Written Assignment — Week 8

2025.10.28

1. Sliced Score Matching (SSM)

We are given

$$L_{\text{SSM}}(\theta) = \mathbb{E}_{x\sim p(x)} \|S(x;\theta)\|^2 + \mathbb{E}_{x\sim p(x)}\mathbb{E}_{v\sim p(v)}\left[\, 2v^\top \nabla_x\big(v^\top S(x;\theta)\big)\right]. \tag{1}$$

Let $v \in \mathbb{R}^d$ be independent of x, and assume

$$\mathbb{E}_v[v] = 0, \qquad \mathbb{E}_v[vv^\top] = I_d.$$

Also, denote $S = S(x;\theta) \in \mathbb{R}^d$ for convenience.
Then

$$\begin{aligned}
\mathbb{E}_x\|S\|^2 &= \mathbb{E}_x\big(S^\top S\big) \\
&= \mathbb{E}_x \operatorname{tr}(SS^\top) \\
&= \mathbb{E}_x \operatorname{tr}\big(SS^\top \mathbb{E}_v[vv^\top]\big) \\
&= \mathbb{E}_x\mathbb{E}_v \operatorname{tr}\big(SS^\top vv^\top\big) \\
&= \mathbb{E}_x\mathbb{E}_v \operatorname{tr}\big(v^\top SS^\top v\big) \\
&= \mathbb{E}_x\mathbb{E}_v \|v^\top S\|^2.
\end{aligned} \tag{2}$$

Substituting (2) into (1), we get

$$L_{\text{SSM}}(\theta) = \mathbb{E}_{x\sim p(x)}\mathbb{E}_{v\sim p(v)}\Big[\|v^\top S(x;\theta)\|^2 + 2v^\top \nabla_x\big(v^\top S(x;\theta)\big)\Big].$$

Thus, the form matches the required expression.

2. Brief Explanation of SDE

A **stochastic differential equation (SDE)** describes the evolution of a random process influenced by both deterministic and stochastic components. The general **Itô form** is

$$dX_t = f(X_t, t)\, dt + g(X_t, t)\, dW_t,$$

where

- $X_t \in \mathbb{R}^d$: state (random variable),
- $f(X_t, t)$: drift term (deterministic trend),

- $g(X_t, t)$: diffusion term (noise strength),
- W_t: standard Brownian motion.

Over a small Δt,

$$X_{t+\Delta t} - X_t \approx f(X_t, t)\,\Delta t + g(X_t, t)\sqrt{\Delta t}\,\varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I).$$

Itô vs. Stratonovich Form

Another equivalent expression is the **Stratonovich form**

$$dX_t = f^{\circ}(X_t, t)\,dt + g(X_t, t) \circ dW_t.$$

They are related by

$$f_{\text{Itô},i} = f_i^{\circ} + \tfrac{1}{2} \sum_{j,k} g_{ik} \frac{\partial g_{jk}}{\partial x_j}.$$

The Itô form is typically used in statistics and numerical simulations, while the Stratonovich form is preferred in physics due to the standard chain rule.

Fokker–Planck Equation

If $p(x, t)$ denotes the probability density of X_t, it satisfies

$$\partial_t p = -\nabla_x \cdot (fp) + \tfrac{1}{2} \sum_{i,j} \partial_{x_i} \partial_{x_j} \Big([gg^{\top}]_{ij} p \Big).$$

This PDE links the stochastic path dynamics to the evolution of the probability density.

Common Examples

1. **Langevin Dynamics:**

$$dX_t = \nabla_x \log \pi(X_t)\,dt + \sqrt{2}\,dW_t.$$

2. **Ornstein–Uhlenbeck Process:**

$$dX_t = -\beta X_t\,dt + \sigma\,dW_t.$$

3. **Geometric Brownian Motion:**

$$dX_t = \mu X_t\,dt + \sigma X_t\,dW_t.$$

Numerical Solution: Euler–Maruyama

The simplest numerical approximation for Itô SDEs is

$$x_{k+1} = x_k + f(x_k, t_k)\,\Delta t + g(x_k, t_k)\sqrt{\Delta t}\,\varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}(0, I).$$

It has strong convergence order 1/2 and weak convergence order 1.

Connection to Diffusion Models

In score-based diffusion models, a **forward SDE** is used to gradually add noise to data:

$$dX_t = f(X_t, t)\,dt + g(t)\,dW_t,$$

which transforms data distribution $p_0(x)$ into nearly Gaussian $p_T(x)$. For instance,

- **VP SDE:** $f = -\frac{1}{2}\beta(t)X_t$, $g = \sqrt{\beta(t)}$;
- **VE SDE:** $f = 0$, $g = \sqrt{\frac{d[\sigma^2(t)]}{dt}}$.

The corresponding **reverse-time SDE** (Anderson, 1982) is

$$dX_t = [f(X_t, t) - g(t)^2 \nabla_x \log p_t(X_t)]\,dt + g(t)\,d\bar{W}_t, \quad t \downarrow 0.$$

In practice, we replace $\nabla_x \log p_t(X_t)$ with a learned score function $s_\theta(x, t)$, then simulate backward using Euler–Maruyama or its deterministic counterpart, the **probability flow ODE**.

Summary

- General form: $dX_t = f\,dt + g\,dW_t$.
- Density evolution: Fokker–Planck equation.
- Numerical method: Euler–Maruyama.
- Diffusion models: forward SDE adds noise, reverse SDE denoises using learned score.