Introduction to Stochastic Calculus

A lot of people...

Contents

Syllabus

The main session will focus on getting an intuitive grasp of stochastic calculus. The main reference here is Oksendal [11], abbreviated as "Ok" below.

Week #	Date	Topic	Reading
1	05/20 - 05/24	Preliminaries	Ok Chapter 1 & 2
2	05/27 - 05/31	The Ito integral	Ok Chapter 3
3	06/03 - 06/07	Ito formula and martingale representation	Ok Chapter 4
4	06/10 - 06/14	Stochastic differential equations	Ok Chapter 5
5	06/17 - 06/21	Properties of diffusion processes	Ok Chapter 7
6	06/24 - 06/28	Feynman-Kac and the martingale problem	Ok Chapter 8.1-8.4
7	07/01 - 07/05	Random time-change and Girsanov	Ok Chapter 8.5-8.6
8	07/08 - 07/12	Boundary value problems	Ok Chapter 9
9	07/15 - 07/19	Optimal stopping	Ok Chapter 10
10	07/22 - 07/26	Stochastic control	Ok Chapter 11
11	07/29 - 08/02	–	
12 and on		find selected topics of interest?	

The auxiliary section will try to establish an in-depth treatment of stochastic calculus. The main reference here will be Le Gall [10], abbreviated as "LG" below.

Week #	Date	Topic	Reading
1	05/20 - 05/24	Gaussian spaces and Brownian motion	LG Chapter 1 & 2
2	05/27 - 05/31	martingales	LG Chapter 3
3	06/03 - 06/07	Continuous semimartingales	LG Chapter 4
4	06/10 - 06/14	–	–
5	06/17 - 06/21	Integration and Ito’s formula	LG Chapter 5.1-5.3
6	06/24 - 06/28	martingale representation and Girsanov	LG Chapter 5.4-5.6
7	07/01 - 07/05	Markov Processes	LG Chapter 6
8	07/08 - 07/12	–	–
9	07/15 - 07/19	Connections to PDEs	LG Chapter 7
10	07/22 - 07/26	Stochastic Differential Equations	LG Chapter 8
11	07/29 - 08/02	Local times	LG Chapter 9
12	08/05 - 08/09	–	–

1 Existence of Brownian Motion

We want to construct a Gaussian process on $\mathbb{R}_+ = [0, \infty)$ that has independent and stationary increments.

1.1 Via Kolmogorov extension

Theorem 1.1 (Kolmogorov's extension theorem). *For all $t_1, \dots, t_k \in \mathbb{T}$, $k \in \mathbb{N}$, let $\nu_{t_1,\dots,t_k}$ be a probability measure on $\mathbb{R}^{nk}$ such that*

$$\nu_{\sigma(t_1),\dots,\sigma(t_k)}(F_1 \times \cdots \times F_k) = \nu_{t_1,\dots,t_k}(F_{\sigma^{-1}(1)} \times \cdots \times F_{\sigma^{-1}(k)}) \tag{1.1}$$

for any permutation σ and for all $t_{k+1}, \dots, t_{k+m} \in \mathbb{T}$

$$\nu_{t_1,\dots,t_k}(F_1, \dots, F_k) = \nu_{t_1,\dots,t_k,t_{k+1},\dots,t_{k+m}}(F_1, \dots, F_k, \mathbb{R}^n, \dots, \mathbb{R}^n). \tag{1.2}$$

Then, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and process $\{X_t : \Omega \to \mathbb{R}^n\}_{t\in\mathbb{T}}$ such that it has finite-dimensional distributions $\nu_{t_1,\dots,t_k}$.

Let $\mathbb{T} = [0, \infty)$, defining

$$p(t, x, y) = \frac{1}{(2\pi t)^{-n/2}} \exp\left(\frac{-\|x - y\|^2}{2t}\right), \tag{1.3}$$

and letting the finite dimensional distribution be

$$\nu_{t_1,\dots,t_k}(F_1, \dots, F_k) = \int_{F_1\times\cdots\times F_k} p(t_1, 0, x_1)p(t_2 - t_1, x_1, x_2) \dots p(t_k - t_{k-1}, x_{k-1}, x_k)dx_1 \dots dx_k, \tag{1.4}$$

Kolmogorov's extension theorem gives us the existence of a process (we call $\{B_t\}_{t\geq 0}$) with such marginals. By construction, we know that it is a Gaussian process with independent increments. Taking another theorem for granted, we get continuity of sample paths.

Theorem 1.2 (Kolmogorov's continuity theorem). *Suppose that $\{X_t\}_{t\geq 0}$ satisfies for all $T > 0$, there exist $\alpha, \beta, D > 0$ such that*

$$\mathbb{E}\,|X_t - X_s|^\alpha \leq D|t - s|^{1+\beta} \tag{1.5}$$

for $0 \leq s, t \leq T$. Then, there exists a $\{\tilde{X}_t\}_{t\geq 0}$ that is continuous version of X, i.e., $\mathbb{P}(X_t = \tilde{X}_t) = 1$ for all $t \geq 0$, with a.s. continuous sample paths.

Picking $\alpha = 2$, $D = 1$, $\beta = 1$ gives the existence of a continuous version.

Definition 1.3 (Canonical Brownian Motion). Let $C[0, \infty)$ be the space of continuous functions on $\mathbb{R}_+$. The canonical Brownian motion, is one such that B_t is taken to be the coordinate map $B_t(x) = x_t$ for $x \in C[0, \infty)$.

1.2 Via weak convergence

For this subsection, we turn our attention to $C[0, T]$ for some finite $T < \infty$. Then, the space equipped with the uniform metric is Polish and we can characterize compactness exactly.

Theorem 1.4 (Arzela-Ascoli). *A set $A \subset C[0, T]$ is pre-compact if and only if it is uniformly bounded, i.e., $\sup_{x\in A} |x(0)| < \infty$, and uniformly equicontinuous, that is,*

$$\lim_{\delta\to 0} \sup_{x\in A} w_x(\delta) = 0 \tag{1.6}$$

where $w_x(\delta) = \sup_{|s-t|\leq\delta} |x_s - x_t|$ is the modulus of continuity.

From compactness, we can relate it to tightness of probability measures via Prokhorov.

Theorem 1.5. *The sequence of probability measures $(P_n)_{n\geq 0}$ is tight if and only if for any $\eta > 0$, there exists an $a < \infty$ such that*

$$P_n(\{x : |x(0)| \geq a\}) \leq \eta \tag{1.7}$$

for all n and for every $\epsilon > 0$,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} P_n(\{x : w_x(\delta) \geq \epsilon\}) = 0. \tag{1.8}$$

Proof. First, suppose that $\{P_n\}$ is tight. Then, for any fixed $\eta > 0$, there is a compact K such that $P_n(K) > 1 - \eta$ for all n. By Arzela-Ascoli, we have that $K \subset \{x : |x(0)| \leq a\}$ for large enough a and $K \subset \{x : w_x(\delta) \leq \epsilon\}$ for small enough δ and any ϵ. Therefore, the two conditions listed follows.

For the other direction, suppose that we have the two conditions listed. Fix some $\eta > 0$. By the first condition, we know that there is an $a < \infty$ such that $P_n(B) \geq 1 - \eta$ for all n where $B = \{x : |x(0)| \leq a\}$. Moreover, for each k, choose δ_k such that $B_k = \{x : w_x(\delta_k) < 1/k\}$ and $P_n(B_k) \geq 1 - \eta/2^k$ for all n. Let $K = \mathrm{cl}(B \cap \bigcap_k B_k)$, then by Arzela-Ascoli, K is compact and $\{P_n\}$ is tight as $\sup_n P_n(K) \leq 1 - 3\eta$. □

Then, we can further relate tightness to convergence in finite-dimensional distributions.

Theorem 1.6. *Let $(P_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $(C, \mathcal{B}_C)$ that is tight. Then, if the finite-dimensional distributions converge, $P_n \to P$.*

Proof. Since $(P_n)_n$ is tight, for any subsequence, there is a further subsequence that converges; denote this limit by $\tilde{P}$. If we can show that $\tilde{P} = P$, then we would have shown that $P_n \to P$. Abusing notation a little, we will denote the sub-sub-sequence with subscripts n.

The proof then follows from the observation that the finite-dimensional sets generates $\mathcal{B}_C$, that is,

$$\mathcal{B}_C = \sigma\left\{\pi^{-1}_{t_1,\ldots,t_k} A : A \in \mathcal{B}_{\mathbb{R}^k}\right\}. \tag{1.9}$$

So, by the assumption that $P_n\pi^{-1}_{t_1,\ldots,t_k} \to P\pi^{-1}_{t_1,\ldots,t_k}$, we have $P = \tilde{P}$ on a set that generates the σ-algebra, and the proof is complete. □

Using these weak convergence techniques, we will construct a tight sequence of measures on $C[0, T]$ that satisfies the increment properties we want. Let $(\xi_k)_{k\in\mathbb{N}}$ be a sequence of independent standard Gaussians. Let $S_n = \sum_{k=1}^n \xi_k$ be the associated random walk. The sequence of measures (P_n) will be the law of the interpolated process X_t:

$$X_t = \begin{cases} 0 & \text{if } t = 0 \\ S_n/\sqrt{n} & \text{if } t = T/n \\ \text{linear interpolation} & \text{otherwise} \end{cases} \tag{1.10}$$

Then, we call the convergent limit of (P_n) the Wiener measure and we've showed the existence of the canonical Brownian motion (unknowingly). By envoking CLT, we need not have ξ_k's be Gaussians—this is known as the Donsker's theorem.

1.3 Via Hilbert space theory

Let $\mathcal{H} = L^2(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+}, m)$ where m is the Lebesgue measure with orthonormal basis $\{e_n\}_{n\in\mathbb{N}}$. Moreover, let $(g_n)_{n\in\mathbb{N}}$ be independent standard Gaussians defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We will define the collection of random variables $\{X(h)\}_{h\in\mathcal{H}}$ to be

$$X(h) = \sum_{n=1}^{\infty} \langle h, e_n\rangle g_n. \tag{1.11}$$

Then, we can see that these random variables inherited the structure of the Hilbert space, i.e., $\mathbb{E}\, X(h)X(h') = \langle h, h'\rangle$. Consequently, $\mathbb{E}\, X(h)^2 = \|h\|^2$ and $X(h)$ and $X(h')$ are independent if and only if h and h' are orthogonal. Moreover, observe that for $F, G \in \mathcal{B}_{\mathbb{R}_+}$ with finite measure, we have

$$\mathbb{E}\, X(1_F)X_{1_G} = \langle 1_F, 1_G\rangle = m(F \cap G). \tag{1.12}$$

Let's call $B_t = X(1_{[0,t]})$. Independent increments follow from the disjointness of time intervals. Gaussianity follows from infinite divisibility and independent increment. The variance $\mathbb{E}\, B_t^2 = t$ by the choice of Lebesgue measure. And we have Brownian motion.

1.4 Via multiscale construction Again, we focus on the finite horizon case, in particular, the $[0, 1]$ case. Let γ be the standard Gaussian, and consider the algorithm:

1. Let $B_0^0 = 0$ and $B_1^0 \sim \gamma$.
2. For each iteration n, do:
3. At location $t = k2^{-n}$ for all $k = 0, \ldots, 2^n$,
   (a) If B_t^n already exist (visited in previous iteration), skip!
   (b) If not, let $Y_0 = B^n_{k2^{-(n-1)}}$ and $Y_1 = B^n_{(k+1)2^{-(n-1)}}$, and let $B^n_{(2k+1)2^{-n}} = (Y_0 + Y_1)/2 + 2^{-(n+1)}\xi$ where $\xi \sim \gamma$ independently of everything else.

The construction above inherits all independence properties needed and can be used to prove existence using the help of wavelet-type multi-resolution analysis. The benefit of this construction is that, by showing B_t^n has a uniform limit almost surely (error estimates + Borel-Cantelli), we get continuity by construction rather than by envoking God-sent theorems.

2 Stochastic Differential Equations

From the theory of ordinary differential equations, we know that evolution equations of the type

$$dX_t = b(t, X_t)dt \tag{2.1}$$

admits a unique solution when b is Lipschitz. The two classic arguments to prove this is: 1) we carry out existence via compactness arguments of the linear interpolation of Euler approximations and uniqueness via Gronwall, or 2) iterate the solution (Picard iteration) and show that it is a contraction.

Now, we're interested in the existence and uniqueness of *stochastic* differential equations

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \tag{2.2}$$

which corresponds to the integral equation

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s. \tag{2.3}$$

It turns out that similar techniques used for the deterministic case gives existence and uniqueness under similar smoothness conditions on b and σ. Moreover, we'll also dive into the caveat that the solution might not be defined pathwise ω-by-ω, but admits a solution in terms of the law on a potentially different probability space.

2.1 Existence and uniqueness of (strong) solutions

Theorem 2.1 (Existence and uniqueness of strong solutions [11, Theorem 5.2.1])**.** *Let $T > 0$ and consider the SDE in (2.2) driven by $\{\mathcal{F}_t\}_{t\geq 0}$-Brownian motion. Let b and σ be such that for any fixed $t \in [0, T]$,*

1. *(linear growth) $|b(t, x)| + |\sigma(t, x)| \leq C(1 + |x|)$, and*
2. *(Lipschitz continuous) $|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq D|x - y|$.*

for $x, y \in \mathbb{R}^d$, $C, D > 0$. Let $Z \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ be a random variable independent of the Brownian motion. Then, the SDE with initial conditions $X_0 = Z$ has a unique solution that is continuous, adapted to the filtration $\{\sigma Z \bigvee \mathcal{F}_t\}_{t\geq 0}$, and $X_\cdot \in L^2(\Omega \times [0, T])$.

Proof. We start by proving existence via showing the sequence of processes obtained from the Picard iteration is Cauchy. Then, using the same estimates, we show uniqueness via Gronwall. Lastly, continuity of the sample paths is immediate from continuity of Brownian motion and the coefficients.

Existence Let $X^{(0)} \equiv X_0$ and define $X^{(k)}$ recursively by

$$X_t^{(k+1)} = X_0 + \int_0^t b(t, X_s^{(k)})ds + \int_0^t \sigma(s, X_s^{(k)})dB_s. \tag{2.4}$$

Then, using Jensen's inequality and Itô's isometry, we can bound the difference between the iterations as

$$\mathbb{E}\,|X_t^{(k+1)} - X_t^{(k)}|^2 = \mathbb{E}\left(\int_0^t b(s, X_s^{(k)}) - b(s, X_s^{(k-1)})ds + \int_0^t \sigma(s, X_s^{(k)}) - \sigma(s, X_s^{(k-1)})dB_s\right)^2 \tag{2.5}$$

$$\leq 3\,\mathbb{E}\left(\int_0^t b(s, X_s^{(k)}) - b(s, X_s^{(k-1)})ds\right)^2 + 3\,\mathbb{E}\left(\int_0^t \sigma(s, X_s^{(k)}) - \sigma(s, X_s^{(k-1)})dB_s\right)^2 \tag{2.6}$$

$$\leq 3t\,\mathbb{E}\int_0^t |b(s, X_s^{(k)}) - b(s, X_s^{(k-1)})|^2 ds + 3\,\mathbb{E}\int_0^t |\sigma(s, X_s^{(k)}) - \sigma(s, X_s^{(k-1)})|^2 ds \tag{2.7}$$

$$\leq 3t(1 + D)\,\mathbb{E}\int_0^t |X_s^{(k)} - X_s^{(k-1)}|^2 ds. \tag{2.8}$$

Then, via Fubini, we can get that

$$\mathbb{E}\,|X_t^{(k+1)} - X_t^{(k)}|^2 \leq \frac{A^{k+1}t^k}{(k+1)!} \tag{2.9}$$

for some constant A depending on C, D and $\mathbb{E}\, X_0^2$. Then, the sequence $(X^{(n)}) \subset L^2(\Omega \times [0,T])$ is Cauchy because for $n > m \in \mathbb{N}$, we have

$$\|X^{(n)} - X^{(m)}\|_{L^2(\Omega\times[0,T]} \leq \sum_{k=n}^{m-1} \left\| X^{(k+1)} - X^{(k)} \right\|_{L^2(\Omega\times[0,T]} \leq \sum_{k=n}^{m-1} \left(\int_0^T \frac{A^{k+1}t^k}{(k+1)!} dt \right)^{1/2} \to 0 \tag{2.10}$$

and $m \to \infty$. Therefore, we have a candidate solution X that is the limit of sequence of outputs from the Picard iteration $X^{(n)}$. Moreover, since for any $t \in [0,T]$, Jensen and Itô's isometry gives

$$\int_0^t b(s, X_s^{(n)}) ds \to \int_0^t b(s, X) ds, \quad \int_0^t \sigma(s, X_s^{(n)}) ds \to \int_0^t \sigma(s, X) ds \tag{2.11}$$

where the convergence is in L^2, we have that X is indeed a solution to the desired SDE.

Uniqueness Suppose there are two solutions to the SDE, X and $\hat{X}$. Then, from some similar computation from before, we get that

$$\mathbb{E}\, |X_t - \hat{X}_t|^2 \leq 3t(1+D)\, \mathbb{E} \int_0^t |X_s - \hat{X}_s|^2 ds. \tag{2.12}$$

Now, recall Gronwall's inequality below.

Lemma 2.2 (Gronwall). *Let $u : \mathbb{R} \to \mathbb{R}$ be a continuous function satisfying*

$$u_t \leq u_0 + \int_0^t \beta_s u_s ds \tag{2.13}$$

for some $\beta : \mathbb{R} \to \mathbb{R}$. Then, we have

$$u_t \leq u_0 \exp\left(\int_0^t \beta_s ds \right). \tag{2.14}$$

It immediately follows that for all $t \in [0,T]$, $X_t = \hat{X}_t$ almost surely. In particular,

$$\mathbb{P}(X_t = \hat{X}_t \ \text{ for } \ t \in \mathbb{Q} \cap [0,T]) = 1. \tag{2.15}$$

Using the continuity of the sample paths, we get uniqueness up to indistinguishability. □

2.2 Tanaka's formula and pathwise non-uniqueness What we've shown just now is that, when the drift and diffusion coefficients are suitably well-behaved, we have a unique solution to the SDE that can be interpreted pathwise. We call this a *strong solution* as the solutions themselves are interpreted much like the ODE analog—it satisfies some integral equation, is adapted to the same filtration as the Brownian motion that is given before hand, and lives in the same probability space.

It turns out that the qualifier "strong" is necessary as there are SDEs which we thought we should amit a solution fails to do so in the typical sense. The protagonist of this subsection will be Tanaka's equation, which is an SDE of the form

$$dX_t = \operatorname{sgn}(X_t) dB_t, \ \ X_0 = 0 \tag{2.16}$$

where $\operatorname{sgn} = 1_{(0,\infty)]} - 1_{(-\infty,0]}$. Just from staring at the equation (and applying Itô's formula), we can see that for the same Brownian motion, if X_t is a solution, $-X_t$ will also be solution—pathwise uniqueness fails immediately! Moreover, for the rest of the subsection, we will show that we don't even have existence of strong solutions.

Proposition 2.3 (Tanaka formula). *Let L_t be the* local time *of a Brownian motion, i.e.,*

$$L_t = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} |\{s \in [0,t] : B_s \in (-\epsilon, \epsilon)\}| \tag{2.17}$$

where $|A|$ denotes the Lebesgue measure of the set A.Then, the local time satisfies

$$|B_t| = \int_0^t \operatorname{sgn}(B_s) dB_s + L_t. \tag{2.18}$$

Proof. First, consider this smoothed version of the absolute value

$$g_\epsilon(x) = \begin{cases} |x| & \text{if } |x| \geq \epsilon, \\ \frac{1}{2}\left(\epsilon + \frac{x^2}{\epsilon}\right) & \text{if } |x| < \epsilon. \end{cases} \tag{2.19}$$

Then, by Itô, we have

$$g_\epsilon(B_t) = \int_0^t g_\epsilon'(B_s)dB_s + \int_0^t 1_{(-\epsilon,\epsilon)}(B_s)ds. \tag{2.20}$$

Looking more carefully at g_ϵ', using Itô isometry and Fubini, we get

$$\mathbb{E}\left(\int_0^t g_\epsilon'(B_s)1_{(-\epsilon,\epsilon)}(B_s)dB_s\right)^2 = \frac{1}{\epsilon^2}\mathbb{E}\int_0^t B_s^2 1_{(-\epsilon,\epsilon)}(B_s)ds = \int_0^t \int_{(-\epsilon,\epsilon)} x^2\gamma_s(dx)ds \tag{2.21}$$

where γ_s is the Gaussian distribution with variance s. We can bound the above by

$$\int_0^t \int_{(-\epsilon,\epsilon)} x^2\gamma_s(dx)ds \leq \int_0^t \epsilon^2\gamma_s((-\epsilon,\epsilon))ds \leq \epsilon^2\int_0^t \frac{2\epsilon}{\sqrt{2\pi s}}ds = c\epsilon^3 \tag{2.22}$$

for some constant $c > 0$. Therefore, we have that

$$\int_0^t g_\epsilon'(B_s)dB_s \longrightarrow 0 \tag{2.23}$$

in L^2 as $\epsilon \to 0$. On the other hand, $g_\epsilon'(x) = \operatorname{sgn}(x)$ for $x \in \mathbb{R} \setminus (-\epsilon,\epsilon)$. Hence, taking $\epsilon \to 0$ gives the desired formula. □

We will take the theorem below for granted, which will be lightly discussed in the next subsection as it is related to the martingale problem.

Lemma 2.4. *An Itô process $dY_t = v_t dB_t$ with initial condition $Y_0 = 0$ is a Brownian motion if and only if $v_t^2 = 1$ almost surely for almost every t.*

Proposition 2.5. *Tanaka's equation* (2.16) *admits no strong solution.*

Proof. We do this by way of contradiction. Suppose X is a strong solution of (2.16) adapted to the filtration of some Brownian motion B. Then, by Lemma 2.4, we know that X is actually be Brownian motion. Moreover, by Itô's formula, we know that

$$dB_t = \operatorname{sgn}(X_t)dX_t, \tag{2.24}$$

which we know the solution—via Tanaka's formula—to be

$$B_t = \int_0^t \operatorname{sgn}(X_s)dX_s = |X_t| - L_t \tag{2.25}$$

with L_t being the local time of X. This means that the Brownian motion B is adapted to the filtration $\sigma\{|X_s| : s \leq t\}$, which is a contradiction. □

2.3 Weak solution and the martingale problem

Now that we've seen the complications that can occur, let's make precise what we mean by solutions.

Definition 2.6 (Solution of SDEs [10, Definition 8.2])**.** A solution to the stochastic differential equation (2.2) is:

- a (complete) filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$,
- a $\mathbb{F}$-Brownian motion B,

- a $\mathbb{F}$-adapted process X with continuous sample paths satisfying (2.3).

Moreover, we say that there is

- weak existence if a solution (in the above sense) exists for every initial condition $X_0 = x$.
- strong existence if the solution is adapted to the complete canonical filtration of B.
- weak uniqueness if all weak solutions have the same law.
- pathwise uniqueness if whenever the probability space and Brownian motion is fixed, solutions of the SDE are indistinguishable.

Proposition 2.7. *Tanaka's equation (2.16) admits a unique weak solution.*

Proof. Let the solution X be a Brownian motion (cf. Lemma 2.4). Define B as

$$B_t = \int_0^t \operatorname{sgn}(X_t) dX_t \quad (dB_t = \operatorname{sgn}(X_t) dX_t). \tag{2.26}$$

Then, it follows that

$$dX_t = \operatorname{sgn}(B_t) dB_t \tag{2.27}$$

and X is a weak solution. Weak uniqueness follows from uniqueless of the law of Brownian motion. □

There is actually a generic way to determining whether an SDE admits a weak solution. First, let's make an observation.

Proposition 2.8. *Let X_t be a solution to the SDE*

$$X_t = X_0 + \int_0^t b(X_s) ds + \sigma(X_s) dB_s. \tag{2.28}$$

Then, for any $f \in C_0^2(\mathbb{R})$, the process

$$f(X_t) - f(X_0) - \int_0^t Af(X_s) ds \tag{2.29}$$

is a martingale, where $Af(x) = b(x)\partial_x f(x) + \frac{1}{2}\sigma^2(x)\partial_{xx}^2 f(x)$ is the generator of the SDE.

Proof. By Itô's formula, we know that

$$f(X_t) = f(X_0) + \int_0^t \left(b(X_s)\partial_x f(X_s) + \frac{1}{2}\sigma^2(x)\partial_{xx}^2 f(X_s) \right) ds + \int_0^t \sigma(X_s)\partial_x f(X_s) dB_s. \tag{2.30}$$

The conclusion then follows from Itô integrals being martingale. □

The clever observation that two great probabilists—Daniel Stroock and Srinivasa Varadhan—made is that the converse can be used to fully characterize weak solutions of SDEs. First, we state the martingale problem.

Definition 2.9. Let $\mathcal{L}$ be the operator of the form $\mathcal{L} = b\partial_x + a\partial_{xx}^2$ where coefficients b and a are locally bounded functions. We say the probability measure $\mathbb{P}^y$ on $(C[0,\infty), \mathcal{B}_{C[0,\infty)})$ solves the martingale problem if $\mathbb{P}^y(x_0 = y) = 1$ and the process

$$f(x_t) - f(x_0) - \int_0^t \mathcal{L}f(x_s) ds \tag{2.31}$$

is a martingale with respect to the filtration $\mathcal{B}_{C[0,t]}$. Moreover, the martingale problem is well-posed if $\mathbb{P}$ is unique.

Theorem 2.10 (Weak solutions and the martingale problem [14, Theorem 20.1])**.** *If the probability measure $\mathbb{P}^y$ solves the martingale problem for coefficients (b, σ^2), then there exists a weak solution for the corresponding SDE whose law is $\mathbb{P}^y$.*

As we've seen, existence of weak solutions does not require Lipschitz coefficients. This is the same more generally for the martingale problem. In fact, Stroock and Varadhan proved that linear growth in addition to some continuity and positive definiteness is enough to gaurantee the well-posedness of the martingale problem [15].

3 Markovian properties of diffusion

In this section, we view solutions to stochastic equations as in (2.2) from the perspective of the more general theory of Markov processes. The reference for general Markov processes (unless specified otherwise) is Le Gall [10, Chapter 6].

3.1 General Markov processes For this subsection, we will put ourselves in a general measure space $(E, \mathcal{E})$. Qualifiers will be added as we go.

Definition 3.1. A collection $(Q_t)_{t\geq 0}$ of stochastic kernels is a transition semigroup if it satisfies:

1. (initial condition) $Q_0(x, dy) = \delta_x(dy)$,
2. (Chapman-Kolmogorov) $Q_{t+s}(x, A) = \int Q_s(y, A)Q_t(x, dy)$,
3. (joint measurability) for every $A \in \mathcal{E}$, the map $(t, x) \mapsto Q_t(x, A)$ is measurable with respect to $\mathcal{B}_{\mathbb{R}^+} \otimes \mathcal{E}$.

Moreover, we say X is a Markov process with semigroup $(Q_t)_{t\geq 0}$ with respect to filtration $\{\mathcal{F}_t\}_{t\geq 0}$ is an $\{\mathcal{F}_t\}_{t\geq 0}$-adapted process such that for any bounded measurable f,

$$\mathbb{E}[f(X_{s+t})|\mathcal{F}_s] = Q_t f(X_s) := \int f(y)Q_t(X_s, dy). \tag{3.1}$$

Intuitively, the Markov property says that we can restart the process whenever and the statistical behavior remains the same. We formalize this intuition below, which would be useful in drawing analogous results for strongly Markov processes.

Theorem 3.2 (Simple Markov property)**.** *Let Y be a Markov process with semigroup $(Q_t)_{t\geq 0}$ with respect to filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Moreover, assume that Y has cádlág sample paths. Then, for any measurable $g : \mathbb{D}(E) \to \mathbb{R}$, we have*

$$\mathbb{E}[g((Y_{t+s})_{t\geq 0})|\mathcal{F}_s] = \mathbb{E}^{Y_s}[g] \tag{3.2}$$

where $\mathbb{E}^y$ is the expectation taken with respect to $\mathrm{Law}(Y^y)$*, the Markov process Y with initial conditions $Y_0 = y$.*

Remark 3.3 (Sample path properties of Markov processes)**.** The space $\mathbb{D}(E)$ above denotes the space of cádlág (right-continuous with left-limits) functions taking values in E equipped with the smallest σ-algebra such that coordinate maps are measurable. In the remainder of the subsection, we will not discuss any sample path properties of Markov processes as we a priori know that diffusion processes (the main interest of the note) has continuous sample paths. In general, we know that Feller processes (cf. Definition 3.6) there is always a modification of Y with cádlág sample paths, along with some modification of the filtration to ensure completeness and right-continuity. For details, see [10, Chapter 6.3].

Proof of Theorem 3.2. From a monotone class argument, we claim that it is sufficient to consider $g = 1_A$ for A depending on a finite number of coordinates, i.e.,

$$A = \{f \in \mathbb{D}(E) : f(t_1) \in B_1, \ldots, f(t_p) \in B_p\} \tag{3.3}$$

for $0 \leq t_1 < \cdots < t_p$ and $\{B_i\}_{i=1}^p \subset E$ that are measurable. We want to show that

$$\mathbb{E}[g(Y_{t+s})|\mathcal{F}_s] = \mathbb{P}[Y_{s+t_1} \in B_1, \ldots, Y_{s+t_p} \in B_p|\mathcal{F}_s] \tag{3.4}$$

$$= \int_{B_1}\int_{B_2}\cdots\int_{B_p} Q_{t_p-t_{p-1}}(y_{p-1}, dy_p)\ldots Q_{t_2-t_1}(y_1, dy_2)Q_{t_1}(Y_s, dy_1). \tag{3.5}$$

Let $\phi_i = 1_{B_i}$ for $i = 1, \ldots, p$. Notice that, by the tower property,

$$\mathbb{E}[\phi_1(Y_{s+t_1})\phi_2(Y_{s+t_2})\cdots\phi_p(Y_{s+t_p})|\mathcal{F}_s] = \mathbb{E}[\phi_1(Y_{s+t_1})\phi_2(Y_{s+t_2})\cdots\mathbb{E}[\phi_p(Y_{s+t_p})|\mathcal{F}_{s+t_{p-1}}]|\mathcal{F}_s] \tag{3.6}$$

$$= \mathbb{E}[\phi_1(Y_{s+t_1})\phi_2(Y_{s+t_2})\cdots Q_{t_p-t_{p-1}}\phi_p(Y_{s+t_{p-1}})|\mathcal{F}_s] \tag{3.7}$$

from which we conclude the proof via induction. □

Alternatively, we can take a functional analytic spin. We can view Q_t as a operator mapping elements from $B(E)$, the set of bounded measurable function equipped with the uniform norm, back to $B(E)$. Then, we can define another useful object called the resolvent.

Definition 3.4. Let $\lambda > 0$. Then, the λ-resolvent of the semigroup $(Q_t)_{t\geq 0}$ is the linear operator $R_\lambda : B(E) \to B(E)$ defined pointwise as

$$R_\lambda f(x) = \int_0^\infty e^{-\lambda t} Q_t f(x) dt. \tag{3.8}$$

The resolvent satisfies a new nice analytic, algebraic, and probabilistic properties. Two of which we will use later; the other is presented for curiosity sake and to draw potential connections with potential theory of martingales.

Proposition 3.5 (Properties of the resolvent operator). *Let $\lambda, \mu > 0$ and $(Q_t)_{t\geq 0}$ be a transition semigroup. Then,*

1. *(boundedness) for any $f \in B(E)$, $\|R_\lambda f\| \leq \lambda^{-1}\|f\|$,*
2. *(resolvent equation) $R_\lambda - R_\mu + (\lambda - \mu)R_\lambda R_\mu$,*
3. *(constructing supermartingales) let X be a Markov process with semigroup $(Q_t)_{t\geq 0}$ with respect to filtration $\{\mathcal{F}_t\}_{t\geq 0}$ and $h \in B(E)$ be non-negative, then $e^{-\lambda t} R_\lambda h(X_t)$ is an $\{\mathcal{F}_t\}_{t\geq 0}$-supermartingale.*

Proof. We prove each item in the order listed.

1. From the fact that $\|Q_t\| \leq 1$, we can bound

$$|R_\lambda f(x)| = \left| \int_0^\infty e^{-\lambda t} Q_t f(x) dt \right| \leq \|f\| \int_0^\infty e^{-\lambda t} dt, \tag{3.9}$$

   and the statement follows.

2. The result follows from several applications of Fubini and change-in-variable:

$$R_\lambda R_\mu f(x) = \int_0^\infty e^{-\lambda s} Q_s \left(\int_0^\infty e^{-\mu t} Q_t f dt \right)(x) ds \tag{3.10}$$

$$= \int_0^\infty e^{-\lambda s} \int_0^\infty e^{-\mu t} Q_{t+s} f(x) dt ds \tag{3.11}$$

$$= \int_0^\infty e^{-(\lambda-\mu)s} \int_0^\infty e^{-\mu(t+s)} Q_{t+s} f(x) dt ds \tag{3.12}$$

$$= \int_0^\infty e^{-(\lambda-\mu)s} \int_s^\infty e^{-\mu t} Q_t f(x) dt ds \tag{3.13}$$

$$= \int_0^\infty e^{-\mu t} Q_t f(x) \int_0^t e^{-(\lambda-\mu)s} ds dt \tag{3.14}$$

$$= \int_0^\infty Q_t f(x) \left(\frac{e^{-\mu t} - e^{-\lambda t}}{\lambda - \mu} \right) dt. \tag{3.15}$$

3. Notice that the process $e^{-\lambda t} R_\lambda h(X_t)$ is uniformly bounded in time; hence, integrable. Therefore, by Fubini, we have that for any $s \geq 0$,

$$e^{-\lambda s} Q_s R_\lambda h = \int_0^\infty e^{-\lambda(t+s)} Q_{s+t} h dt = \int_s^\infty e^{-\lambda t} Q_t h dt \leq R_\lambda h. \tag{3.16}$$

   Finally, for any $t, s \geq 0$, noticing that

$$\mathbb{E}[e^{-\lambda(t+s)} R_\lambda h(X_{t+s}) | \mathcal{F}_t] = e^{-\lambda(t+s)} Q_s R_\lambda h(X_t) \tag{3.17}$$

   and using the preceeding inequality yields the supermartingale property.

□

For the remainder of the notes, we're interested in a particular type of Markov processes with additional structure. In particular, we want to ask for particular continuity in time.

Definition 3.6. Let E be a metrizable, locally compact topological space that is countable at infinity. Let $C_0(E)$ denote the space of continuous functions that vanish at infinity, equipped with the uniform norm. Then, semigroup $(Q_t)_{t\geq 0}$ is a Feller semigroup if

1. (closure) for any $f \in C_0(E)$, $Q_t f \in C_0(E)$,
2. (continuity) for any $f \in C_0(E)$, $\lim_{t\to 0} \|Q_t f - f\| \to 0$.

The Markov process associated with a Feller semigroup is called a Feller process. For each Feller semigroup, we associate the generator $L : D(L) \subset C_0(E) \to C_0(E)$ defined as

$$Lf = \lim_{t\to 0} \frac{Q_t f - f}{t} \tag{3.18}$$

for any $f \in D(L)$, where the domain is defined to be wherever the limit exists.

Remark 3.7. For this note, Feller processes will **always** take values in a locally compact space countable at infinity. We will now drop this qualifier when talking about Feller processes.

Remark 3.8. There are a few different definitions of Feller processes. Some more generally let the operators act on $C_b(E)$, the space of bounded continuous functions equipped with the uniform norm. Some others might require the continuity condition to be with respect to the norm topology as oppose to the strong operator topology. We follow here the standards of [10, Chapter 6.2] and [13, Chapter 7.1].

We can immediately derive two simple identities.

Proposition 3.9. *Let $f \in D(L)$ and $s > 0$. Then, $Q_s f \in D(L)$ and $LQ_s f = Q_s Lf$.*

Proof. Since Q_s is continuous ($\|Q_s\| \leq 1$), we get that

$$\frac{Q_t(Q_s f) - Q_s f}{t} = Q_s \left(\frac{Q_t f - f}{t}\right) \to Q_s Lf \tag{3.19}$$

as $t \to 0$. □

Proposition 3.10. *If $f \in D(L)$, then for every $t \geq 0$, we have*

$$Q_t f = f + \int_0^t Q_s Lf ds. \tag{3.20}$$

Proof. We "differentiate" and get

$$\frac{Q_{t+\epsilon} f - Q_t f}{\epsilon} = Q_t \left(\frac{Q_\epsilon f - f}{\epsilon}\right) \to Q_t Lf \tag{3.21}$$

as $\epsilon \to 0$. □

Using these two identities, we can relate the resolvent operator and the generator.

Theorem 3.11. *Let $\lambda > 0$. Then, $R_\lambda = (\lambda I - L)^{-1}$ in the sense of:*

1. *for every $g \in C_0(E)$, $R_\lambda g \in D(L)$ and $(\lambda I - L)R_\lambda g = g$,*
2. *if $f \in D(L)$, $R_\lambda(\lambda I - L)f = f$,*
3. *$D(L) = \operatorname{range} R_\lambda$, which is dense in $C_0(E)$.*

Proof. We prove each item in the order listed.

1. Let $g \in C_0(E)$ and $\epsilon > 0$. By Fibini and a change-in-variable,

$$\frac{Q_\epsilon(R_\lambda g) - g}{\epsilon} = \frac{1}{\epsilon}\left(\int_0^\infty e^{-\lambda t} Q_{\epsilon+t} g dt - \int_0^\infty e^{-\lambda t} Q_t g dt\right) \tag{3.22}$$

$$= \frac{1}{\epsilon}\left((1 - e^{-\lambda\epsilon})\int_0^\infty e^{-\lambda t} Q_{\epsilon+t} g dt - \int_0^\epsilon e^{-\lambda t} Q_t g dt\right). \tag{3.23}$$

By the continuity property of Feller semigroups, we can take the limit as $\epsilon \to 0$ and get

$$LR_\lambda g = \lim_{\epsilon\to 0} \frac{Q_\epsilon(R_\lambda g) - g}{\epsilon} = \lambda R_\lambda g - g. \tag{3.24}$$

2. Let $f \in D(L)$ and $x \in E$. Then, by the previous proposition, we know that

$$\int_0^\infty e^{-\lambda t} Q_t f(x) dt = \frac{1}{\lambda} f(x) + \int_0^\infty e^{-\lambda t} \int_0^t Q_s L f(x) ds dt. \tag{3.25}$$

Applying Fubini, the above expression becomes

$$\frac{1}{\lambda} f(x) + \int_0^\infty Q_s L f(x) \int_s^\infty e^{-\lambda t} dt ds = \frac{1}{\lambda} f(x) + \int_0^\infty \frac{e^{-\lambda s}}{\lambda} Q_s L f(x) ds. \tag{3.26}$$

Therefore, we get that

$$\lambda R_\lambda f = f + R_\lambda L f. \tag{3.27}$$

3. The fact that $D(L) = \text{range}\, R_\lambda$ is apparent from the inverse relation derived. It is left to show that the range of R_λ is dense. First, we note that for any $\lambda, \mu > 0$, $\text{range}\, R_\lambda = \text{range}\, R_\mu$ because the resolvent equation gives

$$R_\lambda f = R_\mu(f + (\mu - \lambda) R_\lambda f. \tag{3.28}$$

From here, for every $f \in C_0(E)$, we have

$$\lambda R_\lambda f = \lambda \int_0^\infty e^{-\lambda t} Q_t f dt = \int_0^\infty e^{-t} Q_{t/\lambda} f dt \to f \tag{3.29}$$

as $\lambda \to \infty$ where the convergence is due to the continuity property of Feller semigroups and dominated convergence.

□

In fact, the connection bewteen the generator and the resolvent operator is a much deeper theory in functional analysis and PDE considering existence and uniqueness of evolution equations.The Hille-Yosida theorem gives a complete characterization of the two objects—any given Feller (or "strongly continuous"/"C_0" in the functional analytic world) semigroup corresponds to a generator, and every generator generates a unique semigroup through the evolution equations.

Theorem 3.12 (Hille-Yosida [17, Section 7] or [3, Theorem 7.8]). *Let $\mathcal{B}$ be a Banach space, and let operator $L : D(L) \subset \mathcal{B} \to \mathcal{B}$ be an unbounded linear operator. Moreover, assume that $D(L)$ is dense and for every $\lambda > 0$, $\lambda I - L$ is bijective with $\|\lambda I - L\| \le 1$. Then, for any $u_0 \in D(L)$, there is a unique $u : [0, \infty) \to \mathcal{B}$ such that it solves the evolution equation*

$$\frac{du}{dt} = Lu, \quad u(0) = u_0. \tag{3.30}$$

Moreover, the collection of maps $Q_t : u_0 \mapsto u(t)$ is a strongly continuous semigroup. Conversely, given strongly continuous semigroups $(Q_t)_{t\ge 0}$, there exists an operator L satisfying the properties above that solves the evolution equation.

Remark 3.13. There are different versions of Hille-Yosida, some more relevant to the theory of Markov processes than another. The standard reference on this is Ethier and Kurtz [6, Theorem 1.6/Theorem 2.2], which phrased the density of $D(L)$ and invertibility of $\lambda I - L$ as L being *dissipative* or as satisfying certain maximum principle respectively.

3.2 Strong Markov properties of SDEs

We will try to prove the strong Markov property of SDEs two ways. The first uses the particular fact that it is a solution to a differential equation of the form (2.3). The second will start by establishing strong Markov property for Feller processes and we will proceed by showing that the generator and semigroup associated with solutions to SDEs driven by Brownian motion is Feller.

We begin with the first approach.

Theorem 3.14 (Strong Markov property of diffusions). *Let $(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a (complete) filtered space and let X be the unique strong solution to the homogeneous stochastic differential equation*

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t. \tag{3.31}$$

Then, X is a Markov process with respect to $\mathbb{F}$ with semigroup

$$Q_t f(x) = \mathbb{E} f(X_t^x) \tag{3.32}$$

for any $f \in B(\mathbb{R}^d)$ where X^x is any solution to the same SDE with initial condition $X_0^x = x$. Moreover, X is strongly Markov, i.e., for any $\mathbb{F}$-stopping time τ and bounded measureable $g : C(\mathbb{R}^d; \mathbb{R}) \to \mathbb{R}$,

$$\mathbb{E}[1_{\tau<\infty} g((X_{\tau+t})_{t\geq 0})|\mathcal{F}_\tau] = 1_{\tau<\infty} \mathbb{E}^{X_\tau}[g] \tag{3.33}$$

where $\mathbb{E}^x$ is the expectation taken with respect to the measure $\mathrm{Law}(X^x)$.

Remark 3.15. In fact, we do not need the existence of strong solutions. The crucial construction is a map from the desired Brownian motion to the solution of the SDE once we're given the probability space for which the solution lives in. Such map obviously exists when a strong solution is available. However, the existence of such map is more subtle for weak solutions and we refer the reader to [10, Theorem 8.5].

Proof of Theorem 3.14. We want to first show that

$$\mathbb{E}[1_{\tau<\infty} f(X_{\tau+t})|\mathcal{F}_\tau] = 1_{\tau<\infty} Q_t f(X_\tau). \tag{3.34}$$

From here, we will assume that $\tau < \infty$ almost surely for notational simplicity. This does not affect the subsequent pathwise arguments. Notice that

$$X_{\tau+t} = X_\tau + \int_\tau^{\tau+t} b(X_s)ds + \int_\tau^{\tau+t} \sigma(X_s)dB_s. \tag{3.35}$$

Now, define the shifted version of each quantity:

$$\tilde{X}_t = X_{\tau+t}, \quad \tilde{\mathcal{F}}_t = \mathcal{F}_{\tau+t}, \quad \tilde{B}_t = B_{\tau+t} - B_\tau; \tag{3.36}$$

it is worth noting that $\tilde{B}_t$ is still a Brownian motion by its strong Markov property. It follows that $\tilde{X}$ is adapted to $\{\tilde{\mathcal{F}}_t\}_{t\geq 0}$ and

$$\int_\tau^{t+\tau} b(X_s)ds = \int_0^t b(\tilde{X}_s)ds, \quad \int_\tau^{t+\tau} \sigma(X_s)dB_s = \int_0^t \sigma(\tilde{X}_s)d\tilde{B}_s \tag{3.37}$$

where the latter can be argued via approximation (definition of Itô integral). Therefore, $\tilde{X}$ is a strong solution to the same SDE in the "shifted" probability space $(\Omega, \mathcal{F}, \{\tilde{\mathcal{F}}_t\}_{t\geq 0}, \mathbb{P})$ with respect to Brownian motion $\tilde{B}$ and initial condition X_τ. In fact, there is a map $h_{X_\tau} : \tilde{B} \mapsto \tilde{X}$. This means that, for every $t \geq 0$,

$$\mathbb{E}[1_{\tau<\infty} f(X_{\tau+t})|\mathcal{F}_\tau] = \mathbb{E}[1_{\tau<\infty} f(\tilde{X}_t)|\mathcal{F}_\tau] = \mathbb{E}[1_{\tau<\infty} f(h_{X_\tau}(\tilde{B})_t)|\mathcal{F}_\tau] \tag{3.38}$$

$$= 1_{\tau<\infty} \int f(h_{X_\tau}(w)_t) W(dw) \tag{3.39}$$

$$= 1_{\tau<\infty} Q_t f(X_\tau) \tag{3.40}$$

where the third equality is by Fubini and the fact that $\tilde{B}$ is independent of $\mathcal{F}_\tau$ with law W, which is the Wiener measure. The form claimed in (3.33) follows from induction argument as in the proof of Theorem 3.2.

Lastly, we check that $(Q_t)_{t\geq 0}$ is a semigroup.

1. By construction, $Q_0 f(x) = \mathbb{E} f(X_0^x) = x$ for any f.
2. By the tower property,

$$Q_{t+s} f(x) = \mathbb{E} f(X_{s+t}^x) = \mathbb{E}\,\mathbb{E}[f(X_{s+t}^x)|\mathcal{F}_t] = \mathbb{E}\, Q_s f(X_t^x) = \int Q_s f(y) Q_t(x, dy). \tag{3.41}$$

3. The map $(x, t) \mapsto Q_t f(x)$ is measurable is by the continuity of $x \mapsto \mathrm{Law}(X^x)$, cf. [10, Theorem 8.5].

Needless to say, the Markov property follows from the strong Markov property. □

From the proof above, we did not learn much about the structure of the semigroup; rather, we guessed that it is Markovian and it turned out alright. The proposition below digs deeper into the structure of the process and its analytic niceties.

Proposition 3.16 (Itô diffusions are Feller)**.** *The semigroup $(Q_t)_{t\geq 0}$ defined in Theorem 3.14 is Feller with generator $L : D(L) \supset C_0^2(\mathbb{R}^d) \to C_0(\mathbb{R}^d)$ satisfying*

$$Lf(x) = b(x) \cdot \nabla f + \frac{1}{2} \operatorname{tr}[\sigma\sigma^\intercal \nabla^2 f]. \tag{3.42}$$

Proof (assuming b and σ are bounded). The form and domain of the generator follows from an application of Itô's formula and the martingale property of the Itô integral (cf. Theorem 3.18). So, we focus on proving that Q_t is Feller. Fix $f \in C_0(\mathbb{R}^d)$. We first check that $Q_t f \in C_0(\mathbb{R}^d)$. For some $0 < A < \infty$, we can decompose

$$\limsup_{x\to\infty} |Q_t f(x)| = \limsup_{x\to\infty} |\mathbb{E} f(X_t^x)| \leq \limsup_{x\to\infty} |\mathbb{E} f(X_t^x) 1_{|X_t^x - x| \leq A}| + \|f\| \mathbb{P}(|X_t^x - x| > A) \tag{3.43}$$

$$\leq \|f\| \sup_{x\in\mathbb{R}^d} \mathbb{P}(|X_t^x - x| > A). \tag{3.44}$$

However, by boundedness of the coefficients,

$$\mathbb{P}(|X_t^x - x| > A) \leq \frac{\mathbb{E}|X_t^x - x|^2}{A^2} \to 0 \tag{3.45}$$

as $A \to \infty$. Thus, we have that $Q_t f \in C_0(\mathbb{R}^d)$.

Now, we want to show continuity, i.e., $Q_t f \to f$ as $t \to 0$. For any $\epsilon > 0$, we have

$$\limsup_{t\to 0} \sup_{x\in\mathbb{R}^d} |\mathbb{E} f(X_t^x) - f(x)| \leq \limsup_{t\to 0} \sup_{x\in\mathbb{R}^d} |\mathbb{E}(f(X_t^x) - f(x)) 1_{|X_t^x - x|\leq\epsilon}| + 2\|f\| \mathbb{P}(|X_t^x - x| > \epsilon) \tag{3.46}$$

$$\leq \sup_{x,y\in\mathbb{R}^d : |x-y|\leq\epsilon} |f(x) - f(y)| \tag{3.47}$$

where the probability of X_t^x deviating away from x vanishes again by the boundedness of the coefficients and Chebyshev inequality. Moreover, the whole expression vanishes as $\epsilon \to 0$ by the continuity of f. Thus, $(Q_t)_{t\geq 0}$ is a Feller semigroup. □

The point of showing that we're working with Feller processes is that strong Markov property immediately follows (with some modifications to the filtration so that it is right-continuous). The proof, which will not be presented here, is not unfamiliar at all. We argue via approximation, partitioning the time axis and letting the random time take value in some vanishing interval. The continuity properties of Feller semigroups make sure that the limits can be passed through.

Theorem 3.17 (Feller processes are strongly Markov [10, Theroem 6.17])**.** *Let Y be a Markov process with a Feller semigroup $(Q_t)_{t\geq 0}$ with respect to filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Moreover, assume that Y has cádlág sample paths and let τ be a stopping time with respect to the filtration $\{\mathcal{F}_{t^+}\}_{t\geq 0}$. Then, for any measurable $g : \mathbb{D}(E) \to \mathbb{R}$, we have*

$$\mathbb{E}[1_{\tau<\infty} g((Y_{t+\tau})_{t\geq 0}) | \mathcal{F}_\tau] = 1_{\tau<\infty} \mathbb{E}^{Y_\tau}[g]. \tag{3.48}$$

3.3 Martingales and the Feynman-Kac formula

We've seen, in the case of diffusions, that the generator can be used to find us a martingale. The same form actually holds more generally for the Markov processes.

Theorem 3.18 (Generators and martingales). *Let $(Q_t)_{t\geq 0}$ be a Feller semigroup of Markov process X taking values in E. Moreover, let L be the corresponding generator. For $h, g \in C_0(E)$, the following are equivalent:*

1. *$h \in D(L)$ and $Lh = g$,*
2. *for every $x \in E$, the process*

$$h(X_t^x) - \int_0^t g(X_s^x)ds \tag{3.49}$$

   is a martingale.

Proof. We first show item 1 $\implies$ item 2. For any $t, s \geq 0$,

$$\mathbb{E}[h(X_{t+s}^x)|\mathcal{F}_t] = Q_s h(X_t^x) = h(X_t^x) + \int_0^s Q_r g(X_t^x)dr. \tag{3.50}$$

Moreover, by Fubini,

$$\mathbb{E}\left[\int_t^{t+s} g(X_r^x)dr|\mathcal{F}_t\right] = \int_t^{t+s} \mathbb{E}\left[g(X_r^x)|\mathcal{F}_t\right]dr = \int_t^{t+s} Q_{r-t}g(X_t^x)dr = \int_0^s Q_r g(X_t^x)dr. \tag{3.51}$$

Combining the two gives the martingale property:

$$\mathbb{E}\left[h(X_{t+s}^x) - \int_0^{t+s} g(X_r^x)dr|\mathcal{F}_t\right] = h(X_t^x) - \int_0^t g(X_r^x)dr. \tag{3.52}$$

We now show item 2 $\implies$ item 1. Since martingales have constant expectation, for all $t \geq 0$,

$$h(x) = \mathbb{E}\left[h(X_t^x) - \int_0^t g(X_s^x)ds\right]. \tag{3.53}$$

But X is also a Markov process:

$$\mathbb{E}\left[h(X_t^x) - \int_0^t g(X_s^x)ds\right] = Q_t h(x) - \int_0^t Q_s h(x)ds. \tag{3.54}$$

Therefore, putting the two together

$$\frac{Q_t h(x) - h(x)}{t} = \frac{1}{t}\int_0^t Q_s h(x)ds \to h(x) \tag{3.55}$$

by the fact that $(Q_t)_{t\geq 0}$ is Feller. Thus, $h \in D(L)$ and $Lh = g$. □

From here, we immediately get the "Kolmogorov backward equation," which is a weaker version of Dynkin's formula. However, from the same line of reasoning, we also get Dynkin's formula for free via optional sampling theorem. We reproduce the two identities below.

Corollary 3.19 (Kolmogorov & Dynkin). *Let X be a Feller process with generator L. Then, for any $f \in C_0^2(\mathbb{R}^d)$,*

1. *(Kolmogorov's backward equation) the function $u(t, x) = \mathbb{E} f(X_t^x)$ satisfies*

$$\partial_t u = Lu, \quad u(0, x) = f(x). \tag{3.56}$$

2. *(Dynkin's formula) for any integrable stopping time τ, we have*

$$\mathbb{E} f(X_\tau^x) = \mathbb{E}\left[\int_0^\tau Lf(X_s^x)ds\right]. \tag{3.57}$$

So far, we've been converting the probabilistic question to an analytic one. We can do the converse, too. In particular, let L be the generator of an Itô diffusion and consider

$$\partial_t v = Lv - qv, \quad u(0,x) = f(x). \tag{3.58}$$

Can we find a process whose expectation with respect to some function correspond to this PDE?

Example 3.20 (Killed Diffusions). Let X^x be an Itô diffusion in $\mathbb{R}^d$ starting at $x \in \mathbb{R}^d$. We will extend X, denoted by $\tilde{X}^x$, to take values in $\mathbb{R}^d \cup \{☠\}$ where ☠ is the "dead state"; once entered the dead state—or killed—$\tilde{X}^x$ cannot leave. Note that this is not a numerical value. We will kill $\tilde{X}$ at a random time: let τ be exponentially distributed with unit rate independent of $\tilde{X}$, then the killing time ζ is defined to be

$$\zeta = \inf\left\{t \geq 0 : \int_0^t q(\tilde{X}_s^x)ds \leq \tau\right\} \tag{3.59}$$

for some $q \in C(\mathbb{R}^d)$. To summarize, the new process $\tilde{X}^x$ is such that

$$\tilde{X}_t^x = \begin{cases} X_t^x & \text{if } t < \zeta, \\ ☠ & \text{if } t \geq \zeta. \end{cases} \tag{3.60}$$

Let $\{\mathcal{F}_t^x\}_{t\geq 0}$ denote the natural filtration generated by X^x. Now, since we know the distribution of τ, we can write down the distribution of ζ:

$$\mathbb{P}(\zeta > t|\mathcal{F}_\infty^x) = \mathbb{P}\left(\tau \geq \int_0^t q(\tilde{X}_s^x)ds|\mathcal{F}_\infty^x\right) = \exp\left(-\int_0^t q(\tilde{X}_s^x)ds\right) := Z_t^x. \tag{3.61}$$

Moreover, by the Markov property of X (which is inherited by $\tilde{X}$ prior to getting killed), we have

$$\mathbb{P}(\zeta > t+s|\mathcal{F}_t^x) = q(t,x)\,\mathbb{E}\left[\exp\left(-\int_t^{t+s} q(\tilde{X}_s^x)ds\right)|\mathcal{F}_t^x\right] \tag{3.62}$$

$$= Z_t^x\,\mathbb{E}^{\tilde{X}_t^x}\left[\exp\left(-\int_0^s q(x)ds\right)\right]. \tag{3.63}$$

Then, for any $f \in C_0^2(\mathbb{R}^d)$ (with the extension that $f(☠) = 0$), we can see that

$$\mathbb{E}\,f(X_t^x)Z_t^x = \mathbb{E}[f(X_t^x)\mathbb{P}(\zeta > t|\mathcal{F}_\infty^x)] = \mathbb{E}\,f(X_t^x)1_{\zeta>t} = \mathbb{E}\,f(\tilde{X}_t^x). \tag{3.64}$$

Lastly, if we define $v \in C^{1,2}(\mathbb{R}\times\mathbb{R}^d)$ be such that

$$v(t,x) = \mathbb{E}\left[f(X_t^x)Z_t^x\right] = \mathbb{E}\,f(\tilde{X}_t^x), \tag{3.65}$$

it can be shown that v is the unique solution (within some appropriate class of functions) of the PDE (3.58). This is one of many applications (perhaps the less popular) motivations for studying the Feynman-Kac formula.

Theorem 3.21 (Feynman-Kac [11, Theorem 8.2.1]). *Let $f \in C_0^2(\mathbb{R}^d)$ and $q \in C(\mathbb{R}^d)$ be bounded below. Then, letting*

$$v(t,x) = \mathbb{E}\left[f(X_t^x)\exp\left(-\int_0^t q(X_s^x)ds\right)\right], \tag{3.66}$$

then v solves the differential equation

$$\partial_t v = Lv + qv, \quad v(t,x) = f(x). \tag{3.67}$$

Moreover, if $w \in C^{1,2}(\mathbb{R}\times\mathbb{R}^d)$ is a solution to the above equation and is bounded on $K\times\mathbb{R}^n$ for any compact $K \subset \mathbb{R}$, then $w = v$.

4 Girsanov's theorem and consequences

The goal of this section is to demonstrate how the laws of diffusion processes can be related to one another and to motivate change-in-measures as a useful tool in analyzing systems involving diffusions. We will give two examples in-depth: the first concerns with partial large deviation results in the small-noise limit, the second generalizes Bayes' theorem to derive filters for partially-observed diffusions. However, perhaps the most notable application is in finanace where Girsanov's theorem appears as yet another fundamental theorem.

Theorem 4.1 (Fundamental theorem of asset pricing [12, Theorem 2.15]). *A market is arbitrage-free if and only if there exists one equivalent martingale measure.*

We start with an intuitive derivation of the type of results we will be studying. Consider the space of continuous functions on the unit intercal $C([0,1];\mathbb{R})$ and a discretized (perhaps linearly interpolated so that it takes values in $C([0,1];\mathbb{R})$) Brownian motion with grid size δ, which has law

$$\mathbb{P}(B_\delta \in E_1, B_{2\delta} \in E_2, \dots, B_1 \in E_n) = \int_{E_1 \times E_2 \cdots \times E_n} (2\pi\delta)^{-n/2} \exp\left(-\frac{1}{2\delta}\sum_{i=1}^n |x_i - x_{i-1}|^2\right) dx_1 dx_2 \dots dx_n \tag{4.1}$$

$$\int_{E_1 \times E_2 \cdots \times E_n} (2\pi\delta)^{-n/2} \exp\left(-\frac{\delta}{2}\sum_{i=1}^n \left|\frac{x_i - x_{i-1}}{\delta}\right|^2\right) dx_1 dx_2 \dots dx_n \tag{4.2}$$

where $n = 1/\delta$. Now, we take $\delta \to 0$ ($n \to \infty$) to try to write an expression down for the law of Brownian motion of path space, or the Wiener measure $\mathbb{W}$. Notice that the constant term in front converges to one, the sum in the exponent converges to an integral, and the summand converges to the time derivative of the path. Therefore, we "may" write

$$\mathbb{W}(dx) = \exp\left(-\frac{1}{2}\int_0^1 |\dot{x}_t|^2 dt\right) dx = \exp\left(-\frac{1}{2} I(x)\right) dx \tag{4.3}$$

where dx is the Lebesgue measure on $C([0,1];\mathbb{R})$, $\dot{x}$ denotes the time-derivative of the path $x \in C([0,1];\mathbb{R})$, and I we say is the "action functional."

Remark 4.2 (Two contradictions lead to a truth?). The above formulation is wrong in two folds: first, Lebesgue measure on infinite-dimensional spaces do not exist. Moreover, the action function is not well-defined as we know a posteriori that Brownian motion paths are almost surely not differentiable. But perhaps like how product of negative numbers yield a positive one, two false statements can lead to a somewhat-true one.

We're interested in the law of the Brownian motion after shifting by some function h; denote this new measure as $\mathbb{Q}$. In particular, we want to find the Radon-Nikodym derivative such that

$$\frac{d\mathbb{Q}}{d\mathbb{W}}(x)\mathbb{W}(dx) = \exp\left(-\frac{1}{2}\int_0^1 |(x \dot{-} h)_t|^2 dt\right) dx = d\mathbb{Q}(dx). \tag{4.4}$$

But this is easy—we just complete the squares! It turns out that

$$\frac{d\mathbb{Q}}{d\mathbb{W}}(x) = \exp\left(\int_0^1 \dot{h}_t \dot{x}_t dt - \int_0^1 |\dot{h}_t|^2 dt\right) = \exp\left(\int_0^1 \dot{h}_t dx_t - \int_0^1 |\dot{h}_t|^2 dt\right) \tag{4.5}$$

where the first term on the exponential we interpret as the Itô integral. This is the Cameron-Martin formula, which we state more formally below and the proof will follow straightforwardly in the next subsection.

Proposition 4.3 (Cameron-Martin). *Let $h : \mathbb{R}^+ \to \mathbb{R}$ be such that there exists $\dot{h}$ satisfying $h_t = \int_0^t \dot{h}_s ds$ and $\dot{h} \in L^2(\mathbb{R}^+, dt)$. Then, for every nonnegative measurable $g \in C(\mathbb{R}^+;\mathbb{R})$, we have*

$$\int g(x+h)\mathbb{W}(dx) = \int \exp\left(\int_0^\infty \dot{h}_t dx_t - \int_0^\infty |\dot{h}_t|^2 dt\right) g(x)\mathbb{W}(dx). \tag{4.6}$$

Remark 4.4 (Cameron-Martin space and translational invariance)**.** The space of h such that the conditions given above is satisfied is called the Cameron-Martin space. It is in fact a Hilbert space equipped with the inner product

$$\langle f, h \rangle = \int_0^\infty \dot{f}_t \dot{h}_t dt, \tag{4.7}$$

and the space is dense in $C(\mathbb{R}^+; \mathbb{R})$. Due to the lack of translationally-invariant measures in infinite-dimensional spaces, the Wiener measure can be thought of as a good alternative to Lebesgue measures via the Cameron-Martin formula/space. The theory of Gaussian measures and Cameron-Martin space (and related spaces, e.g., the reproducing kernel Hilbert space) reaches far beyond the classical stochastic differential equations story, cf. Hairer [8] for a shorter introduction and Bogachev [2] for a comprehensive treatment.

4.1 Change-in-measure via exponential martingales

While Girsanov's theorem can be proved straightforwardly via Itô's formula and utilizing some characterization theorem, we make an attempt of highlighting the martingale part of the story. We begin with a reminder of a simple result about a special Doob's martingale.

Proposition 4.5 (Radon-Nikodym derivatives as a UI martingale)**.** *Consider a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ *and a measure* $\mathbb{Q}$ *absolutely continuous with respect to* $\mathbb{P}$*. Then, the process* D *defined as*

$$D_t = \frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_{\mathcal{F}_t} \tag{4.8}$$

is a uniformly integrable martingale. Moreover, if the two measures are equivalent, i.e., $\mathbb{P} \ll \mathbb{Q}$ *as well, then* D *is strictly positive.*

Define the map $\mathcal{E}$ such that

$$\mathcal{E}(L) = \exp\left(L - \frac{1}{2}\langle L, L\rangle\right) \tag{4.9}$$

where L is a martingale of the form

$$L_t = \int_0^t b_s dB_s \text{ and } \langle L, L\rangle_t = \int_0^t b_s^2 ds. \tag{4.10}$$

For now, b is taken to be adapted; we will make the integrability conditions more explicit in the theorem statements.

Remark 4.6. The brackets denote the quadratic variation process, or in the case when the arguments are different, the covariation process; more simply, they are called the bracket process. For two martingales X and Y, the bracket process $\langle X, Y\rangle$ is the unique process such that $X_t Y_t - \langle X, Y\rangle_t$ is a martingale. In our case when dealing with diffusions, the bracket process is the "dt" term after taking the product XY. In particular, for Itô processes

$$X = \int_0^\cdot b_s^X ds + \int_0^\cdot \sigma_s^X B_s \tag{4.11}$$

and Y defined similarly, the bracket is

$$\langle X, Y\rangle_t = \int_0^t \sigma_s^X \sigma_s^Y ds. \tag{4.12}$$

Proposition 4.7. *Let* D *be a martingale adapted to the natural filtration of some Brownian motion* B*. Then, there exists an Itô process* L *such that* $D = \mathcal{E}(L)$*. Moreover,* L *takes the form*

$$L_t = \log D_0 + \int_0^t D_s^{-1} dD_s = \log D_0 + \int_0^t b_s D_s^{-1} dB_s. \tag{4.13}$$

Proof. By the martingale representation theorem, we know that there must be a b such that $D = \int_0^\cdot b_s dB_s$. Now, by Itô's formula, we get that

$$\log D_t = \log D_0 + \int_0^t \frac{dD_s}{D_s} - \frac{1}{2}\int_0^t \frac{d\langle D, D\rangle_s}{D_s^2} = \log D_0 + \int_0^t \frac{b_s dB_s}{D_s} - \frac{1}{2}\int_0^t \frac{b_s^2 ds}{D_s^2}. \tag{4.14}$$

Taking L to be of the form claimed verifies that $D = \mathcal{E}(L)$.

□

Remark 4.8 (Local martingales)**.** The above proposition, as well as what is about to proceed, can be generalized. In particular, we can drop the condition that D is adapted to a filtration of Brownian motion and replace it with the general notion of a local martingale—M is a local martingale if there exists a sequence of stopping times $\tau_n \uparrow \infty$ such that the stopped process $M_{t\wedge\tau}$ is a uniformly integrable martingale. However, we will not have Itô integrals appearing and the Lebesgue integration term will be replaced by integration with respect to the quadratic variation process $\langle D, D\rangle$. It happens that we know the quadratic variation process of Itô processes.

Girsanov's theorem identifies a process as Brownian motion. The proposition below states that if the quadratic variation of the process is t, then it is a Brownian motion. This will crucially play a role in the proof.

Proposition 4.9 (Lévy's characterization of Brownian motion)**.** *Let B be an adapted (to some filtration) process with continuous sample paths. Then, the following are equivalent:*

1. *B is a d-dimensional Brownian motion,*
2. *the components B_i's are martingales and $B_iB_j - \delta_{ij}t$ are martingales.*

Proof. We know item 1 implying item 2, so we want to show the converse. The full proof can be found in [9, Theorem 3.3.16] or [10, Theorem 5.12], we highlight some key points below.

1. For any ξ, we know that $\xi \cdot B$ and $(\xi \cdot B)^2 - |\xi|^2 t$ are martingales.
2. By the previous proposition, we know that $\exp(i\xi \cdot B + |\xi|^2 t/2)$ is a martingale.
3. We can then deduce, using characteristic functions, that B has Gaussian increments and increments are independent via a conditioning argument.

□

We are now ready to prove Girsanov's theorem. First, we give a version that uses the tools that we have, which seems slightly awkward as we have avoided generality. To compensate, we will then give a more useful version for diffusion processes after.

Theorem 4.10 (Girsanov)**.** *Let $\mathbb{P}$ and $\mathbb{Q}$ be mutually absolutely continuous. Define the process $D_t = d\mathbb{Q}/d\mathbb{P}|_{\mathcal{F}_t}$ be a martingale adapted to the natural filtration of some Brownian motion B and let L be the process such that $D = \mathcal{E}(L)$. Then,*

$$\tilde{B} = B - \langle B, L\rangle \tag{4.15}$$

is a Brownian motion under $\mathbb{Q}$.

Proof. By Itô's formula (with $f(x, y) = xy$), we have

$$\tilde{B}_t D_t = \int_0^t \tilde{B}_s dD_s + \int_0^t D_s d\tilde{B}_s + \langle D, \tilde{B}\rangle_t \tag{4.16}$$

$$= \int_0^t \tilde{B}_s dD_s + \int_0^t D_s dB_s + \int_0^t D_s d\langle B, L\rangle_s + \langle D, B\rangle_t \tag{4.17}$$

$$= \int_0^t \tilde{B}_s dD_s + \int_0^t D_s dB_s \tag{4.18}$$

where the last equality uses the fact that $dL_s = D_s^{-1} dD_s$. This shows that $\tilde{B}D$ is a martingale in $\mathbb{P}$, which implies that $\tilde{B}$ is a martingale in $\mathbb{Q}$ because for any $A \in \mathcal{F}_s$,

$$\mathbb{E}^{\mathbb{P}}\, 1_A \tilde{B}_t D_t = \mathbb{E}^{\mathbb{P}}\, 1_A \tilde{B}_t D_s = \mathbb{E}^{\mathbb{Q}}\, 1_A \tilde{B}_t. \tag{4.19}$$

Moreover, $\langle \tilde{B}, \tilde{B}\rangle = \langle B, B\rangle$ regardless of the underlying measure since $\mathbb{P}$ and $\mathbb{Q}$ are equivalent and the bracket process can be seen as the probabilistic limit of the sum of pointwise product (cf. [10, Proposition 4.21]). Thus, by Lévy's characterization of Brownian motion, $\tilde{B}$ is a Brownian motion under $\mathbb{Q}$. □

Theorem 4.11 (Girsanov). *Let b be an adapted process such that for $L = \int_0^\cdot b_s dB_s$, the process $\mathcal{E}(L)$ is a uniformly integrable martingale (cf. Proposition 4.13). Then, under the measure $\mathbb{Q}$ defined by the Radon-Nikodym derivative $d\mathbb{Q}/d\mathbb{P} = \mathcal{E}(L)$, the process*

$$\tilde{B}_t = B_t - \int_0^t b_s ds \tag{4.20}$$

is a Brownian motion.

Proof. The proof follows from the first version of Girsanov with $D = \mathcal{E}(L)$. Uniform integrability is used so that $\mathbb{Q} = D_\infty \mathbb{P}$ is a well-defined Radon-Nikodym derivative by martinagle convergence. □

Remark 4.12 (Constructing weak solutions to SDEs). Let's look at the second Girsanov's theorem more carefully. We are given a Brownian motion B and we found a new Brownian motion (under a different measure) where

$$B_t = \int_0^t b_s ds + \tilde{B}_t. \tag{4.21}$$

Now, consider the stochastic differential equation

$$dX_t = b(t, X_t)dt + \tilde{B}_t. \tag{4.22}$$

Then, Girsanov tells us that taking $X = B$ and taking the Brownian motion to be $\tilde{B}$, X is a solution to the SDE under measure $\mathbb{Q}$. In other words, we have found a weak solution X on the probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ and Brownian motion $\tilde{B}$. Putting in on canonical space, we have that

$$\frac{d\mathbb{Q}B^{-1}}{d\mathbb{P}B^{-1}}(x) = \frac{d\mathrm{Law}(\textsc{sde})}{d\mathrm{Law}(\textsc{bm})}(x) = \mathcal{E}\left(\int_0^\cdot b_s dx_s\right). \tag{4.23}$$

The crucial assumption here is that D (or $\mathcal{E}(L)$) is a uniformly integrable martingale. While $\mathcal{E}(L)$ is always a local martingale if L is a local martingale [10, Proposition 5.11], to do a change-in-measure, D must be a true martingale. The benefit of the first statement of Girsanov's theorem is that this is automatically true. Nonetheless, the second version is much more applicable. Below, we give conditions on the drift b so that D is a true (uniform) martingale.

Proposition 4.13 (Novikov's criterion). *Suppose $L = \int_0^\cdot b_s dB_s$ with b such that*

$$\mathbb{E}\exp\left(\frac{1}{2}\langle L, L\rangle_\infty\right) = \mathbb{E}\exp\left(\frac{1}{2}\int_0^\infty b_s^2 ds\right) < \infty, \tag{4.24}$$

then $\mathcal{E}(L)$ is a uniformly integrable martingale.

Remark 4.14 (Uniformly-integrable or not). Uniform integrability above is mainly to make sure that the $\mathcal{E}(L)$ is valid at $t = \infty$. If we only care about a finite time-window $[0, T]$ for some $T < \infty$, it is sufficient to change all the ∞ to T. Moreover, uniform integrability is a non-issue for martingales on a finite-time interval. Nonetheless, Novikov's condition guarantees that $\mathcal{E}(L)$ is a true martingale.

Proof of Proposition 4.13. The argument requires some lengthy estimates and clever uses of stopping times. We refer the reader to [10, Theorem 5.23] for a proof for the general case and [16, Theorem 4.5.8] for the finite time-horizon case. □

Remark 4.15 (Weak solutions continued). From the previous remark, for the finite time-horizon case, we can take any bounded, measurable b and a weak solution exists (no regularity required)! The weak solution can be shown to be unique in law as well.

We conclude this subsection with a yet another useful version of Girsanov's theorem.

Corollary 4.16 (Girsanov). *Let X be the Itô process*

$$X_t = X_0 + \int_0^t b_s^X ds + \int_0^t \sigma_s dB_s. \tag{4.25}$$

where σ_s is invertible for all s. Moreover, suppose $\sigma^{-1}(b^Y - b^X)$ satisfy Novikov's condition and define the measure $\mathbb{Q}$ via the Radon-Nikodym derivative

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}\left(\int_0^\cdot \sigma_s^{-1}(b_s^Y - b_s^X)dB_s\right). \tag{4.26}$$

Then, under $\mathbb{Q}$, the process

$$\tilde{B}_t = B_t - \int_0^t \sigma_s^{-1}(b_s^Y - b_s^X)ds \tag{4.27}$$

is a Brownian motion and X is a (weak) solution to the stochastic differential equation

$$dY_t = b_t^Y dt + \sigma_t d\tilde{B}_t. \tag{4.28}$$

Proof. The fact that $\tilde{B}$ is a Brownian motion under $\mathbb{Q}$ follows from the previous statements of Girsanov's theorem. We only need to check that X (driven by $\tilde{B}$) under $\mathbb{Q}$ has the same law as Y:

$$dX_t = b_t^X dt + \sigma_t dB_t = b_t^X dt + \sigma_t d(\tilde{B}_t + \sigma_t^{-1}(b_t^Y - b_t^X)ds) = b_t^Y dt + \sigma_t d\tilde{B}_t. \tag{4.29}$$

□

4.2 The action functional and large deviations To motivate the usefulness of Girsanov theorem, we take a detour to the theory of large deviations. In particular, let B be a Brownian motion, we're concerned with the convergence:

$$\lim_{\epsilon \to 0} \epsilon B \to 0 \tag{4.30}$$

in the space of continuous functions $C([0,1];\mathbb{R})$ equipped with the uniform metric. More specifically, we're interested in the rate of convergence,

$$\mathbb{P}(\epsilon B \in B_h(\delta)) \approx e^{-\epsilon^{-2}I(h)} \tag{4.31}$$

for some ball $B_h(\delta)$ and rate I that depends on the choice of $h \in C([0,1];\mathbb{R})$. This is the classic Freidlin-Wentzell small-noise limit of Brownian motion, which can be generalized to the convergence of SDEs to the corresponding ODE [7]. In this case, the rate function I is the "action functional" from before:

$$I(h) = \begin{cases} \frac{1}{2}\int_0^1 \dot{h}_t^2 dt & \text{if } h \in \mathcal{H}, \\ +\infty & \text{otherwise} \end{cases} \tag{4.32}$$

where $\mathcal{H}$ is the Cameron-Martin space with elements such that $h_0 = 0$. This restriction is necessary for the sanity check that $I(h) = 0$ if and only if $h \equiv 0$.

The proof of large deviation principles are usually separated into upper and lower bounds as they require different techniques. We give the full proof of the lower bound (often the less obvious one) below as it applies Girsanov's theorem.

Proposition 4.17 (Large deviation lower bound). *For any $\delta > 0$ and $h \in \mathcal{H}$, we have*

$$\liminf_{\epsilon \to 0} \epsilon^2 \log \mathbb{P}(\epsilon B \in B_\delta(h)) \geq -I(h). \tag{4.33}$$

Proof. Using the fact that $\{\epsilon B \in B_\delta(h)\} = \{B \in B_{\delta\epsilon^{-1}}(\epsilon^{-1}h)\}$, by Girsanov/Carmeron-Martin, we have that

$$\mathbb{P}(\epsilon B \in B_\delta(h)) = \int_{B_{\delta\epsilon^{-1}}(0)} \exp\left(-\epsilon^{-1}\int_0^T \dot{h}_t dx_t - \frac{1}{2\epsilon^2}\int_0^T |\dot{h}_t|^2 dt\right)\mathbb{W}(dx) \tag{4.34}$$

$$= \exp\left(-\epsilon^{-2}I(h)\right)\int_{B_{\delta\epsilon^{-1}}(0)} \exp\left(-\epsilon^{-1}\int_0^T \dot{h}_t dx_t\right)\mathbb{W}(dx) \tag{4.35}$$

where $\mathbb{W}$ is the Wiener measure. First, observe that we can choose ϵ such that $\mathbb{W}(B_{\delta\epsilon^{-1}}(0))$ is anything we want; we will take this to be at least $3/4$. Then, by Chebyshev, we see that

$$\mathbb{W}\left(\int_0^T \dot{h}_t dx_t \geq 2\sqrt{2I(h)}\right) \leq \frac{\mathbb{E}\int_0^T \dot{h}_t^2 dt}{8I(h)} = \frac{1}{4}, \tag{4.36}$$

this means that exponentiating both sides gives

$$\mathbb{W}\left(\exp\left(-\epsilon^{-1}\int_0^T \dot{h}_t dx_t\right) \geq \exp\left(-2\epsilon^{-1}\sqrt{2I(h)}\right)\right) \geq \frac{3}{4}. \tag{4.37}$$

Let A be the event above. Combining these two bounds, we can get that

$$\begin{aligned}
\mathbb{P}(\epsilon B \in B_\delta(h)) &\geq \exp\left(-\epsilon^{-2}I(h)\right)\int_{B_{\delta\epsilon^{-1}}(0)\cap A} \exp\left(-\epsilon^{-1}\int_0^T \dot{h}_t dx_t\right)\mathbb{W}(dx) && (4.38)\\
&\geq \exp\left(-\epsilon^{-2}I(h) - 2\epsilon^{-1}\sqrt{2I(h)}\right)\mathbb{W}(B_{\delta\epsilon^{-1}}(0)\cap A) && (4.39)\\
&\geq \frac{1}{2}\exp\left(-\epsilon^{-2}I(h) - 2\epsilon^{-1}\sqrt{2I(h)}\right) && (4.40)
\end{aligned}$$

for small enough ϵ. The statement then follows from taking the logarithm and applying the ϵ^2 scaling (the second term on the exponent vanishes in the limit). □

The upper bound proceeds via approximation, which we will not prove. Interested readers can consult [13, Lemma VIII.2.10]. For completeness, we state the final large deviation principle.

Theorem 4.18 (Wentzell-Friedlin). *For any Borel set A in $C([0,T];\mathbb{R})$, we have*

$$-\inf_{f\in \operatorname{int} A} I(f) \leq \liminf_{\epsilon\to 0}\epsilon^2\log\mathbb{P}(\epsilon B\in B_\delta(h)) \leq \limsup_{\epsilon\to 0}\epsilon^2\log\mathbb{P}(\epsilon B\in B_\delta(h)) \leq -\inf_{f\in \operatorname{cl} A} I(f). \tag{4.41}$$

Remark 4.19 (Exponential tilts in large deviation theory)**.** For those familiar with large deviation principles, the proof technique (on a high level) is not unfamiliar. The typical proof of Cramér's theorem uses an exponential tilt to make the rare event—empirical mean being away from the true mean—not rare; the rate is then extrapolated from the tilt factor. The same idea occurred here: we studied the rare event that ϵB is away from zero by tilting it via Girsanov to a form amendable to analysis.

Remark 4.20 (Most of C space is empty)**.** This large deviation principle makes precise the intuition that most of C space is empty (not attained by Brownian motion). Rather, it clusters around neighborhood of elements in the Carmeron-Martin space, which is much more regular than a "typical" continuous path.

The Freidlin-Wentzell theory follows the classical recipe for proving large deviation principles. A more modern theory due to Dupuis and Ellis [4, 5] utilizes techniques from weak convergence and variational representations. However, Girsanov did not become obsolete in this new paradigm. Rather, due to the intricate connection of variational forms, relative entropy and weak convergence, Girsanov plays an important role in the analysis nonetheless.

Recall that for two measure μ and ν, we define the relative entropy as

$$\mathcal{R}(\nu\|\mu) = \begin{cases}\int \log\frac{d\nu}{d\mu}d\nu & \text{if } \nu \ll \mu\\ +\infty & \text{otherwise .}\end{cases} \tag{4.42}$$

The following proposition represents exponential functionals—a.k.a. moment generating function or partition function in statistical physics—as a problem of minimizing expectation while paying a relative entropy cost.

Proposition 4.21 (Donsker-Varadhan)**.** *Let $\mathcal{X}$ be a Polish space and $\mathcal{P}(\mathcal{X})$ be the set of probability measures on $\mathcal{X}$. Then, for any bounded, measurable $f:\mathcal{X}\to\mathbb{R}$ and measure $\mu\in\mathcal{P}(\mathcal{X})$,*

$$-\log\int e^{-f(x)}\mu(dx) = \inf_{\nu\in\mathcal{P}(\mathcal{X})}\left\{\int f(x)\nu(dx) + \mathcal{R}(\nu\|\mu)\right\}. \tag{4.43}$$

Proof. First, notice that if we let ν^* be the exponential tilt, i.e.,

$$\nu^*(dx) = \frac{1}{Z} e^{-f(x)} \mu(dx) \quad \text{where} \quad Z = \int e^{-f(x)} \mu(dx), \tag{4.44}$$

then the right-hand side becomes

$$\int f(x)\nu^*(dx) + \mathcal{R}(\nu^* \| \mu) = \int f(x)\nu^*(dx) - \int f(x)\nu^*(dx) - \log Z = -\log Z. \tag{4.45}$$

Now, by Jensen, we see that for any $\nu \ll \mu$,

$$-\log \int e^{-f(x)} \mu(dx) = -\log \int e^{-f(x)} \frac{d\nu}{d\mu}(x)\nu(dx) \le \int f(x)\nu(dx) + \mathcal{R}(\nu\|\mu) \tag{4.46}$$

and the equality is achieved at ν^*. □

Theorem 4.22 (Boué-Dupuis)**.** *Let $f : C([0,1];\mathbb{R})$ be a bounded and measurable, and let B be a standard Brownian motion. Then,*

$$-\log \mathbb{E}\, e^{-f(B)} = \inf_{v \in \mathcal{A}} \left\{ \mathbb{E} f\left(B + \int_0^\cdot v_s ds\right) + \frac{1}{2}\int_0^1 |v_s|^2 ds \right\} \tag{4.47}$$

where $\mathcal{A}$ is the set of progressively-measurable process with $E \int_0^1 |v_s|^2 ds < \infty$.

Proof. We provide just the idea of the proof; interested reader can consult the original paper [5, Theorem 3.1] or the reference book [4, Theorem 8.3].

First, we are in the setting where Donsker-Varadhan (Proposition 4.21) applies. Just applying the variational formula gives

$$-\log \mathbb{E}\, e^{-f(B)} = \inf_{\mathbb{Q}} \left\{ \mathbb{E}^{\mathbb{Q}} f(B) + \mathcal{R}(\mathbb{Q}\|\mathbb{P}) \right\}. \tag{4.48}$$

where $\mathbb{Q}$ is some other measure on the probability space. However, we know that the only measures that are absolutely continuous to $\mathbb{P}$ are those that shift the Brownian motion by elements in the Cameron-Martin space. Therefore, to have finite relative entropy cost, we can restrict ourselves to these measures without loss of generality. In particular, we let $\mathbb{Q}$ be such that

$$\frac{d\mathbb{Q}}{d\mathbb{W}}(x) = \exp\left(\int_0^1 v_s dx_s - \frac{1}{2}\int_0^1 |v_s|^2 ds \right). \tag{4.49}$$

for some $v \in \mathcal{H}$.

Now, we parse through the terms inside the infimum. By Girsanov, we know that there is $\tilde{B} = B - \int_0^\cdot v_s ds$ that is a $\mathbb{Q}$-Brownian motion. Therefore,

$$\mathbb{E}^{\mathbb{Q}} f(B) = \mathbb{E}^{\mathbb{Q}} f\left(\tilde{B} + \int_0^\cdot v_s ds\right) \tag{4.50}$$

and

$$\mathcal{R}(\mathbb{Q}\|\mathbb{P}) = \mathbb{E}^{\mathbb{Q}} \log \exp\left(\int_0^1 v_s dB_s - \frac{1}{2}\int_0^1 |v_s|^2 ds \right) = \mathbb{E}^{\mathbb{Q}} \int_0^1 v_s d\tilde{B}_s + \frac{1}{2}\int_0^1 |v_s|^2 ds = \frac{1}{2}\mathbb{E}^{\mathbb{Q}} \int_0^1 |v_s|^2 ds. \tag{4.51}$$

since the Itô integral has zero mean. The only to show now is that

$$\inf_v \left\{ \mathbb{E}^{\mathbb{Q}} f\left(\tilde{B} + \int_0^\cdot v_s ds\right) + \frac{1}{2}\int_0^1 |v_s|^2 ds \right\} \overset{?}{=} \inf_v \left\{ \mathbb{E} f\left(B + \int_0^\cdot v_s ds\right) + \frac{1}{2}\int_0^1 |v_s|^2 ds \right\}, \tag{4.52}$$

and this is in fact non-trivial and requires careful approximations. □

This variational representation turned out indispensible for the proof of small-noise large deviation principle via weak convergence. It also found its way to many other areas of probability as it gives a control-theoretic interpretation to a static yet important quantity. In terms of the main point of the note, we showed that relative entropy works well with Girsanov as the logarithm and exponential cancel. This enables easier entropy estimates that turn out to be useful in other areas of probability as well.

5 Long-time behavior of Markov processes

Suppose we have a discrete-time finite-state Markov chain $(X_t)_{t\in\mathbb{N}} \in \mathcal{X}$. Then, we know that the semigroup can be identified with a matrix $P : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Then, we know that there exists a stationary distribution $\pi \in \mathcal{P}(\mathcal{X})$ such that

$$\pi P = \int P(x, \cdot)\pi(dx) = \pi. \tag{5.1}$$

Moreover, for irreducible (can get to any state) and aperiodic (doesn't go in circles), then $\mathrm{Law}(X_t) \to \pi$ as $t \to \infty$ for any initial conditions. Even better, if we let P be diagonalizable, then we know that

$$d_{\mathrm{TV}}(\mathrm{Law}(X_t), \pi) \le \lambda_1^t d_{\mathrm{TV}}(\mathrm{Law}(X_0), \pi) \tag{5.2}$$

where λ_1 is the second largest eigenvalue of P. We would like to establish similar results for continuous-time Markov chains, that is, existence, uniqueness, and convergence rate towards the stationary distribution.

For this section, we work with a diffusion process X with drift and diffusion coefficients b and σ respectively. As we're working with laws, we only require weak existence of solutions, though more regularity might be implicit in the conditions we state.

5.1 Fokker-Planck and stationarity

Deriving criteria for when stationary distribution exists is actually quite simple. First, by stationarity, we're seeking a measure $\pi \in \mathcal{P}(\mathbb{R}^d)$ such that for any bounded f, if we initialize $X_0 \sim \pi$,

$$\mathbb{E} f(X_t) = \mathbb{E}\,\mathbb{E}[f(X_t)|X_0] = \int \mathbb{E}[f(X_t)|X_0 = x]\pi(dx) = \mathbb{E} f(X_0). \tag{5.3}$$

For reasons we'll see soon, let's temporarily take $f \in C_0^2(\mathbb{R}^d)$. Rearranging the above and adding a factor of t^{-1} gives

$$\frac{1}{t}\int \mathbb{E}[f(X_t)|X_0 = x] - f(x)\pi(dx) = 0. \tag{5.4}$$

Lastly, since f is bounded, we can taking the limit as $t \to 0$ to get

$$\int Lf(x)\pi(dx) = 0 \tag{5.5}$$

where L is the generator of X that takes the form

$$Lf = \sum_{i=1}^d b_i \partial_i f + \frac{1}{2}\sum_{i,j=1}^d \sigma_{ij}\partial_{ij} f = b \cdot \nabla f + \mathrm{tr}[\sigma\sigma^\intercal \nabla^2 f]. \tag{5.6}$$

Definition 5.1. We say π is a stationary distribution of the diffusion X if it satisfies

$$\int Lf(x)\pi(dx) = 0$$

for all $f \in C_0^2(\mathbb{R}^d)$.

Now, consider $L^2 = L^2(\mathbb{R}^d; dx)$ equipped with the Lebesgue measure and the standard inner product. If π admits a density—which, for notational simplicity, we will identify $\frac{d\pi}{dx} \mapsto \pi$—then we can alternatively write the condition for the stationary distribution as

$$\int Lf(x)\pi(dx) = \langle Lf, \pi\rangle = \langle f, L^*\pi\rangle \quad \Longrightarrow \quad L^*\pi = 0 \tag{5.7}$$

where L^* denotes the adjoint of the operator L when viewed as a(n unbounded) linear operator in L^2. The implication holds as the above relation must hold for a dense set of f.

The running example of this section will be the Langevin dynamics, which are solutions to the SDE

$$dX_t = -\nabla U(X_t)dt + \sqrt{\frac{2}{\beta}}dB_t \tag{5.8}$$

where $U : \mathbb{R}^d \to \mathbb{R}$ is some potential (we will add more qualifiers later) and $\beta > 0$ is the inverse temperature. From what we know thus far, we can write the generator as

$$Lf = -\nabla U \cdot \nabla f + \beta^{-1}\Delta f. \tag{5.9}$$

The special property of Langevin diffusions is that the stationary distribution takes the form of a Gibbs measure. We can try to calculate this via the adjoint relation:

$$L^*\pi = \nabla \cdot (\pi \nabla U) + \beta^{-1}\Delta\pi = 0. \tag{5.10}$$

Now, we do some clever manipulations.

$$\beta^{-1}\Delta\pi + \nabla \cdot (\pi\nabla V) = \nabla \cdot (\beta^{-1}\nabla\pi + \pi\nabla V) = \nabla \cdot (\pi(\beta^{-1}\nabla \log \pi + \nabla V)) \tag{5.11}$$

Plugging in $\log \pi = -\beta V + \text{const}$ gives zero, which means that

$$\pi(dx) \propto \exp(-\beta V(x))\, dx. \tag{5.12}$$

The concept of reversibility becomes important in the following two subsections. Broadly speaking, a Markov process is reversible if, once in stationary, the dynamics forward in time is the same as the dynamics backward in time.

Definition 5.2. A Markov process is said to be reversible if the associated semigroup P_t is self-adjoint as a bounded linear operator in $L^2(\pi)$. Moreover, we will say π is a reversible measure.

Remark 5.3 (Non-equilibrium steady-state)**.** In statistical physics, we say a system is at equilibrium when the process is at stationary and the stationary distribution takes the form of a Gibbs distribution. When the stationary distribution is not Gibbs, physicists say that the system is in non-equilibrium steady-state. Terrible terminology, but they make the rules...

In light of the remark, one can check (via integration-by-parts and using the explicit form of π) that Langevin diffusions are always reversible.

We will close this subsection by talking about how the law $\mu_{t,x} := \text{Law}(X_t|X_0 = x)$ evolves in time.

Proposition 5.4 (Fokker-Planck)**.** *Suppose $\mu_{t,x}$ admist a density with respect to the Lebesgue measure for all $t \geq 0, x \in \mathbb{R}^d$. Then, the density obeys*

$$\partial_t \mu_{t,x} = L^*\mu_t. \tag{5.13}$$

Proof. First, we know that for any $f \in C_0^2(\mathbb{R}^d)$, we have

$$\partial_t \int f(y)\mu_{t,\cdot}dy = \partial_t P_t f = P_t L f \tag{5.14}$$

where P_t is the semigroup associated with X/L. Moreover, for every $x \in \mathbb{R}^d$,

$$P_t Lf(x) = \int Lf(y)P_t(x, dy) = \int Lf(y)\mu_{t,x}(y)dy = \int f(y)L^*\mu_{t,x}(y)dy. \tag{5.15}$$

Again, since f is dense in L^2, the above equality implies that $L^*\mu_{t,x} = \partial_t\mu_{t,x}$. □

5.2 Convergence via the log-Sobolev inequality

The existence and uniqueness questions turns out to be questions about the null space of an unbounded operator. The convergence question is a bit more subtle. There are certainly many approach, but we will take a popular one. The statements we will prove will be of the form:

$$\text{if [operator] satisfies [famous name] inequality} \implies \text{[fluctuation] decreases exponentially.} \tag{5.16}$$

First, we need to establish some more operators related to a Markov process beyond the semigroup and generator.

Definition 5.5. Consider a Markov process with generator L. Let $\mathcal{A} \subset D(L)$ such that $fg \in \mathcal{A}$ for $f, g \in D(L)$. Then, the carré du champ operator $\Gamma : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ is a bilinear operator defined as

$$\Gamma(f,g) = \frac{1}{2}\left[L(fg) - fLg - gLf\right]. \tag{5.17}$$

Moreover, we define the Dirichlet energy $\mathcal{E}(f,g) = \int \Gamma(f,g)d\pi$ where π is the stationary distribution of the Markov process.

Remark 5.6 (Square of the field)**.** In French, *carré du champ* means "square of the field," which came from taking L to be the generator of the Langevin diffusion:

$$\Gamma(f,f) = \frac{1}{2}\left[\Delta(f^2) - \nabla U \cdot \nabla(f^2) - 2f(\Delta f - \nabla U \cdot \nabla f)\right] = \|\nabla f\|^2. \tag{5.18}$$

More generally, we have that $\Gamma(f,g) = \nabla f \cdot \nabla g$. We write $\mathcal{E}(f) = \mathcal{E}(f,f)$.

It turns out that under reversibility, the carré du champ operator behaves nicely. We will exploit the properties listed below to define and analyze important operators to come.

Proposition 5.7 (Properties of carré du champ)**.** *Let L be the generator of some Markov process with stationary measure π. Then,*

1. *(positivity)* $\Gamma(f,f) \geq 0$,
2. *(integration-by-parts) if π is reversible, then*
$$\int \Gamma(f,g)d\pi = -\int fLgd\pi. \tag{5.19}$$
3. *(positivity) if π is reversible, $-L$ is a positive operator, i.e., $\langle f, -Lf\rangle_\pi \geq 0$ for all f.*

Proof. We prove each item in the respective order.

1. By Jensen's inequality, we have that $P_t f^2 \geq (P_t f)^2$. Therefore,
$$Lf^2 = \lim_{t\to 0}\frac{1}{t}(P_t f^2 - f^2) \geq \lim_{t\to 0}\frac{1}{t}((P_t f)^2 - f^2) = 2fLf. \tag{5.20}$$
From which, the positivity follows from the definition of Γ.
2. By π being a stationary distribution, $\int L(fg)d\pi = 0$. Then, the integration-by-parts formula simply follows from reversibility.
3. The positivity of $-L$ follows immediately from the positivity of Γ and integration-by-parts.

□

Remark 5.8 (Domain of Dirichlet energy)**.** Under reversibility, the Dirichlet energy actually admits a much simpler form than defined. This means that the domain of $\mathcal{E}$ is actually bigger than that of Γ, i.e., $D(\mathcal{E})$ is at least $D(L)\times D(L)$. In fact, we can be a bit more sophisticated. We can define

$$\mathcal{E}(f) = \lim_{t\to 0}\frac{1}{t}\int f(f - P_t f)d\pi \tag{5.21}$$

for $f \in L^2(\pi)$ and $D(\mathcal{E})$ be the set of functions for which the limit exists. It turns out this domain is even bigger and we can define $\mathcal{E}(f,g)$ via polarization identity of Hilbert spaces. For more details, see [1, Section 1.7.1].

From now on, we will identify a Markov process with the carré du champ operator Γ and stationary distribution π.

The first approach in proving ergodicity of a Markov process is via the log-Sobolev inequality. The protagonist of this functional inequality is the entropy.

Definition 5.9. Let μ be a probability measure, then the Entropy functional is defined as

$$\mathrm{Ent}_\mu[f] = \mathbb{E}_\mu f \log f + \mathbb{E}_\mu f \log(\mathbb{E}_\mu f) \tag{5.22}$$

for $f \geq 0$. Notice that if $f = d\nu/d\mu$ for some $\nu \ll \mu$, $\mathrm{Ent}_\mu[f] = \mathcal{R}(\nu\|\mu)$.

Definition 5.10. We say a Markov process (Γ, π) satisfies a (modified) log-Sobolev inequality with constants $c > 0$, written as LSI(c), if for any $f \in \mathcal{E}(D)$,

$$\mathrm{Ent}_\pi[f] \leq \frac{c}{2}\mathcal{E}(f, \log f). \tag{5.23}$$

Now, we are ready for our first convergence statement.

Theorem 5.11 (Convergence via LSI)**.** *If π satisfies a LSI(c), then for every non-negative integrable f,*

$$\mathrm{Ent}_\pi[P_t f] \leq e^{-2t/c}\,\mathrm{Ent}_\pi[f]. \tag{5.24}$$

Proof. The proof follows from differentiating the entropy functional. First, notice that $\mathbb{E}\, P_t f = \mathbb{E} f$, so the time derivative of the latter term in the entropy vanishes. Therefore, for $f \in D(\mathcal{E})$, we have

$$\frac{d}{dt}\mathrm{Ent}_\pi[P_t f] = \int (1 + \log P_t f) L P_t f \, d\pi = -\int \Gamma(P_t f, 1 + \log P_t f) d\pi = -\mathcal{E}(P_t f, \log P_t f). \tag{5.25}$$

Using the log-Sobolev inequality and Gronwall, we get the desired bound. □

In case the above representation is not clear already, this proves that the exponential convergence to π in the sense of relative entropy.

Corollary 5.12 (Convergence in relative entropy)**.** *For any initial distribution μ_0, if π satisfies a LSI(c), then*

$$\mathcal{R}(\mu_t\|\pi) \leq e^{-2t/c}\mathcal{R}(\mu_0\|\pi) \tag{5.26}$$

where μ_t is the law of the process at time t.

Of course, proving a LSI is a task in and of itself, but with the structure of Langevin diffusions, the problem usually boils down to the convexity of the potential U. We give a well-known result below.

Theorem 5.13 (Bakry-Émery)**.** *A Markov semigroup is said to satisfy the Bakry-Émery criterion with constant $\alpha > 0$ if*

$$\Gamma_2(f, f) = \frac{1}{2}[L\Gamma(f, f) - 2\gamma(f, Lf)] \geq \alpha\Gamma(f, f), \tag{5.27}$$

which implies an LSI with constant at most $1/\alpha$. In particular, a Langevin diffusion satisfies the Bakry-Émery criterion if and only the potential U is α-strongly convex, i.e.,

$$U(y) - U(x) \geq \nabla U^\intercal (y - x) + \frac{\alpha}{2}\|y - x\|^2. \tag{5.28}$$

5.3 Convergence via the Poincaré (spectral gap) inequality On the other hand, the Poincaré inequality deals with convergence in variance. In fact, the Poincaré inequality is more akin to the spectral analysis we did in the discrete-time case. We will make this clear in a second.

Definition 5.14. We say a Markov Process (Γ, π) satisfies a Poincaré inequality, written as P(c), if for all $f \in D(\mathcal{E})$, we have

$$\mathrm{Var}_\pi[f] \leq c\mathcal{E}(f). \tag{5.29}$$

Remark 5.15 (Relation with eigengaps). Suppose f is an eigenfunction of $-L$ corresponding to eigenvalue λ, and (Γ, π) satifies $\mathrm{P}(c)$. Then,

$$\mathrm{Var}_\pi[f] = \int f^2 d\pi \leq c\mathcal{E}(f) = c\int f(-Lf)d\pi = c\lambda \int f^2 d\pi, \tag{5.30}$$

and we have that non-zero eigenvalues of $-L$ satisfies $\lambda \geq 1/c$.

Now, we present a similar convergence theorem as the LSI, but in terms of the variance.

Theorem 5.16 (Convergence via P). *If π is reversible satisfies a $\mathrm{P}(c)$, then for every $f \in L^2(\pi)$, we have that*

$$\mathrm{Var}_\pi[P_t f] \leq e^{-2t/c}\mathrm{Var}_\pi[f]. \tag{5.31}$$

Proof. The same as before, we differentiate with respect to time:

$$\frac{d}{dt}\mathrm{Var}_\pi[P_t f] = 2\int P_t f L P_t f = -2\mathcal{E}(f) \tag{5.32}$$

where the second equality is by the integration-by-parts formula for the Dirichlet energy. Again, using the assumed Poincaré inequality and Gronwall yields the desired bound. □

Bibliographical notes

References

[1] Dominique Bakry, Ivan Gentil, Michel Ledoux, et al., *Analysis and geometry of markov diffusion operators*, Vol. 103, Springer, 2014.

[2] Vladimir Igorevich Bogachev, *Gaussian measures*, American Mathematical Soc., 1998.

[3] Haim Brézis, *Functional analysis, sobolev spaces and partial differential equations*, Vol. 2, Springer, 2011.

[4] Amarjit Budhiraja and Paul Dupuis, *Analysis and approximation of rare events*, Representations and Weak Convergence Methods, Series Probability Theory and Stochastic Modelling **94** (2019).

[5] Paul Dupuis and Richard S Ellis, *A weak convergence approach to the theory of large deviations*, John Wiley & Sons, 2011.

[6] Stewart N Ethier and Thomas G Kurtz, *Markov processes: characterization and convergence*, John Wiley & Sons, 2009.

[7] Mark I. Freidlin and Alexander D. Wentzell, *Random perturbations of dynamical systems*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[8] Martin Hairer, *An introduction to stochastic pdes*, arXiv preprint arXiv:0907.4178 (2009).

[9] Ioannis Karatzas and Steven Shreve, *Brownian motion and stochastic calculus*, Vol. 113, springer, 2014.

[10] Jean-François Le Gall, *Brownian motion, martingales, and stochastic calculus*, Springer, 2016.

[11] Bernt Oksendal, *Stochastic differential equations: an introduction with applications*, Springer Science & Business Media, 2013.

[12] Andrea Pascucci, *Pde and martingale methods in option pricing*, Springer Science & Business Media, 2011.

[13] Daniel Revuz and Marc Yor, *Continuous martingales and brownian motion*, Vol. 293, Springer Science & Business Media, 2013.

[14] L Chris G Rogers and David Williams, *Diffusions, markov processes and martingales: Volume 2, itô calculus*, Vol. 2, Cambridge university press, 2000.

[15] Daniel W Stroock and SR Srinivasa Varadhan, *Multidimensional diffusion processes*, Vol. 233, Springer Science & Business Media, 1997.

[16] Ramon Van Handel, *Stochastic calculus, filtering, and stochastic control*, Course notes., URL http://www. princeton. edu/rvan/acm217/ACM217. pdf **14** (2007).

[17] Kôsaku Yosida, *Functional analysis*, Vol. 123, Springer Science & Business Media, 2012.