Stochastic Calculus

Brian Lee

Spring 2026

These are notes on Stochastic Calculus, broadly based on notes taken from the course 18.676 (Stochastic Calculus) taught at MIT in Spring 2026 by Nike Sun. A brief list of topics includes a formal construction of Brownian Motion, Continuous Time Martingales, Stochastic Integration, Markov Processes, and SDEs. The main reference for the notes are [Le 16] but my notes cover additional material from from [Kal21], [KS91], and [KP92]. In particular, my notes diverge substantially from the lecture material around Lecture 19, where I cover additional topics not covered in lecture such as more general Markov Processes, connections to PDEs, numerical methods, Mallavian Calculus, along with some connections to Quantum Field Theory and Mathematical Finance. The notes are of varying quality, with material towards the beginning somewhat more poorly written than the end. Hopefully in the future, I will have some time to adapt them into a full set of working notes.

Contents

Part I. Foundations

1. (L1) Introduction

A central goal of the class will be the construction of the Brownian Motion and the Stochastic Integral which will give us the machinery for most of the class.

1.1. Brownian Motion as Scaled Simple Random Walks

We start with a motivating example:

Example 1.1 (*Brownian Motion as a Scaled Simple Random Walk*)
Consider a simple random walk $(S_n)_{n\geq 0}$ (ie. $S_n = \sum_{i=1}^n X_i$ where $X_i \sim \text{Unif}\{-1,1\}$. By the CLT, we have $S_n/\sqrt{n} \xrightarrow{d} \mathcal{N}(0,1)$. Now, denote the time and space-scaled SRW:

$$X^{(n)}(t) := \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor} \xrightarrow{d} \mathcal{N}(0,t) \tag{1.1}$$

We have in fact that the entire *process* $X^{(n)} = (X^{(n)}(t))_{t\geq 0}$ converges in law with respect to the metric topology induced by the uniform norm. We will show that this scaled limit is the **Brownian Motion**, ie.

$$(X^{(n)}(t))_{t\geq 0} \xrightarrow{d} (B_t)_{t\geq 0} \tag{1.2}$$

where $B_0 = 0$, $B_t - B_s \sim \mathcal{N}(0, t-s)$ for $t \geq s$ with independent increments.

In fact it turns out all continuous local martingales are of this form and are essentially time-changed Brownian Motions (see the Dambis-Dubins-Schwarz Theorem), which lies at the heart of the reason why the continuous martingale case differs so starkly from the discrete case.

1.2. An Introduction to Ito's Formula

The central result of the course will be that of **Ito's Formula**. Informally, suppose we have a stochastic process X_t that follows a SDE:

$$dX_t = \mu_t dt + \sigma_t dB_t \tag{1.3}$$

where μ_t is known as the **drift** and σ_t the **volatility** (or **diffusion**). Informally, think of this as increments:

$$X_{t+dt} - X_t \approx \mu_t dt + \sigma_t \mathcal{N}(0, dt) \tag{1.4}$$

Consider $f \in C^2(\mathbb{R})$ and consider the evolution of $f(X_t)$. We have (informally)

$$df(X_t) = f'(X_t)[\mu_t dt + \sigma dB_t] + \frac{1}{2} f''(X_t)[\mu_t dt + \sigma_t dB_t]^2 \tag{1.5}$$

where the other terms die in the infinetisimal limit since dB_t^2 is of order dt while the latter terms are $o(dt)$. Ito's formula says in fact that:

$$df(X_t) = \underbrace{\left(f'(X_t)\mu_t + \frac{1}{2}f''(X_t)\sigma_t^2\right)}_{\text{drift}} dt + \underbrace{f'(X_t)}_{\text{volatility}} dB_t \tag{1.6}$$

The reasoning is that the dt^2 is small and thus dies out, as does the cross term since $dtdB_t = \Theta((dt)^{3/2}) = o(dt)$, but the dB_t^2 term is of order dt as well so it remains. In fact, we will see later on in the course in general, that all continuous semi-martingales are essentially of this order ($\approx \sqrt{dt}$) giving rise to the idea of the quadratic variation which remains finite for this class. From here, we derive the Ito integral as an extension of the Stieltjes integral for finite variation processes.

1.3. Application: Conformal Invariance of Planar Brownian Motion

Finally, we give a motivating application in complex analysis by showing that a planar Brownian Motion is conformally invariant. Recall some basic facts from complex analysis:

Definition 1.2

A function $f : \mathbb{C} \to \mathbb{C}$ is **holomorphic** if $\lim_{h\to 0} \frac{f(z+h)-f(z)}{h}$ exists where $h \in \mathbb{C}$.

Theorem 1.3 (*Cauchy-Riemann Equations*)

Let $f : \mathbb{C} \to \mathbb{C}$ be a complex function and let $u = \Re f, v = \Im f$. We have that f is holomorphic at $z = x + iy$ iff it satisfies the **Cauchy-Riemann Equations**:

$$u_x = v_y, \quad u_y = -v_x \tag{1.7}$$

Additionally, u and v are harmonic.

Definition 1.4

Let D, D' be open in $\mathbb{C}$. We say $f : D \to D'$ is **conformal** if it is holomorphic with holomoprhic inverse. Notably, $f'(z) = u_x + iv_y = v_y - iu_x \neq 0 \; \forall z \in D$.

Intuitively, a conformal map rotates and scales but does not shear.

Theorem 1.5 (*Conformal Invariance of Brownian Motion*)

Let $f : D \to D'$ be conformal and let $B_t = (X_t, Y_t)$ be a 2-dimensional Brownian Motion. Then, let $Z_t := X_t + iY_t$, which is a standard Brownian Motion in $\mathbb{C}$. If $u := \Re(f)$ and $v = \Im(f)$, we have

$$du(Z_t) = |f'(Z_t)| O(Z_t) \begin{pmatrix} dX_t \\ dY_t \end{pmatrix} \tag{1.8}$$

where $O(Z_t)$ is a 2×2 orthonormal matrix.

Proof. The proof is a direct application of Ito's Formula and the fact that u and v is harmonic. $\square$

The conclusion is that for conformal f, $f(B_t)$ is a Brownian Motion with a random time-change $|f'(Z_t)|$. The following characterization of the possion distribution is thus an interesting corollary:

Corollary 1.6 (*Exit Time of a Brownian Motion from the Upper Half Plane*)
Consider a planar (complex) Brownian Motion in the upper half plane $\mathbb{H}^2$ starting at $H_0 = iy$ for $y > 0$. Let $\tau = \inf\{t : Z_t \in \mathbb{R}\}$. The law of H_τ is then given by

$$\mathbb{P}(Z_t \in [a, b]) = \int_a^b \frac{y}{\pi(x^2 + y^2)} dx \tag{1.9}$$

ie. it has density given by the Poisson Kernel.

Proof. Consider the map $f : \mathbb{H} \to \mathbb{D}$ sending $iy \mapsto 0$, for example

$$f(z) = (i - z/y)/(i + z/y) \tag{1.10}$$

This function is conformal and maps the upper half plane to the unit circle. By conformal invariance, $f(Z_t)$ is a Brownian motion in the unit disk starting at 0 and the exit time is when it exits the circle between $f(a)$ and $f(b)$. Thus, the exit probability is just the ratio of the arc length, ie.

$$\mathbb{P}(Z_\tau \in [a, b]) = \frac{1}{2\pi} \int_a^b |f'(x)| dx = \int_a^b \frac{y}{\pi(x^2 + y^2)} dx \tag{1.11}$$

□

2. (L2) Gaussian Spaces and Processes, Gaussian White Noise

We review Gaussian random variables quickly.

2.1. Gaussian Random Variables

Definition 2.1
A $\mathbb{R}^d$-**Gaussian vector** is an $\mathbb{R}^d$-valued random variable X such that $\langle X, u \rangle$ is a one-dimensional Gaussian in $\mathbb{R}$ for all $u \in \mathbb{R}^d$.

Lemma 2.2
The law of a Gaussian X are completely determined by its first and second moments, ie. the mean $\mu = \mathbb{E}[X]$ and variance $\Sigma_{ij} = \text{cov}(\langle X, e_i \rangle, \langle X, e_j \rangle)$

Lemma 2.3
We say that $X \sim \mathcal{N}(\mu, \Sigma)$ where $\Sigma \succeq 0$ if X is an $\mathbb{R}^d$-Gaussian. Let $\Sigma = AA^T$ be the Cholesky

factorization. Then, X has density:

$$p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \tag{2.1}$$

Theorem 2.4

For Gaussian, we have independence iff they have covariance zero.

2.2. Gaussian Processes and Spaces

Definition 2.5 (*Gaussian Spaces*)

A (centered) **Gaussian space** is a closed linear subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ which contains only centered Gaussians.

Definition 2.6

Let $(E, \mathcal{E})$ be a measurable space and T be an index set; a **stochastic process** with values E is a collection $(X_t)_{t\in T}$ is a collection of random variables.

Definition 2.7

A stochastic process (X_t) is called a (centered) **Gaussian process** if any finite linear combination of the variables is a centered Gaussian random variable. These variables form a closed linear subspace of $L^2(\Omega)$.

Theorem 2.8

Let H be a centered Gaussian space and (H_i) be a collection of linear subspaces of H. Then H_i are pairwise orthogonal in L^2 iff $\sigma(H_i)$ are independent.

Corollary 2.9

Let H be a Gaussian space and $K \leq H$ be a closed linear subspace. Let p_K be the orthogonal linear projection in the Hilbert Space L^2 and let $X \in H$. We have if $Y \in H$ that

$$Y \mid \sigma(K) \sim \mathcal{N}(\pi_K(Y), \mathbb{E}[(Y - \pi_K(Y))^2]) \tag{2.2}$$

2.3. Gaussian White Noise

Definition 2.10

Let $(E, \mathcal{E})$ be a measurable space and μ a σ-finite measure on $(E, \mathcal{E})$. A **Gaussian white noise** with intensity μ is an isometry G from $L^2(E, \mathcal{E}, \mu)$ to a centered Gaussian space. That is, for $f \in$

$L^2(E, \mathcal{E}, \mu)$ we have $G(f) \sim \mathcal{N}(0, \int f^2 d\mu)$.

Intuitively, each patch dA gets a small Gaussian Noise $\mathcal{N}(0, dx)$.

3. (L3-L5) Brownian Motion

We can construct the standard Brownian Motion in a few different ways: in the homework, we show Levy's Construction building from indicators while in this lecture we use the more standard Kolomogorv construction. The textbook uses the Gaussian White Noise which is equivalent and potentially more illuminating *post-hoc* but is a bit abstract as a first definition.

3.1. Pre-Brownian Motion

We begin with a set of equivalent definitions:

Definition 3.1 (*Pre-Brownian Motion*)
Let G be a Gaussian white noise on $\mathbb{R}^+$ with Lesbegue intensity μ. The random process $(B_t)_{t\in\mathbb{R}^+}$ defined by

$$B_t = G(\mathbb{1}_{[0,t]}) \tag{3.1}$$

is called a **pre-brownian motion**.

The intuition is that the time-derivative of the Brownian Motion is essentially a Gaussian White Noise, ie. an infinetisimal Gaussian.

The following shows a set of equivalent definitions are slightly more illuminating:

Theorem 3.2
Let (X_t) be a real-valued random process. The following are equivlent:

(a) (X_t) is a pre-Brownian motion.

(b) (X_t) is a centered Gaussian process with covariance $K(s, t) = s \wedge t$

(c) $X_0 = 0$ a.s. and for $0 \leq s < t$, $X_t - X_s$ is independent of $\sigma(X_r, r \leq s)$ and distributed according to $\mathcal{N}(0, t - s)$

(d) $X_0 = 0$ a.s. and for every choice of $0 = t_0 < t_1 < \cdots < t_p$ the variables $X_{t_i} - X_{t_{i-1}}$ are independent and distributed according to $\mathcal{N}(0, t_i - t_{i-1})$

We use mainly the third definition:

Definition 3.3 (*Pre-Brownian Motion*)
A **pre-brownian motion** is a stochastic process (B_t) such that $B_0 = 0$, $B_t - B_s \sim \mathcal{N}(0, t - s)$ with independent increments.

3.2. Brownian Motion

Adding a third condition, namely continuity of sample paths gives the definition of a brownian motion:

Definition 3.4 (*Brownian Motion*)
A **Standard Brownian Motion** (or simply just **Brownian Motion**) is a pre-Brownian motion (B_t) with continuous **sample paths**, ie. for a.e. $\omega \in \Omega$, $t \mapsto B_t(\omega)$ is continuous.

A naive construction by Kolomogorov's extension theorem makes no guarantees about the continuity of the sample paths (or even measurability). The pre-Brownian motion (due to the existing properties) are already nearly continuous, and by "slightly" modifying B we can ensure sample path continuity.

Definition 3.5
Let $(X_t)_{t\in T}$ and $(\tilde{X}_t)_{t\in T}$ be two random process indexed by the same set T and with values in the same metric space E. We say $\tilde{X}$ is a **modification** of X if $\forall t \in T$, $\mathbb{P}[\tilde{X}_t = X_t] = 1$. The process $\tilde{X}$ is said to be **indistinguishable** from X if there exists a set if on the other hand $\Omega \backslash \{\omega : X_t(\omega) = \tilde{X}_t(\omega), \forall t \geq 0\}$ is a null set (contained within a measure zero set, although the set itself need not be measurable).

From now on we denote a "unique process" as being up to indistinguishability. The main idea is as follows: for any interval $I \subseteq \mathbb{R}^+$, we have that $X_t = \tilde{X}_t$ a.s. for all $t \in I \cap \mathbb{Q}$. We then use a continuity argument to extend to all $t \in I$. Our proof follows from a set of two lemmas which are combined into the :

- The first, essentially states that if a dyadic function satisfies a Holder-like estimate within each level, that this extends between levels with a suitable constant.
- The second shows that for any stochastic process with a α-Holder mean, we have Holder like differences for the dyadics.

Combining the preceding two lemmas shows that for any stochastic process (X_t) that satisfies those same α-Holder mean conditions, there exists a modification $(\tilde{X}_t)$ which has α-Holder sample paths.

Theorem 3.6 (*Kolomogorov Continuity Lemma*)
Let $X = (X_t)$ be a random process indexed by a bounded interval I of $\mathbb{R}$ and taking values in a complete metric space (E, d). Assume there exists reals $q, \varepsilon, C > 0$ such that, for every $s, t \in I$,

$$\mathbb{E}[|X_s - X_t|^q] \leq C|t-s|^{1+\varepsilon} \tag{3.2}$$

Then, there is a unique (up to indistinguishability) modification $\tilde{X}$ of X whose sample paths are α-Holder continuous for every $\alpha \in (0, \frac{\varepsilon}{q})$. That is, for every $\omega \in \Omega$, $\alpha \in (0, \frac{\varepsilon}{q})$, there exists $C_\alpha(\omega)$ such that for every $s, t \in I$,

$$|\tilde{X}_s(\omega) - \tilde{X}_t(\omega)| \leq C_\alpha(\omega)|t-s|^\alpha \tag{3.3}$$

In particular, $\tilde{X}$ is a modification of X with continuous sample paths.

The construction of the Brownian motion is now straightforward:

Corollary 3.7 (*Construction of the Brownian Motion*)

Let $B = (B_t)$ be a pre-Brownian motion. The process B has a unique modifications whose sample paths are α-Holder continuous $\forall \alpha < 1/2$ (and thus continuous).

Proof. If $s < t$, we have $B_t - B_s \sim \mathcal{N}(0, t-s)$ and has the same law as $\sqrt{t-s}Z$ where $Z \sim \mathcal{N}(0,1)$. So

$$\mathbb{E}\,|B_t - B_s|^q = (t-s)^{q/2}\,\mathbb{E}\,|U|^q = C_q(t-s)^{q/2} \tag{3.4}$$

where $C_q < \infty$. Taking $q > 2$ we can apply the Kolomorov Continuity Lemma with $\varepsilon = q/2 - 1 > 0$ to get a Brownian motion with $(q-2)/(2q) = 1/2 - \frac{1}{q}$ Holder continuity. ■ □

3.3. Wiener Measure and the Canonical Brownian Motion

Having constructed the Brownian motion, we now wish to discuss some properties related to it with regards to a measure space: for example, given a simple random walk X_n scaled by $\sqrt{n}$ and time scaled by n, we want to show some notion of convergence of $(X_n) \to X$ where X is the standard Brownian motion.

Now, recall the idea of weak convergence:

Definition 3.8 (*Weak Convergence*)

We say that X_n converges **in distribution** (or **weakly**) iff $\mathbb{E}[\psi(X_n)] \to \mathbb{E}[\psi(X)]$ for all bounded continuous ψ.

To define a similar notion for stochastic processes, ie. for $(X_t^{(n)}) \to (X_t)$, we need some notion of σ-algebra and measure of our event space along with some topology (to define a valid Borel σ-algebra). More specifically, let $I := [0, \infty)$ and denote $\mathcal{C}(I) = \{\text{continuous functions } I \to \mathbb{R}\} \subseteq \mathbb{R}^I$ where the latter term denotes all functions from $I \to \mathbb{R}$.

A natural σ-algebra can be found by restricting the product σ-algebra $\mathcal{B}(\mathbb{R})^{\otimes I}$ that results from the Kolomogorov extension theorem (see Appendix A) that is generated by sets of the form

$$S = \{f \in \mathbb{R}^I : f(t_1) \in A_1, \cdots, f(t_k) \in A_k; 0 \le t_1 < \cdots < t_k < \infty\} \tag{3.5}$$

Now, because we want continuous sample paths, we restrict to the σ-algebra on continuous functions, ie. we define:

$$\mathcal{G} := \mathcal{B}^{\otimes I} \mid_{C(I)} = \{C(I) \cap A : A \in \mathcal{B}^{\otimes I}\} \tag{3.6}$$

We then arrive at a natural notion of measure:

Definition 3.9 (*Canonical Brownian Motion*)

Let $\mathcal{B} = (B_t)$ be a standard brownian motion. $\mathcal{B}$ defines a measurable mapping $\mathcal{B} : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{C}(I), \mathcal{G}, \mathcal{P})$ where we define the measure to be the pushforward of $\mathbb{P}$:

$$\mathcal{P} := \mathcal{B}_*(\mathbb{P}) = \mathbb{P} \circ \mathcal{B}^{-1} \tag{3.7}$$

We call $\mathcal{P}$ the **Wiener measure** and elements of $(\mathcal{C}(I), \mathcal{G}, \mathcal{P})$ a **canonical Brownian Motion**.

This measure is in fact unique:

Lemma 3.10
The Wiener measure is unique.

Proof. We have by definition

$$\mathcal{P}(B_{t_1} \in A_1, \cdots, B_{t_n} \in A_n) = \mathbb{P}(X_{t_1} \in A_1, \cdots, X_{t_n} \in A_n) \tag{3.8}$$
$$= \int_{\prod A_j} p_{t_1}(x_1) \prod_{i=2}^{n} p_{t_i - t_{i-1}}(x_i - x_{i-1}) dx_1 \cdots dx_n \tag{3.9}$$

This characterizes $\mathcal{P}$ on a π-system generating $\mathcal{G}$ and so it is unique. ■ □

3.4. The Wiener Measure as a Borel Measure

Alternatively, we can interpret $\mathcal{G}$ as being a Borel space arising from a natural topology of $\mathcal{G}$.

A natural candidate for a topology on $C(I)$ is a single uniform norm on the entire space, eg. something like

$$\|f\| = \sup_{T<\infty} \sup_{t\in[0,T]} |f(t)| \tag{3.10}$$

. Unfortunately this does not work as $I = [0, \infty)$ is not compact so the norm can be unbounded.

The natural remedy is the uniform norm on compact sets:

$$d(f, g) = \sum_{n\geq 1} \min\{1, \sup\{|f(t) - g(t)| : 0 \leq t \leq n\}\}/2^n \leq 1 \tag{3.11}$$

Theorem 3.11
$\mathcal{G} = \mathcal{B}^{\otimes I} \mid C_i$ and the Borel σ-field on $C(I)$ with the d-topology coincide.

3.5. Properties of Brownian Motion

3.5.1. Basic Properties

We begin with some basic facts about Brownian Motions:

Theorem 3.12 (*Basic Properties of pre-Brownian Motion*)
Let $B := (B_t)$ be a brownian motion. We have

(a) (Symmetry Property) $-B$ is also a Brownian motion

(b) (Scaling Property) For every $\lambda > 0$, the process $B_t^\lambda = \frac{1}{\lambda} B_{\lambda^2 t}$ is also a Brownian motion.

3.5.2. Brownian Motion as a Markov Process and the Reflection Principle

More critical is the simple Markov property: that is, a Brownian motion is a Markovian process.

Theorem 3.13 (*Simple Markov Property*)
For every $s \geq 0$ the process $B^s_t = (B_{s+t} - B_s)$ is a pre-Brownian motion and is independent of $\sigma(B_r; r \leq s)$.

Proof. The proof is easy and is seen just by noticing that the increments $B^{(s)}_{t_i} - B^{(s)}_{t_{i-1}} = B_{s+t_i} - B_{s+t_{i-1}}$. □

In fact, we can improve on this slightly as follows by using continuity:

Lemma 3.14 (*Simple Markov Property*)
Suppose (B_t) is a simple Brownian motion in $\mathbb{R}^d$. For any fixed $s \geq 0$, the process

$$(W_t)_{t\geq 0} := (B_{s+t} - B_s)_{t\geq 0} \tag{3.12}$$

(this is a shifted Brownian motion) is also a standard Brownian Motion and is independent of $\mathcal{F}_s$. In fact, we have that $W \perp \mathcal{F}_{s+}$.

This leads to the following statement, which roughly states that any right-continuous process with the Simple Markov Property (more generally called a Feller process) has a deterministic beginning:

Theorem 3.15 (*Blumenthal's 0-1 Law*)
If $A \in \mathcal{F}_{0+}$ then $\mathbb{P}(A) \in \{0, 1\}$.

Proof. This follows from the Simple Markov Property which implies that $\mathcal{F}_{0+}$. From here, note that for any $A \in \mathcal{F}_{0+}$ is then independent of itself and so we get that it is trivial. □

Brownian motions additionally has the following interesting property of hitting times, namely that the hitting time for any distance is finite although the expectation is infinite.

Proposition 3.16 (*Hitting Times of Brownian Motion*)
Let (B_t) be a standard Brownian Motion.

1. For every $\varepsilon > 0$, a.s. $\sup(B_s; 0 \leq s \leq \varepsilon) > 0$ and $\inf(B_s; 0 \leq s \leq \varepsilon) < 0$, ie. it crosses zero infinitely many times early on.
2. For all $a \in \mathbb{R}$ the hitting time $T_a = \inf\{t \geq 0 : B_t = a\}$ is finite, ie. $\limsup_{t\to\infty} B_t = \infty$ and $\liminf_{t\to\infty} B_t = -\infty$.

Proof. (a) Take $\varepsilon_p \downarrow 0$ and note $A := \cap_{p=1}^{\infty} \sup_{0\leq s\leq \varepsilon_p} B_s > 0 \in \mathcal{F}_{0+}$. Then note that

$$\mathbb{P}(\sup_{0\leq s\leq \varepsilon} B_s > 0) \geq \mathbb{P}\left(\bigcap_p \sup_{0\leq s\leq \varepsilon_p} B_s > 0\right) \geq 1/2 \tag{3.13}$$

and so by Blumenthal's $0-1$ law it is 1. The infimum case holds similarly but with $-B$.

(b) Use scaling to finish. ■

□

Corollary 3.17

(B_t) is almost surely not nondecreasing on a nontrivial interval.

Proof. Apply the previous corollary on the nontrivial window and use the simple Markov Property. □

In fact, the simple Markov Property can be extended to Stopping Times under some mild assumptions. An important fact will be the following:

$$\{T < t\} = \bigcup_{q \in [0,t)\cap\mathbb{Q}} \{T \leq q\} \in \mathcal{F}_t \tag{3.14}$$

Theorem 3.18 (*Strong Markov Property*)

Let T be a stopping time. We assume $\mathbb{P}(T < \infty) > 0$. Letting

$$B_t^{(T)} := \mathbb{1}_{T<\infty}(B_{T+t} - B_T) \tag{3.15}$$

we have that under the probability measure $\mathbb{P}(\cdot | T < \infty)$ the process (B_t^T) is a Brownian Motion independent of $\mathcal{F}_T$.

Proof. The idea is just a simple extension of the simple Markov Property using dominated convergence on an approximated dyadic set. See Le-Gall page 35. □

One such application of the Strong Markov Property is the Reflection Property. The intuitive picture is as follows:

Theorem 3.19 (*Reflection Principle*)

For every $t > 0$, set $S_t = \sup_{0\leq s\leq t} B_s$. If $a \geq 0$ and $b \leq a$ then

$$\mathbb{P}(S_t \geq a, B_t \leq b) = \mathbb{P}(B_t \geq 2a - b) \tag{3.16}$$

This effectively gives the joint law of (S_t, B_t).

A straightforward corollary is as follows:

Corollary 3.20 (*Hitting Time has Infinite Mean*)

For every $a > 0$, T_a is equidistributed as a^2/B_1^2 and has density:

$$\pi(a) = \frac{a}{\sqrt{2\pi t^3}} \exp\left(-\frac{a^2}{2t}\right) \mathbb{1}_{t>0} \tag{3.17}$$

Thus, taking an expectation gives that $\mathbb{E}[T_a] = \infty$.

Some additional distributions that are explored in Pset 2:

Theorem 3.21

Let B_t be the standard Brownian Motion, $S_t = \sup_{0 \le s \le t} B_s$, $Y_t := S_t - B_t > 0$. We have that

(a) $S_t \overset{(d)}{=} |B_t| \overset{(d)}{=} S_t - B_t$ for each fixed $t \ge 0$

(b) In fact, $(B_t)_{t \ge 0} \overset{(d)}{=} (S_t - B_t)_{t \ge 0}$ as processes.

3.5.3. The Quadratic Variation of Brownian Motion

Although Brownian Motion is α-Holder for any $\alpha \in (0, 1/2)$ on any fixed time interval, one can in fact show that Brownian Motion is not $1/2$-Holder on any nontrivial interval.

Theorem 3.22 (*Levy's Modulus of Continuity Theorem*)

We have for a standard Brownian Motion (B_t) on $[0, 1] \times \Omega \to \mathbb{R}$ that

$$\lim_{h \to 0} \sup_{|t - t'| \le h; t, t' \le 1} \frac{|B_{t'} - B_t|}{\sqrt{2h \log(1/h)}} = 1 \tag{3.18}$$

The proof relies on the Law of the Iterated Logarithmn that we show in Pset2. Intuitively, the statement lies somewhere between the CLT and the (strong) Law of Large Numbers:

Theorem 3.23 (*Law of the Iterated Logarithmn*)

Let (B_t) be a standard BM. Then,

$$\limsup_{t \to \infty} \frac{B_t}{\sqrt{2t \log \log t}} = 1 \tag{3.19}$$

a.s.. The corresponding infimum is given by -1.

However, in some sense B_t is $1/2$-Holder "on-average". More specifically, the Quadratic Variation of the Brownian Motion can be shown to be exactly t.

Theorem 3.24 (*Quadratic Variation of Brownian Motion*)

Let $P_n = \{0 = t_0 < t_1 < \cdots < t_p\}$ be a sequence of increasing subdivisions of $[0, t]$ such that $\mathrm{mesh}(P_n) \to 0$ and let $V^2_{P_n}(B) := \sum_{i=1}^{p} (B_{t_i} - B_{t_{i-1}})^2$. Then, $V^2_{P_n} \xrightarrow{L^2} t$.

Proof. This follows straightforwardly from remarking that $\Delta B_i^2 - \Delta t_i$s are independent by considering their convariances. $\square$

That is, $\langle B \rangle_t = t$ (this is known as the Quadratic Variation).

COROLLARY 3.25

Brownian Motion is not a finite variation process (ie. the total variation is infinite on any compact interval).

Proof. By the Simple Markov Property, it suffices to consider the interval $[0, t]$. Now suppose it was finite variation. Then

$$\sum_{i=1}^{n}(B_{t_i} - B_{t_{i-1}})^2 \leq \sup |B_{t_i} - B_{t_{i-1}}| \sum_{i=1}^{n} |B_{t_i} - B_{t_{i-1}}| \tag{3.20}$$

Taking the mesh to 0, we ge the LHS converges to t while the RHS converges to 0, a contradiction. So it is not finite variation. ■ □

The above is the main reason why we can not directly define the stochastic integral for a Brownian Motion as a simple Stiltjes integral. We show later, however, that a suitable definition of the stochastic integral is possible by defining it as a decomposition through a quadratic variation term and a

4. (L6-L9): Continuous-Time Martingale Theory

We now move onto a continuous-time martingale theory. The basic idea is that under mild regularity assumptions, a process has a cadlag modification (defined below), and all results of discrete-time martingale theory carry over to the continuous setting.

4.1. Filtrations and Processes

We first define a filtration in continuous time:

DEFINITION 4.1

A **filtration** on $(\Omega, \mathcal{F}, \mathbb{P})$ is a collection $(\mathcal{F}_t)_{0 \leq t \leq \infty}$ of nondecreasing sub σ-algebras of $\mathcal{F}$. The **canonical filtration** of random process (X_t) is $\mathcal{F}_t := \sigma(X_s; s \leq t)$.

We will now define notions of "completion" in order to contain all negligible sets and "continuity" which will come useful in order to ensure some regularity properties in order to make much of our stopping time machinery work. The textbook uses the stopping time definition with open intervals $(T < t)$ which will require the use of the right-continuous completion but the lecture uses the closed intervals which can use simply our original filtration.

We can define the infinitesemal look ahead:

DEFINITION 4.2 (*RIGHT-CONTINUOUS FILTRATIONS*)

Define $\mathcal{F}_{t+} := \cap_{s>t} \mathcal{F}_s$. This also defines a filtration. We say a filtration $(\mathcal{F}_t)$ is **right-continuous** if $\mathcal{F}_t = \mathcal{F}_{t+}$ for all $t \geq 0$.

The intuition is mainly as follows: consider a Brownian Motion B_t with resepct to its canonical filtration. Let U be some open set. Now, $\sigma := \inf\{t : B_t \in U\}$ may not be a stopping time with respect to $\mathcal{F}_t$ (although it is with respect to $\mathcal{F}_{t+}$) since it is possible that the Brownian Motion reaches the boundary at time t but can either enter or not which makes $\{T \leq t\}$ not measurable with respect to the filtration. So using the right conitnuous filtration will be preferable in allowing certain stopping times.

Definition 4.3

Let $(\mathcal{F}_t)$ be a filtration and let $\mathcal{N}$ be the class of all $(\mathcal{F}_\infty, \mathbb{P})$-negligible sets (ie. $A \in \mathcal{N}$ if there exists $A' \in \mathcal{F}_\infty$ such that $A \subset A'$ and $\mathbb{P}(A') = 0$). A filtration $\mathcal{F}$ is said to be **complete** if $\mathcal{N} \subseteq \mathcal{F}_0$.

If a filtration is not complete, it can be completed by augmenting as follows:

Definition 4.4 (*Completion*)

The **canonical completion** or **augmentation** of a filtration $(\mathcal{F}_t)$ is given by taking $\mathcal{F}_t' := \mathcal{F}_t \vee \sigma(\mathcal{N})$ where $\vee$ takes the smallest σ-algebra generated by the union.

Some additional notions of measurability which will be required for our continuous theory.

Definition 4.5 (*Measurability of a Process*)

A process $X = (X_t)$ is said to be **measurable** if $(\omega, t) \mapsto X_t(\omega)$ is measurable.

Definition 4.6

A process $X = (X_t)$ that is both adapted and measurable is said to be **progressive**

Intuitively, a progressive process has a filtration such that it is previsible, ie. it can not "look" ahead into the future.

Theorem 4.7

Let (X_t) be an adapted process. If X is right-continuous then it is progressive (and similar if we replace with left continuity)

Proof. One can approximate the Borel Sets with a union of countable rationals and then take a pointwise limit. □

Definition 4.8

The **progressive σ-field** is the collection $\mathcal{P}$ of all sets $A \in \mathcal{F} \otimes \mathcal{B}(\mathbb{R}^+)$ such that the process $X_t(\omega)$ is progressive.

We next define a notion of continuity for the sample paths of stochastic processes which will allow generalization of our discrete time martingale results:

Definition 4.9 (*Càdlàg*)

A stochastic process $(N_t)_{t\geq 0}$ is said to be **rcll** or **càdlàg** if $\lim_{s\downarrow t} N_s = N_t$ and the left limit $\lim_{s\uparrow t} N_s$ exists.

The main thing we want is for our martingale theory to cover two main cases: a brownian motion (B_t) and processes of the form $(N_t - t)$ where N is a Poisson process. Fortunately, both these processes are cadlag.

4.2. Stopping Times

We now review key facts about martingales and extend them to the continuous time setting. All results carry over as long as we have right continuity.

Definition 4.10

A random variable $\tau : \Omega \to [0, \infty)$ is a **stopping time** with regards to $(\mathcal{F}_t)$ if $\{\tau \leq t\} \in \mathcal{F}_t$ for all t. The σ-field of the past up to τ is defined by $\mathcal{F}_\tau := \{A \in \mathcal{F}_\infty : A \cap \{\tau \leq t\}, \forall t \geq 0\}$. One can alternatively define them through the right continuous completion and the open half-intervals instead.

The main properties of stopping times that we need to be familiar about are as follows:

Theorem 4.11 (*Basic Properties of Stopping Times*)

Let $\mathcal{F} = (\mathcal{F}_t)$ be a filtration. Then,

(a) For stopping times $\sigma \leq \tau$ a.s. we have $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.

(b) If σ, τ are stopping times then so is $\sigma \wedge \tau$ and $\sigma \vee \tau$ with $\mathcal{F}_{\sigma \wedge \tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$

(c) If time is discrete, we have for a stopping time σ that X_σ is $\mathcal{F}_\sigma$-measurable.

Proof. We prove only the last property. Note

$$\{X_\sigma \in A\} \cap \{\sigma \leq n\} = \cup_{k \leq n}(\{\sigma = k\} \cap \{X_k \in A\}) \in \mathcal{F}_n \implies \{X_\sigma \in A\} \in \mathcal{F}_\sigma \tag{4.1}$$

□

Note we have that a) and b) hold in both continuous and discrete time. c) on the other hand does not necessarily follow in continuous time as we can not take an uncountable union and we will thus need to have additional assumptions.

Definition 4.12 (*Martingales*)

Let (X_t) be an adapted real-valued process on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$. We say that (X_t) is a **submartingale** (respectively super) if

1. It is integrable $\mathbb{E}\,|X_t| < \infty$ for all $t < \infty$,
2. For every $0 \leq s \leq t$, $X_s \leq \mathbb{E}[X_t | \mathcal{F}_s]$

If a process is both a sub and supermartingale it is said to be a **martingale**.

Consequently in a submartingale the $\mathbb{E}\,X_t$ is non-decreasing while it is non-increasing in the supermartingale case. The way to remember the direction is by thinking that the current value of the function is "less" than the average of your step function.

Theorem 4.13 (*Jensen's Inequality for Martingales*)

If X_t is a submartingale and f is a nondecreasing convex function, then $\forall s \leq t$,

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] \geq f(X_s) \tag{4.2}$$

Namely, we have $f(X_t)$ is also a submartingale.

An important corollary that will become useful in some of the proofs to follow:

Corollary 4.14

If (M_t) is a martingale, $|M_t|$ is a submartingale (namely $|M_t|^p$ is a submartingale for all $p \geq 1$ provided that $M_t \in L^p$). $(M_t)^+ = M_t \wedge 0$ is also a submartingale.

4.3. Doob Martingales and Examples of Non-Doob Martingales

The most intuitively obvious example of continuous-time martingales are as follows:

Definition 4.15 (*Doob Martingale*)

A martingale (X_t) is said to be **closed** (or a **Doob martingale**) iff $\exists Z \in L^1$ such that $X_t = \mathbb{E}[Z \mid \mathcal{F}_t]$ for all $t \geq 0$.

The fact that such a process is indeed a Martingale follows from the Tower Law along with the fact that

$$\mathbb{E}\,|X_t| = \mathbb{E}\,|\mathbb{E}[Z_t \mid \mathcal{F}_t]| \leq \mathbb{E}\,|Z| < \infty \tag{4.3}$$

by conditional Jensen. Not all martingales are, however, of closed or Doob type.

Example 4.16

Brownian Motion is not a closed martingale.

Proof. Note

$$\mathbb{E}\,|B_t| = t^{1/2}|\mathcal{N}(0,1)| \to \infty \tag{4.4}$$

as $t \to \infty$ which is unbounded and hence is not an L^1-bounded process. So it is not closed. □

An important class of such examples (which includes Brownian Motion as a subcase) are generated from processes with independent increments.

Definition 4.17 (*Processes with Independent Increments*)

Consider a real-valued random process $Z = (Z_t)_{t\geq 0}$ adapted to $(\mathcal{F}_t)$. We say that Z has **independent increments** if $Z_t - Z_s \perp \mathcal{F}_s$ for $t \geq s > 0$.

An important class of martingales arises from such types of martingales:

Lemma 4.18

Suppose Z is an $(\mathcal{F}_t)_{t\geq 0}$-adapted real-valued stochastic process with independent increments. We have that

(a) (Centered Process) $Z_t \in L^1$ for every $t \geq 0$, the process $\tilde{Z}_t := Z_t - \mathbb{E}[Z_t]$ is a martingale.

(b) (Centered Variance) If $Z_t \in L^2$ for every $t \geq 0$, the process $Y_t := \tilde{Z}_t^2 - \mathbb{E}[\tilde{Z}_t^2]$ is a martingale.

(c) (Scaled Exponential) If for some $\theta \in \mathbb{R}$ we have $\mathbb{E}[\exp(\theta Z_t)] < \infty$ for every $t \geq 0$, then

$$X_t := \frac{e^{\theta Z_t}}{\mathbb{E}[e^{\theta Z_t}]} \tag{4.5}$$

is a martingale.

The proofs of these are easy: simply show the martingale property.

Corollary 4.19

The processes B_t, $B_t - t$, and $\exp(\theta B_t - \theta^2 t/2)$ are all martingales. The latter types are called the **exponential martingales** of Brownian motion.

4.4. Basic Martingale Properties

We have the following basic result showing that sub/super martingales are L^1-bounded over compact time intervals.

Theorem 4.20

Let (X_t) be a submartingale. Then $\sup\{\mathbb{E}\,|X_s| : 0 \leq s \leq t\} < \infty$ for all $t < \infty$.

Proof. Note $f(x) = x_+ = x \vee 0$ is nondecreasing and convex. From there, simply decompose X_t into positive and negative parts. □

We then show Doob's Martingale Inequalities which will be be helpful in bounding the "size" or "growth" of our martingales. We prove them first for the discrete case and then extend to the continuous-time case using an approximation argument.

Theorem 4.21 (*Doob's Maximal Inequalities in Discrete Time*)

Let (X_n) be a discrete time sub or supermartingale and $\lambda > 0$. Then,

$$\lambda \mathbb{P}\left(\max_{k \leq n} |X_k| \geq \lambda\right) \leq \mathbb{E}[|X_0| + 2|X_n|] \qquad \forall 0 \leq n < \infty, \lambda \geq 0 \tag{4.6}$$

Proof. Let $\tau = \min\{k \geq 0 : k = n \text{ or } |X_n| \geq \lambda\}$ so $|X_\tau| \geq \lambda$ iff $\max\{X_k : 0 \leq k \leq n\} \geq \lambda$. The optional stopping theorem concludes. □

If (X_n) is a martingale we have a slight improvement:

Corollary 4.22

If (X_n) is a discrete time martingale then

$$\lambda \mathbb{P}\left(\max_{k\leq n} |X_k| \geq \lambda\right) \leq \mathbb{E}\,|X_n| \tag{4.7}$$

This immediately implies that for any countable set D that we have the maximal inequality holds, ie.

Corollary 4.23 (*Doob's Maximal Inequality for Countable Subsets*)
Let (X_n) be a discrete time sub or supermartingale and $\lambda > 0$. Let D be a countable subset of $\mathbb{R}$. Then,

$$\lambda \mathbb{P}\left(\sup_{s\in[0,t]\cap D} |X_s| \geq \lambda\right) \leq \mathbb{E}[|X_0| + 2|X_t|] \qquad \forall 0 \leq n < \infty, \lambda \geq 0 \tag{4.8}$$

The maximal inequality still holds if we have right continuity.

Theorem 4.24 (*Doob's Maximal Inequality in Continuous Time*)
Let $(X_t)_{t\geq 0}$ be a submartingale with right-continuous sample paths. Then for $\lambda \geq 0$,

$$\lambda \mathbb{P}(\sup_{0\leq s\leq t} |X_s| \geq \lambda) \leq \mathbb{E}\,|X_0| + 2\,\mathbb{E}\,|X_t| \quad \forall 0 \leq t < \infty \tag{4.9}$$

Proof. Fix t and note for any sequence $0 = t_0 < t_1 < \cdots < t_p = t$, by Doob's Maximal Inequality in Discrete Time we have that $\lambda \mathbb{P}(\max_{0\leq k\leq p} |X_{t_k}|) \leq \mathbb{E}(|X_0| + 2|X_t|)$. Take $D_n \uparrow D$ where D is a countable subset of $[0, t]$ containing 0 and t. We can use monotone convergence to get

$$\lambda \mathbb{P}(\sup_{D\cap[0,t]} |X_s| \geq \lambda) \leq \mathbb{E}(|X_0| + 2|X_t|) \tag{4.10}$$

Then, by right continuity, for any $s \in [0, t]$, we have that we can choose $s_n \downarrow s$ in D (by density) $X_{s_n} \to X_s$ so also $\mathbb{E}[X_{s_n}| \to \mathbb{E}\,|X_s|$ since we have compact sets (so by bounded convergence) while additionally $\sup_{D\cap[0,t)} |X_s| = \sup_{[0,t]} |X_s|$ and since we assume $t \in D$ we are done. □

We can use the same argument for L^p inequalities

Theorem 4.25 (*Doob's L^p-inequality*)
Fix $p > 1$. Let (X_n) be either a martingale in discrete time or a right continuous martingale in continuous time. We have for $0 \leq n < \infty$

$$\left\| \max_{0\leq k\leq n} |X_k| \right\|_p \leq C_p \|X_n\|_p \tag{4.11}$$

where $C_p := p/(p-1)$.

Proof. Let $S_n := \max_{0\leq k\leq n} |X_k|$. Assume $S_n \in L^p$ for simplicity. Then, by Doob's Maximal inequality

$$
\begin{aligned}
\mathbb{E}[S_n^p] &= \int_0^\infty p y^{p-1} \mathbb{P}(S_n \geq y) dy && (4.12)\\
&\leq \int_0^\infty p y^{p-1} \frac{\mathbb{E}[X_n; S_n \geq y]}{y} dy && (4.13)\\
&= \frac{p}{p-1} \mathbb{E}\left[|X_n| \int_0^{S_n} (p-1) y^{p-2} dy \right] && (4.14)\\
&= C_p \mathbb{E}\, |X_n| (S_n)^{p-1} \leq C_p \|X_n\|_p \|S_n\|_p^{p-1} && (4.15)
\end{aligned}
$$

The assumption of L^p-boundedness can be removed from truncation. □

The L^p inequality in the case of $p = 1$ does not hold:

Example 4.26 (*Doob's L^p inequality fails for $p = 1$*)

Let X_n be a simple random walk on $\mathbb{Z}$, let $X_0 = 1, \tau = \min\{n : X_n = 0\}$. Then $M_n := X_{n\wedge\tau}$ is a non-negative martingale with $\|M_n\|_1 = 1$ by the optional stopping theorem. However, $\|\max_{0\leq k\leq n} M_k\|_1$ is unbounded since $\|\max_{0\leq k\leq n} M_k\|_1 \uparrow \mathbb{E}[\max_{k\geq 0} M_k]$ by MCT. But the right hand side goes to ∞ by Borel-Cantelli since $\sum \mathbb{P}(\max_{k\geq 0} M_k \geq a) = \sum 1/a = \infty$ where we use OST or a reflection principle type of argument so the sum goes to ∞.

An additional useful fact:

Theorem 4.27 (*L^p martingale convergence theorem*)

Let (X_n) be a submartingale with each $X_n \in L^p$ where $p > 1$ in either discrete time or continuous time where $t \mapsto X_t$ is right continuous. Then $X_n \to X$ both a.s. and in L^p for some $X \in L^p$.

4.5. Optional Stopping Theorems

The following properties now hold only in discrete time unless the appropriate modifications are made for continuous time.

Theorem 4.28 (*Optional Stopping Theorem in Discrete Time*)

Let (X_n) be a submartingale in discrete time and τ be a bounded stopping time ($\tau \leq n$ a.s.). Then,

$$\mathbb{E}\, X_0 \leq \mathbb{E}\, X_\tau \leq \mathbb{E}\, X_n \tag{4.16}$$

Proof. Define $Y_k = X_{k\wedge\tau}$ to be the "stopped process". It is a submartingale and we get the desired result. □

4.6. Sample Path Regularity via Cadlag Modification

We now show that under certain mild conditions, namely that our filtration is chosen to be right continuous and complete, while our martingale has right continuous expectations, we have that our continuous time martingales will have a cadlag modification which will allow us to specify our martingales by their finite dimensional martingales as in the discrete time case.

Definition 4.29 (*Upcrossing Number*)
Let $I \subseteq [0, \infty)$ and $f : I \to \mathbb{R}$ be given. Take any $-\infty < a < b < \infty$. Now, we define the **upcrossing number**

$$U^f_{a,b} = \sup\{k : \exists s_1 < t_1 < \cdots < s_k < t_k\} \tag{4.17}$$

with $f(s_i) \leq a$ and $f(t_i) \geq b$ for all i.

The upcrossing number can of course be infinite, even for bounded I. The following result shows that control on the upcrossings implies regularity of the process, ie. there exists left and right limits.

Lemma 4.30 (*Control on upcrossings implies regularity*)
Let D be a countable dense subset of $[0, \infty)$ and $f : D \to \mathbb{R}$. Suppose

(a) $\sup\{|f(t)| : t \in D \cap [0, T]\} < \infty$ for all $T \in D$.

(b) $U^f_{a,b}(D \cap [0, T]) < \infty$ for all $T \in D$ and $\forall -\infty < a < b < \infty$ with $a, b \in \mathbb{Q}$.

Then $f(t+) = \lim_{s \downarrow t, s \in D} f(s)$ and $f(t-) = \lim_{s \uparrow t, s \in D} f(s)$ exist and are finite for all t and $g(t) = f(t+)$ is cadlag.

Proof. Suppose $f(t+)$ doesn't exist. Then, we have there exists rational $a < b$ such that

$$\liminf_{s \downarrow t, s \in D} f(s) < a < b < \limsup_{s \downarrow t, s \in D} \implies U^f_{a,b}(D \cap [0, T]) = \infty \quad \forall T > t \tag{4.18}$$

This is a contradiction so we have that the right limit exists and is finite (by a) and so we are done. □

We now define the Doob H-transform, which intuitively represents a "betting strategy" on the underlying process X.

Definition 4.31
On an adapted probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$ we say that X_n is **previsible** if $X_n \in \mathcal{F}_{n-1}$. If X_n is adapted and H_n is previsibile, we call the **Doob transform** or **H-transform** of X, the process $Y = H \star X$, where we define

$$Y_n = (H \star X)_n := \sum_{k=1}^{n} H_k(X_k - X_{k-1}) \in \mathcal{F}_n \tag{4.19}$$

The following result then shows that our betting strategy can not "game" the system if we have a supermartingale which has nonincreasing expectations, with any legal (privisible) strategy H.

Lemma 4.32 (*Doob transform maintains martingale properties*)
If X_n is a supermartingale and H is previsible and nonnegative, then $Y = H \star X$ is also a supermartingale.

Proof. $\mathbb{E}[Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}] = H_n \, \mathbb{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}] \geq 0$. □

We next show a bound on the number of upcrossings of a supermartingale. The main idea is that because supermartingales are nonincreasing in expectation, the upcrossings are bounded.

Lemma 4.33 (*Doob's Upcrossing Inequality in Discrete Time*)
Suppose (X_n) is a supermartingale in discrete time. We have that for any $a < b \in \mathbb{R}$ that

$$\mathbb{E}\, U_{a,b}^X([0,n]) \leq \frac{\mathbb{E}[(X_n - a)_-]}{b-a} \tag{4.20}$$

Proof. Let $\sigma_0 = \tau_0 = 0$ and define stopping times $0 = \tau_0 = \sigma_0 < \tau_1 < \sigma_1 < \cdots$ such that $\sigma_i = \min\{n \geq \tau_i : X_n \leq a\}$ and $\tau_i = \min\{n \geq \sigma_i : X_n \geq b\}$ is defined analogously. Define the betting strategy $H_j := \mathbb{1}[j \in (\sigma_i, \tau_i]$ for some $i \geq 1]$, ie. bet only during upcrossings. Clearly H_j is previsible. We thus have that $Y = H \star X$ is a supermartingale so $0 = \mathbb{E}\, Y_0 \geq \mathbb{E}\, Y_n$. On the other hand, remark

$$Y_n = (b-a)U_{a,b}^X([a,b]) + H_n(X_n - X_{n-1}) \geq (b-a)U_{a,b}^X([a,b]) - (X_n - a)_- \tag{4.21}$$

Taking expectations on both sides yields the result. □

Corollary 4.34
Let X_t be a supermartingale and D a countable dense subset of $[0,\infty)$. Then, there exists a negligible event $N \subseteq \Omega$ such that $\forall \omega \notin N$, $\varphi : t \mapsto X_t(\omega)$, we have

(a) $\sup\{|\varphi(t)| : t \in D \cap [0,T]\} < \infty$ for all $T \in D$.

(b) $U_{a,b}^\varphi(D \cap [0,T]) < \infty$ for all $T \in D$ and $\forall -\infty < a < b < \infty$ with $a, b \in \mathbb{Q}$.

Proof. The maximal inequality implies that $\sup\{|X_t| : t \in [0,T] \cap D\} < \infty$ a.s. □

Corollary 4.35 (*Doob's Upcrossing Inequality for Countable Dense Subsets*)
Let D be a countable dense subset of $[0,\infty)$. Then,

$$\mathbb{E}\, U_{a,b}^X([0,T] \cap D) \leq \mathbb{E}[(X_T - a)_-]/(b-a) < \infty \tag{4.22}$$

for all rational $a < b$.

From the above, we have shown that we can approximate on a countably dense subset for which both right and left limits exist, allowing us to get a cadlag process as follows:

Corollary 4.36
Let (X_t) be a supermartingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$. Let D be a countably dense subset and define

$$Y_t = X_{t+}(\omega) := \lim_{s \downarrow t, s \in D} X_s(\omega) \text{ if it exists or } 0 \text{ otherwise} \tag{4.23}$$

We have that Y_t is cadlag and adapted to $\mathcal{G}_t := \mathcal{F}_{t+}$

Proof. This is straightforward from Corollary 4.35. □

What is left to show is that Y is a supermartingale with respect to $\mathcal{G}_t$ and that it is indeed a valid modification. The first will follow straightforwardly, while the second will require some additional assumptions.

We begin with the first property. We first define the backwards martingale and make some comments about additional properties:

Definition 4.37

We define a **backwards (sub/super)martingale** to be a (sub/super)martingale indexed by $\mathbb{Z}_{\leq 0}$.

Theorem 4.38 (L^1 *backwards-martingale convergence theorem*)

If (X_n) is a backwards submartingale that is L^1-bounded, then $X_n \to X_{-\infty} \in L^1$ a.s. and in L^1 as $n \to -\infty$ as $n \to \infty$.

We will give two proofs of this claim.
Proof 1 (L^1 backwards implies UI): We show that an L^1-bounded backwards submartingale (X_n) is UI. First, note that $(X_n)_+$ is also a backwards submartingale and so $\mathbb{E}\, X_0 \geq \mathbb{E}\, X_{-1} \geq \cdots \mathbb{E}\, X_{-n}$ so $\mathbb{E}\, X_n \downarrow C < \infty$ as $n \to \infty$ since we assume L^1-boundedness.

Now, for $(X_n)_-$

The second proof uses the so called **Doob decomposition** which will lend nicely to the theory of continuous semimartingales later in the course.

Proof 2 (Doob Decomposition): Fix $-\infty < n \leq m \leq 0$. We have

$$X_m - X_n = \sum_{k=n+1}^{m} X_k - X_{k-1} = \sum_{k=n+1}^{m} \Delta_k \tag{4.24}$$

where $\Delta_k \in \mathcal{F}_k$. We have since (X_n) is a submartingale that $X_{k-1} \leq \mathbb{E}[X_k \mid \mathcal{F}_{k-1}]$ so $\mathbb{E}[\Delta_k \mid \mathcal{F}_{k-1}] \geq 0$. Now, consider the following decomposition (which we call the **Doob decomposition**):

$$X_m - X_n = \underbrace{\sum_{k=n+1}^{m} (\Delta_k - \mathbb{E}[\Delta_k \mid \mathcal{F}_{k-1}])}_{M_{n,m}} + \underbrace{\sum_{k=n+1}^{m} \mathbb{E}[\Delta_k \mid \mathcal{F}_{k-1}]}_{A_{n,m}} \tag{4.25}$$

Note by the submartingale property each element in the sum of $A_{n,m}$ is ≥ 0 and $\mathcal{F}_{k-1}$-measurable. So we have the $A_{n,m}$ are monotone in m. On the other hand, note that each $M_{n,m} \in \mathcal{F}_m$ and $\mathbb{E}\, M_{n,m} = 0$. We thus have that letting $A_m := \lim_{n \to -\infty} A_{m,n} \in L^1$ by MCT and since $\lim_{n \to -\infty} X_n = X_{-\infty} \in L^1$ by L^1-a.s. martingale convergence, we have $M_m := \lim_{n \to -\infty} M_{m,n} \in L^1$ and is a reverse-martingale since

$$\mathbb{E}[M_{m+1} - M_m \mid \mathcal{F}_m] = \mathbb{E}[\Delta_{m+1} - \mathbb{E}[\Delta_{m+1} \mid \mathcal{F}_m]] = \tag{4.26}$$

Theorem 4.39 (*The right limit is a martingale*)

If X_t is an adapted supermartingale with respect to $\mathcal{F}_t$ and $Y_t(\omega) := X_{t+}(\omega)$, then Y_t is an adapted supermartingale with respect to $\mathcal{G}_t = \mathcal{F}_{t+}$.

Proof. Both integrability and the martingale properties follow directly from the L^1 martingale convergence theorem. □

The second point to address to get a cadlag modification is to check whether or not the right limit is indeed a modification at all. Indeed, consider the following canonical example:

Example 4.40

Consider the deterministic process $X_t = 1 - \mathbb{1}[t > 1]$. X_t is a supermartingale but it has no cadlag modification.

Proof. Any modification must almost surely agree at $t = 1$ but this implies discontinuity. □

Lemma 4.41

Let (X_t) be an adapted supermartingale with respect to $\mathcal{F}_t$. Defining $Y_t := X_{t+}$, we have $X_t \geq \mathbb{E}[Y_t \mid \mathcal{F}_t]$ with equality if $t \mapsto \mathbb{E}\, X_t$ is right continuous.

Proof. Take $s_k \downarrow t$ in D and remark by backwards L^1 martingale convergence that $X_{s_k} \to Y_t$ a.s. and in L^1 so $X_t \geq \mathbb{E}[Y_t \mid \mathcal{F}_t]$. Furthermore, if we have $t \mapsto \mathbb{E}\, X_t$ is right continuous, we get that $\mathbb{E}\, X_t = \lim_{k\to\infty} \mathbb{E}\, X_{s_k} = \mathbb{E}(\lim_{k\to\infty} X_{s_k}) = \mathbb{E}\, Y_t$. So we get in fact equality. □

We thus get the following:

Corollary 4.42 (*Conditions for the Right Limit of a Supermartingale to be a Modification*)

Suppose $\mathcal{F}_t$ is right continuous and complete. Let X_t be a supermartingale such that $t \mapsto \mathbb{E}\, X_t$ is right continuous. Then X has a cadlag modification $\tilde{X}$ which is also a supermartingale with respect to $\mathcal{F}_t$.

Proof. This is straightforward, simply define $\tilde{X}$ as the right limit a.s. (defining it 0 otherwiese) and note that by the right continuity of the filtration we get that $\tilde{X}_t$ is $\mathcal{F}_t$-adapted. By the previous lemma, we also get that by right continuity of the expectation that $\tilde{X}$ is a modification of X since

$$X_t = \mathbb{E}[X_{t+} \mid \mathcal{F}_t] = \mathbb{E}[\tilde{X}_t \mid \mathcal{F}_t] = \tilde{X}_t \tag{4.27}$$

almost surely. □

In practice, besides the right continuity of the expectation, we are free to simply change our filtration to allow for cadlag modification.

Example 4.43 (*Right Continuity of Filtration is Neccesary*)
Let $\Omega = \{\pm 1\}$ and let $\mathbb{P}$ be the uniform measure on Ω. Define $\mathcal{F}_t := \{\emptyset, \Omega\}$ for $t \leq 1$ and 2^Ω otherwise. Now, $X_t(\omega) = \omega \mathbb{1}[t > 1]$ is an adapted martingale with $\mathbb{E}\, X_t = 0$ but it has no cadlag modification.

Proof. For there to be a cadlag modification, we would have to have $X_1 = U(\{\pm 1\})$ but this is not measurable with respect to $\mathcal{F}_1$. □

TLDR: for continuous time sub/super martingales, we can always create a cadlag martingale by taking right limits on a countable dense subset $D \subseteq [0, \infty)$. To get that it is a modification, we further require right continuity of the mean $t \mapsto \mathbb{E}\, X_t$.

4.7. Optional Stopping in Continuous Time Martingales

We now extend optional stopping for continuous time martingales. The main application is for Brownian motion started at $B_0 = 0$. Letting a stopping time $\tau = \tau_a \wedge \tau_b$ where τ_x is the minimum time to pass x, we have that $0 = B_0 = \mathbb{E}\, B_\tau = a\mathbb{P}(\tau_a < \tau_b) + b\mathbb{P}(\tau_a > \tau_b)$ so $\mathbb{P}(\tau_a < \tau_b) = b/(b - a)$.

The main reason this works is because our expectations are bounded, ie. $\mathbb{E}\, |B_{\tau \wedge n}| \leq \max\{|a|, |b|\}$ so by bounded convergence we can pass to the limit. Without further restrictions, we will see that in continuous time, the OST does not hold: eg. $\tau_1 < \infty$ but $B_0 = 0 \neq 1 = B_1$.

The key idea in discrete time is as follows: let (X_n) be a submartingale and let $Y_n = X_{\tau \wedge n}$. We have $Y_n = X_0 + \sum_{k=1}^n \mathbb{1}[\tau \geq k](X_k - X_{k-1}) = X_0 + H \star X$, ie. it is an offset doob transform with $H_k := \mathbb{1}[\tau \geq k] = 1 - \mathbb{1}[\tau < k] \in H_{k-1}$ which is previsible. Thus, $\mathbb{E}\, X_0 = \mathbb{E}\, Y_0 \leq \mathbb{E}\, Y_n = \mathbb{E}\, X_{n \wedge \tau}$ so the goal is now to take the RHS $\to \infty$. So the main results we need are those regarding L^1-convergence. In the continuous time setting, we additionally need regularity assumptions to extend.

We begin by showing L^1 convergence properties of sub/super martingales. Note the following differs from L^1 martingale convergence in that it is only almost sure. Convergence in L^1 is not neccesarily true, eg. consider $X_t = \exp(B_t - t/2)$.

Theorem 4.44 (*L^1 a.s. martingale convergence theorem*)
Let X_t be an L^1-bounded submartingale ($\sup_t \|X_t\|_1 < \infty$) with right continuous sample paths. Then $X_t \to X_\infty$ a.s. for $\|X_\infty\| \in L^1$.

Proof. By Doob's upcrossing lemma we have $\mathbb{E}\, U_{a,b}([0, T] \cap D) \leq \mathbb{E}[(X_T - a)_-]/(b - a) < \infty$ where $T < \infty$ and D is countably dense. By MCT, we can take a sup and thus get $U_{a,b}(D) < \infty$ a.s.. We can finish with Fatou's lemma to show that X_∞ is L^1. □

On the other hand, for $p > 1$ we have L^p-convergence (see Theorem 4.25).

4.8. Closed Martingales and Uniform Convergence

In the case of martingales, there is a more precise characterization in terms for L^1 convergence in terms of uniform integrability.

Definition 4.45

A collection $(X_i)_{i\in I}$ is **uniformly integrable** if

$$\lim_{M\to\infty}\sup_{i\in I}\mathbb{E}(|X_i|;|X_i|\geq M)=0 \tag{4.28}$$

The most common family of uniformly integrable functions we consider is the set of conditional expectations:

Theorem 4.46 (*Conditional Expectations are UI*)

If $Z\in L^1$ then $\{\mathbb{E}[Z\mid\mathcal{G}]:\mathcal{G}\text{ is a sub }\sigma-\text{algebra}\}$ is UI.

Theorem 4.47 (*L^1-marginale convergence theorem*)

Suppose (X_n) is a discrete stochastic process with $X_n\in L^1$ and $X_n\to X$ in probability. Then, the following are equivalent:

(i.) (X_n) is UI

(ii.) $X_n\to X$ in L^1

(iii.) $\mathbb{E}|X_n|\to\mathbb{E}|X|<\infty$

For continuous time we have $(i)\implies(ii)\implies(iii)\not\Longrightarrow(i)$.

Uniform Integrability for cadlag processes are in fact completely characterized by being closed, as shown by the following theorem:

Theorem 4.48

Let (X_t) be a martingale with right continuous sample paths. The following are equivalent:

1. X is closed
2. $\{X_t\}$ is UI.
3. X_t converges a.s. and in L^1.

From the above, we can go straightforwardly to various stronger Optional Stopping Theorems in discrete time (see Durret Section 4.8). The continuous case is slightly more difficult and relies on some additional regularity assumptions:

Lemma 4.49 (*Finite Stopping Times are Measurable for Progressive Processes*)

Let (X_t) be a $\mathcal{F}_t$-adapted process with right continuous sample paths. Then,

(a) (X_t) is progressive.

(b) If τ is a finite stopping time, then $X_\tau \in \mathcal{F}_\tau$.

Proof. Fix $t > 0$. Denote $F := F_t : \Omega \times [0,t] \to \mathbb{R}$. Consider the mesh: $F^n(\omega, s) := F(\omega, \frac{\lceil sn/t \rceil}{n/t})$. It is easy to see that F^n is measurable wrt $\mathcal{F}_t \times \mathcal{B}([0,t])$. From sample path right continuity, we get $F^n \to F$ as we make $n \to \infty$ (ie. the mesh becomes smaller and smaller) so F is measurable since the limits of measurable functions is measurable.

For the second part, note $X_\tau \in \mathcal{F}_\tau$ implies that $\{X_\tau \in B\} \cup \{\tau \le t\} \in \mathcal{F}_t$ for all Borel B and $t \ge 0$. One can show $X_\tau = F(G(\omega))$ where F is defined above and $G(\omega) = (\omega, t \wedge \tau(\omega))$. □

Way to think about it: right continuous $\implies$ progressive $\implies$ adapted.

The main result for Optional Stopping in Continuous Time is then the following which states that uniform integrability and right continuity gives the following continuous time L^1 optional stopping theorem:

Theorem 4.50 (*Continuous L^1 Optional Stopping Theorem*)
Let (X_t) be a uniformly integrable martingale with right continuous sample paths and let $\sigma \le \tau$ be finite stopping times. Then, $X_\sigma, X_\tau \in L^1$ and $X_\sigma = \mathbb{E}[X_\tau \mid \mathcal{F}_\sigma]$.

Proof. Approximate $\sigma_n \downarrow \sigma$ and $\tau_n \downarrow \tau$ discrete where $\sigma_n \le \tau_n$.

The main idea now is to take the discrete version of the OST: namely, we have that $X_{\sigma_0} = \mathbb{E}[X_{\sigma_0} \mid \mathcal{F}_{\sigma_n}]$ so they are UI. Since $X_{\sigma_n} \to X_\sigma$ a.s. by right continuity we get by L^1-martingale convergence that $X_\sigma \in L^1$ (and similar for X_τ).

Finally, for $A \in \mathcal{F}_\sigma \subseteq \cap \mathcal{F}_{\sigma_n}$ we have

$$\mathbb{E}[X_{\sigma_n}; A] = \mathbb{E}[X_{\tau_n}; A] \implies \mathbb{E}[X_\sigma; A] = \mathbb{E}[X_\tau; A] \tag{4.29}$$

by L^1 convergence. Since $X_\sigma \in \mathcal{F}_\sigma$ by the previous result (since X_t is progressive) w get the result. □

Corollary 4.51 (*Stopped martingales with Right Continous Sample Paths are martingales*)
If X_t is a martingale with right continuous sample paths, then

1. If $\{X_t\}$ is UI then so is $\{X_{t\wedge\tau}\}$ and $X_{t\wedge\tau} = \mathbb{E}[X_\tau \mid \mathcal{F}_t]$
2. $(X_{t\wedge\tau})$ is still a martingale wrt $(\mathcal{F}_t)$

Proof. Let $\sigma = \tau \wedge t \le \tau$. Note since it is a stopping time by the OST above, we have $X_{t\wedge\tau} = \mathbb{E}[X_\tau | \mathcal{F}_{t\wedge\tau}]$ so $X_{t\wedge\tau}$ is UI, being closed. We also have that for $A \in \mathcal{F}_t$, that

$$X_{t\wedge\tau}\mathbb{1}[A \cap \{\tau \le t\}] = X_\tau \mathbb{1}[A \cap \{\tau \le t\}] \implies \mathbb{E}[X_{t\wedge\tau}; A \cap \{\tau \le t\}] = \mathbb{E}[X_\tau; A \cap \{\tau \le t\}] \tag{4.30}$$

Next, for $\tau > t$, just use the fact that $\mathbb{E}[X_\tau | \mathcal{F}_{t\wedge\tau}]$ so we get the first part.

To show the second part, first note it is adapted since $X_{t\wedge\tau} \subseteq \mathcal{F}_{t\wedge\tau} \subseteq \mathcal{F}_\tau$. Next, take any $s \le t$ and note that $X_{t\wedge\tau}$ is UI for for each fixed t. By the first part we thus get that $X_{s\wedge t\wedge\tau}$ is also UI with $X_{s\wedge t\wedge\tau} = \mathbb{E}[X_\tau | \mathcal{F}_s]$. This gives the result. □

4.9. Applications of the Optional Stopping Theorem

In general for many continuous time stochastic processes, we will have that although perhaps (X_t) is not a martingale, we have a sequence of stopping times $\tau_n \uparrow \infty$ and $(X_{t\wedge\tau_n})$ is a ui martingale.

The canonical example of the continuous martingale is Brownian motion although it is not UI since the variance explodes over time. In the spirit of the above paragraph, however, we can choose an appropriate stopping time to create UI martingales to use the OST. Indeed, the OST is powerful in explicit calculation of many of these probability distributions. Consider the following examples:

Example 4.52 (*Law of Exit Points*)
Denote by $\tau_a := \inf\{t \geq 0 : B_t = a\}$. For $a < 0 < b$, we have $\mathbb{P}(\tau_a < \tau_b) = b/(b-a)$ and $\mathbb{P}(\tau_a > \tau_b) = -a/(b-a)$.

Proof. Take $\tau = \tau_a \wedge \tau_b$ □

Example 4.53 (*First Moment of Exit Times*)
Let $\tau_a := \inf\{t \geq 0 : |B_t| = a\}$. We have $\mathbb{E}[\tau_a] = a^2$.

Proof. Take the martingale $X_t = B_t^2 - t$. We have $X_{t\wedge\tau_a}$ is a martingale so $\mathbb{E}[(B_{t\wedge\tau_a})^2] = \mathbb{E}[t \wedge \tau_a] \uparrow \mathbb{E}[\tau_a]$ by MCT while the left hand side $\to \mathbb{E}[B_{\tau_a}^2] = a^2$ by DCT. Remark also that $X_{t\wedge\tau_a}$ is UI since each $|Y_t| \leq a^2 + \tau \in L^1$ so the tails die out. □

Example 4.54 (*Laplace Transform of Exit Times From an Interval*)
Let $\tau = \inf\{t \geq 0 : |B_t| = 1\}$. Fix $\lambda < 0$. We have that the laplace transform

$$\mathbb{E}[\exp(\lambda\tau)] = 1/\cosh(a\sqrt{-2\lambda}) := m(\lambda) \tag{4.31}$$

We call $m(\lambda)$ the **moment generating function** of τ.

Proof. We have for the martingale $M_t = \exp(\theta B_t - \frac{1}{2}\theta^2 t)$ that $M_{t\wedge\tau}$ is a martingale that is also bounded between 0 and $\exp(\theta)$ so it is UI and we can apply the OST. This gives:

$$1 = \mathbb{E}[M_0] = \mathbb{E}[M_\tau] = \frac{\exp(\theta) + \exp(-\theta)}{2}\,\mathbb{E}\left[\exp\left(-\frac{\theta^2\tau}{2}\right)\right] = \cosh(\theta)\,\mathbb{E}\left[\exp\left(-\frac{\theta^2\tau}{2}\right)\right] \tag{4.32}$$

We thus get that taking $\lambda = -\theta^2/2$ that

$$\mathbb{E}[\exp(\lambda\tau)] = 1/\cosh(\sqrt{-2\lambda}) := m(\lambda) \tag{4.33}$$

For $\lambda > 0$, we define $m(\lambda) = \cosh(1/i\sqrt{2\lambda}) = 1/\cos(\sqrt{2\lambda})$ which holds for all $\lambda < \pi^2/8$. □

Corollary 4.55 (*Large Deviations Estimate*)
Defining τ as above, We have

$$\mathbb{P}(\tau \geq t) \leq \exp\left(-(1+o_t(1))\frac{\pi^2 t}{8}\right) \tag{4.34}$$

5. (L10-L13) Continuous Semimartingales

Recall the Doob Decomposition $X_n - X_0 = M_n + A_n$ where M_n was a martingale and A_n was previsible. We develop a continuous time analogue of this, called a continuous *semimartingale* which will allow us to analyze quadratic variation processes.

A continuous semimartingale is the sum of a continuous local martingale and a continuous finite-variation process. Intuitively, this can be thought of as a martingale-like fluctuation term on the order $\sqrt{dt}$ along with a finite-variation drift term which accumulates on the order dt. We give some intuition as to why this is the right class: the main idea is that semimartingales remain closed under smooth transforms due to Ito's formula, the main result of this course.

Lets B_t be a standard Brownian motion (which is a continuous martingale) and $f : \mathbb{R} \to \mathbb{R}$ a smooth bounded function on a compact support, by Ito's Formula we get a decomposition

$$dX_t = f'(B_t)dB_t + \frac{1}{2}f''(B_t)dt \tag{5.1}$$

which can be thought of as a "martingale" term and a finite variation term. More generally, one can show for a continuous semimartingale $X_t = M_t + A_t$ that by Ito's Formula,

$$X_t = \underbrace{\int h'(X_t)dM_t}_{d\tilde{M}_t} + \underbrace{\int h'(X_t)dA_t + \frac{1}{2}h''(X_t)(dM_t)^2}_{d\tilde{A}_t} \tag{5.2}$$

where we recover the case of Brownian motion setting $M_t = B_t$ and $A_t = 0$. Some care is needed of course in the heuristic $dA_t \asymp dt$ which we cover in a discussion of finite variation processes.

5.1. Finite Variation Processes

5.1.1. Functions with Bounded Variation

We begin by defining the notion of a function with finite variation in a deterministic setting.

Definition 5.1 (*Signed Measures*)
Let $(\Omega, \mathcal{F})$ be a measurable space. We call a countably additive function $\alpha : \mathcal{F} \to (-\infty, \infty)$ a (finite) **signed measure**.

Definition 5.2

Given a signed measure α on $(\Omega, \mathcal{F})$ we say $A \in \mathcal{F}$ is **positive** if $\alpha(B) \geq 0$ for all $B \subseteq A$ such that $B \in \mathcal{F}$ and **negative** analogously. A set A is called a **null set** of α if $\alpha(B) = 0$ for all $B \subseteq A$.

The main theorem of signed measures is the following decomposition:

Theorem 5.3 (*Hahn Decomposition*)

For any signed measure α on $(\Omega, \mathcal{F})$, there exists a bipartition $\Omega = A_+ \sqcup A_-$ such that A_+ is positive and A_- is negative. It is essentially unique up to null sets, ie. for any other partition $\Omega = B_+ \sqcup B_-$ we have $A_+ \cap B_-$ and $A_- \cap B_+$ are null sets.

Theorem 5.4 (*Jordan Decomposition*)

For any signed measure α on $(\Omega, \mathcal{F})$ we have there exists a pair of measures (α_+, α_-) such that $\alpha = \alpha_+ - \alpha_-$.

Proof. Just take the positive and negative parts by the Hahn Decomposition. □

This gives an alternative/equivalent characterization of the signed measure.

Definition 5.5 (*Finite Variation Functions*)

A continuous function $a : [0, T] \to \mathbb{R}$ is said to be of **finite variation (FV)** (equivalently **bounded variation (BV)** if $\exists$ a signed measure $\alpha : [0, T] \to \mathbb{R}$ such that $\alpha([0, t]) = a(t)$ for all $t \in [0, T]$.

Note that by the Jordan Decomposition, we have $a(t) = \mu_+([0, t]) - \mu_-([0, t])$ which shows that a is equivalently the difference of two monotone nondecreasing continuous functions that vanish at 0. This is in fact an equivalent defintion.

Definition 5.6 (*Total Variation Measure*)

Let α be a signed measure on $(\Omega, \mathcal{F})$. The measure $|\alpha| = \alpha_+ + \alpha_-$ is called the **Total Variation (TV)** of α

We remark that since $\alpha_+ \ll |\alpha|$ and similarly for α_- so by the Radon-Nikodym theorem we have RN-derivatives $\frac{d\alpha_+}{d|\alpha|} = \mathbb{1}_{\Omega_+}$ and similarly $\frac{d\alpha_-}{d|\alpha|} = \mathbb{1}_{\Omega_-}$.

5.2. The Lesbegue-Stiltjes Integral

We now define a general notion of integration with respect to different integrands for all bounded variation functions:

Definition 5.7 (*Lesbegue-Stieltjes Integral*)

Let a be a finite variation function on $[0, T]$ with corresponding signed measure α and total variation $|\alpha$. Let $f : [0, T] \to \mathbb{R}$ be a measurable function such that $\int_{[0,T]} |f(s)||\mu|(ds) < \infty$. We define the

Lesbegue-Stieltjes Integrals:

$$\int_0^T f(s)da(s) := \int_{[0,T]} f(s)\alpha(ds) \tag{5.3}$$

$$\int_0^T f(s)|da(s)| := \int_{[0,T]} f(s)|\alpha|(ds) \tag{5.4}$$

Note by the triangle inequality (seperating into the different parts) we get

$$\left|\int_0^T f(s)da(s)\right| \leq \int_0^T |f(s)||da(s)| \tag{5.5}$$

Note, we also get the following transformation of measure:

Theorem 5.8 (*Transformation of Finite Variation Function by Integral*)
The function $b(t) := \int_0^T f(s)da(s)$ is also finite variation.

We now show that the Lesbegue-Stieltjes integral can be approximated by an appropriate Riemann sum or mesh.

Lemma 5.9
Let $a : [0,T] \to \mathbb{R}$ be FV and let $P_n = \{(t_1, \cdots, t_{p_n}) : 0 = t_0 < t_1 < \cdots < t_{p_n} = t\}$ be a finite sequence and define

$$V_{P_n}(a) := \sum_{i=1}^{p_n} |a(t_i) - a(t_{i-1})| \tag{5.6}$$

If $P_1 \subseteq P_2 \subseteq \cdots$ is an increasing mesh with mesh$(P_n) \downarrow 0$ (ie. $\sup|t_j - t_{j-1}| \downarrow 0$) then $V_{P_n} \uparrow \int_0^T |da(s)|$

Proof. The $\geq$ direction is trivial: simply note that $|a(t_i) - a(t_{i-1})| = |\alpha((t_{i-1}, t_i])| \leq |\alpha|((t_{i-1}, t_i])$ by the triangle inequality and then take a sum.

For the other direction, we use a martingale probability argument despite the statement being fully deterministic. WLOG scale such that $|\alpha|(([0,t]) = 1$ (exclude the trivial case where $\mu = 0$, in which case the result is trivial) and let $\mathcal{G}_n := \sigma([t_{i-1}^n, t_i^n) : t_i \in P_n)$. Since P_j is increasing, we have $\mathcal{G}_n$ is a filtration. Furthermore since mesh$P_n \downarrow 0$ we have $\sigma((\mathcal{G}_n)_{n\in\mathbb{N}}) = \mathcal{B}([0,t])$. So $\mathcal{G}_\infty = \mathcal{B}([0,t])$.

Now, consider the Hahn Decomposition we have $\Omega := [0,t] = \Omega_+ \sqcup \Omega_-$. Now, define the random variable $X : \Omega \to \mathbb{R}$ by

$$X(t) := \mathbb{1}[t \in \Omega_+] - \mathbb{1}[t \in \Omega_-] = \frac{d\alpha}{d|\alpha|}(t) \tag{5.7}$$

Define the closed martingale $X_n := \mathbb{E}[X|\mathcal{G}_n]$. We thus have $X_n \to X$ in L^1 so of course $\mathbb{E}|X_n| \to \mathbb{E}|X|$. We now show that $\mathbb{E}|X_n| = V_{p_n}(a)$ and $\mathbb{E}|X| = \int_0^t |da(s)|$.

This is easy: recalling that a random variable is constant on its atoms and that $\mathbb{E}[X_n; (t_{i-1}^n, t_i^n]) = X_n(t_i^n)|\alpha|((t_{i-1}^n, t_i^n]) = \alpha((t_{i-1}^n, t_i^n]) = a(t_i^n) - a(t_{i-1}^n) = \mathbb{E}[X; (t_{i-1}^n, t_i^n]]$ from which you can recover the result. $\square$

The converse also holds and is an easy result in real analysis. We now give a discrete approximation to the integral as well:

Theorem 5.10

Suppose $a : [0,T] \to \mathbb{R}$ is finite variation and $f : [0,T] \to \mathbb{R}$ is continuous. If $P_1 \subseteq P_2 \subseteq \cdots$ is an increasing partition (as before) with $\text{mesh}(P_n) \downarrow 0$ then

$$\sum_{i=1}^{p_n} f(t_{i-1}^n)(a(t_i^n) - a(t_i^{n-1})) \to \int_0^t f(s)da(s) \tag{5.8}$$

Proof. Simply take $f^n(t) = f(t_{i-1}^n)$ for $t \in [t_{i-1}^n, t_i^n)$ piecewise and then use DCT. $\square$

Definition 5.11 (*FV on the line*)

A function $a : [0,\infty) \to \mathbb{R}$ is of finite variation if it is finite variation on each compact interval $[0,T]$.

5.3. Finite Variation Processes

Definition 5.12 (*Finite Variation Process*)

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space and A_t an adapted process. A_t is said to be a **finite variation process** if $t \mapsto A_t(\omega)$ is of finite variation $\forall \omega \in \Omega$.

Definition 5.13

If in addition the sample paths are non-decreasing, then the process is said to be an **increasing process**.

Example 5.14

The process $V_t := \int_0^t |dA_s|$ is an increasing process.

Note we have $A_t = \frac{1}{2}(V_t + A_t) - \frac{1}{2}(V_t - A_t)$ which shows that any FV process can be written as a difference of two increasing processes.

Proposition 5.15

Let A be a finite variation process, H be progressive such that $\int_0^t |H_s(\omega)||dA_s(\omega)| < \infty$ for all $\omega \in \Omega$ and $t \in [0,\infty)$. Then,

$$(H \cdot A)_t = \int_0^t H_s dA_s \tag{5.9}$$

is a finite variation process.

Lemma 5.16 (*Associtativity*)

If H and K are progressive with $\int_0^t |H_s||dA_s| < \infty$, we have $K \cdot (H \cdot A) = (KH) \cdot A$

5.4. Continuous Local Martingales

Although in general a continuous time process might not be a martingale even with only a martingale term due to possible unboudedness/variation of the function transform, we can in general show that for some cases we can have the process is "locally" a martingale in that for a sequence of stopping times it reduces towards being a martingale.

Definition 5.17 (*Continuous Local Martingales*)

An adapted process $M = (M_t)_{t\geq 0}$ with continuous sample paths is called a **continuous local martingale** if $\exists$ a nondecreasing sequence (T_n) of stopping times such that $T_n \uparrow \infty$ and for every n, $M_{t\wedge T_n} - M_0$ is a martingale. We say that (T_n) **reduces** M.

Lemma 5.18

Let M be a continuous local martingale.

(a) If σ is a stopping time, $M_{t\wedge\sigma}$is also a continuous local martingale

(b) If $M_t \leq Z \in L^1$ for all t, then M is a u.i. martingale.

(c) $\sigma_n = \inf\{t \geq 0 : |M_t| \geq n\}$ reduces M.

(d) In discrete time, any local martingale is also a martingale.

We start with a motivating example.

Example 5.19 (*Green's Function applied to Brownian Motion*)

Consider **Green's function** $h(x) = 1/|x|^{d-2}$ which is harmonic on $\mathbb{R}^d/\{0\}$ and hence $h(x) =$ mean of h on $B_\delta(x) =$ mean of h on $\partial B_\delta(x)$ if $0 \notin \overline{B_\delta(x)}$.

Let B_t be a standard brownian moption in $\mathbb{R}^d$ starting at e_1. Let $M_t := h(B_t)$. We have that M_t is *not* a martingale in general but it is a local martingale.

Proof. For simplicity fix $d = 3$ and for $p \geq 0$, note

$$\mathbb{E}[M_t^p] = \int_{\mathbb{R}^3} \frac{1}{|x|^p (2\pi t)^{3/2}} \exp\left(-\frac{\|x - e_1\|^2}{2t}\right) dx \tag{5.10}$$

Now note

$$\int_{|x|\le\sqrt{t}} \frac{1}{|x|^p(2\pi t)^{3/2}} \exp\left(-\frac{\|x-e_1\|^2}{2t}\right) dx \le \frac{1}{(2\pi t)^{3/2}} \int_{|x|\le\sqrt{t}} \frac{1}{|x|^p} dx \tag{5.11}$$

$$= \frac{1}{(2\pi t)^{3/2}} \int_0^{\sqrt{t}} C_3 r^{2-p} dr \tag{5.12}$$

$$= \Theta(t^{-p/2}) \tag{5.13}$$

where $C_3 = \int_{S^2} d\omega = 4\pi$. Similarly,

$$\int_{|x|>\sqrt{t}} \frac{1}{|x|^p(2\pi t)^{3/2}} \exp\left(-\frac{\|x-e_1\|^2}{2t}\right) dx \le \frac{1}{\sqrt{t}^p} \int_{|x|\ge\sqrt{t}} \frac{1}{(2\pi t)^{3/2}} \exp\left(-\frac{\|x-e_1\|^2}{2t}\right) dx \tag{5.14}$$

$$= \frac{1}{\sqrt{t}^p} \mathbb{P}(|B_t| \ge \sqrt{t}) \tag{5.15}$$

$$\le \frac{1}{\sqrt{t}^{p/2}} \tag{5.16}$$

so we get $\mathbb{E}[M_t^p] \to 0$ but a martingale would have by the OST $\mathbb{E}[M_0] = \mathbb{E}[M_t] = 1$ but it does not so it is a martingale.

On the other hand, consider $\tau_\epsilon = \inf\{t \ge 0 : |B_t| \le \epsilon\}$. Let $M_t^\epsilon = M_{t\wedge\tau_\epsilon}$. □

Lemma 5.20 (*Nonnegative CLMs are supermartingales*)
If M is a nonnegative continuous local martingale with $M_0 \in L^1$, then M is a supermartingale.

Proof. Let $N_t := M_t - M_0$. There exists a sequence (τ_n) of stopping times that reduces N_t, ie.

$$N_{s\wedge\tau_n} = \mathbb{E}[N_{t\wedge\tau_n}|\mathcal{F}_s] \tag{5.17}$$

so adding M_0 to both sides

$$M_{s\wedge\tau_n} = \mathbb{E}[M_{t\wedge\tau_n}|\mathcal{F}_s] \tag{5.18}$$

Taking $n \to \infty$ and using Fatou's Lemma, we get

$$M_s \ge \mathbb{E}[M_t|\mathcal{F}_s] \tag{5.19}$$

and taking expectation and $s = 0$, we get $\mathbb{E}[M_t] \le \mathbb{E}[M_0] < \infty$ so it is a supermartingale. □

Example 5.21
$M_t := 1/|B_t|^{d-2}$ is a supermartingale.

Theorem 5.22
Let M be a CLM with $M_0 = 0$ and also a finite variation process. Then $M = 0$ a.s.

The main theorem which allows all of the stochastic calculus machinery to work is the idea of the quadratic variation which roughly states that all martingale like quantities have order at most 2:

Theorem 5.23 (*Existence of the Quadratic Variation*)
Let $M = (M_t)$ be a CLM. There exists an increasing process denoted by $\langle M \rangle_t$ which is unique up to indistinguishability, such that $M_t^2 - \langle M \rangle_t$ is a CLM. Furthermore, for each fixed partition V_{P_n} with $0 = t_0^n < \cdots < t_p^n = t$ is an increasing subdivision with mesh going to 0. We then have that

$$\langle M \rangle_t = \lim_{\text{mesh}(V_{p_n}) \downarrow 0} \sum_{i=1}^{p_n} (M_{t_i^n} - M_{t_{i-1}^n})^2 \tag{5.20}$$

in probability. We call $\langle M \rangle_t$ the **quadratic variation** of M.

Sketch. The proof is rather technical but the main ideas are as follows: □

5.5. Quadratic Covariation of Continuous Local Martingales

We can extend the idea of the quadratic variation to a bilinear form between the space of continuous local martingales via polarization:

Definition 5.24
If M and N are two continuous local martingales, the **quadratic covariation** (or **bracket**) $\langle M, N \rangle$ is the finite variation process defined by:

$$\langle M, N \rangle_t = \frac{1}{2}(\langle M + N, M + N \rangle_t - \langle M, M \rangle_t - \langle N, N \rangle_t) \tag{5.21}$$

The main properties are as follows:

Lemma 5.25
Let M and N be two CLMs.

(a) $\langle M, N \rangle$ is the unique (up to indistinguishability) FV process such that $M_t N_t - \langle M, N \rangle_t$ is a continuous local martingale.

(b) $\langle \cdot, \cdot \rangle$ is a bilinear, symmetric form.

(c) If $V_{p_n} = \{t_0 = 0 < t_1 < \cdots < t_p\}$ is an increasing partition of $[0, T]$, we have

$$\lim_{\text{mesh}(V_{p_n}) \downarrow 0} \sum_{i=1}^{p_n} (M_{t_i} - M_{t_{i-1}})(N_{t_i} - N_{t_{i-1}}) \tag{5.22}$$

in probability.

(d) For a stopping time τ, $\langle M^\tau, N^\tau \rangle_t = \langle M^\tau, N \rangle = \langle M, N \rangle_{t \wedge \tau}$

(e) If M and N are two L^2-bounded martingales, $M_t N_t - \langle M, N\rangle_t$ is a UI martingale.

Definition 5.26 (*Orthogonal CLMs*)
If M and N are two CLMs have $\langle M, N\rangle = 0$ we say they are **orthogonal**.

Corollary 5.27
Two independent Brownian Motions are orthogonal.

The covariation satisfies a Cauchy-Schwarz like inequality:

Theorem 5.28 (*Kunita-Watanabe Inequality*)
Let M and N be two continuous local martingales and H, K two measurable processes. Then, a.s.

$$\int_0^\infty |H_s||K_s||d\langle M, N\rangle_s| \le \left(\int_0^\infty |H_s||d\langle M\rangle_s|\right)\left(\int_0^\infty |K_s||d\langle N\rangle_s|\right) \tag{5.23}$$

6. (L14-L17) Stochastic Integration

We now define the stochastic integral which will allow us to extend the idea of the Stiltjes integal to have semimartingale integrators. The core idea is to define it first for square martingales starting with simple progressive integrands and then extending with the Hilbert Space Isometry to more general progressive integrands. The extension to local martingales and then semimartingales is then easy (simply define it for all stopped times and then add a drift term with the normal Stieltjes integral).

6.1. Construction of the Ito Integral

For a local martingale M, we want to define the stochastic integral $(H \cdot M)_t = \int_0^t H_s dM_s$ in a way that is analogous to the Doob Transform in discrete time:

$$(H \cdot M)_n := \sum_{i=1}^n H_i(M_i - M_{i-1}) \tag{6.1}$$

6.1.1. The Ito Integral for Square Martingales

Definition 6.1
Let $\mathbb{H}^2$ be the space of all continuous square martingales M with $M_0 = 0$. Eqivalently, M is a CLM with $\mathbb{E}[\langle M, M\rangle_\infty] < \infty$.

Theorem 6.2 (*Square Continuous Martingales as a Hilbert Space*)
Consider $\mathbb{H}^2$ equipped with the symmetric bilinear form $(M, N)_{\mathbb{H}^2}$ given by the expectation of the

bracket, ie.

$$(M,N)_{\mathbb{H}^2} := \mathbb{E}[\langle M,N\rangle_\infty] = \mathbb{E}[M_\infty N_\infty] \tag{6.2}$$

Then, $\mathbb{H}^2$ equipped with this norm is a Hilbert Space.

Proof. We show the space is complete. Let $(M^n)_{n\geq 1}$ be a sequence in $\mathbb{H}^2$ which is Cauchy, ie.

$$\lim_{m,n\to\infty} \mathbb{E}[(M^n_\infty - M^m_\infty)^2_\infty] = \lim_{m,n\to\infty} \|M^n - M^m\|_{\mathbb{H}^2} = 0 \tag{6.3}$$

Then, we have by the completeness of L^2 that M^n_∞ converges in L^2 to a limit, say Z. Now, fix $t > 0$ and note by Doob's L^p inequality that

$$\mathbb{E}[\sup_{t\geq 0}(M^n_t - M^m_t)^2] \leq 4\,\mathbb{E}[(M^n_\infty - M^m_\infty)^2] \tag{6.4}$$

so taking a limit gives us that $(M^n_t)_{n\geq 1}$ is Cauchy as well. So each (M^n_t) converges in L^2 to a limit. Now, note then we can choose some sequence such that

$$\mathbb{E}[\sum_{k=1}^\infty \sup_{t\geq 0} |M^{n_k}_t - M^{n_{k+1}}_t|] \leq \sum_{k=1}^\infty \mathbb{E}[\sup_{t\geq 0}(M^{n_k}_t - M^{n_{k+1}}_t)^2]^{1/2} < \infty \tag{6.5}$$

and so we get that (M^n_t) converges uniformly a.s. to a limiting process (M_t) so the limiting process is also continuous (take it to be 0 on the null set of non uniformly converging events). Then, from the identity $M^{n_k}_t = \mathbb{E}[M^{n_k}_\infty|\mathcal{F}_t]$ we can take a limit $k\to\infty$ and use the fact that the inner term is uniformly intgerable to get $M_t = \mathbb{E}[Z|\mathcal{F}_t]$ so we have (M_t) is an L^2 bounded continuous martingale, ie. $M \in \mathbb{H}^2$. Since the convergence is uniform we also have that the limit is preserved, ie. $M_\infty = Z$ a.s. and since M^n_∞ converges to Z in L^2 since the a.s. limit and the L^2-limit must coincide we get That $M^n \to M$ in $\mathbb{H}^2$ as desired. ☐

Definition 6.3 *($L^2(M)$ as a Hilbert Space)*

For a continuous square martingale M, let $\mathcal{P}$ denote the progressive σ-field on $\Omega\times\mathbb{R}_+$ and let $L^2(M)$ denote all progressive processes H with

$$\mathbb{E}\left[\int_0^\infty H^2_s d\langle M,M\rangle_s\right] < \infty \tag{6.6}$$

where two processes are equivalent if they are indistinguishable. We can view $L^2(M)$ as a Hilbert Space under the scalar product:

$$(H,K)_{L^2(M)} := \mathbb{E}[\int_0^\infty H_s K_s d\langle M,M\rangle_s] \tag{6.7}$$

and the progressive σ-field.

From here, we now define the notion of the stochastic integral for continuous square martingales by first defining them for elementary progressive processes and then extending with an isometry between $L^2(M) \to \mathbb{H}^2$ (called the *Ito Isometry*).

Definition 6.4 (*Elementary Processes*)

A **elementary process** is a progressive process of the form:

$$H_s(\omega) = \sum_{i=1}^{p} H_{(i)}(\omega)\mathbb{1}[s \in (t_i, t_{i+1}]] \tag{6.8}$$

for $0 = t_0 < t_1 < \cdots < t_p < \infty$ where each $H_{(i)}$ is a bounded $\mathcal{F}_{t_i}$-random variable. We have the space of all such processes $\mathcal{E} \subseteq L^2(M)$ forms a linear subspace of $L^2(M)$.

Theorem 6.5 ($\mathcal{E}$ *is dense in* $L^2(M)$)

For every $M \in \mathbb{H}^2$, $\mathcal{E}$ is dense in $L^2(M)$.

Proof. We show for K orthogonal to $\mathcal{E}$ that it is 0. define the progressive process

$$X_t := \int_0^t K_s d\langle M\rangle_s \tag{6.9}$$

which has that $\mathbb{E}[\int_0^t |K_s| d\langle M, M\rangle_s] \leq \mathbb{E}[\int_0^t |(K_s)^2| d\langle M, M\rangle_s]^{1/2}\,\mathbb{E}[\langle M, M\rangle_\infty]^{1/2} < \infty$ by Cauchy Schwarz and so we have X_t is finite variation.

Next, we show for $A \in \mathcal{F}_s$ we have $\mathbb{E}[(X_t - X_s)\mathbb{1}_A] = 0$. To do this, consider $H \in \mathcal{E}$ defined by $H_r(\omega) = \mathbb{1}\{\omega \in A\}\mathbb{1}[r \in (s,t]]$. Use the fact that $K \perp \mathcal{E}$ to get $(H, K)_{L^2(M)} = \mathbb{E}[\mathbb{1}_A(X_t - X_s)] = 0$. So X_t is FV and a martingale and thus 0 so $K = 0$ a.e. on $d\langle M, M\rangle$. $\square$

The construction of the integral for elementary processes is straightforward:

Definition 6.6 (*Ito Integral for Elementary Processes*)

Let $M \in \mathbb{H}^2$. For an elementary process $H \in \mathcal{E}$ given by

$$H_s(\omega) = \sum_{i=1}^{p} H_{(i)}(\omega)\mathbb{1}[s \in (t_i, t_{i+1}]] \tag{6.10}$$

we define the **Ito Integral**:

$$(H \cdot M)_t = \int_0^t H_s dM_s := \sum_{i=1}^{p} H_{(i)}(\omega)(M_{t_{i+1}\wedge t} - M_{t_i \wedge t}) \tag{6.11}$$

which by previsibility of H also gives a martingale.

We then get that

$$\mathbb{E}[(H \cdot M)_\infty^2] = \mathbb{E}[\sum_{i=1}^{p} H_{(i)}^2 (M_{t_{i+1}}^2 - M_{t_i})^2] = \mathbb{E}[\sum_{i=1}^{p} H_{(i)}^2 (\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i})] = \mathbb{E}[\int_0^\infty H_t^2 d\langle M\rangle_t] \tag{6.12}$$

Note we have a norm in the $\mathbb{H}^2$ on the left and the norm on $L^2(M)$ on the right. We thus get an isometry $\mathcal{E} \to \mathbb{H}^2$. We can then extend by continuity since $\mathcal{E}$ is dense.

Theorem 6.7 (*Isometry between $L^2(M)$ and $\mathbb{H}^2$*)
Let $M \in \mathbb{H}^2$. For $H \in \mathcal{E}$ defined by

$$H_s(\omega) = \sum_{i=1}^{p} H_{(i)}(\omega)\mathbb{1}[s \in (t_i, t_{i+1}]] \tag{6.13}$$

define the isometry $H \mapsto M \cdot H$ as above. This isometry extends uniquely to an isometry $L^2(M) \to \mathbb{H}^2$.

Proof. As shown above, for $H \in \mathcal{E}$ we have that $(H \cdot M) \in \mathbb{H}^2$ with QV $\langle H \cdot M \rangle_t = \int_0^t H_s^2 ds$ so we get that $H \mapsto H \cdot M$ preserves norms. This mapping is linear and thus an isometry and since $\mathcal{E}$ is dense, we can extend uniquely as follows: for general $H \in L^2(M)$ take $H^n \in \mathcal{E}$ with $\|H^n - H\|_{L^2(M)} \to 0$ and define $H \cdot M = \lim_{n\to\infty} H^n \cdot M$ which we can use the $\mathcal{E}$-isometry to show convergence in $\mathbb{H}^2$. $\square$

An alternative characterization of the stochastic integral is as follows and is due to Kunita and Watanabe (1967):

Theorem 6.8 (*Kunita-Watanabe Characterization of the Stochastic Integral*)
For $M \in \mathbb{H}^2$ and $H \in L^2(M)$, $H \cdot M \in \mathbb{H}^2$ is the unique square martingale such that

$$\langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle \tag{6.14}$$

for all $N \in \mathbb{H}^2$. That is,

$$\langle \int_0^{\cdot} H_s dM_s, N \rangle_t = \int_0^t H_s d\langle M, N \rangle_s \quad \forall t \geq 0 \tag{6.15}$$

Proof. For $H \in \mathcal{E}$, the identity can be shown explicitly to hold. For general $H \in \mathbb{H}^2$ by the Kunita-Watanabe Inequality,

$$\mathbb{E}\left[\int_0^\infty |H_s|\,|d\langle M, N\rangle_s|\right] \leq \|H\|_{L^2(M)}\|N\|_{\mathbb{H}^2} < \infty \tag{6.16}$$

and so $\int_0^\infty H_s d\langle M, N\rangle_s = H \cdot \langle M, N\rangle$ is well-defined and finite. Next, take $H^N \overset{L^2(M)}{\to} H$ and note since $M \mapsto H \cdot M$ is an isometry so $H^n \cdot M \to H \cdot M$ in H^2. Next, note the map $X \mapsto \langle X, N\rangle_\infty$ for $X \in \mathbb{H}^2$ into L^1 is continuous, since by Kunita-Watanabe,

$$\mathbb{E}[|\langle X, N\rangle_\infty|] \leq \|N\|_{\mathbb{H}^2}\|X\|_{\mathbb{H}^2} \tag{6.17}$$

We thus get since $H^n \cdot M \to H \cdot M$ in $\mathbb{H}^2$ that by continuity,

$$\langle H \cdot M, N\rangle_\infty = \lim_{n\to\infty} \langle H^n \cdot M, N\rangle_\infty = \lim_{n\to\infty} H^n \cdot \langle M, N\rangle \tag{6.18}$$

Finally, note that by KW again, we have

$$\mathbb{E}\left[\left|\int_0^\infty (H_s^n - H_s)d\langle M, N\rangle_s\right|\right] \leq \mathbb{E}[\langle N, N\rangle_\infty]^{1/2}\|H^n - H\|_{L^2(M)} \to 0 \tag{6.19}$$

so we get

$$\langle H \cdot M, N\rangle_\infty = \lim_{n\to\infty} H^n \cdot \langle M, N\rangle_\infty = (H \cdot \langle M, N\rangle)_\infty \tag{6.20}$$

Stopping at N^t completes the proof. Uniqueness is easy: if $X \cdot \langle M, N \rangle = \langle H \cdot M, N \rangle$, taking $N = X - H \cdot M$ shows that $X = H \cdot M$ a.s. and we are done. □

The idea can be thought of heuristically as follows:

> **Heuristic.** For a martingale $M \in \mathbb{H}^2$ and $H \in L^2(M)$, heuristically we have
>
> $$d(H \cdot M)_t = H_t dM_t \tag{6.21}$$
>
> and
>
> $$d\langle H \cdot M, N \rangle = d\langle H \cdot M, N \rangle_t = d(H \cdot M)_t dN_t = H_t dM_t dN_t = H_t d\langle M, N \rangle_t \tag{6.22}$$

Intuitively, the process H "commutes" with the bracket.

Corollary 6.9

Let $H \in L^2(M)$. If K is progressive, $KH \in L^2(M)$ iff $K \in L^2(H \cdot M)$. In particular,

$$(KH) \cdot M = K \cdot (H \cdot M) \tag{6.23}$$

The most important identity that arises from the above characterizations are those concerning the moments:

Theorem 6.10

Let $M, N \in \mathbb{H}^2$ and $H \in L^2(M)$, $K \in L^2(N)$. We have

$$\mathbb{E}\left[\int_0^t H_s dM_s\right] = 0 \tag{6.24}$$

and

$$\mathbb{E}\left[\left(\int_0^t H_s dM_s\right)\left(\int_0^t K_s dN_s\right)\right] = \mathbb{E}\left[\int_0^t H_s K_s d\langle M, N \rangle_s\right] \tag{6.25}$$

Proof. The first equality holds by the martingale property since $H \cdot M$ is also a martingale. We have since

$$\mathbb{E}\left[\left(\int_0^\cdot H_s dM_s\right)\left(\int_0^\cdot K_s dN_s\right)\right] = \mathbb{E}\left[\langle \int_0^\cdot H_s dM_s, \int_0^\cdot K_s dN_s \rangle_t\right] = \mathbb{E}\left[\int_0^t H_s K_s d\langle M, N \rangle_s\right] \tag{6.26}$$

□

Corollary 6.11 (*The Ito Isometry*)

For $M \in \mathbb{H}^2$ and $H \in L^2(M)$, we have that

$$\mathbb{E}\left[\left(\int_0^t H_s dM_s\right)^2\right] = \mathbb{E}\left[\int_0^t H_s^2 d\langle M \rangle_s\right] \tag{6.27}$$

Proof. Taking $N = M$ and $K_s = M_s$ in the previous theorem:

$$\mathbb{E}\left[\left(\int_0^t H_s dM_s\right)^2\right] = \mathbb{E}\left[\int_0^t H_s^2 d\langle M\rangle_s\right] \tag{6.28}$$

□

6.2. The Ito Integral for Local Martingales

Definition 6.12

Let M be a continuous local martingale and let $L^2_{loc}(M)$ be the set of all progressive processes H such that

$$\mathbb{E}\left[\int_0^\infty H_s^2 d\langle M, M\rangle_s\right] < \infty \tag{6.29}$$

We can now define the Ito Integral for local martingales using the Kunita-Watanabe characterization but for local martingales.

Theorem 6.13 (*The Ito Integral for Local Martingales*)

Let M be a continuous local martingale with $M_0 = 0$. If $H \in L^2_{loc}(M)$, then the integral $H \cdot M$ can be defined, and is the unique continuous local martingale such that

$$\langle H \cdot M, N\rangle = H \cdot \langle M, N\rangle \tag{6.30}$$

for all continuous local martingales N. This definition is consistent with that for $\mathbb{H}^2$.

Proof. Let $\tau_1(n) = \inf\{t \geq 0 : |M_t| \geq n\}$ and $\tau_2(n) = \inf\{t \geq 0 : \int_0^t H_s^2 d\langle M\rangle_s \geq n\}$. Take $\tau_n := \tau_1(n) \wedge \tau_2(n)$. Now consider the stopped process M^{τ_n} which is now bounded by n and hence L^2 bounded. Additionally, $\int_0^\infty H_s^2\, d\langle M\rangle_s \leq n$ so $H \in L^2(M^{\tau_n})$ and thus $H \cdot M^{\tau_n}$ is well-defined and satisfies

$$\langle H \cdot M^{\tau_n}, N\rangle = H \cdot \langle M^{\tau_n}, N\rangle \tag{6.31}$$

Additionally, we remark that for $m > n$ that

$$H \cdot M^{\tau_n} = (H \cdot M)^{\tau_n} = (H \cdot M)^{\tau_m \wedge \tau_n} = (H \cdot M^{\tau_m})^{\tau_n} \tag{6.32}$$

and since $\tau_n \to \infty$ as $n \to \infty$, define $(H \cdot M) = \lim(H \cdot M^{\tau_n})$. Since each of the processes are in $\mathbb{H}^2$, the limiting process is a continuous local martingale. We can then check the commuting condition by taking a monotone limit by extending the characterization for the square martingale case. □

Corollary 6.14 (*Ito Isometry for continuous local martingales*)

If M is a continuous local martingale and $H \in L^2_{loc}(M)$ we have both that

$$\mathbb{E}\left[\left(\int_0^t H_s dM_s\right)^2\right] \leq \mathbb{E}\left[\int_0^t H_s^2 d\langle M\rangle_s\right] \tag{6.33}$$

with equality if the right hand side is $< \infty$.

Finally, for the case in which the integrator is the Wiener process (Brownian Motion) we have the following characterization of the integral:

Theorem 6.15 (*Wiener Integral as Gaussian White Noise*)
Suppose (B_t) is an $(\mathcal{F}_t)$ Brownian Motion and $h \in L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+, dt))$ is a deterministic square function. We then have that under the integral defined earlier, that

$$\int_0^t h(s)dB_s = G(h(s)\mathbb{1}_{[0,t]}) \tag{6.34}$$

where G is the Gaussian White Noise associated with the Brownian Motion. This in turn shows that the integral with respect to a Brownian Motion (Wiener Process) is a Gaussian process.

Proof. Begin with step functions $h(s) = \sum_{k=0}^{p-1} a_k \mathbb{1}_{(t_k, t_{k+1}]}$ which gives precisely the Ito Integral for elementary processes:

$$\int_0^t h(s)dB_s = \sum_{k=0}^{p-1} a_k(B_{t_{k+1}\wedge t} - B_{t_k \wedge t}) \tag{6.35}$$

Then finish with a density argument. □

6.3. The Ito Integral for Semimartingales

The case for semimartingales is easy:

Definition 6.16 (*Locally Bounded Process*)
A process (X_t) is locally bounded if $\sup_{0\leq s\leq t} X_s < \infty$ a.s. for all $t > 0$. Notably, any continuous adapted process is locally bounded.

Lemma 6.17
Let $X = M + V$ be a semimartingale with M a continuous local martingale and V a finite variaiton process, we define the **Ito Integral** by

$$H \cdot X = H \cdot M + H \cdot V \tag{6.36}$$

Such a construction is well-defined.

6.3.1. Convergence Theorems for Stochastic Integrals

We first give an analogous “Dominated Convergence” theorem for Stochastic Integrals:

Theorem 6.18 (*Stochastic Dominated Convergence THeorem*)
Let $X = A + M$ be a semimartingale and H^n, H, K be locally bounded progressive processes with $K \geq 0$. Suppose:

(i) $H_s^n \to H_s$ for a.e. $s \in [0,t]$

(ii) $|H_s^n| \le K$ for every $n \ge 1$ and a.e. $s \in [0,t]$

(iii) $\int_0^t (K_s)^2 d\langle M\rangle_s < \infty$ for all $t \ge 0$ and $\int_0^t K_s |dA_s| < \infty$. Then

$$\lim_{n\to\infty} \int_0^t H_s^n dX_s \to \int_0^t H_s dX_s \tag{6.37}$$

in probability.

Here, by a.e. we mean with respect to $d\langle M, M\rangle_s$ and dA_s.

Proof. The convergence

$$\lim_{n\to\infty} \int_0^t H_s^n dA_s \to \int_0^t H_s dA_s \tag{6.38}$$

is automatic and follows from the usual DCT for Lesbegue integrals □

Finally, we use the previous result to show a Riemann Sum like formula for the stochastic integral:

Theorem 6.19

Let X be a continuous semimartingale and H an adapted process with continuous sample paths. Then for every $t > 0$ and V_{p_n} subdivision of $[0,t]$ with mesh$(V_{p_n}) \downarrow 0$ we have

$$\lim_{n\to\infty} \sum_{i \le |p_n|} H_{t_i^n}(X_{t_{i+1}^n} - X_{t_i^n}) = \int_0^t H_s dX_s \tag{6.39}$$

in probability.

Proof. Define the elementary processes $H_s^n := \sum_{i\le|p_n|} H_{t_i} \mathbb{1}[s \in (t_i, t_{i+1}]]$ which can easily be seen to be progressive. Now note $H_s^n \to H_s$ for all $s \in [0,t]$ a.s. and $K_s := \max_{0\le r\le s} |H_s|$ also bounds each H_s^n. So we get the result by the previous theorem. □

> **Remark.** It is essential in the previous theorem we evaluate at the left end of the interval. For example, let the integrand and integrator be the same, say X_t. Then:
>
> $$\sum_{i=0}^{p_n-1} X_{t_{i+1}}(X_{t_{i+1}} - X_{t_i}) = \int_0^t X_s dX_s + \langle X, X\rangle_t \tag{6.40}$$
>
> If instead the midpoint is used, we get the **Stratonovich integral**. The right endpoint is relatively unused.

6.4. Ito's Formula

Having constructed the Ito integral, now prove Ito's Lemma, the main workhorse of stochastic calculus and the stochastic equivalent of the chain rule. The main idea is that we now have to additionally consider the second order term due to the effects of the continuous local martingale term in the semimartingale.

Theorem 6.20 (*Ito's Formula*)
Let $X^1, \cdots, X^p$ be p continuous semimartingales on a common filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and denote $\bar{X}_t = (X_t^1, \cdots, X_t^p)$. If $f : \mathbb{R}^p \to \mathbb{R}$ is twice differentiable, we have

$$f(\bar{X}_t) - f(\bar{X}_0) = \sum_{i=1}^{p} \int_0^t \frac{\partial f}{\partial x^i}(\bar{X}_s) dX_s^i + \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}(\bar{X}_s)\, d\langle X^i, X^j \rangle_s \tag{6.41}$$

Typically, we abbreviate this as

$$df(\bar{X}_t) = \nabla f(X_t) dX_t + \frac{1}{2} \sum_{i,j=1}^{p} \partial_{ij}^2 f(\bar{X}_t) d\langle X^i, X^j \rangle_t \tag{6.42}$$

Proof. We consider only the case $p = 1$ (the extension is straightforward). Let V_{P_n} be an increasing subsequence of $[0, t]$ with mesh$(V_{P_n}) \to 0$. Now, for every $n > 0$, we have

$$F(X_t) = F(X_0) + \sum_{i=0}^{p_n - 1} (F(X_{t_{i+1}}) - F(X_{t_i})) \tag{6.43}$$

By Taylor-Lagrange on $\theta \to F(X_{t_i} + \theta(X_{t_{i+1}} - X_{t_i}))$ we get

$$f(X_{t_{i+1}}) - f(X_{t_i}) = F'(X_{t_i})(X_{t_{i+1}} - X_{t_i}) + \frac{1}{2} f''_{n,i}(X_{t_{i+1}} - X_{t_i})^2 \tag{6.44}$$

where $f''_{n,i} = f''(X_{t_i} + c(X_{t_{i+1}} - X_{t_i}))$ for some $c \in [0, 1]$. Now the first term $F'(X_{t_i})(X_{t_{i+1}} - X_{t_i}) \to \int_0^t F'(X_s) dX_s$ in probability by Theorem 6.19. Similarly, the second term as we take the mesh to be 0 converges to $\frac{1}{2} \sum_{i=0}^{p_n - 1} f''(X_{t_i})(X_{t_{i+1}} - X_{t_i})^2$ since

$$\left| \frac{1}{2} \sum_{i=0}^{p_n - 1} \left(f''_{n,i} - f''(X_{t_i})\right) \left(X_{t_{i+1}} - X_{t_i}\right)^2 \right| \leq \frac{1}{2} \left(\max_{0 \leq i \leq p_n - 1} |f''_{n,i} - f''(X_{t_i})| \right) \sum_{i=0}^{p_n - 1} \left(X_{t_{i+1}} - X_{t_i}\right)^2 \to 0.$$

by uniform continuity of f''. We now show this converges in probabilty to $\int_0^t f''(X_s) d\langle X \rangle_s$.

The main idea is to show that the random measure μ_n that assigns a weight $(X_{t_{i+1}} - X_{t_i})^2$ to each point $\{t_i\}$ in fact converges in probability to $d\langle X \rangle_s$. Indeed, let D be the set of all t_is in any mesh (and let the mesh increase) and by the Riemann Sum formula for the quadratic covariation we get they agree on all compact intervals $[0, r]$ in probability and thus we can extract a subsequence for which the convergence is almost sure. From here, we can show that $\mu_n \to \mathbb{1}_{[0,t]}(r) d\langle X, X \rangle_r$ in the a.s. sense and we get the desired result. The extension to $p > 1$ is straightforward although it needs some more technical care for the covariation term. $\square$

Remark (Closure of semimartingales under C^2-transforms). Ito's Formula shows in fact that semimartingales are closed under twice differentiable functions since

$$f(X_t) - f(X_0) = \underbrace{\int_0^t f'(X_s) dM_s}_{\text{local}} + \underbrace{\int_0^t f'(X_s) dA_s + \frac{1}{2} \int_0^t f''(X_s) d\langle M \rangle_s}_{\text{FV}} \tag{6.45}$$

Corollary 6.21 (*Stochastic Integration by Parts*)
If X and Y are two semimartingales we have

$$X_tY_t = X_0Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \langle X, Y\rangle_t \tag{6.46}$$

Proof. Take $p = 2$ and $F(x, y) = xy$ in Ito's Formula. □

Corollary 6.22 (*Time-Dependent Ito's Formula*)
For a twice-differentiable time-dependent function $F(t, x) : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$, we have

$$F(t, B_t) = F(0, 0) + \int_0^t \frac{\partial F}{\partial x}(s, B_s)\, dB_s + \int_0^t \left(\frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}\right)(s, B_s)\, ds \tag{6.47}$$

Proof. Use Ito's Lemma with $X_t^1 = t$ and $X_t^2 = B_t$ and note $dt\, dB_t = 0$ while $d\langle B\rangle_s = ds$. □

> **Remark.** We often want to use Ito's Formula for a function that is only twice differentiable on some open set U. One can then take some open set $V \subseteq U$ with $\bar{V} \subseteq U$ and then just stop the process at $\tau = \inf\{t \geq 0 : X_t \notin V\}$.

6.5. Applications of Ito's Formula

We now present some useful consequences of Ito's Formula:

6.5.1. The Expoenntial Martingale

One striking consequence of Ito's Formula is the existence of the exponential (local) martingale which extends the exponential martingale for processes with independent increments:

Theorem 6.23 (*Stochastic Exponential*)
Let M be a continuous local martingale. For $\lambda \in \mathbb{C}$ the process:

$$\mathcal{E}(\lambda M)_t := \exp\{\lambda M_t - \frac{\lambda^2}{2}\langle M\rangle_t\} \tag{6.48}$$

is a $\mathcal{C}$-valued continuous local martingale. This exponential is called the **stochastic exponential** or **Doléans-Dade exponential**.

Proof. Let $X_t = \lambda M_t - \frac{\lambda^2}{2}\langle M\rangle_t$ and $f(x) = \exp(x)$ and apply Ito's Formula. □

6.5.2. The Levy Characterization of Brownian Motion

Ito's Formula gives rise to the Levy Characterization for Brownian Motion which characterizes Brownian Motion entirely by the quadratic variation:

Theorem 6.24 (*Levy's Theorem*)
Let (X_t) be an $(\mathcal{F}_t)$-adapted process. The following are equivalent:

(a) X is a d-dimensional $(\mathcal{F}_t)$-Brownian Motion

(b) the X^i are continuous local martingales with $\langle X^i, X^j \rangle = \delta_{ij}\mathbb{1}\{i = j\}$ for all i, j, t.

In particular, a CLM is a Brownian Motion iff $\langle M \rangle_t = t$ or equivalently iff $M_t^2 - t$ is a CLM.

Proof. $(a) \implies (b)$ is shown. Next, the exponential martingale $\mathcal{E}(i\theta X_t) = \exp(i\theta X_t + \frac{1}{2}|\theta|^2 t)$ which is a CLM and bounded on any compact time interval and hence a true martingale. We thus get

$$\mathbb{E}[i\theta(X_t - X_s)|\mathcal{F}_s] = \exp(-\frac{1}{2}|\theta|^2(t - s)) \tag{6.49}$$

which gives that $X_t - X_s \sim \mathcal{N}(0, t - s)$ and then is a BM by the continuity assupmption. □

6.5.3. Martingales as Brownian Motion under a Time-Change

The next theorem says that strikingly, all continuous local martingales are in fact a Brownian Motion up to a time change:

Theorem 6.25 (*Dambis-Dubin-Schwarz*)
Let M be a continuous local martingale such that $\langle M, M \rangle_\infty$ a.s.. There exists a Brownian Motion $(B_s)_{s \geq 0}$ such that

$$\text{a.s. } \forall t \geq 0, \quad M_t = B_{\langle M,M \rangle_t} \tag{6.50}$$

Proof. First assume $M_0 = 0$. Define the stopping times $\tau_r := \inf\{t \geq 0 : \langle M, M \rangle_t \geq r\}$ which are increasing stopping times by construction. By assumption $\tau_r < \infty$ a.s. and is nondecreasing and left-continuous and thus has a right-limit τ_{r+} except on some negligible set N for which we set it to 0 (it is still adapted because we assume completeness).

Define $B_r := M_{\tau_r}$ for every $r \geq 0$ which is adapted wrt $\mathcal{F}_{\tau_r}$ and the sample paths $r \mapsto B_r$ are left-continuous and thus have right-limits so define B_{r+}. In fact we have $B_{r+} = B_r$ due to the following lemma. The main idea is that although $r \mapsto \tau_r$ may not be continuous (only left-continuous) we have that on the interval for which $\langle M \rangle_t$ remains constant that M should as well.

Lemma 6.26
We have a.s. for every $0 \leq a < b$ that

$$M_t = M_a \quad \forall t \in [a, b] \Leftrightarrow \langle M, M \rangle_b = \langle M, M \rangle_a \tag{6.51}$$

Proof. We have by the continuity of sample paths of M that it suffices to show the result for all fixed $0 \leq a < b$ (otherwise for the null sets, we would have an uncountable union which would not neccesarily have measure zero). One direction is immediately a consequence of the Riemann Sum formula for the Quadratic Variation. For the other direction, consider $N_t := M_t - M_{t\wedge a}$ and consider the stopping time $T_\epsilon = \inf\{t \geq 0 : \langle N \rangle_t \geq \epsilon\}$. Then $\mathbb{E}[N^2_{t\wedge T_\epsilon}] = \mathbb{E}[\langle N \rangle_{t\wedge T_\epsilon}] \leq \epsilon$ (since the stopped process is a true martingale) and thus on $A = \{\langle M, M\rangle_a = \langle M, M\rangle_b\} \subseteq \{T_\epsilon \geq b\}$ we have that

$$\mathbb{E}[\mathbb{1}_A N_t^2] = \mathbb{E}[\mathbb{1}_A N^2_{t\wedge T_\epsilon}] \leq \mathbb{E}[N^2_{t\wedge T_\epsilon}] \leq \epsilon \tag{6.52}$$

so taking $\epsilon \downarrow 0$ we get $N_t = 0$ on A. □

We thus get that sample paths of B_r are in fact continuous. A simple calculation shows that B_s and $B_s^2 - s$ are in fact $(\mathcal{F}_{\tau_r})$-adapted u.i. martingales by the Optional Stopping Theorem. We thus get that

$$\langle B, B\rangle_t = t \tag{6.53}$$

by the uniqueness of the quadratic variation and by the Levy characterization we get that B is a $(\mathcal{F}_{\tau_r})$-Brownian Motion. Finally, note by definition that

$$B_{\langle M,M\rangle_t} = M_{\tau_{\langle M,M\rangle_t}} \tag{6.54}$$

by definition and since $\tau_{\langle M,M\rangle_t} \leq t \leq \tau_{\langle M,M\rangle_t+}$ we get the result if we assume $M_0 = 0$. For $M_0 \neq 0$ just add it back and we are done. □

6.6. Martingales Under Change of Measure

We now study how the notions of martingale and semimartingale change under absolutely continuous change of measure.

6.6.1. Exponential Tilting

We start with the motivating example of exponential tilting which is commonly used in information theory.

Definition 6.27 (*Moment Generating Functions*)
Suppose X is a real-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We call $m(\theta) := \mathbb{E}[e^{\theta X}]$ the **moment generating function (mgf)** and $\kappa(\theta) := \log m(\theta)$ the **cumulant generating function (cgf)**.

Definition 6.28 (*Exponential Tilting*)
Suppose X is a real-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $m(\theta) < \infty$. We define the **exponential tilted measure** $\mathbb{P}_\theta$ by the Radon-Nikodym derivative:

$$\frac{\mathbb{P}_\theta}{\mathbb{P}} = \frac{e^{\theta X}}{m(\theta)} \tag{6.55}$$

that is, $\mathbb{P}_\theta(A) = \mathbb{E}[\frac{d\mathbb{P}}{d\mathbb{P}_\theta}; A]$. In particular,

$$\mathop{\mathbb{E}}_{\theta} X = \mathbb{E}\left[X \frac{e^{\theta X}}{m(\theta)}\right] = \frac{m'(\theta)}{m(\theta)} = \kappa'(\theta) \tag{6.56}$$

Example 6.29 (*Exponential Tilting of a Gaussian is a Mean Shift*)
Suppose $X \sim \mathcal{N}(0,1)$ which has mgf $m(\theta) := \exp(\theta^2/2)$. The exponential tilting is then given by the RN derivative

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}} = \exp(\theta X - \theta^2/2) \tag{6.57}$$

Then, under the tilted distribution $\mathbb{P}_\theta$ we have

$$\mathop{\mathbb{E}}_{\theta}[\exp(itX)] = \mathbb{E}[\exp(itX)\exp(\theta X - \theta^2/2)] = \exp(it\theta - \theta^2/2) \tag{6.58}$$

so $X \sim \mathcal{N}(\theta, 1)$ under $\mathbb{P}_\theta$.

One important application of the exponential tilting is large deviations theory (see Ch.15 of [PW25]) where the tilting is used to show that the tilting is given tightly by the exponential of the rate function. We focus on the simple case of Gaussians in the Chernoff regime:

Example 6.30 (*Large Deviations Analysis of Gaussians*)
Let $X \sim \mathcal{N}(0,1)$. Then,

$$\mathbb{P}(X \geq t) = \exp\left(-\frac{t^2}{2}(1 + o_t(1))\right) \tag{6.59}$$

Proof. By the Chernoff Bound (Markov's Inequality on the exponential), we get

$$\mathbb{P}(X \geq t) \leq \inf_{\lambda \geq 0} \frac{m(\theta)}{\exp(\lambda t)} = \inf_{\lambda \geq 0} \exp(-\lambda^2/2 - \lambda t) = \exp(-t^2/2) \tag{6.60}$$

for all $t > 0$. For the lower bound, fix $\epsilon > 0$ and consider the exponential tilting $\mathbb{P}_t$. We then get:

$$\begin{aligned}
\mathbb{P}(X \geq t) \geq \mathbb{P}(X \in [t, t+\epsilon]) &= \mathop{\mathbb{E}}_{t}\left[\frac{d\mathbb{P}}{d\mathbb{P}_t}; X \in [t, t+\epsilon]\right] && (6.61)\\
&= \mathbb{E}[\exp(-tX + t^2/2)\mathbb{1}_{X \in [t, t+\epsilon]}] && (6.62)\\
&\geq \exp(-t(t+\epsilon) + t^2/2)\mathbb{P}_t(X \in [t, t+\epsilon]) && (6.63)\\
&= \exp(-t^2/2)\exp(-t\epsilon)\mathbb{P}(N(t,1) \in [t, t+\epsilon]) && (6.64)\\
&= \exp(-t^2/2)\exp(-t\epsilon)\mathbb{P}(N(0,1) \in [0, \epsilon]) && (6.65)\\
& && (6.66)
\end{aligned}$$

where for the second to last line we use the fact that under the Exponential Change of Measure the Gaussian gets shifted to a mean t Gaussian and we get the last line by linearity. Taking $\epsilon \downarrow 0$ we get

$$\mathbb{P}(X \geq t) \geq \exp(-t^2/2)(1 + o_t(1)) \tag{6.67}$$

and we get the desired result. □

Example 6.31 (*Exponential Tilting of 2D Gaussian*)

Let $\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}\right)$ under $\mathbb{P}$ and $d\mathbb{Q}/d\mathbb{P} = \exp(\theta Y - \theta^2/2)$. Let $Z = X - aY \perp Y$. Then,

$$\mathop{\mathbb{E}}_{\mathbb{Q}}(e^{itX}) = \exp(ita\theta - t^2/2) \implies X \sim \mathcal{N}(a\theta, 1) \tag{6.68}$$

under $\mathbb{Q}$.

Exercise 6.32

What is the joint law of (X, Y) under $\mathbb{Q}$?

Solution. We have the joint characteristic function:

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathbb{Q}}(e^{i(sX+tY)}) &= \mathbb{E}\left[\exp\left(isX + itY + \theta Y - \frac{\theta^2}{2}\right)\right] \\ &= \exp\left(-\frac{\theta^2}{2}\right)\mathbb{E}\left[e^{isX+(it+\theta)Y}\right]. \end{aligned}$$

and since

$$\mathbb{E}[e^{uX+vY}] = \exp\left(\frac{1}{2}(u^2 + 2auv + v^2)\right)$$

for all $u, v \in \mathbb{C}$. Substituting $u = is$ and $v = it + \theta$ gives

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathbb{Q}}(e^{i(sX+tY)}) &= \exp\left(-\frac{\theta^2}{2} + \frac{1}{2}\left((is)^2 + 2a(is)(it+\theta) + (it+\theta)^2\right)\right) \\ &= \exp\left(i(a\theta)s + i\theta t - \frac{1}{2}(s^2 + 2ast + t^2)\right). \end{aligned}$$

This is the characteristic function of

$$\mathcal{N}\left(\begin{pmatrix} a\theta \\ \theta \end{pmatrix}, \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}\right).$$

Hence under $\mathbb{Q}$,

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} a\theta \\ \theta \end{pmatrix}, \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}\right).$$

$\square$

Example 6.33 (*Tilting of Independent Variables Preserves Tilting*)

Suppose X_i are independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $D_n = \prod_{i=1}^n \frac{\exp(\theta X_i)}{m(\theta)}$ be the product of the exponential tiltings. Now by independence we have D_n is a martingale. Consider the measure $\mathbb{Q}_n$ given by the change of measure D_n. Note the random variables X_i are indepenent under $\mathbb{Q}_n$ as well since

$$\mathop{\mathbb{E}}_{\mathbb{Q}_n}[X_i X_j] = \mathbb{E}[X_i X_j D_n] = \mathbb{E}[X_i X_j \frac{\exp(\theta X_i)}{m(\theta)} \frac{\exp(\theta X_j)}{m(\theta)}] = \mathop{\mathbb{E}}_{\mathbb{Q}_n}[X_i] \mathop{\mathbb{E}}_{\mathbb{Q}_n}[X_j] \tag{6.69}$$

by independence.

The following example is now an informal version of Girsanov's Theorem which essentially gives us a way to construct new continuous local martingales under a change of measure:

Example 6.34 (*Girsanov on Gaussians*)

Suppose we have iid $\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & a_i \\ a_i & 1 \end{pmatrix}\right)$ and $\sigma_i, \tau_i \in \mathcal{F}_{i-1}$ are bounded previsible random variables. Define the martingales:

$$M_k := \sum_{i=1}^{k} \sigma_i X_i \qquad L_k := \sum_{i=1}^{k} \tau_i Y_i \tag{6.70}$$

Consider the process:

$$D_k := \prod_{i=1}^{k} \exp(\tau_i Y_i - \frac{1}{2}\tau_i^2) = \exp(L_k - \frac{1}{2}\sum_{i=1}^{k} \tau_i^2) \tag{6.71}$$

which is also a martingale under $\mathbb{P}$ and we can define $d\mathbb{Q}_n/d\mathbb{P} = D_n$. We have under $\mathbb{Q}_n$ that $X_i \sim \mathcal{N}(a_i\tau_i, 1)$ and M_k behaves like M_k under $\mathbb{P}$ + $\sum_{i=1}^{n} \sigma_i \tau_i a_i$ which acts as the "covariation" of M with L.

6.6.2. Girsanov's Theorem

The informal statement of Girsanov's Theorem is essentially as follows:

Theorem 6.35 (*Girsanov's Theorem (Informal)*)

Suppose M and L are continuous local martingales with respect to a measure $\mathbb{P}$. Define the local martingale $D_t = \mathcal{E}(L)_t := \exp\{L_t - \frac{1}{2}\langle L\rangle_t\}$. Suppose D_t is u.i. so $D_t \xrightarrow{L^1} D$ and we can define $d\mathbb{Q}/d\mathbb{P} = D$. Then, $M - \langle M, L\rangle$ is a continuous local martingale under $\mathbb{Q}$.

There is some nuance however with regards to absolute continuity in the limit $n \to \infty$.

> **Remark.** Note it is still possible for $\nu \not\ll \mu$, ie. the limiting measures to be singular with respect one another!

To deal with this, we first recall the following fact from probability:

Theorem 6.36 (*Limit of Change of Measures*)

Suppose μ and ν are probability measures on $(\Omega, \mathcal{F})$ with $\mathcal{F}_n \uparrow \mathcal{F}$ being a filtration. Suppose for each n, $\mu_n := \mu|_{\mathcal{F}_n}$, $\nu_n := \nu|_{\mathcal{F}_n}$, and $\nu_n \ll \mu_n$ for all n. Then $D_n := d\nu_n/d\mu_n$ is a martingale under μ. If $D = \limsup_{n\to\infty} D_n$ then

$$\nu(A) = \int_A D d\mu + \nu(A \cap \{D = \infty\}) \tag{6.72}$$

Exercise 6.37 (*AC Measures are not AC in the limit*)
Let $\Omega = \{0,1\}^{\mathbb{N}}$ and $\mathcal{F} = \{A \times \prod_{i=n+1}^{\infty}\{0,1\}$ for $A \subseteq \{0,1\}^n\}$. Let $X_i(\omega) = \omega_i$ be the coordinate maps. Now consider the measures:

- $\mu = \otimes_{i=1}^{\infty}\text{Ber}(p)$ the X_i are iid $\text{Ber}(p)$ random variables
- $\nu = \otimes_{i=1}^{\infty}\text{Ber}(q)$ the X_i are iid $\text{Ber}(q)$ random variables.

with $0 < p < q < 1$. Now we have for $\theta > 0$ that the mgf $m(\theta) = \mathbb{E}[\exp(\theta X_i)] = pe^{\theta} + 1 - p$. Choose θ such that under an exponential tilting, the mean aligns, ie.

$$q = \mathbb{E}_{\theta}[X_i] = \frac{pe^{\theta}}{pe^{\theta} + 1 - p} \implies \exp(\theta) = \frac{q(1-p)}{p(1-q)} \tag{6.73}$$

Then, define

$$D_n := \prod_{i=1}^{n} \exp(\theta X_i)/m(\theta) = \left(\frac{q(1-p)}{p(1-q)}\right)^{S_n} \left(\frac{1-q}{1-p}\right)^{n} \tag{6.74}$$

where $S_n = \sum_{i=1}^{n} X_i$. Now, define $d\nu_n/d\mu_n = D_n$ and define $D = \limsup_{n\to\infty} D_n$. Now, note D_n is a martingale under μ and under ν_n but not under ν. To see this, note under ν, $S_n = nq$ so

$$D_n = (q/p)^{nq}[(1-q)/(1-p)]^{n(1-q)}e^{o(n)} = e^{o(n)}\exp\{nh(p;q)\} \tag{6.75}$$

where $h(;)$ is the binary KL divergence. So as $n \to \infty$ we get that $D_n \to \infty$ a.s. under ν so ν is not absoltuely continuous with respect to μ (although $\nu_n \ll \mu_n \ll \nu_n$ for each finite n).

On the other hand, if in the $\mathcal{F}_\infty$ σ-algebra we have mutual absolute continuity, we retain mutual absolute continuity for the restriction. In fact, we get the Radon-Nikodym derivatives on a filtered probability space (in the mutually AC case) forms an (a.s.) positive ui martingale:

Proposition 6.38 (*Radon-Nikodym Derivative of Mutually AC Measures is a Positive Martingale*)
Let $\mathbb{P}, \mathbb{Q}$ be probability measures on a filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$ and suppose $\mathbb{Q} \ll \mathbb{P}$ on $\mathcal{F}_\infty$. For every $t \in [0, \infty]$, let

$$D_t := \frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t} \tag{6.76}$$

be the Radon-Nikodym derivative restricted to $\mathcal{F}_t$ (denote $D_\infty = D$). The process (D_t) is a u.i. martingale with respect to $\mathbb{P}$ and thus has a cadlag modification. For each stopping time T, we additionally have

$$D_T = \frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_T} \tag{6.77}$$

If additionally $\mathbb{P} \ll \mathbb{Q}$ on $\mathcal{F}_\infty$, we have

$$\inf_{t\geq 0} D_t > 0 \qquad \mathbb{P} \text{ a.s.} \tag{6.78}$$

Proof. To show D is a u.i. martingale with respect to $\mathbb{P}$, let $A \in \mathcal{F}_t$ and note

$$\mathbb{Q}(A) = \mathop{\mathbb{E}}_{Q}[\mathbb{1}_A] = \mathop{\mathbb{E}}_{P}[D\mathbb{1}_A] = \mathop{\mathbb{E}}_{P}[\mathbb{1}_A \mathop{\mathbb{E}}_{P}[D \mid \mathcal{F}_t]] \tag{6.79}$$

so by the uniqueness of the Radon-Nikodym derivative we have $D_t = \mathbb{E}_P[D \mid \mathcal{F}_t]$ so (D_t) is closed and thus a ui martingale and we can thus take a cadlag modification of it (we assume as usual, right continuity and completeness of the filtration). For the second assertion, note the (D_t) is a u.i. martingale so we can apply the Optional Stopping Theorem for closed martingales. For the last assertion, define $T_\epsilon := \inf\{t \geq 0 : D_t < \epsilon\}$ and note that

$$\mathbb{Q}(T_\epsilon) = \mathop{\mathbb{E}}_{\mathbb{P}}[\mathbb{1}_{T_\epsilon < \infty} D_T] \leq \epsilon \tag{6.80}$$

so we get $\mathbb{Q}(\cap\{T_{1/n} < \infty\}) = 0$ and the result since $\mathbb{P} \ll \mathbb{Q}$. □

Lemma 6.39 (*Stochastic Logarithmn*)
Let D be a CLM taking (strictly) positive values a.s.. There exists a unique local martingale L such that

$$D_t = \exp\left(L_t - \frac{1}{2}\langle L, L\rangle_t\right) = \mathcal{E}(L)_t \tag{6.81}$$

Moreover,

$$L_t = \log D_0 + \int_0^t D_s^{-1} dD_s \tag{6.82}$$

The CLM L is known as the **stochastic logarithmn** of D.

Proof. Uniqueness follows since if L, L' are two such processes we get that $L_t - L_t' = \frac{1}{2}\langle L'\rangle_t - \langle L\rangle_t$ which is both a CLM and a FV process and is hence 0 giving $L_t = L_t'$ a.s.. Applying Ito's Formula to $f(x) = \log x$ (since $D_t > 0$) gives

$$\log D_t = \log D_0 + \int_0^t \frac{1}{D_s} dD_s - \frac{1}{2}\int_0^t \frac{1}{D_s^2} d\langle D\rangle_t = L_t - \frac{1}{2}\langle L\rangle_t \tag{6.83}$$

letting L be as in the statement. □

We now move on to the formal proof of Girsanov's Theorem. The heuristic idea is essentially that X is a martingale under $\mathbb{Q}$ iff DX is a martingale under $\mathbb{P}$. From here, we have by Ito's product formula that

$$\begin{aligned} d(D\tilde{M}_t) &= \tilde{M}_t dD_t + D_t d\tilde{M}_t + \langle D, \tilde{M}\rangle_t & (6.84)\\ &= \tilde{M}_t dD_t + D_t(dM_t - d\langle M, L\rangle_t) + D_t d\langle M, L\rangle_t & (6.85)\\ &= \tilde{M}_t dD_t + D_t dM_t & (6.86) \end{aligned}$$

which is a CLM.

Theorem 6.40 (*Girsanov's Theorem*)
Let $\mathbb{P}$ and $\mathbb{Q}$ be mutually absolutely continuous measures on $\mathcal{F}_\infty$ and let (D_t) be a cadlag martingale with $(D_t) = \frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t}$ for each $t \geq 0$. Let L be the stochastic logarithmn of D, ie. $D_t = \mathcal{E}(L)_t$. If M is a continuous local martingale under $\mathbb{P}$, then

$$\tilde{M} = M - \langle M, L\rangle \tag{6.87}$$

is a continuous local martingale under $\mathbb{Q}$.

Proof. Suppose $D_t = \mathcal{E}(L)_t$. Let T be a stopping time and X be an adapted process. If $(XD)^T$ is a martingale under $\mathbb{P}$ we claim X^T is a martingale under $\mathbb{Q}$. We have

$$\underset{\mathbb{Q}}{\mathbb{E}}[|X_{T\wedge t}|] = \underset{\mathbb{P}}{\mathbb{E}}[|X_{T\wedge t}D_{T\wedge t}|] < \infty \tag{6.88}$$

and so $X_t^T \in L^1(\mathbb{Q})$. From here, one can simply check the martingale property under $\mathbb{Q}$ by using the definition of a martingale and the stopping time form of Proposition 6.38. Finally, by Ito's Product Formula and bilinearity of the integral operator, we get

$$\tilde{M}_t D_t = M_0 D_0 + \int_0^t \tilde{M}_s dD_s + \int_0^t D_s dM_s - \int_0^t D_s\, d\langle M, L\rangle_s + \langle M, D\rangle_t \tag{6.89}$$

$$= M_0 D_0 + \int_0^t \tilde{M}_s dD_s + \int_0^t D_s dM_s \tag{6.90}$$

using the fact that $d\langle M, L\rangle_s = D_s^{-1} d\langle M, D\rangle_s$. This concludes the proof since now $(\tilde{M})$ is a local martingale under $\mathbb{P}$. □

Corollary 6.41

A continuous local martingale under a mutually absolutely continuous change of measure remains a semimartingale with decomposition $M = \tilde{M} + \langle M, L\rangle$ where L is the stochastic logarithmn.

Corollary 6.42

Stochastic integrals agree under absolutely continuous change of measure.

Corollary 6.43

Brownian motion under AC change of measure remains a Brownian Motion.

Proof. Use Levy's characterization. □

> **Remark.** Typically we work backwards when doing a change of measure and construct $\mathbb{Q}$ and u.i. D as follows: start with a local martingale L such that $L_0 = 0$ and $\langle L, L\rangle_\infty < \infty$. Then $\lim_{t\to\infty} L_t$ exists a.s. and $\mathcal{E}(L_t)$ is a nonnegative CLM and thus a supermartingale which converges a.s. to
>
> $$\mathcal{E}(L)_t = \exp(L_\infty - \frac{1}{2}\langle L\rangle_t) \tag{6.91}$$
>
> which has $\mathbb{E}[\mathcal{E}(L)_t] \leq 1$ by Fatou's Lemma. If in fact $\mathbb{E}[\mathcal{E}(L)_t] = 1$ we get that $\mathcal{E}(L_t)$ is a u.i. martingale and so letting $d\mathbb{Q}/d\mathbb{P} = \mathcal{E}(L)_\infty$ satisfies the conditions of Girsanov.

6.6.3. Novikov and Kazamaki Criterions

The above heuristic is one way of going backwards and constructing a stochastic exponential that corresponds to the Radon-Nikodym derivatives such that the resulting change of measure is a u.i. martingale (and we are thus in the mutually absolutely continuous case rather than the singular one). The following criterions due to Novikov and Kazamaki give more general criterions:

Theorem 6.44 (*Novikov and Kazamaki Criterions*)
Let L be a CLM with $L_0 = 0$. Consider the following properties:

(a) (Novikov's Criterion) $\mathbb{E}[\exp(\frac{1}{2}\langle L\rangle_\infty)] < \infty$

(b) (Kazamaki's Criterion) L is a u.i. martingale and $\mathbb{E}[\exp\frac{1}{2}L_\infty] < \infty$

(c) $\mathcal{E}(L)_t$ is a u.i. martingale

Then a) $\implies$ b) $\implies$ c).

6.6.4. The Cameron-Martin Theorem

A remarkable application of Girsanov's Theorem is the so-called Cameron-Martin formula which gives the law of a a Brownian motion with an additive (deterministic) drift h.

Definition 6.45
The set $\mathcal{H}$ of all functions h such that $\dot{h} \in L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), dt)$ is called the **Cameron-Martin space**.

Theorem 6.46 (*Cameron-Martin Theorem*)
Let $h \in \mathcal{H}$ be a function in the Cameron-Martin space and let $g = \dot{h}$. Let $L_t := \int_0^t g(s)dB_s$. Then the change of measure

$$\begin{aligned} D = \frac{d\mathbb{Q}}{d\mathbb{P}} &= \mathcal{E}(L) = \exp\left(L_t - \frac{1}{2}\langle L\rangle_t\right) \\ &= \exp\left(\int_0^\infty g(s)dB_s - \frac{1}{2}\int_0^\infty g(s)^2 ds\right) \end{aligned}$$

is a u.i. martingale and by Girsanov's Theorem we have that $B_t - \langle B, L\rangle = B_t - h(t)$ is a $(\mathcal{F})_t$ Brownian Motion under $\mathbb{Q}$. In particular, we get the change of measure formula:

$$\mathbb{Q}(\tilde{B} + h \in A) = \mathop{\mathbb{E}}_{\mathbb{P}}[\mathbb{1}_{B\in A}\frac{1}{D}] \tag{6.92}$$

where D is given as above.

Proof. Apply Novikov's Criterion to L and then use Girsanov's Theorem along with the Levy characterization. □

Specialized to Wiener spaces we get the following:

Corollary 6.47 (*Cameron-Martin formula*)

Let $(B_t)_{0\leq t\leq T}$ be a standard Brownian motion, and let $h \in \mathcal{H}$ be a function in the Cameron-Martin space. Then for every nonnegative measurable functional $\Phi : C([0,T],\mathbb{R}) \to [0,\infty]$,

$$\mathbb{E}[\Phi(B+h)] = \mathbb{E}\left[\Phi(B)\exp\left(\int_0^T \dot{h}(s)\,dB_s - \frac{1}{2}\int_0^T \dot{h}(s)^2\,ds\right)\right].$$

Example 6.48 (*Independence of Brownian Motion with Drift from Hitting Times*)

Let $h(t) = \mu t \in \mathcal{H}$ and $g(t) = \dot{h} = \mu$. Let (B_t) be a $(\mathcal{F}_t)$-Brownian Motion and let $\tau = \inf\{t \geq 0 : |B_t| = 1\}$. Then for $L_t = \int_0^t g(s)dB_s$ and change of measure $d\mathbb{Q}/d\mathbb{P} = \exp(L)_t$, we have that $B_\tau \perp \tau$ under $\mathbb{Q}$. That is, under $\mathbb{Q}$ the time for the Brownian Motion under drift $B_t + \mu t$ to hit 1 or -1 is independent of the value at that point (under $\mathbb{Q}$)!

Another particularly illuminating example of Girsanov's Theorem is in Options Pricing. Intuitively, for a portfolio $(\alpha,\beta)_t$ where α is the amount of stock you buy S and β is the amount of bonds M that you buy at a time t with a compounding rate r. In general, the discounted price of an option $e^{-rt}V_t$ may not be a martingale but under an appropriate change of measure through Girsanov, that the discounted price is indeed a martingale.

7. (L18): Stochastic Differential Equations

7.1. Ito SDEs

Definition 7.1 (*Stochastic Differential Equation*)

Let d and m be positive integers, $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \to M_{d\times m}(\mathbb{R})$ and $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ be locally bounded and measurable. An **Ito stochastic differential equation** is an equation of the form

$$dX_t = b_t(X_t)dt + \sigma_t(X_t)dW_t \tag{7.1}$$

where W is a m-dimensional Wiener Process.

The SDE is defined rigorously only in the context of the integral form, and is mainly a notational shorthand. We define a solution to an SDE as follows:

Definition 7.2 (*Weak Solution of an Ito SDE*)

A **weak solution of the Ito SDE** given by Equation 7.1, denoted $E(\sigma, b)$ consists of

(i.) A filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t),\mathbb{P})$ where $\mathcal{F}$ is right continuous and complete.

(ii.) A m-dimensional $(\mathcal{F}_t)$-adapted Wiener process $W = (W^1,\cdots,W^m)$ starting at 0.

(iii.) A d-dimensional $(\mathcal{F}_t)$-adapted process X with continuous sample paths such that

$$X_t = X_0 + \int_0^t b_s(X_s)\,ds + \int_0^t \sigma_s(X_s)\,dW_s \tag{7.2}$$

That is, for each $i \in \{1, \cdots, d\}$, we have

$$X_t^i = X_0^i + \int_0^t b_s^i(X_s)ds + \sum_{j=1}^{m} \int_0^t \sigma_s^{ij}(X_s)dW_s^j + \tag{7.3}$$

If in addition, $X_0 = x \in \mathbb{R}^d$ is deterministic, than we say X is a (weak) solution to $E_x(\sigma, b)$.

Definition 7.3 (*Strong Solution of an Ito SDE*)
A weak solution where X_t is adapted to the filtration generared by the Wiener prcoess is said to be a **strong solution**.

Remark. Intuitively, one can imagine that "all" the randomness of the solution comes from the Wiener process.

Intuitively, one can think of an SDE as describing a system governed by an underlying ODE but perturbed by some noise. There are several notions of existence and uniqueness of an SDE.

Definition 7.4 (*Notions of Existence of Solutions to Ito SDEs*)
For the equation $E(\sigma, b)$ we say that there is:

(a) **weak existence** of a solution if for every $x \in \mathbb{R}^d$, there exists a (weaK) solution of $E_x(\sigma, b)$.

(b) **weak uniqueness** if in addition, for every $x \in \mathbb{R}^d$ all solutions of $E_x(\sigma, b)$ have the same law.

(c) **pathwise uniqueness** if whenever the filtered probability space and the Wiener process are fixed, two solutions X and X' with $X_0 = X_0'$ are a.s. indistinguishable.

Example 7.5
The SDE

$$dX_t = \text{sgn}(X_t)dW_t, \qquad X_0 = y \tag{7.4}$$

where $y \in \mathbb{R}^d$ has weak existence and uniqueness but not pathwise uniqueness.

Proof. We have by the associativity of the integral that for a Brownian Motion starting at $\beta_0 = y$ that

$$B_t := \int_0^t \text{sgn}(\beta_s)d\beta_s \implies \beta_t = y + \int_0^t \text{sgn}(\beta_s)d\beta_s$$

since B is a continuous martingale with $\langle B \rangle_t = t$ (use the Ito Isometry) we have that B is a BM starting at 0. So β_t solves the stochastic differential equation and that any solution must be a BM. Pathwise uniqueness however fails since both β and $-\beta$ solve the equation (take $y = 0$). $\square$

The following theorem shows that pathwise uniqueness implies weak uniqueness:

Theorem 7.6 (*Yamada-Watanabe*)

If both weak existence and pathwise uniqueness hold, then weak uniqueness also holds. Moreover, for any choice of the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and of the $(\mathcal{F}_t)$-Brownian Motion B, there exists for every $x \in \mathbb{R}^d$ a unique strong solution of $E_x(\sigma, b)$.

7.2. The Lipschitz Case

In the case of SDEs with Lipschitz coefficients we will be able to directly construct strong solutions (ie. those adapted to the Brownian Motion) and show pathwise uniqueness.

Part II. The Analytical View

8. (L19): Markov Processes

> **Remark.** In lecture, we covered SDEs first and then discussed Markov Processes but I follow the book and cover Markov Processes first (chapter 6) before jumping back to chapter 8 to make the connection a bit clearer.

8.1. Markov Chains

We first consider the discrete case to motivate our discussion of semigroups and generators.

Definition 8.1 (*Markov Chain*)

A discrete-time Markov process, or **Markov Chain** on a finite state space $E = \{1, \cdots, k\}$ is a collection of random variables $(X_n)_{n \geq 0}$ with $X_n \in E$ with dynamics are specified by a **transition matrix** $P \in R^{k \times k}$ with entries

$$P_{xy} = \mathbb{P}(X_{n+1} = y \mid X_n = x, (X_i)_{i \leq n})$$

for all $n \in \mathbb{N}$.

Lemma 8.2 (*Discrete-Time Chapman-Kolomogorov Equations*)

For any $1 \leq k \leq n - 1$ we have $\mathbb{P}(X_n = y | X_0 = x) = (P^n)_{x,y}$ Namely, the transition matrix for n steps of the chain is simply P^n

Proof. We have

$$\begin{aligned}\mathbb{P}(X_n = y \mid X_0 = x) &= \sum_z \mathbb{P}(X_n = y, X_k = z \mid X_0 = x) \\ &= \sum_z \mathbb{P}(X_k = z \mid X_0 = x)\mathbb{P}(X_0 = x, X_k = z) \\ &= \sum_z (P)^k_{x,z}(P^{n-k})_{z,y} \\ &= P^n_{x,y}\end{aligned}$$

□

Definition 8.3 (*State Space*)

For a Markov Chain with transition matrix P we say two $x \neq y \in [n]$ are part of the same **communicating class** if there exists some $t > 0$ such that $P^t_{xy} > 0$ (ie. there is some path from x to y). The communicating classes form a partition of $[n]$ and we say that the set of such classes are the **state space** S of the Markov Chain.

Definition 8.4 (*Irreducibility*)

A Markov Chain is said to be **irreducible** if the state space S has $|S| = 1$.

Definition 8.5 (*Aperiodicity*)

A Markov Chain is said to be **aperiodic** if $P^t_{ii} < 1$ for all $t > 0$.

Theorem 8.6 (*Perron-Frobenius Theorem*)

If the Markov Chain is irreducible and aperiodic then 1 is a simple eigenvalue of P and all other eigenvalues have modulus less than 1. The associated left eigenvector $\pi^* P = \pi^*$ is positive and is known as the **stationary distribution**, ie. $P^n \to 1\pi^*$ in limit.

8.2. Markov Kernels and Semigroups

We now discuss the general case beyond discrete state spaces:

Definition 8.7 (*Markov Kernel*)

Let $(E, \mathcal{E})$ be a measurable space. A mapping $Q : E \times \mathcal{E} \to [0, 1]$ is called a **Markovian Transition Kernel** (or **Markov Kernel**) if

(a) For every $x \in E$, the map $\mathcal{E} \ni A \mapsto Q(x, A)$ is a proability measure on $(E, \mathcal{E})$.

(b) For every $A \in E$, $E \ni x \to Q(x, A)$ is $\mathcal{E}$-measurable.

Definition 8.8 (Qf)

If $f : E \to \mathbb{R}$ is bounded and measurable, the function Qf defined by

$$Qf(x) = \int_E Q(x, dy) f(y) = \mathop{\mathbb{E}}_{Q_x}[f] \tag{8.1}$$

is also bounded measurable.

Definition 8.9 ($C_b(E)$)

We denote by $C_b(E)$ the vector space of all bounded measurable functions on E with the uniform norm $\|f\| = \sup\{|f(x)| : x \in E\}$.

Definition 8.10 (*Transition Semigroup*)

A collection $(Q_t)_{t\geq 0}$ of transition kernels on E is called a **transition semigroup** if

(i.) (Identity) For every $x \in E, A \in \mathcal{E}$, $Q_0(x, A) = \mathbb{1}[x \in A]$, ie. $Q_0(x, dy) = \delta_x(y)$.

(ii.) (Chapman-Kolomogorov Identity) For every $s, t \geq 0$ and $A \in \mathcal{E}$,

$$Q_{t+s}(x, A) = \int_E Q_t(x, dy) Q_s(y, A) \tag{8.2}$$

(iii.) (Measurability) For every $A \in \mathcal{E}$, the function $(t, x) \mapsto Q_t(x, A)$ is measurable with respect to the σ-algebra $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{E}$.

Now recall the definition of a contraction:

Definition 8.11 (*Contraction*)

A mapping f on a metric space (M, d) to itself is called a **contraction** iff

$$d(F(x), F(y)) \leq k d(x, y) \tag{8.3}$$

for $k \in [0, 1]$

Remark. We use the weaker version of a contraction where k can be equal to 1 in this case.

Lemma 8.12 (Qf *is a contraction*)

Let Q be a Markov Transition Kernel $f \in C_b(E)$. The map $f \mapsto Qf$ is a contraction.

Proof. Note that

$$\|Qf\|_\infty = \sup \left| \int_E Q(x, dy) f(y) \right| \leq |\sup_E f| | \int_E Q(x, dy)| = |\sup_E f| = \|f\|_\infty \tag{8.4}$$

for any $f \in C_b(E)$. □

Remark. Viewed as a contraction between bounded functions, the Chapman-Kolomogorov Identity is equivalent to the assertion that $Q_{t+s} = Q_t Q_s$.

Lemma 8.13 ((Q_t) *is a semigroup*)
The collection (Q_t) where each Q_t is associated with the map $f \mapsto Q_t f$ is a semigroup.

Proof. The identity map is given by Q_0 and associativity follows from the Chapman-Kolomogorov Identity. □

8.3. Markov Processes

Definition 8.14 (*Markov Process*)
Let (Q_t) be a semigroup on E. A **Markov Process** with respect to (Q_t) is an $(\mathcal{F}_t)$-adapted process (X_t) with values in E such that for every $s, t \geq 0$, $f \in C_b(E)$, we have

$$\mathbb{E}[f(X_{s+t})|\mathcal{F}_s] = Q_t f(X_s) \tag{8.5}$$

That is, the conditional law of $f(X_{s+t})$ with respect to $\mathcal{F}_s$ depends only on X_s and the transition matrix.

Definition 8.15
For a Markov process (X_t) with respect to a semigroup (Q_t), we have that

$$\mathbb{P}[X_{s+t} \in A \mid X_r;\ 0 \leq r \leq s] = Q_t(X_s, A). \tag{8.6}$$

This is called the **Markov Property**.

Corollary 8.16
Let X_0 have law $\gamma(dx)$. Then if $0 < t_1 < t_2 < \cdots < t_p$ and $f_0, f_1, \cdots, f_p \in C_b(E)$, then

$$\mathbb{E}[f_0(X_0)f_1(X_{t_1})\cdots f_p(X_{t_p})] = \int f_0(x_0)\gamma(dx0) \int Q_{t_1} f_1(x_1)(x_0, dx_1) f_2(x_2) Q_{t_2-t_1}(x_1, dx_2)$$
$$\cdots \int f_p(x_p) Q_{t_p - t_{p-1}}(x_{p-1}, dx_p)$$

Does a markov process exist for any Markov semigroup? To show this, we make use of the Kolomogorov Extension Theorem. See Appendix A for details.

Theorem 8.17 (*Existence of a Markov Process*)
Let E be a Polish Space with Borel σ-algebra and let (Q_t) be a transition semigroup on E. Let γ be a probability measure on E. There exists a unique probability measure P on $\Omega^* = E^{\mathbb{R}_+}$ under which the canonical process (X_t) is a Markov process with transition semigroup (Q_t) and the law of X_0 is

γ.

Proof. Define a probability measure in the same way as in the previous corollary. The consistency relation follows from the Chapman-Kolomogorov Identity: suppose we have some finite $V = \{(t_1, \cdots, t_p)\}$ and $U = \{(t_{j(1)}, \cdots, t_{j(k)})\}$ is a finite subset. $\square$

8.4. Feller Semigroups, Resolvents, and Generators

A key motivation for us will now be to describe the semigroup (Q_t) without having to specify all the finite dimensional marginals. Indeed, for Markov Chains, we have found that we can fully describe the system with only a *single* transition matrix P whereas for the continuous time case we need to specify an entire collection. Fortunately, for sufficiently nice semigroups (Feller semigroups), such an object *does* exist and is called the generator. Heuristically, we define it as the derivative of the operator $f \mapsto Q_t f$ at time 0. Then,

$$\frac{d}{dt} Q_t = \lim_{s \downarrow 0} \frac{Q_{t+s} - Q_t}{s} = \lim_{s \downarrow 0} \left(\frac{Q_s - I}{s} \right) Q_t = L Q_t$$

which suggests $Q_t = \exp(tL)$ (note we are dealing with operators here so this is purely heuristic). Unfortunately, howeever, the generator L can be unbounded and this does not hold in general. Nevertheless, it is true that $(L, D(L))$ will characterize the Feller semigroup.

> **Remark.** For this section, we assume that all topological spaces E we mention are Polish.

8.4.1. Feller Semigroups

We first define a notion of what it means for the semigroup to be "nice", namely that there are not large jumps for small differences in time.

Definition 8.18

A function $f : E \to \mathbb{R}$ tends to 0 at infinity if, for every $\varepsilon > 0$, there exists a compact subset K of E such that $|f(x)| \leq \varepsilon$ for every $x \in E \backslash K$.

Intuitively, the function vanishes as you exit an increasingly large ball.

Definition 8.19 ($C_0(E)$)

We denote by $C_0(E)$ the set of all continuous real functions on E that tend to 0 at infinity. Under the uniform norm, $C_0(E)$ is a Banach Space.

Definition 8.20 (*Feller Semigroup and Processes*)

A transition semigroup (Q_t) on E is said to be **Feller semigroup** if

(i) $Q_t f \in C_0(E)$ for all $f \in C_0(E)$

(ii) $\|Q_t f - f\| \to 0$ as $t \to 0$ for all $f \in C_0(E)$

A Markov Process with respect to a Feller Semigroup is called a **Feller process**.

Remark. Recall $Q_t f(x) = \mathbb{E}[f(X_t)|X_0 = x]$. The second condition then captures the idea that the process is not likely to make a large jump in small time. It does *not* imply sample path continuity. Indeed, one of the canonical examples of a Feller process is a Jump process.

Lemma 8.21

For a Feller semigroup Q, $f \in C_0(E)$, we have $t \mapsto Q_t f$ is uniformly continuous.

Proof. Use the second property along with the Chapman-Kolomogorov Identity and the fact that $Q_t f$ is a contraction. □

8.4.2. Resolvents

We now define the important notion of the *resolvent*, which will allow us to deal with the fact that L can become unbounded

Definition 8.22

Let $\lambda > 0$. The **λ-resolvent** of the transition semigroup (Q_t) is the linear operator $R_\lambda : C_b(E) \to C_b(E)$ defined by

$$R_\lambda f(x) = \int_0^\infty e^{-\lambda t} Q_t f(x) dt \tag{8.7}$$

for $f \in C_b(E)$ and $x \in E$.

Remark. We typically care about the case where the resolvent maps from $C_0(E)$ to $C_0(E)$ (functions that vanish at infinity) but we take the more general case here. We also remark that we require the measurability of the map $(t, x) \mapsto Q_t(x, A)$ in order to get the integral here to make sense.

Lemma 8.23 (*Properties of the Resolvent*)

We have

(i) $\|R_\lambda f\| \leq \frac{1}{\lambda}\|f\|$.

(ii) If $0 \leq f \leq 1$ then $0 \leq \lambda R_\lambda f \leq 1$

(iii) (Resolvent Equation) If $\lambda, \mu > 0$ then

$$R_\lambda R_\mu = \frac{R_\mu - R_\lambda}{\lambda - \mu} \tag{8.8}$$

Proof. (i) We have $\|R_\lambda f \leq \|f\| \int_0^\infty e^{-\lambda s} ds = \frac{1}{\lambda}\|f\|$

(ii) Apply the previous part.

(iii) Fubini.

□

Example 8.24

For a real Brownian Motion

$$R_\lambda f(x) = \int r_\lambda(y-x) f(y) dy \tag{8.9}$$

where $r_\lambda(y-x) = \frac{1}{\sqrt{2\lambda}} \exp(-|y-x|\sqrt{2\lambda})$.

Lemma 8.25

Let X be a Markov Process with semigroup (Q_t) with respect to the filtration $(\mathcal{F}_t)$. Let $h \in C_b(E)$ be nonnegative and let $\lambda > 0$. THe process $e^{-\lambda t} R_\lambda h(X_t)$ is an $(\mathcal{F}_t)$-supermartingale.

Proof. We have $e^{-\lambda t} R_\lambda h(X_t) \in C_b(E)$ (note the exponential is bounded by 2 and use the fact that $\|R_\lambda h\| \le \|f\|/\lambda$ which is bounded) and thus is L^1. By Fubini, for every $s \ge 0$

$$Q_s R_\lambda h = \int_0^\infty e^{-\lambda t} Q_{s+t} h dt \tag{8.10}$$

We thus get

$$e^{-\lambda s} Q_s R_\lambda h = \int_0^\infty e^{-\lambda(s+t)} Q_{s+t} h dt = \int_s^\infty e^{-\lambda t} Q_t h dt \le R_\lambda h \tag{8.11}$$

from which checking the supermartingale property we get the desired result. □

Lemma 8.26

For $f \in C_0(E)$, we have $R_\lambda f \in C_0(E)$ as well.

Proposition 8.27

Let $\lambda > 0$ and set $\mathcal{R} = \{R_\lambda f : f \in C_0(E)\}$. Then $\mathcal{R}$ does not depend on the choice of $\lambda > 0$ and is dense in $C_0(E)$.

Proof. The resolvent equation gives for $\lambda \neq \mu$ that $R_\lambda f = R_\mu(f + (\mu - \lambda) R_\lambda f)$ so any function of the form $R_\lambda f$ is of the form $R_\mu g$ for some $g \in C_0(E)$ (note we use the fact that $C_0(E)$ is a linear subspace). To see that it is dense, note

$$\lambda R_\lambda f = \lambda \int_0^\infty e^{-\lambda t} Q_t f dt = \int_0^\infty e^{-t} Q_{t/\lambda} f dt \to f \tag{8.12}$$

as $\lambda \to \infty$ which gives the result. □

8.4.3. Generators and The Kolomogorov Equations

Definition 8.28

Define

$$D(L) := \{f \in C_0(E) : \frac{Q_t f - f}{t} \text{ converges in } C_0(E) \text{ when } t \downarrow 0\} \tag{8.13}$$

and for every $f \in D(L)$, we define

$$Lf = \lim_{t \downarrow 0} \frac{Q_t f - f}{t} \tag{8.14}$$

We call $L : D(L) \to C(E)$ the **generator** of the semigroup $(Q_t)_{t \geq 0}$. Note $D(L)$ is a linear subspace of $C_0(E)$. $D(L)$ is referred to as the **domain** of L. We define the **domain**

LEMMA 8.29

Let $f \in D(L)$ and $s > 0$. Then $Q_s(f) \in D(L)$ and L commutes with Q_s on its domain, ie. $L(Q_s f) = Q_s(Lf)$.

Proof. For $f \in D(L)$ We note that

$$\frac{Q_t(Q_s f) - Q_s f}{t} = Q_s \left(\frac{Q_t f - f}{t} \right). \tag{8.15}$$

Taking $t \downarrow 0$ gives $Q_s f \in D(L)$ and

$$L(Q_s f) = Q_s(Lf). \tag{8.16}$$

□

We now show that the generator of a Feller semigroup characterizes the time evolution of any function in the domain of the generator. If in addition, the semigroup (Q_t) admits a density p_t, the equation simplifies further and we show the differential forms. Intuitively, the Kolomogorov Equations highlight how the density of a stochastic process defined by an SDE can be described a corresponding PDE.

PROPOSITION 8.30 (*KOLMOGOROV EQUATIONS FOR A FELLER SEMIGROUP*)

Let $(Q_t)_{t \geq 0}$ be a Feller semigroup with generator $(L, D(L))$. Then for every $f \in D(L)$ and every $t \geq 0$,

$$Q_t f - f = \int_0^t Q_s(Lf)\, ds = \int_0^t L(Q_s f)\, ds. \tag{8.17}$$

Equivalently, for $f \in D(L)$,

$$\partial_t Q_t f = Q_t(Lf) = L(Q_t f). \tag{8.18}$$

The first equality is often denoted the **Kolomogorov Forward Equation (KFE)** while the latter equality is often denoted the **Kolomogorov Backwards Equation (KBE)**.

Proof. Because on the domain L and Q_t commute, it suffices to prove either the forward or backwards equation. Now, for $f \in D(L)$,

$$\frac{Q_{t+s} f - Q_t f}{s} - Q_t L f = Q_t \left(\frac{Q_s f - f}{s} - Lf \right),$$

hence, since Q_t is a contraction on $C_0(E)$,

$$\sup_{t \geq 0} \left\| \frac{Q_{t+s} f - Q_t f}{s} - Q_t L f \right\|_\infty \leq \left\| \frac{Q_s f - f}{s} - Lf \right\|_\infty \to 0.$$

Thus the convergence is uniform in t. So the right derivative of $h(t) = Q_t f(x)$ exists at every point t and is given by $Q_t Lf(x)$. We now show the left derivative exists. Note,

$$|(h(t) - h(t-s))/s - Q_{t-s}Lf(x)| + |Q_{t-s}Lf(x) - Q_t Lf(x)| \to 0$$

where for the second term we use uniform continuity of Q_t. Then, by the Fundmanetal Theorem of Calculus, we get

$$\int_0^t Q_s(Lf) = Q_s f|_0^t = Q_t f - Q_0 f = Q_t f - f \tag{8.19}$$

which gives the desired result. $\square$

> **Remark.** The Kolomogorov Backwards Equation is often written in differential form:
>
> $$\partial_t Q_t f = L(Q_t f) \tag{8.20}$$
>
> If, in addition, the transition kernel admits a density $p_t(x, y)$ with respect to Lebesgue measure, so that
>
> $$Q_t f(x) = \int_{\mathbb{R}^d} p_t(x, y) f(y)\, dy, \tag{8.21}$$
>
> then the backward equation becomes:
>
> $$\partial_t p_t(x, y) = L_x p_t(x, y) \tag{8.22}$$
>
> Meanwhile, the Kolmogorov Forward Equation
>
> $$\partial_t Q_t f = Q_t(Lf) \tag{8.23}$$
>
> becomes, after passing to the (formal) adjoint,
>
> $$\partial_t p_t(x, y) = L_y^* p_t(x, y), \tag{8.24}$$
>
> which is the Kolmogorov forward equation (Note, here we mean by L_x the operator L where we keep y constant and similar for the adjoint). In the case of diffusion processes, the latter equation is known as the *Fokker-Planck Equation*. More on this later.

Note, the Kolmogorov forward and backward equations describe dual ways of encoding the same Markov evolution: the backward equation evolves expected observables (ie. test functions), while the forward equation evolves the law, or density, of the process.

> **Remark.** Suppose we have some Markov process $(X_t)_{t\geq 0}$ with generator L. Then,
>
> (a) (KFE) If $X_0 = x$, we can characterize the time evolution of the the process at a future time t thorugh the (formal) adjoint of the generator L^* is the infinitesimal push. Equivalently, upon integration we can get the density of X_t at a future time by solving the PDE. In other words, the KFE characterizes the time evolution of the process as a PDE with an Initial condition $X_0 = x$.
>
> (b) (KBE) The Kolmogorov backward equation describes how the expected value of a test function depends on the starting point. If
>
> $$u(t, x) = \mathbb{E}_x[f(X_t)],$$
>
> then u is a weak solution of
>
> $$\partial_t u = Lu, \qquad u(0, x) = f(x).$$
>
> Thus the backward equation evolves the test function f by the semigroup and the generator acts on the initial spatial variable x.

This intuition is slightly more easily seen in the context of simpler diffusion prcoesses, which we now discuss:

8.4.4. Diffusion Processes

We make a digression and study a certain class of continuous Markov processes called *Diffusion Processes*. This section is largely adapted from Section 5 of [KS91]. Intuitively, we refer to a solution to an Ito SDE:

$$dX_t = b(t, X_t)dt + \sum_{ij} a_{ij}(t, X_t)d(W_{ij})_t$$

where W_t is a standard Wiener process is called a Diffusion Process but we take a slightly more general take here and characterize them in terms of their generator L.

Definition 8.31 (*Diffusion Process*)
Let $E = \mathbb{R}^d$, and let $X = (X_t)_{t\geq 0}$ be a Feller process on E with semigroup $(Q_t)_{t\geq 0}$ and generator $(L, D(L))$. We say that X is a **(Kolomogorov-Feller) diffusion process** if

(i) $X_0 = x \in \mathbb{R}^d$ is deterministic and the sample paths $t \mapsto X_t$ are a.s. continuous

(ii) $C_c^\infty(\mathbb{R}^d) \subset D(L)$ (here C_c^∞ means it is smooth with compact support)

(iii) there exist functions $b : \mathbb{R}^d \to \mathbb{R}^d$ and $a : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ such that, for every $x \in \mathbb{R}^d$, the matrix $a(x)$ is symmetric and nonnegative definite, and

$$Lf(x) = \sum_{i=1}^{d} b_i(x)\,\partial_i f(x) + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\,\partial_{ij} f(x), \qquad f \in C_c^\infty(\mathbb{R}^d). \tag{8.25}$$

The vector field b is called the **drift** and the matrix-valued function a is called the **diffusion matrix**.

Remark. Intuitively, the drift term is the deterministic change of a particle while the diffusion term represents the random fluctations of a particle in a medium. Such a process can be constructed as the solution of an Ito SDE:

$$dX_t = b(t, X_t)dt + \sigma(t, W_t)dW_t$$

as described in Example 8.33

Some examples of diffusion processes:

Example 8.32 (*Brownian Motion*)
The standard Brownian Motion is a diffusion process with

$$L = \frac{1}{2}\Delta \tag{8.26}$$

where Δ is the Laplacian operator.

Example 8.33 (*Ito SDE*)
For an Ito SDE:
$$dX_t = b(t, X_t)dt + \sigma(t, W_t)dW_t$$
we have all solutions have a generator that can be calculated explicitly as:
$$L_t f(x) = \sum_{i=1}^{d} b_i(x)\partial_i f(x) + \frac{1}{2}\sum_{ij} a_{ij}\partial_{ij} f(x), \qquad a(t,x) = \sigma(t,x)\sigma(t,x)^T$$

We defer the proof. In the case of diffusion processes, the Kolomgorov Equations have a particularly nice form and are commonly used in physics.

Corollary 8.34 (*Fokker–Planck*)
Let $X = (X_t)_{t\geq 0}$ be a diffusion process on $\mathbb{R}^d$ with generator $(L, D(L))$ satisfying
$$Lf(y) = \sum_{i=1}^{d} b_i(y)\partial_i f(y) + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(y)\partial_{ij} f(y), \qquad f \in C_c^\infty(\mathbb{R}^d).$$

Suppose that the transition kernel of X admits a density $p_t(x, y)$ with respect to Lebesgue measure, so that
$$Q_t f(x) = \int_{\mathbb{R}^d} p_t(x,y) f(y)\, dy$$
or in differential form:
$$Q_t(x, dy) = p_t(x,y)dy$$
Then, for each fixed starting point $x \in \mathbb{R}^d$, the function $p_t(x, \cdot)$ satisfies the Kolomogorov Forward Equation
$$\partial_t p_t(x,y) = L_y^* p_t(x,y)$$
weakly, where the formal adjoint L_y^* is given by
$$L_y^* p_t(y) = -\sum_{i=1}^{d} \partial_{y_i}\big(b_i(y)p_t(x,y)\big) + \frac{1}{2}\sum_{i,j=1}^{d} \partial_{y_i y_j}\big(a_{ij}(y)p_t(x,y)\big). \tag{8.27}$$

Equivalently,
$$\partial_t p_t(x,y) = -\sum_{i=1}^{d} \partial_{y_i}\big(b_i(y)p_t(x,y)\big) + \frac{1}{2}\sum_{i,j=1}^{d} \partial_{y_i y_j}\big(a_{ij}(y)p_t(x,y)\big). \tag{8.28}$$

In this form, the PDE is called the **Fokker–Planck equation**.

Proof. Fix $x \in \mathbb{R}^d$ and let $\varphi \in C_c^\infty(\mathbb{R}^d)$. Since $Q_t(x, dy) = p_t(x,y)\, dy$, we have
$$Q_t\varphi(x) = \int_{\mathbb{R}^d} p_t(x,y)\varphi(y)\, dy.$$

By the Kolomogorov forward equation,

$$\partial_t Q_t \varphi(x) = L^*(Q_t\varphi)(x) = Q_t(L\varphi)(x).$$

Therefore

$$\frac{d}{dt}\int_{\mathbb{R}^d} p_t(x,y)\varphi(y)\,dy = \int_{\mathbb{R}^d} p_t(x,y)L\varphi(y)\,dy.$$

We now compute the right-hand side. Using the form of the generator,

$$\int_{\mathbb{R}^d} p_t(x,y)L\varphi(y)\,dy = \int_{\mathbb{R}^d} p_t(x,y)\left[\sum_{i=1}^d b_i(y)\partial_{y_i}\varphi(y) + \frac{1}{2}\sum_{i,j=1}^d a_{ij}(y)\partial_{y_iy_j}\varphi(y)\right]dy.$$

Since φ is compactly supported, all boundary terms vanish when integrating by parts. For the drift term,

$$\int_{\mathbb{R}^d} p_t(x,y)b_i(y)\partial_{y_i}\varphi(y)\,dy = -\int_{\mathbb{R}^d}\varphi(y)\partial_{y_i}\big(b_i(y)p_t(x,y)\big)\,dy.$$

For the second-order term, integrating by parts twice gives

$$\int_{\mathbb{R}^d} p_t(x,y)a_{ij}(y)\partial_{y_iy_j}\varphi(y)\,dy = \int_{\mathbb{R}^d}\varphi(y)\partial_{y_iy_j}\big(a_{ij}(y)p_t(x,y)\big)\,dy.$$

Hence

$$\int_{\mathbb{R}^d} p_t(x,y)L\varphi(y)\,dy = \int_{\mathbb{R}^d}\varphi(y)\left[-\sum_{i=1}^d \partial_{y_i}\big(b_i(y)p_t(x,y)\big) + \frac{1}{2}\sum_{i,j=1}^d \partial_{y_iy_j}\big(a_{ij}(y)p_t(x,y)\big)\right]dy.$$

Therefore, for every $\varphi \in C_c^\infty(\mathbb{R}^d)$,

$$\frac{d}{dt}\int_{\mathbb{R}^d} p_t(x,y)\varphi(y)\,dy = \int_{\mathbb{R}^d}\varphi(y)\left[-\sum_{i=1}^d \partial_{y_i}\big(b_i(y)p_t(x,y)\big) + \frac{1}{2}\sum_{i,j=1}^d \partial_{y_iy_j}\big(a_{ij}(y)p_t(x,y)\big)\right]dy.$$

Which gives precisely the weak formulation:

$$\partial_t p_t(x,y) = -\sum_{i=1}^d \partial_{y_i}\big(b_i(y)p_t(x,y)\big) + \frac{1}{2}\sum_{i,j=1}^d \partial_{y_iy_j}\big(a_{ij}(y)p_t(x,y)\big).$$

□

> **Remark.** Note that we make an implicit assumption that the diffusion and drift coefficients are not time-dependent, ie. are fixed with respect to time. In the case where they depend on a time parameter t, the general relationship still holds although now the generator must also be time-dependent and it is no longer possible to purely describe the system in terms of a single operator.

Later, we will revisit Fokker-Planck but in the context of invariant measures.

Corollary 8.35 (*Kolmogorov backward equation for diffusion processes*)
Let $X = (X_t)_{t\geq 0}$ be a diffusion process on $\mathbb{R}^d$ satisfying the same assumptions as in Corollary 8.34. Additionally, assume the density $p_t(x, y)$ of $Q_t(x, dy)$ is C^2 in each variable. Then, for each fixed fixed terminal point $y \in \mathbb{R}^d$, we get that

$$\partial_t p_t(x, y) = L_x p_t(x, y) \tag{8.29}$$

Equivalently,

$$\partial_t p_t(x, y) = \sum_{i=1}^{d} b_i(x)\partial_{x_i} p_t(x, y) + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\partial_{x_i x_j} p_t(x, y). \tag{8.30}$$

This PDE is called the **Kolmogorov backward equation of the Diffusion Process**.

Proof. Let $f \in C_c^\infty(\mathbb{R}^d)$ and define

$$u(t, x) := Q_t f(x).$$

By the Kolmogorov backward equation for the semigroup,

$$\partial_t Q_t f = L(Q_t f).$$

Therefore

$$\partial_t u(t, x) = L_x u(t, x).$$

Using the explicit form of the diffusion generator, this becomes

$$\partial_t u(t, x) = \sum_{i=1}^{d} b_i(x)\partial_{x_i} u(t, x) + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\partial_{x_i x_j} u(t, x).$$

Since $Q_0 = I$:

$$u(0, x) = Q_0 f(x) = f(x).$$

Now since Q_t admits a transition density $p_t(x, y)$. Then

$$u(t, x) = Q_t f(x) = \int_{\mathbb{R}^d} p_t(x, y) f(y)\, dy.$$

Applying the backward equation gives

$$\partial_t \int_{\mathbb{R}^d} p_t(x, y) f(y)\, dy = L_x \left(\int_{\mathbb{R}^d} p_t(x, y) f(y)\, dy \right).$$

Because we have that p_t is C^2 in the first and second coordinates respectively, we get

$$\int_{\mathbb{R}^d} \partial_t p_t(x, y) f(y)\, dy = \int_{\mathbb{R}^d} L_x p_t(x, y) f(y)\, dy.$$

Since this holds for every $f \in C_c^\infty(\mathbb{R}^d)$, we obtain

$$\partial_t p_t(x, y) = L_x p_t(x, y)$$

in the sense of distributions in the y-variable, and pointwise wherever the derivatives exist. □

8.4.5. The Generator Characterization of a Semigroup

We now identify the domain $D(L)$ in terms of the resolvent R_λ:

Proposition 8.36 (*Resolvent and shifted generator are inverses*)
Let $\lambda > 0$.

(a) For every $g \in C_0(E)$, $R_\lambda g \in D(L)$ and $(\lambda - L)R_\lambda g = g$.

(b) If $f \in D(L)$, $R_\lambda(\lambda - L)f = f$.

Thus, $D(L) = \mathcal{R}$ and $R_\lambda : C_0(E) \to D(L)$ and $\lambda - L : D(L) \to C_0(E)$ are inverses.

Proof. (a) Suppose $g \in C_0(E)$, we have the derivative

$$\begin{aligned}
\lim_{\varepsilon\downarrow 0} \frac{Q_\varepsilon(R_\lambda g) - R_\lambda g}{\varepsilon} &= \lim_{\varepsilon\downarrow 0} \frac{1}{\varepsilon}\left(\int_0^\infty e^{-\lambda t} Q_{\varepsilon+t} g\, dt - \int_0^\infty e^{-\lambda t} Q_t g\, dt\right) \\
&= \lim_{\varepsilon\downarrow 0} \frac{1}{\varepsilon}\left(\int_0^\infty e^{-\lambda t} Q_{\varepsilon+t} g\, dt - \int_\varepsilon^\infty e^{-\lambda t} Q_t g\, dt - \int_0^\varepsilon e^{-\lambda t} Q_t g\, dt\right) \\
&= \lim_{\varepsilon\downarrow 0} \frac{1}{\varepsilon}\left((1 - e^{-\lambda\varepsilon})\int_0^\infty e^{-\lambda t} Q_{\varepsilon+t} g\, dt - \int_0^\varepsilon e^{-\lambda t} Q_t g\, dt\right)
\end{aligned}$$

Now remark that $(1 - e^{-\lambda\varepsilon})/\varepsilon \to \lambda$. Also, by Lemma 8.21 (uniform continuity of $t \mapsto Q_t$) we have that for $T < \infty$ that

$$\begin{aligned}
\left\|\int_0^\infty e^{-\lambda t} Q_{\varepsilon+t} g\, dt - R_\lambda g\right\|_\infty &= \left\|\int_0^\infty e^{-\lambda t}\left(Q_{\varepsilon+t} g - Q_t g\right) dt\right\|_\infty \\
&\le \int_0^T e^{-\lambda t}\left\|Q_{\varepsilon+t} g - Q_t g\right\|_\infty dt + \int_T^\infty e^{-\lambda t}\left\|Q_{\varepsilon+t} g - Q_t g\right\|_\infty dt \\
&\le \left(\int_0^T e^{-\lambda t}\, dt\right)\sup_{0\le t\le T}\|Q_{\varepsilon+t} g - Q_t g\|_\infty + 2\|g\|_\infty \int_T^\infty e^{-\lambda t}\, dt \\
&= \left(\int_0^T e^{-\lambda t}\, dt\right)\sup_{0\le t\le T}\|Q_{\varepsilon+t} g - Q_t g\|_\infty + \frac{2\|g\|_\infty}{\lambda} e^{-\lambda T} \to 0
\end{aligned}$$

where we use the uniform continuity of $t \mapsto Q_t$ and the fact that $g \in C_0(E)$. One similarly gets that the second integral term goes to g and thus we get

$$\frac{d}{dt} Q_t(R_\lambda g)|_{t=0} = \lambda R_\lambda g - g \tag{8.31}$$

which gives the assertion $R_\lambda g \in D(L)$ and notably $(\lambda - L)R_\lambda g = g$.

(b) Suppose $f \in D(L)$. We have by the KFE, $\int_0^t Q_s(Lf)ds = Q_t f - f$. Thus,

$$\begin{aligned}\int_0^\infty e^{-\lambda t} Q_t f(x) dt &= \frac{f(x)}{\lambda} + \int_0^\infty e^{-\lambda t}\left(\int_0^t Q_s(Lf)ds\right) dt \\ &= \frac{f(x)}{\lambda} + \int_0^\infty \int_s^\infty e^{-\lambda t} Q_s(Lf)(x)\, dt\, ds \\ &= \frac{f(x)}{\lambda} + \int_0^\infty Q_s(Lf)(x)\left(\int_s^\infty e^{-\lambda t}\, dt\right) ds \\ &= \frac{f(x)}{\lambda} + \int_0^\infty Q_s(Lf)(x)\frac{e^{-\lambda s}}{\lambda}\, ds.\end{aligned}$$

where we apply Fubini since $|Q_s(Lf)(x)| \le \|Q_s(Lf)\|_\infty \le \|Lf\|_\infty < \infty$. We thus get $R_\lambda \lambda f(x) = f(x) + R_\lambda(Lf)(x)$.

□

The main result is thus as follows, namely that the generator determines the semigroup:

Theorem 8.37 (*Generator determines the Semigroup*)

The semigroup $(Q_t)_{t\ge 0}$ is determined by the generator L (along with the domain $D(L)$).

Proof. Let f be a nonnegative function in $C_0(E)$. We have $R_\lambda f$ is the unique element of $D(L)$ such that $(\lambda - L)R_\lambda f = f$ so $R_\lambda f = \int_0^\infty e^{-\lambda t} Q_t f(x) dt$ is well-determined for every $t > 0$. so $R_\lambda f(x) = \int_0^\infty e^{-\lambda t} Q_t f(x) dt$ is determined for every $\lambda > 0$. We have by Lerch's Theorem that $Q_t f$ is thus uniquely determined for every $f \in C_0(E)$. □

Example 8.38 (*Domain of Brownian Motion*)

For a Brownain motion, we have $L = \frac{1}{2}\Delta$ and $D(L) \subseteq \{h \in C^2(\mathbb{R}) : h, h'' \in C_0(\mathbb{R})\}$.

In general, it is hard to actually identify the domain exactly except in some special cases (such as diffusion processes). The following characterization identifies elements of the domain with martingales:

Theorem 8.39 (*Martingale Characterization of the Generator*)

Let $(Q_t)_{t\ge 0}$ be a Feller semigroup and assume (X_t^x) is a càdlàg Markov Process with respect to the semigroup $(Q_t)_{t\ge 0}$ and such that $\mathbb{P}(X_0^x = x) = 1$. Let $h, g \in C_0(E)$. The following two conditions are equivalent:

(a) $h \in D(L)$ and $Lh = g$.

(b) For every $x \in E$, we have that the process

$$h(X_t^x) - \int_0^t g(X_s^x) ds \tag{8.32}$$

is a $(\mathcal{F}_t)$-martingale.

Proof. (i) $\implies$ (ii) Let $h \in D(L)$ and $g = Lh$. We have by the KFE $Q_s h = h + \int_0^s Q_r g dr$. We thus have

$$\mathbb{E}[h(X_{t+s}^x)|\mathcal{F}_t] = Q_s h(X_t^x) = h(X_t^x) + \int_0^s Q_r g(X_t^x) dr$$

The integral term:

$$\mathbb{E}[\int_t^{t+s} g(X_r^x) dr|\mathcal{F}_t] = \int_t^{t+s} \mathbb{E}[g(X_r^x)|\mathcal{F}_t] dr = \int_t^{t+s} Q_{r-t} g(X_t^x) dr = \int_0^s Q_r g(X_t^x) dr$$

where we use conditional Fubini. Subtracting gives the result.

(ii) $\implies$ (i) We have by the martingale property

$$\mathbb{E}[h(X_t^x) - \int_0^t g(X_r^x) dr] = h(x)$$

Thus,

$$\begin{aligned} \mathbb{E}\left[h(X_t^x) - \int_0^t g(X_r^x)\, dr\right] &= h(x), \\ \mathbb{E}\left[h(X_t^x) - \int_0^t g(X_r^x)\, dr\right] &= Q_t h(x) - \int_0^t Q_r g(x)\, dr. \end{aligned}$$

Consequently,

$$\begin{aligned} Q_t h(x) - h(x) &= \int_0^t Q_r g(x)\, dr, \\ \frac{Q_t h - h}{t} &= \frac{1}{t} \int_0^t Q_r g\, dr \xrightarrow[t \downarrow 0]{C_0(E)} g. \end{aligned}$$

Thus $h \in D(L)$ and $Lh = g$. □

8.5. Sample Path Regularity for Markov Processes

We now show that we can get a càdlàg modification for a Feller process on a Feller semigroup $(Q_t)_{t \geq 0}$.

> **Remark.** Again, we assume E is locally compact and countable and is with respect to a Polish space.

Theorem 8.40

Let $(X_t)_{t \geq 0}$ be a Markov process with semigroup $(Q_t)_{t \geq 0}$ with respect to the filtration $(\mathcal{F}_t)$. Let $\tilde{F}_\infty = F_\infty$ and let $\tilde{F}_t$ be the canonical completion of $\mathcal{F}_{t+}$. Then (X_t) has a càdlàg modification $(\tilde{X}_t)_{t \geq 0}$ which is adapted to $(\tilde{F}_t)$. Moreover, $(\tilde{X}_t)$ is a Markov process with respect to $(\tilde{F}_t)$.

The proof itself is a bit technical so we provide only a sketch.

Sketch. Let E_Δ be the Alexandroff compactification of E where Δ is the point at infinity. Let $C_0^+(E)$ represent the nonnegative functions of $C_0(E)$.

The main idea behind the construction is to regularize by test functions and use topological properties of the space. By seperability, start with a countable family of nonnegative functions $(f_n) \subseteq C_0^+(E)$ that seperates E_Δ (by seperability of the metric space). Then, define

$$\mathcal{H} := \{R_p f_n; p \geq 1, n \geq 0\}$$

which still seperates points of E_Δ since $pR_p f \to f$ uniformly as $p \to \infty$ (use the arguments as in the previous proofs).

We then have for $h \in \mathcal{H}$ that for appropriate p, $e^{-pt}h(X_t)$ is a supermartingale (we have $h = R_p f_n$ for some p so just choose that one). By Doob Regularization, outside of some null set N_h, the right and left limits along a countable dense set $D \subseteq \mathbb{R}$ exists. Now, since $\mathcal{H}$ is countable, letting $N = \cup N_h$ we have right and left limits exist except on the trivial N.

The key idea now is that since $\mathcal{H}$ still seperates E_Δ, so for any there can not be any two accumulation points of either the right or left limit. So let $\tilde{X}_t(\omega) = \lim_{D \ni s \downarrow t} X_s(\omega)$.

A quick application of the definition of Feller semigroups and □

8.6. Examples of Feller Processes

Three important groups of Feller processes beyond Diffusion Processes (Ito processes) are Jump processes, Lèvy processes, and State-Branching Processes.

8.6.1. Jump Processes

We first consider Jump processes which allow for discontinuities in the time evolution of the process albeit at discrete "jumps". An example application is in finance, where one might imagine a stock following geometric Brownian Motion will jump around rather than moving in a purely continuous manner. Although the definition of the jump process is not standardized across the literature, we take the semigroup perspective of [Le 16]

Definition 8.41

Consider a Feller semigroup $(Q_t)_{t \geq 0}$ on a finite set E equipped with the discrete topology. A (pure) **jump process** is a càdlàg Markov process with respect to (Q_t) equipped with a collection of probability measures $(P_x)_{x \in E}$ such that $\mathbb{P}_x(X_0 = x) = 1$.

Remark. Note for any such process we must have a sequence of times $0 = T_0(\omega) < T_1(\omega) \leq T_2(\omega) \leq \cdots \leq \infty$ such that X_t is constant on each $t \in [T_i(\omega), T_{i+1}(\omega))$. These $T_\cdot$ times are also stopping times.

This is illustrated best in Figure 1 below:

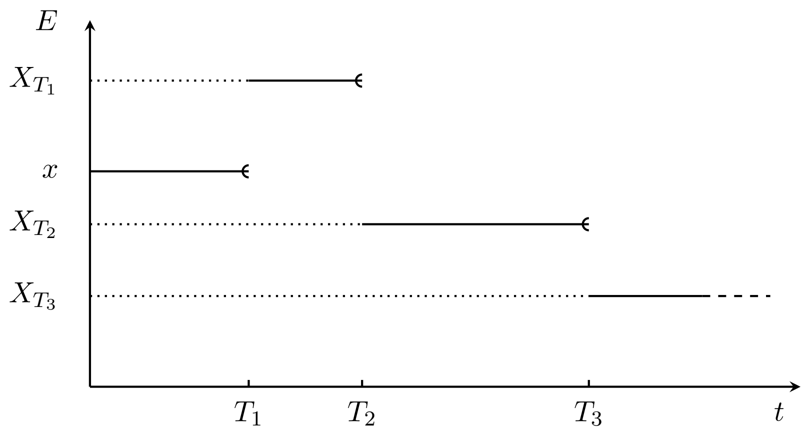

Figure 1: A sample path of a jump process under P_x.

We now characterize the first jump by an exponential law. Recall the definition:

Definition 8.42 (*Exponential Law*)
U is said to be **exponentially distributed** with parameter λ if $\mathbb{P}(U > r) = e^{-\lambda r}$

The lemma intuitively states that the first jump occurs after an exponential amount of time and is independent of how long you waited:

Lemma 8.43
Let $x \in E$. Then, there exists a real number $q(x) \geq 0$ such that T_1 is exponentially distributed with parameter $q(x)$. Additionally, if $q(x) > 0$ then T_1 and X_{T_1} are independent. If $q(x) = 0$, we have $\mathbb{P}(X_t = x, \forall t \geq 0) = 1$.

8.7. Transition and Generator Families

9. (L20) Stroock-Varadhan and Martingale Problems

This section is adapted both from lecture and lecture notes from the old, more analytical version of the course [SC19].

10. (L21) Connections to Partial Differential Equations

The main reference for Feynman-Kac is [Fol08],[KS91], [Kal21], and [KP92]. For general PDE connections we mainly refer to Chapter 7 of [Le 16].

The Kolmogorov equations establish a way of studying the time evolution of a particle undergoing a diffusion process through a parabolic partial differential equation. A beautiful insight is that a partial converse is also true: certain parabolic partial differential equations can be solved probabilistically by running a diffusion process and taking expectations of suitable path functionals and interpreting the PDE as the corresponding Kolmogorov Equation of the process.

We begin with an enlightening example relating the KBE to the heat equation and then proceed to generalize and show the so-called *Feynman-Kac Formula*. Finally, we give a brief introduction to the path integral formulation in Quantum Field Theory and show how under an analytic continuation of the Schrödinger equation with the so-called **Wick Transform** $\tau = it$, that the imaginary Schrödinger equation can be used to find the *ground state* of the wave function.

10.1. The Heat Equation

Throughout this section, we let B be a d-dimensional Brownian motion starting at x with probability measure $\mathbb{P}_x$. Recall $(B_t)_{t\geq 0}$ is Feller with semigroup

$$Q_t\varphi(x) = \int_{\mathbb{R}^d} \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|y-x|^2}{2}\right) \varphi(y)\, dy \tag{10.1}$$

10.2. The Feynman-Kac Formula

Although there are many ways to phrase the Feynman-Kac Formula, we take the perspective of [Kal21] and use it as our primary reference.

We now generalize the preceding discussion to characterize solutions to a more general class of parabolic PDEs by takign the conditional expectation of an exponentially weighted "cost" of the simulated random paths of a diffusion processes. The main result is a simple application of Itô's Lemma and the KBE for a diffusion process.

Theorem 10.1 (*Feynman-Kac Formula*)

10.3. Feynman's Path Integral Formulation and Diffusion Monte Carlo

The intuitive idea of solving parabolic PDEs through conditional expectations of a diffusion process form the basis of an alternative description of Quantum Mechanics: Feynman's **Path Integral Formulation**.

As motivation, consider first the Schrödinger Equation:

$$i\hbar\partial_t\Psi_t(\mathbf{x}) = H_t\Psi_t(\mathbf{x}); \qquad H_t := -\frac{\hbar^2}{2m}\partial_x^2 + V_t(\mathbf{x}) \tag{10.2}$$

Note this is reminiscent of the Kolmogorov Backward Equation, so heuristically we have roughly that the Schrodinger PDE describes the evolution of a Markov Process with generator $L_t = -\frac{i}{\hbar}H_t$ (note again that the generator is now time-varying; see the Remarks after Corollary 8.34). Note, however, that the density of the particle's position is actually given by $|\Psi_t(\mathbf{x})|^2$ and Ψ_t is complex valued (unlike the heat equation), so the generator is actually that of a unitary quantum evolution rather than a Markov semigroup for which our integrals are ill-defined.

The heuristic idea of Feynman's so-called **path integral** formulation is that the solution to the Schrödinger equation should be obtained by summing over all possible (stochastic) particle paths, with each path weighted by a complex phase depending on its classical action,

$$\Psi(t,x) \approx \int_{\text{paths ending at } x} e^{\frac{i}{\hbar}S[\gamma]}\,\mathcal{D}\gamma. \tag{10.3}$$

akin to the exponentially discounted expectation in Theorem 10.1.

Note this characterization is ill-defined, however, because there is no ordinary Lebesgue measure on the space of all paths. Indeed, a mathematically rigorous formulation of the path integral does not exist except for some simple cases, and Feynman-Kac can not be used to directly solve the Schrödinger equation in real time.

On the other hand, Feynman-Kac *can* come useful in understanding the ground state (or the energy levels more generally) of the wave function through analytic continuation, for which numerical methods prove to be exceedingly useful and preferrable to traditional grid-based PDE solvers such as finite differences for high-dimensional problems, where traditional methods tend to suffer from the curse of dimensionality. When paried with numerical SDE solvers, this approach to calculating low-lying energies of a quantum Hamiltonian is called **Diffusion Monte Carlo**.

The main idea is to consider the analytic continuation of the Schrödinger equation through the **Wick Rotation**:

$$t = -i\tau, \tag{10.4}$$

after which, the Schrödinger equation becomes a heat-type equation. In the time-independent case $H = -\frac{\hbar^2}{2m}\Delta + V(x)$, we get

$$\partial_\tau \psi_\tau = -\frac{1}{\hbar}H\psi_\tau = \frac{\hbar}{2m}\Delta\psi_\tau - \frac{1}{\hbar}V(x)\psi_\tau. \tag{10.5}$$

ie., the PDE associated with the KBE of a diffusion process:

$$\partial_\tau \psi = L\psi - c\psi, \qquad L = \frac{\hbar}{2m}\Delta, \qquad c(x) = \frac{1}{\hbar}V(x). \tag{10.6}$$

Note the corresponding Itô SDE of the process is given by:

$$dX_s = \sqrt{\frac{\hbar}{m}}\,dW_s. \tag{10.7}$$

Now, recall that if the transition kernel of X admits a density $p_t(x,y)$, then

$$Q_t(x,dy) = p_t(x,y)\,dy. \tag{10.8}$$

Thus, for any bounded measurable function f,

$$Q_t f(x) = \int_{\mathbb{R}^d} f(y)p_t(x,y)\,dy = \mathop{\mathbb{E}}_x[f(X_t)]. \tag{10.9}$$

Therefore $p_t(x,y)$ is the transition density of X_t given $X_0 = x$, while Q_t is the operator that averages a function against this transition density.

Now if there was no potential term, ie. $V = 0$, the imaginary-time Schrödinger equation reduces to the regular heat equation: equation

$$\partial_\tau \psi = \frac{\hbar}{2m}\Delta\psi = L\psi, \qquad \psi(0,x) = \psi_0(x). \tag{10.10}$$

Since $L = \frac{\hbar}{2m}\Delta$ is the generator of the diffusion in this case the Kolmogorov Backward Equation gives (see the form of the solution from the proof of the KBE for the diffusion case)

$$\psi(\tau,x) = Q_\tau\psi_0(x) = \mathbb{E}_x[\psi_0(X_\tau)]. \tag{10.11}$$

Thus, without a potential, the solution is obtained by averaging the initial wavefunction over all Brownian "paths" starting from x.

If the potential $V \neq 0$, we get an additional zero-order term $-\frac{1}{\hbar}V(x)\psi$ from Feynman-Kac:

$$\psi(\tau,x) = \mathbb{E}_x\left[\exp\left(-\frac{1}{\hbar}\int_0^\tau V(X_s)\,ds\right)\psi_0(X_\tau)\right]. \tag{10.12}$$

This representation is especially useful as a way of filtering out the ground state of the wave function . To see this, suppose for simplicity that H is time-independent, self-adjoint, and has a discrete spectrum

$$H\phi_n = E_n\phi_n, \qquad E_0 \le E_1 \le E_2 \le \cdots,$$

where (ϕ_n) is an orthonormal basis of eigenfunctions. The lowest-energy eigenfunction ϕ_0 is called the **ground state**, and E_0 is the **ground-state energy**. If the initial condition has expansion

$$\psi_0 = \sum_{n=0}^{\infty} a_n\phi_n,$$

then the solution of the imaginary-time Schrödinger equation is given by

$$\psi(\tau) = e^{-\tau H/\hbar}\psi_0 = \sum_{n=0}^{\infty} a_n e^{-\tau E_n/\hbar}\phi_n.$$

Thus the coefficient of ϕ_n is multiplied by $e^{-\tau E_n/\hbar}$. Higher energy modes decay faster than lower energy modes. Factoring out the ground-state contribution gives

$$\psi(\tau) = e^{-\tau E_0/\hbar}\left(a_0\phi_0 + \sum_{n=1}^{\infty} a_n e^{-\tau(E_n-E_0)/\hbar}\phi_n\right).$$

Therefore, provided $a_0 \neq 0$, the normalized imaginary-time solution satisfies formally

$$\frac{\psi(\tau)}{\|\psi(\tau)\|} \longrightarrow \frac{a_0}{|a_0|}\phi_0 \qquad \text{as } \tau \to \infty,$$

up to the irrelevant phase/sign of the eigenfunction. In other words, imaginary-time evolution acts as an energy filter: after repeatedly applying the heat semigroup $e^{-\tau H/\hbar}$ and renormalizing, the shape of the solution converges to the ground state.

Additionally, we can recover the ground-state energy from the exponential decay rate. Formally, if $a_0 \neq 0$, then

$$\|\psi(\tau)\| \sim |a_0| e^{-\tau E_0/\hbar} \qquad \text{as } \tau \to \infty,$$

so

$$E_0 = -\hbar \lim_{\tau\to\infty} \frac{1}{\tau} \log \|\psi(\tau)\|.$$

Equivalently, one can use the Rayleigh quotient

$$\mathcal{E}(\tau) = \frac{\langle \psi(\tau), H\psi(\tau)\rangle}{\langle \psi(\tau), \psi(\tau)\rangle},$$

which converges to E_0 under the same non-orthogonality assumption.

Part III. Advanced Topics

11. (L22) Langevin Dynamics and Long-Time Behavior

11.1. Langevin Dynamics

11.1.1. The Ornsetin-Uhlenbeck Process

These notes are adapted from [Cao25].

Let $(B_t)_{t\geq 0}$ be a d-dimensional Brownian motion adapted to a complete filtration $(\mathcal{F}_t)_{t\geq 0}$.

Definition 11.1

An **Ornstein–Uhlenbeck process** is a continuous adapted process $(X_t)_{t\geq 0}$ solving

$$dX_t = -\theta X_t \, dt + \sigma \, dt \tag{11.1}$$

Equivalently,

$$X_t = X_0 e^{-\theta t} - \theta \int_0^t X_s \, ds + \sigma \, B_t. \tag{11.2}$$

11.1.2. Explicit solution and transition law

Proposition 11.2

Let $X_0 \in \mathcal{F}_0$, and define

$$X_t = e^{-t} X_0 + \sqrt{2} \int_0^t e^{-(t-s)} \, dB_s. \tag{11.3}$$

Then $(X_t)_{t\geq 0}$ is an Ornstein–Uhlenbeck process.

Proof. Apply Itô's formula to

$$e^t X_t = X_0 + \sqrt{2} \int_0^t e^s \, dB_s.$$

Differentiating gives

$$d(e^t X_t) = \sqrt{2} e^t \, dB_t,$$

so after multiplying by e^{-t} we recover

$$dX_t = -X_t \, dt + \sqrt{2} \, dB_t.$$

□

Proposition 11.3

If $(X_t)_{t \geq 0}$ is an Ornstein–Uhlenbeck process, then necessarily

$$X_t = e^{-t} X_0 + \sqrt{2} \int_0^t e^{-(t-s)} \, dB_s. \tag{11.4}$$

In particular, the solution is pathwise unique.

Proof. From the SDE,

$$d(e^t X_t) = e^t X_t \, dt + e^t \, dX_t = \sqrt{2} e^t \, dB_t.$$

Integrating from 0 to t yields

$$e^t X_t = X_0 + \sqrt{2} \int_0^t e^s \, dB_s,$$

which is equivalent to the displayed formula. □

For deterministic initial data $X_0 = x \in \mathbb{R}^d$, write X_t^x for the corresponding solution.

Corollary 11.4

For every $t \geq 0$ and $x \in \mathbb{R}^d$,

$$X_t^x \sim \mathcal{N}\big(e^{-t} x, (1 - e^{-2t}) I_d\big). \tag{11.5}$$

Hence

$$X_t^x \xrightarrow[t \to \infty]{(d)} \mathcal{N}(0, I_d). \tag{11.6}$$

Proof. The stochastic integral is centered Gaussian with covariance

$$2 \int_0^t e^{-2(t-s)} \, ds \, I_d = (1 - e^{-2t}) I_d.$$

The convergence in distribution follows immediately. □

> **Remark.** The factor $\sqrt{2}$ is exactly the normalization that makes the invariant Gaussian be $\mathcal{N}(0, I_d)$. Without it, the limiting covariance would be different.

11.1.3. Semigroup, invariant measure, and generator

Proposition 11.5

The Ornstein–Uhlenbeck process is Markov. Its semigroup is

$$(P_t f)(x) = \mathbb{E}\big[f(X_t^x)\big] = \mathbb{E}\Big[f\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Z\big)\Big], \tag{11.7}$$

where $Z \sim \mathcal{N}(0, I_d)$. Equivalently,

$$(P_t f)(x) = \frac{1}{(2\pi(1-e^{-2t}))^{d/2}} \int_{\mathbb{R}^d} f(y) \exp\Big(-\frac{|y-e^{-t}x|^2}{2(1-e^{-2t})}\Big)\, dy. \tag{11.8}$$

Proof. Fix $0 \le s \le t$. The explicit solution on $[s,t]$ reads

$$X_t = e^{-(t-s)} X_s + \sqrt{2} \int_s^t e^{-(t-u)}\, dB_u.$$

The stochastic integral is independent of $\mathcal{F}_s$ and distributed as $\mathcal{N}(0, (1-e^{-2(t-s)})I_d)$. Therefore

$$\mathbb{E}[f(X_t) \mid \mathcal{F}_s] = (P_{t-s} f)(X_s),$$

which is the Markov property. □

Let γ_d denote the standard Gaussian measure on $\mathbb{R}^d$.

Proposition 11.6

The measure $\gamma\_d = \mathcal{N}(0, I_d)$ is invariant for $(P_t)_{t\ge 0}$, and it is the unique invariant probability measure.

Proof. If $Z \sim \gamma\_d$ and is independent of the Brownian motion, then

$$e^{-t}Z + \sqrt{1-e^{-2t}}\,Z' \sim \gamma\_d,$$

so $P_t^* \gamma\_d = \gamma\_d$. Uniqueness follows from the pointwise convergence

$$P_t f(x) \to \mathbb{E}[f(Z)]$$

for bounded continuous f: if μ is invariant, then

$$\int P_t f\, d\mu = \int f\, d\mu,$$

and sending $t \to \infty$ gives $\int f\, d\mu = \int f\, d\gamma\_d$. □

Proposition 11.7

For $f \in C_b^2(\mathbb{R}^d)$, the generator of (P_t) is

$$Lf(x) = \Delta f(x) - x \cdot \nabla f(x). \tag{11.9}$$

Moreover,

$$f(X_t) - f(X_0) - \int_0^t Lf(X_s)\, ds \tag{11.10}$$

is a martingale.

Proof. Apply Itô's formula:

$$df(X_t) = \nabla f(X_t) \cdot dX_t + \frac{1}{2} \sum_{i,j} \partial_{ij} f(X_t)\, d\langle X^i, X^j \rangle_t.$$

Since $dX_t = -X_t dt + \sqrt{2}\, dB_t$ and $d\langle X^i, X^j \rangle_t = 2\delta_{ij}\, dt$, this becomes

$$df(X_t) = \sqrt{2}\, \nabla f(X_t) \cdot dB_t + \big(\Delta f(X_t) - X_t \cdot \nabla f(X_t)\big) dt.$$

□

PROPOSITION 11.8

The generator L is symmetric and negative semidefinite on $L^2(\gamma\_d)$: for $f, g \in C_b^2(\mathbb{R}^d)$,

$$(Lf, g)_{L^2(\gamma\_d)} = (f, Lg)_{L^2(\gamma\_d)} = -\int_{\mathbb{R}^d} \nabla f \cdot \nabla g \, d\gamma\_d. \tag{11.11}$$

In particular,

$$(Lf, f)_{L^2(\gamma\_d)} = -\|\nabla f\|^2_{L^2(\gamma\_d)} \le 0. \tag{11.12}$$

Proof. Write $d\gamma\_d(x) = (2\pi)^{-d/2} e^{-|x|^2/2}\, dx$. Since

$$\nabla e^{-|x|^2/2} = -x e^{-|x|^2/2},$$

integration by parts gives the cancellation

$$\int (\Delta f - x \cdot \nabla f) g \, d\gamma\_d = -\int \nabla f \cdot \nabla g \, d\gamma\_d.$$

□

PROPOSITION 11.9 (*REVERSIBILITY*)

If $X_0 \sim \gamma\_d$, then for every $t \ge 0$,

$$(X_0, X_t) \overset{(d)}{=} (X_t, X_0). \tag{11.13}$$

Equivalently,

$$(P_t f, g)_{L^2(\gamma\_d)} = (f, P_t g)_{L^2(\gamma\_d)}. \tag{11.14}$$

Proof. From the explicit formula,

$$X_t = e^{-t} X_0 + \sqrt{1 - e^{-2t}}\, Z,$$

with $Z \sim \mathcal{N}(0, I_d)$ independent of X_0. Hence (X_0, X_t) is jointly Gaussian with mean zero and covariance matrix

$$\begin{pmatrix} I_d & e^{-t} I_d \\ e^{-t} I_d & I_d \end{pmatrix},$$

which is symmetric under swapping the two coordinates. □

11.1.4. Commutation, variance decay, and Poincaré

Lemma 11.10 (*Commutation identity*)
For every smooth f,

$$\nabla P_t f = e^{-t} P_t(\nabla f). \tag{11.15}$$

Consequently,

$$\|\nabla P_t f\|_{L^2(\gamma\_d)} \le e^{-t} \|\nabla f\|_{L^2(\gamma\_d)}. \tag{11.16}$$

Proof. Differentiate the explicit formula for $P_t f$ under the integral sign:

$$\partial_i P_t f(x) = e^{-t}\ \mathbb{E}\big[(\partial_i f)(e^{-t}x + \sqrt{1-e^{-2t}}Z)\big].$$

The L^2 estimate follows from Jensen and the invariance of γ_d. □

Proposition 11.11 (*Variance decay*)
For all $f \in L^2(\gamma\_d)$,

$$\mathrm{Var}_{\gamma\_d}(P_t f) \le e^{-2t} \mathrm{Var}_{\gamma\_d}(f). \tag{11.17}$$

Proof. Differentiate the squared L^2 norm of $P_t f - \mathbb{E}_{\gamma\_d}[f]$ and use symmetry of L:

$$\frac{d}{dt}\|P_t f - \mathop{\mathbb{E}}_{\gamma\_d}[f]\|^2_{L^2(\gamma\_d)} = 2(LP_t f, P_t f)_{L^2(\gamma\_d)} = -2\|\nabla P_t f\|^2_{L^2(\gamma\_d)}.$$

The commutation identity bounds the right-hand side by $-2e^{-2t}\|\nabla f\|^2_{L^2(\gamma\_d)}$, which yields exponential decay. □

Theorem 11.12 (*Gaussian Poincaré inequality*)
For every $f \in H^1(\gamma\_d)$,

$$\mathrm{Var}_{\gamma\_d}(f) \le \int_{\mathbb{R}^d} |\nabla f|^2\, d\gamma\_d. \tag{11.18}$$

Proof. Assume first that $\mathbb{E}_{\gamma\_d}[f] = 0$. Since $P_t f \to 0$ in $L^2(\gamma\_d)$,

$$\mathrm{Var}_{\gamma\_d}(f) = \|f\|^2_{L^2(\gamma\_d)} = -\int_0^\infty \frac{d}{dt}\|P_t f\|^2_{L^2(\gamma\_d)}\, dt = 2\int_0^\infty \|\nabla P_t f\|^2_{L^2(\gamma\_d)}\, dt.$$

Using $\|\nabla P_t f\|_{L^2(\gamma\_d)} \le e^{-t}\|\nabla f\|_{L^2(\gamma\_d)}$ gives

$$\mathrm{Var}_{\gamma\_d}(f) \le 2\int_0^\infty e^{-2t}\, dt\, \|\nabla f\|^2_{L^2(\gamma\_d)} = \|\nabla f\|^2_{L^2(\gamma\_d)}.$$

□

11.1.5. Entropy, logarithmic Sobolev, and hypercontractivity

Definition 11.13

For a nonnegative f with $\mathbb{E}_{\gamma\_d}[f] < \infty$, define the relative entropy by

$$\mathrm{KL}_{\gamma\_d}(f) = \mathop{\mathbb{E}}_{\gamma\_d}[f \log f] - \mathop{\mathbb{E}}_{\gamma\_d}[f] \log \mathop{\mathbb{E}}_{\gamma\_d}[f]. \tag{11.19}$$

Proposition 11.14 (*Entropy dissipation*)

Let $f \geq 0$ and set $f_t = P_t f$. Then

$$\frac{d}{dt}\mathrm{KL}_{\gamma\_d}(f_t) = -\mathop{\mathbb{E}}_{\gamma\_d}\left[\frac{|\nabla f_t|^2}{f_t}\right]. \tag{11.20}$$

Proof. Using $\partial_t f_t = L f_t$ and the chain rule for $u \mapsto u \log u$,

$$\frac{d}{dt}\mathrm{KL}_{\gamma\_d}(f_t) = (L f_t, \log f_t + 1)_{L^2(\gamma\_d)} = -\int \nabla f_t \cdot \nabla(\log f_t)\, d\gamma\_d = -\int \frac{|\nabla f_t|^2}{f_t}\, d\gamma\_d.$$

□

Theorem 11.15 (*Gaussian logarithmic Sobolev inequality*)

For every smooth $f \geq 0$,

$$\mathrm{KL}_{\gamma\_d}(f) \leq \frac{1}{2}\mathop{\mathbb{E}}_{\gamma\_d}\left[\frac{|\nabla f|^2}{f}\right]. \tag{11.21}$$

Proof. Since $P_t f \to \mathbb{E}_{\gamma\_d}[f]$ as $t \to \infty$, we have

$$\mathrm{KL}_{\gamma\_d}(f) = -\int_0^\infty \frac{d}{dt}\mathrm{KL}_{\gamma\_d}(P_t f)\, dt = \int_0^\infty \mathop{\mathbb{E}}_{\gamma\_d}\left[\frac{|\nabla P_t f|^2}{P_t f}\right] dt.$$

By the commutation identity and Cauchy–Schwarz,

$$|\nabla P_t f|^2 = e^{-2t}|P_t(\nabla f)|^2 \leq e^{-2t} P_t\left(\frac{|\nabla f|^2}{f}\right) P_t f.$$

Therefore

$$\mathop{\mathbb{E}}_{\gamma\_d}\left[\frac{|\nabla P_t f|^2}{P_t f}\right] \leq e^{-2t}\mathop{\mathbb{E}}_{\gamma\_d}\left[\frac{|\nabla f|^2}{f}\right].$$

Integrating in t gives the result. □

Corollary 11.16 (*Entropy decay*)

For $f \geq 0$,

$$\mathrm{KL}_{\gamma\_d}(P_t f) \leq e^{-2t}\mathrm{KL}_{\gamma\_d}(f). \tag{11.22}$$

Proof. Combine the dissipation identity with the logarithmic Sobolev inequality applied to $P_t f$:

$$\frac{d}{dt}\mathrm{KL}_{\gamma\_d}(P_t f) = -\mathop{\mathbb{E}}_{\gamma\_d}\left[\frac{|\nabla P_t f|^2}{P_t f}\right] \leq -2\mathrm{KL}_{\gamma\_d}(P_t f).$$

Then use Gronwall. ☐

THEOREM 11.17 (*HYPERCONTRACTIVITY*)

Let $1 < p < \infty$ and $t \geq 0$. Define

$$q_t = 1 + e^{2t}(p-1). \tag{11.23}$$

Then for every measurable f,

$$\|P_t f\|_{L^{q_t}(\gamma\_d)} \leq \|f\|_{L^p(\gamma\_d)}. \tag{11.24}$$

Proof sketch. For $f \geq 0$, consider

$$h(t) = \|P_t f\|_{L^{q_t}(\gamma\_d)}.$$

Differentiating $\log h(t)$ produces two terms: one coming from the generator and one from q_t'. The generator term is controlled by entropy dissipation, and the q_t' term is chosen precisely so that the logarithmic Sobolev inequality closes the estimate. The choice $q_t' = 2(q_t - 1)$ is equivalent to $q_t = 1 + e^{2t}(p-1)$, and yields $h'(t) \leq 0$. ☐

11.1.6. The Fokker-Planck equation

Suppose that the law of X_t has density f_t with respect to γ_d. Since the evolution is given by the semigroup,

$$f_t = P_t f_0, \tag{11.25}$$

so formally

$$\partial_t f_t = L f_t = \Delta f_t - x \cdot \nabla f_t. \tag{11.26}$$

If instead g_t denotes the density with respect to Lebesgue measure, then

$$g_t(x) = f_t(x)(2\pi)^{-d/2} e^{-|x|^2/2}, \tag{11.27}$$

which leads after a short computation to the usual Fokker–Planck equation

$$\partial_t g_t = \Delta g_t + \mathrm{div}(x g_t). \tag{11.28}$$

11.2. General Langevin Processes

The Ornstein–Uhlenbeck process is a special case of the much broader class of Langevin SDEs.

DEFINITION 11.18

Let $V : \mathbb{R}^d \to \mathbb{R}$ be a potential. The associated *Langevin dynamics* is

$$dX_t = -\nabla V(X_t)\,dt + \sqrt{2}\,dB_t. \tag{11.29}$$

Whenever

$$Z_V := \int_{\mathbb{R}^d} e^{-V(x)}\,dx < \infty, \tag{11.30}$$

we define the Gibbs measure

$$d\mu\_V(x) = Z_V^{-1} e^{-V(x)}\, dx. \tag{11.31}$$

Example 11.19

If $V(x) = \frac{1}{2}|x|^2$, then $\mu\_V = \gamma\_d$ and the Langevin SDE becomes the Ornstein–Uhlenbeck equation.

11.2.1. Existence, uniqueness, and non-explosion

We first treat the general SDE

$$dX_t = b(X_t)\, dt + \sqrt{2}\, dB_t, \qquad X_0 = Y. \tag{11.32}$$

Proposition 11.20

If $b : \mathbb{R}^d \to \mathbb{R}^d$ is globally Lipschitz and $Y \in \mathcal{F}_0$ with $|Y| < \infty$ a.s., then there exists a unique global adapted solution.

Proof sketch via contraction mapping. Fix $T > 0$ and define $G : C([0,T];\mathbb{R}^d) \to C([0,T];\mathbb{R}^d)$ by

$$(Gf)(t) = Y + \int_0^t b(f(s))\, ds + \sqrt{2}\, B_t. \tag{11.33}$$

With the supremum norm,

$$\|Gf - Gg\|_\infty \le KT \|f - g\|_\infty,$$

where K is the Lipschitz constant of b. Thus G is a strict contraction when $T < (2K)^{-1}$. Banach's fixed-point theorem gives a unique local solution, and since the existence time depends only on K, the solution can be iterated to all of $[0, \infty)$. ☐

Theorem 11.21 (*Local theory for locally Lipschitz drift*)

Assume $(\mathcal{F}_t)$ is right-continuous. If b is locally Lipschitz, then for every $Y \in \mathcal{F}_0$ there exists a unique continuous adapted solution on an interval $[0, T)$, where T is a stopping time. Moreover, on the event $\{T < \infty\}$ one has

$$\lim_{t \uparrow T} |X_t| = \infty. \tag{11.34}$$

Proof sketch. Localize the drift on balls $\{|x| \le R\}$ and solve the truncated equation there by the previous proposition. If T_R is the exit time from the ball of radius R, uniqueness shows that the solutions for different R agree on overlaps. Letting $T = \lim_{R\to\infty} T_R$ produces the maximal solution. If the norm stayed bounded as $t \uparrow T$, one could restart the equation beyond T, contradicting maximality. ☐

For Langevin dynamics one wants sufficient conditions ensuring that the maximal solution is actually global.

Theorem 11.22 (*Global existence criteria for Langevin dynamics*)
Assume ∇V is locally Lipschitz. Suppose at least one of the following holds.

(A) $V(x) \to \infty$ as $|x| \to \infty$, and $|\nabla V|^2 - \Delta V$ is bounded below.

(B) There exist $a, b \in \mathbb{R}$ such that

$$x \cdot \nabla V(x) \geq -a|x|^2 - b \qquad \forall x \in \mathbb{R}^d. \tag{11.35}$$

Then for every deterministic $x_0 \in \mathbb{R}^d$, the Langevin SDE with $X_0 = x_0$ admits a unique global solution.

Proof sketch. Under (A), apply Itô's formula to $V(X_t)$:

$$dV(X_t) = \sqrt{2}\,\nabla V(X_t) \cdot dB_t + \big(-|\nabla V(X_t)|^2 + \Delta V(X_t)\big)dt.$$

The lower bound on $|\nabla V|^2 - \Delta V$ gives a Lyapunov estimate which prevents escape to infinity before any fixed time.
Under (B), apply Itô's formula to $|X_t|^2$:

$$\frac{1}{2}d|X_t|^2 = \sqrt{2}X_t \cdot dB_t + \big(-X_t \cdot \nabla V(X_t) + d\big)dt.$$

The assumption yields an inequality of the form

$$\mathbb{E}\,|X_t|^2 \leq |x_0|^2 + C\int_0^t \mathbb{E}\,|X_s|^2\,ds + Ct,$$

so Gronwall's lemma gives a finite second moment on every finite time interval. This rules out explosion. □

11.2.2. Semigroup, generator, and invariant measure

Assume from now on that the Langevin dynamics is global for every starting point. For each $x \in \mathbb{R}^d$, let X_t^x denote the solution started at x, and define

$$(P_t f)(x) = \mathbb{E}[f(X_t^x)]. \tag{11.36}$$

Then $(P_t)_{t\geq 0}$ is a Markov semigroup.

Proposition 11.23
For $f \in C_b^2(\mathbb{R}^d)$, the generator is

$$L\_V f = \Delta f - \nabla V \cdot \nabla f. \tag{11.37}$$

Moreover,

$$f(X_t) - f(X_0) - \int_0^t L\_V f(X_s)\,ds \tag{11.38}$$

is a martingale.

Proof. Apply Itô's formula exactly as in the Ornstein–Uhlenbeck case, with drift $-\nabla V$ in place of $-x$. $\square$

Proposition 11.24

Assume $Z_V < \infty$. Then for $f, g \in C_b^2(\mathbb{R}^d)$,

$$(L\_V f, g)_{L^2(\mu\_V)} = (f, L\_V g)_{L^2(\mu\_V)} = -\int_{\mathbb{R}^d} \nabla f \cdot \nabla g \, d\mu\_V. \tag{11.39}$$

In particular, μ_V is invariant for (P_t).

Proof. Since $d\mu\_V = Z_V^{-1} e^{-V(x)} dx$ and $\nabla e^{-V} = -e^{-V} \nabla V$, integration by parts gives

$$\int (\Delta f - \nabla V \cdot \nabla f) g \, d\mu\_V = -\int \nabla f \cdot \nabla g \, d\mu\_V.$$

Setting $g \equiv 1$ shows $(L\_V f, 1)_{L^2(\mu\_V)} = 0$, and then

$$\frac{d}{dt} \int P_t f \, d\mu\_V = \int L\_V P_t f \, d\mu\_V = 0,$$

so $\int P_t f \, d\mu\_V = \int f \, d\mu\_V$. $\square$

11.2.3. Reversibility and Girsanov

Proposition 11.25 (*Reversibility*)

Assume $X_0 \sim \mu\_V$. Then for every $t \geq 0$,

$$(X_0, X_t) \overset{(d)}{=} (X_t, X_0). \tag{11.40}$$

Equivalently,

$$(P_t f, g)_{L^2(\mu\_V)} = (f, P_t g)_{L^2(\mu\_V)}. \tag{11.41}$$

Idea of proof. One compares the path law of X^x on $C([0,T]; \mathbb{R}^d)$ with the law of Brownian motion started at x. Girsanov's theorem gives an explicit density involving the drift $-\nabla V$. After rewriting the stochastic integral term with Itô's formula, the density becomes symmetric under time reversal once it is multiplied by the Gibbs weight $e^{-V(x)}$. Integrating over the stationary starting point $X_0 \sim \mu\_V$ yields reversibility. $\square$

11.2.4. Uniform convexity, contraction, and functional inequalities

Assume now that $V \in C^2(\mathbb{R}^d)$ and that there exists $\gamma > 0$ such that

$$\nabla^2 V(x) \succeq \gamma I_d \qquad \forall x \in \mathbb{R}^d. \tag{11.42}$$

Lemma 11.26 (*Strong monotonicity*)
For all $x, y \in \mathbb{R}^d$,
$$\big(\nabla V(x) - \nabla V(y)\big) \cdot (x - y) \geq \gamma |x - y|^2. \tag{11.43}$$

Proof. Let $\ell(t) = (1-t)x + ty$. Then

$$\nabla V(y) - \nabla V(x) = \int_0^1 \nabla^2 V(\ell(t))(y - x)\, dt.$$

Taking the dot product with $y - x$ and using $\nabla^2 V \succeq \gamma I_d$ gives the claim. □

Proposition 11.27 (*Synchronous coupling contraction*)
Let X_t^x and X_t^y solve the Langevin SDE with the same Brownian motion but different initial data x, y. Then
$$|X_t^x - X_t^y|^2 \leq e^{-2\gamma t}|x - y|^2 \qquad \text{for all } t \geq 0. \tag{11.44}$$

Proof. The Brownian terms cancel in the difference, so

$$d(X_t^x - X_t^y) = -\big(\nabla V(X_t^x) - \nabla V(X_t^y)\big)dt.$$

Therefore

$$\frac{1}{2} d|X_t^x - X_t^y|^2 = -\big(\nabla V(X_t^x) - \nabla V(X_t^y)\big) \cdot (X_t^x - X_t^y)\, dt \leq -\gamma |X_t^x - X_t^y|^2 dt.$$

Apply Gronwall. □

Proposition 11.28 (*Sub-commutation*)
Under the same uniform convexity assumption,

$$|\nabla P_t f| \leq e^{-\gamma t} P_t(|\nabla f|), \tag{11.45}$$

and consequently
$$|\nabla P_t f|^2 \leq e^{-2\gamma t} P_t(|\nabla f|^2). \tag{11.46}$$

Proof. Write

$$P_t f(x) - P_t f(y) = \mathbb{E}\big[f(X_t^x) - f(X_t^y)\big] = \mathbb{E}\Big[\int_0^1 \nabla f\big(uX_t^x + (1-u)X_t^y\big) \cdot (X_t^x - X_t^y)\, du\Big].$$

Using the contraction estimate and dividing by $|x - y|$, then sending $y \to x$, gives the first bound. The second follows from Cauchy–Schwarz and Jensen. □

Theorem 11.29 (*Strict convexity implies Poincaré and logarithmic Sobolev*)
If $\nabla^2 V \succeq \gamma I_d$, then μ_V satisfies

$$\mathrm{Var}_{\mu\_V}(f) \le \frac{1}{\gamma} \int |\nabla f|^2 \, d\mu\_V \tag{11.47}$$

for all $f \in H^1(\mu\_V)$, and

$$\mathrm{KL}_{\mu\_V}(f) \le \frac{1}{2\gamma} \int \frac{|\nabla f|^2}{f} \, d\mu\_V \tag{11.48}$$

for all smooth $f \ge 0$.

Proof. The argument is the same semigroup interpolation used for the Gaussian case. For variance, write

$$\mathrm{Var}_{\mu\_V}(f) = 2 \int_0^\infty \|\nabla P_t f\|^2_{L^2(\mu\_V)} \, dt,$$

and then apply the sub-commutation bound:

$$\|\nabla P_t f\|^2_{L^2(\mu\_V)} \le e^{-2\gamma t} \|\nabla f\|^2_{L^2(\mu\_V)}.$$

Integrating $2e^{-2\gamma t}$ over $[0, \infty)$ yields γ^{-1}.
For entropy,

$$\mathrm{KL}_{\mu\_V}(f) = \int_0^\infty \mathop{\mathbb{E}}_{\mu\_V} \left[\frac{|\nabla P_t f|^2}{P_t f} \right] dt,$$

and Cauchy–Schwarz together with sub-commutation gives

$$\mathop{\mathbb{E}}_{\mu\_V} \left[\frac{|\nabla P_t f|^2}{P_t f} \right] \le e^{-2\gamma t} \mathop{\mathbb{E}}_{\mu\_V} \left[\frac{|\nabla f|^2}{f} \right].$$

Integrating once more yields the factor $(2\gamma)^{-1}$. □

Corollary 11.30 (*Exponential convergence*)
Under the uniform convexity assumption,

$$\mathrm{Var}_{\mu\_V}(P_t f) \le e^{-2\gamma t} \mathrm{Var}_{\mu\_V}(f) \tag{11.49}$$

and for $f \ge 0$,

$$\mathrm{KL}_{\mu\_V}(P_t f) \le e^{-2\gamma t} \mathrm{KL}_{\mu\_V}(f). \tag{11.50}$$

Thus the Langevin dynamics converges exponentially fast to equilibrium in both L^2 and relative entropy.

> **Remark.** At the level of intuition, the convexity of V forces trajectories driven by the same noise to move toward one another. This geometric contraction is exactly what feeds the gradient estimates, which in turn give the functional inequalities and the exponential convergence to μ_V.

11.3. Poincaré, Log-Sobolev, and Hypercontractivity

TODO.

11.4. Functional Inequalities and Convergence to Equilibrium

TODO.

11.5. Interacting Particle Systems

TODO.

11.6. Dyson Brownian Motion

TODO.

11.7. Gaussian β-Ensembles

TODO.

12. (L23) Fisk-Stratanovich Calculus

13. (L24) Local Times

TODO.

14. (L24) Brownian Geometry and Potential Theory

TODO.

14.1. Brownian Local Times

14.2. Fractal Brownian Motion and Hausdorff Measure

14.3. Brownian Potential Theory

14.4. Schramm–Loewner Evolution

15. (L25) Modern Applications

15.1. Stochastic Localization

15.2. Concentration and Convex Geometry

15.3. Connections to Sampling and High-Dimensional Probability

16. (L27) Mallavian Calculus

Part IV. Selected Applications

17. (L28) Numerical Methods for SDEs

Of particular practical importance is the use of numerical methods to simulate and/or evaluate Stochastic Differential Equations (SDE). These become of practical importance in both modeling many physical phenomena (which lend naturally to an SDE formulation) but also in creating better high-dimensional numerical approximations to high-dimensional PDE problems due to the connections arising from the Feynman-Kac Formula (and conversely Fokker-Plank/KFE in the reverse direction) which we study in the next two sections (covering Stochastic Filtering and Quantum Monte Carlo Methods specifically).

We begin this section with a brief review of numerical methods for both ODEs and PDEs, reviewing key concepts of consistency, convergence, and stability analysis. We then introduce the Wagner-Platen Expansion, the stochastic analog of the Taylor Expansion which we use as a basis to analyze numerical methods for SDEs.

We then split our analysis of SDEs into two main types of approximations (in a similar vein to the corresponding analyses for ODEs and PDEs):

(a) **Strong Approximations**: where we are focused mainly with the accuracy of the corresponding sample paths $(X_t)_{t\geq 0}$, usually in terms of moments. For example, if we consider L^p random variables and the corresponding L^p norm, and we have some terminal time T and step size h, we have that if our simulated paths satisfy

$$\mathbb{E}[|X_T - \hat{X}_T|^p]^{1/p} \leq Ch^\gamma \tag{17.1}$$

that our solver is of strong order γ.

(b) **Weak approximations**: where we are focused mainly with the expected errors of (sufficiently regular) test functions (ie. statistics) φ. Roughly, if for all regular test functions φ, we have

$$|\mathbb{E}[\varphi(X_T)] - \mathbb{E}[\varphi(\hat{X}_T)]| \leq Ch^\beta \tag{17.2}$$

then our solver is of weak order β.

17.1. Numerical Methods for ODEs

We begin with a review of numerical methods for ODEs. Broadly, we split the space of ODE problems into two types:

Definition 17.1 (*Initial and Boundary Value Problems*)
For an nth-order ODE of the form

$$F(x, y, y', \dots, y^{(n)}) = 0,$$

an **auxiliary condition** is an additional condition imposed on the unknown solution y. We distinguish between:

(a) An **initial value problem** (IVP), where all n conditions are specified at a single point x_0. Typically, this has the form

$$y(x_0) = c_0, \qquad y'(x_0) = c_1, \qquad \dots, \qquad y^{(n-1)}(x_0) = c_{n-1}.$$

Thus an IVP prescribes the initial position and the first $n-1$ derivatives of the solution at one point.

(b) A **boundary value problem** (BVP), where the n conditions are specified at two or more points, usually at the endpoints of an interval $[a, b]$. For example, a second-order BVP may have the form

$$F(x, y, y', y'') = 0, \qquad y(a) = \alpha, \qquad y(b) = \beta.$$

More generally, a BVP prescribes conditions involving y and its derivatives at different points of the domain.

Note, because in the numerical regime, we can often define state variables of the form $\mathbf{y} = [x, y, \cdots, y^{(n-1)}]^T$, we largely consider ODEs of first-order, ie. of the form

$$y' = f(t, y) \tag{17.3}$$

for which our conclusions will largely generalize.

Although other methods based on the Taylor Series or multiderivatives also exist, the most widely applicable numerical method for ODE involves **time discretization** where we make the approximation $dt \approx \Delta t$ for some small constant. Methods that make these time discretizations are referred to as **difference methods**.

The typical lens at which one analyzes time discrete methods are then threefold: consistency, convergence and stability. The former two can often be defined rigorously in terms of an “order” that can be analyzed naturally through the lens of the Taylor series while the last condition is often analyzed through some linear algebra.

In particular, we will make a distinction between *implicit* and *explicit* numerical methods as they differ sharply in their consistency, convergence, and stability properties. We define them rigorously as follows:

Definition 17.2 (*Explicit and Implicit Time-Stepping Methods*)
Consider an IVP written in first-order form

$$y'(t) = f(t, y(t)), \qquad y(t_0) = y_0,$$

and let $t_k = t_0 + k\Delta t$. A difference method that constructs approximations $y_k \approx y(t_k)$ is called

(a) **explicit** if y_{k+1} can be computed directly from previously known information, such as t_k, y_k, and earlier values. In a one-step method, this means

$$y_{k+1} = \Phi_{\Delta t}(t_k, y_k),$$

where the right-hand side depends only on known quantities.

(b) **implicit** if y_{k+1} is defined as the solution of an algebraic equation involving y_{k+1} itself. In a one-step method, this means

$$G_{\Delta t}(t_k, y_k, y_{k+1}) = 0,$$

so that one generally has to solve an algebraic equation at each time step.

The simplest examples of explicit and implicit difference methods are the Forward and Backward Euler methods.

Algorithmn 17.3 (*Euler's Method*)
Consider the IVP

$$y'(t) = f(t, y(t)), \qquad y(t_0) = y_0,$$

and fix a step size $\Delta t > 0$. Let $t_k := t_0 + k\Delta t$.

The **Forward Euler Method** is the explicit difference method

$$y_{k+1} = y_k + \Delta t\, f(t_k, y_k).$$

Equivalently, it approximates the derivative by the forward difference

$$y'(t_k) \approx \frac{y_{k+1} - y_k}{\Delta t}.$$

It has local error of order 2 and global error of order 1.

The **Backward Euler Method** is the implicit difference method

$$y_{k+1} = y_k + \Delta t\, f(t_{k+1}, y_{k+1}).$$

Equivalently, it approximates the derivative by the backward difference

$$y'(t_{k+1}) \approx \frac{y_{k+1} - y_k}{\Delta t}.$$

It is also first-order globally, but typically has better stability properties than Forward Euler.

17.1.1. Consistency and Convergence

Given a difference method, how do we analyze the convergence properties? Specifically, for a time discretization

$$t_0 < t_1 < \cdots < t_N = T, \qquad \Delta t_k := t_{k+1} - t_k, \qquad \Delta := \max_{0 \le k \le N-1} \Delta t_k,$$

a difference method produces approximations $y_k \approx y(t_k)$. We write

$$y(t; s, z)$$

for the exact solution at time t of the ODE starting from z at time s.

Definition 17.4 (*Local and Global Discretization Error*)
Consider an IVP

$$y'(t) = f(t, y(t)), \qquad y(t_0) = y_0,$$

and a time grid

$$t_0 < t_1 < \cdots < t_N.$$

A general difference method produces values $y_k \approx y(t_k)$ by imposing a discrete relation

$$\mathcal{F}_k(t_0, \ldots, t_{k+1}; y_0, \ldots, y_{k+1}) = 0,$$

In the explicit case,

$$y_{k+1} = \Phi_k(t_0, \ldots, t_{k+1}; y_0, \ldots, y_k).$$

The **local discretization error** is the one-step defect obtained by inserting the exact solution into the numerical scheme. In the explicit case, this is

$$\ell_{k+1} := y(t_{k+1}) - \Phi_k(t_0, \ldots, t_{k+1}; y(t_0), \ldots, y(t_k)).$$

The **global discretization error** at time t_k is

$$e_k := y(t_k) - y_k.$$

Intuitively, the local error measures how well the exact solution satisfies the discrete update rule over a single time step. On the other hand, the global error intuitively measures the accumulated difference between the exact solution of the original IVP and the numerical solution generated by the difference method.

As an example, Forward Euler gives

$$y_{k+1} = y_k + \Delta t_k f(t_k, y_k).$$

By Taylor expanding the exact solution starting from y_k at time t_k,

$$y(t_{k+1}; t_k, y_k) = y_k + \Delta t_k f(t_k, y_k) + \frac{1}{2} y''(\theta_k)(\Delta t_k)^2$$

for some $\theta_k \in (t_k, t_{k+1})$. Hence

$$\ell_{k+1} = \frac{1}{2} y''(\theta_k)(\Delta t_k)^2.$$

If $|y''(t)| \le M$ on the relevant interval, then

$$|\ell_{k+1}| \le \frac{M}{2}(\Delta t_k)^2.$$

So Forward Euler has local discretization error on the order of $(\Delta t)^2$.

Now, for a typical explicit one-step method we will have roughly that $y_{k+1} = y_k + \Psi(t_k, y_k, \Delta_k)\Delta_k$ for some explicit increment function $\Psi = \Psi(t, x, \Delta)$. One can imagine that a prerequisite for a solver to converge to zero error is that in the limit of an infinitesimally small discretization error, that the solver is *consistent* with the original ODE. We define this rigorously as follows:

Definition 17.5 (*Consistency of One-Step Difference Methods*)
A one-step method

$$y_{k+1} = y_k + \Psi(t_k, y_k, \Delta t_k)\Delta t_k$$

is called **consistent** with the ODE $y' = f(t, y)$ if

$$\lim_{\Delta t \to 0} \Psi(t, y, \Delta t) = f(t, y).$$

For general ODE solvers, the definition is given more abstractly as follows:

Definition 17.6 (*Consistency for a General Difference Method*)
Consider the IVP

$$y'(t) = f(t, y(t)), \qquad y(t_0) = y_0,$$

on a time grid

$$t_0 < t_1 < \cdots < t_N, \qquad h_k := t_{k+1} - t_k, \qquad h := \max_k h_k.$$

For a general difference method

$$\mathcal{F}_{h,k}(t_0, \ldots, t_N; y_0, \ldots, y_N) = 0,$$

the **local truncation defect** of the method is obtained by inserting the exact solution into the discrete relation:

$$\rho_{h,k} := \mathcal{F}_{h,k}(t_0, \ldots, t_N; y(t_0), \ldots, y(t_N)).$$

After the natural normalization by the step size, the **local truncation error** is

$$\tau_{h,k} := \frac{\rho_{h,k}}{h_k}.$$

The method is called **consistent** with the ODE if, for every sufficiently smooth exact solution y (typically we require Lipschitzness),

$$\max_k \|\tau_{h,k}\| \longrightarrow 0 \qquad \text{as } h \to 0.$$

With consistency, we will have in the limit of $\Delta t \to 0$ that the numerical model will converge to exact solutions to the ODE. We define convergence rigorously in the following manner:

Definition 17.7 (*Convergence*)
A one-step method is said to be **convergent** if the global discretization error converges to 0 in the infinitesimal limit of the time discretization:

$$\lim_{\Delta \downarrow 0} |e_{n+1}| = 0 \tag{17.4}$$

Under suitable regularity conditions for Ψ, one can show that convergence of a one-step solver is equivalent to consistency. This is formalized in the following theorem:

Theorem 17.8 (*Convergence $\Leftrightarrow$ Consistency*)
For a one-step method with increment function Ψ that satisfies the global Lipschitz condition:

$$|\Psi(t', x', \Delta') - \Psi(t, x, \Delta)| \leq K(|t' - t| + |x' - x| + |\Delta' - \Delta|)$$

and a global bound of the form $\lim_{\Delta t \downarrow 0} |\Psi(t, x, \Delta t)| \leq L$, we have the corresponding solver is convergent iff the Ψ is consistent.

Proof. See Theorem 8.3.1 of [KP92]. □

Of more value, however, is the *rate* of convergence, ie. how the error shrinks as a function of the time discretization Δ. We define these rigorously as follows:

Definition 17.9 (*Local and Global Order*)
More generally, the method has **local order** $p + 1$ if

$$|\ell_{k+1}| \leq C(\Delta t_k)^{p+1}$$

uniformly over the time interval of interest. It has **global order** p if

$$\max_{0 \leq k \leq N} |e_k| \leq C\Delta^p.$$

while for the general case is given as follows

Definition 17.10 (*Consistency Order*)
More generally, a general difference order method has **local order** p if the unnormalized defect

$$\max_k \|\rho_{h,k}\| = O(h^{p+1}).$$

while it has **global order** p (more commonly referred to just as having **consistency order** p) if

$$\max_k \|\tau_{h,k}\| = O(h^p).$$

In general, we have the global and local notions of order are largely consistent. For a one-step solver, we can formalize this argument under similar regularity assumptions to Theorem 17.8 as follows:

Theorem 17.11 (*Local Order $p+1$ implies Global Order p*)
A one-step method with increment function Ψ satisfying the conditions of 17.8, we have that if the method has local order $p+1$ that it has global order p.

Proof. See Theorem 8.3.2 of [KP92]. □

Thus Forward Euler has local error of order 2 but global error of order 1.

17.1.2. Examples of Difference Methods

A more accurate implicit method is obtained by averaging the vector field at the two endpoints of the interval.

Algorithmn 17.12 (*Trapezoidal Method*)
The **trapezoidal method** is

$$y_{k+1} = y_k + \frac{\Delta t}{2}\big(f(t_k, y_k) + f(t_{k+1}, y_{k+1})\big).$$

This is implicit because y_{k+1} appears on both sides. Under suitable smoothness assumptions, it has local error of order 3 and global error of order 2.

One can make the trapezoidal method explicit by first predicting y_{k+1} using Forward Euler and then correcting using the trapezoidal rule.

Algorithmn 17.13 (*Heun's Method / Improved Euler Method*)
First compute the Euler predictor

$$\overline{y}_{k+1} = y_k + \Delta t\, f(t_k, y_k).$$

Then define the corrected value

$$y_{k+1} = y_k + \frac{\Delta t}{2}\big(f(t_k, y_k) + f(t_{k+1}, \overline{y}_{k+1})\big).$$

Equivalently,

$$y_{k+1} = y_k + \frac{\Delta t}{2}\left[f(t_k, y_k) + f\big(t_{k+1}, y_k + \Delta t f(t_k, y_k)\big)\right].$$

This is an explicit predictor-corrector method. It has local error of order 3 and global error of order 2.

More generally, an explicit one-step method can be written as

$$y_{k+1} = y_k + \Psi(t_k, y_k, \Delta t)\Delta t,$$

where Ψ is called the **increment function**. For consistency, one requires

$$\lim_{\Delta t \to 0} \Psi(t, y, \Delta t) = f(t, y).$$

Different choices of Ψ give different schemes.

Algorithmn 17.14 (*Classical Fourth-Order Runge–Kutta Method*)

The **classical fourth-order Runge–Kutta method** is the explicit one-step method

$$y_{k+1} = y_k + \frac{\Delta t}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right),$$

where

$$\begin{aligned}
k_1 &= f(t_k, y_k),\\
k_2 &= f\left(t_k + \frac{\Delta t}{2}, y_k + \frac{\Delta t}{2}k_1\right),\\
k_3 &= f\left(t_k + \frac{\Delta t}{2}, y_k + \frac{\Delta t}{2}k_2\right),\\
k_4 &= f(t_k + \Delta t, y_k + \Delta t k_3).
\end{aligned}$$

This method has local error of order 5 and global error of order 4.

Some methods improve accuracy by using values from previous time steps. These are called **multi-step methods**. For example, with equal step size Δt, the three-step Adams–Bashforth method is

$$y_{k+1} = y_k + \frac{\Delta t}{12}\left[23f(t_k, y_k) - 16f(t_{k-1}, y_{k-1}) + 5f(t_{k-2}, y_{k-2})\right].$$

This is explicit, because it only uses previously computed values. Since it requires y_k, y_{k-1}, y_{k-2}, one must first generate starting values using a one-step method such as Heun's method or Runge–Kutta.
In contrast, implicit multi-step methods may involve $f(t_{k+1}, y_{k+1})$. For example, the Adams–Moulton method

$$y_{k+1} = y_k + \frac{\Delta t}{12}\left[5f(t_{k+1}, y_{k+1}) + 8f(t_k, y_k) - f(t_{k-1}, y_{k-1})\right]$$

is implicit and must generally be solved by iteration or by a root-finding method such as Newton's method.

17.1.3. Stability

Consistency and convergence describe what happens as the mesh size tends to zero. In practice, however, a method can still behave poorly for a fixed step size. Small errors from initial data, roundoff, or previous time steps may be amplified by the numerical scheme. This motivates the notion of **stability**.

Definition 17.15 (*Numerical Stability*)

Consider a one-step method

$$y_{k+1} = y_k + \Delta t_k \Psi(t_k, y_k, \Delta t_k)$$

on an interval $[t_0, T]$. We say the method is **numerically stable** on $[t_0, T]$ if there exist constants $\Delta_0 > 0$ and $M > 0$ such that, for every discretization with $\max_k \Delta t_k \leq \Delta_0$, any two numerical solutions (y_k) and $(\widetilde{y}_k)$ satisfy

$$|y_k - \widetilde{y}_k| \leq M\,|y_0 - \widetilde{y}_0|, \qquad t_k \leq T.$$

Thus small perturbations in the initial value remain uniformly controlled over the time interval.

For one-step methods, a sufficient condition for stability is that the increment function Ψ be Lipschitz in the state variable. Indeed, if

$$|\Psi(t,x,\Delta) - \Psi(t,y,\Delta)| \leq L\,|x-y|\,,$$

then

$$|y_{k+1} - \widetilde{y}_{k+1}| \leq (1 + L\Delta t_k)\,|y_k - \widetilde{y}_k|\,.$$

Iterating gives

$$|y_k - \widetilde{y}_k| \leq \exp(L(T-t_0))\,|y_0 - \widetilde{y}_0|\,.$$

So perturbations do not blow up on a finite time interval.
However, numerical stability on finite intervals does not fully explain how to choose a good step size. To study step-size restrictions, one usually tests a method on the scalar linear equation

$$y' = \lambda y, \qquad \lambda \in \mathbb{C}.$$

If $\Re(\lambda) < 0$, the exact solution decays to zero. A stable numerical method should reproduce this decay rather than amplify errors.

Definition 17.16 (*Region of Absolute Stability*)
For the test equation

$$y' = \lambda y,$$

a one-step method often takes the form

$$y_{k+1} = R(z) y_k, \qquad z := \lambda \Delta t,$$

where R is called the **stability function**. The **region of absolute stability** is

$$\mathcal{S} := \{z \in \mathbb{C} : |R(z)| \leq 1\}.$$

Thus $z = \lambda \Delta t$ must lie in $\mathcal{S}$ for numerical errors in the test equation not to grow.

For Forward Euler applied to $y' = \lambda y$, we get

$$y_{k+1} = y_k + \Delta t\,\lambda y_k = (1 + \lambda \Delta t) y_k.$$

Hence

$$R(z) = 1 + z,$$

and the absolute stability region is

$$\mathcal{S}_{\mathrm{FE}} = \{z \in \mathbb{C} : |1 + z| \leq 1\}.$$

This is the disk of radius 1 centered at -1 in the complex plane.
For example, if

$$y' = -16y,$$

then Forward Euler gives

$$y_{k+1} = (1 - 16\Delta t) y_k.$$

For the iterates to decay, we need

$$|1 - 16\Delta t| < 1,$$

which gives

$$0 < \Delta t < \frac{1}{8}.$$

Thus even though the exact solution decays rapidly to zero, Forward Euler becomes unstable if the step size is too large.
Implicit methods often have better stability properties. For Backward Euler,

$$y_{k+1} = y_k + \Delta t\, \lambda y_{k+1},$$

so

$$y_{k+1} = \frac{1}{1 - \lambda \Delta t} y_k.$$

Thus

$$R(z) = \frac{1}{1-z},$$

and the stability region is

$$\mathcal{S}_{\mathrm{BE}} = \left\{ z \in \mathbb{C} : \left| \frac{1}{1-z} \right| \leq 1 \right\} = \{ z \in \mathbb{C} : |1 - z| \geq 1 \}.$$

In particular, the entire left half-plane is contained in $\mathcal{S}_{\mathrm{BE}}$.

DEFINITION 17.17 (A-STABILITY)

A numerical method is called ***A*-stable** if its region of absolute stability contains the entire left half-plane:

$$\{ z \in \mathbb{C} : \Re z \leq 0 \} \subseteq \mathcal{S}.$$

Equivalently, whenever the exact solution of $y' = \lambda y$ decays, the numerical solution does not grow for any step size $\Delta t > 0$.

Backward Euler is A-stable, while Forward Euler is not. The trapezoidal method is also A-stable. Indeed, applying

$$y_{k+1} = y_k + \frac{\Delta t}{2} (\lambda y_k + \lambda y_{k+1})$$

to the test equation gives

$$y_{k+1} = \frac{1 + z/2}{1 - z/2} y_k, \qquad z = \lambda \Delta t.$$

Hence

$$R(z) = \frac{1 + z/2}{1 - z/2}.$$

One checks that

$$|R(z)| \leq 1 \qquad \Longleftrightarrow \qquad \Re z \leq 0.$$

Thus the trapezoidal method is A-stable.
This distinction is especially important for **stiff equations**. Roughly, an ODE is stiff when it has several time scales, some of which decay much more rapidly than others. For example,

$$\frac{d}{dt} \begin{pmatrix} y^1 \\ y^2 \end{pmatrix} = \begin{pmatrix} -\alpha_1 & 0 \\ 0 & -\alpha_2 \end{pmatrix} \begin{pmatrix} y^1 \\ y^2 \end{pmatrix}, \qquad 0 < \alpha_2 \ll \alpha_1.$$

The exact solutions are

$$y^1(t) = y^1(0)e^{-\alpha_1 t}, \qquad y^2(t) = y^2(0)e^{-\alpha_2 t}.$$

The first component decays very quickly, while the second evolves slowly. Forward Euler requires

$$\Delta t \leq \frac{2}{\alpha_1}$$

for stability of the fast component, even if the slow component is the one we care about. Thus explicit methods may require extremely small step sizes on stiff problems. Implicit A-stable methods can often take much larger stable steps.

For multi-step methods, stability has an additional feature. A q-step linear multi-step method can be written as

$$\sum_{j=0}^{q} \alpha_j y_{k+j} = \Delta t \sum_{j=0}^{q} \beta_j f(t_{k+j}, y_{k+j}).$$

Applying this to the test equation $y' = \lambda y$ and trying solutions of the form $y_k = \xi^k$ gives the characteristic equation

$$\sum_{j=0}^{q} (\alpha_j - z\beta_j)\xi^j = 0, \qquad z = \lambda \Delta t.$$

The method is stable for a given z if all roots ξ lie in the unit disk, with any roots on the unit circle being simple. At $z = 0$, this gives the **root condition**

$$\sum_{j=0}^{q} \alpha_j \xi^j = 0$$

has all roots satisfying $|\xi| \leq 1$, and all roots with $|\xi| = 1$ are simple. This condition is called **zero-stability**. For linear multi-step methods, consistency together with zero-stability implies convergence.

17.2. Numerical Methods for PDEs

17.3. The Wagner-Platen Expansion

17.4. Strong Approximations to SDEs

17.5. Weak Approximations to SDEs

18. (L29) Stochastic Filtering and Particle Filters

18.1. Hidden Markov Models and Conditional Distributions

18.2. The Continuous-Time Filtering Problem

18.3. The Innovations Process

18.4. The Kushner–Stratonovich Equation

18.5. The Zakai Equation

18.6. The Kalman–Bucy Filter

18.7. Particle Filters and Sequential Monte Carlo

19. (L30) Applications to Physics

19.1. Quantum Monte Carlo Methods

20. (L31) Applications to Economics

This section is adapted from Section 5.8.1 and 5.8.2 of [KS88] and Section 13.10 of [GS20]. We give a cursory introduction to Portfolio and Consumption pricing and Options Pricing for European Call Options with the Black-Scholes-Merton formula.

The main idea is essentially as follows: assume we can buy **forward-options** or **derivatives** on stock in a market, that is, we can buy an option for a price K (called the **strike price**) to buy a stock at a later time T, called the **exercise date**. If you assume an efficient market (ie. a market with no risk-free arbitrage opportunities) with assets that consist of stocks and bonds, where bonds compound exponentially as

$$dM_t = rM_t dt \implies M_t = \exp(rt) \tag{20.1}$$

and a stock (a more "risky" security) whose values adapt stochastically as:

$$dS_t = S_t(\mu dt + \sigma dW_t) \implies S_t = \exp((u_t - \frac{1}{2}\sigma_t^2)t + \sigma W_t) \tag{20.2}$$

then a self-financing portfolio (ie. one with no external investment) has a value function $w(t, x)$ that satisfies the Black-Scholes PDE:

Theorem 20.1 (*Black-Scholes PDE*)
Suppose $w(t,x)$ is twice continuously differentiable. Then w is the value function of a self-financing portfolio iff

$$\frac{1}{2}\sigma^2 x^2 \omega_{xx} + rx\omega_x + \omega_t - r\omega = 0 \tag{20.3}$$

By solving the PDE and using the Feynman-Kac formula, one can then show that the risk-free "value" of a European Call-Option (ie. an option that can *only* be exercised at the exercise date) is given in closed-form:

Theorem 20.2 (*Black-Scholes Formula*)
Let $t < T$. The value at time t of the European Call Option is

$$S_t \Phi(d_1(t, S_t)) - K \exp(-r(T-t))\Phi(d_2(t, S_t)) \tag{20.4}$$

where $\Phi \sim \mathcal{N}(0,1)$ and

$$d_1(t,x) := \frac{\log(x/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, \qquad d_2(t,x) = d_1(t,x) - \sigma\sqrt{T-t} \tag{20.5}$$

20.1. Portfolio and Consumption Processes

Consider a market with $d+1$ **assets** or **securities** that are traded continuously with a fixed time horizon $0 \leq T < \infty$. Two types of assets comprise this market: a "bond" which grows at a constant compound interest rate r and a "stock" whose value is modeled by a stochastic process.

As a general assumption we assume W to be a d-dimensional Wiener process on the filtration $(\mathcal{F}_t^W)$ and we let the filtration of the underlying process P, denoted $\mathcal{F}_t$ be the canonical completion of $(\mathcal{F}_t^W)$ with respect to P.

Definition 20.3 (*Bonds*)
Of the $d+1$ assets, one of the assets, known as the **bond** has a price $P_0(t)$ which evolves according to the differential equation:

$$dP_0(t) = r(t)P_0(t)dt, \qquad P_0(0) = p_0 \tag{20.6}$$

Definition 20.4 (*Stocks*)
The remaining d assets, called **stocks** are "risky" assets modeled by a stochastic differential equation:

$$dP_i(t) = b_i(t)P_i(t)dt + P_i(t)\sum_{j=1}^{d} \sigma_{ij}(t) W_t^j, \qquad P_i(0) = p_i \tag{20.7}$$

where $r(t)$ is an $\mathcal{F}_t$-adapted process called the **interest rate**, $b(t) = (b_1(t), \cdots, b_d(t))^T$ is an $\mathcal{F}_t^d$-adapted process called the **mean rates of return**, and $\Sigma = [\sigma_{ij}]$ is an $\mathcal{F}_t^{d\times d}$ adapted process called the **dispersion matrix** that we assume are uniformly bounded on for $(t, \omega) \in [0, T] \times \Omega$. Letting $\Lambda = \Sigma\Sigma^T$, we assume for some $\varepsilon > 0$ that

$$\xi^T \Lambda \Xi \geq \varepsilon \|\xi\|^2 \tag{20.8}$$

for all $\xi \in \mathbb{R}^d$

Lemma 20.5

Under the assumptions above, Σ^T has an inverse

We now introduce the notions of portfolios and consumption processes, which intuitively describe how much stock an investor has at any given moment and how much of that money the investor uses. Heuristically, assume an investor starts with some initial endowment $x \geq 0$ who invests in these $d+1$ assets. Let $N_i(t)$ denote the amount of shares of asset i owned at time t. The investor's **wealth** at time t is given by:

$$X_t = \sum_{i=0}^{d} N_i(t) P_i(t) \tag{20.9}$$

where $X_0 = x = \sum_{i=0}^{d} N_i(0) p_i$. We define the portfolio heuristically ast he amount of money given by these stocks at time t: $\pi_i(t) := N_i(t) P_i(t)$. The rigorous definition is as follows:

Definition 20.6 (*Portfolio Process*)

A **portfolio process** $\pi = \{(\pi_1(t), \cdots, \pi_d(t))^T\}$ is a measurable adapted process for which

$$\sum_{i=1}^{d} \int_0^T \pi_i^2(t) dt < \infty \tag{20.10}$$

a.s., ie. its integrable with respect to the Wiener process.

Definition 20.7 (*Consumption Process*)

A **consumption process** C is a measurable adapted non-negative process with

$$\int_0^t C_t dt < \infty \tag{20.11}$$

a.s.

Remark. The portfolio process is allowed to be negative, which would be the equivalent of "short-selling" or "borrowing" the stock.

Now intuitively, the difference in the investor's wealth at a given time t is given by the change in the performance of their stocks minus the amount they consume. That is,

$$dX_t = \sum_{i=0}^{d} N_i(t) dP_i(t) - C_t dt \tag{20.12}$$

Substituting in the form for π, we get the following differential equation for the wealth of the investor:

Definition 20.8 (*Wealth Process*)
A **wealth process** X_t for a pair (π, C) of portfolio and consumption processes is a process that satisfies the SDE:

$$dX_t := (r(t)X_t - C_t)dt + \sum_{i=1}^{d}(b_i(t) - r(t))\pi_i(t)dt + \sum_{i=1}^{d}\sum_{j=1}^{d}\pi_i(t)\sigma_{ij}(t)dW_i^{(j)} \tag{20.13}$$

One can show then that X_t obeys:

$$X_t = e^{\int_0^t r(s)ds}\left[x + \int_0^t e^{-\int_0^s r(u)du}\left(\pi(s)^T(b(s) - r(s)\mathbf{1}) - C_s\right)ds + \int_0^t e^{-\int_0^s r(u)du}\pi(s)^T\sigma(s)dW_s\right] \tag{20.14}$$

20.2. Option Pricing

20.3. Option Consumption and Investment

Part V. Appendix

21. Appendix A: Extension Theorems

We discuss some common extension theorems. The main ones of note are the Dynkin $\pi - \lambda$ lemma and monotone class theorem which allows you to mainly show some property on a generating "cylindrical set" and extending to the entire σ-algebra, and the Kolomogorov Extension Theorem which allows the extension of a collection of measures on a discrete time sequence to a measure on a continuous time process.

21.1. Monotone Class Theorem

Some notes on the monotone class lemma which are useful for some theorems in class and in the textbook. Recall the definition of a monotone class:

Definition 21.1 (*Monotone Class*)
A subset $\mathcal{M} \subseteq \mathcal{P}(E)$ is called a **monotone class** if

(i.) $E \in \mathcal{M}$

(ii.) If $A, B \in \mathcal{M}$, $A \subseteq B$, then $B \backslash A \in \mathcal{M}$

(iii.) If (A_n) is an increasing sequence of subsets of E such that $A_n \in \mathcal{M}$ then $\cup_{n\geq 0}A_n \in \mathcal{M}$.

Definition 21.2 (*Monotone Class generated by a subset of $\mathcal{P}(E)$*)
If $\mathcal{C} \subseteq \mathcal{P}(E)$, the **monotone class generated by** $\mathcal{C}$, denoted $\mathcal{M}(\mathcal{C})$ is defined as the intersection of

all monotone classes containing $\mathcal{C}$.

Lemma 21.3 (*Monotone Class Lemma*)
If $\mathcal{C} \subseteq \mathcal{P}(E)$ is closed under finite unions, then $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$.

Note the monotone class lemma is essentially equivalent to the Dynkin $\pi - \lambda$ lemma, just with $\mathcal{C}$ being a π-system (note a Monotone class is similar to a λ-system except

21.2. Kolomogorov Extension Theorem

We prove the Kolomogorv Extension Theorem for the special case of polish spaces. Throughout this section let $(E, \mathcal{E})$ be a measurable space.

Definition 21.4 (Ω^*)
We denote by $\Omega^* = E^{\mathbb{R}_+}$ be the space of all mappings $\omega : \mathbb{R}_+ \to E$ and let $\mathcal{F}^*$ be the σ-algebra generated by all coordinate mappings $\omega \mapsto \omega(t)$ for $t \in \mathbb{R}_+$.

Definition 21.5
Let $F(\mathbb{R}_+)$ denote the set of all finite subsets of $\mathbb{R}_+$ and for $U \in F(\mathbb{R}_+)$ let $\pi_U : \Omega^* \to E^U$ be the mapping which associates each $\omega : \mathbb{R}_+ \to E$ to its restriction to U, $\omega^U : \mathbb{R}_+ \to E^U$. For $U \subseteq V$ we define $\pi_U^V : E^V \to E^U$ in the natural way (so $\pi_U^V \pi_V = \pi_U$).

Intuitively, we have for some $\omega(t) \in \Omega^*$, we have that if $U = \{t_1, \cdots, t_p\}$ $\pi_U(\omega) = (\omega(t_1), \cdots, \omega(t_n))$ and similarly
Now, recall the definition of a Polish space:

Definition 21.6 (*Polish Space*)
A topological space $\mathcal{X}$ is said to be **polish** if it is seperable (ie. there is a countable dense subsequence) and has a complete metric.

Theorem 21.7 (*Kolomogorov Extension Theorem*)
Assume E is a Polish space equipped with a Borel σ-field $\mathcal{E}$. For every $U \in F(\mathbb{R}_+)$ let μ_U be a probability measure on E^U. Assume that $(\mu_U; u \in F(\mathbb{R}_+)))$ is consistent in the following sense: if $U \subseteq V$ we have $\mu_U = \pi_U^V(\mu_V)$. Then, there exists a unique probability measuyre μ on $(\Omega^*, \mathcal{F}^*)$ such htat $\pi_U(\mu) = \mu_U$ for each $U \in F(\mathbb{R}_+)$.

The intuitive idea is that for a Borel σ-algebra on a Polish Space, we can specify any finite dimensional marginals and then we get a unique measure on the whole space. Note this does not give us any information about the measurability of the consituent sets however.

22. Appendix B: Poisson Processes

We recap the definitions of the standard Poisson process. This is the most natural example of a so-called "jump" process that moves in a discontinuous manner.

The intuition is as follows: informally, suppose we have an infinite number of small Bernoulli random variables $\zeta_j(dt)$ that have a small dt probability of being 1. Let $N_t := \int \zeta(t)dt$ be the sum of the Bernoulli, variables that is $N_t \sim \text{Bin}(t/dt, dt) := \text{Pois}(t)$. We have that N_t is a Poisson random variable with for any $s \leq t$, $N_t - N_s \sim \text{Bin}((t-s)/dt, dt) := \text{Pois}(t-s)$. We have that (N_t) then has independent increments.

More specifically, we define the standard (discrete) Poisson process as follows:

Definition (*Poisson Process*)

Let $\zeta_i \overset{iid}{\sim} \exp(1)$. Define times $T_1 = \zeta_1$, $T_2 = \zeta_1 + \zeta_2$, $T_3 = \zeta_1 + \zeta_2 + \zeta_3, \cdots$. Define the **Poisson process** (N_t) as

$$N_t := \max\{k : T_k \leq t\} \tag{22.1}$$

Then $(N_t)_{t\geq 0}$ is a stochastic process, non-decreasing and right continuous in t.

An alternative definition is as follows:

Definition

A **Poisson process** with rate $\lambda > 0$ is a stochastic process $(N_t)_{t\geq 0}$ such that

1. $N_0 = 0$ a.s.
2. $N_t \in \mathbb{N}_0$ for every t and $t \mapsto N_t$ is almost surely nondecreasing and cadlag.
3. For $0 \leq t_1 < t_2 < \cdots < t_p$, we have
$$N_{t_1} - N_{t_0} \perp N_{t_2} - N_{t_1} \perp \cdots , N_{t_p} - N_{t_{p-1}} \tag{22.2}$$
4. $N_t - N_s \sim \text{Pois}(\lambda(t-s))$ for $0 \leq s \leq t$.

23. Appendix C: Change of Measure and the Radon-Nikodym Theorem

We recall the Radon-Nikodym Theorem that gives us that for absoltuely continuous measures there exists a unique density that changes measure with respect to the other measure.

Definition 23.1 (*Absolute Continuity*)

Let μ, ν be a positive measure on a measurable space $(\Omega, \mathcal{F})$. We say that ν is **absolutely continuous** with respect to μ , denoted $\nu \ll \mu$ if $\mu(A) = 0$ implies $\nu(A) = 0$ where $A \in \mathcal{F}$.

Theorem 23.2 (*Radon-Nikodym Theorem*)
Let μ and ν be a positive σ-finite measure on a measurable space $(\Omega, \mathcal{F})$. If $\nu \ll \mu$ there is a function f on $\mathcal{F}$ so that foir all $A \in \mathcal{F}$

$$\nu(A) = \int_A f d\mu \tag{23.1}$$

f is typically denoted by $d\nu/d\mu$ and is called the **Radon-Nikodym Derivative**

24. Appendix D: PDE Review

We review some material that is useful in the study of PDEs which we make connections to. We first recall the definition of a *formal adjoint* which differs technically from the Hilbert Space adjoint as it is defined in the weak sense in terms of the integration by parts identity on compactly supported test functions.

Definition 24.1 (*Formal adjoint*)
Let $U \subset \mathbb{R}^d$ be open, and let

$$P(x, D)u = \sum_{|\alpha| \leq m} a_\alpha(x) \partial^\alpha u \tag{24.1}$$

be a linear differential operator with smooth coefficients $a_\alpha \in C^\infty(U)$. The **formal adjoint** of P is the differential operator P^* defined by the identity

$$\int_U (Pu)(x)\varphi(x)\, dx = \int_U u(x)(P^*\varphi)(x)\, dx \tag{24.2}$$

for all $u, \varphi \in C_c^\infty(U)$.

Remark. Intuitively, we just consider the formal adjoint in the same way as the Hilbert Space adjoint defined in terms of a functional defined by the commutation relation.

Definition 24.2 (*Weak solution of a PDE*)
Let $U \subset \mathbb{R}^d$ be open, and let

$$P(x, D)u = f \tag{24.3}$$

be a linear PDE on U, where

$$P(x, D)u = \sum_{|\alpha| \leq m} a_\alpha(x) \partial^\alpha u. \tag{24.4}$$

A function $u \in L^1_{\text{loc}}(U)$ is called a **weak solution** of $P(x, D)u = f$ if, for every test function $\varphi \in C_c^\infty(U)$,

$$\int_U u(x) P^*(x, D)\varphi(x)\, dx = \int_U f(x)\varphi(x)\, dx. \tag{24.5}$$

25. Appendix E: Integrability Criterion Cheatsheat

Here is a brief review of the integrability criterions for the various types of stochastic (Ito) integrals that I wrote before the midterm:

Lemma 25.1

Suppose we have a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$.

(a) If M is a continuous square martingale or local martingale, we require that H is progressive and that
$$\mathbb{E}\left[\int_0^\infty H_s^2\, d\langle M, M\rangle_s\right] < \infty.$$

(b) If M is a semimartingale, we require that H is a locally bounded semimartingale.

(c) If V is a finite variation process, we require that $\int_0^t |H_s||dA| < \infty$

References

[Cao25] Sky Cao. *18.676: Stochastic Calculus Lecture Notes (2025)*. 2025.

[Fol08] Gerald B Folland. *Quantum Field Theory: A Tourist Guide for Mathematicians*. Vol. 149. Mathematical Surveys and Monographs. Providence, RI: American Mathematical Society, 2008. ISBN: 978-0-8218-4705-3.

[GS20] Geoffrey R. Grimmett and David R. Stirzaker. *Probability and Random Processes*. 4th ed. Oxford University Press, 2020. ISBN: 978-0-19-884759-5.

[Kal21] O. Kallenberg. *Foundations of Modern Probability*. 3rd ed. Springer Cham, 2021. ISBN: 978-3-030-61870-4. DOI: 10.1007/978-3-030-61871-1.

[KP92] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. 1st ed. Stochastic Modelling and Applied Probability. Springer Berlin, Heidelberg, 1992, pp. XXXVI, 636. ISBN: 978-3-662-12616-5. DOI: 10.1007/978-3-662-12616-5.

[KS88] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Vol. 113. Graduate Texts in Mathematics. Springer, 1988.

[KS91] Ioannis Karatzas and Steven Shreve. *Brownian Motion and Stochastic Calculus*. Vol. 113. Graduate Texts in Mathematics. Springer, 1991. ISBN: 978-0-387-97655-6. DOI: 10.1007/978-1-4612-0949-2.

[Le 16] Jean-François Le Gall. *Brownian Motion, Martingales, and Stochastic Calculus*. Vol. 274. Graduate Texts in Mathematics. Springer Verlag, 2016. ISBN: 978-3-319-31088-6. DOI: 10.1007/978-3-319-31089-3.

[PW25] Yury Polyanskiy and Yihong Wu. *Information Theory: From Coding to Learning*. Cambridge University Press, 2025. URL: https://people.lids.mit.edu/yp/homepage/data/itbook-export.pdf.

[SC19] Daniel Stroock and Sinho Chewi. *18.176: Stochastic Calculus Lecture Notes (Spring 2019)*. 2019.