Lecture 15. Connections with PDE

Tiejun Li[1,2]

[1]School of Mathematical Sciences (SMS),
&
[2]Center for Machine Learning Research (CMLR),
Peking University,
Beijing 100871,
P.R. China
tieli@pku.edu.cn

Office: No. 1 Science Building, Room 1376E

Table of Contents

Particle system

Consider N non-interacting particles moving according to the following deterministic ODEs

$$\frac{d\boldsymbol{X}_t^i}{dt} = \boldsymbol{b}(\boldsymbol{X}_t^i), \quad \boldsymbol{X}_t^i\big|_{t=0} = \boldsymbol{X}_0^i, \quad i = 1, 2, \ldots, N.$$

- Empirical distribution at time t:

$$\mu^N(\boldsymbol{x}, t) = \frac{1}{N} \sum_{i=1}^{N} \delta(\boldsymbol{x} - \boldsymbol{X}_t^i),$$

Particle system

Consider N non-interacting particles moving according to the following deterministic ODEs

$$\frac{d\boldsymbol{X}_t^i}{dt} = \boldsymbol{b}(\boldsymbol{X}_t^i), \quad \boldsymbol{X}_t^i\big|_{t=0} = \boldsymbol{X}_0^i, \quad i = 1, 2, \dots, N.$$

- Empirical distribution at time t:

$$\mu^N(\boldsymbol{x}, t) = \frac{1}{N}\sum_{i=1}^{N} \delta(\boldsymbol{x} - \boldsymbol{X}_t^i),$$

- Problem: What the transition rule for the distribution of these particles is in macroscopic viewpoint, that is, to describe its distributive law when the number of particles N goes to infinity.

Transition rule

For any compactly supported smooth function $\phi(\boldsymbol{x}) \in C_c^\infty(\mathbb{R}^d)$

$$\begin{aligned}
\frac{d}{dt}(\mu^N, \phi) &= \frac{1}{N}\sum_{i=1}^N \frac{d}{dt}\int_{\mathbb{R}^d} \delta(\boldsymbol{x} - \boldsymbol{X}_t^i)\phi(\boldsymbol{x})d\boldsymbol{x} \\
&= \frac{1}{N}\sum_{i=1}^N \frac{d}{dt}\phi(\boldsymbol{X}_t^i) = \frac{1}{N}\sum_{i=1}^N \nabla_{\boldsymbol{x}}\phi(\boldsymbol{X}_t^i)\cdot \boldsymbol{b}(\boldsymbol{X}_t^i) \\
&= \Big(\mu^N, \boldsymbol{b}\cdot\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})\Big),
\end{aligned}$$

where the notation $(\boldsymbol{f}, \boldsymbol{g}) := \int_{\mathbb{R}^d} \boldsymbol{f}(\boldsymbol{x})\cdot\boldsymbol{g}(\boldsymbol{x})d\boldsymbol{x}$ is the inner product of functions.

Transition rule

- Suppose the initial distribution

$$\mu^N(\boldsymbol{x},0) := \frac{1}{N}\sum_{i=1}^{N}\delta(\boldsymbol{x}-\boldsymbol{X}_0^i) \xrightarrow{*} \mu_0(\boldsymbol{x}) \in \mathcal{M}(\mathbb{R}^d) \text{ as } N \to \infty$$

in the sense that $(\mu^N,\phi) \to (\mu,\phi)$ for any $\phi \in C_c^\infty(\mathbb{R}^d)$.

Transition rule

- Suppose the initial distribution

$$\mu^N(\boldsymbol{x},0) := \frac{1}{N}\sum_{i=1}^{N}\delta(\boldsymbol{x}-\boldsymbol{X}_0^i) \xrightarrow{*} \mu_0(\boldsymbol{x}) \in \mathcal{M}(\mathbb{R}^d) \text{ as } N\to\infty$$

in the sense that $(\mu^N,\phi)\to(\mu,\phi)$ for any $\phi\in C_c^\infty(\mathbb{R}^d)$.

- Establish the limit $\mu^N(\boldsymbol{x},t)\xrightarrow{*}\mu(\boldsymbol{x},t)$ and indeed μ satisfies

$$\frac{d}{dt}(\mu,\phi) = (\mu,\boldsymbol{b}\cdot\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})), \quad \mu(\boldsymbol{x},0)=\mu_0(\boldsymbol{x}).$$

Transition rule

- Suppose the initial distribution

$$\mu^N(\boldsymbol{x},0) := \frac{1}{N}\sum_{i=1}^{N}\delta(\boldsymbol{x}-\boldsymbol{X}_0^i) \overset{*}{\to} \mu_0(\boldsymbol{x}) \in \mathcal{M}(\mathbb{R}^d) \text{ as } N\to\infty$$

  in the sense that $(\mu^N,\phi)\to(\mu,\phi)$ for any $\phi\in C_c^\infty(\mathbb{R}^d)$.
- Establish the limit $\mu^N(\boldsymbol{x},t)\overset{*}{\to}\mu(\boldsymbol{x},t)$ and indeed μ satisfies

$$\frac{d}{dt}(\mu,\phi) = (\mu,\boldsymbol{b}\cdot\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})), \quad \mu(\boldsymbol{x},0)=\mu_0(\boldsymbol{x}).$$

- If we assume the probability measure μ has density $\psi(\boldsymbol{x},t)\in C^1(\mathbb{R}^d\times[0,T])$, then we obtain the following hyperbolic equation after integration by parts

$$\partial_t\psi + \nabla_{\boldsymbol{x}}\cdot(\boldsymbol{b}\psi) = 0.$$

Liouville equation

- If the drift vector $\boldsymbol{b}$ satisfies $\nabla_{\boldsymbol{x}} \cdot \boldsymbol{b} = 0$, we get

$$\partial_t \psi + \boldsymbol{b}(\boldsymbol{x}) \cdot \nabla_{\boldsymbol{x}} \psi = 0.$$

This is called the Liouville equation which is well-known in classical mechanics.

Liouville equation

- If the drift vector $\boldsymbol{b}$ satisfies $\nabla_{\boldsymbol{x}} \cdot \boldsymbol{b} = 0$, we get

$$\partial_t \psi + \boldsymbol{b}(\boldsymbol{x}) \cdot \nabla_{\boldsymbol{x}} \psi = 0.$$

  This is called the Liouville equation which is well-known in classical mechanics.

- The orbit of the equation

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{b}(\boldsymbol{x})$$

  is called the characteristics of the above hyperbolic PDE.

Table of Contents

Stochastic case

Replace the deterministic equations with the following SDEs

$$d\boldsymbol{X}_t = \boldsymbol{b}(\boldsymbol{X}_t, t)dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t) \cdot d\boldsymbol{W}_t,$$

- We assume the transition probability density function exists and is defined as $(t \geq s)$

$$p(\boldsymbol{x}, t|\boldsymbol{y}, s)d\boldsymbol{x} = \mathbb{P}\{\boldsymbol{X}_t \in [\boldsymbol{x}, \boldsymbol{x} + d\boldsymbol{x})|\boldsymbol{X}_s = \boldsymbol{y}\}.$$

Stochastic case

Replace the deterministic equations with the following SDEs

$$d\boldsymbol{X}_t = \boldsymbol{b}(\boldsymbol{X}_t, t)dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t) \cdot d\boldsymbol{W}_t,$$

- We assume the transition probability density function exists and is defined as $(t \geq s)$

$$p(\boldsymbol{x}, t|\boldsymbol{y}, s)d\boldsymbol{x} = \mathbb{P}\{\boldsymbol{X}_t \in [\boldsymbol{x}, \boldsymbol{x} + d\boldsymbol{x})|\boldsymbol{X}_s = \boldsymbol{y}\}.$$

- We have the same question on the probability distribution of $\boldsymbol{X}$.

Fokker-Planck equation

- For any function $f \in C_c^\infty(\mathbb{R}^d)$, the Ito formula gives

$$
\begin{aligned}
df(\boldsymbol{X}_t) &= \nabla f(\boldsymbol{X}_t) \cdot d\boldsymbol{X}_t + \frac{1}{2}(d\boldsymbol{X}_t)^T \cdot \nabla^2 f(\boldsymbol{X}_t) \cdot (d\boldsymbol{X}_t) \\
&= (\boldsymbol{b} \cdot \nabla f + \frac{1}{2}\boldsymbol{\sigma}\boldsymbol{\sigma}^T : \nabla^2 f)dt + \nabla f \cdot \boldsymbol{\sigma} \cdot d\boldsymbol{W}_t.
\end{aligned}
$$

Fokker-Planck equation

- For any function $f \in C_c^\infty(\mathbb{R}^d)$, the Ito formula gives

$$
\begin{aligned}
df(\boldsymbol{X}_t) &= \nabla f(\boldsymbol{X}_t) \cdot d\boldsymbol{X}_t + \frac{1}{2}(d\boldsymbol{X}_t)^T \cdot \nabla^2 f(\boldsymbol{X}_t) \cdot (d\boldsymbol{X}_t) \\
&= (\boldsymbol{b} \cdot \nabla f + \frac{1}{2}\boldsymbol{\sigma}\boldsymbol{\sigma}^T : \nabla^2 f)dt + \nabla f \cdot \boldsymbol{\sigma} \cdot d\boldsymbol{W}_t.
\end{aligned}
$$

- Integrating both sides from s to t we get

$$
\begin{aligned}
f(\boldsymbol{X}_t) - f(\boldsymbol{X}_s) &= \int_s^t \nabla f(\boldsymbol{X}_\tau) \cdot \{\boldsymbol{b}(\boldsymbol{X}_\tau, \tau)d\tau + \boldsymbol{\sigma}(\boldsymbol{X}_\tau, \tau)d\boldsymbol{W}_\tau\} \\
&\quad + \frac{1}{2}\int_s^t \sum_{i,j} \partial_{ij}^2 f(\boldsymbol{X}_\tau) a_{ij}(\boldsymbol{X}_\tau, \tau)d\tau,
\end{aligned}
$$

where the diffusion matrix $\boldsymbol{a}(\boldsymbol{x}, t) = \boldsymbol{\sigma}(\boldsymbol{x}, t)\boldsymbol{\sigma}^T(\boldsymbol{x}, t)$.

Fokker-Planck equation

- Now taking expectation on both sides and utilizing the initial condition $\boldsymbol{X}_s = \boldsymbol{y}$, we have

$$\mathbb{E}f(\boldsymbol{X}_t) - f(\boldsymbol{y}) = \mathbb{E}\int_s^t \mathcal{L}f(\boldsymbol{X}_\tau, \tau)d\tau,$$

where the operator $\mathcal{L}$ is defined as

$$\mathcal{L}f(\boldsymbol{x}, t) = \boldsymbol{b}(\boldsymbol{x}, t) \cdot \nabla f(\boldsymbol{x}) + \frac{1}{2}\sum_{i,j} a_{ij}(\boldsymbol{x}, t)\partial_{ij}^2 f(\boldsymbol{x}).$$

Fokker-Planck equation

- Now taking expectation on both sides and utilizing the initial condition $\boldsymbol{X}_s = \boldsymbol{y}$, we have

$$\mathbb{E}f(\boldsymbol{X}_t) - f(\boldsymbol{y}) = \mathbb{E}\int_s^t \mathcal{L}f(\boldsymbol{X}_\tau, \tau)d\tau,$$

  where the operator $\mathcal{L}$ is defined as

$$\mathcal{L}f(\boldsymbol{x}, t) = \boldsymbol{b}(\boldsymbol{x}, t) \cdot \nabla f(\boldsymbol{x}) + \frac{1}{2}\sum_{i,j} a_{ij}(\boldsymbol{x}, t)\partial_{ij}^2 f(\boldsymbol{x}).$$

- In the language of transition pdf $p(\boldsymbol{x}, t|\boldsymbol{y}, s)$, we have

$$\int_{\mathbb{R}^d} f(\boldsymbol{x})p(\boldsymbol{x}, t|\boldsymbol{y}, s)d\boldsymbol{x} - f(\boldsymbol{y}) \quad = \quad \int_s^t \int_{\mathbb{R}^d} \mathcal{L}f(\boldsymbol{x}, \tau)p(\boldsymbol{x}, \tau|\boldsymbol{y}, s)d\boldsymbol{x}d\tau.$$

Fokker-Planck equation

- The adjoint operator $\mathcal{L}^*$ is defined through $(\mathcal{L}f, g)_{L^2} = (f, \mathcal{L}^* g)_{L^2}$. The concrete form of $\mathcal{L}^*$ reads

$$\mathcal{L}^* f(\boldsymbol{x}, t) = -\nabla_{\boldsymbol{x}} \cdot (\boldsymbol{b}(\boldsymbol{x}, t) f(\boldsymbol{x})) + \frac{1}{2}\nabla_{\boldsymbol{x}}^2 : (\boldsymbol{a}(\boldsymbol{x}, t) f(\boldsymbol{x})),$$

where $\nabla_{\boldsymbol{x}}^2 : (\boldsymbol{a} f) = \sum_{ij} \partial_{ij}(a_{ij} f)$.

Fokker-Planck equation

- The adjoint operator $\mathcal{L}^*$ is defined through $(\mathcal{L}f, g)_{L^2} = (f, \mathcal{L}^* g)_{L^2}$. The concrete form of $\mathcal{L}^*$ reads

$$\mathcal{L}^* f(\boldsymbol{x}, t) = -\nabla_{\boldsymbol{x}} \cdot (\boldsymbol{b}(\boldsymbol{x}, t) f(\boldsymbol{x})) + \frac{1}{2} \nabla_{\boldsymbol{x}}^2 : (\boldsymbol{a}(\boldsymbol{x}, t) f(\boldsymbol{x})),$$

  where $\nabla_{\boldsymbol{x}}^2 : (\boldsymbol{a} f) = \sum_{ij} \partial_{ij}(a_{ij} f)$.
- Then the above equation can be simplified to

$$(f, p(\cdot, t|\boldsymbol{y}, s))_{L^2} - f(\boldsymbol{y}) = \int_s^t (f, \mathcal{L}^* p(\cdot, \tau|\boldsymbol{y}, s))_{L^2} d\tau$$

Fokker-Planck equation

- The adjoint operator $\mathcal{L}^*$ is defined through $(\mathcal{L}f, g)_{L^2} = (f, \mathcal{L}^* g)_{L^2}$. The concrete form of $\mathcal{L}^*$ reads

$$\mathcal{L}^* f(\boldsymbol{x}, t) = -\nabla_{\boldsymbol{x}} \cdot (\boldsymbol{b}(\boldsymbol{x}, t) f(\boldsymbol{x})) + \frac{1}{2}\nabla_{\boldsymbol{x}}^2 : (\boldsymbol{a}(\boldsymbol{x}, t) f(\boldsymbol{x})),$$

  where $\nabla_{\boldsymbol{x}}^2 : (\boldsymbol{a} f) = \sum_{ij} \partial_{ij}(a_{ij} f)$.

- Then the above equation can be simplified to

$$(f, p(\cdot, t|\boldsymbol{y}, s))_{L^2} - f(\boldsymbol{y}) = \int_s^t (f, \mathcal{L}^* p(\cdot, \tau|\boldsymbol{y}, s))_{L^2} d\tau$$

- This is exactly the definition of the weak solution of the PDE with respect to t and $\boldsymbol{x}$

$$\partial_t p = \mathcal{L}_{\boldsymbol{x}}^* p(\boldsymbol{x}, t|\boldsymbol{y}, s), \quad p(\boldsymbol{x}, t|\boldsymbol{y}, s)|_{t=s} = \delta(\boldsymbol{x} - \boldsymbol{y}), \quad t \geq s,$$

  in the sense of distribution.

Fokker-Planck equation

The above equation is well-known as the Kolmogorov's forward equation, or the Fokker-Planck equation in physics.

- The "forward" means it is for the forward time variable $t > s$ and its corresponding space variable x.

Fokker-Planck equation

The above equation is well-known as the Kolmogorov's forward equation, or the Fokker-Planck equation in physics.

- The "forward" means it is for the forward time variable $t > s$ and its corresponding space variable $\boldsymbol{x}$.
- When we consider the equation for the backward time variable $s < t$ and $\boldsymbol{y}$, we will call it *backward equation*.

Fokker-Planck equation

The above equation is well-known as the Kolmogorov's forward equation, or the Fokker-Planck equation in physics.

- The "forward" means it is for the forward time variable $t > s$ and its corresponding space variable $\boldsymbol{x}$.
- When we consider the equation for the backward time variable $s < t$ and $\boldsymbol{y}$, we will call it *backward equation*.
- By analogy with the deterministic case, the SDE may be regarded as the "stochastic characteristics" of the parabolic equation.

Fokker-Planck equation

The above equation is well-known as the Kolmogorov's forward equation, or the Fokker-Planck equation in physics.

- The "forward" means it is for the forward time variable $t > s$ and its corresponding space variable $\boldsymbol{x}$.
- When we consider the equation for the backward time variable $s < t$ and $\boldsymbol{y}$, we will call it *backward equation*.
- By analogy with the deterministic case, the SDE may be regarded as the "stochastic characteristics" of the parabolic equation.
- The joint distribution $p(\boldsymbol{x}, t; \boldsymbol{y}, s)$ and the distribution density $p(\boldsymbol{x}, t)$ starting from some initial distribution both satisfy the forward Kolmogorov type equation with respect to $\boldsymbol{x}$ and t.

Brownian motion

- The SDE reads

$$d\boldsymbol{X}_t = d\boldsymbol{W}_t, \quad \boldsymbol{X}_0 = 0.$$

Brownian motion

- The SDE reads

$$d\boldsymbol{X}_t = d\boldsymbol{W}_t, \quad \boldsymbol{X}_0 = 0.$$

- The Fokker-Planck equation is

$$\partial_t p = \frac{1}{2}\Delta p, \quad p(\boldsymbol{x}, 0) = \delta(\boldsymbol{x}).$$

Brownian motion

- The SDE reads

$$d\boldsymbol{X}_t = d\boldsymbol{W}_t, \quad \boldsymbol{X}_0 = 0.$$

- The Fokker-Planck equation is

$$\partial_t p = \frac{1}{2}\Delta p, \quad p(\boldsymbol{x}, 0) = \delta(\boldsymbol{x}).$$

- It is well-known from PDE that its unique solution is the heat kernal

$$p(\boldsymbol{x}, t) = \frac{1}{\sqrt{2\pi t}} \exp\Big(-\frac{\boldsymbol{x}^2}{2t}\Big),$$

which is exactly the pdf of $N(0, t\boldsymbol{I})$. The PDE gives another characterization of the Brownian motion.

Brownian dynamics

- The SDE reads

$$dX_t = -\frac{1}{\gamma}\nabla V(X_t)dt + \sqrt{\frac{2k_BT}{\gamma}}dW_t.$$

Brownian dynamics

- The SDE reads

$$dX_t = -\frac{1}{\gamma}\nabla V(X_t)dt + \sqrt{\frac{2k_BT}{\gamma}}dW_t.$$

- The Fokker-Planck equation is

$$\partial_t p - \nabla \cdot \left(\frac{1}{\gamma}\nabla V(x)p\right) = \frac{k_BT}{\gamma}\Delta p = D\Delta p,$$

where $D = k_BT/\gamma$ is the diffusion coefficient.

Brownian dynamics

- The SDE reads

$$dX_t = -\frac{1}{\gamma}\nabla V(X_t)dt + \sqrt{\frac{2k_BT}{\gamma}}dW_t.$$

- The Fokker-Planck equation is

$$\partial_t p - \nabla \cdot \left(\frac{1}{\gamma}\nabla V(\boldsymbol{x})p\right) = \frac{k_BT}{\gamma}\Delta p = D\Delta p,$$

  where $D = k_BT/\gamma$ is the diffusion coefficient.

- Define the free energy associated with the pdf p as

$$\mathcal{F}(p) = \int_{\mathbb{R}^d} \Big(k_BTp(\boldsymbol{x})\ln p(\boldsymbol{x}) + V(\boldsymbol{x})p(\boldsymbol{x})\Big)d\boldsymbol{x},$$

  where the first term $k_B \int_{\mathbb{R}^d} p(\boldsymbol{x})\ln p(\boldsymbol{x})d\boldsymbol{x}$ corresponds to the negative entropy $-S$ in thermodynamics, and the second term $\int_{\mathbb{R}^d} V(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$ is the internal energy U.

Brownian dynamics

- The chemical potential μ is then given by

$$\mu = \frac{\delta \mathcal{F}}{\delta p} = k_B T(1 + \ln p(\boldsymbol{x})) + V(\boldsymbol{x}).$$

Brownian dynamics

- The chemical potential μ is then given by

$$\mu = \frac{\delta \mathcal{F}}{\delta p} = k_B T(1 + \ln p(\boldsymbol{x})) + V(\boldsymbol{x}).$$

- The velocity field $\boldsymbol{u}(\boldsymbol{x})$ is given by the Fick's Law

$$\boldsymbol{u}(\boldsymbol{x}) = \frac{1}{\gamma}\boldsymbol{f} = -\frac{1}{\gamma}\nabla\mu,$$

where $\boldsymbol{f} = -\nabla\mu$ is the force field.

Brownian dynamics

- The chemical potential μ is then given by

$$\mu = \frac{\delta \mathcal{F}}{\delta p} = k_B T(1 + \ln p(\boldsymbol{x})) + V(\boldsymbol{x}).$$

- The velocity field $\boldsymbol{u}(\boldsymbol{x})$ is given by the Fick's Law

$$\boldsymbol{u}(\boldsymbol{x}) = \frac{1}{\gamma}\boldsymbol{f} = -\frac{1}{\gamma}\nabla\mu,$$

  where $\boldsymbol{f} = -\nabla\mu$ is the force field.
- The current density is defined as

$$\boldsymbol{j}(\boldsymbol{x}) := p(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x})$$

Brownian dynamics

- The chemical potential μ is then given by

$$\mu = \frac{\delta \mathcal{F}}{\delta p} = k_B T(1 + \ln p(\boldsymbol{x})) + V(\boldsymbol{x}).$$

- The velocity field $\boldsymbol{u}(\boldsymbol{x})$ is given by the Fick's Law

$$\boldsymbol{u}(\boldsymbol{x}) = \frac{1}{\gamma}\boldsymbol{f} = -\frac{1}{\gamma}\nabla\mu,$$

  where $\boldsymbol{f} = -\nabla\mu$ is the force field.
- The current density is defined as

$$\boldsymbol{j}(\boldsymbol{x}) := p(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x})$$

- Then the Smoluchowski's equation is a consequence of the continuity equation

$$\partial_t p + \nabla \cdot \boldsymbol{j} = 0.$$

  This approach via deterministic PDE to describe the Brownian dynamics is more common in physics.

Other SDE form

- If the underlying stochastic dynamics is a Stratonovich SDE, we will have its transition pdf satisfies the following type of PDE

$$\partial_t p + \nabla_{\boldsymbol{x}} \cdot (\boldsymbol{b} p) = \frac{1}{2} \nabla_{\boldsymbol{x}} \cdot (\boldsymbol{\sigma} \cdot \nabla_{\boldsymbol{x}} \cdot (\boldsymbol{\sigma} p)),$$

  where $\nabla_{\boldsymbol{x}} \cdot (\boldsymbol{\sigma} \cdot \nabla_{\boldsymbol{x}} \cdot (\boldsymbol{\sigma} p)) = \partial_i(\sigma_{ik} \partial_j (\sigma_{jk} p))$.

Other SDE form

- If the underlying stochastic dynamics is a Stratonovich SDE, we will have its transition pdf satisfies the following type of PDE
$$\partial_t p + \nabla_{\boldsymbol{x}} \cdot (\boldsymbol{b} p) = \frac{1}{2} \nabla_{\boldsymbol{x}} \cdot (\boldsymbol{\sigma} \cdot \nabla_{\boldsymbol{x}} \cdot (\boldsymbol{\sigma} p)),$$
where $\nabla_{\boldsymbol{x}} \cdot (\boldsymbol{\sigma} \cdot \nabla_{\boldsymbol{x}} \cdot (\boldsymbol{\sigma} p)) = \partial_i(\sigma_{ik}\partial_j(\sigma_{jk} p))$.
- If the underlying stochastic dynamics is defined through the backward stochastic integral,
$$d\boldsymbol{X}_t = \boldsymbol{b}(\boldsymbol{x}, t)dt + \boldsymbol{\sigma}(\boldsymbol{x}, t) * d\boldsymbol{W}_t,$$
then $p(\boldsymbol{x}, t)$ satisfies
$$\partial_t p + \partial_i \Big[(b_i + \partial_k \sigma_{ij} \sigma_{kj}) p \Big] = \frac{1}{2} \partial_{ij} : (\sigma_{ik} \sigma_{jk} p),$$
where the Einstein summation convention is assumed. In the one-dimensional case, it can be simplified to
$$\partial_t p + \partial_x (bp) = \frac{1}{2} \partial_x (\sigma^2 \partial_x p).$$

Table of Contents

Probability flux

Many stochastic problems occur in a bounded domain, in which case the boundary conditions are needed.

- To pose suitable boundary conditions in different situations, we need to understand the probability current

$$\boldsymbol{j}(\boldsymbol{x},t) = \boldsymbol{b}(\boldsymbol{x},t)p(\boldsymbol{x},t) - \frac{1}{2}\nabla_{\boldsymbol{x}} \cdot (\boldsymbol{a}(\boldsymbol{x},t)p(\boldsymbol{x},t))$$

  in the Fokker-Planck equation

$$\partial_t p(\boldsymbol{x},t) + \nabla_{\boldsymbol{x}} \cdot \boldsymbol{j}(\boldsymbol{x},t) = 0$$

  more intuitively at first.

Probability flux

Many stochastic problems occur in a bounded domain, in which case the boundary conditions are needed.

- To pose suitable boundary conditions in different situations, we need to understand the probability current

  $$\boldsymbol{j}(\boldsymbol{x},t) = \boldsymbol{b}(\boldsymbol{x},t)p(\boldsymbol{x},t) - \frac{1}{2}\nabla_{\boldsymbol{x}} \cdot (\boldsymbol{a}(\boldsymbol{x},t)p(\boldsymbol{x},t))$$

  in the Fokker-Planck equation

  $$\partial_t p(\boldsymbol{x},t) + \nabla_{\boldsymbol{x}} \cdot \boldsymbol{j}(\boldsymbol{x},t) = 0$$

  more intuitively at first.
- It has the structure of the continuity equations in fluid dynamics.

Boundary Condition

Three commonly used boundary conditions are as follows.

- Reflecting barrier. The particles will be reflected once it hits the boundary ∂D. Thus there will be no probability flux across ∂D, i.e.

$$\boldsymbol{n} \cdot \boldsymbol{j}(\boldsymbol{x}, t) = 0 \quad \boldsymbol{x} \in \partial D.$$

  Note that in this case the total probability is conserved since

$$\frac{d}{dt} \int_D p(\boldsymbol{x}, t) d\boldsymbol{x} = -\int_D \nabla_{\boldsymbol{x}} \cdot \boldsymbol{j}(\boldsymbol{x}, t) d\boldsymbol{x} = -\int_{\partial D} \boldsymbol{n} \cdot \boldsymbol{j}(\boldsymbol{x}, t) dS = 0.$$

Boundary Condition

Three commonly used boundary conditions are as follows.

- Reflecting barrier. The particles will be reflected once it hits the boundary ∂D. Thus there will be no probability flux across ∂D, i.e.

$$\boldsymbol{n} \cdot \boldsymbol{j}(\boldsymbol{x}, t) = 0 \quad \boldsymbol{x} \in \partial D.$$

  Note that in this case the total probability is conserved since

$$\frac{d}{dt} \int_D p(\boldsymbol{x}, t) d\boldsymbol{x} = -\int_D \nabla_{\boldsymbol{x}} \cdot \boldsymbol{j}(\boldsymbol{x}, t) d\boldsymbol{x} = -\int_{\partial D} \boldsymbol{n} \cdot \boldsymbol{j}(\boldsymbol{x}, t) dS = 0.$$

- Absorbing barrier. The particles will be absorbed (or removed) once it hits the boundary ∂D.

$$p(\boldsymbol{x}, t) = 0 \quad \boldsymbol{x} \in \partial D.$$

  The total probability is no longer conserved in this case.

Boundary Condition

Three commonly used boundary conditions are as follows.

- Reflecting barrier. The particles will be reflected once it hits the boundary ∂D. Thus there will be no probability flux across ∂D, i.e.

$$\boldsymbol{n} \cdot \boldsymbol{j}(\boldsymbol{x}, t) = 0 \quad \boldsymbol{x} \in \partial D.$$

  Note that in this case the total probability is conserved since

$$\frac{d}{dt} \int_D p(\boldsymbol{x}, t) d\boldsymbol{x} = - \int_D \nabla_{\boldsymbol{x}} \cdot \boldsymbol{j}(\boldsymbol{x}, t) d\boldsymbol{x} = - \int_{\partial D} \boldsymbol{n} \cdot \boldsymbol{j}(\boldsymbol{x}, t) dS = 0.$$

- Absorbing barrier. The particles will be absorbed (or removed) once it hits the boundary ∂D.

$$p(\boldsymbol{x}, t) = 0 \quad \boldsymbol{x} \in \partial D.$$

  The total probability is no longer conserved in this case.

- Periodic boundary condition. In the periodic case with period L_j in the x_j-direction for $j = 1, \ldots, d$, i.e.

$$p(x_j + L_j, t) = p(x_j, t), \quad j = 1, 2, \ldots, d.$$

Table of Contents

Backward equation

Now let us consider the equation for the transition pdf $p(\boldsymbol{x},t|\boldsymbol{y},s)$ with respect to variable $\boldsymbol{y}$ and s.

- Suppose $\boldsymbol{X}_t$ satisfies the above SDEs. For any given $f(\boldsymbol{x}) \in C_c^\infty(\mathbb{R}^d)$, we define

$$u(\boldsymbol{y},s) = \mathbb{E}^{\boldsymbol{y},s} f(\boldsymbol{X}_t) = \int_{\mathbb{R}^d} f(\boldsymbol{x}) p(\boldsymbol{x},t|\boldsymbol{y},s) d\boldsymbol{x}, \quad s \le t.$$

  Assume that $p(\boldsymbol{x},t|\boldsymbol{y},s)$ is C^1 in s and C^2 in $\boldsymbol{y}$, then we have

$$du(\boldsymbol{X}_\tau,\tau) = (\partial_\tau u + \mathcal{L}u)(\boldsymbol{X}_\tau,\tau)d\tau + \nabla u \cdot \boldsymbol{\sigma} \cdot d\boldsymbol{W}_\tau$$

  by Ito formula.

Backward equation

Now let us consider the equation for the transition pdf $p(\boldsymbol{x}, t|\boldsymbol{y}, s)$ with respect to variable $\boldsymbol{y}$ and s.

- Suppose $\boldsymbol{X}_t$ satisfies the above SDEs. For any given $f(\boldsymbol{x}) \in C_c^\infty(\mathbb{R}^d)$, we define

$$u(\boldsymbol{y}, s) = \mathbb{E}^{\boldsymbol{y},s} f(\boldsymbol{X}_t) = \int_{\mathbb{R}^d} f(\boldsymbol{x}) p(\boldsymbol{x}, t|\boldsymbol{y}, s) d\boldsymbol{x}, \quad s \leq t.$$

  Assume that $p(\boldsymbol{x}, t|\boldsymbol{y}, s)$ is C^1 in s and C^2 in $\boldsymbol{y}$, then we have

$$du(\boldsymbol{X}_\tau, \tau) = (\partial_\tau u + \mathcal{L}u)(\boldsymbol{X}_\tau, \tau) d\tau + \nabla u \cdot \boldsymbol{\sigma} \cdot d\boldsymbol{W}_\tau$$

  by Ito formula.
- Taking expectation we obtain

$$\lim_{t \to s} \frac{1}{t-s} (\mathbb{E}^{\boldsymbol{y},s} u(\boldsymbol{X}_t, t) - u(\boldsymbol{y}, s))$$
$$= \lim_{t \to s} \frac{1}{t-s} \int_s^t \mathbb{E}^{\boldsymbol{y},s} (\partial_\tau u + \mathcal{L}u)(\boldsymbol{X}_\tau, \tau) d\tau = \partial_s u(\boldsymbol{y}, s) + \mathcal{L}u(\boldsymbol{y}, s).$$

Backward equation

- On the other hand it is obvious that

$$\mathbb{E}^{\boldsymbol{y},s} u(\boldsymbol{X}_t, t) = \mathbb{E}^{\boldsymbol{y},s} f(\boldsymbol{X}_t) = u(\boldsymbol{y}, s)$$

and thus

$$\partial_s u(\boldsymbol{y}, s) + \mathcal{L} u(\boldsymbol{y}, s) = 0.$$

Backward equation

- On the other hand it is obvious that

$$\mathbb{E}^{\boldsymbol{y},s} u(\boldsymbol{X}_t, t) = \mathbb{E}^{\boldsymbol{y},s} f(\boldsymbol{X}_t) = u(\boldsymbol{y}, s)$$

and thus

$$\partial_s u(\boldsymbol{y}, s) + \mathcal{L} u(\boldsymbol{y}, s) = 0.$$

- From the arbitrariness of f, we obtain

$$\partial_s p(\boldsymbol{x}, t|\boldsymbol{y}, s) + \mathcal{L}_{\boldsymbol{y}} p(\boldsymbol{x}, t|\boldsymbol{y}, s) = 0, \quad p(\boldsymbol{x}, t|\boldsymbol{y}, t) = \delta(\boldsymbol{x} - \boldsymbol{y}), \quad s < t.$$

This is the well-know Kolmogorov backward equation for the transition density since the time variable s goes backward.

Table of Contents

Invariant distribution

In the following discussion, we consider the situation that the drift $\boldsymbol{b}$ and diffusion coefficient $\boldsymbol{\sigma}$ does not depend on t.

- In this case, the process $\{\boldsymbol{X}_t\}$ is a time-homogeneous Markov process since the transition rule only depends on the states other than the time.

Invariant distribution

In the following discussion, we consider the situation that the drift $\boldsymbol{b}$ and diffusion coefficient $\boldsymbol{\sigma}$ does not depend on t.

- In this case, the process $\{\boldsymbol{X}_t\}$ is a time-homogeneous Markov process since the transition rule only depends on the states other than the time.
- It is interesting to study the case when the system achieves a steady state: that is, the pdf is independent of the time, if the system admits such a solution.

Invariant distribution

In the following discussion, we consider the situation that the drift $\boldsymbol{b}$ and diffusion coefficient $\boldsymbol{\sigma}$ does not depend on t.

- In this case, the process $\{\boldsymbol{X}_t\}$ is a time-homogeneous Markov process since the transition rule only depends on the states other than the time.
- It is interesting to study the case when the system achieves a steady state: that is, the pdf is independent of the time, if the system admits such a solution.
- The steady state pdf satisfies the following PDE

$$\nabla_{\boldsymbol{x}} \cdot (\boldsymbol{b}(\boldsymbol{x}) p_s(\boldsymbol{x})) = \frac{1}{2} \nabla_{\boldsymbol{x}}^2 : (\boldsymbol{a}(\boldsymbol{x}) \; p_s(\boldsymbol{x}))$$

  with suitable boundary conditions. This $p_s(\boldsymbol{x})$ is called the stationary distribution or invariant distribution of the considered system.

Detailed balance

- Specially for the Langevin equation , the invariant distribution satisfies

$$\nabla \cdot \boldsymbol{j}_s(\boldsymbol{x}) = 0,$$

Detailed balance

- Specially for the Langevin equation , the invariant distribution satisfies

$$\nabla \cdot \boldsymbol{j}_s(\boldsymbol{x}) = 0,$$

- In particular, we are interested in the equilibrium solution with a stronger condition $\boldsymbol{j}_s = 0$, i.e. the detailed balance condition in the continuous case, which implies the chemical potential

$$\mu = \text{constant}.$$

Detailed balance

- Specially for the Langevin equation , the invariant distribution satisfies

$$\nabla \cdot \boldsymbol{j}_s(\boldsymbol{x}) = 0,$$

- In particular, we are interested in the equilibrium solution with a stronger condition $\boldsymbol{j}_s = 0$, i.e. the detailed balance condition in the continuous case, which implies the chemical potential

$$\mu = \text{constant}.$$

- It is not difficult to deduce the following well-known Gibbs distribution for the equilibrium

$$p_s(\boldsymbol{x}) = \frac{1}{Z} \exp\Big(- \frac{V(\boldsymbol{x})}{k_B T} \Big)$$

as long as the normalization constant

$$Z = \int_{\mathbb{R}^d} e^{-\frac{V(\boldsymbol{x})}{k_B T}} d\boldsymbol{x}$$

is finite.

Table of Contents

Semigroup

- For the time-homogeneous SDEs, the translational invariance of time for its transition kernel $p(\cdot, t|\boldsymbol{y}, s)$

$$p(A, t+s|\boldsymbol{y}, s) = p(A, t|\boldsymbol{y}, 0), \quad s, t \geq 0$$

for any $\boldsymbol{y} \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$, where $p(A, t|\boldsymbol{y}, s) := \mathbb{E}^{\boldsymbol{y},s} 1_A(\boldsymbol{X}_t) = \int_A p(d\boldsymbol{x}, t|\boldsymbol{y}, s)$.

Semigroup

- For the time-homogeneous SDEs, the translational invariance of time for its transition kernel $p(\cdot, t|\boldsymbol{y}, s)$

$$p(A, t+s|\boldsymbol{y}, s) = p(A, t|\boldsymbol{y}, 0), \quad s, t \geq 0$$

  for any $\boldsymbol{y} \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$, where
  $p(A, t|\boldsymbol{y}, s) := \mathbb{E}^{\boldsymbol{y},s} 1_A(\boldsymbol{X}_t) = \int_A p(d\boldsymbol{x}, t|\boldsymbol{y}, s)$.
- Define the operator T_t on any function $f \in C_0(\mathbb{R}^d)$ as

$$T_t f(\boldsymbol{x}) = \mathbb{E}^{\boldsymbol{x}} f(\boldsymbol{X}_t) = \int_{\mathbb{R}^d} f(\boldsymbol{z}) p(d\boldsymbol{z}, t|\boldsymbol{x}, 0).$$

  Then we have $T_0 f(\boldsymbol{x}) = f(\boldsymbol{x})$ and the following semigroup property for any $t, s \geq 0$

$$\begin{aligned} T_t \circ T_s f(\boldsymbol{x}) &= \mathbb{E}^{\boldsymbol{x}}(\mathbb{E}^{\boldsymbol{X}_t} f(\boldsymbol{X}_s)) \\ &= \int p(d\boldsymbol{y}, t|\boldsymbol{x}, 0) \int f(\boldsymbol{z}) p(d\boldsymbol{z}, s|\boldsymbol{y}, 0) \\ &= \int f(\boldsymbol{z}) \int p(d\boldsymbol{z}, s+t|\boldsymbol{y}, t) p(d\boldsymbol{y}, t|\boldsymbol{x}, 0) \\ &= \mathbb{E}^{\boldsymbol{x}}(f(\boldsymbol{X}_{t+s})) = T_{t+s} f(\boldsymbol{x}). \end{aligned}$$

Semigroup

- Under the condition that $\boldsymbol{b}$ and $\boldsymbol{\sigma}$ are bounded and Lipschitz, one can further show $T_t : C_0(\mathbb{R}^d) \to C_0(\mathbb{R}^d)$ and it is strongly continuous in the sense that

$$\lim_{t \to 0+} \|T_t f - f\|_\infty = 0, \quad \text{for any } f \in C_0(\mathbb{R}^d).$$

T_t is called Feller semigroup in the literature. With this setup, we can utilize the tools from semigroup theory to study T_t .

Semigroup

- Under the condition that $\boldsymbol{b}$ and $\boldsymbol{\sigma}$ are bounded and Lipschitz, one can further show $T_t : C_0(\mathbb{R}^d) \to C_0(\mathbb{R}^d)$ and it is strongly continuous in the sense that

$$\lim_{t \to 0+} \|T_t f - f\|_\infty = 0, \quad \text{for any } f \in C_0(\mathbb{R}^d).$$

  T_t is called Feller semigroup in the literature. With this setup, we can utilize the tools from semigroup theory to study T_t .
- The infinitesimal generator $\mathcal{A}$ of T_t is defined as

$$\mathcal{A}f(\boldsymbol{x}) = \lim_{t \to 0+} \frac{\mathbb{E}^{\boldsymbol{x}} f(\boldsymbol{X}_t) - f(\boldsymbol{x})}{t},$$

  where $f \in D(\mathcal{A}) := \{f \in C_0(\mathbb{R}^d)$ such that the limit exists$\}$.

Semigroup

- Under the condition that $\boldsymbol{b}$ and $\boldsymbol{\sigma}$ are bounded and Lipschitz, one can further show $T_t : C_0(\mathbb{R}^d) \to C_0(\mathbb{R}^d)$ and it is strongly continuous in the sense that

$$\lim_{t\to 0+} \|T_t f - f\|_\infty = 0, \quad \text{for any } f \in C_0(\mathbb{R}^d).$$

  T_t is called Feller semigroup in the literature. With this setup, we can utilize the tools from semigroup theory to study T_t .
- The infinitesimal generator $\mathcal{A}$ of T_t is defined as

$$\mathcal{A}f(\boldsymbol{x}) = \lim_{t\to 0+} \frac{\mathbb{E}^{\boldsymbol{x}} f(\boldsymbol{X}_t) - f(\boldsymbol{x})}{t},$$

  where $f \in D(\mathcal{A}) := \{f \in C_0(\mathbb{R}^d)$ such that the limit exists$\}$.
- For $f \in C_c^2(\mathbb{R}^d) \subset D(\mathcal{A})$ we have

$$\mathcal{A}f(\boldsymbol{x}) = \mathcal{L}f(\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x}) \cdot \nabla f(\boldsymbol{x}) + \frac{1}{2}(\boldsymbol{\sigma}\boldsymbol{\sigma}^T) : \nabla^2 f(\boldsymbol{x}).$$

  from Ito formula.

Backward Equation

We will show that $u(\boldsymbol{x},t) = \mathbb{E}^{\boldsymbol{x}} f(\boldsymbol{X}_t)$ satisfies the *backward equation* for $f \in C_c^2(\mathbb{R}^d)$

$$\partial_t u = \mathcal{A}u(\boldsymbol{x}), \quad u|_{t=0} = f(\boldsymbol{x}).$$

Proof. At first it is not difficult to observe that $u(\boldsymbol{x},t)$ is differentiable with respect to t from Ito's formula and the condition $f \in C_c^2(\mathbb{R}^d)$. For any fixed $t > 0$, define $g(\boldsymbol{x}) = u(\boldsymbol{x},t)$. Then we have

$$\begin{aligned}
\mathcal{A}g(\boldsymbol{x}) &= \lim_{s\to 0+} \frac{1}{s}\Big(\mathbb{E}^{\boldsymbol{x}} g(\boldsymbol{X}_s) - g(\boldsymbol{x})\Big) \\
&= \lim_{s\to 0+} \frac{1}{s}\Big(\mathbb{E}^{\boldsymbol{x}}\mathbb{E}^{\boldsymbol{X}_s} f(\boldsymbol{X}_t) - \mathbb{E}^{\boldsymbol{x}} f(\boldsymbol{X}_t)\Big) \\
&= \lim_{s\to 0+} \frac{1}{s}\Big(\mathbb{E}^{\boldsymbol{x}} f(\boldsymbol{X}_{t+s}) - \mathbb{E}^{\boldsymbol{x}} f(\boldsymbol{X}_t)\Big) \\
&= \lim_{s\to 0+} \frac{1}{s}(u(\boldsymbol{x},t+s) - u(\boldsymbol{x},t)) = \partial_t u(\boldsymbol{x},t).
\end{aligned}$$

This means $u(\cdot,t) \in D(\mathcal{A})$ and the proof is complete.

Feynman-Kac Formula

Theorem

(Feynman-Kac Formula) Let $f \in C_0^2(\mathbb{R}^d)$ and $q \in C(\mathbb{R}^d)$. Assume that q is lower bounded, then

$$v(\boldsymbol{x}, t) = \mathbb{E}^x\Big(\exp(\int_0^t q(\boldsymbol{X}_s)ds) f(\boldsymbol{X}_t)\Big)$$

satisfies the PDE

$$\partial_t v = \mathcal{A}v + qv, \quad v|_{t=0} = f(\boldsymbol{x}).$$

Feynman-Kac Formula

Theorem

(Feynman-Kac Formula) Let $f \in C_0^2(\mathbb{R}^d)$ and $q \in C(\mathbb{R}^d)$. Assume that q is lower bounded, then

$$v(\boldsymbol{x}, t) = \mathbb{E}^x \Big(\exp(\int_0^t q(\boldsymbol{X}_s) ds) f(\boldsymbol{X}_t) \Big)$$

satisfies the PDE

$$\partial_t v = \mathcal{A} v + q v, \quad v|_{t=0} = f(\boldsymbol{x}).$$

- **Intuitive explanation:** In the absence of Brownian motion, the SDE becomes $\frac{d\boldsymbol{X}_t}{dt} = \boldsymbol{b}(\boldsymbol{X}_t), \boldsymbol{X}_0 = \boldsymbol{x}$ and the PDE becomes
$$\partial_t v = \boldsymbol{b} \cdot \nabla v + q v, \quad v|_{t=0} = f(\boldsymbol{x}).$$
The method of characteristics gives us
$$v(\boldsymbol{x}, t) = \exp(\int_0^t q(\boldsymbol{X}_s) ds) f(\boldsymbol{X}_t).$$

Feynman-Kac Formula

The Feynmann-Kac formula tells us the solution of that parabolic PDE can be represented by the ensemble of solution for the ODEs with stochastic characteristics originated from $\boldsymbol{x}$.

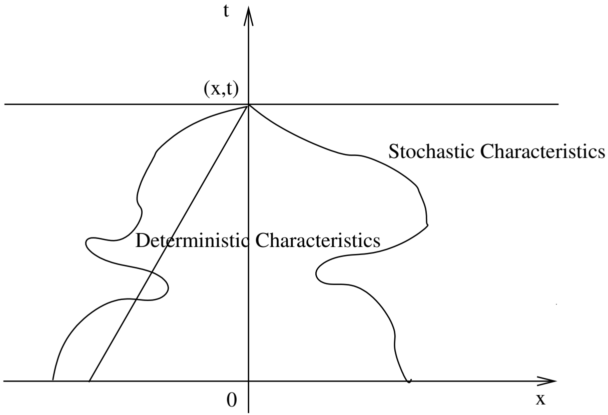

Figure: Schematics of Feynmann-Kac formula.

Proof

Let $Y_t = f(\boldsymbol{X}_t), Z_t = \exp(\int_0^t q(\boldsymbol{X}_s)ds)$, define $v(\boldsymbol{x}, t) = \mathbb{E}^{\boldsymbol{x}}(Y_t Z_t)$. With the similar reason as the previous section, we have $v(\boldsymbol{x}, t)$ is differentiable with respect to t and

$$
\begin{aligned}
&\frac{1}{s}\Big(\mathbb{E}^{\boldsymbol{x}} v(\boldsymbol{X}_s, t) - v(\boldsymbol{x}, t)\Big) \\
&= \frac{1}{s}\Big(\mathbb{E}^{\boldsymbol{x}}\mathbb{E}^{\boldsymbol{X}_s} Z_t f(\boldsymbol{X}_t) - \mathbb{E}^{\boldsymbol{x}} Z_t f(\boldsymbol{X}_t)\Big) \\
&= \frac{1}{s}\Big(\mathbb{E}^{\boldsymbol{x}} \exp(\int_0^t q(\boldsymbol{X}_{r+s})dr) f(\boldsymbol{X}_{t+s}) - \mathbb{E}^{\boldsymbol{x}} Z_t f(\boldsymbol{X}_t)\Big) \\
&= \frac{1}{s}\mathbb{E}^{\boldsymbol{x}}\Big(\exp(-\int_0^s q(\boldsymbol{X}_r)dr) Z_{t+s} f(\boldsymbol{X}_{t+s}) - Z_t f(\boldsymbol{X}_t)\Big) \\
&= \frac{1}{s}\mathbb{E}^{\boldsymbol{x}}\Big(Z_{t+s} f(\boldsymbol{X}_{t+s}) - Z_t f(\boldsymbol{X}_t)\Big) \\
&+ \frac{1}{s}\mathbb{E}^{\boldsymbol{x}}\Big(Z_{t+s} f(\boldsymbol{X}_{t+s})(\exp(-\int_0^s q(\boldsymbol{X}_r)dr) - 1)\Big) \\
&\to \partial_t v - q(\boldsymbol{x}) v(\boldsymbol{x}, t) \qquad \text{as } s \to 0.
\end{aligned}
$$

The left hand side is $\mathcal{A}v(\boldsymbol{x}, t)$ by definition. The proof is complete.

First exit time

Theorem

Suppose $D \subset \mathbb{R}^d$ is a bounded open set and the boundary ∂D is of C^2 type. The coefficients $\boldsymbol{b}, \boldsymbol{\sigma}$ of the SDEs satisfy the Lipschitz condition on $\bar{D}$ and the diffusion matrix $\boldsymbol{a}$ is coercive which is defined as

$$\sum_{i,j} a_{ij}(\boldsymbol{x})\xi_i\xi_j \geq K|\xi|^2 \quad \text{for } \boldsymbol{x} \in D,\ \xi \in \mathbb{R}^d,\ K > 0.$$

Then for $f \in C(\partial D)$, the solution of PDE

$$\mathcal{A}u = 0 \text{ in } D, \quad u = f(\boldsymbol{x}) \text{ on } \partial D$$

can be represented as $u(\boldsymbol{x}) = \mathbb{E}^{\boldsymbol{x}}\big(f(\boldsymbol{X}_{\tau_D})\big)$, where τ_D is the first exit time from domain D defined as $\tau_D := \inf_t\{t \geq 0, \boldsymbol{X}_t \notin D\}$ and thus $\boldsymbol{X}_{\tau_D}$ is the first exit point.
Specially, if $\mathcal{A}u = \frac{1}{2}\Delta u$, then $u(\boldsymbol{x}) = \mathbb{E}^{\boldsymbol{x}}\big(f(\boldsymbol{W}_{\tau_D})\big)$.

Heuristic

- **Heuristic proof.** From PDE theory, one has the solution $u \in C^2(D) \cap C(\bar{D})$. So we can apply the Ito's formula to $u(\boldsymbol{X}_t)$ and take expectation

$$\mathbb{E}^{\boldsymbol{x}} u(\boldsymbol{X}_{\tau_D}) - u(\boldsymbol{x}) = \mathbb{E}^{\boldsymbol{x}} \int_0^{\tau_D} \mathcal{A} u(\boldsymbol{X}_t) dt = 0.$$

Thus

$$u(\boldsymbol{x}) = \mathbb{E}^{\boldsymbol{x}} u(\boldsymbol{X}_{\tau_D}) = \mathbb{E}^{\boldsymbol{x}} \big(f(\boldsymbol{X}_{\tau_D}) \big).$$

Heuristic

- **Heuristic proof.** From PDE theory, one has the solution $u \in C^2(D) \cap C(\bar{D})$. So we can apply the Ito's formula to $u(\boldsymbol{X}_t)$ and take expectation

$$\mathbb{E}^{\boldsymbol{x}} u(\boldsymbol{X}_{\tau_D}) - u(\boldsymbol{x}) = \mathbb{E}^{\boldsymbol{x}} \int_0^{\tau_D} \mathcal{A}u(\boldsymbol{X}_t)dt = 0.$$

Thus

$$u(\boldsymbol{x}) = \mathbb{E}^{\boldsymbol{x}} u(\boldsymbol{X}_{\tau_D}) = \mathbb{E}^{\boldsymbol{x}}\big(f(\boldsymbol{X}_{\tau_D})\big).$$

- Note that in the above derivations we naively take the expectation of the stochastic integral term to be zero. But this is not true in general because τ_D is a random time. In fact, it is the result of the following useful Dynkin's formula.

Heuristic

- **Heuristic proof.** From PDE theory, one has the solution $u \in C^2(D) \cap C(\bar{D})$. So we can apply the Ito's formula to $u(\boldsymbol{X}_t)$ and take expectation

$$\mathbb{E}^{\boldsymbol{x}} u(\boldsymbol{X}_{\tau_D}) - u(\boldsymbol{x}) = \mathbb{E}^{\boldsymbol{x}} \int_0^{\tau_D} \mathcal{A}u(\boldsymbol{X}_t)dt = 0.$$

  Thus

$$u(\boldsymbol{x}) = \mathbb{E}^{\boldsymbol{x}} u(\boldsymbol{X}_{\tau_D}) = \mathbb{E}^{\boldsymbol{x}}\big(f(\boldsymbol{X}_{\tau_D})\big).$$

- Note that in the above derivations we naively take the expectation of the stochastic integral term to be zero. But this is not true in general because τ_D is a random time. In fact, it is the result of the following useful Dynkin's formula.

Lemma (Dynkin's formula)

Let $f \in C_0^2(\mathbb{R}^d)$. Suppose τ is a stopping time with $\mathbb{E}^{\boldsymbol{x}}\tau < \infty$, then

$$\mathbb{E}^{\boldsymbol{x}} f(\boldsymbol{X}_\tau) = f(\boldsymbol{x}) + \mathbb{E}^{\boldsymbol{x}} \int_0^{\tau} \mathcal{A}u(\boldsymbol{X}_t)dt.$$

On the condition $\mathbb{E}^{\boldsymbol{x}}\tau_D < \infty$

- To prove $\mathbb{E}^{\boldsymbol{x}}\tau_D < \infty$, we define an auxiliary function $h(\boldsymbol{x}) = -A\exp(\lambda x_1)$. Then for sufficiently large $A, \lambda > 0$ we have

$$\mathcal{A}h(\boldsymbol{x}) = \frac{1}{2}\sum_{ij} a_{ij}(\boldsymbol{x})\partial_{ij}h(\boldsymbol{x}) + \sum_i b_i(\boldsymbol{x})\partial_i h(\boldsymbol{x}) \le -1, \quad \boldsymbol{x} \in D.$$

On the condition $\mathbb{E}^{\boldsymbol{x}}\tau_D < \infty$

- To prove $\mathbb{E}^{\boldsymbol{x}}\tau_D < \infty$, we define an auxiliary function $h(\boldsymbol{x}) = -A\exp(\lambda x_1)$. Then for sufficiently large $A, \lambda > 0$ we have

$$\mathcal{A}h(\boldsymbol{x}) = \frac{1}{2}\sum_{ij} a_{ij}(\boldsymbol{x})\partial_{ij}h(\boldsymbol{x}) + \sum_i b_i(\boldsymbol{x})\partial_i h(\boldsymbol{x}) \leq -1, \quad \boldsymbol{x} \in D.$$

- By Itô's formula

$$\mathbb{E}^{\boldsymbol{x}} h(\boldsymbol{X}_{\tau_D\wedge T}) - h(\boldsymbol{x}) = \mathbb{E}^{\boldsymbol{x}}\int_0^{\tau_D\wedge T}\mathcal{A}h(\boldsymbol{X}_s)ds \leq -\mathbb{E}^{\boldsymbol{x}}(\tau_D \wedge T)$$

for any fixed $T > 0$.

On the condition $\mathbb{E}^{\boldsymbol{x}}\tau_D < \infty$

- To prove $\mathbb{E}^{\boldsymbol{x}}\tau_D < \infty$, we define an auxiliary function $h(\boldsymbol{x}) = -A\exp(\lambda x_1)$. Then for sufficiently large $A, \lambda > 0$ we have

$$\mathcal{A}h(\boldsymbol{x}) = \frac{1}{2}\sum_{ij} a_{ij}(\boldsymbol{x})\partial_{ij}h(\boldsymbol{x}) + \sum_i b_i(\boldsymbol{x})\partial_i h(\boldsymbol{x}) \leq -1, \quad \boldsymbol{x} \in D.$$

- By Itô's formula

$$\mathbb{E}^{\boldsymbol{x}} h(\boldsymbol{X}_{\tau_D \wedge T}) - h(\boldsymbol{x}) = \mathbb{E}^{\boldsymbol{x}} \int_0^{\tau_D \wedge T} \mathcal{A}h(\boldsymbol{X}_s)ds \leq -\mathbb{E}^{\boldsymbol{x}}(\tau_D \wedge T)$$

  for any fixed $T > 0$.
- Since $|h(\boldsymbol{x})| \leq C$ for $\boldsymbol{x} \in D$, we have

$$\mathbb{E}^{\boldsymbol{x}}(\tau_D \wedge T) \leq 2C.$$

  Taking $T \to \infty$ and using the monotone convergence theorem we obtain $\mathbb{E}^{\boldsymbol{x}}(\tau_D) \leq 2C$.