Q1. Given $q(x_{1-T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1})$

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

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

$x_1, x_2, \dots, x_T$ form a Markov chain when conditioned on x_0.]

$$
\begin{aligned}
q(x_{1-T}|x_0) &= \prod_{t=1}^{T} q(x_t|x_{t-1}) \\
&= \prod_{t=1}^{T} q(x_t|x_{t-1}, x_0) \\
&= q(x_1|x_0, x_0) \prod_{t=2}^{T} q(x_t|x_{t-1}, x_0) \\
&= q(x_1|x_0) \prod_{t=2}^{T} \frac{q(\cancel{x_0})\, q(x_t|x_0)\, q(x_{t-1}|x_t, x_0)}{q(\cancel{x_0})\, q(x_{t-1}|x_0)} \\
&= q(x_1|x_0) \prod_{t=2}^{T} \frac{q(x_t|x_0)}{q(x_{t-1}|x_0)} \prod_{t=2}^{T} q(x_{t-1}|x_t, x_0) \\
&= q(\cancel{x_1|x_0}) \frac{q(x_T|x_0)}{q(\cancel{x_1|x_0})} \prod_{t=2}^{T} q(x_{t-1}|x_t, x_0) \\
&= q(x_T|x_0) \prod_{t=2}^{T} q(x_{t-1}|x_t, x_0)
\end{aligned}
$$

Q2. Prove Eqs. (4), (6), (8) in DDPM paper

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

$a_t = 1-\beta_t$

$\bar{a}_t = \prod_{T}^{t} a_T$

we only need to prove $x_t = \sqrt{\bar{a}_t} x_0 + \sqrt{1-\bar{a}_t}\, z$,

where $t \in [1, T]$ and $z \sim \mathcal{N}(0, I)$

① when $t=1$, it satisfied

$$x_1 = \sqrt{a_1} x_0 + \sqrt{1-a_1} z_1 = \sqrt{\bar{a}_1} x_0 + \sqrt{1-\bar{a}_1} z_1$$

② Let we assume that x_{t-1} can also satisfy the above assumption, we can get

$$x_t = \sqrt{a_t} x_{t-1} + \sqrt{1-a_t} z_{t-1}$$

$$= \sqrt{a_t}(\sqrt{\bar{a}_{t-1}} x_0 + \sqrt{1-\bar{a}_{t-1}} z_{t-1}) + \sqrt{1-a_t} z_t$$

$$= \sqrt{a_t \bar{a}_{t-1}} x_0 + \sqrt{a_t - a_t \bar{a}_{t-1}} z_{t-1} + \sqrt{1-a_t} z_t$$

$$= \underbrace{\sqrt{\bar{a}_t} x_0}_{\text{value}} + \underbrace{\sqrt{a_t - \bar{a}_t} z_{t-1}}_{\sim \mathcal{N}(0, a_t - \bar{a}_t)} + \underbrace{\sqrt{1-a_t} z_t}_{\sim \mathcal{N}(0, 1-a_t)}$$

Based on the property of Gaussian function, we can know

$$x_t \sim \mathcal{N}(\sqrt{\bar{a}_t} x_0, (1-a_t+a_t-\bar{a}_t)I) \sim \mathcal{N}(\sqrt{\bar{a}_t} x_0, (1-\bar{a}_t)I)$$

$$\Rightarrow x_t = \sqrt{\bar{a}_t} x_0 + \sqrt{1-\bar{a}_t}\, \hat{z}_t$$

$\therefore$ this satisfy the initial assumption

(b) $q(x_{t-1}|x_1, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I)$

As we can know, $\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{a}_{t-1}}\beta_t}{1-\bar{a}_t} x_0 + \frac{\sqrt{\bar{a}_t}(1-\bar{a}_{t-1})}{1-\bar{a}_t} x_1$

and $\tilde{\beta}_t := \frac{1-\bar{a}_{t-1}}{1-\bar{a}_t}\beta_t$

$$q(x_{t-1}|x_1, x_0) = \frac{q(x_t|x_{t-1}, x_0)\, q(x_{t-1}|x_0)}{q(x_t|x_0)}$$

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

$$= \frac{\mathcal{N}(x_t; \sqrt{a_t}x_{t-1}, (1-a_t)I)\, \mathcal{N}(x_{t-1}; \sqrt{a_{t-1}}x_0, (1-a_{t-1})I)}{\mathcal{N}(x_t; \sqrt{a_t}x_0, (1-\bar{a}_t)I)}$$

$$\propto \exp\left(-\left(\frac{(x_t-\sqrt{a_t}x_{t-1})^2}{2(1-a_t)} + \frac{(x_{t-1}-\sqrt{\bar{a}_{t-1}}x_0)^2}{2(1-\bar{a}_{t-1})} - \frac{(x_t-\sqrt{\bar{a}_t}x_0)^2}{2(1-\bar{a}_t)}\right)\right)$$

$$= \exp\left(-\frac{1}{2}\left(\frac{(x_t-\sqrt{a_t}x_{t-1})^2}{1-a_t} + \frac{(x_{t-1}-\sqrt{\bar{a}_{t-1}}x_0)^2}{1-\bar{a}_{t-1}} - \frac{(x_t-\sqrt{\bar{a}_t}x_0)^2}{1-\bar{a}_t}\right)\right)$$

$$= \exp\left(-\frac{1}{2}\left(\frac{-2\sqrt{a_t}x_t x_{t-1} + a_t x_{t-1}^2}{1-a_t} + \frac{x_{t-1}^2 - 2\sqrt{\bar{a}_{t-1}}x_{t-1}x_0}{1-\bar{a}_{t-1}} + c(x_t, x_0)\right)\right)$$

$$= \exp\left(-\frac{1}{2}\left(\left(\frac{a_t}{1-a_t} + \frac{1}{1-\bar{a}_{t-1}}\right)x_{t-1}^2 - 2\left(\frac{\sqrt{a_t}x_t}{1-a_t} + \frac{\sqrt{\bar{a}_{t-1}}x_0}{1-\bar{a}_{t-1}}\right)x_{t-1} + c(x_t, x_0)\right)\right)$$

$$= \exp\left(-\frac{1}{2}\left(\frac{a_t - \bar{a}_t + 1 - a_t}{(1-a_t)(1-\bar{a}_{t-1})}x_{t-1}^2 - 2\left(\frac{\sqrt{a_t}x_t}{1-a_t} + \frac{\sqrt{\bar{a}_{t-1}}x_0}{1-\bar{a}_{t-1}}\right)x_{t-1}\right) - \frac{1}{2}c(x_t, x_0)\right)$$

$$= \exp\left(-\frac{1}{2}\left(\frac{1-\bar{a}_t}{(1-a_t)(1-\bar{a}_{t-1})}\right)\left(x_{t-1}^2 - 2\frac{\frac{\sqrt{a_t}x_t}{1-a_t} + \frac{\sqrt{\bar{a}_{t-1}}x_0}{1-\bar{a}_{t-1}}}{\frac{1-\bar{a}_t}{(1-a_t)(1-\bar{a}_{t-1})}}x_{t-1}\right) - \frac{1}{2}c(x_t, x_0)\right)$$

$$= \exp\left(-\frac{1}{2}\left(\frac{1-\bar{a}_t}{(1-a_t)(1-\bar{a}_{t-1})}\right)\left(x_{t-1}^2 - 2\frac{\left(\frac{\sqrt{a_t}x_t}{1-a_t}+\frac{\sqrt{\bar{a}_{t-1}}x_0}{1-\bar{a}_{t-1}}\right)(1-a_t)(1-\bar{a}_{t-1})}{1-\bar{a}_t}x_{t-1}\right) - \frac{1}{2}C(x_t,x_0)\right)$$

$$= \exp\left(-\frac{1}{2}\left(\frac{1}{\frac{1-\bar{a}_t}{(1-a_t)(1-\bar{a}_{t-1})}}\right)\left(x_{t-1}^2 - 2\frac{\sqrt{a_t}(1-\bar{a}_t)x_t+\sqrt{\bar{a}_{t-1}}(1-a_t)x_0}{1-\bar{a}_t}x_{t-1}\right) - \frac{1}{2}C(x_t,x_0)\right)$$

$$= \exp\left(-\frac{\left(x_{t-1}-\frac{\sqrt{a_t}(1-\bar{a}_t)x_t+\sqrt{\bar{a}_{t-1}}(1-a_t)x_0}{1-\bar{a}_t}\right)^2}{2\left(\frac{1-\bar{a}_t}{(1-a_t)(1-\bar{a}_{t-1})}\right)}\right)$$

$$\propto \mathcal{N}\left(x_{t-1};\ \underbrace{\frac{\sqrt{a_t}(1-\bar{a}_{t-1})x_t+\sqrt{\bar{a}_{t-1}}(1-a_t)x_0}{1-\bar{a}_t}}_{\tilde{\mu}_t(x_t,x_0)},\ \underbrace{\frac{(1-a_t)(1-\bar{a}_{t-1})}{1-\bar{a}_t}}_{\tilde{\beta}_t}I\right)$$

(8) $L_{t-1} = E_q\left[\frac{1}{2\sigma_t^2}\|\tilde{\mu}_t(x_t,x_0)-\tilde{\mu}_\theta(x_t,t)\|^2\right]+C$

Based on (6) $q(x_{t-1}|x_1,x_0) = N(x_{t-1};\tilde{\mu}_t(x_t,x_0),\tilde{\beta}_t I)$

and $p_\theta(x_{t-1}|x_t) = N(x_{t-1};\tilde{\mu}_\theta(x_t,t),\sigma_t^2 I)$

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

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

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

② if p, q are Gaussian distribution, KL divergence is

$$KL(p,q) = KL(N(\mu_1,\sigma_1^2)\|N(\mu_2,\sigma_2^2))$$

$$= \int_x \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\log\frac{\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}}{\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}}dx$$

$$= \int_x \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\left[\log\frac{\sigma_2}{\sigma_1}-\frac{(x-\mu_1)^2}{2\sigma_1^2}+\frac{(x-\mu_2)^2}{2\sigma_2^2}\right]dx$$

$$= \log\frac{\sigma_2}{\sigma_1}\int_x \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}dx$$

$$+ (x-\mu_1)^2\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}dx + (x-\mu_2)^2\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}dx$$

$$= \log\frac{\sigma_2}{\sigma_1}-\frac{1}{2}+\frac{\sigma_1^2+(\mu_1-\mu_2)^2}{2\sigma_2^2}$$

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

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

$$= E_{q(x_t|x_0)}\left[\underbrace{KL(q(x_0|x_0,x_t)\|p_\theta(x_0|x_1))}\right]$$

$$KL\,N(\mu,\sigma^2)\|N(\mu_1,\sigma_1^2) = \frac{1}{2}\left(\log\sigma^2+\log\sigma_1^2-1+\frac{\sigma^2}{\sigma_1^2}+\frac{(\mu-\mu_1)^2}{\sigma_1^2}\right)$$

$$\Longleftrightarrow KL(q(x_{t-1}|x_0,x_t)\|p_\theta(x_{t-1}|x_t)) = \frac{1}{2}\left(C+\frac{\|\tilde{\mu}(x_t,x_0)-\mu_\theta(x_t,t)\|^2}{\sigma_t^2}\right)$$