Denoising Diffusion Implicit Models (ICLR'21)

大多人用它来加速采样

复习DDPM

$$\boxed{q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar\alpha_t}x_0, (1-\bar\alpha_t)I)}$$ diffusion 分布

$$x_t = \sqrt{\bar\alpha_t}\,x_0 + \sqrt{1-\bar\alpha_t}\,\varepsilon \quad \varepsilon\sim\mathcal{N}(0,I)$$

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

马尔可夫链 || 与x_0无关

$$\boxed{q(x_{t-1}|x_t,x_0) = \frac{q(x_t|x_{t-1})\,q(x_{t-1}|x_0)}{q(x_t|x_0)} = \mathcal{N}(x_{t-1}; \mu(x_t,x_0), \sigma_t^2 I)}$$ Reverse 分布

用到了x_0，但是我们并不知道x_0

于是用x_t去估计x_0

反写公式：$\hat{x}_0 = \frac{1}{\sqrt{\bar\alpha_t}}(x_t - \sqrt{1-\bar\alpha_t}\,\varepsilon_\theta)$

训练的UNet

用于估计噪声ε

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

$$\mu(x_t,x_0) = \frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})}{1-\bar\alpha_t}x_t + \frac{\sqrt{\bar\alpha_{t-1}}\,\beta_t}{1-\bar\alpha_t}x_0$$

$$\sigma_t^2 = \frac{1-\bar\alpha_{t-1}}{1-\bar\alpha_t}\cdot\beta_t$$

DDPM 采样分两步：① $\hat{x}_{0|t} = \frac{1}{\sqrt{\bar\alpha_t}}\left(x_t - \sqrt{1-\bar\alpha_t}\,\varepsilon_\theta(x_t,t)\right)$

② $x_{t-1} = \mu(x_t, \hat{x}_{0|t}) + \sigma_t\varepsilon \quad \varepsilon\sim\mathcal{N}(0,I)$

DDIM

$$q(x_{t-1}|x_t,x_0) = \sqrt{\bar\alpha_{t-1}}\,x_0 + \sqrt{1-\bar\alpha_{t-1}}\,\varepsilon \quad \varepsilon\sim\mathcal{N}(0,I)$$

$q(x_{t-1}|x_0)$

$$= \sqrt{\bar\alpha_{t-1}}\frac{1}{\sqrt{\bar\alpha_t}}\left(x_t - \sqrt{1-\bar\alpha_t}\,\varepsilon_\theta(x_t,t)\right) + \sqrt{1-\bar\alpha_{t-1}}\,\varepsilon$$

始终没用到 $q(x_t|x_{t-1})$

$$= \sqrt{\bar\alpha_{t-1}}\,\hat{x}_{0|t} + \sqrt{1-\bar\alpha_{t-1}}\,\varepsilon$$

$$= \sqrt{\bar\alpha_{t-1}}\,\hat{x}_{0|t} + \sqrt{1-\bar\alpha_{t-1}}\,\varepsilon_\theta(x_t,t)$$

$$= \sqrt{\bar\alpha_{t-1}}\,\hat{x}_{0|t} + \sqrt{1-\bar\alpha_{t-1}-\sigma_t^2}\,\varepsilon_\theta + \sigma_t\varepsilon$$

$$\left(\sqrt{1-\bar\alpha_{t-1}-\sigma_t^2}\right)^2 + \sigma_t^2 = \left(\sqrt{1-\bar\alpha_{t-1}}\right)^2$$

$q(x_s|x_k,x_0)$

$$\boxed{x_s = \sqrt{\bar\alpha_s}\,\hat{x}_{0|k} + \sqrt{1-\bar\alpha_s-\sigma_k^2}\,\varepsilon_\theta(x_k,k) + \sigma_k\varepsilon}$$

T= 1000 999 998 ... 3 2 1 0 (DDPM)

T= 1000 ... 900 888 666 233 ... 0

把采样公式变了，能实现任意的 $k\to s$

为什么可以不用 $q(x_t|x_{t-1})$？

① train phase

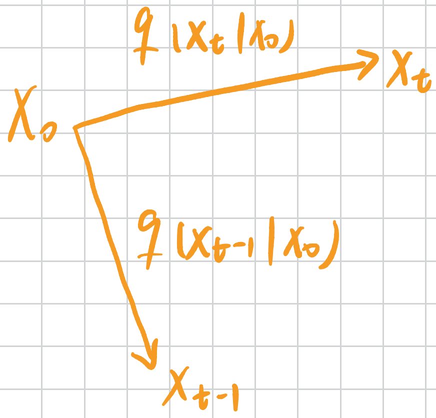
$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}\, x_0, \sqrt{1-\bar{\alpha}_t}\, I)$$

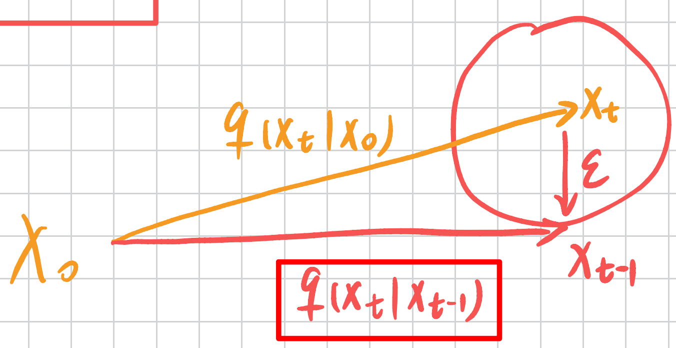
$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t}\, x_{t-1}, (1-\alpha_t) I)$$

② sampling phase

$$q(x_t|x_0) \Rightarrow \hat{x}_{0|t} = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \sqrt{1-\bar{\alpha}_t}\, \varepsilon_\theta(x_t, t))$$

$$\boxed{x_{t-1} = \mu(x_t, \hat{x}_{0|t}) + \sigma_t \varepsilon}$$





DDPM 的这个条件
会给更大的一个限制

$$\text{DDIM} \underset{-q(x_t|x_{t-1})}{\overset{+q(x_t|x_{t-1})}{\rightleftarrows}} \text{DDPM}$$

$q(x_t|x_0)$ $q(x_t|x_0)$ 不能用 $q(x_t|x_{t-1})$

求 $q(x_{t-1}|x_t, x_0)$

$$\begin{cases} x_{t-1} = m_t x_t + n_t x_0 + \sigma_t \varepsilon_1 \\ x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \varepsilon_2 \\ x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\, x_0 + \sqrt{1-\bar{\alpha}_{t-1}}\, \varepsilon_3 \end{cases}$$

$$x_{t-1} = m_t(\sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \varepsilon_2) + n_t x_0 + \sigma_t \varepsilon_1$$

$$= (m_t\sqrt{\bar{\alpha}_t} + n_t) x_0 + m_t\sqrt{1-\bar{\alpha}_t}\, \varepsilon_2 + \sigma_t \varepsilon_1$$

$$\begin{cases} m_t\sqrt{\bar{\alpha}_t} + n_t = \sqrt{\bar{\alpha}_{t-1}} \\ m_t^2(1-\bar{\alpha}_t) + \sigma_t^2 = 1-\bar{\alpha}_{t-1} \end{cases} \qquad n_t = \sqrt{\bar{\alpha}_{t-1}} - \sqrt{\frac{\bar{\alpha}_t}{1-\bar{\alpha}_t}(1-\bar{\alpha}_{t-1}-\sigma_t^2)} \qquad m_t = \frac{1-\bar{\alpha}_{t-1}-\sigma_t^2}{1-\bar{\alpha}_t}$$

$$x_{t-1} = \sqrt{\frac{1-\bar\alpha_{t-1}-\sigma_t^2}{1-\bar\alpha_t}}\, x_t + \left(\sqrt{\bar\alpha_{t-1}} - \sqrt{\frac{\bar\alpha_t}{1-\bar\alpha_t}(1-\bar\alpha_{t-1}-\sigma_t^2)}\right) x_0 + \sigma_t \varepsilon$$

$$= \sqrt{\bar\alpha_{t-1}}\, x_0 + \sqrt{1-\bar\alpha_{t-1}-\sigma_t^2}\left(\frac{1}{\sqrt{1-\bar\alpha_t}} x_t - \frac{\sqrt{\bar\alpha_t}}{\sqrt{1-\bar\alpha_t}} x_0\right) + \sigma_t \varepsilon$$

$$\boxed{= \sqrt{\bar\alpha_{t-1}}\,\underbrace{x_0}_{\hat{x}_{0|t}} + \sqrt{1-\bar\alpha_{t-1}-\sigma_t^2}\;\underbrace{\frac{x_t-\sqrt{\bar\alpha_t}\,x_0}{\sqrt{1-\bar\alpha_t}}}_{\varepsilon_\theta(x_t,t)} + \sigma_t \varepsilon}$$

$\sigma_t = \sigma_{DDPM}$ DDIM $\rightarrow$ DDPM

$q(x_t|x_{t-1})$ $q(x_s|x_k)$

训练目标的角度

DDPM: $\bar\alpha_1$ $\bar\alpha_2$ $\cdots$ $\bar\alpha_t$ $\cdots$ $\bar\alpha_{T-1}$ $\bar\alpha_T$

$\downarrow$

$x_0 \rightarrow x_t$

$\downarrow$

$\varepsilon_\theta(x_1,1)$ $\varepsilon_\theta(x_2,2)$ $\varepsilon_\theta(x_t,t)$ $\cdots$ $\varepsilon_\theta(x_T,T)$

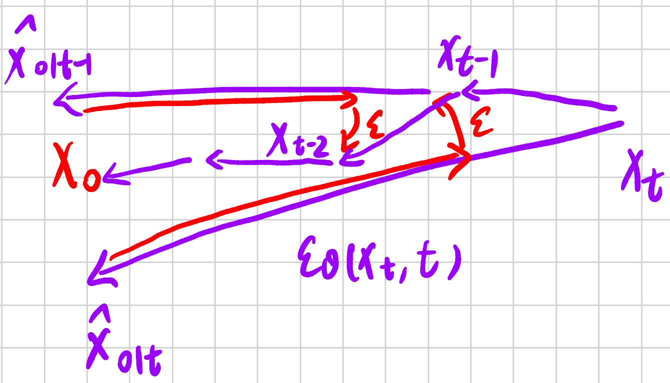
DDIM: $\bar\alpha_{100}$ $\bar\alpha_{200}$ $\cdots$ $\bar\alpha_{900}$ $\bar\alpha_{1000} = \bar\alpha_T$ DDIM参数，训练目标是DDPM的子集

$\varepsilon_\theta(x_{1000},1000)$ $\varepsilon_\theta(x_{2000},2000)$ $\cdots$ $\varepsilon_\theta(x_T,T)$

图示理解

$$x_{t-1} = \underline{\sqrt{\bar\alpha_{t-1}}\,\hat{x}_{0|t}} + \underline{\sqrt{1-\bar\alpha_{t-1}-\sigma_t^2}\;\varepsilon_\theta(x_t,t)} + \sigma_t\varepsilon$$

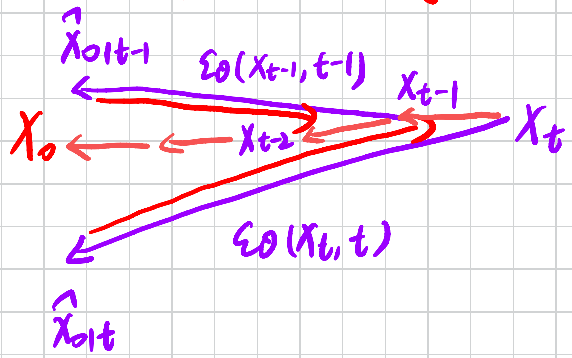
Case 1:



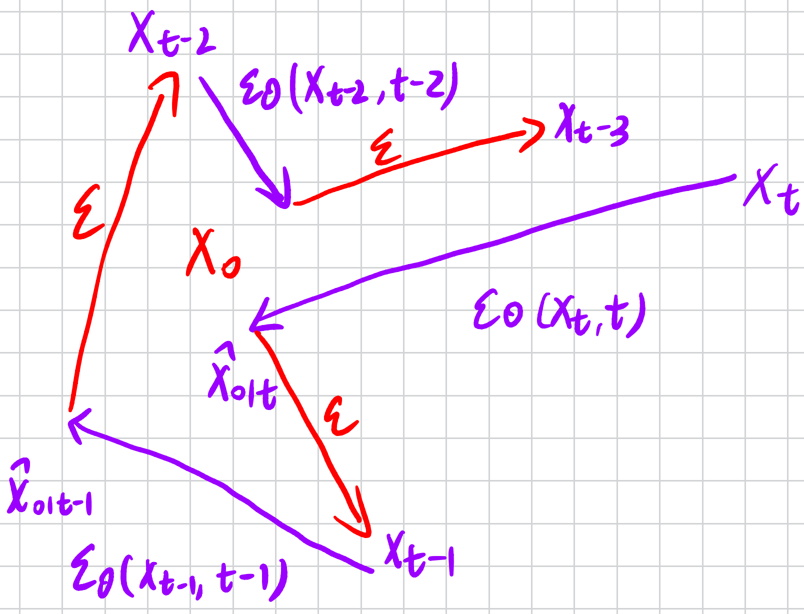
x_t 估计了噪声 $\varepsilon_\theta(x_t,t)$，去除后得到 $\hat{x}_{0|t}$，

再加上比估计噪声小一些的估计噪声，及 $\sigma_t\varepsilon$ 得 x_{t-1}

Case 2 $\sigma_t=0$ 随机性无效了，只与初始采样 x_t 有关



case 3 $\quad \epsilon_\theta$ 系数 $=0$ $\quad$ 剧烈波动



常微分方程

$$x_{t-1}=\sqrt{\bar\alpha_{t-1}}\,x_0+\sqrt{1-\bar\alpha_{t-1}-\sigma_t^2}\,\epsilon_\theta(x_t,t)+\sigma_t\epsilon$$

$\sigma_t=0$
$$x_{t-1}=\sqrt{\bar\alpha_{t-1}}\,x_0+\sqrt{1-\bar\alpha_{t-1}}\,\epsilon_\theta(x_t,t)$$

$$=\sqrt{\bar\alpha_{t-1}}\,\frac{1}{\sqrt{\bar\alpha_t}}\left(x_t-\sqrt{1-\bar\alpha_t}\,\epsilon_\theta(x_t,t)\right)+\sqrt{1-\bar\alpha_{t-1}}\,\epsilon_\theta(x_t,t)$$

$$\frac{x_{t-1}}{\sqrt{\bar\alpha_{t-1}}}=\frac{x_t}{\sqrt{\bar\alpha_t}}-\sqrt{\frac{1-\bar\alpha_t}{\bar\alpha_t}}\,\epsilon_\theta(x_t,t)+\sqrt{\frac{1-\bar\alpha_{t-1}}{\bar\alpha_{t-1}}}\,\epsilon_\theta(x_t,t)$$

$$\frac{x_t}{\sqrt{\bar\alpha_t}}-\frac{x_{t-1}}{\sqrt{\bar\alpha_{t-1}}}=\left(\sqrt{\frac{1-\bar\alpha_t}{\bar\alpha_t}}-\sqrt{\frac{1-\bar\alpha_{t-1}}{\bar\alpha_{t-1}}}\right)\epsilon_\theta(x_t,t)$$

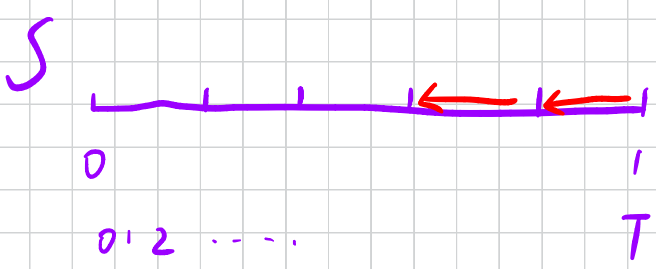
$$a=\sqrt{\bar\alpha}$$
$$\sigma=\sqrt{\frac{1-\bar\alpha}{\bar\alpha}}$$

$$\boxed{\frac{d}{ds}\left(\frac{x(s)}{a(s)}\right)=\frac{d}{ds}\sigma(s)\,\epsilon_\theta(x(s),t(s))}$$

$s\in[0,1]$ 已知 $x(1)\sim N(0,I)$ 求 $x(0)$

$\Updownarrow$

DDPM/DDIM 采样



$x_T=x(1)=\epsilon\sim N(0,I)$

连续的方式求离散

用优化常微分方程的方式去优化采样