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

由定義 $L_{SSM}(\theta) = \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\|S(\mathbf{x};\theta)\|^2 + \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\mathbb{E}_{\mathbf{v}\sim p(\mathbf{v})}\left[2\mathbf{v}^\top\nabla_{\mathbf{x}}(\mathbf{v}^\top S(\mathbf{x};\theta))\right].$ ，我們只需要證明 $\mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\mathbb{E}_{\mathbf{v}\sim p(\mathbf{v})}[\|\mathbf{v}^\top S\|^2] = \mathbb{E}_{\mathbf{x}}[\|S\|^2]$。

$$
\begin{aligned}
\mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\mathbb{E}_{\mathbf{v}\sim p(\mathbf{v})}[\|\mathbf{v}^\top S\|^2] &= \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\mathbb{E}_{\mathbf{v}\sim p(\mathbf{v})}[(\mathbf{v}^\top S)^2] \\
&= \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\mathbb{E}_{\mathbf{v}\sim p(\mathbf{v})}[(S^\top \mathbf{v})(\mathbf{v}^\top S)] \\
&= \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\mathbb{E}_{\mathbf{v}\sim p(\mathbf{v})}[S^\top(\mathbf{v}\mathbf{v}^\top)S] \\
&= \mathbb{E}_{\mathbf{x}}[S^\top\left(\mathbb{E}_{\mathbf{v}}[\mathbf{v}\mathbf{v}^\top]\right)S] && (S \text{ 與 } \mathbf{v} \text{ 無關}) \\
&= \mathbb{E}_{\mathbf{x}}[S^\top I S] && (\text{因為 } \mathbb{E}_{\mathbf{v}}[\mathbf{v}\mathbf{v}^\top] = I) \\
&= \mathbb{E}_{\mathbf{x}}[S^\top S] \\
&= \mathbb{E}_{\mathbf{x}}[\|S\|^2]
\end{aligned}
$$

2 什麼是隨機微分方程 (SDE)？

SDE 是一種數學方程式，用來描述一個隨時間演變的隨機過程 x_t 。它的一般形式（Ito 過程）如下：

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

這個方程式由兩個關鍵部分組成：

漂移項 **(Drift Term):** $f(x_t,t)dt$

這代表系統的確定性趨勢或平均走向。$f(x_t,t)$ 是一個函數，描述在時間 t、狀態為 x_t 時，系統的預期變化率。dt 代表一個極小的時間變化。如果 SDE 只有這一項 ($dx_t = f(x_t,t)dt$)，它就退化為一個普通的確定性微分方程。

擴散項 **(Diffusion Term):** $G(x_t,t)dW_t$

這代表系統的隨機波動或噪聲。$G(x_t,t)$ 是一個函數（可能是矩陣），描述隨機性的規模或強度。dW_t 是最關鍵的部分，它代表 Wiener 過程或布朗運動 (Brownian motion) W_t 的微小變化。

2.1 隨機性的來源：Wiener 過程 (W_t)

驅動 SDE 隨機性的 W_t 是一個具有以下數學特性的隨機過程：

- 初始條件: $W_0 = 0$ 。
- 獨立增量: 在任何不重疊的時間區間內，W_t 的變化（增量）是相互獨立的。

- 平穩高斯增量: 任意一段時間 Δt 內的增量 $\Delta W_t = W(t+\Delta t) - W(t)$，服從平均值為 0、協方差為 ΔtI 的高斯分佈（常態分佈）。即 $W_{t+\Delta t} - W_t \sim N(0, \Delta tI)$ 。
- 連續路徑: W_t 的路徑 $t \mapsto W(t)$ 是連續的，但它幾乎在任何地方都是不可微分的。

2.2 SDE 的積分形式

SDE $dx_t = f(x_t, t)dt + G(x_t, t)dW_t$ 也可以寫成等價的伊藤積分方程：

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

x_0 是初始狀態。$\int_0^t f(x_s, s)ds$ 是一般的積分，代表漂移項隨時間的累積。$\int_0^t G(x_s, s)dW_s$ 是伊藤隨機積分，代表擴散項隨時間的累積。

2.3 如何（近似）求解 SDE？

SDE 很少有精確的解析解。在實務上，我們使用數值方法來近似模擬其路徑，最常用的是 Euler-Maruyama 法。這個方法將時間 t 分割成 N 個寬度為 Δt 的小區間，然後使用以下迭代公式來計算 X_t 的近似值 X_n（其中 $t_n = n\Delta t$）:

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

在電腦模擬中，我們利用 $\Delta W(t_n) \sim N(0, \Delta tI)$ 的特性，將其改寫為：

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

其中 $Z(t_n)$ 是一系列獨立同分佈 (iid) 且服從標準常態分佈 $N(0, 1)$ 的隨機變數。

3 Unanswered Questions

如何衡量 Euler-Maruyama 這個數值方法的好壞?