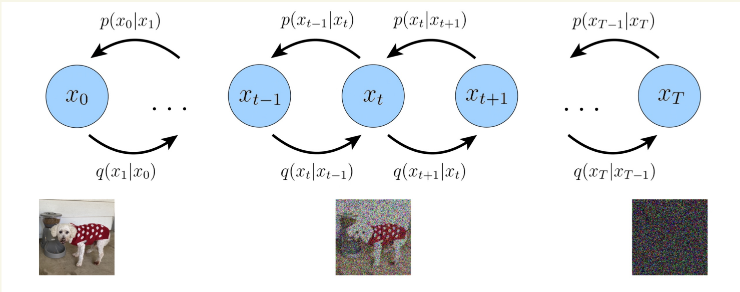
DDPM



Hierachical VAE

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

Volume preserving

- $\mu_t(x_t) = \sqrt{\alpha_t}\, x_t \quad , \qquad \Sigma_t(x_t) = (1-\alpha_t)\, I$

  where, α_t chosen such that variance of latent variable similar scale

Encoder transition

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

Hierachical VAE for variational diffusion model (VDM)

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

  where, $p(x_T) = \mathcal{N}(x_T ; 0, I)$

$$\log \int p(x_{0:T})\, dx_{1:T}$$

$$= \log \int \frac{p(x_{0:T})\, q(x_{1:T}|x_0)}{q(x_{1:T}|x_0)}\, dx_{1:T} = \log \mathbb{E}_{q(x_{1:T}|x_0)}\left[\frac{p(x_{0:T})}{q(x_{1:T}|x_0)}\right]$$

Jensen's equation

$$\geq \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p(x_{0:T})}{q(x_{1:T}|x_0)}\right]$$

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p(x_T)\prod_{i=1}^{T} p_\theta(x_{t-1}|x_t)}{\prod_{i=1}^{T} q(x_t|x_{t-1})}\right]$$

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p(x_T)\, p_\theta(x_0|x_1)\prod_{i=2}^{T} p_\theta(x_{t-1}|x_t)}{q(x_T|x_{T-1})\prod_{i=1}^{T-1} q(x_t|x_{t-1})}\right]$$

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p(x_T)\, p_\theta(x_0|x_1)\prod_{i=1}^{T-1} p_\theta(x_t|x_{t+1})}{q(x_T|x_{T-1})\prod_{i=1}^{T-1} q(x_t|x_{t-1})}\right]$$

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log p_\theta(x_0|x_1)\right]$$

$$+ \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p(x_T)}{q(x_T|x_{T-1})}\right]$$

$$+ \mathbb{E}_{q(x_{1:T}|x_0)}\left[\sum_{i=1}^{T-1} \log \frac{p_\theta(x_t|x_{t+1})}{q(x_t|x_{t-1})}\right]$$

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log p_\theta(x_0|x_1)\right] + \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p(x_T)}{q(x_T|x_{T-1})}\right] + \sum_{i=1}^{T-1} \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p_\theta(x_t|x_{t+1})}{q(x_t|x_{t-1})}\right]$$

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log p_\theta(x_0|x_1)\right] + \mathbb{E}_{\boxed{q(x_{T-1}, x_T|x_0)}}\left[\log \frac{p(x_T)}{q(x_T|x_{T-1})}\right]$$

↑ "$x_{1,\cdots T-2}$" canceled

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

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log p_\theta(x_0|x_1)\right] - \mathbb{E}_{q(x_{T-1}|x_0)}\left[\log \frac{q(x_T|x_{T-1})}{p(x_T)} q(x_T|x_{T-1})\right]$$

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

$$= \underbrace{\mathbb{E}_{q(x_{1:T}|x_0)}\left[\log p_\theta(x_0|x_1)\right]}_{\text{Reconstruct term}} - \underbrace{\mathbb{E}_{q(x_{T-1}|x_0)}\left[D_{KL}(q(x_T|x_{T-1}) \,\|\, p(x_T))\right]}_{\text{Prior matching term}}$$

$$- \underbrace{\sum_{t=1}^{T-1} \mathbb{E}_{q(x_{t-1}, x_{t+1}|x_0)}\left[D_{KL}(q(x_t|x_{t-1}) \,\|\, p_\theta(x_t|x_{t+1}))\right]}_{\text{consistency term}}$$

- **Reconstruct term**
- **Prior matching term** : final latent variable matches gaussian prior
- **consistency term** : make distribution at x_t consistent from both forward and backward process.

∴ consistency term $q(x_{t-1}, x_{t+1}|x_0)$ makes high variance by using ^one sample^ due to two random variables

$\Rightarrow$ key insight : $q(x_t|x_{t-1}) = q(x_t|x_{t-1}, x_0)$

$\uparrow$ with extra conditioning term is superflous due to the Markovian property

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

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

$$\geq \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p(x_{0:T})}{q(x_{1:T}|x_0)}\right]$$

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}{\prod_{t=1}^{T} q(x_t|x_{t-1})}\right]$$

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p(x_T)\, p_\theta(x_0|x_1)\boxed{\prod_{t=2}^{T}} p_\theta(x_{t-1}|x_t)}{q(x_1|x_0)\boxed{\prod_{t=2}^{T}} q(x_t|x_{t-1})}\right]$$

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p(x_T)\, p_\theta(x_0|x_1)\prod_{t=2}^{T} p_\theta(x_{t-1}|x_t)}{q(x_1|x_0)\prod_{t=2}^{T} \boxed{q(x_t|x_{t-1}, x_0)}}\right]$$

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

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

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

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

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

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log p_\theta(x_0|x_1)\right] + \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p(x_T)}{q(x_T|x_0)}\right]$$

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

$$= \mathbb{E}_{q(x_1|x_0)}\left[\log p_\theta(x_0|x_1)\right] + \mathbb{E}_{q(x_T|x_0)}\left[\log \frac{p(x_T)}{q(x_T|x_0)}\right]$$

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

$$\therefore = \underbrace{\mathbb{E}_{q(x_1|x_0)}\left[\log p_\theta(x_0|x_1)\right]}_{\text{Reconstruction term}} - \underbrace{D_{KL}\left(q(x_T|x_0)\,\|\,p(x_T)\right)}_{\text{prior matching}}$$

$$- \underbrace{\sum_{t=2}^{T} \mathbb{E}_{q(x_t|x_0)}\left[D_{KL}\left(q(x_{t-1}|x_t, x_0)\,\|\,p_\theta(x_{t-1}|x_0)\right)\right]}_{\text{denoising matching term}}$$

- **Reconstruction term** : monte-carlo estimation
- **Prior matching term** : no trainable parameter
- **Denoising matching term** : " $q(x_{t-1}|x_t, x_0)$ " : it defines how to denoise a noisy image x_t with access to what the final completely denoised image x_0 should be.

- $D_{KL}(q(x_{t-1} | x_t, x_0) \| p_\theta(x_{t-1} | x_t))$
  - it is difficult to minimize for arbitrary posteriors
  - using Gaussian transition assumption to make tractable

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

- $q(x_t | x_{t-1}, x_0) = q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t}\, x_{t-1}, (1-\alpha_t) I)$

$$x_t = \sqrt{\alpha}\, x_{t-1} + \sqrt{1-\alpha_t}\, \epsilon \quad \text{with} \quad \epsilon \sim \mathcal{N}(0, I)$$

$$x_{t-1} = \sqrt{\alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_{t-1}}\, \epsilon \quad \text{with} \quad \epsilon \sim \mathcal{N}(0, I)$$

$\epsilon_{t-2}, \epsilon_{t-1} \sim \mathcal{N}(0, I)$

$$\begin{aligned}
\boldsymbol{x}_t &= \sqrt{\alpha_t}\boldsymbol{x}_{t-1} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}^*_{t-1} && (61)\\
&= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}\boldsymbol{x}_{t-2} + \sqrt{1-\alpha_{t-1}}\boldsymbol{\epsilon}^*_{t-2}\right) + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}^*_{t-1} && (62)\\
&= \sqrt{\alpha_t\alpha_{t-1}}\boldsymbol{x}_{t-2} + \sqrt{\alpha_t - \alpha_t\alpha_{t-1}}\boldsymbol{\epsilon}^*_{t-2} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}^*_{t-1} && (63)\\
&= \sqrt{\alpha_t\alpha_{t-1}}\boldsymbol{x}_{t-2} + \sqrt{\sqrt{\alpha_t - \alpha_t\alpha_{t-1}}^2 + \sqrt{1-\alpha_t}^2}\boldsymbol{\epsilon}_{t-2} && (64)\\
&= \sqrt{\alpha_t\alpha_{t-1}}\boldsymbol{x}_{t-2} + \sqrt{\alpha_t - \alpha_t\alpha_{t-1} + 1 - \alpha_t}\boldsymbol{\epsilon}_{t-2} && (65)\\
&= \sqrt{\alpha_t\alpha_{t-1}}\boldsymbol{x}_{t-2} + \sqrt{1 - \alpha_t\alpha_{t-1}}\boldsymbol{\epsilon}_{t-2} && (66)\\
&= \ldots && (67)\\
&= \sqrt{\prod_{i=1}^{t}\alpha_i}\boldsymbol{x}_0 + \sqrt{1 - \prod_{i=1}^{t}\alpha_i}\boldsymbol{\epsilon}_0 && (68)\\
&= \sqrt{\bar\alpha_t}\boldsymbol{x}_0 + \sqrt{1-\bar\alpha_t}\boldsymbol{\epsilon}_0 && (69)\\
&\sim \mathcal{N}(\boldsymbol{x}_t; \sqrt{\bar\alpha_t}\boldsymbol{x}_0, (1-\bar\alpha_t)\mathbf{I}) && (70)
\end{aligned}$$

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

gaussian

gaussian

denoising 역시 gaussian.

gaussian

$$\begin{aligned}
q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) &= \frac{q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{x}_0)q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0)}{q(\boldsymbol{x}_t|\boldsymbol{x}_0)} && (71)\\
&= \frac{\mathcal{N}(\boldsymbol{x}_t; \sqrt{\alpha_t}\boldsymbol{x}_{t-1}, (1-\alpha_t)\mathbf{I})\mathcal{N}(\boldsymbol{x}_{t-1}; \sqrt{\bar\alpha_{t-1}}\boldsymbol{x}_0, (1-\bar\alpha_{t-1})\mathbf{I})}{\mathcal{N}(\boldsymbol{x}_t; \sqrt{\bar\alpha_t}\boldsymbol{x}_0, (1-\bar\alpha_t)\mathbf{I})} && (72)\\
&\propto \exp\left\{-\left[\frac{(\boldsymbol{x}_t - \sqrt{\alpha_t}\boldsymbol{x}_{t-1})^2}{2(1-\alpha_t)} + \frac{(\boldsymbol{x}_{t-1} - \sqrt{\bar\alpha_{t-1}}\boldsymbol{x}_0)^2}{2(1-\bar\alpha_{t-1})} - \frac{(\boldsymbol{x}_t - \sqrt{\bar\alpha_t}\boldsymbol{x}_0)^2}{2(1-\bar\alpha_t)}\right]\right\} && (73)\\
&= \exp\left\{-\frac{1}{2}\left[\frac{(\boldsymbol{x}_t - \sqrt{\alpha_t}\boldsymbol{x}_{t-1})^2}{1-\alpha_t} + \frac{(\boldsymbol{x}_{t-1} - \sqrt{\bar\alpha_{t-1}}\boldsymbol{x}_0)^2}{1-\bar\alpha_{t-1}} - \frac{(\boldsymbol{x}_t - \sqrt{\bar\alpha_t}\boldsymbol{x}_0)^2}{1-\bar\alpha_t}\right]\right\} && (74)\\
&= \exp\left\{-\frac{1}{2}\left[\frac{(-2\sqrt{\alpha_t}\boldsymbol{x}_t\boldsymbol{x}_{t-1} + \alpha_t\boldsymbol{x}_{t-1}^2)}{1-\alpha_t} + \frac{(\boldsymbol{x}_{t-1}^2 - 2\sqrt{\bar\alpha_{t-1}}\boldsymbol{x}_{t-1}\boldsymbol{x}_0)}{1-\bar\alpha_{t-1}} + C(\boldsymbol{x}_t, \boldsymbol{x}_0)\right]\right\} && (75)\\
&\propto \exp\left\{-\frac{1}{2}\left[-\frac{2\sqrt{\alpha_t}\boldsymbol{x}_t\boldsymbol{x}_{t-1}}{1-\alpha_t} + \frac{\alpha_t\boldsymbol{x}_{t-1}^2}{1-\alpha_t} + \frac{\boldsymbol{x}_{t-1}^2}{1-\bar\alpha_{t-1}} - \frac{2\sqrt{\bar\alpha_{t-1}}\boldsymbol{x}_{t-1}\boldsymbol{x}_0}{1-\bar\alpha_{t-1}}\right]\right\} && (76)\\
&= \exp\left\{-\frac{1}{2}\left[(\frac{\alpha_t}{1-\alpha_t} + \frac{1}{1-\bar\alpha_{t-1}})\boldsymbol{x}_{t-1}^2 - 2\left(\frac{\sqrt{\alpha_t}\boldsymbol{x}_t}{1-\alpha_t} + \frac{\sqrt{\bar\alpha_{t-1}}\boldsymbol{x}_0}{1-\bar\alpha_{t-1}}\right)\boldsymbol{x}_{t-1}\right]\right\} && (77)\\
&= \exp\left\{-\frac{1}{2}\left[\frac{\alpha_t(1-\bar\alpha_{t-1}) + 1 - \alpha_t}{(1-\alpha_t)(1-\bar\alpha_{t-1})}\boldsymbol{x}_{t-1}^2 - 2\left(\frac{\sqrt{\alpha_t}\boldsymbol{x}_t}{1-\alpha_t} + \frac{\sqrt{\bar\alpha_{t-1}}\boldsymbol{x}_0}{1-\bar\alpha_{t-1}}\right)\boldsymbol{x}_{t-1}\right]\right\} && (78)\\
&= \exp\left\{-\frac{1}{2}\left[\frac{\alpha_t - \bar\alpha_t + 1 - \alpha_t}{(1-\alpha_t)(1-\bar\alpha_{t-1})}\boldsymbol{x}_{t-1}^2 - 2\left(\frac{\sqrt{\alpha_t}\boldsymbol{x}_t}{1-\alpha_t} + \frac{\sqrt{\bar\alpha_{t-1}}\boldsymbol{x}_0}{1-\bar\alpha_{t-1}}\right)\boldsymbol{x}_{t-1}\right]\right\} && (79)\\
&= \exp\left\{-\frac{1}{2}\left[\frac{1-\bar\alpha_t}{(1-\alpha_t)(1-\bar\alpha_{t-1})}\boldsymbol{x}_{t-1}^2 - 2\left(\frac{\sqrt{\alpha_t}\boldsymbol{x}_t}{1-\alpha_t} + \frac{\sqrt{\bar\alpha_{t-1}}\boldsymbol{x}_0}{1-\bar\alpha_{t-1}}\right)\boldsymbol{x}_{t-1}\right]\right\} && (80)\\
&= \exp\left\{-\frac{1}{2}\left(\frac{1-\bar\alpha_t}{(1-\alpha_t)(1-\bar\alpha_{t-1})}\right)\left[\boldsymbol{x}_{t-1}^2 - 2\frac{\left(\frac{\sqrt{\alpha_t}\boldsymbol{x}_t}{1-\alpha_t} + \frac{\sqrt{\bar\alpha_{t-1}}\boldsymbol{x}_0}{1-\bar\alpha_{t-1}}\right)}{\frac{1-\bar\alpha_t}{(1-\alpha_t)(1-\bar\alpha_{t-1})}}\boldsymbol{x}_{t-1}\right]\right\} && (81)\\
&= \exp\left\{-\frac{1}{2}\left(\frac{1-\bar\alpha_t}{(1-\alpha_t)(1-\bar\alpha_{t-1})}\right)\left[\boldsymbol{x}_{t-1}^2 - 2\frac{\left(\frac{\sqrt{\alpha_t}\boldsymbol{x}_t}{1-\alpha_t} + \frac{\sqrt{\bar\alpha_{t-1}}\boldsymbol{x}_0}{1-\bar\alpha_{t-1}}\right)(1-\alpha_t)(1-\bar\alpha_{t-1})}{1-\bar\alpha_t}\boldsymbol{x}_{t-1}\right]\right\} && (82)\\
&= \exp\left\{-\frac{1}{2}\left(\frac{1}{\frac{(1-\alpha_t)(1-\bar\alpha_{t-1})}{1-\bar\alpha_t}}\right)\left[\boldsymbol{x}_{t-1}^2 - 2\frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})\boldsymbol{x}_t + \sqrt{\bar\alpha_{t-1}}(1-\alpha_t)\boldsymbol{x}_0}{1-\bar\alpha_t}\boldsymbol{x}_{t-1}\right]\right\} && (83)\\
&\propto \mathcal{N}(\boldsymbol{x}_{t-1}; \underbrace{\frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})\boldsymbol{x}_t + \sqrt{\bar\alpha_{t-1}}(1-\alpha_t)\boldsymbol{x}_0}{1-\bar\alpha_t}}_{\mu_q(\boldsymbol{x}_t, \boldsymbol{x}_0)}, \underbrace{\frac{(1-\alpha_t)(1-\bar\alpha_{t-1})}{1-\bar\alpha_t}\mathbf{I}}_{\boldsymbol{\Sigma}_q(t)}) && (84)
\end{aligned}$$

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

$$\Sigma_q(t) = \frac{(1-\alpha_t)(1-\bar\alpha_{t-1})}{1-\bar\alpha_t} I$$

where $\bar\alpha_t = \prod_{i=1}^{t} \alpha_i$

$$\underset{\theta}{\arg\min}\ D_{KL}\left(q(x_{t-1}|x_t, x_0) \,\|\, p_\theta(x_{t-1}|x_t) \right)$$

$$= \underset{\theta}{\arg\min}\ D_{KL}\left(\mathcal{N}(x_{t-1}; \mu_q, \Sigma_q(t)) \,\|\, \mathcal{N}(x_{t-1}; \mu_\theta, \Sigma_q(t)) \right)$$

$\mu_\theta(x_t, t)$

$$= \underset{\theta}{\arg\min}\ \frac{1}{2\sigma_q^2(t)} \left[\| \mu_\theta - \mu_q \|_2^2 \right]$$

where, $\mu_q(x_t, x_0) = \dfrac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})x_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)x_0}{1-\bar{\alpha}_t}$

$$\mu_\theta(x_t, t) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})x_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\boxed{\hat{x}_\theta(x_t, t)}}{1-\bar{\alpha}_t}$$

$$= \underset{\theta}{\arg\min}\ \frac{1}{2\sigma_q^2(t)} \frac{\bar{\alpha}_{t-1}(1-\alpha_t)^2}{(1-\bar{\alpha}_t)^2} \left[\| \hat{x}_\theta(x_t, t) - x_0 \|_2^2 \right]$$

$$= \frac{1}{2} \left(SNR(t-1) - SNR(t) \right) \left[\| \hat{x}_\theta(x_t, t) - x_0 \|_2^2 \right]$$

where $SNR(t-1) = \dfrac{\bar{\alpha}_t}{1-\bar{\alpha}_t}$

- Equation 69

$$x_t = \sqrt{\bar{\alpha}}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon_0$$

$$\hookrightarrow \boxed{x_0 = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\, \epsilon_0}{\sqrt{\bar{\alpha}_t}}}$$

$$\mu_q(x_t, x_0) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})x_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\, x_0}{1-\bar{\alpha}_t}$$

$$\therefore \mu_q(x_t, x_0) = \frac{1}{\sqrt{\alpha_t}} x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}} \epsilon_0$$

$$\updownarrow \quad \mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} x_t - \frac{1}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}} \hat{\epsilon}_\theta(x_t, t)$$

$$\arg\min_\theta D_{KL}\left(q(x_{t-1} | x_t, x_0) \,\|\, p_\theta(x_{t-1} | x_t)\right)$$

$$= \arg\min_\theta \frac{1}{2\sigma_q^2(t)} \frac{(1-\alpha_t)^2}{(1-\bar{\alpha}_t)\alpha_t} \left[\| \epsilon_0 - \hat{\epsilon}_\theta(x_t, t) \|_2^2 \right]$$

↳ variance

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

↳ KLD에서의 variance 를 의미

- Tweedie's Formular

  : true mean of an exponential family dist.

  : empirical mean + correction term involving score of estimate.
  (likelihood)

$$\mathbb{E}[\mu_z | z] = z + \Sigma_z \nabla_z \log p(z)$$

where $z \sim \mathcal{N}(z; \mu_z, \Sigma_z)$

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

by tweedie equation.

$$\mathbb{E}[\mu_{x_t} | x_t] = x_t + (1 - \bar{\alpha}_t) \nabla_x \log p(x_t)$$

$$\therefore \sqrt{\bar{\alpha}_t}\, x_0 = x_t + (1 - \bar{\alpha}_t) \nabla \log p(x_t)$$

$$\therefore x_0 = \frac{x_t + (1 - \bar{\alpha}_t) \nabla \log p(x_t)}{\sqrt{\bar{\alpha}_t}}$$

$$q(x_{t-1} | x_t, x_0) = \mathcal{N}(\mu_q(x_t, x_0), \Sigma_q(t))$$

$$\mu_q(x_t, x_0) = \frac{1}{\sqrt{\alpha_t}} x_t + \frac{1 - \alpha_t}{\sqrt{\alpha_t}} \nabla \log p(x)$$

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} x_t + \frac{1 - \alpha_t}{\sqrt{\alpha_t}} s_\theta(x)$$

$$\begin{aligned}
\boldsymbol{\mu}_q(\boldsymbol{x}_t, \boldsymbol{x}_0) &= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\boldsymbol{x}_0}{1-\bar{\alpha}_t} && (134)\\
&= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\frac{\boldsymbol{x}_t + (1-\bar{\alpha}_t)\nabla \log p(\boldsymbol{x}_t)}{\sqrt{\bar{\alpha}_t}}}{1-\bar{\alpha}_t} && (135)\\
&= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t + (1-\alpha_t)\frac{\boldsymbol{x}_t + (1-\bar{\alpha}_t)\nabla \log p(\boldsymbol{x}_t)}{\sqrt{\alpha_t}}}{1-\bar{\alpha}_t} && (136)\\
&= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t}{1-\bar{\alpha}_t} + \frac{(1-\alpha_t)\boldsymbol{x}_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}} + \frac{(1-\alpha_t)(1-\bar{\alpha}_t)\nabla \log p(\boldsymbol{x}_t)}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}} && (137)\\
&= \left(\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} + \frac{1-\alpha_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}}\right)\boldsymbol{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\nabla \log p(\boldsymbol{x}_t) && (138)\\
&= \left(\frac{\alpha_t(1-\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}} + \frac{1-\alpha_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}}\right)\boldsymbol{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\nabla \log p(\boldsymbol{x}_t) && (139)\\
&= \frac{\alpha_t - \bar{\alpha}_t + 1 - \alpha_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}}\boldsymbol{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\nabla \log p(\boldsymbol{x}_t) && (140)\\
&= \frac{1-\bar{\alpha}_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}}\boldsymbol{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\nabla \log p(\boldsymbol{x}_t) && (141)\\
&= \frac{1}{\sqrt{\alpha_t}}\boldsymbol{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\nabla \log p(\boldsymbol{x}_t) && (142)
\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{x}_0 = \frac{\boldsymbol{x}_t + (1-\bar{\alpha}_t)\nabla \log p(\boldsymbol{x}_t)}{\sqrt{\bar{\alpha}_t}} &= \frac{\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_0}{\sqrt{\bar{\alpha}_t}} \\
\therefore (1-\bar{\alpha}_t)\nabla \log p(\boldsymbol{x}_t) &= -\sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_0 \\
\nabla \log p(\boldsymbol{x}_t) &= -\frac{1}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_0
\end{aligned}
$$

score-based generative models

- Langevin dynamics

$$x_{i+1} = x_i + c \nabla \log p(x_i) + \sqrt{2c}\ \varepsilon \qquad i = 0, \cdots, K$$

① all points not on the low dimensional manifold would have 0 probability

↳ log. prob not defined

② estimated score function will not be accurate in low density region

③ Langevin dynamics sampling may not mix even if it's performed using g.t scores

⇒ multiple levels of gaussian noises

↳ adding multiple levels of gaussian noise with increasing variance will result in intermediate distributions that respect the ground truth mixing coefficients.