < -\lambda\Big) \leq \mathsf{E}X_t^+ - \mathsf{E}X_0. \tag{4.6}$$

Proof: Denote the event in (4.5) as A. Note that A is indeed measurable since by continuity, the supremum over $[0,t]$ is the same as the supremum over $[0,t]\cap\mathbb{Q}$, and the later is measurable. Let $T = \inf\{t : X_t \geq \lambda\}$. Then $A = \{T \leq t\}$. Since X is a sub-martingale, X^+ is also a sub-martingale, hence Theorem 4.11 implies that

$$\mathsf{E}X_t^+ \geq \mathsf{E}X_{t\wedge T}^+ \geq \mathsf{E}X_{t\wedge T}^+ \mathbb{1}_{\{T\leq t\}} = \lambda \mathsf{P}(A).$$

This proves (4.5).

Denote the event in (4.6) by B and let $S = \inf\{t : X_t \leq \lambda\}$. Then $B = \{S \leq t\}$. Again by Theorem 4.11, we have

$$\begin{aligned}\mathsf{E}X_0 \leq \mathsf{E}X_{t\wedge S} &= \mathsf{E}X_t \mathbb{1}_{\{T>t\}} + \mathsf{E}X_T \mathbb{1}_{\{T\leq t\}} \\ &\leq \mathsf{E}X_t \mathbb{1}_{\{T>t\}} - \lambda \mathsf{P}(B) \leq \mathsf{E}X_t^+ - \lambda \mathsf{P}(B),\end{aligned}$$

and (4.6) follows. □

Corollary 4.13 *Let $(M_t)_{t\geq 0}$ be a continuous martingale. Then for every $\lambda > 0$,*

$$\lambda \mathsf{P}\Big(\sup_{0\leq s\leq t} |M_t| \geq \lambda\Big) \leq \mathsf{E}|X_t|.$$

Proof: We apply (4.5) in Theorem 4.12 to the sub-martingale $(|M_t|)_{t\geq 0}$. □

For martingales, we also have the control on the maximal of L^p norm.

Theorem 4.14 *Let $(M_t)_{t\geq 0}$ be a continuous martingale. Then for every $p > 1$,*

$$\mathsf{E} \sup_{0\leq s\leq t} |M_s|^p \leq \Big(\frac{p}{p-1}\Big)^p \mathsf{E}|X_t|^p.$$

Proof: Let $Y = \sup\limits_{0\leq s\leq t} |M_s|$. Since $(|M_t|)_{t\geq 0}$ is a continuous sub-martingale, by the proof of (4.5), we have

$$\lambda \mathsf{P}(Y \geq \lambda) + \mathsf{E}|M_t| \mathbb{1}_{\{Y<\lambda\}} \leq \mathsf{E}|M_t|,$$

and hence

$$\mathsf{P}(Y \geq \lambda) \leq \frac{1}{\lambda} \mathsf{E}|M_t| \mathbb{1}_{\{Y\geq\lambda\}}.$$

Now

$$
\begin{aligned}
\mathsf{E}Y^p &= p\int_0^\infty \lambda^{p-1}\mathsf{P}(Y\ge\lambda)\,d\lambda \\
&\le p\int_0^\infty \lambda^{p-2}\mathsf{E}\big(|M_t|\mathbb{1}_{\{Y\ge\lambda\}}\big) \\
&= \mathsf{E}\Big(|M_t|\int_0^Y p\lambda^{p-2}\,d\lambda\Big) \\
&= \frac{p}{p-1}\cdot\mathsf{E}\big(|M_t|\cdot Y^{p-1}\big) \\
&\le \frac{p}{p-1}\big(\mathsf{E}|M_t|^p\big)^{1/p}\big(\mathsf{E}Y^p\big)^{p/(p-1)}.
\end{aligned}
$$

The last inequality is just Hölder inequality. Hence, if $\mathsf{E}Y^p<\infty$, then we can divide both sides by $\big(\mathsf{E}Y^p\big)^{p/(p-1)}$ and then take the p-th power to get $\mathsf{E}Y^p\le\Big(\dfrac{p}{p-1}\Big)^pE|M_t|^p$. To treat the general case where $\mathsf{E}Y^p<\infty$ is not known, we use truncation, that is, we first get the estimate

$$
\mathsf{E}(Y\wedge m)^p\le\big(\frac{p}{p-1}\big)^p\mathsf{E}|M_t|^p
$$

for the bounded r.v $(Y\wedge m)$ with any $m>0$. Then we let $m\to\infty$ and get the desired conclusion. □

As an application of the Doob's L^p-maximal inequality, let us study the continuity of the stochastic integral $M_t=G(\mathbb{1}_{[0,t]}f)$ for $f\in L^2_{\text{loc}}[0,\infty)$. Recall that the Gaussian white noise construction in Theorem 2.5 only ensures that M_t has independent increments, and hence is both a Markov process and a martingale. We can use Theorem 2.7 to get continuity of M if $|f|$ is bounded, but that is still to restrictive. Using martingale argument, we can show that $(M_t)_{t\ge0}$ has a continuous modification as long as $f\in L^2_{\text{loc}}([0,\infty))$. This is essentially the argument that we will use for more general stochastic integral. See XXX.

Fix $T>0$. We just need to show that $(M_t)_{t\in[0,T]}$ has a continuous modification for every $T>0$ and $f\in L^2[0,T]$. By standard argument, there exist piecewise constant functions $f_n\in L^2[0,T]$ such that $\|f_n-f\|_{L^2[0,T]}\to0$. It is easy to check that

$$
M_t^{f_n}=G\Big(\mathbb{1}_{[0,t]}f_n\Big)
$$

is a continuous martingale. Without loss of generality we assume $\|f_n-f_{n+1}\|^2_{L^2}\le 8^{-n}$. For every n, applying Theorem 4.12 to the submartingale $X_t=|M_t^{f_n}-M_t^{f_{n+1}}|^2$, we have

$$
\mathsf{P}\Big(\sup_{0\le t\le T}|M_t^{f_n}-M_t^{f_{n+1}}|\ge\frac{1}{2^n}\Big)\le 4^n\mathsf{E}|M_T^{f_n}-M_T^{f_{n+1}}|^2\le 4^n\|f_n-f_{n+1}\|^2_{L^2[0,T]}\le 2^{-n}.
$$

Then, by Borel–Cantelli, there exists $n_0=n_0(\omega)$ such that for all $n\ge n_0(\omega)$,

$$
\sup_{0\le t\le T}|M_t^{f_n}-M_t^{f_{n+1}}|\le\frac{1}{2^n},
$$

and hence with probability, the infinite function series

$$
M_t^\infty=M_t^{f_0}+\sum_{n=0}^\infty(M_t^{f_{n+1}}-M_t^{f_n})
$$

converges absolutely, and the limiting function is continuous in t. It is easy to check that $(M_t^\infty)_{t\ge0}$ is a continuous modification of $G\big(\mathbb{1}_{[0,t]}f\big)$.

Homework Let $(X_t)_{t\geq 0}$ be a $(\mathcal{F}_t)$-adapted, bounded continuous process. For any partition $\Delta : 0 = t_0 < t_1 < \cdots < t_n = t$ we define $X_s^\Delta = \sum_{k=0}^{n-1} X_{t_k} \mathbb{1}_{[t_k, t_{k+1})}(s)$. By continuity of X, it is easy to see that the limit

$$\lim_{|\Delta|\to 0} \int_0^t X_s^\Delta \, ds = \lim_{|\Delta|\to 0} \sum_{k=0}^{n-1} X_{t_k}(t_{k+1} - t_k) = \int_0^t X_s \, ds$$

exists almost surely, so $\int_0^t X_s \, ds$ is a well-defined r.v.

1. Show that for any sub–σ-field $\mathcal{G} \subset \mathcal{F}_\infty$, there exists a bounded continuous process $(Y_t)_{t\geq 0}$ such that for every $t \geq 0$, $Y_t = \mathsf{E}[X_t \mid \mathcal{G}]$ a.s.

   Hint: define Y_t first for $t \in \mathbb{Q}$ and then consider the extension to $t \in \mathbb{R}$.

2. Show that

$$\mathsf{E}\Big[\int_0^t X_s \, ds \mid \mathcal{G}\Big] = \int_0^t \mathsf{E}[X_s \mid \mathcal{G}] \, ds.$$

   Hint: The identity is true for X_t^Δ; then justify the limit $|\Delta| \to 0$ carefully using boundedness and continuity.

3. Let $i = \sqrt{-1}$. For any $\lambda \in \mathbb{R}$, show that

$$e^{i\lambda B_t} + \int_0^t \frac{1}{2}\lambda^2 e^{i\lambda B_s} \, ds, \quad t \geq 0$$

   is a martingale.

   Note: this implies $f(B_t) - \int_0^t \frac{1}{2} f''(B_s) \, dB_s$ is a martingale if f has a sufficiently nice Fourier transform representation $f(x) = \int e^{i2\pi x\xi} \hat{f}(\xi) \, d\xi$, since $f''(x) = -\int 4\pi^2 \xi^2 e^{i\lambda x\xi} \hat{f}(\xi) \, d\xi$.

5 Local martingales and quadratic variation

Previously we have define the stochastic integral

$$\int_0^t f(s) \, dB_s =: G\big(\mathbb{1}_{[0,t]} f\big)$$

for $f \in L^2_{\text{loc}}[0, \infty)$. At the end of last section we have also seen that $\int_0^t f(s) \, dB_s$ is a continuous martingale, essentially because the prelimiting process

$$\sum_{n=0}^{\infty} f_{t_n} (B_{t\wedge t_{n+1}} - B_{t\wedge t_n}) \tag{5.1}$$

is a continuous martingale. We will consider the following generalizations.

1. First, we want to replace the deterministic function $f(t)$ by a random process. Consider

$$f(t) = \sum_{n=0}^{\infty} \xi_n(\omega) \mathbb{1}_{(t_n, t_{n+1}]}(t) \tag{5.2}$$

and

$$\int_0^t f(s)\,dB_s = \sum_{n=0}^{\infty} \xi_n(\omega)(B_{t\wedge t_{n+1}} - B_{t\wedge t_n}). \tag{5.3}$$

The equation (5.3) defines a martingale as long as $\xi_n \in \mathcal{F}_{t_n}$. For such f, we also have the Itô's isometry:

$$\mathsf{E}\Big(\int_0^t f(s)\,dB_s\Big)^2 = \mathsf{E}\int_0^t f^2(s)\,ds. \tag{5.4}$$

By approximation, we can define $\int_0^t f(s)\,dB_s$ for all processes that can be approximated by processes of the form (5.2), known as the progressively measurable process, such that the right-hand side of (5.4)is finite.

2. But even the Brownian motion B_t is not necessary in (5.3) to define a martingale. We can replace $(B_t)_{t\geq 0}$ by any continuous martingale $(M_t)_{t\geq 0}$. Then the term ds in (5.4) also needs to be adjusted, since $t = \mathsf{E}B_t^2$ is no longer true for other continuous martingales (in fact, it uniquely determines the Brownian motion, see XXX). To this end, we will introduce the quadratic variation of a continuous martingale.

3. Lastly, the condition $\mathsf{E}\int_0^t f^2(s)\,ds < \infty$ can by replaced by a much weaker condition

$$\mathsf{P}\Big(\int_0^t f^2(s)\,ds < \infty\Big) = 1. \tag{5.5}$$

This requires a general technique called "localization". In this context, consider the stopping time

$$T_n = \inf\{t : \int_0^t f^2(s)\,ds \geq n\}.$$

Then $\int_0^{t\wedge T_n} f(s)\,dB_s$ will be a martingale. To define the stochastic integral for all $t > 0$, we only need $T_n \uparrow \infty$ if $n \uparrow \infty$, which follows from (5.5).

5.1 Continuous local martingales

Continuous local martingales form the natural class of processes that will be invariant after stochastic integration. It works well with stopping times.

Definition 5.1 *A process $(M_t)_{t\geq 0}$ is called a* continuous local martingale*, if*

1. *the sample path $t \mapsto M_t(\omega)$ is continuous for all ω, and*

2. *there exists stopping times $T_n \uparrow \infty$ such that $(M_{t\wedge T_n})_{t\geq 0}$ is a (u.i.) martingale.*

Remark 5.1 If $(M_{t\wedge T_n})_{t\geq 0}$ is a u.i. martingale, then $(M_{t\wedge T_n \wedge n})_{t\geq 0}$ is u.i., since it is closed by $X_{T_n\wedge n}$ (see Definition 4.4). This means we can always require T_n to be sequence of bounded stopping times.

Proposition 5.1 *Let $(M_t)_{t\geq 0}$ be a c.l.m. and T be any stopping time. Then $(M_{t\wedge T})_{t\geq 0}$ is also a continuous local martingale.*

Proof: By definition and Remark 5.1, for some bounded stopping time $T_n \uparrow \infty$, $X_t = M_{t\wedge T_n}$ form a u.i. martingale, and hence by Theorem 4.10,

$$M_{(t\wedge T)\wedge T_n} = \mathsf{E}[M_{T_n} \mid \mathcal{F}_{t\wedge T}],$$

so by Proposition 4.7, $(M_{t\wedge T_n \wedge T})_{t\geq 0}$ is u.i. □

Proposition 5.2 *Let M be a c.l.m. If there exists $Z \in L^1$ such that $|M_t| \leq Z$ for all t, then M is a u.i. martingale.*

Proof: Suppose that $(M_{t\wedge T_n})_{t\geq 0}$ is u.i. Then by Theorem 4.10, for $t > s$,

$$M_{s\wedge T_n} = \mathsf{E}\big[M_{t\wedge T_n} \mid \mathcal{F}_s\big]. \tag{5.6}$$

By assumption, $|M_{s\wedge T_n}|, |M_{t\wedge T_n}| \leq Z$, so by Dominated Convergence Theorem, we can take $T_n \uparrow \infty$ in (5.6) to get $\mathsf{E}M_t = [M_t \mid \mathcal{F}_s]$. Uniform integrability follows from the fact that M_t is dominated by Z for all t. □

Proposition 5.3 *Let $S_n = \inf\{t \geq 0 : |M_t| \geq n\}$. Then $(M_{t\wedge S_n})_{t\geq 0}$ is a u.i. martingale.*

Proof: By Proposition 5.1, $(M_{t\wedge S_n})_{t\geq 0}$ is a c.l.m. But $|M_{t\wedge S_n}| \leq n$ for all t, so by Proposition 5.2 it is a u.i. martingale. □

Remark 5.2 This means that we can remove the "uniform integrability" assumption from the definition of continuous local martingales.

5.2 Quadratic variation for continuous local martingales

In this section, for a partition $\Delta : 0 = t_0 < t_1 < \cdots < t_n = t$, $|\Delta|$ will be the maximum length of the intervals in Δ. For a process $(X_t)_{t\geq 0}$, we write $\Delta X_i = X_{t_{i+1}} - X_{t_i}$ for short if there is no ambiguity.

Theorem 5.4 *Let $(M_t)_{t\geq 0}$ be a c.l.m. Then the* quadratic variation process

$$\langle M, M\rangle_t = \langle M\rangle_t = \mathsf{P}\text{-}\lim_{|\Delta|\to 0} \sum_{t_i\in\Delta} (M_{t_{i+1}} - M_{t_i})^2 \tag{5.7}$$

exists, and $M_t^2 - \langle M\rangle_t$ is a c.l.m.

We should compare Theorem 5.4 with the Doob–Meyer decomposition for local sub-martingales.

Theorem 5.5 (Doob-Meyer Decomposition) *Let $(X_t)_{t\geq 0}$ be a continuous local sub-martingale. Then there exists a c.l.m. $(M_t)_{t\geq 0}$ and a continuous increasing process $(A_t)_{t\geq 0}$ such that*

$$X_t = M_t + A_t. \tag{5.8}$$

The decomposition (5.8) is unique up to an additive constant.

For the detailed proof of Theorem 5.5, one can see [KS98, Chap. 1]. Here we only give the proof of uniqueness, which itself is an interesting fact about c.l.m's. Note that the quadratic variation process $\langle M\rangle_t$ in Theorem 5.4 is the increasing process in Theorem 5.5.

Proof of existence of Theorem 5.5: Suppose there are two decompositions

$$X_t = M_t + A_t = M'_t + A'_t.$$

Then

$$Y_t = A'_t - A_t = M_t - M'_t$$

is both a c.l.m. and has finite variation (as it is the difference of two increasing functions). We will show that such process Y_t must be a constant.

Without loss of generality we assume $Y_0 = 0$. Fix K and define

$$T = \inf\{t \geq 0 : |A_t| + |A'_t| \geq K\}.$$

Consider the c.l.m. $Z_t = Y_{t\wedge T}$. Since $|Z_t| \leq K$, by Proposition 5.2 it is in fact a u.i. martingale. Then we have for any partition $0 = t_0 < t_1 < \cdots < t_m = t$,

$$\mathsf{E}Z_t^2 = \sum_{k=0}^{m-1} (Z_{t_{k+1}} - Z_k)^2 \leq K\mathsf{E} \sup_{0\leq k\leq m-1} |Z_{t_{k+1}} - Z_{t_k}|. \tag{5.9}$$

Since Z is continuous, $\sup\limits_{0\leq k\leq m-1} |Z_{t_{k+1}} - Z_{t_k}| \to 0$ a.s., so by Bounded Convergence Theorem, the expectation at the right-hand side of (5.9) goes to zero. Hence $Z_t = Y_{t\wedge T}$ for every K. Letting $K \uparrow \infty$ we obtain $Y_t = 0$ for all t. $\square$

Next we will prove Theorem 5.4. Let us first look at the case of Brownian motion. We already know that for any partition Δ,

$$\mathsf{E} \sum_{t_i\in\Delta} |\Delta B_i|^2 = \sum_{t_i\in\Delta} |\Delta t_i| = t.$$

We will show the L^2-convergence

$$\mathsf{E}\Big| \sum_{t_i\in\Delta} |\Delta B_i|^2 - t\Big|^2 \to 0, \quad |\Delta| \to 0, \tag{5.10}$$

which implies the convergence in probability. Indeed,

$$\begin{aligned}
\mathsf{E}\Big| \sum_{t_i\in\Delta} |\Delta B_i|^2 - t\Big|^2 &= \mathsf{E}\Big\{ \sum_i \Big[|\Delta B_i|^2 - \Delta t_i\Big]\Big\}^2 \\
&= \mathsf{E}\sum_{i,j} \Big[|\Delta B_i|^2 - \Delta t_i\Big]\Big[|\Delta B_j|^2 - \Delta t_j\Big] \\
&= \mathsf{E}\sum_i \Big[|\Delta B_i|^2 - \Delta t_i\Big]^2.
\end{aligned}$$

In the last line we use that all the cross terms are zero, which follows from the fact that $(B_t^2 - t)_{t\geq 0}$ is a martingale. To see this, for $i > j$, we have

$$0 = \mathsf{E}[B_{t_{i+1}}^2 - B_{t_i}^2 - (t_{i+1} - t_i) \mid \mathcal{F}_{t_{j+1}}] = \mathsf{E}[|\Delta B_i|^2 - \Delta t_i \mid \mathcal{F}_{t_{j+1}}], \tag{5.11}$$

and since $|\Delta B_j|^2 - \Delta t_j \in \mathcal{F}_{t_{j+1}}$,

$$\mathsf{E}\Big[|\Delta B_i|^2 - \Delta t_i\Big]\Big[|\Delta B_j|^2 - \Delta t_j\Big] = \mathsf{E}\Big[|\Delta B_j|^2 - \Delta t_j\Big]\mathsf{E}\Big[|\Delta B_j|^2 - \Delta t_j \mid \mathcal{F}_{t_{j+1}}\Big] = 0.$$

Finally, it is easy to see that

$$\sum_i \mathsf{E}\big(|\Delta B_i|^2 - \Delta t_i\big)^2 \leq C\sum_i |\Delta t_i|^2 \leq C|\Delta| \sum_i |\Delta t_i| \leq C|\Delta| t \to 0$$

as desired.

Homework Let $t > 0$ and consider the partition $\Delta_n : t_i = it \cdot 2^{-n}$, $0 \le i \le 2^n$. Show that

$$\lim_{n\to\infty} \sum_{i=0}^{2^n-1} (B_{t_{i+1}} - B_{t_i})^2 = t, \quad \text{a.s.}$$

Hint: Compute the L^2-distance, and then use Markov inequality and Borel–Cantelli.

Proof of Theorem 5.4: Since $(M_t)_{t\ge0}$ is a c.l.m., there are $T_n \uparrow \infty$ such that $(M_{t\wedge T_n})_{t\ge0}$ is a bounded martingales. Then $(M^2_{t\wedge T_n})_{t\ge0}$ is a sub-martingale, and by Theorem 5.5 there exists a continuous martingale N_t and a continuous increasing process A_t such that $M_t^2 = N_t + A_t$. We can further assume that A_t is bounded, otherwise we replace the stopping T_n by

$$\tilde{T}_n = T_n \wedge \inf\{t \ge 0 : A_t \ge n\}.$$

So let us first prove the statement under condition $|M_t|, |A_t| \le K$ for some $K > 0$. Now $N_t = M_t^2 - A_t$ is a bounded c.l.m., so it is a martingale.

We will show

$$\mathsf{E}\Big|\sum_i (\Delta M_i)^2 - A_t\Big|^2 \to 0.$$

In fact, the left-hand side is equal to

$$\sum_{i,j} \mathsf{E}\Big((\Delta M_i)^2 - \Delta A_i\Big)\Big((\Delta M_j)^2 - \Delta A_j\Big).$$

Since $N_t = M_t^2 - A_t$ is a martingale, by the same computation as (5.11), all the cross terms are zero. For the diagonal terms, we have

$$\begin{aligned}\mathsf{E}\sum_i \Big[(\Delta M_i)^2 - \Delta A_i\Big]^2 &\le 2\mathsf{E}\sum_i |\Delta M_i|^4 + 2\mathsf{E}\sum_i |\Delta A_i|^2 \\ &\le 2\mathsf{E}\sup_i |\Delta M_i|^2 \cdot \sum_i |\Delta M_i|^2 + 2\mathsf{E}\sup_i |\Delta A_i| \cdot \sum |\Delta A_i|.\end{aligned}$$

For the second term, $\sup_i |\Delta A_i| \to 0$ a.s. by continuity of A, so the expectation goes to 0 by Bounded Convergence Theorem. For the first term, we use Cauchy–Schwartz and obtain

$$\mathsf{E}\sup_i |\Delta M_i|^2 \cdot \sum_i |\Delta M_i|^2 \le \Big[\mathsf{E}\sup_i |\Delta M_i|^4\Big]^{1/2} \cdot \Big[\mathsf{E}\Big(\sum_i |\Delta M_i|^2\Big)^2\Big]^{1/2}.$$

The first term goes to zero by the continuity of M and Bounded Convergence Theorem. It remains to show that the second term is bounded. In fact, after we expand the square, for the diagonal terms we have

$$\mathsf{E}\sum_i |\Delta M_i|^4 \le 4K^2\mathsf{E}\sum_i |\Delta M_i|^2 = 4K^2\mathsf{E}M_t^2 \le 4K^4,$$

and for the cross terms we have:

$$\mathsf{E}\sum_{j:j>i} |\Delta M_j|^2|\Delta M_i|^2 = \mathsf{E}|\Delta M_i|^2 \cdot \mathsf{E}\big[\sum_{j:j>i} |\Delta M_j|^2 \mid \mathcal{F}_{t_{i+1}}\big] = \mathsf{E}|\Delta M_i|^2 \cdot \mathsf{E}\big[M_t^2 - M_{t_{i+1}}^2 \mid \mathcal{F}_{t_{i+1}}\big] \le 2K^2 \cdot \mathsf{E}|\Delta M_i|^2,$$

and summing over all i we obtain that the sum of all the cross terms are bounded by CK^4.

Let T_K be the corresponding stopping time. Clearly $\lim_{k\to\infty} \mathsf{P}(T_K > t) = 1$. We have just shown that there is an increasing process A_t such that $M^2_{t\wedge T_K} - A_{t\wedge T_K}$ is a martingale, and $\sum_i (M_{t_{i+1}\wedge T_K} - M_{t_i\wedge T_K})^2 \to A_{t\wedge T_K}$ in L^2 and hence in probability. Now

$$\mathsf{P}\Big(|\sum_i |\Delta M_i|^2 - A_t| > \varepsilon\Big) \leq \mathsf{P}\Big(t < T_K; |\sum_i (M_{t_{i+1}\wedge T_K} - M_{t_i\wedge T_K})^2 - A_{t\wedge T_K}| > \varepsilon\Big) + \mathsf{P}(t \geq T_K).$$

For any $\delta > 0$, we first choose K such that $\mathsf{P}(t \geq T_K) < \delta/2$, and then choose $|\Delta|$ small enough such that

$$\mathsf{P}\Big(|\sum_i (M_{t_{i+1}\wedge T_K} - M_{t_i\wedge T_K})^2 - A_{t\wedge T_K}| > \varepsilon\Big) < \delta/2$$

Then $\mathsf{P}\Big(|\sum_i |\Delta M_i|^2 - A_t| > \varepsilon\Big) < \delta$ as desired. This completes the proof. □

5.3 Cross variation and continuous semi-martingales

Definition 5.2 *Let M, N be two c.l.m.'s. The* cross variation*, or* bracket *of M and N is defined by*

$$\langle M, N\rangle_t = \frac{1}{4}\big(\langle M+N\rangle_t - \langle M-N\rangle_t\big).$$

The cross variation has the following properties.

Proposition 5.6 *Let M, N be c.l.m.'s.*

1. *$\langle M, N\rangle$ is the unique (up to indistinguishability) finite variation process such that $M_tN_t - \langle M, N\rangle_t$ is a c.l.m.*
2. *For every $t \geq 0$, we have convergence in probability*
$$\langle M, N\rangle_t = \lim_{|\Delta|\to 0} \sum (M_{t_{i+1}} - M_{t_i})(N_{t_{i+1}} - N_{t_i}).$$
3. *The map $(M, N) \mapsto \langle M, N\rangle$ is bilinear and symmetric.*
4. *For every stopping time,*
$$\langle M^T, N^T\rangle = \langle M^T, N\rangle = \langle M, N^T\rangle.$$

Proof: For Item 1, noting that $M_tN_t = \frac{1}{4}\big((M_t+N_t)^2 - (M_t-N_t)^2\big)$, the difference of two c.l.m.'s

$$M_tN_t - \langle M, N\rangle_t = \frac{1}{4}\Big[\Big((M_t+N_t)^2 - \langle M+N\rangle_t\Big) - \Big((M_t-N_t)^2 - \langle M-N\rangle_t\Big)\Big]$$

is still a c.l.m. The uniqueness follows the same argument as Theorem 5.5.

For Item 2, it suffices to notice that before taking the limit,

$$\sum \Delta M_i \cdot \Delta N_i = \frac{1}{4}\Big[\sum |\Delta(M+N)_i|^2 - \sum |\Delta(M-N)_i|^2\Big].$$

Item 3 follows from Item 3 since each product $\Delta M_i \cdot \Delta N_i$ is symmetric and bilinear.

Item 4 also follows from Item 3 since

$$\Delta M_i^T \cdot \Delta N_i = (M_{T\wedge t_{i+1}} - M_{T\wedge t_i})(N_{t_{i+1}} - N_{t_i}) = \Delta M_i^T \cdot \Delta N_i^T = \Delta M_i \cdot \Delta N_i^T.$$

□

Definition 5.3 *A process $X = (X_t)_{t\geq 0}$ is called a* continuous semi-martingale (c.sm.) *if it has the decomposition*

$$X_t = M_t + A_t,$$

where M_t is a c.l.m. and A_t is a continuous finite variation process.

The cross variation between c.l.m's can be extended to continuous semi-martingales.

Proposition 5.7 *1. If A is a finite variation process and X is a continuous process, then for every $t > 0$ and partition Δ of $[0,t]$,*

$$\lim_{|\Delta|\to 0} \sum \Delta A_i \cdot \Delta X_i = 0, \;\; a.s.$$

2. If $X = M + A$ and $Y = N + A'$ are two continuous semi-martingales, then for every $t > 0$ and partition Δ of $[0,t]$,

$$\lim_{|\Delta|\to 0} \sum \Delta X_i \cdot \Delta Y_i = \lim_{|\Delta|\to 0} \sum_i \Delta M_i \cdot \Delta N_i = \langle M, N\rangle_t, \;\; \textit{in probability.}$$

In particular, we can define $\langle X, Y\rangle_t = \langle M, N\rangle_t$ as the cross variation between X and Y.

Proof: It suffices to prove the first part. We note that

$$\Big|\sum \Delta A_i \Delta X_i\Big| \leq \big(\sup_i |\Delta X_i|\big) \cdot \sum |\Delta A_i|.$$

By continuity of X, as $|\Delta| \to 0$, the first term converges to 0, while by definition of finite variation processes, the second term is bounded a.s. Hence, the left-hand side converges to 0 a.s. □

6 Stochastic integrals

As we have seen in the discussion at the beginning of Section 5, the stochastic integral

$$\int_0^t Y_s \, dX_s$$

is defined by some limit of the left Riemann sum $\sum Y_{t_i}(X_{t_{i+1}} - X_{t_i})$. We have seen the case where Y_t is a deterministic L^2 function and X is the Brownian motion; this is the stochastic integral constructed in the Gaussian white noise expansion Theorem 2.5. In general, we will need more assumptions on the process X (some martingale properties) than Y. Indeed, the appropriate class of processes to consider is the continuous local semi-martingales.

We will first present the celebrated *Itô's Formula*, which says for twice continuously differentiable function f and a continuous semi-martingale X, $f(X_t)$ is also a continuous semi-martingale, and gives the decomposition into local martingale and finite variation processes. This justifies that continuous semi-martingales are the right class of processes to perform stochastic integration. On the other hand, Itô's Formula plays the role of *Fundamental Theorem of Calculus* in classical calculus.

Then we will detail the approximation scheme to define stochastic integrals. It will rely on some Hilbert space theory and the localization techniques.

6.1 Itô's Formula

Let $\mathcal{C}^2(D) = \{f : D \to \mathbb{R} : \nabla f, \nabla^2 f$ exist and are continuous on $D\}$.

Theorem 6.1 *Let $f \in \mathcal{C}^2(\mathbb{R})$ and $X = M + A$ be a continuous semi-martingale Then $f(X_t)$ is also a continuous semi-martingale, such that*

$$f(X_t) - f(X_0) = \int_0^t df(X_s) = \int_0^t f'(X_s)\, dX_s + \frac{1}{2}\int_0^t f''(X_t)\, d\langle X\rangle_t \tag{6.1}$$

$$= \Big[\int_0^t f'(X_s)\, dM_s\Big] + \Big[\int_0^t f'(X_s)\, dA_s + \frac{1}{2}\int_0^t f''(X_t)\, d\langle X\rangle_t\Big]. \tag{6.2}$$

The first and second brackets in (6.2) *are the local martingale and the finite variation term for the continuous semi-martingale $f(X_t)$, respectively.*

We also formally write (6.1) *and* (6.2) *in the derivative form*

$$df(X_t) = f'(X_t)\, dX_t + \frac{1}{2}f''(X_t)d\langle X\rangle_t = f'(X_t)\, dM_t + f'(X_t)\, dA_t + \frac{1}{2}f''(X_t)d\langle X\rangle_t. \tag{6.3}$$

In (6.2), the second and third integrals can be interpreted as Riemann–Stieltjes integral, so only the first integral is new.

Since we have not defined stochastic yet, we will only assume the following fact in our proof of Theorem 6.1: if M is a c.l.m. and Y is a nice process (in the theorem $Y_t = f'(X_t)$), then as a limit in probability the stochastic integral

$$\int_0^t Y_s\, dM_s = \lim_{|\Delta|\to 0} \sum Y_{t_i}(M_{t\wedge t_{i+1}} - M_{t\wedge t_i}) \tag{6.4}$$

can be defined, and is also a c.l.m.

There is also a multi-dimensional version of the Itô's Formula.

Theorem 6.2 *Let $f \in \mathcal{C}^2(\mathbb{R}^d)$ and $X^{(1)}, \dots, X^{(d)}$ be continuous semi-martingales. Then $f(X_t) = f(X_t^{(1)}, \dots, X_t^{(d)})$ is also a continuous semi-martingale, and*

$$df(X_t) = \sum_{j=1}^d \frac{\partial f}{\partial x_j}(X_t)dX_t^{(j)} + \frac{1}{2}\sum_{j,k=1}^d \frac{\partial^2 f}{\partial x_j \partial x_k}(X_t)d\langle X^{(j)}, X^{(k)}\rangle_t. \tag{6.5}$$

Remark 6.1 One can take $X_t^{(1)} = t$, so the function f can also depend on time. In this case, since $X_t^{(1)} = t$ has finite variation, $\langle X^{(1)}, X^{(j)}\rangle_t = 0$ for all $j \neq 1$.

It is not enough to give definition for the stochastic integral (6.4). After applying Itô's formula multiple times, it is inevitable to compute the cross variation between stochastic integrals. Namely, if $dX_t = H_t dM_t$, $dY_t = K_t dN_t$ are two c.l.m.'s given by the stochastic integral, we need to know $\langle X, Y\rangle_t$ to apply Itô's formula again on X and Y. This will be a key property of stochastic integral we need to establish. We will show

$$d\langle X, Y\rangle_t = H_t K_t d\langle M, N\rangle_t$$

in this case.

We will prove Theorem 6.1 assuming (6.4). **Proof of Theorem 6.1:** By localization, we can assume that M_t, A_t, f', f'' are all bounded. If they are not, we can define a stopping time

$$T_K = \inf\{t \geq 0 : |M_t| \geq K, |A_t| \geq K, |f'(X_t)| \geq K, |f''(X_t)| \geq K\}$$

and prove the statement for X^{T_K}, and then let $K \to \infty$. Note that by continuity of X and f', f'', we have $T_K \to \infty$ as $K \to \infty$.

Let $\Delta : 0 = t_0 < t_1 < \cdots < t_n = t$ be a partition of $[0, t]$. Applying Taylor's expansion on each interval $[t_i, t_{i+1}]$ with Lagrangian remainder, we have

$$\begin{aligned}
f(X_t) - f(X_0) &= \sum_{i=0}^{n-1} f(X_{t_{i+1}}) - f(X_{t_i}) \\
&= \sum_{i=0}^{n-1} f'(X_{t_i})\Delta X_i + \frac{1}{2} f''(\tilde{X}_{t_i,t_{i+1}})(\Delta X_i)^2 \\
&= \sum_{i=0}^{n-1} f'(X_{t_i})\Delta X_i + \frac{1}{2}\sum_{i=0}^{n-1} f''(X_{t_i})(\Delta X_i)^2 + \frac{1}{2}\sum_{i=0}^{n-1} \big[f''(X_{t_i}) - f''(\tilde{X}_{t_i,t_{i+1}})\big](\Delta X_i)^2 \\
&=: I_1 + I_2 + I_3.
\end{aligned}$$

Here, $\tilde{X}_{t_i,t_{i+1}}$ is some number between X_{t_i} and $X_{t_{i+1}}$.

By (6.4), $I_1 \to \displaystyle\int_0^t f'(X_s)\, dX_s$ in probability as $|\Delta| \to 0$. Denote the modulus of continuity by

$$\omega(g, \delta) = \sup_{x \neq y,\ |x-y| \leq \delta} |g(x) - g(y)|.$$

For I_3 we have

$$\begin{aligned}
I_3 &\leq \Big(\sup_{0 \leq i \leq n-1} |f''(X_{t_i}) - f''(X_{t_i,t_{i+1}})| \Big) \cdot \sum_{i=0}^{n-1} (\Delta X_i)^2 \\
&\leq \omega\big(f'', \omega(X, |\Delta|)\big) \cdot \sum_{i=0}^{n-1} (\Delta X_i)^2.
\end{aligned}$$

The first term converges to zero a.s. as $|\Delta| \to 0$, since X are bounded and X, f'' are uniformly continuous on compact intervals. The second term converges to $\langle X \rangle_t$ in probability. Hence, their product converges to 0 in probability.

Now it remains to show that

$$I_3 \to \frac{1}{2} \int_0^t f''(X_s)\, d\langle X \rangle_s. \tag{6.6}$$

We will indeed show that (6.6) holds almost surely. We will use some measure theory argument. Recall that a sequence of r.v.'s have a limit in probability if and only if every subsequence has a further subsequence that converges almost surely to that limit. In case of the quadratic variation process $\langle X \rangle$, there exist partition Δ_n on $[0, t]$ with $|\Delta_n| \to 0$ such that with probability one,

$$\sum_{t_i \in \Delta_n} \big(X_{s \wedge t_{i+1}} - X_{s \wedge t_i}\big)^2 \to \langle X \rangle_s \tag{6.7}$$

for a fixed $s > 0$.

By the diagonalization method, we can find require that (6.7) holds *simultaneously* for all $s \in \mathbb{Q} \cap [0, t]$. Indeed, enumerate $\mathbb{Q} \cap [0, t]$ as $q_1, q_2, \ldots$. We first have a sequence of partition $(\Delta_n^{(1)})_{n \geq 1}$ such that (6.7) holds for $t = q_1$. Then, there exists a subsequence $(\Delta_n^{(2)})_{n \geq 1} \subset (\Delta_n^{(1)})_{n \geq 1}$ such that (6.7) holds for $t = q_2$, but being a subsequence, it also holds for $t = q_1$. Continuing this construction we obtain $(\Delta_n^{(k)})_{n\geq 1}$ that (6.7) holds simultaneously for $t = q_1, \ldots, q_k$. Finally, the desired sequence of partitions will be given by the diagonal sequence, $\Delta_n = \Delta_n^{(n)}$, which is a subsequence of every $(\Delta_n^{(k)})_{n\geq 1}$.

Since the limiting process $\langle X\rangle_s$ is increasing and continuous, if (6.7) holds for all $s \in \mathbb{Q} \cap [0,t]$, it holds for all $s \in [0,t]$ since $\mathbb{Q}$ is a dense subset of $\mathbb{R}$. Define the measures μ_n on $[0,t]$ by

$$\mu_n = \sum_{t_i\in\Delta_n} \delta_{X_{t_i}} (X_{t_{i+1}} - X_{t_i})^2$$

Then the distribution of $\mu_n[0,s]$ converges to $\langle X\rangle_s$ for all $s \in [0,t]$. Hence, with probability one, the measure μ_n converge weakly to the measure $\mu(ds) = d\langle X\rangle_s$. By weak convergence, for the continuous function $g(s) = f''(X_s)$, we have

$$\int_0^t g(s)\,\mu_n(ds) = \sum_{t_i\in\Delta_n} f''(X_{t_i})(X_{t_{i+1}} - X_{t_i})^2 \to \int_0^t g(s)\,\mu(ds) = \int_0^t f''(X_s)\,d\langle X\rangle_s.$$

This proves (6.6) and completes the proof of the theorem. $\square$

6.2 Some preparation

We define the space

$$\begin{aligned}\mathbb{H}^2 &= \{M :\ \text{continuous martingale, } \sup_{t\geq 0} \mathsf{E} M_t^2 < \infty,\ M_0 = 0\},\\ &= \{M :\ \text{continuous local martingale, } \mathsf{E}\langle M, M\rangle_\infty < \infty,\ M_0 = 0\}.\end{aligned}$$

This will be the martingale that will replace the Brownian motion. In fact, Brownian motion is not in $\mathbb{H}^2$, but $B^T \in \mathbb{H}$ for all bounded stopping time T.

The equivalence of these two definitions of $\mathbb{H}$ is guaranteed by the following proposition. For its proof, see [LeG16, Theorem 4.13].

Proposition 6.3 *Let M be a c.l.m. with $M_0 = 0$. Then M is a martingale and $\sup_t \mathsf{E} X_t^2 < \infty$ if and only if $\mathsf{E}\langle M\rangle_\infty < \infty$. And when this holds, $M_t^2 - \langle M\rangle_t$ is u.i. and $\mathsf{E}\langle M\rangle_\infty = \mathsf{E} M_\infty^2$.*

The space $\mathbb{H}^2$ is an inner product space, on which the norm and inner product is given by

$$\begin{aligned}\|M\|_{\mathbb{H}^2}^2 &= \mathsf{E}\langle M\rangle_\infty = \mathsf{E} M_\infty^2,\\ \langle M, N\rangle_{\mathbb{H}^2} &= \mathsf{E}\langle M, N\rangle_\infty = \mathsf{E} M_\infty N_\infty.\end{aligned}$$

In fact, $\mathbb{H}$ is a Hilbert space, i.e., an inner product space which is also complete.

Theorem 6.4 *Every Cauchy sequence in $\mathbb{H}^2$ has a limit in $\mathbb{H}^2$. Hence, $\mathbb{H}$ is a Hilbert space.*

Sketch: Let M^n be a Cauchy sequence, i.e., $\mathsf{E}\langle M^m, M^n\rangle_\infty \to 0$ for $n, m \to \infty$. Then by Theorem 4.14, we have

$$\mathsf{E}\sup_{t\geq 0} |M_t^m - M_t^n|^2 \leq 4\mathsf{E}\langle M^m, M^n\rangle_\infty \to 0.$$

The rest is essentially the same as the argument given at the end of Section 4.4. $\square$

Next, we define what should be the integrand process. Let $M \in \mathbb{H}^2$. We define

$$L^2(M) = \{H :\ \text{progressively measurable, } \mathsf{E}\int_0^\infty H_s^2\,d\langle M\rangle_s\}.$$

The space $L^2(M)$ can be identified as a L^2 space. Indeed, define the *progressive σ-field*

$$\mathcal{P} = \{A \in \mathcal{F}_\infty : A \cap \big(\Omega \times [0,t]\big) \in \mathcal{B}([0,t]) \times \mathcal{F}_t,\ \forall t \geq 0\}.$$

Then for $Q = d\mathsf{P}d\langle M\rangle$ defined by

$$Q(A) = \mathsf{E}\int_0^\infty \mathbb{1}_A(\omega,s)\,d\langle M\rangle_s = \int d\mathsf{P}(d\omega)\int_0^\infty \mathbb{1}_A(\omega,s)\,d\langle M^\omega\rangle_s, \quad A \in \mathcal{P}, \tag{6.8}$$

we have

$$L^2(M) = L^2\big(\Omega \times [0,\infty), \mathcal{P}, Q = d\mathsf{P}d\langle M\rangle\big).$$

The condition $M \in \mathbb{H}$ ensures that Q is a finite measure.

Note that the order of integration in (6.8) cannot be changed, since $d\langle M\rangle$ depends on ω. In some sense, (6.8) is more like a conditional expectation decomposition.

A special case is $(M_s = B_{t\wedge s})_{s\geq 0}$, where $d\langle M\rangle_s = ds$ is independent of ω. Then Q will have the product form $Q = \mathsf{P} \otimes ds$.

As an L^2-space, the norm on $L^2(M)$ is defined by

$$\|H\|^2_{L^2(M)} = \mathsf{E}\int_0^\infty H_s^2\,d\langle M\rangle_s.$$

Finally, as we are doing approximation of stochastic integral by left Riemann sum, we need to know that $L^2(M)$ has a dense subset that takes a simple form.

Define the space of elementary functions

$$\mathcal{E} = \{H : H_s(\omega) = H_0(\omega) + \sum_{i=0}^\infty H_{t_i}(\omega)\mathbb{1}_{(t_i,t_{i+1}]}(s),\ H_{t_i} \in \mathcal{F}_{t_i}\}.$$

Theorem 6.5 *Let $M \in \mathbb{H}^2$. The set $\mathcal{E}$ is dense in $L^2(M)$, i.e., for every progressively measurable process H, there exist $H^n \in \mathcal{E} \cap L^2(M)$ such that*

$$\|H^n - H\|_{L^2(M)} \to 0, \quad n \to \infty.$$

Sketch: If H is continuous, we can define

$$H^n(\omega,s) = \sum_{i=0}^{n^2} H_{i/n}(\omega)\mathbb{1}_{(i/n,(i+1)/n]}(s).$$

Then for a.e. ω, since $H(\omega,\cdot) \in L^2(\mathbb{R}_{\geq 0}, d\langle M\rangle)$ and continuous,

$$\int_0^\infty |H^n(\omega,s) - H(\omega,s)|^2\,d\langle M\rangle_s \to 0.$$

It is not hard to show that the limit holds after taking expectation E.

If H is not continuous, we can approximate H by the continuous process,

$$\tilde{H}^m_t = \frac{\int_{(t-1/m)_+}^t H_s\,ds}{(1/m)\vee t},$$

since by Lebesgue Differentiation Theorem, for a.s. ω, $\tilde{H}^m \to H$ a.s. in t and in L^2. Then we can use the approximation on $\tilde{H}^m$ in the first step. □

6.3 Stochastic integral for square integrable martingales

Step 1: $H \in \mathcal{E}$.
Let $M \in \mathbb{H}^2$ and $H \in \mathcal{E} \cap L^2(M)$. It only makes sense to define the stochastic integral as

$$(H \cdot M)_t = \int_0^t H_s \, dM_s := \sum_{i=0}^{\infty} H_{t_i}(\omega)\big(M_{t\wedge t_i} - M_{t\wedge t_{i+1}}\big).$$

One can verify that $H \cdot M \in \mathbb{H}^2$, with

$$\|H \cdot M\|_{\mathbb{H}^2}^2 = \mathsf{E}\Big(\int_0^\infty H_s \, dM_s\Big)^2 = \mathsf{E}\int_0^\infty H_s^2 \, d\langle M\rangle_s = \sum_{i=0}^{\infty} \mathsf{E} H_{t_i}^2 (M_{t_{i+1}} - M_{t_i})^2.$$

The identity

$$\|H \cdot M\|_{\mathbb{H}^2} = \|H\|_{L^2(M)} \tag{6.9}$$

is known as *Itô's isometry*.

From $\mathcal{E}$ to $L^2(M)$.
Let $H \in L^2(M)$. By Theorem 6.5, there are $H^n \in \mathcal{E}$ such that

$$\|H^n - H\|_{L^2(M)}^2 = \mathsf{E}\int_0^\infty (H_s^n - H_s)^2 \, d\langle M\rangle_s \to 0.$$

By (6.9),

$$\|H^n \cdot M - H^m \cdot M\|_{\mathbb{H}^2} = \|H^n - H^m\|_{L^2(M)} \to 0, \quad n, m \to 0,$$

that is, $H^n \cdot M$ forms a Cauchy sequence in $\mathbb{H}^2$. By Theorem 6.4, there is a unique $X \in \mathbb{H}$ such that $H^n \cdot M \to X$ in $\mathbb{H}^2$. We define $H \cdot M = X$. Clearly, $\|H \cdot M\|_{\mathbb{H}^2} = \lim_{n\to\infty} \|H^n \cdot M\|_{\mathbb{H}^2}$. So (6.9) also holds for $H \cdot M$ defined in this way.

The process $H \cdot M$ can be characterized in the following way.

Theorem 6.6 *Let $H \in L^2(M)$. Then $H \cdot M$ is the unique process in $\mathbb{H}^2$ such that*

$$\langle H \cdot M, N\rangle = H \cdot \langle M, N\rangle$$

or, in the integral form,

$$\langle H \cdot M, N\rangle_t = \int_0^t H_s \, d\langle M, N\rangle_s, \quad t \ge 0.$$

Theorem 6.6 can be used to compute the quadratic variation of two stochastic integrals. Indeed, if $dX_t = H_t \, dM_t$ and $Y_t = K_t \, dN_t$, then

$$\langle X, Y\rangle = \langle H \cdot M, K \cdot N\rangle = H \cdot \langle M, K \cdot N\rangle = H \cdot \big(K \cdot \langle M, N\rangle\big) = (HK) \cdot \langle M, N\rangle, \tag{6.10}$$

or in the derivative form,

$$d\langle X, Y\rangle_t = H_t K_t d\langle M, N\rangle_t.$$

In the last step of (6.10), we in fact use the *chain rule* for Riemann–Stieltjes integral.

Another way to interpret Theorem 6.6 is through the general theory of Hilbert space. We recall below the *Riesz Representation Theorem.*

Theorem 6.7 *Let $\mathcal{H}$ be a Hilbert space. Let $\ell : \mathcal{H} \to \mathbb{R}$ be a bounded linear functional. Then there exists a unique $u \in \mathcal{H}$ such that*

$$\ell(x) = \langle u, x\rangle_{\mathcal{H}}, \quad x \in \mathcal{H}.$$

We also need the following Kunita–Watanabe Inequality

Theorem 6.8 *Let H_s and K_s be measurable processes. Then*

$$\left[\int_0^\infty |H_s||K_s||d\langle M,N\rangle_s|\right]^2 \le \int_0^\infty H_s^2 d\langle M\rangle_s \cdot \int_0^\infty K_s^2 d\langle N\rangle_s.$$

Sketch: Consider the case $H = K \equiv 1$. Note that by Cauchy–Schwartz, we have

$$\left[\sum |\Delta M_i|\cdot|\Delta N_i|\right]^2 \le \sum |\Delta M_i|^2 \cdot \sum |\Delta N_i|^2.$$

Hence, by the definition of cross variation and quadratic variation, we have

$$\left|\langle M,N\rangle_s^t\right|^2 \le \langle M\rangle_s^t \langle N\rangle_s^t, \quad s<t.$$

Then, one can show that the inequality holds for all H, K to be simple functions, and then for all measurable H and K. □

To make the connection, we consider the following linear functional

$$N \in \mathbb{H}^2 \mapsto \mathsf{E}\int_0^\infty H_s\, d\langle M,N\rangle_s.$$

By Theorem 6.8, we have (with $K \equiv 1$)

$$\mathsf{E}\int_0^\infty H_s\, d\langle M,N\rangle_s \le \left[\mathsf{E}\int_0^\infty H_s^2\, d\langle M\rangle_s\right]^{1/2}\left[\mathsf{E}\langle N\rangle_\infty\right]^{1/2} = \|H\|_{L^2(M)}\cdot\|N\|_{\mathbb{H}^2}. \tag{6.11}$$

So by Theorem 6.7, there exists $X \in \mathbb{H}^2$ such that

$$\mathsf{E}\langle X,N\rangle_\infty = \mathsf{E}\int_0^\infty H_s d\langle M,N\rangle_s.$$

Then Theorem 6.6 identifies that $X = H\cdot M$.

Proof of Theorem 6.6: Let $H \in \mathcal{E}\cap L^2(M)$. By direct computation we have

$$\langle H_{t_i}(M_{\cdot\wedge t_{i+1}} - M_{\cdot\wedge t_i}), N\rangle_t = H_{t_i}\Big(\langle M,N\rangle_{t\wedge t_{i+1}} - \langle M,N\rangle_{t\wedge t_i}\Big).$$

Summing over all i we have

$$\langle H\cdot M, N\rangle_t = \int_0^t H_s\, d\langle M,N\rangle_s, \quad t\ge 0.$$

By (6.11), this holds for all $H \in L^2(M)$. □

Finally the stochastic integral we have define work well with stopping time.

Theorem 6.9 *Let $M \in \mathbb{H}^2$ and $H \in L^2(M)$. If T is a stopping time, then*

$$\left(\mathbb{1}_{[0,T]}H\right)\cdot M = (H\cdot M)^T = H\cdot M^T,$$

or more explicitly in the integral form,

$$\int_0^\infty \mathbb{1}_{[0,T]} H_s\, dM_s = \int_0^T H_s\, dM_s = \int_0^T H_s\, dM_{s\wedge T},$$

that is, the stopping time and stochastic integrals behave like normal time and integrals.

Proof: We will use the characterization in Theorem 6.6, although a direct approximation approach is also straightforward.

We will use Item 4 in Proposition 5.6 many times.

For the first inequality, we have for any $N \in \mathbb{H}^2$

$$\langle \mathbb{1}_{[0,T]} H \cdot M, N \rangle = \mathbb{1}_{[0,T]} H \langle M, N \rangle,$$

where for the Riemann–Stieltjes integral,

$$\int_0^\infty \mathbb{1}_{[0,T]}(s) H_s \, d\langle M, N \rangle_s = \int_0^\infty H_s \, d\langle M^T, N \rangle_s.$$

Hence, we have

$$\langle \mathbb{1}_{[0,T]} H \cdot M, N \rangle = \mathbb{1}_{[0,T]} H \langle M, N \rangle = H \cdot \langle M^T, N \rangle = \langle H \cdot M^T, N \rangle.$$

For the second identity,

$$\langle (H \cdot M)^T, N \rangle = \langle H \cdot M, N^T \rangle = H \cdot \langle M, N^T \rangle = H \cdot \langle M^T, N \rangle = \langle H \cdot M^T, N \rangle.$$

□

6.4 Stochastic integral for local martingales

Let

$$L^2_{\text{loc}}(M) = \{H \in \mathcal{P} : \int_0^\infty H_s^2 \, d\langle M \rangle_s < \infty\}.$$

Theorem 6.10 *Let M be a c.l.m. and $H \in L^2_{loc}(M)$.*

1. *There exist stopping times $T_n \uparrow \infty$ a.s. such that $M^{T_n} \in \mathbb{H}^2$, $H \in L^2(M^{T_n})$. There exists a continuous local martingale X such that $X_{t \wedge T_n} = (H \cdot M^{T_n})_t$. We define $H \cdot M$ to be the process X.*
2. *For any c.l.m. N,*
$$\langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle.$$
3. *For any stopping time,*
$$(\mathbb{1}_{[0,T]} H) \cdot M = (H \cdot M)^T = H \cdot M^T.$$

Proof: For the first part, consider

$$T_n = \inf\{t \geq 0 : \int_0^t (1 + H_s^2) \, d\langle M \rangle_s \geq n\}.$$

Then $\langle M^{T_n}, M^{T_n} \rangle_t \leq n$ implies that $M^{T_n} \in \mathbb{H}^2$, and

$$\int_0^\infty H_s^2 \, d\langle M^{T_n} \rangle_s = \int_0^{T_n} H_s^2 \, d\langle M \rangle_s \leq n$$

implies that $H \in L^2(M^{T_n})$. So $H \cdot M^{T_n}$ is well-defined.

To check that X is well-defined, we need to show that if $m > n$ and $X_t = (H \cdot M^{T_n})_t$ for $t \leq T_n$ and $\tilde{X}_t = (H \cdot M^{T_m})_t$ for $t \leq T_m$, then $X_t = \tilde{X}_t$ for $t \leq T_n$. This is due to for $t \leq T_n$,

$$\tilde{X}_t = \mathbb{1}_{[0,T_n]} \tilde{X}_t = (H \cdot M^{T_m \wedge T_n})_t = (H \cdot H^{T_n})_t = X_t.$$

The second identity is by Theorem 6.9.

The second and third parts follows from our definition and Theorems 6.6 and 6.9. □

Homework For two c.l.sm.'s X and Y, define the *Stratonovich integral*

$$\int_0^t Y_s \circ dX_s := \int_0^t Y_s \, dX_s + \frac{1}{2}\langle X, Y\rangle_t.$$

1. Show that if Z is another c.l.sm., then

$$\langle Z, \int Y \circ dX\rangle_t = \int_0^t Y_s \, d\langle X, Z\rangle_s.$$

2. Verify the *chain rule* for Stratonovich integrals: for c.l.sm.'s X, Y and Z,

$$\int_0^t Z_s Y_s \circ dX_s = \int_0^t Z_s \circ d\Big(\int_0^s Y_r \circ dX_r\Big).$$

7 Representation of martingales

7.1 Lévy's characterization of Brownian motions

We say that a stochastic process $B_t = \big(B_t^{(1)}, \cdots, B_t^{(d)}\big) \in \mathbb{R}^d$ is a d-dimensional standard Brownian motion if for each coordinate, $B_t^{(j)}$ is a one-dimensional standard motion.

Theorem 7.1 *Let X be a d-dimensional process. Then X is a d-dimensional Brownian motion if and only if $X^{(j)}$ are c.l.m. with quadratic variation*

$$\langle X^{(j)}, X^{(k)}\rangle_t = \delta_{jk} \cdot t = \begin{cases} t, & j = k, \\ 0, & j \neq k. \end{cases}$$

Example 7.1 (Counter-example) The condition on continuity is essential. As an counterexample, consider the Poisson process defined by

$$N_t^\lambda = \max\{k : \xi_1 + \xi_2 + \cdots + \xi_k \le t\},$$

where $\xi_1, \xi_2, \cdots$ are a sequence of i.i.d. $\mathrm{Exp}(\lambda)$ r.v.'s. Then N_t^λ has independent increments and $N_t^\lambda - N_s^\lambda \sim \mathrm{Poi}(\lambda(t-s))$. One can show that $(N_t^\lambda)^2 - \lambda t$ is a martingale, and hence $\langle N^\lambda\rangle_t = \lambda t$. If $\lambda = 1$, the condition of Theorem 6.10 except continuity of the process is satisfied, but obviously N^1 is not the Brownian motion.

Proof: The "$\Rightarrow$" direction is easy, noting that the quadratic variation of two independent Brownian motion is 0 since $\mathsf{E}\Delta B^{(j)}\Delta B^{(k)} = \mathsf{E}\Delta B^{(j)}\mathsf{E}\Delta B^{(k)} = 0$.

For the other direction, we will show that for every $\xi \in \mathbb{R}^d$, $t > s$, we have

$$\mathsf{E}\Big[e^{i\xi\cdot(X_t - X_s)} \mid \mathcal{F}_s\Big] = e^{-\frac{1}{2}|\xi|^2(t-s)} = e^{i\xi\cdot(B_t - B_s)}.$$

If this is true, then (X_t) will have independent increments, and the increments has the same distribution as the d-dimensional Brownian motion, i.e., the standard $\mathcal{N}(0, I_d)$ Gaussian vector. So indeed X will be a d-dimensional Brownian motion.

It suffices to show that

$$M_t = f(t, X_t) = e^{i\xi\cdot X_t + \frac{1}{2}|\xi|^2 t}$$

is a martingale. The Itô's Formula (Theorem 6.2) applies since $X_t^{(j)}$ are c.l.m.'s. We have

$$\partial_t f = \frac{1}{2}|\xi|^2 f, \quad \nabla_x f = i\xi \cdot f, \quad \partial_{jk} f = -\xi_j \xi_k f.$$

Hence,

$$df(t,X_t) = (\partial_t f + \frac{1}{2}\Delta f)dt + (i\xi \cdot f)dX_t = \sum_{j=1}^{d} i\xi_j dX_j,$$

where we used $\langle X^{(j)}, X^{(k)}\rangle_t = \delta_{jk}t$ so only Δf remains in the Itô correction term. Therefore, $M_t = f(t, X_t)$ is a c.l.m. On the other hand,

$$|M_t| = |e^{i\xi X_t + \frac{1}{2}|\xi|^2 t}| \le e^{\frac{1}{2}|\xi|^2 t},$$

So M_t is a true martingale. This completes the proof. $\square$

Homework Let (W_t^1) and (W_t^2) be two independent Brownian motions starts at $W_0^i = x_i$, $x_1 \neq x_2$. Let $T = T(\omega) = \inf\{t \ge 0 : W_t^1 = W_t^2\}$ be their collision time. Define

$$B_t^i(\omega) = \begin{cases} W_t^i, & t < T(\omega), \\ W_T^i + \dfrac{1}{\sqrt{2}}(W_t^1 + W_t^2 - W_T^1 - W_T^2), & t \ge T(\omega), \end{cases} \qquad i = 1, 2.$$

1. Explain why $T(\omega)$ is a stopping time.
2. Find bounded progressively measurable processes (Y_t^i), (Z_t^i), $i = 1, 2$, such that
$$B_t^i = x_i + \int_0^t Y_s^i \, dW_s^1 + \int_0^t Z_s^i \, dW_s^2, \quad i = 1, 2.$$
3. Use Lévy's characterization to show that (B_t^1) and (B_t^2) are Brownian motions (starting at $x_{1,2}$).
4. Show that $\langle B^1, B^2\rangle_t = (t - T) \vee 0$.

7.2 Martingales as stochastic integrals

Let B be the standard Brownian motion on $(\Omega, \mathcal{F}_\infty, (\mathcal{F}_t)_{t\ge 0}, \mathsf{P})$. In this section, we assume that the filtration is given by the augmentation of the natural filtration, namely,

$$\mathcal{F}_t = \sigma\big(\mathcal{F}_t^B \cup \mathcal{N}_\infty\big), \quad \mathcal{N}_\infty = \{A : \exists N \in \mathcal{F}_\infty^B,\ A \subset N,\ \mathsf{P}(N) = 0\}.$$

By discussion in Section 3.3, the augmented filtration satisfies the usual condition.

Theorem 7.2 *Let B be the standard Brownian motion on $(\Omega, \mathcal{F}_\infty, (\mathcal{F}_t)_{t\ge 0}, \mathsf{P})$ where $(\mathcal{F}_t)_{t\ge 0}$ is the augmented filtration.*

1. *For any $Z \in L^2(\Omega, \mathcal{F}_\infty, \mathsf{P})$, there exists a unique $h \in L^2(B)$ s.t.*
$$Z = \mathsf{E}Z + \int_0^\infty h_s \, dB_s. \tag{7.1}$$
2. *For any L^2-bounded martingale M (i.e., $\sup_t \mathsf{E}M_t^2 < \infty$), there exists a unique $h \in L^2(B)$ and constant C s.t.*
$$M_t = C + \int_0^t h_s \, dB_s. \tag{7.2}$$
3. *For any continuous local martingale M, there exists a unique $h \in L^2_{loc}(B)$ and constant C such that (7.2) holds.*

Remark 7.2 Recall that
$$L^2(B) = L^2([0,\infty)\times\Omega, \mathcal{P}, dt\otimes \mathsf{P}).$$
In principle, so elements in the L^2 space are defined up to a $dt\otimes\mathsf{P}$-null set. As we assume the usual condition, any modification on the $dt\otimes\mathsf{P}$-null set will not affect the progressive measurability of the process.

Proof: We will first prove everything except the existence part of Item 1.

Uniqueness of Item 1. Suppose there are two representations in terms of h_s, $\tilde h_s$. We have
$$Z = \mathsf{E}Z + \int_0^\infty h_s\,dB_s = \mathsf{E}Z + \int_0^\infty \tilde h_s\,dB_s,$$
and hence
$$\mathsf{E}\int_0^\infty (h_s - \tilde h_s)\,dB_s.$$
By Itô's isometry,
$$0 = \mathsf{E}\Big[\int_0^\infty (h_s-\tilde h_s)\,dB_s\Big]^2 = \mathsf{E}\int_0^\infty |h_s(\omega)-\tilde h_s(\omega)|^2\,ds = 0.$$
Therefore, $h_s(\omega) = \tilde h_s(\omega)$ for $dt\otimes\mathsf{P}$-a.e. (s,ω) and there are the same element in $L^2(B)$.

From Item 1 to Item 2.
Since $\sup_t \mathsf{E}M_t^2 < \infty$, M_t are u.i., and by Theorem 4.6, there exists M_∞ such that $M_n\to M_\infty$ in L^1 and a.s. Note that $(M_t)_{0\le t\le\infty}$ is a martingale even we do not assume continuity of M, since
$$\mathsf{E}[M_n\mid\mathcal{F}_t] = M_t,\quad n\ge t \qquad\Rightarrow\qquad \mathsf{E}[M_\infty\mid\mathcal{F}_t] = \lim_{n\to\infty}\mathsf{E}[M_n\mid\mathcal{F}_t] = M_t.$$
We also have $M_\infty\in L^2(\Omega,\mathcal{F}_\infty,\mathsf{P})$, since by Fatou,
$$\mathsf{E}M_\infty^2 = \mathsf{E}\lim_{n\to\infty}M_n^2 \le \liminf_{n\to\infty}\mathsf{E}M_n^2 \le \sup_t \mathsf{E}M_t^2 < \infty.$$
Applying Item 1 with $Z = M_\infty$, there exists $h\in L^2(B)$ such that
$$M_\infty = \mathsf{E}M_\infty + \int_0^\infty h_s\,dB_s.$$
Note that $\Big(\int_0^t h_s\,dB_s\Big)_{0\le t\le\infty}$ is a martingale by the construction of stochastic integral, we have
$$M_t = \mathsf{E}[M_\infty\mid\mathcal{F}_t] = \mathsf{E}M_\infty + \int_0^t h_s\,dB_s.$$
The uniqueness of the representation follows from the uniqueness of Item 1.

From Item 2 to Item 3.
Since $\mathcal{F}_0$ is trivial and $M_0\in\mathcal{F}_0$, $M_0 = C$ a.s. for some constant C. For simplicity we assume $C=0$. Let $T_n = \inf\{t\ge 0: |M_t|\ge n\}$. Then M^{T_n} is L^2-bounded, and by Item 2, there exist $h^{(n)}\in L^2(B)$, such that
$$M_t^{T_n} = \int_0^t h_s^{(n)} dB_s. \tag{7.3}$$
Let $m>n$. Note that $M_t^{T_n}$ has two representations. The first one is (7.3), the second one is
$$M_t^{T_n} = M_{t\wedge T_n}^{T_m} = \int_0^{t\wedge T_n} h_s^{(m)}\,dB_s = \int_0^t \mathbb{1}_{[0,T_n]}h_s^{(m)}\,dB_s,$$

where the last equality follows from the property of stochastic integral regarding stopping times, see Theorem 6.9. By the uniqueness of Item 2, we have

$$h_s^{(n)}(\omega) = \mathbb{1}_{[0,T_m(\omega)]}(s) h_s^{(m)}(\omega), \quad \text{for } dt \otimes \mathsf{P}\text{-a.e. } (s,\omega). \tag{7.4}$$

Hence, we can define

$$h_s(\omega) = h_s^{(n)}(\omega), \qquad \text{if } s \le T_n(\omega) \text{ for some } n.$$

Noting $T_n \uparrow \infty$ a.s. and thanks to (7.4), it is easy to see that this gives a consistent definition of an element in $L^2(B)$ up to $dt \otimes \mathsf{P}$-null sets. This proves the existence of the desired representation. The uniqueness follows from a similar localization argument. □

Next we turn to the proof of existence of Item 1 in Theorem 7.2. Consider the set of random variables

$$\mathcal{H} = \{Z \in L^2(\Omega, \mathcal{F}_\infty, \mathsf{P}) : \text{ the representation in (7.1) exists}\}.$$

We observe that $\mathcal{H}$ is a closed subspace of the linear space $L^2(\Omega, \mathcal{F}_\infty, \mathsf{P})$. The linearity is obvious. For the closedness, let

$$Z_n = \mathsf{E}Z_n + \int_0^\infty h_s^{(n)}\, dB_s$$

be a Cauchy sequence in $L^2(\Omega, \mathcal{F}_\infty, \mathsf{P})$. Then by Itô's Isometry,

$$\mathsf{E}\int_0^\infty |h_s^{(n)} - h_s^{(m)}|^2\, ds = \mathsf{E}|Z_n - Z_m|^2,$$

and hence $(h^{(n)})_{n\ge 1}$ is Cauchy in $L^2(B)$. But the space $L^2(B)$ is complete (Theorem 6.5), and hence there exists $h \in L^2(B)$ as the $L^2(B)$-limit of $h^{(n)}$. Then $Z = \lim_{n\to\infty} \mathsf{E}Z_n + \int_0^\infty h_s\, dB_s$ will be the L^2-limit of Z_n and the closedness of $\mathcal{H}$ is proved.

Now, to prove $\mathcal{H} = L^2(\Omega, \mathcal{F}_\infty, \mathsf{P})$, it suffices to show that $\mathcal{H}$ contains a dense subset of $L^2(\Omega, \mathcal{F}_\infty, \mathsf{P})$. The proof of the existence of Item 1 in Theorem 7.2 will be completed by the following two lemmas.

Lemma 7.3 *For all $\lambda_j \in \mathbb{R}$, $0 = t_0 < t_1 < \ldots < t_m$, the real and imaginary parts of*

$$e^{i\cdot\sum_{j=0}^{m-1}\lambda_j(B_{t_{j+1}-B_j})}$$

is element in $\mathcal{H}$.

Lemma 7.4 *The random variables in the form*

$$e^{i\cdot\sum_{j=0}^{m-1}\lambda_j(B_{t_{j+1}-B_j})}, \quad \lambda_j \in \mathbb{R}, 0 = t_0 < t_1 < \cdots < t_m \tag{7.5}$$

are dense in $L^2_{\mathbb{C}}(\Omega, \mathcal{F}_\infty, \mathsf{P})$.

Idea and intuition: Any r.v. measurable with respect to $\mathcal{F}_\infty$ can be approximated by the form $f(B_{t_1}, \ldots, B_{t_m})$ where f is a sufficiently nice function. One the other hand, by the the theory of Fourier transform, we can write

$$f(B_{t_1}, \ldots, B_{t_m}) = \int e^{2\pi i\xi\cdot B} \hat{f}(\xi)\, d\xi$$

for some function $\hat{f}(\omega)$. As integrals can be approximated by Riemann sums, this suggests that the left hand side can be approximated by linear combination of r.v.'s in the form of (7.5).

In principle, one can also approximate a generic r.v. $f(B_{t_1},\dots,B_{t_m})$ by polynomials in $B_{t_1},\dots,B_{t_m}$. But growth of polynomials is difficult to control, while complex exponentials are always bounded, and thus serve as perfect choices for basis functions. $\square$

The proof of Lemma 7.3 follows from a general statement below about exponential martingales, which we will also use a lot when discussing the Girsanov Theorem in Section 8.

Proposition 7.5 *Let L be a c.l.m. Then*

$$\mathcal{E}_t(L) = e^{L_t - \frac{1}{2}\langle L\rangle_t}, \tag{7.6}$$

is a c.l.m., and $d\mathcal{E}_t(L) = \mathcal{E}_t dL_t$.

Proof: Let $X_t = L_t - \dfrac{1}{2}\langle L\rangle_t$. Then $dX_t = dL_t - \dfrac{1}{2}d\langle L\rangle_t$ and

$$d\langle X\rangle_t = d\langle L\rangle_t$$

since $\langle L\rangle_t$ is a finite variational process and does not affect the quadratic variation (Proposition 5.7). By Itô's formula, we have

$$de^{X_t} = e^{X_t}\,dX_t + \frac{1}{2}e^{X_t}d\langle X\rangle_t = e^{X_t}\big(dL_t - \frac{1}{2}d\langle L\rangle_t + \frac{1}{2}d\langle L\rangle_t\big) = e^{X_t}\,dL_t,$$

Hence $\mathcal{E}_t = e^{X_t}$ is a c.l.m. and $d\mathcal{E}_t = \mathcal{E}_t\,dL_t$. $\square$

Proof of Lemma 7.3 : Let

$$f(s) = \sum_{j=0}^{m-1} \lambda_j \mathbb{1}_{(t_j,t_{j+1}]}(s)$$

and $L_t = i\displaystyle\int_0^t f_s\,dB_s$. Then $\langle L\rangle_t = -\displaystyle\sum_{j=0}^{m-1} \lambda_j^2(t\wedge t_{j+1} - t\wedge t_j)$ By Proposition 7.5, we have

$$\mathcal{E}_{t_m}(L) = e^{i\sum\limits_{j=0}^{m-1}\lambda_j(B_{t_{j+1}}-B_{t_j}) + \frac{1}{2}\sum\limits_{j=0}^{m-1}\lambda_j^2(t_{j+1}-t_j)} = 1 + \int_0^{t_m}\mathcal{E}_s(L)f_s\,dB_s.$$

Therefore,

$$e^{i\sum\limits_{j=0}^{m-1}\lambda_j(B_{t_{j+1}}-B_{t_j})} = e^{-\frac{1}{2}\sum\limits_{j=0}^{m-1}\lambda_j^2(t_{j+1}-t_j)}\Big[1 + \int_0^{t_m}\mathcal{E}_s(L)f_s\,dB_s\Big] \in \mathcal{H}.$$

$\square$

A surprising consequence of Theorem 7.2 is that any martingale with respect to the augmented filtration must be continuous. The intuition is that the augmented filtration is generated by a continuous process, namely, the Brownian motion. We have seen the process itself can enforce some properties on the filtration, see for example Theorem 3.18, so it also makes sense that the filtration will determine some properties of an adapted process. The precise statement is the following.

Proposition 7.6 *All martingales adapted to the augmented filtration $(\mathcal{F}_t)_{t\geq 0}$ has a continuous modification.*

Proof: Let $M = (M_t)_{t\geq 0}$ be the martingale. If M_t is L^2-bounded, then the statement follows from Theorem 7.2.

If not, we can assume that M_t is u.i., otherwise we can just discuss the continuity of the martingale $M_{t\wedge a}$ for each a. By Theorem 4.6, uniform integrability implies that there exists M_∞ such that $(M_t)_{0\leq t\leq\infty}$ is a martingale. We can find $M_\infty^{(n)}$ bounded such that $M_\infty^{(n)} \to M_\infty$ in L^1 (for example, $M_\infty^{(n)} = (-n) \vee M_\infty \wedge n$). Then by Theorem 7.2, $M_t^{(n)} = \mathsf{E}[M_\infty^{(n)} \mid \mathcal{F}_t]$ are continuous martingale. Now by Doob's maximal inequality, we can choose a subsequence (which we still denote by $M^{(n)}$) such that

$$\mathsf{P}\Big(\sup_{t\geq 0} |M_t^{(n)} - M_t^{(n+1)}| \geq 2^{-n}\Big) \leq 2^{-n}.$$

Hence, by Borel–Cantelli, almost surely we have

$$\sum_{n=1}^{\infty} \sup_{t\geq 0} |M_t^{(n)} - M_t^{(n+1)}| < \infty,$$

and hence

$$\tilde{M}_t = \lim_{n\to\infty} M_t^{(n)}$$

exists and the limiting process $\tilde{M}_t$ is continuous as the convergence is uniform convergence for continuous function. It remains to show that $\tilde{M}_t$ is a modification of M_t. In fact,

$$M_t = \mathsf{E}[M_\infty \mid \mathcal{F}_t] = \lim_{n\to\infty} \mathsf{E}[M_\infty^{(n)} \mid \mathcal{F}_t] = \lim_{n\to\infty} M_t^{(n)}, \qquad \text{a.s.}$$

□

7.3 Continuous martingale as time-change Brownian motion

In this section we assume that the filtration $(\mathcal{F}_t)$ satisfies the usual condition Definition 3.11.

Theorem 7.7 *Let M be a c.l.m. such that $\langle M\rangle_\infty = \infty$ a.s. Then, there exists a Brownian motion $(\beta_s)_{s\geq 0}$ such that almost surely,*

$$\forall t \geq 0, M_t = \beta_{\langle M\rangle_t}.$$

Proof: Let $\mathcal{N} = \{\langle M\rangle_\infty < \infty\}$. Then $\mathsf{P}(\mathcal{N}) = 0$ and $\mathcal{N} \in \mathcal{F}_t$ for all $t \geq 0$ since $(\mathcal{F}_t)$ satisfies the usual condition. Let

$$\tau_r(\omega) = \begin{cases} \inf\{t \geq 0 : \langle M\rangle_t \geq r\}, & \omega \in \mathcal{N}^c, \\ 0, & \omega \in \mathcal{N}. \end{cases}$$

Then $\mathcal{N} \in \mathcal{F}_t$ implies that $\tau_r(\omega)$ is a stopping time for every $r \geq 0$. Moreover, by definition and the continuity of $\langle M\rangle$, $r \mapsto \tau_r$ is increasing and left-continuous, and its right limit at every point is given by

$$\tau_{r+} = \lim_{s\downarrow r} \tau_s = \begin{cases} \inf\{t \geq 0 : \langle M\rangle_t > r\}, & \omega \in \mathcal{N}^c, \\ 0, & \omega \in \mathcal{N}. \end{cases}$$

Now let

$$\beta_r = M_{\tau_r}. \tag{7.7}$$

We will show that β_r is a BM adapted to the filtration $\mathcal{G}_r = \mathcal{F}_{\tau_r}$.

Since $r \mapsto \tau_r$ is left-continuous with right limits, and $s \mapsto M_s$ is continuous, their composition $r \mapsto M_{\tau_r} = \beta_r$ is left-continuous with right limits. We will show at the end that β_r is in fact right-continuous, so that Lévy's characterization Theorem 7.1 can be applied. Assuming the continuity of β_r, to show that β_r is BM, it suffices to show that $(\beta_s)_{s\geq 0}$, $(\beta_s^2 - s)_{s\geq 0}$ are both c.l.m.'s.

Let $n \geq s \geq r$. For each $n \geq 1$, M^{τ_n} is a u.i. martingale since $\langle M^{\tau_n}\rangle_\infty = n < \infty$ (Proposition 6.3). Since M^{τ_n} is u.i., we can apply Optional Sampling Theorem to the stopping times $\tau_r < \tau_s$ and obtain

$$\mathsf{E}[\beta_s \mid \mathcal{G}_r] = \mathsf{E}\big[M^{\tau_n}_{\tau_s} \mid \mathcal{F}_{\tau_r}\big] = M^{\tau_n}_{\tau_r} = \beta_r.$$

So $(\beta_s)_{s\geq 0}$ is a martingale.

To see that $(\beta_s^2 - s)_{s\geq 0}$ is a martingale, we apply Optional Sampling Theorem to the uniformly integrable martingale $\big[M^{\tau_n}_s\big]^2 - s$. We have

$$E[\beta_s^2 - s \mid \mathcal{G}_r] = \mathsf{E}\Big[\big[M^{\tau_n}_{\tau_s}\big]^2 - \langle M\rangle_{\tau_s} \mid \mathcal{F}_{\tau_r}\Big] = \big[M^{\tau_n}_{\tau_r}\big]^2 - \langle M\rangle_{\tau_r} = \beta_r^2 - r.$$

We have used the continuity of $\langle M\rangle$ to conclude that $\langle M\rangle_{\tau_u} = u$ for all $u \geq 0$.

Having proved that β_s a BM, we need to show that almost surely,

$$\forall t \geq 0, \quad M_t = \beta_{\langle M\rangle_t}. \tag{7.8}$$

In light of (7.7), if $t = \tau_r$ for some r, then (7.8) follows from (7.7) since

$$M_t = M_{\tau_r} = \beta_r = \beta_{\langle M\rangle_{\tau_r}}.$$

But in general, $r \mapsto \tau_r$ is a increasing function that is left-continuous with right limit, the image of τ_r may not be $\mathbb{R}$. That is, it could happen that $\tau_r < t < \tau_{r+}$ for some r. In this case, $\langle M\rangle_t = r$ for $\tau_r < t < \tau_{r+}$. To verify (7.8), it remains to show

$$M_t = M_{\tau_r}, \quad \tau_r \leq t \leq \tau_{r+}. \tag{7.9}$$

Note that $M_{\tau_{r+}} = \lim_{s\downarrow r} \beta_s$, so (7.9) also implies the (right-)continuity of β. We will put this statement in Lemma 7.8. □

Lemma 7.8 *Let M be a c.l.m. Then with probability one,*

$$\forall 0 \leq a < b, \qquad M_t = M_a, \ \forall t \in [a,b] \quad \Leftrightarrow \quad \langle M\rangle_b = \langle M\rangle_a.$$

Proof: Since both M and $\langle M\rangle$ are continuous process, it suffices show that for fixed $a < b$,

$$\{M_t = M_a, \ \forall t \in [a,b]\} = \{\langle M\rangle_b = \langle M\rangle_a\}, \quad \text{a.s.}. \tag{7.10}$$

Then we can take intersection over all $a, b \in \mathbb{Q}$ and use continuity to get the desired result.

The "⇒" direction of (7.10) follows immediately from the construction of quadratic variation, (5.7).

To show the "⇐" direction, letting $N_t = M_t - M_{t\wedge a}$ and $A = \{\langle M\rangle_b = \langle M\rangle_a\}$, it suffices to show $\mathsf{E}\mathbb{1}_A N_t^2 = 0$, for all $t \in [a,b]$. Then $N_t = 0, t \in [a,b]$ a.s. when $\omega \in A$ and (7.10) follows.

Let

$$T_\varepsilon = \inf\{t \geq 0 : \langle N\rangle_t = \langle M\rangle_t - \langle M\rangle_{t\wedge a} \geq \varepsilon\}.$$

Then $A \subset \{T_\varepsilon \geq b\}$ and hence for $t \in [a,b]$,

$$\mathsf{E}\mathbb{1}_A N_t^2 = \mathsf{E}\mathbb{1}_A N^2_{t\wedge T_\varepsilon} \leq \mathsf{E}N^2_{t\wedge T_\varepsilon} = \mathsf{E}\langle N\rangle_{t\wedge T_\varepsilon} \leq \varepsilon.$$

Since the left hand side is independent of ε, and the above inequality holds for all $\varepsilon > 0$, we must have $\mathsf{E}\mathbb{1}_A N_t^2 = 0$, as desired. □

Homework Let $W = (W^1, W^2)$ be a standard two-dimensional Brownian motion starting from the origin. For functions

$$u(x, y) = x^2 - y^2, \quad v(x, y) = 2xy,$$

define

$$B_t^1 = u(W_t^1, W_t^2), \quad B_t^2 = v(W_t^1, W_t^2).$$

1. Show that B_t^i, $i = 1, 2$, are continuous martingales with

$$\langle B^1 \rangle_t = \langle B^2 \rangle_t, \quad \langle B^1, B^2 \rangle_t = 0,$$

2. Show that there exists an random strictly increasing continuous function φ such that

$$(B_t^1, B_t^2) = (\tilde{B}^1_{\varphi(t)}, \tilde{B}^2_{\varphi(t)}),$$

such that $\tilde{B} = (\tilde{B}^1, \tilde{B}^2)$ is a standard two-dimensional Brownian motion.

3. For a two-dimensional continuous process $X = (X_t^1, X_t^2)$, define

$$T_\pm^X = \inf\{t \ge 0 : (X_t^1 \pm 10)^2 + (X_t^2)^2 \le 1\}$$

to be the hitting times of the unit disks centered at $(\pm 10, 0)$ of X. Show that

$$\mathsf{P}(T_+^B > T_-^B) = \mathsf{P}(T_+^W > T_-^W).$$

8 Girsanov Theorem

8.1 Motivation

8.1.1 Gaussian measures on $\mathcal{C}[0, \infty)$

Let P and $\tilde{\mathsf{P}}$ be two (probability) measures on $(\Omega, \mathcal{F})$. Recall the definition of *absolute continuity*: we say that $\tilde{\mathsf{P}}$ is absolutely continuous with respect to P, written $\tilde{P} \ll \mathsf{P}$, if $\mathsf{P}(A) = 0$ implies $\tilde{\mathsf{P}}(A) = 0$. If $\tilde{P} \ll \mathsf{P}$ and $\mathsf{P} \ll \tilde{\mathsf{P}}$, we say that P and $\tilde{\mathsf{P}}$ are *equivalent*, written $\mathsf{P} \sim \tilde{\mathsf{P}}$.

We have the following result on the Radon–Nikodym derivative.

Theorem 8.1 *If $\tilde{\mathsf{P}} \ll \mathsf{P}$, then there exists $Z = \dfrac{d\tilde{\mathsf{P}}}{d\mathsf{P}} \in L^1(\Omega, \mathcal{F}, \mathsf{P})$ such that*

$$\tilde{\mathsf{P}}(A) = \int_A \frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega)\, \mathsf{P}(d\omega) = \mathsf{E}_{\mathsf{P}} Z \mathbb{1}_A. \tag{8.1}$$

The converse is also true: if (8.1) holds for some r.v. $Z \in L^1(\Omega, \mathcal{F}, \mathsf{P})$, then $\tilde{\mathsf{P}} \ll \mathsf{P}$.

One common example of absolutely continuous measures is the continuous r.v. Recall that X is a continuous r.v. if there exists $\rho_X \in L^1(\mathbb{R})$ such that

$$\mathsf{P}_X(A) = \int_A \rho_X(x)\, dx.$$

That means the distribution of X, P_X, is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$. If $\rho_X > 0$ for a.e. x, then P_X and the Lebesgue measure are equivalent.

Now let X and Y be two continuous random variables with positive densities ρ_X and ρ_Y. We have

$$\mathsf{P}_Y(A) = \int_A \rho_Y(y)\, dy = \int_A \rho_X(x) \cdot \left[\frac{\rho_Y(x)}{\rho_X(x)}\right] dx.$$

Then $\mathsf{P}_Y \ll \mathsf{P}_X$ and $\dfrac{d\mathsf{P}_Y}{d\mathsf{P}_X} = \dfrac{\rho_Y(x)}{\rho_X(x)}$. In the case where $X \sim \mathcal{N}(0,1)$ and $Y = X + v = \mathcal{N}(\mu, 1)$, we have

$$\frac{d\mathsf{P}_Y}{d\mathsf{P}_X}(x) = e^{-(x-v)^2/2 + x^2/2} = e^{v \cdot x - \frac{v^2}{2}}.$$

In fact, we have a more general statement for Gaussian vectors.

Proposition 8.2 *Let $X \in \mathbb{R}^d \sim \mathcal{N}(\mu, Q)$ and $Y = X + v$ where $v \in \mathbb{R}^d$. Then P_Y and P_X are mutually absolutely continuous, if and only if $v \in \mathrm{Im}(Q)$ (so that $Q^{-1}v$ is defined). The Radon–Nikodym derivative is*

$$\frac{dP_Y}{dP_X}(x) = e^{-(x-\mu)^T Q^{-1} v - \frac{1}{2} v^T Q^{-1} v}. \tag{8.2}$$

If Q is non-degenerate, then any translation in direction v will produce another absolutely continuous measure on $\mathbb{R}^d$. When Q is degenerate, then only vectors $v \in \mathrm{Im}(Q)$ will produce an absolutely continuous measures. The space $\mathrm{Im}(Q)$ is known as the *Cameron–Martin space* for the Gaussian measure P_X.

The Brownian motion $B = (B_t)_{t \geq 0}$ induces a measure P on $\mathcal{C}[0, \infty)$. Consider a translation

$$\tilde{B}_t = B_t + h(t), \quad h(t) \in \mathcal{C}[0, \infty), \ h(0) = 0.$$

Then $\tilde{B}$ induces another measure $\tilde{\mathsf{P}}$ on $\mathcal{C}[0, \infty)$. The natural question is when the two measures $\tilde{\mathsf{P}}$ and P are mutually absolutely continuous, or equivalently, what is the Cameron–Martin space for Wiener measure?

The answer is $h \in H_0^1[0, \infty)$ where

$$H_0^1[0, \infty) = \{h_t = \int_0^t g(s)\, ds, \ g \in L^2[0, \infty)\}. \tag{8.3}$$

Let us discuss some intuition behind (8.3). First, consider $h_t = t$. Then $h_{t \wedge T} \in H_0^1$ for any $T > 0$ but $h \notin H_0^1$. Hence, $\tilde{\mathsf{P}}$ and P are mutually absolutely continuous on $\mathcal{C}[0, T]$ for any $T > 0$, but not on $\mathcal{C}[0, \infty)$. To illustrate the second point, consider

$$\mathcal{C}[0, \infty) \supset A_c = \{f : \lim_{t \to \infty} \frac{f(t)}{t} = c\}.$$

By Strong Law of Large Numbers and the independent increment property of Brownian motion, it is easy to see that a.s.,

$$\lim_{t \to \infty} \frac{B_t}{t} = \lim_{n \to \infty} \frac{\sum_{k=1}^n B_k - B_{k-1}}{n} = 0, \quad \lim_{t \to \infty} \frac{\tilde{B}_t}{t} = \lim_{t \to \infty} \frac{B_t + t}{t} = 1.$$

Hence, $\mathsf{P}(A_0) = 0$, $\tilde{\mathsf{P}}(A_0) = 1$ and $\mathsf{P}(A_1) = 0$, $\tilde{\mathsf{P}}(A_1) = 1$. So P and $\tilde{\mathsf{P}}$ are in fact singular with respect to each other. On the other hand, if $h \in H_0^1$, then

$$\lim_{t \to \infty} \frac{h(t)}{t} = \lim_{t \to \infty} \frac{\int_0^t g(s)\, ds}{t} = 0$$

since $g \in L^2[0, \infty)$. So we cannot use the event A_0 to as a counterexample for absolute continuity.

The second perspective is that sometimes one can intuitively think of the standard Brownian motion as a Gaussian measure on $\mathcal{C}[0, \infty)$ with "density"

$$e^{-\frac{1}{2} \int_0^\infty |\dot{b}(s)|^2\, ds}. \tag{8.4}$$

(8.4) makes no sense on its own, since first, the derivative $\dot{b}$ is not defined for $b \in \mathcal{C}[0,\infty)$, and second, there is no equivalent thing as the Lebesgue measure on $\mathcal{C}[0,\infty)$ to discuss the density. Nevertheless, we can do some computation using (8.4): formally we have

$$\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}} = \frac{e^{-\frac{1}{2}\int(\dot{b}+\dot{h})^2}}{e^{-\frac{1}{2}\int|\dot{b}|^2}} = e^{\int \dot{b}\dot{h} - \frac{1}{2}\int|\dot{h}|^2}.$$

One then sees that $\displaystyle\int |\dot{h}|^2$ needs to be finite so that the density is a well-defined quantity.

Lastly, the (8.4) could be derived from the following consideration. Recall that a Gaussian measure is such that $\langle v, b\rangle \sim \mathcal{N}(0, \langle Qv, v\rangle)$, where Q is positive symmetric. In the finite dimensional case, the Gaussian density is given by $C\exp(-\frac{1}{2}\langle Q^{-1}b, b\rangle)$. Here, a bounded variation process $V = \displaystyle\int v$ defines a linear functional on $\mathcal{C}[0,\infty)$:

$$\langle v, b\rangle := \int_0^\infty b_s dV_s = -\int_0^\infty V_s\, db_s \sim \mathcal{N}(0, |V|_{L^2}^2).$$

So we have

$$\langle Qv, v\rangle = \int_0^\infty |V|^2\, ds = -\int_0^\infty v(\int V)\, ds.$$

In other words, $Q = (-\partial_{xx})^{-1}$. Hence, $Q^{-1} = (-\partial_{xx})$ and

$$\langle Q^{-1}b, b\rangle = -\int_0^\infty \partial_{xx}b \cdot b = \int_0^\infty |\dot{b}|^2.$$

The above argument can be made rigorous with the general theory of Gaussian measures on Banach spaces.

8.1.2 Brownian motions under change of measure

Let $\mu \in \mathbb{R}^d$. Let ρ and $\tilde{\rho}$ be the density functions for $\mathcal{N}(0, I_d)$ and $\mathcal{N}(\mu, I_d)$:

$$\rho(u) = (2\pi)^{-d/2}e^{-|x-\mu|^2/2}, \quad \tilde{\rho}(u) = (2\pi)^{-d/2}e^{-|x|^2/2}.$$

Consider a random vector $X = (X_1, \dots, X_d)$ with distribution $\mathcal{N}(\mu, I_d)$, viewed as a measurable map from $(\Omega, \mathcal{F})$ to $\mathbb{R}^d$, which induces a probability measure P on $(\Omega, \mathcal{F})$, that is,

$$\mathsf{P}\circ X^{-1}(A) = \mathsf{P}(X \in A) = \int \mathbb{1}_A(u)\rho(u)\, du, \quad \forall A \in \mathcal{B}(\mathbb{R}^d).$$

Now we define another measure $\tilde{\mathsf{P}}$ which is absolutely continuous with respect to P by

$$\tilde{\mathsf{P}}(\Gamma) = \int \mathbb{1}_\Gamma(\omega)\frac{\tilde{\rho}\big(X(\omega)\big)}{\rho\big(X(\omega)\big)}\mathsf{P}(d\omega).$$

Then the measurable map X has a different measure under $\tilde{\mathsf{P}}$. We have the following computation: for $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}
\tilde{\mathsf{P}}\circ X^{-1}(A) = \tilde{\mathsf{P}}(X \in A) &= \int \mathbb{1}_{X(\omega)\in A}(\omega)\frac{\tilde{\rho}\big(X(\omega)\big)}{\rho\big(X(\omega)\big)}\mathsf{P}(d\omega) \\
&= \mathsf{E}\mathbb{1}_A(X)\frac{\tilde{\rho}(X)}{\rho(X)} \\
&= \int \Big[\mathbb{1}_A(u)\frac{\tilde{\rho}(u)}{\rho(u)}\Big]\cdot\rho(u)\, du \qquad = \int \mathbb{1}_A(u)\tilde{\rho}(u).
\end{aligned}$$

That is, $\tilde{\mathsf{P}} \circ X^{-1} = \mathcal{N}(0, I_d)$. The message is that, under a suitable change of measure, we can "normalize" every Gaussian random vector to be a standard Gaussian vector.

This is usually how the Girsanov Theorem is formulated.

Theorem 8.3 (Girsanov) *Let $T \in [0, \infty]$. Let X be a progressively measurable process and define*

$$Z_t(X) = \exp\Big(\int_0^t X_s \, dB_s - \frac{1}{2} \int_0^t |X_s|^2 \, ds \Big), \quad 0 \le t < T.$$

Assume that $\big(Z_t(X)\big)_{0 \le t < T}$ is a martingale so that

$$\tilde{\mathsf{P}}(A) = \mathsf{E} \mathbb{1}_A Z_t(X), \quad \forall A \in \mathcal{F}_t \tag{8.5}$$

defines a new measure.

Then, $\tilde{B}_t = B_t - \int_0^t X_s \, ds$ is a Brownian motion on $[0, T)$ on $(\Omega, \mathcal{F}, \tilde{\mathsf{P}})$.

8.2 Exponential martingale and Radon–Nikodym derivatives

In this section we will discuss the change of measure formula (8.5). We will see that Z_t being a martingale is essential for (8.5) to define a measure, and that Z_t will play the role of the Radon–Nikodym derivative of the new measure with respect the original measure.

Proposition 8.4 *Let $T \in (0, \infty]$*

1. *Let P and $\tilde{\mathsf{P}}$ be two probability measures on a filtered probability space $(\Omega, \mathcal{F}_T, (\mathcal{F}_t)_{0 \le t \le T})$.Let P_t and $\tilde{\mathsf{P}}_t$ be the restriction of P and $\tilde{\mathsf{P}}$ on the smaller σ-field $\mathcal{F}_t \subset \mathcal{F}_T$. Suppose that $\tilde{\mathsf{P}} \ll \mathsf{P}$. Then $\tilde{\mathsf{P}}_t \ll \mathsf{P}_t$ for all t, and*

$$Z_t = \frac{d\tilde{\mathsf{P}}_t}{d\mathsf{P}_t} = \mathsf{E}\Big[\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}} \mid \mathcal{F}_t\Big], \quad 0 \le t \le T \tag{8.6}$$

   is a martingale.

2. *Let $(Z_t)_{0 \le t < T}$ be a P-martingale. Then*

$$\tilde{\mathsf{P}}(A) = \mathsf{E} \mathbb{1}_A(\omega) Z_t(\omega), \quad \forall A \in \mathcal{F}_t, \; 0 \le t < T, \tag{8.7}$$

   defines a probability measure $\tilde{\mathsf{P}}$.

   Moreover, if $(Z_t)_{0 \le t < T}$ is u.i., that is, $Z_T = \lim_{t \to T} Z_t$ exists in L^1 and a.s., then $\tilde{\mathsf{P}} \ll \mathsf{P}$ and $Z_T = \frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}$.

Proof:

1. Let $A \in \mathcal{F}_t$. Then

$$\mathsf{P}_t(A) = 0 \quad \Rightarrow \quad \mathsf{P}(A) = 0 \quad \Rightarrow \quad \tilde{\mathsf{P}}(A) = 0 \quad \Rightarrow \quad \tilde{\mathsf{P}}_t(A) = 0.$$

   Hence, $\tilde{\mathsf{P}} \ll \mathsf{P}$ implies that $\tilde{\mathsf{P}}_t \ll \mathsf{P}_t$.

To show that Z_t is a martingale, it suffices to show the second equality in (8.6). Let $A \in \mathcal{F}_t$. Then by the definition of Radon–Nikodym derivatives,

$$\tilde{\mathsf{P}}_t(A) = \mathsf{E}\mathbb{1}_A \frac{d\tilde{\mathsf{P}}_t}{d\mathsf{P}_t}, \quad \tilde{\mathsf{P}}(A) = \mathsf{E}\mathbb{1}_A \frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}.$$

Therefore,

$$\mathsf{E}\mathbb{1}_A \frac{d\tilde{\mathsf{P}}_t}{d\mathsf{P}_t} = \mathsf{E}\mathbb{1}_A \frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}$$

for all $A \in \mathcal{F}_t$. Hence the second equality in (8.6) holds by the definition of conditional expectation.

2. We need to check that (8.7) gives a consistent definition of a probability measure, since if $A \in \mathcal{F}_s \subset \mathcal{F}_t$, there are two definitions for $\tilde{\mathsf{P}}(A)$:

$$\tilde{\mathsf{P}}(A) = \mathsf{E}\mathbb{1}_A Z_t, \quad \tilde{\mathsf{P}}(A) = \mathsf{E}\mathbb{1}_A Z_s.$$

But $\mathsf{E}\mathbb{1}_A Z_t = \mathsf{E}\mathbb{1}_A Z_s$ just follows from Z_t being martingale.

Suppose now that Z_T exists. For any $A \in \mathcal{F}_t$, $\{\mathbb{1}_A Z_r,\ r \geq t\}$ is u.i. since Z_r are u.i. Then,

$$\tilde{\mathsf{P}}(A) = \lim_{r \to T} \mathsf{E}\mathbb{1}_A Z_r = \mathsf{E}\mathbb{1}_A Z_T.$$

Since $\tilde{\mathsf{P}}(A) = \mathsf{E}\mathbb{1}_A Z_T$ holds for any $A \in \mathcal{F}_t$, $t \geq 0$, it holds for any $A \in \mathcal{F}_T$. Therefore, $\tilde{\mathsf{P}} \ll \mathsf{P}$ and Z_T is the Radon–Nikodym derivative.

□

Remark 8.1 An analog in the case of product measures is the *Kakutani's dichotomy* (see also [Dur07, Example 4.3.7, Theorem 4.3.8]). Let $(\Omega, \mathcal{F}) = (\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$, and consider two product measures

$$\mathsf{P} = G_1 \otimes G_2 \otimes G_3 \otimes \cdots, \quad \tilde{\mathsf{P}} = F_1 \otimes F_2 \otimes F_3 \otimes \cdots$$

Assume that $P_n \ll G_n$, and $q_n = \dfrac{dF_n}{dG_n} > 0$, G_n-a.s. Then, $X_n = \dfrac{d\mathsf{P}_n}{d\tilde{\mathsf{P}}_n}$ is a $\mathcal{F}_n$-martingale. Note that by the nature of the product measure, X_n are independent random variables. Since

$$\{\lim_{n\to\infty} X_n = 0\} = \{\sum_n \log q_n > -\infty\}$$

belongs to the tail σ-algebra, the zero-one law guarantees that $X_n \to X$ P-a.s. for some X. We have either $\mathsf{P} \perp \tilde{\mathsf{P}}$ if $X = 0$, or $\tilde{\mathsf{P}} \ll \mathsf{P}$ if $X > 0$.

Interestingly, this is not too far from our Brownian motion case. Recall the Gaussian white noise construction of Brownian motion

$$B_t(\omega) = \sum_{n=1}^{\infty} \xi_n(\omega) \langle \mathbb{1}_{[0,t]}, e_n \rangle_{L^2},$$

where $\{e_n\}$ is an ONB and ξ_n are i.i.d. $\mathcal{N}(0,1)$ r.v.'s. So Brownian motion also have some product measure structure.

So far we have seen a martingale Z_t plays the role of a Radon–Nikodym derivative of two probability measures on a filtered probability space. The non-trial part of Theorem 8.3 is that Z_t has a special form of an *exponential martingale*, introduced in Proposition 7.5. Here, $dL_t = X_t dB_t$. In fact, for any positive c.l.m., we can express it as an exponential martingale.

Proposition 8.5 *If Z_t is a positive c.l.m., then $Z_t = \mathcal{E}_t(L)$ where*

$$L_t = \log Z_0 + \int_0^t Z_s^{-1}\, dZ_s. \tag{8.8}$$

Proof: Let L be defined by (8.8). Then

$$d \log Z_t = \frac{1}{Z_t}\, dZ_t - \frac{1}{2}\frac{1}{Z_t^2} d\langle Z \rangle_t = dL_t - \frac{1}{2}\langle L \rangle)_t.$$

$\square$

8.3 Girsanov Theorem: proof and applications

8.3.1 Girsanov Transform of c.l.m.'s

In this section, we assume Z_t is a martingale and $\tilde{\mathsf{P}}$ is defined by (8.7).

Lemma 8.6 *An adapted process X_t is a $\tilde{\mathsf{P}}$-martingale if and only if $X_t Z_t$ is a P-martingale.*

Proof: We have

$$\begin{aligned}
X_t \text{ is } \tilde{\mathsf{P}}\text{-martingale} \quad &\Leftrightarrow \quad \tilde{\mathsf{E}} X_t \mathbb{1}_A = \tilde{\mathsf{E}} X_s \mathbb{1}_A, \quad \forall A \in \mathcal{F}_s,\ s < t \\
&\Leftrightarrow \quad \mathsf{E} X_t Z_t \mathbb{1}_A = \mathsf{E} X_s Z_s \mathbb{1}_A, \quad \forall A \in \mathcal{F}_s,\ s < t \\
&\Leftrightarrow \quad X_t Z_t \text{ is } \mathsf{P}\text{-martingale.}
\end{aligned}$$

$\square$

The next proposition describes how continuous semi-martingales behaves under the Girsanov transform.

Proposition 8.7 *Assume that $Z_t = \mathcal{E}_t(L)$ is a martingale.*

1. *If M_t is a P-c.l.m., then $\tilde{M}_t = M_t - \langle M, L \rangle_t$ is a $\tilde{\mathsf{P}}$-c.l.m.*
2. *Let $\tilde{M} = M - \langle M, L \rangle$ and $\tilde{N} = N - \langle N, L \rangle$. Then $\langle \tilde{M}, \tilde{N} \rangle = \langle M, N \rangle$, computed under P or $\tilde{\mathsf{P}}$.*

Proof:

1. After localization, we can assume that Z_t, M_t are bounded martingales. By Lemma 8.6, it suffices to show $Z_t \tilde{M}_t = Z_t M_t - Z_t \langle M, L \rangle$ is a martingale. By Itô's formula, recalling that $dZ_t = Z_t dL_t$, we have

$$\begin{aligned}
d(Z_t M_t) &= Z_t\, dM_t + M_t\, dZ_t + d\langle M, Z \rangle_t \\
&= \text{m.t.} + Z_t\, \langle M, L \rangle_t,
\end{aligned}$$

and

$$d\big[Z_t \langle M, L \rangle_t\big] = Z_t\, d\langle M, L \rangle_t + \langle M, L \rangle_t\, dZ_t.$$

Taking the difference, we see that $dZ_t \tilde{M}_t$ only has martingales terms, and hence $Z_t \tilde{M}_t$ is a martingale.

2. Note that under P, $\tilde{M}$ and $\tilde{N}$ are continuous semi-martingale, and their martingale parts are given by M and N, so $\langle \tilde{M}, \tilde{N} \rangle = \langle M, N \rangle$ under P.

   Since $\tilde{\mathsf{P}} \ll \mathsf{P}$ and $\langle M, N \rangle$ is defined as limit in P-probability and P-a.s. has finite variation, $\langle M, N \rangle$ is also a finite variation process under $\tilde{\mathsf{P}}$. So M and N becomes continuous semi-martingales under $\tilde{\mathsf{P}}$, and their cross variation is defined.

   To show that $\langle \tilde{M}, \tilde{N} \rangle = \langle M, N \rangle$ under $\tilde{\mathsf{P}}$, we need to show that $\tilde{M}\tilde{N} - \langle M, N \rangle$ is a $\tilde{\mathsf{P}}$-c.l.m., i.e., by Lemma 8.6, $Z_t\big(\tilde{M}_t\tilde{N}_t - \langle M, N \rangle_t\big)$ is a P-c.l.m.

   We have

$$\begin{aligned} d(Z_t\tilde{M}_t\tilde{N}_t) &= \tilde{M}_t\tilde{N}_t\, dZ_t + \tilde{M}_t Z_t\big(dN_t - d\langle L, N \rangle_t\big) + \tilde{N}_t Z_t\big(dM_t - d\langle L, M \rangle_t\big) \\ &\quad + Z_t\, d\langle \tilde{M}, \tilde{N} \rangle_t + \tilde{M}_t\, d\langle \tilde{N}, Z \rangle_t + \tilde{N}_t\, d\langle \tilde{M}, Z \rangle_t \\ &= \text{m.t.} + Z_t\, d\langle M, N \rangle_t \end{aligned}$$

   (since $\langle \tilde{M}, \tilde{N} \rangle = \langle M, N \rangle$, $\tilde{M}_t\, d\langle \tilde{N}, Z \rangle_t = \tilde{M}_t\, d\langle N, Z \rangle_t = \tilde{M}_t Z_t\, d\langle N, L \rangle_t$ and likewise for the last term), and

$$d\big[Z_t\langle M, N \rangle_t\big] = Z_t\, d\langle M, N \rangle_t + \langle M, N \rangle_t\, dZ_t,$$

   so $Z_t\big(\tilde{M}_t\tilde{N}_t - \langle M, N \rangle_t\big)$ is indeed a P-c.l.m.

□

Now we are ready to give the proof of Theorem 8.3.

Proof of Theorem 8.3: Let $L = X \cdot B$. Then $Z = \mathcal{E}(L)$ and by Proposition 8.7, $\tilde{B} = B - \langle B, L \rangle$ is a $\tilde{\mathsf{P}}$-c.l.m. Moreover, $\langle \tilde{B}, \tilde{B} \rangle_t = \langle B, B \rangle_t = t$ under $\tilde{\mathsf{P}}$. By Lévy's characterization (Theorem 7.1), $\tilde{B}$ is a Brownian motion under $\tilde{\mathsf{P}}$. □

8.3.2 Application: Brownian motion with drift

Let $\mu \in \mathbb{R}$. By Theorem 8.3, $\tilde{B}_t = B_t - \mu t$ is a Brownian motion under

$$\mathsf{P}^\mu(A) = \mathsf{E}\mathbb{1}_A \exp\Big(\mu B_t - \frac{1}{2}\mu^2 t\Big), \quad A \in \mathcal{F}_t.$$

Let $T_b = \inf\{t : B_t = b\}$. Recall that we have computed $\mathsf{E}e^{-\lambda T_b} = e^{-|b|\sqrt{2a}}$.

We want to compute $\mathsf{P}^\mu(T_b < \infty)$. Note that under $\tilde{\mathsf{P}}$, $B_t = \tilde{B}_t + \mu t$ is a standard Brownian motion with a drift μt. We have

$$\begin{aligned} \mathsf{P}^\mu(T_b \le t) &= \mathsf{E}\mathbb{1}_{\{T_b \le t\}} Z_t \\ &= \mathsf{E}\mathbb{1}_{\{T_b \le t\}} \mathsf{E}[Z_t \mid \mathcal{F}_{t\wedge T_b}] && (\{T_b \le t\} \in \mathcal{F}_{t\wedge T_b}) \\ &= \mathsf{E}\mathbb{1}_{\{T_b \le t\}} Z_{t\wedge T_b} = \mathsf{E}\mathbb{1}_{\{T_b \le t\}} Z_{T_b} \\ &= \mathsf{E}\mathbb{1}_{\{T_b \le t\}} e^{\mu b - \frac{1}{2}\mu^2 b}. \end{aligned}$$

Letting $t \to \infty$, by Dominated Convergence Theorem and noting that $\mathsf{P}(T_b < \infty) = 1$, we have

$$\mathsf{P}^\mu(T_b < \infty) = \mathsf{E}e^{\mu b - \frac{1}{2}\mu^2 b} = e^{\mu b - |\mu b|} = \begin{cases} 1, & \mu b > 0, \\ e^{-2\mu b} < 1, & \mu b < 0. \end{cases}$$

The last result is very intuitive: for example, if $\mu > 0$, then the Brownian motion has a positive drift, which will dominate the typical behavior of $B_t \sim \sqrt{t}$, so it is less likely to hit a negative number $b < 0$.

Proposition 8.8 (Wald's identity) *Let $\mu \in \mathbb{R}$ and T be a P-a.s. finite stopping time. Then*

$$\mathsf{E}e^{\mu B_T - \frac{1}{2}\mu^2 T} = 1 \quad \Leftrightarrow \quad \mathsf{P}^\mu(T < \infty) = 1.$$

Proof: The same argument with T_b replaced by T gives

$$\mathsf{P}^\mu(T \le t) = \mathsf{E}\mathbb{1}_{\{T \le t\}} Z_T = \mathsf{E}\mathbb{1}_{\{T \le t\}} e^{\mu B_T - \frac{1}{2}\mu^2 T}.$$

Letting $t \uparrow \infty$, the desired result follows from Monotone Convergence Theorem. □

8.3.3 Application: Cameron–Martin space

Let $h \in \mathcal{C}[0,\infty)$ with $h(0) = 0$. Consider $\tilde{B}_t = B_t - h_t$. Then $\tilde{B}$ is a Brownian motion on $[0,\infty)$ under $\tilde{\mathsf{P}}$ if and only if $h \in H_0^1[0,\infty)$, where

$$H_0^1[0,\infty) = \{h : \text{weak derivative} \partial_x h \in L^2[0,\infty) \text{ and } h(0) = 0\}.$$

And when $h \in H_0^1[0,\infty)$,

$$\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}} = e^{\int_0^\infty |\dot{h}(s)|\, dB_s - \frac{1}{2}\int_0^\infty |\dot{h}(s)|^2\, ds}.$$

Note that the space $H_0^1[0,\infty)$ is much smaller than $\mathcal{C}[0,\infty)$. It means that although the Wiener measure is defined on $\mathcal{C}[0,\infty)$, not all translations in $\mathcal{C}[0,\infty)$ will generate a new measure that is absolutely continuous with respect to the original measure. Such subspace consisting all such translation for a Gaussian measure is called the *Cameron–Martin* space.

For infinite dimensional Gaussian measures, the Cameron–Martin space is strictly smaller due to the fact there are "too many" directions. For a finite dimensional Gaussian measure, i.e., a Gaussian vector $X \in \mathbb{R}^d \sim \mathcal{N}(\mu, Q)$, the Cameron–Martin space is $\mathbb{R}^d$, unless Q is degenerate, in which case the Cameron–Martin space is the range of Q, or the domain of Q^{-1}. This is also easy to see since otherwise, one cannot write down the Radon–Nikodym derivative (8.2).

8.4 Novikov condition

A key assumption in the Girsanov Theorem Theorem 8.3 is that Z_t is a martingale. In this section we introduce a sufficient condition.

Proposition 8.9 *If Z_t is a positive c.l.m., then Z_t is a super-martingale.*

Proof: Let $T_n \uparrow \infty$ be such that Z^{T_n} is a martingale. Since Z is positive, by Fatou's Lemma, we have for $s < t$,

$$Z_s = \lim_{n\to\infty} Z_s^{T_n} = \lim_{n\to\infty} \mathsf{E}\big[Z_t^{T_n} \mid \mathcal{F}_s\big] \ge \mathsf{E}\Big[\lim_{n\to\infty} Z_t^{T_n} \mid \mathcal{F}_s\Big] = \mathsf{E}\big[Z_t \mid \mathcal{F}_s\big].$$

□

Lemma 8.10 *The process Z_t is a martingale if and only if $\mathsf{E}Z_t = 1$.*

Proof: By Proposition 8.9 Z_t is a super-martingale. By Doob–Meyer Decomposition, $Z_t = M_t - A_t$, where A_t is an increasing process and M_t is a martingale. Then $\mathsf{E}Z_t = 1 \Leftrightarrow A_t \equiv 0$. This proves the lemma. □

Theorem 8.11 (Novikov's condition) *Let L be a c.l.m. and $Z_t = \mathcal{E}_t(L)$. If*

$$\mathsf{E}e^{\frac{1}{2}\langle L\rangle_\infty} < \infty, \tag{8.9}$$

then Z_t is uniformly integrable and hence a martingale.

Proof: We will represent L as the time-change of a standard Brownian motion (Theorem 7.7). Note that (8.9) implies that P-a.s. $\langle L\rangle_\infty < \infty$, so we need to adapt Theorem 7.7 to such case. The Brownian motion will be defined by

$$\beta_s = \begin{cases} L_{\tau_s}, & s < \langle L\rangle_\infty,\ \tau_s = \inf\{t \geq 0 : \langle L\rangle_t \geq s\}, \\ L_\infty + \left(\tilde{\beta}_s - \tilde{\beta}_{\langle L\rangle_\infty}\right), & s \geq \langle L\rangle_\infty, \end{cases}$$

where $\tilde{\beta}$ is a Brownian motion independent of L. Then $T = \langle L\rangle_\infty$ is a stopping time with respect to the filtration

$$\mathcal{G}_s = \mathcal{F}_{\tau_s},$$

since $\{T \geq s\} = \{\tau_s < \infty\} \in \mathcal{F}_{\tau_s}$.

The condition (8.9) can be rewritten as

$$\mathsf{E}e^{\frac{1}{2}T} < \infty. \tag{8.10}$$

In light of Lemma 8.10, it suffices to show that

$$1 = \mathsf{E}e^{B_T - \frac{1}{2}T}. \tag{8.11}$$

Clearly, by Proposition 8.8, (8.11) holds if and only if $\mathsf{P}^{\mu=1}(T < \infty) = 1$. But we know nothing about T except (8.10), Proposition 8.8 is directly useful. The point here is to derive (8.11) from (8.10).

Let $S_n \uparrow \infty$ be a sequence of stopping times such that

$$1 = \mathsf{E}e^{B_{S_n} - \frac{1}{2}S_n} \tag{8.12}$$

By Optional Sampling Theorem applied to the uniformly integrable martingale $e^{B_{S_n\wedge t} - \frac{1}{2}(S_n\wedge t)}$, we have

$$1 = \mathsf{E}e^{B_{S_n\wedge T} - \frac{1}{2}\left(S_n\wedge T\right)} = \mathsf{E}\mathbb{1}_{\{S_n<T\}}e^{B_{S_n} - \frac{1}{2}S_n} + \mathsf{E}\mathbb{1}_{\{S_n\geq T\}}e^{B_T - \frac{1}{2}T}.$$

Letting $S_n \uparrow \infty$, the second term converges to $\mathsf{E}e^{B_T - \frac{1}{2}T}$ by Monotone Convergence Theorem. Hence, to establish (8.11), it remains to show

$$\lim_{n\to\infty} \mathsf{E}\mathbb{1}_{\{S_n<T\}}e^{B_{S_n} - \frac{1}{2}S_n} = 0 \tag{8.13}$$

Now we pick

$$S_n = \inf\{t : B_t = t - n\}.$$

Then (8.13) is bounded by

$$\liminf_{n\to\infty} \mathsf{E}\mathbb{1}_{\{S_n<T\}}e^{B_{S_n} - S_n} \cdot e^{\frac{1}{2}S_n} \leq \liminf_{n\to\infty} e^{-n}\mathsf{E}\mathbb{1}_{\{S_n<T\}}e^{\frac{1}{2}T}.$$

The last limit is 0 by (8.11) and Monotone Convergence Theorem. Also

$$\mathsf{P}^{\mu=1}(S_n < \infty) = \mathsf{P}^{\mu}(\tilde{B}_t \text{ hits } -n \text{ before } \infty) = 1,$$

so (8.12) holds by Proposition 8.8. This completes the proof. □

Remark 8.2 If only for some $\varepsilon > 0$ small, $\mathsf{E}e^{(\frac{1}{2}-\varepsilon)T} < \infty$, then for the argument above we need to consider instead

$$S_n^\varepsilon = \inf\{t : B_t = (1-\varepsilon)t - n\}.$$

However, (8.12) no longer holds since $\tilde{B}_t = B_t - t$ will not hit $-\varepsilon t - n$ a.s. This explains why $\dfrac{1}{2}$ is "sharp".

9 Stochastic differential equations

9.1 Markov semi-groups and diffusions

A Markov process is an adapted process that satisfies the Markov property, see Definition 3.4. To develop the semi-group theory, we take the state space of the Markov process to be a metric space E. Usually $E = \mathbb{Z}, \mathbb{R}^d$ or subsets of them. A *Markov kernel* is a family of probability measures $p_t(x, \cdot) \in \mathcal{M}(E)$, $t \geq 0, x \in E$ that satisfies the following two conditions

- $p_t(x, \cdot) \Longrightarrow \delta_x$, $t \downarrow 0$.
- (Kolmogorov–Chapman) For $t, s \geq 0$ and any $x \in E$,

$$p_{t+s}(x, \cdot) = \int p_t(x, dy) p_s(y, \cdot). \tag{9.1}$$

Any Markov kernel defines a Markov process. Let $\mu \in \mathcal{M}(E)$ be the initial condition. Then the f.d.d. of the Markov process $(X_t)_{t\geq 0}$ is given by

$$\begin{aligned} &\mathsf{P}^\mu(X_{t_0} \in A_0, X_{t_1} \in A_1, X_{t_2} \in A_2, \cdots, X_{t_{n-1}} \in A_{n-1}, X_{t_n} \in A_n) = \int \mu(dx_0) \\ &\cdot \int p_{t_1}(x_0, dx_1) \int p_{t_2-t_1}(x_1, dx_2) \int \cdots \int p_{t_{n-1}-t_{n-2}}(x_{n-2}, dx_{n-1}) p_{t_n-t_{n-1}}(x_{n-1}, A_n). \end{aligned} \tag{9.2}$$

The f.d.d. (9.2) is consistent thanks to (9.1). Then Kolmogorov Extension Theorem guarantees the existence of a stochastic process with (9.2) as its f.d.d. In case that X_t starts from a delta measure $\mu = \delta_x$, it is conventional to write P^x instead of P^{δ_x}.

Formally, given a Markov kernel, the integral operator

$$\mathsf{P}_t f := \int p_t(x, dy) f(y) \tag{9.3}$$

defines a semi-group, since (9.1) implies the semi-group relation

$$\mathsf{P}_t \mathsf{P}_s = \mathsf{P}_{t+s}, \quad t, s \geq 0. \tag{9.4}$$

Most often we impose some regularity assumptions on the Markov kernel so that (9.3) defines an operator on the functional space of continuous functions. To be more precise, let

$$\mathcal{C}_0(E) = \{f \in \mathcal{C}(E) : \lim_{|x|\to\infty} |f(x)| = 0\}.$$

Definition 9.1 *The equation* (9.3) *defines a* Feller semi-group $(\mathsf{P}_t)_{t\geq 0}$ *if*

- *for all* $t \geq 0$, $f \in \mathcal{C}_0(E) \Rightarrow \mathsf{P}_t f \in \mathcal{C}_0(E)$;
- *for all* $f \in \mathcal{C}_0(E)$, $t \mapsto \mathsf{P}_t f$ *is continuous in the topology of* $\mathcal{C}_0(E)$.

It is natural to discuss differentiability once continuity is known. Consider the *operator*

$$\mathcal{L} f := \lim_{t\downarrow 0} \frac{\mathsf{P}_t f - \mathsf{P}_0 f}{t} = \lim_{t\downarrow 0} \frac{\mathsf{P}_t f - f}{t}, \quad \text{in } \mathcal{C}_0(E). \tag{9.5}$$

The operator $\mathcal{L}$ is called the *generator* of the semi-group $(\mathsf{P}_t)_{t\geq 0}$. The limit may not exist for any $f \in \mathcal{C}_0(E)$; the domain of $\mathcal{L}$, denoted by $\mathcal{D}(\mathcal{L})$, consists of all the functions in $\mathcal{C}_0(E)$ such that (9.5) exists. A nice account of the theory can be found in [EK86]. We will not dive deep into the theory of semi-group and generators, but just assume some facts that we will utilize frequently.

Remark 9.1

- $\mathcal{D}(\mathcal{L})$ is dense in $\mathcal{C}_0(E)$.
- When the Markov process is a diffusion, $\mathcal{L}$ is a second-order differential operator and
$$\mathcal{D}(\mathcal{L}) \supset \mathcal{C}_0^2 = \{f \in \mathcal{C}^2(E) : |f(x)|, |(\partial_i f)(x)|, |(\partial_{ij} f)(x)| \to 0, |x| \to \infty\}.$$
- $\mathcal{L}$ and $(\mathsf{P}_t)_{t\geq 0}$ determine each other.

We will elaborate the last point. The key observation is from the semi-group property (9.4):

$$\mathsf{P}_s\mathcal{L}f = \lim_{t\downarrow 0} \mathsf{P}_s\Big(\frac{\mathsf{P}_t f - f}{t}\Big) = \lim_{t\downarrow 0} \frac{\mathsf{P}_{s+t} f - \mathsf{P}_s f}{t} \quad \Rightarrow \quad \frac{d\mathsf{P}_t}{dt} = \mathsf{P}_t\mathcal{L}.$$

The ODE $x'(t) = \lambda x$ has a unique solution $x = x_0 e^{\lambda t}$, so naturally we expect $\mathsf{P}_t = e^{\lambda\mathcal{L}}$ in some sense.

When $\mathcal{L}$ is bounded. This occurs when $(X_t)_{t\geq 0}$ is a finite-state continuous time Markov chain and $\mathcal{L}$ becomes a $N \times N$ matrix which is always bounded. In this case, the matrix exponential can be defined by the Taylor expansion

$$\mathsf{P}_t := e^{t\mathcal{L}} = \sum_{n=0}^{\infty} \frac{t^n \mathcal{L}^n}{n!},$$

and the infinite sum converges in the matrix norm. Clearly, term-by-term differentiation makes sense, and we have

$$\frac{d}{dt}(\mathsf{P}_t f) = \sum_{n=0}^{\infty} \frac{d}{dt}(t^n)\frac{\mathcal{L}^n f}{n!} = \sum_{n=1}^{\infty} \frac{t^{n-1}\mathcal{L}^n f}{(n-1)!}.$$

Hence

$$\frac{d}{dt}(\mathsf{P}_t f) = \mathcal{L}(\mathsf{P}_t f) = \mathsf{P}_t(\mathcal{L}f). \tag{9.6}$$

When $\mathcal{L}$ is unbounded. The relation (9.6) still holds for non-bounded $\mathcal{L}$, even though the exponential via the infinite sum is no longer available.

Proposition 9.1 *If $f \in \mathcal{D}(\mathcal{L})$, then $\mathsf{P}_t f \in \mathcal{D}(L)$ and (9.6) holds, as well as the integral form*

$$\mathsf{P}_t f - f = \int_0^t \mathcal{L}(\mathsf{P}_s f)\, ds = \int_0^t \mathsf{P}_s(\mathcal{L}f)\, ds. \tag{9.7}$$

Remark 9.2 Note: $\mathcal{L}f \in \mathcal{C}_0(E)$ and $s \mapsto \mathsf{P}_s(\mathcal{L}f)$ is continuous in $\mathcal{C}_0(E)$, so the last integral could be defined as a Riemann integral for continuous functions.

Example 9.3 Let $(B_t)_{t\geq 0}$ be the d-dimensional Brownian motion. Then $p_t(x, dy) = (2\pi t)^{-d/2} e^{-\frac{|x-y|^2}{2t}}\, dy$. Using elementary calculus, one can check directly p_t defines a Feller semi-group.

Next we will show that $\mathcal{L}^B = \Delta$. Indeed, let $f \in \mathcal{C}_0^2(\mathbb{R}^d)$, then by Itô's formula,

$$f(B_t) - f(B_0) = \int_0^t \sum_{i=1}^d (\partial_i f)(B_s) dB_s^{(i)} + \frac{1}{2}\int_0^t \sum_{i,j=1}^d (\partial_{ij} f)(B_s) d\langle B^{(i)}, B^{(j)}\rangle_s.$$

Taking expectation, and noting that $|\partial_i f|$ is bounded so that the first term is a true martingale, and $\langle B^{(i)}, B^{(j)}\rangle = \delta_{ij} t$, we have

$$\mathsf{E}^x f(B_t) - f(x) = \int_0^t \mathsf{E}^x(\frac{1}{2}\Delta f)(B_s)\, ds \quad \Rightarrow \quad (\mathsf{P}_t f)(x) - f(x) = \int_0^t \mathsf{P}_s\big(\frac{1}{2}\Delta f\big)(x)\, ds.$$

Hence $f \in \mathcal{D}(\mathcal{L}^B)$ and by (9.7), $\mathcal{L}^B = \frac{1}{2}\Delta$.

9.2 Diffusion and forward/backward Kolmogorov equations

Let us look at a generalization of the calculation at the end of last section. In one dimension, let b, σ be bounded measurable process $\mathbb{R} \to \mathbb{R}$ and $(B_t)_{t\geq 0}$ a standard Brownian motion. Suppose X_t is a process that solves

$$X_t = x + \int_0^t \sigma(X_t)\, dB_t + \int_0^t b(X_t)\, dt. \tag{9.8}$$

Let $f \in \mathcal{C}_0^2(\mathbb{R})$. By Itô's formula, we have

$$f(X_t) = x + \int_0^t f'(X_s)(\sigma(X_s)\, dB_s + b(X_s)\, ds) + \int_0^t \frac{1}{2} f''(X_s)\sigma^2(X_s)\, ds.$$

After taking expectation, the martingale term disappears, and we have

$$\mathsf{E} f(X_t) = f(x) + \int_0^t \mathsf{E}(\mathcal{L}f)(X_s)\, ds, \tag{9.9}$$

where

$$(\mathcal{L}f) = (\mathcal{L}^{b,\sigma} f)(x) = \frac{1}{2}\sigma^2(x)(\partial_{xx} f)(x) + b(x)(\partial_x f)(x). \tag{9.10}$$

If we define $u(t,x) = \mathsf{E} f(X_t)$ (noting that X_t starts from x), then $u(t,x)$ satisfies the *forward Kolmogorov equation*

$$\begin{cases} \partial_t u = \mathcal{L}u, \\ u(0,x) = f(x). \end{cases}$$

Here, $\mathcal{L}$ is a differential operator acting on the x variable, so $(\mathcal{L}u)(t,x) = (\mathcal{L}u(t,\cdot))(x)$.

Now we assume that the distribution of X_t is absolutely continuous with respect to the Lebesgue measure, and denote by $\rho(t,y)$ its density. Then

$$\mathsf{E} f(X_t) = \int \rho(t,y) f(y)\, dy,$$

and

$$\int_0^t \mathsf{E}(\mathcal{L}f)(X_s)\, ds = \int_0^t \int \rho(s,y)(\mathcal{L}f)(y)\, dy = \int_0^t \int (\mathcal{L}^*\rho(s,\cdot)) f(y)\, dy = \int f(y)\, dy \int_0^t (\mathcal{L}^*\rho(s,\cdot))\, ds.$$

Combining these with (9.9), we have for all $f \in \mathcal{C}_0^2(\mathbb{R})$,

$$\int f(y)\, dy \cdot \Big[\rho(t,y) - \int_0^t (\mathcal{L}^*\rho(s,\cdot))\, ds\Big] = f(x).$$

As $\mathcal{C}_0^2(\mathbb{R})$ is dense in $\mathcal{C}(\mathbb{R})$ and hence determines an element in $\mathcal{M}(\mathbb{R})$, the density $\rho(t,y)$ satisfies the *backward Kolmogorov* equation

$$\begin{cases} \partial_t \rho = \mathcal{L}^*\rho, \\ \rho(t,\cdot) \Rightarrow \delta_x, \quad t \downarrow 0. \end{cases}$$

Here, the adjoint operator $\mathcal{L}^*$ is the differential operator such that for all $g, h \in \mathcal{C}_0^\infty(\mathbb{R})$,

$$\int g(x)(\mathcal{L}h)(x)\, dx = \int (\mathcal{L}^* g)(x) h(x)\, dx.$$

When $\mathcal{L}$ is given by (9.10), we have

$$\mathcal{L}^* g = \partial_{xx}\big(\frac{1}{2}\sigma^2(x) g(x)\big) - \partial_x\big(b(x) g(x)\big)$$

from integration by parts.

Proposition 9.2 *If for every $\mu \in \mathcal{M}(\mathbb{R})$, there exists a unique solution $\rho = \rho(\cdot;\mu)$ to the backward Kolmogorov equation, that is, $t \mapsto \rho(t,\cdot)$ is continuous in $\mathcal{M}(E)$ in the weak topology and*

$$\begin{cases} \quad \partial_t \rho = \mathcal{L}^* \rho, \quad t > 0, \\ \rho(t,\cdot) \Rightarrow \mu, \end{cases} \tag{9.11}$$

then $p_t(x,dy) = p(t,dy;\delta_x)$ defines a Markov kernel.

Proof: By definition of a solution, $p_t(x,\cdot) \Rightarrow \delta_x$. It remains to check (9.1).

For (9.1) let us fix $t \geq 0$ and show its validity for all $s \geq 0$. Note that (9.11) is a linear equation, so that

$$\tilde{\rho}(s,\cdot) = \int p_t(x,dz)p_s(z,\cdot)$$

solves

$$\begin{cases} \quad \partial_s \tilde{\rho} = \mathcal{L}^* \tilde{\rho}, \\ \tilde{\rho}(s,\cdot) \Rightarrow p_t(x,\cdot). \end{cases}$$

But $p_{t+s}(x,\cdot)$ solves the same equation since the the evolution PDE (9.11) is well-posed. Hence $p_{t+s} = \tilde{\rho}(s,\cdot)$ and (9.1) is satisfied. □

Now we are ready to define what is a diffusion.

Definition 9.2 *A diffusion $X_t \in \mathbb{R}^d$ is a Markov process such that*

- *The sample path $t \mapsto X_t$ is continuous.*
- *The generator of X_t is*

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

where (a_{ij}) is positive definite.

By Proposition 9.2, such a Markov process exists if (9.11) is well-deposed. The continuity of the path can also be derived from the information on f.d.d. if certain additional assumptions are imposed on $\mathcal{L}$. Let us compute two important quantities for a diffusion. The first is the drift (in the i-th coordinate):

$$\mathsf{E}^x[X_t^{(i)} - x_i] = t \cdot (\mathcal{L}f)(x) + o(t) = t \cdot b_i(x) + o(t), \tag{9.12}$$

where $f(y) = y^i$, the i-th coordinate of the argument y, and we just use $\mathsf{P}_t f - f = \mathcal{L}f \cdot t + o(t)$ from the definition of the generator. Strictly speaking, $f(y) = y^i$ is not in $\mathcal{C}_0(\mathbb{R}^d)$, but we can approximate f by some functions in $\mathcal{C}_0(\mathbb{R}^d)$ and our conclusion will not be affected.

In light of (9.12), it is tempting to guess that $(X_t^{(i)} - x_i)/t = b_i(x)$ a.s. However, this is not true, as the next quantity shows. For $1 \leq i,j \leq d$, let $f(y) = (y^i - x_i)(y^j - x_j)$. Then we have

$$\mathsf{E}^x(X^{(i)} - x_i)(X^{(j)} - x_j) = t \cdot (\mathcal{L}f)(x) + o(t) = \frac{a_{ij}(x) + a_{ji}(x)}{2} \cdot t + o(t). \tag{9.13}$$

First, (9.13) implies that X_t cannot be a.s. differentiable, otherwise the right-hand side should be $O(t^2)$ instead of $O(t)$. Second, one can use (9.13), derived soly from the f.d.d. information, together with the Kolmogorov Continuity Test to say something about the continuity of the process.

In this way, using some knowledge from parabolic PDEs (Proposition 9.2), we know pretty well what is a diffusion process. But it is still meaningful to construct the diffusion as some "stochastic differential equation", like (9.8) This used to be the main motivation behind the stochastic calculus. We have derived the form of $\mathcal{L}$ from (9.8); but we have not yet define what is a solution to (9.8), let alone how to solve it. That will be discussed in the next few sections.

Finally, let us derive the form of $\mathcal{L}$ and $\mathcal{L}^*$ without using Itô's formula. Imagine we are in one dimension and there are some particles on moving on the real line doing "diffusion". We can observe $\rho(t,x)$, the density of the particles at time t, at location x. Fixing a small region $[x, x+\Delta x]$, the first identity we can write down is the conservation of particles, namely,

$$\frac{\partial}{\partial t}\rho(t,x)\Delta x = J(x+\Delta x) - J(x),$$

where $J(y)$ is the flow of particles crossing $\{x = y\}$. The form of J will come from some physical laws. It is reasonable to write

$$J(x) = a(x)\partial_x \rho + b(x)\rho(x).$$

The first term states that the flow should be proportional to the difference of density to the left and to the right of $\{y = x\}$; this is the *principle of diffusion*. We add a term $a(x)$ as a factor. The second term gives some external factor which forces particle at x to move in speed $b(x)$. Using this form of J, one sees that ρ indeed solves a second-order parabolic equation.

9.3 Strong and weak solutions, notion of uniqueness

We start with a general *stochastic differential equation*:

$$\begin{cases} dX_t = b(t,X_t)\,dt + \sigma(t,X_t)\,dB_t \\ X_0 = \xi. \end{cases} \tag{9.14}$$

Here, $X_t \in \mathbb{R}^d$ and B_t is r-dimensional Brownian motion; accordingly, $b(t,X_t) \in \mathbb{R}^{d\times 1}$ and $\sigma(t,X_t) \in \mathbb{R}^{d\times r}$; ξ is a random vector in $\mathbb{R}^d$ with given distribution μ. More explicitly, we could also write down the equation coordinate-wise:

$$dX_t^{(i)} = b_i(t,X_t)\,dt + \sum_{j=1}^{r} \sigma_{ij}(t,X_t)\,dB_t^{(j)}, \quad 1 \le i \le n.$$

The processes X_t, B_t will be adapted on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathsf{P})$ that satisfies the usual condition.

A *strong solution* X_t is a functional of the Brownian motion B_t. We will encode such dependence via adaptedness to the Brownian filtration. More precisely, suppose that on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ there lives a r-dimensional Brownian motion and an independent r.v. ξ with distribution μ. Let $\mathcal{G}_t = \sigma(\xi) \vee \mathcal{F}_t^B$ and $\mathcal{N}$ be the collection of all P-null sets of $(\Omega, \mathcal{G}_\infty, \mathsf{P})$. We define the augmented filtration $\mathcal{F}_t = \sigma(\mathcal{G}_t \cup \mathcal{N})$. This filtration satisfies the usual condition, and it just contains the information of the Brownian motion and the initial condition.

Definition 9.3 *The equation* (9.14) *has a strong solution if there is a process* X_t *satisfying* (9.14) *and is adapted to* $\mathcal{F}_t$ *defined as above.*

We say that strong uniqueness *holds if* X_t' *is another strong solution, then* $\mathsf{P}(X_t = X_t', \forall t \ge 0) = 1$.

Remark 9.4 As a solution, we implicitly assume that the stochastic integral could be defined in the broadest sense given in Section 6.4, e.g., $\int_0^T \sigma^2(t,X_t)\,dt < \infty$ a.s. if we want to consider the solution up to some time $T \in [0,\infty]$.

A weak solution relaxes the condition that X_t is adapted to the filtration generated by the Brownian motion.

Definition 9.4 *The equation* (9.14) *has a weak solution* (X_t, B_t) *if* (X_t, B_t) *are a pair of adapted process on some filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}\mathsf{P})$ *such that* (9.14) *holds, and* B_t *is a Brownian motion under* P.

We will introduce other notions of uniqueness and discuss the distinction between strong and weak solution through the celebrated example by Tanaka.

Consider the SDE

$$dX_t = \operatorname{sgn}(X_t)\, dB_t, \quad X_0 = 0 \tag{9.15}$$

First let us construct a weak solution to (9.15). Let X_t be a Brownian motion on $(\Omega, \mathcal{F}, \mathsf{P})$. Define

$$B_t = \int_0^t \operatorname{sgn}(X_t)\, dX_t.$$

Note that $|\operatorname{sgn}(X_t)| \leq 1$, so the stochastic integral is well-defined and B_t is a martingale. Moreover, $\langle B \rangle_t = |\operatorname{sgn}(X_t)|^2\, d\langle X \rangle_t = dt$, so by Theorem 7.1, B_t is a Brownian motion. Hence (B_t, X_t) is a weak solution to (9.15).

Definition 9.5 *We say that* (9.14) *has* weak uniqueness in law *if for given initial condition* μ, *for any weak solution* (X_t, B_t), *the law* $\mathsf{P}(X_t \in \cdot)$ *as a probability measure on* $\mathcal{C}([0, \infty), \mathbb{R}^d)$ *is unique.*

By Theorem 7.1 again, if (X_t, B_t) is a weak solution to (9.15), then $(X_t)_{t\geq 0}$ is a Brownian motion. Hence the weak uniqueness in law holds for (9.15).

Another notion for uniqueness of weak solutions is the pathwise uniqueness.

Definition 9.6 *We say that* pathwise uniqueness *holds for* (9.14), *if for* (X_t, B_t), (X'_t, B_t) *two weak solutions defined on the same probability space* $(\Omega, \mathcal{F}, \mathsf{P})$, X_t *and* X'_t *are indistinguishable, i.e.,* $\mathsf{P}(X_t = X'_t, \forall t \in [0, \infty)) = 1$.

From the (9.15), if (X_t, B_t) is a weak solution, then $(-X_t, B_t)$ is also a weak solution since $\operatorname{sgn}(x) = -\operatorname{sgn}(-x)$. But it is impossible to have $\mathsf{P}(X_t = -X_t, \forall t \geq 0) = 1$ since $(X_t)_{t\geq 0}$ is known to be the Brownian motion. Hence, pathwise uniqueness fails for (9.15).

Finally, let us also show that strong existence fails for (9.15). Suppose that a Brownian motion (B_t) is given, and X_t is a strong solution. Then

$$dB_t = \operatorname{sgn}(X_t) \cdot \operatorname{sgn}(X_t)\, dB_t = \operatorname{sgn}(X_t) dX_t,$$

and hence

$$B_t = \int_0^t \operatorname{sgn}(X_t) dX_t = |X_t| - L_t^X(0),$$

where $L_t^X(0)$ is the local time of the Brownian motion X at 0. There is various way to define the local time, an increasing process, through limit of certain expression of X, for example,

$$L_t^X = \lim_{\varepsilon \downarrow 0} \int_0^t \frac{\varepsilon}{(\sqrt{|X_s|^2 + \varepsilon})^3}\, ds,$$

which comes from approximating $f(x) = |x|$ by $f_\varepsilon(x) = \sqrt{x^2 + \varepsilon}$. One sees that L_t^X could be defined through $|X|$ (which is intuitive since $X_t = 0$ is the same as $|X_t| = 0$). Therefore, $\mathcal{F}_t^B \subset \mathcal{F}_t^{|X|}$. On the other hand $\mathcal{F}_t^X \subset \mathcal{F}_t^B$ by the definition of strong solution. But then we arrive at the inclusion $\mathcal{F}_t^X \subset \mathcal{F}_t^{|X|}$, which cannot be true since X is a Brownian motion. This shows that (9.15) cannot admit a strong solution. As we will see in (9.15), this is due to the discontinuity of the sign function.

As far as the construction of a diffusion as a solution of SDE, what we need is weak existence and uniqueness. On the other hand, it is usually very easy to work on pathwise uniqueness since it starts on a coupling. The surprising result by Yamada–Watanabe states that pathwise uniqueness plus weak uniqueness lead to strong solvability of (9.14). We will develop the result in Section 10.1.

Theorem 9.3 (Yamada–Watanabe) *If there exists a weak solution to* (9.14) *and pathwise uniqueness holds, then weak uniqueness also holds.*

Moreover, the above assumptions lead to the existence of strong solutions to (9.14).

9.4 Lipschitz case and generalizations

In this section we will prove strong existence and uniqueness of solution to (9.14) under the assumption that the coefficients b, σ are Lipschitz in x: $\exists K > 0$ such that

$$|\sigma(t,x) - \sigma(t,y)| + |b(t,x) - b(t,y)| \le K|x-y| \tag{9.16}$$

$$|\sigma(t,x)|^2 + |b(t,x)|^2 \le K(1+|x|^2). \tag{9.17}$$

Analogous to the solution theory of ordinary differential equation, the solutions are constructed via a Picard iteration scheme. For simplicity the results we state are in one dimension. But the extension to higher dimensions is immediate.

We first state the uniqueness.

Theorem 9.4 *Suppose the coefficients b, σ are locally Lipschitz, i.e., for $n \ge 1$ there exists $K_n > 0$ such that*

$$|\sigma(t,x) - \sigma(t,y)| + |b(t,x) - b(t,y)| \le K|x-y|, \quad \forall |x|, |y| \le n. \tag{9.18}$$

Then pathwise uniqueness holds for (9.14).

We will use the following version of Gronwall's inequality.

Lemma 9.5 *Let $g \ge 0$ be a bounded, measurable function. Let $a, b \ge 0$. If*

$$g(t) \le a + b\int_0^t g(s)\,ds$$

for all $t \ge 0$, then $g(t) \le ae^{bt}$.

Proof: Let (X_t, B_t) and $(\tilde{X}_t, B_t)$ be two weak solutions to (9.14). Let

$$\tau_n = \inf\{t \ge 0 : |X_t| \wedge |\tilde{X}_t| \ge n\}.$$

Then τ_n, $n \ge 1$ are stopping times and $\tau_n \uparrow \infty$ a.s.

Fix $T > 0$. For every $0 \le t \le T$,

$$\tilde{X}_{t\wedge\tau_n} - X_{t\wedge\tau_n} = \int_0^{t\wedge\tau_n} \left[b(s,\tilde{X}_s) - b(s,X_s)\right] ds + \int_0^{t\wedge\tau_n} \left[\sigma(s,\tilde{X}_s) - \sigma(s,X_s)\right] dB_s.$$

Squaring both sides and taking expectation, we have

$$\begin{aligned}
\mathsf{E}|X_{t\wedge\tau_n} - \tilde X_{t\wedge\tau_n}|^2 &\le 2\mathsf{E}\Big[\int_0^{t\wedge\tau_n} \big[b(s,\tilde X_s) - b(s,X_s)\big]\,ds\Big]^2 + 2\mathsf{E}\int_0^{t\wedge\tau_n} \big[\sigma(s,\tilde X_s) - \sigma(s,X_s)\big]^2\,ds \\
&\le 2K_n^2 T\mathsf{E}\int_0^{t\wedge\tau_n} |\tilde X_s - X_s|^2\,ds + 2K_n^2\mathsf{E}\int_0^{t\wedge\tau_n} |\tilde X_s - X_s|^2\,ds \\
&\le 2K_n^2(T+1)\int_0^t \mathsf{E}|\tilde X_{s\wedge\tau_n} - X_{s\wedge\tau_n}|^2\,ds.
\end{aligned}$$

Here, the first line is due to $(a+b)^2 \le 2a^2 + 2b^2$, the second line is by (9.18) and Cauchy–Schwartz on the first term.

Now let $g(t) = \mathsf{E}|\tilde X_{t\wedge\tau_n} - X_{t\wedge\tau_n}|^2$, $a = 0$ and $b = 2K_n^2(T+1)$. Then by Lemma 9.5, $g(t) \equiv 0$. Since $\tilde X_{t\wedge\tau_n}$ and $X_{t\wedge\tau_n}$ are continuous processes, $X_{t\wedge\tau_n} = \tilde X_{t\wedge\tau_n}$ for all $t \le T$. The desired result follows from letting $T \uparrow \infty$ and $\tau_n \uparrow \infty$. □

For the existence of strong solutions, we will consider the following Picard iteration scheme:

$$\begin{aligned}
X_t^{(0)} &= \xi, \\
X_t^{(k+1)} &= \xi + \int_0^t b(s, X_s^{(k)})\,ds + \int_0^t \sigma(s, X_s^{(k)})\,dB_s.
\end{aligned} \tag{9.19}$$

The goal is to show $X^{(k)}$ converges to some strong solution X. We will use the following lemma.

Lemma 9.6 *Let $f_n \ge 0$ be bounded, measurable and $A, B, C \ge 0$. Suppose that*

$$\begin{aligned}
f_0(t) &\le C \\
f_{n+1}(t) &\le A + B\int_0^t f_n(s)\,ds.
\end{aligned} \tag{9.20}$$

Thent

$$f_n(t) \le A\Big[1 + (Bt) + \cdots + \frac{(Bt)^{n-1}}{(n-1)!}\Big] + C\frac{(Bt)^n}{n!}.$$

In particular,

$$\limsup_{n\to\infty} f_n(t) \le Ae^{Bt}$$

and $f_n(t) \le C(Bt)^n/n!$ if $A = 0$.

Proof: Use induction. □

First we assume the initial condition has bounded second moments:

$$\mathsf{E}|\xi|^2 < \infty. \tag{9.21}$$

Lemma 9.7 *Assume* (9.21)*. Then for $T > 0$, there exists $C > 0$ such that*

$$\mathsf{E}|X_t^{(k)}|^2 \le C(1 + \mathsf{E}|\xi|^2)e^{Ct}, \quad 0 \le t \le T.$$

Proof: Let $\tau_n = \inf\{t : |X_t^{(k)}| \ge n\}$. For $0 \le t \le T$, we have

$$\begin{aligned}
\mathsf{E}|X_{t\wedge\tau_n}^{(k+1)}|^2 &\le 3\Big[\mathsf{E}|\xi|^2 + \mathsf{E}\Big(\int_0^{t\wedge\tau_n} b(s, X_s^{(k)})\,ds\Big)^2 + \mathsf{E}\int_0^{t\wedge\tau_n} |\sigma(s, X_s^{(k)})|^2\,ds\Big] \\
&\le 3\Big[\mathsf{E}|\xi|^2 + T\mathsf{E}\int_0^{t\wedge\tau_n} K^2(1 + |X_s^{(k)}|^2)\,ds + \mathsf{E}\int_0^{t\wedge\tau_n} K^2(1 + |X_s^{(k)}|^2)\,ds\Big] \\
&\le C_1(1 + \mathsf{E}|\xi|^2) + C_2\mathsf{E}\int_0^{t\wedge\tau_n} |X_s^{(k)}|^2\,ds.
\end{aligned}$$

Letting $\tau_n \uparrow \infty$, by Monotone Convergence Theorem, we have

$$\mathsf{E}|X_t^{(k+1)}|^2 \le C_1(1+\mathsf{E}|\xi|^2) + C_2 \int_0^t \mathsf{E}|X_s^{(k)}|^2\, ds.$$

The conclusion follows from Lemma 9.6. $\square$

Now let us show that $X_t^{(k)}$ given in (9.19) converges. We have

$$X_t^{(k+1)} - X_t^{(k)} = \int_0^t \left[b(s, X_s^{(k)}) - b(s, X_s^{(k-1)})\right] ds + \int_0^t \left[\sigma(s, X_s^{(k)}) - \sigma(s, X_s^{(k-1)})\right] dB_s.$$

Let

$$M_t^{(k)} = \int_0^t \left[\sigma(s, X_s^{(k)}) - \sigma(s, X_s^{(k-1)})\right] dB_s.$$

By Lemma 9.7 and (9.17), $M_t^{(k)}$ is a martingale. By Doob's Maximal inequality Theorem 4.14,

$$\mathsf{E} \sup_{0\le s\le t} |M_s^{(k)}|^2 \le 4\mathsf{E}|M_t^{(k)}|^2 \le 4(1+K^2)\int_0^t \mathsf{E}\sup_{0\le u\le s} |X_u^{(k)} - X_u^{(k-1)}|^2\, ds.$$

Combined with another simple estimate on the integral of b, we have

$$\mathsf{E} \sup_{0\le s\le t} |X_s^{(k+1)} - X_s^{(k)}|^2 \le C \int_0^t \mathsf{E}\sup_{0\le u\le s} |X_u^{(k)} - X_u^{(k-1)}|^2\, ds$$

for some constant $C > 0$ depending on K, T. Then, by Lemma 9.6,

$$\mathsf{E} \sup_{0\le s\le t} |X_s^{(k+1)} - X_s^{(k)}|^2 \le \Big(\mathsf{E}\sup_{0\le s\le t} |X_t^{(1)} - \xi|^2\Big) \cdot \frac{(Ct)^k}{k!} =: \tilde{C}\frac{(Ct)^k}{k!}.$$

Hence by Markov inequality,

$$\mathsf{P}\Big(\sup_{0\le t\le T} |X_t^{(k+1)} - X_t^{(k)}| \ge 2^{-k}\Big) \le \tilde{C}\frac{(4CT)^k}{k!}.$$

The right-hand side is summable, so by Borel–Cantelli, there exists $k_0 = k_0(\omega)$ such that

$$\sup_{0\le t\le T} |X_t^{(k+1)} - X_t^{(k)}| \le \frac{1}{2^k}$$

for $k \ge k_0(\omega)$. This implies

$$X_t = X_t^{(0)} + \sum_{k=1}^{\infty} X_t^{(k)} - X_t^{(k-1)}$$

converges uniformly to a continuous process almost surely. We can then pass the limit $k \to \infty$ to (9.19) to see that X solves (9.14). Moreover, $X_t^{(k)}$ is a functional of the Brownian motion, so is their limit X_t.

Finally, let us remove the condition (9.21). For $M > 0$, let $\Gamma_M = \{|\xi| \le M\}$ and $\xi_M = \xi \mathbb{1}_{\Gamma_M}$. Then $\xi_M \in L^2$, and hence we know there exists a unique strong solution, X^M, to (9.14) with initial condition ξ_M.

Our goal is show that X^M can be combined to obtain a strong solution to (9.14). We need to show:

- X^M satisfies

$$X_M \mathbb{1}_{\Gamma_M} = X^{M'} \mathbb{1}_{\Gamma_M}, \quad \forall M' > M, \tag{9.22}$$

which implies there exists a process X such that

$$X \mathbb{1}_{\Gamma_M} = X^M \mathbb{1}_{\Gamma_M}. \tag{9.23}$$

- X solves (9.14).

Since X^M is a solution to (9.14), we have

$$X_t^M = \xi_M + \int_0^t b(s, X_s^M)\, ds + \int_0^t \sigma(s, X_s^M)\, dB_s. \tag{9.24}$$

We claim that

$$X_t^M \mathbb{1}_{\Gamma_M} = \xi_M \mathbb{1}_{\Gamma_M} + \int_0^t b(s, X_s^M) \mathbb{1}_{\Gamma_M}\, ds + \int_0^t \sigma(s, X_s^M) \mathbb{1}_{\Gamma_M}\, dB_s. \tag{9.25}$$

This seems trivial It is not as trivial as it seems, to put $\mathbb{1}_{\Gamma_M}$ inside the integral, since the last integral is stochastic integral and not defined pathwise.

We will need some result similar to Theorem 6.9. Indeed, consider

$$T_M(\omega) = \begin{cases} 0, & \omega \notin \Gamma_M, \\ \infty, & \omega \in \Gamma_M, \end{cases}$$

then T_M is a stopping time. And for any $H \in L^2_{\text{loc}}$,

$$\mathbb{1}_{\Gamma_M} \int_0^t H_s\, dB_s = \int_0^{t \wedge T_M} H_s\, dB_s = \int_0^t H_s \mathbb{1}_{[0,T_M]}(s)\, dB_s = \int_0^t H_s \mathbb{1}_{\Gamma_M}\, dB_s. \tag{9.26}$$

And (9.25) follows from using this and (9.24).

Similarly, we also have for $M' > M$,

$$X_t^{M'} \mathbb{1}_{\Gamma_M} = \xi_M \mathbb{1}_{\Gamma_M} + \int_0^t b(s, X_s^{M'}) \mathbb{1}_{\Gamma_M}\, ds + \int_0^t \sigma(s, X_s^{M'}) \mathbb{1}_{\Gamma_M}\, dB_s \tag{9.27}$$

(noting that $\xi_{M'} \mathbb{1}_{\Gamma_M} = \xi_M \mathbb{1}_{\Gamma_M}$). Taking the difference of (9.25) and (9.27), it is routine to get for some $L = L(T) > 0$,

$$\mathsf{E} \sup_{u \in [0,t]} |X_u^M - X_u^{M'}|^2 \mathbb{1}_{\Gamma_M} \le L \int_0^t \mathsf{E} \sup_{u \in [0,s]} |X_u^M - X_u^{M'}|^2 \mathbb{1}_{\Gamma_M}, \quad \forall 0 \le t \le T.$$

And Gronwall's inequality implies that $\mathsf{E} \sup_{u \in [0,t]} |X_u^M - X_u^{M'}|^2 \mathbb{1}_{\Gamma_M} = 0$ which leads to (9.22). So we can find the process X satisfying (9.23).

To show that X is a solution, we notice that (9.23), (9.25) and (9.26) imply that

$$\begin{aligned} X_t \mathbb{1}_{\Gamma_M} = X_t^M \mathbb{1}_{\Gamma_M} &= \xi_M \mathbb{1}_{\Gamma_M} + \int_0^t b(s, X_s^M) \mathbb{1}_{\Gamma_M}\, ds + \int_0^t \sigma(s, X_s^M) \mathbb{1}_{\Gamma_M}\, dB_s \\ &= \xi \mathbb{1}_{\Gamma_M} + \int_0^t b(s, X_s) \mathbb{1}_{\Gamma_M}\, ds + \int_0^t \sigma(s, X_s) \mathbb{1}_{\Gamma_M}\, dB_s \\ &= \mathbb{1}_{\Gamma_M} \Big[\xi + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s \Big]. \end{aligned}$$

Since $\mathbb{1}_{\Gamma_M} \uparrow 1$, this shows that X is indeed a solution to (9.14).

9.5 Pathwise uniqueness for 1D SDE

In general, uniqueness and existence of solutions are established by very different techniques; that is how Theorem 9.3 plays a role. In this section, we present some general results on the pathwise uniqueness for 1D SDEs:

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dB_t, \quad b, \sigma : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}. \tag{9.28}$$

Proposition 9.8 *Assume that b is locally bounded and*

$$b(t, x) \le b(t, y), \quad x > y. \tag{9.29}$$

, and that $\sigma \equiv 1$. Then pathwise uniqueness holds for (9.28)*.*

Proof: Let $(X^{(j)}, B)$, $j = 1, 2$, be two weak solutions (defined on the same probability space). Let $\Delta_t = X_t^{(1)} - X_t^{(2)}$. Then

$$d\Delta_t = \big(b(t, X_t^{(1)}) - b(t, X_t^{(2)})\big)\,dt$$

and

$$d\Delta_t^2 = 2\Delta_t \cdot d\Delta_t = 2\Delta_t\big(b(t, X_t^{(1)}) - b(t, X_t^{(2)})\big)\,dt.$$

Note that (9.29) implies $(x - y)(b(t, x) - b(t, y)) \le 0$ for all x, y. Hence,

$$\Delta_t^2 = 2\int_0^t \big(X_t^{(1)} - X_t^{(2)}\big)\big(b(t, X_t^{(1)}) - b(t, X_t^{(2)})\big)\,dt \le 0.$$

So $\Delta_t \equiv 0$ a.s. and we get the desired conclusion. □

A more general statement than the pathwise uniqueness is comparison between different solutions.

Proposition 9.9 *Suppose $(X^{(j)}, B)$, $j = 1, 2$, are weak solutions to*

$$dX_t^{(j)} = b_j(t, X_t^{(j)})\,dt + \sigma(t, X_t^{(j)})\,dB_t, \quad X_0^{(j)} = \xi_j.$$

on the same probability space. Suppose that the coefficients and initial conditions of the SDEs satisfy the following.

- $\xi_1 \le \xi_2$ *almost surely.*
- *for all (t, x),*

$$b_1(t, x) \le b_2(t, x). \tag{9.30}$$

- b_1 *(or b_2) satisfies the global Lipschitz condition*

$$|b_1(t, x) - b_1(t, y)| \le K|x - y|, \quad x, y \in \mathbb{R}. \tag{9.31}$$

- σ *satisfies*

$$|\sigma(t, x) - \sigma(t, y)| \le h(|x - y|), \quad x, y \in \mathbb{R}, \tag{9.32}$$

  where $h : \mathbb{R}_+ \to \mathbb{R}_+$ is strictly increasing functions such that $h(0) = 0$ and

$$\int_0^1 h^{-2}(u)\,du = \infty. \tag{9.33}$$

Then $X_t^{(1)} \leq X_t^{(2)}$ for all $t \geq 0$, almost surely.

As a corollary, we also get pathwise uniqueness.

Corollary 9.10 *Suppose that in* (9.28)*, b is globally Lipschitz (i.e.,* (9.31)*) and σ satisfies* (9.32)*. Then pathwise uniqueness holds.*

Proof: By Proposition 9.9, for two weak solutions $(X^{(j)}, B)$, $j = 1, 2$, almost surely $X_t^{(1)} - X_t^{(2)}$ is both non-negative and non-positive, and hence is constantly 0. This proves the claim. □

Now let us return to the proof of Proposition 9.9. In general, results of pathwise uniqueness and comparison starts from applying Itô calculus to $\varphi(\Delta_t)$, and then using Gronwall's inequality to deduce $\varphi(\Delta_t) = 0$. For uniqueness, a common choice is $\varphi(w) = w^2$. For comparison, a natural choice is $\varphi(w) = w_+$, the positive part, since $(X_t^{(1)} - X_t^{(2)})_+ = 0$ will imply $X_t^{(1)} \leq X_t^{(2)}$. Formally, using Itô calculus on $\varphi(\Delta_t)$, one has

$$d(\Delta_t)_+ = \mathbb{1}_{\{\Delta_t \geq 0\}} \Big[(b_1(t, X_t^{(1)} - b_2(t, X_t^{(2)}))) \, dt + (\sigma(t, X_t^{(1)}) - \sigma(t, X_t^{(2)})) \, dB_t \Big], \tag{9.34}$$

where the Itô correction term is zero since $\varphi''(w) = 0$ except at $w = 0$. Taking expectation gives $\mathsf{E}(\Delta_t)_+ \leq K \displaystyle\int_0^t (\Delta_s)_+ \, ds$, and then Gronwall's inequality will finish the proof. The hole in this argument is that φ is not C^2 at $w = 0$, and Itô's formula is very sensitive regarding this. To fix this, one should approximate φ by a family of C^2-functions φ_n, and hope that (9.34) holds for φ_n with small errors terms, which can be got rid of in the $n \to \infty$ limit.

To approximate $\varphi(w) = w_+$ by C^2-functions, we impose the following constraints:

- $\varphi_n(x) = \varphi(x)$ when $x \notin (0, 1/n)$.
- $|\varphi_n'(x)| \leq 1$ for all x, and $\varphi'(x)$ is increasing from $x = 0$ to $x = 1/n$.

Anything such function is uniquely determined by its second derivative $\varphi''(x) =: \eta(x)$ that satisfying

$$\eta(x) = 0, \quad x \notin (0, 1/n), \qquad \eta(x) \geq 0, \quad \int_0^{1/n} \eta(x) \, dx = 1. \tag{9.35}$$

We have the freedom to choose any function η, and we will take advantage of this in the proof.

Proof of Proposition 9.9: After localization we can assume that

$$\mathsf{E} \int_0^t \sigma^2(s, X_s^{(j)}) \, ds < \infty$$

for all $t \geq 0$.

Let $\varphi_n \in \mathcal{C}^2$ such that $\varphi_n(0) = \varphi_n'(0) = 0$ and φ'' satisfies (9.35). Let $\Delta_t = X_t^{(1)} - X_t^{(2)}$. Applying Itô's formula to $\varphi_n(\Delta_t)$ and then taking expectation, we obtain

$$\mathsf{E}\varphi_n(\Delta_t) = \mathsf{E} \int_0^t \varphi'(\Delta_s) \big[b_1(s, X_s^{(1)}) - b_2(s, X_s^{(2)}) \big] \, ds + \frac{1}{2} \mathsf{E} \int_0^t \varphi_n''(\Delta_s) \Big[\sigma(s, X_s^{(1)}) - \sigma(s, X_s^{(2)}) \Big]^2 \, ds =: I_1 + I_2.$$

For the integrand inside I_1, we have

$$\varphi'(\Delta_s) \big[b_1(s, X_s^{(1)}) - b_2(s, X_s^{(2)}) \big] \leq \varphi'(\Delta_s) \big[b_1(s, X_s^{(1)}) - b_1(s, X_s^{(2)}) \big] + \varphi'(\Delta_s) \big[b_1(s, X_s^{(2)}) - b_2(s, X_s^{(2)}) \big] \leq K(\Delta_s)_+.$$

The condition(9.31) to bound The first term is estimated using (9.31), noting that it is zero if $\Delta_s \leq 0$, and the second term is non-positive due to (9.30).

For I_2, we first have the following observation: if $f(x)$ is a decreasing positive function on $(0,\delta]$ such that $\int_0^\delta f(x)\,dx = \infty$, then for every $\varepsilon > 0$, there exists an positive function $g(x)$ such that $g(x) \le \varepsilon f(x)$ and $\int_0^\delta g(x)\,dx = 1$.

Now we choose $\varphi''(x)$ and hence φ_n as follows. For any $\varepsilon_n \downarrow 0$, let $\varphi''(x)$ satisfy (9.35) and $\varphi''(x) \le \varepsilon_n h^{-2}(x)$, which is possible by the observation above and the assumption (9.33). Then, by (9.32),

$$I_2 \le \frac{1}{2}\mathsf{E}\int_0^t \varphi''(\Delta_s)h^2(\Delta_s)\,ds \le \frac{\varepsilon_n t}{2}.$$

Combining all these, we have

$$\mathsf{E}\varphi_n(\Delta_t) \le K\int_0^t \mathsf{E}(\Delta_s)_+\,ds + \frac{\varepsilon_n t}{2}.$$

Letting $n\to\infty$, one has

$$\mathsf{E}(\Delta_t)_+ \le K\int_0^t \mathsf{E}(\Delta_s)_+,$$

and the desired conclusion will follow from Gronwall's inequality. □

Remark 9.5 A direct proof of the pathwise uniqueness could be done by applying Itô's formula to $\tilde\varphi_n(\Delta_t)$, where $\tilde\varphi_n(w)$ approximate $|w|$ in a similar way. In fact, one can just define $\tilde\varphi_n$ to be an even function such that $\tilde\varphi_n(w) = \varphi_n(w)$ when $w \ge 0$.

9.6 Some examples of SDEs

9.6.1 Linear equations

Here we consider a multi-dimensional SDE

$$dX_t = \big[A(t)X_t + a(t)\big]\,dt + \sigma(t)\,dB_t, \tag{9.36}$$

where $A(t)\in\mathbb{R}^{d\times d}$, $a(t)\in\mathbb{R}^{d\times 1}$, $\sigma(t)\in\mathbb{R}^{d\times r}$ and B is a r-dimensional Brownian motion. Assume that all the coefficients are locally bounded.

To solve (9.36) we will borrow the idea of the method "variation of constant" in ODE theory. Suppose that $\Phi(t)\in\mathbb{R}^{d\times d}$ solves the matrix equation (which is called a *fundamental solution*)

$$\dot\Phi(t) = A(t)\Phi(t), \quad \Phi(0) = I.$$

In variation of constant, we expect $X(t) = \Phi(t)v(t)$ for some "varying" vector $v(t)$. Effectively, we need to deduce what equation $v(t) = \Phi^{-1}(t)X(t)$ solves. We have

$$d(\Phi^{-1}(t)X_t) = \Phi^{-1}(t)dX_t + \big[\Phi^{-1}(t)\big]'X_t\,dt. \tag{9.37}$$

Noting that $I \equiv \Phi(t)\cdot\Phi^{-1}(t)$, so by product rule

$$0 = \Phi'(t)\cdot\Phi^{-1}(t) + \Phi(t)\cdot\big[\Phi^{-1}(t)\big]' \quad\Rightarrow\quad \big[\Phi^{-1}(t)\big]' = \Phi^{-1}(t)\Phi'(t)\Phi^{-1}(t).$$

Back to (9.37) we then have

$$\begin{aligned} d(\Phi^{-1}(t)X_t) &= \Phi^{-1}(t)dX_t + \Phi^{-1}(t)\Phi'(t)\Phi^{-1}(t)X_t\,dt = \Phi^{-1}(t)\Big[dX_t + A(t)X_t\,dt\Big] \\ &= \Phi^{-1}(t)\Big[a(t)\,dt + \sigma(t)\,dB_t\Big]. \end{aligned}$$

The right-hand side does not depend on X and can be integrated. We arrive at the explicit form of the strong solution

$$X_t = X_0 + \Phi(t) \int_0^t \Phi^{-1}(s) \Big[a(s)\, ds + \sigma(s)\, dB_s \Big]. \tag{9.38}$$

9.6.2 Ornstein-Uhlenbeck process

The OU process is the strong solution to the SDE

$$dX_t = dB_t - \lambda X_t dt, \quad X_0 = \xi, \tag{9.39}$$

where $\lambda > 0$ is a constant. It is a special linear SDE so strong solution exists and is unique. The expression (9.38) specialized to the OU process gives

$$X_t = e^{-\lambda t} X_0 - \int_0^t e^{-\lambda(t-s)} dB_s. \tag{9.40}$$

Intuitively, OU process is like the Brownian motion with a drift towards the origin. So, in contrast with Brownian motion which is a Markov process without stationary distribution, the OU process has a stationary distribution, which is Gaussian. In general, for a Markov process $X_t \in \mathbb{R}$, a stationary distribution is a probability measure on $\mathbb{R}$ such that if $X_0 \sim \mu$, then $X_t \sim \mu$ for all $t \geq 0$.

There are several ways to understand the stationary distribution. First, imagine we are solving the equation from $-\infty$ instead of time 0, then (9.40) implies

$$X_t = \lim_{t_0 \to -\infty} e^{-\lambda(t-t_0)} X_{t_0} - \int_{t_0}^t e^{-\lambda(t-s)} dB_s = - \int_{-\infty}^t e^{-\lambda(t-s)}\, dB_s.$$

(We need to define what is Brownian motion B_t for $t < 0$; this is done by the so-called *two-sided Brownian motion* by running an independent copy of Brownian motion backward from 0.) The key point of the analysis is that the initial condition is "forgot" by the limiting procedure due to the exponential decaying factor, and the stochastic integral, even defined on an infinite interval, still makes sense. In fact, since the integrand is deterministic, one has

$$- \int_{-\infty}^t e^{-\lambda(t-s)}\, dB_s \sim \mathcal{N}(0, \int_{-\infty}^t e^{-2(\lambda-s)}\, ds) = \mathcal{N}(0, \frac{1}{2\lambda}).$$

And $\mathcal{N}(0, (2\lambda)^{-1})$ is the stationary distribution.

We will use the semi-group theory to check that $\mathcal{N}(0, (2\lambda)^{-1})$ is indeed stationary. Note that the density of a Markov process evolves according to (9.11). It is easy to compute $\mathcal{L}$ and $\mathcal{L}^*$ for the OU process:

$$\mathcal{L} f = \frac{1}{2} \partial_{xx} f - \lambda x \partial_x f, \quad \mathcal{L}^* g = \frac{1}{2} \partial_{xx} g + \partial_x \big(\lambda x g \big) = \frac{1}{2} \partial_{xx} g + \lambda g + \lambda x \partial_x g.$$

Then for $\rho(x) = C e^{-\lambda x^2}$,

$$\partial_x \rho = \rho \cdot (-2\lambda x), \quad \partial_{xx} \rho = \rho \cdot \big[(-2\lambda x)^2 - 2\lambda \big].$$

So

$$\mathcal{L}^* \rho = \rho \Big[\frac{1}{2} (4\lambda^2 x^2 - 2\lambda) + \lambda + \lambda x (-2\lambda x) \Big] = 0.$$

9.6.3 Brownian Bridge

Let $T > 0$. The following SDE is called Brownian bridge over $[0, T]$:

$$dX_t = \frac{-X_t}{T-t}\,dt + dB_t,\ t \in [0, T) \quad X_0 = 0. \tag{9.41}$$

Note that when $t \uparrow T$, there is a singularity for the drift term $b(t, x) = \dfrac{-x}{T-t}$, but otherwise (9.41) is a linear SDE. To define construct a solution to (9.41), we first obtain the existence of strong solution over each interval $[0, T-\varepsilon]$, $\varepsilon > 0$. By uniqueness of the strong solutions, all these solutions will be consistent with each other, so that we obtain a solution over $[0, T)$. Indeed, $b(t, x) = -\dfrac{x}{T-t}$ is still "locally finite", where "local" means for every compact subset inside $t \in [0, T)$. The solution can be expressed by

$$X_t = (T-t)\int_0^t \frac{dB_s}{T-s}, \quad 0 \le s < T. \tag{9.42}$$

There is another definition of Brownian bridge: let W_t be a Brownian motion, then

$$X_t = W_t - t/T \cdot W_T, \quad t \in [0, T] \tag{9.43}$$

is a Brownian bridge. Of course, the Brownian motions W_t in (9.43) and B_t in (9.41) are differently. From (9.43) one can easily read the f.d.d. of Brownian bridge, but it is not clear that B_t is a Markov process. Another point is that (9.43) implies that $X_T = 0$, which is not clear from (9.41).

Let us prove that X_t defined in (9.43) satisfies $\lim_{t\uparrow T} X_t = 0$. Indeed, let

$$M_t = \int_0^t \frac{dB_s}{T-s}.$$

Then

$$\langle M\rangle_t = \frac{1}{T-t} - \frac{1}{T}.$$

We can express M as a time-changed Brownian motion: $M_t = \beta_{\langle M\rangle_t}$, so that

$$X_t = \frac{\beta_{\langle M\rangle_t}}{\langle M\rangle_t + \frac{1}{T}}.$$

But for any Brownian motion β, by Strong Law of Large Numbers one has

$$\lim_{s\uparrow\infty} \frac{\beta_s}{s} = 0.$$

So

$$\lim_{t\uparrow T} X_t = \lim_{s\to\infty} \frac{\beta_s}{s + \frac{1}{T}} = 0.$$

10 Weak solution and martingale problem

10.1 Yamada-Watanabe Theorem

In this section we will prove Theorem 9.3.

10.1.1 Weak uniqueness

For simplicity we will work in one dimension: $d = r = 1$. To start, let us assume that we have two weak solutions $(X^{(j)}, W^{(j)})$, defined on two probability spaces $(\Omega^{(j)}, \mathcal{F}^{(j)}, \nu_j)$, with filtrations $(\mathcal{F}_t^{(j)})$. To separate the initial condition, we introduce the processes $Y_t^{(j)} = X_t^{(j)} - X_0^{(j)}$ and write the weak solutions as a triple $(X_0^{(j)}, W^{(j)}, Y^{(j)})$. We think of the triple as a random element taking value in

$$\Theta := \mathbb{R} \times \mathcal{C}_0[0,\infty) \times \mathcal{C}_0[0,\infty),$$

where the subscript 0 means that the continuous process starts from 0. The space Θ is a Cartesian product of three metric spaces, and hence is also a metric space. The measurable sets are just the all the Borel sets, denoted by $\mathcal{B}(\Theta)$. The initial condition $X_0^{(j)}$ will have distribution μ. We will write a general element in Θ by $\theta = (x, w, y)$

To apply pathwise uniqueness, we need to have two weak solutions defined on the same probability space. Since weak solutions only care about the distribution, it is natural to consider the measures on Θ induced by ν_j, namely,

$$\mathsf{P}_j(A) = \nu_j\big((X_0^{(j)}, W^{(j)}, Y^{(j)}) \in A\big), \quad A \in \Theta, \qquad j = 1, 2.$$

Another crucial point in pathwise uniqueness is that the driving Brownian motion must be the same. Although as weak solutions, $Y^{(j)}$ is generally NOT a functional of $W^{(j)}$, but the distribution of $Y^{(j)}$ will depend on $W^{(j)}$; this is the idea of conditional probability. More precisely, we are trying to decompose P_j into

$$\mathsf{P}_j(A) = \int_A Q_j(x, w; dy)\mathsf{P}^W(dw)\mu(dx), \tag{10.1}$$

where $\mathsf{P}^W(\cdot)$ is the Wiener measure on $\mathcal{C}_0[0,\infty)$, i.e., the law of standard Brownian motion on $\mathcal{C}_0[0,\infty)$, and for each (x, w), $Q_j(x, w; dy)$ is a probability measure on $\mathcal{C}_0[0,\infty)$, which is the conditional distribution of $Y^{(j)}$ given $(X_0^{(j)}, W^{(j)}) = (x, w)$. The decomposition (10.1) is rigorously defined via the *regular conditional probability*.

Definition 10.1 *Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and $\mathcal{G} \subset \mathcal{F}$. A regular condition probability is a functional $Q(\omega; A) : \Omega \times \mathcal{F} \to [0, 1]$ such that*

1. *$\forall \omega \in \Omega$, $Q(\omega; \cdot)$ is a probability measure on $(\Omega, \mathcal{F})$.*
2. *$\forall A \in \mathcal{F}$, $\omega \mapsto Q(\omega; A)$ is a $\mathcal{G}$-measurable.*
3. *$\forall A \in \mathcal{F}$, $Q(\omega; A) = \mathsf{P}[A \mid \mathcal{G}](\omega)$, P-a.e. $\omega \in \Omega$.*

A sufficient condition for the existence of regular condition probability is that $(\Omega, \mathcal{F})$ is a *Borel space*, i.e., there exists a one-to-one bijection $\varphi : (\Omega, \mathcal{F}) \to ([0, 1], \mathcal{B}([0, 1]))$ such that φ and φ^{-1} are both measurable. A *Polish space*, i.e., a complete separable metric space, equipped with its Borel σ-algebra is a Borel space.

As a special case of the regular condition probability, we consider a probability measure P on a product space $\Omega = \Omega_1 \times \Omega_2$ where Ω_j are metric spaces equipped with their Borel σ-algebra $\mathcal{F}_j$. The coordinate map $(X(\omega), Y(\omega)) := (\omega_1, \omega_2)$ may be regarded as a pair of random elements on $(\Omega, \mathcal{F}, \mathsf{P})$ where $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$. We want to have such decomposition

$$\mathsf{P}(d\omega_1, d\omega_2) = Q(\omega_1; d\omega_2)\mathsf{P} \circ X^{-1}(d\omega_1). \tag{10.2}$$

In the case where P is a probability measure on $\mathbb{R}^2$ with joint density $\rho(x, y) > 0$, it is very clear that

$$\mathsf{P} \circ X^{-1}(dx) = \rho_X(x) = \int_{\mathbb{R}} \rho(x, y)\, dy,$$

and

$$Q(x; dy) = \rho_{Y|X}(y \mid x)\, dy, \quad \rho_{Y|X}(y \mid x) = \frac{\rho(x, y)}{\rho_X(x)},$$

where $Q(x; \cdot)$ is a probability measure and $(x, y) \mapsto \rho_{Y|X}(y \mid x)$ is measurable.

Now let us apply the regular conditional probability to obtain (10.2) in the general setting. Let $\mathcal{G} = \sigma(X)$. Assuming Ω is Polish, there exists a regular condition probability $\tilde{Q}(\omega; A)$. Since $\omega \mapsto \tilde{Q}(\omega; A)$ is $\mathcal{G}$-measurable, we can write

$$\tilde{Q}(\omega; A) = \tilde{Q}(X(\omega); A) = \tilde{Q}(\omega_1; A).$$

For $A = G \times F$ where $G \in \mathcal{G}$, we have

$$\mathsf{P}(G \times F) = \int \tilde{Q}(\omega_1; G \times F) \mathsf{P} \circ X^{-1}(d\omega_1).$$

But $\tilde{Q}$ is a functional on $\Omega_1 \times \mathcal{F}$. Compared to (10.2), we need a functional Q on $\Omega_1 \times \mathcal{F}_2$ such that

$$\tilde{Q}(\omega_1; G \times F) = \mathbb{1}_G(\omega_1) \cdot Q(\omega_1; F), \quad G \in \mathcal{G},\ F \in \mathcal{F}_2, \tag{10.3}$$

such that $Q(\omega_1; \cdot)$ is a probability measure and $\omega_1 \mapsto Q(\omega_1; F)$ is $\mathcal{G}$-measurable. Note that $\tilde{Q}(\omega_1; \cdot)$ is already a probability measure, so $\tilde{Q}(\omega_1; G \times \cdot)$ is a measure on $\mathcal{F}_2$. To be a probability measure, it must have total mass 1 when $\omega_1 \in G$, i.e., we need

$$\tilde{Q}(\omega_1; G \times \Omega_2) = \mathbb{1}_G(\omega_1). \tag{10.4}$$

For every $G \in \mathcal{G}$, (10.4) holds for $\mathsf{P} \circ X^{-1}$-a.e. ω_1 due to Item 3 in Definition 10.1. We can then define Q via (10.4). To get the measurability of $\omega_1 \mapsto Q(\omega_1; F)$, we need to be careful about the exceptional zero-measure sets where (10.4) fails. These sets depends on G. In order to obtain a *common* exceptional set, a sufficient condition is that $\mathcal{G}$ is *countably determined*. Note that the Borel σ-algebra of a Polish space is always countably determined (due to separability).

Finally we can put everything together. Note that Θ is a Polish space, so that regular conditional probability exists and that a common zero-measure exceptional set for (10.4) can be found (we use separability twice in for different purposes.) For P_j, there exists $Q_j(x, w; F) : \mathbb{R} \times \mathcal{C}_0[0, \infty) \times \mathcal{B}(\mathcal{C}_0[0, \infty))$ such that

- for every (x, w), $Q_j(x, w; \cdot)$ is a probability measure.
- for every F, $(x, w) \mapsto Q_j(x, w; F)$ is $\mathcal{B}(\mathbb{R} \times \mathcal{C}_0[0, \infty))$-measurable.
- for every $G \in \mathbb{R} \times \mathcal{C}_0[0, \infty)$ and $F \in \mathcal{C}_0[0, \infty)$,

$$\mathsf{P}_j(G \times F) = \int_G Q_j(x, w; F) \mu(dx) \mathsf{P}^W(dw). \tag{10.5}$$

Now we turn to the proof of weak uniqueness. Let $\tilde{\Theta} = \mathbb{R} \times \left[\mathcal{C}_0[0, \infty)\right]^3$. Define the following probability measure on $(\tilde{\Theta}, \mathcal{B}(\tilde{\Theta}))$:

$$\mathsf{P}(dx, dw, dy_1, dy_2) = Q_1(x, w; dy_1) \cdot Q_2(x, w; dy_2) \mu(dx) \mathsf{P}^W(dw).$$

We can equip the probability space with a proper augmentation of the filtration $\mathcal{B}_t = \mathcal{B}(\mathbb{R}\times\big[\mathcal{C}_0[0,t]\big]^3)$. On $(\tilde{\Theta}, \mathcal{B}(\tilde{\Theta}), \mathsf{P})$, by (10.5), we have

$$\mathsf{P}((x,w,y_j)\in A) = \int \mathbb{1}_{\{(x,w,y_j)\in A\}} Q_j(x,w;F)\mu(dx)\mathsf{P}^W(dw) = \mathsf{P}_j(A), \quad A\in\mathcal{B}(\tilde{\Theta}).$$

(First prove this for $A = G\times F$ and then for general A by standard argument.) But $(X^{(j)}, B^{(j)}) = (x+y_j, w)$, $j=1,2$ are two weak solutions on $(\tilde{\Theta}, \mathcal{B}(\tilde{\Theta}), \mathsf{P})$. By pathwise uniqueness, $X^{(1)} = X^{(2)}$ under P, so

$$\mathsf{P}_1(A) = \mathsf{P}((x,w,y_1)\in A) = \mathsf{P}((x,w,y_2)\in A) = \mathsf{P}_2(A), \quad A\in\mathcal{B}(\tilde{\Theta}).$$

This is the weak uniqueness.

10.1.2 Strong existence

Continuing the discussion of Section 10.1.1, let $B = \{(y_1,y_2): y_1(t) = y_2(t)\}\subset\mathcal{B}\big(\mathcal{C}_0[0,\infty)^2\big)$ and

$$Q(x,w;dy_1dy_2) = Q_1(x,w;dy_1)\cdot Q_2(x,w;dy_2).$$

Then pathwise uniqueness implies that

$$1 = \mathsf{P}((y_1,y_2)\in B) = \int Q(x,w;B)\mu(dx)\mathsf{P}^W(dw).$$

Therefore, $Q(x,w;B) = 1$ for all $(x,w)\in N^c$ where N is some P-null set in $\mathcal{B}(\mathbb{R}\times\mathcal{C}_0[0,\infty))$.

We first note that since for every $(x,w)\in N^c$, $Q(x,w;\cdot)$ is a product measure of Q_1 and Q_2. So $Q(x,w;B)=1$ implies that there exists $k(x,w)\in\mathcal{C}_0[0,\infty)$ such that

$$Q_j(x,w;\{k(x,w)\}) = 1, \quad j=1,2.$$

In other words, pathwise uniqueness forces the solution process $y(t)$ to be a functional of (x,w).

It remains to show that $(x,w)\mapsto k(x,w)$ is progressively measurable. Let us first show that it is measurable. In fact, for any $\Gamma\in\mathcal{B}\big(\mathcal{C}_0[0,\infty)\big)$, since $Q_1(x,w;\cdot)$ a.s. assign full measure to a singleton,

$$\{(x,w): k(x,w)\in\Gamma\} \overset{\text{a.s.}}{=} \{Q_1(x,w;\Gamma)=1\}\in\mathcal{B}\big(\mathbb{R}\times\mathcal{C}_0[0,\infty)\big)$$

by measurability of the regular condition probability. This shows measurability of $(x,w)\mapsto k(x,w)$.

To obtain progressive measurability, we need to first establish that $(x,w)\mapsto Q_j(x,w;\cdot)$ are progressively measurable. Then the same argument as above will lead to progressive measurability of $(x,w)\mapsto k(x,w)$. We skip the details here.

Now it is clear that we have strong uniqueness. Let (B_t) be a Brownian motion on a probability space $(\Omega,\mathcal{F},\mathsf{P})$ and ξ be a r.v. independent of B with distribution μ. Then

$$X_t = \xi + \big[k(x,B)\big](t)$$

is a strong solution to the problem.

10.2 Martingales from SDEs

Let X_t be a weak solution to (9.14). For any $f \in \mathcal{C}^{1,2}\big([0,\infty) \times \mathbb{R}^d\big)$, by Itô's formula,

$$df(X_t) = \sum_{i=1}^{d} (\partial_i f)(X_t) \cdot \big[b_i(t, X_t)\, dt + \sum_{k=1}^{r} \sigma_{ik}(t, X_t)\, dB_t^{(k)}\big] + \frac{1}{2} \sum_{i,j=1}^{d} (\partial_{ij} f)(X_t) \sum_{k=1}^{r} (\sigma_{ik}\sigma_{kj})(t, X_t)\, dt. \tag{10.6}$$

This leads to the definition of the generator

$$(\mathcal{L}_t f)(x) = \sum_{i=1}^{d} b_i(t,x)(\partial_i f) + \frac{1}{2} \sum_{i,j=1}^{r} a_{ij}(t,x) \partial_{ij} f,$$

where

$$A = (a_{ij})_{i,j=1}^{d} =: \sigma\sigma^T = \Big(\sum_{k=1}^{r} \sigma_{ik}\sigma_{kj}\Big)_{i,j=1}^{d}$$

is the *diffusion* matrix. We have the following observation.

Proposition 10.1 *Let (X, W) be a weak solution and $f \in \mathcal{C}^{1,2}([0,\infty) \times \mathbb{R}^d)$. Then*

$$M_t^f := f(t, X_t) - f(0, X_0) - \int_0^t (\partial_s + \mathcal{L}_s) f(s, X_s)\, ds \tag{10.7}$$

is a c.l.m.

Moreover, let $f, g \in \mathcal{C}^{1,2}$ and M^f, M^g be defined by (10.7), then

$$\langle M^f, M^g \rangle_t = \sum_{i,j=1}^{d} \int_0^t a_{ij}(s, X_s)(\partial_i f)(s, X_s)(\partial_j g)(s, X_s)\, ds.$$

Proof: Consider the process stopped at the stopping times

$$\tau_n = \inf\{t \geq 0 : |X_t| \geq n \text{ or } \sup_{i,k} \int_0^t \sigma_{ik}^2(s, X_s)\, ds \geq n\}.$$

As a weak solution one must have $\tau_n \uparrow \infty$ a.s. The proposition essentially follows from the computation (10.6). $\square$

We will see that the martingales defined in (10.7) also characterize the process X. To start, we investigate the case of Brownian motion.

Proposition 10.2 *Let X be a adapted, continuous process. Then X is a Brownian motion if and only if*

$$f(X_t) - f(X_0) - \int_0^t \big(\frac{1}{2}\Delta f\big)(X_s)\, ds$$

is a c.l.m. for every $f \in \mathcal{C}^2(\mathbb{R}^d)$.

Proof: The "$\Rightarrow$" direction follows from Proposition 10.1. For the other direction, we consider two special classes of $\mathcal{C}^2$-functions:

$$f_i(x) = x_i, \quad f_{ij}(x) = x_i x_j, \qquad i, j = 1, 2, ..., d.$$

Note that $\Delta f_i = 0$ and $\Delta f_{ij} = 2\delta_{ij}$. By the assumption, X^i are c.l.m.'s and $X^iX^j - \delta_{ij}t$ are c.l.m.'s, i.e., $\langle X^i, X^j\rangle_t$. Hence, by Theorem 7.1, X^i are independent Brownian motions. This completes the proof. $\square$

A local martingale solution is basically a probability measure on $\mathcal{C}[0,\infty)^d$. Let $(\Omega, \mathcal{F}_0) = (\mathcal{C}[0,\infty)^d, \mathcal{B}(\mathcal{C}[0,\infty)^d))$ and P be a probability measure on $(\Omega, \mathcal{F}_0)$. Consider the following procedure of augmentation. Let $\mathcal{N}$ be the collection of all P-null sets. Let $\mathcal{B}_t = \mathcal{B}_t(\mathcal{C}[0,\infty)^d)$ be the natural filtration. Let $\mathcal{G}_t = \sigma(\mathcal{B}_t \cup \mathcal{N})$. Let

$$\mathcal{F} = \sigma(\mathcal{F}_0 \cup \mathcal{N}), \quad \mathcal{F}_t = \mathcal{G}_{t+}.$$

Then $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathsf{P})$ is a probability space with a filtration satisfying the usual condition.

Definition 10.2 *A solution to the local martingale problem* (10.7) *is a probabilistic distribution* P *on* $(\Omega, \mathcal{F}_0)$ *such that for all* $f \in \mathcal{C}^2(\mathbb{R}^d)$, *if* $\mathbf{y} = (y(t))_{t\geq 0}$ *has the law* P, *then*

$$M_t^f = f(y(t)) - f(y(0)) - \int_0^t (\mathcal{L}_s f)(y(s))\, ds,$$

is a $(\mathcal{F}_t)$*-c.l.m., where* $\mathcal{F}_t$ *are given by the above augmenting procedure.*

It turns out that to verify a solution to the local martingale problem, we do not need to check every $f \in \mathcal{C}^2$, but only all the polynomials in x_i of degree one and two, which is already the case in Proposition 10.2.

Theorem 10.3 *If* M^f *is a c.l.m. for* f *being* $f_i(x) = x_i$ *and* $f_{ij}(x) = x_ix_j$, *then there exists a Brownian motion* $(B_t)_{t\geq 0}$ *on* $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t), \tilde{\mathsf{P}})$, *an extension of* $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathsf{P})$, *such that* $(X_t = y(t), B_t)$ *is a weak solution.*

Consequently, M^f *is a c.l.m. for every* $f \in \mathcal{C}^2$.

Proof: **Step 1.** Let

$$M_t^{(i)} = X_t^{(i)} - X_0^{(i)} - \int_0^t b_i(s, X_s)\, ds.$$

We claim that

$$M_t^{(i)} M_t^{(j)} - \int_0^t a_{ij}(s, X_s)\, ds$$

is a c.l.m., which implies that

$$\langle M^{(i)}, M^{(j)}\rangle_t = \int_0^t a_{ij}(s, X_s)\, ds. \tag{10.8}$$

Indeed, from $dM_t^{(i)} = dX_t - b_i(t, X_t)\, dt$, $M^{(i)}$ and $X^{(i)}$ differ by a finite variation process, and hence

$$\begin{aligned} dX_t^{(i)} X_t^{(j)} &= X_t^{(i)}\, dX_t^{(j)} + X_t^{(j)}\, dX_t^{(i)} + d\langle M^{(i)}, M^{(j)}\rangle_t \\ &= X_t^{(i)}\, dM_t^{(j)} + X_t^{(j)}\, dM_t^{(i)} + b_j(t, X_t) X_t^{(i)}\, dt + b_i(t, X_t) X^{(j)}\, dt + d\langle M^{(i)}, M^{(j)}\rangle_t. \end{aligned}$$

One the other hand, using the assumption on $f = f_{ij}$,

$$dX_t^{(i)} X_t^{(j)} = \text{m.t.} + b_i(t, X_t) dX_t^{(j)} + b_j(t, X_t) dX_t^{(i)} + a_{ij}(t, X_t)\, dt.$$

Comparing these two displays proves the claim.

Step 2. By [KS98, Theorem 3.4.2], if $M^{(i)}$ are c.l.m.'s with cross variation (10.8), then there exists an extended probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t), \tilde{\mathsf{P}})$ on which there are a d-dimensional Brownian motion

$\tilde{W} = (\tilde{W}^{(1)}, \dots, \tilde{W}^{(d)})$ and $(\tilde{\mathcal{F}}_t)$-adapted processes $(\rho_{ij}(t))_{t\geq 0}$, $1 \leq i, j \leq d$ such that $\tilde{\mathsf{P}}\big(\int_0^t \rho_{ij}^2(s)\, ds < \infty\big) = 1$ for all $t > 0$ and

$$M_t^{(i)} = \sum_{j=1}^d \int_0^t \rho_{ij}(s)\, d\tilde{W}_s^{(j)}. \tag{10.9}$$

We illustrate the idea in the case $d = 1$. We can define

$$\tilde{W}_t^{(1)} = \int_0^t \frac{1}{\sqrt{a_{11}(s)}}\, dM_s^{(1)}.$$

Clearly, $\tilde{W}^{(1)}$ is a c.l.m. and $\langle \tilde{W}^{(1)} \rangle_t = t$, so by Theorem 7.1, $\tilde{W}^{(1)}$ is Brownian motion and (10.9) holds with $\rho_{11}(t) = \sqrt{a_{11}(t)}$. One may need to extend the probability space to handle the singular case where $a_{11}(t) = 0$ for some t.

Step 3. Assuming (10.9), we need to show there exists r-dimensional Brownian motion W such that

$$\int_0^t \rho(s)\, d\tilde{W}_s = \int_0^t \sigma(s, X_s)\, dW_s,$$

so that

$$M_t^{(i)} = \sum_{j=1}^d \int_0^t \rho_{ij}(s)\, d\tilde{W}_s^{(j)} = \sum_{j=1}^r \int_0^t \sigma(s, X_s)\, dW_s. \tag{10.10}$$

So that (X, W) indeed is a weak solution.

Note that the only relation between the matrices ρ and σ is

$$\rho\rho^T = A = \sigma\sigma^T, \quad \mathsf{P}\text{-a.s.}$$

The claim follows from a linear algebra construction: there exists a Borel-measurable map

$$\begin{aligned} R : \mathbb{R}^{d\times d} \times \mathbb{R}^{d\times r} &\to \mathbb{R}^{r\times d} \\ (\rho, \sigma) &\mapsto R(\rho, \sigma) \end{aligned}$$

such that if $\rho\rho^T = \sigma\sigma^T$, then $\rho = \sigma R$. We skip the construction here. With this at hand, we can define

$$W_t = \int_0^t R\Big(\rho(s), \sigma(s, X_s)\Big)\, d\tilde{W}_s$$

and (10.10) holds.

□

We will finish this section by two generalizations of the local martingale problem.

The first is that we can also consider the more general *functional SDEs*, namely, the coefficients b, σ depend on the trajectories $t \mapsto y(t)$ but not just $y(t)$. To fix the idea, let $b_i(t, \mathbf{y})$, $\sigma_{ij}(t, \mathbf{y})$ be progressively measurable functions. Then for $u \in \mathcal{C}^2$, we can define $\mathcal{L}'_t$ by

$$(\mathcal{L}'_t u)(\mathbf{y}) = \frac{1}{2} \sum_{i=1}^d \sum_{k=1}^r a_{ik}(t, \mathbf{y}) \partial_{ik} u\big(y(t)\big) + \sum_{i=1}^d b_i(t, \mathbf{y}) \partial_i u\big(y(t)\big),$$

and accordingly for $f \in \mathcal{C}^{1,2}$,

$$M_t^f = f(t, y(t)) - f(0, y(0)) - \int_0^t (\partial_s f + \mathcal{L}'_s f)(s, y_s)\, ds.$$

The second is the *martingale problem*, in which instead of taking $f \in \mathcal{C}^2(\mathbb{R}^d)$, we test all functions $f \in \mathcal{C}_0^2(\mathbb{R}^d)$, and the process M^f will be continuous martingales instead of c.l.m.'s. Note that we can always approximate f_i, f_{ij} by some $f_i^N, f_{ij}^N \in \mathcal{C}_0^2(\mathbb{R}^d)$ that coincide with them for $|x| \le N$. To justify the limiting process $N \uparrow \infty$ and obtain the equivalency between the martingale problem and the local version, we need the **local boundedness of** σ, i.e.,

- in the functional SDE case,

$$|\sigma(t, \mathbf{y})| \le K_T, \quad \forall 0 \le t \le T, \ \mathbf{y} \in \mathcal{C}[0, \infty)^d,$$

- or, in the non-functional case, $\sigma(t, y)$ is locally bounded.

10.3 Existence for martingale problem solution

In this section we consider the case $d = 1$ and the time-homogeneous SDE

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dW_t. \tag{10.11}$$

We assume that b, σ are bounded and continuous functions. We also assume the initial distribution μ has $2m$-th moment for some $m > 1$.

Theorem 10.4 (Varadhan–Stroock) *For the above SDE, there exists a solution to the martingale problem, and hence a weak solution exists.*

The starting point is to use Euler scheme to construct approximated solutions. In fact, we will see a story parallel to what happens in the theory of ODEs: for ODEs, when the coefficients are Lipschitz continuous, then one can use Picard iteration to show existence and uniqueness of solutions to the ODE, while when the coefficients are merely continuous, one can use Euler scheme to construct a family of piecewise linear approximate solutions and subtract a converging subsequence in the topology of continuous functions. Here, for SDE, we will use Euler scheme to construct a family of random functions and try to subtract convergence subsequence in the topology of weak convergence of probabilistic measures on continuous functions.

Step 1: approximation (Euler scheme).
For each n, we discretize the time by $t = 0, 1/2^n, \ldots$, and consider the process

$$X_0^{(n)} \sim \mu, \qquad X_t^{(n)} = X_{j/2^n}^{(n)} + b(X_{j/2^n}^{(n)})(t - j/2^n) + \sigma(X_{j/2^n}^{(n)})(W_t - W_{j/2^n}), \quad t \in (j/2^n, (j+1)/2^n].$$

Then $X_t^{(n)}$ solves the functional SDE

$$X_t^{(n)} = X_0^{(n)} + \int_0^t b^{(n)}(s, X^{(n)})\, ds + \sigma^{(n)}(s, X^{(n)})\, dW_s, \tag{10.12}$$

where

$$b^{(n)}(t, \mathbf{y}) = b\big(y(2^{-n}[2^n t])\big), \quad \sigma^{(n)}(t, \mathbf{y}) = \sigma\big(y(2^{-n}[2^n t])\big).$$

Step 2: estimate.
We need the following lemma.

Lemma 10.5 *Suppose that X solves the functional SDE*

$$X_t = X_0 + \int_0^t b(s, X)\, ds + \sigma(s, X)\, dW_s, \quad X_0 = \xi. \tag{10.13}$$

where for each $T > 0$, the coefficients b, σ satisfy

$$|b(t, \mathbf{y})|^2 + |\sigma(t, \mathbf{y})|^2 \le K_T(1 + \max_{0 \le s \le t} |y(s)|^2), \quad \forall 0 < t \le T,$$

for some $K > 0$.

Then for any $m \ge 1$, there exists $C = C(T, K)$ such that

$$\mathsf{E} \max_{0 \le s \le t} |X_s|^{2m} \le C\big(1 + \mathsf{E}|\xi|^{2m}\big)e^{Ct}, \quad 0 \le t \le T, \tag{10.14}$$

and

$$\mathsf{E}|X_t - X_s|^{2m} \le C(1 + \mathsf{E}|\xi|^{2m})(t - s)^m, \quad 0 \le s < t \le T. \tag{10.15}$$

We will need to use the martingale moment inequality by Burkholder–Davis–Gundy.

Theorem 10.6 (Burkholder–Davis–Gundy) *For any $m > 0$, there are universal constants c_m, C_m such that for any c.l.m. M and any stopping time T,*

$$c_m \mathsf{E}\langle M\rangle_T^m \le \mathsf{E}(M_T^*)^{2m} \le C_m \mathsf{E}\langle M\rangle_T^m, \tag{10.16}$$

where $M_t^* = \sup_{0 \le s \le t} |M_s|$ is the maximal process of M.

For the proof, see [KS98, Theorem 3.3.28].

Sketch of the proof of Lemma 10.5: (10.14) follows from applying Gronwall's inequality on the left-hand side. To set up the conditions for Gronwall's inequality, we first observe that

$$|X_t|^{2m} \le C\big(|\xi|^{2m} + \big[\int_0^t b(s, X)\, ds\big]^{2m} + \big[\int_0^t \sigma(s, X)\, dW_s\big]^{2m}\big).$$

Taking supremum over an interval $[0, t]$, only the supremum of the c.l.m.

$$M_t = \int_0^t \sigma(s, X)\, dW_s$$

needs special attention to close the inequality. Let $M_t^* = \sup_{0 \le s \le t} |M_t|$. Theorem 10.6 ensures that

$$\mathsf{E}(M_t^*)^{2m} \le C\mathsf{E}\langle M\rangle_t^m.$$

Combining this with

$$\mathsf{E}\langle M\rangle_t^m = \mathsf{E}\big|\int_0^t |\sigma(s, X)|^2\, ds\big|^m \le \mathsf{E}t^m \cdot C(1 + \max_{0 \le s \le t} |X_s|^2)^m,$$

the rest is routine.

(10.15) follows from applying a similar estimate on

$$|X_t - X_s|^{2m} \le C\big(\big[\int_s^t b(r, X)\, dr\big]^{2m} + \big[\int_s^t \sigma(r, X)\, dW_r\big]^{2m}\big).$$

and (10.14). □

Since we assume that b, σ are bounded and continuous, $b^{(n)}$ and $\sigma^{(n)}$ will satisfy the condition of Lemma 10.5. Noting μ has $2m$-th moment for some $m \ge 1$, by Lemma 10.5, for each $T > 0$, there exists $C = C_T$ such that

$$\mathsf{E}|\max_{0 \le s \le t} X_t|^{2m} \le C, \quad \mathsf{E}|X_t - X_s|^{2m} \le C(t - s)^m, \qquad 0 \le s < t \le T. \tag{10.17}$$

Step 3: extract convergence subsequence.

Let $\mathsf{P}^{(n)}$ be the law of $X^{(n)}$ on $\mathcal{C}[0, \infty)$. We recall the Prohorov Theorem.

Theorem 10.7 (Prohorov; see e.g. [Bil99]) *Let E be a metric space and μ_n be a sequence of probability measures on $(E, \mathcal{B}(E))$. Then μ_n have a convergence subsequence in the topology of weak convergence of probability measures if and only if μ_n is* tight, *that is, for every $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset E$ such that $\mu_n(K_\varepsilon^c) \le \varepsilon$ for all n.*

We also recall from real analysis that $F \subset \mathcal{C}[0,\infty)$ is pre-compact if and only if all functions $\mathbf{y} \in F$ are *uniformly bounded and equi-continuous* on every interval $[0,t]$.

The tightness of $\mathsf{P}^{(n)}$ will follow from the two conditions in (10.17): the first inequality implies that $X^{(n)}$ are uniformly bounded with high probability, the second inequality and argument similar to Theorem 2.7 implies equi-continuity holds with high probability. Hence, there exists P^* as a (subsequential) weak limit of $\mathsf{P}^{(n)}$. This means that

$$\int \Phi(\mathbf{y}) \mathsf{P}^{(n)}(d\mathbf{y}) \to \int \Phi(\mathbf{y})\ \mathsf{P}^*(d\mathbf{y}), \quad n \to \infty \tag{10.18}$$

for every bounded continuous functional $\Phi : \mathcal{C}[0,\infty) \to \mathbb{R}$.

Step 4: P^* solves the martingale problem.
First, let us verify that P^* gives the desired initial condition. Let $f \in \mathcal{C}_b(\mathbb{R})$. Then

$$\Phi_f(\mathbf{y}) := f\big(y(0)\big)$$

is a bounded continuous functional. Hence by (10.18),

$$\mathsf{E}^* f(y(0)) = \lim_{n\to\infty} \mathsf{E}^{(n)} f(y(0)) = \int_{\mathbb{R}} \mu(dr) \cdot f(r).$$

Since this holds for every $f \in \mathcal{C}_b(\mathbb{R})$, we have

$$\mathsf{P}^*\Big(\mathbf{y} \in \mathcal{C}[0,\infty) : y(0) \in \Gamma\Big) = \mu(\Gamma), \quad \Gamma \in \mathcal{B}(\mathbb{R}),$$

that is, P^* has the correct initial contidition.

Second we need to check that P^* solves the martingale problem, that is, for every $f \in \mathcal{C}_0^2(\mathbb{R})$ and $0 \le s < t$,

$$\mathsf{E}^*\Big[f\big(y(t)\big) - f\big(y(0)\big) - \int_s^t (\mathcal{L}f)\big(y(u)\big)\, du \mid \mathcal{B}_s\Big] = 0,$$

or equivalently, for every bounded continuous $\mathcal{B}_s$-measurable functional $g : \mathcal{C}[0,\infty) \to \mathbb{R}$,

$$\mathsf{E}^*\Big[f\big(y(t)\big) - f\big(y(0)\big) - \int_s^t (\mathcal{L}f)\big(y(u)\big)\, du\Big] g(\mathbf{y}) = 0, \tag{10.19}$$

Let

$$F_n(\mathbf{y}) = f\big(y(t)\big) - f\big(y(0)\big) - \int_s^t (\mathcal{L}_u^{(n)} f)\big(y(u)\big)\, du, \quad F(\mathbf{y}) = f\big(y(t)\big) - f\big(y(0)\big) - \int_s^t (\mathcal{L}f)\big(y(u)\big)\, du.$$

By triangle inequality,

$$|\mathsf{E}^{(n)} F_n(\mathbf{y}) g(\mathbf{y}) - \mathsf{E}^* F(\mathbf{y}) g(\mathbf{y})| \le |\mathsf{E}^{(n)} F_n(\mathbf{y}) g(\mathbf{y}) - \mathsf{E}^{(n)} F(\mathbf{y}) g(\mathbf{y})| + |\mathsf{E}^{(n)} F(\mathbf{y}) g(\mathbf{y}) - \mathsf{E}^* F(\mathbf{y}) g(\mathbf{y})|.$$

The second term goes to 0 by weak convergence $\mathsf{P}^{(n)} \to \mathsf{P}^*$. To control the first term, it suffices to show that $F_n(\mathbf{y}) \to F(\mathbf{y})$ uniformly on compact sets in $\mathcal{C}[0,\infty)$, which follows from the continuity of b and σ. Combining all these we can establish (10.19) and hence P^* indeed solves the martingale problem.

10.4 Uniqueness for martingale problem solution and strong Markov property

To simplify the discussion, we assume in this section that our coefficients b, σs b, σs b, σs b, σs b, σs b, σs b, σs b, σ does not depend on t. In the most general setting, the uniqueness of martingale problem solutions is usually guaranteed by the *existence* of corresponding PDE solutions.

To be more precise, consider the Cauchy problem

$$\begin{cases} \qquad \partial_t u = \mathcal{L}u, & (0,\infty)\times\mathbb{R}^d, \\ u(t=0,\cdot) = f \in \mathcal{C}_0^\infty(\mathbb{R}^d). & \end{cases} \tag{10.20}$$

We say that the Cauchy problem admits a solution if there is a function $u_f \in \mathcal{C}\big([0,\infty)\times\mathbb{R}^d\big) \cap \mathcal{C}^{1,2}\big((0,\infty)\times\mathbb{R}^d\big)$ solves (10.20), which is in addition bounded on $[0,T]\times\mathbb{R}^d$ for every $T>0$.

The solution of (10.20) exists under very mild condition on b,σ. For example, a sufficient condition is that the diffusion matrix $A(x) = (a_{ij}(x))$ is uniformly elliptic on compact sets, and b,σ are bounded and Borel-measurable. In $d=1,2$, this condition also implies uniqueness of solutions. Since we assume b,σ are merely measurable, it is weaker than the continuity condition we imposed in Section 10.3.

Proposition 10.8 *Assume that* (10.20) *has a solution. Then the one-dimensional marginal of the solution to the martingale problem is unique. Precisely, let* P^x *and* $\tilde{\mathsf{P}}^x$ *be two solutions to the martingale problem with initial condition* $x\in\mathbb{R}^d$*. Then for every* $t>0$ *and* $\Gamma\in\mathcal{B}(\mathbb{R})$*,*

$$\mathsf{P}^*\Big(y(t)\in\Gamma\Big) = \tilde{\mathsf{P}}^*\Big(y(t)\in\Gamma\Big).$$

Proof: Fix any $f\in\mathcal{C}_0^\infty(\mathbb{R}^d)$ and $T>0$. Let $g(t,x) := u_f(T-t,x)$. Then $\partial_t g + \mathcal{L}g = 0$ and $g(T,\cdot) = f$. Also g is bounded. This implies that $g\big(t,y(t)\big)$ is a martingale under both P^x and $\tilde{\mathsf{P}}^x$.

We have

$$\mathsf{E}^x f\big(y(T)\big) = \mathsf{E}^x g\big(T,y(T)\big) = \mathsf{E}^x g\big(0,y(0)\big) = g(0,x).$$

and similarly $\tilde{\mathsf{E}}^x f\big(y(T)\big) = g(0,x)$. Hence,

$$\mathsf{E}^x f\big(y(T)\big) = \tilde{\mathsf{E}}^x f\big(y(T)\big).$$

As this holds for every $f\in\mathcal{C}_0^\infty(\mathbb{R}^d)$, the marginal distribution of P^x and $\tilde{\mathsf{P}}^x$ is the same at time T. This completes the proof. □

Of course, it is not enough to have only one-dimensional marginals to agree. But before proving a similar statement for f.d.d., we will take a detour to talk about the strong Markov property of the martingale solution.

For $s\ge 0$, we define the *shift operator* θ_s on $\mathcal{C}[0,\infty)^d$ to be

$$\theta_s\mathbf{y} = \big(y(s+t)\big)_{t\ge 0}.$$

Note that θ_s is a bounded continuous functional, and $(s,\mathbf{y})\mapsto\theta_s(\mathbf{y})$ is jointly measurable in s and $\mathbf{y}$.

We will state a technical lemma, whose proof will be postponed to the end of this section.

Lemma 10.9 *Let* P *be a solution to the time-homogeneous martingale problem, i.e., for every* $0\le s<t$ *and* $f\in\mathcal{C}_0^2(\mathbb{R}^d)$

$$\mathsf{E}\Big[f\big(y(t)\big) - f\big(y(s)\big) - \int_s^t (\mathcal{L}f)\big(y(u)\big)\,du \mid \mathcal{B}_s\Big] = 0. \tag{10.21}$$

Let T *be a bounded stopping time and write* $\mathcal{G} = \mathcal{B}_T$*. Let* $Q_\omega(F) := Q(\omega;F) : \Omega\times\mathcal{B}\to[0,1]$ *be the regular conditional probability for* $\mathcal{B}$ *given* $\mathcal{G}$*.*

Then, there exists a P*-null set* $N \subset \mathcal{G}$ *s.t.*

$$\mathsf{P}_\omega := Q_\omega \circ \theta_T^{-1} \tag{10.22}$$

is a solution to (10.21) *and satisfies*

$$\mathsf{P}\Big(\mathbf{y} \in \mathcal{C}[0,\infty)^d : y(0) = x\Big) = 1 \tag{10.23}$$

with $x = \omega\big(T(\omega)\big)$*, for all* $\omega \notin N$*.*

Theorem 10.10 *The setting is the same as Proposition 10.8. Then* P^x *and* $\tilde{\mathsf{P}}^x$ *has the same f.d.d.*

Proof: For every $0 \le t_1 < t_2 < \cdots < t_n$, We will show that P^x and $\tilde{\mathsf{P}}^x$ agree on $\sigma(t_1, \dots, t_n)$. We will show this by induction on n. The base case $n = 1$ is given by Proposition 10.8.

Assuming that P^x and $\tilde{\mathsf{P}}^x$ agree on $\sigma(t_1, \cdots, t_{n-1}) =: \mathcal{G}$. Then by Lemma 10.9, there exists a P^x-null set $N \in \mathcal{G}$ such that $\mathsf{P}_{\mathbf{y}} := Q_{\mathbf{y}} \circ \theta_{t_{n-1}}^{-1}$ is a solution to (10.21) and satisfies (10.23) with $x = y(t_{n-1})$ for all $\mathbf{y} \notin N$. There is also a $\tilde{\mathsf{P}}^x$-null set $\tilde{N} \in \mathcal{G}$ and a similarly defined $\tilde{\mathsf{P}}_{\mathbf{y}}$ on $y \notin \tilde{N}$. Note that by induction hypothesis, P^x and $\tilde{\mathsf{P}}^x$ agrees on $\mathcal{G}$, so N, $\tilde{N}$ are null sets under both P^x and $\tilde{\mathsf{P}}^x$.

For any $A \in \mathcal{B}(\mathbb{R}^{d(n-1)})$ and $B \in \mathcal{B}(\mathbb{R}^d)$, we have

$$\begin{aligned}
&\mathsf{P}^x\Big(\big(y(t_1), \dots, y(t_{n-1})\big) \in A,\ y(t_n) \in B\Big) \\
&= \int_A \mathsf{P}^x \circ \pi_{n-1}^{-1}(dy_1 \cdots dy_{t_{n-1}}) \cdot \mathsf{P}_{\mathbf{y}}(\omega : \omega(t_n - t_{n-1}) \in B) \\
&= \int_A \mathsf{P}^x \circ \pi_{n-1}^{-1}(dy_1 \cdots dy_{t_{n-1}}) \cdot \tilde{\mathsf{P}}_{\mathbf{y}}(\omega : \omega(t_n - t_{n-1}) \in B) \\
&= \int_A \tilde{\mathsf{P}}^x \circ \pi_{n-1}^{-1}(dy_1 \cdots dy_{t_{n-1}}) \cdot \tilde{\mathsf{P}}_{\mathbf{y}}(\omega : \omega(t_n - t_{n-1}) \in B) \\
&= \tilde{\mathsf{P}}^x\Big(\big(y(t_1), \dots, y(t_{n-1})\big) \in A,\ y(t_n) \in B\Big).
\end{aligned}$$

Here, the second equality is due to the fact that for $\mathbf{y} \in (N \cup \tilde{N})^c$ (i.e., P^x-a.e. $\mathbf{y}$), $\mathsf{P}_{\mathbf{y}}$ and $\tilde{\mathsf{P}}_{\mathbf{y}}$ are martingale solutions satisfying (10.23) with $x = y(t_{n-1})$, and hence by Proposition 10.8, they have the same one-dimensional marginals. The third equality is the induction hypothesis. Therefore, P^x and $\tilde{\mathsf{P}}^x$ agrees on $\sigma(t_1, \cdots, t_n)$, and this completes the induction step. □

Now with uniqueness of solution to the martingale problem, we can use the notation P^x to denote the unique solution to (10.21) such that (10.23) holds.

The following strong Markov property holds.

Theorem 10.11 *Let* T *be a bounded stopping time and* $x \in \mathbb{R}^d$*. Then for every* $F \in \mathcal{B}(\mathcal{C}[0,\infty)^d)$*,*

$$\mathsf{P}^x\Big[\theta_T^{-1} F \mid \mathcal{B}_T\Big](\omega) = \mathsf{P}^y(F)\Big|_{y=\omega(T)}, \quad \mathsf{P}^x\text{-a.e. } \omega.$$

Proof: With the notation in Lemma 10.9,

$$\mathsf{P}^x\Big[\theta_T^{-1} F \mid \mathcal{B}_T\Big](\omega) = Q(\omega; \theta_T^{-1} F) = \mathsf{P}_\omega(F) = \mathsf{P}^{\omega(T(\omega))}(F)$$

as desired. □

Proof of Lemma 10.9:

Since $Q_\omega(\cdot)$ is a regular condition probability, for every $\Gamma \in \mathcal{G}$, for P-a.e. ω it holds

$$Q_\omega(\Gamma) = \mathbb{1}_\Gamma(\omega). \tag{10.24}$$

Since $\mathcal{G} \subset \mathcal{B}(\mathcal{C}[0,\infty)^d)$ is countably determined, we can find a common exceptional set N such that (10.24) holds for all $\omega \notin \Gamma$. For $u \in \mathbb{R}^d$, let $\Gamma_u = \{\mathbf{y} \in \Omega : y\big(T(\mathbf{y})\big) = u\} \in \mathcal{G}$. Then

$$\mathsf{P}_\omega\big(\mathbf{y} \in \Omega : y(0) = \omega\big(T(\omega)\big)\big) = Q_\omega(\omega; \Gamma_{\omega(T(\omega))}) = \mathbb{1}_{\Gamma_{\omega(T(\omega))}}(\omega) = 1$$

for all $\omega \notin N$. This verifies (10.23) with $x = \omega\big(T(\omega)\big)$.

Next we will show that P_ω solves (10.21). Let $f \in \mathcal{C}_0^2[0,\infty)^d$ and $F \in \mathcal{B}_s$. Let

$$Z(\mathbf{y}) := f\big(y(t)\big) - f\big(y(s)\big) - \int_s^t (\mathcal{L}f)\big(y(u)\big)\, du.$$

For $\omega \notin N$, we have the a.s. equalities

$$\begin{aligned}
&\int_\Omega Z(\mathbf{y}) \mathbb{1}_F(\mathbf{y}) \mathsf{P}_\omega(\mathbf{y}) \\
&= \int_\Omega Z(\theta_{T(\mathbf{y})}\mathbf{y}) \mathbb{1}_F(\mathbf{y})(\theta_{T(\mathbf{y})}\mathbf{y}) Q(\omega; d\mathbf{y}) && \text{(definition of } \mathsf{P}_\omega) \\
&= \mathsf{E}\Big[(Z \circ \theta_T) \cdot \mathbb{1}_{\theta_T^{-1}F} \mid \mathcal{G}\Big](\omega) && \text{(definition of reg. cond. prob.)} \\
&= \mathsf{E}\Big[\mathsf{E}[Z \circ \theta_T \mid \mathcal{B}_{T+s}] \cdot \mathbb{1}_{\theta_T^{-1}F} \mid \mathcal{G}\Big](\omega) && (F \in \mathcal{B}_s \Rightarrow \theta_T^{-1}F \in \mathcal{B}_{T+s}) \\
&= \mathsf{E}[0 \cdot \mathbb{1}_{\theta_T^{-1}F} \mid \mathcal{G}](\omega) && \text{(OST applied to } M_t^f \text{ at } s+T, t+T) \\
&= 0.
\end{aligned}$$

The previous computation shows that there exist a null set $N(s,t,f,F)$ such that

$$\int_F Z(\mathbf{y}) \mathsf{P}_\omega(d\mathbf{y}) = 0, \quad \forall \omega \notin N(s,t,f,F).$$

Since $\mathcal{G}$ is countably determined, and $Z(\mathbf{y})$ is continuous in s and t, we can find a null set $N(f)$ such that M_t^f is a martingale under P_ω for all $\omega \in N^c$.

Note that Theorem 10.3 and the discussion that follows implies that, if M_t^f is martingales for some countable family of functions f (namely, those $\mathcal{C}_0^2$-approximations of $x_i, x_i x_j$), then it is a martingales for all $f \in \mathcal{C}_0^2$. This shows that P is indeed a solution to the martingale problem. □

11 Diffussion and PDEs

11.1 Representation of PDEs solutions

11.1.1 Elliptic equation

In this section we assume that b, σ are continuous and independent of t, so that $\mathcal{L}_t = \mathcal{L}$. We consider the following *Dirichlet problem*

$$\begin{cases} -(\mathcal{L}u)(x) = g(x) - k(x)u(x), & x \in D, \\ \qquad u|_{\partial D} f, \end{cases} \tag{11.1}$$

where D is a bounded open domain, $k \geq 0$, g are continuous functions on $\bar{D}$ and f is a continuous function on ∂D. A (classical) solution to (11.1) is a function $u \in \mathcal{C}^2(D) \cap \mathcal{C}(\bar{D})$ that satisfies (11.1).

Theorem 11.1 *Let u be a solution to (11.1). Let X be a solution to the SDE with generator $\mathcal{L}$ and*

$$\tau_D = \inf\{t \geq 0 : X_t \notin D\}$$

be the exit time of D. If $\mathsf{E}^x \tau_D < \infty$, then

$$u(x) = \mathsf{E}^x \Big[f(X_{\tau_D}) e^{-\int_0^{\tau_D} k(X_s)\, ds} + \int_0^{\tau_D} g(X_t) e^{-\int_0^t k(X_s)\, ds}\, dt \Big].$$

Proof: Let $U_t = u(X_t)$, $\mathcal{E}_t = e^{-\int_0^t k(X_s)\, ds}$. Then

$$dU_t = \text{m.t.} + (\mathcal{L}u)(X_t)\, dt, \quad d\mathcal{E}_t = \mathcal{E}_t \cdot (-k(X_t))\, dt.$$

Hence

$$d(U_t \mathcal{E}_t) = \text{m.t.} + \mathcal{E}_t \cdot \Big[\mathcal{L}u - ku \Big]\, dt = \text{m.t.} - \mathcal{E}_t g(X_t)\, dt.$$

Therefore,

$$Y_t = u(X_t) e^{-\int_0^t k(X_s)\, ds} + \int_0^t g(X_s) e^{-\int_0^s k(X_\theta)\, d\theta}\, ds$$

is a c.l.m.

We have

$$|Y_{t \wedge \tau_D}| \leq \sup_{\bar{D}} |u| + |\tau_D| \cdot \sup_{\bar{D}} |g|.$$

Since $\mathsf{E}^x \tau_D < \infty$, $(Y_{t \wedge \tau_D})_{t \geq 0}$ is a uniformly integrable martingale under P^x. By Optional Sampling Theorem, we have

$$u(x) = Y_0 = \mathsf{E}^x Y_{\tau_D} = \mathsf{E}^x \Big[f(X_{\tau_D}) e^{-\int_0^{\tau_D} k(X_s)\, ds} + \int_0^{\tau_D} g(X_t) e^{-\int_0^t k(X_s)\, ds}\, dt \Big]$$

as desired. □

Remark 11.1 If $g \equiv 0$, then we only need $\mathsf{P}^x(\tau_D < \infty) = 1$.

One may ask when τ_D has finite expectation. A sufficient condition is that X diffuses in at least one direction so that the exit time is not larger that that of a one-dimensional Brownian motion from a bounded set. A precise statement is the following.

Lemma 11.2 *Suppose that*

$$L = \sup_{x \in D} |x_1| < \infty, \quad a = \min_{x \in \bar{D}} a_{11}(x) > 0, \quad b = \max_{x \in \bar{D}} |b_1(x)| < \infty.$$

Then $\mathsf{E}^x \tau_D < \infty$.

Proof: We consider a test function $h(x) = \mu e^{-\nu x_1} \in \mathcal{C}^\infty(\mathbb{R}^d)$ with μ,ν to be determined. Then under the assumptions,

$$\mathcal{L}h = vb_1(x)h(x) + \frac{1}{2}\nu^2 a_{11}(x) h(x) \geq h(x)\big[-\nu L + \frac{1}{2}\nu^2 a \big] \geq \mu e^{-\nu L} \nu (\nu a/2 - L).$$

We first choose $\nu > 2L/a$ so that the right hand side is positive, then choose L large enough so that $\mathcal{L}h \geq 1$. Then,

$$M_t^h = h(X_t) - \int_0^t (\mathcal{L}h)(X_s)\, ds$$

is a bounded martingale and

$$\mathsf{E}^x h(X_{t \wedge \tau_D}) - h(x) = \mathsf{E}^x \int_0^{t \wedge \tau_D} (\mathcal{L}h)(X_s)\, ds \geq \mathsf{E}^x (t \wedge \tau_D).$$

By sending $t \uparrow \infty$, we see that $\mathsf{E}^x \tau_D \leq 2 \sup_D |h| \leq 2e^{\nu L} < \infty$. □

11.1.2 Parabolic equation; Feynman–Kac

In this section, we assume that b, σ are continuous and satisfy the lienar growth condition (in x). We consider the Cauchy problem

$$\begin{cases} \partial_t u = \mathcal{L}_t u - ku + g, & (t,x) \in (0,\infty) \times \mathbb{R}^d \\ u|_{t=0} = f, & x \in \mathbb{R}^d, \end{cases} \tag{11.2}$$

where $k \geq 0$, f, g are continuous functions on their domains. Moreover, g satisfy either

$$|g(t,x)| \leq L(1 + |x|^{2\lambda}) \tag{11.3}$$

for some $L > 0$, $\lambda \geq 1$ or

$$g \geq 0. \tag{11.4}$$

Theorem 11.3 *If $u \in \mathcal{C}([0,T] \times \mathbb{R}^d) \cap \mathcal{C}^{1,2}((0,T] \times \mathbb{R}^d)$ solves (11.2), and for some $\mu > 1$, $M > 0$,*

$$\max_{0 \leq t \leq T} |u(t,x)| \leq M(1 + |x|^{2\mu}).$$

Then

$$u(T,x) = \mathsf{E}^x \Big[f(X_T) e^{-\int_0^T k(T-s,X_s)\,ds} + \int_0^T g(T-t,X_t) e^{-\int_0^t k(T-s,X_s)\,ds}\,dt \Big].$$

Proof: By a similar computation to the proof of Theorem 11.1, one can show that

$$Y_t = u(T-t,X_t) e^{-\int_0^t k(T-s,X_s)\,ds} + \int_0^t g(T-s,X_s) e^{-\int_0^s k(T-\theta,X_\theta)\,d\theta}\,ds$$

is a c.l.m. on $t \in [0,T)$ under P^x, since $(Y_{t \wedge \tau_n})$ is a martingale for

$$\tau_n = \inf\{t \geq 0 : |X_t| \geq n\}.$$

(The martingale part involves integrals like $\int \partial_i u dX$, but $\partial_i u(t,x)$ could be unbounded near $t = 0$.)
We need to show $\mathsf{E}Y_0 = \mathsf{E}Y_T$

First let us Assume (11.3). By the growth condition on u and g, we have

$$|Y_t| \leq C(1 + \sup_{0 \leq s \leq T} |X_s|^{2\mu}) + CT \cdot (1 + \sup_{0 \leq s \leq T} |X_s|^{2\lambda}), \quad t \in [0,T). \tag{11.5}$$

By Lemma 10.5, the right-hand side is integrable, hence $(Y_t)_{0 \leq t < T}$ is a u.i. martingale and $\mathsf{E}Y_0 = \mathsf{E}Y_T$.

Next let us assume (11.4). For every $t < T$ and τ_n, by Optional Sampling Theorem we have

$$\mathsf{E}Y_0 = \mathsf{E}Y_{t \wedge \tau_n} = \mathsf{E}^x \Big[u(T - t \wedge \tau_n, X_{t \wedge \tau_n}) e^{-\int_0^{t \wedge \tau_n} k(T-s,X_s)\,ds} + \int_0^{t \wedge \tau_n} g(T-t,X_t) e^{-\int_0^t k(T-s,X_s)\,ds}\,dt \Big].$$

We first let $t \uparrow T$ and then $\tau_n \to \infty$. The first expectation converges due to continuity of u,

$$|u(T - t \wedge \tau_n, X_{t \wedge \tau_n}) e^{-\int_0^{t \wedge \tau_n} k(T-s,X_s)\,ds}| \leq C(1 + \sup_{0 \leq s \leq T} |X_s|^{2\mu})$$

and Lemma 10.5. The second expectation converges due to $g \geq 0$ and Monotone Convergence Theorem. $\square$

Remark 11.2 Theorem 11.3 can be viewed as a special case of Theorem 11.1

11.2 Harmonic functions

We say that a function u is *harmonic* in a domain $D \subset \mathbb{R}^d$ if $u \in \mathcal{C}^2(D)$ and $\Delta u = 0$ in D.

We will use

$$\fint_A f(x)\,dx = \frac{1}{|A|}\int_A f(x)\,dx$$

to denote the average of f over a set A (w.r.t. to the Lebesgue measure). Harmonic functions enjoy the celebrated *mean-value property below.*

Theorem 11.4 *A function u is harmonic in D if and only if for all ball $B_r(x) \subset D$,*

$$u(x) = \fint_{\partial B_r(x)} u(y)\,dy = \fint_{B_r(x)} u(y)\,dy. \tag{11.6}$$

Proof: We only illustrate the "$\Longrightarrow$" direction as it has a simple probabilistic proof. The other direction can only be proved analytically and can be found in any undergraduate PDE text.

Let B be a Brownian motion and let τ be its exit time of D. Then by Theorem 11.1,

$$\mathsf{E}u(x) = \mathsf{E}^x u(B_\tau) = \int u(y)\,\mu(dy),$$

where $\mu(dy) = \mathsf{P}^x(B_\tau \in dy)$ is the *exit* measure of B. Clearly, μ is a measure on $\partial B_r(x)$; also μ must be rotationally invariant since both the set D and the process B are. The only such measure on $\partial B_r(x)$ is the uniform measure, and hence

$$\int u(y)\,\mu(dy) = \fint_{\partial B_r(x)} u(y)\,dy.$$

The second equality of (11.6) follows from the first one since for some dimensional constant d,

$$\int_{B_r(x)} u(y)\,dy = \int_0^r dr' \cdot c_d (r')^{d-1} \cdot \fint_{\partial B_{r'}(x)} u(y)\,dy = u(x)\cdot|B_r(x)|.$$

$\square$

Now we consider the Dirichlet problem

$$\begin{cases} \Delta u(x) = 0, & x \in D, \\ \lim\limits_{D \ni x \to y} u(y) = f(y), & y \in \partial\Omega, \end{cases} \tag{11.7}$$

where f is a continuous function on D and $u \in \mathcal{C}^2(D)$. We do not assume that D is a bounded domain; if it is, then u must also be in $\mathcal{C}(\bar{D})$ and we in the same situation as Section 11.1.1. Since we do not assume the boundedness of the domain, u can also be unbounded over D.

Proposition 11.5 *Assume u is bounded and $\mathsf{P}^x(\tau_D < \infty) = 1$ for all $x \in D$. Then, any bounded solution of* (11.7) *can be represented by*

$$u(x) = \mathsf{E}^x f(B_{\tau_D}). \tag{11.8}$$

Proof: Let

$$D_n = \{x : \mathrm{dist}(x, \partial D) \ge \frac{1}{n}\}, \quad B_n = \{x : |x| \le n\}.$$

Then for every n, $|\partial_i u(B_t)|$ is bounded when $t \le \tau_{D_n} \wedge \tau_{B_n}$, and hence

$$u(B_{t\wedge\tau_{D_n}\wedge\tau_{B_n}}) - u(x) = \sum_{i=1}^{d} \int_0^{t\wedge\tau_{D_n}\wedge\tau_{B_n}} \partial_i u(B_s)\, dB_s$$

is a martingale. Hence,

$$u(x) = \mathsf{E}^x u(B_{t\wedge\tau_{D_n}\wedge\tau_{B_n}})$$

for every t and n.

Since $\tau_{D_n} < \tau_D < \infty$,

$$u(B_{t\wedge\tau_{D_n}\wedge\tau_{B_n}}) \to u(B_{\tau_{D_n}\wedge\tau_{B_n}}), \quad t \to \infty.$$

Since $\tau_{B_n} \to \infty$ a.s., and $\tau_{D_n} \uparrow \tau_D < \infty$,

$$u(B_{\tau_{D_n}\wedge\tau_{B_n}}) \to f(B_{\tau_D}), \quad n \to \infty$$

by the continuity of u at the boundary. Passing the limit under the expectation is justified by the assumption that u is bounded and the Dominated Convergence Theorem. □

Any function taking the form (11.8) is also harmonic under minimum assumption on f.

Proposition 11.6 *If $\mathsf{E}^x|f(B_{\tau_D})| < \infty$ for every $x \in D$, then u given by* (11.8) *is harmonic in D.*

Proof: By strong Markov property, for $B_r(x) \subset D$, we have

$$u(x) = \mathsf{E}^x f(B_{\tau_D}) = \mathsf{E}^x\Big(\mathsf{E}^y\big[f(B_{\tau_D}) \mid B_{\tau_{B_r(x)}} = y\big]\Big) = \mathsf{E}^x u(B_{\tau_{B_r(x)}}).$$

Therefore, u has the mean-value property in D and hence by Theorem 11.4, it is harmonic in D. □

Let $f : \partial D \to \mathbb{R}$ be a bounded, measurable function. Assume that f is continuous at $a \in \partial D$. The natural question to when the following limit holds true:

$$\lim_{D \ni x \to a} \mathsf{E}^x f(B_{\tau_D}) = f(a). \tag{11.9}$$

It turns out the validity of the limit only depends on the geometry of ∂D.

Definition 11.1 *Let $\sigma_D = \inf\{t > 0 : B_t \notin D\}$. A point $a \in \partial D$ is* regular *if $\mathsf{P}^a(\sigma_D = 0) = 1$, and* irregular *if $\mathsf{P}^a(\sigma_D = 0) = 0$.*

Remark 11.3 Note that $\{\sigma_D = 0\} \in \mathcal{F}^B_{0+}$. So by Blumenthal's 0-1 law (Theorem 3.5), $\mathsf{P}^a(\sigma_D = 0) \in \{0, 1\}$. So any boundary point is either regular or irregular.

Let us discuss some examples of regular/irregular points.

Example 11.4 When $d = 1$, all boundary points are regular. This follows from the fact that P^a-a.s., (see also (3.12))

$$\sup_{0\le t\le\varepsilon}(B_t - a) > 0 > \inf_{0\le t\le\varepsilon}(B_t - a), \quad \forall\varepsilon > 0.$$

Example 11.5 When $d \ge 2$, an isolated boundary point is irregular, as the following example shows.

Consider the *punctured disk* $D = \{x : 0 < |x| < 1\} \subset \mathbb{R}^d$, $d \ge 2$ and $a = 0 \in \partial D$. Since d-dimensional Brownian motion, $d \ge 2$, is point-transient, i.e., the probability of hitting every fixed point is 0, we see that

$$\mathsf{P}^0(\sigma_D = 0) \le \mathsf{P}^0(\exists t > 0 : B_t = 0) = 0.$$

Example 11.6 When $d \geq 3$, irregular points can be found at *cusp point*, which is connected to D^c via a very "thin" tunnel behaving like $|r|^\alpha$, $\alpha < 1$. This is a small relaxation from the isolation condition, but does not satisfy the well-known "exterior cone condition" (see).

We will give the example of *Lebesgue's Thorn* in dimension $d = 3$. The domain will be rotationally symmetric around the x_1-axis. Let

$$E = \{|x_1| < 1, x_2^2 + x_3^2 < 1\}, \quad F_n = F_n^{\varepsilon_n} = \{2^{-n} \leq x_1 \leq 2^{-n+1},\ x_2^2 + x_3^2 \leq \varepsilon_n^2\}$$

where ε_n are sufficiently small numbers to be determined. We set $D = E \setminus \Big(\bigcup_{n=1}^{\infty} F_n\Big)$ and will show that when ε_n is suitably chosen, then $a = (0,0,0)$ is an irregular point. Here, a is connected to D^c through the "tunnel" $\bigcup_{n=1}^{\infty} F_n$, and by choosing ε_n small we enforce the cusp-like behavior.

We have

$$\mathsf{P}^0(\sigma_D = 0) \leq \mathsf{P}^0(B_t \in F_n,\ \text{for some } n \geq 1) \leq \sum_{n=1}^{\infty} \mathsf{P}^0\big(\exists t > 0,\ B_t \in F_n\big). \tag{11.10}$$

Note that by point-transient of Brownian motion in two dimension,

$$\lim_{\varepsilon \downarrow 0} \mathsf{P}^0(\exists t > 0:\ B_t \in F_n^{\varepsilon_n}) = \mathsf{P}^0\big(\exists t > 0 : B_t \in F_n^0\big) = \mathsf{P}^0\big(\exists t > 0,\ (B_t^{(2)}, B_t^{(3)}) = (0,0)\big) = 0.$$

Hence, by choosing ε_n small, we can have

$$\mathsf{P}^0\big(\exists t > 0,\ B_t \in F_n^{\varepsilon_n}\big) \leq 3^{-n}.$$

Combining with (11.10) we have $\mathsf{P}^0(\sigma_D = 0) < 1$, and hence $a = 0$ is irregular (remark after Definition 11.1).

Theorem 11.7 *Let $d \geq 2$ and $a \in \partial D$. The following statements are equivalent:*

1) (11.9) *holds;*

2) a is regular;

3) $\forall \varepsilon > 0$,

$$\lim_{D \ni x \to a} \mathsf{P}^x(\tau_D > \varepsilon) = 0. \tag{11.11}$$

Proof: Here we will only show Item 2 ⇒ Item 3 ⇒ Item 1. The implication Item 1 ⇒ Item 2 requires an explicit construction of counter-example; we refer to [KS98, Theorem 4.2.12] for the complete proof.

Without loss of generality we assume $a = 0$.

Item 2 ⇒ Item 3. We have

$$\begin{aligned}
\lim_{D \ni x \to 0} \mathsf{P}^x\big(\tau_D > \varepsilon\big) &\leq \limsup_{D \ni x \to 0} \mathsf{P}^x\big(\sigma_D > \varepsilon\big) && (\tau_D \leq \sigma_D \text{ by definition}) \\
&\leq \limsup_{D \ni x \to 0} \mathsf{P}^x\big(B_t \in D;\ 0 < t \leq \varepsilon\big) \\
&\leq limsup_{D \ni x \to 0} \mathsf{P}^x\big(B_t \leq D;\ \delta \leq t \leq \varepsilon\big) && (\text{choose any } \delta \leq \varepsilon) \\
&= \limsup_{D \ni x \to 0} \int \mathsf{P}^x\big(B_\delta \in dy\big) \mathsf{P}^y\big(\tau_D > \varepsilon - \delta\big).
\end{aligned}$$

Note that although $\mathsf{P}^y(\tau_D > \varepsilon - \delta)$ is merely measurable, since the transition probability $\mathsf{P}^x\big(B_\delta \in \cdot\big)$ is nice, the integral on the last line is in fact continuous in x, and hence we have

$$\lim_{D \ni x \to 0} \mathsf{P}^x\big(\tau_D > \varepsilon\big) \leq \mathsf{P}^0(B_t \in D;\ \delta \leq t \leq \varepsilon).$$

By taking $\delta \downarrow 0$, the right hand side converges to $\mathsf{P}^0(\sigma_D \geq \varepsilon)$ and is zero since 0 is regular. This proves (11.11).

Item 3 $\Rightarrow$ Item 1. Without loss of generality we can assume that f is bounded.

For any $r > 0$, we have

$$\mathsf{E}^x|f(B_{\tau_D}) - f(0)| \leq \mathsf{E}^x|f(B_{\tau_D}) - f(0)|\mathbb{1}_{\{|B_{\tau_D}|\leq r\}} + \|f\|_\infty \mathsf{P}^x(|B_{\tau_D}| > r)$$

As $r \downarrow 0$, the first term goes to 0 by the continuity of f at $x = 0$. For the second term, we have

$$\begin{aligned}
&\lim_{D\ni x\to 0} \mathsf{P}^x\big(|B_{\tau_D}| > r\big) \\
&\leq \limsup_{D\ni x\to 0} \mathsf{P}^x\big(\tau_D < \varepsilon,\ |B_{\tau_D}| > r\big) + \limsup_{D\ni x\to 0} \mathsf{P}^x(\tau_D > \varepsilon) && \text{(choose any } \varepsilon > 0\text{)} \\
&\leq \limsup_{D\ni x\to 0} \mathsf{P}^x\big(\sup_{0\leq t\leq\varepsilon} |B_t - B_0| \geq r/2\big) && \text{(the second term is 0 by (11.11))} \\
&= \mathsf{P}^{x=0}(\sup_{0\leq t\leq\varepsilon} |B_t| \geq r/2).
\end{aligned}$$

The probability in the last but one line does not depend on x since the law of Brownian motion is translational invariant. Clearly, the last line goes to 0 as $\varepsilon \to 0$. Combining all these we finish the proof. $\square$

Let $y \neq 0$ be a direction and $\theta \in (0, \pi/2]$. We define the cone

$$\mathrm{Co}(y, \theta) = \{x \in \mathbb{R}^d : \mathrm{angle}\langle x, y\rangle \leq \theta\}.$$

Definition 11.2 *A point $a \in \partial D$ satisfies the* exterior cone condition *if $a + \mathrm{Co}(y, \theta) \subset \mathbb{R}^d \setminus D$ for some $y \neq 0$ and $\theta \in (0, \pi/2]$.*

Theorem 11.8 *If a satisfies the exterior cone condition, then a is regular.*

Proof: Without loss of generality, set $a = 0$. We have

$$\mathsf{P}^0(\sigma_D \leq t) \geq \mathsf{P}^0(B_t \in \mathrm{Co}(y, \theta)).$$

But $\lambda B_{\lambda^{-2}t} \stackrel{d}{=} B_t$ and $\mathrm{Co}(y, \theta)$ is invariant under dilation, so

$$\mathsf{P}^0(B_t \in \mathrm{Co}(y, \theta)) = \mathsf{P}^0(B_1 \in \mathrm{Co}(y, \theta)) > 0.$$

Hence 0 is regular. $\square$

11.3 Some computations about hitting time

In this section we will show some examples of using PDE to compute $\mathsf{P}^x(\tau_D < \infty)$ and $\mathsf{E}^x\tau_D$ for Brownian motion.

The key observation is the following.

Proposition 11.9 *Let D be a set and $u(x) = \mathsf{P}^x(\tau_D < \infty)$. Then $h(x)$ is a harmonic function in D.*

Proof: By strong Markov property, for $B_r(x) \subset D$,

$$u(x) = \mathsf{P}^x(\tau_D < \infty) = \mathsf{E}^x\mathsf{P}^y\Big(\tau_D < \infty \mid B_{\tau_{B_r(x)}} = y\Big) = \mathsf{E}^x u(B_{\tau_{B_r(x)}}).$$

Hence, h has the mean value property and by Theorem 11.4 it is harmonic in D. $\square$

Let B be Brownian motion in d dimension. For $D = B_r(0)$ and $|x| > r$, let us compute $u(x) = \mathsf{P}^x(\tau_D < \infty)$. By Proposition 11.9, u is a classical solution to the PDE

$$\begin{cases} \Delta u(x) = 0, & |x| > r, \\ u(x) = 1, & |x| = r. \end{cases} \tag{11.12}$$

However, as $\{x : |x| > r\}$ is now an unbounded domain, the solution to (11.12) is no longer unique. An obvious solution to (11.12) is $u_1(x) \equiv 1$. Another solution is the so-called *spherically symmetric harmonic function*, given by

$$u_2(x) = \begin{cases} \dfrac{|x|}{r}, & d = 1, \\ \dfrac{\log |x|}{\log r}, & d = 2, \\ \dfrac{|x|^{2-d}}{r^{2-d}}, & d \geq 3. \end{cases} \tag{11.13}$$

Clearly, $u(x) = \mathsf{P}^x(\tau_D < \infty)$ is a bounded solution since probability is bounded by 1. In fact one can show that under the additional assumption of boundedness, for $d = 1, 2$ the solution to (11.12) is unique, i.e., $u(x) = u_1(x) \equiv 1$. This means that in dimensions 1 and 2, Brownian motion is neighborhood-recurrent, i.e., it will hit any open ball almost surely.

However, when $d \geq 3$ (11.13) also gives another bounded solution to (11.12), and in fact it is the correct form of $\mathsf{P}^x(\tau_D < \infty)$. To prove this, we start with a modified problem. For $r \leq |x| \leq R$, let us consider $\tilde{u}_{r,R}(x) = \tilde{u}(x) = \mathsf{P}^x(\tau_{B_r} < \tau_{B_R})$. The function $\tilde{u}$ is the solution to the Dirichlet problem

$$\begin{cases} \Delta u(x) = 0, & r < |x| < R, \\ u(x) = 0, & |x| = R, \\ u(x) = 1, & |x| = r. \end{cases} \tag{11.14}$$

This is a special case of (11.7) with

$$f(x) = \mathbb{1}_{|x|=r}(x). \tag{11.15}$$

Clearly all points in $\{x : |x| = r, R\}$ are regular, and f in (11.15) is continuous at every point of the boundary. Hence, the solution to (11.14) is unique, and is given by

$$u(x) = \mathsf{E}^x f(B_{\tau_{B_r} \wedge \tau_{B_R}}) = \mathsf{E}^x \mathbb{1}_{\tau_{B_r} < \tau_{B_R}}$$

as desired.

Now using (11.13) we can easily write down solutions to (11.14) by performing a linear transform:

$$\tilde{u}(x) = \begin{cases} \dfrac{R - |x|}{R - r}, & d = 1, \\ \dfrac{\log |x| - \log R}{\log r - \log R}, & d = 2, \\ \dfrac{R^{2-d} - |x|^{2-d}}{R^{2-d} - r^{2-d}}, & d \geq 3. \end{cases}$$

By uniqueness, now we know that it is the only solution to (11.14).

Now we can compute $\mathsf{P}^x(\tau_{B_r} < \infty)$ by sending $R \uparrow \infty$. Note that we use the fact that

$$\lim_{R \uparrow \infty} \tau_{B_R} = \infty, \quad \text{a.s.}$$

since paths of Brownian motion is continuous.

We have

$$\mathsf{P}^x(\tau_{B_r} < \infty) = \lim_{R\to\infty} \mathsf{P}^x(\tau_{B_r} < \tau_{B_R}) = \begin{cases} 1, & d = 1, 2, \\ \dfrac{|x|^{2-d}}{r^{2-d}}, & d \geq 3. \end{cases}$$

Hence, for $d \geq 3$, the Brownian motion is not neighborhood recurrent.

When $d = 2$, we can also let $r \to 0$ first to obtain

$$\mathsf{P}^x(\tau_0 < \tau_{B_R}) = \lim_{r\downarrow 0} \frac{\log|x| - \log R}{\log|r| - \log R} = 0.$$

Then by sending $R \uparrow \infty$ we obatin $\mathsf{P}^x(\tau_0 < \infty) = 0$. This means that two-dimensional Brownian motion is point-transient.

In general, if X is a diffusion with generator $\mathcal{L}$, then $u(x) = \mathsf{P}^x(\tau_{D_1} < \tau_{D_2})$ is a $\mathcal{L}$-harmonic function (i.e., $\mathcal{L}u = 0$) with suitable boundary condition. However, such PDEs are usually difficult to solve. A manageable case is when the PDE becomes an ODE in dimension one.

Let $X_t = B_t + \mu t$, $\mu > 0$ be the Brownian motion with drift. Then $\mathcal{L} = \frac{1}{2}\partial_{xx} + \mu\partial_x$. We want to compute $\mathsf{P}^x(\tau_0 < \infty)$ using PDE/ODE method.

For $R > 0$, let $u_R(x) = \mathsf{P}^x(\tau_0 < \tau_R)$, Then $u = u_R(x)$ solves

$$\begin{cases} \frac{1}{2}\partial_{xx}u(x) + \mu\partial_x u(x) = 0, \quad x \in (0, R), \\ u(0) = 0, u(R) = 0. \end{cases}$$

This is second-order linear ODE with constant coefficients. The two roots of the characteristic function $\frac{1}{2}\lambda^2 + \mu\lambda = 0$ are $\lambda = 0, -2\mu$, so any solution can be written as $c_1 + c_2 e^{-2\mu x}$. With some effort we can find

$$\mathsf{P}^x(\tau_0 < \tau_R) = u_R(x) = \frac{e^{-2\mu x} - e^{-2\mu R}}{1 - e^{-2\mu R}}.$$

By sending $R \to \infty$, we see that $\mathsf{P}^x(\tau_0 < \infty) = e^{-2\mu R}$.

For the next example, we will compute $\mathsf{E}^x \tau_0 \wedge \tau_R$ for 1d Brownian motion (without drift). The idea to solve the ODE

$$\begin{cases} \frac{1}{2}\partial_{xx}u(x) = -1, \quad x \in (0, R), \\ u(0) = u(R) = 0. \end{cases}$$

By Theorem 11.1,

$$u(x) = \mathsf{E}^x \int_0^{\tau_0\wedge\tau_R} 1\, dt = \mathsf{E}^x \tau_0 \wedge \tau_R.$$

One can verify that the solution to the ODE is $u(x) = x(R - x)$, and hence $\mathsf{E}^x \tau_0 \wedge \tau_R = x(R - x)$.

11.4 A brief introduction to Doob's h-transform

The goal of this section is to how to condition to zero-probability events to get new processes. Two notable examples are:

1. the Bessel-3 process as "one-dimensional Brownian motion conditioned on never hitting zero",
2. the Brownian bridge as "one-dimensional Brownian motion conditioned on hitting zero at time T".

This is done via the Doob's h-transform. Here we will not give a full account of whole theory but just focus on the computation part.

Let $X_t = \omega_t$ be a Markov process on a state space S. Recall the shift operator $(\theta_t\omega)_s = \omega_{s+t}$ The *invariant σ-field* is

$$\mathcal{I} = \{A : A = \theta_t^{-1}A,\ \forall t > 0\} = \{\text{ sets invariant under } \theta_t\}.$$

Elements in $\mathcal{I}$ are also called *invariant sets*. Typical invariant sets are

$$\{\tau_\Gamma < \infty\}, \quad \{\tau_{\Gamma_1} < \tau_{\Gamma_2}\}.$$

In general, an invariant set should only depend on the infinite future.

We also say a function h is harmonic if $\mathsf{P}_t h \equiv h$. If the Markov process is in continuous time and has a generator $\mathcal{L}$, then this implies $\mathcal{L}h = 0$, which justifies the term "harmonic". We may also say that h is $\mathcal{L}$-harmonic if we are dealing with more than one generators.

The first observation is that harmonic functions are linked to invariant sets.

Proposition 11.10 *Let $A \in \mathcal{I}$. Then $h(x) = \mathsf{P}^x(A)$ is harmonic.*

Proof: We have

$$h(x) = \mathsf{P}^x(\omega : \omega \in A) = \mathsf{P}^x(\omega : \omega \in \theta_t^{-1}A) = \mathsf{P}^x(\omega : \theta_t\omega \in A) = \mathsf{P}^x(X_t \in A) = (\mathsf{P}_t h)(x).$$

$\square$

Let A be an invariant set and $h(x) = \mathsf{P}^x(A)$. Let $\tilde{S} = \{x \in S : h(x) > 0\}$. We can define a new measure $\tilde{\mathsf{P}}^x \ll \mathsf{P}^x$ by specifying the Rydon–Nikodym derivative

$$\tilde{\mathsf{P}}^x(d\omega) = \frac{\mathbb{1}_A(\omega)}{h(x)}\mathsf{P}^x(d\omega). \tag{11.16}$$

Theorem 11.11 (Doob's h-transform) *Let $x \in \tilde{S}$. The process X is again a Markov process under $\tilde{\mathsf{P}}^x$. Moreover, the transition kernel for the new Markov process is*

$$\tilde{p}_t(x, dy) = \frac{h(y)}{h(x)}p_t(x, dy). \tag{11.17}$$

Proof:

Let us first verify that (11.17) indeed gives a Markov transition kernel on $\tilde{S}$. First, $\tilde{p}_t(x, \cdot)$ is a probability measure on $\tilde{S}$, since

$$\tilde{p}_t(x, \tilde{S}) = \int_{\tilde{S}} \frac{h(y)}{h(x)}p_t(x, dy) = \frac{1}{h(x)}\int_S h(y)\,p_t(x, dy) = \frac{1}{h(x)}\mathsf{E}^x h(X_t) = 1$$

as h is harmonic. Second, h satisfies the Komolgorov–Chapman equation: noting that $p_s(z, S) = 0$ if $h(z) = 0$

$$\begin{aligned}\tilde{p}_{t+s}(x, dy) &= \int_S \frac{h(y)}{h(x)}p_t(x, dz)p_s(z, dy) = \int_{\tilde{S}} \frac{h(y)}{h(x)}p_t(x, dz)p_s(z, dy)\\ &= \int_{\tilde{S}} \frac{h(z)}{h(x)}p_t(x, dz)\frac{h(z)}{h(x)}p_s(z, dy) = \int_{\tilde{S}} \tilde{p}_t(x, dz)\tilde{p}_t(z, dy).\end{aligned}$$

Next, we will show that (11.16) defines a Markov process with kernel (11.17). Note that restrict to any $\mathcal{F}_t$, the Markov property implies that

$$\tilde{\mathsf{P}}^x|_{\mathcal{F}_t}(d\omega) = \mathsf{E}^x\Big[\frac{\mathbb{1}_A(\omega)}{h(x)} \mid \mathcal{F}_t\Big]\mathsf{P}^x|_{\mathcal{F}_t}(d\omega) = \frac{h(X_t)}{h(x)}\mathsf{P}^x|_{\mathcal{F}_t}(d\omega) \tag{11.18}$$

For the Markov property of X_t under $\tilde{\mathsf{P}}^x$, it suffices to check that for any bounded continuous function f,

$$\tilde{\mathsf{E}}^x\big[f(X_{t+s}) \mid \mathcal{F}_t\big] = \Big[\int \tilde{p}_s(y, dz) f(z)\Big]|_{y=X_t}.$$

Take $\Gamma \in \mathcal{F}_t$. We have

$$\begin{aligned}
\tilde{\mathsf{E}}^x\Big(\tilde{\mathsf{E}}^x\big[f(X_{t+s}) \mid \mathcal{F}_t\big]\mathbb{1}_\Gamma\Big) &= \tilde{\mathsf{E}}^x f(X_{t+s})\mathbb{1}_\Gamma \\
&= \mathsf{E}^x \frac{h(X_{t+s})}{h(x)} f(X_{t+s})\mathbb{1}_\Gamma \\
&= \mathsf{E}^x \frac{h(X_t)}{h(x)} \cdot \frac{h(X_{t+s})}{h(X_t)} f(X_{t+s})\mathbb{1}_\Gamma \\
&= \mathsf{E}^x \frac{h(X_t)}{h(x)}\mathbb{1}_\Gamma \cdot \mathsf{E}^y\Big[\frac{h(X_{t+s})}{h(y)} f(X_{t+s}) \mid y = X_t\Big] \\
&= \mathsf{E}^x \frac{h(X_t)}{h(x)}\mathbb{1}_\Gamma\Big(\int \tilde{p}_s(y, dz) f(z)\Big)_{y=X_t} \\
&= \tilde{\mathsf{E}}^x \mathbb{1}_\Gamma\Big(\int \tilde{p}_s(y, dz) f(z)\Big)_{y=X_t},
\end{aligned}$$

as desired. □

If the original process X has a generator $\mathcal{L}$, then by (11.17) the generator for the new process is given by

$$(\tilde{\mathcal{L}}f)(x) = \frac{1}{h(x)}(\mathcal{L}hf)(x).$$

If X is a diffusion and $\mathcal{L} = \frac{1}{2}\sum a_{ij}\partial_{ij} + \sum b_i\partial_i$, then direct computation gives

$$\begin{aligned}
(\tilde{\mathcal{L}}f)(x) &= \frac{1}{h(x)}\Big[\frac{1}{2}\sum a_{ij}\partial_{ij}(h(x)f(x)) + \sum b_i(x)\partial_i\big(h(x)f(x)\big)\Big] \\
&= \frac{1}{h(x)}\Big(h(x)\cdot(\mathcal{L}f)(x) + f(x)\cdot(\mathcal{L}h)(x) + \frac{1}{2}\sum a_{ij}\partial_i h\partial_j f\Big) \\
&=: \Big(\mathcal{L} + \tilde{b}\cdot\nabla\Big)f,
\end{aligned}$$

where

$$\tilde{b}_i(x) = \frac{1}{2h(x)}\sum_j a_{ij}(\partial_j h)(x).$$

That is, the effect of Doob's h-transform is to add a drift to the diffusion.

What if $\mathsf{P}^x(A) = 0$? An intuitive idea is to consider $A_\varepsilon \downarrow 0$ with $\mathsf{P}^x(A_\varepsilon) > 0$, and taking the $\varepsilon \downarrow 0$ limit of the conditioned processes. What will happen is that as $\varepsilon \downarrow 0$, there will be some function harmonic function $h(x)$ such that

$$\lim_{\varepsilon\downarrow 0}\frac{h_\varepsilon(x)}{h_\varepsilon(y)} = \frac{h(x)}{h(y)}, \quad \forall x, y.$$

In light of (11.17) or (11.18), the Doob's h-transform makes sense for any harmonic function h.

Let us return to the examples we mention at the beginning of this section.

One-dimensional conditioned on never hitting 0.
A candidate for A_ε is $A_\varepsilon = \{\tau_0 > \tau_{1/\varepsilon}\}$. One can easily find that

$$h_\varepsilon(x) = \mathsf{P}^x(\tau_0 > \tau_{1/\varepsilon}) = \varepsilon x.$$

Therefore, we should choose $h(x) = x$ and consequently,

$$\tilde{\mathcal{L}} = \frac{1}{2}\partial_{xx} + \frac{1}{x}\partial_x.$$

This is the generator for Bessel-3 process.

Brownian bridge.
A candidate for A_ε is $A_\varepsilon = \{B_T \in (-\varepsilon, \varepsilon)\}$. Note that to use the invariant set setting, we need to lift the Brownian motion to $X_t = (t, B_t)$ on the state space $(0,\infty)\times\mathbb{R}$, so that A_ε is an invariant set. One can check that now the corresponding $h(t,x)$ is given by

$$h(t,x) = \frac{1}{\sqrt{2\pi(T-t)}} e^{-\frac{x^2}{2(T-t)}}.$$

Hence

$$\tilde{b}(t,x) = \frac{1}{h(t,x)}\Big(-\frac{x}{T-t}\Big)h = -\frac{x}{T-t}.$$

Therefore, the generator for the conditioned process is $\mathcal{L}_t = \frac{1}{2}\partial_{xx} - \frac{x}{T-t}\partial_x$, and we recover the SDE (9.41) satisfied by the Brownian bridge.

12 Local time and Brownian excursion

12.1 Local time for continuous semi-martingale

One way to understand the the Itô's formula (Theorem 6.1) is that continuous semi-martingales as a class of processes are invariant under $\mathcal{C}^2$ transforms. One can ask if the $\mathcal{C}^2$ condition can be relaxed. The first result is to generalize this to convex functions.

Proposition 12.1 *If f is convex and X is a continuous semi-martingale, then $f(X_t)$ is also a continuous semi-martingale and we have*

$$f(X_t) = f(X_0) + \int_0^t f'_-(X_s)\,dX_s + A_t^f, \tag{12.1}$$

where $(A_t^f)_{t\ge 0}$ is some increasing process.

Proof: We will try to approximate f by $f_n \in \mathcal{C}^2$ and investigate what will be the limit of the Itô's formula applied to f_n.

The approximation is a standard argument using *mollifiers*. Let $h(x) \in \mathcal{C}_0^\infty(\mathbb{R})$ be a such that

$$h(x) \ge 0, \quad \operatorname{supp} h \subset [0,1], \quad \int_0^1 h(x)\,dx = 1.$$

Such function exists, one example being $h(x) = ce^{-\frac{1}{x^2(1-x)^2}}$. The mollification

$$g_n(x) = (h_n * g)(x) = \int h_n(y)g(x-y)\,dy = \int h_n(x-y)g(y)\,dy$$

enjoys many good properties:

1. if g is locally integrable, then $g_n \in \mathcal{C}^\infty$;
2. $g_n(x) \to g(x)$ for a.e. x;
3. if g is continuous on $[a,b]$, then $g_n \to g$ uniformly on $[a,b]$;
4. if $f \in L^p_{\text{loc}}$, $1 \le p < \infty$, then $g_n \to g$ in L^p_{loc}.

Write the decomposition of the continuous semi-martingale X as $X_t = M_t + V_t$ where M is a c.l.m. and V_t a finite-variation process. Consider the stopping time

$$\tau_K = \inf\{t \ge 0 : |X_t| + \langle M\rangle_t + \int_0^t |dV_s| \ge K\}.$$

We first look at the stopped process $X_t^{\tau_K}$. We can assume that f, f'_- are all bounded since we only care about their values in a finite interval.

Let $f_n = h_n * f$. Note that f is a convex function, so f is continuous and f'_- is left-continuous. We have $f_n \to f$ uniformly and in L^p, $p \ge 1$. Also, we have $f'_n = h_n * f'_-$ ($= h_n * f'_+$, but $f'_\pm$ differ at countably many points, so the integration is the same), and since f'_- is left-continuous and h_n is supported on $[0, 1/n]$, we know that $f'_n \to f'_-$ at every point. Finally, f_n is a convex function, since f_n is a convex integration of the convex functions $f(\cdot - y)$, $y \in \mathbb{R}$, and convexity is preserved under convex combination. As a result, $f''_n \ge 0$.

Applying Itô's formula to $f_n \in \mathcal{C}^2$, we have

$$f_n(X_{t\wedge\tau_K}) = f_n(X_0) + \int_0^{t\wedge\tau_K} f'_n(X_s)\,dM_s + \int_0^{t\wedge\tau_K} f'_n(X_s)\,dV_s + \frac{1}{2}\int_0^{t\wedge\tau_K} f''_n(X_s)\,d\langle X\rangle_s. \tag{12.2}$$

The left-hand side and the first, third third term on the right-hand side is defined path-wise. Moreover, almost surely, we have

$$f_n(X_{t\wedge\tau_K}) \to f(X_{t\wedge\tau_K}), \quad f_n(X_0) \to f(X_0)$$

since $f_n(z) \to f(z)$ for every z and

$$\int_0^{t\wedge\tau_K} f'_n(X_s)\,dV_s \to \int_0^{t\wedge\tau_K} f'_-(X_s)\,dV_s$$

by $f'_n(X_s) \to f'_-(X_s)$ for every s and Bounded Convergence Theorem applied to the measure dV_s. The second term on the right-hand side is a square-integrable martingale defined in the L^2-sense, and we have by Itô's isometry,

$$\mathsf{E}\Big|\int_0^{t\wedge\tau_K} f'_n(X_s)\,dM_s - \int_0^{t\wedge\tau_K} f'_-(X_s)\,dM_s\Big|^2 = \mathsf{E}\int_0^{t\wedge\tau_K} |f'_n(X_s) - f'_-(X_s)|^2\,d\langle M\rangle_s.$$

The right-hand side converges to 0 Inside the expectation, since $|f'_n(X_s) - f'_s(X_s)| \to 0$ for every s and is bounded, by Bounded Convergence Theorem applied on the measure $d\langle M\rangle_s$, we know that

$$\int_0^{t\wedge\tau_K} |f'_n(X_s) - f'_-(X_s)|^2\,d\langle M\rangle_s \to 0.$$

This integral is bounded due to our localization. Then by Bounded Convergence Theorem again, we know the second term in (12.2) converges in probability to $\int_0^{t\wedge\tau_K} f'_-(X_s)\,dM_s$.

Therefore, there exists a process $A_t^{f,K}$ such that

$$\frac{1}{2}\int_0^{t\wedge\tau_K} f''_n(X_s)\,d\langle M\rangle_s \to A_t^{f,K}$$

in probability for every $t\geq 0$. Since the pre-limiting process is increasing in t, the limit $A_t^{f,K}$ is also increasing in t.

The remaining procedure is to remove the localization. We omit the proof here. We will obtain an increasing process A_t^f such that $A_t^{f,K} = A_t^f$ for $t\leq \tau_K$. This completes the proof. □

Remark 12.1 We use $f'_n(x)\to f'_-(x)$ *for every* x. The above argument does not work if only assuming $f'_n\to f'_-$ a.e. (in fact, we also have a.s. $f'_n(x)\to f'_+(x)$ since $f'_- = f'_+$ except at countably many points), since the random measures $A\mapsto \int_0^t \mathbb{1}_A(X_s)\,dV_s$ and $A\mapsto \int_0^t \mathbb{1}_A(X_s)\,d\langle M\rangle_s$ do not have to be absolutely continuous w.r.t. to the Lebesgue measure.

Remark 12.2 One can get a similar statement if we consider the mollifier $\tilde{h}(x) = h(x-1)$, then $\tilde{h}$ will be supported on $[-1,0]$ and $f'_n\to f'_+$. Correspondingly we will get another increasing process $\tilde{A}_t^f$ such that

$$f(X_t) = f(X_0) + \int_0^t f'_+(X_s)\,dX_s + \tilde{A}_t^f.$$

In general $A^f\neq \tilde{A}^f$. When $f\in\mathcal{C}^2$, then

$$A_t^f = \tilde{A}_t^f = \frac{1}{2}\int_0^t f''(X_s)\,d\langle M\rangle_s.$$

Let $\operatorname{sgn}(x) = \mathbb{1}_{x>0} - \mathbb{1}_{x\leq 0} = (|x|)'_-$. The Tanaka's formula gives the definition of local time.

Theorem 12.2 *Let X be a continuous semi-martingale For every $a\in\mathbb{R}$, there exists an increasing process $L^a(X)$ such that*

$$|X_t - a| = |X_0 - a| + \int_0^t \operatorname{sgn}(X_s - a)\,dX_s + L_t^a(X), \tag{12.3}$$

$$(X_t - a)_+ = (X_0 - a)_+ + \int_0^t \mathbb{1}_{\{X_s>a\}}\,dX_s + \frac{1}{2}L_t^a(X), \tag{12.4}$$

$$(X_t - a)_- = (X_0 - a)_- - \int_0^t \mathbb{1}_{\{X_s\leq a\}}\,dX_s + \frac{1}{2}L_t^a(X).$$

Proof: Applying Proposition 12.1 to the convex function $f(x) = |x-a|$, we obtain an increasing process A_t^f, which we will call $L_t^a(X)$ and (12.3) holds. Applying Proposition 12.1 to the convex functions $f(x) = (x-a)_\pm$, we obtain another two increasing process

$$(X_t - a)_+ = (X_0 - a)_+ + \int_0^t \mathbb{1}_{\{X_s>a\}}\,dX_s + A_t^+ \tag{12.5}$$

$$(X_t - a)_- = (X_0 - a)_- - \int_0^t \mathbb{1}_{\{X_s\leq a\}}\,dX_s + A_t^-. \tag{12.6}$$

It remains to show that $A_t^\pm = \dfrac{1}{2} L_t^a(X)$.

In fact, taking the difference of (12.5), (12.6) and noting that $x = x_+ - x_-$, we obtain

$$(X_t - a) = X_0 - a + \int_0^t \left(\mathbb{1}_{\{X_s > a\}} + \mathbb{1}_{\{X_s \le a\}}\right) dX_s + (A_t^+ - A_t^-).$$

From this we have $A_t^+ = A_t^-$. On the other hand, taking the sum of (12.5) and (12.6) and noting that $|x| = x_+ - x_-$, we have

$$|X_t - a| = |X_0 - a| + \int_0^t \operatorname{sgn}(X_s - a)\, dX_s + (A_t^+ + A_t^-).$$

Comparing with (12.3) we see that $A_t^+ = A_t^- = \dfrac{1}{2} L_t^a(X)$ as desired. $\square$

Definition 12.1 *Let X be a continuous semi-martingale. For $a \in \mathbb{R}$, $L^a(X) = (L_t^a(X))_{t \ge 0}$ given by* (12.3) *is called the* local time *of X.*

12.2 Continuity of local time and other properties

The main result for this section is the *Generalized Itô's formula.*

Theorem 12.3 *Let X be a continuous semi-martingale and f be a convex function (or difference of two convex function). Then*

$$f(X_t) = f(X_0) + \int_0^t f'_-(X_s)\, dX_s + \frac{1}{2} \int_{\mathbb{R}} L_t^a(X) f''(da), \tag{12.7}$$

where $f''(da) = df'_-(a)$ is the signed measure generated by the finite-variation f'_-.

First, we note that when $f(x) = |x-a|, (x-a)_+$ and $(x-a)_-$, Theorem 12.3 reduces to Theorem 12.2. In fact, any convex function is more or less a convex integration of $(x-a)_+$, so Theorem 12.3 just comes from a convex integration of Theorem 12.2.

Lemma 12.4 *Let f be a convex function with $\lim\limits_{x\to\infty} f(x) = 0$. Then*

$$f(x) = \int (x-a)_+ f''(da), \quad f'(x) = \int \mathbb{1}_{\{x > a\}} f''(da).$$

Proof: The convexity of f and the assumption also imply that $\lim\limits_{x \to -\infty} f'(x) = 0$, so we have

$$f'(x) = \int_{-\infty}^x f''(da). \tag{12.8}$$

For the first identity, by Fubini and (12.8) we have

$$\begin{aligned} f(x) &= \int \mathbb{1}_{\{y<x\}} f'(y) dy = \int \mathbb{1}_{\{y<x\}} dy \cdot \int \mathbb{1}_{\{a \le y\}} f''(da) \\ &= \int f''(da) \int \mathbb{1}_{\{a \le y < x\}} dy = \int (x-a)_+ f''(da). \end{aligned}$$

$\square$

Second, since $f''(\cdot)$ is only a signed measure, we need the $f''(\cdot)$-measurability of $a \mapsto L_t^a(X)$ for the last integral in (12.7) to make sense. In fact, we will show that $a \mapsto L_t^a(X)$ is cadlag, so that it is $f''(\cdot)$-measurable for any convex f, and for many scenarios it is even continuous. Such regularity is not obvious from the definition of local time (12.3).

Let X be a continuous semi-martingale with decomposition $X = M + V$. Let

$$Y_t^a(X) = \int_0^t \mathbb{1}_{\{X_s > a\}}\, dM_s, \quad Z_t^a(X) = \int_0^t \mathbb{1}_{\{X_s > a\}}\, dV_s$$

We note that $Y^a = (Y_t^a)_{t\geq 0}$ and $Z^a = (Z_t^a)_{t \geq 0}$ are continuous processes. We can view $a \mapsto Y^a$ and $a \mapsto Z^a$ as a stochastic process taking values in $\mathcal{C}(\mathbb{R}_+)$, equipped with the *locally uniform (LU)* topology:

$$\mathbf{x}_n \to \mathbf{y} \text{ in } \mathcal{C}(\mathbb{R}_+) \quad \Leftrightarrow \quad \lim_{n\to\infty} \sup_{0\leq t \leq T} |x_n(t) - y(t)| = 0, \ \forall T > 0.$$

Equivalently, the LU topology is generated by open sets given by the metric

$$d_{\mathrm{LU}}(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{\infty} \frac{|x_n - y|_{L^\infty[0,n]} \wedge 1}{2^n}.$$

Lemma 12.5 *The mapping $a \mapsto Z^a$ is cadlag in $\mathcal{C}(\mathbb{R}_+)$.*

Proof: For all $s \geq 0$, we have the convergence

$$\lim_{a \downarrow a_0} \mathbb{1}_{\{X_s > a\}} = \mathbb{1}_{\{X_s > a_0\}}, \ \lim_{a\uparrow a_0} \mathbb{1}_{\{X_s > a\}} = \mathbb{1}_{\{X_s \geq a_0\}}.$$

Using Bounded Convergence Theorem w.r.t. the measure dV_s, we have

$$\lim_{a\downarrow a_0} \int_0^t \mathbb{1}_{\{X_s > a\}}\, dV_s = \int_0^t \mathbb{1}_{\{X_s > a_0\}}\, dV_s,$$

and

$$\lim_{a \uparrow a_0} \int_0^t \mathbb{1}_{\{X_s > a\}}\, dV_s = \int_0^t \mathbb{1}_{\{X_s \geq a_0\}}\, dV_s =: Z_t^{a-}. \tag{12.9}$$

$\square$

Lemma 12.6 *The process $a \mapsto Y^a \in \mathcal{C}(\mathbb{R}_+)$ has a continuous modification.*

Proof: Let

$$T_n = \inf\{t : \langle M \rangle_t + \int_0^t |dV_s| \geq n\}$$

and consider the stopped process $Y_{t\wedge T_n}^a$. By standard localization argument, it suffices to show that $a \mapsto (Y_{t\wedge T_n}^a)_{t\geq 0}$ has continuous modification for every n. Therefore, without loss of generality, we assume that for some $K > 0$,

$$\langle M \rangle_t + \int_0^t |dV_s| \leq K, \quad \forall t > 0.$$

Our main tool is Theorem 2.7. Although Theorem 2.7 is stated for a stochastic process on $\mathbb{R}$, the theorem also holds for stochastic processes taking value in any metric space, for example $\mathcal{C}(\mathbb{R}_+)$. Taking into account the definition of local uniform topology, it suffices to show

$$\mathsf{E} \sup_{0 \leq t \leq T} |Y_t^b - Y_t^a|^\alpha \leq C|b-a|^{1+\beta} \tag{12.10}$$

for some $\alpha, \beta > 0$; then Theorem 2.7 will ensure that $a \mapsto Y^a$ has a continuous modification.

Let $b > a$. Since

$$Y_t^b - Y_t^a = \int_0^t \mathbb{1}_{\{a<X_s\le b\}}\, dM_s$$

is a martingale, by Theorem 10.6 we have for any $p > 0$,

$$\mathsf{E} \sup_{0\le t\le T} |Y_t^b - Y_t^a|^p \le \Big[\mathsf{E} \int_0^T \mathbb{1}_{\{a<X_s\le b\}}\, d\langle M\rangle_s\Big]^{p/2}.$$

We will try to bound the integral on the right-hand side. This integral is almost the Itô's correction term, but $\mathbb{1}_{\{a<X_s\le b\}}$ is not continuous so it is not the second derivative of a $\mathcal{C}^2$-function.

To fix this, let us introduce

$$\varphi_{a,b}(x) = \begin{cases} \dfrac{x-a_1}{a-a_1}, & x < a, \\ 1, & x \in [a,b], \qquad a_1 = a - |b-a|, \quad b_1 = b + |b-a|. \\ \dfrac{b_1 - x}{b_1 - b}, & x > b, \end{cases}$$

Then $f(x) = \displaystyle\int_0^x \int_0^y \varphi_{a,b}(z)\, dz \in \mathcal{C}^2$ and $f'' = \varphi_{a,b}$. Moreover,

$$|f'(x)| \le |\varphi_{a,b}|_{L^1} = 2|b-a|.$$

By Itô's formula, we have

$$0 \le \int_0^T \mathbb{1}_{\{a<X_t\le b\}}\, d\langle M\rangle_t \le f(X_T) - f(X_0) - \int_0^T f'(X_t)\, dM_t - \int_0^T f'(X_t)\, dV_t.$$

For $q \ge 1$, we have

$$\begin{aligned} \mathsf{E}|f(X_T) - f(X_0)|^q &\le \sup |f'|^q \cdot \mathsf{E}|X_T - X_0|^q \le C|b-a|^q \mathsf{E}\Big[|M_T - M_0| + \int_0^T |dV_t|\Big]^q \\ &\le C|b-a|^q \mathsf{E}\Big[\langle M\rangle_T^{q/2} + \Big(\int_0^T |dV_s|\Big)^q\Big] \le C_K |b-a|^q, \end{aligned}$$

$$\mathsf{E}\Big|\int_0^T f'(X_t)\, dM_t\Big|^q \le C\mathsf{E}\Big|\int_0^T |f'(X_t)|^2\, d\langle M\rangle_t\Big|^{q/2} \le C_K |b-a|^q,$$

and

$$\mathsf{E}\Big|\int_0^T f'(X_t)\, dV_t\Big|^q \le C\mathsf{E}\Big|\int_0^T |f'(X_t)|\, |dV_t|\Big|^q \le C_K|b-a|^q.$$

Combining all these, we see that

$$\mathsf{E} \sup_{0\le t\le T} |Y_t^b - Y_t^a|^p \le \Big[\mathsf{E} \int_0^T \mathbb{1}_{\{a<X_s\le b\}}\, d\langle M\rangle_s\Big]^{p/2} \le C_K |b-a|^{p/2}$$

if $p \ge 2$. This completes the proof. □

With all these preparations, we have the regularity for the local time $L_t^a(X)$.

Theorem 12.7 *Let X be a continuous semi-martingale and $L_t^a(X)$ be its local time. Then $a \mapsto L^a$ has a cadlag modification in $\mathcal{C}(\mathbb{R}_+)$. Moreover,*

$$L_t^a(X) - L_t^{a-}(X) = 2\int_0^t \mathbb{1}_{\{X_s=a\}}\,dV_s,$$

where $X = M + V$ is the decomposition for X.

In particular, if X is a c.l.m. so that $V \equiv 0$, then $L_t^a(X)$ is jointly continuous in (a, x).

Proof: By Theorem 12.2, we have

$$L_t^a(X) = 2\cdot\Big[(X_t-a)_+ - (X_0-a)_+ - \int_0^t \mathbb{1}_{\{X_s>a\}}\,dX_s\Big].$$

The two functions $(X_t - a)_+$ and $(X_0 - a)_+$ are obviously continuous in (a, x). The conclusion then follows from Lemmas 12.5 and 12.6. $\square$

Now we are ready to prove the Generalized Itô's formula. **Proof of Theorem 12.3:** After localization, we can assume that X is bounded, so without loss of generality we can assume that the support of f'' is contained in $[-K, K]$ for some $K > 0$. Moreover, (12.7) is invariant after adding a linear function to f, so we can further assume that $f(x) \equiv 0$ for $x < -K$. Noting Lemma 12.4, integrating (12.4) w.r.t. the measure $f''(\cdot)$ gives

$$f(X_t) = f(X_0) + \int f''(a)da\int_0^t \mathbb{1}_{\{X_s>a\}}\,dX_s + \frac{1}{2}\int L_t^a(X)f''(da).$$

We only need to justify the following Fubini Theorem holds:

$$\int f''(a)da\int_0^t \mathbb{1}_{\{X_s>a\}}\,dX_s = \int_0^t dX_s\int f''(a)da\cdot\mathbb{1}_{\{X_s>a\}} = \int_0^t f'_-(X_s)\,dX_s.$$

Note that $\int dX_s = \int dV_s + \int dM_s$. The integral w.r.t. dV_s is a Riemann–Stieltjes integral, which is defined pathwise and Fubini Theorem holds. On the other hand, the stochastic integral is NOT defined pathwise and justifying the change of order of integration has to be more careful. This is also knows as the *stochastic Fubini Theorem* and we give a proof here in our setting.

For simplicity let us assume that $\langle M\rangle_\infty \le K$ (which can be achieved by localization). Let

$$\Phi_t = \int_{[-K,K]} f''(da)\int_0^t \mathbb{1}_{\{X_s>a\}}\,dM_s = \int f''(da)Y_t^a$$

Note that Φ_t is a martingale, since $a \mapsto Y^a$ is cadlag and hence Φ_t can be approximated by Riemann sums, each of which is a linear combination of $Y_t^{a_i}$ and hence a martingale. So Φ_t is the limit of martingales, and must also be a martingale. It is not hard to see that Φ_t is square-integrable, so $\Phi \in \mathbb{H}^2$.

Take any $N \in \mathbb{H}^2$. Then

$$\begin{aligned}\mathsf{E}\langle\Phi, N\rangle_\infty &= \mathsf{E}\int_{[-K,K]} f''(da)\cdot\int_0^\infty \mathbb{1}_{\{X_s>a\}}\,d\langle M,N\rangle_s\\ &= \mathsf{E}\int_0^\infty d\langle M,N\rangle_s\cdot\int_{[-K,K]} f''(da)\mathbb{1}_{\{X_s>a\}}\\ &= \mathsf{E}\int_0^\infty f'_-(X_s)\,d\langle M,N\rangle_s\end{aligned}$$

$$= \mathsf{E}\langle\int_0^\cdot f'_-(X_s)\,dM_s, N\rangle_\infty.$$

Hence by Theorem 6.6,

$$\Phi_t = \int_0^t f'_-(X_s)\, dM_s.$$

This completes the proof. $\square$

We can use Theorem 12.3 to characterize the local time as *density of occupation time.*

Theorem 12.8 *Let X be a continuous semi-martingale and $L_t^a(X)$ be its local time. Almost surely, for any $t \geq 0$ and $\varphi \geq 0$ measurable, we have*

$$\int_0^t \varphi(X_s)\, d\langle X\rangle_s = \int_{\mathbb{R}} \varphi(a) L_t^a(X)\, da. \tag{12.11}$$

In another word, the random measure

$$A \mapsto \int_0^t \mathbb{1}_A(X_s)\, d\langle X\rangle_s$$

almost surely has density $L_t^a(X)$.

Proof: Since a signed measure is determined by the integral of countable many compactly supported continuous functions against it, it suffices to show that (12.11) holds almost surely for a fixed compactly supported continuous function.

Let $f \in \mathcal{C}^2$ such that $f'' = \varphi$. Then comparing the Itô's formula and Generalized Itô's formula applied to $f(X_t)$ gives the desired result. $\square$

Proposition 12.9 *Let X be a continuous semi-martingale. Then almost surely, for all a and $t \geq 0$,*

$$L_t^a(X) = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^t \mathbb{1}_{\{a \leq X_s \leq a+\varepsilon\}}\, d\langle X\rangle_s.$$

Proof: By Theorem 12.8 with $\varphi = \mathbb{1}_{[a,a+\varepsilon]}$, we have

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_0^t \mathbb{1}_{\{a \leq X_s \leq a+\varepsilon\}}\, d\langle X\rangle_s = \frac{1}{\varepsilon} \int_a^{a+\varepsilon} L_t^b(X)\, db.$$

The desired conclusion follows from the right continuity of $b \mapsto L_t^b(X)$. $\square$

Corollary 12.10 *Almost surely, for all a and $t > 0$,*

$$\frac{1}{2}(L_t^a(X) + L_t^{a-}(X)) = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^t \mathbb{1}_{\{a-\varepsilon \leq X_s \leq a+\varepsilon\}}\, d\langle X\rangle_s.$$

We call $\tilde{L}_a^a(X) = \frac{1}{2}(L_t^a(X) + L_t^{a-}(X))$ the *symmetric local time.* We also have the generalized Itô's formula for symmetric local time:

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\, dX_s + \frac{1}{2}\tilde{L}_t^a(X) f''(da),$$

where $f'(x) := \frac{1}{2}[f'_+(x) + f'_-(x)]$.

12.3 Brownian excursions

We recall the Lebesgue decomposition for an increasing function $f : \mathbb{R} \to \mathbb{R}$: f can be uniquely written as the sum of three functions

$$f = f^{\text{jump}} + f^{\text{abs}} + f^{\text{sing}},$$

such that

- f^{jump} increases only by jumps, that is, $df^{\text{jump}} = \sum_{i=1}^{\infty} a_i \delta_{b_i}$;
- f^{abs} is absolutely continuous and increasing, so it has a derivative $g(x)$ a.s. and $df^{\text{abs}} = g(x)\,dx$;
- f^{sing} is continuous, increasing, and it is differentiable almost everywhere but the derivative equals to 0, that is, df^{sing} is a mutually singular with respect to the Lebesgue measure.

Note that an example for f^{sing} is the *Cantor's function*:

$$\varphi(x) = \sum_{n=1}^{\infty} \frac{2x_n}{3^n}, \quad x = \sum_{n=1}^{\infty} \frac{x_n}{2^n},\ x \in [0,1],\ x_n \in \{0,1\}.$$

We recall the notation for level sets of a process X:

$$\mathcal{Z}_a^X(\omega) = \{t \geq 0 : X_t(\omega) = a\}.$$

Also, for a measure μ on $\mathbb{R}$, we denote by $\operatorname{supp}\mu$ the *support of* μ, defined

$$\operatorname{supp}\mu = \{x \in \mathbb{R} : \mu(x - \varepsilon, x + \varepsilon) > 0,\ \forall \varepsilon > 0\}.$$

It is easy to check that $\operatorname{supp}\mu$ is always a closed set.

Proposition 12.11 *Let X be a continuous semi-martingale and let $L_t^a(X)$ be its local time at level a. Then for a.e. ω, $d_s L_s^a(X)$ is supported on $\mathcal{Z}_a^X(\omega)$.*

Proof: Let $Y_t = |X_t - a|$. Then by Theorem 12.2, Y_t is a continuous semi-martingale and

$$dY_t = \operatorname{sgn}(X_t - a)\,dX_t + dL_t^a(X).$$

Applying Itô's formula to Y_t^2, we have

$$Y_t^2 = Y_0^2 + \int_0^t 2Y_t\,dY_t + \int_0^t d\langle X\rangle_s = Y_0^2 + \int_0^t 2(X_t - a)\,dX_t + \int_0^t 2|X_t - a|\,dL_t^a(X) + \int_0^t d\langle X\rangle_s. \quad (12.12)$$

Here, we used $\operatorname{sgn}(x) \cdot |x| = x$.

On the other hand, $Y_t^2 = (X_t - a)^2$. Comparing the Itô's formula for this expression with (12.12), we obtain

$$\int_0^t |X_s - a|\,dL_s^a(X) = 0.$$

Then, for every $\varepsilon > 0$, $dL_s^a(X)\Big(\{s \geq 0 : |X_s - a| > \varepsilon\}\Big) = 0$. So if $X_s \neq a$, then $s \notin \operatorname{supp}\mu$. This completes the proof. □

In what follows we focus on the case where X is the Brownian motion. Recalling Proposition 3.22, we first have the following.

Corollary 12.12 *For every a, with probability one,* $\mathrm{Leb}(\mathcal{Z}_a^B(\omega)) = 0$.

Since the Brownian motion has continuous sample path, its zero set $\mathcal{Z} := \mathcal{Z}_0^B(\omega)$ is a close set, and hence $\mathcal{Z}^c$ is an open subset of $\mathbb{R}$. As any open set in $\mathbb{R}$, $\mathcal{Z}^c$ can be written as a disjoint union

$$\mathcal{Z}^c = \bigcup_{n=1}^{\infty} (a_n, b_n),$$

where on each (a_n, b_n), the Brownian motion B is either strictly positive or negative. We call these intervals (a_n, b_n) *excursion intervals* of the Brownian motion, and for each n, $(B_t)_{t\in[a_n,b_n]}$ *an excursion* of the Brownian motion. Consider the renormalized excursion

$$\tilde{e}_n^s = \frac{1}{\sqrt{b_n - a_n}}\Big(B_{a_n+s(b_n-a_n)} - B_{a_n}\Big), \quad s \in [0,1].$$

In fact,

- $(\tilde{e}_n)_{n\geq 1}$ are i.i.d. processes.
- $\tilde{e}$ solves the SDE

$$\tilde{e}_0 = 0, \quad d\tilde{e}_t = \frac{1}{\tilde{e}_t} - \frac{\tilde{e}_t}{1-t}\, dt + dW_t.$$

  The SDE can be derived from the Doob-h transform of the Brownian motion with

$$h(x,t) = \frac{x}{(1-t)^{3/2}} e^{-\frac{x^2}{2(1-t)}},$$

  interpreted as "conditioned on hitting 0 at time 1", and $h(x,t)$ is the hitting time density.

One of the goals of studying Brownian excursion is to reconstruct Brownian motions path from excursions. Intuitive, this is achieved in two steps:

1. Determine the zero set $\mathcal{Z}$, and write $\mathcal{Z}^c = \bigcup_{n=1}^{\infty} (a_n, b_n)$.
2. Sample i.i.d. excursion processes $\tilde{e}_n$ and on each interval (a_n, b_n), define

$$B_t = \sqrt{b_n - a_n} \cdot \tilde{e}_n\Big(\frac{t-a_n}{b_n-a_n}\Big), \quad t \in (a_n, b_n).$$

We will focus on the first step, and we will see that $\mathcal{Z}$ already determines the local time of Brownian motion at 0. This is remarkable since previously we see that the local time of a continuous semi-martingale X is determined by the behavior of X near a, see Proposition 12.9.

12.3.1 Lévy's Theorem and Skorokhod equation

We use the notion $X_t^* = \sup\limits_{0\leq s\leq t} |X_t|$ to denote the *maximal process* of X.

Theorem 12.13 (Paul Lévy)

$$\big(B_t^*, B_t^* - B_t,\ t\geq 0\big) \overset{\mathrm{d}}{=} (L_t^0(B), |B_t|,\ t \geq 0).$$

We will prove the theorem by studying the *Skorokhod equation*. Let $y \in \mathcal{C}[0,\infty)$ with $y(0)=0$. A pair of functions $\big(z(t),a(t)\big)_{t\ge 0}$ is a solution to the Skorokhod equation if they solve

$$z(t) = -y(t) + a(t), \quad t \ge 0, \tag{12.13}$$

and satisfy

- $z(t) \ge 0$, $\forall t \ge 0$;
- $a(t)$ is continuous, increasing, $\lim_{t\to 0} a(t) = 0$, and
$$\operatorname{supp} da_s \subset \{s \ge 0 : z(s) = 0\}.$$

Lemma 12.14 *There exists a unique pair of solution* (z,a) *that solve* (12.13). *Moreover, the function* a *can be represented as*

$$a(t) = \sup_{s\le t} y(s). \tag{12.14}$$

Proof: **Uniqueness.**
Suppose there are two pairs of functions (z,a) and $(\tilde z,\tilde a)$ that solves (12.13). Then, $z-\tilde z = a-\tilde a$ has bounded variation. We have

$$\begin{aligned} 0 \le \frac{1}{2}(z-\tilde z)^2(t) &= \int_0^t (z-\tilde z)(s)\, d(a-\tilde a)(s) \\ &= \Big[\int_0^t z\, da + \int_0^t \tilde z d\tilde a\Big] - \int_0^t \tilde z da - \int_0^t z\, d\tilde a \le 0. \end{aligned}$$

Here, the two integrals in the bracket are zero since $\operatorname{supp} da_s$. The other two integrals are non-negative since $z,\tilde z\ge 0$ and $a,\tilde a$ are increasing. Hence $\tilde z - z \equiv 0$ and this proves uniqueness.

Uniqueness.
It suffices to verify (12.14) and $z(t) = a(t) - y(t)$ indeed give the solution. Clearly, by the definition of supremum, $z\ge 0$ and a is increasing. Also, a is continuous and $\lim_{t\downarrow 0} a(t) = 0$ since y is continuous and $y(0)=0$.

The most difficult part is to verify (12.14). Let $z(t) = a(t)-y(t) > 0$. Then by continuity of y, there exists $t_1 < t$ such that $a(t) = y(t_1) > y(t)$. By continuity of y, there exists $\varepsilon < t - t_1$ such that

$$y(t_1) > y(s), \quad s \in (t-\varepsilon, t+\varepsilon),$$

and hence $a(s) = y(t_1)$, $\forall s \in (t-\varepsilon,t+\varepsilon)$, that is, $t \in (\operatorname{supp} da_s)^c$. This proves (12.14). □

Proof of Theorem 12.13: By Theorem 12.2, we have

$$|B_t| = L_t^0(B) + \int_0^t \operatorname{sgn}(B_s)\, dB_s =: L_t^0(B) - \beta_t,$$

and by Theorem 7.1, β_t is a Brownian motion. By Proposition 12.11,

$$(y,z,a) = (\beta, |B|, L^0(B))$$

solves (12.13). By Lemma 12.14, $\beta_t^* = L_t^0(B)$ and the theorem is proved. □

Now let us define two new clocks for the Brownian motion β:

$$\begin{aligned} S_b &= \inf\{t\ge 0 : \beta_t^* > b\} = \inf\{t \ge 0 : \beta_t > b\} = \inf\{t\ge 0 : L_t^0(B) > b\}, \\ T_b &= \inf\{t\ge 0 : \beta_t^* = b\} = \inf\{t \ge 0 : \beta_t = b\} = \inf\{t\ge 0 : L_t^0(B) = b\}. \end{aligned}$$

It is not hard to see that $(S_b)_{b\ge 0}$, $(T_b)_{b\ge 0}$ are the right-continuous and left-continuous inverse of the increasing function $t \mapsto \beta_t^* = L_t^0(B)$.

Proposition 12.15 *Almost surely, the process $b \mapsto S_b$ grows only by jumps.*

Proof: Since almost surely, $\mathrm{Leb}(\mathcal{Z}_0^B(\omega)) = 0$ and T, S are the left/right-continuous inverse of $t \mapsto L_t^0(B)$, we have

$$S_b = \sum_{b' \le b} (S_{b'} - T_{b'}).$$

This proves the claim. □

The process S_b is known to be a *stable process*. In the last part of this section, we will collect some fact about infinite divisible laws, stable process and subordinates.

Definition 12.2 *A probability measure μ is* infinite divisible *if for all n, $\exists \mu_n$ such that $\mu = \mu_n^{*n}$, that is, if $X_1, \ldots, X_n$ are i.i.d. with distribution μ_n, then $X_1 + \cdots + X_n \sim \mu$.*

Examples for infinite divisible distributions are Poisson, Gaussian, and Cauchy distribution, which can be verified easily from their characteristic functions.

Definition 12.3 *A stochastic process X is a* Lévy process *if X has stationary, independent increment.*

If X is Lévy, then for every t, $X_t - X_0$ is infinite divisible.

Theorem 12.16 (Lévy–Khintchine) *A r.v. X has infinite divisible distribution if and only if its characteristic function has the representation*

$$\mathsf{E}e^{iuX} = \exp\Big(i\beta u - \frac{\sigma^2 u}{2} + \int (e^{iux} - 1 - \frac{iux}{1+x^2})\,\nu(dx)\Big), \tag{12.15}$$

where $\beta \in \mathbb{R}$, $\sigma \ge 0$ and ν is a measure on $\mathbb{R}$ such that $\displaystyle\int \frac{|x|}{1+x^2}\nu(dx) < \infty$.

If X comes from a Lévy process, then the three parts in the characteristic function corresponds to the linear part ($i\beta u$), Brownian motion ($\sigma^2 u/2$) and the pure jump process.

Definition 12.4 *A r.v. Y is stable if for every k, there exists a_k, b_k such that for $Y_i \overset{\mathrm{d}}{=} Y$,*

$$Y_1 + \cdots + Y_k \overset{\mathrm{d}}{=} a_k Y + b_k.$$

If Y is stable, then necessarily $a_k = k^{1/\alpha}$ for some $\alpha \in (0,2]$. The number α is called the stable index.

A Lévy process with stable increment is called a stable process.

Proposition 12.17 *If X is stable, then in (12.15), $\sigma = 0$ and*

$$\nu = \big(m_1 \mathbb{1}_{\{x<0\}} + m_2 \mathbb{1}_{\{x_2>0\}}\big)|x|^{-1-\alpha}$$

where α is the stable index of X.

In fact, S_b is a stable process with index $1/2$. We have

$$S_{nb} = \inf\{t \ge 0 : \beta_t > nb\} \overset{\mathrm{d}}{=} S_b^{(1)} + \cdots + S_b^{(n)},$$

where

$$S_b^{(k)} = \inf\{t \ge S_{(k-1)b} : \beta_t > kb\} - S_{(k-1)b} \overset{\mathrm{d}}{=} S_b.$$

On the other hand, by the diffusion scaling of Brownian motion,

$$S_{nb} = \inf\{t \ge 0 : \beta_t > nb\} \overset{\mathrm{d}}{=} n^2 \inf\{t \ge 0 : \beta_t > b\}.$$

Definition 12.5 *A stable process X is called a subordinates if X_t is non-decreasing.*

Although it is possible to study S_b using the tools from stable law, in this note we will study it using Poisson point process, with the knowledge that $b \mapsto S_b$ only have jumps. This will be done in the next section.

12.3.2 Poisson point process and excursion

Since S_b grows only by jumps, it is completely characterized by the information of all jumps. Each jump can be represented by a point $(b, j(b)) \in (0, \infty)^2$. The first coordinate is the *location* of the jump, and the second coordinate $j(b) = S_b - T_b$ is the size of the jump. It is still consistent to write $j(b) = 0$ if S_b is continuous at b.

All the jumps constitutes a *simple point process* on $\mathbb{R}_+^2$. A standard way to study such point process is to regard it as a random counting measure

$$\nu = \sum_{j(b) \neq 0} \delta_{(b, j(b))}.$$

The minimum measurability assumption on the random measure ν is that we can count how many points there are in a given set. This leads to the following definition.

Definition 12.6 *Let $(H, \mathcal{H})$ be a measurable space. A (simple) point process is a random counting measure ν on $(H, \mathcal{H})$ such that for all $C \in \mathcal{H}$, $\nu(C) \in \{0, 1, \cdots\} \cup \{\infty\}$ is a r.v.*

Note that we are thinking of $H = (0, \infty)^2$.

Definition 12.7 *A random counting measure ν is a* Poisson point process (PPP) *on $(H, \mathcal{H})$ if*

1. *for every $C \in \mathcal{H}$, either $\nu(C) = \infty$ almost surely when $\mathsf{E}\nu(C) = \infty$, or $\nu(C) \sim \mathrm{Poi}\big(\lambda(C)\big)$ where $\lambda(C) := \mathsf{E}\nu(C) < \infty$;*
2. *for any disjoint $C_1, C_2, \cdots, C_n$, $\nu(C_1)$, $\nu(C_2), \cdots, \nu(C_n)$ are independent Poisson random variables.*

The measure $\lambda(C) := \mathsf{E}\nu(C)$ is called the intensity measure *of the PPP.*

The distribution of ν is completely determined by its f.d.d.

$$\mathcal{L}\big(\nu(C_1), \nu(C_2), \cdots, \nu(C_n)\big), \quad C_1, C_2, \cdots, C_n \in \mathcal{H}.$$

Example 12.3 (Poisson process) Let $\lambda > 0$. A Poisson process with intensity λ is defined by

$$N_t = \max\{k : \xi_1 + \xi_2 + \cdots + \xi_k \le t\},$$

where ξ_i are i.i.d. $\mathrm{Exp}(\lambda)$ r.v.'s. The process N_t is counting how many independent exponential clocks have rang in the interval $[0, t]$. The times when these clocks ring form a PPP on $[0, \infty)$ with intensity measure $\lambda\, dt$. We can write $N_t = \nu([0, t])$.

Example 12.4 (Compound poisson) Let η_i be i.i.d. r.v.'s and N_t be a Poisson process with intensity λ. We define

$$Z_t(\omega) = \sum_{n=1}^{N_t} \eta(k).$$

We can represent Z_t as

$$Z_t = \sum_{0 \le s \le t,\ (s,\ell) \in \operatorname{supp} \nu} \ell = \int_{[0,t] \times \mathbb{R}_+} \ell \nu(ds d\ell),$$

where ν is a PPP on $[0, \infty) \times \mathbb{R}_+$ constructed as follows: for each $s = \xi_1 + \cdots + \xi_k$, place a point at (s, η_k), where η_k are i.i.d. random variables. In fact, ν is a PPP with intensity $\lambda dt \otimes \pi$, where π is the common distribution of η_i.

It is standard to compute the Laplace transform of the compound Poisson r.v. Z_t. We have

$$\begin{aligned}
\mathsf{E}e^{-\alpha Z_t} &= \mathsf{E}e^{-\alpha \sum_{n=1}^{N_t} \eta_n} \\
&= \mathsf{E}\Big(\mathsf{E}e^{-\alpha\eta}\Big)^{N_t} =: \mathsf{E}a^{N_t} \\
&= \sum_{k=0}^{\infty} \frac{a^k(\lambda t)^k}{k!} e^{-\lambda t} \\
&= e^{-\lambda t(1-a)} = \exp\Big(-\lambda t \int (1-e^{\alpha\ell})\pi(d\ell)\Big).
\end{aligned}$$

In general, we can consider any integrable function $f(s,\ell)$ instead of $\ell \mathbb{1}_{\{s\le t\}}$ and obtain

$$\mathsf{E}e^{-\alpha \int f(s,\ell)\nu(dsd\ell)} = \exp\Big(-\int (1-e^{-\alpha f(s,\ell)})\lambda(dsd\ell)\Big), \tag{12.16}$$

where λ is intensity measure of ν.

Now we return to our study of S_b. Consider the random measure

$$\nu_\omega = \sum_{j(b)>0} \delta_{(b,j(b))}. \tag{12.17}$$

For $\Gamma \in \mathcal{B}(0,\infty)$, we also define

$$N_t^\Gamma = \sum_{b<t} \mathbb{1}_\Gamma\big(j(b)\big) = \nu\big([0,t]\times\Gamma\big).$$

Proposition 12.18 *For any Γ, N_t^Γ has stationary and independent increments.*

Proof: Note that $N_t^\Gamma - N_s^\Gamma$ depends only on $\big(\beta_r\big)_{r\ge S_t}$. The statement follows from the fact that S_t is a stopping time and the strong Markov property for the Brownian motion β. □

Our goal is to show that ν is a PPP. We can identify the intensity measure (in variable ℓ) as

$$\rho(\Gamma) := \frac{1}{t}\mathsf{E}N_t^\Gamma.$$

The definition of ρ is independent of t by Proposition 12.18. Also, ρ is a measure on $\mathcal{B}(0,\infty)$.

Theorem 12.19 *The random measure ν_ω is a PPP on $(0,\infty)^2$ with intensity $dt\otimes\rho$.*

We need a lemma.

Lemma 12.20 *Let $\Phi : (\omega,t,\ell) \to \mathbb{R}_+$ be predictable. Then*

$$\mathsf{E}\int \Phi(\omega,t,\ell)\nu(dtd\ell) = \int_0^\infty dt\cdot \mathsf{E}\int \Phi(\omega,t,\ell)\rho(d\ell).$$

Here, the *predictable σ-field* is the smallest σ-algebra such that all left-continuous in t map $\varphi(\omega,t)$ are measurable; Φ is predictable if $(\omega,t)\mapsto\Phi(\omega,t,\ell)$ is measurable w.r.t. the predictable σ-algebra. Below we will use such a property: if $\varphi(\omega,t)$ is predictable and M_t is a right-continuous martingale, then

$$\int_0^t \varphi(\omega,t)\,dM_t$$

is well-defined and is also a martingale.

Proof: It suffices to check for Φ taking the form $\Phi(\omega,t,\ell)=\varphi(\omega,t)\mathbb{1}_\Gamma(\ell)$ where φ is predictable and $\Gamma\in\mathcal{B}(0,\infty)$. Then $N_t^\Gamma-t\rho(\Gamma)$ is a right-continuous martingale, and hence

$$0=\mathsf{E}\int_0^\infty \varphi(\omega,t)(dN_t^\Gamma-\rho(\Gamma)dt) \quad\Rightarrow\quad \mathsf{E}\int\Phi(\omega,t,\ell)\nu(dtd\ell)=\rho(\Gamma)\int_0^\infty dt\,\mathsf{E}\varphi(\omega,t).$$

□

Proof of Theorem 12.19: We will show (c.f. (12.16)) for every $f(t,\ell)\geq 0$ with

$$\int_0^\infty dt\int f(t,\ell)\,d\ell<\infty,$$

we have

$$\mathsf{E}\exp\Big(-\int f(s,\ell)\mathbb{1}_{[0,t]}(s)\nu_\omega(dsd\ell)\Big)=\exp\Big(-\int_0^t ds\int(1-e^{f(s,\ell)})\rho(d\ell)\Big). \tag{12.18}$$

This will implies ν_ω has the same f.d.d. distribution as a PPP with intensity measure $dt\otimes\rho$.

Let

$$X_t=\int f(s,\ell)\mathbb{1}_{[0,t]}\nu_\omega(dsd\ell)=\sum_{s\leq t,\ (s,\ell)\in\operatorname{supp}\nu_\omega} f(s,\ell).$$

The

$$\begin{aligned}H(t):=\mathsf{E}e^{-X_t}-1&=\sum_{s\leq t,}\mathsf{E}e^{-X_s}-e^{-X_{s-}}\\&=\mathsf{E}\sum_{s\leq t,\ (s,\ell)\in\operatorname{supp}\nu_\omega}e^{-X_{s-}}\big(e^{-f(s,\ell)}-1\big)\\&=\mathsf{E}\int_0^t ds\,e^{-X_{s-}}\int(e^{-f(s,\ell)}-1)\rho(d\ell).\end{aligned}$$

In the last equality, we use Lemma 12.20, and the predictability of the functional in the last but one line follows from the strong Markov property of β.

Write $G(t)=\displaystyle\int(e^{-f(t,\ell)}-1)\rho(d\ell)$. We have

$$H(t)=1+\int_0^t H(s-)G(s)\,ds, \tag{12.19}$$

and $G\in L^1(\mathbb{R})$ by our assumption on f. Since $|H(s)|\leq 1$, the integrand in (12.19) is also L^1 and hence $H(t)$ is continuous, and $H(s-)=H(s)$. Then the integral equation has a unique solution $H(t)=\exp(G(t))$, and this completes the proof. □

We know very precise information about the distribution of S_b.

$$\mathsf{P}(S_b\in dt)=\frac{b}{\sqrt{2\pi t^3}}e^{-\frac{b^2}{2t}},\quad \mathsf{E}e^{-S_b}=e^{-\sqrt{2}b}.$$

From either of them, we can compute $\rho(d\ell)=\dfrac{d\ell}{\sqrt{2\pi\ell^3}}$, $\ell>0$.

Now we will give another description of the local time using only the set $\mathcal{Z}=\mathcal{Z}_0^B(\omega)$. Let

$$\begin{aligned}N_b^{\delta,\varepsilon}&=\nu\big([0,b]\times[\varepsilon,\infty)\big)\sim\operatorname{Poi}\Big(\frac{2}{\pi}\frac{b}{\sqrt{\varepsilon}}\Big)\\&=\#\text{ of jumps of size }\ell\geq\varepsilon\text{ for }(S_a)_{0\leq a\leq b}.\end{aligned}$$

Proposition 12.21 *Almost surely, for all $b \geq 0$,*

$$\lim_{\varepsilon\downarrow 0} \frac{\pi\varepsilon}{2} N_b^\varepsilon = b.$$

Proof: It suffices to prove it for a fixed $b \geq 0$. Then it simultaneously holds for all $b \in \mathbb{Q}$, and the conclusion follows from the monotonicity of $b \mapsto N_b^\varepsilon$.

Let $Q_t = \nu([0,b] \times [t^{-2}, \infty))$. Then Q_t is a Poisson process with parameter $\mathsf{E}Q_1 = \sqrt{\frac{2}{\pi}}b$. Let $\xi_1, \xi_2, \ldots$ be the exponential r.v.'s that build (Q_t). For $Q_t = n$, we have

$$\frac{n}{\xi_1 + \xi_2 + \cdots + \xi_n + \xi_{n+1}} \leq \frac{Q_t}{t} \leq \frac{n}{\xi_1 + \xi_2 + \cdots + \xi_n}.$$

By strong law of large numbers, we have almost surely,

$$\lim_{t\to\infty} \frac{Q_t}{t} = \frac{1}{\mathsf{E}\xi} = \sqrt{\frac{2}{\pi}}b.$$

Hence,

$$\lim_{\delta\downarrow 0} \frac{N_b^\delta}{\sqrt{1/\delta}} = \sqrt{\frac{2}{\pi}}b.$$

This completes the proof. $\square$

As corollary, we obtain a description of the local time using only the zero set:

$$\begin{aligned}
L_t^B(0) = \beta^*(t) &= \lim_{\varepsilon\downarrow 0} \frac{\pi\varepsilon}{2} \nu\big((0, \beta_t^*] \times [\varepsilon, \infty)\big) \\
&= \lim_{\varepsilon\downarrow 0} \frac{\pi\varepsilon}{2} \cdot \#\{ \text{ jumps} \geq \varepsilon \text{ made by } (S_b)_{b \leq \beta_t^*} \} \\
&= \lim_{\varepsilon\downarrow 0} \frac{\pi\varepsilon}{2} \cdot \#\{ \text{ excursion interval} \geq \varepsilon \text{ made by } (B_s)_{s \leq S_{\beta_t^*}} \} \\
&= \lim_{\varepsilon\downarrow 0} \frac{\pi\varepsilon}{2} \cdot \#\{ \text{ excursion interval} \geq \varepsilon \text{ made by } B \text{ before time } t \}.
\end{aligned}$$

12.4 Ray–Knight Theorem

In this section we state the Ray–Knight Theorems, which give information of the joint distribution of local time at different levels.

The *square Bessel process*, denoted by $\text{BESQ}^\delta(x)$, is the unique strong solution to the SDE

$$Z_t = x + \int_0^t \sqrt{Z_s}\, dW_s + \delta t.$$

Note that the pathwise uniqueness holds for this SDE since $\int_0^1 \left(\frac{1}{\sqrt{x}}\right)^2 dx = \infty$. If $\delta \in \mathbb{Z}_+$, then $Z_t = \big(B_t^{(1)}\big)^2 + \cdots + (B_t^{(\delta)})^2$ is a solution. If $\delta = 0$, we do not have such representation, but it is easy to see that $Z_t \equiv 0$ is always a solution. So pathwise uniqueness implies that 0 is an absorbing state, i.e., if $Z_{t_0} = 0$ for some t_0, then $Z_t = 0$ for all $t \geq t_0$.

Theorem 12.22 (First Ray–Knight Theorem) *Let B be Brownian motion and $T_s = \inf\{t : B_t = 1\}$. Let $Z_a = L_{T_1}^{1-a}(B)$. Then $(Z_a, 0 \leq a \leq 1)$ is $\text{BESQ}^2(0)$.*

Intuitively, after the Brownian motion hits 1, the local time L^a should increase as a decreases from 1. This monotonicity is captured by the BES$Q^2(0)$ process, which is a sub-martingale starting from 0.

Theorem 12.23 (Second Ray–Knight Theorem) *Let $\tau_x = \inf\{t : L^0_t(B) \geq x\}$. Then $(L^a_{\tau_x}(B), a \geq 0)$ is BES$Q^0(x)$.*

The second Ray–Knight Theorem views the local time using the a "different clock", i.e., the local time at 0.

13 Stratonovich integral

Let us consider the SDE

$$X_t = \xi + \int_0^t b(X_s)\, ds + \int_0^t \sigma(X_s)\, dW_s.$$

For simplicity, in this section we assume that b is bounded Lipschitz, and $\sigma \in \mathcal{C}^2_b$ (σ, σ', σ'' bounded continuous).

Note that if $s \mapsto W_s$ has finite variation, the we can define the last integral as a Riemann–Stieltjes integral and the integral equation can make sense. But we know that Brownian motion does not have finite variation, and that is the reason why we need to develop the theory of stochastic integral with the help of martingale theory. However, this is not the only approach; in problems such as singular SPDEs and Gaussian free field, Liouville quantum gravity, etc, it is common to smooth a rough object like the Brownian motion and study the limit when the smoothing is removed.

In our context, it is natural ask the following question. If $W^N \to W$ locally uniformly and W^N has finite variation, and X^N solves the integral equation

$$X^N_t = \xi + \int_0^t b(X^N_s)\, ds + \int_0^t \sigma(X^N_s) dW^N_s,$$

what can be said about the limit $\lim\limits_{N\to\infty} X^N$?

The answer is that $X_N \to \tilde{X}$ where $\tilde{X}$ solves

$$\tilde{X}_t = \xi + \int_0^t b(\tilde{X}_s)\, ds + \int_0^t \sigma(\tilde{X}_s) \circ dW_s. \tag{13.1}$$

The $\circ$ symbol denotes the *Stratonovich integral*, which is defined as follows: if Y, Z are two continuous semi-martingale, then

$$\int_0^t Y_s \circ dZ_s = \int_0^t Y_s\, dZ_s + \frac{1}{2}\langle Y, Z\rangle_t, \tag{13.2}$$

where the first integral is the Itô integral. We will rigorously establish this approximation result in one dimension in Section 13.1.

A notable feature for the Stratonovich integral is that normal chain rule holds, i.e., without Itô's correction term:

$$f(X_t) = \int_0^t \sum_i (\partial_i f)(X_t) \circ dX^{(i)}_t, \quad X^{(i)}_t \text{ continuous semi-martingale}, \quad f \in \mathcal{C}^3. \tag{13.3}$$

The $\mathcal{C}^3$ condition is unnatural since the chain rule only involves the first derivative. The assumption cannot be removed here if we rely on (13.2), since we need to apply Itô's formula to $\partial_i f(X)$ to obtain their cross variation.

In Section 13.2 we will see another definition of the Stratonovich integrals, with which we can establish the chain rule under the assumption $f \in \mathcal{C}^1$. The alternative approach also explains why $1/2$ comes out from smooth approximation but not other factors $1/3$, $2/3$, etc.

13.1 One-dimension result

Let V_t be a function with bounded variation. Consider the integral equation

$$X_t = x + \int_0^t b(X_s)\, ds + \int_0^t \sigma(X_s)\, dV_s. \tag{13.4}$$

Proposition 13.1 *There exists a unique solution to* (13.4)*, and is given by a continuous functional*

$$\Phi : \mathbb{R} \times \mathcal{C}(\mathbb{R}) \to \mathcal{C}(\mathbb{R}), \quad (x, (V_t)_{t\geq 0}) \mapsto (X_t)_{t\geq 0}.$$

For the uniqueness, one just looks at the difference of two solutions $|X_t^1 - X_t^2|$ and finds a way to apply the Gronwall's inequality. Let us illustrate how to solve the integral equation and construct the functional Φ.

To motivate, let us look at the simplest case where $\sigma(X_t) \equiv c$. To solve the equation

$$dX_t = b(X_t)\, dt + c dV_t,$$

a simple idea is to define the substitution $X_t = Y_t + cV_t$, and then solves

$$dY_t = b(Y_t + cV_t)\, dt$$

which is a normal ODE. The gain here is for the Y-ODE, the dependence on V_t is no longer through the Riemann–Stieltjes integral dV_t, but V_t becomes a part of the ODE coefficient.

In the general case, let $u(x, y)$ solves

$$\partial_x u(x, y) = \sigma\big(u(x, y)\big), \quad u(0, y) = y,$$

and let Y_t solves

$$dY_t = f(V_t, Y_t), \quad Y_0 = x, \quad f(x, y) = \frac{1}{\partial_y u(x, y)} b\big(u(x, y)\big).$$

The solution to (13.4) is then given by $X_t = u(V_t, Y_t)$. It is not hard to check the continuous dependency of X_t on x, V.

To see that X_t is indeed a solution, we have

$$\begin{aligned} dX_t &= \partial_x u(V_t, Y_t)\, dV_t + (\partial_y u)(V_t, Y_t)\, dY_t \\ &= \sigma\big(u(V_t, Y_t)\big)\, dV_t + b\big(u(V_t, Y_t)\big)\, dt \end{aligned}$$

as desired.

Now what is $X_t = \Phi(x, W_t)$ where W_t is a Brownian motion? We have

$$\begin{aligned} dX_t &= \partial_x u(W_t, Y_t)\, dW_t + \frac{1}{2}\partial_{xx} u(W_t, Y_t)\, dt + \partial_y u(W_t, Y_t) f(W_t, Y_t)\, dt \\ &= \sigma(X_t)\, dW_t + b(X_t)\, dt + \frac{1}{2}\partial_{xx} u(W_t, Y_t)\, dt. \end{aligned}$$

Note that $\partial_{xx} u = (\partial_x u) \cdot \partial_x \sigma$ and

$$\sigma(X_t) = \int \sigma'(X_t)\, dX_t + \frac{1}{2}\int \sigma''(X_t)\, d\langle X\rangle_t,$$

so

$$d\langle \sigma(X.), W\rangle_t = \sigma'(X_t)\sigma(X_t)\, dt.$$

Therefore, $X_t = \Phi(x, W_t)$ solves (13.1).

In higher dimension, the integral equation (13.4) no longer has such nice presentation as Proposition 13.1. The approximation result only holds for linear interpolation.

13.2 Alternative construction of Stratonovich integrals

The goal of this section is to give an alternative construction of the Stratonovich integral so that the chain rule (13.3) holds for $f \in \mathcal{C}^1$.

Fix $T > 0$. For $N \geq 1$ and a process Z, we define the linear interpolation process

$$Z_t^N = \begin{cases} Z_t, & t = 2^{-N} \cdot kT, \\ \text{linear interpolation between} Z_{2^{-N}kT}, Z_{2^{-N}(k+1)T}, & t \in (2^{-N}kT, 2^{-N}(k+1)T). \end{cases}$$

Note that Z^N will have bounded variation, and for continuous semi-martingales X, Y, we have

$$\begin{aligned} \int_0^t Y_s^N dX_s^N &= \sum_i \frac{Y_{t_{i+1}} + Y_{t_i}}{2}(X_{t_{i+1}} - X_{t_i}) \\ &= \sum_i Y_{t_i}(X_{t_{i+1}} - X_{t_i}) + \sum_i \frac{Y_{t_{i+1}} - Y_{t_i}}{2}(X_{t_{i+1}} - X_{t_i}) \\ &\to \int_0^t Y_s\, dX_s + \frac{1}{2}\langle X, Y\rangle_t, \end{aligned}$$

where $t_i = i \cdot 2^{-N}T$.

This motivates the following definition. Let X be a continuous semi-martingale. We say that

$$Y_t(\omega) : [0, T] \times \Omega \to \mathbb{R}$$

is *Stratonovich integrable* w.r.t. X if and only if the limit

$$\lim_{N\to\infty} \sum_{m=0}^{2^N-1} \frac{Y_{t_{i+1}} + Y_{t_i}}{2}(X_{t_{i+1}} - X_{t_i}) =: \int_0^T Y_t \circ dX_t$$

exists in probability. Note that here $Y_t(\omega)$ does not have to be a semi-martingale.

To perform calculation we still need to use the Itô's theory. Let $\check{X}_t^T = X\big((T-t)_+\big)$ and assume that $\check{X}^T$ is a continuous semi-martingale w.r.t. some filtration $\check{\mathcal{F}}_t^T$. If $Y = (Y_t(\omega)) \in \mathcal{C}[0, T]$ a.s. and $Y_t(\omega) \in \mathcal{F}_t \cap \check{\mathcal{F}}_{T-t}^T$, then Y is Stratonovich integrable on $[0, T]$ w.r.t. X and

$$\int_0^T Y_t \circ dX_t = \frac{1}{2}\int_0^T Y(t)\, dX_t - \frac{1}{2}\int_0^T Y(T-t)\, d\check{X}_t^T.$$

Here, $Y \in \mathcal{C}[0, T]$ so Y_t is both $(\mathcal{F}_t)$-progressively measurable and Y_{T-t} is $\check{\mathcal{F}}_t^T$-progressively measurable, and hence both Itô integrals make sense.

Now let W_t, $t \in [0, T]$ be Brownian motion. Clearly, $\check{W}_t = W\big((T-t)_+\big)$ is semi-martingale, with

$$\langle \check{W}\rangle_t = t \wedge T, \quad \text{finite variation part } V_t = -\int_0^t \frac{\check{W}_s^T}{T-s}\, ds.$$

Here, the finite variation part is non-zero since the terminal condition $\check{W}_T = W_0 = 0$ is enforced. Then, $g(W_t)$ is Stratonovich integrable on $[0, T]$ w.r.t. W for every $g \in \mathcal{C}(\mathbb{R})$. Moreover, we have the following approximation result for Stratonovich integrals.

Proposition 13.2 *If $g_n, g \in \mathcal{C}_b(\mathbb{R})$ such that $g_n \to g$ locally uniformly, then*

$$\int_0^T g_n(W_t) \circ dW_t \to \int_0^T g(W_t) \circ dW_t$$

in probability.

Proof: Let $W^* = \sup_{0 \le t \le T} |W_t| = \sup_{0 \le t \le T} |\check{W}_t^T|$. We have

$$\begin{aligned}
&\mathsf{E}\Big|\int_0^T g_n(W_t)\, dW_t - \int_0^t g(W_t) dW_t\Big| \le \Big[\mathsf{E}\int_0^T |g_n(W_t) - g(W_t)|^2\, dt\Big]^{1/2} \\
&\le \Big[\mathsf{E}\mathbb{1}_{W^* \le M}\int_0^T |g_n(W_t) - g(W_t)|^2\, dt + \mathsf{E}\mathbb{1}_{W^* \ge M}\int_0^T |g_n(W_t) - g(W_t)|^2\, dt\Big]^{1/2}.
\end{aligned}$$

By standard argument we can show that the expectation on the left-hand side converges to 0 as $n \to \infty$.

It is similar for the other integral, but now $d\check{W}_t = d\check{M}_t - \dfrac{\check{W}_t}{T-t}\, dt$. □

Finally, we are ready to prove (13.3) for X being a multi-dimensional Brownian motion, that is,

$$f(W_T) - f(W_0) = \sum_{i=1}^d \int_0^T (\partial_i f)(W_t) \circ dW_t^{(i)}, \quad f \in \mathcal{C}^1. \tag{13.5}$$

Let $f_n \in \mathcal{C}_b^2$ such that $f_n' \to f'$ locally uniformly. We have

$$\begin{aligned}
\sum_{i=1}^d \int_0^T (\partial_i f_n)(W_t) \circ dW_t^{(i)} &= \frac{1}{2}\sum_{i=1}^d \int_0^T (\partial_i f_n)(W_t)\, dW_t^{(i)} - \frac{1}{2}\sum_{i=1}^d \int_0^T (\partial_i f_n)(\check{W}_t)\, d\check{W}_t^{(i)} \\
&= \frac{1}{2}\Big[f_n(W_T) - f_n(W_0) - \sum_{i,j}\frac{1}{2}\int_0^T \partial_{ij} f_n(W_t)\, dt\Big] \\
&\quad - \frac{1}{2}\Big[f_n(\check{W}_T) - f_n(\check{W}_0) - \sum_{i,j}\frac{1}{2}\int_0^T \partial_{ij} f_n(\check{W}_t)\, dt\Big] \\
&= f_n(W_T) - f_n(W_0).
\end{aligned}$$

Note that the two Itô correction terms cancel since $\check{W}$ is just the time-reverse of W. This also explains why $1/2$ appears in front of $\langle X, Y\rangle_t$ in the definition of Stratonovich integral: this is the only factor so that the Itô correction terms in the forward and backward integrals can cancel. Letting $f_n \to f$, the chain rule (13.5) follows from Proposition 13.2.

14 Notations

14.1 Abbreviations

i.i.d.	independent, identically distributed
r.v.	random variable
f.d.d.	finite-dimensional distribution
ch.f.	characteristic function
u.i.	uniformly integrable
c.l.m.	continuous local martingale
m.t.	martingale term(s)

14.2 Sets

$\mathbb{Z}$	set of integers
$\mathbb{N}$	set of natural numbers $\{0, 1, 2, \ldots\}$
$\mathbb{Q}$	set of rational numbers
$\mathbb{R}$	set of real numbers
$\mathbb{R}_+$ (resp. $\mathbb{R}_-$)	set of non-negative (resp. non-positive) real numbers

14.3 Relations

$\Rightarrow_d$ or $\Rightarrow$	convergence in distribution/law
$\stackrel{\mathrm{d}}{=}$	equal in law

14.4 Functional spaces

$\mathcal{C}[a, b]$	continuous function defined on the interval $[a, b]$
$\mathcal{C}^\alpha[a, b]$	α-Hölder continuous function defined on the interval $[a, b]$
$\mathcal{M}(E)$	probability measures on a metric space E

14.5 Operations

$a \wedge b$	$\min(a, b)$
$a \vee b$	$\max(a, b)$
$\langle a, b\rangle$	inner product in a Euclidean space/Hilbert space (or) a linear functional a in the dual space $\mathcal{X}^*$ acting on an element b in a Banach space $\mathcal{X}$
$A\Delta B = (A \setminus B) \cup (B \setminus A)$	the difference set.

14.6 Miscellaneous

$\mathcal{L}(X)$	distribution/law of a random variable/element X
$\mathcal{N}(\mu, \sigma^2)$	normal distribution
$\mathrm{Exp}(\lambda)$	exponential distribution
$\mathrm{Poi}(\lambda)$	Poisson distribution

References

[Bil99] Patrick Billingsley. *Convergence of Probability Measures.* Wiley Series in Probability and Statistics. Probability and Statistics. Wiley, New York, 2nd ed edition, 1999.

[Chu74] Kai Lai Chung. *A Course in Probability Theory.* Probability and Mathematical Statistics, 21. Academic Press, New York, 2d ed edition, 1974.

[Dur07] Richard Durrett. *Probability: Theory and Examples.* Duxbury Advanced Series. Thomson Brooks/Cole, Belmont, Calif, 3. ed. edition, 2007.

[EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence.* Wiley Series in Probability and Statistics. Wiley, 1 edition, March 1986.

[Hai] Martin Hairer. An Introduction to Stochastic PDEs. `https://www.hairer.org/notes/SPDEs.pdf`.

[KS98] Ioannis Karatzas and Steven Shreve. *Brownian Motion and Stochastic Calculus.* Graduate Texts in Mathematics. Springer-Verlag, New York, 2 edition, 1998.

[LeG16] Jean-François LeGall. *Brownian Motion, Martingales, and Stochastic Calculus*, volume 274 of *Graduate Texts in Mathematics.* Springer International Publishing, Cham, 2016.

[PZ14] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic Equations in Infinite Dimensions.* Cambridge University Press, 2014.

[Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics.* Springer New York, New York, NY, 1996.