Connecting Discrete and Continuous Diffusion Models

Liu Yang

Recall the Markov chain with Gaussian noise:

$$q(x_s|x_t) = \mathcal{N}(A_{s,t}x_t, B_{s,t}^2)$$

for $s > t$, we have

$$\begin{aligned} A_{r,t} &= A_{r,s}A_{s,t} \\ B_{r,t}^2 &= A_{r,s}^2 B_{s,t}^2 + B_{r,s}^2 \end{aligned}$$

for $r > s > t$. And

$$q(x_t|x_s, x_0) = \mathcal{N}(C_{s,t}x_s + D_{s,t}x_0, E_{s,t}^2)$$

for $s > t > 0$ with

$$\begin{aligned} C_{s,t} &= A_{s,t}B_{t,0}^2/B_{s,0}^2, \\ D_{s,t} &= A_{t,0}B_{s,t}^2/B_{s,0}^2, \\ E_{s,t}^2 &= B_{s,t}^2 B_{t,0}^2/B_{s,0}^2. \end{aligned}$$

Now, consider a forward diffusion process:

$$q(x_{t+\Delta t}|x_t) = \mathcal{N}(A_{t+\Delta t,t}x_t, B_{t+\Delta t,t}^2)$$

We set

$$\begin{aligned} A_{t+\Delta t,t} &= 1 - f_t \Delta t, \\ B_{t+\Delta t,t} &= g_t \sqrt{\Delta t}, \end{aligned}$$

which is consistent with the Euler scheme of Langevin equation:

$$dX_t = -f_t X_t dt + g_t dW_t$$

In DDPM, we require $A_{t+\Delta t,t}^2 + B_{t+\Delta t,t}^2 = 1$. This basically says $f_t = g_t^2/2$ as $\Delta t \to 0$.

Also recall

$$q(x_{t-\Delta t}|x_t, x_0) = \mathcal{N}(C_{t,t-\Delta t}x_t + D_{t,t-\Delta t}x_0, E_{t,t-\Delta t}^2)$$

where

$$
\begin{aligned}
C_{t,t-\Delta t} &= A_{t,t-\Delta t} B^2_{t-\Delta t,0} / B^2_{t,0}, \\
&= A_{t,t-\Delta t} \frac{B^2_{t,0} - B^2_{t,t-\Delta t}}{A^2_{t,t-\Delta t} B^2_{t,0}} \\
&= \frac{1}{A_{t,t-\Delta t}} \frac{B^2_{t,0} - B^2_{t,t-\Delta t}}{B^2_{t,0}} \\
&= \frac{1}{1 - f_{t-\Delta t}\Delta t} \left(1 - \frac{g^2_{t-\Delta t}\Delta t}{B^2_{t,0}}\right) \\
&\approx 1 + f_{t-\Delta t}\Delta t - g^2_{t-\Delta t}\Delta t / B^2_{t,0} \\
&\approx 1 + f_t \Delta t - g^2_t \Delta t / B^2_{t,0}
\end{aligned}
$$

$$
\begin{aligned}
D_{t,t-\Delta t} &= A_{t-\Delta t,0} B^2_{t,t-\Delta t} / B^2_{t,0} \\
&= A_{t-\Delta t,0} g^2_{t-\Delta t} \Delta t / B^2_{t,0} \\
&\approx A_{t,0} g^2_t \Delta t / B^2_{t,0}
\end{aligned}
$$

$$
\begin{aligned}
E^2_{t,t-\Delta t} &= B^2_{t,t-\Delta t} B^2_{t-\Delta t,0} / B^2_{t,0} \\
&\approx g^2_{t-\Delta t}\Delta t \\
&\approx g^2_t \Delta t
\end{aligned}
$$

Here we keep the first order Δt in approximation. So we have

$$
q(x_{t-\Delta t}|x_t, x_0) = \mathcal{N}(x_t + f_t x_t \Delta t + g^2_t \frac{A_{t,0}x_0 - x_t}{B^2_{t,0}} \Delta t, g^2_t \Delta t)
$$

Now we set

$$
p_\theta(x_{t-\Delta t}|x_t) = \mathcal{N}(x_t + f_t x_t \Delta t + g^2_t h^*(x_t)\Delta t, g^2_t \Delta t)
$$

In DDPM, we defined the mean squared loss:

$$
h^*(x_t) = \arg\min_{h(x_t)} E_{q(x_t,x_0)} \|h(x_t) - \frac{A_{t,0}x_0 - x_t}{B^2_{t,0}}\|^2
$$

Note that

$$
\frac{A_{t,0}x_0 - x_t}{B^2_{t,0}} = \nabla_{x_t} \log q(x_t|x_0)
$$

we recovered the score match loss in continuous diffusion.

The optimizer

$$
h^*(x_t) = \nabla_{x_t} \log q(x_t, t)
$$

so $p_\theta(x_{t-\Delta t}|x_t)$ corresponds to the reverse SODE

$$
dX_\tau = (f_t X_\tau + g^2_t \nabla_{X_\tau} \log q(X_\tau, t))d\tau + g_t dW_\tau
$$

where $\tau = T - t$.

Actually, if we use the parameterization in DDPM, then the noise ϵ is indeed

$$\epsilon = -\frac{A_{t,0}x_0 - x_t}{B_{t,0}} = -B_{t,0}\nabla_{x_t} \log q(x_t|x_0)$$

And the optimal neural network

$$\epsilon_\theta^* = -B_{t,0}\nabla_{x_t} \log q(x_t, t)$$

In other words, the prediction of ϵ is effectively an approximation of the (negative) score function scaled by the noise level.

Now we look into the reverse diffusion process. In DDPM, we have

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sigma_t z$$

In SODE perspective, suppose the forward process

$$dX_t = -f_t X_t dt + g_t dW_t$$

Then the reverse SODE is

$$dX_t = [f_t X_t + g_t^2 \nabla_x \log q(x, t)]dt + g_t dW_t$$

Clearly the diffusion parts are consistent, since $\sigma_t = \sqrt{\beta_t} = B_{t,t-\Delta t} \approx g_t\sqrt{\Delta t}$.

For the drift part $\alpha_t = A_{t,t-\Delta t}^2$, therefore

$$\frac{1}{\sqrt{\alpha_t}}x_t = \frac{1}{A_{t,t-\Delta t}}x_t \approx (1 + f_t\Delta t)x_t$$

Moreover,

$$\begin{aligned}
&\frac{1}{\sqrt{\alpha_t}}\left(-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right) \\
=&\frac{1}{\sqrt{\alpha_t}}\left(\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}B_{t,0}\nabla_{x_t} \log q(x_t, t)\right) \\
=&\frac{1-\alpha_t}{\sqrt{\alpha_t}}\nabla_{x_t} \log q(x_t, t) \\
=&\frac{\beta_t}{\sqrt{\alpha_t}}\nabla_{x_t} \log q(x_t, t) \\
\approx& g_t^2\Delta t\nabla_{x_t} \log q(x_t, t)
\end{aligned}$$

We conclude that the discrete and continuous reverse diffusion processes are consistent.