Given $q(X_{1-T}|X_0) = \prod_{t=1}^{T} q(X_t|X_{t-1})$

Show $q(X_{1-T}|X_0) = q(X_T|X_0)\prod_{t=T}^{2} q(X_{t-1}|X_t, X_0)$

since $X_1, X_2, \ldots\ldots, X_T$ from a Markov chain when conditioned on X_0.:

$$\Rightarrow q(X_{1-T}|X_0) = \prod_{t=1}^{T} q(X_t|X_{t-1}, X_0)$$

$$\Rightarrow q(X_{1-T}|X_0) = \frac{q(X_T, X_{T-1}, X_0)}{q(X_{T-1}, X_0)} \times \frac{q(X_{T-1}, X_{T-2}, X_0)}{(X_{T-2}, X_0)} \times \cdots\cdots\cdots$$

$$\times \frac{q(X_3, X_2, X_0)}{q(X_2, X_0)} \times \frac{q(X_2, X_1, X_0)}{q(\cancel{X_1, X_0})} \times \frac{\cancel{q(X_1, X_0)}}{q(X_0)}$$

$$\Rightarrow q(X_{1-T}|X_0) = \frac{q(X_T, X_{T-1}, X_0)}{q(X_0)} \times \frac{q(X_{T-1}, X_{T-2}, X_0)}{q(X_{T-1}, X_0)} \times \cdots\cdots\cdots \times \frac{q(X_2, X_1, X_0)}{q(X_2, X_0)}$$

$$= \frac{q(X_T, X_{T-1}, X_0)}{q(X_0)} \times \left[q(X_{T-2}|X_{T-1}, X_0) \times \cdots\cdots \times q(X_1|X_2, X_0) \right]$$

$$= \frac{q(X_T, X_0)}{q(X_0)} \times \frac{q(X_T, X_{T-1}, X_0)}{q(X_T, X_0)} \times \prod_{t=T-1}^{2} q(X_{t-1}|X_t, X_0)$$

$$= q(X_T|X_0) \times \left[q(X_{T-1}|X_T, X_0) \times \prod_{t=T-1}^{2} q(X_{t-1}|X_t, X_0) \right]$$

$$= q(X_T|X_0) \prod_{t=T}^{2} q(X_{t-1}|X_t, X_0) \quad ※.$$

eq(4). Define $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$

Then $q(X_t|X_{t-1}) = N(\sqrt{1-\beta_t}\, X_{t-1}, \beta_t I)$

$$X_t = \sqrt{1-\beta_t}\, X_{t-1} + \sqrt{\beta_t}\, \varepsilon, \quad \varepsilon \sim N(0, I)$$

$$= \sqrt{\alpha_t}\, X_{t-1} + \sqrt{1-\alpha_t}\, \varepsilon$$

$$= \sqrt{\alpha_t \alpha_{t-1}}\, X_{t-2} + \sqrt{1-\alpha_t \alpha_{t-1}}\, \varepsilon$$

$$= \cdots\cdots$$

$$= \sqrt{\bar{\alpha}_t}\, X_0 + \sqrt{1-\bar{\alpha}_t}\, \varepsilon$$

$\therefore q(X_t|X_0) = N(\sqrt{\bar{\alpha}_t}\, X_0, (1-\bar{\alpha}_t) I)$

eq(6). We know that $q(X_t|X_{t-1}, X_0) = q(X_t|X_{t-1})$ since Markov Chain.

$$q(X_t|X_{t-1}, X_0) = q(X_t|X_{t-1}) = N(\sqrt{1-\beta_t}\,X_{t-1}, \beta_t I)$$

And we also know that $q(X_t|X_0) = N(\sqrt{\bar{\alpha}_t}\,X_0, (1-\bar{\alpha}_t)I)$ from eq(4).

Thus, we get :

$$q(X_{t-1}|X_t, X_0) = q(X_t|X_{t-1}, X_0)\cdot\frac{q(X_{t-1}|X_0)}{q(X_t|X_0)}$$

$$\propto \exp\left[-\frac{1}{2}\left(\frac{(X_t-\sqrt{\alpha_t}X_{t-1})^2}{\beta_t}+\frac{(X_{t-1}-\sqrt{\bar{\alpha}_{t-1}}X_0)^2}{1-\bar{\alpha}_{t-1}}-\frac{(X_t-\sqrt{\bar{\alpha}_t}X_0)^2}{1-\bar{\alpha}_t}\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\frac{X_t^2-2\sqrt{\alpha_t}X_tX_{t-1}+\alpha_t X_{t-1}^2}{\beta_t}+\frac{X_{t-1}^2-2\sqrt{\bar{\alpha}_{t-1}}X_0X_{t-1}+\bar{\alpha}_{t-1}X_0^2}{1-\bar{\alpha}_{t-1}}-\frac{(X_t-\sqrt{\bar{\alpha}_t}X_0)^2}{1-\bar{\alpha}_t}\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\left(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right)X_{t-1}^2-\left(\frac{2\sqrt{\alpha_t}}{\beta_t}X_t+\frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}X_0\right)X_{t-1}+C(X_t,X_0)\right)\right]$$

$C(X_t, X_0)$ is an irrelevant part of X_{t-1}, so we can omit it.

Thus, $q(X_{t-1}|X_t, X_0)$ is Gaussian Distribution.

And $$\tilde{\beta}_t = \frac{1}{\left(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right)} = \frac{1}{\left(\frac{\alpha_t-\bar{\alpha}_t+\beta_t}{\beta_t(1-\bar{\alpha}_{t-1})}\right)} = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t$$

$$\tilde{\mu}_t(X_t, X_0) = \frac{\left(\frac{\sqrt{\alpha_t}}{\beta_t}X_t+\frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}X_0\right)}{\left(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right)} = \left(\frac{\sqrt{\alpha_t}}{\beta_t}X_t+\frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}X_0\right)\cdot\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t$$

$$= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}X_t+\frac{\sqrt{\bar{\alpha}_{t-1}}\,\beta_t}{1-\bar{\alpha}_t}X_0$$

$$\therefore\ q(X_{t-1}|X_t, X_0) = N(X_{t-1}; \tilde{\mu}_t(X_t, X_0), \tilde{\beta}_t I)$$

where $\tilde{\mu}_t(X_t, X_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\,\beta_t}{1-\bar{\alpha}_t}X_0+\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}X_t$ and $\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$

$$eq(8.).\quad L = \mathbb{E}_q\left[-\log \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right]$$

$$= \mathbb{E}_q\left[-\log p(x_T) - \sum_{t\geq 1} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_t|x_{t-1})}\right]$$

$$= \mathbb{E}_q\left[-\log p(x_T) - \sum_{t\geq 1} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_t|x_{t-1})} - \log \frac{p_\theta(x_0|x_1)}{q(x_1|x_0)}\right.$$

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

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

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

Then $D_{KL}(q(x_{t-1}|x_t, x_0)\,\|\,p_\theta(x_{t-1}|x_t))$

$$= D_{KL}(\mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \sigma_t^2 I)\,\|\,\mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 I))$$

$$\left(\because D_{KL}(p_1\|p_2) = \frac{1}{2}\left(tr(\Sigma_2^{-1}\Sigma_1) + (\mu_2-\mu_1)^T \Sigma_2^{-1}(\mu_2-\mu_1) - n + \log \frac{\det(\Sigma_2)}{\det(\Sigma_1)}\right)\right).$$

$$= \frac{1}{2}\left(n + \frac{1}{\sigma_t^2}\|\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t)\|^2 - n + \log 1\right)$$

$$= \frac{1}{2\sigma_t^2}\|\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t)\|^2$$

$$\therefore L_{t-1} = \mathbb{E}_q\left[\frac{1}{2\sigma_t^2}\|\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t)\|^2\right] + C.$$