Sliced Score Matching (SSM) 損失函數的等價形式證明

1. 初始定義 (Starting Definition)

Sliced Score Matching 的目標是最小化模型分數 $s_m(x;\theta)$ 與真實資料分數 $s_d(x)$ 在隨機方向向量 **v** 上投影的平方差期望值。其損失函數 $L(\theta;p_v)$ 被定義為 Fisher divergence 的一種形式：

$$L(\theta;p_v) \triangleq \frac{1}{2}\mathbb{E}_{p_v}\mathbb{E}_{p_d}[(v^T s_m(x;\theta) - v^T s_d(x))^2]$$

其中：

- $p_d(x)$ 是真實的資料分佈。
- $s_m(x;\theta) = \nabla_x \log p_m(x;\theta)$ 是模型的 score function。
- $s_d(x) = \nabla_x \log p_d(x)$ 是資料的 score function。
- **v** 是從分佈 p_v 中抽樣的隨機投影向量。

這個初始定義中包含了未知的資料分數 $s_d(x)$，使其無法直接計算。

2. 展開損失函數 (Expansion of the Loss Function)

首先，將平方項展開：

$$L(\theta;p_v) = \frac{1}{2}\mathbb{E}_{p_v}\mathbb{E}_{p_d}[(v^T s_m(x;\theta))^2 - 2(v^T s_m(x;\theta))(v^T s_d(x)) + (v^T s_d(x))^2]$$

利用期望的線性性質，可以將其拆分為三項：

$$L(\theta;p_v) = \frac{1}{2}\mathbb{E}_{p_v}\mathbb{E}_{p_d}[(v^T s_m)^2] - \mathbb{E}_{p_v}\mathbb{E}_{p_d}[(v^T s_m)(v^T s_d)] + \frac{1}{2}\mathbb{E}_{p_v}\mathbb{E}_{p_d}[(v^T s_d)^2]$$

由於最後一項 $\frac{1}{2}\mathbb{E}_{p_v}\mathbb{E}_{p_d}[(v^T s_d)^2]$ 與模型參數 θ 無關，因此在最佳化過程中可視為一個常數 C。於是，與最佳化相關的部分簡化為：

$$L(\theta;p_v) = \mathbb{E}_{p_v}\mathbb{E}_{p_d}\left[\frac{1}{2}(v^T s_m(x;\theta))^2 - (v^T s_m(x;\theta))(v^T s_d(x))\right] + C$$

3. 透過分部積分法消除 $s_d(x)$ (Eliminating $s_d(x)$ via Integration by Parts)

此步驟的關鍵是消除對 $s_d(x)$ 的依賴。專注於處理交叉項 $-\mathbb{E}_{p_v}\mathbb{E}_{p_d}[(v^T s_m)(v^T s_d)]$。

1. 將期望寫成積分形式，並利用 $s_d(x) = \nabla_x \log p_d(x) = \frac{\nabla_x p_d(x)}{p_d(x)}$ 的關係：

$$-\mathbb{E}_{p_v}\mathbb{E}_{p_d}[(v^T s_m)(v^T s_d)] = -\mathbb{E}_{p_v}\int p_d(x)(v^T s_m(x;\theta))(v^T s_d(x))dx$$

$$= -\mathbb{E}_{p_v}\int (v^T s_m(x;\theta))(v^T \nabla_x p_d(x))dx$$

2. 在滿足特定邊界條件（即 $\lim_{||x||\to\infty} s_m(x;\theta)p_d(x) = 0$）的假設下，可以對上式使用多變量分部積分法 (multivariate integration by parts)。這使得交叉項可以轉換為：

$$-\mathbb{E}_{p_v}\int (v^T s_m)(v^T \nabla_x p_d(x))dx = \mathbb{E}_{p_v}\int p_d(x)v^T(\nabla_x s_m(x;\theta))v dx$$

3. 將結果寫回期望的形式：

$$\mathbb{E}_{p_v}\mathbb{E}_{p_d}[v^T \nabla_x s_m(x;\theta)v]$$

4. 最終形式 (Final Form)

將分部積分的結果代換回第 2 步的表達式中，得到一個不再依賴於 $s_d(x)$ 的新目標函數，稱之為 $J(\theta;p_v)$：

$$L(\theta;p_v) = \mathbb{E}_{p_v}\mathbb{E}_{p_d}\left[v^T \nabla_x s_m(x;\theta)v + \frac{1}{2}(v^T s_m(x;\theta))^2\right] + C = J(\theta;p_v) + C$$

最小化 $L(\theta;p_v)$ 等價於最小化 $J(\theta;p_v)$。

等價形式的證明：

$$L_{SSM} = \mathbb{E}_{x\sim p(x)}\mathbb{E}_{v\sim p(v)}[\|v^T S(x;\theta)\|_2^2 + 2v^T \nabla_x(v^T S(x;\theta))]$$

解析這個表達式：

- 由於 $v^T S(x;\theta)$ 是一個純量 (scalar)，其 L2 範數的平方 $\|v^T S(x;\theta)\|_2^2$ 就等於 $(v^T S(x;\theta))^2$。
- 第二項 $v^T \nabla_x(v^T S(x;\theta))$ 是計算 Hessian-vector product $v^T(\nabla_x S(x;\theta))v$ 的一種有效方式，這在【Sliced Score Matching: A Scalable Approach to Density and Score Estimation】的 Algorithm 1 中有所體現。

因此，可以將表達式可以寫成：

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

恰好為上面推導出的 $J(\theta;p_v)$ 的兩倍：

$$2\times J(\theta;p_v) = 2\times\mathbb{E}_{p_v}\mathbb{E}_{p_d}\left[\frac{1}{2}(v^T s_m)^2 + v^T \nabla_x s_m v\right] = \mathbb{E}_{p_v}\mathbb{E}_{p_d}\left[(v^T s_m)^2 + 2v^T \nabla_x s_m v\right]$$

結論

原始的 SSM 損失 $L(\theta; p_v)$ 在最佳化上等價於 $J(\theta; p_v)$。HW8中的表達式是 $2 \times J(\theta; p_v)$。由於乘以一個正的常數 2 並不會改變最佳解的位置，因此它也是一個用於 Sliced Score Matching 的有效損失函數。

Briefly explain SDE.

什麼是隨機微分方程 (SDE)？

想像一下，描述一個物體的運動，比如一片葉子在空中飄落。在一個理想、沒有風的環境下，可以用一個普通的微分方程 (ODE) 來精確描述它的路徑，只考慮重力。這就像一個完全可以預測的系統。

然而，現實世界充滿了隨機性。葉子會被陣風隨機地吹動，它的路徑變得不再那麼確定。**隨機微分方程 (SDE)** 就是用來描述這種**包含隨機變化**的系統的數學工具。

簡單來說，SDE 是一種微分方程，它不僅包含像重力這樣可預測的、確定的部分，還額外加入了一個**隨機的部分**來模擬不可預測的波動或“噪聲”。

SDE 的兩個核心部分

一個典型的 SDE 可以表達成：

$$dx_t = \underbrace{f(x_t, t)}_{\text{漂移 (Drift)}} \, dt + \underbrace{G(x_t, t)}_{\text{擴散 (Diffusion)}} \, dW_t$$

其中：

1. **漂移項 (Drift):** $f(x_t, t)\,dt$
    - 這部分代表了系統的**確定性趨勢**或平均行為。
    - 可以想像成河流的主流方向。就算河面上有各種波紋，河水整體的趨勢還是朝著下游流動。
    - 在葉子的例子中，漂移項就代表了重力，使葉子總體上朝著地面飄落。
2. **擴散項 (Diffusion):** $G(x_t, t)\,dW_t$
    - 這部分代表了系統的**隨機波動**或噪聲。
    - 可以想像成隨機吹來的陣風，它的大小和方向都是不確定的，會讓葉子的實際路徑偏離其平均下落軌跡。
    - 其中，dW_t 代表**維納過程** (Wiener Process) 或**布朗運動** (Brownian Motion)，這是描述隨機遊走最核心的數學模型。$G(x_t, t)$ 則決定了這個隨機波動的強度。

總結

	普通微分方程 (ODE)	**隨機微分方程 (SDE)**
描述對象	可預測、確定的系統	包含隨機性、不確定的系統
例子	理想環境下的拋物線運動	股票價格的波動、花粉在水中的運動
組成	只有漂移 (Drift) 項	**漂移 (Drift) 項** + **擴散 (Diffusion) 項**

總而言之，SDE 透過結合一個**可預測的趨勢**和一個**不可預測的隨機波動**，能夠更真實地模擬和分析金融市場、物理學、生物學等領域中各種動態變化的複雜系統。

Unanswered Questions

請問隨機微分方程 (SDE) 在機器學習領域中有哪些具體的應用？例如，它如何與近年來備受關注的生成模型 (Generative Models) 或擴散模型 (Diffusion Models) 產生關聯？