Understanding Score Matching and Its Role in Score-Based Generative Models

Chia Yu Ou

October 2025

1 Introduction

Generative models aim to learn the underlying data distribution $p(x)$ from which observed data are sampled. Unlike discriminative models that focus on conditional distributions $p(y|x)$, generative models try to capture the full structure of the input space. However, directly modeling $p(x)$ is challenging due to the presence of the intractable normalizing constant in energy-based models. To overcome this, score-based methods learn the *score function*—the gradient of the log-density, denoted as $\nabla_x \log p(x)$—instead of the density itself. This technique is known as **score matching**.

2 The Concept of Score Matching

The idea of score matching, introduced by Hyvärinen (2005), is to fit a parameterized model $S(x;\theta)$ to the true score $\nabla_x \log p(x)$ by minimizing the expected squared difference:

$$L_{\text{ESM}}(\theta) = \mathbb{E}_{p(x)}\left[\|S(x;\theta) - \nabla_x \log p(x)\|^2\right]$$

However, since the true score $\nabla_x \log p(x)$ is unknown, the above objective is impractical. By integrating by parts and assuming appropriate boundary conditions, the loss can be rewritten as:

$$L_{\text{ISM}}(\theta) = \mathbb{E}_{p(x)}\left[\|S(x;\theta)\|^2 + 2\nabla_x \cdot S(x;\theta)\right]$$

This is known as **implicit score matching (ISM)** and avoids explicit access to $\nabla_x \log p(x)$.

3 Denoising Score Matching (DSM)

To make score matching applicable in practice, particularly for high-dimensional data, **Denoising Score Matching (DSM)** is introduced. DSM assumes access to clean data $x_0 \sim p_0(x_0)$, and generates noisy observations $x \sim p(x|x_0)$ by adding Gaussian noise:

$$x = x_0 + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I)$$

The conditional distribution $p(x|x_0)$ is known, so its score can be computed:

$$\nabla_x \log p(x|x_0) = -\frac{1}{\sigma^2}(x - x_0)$$

The DSM loss becomes:

$$L_{\text{DSM}}(\theta) = \mathbb{E}_{x_0 \sim p_0(x_0)} \mathbb{E}_{x \sim p(x|x_0)} \left\| S_\sigma(x;\theta) + \frac{1}{\sigma^2}(x - x_0) \right\|^2$$

Substituting $x = x_0 + \sigma\epsilon$, the loss simplifies to:

$$L_{\text{DSM}}(\theta) = \mathbb{E}_{x_0,\epsilon} \left\| \sigma S_\sigma(x_0 + \sigma\epsilon;\theta) + \epsilon \right\|^2$$

This form is widely used in modern score-based generative models.

4 Application in Diffusion Generative Models

Score matching plays a critical role in score-based diffusion models, also known as **score-based generative models**. These models consist of two stages:

1. **Forward process:** Gradually add Gaussian noise to the data over time, transforming the data distribution $p_0(x)$ into an isotropic Gaussian.
2. **Reverse process:** Use the learned score function $\nabla_x \log p_t(x)$ to iteratively denoise the data, sampling from complex data distributions by reversing the noise process.

The score function at each noise level is trained using DSM. When we train the model to learn the score function $S(x;\theta) \approx \nabla_x \log p(x)$, the resulting vector field tells us the direction in which each noisy point should move to become more likely under the data distribution. This allows us to start from pure Gaussian noise and, by following the score directions step by step, gradually generate clean data samples.

5 Conclusion

Score matching provides a powerful alternative to likelihood-based training in generative models, particularly when the density normalization constant is intractable. Denoising Score Matching offers a practical implementation by leveraging known conditional distributions. Its integration into diffusion models enables high-quality sample generation and has become a cornerstone of modern generative modeling.