Problem 1

Given

$$q(\mathbf{x}_{1-T}|\mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})$$

Show that

$$q(\mathbf{x}_{1-T}|\mathbf{x}_0) = q(\mathbf{x}_T|\mathbf{x}_0)\prod_{t=T}^{2} q(\mathbf{x}_{t-1}|\mathbf{x}_T, \mathbf{x}_0)$$

Solution

$$\begin{aligned}
& q(\mathbf{x}_T|\mathbf{x}_0)\prod_{t=T}^{2} q(\mathbf{x}_{t-1}|\mathbf{x}_T, \mathbf{x}_0) \\
&= q(\mathbf{x}_T|\mathbf{x}_0)\left[q(\mathbf{x}_{T-1}|\mathbf{x}_T, \mathbf{x}_0)q(\mathbf{x}_{T-2}|\mathbf{x}_{T-1}, \mathbf{x}_0)\cdots q(\mathbf{x}_2|\mathbf{x}_3, \mathbf{x}_0)q(\mathbf{x}_1|\mathbf{x}_2, \mathbf{x}_0)\right] \\
&= q(\mathbf{x}_T|\mathbf{x}_0)\left[q(\mathbf{x}_T|\mathbf{x}_{T-1}, \mathbf{x}_0)\frac{q(\mathbf{x}_{T-1}|\mathbf{x}_0)}{q(\mathbf{x}_T|\mathbf{x}_0)}q(\mathbf{x}_{T-1}|\mathbf{x}_{T-2}, \mathbf{x}_0)\frac{q(\mathbf{x}_{T-2}|\mathbf{x}_0)}{q(\mathbf{x}_{T-1}|\mathbf{x}_0)}\cdots\right. \\
&\quad \left. q(\mathbf{x}_3|\mathbf{x}_2, \mathbf{x}_0)\frac{q(\mathbf{x}_2|\mathbf{x}_0)}{q(\mathbf{x}_3|\mathbf{x}_0)}q(\mathbf{x}_2|\mathbf{x}_1, \mathbf{x}_0)\frac{q(\mathbf{x}_1|\mathbf{x}_0)}{q(\mathbf{x}_2|\mathbf{x}_0)}\right] \\
&= q(\mathbf{x}_T|\mathbf{x}_0)\left[\prod_{t=2}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})\frac{q(\mathbf{x}_1|\mathbf{x}_0)}{q(\mathbf{x}_T|\mathbf{x}_0)}\right] \\
&= q(\mathbf{x}_1|\mathbf{x}_0)\prod_{t=2}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}) \\
&= \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}) \\
&= q(\mathbf{x}_{1-T}|\mathbf{x}_0)
\end{aligned} \tag{1}$$

Problem 2

Prove Equation (4)

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$$

Solution

Suppose $\epsilon_i, \forall\ i \in \{t-1, t-2, \cdots\} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Then,

$$\begin{aligned}
\mathbf{x}_t &= \sqrt{\alpha_t}\mathbf{x}_{t-1} + \sqrt{1-\alpha_t}\epsilon_{t-1} \\
&= \sqrt{\alpha_t}(\sqrt{\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_{t-1}}\epsilon_{t-2}) + \sqrt{1-\alpha_t}\epsilon_{t-1} \\
&= \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{\sqrt{\alpha_t - \alpha_t\alpha_{t-1}}^2 + \sqrt{1-\alpha_t}^2}\bar{\epsilon}_{t-2} \\
&= \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\bar{\epsilon}_{t-2} \\
&= \cdots \\
&= \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon
\end{aligned} \tag{2}$$

In 3rd line, $\bar{\epsilon}_{t-2}$ merges two Gaussian matrices ϵ_{t-1} and ϵ_{t-2}.
Thus,

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$$

Problem 3

Prove Equation (6)

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1}; \tilde{\mu}_t(\mathbf{x}_t, x_0), \tilde{\beta}_t\mathbf{I})$$

Solution

From Equation (4), we know that

$$\begin{aligned} q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0) &= q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}) \\ q(\mathbf{x}_{t-1}|\mathbf{x}_0) &= \mathcal{N}(\mathbf{x}_{t-1}; \sqrt{\bar{\alpha}_{t-1}}\mathbf{x}_0, (1-\bar{\alpha}_{t-1})\mathbf{I}) \\ q(\mathbf{x}_t|\mathbf{x}_0) &= \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}) \end{aligned} \tag{3}$$

Thus,

$$\begin{aligned} & q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \\ = \; & q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)\frac{q(\mathbf{x}_{t-1}|\mathbf{x}_0)}{q(\mathbf{x}_t|\mathbf{x}_0)} \\ \propto \; & \exp\left(-\frac{1}{2}\left(\frac{(\mathbf{x}_t - \sqrt{\alpha_t}\mathbf{x}_{t-1})^2}{\beta_t} + \frac{(\mathbf{x}_{t-1} - \sqrt{\bar{\alpha}_{t-1}}\mathbf{x}_0)^2}{1-\bar{\alpha}_{t-1}} + \frac{(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0)^2}{1-\bar{\alpha}_t}\right)\right) \\ = \; & \exp\left(-\frac{1}{2}\left(\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right)\mathbf{x}_{t-1}^2 - \left(\frac{2\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0\right)\mathbf{x}_{t-1} + C(\mathbf{x}_t, \mathbf{x}_0)\right)\right) \end{aligned} \tag{4}$$

$C(\mathbf{x}_t, \mathbf{x}_0)$ does not depend on $\mathbf{x}_{t-1}$, can be omitted. Then,

$$\begin{aligned} \tilde{\beta}_t &= 1/\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right) = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t \\ \tilde{\mu}(\mathbf{x}_t, \mathbf{x}_0) &= \left(\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0\right) / \left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right) \\ &= \left(\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0\right)\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t \\ &= \frac{\sqrt{\alpha}_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0 \\ &\text{where } q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1}; \tilde{\mu}_t(\mathbf{x}_t, x_0), \tilde{\beta}_t\mathbf{I}) \end{aligned} \tag{5}$$

Problem 4

Prove Equation (8)

$$L_{t-1} = \mathbb{E}_q \left[\frac{1}{2\sigma_t^2} ||\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0) - \mu_\theta(\mathbf{x}_t, t)||^2 \right] + C$$

Solution

From Equation (16), we have

$$\begin{aligned} L = & \underbrace{D_{KL}(q(\mathbf{x}_T|\mathbf{x}_0)||p_\theta(\mathbf{x}_T))}_{L_T} - \underbrace{\mathbb{E}_{q(\mathbf{x}_1|\mathbf{x}_0)} \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0} \\ & + \sum_{t=2}^{T} \underbrace{\mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)} \left[D_{KL}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)) \right]}_{L_{t-1}} \end{aligned} \tag{6}$$

Fix $\Sigma_\theta(\mathbf{x}_t, t) = \sigma_t^2 \mathbf{I}$. And KL divergence of two Gaussian distributions p_1 and p_2 is

$$KL(p_1||p_2) = \frac{1}{2}[tr(\Sigma_2^{-1}\Sigma_1) + (\mu_2 - \mu_1)^\intercal \Sigma_2^{-1}(\mu_2 - \mu_1) - n + \log \frac{\det(\Sigma_2)}{\det(\Sigma_1)}]$$

Thus,

$$\begin{aligned} & D_{KL}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)) \\ = & D_{KL}(\mathcal{N}(\mathbf{x}_{t-1}; \tilde{\mu}(\mathbf{x}_t, \mathbf{x}_0), \sigma_t^2 \mathbf{I})||\mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \sigma_t^2 \mathbf{I})) \\ = & \frac{1}{2} \left(n + \frac{1}{\sigma_t^2} ||\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0) - \mu_\theta(\mathbf{x}_t, t)||^2 - n + \log 1 \right) \\ = & \frac{1}{2\sigma_t^2} ||\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0) - \mu_\theta(\mathbf{x}_t, t)||^2 \\ \rightarrow & L_{t-1} = \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)} \left[\frac{1}{2\sigma_t^2} ||\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0) - \mu_\theta(\mathbf{x}_t, t)||^2 \right] \end{aligned} \tag{7}$$