Diffusion Processes

Jyunyi Liao

Contents

1 Markov Processes

In this section, we study the Markov processes, which cover a wide range of stochastic processes. Roughly speaking, a Markov process is a stochastic process such that, given the current information, the future states are conditionally independent of the past states.

Our discussion is based on a filtered probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The state space of stochastic processes is a measurable space $(E, \mathscr{E})$. Usually, E is a topological space, and $\mathscr{E}$ is the Borel σ-algebra on E.

1.1 Transition Functions

Definition 1.1 (Transition probability)**.** Let $(E, \mathscr{E})$ be a measurable space. A *transition probability on* E is a mapping $P : E \times \mathscr{E} \to [0, 1]$ such that

(i) For every $x \in E$, the mapping $\mathscr{E} \ni A \mapsto P(x, A)$ is a probability measure on $(E, \mathscr{E})$; and

(ii) For every $A \in \mathscr{E}$, the mapping $E \ni x \mapsto P(x, A)$ is $\mathscr{E}$-measurable.

Remark. If $f \in B(E)$, i.e. f is a bounded measurable function on E, we define

$$(Pf)(x) = \int_E P(x, dy)f(y).$$

Then Pf is also a bounded measurable function on E. For this reason, P is also viewed as a linear operator on the Banach space $B(E)$. Furthermore, P satisfies the following properties:

(i) P is positive, i.e. $Pf \geq 0$ for each $f \geq 0$;

(ii) P is a contraction, i.e. $\|Pf\|_\infty \leq \|f\|_\infty$, where $\|\cdot\|_\infty$ is the uniform norm on $B(E)$.

Furthermore, if both P and Q are transition probabilities, we define the *multiplication* operation:

$$PQ(x, A) = \int_E P(x, dy)Q(y, A).$$

Then PQ is also a transition probability on E, and can be viewed as the composition of operators P and Q.

Definition 1.2 (Transition function)**.** A *transition function* on E is a family $\{P_{s,t} : 0 \leq s < t\}$ of transition probabilites such that for all $0 \leq r < s < t$,

$$P_{r,t}(x, A) = \int_E P_{r,s}(x, dy)P_{s,t}(y, A) \tag{1.1}$$

for each $x \in E$ and each $A \in \mathscr{E}$. The relation (1.1) is called the *Chapman-Kolmogorov equation*. The transition function is said to be *homogeneous* if $P_{s,t}$ depends on s and t only through the difference $t - s$. In that case, we write P_t for $P_{0,t}$, and the Chapman-Kolmogorov equation writes

$$P_{t+s}(x, A) = \int_E P_s(x, dy)P_t(y, A).$$

In other words, the family $\{P_t, t \geq 0\}$ forms a semigroup.

Definition 1.3 (Markov process)**.** Let $(\Omega, \mathscr{F}, (\mathscr{F})_{t\geq 0}, \mathbb{P})$ be a filtered probability space. An adapted stochastic process $(X_t)_{t\geq 0}$ is said to be a *Markov process* with respect to $(\mathscr{F}_t)_{t\geq 0}$ with transition function $(P_{s,t})$, if for any nonnegative measurable function $f : E \to \mathbb{R}_+$ and any pair (s, t) with $s < t$,

$$\mathbb{E}\left[f(X_t)|\mathscr{F}_s\right] = P_{s,t}f(X_s) \quad \mathbb{P}\text{-a.s.}. \tag{1.2}$$

Remark. If the filtration is not specified, we usually use the canonical filtration

$$\mathscr{F}_t^X = \sigma(X_s, 0 \leq s \leq t).$$

Proposition 1.4. *A stochastic process $(X_t)_{t\geq 0}$ is a Markov process with transition function $P_{s,t}$ and initial measure μ, if and only if for any $0=t_0<t_1<\cdots<t_n$ and $A_0, A_1, \cdots, A_n \in \mathscr{E}$,*

$$\mathbb{P}\left(X_{t_0}\in A_0, X_{t_1}\in A_1,\cdots, X_{t_n}\in A_n\right)=\int_{A_0}\gamma(dx_0)\int_{A_1}P_{t_0,t_1}(x_0,dx_1)\cdots\int_{A_n}P_{t_{n-1},t_n}(x_{n-1},dx_n). \tag{1.3}$$

Proof. If $(X_t)_{t\geq 0}$ is a Markov process,

$$\begin{aligned}
&\mathbb{P}\left(X_{t_0}\in A_0, X_{t_1}\in A_1,\cdots, X_{t_n}\in A_n\right)=\mathbb{E}\left[\mathbb{1}_{A_0}(X_{t_0})\mathbb{1}_{A_1}(X_{t_1})\cdots\mathbb{1}_{A_n}(X_{t_n})\right]\\
&\quad=\mathbb{E}\left[\mathbb{1}_{A_0}(X_{t_0})\mathbb{1}_{A_1}(X_{t_1})\cdots\mathbb{1}_{A_{n-1}}(X_{t_{n-1}})\mathbb{E}\left[\mathbb{1}_{A_n}(X_n)|\mathscr{F}^X_{t_{n-1}}\right]\right]\\
&\quad=\mathbb{E}\left[\mathbb{1}_{A_0}(X_{t_0})\mathbb{1}_{A_1}(X_{t_1})\cdots\mathbb{1}_{A_{n-1}}(X_{t_{n-1}})P_{t_{n-1},t_n}\mathbb{1}_{A_n}(X_{t_{n-1}})\right]\\
&\quad=\mathbb{E}\left[\mathbb{1}_{A_0}(X_{t_0})\mathbb{1}_{A_1}(X_{t_1})\cdots\mathbb{1}_{A_{n-1}}(X_{t_{n-1}})\int_{A_n}P_{t_{n-1},t_n}(X_{t_{n-1}},x_n)\,dx_n\right].
\end{aligned}$$

Repeating this procedure, we have

$$\begin{aligned}
\mathbb{P}\left(X_{t_0}\in A_0, X_{t_1}\in A_1,\cdots, X_{t_n}\in A_n\right)&=\mathbb{E}\left[\mathbb{1}_{A_0}(X_{t_0})\int_{A_1}P_{t_0,t_1}(X_{t_0},dx_1)\cdots\int_{A_n}P_{t_{n-1},t_n}(x_{n-1},dx_n)\right]\\
&=\int_{A_0}\gamma(dx_0)\int_{A_1}P_{t_0,t_1}(x_0,dx_1)\cdots\int_{A_n}P_{t_{n-1},t_n}(x_{n-1},dx_n).
\end{aligned}$$

Now we assume that $(X_t)_{t\geq 0}$ satisfies (1.3). To prove (1.2), we must show that for all $A\in\mathscr{F}^X_s$,

$$\mathbb{E}\left[f(X_t)\mathbb{1}_A\right]=\mathbb{E}\left[P_{s,t}f(X_s)\mathbb{1}_A\right]. \tag{1.4}$$

Since the sets A satisfying form a λ-system, by π-λ theorem, it suffices to show (1.4) for all cylinder sets

$$A=\left\{X_{t_0}\in A_0, X_{t_1}\in A_1,\cdots,X_{t_n}\in A_n\right\},$$

where $0=t_0<t_1<\cdots<t_n=s$ and $A_0,A_1,\cdots,A_n\in\mathscr{E}$. If $f=\mathbb{1}_B$, where $B\in\mathscr{E}$, by (1.3), we have

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}_B(X_t)\mathbb{1}_A\right]&=\mathbb{P}\left(X_{t_0}\in A_0, X_{t_1}\in A_1,\cdots,X_{t_n}\in A_n, X_t\in B\right)\\
&=\int_{A_0}\gamma(dx_0)\int_{A_1}P_{t_0,t_1}(x_0,dx_1)\cdots\int_{A_n}P_{t_{n-1},t_n}(x_{n-1},dx_n)\int_B P_{t_n,t}(x_n,dy)\\
&=\int_{A_0}\gamma(dx_0)\int_{A_1}P_{t_0,t_1}(x_0,dx_1)\cdots\int_{A_n}P_{t_{n-1},t_n}(x_{n-1},dx_n)P_{s,t}\mathbb{1}_B(x_n)\\
&=\mathbb{E}\left[\mathbb{1}_{A_0}(X_{t_0})\int_{A_1}P_{t_0,t_1}(X_0,dx_1)\cdots\int_{A_n}P_{t_{n-1},t_n}(x_{n-1},dx_n)P_{s,t}\mathbb{1}_B(x_n)\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{A_0}(X_{t_0})\mathbb{1}_{A_1}(X_{t_1})\int_{A_2}P_{t_1,t_2}(X_1,dx_2)\cdots\int_{A_n}P_{t_{n-1},t_n}(x_{n-1},dx_n)P_{s,t}\mathbb{1}_B(x_n)|\mathscr{F}^X_0\right]\right]\\
&=\cdots=\mathbb{E}\left[\mathbb{E}\left[\cdots\mathbb{E}\left[P_{s,t}\mathbb{1}_B(X_s)\mathbb{1}_A|\mathscr{F}^X_{t_{n-1}}\right]\cdots|\mathscr{F}^X_0\right]\right]=\mathbb{E}\left[P_{s,t}(X_s)\mathbb{1}_A\right].
\end{aligned}$$

Thus (1.4) holds for all indicators, and hence for all simple functions. By simple function approximation and monotone convergence theorem, (1.4) holds for all nonnegative meaesurable functions $f:E\to\mathbb{R}_+$. ☐

Remark. We can construct a Markov process with transition function $P_{s,t}$ and initial measure γ with this proposition. We define $(\Omega,\mathscr{F})=(E^{\mathbb{R}_+},\mathscr{E}^{\mathbb{R}_+})$, and let $X_t(\omega)=\omega_t$ be the projction map. Then (1.3) defines a compatible family of finite marginal distributions on $(E^{\mathbb{R}_+},\mathscr{E}^{\mathbb{R}_+})$. If all marginal distributions $(\mathbb{P}(\omega_t), t\geq 0)$ are inner-regular (which is the case, for example, if E is locally compact, Hasudorff and second countable), by the Kolmogorov extension theorem, the transition function $P_{s,t}$ extends to a unique probability measure P_γ on $(E^{\mathbb{R}_+},\mathscr{E}^{\mathbb{R}_+})$. The canonical process $(X_t)_{t\geq 0}$ is a Markov process under this probability measure.

Notation. For every probability measure γ on $(E,\mathscr{E})$, we denote by P_γ the law of the Markov process with transition function $P_{s,t}$ and initial measure γ, which is a probability measure on $(E^{\mathbb{R}_+},\mathscr{E}^{\mathbb{R}_+})$. For any nonnegrative measurable function $\Phi : E^{\mathbb{R}_+} \to \mathbb{R}_+$, we write

$$\mathbb{E}_\gamma[\Phi] = \int \Phi \, dP_\gamma.$$

For any fixed $x \in E$, we we write $P_x = P_{\delta_x}$, and

$$\mathbb{E}_x[\Phi] = \int \Phi \, dP_x.$$

We have the following useful conclusion.

Proposition 1.5. *For any nonnegrative or bounded measurable function $\Phi : E^{\mathbb{R}_+} \to \mathbb{R}_+$, the map $x \mapsto \mathbb{E}_x[\Phi]$ is $\mathscr{E}$-measurable, and for any probability measure γ on $(E,\mathscr{E})$,*

$$\mathbb{E}_\gamma[\Phi] = \int_E \gamma(dx)\mathbb{E}_x[\Phi].$$

Proof. For simplicity, we fix the initial measure γ. Let $\mathscr{A}$ be the class of cylinder sets

$$A = \{(x_t)_{t\geq 0} : x_{t_0} \in A_0, x_{t_1} \in A_1, \cdots, x_{t_n} \in A_n\} \in \mathscr{E}^{\mathbb{R}_+},$$

where $0 = t_0 < t_1 < \cdots < t_n$, and $A_0, A_1, \cdots, A_n \in \mathscr{E}$. Then $\mathscr{A}$ is a π-system, and

$$\begin{aligned}\mathbb{E}_x[\mathbb{1}_A] &= \int_{A_0} \delta_x(dx_0) \int_{A_1} P_{t_0,t_1}(x_0, dx_1) \cdots \int_{A_n} P_{t_{n-1},t_n}(x_{n-1}, dx_n) \\ &= \mathbb{1}_{\{x\in A_0\}} \int_{A_1} P_{t_0,t_1}(x, dx_1) \cdots \int_{A_n} P_{t_{n-1},t_n}(x_{n-1}, dx_n),\end{aligned}$$

which is a $\mathscr{E}$-measurable function of $x \in E$. Also,

$$\begin{aligned}\mathbb{E}_\gamma[\mathbb{1}_A] &= \int_{A_0} \gamma(dx_0) \int_{A_1} P_{t_0,t_1}(x_0, dx_1) \cdots \int_{A_n} P_{t_{n-1},t_n}(x_{n-1}, dx_n) \\ &= \int_E \gamma(dx_0)\mathbb{1}_{\{x_0\in A_0\}} \int_{A_1} P_{t_0,t_1}(x_0, dx_1) \cdots \int_{A_n} P_{t_{n-1},t_n}(x_{n-1}, dx_n) = \int_E \gamma(dx_0)\mathbb{E}_{x_0}[\Phi].\end{aligned}$$

Hence the proposition is true for all $\Phi = \mathbb{1}_A$, where $A \in \mathscr{A}$ is a cylinder set. Since the class of functions Φ satisfying this proposition is closed under addition, scalar multiplication and increasing limits (by monotone convergence theorem), we conclude the proof by applying the monotone convergence theorem for funtions. $\square$

2 Diffusion Process

In this section, we study the stochastic differential equation (SDE)

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

where $\sigma = (\sigma_{ij})_{i\in[p],j\in[q]} : \mathbb{R}_+ \times \mathbb{R}^p \to \mathbb{R}^{p\times q}$ and $b = (b_i)_{i\in[p]} : \mathbb{R}_+ \times \mathbb{R}^p \to \mathbb{R}^p$ be locally bounded measurable functions. The function $b : \mathbb{R}^p \to \mathbb{R}^p$ is called the *drift coefficient*, and $\sigma : \mathbb{R}^p \to \mathbb{R}^{p\times q}$ is called the *diffusion coefficient*. The soltions of (2.1) with continuous sample paths are called *diffusion processes*.

For notation simplicity, we also write $E(\sigma, b)$ for the SDE (2.1). For each $x \in \mathbb{R}^p$, we write $E_x(\sigma, b)$ for the SDE (2.1) together with the initial value $X_0 = x$. We use $|\cdot|$ to denote the Euclidean norm of vectors and the Frobenius norm of matrices. Throughout this section, we assume that the coefficients of SDE (2.1) are Lipschitz continuous, i.e. there exists a constant $K > 0$ such that for all $x, y \in \mathbb{R}^p$,

$$|\sigma(t, x) - \sigma(s, y)| \le K|t - s| + K|x - y|, \qquad |b(t, x) - b(s, y)| \le K|t - s| + K|x - y|.$$

2.1 Solutions of SDE

In this subsection, we discuss the solvability of stochastic differential equations.

Definition 2.1. A *solution of the stochastic equation* (2.1) consists of:

- A filtered probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a complete filtration $(\mathscr{F})_{t\ge 0}$;
- A q-dimensional $(\mathscr{F}_t)$-Brownian motion $B = (B^1, \cdots, B^q)$ starting from 0;
- An $(\mathscr{F}_t)$-adapted and sample-continuous process $X = (X^1, \cdots, X^p)$ taking values in $\mathbb{R}^p$, such that

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

Definition 2.2. For the stochastic differential equation $E(\sigma, b)$, we say that there is

- *weak existence*, if for every $x \in \mathbb{R}^p$, there exists a solution of $E_x(\sigma, b)$;
- *weak existence* and *weak uniqueness*, if in addition, for every $x \in \mathbb{R}^p$, all solutions of $E_x(\sigma, b)$ have the same law;
- *pathwise uniqueness*, if, whenever the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\ge 0}, \mathbb{P})$ and the $(\mathscr{F}_t)$-Brownian motion B are fixed, two solutions X and Y such that $X_0 = Y_0$ *a.s.* are indistinguishable.

Furthermore, we say that a solution X of $E(\sigma, b)$ is a *strong solution* if X is adapted with respect to the completed canonical filtration of B.

For the completeness of our discussion, we state (without proof) two theorems concerning the existence and uniqueness of the solution of SDEs with Lipschitz continuous coefficients.

Theorem 2.3. *Let σ and b be Lipschitz continuous. Then pathwise uniqueness holds for the SDE $E(\sigma, b)$. Furthermore, for every complete filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\ge 0}, \mathbb{P})$, every $(\mathscr{F}_t)$-Brownian motion B and every $x \in \mathbb{R}^p$, there exists a unique strong solution of $E_x(\sigma, b)$.*

Theorem 2.4. *Equip both $C(\mathbb{R}_+, \mathbb{R}^p)$ and $C(\mathbb{R}_+, \mathbb{R}^q)$ with the Borel σ-algebra of the compact convergence topology, and complete the σ-algebra on $C(\mathbb{R}_+, \mathbb{R}^p)$ by W-negligible sets, where W is the Wiener measure. Then for all $x \in \mathbb{R}^p$, there exists a measurable mapping $F_x : C(\mathbb{R}_+, \mathbb{R}^q) \to C(\mathbb{R}_+, \mathbb{R}^p)$ satisfying*

(i) for every $t \ge 0$, the mapping $\mathbf{w} \mapsto F_x(\mathbf{w})_t$ coincides W-a.s. a measurable function of $(\mathbf{w}(r))_{0\le r\le t}$;

(ii) for every $\mathbf{w} \in C(\mathbb{R}_+, \mathbb{R}^q)$, the mapping $x \mapsto F_x(\mathbf{w})$ is continuous;

(iii) for every $t \ge 0$, and for every choice of the complete filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\ge 0}, \mathbb{P})$ and of the $(\mathscr{F}_t)$-Brownian motion B with $B_0 = 0$, the process $X_t = F_x(B)_t$ is the unique solution of $E(\sigma, b)$. Furthermore, if Z is an $\mathscr{F}_0$-measurable $\mathbb{R}^p$-valued random variable, the process $F_Z(B)_t$ is the unique solution of $E(\sigma, b)$ with $X_0 = Z$.

Theorem 2.5 (Markov property of diffusion processes). *Assume that $X = (X_t)_{t\geq 0}$ is a solution of SDE* (2.1) *on a complete filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$. Then $(X_t)_{t\geq 0}$ is a Markov process with respect to the filtration $(\mathscr{F}_t)_{t\geq 0}$. For each $s \geq 0$ and $t > 0$, the transition function $P_{s,s+t}$ defined by*

$$P_{s,s+t}f(x) = \mathbb{E}\left[f(Y_t)\right],$$

where $Y = (Y_t)_{t\geq 0}$ is an arbitrary solution of

$$\begin{cases} dY_t = \sigma(s+t, Y_t)\, dB_t + b(s+t, Y_t)\, dt, \\ Y_0 = x. \end{cases}$$

Let F_x be the mapping given by Theorem 2.4 corresponding to the above SDE. Then we also write

$$P_{s,s+t}f(x) = \int f(F_x(\mathbf{w})_t)\, W(d\mathbf{w}).$$

Proof. We first prove that, for any $f \in B(\mathbb{R}^p)$ and any $s \geq 0, t > 0$, we have

$$\mathbb{E}\left[f(X_{s+t})|\mathscr{F}_s\right] = P_{s,s+t}f(X_s),$$

To deal the time shift s, we define filtration $(\mathscr{F}'_t)_{t\geq 0}$ and processes $(X'_t)_{t\geq 0}$, $(B'_t)_{t\geq 0}$ as follows:

$$\mathscr{F}'_t = \mathscr{F}_{s+t}, \quad X'_t = X_{s+t}, \quad B'_t = B_{t+s} - B_s.$$

Then $(\mathscr{F}'_t)_{t\geq 0}$ is a complete filtration, X' is adapted to $(\mathscr{F}'_t)_{t\geq 0}$, and B' is a q-dimensional $(\mathscr{F}'_t)$-Brownian motion. Furthermore,

$$X'_t = X_{s+t} = X_s + \int_s^{s+t} \sigma(r, X_r)\, dB_r + \int_s^{s+t} b(r, X_r)\, dr = X_s + \int_0^t \sigma(s+r, X'_r)\, dB'_r + \int_0^t b(s+r, X'_r)\, dr.$$

Consequently, X' solves $E(\sigma, b)$ on the space $(\Omega, \mathscr{F}, (\mathscr{F}'_t)_{t\geq 0}, \mathbb{P})$ and with Brownian motion B', with $X'_0 = X_s$. By Theorem 2.4 (iii), we have $X' = F_{X_s}(B')$ *a.s.*, which implies

$$\mathbb{E}\left[f(X_{s+t})|\mathscr{F}_s\right] = \mathbb{E}\left[f(X'_t)|\mathscr{F}_s\right] = \mathbb{E}\left[f(F_{X_s}(B')_t)|\mathscr{F}_s\right] = \int f(F_{X_s}(\mathbf{w})_t)\, W(d\mathbf{w}) = P_{s,s+t}f(X_s),$$

where the third equality follows from the independence of B' and $\mathscr{F}_s$.

Now it remains to verify that $P_{s,s+t}$ is a transition function. Clearly, $x \mapsto P_{s,s+t}f(x)$ is a continuous map, hence is measurable. Finally, note that

$$P_{s,s+t+v}f(x) = \mathbb{E}\left[f(Y_{t+v})\right] = \mathbb{E}\left[\mathbb{E}\left[f(Y_{t+v})|\mathscr{F}_{s+t}\right]\right] = \mathbb{E}\left[P_{s+t,s+t+v}f(Y_t)\right] = \int P_{s,s+t}(x, dy)P_{s+t,s+t+v}f(y),$$

which is the Chapman-Kolmogorov equation. This completes the proof. $\square$

Remark. According to Theorem 2.4 (ii), if we take $f \in C_b(\mathbb{R}^p)$, by dominated convergence theorem,

$$P_{s,s+t}f(x) = \int f(F_x(\mathbf{w})_t)\, W(d\mathbf{w})$$

is in $C_b(\mathbb{R}^p)$ as well.

In our later discussion, we use $P_{s,t}$ to denote the transition function of the diffusion processes $(X_t)_{t\geq 0}$ defined by SDE (2.1).

2.2 Transition Functions

In this subsection, we discuss the properties of the transition function of a diffusion process.

Theorem 2.6. *Let $P_{s,t}$ be the transition function of the diffusion process defined by*

$$dX_t = \sigma(t, X_t)\, dB_t + b(t, X_t)\, dt.$$

For every $\varphi \in C_b^2(\mathbb{R}^p)$,

$$d\varphi(X_t) = \sigma^*(t, X_t)\nabla\varphi(X_t)\, dB_t + \left(\frac{1}{2}\sigma\sigma^*(t, X_t) \cdot \nabla^2\varphi(X_t) + b(t, X_t) \cdot \nabla\varphi(X_t)\right) dt. \tag{2.2}$$

where σ^* is the matrix transpose of σ. Furthermore, for every $x \in \mathbb{R}^p$, we have

$$\lim_{h\downarrow 0} \frac{P_{t,t+h}\varphi(x) - \varphi(x)}{h} = \frac{1}{2}\sigma\sigma^*(t, x) \cdot \nabla^2\varphi(x) + b(t, x) \cdot \nabla\varphi(x), \tag{2.3}$$

and

$$\lim_{h\downarrow 0} \frac{P_{t-h,t}\varphi(x) - \varphi(x)}{h} = \frac{1}{2}\sigma\sigma^*(t, x) \cdot \nabla^2\varphi(x) + b(t, x) \cdot \nabla\varphi(x). \tag{2.4}$$

We will give a formal proof of this Theorem later. For convenience of our analysis, we introduce a useful lemma in the study of differential equations.

Lemma 2.7 (Gronwall's lemma)**.** *Let $T > 0$, and $g : [0, T] \to \mathbb{R}_+$ is a bounded measurable function. If there exist two constants $a \geq 0$ and $b \geq 0$ such that*

$$g(t) \leq a + b\int_0^t g(s)\, ds, \quad \forall t \in [0, T],$$

then we have $g(t) \leq ae^{bt}$ for all $t \in [0, T]$.

Proof. A simple recursion on g gives

$$\begin{aligned}
g(t) &\leq a + b\int_0^t \left(a + b\int_0^{s_1} g(s_2)\, ds\right) ds_1 \\
&= a + a(bt) + b^2\int_0^t ds_1 \int_0^{s_1} ds_2\, g(s_2) \\
&\leq a + a(bt) + b^2\int_0^t ds_1 \int_0^{s_1} ds_2 \left(a + b\int_0^{s_2} g(s_3)\, ds_3\right) \\
&= a + a(bt) + a\frac{(bt)^2}{2} + b^3\int_0^t ds_1 \int_0^{s_1} ds_2 \int_0^{s_2} ds_3 g(s_3) \leq \cdots \\
&\leq a + a(bt) + a\frac{(bt)^2}{2} + \cdots + a\frac{(bt)^n}{n!} + b^{n+1}\int_0^t ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{n+1}} ds_{n+1}\, g(s_{n+1}).
\end{aligned}$$

Since g is bounded, we let $0 \leq g(t) \leq M$ for all $t \in [0, T]$. Then

$$g(t) \leq a\sum_{k=0}^n \frac{(bt)^n}{n!} + \frac{M(bt)^{n+1}}{(n+1)!}.$$

The desired result follows by letting $n \to \infty$. □

Remark. A usseful case of this lemma is that when $a = 0$, we have $g(t) = 0$.

We then give an estimate for the second moment of a diffusion process.

Lemma 2.8. *Fix $x \in \mathbb{R}^p$, and let $(X_t^x)_{t\geq 0}$ be a solution of the SDE $E_x(\sigma, b)$. Then there exists a constant $C_x > 0$ depending only on x, such that for all $t \geq 0$,*

$$\mathbb{E}\left[|X_t^x - x|^2\right] \leq C_x e^{C_x(t+t^2)}(t + t^2 + t^3 + t^4).$$

Proof. By the triangle inequality and Lipschitz continuity, for all $t \geq 0$, we have

$$\begin{aligned}|\sigma(t, X_t^x)|^2 &\leq (|\sigma(0,x)| + |\sigma(0,x) - \sigma(t,x)| + |\sigma(t, X_t^x) - \sigma(t,x)|)^2 \\ &\leq 3|\sigma(0,x)|^2 + 3K^2t^2 + 3K^2|X_t^x - x|^2.\end{aligned} \tag{2.5}$$

A similar formula also holds for $|b(t, X_t^x)|^2$. We define a stopping time $\tau = \inf\{t \geq 0 : |X_t^x - x| > M\}$ for some $M > 0$, and fix $T > 0$. Then the function $t \mapsto \mathbb{E}\left[|X_{t\wedge\tau}^x - x|^2\right]$ is bounded on $[0, T]$. For any $t \in [0, T]$, we have

$$\begin{aligned}\mathbb{E}\left[|X_{t\wedge\tau}^x - x|^2\right] &\leq 2\mathbb{E}\left[\left(\int_0^{t\wedge\tau} \sigma(s, X_s^x)\, dB_s\right)^2\right] + 2\mathbb{E}\left[\left(\int_0^{t\wedge\tau} b(s, X_s^x)\, ds\right)^2\right] \\ &\leq 2\mathbb{E}\left[\int_0^{t\wedge\tau} |\sigma(s, X_s^x)|^2\, ds\right] + 2\mathbb{E}\left[T\int_0^{t\wedge\tau} |b(s, X_s^x)|^2\, ds\right] \\ &\leq 6T(|\sigma(0,x)|^2 + T|b(0,x)|^2) + 2K^2T^3(1+T) + 6K^2(1+T)\int_0^{t\wedge\tau} \mathbb{E}\left[|X_s^x - x|^2\right] ds \\ &\leq 6T(|\sigma(0,x)|^2 + T|b(0,x)|^2) + 2K^2T^3(1+T) + 6K^2(1+T)\int_0^{t} \mathbb{E}\left[|X_{s\wedge\tau}^x - x|^2\right] ds.\end{aligned}$$

where we use (2.5) in the third inequality. By Gronwall's lemma, we have

$$\mathbb{E}\left[|X_{t\wedge\tau}^x - x|^2\right] \leq \left(6T|\sigma(0,x)|^2 + 6T^2|b(0,x)|^2 + 2K^2T^3 + 2K^2T^4\right) e^{6K^2(1+T)T}, \quad \forall t \in [0, T].$$

We let $M \to \infty$, and apply the monotone convergence theorem to obtain

$$\mathbb{E}\left[|X_T^x - x|^2\right] \leq \left(6T|\sigma(0,x)|^2 + 6T^2|b(0,x)|^2 + 2K^2T^3 + 2K^2T^4\right) e^{4K^2(T+T^2)}.$$

Setting $C_x = 6\max\left\{|\sigma(0,x)|^2, |b(0,x)|^2, K^2\right\}$ concludes the proof. $\square$

Now we prove the main result.

Proof of Theorem 2.6. (i) The covariation of the process $(X_t)_{t\geq 0}$ is

$$\langle X^i, X^j\rangle_t = \sum_{k=1}^q \int_0^t \sigma_{ik}(s, X_s)\sigma_{jk}(s, X_s)\, ds = \int_0^t (\sigma\sigma^*)_{ij}(s, X_s)\, ds.$$

Then we apply Itô's formula to the semimartingale $\varphi(Y_t)$:

$$\begin{aligned}\varphi(X_t) - \varphi(X_0) &= \sum_{i=1}^p \int_0^t \frac{\partial\varphi}{\partial x_i}(X_s)\, dX_s^i + \frac{1}{2}\sum_{i,j=1}^p \int_0^t \frac{\partial^2\varphi}{\partial x_i \partial x_j}(X_s)\, d\langle X^i, X^j\rangle_s \\ &= \sum_{k=1}^q \sum_{i=1}^p \int_0^t \sigma_{ik}(s, X_s)\frac{\partial\varphi}{\partial x_i}(X_s)\, dB_s^k \\ &\quad + \int_0^t \left[\sum_{i=1}^p b_i(s, X_s)\frac{\partial\varphi}{\partial x_i}(X_s) + \frac{1}{2}\sum_{i,j=1}^p (\sigma\sigma^*)_{ij}(s, X_s)\frac{\partial^2\varphi}{\partial x_i\partial x_j}(X_s)\right] ds.\end{aligned}$$

This is equivalent to (2.2).

(ii) Next, we verify that $P_{s,t}\varphi(x) \to \varphi(x)$ as $t-s \downarrow 0$. Let (Y_t) be a Markov process with transition function $\{P_{s+t,s+v}, v > t \geq 0\}$ beginning from x. For any $\lambda > 0$,

$$\begin{aligned}|P_{s,t}\varphi(x) - \varphi(x)| &= |\mathbb{E}[\varphi(Y_{t-s})] - \varphi(x)| \\ &\leq \sup_{y:|y-x|\leq\lambda} |\varphi(y) - \varphi(x)| + 2\,\|\varphi\|_\infty\, \mathbb{P}(|Y_{t-s} - x| > \lambda) \\ &\leq \sup_{y:|y-x|\leq\lambda} |\varphi(y) - \varphi(x)| + \frac{2C_x}{\lambda^2} e^{C_x(t-s)(1+t-s)} \sum_{k=1}^{4} (t-s)^k \|\varphi\|_\infty.\end{aligned}$$

Hence

$$\lim_{t-s\downarrow 0} |P_{s,t}\varphi(x) - \varphi(x)| \leq \sup_{y:|y-x|\leq\lambda} |\varphi(y) - \varphi(x)|,$$

which holds for all $\lambda > 0$. Since φ is continuous, taking $\lambda \downarrow 0$ gives $P_{s,t}\varphi(x) \to \varphi(x)$. Furthermore, the convergence rate depends on s,t only through their difference $t-s$, i.e.

$$\lim_{h\downarrow 0} \sup_{s\geq 0} |P_{s,s+h}\varphi(x) - \varphi(x)| = 0.$$

(iii) We fix $T \geq 0$ and a Brownian motion $(B_t)_{t\geq 0}$. Then we define a Markov process $(Y_t)_{t\geq 0}$ beginning from $Y_0 = x$ with transition function $\{P_{T+s,T+t}, t > s \geq 0\}$ by

$$Y_t = x + \int_0^t \sigma(T+s, Y_s)\, dB_s + \int_0^t b(T+s, Y_s)\, ds.$$

Similar to (i), we apply Itô's formula to the semimartingale $\varphi(Y_t)$:

$$\begin{aligned}\varphi(Y_t) = \varphi(x) &+ \sum_{i=1}^{p}\sum_{k=1}^{q} \int_0^t \sigma_{ik}(T+s, Y_s) \frac{\partial \varphi}{\partial x_i}(Y_s)\, dB_s^k \\ &+ \int_0^t \left[\sum_{i=1}^{p} b_i(T+s, Y_s)\frac{\partial \varphi}{\partial x_i}(Y_s) + \frac{1}{2}\sum_{i,j=1}^{p} (\sigma\sigma^*)_{ij}(T+s, Y_s) \frac{\partial^2 \varphi}{\partial x_i \partial x_j}(Y_s)\right] ds.\end{aligned}$$

Define function $g : \mathbb{R}_+ \times \mathbb{R}^p \to \mathbb{R}$ by

$$g(t,y) = \frac{1}{2}\sum_{i,j=1}^{p} (\sigma\sigma^*)_{ij}(t,y)\frac{\partial^2\varphi}{\partial x_i \partial x_j}(y) + \sum_{i=1}^{p} b_i(t,y)\frac{\partial\varphi}{\partial x_i}(y) = \frac{1}{2}\sigma\sigma^*(t,y)\cdot\nabla^2\varphi(y) + b(t,y)\nabla\varphi(y).$$

Then g is also a continuous function, and

$$|g(t,y) - g(s,y)| \leq \frac{1}{2}K^2\|\nabla^2\varphi\|_\infty + K\|\nabla\varphi\|_\infty =: R.$$

We then take expectation on both sides of the identity

$$\varphi(Y_t) - \varphi(x) - \int_0^t g(T+s, X_s)\, ds = \sum_{i=1}^{p}\sum_{k=1}^{q}\int_t^{t+\epsilon} \sigma_{ik}(T+s, Y_s)\frac{\partial\varphi}{\partial x_i}(Y_s)\, dB_s^k.$$

to get

$$P_{T,T+t}\,\varphi(x) - \varphi(x) - \int_0^t P_{T,T+s}\, g(T+s, x)\, ds = 0. \tag{2.6}$$

We also note that

$$
\begin{aligned}
&\frac{1}{t}\int_0^t |P_{T,T+s}\, g(T+s,x) - g(T,x)|\, ds \\
&\le \frac{1}{t}\int_0^t |P_{T,T+s}\, g(T+s,x) - P_{T,T+s} g(T,x)|\, ds + \frac{1}{t}\int_0^t |P_{T,T+s}\, g(T,x) - g(T,x)|\, ds \\
&\le \frac{1}{t}\int_0^t \|g(T+s,\cdot) - g(T,\cdot)\|_\infty\, ds + \frac{1}{t}\int_0^t |P_{T,T+s}\, g(T,x) - g(T,x)|\, ds \\
&\le \frac{Rt}{2} + \sup_{s\in[0,t]} |P_{T,T+s} g(T,x) - g(T,x)|,
\end{aligned}
\tag{2.7}
$$

which converges to 0 as $t \downarrow 0$. Combining (2.6) and (2.7), we obtain

$$
\lim_{t\downarrow 0} \frac{P_{T,T+t}\, \varphi(x) - \varphi(x)}{t} = g(T,x).
$$

Replacing T with $T-t$, (2.6) becomes

$$
P_{T-t,T}\, \varphi(x) - \varphi(x) - \int_0^t P_{T-t,T-t+s}\, g(T-t+s,x)\, ds = 0.
$$

Similar to (2.7), we have

$$
\begin{aligned}
&\frac{1}{t}\int_0^t |P_{T-t,T-t+s}\, g(T-t+s,x) - g(T,x)|\, ds \\
&\le \frac{1}{t}\int_0^t |P_{T-t,T-t+s}\, g(T-t+s,x) - P_{T-t,T-t+s} g(T,x)|\, ds + \frac{1}{t}\int_0^t |P_{T-t,T-t+s}\, g(T,x) - g(T,x)|\, ds \\
&\le \frac{1}{t}\int_0^t \|g(T-t+s,\cdot) - g(T,\cdot)\|_\infty\, ds + \frac{1}{t}\int_0^t |P_{T-t,T-t+s}\, g(T,x) - g(T,x)|\, ds \\
&\le \frac{Rt}{2} + \sup_{s\in[0,t]} |P_{T-t,T-t+s} g(T,x) - g(T,x)|.
\end{aligned}
$$

Combining the last two display gives

$$
\lim_{t\downarrow 0} \frac{P_{T-t,T}\, \varphi(x) - \varphi(x)}{t} = g(T,x).
$$

Thus we complete the proof. □

Remark. For notation simplicity, we denote by $\Sigma = \sigma\sigma^*$, and write

$$
\mathscr{L}_t\varphi = \frac{1}{2}\Sigma(t,\cdot)\cdot\nabla^2\varphi + b(t,\cdot)\cdot\nabla\varphi = \frac{1}{2}\sum_{i,j=1}^p (\sigma\sigma^*)_{ij}(t,\cdot)\frac{\partial^2\varphi}{\partial x_i \partial x_j} + \sum_{i=1}^p b_i(t,\cdot)\frac{\partial\varphi}{\partial x_i}, \quad \varphi \in C^2(\mathbb{R}^p).
$$

Using integration by parts, the adjoint of $\mathscr{L}_t$ is given by

$$
\mathscr{L}_t^*\psi = \frac{1}{2}\nabla_x^2\cdot\Sigma(t,\cdot)\psi - \nabla_x\cdot b(t,\cdot)\psi = \frac{1}{2}\sum_{i,j=1}^p \frac{\partial^2}{\partial x_i \partial x_j}(\sigma\sigma^*)_{ij}(t,\cdot)\psi - \sum_{i=1}^p \frac{\partial}{\partial x_i} b_i(t,\cdot)\psi, \quad \psi \in C^2(\mathbb{R}^p).
$$

The adjoint property holds in the sense of integration

$$
\int_{\mathbb{R}^p} \mathscr{L}_t\varphi(x)\, \psi(x)\, dx = \int_{\mathbb{R}^p} \varphi(x)\, \mathscr{L}_t^*\psi(x)\, dx, \quad \varphi,\psi \in C_c^\infty(\mathbb{R}^p).
$$

2.3 The Kolmogorov Backward and Forward Equations

Theorem 2.9 (Kolmogorov backward equation)**.** *Fix $T > 0$ and $\varphi \in C_b^2(\mathbb{R}^p)$. Define $u : \mathbb{R}^+ \times \mathbb{R}^p \to \mathbb{R}$ by*

$$u(t,x) = P_{t,T}\varphi(x).$$

Suppose that $u \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^p)$. Then $u(t,x)$ solves the following Kolmogorov backward equation:

$$\begin{cases} \dfrac{\partial u}{\partial t} + \dfrac{1}{2}\Sigma \cdot \nabla_x^2 u + b \cdot \nabla_x u = 0, \quad 0 \le t \le T, \\ u(T,x) = \varphi(x). \end{cases} \tag{2.8}$$

Proof. By definition, the final value $u(T,x) = \varphi(x)$. To recover the PDE, note that

$$-\frac{\partial u}{\partial t}(t,x) = \lim_{h\downarrow 0} \frac{P_{t-h,T}(x)\varphi - P_{t,T}\varphi(x)}{h} = \lim_{h\downarrow 0} \frac{P_{t-h,t}u(t,x) - u(t,x)}{h}.$$

The result follows from (2.4). □

Remark. We define $\overline{u}(s,x) = u(T-s,x) = P_{T-s,T}\varphi(x)$. Then we turn (2.8) to an initial value problem:

$$\begin{cases} \dfrac{\partial \overline{u}}{\partial t} = \dfrac{1}{2}\Sigma \cdot \nabla_x^2 \overline{u} + b \cdot \nabla_x \overline{u}, \quad 0 \le t \le T, \\ \overline{u}(0,x) = \varphi(x). \end{cases}$$

Theorem 2.10 (Kolmogorov forward equation)**.** *Let $p_{s,t}(\cdot|x)$ be the probability density of $P_{s,t}(x,\cdot)$, and assume that $p_{s,t}(\cdot|x) \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^p)$ for all $t > s \ge 0$ and $x \in \mathbb{R}^p$. Then*

$$\begin{cases} \dfrac{\partial}{\partial t} p_{s,t}(y|x) = \dfrac{1}{2}\nabla_y^2 \cdot \Sigma(t,y)p_{s,t}(y|x) - \nabla_y \cdot b(t,y)p_{s,t}(y|x), \\ \lim_{t\downarrow s} p_{s,t}(y|x) = \delta(y-x). \end{cases} \tag{2.9}$$

Proof. Let $p_{s,t}(\cdot|x)$ be the probability density function of $P_{0,t}(x,\cdot)$, where $t > 0$. Let $\varphi \in C_c^\infty(\mathbb{R}^p)$. Then

$$\lim_{h\downarrow 0} \frac{1}{h}\left(\int_{\mathbb{R}^p} \varphi(y)p_{t,t+h}(y|x)\, ds - \varphi(x) \right) = \mathscr{L}_t\varphi(x).$$

For any $t > 0$, we use interchangeability of derivative and integration:

$$\begin{aligned} \int_{\mathbb{R}^p} \varphi(y)\frac{\partial}{\partial t}p_{s,t}(y|x)\, dy &= \frac{\partial}{\partial t}\int_{\mathbb{R}^p} \varphi(y)p_{s,t}(y|x)\, dy = \lim_{h\downarrow 0}\frac{1}{h}\int_{\mathbb{R}^p} \varphi(y)\left(p_{s,t+h}(y|x) - p_{s,t}(y|x)\right) dy \\ &= \lim_{h\downarrow 0}\frac{1}{h}\left(\int_{\mathbb{R}^p}\int_{\mathbb{R}^p} p_{s,t}(z|x)p_{t,t+h}(y|z)\varphi(y)\, dz\, dy - \int_{\mathbb{R}^p} p_{s,t}(z|x)\varphi(z)\, dz \right) \\ &= \lim_{h\downarrow 0}\int_{\mathbb{R}^p} p_{s,t}(z|x)\cdot \frac{1}{h}\left(\int_{\mathbb{R}^p} \varphi(y)p_{t,t+h}(y|z)\, dy - \varphi(z) \right) dz \\ &= \int_{\mathbb{R}^p} \mathscr{L}_t\varphi(z)p_{s,t}(z|x)\, dz = \int_{\mathbb{R}^p} \varphi(z)\mathscr{L}_t^* p_{s,t}(z|x)\, dz, \end{aligned}$$

where the second line follows from the Chapman-Kolmogorov equation, and in the fifth equality we apply the dominated convergence theorem. Since the above equation holds for all $\varphi \in C_c^\infty(\mathbb{R})$,

$$\frac{\partial}{\partial t}p_{s,t}(y|x) = \mathscr{L}_t^* p_{s,t}(y|x) = \frac{1}{2}\nabla_y^2 \cdot \Sigma(t,y)p_{s,t}(y|x) - \nabla_y \cdot b(t,y)p_{s,t}(y|x). \tag{2.10}$$

Then we finish the proof. □

2.4 The Feynman-Kac Formula

The Feynman-Kac formula is a generalization of the Kolmogorov backward equation. It reveals a connection between parabolic partial differentiabl equations and stochastic differential equations.

Theorem 2.11 (Feynman-Kac formula). *Fix $T > 0$. Let $(X_t)_{0\le t\le T}$ be a diffusion process defined by*

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t.$$

For each $t \ge 0$ and $x \in \mathbb{R}^p$, let

$$u(t, x) = \mathbb{E}\left[e^{-\int_t^T V(s,X_s)\, ds} \varphi(X_T) \,\middle|\, X_t = x \right].$$

Suppose that $u \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^p)$. Then $u(t, x)$ solves the final value problem

$$\begin{cases} \dfrac{\partial u}{\partial t} + b \cdot \nabla_x u + \dfrac{1}{2}\Sigma \cdot \nabla_x^2 u = V u, \quad 0 \le t \le T, \\ u(T, x) = \varphi(x). \end{cases} \tag{2.11}$$

Proof. We let u be a solution of (2.11). By Itô's lemma, we have

$$\begin{aligned} du(t, X_t) &= \frac{\partial u}{\partial t}(t, X_t)\, dt + \nabla_x u(t, X_t) \cdot dX_t + \frac{1}{2}\nabla_x^2 u(t, X_t) \cdot\, d\langle X, X\rangle_t \\ &= \left[\frac{\partial u}{\partial t}(t, X_t) + b(t, X_t) \cdot \nabla_x u(t, X_t) + \frac{1}{2}\Sigma(t, X_t) \cdot \nabla_x^2 u(t, X_t) \right] dt + \nabla_x u(t, X_t) \cdot \sigma(t, X_t)\, dB_t. \end{aligned}$$

Since u satisfies (2.11), we have

$$du(t, X_t) = V(t, X_t) u(t, X_t)\, dt + \nabla_x u(t, X_t) \cdot \sigma(t, X_t)\, dB_t.$$

Wo solve the SDE by integrating both sides on $[t, T]$. For the part involving dt, we note that the integrating factor $e^{\int_0^t V(s,X_s)\, ds}$ is a finite variation process. We multiply both sides of the SDE by the factor and apply methods for solving ordinary differential equations. Then

$$u(T, X_T)\, e^{-\int_0^T V(s,X_s)\, ds} - u(t, X_t)\, e^{-\int_0^t V(s,X_s)\, ds} = \int_t^T e^{-\int_0^s V(\tau,X_\tau)\, d\tau} \nabla_x^2 u(s, X_s) \cdot \sigma(s, X_s)\, dB_s,$$

and

$$u(t, X_t) = \varphi(X_T) e^{-\int_t^T V(s,X_s)\, ds} - \int_t^T e^{-\int_t^s V(\tau,X_\tau)\, d\tau} \nabla_x^2 u(s, X_s) \cdot \sigma(s, X_s)\, dB_s. \tag{2.12}$$

Conditional on $X_t = x$, the process $\int_t^{\cdot} e^{-\int_t^s V(\tau,X_\tau)\, d\tau} \nabla_x^2 u(s, X_s) \cdot \sigma(s, X_s)\, dB_s$ is a continuous local martingale, and it becomes a martingale when stopped by T. By the martingale property,

$$\mathbb{E}\left[\int_t^T e^{-\int_t^s V(\tau,X_\tau)\, d\tau} \nabla_x^2 u(s, X_s) \cdot \sigma(s, X_s)\, dB_s \,\middle|\, X_t = x \right] = 0.$$

Hence by taking the expectation conditional on $X_t = x$ on both sides of (2.12), we get

$$u(t, x) = \mathbb{E}\left[\varphi(X_T) e^{-\int_t^T V(s,X_s)\, ds} \,\middle|\, X_t = x \right],$$

which completes the proof. ☐

Remark. (i) We define

$$\overline{u}(t,x) = u(T-t,x) = \mathbb{E}\left[e^{-\int_{T-t}^{T} V(s,X_s)\,ds} \varphi(X_T) \,\middle|\, X_{T-t} = x \right].$$

Then we turn (2.11) into an initial value problem

$$\begin{cases} \dfrac{\partial \overline{u}}{\partial t} = b \cdot \nabla_x \overline{u} + \dfrac{1}{2}\Sigma \cdot \nabla_x^2 \overline{u} - V\overline{u}, \quad 0 \le t \le T, \\ \overline{u}(0,x) = \varphi(x). \end{cases}$$

(ii) More generally, we define

$$u(t,x) = \mathbb{E}\left[e^{-\int_t^T V(s,X_s)\,ds} \varphi(X_T) + \int_t^T e^{-\int_t^s V(\tau,X_\tau)\,d\tau} g(s,X_s)\,ds \,\middle|\, X_t = x \right].$$

Suppose that $u \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^p)$. Then $u(t,x)$ solves the final value problem

$$\begin{cases} \dfrac{\partial u}{\partial t} + b \cdot \nabla_x u + \dfrac{1}{2}\Sigma \cdot \nabla_x^2 u = Vu + g, \quad 0 \le t \le T, \\ u(T,x) = \varphi(x). \end{cases}$$

This is the Feynman-Kac formula for inhomogeneous parabolic PDEs.

2.5 The Fokker-Planck Equation

The Fokker-Planck equation describes the evolution of the probabiity density in a diffusion process.

Theorem 2.12 (Fokker-Planck equation)**.** *Consider the diffusion process*

$$dX_t = \sigma(t,X_t)\,dB_t + b(t,X_t)\,dt.$$

Let the Assumption of Theorem 2.9 holds. Let $\rho(t,x)$ the probability density of X_t for each $t \ge 0$, and write $\rho_0 = \rho(0,\cdot)$. If $\rho \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^p)$, then ρ solves the following Fokker-Planck equation:

$$\begin{cases} \dfrac{\partial \rho}{\partial t} = \dfrac{1}{2}\nabla_x^2 \cdot \Sigma\rho - \nabla_x \cdot b\rho, \quad t > 0, \\ \rho(0,x) = \rho_0(x). \end{cases} \tag{2.13}$$

Proof. If $X_0 \sim \rho_0$, we have

$$\rho(t,y) = \int_{\mathbb{R}^p} p_{0,t}(y|x)\rho_0(x)\,\mathrm{d}x.$$

Integrating both sides of the Kolmogorov forward equation (2.9), we obtain

$$\int_{\mathbb{R}^p} \frac{\partial}{\partial t} p_{0,t}(y|x)\rho_0(x)\,dx = \frac{1}{2}\int_{\mathbb{R}^p} \nabla_y^2 \cdot \Sigma(t,y) p_{0,t}(y|x)\rho_0(x)\,dx - \int_{\mathbb{R}^p} \nabla_y \cdot b(t,y) p_{0,t}(y|x)\rho_0(x)\,dx$$

The dominated convergence theorem gives the guarantee of exchangeability of derivative and integration. $\square$

2.6 Anderson's Reverse-time SDE

In this subsection, we study the time-reversal of diffusion processes. A reverse-time SDE

$$d\overline{X}_t = \overline{\sigma}(t, \overline{X}_t)\, \overleftarrow{d}\overline{B}_t + \overline{b}(t, \overline{X}_t)\, dt,$$

consists of

- A diffusion coefficient $\overline{\sigma} : \mathbb{R}_+ \times \mathbb{R}^p \to \mathbb{R}^{p\times q}$ and a drift coefficicent $\overline{b} : \mathbb{R}_+ \times \mathbb{R}^p \to \mathbb{R}^p$;
- A backward q-dimensional Brownian motion $(\overline{B}_t)$ with respect to a backward filtration $(\mathscr{G}_t)_{t\geq 0}$, which is a continuous stochastic process such that $\overline{B}_0 = 0$, and for any $t > s \geq 0$, the increment

  $$\overline{B}_t - \overline{B}_s \sim N(0, (t-s)\,\mathrm{Id})$$

  and is independent of $\mathscr{G}_t$;
- A stochastic process $(\overline{X}_t)_{t\geq 0}$ satisfying

  $$\overline{X}_t = \overline{X}_0 + \int_0^t \overline{\sigma}(s, \overline{X}_s)\, \overleftarrow{d}\overline{B}_s + \int_0^t \overline{b}(s, \overline{X}_s)\, ds,$$

  where the backward Itô integral is defined by

  $$\int_0^t \overline{\sigma}(s, \overline{X}_s)\, \overleftarrow{d}\overline{B}_s = \lim_{m\to\infty} \sum_{k=1}^{n_m} \overline{\sigma}\left(t_k^m, \overline{X}_{t_k^m}\right)\left(\overline{B}_{t_k^m} - \overline{B}_{t_{k-1}^m}\right), \tag{2.14}$$

  where $0 = t_0^m < t_1^m < \cdots < t_{n_m}^m = t$ is an increasing sequence of partitions of $[0, t]$ such that the mesh $\max_{1\leq k\leq m_n} |t_k^m - t_{k-1}^m| \downarrow 0$ as $m \to \infty$, and the limit holds in probability. Note that we evaluate the integrad $\overline{\sigma}$ at the right end of each subinterval $[t_{k-1}^m, t_k^m]$ in the Riemann sum (2.14). In comparision, we evaluate the integrand at the left end in the (forward) Itô integral.

Theorem 2.13 (Anderson's reverse-time SDE Theorem)**.** *Let $(X_t)_{t\geq 0}$ be the process defined by*

$$dX_t = \sigma(t, X_t)\, dB_t + b(t, X_t)\, dt,$$

where $\sigma : \mathbb{R}^+ \times \mathbb{R}^p \to \mathbb{R}^{p\times q}$ and $b : \mathbb{R}^+ \times \mathbb{R}^p \to \mathbb{R}^p$ are such as to guarantee the existence of a probability density $\rho(t, x)$ for $t \geq 0$ as a smooth and unique solution of its associated Fokker-Planck equation (2.13)*. Define a q-dimensional process $(\overline{B}_t)_{t\geq 0}$ by $\overline{B}_0 = 0$, and*

$$d\overline{B}_t^j = dB_t^j + \frac{1}{\rho(t, X_t)} \sum_{i=1}^p \frac{\partial(\sigma_{ij}\rho)}{\partial x_i}(t, X_t)\, dt, \quad j = 1, 2, \cdots, q,$$

and suppose further that the Fokker-Planck equation associated with the joint process $(X_t, \overline{B}_t)_{t\geq 0}$ yields a smooth and unique solution for $t \geq 0$. Then

(i) $(\overline{B}_t)_{t\geq 0}$ is a backward Brownian motion with respect to the backward filtration $(\mathscr{G}_t)_{t\geq 0}$, where $\mathscr{G}_t$ is the sub-σ-algebra generated by $(X_s, \overline{B}_s)_{s\geq t}$.

(ii) $(X_t)_{t\geq 0}$ satisfies the reversed-time SDE

$$dX_t = \sigma(t, X_t)\, \overleftarrow{d}\overline{B}_t + \overline{b}(t, X_t)\, dt,$$

where the reverse-time drift coefficient $\overline{b} : \mathbb{R}_+ \times \mathbb{R}^p \to \mathbb{R}^p$ is defined by

$$\overline{b}^i(t, x) = b^i(t, x) - \sum_{k=1}^p \frac{\partial(\sigma\sigma^*)_{ik}}{\partial x_k}(t, x) - \sum_{k=1}^p (\sigma\sigma^*)_{ik} \frac{\partial \log\rho}{\partial x_k}(t, x), \quad i = 1, 2, \cdots, p.$$

Remark. For notation simplicity we write

$$d\overline{B}_t = d\overline{B}_0 + \frac{1}{\rho(t,X_t)} \nabla_x \cdot \rho\sigma^*(t,X_t)\, dt, \quad \text{and} \quad \overline{b}(t,x) = b(t,x) - \nabla_x \cdot \sigma\sigma^*(t,x) - \sigma\sigma^*(t,x)\nabla_x \log \rho(t,x).$$

We let $\rho(t,x,w)$ be the joint density of (X_t, B_t), which solves the following Fokker-Planck equation:

$$\begin{aligned}
\frac{\partial \rho}{\partial t} &= \frac{1}{2}\nabla^2_{x,w} \cdot \begin{pmatrix} \sigma\sigma^* & \sigma \\ \sigma^* & \mathrm{Id} \end{pmatrix} \rho - \nabla_{x,w} \cdot \begin{pmatrix} b\rho \\ \nabla_x \cdot \sigma^* \rho \end{pmatrix} \\
&= \frac{1}{2}\nabla^2_x \cdot \sigma\sigma^*\rho + \sum_{i=1}^{p}\sum_{j=1}^{q} \frac{\partial^2(\sigma_{ij}\rho)}{\partial x_i \partial w_j} + \frac{1}{2}\operatorname{tr}(\nabla^2_w \rho) - \nabla_x \cdot b\rho - \sum_{j=1}^{q} \frac{\partial}{\partial w_j}\left(\sum_{i=1}^{p} \frac{\partial(\sigma_{ij}\rho)}{\partial x_i}\right) \\
&= \frac{1}{2}\nabla^2_x \cdot \sigma\sigma^*\rho + \frac{1}{2}\operatorname{tr}(\nabla^2_w \rho) - \nabla_x \cdot b\rho.
\end{aligned}$$

Let $X_0 \sim \rho_0$ be a random variable independent of the Brownina motion $(B_t)_{t\geq 0}$. Then the full problem is

$$\begin{cases} \frac{\partial \rho}{\partial t} = \frac{1}{2}\nabla^2_x \cdot \sigma\sigma^*\rho + \frac{1}{2}\operatorname{tr}(\nabla^2_w \rho) - \nabla_x \cdot b\rho, \\ \rho(0,x,w) = \rho_0(x)\delta(w). \end{cases} \tag{2.15}$$

Lemma 2.14. *Let $\widetilde{\rho}(t,x)$ be a time-varying density function that solves the Fokker-Planck equation* (2.13) *for the process $(X_t)_{t\geq 0}$. Define*

$$\phi(t,w) = \frac{1}{(2\pi t)^{q/2}} e^{-\frac{|w|^2}{2t}}, \quad t > 0,\ w \in \mathbb{R}^q.$$

Then the solution $\rho(t,x,w)$ of the Fokker-Planck equation for $(X_t, \overline{B}_t)_{t\geq 0}$ is given by

$$\rho(t,x,w) = \widetilde{\rho}(t,x)\phi(t,w).$$

Proof. We note that

$$\frac{1}{2}\nabla^2_x \cdot (\sigma\sigma^*\rho) - \nabla_x \cdot b\rho = \phi \cdot \left(\frac{1}{2}\nabla^2_x \cdot \sigma\sigma^*\widetilde{\rho} - \nabla_x \cdot b\widetilde{\rho}\right),$$

and

$$\frac{1}{2}\operatorname{tr}\left(\nabla^2_w \rho\right) = \frac{1}{2}\sum_{j=1}^{q} \widetilde{\rho} \cdot \left(\frac{w_j^2}{t^2} - \frac{1}{t}\right)\phi = \widetilde{\rho} \cdot \left(\frac{|w|^2}{2t^2} - \frac{q}{2t}\right) \cdot \phi.$$

Therefore ρ satisfy the Fokker-Planck equation:

$$\begin{aligned}
\frac{\partial \rho}{\partial t} &= \phi\frac{\partial \widetilde{\rho}}{\partial t} + \widetilde{\rho}\frac{\partial \phi}{\partial t} = \phi \cdot \left(\frac{1}{2}\nabla^2_x \cdot \sigma\sigma^*\widetilde{\rho} - \nabla_x \cdot b\widetilde{\rho}\right) + \widetilde{\rho} \cdot \left(\frac{|w|^2}{2t^2} - \frac{q}{2t}\right) \cdot \phi \\
&= \frac{1}{2}\nabla^2_x \cdot (\sigma\sigma^*\rho) - \nabla_x \cdot b\rho + \frac{1}{2}\operatorname{tr}\left(\nabla^2_w \rho\right).
\end{aligned}$$

Since the initial condition $\rho(0,t,w) = \rho_0(x)\delta(w)$ is clear, and we complete the proof. $\square$

Remark. An elementary application of Bayes' theorem shows that the conditional distribution of $\overline{B}_t$ given X_t is described by the density function

$$\rho_t(w|x) = \phi(w,t) = \frac{1}{(2\pi t)^{q/2}} e^{-\frac{|w|^2}{2t}}.$$

Hence $\overline{B}_t$ is independent of X_t.

Lemma 2.15. *Let $\rho(t,x,w) = \widetilde{\rho}(t,x)\phi(t,w)$ be the solution of the Fokker-Planck equation* (2.15)*. Then for any $t > s \geq 0$, the conditional density associated with* (2.15) *is*

$$\rho(t, x_t, w_t|s, w_s) = \widetilde{\rho}(t, x_t)\phi(t-s, w_t - w_s) \tag{2.16}$$

Proof. We first consider the conditional density $\rho(x_t, w_t|x_s, w_s)$, which satisfies the Fokker-Planck equation (2.15) with initial condition $\rho(s, x, w|s, x_s, w_s) = \delta(x - x_s)\delta(w - w_s)$. By Lemma 2.14,

$$\rho(t, x_t, w_t|s, x_s, w_s) = p_{s,t}(x_t|x_s)\phi(t-s, w_t - w_s).$$

Then

$$\begin{aligned}\rho(t, x_t, w_t|s, w_s) &= \int_{\mathbb{R}^p} \rho(s, x_s|s, w_s)\rho(t, x_t, w_t|s, x_s, w_s)\, dx_s \\ &= \int_{\mathbb{R}^p} \widetilde{\rho}(s, x_s)p_{s,t}(x_t|x_s)\phi(t-s, w_t - w_s)\, dx_s = \widetilde{\rho}(t, x_t)\phi(t-s, w_t - w_s).\end{aligned}$$

Also, as $t \downarrow s$, both sides of (2.16) has the same limit $\widetilde{\rho}(s, x_t)\delta(w_t - w_s)$. □

Remark. If $t > s \geq 0$, the joint conditional density of X_t and $\overline{B}_t - \overline{B}_s$ given $\overline{B}_s$ is also given by (2.16), which depends on w_t and w_s only through their difference. Therefore, the joint distribution of X_t and $\overline{B}_t - \overline{B}_s$ does not depend on the value of $\overline{B}_s$. Furthermore, X_t, $\overline{B}_t$ and $\overline{B}_t - \overline{B}_s$ are mutually independent:

$$\rho_{s,t}(x_t, w_t - w_s|w_s) = \rho_{s,t}(x_t, w_t - w_s) = \widetilde{\rho}(t, x_t)\phi(t-s, w_t - w_s).$$

Based on this property, we define $\mathscr{G}_s$ to be the σ-algebra generated by random varibles $(X_t)_{t\geq s}$ and $(\overline{B}_t)_{t\geq s}$. Then the continuous semimartingale $(\overline{B}_t)_{t\geq 0}$ satisfies $\overline{B}_t - \overline{B}_s \sim N(0, (t-s)\,\mathrm{Id})$ for all $t > s \geq 0$. Since the increment $\overline{B}_t - \overline{B}_s$ is independent of $\mathscr{G}_t$, we can view the stochastic process $(\overline{B}_t)_{t\geq 0}$ as a backward Brownian motion with respect to the backward filtration $(\mathscr{G}_t)_{t\geq 0}$.

Proof of Theorem 2.13. The result (i) is shown in the above remark, and it remains to shown (ii). By definition of $(X_t, \overline{B}_t)_{t\geq 0}$, we can write $(X_t)_{t\geq 0}$ as the following forward Itô integral:

$$\begin{aligned}X_t &= X_0 + \int_0^t \sigma(s, X_s)\, dB_s + \int_0^t b(s, X_s)\, ds \\ &= X_0 + \int_0^t \sigma(s, X_s)\, d\overline{B}_s + \int_0^t \left(b(s, X_s) - \sigma(s, X_s)\frac{\nabla_x \cdot \sigma^* \rho(s, X_s)}{\rho(s, X_s)}\right) ds.\end{aligned} \tag{2.17}$$

To reverse the process, we need to compute the backward Itô integral. Let $0 = t_0^m < t_1^m < \cdots < t_{n_m}^m = t$ be an increasing sequence of partitions of $[0,t]$ such that $\max_{1\leq k\leq m_n} |t_k^m - t_{k-1}^m| \downarrow 0$ as $m \to \infty$. Then

$$\begin{aligned}&\int_0^t \sigma_{ij}(s, X_s)\, \overleftarrow{d}\overline{B}_s^j = \lim_{m\to\infty} \sum_{k=1}^{n_m} \sigma_{ij}\left(t_k^m, X_{t_k^m}\right)\left(\overline{B}_{t_k^m}^j - \overline{B}_{t_{k-1}^m}^j\right) \\ &= \lim_{m\to\infty} \sum_{k=1}^{n_m} \left(\overline{B}_{t_k^m}^j - \overline{B}_{t_{k-1}^m}^j\right) \\ &\quad \times \left(\sigma_{ij}\left(t_{k-1}^m, X_{t_{k-1}^m}\right) + (t_k^m - t_{k-1}^m)\frac{\partial \sigma_{ij}}{\partial t}\left(t_{k-1}^m, X_{t_{k-1}^m}\right) + \nabla_x \sigma_{ij}\left(t_{k-1}^m, X_{t_{k-1}^m}\right) \cdot \left(X_{t_k^m} - X_{t_{k-1}^m}\right)\right) \\ &= \int_0^t \sigma_{ij}(s, X_s)\, d\overline{B}_s^j + \sum_{k=1}^p \int_0^t \frac{\partial \sigma_{ij}}{\partial x_k}(s, X_s)\, d\langle X^k, \overline{B}^j\rangle_s \\ &= \int_0^t \sigma_{ij}(s, X_s)\, d\overline{B}_s^j + \sum_{k=1}^p \int_0^t \frac{\partial \sigma_{ij}}{\partial x_k}(s, X_s)\, \sigma_{kj}(s, X_s)\, ds.\end{aligned}$$

Hence, to convert (2.17) to a backward Itô integral, we need to subtract a correction term:

$$
\begin{aligned}
X_t^i &= X_0^i + \sum_{j=1}^{q} \int_0^t \sigma_{ij}(s, X_s) \overleftarrow{dB}_s^j \\
&\quad + \int_0^t \left[b_i(s, X_s) - \sum_{j=1}^{q} \sum_{k=1}^{p} \frac{\partial \sigma_{ij}}{\partial x_k}(s, X_s)\, \sigma_{kj}(s, X_s) - \sum_{j=1}^{q} \frac{\sigma_{ij}(s, X_s)}{\rho(s, X_s)} \sum_{k=1}^{p} \frac{\partial(\sigma_{kj}\rho)}{\partial x_k}(s, X_s) \right] ds \\
&= X_0^i + \sum_{j=1}^{q} \int_0^t \sigma_{ij}(s, X_s) \overleftarrow{dB}_s^j + \int_0^t \left[b_i(s, X_s) - \sum_{j=1}^{q} \sum_{k=1}^{p} \left(\frac{\partial(\sigma_{ij}\sigma_{kj})}{\partial x_k}(s, X_s) + \sigma_{ij}\sigma_{kj} \frac{\partial \log \rho}{\partial x_k}(s, X_s) \right) \right] ds \\
&= X_0^i + \sum_{j=1}^{q} \int_0^t \sigma_{ij}(s, X_s) \overleftarrow{dB}_s^j + \int_0^t \left[b_i(s, X_s) - \sum_{k=1}^{p} \left(\frac{\partial(\sigma\sigma^*)_{ik}}{\partial x_k}(s, X_s) + (\sigma\sigma^*)_{ik} \frac{\partial \log \rho}{\partial x_k}(s, X_s) \right) \right] ds.
\end{aligned}
$$

We write this integral to a compact form:

$$
X_t = X_0 + \int_0^t \sigma(s, X_s) \overleftarrow{dB}_s + \int_0^t \left[b(s, X_s) - (\nabla_x \cdot \Sigma)(s, X_s) - \Sigma(s, X_s) \nabla_x \log \rho(s, X_s) \right] ds,
$$

where $\Sigma = \sigma\sigma^*$, and

$$
\nabla_x \cdot \Sigma = \sum_{k=1}^{p} \begin{pmatrix} \frac{\partial \Sigma_{1k}}{\partial x_k} \\ \vdots \\ \frac{\partial \Sigma_{pk}}{\partial x_k} \end{pmatrix}. \tag{2.18}
$$

Equivalently, we write the reverse-time SDE as

$$
dX_t = \sigma(t, X_t) \overleftarrow{dB}_t + \bar{b}(t, X_t)\, dt,
$$

where $\bar{b} = b - \nabla_x \cdot \Sigma - \Sigma \nabla_x \log \rho$ is the reversed-time drift coefficient. □

Remark. We can prove a weaker analogue of Theorem 2.13 in the sense of marginal distribution. We fix $T > 0$, and let $(\rho_t)_{0 \le t \le T}$ be the marginal densities of the process $(X_t)_{0 \le t \le T}$ defined by

$$
dX_t = \sigma(t, X_t)\, dB_t + b(t, X_t)\, dt, \quad X_0 \sim \rho_0. \tag{2.19}
$$

Then the reverse-time SDE is

$$
d\overline{X}_t = \sigma(t, \overline{X}_t) \overleftarrow{dB}_t + [b(t, \overline{X}_t) - \nabla_x \cdot \Sigma(t, \overline{X}_t) - \Sigma(t, \overline{X}_t) \nabla_x \log \rho_t(\overline{X}_t)]\, dt, \quad \overline{X}_T \sim \rho_T.
$$

We can convert this to a forward-time SDE by defining $Y_{T-t} = \overline{X}_t$. Then

$$
dY_t = \sigma(T-t, Y_t)\, dB_t - [b(T-t, Y_t) - \nabla_x \cdot \Sigma(T-t, Y_t) - \Sigma(T-t, Y_t) \nabla_x \log \rho_{T-t}(Y_t)]\, dt, \quad Y_0 \sim \rho_T. \tag{2.20}
$$

Let $(\lambda_t)_{0 \le t \le T}$ be the marginal densities of the process $(Y_t)_{0 \le t \le T}$. Then $(\lambda_t)_{0 \le t \le T}$ satisfies the following Fokker-Planck equation with $\lambda_0 = \rho_T$:

$$
\frac{\partial \lambda_t}{\partial t}(y) = \frac{1}{2} \nabla_y^2 \cdot \Sigma(T-t, y) \lambda_t(y) + \nabla_y \cdot [b(T-t, y) - \nabla_y \cdot \Sigma(T-t, y) - \Sigma(T-t, y) \nabla_y \log \rho_{T-t}(y)]\, \lambda_t(y).
$$

Also, let $\overline{\rho}_t = \lambda_{T-t}$ be the marginal densities of $(\overline{X}_t)_{0 \le t \le T}$. Then $\overline{\rho}_T = \rho_T$, and

$$
\frac{\partial \overline{\rho}_t}{\partial t}(x) = -\frac{1}{2} \nabla_x^2 \cdot \Sigma(t, x) \overline{\rho}_t(x) - \nabla_x \cdot [b(t, x) - \nabla_x \cdot \Sigma(t, x) - \Sigma(t, x) \nabla_x \log \rho_t(x)]\, \overline{\rho}_t(x). \tag{2.21}
$$

We plug-in $(\overline{\rho}_t)_{0\le t\le T} = (\rho_t)_{0\le t\le T}$ into (2.21) to obtain

$$
\begin{aligned}
\frac{\partial \rho_t}{\partial t}(x) &= -\frac{1}{2}\nabla_x^2 \cdot \Sigma(t,x)\rho_t(x) - \nabla_x \cdot \left[b(t,x) - \nabla_x \cdot \Sigma(t,x) - \Sigma(t,x)\nabla_x \log \rho_t(x)\right]\rho_t(x) \\
&= -\frac{1}{2}\nabla_x^2 \cdot \Sigma(t,x)\rho_t(x) - \left[b(t,x) - \nabla_x \cdot \Sigma(t,x) - \Sigma(t,x)\nabla_x \log \rho_t(x)\right] \cdot \nabla_x \rho_t(x) \\
&\quad - \left[\nabla_x \cdot b(t,x) - \nabla_x^2 \cdot \Sigma(t,x) - (\nabla_x \cdot \Sigma)(t,x)\nabla_x \log \rho_t(x) - \Sigma(t,x) \cdot \nabla_x^2 \log \rho_t(x)\right]\rho_t(x) \\
&= -\frac{1}{2}\nabla_x^2 \cdot \Sigma(t,x)\rho_t(x) - b(t,x) \cdot \nabla_x \rho_t(x) + (\nabla_x \cdot \Sigma)(t,x) \cdot \nabla_x \rho_t(x) + \frac{\nabla_x \rho_t(x)^\top}{\rho_t(x)}\Sigma(t,x)\nabla_x \rho_t(x) \\
&\quad - \rho_t(x)\nabla_x \cdot b(t,x) + \rho_t(x)\nabla_x^2 \cdot \Sigma(t,x) + (\nabla_x \cdot \Sigma)(t,x)\nabla_x \rho_t(x) \\
&\quad + \Sigma(t,x) \cdot \nabla_x^2 \rho_t(x) - \frac{\Sigma(t,x)}{\rho_t(x)} \cdot \nabla_x \rho_t(x) \nabla_x \rho_t(x)^\top \\
&= -\frac{1}{2}\nabla_x^2 \cdot \Sigma(t,x)\rho_t(x) + \rho_t(x)\nabla_x^2 \cdot \Sigma(t,x) + 2(\nabla_x \cdot \Sigma)(t,x) \cdot \nabla_x \rho_t(x) + \Sigma(t,x) \cdot \nabla_x^2 \rho_t(x) \\
&\quad - b(t,x) \cdot \nabla_x \rho_t(x) - \rho_t(x)\nabla_x \cdot b(t,x) \\
&= \frac{1}{2}\nabla_x^2 \cdot \Sigma(t,x)\rho_t(x) - \nabla_x \cdot b(t,x)\rho_t(x),
\end{aligned}
$$

which is the Fokker-Planck equation for $(\rho_t)_{0\le t\le T}$. Since $\rho_T = \overline{\rho_T} = \lambda_0$, the marginal densities $(\rho_t)_{0\le t\le T}$ solves the Fokker-Planck equation (2.21), and $\rho_t = \overline{\rho}_t = \lambda_{T-t}$. This implies

$$
X_t \overset{d}{=} \overline{X}_t = Y_{T-t}, \quad 0 \le t \le T.
$$

To summarize, the marginal densities of $(X_t)_{t\ge 0}$ and $(Y_t)_{t\ge 0}$ are reversely aligned.

References

[1] B. D. Anderson, *Reverse-time diffusion equation models*, Stochastic Processes and their Applications, 12 (1982), pp. 313–326.

[2] G. A. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, Springer, 2014.