DDPM

1. Forward:

$$q(x_t \mid x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}\, x_{t-1}, \beta_t I) \qquad \beta_t \in (0,1) \qquad (1)$$

$$q(x_{1:T} \mid x_0) = \prod_{t=1}^{T} q(x_t \mid x_{t-1})$$

$\Rightarrow$ define $\alpha_t = 1-\beta_t$. $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$

$$x_t = \sqrt{\alpha_t}\, x_{t-1} + \sqrt{1-\alpha_t}\, z_{t-1}, \quad z_t \sim \mathcal{N}(0, I)$$

$$= \sqrt{\alpha_t \alpha_{t-1}}\, x_{t-2} + \sqrt{\alpha_t} \cdot \underbrace{\sqrt{1-\alpha_{t-1}}\, z_{t-2}}_{\text{Standard normal}} + \underbrace{\sqrt{1-\alpha_t}\, z_{t-1}}_{\text{standard normal.}}$$

merge two gaussian

$$= \sqrt{\alpha_t \alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_t \alpha_{t-1}}\, \bar{z}_{t-2}, \quad \bar{z}_{t-2} \sim \mathcal{N}(0, I)$$

$$= \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, z$$

$$\Rightarrow q(x_t \mid x_0) = \mathcal{N}(x_0; \sqrt{\bar{\alpha}_t}, (1-\bar{\alpha}_t) I) \qquad (2)$$

also $0 < \beta_1 < \beta_2 < \cdots < \beta_T < 1$, hence. $\bar{\alpha}_1 > \bar{\alpha}_2 > \cdots > \bar{\alpha}_T$

2. Reverse. (learn p_θ)

$$p_\theta(x_{0:T}) = p_\theta(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1} \mid x_t)$$

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

$$q(x_{t-1} \mid x_t, \underset{\text{given}}{x_0}) = q(x_t \mid x_{t-1}, x_0) \cdot \frac{q(x_{t-1} \mid x_0)}{q(x_t \mid x_0)} \quad \cdots\text{ Bayes.}$$

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

$$\Rightarrow \tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t. \qquad \tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} x_0$$

$$x_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \sqrt{1-\bar{\alpha}_t}\, z_t) \nearrow \quad = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} z_t\right)$$

3. Loss.

$\boxed{1}$. $$-\log p_\theta(x_0) \le -\log p_\theta(x_0) + KL(q(x_{1:T} \mid x_0) \,\|\, p(x_{1:T} \mid x_0))$$

$$= -\log p_\theta(x_0) + E_q\left[\log \frac{q(x_{1:T}|x_0)}{p(x_{0:T})/p(x_0)}\right]$$

$$= -\log p_\theta(x_0) + E_q\left[\log \frac{q(x_{1:T}|x_0)}{p(x_{0:T})}\right] + \underbrace{E_{x_{1:T}\sim q}\left[\log p(x_0)\right]}_{=\log p(x_0)}$$

$$= E_q\left[\log \frac{q(x_{1:T}|x_0)}{p(x_{0:T})}\right]$$

let $L_{VLB} = E_{q(x_{0:T})}\left[\log \frac{q(x_{1:T}|x_0)}{p(x_{0:T})}\right] \geq -E_{q(x_0)} \log p_\theta(x_0)$

$\boxed{2}$ same as $\boxed{1}$ $L_{CE} = -E_{q(x_0)} \log p_\theta(x_0)$ minimize cross entropy.

$$= -E_{q(x_0)} \log\left(\int p_\theta(x_{0:T})\, dx_{1:T}\right)$$

$$= -E_{q(x_0)} \log\left(\int q(x_{1:T}|x_0) \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\, dx_{1:T}\right)$$

Jensen's inequality

$$\leq -E_{q(x_0)} \int q(x_{1:T}|x_0) \log \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\, dx_{1:T}$$

$$= -E_{q(x_{0:T})}\left[\log \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right]$$

$$= E_{q(x_{0:T})}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] = L_{VLB}.$$

Note $L_{VLB} = E_{q(x_{0:T})}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] \geq -E_{q(x_0)}(\log p(x_0))$

$$L_{VLB} = E_q\left[\log \frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p_\theta(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}\right]$$

forward

reverse.

$$= E_q\left[-\log p_\theta(x_T) + \sum_{t=1}^{T} \log \frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}\right]$$

$$= E_q\left[-\log p_\theta(x_T) + \log \frac{q(x_1|x_0)}{p_\theta(x_0|x_1)} + \sum_{t=2}^{T} \log \frac{q}{p_\theta}\right]$$

Bayes

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

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

$$= \log \frac{q(x_T|x_0)}{q(x_1|x_0)}$$

$$= \mathbb{E}_q\left[\log \frac{q(x_T|x_0)}{p_\theta(x_T)} + \sum_{t\geq 2}^{T} \log \frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)} - \log p_\theta(x_0|x_1)\right]$$

?

$$= \mathbb{E}_q\Big[\underbrace{KL(q(x_T|x_0)\,\|\,p_\theta(x_T))}_{L_T} + \sum_{t=2}^{T} \underbrace{KL(q(x_{t-1}|x_t, x_0)\,\|\,p_\theta(x_{t-1}|x_t))}_{L_{t-1}} \underbrace{- \log p_\theta(x_0|x_1)}_{L_0}\Big]$$

$$\Rightarrow L_{simple} = \mathbb{E}_{x_0, x_t}\left[\| z_t - z_\theta(\sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, z_t, t)\|^2\right] + C.$$

Training.

1. Repeat.
2. $x_0 \sim q(x_0)$
3. $t \sim \text{Uniform}(\{1, \dots, T\})$
4. $\epsilon \sim \mathcal{N}(0, I)$
5. Take gradient descent step on
   $\nabla_\theta \| \epsilon - z_\theta(\sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon, t)\|^2$ ← Loss update.
6. until converge

Sampling

1. $x_T \sim \mathcal{N}(0, I)$
2. for $t = T \dots 1$ do
3. $z \sim \mathcal{N}(0, I)$ if $t > 1$, else $z = 0$
4. $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} z_\theta(x_t, t)\right) + \sigma_t z$ ← reverse diffusion
5. end for
6. return x_0.