1.

The original objective func. (loss) is defined as $L_{SSM} = \mathbb{E}_{x\sim p(x)}\mathbb{E}_{v\sim p(v)}[\frac{1}{2}\|v^T\nabla_x \log p(x) - v^T S(x;\theta)\|^2]$.

Since $v^T\nabla_x \log p(x)$ and $v^T S(x;\theta)$ are both scalars, the squared norm $\|\cdot\|^2$ is simply the scalar squared.

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

We expand the squared term:

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

$$= \underbrace{\mathbb{E}_{x,v}[\frac{1}{2}(v^T\nabla_x \log p(x))^2]}_{\text{constant}} - \underbrace{\mathbb{E}_{x,v}[(v^T\nabla_x \log p(x))(v^T S(x;\theta))]}_{\text{transform this term}} + \underbrace{\mathbb{E}_{x,v}[\frac{1}{2}(v^T S(x;\theta))^2]}_{\text{computable}}$$

Let $f(x) = v^T S(x;\theta)$

$$\Rightarrow \mathbb{E}_{x,v}[(v^T\nabla_x \log p(x))(v^T S(x;\theta))] = \mathbb{E}_{x,v}[(v^T\nabla_x \log p(x)) f(x)] = \int p(x)\left(v^T \frac{\nabla_x p(x)}{p(x)}\right) f(x)\, dx$$

$$= \int f(x)(v^T\nabla_x p(x))\, dx$$

$$v^T\nabla_x(f(x)p(x)) = (v^T\nabla_x f(x))p(x) + f(x)(v^T\nabla_x p(x))$$

$$\Rightarrow \int v^T\nabla_x(f(x)p(x))\, dx = \int (v^T\nabla_x f(x))p(x)\, dx + \int f(x)(v^T\nabla_x p(x))\, dx$$

By the divergence thm, the left hand side is a boundary integral.

Under the standard assumption for score matching, we assume $p(x)\to 0$ (and $p(x)f(x)\to 0$) as x approaches the boundary (infinity). Thus, the boundary term vanishes (equals 0.)

$$0 = \int (v^T\nabla_x f(x))\, p(x)\, dx + \int f(x)(v^T\nabla_x p(x))\, dx$$

$$\int f(x)(v^T\nabla_x p(x))\, dx = -\int p(x)(v^T\nabla_x f(x))\, dx$$

$\mathbb{E}_x[(v^T\nabla_x \log p(x)) f(x)] = -\mathbb{E}_x[v^T\nabla_x f(x)]$ subst. $f(x) = v^T S(x;\theta)$ back in,

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

Minimizing L_{SSM} is equivalent to minimizing. Let this new objective be $J(\theta)$:

$$J(\theta) = \mathbb{E}_{x,v}[\frac{1}{2}(v^T S(x;\theta))^2] - \mathbb{E}_{x,v}[(v^T\nabla_x \log p(x))(v^T S(x;\theta))]$$

$$\Rightarrow J(\theta) = \mathbb{E}_{x,v}[\frac{1}{2}(v^T S(x;\theta))^2] - (-\mathbb{E}_{x,v}[v^T\nabla_x(v^T S(x;\theta))])$$

$$= \mathbb{E}_{x,v}[\frac{1}{2}(v^T S(x;\theta))^2 + v^T\nabla_x(v^T S(x;\theta))]$$

The loss expression given in the question is:

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

Since $v^T S(x;\theta)$ is a scalar, $\|v^T S(x;\theta)\|^2 = (v^T S(x;\theta))^2$.

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

$$\Rightarrow L_{SSM} = 2J(\theta)$$

Since the L_{SSM} in the question is simply 2 times our derived tractable loss $J(\theta)$, and $J(\theta)$ is equivalent to the original loss L_{SSM}, minimizing L_{SSM} is equivalent to minimizing the original L_{SSM}.

Conclusion : We have shown that the original Sliced Score Matching loss L_{SSM} is equivalent to minimizing $J(\theta)$. The expression L_{SSM} given in the question is a positive constant multiple of $J(\theta)$, and is therefore an equivalent and tractable objective for Sliced Score Matching.

2. Briefly explain SDE.

A Stochastic Differential Equation (SDE) is a differential equation that describes the evolution of a stochastic process, meaning it involves both deterministic and random components.

The general form of an SDE is written as : $dx_t = f(x_t, t)\,dt + G(x_t, t)\,dW_t$

This equation consists of two main parts :

① Drift term : $f(x_t, t)\,dt$

- This is the deterministic part of the process.
- It describes the average, or expected, direction and speed of the process's evolution.
- If $G(x_t, t) = 0$, the SDE reduces to a deterministic ordinary differential eqn. (ODE).

② Diffusion term : $G(x_t, t)\,dW_t$

- This is the stochastic (random) part of the process.
- It describes the random fluctuations or "diffusion" driven by noise.
- The func. $G(x_t, t)$ scales the magnitude of the random noise.
- dW_t is the increment of a Wiener process (or Brownian motion), which is the source of randomness.

The key property of the Wiener process (W_t) is that it has stationary and independent. Gaussian increments. This means the change over any time interval u, $W_{t+u} - W_t$, follows a zero-mean Gaussian distribution $N(0, uI)$.

The solution to an SDE, x_t, is itself a stochastic process, typically represented in its integral form (the Ito integral eqn.) : $x_t = x_0 + \int_0^t f(x_s, s)\,ds + \int_0^t G(x_s, s)\,dW_s$

SDEs are powerful tools for modeling systems that evolve over time subject to both a deterministic trend and random volatility. They are widely used in pathysis, finance, and for defining diffusion-based generative models in machine learning.

3. Question:

The Ito integral $\int G(X_s, s)\, dW_s$ is defined as a limit of a sum, where the term $G(X(t_k), t_k)[W(t_{k+1}) - W(t_k)]$ uses the left endpoint t_k of the time interval $[t_k, t_{k+1}]$ to evaluate the func. G.

- Is the choice unique?
- Why is the left endpoint specifically chosen?

  Does this choice lead to a different set of calculus rules for SDEs?