Notes for Flow Matching

Jingxuan Zhang
Department of Computer Science and Engineering
East China University of Science and Technology
Shanghai, China
y21220033@mail.ecust.edu.cn

1 Theory

According to the Flow Matching paper [1], we need to find a path between the noise distribution $p_0(x_0)$, such as Gaussian, to the data distribution $p_1(x_1)$. To go through the whole path, one possible way is to interpolate the two distributions linearly:

$$x_t = (1-t)x_0 + tx_1, \quad t \in [0,1]. \tag{1}$$

Here $x_0 \sim p_0(x_0)$ and $x_1 \sim p_1(x_1)$. Then we can derive the velocity field $v_t(x_t)$ as:

$$v_t(x_t) = \frac{dx_t}{dt} = x_1 - x_0. \tag{2}$$

In the original paper, we use ψ_t to model x and u_t to model the derivative, such as $x_t = \psi_t(x)$ and $\frac{d}{dt}x_t = \frac{d}{dt}\psi_t(x) = u_t(\psi_t(x))$. We construct a neural network $v_t(x_t)$ to approximate the velocity field $u_t(x_t)$. The loss function is defined as:

$$L = \mathbb{E}_{x_t \sim p_t(x_t)}\left[\left\|v_t\left(x_t\right) - u_t\left(x_t\right)\right\|^2\right]. \tag{3}$$

However, we do not know how to sample from $p_t(x_t)$ nor the ground truth $u_t(x_t)$. To solve this problem, we can represent the loss function as:

$$\begin{aligned} L &= \mathbb{E}_{x_t \sim p_t(x_t)}\left[\left\|v_t\left(x_t\right) - u_t\left(x_t\right)\right\|^2\right] \\ &= \mathbb{E}_{x_t \sim p_t(x_t)}\left[\left\|v_t\left(x_t\right)\right\|^2 - 2 \cdot v_t\left(x_t\right) \cdot u_t\left(x_t\right) + \left\|u_t\left(x_t\right)\right\|^2\right] \end{aligned} \tag{4}$$

The middle term can be transformed using the expection notation:

$$\begin{aligned} &\mathbb{E}_{x_t \sim p_t(x)}\left[2 \cdot v_t\left(x_t\right) \cdot u_t\left(x_t\right)\right] = 2 \cdot \int v_t\left(x_t\right) \cdot u_t\left(x_t\right) \cdot p_t\left(x_t\right) \mathrm{d}x_t \\ =&2 \cdot \int v_t\left(x_t\right) \cdot \frac{\int u_t\left(x_t \mid x_1\right) p_t\left(x_t \mid x_1\right) q\left(x_1\right) \mathrm{d}x_1}{p_t\left(x_t\right)} \cdot p_t\left(x_t\right) \mathrm{d}x_t \\ =&2 \cdot \iint v_t\left(x_t\right) \cdot u_t\left(x_t \mid x_1\right) \cdot p_t\left(x_t \mid x_1\right) q\left(x_1\right) \mathrm{d}x_1 \mathrm{d}x_t \\ =&2 \cdot \mathbb{E}_{x_1 \sim q(x_1), x_t \sim p_t(x_t \mid x_1)}\left[v_t\left(x_t\right) \cdot u_t\left(x_t \mid x_1\right)\right] \end{aligned}. \tag{5}$$

After we plug in the above equation, we have:

$$\begin{aligned} L =& \mathbb{E}_{x_t \sim p_t(x_t)}\left[\left\|v_t\left(x_t\right)\right\|^2\right] - 2 \cdot \mathbb{E}_{x_1 \sim q(x_1), x_t \sim p_t(x_t \mid x_1)}\left[v_t\left(x_t\right) \cdot u_t\left(x_t \mid x_1\right)\right] \\ &+ \mathbb{E}_{x_t \sim p_t(x_t)}\left[\left\|u_t\left(x_t\right)\right\|^2\right] \\ =& \mathbb{E}_{x_t \sim p_t(x_t)}\left[\left\|v_t\left(x_t\right)\right\|^2\right] - 2 \cdot \mathbb{E}_{x_1 \sim q(x_1), x_t \sim p_t(x_t \mid x_1)}\left[v_t\left(x_t\right) \cdot u_t\left(x_t \mid x_1\right)\right] \\ &+ \mathbb{E}_{x_t \sim p_t(x_t)}\left[\left\|u_t\left(x_t\right)\right\|^2\right] + \left\|u_t\left(x_t \mid x_1\right)\right\|^2 - \left\|u_t\left(x_t \mid x_1\right)\right\|^2\right] \end{aligned} \tag{6}$$

After we swap the last two terms, we have:

$$
\begin{aligned}
L &= \mathbb{E}_{x_t \sim p_t(x_t)}\left[\left\|v_t\left(x_t\right)\right\|^2\right] - 2 \cdot \mathbb{E}_{x_1 \sim q(x_1), x_t \sim p_t(x_t \mid x_1)}\left[v_t\left(x_t\right) \cdot u_t\left(x_t \mid x_1\right)\right] \\
&\quad + \mathbb{E}_{x_t \sim p_t(x_t)}\left[\left\|u_t\left(x_t \mid x_1\right)\right\|^2\right] + \mathbb{E}_{x_t \sim p_t(x_t)}\left[\left\|u_t\left(x_t\right)\right\|^2 - \left\|u_t\left(x_t \mid x_1\right)\right\|^2\right] \\
&= \mathbb{E}_{x_1 \sim q(x_1), x_t \sim p_t(x_t \mid x_1)}\left[\left\|v_t\left(x_t\right)\right\|^2 - 2 \cdot v_t\left(x_t\right) \cdot u_t\left(x_t \mid x_1\right) + \left\|u_t\left(x_t \mid x_1\right)\right\|^2\right] \\
&\quad + \mathbb{E}_{x_t \sim p_t(x_t)}\left[\left\|u_t\left(x_t\right)\right\|^2 - \left\|u_t\left(x_t \mid x_1\right)\right\|^2\right] \\
&= \mathbb{E}_{x_1 \sim q(x_1), x_t \sim p_t(x_t \mid x_1)}\left[\left\|v_t\left(x_t\right) - u_t\left(x_t \mid x_1\right)\right\|^2\right] \\
&\quad + \mathbb{E}_{x_t \sim p_t(x_t)}\left[\left\|u_t\left(x_t\right)\right\|^2 - \left\|u_t\left(x_t \mid x_1\right)\right\|^2\right]
\end{aligned} \tag{7}
$$

The last term is a constant that does not depend on v_t. Therefore, we can minimize the following loss function instead:

$$
L = \mathbb{E}_{x_1 \sim q(x_1), x_t \sim p_t(x_t \mid x_1)}\left[\left\|v_t\left(x_t\right) - u_t\left(x_t \mid x_1\right)\right\|^2\right]. \tag{8}
$$

Suppose $\psi_t(x) = \sigma_t\left(x_1\right) x + \mu_t\left(x_1\right)$, we have $\psi_t\left(x_0\right) = \sigma_t\left(x_1\right) x_0 + \mu_t\left(x_1\right)$. Let $\mu_t\left(x_1\right) = t x_1, \quad \sigma_t\left(x_1\right) = 1 - t$, then $x_t = \psi_t\left(x_0\right) = (1-t) x_0 + t x_1$, which is the same as our linear interpolation.

Now, remember that $\frac{d}{dt} x_t = \frac{d}{dt} \psi_t(x) = u_t(\psi_t(x))$. We can add a condition to the both sides, and then we have $\frac{\mathrm{d}}{\mathrm{d}t} \psi_t(x) = u_t\left(\psi_t(x) \mid x_1\right)$ or $\frac{\mathrm{d}}{\mathrm{d}t} x_t = u_t\left(x_t \mid x_1\right)$ or $\frac{\mathrm{d}}{\mathrm{d}t} \psi_t\left(x_0\right) = u_t\left(\psi_t\left(x_0\right) \mid x_1\right)$.

Therefore, the loss function can be rewritten as:

$$
L = \mathbb{E}_{x_1 \sim q(x_1), x_0 \sim p_0(x_0)}\left[\left\|v_t\left(\psi_t\left(x_0\right)\right) - \frac{\mathrm{d}}{\mathrm{d}t} \psi_t\left(x_0\right)\right\|^2\right]. \tag{9}
$$

After calculating the derivative, we have:

$$
L = \mathbb{E}_{x_1 \sim q(x_1), x_0 \sim p_0(x_0)}\left[\left\|v_t\left((1-t) x_0 + t x_1\right) - \left(x_1 - x_0\right)\right\|^2\right]. \tag{10}
$$

This is the final loss function used in the Flow Matching paper.

References

[1] Y. Lipman, R. T. Chen, H. Ben-Hamu, M. Nickel, and M. Le, “Flow matching for generative modeling,” in *ICLR*, 2023, pp. 1–28.