

$$x_0 \longrightarrow x_1 \dots \longrightarrow x_{t-1} \xrightarrow{q(x_t|x_{t-1})} x_t \longrightarrow \dots \longrightarrow x_T$$

~~$q(x_t|x_{t-1}) = \mathcal{N}(\sqrt{\alpha_t}x_{t-1}, (1-\alpha_t)\mathbb{I})$~~

$$x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}\varepsilon.$$

$$q(x_t|x_{t-1}) = \mathcal{N}(\sqrt{\alpha_t}x_{t-1}, (1-\alpha_t)\mathbb{I})$$

$$q(x_T) = \int q(x_0) q(x_1|x_0) q(x_2|x_1) \dots q(x_T|x_{T-1}) dx_0:T$$

如果用真实  $q(x_{t-1}|x_t) = \frac{q(x_t|x_{t-1})}{q(x_t)}$  进行采样

则逆过程的

$$\begin{aligned} p(x_0) &= \int q(x_T) \prod_{t=1}^T q(x_{t-1}|x_t) dx_0:T \\ &= \int q(x_T) \cdot \prod_{t=1}^T \frac{q(x_t|x_{t-1})}{q(x_t)} dx_0:T \\ &= \int q(x_0) \cdot \prod_{t=1}^T q(x_t|x_{t-1}) dx_0:T = q(x_0) \end{aligned}$$

能成功还原  $q(x_0)$

为了完成这个目标，有两种方式。①有

方法：①  $p_\theta(x_t|x_0)$  存在于

也就是说如果有网络能够拟合  ~~$q(x_t|x_{t-1})$~~   $q(x_t|x_{t-1})$   
我们能完成完美采样

定义网络  $q(p_\theta(x_t|x_0))$ . 如何优化？

$$\text{极大似然估计. } p_\theta(x_0) = \int p_\theta(x_0:T) dx_0:T$$

$$= \int p_\theta(x_0|x_{1:T}) p_\theta(x_{1:T}) dx_{1:T}$$

$$= \frac{1}{N} \sum_{i=1}^N p_\theta(x_0|x_{1:T}^i)$$

① 采样非常困难.

② 效率低下， $x_0$  与  $x_{1:T}$  无关联.

采用重要性采样.

$$P_\theta(x_0) = \int p_\theta(x_0; T) / q(x_{1:T} | x_0) \cdot q(x_{1:T} | x_0) dx_{1:T}.$$

好处: ① 引入  $x_{1:T}$  与  $x_0$  的关联, 提高网络选取采样到有意义区域

② 提高采取样本速度, 正向过程中的  $X_{1:T}$  计算方便

上式利用蒙特卡洛方法可近似计算

$$P_\theta(x_0) \approx \frac{1}{N} \sum_{i=1}^N \frac{p_\theta(x_0; T)}{q(x_{1:T} | x_0)}.$$

问题: 采样估计存在偏差, 上述目标可能不稳定.

利用不等式进行放缩, 最大化下界.

$$\begin{aligned} \log P_\theta(x_0) &= \log \int \frac{p_\theta(x_0; T)}{q(x_{1:T} | x_0)} \cdot q(x_{1:T} | x_0) dx_{1:T} \\ &\geq \cancel{\int q(x_{1:T} | x_0) \log \frac{p_\theta(x_0; T)}{q(x_{1:T} | x_0)} dx_{1:T}} \\ &= E_{x_{1:T} \sim q} [\log \frac{p_\theta(x_0; T)}{q(x_{1:T} | x_0)}] \end{aligned}$$

$$P_\theta(x_0; T) = \cancel{p_\theta(x_T)} P_\theta(x_T) \cdot \prod_{t=1}^{T-1} P_\theta(x_{t+1} | x_t)$$

$$q(x_{1:T} | x_0) = \prod_{t=1}^T q(x_t | x_{t-1}) = q(x_1 | x_0) \prod_{t=2}^T q(x_t | x_{t-1}, x_0)$$

如果按式①展开.

$$\log P_\theta(x_0) \geq E_{x_{1:T} \sim q} \left[ \log \frac{P_\theta(x_1) \prod_{t=1}^{T-1} P_\theta(x_{t+1} | x_t)}{\prod_{t=1}^T q(x_t | x_{t-1})} \right]$$

$$\begin{aligned} &= E_{x_{1:T} \sim q} \left[ \log \frac{P_\theta(x_T) \prod_{t=1}^{T-1} P_\theta(x_{t+1} | x_t)}{\frac{q(x_{t+1} | x_t) q(x_t)}{q(x_{t+1})}} \right] \\ &= E_{x_{1:T} \sim q} \left[ \underbrace{\log \frac{P_\theta(x_T)}{q(x_T)} + \sum_{t=1}^{T-1} \log \frac{P_\theta(x_{t+1} | x_t)}{q(x_{t+1} | x_t)} + \log q(x_0)}_{\text{与 } E[\text{采样分布}] \text{ 不同} \rightarrow \text{负的KL}} \right] \end{aligned}$$

② 式展开：

$$\begin{aligned}
 & \log p_{\theta}(x_0) = E_{X_1:T \sim q} \left[ \log \frac{\prod_{t=1}^T p_{\theta}(x_t | x_{t-1})}{\prod_{t=2}^T q(x_t | x_{t-1}, x_0)} \right] \\
 & = E_{X_1:T \sim q} \left[ \log \frac{\prod_{t=1}^T p_{\theta}(x_t | x_{t-1})}{\prod_{t=2}^T \frac{q(x_t | x_{t-1}, x_0) q(x_t | x_0)}{q(x_{t-1} | x_0)}} \right] \\
 & = E_{X_1:T \sim q} \left[ \log \frac{\prod_{t=1}^T p_{\theta}(x_t | x_{t-1}) \cdot q(x_t | x_0)}{\cancel{q(x_t | x_0)} \prod_{t=2}^T q(x_{t-1} | x_t, x_0)} \right] \\
 & = E_{X_1:T \sim q} \left[ \log \frac{p_{\theta}(x_T)}{q(x_T | x_0)} + \sum_{t=2}^T \log \frac{p_{\theta}(x_t | x_{t-1})}{q(x_t | x_{t-1}, x_0)} + \log p_{\theta}(x_0 | x_1) \right]
 \end{aligned}$$

~~$E_{X_1:T \sim q}$~~

$$E_{x_0} [\log p_{\theta}(x_0)] = E_q \left[ -KL(q(x_T | x_0) || p_{\theta}(x_T)) - \sum_{t=2}^T KL(q(x_t | x_{t-1}, x_0) || p_{\theta}(x_t | x_{t-1})) + \log p_{\theta}(x_0 | x_1) \right].$$

$$q(x_{t-1} | x_t, x_0) = \frac{q(x_t | x_{t-1}, x_0) q(x_{t-1} | x_0)}{q(x_t | x_0)}$$

$$\propto \exp \left[ -\frac{(x_t - \sqrt{\alpha_t} x_{t-1})^2}{2(1-\alpha_t)} - \frac{(x_{t-1} - \sqrt{\alpha_{t-1}} x_0)^2}{2(1-\bar{\alpha}_{t-1})} + \frac{(x_t - \sqrt{\alpha_t} x_0)^2}{2(1-\bar{\alpha}_t)} \right]$$

快速 ~~暴力~~ 漆平方。

$$2 \frac{\log q(x_{t-1} | x_t, x_0)}{\partial x_{t-1}} = \frac{\sqrt{\alpha_t} (x_t - \sqrt{\alpha_t} x_{t-1})}{1-\alpha_t} - \frac{x_{t-1} - \sqrt{\alpha_{t-1}} x_0}{1-\bar{\alpha}_{t-1}} = 0$$

$$\begin{aligned}
 x_t &= \frac{\sqrt{\alpha_t} (1-\bar{\alpha}_{t-1}) x_t + (1-\alpha_t) \sqrt{\alpha_{t-1}} x_0}{1-\bar{\alpha}_t} \rightarrow x_0 = \frac{x_t - \sqrt{1-\bar{\alpha}_t} \varepsilon}{\sqrt{\alpha_t}} \\
 &= \frac{\alpha_t (1-\bar{\alpha}_{t-1}) x_t + (1-\alpha_t) (x_t - \sqrt{1-\bar{\alpha}_t} \varepsilon)}{\sqrt{\alpha_t} (1-\bar{\alpha}_t)}
 \end{aligned}$$

$$= \frac{1}{\sqrt{\alpha_t}} \left( x_t - \frac{(1-\alpha_t)}{\sqrt{1-\bar{\alpha}_t}} \varepsilon \right)$$

~~状态转移~~ 式子差

$$\frac{\partial^2 \log}{\partial x_t^2} = -\frac{\alpha_t}{1-\alpha_t} - \frac{1}{1-\bar{\alpha}_{t-1}} = \frac{-\alpha_t + \bar{\alpha}_t - 1 + \beta_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})} = 2\alpha.$$

$$-\frac{1}{2\alpha} = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} (1-\alpha_t)$$

$$-\frac{(x-u)^2}{2 \cdot (-\frac{1}{2\alpha})}$$

求得  $u = \frac{1}{\sqrt{\alpha_t}} [x_t - \frac{(1-\alpha_t)}{\sqrt{1-\bar{\alpha}_t}} \varepsilon]$  ( $\text{ps: } u = \frac{1}{\sqrt{\alpha_t}} (x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \varepsilon)$ )

$$\sigma = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} (1-\alpha_t)$$

定  $p_\theta(x_{t+1}|x_t) = \text{norm}(x_t; u_\theta(x_t, t), \sigma(x_t, t))$

$$u_\theta(x_t, t) = \frac{1}{\sqrt{2\pi}} (x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} g_\theta(x_t, t)).$$

$$\sigma(x_t, t) = \sigma$$

IMI  $KL(p_\theta \| q(x_{t-1}|x_t, x_0) | p_\theta(x_{t-1}|x_t))$

$$= \frac{1}{2\sigma} ||u_\theta - u||^2$$

$$= \frac{\beta_t^2}{2\sigma \alpha_t (1-\alpha_t)} ||\varepsilon_\theta - \varepsilon||^2 \quad \text{权重一般是递减的}$$

简化 RSS =  $E_{x_0, \varepsilon, t} ||\varepsilon_\theta(\beta x_0 + \sqrt{1-\bar{\alpha}_t} \varepsilon, t) - \varepsilon||^2$

↓  
简化 RSS 在  $t \rightarrow 0$  处约束变大  
 $t \rightarrow 1$  处约束变小

随机过程  $\{X_t, t \in T\}$  可以看作是时间序列上的一系列随机变量：

$B_t$ , or  $W_t$ , 布朗运动 或 维纳过程.

$$\textcircled{1} \quad B_{t=0} = 0$$

\textcircled{2} 独立增量:  $t_1 > s_1, t_2 > s_2$  且  $[t_1, s_1], [t_2, s_2]$  没有交集  
则  $B(t_1) - B(s_1)$  与  $B(t_2) - B(s_2)$  独立.

\textcircled{3} 正态分布增量:  $t > s, B(t) - B(s) \sim N(0, t-s)$

\textcircled{4} 无记忆性: ~~增量~~  $dB = B(t) - B(s)$ , 只与  $t-s$  有关 与  $B(s)$

$B(t)$   
值无关

随机微分方程 (SDE)

$$dx = \underbrace{f(x, t) dt}_{\text{确定性}} + \underbrace{g(x, t) dB}_{\text{随机噪声.}}$$

$$dB \sim N(0, dt), \quad (dB)^2 = dt \quad (\text{ps: } E(dB^2) = dt)$$

看一个例子.  $dx = g(t) dB$ .

易解得  $x = x_0 + \underbrace{\int_0^t g(u) dB}_\text{有什么性质?}$

$$I(t) = \int_0^t g(u) dB$$

\textcircled{5} 由于  $dB$  每个时间  $t$  都相互独立,  $I(t)$  可以看作相互独立的高斯分布的卷加 而知  $I(t) \sim N(\mu, \sigma^2)$ . (正态)

$$\mu = E(I(t)) = E\left(\int_0^t g(u) dB\right) = \int_0^t E(g(u)) dB = \int_0^t 0 = 0$$

$$\sigma^2 = D(I(t)) = E(I^2(t)) - [E(I(t))]^2 = E\left(\int_0^t g(u) dB \int_0^t g(v) dB\right)$$

$$\begin{aligned}
 E(I(t)) &= \int_0^t g(u) dB(u) \int_0^t g(u) dB(u) = E \int_0^t \int_0^t g(u) g(v) dB(u) dB(v) \\
 &= \int_0^t \int_{u=v} E(g(u) g(v) dB(u) dB(v)) + \int_{u \neq v} E(g(u) g(v) dB(u) dB(v)) \\
 &= \int_0^t g(u)^2 du.
 \end{aligned}$$

$$\sigma^2 = \int_0^t g(u)^2 du. \quad \left( \begin{array}{l} \text{由 } dB \sim N(0, dt). \\ \text{且 } g(u) dB \sim N(0, g(u)^2 dt). \\ \text{方差相加 } D(I(t)) = \int_0^t g(u)^2 du \end{array} \right)$$

再看一个ODE:

$$dx = f(t)x dt.$$

$$\text{不难得解得: } x = x_0 \exp\left(\int_0^t f(u) du\right).$$

如何理解  $\exp\left(\int_0^t f(u) du\right)$  这一项?

$$dx = f(x) x dt = f(t) x dt$$

$$\underline{x_{t+dt}} = (1 + f(t) dt) x_t.$$

$$\cancel{x_t = (1 + f(t))}$$

$$x_{dt} = (1 + \cancel{f(t) dt}) x_0$$

$$x_{2dt} = (1 + f(dt) dt) x_{dt} = (1 + f(dt) dt) ((1 + f(dt) dt) x_0).$$

$$x_t = \underbrace{(1 + f(t-dt) dt)}_{\cancel{\text{taylor}}} \dots \dots \underbrace{(1 + f(t, dt)}_{\cancel{\text{taylor}}} x_0.$$

$$= \exp(f(t-dt) dt) \dots \exp(f(t, dt))$$

$$= x_0 \exp\left(\sum_k f(kdt) dt\right) \stackrel{k \rightarrow \infty}{=} x_0 \exp\left(\int_0^t f(u) du\right)$$

再看一个 SDE:  $dx = f(t)Xdt + g(t)dB$ .

$$ODE = dx = f(t)Xdt \text{ 解为 } X = X_0 \exp(\int_0^t f(u)du)$$

$$\text{构造 } M(t) = \exp(\int_0^t -f(u)du)$$

$$\begin{aligned} M(t)dx - M(t)f(t)Xdt &= g(t)M(t)dB \\ d(M(t)X) &= g(t)M(t)dB \\ d(M(t)X) &= dM(t) \cdot X + M(t)dx \\ &= M(t) \cdot (-f(t))Xdt + M(t)dx. \end{aligned}$$

$$M(t)X = \int_0^t g(u)M(u)dB + M(0)X_0.$$

$$\begin{aligned} X &\leftarrow \cancel{M(t)X} \quad \cancel{M(t)} \neq \cancel{\exp(\int_0^t -f(u)du)} \\ \frac{M(s)}{M(t)} &= \frac{\exp(\int_0^s -f(u)du)}{\exp(\int_0^t -f(u)du)} = \exp(\int_s^t f(u)du). \\ X &= \underbrace{X_0 \exp(\int_0^t f(u)du)}_{\text{确定性项}} + \underbrace{\int_0^t \exp(\int_u^t f(s)ds) g(u) dB}_{\text{随机性项}} \end{aligned}$$

虽然将  $dx = f(t)Xdt + g(t)dB$  按照前面 ODE 的写法.

$$\text{可以写成 } X = \cancel{X_0 \exp}$$

$$\begin{aligned} X &= X_0 \prod_{u=0}^t (1 + f(u)du) + \sum_{u=0}^t \left( \prod_{s=u}^t (1 + f(s)ds) \right) g(u) dB \\ &\stackrel{\text{taylor}}{=} X_0 \exp\left(\sum_{u=0}^t f(u)du\right) + \sum_{u=0}^t \exp\left(\sum_{s=u}^t f(s)ds\right) g(u) dB \\ &= X_0 \exp\left(\int_0^t f(u)du\right) + \int_0^t \exp\left(\int_u^t f(s)ds\right) g(u) dB \end{aligned}$$

回归问题.

$$dx = f(x_t, t) dt + g_t dB.$$

$$\text{则 } X_{t+dt} = X_t + f(x_t, t) dt + g_t dB.$$

$$p(X_{t+dt} | X_t) = \mathcal{N}(X_{t+dt}; X_t + f(x_t, t) dt, g_t^2 dt)$$

按照前面的双向加 Bayes, 用真是的  $p(X_t | X_{t+dt})$  逆采样而得  $p(X_t)$

$$p(X_t | X_{t+dt}) = \frac{p(X_{t+dt} | X_t) p(X_t)}{p(X_{t+dt})}$$

$$\propto \exp\left(-\frac{(X_{t+dt} - (X_t + f(x_t, t) dt))^2}{2g_t^2 dt} + \log p(X_t) - \log p(X_{t+dt})\right).$$

$$\log p(X_t) - \log p(X_{t+dt}) = -d \log p(X_t)$$

$$= -(\nabla_X \log p(X_t) dx + \frac{\partial \log p(X_t)}{\partial t} dt)$$

$$= (x_t - X_{t+dt}) \nabla_X \log p(X_t) + O(t).$$

$$\Rightarrow \exp\left(-\frac{(X_{t+dt} - (X_t + f(x_t, t) dt))^2}{2g_t^2 dt} + (x_t - X_{t+dt}) \nabla_X \log p(X_t)\right)$$

$$\text{快速配方法: } \frac{\partial \log p(X_t | X_{t+dt})}{\partial x_t} = \frac{X_{t+dt} - (X_t + f(x_t, t) dt)}{g_t^2 dt} + \nabla_X \log p(X_t) = 0$$

$$\Rightarrow X_t = X_{t+dt} - (f(x_t, t) - g_t^2 \nabla_X \log p(X_t)) dt.$$

取  $\Delta t = -dt$ , ~~附近似然  $h(t+dt)dt = h(t)dt$~~ ,  
 $\frac{\partial^2 \log p}{\partial x^2} = -\frac{1}{g_t^2 dt} = 2a$

$$g_t^2 = -\frac{1}{2a} = g_{t+dt}^2 dB \quad \text{有 } dx = [f(x_t, t) - g_t^2 \nabla_X \log p(X_t)] dt + g_t dB.$$

$$\text{即 } dx = f(x_t, t) dt + g_t dB.$$

$$\text{对应逆向 } dx = [f(x_t, t) - g_t^2 \nabla_X \log p(X_t)] dt + g_t dB.$$

NCSN:

$$X_{i+1} = X_i + \sqrt{\sigma^2(t+1) - \sigma^2(t)} \varepsilon, \quad i=1, \dots, N.$$

取  $\Delta t = \frac{1}{N}$ ,  $t = \frac{i}{N}$ ,  $\Delta t = \frac{1}{N}$ .

$$\text{Rif } X_{t+\Delta t} = X_t + \sqrt{\sigma^2(t+\Delta t) - \sigma^2(t)} \varepsilon$$

$$= X_t + \sqrt{\frac{d[\sigma^2(t)]}{dt} \Delta t} \varepsilon$$

$$= X_t + \sqrt{\frac{d[\sigma^2(t)]}{dt}} dB$$

$$\Rightarrow dX = \sqrt{\frac{d[\sigma^2(t)]}{dt}} dB.$$

解得.  $X = X_0 + \int_0^t \frac{d[\sigma^2(u)]}{du} du \rightarrow,$

$= X_0 + \sigma^2(t)$ . 符合原来离散形式.

DDPM:  $X_i = \sqrt{1-\beta_i} X_{i-1} + \sqrt{\beta_i} \varepsilon, \quad i=1, \dots, N$

取  $\beta(t) = N\beta/N$ ,  $\Delta t = \frac{1}{N}$

$$\Delta t = \frac{1}{N}$$

$$\boxed{\beta_i = \beta_0 + (t-i)\Delta t}$$

$$\text{Rif } X_t = \sqrt{1-\beta(t)\Delta t} X_{t-\Delta t} + \sqrt{\beta(t)\Delta t} \varepsilon$$

$$\approx [1 - \frac{1}{2}\beta(t)\Delta t] X_{t-\Delta t} + \sqrt{\beta(t)} dB$$

$$\begin{aligned} h(x) &= \sqrt{1-x} \\ h(x) &= h(0) + \frac{h'(0)}{1!}(x-0) \\ &\approx 1 - \frac{1}{2}x \end{aligned}$$

$$\therefore dX = -\frac{1}{2}\beta(t)X dt + \sqrt{\beta(t)}dB$$

解得  $X = X_0 \exp\left(\int_0^t -\frac{1}{2}\beta(s)du\right) + \int_0^t \exp\left(\int_u^t -\frac{1}{2}\beta(s)ds\right) \sqrt{\beta(u)} dB$

$$= X_0 \exp\left(\sum_{u=1}^t -\frac{1}{2}\beta(u)du\right) + \sum_{u=1}^t \exp\left(\sum_{s=u+1}^t -\frac{1}{2}\beta(s)ds\right) \sqrt{\beta(u)} dB$$

$$= X_0 \prod_{u=1}^t \exp\left(-\frac{1}{2}\beta(u)du\right) + \sum_{u=1}^t \left( \prod_{s=u+1}^t \exp\left(-\frac{1}{2}\beta(s)ds\right) \right) \sqrt{\beta(u)} dB$$

$$= X_0 \prod_{u=1}^t \sqrt{1-\beta(u)} du + \sum_{u=1}^t \left( \prod_{s=u+1}^t \sqrt{1-\beta(s)} ds \right) \sqrt{\beta(u)} dB$$

$$= X_0 \cdot \sqrt{\alpha_t} + \sum_{n=1}^t \left( \prod_{s=n+1}^t \sqrt{\alpha_s} \right) \sqrt{\beta_n} dB$$

$$= X_0 \sqrt{\alpha_t} + \underbrace{\sum_{n=1}^t \left( \prod_{s=n+1}^t \sqrt{\alpha_s} \right) \sqrt{\beta_n} \varepsilon_n}_{\text{(含高斯噪声)}} \quad (\text{含高斯噪声})$$

合并高斯噪声

$$n=0. \quad \sigma^2 = \alpha_2 - \alpha_t(1-\alpha_1) + \alpha_3 \cdots \alpha_t(1-\alpha_2) + \cdots + 1 - \alpha_t.$$

$$= \cancel{\alpha_2 \cdots \alpha_t} - \alpha_1 \cdots \alpha_t + \cancel{\alpha_3 \cdots \alpha_t} - \cancel{\alpha_1 \cdots \alpha_t} + \cdots + 1 - \cancel{\alpha_t}.$$

$$= 1 - \alpha_1 \cdots \alpha_t = 1 - \bar{\alpha}_t$$

$$= \sqrt{\bar{\alpha}_t} X_0 + \sqrt{1 - \bar{\alpha}_t} \varepsilon_t \quad \text{符合原式.}$$

$$dx = -\frac{1}{2} \beta(t) dt + \sqrt{\beta(t)} dB \rightarrow \text{逆向} \Rightarrow dx = \left[ -\frac{1}{2} \beta(t) X - \beta(t) \nabla \log p(x) \right] dt + \sqrt{\beta(t)} dB$$

$$\begin{aligned} X_{t-dt} &= X_t + \frac{1}{2} \beta(t) dt X_t + \beta(t) dt \nabla \log p(x) + \sqrt{\beta(t)} dB \\ &= (1 + \frac{1}{2} \beta(t) dt) X_t + \beta(t) dt \nabla \log p(x) + \sqrt{\beta(t)} dB. \\ &= \frac{1}{\sqrt{1 - \beta(t) dt}} (X_t + \beta(t) dt \cdot \sqrt{1 - \beta(t) dt} \nabla \log p(x)) + \text{noise}. \\ &= \frac{1}{\sqrt{1 - \beta(t) dt}} (X_t + \beta(t) dt \cdot (1 - \frac{1}{2} \beta(t) dt) \text{Score}) + \text{noise}. \end{aligned}$$

$$= \frac{1}{\sqrt{1 - \beta(t) dt}} (X_t + \beta(t) dt \text{Score}) + \text{noise}.$$

$$= \frac{1}{\sqrt{\alpha_t}} (X_t + \beta_t \text{Score}) + \sqrt{\beta_t} \varepsilon.$$

$$\begin{aligned}
\text{Score} &= \nabla_{x_t} \log p(x_t) \\
&= \frac{1}{p(x_t)} \nabla p(x_t) = \frac{1}{p(x_t)} \nabla_x \int p(x_t | x_0) p(x_0) dx_0 \\
&= \frac{1}{p(x_t)} \int \nabla_{x_t} p(x_t | x_0) p(x_0) dx_0 \\
&= \frac{1}{p(x_t)} \int [\nabla_{x_t} \log p(x_t | x_0)] p(x_t | x_0) p(x_0) dx_0 \\
&= \frac{1}{p(x_t)} \int \left[ \nabla_{x_t} \left( \frac{-(x_t - u)^2}{2\sigma^2} \right) \right] p(x_t, x_0) dx_0 \\
&= \frac{1}{p(x_t)} \int -\frac{\bar{\epsilon}}{\sigma} p(x_t, x_0) dx_0 \\
&= -\frac{\bar{\epsilon}}{\sigma} = -\frac{\bar{\epsilon}}{\sqrt{1-\alpha_t}}
\end{aligned}$$

$$X_{t+1} = \frac{1}{\sqrt{\alpha_t}} (X_t - \beta \frac{\bar{\epsilon}}{\sqrt{1-\alpha_t}}) + \sqrt{\beta_t} \bar{\epsilon}$$

~~prop prop~~

根元单流 ODE:

$$dx = f(x, t) dt + g(t) dB \rightarrow \text{逆向 } dx = [f - g^2 \text{score}] dt + \text{noise}$$

F-K 方程:

$$\frac{\partial p(x_t, t)}{\partial t} = -\frac{\partial (f(x_t, t) p(x_t, t))}{\partial x} + \frac{1}{2} g^2 \frac{\partial^2 p(x_t, t)}{\partial x^2}$$

等价于  $dx = (f(x, t) - \frac{1}{2}(g^2 - \sigma^2) \nabla \log p(x)) dt + \cancel{\sigma} \sigma dB$   
 $\forall \sigma(t) < g(t)$

反向:  $dx = [f(x, t) - \frac{1}{2}(g^2 + \sigma^2) \nabla \log p(x)] dt + \sigma dB$

$\sigma \rightarrow 0$ .

得 ~~ODE~~  $= dx = [f(x, t) - \frac{1}{2} g^2 \nabla \log p(x)] dt$

DDIM.

$$L = \text{Ex}_{x_0} \left[ \left( \mathbb{E}_{\theta}(\sqrt{\bar{\alpha}_t} x_t + \sqrt{1 - \bar{\alpha}_t} \epsilon) - \epsilon \right)^2 \right]$$

~~直接~~ 简单变换易懂

$$\begin{aligned} L &= \text{Ex}_{x_t, x_0} \left[ \left( s_{\theta}(x_t) - \nabla \log p(x_t | x_0) \right)^2 \right] \\ &= k \epsilon_0 \\ &= -\frac{\epsilon}{\sigma} \end{aligned}$$

$$\begin{aligned} \nabla_{x_t} \log p(x_t) &= \frac{1}{p(x_t)} \nabla_{x_t} \int p(x_t | x_0) p(x_0) dx_0 \\ &= \frac{1}{p(x_t)} \int [\nabla_{x_t} p(x_t | x_0)] p(x_0) dx_0 \\ &= \frac{1}{p(x_t)} \int \nabla_{x_t} \log p(x_t | x_0) p(x_t, x_0) dx_0 \\ &= \text{Ex}_0[x_t] [\nabla_{x_t} \log p(x_t | x_0)] \end{aligned}$$

$$\Rightarrow L = \text{Ex}_{x_t, x_0} = \text{Ex}_t \text{Ex}_0[x_t]$$

也就是说 DDPM 实际上在学  $\nabla_{x_t} \log p(x_t)$

$$\text{有 } \nabla_{x_t} \log p(x_t) = -\frac{\epsilon}{\sigma} = -\frac{\epsilon_{\theta}(x_t)}{\sqrt{1 - \bar{\alpha}_t}}$$

而这个过程中只需要用到  $p(x_t | x_0)$ .

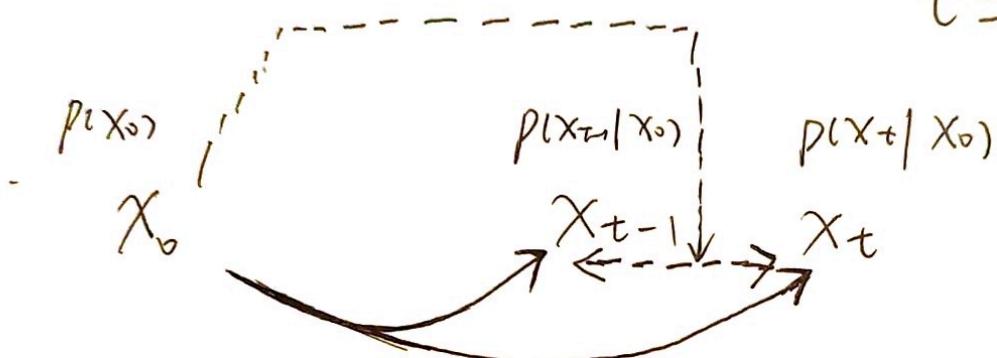
没有用到  $p(x_t | x_{t-1})$  这个 Markov 前向

也就是说我们只要保持边缘分布  $p(x_t | x_0)$  就可以而不用考虑  $p(x_{t-1} | x_t)$  或  $p(x_t | x_{t-1})$  是否 markov.

定义  $\bar{x}_t, t=1, \dots, T$

边缘分布  $p(x_t | x_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t} x_0, (1-\bar{\alpha}_t)I)$

$t = 1, \dots, T$



显然光有边缘分布是没办法进行采样的  
我们仍需要  $x_t \rightarrow x_{t-1}$  的变换.

$$\int \underbrace{p(x_{t-1} | x_t, x_0)}_{p(x_{t-1} | x_t, x_0)} p(x_t | x_0) dx_0 = p(x_{t-1} | x_0)$$

$p(x_{t-1} | x_t, x_0)$  需满足边缘分布固定.

$$p(x_t | x_0) = X_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t} \varepsilon_t$$

$$p(x_{t-1} | x_t, x_0) := X_{t-1} = k_1 X_t + k_2 x_0 + \sigma \varepsilon$$

$$\begin{aligned} \text{消去 } X_t \\ X_{t-1} &= k_1 (\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t} \varepsilon_t) \\ &\quad + k_2 x_0 + \sigma \varepsilon \\ &= (k_1 \sqrt{\bar{\alpha}_t} + k_2) x_0 + k_1 \sqrt{1-\bar{\alpha}_t} \varepsilon_t + \sigma \varepsilon \end{aligned}$$

$$P(X_{t-1} | X_0) : X_{t-1} = \sqrt{\bar{\alpha}_{t-1}} X_0 + \sqrt{1 - \bar{\alpha}_{t-1}} \varepsilon_{t-1}$$

$$\begin{cases} k_1 \sqrt{\bar{\alpha}_t} + k_2 = \sqrt{\bar{\alpha}_{t-1}} \\ (k_1 \sqrt{1 - \bar{\alpha}_t})^2 + \delta^2 = (\sqrt{1 - \bar{\alpha}_{t-1}})^2 \end{cases}$$

$$\Rightarrow \begin{cases} k_1 = \frac{\sqrt{1 - \bar{\alpha}_{t-1} - \delta^2}}{\sqrt{1 - \bar{\alpha}_t}} \\ k_2 = \sqrt{\bar{\alpha}_{t-1}} - \sqrt{\bar{\alpha}_t} \frac{\sqrt{1 - \bar{\alpha}_{t-1} - \delta^2}}{\sqrt{1 - \bar{\alpha}_t}} \end{cases}$$

$$\Rightarrow P(X_{t-1} | X_t, X_0) = N(\mu, \sigma^2 I)$$

$$= N(k_1 X_t + k_2 X_0, \sigma^2 I)$$

$$= N(\sqrt{\bar{\alpha}_{t-1}} X_0 + \sqrt{1 - \bar{\alpha}_{t-1} - \delta^2} \frac{X_t - \sqrt{\bar{\alpha}_t} X_0}{\sqrt{1 - \bar{\alpha}_t}}, \sigma^2 I)$$

$$X_0 = \frac{X_t - \sqrt{1 - \bar{\alpha}_t} \varepsilon_0(X_t, t)}{\sqrt{\bar{\alpha}_t}}$$

$$\Rightarrow X_{t-1} = \underbrace{\sqrt{\bar{\alpha}_{t-1}} \frac{X_t - \sqrt{1 - \bar{\alpha}_t} \varepsilon_0(X_t, t)}{\sqrt{\bar{\alpha}_t}}}_{\text{predict } X_0} + \underbrace{\sqrt{1 - \bar{\alpha}_{t-1} - \delta^2} \varepsilon_0(X_t, t) + \bar{\delta} \varepsilon}_{\text{point to } X_t}$$

加速采样：

在推导边际分布的维持中，没有归一化

必须为  $X_t \rightarrow X_{t-1}$ .

$$P(X_t | X_0) = \sqrt{\alpha_t} X_0 + \sqrt{1-\alpha_t}$$

$$P(X_t | X_0) = X_t = \sqrt{\alpha_t} X_0 + \sqrt{1-\alpha_t} \epsilon_t$$

考虑  $\tau_1 < \tau_2$ .

$$\cancel{P(\tau_1 | \tau_2)}$$

$$\int P(X_{\tau_1} | X_{\tau_2}, X_0) P(X_{\tau_2} | X_0) dX_0 = P(X_{\tau_1} | X_0)$$

$$\Rightarrow X_{\tau_1} = \sqrt{\alpha_{\tau_1}} X_0 + \sqrt{1-\alpha_{\tau_1}-\sigma^2} \frac{X_{\tau_2} - \sqrt{\alpha_{\tau_2}} X_0}{\sqrt{1-\alpha_{\tau_2}}} + \sigma \xi$$

$$= \sqrt{\alpha_{\tau_1}} \frac{X_{\tau_2} - \sqrt{1-\alpha_{\tau_2}} \epsilon_0(X_{\tau_2}, \tau_2)}{\sqrt{\alpha_{\tau_2}}} + \sqrt{1-\alpha_{\tau_1}-\sigma^2} \epsilon_0(X_{\tau_2}, \tau_2) + \sigma \xi$$

即可选  $(1, 2, \dots, T)$  的子序列  $(\tau_1, \tau_2, \dots, \tau_s)$

进行采样.

考虑  $\sigma_t = 0$ . 即.

$$X_{t-dt} = \sqrt{\alpha_{t-dt}} \frac{X_t - \sqrt{1-\alpha_t} \xi_0}{\sqrt{\alpha_t}} + \sqrt{1-\alpha_{t-dt}} \xi_0.$$

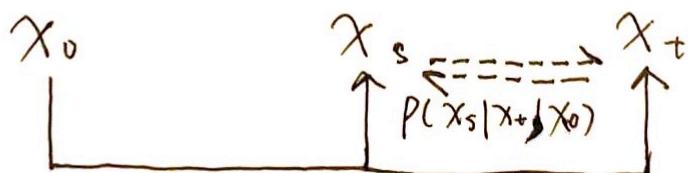
整理有

$$\frac{X_{t-d\tau}}{\sqrt{\alpha_{t-d\tau}}} = \frac{X_t}{\sqrt{\alpha_t}} + \left( \sqrt{\frac{1-\bar{\alpha}_{t-d\tau}}{\bar{\alpha}_{t-d\tau}}} - \sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}} \right) \epsilon_\theta.$$

$$\Rightarrow d\left(\frac{X_t}{\sqrt{\alpha_t}}\right) = \epsilon_\theta d\left(\sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}\right)$$

即 DDIM 在  $\sigma=0$  时变成 ODE

OK. 继续着 DDIM 的思路. 我们看一种特殊情况  $= p(x_t | x_0) = X_t = \alpha(t) X_0 + \bar{\alpha}(t) \epsilon_t$



$$\int p(x_s | x_t, x_0) p(x_t | x_0) dx_0 = p(x_s | x_t)$$
$$x_s = k_1 x_t + k_2 x_0 + \cancel{\alpha} + \cancel{\bar{\alpha}} + \gamma \epsilon$$
$$+ \gamma \epsilon$$

$$= k_1 (\alpha(t) x_0 + \bar{\alpha}(t) \epsilon_t) + k_2 x_0 + \gamma \epsilon$$

$$= (k_1 \alpha(t) + k_2) x_0 + k_1 \bar{\alpha}(t) \epsilon_t + \gamma \epsilon$$

$$P(X_s|x_0) = X_s = \alpha(s, x_0 + \sigma(s)) \varepsilon,$$

$$\Rightarrow \begin{cases} k_1 \alpha(t) + k_2 = \alpha(s) \\ [k_1 \sigma(t)]^2 + r^2 = \sigma(s)^2 \end{cases}$$

$$\Rightarrow \begin{cases} k_1 = \frac{\sqrt{\sigma(s)^2 - r^2}}{\sigma(t)} \\ k_2 = \alpha(s) - \frac{\alpha(t)}{\sigma(t)} \sqrt{\sigma(s)^2 - r^2} \end{cases}$$

$$P(X_s|x_t, x_0) = X_s = k_1 x_t + k_2 x_0 + r \varepsilon$$

$$= \alpha(s, x_0 + \sqrt{\sigma(s)^2 - r^2} \frac{x_t - \alpha(t)x_0}{\sigma(t)}) + r \varepsilon$$

$$\text{令 } \tilde{x}_0 = X_0^\theta(x_t, t) = \frac{x_t - \sigma(t) e_\theta(x_t, t)}{\alpha(t)}$$

$$P_\theta(X_s|x_t) = P(X_s|x_t, X_0^\theta(x_t, t))$$

$$KL(P(X_s|x_t, x_0) || P_\theta(X_s|x_t)) = KL(P(X_s|x_t, x_0) || P(X_s|x_t, X_0^\theta(x_t, t)))$$

$$= \frac{k_2^2 ||x_0 - X_0^\theta(x_t, t)||}{2r^2} \quad \text{代入 } x_0 \text{ 和 } X_0^\theta$$

$$E_{x_0, x_t}(KL) = E_{x_0, x_t} \frac{k_2^2 ||x_0 - X_0^\theta(x_t, t)||}{2r^2} = E_{x_0, \varepsilon} \left[ \frac{k_2^2 \sigma^2(t)}{2r^2 \alpha^2(t)} ||\varepsilon - e_\theta||^2 \right]$$

$$\text{也就说 } P(X_t|x_0) = X_t = \alpha(t, x_0 + \sigma(t)) \varepsilon$$

不用保证  $\alpha(t)^2 + \sigma(t)^2$  也能用 Denoising Score Matching  
的一个DDPM(DDIM)

再看看 SDE

$$dx = f(t)x dt + g(t)dB$$

解得  $X(t) = X_0 \exp(\int_0^t f(s)ds) + \int_0^t \exp(\int_u^t f(s)ds) g(s) dB$

而我们有  $X(t) = \alpha(t)X_0 + \beta(t)\varepsilon$

$$\Rightarrow \alpha(t) = \exp(\int_0^t f(s)ds) \Rightarrow f(t) = \frac{d \ln \alpha(t)}{dt}$$

(PS: 如果有  $f(s) = -\frac{1}{2}\beta(s)$ , 然后离散化 + taylor.)  
且有  $\alpha(t) = \sqrt{\alpha_0} e^{\int_0^t f(u)du}$   
 $\alpha(t) = \sqrt{1 - \beta^2} = \sqrt{\alpha_0}$ )

$$\sigma^2(t) = \int_0^t \left[ \frac{\alpha(t)}{\alpha(u)} \right]^2 g^2(u) du$$

$$\frac{d\sigma^2(t)}{dt} = \frac{d \int_0^t \left( \frac{\alpha(t)}{\alpha(u)} \right)^2 g^2(u) du}{dt} = \frac{d \left[ \alpha(t) \int_0^t \frac{g^2(u)}{\alpha(u)} du \right]}{dt}$$

$$= 2\alpha(t) \frac{d\alpha(t)}{dt} \int_0^t \left[ \frac{g^2(u)}{\alpha(u)} \right] du + \alpha^2(t) \left( \frac{g(t)}{\alpha(t)} \right)^2$$

$$= 2 \frac{1}{\alpha(t)} \frac{d\alpha(t)}{dt} \int_0^t \alpha^2(t) \frac{g^2(u)}{\alpha^2(u)} du + g^2(t)$$

$$= 2 \frac{d \ln \alpha(t)}{dt} \sigma^2(t) + g^2(t)$$

$$\Rightarrow \begin{cases} f(t) = \frac{d \ln \alpha(t)}{dt} \\ g^2(t) = \frac{d\sigma^2(t)}{dt} - 2 \frac{d \ln \alpha(t)}{dt} \sigma^2(t) \end{cases}$$

我们已知  $dx = f(x, t) dt + g(t) dB$   
的逆向 SDE 为

$$dx = [f(x, t) - g^2(t) \text{Score}] dt + g(t) dB$$

逆向 probability flow SDE 为

$$dx = [f(x, t) - \frac{1}{2}g^2(t) \text{Score}] dt$$

$$\begin{cases} f(x, t) = x \cdot \frac{d \ln \alpha(t)}{dt} \\ g^2(t) = \frac{d \sigma^2(t)}{dt} - 2 \frac{d \ln \alpha(t)}{dt} \sigma^2(t) \end{cases}$$

$$dx = [x \cdot \frac{d \ln \alpha(t)}{dt} - \frac{1}{2}(\frac{d \sigma^2(t)}{dt} - 2 \frac{d \ln \alpha(t)}{dt} \sigma^2(t)) \text{Score}] dt$$

$$\frac{dx}{dt} - x \cdot \frac{d \ln \alpha(t)}{dt} = -\frac{1}{2} \left[ \frac{d \sigma^2(t)}{dt} - 2 \frac{d \ln \alpha(t)}{dt} \sigma^2(t) \right] \text{Score}$$

$$\text{先看左边 } \frac{dx}{dt} - x \cdot \frac{d \ln \alpha(t)}{dt} = \frac{dx}{dt} - x \cdot \frac{1}{\alpha(t)} \frac{d \alpha(t)}{dt}$$

$$= \alpha(t) \left[ \frac{d(x)}{dt} \cdot \frac{1}{\alpha(t)} + x \cdot \left( -\frac{1}{\alpha^2(t)} \right) \cdot \frac{d \alpha(t)}{dt} \right]$$

$$= \alpha(t) \cdot \frac{d(\frac{x}{\alpha(t)})}{dt}$$

$$\text{再看右边 } -\frac{1}{2} \left[ \frac{d \sigma^2(t)}{dt} - 2 \frac{d \ln \alpha(t)}{dt} \sigma^2(t) \right] \text{Score}$$

$$= -\frac{1}{2} \left[ 2\sigma(t) \frac{d \sigma(t)}{dt} - 2 \frac{1}{\alpha(t)} \frac{d \alpha(t)}{dt} \sigma^2(t) \right] \text{Score}$$

$$= -\alpha(t) \sigma(t) \left[ \frac{d\sigma(t)}{dt} \cdot \frac{1}{\alpha(t)} + \sigma(t) \cdot \left( -\frac{1}{\alpha^2(t)} \right) \frac{d\alpha(t)}{dt} \right] \text{Score}$$

$$= -\alpha(t) \sigma(t) \frac{d \left[ \frac{\sigma(t)}{\alpha(t)} \right]}{dt} \text{Score}$$

$$\Rightarrow \alpha(t) \cdot d \left[ \frac{x}{\alpha(t)} \right] = -\alpha(t) \sigma(t) d \left[ \frac{\sigma(t)}{\alpha(t)} \right] \text{Score} .$$

$$\text{Score} = \sum_{t+1} \log p(x_t) = -\frac{\epsilon}{\sigma} = -\frac{\epsilon \theta(x_t, t)}{\sigma(t)}$$

$$\Rightarrow d \left[ \frac{x}{\alpha(t)} \right] = \epsilon_\theta d \left[ \frac{\sigma(t)}{\alpha(t)} \right]$$

再看  $x_s = \sqrt{\alpha_s} \frac{x_t - \alpha(t)x_0}{\sqrt{\sigma_t}} + \epsilon$

$$x_s = \alpha(s) x_0 + \sqrt{\sigma^2(s) - r^2} \frac{x_t - \alpha(t)x_0}{\sigma(t)} + \gamma \epsilon$$

$$\text{令 } r=0, \text{ 并代入 } x_0 = \frac{x_t - \sigma(t)\epsilon_\theta}{\alpha(t)}$$

$$x_s = \alpha(s) \frac{x_t - \sigma(t)\epsilon_\theta}{\alpha(t)} + \frac{\sigma(s) - \sigma(t)}{\sigma(t)} \cdot \alpha(s) \cdot \sigma(s) \epsilon_\theta$$

$$\Rightarrow \frac{x_s}{\alpha(s)} - \frac{x_t}{\alpha(t)} = \left[ \frac{\sigma(s)}{\alpha(s)} - \frac{\sigma(t)}{\alpha(t)} \right] \epsilon_\theta$$

$$\Rightarrow d \left[ \frac{x}{\alpha(t)} \right] = \epsilon_\theta d \left[ \frac{\sigma(t)}{\alpha(t)} \right]$$

如果令  $\alpha^2(t) + \sigma^2(t) = 1 \Rightarrow \cancel{\text{DDPM}} \text{ 以及 } \cancel{\text{ODE}}$

可以看到 DDPM 的 ODE  
就是 DDPM 的 DDE

$$X_t = \alpha_t + X_0 + \sigma_t \epsilon, \epsilon \sim N(0, 1)$$

$$\text{令 } Z_t = \frac{X_t}{\alpha_t} \cdot = Z_0 + \frac{\sigma_t}{\alpha_t} \epsilon$$

按照 NCSN 的思路有

$$dZ = \sqrt{\frac{d(\frac{\sigma_t}{\alpha_t})^2}{dt}} dB$$

$$\text{即 } g(z) = \frac{d[\frac{\sigma_t}{\alpha_t}]^2}{dt} = 2 \frac{\sigma_t}{\alpha_t} \cdot \frac{d[\frac{\sigma_t}{\alpha_t}]}{dt}$$

则对应 ODE 有

$$d\dot{z} = -\frac{1}{2} g(z), \nabla_z \log p_z(z), dt$$

$$= -\frac{\sigma_t}{\alpha_t} d(\frac{\sigma_t}{\alpha_t}) \nabla_z \log p_z(z)$$

$$\begin{aligned} z_t &= \frac{x_t}{\alpha_t} \quad . \quad p_z(z) = \frac{dF_z(z)}{dz} = \frac{dF_x(f^{-1}(z))}{dz} \\ &= \frac{dF_x(f^{-1}(z))}{df^{-1}(z)} \cdot \frac{df^{-1}(z)}{dz} \\ &= p_x(\alpha_t z) \cdot \alpha_t \end{aligned}$$

$$\begin{aligned} \nabla_z \log p_z(z) &= \nabla_z [\alpha_t z] \nabla_x \log p_x(x) + \underbrace{\partial_x \log}_{\partial_x \log \alpha_t} \alpha_t \\ &= \alpha_t \nabla_x \log p_x(x) \end{aligned}$$

$$\alpha_t \left[ -\frac{\sigma_t}{\alpha_t} \nabla_z \log p_z(z) \right] = \hat{f}_p(x, t)$$

$$d(\frac{x}{\alpha_t}) = \epsilon d(\frac{\sigma_t}{\alpha_t})$$