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

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

①

Let $S: \mathbb{R}^d \to \mathbb{R}^d$ be a continuously differentiable vector field (the score network output)
Define the Implicit Score Matching (ISM) loss as: $\mathcal{L}_{ISM}(\theta) = E_{x\sim p(x)}[\|S(x;\theta)\|^2 + 2\nabla_x \cdot S(x;\theta)]$
Let $v \in \mathbb{R}^d$ be a random vector independent of x, satisfying $E[v]=0$, $E[vv^T]\, I_d$.
Then, $\mathcal{L}_{ISM}(\theta) = E_{x\sim p(x)} E_v[\|v^T S(x;\theta)\|^2 + 2v^T\nabla_x(v^T S(x;\theta))] = \mathcal{L}_{SSM}(\theta)$

<pf.>

1° Using Hutchinson's Trace Identity: ①
For any matrix $A \in \mathbb{R}^{d\times d}$, we have $E_v[v^T A v] = E_v[tr(v^T A v)] = E_v[tr(A vv^T)] = tr(A)$
This identity holds whenever $E_v[vv^T] = I_d$, such as $v \sim N(0, I_d)$

2° By ① with $A = S(x)S(x)^T$,
then, $\|S(x)\|^2 = tr(S(x)S(x)^T) = E_v[v^T S(x)S(x)^T v] = E_v\|S(x)^T v\|^2 = E_v\|v^T S(x)\|^2$ — ②

3° Let $J_S(x) = \nabla_x S(x) \in \mathbb{R}^{d\times d}$ denote the Jacobian matrix of S, then,
$\nabla_x(v^T S(x)) = J_S(x)^T v$, hence, $v^T\nabla_x(v^T S(x)) = v^T J_S(x)^T v$
Taking expectation and applying ① again: $E_v[v^T J_S(x)^T v] = tr(J_S(x)^T) = tr(J_S(x)) = \nabla_x \cdot S(x)$ ③

4° Substituting ② & ③ into ISM definition, and apply Fubini's theorem:

$$\mathcal{L}_{ISM}(\theta) = E_x[E_v\|v^T S(x;\theta)\|^2 + 2E_v v^T\nabla_x(v^T S(x;\theta))] = E_x E_v[\|v^T S(x;\theta)\|^2 + 2v^T\nabla_x(v^T S(x;\theta))]$$

This matches the definition of the sliced score matching loss:
$\mathcal{L}_{SSM}(\theta) = E_{x\sim p(x)} E_{v\sim p(v)}[\|v^T S(x;\theta)\|^2 + 2v^T\nabla_x(v^T S(x;\theta))]$. ▨

2. Briefly explain SDE.

②

An SDE (stochastic differential equation) describes the continuous-time evolution of a random process influenced by both deterministic drift and random noise:
$dx_t = f(x_t,t)dt + g(t)dW_t$, where ① $f(x_t,t)$ is the drift term ② $g(t)dW_t$ is the diffusion term ③ W_t is a Wiener process (Brownian motion).
Intuitively, the SDE says that the system moves deterministically according to f, but also experience continuous random perturbations determined by g.

3. Unanswered Questions

There are unanswered questions from the lecture, and there are likely more questions we haven't covered.

- Take a moment to think about these questions.
- Write down the ones you find important, confusing, or interesting.
- You do **not** need to answer them—just state them clearly.

Q: ① How many random direction v are typically needed in practice for SSM to be accurate?

② Are there better distributions for v than Gaussian or Rademacher?