

# Assignment 8

314652021羅瑋翔

## Written Assignment

Q1.

The screenshot shows a presentation slide titled "Sliced score matching" under the heading "Flexible models". The slide contains the following content:

- **Intuition:** one dimensional problems should be easier
- **Idea:** project onto random directions
- **Randomized objective: Sliced Fisher Divergence**

$$\frac{1}{2} \mathbb{E}_{p_{\mathbf{v}}} \mathbb{E}_{p_{\text{data}}(\mathbf{x})} [(\mathbf{v}^T \nabla_{\mathbf{x}} \log p_{\text{data}}(\mathbf{x}) - \mathbf{v}^T \mathbf{s}_{\theta}(\mathbf{x}))^2]$$

- Integration by parts  $\rightarrow$  **Sliced Score Matching:**

$$\mathbb{E}_{p_{\mathbf{v}}} \mathbb{E}_{p_{\text{data}}(\mathbf{x})} \left[ \mathbf{v}^T \nabla_{\mathbf{x}} \mathbf{s}_{\theta}(\mathbf{x}) \mathbf{v} + \frac{1}{2} (\mathbf{v}^T \mathbf{s}_{\theta}(\mathbf{x}))^2 \right]$$

On the right side of the slide, there is a video feed of a person wearing headphones, labeled "Yang S". Below the video feed is the logo for the "CENTER FOR Brains Minds+ Machines". Further down, the date "December 12, 2022" is displayed. At the bottom, the text "Diffusion and Score-Based Generative Models" is shown, followed by the name "Yang Song" and "Stanford University".

By sliced Fisher Divergence:

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

First,

$$\mathbb{E}_{x \sim p(x)} [(v^T S(x; \theta)) (v^T \nabla_x \log p(x))] = \int (v^T S(x; \theta)) (v^T \nabla_x \log p(x)) p(x) dx$$

And,

$$\nabla_x \log p(x) = \nabla_x (\ln p(x)) = \frac{1}{p(x)} \nabla_x p(x)$$

we replace  $\nabla_x \log p(x)$  with  $\frac{\nabla_x p(x)}{p(x)}$ .

Then we have:

$$\mathbb{E}_{x \sim p(x)}[(v^T S(x; \theta))(v^T \nabla_x \log p(x))] = \int (v^T S(x; \theta))(v^T \nabla_x p(x)) dx$$

The following is integration by part:

$$\int (v^T S(x; \theta))(v^T \nabla_x p(x)) dx = - \int p(x)(v^T \nabla_x (v^T S(x; \theta))) dx$$

So, we will have:

$$\begin{aligned} & \mathbb{E}_{x \sim p(x)} \mathbb{E}_{v \sim p(v)} [(v^T \nabla_x \log p(x) - v^T S(x; \theta))^2] \\ &= \mathbb{E}_{x \sim p(x)} \mathbb{E}_{v \sim p(v)} [(v^T S(x; \theta))^2 - 2(v^T S(x; \theta))(v^T \nabla_x \log p(x)) + (v^T \nabla_x \log p(x))^2] \\ &= \mathbb{E}_{x \sim p(x)} \mathbb{E}_{v \sim p(v)} [\|v^T S(x; \theta)\|^2 + 2v^T \nabla_x (v^T S(x; \theta))] \end{aligned}$$

Since the last term  $(v^T \nabla_x \log p(x))^2$  is constant, we omit it while optimizing.  
Then we have this result.

## Q2.

A stochastic differential equation is an ODE contaminated through a stochastic term. And it is of the form:

$$dX_t = f(X_t, t) + G(X_t, t)dW_t$$

where  $f(X_t, t)$  is the drift(deterministic trend) and  $G(X_t, t)dW_t$  is the diffusion term driven by Brownian motion  $W_t$ .

The main method of solution is to find the PDF as a function of time using the equivalent Fokker-Planck equation(FPE). Since FPE is a deterministic PDE. It tells how the PDF evolves in time.

## Q3.

I cannot understand what is p mean in:

$$\frac{dp}{dt} = \frac{\sigma^2}{2} P_{xx}$$

Is it the probability at time t?

I don't know how to implement Fokker-Planck equation to solve SDE. What is the relationship between them?

And can we use FPE to solve the SDEs with different kinds of different PDFs?