1. 證明

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

Start from the ISM loss

$$L_{\text{ISM}}(\theta) = \mathbb{E}_{x\sim p(x)} \ [\| \ S(x;\theta) \ \|^2 + 2 \ \nabla_x \ \cdot \ S(x;\theta)],$$

and use the matrix form of the divergence:

$$\nabla_x \ \cdot \ S(x;\theta) = \text{tr}\ (\nabla_x S(x;\theta)),$$

where $\nabla_x S$ is the Jacobian matrix with entries $(\nabla_x S)_{ij} = \frac{\partial S_i}{\partial x_j}$.

Let $v \in \mathbb{R}^d$ be a random vector with
$\mathbb{E}[vv^T] = I$(e.g. $v \sim \mathcal{N}(0, I)$or Rademacher entries ± 1iid). By Hutchinson's trace estimator,

$$\text{tr}\ (\nabla_x S) = \mathbb{E}_v[v^T(\nabla_x S)\ v].$$

Also note that for the scalar $\phi(x) \equiv v^T S(x;\theta)$,

$$\nabla_x \phi(x) = (\nabla_x S(x;\theta))^T v,$$

so

$$v^T \nabla_x \phi(x) = v^T (\nabla_x S)^T v = \sum_{i,j} v_i v_j \frac{\partial S_i}{\partial x_j} = v^T (\nabla_x S)\ v.$$

Thus

$$\text{tr}\ (\nabla_x S) = \mathbb{E}_v[\ v^T \nabla_x (v^T S)\ ].$$

Finally, express the squared norm $\| S \|^2$also as an expectation over v:

$$\mathbb{E}_v[(v^T S)^2] = S^T \mathbb{E}_v[v v^T] S = S^T I S = \| S \|^2.$$

Putting these together,

$$\begin{aligned} L_{\mathrm{ISM}}(\theta) &= \mathbb{E}_x[\| S(x;\theta) \|^2 + 2\ \mathrm{tr}\ (\nabla_x S(x;\theta))] \\ &= \mathbb{E}_x[\mathbb{E}_v[(v^T S(x;\theta))^2] + 2\ \mathbb{E}_v[v^T \nabla_x (v^T S(x;\theta))]] \\ &= \mathbb{E}_{x,v}[(v^T S(x;\theta))^2 + 2\ v^T \nabla_x (v^T S(x;\theta))]. \end{aligned}$$

This last expression is exactly the sliced score matching (SSM) loss:

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

2.SDE 簡單說明

一、SDE 的基本形式

隨機微分方程是用來描述**受隨機擾動影響的動態系統**。它的一般形式為：

$$dx_t = f(x_t, t)\,dt + G(x_t, t)\,dW_t, x(0) = x_0,$$

其中：

- $x_t \in \mathbb{R}^d$：系統狀態（未知的隨機過程）；
- $f(x_t, t)$：**漂移項（drift）**，描述平均變化趨勢；
- $G(x_t, t)$：**擴散項（diffusion）**，控制隨機擾動強度；
- W_t：**布朗運動（Wiener process）**，代表隨機噪音；
- dW_t是「隨機微分」，具有統計性質：

$$E[dW_t] = 0, Var(dW_t) = dt.$$

二、等價的積分形式（Itô 積分）

SDE 可以寫成 Itô 積分方程：

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

第二項是「確定性積分」，第三項是「隨機積分（Itô integral）」，它處理布朗運動的隨機變化。

三、直觀理解

- **漂移項** $f(x_t, t)\, dt$：描述系統的平均運動（類似牛頓運動方程）。
- **擴散項** $G(x_t, t)\, dW_t$：描述系統的隨機擾動，讓軌跡出現隨機性。

因此，SDE 可以視為「普通微分方程 + 隨機噪音」。

四、例子

1. **純擴散過程**：

$$dx_t = \sigma\, dW_t \Rightarrow x_t = x_0 + \sigma W_t,$$

表示隨機遊走，$x_t \sim \mathcal{N}(x_0, \sigma^2 t)$。

2. **漂移 + 擴散（隨機游走）**：

$$dx_t = \mu\, dt + \sigma\, dW_t \Rightarrow x_t = x_0 + \mu t + \sigma W_t,$$

表示平均以速度 μ移動，同時伴隨隨機擾動。

3. **Ornstein–Uhlenbeck（OU）過程**：

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

表示具有「回復趨勢」的隨機過程，是許多物理與金融模型的基礎。

五、數值模擬（Euler–Maruyama 方法）

因為 SDE 通常無法解析求解，我們用數值方法模擬：

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

其中 $\Delta W_n \sim \mathcal{N}(0, \Delta t)$。

六、在生成模型中的應用（如 Score-based SDE）

在 **Score-based generative modeling (Song et al., 2021)** 中，

- **Forward SDE**：把資料分布 $p_0(x)$逐漸擴散成高斯噪音；
- **Reverse SDE**：根據分布的梯度（score function）逆推資料分布；
- 利用神經網路估計 $\nabla_x \log\, p_t(x)$，即可從噪音「生成」真實資料。

總結

概念	直覺	例子
漂移項 $f(\mathrm{x}_{\mathrm{t}}, t)$	平均運動方向	OU 過程的 $-\beta x_{\mathrm{t}}$
擴散項 $G(x_{\mathrm{t}}, t)$	噪音強度	常數 σ
布朗運動 W_{t}	隨機擾動來源	高斯隨機增量
模擬方法	Euler–Maruyama	離散近似
應用	Score-based generative model	DDPM、SDE-based diffusion

3.問題

1. 對任意 SDE，反向時間 SDE 存在且唯一解的充分條件是什麼？有哪些技術細節（例如正則性、邊界條件）需要注意？
2. 推導反向 SDE 的嚴格假設有哪些？在擴散係數依賴狀態（非齊次矩陣）時，公式如何改寫？