Diffusion-based Models

$p(\bar x_{k-1}|\bar x_k) \simeq q(\bar x_{k-1}|\bar x_k)$

$a\bar\varepsilon_1 + b\bar\varepsilon_2 \sim \mathcal{N}(\bar 0, \sqrt{a^2+b^2})$

$\sim p(\bar x_0)$

$\bar x_0 \to \bar x_1 \ldots \to \bar x_{k-1} \to \bar x_k \to \bar x_{k+1} \to \ldots \to \bar x_N = \bar z$

$p(\bar x_1|\bar x_0) = \mathcal{N}(\_)$ $\qquad$ $p(\bar x_k|\bar x_{k-1}) = \mathcal{N}(\_)$

Reparametrization trick

$$p(\bar x_t|\bar x_{t-1}) = \mathcal{N}(\bar x_t \,|\, \sqrt{1-\beta_t}\cdot \bar x_{t-1}, \beta_t I), \qquad \bar\varepsilon_t \sim \mathcal{N}(\bar 0, I)$$

$$\alpha_t = 1-\beta_t$$

$$\bar x_t = \sqrt{1-\beta_t}\cdot\bar x_{t-1} + \sqrt{\beta_t}\cdot\bar\varepsilon_t =$$

$$= \sqrt{(1-\beta_t)(1-\beta_{t-1})}\,\bar x_{t-2} + \overset{\alpha_t(1-\alpha_{t-1})}{\sqrt{\beta_{t-1}(1-\beta_t)}}\,\bar\varepsilon_{t-1} + \overset{1-\alpha_t}{\sqrt{\beta_t}}\,\bar\varepsilon_t =$$

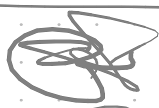
$$= \sqrt{\alpha_t\alpha_{t-1}}\cdot\bar x_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\cdot\bar\varepsilon = \ldots$$

$$\ldots = \sqrt{\underbrace{\alpha_t\alpha_{t-1}\ldots\alpha_1}_{A_t}}\cdot\bar x_0 + \sqrt{1-\underbrace{\alpha_t\alpha_{t-1}\ldots\alpha_1}_{A_t}}\cdot\bar\varepsilon$$

$$\bar x_t = \sqrt{A_t}\cdot\bar x_0 + \sqrt{1-A_t}\cdot\bar\varepsilon$$

forward diffusion

$$\boxed{p(\bar x_t|\bar x_0) = \mathcal{N}(\sqrt{A_t}\cdot\bar x_0, (1-A_t)I)}$$



$$p(\bar x_0|\bar x_T)$$

$$p(\bar x_0, \bar x_1 \ldots \bar x_T) = p(\bar x_0)\cdot\prod_{t=1}^{T} p(\bar x_t|\bar x_{t-1})$$

$$= p(\bar x_T)\prod_{t=1}^{T} p(\bar x_{t-1}|\bar x_t)$$

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

$$= \frac{1}{(\beta_t)^{d/2}\cdot(2\pi)^{d/2}}\cdot e^{-\frac{1}{2\beta_t}(\bar x_t-\sqrt{\alpha_t}\,\bar x_{t-1})^2}\cdot\frac{1}{(1-A_{t-1})^{d/2}(2\pi)^{d/2}}\cdot e^{-\frac{1}{2(1-A_{t-1})}(\bar x_{t-1}-\sqrt{A_{t-1}}\,\bar x_0)^2}$$

$$\propto$$

$- - -$

$$\propto e^{-\frac{1}{2}\left(\frac{(\bar x_t-\sqrt{\alpha_t}\,\bar x_{t-1})^2}{\beta_t}+\frac{(\bar x_{t-1}-\sqrt{A_{t-1}}\,\bar x_0)^2}{1-A_{t-1}}-\cancel{\frac{(\bar x_t-\sqrt{A_t}\,\bar x_0)^2}{1-A_t}}\right)}$$

$$\underbrace{\left(\frac{1}{1-A_{t-1}}+\frac{\alpha_t}{\beta_t}\right)}_{=\frac{1}{\tilde\beta_t}}\bar x_{t-1}^T\bar x_{t-1}-2\underbrace{\left(\frac{\sqrt{\alpha_t}}{\beta_t}\bar x_t+\frac{\sqrt{A_{t-1}}}{1-A_{t-1}}\bar x_0\right)}_{=\frac{1}{\tilde\beta_t}\cdot\tilde\mu(\bar x_0,\bar x_t)}^T\bar x_{t-1}+\text{Const}$$

$$p(\bar x_{t-1}|\bar x_t,\bar x_0)=\mathcal{N}(\bar x_{t-1}|\tilde\mu(\bar x_0,\bar x_t),\tilde\beta_t I)$$

$$\frac{1}{\tilde\beta_t}=\frac{1}{1-A_{t-1}}+\frac{\alpha_t}{1-\alpha_t}=\frac{(1-\cancel{\alpha_t})+\alpha_t(\cancel{1}-A_{t-1})}{(1-A_{t-1})(1-\alpha_t)}=\frac{1-\overbrace{\alpha_t A_{t-1}}^{A_t}}{(1-A_{t-1})\beta_t}$$

$$\boxed{\tilde\beta_t=\frac{1-A_{t-1}}{1-A_t}\cdot\beta_t}$$

$$\tilde\mu(\bar x_0,\bar x_t)=\underline{\quad}\cdot\bar x_0+\underline{\quad}\cdot\bar x_t$$

$$\bar x_{t-1}=\underline{\quad}\cdot\bar x_0+\underline{\quad}\cdot\bar x_t+\underline{\quad}\cdot\bar\varepsilon$$

$$\bar x_t=\underline{\quad}\,\bar x_0+\underline{\quad}\cdot\bar\varepsilon$$

$$\bar x_0=\underline{\quad}\,\bar x_t+\underline{\quad}\cdot\bar\varepsilon_t$$

$$\cancel{q(\bar x_{t-1}|\bar x_t)}\qquad\qquad q(\bar x_1,\dots,\bar x_T|\bar x_0)\approx p(\bar x_1\dots\bar x_T|\bar x_0)$$

$$\cancel{\mathbb{E}_{q(\bar x_0,\bar x_1\dots\bar x_T)}}\ \log p(\bar x_0)=\log p(\bar x_0\dots\bar x_T)-\log p(\bar x_1\dots\bar x_T|\bar x_0)\ \mathbb{E}_{q(\bar x_0\dots\bar x_T)}$$

$$\mathbb{E}_{q(\bar x_0)}[\log p(\bar x_0)]=\mathbb{E}_q\left[\log p(\bar x_{0:T})-\log p(\bar x_{1..T}|\bar x_0)\right]$$

$$\pm\log q(\bar x_{1..T}|\bar x_0)$$

$$\underbrace{\mathbb{E}_{\bar{x}_0}[\log p(\bar{x}_0)]}_{\text{Const}(\bar{x}_0)} = \underbrace{\mathbb{E}_q\left[\log \frac{p(\bar{x}_{0..T})}{q(\bar{x}_{1..T}|\bar{x}_0)}\right]}_{= -L(q)} - \underbrace{\mathbb{E}_q\left[\log \frac{q(\bar{x}_{1..T}|\bar{x}_0)}{p(\bar{x}_{1..T}|\bar{x}_0)}\right]}_{KL(q(\bar{x}_{1..T}|\bar{x}_0)\|p(\bar{x}_{1..T}|\bar{x}_0))}$$

$$\text{Const}(\bar{x}_0) = -L(q) + KL(q(\bar{x}_{1..T}|\bar{x}_0)\|p(\bar{x}_{1..T}|\bar{x}_0))$$

$q \to \min$ $\qquad$ $q \to \min$

$$L(q) = \mathbb{E}_q\left[\log \frac{q(\bar{x}_{1..T}|\bar{x}_0)}{p(\bar{x}_0..\bar{x}_T)}\right]$$

$$\boxed{p(\bar{x}_{t-1}|\bar{x}_t, \bar{x}_0) = N(\bar{x}_{t-1}|\tilde{\mu}, \hat{\beta}_t I)}$$

$$\bar{x}_{t-1} \longrightarrow \bar{x}_t \longrightarrow \bar{x}_{t+1}$$

$$p(\bar{x}_{t-1}|\bar{x}_t) \simeq q_\theta(\bar{x}_{t-1}|\bar{x}_t)$$

$$q_\theta(\bar{x}_0, \dots, \bar{x}_T) = q_\theta(\bar{x}_T) \cdot \prod q_\theta(\bar{x}_{t-1}|\bar{x}_t)$$

$$q_\theta(\bar{x}_1..\bar{x}_T|\bar{x}_0) \approx p(\bar{x}_1..\bar{x}_T|\bar{x}_0)$$

$$\log q_\theta(\bar{x}_0) = \log q_\theta(\bar{x}_{0:T}) - \log q_\theta(\bar{x}_{1..T}|\bar{x}_0)$$

$$\mathbb{E}_{p(\bar{x}_0)}[\log q_\theta(\bar{x}_0)] = \underbrace{\mathbb{E}_p\left[\log \frac{q_\theta(\bar{x}_{0..T})}{p(\bar{x}_1..\bar{x}_T|\bar{x}_0)}\right]}_{\theta \to \max} - \underbrace{\mathbb{E}_p\left[\log \frac{q_\theta(\bar{x}_1..\bar{x}_T|\bar{x}_0)}{p(\bar{x}_1..\bar{x}_T|\bar{x}_0)}\right]}$$

$$\text{Const}(\bar{x}_0) = \underbrace{-L[q_\theta]}_{\theta \to \min} + \underbrace{KL(p\|q_\theta)}_{\theta \to \min}$$

$$L(\bar\theta) = \mathbb{E}_p\left[\log \frac{\prod_{t=1}^{T} p(\bar x_t|\bar x_{t-1})}{q_\theta(\bar x_0, \dots \bar x_T) = q_\theta(\bar x_T)\cdot \prod q_\theta(\bar x_{t-1}|\bar x_t)}\right] =$$

$$= \mathbb{E}_p\left[-\log q_\theta(\bar x_T) + \underbrace{\sum_{t=2}^{T} \log \frac{p(\bar x_t|\bar x_{t-1})}{q_\theta(\bar x_{t-1}|\bar x_t)}} + \log \frac{p(\bar x_1|\bar x_0)}{q_\theta(\bar x_0|\bar x_1)}\right] =$$

$$KL(\dots(\bar x_{t-1}|\dots)\,\|\,q_\theta(\bar x_{t-1}|\dots))$$

$$\boxed{p(\bar x_{t-1}|\bar x_t, \bar x_0)}$$

$$\frac{p(\bar x_t|\bar x_{t-1}, \bar x_0)}{q_\theta(\bar x_{t-1}|\bar x_t)} = \frac{p(\bar x_{t-1}|\bar x_t, \bar x_0)\cdot p(\bar x_t|\bar x_0)}{p(\bar x_{t-1}|\bar x_0)\cdot q_\theta(\bar x_{t-1}|\bar x_t)} =$$

$$= \frac{p(\bar x_{t-1}|\bar x_t, \bar x_0)}{q_\theta(\bar x_{t-1}|\bar x_t)} \cdot \frac{p(\bar x_t|\bar x_0)}{p(\bar x_{t-1}|\bar x_0)}$$

$$= \mathbb{E}_p\left[-\log q_\theta(\bar x_T) + \log \frac{p(\bar x_1|\bar x_0)}{q_\theta(\bar x_0|\bar x_1)} + \sum_{t=2}^{T}\left[\log \frac{p(\bar x_{t-1}|\bar x_t, \bar x_0)}{q_\theta(\bar x_{t-1}|\bar x_t)} + \overline{\log \frac{p(\bar x_t|\bar x_0)}{p(\bar x_{t-1}|\bar x_0)}}\right]\right]$$

$$= \mathbb{E}_p\left[-\log q_\theta(\bar x_T) + \log \frac{\cancel{p(\bar x_1|\bar x_0)}}{q_\theta(\bar x_0|\bar x_1)} + \sum_{t=2}^{T} \log \frac{p(\bar x_{t-1}|\bar x_t, \bar x_0)}{q_\theta(\bar x_{t-1}|\bar x_t)} + \log p(\bar x_T|\bar x_0) \cancel{- \log p(\bar x_1|\bar x_0)}\right]$$

$$L_T = \mathbb{E}[-]$$

$$= \mathbb{E}_{p(\bar x_T|\bar x_0)}\left[\overbrace{\log \frac{p(\bar x_T|\bar x_0)}{q_\theta(\bar x_T)}} + \sum_{t=2}^{T} \overbrace{\log \frac{p(\bar x_{t-1}|\bar x_t, \bar x_0)}{q_\theta(\bar x_{t-1}|\bar x_t)}}^{L_{t-1}} - \overbrace{\log q_\theta(\bar x_0|\bar x_1)}^{L_0}\right]$$

$L = L_T + L_{T-1} + \dots + L_1 + L_0$, где:

$L_0 = -\log q_\theta(\bar x_0|\bar x_1)$

$L_T = KL(p(\bar x_T|\bar x_0)\,\|\,q_\theta(\bar x_T))$

$L_t = KL(p(\bar x_t|\bar x_{t+1}, \bar x_0)\,\|\,q_\theta(\bar x_t|\bar x_{t+1}))$

$$p(\underline{\bar{x}_1 \ldots \bar{x}_T} \mid \bar{x}_0) \approx q_\theta(\bar{x}_1 \ldots \bar{x}_T)$$

$$\shortparallel$$

$$p(\bar{x}_1 \mid \bar{x}_0)\, p(\bar{x}_2 \mid \bar{x}_1) \ldots$$

$$p(\bar{x}_2 \mid \bar{x}_1, \bar{x}_0)$$