Propagating Diffusion Algebras: One Formula to All

Ziyu Wei

December 28, 2024

神给了我们改变世界的武器，蕴藏了翻云覆雨、惊天动地之力量，此之谓剑；驾驭此剑，发挥其力量，以惊天地、泣鬼神者，谓之执剑人。

目录

1 On the top

2024.12.1. Now I begin to write my deduction of the diffusion algebras. Referring mainly to [1][3][4][5][2] and blog [https://spaces.ac.cn/archives/9228]

2 God and Swordman

God said: 诞生吧！diffusion model！, and said:

$$dx = \left(f(x,t) - \frac{1}{2}(g_t^2 - \sigma_t^2)\nabla_x \log p_t(x) \right) dt + \sigma_t dw \tag{1}$$

The swordman said:

$$\begin{aligned} f(x,t) &= \frac{1}{\lambda_t}\frac{d\lambda_t}{dt}x \\ g(t)^2 &= \lambda_t^2 \frac{d}{dt}\left(\frac{\mu_t}{\lambda_t}\right)^2 \\ dw &= \sqrt{dt}\epsilon \end{aligned} \tag{2}$$

And he said: What god said is an SDE, witch depict the evolution of a random process. The item $p_t(x)$ is the Marginal distribution of x at time t. If we let the diffusion coefficient dw be a Gaussian distribution, then the every x_t(x at time t) will sampled from a Gaussian distribution, as below

$$p(x_{t+dt}|x_t) \sim N\left(x_{t+dt};\ x_t + (f(x,t) - \frac{1}{2}(g_t^2 - \sigma_t^2)\nabla_x \log p_t(x_t))dt, \sigma_t^2 dt \right) \tag{3}$$

3 SDE 的逆过程

对于任意一个 SDE，

$$dx = f(x,t))dt + \sigma_t dw \tag{4}$$

其逆过程的演化关系可以由贝叶斯公式得到

$$p(x_t|x_{t+dt}) = \frac{p(x_{t+dt}|x_t)p(x_t)}{p(x_{t+dt})} \tag{5}$$

特别地，当 dw 选取为高斯噪声，以至于每一步的 x_t 都满足高斯分布时，有

$$\begin{aligned} p(x_{t+dt}|x_t) &= N\left(x_{t+dt};\ x_t + f(x,t)dt, \sigma_t^2 dt\right) \\ &\propto exp(-\frac{(x_{t+dt} - x_t - f^(x,t)dt)^2}{2\sigma_t^2 dt}) \end{aligned} \tag{6}$$

$$p(x_t|x_{t+dt}) \propto exp\left\{ -\frac{(x_{t+dt} - x_t - f(x,t)dt)^2}{2\sigma_t^2 dt} + \log p(x_t) - \log p(x_{t+dt}) \right\} \tag{7}$$

注意到，我们当谈及概率 p 时，由于我们研究的是一个随 t 的演化过程，所以不同 t 时刻对应的 $p(x)$ 也是不一样的，如果要区分开来，则应写作 $p(t, x_t)$。所以式 (5) 严格来说应该写作

$$p(t, x_t|t + dt, x_{t+dt}) = \frac{p(t + dt, x_{t+dt}|t, x_t)p(t, x_t)}{p(t + dt, x_{t+dt})} \tag{8}$$

同样的，结合微分关系，有

$$\begin{aligned}\log p(x_{t+dt}) - \log p(x_t) &= \log p(t + dt, x_{t+dt}) - \log p(t, x_t) \\ &= \frac{\partial(\log(p(t, x_t))}{\partial t} \times dt + \frac{\partial(\log(p(t, x_t))}{\partial x} \times dx \\ &= \frac{\partial(\log(p(t, x_t))}{\partial t} \times dt + \nabla_x \log(p(t, x_t))dx \\ &= \frac{\partial(\log(p(t, x_t))}{\partial t} \times dt + \nabla_x \log(p(t, x_t))(x_{t+dt} - x_t) \\ &\sim \frac{\partial(\log(p(x_t))}{\partial t} \times dt + \nabla_x \log(p(x_t))(x_{t+dt} - x_t)\end{aligned} \tag{9}$$

带入 (7)，有

$$\begin{aligned}p(x_t|x_{t+dt}) &\propto exp\left\{-\frac{(x_{t+dt} - x_t - f(x, t)dt)^2}{2\sigma_t^2 dt} - \frac{\partial(\log(p(x_t))}{\partial t} \times dt - \nabla_x \log(p(x_t))(x_{t+dt} - x_t)\right\} \\ &= exp\left\{-\frac{1}{2\sigma_t^2 dt}\left((x_{t+dt} - x_t - f(x, t)dt)^2 + \frac{\partial(\log(p(x_t))}{\partial t}2\sigma_t^2 dt^2 + \nabla_x \log(p(x_t))(x_{t+dt} - x_t)2\sigma_t^2 dt\right)\right\} \\ &= exp\left\{-\frac{1}{2\sigma_t^2 dt}\left((x_{t+dt} - x_t - f(x, t)dt - \sigma_t^2\nabla_x \log(p(x_t)dt)^2 + C\right)\right\}\end{aligned} \tag{10}$$

上式最后一步对 $(x_{t+dt} - x_t)$ 进行了配方，且

$$C = s\sigma_t^2 f(x, t)\nabla_x \log(p(x_t)dt^2 - \sigma_t^4(\nabla_x \log(p(x_t))^2 dt^2 - s\sigma_t^2\frac{\partial(\log(p(x_t))}{\partial t}dt^2 \tag{11}$$

注意到

$$\lim_{dt\to 0} C = 0 \tag{12}$$

所以

$$\begin{aligned}p(x_t|x_{t+dt}) &\propto exp\left\{-\frac{1}{2\sigma_t^2 dt}\left((x_{t+dt} - x_t - f(x, t)dt + \sigma_t^2\nabla_x \log(p(x_t)dt)^2\right)\right\} \\ &\sim N\left(x_t;\ x_{t+dt} + (-f(x_{t+dt}, t + dt) + \sigma_{t+dt}^2\nabla_x \log(p(x_{t+dt}))dt,\ \sigma_{t+dt}^2 dt\right)\end{aligned} \tag{13}$$

所以逆过程的 SDE 为

$$dx = (f(x, t) - \sigma_t^2\nabla_x \log(p(x_{t+dt}))dt + \sigma_t dw \tag{14}$$

god 给的 SDE (1) 相当于一般 SDE (4) 中

$$f(x,t) \to f(x,t) - \frac{1}{2}(g_t^2 - \sigma_t^2)\nabla_x \log p_t(x) \tag{15}$$

他的逆过程为

$$dx = \left(f(x,t) - \frac{1}{2}(g_t^2 + \sigma_t^2)\nabla_x \log p_t(x)\right)dt + \sigma_t dw \tag{16}$$

4 Explanation of what god said: 边缘分布的一致性

4.1 Dirac δ Function

Dirac δ 函数定义为

$$\begin{gathered}\delta(x-x_0) = \begin{cases} +\infty & \text{if } x == x_0 \\ 0 & \text{if } x! = x_0 \end{cases} \\ \text{and} \int \delta(x-x_0)dx = 1\end{gathered} \tag{17}$$

如果将 $\delta(x-x_0)$ 在 $x=\xi$ 附近进行泰勒展开，则

$$\delta(x-x_0) = \delta(\xi - x_0) + \nabla_x\delta(x-x_0)|_{x=\xi}\xi + \frac{1}{2}\nabla_x^2\delta(x-x_0)|_{x=\xi}\xi^2 + \dots \tag{18}$$

假如说我们现在有一个分布 $p(x)$，则根据 δ 函数的性质

$$p(x) = \int \delta(x-y)p(y)dy = E_y(\delta(x-y)) \tag{19}$$

这个式子还好，他帮我们把一个分布转化成一个期望的形式。
同样的，还有

$$p(x)f(x) = \int \delta(x-y)p(y)f(y)dy = E_y(\delta(x-y)f(y)) \tag{20}$$

对上式两边思考，有

$$\nabla_x[p(x)f(x)] = E_y(f(y)\nabla_x\delta(x-y)) \tag{21}$$

4.2 Fokker-Plank equation

根据式 (19)，有

$$\begin{aligned} p_{x_{t+dt}}(x) &= E_{x_{t+dt}}[\delta(x - x_{t+dt})] \\ &= E_{x_t,w}[\delta(x - x_t - f(x,t)dt - \sigma_t dw)] \\ &= E_{x_t,w}[\delta(x-x_t) + (-f(x,t)dt - \sigma_t dw)\nabla_x\delta(x-x_t) + \frac{1}{2}(f(x,t)dt + \sigma_t dw)^2\nabla_x^2\delta(x-x_t) + \dots] \end{aligned} \tag{22}$$

当取 $dw=\sqrt{dt}\epsilon$，有

$$
\begin{aligned}
p_{x_{t+dt}}(x) &= E_{x_t,w}[\delta(x-x_t)+(-f(x,t)dt-\sigma_t dw)\nabla_x\delta(x-x_t)+\frac{1}{2}(f(x,t)dt+\sigma_t dw)^2\nabla_x^2\delta(x-x_t)+O(dt)] \\
&= E_{x_t,\epsilon}[\delta(x-x_t)-(f(x,t)dt+\sigma_t\sqrt{dt}\epsilon)\nabla_x\delta(x-x_t)+\frac{1}{2}\sigma_t^2 dt^2\epsilon^2\nabla_x^2\delta(x-x_t)+O(dt)] \\
&\approx E_{x_t}[\delta(x-x_t)-(f(x,t)dt)\nabla_x\delta(x-x_t)+\frac{1}{2}\sigma_t^2 dt^2\nabla_x^2\delta(x-x_t)] \\
&= p_t(x)-\nabla_x(f(x,t)\sigma_t)dt+\frac{1}{2}\sigma_t^2\nabla_x^2 p_t(x)dt
\end{aligned}
\tag{23}
$$

从而

$$
\frac{\partial}{\partial t}p_t(x)=-\nabla_x(f(x,t)p_t(x))+\frac{1}{2}\sigma_t^2\nabla_x^2 p_t(x) \tag{24}
$$

这个式子称为 F-P 方程，它描述了 SDE (4) 的边缘分布随时间 t 的变化

4.3 等价变换

我们对 F-P 方程 (24) 做一个等价变换

$$
\begin{aligned}
\frac{\partial}{\partial t}p_t(x) &= -\nabla_x(f(x,t)p_t(x))+\frac{1}{2}\sigma_t^2\nabla_x^2 p_t(x) \\
&= -\nabla_x(f(x,t)p_t(x))+\frac{1}{2}\sigma_t^2\nabla_x^2 p_t(x)-\frac{1}{2}\lambda_t^2\nabla_x^2 p_t(x)+\frac{1}{2}\lambda_t^2\nabla_x^2 p_t(x) \\
&= -\nabla_x\left(f(x,t)p_t(x)-\frac{1}{2}(\sigma_t^2-\lambda_t^2)\nabla_x p_t(x)\right)+\frac{1}{2}\lambda_t^2\nabla_x^2 p_t(x) \\
&= -\nabla_x\left(p_t(x)[f(x,t)-\frac{1}{2}(\sigma_t^2-\lambda_t^2)\nabla_x\log p_t(x)]\right)+\frac{1}{2}\lambda_t^2\nabla_x^2 p_t(x)
\end{aligned}
\tag{25}
$$

注意：式 (24) 和式 (25) 之间是完全的等价变换，也就是说，无论我们新引入的 λ_t 取何值，二者描述的边缘分布是完全相同的。通过对比式 (24) 和其对应的 SDE(4) 之间的参数可以知道，式 (25) 对应的 SDE 为

$$
dx=\left(f(x,t)-\frac{1}{2}(\sigma_t^2-\lambda_t^2)\nabla_x\log p_t(x)\right)dt+\lambda_t dw \tag{26}
$$

进行一个变量代换

$$
\begin{aligned}
\sigma_t &\to g_t \\
\lambda_t &\to \sigma_t
\end{aligned}
\tag{27}
$$

则上式可重写为

$$
dx=\left(f(x,t)-\frac{1}{2}(g_t^2-\sigma_t^2)\nabla_x\log p_t(x)\right)dt+\sigma_t dw \tag{28}
$$

这正是 god 给我们的式子!
无论 σ_t 取何值，其对应的 F-P 方程都是等价的，即边缘分布总是相同的。换句话说，这个式子描述了一族边缘分布相同的 SDE。
特别地，当取 $\sigma_t = g_t$ 时，上式为

$$dx = f(x,t)dt + g_t dw \tag{29}$$

4.4 逆过程的等价性

(28) 的逆过程

$$x_{t+dt} - x_t = dx = \left(f(x,t) - \frac{1}{2}(g_t^2 + \sigma_t^2)\nabla_x \log p_t(x) \right) dt + \sigma_t dw \tag{30}$$

这是表示逆过程的，只允许从 x_{t+dt} 到 x_t 进行变化。把逆过程视为正过程，即

$$\begin{aligned} &x_{t+dt} \to x_t \\ &x_t \to x_{t+dt} \end{aligned} \tag{31}$$

$$\nabla_x p_t(x) = \frac{p_t(x_{t+dt}) - p_t(x_t)}{x_{t+dt} - x_t} \tag{32}$$

如何证明逆过程也是一个边缘分布不变的分布族？并且其边缘分布和正过程相同？

5 Birth of Diffusion

5.1 Margin Distribution

对于式 (1) 和式 (2)，我们取

$$\begin{aligned} &\lambda_t^2 + \mu_t^2 = 1 \\ &\lambda_t^2 = \prod_{i=1}^{t} \alpha_i \equiv \bar{\alpha}_t \quad (seperate \quad condition) \end{aligned} \tag{33}$$

由于式 (1) 描述了边缘分布相同的 SDE 族，则其边缘分布和 $\sigma_t = g_t$ 的特殊情况相同。
当 $\sigma_t = g_t$

$$dx = f(x,t)dt + g_t dw \tag{34}$$

我们将 swordman 所提供的参数 (2) 带入，

$$
\begin{aligned}
dx &= \frac{1}{\lambda_t}d\lambda_t x + \sqrt{\lambda_t^2 \frac{d}{dt}\left(\frac{\mu_t}{\lambda_t}\right)^2}\sqrt{dt}\epsilon \\
\Rightarrow dx - \frac{1}{\lambda_t}d\lambda_t x &= \lambda_t\sqrt{d\left(\frac{\mu_t}{\lambda_t}\right)^2}\epsilon \\
\Rightarrow \frac{dx}{\lambda_t} - \frac{1}{\lambda_t^2}d\lambda_t x &= \sqrt{d\left(\frac{\mu_t}{\lambda_t}\right)^2}\epsilon \\
\Rightarrow d\left(\frac{x}{\lambda_t}\right) &= \sqrt{d\left(\frac{\mu_t}{\lambda_t}\right)^2}\epsilon
\end{aligned}
\tag{35}
$$

考虑 $\lambda_t^2 + \mu_t^2 = 1$ 的特殊情况

$$
\begin{aligned}
d\left(\frac{x}{\lambda_t}\right) &= \sqrt{d\left(\frac{1-\lambda_t^2}{\lambda_t^2}\right)}\epsilon \\
\Rightarrow d\left(\frac{x}{\lambda_t}\right) &= \sqrt{d\left(\frac{1}{\lambda_t^2}\right)}\epsilon
\end{aligned}
\tag{36}
$$

对于离散的情况

$$
\left(\frac{x_{t+\Delta t}}{\lambda_{t+\Delta t}} - \frac{x_t}{\lambda_t}\right) = \sqrt{\left(\frac{1}{\lambda_{t+\Delta t}}\right)^2 - \left(\frac{1}{\lambda_t}\right)^2}\epsilon \tag{37}
$$

当 $\Delta t \to 1$，取

$$
\lambda_t^2 = \prod_{i=1}^{t}\alpha_i \equiv \bar{\alpha}_t \tag{38}
$$

则

$$
\begin{aligned}
\left(\frac{x_{t+1}}{\sqrt{\bar{\alpha}_{t+1}}} - \frac{x_t}{\sqrt{\bar{\alpha}_t}}\right) &= \sqrt{\left(\frac{1}{\sqrt{\bar{\alpha}_{t+1}}}\right)^2 - \left(\frac{1}{\sqrt{\bar{\alpha}_t}}\right)^2}\epsilon \\
\Rightarrow \frac{x_{t+1}}{\sqrt{\bar{\alpha}_{t+1}}} - \frac{x_t}{\sqrt{\bar{\alpha}_t}} &= \sqrt{\frac{1}{\bar{\alpha}_{t+1}} - \frac{1}{\bar{\alpha}_t}}\epsilon \\
\Rightarrow \sqrt{\bar{\alpha}_{t+1}}\frac{x_{t+1}}{\sqrt{\bar{\alpha}_{t+1}}} - \sqrt{\bar{\alpha}_{t+1}}\frac{x_t}{\sqrt{\bar{\alpha}_t}} &= \sqrt{\bar{\alpha}_{t+1}}\sqrt{\frac{1}{\bar{\alpha}_{t+1}} - \frac{1}{\bar{\alpha}_t}}\epsilon \\
\Rightarrow x_{t+1} - \sqrt{\frac{\bar{\alpha}_{t+1}}{\bar{\alpha}_t}}x_t &= \sqrt{1 - \frac{\bar{\alpha}_{t+1}}{\bar{\alpha}_t}}\epsilon
\end{aligned}
\tag{39}
$$

由 (33), 注意到 $\bar{\alpha}_{t+1}/\bar{\alpha}_t = \alpha_{t+1}$，则

$$
\Rightarrow x_{t+1} = \sqrt{\alpha_{t+1}}x_t + \sqrt{1-\alpha_{t+1}}\epsilon \tag{40}
$$

当 $\epsilon \sim N(0,1)$，从初始确定的 x_0 演化而来的每一个 x_t 都满足正态分布

$$\begin{aligned} x_{t+1} &\sim N(x_{t+1}; \sqrt{\alpha_{t+1}}x_t, 1-\alpha_{t+1}) \\ x_{t+1} &\sim N(x_{t+1}; \sqrt{\bar{\alpha}_{t+1}}x_0, 1-\bar{\alpha}_{t+1}) \end{aligned} \tag{41}$$

综上所述，具有相同边缘分布的 SDE 族 (1)，其边缘分布等于 $\sigma_t = g_t$ 的特殊情况下 (34) 的边缘分布，且在 $\lambda_t^2 + \mu_t^2 = 1$，并通过 (33) 量化时，其边缘分布满足正态分布 (41)。
根据边缘分布 (41)，有

$$p(x_t) \propto exp\left(-\frac{(x_t - \sqrt{\bar{\alpha}_t}x_0)^2}{2(1-\bar{\alpha}_t)}\right) \tag{42}$$

则

$$\nabla_x \log p(x_t) = -\frac{(x_t - \sqrt{\bar{\alpha}_t}x_0)}{2(1-\bar{\alpha}_t)} = -\frac{\sqrt{1-\bar{\alpha}_t}\epsilon(x_t,t)}{1-\bar{\alpha}_t} \tag{43}$$

其中 $\epsilon(x_t,t)$ 为从第 0 到第 t 步的噪声。

5.2 离散化 SDE

将 SDE(1) 离散化，

$$x_{t+\Delta t} - x_t = \left(f(x,t) - \frac{1}{2}(g_t^2 - \sigma_t^2)\nabla_x \log p_t(x)\right)\Delta t + \sigma_t\sqrt{\Delta t}\epsilon \tag{44}$$

在 (33) 的条件下将 (2) 离散化

$$\begin{aligned} f(x,t) &= \frac{1}{\lambda_t}\frac{d\lambda_t}{dt}x \\ &= \frac{1}{\sqrt{\bar{\alpha}_t}}\frac{d\sqrt{\bar{\alpha}_t}}{dt}x \\ &= \frac{1}{\sqrt{\bar{\alpha}_t}}\frac{\sqrt{\bar{\alpha}_{t+\Delta t}} - \sqrt{\bar{\alpha}_t}}{\Delta t}x \\ &= \frac{\sqrt{\bar{\alpha}_{t+\Delta t}/\bar{\alpha}_t} - 1}{\Delta t}x \equiv f_t x \end{aligned} \tag{45}$$

$$\begin{aligned} g^2(t) &= \lambda_t^2\frac{d}{dt}\left(\frac{\mu_t}{\lambda_t}\right)^2 = 2\lambda_t\mu_t\frac{d}{dt}\left(\frac{\mu_t}{\lambda_t}\right) \\ &= \bar{\alpha}_t\frac{d}{dt}\left(\frac{1}{\bar{\alpha}_t}\right) = 2\bar{\alpha}_t\sqrt{1-\bar{\alpha}_t}\frac{d}{dt}\left(\frac{\sqrt{1-\bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}}\right) \\ &= \bar{\alpha}_{t+\Delta t}\left(\frac{1}{\bar{\alpha}_{t+\Delta t}} - \frac{1}{\bar{\alpha}_t}\right)\frac{1}{\Delta t} = 2\bar{\alpha}_t\sqrt{1-\bar{\alpha}_t}\left(\frac{\sqrt{1-\bar{\alpha}_{t+\Delta t}}}{\sqrt{\bar{\alpha}_{t+\Delta t}}} - \frac{\sqrt{1-\bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}}\right)\frac{1}{\Delta t} \end{aligned} \tag{46}$$

将 SDE 的逆过程 (14) 离散化，由于在 SDE 逆过程推导中，我们用了很多 $\Delta t \to 0$ 的近似，但是在离散化的时候，这些近似会导致结果的偏差，所以我们不能直接从 (14) 入手离散化，而是要从贝

叶斯公式 (5) 出发，重新推导其离散化过程。

$$
\begin{aligned}
p(x_t|x_{t+dt}) &= \frac{p(x_{t+dt}|x_t)p(x_t)}{p(x_{t+dt})} \\
&\propto exp\left(-\frac{\left(-\Delta t(f_t x_t - \frac{1}{2}(g_t^2-\sigma_t^2)\nabla_x \log p_t(x)) + x_{\Delta \mathrm{t}+t} - x_t\right)^2}{2\Delta \mathrm{t}\sigma^2} + \frac{\left(x_{\Delta \mathrm{t}+t} - x_0\sqrt{\bar{\alpha}_{\Delta \mathrm{t}+t}}\right)^2}{2\left(1-\bar{\alpha}_{\Delta \mathrm{t}+t}\right)} - \frac{\left(x_t - x_0\sqrt{\bar{\alpha}_t}\right)^2}{2\left(1-\bar{\alpha}_t\right)}\right)
\end{aligned} \tag{47}
$$

由 mathematica 计算，x_t 的二次项系数为

$$
-\frac{(f_t\Delta t+1)^2}{2\sigma^2\Delta t} - \frac{1}{2\left(1-\bar{\alpha}_t\right)} \tag{48}
$$

x_t 的一次项系数为

$$
-\frac{f_t\frac{1}{2}(g_t^2-\sigma_t^2)\nabla_x \log p_t(x)}{\sigma^2} - \frac{\frac{1}{2}(g_t^2-\sigma_t^2)\nabla_x \log p_t(x)}{\sigma^2\Delta t} + \frac{f_t x_{t+\Delta t}}{\sigma^2} + \frac{x_0\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t} + \frac{x_{\Delta \mathrm{t}+t}}{\Delta \mathrm{t}\sigma^2} \tag{49}
$$

所以

$$
\begin{aligned}
& p(x_t|x_{t+dt}) \\
\propto\ & exp\left(-\left(\frac{(f_t\Delta t+1)^2}{2\sigma^2\Delta t} + \frac{1}{2\left(1-\bar{\alpha}_t\right)}\right)\left(x_t - \frac{-\frac{f_t\frac{1}{2}(g_t^2-\sigma_t^2)\nabla_x \log p_t(x)}{\sigma^2} - \frac{\frac{1}{2}(g_t^2-\sigma_t^2)\nabla_x \log p_t(x)}{\sigma^2\Delta t} + \frac{f_t x_{t+\Delta t}}{\sigma^2} + \frac{x_0\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t} + \frac{x_{\Delta \mathrm{t}+t}}{\Delta \mathrm{t}\sigma^2}}{\frac{(f_t\Delta t+1)^2}{\sigma^2\Delta t} + \frac{1}{(1-\bar{\alpha}_t)}}\right)^2\right)
\end{aligned} \tag{50}
$$

5.3 Birth of DDPM

There be an YHC, for the equation (1), he said, "the $\nabla_x p_t(x)$ is ugly, i donot want it.", then he let $\sigma_t = g_t$, and it vanished.

$$
dx = f(x,t)dt + g_t dw \tag{51}
$$

这和 (34) 相同，在离散情况下，其描述的 x_t 之间的传递关系为 (40)

$$
x_{t+1} = \sqrt{\alpha_{t+1}}x_t + \sqrt{1-\alpha_{t+1}}\epsilon \tag{52}
$$

由 (41)，有

$$
x_{t+1} = \sqrt{\bar{\alpha}_{t+1}}x_0 + \sqrt{1-\bar{\alpha}_{t+1}}\epsilon(x_{t+1},t) \tag{53}
$$

注意，(52) 和 (53) 中的 ϵ 都是正态分布，但有所不同这正是 DDPM 的前向关系式！

对于 SDE 逆过程 (14)，使用离散化的 SDE 逆过程 (50), 并带入离散化参数 (45),(46)，有

$$
\begin{aligned}
& p(x_t|x_{t+dt}) \\
\propto\, & exp\left(-(\frac{(f_t\Delta t+1)^2}{2\sigma^2\Delta t}+\frac{1}{2\,(1-\bar{\alpha}_t)})\left(x_t-\frac{-\frac{f_t\frac{1}{2}(g_t^2-\sigma_t^2)\nabla_x\log p_t(x)}{\sigma^2}-\frac{\frac{1}{2}(g_t^2-\sigma_t^2)\nabla_x\log p_t(x)}{\sigma^2\Delta t}+\frac{f_t x_{t+\Delta t}}{\sigma^2}+\frac{x_0\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}+\frac{x_{\Delta t+t}}{\Delta t\sigma^2}}{\frac{(f_t\Delta t+1)^2}{\sigma^2\Delta t}+\frac{1}{(1-\bar{\alpha}_t)}}\right)^2\right) \\
=\, & exp\left(-\frac{1}{2}((\frac{\sqrt{\bar{\alpha}_{t+\Delta t}/\bar{\alpha}_t}-1+1)^2}{\bar{\alpha}_{t+\Delta t}\left(\frac{1}{\bar{\alpha}_{t+\Delta t}}-\frac{1}{\bar{\alpha}_t}\right)}+\frac{1}{(1-\bar{\alpha}_t)})\left(x_t-\frac{\frac{\sqrt{\bar{\alpha}_{t+\Delta t}/\bar{\alpha}_t}-1}{\bar{\alpha}_{t+\Delta t}(\frac{1}{\bar{\alpha}_{t+\Delta t}}-\frac{1}{\bar{\alpha}_t})}x_{t+\Delta t}+\frac{x_0\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}+\frac{x_{\Delta t+t}}{\bar{\alpha}_{t+\Delta t}(\frac{1}{\bar{\alpha}_{t+\Delta t}}-\frac{1}{\bar{\alpha}_t})}}{\frac{\sqrt{\bar{\alpha}_{t+\Delta t}/\bar{\alpha}_t}-1+1)^2}{\bar{\alpha}_{t+\Delta t}\left(\frac{1}{\bar{\alpha}_{t+\Delta t}}-\frac{1}{\bar{\alpha}_t}\right)}+\frac{1}{(1-\bar{\alpha}_t)}}\right)^2\right)
\end{aligned}
\tag{54}
$$

取 $\Delta t=1$，考虑到 $\bar{\alpha}_{t+1}/\bar{\alpha}_t=\alpha_t$。有

$$
\begin{aligned}
& p(x_t|x_{t+1}) \\
\propto\, & exp\left(-\frac{1}{2}(\frac{1-\bar{\alpha}_{t+1}}{(1-\alpha_{t+1})(1-\bar{\alpha}_t)})\left(x_t-\frac{\frac{x_0\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}+\frac{\sqrt{\alpha_{t+1}}x_{\Delta t+t}}{1-\alpha_{t+1}}}{\frac{1-\bar{\alpha}_{t+1}}{(1-\alpha_{t+1})(1-\bar{\alpha}_t)}}\right)^2\right) \\
=\, & exp\left(\frac{\left(x_t-\frac{x_0\sqrt{\bar{\alpha}_t}(1-\alpha_{t+1})+\sqrt{\alpha_{t+1}}(1-\bar{\alpha}_t)x_{t+1}}{1-\alpha_{t+1}}\right)^2}{2\frac{(1-\alpha_{t+1})(1-\bar{\alpha}_t)}{1-\bar{\alpha}_{t+1}}}\right) \\
\sim\, & N(x_t;\frac{x_0\sqrt{\bar{\alpha}_t}(1-\alpha_{t+1})+\sqrt{\alpha_{t+1}}(1-\bar{\alpha}_t)x_{t+1}}{1-\alpha_{t+1}},\frac{(1-\alpha_{t+1})(1-\bar{\alpha}_t)}{1-\bar{\alpha}_{t+1}})
\end{aligned}
\tag{55}
$$

这正是 DDPM 的逆向过程!
证明 $\Delta t\to 0$ 时，离散的逆过程 SDE 表达式和 (14) 相同

5.4 Birth of DDIM

There be an SHC, and he said, "Oh no, i donnot want any uncertainty, i want anything to be certain.", then he let $\sigma_t\equiv 0$.
这种情况下，SDE(1) 写为

$$
dx=\left(f(x,t)-\frac{1}{2}g_t^2\nabla_x\log p_t(x)\right)dt \tag{56}
$$

SDE 逆过程 (16) 写为

$$
dx=\left(f(x,t)-\frac{1}{2}g_t^2\nabla_x\log p_t(x)\right)dt \tag{57}
$$

可以看到，这种情况下，方程中没有任何随机项出现，SDE 变成了 ODE！且正过程和逆过程的表达式相同，也就是说这是一个完全可逆的过程!

考虑该方程描述的离散化情景，由 (43),(2),(33), 以及离散化参数 (45)46)，式子 (56) 可写为

$$
\begin{aligned}
dx &= \frac{1}{\lambda_t} d\lambda_t x - \frac{1}{2}\lambda_t^2 \frac{d}{dt}\left(\frac{\mu_t}{\lambda_t}\right)^2 \nabla_x \log p_t(x) dt \\
\Rightarrow dx &= \frac{1}{\lambda_t} d\lambda_t x - \lambda_t \mu_t \frac{d}{dt}\left(\frac{\mu_t}{\lambda_t}\right)(-\frac{\epsilon(x_t,t)}{\mu_t}) dt \\
\Rightarrow d(\frac{x}{\lambda_t}) &= d\left(\frac{\mu_t}{\lambda_t}\right)\epsilon(x_t,t) \\
\Rightarrow \frac{x_{t+\Delta t}}{\lambda_{t+\Delta t}} - \frac{x_t}{\lambda_t} &= (\frac{\mu_{t+\Delta t}}{\lambda_{t+\Delta t}} - \frac{\mu_t}{\lambda_t})\epsilon(x_t,t)
\end{aligned} \tag{58}
$$

取 $\Delta t = 1$，并结合 (33)，可得

$$
\frac{x_{t+1}}{\sqrt{\bar{\alpha}_{t+1}}} - \frac{x_t}{\sqrt{\bar{\alpha}_t}} = (\frac{\sqrt{1-\bar{\alpha}_{t+1}}}{\sqrt{\bar{\alpha}_{t+1}}} - \frac{\sqrt{1-\bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}})\epsilon(x_t,t) \tag{59}
$$

整理得

$$
\begin{aligned}
x_{t+1} &= \sqrt{\alpha_{t+1}} x_t + (\sqrt{1-\bar{\alpha}_{t+1}} - \sqrt{\alpha_{t+1}}\sqrt{1-\bar{\alpha}_t})\epsilon(x_t,t) \\
x_t &= \frac{x_{t+1}}{\sqrt{\alpha_{t+1}}} - (\frac{\sqrt{1-\bar{\alpha}_{t+1}}}{\sqrt{\alpha_{t+1}}} - \sqrt{1-\bar{\alpha}_t})\epsilon(x_t,t)
\end{aligned} \tag{60}
$$

这正是 DDIM 的去噪方程!

5.5 DDPM 和 DDIM 的等效性

DDIM 和 DDPM 是等效的，或者说 1所描述的这个 SDE 族所描述的随机过程，在每一时刻的观测意义上是等效的。

虽然每一步 $t \to t + \Delta t$, 对于特定的 x_t，他所产生的变化是不同的，（随 σ 的不同而不同，最显而易见的例子就是可以有随机噪声，也可以没有随机噪声），但是综合所有的 $\{x_t\}$，也就是说，在每个 t 时刻，观测到的 $\{x_t\}$ 的分布总是相同的，与 σ 无关。

换句话说，这个分布族描述的一系列随机过程，虽然在每一步，个体的变化不同，但总体的分布（边缘分布）总是相同的。

特别地，当 $\sigma_t \equiv 0$ 时，这个完全可逆的 ODE 描述的变化过程，与 SDE 描述的变化过程在每一时刻具有相同的边缘分布。

Song 在文章中给出一张图，生动形象地展示了这一点 [5]

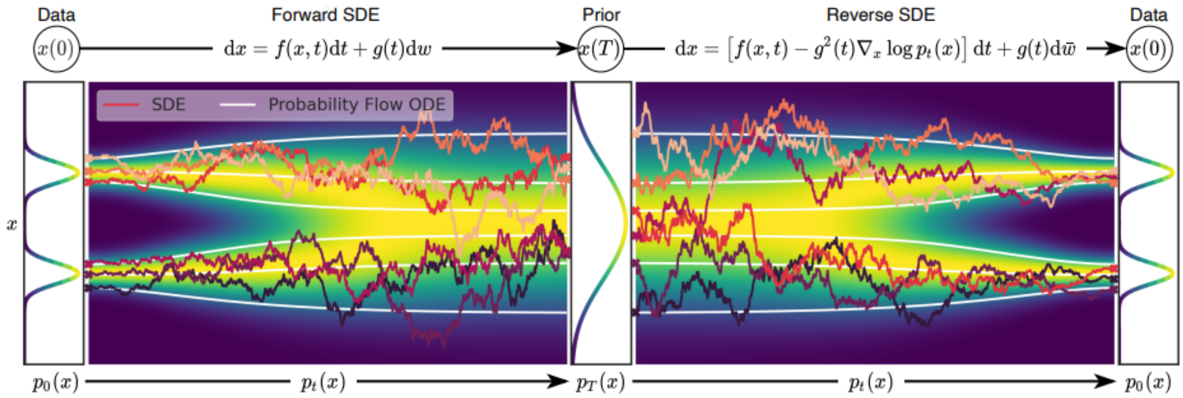

Figure 2: **Overview of score-based generative modeling through SDEs**. We can map data to a noise distribution (the prior) with an SDE (Section 3.1), and reverse this SDE for generative modeling (Section 3.2). We can also reverse the associated probability flow ODE (Section 4.3), which yields a deterministic process that samples from the same distribution as the SDE. Both the reverse-time SDE and probability flow ODE can be obtained by estimating the score $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ (Section 3.3).

图 1: 不同 SDE 描述的变化过程

5.6 Loss Function

6 Flow-Based Model

6.1 Understand Flow

基本概念

在概率流中，有几个重要的概念

1. 联合概率密度路径
$$p_t(x_1, x_2, \ldots, x_d) = p_t(x) : [0,1] \times \mathcal{R}^d \to \mathcal{R}_{>0} \tag{61}$$
描述了随变化时间，不同样本点的概率密度

2. 向量场
$$v = v_t(x_t) : [0,1] \times \mathcal{R}^d \to \mathcal{R}^d \tag{62}$$
描述了样本初始点在不同时间的变化方向及变化率

3. CNF 连续正规化流 (flow map)
$$\begin{aligned} &\phi_t(x) : [0,1] \times \mathcal{R}^d \to \mathcal{R}^d \\ &\phi_0(x) = x \end{aligned} \tag{63}$$

将 $t=0$ 时的样本初始点，映射到 $t=t$ 时刻他所对应的位置上。

这里可以这样理解，之前的 ODE 描述的是一个点的 ODE，t=0 时 $x_t=x_0$ 描述了初始位置，$t=t$ 时 x_t 描述了在 t 时刻这个样本的位置。但这里不再描述单一样本点，而是描述所有样本点 (从而具有多个初始点)，也就是说，不同的 x 表示的是不同的初始点，而不能表示一个样本点在不同时刻的值，所以我们需要一个 $\phi_t(x)$ 来表示初始点值为 x 的在 t=t 时的位置。这里的 x 相当于相当于是 x_0，描述 0 时刻所有样本点，$\phi_t(x)$ 相当于时 x_t，但 x 取任意值，描述 t 时刻所有样本点。也就是说，这里的 $\phi_t(1)$ 等于之前的 1_t，之前对应于 x_t 的 x 表示初始点，是不会变的，这里的 $phi_t(x)$，x 也为初始点，但可以取为全空间的任意值。换句话来说，以前的 ODE 描述了单一的样本点变化，这里的 ODE 同时描述了全空间所有样本点的变化。

概率流和样本点的协同变化

向量场 $v_t(x)$ 描述了样本点 x 的变化率，数学表示为

$$\begin{aligned}\frac{d}{dt}\phi_t(x)&=v_t(\phi_t(x))\\ \phi_0(x)&=x\end{aligned}\tag{64}$$

根据概率协同不变性

$$\begin{aligned}&p_0(x)d\phi_0(x)=p_t(\phi_t(x))d(\phi_t(x))\\ &p(x)dx=p_t(\phi_t(x))d(\phi_t(x))\\ &p_t(x_t)=p_0(\phi_t^{-1}(x_t))\det\left[\frac{\partial\phi_t^{-1}(x_t)}{\partial x_t}\right]\end{aligned}\tag{65}$$

所以我们可以定义一个推前映射

$$p_t=[\phi_t]_*p_0(x)=p_0(\phi_t^{-1}(x))\det\left[\frac{\partial\phi_t^{-1}(x)}{\partial x}\right]\tag{66}$$

概率流连续性方程

$$\frac{\partial p}{\partial t}+\nabla_x(pv)=0\tag{67}$$

单位时间内体积内 p 的减少等于流出体积的概率量。p 在 4-空间内是恒定的。

DDIM 的 ODE 具有自洽的概率流和向量场

DDPM 的 ODE(56)

$$dx=\left(f(x,t)-\frac{1}{2}g_t^2\nabla_x\log p_t(x)\right)dt\tag{68}$$

对应的 F-P 方程 (25)

$$\begin{aligned}
&\frac{\partial}{\partial t}p_t(x) = -\nabla_x(f(x,t)p_t(x)) + \frac{1}{2}g_t^2\nabla_x^2 p_t(x)\\
\Rightarrow &\frac{\partial}{\partial t}p_t(x) = -\nabla_x(f(x,t)p_t(x) - \frac{1}{2}g_t^2\nabla_x p_t(x))\\
&\qquad = -\nabla_x[p_t(x)\cdot(f(x,t) - \frac{1}{2}g_t^2\nabla_x \log p_t(x))]
\end{aligned} \tag{69}$$

满足概率连续性方程，且对应的

$$v_t(x) = f(x,t) - \frac{1}{2}g_t^2\nabla_x \log p_t(x) \tag{70}$$

6.2 一个变换带来的分布

对于 $t \in [0,1]$ 假如我们有

$$\begin{aligned}
&p_0(x) = \mathcal{N}(0,I)\\
&p_1(x) = q(x) \quad (\text{特定分布})
\end{aligned} \tag{71}$$

我们想找一组自洽的 (ϕ_t, v_t, p_t)，从而构建从 p_0 到 p_1 的一组概率路径。
在给定 $t=1$ 时的 x_1 的情况下，寻找条件概率路径 $(\phi_t(x), v_t(x|x_1), p_t(x|x_1))$，使得

$$\begin{aligned}
&p_0(x|x_1) = \mathcal{N}(0,I)\\
&p_1(x|x_1) = \mathcal{N}(x_1, \sigma_{min}^2 I) \sim \delta(x - x_1)
\end{aligned} \tag{72}$$

从而他们的所产生的边缘分布满足式 (71)
我们令

$$\begin{aligned}
&\phi_t(x|x_1) = \sigma_t(x_1)x + \mu_t(x_1)\\
&\sigma_0 = 1, \sigma_1 = \sigma_{min}\\
&\mu_0 = 0, \mu_1 = x_1
\end{aligned} \tag{73}$$

则

$$\begin{aligned}
&p_t(x|x_1) = ([\phi_t]_* p_0)(x) = p_0((x-\mu)/\sigma)/\sigma = \mathcal{N}(\mu_t(x_1), \sigma_t^2(x_1))\\
&p_1(x|x_1) = \mathcal{N}(x_1, \sigma_{min}^2)
\end{aligned} \tag{74}$$

满足我们所需条件，其对应的向量场为

$$\begin{aligned}
&v_t(\phi_t(x)|x_1) = d\phi_t(x)/dt = \sigma_t'(x_1)x + \mu_t'(x_1)\\
&= \sigma_t'(x_1)(\phi_t(x) - \mu_t(x_1))/\sigma_t(x_1) + \mu_t'(x_1)\\
\Rightarrow &v_t(x|x_1) = \frac{\sigma_t'(x_1)(x - \mu_t(x_1))}{\sigma_t(x_1)} + \mu_t'(x_1)
\end{aligned} \tag{75}$$

6.3 Flow Matching

由于向量场可以生成概率路径，所以我们可以找到一个向量场，使其建立标准高斯分布到数据空间分布的概率路径。由于 ODE 变换是可逆的，所以可以用来做生成。
那么由此，我们希望优化

$$\mathcal{L}_{FM}(\theta) = E_{t,p_t(x)}||u_t(x) - v_t(x)||^2 \tag{76}$$

其中，u_t 可以由网络预测，v_t 是 gt 值。
但是现在我们没有一个 gt 值，我们可以构造一个 gt 值。使其产生的概率路径满足
$N(0,1) \to q_{data}(x)$。但是 $q_{data}(x)$ 未知，所以我们需要采取一些技巧：通过混合简单路径来构造目标概率路径。给定一个特殊的数据点 x_1，构造 CFM 使得，$p_0(x|x_1) = N(0,1)$,
$p_1(x|x_1) = N(x|x_1, \sigma^2 I)$，从而其边缘分部满足 FM 中的形式。其目标函数为

$$\mathcal{L}_{CFM}(\theta) = E_{t,p_t(x\ x_1)}||u_t(x) - v_t(x|x_1)||^2 \tag{77}$$

可以证明

$$\nabla_\theta \mathcal{L}_{FM} = \nabla_\theta \mathcal{L}_{CFM} \tag{78}$$

而且 gt 值 $v_t(x|x_1)$ 我们已经在上一节中找到了:(75)
Flow Matching 原文 [2] 中给出了 VE、VP 两种情况的事例，以及 Optimal Transport 情况。在附录中给出了极限情况下的 Diffusion 理论中的 v_t(70) 和 flow-matching 中式 (75) 的等价。我懒得推了，直接 copy 过来

Example I: Diffusion conditional VFs. Diffusion models start with data points and gradually add noise until it approximates pure noise. These can be formulated as stochastic processes, which have strict requirements in order to obtain closed form representation at arbitrary times t, resulting in Gaussian conditional probability paths $p_t(x|x_1)$ with specific choices of mean $\mu_t(x_1)$ and std $\sigma_t(x_1)$ (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2020b). For example, the reversed (noise→data) Variance Exploding (VE) path has the form

$$p_t(x) = \mathcal{N}(x|x_1, \sigma_{1-t}^2 I), \tag{16}$$

where σ_t is an increasing function, $\sigma_0 = 0$, and $\sigma_1 \gg 1$. Next, equation 16 provides the choices of $\mu_t(x_1) = x_1$ and $\sigma_t(x_1) = \sigma_{1-t}$. Plugging these into equation 15 of Theorem 3 we get

$$u_t(x|x_1) = -\frac{\sigma'_{1-t}}{\sigma_{1-t}}(x - x_1). \tag{17}$$

The reversed (noise→data) Variance Preserving (VP) diffusion path has the form

$$p_t(x|x_1) = \mathcal{N}(x \,|\, \alpha_{1-t}x_1, \left(1 - \alpha_{1-t}^2\right) I), \text{where } \alpha_t = e^{-\frac{1}{2}T(t)}, T(t) = \int_0^t \beta(s)ds, \tag{18}$$

and β is the noise scale function. Equation 18 provides the choices of $\mu_t(x_1) = \alpha_{1-t}x_1$ and $\sigma_t(x_1) = \sqrt{1 - \alpha_{1-t}^2}$. Plugging these into equation 15 of Theorem 3 we get

$$u_t(x|x_1) = \frac{\alpha'_{1-t}}{1 - \alpha_{1-t}^2}(\alpha_{1-t}x - x_1) = -\frac{T'(1-t)}{2}\left[\frac{e^{-T(1-t)}x - e^{-\frac{1}{2}T(1-t)}x_1}{1 - e^{-T(1-t)}}\right]. \tag{19}$$

图 2

Example II: Optimal Transport conditional VFs. An arguably more natural choice for conditional probability paths is to define the mean and the std to simply change linearly in time, *i.e.*,

$$\mu_t(x) = tx_1, \text{ and } \sigma_t(x) = 1 - (1 - \sigma_{\min})t. \tag{20}$$

According to Theorem 3 this path is generated by the VF

$$u_t(x|x_1) = \frac{x_1 - (1 - \sigma_{\min})x}{1 - (1 - \sigma_{\min})t}, \tag{21}$$

图 3

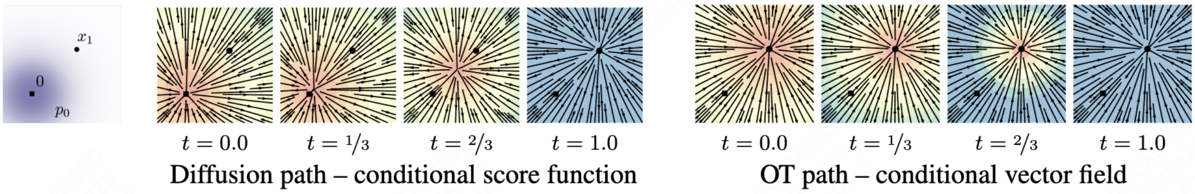

Figure 2: Compared to the diffusion path's conditional score function, the OT path's conditional vector field has constant direction in time and is arguably simpler to fit with a parametric model. Note the blue color denotes larger magnitude while red color denotes smaller magnitude.

图 4

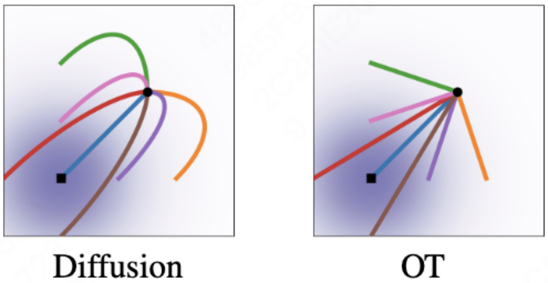

Figure 3: Diffusion and OT trajectories.

图 5

Variance Preserving (VP) path The SDE for the VP path is

$$dy = -\frac{T'(t)}{2}y + \sqrt{T'(t)}dw,$$

where $T(t) = \int_0^t \beta(s)ds$, $t \in [0,1]$. The SDE coefficients are therefore

$$f_s(y) = -\frac{T'(s)}{2}y, \quad g_s = \sqrt{T'(s)}$$

and

$$p_t(y|y_0) = \mathcal{N}(y|e^{-\frac{1}{2}T(t)}y_0, (1 - e^{-T(t)})I).$$

Plugging these choices in equation 40 we get the conditional VF

$$w_t(y|y_0) = \frac{T'(t)}{2}\left(\frac{y - e^{-\frac{1}{2}T(t)}y_0}{1 - e^{-T(t)}} - y\right) \tag{41}$$

Using Lemma 1 to reverse the time we get the conditional VF for the reverse probability path:

$$\begin{aligned}\tilde{w}_t(y|y_0) &= -\frac{T'(1-t)}{2}\left(\frac{y - e^{-\frac{1}{2}T(1-t)}y_0}{1 - e^{-T(1-t)}} - y\right)\\ &= -\frac{T'(1-t)}{2}\left[\frac{e^{-T(1-t)}y - e^{-\frac{1}{2}T(1-t)}y_0}{1 - e^{-T(1-t)}}\right],\end{aligned}$$

which coincides with equation 19.

图 6

参考文献

[1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.

[2] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

[3] Cheng Luo. Understanding diffusion models: A unified perspective. *arXiv preprint arXiv:2208.11970*, 2022. arXiv:2208.11970.

[4] Jiaming Song, Cheng Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. arXiv:2010.02502.

[5] Yang Song, Jon Sohl-Dickstein, D.P. Kingma, and et al. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. arXiv:2011.13456.