An Introduction to Diffusion Processes

Guohuan Zhao

THE ACADEMY OF MATHEMATICS AND SYSTEMS SCIENCE OF CHINESE ACADEMY OF SCIENCES, BEIJING 100190, CHINA
Email address: gzhao@amss.ac.cn

Contents

Preface

The primary motivation for writing this lecture note is to introduce graduate students to the historical background, modern mathematical treatment, and certain analytical applications of **diffusion processes**. During the preparation, the main reference was Richard Bass's book *Diffusions and Elliptic Operators*, which provided a solid foundation and valuable insights.

Given the time constraint of a ten-week short course, it is impossible to cover all aspects of this rich subject in depth. Therefore, I have selected a number of fundamental and interesting topics and reorganized them in my own way to form the content of these notes. My hope is that this material will offer a clear and engaging entry point into the theory of diffusion processes for students encountering it for the first time.

Guohuan Zhao

CHAPTER 1

Construction of Diffusion processes I

1.1. Brownian Motion

The *Brownian motion* is a continuous stochastic process characterized by independent increments that follow a normal distribution. It is widely used to model the irregular motion of tiny particles suspended in a fluid. As one of the fundamental concepts in stochastic analysis, Brownian motion serves as a cornerstone for understanding more complicate stochastic processes.

In physics, Brownian motion was discovered in 1827 by the British botanist Robert Brown. While observing pollen particles suspended in water under a conventional microscope, he noticed their irregular motion. Since 1860, numerous scientists have studied this phenomenon and identified the following key characteristics of Brownian motion:

(1) The motion of the particles consists of translation and rotation;
(2) The movements of the particles are apparently uncorrelated, even when the particles approach each other to distances smaller than their diameters;
(3) The smaller the particles, the lower the viscosity of the liquid, or the higher the temperature, the more active the motion of the particles;
(4) The composition and density of the particles have no effect on their motion;
(5) The motion of the particles never stops.

In 1905, Einstein proposed a related theory. His theory has two parts: the first part defines the diffusion equation for Brownian particles, where the diffusion coefficient is related to the mean square displacement of the Brownian particles, and the second part describes the relationship between the diffusion coefficient and measurable physical quantities. Here we briefly introduce the first part: determining the distance a Brownian particle moves in a given time. Classical mechanics cannot determine this distance because a Brownian particle will be subjected to a large number of collisions, approximately 10^{14} collisions per second. Einstein considered the position of the particle in space at time t as a random variable X_t, and let $\rho(t, x)$ be the density of X_t. Assume τ_B is the relaxation time, and $\Delta t \gg \tau_B$. The increment $X_{t+\Delta t} - X_t$ over the time interval Δt is also a random variable, and its probability density is assumed to be $\varphi_{\Delta t}$ (depending only on Δt). For a homogeneous liquid, we can naturally assume that $\varphi_{\Delta t}$ is rotationally symmetric.

Using Taylor expansion:

$$\rho(t,x) + \partial_t\rho(t,x)\Delta t \approx \rho(t+\Delta t, x) = \int_{\mathbb{R}^3} \rho(t, x-y)\varphi_{\Delta t}(y)\mathrm{d}y$$

$$\approx \rho(t,x)\int_{\mathbb{R}^3}\varphi_{\Delta t}(y)\mathrm{d}y - \nabla\rho(t,x)\cdot\int_{\mathbb{R}^3}y\varphi_{\Delta t}(y)\mathrm{d}y$$

$$+ \frac{1}{2}\partial^2_{y_i y_j}\rho(t,x)\int_{\mathbb{R}^3}y_i y_j\varphi_{\Delta t}(y)\mathrm{d}y$$

$$= \rho(t,x) + \frac{1}{2}\int_{\mathbb{R}^3}|y|^2\varphi_{\Delta t}(y)dy\Delta\rho(t,x).$$

Therefore,

$$\partial_t\rho = \frac{\int_{\mathbb{R}^3}|y|^2\varphi_{\Delta t}(y)dy}{2\Delta t}\Delta\rho.$$

From both theoretical and experimental perspectives, it is reasonable to assume that $\nu = \frac{1}{2\Delta t}\int_{\mathbb{R}^3}|y|^2\varphi_{\Delta t}(y)dy$ is a constant, called the diffusion coefficient of the Brownian particle. Thus, the above equation can be written as:

$$\partial_t\rho = \nu\Delta\rho, \quad \rho(0,x) = f(x).$$

The solution to this heat equation is:

$$\rho(t,x) = \int_{\mathbb{R}^3}\frac{1}{(4\pi\nu t)^{3/2}}e^{-\frac{(x-y)^2}{4\nu t}}f(y)dy.$$

From this, we obtain that if $X_0 = x$, then the distribution of X_t is a standard Gaussian distribution. The second part of Einstein's theory relates the diffusion constant to physically measurable quantities, such as the mean square displacement of the particle over a given time interval. This result allows for the experimental determination of Avogadro's number and, consequently, the size of molecules. However, we will not discuss this further here.

1.1.1. Mathematical Definition of Brownian Motion. Note that Einstein did not explicitly establish a mathematical model for Brownian motion. This problem was solved by Wiener.

Definition 1.1. $(W_t)_{t\geqslant 0}$ is a stochastic process satisfying:
   (1) **Stationary independent increments and Gaussian property**: For $t > s$, the increment $W_t - W_s$ follows a normal distribution with mean 0 and variance $(t-s)I_{d\times d}$, and the increment $W_t - W_s$ is independent of the process $(W_u)_{0\leqslant u\leqslant s}$ before time s;
   (2) **Path continuity**: $(W_t)_{t\geqslant 0}$ is almost surely continuous;
Usually, we assume $W_0 = 0$, in which case, W is called standard Brownian motion.

Of course, a natural mathematical question is whether such a stochastic process exists.

A stochastic process defined on $(\Omega, \mathcal{F}, \mathbf{P})$ taking value in a measurable space $(E, \mathcal{E})$ can be understood in various ways. It involves a collection of random variables $X_t \in E$ indexed by a parameter set $\mathbf{T}$ (usually, $\mathbf{T} = \mathbb{N}$ or $\mathbb{R}_+$), where X_t is a measurable map from $(\Omega, \mathcal{F}, \mathbf{P})$ to $(E, \mathcal{E})$ for each $t \in \mathbf{T}$. The parameter set $\mathbf{T}$ typically represents time and can be discrete or continuous. The process can also be regard as a measurable map from $(\Omega, \mathcal{F}, \mathbf{P})$ to the space of functions

$E^{\mathbf{T}}$. The Kolmogorov σ-field on $E^{\mathbf{T}}$ is the smallest σ-field making the projections $\pi_t : E^{\mathbf{T}} \ni f \mapsto f(t) \in E$ measurable. This definition ensures that a random map $\Omega \ni \omega \mapsto X_.(\omega) \in E^{\mathbf{T}}$ is measurable if its component random variables $X_t : \Omega \to E$ are measurable for all $t \in \mathbf{T}$. Therefore, the mapping $\omega \mapsto X_.(\omega)$ induces a measure on $(E^{\mathbf{T}}, \mathcal{E}^{\mathbf{T}})$ denoted by $\mathbb{P}$. The underlying probability model $(\Omega, \mathcal{F}, \mathbf{P})$ is replaceable by the canonical model $(\mathbb{P}, E^{\mathbf{T}}, \mathcal{E}^{\mathbf{T}})$ with a specific choice of $X_t(f) = \pi_t(f) = f(t)$. In simpler terms, a stochastic process is just a probability measure $\mathbb{P}$ on $(E^{\mathbf{T}}, \mathcal{E}^{\mathbf{T}})$.

Another point of view is that the only relevant objects are the joint distributions of $(X_{t_1}, X_{t_2}, \cdots, X_{t_n})$ for every n and every finite subset $I = (t_1, t_2, ..., t_n)$ of $\mathbf{T}$. These can be specified as probability measures μ_I on $\mathbb{R}^n$. These μ_I cannot be totally arbitrary. If we allow different permutations of the same set, so that I and I' are permutations of each other then μ_I and $\mu_{I'}$ should be related by the same permutation. If $I \subseteq I'$, then we can obtain the joint distribution of $(X_t)_{t \in I}$ by projecting the joint distribution of $(X_t)_{t \in I'}$ from $\mathbb{R}^{n'}$ to $\mathbb{R}^n$ where n and n' are the cardinalities of I and I' respectively. A stochastic process can then be viewed as a family (μ_I) of distributions on various finite dimensional spaces that satisfy the consistency conditions. A theorem of Kolmogorov says that this is not all that different. Any such consistent family arises from a $\mathbb{P}$ on $(E^{\mathbf{T}}, \mathcal{E}^{\mathbf{T}})$ which is uniquely determined by the family (μ_I).

Definition 1.2. We say A measurable space $(E, \mathcal{E})$ is said to be standard if there exists a Polish space X such that $(E, \mathcal{E})$ is isomorphic (as a measurable space) to $(X, \mathcal{B}(X))$.

THEOREM 1.3 (Kolmogorov's consistency Theorem, cf. [**Yan21**]). *Let E be a standard measure space. Assume that we are given for every $t_1, ..., t_n \in \mathbf{T}$ a probability measure $\mu_{t_1 \cdots t_n}$ on E^n, and that these probability measures satisfy:*

(i) for each $\tau \in S_n$ and $A_i \in \mathcal{E}$,

$$\mu_{t_1 \cdots t_n}(A_1 \times ... \times A_n) = \mu_{t_{\tau(1)} \cdots t_{\tau(n)}}(A_{\tau(1)} \times ... \times A_{\tau(n)});$$

(ii) for each $A_i \in \mathcal{E}$,

$$\mu_{t_1 \cdots t_n}(A_1 \times ... \times A_{n-1} \times E) = \mu_{t_1 \cdots t_{n-1}}(A_1 \times ... \times A_{n-1}).$$

Then, there is a unique probability measure $\mathbf{P}$ on $(E^{\mathbf{T}}, \mathcal{E}^{\mathbf{T}})$ such that for $t_1, ..., t_n \in \mathbf{T}, A_1, ..., A_n \in \mathcal{E}$: $\mathbf{P}(f(t_1) \in A_1, ..., f(t_n) \in A_n) = \mu_{t_1, ..., t_n}(A_1 \times ... \times A_n).$

Let $\mathbf{T} = \mathbb{R}_+$ and E be a Polish space. By Theorem 1.3, we can define a probability measure $\mathbb{P}$ on $E^{\mathbb{R}_+}$ such that the canonical process $X_t(f) = f(t)$ satisfies the conditions in Theorem 1.3. However, whether the measure is concentrated on the space of continuous functions is not a simple question. In fact, since $\mathbf{T} = \mathbb{R}_+$ is uncountable the space of bounded functions, continuous functions, etc., are **not** measurable sets of $E^{\mathbb{R}_+}$. They do not belong to the natural σ-field. Essentially, in probability theory, the rules involve only a countable collection of sets at one time, and any information that involves the values of an uncountable number of measurable functions is beyond reach. There is an intrinsic reason for this. In probability theory, we can always change the values of a random variable on a set of measure 0, and we have not changed anything significant. Since we are allowed to mess up each function on a set of measure 0, we have to assume that each function has indeed been messed up on a set of measure 0. If we are dealing with

a countable number of functions, the 'mess up' has occurred only on the countable union of these individual sets of measure 0, which, by the properties of a measure, is again a set of measure 0. On the other hand, if we are dealing with an uncountable set of functions, then these sets of measure 0 can possibly gang up on us.

Often, we aim to find a version of stochastic process with continuous trajectories, or equivalently, to establish a measure $\mathbb{P}$ on $C(\mathbb{R}_+; \mathbb{R}^d)$ with the natural σ-field. However, this is not always achievable. We are looking for sufficient conditions on the finite dimensional distributions (μ_I) to ensure the existence of $\mathbb{P}$ on $C(\mathbb{R}_+; \mathbb{R}^d)$.

THEOREM 1.4 (Kolmogorov). *Let* $I = [0, T]$, *and let* $p > 1$ *and* $\beta \in (1/p, 1)$. *Assume* $(Y_t \in \mathbb{R}^d)_{t \in I}$ *satisfies*

(1.1) $$\mathbf{E}|Y_s - Y_t|^p \leqslant c|t - s|^{1+\beta p}, \ \forall t, s \in I.$$

Then there exists a version of Y, *say* X *(for each* $t \in I$, $\mathbf{P}(X_t = Y_t) = 1$*), such that*

$$\mathbf{P}\left(\sup_{t \in I} \frac{|X_t - X_s|}{|t - s|^\alpha} \leqslant K\right) = 1,$$

where $\alpha \in (0, \beta - 1/p)$, $K = K(\alpha, \beta, p, c, I, \omega)$ *and* $\mathbf{E}K^p < \infty$.

PROOF. Regard Y as a measurable function from $\Omega \times I$ to $\mathbb{R}^d$. By Lemma 1.5 below, there is a null set $\mathcal{N} \subseteq \Omega$ and a measurable function $X : \Omega \times I \to \mathbb{R}^d$, such that for each $\omega \notin \mathcal{N}$,

$$\mathscr{L}^1\left(\{t \in I : X_t(\omega) \neq Y_t(\omega)\}\right) = 0,$$

and $X(\omega)$ is a continuous function. Moreover,

$$\|X_\cdot(\omega)\|_{C^\alpha(I)} \lesssim K(\omega) := \left(\iint_{I \times I} \frac{|Y_t(\omega) - Y_s(\omega)|^p}{|t - s|^{2+\alpha p}} \, \mathrm{d}s\mathrm{d}t\right)^{1/p} \in L^p(\mathbf{P}).$$

By Fubini theorem, there exists a $\mathscr{L}^1$-null set $N \subseteq I$, such that for each $t \notin N$, $\mathbf{P}(X_t \neq Y_t) = 0$. For any $t_0 \in N$, by (1.1), one can see that $Y_{t_n} \xrightarrow[I \backslash N \ni t_n \to t_0]{\mathbf{P}} Y_{t_0}$. On the other hand, $Y_{t_n} \overset{\text{a.s}}{=} X_{t_n} \to X_{t_0}$, so we have $X_{t_0} \overset{\text{a.s}}{=} Y_{t_0}$. Therefore, X is a continuous version of Y. □

Lemma 1.5 (Fractional Sobolev inequality). *Let* $R > 0$, $p > n$ *and* $s \in (n/p, 1)$. *Let* $f : B_R \to \mathbb{R}^d$ *be a measurable function. Assume*

$$\iint_{B_R \times B_R} \frac{|f(x) - f(y)|^p}{|x - y|^{n+sp}} \, \mathrm{d}x\mathrm{d}y < \infty.$$

Then there exists a version of f, *say* $\tilde{f}$, *such that*

(1.2) $$\sup_{x,y \in B_R} \frac{|\tilde{f}(x) - \tilde{f}(y)|}{|x - y|^{s - \frac{n}{p}}} \leqslant C \left(\iint_{B_R \times B_R} \frac{|f(x) - f(y)|^p}{|x - y|^{n+sp}} \, \mathrm{d}x\mathrm{d}y\right)^{1/p},$$

Here C *only depends on* n, s, p *and* R.

Thanks to Theorem 1.4 and the discussion after Theorem 1.3, we can first construct a probability measure $\mathbf{P}$ on $\Omega = (\mathbb{R}^d)^{\mathbb{R}_+}$ such that

$$\mathbf{P}(Y_{t_1} \in A_1, \cdots, Y_{t_n} \in A_n)$$
$$= \int_{A_1} \cdots \int_{A_n} p_{t_1}(x_1) p_{t_2 - t_1}(x_1 - x_2) \cdots p_{t_n - t_{n-1}}(x_{n-1} - x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_n,$$

where $Y_t(\omega) = \omega$ is the canonical process, and $p_t(x) = (2\pi t)^{-\frac{d}{2}} \exp(-|x|^2/(2t))$. Then using Theorem 1.4, one can establish the existence of an α-Hölder continuous version of Y with $\alpha \in (0, \frac{1}{2})$, which is a Brownian motion. Once we get such a continuous version, in fact we obtain a probability measure $\mathbb{P}$ on $C(\mathbb{R}_+; \mathbb{R}^d)$, under which the canonical process is a Brownian motion.

1.2. Markov Processes

Intuitively speaking, a process X is Markov if, given its whole past up until some time s, the future behaviour depends only its state at time s. To make this precise, let us suppose that X takes values in a measurable space $(E, \mathcal{E})$ and, to denote the past, let $\mathcal{F}_t$ be the sigma-algebra generated by $\{X_s : s \leqslant t\}$. The Markov property then says that, for any times $s \leqslant t$ and bounded measurable function $f : E \to \mathbb{R}$, the expected value of $f(X_t)$ conditional on $\mathcal{F}_s$ is a function of X_s. Equivalently,

$$(1.3) \qquad \mathbf{E}\left[f(X_t) \mid \mathcal{F}_s\right] = \mathbf{E}\left[f(X_t) \mid X_s\right], \quad \text{a.s.}$$

More generally, this idea makes sense with respect to any filtered probability space $\mathbb{F} = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geqslant 0}, \mathbf{P})$. A process X is Markov with respect to $\mathbb{F}$ if it is adapted and (1.3) holds for times $s \leqslant t$.

Continuous time Markov processes are usually defined in terms of transition functions. These specify how the distribution of X_t is determined by its value at an earlier time s. To state the definition of transition functions, it is necessary to introduce the concept of transition probabilities.

Definition 1.6. A (transition) kernel Q on a measurable space $(E, \mathcal{E})$ is a map

$$Q : E \times \mathcal{E} \to \mathbb{R}_+ \cup \{\infty\},$$
$$(x, A) \mapsto N(x, A)$$

such that for each $x \in E$, the map $A \mapsto Q(x, A)$ is a measure, and for each $A \in \mathcal{E}$, the map $x \mapsto N(x, A)$ is measurable. If, furthermore, $Q(x, E) = 1$ for all $x \in E$, then Q is a transition probability.

For any $f \in B(E)$, we set

$$Qf(x) = \int_E f(y) Q(x, \mathrm{d}y).$$

A transition probability, then, associates to each $x \in E$ is a probability measure on $(E, \mathcal{E})$. This can be used to describe how the conditional distribution of a process at a time t depends on its value at an earlier time s by

$$\mathbf{P}(X_t \in A \mid \mathcal{F}_s) = Q(X_s, A).$$

A Markov process is defined by a collection of transition probabilities $(P_{s,t})_{t \geqslant s}$, describing how it goes from its state at time s to a distribution at time t. We only consider the homogeneous case here, meaning that $P_{s,t}$ depends only on the size $t - s$ of the time increment, so the notation $P_{s,t}$ can be replaced by P_{t-s}.

Definition 1.7. A homogeneous transition function on $(E, \mathcal{E})$ is a collection P_t, $t \geq 0$ of transition probabilities on $(E, \mathcal{E})$ such that

$$P_{s+t} = P_s P_t, \quad s, t \geqslant 0$$

A process X is Markov with transition function $P = (P_t)_{t \geqslant 0}$, and with respect to a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbf{P})$ if it is adapted and

$$\mathbf{E}\left(f(X_t) \mid \mathcal{F}_s\right) = P_{t-s}f(X_s), \quad t > s.$$

The identity $P_{s+t} = P_s P_t$ is known as the Chapman-Kolmogorov equation, and is required so that the transition probabilities are consistent with the tower rule for conditional expectations. Alternatively $(P_t)_{t \geqslant 0}$ forms a semigroup.

The distribution of a Markov process is determined uniquely by its transition function and initial distribution.

Proposition 1.8. *Suppose that X is a Markov process on $(E, \mathcal{E})$ with transition function P such that X_0 has distribution μ. Then, for any times $0 = t_0 < t_1 < \cdots < t_n$ and bounded measurable function $f \colon E^{n+1} \to \mathbb{R}$,*

$$\mathbf{E}[f(X_{t_0}, \ldots, X_{t_n})]$$
$$= \iint \cdots \int f(x_0, \ldots, x_n) \, P_{t_n - t_{n-1}}(x_{n-1}, \mathrm{d}x_n) \cdots P_{t_1 - t_0}(x_0, \mathrm{d}x_1) \mu(\mathrm{d}x_0).$$

Proposition 1.9. *Let $(E, \mathcal{E})$ be a measurable space, and $\Omega = E^{\mathbb{R}_+}$. Denote its coordinate process by X,*

$$X_t \colon \Omega \to E, \quad \omega \mapsto X_t(\omega) = \omega(t).$$

Also, let $\mathcal{F}^0$ be the σ-algebra generated by $\{X_t \colon t \in \mathbb{R}_+\}$ and, for each $t \geqslant 0$, let $\mathcal{F}_t^0$ be the σ-algebra generated by $\{X_s \colon s \leq t\}$. So, $(\mathcal{F}_t^0)_{t \geqslant 0}$ is a filtration on the measurable space $(\Omega, \mathcal{F}^0)$ with respect to which X is adapted.

Then, for every transition function $(P_t)_{t \geqslant 0}$ and probability distribution μ on E, there is a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F}^0)$ under which X is a Markov process with transition function $(P_t)_{t \geqslant 0}$ and initial distribution μ.

Remark 1.10. The superscripts $'0'$ just denote the fact that we are using the un-completed σ-algebras. Once the probability measure has been defined, it is standard practice to complete the filtration, which does not affect the Markov property.

The unique measure with respect to which X is Markov with the given transition function and initial distribution is denoted by $\mathbb{P}_\mu$, and expectation with respect to this measure is denoted by $\mathbb{E}_\mu$. In particular, if $\mu = \delta_x$ then we write $\mathbb{P}_x \equiv \mathbb{P}_{\delta_x}$ and, similarly, write $\mathbb{E}_x$ for $\mathbb{E}_{\delta_x}$.

1.3. Diffusions

Diffusion is a physical phenomenon that describes the process by which several substances mixed together tend to move towards equilibrium. For example, Brownian motion describes the process by which pollen particles suspended in a liquid gradually "diffuse" to a "uniform" distribution. A natural question arises: if the physical properties of the liquid at different times and locations affect the pollen particles differently, for instance, in a flowing liquid, what motion laws will the pollen particles follow?

The diffusion process does not have a unified mathematical definition, but its core is a *Markov process* with continuous trajectories. Similar to Brownian motion, the evolution of its macroscopic properties can be characterized by establishing equations that satisfy the transition probabilities. Alternatively, by tracking the

trajectory of each pollen particle, a probability space can be constructed, and *stochastic differential equations (SDE)* can be established to describe the motion laws they obey from a microscopic perspective.

We will first present its construction using the first method, which can be traced back to Kolmogorov's early groundbreaking papers on Markov processes.

1.3.1. Fokker-Planck-Kolmogorov equations. Compared to Brownian motion, we give three conditions for a *time-homogeneous* diffusion process X_t: for any $\varepsilon > 0$,

$$(1.4) \qquad \lim_{t \to 0} t^{-1} \sup_{x \in \mathbb{R}^d} P_t(x, B_\varepsilon^c(x)) = 0,$$

$$(1.5) \qquad \lim_{t \to 0} t^{-1} \int_{|y-x| \leqslant \varepsilon} (y - x) P_t(x, \mathrm{d}y) = b(x),$$

$$(1.6) \qquad \lim_{t \to 0} (2t)^{-1} \int_{|y-x| \leqslant} (y - x)_i \, (y - x)_j \, P_t(x, \mathrm{d}y) = a_{ij}(x) \quad i, j = 1, \cdots, d.$$

b and a are called the drift coefficient and diffusion coefficient of the diffusion process $(X_t)_{t \geqslant 0}$, respectively. In the sequel, we always assume that

$$a, b \in L^\infty.$$

We want to derive the evolution laws that the transition probabilities should satisfy. Let $f \in C_b^2(\mathbb{R}^d)$. Then

$$\frac{P_t f(x) - f(x)}{t}$$
$$= \frac{1}{t} \int_{|y-x| \leqslant \varepsilon} (f(y) - f(x)) \, P_t(x, \mathrm{d}y) + \frac{1}{t} \int_{|y-x| > \varepsilon} (f(y) - f(x)) \, P_t(x, \mathrm{d}y) =: I_1 + I_2.$$

By Taylor's expansion theorem, and using (1.5) and (1.6), we have

$$I_1 = \frac{1}{t} \int_{|y-x| \leqslant \varepsilon} (y - x)_i P_t(x, \mathrm{d}y) \, \partial_i f(x)$$
$$+ \frac{1}{2t} \int_{|y-x| \leqslant \varepsilon} (y - x)_i (y - x)_j P_t(x, \mathrm{d}y) \, \partial_{ij} f(x) + o(1)$$
$$\longrightarrow b(x) \cdot \nabla f(x) + a_{ij}(x) \partial_{ij} f(x), \ t \to 0.$$

Applying (1.4), we have $I_2 \to 0$, as $t \to 0$. Therefore,

$$(1.7) \ \lim_{t \to \infty} \frac{P_t f(x) - f(x)}{t} = a_{ij}(x) \partial_{ij} f(x) + b(x) \cdot \nabla f(x) =: L f(x), \quad f \in C_b^2(\mathbb{R}^d).$$

Assume that

$$P_t(x, \mathrm{d}y) = p(t, x, y) \mathrm{d}y \ \text{ and } P_t f \in C_b^2, \quad t \geqslant 0.$$

Thanks to the Chapman-Kolmogorov equation and (1.7), one can verify that

$$\partial_t P_t f(x) = L P_t f(x), \quad \lim_{t \to 0} P_t f(x) = f(x),$$

which can be read as

$$(1.8) \qquad \partial_t p(\cdot, y) = L p(\cdot, y), \quad \lim_{t \to 0} p(t, \cdot, y) = \delta_y.$$

Kolmogorov's idea for constructing the diffusion process corresponding to L involves solving the partial differential equation (PDE) (1.8) (in fact his solves the forward equation in his paper) to obtain the density of the process, $p(t, x, \cdot)$.

We need to introduce some notations. For any $y \in \mathbb{R}^d$, put

$$L_{a(y)} f(x) := a_{ij}(y) \partial_{ij} f(x)$$

and

$$p_0^y(t, x, z) := \left[\pi t \sqrt[d]{\det(a(y))} \right]^{-\frac{d}{2}} \exp\left(-\frac{a_{ij}^{-1}(y)(x-z)_i(x-z)_j}{t} \right).$$

Let

$$\mathbb{D} = \{(t, x, y) : 0 \leqslant t \leqslant 1, x, y \in \mathbb{R}^d, x \neq y\}.$$

For any $\lambda > 0, \gamma \in \mathbb{R}$, put

$$\varrho_{\lambda,\gamma}(t, x) := t^{(-d+\gamma)/2} e^{-\frac{\lambda|x|^2}{t}}, \quad t > 0, x \in \mathbb{R}^d.$$

ϱ_λ is denoted by ϱ_λ for simplicity. For any $p^{(1)}, p^{(2)}, \cdots, p^{(n)} : \mathbb{D} \to \mathbb{R}$, define

$$\left[p^{(n)} \otimes \cdots \otimes p^{(2)} \otimes p^{(1)} \right](t, x, y)$$

$$:= \int_{0 < \tau_1 < \cdots < \tau_{n-1} < t} \int_{\mathbb{R}^{nd}} p^{(n)}(t - \tau_{n-1}, x, z_{n-1}) \cdots$$

$$p^{(2)}(\tau_2 - \tau_1, z_2, z_1) p^{(1)}(\tau_1, z_1, y) dz_1 \cdots dz_{n_1} \, d\tau_1 \cdots d\tau_{n-1}.$$

It is easy to verify that

$$\partial_t p_0^y(\cdot, z) = L_{a(y)} p_0^y(\cdot, z), \quad y, z \in \mathbb{R}^d.$$

Recall that p satisfies (1.8), therefore,

$$\partial_t p = L_{a(y)} p + (L - L_{a(y)}) p.$$

Formally, using Duhamel's formula, we have

$$p(t, x, y) = p_0^y(t, x, y) + [p_0^y \otimes (L - L_{a(y)}) p](t, x, y),$$

(1.9)

$$p(\cdot, y) = \sum_{n=0}^\infty \left[\underbrace{p_0^y \otimes \overbrace{[(L - L_{a(y)}) p_0^y]^{\otimes^n}}^{=:q_n}}_{=:p_n} \right](\cdot, y) = [p_0^y + p_0^y \otimes q](\cdot, y), \quad q = \sum_{n=1}^\infty q_n$$

and

(1.10) $$p(\cdot, y) = \left[p_0^y + p_0^y \otimes (L - L_{a(y)}) p \right](\cdot, y), \quad y \in \mathbb{R}^d.$$

For notion simplicity, we omit the superscript y below.

We attempt to show that the infinite series in (1.9) do convergence (in some sense), and p given by (1.9) satisfying (1.10) is a fundamental solution to (1.8), provided that the coefficients satisfies

ASSUMPTION 1. There exists $\alpha \in (0, 1)$ and $\Lambda > 1$ such that

$$\Lambda^{-1} |\xi|^2 \leqslant a_{ij} \xi_i \xi_j \leqslant \Lambda |\xi|^2$$

and

$$\|a\|_{C^\alpha} = N_1 < \infty, \quad \|b\|_{L^\infty} = N_2 < \infty.$$

1.3.2. Heat Kernel Estimate. In this subsection, we use the classical Levi's freezing coefficients method to prove that (1.8) admits a nice solution provided that the coefficients a and b satisfies Assumption 1.

For simplicity, we always assume $b = 0$ in the sequel. Readers interested in the general case can work out the details themselves.

THEOREM 1.11. *Under Assumption 1, there is a unique continuous transition density function $p(t, x, y) \in \mathbb{D}$ such that*

$$\partial_t p(\cdot, y) = L p(\cdot, y) \quad and \quad \lim_{t \to 0} p(t, x, y) = \delta_y, \quad y \in \mathbb{R}^d.$$

Moreover,

(i) *for any $f \in C_0(\mathbb{R}^d)$, $P_t f \to f$ uniformly;*

(ii) *for any $t \in [0, 1], x \in \mathbb{R}^d$*

$$(1.11) \qquad p(t, x, \cdot) \geqslant 0 \quad and \quad \int_{\mathbb{R}^d} p(t, x, y) \mathrm{d}y = 1;$$

(iii) *for any $t, s \in [0, 1]$ and $x, y \in \mathbb{R}^d$*

$$(1.12) \qquad p(t + s, x, y) = \int_{\mathbb{R}^d} p(t, x, z) p(s, z, y) \mathrm{d}z;$$

(iv) *there is a constant $C > 1$ only depending on d, α, Λ and N_i such that such that for any $(t, x, y) \in \mathbb{D}$,*

$$(1.13) \qquad C^{-1} t^{-\frac{d}{2}} \exp(-C|x|^2/t)) \leqslant p(t, x, y) \leqslant C t^{-\frac{d}{2}} \exp(-|x|^2/(Ct))).$$

Lemma 1.12. *For any $k \in \mathbb{N}$, there is a constant $\lambda_k > 0$ such that*

$$|\nabla_x^k p_0| \lesssim \varrho_{\lambda_k, -k}$$

and for any $t \in [0, 1]$, $x_1, x_2, z \in \mathbb{R}^d$ and $\beta \in (0, 1)$, it holds that

$$|\nabla_x^k p_0(t, x_1, z) - \nabla_x^k p_0(t, x_2, z)|$$
$$\lesssim |x_1 - x_2|^\beta \left[\varrho_{\lambda_k, -k-\beta}(t, x_1, z) + \varrho_{\lambda_k, -k-\beta}(t, x_2, z) \right].$$

Exercise 1.13. Prove Lemma 1.12.

Lemma 1.14. *It holds that*

$$(1.14) \qquad |q| \lesssim \varrho_{\lambda, \alpha-2};$$

For any $t \in [0, 1]$, $x_1, x_2, y \in \mathbb{R}^d$ and $\beta \in (0, 1)$, it holds that

$$(1.15) \qquad |q(t, x_1, y) - q(t, x_2, y)| \lesssim |x_1 - x_2|^\beta \sum_{i=1,2} \varrho_{\lambda, \alpha-\beta-2}(t, x_i, y).$$

PROOF. We Claim that

$$(1.16) \qquad |q_n(t, x, y)| \leqslant \underbrace{\frac{(C_1 \Gamma(\alpha/2))^n}{(\lambda/\pi)^{d(n-1)/2} \Gamma(n\alpha/2)}}_{=: \gamma_n} \varrho_{\lambda, n\alpha-2}(t, x - y),$$

where Γ is the Gamma function, and C_1 and λ only depends on d, α, Λ and N_i.

By Lemma 1.12 and the Hölder regularity of a, we have

$$|q_1(t, x, y)| = |[(L - L_{a(y)})p_0](t, x, y)| \leqslant C_1 \varrho_{\lambda, \alpha-2}(t, x, y).$$

Assume (1.16) holds. This together with the fact that $q_{n+1} = q_1 \otimes q_n$ yields

$$|q_{n+1}(t, x, y)|$$

$$\leqslant C_1 \gamma_n \int_0^t (t - \tau)^{\frac{\alpha}{2}-1} \tau^{\frac{n\alpha}{2}-1} d\tau \int_{\mathbb{R}^d} \varrho_\lambda(t - \tau, x - z) \varrho_\lambda(\tau, z - y) dz$$

$$= C_1 \left(\pi\lambda^{-1}\right)^{d/2} \gamma_n \varrho_\lambda(t, x - y) \int_0^t (t - \tau)^{\frac{\alpha}{2}-1} \tau^{\frac{n\alpha}{2}-1} d\tau$$

$$= C_1 \left(\pi\lambda^{-1}\right)^{d/2} \gamma_n \varrho_{\lambda, (n+1)\alpha-2}(t, x - y) B\left(\frac{n\alpha}{2}, \frac{\alpha}{2}\right)$$

$$= \gamma_{n+1} \varrho_{\lambda, (n+1)\alpha-2}(t, x - y).$$

Therefore, we finish the proof for (1.16), which also implies

$$q = \sum_{n=1}^{\infty} q_n \lesssim \varrho_{\lambda, \alpha-2}.$$

Next, we verify that

(1.17)
$$|q_1(t, x_1, y) - q_1(t, x_2, y)| \lesssim |x_1 - x_2|^\beta \sum_{i=1,2} \varrho_{\lambda, \alpha-\beta-2}(t, x_i - y), \quad (t, x, y) \in \mathbb{D}.$$

If $|x_1 - x_2| > \sqrt{t}$, then it is a consequence of (1.16). When $|x_1 - x_2| \leqslant \sqrt{t}$, we have

$$|q_1(t, x_1, y) - q_1(t, x_2, y)|$$

$$\leqslant |a(x_1) - a(x_2)| \cdot |\nabla_x^2 p_0(t, x_1, y)| + |a(x_2) - a(y)| \cdot |\nabla_x^2 p_0(t, x_1, y) - \nabla_x^2 p_0(t, x_2, y)|$$

$$\lesssim |x_1 - x_2|^\alpha \varrho_{\lambda_2, -2}(t, x_1 - y) + |x_2 - y|^\alpha |x_1 - x_2| \varrho_{\lambda_3, -3}(t, x_2 - y - \theta(x_1 - x_2))$$

$$\lesssim |x_1 - x_2|^\beta \varrho_{\lambda, \alpha-\beta-2}(t, x_1 - y) + |x_2 - y|^\alpha |x_1 - x_2|^\beta \varrho_{\lambda, -\beta-2}(t, x_2 - y)$$

$$\lesssim |x_1 - x_2|^\beta \sum_{i=1,2} \varrho_{\lambda, \alpha-\beta-2}(t, x_i - y).$$

Therefore, (1.17) holds for any $(t, x, y) \in \mathbb{D}$. Noting that $q = q_1 + q_1 \otimes q$, we have

$$|q(t, x_1, y) - q(t, x_2, y)|$$

$$\lesssim |x_1 - x_2|^\beta \sum_{i=1,2} \varrho_{\lambda, \alpha-\beta-2}(t, x_i - y)$$

$$+ |x_1 - x_2|^\beta \int_0^t (t - \tau)^{\frac{\alpha-\beta}{2}-1} \tau^{\frac{\alpha}{2}-1} d\tau \sum_{i=1,2} \varrho_\lambda(t, x_i - y)$$

$$\lesssim |x_1 - x_2|^\beta \sum_{i=1,2} \varrho_{\lambda, \alpha-\beta-2}(t, x_i - y).$$

$$\square$$

The above lemma implies that the infinite series in (1.9) do convergence, and p given by (1.9) satisfying (1.10).

Lemma 1.15. *There is a constant $\lambda > 0$ such that*

(1.18)
$$|p| \lesssim \varrho_\lambda, \quad |\partial_t p| + |\nabla_x^2 p| \lesssim \varrho_{\lambda, -2}$$

and

(1.19) $$[L - L_{a(y)})p](t, x, y) = q(t, x, y).$$

PROOF. Recalling that $p = p_0 + p_0 \otimes q$, by Lemma 1.12, we only need to prove

(1.20) $$|\nabla_x^2 p_0 \otimes q| \lesssim \varrho_{\lambda, \alpha - 2}.$$

Note

$$\nabla_x^2 (p_0 \otimes q)(t, x, y) = \int_0^t \int_{\mathbb{R}^d} \nabla_x^2 p_0(t - \tau, x, z) \, q(\tau, z, y) \, \mathrm{d}z \, \mathrm{d}\tau$$

$$= \int_0^{\frac{t}{2}} \cdots + \int_{\frac{t}{2}}^t \cdots =: I_1 + I_2.$$

Thanks to Lemma 1.12 and Lemma 1.14,

$$|I_1| \lesssim \int_0^{\frac{t}{2}} \varrho_{\lambda, -2}(t - \tau, x - z) \varrho_{\lambda, \alpha - 2}(\tau, z - y) \, \mathrm{d}z \mathrm{d}\tau$$

$$\lesssim \varrho_\lambda(t, x - y) \int_0^{\frac{t}{2}} (t - \tau)^{-1} \tau^{\frac{\alpha}{2} - 1} \mathrm{d}\tau \lesssim \varrho_{\lambda, \alpha - 2}(t, x - y)$$

Noting that $p_0(t, x, z) = p_0(t, x - z)$, we have

$$\int_{\mathbb{R}^d} \nabla_x^k p_0(t, x, z) \mathrm{d}z = 0, \quad k \in \mathbb{N}.$$

In view of (1.15), we get

$$|I_2| = \left| \int_{\frac{t}{2}}^t \mathrm{d}\tau \int_{\mathbb{R}^d} \nabla_x^2 p_0(t - \tau, x, z) \left[q(\tau, z, y) - q(\tau, x, y) \right] \mathrm{d}z \right|$$

$$\lesssim \int_{\frac{t}{2}}^t (t - \tau)^{\frac{\beta}{2} - 1} \tau^{\frac{\alpha - \beta}{2} - 1} \mathrm{d}\tau \int_{\mathbb{R}^d} \left[\varrho_\lambda(\tau, y - z) + \varrho_\lambda(\tau, x - y) \right] \varrho_\lambda(t - \tau, x - z) \mathrm{d}z$$

$$\lesssim \varrho_{\lambda, \alpha - 2}(t, x - y).$$

Therefore, $|\nabla_x^2 p| \lesssim \varrho_{\lambda, -2}$. Similarly, one can verify that $|\partial_t p| \lesssim \varrho_{\lambda, -2}$. □

The above lemma implies

$$p(t, x, y) = p_0(t, x, y) + [p_0 \otimes (L - L_{a(y)})p](t, x, y),$$

which yields that p satisfies $\partial_t p = Lp$.

PROOF OF THEOREM 1.11. (i). It is easy to verify that

$$v(t, x) := \int_{\mathbb{R}^d} p_0(t, x, z) f(z) \mathrm{d}z$$

convergence to f uniformly when $f \in C_0(\mathbb{R}^d)$ as $t \to 0$. In the light of (1.14), we have

(1.21) $$|p_0 \otimes q| \lesssim \varrho_{\lambda, \alpha},$$

which yields that

$$\left| \int_{\mathbb{R}^d} (p_0 \otimes q)(t, x, y) f(y) \mathrm{d}y \right| \lesssim t^{\frac{\alpha}{2}} \|f\|_{L^\infty} \to 0.$$

Therefore, our desired assertion holds, due to the fact that $p = p_0 + p_0 \otimes q$.

(ii) and (iii) follow directly as consequences of the maximum principle for parabolic equations.

(iv) Thanks to Lemma 1.15, we only need to prove the lower bound estimate. There exists constant $T > 0$ such that
(1.22)
$$p(t,x,y) \geqslant p_0(t,x,y) - |p_0 \otimes q|(t,x,y) \gtrsim t^{-\frac{d}{2}} - t^{\frac{-d+\alpha}{2}} \gtrsim t^{-\frac{d}{2}}, \quad |x-y| < \sqrt{t}, \ t \in [0,T].$$

If $|x - y| > \sqrt{t}$, let Let n be the least integer greater than $4|x - y|/\sqrt{t}$, i.e. $n - 1 \leqslant 4|x - y|^2/t < n$.

$$x_i = x + (y - x)i/n, \quad B_i := B\left(x_i, 8^{-1}\sqrt{t/n}\right) \quad \text{and} \quad t_i = it/n.$$

Noting that for all $z_i \in B_i$,

$$|z_i - z_{i+1}| \leqslant |z_i - x_i| + |x_i - x_{i+1}| + |x_{i+1} - z_{i+1}| \leqslant \frac{\sqrt{t/n}}{2},$$

by the on-digonal estimate (1.22), we have

$$p(t_{i+1} - t_i, z_i, z_{i+1}) \geqslant c_2(t/n)^{-\frac{d}{2}}.$$

Hence, by the C-K equation, there is a constant $c_3 \in (0,1)$ such that

$$p(t,x,y) \geqslant \int_{B_{n-1}} \cdots \int_{B_1} p(t_1,x,z_1) \cdots p(t_n - t_{n-1}, z_{n-1}, y) \mathrm{d}z_1 \cdots \mathrm{d}z_{n-1}$$

$$\geqslant \left[c_2(t/n)^{-\frac{d}{2}}\right]^n \left[\omega_d \left(\sqrt{t/(64n)}\right)^d\right]^{n-1} \geqslant t^{-\frac{d}{2}} c_3^n n^{\frac{d}{2}}$$

$$\geqslant t^{-\frac{d}{2}} c_3^{4|x-y|^2/t} (|x-y|^2/t)^{\frac{d}{2}} \gtrsim \varrho_\lambda(t,x,y).$$

$\square$

Construction of Diffusion processes II

2.1. Motivation

Kolmogorov's approach to constructing diffusion processes is purely analytical. A natural question that arises is whether it is possible to provide a "microscopic" construction at the level of trajectories. This was one of the original motivations for Itô's introduction of stochastic integrals and stochastic differential equations.

Assume $d = 1$. Intutively, the diffusion process can be constructed as follows: Let Δt be a fixed, small time interval, and consider the following approximation process:

$$X_t = X_0 + b(X_0)t + \sqrt{2a(X_0)}(W_t - W_0), \qquad\qquad t \in [0, \Delta t);$$

$$X_t = X_{\Delta t} + b(X_{\Delta t})(t - \Delta t) + \sqrt{2a(X_{\Delta t})}(W_t - W_{\Delta t}), \qquad\qquad t \in [\Delta t, 2\Delta t);$$

$$\vdots$$

$$X_t = X_{k\Delta t} + b(X_{k\Delta t})(t - k\Delta t) + \sqrt{2a(X_{k\Delta t})}(W_t - W_{k\Delta t}), \qquad t \in [k\Delta t, (k+1)\Delta t).$$

If, as $\Delta t \to 0$, the process X_t (which depends on Δt) converges (in some sense) to a stochastic process, then formally, the limiting process (still denoted as X_t) satisfies:

$$(2.1) \qquad\qquad X_t = X_0 + \int_0^t b(X_t)\mathrm{d}t + \underbrace{\int_0^t \sqrt{2a(X_t)}\mathrm{d}W_t}_{?}.$$

One issue is how to understand the last term in the above equation.

Exercise 2.1. Prove that

$$\mathbf{P}\left(\limsup_{t \to 0} \frac{|W_t|}{t^{1/2}} = \infty\right) = 1.$$

When h_t is β-Hölder continuous and x_t is α-Hölder continuous, if $\alpha + \beta > 1$, then we can prove that the following Riemann sum converges:

$$\sum_{k=0}^{n-1} h_k(x_{k+1} - x_k).$$

However, when $\alpha + \beta < 1$, in general, we cannot mathematically prove the convergence of the above Riemann sum.

Note that the paths of Brownian motion are only α-Hölder continuous for $\alpha < 1/2$. It is expected that the Hölder exponent of the paths of X will not exceed $1/2$ either. Therefore, unless we uncover more information about the paths

of Brownian motion, we cannot define the second term on the right-hand side of
(2.1) pathwise.

Itô considered a more general problem: Suppose H is an adapted process, how
can we define the following integral

$$\int_0^1 H_s \mathrm{d}W_s.$$

Assume $F_t = \sigma(W_s : s \in [0, t])$, and H is a bounded simple process with respect to
F_t, i.e., $H_t = H_0 I_{\{0\}}(t) + \sum_{i=0}^{n-1} H_{t_i} I_{(t_i, t_{i+1}]}(t)$, where $0 = t_0 < t_1 < \cdots < t_n = 1$,
and $H_{t_i} \in F_{t_i}$. Naturally, we can define $\int_0^t H_s \mathrm{d}W_s = \sum_{i=0}^{n-1} H_{t_i}(W_{t \wedge t_{i+1}} - W_{t \wedge t_i})$.
Itô observed that $t \mapsto \int_0^t H_s \mathrm{d}W_s$ is a martingale and satisfies the isometry property

$$\mathbf{E}\left(\int_0^t H_s \mathrm{d}W_s\right)^2 = \mathbf{E}\int_0^t H_s^2 \mathrm{d}s.$$

If H is a general adapted process, and there exists a sequence of simple processes
$(H^n)_{n \in \mathbb{N}}$ such that

$$\mathbf{E}\left(\int_0^1 (H_s^n - H_s)^2 \mathrm{d}s\right) \to 0, \quad n \to \infty,$$

then by Doob's inequality,

$$\mathbf{E}\left(\sup_{t \in [0,1]}\left|\int_0^t H_s^n \mathrm{d}W_s - \int_0^t H_s^m \mathrm{d}W_s\right|^2\right) \leqslant C\mathbf{E}\int_0^1 (H_s^n - H_s^m)^2 \mathrm{d}s \to 0.$$

Thus, $\int_0^t H_s^n \mathrm{d}W_s$ converges to a continuous martingale, which we define as $\int_0^t H_s \mathrm{d}W_s$.
In fact, later we can argue that we can define the integral of a very general adapted
process H with respect to Brownian motion W. Once the integral with respect to
Brownian motion is defined, under very general conditions, we can solve equation
(2.1), providing a probabilistic construction of diffusion processes.

Essentially, Itô utilized the adaptability of the integrand and the martingale
property of Brownian motion. Thus, stochastic analysis injected new vitality into
the development of martingale theory. Kunita and Watanabe extended the theory
of stochastic integrals from the case of Brownian motion to general square-integrable
martingales using the Doob-Meyer decomposition. The Strasbourg school in France
further generalized it to the most general case of semimartingales and established
a general theory of stochastic processes.

Basic results in stochastic analysis will be used in this note are presented in
the next section.

2.2. Basic Stochastic Analysis

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a standard probability space. Let $\mathcal{F}_n$ $(n \in \mathbb{N})$ be an increasing
sequence of σ-fields. A sequence of random variables X_n is **adapted** to $\mathcal{F}_n$ if for
each n, X_n is $\mathcal{F}_n$ measurable. Similarly a collection of random variables X_t $(t \in \mathbb{R}_+)$
is adapted to $\mathcal{F}_t$ if each X_t is $\mathcal{F}_t$ measurable. We say the filtration $\mathcal{F}_t$ satisfies
the usual conditions if $\mathcal{F}_t$ is **right continuous** (i.e., $\mathcal{F}_t = \mathcal{F}_{t+}$ for all t, where
$\mathcal{F}_{t+} = \cap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}$) and each $\mathcal{F}_t$ is **complete** (i.e., $\mathcal{F}_t$ contains all $\mathbf{P}$-null sets).

We say $\tau : \Omega \to \mathbb{N} \, (\mathbb{R}_+) \cup \{\infty\}$ is a **stopping time** if τ satisfying $\{\tau \leqslant n\} \in$
$\mathcal{F}_n$ $(\{\tau \leqslant t\} \in \mathcal{F}_t)$, for each $n \in \mathbb{N}$ $(t \in \mathbb{R}_+)$.

$\mathcal{F}_\tau$ is a σ-field containing all measurable sets $A \in cF$ such that $A \cap \{\tau \leqslant n\} \in \mathcal{F}_n$ $(A \cap \{\tau \leqslant t\} \in \mathcal{F}_t)$ for all $n \in \mathbb{N}$ $(t \in \mathbb{R}_+)$.

Definition 2.2. Let X_t be a real-valued $\mathcal{F}_t$-adapted processes. If for each t and $s < t$, X_t is integrable and $\mathbf{E}(X_t|\mathcal{F}_s) \geqslant (\leqslant) X_s$ a.s., then we call X_t is a **submartingale** **(supermartingale)**. We say X_t is a **martingale** if it is both a submartingale and a supermartingale.

Example 2.3. Let $\xi_1, \xi_2, \cdots$ be a sequence of i.i.d random variable. Set $X_n := \sum_{i=0}^n \xi_i$ and $\mathcal{F}_n := \sigma(\xi_0, \cdots \xi_n)$.

Below we recall the results about discrete time martingales and submartingales that will be used. The proof of the subsequent statements can be found in Durrett's book [**Dur19**], and in many other books dealing with discrete time martingales.

THEOREM 2.4 (Doob). *If $X_n \in \mathcal{F}_n$ is a submartingale then it can be uniquely decomposed as $X_n = M_n + A_n$, where $M_n \in \mathcal{F}_n$ is martingale, $A_n = 0$, $A_{n+1} \geqslant A_n$ almost surely and A_n is $\mathcal{F}_{n-1}$-measurable.*

The following theorem lies at the basis of all other results for martingales.

THEOREM 2.5 (Doob's Optional stopping theorem). *Assume that σ and τ are two bounded stopping time, and X_t is a submartingale, then $\mathbf{E}(X_\tau|\mathcal{F}_\sigma) \geqslant X_{\sigma \wedge \tau}$.*

Lemma 2.6. *Let X_n be a submartingale, and τ be a bounded stopping time and $\tau \leqslant K$ (constant). Then*

(i) $\mathbf{E}(X_K|\mathcal{F}_\tau) \geqslant X_\tau$;
(ii) $X_{\tau \wedge n}$ is a $\mathcal{F}_n$-submartingale.

PROOF. (i). for each $A \in \mathcal{F}_\tau$, we will show that $\mathbf{E}(X_K; A) \geqslant \mathbf{E}(X_\tau; A)$. In fact,

$$\mathbf{E}(X_\tau; A) = \sum_{k=0}^K \mathbf{E}(X_k; \underbrace{A \cap \{\tau = k\}}_{\in \mathcal{F}_k}) \leqslant \sum_{k=0}^K \mathbf{E}(X_K; A \cap \{\tau = k\}) = \mathbf{E}(X_K; A).$$

(ii). For each $A \in \mathcal{F}_{n-1}$,

$$\begin{aligned}
\mathbf{E}(X_{\tau \wedge n}; A) &= \mathbf{E}(X_{\tau \wedge n}; A \cap \{\tau \leqslant n-1\}) + \mathbf{E}(X_{\tau \wedge n}; A \cap \{\tau > n-1\}) \\
&= \mathbf{E}(X_\tau; A \cap \{\tau \leqslant n-1\}) + \mathbf{E}(X_n; \underbrace{A \cap \{\tau > n-1\}}_{\in \mathcal{F}_{n-1}}) \\
&\geqslant \mathbf{E}(X_\tau; A \cap \{\tau \leqslant n-1\}) + \mathbf{E}(X_{n-1}; A \cap \{\tau > n-1\}) \\
&= \mathbf{E}(X_{\tau \wedge (n-1)}; A).
\end{aligned}$$

$\square$

PROOF OF THEOREM 2.5. By the above lemma, we have $\mathbf{E}(X_\tau|\mathcal{F}_\sigma) = \mathbf{E}(X_{K \wedge \tau}|\mathcal{F}_\sigma) \geqslant X_{\sigma \wedge \tau}$.

$\square$

THEOREM 2.7 (Doob's inequality). *Let M_n be a martingale. If $M_n^* := \sup_{k \leqslant n} |M_k|$, then*

$$\mathbf{P}(M_n^* > \lambda) \leqslant \lambda^{-1} \mathbf{E}(|M_n|; M_n^* > \lambda).$$

PROOF. Let $\tau = \inf\{k : |M_k| > \lambda\}$. Noting that $\{M_n^* > \lambda\}\} = \{\tau \leqslant n\}$, we have

$$\lambda \mathbf{P}(M_n^* > \lambda) = \lambda \mathbf{P}(\tau \leqslant n) \leqslant \mathbf{E}(|M_\tau|; \tau \leqslant n)$$
$$\leqslant \mathbf{E}(|M_{\tau \wedge n}|; \tau \leqslant n) \leqslant \mathbf{E}(|M_n|; M_n^* > \lambda).$$

$\square$

Corollary 2.8. *Let M_n be a martingale and T be a stopping time. For each $p > 1$, $\mathbf{E}|M_T^*|^p \leqslant C_p \mathbf{E}|M_T|^p$.*

Let $a \leqslant b$. Set $\sigma_1 = \inf\{n \geqslant 0 : X_n \leqslant a\}$, $\tau_1 = \inf\{n > \sigma_1 : X_n \geqslant b\}$, $\sigma_2 = \inf\{n > \tau_1 : M_n \leqslant a\}$, $\tau_2 = \inf\{n > \sigma_2 : X_n \geqslant b\}$, ..., and $U_N := \max\{k : \tau_k \leqslant N\}$.

Lemma 2.9 (Upcrossing inequality). *Suppose that X_N is a submartingale, then*

$$(b - a)\mathbf{E}U_N(a, b) \leqslant \mathbf{E}(X_N - a)^+.$$

PROOF. We only prove the case that $a = 0$ and $X_k \geqslant 0$.

$$X_N = \underbrace{X_{S_1 \wedge N}}_{\geqslant 0} + \underbrace{\sum_{i=1}^\infty X_{T_i \wedge N} - X_{S_i \wedge N}}_{\geqslant bU_N(0,b)} + \sum_{i=1}^\infty \underbrace{X_{S_{i+1} \wedge N} - X_{T_i \wedge N}}_{\text{positive expectation}} .$$

$\square$

Upcrossing inequality leads to

THEOREM 2.10. *If X_n is a submartingale such that $\sup_n \mathbf{E}X_n^+ < \infty$, then X_n converges a.s. as $n \to \infty$.*

Corollary 2.11. *Suppose that $X \in L^1(\mathbf{P}, \Omega)$, $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, then*

$$\lim_{n \to \infty} \mathbf{E}(X|\mathcal{F}_n) = \mathbf{E}(X|\mathcal{F}_\infty), \quad \text{a.s. and in } L^1.$$

Example 2.12. For an example of a discrete martingale, let $\Omega = [0, 1], \mathbf{P}$ Lebesgue measure, and f an integrable function on $[0, 1]$. Let $\mathcal{F}_n$ be the σ-field generated by the sets

$$\{[k/2^n, (k + 1)/2^n), k = 0, 1, \ldots, 2^n - 1\}.$$

Let $f_n = \mathbf{E}[f \mid \mathcal{F}_n]$. If I is an interval in $\mathcal{F}_n$, shows that

$$f_n(x) = \frac{1}{|I|} \int_I f(y)dy \quad \text{if } x \in I.$$

f_n is a particular example of what is known as a dyadic martingale. Of course, $[0, 1]$ could be replaced by any interval as long as we normalize so that the total mass of the interval is 1. We could also divide cubes in $\mathbb{R}^d$ into 2^d subcubes at each step and define f_n analogously. Such martingales are called dyadic martingales. In fact, we could replace Lebesgue measure by any finite measure μ, and instead of decomposing into equal subcubes, we could use any nested partition of sets we like, provided none of these sets had μ measure 0.

All of the above results also hold for all right continuous martingale (submartingales) (see* [Hua01]*).

THEOREM 2.13. *Assume X is a continuous submartingale, then there exists a unique martingale M and a unique continuous increasing adapted process A such that*

$$A_0 = 0, \quad X_t = M_t + A_t.$$

If M is a continuous square integrable martingale, then M^2 is a submartingale. Thus, there exists a continuous increasing process, denoted by $\langle M \rangle$, the **quadratic variation** of M, such that $M^2 - \langle M \rangle$ is a martingale. Particularly, $\mathbf{E}M_t^2 - \mathbf{E}M_0^2 = \mathbf{E}\langle M \rangle_t$.

2.2.1. Stochastic Integral. From now on, unless stated otherwise, our processes have continuous paths.

Lemma 2.14. *Let M_t be a square integrable martingale (that is, $M_t \in L^2$ for every $t \geqslant 0$). Let $0 \leqslant s < t$ and let $s = t_0 < t_1 < \cdots < t_n = t$ be a division of the interval $[s, t]$. Then,*

$$\mathbf{E}\left[\sum_{i=1}^{n} \left(M_{t_i} - M_{t_{i-1}} \right)^2 \mid \mathcal{F}_s \right] = \mathbf{E}\left[M_t^2 - M_s^2 \mid \mathcal{F}_s \right] = \mathbf{E}\left[(M_t - M_s)^2 \mid \mathcal{F}_s \right].$$

PROOF. For every $i = 1, \ldots, n$,

$$\begin{aligned}
\mathbf{E}\left[\left(M_{t_i} - M_{t_{i-1}} \right)^2 \mid \mathcal{F}_s \right] &= \mathbf{E}\left[\mathbf{E}\left[\left(M_{t_i} - M_{t_{i-1}} \right)^2 \mid \mathcal{F}_{t_{i-1}} \right] \mid \mathcal{F}_s \right] \\
&= \mathbf{E}\left[\mathbf{E}\left[M_{t_i}^2 \mid \mathcal{F}_{t_{i-1}} \right] - 2M_{t_{i-1}} \mathbf{E}\left[M_{t_i} \mid \mathcal{F}_{t_{i-1}} \right] + M_{t_{i-1}}^2 \mid \mathcal{F}_s \right] \\
&= \mathbf{E}\left[\mathbf{E}\left[M_{t_i}^2 \mid \mathcal{F}_{t_{i-1}} \right] - M_{t_{i-1}}^2 \mid \mathcal{F}_s \right] \\
&= \mathbf{E}\left[M_{t_i}^2 - M_{t_{i-1}}^2 \mid \mathcal{F}_s \right]
\end{aligned}$$

and the desired result follows by summing over i. $\qquad\square$

We say that M_t if a **local martingale** if there exist stopping times $\tau_n \uparrow \infty$ such that $X_{\tau_n \wedge t}$ is a martingale for each $n \in \mathbb{N}$.

THEOREM 2.15. *Let M_t be a continuous local martingale. 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 continuous local martingale. Furthermore, for every fixed $t > 0$, if $\pi^n = \{(t_0^n, \cdots, t_{k_n}^n) : 0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t\}$ is an increasing sequence of subdivisions of $[0, t]$ with mesh going to 0 , then we have*

$$\langle M \rangle_t = \lim_{n \to \infty} \sum_{i=1}^{k_n} \left(M_{t_i^n} - M_{t_{i-1}^n} \right)^2$$

in probability. The process $\langle M \rangle_t$ is called the quadratic variation of M_t.

Theorem 2.15 is a consequence of the following lemma.

Lemma 2.16. *Let M_t be a continuous bounded martingale. Let $\pi^n = \{(t_0^n, \cdots, t_{k_n}^n) : 0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = T\}$ be an increasing sequence of subdivisions of $[0, T]$ with mesh going to 0 , then for each n,*

$$N_t^n := \sum_{i=1}^{k_n} M_{t_{i-1}}(M_{t_i \wedge t} - M_{t_{i-1} \wedge t})$$

is a martingale, and N_t^n convergent uniformly on compacts, with probability one to some square integrable martingale N_t.

PROOF. It is easy to verify that N_t^n is a martingale. Let us fix $n \leqslant m$ and evaluate the product $\mathbf{E}\left(N_T^n N_T^m\right)$. This product is equal to

$$\sum_{i=1}^{k_n}\sum_{j=1}^{k_m}\mathbf{E}\left[M_{t_{i-1}^n}\left(M_{t_i^n}-M_{t_{i-1}^n}\right)M_{t_{j-1}^m}\left(M_{t_j^m}-M_{t_{j-1}^m}\right)\right].$$

In this double sum, the only terms that may be nonzero are those corresponding to indices i and j such that the interval $\left(t_{j-1}^m,t_j^m\right]$ is contained in $\left(t_{i-1}^n,t_i^n\right]$. Indeed, suppose that $t_i^n \leqslant t_{j-1}^m$ (the symmetric case $t_j^m \leqslant t_{i-1}^n$ is treated in an analogous way).

Then, conditioning on the σ-field $\mathscr{F}_{t_{j-1}^m}$, we have

$$\mathbf{E}\left[M_{t_{i-1}^n}\left(M_{t_i^n}-M_{t_{i-1}^n}\right)M_{t_{j-1}^m}\left(M_{t_j^m}-M_{t_{j-1}^m}\right)\right]$$
$$=\mathbf{E}\left[M_{t_{i-1}^n}\left(M_{t_i^n}-M_{t_{i-1}^n}\right)M_{t_{j-1}^m}\mathbf{E}\left[M_{t_j^m}-M_{t_{j-1}^m}\mid\mathscr{F}_{t_{j-1}^m}\right]\right]=0.$$

For every $j=1,\ldots,k_m$, write $i_{n,m}(j)$ for the unique index i such that $\left(t_{j-1}^m,t_j^m\right]\subset\left(t_{i-1}^n,t_i^n\right]$. It follows from the previous considerations that

$$\mathbf{E}\left[N_T^n N_T^m\right]=\sum_{1\leqslant j\leqslant k_m,i=i_{n,m}(j)}\mathbf{E}\left[M_{t_{i-1}^n}\left(M_{t_i^n}-M_{t_{i-1}^n}\right)M_{t_{j-1}^m}\left(M_{t_j^m}-M_{t_{j-1}^m}\right)\right].$$

In each term $\mathbf{E}\left[M_{t_{i-1}^n}\left(M_{t_i^n}-M_{t_{i-1}^n}\right)M_{t_{j-1}^m}\left(M_{t_j^m}-M_{t_{j-1}^m}\right)\right]$, we can now decompose

$$M_{t_i^n}-M_{t_{i-1}^n}=\sum_{k:i_{n,m}(k)=i}\left(M_{t_k^m}-M_{t_{k-1}^m}\right)$$

and we observe that, if k is such that $i_{n,m}(k)=i$ but $k\neq j$,

$$\mathbf{E}\left[M_{t_{i-1}^n}\left(M_{t_k^m}-M_{t_{k-1}^m}\right)M_{t_{j-1}^m}\left(M_{t_j^m}-M_{t_{j-1}^m}\right)\right]=0$$

(condition on $\mathscr{F}_{t_{k-1}^m}$ if $k>j$ and on $\mathscr{F}_{t_{j-1}^m}$ if $k<j$). The only case that remains is $k=j$, and we have thus obtained

$$\mathbf{E}\left[N_T^n N_T^m\right]=\sum_{1\leqslant j\leqslant k_m,i=i_{n,m}(j)}\mathbf{E}\left[M_{t_{i-1}^n}M_{t_{j-1}^m}\left(M_{t_j^m}-M_{t_{j-1}^m}\right)^2\right].$$

As a special case of this relation, we have

$$\mathbf{E}\left[\left(N_T^m\right)^2\right]=\sum_{1\leqslant j\leqslant k_m}\mathbf{E}\left[M_{t_{j-1}^m}^2\left(M_{t_j^m}-M_{t_{j-1}^m}\right)^2\right].$$

Furthermore,

$$
\begin{aligned}
\mathbf{E}\left[(N_T^n)^2\right] &= \sum_{1 \leqslant i \leqslant n} \mathbf{E}\left[M_{t_{i-1}^n}^2 \left(M_{t_i^n} - M_{t_{i-1}^n}\right)^2\right] \\
&= \sum_{1 \leqslant i \leqslant n} \mathbf{E}\left[M_{t_{i-1}^n}^2 \mathbf{E}\left[\left(M_{t_i^n} - M_{t_{i-1}^n}\right)^2 \mid \mathscr{F}_{i_{i-1}^n}\right]\right] \\
&= \sum_{1 \leqslant i \leqslant k_n} \mathbf{E}\left[M_{t_{i-1}^n}^2 \sum_{j:i_{n,m}(j)=i} \mathbf{E}\left[\left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2 \mid \mathscr{F}_{t_{i-1}^n}\right]\right] \\
&= \sum_{1 \leqslant j \leqslant k_m, i=i_{n,m}(j)} \mathbf{E}\left[M_{t_{i-1}^n}^2 \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right],
\end{aligned}
$$

If we combine the last three displays, we get

$$
\mathbf{E}\left[(N_T^n - N_T^m)^2\right] = \mathbf{E}\left[\sum_{1 \leqslant j \leqslant k_m, i=i_{n,m}(j)} \left(M_{t_{i-1}^n} - M_{t_{j-1}^m}\right)^2 \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right].
$$

Using the Cauchy-Schwarz inequality, we then have

$$
\begin{aligned}
\mathbf{E}\left[(N_T^n - N_T^m)^2\right] \leqslant &\mathbf{E}\left[\sup_{1 \leqslant j \leqslant k_m, i=i_{n,m}(j)} \left(M_{t_{i-1}^n} - M_{t_{j-1}^m}\right)^4\right]^{1/2} \\
&\times \mathbf{E}\left[\left(\sum_{1 \leqslant j \leqslant k_m} \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right)^2\right]^{1/2}.
\end{aligned}
$$

By the continuity of sample paths (together with the fact that the mesh of our subdivisions tends to 0) and dominated convergence, we have

$$
\lim_{n,m \to \infty, n \leqslant m} \mathbf{E}\left[\sup_{1 \leqslant j \leqslant k_m, i=i_{n,m}(j)} \left(M_{t_{i-1}^n} - M_{t_{j-1}^m}\right)^4\right] = 0.
$$

To complete the proof of the lemma, it is then enough to prove the existence of a finite constant C such that, for every m,

$$
\mathbf{E}\left[\left(\sum_{1 \leqslant j \leqslant k_m} \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right)^2\right] \leqslant C.
$$

Let A be a constant such that $|M_t| \leqslant A$ for every $t \geqslant 0$. Expanding the square and using Proposition 3.14 twice, we have

$$\mathbf{E}\left[\left(\sum_{1 \leqslant j \leqslant k_m} \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right)^2\right]$$

$$= \mathbf{E}\left[\sum_{1 \leqslant j \leqslant k_m} \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^4\right] + 2\mathbf{E}\left[\sum_{1 \leqslant j < k \leqslant k_m} \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2 \left(M_{t_k^m} - M_{t_{k-1}^m}\right)^2\right]$$

$$\leqslant 4A^2 \mathbf{E}\left[\sum_{1 \leqslant j \leqslant k_m} \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right]$$

$$+ 2\sum_{j=1}^{k_m-1} \mathbf{E}\left[\left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2 \mathbf{E}\left[\sum_{k=j+1}^{k_m} \left(M_{t_k^m} - M_{t_{k-1}^m}\right)^2 \mid \mathscr{F}_{t_j^m}\right]\right]$$

$$= 4A^2 \mathbf{E}\left[\sum_{1 \leqslant j \leqslant k_m} \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right]$$

$$+ 2\sum_{j=1}^{k_m-1} \mathbf{E}\left[\left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2 \mathbf{E}\left[\left(M_T - M_{t_j^m}\right)^2 \mid \mathscr{F}_{t_j^m}\right]\right]$$

$$\square$$

Let M_t be a square integrable martingale, $0 = t_0 \leqslant t_1 \leqslant \cdots \leqslant t_n = T$ and $H_s(\omega) = \sum_{i=0}^{n-1} H_{t_i}(\omega)\mathbf{1}_{(t_i, t_{i+1}]}(s)$, where F_i is bounded and $\mathcal{F}_{t_i}$-measurable. Define

$$\int_0^t H_s \mathrm{d}M_s := \sum_{i=0}^{n-1} H_{t_i}(M_{t \wedge t_{i+1}} - M_{t \wedge t_i}).$$

Then

Lemma 2.17. $t \mapsto \int_0^t H_s \mathrm{d}M_s$ is a L^2-martingale. Moreover, we have the following Itô isometry:

$$(2.2) \qquad \mathbf{E}\left(\int_0^t H_s \mathrm{d}M_s\right)^2 = \mathbf{E}\int_0^t H_s^2 \mathrm{d}\langle M\rangle_s.$$

PROOF.

$$\mathbf{E}\left(\int_0^1 H_s \mathrm{d}M_s\right)^2 = \mathbf{E}\sum_i H_{t_i}^2 (M_{t_{i+1}} - M_{t_i})^2 + 2\mathbf{E}\sum_{i<j} H_{t_i} H_{t_j} (M_{t_{i+1}} - M_{t_i})(M_{t_{j+1}} - M_{t_j})$$

$$=: I_1 + I_2.$$

$$I_1 = \sum_i \mathbf{E}\mathbf{E}\left(H_{t_i}^2 (M_{t_{i+1}} - M_{t_i})^2 \big| \mathcal{F}_{t_i}\right) = \sum_i \mathbf{E}\left[H_{t_i}^2 \mathbf{E}\left((M_{t_{i+1}} - M_{t_i})^2 \big| \mathcal{F}_{t_i}\right)\right]$$

$$= \sum_i \mathbf{E}H_{t_i}^2 (\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i}) = \mathbf{E}\int_0^1 H_s^2 \mathrm{d}\langle M\rangle_s,$$

$$I_2 = 2 \sum_{i<j} \mathbf{E} \left[H_{t_i} H_{t_j} (M_{t_{i+1}} - M_{t_i}) \mathbf{E} \left((M_{t_{j+1}} - M_{t_j}) \big| \mathcal{F}_{t_j} \right) \right] = 0.$$

Therefore,

$$\mathbf{E} \left(\int_0^1 H_s \mathrm{d}M_s \right)^2 = \mathbf{E} \int_0^1 H_s^2 \mathrm{d}\langle M \rangle_s.$$

$t \mapsto \int_0^t H_s \mathrm{d}W_s = \sum_{i=0}^{n-1} H_{t_i}(W_{t \wedge t_{i+1}} - W_{t \wedge t_i})$ is a continuous martingale. $\qquad \square$

We then can use this to extend the above construction to more general H_s satisfying $\int_0^t H_s^2 \mathrm{d}\langle M \rangle_s < \infty$ by taking limits in L^2. For general continuous local martingale, we can employ standard localization argument to define the above integral. For $X_t = M_t + A_t$, a semimartingale, $\int_0^t H_s \mathrm{d}X_s$ is given by

$$\int_0^t H_s \mathrm{d}X_s = \int_0^t H_s \mathrm{d}M_s + \int_0^t H_s \mathrm{d}A_s,$$

where the first integral on the right is a stochastic integral and the second integral on the right is a Riemann-Stieltjes integral.

Proposition 2.18.

$$\left\langle \int_0^\cdot H_s \mathrm{d}M_s \right\rangle_t = \int_0^t H_s^2 \mathrm{d}\langle M \rangle_s.$$

Let $N_t = \int_0^t H_s \mathrm{d}M_s$. Then

$$\int_0^t K_s \mathrm{d}N_s = \int_0^t K_s H_s \mathrm{d}M_s.$$

2.3. Itô's formula and its applications

We list some important results in stochastic calculus.

THEOREM 2.19 (Itô's formula). *If each X_t^i (for each $i \in 1, \cdots d\}$) is a continuous semimartingale and $f \in C^2(\mathbb{R}^d)$, then*

$$f(X_t) - f(X_0)$$

(2.3)
$$= \int_0^t \sum_{i=1}^d \partial_i f(X_s) \, \mathrm{d}X_s^i + \frac{1}{2} \int_0^t \sum_{i,j=1}^d \partial_{ij} f(X_s) \, \mathrm{d} \left\langle X^i, X^j \right\rangle_s$$

(see [**Hua01**, Theorem 13.5]).

It is often useful to use the language of Stratonovitch's integration to study stochastic differential equations because the Itô's formula takes a much nicer form. If M_t is an $\mathcal{F}_t$-adapted real valued local martingale and if H_t is an $\mathcal{F}_t$-adapted continuous semimartingale satisfying $\mathbf{P} \left(\int_0^T H_s \mathrm{d}\langle M \rangle_s < \infty \right) = 1$, then by definition the Stratonovitch integral of H_t with respect to M_t is defined as

$$\int_0^T H_t \circ \mathrm{d}M_t = \int_0^T H_t \mathrm{d}M_t + \frac{1}{2} \langle H, M \rangle_T.$$

By using Stratonovitch integral instead of Itô's, the Itô formula reduces to the classical change of variable formula.

THEOREM 2.20. *Let M_t be a d–dimensional continuous semimartingale. Let now f be a C^2 function. We have*

$$f(M_t) = f(M_0) + \int_0^t \partial_i f(X_s) \circ \mathrm{d}M_s^i, \quad t \geqslant 0.$$

THEOREM 2.21 (Burkholder-Davis-Gundy inequalities). *If M_t is a continuous martingale with $M_0 = 0$, and τ is a stopping time, then*

(2.4)
$$\mathbf{E} \sup_{t \in [0, \tau]} |M_t|^p \asymp_p \mathbf{E} \langle M \rangle_\tau^{p/2}, \quad p \in (0, \infty)$$

PROOF. *Step 1*: for any $p \geqslant 2$, by Itô's formula

$$|M_\tau|^p = p \int_0^\tau \mathrm{sgn}(M_t)|M_t|^{p-2} M_t \mathrm{d}M_t + \frac{p(p-1)}{2} \int_0^T |M_t|^{p-2} \mathrm{d}\langle M \rangle_t;$$

By Doob's inequality and Hölder's inequality,

$$\mathbf{E}(M_\tau^*)^p \lesssim_p \mathbf{E}|M_\tau|^p \lesssim_p \mathbf{E}((M_\tau^*)^{p-2} \langle M \rangle_\tau)$$

$$\leqslant (\mathbf{E}(M_\tau^*)^p)^{1 - \frac{2}{p}} (\mathbf{E} \langle M \rangle_\tau^{\frac{p}{2}})^{\frac{2}{p}};$$

Step 2: using Lenglart's domination inequality, we can get the proof for the case $p \in (0, 2)$.

We proceed now to the proof of the left hand side inequality. We have,

$$M_t^2 = \langle M \rangle_t + 2 \int_0^t M_s dM_s.$$

Therefore, we get

$$\mathbf{E} \left(\langle M \rangle_T^{\frac{p}{2}} \right) \lesssim \mathbf{E}(M_T^*)^p + \mathbf{E} \left(\sup_{0 \leq t \leq T} \left| \int_0^t M_s dM_s \right|^{p/2} \right).$$

By using the previous argument, we now have

$$2^{\frac{p}{2}} \mathbf{E} \left(\sup_{0 \leq t \leq T} \left| \int_0^t M_s dM_s \right|^{p/2} \right) \leq C\mathbf{E} \left(\left(\int_0^T M_s^2 d\langle M \rangle_s \right)^{p/4} \right)$$

$$\leqslant C\mathbf{E} \left((M_T^*)^{p/2} \langle M \rangle_T^{p/4} \right) \leqslant C \left(\mathbf{E}(M_T^*)^p \right)^{1/2} \left(\mathbf{E} \langle M \rangle_T^{p/2} \right)^{1/2}$$

$$\leqslant \varepsilon' \mathbf{E}(M_T^*)^p + C_{\varepsilon'} \mathbf{E} \langle M \rangle_T^{p/2} \leqslant \varepsilon.$$

As a conclusion, we obtained that d $\hspace{2cm}$ $\square$

Proposition 2.22 (Lenglart). *Let X_t be a positive adapted right-continuous process and A_t be an increasing process. Assume that for every bounded stopping time τ, $\mathbf{E}(X_\tau \mid \mathcal{F}_0) \leqslant \mathbf{E}(A_\tau \mid \mathcal{F}_0)$. Then, for every $\kappa \in (0, 1)$,*

$$\mathbf{E} \left(X_T^* \right)^\kappa \leqslant \frac{2 - \kappa}{1 - \kappa} \mathbf{E} \left(A_T^\kappa \right).$$

We shall use this lemma to prove the following

Another approach to proving (2.4) is utilizing "good-λ" inequality (cf. [**RY13**]).

THEOREM 2.23 (Lévy's theorem). *If X_t is a d-dimensional $(\mathcal{F}_t)_{t \geqslant 0}$-adapted process, each of whose coordinates is a continuous local martingale, and $\langle X^i, X^j \rangle_t = \delta_{ij} t$, then X_t is a d-dimensional $(\mathcal{F}_t)_{t \geqslant 0}$-Brownian motion.*

PROOF. Let $\xi \in \mathbb{R}^d$. Then $\xi \cdot X_t$ is a continuous local martingale with quadratic variation $\langle \xi \cdot X \rangle_t = |\xi|^2 t$. By Itô's formula, $\exp(i\xi \cdot X_t + \frac{1}{2}|\xi|^2 t)$ is a continuous local martingale. This complex continuous local martingale is bounded on every finite interval and is therefore a (true) martingale, in the sense that its real and imaginary parts are both martingales. Hence, for every $s < t$,

$$\mathbf{E}\left[\exp\left(i\xi \cdot X_t + \frac{1}{2}|\xi|^2 t \right) \middle| \mathcal{F}_s \right] = \exp\left(i\xi \cdot X_s + \frac{1}{2}|\xi|^2 s \right)$$

Thus,

$$\mathbf{E}\left[\exp\left(i\xi \cdot (X_t - X_s) \right) \middle| \mathcal{F}_s \right] = \exp\left(\frac{1}{2}|\xi|^2 (t - s) \right).$$

This implies $X_t - X_s$ is independent with $\mathcal{F}_s$ and $X_t - X_s \sim \mathcal{N}(0, t - s)$.

Finally, X is adapted and has independent increments with respect to the filtration $(\mathcal{F}_t)_{t \geqslant 0}$ so that X is a s-dimensional $(\mathcal{F}_t)_{t \geqslant 0}$-Brownian motion. $\square$

Let M_t be a continuous local martingale with $M_0 = 0$. Set $\mathscr{E}(M)_t := \exp(M_t - \langle M \rangle_t / 2)$.

Proposition 2.24. $\mathscr{E}(M)_t$ *is a continuous local martingale, and is the unique solution to*

$$\mathrm{d}X_t = X_t \mathrm{d}M_t, \quad X_0 = 1.$$

THEOREM 2.25 (Girsanov theorem). *Let X_t and M_t be two continuous local martingales under $\mathbb{P}$ with $M_0 = 0$ $\mathbb{P}$-a.s.. Assume that $\mathscr{E}(M)_t$ is a martingale, we define a new probability measure $\mathbb{Q}$ by setting the restriction of $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$ to $\mathcal{F}_t$ to be $\mathscr{E}(M)_t$, then $X_t - \langle X, M \rangle_t$ is a martingale under $\mathbb{Q}$ and the quadratic variation of X_t is the same under $\mathbb{P}$ and $\mathbb{Q}$.*

PROOF. By localization, we can assume X is a martingale. Set $Y_t = X_t - \langle X, M \rangle_t$. We only need to verify that $Y_t \mathscr{E}(M)_t$ is a martingale under $\mathbb{P}$. By Itô's formula,

$$\begin{aligned}
\mathrm{d}Y_t \mathscr{E}(M)_t =& \mathscr{E}(M)_t \mathrm{d}X_t - \mathscr{E}(M)_t \mathrm{d}\langle X, M \rangle_t + Y_t \mathscr{E}(M)_t \mathrm{d}M_t + \mathrm{d}\langle X, \mathscr{E}(M) \rangle_t \\
=& \mathscr{E}(M)_t \mathrm{d}X_t + Y_t \mathscr{E}(M)_t \mathrm{d}M_t.
\end{aligned}$$

Therefore, $Y_t \mathscr{E}(M)_t$ is a martingale, which implies

$$\mathbf{E}_{\mathbb{Q}}(Y_t; A) = \mathbf{E}_{\mathbb{Q}}(Y_s; A), \quad \forall A \in \mathcal{F}_s,$$

i.e.

$$\mathbf{E}_{\mathbb{Q}}(Y_t | \mathcal{F}_s) = Y_s.$$

$\square$

THEOREM 2.26 (Dambis-Dubins-Schwarz's Theorem). *Let M be a continuous local martingale with respect to a filtration $(\mathcal{F}_t)_{t \geqslant 0}$, such that $M_0 = 0$ and $\langle M \rangle_\infty = \infty$ almost surely. For all $t \geqslant 0$, let*

$$T_t = \inf\{ s \geqslant 0 : \langle M \rangle_s > t \} = \langle M \rangle_t^{-1}$$

be the generalized inverse of the non-decreasing process $\langle M \rangle$ issued from 0. Then

(i) $B = (M_{T_t})_{t \geqslant 0}$ is a Brownian motion with respect to the filtration $(\mathcal{F}_{T_t})_{t \geqslant 0}$.
(ii) $(B_{\langle M \rangle_t})_{t \geqslant 0} = (M_t)_{t \geqslant 0}$.

Since $\langle M \rangle$ can be flat on an interval, the map $t \mapsto T_t$ can be discontinuous. But this does not contradict the continuity of $t \mapsto M_{T_t}$. Indeed, the flatness lemma states that

Lemma 2.27 (Flatness Lemma). *M and $\langle M \rangle$ are constant on the same intervals in the sense that almost surely, for all $0 \leqslant a < b$,*

$$\forall t \in [a, b], M_t = M_a \quad \text{if and only if} \quad \langle M \rangle_b = \langle M \rangle_a.$$

PROOF. Since M and $\langle M \rangle$ are continuous, it suffices to show that for all $0 \leqslant a \leqslant b$, almost surely,

$$\{\forall t \in [a, b] : M_t = M_a\} = \{\langle M \rangle_b = \langle M \rangle_a\}.$$

The inclusion $\subset$ comes from the approximation of the quadratic variation. Let us prove the converse. To this end, we consider the continuous local martingale $(N_t)_{t \geqslant 0} = (M_t - M_{t \wedge a})_{t \geqslant 0}$. We have

$$\langle N \rangle = \langle M \rangle - 2\langle M, M^a \rangle + \langle M^a \rangle = \langle M \rangle - 2\langle M \rangle^a + \langle M \rangle^a = \langle M \rangle - \langle M \rangle^a.$$

For all $\epsilon > 0$, we set the stopping time $T_\epsilon = \inf\{t \geqslant 0 : \langle N \rangle_t > \epsilon\}$. The continuous semi-martingale N^{T_ϵ} satisfies $N_0^{T_\epsilon} = 0$ and $\langle N^{T_\epsilon} \rangle_\infty = \langle N \rangle_{T_\epsilon} \leqslant \epsilon$. It follows that N^{T_ϵ} is a martingale bounded in L^2, and for all $t \geqslant 0$,

$$\mathbf{E}(N_{t \wedge T_\epsilon}^2) = \mathbf{E}(\langle N \rangle_{t \wedge T_\epsilon}) \leqslant \epsilon.$$

Let us define the event $A = \{\langle M \rangle_b = \langle M \rangle_a\}$. Then $A \subset \{T_\epsilon \geqslant b\}$ and, for all $t \in [a, b]$,

$$\mathbf{E}(1_A N_t^2) = \mathbf{E}(1_A N_{t \wedge T_\epsilon}^2) \leqslant \mathbf{E}(N_{t \wedge T_\epsilon}^2) \leqslant \epsilon.$$

By sending ϵ to 0 we obtain $\mathbf{E}(1_A N_t^2) = 0$ and thus $N_t = 0$ almost surely on A. This ends the proof of the flatness lemma, which is of independent interest. $\square$

PROOF OF THEOREM 2.26. For all $t \geqslant 0$, the random variable T_t is a stopping time with respect to $(\mathcal{F}_u)_{u \geqslant 0}$, and $s \mapsto T_s$ is non-decreasing. It follows that for all $0 \leqslant s \leqslant t$, $\mathcal{F}_{T_s} \subset \mathcal{F}_{T_t}$, and thus $(\mathcal{F}_{T_u})_{u \geqslant 0}$ is a filtration. Moreover, for all $t \geqslant 0$, T_t is a stopping time for the filtration $(\mathcal{F}_{T_u})_{u \geqslant 0}$. We have $T_t < \infty$ for all $t \geqslant 0$ on the almost sure event $\{\langle M \rangle_\infty = \infty\}$. By construction, $(T_t)_{t \geqslant 0}$ is right continuous, non-decreasing (and thus with left limits), and adapted with respect to $(\mathcal{F}_{T_t})_{t \geqslant 0}$. Since M is continuous, $B = (M_{T_t})_{t \geqslant 0}$ is right continuous with left limits. Moreover, for all $t \geqslant 0$,

$$B_{t^-} = \lim_{s \to t^-} B_s = M_{T_{t^-}}.$$

By the flatness lemma, almost surely $B_{t^-} = B_t$ for all $t \geqslant 0$, hence B is continuous.

Let us show that B is a Brownian motion for $(\mathcal{F}_{T_t})_{t \geqslant 0}$. For all $n \geqslant 0$, M^{T_n} is a continuous local martingale issued from the origin and $\langle M^{T_n} \rangle_\infty = \langle M \rangle_{T_n} = n$ almost surely. It follows that for all $n \geqslant 0$, the processes

$$M^{T_n} \quad \text{and} \quad (M^{T_n})^2 - \langle M \rangle^{T_n}$$

are uniformly integrable martingales. Now, for all $0 \leqslant s \leqslant t \leqslant n$, and by the Doob stopping theorem for uniformly integrable martingales, using $T_s \leqslant T_t \leqslant T_n$,

$$\mathbf{E}(B_t | \mathcal{F}_{T_s}) = \mathbf{E}(M_{T_t}^{T_n} | \mathcal{F}_{T_s}) = M_{T_s}^{T_n} = M_{T_n \wedge T_s} = B_s$$

and similarly, using additionally the property $\langle M \rangle_{T_t}^{T_n} = \langle M \rangle_{T_n \wedge T_t} = \langle M \rangle_{T_t} = t$,

$$\mathbf{E}(B_t^2 - t | \mathcal{F}_{T_s}) = \mathbf{E}((M_{T_t}^{T_n})^2 - \langle M^{T_n} \rangle_{T_t} | \mathcal{F}_{T_s}) = (M_{T_s}^{T_n})^2 - \langle M^{T_n} \rangle_{T_s} = B_s^2 - s.$$

Thus, B and $(B_t^2 - t)_{t \geqslant 0}$ are martingales with respect to the filtration $(\mathcal{F}_{T_t})_{t \geqslant 0}$. It follows now from the Lévy characterization that B is a Brownian motion for $(\mathcal{F}_{T_t})_{t \geqslant 0}$.

Let us show that $M = B_{\langle M \rangle}$. By definition of B, almost surely, for all $t \geqslant 0$,

$$B_{\langle M \rangle_t} = M_{T_{\langle M \rangle_t}}.$$

Now, $T_{\langle M \rangle_t} \leqslant t \leqslant T_{\langle M \rangle_t}$, and since $\langle M \rangle$ takes the same value at $T_{\langle M \rangle_t}$ and $T_{\langle M \rangle_t}$, we get $t = T_{\langle M \rangle_t}$, and the flatness lemma gives $M_t = M_{T_{\langle M \rangle_t}}$ for all $t \geqslant 0$ almost surely. In other words, using the definition of B, this means that almost surely, for all $t \geqslant 0$,

$$M_t = M_{T_{\langle M \rangle_t}} = B_{\langle M \rangle_t}.$$

$\square$

2.4. Stochastic Differential Equations

One of the main object in this note is the following SDE:

$$(2.5) \qquad \mathrm{d}X_t^i = \sigma_k^i(X_t)\mathrm{d}W_t^k + b^i(X_t)\mathrm{d}t, \quad X_0 = \xi \in \mathcal{F}_0.$$

Given $(\Omega, \mathcal{F}, \mathbf{P}, (\mathcal{F}_t)_{t \geqslant 0}, W_t)$, we say (2.5) has a **pathwise solution** if there exists a continuous $\mathcal{F}_t$-adapted process X_t satisfying (2.5). We say that we have **pathwise uniqueness** for (2.5) if whenever X_t and Y_t are two solutions, then there exists a set $\mathcal{N}$ such that $\mathbf{P}(\mathcal{N}) = 0$ and for all $\omega \notin \mathcal{N}$, we have $(X_t(\omega))_{t \geqslant 0} = (Y_t(\omega))_{t \geqslant 0}$.

2.4.1. Lipschitz conditions.

THEOREM 2.28 (Itô). *Suppose σ and b are Lipschitz. Then there exists a unique pathwise solution to the SDE* (2.5) *for any $X_0 \in L^2(\Omega, \mathcal{F}_0, \mathbf{P})$.*

PROOF. Let $\mathcal{B}$ denote the set of all continuous processes ξ that are adapted to the filtration $\mathcal{F}_t$ and satisfy

$$\|\xi\|_{\mathcal{B}} := \left(\mathbf{E} \sup_{t \in [0,T]} |\xi_t|^2 \right)^{1/2} < \infty.$$

Here T is a positive number which will be determined later. It is not hard to verify that $\mathcal{B}$ is a Banach space. Define a map $\mathscr{A}$ on $\mathcal{B}$ by

$$(\mathscr{A}(\xi))_t := X_0 + \int_0^t \sigma(\xi_s) \cdot \mathrm{d}W_s + \int_0^t b(\xi_s)\mathrm{d}s, \quad t \in [0,T].$$

(Verify that $\mathcal{A}(\xi) \in \mathcal{B}$). By (2.4) (or Doob's inequality) and Lipschitz condition on the coefficients,

$$\|\mathscr{A}(\xi) - \mathscr{A}(\eta)\|_{\mathcal{B}}^2 = \mathbf{E} \sup_{t \in [0,T]} |\mathscr{A}(\xi)_t - \mathscr{A}(\eta)_t|^2$$

$$\leqslant 2\mathbf{E} \sup_{t \in [0,T]} \left| \int_0^t (\sigma(\xi_s) - \sigma(\eta_s)) \mathrm{d}W_s \right|^2 + 2\mathbf{E} \sup_{t \in [0,T]} \left| \int_0^t (b(\xi_s) - b(\eta_s)) \mathrm{d}s \right|^2$$

$$\overset{(2.4)}{\leqslant} C\mathbf{E} \int_0^T |\sigma(\xi_s) - \sigma(\eta_s)|^2 \mathrm{d}s + C\mathbf{E} \left(\int_0^T |b(\xi_s) - b(\eta_s)| \mathrm{d}s \right)^2$$

$$\leqslant C(T + T^2)\mathbf{E} \sup_{t \in [0,T]} |\xi_t - \eta_t|^2 = C_1(T + T^2)\|\xi - \eta\|_{\mathcal{B}}^2.$$

Choosing $T > 0$ sufficiently small such that $C_1(T + T^2) \leqslant 1/2$, then $\mathscr{A}$ is a Contraction mapping on $\mathcal{B}$. Banach fixed-point theorem yields that $\mathscr{A}$ has a unique fixed point, which is the unique pathwise solution to (2.5). We can extend the same result to arbitrarily time intervals. $\square$

2.4.2. Definitions of solutions.

(1) **strong solution exists** to (2.5): if given the Brownian motion W_t there exists a process X_t satisfying (2.5) such that X_t is adapted to the filtration generated by W_t.
(2) **weak solution exists** to (2.5): if there exists $(\Omega, \mathcal{F}, \mathbf{P}, (\mathcal{F}_t)_{t \geqslant 0}; X_t, W_t)$ such that W_t is a $(\mathcal{F}_t)_{t \geqslant 0}$-Brownian motion and the equation (2.5) holds.
(3) **weak uniqueness**: if whenever $(\Omega, \mathcal{F}, \mathbf{P}, (\mathcal{F}_t)_{t \geqslant 0}; X_t, W_t)$ and $(\Theta, \mathcal{G}, \mathbf{Q}, (\mathcal{G}_t)_{t \geqslant 0}; Y_t, B_t)$ are two weak solutions, then the laws of the processes X and Y are equal; **Joint uniqueness in law** means the joint law of (X, W) and (Y, B) are equal.

A fundamental result is

THEOREM 2.29 (Yamada-Watanabe-Engelbert [**Eng91**]). *The following two conditions are equivalent.*

(i) *For every initial distribution, there exists a weak strong solution to (2.5) and the solution to (2.5) is pathwise unique.*
(ii) *For every initial distribution, there exists a strong strong solution to (2.5) and the solution to (2.5) is jointly unique in law.*

If one (and therefore both) of these conditions is satisfied then every solution to (2.5) is a strong solution.

Proposition 2.30.

2.4.3. SDEs with Hölder drifts.
For strong well-posedness, if the diffusion coefficient σ is non-degenerate, then the condition on b can be weakened.

Let $\delta \in (0, 1)$. Define

$$\mathbb{S}_{\delta}^d = \left\{ A \in S(d) : \delta I_d \leqslant A \leqslant \delta^{-1} I_d \right\}$$

THEOREM 2.31 (Krylov [**Kry21b**]). *Suppose that $a \in \mathbb{S}_{\delta}^d$ and $\nabla\sigma, b \in L^d(\mathbb{R}^d)$, then equation (2.5) admits a unique strong solution.*

Of course, we will not to prove such a strong result here, but a simper one below.

THEOREM 2.32 (Flandoli-Gubinelli-Priola [**FGP10**]). *Equation* (2.5) *admits a unique strong solution, provided that* $a \in \mathbb{S}_\delta^d$, σ *is Lipschitz, and* $b \in C^\alpha(\mathbb{R}^d)$ ($\forall \alpha > 0$).

Let

$$Lu = a_{ij}\partial_{ij}u + b_i\partial_i u.$$

We will consider

$$(2.6) \qquad\qquad \lambda u - Lu = f, \quad \lambda > 0.$$

We need the following apriori estimate.

Lemma 2.33. *Suppose* $a \in \mathbb{S}_\delta^d$ *and* $a, b \in C^\alpha$. *Then for any* $\lambda > 0$ *and* $u \in C^{2,\alpha}$, *it holds that*

$$(2.7) \qquad\qquad \lambda\|u\|_\alpha + \|u\|_{2+\alpha} \leqslant C\|\lambda u - Lu\|_\alpha,$$

where C *only depends on* d, δ, α, *and* $\|a\|_\alpha$ *and* $\|b\|_\alpha$.

PROOF. The proof for the above lemma for $L = \Delta$ can be founded in Appendix 7.3.

For general elliptic operators, let $\zeta \in C_c^\infty(B_2)$ such that $\zeta \geqslant 0$, $\zeta \equiv 1$ in B_1. Set $\zeta_\varepsilon^z = \zeta((x-z)/\varepsilon)$, and $f = \lambda u - Lu$. Then

$$\lambda(u\zeta_\varepsilon^z) - a_{ij}(z)\partial_{ij}(u\zeta_\varepsilon^z)$$
$$= (a_{ij} - a_{ij}(z))\partial_{ij}(u\zeta_\varepsilon^z) - 2a_{ij}\partial_i u\partial_j\zeta_\varepsilon^z - a_{ij}\partial_{ij}\zeta_\varepsilon^z u$$
$$+ b_i\partial_i u\, \zeta_\varepsilon^z + f\zeta_\varepsilon^z,$$

In virtue of Lemma 7.9, we have

$$\lambda[u\zeta_\varepsilon^z]_\alpha + [\nabla^2(u\zeta_\varepsilon^z)]_\alpha \leqslant C\varepsilon[\nabla^2(u\zeta_\varepsilon^z)]_\alpha + C[f]_\alpha + C\varepsilon^{-\alpha}\|f\|_0$$
$$+ C\|\nabla^2 u\|_0 + C\varepsilon^{-1}[\nabla u]_\alpha + C\varepsilon^{-1-\alpha}\|\nabla u\|_0 + C\varepsilon^{-2}[u]_\alpha + C\varepsilon^{-2-\varepsilon}\|u\|_0.$$

Here we use the fact that

$$[fg]_\alpha \leqslant \|f\|_0[g]_\alpha + [f]_\alpha\|g\|_0.$$

Choosing $\varepsilon_0 > 0$ sufficiently small so that $C\varepsilon_0 \leqslant 1/2$, we get

$$\lambda[u]_\alpha + [\nabla^2 u]_\alpha \leqslant \sup_{z\in\mathbb{R}^d}\left(\lambda[u\zeta_{\varepsilon_0}^z]_\alpha + [\nabla^2(u\zeta_{\varepsilon_0}^z)]_\alpha\right) \leqslant C_{\varepsilon_0}(\|f\|_\alpha + \|u\|_2).$$

Noting that $\|u\|_2 \leqslant \delta[\nabla^2 u]_\alpha + C_\delta\|u\|_0$, $\delta > 0$, we obtain that

$$\lambda[u]_\alpha + [\nabla^2 u]_\alpha \leqslant C(\|f\|_\alpha + \|u\|_0).$$

Since $\lambda\|u\|_0 \leqslant \|f\|_0$, by interpolation, one sees that

$$\lambda\|u\|_\alpha + \|u\|_{2+\alpha} \leqslant C(1 + \lambda^{-1})\|f\|_\alpha.$$

So we obtain our desired assertion. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

THEOREM 2.34. *Suppose* $a \in \mathbb{S}_\delta^d$ *and* $a, b \in C^\alpha$. *Then for any* $\lambda > 0$ *and* $f \in C^\alpha$, *equation* (2.6) *admits a unique solution in* $C^{2,\alpha}$. *Moreover,*

$$(2.8) \qquad\qquad \lambda\|u\|_\alpha + \|u\|_{2+\alpha} \leqslant C(1 + \lambda^{-1})\|f\|_\alpha,$$

where C *only depends on* d, δ, α, *and* $\|a\|_\alpha$ *and* $\|b\|_\alpha$.

Sketch of the proof for Theorem 2.34:

(i) If $L = \Delta$ and $f \in \mathscr{S}(\mathbb{R}^d)$, then for each $\lambda > 0$, one can use Fourier transformation to solve (2.6), i.e. $u = \mathcal{F}^{-1}\left[\mathcal{F}(f) \cdot (\lambda + 4\pi^2|\cdot|^2)\right] \in \cap_{s>0}H^s \subseteq C_b^\infty$. Moreover, (2.8) can also be proved by Fourier analysis method (see Appendix 7.3);

(ii) For any L satisfying the conditions in Theorem 2.34, and any $u \in C^{2,\alpha}$, by Lemma 2.33, (2.8) holds true for any $\lambda > 0$;

(iii) Let χ be a cutoff function and ζ be a mollifier. For any $f \in C^\alpha$, we set $f_\varepsilon = \chi_\varepsilon(f * \zeta_\varepsilon)$. Here $\chi_\varepsilon(x) = \chi(x/\varepsilon)$ and $\zeta_\varepsilon(x) = \varepsilon^{-d}\zeta(x/\varepsilon)$. Using (i), for each $\varepsilon > 0$, there is a smooth soluiton, say u_ε, to (2.6) with L and f replaced by Δ and f_ε. The limit of (u_ε), u, satisfies $\lambda u - \Delta u = f$, and u also satisfies (2.8);

(iv) In the light of (2.8) and the method of continuity (see Lemma 2.35 below), one can obtain the solvability of (2.6) in $C^{2,\alpha}$.

Lemma 2.35 (Method of continuity). *Let B be a Banach space, V a normed vector space, and T_t a norm continuous family of bounded linear operators from B into V. Assume that there exists a positive constant C such that for every $t \in [0,1]$ and every $x \in B$,*

$$\|x\|_B \leqslant C\|T_t x\|_V.$$

Then T_0 is a surjective if and only if T_1 is surjective as well.

PROOF OF THEOREM 2.32. Since σ and b are bounded continuous, weak solution exists to (2.5) (see [**Hua01**]). Thanks to Theorem 2.29, we only need to prove the pathwise uniqueness.

Let $\lambda \gg 1$ Consider the following equation

$$\lambda \mathbf{u}_\lambda - L\mathbf{u}_\lambda = b.$$

By Lemma 2.33 and interpolation theorem

$$\|\nabla \mathbf{u}\|_0 \leqslant \|\mathbf{u}\|_0^{\frac{1}{2}}\|\nabla^2 \mathbf{u}\|_0^{\frac{1}{2}} \leqslant C\lambda^{-\frac{1}{2}}\|b\|_\alpha.$$

Choosing λ sufficiently large so that $C\lambda^{-\frac{1}{2}} < 1/2$. Set $\phi(x) = x + \mathbf{u}(x)$, then $\phi : \mathbb{R}^d \to \mathbb{R}^d$ is a $C^{1,\alpha}$-homeomorphism.

Assume that X and X' are two solutions to (2.5). Set $Y_t = \phi(X_t)$ and $Y_t' = \phi(X_t')$. Then by Itô's formula,

$$dY_t^i = (\delta_j^i + \partial_j \mathbf{u}^i)(X_t)\sigma_{jk}(X_t)dW_t^k + \left[a_{jk}(X_t)\partial_{jk}\mathbf{u}^i(X_t) + (\delta_j^i + \partial_j\mathbf{u}^i)(X_t)b^j(X_t)\right]dt$$

i.e.

$$\begin{aligned}
dY_t &= [(I + \nabla\mathbf{u})\sigma] \circ \phi^{-1}(Y_t)dW_t + \left[a : \nabla^2\mathbf{u} + (I + \nabla\mathbf{u})b\right] \circ \phi^{-1}(Y_t)dt \\
&= \underbrace{[(I + \nabla\mathbf{u})\sigma] \circ \phi^{-1}}_{=:\widetilde{\sigma}}(Y_t)dW_t + \underbrace{\lambda u \circ \phi^{-1}}_{=:\widetilde{b}}(Y_t)dt.
\end{aligned}$$

Similarly, $dY_t' = \widetilde{\sigma}(Y_t')dW_t + \widetilde{b}(Y_t')dt$. Since $\widetilde{\sigma}$ and $\widetilde{b}$ are both $C^{1,\alpha}$ functions, as in the proof for Theorem 2.28, we have

$$\mathbf{E}|Y_t - Y_t'|^2 \leqslant C\int_0^t \mathbf{E}|Y_s - Y_s'|^2 ds.$$

This yields $Y_t = Y_t'$, due to Gronwall's inequality. Since ϕ is one-to-one, $X_t = X_t'$. $\qquad\square$

2.4.4. Stochastic Flow. Consider (2.5) .

THEOREM 2.36. *If σ and b are Lipschitz, then there exists a version of $X_t(x)$ that is continuous in (t, x) a.s.*

PROOF.

$$X_t(x) - X_t(y) = x - y + \int_0^t [\sigma(X_s(x)) - \sigma(X_s(y))]\, \mathrm{d}W_s + \int_0^t [b(X_s(x)) - b(X_s(y))]\, \mathrm{d}s$$

By the Burkholder-Davis-Gundy inequalities, for any $t \in [0, 1]$ and $p \geqslant 2$,

$$\mathbf{E} \sup_{s \in [0,t]} \left| \int_0^s [\sigma(X_r(x)) - \sigma(X_r(y))]\, \mathrm{d}W_r \right|^p$$

$$\leqslant C\mathbf{E} \left(\int_0^t |X_s(x) - X_s(y)|^2 \mathrm{d}s \right)^{p/2}$$

$$\leqslant C\mathbf{E} \int_0^t |X_s(x) - X_s(y)|^p \mathrm{d}s.$$

Set $g(t) = \mathbf{E} \sup_{s \in [0,t]} |X_s(x) - X_s(y)|^p$. Then for any $T > 0$, we have

$$g(t) \leqslant C|x - y|^p + C \int_0^t g(s)\mathrm{d}s, \quad t \in [0, T],$$

where C only depends on d, p and T. Gronwall's inequality yields

$$\mathbf{E} \sup_{t \in [0,1]} |X_t(x) - X_t(y)|^p \leqslant C|x - y|^p, \quad \forall p \geqslant 2.$$

Further, one can verify that

$$\mathbf{E}\, |X_t(x) - X_s(y)|^p \leqslant C \left(|x - y| + |t - s|^{\frac{1}{2}} \right)^p, \quad x, y \in \mathbb{R}^d,\ t, s \in [0, T],\ p \geqslant 2.$$

This together with Kolmogorov's continuity theorem implies that there is a continuous version of $(t, x) \mapsto X_t(x)$ such that

$$\|X(\omega)\|_{C^\alpha([0,1];\dot{C}^\beta(B_R))} \leqslant K(\omega)$$

with $\alpha \in (0, 1/2)$ and $\beta \in (0, 1)$, and $K \in L^p$ for all $p \geqslant 1$. $\qquad\square$

Remark 2.37. The above result also holds if σ and b are ω-dependent and $\|\sigma\|_{C^1} + \|b\|_{C^1} \leqslant L$ a.s., for some constant L.

The collection of processes $X_t(x)$ is called a flow. If σ and b are smoother functions, then $X_t(x)$ will be smoother in x. Taking derivative, and using the chain rule, formally we have

$$\partial_j X_t^i(x) = \delta_j^i + \int_0^t \partial_l \sigma_k^i(X_s(x))\partial_j X_s^l(x)\mathrm{d}W_s^k$$

$$+ \int_0^t \partial_l b^i(X_s(x))\partial_j X_s^l(x)\mathrm{d}s.$$

Suppose that σ and b are in C_b^1, we consider the SDE

$$(2.9)\ \ \mathrm{d}J_j^i(t, x) = \partial_l \sigma_k^i(X_t(x))J_j^l(t, x)\mathrm{d}W_t^k + \partial_l b^i(X_t(x))J_j^l(t, x)\mathrm{d}t, \quad J_j^i(0, x) = \delta_j^i.$$

Follow the proof of Theorem 2.36, we have

Proposition 2.38. *Assume* $\sigma, b \in C_b^2$. *A strong solution to* (2.9) *exists and is pathwise unique. The solution has moments of all orders. Moreover,* $J(t, x)$ *has a Hölder continuous version, and*

$$\mathbf{E} \sup_{t \in [0,T]} |J(t,x) - J(t,y)|^p \leqslant C|x-y|^p, \quad x, y \in \mathbb{R}^d, \ p \geqslant 1.$$

Exercise 2.39. Prove Proposition 2.38.

We now prove the differentiability of $X_t(x)$.

THEOREM 2.40. *Suppose* $\sigma, b \in C_b^k$. *Then* $x \mapsto X_t(x)$ *is* $C^{k-1,\alpha}$ *a.s., and* $\nabla X_t(x) = J(t,x)$.

PROOF. For simplicity, we take $b = 0$ and $k = 2$. Then

$$X_t^i(x) = x^i + \int_0^t \sigma_k^i(X_s(x)) \mathrm{d}W_s^k.$$

Set

$$S_j^i(t,x,h) := X_t^i(x + e_j h), \quad Y_j^i(t,x,h) := \frac{X_t^i(x + e_j h) - X_t^i(x)}{|h|}, \quad h \neq 0.$$

Then

$$S_j^i(t,x,h) = x^i + e_j h + \int_0^t \sigma_k^i(S_j(s,x,h)) \mathrm{d}W_s^k$$

and

$$Y_j^i(t,x,h) = \delta_j^i + \int_0^t \underbrace{\left[\int_0^1 \partial_l \sigma_k^i(\theta S_j(s,x,h) + (1-\theta)X_s(x)) \mathrm{d}\theta\right]}_{\leqslant \|\nabla \sigma\|_0} Y_j^l(s,x,h) \mathrm{d}W_s^k.$$

Set

$$Z(t,x,h) = (X(t,x), S(t,x,h), Y(t,x,h)).$$

Then

$$Z(0,x,h) = \left(x, (x^i + e_j h), (\delta_j^i)\right)$$

and Z satisfies an SDE with Lipschitz continuous coefficients. Noting that

$$|Z(0,x,h) - Z(0,x',h')| \leqslant C(|x-x'| + |h-h'|),$$

following the arguments in Theorem 2.36, we can obtain that

$$\mathbf{E} \sup_{t \in [0,T]} |Z(t,x,h) - Z(t,x',h')|^p \leqslant C(|x-x'|^p + |h-h'|^p), \quad p \gg 1.$$

This implies that $Z(t,x,h)$ admits a locally Hölder continuous version. Consequently, for almost every $\omega \in \Omega$, the limit $\lim_{h \to 0} Y(t,x,h)(\omega)$ exists for each $t \geqslant 0$ and $x \in \mathbb{R}^d$. Furthermore, it is straightforward to verify that this limit coincides with J, as both satisfy the same equation. $\qquad\square$

One can also show (see Ikeda and Watanabe [**IW14**]) that the map $x \mapsto X_t(x)$ is one-to-one and onto $\mathbb{R}^d$.

Path properties of Itô Processes

3.1. Some Properties of Brownian Motion

Let $d = 1$.

Lemma 3.1. *Let (W_t) be a 1-dimensional Brownian motion. For any $0 < \delta \leqslant 10^{-4}$, it holds that*

(i)

$$(3.1) \qquad \inf_{|x| \leqslant 1/3} \mathbf{P}_x \left(\sup_{t \in [0,\delta]} |W_t| \leqslant 1 \right) \geqslant 5/6.$$

(ii)

$$(3.2) \qquad \inf_{|x| \leqslant 1/3} \mathbf{P}_x(|W_\delta| \leqslant 1/3) \geqslant 1/3$$

PROOF. For (3.1), using Doob's inequality, we have

$$\mathbf{P}_0 \left(\sup_{t \in [0,\delta]} |W_t| > \frac{2}{3} \right) \leqslant \frac{3}{2} \mathbf{E}_0 |W_\delta| \leqslant \frac{3\sqrt{\delta}}{2}.$$

Thus,

$$\mathbf{P}_0 \left(\sup_{t \in [0,\delta]} |W_t| > \frac{2}{3} \right) \leqslant \frac{1}{6}, \quad \delta < 10^{-2},$$

which yields

$$\inf_{|x| \leqslant 1/3} \mathbf{P}_x \left(\sup_{t \in [0,\delta]} |W_t| \leqslant 1 \right) \geqslant \mathbf{P}_0 \left(\sup_{t \in [0,\delta]} |W_t| \leqslant \frac{2}{3} \right) \geqslant \frac{5}{6}, \quad \delta < 10^{-2}.$$

For (3.2), we have

$$\inf_{|x| \leqslant 1/3} \mathbf{P}_x(|W_\delta| \leqslant 1/3) \geqslant \mathbf{P}_0(0 \leqslant W_\delta \leqslant 1/3) \geqslant \frac{1}{3}, \quad 0 < \delta \leqslant 10^{-4}.$$

$\square$

Proposition 3.2. *Let W be a 1-dimensional Brownian motion. For any $\varepsilon > 0$ and $T > 0$, there is a constant $c(\varepsilon, T) > 0$ such that*

$$\mathbf{P}_0 \left(\sup_{t \in [0,T]} |W_t| \leqslant \varepsilon \right) \geqslant c(\varepsilon, T).$$

PROOF. By the scaling property of Brownian motion ($\varepsilon^{-1}W_t \overset{d}{=} W_{\varepsilon^{-2}t}$), we only need to show

$$\mathbf{P}_0 \left(\sup_{t \in [0, T\varepsilon^{-2}]} |W_t| \leqslant 1 \right) \geqslant c(\varepsilon, T) > 0.$$

Set $\delta = 10^{-4}$. (3.1) and (3.2) imply that

$$\inf_{|x| \leqslant 1/3} \mathbf{P}_x \left(\sup_{t \in [0, \delta]} |W_t| \leqslant 1, \ |W_\delta| \leqslant 1/3 \right) \geqslant \frac{1}{6}.$$

Letting $k = [T\varepsilon^{-2}\delta^{-1}] = [10^4 T\varepsilon^{-2}]$, we have

$$\mathbf{P}_0 \left(\sup_{t \in [0, T\varepsilon^{-2}]} |W_t| \leqslant 1 \right)$$

$$\geqslant \mathbf{P}_0 \left(\sup_{t \in [i\delta, (i+1)\delta]} |W_t| \leqslant 1 \ \& \ |W_{i\delta}| \leqslant 1/3, \quad i = 0, 1, \cdots k \right)$$

$$\geqslant 6^{-k} =: c(\varepsilon, T) > 0.$$

$\square$

Proposition 3.3. *Let W be a 1-dimensional Brownian motion. Then for any $\lambda, t > 0$*

$$\mathbf{P} \left(\sup_{s \in [0, t]} |W_s| > \lambda \right) \leqslant 2 e^{-\frac{\lambda^2}{2t}}$$

PROOF. Let $X_t = e^{a|W_t|}$ with $a > 0$. Since $x \mapsto e^{a|x|}$ is a convex function, X_t is a submartingale. By Doob's inequality, we have

$$\mathbf{P}(W_t^* > \lambda) = \mathbf{P}\left(X_t^* > e^{a\lambda} \right) \leqslant e^{-a\lambda} \mathbf{E} X_t = \frac{2e^{-a\lambda}}{\sqrt{2\pi t}} \int_0^\infty e^{ax - \frac{x^2}{2t}} \, \mathrm{d}x = 2 e^{\frac{a^2 t}{2} - a\lambda}.$$

Taking $a = \lambda/t$, we obtain

$$\mathbf{P}(W_t^* > \lambda) \leqslant 2 e^{-\frac{\lambda^2}{2t}}.$$

$\square$

Corollary 3.4 (Exponential martingale inequality). *Let M_t be a continuous martingale with $M_0 = 0$, and τ be a bounded stopping time. Then*

$$\mathbf{P} \left(\sup_{t \leqslant \tau} |M_t| > \lambda \ \& \ \langle M \rangle_\tau \leqslant \mu \right) \leqslant 2 e^{-\frac{\lambda^2}{2\mu}}.$$

PROOF. By Dambis-Dubins-Schwarz Theorem, M_t is a time change of a Brownian motion W_t. So the desired probability is bounded by

$$\mathbf{P} \left(\sup_{t \leqslant T} |W_t| > \lambda \ \& \ \langle W \rangle_T < \mu \right),$$

where T is a stopping time. Since $\langle W \rangle_T = T$, the probability above is in turn bounded by

$$\mathbf{P} \left(\sup_{t \leqslant \mu} |W_t| > \lambda \right) \leqslant 2 e^{-\frac{\lambda^2}{2\mu}},$$

due to Proposition 3.3.

$\square$

The next result, which is known as the law of iterated logarithm shows in particular that Brownian paths are not $\frac{1}{2}$-Hölder continuous.

THEOREM 3.5 (law of iterated logarithm). *Let $(W_t)_{t \geq 0}$ be a Brownian motion. For $s \geq 0$,*

$$\mathbf{P}\left(\liminf_{t \to 0} \frac{W_{t+s} - W_s}{\sqrt{2t \log\log\frac{1}{t}}} = -1, \limsup_{t \to 0} \frac{W_{t+s} - W_s}{\sqrt{2t \log\log\frac{1}{t}}} = 1\right) = 1.$$

PROOF. Thanks to the symmetry and invariance by translation of the Brownian motion, it suffices to show that:

$$\mathbf{P}\left(\limsup_{t \to 0} \frac{W_t}{\sqrt{2t \log\log\frac{1}{t}}} = 1\right) = 1.$$

Let us first prove that

$$\mathbf{P}\left(\limsup_{t \to 0} \frac{W_t}{\sqrt{2t \log\log\frac{1}{t}}} \leqslant 1\right) = 1.$$

Let us denote $h(t) = \sqrt{2t \log\log\frac{1}{t}}$. Let $\alpha, \beta > 0$, from Doob's maximal inequality applied to the martingale $\left(e^{\alpha W_t - \frac{\alpha^2}{2}t}\right)_{t \geq 0}$, we have for $t \geq 0$:

$$\mathbf{P}\left(\sup_{0 \leqslant s \leqslant t} \left(W_s - \frac{\alpha}{2}s\right) > \beta\right) = \mathbf{P}\left(\sup_{0 \leqslant s \leqslant t} e^{\alpha W_s - \frac{\alpha^2}{2}s} > e^{\alpha\beta}\right) \leqslant e^{-\alpha\beta}.$$

Let now $\theta, \delta \in (0, 1)$. Using the previous inequality for every $n \in \mathbb{N}$ with $t = \theta^n, \alpha = \frac{(1+\delta)h(\theta^n)}{\theta^n}, \beta = \frac{1}{2}h(\theta^n)$, yields when $n \to +\infty$,

$$\mathbf{P}\left(\sup_{0 \leqslant s \leqslant \theta^n} \left(W_s - \frac{(1+\delta)h(\theta^n)}{2\theta^n}s\right) > \frac{1}{2}h(\theta^n)\right) = O\left(\frac{1}{n^{1+\delta}}\right).$$

Therefore from Borel-Cantelli lemma, for almost every $\omega \in \Omega$, we may find $N(\omega) \in \mathbb{N}$ such that for $n \geq N(\omega)$,

$$\sup_{0 \leqslant s \leqslant \theta^n} \left(W_s(\omega) - \frac{(1+\delta)h(\theta^n)}{2\theta^n}s\right) \leqslant \frac{1}{2}h(\theta^n).$$

But,

$$\sup_{0 \leqslant s \leqslant \theta^n} \left(W_s(\omega) - \frac{(1+\delta)h(\theta^n)}{2\theta^n}s\right) \leqslant \frac{1}{2}h(\theta^n)$$

implies that for $\theta^{n+1} \leqslant t \leqslant \theta^n$,

$$W_t(\omega) \leqslant \sup_{0 \leqslant s \leqslant \theta^n} W_s(\omega) \leqslant \frac{1}{2}(2+\delta)h(\theta^n) \leqslant \frac{(2+\delta)h(t)}{2\sqrt{\theta}}.$$

We conclude:

$$\mathbf{P}\left(\limsup_{t \to 0} \frac{W_t}{\sqrt{2t \log\log\frac{1}{t}}} \leqslant \frac{2+\delta}{2\sqrt{\theta}}\right) = 1.$$

Letting now $\theta \to 1$ and $\delta \to 0$ yields

$$\mathbf{P}\left(\limsup_{t \to 0} \frac{W_t}{\sqrt{2t \log \log \frac{1}{t}}} \leqslant 1\right) = 1.$$

Let us now prove that

$$\mathbf{P}\left(\limsup_{t \to 0} \frac{W_t}{\sqrt{2t \log \log \frac{1}{t}}} \geq 1\right) = 1.$$

Let $\theta \in (0, 1)$. For $n \in \mathbb{N}$, we denote

$$A_n = \left\{\omega, W_{\theta^n}(\omega) - W_{\theta^{n+1}}(\omega) \geq (1 - \sqrt{\theta})h(\theta^n)\right\}.$$

Let us prove that $\sum \mathbf{P}(A_n) = +\infty$. The basic inequality

$$\int_a^{+\infty} e^{-\frac{u^2}{2}} du \geq \frac{a}{1 + a^2} e^{-\frac{a^2}{2}},$$

implies

$$\mathbf{P}(A_n) = \frac{1}{\sqrt{2\pi}} \int_{a_n}^{+\infty} e^{-\frac{u^2}{2}} du \geq \frac{a_n}{1 + a_n^2} e^{-\frac{a_n^2}{2}},$$

with

$$a_n = \frac{(1 - \sqrt{\theta})h(\theta^n)}{\theta^{n/2}\sqrt{1 - \theta}}.$$

When $n \to +\infty$,

$$\frac{a_n}{1 + a_n^2} e^{-\frac{a_n^2}{2}} = O\left(\frac{1}{n^{\frac{1 + \theta - 2\sqrt{\theta}}{1 - \theta}}}\right),$$

therefore,

$$\sum \mathbf{P}(A_n) = +\infty.$$

As a consequence of the independence of the Brownian increments and of Borel-Cantelli lemma, the event

$$W_{\theta^n} - W_{\theta^{n+1}} \geq (1 - \sqrt{\theta})h(\theta^n)$$

will occur almost surely for infinitely many n's. But, thanks to the first part of the proof, for almost every ω, we may find $N(\omega)$ such that for $n \geq N(\omega)$,

$$W_{\theta^{n+1}} > -2h(\theta^{n+1}) \geq -2\sqrt{\theta}h(\theta^n).$$

Thus, almost surely, the event $W_{\theta^n} > h(\theta^n)(1 - 3\sqrt{\theta})$ will occur for infinitely many n's. This implies

$$\mathbf{P}\left(\limsup_{t \to 0} \frac{W_t}{\sqrt{2t \log \log \frac{1}{t}}} \geq 1 - 3\sqrt{\theta}\right) = 1.$$

We finally get

$$\mathbf{P}\left(\limsup_{t \to 0} \frac{W_t}{\sqrt{2t \log \log \frac{1}{t}}} \geq 1\right) = 1.$$

by letting $\theta \to 0$. $\square$

As a straightforward consequence, we may observe that the time inversion invariance property of Brownian motion implies:

Corollary 3.6. *Let* $(W_t)_{t\geq 0}$ *be a standard Brownian motion.*

$$\mathbf{P}\left(\liminf_{t\to +\infty}\frac{W_t}{\sqrt{2t\log\log t}} = -1, \limsup_{t\to +\infty}\frac{W_t}{\sqrt{2t\log\log t}} = 1\right) = 1.$$

3.2. Support theorem

Let

$$\sigma : \mathbb{R}_+ \times \Omega \to \mathbb{R}^{d\times d},\ b : \mathbb{R}_+ \times \Omega \to \mathbb{R}^d\ \text{ and }\ a_t = \frac{1}{2}\sigma_t\sigma_t^T.$$

Set

$$(3.3) \qquad\qquad x_t = \int_0^t \sigma_s \cdot \mathrm{d}W_s + \int_0^t b_s \mathrm{d}s.$$

For simplicity, we always assume that $a \in \mathbb{S}_\delta^d$.

The following result is a simplify version of Stroock-Varadhan's support theorem, which is taken from [**Bas98**].

THEOREM 3.7 (Support theorem). *Suppose* σ, σ^{-1} *and* b *are bounded,* x_t *is given by* (3.3). *Suppose* $\varphi : [0,1] \to \mathbb{R}^d$ *is continuous with* $\varphi(0) = 0$. *Then for each* $\varepsilon > 0$, *there exists a constant* $c > 0$ *depending only on* ε, *the modulus of continuity of* φ, *and the bounds on* b, σ *and* σ^{-1} *such that*

$$(3.4) \qquad\qquad \mathbf{P}\left(\sup_{t\in[0,1]}|x_t - \varphi(t)| \leqslant \varepsilon\right) \geqslant c.$$

This can be interpreted as saying that the graph of x_t remains within an ε-tube around φ with positive probability.

To prove Theorem 3.2, we need some auxiliary lemmas.

Lemma 3.8. *Suppose* $X_0 = 0$, $X_t = M_t + A_t$ *is a continuous semimartingale with* $\mathrm{d}A_t/\mathrm{d}t$ *and* $\mathrm{d}\langle M\rangle_t/\mathrm{d}t$ *bounded above by* N_1 *and* $\mathrm{d}\langle M\rangle_t/\mathrm{d}t$ *bounded below by* $N_2 > 0$. *If* $\varepsilon > 0$ *and* $T > 0$, *then*

$$\mathbf{P}\left(\sup_{t\in[0,T]}|X_t| < \varepsilon\right) \geqslant c(\varepsilon, T, N_1, N_2) > 0.$$

PROOF. Let $\tau_t = \langle M\rangle_t^{-1} := \inf\{s > 0 : \langle M\rangle_s > t\}$. In virtue of Dambis-Dubins-Schwarz Theorem, $B_t := M_{\tau_t}$ is a Brownian motion. By our assumptions on $\langle M\rangle$, $\tau_t \asymp t$, and $Y_t := X_{\tau_t} = B_t + \int_0^t b_s \mathrm{d}s$ with $|b_s| \leqslant C(N_1, N_2)$. Our assertion will follow if we can show

$$\mathbf{P}\left(\sup_{t\in[0,T]}|Y_t| \leqslant \varepsilon\right) \geqslant c > 0.$$

We now use Girsanov's theorem. Define a probability measure $\mathbf{Q}$ by

$$\mathrm{d}\mathbf{Q}/\mathrm{d}\mathbf{P} = \mathcal{E}_T(-b) := \exp\left(-\int_0^T b_s \mathrm{d}B_s - \frac{1}{2}\int_0^T |b_s|^2 \mathrm{d}s\right)\ \text{ on }\ \mathcal{F}_T.$$

By Girsanov's theorem, under $\mathbf{Q}$, Y_t is a Brownian motion. Therefore,

$$\mathbf{Q}(A) \geqslant c > 0, \quad A = \left\{ \sup_{t \in [0,T]} |Y_t| \leqslant \varepsilon \right\}.$$

By Hölder's inequality,

$$c \leqslant \mathbf{Q}(A) \leqslant E_{\mathbf{P}}(\mathcal{E}_T(-b)\mathbf{1}_A) \leqslant [E_{\mathbf{P}}\mathcal{E}_T^2(-b)]^{\frac{1}{2}}[\mathbf{P}(A)]^{\frac{1}{2}}.$$

Since b is bounded, it is easy to verify that $E_{\mathbf{P}}\mathcal{E}_T^2(-b) < \infty$. This yields $\mathbf{P}(A) \geqslant c > 0$. □

Now we are on the point to give

PROOF OF THEOREM 3.2. *Step 1*: We first consider the case and $\varphi = 0$. Fix $z \in \partial B_{\varepsilon/4}$. Applying Itô's formula with $f(x) = |x - z|^2$ and setting $y_t = |x_t - z|^2$, then

$$y_t = z^2 + \int_0^t (x_s - z) \cdot \mathrm{d}x_s + 2 \int_0^t \mathrm{tr}a_s \mathrm{d}s, \quad \frac{\mathrm{d}}{\mathrm{d}t}\langle y \rangle_t = (x_t - z)^T a_s(x_t - z) \asymp y_t.$$

Set $\tau := \inf\{s > 0 : |y_s - y_0| \geqslant (\varepsilon/8)^2\}$, then $c\varepsilon^2 \leqslant \mathrm{d}\langle y \rangle_t/\mathrm{d}t \leqslant C\varepsilon^2$, $t \in [0, \tau]$. If we set z_t equal to y_t for $t \leqslant \tau$ and equal to some Brownian motion for t larger than this stopping time, then Lemma 3.8 applies (for z_t) and

$$\mathbf{P}\left(\sup_{t \in [0,T]} |x_t| \leqslant \varepsilon \right) \geqslant \mathbf{P}\left(\sup_{t \in [0,T]} |y_t - y_0| \leqslant (\varepsilon/8)^2 \right)$$

$$= \mathbf{P}\left(\sup_{t \in [0,T]} |z_t - z_0| \leqslant (\varepsilon/8)^2 \right) > 0.$$

Step 2: Without loss of generality, we may assume φ is differentiable with a derivative bounded by a constant. Define a new probability measure $\mathbf{Q}$ by

$$\mathrm{d}\mathbf{Q}/\mathrm{d}\mathbf{P} = \exp\left(-\int_0^T \varphi'(s)\sigma_s^{-1}\mathrm{d}W_s - \frac{1}{2}\int_0^T |\varphi'(s)\sigma_s^{-1}|^2 \mathrm{d}s \right) \quad \text{on } \mathcal{F}_T.$$

Noting that

$$\left\langle -\int_0^{\cdot} \varphi'(s)\sigma_s^{-1}\mathrm{d}W_s, x \right\rangle_t = \int_0^t \varphi'(s)\mathrm{d}s = -\varphi(t).$$

So by the Girsanov theorem, under $\mathbf{Q}$ each component of x_t is a semimartingale and $n_t^i := x_t^i - \int_0^t b_s^i \mathrm{d}s - \varphi^i(t)$ is a martingale for each $i = 1, \cdots, d$, and $\langle n^i, n^j \rangle_t = \int_0^t \sigma_k^i(s)\sigma_k^j(s)\mathrm{d}s$. Therefore,

$$B_t := \int_0^t \sigma_s^{-1}\mathrm{d}n_s$$

is a continuous local martingale with $\langle B^i, B^j \rangle_t = \delta_{ij}t$ under $\mathbf{Q}$. Thanks to Lévy's Theorem, B_t is a d-dimensional Brownian motion udner $\mathbf{Q}$. Since

$$x_t - \varphi(t) = \int_0^t \sigma_s \mathrm{d}B_s + \int_0^t b_s \mathrm{d}s,$$

by *Step 1*, $\mathbf{Q}(\sup_{t\in[0,T]}|x_t - \varphi(t)| < \varepsilon) \geqslant c > 0$. similarly to the last paragraph of the proof for Lemma 3.8, we conclude

$$\mathbf{P}\left(\sup_{t\in[0,T]}|x_t - \varphi(t)| < \varepsilon\right) \geqslant c > 0.$$

$\square$

3.3. ABP estimate and Generalized Itô's formula

Below we will use the an analytic result due to Alexsandroff to study the Itô process given by (3.3). Below, we employ an analytic result due to Alexandroff to study the Itô process defined in (3.3). For simplicity, in this section, *we assume that $b = 0$.* However, all results except Proposition 3.9 remain valid if b is uniformly bounded. In that case, the constant C appearing in the estimates below may also depend on the upper bound of $|b|$.

Proposition 3.9 (Alexsandroff)**.** *Let f be a nonnegative function on B_1 such that f^d has finite integral over B_1 and $f = 0$ outside B_1. Then there exists a* **nonpositive convex** *function u on B_2 such that*

(i) for any $x \in B_2$,

$$(3.5) \qquad |u(x)| \leqslant C\left(\int_{B_1} f^d dx\right)^{\frac{1}{d}};$$

(ii) for any symmetric positive definite matrix $a \in \mathbb{R}^{d\times d}$, $0 < \varepsilon < 1$ and $x \in B_1$,

$$(3.6) \qquad a_{ij}\partial_{ij}u_\varepsilon(x) \geqslant d\sqrt[d]{\det a}\, f_\varepsilon(x),$$

where $u_\varepsilon = u * \zeta_\varepsilon$, and ζ_ε is a standard mollifier.

(3.5) is called Alexandroff–Bakelman–Pucci estimate in PDE literature.

In Section 7.4, we provide the proof for Proposition 3.9 based on the very initial knowledge of the solvability of the following Monge–Ampère equations and estimates of its solutions:

$$(3.7) \qquad \det \nabla^2 u(x) = f \quad \text{in } D,$$

which, actually, after a long development became also one of the cornerstones of the theory of fully nonlinear elliptic partial differential equations.

Let

$$\sigma : \mathbb{R}_+ \times \Omega \to \mathbb{R}^{d\times d} \quad \text{and} \quad a_t = \frac{1}{2}\sigma_t\sigma_t^T.$$

Set

$$(3.8) \qquad x_t = \int_0^t \sigma_s \cdot \mathrm{d}W_s$$

and

$$\tau_R(x) = \inf\{t > 0 : x + x_t \notin B_R\}.$$

Proposition 3.9 implies

THEOREM 3.10 (Krylov [**Kry09**]). *There is a constant $C_4 = C_4(d)$ such that for any $R > 0$, and nonnegative Borel f given on $\mathbb{R}^d$, we have*

$$(3.9) \qquad \mathbf{E} \int_0^{\tau_R(x)} f(x + x_t) \sqrt[d]{\det a_t}\, \mathrm{d}t \leqslant C_4 R \|f\|_{L^d(B_R)},$$

PROOF. By scaling, we only need to consider the case $R = 1$. We can also assume $f \in C_c^\infty(B_1)$.

By Itô's formula,

$$u_\varepsilon(x + x_{t \wedge \tau_1(x)}) - u_\varepsilon(x) = \int_0^{t \wedge \tau_1(x)} a_s^{ij} \partial_{ij} u_\varepsilon(x + x_s)\mathrm{d}s + m_{t \wedge \tau_1(x)},$$

where m is a local martingale with $m_0 = 0$. Selecting an appropriate stopping time sequence, taking expectation, letting $t \to \infty$ and using Proposition 3.9, we get

$$\int_0^{\tau_1(x)} \sqrt[d]{\det a_t}\, f_\varepsilon(x + x_t)\mathrm{d}t \leqslant d^{-1} \int_0^{\tau_1(x)} a_t^{ij} \partial_{ij} u_\varepsilon(x + x_t)\mathrm{d}t$$
$$\leqslant \frac{2}{d} \sup_{x \in B_1} |u(x)| \leqslant C_4 \|f\|_{L^d(B_1)}.$$

Letting $\varepsilon \to 0$, we obtain our assertion.   □

We should point out that here we do not need to assume $a \in \mathbb{S}_\delta^d$.

Remark 3.11.   (i) (3.9) implies that if x_t is a Itô's process given by (3.3) with σ non-degenerate, then the process $t \mapsto \int_0^t f(x_s)\mathrm{d}s$ is well-defined.

(ii) Suppose x_t is a Itô process given by (3.3), $a \in \mathbb{S}_\delta^d$ and b satisfying $|b_t| \leqslant \mathfrak{b}(x_t)$ with some $\mathfrak{b} \in L^d$. In this case, Krylov [**Kry21a**] also proved (3.9) with $\|f\|_{L^d(D)}$ replaced by $\|f\|_{L^{d-\varepsilon}(D)}$ for some $\varepsilon = \varepsilon(d, \delta, \|\mathfrak{b}\|) > 0$.

Theorem 3.10 as many results below admits a natural generalization with conditional expectations. This generalization is obtained by tedious and not informative repeating the proof with obvious changes. We mean the following which we call the conditional version of Theorem 3.10 . Let γ be a finite stopping time, then

$$(3.10) \qquad \mathbf{E}\left[\int_\gamma^{\tau_R(x)} f(x + x_t) \sqrt[d]{\det a_t}\, \mathbf{1}_{\{\gamma \leqslant \tau_R(x)\}}\, \mathrm{d}t \Big| \mathcal{F}_\gamma\right] \leqslant C_4 R \|f\|_{L_d(B_R)}.$$

Lemma 3.12. *Assume that $a \in \mathbb{S}_\delta^d$. Then for any $R > 0$ and $x \in B_R$, it holds that*

$$\mathbf{E}\tau_R(x)^n \leqslant n!(C_5 R^2/\delta)^n,$$

where C_5 only depends on d.

PROOF. We can assume $x = 0$ and set $\tau_R = \tau_R(0)$.

We claim that

$$(3.11) \qquad I_n(t) := \mathbf{E}\left([\tau_R - t]_+^n | \mathcal{F}_t\right) \leqslant n!(C_5 R^2/\delta)^n.$$

Of course, (3.11) implies our desired result.

When $n = 1$, (3.10) implies (3.11). If our assertion is true for a given n, then

$$
\begin{aligned}
I_{n+1}(t) =& (n+1)! \mathbf{E}\left(\int \mathbf{1}_{t < t_1 < \cdots < t_{n+1} < \tau_R} \mathrm{d}t_1 \cdots \mathrm{d}t_{n+1} \Big| \mathcal{F}_t\right) \\
=& (n+1)! \int \mathrm{d}t_1 \cdots t_{n+1} \mathbf{E}\left(\mathbf{1}_{t < t_1 < \cdots < t_n < \tau_R} \mathbf{1}_{t_n < t_{n+1} < \tau_R} \Big| \mathcal{F}_t\right) \\
=& (n+1)! \int \mathrm{d}t_1 \cdots t_{n+1} \mathbf{E}\left[\mathbf{1}_{t < t_1 < \cdots < t_n < \tau_R} \mathbf{E}\left(\mathbf{1}_{t_n < t_{n+1} < \tau_R} \Big| \mathcal{F}_{t_n}\right) \Big| \mathcal{F}_t\right] \\
=& (n+1) \mathbf{E}\left[n! \int \mathbf{1}_{t < t_1 < \cdots < t_n < \tau_R} \mathrm{d}t_1 \cdots t_n \int \mathbf{E}\left(\mathbf{1}_{t_n < t_{n+1} < \tau_R} \Big| \mathcal{F}_{t_n}\right) \mathrm{d}t_{n+1} \Big| \mathcal{F}_t\right] \\
=& (n+1) \mathbf{E}\left\{[\tau_R - t]_+^n \mathbf{E}\left[\int_{t_n}^{\tau_R} \mathbf{1}_{B_R}(x_{t_{n+1}}) \mathrm{d}t_{n+1} \Big| \mathcal{F}_{t_n}\right] \Big| \mathcal{F}_t\right\} \\
\overset{(3.10)}{\leqslant}& (n+1) C_5 \delta^{-1} R^2 I_n(t) \overset{(3.11)}{\leqslant} (n+1)! (C_5 R^2/\delta)^{n+1}.
\end{aligned}
$$

So we get what we desired. $\qquad\square$

THEOREM 3.13. *Assume that $a \in \mathbb{S}_\delta^d$. Then for any $\mu < \delta/C_5$, $R \in (0, \infty)$ and $x \in B_R$,*

$$
(3.12) \qquad\qquad \mathbf{E} \exp\left(\frac{\mu \tau_R(x)}{\delta R^2}\right) \leqslant (1 - C_5 \mu/\delta)^{-1}.
$$

In particular, for each $\lambda > 0$,

$$
(3.13) \qquad\qquad \mathbf{P}\left(\tau_R(x) \geqslant \lambda\right) \leqslant 2 \exp\left(-\frac{\lambda}{2 C_5 R^2}\right).
$$

Exercise 3.14. Let B be a one-dimensional BM. Let $I = (-1, 1)$. Prove that

$$
\mathbf{E}\tau_I^n \leqslant C^n n!.
$$

Using this to give another proof for (3.12).

Put

$$
\tau_R := \tau_R(0).
$$

Theorem 3.13 says that τ_R is smaller than a constant times R^2 with high probability. We want to show that in a sense the converse is also true: R^2 is basically smaller than a constant times τ_R with high probability.

Lemma 3.15. *Assume that $a \in \mathbb{S}_\delta^d$. There exists C_6 depending only on d such that*

$$
(3.14) \qquad\qquad \mathbf{P}(\tau_R/R^2 \leqslant t) \leqslant C_6 \delta^{-1} t, \quad t, R > 0.
$$

PROOF. We only need to prove the case $R = 1$. Let ϕ be a C^2 function that is zero at 0, one on ∂B_1, with $\partial_{ij}\phi$ bounded by a constant. By Itô's formula

$$
\mathrm{d}\phi(x_t) = \nabla\phi(x_t) \cdot \sigma_t \mathrm{d}W_t + a_t^{ij} \partial_{ij}\phi(x_t) \mathrm{d}t,
$$

which yields that

$$
\phi(x_{t \wedge \tau_1}) = \mathbf{E}\int_0^{t \wedge \tau_1} a_s^{ij} \partial_{ij}\phi(x_s) \mathrm{d}s \leqslant C_6 \delta^{-1} t.
$$

Since $\phi(x_{t \wedge \tau_1}) \geqslant \mathbf{1}_{\{\tau_1 \leqslant t\}}$, we get $\mathbf{P}(\tau_1 \leqslant t) \leqslant C_6 \delta^{-1} t$. $\qquad\square$

Lemma 3.16. *Assume that $a \in \mathbb{S}_\delta^d$. There is a constant $\mathfrak{R} = R(d, \delta)$ such that*

$$
\mathbf{E} \exp(-\tau_\mathfrak{R}) \leqslant 1/2.
$$

PROOF. **Fact**: Let X be a non-negative random variable, and let $F : \mathbb{R}_+ \to \mathbb{R}$ be a decreasing function with $F(\infty) = 0$. Then

$$\mathbf{E}F(X) = -\int_0^\infty F'(t)\mathbf{P}(X \leqslant t)\mathrm{d}t.$$

Set $X = \tau_R$ and $F(t) = \mathrm{e}^{-t}$. In virtue of (3.14),

$$\mathbf{E}\mathrm{e}^{-\tau_R} = \int_0^\infty \mathrm{e}^{-t}\mathbf{P}(\tau_R \leqslant t)\mathrm{d}t \leqslant \int_0^\infty \mathrm{e}^{-t}[1 \wedge (C_3\delta^{-1}R^{-2}t)]\mathrm{d}t \leqslant C_7\delta^{-1}R^{-2}.$$

We set $\mathfrak{R} = \sqrt{2C_7/\delta}$.   □

Exercise 3.17. For any $R \in (0, \infty)$

(3.15) $\mathbf{E}\exp\left(-\mathfrak{R}^2\tau_R/R^2\right) \leqslant 1/2.$

THEOREM 3.18. *Assume that $a \in \mathbb{S}_\delta^d$. For any $\kappa \in (0, 1), R \in (0, \infty), x \in B_{\kappa R}$, and $\lambda \geqslant 0$,*

(3.16) $\mathbf{E}\exp\left(-\lambda\tau_R(x)\right) \leqslant 2e^{-\sqrt{\lambda}(1-\kappa)R/K},$

where $K = \mathfrak{R}/\log 2$. Consequently,

(3.17) $\mathbf{P}\left(\tau_R(x) \leqslant tR^2\right) \leqslant 2\exp\left(-\frac{\beta(1-\kappa)^2}{t}\right),$

where $\beta = \beta(\mathfrak{R}) = K^{-2}(\mathfrak{R})/4 \in (0, 1)$.

PROOF. Recall that $\tau_R(x)$ is the first exit time of $x + x_t$ from B_R. Let $\tau_R'(x)$ be the first exit time of $x + x_t$ from $B_{(1-\kappa)R}(x)$.

We again assume that $R = 1$, $x = 0$ and $\kappa = 0$. Take $N \in \mathbb{N}$, to be specified later, and introduce τ^k, $k = 1, \cdots, N$, as the first exit time of x_t from $B_{k/N}$. We also set γ^k be the first exit times of x_t from $B_{N^{-1}}(x_{\tau^{k-1}})$ after τ^{k-1}, then

$$\tau^{k-1} \leqslant \gamma^k \leqslant \tau^k$$

and

$$\tau_1 \geqslant (\gamma^1 - \tau_0) + (\gamma^2 - \tau^1) + \cdots + (\gamma^N - \tau^{N-1}).$$

By the conditional version of (3.15),

$$\mathbf{E}\left\{\exp\left[-\mathfrak{R}^2N^2(\gamma^k - \tau^{k-1})\right]|\mathcal{F}_{\tau^{k-1}}\right\} \leqslant 1/2.$$

Therefore,
(3.18)
$$\mathbf{E}\left[\exp\left(-\mathfrak{R}^2N^2\tau_1\right)\right]$$
$$\leqslant \mathbf{E}\left[\prod_{k=1}^N \exp\left(-\mathfrak{R}^2N^2(\gamma^k - \tau^{k-1})\right)\right]$$
$$\leqslant \mathbf{E}\left\{\prod_{k=1}^{N-1} \exp\left(-\mathfrak{R}^2N^2(\gamma^k - \tau^{k-1})\right)\mathbf{E}\left[\exp\left(-\mathfrak{R}^2N^2(\gamma^N - \tau^{N-1})\right)\Big|\mathcal{F}_{\tau^{N-1}}\right]\right\}$$
$$\leqslant \frac{1}{2}\mathbf{E}\left[\prod_{k=1}^{N-1} \exp\left(-\mathfrak{R}^2N^2(\gamma^k - \tau^{k-1})\right)\right] \leqslant \cdots \leqslant (1/2)^N.$$

Choosing $N = [\sqrt{\lambda}/\mathfrak{R}]$, we get (3.16).

For (3.17). Thanks to (3.16),

$$\mathbf{P}\left(\tau_R(x) \leqslant tR^2\right) = \mathbf{P}\left(e^{-\lambda \tau_R(x)} \geqslant e^{-\lambda tR^2}\right) \leqslant 2e^{\lambda tR^2 - \sqrt{\lambda}(1-\kappa)R/K},$$

Choosing $\lambda = (\frac{1-\kappa}{2tRK})^2$, we obtain the desired estimate. $\qquad\square$

The above estimates for first exit times have many important applications. Let

$$\sigma_\Gamma(x) = \inf\{t > 0 : x + x_t \in \Gamma\}$$

be the first time the process $x + x_t$ hits Γ.

Proposition 3.19. *Assume that $a \in \mathbb{S}_\delta^d$. For any $\kappa \in (0,1)$ there is a function $q(\gamma), \gamma \in (0,1)$, depending only on d, δ, κ and naturally, also on γ, such that for any $R \in (0,\infty), x \in B_{\kappa R}$, and closed $\Gamma \subset B_R$ satisfying $|\Gamma| \geqslant \gamma |B_R|$, it holds that*

$$\mathbf{P}\left(\sigma_\Gamma(x) \leqslant \tau_R(x)\right) \geqslant q(\gamma).$$

Furthermore, $q(\gamma) \to 1$ as $\gamma \uparrow 1$.

PROOF. By using scaling as before we reduce the general case to the one in which $R = 1$. For any $\varepsilon > 0$, we have

$$\mathbf{P}\left(\sigma_\Gamma(x) > \tau_1(x)\right) \leqslant \mathbf{P}\left(\tau_1(x) = \int_0^{\tau_1(x)} \mathbf{1}_{B_1 \setminus \Gamma}(x + x_t)\, dt\right)$$

$$\leqslant \mathbf{P}\left(\tau_1(x) \leqslant \varepsilon\right) + \varepsilon^{-1}\mathbf{E}\int_0^{\tau_1(x)} I_{B_1 \setminus \Gamma}(x + x_t)\, dt.$$

In virtue of (3.17) and (3.9), for any $x \in B_\kappa$ and any $\varepsilon > 0$, it holds that

$$\mathbf{P}\left(\sigma_\Gamma(x) > \tau_1(x)\right) \leqslant 2e^{-\frac{\beta(1-\kappa)^2}{\varepsilon}} + C\varepsilon^{-1}|B_1 \setminus \Gamma|^{1/d}$$

$$\leqslant 2e^{-\frac{1}{C\varepsilon}} + C\varepsilon^{-1}(1-\gamma)^{1/d},$$

where the constants C depend only on d, δ, κ. By denoting

$$q(\gamma) = 1 - \inf_{\varepsilon > 0}\left(2e^{-\frac{1}{C\varepsilon}} + C\varepsilon^{-1}(1-\gamma)^{1/d}\right),$$

we obtain our desired assertion. $\qquad\square$

Note that in the above result, we have no assumption on the shape of the set Γ.

Exercise 3.20. For any $\kappa \in (0,1)$, $R \in (0,\infty)$. For any $x \in B_{R/2}$ and $B_{\kappa R}(y) \subseteq B_R$, we have

$$\mathbf{P}\left(\sigma_{B_{\kappa R}(y)}(x) < \tau_R(x)\right) \geqslant \zeta(\kappa) > 0,$$

where $\zeta(\kappa) > 0$ depends only on d, δ, and naturally, also on κ.

Hint: Using support theorem.

THEOREM 3.21. *Let $p \geqslant d$. Then there exists constants C depending only on d, δ, such that for any $\lambda > 0$ and Borel nonnegative f given on $\mathbb{R}^d$ we have*

$$(3.19) \qquad \mathbf{E}\int_0^\infty e^{-\lambda t}f(x_t)\, dt \leqslant C\lambda^{\frac{d}{2p}-1}\|f\|_p.$$

PROOF. Let γ be a stopping time and γ' be the first exit time of x_t from $B_R(x_\gamma)$ after γ. By the conditional version of (3.16),

$$\mathbf{E}\left[\exp\left(-\lambda(\gamma'-\gamma)\right)\Big|\mathcal{F}_\gamma\right] \leqslant 2e^{-\sqrt{\lambda}R/K}.$$

Choosing $R = K/\sqrt{\lambda}$, then

$$\mathbf{E}\left[\exp\left(-\lambda(\gamma'-\gamma)\right)\Big|\mathcal{F}_\gamma\right] \leqslant 2/e < 1.$$

Let $\tau^0 = 0$ and τ^k be the first exit time of x_t from $B_R(x_{\tau^{k-1}})$ after τ^{k-1}. As the proof for (3.18), we have

$$(3.20) \qquad \mathbf{E}e^{-\lambda\tau^k} = \mathbf{E}\prod_{i=1}^{k}e^{-\lambda(\tau^k-\tau^{k-1})} \leqslant (2/e)^k.$$

If (3.20) holds, then

$$\mathbf{E}\int_0^\infty e^{-\lambda t}f(x_t)\,\mathrm{d}t \leqslant \sum_{k=1}^{\infty}\mathbf{E}\left[e^{-\lambda\tau^{k-1}}\mathbf{E}\left(\int_{\tau^{k-1}}^{\tau^k}f(x_t)\,\mathrm{d}t\Big|\mathcal{F}_{\tau^{k-1}}\right)\right]$$

$$\overset{(3.9)}{\leqslant} \sum_{k=1}^{\infty}\mathbf{E}\left(C\delta^{-1}R\|f\|_{L^d(B_R(x_{\tau^{k-1}}))}e^{-\lambda\tau^{k-1}}\right)$$

$$\leqslant C\delta^{-1}R^{2-\frac{d}{p}}\|f\|_p\sum_{k=0}^{\infty}\mathbf{E}e^{-\lambda\tau^k}$$

$$\leqslant C\delta^{-1}(K/\sqrt{\lambda})^{2-\frac{d}{p}}\|f\|_p\sum_{k=0}^{\infty}(2/e)^k$$

$$\leqslant C\lambda^{\frac{d}{2p}-1}\|f\|_p.$$

$\square$

THEOREM 3.22 (Generalized Itô's formula, see Krylov-[**Kry09**]). *Let x_t be a Itô process given by* (3.8). *Suppose that $a \in \mathbb{S}_\delta^d$, then for any $u \in W_{loc}^{2,p}$ with $p \geqslant d$, we have*

$$(3.21) \qquad \begin{aligned} &u(x+x_t) - u(x) \\ &= \int_0^t \nabla u(x+x_s)\sigma_s\mathrm{d}W_s + \int_0^t a_s^{ij}\partial_{ij}u(x+x_s)\mathrm{d}s \end{aligned}$$

PROOF. We only consider the case $x = 0$ and $u \in W^{2,d}$. Let $\eta \in C_c^\infty(B_1)$ with $\int\eta = 1$. Set $\eta_\varepsilon(x) = \varepsilon^{-d}\eta(x/\varepsilon)$ and $u_\varepsilon = u * \eta_\varepsilon$. By Itô's formula,

$$(3.22) \qquad u_\varepsilon(x_t) - u_\varepsilon(x_0) = \int_0^t \nabla u_\varepsilon(x_s)\sigma_s\mathrm{d}W_s + \int_0^t a_s^{ij}\partial_{ij}u_\varepsilon(x_s))\mathrm{d}s.$$

Fact: by Sobolev embedding theorem, we have

$$(3.23) \qquad W^{2,d} \hookrightarrow C_b; \quad \|\nabla u\|_{2d} \leqslant C(\|\nabla^2 u\|_{L^d} + \|\nabla u\|_{L^d}).$$

Since $u \in C_b$, by letting $\varepsilon \to 0$, one sees the left-hand side of (3.22) goes to $u(x_t) - u(x_0)$ as $\varepsilon \to 0$. For the right-hand side of (3.22). By Doob's maximal inequality

$$\mathbf{E} \sup_{t \in [0,T]} \left| \int_0^t \nabla u_\varepsilon(x_s) \sigma_s \mathrm{d}W_s - \int_0^t \nabla u_{\varepsilon'}(x_s) \sigma_s \mathrm{d}W_s \right|^2$$

$$\leqslant C\mathbf{E} \int_0^T |\nabla u_\varepsilon - \nabla u_{\varepsilon'}|^2(x_s) \mathrm{d}s \leqslant C \|\nabla u_\varepsilon - \nabla u_{\varepsilon'}\|_{L^{2d}}^2$$

$$\overset{(3.23)}{\leqslant} C \|u_\varepsilon - u_{\varepsilon'}\|_{W^{2,d}} \to 0, \quad \varepsilon, \varepsilon' \to 0.$$

Similarly, we can also show that the second integral on the right-hand side of (3.22) also converges to $\int_0^t a_s^{ij} \partial_{ij} u(x_s) \mathrm{d}s$. $\qquad\square$

Remark 3.23. The above generalized Itô's formula also holds for Itô process given by (3.3), where $a \in \mathbb{S}_\delta^d$, and b satisfying $|b_t| \leqslant \mathfrak{b}(x_t)$ with $\mathfrak{b} \in L^d$.

Exercise 3.24. Let $d \geqslant 1$, $p > 1 \vee \frac{d}{2}$ and W be a d-dimensional Brownian motion. Prove that

(1) if $f \in L^p$, then $\int_0^t f(W_s) \mathrm{d}s$ is well-defined;
(2) if $u \in W^{2,p}$, then it holds that

$$u(W_t) - u(0) = \int_0^t \nabla u(W_s) \mathrm{d}W_s + \int_0^t \frac{1}{2} \Delta u(W_s) \mathrm{d}s.$$

Weak well-posedness

In this section, we study the weak well-posedness of (2.5) (with $b = 0$). The core is to study the regularity of following resolvent equation:

$$\lambda u - Lu = f, \tag{4.1}$$

where $Lu = a_{ij}\partial_{ij}u$, $a \in \mathbb{S}^d_\delta$ and uniformly continuous.

4.1. Uniqueness in law

THEOREM 4.1 (Stroock-Varadhan). *Under the assumptions that σ is continuous and bounded, and $\sigma(x)\sigma^t(x) > 0$ for each $x \in \mathbb{R}^d$. Then SDE (2.5) (with $b = 0$) has a weak solution, and the distribution of such solution is unique.*

Our strategy is

(a) Using generalized Itô's formual and L^p-estimate for the resolvent equation to show the uniqueness of law(X_t).
(b) Proving the finite-dimensional distribution of $(X_t)_{t \geqslant 0}$ is unique by induction.

Lemma 4.2. *Let $L = \Delta$ and $p \in (1, \infty)$. For any $f \in L^p$, there exists a unique solution $u \in W^{2,p}$ solving (4.1). Moreover, u satisfies*

$$\lambda \|u\|_p + \|\nabla u^2\|_p \leqslant C\|f\|_p, \tag{4.2}$$

where C only depends on d and p.

THEOREM 4.3. *Let $p \in (1, \infty)$. There exists a constant $\lambda_0 = \lambda_0(d, p, \omega_a) > 0$ such that for any $\lambda \geqslant \lambda_0$ and $f \in L^p$, equation (4.1) admits a unique solution $u \in W^{2,p}$.*

PROOF. Assume that $u \in W^{2,p}$. We want to show that for sufficiently large λ, it holds that

$$\lambda \|u\|_p + \|u\|_{W^{2,p}} \leqslant C\|\lambda u - Lu\|_p. \tag{4.3}$$

Suppose we have (4.3). Let $T_0 = \lambda - \Delta$ and $T_1 = \lambda - L$, and $B = W^{2,p}$ and $V = L^p$. Utilizing Lemma 2.35 and Lemma 4.2, we can see that (4.1) has a solution in $W^{2,p}$.

Now let us prove (4.3) for $L = a^{ij}\partial_{ij}$. Let $f := \lambda u - Lu$. Assume $\zeta \in C_c^\infty(B_2)$ such that $\zeta \geqslant 0$, $\zeta \equiv 1$ in B_1. Set $\zeta_\varepsilon^z = \zeta((x - z)/\varepsilon)$. Then

$$\lambda(u\zeta_\varepsilon^z) - a_{ij}(z)\partial_{ij}(u\zeta_\varepsilon^z) = f\zeta_\varepsilon^z - 2a_{ij}\partial_i u\partial_j\zeta_\varepsilon^z - a_{ij}\partial_{ij}\zeta_\varepsilon^z u + (a_{ij} - a_{ij}(z))\partial_{ij}(u\zeta_\varepsilon^z).$$

By Lemma 4.2, we get

$$\begin{aligned}
\lambda\|u\zeta_\varepsilon^z\|_p + \|\nabla^2(u\zeta_\varepsilon^z)\|_p \leqslant{}& C\omega_a(2\varepsilon)\|\nabla^2(u\zeta_\varepsilon^z)\|_p + C\|f\|_{L^p(B_{2\varepsilon}(z))} \\
&+ C\varepsilon^{-1}\|\nabla u\|_{L^p(B_{2\varepsilon}(z))} + C\varepsilon^{-2}\|u\|_{L^p(B_{2\varepsilon}(z))}.
\end{aligned}$$

Choosing $\varepsilon_0 > 0$ sufficiently small such that $C\omega_a(2\varepsilon_0) \leqslant 1/2$, then

$$
\begin{aligned}
(4.4) \quad & \lambda\|u\|_{L^p(B_{\varepsilon_0}(z))} + \|\nabla^2 u\|_{L^p(B_{\varepsilon_0}(z))} \\
& \leqslant C\|f\|_{L^p(B_{2\varepsilon_0}(z))} + C\varepsilon_0^{-1}\|\nabla u\|_{L^p(B_{2\varepsilon_0}(z))} + C\varepsilon_0^{-2}\|u\|_{L^p(B_{2\varepsilon_0}(z))}.
\end{aligned}
$$

Fact: There exist constants $c = c(d, p, \varepsilon) > 0$ and $C = C(d, p, \varepsilon) > 0$, and a sequence $\{z_i\}_{i \in \mathbb{N}} \subseteq \mathbb{R}^d$ such that

$$
(4.5) \qquad c\sum_i \int |h\zeta_\varepsilon^{z_i}|^p \leqslant \int |h|^p \leqslant C\sum_i \int |h\zeta_\varepsilon^{z_i}|^p.
$$

By (4.4) and (4.5), we obtain

$$
\lambda\|u\|_p^p + \|\nabla^2 u\|_p^p \leqslant C\|f\|_p^p + C\|\nabla u\|_p^p + C\|u\|_p^p,
$$

where C only depends on d, p and ω_a. Using intepolation theorem, one can see that

$$
\lambda\|u\|_p + \|\nabla u\|_p + \|\nabla^2 u\|_p^p \leqslant \frac{1}{2}\|\nabla^2 u\|_p + \frac{\lambda_0}{2}\|u\|_p + C\|f\|_p,
$$

where $\lambda_0 \geqslant 1$ is a constant only depends on d, p and ω_a. Therefore, for any $\lambda \geqslant \lambda_0 \geqslant 1$, we have

$$
\lambda\|u\|_p + \|u\|_{W^{2,p}} \leqslant C\|f\|_p.
$$

$\square$

Now let $f \in C_c^\infty(\mathbb{R}^d)$. Assume that $u \in W^{2,d}$ is a solution to (4.1) for some $\lambda \geqslant \lambda_0$. Applying Generalized Itô's formula, one can see that

$$
\mathrm{d}\left(\mathrm{e}^{-\lambda t}u(X_{s+t})\right) = \mathrm{e}^{-\lambda t}\left[-\lambda u(X_{s+t}) + Lu(X_{s+t})\right] + \mathrm{e}^{-\lambda t}\nabla u(X_{s+t})\sigma(X_{s+t})\mathrm{d}W_{s+t}.
$$

Taking expection conditional on $\mathcal{F}_s$, we get

$$
u(X_s) = \mathbf{E}(u(X_s)|\mathcal{F}_s) = \int_0^\infty \mathrm{e}^{-\lambda t}\mathbf{E}\left(f(X_{s+t})\big|\mathcal{F}_s\right)\mathrm{d}t, \quad \forall \lambda \gg 1.
$$

This implies that $\mathbf{P}(X_{s+t} \in \cdot|\mathcal{F}_s)$ is unique and $\mathbf{P}(X_{s+t} \in \cdot|\mathcal{F}_s) = \mathbf{P}(X_{s+t} \in \cdot|X_s)$. Using this fact, then the uniqueness in law of X_t can be obtained by induction.

4.2. Strong Markov property

Define $\mathcal{W}$ to be the set of all continuous functions from $\mathbb{R}_+$ to $\mathbb{R}^d$. Suppose that for each starting point x the SDE (2.5) has a solution that is unique in law. Let us denote the solution by $X(x, t, \omega)$. For each x define a probability measure $\mathbb{P}_x$ on $\mathcal{W}$ so that

$$
\begin{aligned}
& \mathbf{P}\left(X(x, t_1) \in A_1, \cdots, X(x, t_n) \in A_n\right) \\
& = \mathbb{P}_x(\omega(t_1) \in A_1, \cdots, \omega(t_n) \in A_n).
\end{aligned}
$$

Let $\mathcal{G}_t^0$ be the σ-algebra generated by $\{\omega_s : s \leqslant t\}$. We complete these σ-fields by considering all sets that are in the $\mathbb{P}_x$ completion of $\mathcal{G}_t^0$ for all x. Finally, we obtain a right continuous filtration by letting $\mathcal{G}_t := \cap_{\varepsilon > 0}\mathcal{G}_{t+\varepsilon}^0$. We then extend $\mathbb{P}_x$ to $\mathcal{G}_\infty$.

For each $t \geqslant 0$, the shift operator $\theta_t : \mathcal{W} \to \mathcal{W}$ is given by $\theta_t\omega(s) = \omega(t+s), s \geqslant 0$. For any finite stopping time τ, we also set $\theta_\tau(\omega)(s) = \omega(\tau(\omega) + s), s \geqslant 0$.

The **strong Markov property** of $(\mathbb{P}_x)_{x \in \mathbb{R}^d}$ is the assertion that

$$\mathbb{E}_x(Y \circ \theta_\tau | \mathcal{G}_\tau) = \mathbb{E}_{X_\tau}(Y),$$

whenever $x \in \mathbb{R}^d$, $Y \in \mathcal{G}_\infty$ is bounded, τ a finite stopping time and $X_t(\omega) = \omega(t)$ is the canonical process.

From now on, by a slight abuse of notation, we will say $(\mathbb{P}_x, X_t)$ is a strong Markov process if $(\mathbb{P}_x)_{x \in \mathbb{R}^d}$ satisfies the strong Markov property.

To prove the strong Markov property it suffices to show

$$(4.6) \qquad \mathbb{E}_x(f(X_{\tau+t}) | \mathcal{G}_\tau) = \mathbb{E}_{X_\tau} f(X_t),$$

for all $x \in \mathbb{R}^d$, $f \in C_c(\mathbb{R}^d)$ and τ a bounded stopping time.

THEOREM 4.4. *Suppose the solution to* (2.5) *is weakly unique for each* $x \in \mathbb{R}^d$. *Then* $(\mathbb{P}_x, X_t)$ *is a strong Markov process.*

PROOF. Let $(\mathcal{W}, \mathcal{G}, (\mathcal{G}_t)_{t \geqslant 0}, \mathbb{P}_x; X, W)$ be a weak solution to (2.5), where X_t is the canonical process and W_t is a $\mathcal{G}_t$-Brownian motion. By definition

$$X_t = X_0 + \int_0^t \sigma(X_s) \, dW_s + \int_0^t b(X_s) \, ds.$$

Set $X_t' = X_{\tau+t}$ and $W_t' = W_{\tau+t} - W_\tau$. By definition, we have

$$(4.7) \qquad X_t' = X_0' + \int_0^t \sigma(X_s') \, dW_s' + \int_0^t b(X_s') \, ds.$$

Let $Q_\tau : \mathcal{W} \times \mathscr{B}(\mathcal{W}) \to [0,1]$ be the regular conditional probability given $\mathcal{G}_\tau$. Then for each bounded function $Y \in \mathcal{G}$, it holds that

$$(4.8) \qquad \mathbb{E}_x(Y | \mathcal{G}_\tau)(\omega) = \int_{\mathcal{W}} Y(\omega') Q_\tau(\omega, d\omega') =: E_{Q_\tau(\omega, \cdot)}(Y), \quad \mathbb{P}_x\text{-a.s.}$$

> **Claim**: For $\mathbb{P}_x$-a.s. $\omega \in \mathcal{W}$, it holds that
> (a) $X_0'(\omega') = X_{\tau(\omega)}(\omega)$, for $Q_\tau(\omega, \cdot)$-a.s. $\omega' \in \mathcal{W}$.
> (b) W' is a Brownian motion with respect to the measure $Q_\tau(\omega, \cdot)$.

By our claim above, X_t' is a weak solution to (2.5) with initial data $X_0' = X_{\tau(\omega)}(\omega)$ under $Q_\tau(\omega, \cdot)$. The uniqueness in law implies that

$$E_{Q_\tau} f(X_t') = \mathbb{E}_{X_\tau} f(X_t), \quad \mathbb{P}_x\text{-a.s.}$$

On the other hand, by definition

$$E_{Q_\tau} f(X_t') = E_{Q_\tau} f(X_{\tau+t}) = \mathbb{E}_x(f(X_{\tau+t}) | \mathcal{G}_\tau).$$

Thus, we get (4.6).

Our task now is to prove the claim.

For (a). Let $B = \{\omega' \in \mathcal{W} : \omega'(\tau(\omega')) \in A \subseteq \mathcal{B}(\mathbb{R}^d)\} \in \mathcal{G}_\tau$. Then by the definition of Q_τ, for $\mathbb{P}_x$-a.s. $\omega \in \mathcal{W}$, we have

$$Q_\tau(\omega, B) = \mathbb{P}_x(B | \mathcal{G}_\tau)(\omega) = \mathbf{1}_B(\omega) = \begin{cases} 1, & \text{if } \omega(\tau(\omega)) \in A \\ 0, & \text{if } \omega(\tau(\omega)) \notin A, \end{cases}$$

which yields claim (a) above.

For (b), we need the following

> **Fact**: If W_t is a $\mathcal{G}_t$-Brownian motion and τ is a finite stopping time, then $W_{\tau+t} - W_\tau$ is a $\mathcal{G}_{\tau+t}$-Brownian motion. (This can be proved by Lévy's Characterization Theorem)

Using the above fact, we have

$$E_{Q_\tau} \exp\left(i \sum_{k=1}^{n-1} \lambda_k \cdot \left(W_{\tau+t_{k+1}} - W_{\tau+t_k} \right) \right)$$

$$= \mathbb{E}_x \left[\exp\left(i \sum_{k=1}^{n-1} \lambda_k \cdot \left(W_{\tau+t_{k+1}} - W_{\tau+t_k} \right) \right) \Big| \mathcal{G}_\tau \right]$$

$$= \exp\left(\sum_{k=1}^{n-1} |\lambda_k|^2 \left(t_{k+1} - t_k \right) / 2 \right),$$

we get what we claimed. □

CHAPTER 5

Applications to Elliptic equations

Let X_t be the solution to (2.5) with $X_0 = x$. We will write $(\mathbb{P}_x, X_t)$ for the strong Markov process corresponding to σ and b (This can be ensured by assuming $\sigma, b \in C_b^1$, or $a \in \mathbb{S}_\delta$, a is continuous and b is bounded).

Put $L = a_{ij}\partial_{ij} + b_i\partial_i$. We always assume a and b are bounded.

5.1. Poisson equations

Consider the following Poisson equation:

$$(5.1) \qquad \lambda u - Lu = f, \quad \lambda > 0, f \in C_b.$$

The relationship between these two subjects can be easily established by Itô's formula:

THEOREM 5.1. *Suppose u is a C_b^2 function satisfying the above Poisson equation. Then*

$$u(x) = \mathbb{E}_x \int_0^\infty e^{-\lambda t} f(X_t)\,\mathrm{d}t$$

PROOF. Applying Itô's formula, we have $\mathrm{d}u(X_t) = \mathrm{d}M_t + Lu(X_t)\mathrm{d}t$, where M is a L^2-martingale. So

$$e^{-\lambda t} u(X_t) - u(x) = \int_0^t e^{-\lambda s}\mathrm{d}M_s + \int_0^t e^{-\lambda s} Lu(X_s)\,\mathrm{d}s$$
$$- \lambda \int_0^t e^{-\lambda s} u(X_s)\,\mathrm{d}s.$$

Taking expectation, we get what we claimed. $\square$

Let us now let D be a nice bounded domain, e.g., a ball. Poisson's equation in D requires one to find a function u such that

$$\begin{cases} \lambda u - Lu = f & \text{in } D \\ u = 0 & \text{on } \partial D, \end{cases}$$

where $\lambda \geqslant 0$. Here we can allow λ to be equal to 0. Recall that if $L = \Delta$ (X_t is a Brownian motion), then the time to exit D, namely, $\tau_D := \inf\{t > 0 : X_t \notin D\}$, is finite almost surely.

THEOREM 5.2. *Suppose u is a solution to Poisson's equation in a bounded domain D that is C^2 in D and continuous on $\bar{D}$. Assume also that*

$$\mathbb{P}_x(\tau_D < \infty) = 1, \quad x \in D$$

where $\tau_D = \inf\{t > 0 : x \notin D\}$. Then

$$u(x) = \mathbb{E}_x \int_0^{\tau_D} e^{-\lambda s} f(X_s)\,\mathrm{d}s.$$

49

Exercise 5.3. Prove Theorem 5.2.

5.2. Dirichlet Problems and Harmonic functions

Let D be a ball (or other nice bounded domain) and let us consider the solution to the Dirichlet problem: given g a continuous function on ∂D, find $u \in C(\bar{D})$ such that u is C^2 in D and

$$(5.2) \qquad \begin{cases} Lu = 0 \text{ in } D \\ u = g \text{ on } \partial D. \end{cases}$$

If $Lu = 0$ in D, we say u is L-harmonic in D.

THEOREM 5.4. *Assume that* $\mathbb{P}_x(\tau_D < \infty) = 1$ *for each* $x \in D$. *Suppose that* $u \in C^2(D) \cap C(\bar{D})$ *satisfies* (5.2), *then*

$$u(x) = \mathbb{E}_x g\left(X_{\tau_D}\right).$$

PROOF. Let $\tau_n = \inf\{t : \text{dist}\,(X_t, \partial D) < 1/n\}$. By Itô's formula,

$$u\left(X_{t \wedge \tau_n}\right) = u\left(X_0\right) + M_{t \wedge \tau_n} + \int_0^{t \wedge \tau_n} Lu\left(X_s\right) ds.$$

Since $Lu = 0$ inside D, taking expectations shows

$$u(x) = \mathbb{E}_x u\left(X_{t \wedge \tau_n}\right).$$

We let $t \to \infty$ and then $n \to \infty$. By dominated convergence, we obtain $u(x) = \mathbb{E}_x u\left(X_{\tau_D}\right)$. This is what we want since $u = g$ on ∂D. □

Exercise 5.5. Theorem 5.4 implies the weak maximum principle: $\max_D u \leqslant \max_{\partial D} u$.

Exercise 5.6. Theorem 3.2 implies the strong maximum principle: if u is not a constant function, then for each $x \in D$, $u(x) < \max_{\partial D} u$

THEOREM 5.7. *Assume that* $(\mathbb{P}_x, X_t)$ *is a strong Markov process and that* $\mathbb{P}_x\left(\tau_D < \infty\right) = 1$, $x \in D$. *Suppose that* $a, b \in C(D)$, $g \in C(\partial D)$, *and* $u(x) := \mathbb{E}_x g(X_{\tau_D}) \in C^2(D)$. *Then* $Lu = 0$ *in* D.

PROOF. Let $B_r(x) \subseteq D$. By the strong Markov property, we have

$$u(x) = \mathbb{E}_x g(X_{\tau_D}) = \mathbb{E}_x g(X_{\tau_D} \circ \theta_{\tau_{B_r(x)}}) = \mathbb{E}_x \left\{ \mathbb{E}_x \left[g(X_{\tau_D} \circ \theta_{\tau_{B_r(x)}}) \middle| \mathcal{F}_{\tau_{B_r(x)}} \right] \right\}$$
$$= \mathbb{E}_x \left[\mathbb{E}_{X_{\tau_{B_r(x)}}} g(X_{\tau_D}) \right] = \mathbb{E}_x u(X_{\tau_{B_r(x)}}).$$

Noting that $u \in C^2(D)$, by Itô's formula,

$$u(X_{t \wedge \tau_{B_r(x)}}) - u(x) = \int_0^{t \wedge \tau_{B_r(x)}} Lu(X_s) ds + M_{t \wedge \tau_{B_r(x)}},$$

where M is a martingale. Taking expectations and letting $t \to \infty$,

$$0 = \frac{1}{\mathbb{E}_x \tau_{B_r(x)}} \mathbb{E}_x \int_0^{\tau_{B_r(x)}} Lu(X_s) ds.$$

By the continuity of Lu and letting $r \to 0$, we get $Lu(x) = 0$. □

One can also consider the following Schrödinger type operator:

$$L_q u = Lu + cu.$$

Equation involving the above operator are considerably simpler than the quantum mechanics Schrödinger equation because here all terms are real-valued.

THEOREM 5.8. *Let D be a nice bounded domain, and $q \in C(\bar{D})$ and $g \in C(\partial D)$. Let $u \in C^2(D) \cap C(\bar{D})$ that agrees with g on ∂D and satisfies $L_q u = 0$ in D. If*

$$\mathbb{E}_x \exp\left(\int_0^{\tau_D} q^+(X_s)\mathrm{d}s\right) < \infty,$$

then

$$u(x) = \mathbb{E}_x\left[g(X_{\tau_D})\exp\left(\int_0^{\tau_D} q(X_s)\mathrm{d}s\right)\right].$$

Exercise 5.9. Prove Theorem 5.8.

Exercise 5.10. Using (3.9) to show: there exists $\varepsilon > 0$ such that if $B \subseteq Q_1, x \in Q_{1/2}$, and $|Q_1 - B| < \varepsilon$, then

$$\mathbb{E}_x \int_0^{\tau_{Q_1}} 1_B(X_s)\,ds \geqslant c > 0,$$

where c is a constant only depends on d, δ and ε.

5.3. Once again on the hitting probability

Recall that

$$x_t = \int_0^t \sigma_s \mathrm{d}W_s, \quad a = \frac{1}{2}\sigma\sigma^t \in \mathbb{S}_\delta^d.$$

In this section, we want to prove following important hitting probability estimate, which is a refined version of Proposition 3.19. This was first found by Krylov-Safonov [**KS79**].

Recall that

$$\sigma_\Gamma(x) = \inf\{t > 0 : x + x_t \in \Gamma\} \quad \text{and} \quad \tau_Q = \inf\{t > 0 : x + x_t \notin Q\}.$$

THEOREM 5.11. *There is a increasing function $p : (0,1) \to (0,1)$, which only depends on d and δ, such that for any $\Gamma \subset Q_1$ and $x \in Q_{1/2}$,*

(5.3) $$\mathbf{P}(\sigma_\Gamma(x) < \tau_{Q_1}(x)) \geqslant p(|\Gamma|).$$

Before prove Theorem 5.11, we need some preparation.

One tool is a corollary of the Calderón-Zygmund cube decomposition. Let Q_1 be the unit cube. We split it into 2^n cubes of half side. We do the same splitting with each one of these 2^n cubes and we iterate this process. The cubes obtained in this way are called dyadic cubes.

If Q is a dyadic cube different from Q_1, we say that $\widetilde{Q}$ is the predecessor of Q if Q is one of the 2^n cubes obtained from dividing $\widetilde{Q}$.

We also let $Q(\kappa)$ denote the cube with the same center as Q but side length κ times as long.

Lemma 5.12 (Krylov-Safonov [**KS79**]). *Let $\gamma \in (0,1)$. If $\Gamma \subset Q_1$ and $|\Gamma| \leqslant \gamma$, then there exists a sequence of dyadic cubes, say $\{Q^i\}_{i \in \mathcal{I}}$ such that*

(1) the interiors of the Q^i are pairwise disjoint;

(2) $|\Gamma \cap Q^i| > \gamma |Q^i|$ and $|\Gamma \cap \widetilde{Q}^i| \leqslant \gamma |\widetilde{Q}^i|$, for each $i \in \mathcal{I}$;

(3) $|\Gamma| \leqslant \gamma |E|$ and $|\Gamma \backslash E| = 0$, where $E = \cup_{i \in \mathcal{I}} \widetilde{Q}^i$.

PROOF. We use the Calderón-Zygmund decomposition. We have that

$$\frac{|Q_1 \cap \Gamma|}{|Q_1|} = |\Gamma| \leqslant \gamma.$$

We subdivide Q_1 into 2^n dyadic cubes. If Q is one of these 2^n subcubes of Q_1 and satisfies $|Q \cap \Gamma|/|Q| \leqslant \gamma$, we then split Q into 2^n dyadic cubes. We iterate this process. In this way, we pick a family $Q^1, Q^2 \cdots$ of dyadic cubes (different from Q_1) such that

$$\frac{|Q^i \cap \Gamma|}{|Q^i|} > \gamma, \quad \forall i \in \mathcal{I}$$

If $x \notin \cup_{i \in \mathcal{I}} Q^i$ then x belongs to an infinite number of closed dyadic cubes Q with diameters tending to zero, such that $|Q \cap \Gamma|/|Q| \leqslant \gamma < 1$. Applying the Lebesgue differentiation theorem to $\mathbf{1}_\Gamma$, we get that $\mathbf{1}_\Gamma(x) \leqslant \gamma < 1$ for a.e. $x \notin \cup_{i \in \mathcal{I}} Q^i$. Hence $\Gamma \subseteq \cup_{i \in \mathcal{I}} Q^i$, except for a set of measure zero.

Consider the family of predecessors of the cubes $\{Q^i\}$, and relabel them as $\{\widetilde{Q}^i\}_{i \in \widetilde{\mathcal{I}}}$ to ensure pairwise disjointness. We immediately observe that:

$$\Gamma \subseteq \cup_{i \in \mathcal{I}} Q^i \subseteq \cup_{i \in \widetilde{\mathcal{I}}} \widetilde{Q}^i =: E,$$

except for a set of measure zero. From the way we chose the cubes Q^i,

$$\frac{\left|\widetilde{Q}^i \cap \Gamma\right|}{\left|\widetilde{Q}^i\right|} \leqslant \gamma, \quad \forall i \in \widetilde{\mathcal{I}}.$$

We conclude that

$$|\Gamma| \leqslant \sum_{i \in \widetilde{\mathcal{I}}} \left|\widetilde{Q}^i \cap \Gamma\right| \leqslant \gamma \sum \left|\widetilde{Q}^i\right| = \gamma \left|\bigcup_{i \in \widetilde{\mathcal{I}}} \widetilde{Q}^i\right| \leqslant \gamma |E|,$$

that finishes the proof of Lemma 5.12. $\qquad \square$

The second tool is support theorem, which implies

Lemma 5.13. Let $\kappa \in (3/4, 1)$. Suppose that $\widetilde{Q}$ is the predecessor of Q, then for each $x \in \widetilde{Q}(\kappa)$,

$$\mathbf{P}\left(\sigma_{Q(\frac{1}{2})}(x) < \tau_{\widetilde{Q}}(x)\right) \geqslant p'(\kappa) > 0,$$

where $p'(\kappa)$ only depends on d, δ and κ.

PROOF OF THEOREM 5.11. Define

$$p(\gamma) = \inf\left\{\mathbf{P}\left(\sigma_\Gamma(x) < \tau_{Q_1}(x)\right) : a \in \mathbb{S}_\delta^d, x \in Q_{1/2}, \Gamma \subset Q_1, |\Gamma| \geqslant \gamma\right\}.$$

By Proposition 3.19, we know that there exists a constant $b \in (0,1)$ such that $p(b) > 0$.

We want to prove that for each $\gamma \in (0, b]$,

$$p(\gamma) > 0 \ \text{ implies } \ p(\theta \gamma) > 0, \ \text{where } \theta = \frac{1+b}{2} < 1.$$

Assume that $p(\gamma) > 0$ for some $\gamma \in (0, b]$, and $\Gamma \subseteq Q_1$ with $\gamma \geqslant |\Gamma| \geqslant \theta\gamma$. Let Q^i and $E = \cup_{i \in \mathcal{I}} \widetilde{Q}^i$ be the sets in Lemma 5.12. Then

$$|E| \geqslant |\Gamma|/\gamma \geqslant \theta = \frac{1+b}{2}.$$

Therefore, we can find a finite subset of $\mathcal{I}$, say $\mathcal{I}_0$, and $\kappa \in (3/4, 1)$ such that

$$A := \bigcup_{i \in \mathcal{I}_0} \widetilde{Q}^i(\kappa) \text{ with } |A| \geqslant b.$$

Since $|A| \geqslant b|Q_1|$, by Proposition 3.19,

$$(5.4) \qquad \mathbf{P}\left(\sigma_A(x) < \tau_{Q_1}(x)\right) \geqslant p(b) > 0, \quad \forall x \in Q_{1/2}.$$

Suppose that $y \in \partial A = \cup_{i \in \mathcal{I}_0} \partial \widetilde{Q}^i(\kappa)$, then $y \in \partial \widetilde{Q}^i(\kappa)$ for some $i \in \mathcal{I}_0$. In this case,

$$\mathbf{P}\left(\sigma_{Q^i(1/2)}(y) < \tau_{Q_1}(y)\right) \geqslant \mathbf{P}\left(\sigma_{Q^i(1/2)}(y) < \tau_{\widetilde{Q}^i}(y)\right) \geqslant p'(\kappa) > 0,$$

due to Lemma 5.13. Set

$$B = \bigcup_{i \in \mathcal{I}_0} Q^i(1/2).$$

Then

$$\mathbf{P}\left(\sigma_B(y) < \tau_{Q_1}(y)\right) \geqslant \inf_{y \in \partial A} \mathbf{P}\left(\sigma_{Q^i(1/2)}(y) < \tau_{\widetilde{Q}^i}(y)\right)$$
$$\geqslant p'(\kappa) > 0, \quad \forall y \in \partial A.$$

The conditional version of above estimate we need below is

$$(5.5) \qquad \mathbf{P}\left(\sigma'_B < \tau'_{Q_1} \middle| \mathcal{F}_{\sigma_A(x)}\right) \geqslant p'(\kappa) > 0.$$

where

$$\sigma'_B := \inf\left\{t > \sigma_A(x) : x + x_t \in B\right\} \quad \text{and} \quad \tau'_{Q_1} := \inf\left\{t > \sigma_A(x) : x + x_t \notin Q_1\right\}.$$

Suppose that $z \in \partial B$, then $z \in \partial Q^i(1/2)$ for some $i \in \mathcal{I}_0$. Since $|\Gamma \cap Q^i| > \gamma|Q^i|$, by our assumption

$$\mathbf{P}\left(\sigma_{\Gamma \cap Q^i}(z) < \tau_{Q_1}(y)\right) \geqslant \mathbf{P}(\sigma_{\Gamma \cap Q^i}(z) < \tau_{Q^i}(z)) \geqslant p(\gamma) > 0, \quad \forall z \in \partial B.$$

Set

$$D = \bigcup_{i \in \mathcal{I}_0} Q_i.$$

Then

$$\mathbf{P}(\sigma_\Gamma(z) < \tau_{Q_1}(z)) \geqslant \mathbf{P}\left(\sigma_{\Gamma \cap D}(z) < \tau_{Q_1}(z)\right)$$
$$\geqslant \inf_{i \in \mathcal{I}_0} \mathbf{P}(\sigma_{\Gamma \cap Q^i}(z) < \tau_{Q^i}(z)) \geqslant p(\gamma) > 0, \quad \forall z \in \partial B.$$

The conditional version of above estimate we need below is

$$(5.6) \qquad \mathbf{P}\left(\sigma''_\Gamma < \tau''_{Q_1} \middle| \mathcal{F}_{\sigma'_B}\right) \geqslant p(\gamma) > 0.$$

where

$$\sigma''_\Gamma := \inf\left\{t > \sigma'_B : x + x_t \in \Gamma\right\} \quad \text{and} \quad \tau''_{Q_1} := \inf\left\{t > \sigma'_B : x + x_t \notin Q_1\right\}$$
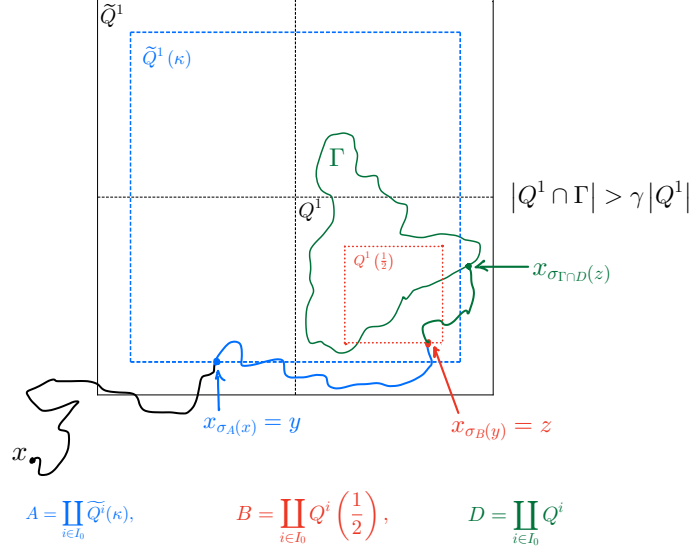
FIGURE 1. Hitting Prob.

Therefore, for each $x \in Q_{1/2}$,

$$
\begin{aligned}
& \mathbf{P}\left(\sigma_\Gamma(x) < \tau_{Q_1}(x)\right) \\
\geqslant\ & \mathbf{E}\left[\mathbf{P}\left(\sigma_A(x) < \tau_{Q_1}(x); \sigma'_\Gamma < \tau'_{Q_1} \big| \mathcal{F}_{\sigma_A(x)}\right)\right] \\
=\ & \mathbf{E}\left[\mathbf{1}_{\sigma_A(x)<\tau_{Q_1}(x)} \mathbf{P}\left(\sigma'_\Gamma < \tau'_{Q_1} \big| \mathcal{F}_{\sigma_A(x)}\right)\right] \\
=\ & \mathbf{E}\left[\mathbf{1}_{\sigma_A(x)<\tau_{Q_1}(x)} \mathbf{P}\left(\sigma'_B < \tau'_{Q_1}; \sigma''_\Gamma < \tau''_{Q_1} \big| \mathcal{F}_{\sigma_A(x)}\right)\right] \\
=\ & \mathbf{E}\left\{\mathbf{1}_{\sigma_A(x)<\tau_{Q_1}(x)} \mathbf{E}\left[\mathbf{1}_{\sigma'_B<\tau'_{Q_1}} \mathbf{P}\left(\sigma''_\Gamma < \tau''_{Q_1} \big| \mathcal{F}_{\sigma'_B}\right) \big| \mathcal{F}_{\sigma_A(x)}\right]\right\} \\
\overset{(5.6)}{\geqslant}\ & p(\gamma) \mathbf{E}\left[\mathbf{1}_{\sigma_A(x)<\tau_{Q_1}(x)} \mathbf{P}\left(\sigma'_B < \tau'_{Q_1} \big| \mathcal{F}_{\sigma_A(x)}\right)\right] \\
\overset{(5.5)}{\geqslant}\ & p(\gamma) p'(\kappa) \mathbf{P}\left(\sigma_A(x) < \tau_{Q_1}(x)\right) \\
\overset{(5.4)}{\geqslant}\ & p(b) p(\gamma) p'(\kappa) \geqslant p'(\kappa) p^2(\gamma) > 0.
\end{aligned}
$$

Since the above estimate holds for any $\Gamma \subseteq Q_1$ with $|\Gamma| \geqslant \theta\gamma$, we get $p(\theta\gamma) > 0$, provided that $p(\gamma) > 0$. Noting that $\theta < 1$, we obtain that $p(\gamma) > 0$ for all $\gamma \in (0,1)$. $\qquad \square$

5.4. Harnack Inequality and Hölder estimate

In this section, we prove some theorems of Krylov and Safonov [**KS81**] concerning (positive) L-harmonic functions. Let $\delta \in (0,1)$. Set

$$\mathscr{P}(\delta) := \Big\{ \{\mathbb{P}_x\}_{x \in \mathbb{R}^n} : (\mathbb{P}_x, X) \text{ is the strong Markov process}$$

$$\text{associate with some } a(\cdot) \in \mathbb{S}_\delta^d \Big\}.$$

Let

$$[u]_{\alpha;D} := \sup_{x,y \in D} \frac{|u(x) - u(y)|}{|x - y|^\alpha} \quad and \quad \operatorname*{osc}_D u := \sup_{x \in D} u(x) - \inf_{x \in D} u(x).$$

THEOREM 5.14 (Hölder estimate). *Suppose u is bounded in Q_1 and $Lu = 0$ in Q_1. Then there exist α and C only depending on d and δ such that*

$$(5.7) \qquad\qquad [u]_{\alpha;Q_{1/2}} \leqslant C \operatorname*{osc}_{Q_1} u.$$

PROOF. **Claim:** there exists a constant $\rho \in (0,1)$ such that for any $z \in Q_{1/2}$, $r \leqslant 1/2$,

$$(5.8) \qquad\qquad \operatorname*{osc}_{Q_{r/2}(z)} u \leqslant \rho \operatorname*{osc}_{Q_r(z)} u.$$

Assume the claim is true. Suppose that $x,y \in Q_{1/2}$ and $|x-y| \ll 1$, let $k \in \mathbb{N}$ such that $2^{-k-1} \leqslant |x-y| < 2^{-k}$.

$$|u(x) - u(y)| \leqslant \operatorname*{osc}_{Q_{2^{-k}}(x)} u \leqslant \rho \operatorname*{osc}_{Q_{2^{-k+1}}(x)} u \leqslant \cdots \leqslant C\rho^k \operatorname*{osc}_{Q_1} u$$

$$\leqslant C\rho^{-\log_2 |x-y|} \operatorname*{osc}_{Q_1} u \leqslant C|x-y|^{-\log_2 \rho} \operatorname*{osc}_{Q_1} u.$$

Therefore, the above claim implies (5.7) with $\alpha = \log_2 \rho^{-1}$.

To prove (5.8). Without loss of generality, we can assume $\inf_{x \in Q_r(z)} u = 0$ and $\sup_{x \in Q_r(z)} u = 1$. In this case, $\operatorname*{osc}_{Q_r(z)} u = 1$. Let $\Gamma := \{x \in Q_{r/2} : u(x) \geqslant 1/2\}$, we may assume $|\Gamma| \geqslant \frac{1}{2}|Q_{r/2}|$, if not, we replace u by $1-u$. For any $x \in Q_{r/2}$, by Itô's formula, Theorem 5.11 and scaling,

$$u(x) = \mathbb{E}_x u(X_{\tau_{Q_r} \wedge \sigma_\Gamma}) \geqslant \frac{1}{2}\mathbb{P}_x(\sigma_\Gamma < \tau_{Q_r}) \geqslant \frac{1}{2}p(2^{-d-1}).$$

Hence we get

$$\operatorname*{osc}_{Q_{r/2}(z)} u \leqslant 1 - \frac{1}{2}p(2^{-d-1}) =: \rho = \rho \operatorname*{osc}_{Q_r(z)} u.$$

$\square$

THEOREM 5.15 (Harnack inequality). *Suppose $a \in \mathbb{S}_\delta^d$ and $L = a_{ij}\partial_{ij}$. There exists C depending only on δ such that if u is nonnegative, bounded in Q_4, and $u(X_{t \wedge \tau_{Q_4}})$ is a martingale, then $u(x) \leqslant Cu(y)$ if $x,y \in Q_1$.*

PROOF. If we look at $u + \delta$ and let $\delta \to 0$, we may assume $u > 0$. By looking at Cu, we may assume $\inf_{Q_{1/2}} u = 1$. By Theorem , we know that u is Hölder continuous in Q_1, so there exists

$$y \in Q_{1/2} \quad such \ that \quad u(y) = 1.$$

We want to show that u is bounded above by a constant in Q_1, where the constant depends only on δ.

By the support theorem and scaling, if $x \in Q_{1/2}$, there exists δ such that

$$\mathbb{P}_y \left(\sigma_{Q_{1/2}(x)} < \tau_{Q_2} \right) \geqslant \delta.$$

By scaling, if $z \in Q_{1/2}(x)$, then $\mathbb{P}_z \left(\sigma_{Q_{1/4}(x)} < \tau_{Q_2} \right) \geqslant \delta$. So by the strong Markov property,

$$\mathbb{P}_z \left(\sigma_{Q_{1/4}(x)} < \tau_{Q_2} \right) \geqslant \delta^2.$$

Repeating and using induction,

$$\mathbb{P}_y \left(\sigma_{Q_{2^{-k}}(x)} < \tau_{Q_2} \right) \geqslant \delta^k.$$

Then

$$1 = u(y) \geqslant \mathbb{E}_y \left[u \left(X_{\sigma_{Q_{2^{-k}}(x)}} \right) ; \sigma_{Q_{2^{-k}}(x)} < \tau_{Q_2} \right]$$

$$\geqslant \delta^k \left(\inf_{Q_{2^{-k}}(x)} u \right),$$

or

(5.9) $$\inf_{Q_{2^{-k}}(x)} u \leqslant \delta^{-k}, \quad \forall k \geqslant 1.$$

By the proof of Theorem 5.14, there exists $\rho < 1$ such that

$$\underset{Q_{2^{-k-1}}(x)}{\mathrm{osc}} \, u \leqslant \rho \underset{Q_{2^{-k}}(x)}{\mathrm{osc}} \, u.$$

Take N large so that $\rho^{-N} \geqslant 1/\left(\delta - \delta^2 \right)$. Then

$$\underset{Q_{2^{N-k}}(x)}{\mathrm{Osc}} \, u \geqslant \rho^{-N} \underset{Q_{2^{-k}}(x)}{\mathrm{Osc}} \, u \geqslant \frac{1}{\delta - \delta^2} \underset{Q_{2^{-k}}(x)}{\mathrm{Osc}} \, u.$$

Take K large so that $\sqrt{d}2^{N-K} < 1/8$. Suppose there exists $x_0 \in Q_1(y)$ such that $u(x_0) \geqslant \delta^{-K-1}$.

We will construct a sequence $x_1, x_2, \ldots$ by induction such that $u(x_j) \geqslant \delta^{-K-j-1}$.

Suppose we have $x_j \in Q_{2^{N+1-K-j}}(x_{j-1})$ with $u(x_j) \geqslant \delta^{-K-j-1}$, $j \leqslant n$. Since $|x_j - x_{j-1}| < \sqrt{d}2^{N+1-K-j}, 1 \leqslant j \leqslant n$, and $|x_0 - y| \leqslant 1$, then $|x_n - y| < 2$. Since $u(x_n) \geqslant \delta^{-K-n-1}$ and by (5.9), $\inf_{Q_{2^{-K-n}}(x_n)} u \leqslant \delta^{-K-n}$,

$$\underset{Q_{2^{-K-n}}(x_n)}{\mathrm{Osc}} \, u \geqslant \delta^{-K-n} \left(\delta^{-1} - 1 \right).$$

So $\mathrm{osc}_{Q_{2^{N-K-n}}(x_n)} u \geqslant \delta^{-K-n-2}$, which implies that there exists $x_{n+1} \in Q_{2^{N-K-n}}(x_n)$ with $u(x_{n+1}) \geqslant \delta^{-K-n-2}$ because u is nonnegative. By induction we obtain a sequence x_n with $x_n \in Q_3(y)$ and $u(x_n) \to \infty$. This contradicts the boundedness of u on Q_4. Therefore u is bounded on Q_1 by δ^{-K-1}. $\qquad \square$

CHAPTER 6

Divergence Form Operators

Let $a \in \mathbb{S}_\delta^d$. One can also consider the following Dirichlet form

$$\mathcal{E}(u,v) := \int_{\mathbb{R}^d} a_{ij}\partial_i u \partial_i v = \mathcal{E}(v,u), \quad u,v \in H^1.$$

Then,

$$\mathcal{E}(u,v) = -\langle Lu, v \rangle = -\langle u, Lv \rangle, \quad u, v \in C_c^\infty,$$

where

$$Lu = \partial_j(a_{ij}\partial_i u).$$

By our assumption on a,

$$\mathcal{E}(u,u) \asymp \int_{\mathbb{R}^d} |\nabla u|^2 = \|u\|_{H^1}^2.$$

Assume that $a \in \mathbb{S}_\delta^d \cap C^\infty$, and $(\mathbb{P}_x, X_t)$ be the diffusion process associated to For any $f \in L^1 \cap L^\infty$, set $P_t f(x) = \mathbb{E}_x f(X_t)$. Then

$$\partial_t P_t f = L P_t f, \quad f \in C_c^\infty.$$

By the symmetricity of L, we have

$$\partial_t \langle f, P_t^* g \rangle = \partial_t \langle P_t f, g \rangle = \langle L P_t f, g \rangle = \langle P_t L f, g \rangle = \langle f, L P_t^* g \rangle, \quad f, g \in C_c^\infty.$$

This yields that $P_t^* f = P_t f$. Consequently,

$$\int P_t f = \partial_t \langle P_t f, 1 \rangle = \langle f, P_t 1 \rangle = \int f.$$

Assume $f \geqslant 0$, $f \in C_c^\infty$ and $\|f\|_1 = 1$.

$$\frac{\mathrm{d}}{\mathrm{d}t}\|P_t f\|_2^2 = \partial_t \langle P_t f, \partial_t P_t f \rangle = 2\langle P_t f, L P_t f \rangle \leqslant -2\delta \|\nabla P_t f\|_2^2$$

By Nash's inequality:

$$\|u\|_2 \leqslant C(d)\|u\|_1^{\frac{2}{d+2}}\|\nabla u\|_2^{\frac{d}{d+2}}$$

and the fact that $\|P_t f\|_1 \leqslant \|f\|_1 = 1$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\|P_t f\|_2^2 \leqslant -c\left(\frac{\|P_t f\|_2}{\|P_t f\|_1^{\frac{2}{d+2}}}\right)^{\frac{2(d+2)}{d}} \leqslant -c\left(\|P_t f\|_2^2\right)^{1+\frac{2}{d}}$$

Put $\theta(t) = \|P_t f\|_2^2$. Then $\theta'(t) \leqslant -c\left(\theta(t)\right)^{1+\frac{2}{d}}$ and $\theta(t) > 0$. This yields

$$\left[\theta^{-\frac{2}{d}}(t)\right]' \geqslant c > 0 \Rightarrow \theta^{-\frac{2}{d}}(t) \geqslant ct + \theta^{-\frac{2}{d}}(0).$$

Thus,

$$\|P_t f\|_2^2 \leqslant \theta(t) \leqslant Ct^{-\frac{d}{2}}, \quad \|f\|_1 = 1.$$

So we obtain

(6.1) $$\|P_t\|_{L^1 \to L^2} \leqslant Ct^{-\frac{d}{4}}$$

For any $f \in L^2$, $g \in L^1$,

$$\langle P_t f, g \rangle = \langle f, P_t g \rangle \leqslant C\|f\|_2 \|P_t g\|_2 \leqslant Ct^{-\frac{d}{4}} \|g\|_1 \|f\|_2.$$

This yields

(6.2) $$\|P_t\|_{L^2 \to L^\infty} \leqslant Ct^{-\frac{d}{4}}$$

(6.1) and (6.2) imply that

$$\|P_t\|_{L^1 \to L^\infty} \leqslant \|P_{\frac{t}{2}}\|_{L^1 \to L^2} \|P_{\frac{t}{2}}\|_{L^2 \to L^\infty} \leqslant Ct^{-\frac{d}{2}}.$$

Therefore,

$$p(t, x, y) \leqslant Ct^{-\frac{d}{2}}.$$

Set

$$P_t^\varphi f(x) := e^{\varphi(x)} P_t(e^{-\varphi} f)(x)$$

$$\langle L^\varphi f, g \rangle = \langle \partial_t P_t^\varphi f, g \rangle|_{t=0} = \langle e^\varphi \partial_t P_t(e^{-\varphi} f), g \rangle|_{t=0}$$

(6.3)
$$= -\int_{\mathbb{R}^d} a_{ij} \partial_i(e^{-\varphi} f) \partial_j(e^\varphi g)$$

$$= \langle Lf, g \rangle + \int a_{ij}(\partial_i g \partial_j \varphi f - \partial_i f \partial_j \varphi g) + \int a_{ij} \partial_i \varphi \partial_j \varphi f g$$

$$\frac{\mathrm{d}}{\mathrm{d}t} \|P_t^\varphi f\|_2^2 = \langle L^\varphi P_t^\varphi f, P_t^\varphi f \rangle + \int a_{ij} \partial_i \varphi \partial_j \varphi f^2.$$

Let $\varphi(x) = \beta \cdot x$ with $\beta \in \mathbb{R}^d$, and $\|f\|_1 = 1$. Then

$$\frac{\mathrm{d}}{\mathrm{d}t} \|P_t^\varphi f\|_2^2 \leqslant C|\beta|^2 \|P_t f\|_2^2 - c\left(\|P_t f\|_2^2\right)^{1+\frac{2}{d}}$$

$$\theta'(t) \leqslant C|\beta|^2 \theta(t) - c(\theta(t))^{1+\frac{2}{d}}$$

CHAPTER 7

Appendix

7.1. Probabilistic terminology

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and $(E, \mathcal{E})$ be a measurable space. $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ a measurable map, and $\mathcal{G}$ a σ-field $\subseteq \mathcal{F}$.

When $E = \mathbb{R}$, we define the **conditional expectation** of X given $\mathcal{G}$, $\mathbf{E}(X|\mathcal{G})$, to be any random variable Y that satisfies

(a) $Y \in \mathcal{G}$;
(b) for all $A \in \mathcal{G}$, $\mathbf{E}(X; A) = \mathbf{E}(Y; A)$.

$Q_{\mathcal{G}} : \Omega \times \mathcal{E} \to [0, 1]$ is said to be a **regular conditional distribution** (RCD) for X given $\mathcal{G}$ if

(a) For each $A \in \mathcal{E}$, $\omega \mapsto Q_{\mathcal{G}}(\omega, A)$ is a version of $\mathbf{E}(\mathbf{1}_A(X)|\mathcal{G})$;
(b) For a.e. $\omega \in \Omega$, $A \mapsto Q_{\mathcal{G}}(\omega, A)$ is a probability measure.

If $E = \Omega$, $X(\omega) = \omega$, then $Q_{\mathcal{G}}$ is called a **regular conditional probability**.

The following results can be found in Durrett's book [**Dur19**].

Proposition 7.1.   *(i) If $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$, then*

$$(7.1) \qquad \mathbf{E}[(X|\mathcal{G}_2)|\mathcal{G}_1] = \mathbf{E}(X|\mathcal{G}_1)$$

 (ii) Assume that $X \in \mathcal{F}$ and $Y \in \mathcal{G} \subseteq \mathcal{F}$, then

$$(7.2) \qquad \mathbf{E}(XY|\mathcal{G}) = \mathbf{E}(X|\mathcal{G})Y.$$

(iii) (Jesen's inequality) If φ is a convex function, then

$$(7.3) \qquad \mathbf{E}(\varphi(X)|\mathcal{G}) \leqslant \varphi(\mathbf{E}(X|\mathcal{G})).$$

Proposition 7.2. *Let $Q_{\mathcal{G}}$ be a RCD for X given $\mathcal{G}$. If $f : E \to \mathbb{R}$ satisfying $\mathbf{E}|f(X)| < \infty$, then*

$$\mathbf{E}(f(X)|\mathcal{G})(\omega) = \int_E f(x) Q_{\mathcal{G}}(\omega, \mathrm{d}x) \quad a.s..$$

THEOREM 7.3. *RCD exists if E is a **standard measure space** and $\mathcal{E} = \mathcal{B}(E)$.*

Proposition 7.4. *Assume $X \geqslant 0$, $f : \mathbb{R}_+ \to \mathbb{R}_+$ such that $f \in C^1(\mathbb{R}_+)$ and $f(0) = 0$. Then*

$$(7.4) \qquad \mathbf{E}f(X) = \int_0^\infty f'(t)\mathbf{P}(X > t)\mathrm{d}t.$$

Exercise 7.5. If $X \geqslant 0$, $f : \mathbb{R}_+ \to \mathbb{R}_+$ such that $f \in C^1(\mathbb{R}_+)$ and $f(\infty) = 0$. Then

$$(7.5) \qquad \mathbf{E}f(X) = -\int_0^\infty f'(t)\mathbf{P}(X \leqslant t)\mathrm{d}t.$$

7.2. Maximal Principle

Consider the linear parabolic equation:

$$\partial_t u = a_{ij}\partial_{ij}u + b_i\partial_i u + cu,$$

where a_{ij} is uniformly elliptic, b_i and c are bounded and continuous.

Proposition 7.6 (Weak Maximum Principle). *If $c(x,t) \leqslant 0$ and u is bounded, then*

$$\sup_{\mathbb{R}_+ \times \mathbb{R}^d} u = \sup_{\{0\} \times \mathbb{R}^d} u.$$

Proposition 7.7 (Strong Maximum Principle). *If u achieves its maximum (or minimum) at a point $(t_0, x_0) \in \mathbb{R}_+ \times \mathbb{R}^d$, then u is constant in $[0, t_0] \times \mathbb{R}^d$.*

7.3. Schauder estimate

Let $\mathscr{S}$ be the Schwartz space of all rapidly decreasing functions, and $\mathscr{S}'$ the dual space of $\mathscr{S}$ called Schwartz generalized function (or tempered distribution) space. Given $f \in \mathscr{S}$, let $\mathscr{F}f = \hat{f}$ be the Fourier transform defined by

$$\hat{f}(\xi) := \int_{\mathbb{R}^d} e^{-i2\pi\xi\cdot x} f(x)\mathrm{d}x.$$

Let $\chi : \mathbb{R}^d \to [0,1]$ be a smooth radial function with

$$\chi(\xi) = 1, \ |\xi| \leqslant 1, \ \chi(\xi) = 0, \ |\xi| \geqslant 3/2.$$

Define

$$\varphi(\xi) := \chi(\xi) - \chi(2\xi).$$

It is easy to see that $\varphi \geqslant 0$ and supp $\varphi \subset B_{3/2} \setminus B_{1/2}$ and formally

$$(7.6) \qquad \sum_{j=-k}^{k} \varphi(2^{-j}\xi) = \chi(2^{-k}\xi) - \chi(2^{k+1}\xi) \overset{k\to\infty}{\to} 1.$$

In particular, if $|j - j'| \geqslant 2$, then

$$\mathrm{supp}\varphi(2^{-j}\cdot) \cap \mathrm{supp}\varphi(2^{-j'}\cdot) = \varnothing.$$

From now on we shall fix such χ and φ and define

$$\Delta_j f := \mathscr{F}^{-1}(\varphi(2^{-j}\cdot)\mathscr{F}f), \quad j \in \mathbb{Z}.$$

Set $h := \mathscr{F}^{-1}(\varphi)$, then $h_j := \mathscr{F}^{-1}(\varphi(2^{-j}\cdot)) = 2^{jd}h(2^j\cdot)$. Noting that we have

$$\int_{\mathbb{R}^d} h_j = \varphi(0) = 0.$$

We first recall the following useful lemmas.

Lemma 7.8. *Let $\alpha \in (0,1)$. For any $u \in C^\alpha$, it holds that*

$$(7.7) \qquad \frac{1}{C}\sup_{j\in\mathbb{Z}} 2^{j\alpha}\|\Delta_j u\|_0 \leqslant [u]_\alpha \leqslant C\sup_{j\in\mathbb{Z}} 2^{j\alpha}\|\Delta_j u\|_0,$$

where

$$[u]_\alpha := \sup_{x\neq y} \frac{|u(x) - u(y)|}{|x-y|^\alpha},$$

and C only depends on d and α.

PROOF. 1) For any $x \neq y$, we have

$$|u(x) - u(y)| \leqslant \sum_j |\Delta_j u(x) - \Delta_j u(y)| \leqslant \sum_j (|x-y| \|\nabla \Delta_j u\|_0) \wedge (2\|\Delta_j u\|_0)$$

$$\leqslant C \sum_j (2^j |x-y| \wedge 1) \|\Delta_j u\|_0$$

$$\leqslant C 2^{j\alpha} \|\Delta_j u\|_0 \left(|x-y| \sum_{j \leqslant \log_2 |x-y|} 2^{j(1-\alpha)} + \sum_{j \geqslant \log_2 |x-y|} 2^{-j\alpha} \right)$$

$$\leqslant C |x-y|^{\alpha} \sup_j 2^{j\alpha} \|\Delta_j u\|_0.$$

2) For any $j \in \mathbb{Z}$ and $x \in \mathbb{R}^d$, we have

$$|\Delta_j u(x)| = \left| \int_{\mathbb{R}^d} u(x-y) h_j(y) \mathrm{d}y \right| = \left| \int_{\mathbb{R}^d} [u(x-y) - u(x)] h_j(y) \mathrm{d}y \right|$$

$$\leqslant [u]_\alpha 2^{jd} \int_{\mathbb{R}^d} |y|^\alpha |h(2^j y)| \mathrm{d}y \leqslant C 2^{-j\alpha} [u]_\alpha.$$

$\square$

Lemma 7.9. *There is a constant* $C = C(d, \alpha)$, *such that for any* $u \in C^{2,\alpha}$ *and* $\lambda \geqslant 0$,

(7.8) $$\lambda \|u\|_0 \leqslant C \|\lambda u - \Delta u\|_0$$

and

(7.9) $$\lambda [u]_\alpha + [\nabla^2 u]_\alpha \leqslant C[\lambda u - \Delta u]_\alpha$$

PROOF. 1) If there exits $x_0 \in \mathbb{R}^d$ such that $u(x_0)($ or $-u(x_0)) = \|u\|_0$, then $\Delta u(x_0) \leqslant 0$ $(\Delta u(x_0) \geqslant 0)$. This implies $|\lambda u(x_0)| \leqslant |\lambda u(x_0) - \Delta u(x_0)|$; If such x_0 does not exist, then we can consider function $u_R = u\chi(\cdot/R)$ $(R \gg 1)$.

2) We only prove the case $\lambda = 0$. Let $f = \Delta u$. Define

$$\varphi^{kl}(\xi) := \frac{\xi_k \xi_l}{|\xi|^2} \varphi(\xi), \ h^{kl}(x) := \mathscr{F}^{-1}(\varphi^{kl})(x); \ \varphi_j^{kl}(\xi) := \varphi^{kl}(2^{-j}\xi), \ h_j^{kl}(x) := 2^{jd} h^{kl}(2^j x).$$

It is easy to see

$$\partial_{kl} u = \sum_{j \in \mathbb{Z}} u_j^{kl} := \sum_{j \in \mathbb{Z}} \varphi_j^{kl}(D) f = \sum_{j \in \mathbb{Z}} h_j^{kl} * f.$$

For any $x \in \mathbb{R}^d$, noticing $h^{kl} \in \mathscr{S}(\mathbb{R}^d)$ and $\int h^{kl} = \varphi(0) = 0$, we get

$$|u_j^{kl}(x)| = \left| \int_{\mathbb{R}^d} h_j^{kl}(y) f(x-y) \mathrm{d}y \right| = \left| \int_{\mathbb{R}^d} h^{kl}(z)(f(x-2^{-j}z) - f(x)) \mathrm{d}z \right|$$

$$\leqslant \int_{\mathbb{R}^d} |h^{kl}(z)| \cdot [f]_\alpha |2^{-j}z|^\alpha \mathrm{d}z \leqslant C[f]_\alpha 2^{-j\alpha}$$

This together with Lemma 7.8 yields that

$$[\nabla^2 u]_\alpha \leqslant C \sup_{j,k,l} 2^{j\alpha} \|\Delta_j \partial_{kl} u\|_0 \leqslant C[f]_\alpha.$$

$\square$

7.4. Monge-Ampère Equation

Lemma 7.10 (Area formula). *Consider a locally Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}^d$ and a Borel set $A \subseteq \mathbb{R}^d$. Then the function $y \mapsto N_A(y) := \mathrm{card}\{f^{-1}(y) \cap A\}\}$ is measurable and*

$$\int_A |\det(\nabla f(x))| \mathrm{d}x = \int_{\mathbb{R}^n} N_A(y) \mathrm{d}y \geqslant \mathscr{L}^d(f(A)).$$

Consequently, for any $g \geqslant 0$,

(7.10) $$\int_{f(A)} g(y) \mathrm{d}y \leqslant \int_A g(f(x)) |\det \nabla f(x)| \mathrm{d}x.$$

To motivate the definition of weak solutions to (3.7), given an open set $D \subset \mathbb{R}^n$, consider $u : D \to \mathbb{R}$ a convex function of class C^2 satisfying (3.7) for some $f : D \to \mathbb{R}^+$. Then given any Borel set $E \subset D$, it follows by the area formula that

$$\int_E f \, \mathrm{d}x = \int_E \det D^2 u \, \mathrm{d}x = |\nabla u(E)|.$$

Notice that while the above argument needs u to be of class C^2, the identity

$$\int_E f = |\nabla u(E)|$$

makes sense if u is only of class C^1. To find a definition when u is merely convex one could try to replace the gradient $\nabla u(x)$ with the subdifferential $\partial u(x)$ and ask for the above equality to hold for any Borel set E. Here $\partial u(x)$ is given by

$$\partial u(x) := \left\{ p \in \mathbb{R}^d : u(y) \geqslant u(x) + \langle p, y - x \rangle \quad \forall y \in D \right\}.$$

This motivates the following definition:

Definition 7.11. Given an open set $D \subset \mathbb{R}^n$ and a convex function $u : D \to \mathbb{R}$, we define the Monge-Ampère measure associated to u by

$$\mu_u(E) := \left| \bigcup_{x \in E} \partial u(x) \right|$$

The basic idea of Alexandrov was to say that u is a weak solution of (3.7) if $\mu_u|_D = \nu|_D$.

Lemma 7.12. *Let $u, v : D \to \mathbb{R}$ be convex functions. Then*

$$\mu_{u+v} \geqslant \mu_u + \mu_v \quad and \quad \mu_{\lambda u} = \lambda^n \mu_u \quad \forall \lambda > 0.$$

The following result is the celebrated Alexandrov maximum principle.

THEOREM 7.13. *Let D be an open bounded convex set, and let $u : D \to \mathbb{R}$ be a convex function such that $u|_{\partial D} = 0$. Then there exists a dimensional constant $C = C(d)$ such that*

(7.11) $$|u(x)| \leqslant C(d) \, \mathrm{diam}(D)^{\frac{d-1}{d}} \, \mathrm{dist}(x, \partial D)^{\frac{1}{d}} |\partial u(D)|^{\frac{1}{d}}, \quad \forall x \in D.$$

PROOF. Let $(x, u(x))$ be a point on the graph of u, and consider the convex "conical" function $y \mapsto \widehat{C}_x(y)$ with vertex at $(x, u(x))$ that vanishes on ∂D. Since $u \leqslant \widehat{C}_x$ in D (by the convexity of u), Lemma 2.7 implies that

$$\left| \partial \widehat{C}_x(x) \right| \leqslant \left| \partial \widehat{C}_x(D) \right| \leqslant |\partial u(D)|;$$

so, to conclude the proof, it suffices to bound $|\partial \widehat{C}_x(x)|$ from below. It is not hard to see

- $\partial \widehat{C}_x(x)$ contains the ball B_ρ with $\rho = |u(x)|/\mathrm{diam}(D)$
- $\partial \widehat{C}_x(x)$ contains a vector of norm $|u(x)|/\mathrm{dist}(x, \partial D)$

Thus,

$$\partial \widehat{C}_x(x) \supset B_\varrho(0) \cup \{q\}, \quad |q| = |u(x)|/\mathrm{dist}(x, \partial D).$$

Since $\partial \widehat{C}_x(x)$ is convex, it follows that $\partial \widehat{C}_x(x)$ contains the cone $\mathcal{C}$ generated by q and $\Sigma_q := \{p \in B_\rho : \langle p, q \rangle = 0\}$. Therefore

$$c(d)\rho^{d-1}|q| = |\mathcal{C}| \leqslant |\partial u(D)|.$$

$\square$

THEOREM 7.14. *Let D be an open bounded convex set, and let ν be a Borel measure on D with $\nu(D) < \infty$. Then there exists a unique convex function $u : D \to \mathbb{R}$ solving the Dirichlet problem*

$$\begin{cases} \mu_u = \nu & \text{in } D \\ u = 0 & \text{on } \partial D \end{cases}$$

PROOF. By the stability result proved in Lemma below, since any finite measure can be approximated in the weak* topology by a finite sum of Dirac deltas, we only need to solve the Dirichlet problem when $\nu = \sum_{i=1}^N \alpha_i \delta_{x_i}$ with $x_i \in D$ and $\alpha_i > 0$. To prove existence of a solution, we use the so-called Perron method: we define

$$\mathcal{S}[\nu] := \{v : \Omega \to \mathbb{R} \text{ convex} : \ v|_{\partial \Omega} = 0, \mu_v \geqslant \nu \text{ in } \Omega\}$$

and we show that the largest element in $\mathcal{S}[\nu]$ is the desired solution. We split the argument into several steps.

Step 1: $\mathcal{S}[\nu] \neq \emptyset$. To construct an element of $\mathcal{S}[\nu]$, we consider the "conical" function C_{x_i}, that is 0 on $\partial \Omega$ and takes the value -1 at its vertex x_i. The Monge–Ampère measure of this function is concentrated at x_i and has mass equal to some positive number β_i corresponding to the measure of the set of supporting hyperplanes at x_i. Now, consider the convex function $\bar{v} = \sum_{i=1}^N \lambda C_{x_i}$, where λ has to be chosen. We notice that $\bar{v}|_{\partial \Omega} = 0$. In addition, provided λ is sufficiently large, Lemma below implies that

$$\mu_{\bar{v}} \geqslant \sum_{i=1}^N \mu_{\lambda \hat{C}_{x_i}} = \sum_{i=1}^N \lambda^d \mu_{\hat{C}_{x_i}} = \sum_{i=1}^N \lambda^d \beta_i \delta_{x_i} \geqslant \sum_{i=1}^N \alpha_i \delta_{x_i} = \nu.$$

This yields $\bar{v} \in \mathcal{S}[\nu]$.

Step 2: $v_1, v_2 \in \mathcal{S}[\nu] \Rightarrow w := \max\{v_1, v_2\} \in \mathcal{S}[\nu]$. Set

$$\Omega_0 := \{v_1 = v_2\}, \quad \Omega_1 := \{v_1 > v_2\}, \quad \text{and} \quad \Omega_2 := \{v_1 < v_2\}$$

Also, given a Borel set $E \subseteq \Omega$, consider $E_i = E \cap \Omega_i$.

Since Ω_1 and Ω_2 are open sets, $w|_{\Omega_1} = v_1$ and $w|_{\Omega_2} = v_2$,

$$\partial w(E_1) = \partial v_1(E_1), \quad \partial w(E_2) = \partial v_2(E_2).$$

In addition, since $w = v_1$ on Ω_0 and $w \geqslant v_1$ everywhere else, we have

$$\partial v_1(E) \subseteq \partial w(E_0).$$

Therefore,
$$\mu_w(E) \geqslant \mu_{v_1}(E_0 \cup E_1) + \mu_{v_2}(E_2) \geqslant \nu(E).$$

Step 3: $u := \sup_{v \in S[\nu]} v$ belongs to $S[\nu]$. Let $w_m \uparrow u$ locally uniformly. Then $\mu_{w_m} \rightharpoonup *\mu_u$. Also, we deduce immediately that $u|_{\partial\Omega} = 0$ by construction; hence, $u \in \mathcal{S}[\nu]$.

Step 4: The measure μ_u is supported at the points $\{x_1, \cdots x_N\}$. Otherwise, there exists a set $E \subseteq D$ such that
$$E \cap \{x_1, \ldots, x_N\} = \emptyset \quad \text{and} \quad |\partial u(E)| = \mu_u(E) > 0$$

Therefore,
$$\left|\partial u(E) \backslash [\cup_{i=1}^N \partial u(x_i) \cup \partial u(\partial D)]\right| = |\partial u(E)| > 0$$

Let $x_0 \in E$ and $p \in \partial u(x_0) \backslash [\cup_{i=1}^N \partial u(x_i) \cup \partial u(\partial D)]$. Then there exists $\delta > 0$ such that

(7.12)
$$u \geqslant \ell_{x_0,p} + 2\delta \quad \text{on } \{x_1, \ldots, x_N\} \cup \partial\Omega,$$

where $\ell_{x_0,p}(x) = u(x_0) + p \cdot (x - x_0)$. Set $\bar{u} := \max\{\ell_{x_0,p} + \delta, u\} \gneqq u$. Notice that $\bar{u}$ is convex, $\bar{u} \geqslant u$, and it follows by (7.12) that $\bar{u} = u$ in a neighborhood of $\{x_1, \ldots, x_N\} \cup \partial\Omega$. In particular, $\bar{u}|_{\partial\Omega} = 0$ and $\partial\bar{u}(x_i) = \partial u(x_i)$, which implies that $u \log eqq\bar{u} \in \mathcal{S}[\nu]$. This is a contradiction.

Step 5: $\mu_u = \nu$. By *Step 3* and *Step 4*, we know that $\mu_u = \sum_{i=1}^N \beta_i \delta_{x_i}$ with $\beta_i \geqslant \alpha_i$. Assume that $\beta_1 = \mu_u(x_1) > \nu(x_1) = \alpha_1$. $\qquad\square$

Bibliography

[Bas98]   Richard F Bass. *Diffusions and elliptic operators*. Springer Science & Business Media, 1998.

[Dur19]   Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[Eng91]   Hans-Jürgen Engelbert. On the theorem of t. yamada and s. watanabe. *Stochastics: An International Journal of Probability and Stochastic Processes*, 36(3-4):205–216, 1991.

[FGP10]   Franco Flandoli, Massimiliano Gubinelli, and Enrico Priola. Well-posedness of the transport equation by stochastic perturbation. *Inventiones mathematicae*, 180(1):1–53, 2010.

[Hua01]   Zhiyuan Huang. *Fundamentals of stochastic analysis, version 2 (in Chinese)*. Science Press, 2001.

[IW14]   Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*, volume 24. Elsevier, 2014.

[Kry09]   N. V. Krylov. *Controlled diffusion processes*, volume 14 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2009. Translated from the 1977 Russian original by A. B. Aries, Reprint of the 1980 edition.

[Kry21a] NV Krylov. On diffusion processes with drift in L_d. *Probability Theory and Related Fields*, 179(1):165–199, 2021.

[Kry21b] NV Krylov. On strong solutions of Itô's equations with $\sigma \in W_d^1$ and $b \in L_d$. *The Annals of Probability*, 49(6):3142–3167, 2021.

[KS79]   N. V. Krylov and M. V. Safonov. An estimate for the probability of a diffusion process hitting a set of positive measure. *Dokl. Akad. Nauk SSSR*, 245(1):18–20, 1979.

[KS81]   N V Krylov and M V Safonov. A certain property of solutions of parabolic equations with measurable coefficients. *Mathematics of the USSR-Izvestiya*, 16(1):151, feb 1981.

[RY13]   Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

[Yan21]   Jiaan Yan. *Lecture note on Measure Theory*. Science Press, China, third edition, 2021.