Week 8 written assignment: Sliced Score Matching and SDE

Kai-Hung Cheng

2025/10/28

1. Proof of the Sliced Score Matching (SSM) Loss Formulation

We start from the standard definition of the sliced score matching (SSM) loss:

$$L_{\text{SSM}} = \mathbb{E}_{x\sim p(x)}\mathbb{E}_{v\sim p(v)}\left[\frac{1}{2}\left(v^\top S(x;\theta) - v^\top \nabla_x \log p(x)\right)^2\right],$$

where $S(x;\theta)$ denotes the model score estimator, and v is a random direction vector drawn from $p(v)$ (e.g., standard Gaussian or uniform on the unit sphere).

Step 1. Expand the square term

Expanding the squared term gives

$$\frac{1}{2}(a-b)^2 = \frac{1}{2}a^2 - ab + \frac{1}{2}b^2,$$

where $a = v^\top S(x;\theta)$ and $b = v^\top \nabla_x \log p(x)$. Thus,

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

The last term depends only on the true data distribution $p(x)$ and is independent of θ; hence it can be dropped as a constant.

Step 2. Integration by parts

We now focus on the cross term:

$$T = \mathbb{E}_{x,v}\left[(v^\top S(x;\theta))(v^\top \nabla_x \log p(x))\right].$$

Fixing v, write $g_v(x) = v^\top S(x;\theta)$, then

$$T = \mathbb{E}_{x\sim p(x)}\left[g_v(x)\, v^\top \nabla_x \log p(x)\right] = \int g_v(x)\, v^\top \nabla_x p(x)\, dx.$$

Applying integration by parts (assuming $p(x)g_v(x) \to 0$ at the boundary),

$$\int g_v(x)\, v^\top \nabla_x p(x)\, dx = -\int p(x)\, v^\top \nabla_x g_v(x)\, dx.$$

Hence,

$$T = -\mathbb{E}_{x,v}\left[v^\top \nabla_x (v^\top S(x;\theta))\right].$$

Step 3. Substitute back and simplify

Plugging T back into the loss function gives

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

Multiplying by a positive constant factor (which does not change the minimizer) yields the equivalent form:

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

Step 4. Remarks

The derivation assumes:

- $S(x;\theta)$ is differentiable with respect to x.
- $p(x)$ decays sufficiently fast so that boundary terms vanish.

This form eliminates $\nabla_x \log p(x)$, allowing unsupervised training of the score network without requiring the true density.

—

2. Brief Explanation of Stochastic Differential Equation (SDE)

A **stochastic differential equation (SDE)** describes the evolution of a random process with both deterministic and stochastic components:

$$dx_t = f(x_t, t)\, dt + G(x_t, t)\, dW_t,$$

where:

- $f(x_t, t)$ is the **drift term**, determining the average direction of change.
- $G(x_t, t)$ is the **diffusion term**, scaling the random noise.
- W_t is a **Brownian motion** (or Wiener process) with $W_t - W_s \sim \mathcal{N}(0, (t-s)I)$.

The solution x_t is called an **Itô process** and satisfies the integral form:

$$x_t = x_0 + \int_0^t f(x_s, s)\, ds + \int_0^t G(x_s, s)\, dW_s.$$

SDEs are widely used in physics, finance, and diffusion models to represent systems affected by both deterministic dynamics and random fluctuations.

3. Unanswered Questions

- In what sense is Brownian motion nowhere differentiable, yet we can still define dW_t and integrate with respect to it?
- What is the intuition behind the diffusion term $G(x_t, t)\, dW_t$ in modeling real-world systems like stock prices or physical diffusion?