Chapter (1)

Brownian Motion and Local Martingales

Section (1.1): Brownian Motion and Brownian Paths

This section is devoted to the construction and some properties of one of probability theory's most fundamental objects. Brownian motion earned its name after R. Brown, who observed around 1827 that tiny particles of pollen in water have an extremely erratic motion. It was observed by Physicists that this was due to an important number of random shocks undertaken by the particles from the (much smaller) water molecules in motion in the liquid. A. Einstein established in 1905 the first mathematical basis for Brownian motion, by showing that it must be an isotropic Gaussian process. The first rigorous mathematical construction of Brownian motion is due to N. Wiener in 1923, using Fourier theory.

In order to motivate the introduction of this object, we first begin by a "microscopical" depiction of Brownian motion. Suppose $(X_n, n \geq 0)$ is a sequence of $\mathbb{R}^d$ valued random variables with mean 0 and covariance matrix $\sigma^2 I_d$, which is the identity matrix in d dimensions, for some $\sigma^2 > 0$. Namely, if $X_1 = (X_1^1, \ldots, X_1^d)$,

$$E\left[X_1^i\right] = 0,\ E\left[X_1^i X_1^j\right] = \sigma^2 \delta_{ij},\ 1 \leq i,\ j \leq d.$$

We interpret X_n as the spatial displacement resulting from the shocks due to water molecules during the *n-th* time interval, and the fact that the covariance matrix is scalar stands for an isotropy assumption (no direction of space is privileged).

From this, we let $S_n = X_1 + \cdots + X_n$ and we embed this discrete-time process into continuous time by letting

$$B_t^{(n)} = n^{-1/2} S_{[nt]},\ t \geq 0.$$

Let $|\cdot|$ be the Euclidean norm on R^d and for $t > 0$ and $X, y \in \mathrm{R}^d$, define

$$pt(x) = \frac{1}{(2\pi t)^{d/2}} \exp\left(\frac{|x|^2}{2t}\right),$$

Which is the density of the Gaussian distribution $\mathrm{N}(0, tI_d)$ with mean 0 and covariance matrix tI_d. By convention, the Gaussian law $\mathrm{N}(m, 0)$ is the Dirac mass at m.

Proposition (1.1.1):

Let $0 = t_1 \geq t_2 < \ldots < t_k$. Then the finite marginal distributions of $B^{(n)}$ with respect to times $t_1, \ldots, t_k$ converge weakly as $n \to \infty$. More precisely, if F is a bounded continuous function, and letting $x_0 = 0,\quad t_0 = 0$,

$$\mathrm{E}\left[F(B_{t_1}^{(n)}, \ldots, B_{t_k}^{(n)})\right] \underset{n \to \infty}{\longrightarrow} \int_{(\mathrm{R}^d)^k} F(x_1, \ldots x_k) \prod_{1 \leq i \leq k} p_{\sigma^2}(t_i - t_{i-1})(x_i - x_{i-1}) dx_i.$$

Otherwise said, $(B_{t_1}^{(n)}, \ldots, B_{t_k}^{(n)})$ converges in distribution to $\left(G_1, G_2, \ldots, G_k\right)$, which is a random vector whose law is characterized by the fact that $\left(G_1, G_2 - G_1, \ldots, G_k - G_{k-1}\right)$ are independent centered Gaussian random variables with respective covariance matrices $\sigma^2(t_i - t_{i-1}) I_d$.

Proof: With the notations of the theorem, we first check that $\left(B_{t_1}^{(n)}, B_{t_2}^{(n)} - B_{t_1}^{(n)}, \ldots, B_{t_k}^{(n)} - B_{t_{k-1}}^{(n)}\right)$ is a sequence of independent random variables. Indeed, one has for $1 \leq i \leq k$

$$B^{(n)}_{t_i} - B^{(n)}_{t_{i-1}} = \frac{1}{\sqrt{n}} \sum_{i=[nt_{i-1}]+1}^{[nt_i]} X_j,$$

and the independence follows by the fact that $\left(X_j \ , j \geq 0\right)$is an i.i.d. family. Even better, we have the identity in distribution for the ***i***-th increment

$$B^{(n)}_{t_i} - B^{(n)}_{t_{i-1}} \underline{\underline{d}} \frac{\sqrt{[nt_i]-[nt_{i-1}]}}{\sqrt{n}} \frac{1}{\sqrt{[nt_i]-[nt_{i-1}]}} \sum^{[nt_i]-[nt_{i-1}]} X_j$$

and the central limit theorem shows that this converges in distribution to a Gaussian law $\mathrm{N}\ (0,\sigma^2(t_i - t_{i-1})I_d$. Summing up our study, and introducing characteristic functions, we have shown that for every $\xi=(\xi_j,1\leq j \leq k)$,

$$\mathrm{E}\left[\exp\left(i\sum_{j=1}^{k}\xi_j(B^{(n)}_{t_j} - B^{(n)}_{t_{j-1}})\right)\right] = \prod_{j=1}^{k}\exp\left(i\xi_j(B^{(n)}_{t_j} - B^{(n)}_{t_{j-1}})\right)$$

$$\underset{n\to\infty}{\longrightarrow}\prod_{j=1}^{k}\exp\left(i\xi_j(G_j - G_{j-1})\right)$$

$$= = = \equiv \mathrm{E}\left[\exp\left(i\prod_{j=1}^{k}\xi_j(G_j - G_{j-1})\right)\right],$$

where G_1 , . . . , $G_k$is distributed as in the statement of the proposition. By Levy's convergence theorem we deduce that increments of $B^{(n)}$between times $t_i$converge to increments of the sequence G_i, which is easily equivalent to the statement.

This suggests that $B^{(n)}$should converge to a process B whose increments are independent and Gaussian with covariances dictated by the above formula. This will be set in a rigorous way later in the research, with Donsker's invariance theorem.

Definition (1.1.2):

An $\mathbb{R}^d$-valued stochastic process $(B_t \ , \ t \geq 0)$ is called a standard Brownian motion if it is a continuous process, that satisfies the following conditions:

(i) $B_0 = 0$ a.s.,

(ii) For every $0 = t_0 \leq t_1 \leq t_2 \leq ... \leq t_k$, the increments $(B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, ..., B_{t_k} - B_{t_{k-1}})$ are independent, and

(iii) for every $t, s \geq 0$, the law of $B_{t+s} - B_t$ is Gaussian with mean 0 and covariance sI_d.

The term "standard" refers to the fact that B_1 is normalized to have variance I_d, and the choice $B_0 = 0$.

The characteristic properties (i), (ii), (iii) exactly amount to say that the finite dimensional marginal's of a Brownian motion are given by the formula of Proposition (1.1.1). Therefore the law of the Brownian motion is uniquely determined. We now show Wiener's theorem that Brownian motion exists!

Theorem (1.1.3): (Wiener)

There exists a Brownian motion on some probability space. We will first prove the theorem in dimension $d = 1$ and construct a process $(B_t, 0 \leq t \leq 1)$ satisfying the properties of a Brownian motion. This proof is essentially due to P. Levy in 1948. Before we start, we will need the following lemma.

Lemma (1.1.4):

Let N be a standard Gaussian random variable. Then

$$\frac{x^{-1}-x^{-3}}{\sqrt{2\pi}}e^{-x^2/2} \le P(\mathcal{N} > x) \le \frac{x^{-1}}{\sqrt{2\pi}}e^{-x^2/2}.$$

(1.1)

Let $D_0 = \{0,\ 1\}$, $D_n = \{k2^{-n}\ ,\ 0 \le k \le 2^n\}$ for $n \ge 1$, and $D = U_{n\ge 0} D_n$ be the set of dyadic rational numbers in [0, 1]. On some probability space (Ω, $\mathcal{F}$, P), let $(Z_d, d \in D)$ be a collection of i.i.d. random variables all having a Gaussian distribution $\mathcal{N}$ (0, 1) with mean 0 and variance 1. It is a well-known and important fact that if the random variables $X_1, X_2, \ldots$" are linear combinations of independent centered Gaussian random variables, then $X_1, X_2, \ldots$" are independent if and only if they are pair-wise uncorrelated, namely $Cov(X_i\ , X_j) = \mathrm{E}[X_i\ X_j] = 0$ for every $i \ne j$.

We set $X_0 = 0$ and $X_1 = Z_1$. Inductively, given ($X_d^{n-1}, d \in D_{n-1}$), we build $(X_d^n, d \in D_n)$ in such a way that $(X_d^n, d \in D_n)$satisfies (i), (ii), (iii) in the definition of the Brownian motion (where the instants under consideration are taken in D_n).

To this end, take $d \in D_n \setminus D_{n-1}$, and let $d_- = d - 2^{-n}$ and $d_+ = d + 2^{-n}$ so that d_-, d_+ are consecutive dyadic numbers in D_{n-1}. Then define:

$$X_d^n = \frac{X_{d_-}^{n-1} + X_{d_+}^{n-1}}{2} + \frac{Z_d}{2^{(n+1)/2}}$$

and put $X_{d-}^n = X_{d-}^{n-1}$ and $X_{d_+}^n = X_{d_+}^{n-1}$. Note that with these definitions,

$$X_d^n - X_{d_-}^n = N_d + N_d^{'}$$

$$X_{d_+}^n - X_d^n = N_d - N_d^{'}$$

Where

$$N_d = (X^{n-1}_{d_+} - X^{n-1}_{d_-})/2,\ N'_d = Z_d / 2^{(n+1)/2}$$

are by the induction hypothesis two independent centered Gaussian random variables with variance 2^{-n-1}. From this, one deduces

$$Cov(N_d + N'_d, N_d - N'_d) = Var(N_d) - Var(N'_d) = 0,$$

so that the increments $X^n_d - X^n_{d_-}$ and $X^n_{d_+} - X^n_d$ are independent with variance 2^{-n}, as should be. Moreover, these increments are independent of the increments $X^n_{d'+2^{-n-1}} - X^n_{d'}$ for $d' \in D_{n-1}, d' \neq d_-$ and of $Z_{d'}, d' \in D_n \setminus D_{n-1}, d' \neq D$ so they are independent of the increments $X_{d''+2^{-n}} - X_{d''}$ for $d'' \in D_n, d'' \notin \{d_-, d\}$. This allows the induction argument to proceed one step further.

We have thus defined a process $(X^n_d, d \in D_n)$, which satisfies properties (i), (ii) and (iii) for all dydadic times $t_1,\ t_2\ , \ldots,\ t_k \in D_n$. Observe that if $D \in D_n, X^m_d = X^n_d$ for all $m \geq n$. Hence for all $d \in D$,

$$B_d = \lim_{m\to\infty} X^m_d$$

is well-defined and the process $(B_d, d \in D)$ obviously satisfies (i), (ii) and (iii). To extend this to a process defined on the entire interval [0, 1], we proceed as follows. Define, for each $n \geq 0$, a process $X_n(t), 0 \leq t \leq 1$, to be the linear interpolation of the values $(B_d, d \in D_n)$ the dyadic times at level n. Note that if $d \in D$, say $d \in D_m$ with $m \geq 0$, then for any $n \geq m$, $X_n(d) = X_m(d) = B_d$. Furthermore, define an event A_n by

$$A_n = \left\{ \sup_{0\leq t\leq 1} | X_n(t) - X_{n-1}(t) | > 2^{-n/4} \right\}.$$

We then have, by Lemma (1.1.4), if N is a standard gaussian random variable:

$$\mathrm{P}(A_n) = \mathrm{P}\left(\bigcup_{j=0}^{2^{n-1}} \sup_{t\in[(2j)2^{-n},(2j+2)2^{-n}]} |X_n(t) - X_{n-1}(t)| > 2^{-n/4} \right).$$

$$\leq \sum_{j=0}^{2^{n-1}} \mathrm{P}\left(\frac{|Z_{(2j+1)2^{-n}}|}{2^{(n+1)/2}} > 2^{-n/4} \right)$$

$$\leq \sum_{k=0}^{2^{n-1}} \mathrm{P}(|\mathrm{N}| > 2^{(n+2)/4})$$

$$\leq \pi^{-1/2} 2^{3n/4} \exp(-2^{n/2})$$

We conclude that

$$\sum_{n=0}^{\infty} \mathrm{P}(A_n) < \infty$$

and by Borel-Cantelli, the events A_n occur only finitely often. We deduce immediately that the sequence of functions X_n is almost surely Cauchy in $C(0, 1)$ equipped with the topology of uniform convergence, and hence X_n converges toward a continuous limit function $(\tilde{X}(t), 0 \leq t \leq 1)$ uniformly, almost surely. Since $X_n(t)$ is constantly equal to X_t for $t \in D$ and for n large enough, it must be that $X_t = \tilde{X}(t)$ for all $t \in D$. Thus $\tilde{X}$ is a continuous extension of X, and we still denote this extension by X. We now deduce properties (i), (ii) and (iii) for X by continuity and the fact that X_n satisfies these properties. Indeed, let $k \geq 1$ and let $0 < t_1 < t_2 ... < t_k < 1.$ Fix $\alpha_1, \alpha_2, ..., \alpha_k \geq 0.$ For every $1 \geq i \geq k,$ fix a sequence $(d_i^{(n)})_t^{\infty}$ such that $\lim_{n\to\infty} d_i^{(n)} = t_i,$ and assume (since D is dense in [0, 1]) that $d^{(n)} \in D$ and $t_{i-1} < d_i^{(n)} \leq t_i$. Then by Lebesgue's dominated convergence theorem:

$$\mathrm{E}\left[\exp\{i\alpha_1 X_{t_1} + i\alpha_2 (X_{t_2} - X_{t_1}) + ... + i\alpha_k (X_{t_k} - X_{t_{k-1}})\} \right]$$

$$= \lim_{n\to\infty} \mathrm{E}\left[\exp\{i\alpha_1 X_{d_1^{(n)}} + i\alpha_2 (X_{d_2^{(n)}} - X_{d_1^{(n)}}) + ... + i\alpha_k (X_{d_k^{(n)}} - X_{d_{k-1}^{(n)}})\} \right]$$

$$= \lim_{n\to\infty} \exp\left\{ -\frac{\alpha_1^2}{2} d_1^{(n)} - ... - \frac{\alpha_k^2}{2} (d_k^{(n)} - d_{k-1}^{(n)}) \right\}$$

$$= \exp\left\{ -\frac{\alpha_1^2}{2} t_1 - ... - \frac{\alpha_k^2}{2} (t_k - t_{k-1}) \right\}.$$

It is now easy to construct a Brownian motion indexed by R_+ . Simply take independent standard Brownian motions $(B_t^i, 0 \le t \le 1), i \ge 0$ as we just constructed, and let

$$B_t = \sum_{i=0}^{\lfloor t \rfloor - 1} B_1^i + B_{t-\lfloor t \rfloor}^{\lfloor t \rfloor},\ t \ge 0.$$

It is easy to check that this has the wanted properties. Finally, it is straightforward to build a Brownian motion in $\mathbb{R}_d$, by taking d independent copies $B^1,...,\mathrm{B}^d$ of B and checking that $((B_t^1,...,B_t^d), t \ge 0)$ is a Brownian motion in $\mathbb{R}^d$.

Remark(1.1.5):

The extension of $(B_d, d \in D)$ could have been obtained by appealing to the existence of a continuous modification, whose existence is provided by Kolmogorov's criterion below.

Now- we discuss continuity and Holder continuity of Brownian paths.

We construct Brownian motion which directly yields a random process satisfying the three properties defining a Brownian motion, and which was at the same time continuous. In fact, and that is the reason why continuity is part of Definition (1.1.2), the next theorem will imply that any process satisfying (i), (ii) and (iii) can be slightly modified so that its trajectories are a.s continuous. The result is in fact much more general than that. As a consequence, we establish stronger regularity properties for Brownian motion than mere continuity: we prove that the path is almost surely Holder with exponent $1/2-\varepsilon$ for all $\varepsilon>0$. To start with, we need to introduce the concept of version (modification) and indistinguishable versions.

Definition (1.1.6):

If X and X' are two processes defined on some common probability space $(\Omega, \mathcal{F}, P)$, we say that is a version of X if for every t, $p(X_t(\omega)) = X_t^{'}(\omega)) = 1$.

In particular, two versions X and X' of the same process share the same finite-dimensional distribution, however, this does not say that there exists an ω so that $X_t(\omega) = X_t^{'}(\omega)$ for every t. This becomes true if both X and X' are a *priori* known to be continuous or cadlag, for instance. When the two trajectories coincide almost surely for all $t \geq 0$, we say that X and X' are indistinguishable:

Definition (1.1.7):

If X and X' are two processes defined on some common probability space $(\Omega, \mathcal{F}, \mathrm{P})$, we say that X' is an indistinguishable version of X

$$\mathrm{P}\ (X_t(\omega) = X_t^{'}(\omega) \quad \text{for all t})=1$$

Note that, up to in distinguish ability, there exists at most one continuous modification of a given process $(X_t, t \geq 0)$. Kolmogorov's criterion is a fundamental result which guarantees the existence of a continuous version (but not necessarily indistinguishable version) based solely on an L^p control of the two-dimensional distributions. We will apply to Brownian motion below, but it is useful in many other contexts.

Theorem (1.1.8): (Kolmogorov's Continuity Criterion)

Let $(X_t, 0 \le t \le 1)$ be a stochastic process with real values. Suppose there exist $p > 0, c > 0, \varepsilon > 0$ so that for every $s,t \ge 0$,

$$\mathrm{E}\left[\left|X_t - X_s\right|^p\right] \le c\left|t-s\right|^{1+\varepsilon}$$

Then, there exists a modification $\tilde{X}$ of X which is a.s. continuous, and even $\alpha - H\ddot{o}lder$ continuous for any $\alpha \in (0, \varepsilon / p)$.

Proof: Let $D_n = \{k.2^{-n}, 0 \le k \le 2^n\}$ denote the dyadic numbers of $[0,1]$ with level *n*, so D_n increases as n increases. Then letting $\alpha \in (0, \varepsilon / p)$, Markov's inequality gives for $0 \le k < 2^n$,

$$\mathrm{P}\left(\left|X_{k2^{-n}} - X_{(k+1)2^{-n}}\right| > 2^{-n\alpha}\right) \le \mathrm{E}\left[\left|X_{k2^{-n}} - X_{(k+1)2^{-n}}\right|^p\right] c2^{np\alpha}2^{-n-n\varepsilon} \le c2^{-n}2^{-(\varepsilon - p\alpha)n}.$$

Summing over D_n we obtain

$$\mathrm{P}\left(\sup_{0 \le k < 2^n}\left|X_{k2^{-n}} - X_{(k+1)2^{-n}}\right| > 2^{-n\alpha}\right) \le c2^{-n(\varepsilon - p\alpha)},$$

which is summable. Therefore, the Borel-Cantelli lemma shows that for a.a. ω, there exits N_ω so that if $n \ge N_\omega$, the supremum under consideration is $\le 2^{-n\alpha}$. Otherwise said, a.s.,

$$\sup_{n \ge 0} \sup_{k \in \{0...,2^n - 1\}} \frac{\left|X_{k2^{-n}} - X_{(k+1)2^{-n}}\right|}{2^{-n\alpha}} \le M(\omega) < \infty, a.s.$$

We claim that this implies that for every

$$s,t \in D = U_{n \ge 0} D_n, | X_s - X_t | \le M'(\omega) | t - s |^\alpha, \text{ for some } M'(\omega) < \infty \text{ a.s.}$$

Indeed, if $s,t \in D, s < t$, and let r is the least integer such that $t - s > 2^{-r-1}$. Then there exists $0 \le k \le 2^r$ and integers $l, m \ge 0$ such that

$$s = k2^r - \varepsilon_1 2^{-r-1} - \ldots - \varepsilon_l 2^{-r-l}$$

and

$$t = k2^{-r} + \varepsilon_0^{'} 2^{-r} + \varepsilon_1^{'} 2^{-r-1} + \ldots + \varepsilon_m^{'} 2^{-r-m}$$

with $\varepsilon_i \varepsilon_i^{'} \in \{0,1\}$. For $0 \le i \le l$, let

$$s_i = k2^{-r} - \varepsilon_1 2^{-r-1} - \ldots - \varepsilon_i 2^{-r-i}.$$

By the triangular inequality

$$|X_t - X_s| = |X_{t_m} - X_{s_l}| \le |X_{t_0} - X_{s_0}| + \sum_{i=1}^{l} |X_{t_i} - X_{t_{i-1}}| + \sum_{j=1}^{m} |X_{s_j} - X_{s_{j-1}}|$$

$$\le M(\omega)2^{-r\alpha} + \sum_{i=1}^{l} M(\omega) 2^{-(r+i)\alpha} + \sum_{j=1}^{m} 2^{-(r+j)\alpha} M(\omega)$$

$$\le M(\omega)2^{-r\alpha}(1 + 2(1-2^{-\alpha})^{-1})$$

$$\le M'(\omega)|t-s|^{\alpha}$$

where $M'(\omega) = M(\omega)2^{\alpha}(1+2(1-2^{-\alpha})^{-1})$. Therefore, the process $(X_t, t \in D)$ is a.s. uniformly continuous (and even $\alpha - \textit{Hölder}$ continuous). Since D is an everywhere dense set in [0, 1], the latter process a.s. admits a unique continuous extension $\tilde{X}$ on [0, 1], which is also $\alpha - \textit{Hölder}$ continuous (it is consistently defined by $\tilde{X}_t = \lim_n X_{t_n}$, where $(t_n, n \ge 0)$ is any *D-valued* sequence converging to t). On the exceptional set where $(X_d, d \in D)$ is not uniformly continuous, we let $\tilde{X}_t = 0,\ 0 \le t \le 1$, so $\tilde{X}$ is continuous. It remains to show that $\tilde{X}$ is a version of *X*. To this end, we estimate by Fatou's lemma

$$\mathrm{E}\left[\left|X_t - \tilde{X}_t\right|^p\right] \le \lim_n \inf \mathrm{E}\left[\left|X_t - X_{t_n}\right|^p\right]$$

where $(t_n, n \ge 0)$ is any D-valued sequence converging to t. But since $\mathrm{E}\left[\left|X_t - X_{t_n}\right|^p\right] \le c\left|t - t_n\right|^{1+\varepsilon}$, this converges to 0 as $n \to \infty$. Therefore, $X_t = \tilde{X}_t$ a.s. for every t. From now on we will consider exclusively a continuous modification of Brownian motion, which is unique up to indistinguishability. As a corollary to Kolmogorov's criterion, we obtain the aforementioned result on the $\alpha - \textit{Hölder}$ properties of Brownian motion:

Corollary (1.1.9):

Let $(B_t, t \ge 0)$ be a standard Brownian motion in dimension 1. Almost surely, B is *Hölder*-continuous of order for any $0 < \alpha < 1/2$. More precisely, with probability 1, for

$$\sup_{n \le t, s \le n+1} \frac{|B_t - B_s|}{|t - s|^\alpha} < \infty \tag{1.2}$$

Proof: Let $s \le t \in D$, and notice that for every $p > 0$, since $B_t - B_s$ has the same law as $\sqrt{(t-s)N}$, (where N is a standard Gaussian random variable), we have $\mathrm{E}(|B_t - B_s|^p) \le M\,|t - s|^{1+\varepsilon}$ with $\varepsilon = p/2 - 1$ and $M = \mathrm{E}(|N|^p) < \infty$. For $p > 2, \varepsilon > 0$ and thus X is *Hölder* of order α for $\alpha < \varepsilon / p = 1/2 - 1/p$. Since $p > 2$ is arbitrary, then B is $\alpha - \textit{Hölder}$ for any $\alpha < 1/2$, almost surely.

Notice that the above corollary does not say anything about higher-order *Hölder* continuity: all we know is that the path is a.s. *Hölder* of order $\alpha < 1/2$. The next result tells us that this is, in some sense, sharp.

Theorem (1.1.10):

Let B be a continuous modification of Brownian motion. Let $\gamma > 1/2$. Then it holds:

$$\mathrm{P}\left(\forall t \geq 0 : \limsup_{h \to 0^+} \frac{|B_{t+h} - B_t|}{h^\gamma} = +\infty \right) = 1$$

Proof: We first observe that

$$\left\{ \exists t \geq 0 : \limsup_{h \to 0^+} \frac{|B_{t+h} - B_t|}{h^\gamma} < \infty \right\}$$

$$\subseteq \bigcup_{p=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \{ \exists t \in [0,m] : |B_{t+h} - B_t| \leq p h^\gamma, \forall h \in (0, 1/k) \}$$

Therefore, it suffices to show that

$$\mathrm{P}(\exists t \in [0,m] : |B_{t+h} - B_t| \leq p h^\gamma, \forall h \in (0,\delta)) = 0$$

For all $p \geq 1, m \geq 1, \delta > 0$. For $n \geq 1, 1 \leq i \leq mn-1$, define:

$$A_{i,n} = \left\{ \exists s \in \left[\frac{i}{n}, \frac{i+1}{n} \right] : |B_{s+h} - B_s| \leq p h^\gamma, \forall h \in (0,\delta) \right\}.$$

It suffices to show:

$$\lim_{n \to \infty} \sum_{i=0}^{mn-1} \mathrm{P}(A_{i,n}) = 0 \qquad (1.3)$$

Fix a large constant $K > 0$ to be chosen suitably later. We wish to exploit the fact that on the event $A_{i,n}$ many increments must be small. The trick is to be able to fix in advance the times at which these increments will be too small. More precisely, on $A_{i,n}$, as long as $n \geq (K+1)/\delta$, for all $1 \leq j \leq K$:

$$0 \le \frac{i+j}{n} - s \le \frac{K+1}{n} \le \delta$$

where s is as in the definition of $A_{i,n}$. Thus, taking $h = (i+j)/n - s$, on $A_{i,n}$:

$$| B_{\frac{i+j}{n}} - B_s | \le p\left(\frac{i+j}{n} - s\right)^{\gamma} \le p\left(\frac{K+1}{n}\right)^{\gamma}$$

If $2 \le j \le K$, by the triangular inequality:

$$\left| B_{\frac{i+j}{n}} - B_{\frac{i+j-1}{n}} \right| \le 2p\left(\frac{K+1}{n}\right)^{\gamma}$$

Therefore, there exists $C > 0$ such that for all $n \ge (K+1)/\delta$

$$\mathrm{P}(A_{i,n}) \le \mathrm{P}\left(\bigcap_{j=2}^{K}\left\{\left| B_{\frac{i+j}{n}} - B_{\frac{i+j-1}{n}} \right| \le 2p\left(\frac{K+1}{n}\right)^{\gamma}\right\}\right)$$

$$\le \prod_{j=2}^{K} \mathrm{P}\left(| \mathrm{N}\ (0,1/n) | \le 2p\left(\frac{k+1}{n}\right)^{r}\right)$$

$$\le \left[\mathrm{P}\left(| \mathrm{N}\ (0,1) | \le 2p\left(\frac{k+1}{n}\right)^{r} n^{\frac{1}{2}}\right)\right]^{K-1}$$

$$\le \left[2p\left(\frac{K+1}{n}\right)^{\gamma} n^{1/2}\right]^{K-1} = \frac{C}{n^{(\gamma-1/2)(K-1)}}$$

It follows that for all $n \ge (K+1)/\delta$:

$$\sum_{i=1}^{mn-1} \mathrm{P}(A_{i,n}) \le \frac{Cm}{n^{(\gamma-1/2)(K-1)-1}}$$

Thus if K is large enough that $(\gamma - 1/2)(K-1) > 1$, the right-hand side tends to 0 for all $n \ge n_0 = \lceil (K + 1/\delta) \rceil + 1$. This proves Equation(1.3), and, as a consequence, Theorem (1.1.8).

As a corollary to the last result, we obtain the celebrated Paley-Wiener-Zygmund theorem:

Corollary (1.1.11):

Almost surely, $t \mapsto B_t$ is nowhere differentiable

Now we study some Basic Properties.

Let $\Omega = C(\mathbb{R}_+, \mathbb{R}^d)$ be the 'Wiener space' of continuous functions, endowed with the product σ-algebra $\mathcal{W}$ (or the Borel σ-algebra associated with the compact-open topology).

Definition (1.1.12): (Wiener's Measure)

Let $(B_t, t \geq 0)$ be a Brownian motion, and let W be the law on Ω of B: that is, for any $A \in \mathcal{W}$,

$$\mathrm{W}(A) = \mathrm{P}((B_t, t \geq 0) \in A)$$

W is called Wiener's measure.

Of course, we must check that this definition makes sense, i.e. that W does not depend on the construction of B. To see this, note that the finite-dimensional (i.e, the joint law of $(B_{t_1}, ..., B_{t_n})$ are entirely specified by the definition of a Brownian motion. Since the σ-field $\mathcal{W}$ is generatedbycylinder events of the form $\{X_{t_1} \in A_1, ..., X_{t_n} \in A_n\}$, the right-hand side in the above display is indeed uniquely specifies.

Definition (1.1.13):

We now think of Ω as our probability space. For $\omega \in \Omega$ define:

$$X_t(\omega) = \omega(t), t \geq 0$$

We call $(X_t(\omega), t \geq 0)$ the canonical process. Then $(X_t, t \geq 0)$, under the probability measure W, is a Brownian motion. This is the canonical construction of Brownian motion.

Remark (1.1.14):

It is rarely that case in probability theory that we put some emphasis on the probability space on which a certain random process is constructed. (In all practical cases, we usually assume that such a random process is given to us). However the full advantage of specifying the probability space and measure will come to light we deal with Girsanov's change of measure theorem.

For $x \in \mathbb{R}^d$ we also let $\mathrm{W}_x(dw)$ be the image measure of W by $(w_t, t \geq 0) \mapsto (x + w_t, t \geq 0)$. A (continuous) process with law $\mathrm{W}_x(dw)$ is called a Brownian motion started at x. We let $(\mathcal{F}_t^B, t \geq 0)$ be the natural filtration of $(B_t, t \geq 0)$, completed by zero-probability events.

Definition (1.1.15):

We say that state B is Brownian motion (started at X) if $(B_t - X, t \geq 0)$ is a standard Brownian motion which is independent of X.

Otherwise said, it is the same as the definition as a standard Brownian motion, except that we do not require that $B_0 = 0$. if we want to express this on the Wiener space with the Wiener measure, we have for every measurable functional $F : \Omega \to \mathbb{R}_+$,

$$\mathbf{E}[F(B_t, t \geq 0)] = \int_{\mathbb{R}^d} \mathrm{P}(X \in dx) \int_{\Omega} \mathrm{W}(dw) F(x + w(t), t \geq 0) = \int_{\mathbb{R}^d} \mathrm{P}(X \in dx) \mathrm{W}_x(F).$$

It will be handy to use the notation $\mathrm{W}_X(F)$ for the random variable $\omega \mapsto W_{X(\omega)}(F)$, so that the right-hand side can be shortened as $\mathrm{E}(\mathrm{W}_X(F))$.

We now state some fundamental results, which are often referred to as the scaling properties of Brownian motion, or scale-invariance of Brownian motion.

Proposition (1.1.16):

Let B be a standard Brownian motion in $\mathbb{R}^d$.

1. Rotational invariance: If $U \in O(d)$ is an orthogonal matrix, then $UB = (UB_t, t \geq 0)$ is again a Brownian motion. In particular, $-B$ is a Brownian motion.
2. Scaling property: If $\lambda > 0$ then $(\lambda^{-1/2} B_{\lambda t}, t \geq 0)$ is a standard Brownian motion
3. Time-inversion: $(tB_{1/t}, t \geq 0)$ is also a Brownian motion (at $t = 0$ the process is defined by its value 0).

We now start to discuss ideas revolving around the Markov property of Brownian motion and its applications to path properties. We begin with the simple Markov property, which takes a particularly nice form in this context.

Theorem (1.1.17):

Let $(B_t, t \geq 0)$ be a Brownian motion, and let $s > 0$. Then

$$(\tilde{B}_t = B_{t+s} - B_s, t \geq 0)$$

s a Brownian motion, independent of the σ-field $\mathcal{F}_{s^+}^B = \bigcap_{t>s} \mathcal{F}_t^B$.

Proof: Since $\tilde{B}$ is continuous and $\tilde{B}_0 = 0$, to show that $\tilde{B}$ is a Brownian motion it suffices to check that the increments have the correct distribution. However if $t \geq u, \tilde{B}_t - \tilde{B}_u = B_{s+t} - B_{s+u}$ so this follows directly from the fact that B itself is a Brownian motion. It remains to show that $\tilde{B}$ is independent from $\mathcal{F}_{s^+}$. We start by checking independence with respect to $\mathcal{F}_s$, for which we can assume $d=1$. We will use this easy lemma, which is an important property worth remembering:

Lemma (1.1.18):

Let $s,t \geq 0$ Then

$$\mathrm{cov}(B_s, B_t) = s \wedge t.$$

Now, to prove independence of $\tilde{B}$ with respect to $\mathcal{F}_s$, it suffices to check that the finite-dimensional marginal's are independent: i.e., if $s_1 \leq ... s_m \leq s$ and $t_1 \leq ... t_n$, we want to show that

$$(B_{s_1}, ..., B_{s_m}) \text{ and } (\tilde{B}_{t_1}, ..., \tilde{B}_{t_n})$$

are independent. However, the $m+n$-coordinate vector $(B_{s_1}, ..., B_{s_m}, \tilde{B}_{t_1}, ..., \tilde{B}_{t_n})$ is a Gaussian vector (since it is the image by a linear application of a Gaussian vector), and it suffices to check that the covariance of two distinct terms is 0. Since each term has zero expectation:

$$Cov(\tilde{B}_{t_i}, B_{s_j}) = \mathrm{E}(\tilde{B}_{t_i} B_{s_j})$$

$$= \mathrm{E}(B_{s+t_i} B_{s_j}) - \mathrm{E}(B_s B_{s_j})$$

$$= s_j \wedge (s+t_i) - (s_j \wedge s) = s_j - s_j = 0$$

which proves the independence with respect to $\mathcal{F}_s$. If $A \in \mathcal{F}_{s^+}$, we wish to show that for every continuous functional $F : (\mathbb{R}^d)^k \to \mathbb{R}$ continuous and bounded,

$$\mathrm{E}(1_{\{A\}} F(\tilde{B}_{t_1}, ..., \tilde{B}_{t_k})) = \mathrm{P}(A)\mathrm{E}(F(\tilde{B}_{t_1}, ..., \tilde{B}_{t_k}))$$

Now, for any $\varepsilon > 0$, $A \in \mathcal{F}_{s^+} \subseteq \mathcal{F}_{s+\varepsilon}$, thus, using the property just proved:

$$E(1_A F(B_{t_1+s+\varepsilon} - B_{s+\varepsilon}, \ldots, B_{t_k+s+\varepsilon} - B_{s+\varepsilon})) = P(A)E(F(B_{t_1+s+\varepsilon} - B_{s+\varepsilon}, \ldots, B_{t_k+s+\varepsilon} - B_{s+\varepsilon}))$$

Letting $\varepsilon \to 0$ in the above identity, since B is continuous and F is bounded and continuous, we have (by Lebesgue's dominated convergence theorem),

$$E(1_{\{A\}} F(\tilde{B}_{t_1}, \ldots, \tilde{B}_{t_k})) = P(A)E(F(\tilde{B}_{t_1}, \ldots, \tilde{B}_{t_k}))$$

as required.

Theorem (1.1.19): (Blumenthal's Zero-One Law)

Let B be a standard Brownian motion. The σ-algebra $F_{0^+}^B = \bigcap_{\varepsilon>0} F_\varepsilon^B$ is trivial i.e. constituted of events of probability 0 or 1.

Proof: By the previous result, $(B_t, t \geq 0)$ is independent from F_{0+}. However F_∞^B contains F_{0+}^B, so this implies that the σ-field F_{0+} is independent of itself, and $P(A) = P(A \cap A) = P(A)^2$ by independence. Thus $P(A)$ is solution to the equation $x = x^2$ whose roots are precisely 0 and 1.

Proposition (1.1.20):

(i) For $d = 1$ and $t \geq 0$, let $S_t = \sup_{0 \leq s \leq t} B_s$ and $I_t = \inf_{0 \leq s \leq t} B_s$ (these are random variables because B is continuous). Then almost surely, for every $\varepsilon > 0$, one has

$$S_\varepsilon > 0 \text{ and } I_\varepsilon < 0$$

In particular, a.s. there exists a zero of B in any interval of the form $(0, \varepsilon), \varepsilon > 0$

(ii) A.s.,

$$\sup_{t \geq 0} B_t = -\inf_{t \geq 0} B_t = +\infty$$

(iii) Let C be an open cone in R^d with non-empty interior and origin at 0 (i.e., a set of the form $\{tu : t > 0, u \in A\}$, where A is a non-empty open subset of the unit sphere of R^d). If

$$H_C = \inf\{t > 0 : B_t \in C\}$$

is the first hitting time of C, then $H_C = 0$ a.s.

Proof:

(i) The probability that $B_t > 0$ is 1/2 for every t, so $P(S_t > 0) \geq 1/2)$, and therefore if $t_n, n \geq 0$ is any sequence decreasing to 0, $P(\limsup_n \{B_{t_n} > 0\}) \geq \limsup_n P(B_{t_n} > 0) = 1/2$ .Since the event $\limsup_n \{B_{t_n} > 0\}$ is in F_{0+} , Blumenthal's law shows that its probability must be 1. The same is true for the infimum by considering the Brownian motion -B.

(ii) Let $S_\infty = \sup_{t \geq 0} B_t$. By scaling invariance, for every $à > 0$, $à \cdot S_\infty = \sup_{t>0} à \cdot B_t$ has same law as $\sup_{t \geq 0} B_{à \cdot t} = S_\infty$. This is possible only if $S_\infty \in \{0, \infty\}$ a.s., however, it cannot be 0 by (i).

(iii) The cone C is invariant by multiplication by a positive scalar, so that $P(B_t \in C)$ is the same as $P(B_1 \in C)$ for every t by the scaling invariance of Brownian motion. Now, if C has nonempty interior, it is straightforward to check that $P(B_1 \in C) > 0$, and one concludes similarly as above.

We now want to prove an important analog of the simple Markov property, where deterministic times are replaced by stopping times. To begin with, we extend a little the definition of Brownian motion, by allowing it to start from a random location, and by working with filtrations that are (slightly) larger than the natural filtration of a standard Brownian motion.

Definition (1.1.21):

Let $(\mathcal{F}_t, t \geq 0)$ be a filtration. We say that a Brownian motion B is an $(\mathcal{F}_t)$-Brownian motion if B is adapted to $(\mathcal{F}_t)$, and if $B^{(t)} = (B_{t+s} - B_t, s \geq 0)$ is independent of $\mathcal{F}_t$ for every $t \geq 0$.

For instance, if $(\mathcal{F}_t)$ is the natural filtration of a 2-dimensional Brownian motion $(B_t^1, B_t^2, t \geq 0)$ then $(B_t^1, t \geq 0)$ is an $(\mathcal{F}_t)$-Brownian motion. If B' is a standard Brownian motion and X is a random variable independent of B', the $B = (X + B_t', t \geq 0)$ is a Brownian motion (started at $B_0 = X$), and is an $(\mathcal{F}_t = (\sigma(X) \vee \mathcal{F}_t^{B'})$-Brownian motion. A Brownian motion is always an $\mathcal{F}_t^B)$-Brownian motion. If B is a standard Brownian motion, then the completed filtration $\mathcal{F}_t = \mathcal{F}_t^B \vee \mathcal{N}$ ($\mathcal{N}$ being the set of events of probability 0) can be shown to be right-continuous, i.e. $\mathcal{F}_{t+} = \mathcal{F}_t$ for every $t \geq 0$, and B is an $(\mathcal{F}_t)$-Brownian motion.

Definition (1.1.22):

Let $\mathcal{F}$ be a filtration and let T be a stopping time. The σ-field $\mathcal{F}_T$ is defined by

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0\}$$

It is elementary (but tedious) that in the case of filtration generated by a process X, $\mathcal{F}_T = \sigma(X_{s \wedge T}, s \geq 0)$. In particular T and X_T are $\mathcal{F}_T$-measurable. This corroborates the intuition that $\mathcal{F}_T$ is the -algebra generated by all the events occurring prior to time T. We may now state the strong Markov property.

Theorem (1.1.23): (Strong Markov Property)

Let $(B_t, t \geq 0)$ be an $(\mathcal{F}_t)$-Brownian motion in $\mathbb{R}^d$ and T be an $(\mathcal{F}_t)$-stopping time. We let $B_t^{(T)} = B_{T+t} - B_T$ for every $t \geq 0$ on the event $\{T < \infty\}$, and 0 otherwise. Conditionally on $\{T < \infty\}$, the process $B^{(T)}$ is a standard Brownian motion, which is independent of $\mathcal{F}_T$. Otherwise said, conditionally given $\mathcal{F}_T$ and $\{T < \infty\}$, the process $(B_{T+t}, t \geq 0)$ is an $(\mathcal{F}_{T+t})$-Brownian motion started at B_T .

Proof: Suppose first that $T < \infty$ a.s. Let $A \in \mathcal{F}_T$, and consider times $t_1 < t_2 < \ldots < t_p$. We want to show that for every bounded continuous function F on $(\mathbb{R}^d)^p$,

$$\mathrm{E}\left[1_{\{A\}} F(B_{t_1}^{(T)}, \ldots, B_{t_p}^{(T)})\right] = P(A) E\left[F(B_{t_1}, \ldots, B_{t_p})\right] \qquad (1.4)$$

Indeed, taking $A = \Omega$ entails that $B^{(T)}$ is a Brownian motion, while letting A vary in $\mathcal{F}_T$ entails the independence of $(B_{t_1}^{(T)}, \ldots, B_{t_k}^{(T)})$ and $\mathcal{F}_T$ for every $t_1, \ldots, t_k$, hence of $B^{(T)}$ and $\mathcal{F}_T$.

$$\mathrm{E}\left[1_{\{A\}} F(B_{t_1}^{(T)}, \ldots B_{t_p}^{(T)})\right] = \lim_{n\to\infty} \sum_{k=1}^{\infty} \mathrm{E}\left[1_{\{A \cap \{(k-1)2^n < T \leq k2^{-n}\}\}}\right] F(B_{t_1}^{(k2^{-n})}, \ldots, B_{t_p}^{(k2^{-n})})$$

$$= \lim_{n\to\infty} \sum_{k=1}^{\infty} P(A \cap \{(k-1)2^{-n} < T \leq k2^{-n}\}) \mathrm{E}\left[F(B_{t_1}, \ldots, B_{t_p})\right]$$

$$= \mathrm{P}(A) \mathrm{E}\left[F(B_{t_1}, \ldots, B_{t_p})\right].$$

where we used the simple Markov property and the fact that $A \cap \{(k-1)2^{-n} < T \leq k2^{-n}s\} \in \mathcal{F}_{k2^{-n}}$ by definition. Finally, if $P(T = \infty) > 0$, check that Equation(1.4) remains true when replacing A by $A \cap \{T < \infty\}$, and divide by $\mathrm{P}(\{T < \infty\})$.

An important example of application of the strong Markov property is the so-called reflection principle. Recall that $S_t = \sup_{0\le s\le t} B_s$.

Theorem (1.1.24): (Reflection principle)

Let $0 < a$ and $b \le a$, then for every $t \ge 0$, $\mathrm{P}(S_t \ge a, B_t \le b) = \mathrm{P}(B_t \ge 2a - b)$

Proof: Let $T_a = \inf\{t \ge 0 : B_t \ge a\}$ be the first entrance time of B_t in $[a,\infty]$ for $a > 0$. Then T_a is an (F_T^B)-stopping time for every a and $T_a < \infty$ a.s. since $S_\infty = \infty$ a.s. where $S_\infty = \lim_{t\to\infty} S_t$..

Now by continuity of $B, B_{T_a} = a$ for every a. We thus have:

$$\mathrm{P}(S_t \ge a, B_t \le b) = \mathrm{P}(T_a \le t, B_t \le b)$$

$$= \mathrm{P}(T_a \le t, B^{(T_a)}_{t-T_a} \le b - a)$$

$$= \mathrm{P}(T_a \le t, -B^{(T_a)} \ge a - b).$$

Now, by the strong Markov property at time T_a, B^{T_a} is a Brownian motion independent of F_{T_a} and thus of T_a. In particular, we deduce that the joint law of $(T_a, B^{(T_a)})$ is identical to the joint law of $(T_a, -B^{(T_a)})$ by symmetry of Brownian motion. It follows that

$$\mathrm{P}(S_t \ge a, B_t \le b) = \mathrm{P}(T_a \le t, -B^{(T_a)} \ge a - b)$$

$$= \mathrm{P}(T_a \le t, B_t \ge 2a - b)$$

$$= \mathrm{P}(B_t \ge 2a - b).$$

Corollary (1.1.25):

We have the following identities in distribution: for all $t \ge 0$,

$$S_t =_d |B_t| =_d |\mathrm{N}(0,t)|.$$

Moreover, for every $x>0$, the random time T_x has same law as $(x/B_1)^2$.

Proof: We write, for all $t\geq 0$ and all $a\geq 0$,

$$\mathrm{P}(S_t\geq a)=\mathrm{P}(S_t\geq a,B_t\leq a)+\mathrm{P}(S_t\geq a,B_t\geq a)$$

$$=\mathrm{P}(B_t\geq 2a-a)+\mathrm{P}(B_t\geq a)$$

$$=2\mathrm{P}(B_t\geq a)=P(|B_t|\geq a)$$

since when $B_t\geq a, S_t\geq a$ automatically as well.

We end with a famous result of P . Levy on the quadratic variation of Brownian e motion. This result plays a fundamental role in the development of the stochastic integral. Let $(B_t,t\geq 0)$ be a standard Brownian motion. Let $t>0$ be fixed and for $n\geq 1$ let $\square_n=\{0=t_0(n)<t_1(n)<....<t_{m_n}(n)=t\}$ be a subdivision of $[0,t]$, such that

$$\eta_n=\max_{1\leq i\leq m_n}(t_i(n)-t_{i-1}(n))\xrightarrow[n\to\infty]{}0.$$

Theorem (1.1.26): (Levy)

$$\lim_{n\to\infty}\sum_{i=1}^{m_n}(B_{t_i}-B_{t_{i-1}})^2=t$$

Section(1.2): Finite Variation and Local Martingales

We now discuss the bigger picture in a nutshell.

We will now spend rather a lot of time to give a precise and rigorous construction of the stochastic integral, for as large a class of processes as possible, subject to continuity. This level of generality has a price, which is that the construction can appear quite technical without shedding any light on the sort of processes we are talking about. So before we embark on this journey, here are a few points which, in my opinion, guide the whole construction and should also guide your intuition throughout. What follows is only informal, and in particular, we do not describe issues related to measurability, and finiteness of the integral.

The real difficulty in the construction of the integral is in how to make sense of

$$\int_0^t H_s dM_s \tag{1.5}$$

where M is a martingale and H is, say, left-continuous or continuous. Even though dM does not make sense as a measure (the paths of martingales, just like Brownian paths, have too wild oscillations for that), it is easy to cook up a definition which makes intuitive sense when H is a simple process, that is, H takes only finitely many (bounded) values. Indeed, it suffices to require that the integral process in Equation (1.5) varies in the same way as M on the intervals over which H is constant, and has jumps when H does. A natural approach is then to try to extend this definition to more general classes of processes by "taking a limit" of integrals

$$\int_0^t H_s^n dM_s \to \int_0^t H_s dM_s \tag{1.6}$$

where the integrands in the left-hand side are simple and approximate H. In implementing this method, one faces several technical difficulties. The strategy is to construct a suitable function space where the sequence on the left-hand side of Equation (1.6) forms a Cauchy sequence. If the function space is complete, the sequence of integrals has a limit, which we may call the integral of H with respect to M. But we must also guarantee that this limit does not depend on the approximating sequence. It remains to find a space which has the desired properties. The key property which we will use (over and over again) is that martingales have a finite quadratic variation:

$$[M]_t = \lim_{n\to\infty} \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} (M_{(k+1)2^{-n}} - M_{k2^{-n}})^2 \tag{1.7}$$

exists and is finite, and is non-decreasing in *t*.

Furthermore, one can show (Theorem (2.1.10)) that $M_t^2 - [M]_t$ is a martingale. Now, when H is simple, it is not hard to convince yourself that the Equation (1.5) must also be a martingale. So what should be the t quadratic variation of $\int_0^t H_s dM_s$? Based on the approximation Equation (1.7), the amount of quadratic variation that we add to the integral between t and $t = dt$ is approximately $H_t^2 d[M]_t$. Hence any sensible definition of the stochastic integral must satisfy

$$\left[\int_0 H_s dM_s\right]_t = \int_0^t H_s^2 d[M]_s \tag{1.8}$$

The key insight of Ito was the realization that this property was sufficient to define the integral. Indeed, using the optional stopping theorem, this is essentially the same as requiring:

$$\mathrm{E}\left(\left(\int_0^\infty H_s dM_s\right)^2\right) = \mathrm{E}\left(\int_0^\infty H_s^2 d[M]_s\right). \tag{1.9}$$

Interpreting the right-hand side as an L^2 norm on the space of bounded integrands, this statement is saying that the stochastic integral must be a certain isometry between Hilbert spaces. The left-hand side shows that the correct space of martingales is the set of martingales with $\mathrm{E}([M]_\infty^2) < \infty,$, or, equivalently (as it turns out), martingales which are bounded in L^2. This space, endowed with the norm on the left-hand side of Equation (1.9) is indeed complete and simple processes are dense in it. Equation (1.9) is then relatively easy to prove for simple processes. This implies, at once, that the sequence in the left-hand side of Equation (1.6) is Cauchy (and hence has a limit), and the isometry property shows that this limit cannot depend on the approximating sequence.

At this point we have finished the construction of the stochastic integral for martingales which are bounded in L^2. Stopping at suitable stopping times, it is then easy to extend this definition to general martingales, or indeed to processes known as local martingales. Adding a "finite variation" component for which the Equation (1.5) is defined as a good old Stieltjes-Lebesgue integral finishes the construction for semi-martingales.

Having spoken about the bigger picture in a nutshell, it is now time to rewind the tape and go back to the beginning.

We now start to illustrate Finite variation integrals.

Finite variation processes are essentially those for which the standard notion of integral (the one you learn about in measure theory courses) is well-defined. Since finite variation is a path-wise property, we will first establish integrals with respect to deterministic integrants and lift it to stochastic processes in the following.

Recall that a function $f:\mathbb{R}\to\mathbb{R}$ is *cadlag* or *rcll* if it is right-continuous and has left limits. For such functions we write $\Delta f(t)=f(t)-f(t-)$ where $f(t-)=\lim_{s\uparrow t} f(s)$. Suppose $a:[0,\infty)\to\mathbb{R}$ is an increasing cadlag function. Then there exists a unique Borel measure da on $(0,\infty)$ such that $da((s,t])=a(t)-a(s),$, the Lebesgue-Stieltjes measure with distribution function a. Since da is a proper measure, there is no problem in defining, for any non-negative measurable function h and $t \geq 0$:

$$(h.a)(t) = \int_{(0,t]} h(s)da(s). \tag{1.10}$$

We may extend this definition to a cadlag function $a = a' - a''$, where a' and a'' are both increasing cadlag, and to integrable $h:[0,\infty) \to \mathbb{R}$. Subject to the finiteness of all the terms on the right we define

$$h.a = h^+.a' - h^+.a'' - h^-.a' + h^-.a''. \tag{1.11}$$

Where $h^{\pm} = \max\{\pm h, 0\}$ are the positive and negative part of h.

To be able to make this definition we have assumed that a was the difference between two non-decreasing functions. We now ask for an analytic characterization of those functions which have this property. If a is a measurable function and I an interval, we define (with a slight abuse of notation) $da(I) = a(\sup I) - a(\inf I)$ even though da is not really a measure.

Lemma (1.2.1):

Let $a:[0,\infty) \to \mathbb{R}$ be cadlag and define $\upsilon^n(0) = 0$, and for all $t > 0$

$$\upsilon^n(t) = \sum_{k=0}^{\lceil 2^n t \rceil - 1} | a((k+1)2^{-n}) - a(k2^{-n}) |. \tag{1.12}$$

Then $\upsilon(t) = \lim_{n\to\infty} \upsilon^n(t)$ exists for all $t \geq 0$ and is non-decreasing in t.

Proof: Let $t_n^+ = 2^{-n}\lceil 2^n t \rceil$ and $t_n^- = 2^{-n}\left(\lceil 2^n t \rceil - 1\right)$ and write

$$\upsilon^n(t) = \sum_{\substack{I \in \mathbb{D}_n \\ \inf I < t}} | da(I) | = \sum_{\substack{I \in \mathbb{D}_n \\ \sup I < t}} | da(I) | + | a(t_n^+) - a(t_n^-) |. \tag{1.13}$$

where $\Delta_n = \{(k2^{-n},(k+1)2^{-n}] : k \in \mathbb{Z}\}$. The first term is non-decreasing in n by the triangle inequality, and so has a limit as $n \to \infty$. The second converges to $|\Delta a(t)| = |a(t) - a(t-)|$ as a is cadlag, and so $\upsilon(t)$ exists for all $t \geq 0$.

Since $\upsilon^n(t)$ is non-decreasing in t for all n, the same holds for $\upsilon(t)$.

Definition (1.2.2):

$\upsilon(t)$ is called the total variation of a over $(0,t]$ and a is said to be of finite variation if $\upsilon(t) < \infty$ for all $t \geq 0$.

Proposition (1.2.3):

A cadlag function $a:[0,\infty) \to \mathbb{R}$ can be expressed as $a = a' - a''$, with a', a'' increasing and cadlag, iff a is of finite variation. In this case, $t \mapsto \upsilon(t)$ is a a cadlag with $\Delta\upsilon(t) = |\Delta a(t)|$ and $a^{\pm} = \frac{1}{2}(\upsilon \pm a)$ are the smallest functions a' and a'' with that property.

Proof: Suppose $\upsilon(t) < \infty$ for all $t \geq 0$.

Direction 1. Assume that $a = a' - a''$ for two cadlag non-decreasing functions a', a'', and let us show that $\upsilon(t) < \infty$. This is the easy direction: if $I \in \Delta_n$, then we have by the triangle inequality (since a' and a'' are non-decreasing).

$$|da(I) \leq da'(I) + da''(I),$$

So summing over all intervals $I \in \Delta_n$ with $\inf I < t$, by noting that the sums on the right-hand side telescope:

$$\upsilon^n(t) \leq a'(t_n^+) + a''(t_n^+)$$

Since a' and a'' are cadlag, the right-hand side converges to $a'(t)+a''(t)$ as $n \to \infty$ and is in particular bounded. Thus $\upsilon(t) < \infty$ for all $t \geq 0$.

Converse: Assume that $\upsilon(t) < \infty$ for all $t \geq 0$. Let us first show that υ is cadlag. Fix $T > 0$ and consider

$$u^n(t) = \sum_{\substack{I \in \Delta_n \\ t < \inf I < \sup I < T}} |da(I)| \text{ for all } \quad t \leq T. \tag{1.14}$$

$u^n(t)$ is clearly non-increasing in t, and it is easy to see that it is also right-continuous. These two properties imply that $\{t \in [0,T] : u^n(t) \leq x\}$ is closed for all $x \geq 0$. Now, just as for $\upsilon^n(t)$, the sum defining $u^n(t)$ is non-decreasing in n by the triangular inequality. Thus for all $t \geq 0, u^n(t)$, has a limit as $n \to \infty$ which we may call $u(t)$. We have that

$$\{t \in [0,T] : u(t) \leq x\} = \bigcap_{n \in \mathbb{N}} \{t \in [0,T] \, u^n(t) \leq x\} \tag{1.15}$$

is closed as a countable intersection of closed sets. This, together with the fact that $u(t)$ is non-increasing in *t*, implies that u is right-continuous. Furthermore, observe that for all $t < T$:

$$\upsilon^n(T) = \upsilon^n(t) + u^n(t) + |a(T_n^+) - a(T_n^-)|, \tag{1.16}$$

The final term on the right converges to $|\Delta a(T)|$ as $n \to \infty$ because a is right-continuous. Hence for all $t < T$ we have $\upsilon(T) = \upsilon(T) - u(T) + |\Delta a(T)|$ and since T was arbitrary, υ is right-continuous. υ has left limits since it is non-decreasing, and taking the limit $n \to \infty$ in (1.13) we get $\upsilon^n(T) = \upsilon(t-) + |\Delta a(t)|$.

Step2: $\upsilon = a^+ - a^-$. Having made these observations, define two functions a^+ and a^- by

$$a^+ = \frac{1}{2}(\upsilon + a) \text{ and } a^- = \frac{1}{2}(\upsilon - a) \text{ are cadlag.} \tag{1.17}$$

Since υ is cadlag, then a^+ and a^- are also cadlag. It thus suffices to prove that they are non-decreasing. However, note that for each $m \in \mathbb{N}^+$,

$$d\upsilon^m(I) = |da(I)| \text{ for all } I \in \Delta_m$$

$$\text{and } d\upsilon^n(I) \geq |da(I) \text{ for all } I \in \Delta_m \text{ if } n \geq m \qquad (1.18)$$

Thus $da^{\pm}(I) = \frac{1}{2} d\upsilon(I) \pm \frac{1}{2} da(I) \geq 0$ for all $I \in \bigcup_{m \geq 1} \Delta_m$ (by right-continuity) that a^+ and a^- are non-decreasing.

***Step* 3**: minimality. Suppose now $a = a' - a''$ where a', a'' are non-decreasing with $a(0) = a'(0) = a''(0) = 0$ without loss of generality. Then for any $I \in \Delta_n, n \geq 0$,

$$|da(I)| \leq da'(I) + da''(I). \qquad (1.19)$$

Summing over $I \in \Delta_n$ with $\sup I < t$ in Equation (1.13), the terms in the sum telescope and we obtain

$$\upsilon^n(t) \leq a'(t_n^+) + a''(t_n^+). \qquad (1.20)$$

Letting $n \to \infty$, the left-hand side converges to $\upsilon(t)$ by definition, and the right-hand side converges to $a'(t) + a''(t)$ since a' and a'' are right-continuous. Note that we can also write $\upsilon(t) = a^+(t) + a^-(t)$ and hence the last inequality shows

$$a^+(t) + a^-(t) \leq a'(t) + a''(t)$$

for all $t \geq 0$, Adding and substracting $a = a^+ - a^- = a' - a''$ on both sides we get $a^+(t) \leq a'(t)$ and $a^-(t) \leq a''(t)$ for all $t \geq 0$, as required.

Suppose now that we have a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Recall that a process $X : \Omega \times [0,\infty) \to \mathbb{R}$ is adapted to $(\mathcal{F}_t)_{t\geq 0}$ if $X_t = X(.,t)$ is $\mathcal{F}_t$-measurable for all $t \geq 0$, and X is cadlag if $X(\omega,.)$ is cadlag for all $\omega \in \Omega$

Definition (1.2.4):

Let A be a cadlag adapted process. Its total variation process V is defined path-wise (for each $\omega \in \Omega$) as the total variation of $A(\omega,.)$. We say that A is of finite variation if $A(\omega,.)$ is of finite variation for all $\omega \in \Omega$.

Lemma (1.2.5):

Let A be a cadlag adapted process with finite total variation V. Then V is cadlag adapted and path-wise non-decreasing.

Proof: Using the same partition as in Equation (1.13) we get

$$V_t = \lim_{n\to\infty} \tilde{V}_t^n + |\Delta A_t| \tag{1.21}$$

Where $\tilde{V}_t^n = \sum_{k=0}^{2^n t_n^- - 1} |A_{(k+1)2^{-n}} - A_{k2^{-n}}|$ is adapted for all $n \in \mathbb{N}$ since $t_n^- \leq t$ and ΔAt is $\mathcal{F}_t$-measurable since A is cadlag adapted. Thus V is adapted and it is cadlag and increasing because $V(\omega,.)$ is cadlag and increasing for all $\omega \in \Omega$.

In the later we will introduce a suitable class of integrands H for a path-wise definition of the stochastic integral

$$(H.A)(\omega,t) = \int_{(0,t]} H(\omega,s)dA(\omega,s). \tag{1.22}$$

We now started to study previsid processes.

Definition (1.2.6):

The previsible σ-algebra $\mathcal{P}$ is the σ-algebra generated by sets of the form $E\times(s,t]$ where $E\in\mathcal{F}_s$ and s < t. A previsible process H is a $\mathcal{P}$-measurable map $H:\Omega\times(0,\infty)\to\mathbb{R}$.

Proposition (1.2.7):

Let X be cadlag adapted and $H_t = X_{t-}, t>0$. Then H is previsible.

Proof: $H:\Omega\times(0,\infty)\to\mathbb{R}$. is left-continuous and adapted.

Set $t_n^- = k2^{-n}$ when $k2^{-n} < t \le (k+1)2^{-n}$ and

$$H_t^n = H_{t_n^-} = \sum_{k=0}^{\infty} H_{k2^{-n}} 1_{\{(k2^{-n},(k+1)2^{-n}]\}}(t) \tag{1.23}$$

So H^n is previsible for all $n\in\mathbb{N}$ since $H_{t_n^-}$ is $F_{t_n^-}$-measurable as H is adapted and $t_n^- < t$. But $t_n^- \uparrow t$ and so $H_t^n \to H_t$ as $n\to\infty$ by left-continuity and H is also n previsible.

Remark (1.2.8):

$\mathcal{P}$ is the smallest σ-algebra such that all adapted left-continuous processes are measurable.

Examples(1.2.9):

(i) Brownian motion is previsible by Proposition (1.2.7), since it is continuous.

(ii) A Poisson process $(N_t)_{t\ge 0}$ or, indeed, any other continuous-time Markov chain with discrete state space is not previsible, since N_t is not F_{t-}-measurable.

Proposition (1.2.10):

Let A be a cadlag adapted finite variation process with total variation V. Let H be previsible such that for all $t \geq 0$ and all $\omega \in \Omega$

$$\int_{(0,t]} |H(\omega,s)|\, dV(\omega,s) < \infty \tag{1.24}$$

Then the process defined path-wise by

$$(H.A)_t = \int_{(0,t]} H_s dA_s \tag{1.25}$$

is well-defined, cadlag, adapted and of finite variation.

Proof: First note that in Equation (1.25)is well-defined for all t due to the finiteness of the Equation (1.24). (More precisely, Equation (1.24) implies that all four terms defining Equation (1.25) in Equation (1.11) are finite). By referring to Equation (1.11) we may assume without loss of generality in the rest of the proof that H is non-negative and Anon-decreasing.

We now show that $(H.A)$is cadlag for each fixed $\omega \in \Omega$. We have $1_{\{(0,s)\}} \to 1_{\{(0,t)\}}$ as $s \uparrow t$, $1_{\{(0,s)\}} \to 1_{\{(0,t)\}}$ as $s \downarrow t$, and

$$(H.A)_t = \int_{(0,\infty)} H_s 1_{\{(0,t)\}}(s) dA_s \tag{1.26}$$

Hence, by dominated convergence, the following limits exist

$$(H.A)_{t+} = (H.A)_t \text{ and } (H.A)_{t-} = \int_{(0,\infty)} H_s 1_{\{(0,t)\}}(s) dA_s \tag{1.27}$$

and $H.A$ is cadlag with

$$\Delta(H.A)_t = \int_{(0,\infty)} H_s 1_{\{(t)\}}(s) dA_s = H_t \Delta A_t.$$

Next, we show that $H.A$ is adapted via a monotone class argument. Suppose first $H = 1_{\{B\times(s,u]\}}$ where $B \in \mathcal{F}_s$. Then $(H.A)_t = 1_{\{B\}}(A_{t\wedge u} - A_{t\wedge s})$ which is clearly $\mathcal{F}_t$ -measurable. Now let

$$\prod = \{B\times(s,u] : B \in \mathcal{F}_s, s < u\} \text{ and} \tag{1.28}$$

$$\mathcal{A} = \{C \in \mathcal{P} : (1_{\{C\}}.A)_t \text{ is } \mathcal{F}_s\text{-measurable}\} \tag{1.29}$$

so that $\prod$ is a π-system and $\prod \subseteq \mathcal{A}$. But $\mathcal{A} \subseteq \mathcal{P} = \sigma(\prod)$ and $\mathcal{A}$ is a λ-system.

[Recall: A a π-system contains $\emptyset$ and is stable by intersection. A λ-system (or d-system) is stable by taking the difference and countable unions. To see that $\mathcal{A}$ is λ-system. Not that if $C \subseteq D \in \mathcal{A}$ then $((1_{\{D\}} - 1_{\{C\}}).A)_t$ is $\mathcal{F}_t$ -measurable, which gives $D\setminus C \in \mathcal{A}$; and if $C_n \in \mathcal{A}$ with $C_n \uparrow C$ then $C \in \mathcal{A}$ since a limit of measurable functions is measurable.]

Hence, by Dynkin's lemma, $\sigma(\prod) \subseteq \mathcal{A}$. But by definition, $\sigma(\prod) = \mathcal{P}$ and $\mathcal{A} \subseteq \mathcal{P}$.Thus $\mathcal{A} = \mathcal{P}$.. Suppose now that H is non-negative, $\mathcal{P}$.-measurable. For all $n \in \mathbb{N}$ set

$$H^n = 2^{-n}\lfloor 2^n H \rfloor = \sum_{k=1}^{\infty} 2^{-n} k 1_{\underbrace{\{\{H\in[2^{-n}k, 2^{-n}(k+1))\}}_{\in \mathcal{P}}} \tag{1.30}$$

so that $(H^n.A)_t$ is $\mathcal{F}_t$ -measurable. We have $(H^n.A)_t \uparrow (H.A)_t$ by monotone convergence (applied for each ω). Hence, $(H.A)_t$ is $\mathcal{F}_t$ -measurable.

Example (1.2.11):

Suppose that H is a previsible process, such as Brownian motion, and that

$$\int_{(0,t]} |H_s|\, ds < \infty \text{ for all } \omega \in \Omega \text{ and } t \geq 0 \tag{1.31}$$

Then $\int_{(0,t]} H_s ds$ is cadlag, adapted and of finite variation.

Now we discuss local martingales.

We work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ where $(\mathcal{F}_t)_{t\geq 0}$ satisfies what is technically known as the usual conditions, i.e. $\mathcal{F}$ is $\mathbb{P}$-complete (equivalently, $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets), and $(\mathcal{F}_t)_{t\geq 0}$ is right-continuous in the sense that

$$\mathcal{F}_t = \mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s \text{ for all } t \geq 0 \tag{1.32}$$

Note for instance that the filtration generated by Brownian motion completed by zero-probability events satisfies the usual conditions (this is essentially a consequence of the simple Markov property and Blumenthal's zero-one law).

Recall that an adapted process X is a martingale if it is integrable $(\mathbb{E}(|X_t|) < \infty$ for all $t \geq 0)$ and if

$$\mathbb{E}(X_t \mid \mathcal{F}_s) = X_s \text{ a.s for all } s \leq t \tag{1.33}$$

We write $\mathcal{M}$ for the set of all cadlag martingales. The following result is fundamental and will be used repeatedly in this research.

Theorem (1.1.12):(Optional Stopping Theorem)(OST)

Let X be a cadlag adapted integrable process. Then the following are equivalent:

(i) X is a martingale

(ii) $X^T = (X_{t\wedge T}, t \geq 0)$ is a martingale for all bounded stopping times T .

(iii) For all bounded stopping times S, T, $\mathbb{E}(X_T \mid \mathcal{F}_S) = X_{S\wedge T}$ a.s .

(iv) $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ for all bounded stopping times T .

Proof: It is well known that (i) $\Rightarrow \ldots \Rightarrow$ (iv). We show how (iv) implies (i). Let $s < t$ and fix $u > t$. Let $A \in \mathrm{F}_S$,, and define a random time T by saying $T = t$ if A occurs, or $T = u$ otherwise. Similarly, define $S = s$ otherwise. Note that both S and T are stopping times, and are bounded. Thus by (iv):

$$\mathrm{E}(X_T) = \mathrm{E}(X_0) = \mathrm{E}(X_S) \qquad (1.34)$$

On the other hand,

$$\mathrm{E}(X_T) = \mathrm{E}(X_t 1_{\{A\}}) + \mathrm{E}(X_u 1_{\{A^c\}})$$

and similarly:

$$\mathrm{E}(X_S) = \mathrm{E}(X_s 1_{\{A\}}) + \mathrm{E}(X_u 1_{\{A^c\}})$$

Plugging this intoEquation (1.34) and cancelling the terms $\mathrm{E}(X_u 1_{\{A^c\}})$, we find:

$$\mathrm{E}(X_t 1_{\{A\}}) = \mathrm{E}(X_s 1_{\{A\}})$$

for all $s < t$ and all $A \in \mathrm{F}_S$,. This means (by definition) that

$$\mathrm{E}(X_t \mid \mathrm{F}_s) = X_s, \text{ a.s.}$$

as required. Hence, since X is adapted and integrable, X is a martingale.

It is also the case that M is stable under stopping. This observation leads us to define a slightly more general class of processes, called local martingales.

Definition (1.2.13):

A cadlag adapted process X is a local martingale, $X \in \mathrm{M}_{loc}$, if there exists a sequence $(T_n)_{n\in\mathbb{N}}$ of stopping times with $T_n \uparrow \infty$ such that $(X_t^{T_n})_{t\geq 0} \in \mathrm{M}$ for all $n \in N$. We say that the sequence $(T_n)_{n\in\mathbb{N}}$ reduces X.

In particular $\mathrm{M} \subseteq \mathrm{M}_{loc}$, since any sequence $(T_n)_{n\in\mathbb{N}}$ of stopping times reduces X by OST(ii). Recall that a family $\mathrm{X} \equiv (X_i)_{i\in I}$ of random variables is called uniformly integrable (UI) if

$$\sup_{i\in I} \mathrm{E}(|X_i| 1_{\{|X_i|\geq \lambda\}}) \to 0 \quad \text{as } \lambda \to \infty \tag{1.35}$$

We now give necessary and sufficient conditions for a local martingale to be a martingale.

Proposition (1.2.14):

The following statements are equivalent:

(i) X is a martingale

(ii) X is a local martingale and for all $t \geq 0$ the set

$\mathrm{X}_t \equiv \{X_T : T$ is a stopping time, $T \leq t\}$ is UI . (1.36)

Proof: Suppose (i) holds. By the Optional Stopping Theorem, if T is a stopping time with $T \leq t$, then $X_T = \mathrm{E}(X_t \mid \mathrm{F}_T)$ a.s.. Thus by definition X_t is uniformly integrable.

If (ii) holds, suppose $(T_n)_{n\geq 0}$ reduces X. Let T be any bounded stopping time, $T \leq t$ say. By the Optional Stopping Theorem applied to the martingale X^{T_n},

$$\mathrm{E}(X_0) = \mathrm{E}(X_0^{T_n}) = \mathrm{E}(X_T^{T_n}) = \mathrm{E}(X_{T\wedge T_n}) \tag{1.37}$$

Since $\{X_{T\wedge T_n} : n \in \mathbb{N}\}$ is uniformly integrable by assumption, $\mathrm{E}(X_{T\wedge T_n}) \to \mathrm{E}(X_T)$ as $n \to \infty$. Therefore,

$$\mathrm{E}(X_T) = \mathrm{E}(X_0).$$

But then by the Optional Stopping Theorem again, X must be a martingale.

An extremely useful consequence of the above is the following:

Corollary (1.2.15):

Let M be a local martingale, and assume that M is bounded. Then M is a true martingale. More generally, if M is a local martingale such that for all $t \geq 0, |M_t| \leq Z$, for some $Z \in L^1$, then M is a true martingale.

Remark (1.2.16):

Occasionally, we will need the following stronger version of (iii) in the Optional stopping theorem: if X is a uniformly integrable martingale, then for any stopping times S,T

$$\mathrm{E}(X_T \mid \mathrm{F}_S) = X_{S \wedge T} \tag{1.38}$$

almost surely.

Proposition (1.2.17):

A nonnegative local martingale M is a super-martingale.

Proof: This follows simply from the definition of local martingales and Fatou's lemma for conditional expectations.

Remark (1.2.18):

A martingale can be interpreted as the fortune of a player in a fair game. A local martingale which is not a true martingale, on the other hand, is the fortune of a player in a game which looks locally fair: unfortunately, this is only because there are going to be times of huge increases of X followed by an eventual ruin. Overall, as the above proposition shows, the expected fortune decreases. A local martingale is thus something akin to a bubble in the market. (Thanks are due to M . Tehranchi for this analogy).

Proposition (1.2.19):

Let M be a continuous local martingale ($M \in \mathcal{M}_{c,loc}$) starting from 0. Set $S_n = \inf\{t \geq 0 : |M_t| = n\}$.. Then $(S_n)_{n\geq 0}$ reduces M .

Proof: Note that $\{S_n \leq t\} = \bigcap_{k\in\mathbb{N}} \bigcup_{\substack{s\in\mathbb{Q} \\ s\leq t}} \{|M_s| > n - 1/k\} \in \mathcal{F}_t$,

and so S_n is a stopping time. For each $\omega \in \Omega$, $S_n(\omega), n \geq 0)$ must be non-decreasing by the mean-value theorem since M is continuous, and $\lim_{n\to\infty} S_n$ can only be infinite by continuity as well. Hence $S_n \uparrow \infty$ a.s.. Let $(T_k)_{k\in\mathbb{N}}$ be a reducing sequence for M , i.e. $M^{T_k} \in \mathcal{M}$. By OST, also $M^{S_n \wedge T_k}\mathcal{M} \in \mathcal{M}$ and so $M^{S_n} \in \mathcal{M}_{loc}$ for each $n \in \mathbb{N}$. But M^{S_n} Sn is bounded and so also a martingale.

Theorem (1.2.20):

Let M be a continuous local martingale which is also of finite variation, and such that $M_0 = 0$ a.s. Then M is indistinguishable from 0.

Remarks.

(i) In particular Brownian motion is not of finite variation.

(ii) This makes it clear that the theory of finite variation integrals we have developed is useless for integrating with respect to continuous local martingales.

(iii) It is essential to assume that M is continuous in this theorem.

Proof. Let V denote the total variation process of M . Then V is continuous and adapted with $V_0 = 0$.. Set $S_n = \inf\{t \geq 0 : V_t = n\}$.. Then S_n is a stopping time for all $n \in \mathbb{N}$ since V is adapted, and $S_n \uparrow \infty$ as $n \to \infty$ since V_t is non-decreasing and finite for all $t \geq 0$.. It suffices to show $M^{S_n} \equiv 0$ for all $n \in \mathbb{N}$. By OST, $M^{S_n} \in \mathcal{M}_{loc}$. Also

$$|M_t^{S_n}| \leq |V_t^{S_n}| \leq n, \tag{1.39}$$

and so, by Proposition (1.2.12), $M^{S_n} \in \mathcal{M}$.

Replacing M by M^{S_n} we can reduce to the case where M is a bounded martingale of bounded variation, i.e. V is bounded.

Lemma (1.2.21):

Let M be a martingale and such that for some given $s < t$, $\mathrm{E}(M_s^2) < \infty$. and $\mathrm{E}(M_t^2) < \infty$.then

$$\mathrm{E}(M_t^2 - M_s^2 \mid \mathcal{F}_s) = \mathrm{E}((M_t - M_s)^2 \mid \mathcal{F}_s), \text{ a.s} \tag{1.40}$$

(This trick will be used over and over again in what follows, so it is a good point to memorize it).

Proof: By expanding the square $(M_t - M_s)^2$, the right-hand side is equal to

$$\mathrm{E}((M_t - M_s)^2 \mid \mathcal{F}_s) = \mathrm{E}(M_t^2 \mid \mathcal{F}_s) - 2M_s \mathrm{E}(M_t \mid \mathcal{F}_s) + M_s^2$$

$$= \mathrm{E}(M_t^2 \mid \mathcal{F}_s) - 2M_s^2 + M_s^2$$

$$= \mathrm{E}(M_t^2 - M_s^2 \mid \mathcal{F}_s)$$

as required.

Coming back to the proof of the theorem, fix $t > 0$ and set $t_k = kt / N$ for $0 \le k \le N$. By Equation (1.40),

$$\mathrm{E}(M_t^2) = \mathrm{E}\left(\sum_{k=0}^{N-1} (M_{t_{k+1}}^2 - M_{t_k}^2)\right) = \left(\sum_{k=0}^{N-1} (M_{t_{k+1}} - M_{t_k})^2\right)$$

$$\le \mathrm{E}(\underbrace{\sup_{k<N} | M_{t_{k+1}} - M_{t_k} |}_{\le V_t \le n} \underbrace{\sum_{k=0}^{N-1} |M_{t_{k+1}} - M_{t_k}}_{\le V_t \le n}) \tag{1.41}$$

As M is bounded and continuous,

$$\sup_{k<N} | M_{t_{k+1}} - M_{t_k} | \to 0 \text{ as } N \to \infty \tag{1.42}$$

and so, by bounded convergence,

$$\mathrm{E}\left(\sup_{k<N} | M_{t_{k+1}} - M_{t_k} | \sum_{k=0}^{N-1} | M_{t_{k+1}} - M_{t_k} |\right) \to 0 \text{ as } N \to \infty \tag{1.43}$$

Hence, $\mathrm{E}(M_t^2) = 0$ for all $t \geq 0$. Since M is continuous, M is indistinguishable from 0.

Definition (1.2.22):

A continuous semimartingale X is an adapted continuous process which may be written as

$$X = X_0 + M + A \text{ with } M_0 = A_0 = 0 \tag{1.44}$$

where $M \in \mathrm{M}_{c,loc}$ and A is a finite variation continuous process.

Note that as a consequence of Theorem (1.1.18), the decomposition is unique up to indistinguishability. This is known as the Doob-Meyer decomposition.

Remark (1.2.23):

The proof of the last theorem tells us something extremely useful for the following. If t_k is the dyadic subdivision, then the calculation shows that

$$\mathrm{E}(M_t^2) = \mathrm{E}\left(\sum_{k:t_k \leq t} (M_{t_{k+1}} - M_{t_k})^2\right) \tag{1.45}$$

so there is good reason to believe that if M is say, bounded in L^2 , then it has finite quadratic variation Q_t and moreover

$$M_t^2 - Q_t$$

has constant expectation 0. In fact, we will see that this is indeed the case and $M_t^2 - Q_t$ is also a martingale.

Chapter 2
The stochastic integral

In this chapter we establish the stochastic integral with respect to continuous semi-martingales. In places, we develop parts of the theory also for càdlàg semi-martingales, where this involves no extra work. However, parts of the construction will use crucially the assumption of continuity. A more general theory exists, but it is beyond the scope of this research.

Recall that we say a process X is *bounded in* L^2 if

$$\sup_{t\geq 0} \| X_t \|_2 < \infty \tag{2.1}$$

where here and in the rest of the course, for a random variable X:

$$\| X \|_2 = \mathrm{E}(| X |^2)^{1/2} \tag{2.2}$$

Write $\mathcal{M}^2$ for the set of all càdlàg L^2-bounded martingales, and $\mathcal{M}_c^2$ for the set of continuous martingales bounded in L^2. Recall the following two fundamental results from Advanced probability:

Section (2.1): Integral on L^2 and Quadratic Variation

We start this section by simple integrands and L^2 properties.

Definition (2.1.1):

A *simple process* is any map $H : \Omega \times (0,\infty) \to \mathbb{R}$ of the form

$$H(\omega,t) = \sum_{k=0}^{n-1} Z_k(\omega) 1_{\{(t_k, t_k+1]\}}(t) \tag{2.3}$$

where $n \in \mathbb{N}, 0 = t_0 < ... < t_n < \infty$ and Z_k is a *bounded* $\mathcal{F}_{t_k}$-measurable random variable for all k. We denote the set of simple processes by $\mathcal{S}$. Given $H \in \mathcal{S}$ we denote $\| H \|_\infty = es\sup | H |$ the essential supremum of H, i.e., the smallest $M > 0$ such that $\sup_{t\geq 0} | H(t,\omega) | \leq M$ almost surely.

Note that $\mathcal{S}$ is a vector space and that (by definition) every simple process is previsible. We now define the stochastic integral for simple processes.

Definition (2.1.2):

For $H = \sum_{k=0}^{n-1} Z_k 1_{\{(t_k, t_{k+1}]\}} \in \mathrm{S}$ and $M \in \mathrm{M}^2$ set

$$(H.M)_t = \sum_{k=0}^{n-1} Z_k (M_{t_{k+1}\wedge t} - M_{t_k \wedge t}) \tag{2.4}$$

Proposition (2.1.3):

Let $H \in \mathrm{S}$ *and* $M \in \mathrm{M}^2$. *Let T be a stopping time. Then*

(i) $H.M^T = (H.M)^T$

(ii) $H.M \in \mathrm{M}^2$

(iii) $\mathrm{E}((H.M)_\infty^2 \leq \| H \|_\infty^2 \mathrm{E}((M_\infty - M_0)^2)$

Proof:

(i) For all $t \geq 0$ we have

$$(H.M^T)_t = \sum_{k=0}^{n-1} Z_k (M_{t_{k+1}\wedge t}^T - M_{t_k \wedge t}^T)$$

$$= \sum_{k=0}^{n-1} Z_k (M_{t_{k+1}\wedge t \wedge T} - M_{t_k \wedge t \wedge T}) = (H.M)_{t\wedge T} = (H.M)_t^T \tag{2.5}$$

(ii) For $t_k \leq \mathrm{S} \leq t < t_{k+1}, (H.M)_t - (H.M)_\mathrm{S} = Z_k (M_t - M_\mathrm{S})$, so that

$$\mathbf{E}((H.M)_t - (H.M)_\mathrm{S} \mid \mathrm{F}_\mathrm{S}) = Z_k \mathrm{E}(M_t - M_\mathrm{S} \mid \mathrm{F}_\mathrm{S}) = 0 \tag{2.6}$$

This extends easily to general $\mathrm{S} \leq t$ and hence $H.M$ is a martingale. To show it is bounded in L^2, note that if $j<k$ we have the following "orthogonality relation":

$$\mathrm{E}(Z_j (M_{t_{j+1}} - M_{t_j}) Z_k (M_{t_{k+1}} - M_{t_k})) =$$

$$\mathbf{E}(Z_j (M_{t_{j+1}} - M_{t_j}) Z_k \mathrm{E}(M_{t_{k+1}} - M_{t_k} \mid \mathrm{F}_{t_k})) = 0 \tag{2.7}$$

Thus let $t \geq 0$ and assume that $t_n \leq t$ for simplicity. To compute $\mathbf{E}((H.M)_t^2)$, we expand the square and use the above orthogonality relation:

$$\mathrm{E}((H.M)_t^2 = \mathrm{E}\left(\left(\sum_{k=0}^{n-1} Z_k (M_{t_{k+1}} - M_{t_k})\right)^2\right) = \sum_{k=0}^{n-1} \mathrm{E}(Z_k^2 (M_{t_{k+1}} - M_{t_k})^2)$$

$$\leq \| H \|_\infty^2 \sum_{k=0}^{n-1} \mathbf{E}((M_{t_{k+1}} - M_{t_k})^2) = \| H \|_\infty^2 \mathbf{E}((M_{t_n} - M_0)^2) \qquad (2.8)$$

(On two occasions, we used the trick (1.40)).

Similarly, if $t_j \leq t < t_{j+1}$, then the same calculation gives:

$$\mathrm{E}((H.M)_t^2) \leq \| H \|_\infty^2 \mathrm{E}((M_t - M_0)^2)$$

But note that since $M \in \mathcal{M}$,then $(M_t - M_0)^2$ is a sub-martingale, so if $u \geq t$ then $\mathbf{E}((M_t - M_0)^2) \leq \mathbf{E}((M_u - M_0)^2)$. Since $M \in \mathcal{M}^2$, then the convergence of M_u to M_∞ holds in L^2 as $u \to \infty$, hence we deduce $\mathbf{E}((M_t - M_0)^2) \leq \mathbf{E}((M_\infty - M_0)^2)$ by letting $u \to \infty$.

Thus for all $t \geq 0$,

$$\mathbf{E}((H.M)_t^2) \leq \| H \|_\infty^2 \mathbf{E}(M_\infty - M_0)^2)$$

Thus $H.M \in \mathcal{M}^2$ and letting $t \to \infty$ in the above inequality (which we may since we now know $H.M \in \mathcal{M}^2$) we obtain the desired (iii).

To extend the simple integral defined before, we will need some Hilbert space properties of the set of integrators we are considering. As before, we work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_2)_{t \geq 0}, \mathrm{P})$ where $(\mathcal{F}_t)_{t \geq 0}$ satisfies the usual conditions.

Definition (2.1.4):

For all càdlàg adapted processes X define the triple norm

$$||| X ||| = \| \sup_{t \geq 0} | X_t | \|_2$$

We write $\mathcal{C}^2$ for the set of all càdlàg adapted processes X such that $||| X ||| < \infty$.

On $\mathcal{M}^2$, define the norm $\| X \| = \| X_\infty \|_2$

Remark(2.1.5):

Note that the function. $\| . \|$ on $\mathcal{M}^2$ defines indeed a norm. The only point which demands justification is the requirement that if $\| M \| = 0$, then M is indistinguishable from 0. But $\| M \| = 0$, then $\mathrm{E}(M_\infty^2) = 0$ so $M_\infty = 0$ a.s. By the martingale convergence theorem

$$M_t = \mathrm{E}(M_\infty | \mathcal{F}_t) \text{ a.s.}$$

so $M_t = 0$ a.s. as well. Since M is càdlàg, it is indistinguishable from 0.

We may now state some L^2 properties which show that the space of square-integrable martingales can be seen as a Hilbert space. As we will see later, this underlying Hubert structure is the basis of the formal definition of the stochastic integral in the general case. (Formally, it is defined as an isometry between Hilbert spaces).

Theorem (2.1.6):

Let $X \in \mathcal{M}^2$. There exists $X_\infty \in L^2$ such that

$$X_t \to X_\infty \; a.s. \text{ and in } L^2; \text{ as } t \to \infty \tag{2.9}$$

Moreover, $X_t = \mathrm{E}(X_\infty | \mathcal{F}_t)$ a.s. for all $t \geq 0$

The second result which we will need is Doob's L^2 inequality:

Theorem (2.1.7):

For $X \in \mathcal{M}^2$,

$$\mathbf{E}(\sup_{t \geq 0} | X_t |^2) \leq 4\mathrm{E}(X_\infty^2) \tag{2.10}$$

Similar to the construction to the Lebesgue integral in measure theory.

Proposition (2.1.8):

We have

(i) $(\mathcal{C}^2, \||\cdot\||)$ is complete

(ii) $\mathcal{M}^2 = \mathcal{M} \cap \mathcal{C}^2$

(iii) $(\mathcal{M}^2, \| \cdot \|)$ is a Hubert space with $\mathcal{M}_C^2 = \mathcal{M}_C \cap \mathcal{M}^2$ as a closed subspace

(iv) $X \mapsto X_\infty : \mathcal{M}^2 \to L^2(\mathcal{F}_\infty)$ is an isometry

Proof:

(i) Suppose $(X^n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $(\mathrm{C}^2, |||\,.\,|||)$. Then we can find a subsequence $(n_k)_{k\in\mathbb{N}}$ such that $\sum |||X^{n_{k+1}} - X^{n_k}||| < \infty$. Then by the triangular inequality,

$$\left\| \sum_{k=1}^{\infty} \sup_{t\ge 0} |X_t^{n_{k+1}} - X_t^{n_k}| \right\| \le \sum_{k=1}^{\infty} |||X^{n_{k+1}} - X^{n_k}||| < \infty \qquad (2.11)$$

and so for almost every $\omega \in \Omega$,

$$\sum_{k=1}^{\infty} \sup_{t\ge 0} |X_t^{n_{k+1}}(\omega) - X_t^{n_k}(\omega)| < \infty$$

Since the space of càdlàg functions equipped with the $\|\cdot\|_\infty$ norm is complete, there exists a càdlàg process X such that $(X_t^{n_k}(\omega))_{k\in\mathbb{N}} \to X(\omega)$ as $k \to \infty$ uniformly in $t \ge 0$. Now

$$|||X^n - X|||^2 = \mathrm{E}(\sup_{t\ge 0} |X_t^n - X_t|^2) \qquad (2.12)$$

$$\le \liminf_{k\to\infty} \mathrm{E}(\sup_{t\ge 0} |X_t^n - X_t^{n_k}|^2) \text{ by Fatou's lemma}$$

$$= \liminf_{k\to\infty} |||X^n - X^{n_k}|||^2 \to 0 \text{ as } n \to \infty \qquad (2.13)$$

because $(X^n)_{n\in\mathbb{N}}$ is a Cauchy sequence. Hence $(\mathrm{C}^2, |||\,.\,|||)$ is complete.

(ii) For $X \in \mathrm{C}^2 \cap \mathrm{M}$ we have

$$\sup_{t\ge 0} \|X_t\|_2 \le \| \sup_{t\ge 0} |X_t| \|_2 = |||X||| < \infty \qquad (2.14)$$

and so $X \in \mathrm{M}^2$. On the other hand, if $X \in \mathrm{M}^2$, by Doob's inequality,

$$|||X||| \le 2\|X\| < \infty, \text{ and so } X \in \mathrm{C}^2 \cap \mathrm{M} \qquad (2.15)$$

(iii) $(X, Y) \mapsto \mathrm{E}(X_\infty Y_\infty)$ defines an inner product on M^2 whose associated norm is precisely the double norm $\|\cdot\|$. Moreover, for $X \in \mathrm{M}^2$, we have shown in (ii) that

$$\|X\| \le |||X||| \le 2\|X\|, \qquad (2.16)$$

that is, $\|\cdot\|$ and $|||\cdot|||$ are equivalent on $\mathcal{M}^2$. Thus $\mathcal{M}^2$ is complete for $\|\cdot\|$ if and only if it is complete for $|||\cdot|||$, and by (i) it is thus sufficient to show that $\mathcal{M}^2$ is closed in $(\mathcal{C}^2, |||.|||)$. If $X^n \in \mathcal{M}^2$ and $|||X^n - X||| \to 0$ as $n \to \infty$ for some X, then X is certainly càdlàg adapted and L^2-bounded. Furthermore, by Jensen's inequality for conditional expectations,

$$\|\mathrm{E}(X_t|\mathcal{F}_s) - X_s\|_2 \le \|\mathrm{E}(X_t - X_t^n|\mathcal{F}_s)\|_2 + \|X_s^n - X_s\|_2$$

$$\le \|X_t - X_t^n\|_2 - \|X_s^n - X_s\|_2 \tag{2.17}$$

$$\le 2|||X^n - X||| \to 0 \tag{2.18}$$

as $n \to \infty$ and so $X \in \mathcal{M}^2$. By the same argument $\mathcal{M}_C^2$ is closed in $(\mathcal{M}^2, \|\cdot\|)$ where continuity of $t \mapsto X_t(\omega)$ follows by uniform convergence in t.

(iv) For $X, Y \in \mathcal{M}^2, \|X - Y\| = \|X_\infty - Y_\infty\|_2$ by definition.

Now we study Quadratic variation.

Definition (2.1.9):

For a sequence $(X^n)_{n\in\mathbb{N}}$ we say that $X^n \to X$ uniformly on compacts in probability $(u.c.p.)$ if

$$\forall \varepsilon > 0\ \forall t \ge 0 : \mathrm{P}(\sup_{s\le t} |X_s^n - X_s| > \epsilon) \to 0 \text{ as } n \to \infty. \tag{2.19}$$

Theorem (2.1.10): (Quadratic Variation)

For each $M \in \mathcal{M}_{C,loc}$ there exists a unique (up to indistinguishability) continuous adapted non-decreasing process $[M]$ such that $M^2 - [M] \in \mathcal{M}_{c,loc}$ and such that $[M]_0 = 0$ a.s. Moreover, for

$$[M]_t^n = \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} (M_{(k+1)2^{-n}} - M_{k2^{-n}})^n \tag{2.20}$$

we have

$$[M]^n \to [M]\ u.c.p. \text{ as } n \to \infty.$$

We call $[M]$ the quadratic variation process of M.

Example (2.1.11):

Let B be a standard Brownian motion. Then we know that $[B]_t^n \to t$ for every $t \geq 0$ in probability. Thus $[B_t] = t$ and we deduce that $B_t^2 - t$ is a local martingale.

Alternatively, it is not hard to see that $(B_t^2 - t)_{t \geq 0}$ is a (true) continuous martingale. But then by Theorem (2.1.10), $t = [B]_t$.

Proof: [Proof of Theorem (2.1.10)] Wlog we will consider the case $M_0 = 0$. Uniqueness is easy: if A and A' are two increasing processes satisfying the conditions for $[M]$ then

$$A_t - A_t^{'} = (M_t^2 - A_t^2) - (M_t^2 - A_t) \in \mathrm{M}_{c,loc} \tag{2.21}$$

is of finite variation and thus $A \equiv A'$ a.s. by Theorem (1.2.18)

Existence: First we assume that M is bounded, which implies $M \in \mathrm{M}_c^2$. Fix $T > 0$ deterministic. Let

$$H_t^n = M_{2^{-n}\lfloor 2^n t \rfloor}^T = \sum_{k=0}^{\lfloor 2^n T \rfloor - 1} M_{k 2^{-n}} 1_{\{k 2^{-n}, (k+1)2^{-n}]\}}(t). \tag{2.22}$$

Then $H^n \in \mathrm{S}$ for all $n \in \mathbb{N}$. Hence X^n defined by

$$X_t^n = (H^n.M)_t = \sum_{k=0}^{\lfloor 2^n T \rfloor - 1} M_{k 2^{-n}} (M_{(k+1)2^{-n} \wedge t} - M_{k 2^{-n} \wedge t}). \tag{2.23}$$

is in M_c^2 by Proposition(2.1.3) and by continuity of M. Recall that $\| X^n \| = \| X_\infty^n \|_2 = \| X_T \|_2$ since X_t^n is constant for $t \geq T$. For $n, m \geq 0$ we have by linearity of the stochastic integral,

$$H^n.M - H^m.M = (H^n - H^m).M,$$

hence letting $H = H^n - H^m$ for ease of notations

$$\| X^n - X^m \|^2 = \mathrm{E}[(H.M)_T^2].$$

Therefore, computing this as in Proposition (2.1.3)

$$\| X^n - X^m \|^2 = \mathrm{E}\left[\left(\sum_{k=0}^{\lfloor 2^n T \rfloor - 1} H_{k 2^{-n}} (M_{(k+1)2^{-n}} - M_{k 2^{-n}}) \right)^2 \right]$$

$$= \mathrm{E}\left[\sum_{k=0}^{\lfloor 2^n T \rfloor -1} H^2_{k2^{-n}} (M_{(k+1)2^{-n}} - M_{k2^{-n}})^2\right]$$

$$\le \mathrm{E}\left(\sup_{0\le t\le T} |H_t|^2 \sum_{k=0}^{\lfloor 2^n T \rfloor -1} (M_{(k+1)2^{-n}} - M_{k2^{-n}})^2\right)$$

Recalling that $H = H^n - H^m$ where both H^n arid H^m are dyadic approximations of M , so that the first term in the above expectation tends to 0 almost surely as $n,m \to \infty$ (by uniform continuity of M on $[0,T]$), and that moreover the expectation of the second term is bounded by Equation(1.45), it is tempting to conclude that the left-hand side of the above inequality also tends to 0. This turns out to be true be requires stronger arguments than what we just sketched. In fact, we will deal with both terms separately: by the Cauchy-Schwartz inequality,

$$\| X^n - X^m \|^2 \le \mathrm{E}\left[\sup_{0\le t\le T} |H^n_t - H^m_t|^4\right]^{1/2} \mathrm{E}\left[\left(\sum_{k=0}^{\lfloor 2^n T \rfloor -1} (M_{(k+1)2^{-n}} - M_{k2^{-n}})^2\right)^2\right]^{1/2} \quad (2.24)$$

The first term tends to 0 by the above discussion and Lebesgue's convergence theorem since M is bounded, and the second term is bounded because of the following lemma.

Lemma (2.1.12):

Let $M \in \mathrm{M}$ be bounded. Suppose that $l \in$ □ and $0 = t_0 < t_1 < ... < t_l < \infty$..

Then $\mathrm{E}\left(\left(\sum_{k=0}^{l-1} (M_{t_{k+1}} - M_{t_k})^2\right)^2\right)$ is bounded.

Proof: [Proof of Lemma (2.1.12) First note that

$$\mathrm{E}\left(\left(\sum_{k=0}^{l-1} (M_{t_{k+1}} - M_{t_k})^2\right)^2\right) = \sum_{k=0}^{l-1} \mathrm{E}((M_{t_{k+1}} - M_{t_k})^4)$$

$$+2\sum_{k=0}^{l-1} \mathrm{E}\left((M_{t_{k+1}} - M_{t_k})^2 \sum_{j=k+1}^{l-1} (M_{t_{j+1}} - M_{t_j})^2\right) \quad (2.25)$$

For each fixed k we have

$$\mathrm{E}\left((M_{t_{k+1}} - M_{t_k})^2 \sum_{j=k+1}^{l-1} (M_{t_{j+1}} - M_{t_j})^2 \right) =$$

$$\mathrm{E}\left((M_{t_{k+1}} - M_{t_k})^2 \mathrm{E}\left(\sum_{j=k+1}^{l-1} (M_{t_{j+1}} - M_{t_j})^2 \middle| \mathrm{F}_{t_{k+1}} \right) \right) =$$

$$\mathrm{E}\left((M_{t_{k+1}} - M_{t_k})^2 \mathrm{E}\left(\sum_{j=k+1}^{l-1} (M^2_{t_{j+1}} - M^2_{t_j}) \middle| \mathrm{F}_{t_{k+1}} \right) \right) =$$

$$\mathrm{E}\left((M_{t_{k+1}} - M_{t_k})^2 \mathrm{E}(M^2_{t_l} - M^2_{t_{k+1}} \middle| \mathrm{F}_{t_{k+1}}) \right) =$$

$$\mathrm{E}\left((M_{t_{k+1}} - M_{t_k})^2 (M^2_{t_l} - M^2_{t_{k+1}})^2 \right) = \qquad (2.26)$$

After inserting this in Equation(2.25) we get the estimate

$$\mathrm{E}\left(\left(\sum_{k=0}^{l-1} (M_{t_{k+1}} - M_{t_k})^2 \right)^2 \right) \leq$$

$$\leq \mathrm{E}\left(\left(\sup_j \left| M_{t_{j+1}} - M_{t_j} \right|^2 + 2 \sup_j \left| M_{t_l} - M_{t_j} \right|^2 \right) \sum_{k=0}^{l-1} (M_{t_{k+1}} - M_{t_k})^2 \right) \qquad (2.27)$$

Now, M is uniformly bounded by C, say. So using the inequality $(x-y)^2 \leq 2(x^2+y^2)$, we obtain

$$\mathrm{E}\left(\left(\sum_{k=0}^{l-1} (M_{t_{k+1}} - M_{t_k})^2 \right)^2 \right) \leq 12C^2 \mathrm{E}\left(\sum_{k=0}^{l-1} (M_{t_{k+1}} - M_{t_k})^2 \right) =$$

$$12C^2 \mathrm{E}((M_{t_l} - M_{t_0})^2) \leq 48C^4 \; (2.28)$$

Returning to Equation(2.24), it follows from Lemma (2.1.12), that X^n is a Cauchy sequence in $(\mathrm{M}^2_c, \|\cdot\|)$

and so, by Proposition (2.1.8), converse to a limit $Y = (Y_t, t \geq 0) \in \mathrm{M}^2_c$. Now, for any n and $1 \leq k \leq \lfloor 2^n T \rfloor$

$$(M_{k2^{-n}})^2 - 2X^n_{k2^{-n}} = \sum_{j=0}^{k-1} (M^2_{(j+1)2^{-n}} - M^2_{j2^{-n}}) - \sum_{j=0}^{k-1} 2M_{j2^{-n}} (M_{(j+1)2^{-n}} - M_{j2^{-n}})$$

$$= \sum_{j=0}^{k-1} (M_{(j+1)2^{-n}} - M_{j2^{-n}})^2 = [M]^n_{k2^{-n}}$$

Hence, $M_t^2 - 2X_t^n$ is increasing along the sequence of times $(k2^{-n}, 1 \le k \le \lfloor 2^n T \rfloor$ Passing to the limit $n \to \infty$, $M_t^2 - 2X_t^n$ must be a.s. increasing. Set

$$[M]_t = M_t^2 - 2Y_t, t \in [0,T]$$

on the set where $M^2 - 2Y$ is increasing and $[M] \equiv 0$ otherwise. Hence, $[M]$ is a continuous increasing process and $M^2 - [M] = 2Y$ is a martingale on $[0,T]$.

We extend the definition of $[M]_t$ to $t \in [0,\infty)$ by applying the foregoing for all $T \in \mathbb{N}$.

Note that the process $[M]$ obtained with T is the restriction to $[0,T]$ of $[M]$ defined with $T+1$

Now, note that $M^2_{2^{-n}\lfloor t2^n \rfloor}$ converges to M^2 u.c.p. by uniform continuity, and convergence of X^n towards X also holds in the u.c.p. sense since it holds in the stronger $(\mathrm{M}_c^2, \|\cdot\|)$ sense.

Thus the theorem is proved when M is bounded.

Now we turn to the general case $M \in \mathrm{M}_{c,loc}$. Define

$$T_n = \inf\{t \ge 0 : |M_t| \ge n\}.$$

Then $(T_n)_{n\in\mathbb{N}}$ reduces M and we can apply the bounded case to M^{T_n}, writing $A^n = [M^{T_n}]$. By uniqueness, $A^{n+1}_{t\wedge T_n}$ and A^n_t are indistinguishable. Thus there exists an increasing process A such that for all $n \in \mathbb{N}$, $A_{t\wedge T_n}$ and A^n_t are indistinguishable. Define $[M]_t = A_t$. By construction, $(M^2_{t\wedge T_n} - A_{t\wedge T_n})_{t\ge 0} \in \mathrm{M}_c$ and so $(M_t^2 - A_t)_{t \ge 0} \in \mathrm{M}_{c,loc}$ as required. It remains to show that $[M]$ is the u.c.p. limit of its dyadic approximations. Let $[M]^{(m)}$ be the dyadic approximation at stage m m. Note that for fixed $n \ge 1$, , we have $[M^{T_n}]^m \to [M^{T_n}]$ u.c.p as $m \to \infty$ by the bounded case. Since for all fixed $t \ge 0$; $\mathrm{P}(T_n \ge 1) \to 1$ as $n \to \infty$, we obtain that $[M]^m \to [M]$ u.c.p. as $m \to \infty$.

Remark (2.1.13):

The idea of the proof itself is based on Ito's formula. Indeed, as we will soon prove, if $M \in \mathrm{M}_{c,loc}$. then

$$M^2 = 2M.M + [M]$$

so the idea of the definition of H^n and then $X^n = H^n.M$ is to take an approximation of M and $M.M$. So the martingale Y in the proof really is $M.M$, and this is why we define $[M] = M^2 - 2Y$.

Remark (2.1.14):

Note that $[M]$ is non-decreasing and thus of finite variation, and that if T is any possibly random time, $[M^T] = [M]^T$.

Theorem(2.1.15):

If $M \in \mathrm{M}_c^2, M^2 - [M]$ is a uniformly integrable martingale. Moreover, $\mathrm{E}([M]_\infty) = \mathrm{E}((M_\infty - M_0)^2)$ Conversely, if $M \in \mathrm{M}_{c,loc}$. with $\mathrm{E}(M_0^2) < \infty$ and $\mathrm{E}([M]_\infty) < \infty$, Then $M \in \mathrm{M}_c^2$.

Proof: Since the quadratic variation is unchanged by the addition of a constant to M, we may assume wlog that $M_0 = 0$ a.s. For the first direction, assume $M \in \mathrm{M}_c^2$. Let $S_n = \inf\{t \geq 0 : [M]_t \geq n\}$. S_n is a stopping time and $[M]_{t \wedge S_n} \leq n$. Thus, the stopped local martingale satisfies

$$\left| M_{t \wedge S_n}^2 - [M]_{t \wedge S_n} \right| \leq n + \sup_{t \geq 0} M_t^2 \tag{2.29}$$

is bounded by an integrable random variable and thus a true martingale (see remark after Proposition (1.2.14). Thus

$$\mathrm{E}([M]_{t \wedge S_n}) = \mathrm{E}(M_{t \wedge S_n}^2) < \infty \text{ for all } t \geq 0 \tag{2.30}$$

We take the limit $t \to \infty$, using monotone convergence on the left and dominated convergence on the right, and then $n \to \infty$ by the same arguments to get

$$\mathrm{E}([M]_\infty) = \mathrm{E}(M_\infty^2) < \infty \tag{2.31}$$

Hence, $|M_t^2-[M]_t|$ is dominated by $\sup_{t\geq\infty} M_t^2+[M]_\infty$, which is integrable. Thus $M^2-[M]$ is a true martingale and is uniformly integrable since:

$$\mathrm{E}(\sup_{t\geq 0}|M_t^2-[M]_t|)\leq \mathrm{E}((\sup_{t\geq 0} M_t)^2+[M]_\infty)\leq 5\mathrm{E}(M_\infty^2)<\infty \quad (2.32)$$

Remark (2.1.16):

Some textbooks use the notation $\langle M\rangle$ rather than $[M]$ for the quadratic variation. In general (i.e., in the discontinuous case), $\langle M\rangle$ should be previsible and means something slightly different (beyond the scope of this course), but it coincides with $[M]$ when M is continuous.

Section (2.2): $\text{It}\hat{\text{o}}'$sIntegrals and $\text{It}\hat{\text{o}}'$s Formula

Now we Start to Study Ito's Integrals.

Proposition (2.2.1):

Let μ be a finite measure on the previsible σ-algebra $\mathcal{P}$. Then $\mathcal{S}$ is a dense subspace of $L^2(\mathcal{P},\mu)$.

Proof: If $H\in\mathcal{S}$ then H is bounded so $H\in L^2(\mathcal{P},\mu)$. Thus $\mathcal{S}\subseteq L^2(\mathcal{P},\mu)$.. Denote by $\bar{\mathcal{S}}$ the closure of $\mathcal{S}$ in $L^2(\mathcal{P},\mu)$. Since linear combinations of indicator functions of the form $1_{\{A\}}$ for $A\in\mathcal{P}$ are dense in $L^2(\mathcal{P},\mu)$. by measure theory, it suffices to prove that if $A\in\mathcal{P}$, then $A\in\bar{\mathcal{S}}$. Set

$$\mathcal{A}:=\{A\in\mathcal{P}:1_{\{A\}}\in\bar{\mathcal{S}}\}. \quad (2.33)$$

Then $\mathcal{A}$ is a d-system. [Check: $1_{\{\Omega\times(0,n)\}}\in\mathcal{S}$ so $1_{\{\Omega\times(0,\infty)\}}\in\bar{\mathcal{S}}$ and $\Omega\times(0,\infty)\in\mathcal{A}$; if $C\subseteq D\in\mathcal{A}$ then $D\setminus C\in\mathcal{A}$; if $C_n\in\mathcal{A}$ and $C_n \uparrow C$ then $C\in\mathcal{A}$ since $\bar{\mathcal{S}}$ is the closure of $\mathcal{S}$ in $L^2(\mathcal{P},\mu)$]. Moreover $\mathcal{A}$ contains the π-system $\{B\times(s,t]$: $B\in\mathcal{F}_s, s\leq t\}$, which generates $\mathcal{P}$. Hence, by Dynkin's lemma, $\mathcal{A}=\mathcal{P}$.

Given $M\in\mathcal{M}_c^2$, define a measure μ on $\mathcal{P}$ by

$$\mu(A\times(s,t])=\mathrm{E}(1_{\{A\}}([M]_t-[M]_s)) \text{ for all } s<t, A\in\mathcal{F}_s. \quad (2.34)$$

Since $\mathcal{P}$ is generated by the π-system of events of this form, this uniquely specifies μ. Alternatively, write

$$\mu(d\omega \otimes dt) = d[M](\omega, dt)\mathrm{P}(d\omega), \tag{2.35}$$

where for a fixed ω, $d[M](\omega,.)$ is the Lebesgue-Stieltjes measure associated to the non-decreasing function $[M](\omega)$. Thus, for a previsible process $H \geq 0$,

$$\int_{\Omega\times(0,\infty)} H d\mu = \mathrm{E}\left(\int_0^\infty H_s d[M]_s\right) \tag{2.36}$$

Definition (2.2.2):

Set $L^2(M) = L^2(\Omega\times(0,\infty), \mathcal{P}, \mu)$ and write

$$\| H \|_M^2 = \| H \|_{L^2(M)}^2 = \mathrm{E}\left(\int_0^\infty H_s^2 d[M]_s\right) \tag{2.37}$$

so that $L^2(M)$ is the space of previsible processes H such that $\| H \|_M^2 < \infty$. Note that the simple processes $\mathcal{S} \subseteq L^2(M)$ for all $M \in \mathcal{M}_c^2$. Now, recall that if $M \in \mathcal{M}_c^2$ we defined $\| M \|^2 = \mathrm{E}(M_\infty^2)$.

Theorem (2.2.3): (Itô Isometry.)

For every $M \in \mathcal{M}_c^2$ there exists a unique isometry

$$I = (L^2(M), \|.\|_M) \to (\mathcal{M}_c^2, \|.\|)$$

such that

$$I(H) = H.M \text{ for all } H \in \mathcal{S}.$$

Proof: Let $H = \sum_{k=0}^{n-1} Z_k 1_{\{(t_k, t_{k+1}]\}} \in \mathcal{S}$. By Proposition (2.1.3), $H.M \in \mathcal{M}_c^2$ with

$$\| H.M \|^2 = \sum_{k=0}^{n-1} \mathrm{E}(Z_k^2(M_{t_{k+1}} - M_{t_k})^2). \tag{2.38}$$

But $M^2 - [M]$ is a martingale so that

$$\mathrm{E}(Z_k^2(M_{t_{k+1}} - M_{t_k})^2) = \mathrm{E}(Z_k^2 \mathrm{E}((M_{t_{k+1}} - M_{t_k})^2 \mid \mathcal{F}_{t_k})) =$$

$$= \mathrm{E}(Z_k^2 \mathrm{E}(M_{t_{k+1}}^2 - M_{t_k}^2 \mid \mathcal{F}_{t_k})) = \mathrm{E}(Z_k^2([M]_{t_{k+1}} - [M]_{t_k})), \tag{2.39}$$

And so $\| H.M \|^2 = \mathrm{E}\left(\int_0^\infty H_s^2 d[M]_s\right) = \| H \|_M^2$.

Now let $H \in L^2(M)$. We have thus defined a function I from $\mathcal{S}$ to $\mathcal{M}_c^2$, which is an isometry. However, $\mathcal{S}$ is dense in $L^2(M) = L^2(\mathcal{P}, \mu)$ by Proposition (2.2.1). This implies that there is a unique way to extend I to $H \in L^2(M)$ which makes I into an isometry. Indeed, let $H \in L^2(M)$. Then there exists H^n a sequence of simple processes such that $H^n \to H$ in $L^2(M)$. Then by linearity:

$$\| I(H^n) - I(H^m)) \| = \| I(H^n - H^m) \| = \| H^n - H^m \|_M$$

so $I(H^n)$ is a Cauchy sequence in $(\mathcal{M}_c^2, \|.\|)$, which is complete. Therefore, $I(H^n)$ converges to some limit which we may denote by $I(H)$. It is easy to check that $I(H)$ does not depend on the sequence H^n chosen to approximate H if $H^n \to H$ and $K^n \to H$ in $L^2(M)$,

then

$\| I(H^n) - I(K^n) \| = \| H^n - K^n \|_M \to 0$ as $n \to \infty$, so the limits of $I(H^n)$ and $I(K^n)$ must be indistinguishable. $I(H)$ is then, indeed, an isometry on $L^2(M)$. For $H \in \mathcal{S}$ we have consistently $I(H) = H.M$ by choosing $H^n \equiv H$.

Definition (2.2.4):

We write

$$I(H)_t = (H.M)_t = \int_0^t H_s dM_s$$

for all $H \in L^2(M)$.. The process $H.M$ is Itô's stochastic integral of H with respect to M .

Remark (2.2.5):

By Theorem (2.2.3), this is consistent with our previous definition of $H.M$ for $H \in \mathcal{S}$.

Proposition (2.2.6):

Let $M \in \mathcal{M}_c^2$ and $H \in L^2(M)$. Let T be a stopping time. Then $H 1_{(0.T]} \in L^2(M)$ and $H \in L^2(M^T)$, and we have:

$$(H.M)^T = (H1_{\{(0,T]\}}).M = H.(M^T). \tag{2.40}$$

Proof: Let $H \in L^2(M)$. It is trivial to check that $H1_{(0,T]} \in L^2(M)$. (to see that it is previsible, note that $1_{(0,T]}(t)$ is left-continuous and hence previsible). To see that $H \in L^2(M^T)$, note that $[M^T] = [M]^T$ by the discrete approximation in the definition of quadratic variation, and thus

$$\mathrm{E}\int_0^\infty H_s^2 d[M^T]_s = \mathrm{E}\int_0^T H_s^2 d[M]_s \le \mathrm{E}\int_0^\infty H_s^2 d[M]_s < \infty$$

Step 1. Take $M \in \mathrm{M}_c^2$ and suppose first that $H \in \mathrm{S}$. If T takes only finitely many values, $H1_{\{(0,T]\}} \in \mathrm{S}$ and $(H.M)^T = (H1_{(0,T]}).M$ is easily checked. For general T, set $T_n = (2^{-n}\lceil 2^n T \rceil \wedge n$ which is a stopping time that takes only finitely many values.

Then $T_n \downarrow T$ as $n \to \infty$ and so

$$\| H1_{\{(0,T_n]\}} - H1_{\{(0,T]\}} \|_M^2 = \mathrm{E}\left(\int_0^\infty H_t^2 (1_{\{(0,T_n]\}} - 1_{\{(0,T]\}})^2(t) d[M]_t\right) \to 0 \tag{2.41}$$

As $n \to \infty$, by dominated convergence, and so $H1_{\{(0,T_n]\}}.M \to H1_{\{(0,T]\}}.M$ in M_c^2 by Theorem (2.2.3). But $(H.M)_t^{T_n} \to (H.M)_t^T$ a.s. by continuity and hence, $(H.M)^T = (H1_{\{(0,T]\}}).M$ since $(H.M)^{T_n} = (H1_{\{(0,T_n]\}}).M$. for all $n \in \mathbb{N}$ by the first part. On the other hand we already know $(H.M)^T = H.(M^T)$ by Proposition (2.1.3).

Step 2. Now for $H \in L^2(M)$.choose $H^n \in \mathrm{S}$ such that $H^n \to H$ in $L^2(M)$. Then $H^n.M \to H.M$ in M_c^2 so $(H^n.M)^T \to (H.M)^T$ in M_c^2 by Doob's inequality.

Also,

$$\| H^n 1_{\{(0,T]\}} - H1_{\{(0,T]\}} \|_M^2 = \mathrm{E}\left(\int_0^T (H^n - H)_s^2 d[M]_s\right) \le \left\| H^n - H \right\|_M^2 \to 0 \tag{2.42}$$

as $n \to \infty$, so $(H^n 1_{\{(0,T]\}}).M \to (H1_{\{(0,T]\}}).M$ in M_c^2 by the isometry property of theorem (2.2.3). Again, by equating the limits of both sequences we get $(H.M)^T = (H1_{(0,T]}).M$ Moreover,

$$\| H^n - H \|_{M^T}^2 = \mathrm{E}\left(\int_0^\infty (H^n - H)_s^2 d[M^T]_s \right) =$$

$$= \mathrm{E}\left(\int_0^T (H^n - H)_s^2 d[M]_s \right) \le \| H^n - H \|_M \to 0 \qquad (2.43)$$

so $H^n.(M^T) \to H.(M^T)$ in M_c^2. Hence, $(H.M)^T \to H.(M)^T$.

Proposition (2.2.6) allows us to make a final extension of Itȏ's integral to locally bounded, pre-visible integrands.

Definition (2.2.7):

Let H be previsible. Say that H is locally bounded if there exist stopping times $S_n \uparrow \infty$ a.s. such that $H1_{(0,S_n]}$ is bounded for all $n \in \mathbb{N}$, i.e. there exists $C_n < \infty$ nonrandom such that $\sup_{t\ge 0} \left| H_t 1_{(0,S_n]}(t) \right| \le C_n$ a.s..

Note that if H_t is càdlàg and adapted, then H_{t^-} is previsible and locally bounded.

Definition(2.2.8):

Let H be a previsible locally bounded process and let $M \in \mathrm{M}_{c,loc}$ Choose stopping times $S_n^{'} = \inf\{t \ge 0 : |M_t| \ge n\} \uparrow \infty$ a.s., and note that $M^{S_n^{'}} \in \mathrm{M}_c^2$ for all $n \in \mathbb{N}$. Set $T_n = S_n \wedge S_n^{'}$, arid define

$$(H.M)_t = (H1_{(0,T_n]}.M^{T_n})_t \text{ for all } t \le T_n \qquad (2.44)$$

Remarks (2.2.9):

(i) The stochastic integral in the right-hand side ofEquation(2.44) is well-defined: indeed, every bounded previsible process is in $L^2(M)$ whenever $M \in \mathrm{M}_c^2$. Moreover, $H1_{(0,T_n]}$ bounded and

$M^{T_n} = (M^{S'_n})^{T_n} \in \mathrm{M}_c^2$, so $H1_{(0,T_n]}.M^{T_n}$ makes sense (it falls within the category of processes covered by Theorem (2.2.3)).

(ii) Proposition (2.2.6) ensures that the right-hand side does not depend on n for all n large enough that $T_n \geq t$.

(iii) Note also that the definition does not depend on the sequence of stopping times $(T_n)_{n\geq 0}$ used to reduce M and H, so long as H^{T_n}, is bounded and $M^{T_n} \in \mathrm{M}_c^2$ for all $n \geq 0$.

(iv) It is furthermore consistent with our previous definitions of stochastic integral when $M \in \mathrm{M}_c^2$ and $H \in L^2(M)$.

Proposition (2.2.10):

If H is locally bounded previsible and $M \in \mathrm{M}_{c,loc}^2$ then for all stopping times T we have $(H.M)^T = (H1_{(0,T]}).M = H.(M^T)..$

Proof: Let us start by checking the first of these equalities. By Proposition (2.2.6), we know that

$$(H1_{(0,T_n]}.M^{T_n})^T = H1_{(0,T]}1_{(0,T_n]}.M^{T_n}$$

As $n \to \infty$, the left-hand side converges pointwise a.s. to $(H.M)^T$ by definition, while the right-hand side also converges pointwise a.s. to $H1_{(0,T]}.M$ since the sequence (T_n) also "reduce" $H1_{(0,T]}$ and M in the sense of Definition (2.2.8). The second equality follows the same argument.

Theorem (2.2.11): (Quadratic Variation of Stochastic Integral)

Let $M \in \mathrm{M}_{\mathrm{C},loc}$ and H be locally bounded previsible. Then $H.M$ is a continuous local martingale, whose quadratic variation is given by

$$[H.M] = H^2.[M].$$

Remark (2.2.12):

In practice, we often use this theorem in combination with Theorem (2.1.15) to conclude that $H.M$ is a true martingale. In addition, as already discussed informally at the very beginning of the construction, this is the property which in some sense motivates the entire construction of the stochastic integral.

Proof: Let T_n be a sequence of stopping; times which reduces both H and $M : H^{T_n}$ is bounded and $M^{T_n} \in \mathcal{M}_c^2$. By Proposition (2.2.10),

$$(H.M)^{T_n} = (H1_{(0,T_n]}).M^{T_n} \in \mathcal{M}_c^2 \tag{2.45}$$

which implies that $H.M$ is a continuous local martingale. To compute the quadratic variation, assume first that $M \in \mathcal{M}_c^2$ and that H is uniformly pounded in time. For any stopping time T, we have by the isometry property of Theorem (2.2.3):

$$\mathrm{E}((H.M)_T^2) = \mathrm{E}((H1_{(0,T]}.M)_\infty^2) =$$

$$= \mathrm{E}((H^2 1_{(0,T]}.[M])_\infty) = \mathrm{E}((H^2.[M])_T) \tag{2.46}$$

By the optional stopping theorem, we conclude that $(H.M)^2 - H^2.[M]$ is a martingale. Moreover, since H is locally bounded and $[M]$ continuous one also shows that $H^2.[M]$ is continuous with probability 1. Therefore, by Theorem (2.1.10), we have $[H.M] \equiv H^2.[M]$. In the general case, note that as a consequence of (i) and of the fact that $[M^T] = [M]^T$, we may write

$$[H.M] = \lim_{n\to\infty} [H.M]^{T_n}$$

$$= \lim_{n\to\infty} \left[(H.M)^{T_n} \right]$$

$$= \lim_{n\to\infty} \left[H1_{(0,T_n]}.M^{T_n} \right]$$

$$= \lim_{n\to\infty} H^2 1_{(0,T_n]}.[M]^{T_n} \text{ (by the above)}$$

$$= H^2.[M] \text{ (by monotone convergence)}$$

where the limits in these equalities are a.s. pointwise limits.

Theorem (2.2.13): (Stochastic Chain Rule)

Let H, K be locally bounded and previsible and $M \in \mathcal{M}_{c,loc}^2$. Then

$$H.(K.M) = HK.M.$$

We view this result as a stochastic chain rule, since it is telling us that:

$$d\left(\int_0^t K_s dM_s\right) = K_t dM_t.$$

This is a rule that is extremely useful in the practice of computing stochastic integrals. E.g., if $dY_t = H_t dX_t$ then $dX_t = (1/H_t)dY_t$.

Proof: The case $H, K \in \mathrm{S}$ is tedious but elementary. For H, K uniformly bounded and $M \in \mathrm{M}_c^2$, there exist $H^n, K^n \in \mathrm{S}, n \in \mathbb{N}$ such that. $H^n \to H$ and $K^n \to K$ in $L^2(M)$. Furthermore, we may also assume that $\| H^n \|_\infty$ and $\| K^n \|_\infty$ are uniformly bounded in n

(indeed, truncating K^n at $\| K \|_\infty + 1$ can only improve the L^2 difference between K^n and K). We first prove an upper bound on $\| H \|_{L^2(K.M)}$:

$$\begin{aligned}
\| H \|_{L^2(K.M)}^2 &= \mathrm{E}((H^2.[K.M])_\infty) \\
&= \mathrm{E}((H^2.(K^2.[M]))_\infty) \\
&\overset{*}{=} \mathrm{E}(((HK)^2.[M])_\infty) \\
&= \| HK \|_{L^2(M)}^2 \\
&\le \min\left\{ \| H \|_\infty^2 \| K \|_{L^2(M)}^2, \| H \|_{L^2(M)}^2 \| K \|_\infty^2 \right\},
\end{aligned} \tag{2.47}$$

since $[M]$ is non-decreasing and thus of finite variation. We have $H^n.(K^n.M) = (H^n K^n).M$ and using Equation(2.47)

$$\begin{aligned}
\left\| H^n.(K^n.M) - H.(K.M) \right\| &\le \left\| (H^n - H).(K^n.M) \right\| + \left\| H.((K^n - K).M) \right\| \\
&= \left\| H^n - H) \right\|_{L^2(K^n.M)} + \left\| H \right\|_{L^2((K^n-K).M)} \\
&\le \left\| H^n - H) \right\|_{L^2(M)} \left\| K^n \right\|_\infty + \left\| H \right\|_\infty \left\| K^n - K \right\|_{L^2(M)} \to 0 \quad \text{as } n \to \infty
\end{aligned}$$

So $H^n.(K^n.M) \to H.(K.M)$ in M_c^2. Similarly, $(H^n K^n).M \to (HK).M$ in M_c^2,. which implies the result.

Definition (2.2.14):

Let X be a continuous semimartingale $X = X_0 + M + A$ with $M \in \mathrm{M}_{c,loc}, A$. finite variation process and $M_0 = A_0 = 0$. We set the quadratic variation of X to be that its martingale part, $[X] = [M]$, independently of A.

This definition finds its justification in the fact that

$$\sum_{k=0}^{\lfloor 2^n t \rfloor - 1} (X_{(k+1)2^{-n}} - X_{k2^{-n}})^2 \to [X]_t \text{ u.c.p.} \tag{2.48}$$

as $n \to \infty$, as is not hard to show.

Definition (2.2.15):

For a continuous semimartingale X and H locally bounded and previsible, we define the stochastic integral

$$H.X = H.M + H.A, \text{ writing also } (H.X)_t = \int_0^t H_s dX_s, \tag{2.49}$$

where $H.M$ is Itȏ's integral from Definition (2.2.8) and $H.A$ is the finite variation integral defined in Proposition (1.2.10). We agree that

$$dZ_t = H_t dX_t \text{ means } Z_t - Z_0 = \int_0^t H_s dX_s \tag{2.50}$$

Note that $H.X$ is already given in Doob-Meyer decomposition and is thus obviously a continuous semimartingale. Under the additional assumption that H is left-continuous, one can show that the Riemann sum approximation to the integral converges.

Proposition (2.2.16):

Let X be a continuous semimartingale and H be a left-continuous, adapted and locally bounded process. Then

$$\sum_{k=0}^{\lfloor 2^n t \rfloor - 1} H_{k2^{-n}} (X_{(k+1)2^{-n}} - X_{k2^{-n}}) \to \int_0^t H_s dX_s \text{ u.c.p. as } n \to \infty \tag{2.51}$$

Proof: We can treat the finite variation part $X_0 + A$ and the local martingale part M separately. It suffices to show that

$$\sum_{k=0}^{\lfloor 2^n t \rfloor -1} H_{k2^{-n}}(M_{(k+1)2^{-n}} - M_{k2^{-n}}) \to (H.M)_t \text{ u.c.p. as } n \to \infty \tag{2.52}$$

when $M \in \mathrm{M}_{c,loc}$ with $M_0 = 0$. By localization, we can reduce to the case where $M \in \mathrm{M}_c^2$ and H_t is bounded uniformly for $t > 0$. Let $H_t^n = H_{2^{-n}\lfloor 2^n t \rfloor}$. Then $H_t^n \to H_t$ as $n \to \infty$ by left continuity. Now,

$$(H^n.M)_t = \sum_{k=0}^{\lfloor 2^n t \rfloor -1} H_{k2^{-n}}(M_{(k+1)2^{-n}} - M_{k2^{-n}}) + H_{2^{-n}\lfloor 2^n t \rfloor}(M_t - M_{2^{-n}\lfloor 2^n t \rfloor}) \tag{2.53}$$

where, since M is continuous (and therefore almost surely uniformly continuous on any compact interval), $M_t - M_{2^{-n}\lfloor 2^n t \rfloor} \to 0$ u.c.p. as $n \to \infty$. We can thus ignore the second term on the right. Now

$$\| H^n - H \|_M = E\left(\int_0^\infty (H_t^n - H_t)^2 d[M]_t\right) \to 0 \text{ as } n \to \infty \tag{2.54}$$

by bounded convergence and the fact that $H_t^n \to H_t$ for every t as $n \to \infty$. By the isometry property, $H^n.M \to H.M$ in M_c^2. Using Doob's inequality, it is easy to see that this implies u.c.p. convergence.

To step away from the theory for a moment and look at a concrete example, you should try your hands at proving the following result. This will be generalized in a moment in Theorem (2.2.27) so you can go look for some inspiration there if you are stuck.

Proposition (2.2.17):

Let $(M_t, t \geq 0)$ be a continuous local martingale. Then for all $t \geq 0$,

$$M_t^2 = M_0^2 + 2\int_0^t M_s dM_s + [M]_t.$$

In particular if $(B_t, t \geq 0)$ is a one-dimensional standard Brownian motion, then

$$B_t^2 = 2\int_0^t B_s dB_s + t$$

is a semi-martingale.

Now we study Covariation.

In practice we do not calculate integrals from first principles, but rather se tools of calculus such as integration by parts or the chin rule. In this section we derive these tools for stochastic integrals, which differ from ordinary calculus in certain correction terms. A useful tool for deriving these rules will be the covariation of two local martingales.

Definition(2.2.18):

Let $M, N \in \mathcal{M}_{c,loc}$ adapted to a common filtration $(\mathcal{F}_t, t \geq 0)$ satisfying the usual conditions, and set

$$[M,N] = \tfrac{1}{4}([M+N]-[M-N]) \qquad (\textit{polarization identity}) \tag{2.55}$$

$[M,N]$ *is called the covariation of* M *and* N.

Proposition (2.2.19):

Let $M, N \in \mathcal{M}_{c,loc}$ Then we have:

(i) $[M,N]$ is the unique (up to indistinguishability) continuous adapted process with finite variation such that $MN-[M,N]$ is a continuous local martingale started from 0.

(ii) For $n \geq 1$ and for all $t \geq 0$, let

$$[M,N]_t^n = \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} (M_{(k+1)2^{-n}} - M_{k2^{-n}})(N_{(k+1)2^{-n}} - N_{k2^{-n}}) \tag{2.56}$$

Then $[M,N]^n \to [M,N]$ u.c.p. as $n \to \infty$,

(iii) for $M, N \in \mathcal{M}_c^2, MN-[M,N]$ is a *UI* martingale

(iv) $[M,N]$ is a symmetric bilinear form.

Proof: (i) Note that $MN = \frac{1}{4}((M+N)^2-(M-N)^2)$ It is thus obvious that $MN-[M,N]$ is a continuous local martingale. Moreover, finite variation is an obvious consequence of the definition and uniqueness follows easily from Theorem (1.2.18).

(ii) and (iii) follow form polarizing the sum Equation(2.56) just as in Equation(2.55) and applying Theorems (2.1.9) and (2.1.14).

For (iv), the symmetry comes from the uniqueness in (i), while the bilinearity also follows from (i).

Remark (2.2.20):

Of course, $[M,M]=[M]$.

Theorem (2.2.21): (Kunita-Watanabe Identity)

Let $M,N\in \mathrm{M}_{c,loc}$ and H be a locally bounded previsible process. Then

$$[H.M,N]=H.[M,N]. \tag{2.57}$$

Proof: We may assume by localization that $M,N\in \mathrm{M}_c^2$ and that H is uniformly bounded in time. Note that $H.[M,N]$ is of finite variation, and thus by the uniqueness of Proposition (2.2.19), it suffices to prove that

$$(H.M)N-H.[M,N]\in \mathrm{M}_{c,loc}$$

By the optional stopping theorem, it suffices to prove that for all bounded stopping times T,

$$\mathrm{E}((H.M)_T N_T)=\mathrm{E}((H.[M,N])_T) \tag{2.58}$$

and by considering the stopped processes H^T, M^T and N^T it suffices to prove that $\mathrm{E}((H.M)_\infty N_\infty)=\mathrm{E}((H.[M,N])_\infty)$. If H is of the form $Z1_{\{(s,t]\}}$ with Z bounded F_s measurable, then this identity becomes

$$\mathrm{E}\{Z(M_t-M_S)N_\infty\}=\mathrm{E}\{Z([M,N]_t-[M,N]_S)\}.$$

However, note that since $MN-[M,N]$ is a martingale, we have:

$$\begin{aligned}\mathrm{E}\{Z(M_t-M_S)N_\infty\}&=\mathrm{E}\{ZM_t\mathrm{E}(N_\infty\mid \mathrm{F}_S)\}-\mathrm{E}\{ZM_S\mathrm{E}(N_\infty\mid \mathrm{F}_S)\}\\&=\mathrm{E}\{Z\mathrm{E}(M_tN_t-M_SN_S\mid \mathrm{F}_S)\}\\&=\mathrm{E}\{Z\mathrm{E}([M,N]_t-[M,N]_S\mid \mathrm{F}_S)\}\\&=\mathrm{E}\{Z([M,N]_t-[M,N]_S)\},\end{aligned}$$

as required. Equation(2.58) then extends by linearity to all $H \in \mathcal{S}$. If H is bounded, we may find a L sequence $H^n \to H$ in $L^2(M)$ such that $H^n \in \mathcal{S}$ and is uniformly bounded. The Lebesgue convergence theorem then shows that Equation(2.58) holds. This proves the result.

Remark (2.2.22):

Note that a consequence of this identity is that $[H.M, H.N] = H^2.[M,N]$.

Definition (2.2.23):

Let X,Y be continuous semi-martingales: We define their covariation $[X,Y]$ to be the covariation of their respective martingale parts in the Doob-Meyer decomposition.

It is not hard to see that $\lim_{n\to\infty}[X,Y]^n = [X,Y]$ u.c.p where

$$[X,Y]_t^n = \sum_{k=0}^{\lfloor 2^n t \rfloor -1} (X_{(k+1)2^{-n}} - X_{k2^{-n}})(Y_{(k+1)2^{-n}} - Y_{k2^{-n}})$$

An important property of the covariation is that two independent semi-martingales have zero covariation. However, just as there exist many pairs of random variables with zero correlation which are not independent, the converse is false. A notable exception is the Levy characterization of Brownian motion.

Proposition (2.2.24):

Let X,Y be continuous local martingales in a common filtration $(\mathcal{F}_t, t \geq 0)$ satisfying, the usual conditions, and assume that X and Y are independent, i.e., $\sigma(X_s, s \geq 0)$ and $\sigma(Y_s, s \geq 0)$ are independent. Then $[X,Y] \equiv 0$.

Remark (2.2.25):

In particular, if $B = (B^1, \ldots, B^d)$ is a d-dimensional $\mathcal{F}$ -Brownian motion, then $[B^i, B^j]_t = \delta_{i,j} t$.

Remark (2.2.26):

Note that the Kunita-Watanabe identity $[H.X,Y]=H.\{X,Y]=[X,H.Y]$ also holds for continuous semi-martingales.

Now we study Itô's formula.

Theorem (2.2.27): (Integration by parts)

Let X,Y be continuous semimartingales. Then

$$X_tY_t - X_0Y_0 = \int_0^t X_s dY_s + \int_0^t Y_s dX_s + [X,Y]_t. \tag{2.59}$$

Proof: Since both sides are continuous in t, it suffices to consider $t = M2^{-N}$ for $M,N \geq 1$. Note that

$$X_tY_t - X_sY_s = X_s(Y_t - Y_s) + Y_s(X_t - X_s) + (X_t - X_s)(Y_t - Y_s) \tag{2.60}$$

so for $n \geq N$

$$X_tY_t - X_0Y_0 = \sum_{k=0}^{M2^{n-N}-1} (X_{k2^{-n}}(Y_{(k+1)2^{-n}} - Y_{k2^{-n}}) + Y_{k2^{-n}}(X_{(k+1)2^{-n}} - X_{k2^{-n}}) +$$

$$(X_{(k+1)2^{-n}} - X_{k2^{-n}})(Y_{(k+1)2^{-n}} - Y_{k2^{-n}})$$

$$\xrightarrow{u.c.p} (X.Y)_t + (Y.X)_t + [X,Y]_t \text{ as } n \to \infty \tag{2.61}$$

by Proposition (2.2.16)

Note the extra covariation term which we do not get in the deterministic case. The next result, Itô's formula, tells us that a smooth function of a continuous semimartingale is again a continuous semimartingale and gives us its precise decomposition in a sort of chain rule.

Theorem (2.2.28): (Itô's Formula)

Let $X^1, X^2, ..., X^d$ be continuous semimartingales and set $X = (X^1, ..., X^d)$.

Let $f \in C^2(\mathbb{R}^d, \mathbb{R})$.. Then

$$f(X_t) = f(X_0) + \sum_{i=1}^{d} \int_0^t \tfrac{\partial f}{\partial x^i}(X_s) dX_s^i + \tfrac{1}{2} \sum_{i.j=1}^{d} \int_0^t \tfrac{\partial^2 f}{\partial x^i \partial x^j}(X_s) d[X^i, X^j]_s. \tag{2.62}$$

Remarks: (i)In particular, $f(X)$ is a continuous semimartingale with decomposition

$$f(X_t) = f(X_0) + \underbrace{\sum_{i=1}^{d}\int_0^t \tfrac{\partial f}{\partial x^i}(X_s)dM_s^i}_{\in \mathcal{M}_{c,loc}}$$

$$+\underbrace{\sum_{i=1}^{d}\int_0^t \tfrac{\partial f}{\partial x^i}(X_s)dA_s^i + \tfrac{1}{2}\sum_{i,j=1}^{d}\int_0^t \tfrac{\partial^2 f}{\partial x^i \partial x^j}(X_s)d[M^i, M^j]_s}_{\text{finite variation}}. \tag{2.63}$$

where the covariation of the $\mathbb{R}^d$-valued semimartingale $X = X_0 + A + M$ is $[X^i, X^j] = [M^i, M^j]$, due to quadratic variation and the polarization identity Equation(2.55).

(ii) Intuitive proof by Taylor expansion for $d = 1$:

$$f(X_t) = f(X_0) + \sum_{k=0}^{\lfloor 2^n t \rfloor - 1}(f(X_{(k+1)2^{-n}}) - f(X_{k2^{-n}})) + (f(X_t) - f(X_{\lfloor 2^n t \rfloor 2^{-n}})) =$$

$$= f(X_0) + \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} f'(X_{k2^{-n}})(X_{(k+1)2^{-n}} - X_{k2^{-n}}) +$$

$$+\tfrac{1}{2}\sum_{k=0}^{\lfloor 2^n t \rfloor - 1} f''(X_{k2^{-n}})(X_{(k+1)2^{-n}} - X_{k2^{-n}}) + \text{ error terms}$$

$$\xrightarrow{u.c.p} f(X_0) + \int_0^t f'(X_s)dX_s + \tfrac{1}{2}\int_0^t f''(X_s)d[X]_s \tag{2.64}$$

We will not follow this method of proof, because the error terms are hard to deal with.

Proof: (for $d = 1$)

Write $X = X_0 + A + M$, where A has total variation process V. Let

$$T_r = \inf\{t \geq 0 : |X_t| + V_t + [M]_t > r\}. \tag{2.65}$$

Then $(T_r)_{r\geq 0}$ is a family of stopping times with $T_r \uparrow \infty$. It is sufficient to prove Equation(2.62) on the time intervals $[0, T_r]$. Let $\mathcal{A} \subseteq C^2(\mathbb{R}, \mathbb{R})$ denote the subset of functions f for which the formula holds. Then

(i) $\mathcal{A}$ contains the functions $f(x) \equiv 1$ and $f(x) = x$.

(ii) $\mathcal{A}$ is a vector space. Below we will show that $\mathcal{A}$ is, in fact, an algebra, i.e. in addition

(iii) $f, g \in A \Rightarrow fg \in A$

Finally we will show that

(iv) if $f_n \in A$ and $f_n \to f$ in $C^2(B_r, \mathbb{R})$ for all $r > 0$ then $f \in A$, where $f_n \to f$ in $C^2(B_r, R)$ means that $\Delta_{n,r} \to 0$ as $n \to \infty$ with $B_r = \{x : |x| \le r\}$ and

$$\Delta_{n,r} = \max\left\{ \sup_{x\in B_r} |f_n(x) - f(x)|, \sup_{x\in B_r} |f_n^{'}(x) - f^{'}(x)|, \sup_{x\in B_r} |f_n^{"}(x) - f^{"}(x)| \right\} \tag{2.66}$$

(i) -(iii) imply that A contains all polynomials. By Weierstrass' approximation theorem, these are dense in $C^2(B_r, \mathbb{R})$ and so (iv) implies $A = C^2(B_r, \mathbb{R})$.

Proof of (iii): Suppose $f, g \in A$ and set $F_t = f(X_t), G_t = g(X_t)$. Since the formula holds for f and g, F and G are continuous semimartingales. Integration by parts (Theorem 2.2.27). yields

$$F_t G_t - F_0 G_0 = \int_0^t F_s dG_s + \int_0^t G_s dF_s + [F, G]_t \tag{2.67}$$

we have $F.G = F.(1.G)$ and using Itô's formula for $(1.G)_s = g(X_s) - g(X_0)$ we get by

$$\int_0^t F_s dG_s = \int_0^t f(X_s) g^{'}(X_s) dX_s + \tfrac{1}{2}\int_0^t f(X_s) g^{"}(X_s) d[X]_s \tag{2.68}$$

By the Kunita-Watanabe identity Theorem (2.2.21) we have $[f^{'}.X, G] = f^{'},[X, G]$. Applying this a second time for G leads to

$$[F, G]_t = [f^{'}(X).X, g^{'}(X).X]_t = \int_0^t f^{'}(X_s) g^{'}(X_s) d[X]_s \tag{2.69}$$

Substituting these into Equation(2.67), we obtain Itô's formula for fg.

Proof of (iv): Let $f_n \in A$ such that $f_n \to f$ in $C^2(B_r, \mathbb{R})$. Then

$$\int_0^{t\wedge T_r} |f_n^{'}(X_s) - f^{'}(X_s)| \, dV_s + \tfrac{1}{2}\int_0^{t\wedge T_r} |f_n^{"}(X_s) - f^{"}(X_s)| \, d[M]_s \le$$
$$\le \Delta_{n,r}(V_{t\wedge T_r} + \tfrac{1}{2}[M]_{t\wedge T_r}) \le r\Delta_{n,r} \to 0 \quad \text{as } n \to \infty \tag{2.70}$$

and so

$$\int_0^{t\wedge T_r} f_n'(X_S)dA_S + \tfrac{1}{2}\int_0^{t\wedge T_r} f_n''(X_S)d[M]_S \to \int_0^{t\wedge T_r} f'(X_S)dA_S + \tfrac{1}{2}\int_0^{t\wedge T_r} f''(X_S)d[M]_S$$

Moreover, $M^{T_r} \in \mathrm{M}_c^2$ and so

$$\left\|(f_n'(X).M)^{T_r} - (f'(X).M)^{T_r}\right\|^2 = \mathrm{E}\left(\int_0^{T_r}(f_n'(X_s) - f'(X_s))^2 d[M]_s\right)$$

$$\leq \Delta_{n,r}^2 \mathrm{E}([M]_{T_r}) \leq r\Delta_{n,r}^2 \to 0, \tag{2.71}$$

as $n \to \infty$ and so $(f_n'(X).M)^{T_r} \to (f'(X).M)^{T_r}$ in M_c^2. For any fixed r, $X^{T_r} \in B_r$ and taking the limit $n \to \infty$ in Itô's formula for f_n we obtain

$$f(X_{t\wedge T_r}) = f(X_0) + \int_0^{t\wedge T_r} f'(X_s)dX_s + \tfrac{1}{2}\int_0^{t\wedge T_r} f''(X_s)d[X]_s \tag{2.72}$$

Remark (2.2.29):

For $d > 1$, (i)becomes ' A contains the constant 1 and the coordinate functions $f_1(x) = x^1, \ldots, f_d(x) = x^d$'. Check that you can then follow the same argument, dealing with all the different components $X^i, M^i, [M^i, M^j]$ etc.

Corollary (2.2.30):

Let $X^1, X^2, \ldots, X^d$ be continuous semimaingales and set $X = (X^1, \ldots, X^d)$. Let $f \in C^2(\mathbb{R}_+ \times \mathbb{R}^d, \mathbb{R})$. Then

$$f(t, X_t) = f(0, X_0) + \int_0^t \frac{\partial f}{\partial t}(s, X_s)ds + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x^i}(X_s)dX_s^i$$

$$\frac{1}{2}\sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s)d[X^i, X^j]_s$$

Proof: This is an immediate consequence of Equation(2.62). Indeed, the process $t \mapsto t$ is non-decreasing and so of finite variation, so $(t, X_t^1, \ldots, X_t^d)$ is a $d+1$-dimensional semi-martingale. The result follows by applying Itô's formula to this d + 1-dimensional process, and observing that since $t \mapsto t$ is of finite variation, it does not contribute to any of the covariation terms.

We will also occasionally need the following generalization of Itô's formula which allows us to localise the process in some open set:

Proposition (2.2.31):

Let D be a domain (open and connected subset of $\mathbb{R}^d$) which is a proper subset. Let $f : D \to \mathbb{R}$ be a C^2 function on D. Then if X is a semimartingale such that

$X_0 \in D$ almost surely, and if $T = \inf\{t \geq 0 : X_t \notin D\}$ then we have:

$$f(X_t) = \int_0^t \sum_{i=1}^d \frac{\partial f}{\partial x_i}(X_s)dX_s^i + \frac{1}{2}\sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)d[X^i, X^j]_s$$

almost surely for all $t < T$.

Proof: We may assume without loss of generality that $X^i \in \mathcal{M}_c^2$ for each $1 \leq i \leq d$ Let $n \geq 1$ and defines $\tau_n = \inf\{t \geq 0 : d(X_t, D^c) \leq 1/n\}$. Then $\tau_n \leq T$ almost surely and τ_n is nondecreasing, hence $L = \lim_{n \to \infty} \tau_n$, exists. We have $L \leq T$ by passing to the limit in $\tau_n \leq T$, and we also claim that $L \leq T$. Indeed, since the distance is a continuous function, $d(X_L, D^c) = 0$.

Note that $d(X_L, D^c) = \inf_{y \in D^c} d(X_L, y) = \inf_{y \in D^c \cap \bar{B}(X_L, 1)} d(X_L, y)$.

Sinc D is open, $D^c \cap \bar{B}(X_L, 1)$ is compact and thus this distance is attained. This means that $X_L \in D^c$ which implies $L \leq T$ .Thus

$L = T$ .Let $\tau_n' = \inf\{t \geq 0 : \| X_t \| + \sum_{1 \leq i,j \leq d} V([X^i, X^j]_t) \geq n\}$,

where $V(X)$ denotes the total variation of the process X. Let $T_n = \tau_n \wedge \tau_n'$. Then $T_n < T$ for all $n \geq 0$ and T_n increases towards T almost surely.

Now, let us introduce a sequence of functions $(\varphi_m)_{m \geq 1}$ which are C^∞-approximations of the identity (such as the Gaussian density with mean 0 and covariance matrix $(1/m)I$.) Consider the function

$$f_{n,m} = (f 1_{\{D_n\}}) * \varphi_m$$

where D_n is the subdomain $D_n = \{x \in D : d(x, D^c) > 1/n\}$ and $*$ denotes the convolution of two functions, i.e., $f * g(x) = \int_{\mathbb{R}^d} f(y)g(x-y)dy$. Since φ_m is C^∞, and since convolution is a regularizing operation, the function $f_{n,m}$ is C^∞ for all n, m . Thus we can apply Itô's formula to $f_{n,m}$. Stopping at time T_n , we get:

$$f_{n,m}(X_{t\wedge T_n}) = \int_0^{t\wedge T_n} \sum_{i=1}^{d} \frac{\partial f_{n,m}}{\partial x_i}(X_s)dX_s^i + \frac{1}{2}\sum_{i,j=1}^{d}\int_0^{t\wedge T_n} \frac{\partial^2 f_{n,m}}{\partial x_i \partial x_j}(X_s)d[X^i, X^j]_s \quad (2.73)$$

However, since f is C^2 inside D , we have for all $x \in D_n$:

$$\frac{\partial f_{n,m}}{\partial x_i}(x) = \varphi_m * \frac{\partial f}{\partial x_i}$$

and

$$\frac{\partial^2 f_{n,m}}{\partial x_i \partial x_j}(X) = \varphi_m * \frac{\partial^2 f}{\partial x_i \partial x_j}(X)$$

Since φ_m is an approximation of the identity, this means that as $m \to \infty$,

$$\frac{\partial f_{n,m}}{\partial x_i}(x) \to \frac{\partial f}{\partial x_i}(x), \text{ and } \frac{\partial^2 f_{n,m}}{\partial x_i \partial x_j}(x) \to \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$$

pointwise in D . This implies that one can take the limit $m \to \infty$ in Equation(2.73). Indeed, the second term

$$\int_0^{t\wedge T_n} \frac{\partial^2 f_{n,m}}{\partial x_i \partial x_j}(X_s)d[X^i, X^j]_s \to \int_0^{t\wedge T_n} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)d[X^i, X^j]_s$$

converges because of the Lebesgue convergence theorem since each $(X^i)^{T_n} \in \mathrm{M}_c^2$ $(1 \le i \le d$ and $n \ge 1)$. To see that the first term also converges, applying the isometry property of the stochastic integral:

$$\left\| \left(\frac{\partial f_{n,m}}{\partial x_i} - \frac{\partial f}{\partial x_i} \right)(X).(X^i)^{T_n} \right\|_X^2 = \mathrm{E}\left(\int_0^\infty \left(\frac{\partial f_{n,m}}{\partial x_i} - \frac{\partial f}{\partial x_i} \right)^2 (X_s)d[(X^i)^{T_n}]_s \right)$$

Since for a fixed n, $[(X^i)^{T_n}]$ is bounded by n, and since $\frac{\partial f_{n,m}}{\partial x_i}(x) \to \frac{\partial f}{\partial x_i}(x)$ pointwise in $x \in D_n$, and these functions are uniformly bounded in m , we may apply the Lebesgue dominated convergence theorem, and get that the right-hand side converges to 0.

Thus we get the desired Ito formula for all $t \leq T_n$ almost surely. Letting $n \to \infty$ finishes the proof.

Chapter Three

Applications to Brownian Motion and Martingales

Section (3.1): Brownian Martingales and Dubins-Schwarz Theorem

As we will see in a few moments, martingales are very useful to understand (and ultimately prove results about) the behaviour of random processes.

We start with the following very useful observation.

Theorem (3.1.1): (Exponential Martingale)

Let $M \in \mathrm{M}_{c,loc}$ with $M_0 = 0$. Then $Z_t = \exp\left(M_t - \frac{1}{2}[M]_t\right)$ defines a continuous local martingale. We call Z the exponential (local) martingale of M.

Proof: The function $f(x,y) = \exp(x - y/2)$ is C^2 and M_t and $[M]_t$ are both semi martingales. By Itô's formula,

$$dZ_t = Z_t\left(dM_t - \frac{1}{2}d[M]_t\right) + \frac{1}{2}Z_t d[M]_t = Z_t dM_t \tag{3.1}$$

so $Z_t = Z_0 + \int_0^t Z_s dM_s$ is a local Martingale.

Applying this to Brownian motion, we find the following martingales, which are the basis of a finer study of Brownian motion. If $x, y \in \mathrm{C}^d$, we note $\langle x, y \rangle = \sum_{i=1}^d x_i \overline{y}_i$ their complex scalar product.

Theorem (3.1.2):

Let $(B_t, t \geq 0)$ be an (F_t)-*Brownian* motion in $d \geq 1$ dimensions.

(i) If $d = 1$ and $(B_0 \in L^1)$, the process $(B_t, t \geq 0)$ is a (F_t) -martingale.

(ii) If $d \geq 1$ and $(B_0 \in L^2)$, the process $(\|B_t\|^2 - dt, t \geq 0)$ is a (F_t) - martingale.

(iii) Let $d \geq 1$ and $u = (u_1, ..., u_d) \in \mathbf{C}^d$. Assume that $E\left[\left|\exp(\langle u, B_0 \rangle)\right|\right] < \infty$, the process defined by

$$M_t = \exp\left(\langle u, B_t \rangle - tu^2/2\right)$$

is also a (F_t) -martingale for every $u^2 \in \mathrm{C}^d$ where u^2 is a notation for $\sum_{i=1}^d u_i^2$.

Notice that in (iii), we are dealing with ℂ-valued processes. The definition of $E\left[X \,|\mathrm{G}\right]$ the conditional expectation for a random variable $X \in L^1(\mathrm{C})$ is $E\left[\Re X \,|\mathrm{G}\right] + iE\left|\Im X \,|\mathrm{G}\right|$, and we say that an integrable process $(X_t, t \geq 0)$ with values in C, and adapted to a filtrtion (F_t), is a martingale if its real and imaginary parts are. Notice that the hypothesis on B_0 in (iii) is automatically satisfied whenever it $u = iv$ is purely imaginary, i.e., $v \in \mathrm{R}$.

Proof: (i) if $s < t, \mathrm{E}\left[B_t - B_s \middle| \mathrm{F}_s\right] = \mathrm{E}\left[B_{t-s}^{(s)}\right] = 0,$ where $B_u^{(s)} = B_{u+s} - B_s$ has mean 0 and is independent of F_s, by the simple Markov property. The integrability of the process is obvious by assumption on B_0.

(ii) Since $B_0 \in L^2$ and $B_t - B_0$ is a normal random variable, we have by the triangle inequality that $B_t \in L^2$. For $s \leq t, B_t^2 = (B_t - B_s)^2 + 2B_s(B_t - B_s) + B_s^2$. Taking conditional expectation given F_s and using the simple Markov property gives $\mathrm{E}\left[B_t^2\right] = (t - s) + B_s^2$, hence the result. A proof using Ito's formula is to say that B is an F -local martingale and

$$B_t^2 = 2\int_0^t B_s dB_s + t.$$

Thus $B_t^2 - t = M_t = 2\int_0^t B_s dB_S$ is an F -local martingale. It thus suffices to show that it is a true martingale, which can be proved for instance by observing that the quadratic variation is

$$[M]_t = \int_0^t B_s^2 ds$$

which has finite expectation for all $t > 0$ by Fubini's theorem. By Theorem (2.1.15) $(M_{s\wedge t}, s > 0)$ is a martingale bounded in L^2 and hence M is a true martingale.

(iii) To check integrability, note that $E\left[exp\left(\lambda B_t\right)\right] = exp\left(t\lambda^2/2\right)$ whenever B is a standard Brownian motion, and since $\left|e^z\right| = e^{\Re s}$, then we have

$$\mathbf{E}\left(\left|exp\langle u, B_t\rangle\right|\right) = \mathbf{E}\left(\left|exp\langle \in u, (B_t - B_0 + B_0)\rangle\right|\right)$$

$$= \mathbf{E}\left(\left|exp\left(u, B_t - B_0\right)\right|\right)\mathbf{E}\left(\left|exp\langle u, B_0\rangle\right|\right)$$

$$= exp\left(\sum_{i=1}^{d} t\left(\Re \mathrm{u}_i\right)^2/2\right)\mathrm{E}\left|exp\left(u, B_0\right)\right| < \infty.$$

To show that M is a martingale, consider $X_{d+1} = t$ which is a continuous semi-martingale.

Let $f\left(x_1,...,x_d,x_{d+1}\right) = exp\left(\sum_{i=1}^{d} u_i x_i - (1/2)u^2 x_{d+1}\right). f \in C^2\left(\mathrm{R}^{d+1},\mathrm{C}\right)$ so we may apply Itô's formula and obtain:

$$M_t = M_0 + \int \sum_{\mathrm{i}=1}^{\mathrm{d}} \mathrm{u}_\mathrm{i} exp\left(\langle u, B_s\rangle - su^2/2\right) dB_s^i$$

Since $d\left[B^{\mathrm{i}}, B^{\mathrm{j}}\right]_{\mathrm{t}} = \delta_{i,j} dt$ and $\left[B^i, t\right] = 0$ for all $1 \le i, j \le d$, so that the finite variations term cancel. It thus suffices to show that: $\int_0^t u_i exp\left(\langle u, B_s\rangle - su^2/2\right) dB_s^i$

is a true martingale. We take the quadratic variation of the real and imaginary parts, and it suffices by Fubini's theorem to show that

$$\int_0^t \mathrm{E}\left(\exp\left(\sum_{i=1}^{d} 2r_i B_s^i - s\langle u,\bar{u}\rangle^2\right)\right)ds < \infty \qquad (3.2)$$

where r_i is the complex modulus of u_i and $\langle u,\bar{u}\rangle^2 = \sum_{i=1}^{d} r_i^2$. Equation (3.2) follows instantly from the independence of the coordinates and the fact that $E\left[exp\left(rB_t\right)\right] = exp\left(tr^2/2\right)$.

A classical application of these martingales is to show the following result, often referred to as the gambler's ruin estimates.

Theorem (3.1.3):

Let $(B_t, t \geq 0)$ be a standard Brownian motion and $T_x = inf\{t \geq 0 : B_t = x\}$.Then for $x, y > 0$, one has

$$\mathrm{P}\left(T_{-y} < T_x\right) = \frac{x}{x+y},\ \mathrm{E}\left[T_x \wedge T_{-y}\right] = xy$$

Proof: Let $T = T_{-y} \wedge T_x$, which is a stopping time. Moreover, B^T is bounded $\left(\text{by } \max(x, y)\right)$ so we may apply the optional stopping theorem to find that $\mathrm{E}(B_T) = \mathrm{E}(B_0) = 0$. On the other hand, $\mathrm{E}(B_T) = -yp + x(1-p)$, where $p = P\left(T_{-y} < T_x\right)$ is the probability of interest to us. Thus $py = (1-p)x$ and the first statement follows easily. For the second statement, observe that $B_{t\wedge T}^2 - (t \wedge T)$ is a martingale (since martingales are stable by stopping) and thus

$$\mathbf{E}\left(B_{t\wedge T}^2\right) = \mathbf{E}\left(t \wedge T\right)$$

We may let $t \to \infty$ since the left-hand side is bounded and the right-hand side is monotone, and deduce that

$$\mathbf{E}\left(B_T^2\right) = \mathbf{E}(T)$$

Using the first statement,

$$\mathbf{E}\left(B_T^2\right) = \frac{x}{x+y} y^2 + \frac{y}{y+x} x^2 = xy$$

and the claim follows.

Similarly,

Theorem (3.1.4):

The Laplace transform of T_x for $x \in \mathrm{R}$ is given by $\mathrm{E}\left(e^{-qTx}\right) = e^{-|x|\sqrt{2q}}$. Moreover, the random variable $T = T_x \wedge T_{-y}$ has a Laplace transform given by

$$\mathbf{E}\left(e^{-qT}\right) = \frac{sinh\sqrt{2q}x + \sinh\sqrt{2q}y}{\sinh(\sqrt{2q}(x+y))}$$

and when $y = x, T$ is independent from the event $\{T_{-x} < T_x\}$.

Proof: The first statement follows directly from the optional stopping theorem and the fact that $e^{\lambda B_t - (\lambda^2/2)_t}$ is a martingale which is bounded when stopped at T_x if $x \geq 0$ (which we may assume by symmetry). The second statement is a bit more involved. Let

$$M_t = e^{-\lambda^2 t/2} \, sinh\left(\lambda\left(B_t + y\right)\right)$$

is also a martingale since it can be written as

$$\frac{1}{2} e^{-\lambda^2 t/2} \, \mathrm{e}^{\lambda\left(B_t + y\right)} - \frac{1}{2} e^{-\lambda^2 t/2} \, \mathrm{e}^{-\lambda\left(B_t + y\right)}$$

which is the sum of two martingales. Now, stopping at $T = T_{-x} \wedge T_y, M$ is bounded so we can use the optional stopping theorem to obtain:

$$\sinh(\lambda y) = \mathrm{E}\left(\sinh\left(\lambda\left(B_T + y\right)\right)e^{-T\lambda^2/2}\right)$$

$$= \mathrm{E}\left(\sinh\left(\lambda(x+y)\right)e^{-T\lambda^2/2}1_{\{T_x > T_{-y}\}}\right)$$

Thus:

$$\mathrm{E}\left(e^{-T\lambda^2/2}1_{\{T_x > T_{-y}\}}\right) = \frac{\sinh(\lambda y)}{\sinh\left(\lambda(x+y)\right)}$$

By symmetry,

$$\mathrm{E}\left(e^{-T\lambda^2/2}1_{\{T_{-y} > T_x\}}\right) = \frac{\sinh(\lambda x)}{\sinh\left(\lambda(x+y)\right)}$$

Adding up the two terms,

$$\mathrm{E}\left(e^{-T\lambda^2/2}\right) = \frac{\sinh(\lambda y) + \sinh(\lambda x)}{\sinh\left(\lambda(x+y)\right)}$$

When $x = y$, it suffices to check that

$$\mathrm{E}\left(e^{-T\lambda^2/2}1_{\{T_{-y} > T_x\}}\right) = \mathrm{E}\left(e^{-T\lambda^2/2}\right)P(T_{-y} < T_x) = \frac{1}{2}\mathrm{E}\left(e^{-T\lambda^2/2}\right)$$

Which is easy to check.

Another family of martingales is provided by the result below. This is the first hint of a deep connection between Brownian motion and second-order elliptic partial differential operators, a theme which we will explore in greater detail later on in the research (This also connects to the theory of martingale problems developed by Stroock and Varadhan, which has proved to be one of the most successful tools in probability theory).

Theorem (3.1.5):

Let $(B_t, t>0)$ be a $(\mathcal{F}_t)$-Brownian motion. Let $f(t,x): \mathbb{R}_+ \times \mathbb{R}_d \to \mathbb{C}$ be continuously differentiable in the variable t and twice continuously differentiable in x. Then,

$$M_t^f = f(t,B_t) - f(0,B_0) - \int_0^t ds\left(\frac{\partial}{\partial t} + \frac{1}{2}\Delta\right) f(s,B_s),\ t \ge 0$$

is a $(\mathcal{F}_t)$-local martingale, where $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$ is the Laplacian operator acting on the spatial coordinate of f. If moreover, the first derivatives are uniformly bounded on every compact interval (that is, for all $T>0$,

$$\sup_{t\in[0,T]} \sup_{x\in\mathbb{R}^d} \left|\frac{\partial f}{\partial x_i}(t,x)\right| < \infty$$

for all $1 \le i \le d$), then M^f is a true martingale.

Proof: By Itô's formula,

$$M_t^f = \int_0^t \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(s,B_s)\,dB_s^i$$

is indeed a local martingale. The fact it is a true martingale when the first partial derivatives are uniformly bounded on every compact time interval, follows from the fact that the quadratic variation of M^f is bounded on every compact time interval, and hence it is a true martingale (even bounded in L^2) on every compact time interval.

Now we discuss Dubins-Schwarz Theorem.

We work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbf{P})$ where $(\mathcal{F}_t)_{t\ge 0}$ satisfies the usual conditions.

Theorem (3.1.6):(Levy's Characterization of Brownian Motion)

Let $X^1,\dots,X^d \in \mathrm{M}_{c,loc}$ The two following statements are equivalent.

(i) For all $t \geq 0$, $\left[X^i, X^j\right]_t = \delta_{ij} t$.

(ii) $X = \left(X^1,\dots, X^d\right)$ is a Brownian motion in $\mathbf{R}^d$.

Proof: It suffices to show that, for $0 \leq s \leq t, X_t - X_s \sim N\left(0,(t\text{-}s)I\right)$ and the increment is independent of F_s . By uniqueness of characteristic functions, this is equivalent to showing that for all $s \leq t$ and for all $\theta \in \mathbf{R}^d$.

$$\mathbf{E}\left(exp\left(i\langle\theta, X_t - X_s\rangle\right)\middle|\mathrm{F}_s\right) = exp\left(-\frac{1}{2}\|\theta\|^2 (t-s)\right). \tag{3.3}$$

(Here $\langle .,. \rangle$ is the usual scalar product on $\mathbf{R}^d$ and $\|\theta\|$ is the Euclidean norm). $Fix\, \theta \in \mathrm{R}^d$ and set $Y_t = \langle\theta, X_t\rangle = \theta_1 X_t^1 + \dots + \theta_d X_t^d$.

Then Y is a local martingale, and by the assumptions and the bilinearity of the covariation, $[Y] = \sum_{i=1}^{d} \theta_i^2 t = t\|\theta\|^2$. Define also

$$Z_t = exp\left(iY_t + \frac{1}{2}[\mathrm{Y}]_t\right) = exp\left(i\langle\theta, X_t\rangle + \frac{1}{2}|\theta|^2 t\right).$$

Z is the exponential martingale associated with iY_t which is a local martingale, so $Z \in \mathrm{M}_{c,loc}$. Moreover, Z is bounded on $[0, t]$ for all $t \geq 0$ $\left(\text{since } [Y]_t = \|\theta\|^2 t\right)$ and so is a true martingale by Proposition (1.2.14). Hence, $\mathbf{E}(\mathrm{Z}_t | \mathrm{F}_s) = \mathrm{Z}_s$, or equivalently:

$$\mathrm{E}\left(\frac{Z_t}{Z_s}\middle|\mathrm{F}_s\right) = 1, a.s.$$

Equation (3.3) follows directly.

Theorem (3.1.7): (Dubins-Schwarz Theorem)

Let $M \in \mathcal{M}_{c,loc}$ with $M_0 = 0$ and $[M]_\infty = \infty$ *a.s.*. Set

$T_s = \inf\{t \geq 0 : [M]_t > s\}, B_s = M_{T_s}$. Then T_s is an $(\mathcal{F}_t)_{t>0}$-stopping time. If $\mathcal{G}_s = \mathcal{F}_{T_s}$ then $(\mathcal{G}_s)_{s\geq 0}$ is a filtration and B is a $(\mathcal{G}_t)_{t\geq 0}$-Brownian motion. Moreover $M_t = B_{[M]_t}$.

Remark (3.1.8):

So any continuous local martingale is a (stochastic) time-change of Brownian motion. In this sense, Brownian motion is the most general continuous local martingale.

Proof: Since $[M]$ is continuous and adapted, T_s is a stopping time, and since $[M]_\infty = \infty$ it must be that $T_s < \infty$ *a.s.* for all $s \geq 0$. We start the proof by the following lemma.

Lemma (3.1.9):

B is *a.s. continuous.*

Proof: Note that $s \mapsto T_s$ is càdlàg and nondecreasing and thus B is càdlàg. So it remains to show $B_{s-} = B_s$ for all $s > 0$, or equivalently $M_{T_{s-}} = M_{T_s}$, where

$$T_{s-} = \inf\{t \geq 0 : [M]_t = s\}$$

and note that T_{s-} is also a stopping time. Let s > 0. We need to show that M is constant between T_{s-} and T_s whenever $T_{s-} < T_s$, i.e. whenever $[M]$ is constant. Note that by Theorem (2.1.15) $(M^2 - [M])^{T_s}$ is uniformly integrable since $\mathbf{E}([M^{T_s}]_\infty) < \infty$. Hence, by the optional stopping theorem (the uniformly integrable version, see Equation (1.38) we get:

$$\mathbf{E}(M^2_{T_s} - [M]_{T_s} \big| F_{T_{s-}}) = M^2_{T_{s-}} - [M]_{T_{s-}}$$

But by assumption, $[M]_{T_s} = [M]_{T_{s-}}$ and M is a martingale, we obtain

$$\mathbf{E}(M^2_{T_s} - M^2_{T_{s-}} \big| F_{T_{s-}}) = \mathbf{E}(M_{T_s} - M_{T_{s-}} \big|^2 F_{T_{s-}}) = 0$$

and so M is a.s. constant between T_{s-} and T_s. This proves that B is almost surely continuous at time s. To prove that B is a.s. continuous simultaneously for all $s \geq 0$, note that if $T_r = \inf\{t > 0 : M_t \neq M_r\}$ and $S_r = \inf\{t > 0 : [M]_t \neq M_r\}$ then the previous argument says that for all fixed $r > 0$ (and hence for all $r \in \mathbf{Q}_+$), $T_r = S_r$. a.s. But observe that T_r and S_r are both càdlag. Thus equality holds almost surely for all $r \geq 0$ and hence almost surely, M and $[M]$ are constant on the same intervals. This implies the result.

We also need the following lemma.

Lemma (3.1.10):

B is adapted to $(G_t)_{t \geq 0}$.

Proof: It is trivial to check that $(G_s)_{s \geq 0}$ is a filtration. Indeed, if $S \leq T$ a.s. are two stopping times for the complete filtration (F_t), and if $A \in Fs$, then for all $t \geq 0$,

$$A \cap \{T \leq t\} = (A \cap \{S \leq t\}) \cap \{T \leq t\}$$

up to zero-probability events. The first event in the right-hand side is in F_t by assumption, and the second is also in F_t since T is a stopping time. Since (F_t) is complete, we conclude that $A \in F_T$ as well, and hence $Fs \subseteq F_T$. From this, since $T_r \leq T_s$ almost surely if $r \leq s$, (G_s) is a filtration. It thus suffices to show that if X is a càdlàg adapted process and T is a stopping time, then $X_T 1_{\{T < \infty\}}$ is

$\mathcal{F}_T$ -measurable. Note that a random variable Z is $\mathcal{F}_T$ -measurable if $Z 1_{\{T \le t\}} \in \mathcal{F}_t$ for every $t \ge 0$. If T only takes countably many values $\{t_k\}_{k=1}^{\infty}$, then

$$X_T 1_{\{T<\infty\}} = \sum_{k=1}^{\infty} X_{t_k} 1_{\{T=t_k\}}$$

so it is trivial to check that $X_T 1_{\{T<\infty\}}$ is $\mathcal{F}_T$ -measurable, since every term in the above sum is. In the general case, let $T_n = 2^{-n}[2^n T]$ where $[x]$ denotes smallest $n \in \mathbf{Z}_+$ with $n \ge x$. Then T_n is also a stopping time, finite whenever $T < \infty$, and such that $T_n \ge T$ while $T_n \to T$ almost surely. Thus for all $u \ge 0$, and for all $n \ge 1$, $X_{T_n} 1_{\{T_n \le u\}}$ is $\mathcal{F}_u$ -measurable. Furthermore, by right-continuity of X , $\lim_{n\to\infty} X_{T_n} 1_{\{T_n \le u\}} = X_T 1_{\{T<u\}}$. Thus $X_T 1_{\{T_n<u\}}$ is $\mathcal{F}_u$ -measurable. Naturally, $X_T 1_{\{T=u\}} = X_u 1_{\{T=u\}}$ is also $\mathcal{F}_u$ -measurable, so we deduce that $X_T 1_{\{T_n \le u\}}$ is $\mathcal{F}_u$ -measurable.

Having proved this lemma, we can now finish the proof of the Dubins-Schwarz theorem .Fix $s > 0$. Then $[M^{T_s}]_\infty = [M]_{T_s} = s$, by continuity of $[M]$. Thus by Theorem (2.1.15), $M^{T_s} \in \mathcal{M}_c^2$ since $\mathbf{E}\left(\left[M^{T_s}\right]_\infty\right) < \infty$. In particular, $(M_{t \wedge T_s}, s \ge 0)$ is uniformly integrable by Doob's inequality. In particular, we get that $M_{T_r} \in L^2(\mathbf{P})$ for $r \le s$ (and s was arbitrary). Applying the uniformly integrable version of the optional stopping theorem Equation (see Equation (1.38) a first time, we obtain

$$\mathrm{E}\left(M_{T_r} | \mathcal{F}_{T_r}\right) = M_{T_r}$$

a.s. and thus B is a $\mathcal{G}$-martingale. Furthermore, since $M^{T_s} \in \mathcal{M}_c^2$ by Theorem (2.1.15), $(M^2 - [M])^{T_s}$ is also a uniformly integrable martingale. By Equation (1.38) again, for $r \le s$,

$$\mathrm{E}\left(B_s^2 - s \,\middle|\, \mathcal{G}_r\right) = \mathbf{E}\left((M^2 - [M])_{T_s} \,\middle|\, \mathcal{F}_{T_r}\right)$$

$$= M_{T_r}^2 - [M])_{T_r} = B_r^2 - r$$

Hence, $B \in \mathcal{M}_c$ with $[B]_s = s$ and so, by Levy's characterization, B is a $(\mathcal{G}_t)_{t>0}$ -Brownian motion.

Before we head on to our next topic, here are a few complements to this theorem, given without proof. The first result is a strengthening of the Dubins-Schwarz theorem.

Theorem (3.1.11):

Let M be a continuous local martingale with $M_0 = 0$ a.s. Then we may enlarge the probability space and define a Brownian motion B on it in such a way that

$$M = B_{[M]t} \ a.s. \ for \ all \ t \geq 0.$$

More precisely, taking an independent Brownian motion β , if

$$B_s = \begin{cases} M_{T_s} & for \quad s \leq [M]_\infty \\ M_\infty + \beta_s - [M]_\infty & for \ all \quad s \geq [M]_\infty \end{cases}$$

then B is a Brownian motion and for all $t \geq 0$, $M_t = B_{[M]_t}$.

Remark (3.1.12):

One informal but very informative!) conclusion that one can draw from this theorem is the fact that the quadratic variation should be regarded as a natural clock for the martingale. This is demonstrated for instance in the following theorem.

Example (3.1.13):

Let B be a Brownian motion and let h be a deterministic function in $L^2\left(\mathbf{R}^+, \mathrm{B}, \lambda\right)$ with Lebesgue measure λ. Set $X = \int_0^\infty h_s dB_s$. Then $X \sim N\left(0, \|h\|_2^2\right)$.

Theorem (3.1.14):

Let M be a continuous local martingale. Then

(a) $\mathrm{P}\left(\lim_{t\to\infty} |M|_t = \infty\right) = 0$.

(b) $\left\{\omega : \lim_{t\to\infty} M_t \text{ exists and is finite}\right\} = \left\{\omega : [M]_\infty < \infty\right\}$ *up to null events.*

(c) $\left\{[M]_\infty = \infty\right\} = \left\{\lim_{t\to\infty} \sup M_t = +\infty \text{ and } \lim_{t\to\infty} \inf M_t = -\infty\right\}$ *up to null events.*

Section (3.2): P1anar Brownian Motion and Dobiskr's Invariance Principle

As explained before, Brownian motion is the scaling limit of d-dimensional random walks (this theorem will actually be proved in its strong form in the Later). One of the most striking results about random walks is Polya’s theorem which says that simple random walk is recurrent in dimension 1 and 2, while it is transient in dimension 3. What is the situation for Brownian motion? Being the scaling limit of simple random walk, one might expect the answer to be the same for Brownian motion. It turns out that this is almost the case: there is however something subtle happening in dimension 2. In the planar case, Brownian motion is neighbourhood-recurrent (it visits any neighbourhood of any point “infinitely often”) but almost surely does not hit any point chosen in advance.

We work with the Wiener measure $\mathbf{W}$ on the space of continuous functions, and recall that $\mathbf{W}_x$, denote the law of a Brownian motion started at x. Let E_x denote the expectation under this probability measure. In the sequel, $B(x, r)$ and $\overline{B}(x, r)$ denote the Euclidean ball of radius r about $x \in \mathrm{R}^d$.

Theorem (3.2.1): (Recurrence/Transience Classification.)

Let $d \geq 1$ We have the following dimension-dependent behaviour.

(a) If $d = 1$, Brownian motion is point-recurrent in the sense that:

$$\mathbf{W}_0 \text{ - a.s. } \textit{for all } x \in \mathbb{R}, \{t \geq 0 : B_t = x\} \textit{ is unbounded}$$

(b) If $d \geq 3$, Brownian motion is transient, in the sense that $\|B_t\| \to \infty$ almost surely as $t \to \infty$.

(c) If $d = 2$, Brownian, motion is neighbourhood-recurrent, in the sense that for every $x \in \mathbb{R}^d$, every open set is visited infinitely often $\mathbf{W}_x$-almost surely. Equivalently, for any $\varepsilon > 0$,

$$\{t \geq 0 : \|B_t\| < \varepsilon\} \textit{ is unbounded}$$

$\mathbf{W}_x$-almost surely for every $x \in \mathbb{R}^2$. However, points are polar in the sense that for every $x \in \mathbb{R}^2$

$$\mathbf{W}_0(B_t = s \textit{ for some } t > 0) = 0.$$

Proof: (a) is a consequence of (ii) in proposition (1.1.20).

(b) Let $B = (B^1, \ldots, B^d)$ be a d-dimensional Brownian motion with $d = 3$. Clearly it suffices to prove the result for $d \geq 3$ since

$$\|B_t\|^2 \geq R_t^2 := \sum_{i=1}^{3} \|B_t^i\|^2$$

and we are precisely claiming that the right-hand side tends to infinity as $t \to \infty$. Now, for $d = 3$, a simple calculation shows that if $f(x) = 1/\|x\|$, then if $\Delta f = 0 \ in \ \mathbb{R}^3 \setminus \{0\}$. Thus by the local Ito's formula,

$$1/R_{t \wedge T} \text{ is a local martingale,}$$

where T is the hitting time of 0. Since it is nonnegative, it follows from Proposition (1.2.17) that it is a supermartingale. Being nonnegative, the martingale convergence theorem tel1 us that it has an almost sure limit M as $t \to \infty$, and it suffices to prove that $M = 0$ almost surely. Note that $E(M) < E(1/R_0) < \infty$, so that $M < \infty$ almost surely and thus $T = \infty$ almost surely. Now on the event $\{M > 0\}, R$ must be bounded, and thus so is $|B_t^1|$. This has probability 0 by (i) and hence $M = 0$ a.s.

(c) Let $d=2$ and let B be a planar Brownian motion. Assume without loss of generality that $B_0 = 1$. We are going to establish that starting from there, B will never hit 0 but will come close to it "infinitely often" (or rather, "unboundedly often"). For $k \in \mathbf{Z}$, let $R_k = e^k$ and let

$$T_k = inf \ \{t \geq 0 : \|B_t\| = R_k\}$$

and let

$$T = T_{-\infty} = inf \ \{t \geq 0 : \|B_t\| = 0\}.$$

Our first goal will be to show that $T = \infty$, almost surely. Define a sequence of stopping times T_n as follows $T_0 = 0$, and by induction if $Z_n = \|B_{T_n}\|$ then

$$T_{n+1} \ inf \ \{t \geq T_n, \|B_t\| \in \{e^{-1} Z_n, \ e Z_n\}\}.$$

Notice that if $k,\ m \geq 1$ are two large integers, the probability that $T_{-k} < T_m$ is the probability that Z_n visits e^{-k} before e^m. Put it another way, it is also the probability that $(logZ_n,\ n \geq 0)$ visits $-k$ before *m*.

On the other hand, by Itô's formula, $M_t = log\,\|B_{t \wedge T}\|$ is a local martingale since

$$(x, y) \to log\left(x^2 + y^2\right) \text{ is harmonic on } \mathbb{R}^2 \setminus \{(0,\ 0)\}.$$

Since M_t is bounded on $[T_n; T_{n+1}]$, it follows from the Optional Stopping Theorem that given $\log Z_n = k \in \mathbf{Z}$,

$$P\left(\log Z_{n+1} = k + 1 \mid logZ_n = k\right) = P\left(\log Z_{n+1} = k - 1 \mid \log Z_n = k\right) = 1/2.$$

Moreover, the strong Markov property of Brownian motion implies that $(\log Z_n,\ n \geq 0)$ is a Markov chain. In other words, $(\log Z_n,\ n \geq 0)$ is nothing but simple random walk on $\mathbf{Z}$. In particular, it is recurrent. Therefore, for any $m \geq 0$,

$$\mathbf{P}\left(T_{-k} < T_m\right) \to 0$$

as $k \to \infty$. Therefore,

$$\mathbf{P}\left(T_{-k} < T_m\right) = 0$$

for all $m \geq 0$. This implies that $T = \infty$ almost surely since $T_m \to \infty$ as $m \to \infty$. On the other hand, this argument shows that $T_k < \infty$ for all $k \in \mathbf{Z}$, and there are infinitely many times that B visits this ball after returning to a radius greater than 1. Thus the set of times such that $B_t \in B\left(0, R_k\right)$ is unbounded a.s.

Remark (3.2.2):

Notice that (iii) implies the fact that $\{t > 0 : B_t \in B(x, \varepsilon)\}$ is unbounded for every $x \in \mathbb{R}^2$ and every $\varepsilon > 0$, almost surely. Indeed, one can cover $\mathbb{R}^2$ by a countable union of balls of a fixed radius. In particular, the trajectory of a 2-dimensional Brownian motion is everywhere dense. On the other hand, it will a.s. never hit a fixed countable family of points (except maybe at time 0), like the points with rational coordinates!

Theorem (3.2.3):

Let $d = 2$ and identify $\mathbb{R}^2$ with the complex field $\mathbb{C}$. Let $f : D \to D'$ be analytic (i.e.. complex differentiable). Let $z \in D$ and consider a Brownian motion $(B_t, t \geq 0)$ started at z. Let $\mathrm{T}_D = inf\{t > 0 : B_t \notin D\}$. Then there exists a Brownian motion B' started at $f(z)$, and a nondecreasing random function $\sigma(t)$, such that

$$f(B_{t \wedge \mathrm{T}_D}) = B'_{\sigma(t) \wedge \mathrm{T}'_{D'}}$$

Where $\sigma(t) = \int_0^t |f'(B_s)|^2 ds$. In other words, $f(B)$ is a time-changed Brownian motion stopped when it leaves D'.

It will be useful to recall the Cauchy-Riemann equations for complex analytic functions: if $f = u + iv$ is a complex-differentiable function with real and imaginary parts u and v, then:

$$\begin{cases} \dfrac{\delta u}{\delta x} = \dfrac{\delta v}{\delta y} \\[2ex] \dfrac{\delta u}{\delta y} = \dfrac{-\delta v}{\delta x} \end{cases}$$

from which it follows by further differentiation that both u and v are harmonic functions (i.e., $\Delta u \equiv \Delta v \equiv 0$) all over D). Applying Ito's formula and the Cauchy-Riemann equations shows that the real and imaginary parts of $f(B_t)$ have zero covariation and that they have identical quadratic variation. Applying the ideas of the Dubins-Schwartz theorem yields the result.

In principle, Theorem (3.2.3) (in combination with the famous Riemann mapping theorem) can be used to extract all the information we need about the behaviour of Brownian motion. For instance, the exit distribution from a domain D is simply the conformal image of the uniform measure of the circle by a map from the disc to this domain.

Remark (3.2.4):

The ramifications of this result are huge. On the one hand, it serves as a bridge between probability and complex analysis. This is one aspect of the deep connection between random processes ad harmonic analysis (which will be further developed later on). On the other hand, conformal invariance of Brownian motion, already observed by Paul Levy in the 1940's, can be seen as the starting poir1t of Schramm-Loewner Evolution (SLE), one of the most fascinating recent theories developed in probability, which may be seen as a study random processes in the complex plane that possess the conformal invariance property.

As an example of application of conformal invariance to planar Brownian motion, we will mention the following result, due to Spitzer. $Let\,(B_t,\ t \geq 0)$ denote a planar Brownian motion started from (1,0). Let $\mathbf{W}_t$ denote the total number of windigs of the curve B about 0 up to time t. This counts $+1$ for each clockwise winding around 0 and -1 for each counterclockwise winding.

Theorem (3.2.5): (Spitzer's Winding Number Theorem)

We have the following convergence in distribution:

$\frac{4\pi W_t}{\log t} \to_d C$ a Cauchy distribution with parameter 1.

Recall thata Cauchy distribution(with parameter 1) has density $\frac{1}{\pi(1+x^2)}$ and has the same distribution as $.N/N^{`}$, where these are two independent standard Gaussian random variables. To deduce Theorem (3.2.5), observe that by scaling, W_t, has the same distribution as $\overline{W}_\varepsilon$, where $\overline{W}_\varepsilon$ is the number of windings by time 1 of a Brownian motion started from $(\varepsilon,0)$ and $\varepsilon = 1/\sqrt{t}$. It is a simple estimate that $|\overline{W}_\varepsilon - \theta_\varepsilon|$ is bounded in probability, where θ_ε, is the number of windings up to time T, the hitting time of the unit sphere $\{z : |z| = 1\}$ The result follows *since* $\varepsilon = 1/\sqrt{t}$ and $C$is symmetric about 0.

Now we discuss Donsker's invariance principle.

The following theorem completes the description of Brownian motion as a 'limit' of centered random walks as depicted in the beginning of the chapter, and strengthens the convergence of finite-dimensional marginals to that convergence in distribution.

We endow $C([0, 1], \mathbf{R})$, with the supremum norm, and recall that the product σ-algebra associated with it coincides with the Borel σ-algebra associated with this norm. We say that a function $F : C([0, 1]) \to \mathbb{R}$ is continuous if it is continuous with respect to this norm. Often, functions F defined on C will be called functionals. For instance, the supremum of a path on the interval [0,1] is a (continuous) functional.

Theorem (3.2.6): (Donsker's Invariance Principle)

Let $(X_n, n \geq 1)$ be a sequence of $\mathbb{R}$-valued integrable independent random variables with common law μ, such that

$$\int x\mu(dx)=0 \text{ and} \int x^2\mu(dx)=\sigma^2 \in (0,\infty) .$$

Let $S_0 = 0$ and $S_n = X_1 + ... + X_n$, and define a continuous process that interpolates linearly between values of S, namely

$$S_t = (1-\{t\})S_{\lfloor t \rfloor} + \{t\}S_{\lfloor t \rfloor+1} \; t \geq 0 ,$$

where $\lfloor t \rfloor$ denotes the integer part of t and $\{t\} = t - \lfloor t \rfloor$. Then

$$S^{[N]} := \left(\frac{S_{Nt}}{\sqrt{\sigma^2 N}}, 0 \leq t \leq 1 \right)$$

converges in distribution to a standard Brownian motion between times 0 and 1, i.e. for every bounded continuous function $F : C([0,\ 1]) \to \mathbb{R}$,

$$\mathbf{E}\left[F(S^{[N]})\right]_{n\to\infty} \to \mathbf{E}_0\left[F(B)\right] .$$

Notice that this is much stronger than what Proposition (1.1.1) says. Despite the slight difference of framework between these two results (one uses cadlage continuous-time version of the random walk, and the other uses an interpolated continuous version), Donsker's invariance principle is stronger. For instance, one can infer from this theorem that the random variable $N^{-1/2} \sup_{0\leq n\leq N} S_n$ converges to $\sup_{0\leq t\leq 1} B_t$ in distribution, because $f \to sup\ f$ is a continuous operation on $C([0,\ 1],\mathbb{R})$. Proposition (1.1.1) would be powerless to address this issue.

The proof we give here is an elegant demonstration that makes use of a coupling of the random walk with a Brownian motion, called the Skorokhod embedding theorem. It is however specific to dimension $d=1$. Suppose we are given a Brownian motion $(B_t, t \geq 0)$ on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

Let $\mu+(dx) = \mathbf{P}(X_1 \in dx)1_{\{x \geq 0\}}$ and $\mu-(dy) = \mathbf{P}(-X_1 \in dy)1_{\{y > 0\}}$ define two nonnegative measures. Assume that $(\Omega, \mathcal{F}, \mathbf{P})$ is a rich enough probability space so that we can further define on it, independently of $(B_t, t \geq 0)$, a sequence of independent identically distributed $\mathbb{R}^2$-valued random variables $((Y_n, Z_n), n \geq 1)$ with distribution

$$\mathbf{P}((Y_n, Z_n) \in dxdy) = \frac{1}{C}(x+y)\,\mu+(dx)\,\mu-(dy),$$

where $C > 0$ is the appropriate normalizing constant that makes this expression a probability measure (this is possible because X has a well-defined expectation).

We define a sequence of random times, by $T_0 = 0$, $T_1 = inf\{t \geq 0 : B_t \in \{Y_1, -Z_1\}\}$, and recursively,

$$T_n = inf\{t \geq T_{n-1} : B_t - B_{T_{n-1}} \in \{Y_1, -Z_1\}\}$$

By (ii) in Proposition (1.1.20), these times are a.s. finite. We claim that

Lemma (3.2.7): (Skorokhod's Embedding).

The sequence $(B_{T_n}, n \geq 0)$ has the same law as $(S_n, n \geq 0)$. Moreover, the intertimes $(T_n \text{-} T_{n-1}, n \geq 1)$ form a sequence of independent random variables with same distribution, and expectation $\mathbf{E}[T_1] = \sigma^2$

Proof: Let $\mathcal{F}^B$ be the filtration of the Brownian motion, and for each $n \geq 0$, introduce the filtration $\mathcal{G}^n = (\mathcal{G}_t^n, t \geq 0)$ defined by

$$\mathcal{G}_t^n = \mathcal{F}_t^B \vee \sigma(Y_1, Z_1, \ldots, Y_n, Z_n).$$

Since (Y_i, Z_i) are independent from $\mathcal{F}^B, (B_t, t \geq 0)$ is a $\mathcal{G}^n$-Brownian motion for every $n \geq 0$. Moreover, T_n is a stopping time for $\mathcal{G}^n$. It follows that if $\tilde{B}_t = \left(B_{T_n+t} - B_{T_n}, t \geq 0\right)$ then is $\tilde{B}_t$ independent from $\mathcal{G}^n_{T_n}$. Moreover, (Y_{n+1}, Z_{n+1}) is independent both from $\mathcal{G}^n_{T_n}$ and from $\tilde{B}$, therefore $(T_{n+1} - T_n)$, which depends only on $\tilde{B}$ and (Y_{n+1}, Z_{n+1}) is independent from $\mathcal{G}^n_{T_n}$. In particular, $(T_{n+1} - T_n)$ is independent from $\sigma(T_0, T_1, ..., T_n)$. More generally, we obtain that the processes $\left(B_{t+T_{n-1}} - B_{T_{n-1}}, 0 \leq t \leq T_n - T_{n-1}\right)$ are independent with the same distribution.

It therefore remains to check that B_{T_1} has the same law as X_1 and $\mathrm{E}[T_1] = \sigma^2$. Remember from Theorem(3.1.3) that given Y_1, Z_1, the probability that $B_{T_1} = Y_1$ *is* $Z_1 / (Y_1 + Z_1)$, as follows from the optional stopping theorem. Therefore, for every non-negative measurable function *f*, by first conditioning on $(Y_1 + Z_1)$, we get

$$\mathrm{E}\left[f(B_{T_1})\right] = \mathrm{E}\left[f\left(Y_1\right)\frac{Z_1}{Y_1 + Z_1} + f(-Z_1)\frac{Y_1}{Y_1 + Z_1}\right]$$

$$= \int_{\mathbb{R}_+ \times \mathbb{R}_+^*} \frac{1}{C}(x+y)\mu_+(dx)\mu_-(dy)\left(f(x)\frac{y}{x+y} + f(-y)\frac{x}{x+y}\right)$$

$$= \frac{1}{C}\int_{\mathbb{R}_+ \times \mathbb{R}_+^*} \mu_+(dx)\mu_-(dy)\left(yf(x) + xf(-y)\right)$$

$$= \frac{1}{C}\int_{\mathbb{R}_+} \mu_+(dx) f(x) \int_{\mathbb{R}_+^*} y\,\mu-(dy) + \frac{1}{C}\int_{\mathbb{R}_+^*} \mu-(dy) f(-y) \int_{\mathbb{R}_+} x\,\mu+(dx)$$

Now observe that since $\mathrm{E}[X_1] = 0$, it must be the case that

$$\int_{\mathbb{R}_+} x\mu+(dx) = \int_{\mathbb{R}_+^*} y\mu-(dy) = C'$$

say, and thus, the left hand side is equal to

$$\mathrm{E}\left[f\,(B_{T_1})\right]=\frac{C^{'}}{C}\int_{\mathrm{R}+}\left(f\,\left(x\right)\mu+\left(dx\right)+f\,\left(-x\right)\mu-\left(dx\right)\right)$$

$$=\frac{C^{'}}{C}\int_{\mathrm{R}+}f\,\left(x\right)\mu\left(dx\right)$$

$$=\frac{C^{'}}{C}\mathrm{E}\left[f\,(X_1)\right].$$

By taking $f\equiv 1$, it must be that $C^{'}=C$, and hence B_{T_1} has the same law as X_1

For $\mathrm{E}\left[T_1\right]$, recall from Theorem (3.1.3) that $\mathrm{E}\left[inf\,\{t\geq 0:B_t\in\{x,-y\}\}\right]=xy$, so by a similar conditioning argument as above,

$$\mathrm{E}\left[T_1\right]==\int_{\mathrm{R}_+\times\mathrm{R}_+^*}\frac{1}{C}(x+y)xy\mu+(dx)\mu-(dy)$$

$$=\frac{1}{C}\int_{\mathrm{R}_+}x^{2}\mu+(dx)\int_{\mathrm{R}_+^*}y\,\mu-(dy)+\frac{1}{C}\int_{\mathrm{R}_+^*}y^{2}\mu-(dy)\int_{\mathrm{R}_+}x\,\mu+(dx)$$

$$=\frac{C^{'}}{C}\int_{\mathrm{R}+}x^{2}\mu\left(dx\right)$$

$$=\frac{C^{'}}{C}\sigma^2$$

but since we already know that $C^{'}=C$, this shows that $\mathrm{E}\left[T_1\right]=\sigma^2$, as claimed.

We will need another lemma, which tells us that the times T_m are in a fairly strong sense localized around there mean $m\sigma^2$.

Lemma (3.2.8):

We have the following convergence as $N\to\infty$:

$$N^{-1}\sup_{0\leq n\leq N}\left|T_n-\sigma^2 n\right|\to 0\ a.s. \tag{3.5}$$

Proof: By the strong law of large numbers, note that $T_n / n \to \sigma^2$ almost surely. Thus, fix $\varepsilon > 0$. Then there exists $N_0 = N_0(\varepsilon, \omega)$ such that if $n \geq N_0, \left|n^{-1}T_n - \sigma^2\right| \leq \varepsilon$. Thus if $N_0 \leq n \leq N$, then

$$N^{-1}\left|T_n - n\sigma^2\right| \leq \frac{n}{N}\varepsilon \leq \varepsilon$$

Moreover, $N^{-1}\sup_{0 \leq n \leq N_0}\left|T_n - n\sigma^2\right|$ tends to 0 almost surely as $N \to \infty$, so this implies Equation (3.5).

Proof of Donsker's Invariance Principle: We suppose given a Brownian motion *B*. For $N \geq 1$, define $B_t^{(N)} = N^{1/2}B_{N^{-1}t}, t \geq 0$, which is a Brownian motion by scaling invariance. Perform the Skorokhod embedding construction on $B^{(N)}$ to obtain variables $\left(T_n^{(N)}, n \geq 0\right)$. Then, let $S^{(N)} = B_{T_n^{(N)}}^{(N)}$ Then by Lemma (3.2.7), $\left(S_n^{(N)}, n \geq 0\right)$ is a random walk with same law as $\left(S_n^{(N)}, n \geq 0\right)$. We interpolate linearly between integers to obtain a continuous process $\left(S_t^{(N)}, 0 \leq t \leq 1\right)$ which thus has the distribution as $\left(S_t, 0 \leq t \leq 1\right)$. Finally, let

$$\tilde{S}_t^{(N)} = \frac{S_{Nt}^{(N)}}{\sqrt{\sigma^2 N}}, t \geq 0$$

and $\tilde{T}_n^{(N)} = N^{-1}T_n^{(N)}$. Finally, let $B_t^{'} = B_{\sigma^2 t} / \sqrt{\sigma^2}$, which is also a Brownian motion. We are going to show that the supremum norm

$$\left\|B_t^{'} - \tilde{S}_t^{(N)}\right\|_\infty \to_p 0$$

is $N \to \infty$, where $\to_p$, denotes convergence in, probability.

First recall what we have proved in Equation (3.5), and note that this implies convergence in probability. Since $\left(T_n^{(N)}, n \geq 0\right)$has the same distribution as $(T_n, n \geq 0)$ we infer from this that for every $\delta > 0$, letting $\delta' = \delta\sigma^2 > 0$, we have:

$$\mathrm{P}\left(N^{-1} \sup_{0 \leq n \leq N} \left| T_n^{(N)} - n\sigma^2 \right| \geq \delta' \right) \underset{N \to \infty}{\to} 0.$$

Therefore dividing by σ^2:

$$\mathrm{P}\left(\sup_{0 \leq n \leq N} \left| \tilde{T}_n^{(N)} / \sigma^2 - n / N \right| \geq \delta \right) \underset{N \to \infty}{\to} 0.$$

Now, note that if $t = n / N$, then

$$\tilde{S}_t^{(N)} = \frac{S_t^{(N)}}{\sqrt{N\sigma^2}} = \frac{B_{\tilde{T}_n^{(N)}}^{(N)}}{\sqrt{\sigma^2}} = B'_{\tilde{T}_n^{(N)}/\sigma^2}.$$

Thus, by continuity, if $t \in \left[n / N, (n+1) / N \right]$, there exists $u \in \left[T_n^{(N)} / \sigma^2, T_{n+1}^{(N)} / \sigma^2 \right]$. such that $S_t^{(N)} = B'_{u}$. Therefore, for all $\varepsilon > 0$ and all $\delta > 0$, the event

$$\left\{ \sup_{0 \leq t \leq 1} \left| \tilde{S}_t^{(N)} - B'_t \right| > \varepsilon \right\} \subseteq K_\delta^N \cup L_{\delta,\varepsilon,}^N$$

where

$$K_\delta^N = \left\{ \sup_{0 \leq n \leq N} \left| \tilde{T}_n^{(N)} / \sigma^2 - n / N \right| > \delta \right\}$$

and

$$L_{\delta,\varepsilon} = \left\{ \exists t \in [0,1], \exists u \in [t - \delta, t + \delta + 1/N] : \left| B'_t - B'_u \right| > \varepsilon \right\}.$$

We already know that $P\left(K_\delta^{(N)}\right) \to 0$ *as* $N \to \infty$. For $L_{\delta,\varepsilon,}^{(N)}$, since B' is a.s. uniformly continuous on [0,1], by taking δ small enough and then N large enough, we can make $P\left(L^N\right)$ as small as wanted. More precisely, if

$$L_{2\delta,\varepsilon}^N = \left\{ \exists t \in [0,1], \exists u \in [t - 2\delta, t + 2\delta + 1/N] : \left| B'_t - B'_u \right| > \varepsilon \right\}.$$

then for $N \geq 1/\delta, L_{\delta,\varepsilon}^{N} \subseteq L_{2\delta,\varepsilon}$, and thus for all $\delta > 0$:

$$\limsup_{N\to\infty} \mathrm{P}\left(\left\|\tilde{S}^{(N)} - B^{'}\right\|_{\infty} > \varepsilon\right) \leq \mathrm{P}\left(L_{2\delta,\varepsilon}\right)$$

However, as $\delta \to 0$, $P\left(L_{2\delta,\varepsilon}\right) \to 0$ by almost sure continuity of $B^{'}$ on (0,1) and the fact that these events are decreasing. Hence it must be that

$$\limsup_{N\to\infty} \mathrm{P}\left(\left\|\tilde{S}^{(N)} - B^{'}\right\|_{\infty} > \varepsilon\right) = 0.$$

Therefore, $\left(\tilde{S}^{(N)}, 0 \leq t \leq 1\right)$ converges in probability for the uniform norm to $\left(B_t, 0 \leq t \leq 1\right)$, which entails convergence in distribution. This concludes the proof.

Section (3.3): Dirichiet Problem and Girsanov's Theorem

Let D be a connected open subset of R^d for some $d \geq 1$ (though the story is interesting only for $d \geq 2$). We will say that D is a domain. Let ∂D be the boundary of D. We denote by Δ the Laplacian on R^d.

Definition (3.3.1):

Let $g : \partial D \to \mathrm{R}$ be a continuous function. A solution of the Dirich Let problem with boundary condition g on D is a function $u : \bar{D} \to \mathrm{R}$ of class $\mathrm{C}^2(D) \cap \mathrm{C}(\bar{D})$, such that

$$\begin{cases} \Delta u = 0 \quad on \quad D \\ u|\partial D = \mathrm{g}. \end{cases} \tag{3.6}$$

A solution of the Dirichlet problem is the mathematical counterpart of the following physical problem: given an object made of homogeneous material, such that the temperature $\mathrm{g}(y)$ is imposed at point y of its boundary, the solution $u(x)$ of the Dirichlet problem gives g the temperature at the point x in the object when equilibrium is attained.

As we will see, it is possible to give a probabilistic resolution of the Dirichlet problem with the help of Brownian motion. This is essentially due to Kakutani. We let E_x be the law of the Brownian motion in R^d started at x. In the remaining of the section, let $T = \inf\{t \geq 0 : B_t \notin D\}$ be the first exit time from D. It is a stopping time, as it is the first entrance time in the closed set D^{C}. We will assume that the domain D is such that $\mathrm{P}(T < \infty) = 1$ to avoid complications. Hence B_T is a well-defined random variable.

In the sequel, $|\square|$ is the Euclidean norm on R^d. The goal now is prove the following result:

Theorem (3.3.2):

Suppose that $g \in \mathrm{C}(\partial D, \mathrm{R})$ is bounded, and assume that D satisfies a local exterior cone condition (l.e.c.c.), i.e. for every $y \in \partial D$, there exists a nonempty open convex cone with origin at y such that $C \cap B(y,r) \subseteq D^C$ for some $r > 0$. Then the function

$$u : x \to \mathrm{E}_x[g(B_T)]$$

is the unique bounded solution. In particular, if D is bounded and satisfies the l.e.c.c. then u is the unique solution of the Dirichlet problem.

We start with a uniqueness statement.

Proposition (3.3.3):

Let g be a bounded function in $\mathrm{C}(\partial D, R)$. Set

$$u(x) = \mathrm{E}_x[g(B_T)].$$

If v is a bounded solution of the Dirichlet problem, then $v = u$.

In particular, we obtain uniqueness when D is bounded. Notice that we do not make any assumption on the regularity of D here besides the fact that $T < \infty$ a.s.

Proof: Let v be a bounded solution of the Dirichlet problem. Let $T_n = \inf\left\{t \geq 0 : d\left(X_t, D^C\right) < \frac{1}{n}\right\}$. Since $\Delta v = 0$ inside D, we know by proposition (2.2.31) that $M_t = v\left(B_{t\wedge T_n}\right) - v\left(B_0\right)$ is a local martingale started at 0 (here, $B_0 = x$ almost surely). Moreover, since v is bounded, M is a true martingale which is uniformly integrable. Applying the optional stopping theorem (1.2.12) at the stopping time T_n,

$$\mathrm{E}\left(M_{T_n}\right) = \mathrm{E}_x\left(v\left(B_{T_n}\right) - v(x)\right) = 0.$$

Since $T_n \to T$ almost surely as $n \to \infty$, and since v is continuous on $C(\bar{D})$, we get:

$$v(x) = \mathrm{E}_x\left(g\left(B_T\right)\right).$$

as claimed.

For every $x \in \mathrm{R}^d$ and $r > 0$, let $\sigma_{x,r}$ be the uniform probability measure on the sphere $\mathrm{S}_{x,r} = \left\{y \in \mathbf{R}^d : |y - r| = r\right\}$. It is the unique probability measure on $\mathrm{S}_{x,r}$. that is invariant under isometrics of $\mathrm{S}_{x,r}$. We say that a locally bounded measurable function $h : D \to \mathbf{R}$ is harmonic on D if for every $x \in D$ and every $r > 0$ such that the closed ball $\bar{B}(x,r)$ with center x and radius r is contained in D,

$$h(x) = \int_{\mathrm{S}_{x,r}} h(y)\sigma_{x,r}(dy).$$

Proposition (3.3.4):

Let h be harmonic on a domain D. Then $h \in C^{\infty}(D, \mathbf{R})$, and $\Delta h = 0$ = 0 on D.

Proof: Let $x \in D$ and $\varepsilon > 0$ such that $\overline{B}(x,\varepsilon) \subseteq D$. Then let $\varphi \in C^{\infty}(\mathbf{R},\mathbf{R})$, be non-negative with non-empty compact support in $[0,\varepsilon[$. We have, for $0 < r < \varepsilon$,

$$h(x) = \int_{\mathbf{S}(0,r)} h(x+y)\sigma_{0,r}(dy).$$

Multiplying by $\varphi(r)r^{d-1}$ and integrating over $r \in (0,\varepsilon)$ gives

$$ch(x) = \int_{B(0,\varepsilon)} \varphi(|z|)h(x+z)\,dz,$$

where $C > 0$, is some constant, where we have used the fact that

$$\int_{\mathbf{R}^d} f(x)dx = c\int_{\mathbf{R}_+} r^{d-1}dr \int_{\mathbf{S}(0,r)} f(ry)\sigma_{0,r}(dy)$$

for some $C > 0$. Therefore,

$$ch(x) = \int_{B(x,\varepsilon)} \varphi(|z-x|)h(z)\,dz = \int_{\mathrm{R}^d} \varphi(|z-x|)h(z)\,dz$$

since φ is supported on $B(0,\varepsilon)$. By derivation under the $\int$ sign, we easily get that h is C^{∞}. (Indeed, we may assume that $r \to \varphi(r^{1/2})$ is C^{∞}). Another way to say this is to say that $ch(x) = \overline{\varphi} * h$ where $\tilde{\varphi}(z) = \varphi(|z|)$. If $r \to \varphi(r^{1/2})$ is C^{∞}, then $\tilde{\varphi} \in C^{\infty}(\mathbf{R}^{d}, \mathbf{R})$ and thus, the convolution being a regularizing operation, this implies $\tilde{\varphi} * h \in C^{\infty}(D,\mathbf{R})$. Thus $h \in C^{\infty}(D,\mathbf{R})$.

Next, by translation we may suppose that $0 \in D$ and show only that $\Delta h(0) = 0$. we may apply Taylor's formula to h, obtaining, as $x \to 0$,

$$h(x) = h(0) + (\nabla h(0), x) + \sum_{i=1}^{d} x_i^2 \frac{\partial^2 h}{\partial x_i^2}(0) + \sum_{i \neq j}^{d} x_i x_j \frac{\partial^2 h}{\partial x_i \partial x_j}(0) + o(|x|^2).$$

Now, integration over $\mathbf{S}_{0,r}$ for r small enough yields

$$\int_{\mathbf{S}_{x,r}} h(x)\sigma_{0,r}(dx) = h(0) + C_r \Delta h(0) + o(r^2)$$

Where $C_r = \int_{\mathbf{S}_{0,r}} x_1^2 \sigma_{0,r}(dx)$. Now, it is easy to see that there exists $c > 0$ such that $C_r > cr^2$ for all $0 \le r \le 1$. Since the left-hand side is $h(0)$ and the error term on the right-hand side is $o(r^2) = o(C_r)$, it follows that $\Delta h(0) = 0$.

Therefore, harmonic functions are solutions of certain Dirichlet problems.

Proposition (3.3.5):

Let g be a bounded measurable function on ∂D, and let $T = inf\{t \ge 0 : B_t \notin D\}$. Then the function. $h : x \in D \to \mathbf{E}_x[g(B_T)]$ is harmonic on D, and hence $\Delta h = 0$ on D.

Proof: For every Bore subsets $A_1, \ldots, A_k$ of $\mathbf{R}^d$ and times $t_1 < \ldots < t_k$, the map

$$x \to \mathbf{P}_x\left(B_{t_1} \in A_1, \ldots, B_{t_n} \in A_n\right)$$

is measurable by Fubini's theorem, once one has written the explicit formula for this probability. Therefore, by the monotone class theorem, $x \to E_x[F]$, is measurable for every integrable random variable F, which is measurable with respect to the product or-algebra on $C(\mathbf{R}, \mathbf{R}^d)$. Moreover, h is bounded by assumption.

Now, let $S = inf\{t \ge 0 : |B_t - x| \ge r\}$ the first exit time of B form the ball of center x and radius r. Then by (ii), Proposition (1.1.20), $S < \infty$ a.s. By the strong Markov property, $\tilde{B} = (Bs + t,\ t > 0)$ is an (F_{S+t}) Brownian motion started at B. Moreover, the first hitting time of ∂D for $\tilde{B}$ is $\tilde{T} = T \to S$. Moreover, $\tilde{B}_{\tilde{T}} = B_T$, so that

$$\mathbf{E}_x\left[g\left(B_T\right)\right] = \mathbf{E}_x[g(\tilde{B}_{\tilde{T}})] = \int_{\mathbf{R}^d} \mathbf{P}_x\left(B_s \in \mathrm{dy}\right)\mathbf{E}_y\left[g\left(B_T\right)1_{\{T<\infty\}}\right],$$

we recognize $\int_{\mathbf{R}^d} \mathbf{P}_x\left(B_s \in \mathrm{dy}\right)h(y)$ in the last expression.

Since B starts from x under P_x the rotation invariance of Brownian motion shows that $Bs - x$ has a distribution on the sphere of center 0 and radius r which is invariant under the orthogonal group, so we conclude that the distribution of Bs is the uniform measure on the sphere of center x and radius r, and therefore that h is harmonic on D.

It remains to understand whether the function u of Theorem (3.3.2). is actually a solution of the Dirichlet problem. Indeed, is not the case in general that $u(x)$ has limit $g(y)$ as $x \in D, x \to y$, and the reason is that some points of ∂D may be 'invisible' to Brownian motion. The researcher can convince himself, for example, that if $D = B(0,1)\setminus\{0\}$ is the open ball of $\mathbf{R}^2$ with center 0 and radius 1, whose origin has been removed, and if $g = 1_{\{0\}}$, then no solution of the Dirichlet problem with boundary constraint g exists. The probabilistic reason for that is that Brownian motion does not see the boundary point 0. This is the reason why we have to make regularity assumptions on D in the following theorem.

Proof of Theorem (3.3.2):

It remains to prove that under the l.e.c.c., u is continuous on $\bar{D}$, i.e. $u(x)$ converges to $g(y)$ as $x \in D$ converges to $y \in \partial D$. In order to do that, we need a preliminary lemma. Recall that T is the first exit time of D for the Brownian path.

Lemma (3.3.6):

Let D be a domain satisfying the l.e.c.c., and let $y \in \partial D.$ Then for every $\eta > 0, \mathbf{P}_x(T < \eta) \to 1$ as $D \to y.$

Proof: Let $C_y = y + C$ be a nonempty open convex cone with origin at y such that $C_y \subseteq D^c$. Then it is an elementary geometrical fact that there exists anonempty open convex cone $C^{'}$ with origin at 0 such that for every $\delta > 0$ small enough, we can find an $\varepsilon = \varepsilon(\delta) > 0$ such that if $C_x^{'} = \text{x} + \text{C}^{'}$, then $\left(C_x^{'} \setminus \bar{B}(x, \delta)\right) \subseteq C_y$ for all $x \in B(y, \varepsilon)$.

Here is a justification. Assume without loss of generality that $y = 0$ to simplify, and fix $\delta > 0$. Let 0 be an open set in the unit sphere such that $C = \{\lambda z, z \in O, \lambda > 0\}$

There exists $\alpha > 0$ and another open set O' in the unit sphere such that $O' \subseteq O$ and if $z \in \mathbf{S}_{0,1}$ with $d(z, O') \leq \alpha$ then $z \in O$. (For instance consider the intersection of the sphere with two concentric open balls centered at some $z_0 \in O$, and take O' the smaller of the two balls intersected with $\mathbf{S}$). Now, choose $\varepsilon = \delta\alpha / 4.$ *Let* $x \in B(0, \varepsilon)$, and let us show that $(x + C') \setminus \bar{B}(x, \delta) \subseteq C$ where C' is the cone generated by O' originating at 0 (which does not depend on δ) For $y \in O'$, let $z = (x + \lambda y)/r$ where $r = \|x + \lambda y\|$, then $z \in \mathbf{S}_{0,1}$. Moreover, note that by the triangular inequality $|r - \lambda| \leq \varepsilon$. Thus if $r \geq \delta / 2$,

$$\|y - z\| = \left\| y - \frac{1}{r}(\lambda y + x) \right\|$$

$$= \frac{1}{r}\|(r - \lambda)y - x\|$$

$$\leq \frac{2}{\delta}\|(r - \lambda) + \varepsilon\|$$

$$\leq \frac{4\varepsilon}{\delta} \leq \alpha$$

by definition of ε. Hence $z \in O$ and hence $x + \lambda y = rz \in C$. Now, if ε is further chosen such that $\varepsilon \le \delta / 2$, then for all $x \in B(0, \varepsilon)$ and for all $u \in (x + C') \setminus \bar{B}(x, \delta)$, $r = \|u\| \le \delta / 2$ automatically by the triangular inequality, and thus the previous conclusion $n \in C$ holds. We have shown that $(x + C') \setminus \bar{B}(x, \delta) \subseteq C$ as desired.]

Now by (iii) in Proposition (1.1.20), if

$$H_{C'}^{\delta} = \inf\{t > 0 : B_t \in C' \setminus \bar{B}(0, \delta)\}$$

then

$$\mathbf{P}_0\left(H_{C'}^{\delta} < \eta\right) \to \mathbf{P}_0\left(H_{C'} < \eta\right) = 1 \; as \; \delta \downarrow 0.$$

Therefore, for all $\alpha > 0$ there exists $\delta > 0$ such that $\mathbf{P}\left(H_{C'}^{\delta} \le \eta\right) \ge 1 - \alpha$. Choosing $\varepsilon = \varepsilon(\delta)$ associated with this δ, we find that for every $x \in B(y, \varepsilon)$, we have by translation invariance, and letting T_K be the hitting time of a set K,

$$\begin{aligned}\mathbf{P}_x\left(T > \eta\right) &\le \mathbf{P}_x(T_{C_x' \setminus \bar{B}(x,\delta)} > \eta)\\ &= \mathbf{P}_0(H_{C'}^{\delta} > \eta)\\ &\le \alpha \text{ (by our choice of } \delta\text{)}.\end{aligned}$$

This means that $\mathbf{P}_x\left(T > \eta\right) \to 0$, $x \to y$, which proves the lemma.

We can now finish the proof of Theorem (3.3.2) Let $y \in \partial D$. We want to estimate the quantity $\mathbf{E}_x\left[g\left(B_T\right)\right] - g\left(y\right)$ for some $x \in D$. For $\eta, \delta > 0$, let

$$A_{\eta,\delta} = \left\{ \sup_{0 \le t \le \eta} \left|B_t - x\right| \ge \delta / 2 \right\} .$$

This event decreases to $\varnothing$ as $\eta \downarrow 0$ Because B has continuous paths. Now, for any $\delta, \eta > 0$,

$$\mathbf{E}_x\left[\left|g(B_T)-g(y)\right|\right]=\mathbf{E}_x\left[\left|g(B_T)-g(y)\right|;\{T\le\eta\}\cap A^c_{\delta,\eta}\right]$$

$$+\mathbf{E}_x\left[\left|g(B_T)-g(y)\right|;\{T\le\eta\}\cap A_{\delta,\eta}\right]$$

$$+\mathbf{E}_x\left[\left|g(B_T)-g(y)\right|;\{T\ge\eta\}\right]$$

Fix $\varepsilon>0$. We are going to show that each of these three quantities can be made $<\varepsilon/3$ for x close enough to y. Since g is continuous at y, for some $6>0,\left|y-z\right|<\delta$ with $y,z\in\partial D$ implies $\left|g(y)-g(z)\right|<\varepsilon/3$. Moreover, on the event $\{T\le\eta\}\cap A^c_{\delta,\eta}$, we know that $\left|B_T-x\right|<\delta/2$, and thus $\left|B_T-x\right|\le\delta$ as soon as $\left|x-y\right|<\delta/2$. Therefore, for every $\eta>0$, the first quantity is less than $\varepsilon/3$ for $x\in\bar{B}\left(y,\delta/2\right)$.

Next, if M is an upper bound for $\left|g\right|$, the second quantity is bounded by $2M\mathbf{P}\left(A_{\delta,\eta}\right)$. Hence, by now choosing η small enough, this is $<\varepsilon/3$.

Finally, with δ,η fixed as above, the third quantity is bounded by $2M\mathbf{P}_x(T\ge\eta)$. By the previous lemma, this is $<\varepsilon/3$ as soon as $x\in B\left(y,\alpha\right)\cap D$ for some $\alpha>0$. Therefore, for any $x\in B\left(y,\alpha\wedge\delta/2\right)\cap D, \left|u\left(x\right)-g\left(y\right)\right|<\varepsilon$. This entails the result.

Corollary (3.3.7):

A function $u:D\to\mathbf{R}$ is harmonic in D if and only if it is in $C^2\left(D,\mathbf{R}\right)$, and satisfies $\Delta u=0$.

Proof: Let u be of class $C^2\left(D,\mathbf{R}\right)$ and be of zero Laplacian, and let $x\in D$. Let ε be such that $B\left(x,\varepsilon\right)\subseteq D$, and notice that $u|_{\bar{B}(x,\varepsilon)}$ is a solution of the Dirichlet problem on $\bar{B}\left(x,\varepsilon\right)$ with boundary values $u|_{\partial B(x,\varepsilon)}$. Then $B\left(x,\varepsilon\right)$ satisfies the

$l.e.c.c.$, so that $u|_{B(x,\varepsilon)}$ is the unique such solution, which is also given by the harmonic function of Theorem (3.3.2). Therefore, u is harmonic on D.

No we illustrate Girsanov's theorem.

Given a local martingale M, recall the definition of its exponential martingale (Theorem 3.1.1), which will play a crucial role in what follows. Recall that if $M \in \mathcal{M}_{c,loc}$ with $M_0 = 0$, then $Z_t = exp\left(M_t - \frac{1}{2}[M]_t\right)$ defines a continuous local martingale by Itô's formula. Z is the exponential (local) martingale of M (sometimes also called the stochastic exponential of M) and we write $Z = \varepsilon(M)$.

We start by the following inequality which will be useful in the proof of Girsanov's theorem, but is also interesting in its own right.

Proposition (3.3.8): (Exponential martingale inequality)

Let $M \in \mathcal{M}_{c,loc}$ with $M_0 = 0$. Then for all $x > 0, u > 0$,

$$\mathbf{P}\left(\sup_{t\geq 0} M_t > x, [M]_\infty \leq u\right) \leq e^{-x^2} / (2u) \tag{3.7}$$

Proof: $Fix\ x > 0$ and set $T = inf\left\{t \geq 0 : [M]_t > x\right\}$. Fix $\theta \in \mathbf{R}$ and set

$$Z_t = exp\left(\theta M_t^T - \frac{1}{2}\theta^2 [M]_t^T\right). \tag{3.8}$$

Then $Z \in \mathcal{M}_{c,loc}$ and $|Z| \leq e^{\theta x}$. Hence, $Z \in \mathcal{M}_c^2$ and, by OST, $\mathbf{E}(Z_\infty) = \mathbf{E}(Z_0) = 1$. For $u > 0$ we get by Markov's inequality

$$\mathbf{P}\left(\sup_{t\geq 0} M_t > x, [M]_\infty \leq u\right) \leq \mathbf{P}\left(Z_\infty \geq e^{\theta x - \frac{1}{2}\theta^2 u}\right) \leq e^{-\theta x + \frac{1}{2}\theta^2 u} \tag{3.9}$$

Optimizing in θ gives $\theta = x/u$ and the result follows. (It is also possible to use the DubinsSchwarz theorem, as a calculus argument shows that $\mathbf{P}(|Z| > \lambda) \le e^{-\lambda^2/2}$ for all $\lambda \ge 0$, when Z is a standard Gaussian random variable).

Proposition (3.3.9):

Let $M \in \mathrm{M}_{c,loc}$ with $M_0 = 0$ and suppose that $[M]$ is a.s. uniformly bounded. Then $\mathrm{E}(M)$ is a *UI* martingale.

Proof: Let C be such that $[M]_\infty \le C$ a.s. By the exponential martingale inequality, for all $x > 0$

$$\mathbf{P}\left(\sup_{t\ge 0} M_t > x\right) = \mathbf{P}\left(\sup_{t\ge 0} M_t > x, [M]_\infty \le C\right) \le e^{-x^2/(2C)} \qquad (3.10)$$

Now, $\sup_{t\ge 0} \mathrm{E}(M)_t \le \exp\left(\sup_{t\ge 0} M_t\right)$ and

$$\mathbf{E}\left(\exp\left(\sup_{t\ge 0} M_t\right)\right) = \int_0^\infty \mathbf{P}\left(\sup_{t\ge 0} M_t > log\lambda\right) \mathrm{d}\lambda$$

$$\le 1 + \int_0^\infty e^{-(log\lambda)2/(2C)} \mathrm{d}\lambda < \infty \qquad (3.11)$$

Hence, $\mathrm{E}(M)$ is *UI* and, by Proposition (1.2.4), $\mathrm{E}(M)$ is a martingale.

Girsanov's theorem is a result which relates absolute continuous changes of the underlying probability measure $\mathbf{P}$ to changes in the drift of the process. The starting point of the question could be formulated as follows. Suppose we are given realizations of two processes X and Y, where X is a Brownian motion and Y is a Brownian motion with drift b. However, we do not know which is which. Can we tell them apart with probability 1 just by looking at the sample paths? If we get to observe them up to time ∞ then we can, since $\lim_{t\to\infty} Y_t / t = b$ almost surely. However, if we get to observe them only on a finite window, it will not be possible to distinguish them with probability 1: we aye that their law (restricted to $[0,T]$ for any $T>0$) is absolutely continuous with respect to one another. When such is the case, there is a "density" of the law of one process with respective to the other. This density is a random variable which depends on T, and which will turn out to be a certain exponential martingale.

Recall that for two probability measures $\mathbf{P}_1, \mathbf{P}_2$ on a measurable space $(\Omega, \mathcal{F}), \mathbf{P}_1$ is absolutely continuous with respect to $\mathbf{P}_2, \mathbf{P}_1 \ll \mathbf{P}_2$, if

$$\mathbf{P}_2(\mathrm{A})=0 \quad \Rightarrow \mathbf{P}_1(\mathrm{A})=0 \text{ for all } A \in \mathcal{F} \tag{3.12}$$

In this case, by the Radon-Nikodym theorem, there exists a density $f:\Omega \to [0,\infty)$ which is $\mathcal{F}$ -measurable and $\mathbf{P}_2$ unique almost surely (and hence $\mathbf{P}_1$ unique almost surely as well), such that $\mathbf{P}_1 = f\mathbf{P}_2$. That is, for all $A \in \mathcal{F}$,

$$\mathbf{P}_1(\mathrm{A})=\int_\Omega f(\omega) 1_{\{A\}} \mathrm{d}\mathbf{P}_2(\omega)$$

f is also called the Radon-Nikodym derivative, and we denote:

$$\left.\frac{\mathrm{d}\mathbf{P}_1}{\mathrm{d}\mathbf{P}_2}\right|_{\mathcal{F}} = f$$

(Note that in general, the density f depends on the σ-field $\mathcal{F}$).

In order to see where Girsanov's theorem comes from on a simple example where one can compute everything by hand, consider the following. Let $\sigma>0$ and $b\neq 0$, and let $X_t = \sigma B_t + bt$. Then we claim that the law of X is absolutely continuous with respect to the law of Brownian motion $Y_t = \sigma B_t$ with speed σ (but without drift), so long as we restrict ourselves to events of

F_t for some fixed $t>0$. Indeed, if $n\geq 1$ and $0=t_0<t_1<...<t_n=t$ and $x_0=0, x_1,...,x_n\in\mathbf{R}$, then we have:

$$\mathbf{P}(X_{t_1}=x_1,...,X_{t_n}=x_n)=C\exp\left(-\sum_{i=0}^{n-1}\frac{\left(x_{i+1}-x_i-b(t_{i+1}-t_i)\right)^2}{2\sigma^2(t_{i+1}-t_i)}\right)\prod_{i=1}^{n}dx_i$$

where C is a factor depending on $t_1,...,t_n$ and σ, whose value is of no interest to us. It follows that

$$\frac{\mathbf{P}(X_{t_1}=x_1,...,X_{t_n}=x_n)}{\mathbf{P}(Y_{t_1}=x_1,...,Y_{t_n}=x_n)}=e^{-z}$$

Where

$$Z=\sum_{i=0}^{n-1}\frac{\left(x_{i+1}-x_i-b(t_{i+1}-t_i)\right)^2}{2\sigma^2(t_{i+1}-t_i)}-\frac{\left(x_{i+1}-x_i\right)^2}{2\sigma^2(t_{i+1}-t_i)}$$

$$=\sum_{i=0}^{n-1}-\frac{b}{\sigma^2}\left(x_{i+1}-x_i\right)+\frac{1}{2\sigma^2}b^2\left(t_{i+1}-t_i\right)$$

$$\to -\int_0^t\sigma^{-2}bdY_s+\frac{1}{2}\int_0^t b^2\sigma^{-2}ds \text{ as } \quad n\to\infty \ .$$

(We have written the last bit as a convergence although there is an exact equality. This makes it clear that when σ and b depend on the position x – which is precisely what defines the SDE's developed in the next chapter then a similar calculation holds and Girsanov's theorem will hold.) Thus if $\mathbf{Q}$ is the law of X, and $\mathbf{P}$ the law of Y,

$$\left.\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\right|_{F_t}=\exp\left(\int_0^t\frac{b}{\sigma}dB_s-\int_0^t\frac{b^2}{2\sigma^2}ds\right)$$

$$=\mathrm{E}(b\sigma^{-2}Y)_t$$

So we have written the density of X with respect to Y as an exponential martingale.

The point of view of Girsanov's theorem is a slightly different perspective, essentially the other side of the same coin. We will consider changes of measures given by a suitable exponential martingale, and observe the effect on the drift. It is of fundamental importance in mathematical finance (in the context of "risk neutral measures").

Theorem (3.3.10):(Girsanov's Theorem)

Let $M \in \mathcal{M}_{c,loc}$ with $M_0 = 0$. Suppose that $Z = \mathcal{E}(M)$ is a *UI* martingale. We can define a new probability measure $\tilde{\mathbf{P}} \ll \mathbf{P}$, on $(\Omega, \mathcal{F})$ by:

$$\tilde{\mathbf{P}}(A) = \mathbf{E}(Z_\infty 1_A),\ A \in \mathcal{F}. \tag{3.13}$$

Then for every $X \in \mathcal{M}_{c,loc}(\mathbf{P}), X - [X, M] \in \mathcal{M}_{c,loc}(\tilde{\mathbf{P}})$. Moreover the quadratic variation of X under $\mathbf{P}$ and of $X - [X, M]$ under $\tilde{\mathbf{P}}$ are identical $\mathbf{P}$ and $\tilde{\mathbf{P}}$ almost surely.

Proof: Since Z is UI, the limit $Z_\infty = \lim_{t\to\infty} Z_t$, exists $\mathbf{P}$-almost surely, $Z_\infty \geq 0$ and $\mathbf{E}(Z_\infty) = \mathbf{E}(Z_0) = 1.$ thus $\tilde{\mathbf{P}}(\Omega) = 1, \tilde{\mathbf{P}}(\emptyset) = 0$ and countable additively follows by linearity of expectation and the monotone convergence theorem. I$\mathbf{P}(A) = 0$ then $\tilde{\mathbf{P}}(A) = \int_A Z_\infty d\mathbf{P} = 0$, so $\tilde{\mathbf{P}} \ll \mathbf{P}$. Let $X \in \mathcal{M}_{c,loc}$ and set

$$T_n = \inf\{t \geq 0 : |X_t - [X, M]_t| \geq n\}. \tag{3.14}$$

Since $X - [X, M]$ is continuous, $\mathbf{P}(T_n \uparrow \infty) = 1$ will implies $\tilde{\mathbf{P}}(T_n \uparrow \infty) = 1$. So to show that $Y = X - [X, M] \in \mathcal{M}_{c,loc}(\tilde{\mathbf{P}})$, it suffices to show that

$$Y^{T_n} = X^{T_n} - [X^{T_n}, M] \in \mathcal{M}_c(\tilde{\mathbf{P}}) \text{ for all } n \in \mathbf{N} \tag{3.15}$$

Replacing X by X^{T_n}, we reduce to the case where Y is uniformly bounded. By the integration by parts formula and the Kunita-Watanabe identity,

$$d(Z_tY_t) = Y_t dZ_t + Z_t dY_t + d[Z_t, Y_t]$$

$$= Y_t Z_t dM_t + Z_t\left(dX_t - d\left[X_t, M_t\right] + Z_t d\left[M_t, X_t\right]\right)$$

$$= Y_t Z_t dM_t + Z_t dX_t \tag{3.16}$$

and so $ZY \in \mathcal{M}_{c,loc}(\mathbf{P})$. Also $\{Z_T \ : T \text{ is a stopping time}\}$ is UI.

So since Y is bounded, $\{Z_T Y_T \ : T \text{ is a stopping time}\}$ is UI. Hence, $ZY \in \mathcal{M}_c(\mathbf{P})$. But then for $s \le t$, if $A \in \mathcal{F}_s$,

$$\tilde{\mathbf{E}}\left((Y_t - Y_s)1_A\right) = \mathbf{E}\left(Z_\infty (Y_t - Y_s)1_A\right)$$

$$= \mathbf{E}\left[1_A\left(\mathbf{E}\left(Z_\infty Y_t \mid \mathcal{F}_t\right) - \mathbf{E}\left(Z_\infty Y_s \mid \mathcal{F}_s\right)\right)\right]$$

$$= \mathbf{E}\left[1_A\left(Z_t Y_t - Z_s Y_s\right)\right] = 0$$

Since $ZY \in \mathcal{M}_c(\mathbf{P})$. Therefore, $Y \in \mathcal{M}_c\left(\tilde{\mathbf{P}}\right)$ as required. The fact that the quadratic variation $[Y]$ is the same under $\tilde{\mathbf{P}}$ as it comes from the discrete approximation under $\mathbf{P}$:

$$[Y]_t = [X]_t = \lim_{n\to\infty} \sum_{k=0}^{[2^n t]-1} \left(X_{(k+1)2-n} - X_{k2-n}\right)^2$$

$\mathbf{P}$-u.c.p. Thus there exists a subsequence n_k for which the convergence holds $\mathbf{P}$-almost surely uniformly on compacts. Since $\tilde{\mathbf{P}}$ is absolutely continuous with respect to $\mathbf{P}$ this limit also holds $\tilde{\mathbf{P}}$ almost surely for this particular subsequence. Since the whole sequence converges in probability to $[Y]$ in the $\tilde{\mathbf{P}}$-u.c.p. sense (by general theory, since $Y \in \mathcal{M}_{c,loc}\left(\tilde{\mathbf{P}}\right)$, this uniquely identifies the limit, and hence the quadratic variation $[Y]$ under $\tilde{\mathbf{P}}$ has the same value as under $\mathbf{P}$.

Corollary (3.3.11):

Let B be a standard Brownian motion under $\mathbf{P}$ and $M \in \mathcal{M}_{c,loc}$ such that $M_0 = 0$. Suppose $Z = \mathcal{E}(M)$ is a UI martingale and $\tilde{\mathbf{P}}(A) = \mathbf{E}(Z_\infty 1_A)$ for all $A \in \mathcal{F}$. Then $\tilde{B} := B - [B, M]$ is a $\tilde{\mathbf{P}}$-Brownian motion.

Proof: Since $\tilde{B} \in \mathcal{M}_{c,loc}(\tilde{\mathbf{P}})$ by Theorem (3.3.10) and has $[\tilde{B}]_t = [B]_t = t$, by Levy's characterization, it is a Brownian motion.

Remark (3.3.12):

This corollary should be read backward if X is a Brownian motion, then changing the measure by the exponential martingale $\mathcal{E}(M)$, $X = \tilde{X} + [X, M]$ where $\tilde{X}$ is a Brownian motion under the new measure. So the old process (which was just Brownian motion) becomes under the new measure a Brownian motion plus a "drift" term given by the conversatio $[X, M]$.

Let $(\mathcal{W}, \mathcal{W}, \mathbf{W})$ be the Wiener space. (Recall that $W = C([0,\infty), \mathbf{R}), \mathcal{W} = \sigma(X_t : \geq 0)$ where $X_t : W \to \mathbf{R}$ with $X_t(w) = w(t)$. The Wiener measure $\mathbf{W}$ is the unique probability measure on $(\mathcal{W}, \mathcal{W})$ such that $(X_t)_{t \geq 0}$ is a Brownian motion started from 0.)

Definition (3.3.13):

Define the Cameron-Martin space

$$H = \left\{ h \in W : h(t) = \int_0^t \varnothing(s)\ ds \text{ for some } \varnothing \in L^2([0,\infty)) \right\} . \qquad (3.17)$$

For $h \in H$, write $h = \varnothing$ the derivative of h.

Theorem (3.3.14): (Girsanov, Cameron-Martin Theorem)

Fix $h \in H$ and set $\mathbf{W}^h$ to be the law on $(\mathrm{W}, \mathcal{W})$ of $(B_t + h(t), t \geq 0)$ where B_t is a Brownian motion: that is, for all $A \in \mathcal{W}$,

$$\mathbf{W}^h(A) = \mathbf{W}(\{w \in W : w + h \in A\}). \tag{3.18}$$

Then $\mathbf{W}^h$ is a probability measure on $(\mathrm{W}, \mathcal{W})$ and $\mathbf{W}^h \ll \mathbf{W}$ with Radon-Nikodym density

$$\left.\frac{\mathrm{d}\mathbf{W}^h}{\mathrm{d}\mathbf{W}}\right|_{\mathcal{W}} = \exp\left(\int_0^\infty h(s)dX_s - \frac{1}{2}\int_0^\infty h(s)^2 \mathrm{ds}\right). \tag{3.19}$$

Remark (3.3.15):

So if we take a Brownian motion and shift it by a deterministic function $h \in H$ then the resulting process has a law which is absolutely continuous with respect to that of the original Brownian motion.

Proof: Set $\mathcal{W}_t = \sigma(X_s, s \leq t)$ and $M_t = \int_0^t \varnothing(s) dX_s$. Then $M \in \mathcal{M}_c^2(W, \mathcal{W}, (\mathcal{W})_{t\geq 0}, \mathbf{W})$ and

$$[M]_\infty = \int_0^\infty \varnothing(s)^2 \, ds =: C < \infty. \tag{3.20}$$

By Proposition (3.3.9), $\mathcal{E}(M)$ is a UI martingale, so we can define a new probability measure $\mathrm{P} \ll \mathbf{W}$ on $(\mathrm{W}, \mathcal{W})$ by

$$\frac{\mathrm{dP}}{\mathrm{d}\mathbf{W}}(w) = \exp\left(M_\infty(w) - \frac{1}{2}[M]_\infty(w)\right) = \exp\left(\int_0^\infty \varnothing(s) dX_s(w) - \frac{1}{2}\int_0^\infty \varnothing(s)^2 \mathrm{ds}\right). \tag{3.21}$$

and $\tilde{X} = X - [X, M] \in \mathcal{M}_{c.loc}(\mathrm{P})$ by Girsanov's theorem. Since X is a $\mathbf{W}$-Brownian motion, by Corollary (3.3.11) $\tilde{X}$ is a P-Brownian motion. But by the Kunita-Watanabe identity,

$$[X,M]_t=[X,\varnothing.X]_t$$

$$=\varnothing.[X,X]_t$$

$$=\int_0^t \varnothing(s)ds=h(t)$$

hence we get that $\tilde{X}(\omega)=X(\omega)-h=\omega-h$. Hence, under P, $X=\tilde{X}+h$, where $\tilde{X}$ is a P -Brownian motion. Therefore, $\mathbf{W}^h=\mathrm{P}$ on $\mathrm{F}_\infty=\mathrm{W}$:

$$\mathrm{P}\left(A\right)=\mathrm{P}\left(\{\omega:\omega\in A\}\right)=\mathrm{P}\left(\{\omega:\tilde{X}\left(\omega\right)+h\in A\}\right)$$

$$=\mathrm{W}\left(\{\omega:\omega+h\in A\}\right)=\mathrm{W}^h\left(A\right)$$

as required.

One of the most important applications of Girsanov's theorem is to the study of Brownian motion with constant drift. Indeed, applying the previous result with $\phi(s)=1_{\{s\le t\}}$ gives us the following corollary.

Corollary (3.3.16):

Let $\gamma\neq 0$ and let W^γ be the law of $\left(\mathrm{X}_t+\gamma t,\ t\geq 0\right)$ under W. *Then for all* $t>0$, and for any $\mathrm{A}\in\mathrm{F}_t$,

$$\mathrm{W}^\gamma\left(A\right)=\mathrm{E}_{\mathrm{W}}\left(1_A\exp\left(\gamma X_t-\frac{1}{2}\gamma^2 t\right)\right) \qquad (3.22)$$

This allows us to compute functionals of Brownian motion with drift in terms of Brownian motion without drift - a very powerful technique.

Chapter Four

Stochastic Differential Equations

Suppose we have a differential equation, say $\frac{dx(t)}{dt} = b(x(t))$ or, in integral form,

$$x(t) = x(0) + \int_0^t b(x(s))ds \tag{4.1}$$

Which describes a system evolving in time, be it the growth of a population, the trajectory of a moving object or the price of an asset. Taking into account random perturbations, it may be more realistic to add a noise term:

$$X_t = X_0 + \int_0^t b(X_s)ds + \sigma B_t \tag{4.2}$$

Where B is a Brownian motion and σ is a constant controlling the intensity of the noise. It may be that this intensity depends on the state of the system, in which case we have to consider an equation of the form:

$$X_t = X_0 + \int_0^t b(x_s)ds + \int_0^t \sigma(x_s)dB_s \tag{4.3}$$

Where the last term is, of course, an Itô integral. Equation (4.3) is a stochastic differential equation and may also be written

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t \tag{4.4}$$

Section (4.1): Lipschitz Coefficients, Strong Markov Property and Definitions Processes

We start by general definitions.

Let B be a Brownian motion in R^m with $m \geq 1$. Let $d \geq 1$ and suppose

$$\sigma(x) = (\sigma_{ij}(x))_{\substack{1\leq i\leq d\\ 1\leq j\leq m}} : \mathbf{R}^d \to \mathbf{R}^{d\times m}$$

and

$$b(x) = (b_i(x))_{1\leq i\leq d} : \mathbf{R}^d \to \mathbf{R}^d$$

are given Borel functions, bounded on compact sets. Consider the equation in $\mathbf{R}^d$:

$$dX_t = \sigma(X_t)\,dB_t + b(X_t)dt \tag{4.5}$$

which may be written component wise as

$$dX_t^i = \sum_{j=1}^{m} \sigma_{ij}(X_t)dB_t^j + b_j(X_t)dt \quad , 1 \leq i \leq d \tag{4.6}$$

This general SDE will be called $E(\sigma,b)$. A solution to $E(\sigma,b)$ in Equation (4.5) consists of:

- a filtered probability space $(\Omega, F\,(F_t)_{t\geq 0}, P)$ satisfying the usual conditions;
- an $(F_t)_{t\geq 0}$ -Brownian motion $B=(B^1,\ldots,B^m)$ taking values in $\mathbf{R}^m$;
- an $(F_t)_{t\geq 0}$ -adapted continuous process $X=(X^1,\ldots,X^d) \in \mathbf{R}^d$ such that

$$X_t = X_0 + \int_0^t \sigma(X_s)dB_s + \int_0^t b(X_s)ds \tag{4.7}$$

When, in addition, $X_0 = x \in \mathbf{R}^d$, we say that X is a solution started from x.

Definition (4.1.1) :

Let $E(\sigma,b)$ be the SDE in (4.5).

- We say that $E(\sigma,b)$ has a solution if for all $x \in \mathrm{R}^d$, there exists a solution to the SDE started from x.
- There is uniqueness in law if all solutions to $E(\sigma,b)$ started from x have the same distribution.
- There is pathwise uniqueness if, when we fix $(\Omega, \mathrm{F}(\mathrm{F}_t)_{t\geq 0}, \mathrm{P})$ and B then any two solutions X and X' satisfying $X_0 = X'_0$ a:s: are indistinguishable from one another.
- We say that a solution X of $E(\sigma,b)$ started from x is a strong solution if X is adapted to the natural filtration of B.

Remark (4.1.2):

In general, $\sigma(B_s, s \leq t) \subseteq \mathrm{F}_t$ and a solution might not be measurable with respect to the Brownian motion B. A strong solution depends only on $x \in \mathrm{R}^{\mathrm{d}}$ and the Brownian motion B, and is moreover non-anticipative: if the path of B is known up to time t, then so is the path of X up to time t. We will term weak any solution that is not strong.

Remark (4.1.3):

If every solution is strong, then pathwise uniqueness holds. Indeed, any solution must then be a certain measurable functional of the path B. If two functionals F_1 and F_2 of B gave two solutions to the SDE, then we would construct a third one by tossing a coin and choosing X_1 or X_2. This third solution would then not be adapted to F^B.

Example (4.1.4):

It is possible to have existence of a weak solution and uniqueness in law without pathwise uniqueness. Suppose β is a Brownian motion in $\mathbb{R}$ with $\beta_0 = x$. Set

$$B_t = \int_0^t \operatorname{sgn}(\beta_s)\, d\beta_s \quad \text{where } sgn\ (x) = \begin{cases} -1 \text{ if } x < 0 \\ 1 \text{ if } x > 0 \end{cases} \tag{4.8}$$

Since sgn is left-continuous, $(sgn\,(\beta_t))_{t\geq 0}$ is previsible, so that the Itô integral is well defined and $B \in M_{c,loc}$. By Levy's characterization, B is a Brownian motion started from 0, since $[B]_t = [\beta]_t = t$. It is also true that

$$\beta_t = x + \int_0^t \operatorname{sgn}(\beta_s)\, dB_s \tag{4.9}$$

(Indeed, by definition $dB_s = sgn(\beta_s)dB_s$ so multiplying by $sgn(\beta_s)$ yields, by the stochastic chain rule, $sgn(\beta_s)dB_s = d\beta_s$ Hence, β is a solution to the SDE

$$dX_t = \operatorname{sgn}(X_t)dB_t,\ X_0 = x \tag{4.10}$$

Thus Equation (4.10) has a weak solution. Applying Levy's characterization again, it is clear that any solution must be a Brownian motion and so there is uniqueness in law. On the other hand, path wise uniqueness does not hold: Suppose that β is a solution to Equation (4.10) with $\beta_0 = 0$. Then both β and $-\beta$ are solutions to Equation (4.10) started from 0. Indeed, we may write

$$-\beta_t = -\int_0^t \operatorname{sgn}(\beta_s)\, dB_s = \int_0^t \operatorname{sgn}(-\beta_s)\, dB_s + 2\int_0^t 1_{\{\beta_s = 0\}}\, dB_s$$

The second term is a local martingale since it is an integral with respect to B . The quadratic variation of this local martingale is $4\int_0^t 1_{\{\beta_s=0\}} ds$ which is 0 almost surely by Fubini's theorem (since β must be a Brownian motion by Lévy's characterization). Hence this local martingale is indistinguishable from 0, and $-\beta$ is a solution to Equation (4.10).

It also turns out that β is not a strong solution to (4.10).

Theorem (4.1.5): (Yamada-Watanabe)

Let σ, b be measurable functions. If path wise uniqueness holds for $E(\sigma, b)$ and there exist solutions, then there is also uniqueness in law. In this case, for every filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}; \mathrm{P})$ and every $\mathcal{F}_t$-Brownian motion $B = (B-t, t \geq 0)$, and for every $x \in \mathrm{R}^d$, the unique solution X to $E_x(\sigma, b)$ is strong.

In particular path wise uniqueness is stronger than weak uniqueness, provided that there exist solutions.

Now we study Lipschitz coefficients.

For $U \subseteq \mathbb{R}^d$ and $f : U \to \mathbb{R}^d$, say f is Lipschitz with Lipschitz constant $K < \infty$ if

$$|f(x) - f(y)| \leq K|x - y| \quad \textit{for all} \quad x, y \in U, \tag{4.11}$$

Where $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^d$. (If $f : U \to \mathbb{R}^{d\times m}$ then the left-hand side is the Euclidean norm in $\mathbb{R}^{d\times m}$). The key result of this part of section will be that SDE with Lipschitz coefficients have path wise unique solutions which are furthermore always strong.

We start preparing for this result by recalling two important results which will be used in the proof.

Theorem (4.1.6): (Contraction Mapping Theorem)

Let (X,d) be a complete metric space and $F: X \to X$. Suppose that the iterated function F^n is a contraction for some $n \in \mathbb{N}$, i.e.

$$\exists_{r<1} \forall x, y \in X : d(F^n(x), F^n(y)) \le rd(x, y). \tag{4.12}$$

Then *F* has a unique fixed point.

Remark (4.1.7):

This theorem is most well-known when *F* itself is contractive, rather than F^n . However, the theorem for $n \ge 1$ easily follows from the $n = 1$ result. Indeed, if $n \ge 1$ and F^n is contractive, then (by the theorem for $n = 1$) F^n must have a unique fixed point *x*. We claim that *x* is also a fixed point of *F*. Indeed, let $x_1 = F(x),\ x_2 = F^2(x), \dots,\ x_{n-1} = F^{n-1}(x)$. Then since $F^n(x) = x,$ we have

$$F^n(x_1) = F(F^{n-1}(x_1))$$

$$= F(F^n(x))$$

$$= F(x) = x_1$$

so x_1 is a fixed point of F^n as well. But F^n has a unique fixed point, so $x = x_1$. Therefore, $F(x) = x_1 = x$ and x_1 is a fixed point of *F*.

Lemma (4.1.8): (Gronwall's Lemma):

Let $T > 0$ and let *f* be a non-negative bounded measurable function on $[0,T]$. Suppose that for some $a, b \ge 0$:

$$f(t) \le a + b\int_0^t f(s)\, ds \quad 0 \le t \le T. \tag{4.13}$$

Then $f(t) \le a\ exp(bt)$ for all $t \in [0,T]$. In particular if $a = 0$ then $f = 0$.

Proof: The proof uses a trick which is close to what we will do in the proof of the next theorem. The idea is to iterate the inequality Equation (4.13). We get:

$$f(t) \le a + b\int_0^t a + b\int_0^s f(u)duds$$

$$= a + abt + b^2 \int \int_{0\le t_1 \le t_2} f(t_1)dt_1 dt_2$$

$$\le \ldots \le a + abt + a\frac{b^2t^2}{2!} + \ldots + a\frac{b^n t^n}{n!} + b^{n+1}\int_{0\le t_1\le \ldots \le t_{n+1}\le t} f(t_1)dt_1 \ldots dt_{n+1}$$

where the term $b^n t^n / n!$ comes from the fact that $\int_{0\le t_1 \ldots t_{n+1}\le t} dt_1 \ldots dt_n = t^n / n!$, since the volume of the cube is t^n and the ordering $t_1 \le \ldots t_n$ is one of $n!$possible ordering of the variables, with each ordering contributing the same fraction to the total volume. This argument shows that the last term in the right-hand side of the inequality tends to 0 (since f is bounded). We recognize the Taylor expansion of the exponential function in all the other terms when $n \to \infty$. Thus $f(t) \le ae^{bt}$.

Theorem (4.1.9):

Suppose that $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ and $b : \mathbb{R}^d \to \mathbb{R}^d$ are Lipschitz. Then there is

pathwise uniqueness for the *SDE*

$$dX_t = \sigma(X_t)dB_t + b(X_t)dt. \tag{4.14}$$

Moreover, for each filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}; \mathbb{P})$and each $(\mathcal{F}_t)_{t\ge 0}$-Brownian motion B, there exists a strong solution to the *SDE* started from x, for any $x \in \mathbb{R}$.

Proof: (for $d = m = 1$). Fix $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}; \mathrm{P}\right)$ and B. Let $\left(\mathcal{F}_t^{B}\right)_{t\leq 0}$ be the natural filtration generated by B so that $\mathcal{F}_t^{B} \subseteq \mathcal{F}_t$ for all $t \geq 0$. Suppose that K is the Lipschitz constant for σ and b.

Pathwise uniqueness:

Suppose X and X' are two solutions on $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}; \mathrm{P}\right)$ with

$X_0 = X_0^{'}$ a.s.. Fix M and let

$$\mathrm{T} = \inf\left\{t \geq 0 : \left|x_t\right| \vee \left|x_t^{'}\right| \geq M\right\}. \tag{4.15}$$

Then $X_{t\wedge\mathrm{T}} = X_0 + \int_0^{t\wedge\mathrm{T}} \sigma(X_s)dB_s + \int_0^{t\wedge\mathrm{T}} b(X_s)ds,$ and similarly for $X^{`}$.

Let $T > 0$. If $0 \leq t \leq T$ then, since

$$(x+y)^2 \leq 2(x^2+y^2) \tag{4.16}$$

for all $x, y \in \mathbb{R}$, we have:

$$\mathrm{E}\left(\left(X_{t\wedge\mathrm{T}} - X_{t\wedge\mathrm{T}}^{'}\right)^2\right)$$

$$\leq 2\mathrm{E}\left(\left(\int_0^{t\wedge\mathrm{T}}\left(\sigma(X_s)-\sigma(X_s^{'})\right)dB_s\right)^2\right) + 2\mathrm{E}\left(\left(\int_0^{t\wedge\mathrm{T}}\left(b(X_s)-b(X_s^{'})\right)ds\right)^2\right)$$

$$\leq 2\mathrm{E}\left(\int_0^{t\wedge\mathrm{T}}\left(\sigma(X_s)-\sigma(X_s^{'})\right)^2 ds\right) + 2T\mathrm{E}\left(\int_0^{t\wedge\mathrm{T}}\left(b(X_s)-b(X_s^{'})^2\right)ds\right)$$

(by the Itô isometry and the Cauchy-Schwarz inequality).

$\leq 2K^2(1+T)\mathrm{E}\left(\int_0^{t\wedge\mathrm{T}}\left(X_s - X_s^{`}\right)^2 ds\right)$ (by the Lipschitz property)

$$\leq 2K^2(1+T)\int_0^t \mathrm{E}\left(\left(X_{s\wedge t} - X_{s\wedge\mathrm{T}}^{`}\right)^2\right)ds. \tag{4.17}$$

Let $f(t) = \mathrm{E}\left(\left(X_{s\wedge T} - X^{`}_{s\wedge T}\right)^2\right)$. Then $f(t)$ is bounded by $4M^2$ and

$$f(t) \le 2K^2(1+T)\int_0^t f(s)ds. \tag{4.18}$$

Hence, by Gronwall's lemma, $f(t)=0$ for all $t\in[0,T]$. So $X_{t\wedge T} = X^{`}_{t\wedge T}$

a.s. and, letting $M,T\to\infty$, we obtain that X and X`are indistinguishable.

Existence of a strong solution:

We start by constructing a weak solution as a fixed point of a certain mapping. Let Fix $(\Omega,\mathcal{F},(\mathcal{F}_t);\mathrm{P})$ be a filtered probability space and let B be a Brownian motion. Write $\mathcal{C}_T$ for the set of continuous processes $X:[0,T]\to\mathbb{R}$ adapted to $(\mathcal{F}_t)$ such that

$$|||X|||_T := \left\|\sup_{t\le T}|X_t|\right\|_2 < \infty. \tag{4.19}$$

and $\mathcal{C}$ for the set of continuous adapted processes $X:[0,\infty)\to\mathbb{R}$ such that

$$|||X|||_T < \infty \text{ for all } T>0 . \tag{4.20}$$

Recall from Proposition (2.1.8) that $\left(\mathcal{C}_T, |||.|||_T\right)$ is complete. Let $\mathcal{C}^{'}_T = \mathcal{C}_T \cap \{X_0 = x\}$, and let Φ be a mapping defined on $\mathcal{C}^{'}_T$ by;

$$\Phi(X)_t = x + \int_0^t \sigma(X_s)dB_s + \int_0^t b(X_s)ds \text{ for all } t\le T . \tag{4.21}$$

Note that a solution to $E(\sigma,b)$ is a fixed point of Φ. We start by showing that if $X\in\mathcal{C}^{'}_T$, then so is $\Phi(X)$. For all $y\in\mathbb{R}$,

$$|\sigma(y)| \le |\sigma(0)| + K|y|,\ |b(y)| \le |b(0)| + K|y|. \tag{4.22}$$

Suppose $X \in C_T$ for some T. Let $M_t = \int_0^t \sigma(X_s)dB_s, 0 \le t \le T$. then $[M]_T = \int_0^T \sigma(X_s)ds$ and by Equation (4.16)

$$\mathrm{E}\left([M]_T\right) \le 2T\left(|\sigma(0)|^2 + K^2 \|\|X\|\|_T^2\right) < \infty \ . \tag{4.23}$$

Hence, by Theorem (2.1.15), $(M_t)_{0\le t\le T}$ is a martingale bounded in L^2. So by Doob's L^2 inequality and Equation (4.23):

$$E\left(\sup_{t\le T}\left|\int_0^t \sigma(X_s)dB_s\right|^2\right) \le 8T\left(|\sigma(0)|^2 + K^2 \|\|X\|\|_T^2\right) < \infty \ . \tag{4.24}$$

Therefore $(M_t, t \ge 0)$ belongs to C_T. Similarly, by Equations (4.22), (4.16) and the Cauchy-Schwarz inequality:

$$E\left(\sup_{t\le T}\left|\int_0^t \sigma(X_s)dB_s\right|^2\right) \le TE\left(\int_0^T |b(X_s)|^2 \, ds\right)$$

$$\le 2T^2\left(|b(0)|^2 + K^2 \|\|X\|\|_T^2\right) < \infty. \tag{4.25}$$

Therefore, $\left(\int_0^t b(X_s)\,ds, t \ge 0\right)$ belongs to C_T as well. By the triangular inequality, it follows that $\Phi(X) \in C_T$ and thus $\Phi(X) \in C_T^{'}$ since by definition $\Phi(X)_0 = x$. Now, let $X, Y \in C_T^{'}$. By Doob's inequality again and Equation (4.17),

$$\|\|\Phi(X) - \Phi(Y)\|\|_T^2 = E\left(\sup_{0\le t\le T}|\Phi(X)_t - \Phi(Y)_t|^2\right) \tag{4.26}$$

$$\le 2E\left(\sup_{0\le t\le T}\left|\int_0^t \sigma(X_s)dB_s - \int_0^t \sigma(Y_s)dB_s\right|^2\right) \tag{4.27}$$

$$+2E\left(\sup_{0\le t\le T}\left|\int_0^t b(X_s)ds - \int_0^t b(Y_s)ds\right|^2\right) \tag{4.28}$$

$$\le 2K^2(4+T)\int_0^t E\left(|X_t - Y_t|^2\right)dt$$

$$\le \underbrace{2K^2(4+T)}_{C_T}\int_0^t \left(\|\|X - Y\|\|_t^2\right)dt \tag{4.29}$$

By induction using Equation (4.29), we have for all $n \geq 0$ that

$$\left\|\Phi^n(X)-\Phi^n(Y)\right\|_T^2 \leq C_T^n \int \dots \int \left\|X-Y\right\|_{t_n}^2 1_{\{0\leq t_n \leq \dots \leq t_1 \leq T\}} dt_1 \dots dt_n$$

$$=\frac{C_T^n T^n}{n!}\left\|X-Y\right\|_t^2 \text{ by symmetry (see Gronwall's lemma)} \tag{4.30}$$

For n sufficiently large, Φ^n is a contraction on the complete metric space $\left(\mathcal{C}_T^{'}, \left\|\cdot\right\|_T\right)$. Hence, by the Contraction Mapping Theorem, Φ has a unique fixed point which we may call $X^{(T)} \in \mathcal{C}_T^{'}$.

By uniqueness of this fixed point, $X_t^{(T)} = X_t^{(T')}$ for all $t \leq T \wedge T'$ a.s. and so we may consistently define $X \in \mathcal{C}$ by

$$X_t = X_t^{(N)} \text{ for } t \leq N, N \in \mathbb{N} . \tag{4.31}$$

This is the pathwise unique solution to the SDE started from x. It remains to prove that it is $\left(\mathcal{F}_t^B\right)_{t\geq 0}$-adapted. Define a sequence $\left(Y^n\right)_{n\geq 0}$ in $\mathcal{C}_T$ by

$$Y^0 \equiv x,\ Y^n = \Phi\left(Y^{n-1}\right) \text{ for } n \geq 1. \tag{4.32}$$

Then Y^n is $\left(\mathcal{F}_t^B\right)_{t\geq 0}$-adapted for each $n \geq 0$. Since $X = \Phi^n(X)$ for all $n \geq 0$. by Equation (4.30) we have

$$\left\|X-Y^n\right\|_T^2 \leq \frac{C_T^n T^n}{n!}\left\|X-x\right\|_T^2 . \tag{4.33}$$

Hence, $Y^n \to X$ in $\mathcal{C}_T$ and thus $Y_t^n \to X_t$ in probability for a fixed $t > 0$. Thus there exists a subsequence n_k such that $Y_t^{n_k} \to X_t$ almost surely. Since $Y_t^{n_k}$ is $\mathcal{F}_t^B$-measurable, then so is X_t. Therefore X is $\left(\mathcal{F}_t^B\right)_{t\geq 0}$-adapted and the proof of this theorem is finished.

Remark (4.1.10):

The uniqueness of the fixed point in the contraction mapping theorem cannot be invoked directly to prove path wise uniqueness of the solutions. What this result give is path wise uniqueness of solutions in C_T for any $T > 0$. So if we knew a priori that any solution belongs to C_T, we could invoke this result. (Note that our proof that $\Phi(x) \in C_T$ relies on the fact that X is already assumed to be in C_T). Thus a byproduct of our proof is that indeed any solution belongs to C_T for any $T > 0$.

Corollary (4.1.11):

Let σ_{ij}, b_i be Lipschitz functions on $\mathbb{R}^d$ for $1 \le i, j \le d$. Then every solution to $E_x(\sigma, b)$ is strong, and there is uniqueness in distribution for the solutions to E(σ, b).

Proof: The proof of the theorem constructs a strong solution for every filtered probability space and Brownian motion defined on it. On the other hand there is path wise uniqueness of solutions so any solution must be strong. By the Yamada-Watanabe theorem, it also follow from existence of solutions and path wise uniqueness that uniqueness in distribution holds.

Example (4.1.12): (Ornstein-Hollenbeck Process)

Fix $\lambda \in R$ and consider the SDE in $\mathbb{R}^2$

$$dV_t = dB_t - \lambda V_t dt,\ V_0 = v_0,\ dx_t = V_t dt,\ X_0 = x_0 \tag{4.34}$$

When $\lambda > 0$ this models the motion of a pollen grain on the surface of a liquid, and λ then represents the viscosity of that liquid. x represents the x-coordinate of the grain's position and V represents its velocity in the x-direction. $-\lambda V$ is the friction force due to viscosity.

Whenever $|V|$ becomes large, the system acts to reduce it. (This is a much more realistic model of random motion from a physical point of view than Brownian motion which oscillates too wildly!) V is called the Ornstein-Uhlenbeck (velocity) process. Then there is path wise uniqueness for this SDE. In fact, this is a rare example of a SDE we can solve explicitly.

Remark (4.1.13):

If σ and b are only defined on a closed set k, then there is strong existence and path wise uniqueness at least up until the time $\mathrm{T} = \inf\left\{t \geq 0, X_t \in k^c\right\}$.

Now we discuss strong Markov property and diffusion processes.

In an ordinary differential equation, the future of the trajectory of a particle is entirely determined by its present position. The stochastic analogue for stochastic differential equations is true as well: solutions to SDE's have the strong Markov property, i.e., the distribution of their future depends only on their present position. (In fact, SDE solutions should be viewed as the prototypical example of a strong Markov process.)

Theorem (4.1.14): (Strong Markov Property)

Assume that σ and b are two Lipschitz functions.

Then for all $x \in \mathbb{R}^d$, if X^x denotes a weak solution started from x to $E(\sigma, b)$, if F is any measurable nonnegative functional on $C\left([0,\infty],\mathbb{R}^d\right)$ then almost surely, for any stopping time T:

$$\mathrm{E}\left[F\left(X^x_{T+t}, t \geq 0\right) | \mathcal{F}_T\right] = \mathrm{E}\left[F\left(X^y_t, t \geq 0\right)\right]_{y=X_T} \qquad (4.35)$$

on the event $\{T < \infty\}$.

Proof. By considering $T\Lambda n$, it suffices to consider the case where $T<\infty$ a.s.. As we will see, the strong Markov property is a relatively straightforward consequence of Corollary (4.1.11).

Let $Y_t = X^x_{T+t}$. Since X is a solution to $E_x(\sigma,b)$, we have

$$X^x_{T+t} - X^x_T = \int_T^{T+t} \sigma\left(X^x_s\right).dB_s + \int_T^{T+t} b\left(X^x_s\right).ds \tag{4.36}$$

To make the change of variable $u=t+T$, we use the following Lemma:

Lemma (4.1.15):

Let H be a previsible locally bounded process, and let X be a continuous local martingale. If T is a stopping time and $X^{(T)} = \left(X_{t+T} - X_T, t \geq 0\right)$ then

$$\int_T^{T+t} H_S dX_S = \int_0^t H_{T+u} dX^{(T)}_u$$

Proof: Only the case where X is a local martingale needs to be discussed. The statement is trivial for processes of the form $H = 1_{\{Ax(s,t)\}}$ where $A \in \mathrm{F}_s$ and the general result follows by linearity and the *Itô* isometry when $M \in \mathrm{M}^2_c$ and $H \in L^2(M)$. Finally the general result follows by localization.

Thus, if $y = X_T$, then making the change of variable in Equation (4.36) we get:

$$Y_t = y + \int_0^t \sigma(y_u).dB^{(T)}_u + \int_0^t b(y_u).du$$

where $B^{(T)}_t = B_{T+t} - B_T$ is a Brownian motion independent from F_t. Y is adapted to the filtration $(\mathrm{F}_{t+u}, u \geq 0)$ which satisfies the usual conditions. Therefore, the previous theorem applies and Y is adapted to $(\mathcal{G}_t)_{t\geq 0}$, where for all t > 0, $\mathcal{G}_t$ is the σ-field generated by X_T and $B_S^{(T)}$, $S \le t$. Thus, we can write $(Y_t)_{t\ge 0}$ as a certain deterministic and measurable functional Φ of its starting point X_T and the driving Brownian motion, $\Phi\left(X_T, B^{(T)}\right)$. Furthermore, note that by definition $\Phi(y,B)$ is the unique solution to $E_y(\sigma,b)$ corresponding to the driving Brownian motion B. Hence (by weak uniqueness) $\Phi(y,B)$ has the same law as X^Y. Since $B^{(T)}$ is independent from $\mathcal{F}_T$, it is independent from X_T (because X is adapted to $\mathcal{F}$). It follows that the left-hand side of (4.35) may be computed as:

$$E\left[F(Y_t, t \ge 0) \mid \mathcal{F}_T\right] = E\left[F\left(\Phi\left(X_T, B^{(T)}\right)\right) \mid \mathcal{F}_T\right]$$

$$= E\left[F\left(\Phi\left(y, B^{(T)}\right)\right)\right]|_{y=X_T} \ a.s$$

$$= E\left[F\left(X_t^y, t \ge 0\right)\right]|_{y=X_T} \ a.s$$

Which is exactly the content of the Strong Markov Property.

In the remainder of this research we now provide a brief introduction to the theory of diffusion processes, which are Markov processes characterized by martingale properties. We first construct these processes with SDE's and then move on to describe some fundamental connection with PDE's. In the later we show how diffusions arise as scaling limits of Markov chains.

$$Lf(x) = \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(x)\frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b_i(x)\frac{\partial f}{\partial x_i} \tag{4.37}$$

where $a_{i,j}(x)$ is a measurable function called the diffusivity and $b(x)$, another measurable function, is called the drift. We assume that $\left(a_{i,j}(x)\right)_{i,j}$ is a symmetric nonnegative matrix for all $x \in \mathbb{R}^d$.

Definition (4.1.16):

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ be a filtered probability space satisfying the usual conditions. Say that a process $X = (X_t, t \geq 0)$ is an L-diffusion (or diffusion generated by L) if for all $f \in C_b^2(\mathbb{R}^d)$, the process M^f is a local martingale, where for all $t \geq 0$:

$$M_t^f = f(x_t) - f(x_0) - \int_0^t Lf(X_s).ds \tag{4.38}$$

For the moment, we don't know whether such processes exist, and we haven't shown any sort of uniqueness. The following result takes care of the existence part.

Theorem (4.1.17):

Let be a solution (in $\mathbb{R}^d$) to the SDE

$$dX_t = \sigma(x_t)dB_t + b(x_x).dt$$

where B is a $(\mathcal{F}_t)$ -Brownian motion in $\mathbb{R}^d$ and where $\sigma = (\sigma_{i,j}(x))_{1 \leq i,j \leq d}$ and $b = (b_i(x))_{1 \leq i \leq d}$ are measurable. Then for all $f \in C^{1,2}(R_+ \times R^d)$,

$$M_t^f - f(t, x_t) - f(0, X_0) - \int_0^t \left(\frac{\partial f}{\partial t} + Lf \right)(s, X_s).ds \tag{4.39}$$

is a local martingale, where X has the form Equation (4.37) and $a = \sigma.\sigma^T$. In particular, if the coefficients σ, b are bounded, then X is an L-diffusion.

This results follows simply from an application of $It\hat{o}$,s formula.

Remarks (4.1.18):

(1) If $a_{i,j}$ is uniformly positive definite (that is, there exists $\zeta > 0$ such that

$$\langle A\zeta,\zeta\rangle=\sum_{i,j=1}^{d}\zeta_i a_{i,j}(x)\zeta_j\geq\zeta\|\varepsilon\|$$

for all $\zeta\in\mathbb{R}^d$ and all $x\in\mathbb{R}^d$), then a has a positive-definite square root matrix σ. If a is furthermore Lipschitz, then it can be shown that $\sigma(x)$ is also Lipschitz. It follows that if a,b are bounded Lipschitz functions and a is uniformly positive definite, then L-diffusions exist, by Theorem (4.1.9), for any given starting point X_0.

(2) Brownian motion in $\mathbb{R}^d$ is an L-diffusion for $L=\frac{1}{2}\Delta$.

(3) In the language of Markov Processes, we say that L is the infinitesimal generator of X. Intuitively, $Lf(x)$ describes the infinitesimal expected change in $f(X)$ given that $X_t=x$. That is,

$$\lim_{\varepsilon\to 0}\mathrm{E}\left(\frac{f(X_{t+\varepsilon})-f(X_t)}{\varepsilon}\middle|\mathrm{F}_t,X_t=x\right)=Lf(x)$$

For every $f\in C_b^2(\mathbb{R}^d)$.

Now we Illustrate Some links with PDEs.

Theorem (4.1.19):

Let D be an open set in $\mathbb{R}^d$. Let L be defined by Equation (4.37) for uniformly positive definite Lipschitz bounded coefficients a,b. Let $g\in C(\partial D)$ and let $\phi\in C(\overline{D})$ such that both ϕ and g are bounded. Define:

$$u(x)=\mathrm{E}_x\left(\int_0^T\phi(X_s)ds+g(X_T)\right),x\in D$$

where X is an L-diffusion and $T = \inf\{t > 0 : X_t \notin D\}$. Then u is the unique continuous function on $\bar{D}$ which is solution to the Dirichlet problem:

$$\begin{cases} Lu + \phi = 0 & in \quad D \\ u = g & on \quad \partial D \end{cases}$$

Another link is provided by the following Cauchy problem- that is, an evolution problem for which the initial condition is prescribed.

Theorem (4.1.20)

Let $g : \mathbb{R}^d \to \mathbb{R}$ be a given continuous bounded function, and let X be an L diffusion where Lsatisfies the same assumptions as in Theorem (4.1.19). Then if we define:

$$u(t,x) = \mathrm{E}_x\left(g\left(x_t\right)\right) \quad \text{for all } t \geq 0, x \in \mathbb{R}^d$$

then u is the unique solution in $C^{1,2}\left(\mathbb{R}_+ \times \mathbb{R}^d\right)$ to the problem:

$$\begin{cases} \dfrac{\partial u}{\partial t} = Lu \; on & \mathbb{R}_+ \times \mathbb{R}^d \\ u(0,.) = g \; on & \mathbb{R}^d \end{cases}$$

One word about the proof of the uniqueness part: let v be a solution to this problem, and let u be our candidate. Let us show that $v = u,$ Fix $T > 0$ and let $F(t,x) = v(T-t,x)$ .Applying Equation (4.39) to the function f , we see that

$$M_t = v\left(T - t, X_t\right), 0 \leq t \leq T$$

is a martingale. Thus

$$\mathrm{E}\left(M_0\right) = \mathrm{E}\left(M_T\right)$$

and it follows that $v(T,x) = \mathrm{E}_x\left(g\left(X_T\right)\right)$. The uniqueness part of the Theorem is proved.

Remark (4.1.21):

The application of Equation (4.39) (as opposed to Equation (4.38), which defines diffusions) is a bit tricky to justify at this point. We will soon see that diffusions solve suitable SDE's (see Theorem (4.2.2)) from which Theorem (4.1.17) follows. Alternatively, if is a diffusion then by the integration by parts formula, the process M^f of Equation (4.39) is a local martingale as soon as $f(t,x) = f_1(t)f_2(x)$ for some C^2 functions f_1 and f_2. Thus the class of f for which M^f is a local martingale contains all linear combinations of product functions $f_1(t)f_2(x)$. That Equation (4.39) holds for general functions f now follows from an approximation argument.

Remark (4.1.22):

Note that the Cauchy problem may reformulated as a Dirichlet problem in

$\mathbb{R}^{d+1}$ by changing L into

$$\tilde{L} = \overline{L} - \frac{\partial}{\partial t}$$

Fix a point $(t,x) \in \mathbb{R}^{d+1}$. By Theorem (4.1.19), the solution $u(t, x)$ is given by $E(\tilde{X}_T)$ where X is the diffusion with generator $\tilde{L}$. This corresponds to adding a coordinate X^{d+1} to the diffusion , such that $X_s^{d+1} = X_0^{d+1} - s$, that is, time is decreasing at speed 1. The time corresponds to the first time that the "time" coordinate hits 0, i.e., time t if we start from (x,t). The other d coordinates are then distributed according to $P_x(x_t \in .)$. This proves Theorem (4.1.20), given Theorem (4.1.19).

Example (4.1.23):

Let $(B_t, t \geq 0)$ be a 3-dimensional Brownian motion with $B_0 = 0$ let $\mathrm{T} = \inf\{t > 0 : \|B_t\| = 1\}$. Compute $\mathrm{E}(T)$. Answer: Let $R_t = \|B_t\|$. Then an application of $It\hat{o}$'s formula shows that

$$dR_t = dB_t + \frac{1}{R_t}.dt$$

It follows that $(R_t, t \geq 0)$ is a diffusion process on $(0, \infty)$, with generator:

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

Thus if $\Phi \equiv 1$ and $g \equiv 0$ in the previous theorem, $\mathrm{E}_0(\mathrm{T}) = u(o)$ where $u(x) = \mathrm{E}_x(\mathrm{T})$ is a function solving:

$$\bar{L}u = -1 \text{ for all } x \in (0,1)$$

Solving this ODE yields that if $f = u'$ then

$$f(x) = \left(-\frac{2}{3}x^3 - c\right)x^{-2}$$

for some constant $c \in \mathbb{R}$, so integrating:

$$u(x) = -\frac{1}{3}x^2 + \frac{c}{x} + c'$$

for a constant $c' \in R$: But we note that c must be equal to 0. Indeed, otherwise $\mathrm{E}_0(\mathrm{T}) = \infty$ by the monotone convergence theorem, which is impossible by comparison with a one-dimensional Brownian motion and Theorem (3.1.3).

Thus $u(x) = -\frac{1}{3}x^2 + c'$

and since $u(1) = 0$ we have $u(x) = \frac{1}{3}(1 - x^2)$. Hence $\mathrm{E}_0(T) = \frac{1}{3}$.

Theorem (4.1.24):

Let $f \in C_b^2(\mathbb{R}^d)$ and let $V \in L^\infty(\mathbb{R}^d)$, that is, V is uniformly bounded. For all $t, x \geq 0$, let

$$u(t,x) = \mathrm{E}_x\left(f(B_t)\exp\left(\int_0^t V(B_s)\,ds\right)\right)$$

Then u is the unique solution $w \in C_b^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$:

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{1}{2}\Delta u + Vu & on \quad \mathbb{R}_+ \times \mathbb{R}^d \\ u(0,.) = f & on \quad \mathbb{R}^d \end{cases}$$

Proof: Here again, the uniqueness part is an easy application of $It\hat{o}$'s formula. Let u be a solution and let $M_t = u(T-t, B_t)E_t$ where $E_t = \exp\left(\int_0^t V(B_s)ds\right)$ is of finite variation. By $It\hat{o}$'s formula:

$$dM_t = \nabla u(T-t, B_t)E_t dB_t + \left(-u + \frac{1}{2}\Delta u + Vu\right)(T-t, B_t)E_t dt$$

$$= \nabla u(T-t, B_t)E_t dB_t$$

since the second term is equal to 0 (because u is a solution to the PDE problem). Thus M is a local martingale, and it is uniformly bounded on [0, T], hence a true martingale. By the Optional Stopping Theorem:

$$u(T,x) = \mathrm{E}_x(M_0) = \mathrm{E}_x(M_T) = \mathrm{E}_x\left(f(B_T)\mathrm{E}_T\right)$$

which is precisely the claim.

Remark (4.1.25):

This formula turns out to be very useful when applied the other way round: in fact, it was originally introduced to compute expectations involving exponential functional s of Brownian motion, which tend to occur frequently in mathematical finance and in statistical mechanics, where V is a potential. (This is presumably why Feynman got interested in this problem). Then we can write:

$$\mathrm{E}_x\left(\exp\left\{-\beta\int_0^T V\left(X_s\right)ds\right\} f\left(X_T\right)\right)=u\left(x,T\right)$$

where

$$\frac{\partial u}{\partial t}=\frac{1}{2}\Delta u-\beta uV \quad \text{on } \mathbb{R}_+\times\mathbb{R}^d$$

and $u\left(x,0\right)=f\left(x\right)$ for all $x\in\mathbb{R}^d$. This often makes these computations easier, by bringing in techniques that were developed in analysis (e.g., Fourier analysis). In mathematical finance, the Feynman-Kac formula allows to compute the Black-Scholes formula for the price of a call in terms of a certain PDE. This point of view is in some sense dual to ours, and it is a great advantage to have these two approaches for what is, fundamentally, the same object.

Section (4.2): Stroock–Varadhan Theory of Diffusion Approximation

We start to study Martingale problems.

Let $\sigma_{i,j}(x)_{1\le i,j\le d}$ and $\left(b_i(x)\right)_{1\le i\le d}$ be a family of measurable functions with values in $\mathbb{R}$.

Let $a(x)=6(x)6^T(x)$. (Here we assume for simplicity that $m=d$).

Definition (4.2.1):

We say that a process $X = (X_t, t \geq 0)$ with values in $\mathbb{R}^d$, together with a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, solves the martingale problem $M(a,b)$ if for all $1 \leq i, j \leq d$,

$$Y^i = \left(x_t^i - \int_0^t b_i(X_s)\,ds; t \geq 0\right)$$

and

$$\left(Y_t^i Y_t^j - \int_0^t a_{i,j}(X_s)\,ds; t \geq 0\right)$$

are local martingales.

Of course, the second condition implies that

$$\left[Y^i, Y^j\right]_t = \left[X^i, X^j\right]_t = \int_0^t a_{i,j}(X_s)\,ds.$$

For instance, if σ, b are in addition Lipschitz, then there exists $\left(\Omega, x, (\mathcal{F})_{t\geq 0}\right)$ and an $\mathcal{F}_t$-Brownian motion $(B_t, t \geq 0)$ solution to the stochastic differential equation:

$$dx_t = \sigma(x_t).dB_t + b(x_t)\,dt.$$

X then solves the martingale problem $M(a, b)$. In fact, note that any (weak) solution to $E(\sigma, b)$ gives a solution to the martingale problem $\mathcal{M}$(a, b). More generally even, any L-diffusion will solve the martingale problem:

Theorem (4.2.2):

Let $a=\sigma\sigma^T$ and let X and $(\Omega,\mathcal{F},(\mathcal{F}_t),P)$ be a solution to $M(a,b)$. Then there exists an $\mathcal{F}_t$-Brownian motion $(B_t, t \geq 0)$ in R^d defined on an enlarged probability space, such that (X,B) solves $E(\sigma,\ b)$.

Proof: Assume first that σ is invertible for every $x\in\mathbb{R}^d$. Then define $Y_t^i = X_t^j - \int_0^t b_s^i(X_s)ds$, so that $Y^i = M_{c,loc}$ and by definition we have $d[Y^i,Y^j]=a_{i,j}(x_t)dt$. Define:

$$B_t^i = \int_0^t \sum_{k=1}^d \left(\sigma^{-1}\right)_{i,k}(X_s)dY_s^k$$

Thus $B^i \in M_{c,loc}$. Since $a=\sigma\sigma^T$ and thus $\rho a \rho^T = I$ where $\rho = \sigma^{-1}$, or, in coefficients, $\sum_{k,\ell}\rho_{i,k}a_{k,\ell}\rho_{j,\ell}=\delta_{i,j}$, we have

$$\left[B^i,B^j\right]_t = \sum_{k.\ell=1}^{d}\int_0^t \rho_{i.k}(X_s)a_{k,\ell}(X_s)ds = \delta_{i,j}t$$

so by Levy's characterization, B is an Brownian motion in $\mathbb{R}^d$.

Moreover, by the stochastic chain rule (Theorem 2.2.13),

$$\int_0^t \sigma(X_s)dB_s = Y_t - Y_0 = X_t - \int_0^t b(X_t)dt \qquad (4.40)$$

Indeed the i^{th} component of the left-hand side may be written as

$$\sum_{j=1}^{d}\int_0^t \sigma_{i,j}(X_s)dB_s^j = \int_0^t \sum_{j,k=1}^{d}\sigma_{i,j}\sigma_{i,k}^{-1}dY_s^k = \int_0^t dY_s$$

But Equation (4.40) is simply the statement that (X,B) solves $E(\sigma,b)$.

When σ is not everywhere invertible, we proceed like in the generalized version of Dubins- Schwartz's (when $[M]_\infty < \infty$) and let the

Brownian motion evolve independently when s is such that σX_s is not invertible.

Theorem (4.2.2) Shows that there is a one-to-one correspondence between solutions to the stochastic differential equation E(σ,b) and the martingale problem $M(a,b)$. In particular, there is uniqueness in distribution to the solutions of $\mathrm{E}(\sigma,b)$, if and only if the solutions to the martingale problem $M(a,\ b)$ are unique, where uniqueness means that all solutions to $M(a,b)$ with identical starting points have the same law.

Now we Notions of Weak Convergence of Processes.

In the following we describe some basic results in the theory of weak convergence of processes, which we do not prove due to the time constraints. We will however use these results in this section to discuss the convergence of Markov chains towards solutions of certain stochastic differential equations.

The point of view here is similar to the one in Donsker's theorem. We view a process as a random variable with values in the space Ω of trajectories. We thus need to recall a few notions about weak convergence in general metric space. Let (S,d) be a metric space. The distance function $d(x,y)$ satisfies $d(x,y)=0$ if and only if $x=y; d(x,y)\geq 0;\ d(x,z)\leq d(x,y)+d(y,z)$.

The open ball $B(x,r)$ is the set $\{y\in S: d(x,y)<r\}$. The Borel σ-field is the field generated by all open sets.

The notion of convergence in distribution is defined in terms of test functions, which are only required to be bounded and continuous (for the topology of S):

Definition (4.2.3):

Let $(\mu_n)_{n\geq 1}$ be a sequence of probability distributions on S. We say that $\mu_n \to \mu$ weakly as $n \to \infty$, if $\int_S f d\mu_n \to \int_S f d\mu$ as $n \to \infty$ for all bounded continuous functions f. If μ_n is the law of a random variable X_n and μ that of a random variable X, we say that $X_n \to X$ in distribution (or in law).

There are a number of ways one can reformulate the notion of weak convergence in terms of the mass assigned to events that are either closed or open. If $A \subseteq S$, we recall the definition of the frontier of A, which is the set $\partial A := \bar{A} \setminus \text{int}(A)$.

Theorem (4.2.4):

Let $(X_n)_{n\geq 1}$ be a sequence of random variables with values in S. The following are equivalent.

(i) $X_n \to X$ in distribution.

(ii) For all closed set K, $\limsup_{n\to\infty} \mathrm{P}(X_n \in K) \leq \mathrm{P}(X \in K)$.

(iii) For all open set 0, $\liminf_{n\to\infty} \mathrm{P}(x_n \in 0) \geq \mathrm{P}(x \in 0)$.

(iv) For all sets A such that $\mathrm{P}(X \in \partial A) = 0$,

$\limsup_{n\to\theta} \mathrm{P}(X \in A) = \mathrm{P}(X \in A)$

(v) For all sets A such that $\mathrm{P}(X \in \partial A) = 0$,

$\lim_{n\to\infty} \mathrm{P}(x \in A) = \mathrm{P}(x \in A)$

(vi) For any bounded function f, denote by D_f the set of discontinuities of f. Then for any f such that $\mathrm{P}(X \in D_f) = 0$, $\mathbf{E}(f(X_n)) \to \mathbf{E}(f(X))$ as $n \to \infty$.

It is important to note that the random variables X_n need not be related in any particular way. In fact they may even be defined on different probability spaces. However, it turns out that (provided the metric space is sufficiently nice), one can always choose a common probability space for the random variables and define a sequence of random variables Y_n with law identical to X_n, in such a way that convergence occurs almost surely. This is the content of the "Skorokhod representation theorem", which we may occasionally need.

Lemma (4.2.5):

Suppose S is complete and separable. If $\mu_n \to \mu$ weakly then there exists random variables Y_n defined on $\Omega = [0,1]$ equipped with the Lebesgue measure P, such that $Y_n \overset{d}{=} \mu$ for all $n \geq 1$, and $\lim_{n\to\infty} Y_n = Y$, P-almost surely, where $Y_n \overset{d}{=} \mu$.

We now specialize to the case where the random variables X_n take values in the space C of continuous trajectories over the compact interval [0,1]. This is precisely the point of view in Donsker's theorem. We equip C with the distance of the sup-norm:

$$d(f,g) = \|f-g\|_\infty = \sup_{t\in[0,1]} |f(t)-g(t)|$$

This turns C into a complete, separable metric space, on which it makes sense to talk about weak convergence.

Example (4.2.6):

If $(S_n, n \geq 0)$ is a simple random walk on $\mathbf{Z}$, then by Donsker's theorem: $(S_t^{[N]}, 0 \leq t \leq 1)$, converges weakly towards a Brownian motion on [0, 1], where $\left(S_t^{[N]} = N^{-1/2} S_{Nt}\right)$.

A classical trick in analysis for proving convergence of a sequence x_n towards a limit x is to prove that (a) the sequence takes its values in a compact set, and (b) there can only be one sub sequential limit. It is usually part (a) which demands slightly harder work to establish, as part (b) follows from usually softer arguments (we typically have identified the limit at this stage). Fortunately there is a general criterion and fairly easy to use in practice, which tells us when the set $k = \{x_n\}_{n=1}^{\infty}$ is compact (or, actually, pre-compact, meaning that $\bar{k}$ is compact). When this happens, we say that the sequence of processes $(x_n)_{n\geq 1}$ is tight. This criterion consists in, roughly speaking, showing that the process doesn't oscillate too wildly.

This is the content of the following theorem. For a continuous path $w(t), t \in [0,1]$ let $osc_{\delta}(w) = \sup\{|w(s) - w(t)| : |s-t| \leq \delta\}$ osc_{δ} is simply the modulus of continuity of the path w, at precision δ.

Theorem (4.2.7):

Suppose that $(X^n)_{n\geq 1}$ is a sequence of processes with values in C. Then X^n is tight, if and only if for each $\varepsilon > 0$, there exists $n_o, M \geq 1$ and $\delta > 0$ such that:

(i) $\mathbf{P}\left(\left|X^n(0)\right| > M\right) \leq \varepsilon$ for all $n \geq n_0$.

(ii) $\mathbf{P}\left(osc_{\delta} > \varepsilon\right) \leq \varepsilon$

To summarize, to show that a sequence converges weakly in C, it suffices to prove that (i) and (ii) hold above and that there is a unique weak sub sequential limit. This is for instance the case if one has already established convergence of the finite-dimensional distributions, i.e., convergence of the k-dimensional vector $\left(X_{t_1}^n, ..., X_{t_k}^n\right)$ towards $\left(X_{t_1}, ..., X_{t_k}\right)$ for any $k \geq 1$ and any choice of "test times" $t_1, ..., t_k$. This could have been a possible route for proving Donsker's theorem, as convergence of finite-dimensional distributions is easy to establish.

Note that condition (i) in the above theorem says that the starting point of the process $X^n(0)$ takes values in a compact set with arbitrarily high probability. This is usually trivial since very often, the starting point of a process is a deterministic point such as 0.

In the later, we will prove weak convergence of certain rescaled Markov chains towards diffusion processes. For this, we will usually use the fact that any weak sub sequential limit must satisfy the associated martingale problem $M(a,b)$ for which sufficient smoothness of the coefficients proves uniqueness in distribution. However there is one (small) additional difficulty in this case: it will be more natural to work with right-continuous processes X^n rather than with the linear-interpolation of X_n, which typically loses some of the Markov property.

Let D be the space of right-continuous paths on [0,1]. Without entering into the details, D can also be equipped with a complete separable metric d, which is called the Skorokhod topology. It can also be proved that if a sequence of right-continuous processes X^n satisfy (i) and (ii) in Theorem (4.2.7), then X^n is also tight and any sub sequential limit Xmust be continuous, in the sense that $\mathrm{P}(X \in C) = 1$. Furthermore, weak convergence with respect to the Skorkhod topology towards a continuous process , implies weak convergence in C of the linear interpolations. Another fact which will be needed is that if $x_n \to x$ in the Skorokhod topology, the $x_n(t) \to x(t)$ for all $t \geq 0$.

Now we need to study Markov chains and diffusions.

The result which we now discuss is due to Stroock and Varadhan, and shows a link between rescaled Markov chains and certain diffusion processes. It is applicable in a remarkably wide variety of contexts, of which we will only have the time to give one example.

While the idea for the statement of the result is in fact fairly simple, there is quite a bit of notation to introduce. We assume that a certain Markov chain is given to us. A certain scaling parameter $h > 0$ is going to 0, and we assume that the chain has already been rescaled, so it takes it values in a certain set $S_h \subseteq \mathbb{R}^d$. We will denote this Markov chain by $\left(Y_n^h, n \geq 1\right)$. The transition probabilities of Y are given by a transition kernel $\prod_h$ which may depend on $h > 0$:

$$\mathbf{P}\left(Y_{n+1}^h \in A \mid Y_n^h = x\right) = \Pi_h(x, A)$$

We define the random process on [0,1] by

$$X_t^h = Y_{\lfloor t/h \rfloor}^h, t \in [0,1]$$

so that X^h is almost surely right-continuous and is constant between two successive jumps of the chain, which may occur every h units of time for the process X^h. We let K_h denote the rescaled transition kernel:

$$K_h(x, dy) = \frac{1}{h}\Pi^h(x, dy)$$

Roughly, the conditions of the theorem states that "the infinitesimal mean variance of the jumps of X when $X = x$ are approximately given by b(x) and $\sigma(x)$, respectively". The conclusion states that X^h converges weakly towards the solution of $M(a,b)$.

For $1 \leq i, j \leq d$, define:

$$a_{i,j}^h = \int_{|y-x|\le 1} (y_i - x_i)(y_j - x_j) K_h(x, dy)$$

$$b_i^h(x) = \int_{|y-x|\le 1} (y_i - x_i) K_h(x, dy)$$

$$\Delta_\zeta^h(x) = k^h(x, B(x,\varepsilon)^c)$$

Suppose that a_{ij} and b_i are continuous coefficients on $\mathbb{R}^d$ for which the martingale problem $M(a,b)$ is well posed, i.e., for each $x \in \mathbb{R}^d$ there is a unique in distribution process $(X_t, 0 \le t \le 1)$ such that $X_0 = x$ almost surely, and

$$Y_t^i = X_t^i - \int_0^t b_i(X_s)ds \quad \text{and} \quad Y_t^i Y_t^j - \int_0^t a_{ij}(X_s)ds$$

are both local martingales.

Theorem (4.2.8):

Suppose that the above holds, and that for every $1 \le i, j \le d$, and every $R > 0$, every $\varepsilon > 0$,

(i) $\lim_{h\to 0} \sup_{|x|\le R} \left|a_{ij}^h(x) - a_{ij}(x)\right| = 0$

(ii) $\lim_{h\to 0} \sup_{|x|\le R} \left|b_i^h(x) - b_i(x)\right| = 0$

(iii) $\lim_{h\to 0} \sup_{|x|\le R} \Delta_\varepsilon^h(x) = 0$

Then if $X_0^h = x_h \to x_0$, we have $(X_t^h, 0 \le t \le 1) \to (X_t, 0 \le t \le 1)$ weakly in D, and in particular, the linear interpolations of Y^h converge weakly in C.

The rest of this part is devoted to a proof of this result. By localization, one may replace (i), (ii) and (iii) by the following stronger conditions:

(i) $\lim_{h\to 0} \sup_{x\in\mathbb{R}^d} \left|a_{ij}^h(x) - a_{ij}(x)\right| = 0$

(ii) $\lim_{h\to 0}\sup_{x\in\mathbb{R}^d}\left|b_i^h(x)-b_i(x)\right|=0$

(iii) $\lim_{h\to 0}\sup_{x\leq\mathbb{R}^d}\Delta_\varepsilon^h(x)=0$

(iv) Moreover $a_{i,j}^h, b_i^h, \Delta_\zeta^h$ are uniformly bounded in h and x.

Step 1. Tightness:

Let f be a bounded and measurable function. Define the operator L^h by

$$L^h f(x)=\int K_h(x,dy)\big(f(y)-f(x)\big) \tag{4.41}$$

This is the "generator" of the process: this represents the infinitesimal change in the function f when the process is at x. In particular, note that the process

$$f(Y_k^h)-\sum_{j=0}^{k-1} hL^h f(Y_j^h), k=0,1,2,... \tag{4.42}$$

is a (discrete-time) martingale. For our proof of tightness we are going to need an estimate on the time needed by the chain to make a deviation of size roughly $\varepsilon>0$, when it starts at position $y\in\mathbb{R}^d$. To do this, we introduce a function $g:\mathbb{R}\to\mathbb{R}$ such that $g\in C^2, 0\leq g\leq 1$ and $g(x)=0$ if $x\in\mathbb{R}^d$, while $g(0)=1$. We also define for $x\in\mathbb{R}^d$ $f_\varepsilon(x)=g(|x|^2/\varepsilon^2)$ which is also C^2, and becomes 0 when $|x|\geq\varepsilon$, and for $a\in\mathbb{R}^d$, let $a\in\mathbb{R}^d$ let $f_{a,\varepsilon}(x)=f_\varepsilon(x-a)$.

Lemma (4.2.9):

There exists $C_\varepsilon<\infty$, independent of h, such that $\left|L^h f_{a,\varepsilon}\right|<C_\varepsilon$ for all $a,x\in\mathbb{R}^d$.

Proof: This is simply an application of Taylor expansion. For $t\in[0,1]$ and $a,x\in\mathbb{R}^d$, let $\phi(t)=f_{a,\varepsilon}=\big(x+t(y-x)\big)$. Then by Taylor's theorem, there exist $C_{xy}\in[0,1]$ such that

$$f_{a,\varepsilon}(y)-f_{a,\varepsilon}(x)=\phi(1)-\phi(0)=\phi'(0)+\frac{1}{2!}\phi''(c_{xy})$$

$$=\sum_{i=1}^{d}(y_i-x_i)D_i f(x)+\sum_{1\le i,j}^{d}(y_i-x_i)(y_j-x_j)D_{ij}f(z_{xy})$$

where $f\equiv f_{a,\varepsilon}$ and D_i and D_{ij} stand for $\frac{\partial f}{\partial x_i}$ and $\frac{\partial^2 f}{\partial x_i \partial x_j}$ respectively, while $z_{xy}=x+c_{xy}(y-x)\in[x,y]$.

To obtain $L^h f(x)$, we integrate the above with respect to $k_h(x,dy)$, and get:

$$L^h f_{a,\varepsilon}(x)=\int k_h(x,dy)\left(f_{a,\varepsilon}(y)-f_{a,\varepsilon}(x)\right)\le\left|\nabla f_{a,\varepsilon}(x).\int_{|y-x|\le 1}(y-x)k_h(x,dy)\right|$$
$$+\left|\int_{|y-x|\le 1}\sum_{i,j}(y_i-x_i)(y_j-x_j)D_{ij}f_{a,\varepsilon}(\square_{xy})K_h(x,dy)\right|$$
$$+2\|f_{a,\varepsilon}\|_\infty K_h\left(x,b(x,1)^c\right)$$

Let $A_\varepsilon=\sup\left|\nabla f_{a,\varepsilon}(x)\right|$, let $B_\varepsilon=\sup_z\|Df(z)\|$, where $Df=(D_{ij}f)_{1\le i,j\le d}$ the Hessian matrix of f and for a matrix $\mathrm{M}=(m_{ij})$ we note

$$\|m_{ij}\|=\sup_{u\in\square^d:|u|=1}\left|\langle u,Mu\rangle\right|$$

Thus

$$\left|\sum_{i,j}(y_i-x_i)(y_j-x_j)D_{ij}f_{a,\varepsilon}(Z_{xy})\right|\le|y-x|^2 B_\varepsilon$$

hence by Cauchy-Schwarz

$$L^h f_{a,\varepsilon}(x)\le A_\varepsilon\left|b^h(x)\right|+B_\varepsilon\int_{|y-x|\le 1}|y-x|^2K_h(x,dy)+2K_h\left(x,B(x,1)^c\right)$$

Since $\int\limits_{|y-x|\le 1} |y-x|^2 K_h(x,dy) = \sum_i a_{ij}^h$ and since we have assumed in (iv) that all those quantities were uniformly bounded, we have proved the lemma.

To estimate $osc_\delta(x^h)$, we introduce the following random variables:

$$\mathrm{T}_0 = 0\,,$$

$$\mathrm{T}_n = \inf\left\{t \ge \mathrm{T}_{n-1} : \left|X_t^h - X_{Tn-1}\right| \ge \varepsilon\right\}$$

$$\square = \min\{n : \mathrm{T}_n > 1\}$$

$$\sigma = \min\{\mathrm{T}_n - \mathrm{T}_{n-1} : 1 \le n \le \square\}$$

and, finally

$$\theta = \max\left\{\left|X^h(t) - X^h(t^-)\right| : 0 \le t \le 1\right\}$$

The relation between these random variables and tightness is provided by the following lemma.

Lemma (4.2.10):

Assume that $\sigma > \delta$ and that $\theta < \varepsilon$. Then $osc_\delta(x^h) \le 4\varepsilon$.

Proof: The proof is straightforward. We want to show that for all s; $s,t \in [0,1]$ with $|s-t| \le \delta$, $\left|X^h(s) - X^h(t)\right| \le 4\varepsilon$. The point is that since $|s-t| < \delta < \sigma$, s and t can only span at most one of the interval $[\mathrm{T}_{n-1}, \mathrm{T}_n]$, and by definition of these stopping times, everything behaves well on those intervals. Thus if $\mathrm{T}_{n-1} \le s \le t < \mathrm{T}_n$, then $|f(s) - f(t)| \le 2\varepsilon$. If on the other hand, $\mathrm{T}_{n-1} \le s \le \mathrm{T}_n < t$, then

$$\left|f(s)-f(t)\right| \le \left|f(s)-f\left(\mathrm{T}_{n-1}\right)\right| + \left|f(t)-f\left(\mathrm{T}_n\right)\right| + \left|f\left(\mathrm{T}_n\right)-f\left(\mathrm{T}_n^-\right)\right|$$
$$+\left|f\left(\mathrm{T}_n^-\right)-f\left(\mathrm{T}_{n-1}\right)\right| \le 4\varepsilon$$

We now use this to prove the tightness estimate. Since it is assumed that the starting point $X_\circ^h = x^h$ is nonrandom and converges towards a fixed x_0, it suffices to prove the statement about oscillations: for all ε, there exists $\delta > 0$ and such that for all $h \le h_0$,

$$\mathrm{P}\left(osc_\delta\left(x^h\right) \ge \varepsilon\right) \le \varepsilon$$

Thus it follows to prove that for all h sufficiently small and for δ small enough, $\mathbf{P}_x(\theta > \varepsilon/4) \to 0$ as $h \to 0$, and $\mathrm{P}_x(\sigma > \delta) \to 0$ for $h \to 0$ for all $x \in R^d$. The first one is very simple: since there are at most $\frac{1}{h}$ time steps in the unit interval [0, 1], a simple union bound yields

$$\mathrm{P}_x(\theta > \varepsilon) \le \frac{1}{h}\sup_y \prod\nolimits_h \left(y, B(y,\varepsilon)^c\right) \le \sup_y \Delta_\varepsilon^h(y) \to 0$$

by (iii). The second one requires more arguments. We follow the elegant argument introduced by Stroock and Varadhan. The first step is to estimate $P_x\left(T_{1`} \le u\right)$ for small u. Note that by Lemma (4.2.9), the process

$$f_{x,\varepsilon}\left(\mathrm{Y}_k^h\right) + \square_\varepsilon hk,\ k = 0,1,2,\ldots$$

is a submartingale. Thus letting

$\mathrm{T} = \inf\left\{k \ge 1 : \left|\mathrm{Y}_k^h - x\right| > \varepsilon\right\}$ so that $\mathrm{T}_1 = h\mathrm{T}$ .Using the Optional stopping theorem at $\mathrm{T}\Lambda u'$ with $u' = u/h$,

$$\mathrm{E}_x\left\{f_{x,\varepsilon}\left(\mathrm{Y}_{\mathrm{T}\Lambda u'}^h\right) + \square_\varepsilon h\left(Y_{T\Lambda u'}^h\right)\right\} \ge 1$$

Since $T \Lambda u' \leq u$ and since on the event that $T \leq u'$, we have that $\left|Y_{T \Lambda u'} - x\right| \geq \varepsilon$, so $f_{x,\varepsilon}\left(Y^{h}_{\mathrm{T}\Lambda u'}\right) = 0$, we have:

$$\mathrm{P}_x\left(\mathrm{T}_1 \leq u\right) = \mathrm{P}\left(\mathrm{T} \leq u'\right) \leq \mathrm{E}_x\left\{1 - f_{x,\varepsilon}\left(Y^{h}_{\mathrm{T}\Lambda u'}\right)\right\} \leq h\mathrm{C}_{\varepsilon} u' = \mathrm{C}_{\varepsilon} u$$

This has the following consequence: for all u > 0, letting $\mathrm{P} = \mathrm{P}_x\left(\mathrm{T} \leq u\right)$:

$$\begin{aligned}
&\mathrm{E}_x\left(e^{-\mathrm{T}}\right) \leq \mathrm{P}_x\left(\mathrm{T} \leq u\right) + e^{-u}\mathrm{P}_x\left(\mathrm{T} \geq u\right) \leq P + e^{-u}\left(1 - P\right) \leq e^{-u} + P\left(1 - e^{-u}\right) \\
&\leq e^{-u} + Pu \leq 1 - u + \square_{\varepsilon} u^2
\end{aligned}$$

Thus by choosing u small enough, we can find $\lambda < 1$, independent of x or δ (depending solely on ε through C_{ε}), such that $\mathbf{E}_x\left(e^{-\mathrm{T}}\right) \leq \lambda$. Now, iterating and using the strong Markov property at times $\mathrm{T}_1, \ldots, \mathrm{T}_n$, which are stopping times,

$$\mathrm{E}_x\left(e^{-\mathrm{T}_n}\right) \leq \lambda$$

since λ does not depend on x, and thus by Markov's inequality:

$$\begin{aligned}
&\mathrm{P}_x\left(N > n\right) = \mathrm{P}_x\left(\mathrm{T}_n < 1\right) \leq \mathrm{P}_x\left(e^{-\mathrm{T}_n} \geq e^{-1}\right) \\
&\mathrm{P} \qquad\qquad \leq e\mathrm{E}_x\left(e^{-\mathrm{T}_n}\right) \leq e\lambda^n
\end{aligned}$$

We finish by saying that

$$\mathrm{P}_x\left(\sigma \leq \delta\right) \leq k \sup_y \mathrm{P}_y\left(\mathrm{T} \leq \delta\right) + \mathrm{P}_x\left(\square > k\right) \leq \square_{\varepsilon} k\delta + e\lambda^k$$

Thus we take k large enough that $e\lambda^k < \varepsilon/2$ and then pick δ small enough that $\square_{\varepsilon} k\,\delta < \varepsilon/2$.

We are then done for the proof of tightness.

Step 2. Uniqueness of the weak subsequential limits.

Since we have assumed that the martingale problem $\mathrm{M}(a,b)$ was well posed, it suffices to show that the limit of any weakly convergent subsequence solves the martingale problem $\mathrm{M}(a,b)$. Our first step for doing so is to show that the generator of the Markov chain L^h converges in a suitable sense to the generator L of the diffusion:

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

Lemma (4.2.11):

Let $f\in C_k^2$ be twice differentiable and with compact support. Then $L^h f(x)\to L(x)$ uniformly over $x\in\mathbb{R}^d$ as $h\to 0$.

Proof: Going back to our Taylor expansion of $L^h f(x)$, and recalling the definition of $b_i^h(x)$ and $a_{ij}(x)$, we may write:

$$\begin{aligned}L^h f(x)&=\sum_{i=1}^{d}b_i^h(x)D_i f(x)\\&+\int_{|y-x|\le 1}\sum_{i,j=1}^{d}(y_i-x_i)(y_j-x_j)D_{ij}f(z_{xy})K_h(x,dy)\\&+\int_{|y-x|>1}\left[f(y)-f(x)\right]K_h(x,dy)\end{aligned}$$

The final term in the right-hand side converges to 0 uniformly in x by assumption (iii) with $\varepsilon=1$. To deal with the first term, note that

$$\begin{aligned}&\left|\sum_{i=1}^{d}b_i^h(x)D_i f(x)-\sum_{i=1}^{d}b_i(x)D_i f(x)\right|\le\\&\sup_{1\le i\le d}\left|b_i^h(x)-b_i(x)\right|\sum_{i=1}^{d}\|D_i f\|_\infty\end{aligned}$$

Which converges to 0 uniformly in x by assumption (ii) (since $f \in C_k^2$). It remains to deal with the central term. Recalling the definition of $a_{ij}^h(x)$, we get:

$$\begin{aligned}
&\left| \int_{|y-x|\le 1} \sum_{i,j=1}^{d} (y_i - x_i)(y_j - x_j) D_i f(z_{xy}) k_h(x,dy) - \sum_{i,j=1}^{d} a_{ij}(x) D_{ij} f(x) \right| \\
&\le \left| \sum_{i,j=1}^{d} a_{ij}^h(x) D_{ij} f(x) - \sum_{i,j=1}^{d} a_{ij}(x) D_{ij} f(x) \right| \\
&+ \left| \int_{|y-x|\le 1} \sum_{i,j=1}^{d} (y_i - x_i)(y_j - x_j) \left[D_{ij} f(z_{xy}) - D_{ij} f(x) \right] k_h(x,dy) \right|
\end{aligned}$$

The first term converges to 0 uniformly in x by (i) and the fact that the derivatives of f are uniformly bounded. The second term can be split in an integral over $|y-x| > \varepsilon$ and $|y-x| < \varepsilon$. The first one converges to 0 uniformly in $x \in \mathbb{R}^d$ thanks to (iii) and the fact that the integrand is bounded. For the other term, let

$$\Gamma(\varepsilon) = \sup_{1\le i,j\le d} \sup_{|y-x|\le\varepsilon} \left| D_{ij} f(z_{xy}) - D_{ij} f(x) \right|$$

Then since z_{xy} lies on the segment between x and y, and since $D_{ij}f$ is continuous on the compact set K (and hence uniformly continuous), $\Gamma(\varepsilon) \to 0$ as $\varepsilon \to 0$. On the other hand,

$$\begin{aligned}
&\left| \int_{|y-x|\le 1} \sum_{i,j=1}^{d} (y_i - x_i)(y_j - x_j) \left[D_{ij} f(z_{xy}) - D_{ij} f(x) \right] k_h(x,dy) \right| \\
&\le \Gamma(\varepsilon) \int_{|y-x|\le\varepsilon} |y-x|^2 k_h(x,dy)
\end{aligned}$$

by Cauchy-Schwarz's inequality,

so the proof of the lemma is complete.

We now use this lemma to conclude the proof of Theorem (4.2.8). Fix $h_n \to 0$ such that $X^{h_n} \to x$ weakly (in D) as $n \to \infty$. (Recall that X^h is defined as $X_t^h = Y_{\lfloor t/h \rfloor}$. Fix $s < t$. Then for any continuous functional $F : D \to R$ which is measurable with respect to Fs, we have, since L^h is the discrete generator of Y,

$$f\left(X_{kh_n}^{h_n}\right) - \sum_{j=0}^{k-1} h_n L^{h_n} f\left(X_{jh_n}^{h_n}\right), k = 0,1,...$$

is a martingale. In particular, taking $k = k_n$ such that $kh_n > s$, i.e., $k_n = \lceil s/h_n \rceil$, and taking ℓ_n similarly so that $\ell_n h_n > t$, i.e., $\ell_n = \lceil t/h_n \rceil$, we get

$$\mathbf{E}_x\left(F\left(x^{h_n}\right)\left\{f\left(X_{\ell_n h_n}^{h_n}\right) - f\left(x_{k_n h_n}^{h_n}\right) - \sum_{j=k_n}^{\ell_{n-1}} h_n L^{h_n} f\left(X_{jh_n}^{h_n}\right)\right\}\right) = 0$$

By using the Skorokhod representation theorem, one may find Y^n such that $Y^n \overset{d}{=} X^{hn}$ and $Y^n \to Y$ almost surely, where $Y \overset{d}{=} X$. We recognize a Riemann sum in this expectation. Since almost sure convergence in D implies almost sure convergence of the martingals, we conclude by the Lebesgue convergence theorem that

$$\mathrm{E}_x\left(F(n)\left\{F(X_t) - F(X_s) - \int_s^t Lf\left(X_u\right) du\right\}\right) = 0$$

Since F is an arbitrary continuous function on D, it follows that

$$f(X_t) - \int_0^t Lf(X_u) du, t \geq 0$$

is a martingale for all $f \in C_k^2$. Since the martingale problem has a unique solution, the desired conclusion follows. This ends the proof of Theorem (4.2.8).

Example (4.2.12):

This result has literally thousands of practical applications, and we show one particularly simple such application.

Now we discuss The Ehrenfest chain: This is a Markov chain which models a box filled with gas molecules which is divided in two equal pieces, and where gas molecules can be exchanged between the two pieces through a small hole. Mathematically, we have two urns with a total of $2n$ balls (molecules). At each time step we pick one ball uniformly at random among the $2n$ balls of the urn, and move it to the other urn (we think of this event as a certain gas molecule going through that hole). Let Y_t^n denote the number of molecules in the left urn.

Define a normalized process $X_t^n = \left(Y_{\lfloor tn \rfloor}^n - n\right)/\sqrt{n}$, and assume for instance that $Y_0^n = n$, i.e., equal number of molecules in each urn.

Theorem (4.2.13):

The process $\left(X_t^n, 0 \le t \le 1\right)$ converges weakly to an Orstein-Uhlenbeck diffusion $\left(X_t, 0 \le t \le 1\right)$ with unit viscosity, i.e., the pathwise unique solution to

$$dX_t = -X_t dt + dB_t, X_0 = 0$$

Thus the number of molecules in each urn never deviates too much from n. Writing $K^n(x, dy) = n\prod^n(x, dy)$.

Proof: The state space for Y^n is $S_n = \left\{k/\sqrt{n} : -n \le k \le n\right\}$. The transition probability $\prod^n$ of Y^n is given

$$\prod{}^n\left(x, x+n^{-1/2}\right) = \frac{n - x\sqrt{n}}{2n}, \prod{}^n\left(x, x+n^{-1/2}\right) = \frac{n + x\sqrt{n}}{2n}$$

Here $d=1$, and the expected infinitesimal drift

$$\hat{b}^n(x)=\int (y-x)k^n\left(x,dy\right)=n\left\{n^{-1/2}\frac{n-x\sqrt{n}}{2n}-n^{-1/2}\frac{n+x\sqrt{n}}{2n}\right\}=-x$$

While the infinitesimal variance

$$\hat{a}^n(x)=\int (y-x)^2K^n(x,dy)=n\left\{n^{-\frac{1}{2}}\frac{n-x\sqrt{n}}{2n}+n^{-\frac{1}{2}}\frac{n-x\sqrt{n}}{2n}\right\}=1$$

It follows without difficulty that the truncated expected drift and variance, respectively

$b^n(x)=\int_{|y-x|\leq 1}\left(y-x\right)^2k^n(x,dy)$ and $a^n(x)=\int_{|y-x|\leq 1}\left(y-x\right)^2k^n(x,dy),$ satisfy:

$a^n\left(x\right)\to 1; b^n\left(x\right)\to -x$

Uniformly on every compact set. Since the coefficients of the Ornstein-Uhlenbeck diffusion are Lipschitz, there is pathwise uniqueness for the associated SDE and thus uniqueness in distribution. Therefore, $\left(X_t^n, 0\leq t\leq 1\right)$ converges to $\left(X_t, 0\leq t\leq 1\right)$ weakly, by Theorem (4.2.8).

References

[1] B. Øksendal. Stochastic Differential Equations, An Introduction with Applications, 6th ed. Springer-Verlag Berlin Heidelberg New-York, 2003.

[2] D. Revuz and M. Yor (1999). Continuous martingales and Brownian motion. 3^{rd} edition.Springer, Grundlehren der mathematischenWissenschaften, Vol. 293.

[3] D.W. Stroock and S.R.S. Varadhan (1997). Multidimensional Diffusion Processes.Springer, Classics in Mathematics. Reprint of Grundlehren dermathematischenWissenschaften, Vol. 233.

[4] Hamza K. and Klebaner F.C. (1995) "Conditions for integrability of Markov Chains," J. Appl. Probab.

[5] J. Bertoin. Lévy Processes. Cambridge University Press, 121, 1996.

[6] Kallianpur G. Stochastic Filtering Theory. Springer-Verlag, 1980.

[7] Karatzas I. and Shreve S.E. Brownian Motion and Stochastic Calculus. Springer-Verlag, 1988.

[8] L.C.G. Rogers and D. Williams (1987). Diffusions, Markov Processes and Martingales, Volume 2. 2^{nd} edition. Cambridge University Press.

[9] Oksendal B. Stochastic Differential Equations. Springer-Verlag, 1995.

[10] R. Durrett. Stochastic calculus: a practical introduction. CRC press, Probability and Stochastics Series, 1996.

[11] Revuz D. and Yor M. Continuous Martingales and Brownian Motion. Springer-Verlag, 1991.