Stochastic Calculus

Based on lecture notes by Prof. Xunyu Zhou

Michaelmas Term, 2015

Contents

0 Conventions and abbreviations **2**

1 Preliminaries **2**

1.1 Probability Space 2

1.2 Convergence and Uniform Integrability 4

1.3 Stochastic Processes and Filtration 6

1.4 Martingales 7

2 Brownian Motion **9**

2.1 Definition 9

2.2 Markov property and finite-dimensional distributions 10

2.3 The reflection principle and martingale property 12

2.4 Quadratic variation process 13

3 Itô's calculus **15**

3.1 Stochastic integrals for BM 15

3.2 Stochastic integrals for adapted processes 16

3.3 Itô's integration for semimartingales 17

3.4 Itô's formula 18

3.5 Stochastic exponentials and Girsanov theorem 20

4 Stochastic Differential Equations **23**

4.1 Introduction 23

4.2 Several examples 24

4.2.1 Linear SDEs 24

4.2.2 Geometric BM 24

4.2.3 Cameron-Martin's formula 24

4.3 Existence and uniqueness 25

0 Conventions and abbreviations

Vectors are column vectors.

a.s.	almost surely
BM	Brownian motion
CDF	cumulative distribution function
r.v.	random variable / random vector
s.t.	such that
PDF	probability density function
SDE	stochastic differential equation
w.r.t.	with respect to

1 Preliminaries

1.1 Probability Space

Definition 1.1.1. *Let a set Ω be non-empty and $\mathcal{F} \subseteq 2^{\Omega}$, called a class, be non-empty. We call $\mathcal{F}$*

(i) a π-system if $A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}$.

(ii) a λ-system if

$$\begin{cases} \Omega \in \mathcal{F} \\ A, B \in \mathcal{F} \text{ and } A \subseteq B \implies B \backslash A \in \mathcal{F} \\ A_i \in \mathcal{F}, A_i \uparrow A, i = 1, 2, \ldots^{1} \implies A \in \mathcal{F} \end{cases}$$

(iii) a σ-field (or σ-algebra) if

$$\begin{cases} \Omega \in \mathcal{F} \\ A, B \in \mathcal{F} \implies B \backslash A \in \mathcal{F} \\ A_i \in \mathcal{F}, i = 1, 2, \ldots \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F} \end{cases}$$

Remark. $\mathcal{F}$ is a σ-field if and only if $\mathcal{F}$ is both a π-system and a λ-system.

For any class $\mathcal{A} \subseteq 2^{\Omega}$, $\sigma(\mathcal{A})$ is the smallest σ-field containing $\mathcal{A}$, which is called the σ-field *generated* by $\mathcal{A}$.

Lemma 1.1.2 (Monotone Class Theorem)**.** *Let $\mathcal{A} \subseteq \mathcal{F} \subseteq 2^{\Omega}$.*

If $\mathcal{A}$ is a π-system and $\mathcal{F}$ is a λ-system, then $\sigma(\mathcal{A}) \subseteq \mathcal{F}$.

Definition 1.1.3. *Let $\{\mathcal{F}_\alpha\}$ be a (possibly uncountable) family of σ-fields on Ω. Then we define*

$$\bigvee_{\alpha} \mathcal{F}_\alpha := \sigma\left(\bigcup_{\alpha} \mathcal{F}_\alpha\right) \qquad \textit{the smallest } \sigma\textit{-field} \text{ containing } \textit{all } \mathcal{F}_a;$$

$$\bigwedge_{\alpha} \mathcal{F}_\alpha := \bigcap_{\alpha} \mathcal{F}_\alpha \qquad \textit{the largest } \sigma\textit{-field} \text{ contained } \textit{in } \mathcal{F}_\alpha.$$

Definition 1.1.4. *Let Ω be non-empty and let $\mathcal{F}$ be a σ-field on Ω.*

Then we call $(\Omega, \mathcal{F})$ a measurable space.

A map $\mathbf{P} : \mathcal{F} \to [0, 1]$ is called a probability measure *on $(\Omega, \mathcal{F})$ if*

(i) $\mathbf{P}(\Omega) = 1, \mathbf{P}(\emptyset) = 0$;

(ii) $\forall A_i \in \mathcal{F}$ with $A_i \cap A_j = \emptyset\ \forall j \neq i \implies \mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(A_i)$.

Then we call $(\Omega, \mathcal{F}, \mathbf{P})$ a probability space*.*

We call Ω a sample space*, $A \in \mathcal{F}$ an* event *and $\omega \in \Omega$ a* sample *or* outcome*.*

[1] $A_i \uparrow A$ means $A_1 \subseteq A_2 \subseteq \cdots$ and $A = \bigcup_{i=1}^{\infty} A_i$.

Definition 1.1.5 (Independence)**.** *Two events A, B are called* independent *if* $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.
An event A is independent of a σ-field $\mathcal{F}$ if A is independent of B $\forall B \in \mathcal{F}$.
Two σ-fields $\mathcal{F}, \mathcal{G}$ are independent if all $A \in \mathcal{F}$ are independent of $\mathcal{G}$.

If an event A is s.t. $\mathbf{P}(A) = 1$, we may denote it as "A holds, $\mathbf{P}$-a.s." or just a.s. (almost surely) if the probability measure in question is clear.
A is called $\mathbf{P}$-null if $\mathbf{P}(A) = 0$.

Definition 1.1.6. $(\Omega, \mathcal{F}, \mathbf{P})$ *is called* complete *if for any $\mathbf{P}$-null set $A \in \mathcal{F}$, we have $B \in \mathcal{F}$ whenever $B \subseteq A$.*

Definition 1.1.7. *Given $(\Omega, \mathcal{F}, \mathbf{P})$.*
Let $X : \Omega \to \mathbb{R}^d$ be measurable, i.e. for every Borel measurable set B of $\mathbb{R}^d$,

$$X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F}.$$

Then X is called a random variable *(or* random vector*, if $d > 1$).*

If X is integrable, denoted as $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$, i.e. $\int_\Omega |X(\omega)| d\mathbf{P}(\omega) < \infty$, then its *mean* (or *expectation*) is

$$E[X] := \int_\Omega X(\omega) d\mathbf{P}(\omega).$$

In general, for $p \geq 1$, X is called p-th integrable, denoted as $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbf{P})$ if $\int_\Omega |X(\omega)|^p d\mathbf{P}(\omega) < \infty$.

Definition 1.1.8. *If $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$ then* $\mathrm{Cov}(X, Y) := \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])^T]$.
We can then define the variance as $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.

Definition 1.1.9. *Let $X \equiv (X_1, X_2, \dots, X_d)^T$ be a r.v..*
We define the cumulative distribution function (CDF)

$$F(x) \equiv F(x_1, x_2, \dots, x_d) := \mathbf{P}(\{\omega : X_i(\omega) \leq x_i, i = 1, 2, \dots, d\}).$$

Definition 1.1.10. *If there exists an $f \in \mathcal{L}^1(\mathbb{R}^d)$ s.t.*

$$F(x) = \int_{-\infty}^{x_d} \cdots \int_{-\infty}^{x_1} f(y_1, \dots, y_d) dy_1 \dots dy_d,$$

then f is called the probability density function (PDF) *of F.*

If X is a one-dimensional r.v., then

$$\begin{aligned} \mathbf{E}[X] &= \int_{\mathbb{R}} x \; dF(x) = \int_{-\infty}^{\infty} x f(x) \; dx && \text{if } f \text{ exists} \\ \mathbf{E}[g(X)] &= \int_{\mathbb{R}} g(x) \; dF(x) = \int_{-\infty}^{\infty} g(x) f(x) \; dx && \text{if } f \text{ exists and } g : \mathbb{R} \to \mathbb{R} \text{ measurable} \end{aligned}$$

Example 1. If

$$f(x) = \left[(2\pi)^d |\Sigma|\right]^{-1/2} \exp\left[-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right], \; x \in \mathbb{R}^d$$

where $\mu \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d \times d}$ and $\Sigma^T = \Sigma > 0$[2], then we say that X is a *normal r.v.*, denoted $X \sim \mathrm{N}(\mu, \Sigma)$. X is also called *Gaussian*.

Definition 1.1.11. *Let X be an* integrable *r.v. on $(\Omega, \mathcal{F}, \mathbf{P})$ and $\mathcal{G} \subseteq \mathcal{F}$.*
The conditional expectation of $\mathbf{E}[X|\mathcal{G}]$ of X given $\mathcal{G}$ is a r.v. (unique up to a.s.) satisfying

(i) $\mathbf{E}[X|\mathcal{G}]$ *is measurable w.r.t. $\mathcal{G}$ (also called a $\mathcal{G}$-r.v.);*

[2] $\Sigma > 0$ means that Σ is positive definite, i.e. $x^T \Sigma x > 0$ $\forall x \in \mathbb{R}^d \backslash \{0\}$.

(ii) $\mathbf{E}[X; A] := \mathbf{E}[X\mathbb{1}_A] = \mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A]\ \forall A \in \mathcal{G}$.

Remark. (i) If X is $\mathcal{F}$-measurable and Y is $\mathcal{G}$-measurable then

$$\mathbf{E}[XY|\mathcal{G}] = Y\mathbf{E}[X|\mathcal{G}], \text{ a.s.,}$$

so long as both sides are well defined (cf. integrability requirement for a conditional r.v.).

(ii) Let $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$, then

$$\mathbf{E}[\mathbf{E}[X|\mathcal{G}_1]|\mathcal{G}_2] = \mathbf{E}[X|\mathcal{G}_1] = \mathbf{E}[\mathbf{E}[X|\mathcal{G}_2]|\mathcal{G}_1], \text{ a.s..}$$

(iii) Jensen's inequality. Suppose $\varphi : \mathbb{R}^d \to \mathbb{R}$ is a convex function[3], then

$$\varphi(\mathbf{E}[X|\mathcal{G}]) \leq \mathbf{E}[\varphi(X)|\mathcal{G}].$$

A σ-field $\sigma(X)$ generated by a r.v. X is the smallest σ-field contained in $\mathcal{F}$ w.r.t. which X is measurable. Similiarly, we can define $\sigma(X_\alpha : \alpha \in A)$[4], for a collection of random variables.
We define $\mathbf{E}[X|Y] := \mathbf{E}[X|\sigma(Y)]$.
X is independent of $\mathcal{G}$ if $\sigma(X)$ is independent of $\mathcal{G}$, which is equivalent to

$$\mathbf{E}[f(X)|\mathcal{G}] = \mathbf{E}[f(X)] \text{ for all bounded Borel-measurable functions } f.$$

Random variables $X_1, \ldots, X_n$ are mutually independent if $\sigma(X_1), \ldots, \sigma(X_n)$ are mutually independent, which is equivalent to

$$\mathbf{E}\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbf{E}[f_i(X_i)] \text{ for all bounded Borel-measurable functions } f_1, \ldots, f_n.$$

In particular, if X, Y are independent, then $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.

1.2 Convergence and Uniform Integrability

Definition 1.2.1. *Let X_n, $n = 1, 2, 3, \ldots$ and $X : (\Omega, \mathcal{F}, \mathbf{P}) \to \mathbb{R}^d$ be r.v.s. We say X_n converges to X*

- almost surely *(a.s.) if* $\lim_{n\to\infty} |X_n - X| = 0$ *a.s.,*
- in probability *if* $\forall\varepsilon > 0$, $\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$,
- in $\mathcal{L}^1$ *if* $\lim_{n\to\infty} \mathbf{E}|X_n - X| = 0$.

Remark.

$$\begin{aligned}
\text{a.s. convergence} &\implies \text{convergence in probability}\\
\mathcal{L}^1 \text{ convergence} &\implies \text{convergence in probability}\\
\text{convergence in probability} &\implies \text{there exists a subsequence that converges a.s.}
\end{aligned}$$

Lemma 1.2.2 (Fatou's Lemma)**.** *If $X_n \geq 0$[5], then*

$$\mathbf{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} \mathbf{E}[X_n].$$

Remark. 1. Fatou's lemma holds when $X_n \geq Y$ and $\mathbf{E}[Y] > -\infty$ (consider that $X_n - Y \geq 0$).

[3] A function f is convex if $f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)\ \forall x_1, x_2 \in \mathbb{R},\ \forall t \in [0, 1]$
[4] $\sigma(X_\alpha : \alpha \in A) = \sigma(\{\omega \in \Omega : X_\alpha(\omega) \in B\} : \alpha \in A, B \in \mathcal{B})$.
[5] a.s. is implied here and in all other inequalities involving r.v.s.

2. When $X_n \le 0$, applying the lemma to $-X_n$ gives us

$$\mathbf{E}\left[\limsup_{n\to\infty} X_n\right] \ge \limsup_{n\to\infty} \mathbf{E}[X_n].$$

Lemma 1.2.3 (Monotone Convergence Theorem)**.** *If $X_n \uparrow X$ and $\mathbf{E}[X_1] > -\infty$, then*

$$\mathbf{E}[X] = \mathbf{E}\left[\lim_{n\to\infty} X_n\right] = \lim_{n\to\infty} \mathbf{E}[X_n].$$

Lemma 1.2.4 (Dominated Convergence Theorem)**.** *If $X_n \to X$ a.s. and $|X_n| \le Y$ with $\mathbf{E}[Y] < \infty$, then*

$$\mathbf{E}[X] = \mathbf{E}\left[\lim_{n\to\infty} X_n\right] = \lim_{n\to\infty} \mathbf{E}[X_n].$$

$$\text{If } X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P}), \text{ then } \lim_{N\to\infty} \int_{\{|X|\ge N\}} |X| \, d\mathbf{P} = 0. \tag{1.2.1}$$

$$\text{If } X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P}), \text{ then } \int_A |X| \, d\mathbf{P} \to 0 \text{ as } \mu(A) \to \infty. \tag{1.2.2}$$

Definition 1.2.5. *Let $\Gamma \subseteq \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$ be given.*
Γ *is said to be* uniformly integrable (UI) *if*

$$\lim_{N\to\infty} \sup_{X\in\Gamma} \int_{\{|X|\ge N\}} |X| \, d\mathbf{P} = 0.$$

Remark.

1. Any finite family of integrable r.v.s is UI. (Note that because the family is finite, the supremum becomes a maximum.)
2. Let $\Gamma \subseteq \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$, and suppose there is $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$, s.t. $|X| \le Y \ \forall X \in \Gamma$, then Γ is UI by (1.2.1).
3. Let $\Gamma \subseteq \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$ and suppose $\exists p > 1$ s.t. $\sup_{X\in\Gamma} \mathbf{E}|X|^p < \infty$, then Γ is UI.

Theorem 1.2.6. *$\Gamma \subseteq \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$ is UI if and only if*

1. $\sup_{X\in\Gamma} \mathbf{E}|X| < \infty$,
2. $\forall \varepsilon > 0, \ \exists \delta > 0$ *s.t.* $\int_A |X| \, d\mathbf{P} < \varepsilon$, $\forall X \in \Gamma$, *whenever* $A \in \mathcal{F}$[6] *and* $\mathbf{P}(A) < \delta$.

[6] A here is like a more general form of $\{|X| \ge N\}$ in the definition of UI.

Proof. **Necessity**

For any $A \in \mathcal{F}$ and $N > 0$,

$$\begin{aligned}\int_A |X| \; d\mathbf{P} &= \int_{A\cap\{|X|<N\}} |X| \; d\mathbf{P} + \int_{A\cap\{|X|\geq N\}} |X| \; d\mathbf{P} \\ &\leq N\mathbf{P}(A\cap\{|X|<N\}) + \int_{A\cap\{|X|\geq N\}} |X| \; d\mathbf{P} \\ &\leq N\mathbf{P}(A) + \underbrace{\int_{A\cap\{|X|\geq N\}} |X| \; d\mathbf{P}}_{\text{c.f. UI definition}}.\end{aligned}$$

Let $A = \Omega$. Then, for N sufficiently large,

$$\begin{aligned}\mathbf{E}[X] &\leq NP(\Omega) + \int_{\{|X|\geq N\}} |X| \; d\mathbf{P} \\ &\leq N+1.\end{aligned}$$

For any $\varepsilon > 0$, choose $N > 0$ s.t.

$$\sup_{X\in\Gamma} \int_{\{|X|\geq N\}} |X| \; d\mathbf{P} \leq \frac{\varepsilon}{2}.$$

Let $\delta = \varepsilon/(2N)$. Then, so long as $P(A) \leq \delta$, we have

$$\int_{|A|} |X| \; d\mathbf{P} \leq N \times \delta + \frac{\varepsilon}{2} = N \times \frac{\varepsilon}{2N} + \frac{\varepsilon}{2} = \varepsilon, \; \forall X \in \Gamma.$$

Sufficiency Refer to exercise sheet 2, question 3. ☐

Theorem 1.2.7. *Let $\{X_n\} \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$. Then $\mathbf{E}|X_n - X| \to 0$ as $n \to \infty$ for some X if and only if $\{X_n\}$ is UI and $X_n \to X$ in probability as $n \to \infty$.*

Let $\{A_n\} \subseteq \mathcal{F}$. Define

$$\{A_n \text{ i.o.}\} \equiv \{A_n \text{ occurs infinitely often}\} \equiv \limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Lemma 1.2.8 (Borel-Cantelli Lemma).

1. *If $\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty$, then $\mathbf{P}(\{A_n \text{ i.o.}\}) = 0$,*
2. *If $\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty$ and $A_1, A_2, \ldots$ are independent, then $\mathbf{P}(\{A_n \text{ i.o.}\}) = 1$.*

1.3 Stochastic Processes and Filtration

Let T denote either $\mathbb{Z}^+ := \{0, 1, 2, \ldots\}$ or $\mathbb{R}^+ = [0, \infty)$.

Definition 1.3.1. *A family $X \equiv (X_t)_{t\in T}$ of r.v.s from $(\Omega, \mathcal{F}, \mathbf{P})$ to $\mathbb{R}^d$ is called a* stochastic process.

$(X_t)_{t\in T}$ is also a mapping from $T \times \Omega \to \mathbb{R}^d$, which is sometimes also called a *random function.*

For any $\omega \in \Omega$, the function $t \to X_t(\omega)$ is called a *sample path.* The process is called *continuous* (or càdlàg / RCLL[7]) if $T = [0, \infty)$ and the sample is continuous (resp. càdlàg / RCLL) for almost all $\omega \in \Omega$.

Let $X = (X_t)_{t\geq 0}$ be a stochastic process and $0 \leq t_1 < t_2 < \cdots < t_n$ be an arbitrary partition. The joint distribution of r.v.s $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$:

$$\mu_{t_1,\ldots,t_n}(dx_1, \ldots, dx_n) = \mathbf{P}(X_{t_1} \in dx_1, \ldots, X_{t_n} \in dx_n)$$

is called a *finite dimensional distribution* of $(X_t)_{t\geq 0}$.

[7] Right continuous with left limits.

Definition 1.3.2. *Two processes X_t and Y_t are* (stochastically) equivalent *if $X_t = Y_t$ a.s., $\forall t \in T$.*

For a measurable space $(\Omega, \mathcal{F})$, we introduce a montone family of σ-fields $\mathcal{F}_t \subseteq \mathcal{F}$, $\forall t \in T$, satisfying

$$\mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2} \ \forall 0 \leq t_1 \leq t_2, \ t_1, t_2 \in T.$$

Such a family is called a *filtration*, and $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_t)_{t \in T})$ is called a *filtered probability space.*
Set

$$\begin{aligned} \mathcal{F}_{t+} &:= \bigcap_{s>t} \mathcal{F}_s, \ \forall t \geq 0, \\ \mathcal{F}_{t-} &:= \bigcup_{s<t} \mathcal{F}_s, \ \forall t \geq 0. \end{aligned}$$

If $\mathcal{F}_{t+} = \mathcal{F}_t$ (resp. $\mathcal{F}_{t-} = \mathcal{F}_t$), then we say that the filtration is right (resp. left) continuous.
If $(X_t)_{t \in T}$ is a process, then $\mathcal{F}_t := \sigma\{X_s, \ s \in T, s \leq t\}$ is called the filtration generated by $(X_t)_{t \in T}$.
If $(X_t)_{t \geq 0}$ is càdlàg, then it generates a right continuous filtration.

Definition 1.3.3. *We say that $(\Omega, \mathcal{F}, \mathbf{P})$ satifies the* usual conditions *if it is complete, $\mathcal{F}_0$ contains all the $\mathbf{P}$-null sets in $\mathcal{F}$, and $(\mathcal{F}_t)_{t \geq 0}$ is right-continuous.*

Remark. In future, we always assume a filtered probability space satifies the usual conditions.

Definition 1.3.4. *Let $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_t)_{t \geq 0})$ be a filtered probability space, and $(X_t)_{t \geq 0}$ a process.*

1. *$(X_t)_{t \geq 0}$ is called* measurable *if the map $(t, \omega) \mapsto X_t(\omega)$ is $(\mathcal{B}(\mathbb{R}^+) \times \mathcal{F})/\mathcal{B}(\mathbb{R}^d)$-measurable.*
2. *$(X_t)_{t \geq 0}$ is called* $\mathcal{F}_t$-adapted *if $\forall t \geq 0$, the map $\omega \mapsto X_t(\omega)$ is $\mathcal{F}_t/\mathcal{B}(\mathbb{R}^d)$-measurable.*

1.4 Martingales

Let $\{X_n : n \in \mathbb{Z}\}$ be a discrete time process on a filtered probability space $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_n)_{n \in \mathbb{Z}^+})$.
If X_n is $\mathcal{F}_{n-1}$ measurable $\forall n \geq 1$ and X_0 is $\mathcal{F}_0$ measurable, then we say that $\{X_n\}$ is *predictable* or *previsible.*
On $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_n)_{n \in \mathbb{Z}^+})$, a measurable function $\tau : \Omega \to \mathbb{Z}^+ \cup \{+\infty\}$ is called a *stopping time* (or *random time*) *w.r.t. the filtration* $\{\mathcal{F}_n\}$ if $\{\omega : \tau(\omega) = n\} \in \mathcal{F}_n \ \forall n \in \mathbb{Z}^+$.
Given τ, the σ-field

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \leq n\} \in \mathcal{F}_n \ \forall n \in \mathbb{Z}^+\}$$

represents the information available up to the random time τ.

Definition 1.4.1. *Let $\{X_n\}$ be adapted to $(\mathcal{F}_n)$ on $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_n)_{n \in \mathbb{Z}^+})$. Assume that each $X_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$.*

1. *$\{X_n\}$ is called a* martingale *if $\mathbf{E}[X_{n+1} | \mathcal{F}_n] = X_n \ \forall n \in \mathbb{Z}^+$.*
2. *$\{X_n\}$ is called a* supermartingale *(resp.* submartingale*) if $\mathbf{E}[X_{n+1} | \mathcal{F}_n] \leq X_n$ (resp. $\mathbf{E}[X_{n+1} | \mathcal{F}_n] \geq X_n$).*

Theorem 1.4.2 (Doob's optional sampling theorem / optional stopping theorem)**.** *Let $\{X_n\}$ be a martingale (resp. supermartingale) and σ, τ be two bounded stopping times with $\sigma \leq \tau$[8]. Then $\mathbf{E}[X_\tau | \mathcal{F}_\sigma] = X_\sigma$ a.s. (resp. $\mathbf{E}[X_\tau | \mathcal{F}_\sigma] \leq X_\sigma$ a.s.).*

Theorem 1.4.3 (Kolmogrov's inequality)**.** *Let $\{X_n\}$ be a martingale and $X_n \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P}) \ \forall n \in \mathbb{Z}^+$. Then*

$$P\left(\max_{k \leq n} |X_k| \geq \lambda\right) \leq \frac{\mathbf{E}[|X_n|^2]}{\lambda^2}.$$

[8] Recall that this is a.s. because σ and τ are random.

Remark. Compare this with Markov's inequality,

$$P\left(\max_{k\leq n}|X_k|\geq\lambda\right)\leq\frac{\mathbf{E}[\max_{k\leq n}|X_k|^2]}{\lambda^2}.$$

Note that Kolmogrove's inequality is "sharper", because we have been able to take advantage of X_n being a martingale.

Theorem 1.4.4 (Convergence Theorem)**.** *1. Let $\{X_n\}$ be a supermartingale. If $\sup_n \mathbf{E}|X_n|<\infty$, then $\exists X_\infty\in\mathcal{L}^1(\Omega,\mathcal{F},\mathbf{P})$ s.t. $X_n\to X_\infty$ a.s. as $n\to\infty$. Moreover, if $\{X_n\}$ is non-negative, then $\mathbf{E}[X_\infty|\mathcal{F}_n]\leq X_n$ a.s. $\forall n$.*

2. Let $\{X_n\}$ be a martingale and also be UI. Then $\exists X_\infty\in\mathcal{L}^1(\Omega,\mathcal{F},\mathbf{P})$ s.t. $X_n\to\infty$ a.s. and in $\mathcal{L}^1$. Moreover, $X_n=\mathbf{E}[X_\infty|\mathcal{F}_n]$.

Remark. Applying Fatou's lemma (note that we have X_n non-negative), we have

$$\mathbf{E}[X_\infty|\mathcal{F}_n]=\mathbf{E}\left[\lim_{m\to\infty}X_m|\mathcal{F}_n\right]\leq\lim_{m\to\infty}\mathbf{E}[X_m|\mathcal{F}_n]=\lim_{m\to\infty}X_n=X_n.$$

Note that $\mathbf{E}[X_m|\mathcal{F}_n]\leq X_n$ is only true for $m>n$, but we're taking the limit as $m\to\infty$ so this isn't a problem.

Now we consider the continuous time case. Let $(\Omega,\mathcal{F},\mathbf{P};(\mathcal{F}_t)_{t\geq0})$ be a filtered probability space. A map $\tau:\Omega\to[0,\infty]$ is an $(\mathcal{F}_t)$-stopping time if

$$\{\omega:\tau(\omega)\leq t\}\in\mathcal{F}_t,\ \forall t\geq0.$$

Define

$$\mathcal{F}_\tau:=\{A\in\mathcal{F}:\{\tau\leq t\}\cap A\in\mathcal{F}_t,\ \forall t\geq0\}.$$

Definition 1.4.5. *A real-valued $(\mathcal{F}_t)_{t\geq0}$-adapted process $(X_t)_{t\geq0}$ is called an $(\mathcal{F}_t)$-martingale (resp. supermartingale, submartingale) if $\forall t\geq0$, $X_t\in\mathcal{L}^1(\Omega,\mathcal{F},\mathbf{P})$ with $\mathbf{E}[X_t|\mathcal{F}_s]=X_s$ (resp. $\leq$, $\geq$) a.s., $\forall s\leq t$.*

Theorem 1.4.6. *Let $p\geq1$ and $(X_t)_{t\geq0}$ be a right continuous $(\mathcal{F}_t)$-martingale with $\mathbf{E}|X_t|^p<\infty$ $\forall t\geq0$. Then for any $T>0$:*

1. *(Doob's martingale inequality) $P(\sup_{0\leq t\leq T}|X_t|\geq\lambda)\leq\mathbf{E}|X_T|^p/\lambda^p$ $\forall\lambda>0$.*
2. *For $p>1$, $\mathbf{E}[\sup_{0\leq t\leq T}|X_t|^p]\leq(p/[p-1])^p\mathbf{E}|X_t|^p$.*

Theorem 1.4.7 (Doob's optional sampling theorem / optional stopping theorem)**.** *Let $(X_t)_{t\geq0}$ be a right continuous martingale (resp. supermartingale, submartingale) on $(\Omega,\mathcal{F},\mathbf{P};(\mathcal{F}_t)_{t\geq0})$.*

1. *If $\sigma\leq\tau$ are two bounded stopping times, then*

$$E[X_\tau|\mathcal{F}_\sigma]=X_\sigma\ \text{a.s. (resp. }\leq,\ \geq). \tag{1.4.1}$$

2. *If $(X_t)_{t\geq0}$ are UI and $\sigma\leq\tau\leq\infty$, then* (1.4.1) *holds.*

Remark. Suppose you're playing a fair game, and you have some strategy for when you stop playing (τ). The optional sampling theorem tells us that you cannot devise a winning strategy (or a losing strategy, for that matter).

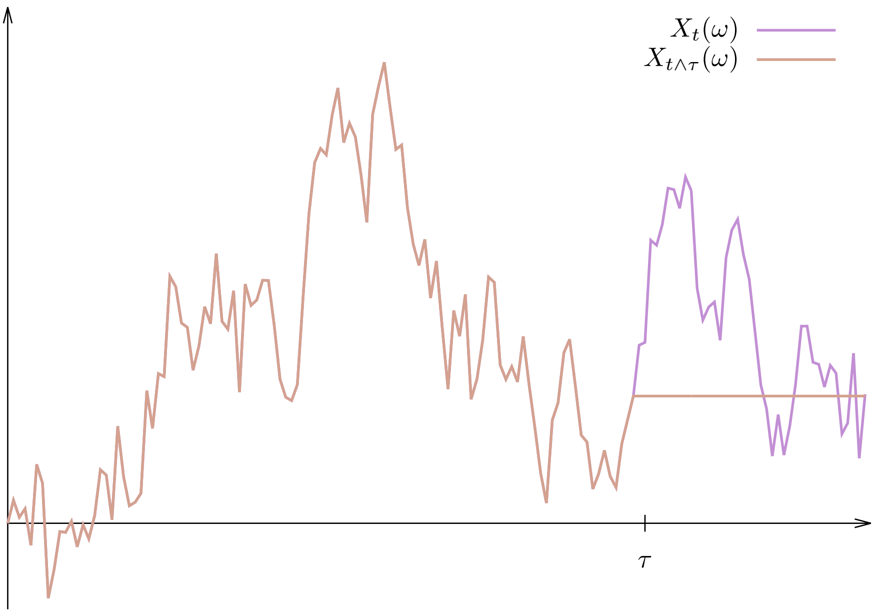
Let $(X_t)_{t\geq0}$ be a process and τ a stopping time. Then

$$X_{t\wedge\tau}(\omega):=\begin{cases}X_t(\omega) & \text{if } t<\tau(\omega)\\ X_{\tau(\omega)}(\omega) & \text{if } t\geq\tau(\omega)\end{cases}.$$

$X_{t\wedge\tau}(\omega)$ is called a stopped process, refer to figure 1.

Definition 1.4.8. *An $(\mathcal{F}_t)$-adapted process $(X_t)_{t\geq0}$ on a filtered space $(\Omega,\mathcal{F},\mathbf{P};(\mathcal{F}_t)_{t\geq0})$ is called a* local martingale *if there is an increasing sequence $\{\tau_n\}$ of* finite *stopping times $\tau_n\uparrow\infty$ as $n\to\infty$ a.s. and s.t. $(X_{t\wedge\tau_n})_{t\geq0}$ is a martingale for each n.*

FIGURE 1: Stopped process.



2 Brownian Motion

2.1 Definition

Definition 2.1.1. *A process $B = (B_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbf{P})$*[9] *taking values in $\mathbb{R}^d$ is called* Brownian motion (BM) *if*

1. *(Independent Increments)*[10] *For any $0 \leq t_0 < t_1 < \cdots < t_n$, the r.v.s B_{t_0}, $B_{t_1} - B_{t_0}$, $\dots$, $B_{t_n} - B_{t_{n-1}}$ are mutually independent.*
2. *(Normality) For any $t > s \geq 0$, the r.v. $B_t - B_s \sim \mathrm{N}(0, (t-s)I)$, that is, $B_t - B_s$ has pdf*

$$p(t-s, x) = \frac{1}{(2\pi[t-s])^{d/2}} \exp\left(-\frac{|x|^2}{2(t-s)}\right), \ x \in \mathbb{R}^d.$$

3. *(Continuity) $(B_t)_{t\geq 0}$ is continuous a.s..*

If $\mathbf{P}(B_0 = 0) = 1$, then $(B_t)_{t\geq 0}$ is called *standard Brownian motion.* We typically assume that this is what we are working with.

Example 2. Let $(B_t)_{t\geq 0}$ be a BM in $\mathbb{R}$. Then

$$\mathbf{E}|B_t - B_s|^p = C_p|t-s|^{p/2} \ \forall t > s \geq 0, \ p \geq 0,$$

where

$$C_p := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |x|^p e^{-|x|^2/2} \, dx.$$

Indeed,

$$\begin{aligned}
\mathbf{E}|B_t - B_s|^p =& \frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{\infty} |x|^p \exp\left(-\frac{|x|^2}{2(t-s)}\right) \, dx \\
=& \frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{\infty} (t-s)^p |y|^p e^{-|y|^2/2} \sqrt{t-s} \, dy \\
=& C_p (t-s)^{p/2},
\end{aligned}$$

[9] Note that is *not* a filtered probability space.

[10] This property is very important.

where we made the change of variables $y = x/\sqrt{t-s}$.

Properties of BM

1. $(B_t)_{t\geq 0}$ is nowhere differentiable, a.s..
2. $\lim_{t\to\infty} B_t/t = 0$, a.s..
3. $\mathrm{Cov}(B_t, B_s) = s \wedge t$.
4. $\tilde{B}_t := B_{t+s} - B_s$, for any given $s \geq 0$, is also a standard Brownian motion. This property is also true for random $s \geq 0$.

Remark. For point 2, consider

$$\frac{B_n}{n} = \frac{(B_n - B_{n-1}) + \cdots + (B_1 - B_0)}{n} \to \mathbf{E}[\mathrm{N}(0,1)] = 0 \text{ a.s..}$$

The result holds due to the strong law of large numbers and the normality property of BM (in particular, $B_i - B_{i-1} \sim \mathrm{N}(0,1)$).

This is not a proof! It just provides some intuition.

For point 3, suppose $t \geq s$. Then

$$\begin{aligned}\mathrm{Cov}(B_t, B_s) &= \mathrm{Cov}([B_t - B_s] + B_s, B_s)\\ &= \mathrm{Cov}(B_t - B_s, B_s) + \mathrm{Cov}(B_s, B_s)\\ &= 0 + \mathrm{Var}(B_s)\\ &= s,\end{aligned}$$

Where we have used independent increments and normality of BM. The result as given above follows from symmetry.

Remark (Scaling).
1. $M_t := \lambda B_{t/\lambda^2}$ $\forall \lambda \neq 0$ is also a standard BM. In particular, $-B_t$ is a standard BM.
2. Let $d = 1$. Define
$$W_t = \begin{cases} tB_{1/t} & t > 0 \\ 0 & t = 0 \end{cases}.$$
   Then W_t is a BM.

   It is easy to check that $\mathrm{Var}(W_t) = t$.

   For continuity, we need to check that $\lim_{t\to 0} tB_{1/t} = 0$. Then
$$\lim_{to\to 0} tB_{1/t} = \lim_{t\to 0} \frac{B_{1/t}}{1/t} = 0 \text{ a.s..}$$
   The final equality follows from the second property of BM described above.

2.2 Markov property and finite-dimensional distributions

Recall:

$$p(t,x) = \frac{1}{(2\pi t)^{d/2}} e^{-|x|^2/2t},\ t > 0,\ x \in \mathbb{R}^d.$$

We also have that

$$p(t+s, x-y) = p(t, x-z)p(s, z-y)\ \forall t, s > 0, x, y, z \in \mathbb{R}^d.$$

Refer to exercise sheet 4, problem 2.

Lemma 2.2.1. *If $t > s > 0$, then the joint distribution (B_t, B_s) is given by*

$$P(B_s \in dx, B_t \in dy) = p(s,x)p(t-s, y-x)\ dxdy.$$

Proof. Since B_s and $B_t - B_s$ are independent, $(B_s, B_t - B_s)$ has joint density function $p(s, x_1)p(t-s, x_2)$. Hence, for all bounded Borel measurable functions f:

$$\begin{aligned} E[f(B_s, B_t)] &= \mathbf{E}[f(B_s, B_t - B_s + B_s)] \\ &= \int\int f(x_1, x_2 + x_1)p(s, x_1)p(t-s, x_2)\ dx_1 dx_2 \\ &= \int\int f(x, y)p(s, x)p(t-s, y-x)\ dxdy. \end{aligned}$$

In the final equality, we performed the change of variables $x_1 = x$, $x_1 + x_2 = y$. In this case, the Jacobian is 1, so $dx_1 dx_2 = dxdy$.

This proves the claim. $\square$

Let $\mathcal{F}_t^0 := \sigma\{B_s, s \leq t\},\ t \geq 0$. This is the *natural filtration generated by the BM.*

$B_t - B_s$ is independent of $\mathcal{F}_s^0$ $\forall t > s \geq 0$, because of independent increments.

Theorem 2.2.2 (Markov Property)**.** *Let $t > s \geq 0$ and f be a bounded Borel measurable function. Then*

$$\mathbf{E}[f(B_t)|\mathcal{F}_s^0] = \mathbf{E}[f(B_t)|B_s].$$

This is sometimes called the "history independent" or "memoryless" property.

Proof. It suffices to prove for $f \in C_b^\infty(\mathbb{R}^d)$, the set of infinitely differentiable ("smooth"), bounded functions.[11]

For such f, we have the Taylor expansion

$$f(x+y) = \lim_{N\to\infty} \sum_{n=1}^{N} f_n(x)g_n(y) \text{ for some functions } f_n,\ g_n.$$

Then

$$\begin{aligned} \mathbf{E}[f(B_t)|\mathcal{F}_s^0] &= \mathbf{E}[f(B_t - B_s + B_s)|\mathcal{F}_s^0] \\ &= \mathbf{E}\left[\lim_{N\to\infty} \sum_{n=1}^{N} f_n(B_s)g_n(B_t - B_s)|\mathcal{F}_s^0\right] \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} \mathbf{E}[f_n(B_s)g_n(B_t - B_s)|\mathcal{F}_s^0] \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} \mathbf{E}[g_n(B_t - B_s)|\mathcal{F}_s^0]f_n(B_s) \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} \mathbf{E}[g_n(B_t - B_s)]f_n(B_s) \end{aligned}$$

where we move the limit out of the expectation by the dominated conergence theorem, we move the $f_n(B_s)$ term out because B_s is $\mathcal{F}_s^0$ measurable and we remove the condition on $\mathcal{F}_s^0$ because of independent increments.

If we repeat the above derivation, but with $\mathcal{F}_s^0$ replaced by B_s, we get the same result. $\square$

We now consider a more general version of lemma 2.2.1.

Theorem 2.2.3 (Finite-Dimensional Distributions)**.** *For any $0 < t_1 < t_2 < \cdots < t_n$, the $\mathbb{R}^{nd}$ valued r.v. $(B_{t_1}, B_{t_2}, \ldots, B_{t_n})$ has a joint density function*

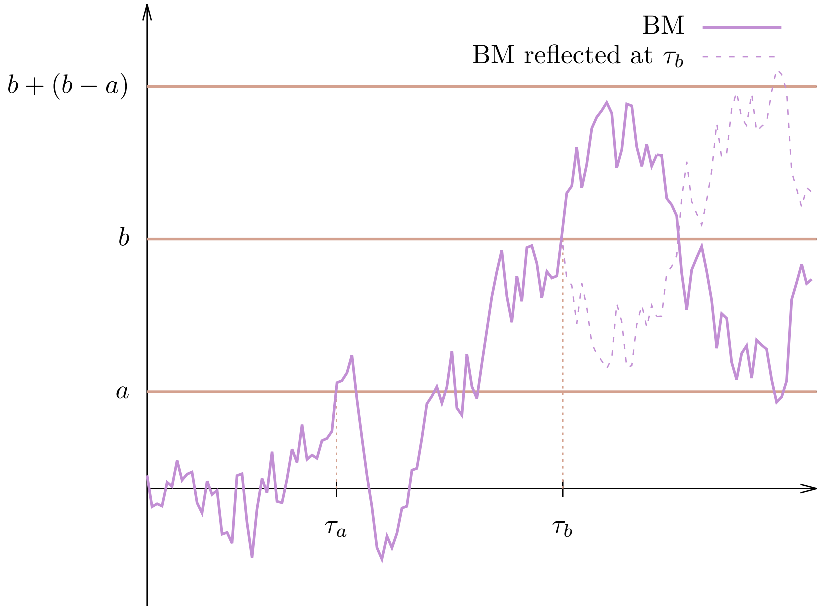
$$p(t_1, x_1)p(t_2 - t_1, x_2 - x_2)\ldots p(t_n - t_{n-1}, x_n - x_{n-1}).$$

[11] Any Borel measurable function can be approximated by functions in this space, cf. pp. 27 - 28 in Stochastic Differential Equations, Øksendal (2000).

2.3 The reflection principle and martingale property

Let $B = (B_t)_{t\geq 0}$ be a standard BM on $(\Omega, \mathcal{F}, \mathbf{P}; \mathcal{F}_t^0)$ in $\mathbb{R}$. Let $b > 0$, $b > a$ and $\tau_b = \inf\{t \geq 0 : B_t = b\}$ (a stopping time).

FIGURE 2: The reflection principle.



Consider

$$\mathbf{P}\left(\max_{0\leq s\leq t} B_s \geq b, B_t \leq a\right) = \mathbf{P}\left(\max_{0\leq s\leq t} B_s \geq b, B_t \geq b + (b-a)\right) = \mathbf{P}\left(B_t \geq 2b - a\right) = \frac{1}{\sqrt{2\pi t}} \int_{2b-a}^{\infty} e^{-x^2/2t}\, dx.$$

The first equality is clear from figure 2, the second follows because the second event in implies the first, and the third equality holds because $B_t \sim \mathrm{N}(0, t)$.

Now we consider the case where $\max_{0\leq s\leq t} B_s \leq b$,

$$\mathbf{P}\left(\max_{0\leq s\leq t} B_s \leq b, B_t \leq a\right) = \mathbf{P}\left(B_t \leq a\right) - \mathbf{P}(B_t \leq a - 2b) = \mathbf{P}\left(a - 2b \leq B_t \leq a\right) = \int_{a-2b}^{a} \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}\, dx.$$

Let $M_t := \max_{0\leq s\leq t} B_s$. This is called the *running maximum* process.

We now consider the joint distribution of M_t and B_t, which are highly correlated with each other.

Theorem 2.3.1. *The joint distribution of* (M_t, B_t) *is*

$$P(M_t \in db, B_t \in da) = \frac{2(2b-a)}{\sqrt{2\pi t^3}} \exp\left(-\frac{(2b-a)^2}{2t}\right)\, dadb,$$

over the region $\{b - a : b \geq a, b > 0\}$.

Now,

$$\begin{aligned}
P(M_t \geq c) &= \frac{2}{\sqrt{2\pi t^3}} \int_c^{\infty} \int_{-\infty}^{b} (2b-a) \exp\left(-\frac{(2b-a)^2}{2t}\right)\, dadb \\
&= \frac{2}{\sqrt{2\pi t^3}} \int_c^{\infty} \int_b^{\infty} x \exp\left(-\frac{x^2}{2t}\right)\, dxdb \\
&= \frac{2}{\sqrt{2\pi}} \int_{c/\sqrt{t}}^{\infty} \exp\left(-\frac{x^2}{2}\right)\, dx \\
&= 2[1 - \mathrm{N}(c/\sqrt{t})] \\
&= 2\mathbf{P}(B_t \geq c),
\end{aligned}$$

where $\mathrm{N}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$.

In the above, we first used the change of variable $x := 2b - a$, followed by $x := b/\sqrt{t}$.

A martingale $(M_t)_{t\geq 0}$ is called *square integrable* if $\mathbf{E}[M_t^2] < \infty\ \forall t \geq 0$.

We define the class $\mathcal{M}_2^c := \{$all continuous, square integrable martingales$\}$.

Theorem 2.3.2. *1. $B_t \equiv (B_t^1\ B_t^2\ \dots\ B_t^d)$ is a continuous, square integrable martingale.*

2. $M_t := B_t^i B_t^j - \delta_{ij} t$ is a continuous martinagle, where δ_{ij} is Kronecker's delta.

Proof. 1. We only show that B_t is a martingale:

$$\begin{aligned}
\mathbf{E}[B_t|\mathcal{F}_s^0] &= \mathbf{E}[B_t - B_s + B_s|\mathcal{F}_s^0] \\
&= \mathbf{E}[B_t - B_s|\mathcal{F}_s^0] + \mathbf{E}[B_s|\mathcal{F}_s^0] \\
&= \mathbf{E}[B_t - B_s] + B_s \\
&= 0 + B_s \\
&= B_s.
\end{aligned}$$

2. We only need to show for BM in $\mathbb{R}$ that

$$\mathbf{E}[B_t^2 - t|\mathcal{F}_s^0] = B_s^2 - s.$$

Rearranging, we see that

$$\begin{aligned}
\mathbf{E}[B_t^2 - B_s^2|\mathcal{F}_s^0] &= \mathbf{E}[(B_t - B_s)^2|\mathcal{F}_s^0] + \mathbf{E}[2B_s(B_t - B_s)|\mathcal{F}_s^0] \\
&= \mathbf{E}[(B_t - B_s)^t] + 2B_s \underbrace{\mathbf{E}[B_t - B_s|\mathcal{F}_s^0]}_{=0} \\
&= \mathrm{Var}(B_t - B_s) + \mathbf{E}[B_t - B_s]^2 \\
&= t - s + 0^2 \\
&= t - s.
\end{aligned}$$

□

2.4 Quadratic variation process

Lemma 2.4.1. *Let $\mathcal{D} := \{0 = t_0 < t_1 < \cdots < t_n = t\}$ be a finite partition of $[0, t]$, and let $V_D := \sum_{l=1}^{n} |B_{t_l} - B_{t_l - 1}|^2$, the quadratic variation of B over D. Then $E[V_D] = t$, $\mathrm{Var}(V_D) = 2\sum_{l=1}^{n}(t_l - t_{l-1})^2$.*

Proof. The expectation is

$$\begin{aligned}
\mathbf{E}[V_D] &= \sum_{l=1}^{n} \mathbf{E}|B_{t_l} - B_{t_{l-1}}|^2 \\
&= \sum_{l=1}^{n} |t_l - t_{l-1}| \\
&= t,
\end{aligned}$$

and the variance is

$$\begin{aligned}
\mathrm{Var}(V_D) =& \mathbf{E}\left[(V_D - \mathbf{E}V_D)^2\right] \\
=& \mathbf{E}\left[\left(\sum_{l=1}^{n} |B_{t_l} - B_{t_{l-1}}|^2 - |t_l - t_{l-1}|\right)^2\right] \\
=& \sum_{k,l=1}^{n} \mathbf{E}\left[\left(|B_{t_k} - B_{t_{k-1}}|^2 - |t_k - t_{k-1}|\right)\left(|B_{t_l} - B_{t_{l-1}}|^2 - |t_l - t_{l-1}|\right)\right] \\
=& \sum_{k=1}^{n} \mathbf{E}\left[\left(|B_{t_k} - B_{t_{k-1}}|^2 - |t_k - t_{k-1}|\right)^2\right] + \\
& \sum_{k\neq l} \mathbf{E}\left[\left(|B_{t_k} - B_{t_{k-1}}|^2 - |t_k - t_{k-1}|\right)\left(|B_{t_l} - B_{t_{l-1}}|^2 - |t_l - t_{l-1}|\right)\right] \\
=& \sum_{k=1}^{n} \mathbf{E}\left[(B_{t_k} - B_{t_{k-1}})^4 - 2(t_k - t_{k-1})(B_{t_k} - B_{t_{k-1}}) + (t_k - t_{k-1})^2\right] \\
=& \sum_{k=1}^{n} \left\{3(t_k - t_{k-1})^2 - 2(t_k - t_{k-1})^2 + (t_k - t_{k-1})^2\right\} \\
=& 2\sum_{l=1}^{n} (t_k - t_{k-1})^2.
\end{aligned}$$

□

Theorem 2.4.2. *We have* $\lim_{m(D)\to 0} \sum_{l=1}^{n} |B_{t_l} - B_{t_{l-1}}|^2 = t$ *in* $\mathcal{L}^2$ *and in probability, for any* $t \geq 0$, *where* $\mathcal{D}$ *runs over all finite partitions of* $[0,t]$ *and* $m(D) := \max_l |t_l - t_{l-1}|$.

Proof. We only need to show convergence in $\mathcal{L}^2$ as this implies convergence in $\mathcal{L}^1$ which in turn implies convergence in probability.

$$\begin{aligned}
\mathbf{E}\left|\sum_{l=1}^{n} |B_{t_l} - B_{t_{l-1}}|^2 - t\right|^2 &= \mathbf{E}|V_D - \mathbf{E}V_D|^2 \\
&= \mathrm{Var}(V_D) \\
&= 2\sum_{l=1}^{n} (t_l - t_{l-1})^2 \\
&\leq 2m(D)\sum_{l=1}^{n} (t_l - t_{l-1}) \\
&= 2t \cdot m(D) \to 0 \text{ as } m(D) \to 0.
\end{aligned}$$

□

Let $M = (M_t)_{t\geq 0} \in \mathcal{M}_c^2$. As with BM, the limit

$$\langle M \rangle_t := \lim_{m(D)\to 0} \sum_{l=1}^{n} |M_{t_l} - M_{t_{l-1}}|^2$$

exists in $\mathcal{L}^2$ and hence also in probability.

$\{\langle M \rangle_t\}_{t\geq 0}$ is called the *quadratic variation process* of $(M_t)_{t\geq 0}$, or simply the *bracket process*. It is an adapted, continuous, increasing process with $\langle M \rangle_0 = 0$.

Doob-Meyer decomposition If $M = (M_t)_{t\geq 0}$ is a continuous martingale and $A = (A_t)_{t\geq 0}$ is adapted, continuous and increasing, then $X := M + A$ is a submartingale.

The reverse is the Doob-Meyer decomposition. In particular, the decompositon is unique.

Theorem 2.4.3. *Let $M \in \mathcal{M}_c^2$ be given. Then $\langle M \rangle_t$ is the unique, continuous, adapted and increasing process initially zero s.t. $M_t^2 - \langle M \rangle_t$ is a martingale.*

Proof. By Jensen's inequality, $\mathbf{E}[M_t^2|\mathcal{F}_s] \geq \mathbf{E}[M_t|\mathcal{F}_s]^2 = M_s^2$, so M_t^2 is a submartingale. Hence, by the Doob-Meyer decomposition,

$$M_t^2 = N_t + A_t,$$

where N_t is a martingale and A_t is an increasing process. This decomposition is unique.

On the other hand, $M_t^2 = (M_t^2 - \langle M \rangle_t) + \langle M \rangle_t$, and since the decomposition above is unique and $\langle M \rangle_t$ is increasing, $\langle M \rangle_t$ is the unique process described in the theorem statement. $\square$

Theorem 2.4.4. *Let $M, N \in \mathcal{M}_c^2$ and let*

$$\langle M, N \rangle_t := \frac{1}{4}(\langle M + N \rangle_t - \langle M - N \rangle_t),$$

called the bracket process *of M, N. Then $\langle M, N \rangle_t$ is the unique, adapted, continuous process having finite variation almost surely and with initial zero, s.t. $M_t N_t - \langle M, N \rangle_t$ is a martingale.*

Moreover, $\sum_{m(\mathcal{D}) \to 0} (M_{t_l} - M_{t_{l-1}})(N_{t_l} - N_{t_{l-1}}) = \langle M, N \rangle_t$ in probability where $\mathcal{D} = \{0 = t_0 < t_1 < \dots t_n = t\}$ and $m(D) = \max_l |t_l - t_{l-1}|$.

3 Itô's calculus

$B = (B_t)_{t \geq 0}$ is a standard BM on $(\Omega.\mathcal{F}, \mathbf{P})$ and $\mathcal{F}_t^0 = \sigma\{B_s, 0 \leq s \leq t\}$ is the natural filtration.

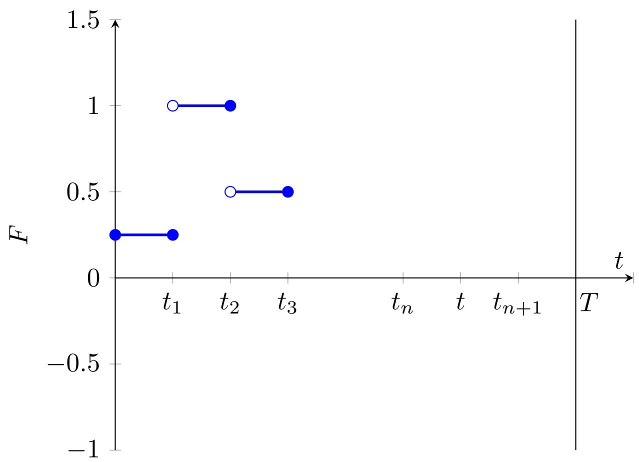
Our objective is to define $\int_0^t F_s dB_s$ $\forall t \geq 0$ for some process $(F_t)_{t \geq 0}$.

3.1 Stochastic integrals for BM

By convention, we shall call a process *adapted* if it is measurable and $\mathcal{F}_t^0$-adapted.

An adapted process $F = (F_t)_{t \geq 0}$ is called a *simple process* if $F_t(\omega) = f_0(\omega)\mathbb{1}_{\{0\}}(t) + \sum_{i=0}^{\infty} \mathbb{1}_{(t_i, t_{i+1}]}(t) f_i(\omega)$, where $0 = t_0 < t_1 < \dots$ so that for any finite $T \geq 0$, there are finitely many $t_i \in [0, T]$, each f_i is a $\mathcal{F}_{t_i}^0$-measurable r.v., f_0 is $\mathcal{F}_0^0$-measurable, and F is uniformly bounded.

FIGURE 3: A simple process.



Let $\mathcal{L}_0$ denote the space of all simple processes.

Let $F \in \mathcal{L}_0$, then define

$$\int_0^t F_s dB_s \equiv I(F)_t := \sum_{i=0}^{\infty} f_i (B_{t_{i+1} \wedge t} - B_{t_i \wedge t}).$$

Note that this sum is finite, since we stop at t.

Clearly, $I(F)$ is continuous in t and $\mathcal{F}_t^0$-adapted (f_i is $\mathcal{F}_{t_i}$-measurable, therefore $\mathcal{F}_t$-measurable, since $\mathcal{F}_{t_i} \subseteq \mathcal{F}_t$ if $t_i \leq t$).

Lemma 3.1.1. $(I(F))_{t\geq 0}$ *is a martingale, i.e.*

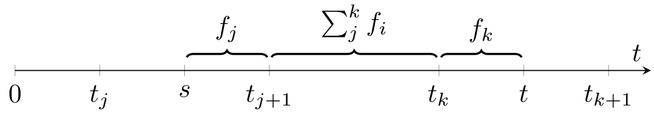
$$\mathbf{E}[I(F)_t - I(F)_s | \mathcal{F}_s^0] = 0, \ \forall t \geq s \geq 0.$$

Proof. Assume $t_j < s \leq t_{j+1}$, $t_k < t \leq t_{k+1}$ for some j, k.

Then $j \leq k$. Consider $j + 1 < k$, in which case

$$I(F)_t - I(F)_s = \sum_{i=j+1}^{k-1} f_i(B_{t_i+1} - B_{t_i}) + f_j(B_{t_{j+1}} - B_s) + f_k(B_t - B_{t_k}).$$

FIGURE 4



Consider the expectation of the terms in the sum. By the tower property, we have

$\mathbf{E}[f_i(B_{t_{i+1}} - B_{t_i})|\mathcal{F}_s^0] = \mathbf{E}\{\mathbf{E}[f_i(B_{t_{i+1}} - B_{t_i})|\mathcal{F}_{t_i}^0]|\mathcal{F}_s^0\} = \mathbf{E}\{f_i\mathbf{E}[(B_{t_{i+1}} - B_{t_i})|\mathcal{F}_{t_i}^0]|\mathcal{F}_s^0\} = 0, \ \forall j+1 \leq i \leq k-1.$

Similarly, the terms involving f_j and f_k have conditional expectation zero also. Hence the expectation of the difference is zero and so the given process is a martingale. ☐

Lemma 3.1.2. $I(F) \in \mathcal{M}_c^2$ *and* $\langle I(F) \rangle_t = \int_0^t F_s^2 \ ds$.

Proof. To prove $I(F)_t^2 - \int_0^t F_s^2 \ ds$ is a martingale, use the same idea as in the proof of lemma 3.1.1. ☐

Lemma 3.1.3 (Itô's isometry)**.** $F \mapsto I(F)$ *is linear and* $\mathbf{E}[I(F)_t^2] = \mathbf{E}\left[\int_0^t F_s^2 \ ds\right]$.

Proof. $I(F)_t^2 - \int_0^t F_s^2 \ ds$ is a martingale, hence $\mathbf{E}[I(F)_t^2 - \int_0^t F_s^2 \ ds] = 0$. Then by linearity of expectation and rearranging, we have the result. ☐

3.2 Stochastic integrals for adapted processes

Now, for $T > 0$ define

$$\mathcal{L}_{\mathcal{F}}^2(0, T; \mathbb{R}) = \left\{ F = (F_t)_{t\geq 0} : F \text{ is adapted and } \mathbf{E}\left[\int_0^T F_t^2 \ dt\right] < \infty \right\} \supseteq \mathcal{L}_0.$$

In fact, $\mathcal{L}_0$ is dense in $\mathcal{L}_{\mathcal{F}}^2$.

$\forall F \in \mathcal{L}_{\mathcal{F}}^2(0, T; \mathbb{R})$, $\exists F_n \in \mathcal{L}_0$ s.t. $\mathbf{E}\left[\int_0^T |F_{n,t} - F_t|^2 \ dt\right] \to 0$ as $n \to \infty$.

$$\mathbf{E}|I(F_n)_T - I(F_m)_T|^2 = \mathbf{E}|I(F_n - F_m)_T|^2 = \mathbf{E}\int_0^T |F_{n,s} - F_{m,s}|^2 \ ds \to 0 \text{ as } n, m \to \infty.$$

$$I(F_n) \to I(F).$$

We define $I(F) := \lim_{n\to\infty} I(F_n)$ as the Itô integral against B, denoted as $I(F)_t = \int_0^t F_s \ dB_s$.

Theorem 3.2.1. *If $F = (F_t)_{t\geq 0} \in \mathcal{L}^2_{\mathcal{F}}(0,T;\mathbb{R})$, then both $\int_0^t F_s\ dB_s$ and $\left(\int_0^t F_s\ dB_s\right)^2 - \int_0^t F_s^2\ ds$ are martingales.*

Moreover, $\mathbf{E}[\int_0^t F_s\ dB_s]^2 = \mathbf{E}\int_0^t F_s^2\ ds\ \forall t \geq 0$, and $\mathbf{E}\left[\left(\int_s^t F_u\ dB_u\right)^2 \middle| \mathcal{F}_s\right] = \mathbf{E}[\int_s^t F_u^2\ du|\mathcal{F}_s]\ \forall t \geq s > 0$.

Finally, $\langle \int_0^t F_s\ dB_s, \int_0^t G_s\ dB_s\rangle_t = \int_0^t F_s G_s\ ds,\ \forall F, G \in \mathcal{L}^2_{\mathcal{F}}(0,T;\mathbb{R})$.

We now define $\int_0^t F_s\ dM_s$ where $M \in \mathcal{M}_c^2$.
Suppose $F \in \mathcal{L}_0$: $F_t = f_0 \mathbb{1}_{\{0\}}(t) + \sum_{i=0}^\infty f_i \mathbb{1}_{(t_i,t_{i+1}]}(t)$.
Define $I^M(F) := \sum_{i=0}^\infty f_i(M_{t\wedge t_{i+1}} - M_{t\wedge t_i})$.
Then

1. $I^M(F) \in \mathcal{M}_c^2$;
2. $\langle I^M(F)\rangle_t = \int_0^t F_s^2\ d\langle M\rangle_s$;
3. (Itô's isometry) $E(\int_0^t F_s\ dM_s)^2 = \mathbf{E}[\int_0^t F_s^2\ d\langle M\rangle_s]$.

$$\mathcal{L}_F^{2,M}(0,T;\mathbb{R}) := \left\{F = (F_t)_{t\geq 0} : F \text{ is adapted, and } \mathbf{E}\int_0^t F_s^2\ d\langle M\rangle_s < \infty\right\}.$$

One can define $\int_0^t F_s\ dM_s$ for $F \in \mathcal{L}_{\mathcal{F}}^{2,M}(0,T;\mathbb{R})$ via a limiting argument.
$\int_0^t F_s\ dM_s$ is a continuous martingale, $\mathbf{E}(\int_0^t F_s\ dM_s)^2 = \mathbf{E}[\int_0^t F_s^2\ d\langle M\rangle_s]$.
$\langle \int_0^t F_s\ dM_s\rangle = \int_0^t F_s^2\ d\langle M\rangle_s$.
$\langle \int_0^t F_s\ dM_s, \int_0^t G_s\ dN_s\rangle = \int_0^t F_s G_s\ d\langle M, N\rangle_s$.

3.3 Itô's integration for semimartingales

Let $M = (M_t)_{t\geq 0}$ be a continuous local martingale with $M_0 = 0$. Then there exists $\{\tau_n\}$ with $\tau_n \uparrow \infty$ as $n \to \infty$ a.s. and $(M_n)_{t\geq 0} := (M_{t\wedge\tau_n})_{t\geq 0}$ is a continuous martingale.
Define $\langle M\rangle_t := \langle M_n\rangle_t\ \forall t \leq \tau_n$, which is an adapted, continuous and increasing process, with $\langle M\rangle_0 = 0$ and $M_t^2 - \langle M\rangle_t$ is a local martingale.
Now we define $\int_0^t F_s\ dM_s$ for F which is left-continuous, adapted and $\int_0^T F_s^2\ d\langle M\rangle_s < \infty$ a.s. $\forall T > 0$.
$\int_0^t F_s\ dM_s := \int_0^t F_s d(M_n)_s$ if $t \leq \tilde{\tau}_n$ where $\tilde{\tau}_n \uparrow \infty$.
Both $\int_0^t F_s\ dM_s$ and $(\int_0^t F_s\ dM_s)^2 - \int_0^t F_s^2\ d\langle M\rangle_s$ are local martingales.

Definition 3.3.1. *An adapted continuous process $X = (X_t)_{t\geq 0}$ is called a semimartingale if $X_t = M_t + V_t$ where $(M_t)_{t\geq 0}$ is a continuous local martingale and $(V_t)_{t\geq 0}$ is a process with finite variation in any finite interval.*

$\int_0^t F_s\ dX_s := \int_0^t F_s\ dM_s + \int_0^t F_s\ dV_s$.
$\int_0^t F_s\ dX_s = \lim_{m(\mathcal{D})\to 0} \sum_l \left(F_{l-1}(X_{t_l} - X_{t_{l-1}})\right)$ in probability.
If X, Y are semimartingales, then

$$\langle X, Y\rangle_t := \lim_{m(\mathcal{D})\to 0} \sum_{l=1}^n (X_{t_l} - X_{t_{l-1}})(Y_{t_l} - Y_{t_{l-1}}) \text{ in probability}$$

where $\mathcal{D} = \{0 = t_0 < t_1 < \cdots < t_n = t\}$.

Properties

1. (Bilinearity) $\langle \alpha X+Y, \beta U+V\rangle = \alpha\beta\langle X,U\rangle + \alpha\langle X,V\rangle + \beta\langle Y,U\rangle + \langle Y,V\rangle$.
2. $\langle X,Y\rangle = 0$ if X is continuous with finite variation.
3. If $X_t = M_t + V_t$, $Y_t = N_t + U_t$, where M,N are martingales, V,U have finite variation and so X,Y are semimartingales, then $\langle X,Y\rangle_t = \langle M,N\rangle_t$.

3.4 Itô's formula

Let $(X_t)_{t\geq 0}$ be a continuous semimartingale. For a partition $0=t_0<t_1<\cdots<t_n=t$,

$$\begin{aligned} X_{t_j}^2 - X_{t_{j-1}}^2 &= (X_{t_j}-X_{t_{j-1}})^2 + 2X_{t_{j-1}}(X_{t_j}-X_{t_{j-1}}) \\ \implies X_t^2 - X_0^2 &= 2\sum_{j=1}^n X_{t_{j-1}}(X_{t_j}-X_{t_{j-1}}) + \sum_{j=1}^n (X_{t_j}-X_{t_{j-1}})^2. \end{aligned}$$

Letting $m(\mathcal{D})\to 0$,

$$X_t^2 - X_0^2 = 2\int_0^t X_s\ dX_s + \langle X\rangle_t. \tag{3.4.1}$$

Compare this with the formula for integration by parts:

$$\int_0^t x_s\ dy_s = x_s y_s\big|_0^t - \int_0^t y_s\ dx_s.$$

Setting $y_s := x_s$ and rearranging, we get

$$x_t^2 - x_0^2 = 2\int_0^t x_s\ dx_s.$$

Note that in the stochastic version there is a quadratic variation term, but this is absent in the deterministic version.

This is the simplest case of Itô's formula.

Lemma 3.4.1 (Integration by parts)**.** *Let X,Y be two continuous semimartingales. Then*

$$X_tY_t - X_0Y_0 = \int_0^t X_s\ dY_s + \int_0^t Y_s\ dX_s + \langle X,Y\rangle_t.$$

Theorem 3.4.2 (Itô's formula)**.** *Let $(X_t)_{t\geq 0} = (X_t^1,\dots,X_t^d)_{t\geq 0}$ be a continuous, semimartingale in $\mathbb{R}^d$ with $X_t^i = M_t^i + A_t^i$ where $M_t^1,\dots,M_t^d$ are continuous local martingales and $A_t^1,\dots,A_t^d$ are continuous, adapted processes with finite variation.*
Let $f\in C^2(\mathbb{R}^d;\mathbb{R})$[12]. Then

$$f(X_t) - f(X_0) = \sum_{i=1}^d \int_0^t \frac{\partial f(X_s)}{\partial x_i}\ dX_s^i + \frac{1}{2}\sum_{i,j=1}^d \int_0^t \frac{\partial^2 f(X_s)}{\partial x_i \partial x_j}\ d\langle M^i, M^j\rangle_t.^{13}$$

Proof. We prove in the case $d=1$ and $X_t = M_t \in \mathcal{M}_c^2$.

We need to show

$$f(M_t) - f(M_0) = \int_0^t f'(M_s)\ dM_s + \frac{1}{2}\int_0^t f''(M_s)\ d\langle M\rangle_s. \tag{3.4.2}$$

This is true for $f(x)=x^2$ (this is just (3.4.1)).

[12] $f:\mathbb{R}^d\to\mathbb{R}$ and $f\in C^2$.
[13] ∂x_i is differentiating with respect to the i-th parameter of f.

Now suppose (3.4.2) is true for $f(x) = x^n$:

$$M_t^n - M_0^n = n \int_0^t M_s^{n-1}\ dM_s + \frac{1}{2} n(n-1) \int_0^t M_s^{n-2}\ d\langle M \rangle_s.$$

Applying integration by parts (lemma 3.4.1) to M^n and M, we have

$$\begin{aligned}
M_t^{n+1} - M_0^{n+1} =& \int_0^t M_s^n\ dM_s + \int_0^t M_s\ dM_s^n + \langle M^n, M \rangle_t \\
=& \int_0^t M_s^n\ dM_s + \int_0^t M_s (n M_s^{n-1} + \tfrac{1}{2} n(n-1) M_s^{n-2}\ d\langle M \rangle_s) \\
&+ \left\langle M_0^n + n \int_0^t M_s^{n-1}\ dM_s + \tfrac{1}{2} n(n-1) \int_0^t M_s^{n-2}\ d\langle M \rangle_s, \int_0^t dM_s \right\rangle_t \\
=& \int_0^t M_s^n\ dM_s + n \int_0^t M_s^n\ dM_s + \tfrac{1}{2} n(n-1) \int_0^t M_s^{n-1}\ d\langle M \rangle_s \\
&+ n \int_0^t M_s^{n-1}\ d\langle M \rangle_s \\
=& (n+1) \int_0^t M_s^n\ dM_s + \tfrac{1}{2} n(n+1) \int_0^t M_s^{n-1}\ d\langle M \rangle_s
\end{aligned}$$

□

Corollary 3.4.3 (Itô's formula for diffusion processes)**.** *Let* $(X_t)_{t \geq 0} = (X_t^1, \ldots, X_t^d)_{t \geq 0}$ *be given as*

$$X_t^i = X_0^i + \int_0^t b_s^i\ ds + \sum_{j=1}^n \int_0^t \sigma_s^{i,j}\ dB_s^j,\ i = 1, 2, \ldots, d$$

where $(B_t)_{t \geq 0} = (B_t^1, \ldots, B_t^n)_{t \geq 0}$ *is a standard BM in* $\mathbb{R}^n$*, and* $f \in C^{1,2}([0, \infty] \times \mathbb{R}^d; \mathbb{R})$*. Then*

$$f(t, X_t) - f(0, X_0) = \int_0^t \frac{\partial f(s, X_s)}{\partial s}\ ds + \sum_{i=1}^d \int_0^t \frac{\partial f(s, X_s)}{\partial x_i}\ dX_s^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t a_s^{i,j} \frac{\partial^2 f(s, X_s)}{\partial x_i \partial x_j}\ ds,$$

where $a_s^{i,j} = \sum_{k=1}^n \sigma_s^{ik} \sigma_s^{kj},\ i, j = 1, 2, \ldots, d.$

Example 3. Let $B = (B^1, \ldots, B_n)$ be a standard BM. Then for any $f \in C^2(\mathbb{R}^n; \mathbb{R})$:

$$\begin{aligned}
f(B_t) - f(B_0) &= \sum_{i=1}^n \int_0^t \frac{\partial f(B_s)}{\partial x_i}\ dB_s^i + \frac{1}{2} \sum_{i=1}^n \int_0^t \frac{\partial^2 f(B_s)}{\partial {x_i}^2}\ ds \\
&= \int_0^t \nabla f(B_s) \cdot dB_s + \frac{1}{2} \int_0^t \Delta f(B_s)\ ds.
\end{aligned}$$

Hence, $f(B_t) - f(B_0) - \frac{1}{2} \int_0^t \Delta f(B_s)\ ds$ is a local martingale.

Example 4. We will show that $\mathbf{E}| \int_0^t \sigma_s\ dB_s|^2 = \mathbf{E} \int_0^t \sigma_s^2\ ds$ for bounded σ.

Let $M_t := \int_0^t \sigma_s\ dB_s$.

$f(x) = x^2,\ f'(x) = 2x,\ f''(x) = 2,\ \langle M \rangle_t = \int_0^t \sigma_s^2\ ds.$

$$\begin{aligned}
f(M_t) - f(M_0) &= 2 \int_0^t M_s\ dM_s + \frac{1}{2} \cdot 2 \int_0^t \sigma_s^2\ ds \\
\implies \quad \left| \int_0^t \sigma_s\ dB_s \right|^2 &= 2 \int_0^t M_s\ dM_s + \int_0^t \sigma_s^2\ ds \\
\implies \mathbf{E} \left| \int_0^t \sigma_s\ dB_s \right|^2 &= \mathbf{E} \int_0^t \sigma_s^2\ ds.
\end{aligned}$$

3.5 Stochastic exponentials and Girsanov theorem

Consider a deterministic ODE,

$$\begin{cases} dz_t &= z_t \ dx_t \\ z_0 &= 1 \end{cases} \implies z_t = e^{x_t}.$$

What about the stochastic case?

$$\begin{cases} dZ_t &= Z_t \ dX_t \\ Z_0 &= 1 \end{cases}, \tag{3.5.1}$$

where X_t is a continuous semimartingale.

The real meaning of (3.5.1) is

$$Z_t = 1 + \int_0^t Z_s \ dX_s. \tag{3.5.2}$$

Try $Z_t = \exp(X_t + V_t)$, where $(V_t)_{t\geq 0}$, a finite variation process, is to be determined.

Applying Itô's formula,

$$\begin{aligned} Z_t &= Z_0 + \int_0^t \exp(X_s + V_s) \ d(X_s + V_s) + \frac{1}{2} \int_0^t \exp(X_s + V_s) \ d\langle X \rangle_s{}^{14} \\ &= 1 + \int_0^t Z_s \ d(X_s + V_s) + \frac{1}{2} \int_0^t Z_s \ d\langle X \rangle_s. \end{aligned} \tag{3.5.3}$$

This suggests we should set $V_t := -\frac{1}{2}\langle X \rangle_t$. Compare (3.5.2) with (3.5.3).

Lemma 3.5.1. *Let X_t be a semimartingale with $X_0 = 0$. Then*

$$\mathcal{E}(X) := \exp(X_t - \tfrac{1}{2}\langle X \rangle_t), \tag{3.5.4}$$

called the stochastic exponential of X, *is the solution to* (3.5.2).

Lemma 3.5.2. *Any continuous, non-negative local martingale is a supermartingale.*

Proof. Refer to exercise sheet 7, question 1. □

Corollary 3.5.3. *Let $M = (M_t)_{t\geq 0}$ be a continuous local martingale with $M_0 = 0$. Then $\mathcal{E}(M)$ is a supermartingale. In particular,*

$$\mathbf{E}[\exp(M_t - \tfrac{1}{2}\langle M \rangle_t)] \leq 1 \ \forall t \geq 0.$$

Moreover, $\mathcal{E}(M)$ is a martingale up to time $T > 0$ if and only if

$$\mathbf{E}[\exp(M_T - \tfrac{1}{2}\langle M \rangle_T)] = 1. \tag{3.5.5}$$

Theorem 3.5.4 (Novikov's condition)**.** *Let $M = (M_t)_{t\geq 0}$ be a continuous local martingale with $M_0 = 0$. If $\mathbf{E}\exp(\frac{1}{2}\langle M \rangle_T) < \infty$, then $\mathcal{E}(M)$ is a martingale up to time T.*

Proof. Refer to Prof. Qian's lecture notes, theorem 4.6.9. □

Example 5. Let $B = (B_t)_{t\geq 0}$ be a standard BM and $F = (F_t)_{t\geq 0} \in \mathcal{L}^2_{\mathcal{F}}(0, T; \mathbb{R})$. Assume

$$\mathbf{E}\exp(\frac{1}{2}\int_0^T F_t^2 \ dt) < \infty,$$

then $X_t = \mathcal{E}\left[\int_0^t F_s \ dB_s\right] \equiv \exp(\int_0^t F_s \ dB_s - \frac{1}{2}\int_0^t F_s^2)$ is a martingale up to T.

Theorem 3.5.5 (Lévy's characterisation)**.** *A process $M = (M_t)_{t\geq 0}$ with $M_0 = 0$ is a standard BM if and only if it is a continuous local martingale with $\langle M \rangle_t = t$.*

[14] $d\langle X + V \rangle_s = d\langle X \rangle_s$ because V has finite variation.

Proof. We prove the backwards direction.

$\forall T > 0,\ \lambda \in \mathbb{R}$:

$$\mathbf{E}\exp(\tfrac{1}{2}\langle \lambda M\rangle_T) = \mathbf{E}\exp(\tfrac{1}{2}\lambda^2\langle M\rangle_T) = \mathbf{E}\exp(\tfrac{1}{2}\lambda^2 T) < \infty.$$

Hence, $\mathcal{E}(\lambda M) = \exp(\lambda M_t - \frac{1}{2}\lambda^2 t)$ is a martingale up to T, by Novikov's condition.

$$\begin{aligned} &\mathbf{E}[e^{\lambda M_t - \frac{1}{2}\lambda^2 t}|\mathcal{F}_s] = e^{\lambda M_s - \frac{1}{2}\lambda^2 s}\ \forall 0 \le s \le t \le T \\ \iff &\mathbf{E}[e^{\lambda(M_t - M_s)}|\mathcal{F}_s] = e^{\frac{1}{2}\lambda^2(t-s)} \\ \iff &\quad \mathbf{E}[e^{\lambda(M_t - M_s)}] = e^{\frac{1}{2}\lambda^2(t-s)}. \end{aligned}$$

So, $(M_t)_{t\ge 0}$ has independent increments.

Note that the last line is just the expression for the MGF of a $\mathrm{N}(0, t-s)$ r.v., so $M_t - M_s \sim \mathrm{N}(0, t-s)$. $\square$

Suppose we have a filtered probability space $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_t)_{t\ge 0})$ and $T > 0$.

Let Q be another probability measure s.t.

$$\int_\Omega X(\omega)\ d\mathbf{Q}(\omega) = \int_\Omega X(\omega)\ \xi(\omega)\ d\mathbf{P}(\omega)\ \forall \text{ bounded } \mathcal{F}_T\text{-measurable r.v.s } X \tag{3.5.6}$$

where $0 < \xi \in \mathcal{L}^1(\Omega, \mathcal{F}_T, \mathbf{P})$.

We write this as

$$\left.\frac{d\mathbf{Q}}{d\mathbf{P}}\right|_{\mathcal{F}_t} = \xi.$$

We can rewrite (3.5.6) as

$$\mathbf{E}^{\mathbf{Q}}[X] = \mathbf{E}^P[X\xi] \implies \mathbf{E}^{\mathbf{P}}[\xi] = 1,$$

where the implication follows by setting $X \equiv 1$.

Now, let X be $\mathcal{F}_t$-measurable, $\forall t \le T$. Then

$$\mathbf{E}^{\mathbf{Q}}[X] = \mathbf{E}^{\mathbf{P}}[X\xi] = \mathbf{E}^{\mathbf{P}}[\mathbf{E}^{\mathbf{P}}[X\xi|\mathcal{F}_t]] = \mathbf{E}^{\mathbf{P}}[X\mathbf{E}^{\mathbf{P}}[\xi|\mathcal{F}_t]].$$

So, $\forall t \le T$, $\frac{d\mathbf{Q}}{d\mathbf{P}}|_{\mathcal{F}_t} = \mathbf{E}^{\mathbf{P}}[\xi|\mathcal{F}_t] > 0$[15].

Conversely, let $Z = (Z_t)_{t\ge 0}$ be a continuous, positive martingale up to T with $Z_0 = 1$, on $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_t)_{t\ge 0})$.

Define a measure on $(\Omega, \mathcal{F}_T)$ by

$$\left.\frac{d\mathbf{Q}}{d\mathbf{P}}\right|_{\mathcal{F}_t} = Z_t.$$

For any $\mathcal{F}_t$-measurable r.v. X,

$$\mathbf{E}^{\mathbf{Q}}[X] = \mathbf{E}^{\mathbf{P}}[\mathbf{E}^{\mathbf{P}}[Z_T X|\mathcal{F}_t]] = \mathbf{E}^{\mathbf{P}}[X\mathbf{E}^{\mathbf{P}}[Z_T|\mathcal{F}_t]] = \mathbf{E}^{\mathbf{P}}[XZ_t].$$

Theorem 3.5.6 (Girsanov's theorem)**.** *Let $(M_t)_{t\ge 0}$ be a continuous local martingale and $(Z_t)_{t\ge 0}$ a continuous, positive martingale up to T on $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_t)_{t\ge 0})$ with $Z_0 = 1$. Then*

$$X_t := M_t - \int_0^t \frac{1}{Z_s}\ d\langle M, Z\rangle_s$$

is a continuous local martingale on $(\Omega, \mathcal{F}, \mathbf{Q}; (\mathcal{F}_t)_{t\ge 0})$ up to T.

[15] This expectation is a process in E, in fact, is a martingale up to T. Put $Y_t := \mathbf{E}[\xi|\mathcal{F}_t]$. Then

$$\mathbf{E}[Y_t|\mathcal{F}_s] = \mathbf{E}[\mathbf{E}[\xi|\mathcal{F}_t]|\mathcal{F}_s] = \mathbf{E}[\xi|\mathcal{F}_s] = Y_s,\ \forall s \le t \le T.$$

Proof. We prove for the case when Z and X are bounded and $M \in \mathcal{M}_c^2$. Using integration by parts,

$$\begin{aligned}
Z_t X_t &= Z_0 X_0 + \int_0^t Z_s \, dX_s + \int_0^t X_s \, dZ_s + \langle Z, X \rangle_t \\
&= Z_0 X_0 + \int_0^t Z_s \, dM_s - \underbrace{\int_0^t Z_s \cdot \frac{1}{Z_s} \, d\langle M, Z \rangle_s}_{=\langle M,Z\rangle_t - \langle M,Z\rangle_0} + \int_0^t X_s \, dZ_s + \underbrace{\langle Z, X \rangle_t}_{=\langle M,Z\rangle_t} \\
&= Z_0 X_0 + \int_0^t Z_s \, dM_s + \int_0^t X_s \, dZ_s + \langle M, Z \rangle_0,
\end{aligned}$$

which is a $\mathbf{P}$-martingale.

Now, $\forall A \in \mathcal{F}_s$, $s \le t$,

$$\begin{aligned}
\mathbf{E}^{\mathbf{Q}}[\mathbb{1}_A (X_t - X_s)] &= \mathbf{E}^{\mathbf{Q}}[\mathbb{1}_A X_t] - \mathbf{E}^{\mathbf{Q}}[\mathbb{1}_A X_s] \\
&= \mathbf{E}^{\mathbf{P}}[\mathbb{1}_A X_t Z_t] - \mathbf{E}^{\mathbf{P}}[\mathbb{1}_A X_s Z_s] \\
&= \mathbf{E}^{\mathbf{P}}[\mathbb{1}_A (X_t Z_t - X_s Z_s)] \\
&= 0.
\end{aligned}$$

□

Since $Z_t > 0$, applying Itô's lemma to $\log Z_t$

$$\log Z_t - \log Z_0 = \int_0^t \frac{1}{Z_s} \, dZ_s - \frac{1}{2} \int_0^t \frac{1}{Z_s^2} \, d\langle Z \rangle_s .$$

Hence, $Z_t = \mathcal{E}(N)_t$ where $N_t := \int_0^t \frac{1}{Z_s} \, dZ_s$ is a continuous local martingale.

$Z_t = 1 + \int_0^t Z_s \, dN_s$.

$$\langle M, Z \rangle_t = \left\langle \int_0^t \, dM_s, \int_0^t Z_s \, dN_s \right\rangle_t = \int_0^t Z_s \, d\langle M, N \rangle_s .$$

Corollary 3.5.7. *Let N_t be a continuous martingale on $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_t)_{t\ge 0})$, $N_0 = 0$, s.t. $\mathcal{E}(N)_t$ is a continuous martingale up to T.*

Define a probability measure $\mathbf{Q}$ on $(\Omega, \mathcal{F}_T)$ by

$$\left. \frac{d\mathbf{Q}}{d\mathbf{P}} \right|_{\mathcal{F}_t} = \mathcal{E}(N)_t, \forall t \le T.$$

If $(M_t)_{t\ge 0}$ is a continuous local martingale under P, then $X_t := M_t - \langle N, M \rangle_t$ is a continuous local martingale under Q up to T.

Example 6. Consider the stock price

$$dS_t = S_t (r_t dt + dW_t) = S_t d\tilde{W}_t .$$

Let $N_t := -\int_0^t r_s \, dW_s$.

$\mathbf{E}[\exp(\frac{1}{2}\langle N \rangle_t)] = \mathbf{E}[\exp(\frac{1}{2}\int_0^t r_s^2 \, ds)] < \infty$, then $\mathcal{E}(N)_t = \exp(-\int_0^t r_s \, dW_s + \frac{1}{2}\int_0^t r_s^2 \, ds)$ is a martingale up to T, by Novkikov's condition.

Under $\frac{d\mathbf{Q}}{d\mathbf{P}}|_{\mathcal{F}_t} = \mathcal{E}(N)_t$, $\tilde{W}_t := W_t - \langle N, W \rangle_t = W_t + \int_0^t r_s \, ds$[16] is a continuous local martingale under $\mathbf{Q}$.

However, $\langle \tilde{W} \rangle_t = \langle W \rangle_t = t \implies \tilde{W}$ is a BM under $\mathbf{Q}$.

In this case, $dS_t = S_t \, d\tilde{W}_t \implies S_t$ is a martingale under $\mathbf{Q}$.

So, e.g. $\mathbf{E}^{\mathbf{Q}}[S_T] = \mathbf{E}^{\mathbf{Q}}[S_0]$, where S_T is the future price and S_0 is the current price. Q is called the "risk-neutral measure" and $\mathbf{P}$ is called the "physical measure".

[16] $\langle N, W \rangle_t = -\left\langle \int_0^t r_s \, dW_s, \int_0^t dW_s \right\rangle_t = -\int_0^t r_s \, ds$.

$M_t := \int_0^t F_s \, dB_s$ is an $\mathcal{F}_t^0$-martingale if $F \in \mathcal{L}^2_{\mathcal{F}}(0,T;\mathbb{R})$.

Theorem 3.5.8 (Martingale representation theorem)**.** *Let $M = (M_t)_{t\geq 0}$ be a square-integrable martingale on $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_t^0)_{t\geq 0})$. Then there exists $F = (F_t)_{t\geq 0} \in \mathcal{L}^2_{\mathcal{F}}(0,T;\mathbb{R})$ s.t. $M_t = M_0 + \int_0^t F_s \, dB_s$.*

4 Stochastic Differential Equations

4.1 Introduction

Consider

$$dX_t^j = f_0^j(t, X_t)dt + \sum_{i=1}^n f_i^j(t, X_t)dB_t^i, \; j = 1, 2, \dots, N \tag{4.1.1}$$

where $B_t = (B_t^1, \dots, B_t^n)_{t\geq 0}$ is a standard BM in $\mathbb{R}^n$ on $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_t)_{t\geq 0})$, and

$$f_i^j : [0, \infty) \times \mathbb{R}^N \to \mathbb{R}, \; i = 1, 2, \dots, n, \; j = 1, 2, \dots, N$$

are Borel measurable functions.

Note that (4.1.1) should be interpreted as

$$X_t^j = X_0^j + \int_0^t f_0^j(s, X_s)ds + \sum_{i=1}^n \int_0^t f_i^j(s, X_s)dB_s^i. \tag{4.1.2}$$ [17]

Definition 4.1.1. 1. *An adapted, continuous, $\mathbb{R}^N$-valued process $X = (X_t)_{t\geq 0} = (X_t^1, \dots, X_t^N)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_t)_{t\geq 0})$ is a* weak solution *of* (4.1.1) *if there is a standard BM $W = (W_t)_{t\geq 0}$ in $\mathbb{R}^n$, adapted to $(\mathcal{F}_t)_{t\geq 0}$, s.t.* (4.1.2) *holds with B replaced by W.*

2. *Given a standard BM $B = (B_t)_{t\geq 0}$ in $\mathbb{R}^n$ on $(\Omega, \mathcal{F}, \mathbf{P})$ with filtration $(\mathcal{F}_t^0)_{t\geq 0}$, an adapted continuous process $X = (X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_t^0)_{t\geq 0})$ is a* strong solution *of* (4.1.1) *if* (4.1.2) *holds.*

Definition 4.1.2. 1. *We say that* (4.1.1) *has* pathwise uniqueness *if whenever (X, B) and $(\tilde{X}, B)$ are two weak solutions defined on the same space $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_t)_{t\geq 0})$, and $X_0 = \tilde{X}_0$, then $X = \tilde{X}$ a.s..*

2. *We say that* (4.1.1) *has* uniqueness in law *if whenever (X, B) and $(\tilde{X}, \tilde{B})$ are two weak solutions (possibly defined on different probability spaces) and X_0 and $\tilde{X}_0$ have the same distribution, then X and $\tilde{X}$ have all the same finite-dimensional distributions.*

Theorem 4.1.3 (Yamada-Watanabe)**.** *Pathwise uniqueness implies uniqueness in law.*

Example 7 (Hiroshi Tanaka)*.* Consider a one-dimensional SDE

$$X_t = \int_0^t \operatorname{sgn}(X_s)dB_s, \; 0 \leq t < \infty$$

where $\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$

Then

1. Uniqueness in law holds: X_t is a continuous martingale and $\langle X \rangle_t = \int_0^t 1ds = t \implies X_t$ is BM.
2. If (X, B) is a weak solution:

   $$-X_t = \int_0^t -\operatorname{sgn}(X_s)dB_s = \int_0^t \operatorname{sgn}(-X_s)dB_s \implies (-X, B) \text{ is also a weak solution.}$$

   Hence pathwise uniquess does not hold.
3. Let W_t be any standard BM on $(\Omega, \mathcal{F}, \mathbf{P})$ and then let $B_t := \int_0^t \operatorname{sgn}(W_t)dW_t$. $(B_t)_{t\geq 0}$ is BM and $\int_0^t \operatorname{sgn}(W_s)dB_s = \int_0^t \underbrace{\operatorname{sgn}(W_s)^2}_{=1} dW_s = W_t$.

   Hence (W_t, B_t) is a weak solution.
4. There is no strong solution (see Roger & Williams 1990, p. 151).

[17] The stochastic differential equation (SDE) is defined by the fs. The underlying probability space and the BM are *not* part of the definition of the SDE.

4.2 Several examples

4.2.1 Linear SDEs

Consider

$$dX_t = \beta X_t dt + \sigma dB_t, \tag{4.2.1}$$

where X_t and B_t are $N \times 1$ column vectors, β is an $N \times N$ matrix and σ is an $N \times n$ matrix.
Using integration by parts,

$$\begin{aligned}
e^{-\beta t}X_t - e^{-\beta 0}X_0 &= \int_0^t e^{-\beta s} dX_s + \int_0^t d(e^{-\beta s})X_s{}^{18} \\
&= \int_0^t e^{-\beta s} dX_s - \int_0^t e^{-\beta s}\beta X_s ds \\
&= \int_0^t e^{-\beta s}(dX_s - \beta X_s ds) \\
&= \int_0^t e^{-\beta s}\sigma dB_s \\
\implies \qquad X_t &= e^{\beta t}X_0 + \int_0^t e^{\beta(t-s)}\sigma dB_s.
\end{aligned}$$

Ornstein-Uhlenbeck process

$$dX_t = dB_t - (AX_t)dt.$$

4.2.2 Geometric BM

The Black-Scholes model is

$$dS_t = S_t(\mu_t dt + \sigma_t dB_t).$$

Regarding the parenthesised term as "X_t", we see that we have a stochastic exponential:

$$\begin{aligned}
S_t &= \mathcal{E}(X)_t \\
&= \mathcal{E}\left(\int_0^t \mu_s ds + \int_0^t \sigma_s dB_s\right) \\
&= S_0 \exp\left(\int_0^t \mu_s ds + \int_0^t \sigma_s dB_s - \frac{1}{2}\int_0^t \sigma_s^2 ds\right) \\
&= S_0 \exp\left(\int_0^t [\mu_s - \tfrac{1}{2}\sigma_s^2] ds + \int_0^t \sigma_s dB_s\right).
\end{aligned}$$

If $\mu_t \equiv \mu$ and $\sigma_t \equiv \sigma$ then

$$S_t = S_0 \exp(\sigma B_t + [\mu - \tfrac{1}{2}\sigma^2]t),$$

hence S_t follows a lognormal distribution, because $B_t \sim \mathrm{N}(0, t)$ and the remaining terms are constants.

4.2.3 Cameron-Martin's formula

Consider the SDE

$$dX_t = dB_t + b(t, X_t)dt \tag{4.2.2}$$

where b is bounded and measurable on $[0, T] \times \mathbb{R}$.
Let $(W_t)_{t\geq 0}$ be a standard BM on $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_t)_{t\geq 0})$.
Define

$$\left.\frac{d\mathbf{Q}}{d\mathbf{P}}\right|_{\mathcal{F}_t} = \mathcal{E}(N)_t,\ t \geq 0,$$

[18]Recall that the matrix exponential is defined as $e^{\beta t} := \sum_{k=0}^{\infty} \frac{t^k}{k!}\beta^k$ and so we have $\frac{d}{dt}e^{\beta t} = \beta e^{\beta t} = e^{\beta t}\beta$.

where $N_t := \int_0^t b(s, W_s) dW_s$ is a **P**-martingale with $\langle N \rangle_t = \int_0^t b^2(s, X_s)\ ds$. By Novikov's condition,

$$\mathcal{E}(N)_t \equiv \exp\left(\int_0^t b(s, W_s) dW_s - \frac{1}{2}\int_0^t b^2(s, W_s) ds\right)$$

is a **P**-martingale.

By corollary 3.5.7,

$$B_t := W_t - \langle W, N \rangle_t = W_t - \int_0^t b(s, W_s) ds$$

is a continuous local **Q**-martingale. Moreover, $\langle B \rangle_t = \langle W \rangle_t = t$.

Hence, $(B_t)_{t \geq 0}$ is a standard BM under **Q**.

However, $dW_t = dB_t + b(t, W_t)dt \implies (W, B)$ is a weak solution of (4.2.2) on $(\Omega, \mathcal{F}, \mathbf{Q})$.

Theorem 4.2.1 (Cameron-Martin's formula)**.** *Let $b(t,x) = (b^1(t,x), \ldots, b^n(t,x))$ be bounded measurable functions on $[0,T] \times \mathbb{R}^n$ and $W_t = (W_t^1, \ldots, W_t^n)$ be a standard BM on $(\Omega, \mathcal{F}, \mathbf{P}; (\mathcal{F}_t)_{t\geq 0})$. Define probability measure **Q** by*

$$\left.\frac{d\mathbf{Q}}{d\mathbf{P}}\right|_{\mathcal{F}_t} := \exp\left(\sum_{k=1}^n \int_0^t b^k(s, W_s) dW_s^k - \frac{1}{2}\sum_{k=1}^n \int_0^t |b^k(s, W_s)|^2 ds\right),\ 0 \leq t \leq T.$$

*Then $(W_t)_{t\geq 0}$ under **Q** is a solution to*

$$dX_t^j = dB_t^j + b^j(t, X_t)dt$$

*for some BM $B_t = (B_t^1, \ldots, B_t^n)$ under **Q**.*

4.3 Existence and uniqueness

Consider the SDE

$$dX_t^j = f_0^j(t, X_t)dt + \sum_{i=1}^n f_i^j(t, X_t) dB_t^j,\ j = 1, 2, \ldots, N \tag{4.3.1}$$

Theorem 4.3.1. *Suppose that f_i^j $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, N$ satisfy the Lipschitz condition*

$$|f_i^j(t,x) - f_i^j(t,y)| \leq c|x-y|\ \forall x, y \in \mathbb{R}^N,\ t \geq 0$$

for some constant $c > 0$ and the linear growth condition

$$f_i^j(t,x)| \leq c(1 + |x|)\ \forall x \in \mathbb{R}^N,\ t \geq 0.$$

Then for any $\eta \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$ and a standard BM B_t in $\mathbb{R}^n$, there is a strong solution $(X_t)_{t\geq 0}$ of (4.3.1) with $X_0 = \eta$. Moreover, the pathwise uniqueness holds, which implies uniqueness in law by theorem 4.1.3 (Yamada-Watanabe).

Consider example 7 (Hiroshi Tanaka). There is no strong solution and no pathwise uniqueness. Note that sgn doesn't satsifiy the Lipschitz condition.

Example 8 (Itô and Watanabe, 1978)*.* The SDE

$$dX_t = 3X_t^{2/3} dB_t + 3X_t^{1/3} dt$$

has infinitely many strong solutions of the form

$$X_t^{(\theta)} := \begin{cases} 0 & \text{if } 0 \leq t < \tau_\theta \\ B_t^3 & \text{if } \tau_\theta \leq t < \infty \end{cases}$$

where $0 \leq \theta < \infty$ and $\tau_\theta := \inf\{t \geq \theta : B_t = 0\}$.

If $t < \tau_\theta$, then $X_t^{(\theta)} = 0$ and so clearly it satisfies the SDE.

When $t \geq \tau_\theta$, we see that

$$dB_t^3 = 3B_t^2 dB_t + \frac{1}{2} \cdot 3 \cdot 2B_t dt = 3(B_t^3)^{2/3} + 3(B_t^3)^{1/3} dt.$$

by applying Itô's formula.