Assumptions for all problems

1. $q(z_n \mid z_{n-1}) = \mathcal{N}(z_n; z_{n-1}, \sigma_n^2 I)$, or $z_n = z_{n-1} + \sigma_n \varepsilon$, $\varepsilon \sim \mathcal{N}(0, I)$.
2. $z_0, \ldots, z_{n-1}$ are conditionally independent of $z_{n+1}, \ldots, z_N$ given z_n.
3. $V_n = \sigma_1^2 + \ldots + \sigma_n^2$, so, by the previous two assumptions, $q(z_n \mid z_0) = \mathcal{N}(z_n; z_0, V_n I)$, or $z_n = z_0 + \sqrt{V_n}\varepsilon$, $\varepsilon \sim \mathcal{N}(0, I)$.
4. Unless stated otherwise, $\mathbb{E}$ is taken over $p(z_0)$, $q(\cdot \mid z_0)$, and $n \sim \text{Unif}(\{1, \ldots, N\})$.

Problem 1: Closed form for ELBO continued

In Lecture 8 we showed that the diffusion model ELBO can be written as

$$\mathcal{L}_{\text{ELBO}}(\theta) = \mathbb{E}_{z_0 \sim p(z_0), z_{1,\ldots,N} \sim q(\cdot|z_0)} \left[\log p(z_N) + \sum_{n=1}^{N} \log \frac{p_\theta(z_{n-1} \mid z_n)}{q(z_n \mid z_{n-1})} \right]$$

Show the following Monte-Carlo estimator of $\mathcal{L}_{\text{ELBO}}(\theta)$:

$$\mathcal{L}_{\text{ELBO}}(\theta) = \underbrace{-\,\mathbb{E}\left[\frac{1}{2\gamma_{n-1}^2} \|\mu_n(z_n, z_0) - \mu_n(z_n; \theta)\|^2 \right]}_{L_n} + \underbrace{\mathbb{E}\left[\log p_\theta(z_0 \mid z_1)\right]}_{L_0} + \text{const.}$$

where $\mu_n(z_n, z_0)$ is given by a linear combination of z_n and z_0, $\mu_n(z_n; \theta)$ is the predicted mean of the Gaussian defining $p_\theta(z_{n-1} \mid x_n)$, and the γ_{n-1}^2 are some constants. Use the following steps:

1. Using the fact that $q(z_n \mid z_{n-1}) = \frac{q(z_{n-1}|z_n, z_0) q(z_n|z_0)}{q(z_{n-1}|z_0)}$ for $2 \leq n \leq N$, show that $q(z_{n-1} \mid z_n, z_0) = \mathcal{N}(z_{n-1}; \mu_n(z_n, z_0), \gamma_{n-1}^2 I)$. Provide the expressions for $\mu(z_n, z_0)$ and γ_{n-1}^2.
2. Compute $\log \frac{p_\theta(z_{n-1}|z_n)}{q(z_{n-1}|z_n, z_0)}$, assuming that $p_\theta(z_{n-1} \mid z_n) = \mathcal{N}(z_{n-1}; \mu_n(z_n; \theta), \gamma_{n-1}^2 I)$;

How is this version of the ELBO related to the one derived in Problem 2 of Week 8?

Problem 2: Equivalent losses

Using the fact that $\mu_n(z_n, z_0)$ is a linear combination of z_n and z_0, show that L_n can be equivalently rewritten as

1. **(denoiser)** $\mathbb{E}\left[\omega(n)\|z_0 - f_n(z_n; \theta)\|^2\right]$ using that $\mu_n(z_n, z_0)$ is a linear combination of z_n and z_0;
2. **(noise predictor)** $\mathbb{E}\left[\omega(n)\|\varepsilon - f_n(z_n; \theta)\|^2\right]$ using $z_n = z_0 + \sqrt{V_n}\varepsilon$; $\varepsilon \sim \mathcal{N}(0, I)$;
3. **(score predictor)** $\mathbb{E}\left[\omega(n)\|s_n - f_n(z_n; \theta)\|^2\right]$ using, given fixed z_0, $s_n = \frac{1}{V_n}(z_0 - z_n)$;

where in each case $f_n(z_n; \theta)$ is a neural network. In each case, express $\mu_n(z_n; \theta)$ via $f_n(z_n; \theta)$ and compute the weight $\omega(n)$ that gives the equivalent loss.
Note: for the continuous versions of the provided losses, the weight ω can be given a particular meaning. To learn more about it, see the following papers: [1], [2].

Problem 3: Discrete sampling using score / noise / denoiser predictions

For each of the equivalent definitions of L_n from the previous problem and corresponding trained models f_θ, write the probabilistic program for sampling z_0.