STAT 37601 Final Project

Albert Chu

May 2025

1 Theory I

1. It is clear that we can recursively substitute the forward kernels:

$$\begin{aligned}
x_t &= \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}\epsilon_t \\
&= \sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \sqrt{\alpha_t(1-\alpha_{t-1})}\epsilon_{t-1} + \sqrt{1-\alpha_t}\epsilon_t \\
&= ... \\
&= \sqrt{\bar{\alpha}_t}x_0 + \sum_{s=1}^{t}\sqrt{1-\alpha_s}\sqrt{\prod_{j=s+1}^{t}\alpha_j\epsilon_s}
\end{aligned}$$

   The first term is deterministic and the sum is a linear combination of independent gaussians, hence it is Gaussian with mean 0 and covariance of the form
$$\sum_{s=1}^{t}(1-\alpha_s)(\prod_{j=s+1}^{t}\alpha_j)I_d = (1-\bar{\alpha}_t)I_d$$
   Therefore, $x_t|x_0 \sim \mathcal{N}(\sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I_d)$

2. Given (1), we can expand it by splitting the logarithm and extracting the sum from the expectation.
$$\log\frac{\prod_{t=1}^{T}p_{t-1|t}(x_{t-1}|x_t;\theta)p_T(x_T)}{\prod_{t=1}^{T}q_{t|t-1}(x_t|x_{t-1})} = \sum_{t=1}^{T}\log\frac{p_{t-1|t}(x_{t-1}|x_t;\theta)}{\prod_{t=1}^{T}q_{t|t-1}(x_t|x_{t-1})} + \log p_T(x_T)$$
   The term inside the first integral depends only on (x_{t-1}, x_t), so we can integrate out all other variables and write the expectation wrt to the two-step marginal as
$$q_{t-1,t|0}(x_t, x_{t-1}|x_0) = q_{t-1|0}(x_{t-1}|x_0)q_{t|t-1}(x_t|x_{t-1})$$
   We can put these all together to get exactly
$$L(\theta, X_0) = \sum_{t=1}^{T}\int_{x_t,x_{t-1}}\log\frac{p_{t-1|t}(x_{t-1}|x_t;\theta)}{q_{t|t-1}(x_t|x_{t-1})}q_{t-1,t|0}(x_t,x_{t-1}|x_0)dx_{t-1}dx_t + \int_{x_T}q_T(x_T|x_0)\log p(x_T)dx$$

3. To get to the first equation, since we're setting up a loss for training, we simply take the negative ELBO and drop irrelevant constants like the $q_{t|t-1}$ and $p_T(x_T)$ to get

$$\sum_{t=1}^{T}\int_{x_{t-1},x_t} -\log p(x_{t-1}|x_t;\theta)q_{t-1,t|0}(x_{t-1},x_t|X_0)dx_{t-1}dx_t$$

We can then insert the Gaussian form of $p_{t-1|t}$ to get to the second line,

$$-\log p_{t-1|t}(x_{t-1}|x_t;\theta) = \frac{|x_{t-1}-\mu(x_t,t;\theta)|^2}{2(1-\alpha_t)} + \frac{d}{2}\log 2\pi(1-\alpha_t)$$

Note that we can drop the final term since it does not depend on θ,

$$= \sum_{t=1}^{T}\int \frac{|x_{t-1}-\mu(x_t,t;\theta)|^2}{2(1-\alpha_t)} q_{t-1,t|0}(x_{t-1},x_t|X_0)dx_{t-1}dx_t + C$$

The final line comes from recognizing the integral as an expectation to get a compact form.

4. Since problem 1 already gives $q_{t|0}$ and the entire forward chain is Gaussian, the pair (X_{t-1},X_t) is jointly Gaussian with

$$E[X_{t-1}|x_0] = \sqrt{\bar{\alpha}_{t-1}}x_0$$
$$E[X_t|x_0] = \sqrt{\bar{\alpha}_t}x_0$$
$$Cov(X_{t-1},X_t|x_0) = \sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})I_d$$

Therefore, plugging those into the conditioning formula yields

$$X_t \sim \mathcal{N}(\sqrt{\bar{\alpha}_t}x_0,(1-\bar{\alpha}_t)I_d)$$
$$X_{t-1}|X_t,x_0 \sim \mathcal{N}(\tilde{\mu}_{t-1},\tilde{\sigma}_t^2 I_d)$$
$$\tilde{\sigma}_t^2 = \frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}$$
$$\tilde{\mu}_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}X_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}x_0$$

2 Theory II

1. Re-express $q(x_t|x_{t-1})$ as a Gaussian,

$$q(x_{t-1}|x_t) = \mathcal{N}\left(\frac{x_t}{\sqrt{\alpha_t}},\frac{\beta_t}{\sqrt{\alpha_t}}I_d\right)$$

Since we're multiplying two Gaussians with $\mathcal{N}(m_1, S_1)$ and $\mathcal{N}(m_2, S_2)$, we can compute Σ,

$$
\begin{aligned}
S_1^{-1} + S_2^{-1} &= \frac{\alpha_t}{\beta_t} I_d + \frac{1}{1-\bar{\alpha}_{t-1}} I_d \\
&= \frac{1-\bar{\alpha}_t}{\beta_t(1-\bar{\alpha}_{t-1})} I_d
\end{aligned}
$$

Therefore, $\Sigma = \frac{\beta_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} I_d$ We can also see,

$$
\begin{aligned}
\mu &= \rho_t(S_1^{-1} m_1 + S_2^{-1} m_2) \\
&= \rho_t(\frac{\alpha_t}{\beta_t} \frac{x_t}{\sqrt{\alpha_t}} + \frac{1}{1-\bar{\alpha}_{t-1}} \sqrt{\bar{\alpha}_{t-1}} x_0) \\
&= \rho_t(\frac{\sqrt{\alpha}_t}{1-\alpha_t} x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}} x_0) \\
&= \frac{(1-\bar{\alpha}_{t-1})\sqrt{\alpha_t} x_t + (1-\alpha_t)\sqrt{\bar{\alpha}_{t-1} x_0}}{1-\bar{\alpha}_t}
\end{aligned}
$$

2. The first equality comes from marginalizing out since the forward process is Markov with fixed x_0,

$$
q_{t-1,t|0}(x_{t-1}, x_t|x_0) = q_{t-1|t,0}(x_{t-1}|x_t, x_0) q_{t|0}(x_t|x_0)
$$

The second equality comes from evaluating the inner integral over x_{t-1},

$$
E_{x_{t-1} \sim q_{t-1|t,0}}[|x_{t-1} - \mu(x_t, t; \theta)|^2] = |\tilde{\mu}_t(x_t, x_0) - \mu(x_t, t; \theta)|^2 + p_t d
$$

The third equality comes from recognizing the remaining integral as an expectation.

3. We substitute X_0 in terms of X_T and ϵ_t,

$$
X_0 = \frac{X_t - \sqrt{1-\bar{\alpha}_t}\epsilon_t}{\sqrt{\bar{\alpha}_t}}
$$

We can then plug in to the earlier equation for the closed-form posterior mean to get

$$
\sqrt{\bar{\alpha}_{t-1}} X_0 = \frac{1}{\sqrt{\alpha}_t}(X_t - \sqrt{1-\bar{\alpha}_t}\epsilon_t)
$$

$$
\begin{aligned}
\tilde{\mu}(X_t, X_0) &= \frac{1}{1-\bar{\alpha}_t}\left(\frac{(1-\alpha_t)}{\sqrt{\alpha_t}}(X_t - \sqrt{1-\bar{\alpha}_t}\epsilon_t) + \frac{\sqrt{\alpha_t}(\alpha_t - \bar{\alpha}_t)}{\alpha_t} X_t\right) \\
&= \frac{1}{\sqrt{\alpha_t}}(X_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t)
\end{aligned}
$$

4. We substitute the two means,

$$\tilde{\mu} - \mu = frac1 - \alpha_t \sqrt{\alpha_t(1 - \bar{\alpha_t})}[\epsilon_t - e_t(X_t, t; \theta)]$$

We then square the norm and divide by $2(1 - \alpha_t)$,

$$\rightarrow \frac{(1 - \alpha_t)^2}{2\alpha_t(1 - \bar{\alpha_t})} |\epsilon_t - e_t(X_t, t; \theta)|^2$$

And summing over $t = 1, ..., T$ yields the given loss equation.