1. Goal: $\mathcal{L}_{SSM}(\theta) = E_{x\sim p(x)} E_{v\sim p(v)} [\, \|v^T s(x;\theta)\|^2 + 2v^T \nabla_x (v^T s(x;\theta))]$

Proof:

Since $\mathcal{L}_{ISM} = E_{x\sim p(x)} [\, \|s(x;\theta)\|^2 + 2\nabla_x \cdot s(x;\theta)]$

first term:

$$v^T s = \sum_{k=1}^{n} v_k s_k$$

$$\Rightarrow E_{v\sim p(v)} [\, \|v^T s\|^2] = E_{v\sim p(v)} [\, v^T (s s^T) v]$$

$= E_{v\sim p(v)} [\, \mathrm{trace}(v v^T s s^T)]$ by assigment 5 problem 2

$$= \mathrm{trace}(s s^T)$$

$$= s_1^2 + s_2^2 + \dots + s_n^2$$

$$= \|s\|^2$$

Thus, $\|s(x;\theta)\|^2 = E_{v\sim p(v)} [\, \|v^T s(x;\theta)\|^2]$

Second term:

$$\nabla_x \cdot s(x;\theta) = tr(\nabla_x s(x;\theta))$$

$= E_{v\sim p(v)} [v^T \nabla_x s(x;\theta) v]$ by Hutchinson's trace estimator, $v \in \mathbb{R}^d$ be a random vector s.t. $E_{v\sim p(v)}[v v^T] = I$

$= E_{v\sim p(v)} [v^T \nabla_x (v^T s(x;\theta))]$ by (*)

Thus, $2\nabla \cdot s(x;\theta) = E_{v\sim p(v)} [v^T \nabla_x (v^T s(x;\theta))]$

Therefore, $\mathcal{L}_{SSM}(\theta) = E_{x\sim p(x)} E_{v\sim p(v)} [\, \|v^T s(x;\theta)\|^2 + 2v^T \nabla_x (v^T s(x;\theta))]$

2. Briefly explain SDE:

Stochastic differential equation:

$$dx_t = \underbrace{f(x_t, t)}_{\text{drift}} dt + \underbrace{G(x_t, t)}_{\text{diffusion}} dW_t, \quad x(0) = x_0$$

where $x_t \in \mathbb{R}^d$, $f \in \mathbb{R}^d$ and $G \in \mathbb{R}^{d\times d}$

x_t is the stochastic process

W_t is a standard Brownian motion

It describes the evolution of a system over time, where the system is influenced by random noise or uncertainty.

It models how a variable changes over time with both deterministic and random components

(*): $\nabla_x s(x;\theta) v = \nabla_x (v^T s(x;\theta))$

proof:

$$\nabla_x s = \begin{bmatrix} \frac{\partial s_1}{\partial x_1} & \cdots & \frac{\partial s_n}{\partial x_1} \\ \vdots & & \\ \frac{\partial s_1}{\partial x_n} & \cdots & \frac{\partial s_n}{\partial x_n} \end{bmatrix}$$

$$\Rightarrow \nabla_x s\, v = \begin{bmatrix} \sum_{k=1}^{n} \frac{\partial s_k}{\partial x_1} v_k \\ \vdots \\ \sum_{k=1}^{n} \frac{\partial s_k}{\partial x_n} v_k \end{bmatrix}$$

$$v^T s(x;\theta) = \sum_{k=1}^{n} v_k s_k$$

$$\Rightarrow \nabla_x (v^T s(x;\theta)) = \begin{bmatrix} \sum_{k=1}^{n} \frac{\partial s_k}{\partial x_1} v_k \\ \vdots \\ \sum_{k=1}^{n} \frac{\partial s_k}{\partial x_n} v_k \end{bmatrix}$$

Thus, $(\nabla_x s(x;\theta)) v = \nabla_x (v^T s(x;\theta))$