

HW8

112652011 廖晨鈞

The solution and ideas are based on Gemini and adapt by myself with fully comprehension.

Q1

Show that the sliced score matching (SSM) loss can also be written as

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

Solution

The core idea of SSM is to not match the entire score vectors $S(x; \theta)$ and $\nabla_x \log p(x)$, but to match their one-dimensional projections onto a set of random direction vectors v .

The explicit loss is the expected squared difference between these projections:

$$L_{SSM,explicit}(\theta) = \mathbb{E}_{x \sim p(x)} \mathbb{E}_{v \sim p(v)} [(v^T S(x; \theta) - v^T \nabla_x \log p(x))^2]$$

1. Expanding the squared term

For simplicity, let's omit the function arguments $(x; \theta)$

$$\begin{aligned} L_{SSM,explicit} &= \mathbb{E}_{x,v} [(v^T S - v^T \nabla_x \log p(x))^2] \\ &= \mathbb{E}_{x,v} [(v^T S)^2 - 2(v^T S)(v^T \nabla_x \log p(x)) + (v^T \nabla_x \log p(x))^2] \end{aligned}$$

Using the linearity of expectation, we can break this into three terms:

$$L_{SSM,explicit} = \underbrace{\mathbb{E}_{x,v} [(v^T S)^2]}_{\text{Term 1}} - \underbrace{2 \mathbb{E}_{x,v} [(v^T S)(v^T \nabla_x \log p(x))]}_{\text{Term 2}} + \underbrace{\mathbb{E}_{x,v} [(v^T \nabla_x \log p(x))^2]}_{\text{Term 3}}$$

Now let's analyze each term:

- **Term 1** is already present in our target expression. We don't need to change it.
- **Term 2** is the one we need to rewrite.
- **Term 3** does not depend on the parameter θ . Therefore, it is a constant C.

2. Rewriting the second term

First, let's fix a random vector v and only consider the expectation over x :

$$\mathbb{E}_{x \sim p(x)} [(v^T S)(v^T \nabla_x \log p(x))]$$

We use the identity $\nabla_x \log p(x) = \frac{\nabla_x p(x)}{p(x)}$:

$$= \int_{\mathbb{R}^d} p(x) \left((v^T S(x)) \left(v^T \frac{\nabla_x p(x)}{p(x)} \right) \right) dx = \int_{\mathbb{R}^d} (v^T S(x)) (v^T \nabla_x p(x)) dx$$

Since $v^T S$ is a scalar, let the scalar function be $h(x) = v^T S(x)$. The integral is $\int h(x) (v^T \nabla_x p(x)) dx$. Let's write out the dot product $v^T \nabla_x p(x) = \sum_i v_i \frac{\partial p(x)}{\partial x_i}$.

$$= \int h(x) \left(\sum_i v_i \frac{\partial p(x)}{\partial x_i} \right) dx = \sum_i v_i \int h(x) \frac{\partial p(x)}{\partial x_i} dx$$

2.1 Integration by parts

Now, we focus on the integral inside the summation, which is a one-dimensional integral with respect to the variable x_i . We apply the standard integration by parts formula:

$$\int u dv = uv - \int v du$$

For the integral $\int h(x) \frac{\partial p(x)}{\partial x_i} dx_i$, we set:

- $u = h(x)$
- $dv = \frac{\partial p(x)}{\partial x_i} dx_i$

This gives us:

- $du = \frac{\partial h(x)}{\partial x_i} dx_i$
- $v = \int \frac{\partial p(x)}{\partial x_i} dx_i = p(x)$

Plugging these into the formula, we get:

$$\int_{-\infty}^{+\infty} h(x) \frac{\partial p(x)}{\partial x_i} dx_i = \underbrace{[h(x)p(x)]_{x_i=-\infty}^{x_i=+\infty}}_{\text{Boundary Term}} - \int_{-\infty}^{+\infty} p(x) \frac{\partial h(x)}{\partial x_i} dx_i$$

The crucial step is to analyze the Boundary Term. A fundamental property of any well-behaved probability density function $p(x)$ defined over \mathbb{R}^d is that it must vanish at infinity. That is:

$$\lim_{\|x\| \rightarrow \infty} p(x) = 0$$

We also assume that $S(x)$ (and $h(x)$) does not grow faster than $p(x)$ decays. Thus, the product $h(x)p(x)$ goes to zero at the boundaries:

$$\lim_{x_i \rightarrow \pm\infty} (h(x)p(x)) = 0$$

Therefore, the boundary term is $0 - 0 = 0$ and then vanishes. This simplifies our result to:

$$\int h(x) \frac{\partial p(x)}{\partial x_i} dx_i = - \int p(x) \frac{\partial h(x)}{\partial x_i} dx_i$$

Substituting this back into the sum:

$$= \sum_i v_i \left(- \int p(x) \frac{\partial h(x)}{\partial x_i} dx \right) = - \int p(x) \left(\sum_i v_i \frac{\partial h(x)}{\partial x_i} \right) dx$$

The term $\sum_i v_i \frac{\partial h(x)}{\partial x_i}$ is exactly the dot product $v^T \nabla_x h(x)$.

$$= - \int p(x) (v^T \nabla_x h(x)) dx = - \mathbb{E}_{x \sim p(x)} [v^T \nabla_x h(x)]$$

Now, substitute $h(x) = v^T S(x)$ back in:

$$\mathbb{E}_{x \sim p(x)} [(v^T S)(v^T \nabla_x \log p(x))] = - \mathbb{E}_{x \sim p(x)} [v^T \nabla_x (v^T S(x))]$$

So we rewrite the term 2 into:

$$\text{New Term 2} = -2 \mathbb{E}_{x \sim p(x)} [v^T \nabla_x (v^T S(x))]$$

3. Putting it all together

Let's go back to our expanded loss function:

$$\begin{aligned} L_{SSM} &= \text{Term 1} + \text{New Term 2} + \text{Term 3} \\ &= \mathbb{E}_{x,v} [(v^T S)^2] - 2 (-\mathbb{E}_{x,v} [v^T \nabla_x (v^T S)]) + C \\ &= \mathbb{E}_{x,v} [(v^T S(x; \theta))^2 + 2v^T \nabla_x (v^T S(x; \theta))] + C \end{aligned}$$

Since minimizing $L_{SSM}(\theta) + C$ with respect to θ is equivalent to minimizing $L_{SSM}(\theta)$, we can omit the constant C . Thus, we get the desired result.

Q2

Briefly explain SDE.

Solution

Stochastic differential equation (SDE)

An SDE models a system that evolves over time under both deterministic drift and random diffusion.

- **SDE General Form:**

$$dx_t = \underbrace{f(x_t, t) dt}_{\text{Drift}} + \underbrace{G(x_t, t) dW_t}_{\text{Diffusion}}$$

- * x_t : The state of the system at time t .
- * $f(x_t, t)$: The **drift** coefficient, a vector describing the deterministic trend of the process.
- * $G(x_t, t)$: The **diffusion** coefficient, a matrix defining the magnitude of the random noise.
- * W_t : A standard **Wiener process** (or Brownian motion).

- **Integral Form (Ito Integral):**

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

The second integral, $\int G dW_s$, is an Ito integral, defined as a limit of sums over random increments of W_t .

Wiener Process (Brownian Motion) W_t

A continuous stochastic process modeling random motion.

1. Start at origin: $W_0 = 0$.
2. Increments: Has independent and stationary Gaussian increments, where $W_{t+\Delta t} - W_t \sim N(0, \Delta t I)$.
3. Paths: Sample paths are continuous but nowhere differentiable with probability 1.

Numerical Simulation: Euler-Maruyama Method

SDEs are typically solved numerically by discretizing time into steps of size Δt .

- **Update Rule:**

$$X_{n+1} = X_n + f(X_n, t_n) \Delta t + G(X_n, t_n) \sqrt{\Delta t} Z_n$$

where Z_n are independent random variables drawn from a standard normal distribution, $N(0, I)$.

Key Examples

- **Brownian Motion with Drift:**

$$dx_t = \mu dt + \sigma dW_t$$

Solution $x(t) \sim N(x_0 + \mu t, \sigma^2 t)$. It follows a linear trend with added random noise.

- **Ornstein-Uhlenbeck (OU) process:**

$$dx_t = -\beta x_t dt + \sigma dW_t$$

A **mean-reverting** process. The drift term $-\beta x_t$ acts as a restoring force, pulling the process back towards its mean (in this case, 0).