

CONSTRUCTION OF DIFFUSION PROCESSES I

GUOHUAN ZHAO

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:

$$\begin{aligned} \rho(t, x) + \partial_t \rho(t, x) \Delta t &\approx \rho(t + \Delta t, x) = \int_{\mathbb{R}^3} \rho(t, x - y) \varphi_{\Delta t}(y) dy \\ &\approx \rho(t, x) \int_{\mathbb{R}^3} \varphi_{\Delta t}(y) dy - \nabla \rho(t, x) \cdot \int_{\mathbb{R}^3} y \varphi_{\Delta t}(y) dy \end{aligned}$$

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

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. 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. $(W_t)_{t \geq 0}$ 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 $(W_u)_{0 \leq u \leq s}$ before time s ;
- (2) **Path continuity:** $(W_t)_{t \geq 0}$ 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 regarded 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 $X_t : \Omega \rightarrow E$ 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 $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ for every n and every finite subset $I = (t_1, t_2, \dots, t_n)$ 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 1.1 (Kolmogorov's consistency Theorem, cf. [Yan21]). *Let E be a standard measure space. Assume that we are given for every $t_1, \dots, t_n \in \mathbf{T}$ a probability measure $\mu_{t_1 \dots t_n}$ on E^n , and that these probability measures satisfy:*

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

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

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

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

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

Let $\mathbf{T} = \mathbb{R}_+$ and E be a Polish space. By Theorem 1.1, 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 1.1. 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 1.2 (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*

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

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

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

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 1.3 below, there is a null set $\mathcal{N} \subseteq \Omega$ and a measurable function $X : \Omega \times I \rightarrow \mathbb{R}^d$, such that for each $\omega \notin \mathcal{N}$,

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

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

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

By Fubini theorem, there exists a \mathcal{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 $Y_{t_n} \xrightarrow[\mathbb{I} \setminus N \ni t_n \rightarrow t_0]{\mathbf{P}} Y_{t_0}$. On the other hand, $Y_{t_n} \stackrel{\text{a.s.}}{=} X_{t_n} \rightarrow X_{t_0}$, so we have $X_{t_0} \stackrel{\text{a.s.}}{=} Y_{t_0}$. Therefore, X is a continuous version of Y . \square

Lemma 1.3 (Fractional Sobolev inequality). *Let D be an open set in \mathbb{R}^n , $p > n$ and $s \in (n/p, 1)$. Let $f : D \rightarrow \mathbb{R}^d$ be a measurable function. Assume*

$$\iint_{D \times D} \frac{|f(x) - f(y)|^p}{|x - y|^{n+sp}} dx dy < \infty.$$

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

$$\sup_{x, y \in D} \frac{|\tilde{f}(x) - \tilde{f}(y)|}{|x - y|^{s - \frac{n}{p}}} \leq C \left(\iint_{D \times D} \frac{|f(x) - f(y)|^p}{|x - y|^{n+sp}} dx dy \right)^{1/p}, \quad (1.2)$$

Here C only depends on n, s, p and D .

Thanks to Theorem 1.2 and the discussion after Theorem 1.1, we can first construct a probability measure \mathbf{P} on $\Omega = (\mathbb{R}^d)^{\mathbb{R}_+}$ such that

$$\begin{aligned} & \mathbf{P}(Y_{t_1} \in A_1, \dots, Y_{t_n} \in A_n) \\ &= \int_{A_1} \cdots \int_{A_n} p_{t_1}(x_1) p_{t_2-t_1}(x_1 - x_2) \cdots p_{t_n-t_{n-1}}(x_{n-1} - x_n) dx_1 \cdots dx_n, \end{aligned}$$

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 1.2, 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.

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 $\{X_s: s \leq t\}$. The Markov property then says that, for any times $s \leq t$ and bounded measurable function $f: E \rightarrow \mathbb{R}$, the expected value of $f(X_t)$ conditional on \mathcal{F}_s is a function of X_s . Equivalently,

$$\mathbf{E}[f(X_t) \mid \mathcal{F}_s] = \mathbf{E}[f(X_t) \mid X_s], \quad \text{a.s.} \quad (2.1)$$

More generally, this idea makes sense with respect to any filtered probability space $\mathbb{F} = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbf{P})$. A process X is Markov with respect to \mathbb{F} if it is adapted and (2.1) holds for times $s \leq t$.

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

$$\begin{aligned} Q: E \times \mathcal{E} &\rightarrow \mathbb{R}_+ \cup \{\infty\}, \\ (x, A) &\mapsto N(x, A) \end{aligned}$$

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

$$Qf(x) = \int_E f(y)Q(x, dy).$$

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 $(P_{s,t})_{t \geq s}$, 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 2.2. 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

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

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

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

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 $(P_t)_{t \geq 0}$ forms a semigroup.

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

Proposition 2.1. *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 $0 = t_0 < t_1 < \dots < t_n$ and bounded measurable function $f: E^{n+1} \rightarrow \mathbb{R}$,*

$$\begin{aligned} & \mathbf{E}[f(X_{t_0}, \dots, X_{t_n})] \\ &= \int \int \dots \int f(x_0, \dots, x_n) P_{t_n - t_{n-1}}(x_{n-1}, dx_n) \dots P_{t_1 - t_0}(x_0, dx_1) \mu(dx_0). \end{aligned}$$

Proposition 2.2. *Let (E, \mathcal{E}) be a measurable space, and $\Omega = E^{\mathbb{R}_+}$. Denote its coordinate process by X ,*

$$X_t: \Omega \rightarrow E, \quad \omega \mapsto X_t(\omega) = \omega(t).$$

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

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

Remark 2.1. *The superscripts '0' just denote the fact that we are using the uncompleted 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}_μ , and expectation with respect to this measure is denoted by \mathbb{E}_μ . 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}_{δ_x} .

3. DIFFUSION

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.

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

$$\lim_{t \rightarrow 0} t^{-1} \sup_{x \in \mathbb{R}^d} P_t(x, B_\varepsilon^c(x)) = 0, \quad (3.1)$$

$$\lim_{t \rightarrow 0} t^{-1} \int_{|y-x| \leq \varepsilon} (y-x) P_t(x, dy) = b(x), \quad (3.2)$$

$$\lim_{t \rightarrow 0} (2t)^{-1} \int_{|y-x| \leq \varepsilon} (y-x)_i (y-x)_j P_t(x, dy) = a_{ij}(x) \quad i, j = 1, \dots, d. \quad (3.3)$$

b and a are called the drift coefficient and diffusion coefficient of the diffusion process $(X_t)_{t \geq 0}$, 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

$$\begin{aligned} & \frac{P_t f(x) - f(x)}{t} \\ &= \frac{1}{t} \int_{|y-x| \leq \varepsilon} (f(y) - f(x)) P_t(x, dy) + \frac{1}{t} \int_{|y-x| > \varepsilon} (f(y) - f(x)) P_t(x, dy) =: I_1 + I_2. \end{aligned}$$

By Taylor's expansion theorem, and using (3.2) and (3.3), we have

$$\begin{aligned} I_1 &= \frac{1}{t} \int_{|y-x| \leq \varepsilon} (y-x)_i P_t(x, dy) \partial_i f(x) \\ &\quad + \frac{1}{2t} \int_{|y-x| \leq \varepsilon} (y-x)_i (y-x)_j P_t(x, dy) \partial_{ij} f(x) + o(1) \\ &\rightarrow b(x) \cdot \nabla f(x) + a(x) : \nabla^2 f(x), \quad t \rightarrow 0. \end{aligned}$$

Applying (3.1), we have $I_2 \rightarrow 0$, $t \rightarrow 0$. Therefore,

$$\lim_{t \rightarrow \infty} \frac{P_t f(x) - f(x)}{t} = a(x) : \nabla^2 f(x) + b(x) \cdot \nabla f(x) =: Lf(x), \quad f \in C_b^2(\mathbb{R}^d). \quad (3.4)$$

Assume that

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

Thanks to the Chapman-Kolmogorov equation and (3.4), one can verify that

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

which can be read as

$$\partial_t p(\cdot, y) = L p(\cdot, y), \quad \lim_{t \rightarrow 0} p(t, \cdot, y) = \delta_y. \quad (3.5)$$

Kolmogorov's idea for constructing the diffusion process corresponding to L involves solving the partial differential equation (PDE) (3.5) (in fact he solves the forward equation in his paper) to obtain the density of the process, $p(t, x, \cdot)$.

Fix $y \in \mathbb{R}^d$, put

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

and

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

Then

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

Recall that p satisfies (3.5), therefore,

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

Formally, using Duhamel's formula, we have

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

$$p = \sum_{n=0}^{\infty} \underbrace{p_0 \otimes [(L - L_{a(y)}) p_0]^{\otimes n}}_{=: p_n} = p_0 + p_0 \otimes q, \quad q = \sum_{n=1}^{\infty} q_n. \quad (3.6)$$

and

$$p = p_0 + p_0 \otimes (L - L_{a(y)}) p \quad (3.7)$$

For notion simplicity, we omit the superscript y below.

We attempt to show that the infinite series in (3.6) do convergence (in some sense), and p given by (3.6) satisfying (3.7) is a fundamental solution to (3.5), provided that the coefficients satisfies

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

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

and

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

3.2. Heat Kernel Estimate I. In this subsection, we use the classical Levi's freezing coefficients method to prove that (3.5) 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.

Put

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

Theorem 3.1. *Under Assumption 1, there is a unique continuous function $p(t, x, y) \in \mathbb{D}$ such that*

$$\partial_t p(\cdot, y) = L p(\cdot, y), \quad y \in \mathbb{R}^d.$$

Moreover,

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

(ii)

$$p \geq 0 \quad \text{and} \quad \int_{\mathbb{R}^d} p(t, x, y) dy = 1, \quad t \geq 0, x \in \mathbb{R}^d; \quad (3.8)$$

(iii)

$$p(t+s, x, y) = \int_{\mathbb{R}^d} p(t, x, z) p(s, z, y) dz; \quad (3.9)$$

(iv)

$$t^{-\frac{d}{2}} \exp(-C|x|^2/t) \lesssim p(t, x, y) \lesssim t^{-\frac{d}{2}} \exp(-|x|^2/(Ct)). \quad (3.10)$$

Before giving the full proof for Theorem 3.1, we introduce some notations. 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 $p^{(1)}, p^{(2)}, \dots, p^{(n)} : \mathbb{D} \rightarrow \mathbb{R}$, define

$$\begin{aligned} & \left[p^{(n)} \otimes \dots \otimes p^{(2)} \otimes p^{(1)} \right] (t, x, y) \\ & := \int_{0 < \tau_1 < \dots < \tau_{n-1} < t} \int_{\mathbb{R}^{nd}} p^{(n)}(t - \tau_{n-1}, x, z_{n-1}) \dots \\ & \quad p^{(2)}(\tau_2 - \tau_1, z_2, z_1) p^{(1)}(\tau_1, z_1, y) dz_1 \dots dz_{n-1} d\tau_1 \dots d\tau_{n-1}. \end{aligned}$$

Recall that

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

Lemma 3.2. *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

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

Exercise 3.1. *Prove Lemma 3.2.*

Lemma 3.3. *It holds that*

$$|q| \lesssim \varrho_{\lambda, \alpha-2}; \quad (3.11)$$

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

$$|q(t, x_1, y) - q(t, x_2, y)| \lesssim |x_1 - x_2|^\beta \sum_{i=1,2} \varrho_{\lambda, \alpha-\beta-2}(t, x_i, y). \quad (3.12)$$

Proof. We Claim that

$$|q_n(t, x, y)| \leq \underbrace{\frac{(C_1 \Gamma(\alpha/2))^n}{(\lambda/\pi)^{d(n-1)/2} \Gamma(n\alpha/2)}}_{=: \gamma_n} \varrho_{\lambda, n\alpha-2}(t, x - y), \quad (3.13)$$

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

By Lemma 3.2 and the Hölder regularity of a , we have

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

Assume (3.13) holds. This together with the fact that $q_{n+1} = q_1 \otimes q_n$ yields

$$\begin{aligned}
& |q_{n+1}(t, x, y)| \\
& \leq C_1 \gamma_n \int_0^t (t-\tau)^{\frac{\alpha}{2}-1} \tau^{\frac{n\alpha}{2}-1} d\tau \int_{\mathbb{R}^d} \varrho_\lambda(t-\tau, x-z) \varrho_\lambda(\tau, z-y) dz \\
& = C_1 (\pi \lambda^{-1})^{d/2} \gamma_n \varrho_\lambda(t, x-y) \int_0^t (t-\tau)^{\frac{\alpha}{2}-1} \tau^{\frac{n\alpha}{2}-1} d\tau \\
& = C_1 (\pi \lambda^{-1})^{d/2} \gamma_n \varrho_{\lambda, (n+1)\alpha-2}(t, x-y) B\left(\frac{n\alpha}{2}, \frac{\alpha}{2}\right) \\
& = \gamma_{n+1} \varrho_{\lambda, (n+1)\alpha-2}(t, x-y).
\end{aligned}$$

Therefore, we finish the proof for (3.13), which also implies

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

Next, we verify that

$$|q_1(t, x_1, y) - q_1(t, x_2, y)| \lesssim |x_1 - x_2|^\beta \sum_{i=1,2} \varrho_{\lambda, \alpha-\beta-2}(t, x_i - y), \quad (t, x, y) \in \mathbb{D}. \quad (3.14)$$

If $|x_1 - x_2| > \sqrt{t}$, then it is a consequence of (3.13). When $|x_1 - x_2| \leq \sqrt{t}$, we have

$$\begin{aligned}
& |q_1(t, x_1, y) - q_1(t, x_2, y)| \\
& \leq |a(x_1) - a(x_2)| \cdot |\nabla_x^2 p_0(t, x_1, y)| + |a(x_2) - a(y)| \cdot |\nabla_x^2 p_0(t, x_1, y) - \nabla_x^2 p_0(t, x_2, y)| \\
& \lesssim |x_1 - x_2|^\alpha \varrho_{\lambda_2, -2}(t, x_1 - y) + |x_2 - y|^\alpha |x_1 - x_2| \varrho_{\lambda_3, -3}(t, x_2 - y - \theta(x_1 - x_2)) \\
& \lesssim |x_1 - x_2|^\beta \varrho_{\lambda, \alpha-\beta-2}(t, x_1 - y) + |x_2 - y|^\alpha |x_1 - x_2|^\beta \varrho_{\lambda, -\beta-2}(t, x_2 - y) \\
& \lesssim |x_1 - x_2|^\beta \sum_{i=1,2} \varrho_{\lambda, \alpha-\beta-2}(t, x_i - y).
\end{aligned}$$

Therefore, (3.14) holds for any $(t, x, y) \in \mathbb{D}$. Noting that $q = q_1 + q_1 \otimes q$, we have

$$\begin{aligned}
& |q(t, x_1, y) - q(t, x_2, y)| \\
& \lesssim |x_1 - x_2|^\beta \sum_{i=1,2} \varrho_{\lambda, \alpha-\beta-2}(t, x_i - y) \\
& \quad + |x_1 - x_2|^\beta \int_0^t (t-\tau)^{\frac{\alpha-\beta}{2}-1} \tau^{\frac{\alpha}{2}-1} d\tau \sum_{i=1,2} \varrho_\lambda(t, x_i - y) \\
& \lesssim |x_1 - x_2|^\beta \sum_{i=1,2} \varrho_{\lambda, \alpha-\beta-2}(t, x_i - y).
\end{aligned}$$

□

The above lemma implies that the infinite series in (3.6) do convergence, and p given by (3.6) satisfying (3.7).

Lemma 3.4. *There is a constant $\lambda > 0$ such that*

$$|p| \lesssim \varrho_\lambda, \quad |\partial_t p| + |\nabla_x^2 p| \lesssim \varrho_{\lambda, -2} \quad (3.15)$$

and

$$[L - L_{a(y)}]p(t, x, y) = q(t, x, y). \quad (3.16)$$

Proof. Recalling that $p = p_0 + p_0 \otimes q$, by Lemma 3.2, we only need to prove

$$|\nabla_x^2 p_0 \otimes q| \lesssim \varrho_{\lambda, \alpha-2}. \quad (3.17)$$

Note

$$\begin{aligned} \nabla_x^2(p_0 \otimes q)(t, x, y) &= \int_0^t \int_{\mathbb{R}^d} \nabla_x^2 p_0(t-\tau, x, z) q(\tau, z, y) dz d\tau \\ &= \int_0^{\frac{t}{2}} \cdots + \int_{\frac{t}{2}}^t \cdots =: I_1 + I_2. \end{aligned}$$

Thanks to Lemma 3.2 and Lemma 3.3,

$$\begin{aligned} |I_1| &\lesssim \int_0^{\frac{t}{2}} \varrho_{\lambda, -2}(t-\tau, x-z) \varrho_{\lambda, \alpha-2}(\tau, z-y) dz d\tau \\ &\lesssim \varrho_{\lambda}(t, x-y) \int_0^{\frac{t}{2}} (t-\tau)^{-1} \tau^{\frac{\alpha}{2}-1} d\tau \lesssim \varrho_{\lambda, \alpha-2}(t, x-y) \end{aligned}$$

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

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

In view of (3.12), we get

$$\begin{aligned} |I_2| &= \left| \int_{\frac{t}{2}}^t d\tau \int_{\mathbb{R}^d} \nabla_x^2 p_0(t-\tau, x, z) [q(\tau, z, y) - q(\tau, x, y)] dz \right| \\ &\lesssim \int_{\frac{t}{2}}^t (t-\tau)^{\frac{\beta}{2}-1} \tau^{\frac{\alpha-\beta}{2}-1} d\tau \int_{\mathbb{R}^d} [\varrho_{\lambda}(\tau, y-z) + \varrho_{\lambda}(\tau, x-y)] \varrho_{\lambda}(t-\tau, x-z) dz \\ &\lesssim \varrho_{\lambda, \alpha-2}(t, x-y). \end{aligned}$$

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

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 3.1. (i). It is easy to verify that

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

convergence to f uniformly when $f \in C_0(\mathbb{R}^d)$ as $t \rightarrow 0$. In the light of (3.11), we have

$$|p_0 \otimes q| \lesssim \varrho_{\lambda, \alpha}, \quad (3.18)$$

which yields that

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

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 3.4, we only need to prove the lower bound estimate. There exists constant $T > 0$ such that

$$p(t, x, y) \geq p_0(t, x, y) - |p_0 \otimes q|(t, x, y) \gtrsim t^{-\frac{d}{2}} - t^{-\frac{d+\alpha}{2}} \gtrsim t^{-\frac{d}{2}}, \quad |x-y| < \sqrt{t}, \quad t \in [0, T]. \quad (3.19)$$

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

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

Noting that for all $z_i \in B_i$,

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

by the on-diagonal estimate (3.19), we have

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

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

$$\begin{aligned} p(t, x, y) &\geq \int_{B_{n-1}} \cdots \int_{B_1} p(t_1, x, z_1) \cdots p(t_n - t_{n-1}, z_{n-1}, y) dz_1 \cdots dz_{n-1} \\ &\geq \left[c_2(t/n)^{-\frac{d}{2}} \right]^n \left[\omega_d \left(\sqrt{t/(64n)} \right)^d \right]^{n-1} \geq t^{-\frac{d}{2}} c_3^n n^{\frac{d}{2}} \\ &\geq t^{-\frac{d}{2}} c_3^{4|x-y|^2/t} (|x-y|^2/t)^{\frac{d}{2}} \gtrsim \varrho_\lambda(t, x, y). \end{aligned}$$

□

REFERENCES

- [Dur19] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.
 [Yan21] Jiaan Yan. *Lecture note on Measure Theory*. Science Press, China, third edition, 2021.

APPENDIX A. PROBABILISTIC TERMINOLOGY

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and (E, \mathcal{E}) be a measurable space. $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ 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 $A \in \mathcal{G}$, $\mathbf{E}(X; A) = \mathbf{E}(Y; A)$.

$Q_{\mathcal{G}} : \Omega \times \mathcal{E} \rightarrow [0, 1]$ 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 A.1. (i) If $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$, then

$$\mathbf{E}[(X|\mathcal{G}_2)|\mathcal{G}_1] = \mathbf{E}(X|\mathcal{G}_1) \quad (\text{A.1})$$

(ii) Assume that $X \in \mathcal{F}$ and $Y \in \mathcal{G} \subseteq \mathcal{F}$, then

$$\mathbf{E}(XY|\mathcal{G}) = \mathbf{E}(X|\mathcal{G})Y. \quad (\text{A.2})$$

(iii) (Jesen's inequality) If φ is a convex function, then

$$\mathbf{E}(\varphi(X)|\mathcal{G}) \leq \varphi(\mathbf{E}(X|\mathcal{G})). \quad (\text{A.3})$$

Proposition A.2. Let $Q_{\mathcal{G}}$ be a RCD for X given \mathcal{G} . If $f : E \rightarrow \mathbb{R}$ satisfying $\mathbf{E}|f(X)| < \infty$, then

$$\mathbf{E}(f(X)|\mathcal{G})(\omega) = \int_E f(x)Q_{\mathcal{G}}(\omega, dx) \quad \text{a.s.}$$

Theorem A.3. RCD exists if E is a standard measure space and $\mathcal{E} = \mathcal{B}(E)$.

Proposition A.4. Assume $X \geq 0$, $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that $f \in C^1(\mathbb{R}_+)$ and $f(0) = 0$. Then

$$\mathbf{E}f(X) = \int_0^\infty f'(t)\mathbf{P}(X > t)dt. \quad (\text{A.4})$$

Exercise A.1. If $X \geq 0$, $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that $f \in C^1(\mathbb{R}_+)$ and $f(\infty) = 0$. Then

$$\mathbf{E}f(X) = - \int_0^\infty f'(t)\mathbf{P}(X \leq t)dt. \quad (\text{A.5})$$

APPENDIX B. 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 B.1 (Weak Maximum Principle). If $c(x, t) \leq 0$ and u is bounded, then

$$\sup_{\mathbb{R}_+ \times \mathbb{R}^d} u = \sup_{\{0\} \times \mathbb{R}^d} u.$$

Proposition B.2 (Strong Maximum Principle). If u achieves its maximum (or minimum) at an interior point $(t_0, x_0) \in \mathbb{R}_+ \times \mathbb{R}^d$, then u is constant in $[0, t_0] \times \mathbb{R}^d$.