# ML HW8

**Date:** 10/22/2025
**Week:** 8
**Author:** Alvin B. Lin
**Student ID:** 112652040

**Problem 1: Form Transformation of SSM**

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

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

**Solution:**

Given the sliced score matching loss is defined as (from the note)

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

Meaning that we only need to show

$$\mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\mathbb{E}_{\mathbf{v}\sim p(\mathbf{v})}\left\|\mathbf{v}^\top S(\mathbf{x};\theta)\right\|^2 = \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\left\|S(\mathbf{x};\theta)\right\|^2$$

*Proof.* First since $\mathbf{x}$ is independent of $\mathbf{v}$, thus the iterated expectation can be written as

$$\mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\mathbb{E}_{\mathbf{v}\sim p(\mathbf{v})}\left\|\mathbf{v}^\top S(\mathbf{x};\theta)\right\|^2 = \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\left(\mathbb{E}_{\mathbf{v}\sim p(\mathbf{v})}\left\|\mathbf{v}^\top S(\mathbf{x};\theta)\right\|^2\right) \tag{1}$$

**Lemma 1**

Given $\mathbf{v} \sim \mathcal{N}(\mathbf{0},\ I)$, we have

$$\mathbb{E}_{\mathbf{v}\sim p(\mathbf{v})}\left(\mathbf{v}\mathbf{v}^\top\right) = I$$

*Proof.* We know that if $\mathbf{v} \sim \mathcal{N}(\mu, \mathbf{\Sigma})$, then

$$\mathbb{E}_{\mathbf{v}\sim p(\mathbf{v})}\left((\mathbf{v}-\mu)(\mathbf{v}-\mu)^\top\right) = \mathbf{\Sigma}$$

Take $\mu = \mathbf{0}$ and $\mathbf{\Sigma} = I$, we get

$$\mathbb{E}_{\mathbf{v}\sim p(\mathbf{v})}\left(\mathbf{v}\mathbf{v}^\top\right) = I$$

□

Equation (1) can also be written as

$$\mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\left(S(\mathbf{x};\theta)^\top\left(\mathbb{E}_{\mathbf{v}\sim p(\mathbf{v})}\mathbf{v}\mathbf{v}^\top\right)S(\mathbf{x};\theta)\right) = \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\left(S(\mathbf{x};\theta)^\top S(\mathbf{x};\theta)\right) = \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\left\|S(\mathbf{x};\theta)\right\|^2,$$

which completes the proof. □

**Problem 2: Understanding of SDE**

Briefly explain SDE.

### From Time-Series to Start

A better way to understand the **SDE** is from **time-series**, a modeling way of time-related discrete data. We first look at an example of a time-series case, the stock price of Apple's from 2018 to 2024 (Figure 1).

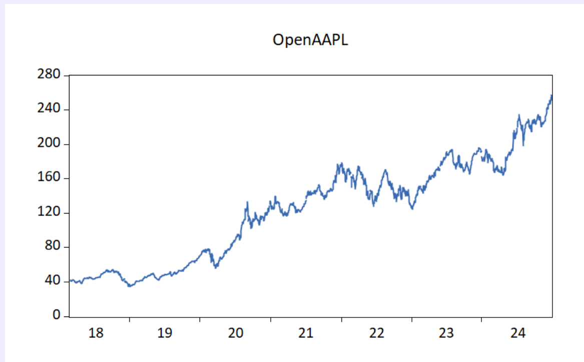


**Figure 1:** Stock price of Apple Inc (AAPL) from 2018 to 2024 (Data source: Yahoo Finance)

We can see from the graph that there is a trend that the price is rising with a linear relation, with a volatility/diffusion term. One can easily model the price $X(t)$ by AR(1), auto-regressing model,

$$X(t) = \mu + X(t-1) + \sigma\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,1), \quad X(0) = x_0$$

Can also be arranged to:

$$\Delta X(t) = \mu\Delta t + \sigma\Delta\varepsilon_t$$

If we take $\Delta t \to 0$, we get:

$$\mathrm{d}X(t) = \mu\mathrm{d}t + \sigma\mathrm{d}W_t, \quad X(0) = x_0 \tag{2}$$

which is the form of **SDE**, $W_t$ be the Brownian motion/Wiener process. (**Note:** Taking $\Delta t \to 0$ here is only intuitive, but not rigorous.)

### Drifting and Diffusion/Volatility

If we integrate equation (2) from 0 to $T$, then:

$$X(T) = X(0) + \mu T + \sigma W(T)$$

We can understand the formula by: $X(0)$ be the stock price at time 0 and $\mu$ be the expected return and $\sigma W(T)$ be the volatility of the stock price. Here we call such $\mu$ as drifting term and $\sigma$ as volatility term or diffusion term.

Continuing the stock price for Apple case, if we set 1/1/2020 as $t = 0$, and use a simple SDE (equation 2) to model, we get the result (Figure 2), the volatility is proportional to $\sqrt{T}$; the actual stock price path can be seen as a possibility of the Stochastic process.

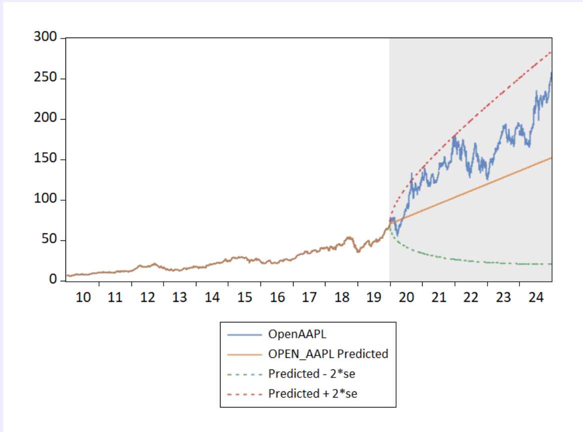


**Figure 2:** Prediction of the stock price by history data

### More on SDE

With the help of **Itô calculus**, many things that have randomness involved can be modelled by SDE, and have some similar properties to the general calculus. One major use is the **Black-Scholes model**, a model that can estimate the option in the market.

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

The equation is very analogous to the ordinary partial differential equation, but with some special rules that:

$$(\mathrm{d}W_t)^2 = \mathrm{d}t; \quad \mathrm{d}W_t \cdot \mathrm{d}t = (\mathrm{d}W_t)^n = 0, \quad n \geq 3$$