Sliced Score Matching (SSM) 等價形式與 Hutchinson’s Trick

令資料分佈 $p(x)$，模型的 score 為

$$S(x;\theta) = \nabla_x \log q_\theta(x).$$

1. SSM 等價形式（到常數為止）

考慮 sliced Fisher divergence（丟掉與 θ 無關常數），並且引入一個投影矩陣$v, v \in \mathbb{R}^{d\times 1}$投影矩陣$v$服從某一分布$p_v$(可以為高斯分布或其它分布)，即$v \sim p_v$，$x$和$v$是互相獨立的。:

：

$$\tilde{L}_{\mathrm{SSM}}(\theta) = \mathbb{E}_{x\sim p}\,\mathbb{E}_{v\sim p(v)}\left[\frac{1}{2}\left(v^\top\big(S(x;\theta) - s_p(x)\big)\right)^2\right].$$

展開並用分部積分（Stein 恒等式；假設邊界項為 0）：

$$\mathbb{E}_x\left[(v^\top S)(v^\top s_p)\right] = -\mathbb{E}_x\left[v^\top \nabla_x (v^\top S)\right].$$

因此（到常數與正比例因子為止）：

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

2. 與 Hutchinson’s Trick 的連結

Hutchinson’s trick：若隨機向量 v 的分佈滿足

$$\mathbb{E}_v[vv^\top] = I,$$

則對任何方陣 A,

$$\mathrm{tr}(A) = \mathbb{E}_v\left[v^\top A v\right].$$

常見選擇：Rademacher（每一維 ± 1 均機率）或標準常態 $v \sim \mathcal{N}(0, I)$。

在原始（非 sliced）score matching 目標中會出現

$$\operatorname{tr}\big(\nabla_x S(x;\theta)\big) = \sum_{i=1}^{d} \frac{\partial S_i(x;\theta)}{\partial x_i} = \nabla_x \cdot S(x;\theta).$$

用 Hutchinson's trick 可把上述**昂貴的散度/跡**改寫為

$$\operatorname{tr}\big(\nabla_x S(x;\theta)\big) = \mathbb{E}_v\left[v^\top \nabla_x S(x;\theta)\, v\right].$$

注意到

$$v^\top \nabla_x S(x;\theta)\, v = v^\top \nabla_x\big(v^\top S(x;\theta)\big),$$

因此「跡項」可被**一次一個方向**的方向導數所近似。這正是 SSM 中
$v^\top \nabla_x\big(v^\top S(x;\theta)\big)$ 的來源：它把
$\operatorname{tr}(\nabla_x S)$ 的計算變成對隨機方向 v 的期望，避免了全梯度散度的顯式求和。

3. 條件與備註

- 需要 $S(x;\theta)$ 與 $p(x)$ 足夠光滑，且邊界項消失（例如 $p(x)$ 在無窮遠足夠快地衰減）。
- $p(v)$ 只需滿足 $\mathbb{E}[vv^\top] = I$，Rademacher 與 Gaussian 均可。
- 自動微分下，$v^\top \nabla_x(v^\top S)$ 可用「先對 x 做方向導數再內積」實作，避免整個 Jacobian。

SDE 是什麼？

「**規律趨勢 + 隨機亂流**」的連續時間模型。常見寫法：

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

- x_t：系統狀態
- **drift** f：決定「往哪裡走」
- **diffusion** G：決定「抖多大」
- W_t：布朗運動（持續亂晃的隨機來源）

布朗運動 W_t

- 路徑連續但到處不平順
- 增量獨立，且每段增量 $\sim \mathcal{N}(0,\ \Delta t)$
- 直觀：每個瞬間都在微小亂晃，但整體有統計規律

白噪音 vs. 布朗運動

- 可把「白噪音」理解成布朗運動的**形式導數**
- 在 SDE 裡代表「每個時刻互不相關、平均 0 的擾動」

怎麼讀 SDE 裡的隨機積分？

$$\int G\,dW \quad \approx \quad \sum G(\text{當下}) \times \underbrace{\Delta W}_{\sim\mathcal{N}(0,\,\Delta t)}$$

把時間切很細，每小段抓一個常態增量，累加後取極限。

數值模擬（Euler–Maruyama）

最基本離散化：

$$X_{n+1} \;=\; X_n \;+\; f(X_n,t_n)\,\Delta t \;+\; G(X_n,t_n)\,\sqrt{\Delta t}\,Z_n, \qquad Z_n \sim \mathcal{N}(0,1)$$

像歐拉法在每一步再**加一口常態雜訊**。

直觀範例

- **純擾動**：$dx_t = \sigma\,dW_t$
  - 平均不變，但不確定性隨時間增大（變異數 $= \sigma^2 t$）。
- **常數趨勢 + 擾動**：$dx_t = \mu\,dt + \sigma\,dW_t$
  - 一路往前推進（速度 μ），同時持續抖動。
- **均值回復（OU 過程）**：$dx_t = -\beta x_t\,dt + \sigma\,dW_t$
  - 偏離 0 就被拉回，但一路都有噪音干擾；常用於有「回到均衡」傾向的動態。

一句話總結

SDE 把「可預測的趨勢」與「不可預測的隨機波動」合在一起描述系統的時間演化