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

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

1. 目標是最小化模型預測的對數機率密度梯度（分數）$S(x;\theta)$ 與真實數據分佈 $p(x)$ 的分數 $\nabla_x logp(x)$ 之間的差距。目標函數為：

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

   因為 $\nabla_x logp(x)$ 未知所以無法直接計算。

2. 透過將分數向量投影到隨機的單位向量 v 上，將高維度的向量匹配問題轉化為一維標量的匹配問題，從而簡化計算。SSM 的目標函數定義為：

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

3. 展開平方項：

   $$(v^T S - v^T\nabla_x \log p)^2 = (v^T S)^2 - 2(v^T S)(v^T\nabla_x \log p) + (v^T\nabla_x \log p)^2$$

   對 x 取期望後，目標函數變為（忽略常數 和對 v 的期望）：

   $$\mathbb{E}_x[(v^T S)^2] - 2\mathbb{E}_x[(v^T S)(v^T\nabla_x \log p)] + \mathbb{E}_x[(v^T\nabla_x \log p)^2]$$

   - 第三項 $E_x[(v^T\nabla_x logp)^2]$ 是數據真實分佈的 Fisher divergence，它與模型參數 θ 無關，因此可忽略。
   - 第一項 $(v^T S)^2$ 是一個標量的平方，等價於其 L2 範數的平方 $||v^T S||^2$。

   所以需要再處理：

   $$L(\theta) \propto \mathbb{E}_x[\|v^T S(x;\theta)\|^2] - 2\mathbb{E}_x[(v^T S(x;\theta))(v^T\nabla_x \log p(x))]$$

4. 使用Integration by Parts

   因為 $h(x)$ 和機率密度 $p(x)$（在邊界處 $p(x)h(x) = 0$），有：

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

   讓$h(x) = (v^T S(x;\theta))v$。因為$h(x)$ 是一個向量函數。 所以 $h(x)^T\nabla_x \log p(x) = ((v^T S)v)^T\nabla_x \log p(x) = (v^T S)(v^T\nabla_x \log p(x))$。

   因此我們可以得到：

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

   因為 v 是一個常數向量，$\nabla_x^T((v^T S)v) = v^T\nabla_x(v^T S)$。 所以，

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

5. 將結果代回 $L(\theta)$：

$$L(\theta) \propto \mathbb{E}_x[\|v^T S(x;\theta)\|^2] - 2(-\mathbb{E}_x[v^T \nabla_x (v^T S(x;\theta))])$$

$$L(\theta) \propto \mathbb{E}_x[\|v^T S(x;\theta)\|^2 + 2v^T \nabla_x (v^T S(x;\theta))]$$

最後，將對隨機向量 v 的期望加回來，就得到了最終的表達式：

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

2. Briefly explain SDE

*(Stochastic Differential Equation, SDE) 是一種特殊的微分方程，它的方程中包含至少一個隨機過程*項（通常是白噪音的積分）。因此，SDE 的解本身也是一個隨機過程，而不像普通微分方程 (ODE) 的解是一個確定的函數。

一般形式：

一個 SDE 通常可以寫成如下形式：

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t$$

這裡的各項代表：

- X_t：表示一個隨機過程在時間 t 的狀態。
- dX_t：表示狀態 X_t 在一個極小的時間 dt 內的變化量。
- $f(X_t, t)dt$：稱為**Drift Term**。它描述了 X_t 隨時間演變的確定性、可預測的平均趨勢。
- $g(X_t, t)dW_t$：稱為**Diffusion Term**。它描述了 X_t 演變過程中的隨機波動或不確定性。
  - $g(X_t, t)$ 決定了隨機波動的幅度。
  - dW_t 是一個維納過程（或布朗運動）的無窮小增量，可以理解為一個服從正態分佈的隨機噪音項。

SDE 的目標是將一個連續變化的過程分解為兩部分：一部分是**確定性的平均行為（漂移）**，另一部分是**不可預測的隨機擾動（擴散）**。

3.Unanswered Questions

在 SSM 中，會將score投影到一個隨機向量 v 上。p(v) 除了球面均勻分佈，能否使用其他分佈？不同的選擇會有顯著影響模型的訓練效率和性能？