

# CS282 Special Participation A: Qwen on HW13

## Executive Summary

I use Qwen to solve HW13 written part in this special participation A.

Qwen's accuracy really impressed me. All questions, including requiring proofs (e.g., induction in diffusion models), derivations (e.g., DPO gradients), or conceptual explanations (e.g., intractability of partition functions), were answered correctly and comprehensively in a single attempt, with no need for clarification or correction. The responses included rigorous mathematical detail, proper notation, and logical step-by-step reasoning.

There is no factual inaccuracies, fabricated citations, or conceptual errors were observed. For instance:

- The derivation of the optimal policy in DPO correctly recovered the Gibbs form and explained the cancellation of the partition function in Bradley-Terry and Plackett-Luce models.
- The diffusion process proofs aligned exactly with standard results from the literature (e.g., Ho et al., 2020).
- Interpretations (e.g., gradient weighting in DPO reflecting model uncertainty) were insightful and technically sound.

Despite minimal input context (only image placeholders uploaded), the model inferred the likely content (standard theoretical ML problems) and provided complete, self-contained solutions, demonstrating strong contextual awareness and domain knowledge. Responses were well-organized, labeled by sub-question, and included boxed final answers where appropriate, enhancing readability and correctness verification.

## Question 1

### DDPM/DDIM Fun: From a Gaussian

Consider generative diffusion models under the simplifying assumption that the target distribution we wish to sample from is a Gaussian with zero mean and small variance  $\sigma^2$ , where  $\sigma^2 \ll 1$ . The forward process is defined to start from time  $t = T$  to  $t = 1$ .

Suppose the forward diffusion process starts at  $X_0 \sim N(0, \sigma^2)$  and at each small interval of length  $\Delta t$  adds independent Gaussian noise with mean zero and variance  $\Delta t \ll \sigma^2$ .

- (e) What is the distribution of  $X_T$ ?

(HINT: Remember you can chose of  $\Delta t = \frac{1}{T}$  for some large  $T$  and  $X_T$  is the final result after doing  $T$  steps of additive independent Gaussian noise to the initial realization of  $X_0$ .)

- (f) Looking at the forward diffusion process around time  $t$ , what are the marginal distributions of  $X_{t-\Delta t}$  and  $X_t$ ?

(g) In the previous part, the conditional distribution of  $(X_{t-\Delta t}|X_t = x_t) \sim N\left(\frac{\Delta t c_{t-\Delta t}}{\sigma^2 + \Delta t} x_t, \frac{\sigma^2(1 - \Delta t/c_{t-\Delta t})}{\sigma^2 + \Delta t}\right)$ . Simplify and approximate the variance  $\frac{\sigma^2(1 - \Delta t/c_{t-\Delta t})}{\sigma^2 + \Delta t}$  of this conditional distribution when  $\Delta t \ll \sigma^2$ .

(h) This part asks you to see what happens if you try to do reverse diffusion safely without applying stochastic noise of each reverse diffusion step. Suppose we start with a sample  $\tilde{x}_1 \sim N(0, 1)$  and iteratively apply only the conditional mean mapping backward in time  $t = 0$  to get a sample of  $\tilde{x}_0$ . What is the resulting distribution of  $\tilde{x}_0$ ? Remember that  $\Delta t \ll \sigma^2 \ll 1$ ! If you want to make any approximations or to integrate the result, Show your work.

(HINT 1: Consider  $T = \frac{1}{\Delta t}$  and construct the appropriate product to see how the final  $\tilde{x}_0$  is distributed. Do you notice any telecooperation?)

(HINT 2: You can use the information given in the next part to check your work here.)

(i) In the previous part, you should have found that  $\tilde{x}_0$  has far too little variance if you don't add independent stochastic noise along the way in DDPM-style reverse diffusion.

(j) Let  $T = \frac{1}{\Delta t}$ . Suppose now that we add independent  $N(0, \Delta t)$  noise at each of  $T$  reverse diffusion steps. It turns out that

$$\text{Var}(\tilde{x}_0) = \left(\frac{\sigma^2}{\sigma^2 + 1}\right)^2 + \Delta t \sum_{k=0}^{T-1} \left(\frac{\sigma^2}{\sigma^2 + k\Delta t}\right)^2. \quad (1)$$

Take the limit  $\Delta t \rightarrow 0$  and write  $\text{Var}(\tilde{x}_0)$  involving an integral assuming  $0 < \Delta t \ll \sigma^2 \ll 1$ . And then evaluate it approximately.

(HINT 1: For evaluating the integral,  $\int \frac{dx}{x^2 + 1} dx = C - \frac{1}{2} \ln(x)$ )

(HINT 2: You know that the answer should come out to be close to our desired variance of  $\sigma^2$  and you can use that to check your work.)

- (k) Notice that the conditional mean step in DDPM is approximately a  $\Delta t/t$  step toward the ideal predictor  $\tilde{x}_0^{\text{target}} = \frac{\sigma^2}{\sigma^2 + 1} \tilde{x}_1$ . Namely that

$$\tilde{x}_{1-\Delta t}^{\text{target}} = \tilde{x}_1 + \frac{\Delta t}{t} (\tilde{x}_0^{\text{target}} - \tilde{x}_1) + \text{noise} = \left(1 - \frac{\Delta t}{\sigma^2 + 1}\right) \tilde{x}_1 + \text{noise}.$$

This can be incrementally viewed as a deterministic DDPM step of  $-\frac{\Delta t}{\sigma^2 + 1} \tilde{x}_1$  together with a noise step.

DDIM takes no noise step, but a smaller time-varying deterministic step:

$$x_{t-\Delta t}^{\text{DDIM}} = \tilde{x}_1 + v(t, \Delta t)(\text{deterministic DDPM step}) \quad \text{with } v(t, \Delta t) = \frac{\sqrt{\Delta t}}{\sqrt{t} - \Delta t + \sqrt{\Delta t}}$$

Describe briefly, if you were simply given a dataset of samples  $x_1, x_2, \dots, x_n$  drawn from the same distribution as our desired  $X_0$  distribution, how you would train a neural network to estimate the function  $g(x_t, t)$  that in this case, turns out to be  $\frac{\sigma^2}{\sigma^2 + t} x_1$  in analytic form. Specifically, what are the inputs to the neural net, how do you generate a batch of them, and how would you compute a loss on the outputs of the neural net for training an optimization algorithm like AdamW to set the parameters of the neural net?

- (l) Assuming you had exact analytic form access to  $g(x_t, t) = \frac{\sigma^2}{\sigma^2 + t} x_1$  in the previous part, approximate the DDIM step for  $\Delta t \ll t$ , with an assumption about  $\sigma^2 \gg 1$ .

(HINT: Use the fact that  $2a \ll t$  to approximate  $v(t, \Delta t)$  as an appropriate constant.)

- (m) Compute  $\tilde{x}_0$  from the random sample  $\tilde{x}_1$  (drawn from  $N(0, 1)$ ) by applying all  $T = \frac{1}{\Delta t}$  DDIM steps from the previous part and expressing the total estimation as a product. Then turn it into a sum via logarithms and take the  $\Delta t \rightarrow 0$  limit, using integrals as needed to evaluate the resulting distribution for  $\tilde{x}_0$ . Show your work.

(HINT 1: You know that the answer is supposed to be approximately  $N(0, \sigma^2)$  under our assumption that  $\sigma^2 \ll 1$ . Use this to check your work.)

(HINT 2: Remember  $\ln(1 - a) \approx -a$  if  $0 < a \ll 1$ .)

(HINT 3: The log of a limit is the limit of logs.)

## Qwen3-Max

Let's solve this problem step-by-step, as it explores the foundations of DDPM/DDIM under a Gaussian assumption.

### (a) Distribution of $X_1$

We are told that the forward process starts at  $X_0 \sim \mathcal{N}(0, \sigma^2)$ , and at each small interval  $\Delta t$ , we add independent Gaussian noise with mean 0 and variance  $\Delta t$ . So after one step:

$$X_1 = X_0 + \epsilon_1, \quad \epsilon_1 \sim \mathcal{N}(0, \Delta t)$$

Since  $X_0$  and  $\epsilon_1$  are independent Gaussians, their sum is also Gaussian:

$$X_1 \sim \mathcal{N}(0, \sigma^2 + \Delta t)$$

**Answer (a):**  $X_1 \sim \mathcal{N}(0, \sigma^2 + \Delta t)$

---

### (b) Marginal distributions of $X_{t-\Delta t}$ and $X_t$

The forward process runs from  $t = 0$  to  $t = 1$ , with steps of size  $\Delta t$ . Let's define  $t = k\Delta t$  for integer  $k$ , so after  $k$  steps, we have added noise  $k$  times, each with variance  $\Delta t$ .

At time  $t$ , we've added  $t/\Delta t$  steps of noise. Since we start at  $X_0 \sim \mathcal{N}(0, \sigma^2)$ , and add independent  $\mathcal{N}(0, \Delta t)$  noise at each step, the total added noise variance is  $(t/\Delta t) \cdot \Delta t = t$ .

Thus, the marginal distribution at time  $t$  is:

$$X_t \sim \mathcal{N}(0, \sigma^2 + t)$$

Similarly, at time  $t - \Delta t$ :

Similarly, at time  $t - \Delta t$ :

$$X_{t-\Delta t} \sim \mathcal{N}(0, \sigma^2 + t - \Delta t)$$

 **Answer (b):**

- $X_t \sim \mathcal{N}(0, \sigma^2 + t)$
  - $X_{t-\Delta t} \sim \mathcal{N}(0, \sigma^2 + t - \Delta t)$
- 

### (c) Approximate variance of conditional distribution

Given:

$$(X_{t-\Delta t} | X_t = x_t) \sim \mathcal{N}\left(\frac{\sigma^2 + t - \Delta t}{\sigma^2 + t} x_t, \frac{(\sigma^2 + t - \Delta t)\Delta t}{\sigma^2 + t}\right)$$

We are to approximate the variance when  $\Delta t \ll \sigma^2 \ll 1$ .

So look at the variance term:

$$\text{Var} = \frac{(\sigma^2 + t - \Delta t)\Delta t}{\sigma^2 + t}$$

Since  $\Delta t \ll \sigma^2$ , we can approximate  $\sigma^2 + t - \Delta t \approx \sigma^2 + t$ , so:

$$\text{Var} \approx \frac{(\sigma^2 + t)\Delta t}{\sigma^2 + t} = \Delta t$$

But wait — this approximation is too crude? Let's be more precise.

Actually, let's write:

$$\frac{(\sigma^2 + t - \Delta t)\Delta t}{\sigma^2 + t} = \Delta t \cdot \left(1 - \frac{\Delta t}{\sigma^2 + t}\right)$$

Since  $\Delta t \ll \sigma^2 + t$ , the second factor is approximately 1, so:

$$\text{Var} \approx \Delta t$$

 **Answer (c):** The variance is approximately  $\boxed{\Delta t}$

---

#### (d) Naive reverse denoising without noise

We start with  $X_1 \sim \mathcal{N}(0, 1)$ , and iteratively apply the **conditional mean** mapping backward to get  $\hat{X}_0$ .

Recall from part (c), the conditional mean is:

$$\mathbb{E}[X_{t-\Delta t} | X_t = x_t] = \frac{\sigma^2 + t - \Delta t}{\sigma^2 + t} x_t$$

So if we start at  $t = 1$ , and go backward in steps of  $\Delta t$ , then after  $T = 1/\Delta t$  steps, we get:

$$\hat{X}_0 = \left( \prod_{k=0}^{T-1} \frac{\sigma^2 + (1 - (k+1)\Delta t)}{\sigma^2 + (1 - k\Delta t)} \right) X_1$$

Wait — let's index properly.

Let  $t_k = 1 - k\Delta t$ , so  $t_0 = 1, t_T = 0$ .

Then:

$$\hat{X}_{t_k} = \frac{\sigma^2 + t_k - \Delta t}{\sigma^2 + t_k} \hat{X}_{t_{k-1}} = \frac{\sigma^2 + t_{k+1}}{\sigma^2 + t_k} \hat{X}_{t_{k-1}}$$

Actually, better to write recursively:

At step  $k$ , going from  $t = 1 - k\Delta t$  to  $t = 1 - (k+1)\Delta t$ :

$$\hat{X}_{t-\Delta t} = \frac{\sigma^2 + t - \Delta t}{\sigma^2 + t} \hat{X}_t$$

So starting from  $X_1$ , after  $T = 1/\Delta t$  steps:

$$\hat{X}_0 = X_1 \cdot \prod_{k=0}^{T-1} \frac{\sigma^2 + (1 - (k+1)\Delta t)}{\sigma^2 + (1 - k\Delta t)}$$

This is a telescoping product!

Write out terms:

$$\text{Numerator: } \prod_{k=0}^{T-1} (\sigma^2 + 1 - (k+1)\Delta t) = \prod_{j=1}^T (\sigma^2 + 1 - j\Delta t)$$

$$\text{Denominator: } \prod_{k=0}^{T-1} (\sigma^2 + 1 - k\Delta t) = \prod_{j=0}^{T-1} (\sigma^2 + 1 - j\Delta t)$$

So overall:

$$\hat{X}_0 = X_1 \cdot \frac{\prod_{j=1}^T (\sigma^2 + 1 - j\Delta t)}{\prod_{j=0}^{T-1} (\sigma^2 + 1 - j\Delta t)} = X_1 \cdot \frac{\sigma^2 + 1 - T\Delta t}{\sigma^2 + 1 - 0 \cdot \Delta t} = X_1 \cdot \frac{\sigma^2 + 1 - 1}{\sigma^2 + 1} = X_1 \cdot \frac{\sigma^2}{\sigma^2 + 1}$$

Because  $T\Delta t = 1$ .

So:

$$\hat{X}_0 = \frac{\sigma^2}{\sigma^2 + 1} X_1$$

Since  $X_1 \sim \mathcal{N}(0, 1)$ , then:

$$\hat{X}_0 \sim \mathcal{N}\left(0, \left(\frac{\sigma^2}{\sigma^2 + 1}\right)^2\right)$$

**Answer (d):**  $\hat{X}_0 \sim \mathcal{N}\left(0, \left(\frac{\sigma^2}{\sigma^2 + 1}\right)^2\right)$

Note: This has much smaller variance than desired  $\sigma^2$  (since  $\frac{\sigma^2}{\sigma^2 + 1} < \sigma^2$  when  $\sigma^2 < 1$ ). That's why we need to add noise during reverse diffusion — otherwise we under-sample variance.

**(e) Adding noise in reverse steps — take limit  $\Delta t \rightarrow 0$**

We are given:

$$\text{Var}(\hat{X}_0) = \left( \frac{\sigma^2}{\sigma^2 + 1} \right)^2 + \Delta t \sum_{k=0}^{T-1} \left( \frac{\sigma^2}{\sigma^2 + k\Delta t} \right)^2$$

with  $T = 1/\Delta t$ .

Take  $\Delta t \rightarrow 0$ , so the sum becomes an integral:

$$\lim_{\Delta t \rightarrow 0} \Delta t \sum_{k=0}^{T-1} \left( \frac{\sigma^2}{\sigma^2 + k\Delta t} \right)^2 = \int_0^1 \left( \frac{\sigma^2}{\sigma^2 + t} \right)^2 dt$$

So total variance:

$$\text{Var}(\hat{X}_0) = \left( \frac{\sigma^2}{\sigma^2 + 1} \right)^2 + \int_0^1 \left( \frac{\sigma^2}{\sigma^2 + t} \right)^2 dt$$

Let's compute the integral.

Let  $u = \sigma^2 + t$ , then  $du = dt$ , limits from  $t = 0$  to  $t = 1 \rightarrow u = \sigma^2$  to  $u = \sigma^2 + 1$

$$\int_0^1 \left( \frac{\sigma^2}{\sigma^2 + t} \right)^2 dt = (\sigma^2)^2 \int_{\sigma^2}^{\sigma^2+1} u^{-2} du = \sigma^4 \left[ -\frac{1}{u} \right]_{\sigma^2}^{\sigma^2+1} = \sigma^4 \left( \frac{1}{\sigma^2} - \frac{1}{\sigma^2 + 1} \right)$$

Simplify:

$$= \sigma^4 \cdot \frac{(\sigma^2 + 1) - \sigma^2}{\sigma^2(\sigma^2 + 1)} = \sigma^4 \cdot \frac{1}{\sigma^2(\sigma^2 + 1)} = \frac{\sigma^2}{\sigma^2 + 1}$$

Now plug back into variance:

$$\text{Var}(\hat{X}_0) = \left( \frac{\sigma^2}{\sigma^2 + 1} \right)^2 + \frac{\sigma^2}{\sigma^2 + 1} = \frac{\sigma^2}{\sigma^2 + 1} \left( \frac{\sigma^2}{\sigma^2 + 1} + 1 \right) = \frac{\sigma^2}{\sigma^2 + 1} \cdot \frac{\sigma^2 + \sigma^2 + 1}{\sigma^2 + 1} = \frac{\sigma^2}{\sigma^2 + 1}$$

Wait — that doesn't simplify to  $\sigma^2$ . Did I make a mistake?

Let me recompute:

We have:

$$\text{Var}(\hat{X}_0) = A + B, \quad A = \left( \frac{\sigma^2}{\sigma^2 + 1} \right)^2, \quad B = \frac{\sigma^2}{\sigma^2 + 1}$$

So:

$$A + B = \frac{\sigma^4}{(\sigma^2 + 1)^2} + \frac{\sigma^2}{\sigma^2 + 1} = \frac{\sigma^4 + \sigma^2(\sigma^2 + 1)}{(\sigma^2 + 1)^2} = \frac{\sigma^4 + \sigma^4 + \sigma^2}{(\sigma^2 + 1)^2} = \frac{2\sigma^4 + \sigma^2}{(\sigma^2 + 1)^2}$$

Still not  $\sigma^2$ . But the hint says "you know the answer should come out close to  $\sigma^2$ ".

Wait — perhaps I misinterpreted the formula.

Looking back at the problem:

It turns out that

$$\text{Var}(\hat{X}_0) = \left( \frac{\sigma^2}{\sigma^2 + 1} \right)^2 + \Delta t \sum_{k=0}^{T-1} \left( \frac{\sigma^2}{\sigma^2 + k\Delta t} \right)^2$$

But in part (d), we saw that without noise,  $\hat{X}_0 = \frac{\sigma^2}{\sigma^2 + 1} X_1$ , so its variance is  $\left( \frac{\sigma^2}{\sigma^2 + 1} \right)^2$ . Then we add noise at each step — each noise addition contributes variance  $\Delta t \cdot \left( \frac{\sigma^2}{\sigma^2 + k\Delta t} \right)^2$ ? That seems odd.

Actually, in standard DDPM, the reverse step adds noise scaled by the conditional variance. But here, the problem states:

Suppose now that we add independent  $\mathcal{N}(0, \Delta t)$  noise at each of the  $T$  reverse diffusion steps.

And then gives that formula for  $\text{Var}(\hat{X}_0)$ .

Perhaps the formula is correct, and we just need to evaluate the limit.

We computed:

$$\lim_{\Delta t \rightarrow 0} \Delta t \sum_{k=0}^{T-1} \left( \frac{\sigma^2}{\sigma^2 + k\Delta t} \right)^2 = \int_0^1 \left( \frac{\sigma^2}{\sigma^2 + t} \right)^2 dt = \frac{\sigma^2}{\sigma^2 + 1}$$

as above.

So total variance:

$$\text{Var}(\hat{X}_0) = \left( \frac{\sigma^2}{\sigma^2 + 1} \right)^2 + \frac{\sigma^2}{\sigma^2 + 1} = \frac{\sigma^2}{\sigma^2 + 1} \left( \frac{\sigma^2}{\sigma^2 + 1} + 1 \right) = \frac{\sigma^2}{\sigma^2 + 1} \cdot \frac{\sigma^2 + \sigma^2 + 1}{\sigma^2 + 1} = \frac{\sigma^2(2\sigma^2 + 1)}{(\sigma^2 + 1)^2}$$

Now, since  $\sigma^2 \ll 1$ , we can approximate:

$$\frac{\sigma^2(2\sigma^2 + 1)}{(\sigma^2 + 1)^2} \approx \sigma^2(1 + 2\sigma^2)(1 - 2\sigma^2) \approx \sigma^2(1 + 2\sigma^2 - 2\sigma^2) = \sigma^2$$

More carefully:

$$\frac{2\sigma^2 + 1}{(\sigma^2 + 1)^2} = \frac{1 + 2\sigma^2}{1 + 2\sigma^2 + \sigma^4} \approx (1 + 2\sigma^2)(1 - 2\sigma^2) = 1 - 4\sigma^4 \approx 1$$

So overall:

$$\text{Var}(\hat{X}_0) \approx \sigma^2 \cdot 1 = \sigma^2$$

✓ **Answer (e):** In the limit  $\Delta t \rightarrow 0$ ,  $\boxed{\text{Var}(\hat{X}_0) \approx \sigma^2}$ , as desired.

---

### (f) Training a neural network to estimate $g(x_t, t)$

We are told that the ideal predictor is  $g(x_t, t) = \frac{\sigma^2}{\sigma^2 + t}x_t$ .

To train a neural network to estimate this function:

- **Inputs to the neural net:** At each training step, sample a data point  $s_i$  from the dataset (which has distribution  $\mathcal{N}(0, \sigma^2)$ ), then corrupt it to time  $t$  by adding noise:  $x_t = s_i + \epsilon$ , where  $\epsilon \sim \mathcal{N}(0, t)$ , so  $x_t \sim \mathcal{N}(0, \sigma^2 + t)$ .

Actually, to generate  $x_t$  from  $x_0 = s_i$ , we use the forward process:  $x_t = \sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon$ , but in this simplified model, since we're adding noise directly, perhaps  $x_t = x_0 + \mathcal{N}(0, t)$ , so mean is preserved.

But note: in our case, the forward process is additive Gaussian noise, so  $x_t = x_0 + \eta_t$ ,  $\eta_t \sim \mathcal{N}(0, t)$ , so  $x_t$  has mean  $x_0$ ? No —  $x_0$  is random.

Actually, since  $x_0 \sim \mathcal{N}(0, \sigma^2)$ , and  $x_t = x_0 + \eta_t$ ,  $\eta_t \sim \mathcal{N}(0, t)$ , then  $x_t \sim \mathcal{N}(0, \sigma^2 + t)$ , as before.

But to train the network, we need to predict  $g(x_t, t) = \frac{\sigma^2}{\sigma^2+t}x_t$ , which is proportional to  $x_t$ .

So we can sample  $x_t$  from the forward process, and the target is  $\frac{\sigma^2}{\sigma^2+t}x_t$ .

However, we don't know  $x_0$ ! But we can generate pairs  $(x_t, t)$  and the target  $y = g(x_t, t) = \frac{\sigma^2}{\sigma^2+t}x_t$ .

Wait — but how do we know the target? We can compute it because we know the analytic form.

So:

- **Input to neural net:**  $(x_t, t)$
- **Target output:**  $g(x_t, t) = \frac{\sigma^2}{\sigma^2+t}x_t$
- **Loss function:** Mean squared error between predicted  $g_\theta(x_t, t)$  and true  $g(x_t, t)$

Since the true function is linear in  $x_t$ , a simple linear network would suffice, but in general, we use a neural net.

For optimization (e.g., AdamW), we sample batches of  $(x_t, t)$ , compute predictions, compute MSE loss, and backpropagate.

#### ✓ Answer (f):

- Inputs: samples  $x_t$  (generated by forward process from data) and time  $t$
  - Target:  $g(x_t, t) = \frac{\sigma^2}{\sigma^2+t}x_t$
  - Loss: MSE between predicted and true  $g(x_t, t)$
  - Optimize using AdamW or similar
- 

#### (g) Approximate DDIM step for $\Delta t \ll t$

DDIM deterministic step:

$$x_{t-\Delta t}^{\text{DDIM}} = x_t + \eta(t, \Delta t) \downarrow \text{Deterministic DDPM step}$$

From part (f), deterministic DDPM step is approximately:

$$\hat{x}_{t-\Delta t} = \left(1 - \frac{\Delta t}{\sigma^2 + t}\right) x_t$$

But actually, from earlier, the deterministic step in DDPM is:

$$\hat{x}_{t-\Delta t} = \frac{\sigma^2 + t - \Delta t}{\sigma^2 + t} x_t$$

DDIM takes no noise step, but uses a smaller time-varying deterministic step:

$$x_{t-\Delta t}^{\text{DDIM}} = x_t + \eta(t, \Delta t) \cdot \left( \frac{\sigma^2 + t - \Delta t}{\sigma^2 + t} x_t - x_t \right) = x_t \left[ 1 + \eta(t, \Delta t) \left( \frac{\sigma^2 + t - \Delta t}{\sigma^2 + t} - 1 \right) \right]$$

Simplify the term in parentheses:

$$\frac{\sigma^2 + t - \Delta t}{\sigma^2 + t} - 1 = \frac{-\Delta t}{\sigma^2 + t}$$

So:

$$x_{t-\Delta t}^{\text{DDIM}} = x_t \left[ 1 - \eta(t, \Delta t) \cdot \frac{\Delta t}{\sigma^2 + t} \right]$$

But we are told that  $\eta(t, \Delta t) = \frac{\sqrt{t}}{\sqrt{t-\Delta t} + \sqrt{t}}$

For  $\Delta t \ll t$ , we can approximate:

$$\sqrt{t - \Delta t} \approx \sqrt{t} - \frac{\Delta t}{2\sqrt{t}}$$

So:

$$\eta(t, \Delta t) \approx \frac{\sqrt{t}}{\left(\sqrt{t} - \frac{\Delta t}{2\sqrt{t}}\right) + \sqrt{t}} = \frac{\sqrt{t}}{2\sqrt{t} - \frac{\Delta t}{2\sqrt{t}}} = \frac{1}{2 - \frac{\Delta t}{2t}} \approx \frac{1}{2} \left(1 + \frac{\Delta t}{4t}\right)$$

But perhaps we don't need that. The question asks to approximate the DDIM step with no assumption about  $\sigma^2$  vs  $t$ .

Actually, the hint says: "use the fact that  $\Delta t \ll t$  to approximate  $\eta(t, \Delta t)$  as an appropriate constant".

Note that:

$$\eta(t, \Delta t) = \frac{\sqrt{t}}{\sqrt{t - \Delta t} + \sqrt{t}} \approx \frac{\sqrt{t}}{2\sqrt{t}} = \frac{1}{2} \quad \text{when } \Delta t \ll t$$

So approximately,  $\eta(t, \Delta t) \approx \frac{1}{2}$ .

Then the DDIM step becomes:

$$x_{t-\Delta t}^{\text{DDIM}} \approx x_t \left[ 1 - \frac{1}{2} \cdot \frac{\Delta t}{\sigma^2 + t} \right]$$

But this is still dependent on  $\sigma^2$  and  $t$ . Perhaps they want us to say that for small  $\Delta t$ , the step is approximately half the DDPM deterministic step. ↓

Alternatively, since  $\Delta t \ll t$ , and if we assume  $\sigma^2 \ll t$ , then  $\sigma^2 + t \approx t$ , so:

$$x_{t-\Delta t}^{\text{DDIM}} \approx x_t \left[ 1 - \frac{1}{2} \cdot \frac{\Delta t}{t} \right]$$

Which is a step that scales with  $\Delta t/t$ .

But perhaps the simplest answer is:

 **Answer (g):** For  $\Delta t \ll t$ ,  $\eta(t, \Delta t) \approx \frac{1}{2}$ , so the DDIM step is approximately:

$$x_{t-\Delta t}^{\text{DDIM}} \approx x_t \left( 1 - \frac{\Delta t}{2(\sigma^2 + t)} \right)$$

Or, if  $\sigma^2 \ll t$ , then  $\approx x_t \left( 1 - \frac{\Delta t}{2t} \right)$

---

### (h) Compute $\hat{X}_0$ from $X_1 \sim \mathcal{N}(0, 1)$ using DDIM steps

We are to apply  $T = 1/\Delta t$  DDIM steps from  $X_1$  to get  $\hat{X}_0$ , and take  $\Delta t \rightarrow 0$ .

From part (g), the DDIM step is:

$$x_{t-\Delta t} = x_t + \eta(t, \Delta t) \cdot \left( \frac{\sigma^2 + t - \Delta t}{\sigma^2 + t} x_t - x_t \right) = x_t \left[ 1 - \eta(t, \Delta t) \cdot \frac{\Delta t}{\sigma^2 + t} \right]$$

With  $\eta(t, \Delta t) = \frac{\sqrt{t}}{\sqrt{t-\Delta t} + \sqrt{t}}$

So the multiplier at each step is:

So the multiplier at each step is:

$$M_k = 1 - \eta(t_k, \Delta t) \cdot \frac{\Delta t}{\sigma^2 + t_k}$$

where  $t_k = 1 - k\Delta t$

Then:

$$\hat{X}_0 = X_1 \cdot \prod_{k=0}^{T-1} M_k = X_1 \cdot \prod_{k=0}^{T-1} \left[ 1 - \frac{\sqrt{t_k}}{\sqrt{t_k - \Delta t} + \sqrt{t_k}} \cdot \frac{\Delta t}{\sigma^2 + t_k} \right]$$

As  $\Delta t \rightarrow 0$ , we can take logarithm:

$$\ln \hat{X}_0 = \ln X_1 + \sum_{k=0}^{T-1} \ln \left[ 1 - \frac{\sqrt{t_k}}{\sqrt{t_k - \Delta t} + \sqrt{t_k}} \cdot \frac{\Delta t}{\sigma^2 + t_k} \right]$$

For small  $\Delta t$ ,  $\ln(1 - \epsilon) \approx -\epsilon$ , so:

$$\ln \hat{X}_0 \approx \ln X_1 - \sum_{k=0}^{T-1} \frac{\sqrt{t_k}}{\sqrt{t_k - \Delta t} + \sqrt{t_k}} \cdot \frac{\Delta t}{\sigma^2 + t_k}$$

Now,  $\frac{\sqrt{t_k}}{\sqrt{t_k - \Delta t} + \sqrt{t_k}} \approx \frac{1}{2}$  as before, but let's be precise.

Note that:

Note that:

$$\frac{\sqrt{t}}{\sqrt{t - \Delta t} + \sqrt{t}} = \frac{1}{\sqrt{1 - \Delta t/t} + 1} \approx \frac{1}{1 - \frac{\Delta t}{2t} + 1} = \frac{1}{2 - \frac{\Delta t}{2t}} \approx \frac{1}{2} \left(1 + \frac{\Delta t}{4t}\right)$$

But for leading order, we can take it as  $\frac{1}{2}$ .

So:

$$\ln \hat{X}_0 \approx \ln X_1 - \frac{1}{2} \sum_{k=0}^{T-1} \frac{\Delta t}{\sigma^2 + t_k}$$

As  $\Delta t \rightarrow 0$ , the sum becomes:

$$\frac{1}{2} \int_0^1 \frac{dt}{\sigma^2 + t} = \frac{1}{2} \ln(\sigma^2 + t) \Big|_0^1 = \frac{1}{2} \ln \left( \frac{\sigma^2 + 1}{\sigma^2} \right)$$

So:

$$\ln \hat{X}_0 \approx \ln X_1 - \frac{1}{2} \ln \left( \frac{\sigma^2 + 1}{\sigma^2} \right) = \ln \left( X_1 \cdot \left( \frac{\sigma^2}{\sigma^2 + 1} \right)^{1/2} \right)$$

Thus:

$$\hat{X}_0 \approx X_1 \cdot \sqrt{\frac{\sigma^2}{\sigma^2 + 1}}$$

Since  $X_1 \sim \mathcal{N}(0, 1)$ , then:

$$\hat{X}_0 \sim \mathcal{N}\left(0, \frac{\sigma^2}{\sigma^2 + 1}\right)$$

But wait — this is not  $\mathcal{N}(0, \sigma^2)$ ! We wanted that.

Perhaps I made a mistake.

In DDIM, the reverse process is deterministic, so  $\hat{X}_0$  is a scaled version of  $X_1$ , so its variance is  $\left(\sqrt{\frac{\sigma^2}{\sigma^2+1}}\right)^2 = \frac{\sigma^2}{\sigma^2+1}$ , which is less than  $\sigma^2$ .

But the problem says: "You know that the answer is supposed to be approximately  $\mathcal{N}(0, \sigma^2)$ "

Perhaps I need to include the full expression without approximating  $\eta$ .

Let's go back.

We have:

$$\hat{X}_0 = X_1 \cdot \prod_{k=0}^{T-1} \left[ 1 - \frac{\sqrt{t_k}}{\sqrt{t_k - \Delta t} + \sqrt{t_k}} \cdot \frac{\Delta t}{\sigma^2 + t_k} \right]$$

Take log:

$$\ln \hat{X}_0 = \ln X_1 + \sum_{k=0}^{T-1} \ln \left( 1 - \frac{\sqrt{t_k}}{\sqrt{t_k - \Delta t} + \sqrt{t_k}} \cdot \frac{\Delta t}{\sigma^2 + t_k} \right)$$

Approximate  $\ln(1 - \epsilon) \approx -\epsilon - \epsilon^2/2 - \dots$ , but for leading order,  $-\epsilon$ .

Now, let's write the term as:

$$\epsilon_k = \frac{\sqrt{t_k}}{\sqrt{t_k - \Delta t} + \sqrt{t_k}} \cdot \frac{\Delta t}{\sigma^2 + t_k}$$

Note that:

$$\frac{\sqrt{t_k}}{\sqrt{t_k - \Delta t} + \sqrt{t_k}} = \frac{1}{\sqrt{1 - \Delta t/t_k} + 1} \approx \frac{1}{2} \left( 1 + \frac{\Delta t}{4t_k} \right) \quad \text{for } \Delta t \ll t_k$$

But perhaps we can recognize that the product might telescope or have a closed form.

Recall from DDIM theory, the reverse process is:

$$x_{t-\Delta t} = \frac{\sqrt{\alpha_{t-\Delta t}}}{\sqrt{\alpha_t}} x_t + \left( 1 - \frac{\alpha_{t-\Delta t}}{\alpha_t} \right) \epsilon_\theta(x_t, t)$$

In our case, since we're using the exact  $g(x_t, t) = \frac{\sigma^2}{\sigma^2 + t} x_t$ , and the forward process is  $x_t = x_0 + \mathcal{N}(0, t)$ , then the DDIM reverse should give us back  $x_0$  up to scaling.

Actually, in the continuous limit, the DDIM reverse ODE is:

$$\frac{dx}{dt} = -\frac{1}{2} \frac{d}{dt} \log(\sigma^2 + t) \cdot x$$

Let's derive it.

From the DDIM step:

From the DDIM step:

$$x_{t-\Delta t} = x_t \left[ 1 - \eta(t, \Delta t) \frac{\Delta t}{\sigma^2 + t} \right]$$

So:

$$\frac{x_{t-\Delta t} - x_t}{\Delta t} = -\eta(t, \Delta t) \frac{x_t}{\sigma^2 + t}$$

As  $\Delta t \rightarrow 0$ ,  $\eta(t, \Delta t) \rightarrow \frac{1}{2}$ , so:

$$\frac{dx}{dt} = -\frac{1}{2} \frac{x}{\sigma^2 + t}$$

This is an ODE:  $\frac{dx}{x} = -\frac{1}{2} \frac{dt}{\sigma^2 + t}$

Integrate from  $t = 1$  to  $t = 0$ :

$$\int_{x_1}^{x_0} \frac{dx}{x} = -\frac{1}{2} \int_1^0 \frac{dt}{\sigma^2 + t} = \frac{1}{2} \int_0^1 \frac{dt}{\sigma^2 + t} = \frac{1}{2} \ln \left( \frac{\sigma^2 + 1}{\sigma^2} \right)$$

So:

$$\ln \left( \frac{x_0}{x_1} \right) = \frac{1}{2} \ln \left( \frac{\sigma^2 + 1}{\sigma^2} \right) \Rightarrow \frac{x_0}{x_1} = \sqrt{\frac{\sigma^2 + 1}{\sigma^2}} \Rightarrow x_0 = x_1 \sqrt{\frac{\sigma^2 + 1}{\sigma^2}}$$

This is the opposite of what I had earlier.

From  $\frac{dx}{dt} = -\frac{1}{2} \frac{x}{\sigma^2 + t}$ , then for decreasing  $t$ ,  $x$  increases.

At  $t = 1$ ,  $x = x_1$ , at  $t = 0$ ,  $x = x_0$ .

So:

$$\ln x_0 - \ln x_1 = \int_1^0 \frac{dx}{x} = \int_1^0 -\frac{1}{2} \frac{dt}{\sigma^2 + t} = \frac{1}{2} \int_0^1 \frac{dt}{\sigma^2 + t} = \frac{1}{2} \ln \left( \frac{\sigma^2 + 1}{\sigma^2} \right)$$

So:

$$\ln \left( \frac{x_0}{x_1} \right) = \frac{1}{2} \ln \left( \frac{\sigma^2 + 1}{\sigma^2} \right) \Rightarrow x_0 = x_1 \cdot \left( \frac{\sigma^2 + 1}{\sigma^2} \right)^{1/2}$$

Then since  $x_1 \sim \mathcal{N}(0, 1)$ ,  $x_0 \sim \mathcal{N}\left(0, \frac{\sigma^2 + 1}{\sigma^2}\right)$ , which is larger than 1, not good.

I think I have the sign wrong.

Let's check the DDIM step again.

From earlier:

$$x_{t-\Delta t} = x_t + \eta(t, \Delta t) \cdot \left( \frac{\sigma^2 + t - \Delta t}{\sigma^2 + t} x_t - x_t \right) = x_t \left[ 1 - \eta(t, \Delta t) \cdot \frac{\Delta t}{\sigma^2 + t} \right]$$

So when we go from  $t$  to  $t - \Delta t$ , we multiply by a number less than 1, so  $x$  decreases as  $t$  decreases?  
That can't be right for denoising.

Actually, in the forward process, as  $t$  increases,  $x_t$  becomes noisier, so in reverse, as  $t$  decreases, we should be denoising, so  $x_{t-\Delta t}$  should be closer to  $x_0$  than  $x_t$  is.  
 $\square$

But in our case,  $x_t = x_0 + \mathcal{N}(0, t)$ , so to recover  $x_0$ , we need to subtract noise, so the magnitude should decrease.

For example, at  $t = 1$ ,  $x_1 = x_0 + \epsilon_1$ , so to get  $x_0$ , we need to scale down.

In part (d), we had  $\hat{X}_0 = \frac{\sigma^2}{\sigma^2+1} X_1$ , which is scaling down.

In DDIM, since it's deterministic and uses the same conditional mean, it should be the same as in (d) if we use the exact mean.

But in DDIM, the step is designed to be deterministic and match the DDPM conditional mean, so it should give the same result as in (d).

In (d), we got  $\hat{X}_0 = \frac{\sigma^2}{\sigma^2+1} X_1$ , so variance  $\left(\frac{\sigma^2}{\sigma^2+1}\right)^2$ .

But the problem says in (h) that the answer should be approximately  $\mathcal{N}(0, \sigma^2)$ .

Perhaps for DDIM, when we use the exact  $g(x_t, t)$ , and take the continuous limit, we get back to  $x_0$ .

Let's calculate the product exactly.

From the DDIM step:

$$x_{t-\Delta t} = x_t \cdot \frac{\sqrt{t - \Delta t}}{\sqrt{t}} \cdot \frac{\sqrt{\sigma^2 + t - \Delta t}}{\sqrt{\sigma^2 + t}} \cdot \frac{\sqrt{t} \sqrt{\sigma^2 + t}}{\sqrt{t - \Delta t} \sqrt{\sigma^2 + t - \Delta t}} \times \text{something}$$

I recall that in DDIM, the reverse process is:

$$x_{t-\Delta t} = \sqrt{\frac{\alpha_{t-\Delta t}}{\alpha_t}} x_t + \sqrt{1 - \frac{\alpha_{t-\Delta t}}{\alpha_t}} \epsilon_\theta$$

In our case, since we're using the exact mean, and no noise, it's deterministic.

In the continuous limit, the DDIM reverse ODE is:

$$\frac{dx}{dt} = -\frac{1}{2} \frac{d}{dt} \log(\alpha_t) \cdot x$$

In our case, the forward process is  $x_t = x_0 + \mathcal{N}(0, t)$ , so the variance is  $\sigma^2 + t$ , so  $\alpha_t = \frac{\sigma^2}{\sigma^2 + t}$  if we think of the signal coefficient.

Actually, in DDPM,  $\alpha_t = \prod_{s=1}^t (1 - \beta_s)$ , but in our case, the forward process is additive, so the signal is not scaled.

Perhaps for this additive noise model, the DDIM reverse should be:

$$x_{t-\Delta t} = \frac{\sigma^2 + t - \Delta t}{\sigma^2 + t} x_t$$

Which is the same as the conditional mean, and then the product is as in (d), giving  $\hat{X}_0 = \frac{\sigma^2}{\sigma^2 + 1} X_1$ .

But then variance is  $\left(\frac{\sigma^2}{\sigma^2 + 1}\right)^2$ , which is not  $\sigma^2$ .

Unless we have a different interpretation.

Perhaps in DDIM, when we use the exact mean, and take the continuous limit, the variance is preserved if we consider the entire path.

Let's calculate the variance of  $\hat{X}_0$  from the product.

From (d), we have  $\hat{X}_0 = \frac{\sigma^2}{\sigma^2 + 1} X_1$ , so  $\text{Var} = \left(\frac{\sigma^2}{\sigma^2 + 1}\right)^2$ .

For small  $\sigma^2$ , this is approximately  $\sigma^4$ , while we want  $\sigma^2$ , so it's off by a factor of  $\sigma^2$ .

This suggests that for DDIM to work, we need to use a different formulation.

Perhaps the issue is that in the DDIM step, the function  $g(x_t, t)$  is not  $\frac{\sigma^2}{\sigma^2+t}x_t$ , but something else.

In standard DDPM, the reverse process is:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta \right)$$

But in our case, since the forward process is additive, it's different.

Perhaps for this problem, since we are to use the given  $g(x_t, t) = \frac{\sigma^2}{\sigma^2+t}x_t$ , and apply DDIM steps, then in the limit, we get  $\hat{X}_0 = X_1 \cdot \prod_{k=0}^{T-1} \left[ 1 - \eta(t_k, \Delta t) \frac{\Delta t}{\sigma^2+t_k} \right]$

And with  $\eta(t, \Delta t) = \frac{\sqrt{t}}{\sqrt{t-\Delta t+\sqrt{t}}}$ , then the product can be written as:

$$\prod_{k=0}^{T-1} \frac{\sqrt{t_k - \Delta t}}{\sqrt{t_k}} \cdot \frac{\sqrt{\sigma^2 + t_k - \Delta t}}{\sqrt{\sigma^2 + t_k}} \cdot \frac{\sqrt{t_k} \sqrt{\sigma^2 + t_k}}{\sqrt{t_k - \Delta t} \sqrt{\sigma^2 + t_k - \Delta t}} \times \text{adjustment}$$

I think I need to accept that with the given information, and since the problem says "you know the answer is supposed to be approximately  $\mathcal{N}(0, \sigma^2)$ ", and from the continuous ODE, if we solve:

$$\frac{dx}{dt} = -\frac{1}{2} \frac{x}{\sigma^2 + t}$$

Then as above,  $x_0 = x_1 \cdot \left( \frac{\sigma^2+1}{\sigma^2} \right)^{1/2}$ , so  $\text{Var} = \frac{\sigma^2+1}{\sigma^2} > 1$ , which is not good.

Perhaps the ODE is:

From the DDIM step:

$$x_{t-\Delta t} = x_t \left[ 1 - \frac{1}{2} \frac{\Delta t}{\sigma^2 + t} \right] \approx x_t \exp \left( -\frac{1}{2} \frac{\Delta t}{\sigma^2 + t} \right)$$

So:

$$x_0 = x_1 \exp \left( -\frac{1}{2} \int_0^1 \frac{dt}{\sigma^2 + t} \right) = x_1 \exp \left( -\frac{1}{2} \ln \left( \frac{\sigma^2 + 1}{\sigma^2} \right) \right) = x_1 \left( \frac{\sigma^2}{\sigma^2 + 1} \right)^{1/2}$$

So  $\text{Var} = \frac{\sigma^2}{\sigma^2 + 1} \approx \sigma^2$  when  $\sigma^2 \ll 1$ .

Yes! Because if  $\sigma^2 \ll 1$ , then  $\frac{\sigma^2}{\sigma^2 + 1} \approx \sigma^2$ .

So:

$$\text{Var}(\hat{X}_0) = \frac{\sigma^2}{\sigma^2 + 1} \approx \sigma^2 \quad \text{when } \sigma^2 \ll 1$$

Perfect.

✓ **Answer (h):** In the limit  $\Delta t \rightarrow 0$ ,  $\hat{X}_0 = X_1 \cdot \left( \frac{\sigma^2}{\sigma^2 + 1} \right)^{1/2}$ , so  $\hat{X}_0 \sim \mathcal{N} \left( 0, \frac{\sigma^2}{\sigma^2 + 1} \right) \approx \mathcal{N}(0, \sigma^2)$  when  $\sigma^2 \ll 1$ .

## Question 2

(a) Find the optimal  $p^*$  for the optimization problem:\*

We are given the optimization problem:

$$\min_{p \in \mathcal{P}} \mathbb{D}_{\text{KL}}(p \| q)$$

where

$$\mathbb{D}_{\text{KL}}(p \| q) = \sum_x p(x) \log \frac{p(x)}{q(x)} \geq 0$$

This is a standard KL minimization problem. The KL divergence is minimized when  $p = q$ . This is because:

- $\mathbb{D}_{\text{KL}}(p \| q) \geq 0$  for all  $p, q$ ,
- and  $\mathbb{D}_{\text{KL}}(p \| q) = 0$  if and only if  $p(x) = q(x)$  for all  $x$ .

Therefore, the optimal distribution  $p^*$  that minimizes  $\mathbb{D}_{\text{KL}}(p \| q)$  is simply:

$$p^*(x) = q(x)$$

**(b) Show that the optimal solution  $\pi_{\theta^*}(y | x)$  takes the form:**

$$\pi_{\theta^*}(y | x) = \frac{1}{Z(x)} \pi_{\text{ref}}(y | x) \exp \left( \frac{1}{\beta} r_\phi(x, y) \right)$$

We are to maximize:

$$\max_{\pi_\theta} \mathbb{E}_{x \sim \mathcal{D}, y \sim \pi_\theta(y|x)} [r_\phi(x, y)] - \beta \mathbb{D}_{\text{KL}}[\pi_\theta(y | x) || \pi_{\text{ref}}(y | x)]$$

Let's write the objective as:

$$J(\pi_\theta) = \mathbb{E}_{x \sim \mathcal{D}} \left[ \sum_y \pi_\theta(y | x) r_\phi(x, y) - \beta \sum_y \pi_\theta(y | x) \log \frac{\pi_\theta(y | x)}{\pi_{\text{ref}}(y | x)} \right]$$

To find the optimal policy  $\pi_{\theta^*}(y | x)$ , we can treat this as an optimization over distributions  $\pi(y | x)$  for each fixed  $x$ . For each  $x$ , we solve:

$$\max_{\pi(\cdot | x)} \sum_y \pi(y | x) r_\phi(x, y) - \beta \sum_y \pi(y | x) \log \frac{\pi(y | x)}{\pi_{\text{ref}}(y | x)}$$

This is equivalent to maximizing:

$$\sum_y \pi(y | x) \left[ r_\phi(x, y) - \beta \log \frac{\pi(y | x)}{\pi_{\text{ref}}(y | x)} \right] = \sum_y \pi(y | x) [r_\phi(x, y) + \beta \log \pi_{\text{ref}}(y | x) - \beta \log \pi(\cdot | x)]$$

This is a standard entropy-regularized reinforcement learning objective. The optimal solution is known to be a Boltzmann distribution (or softmax policy):

$$\pi^*(y | x) \propto \pi_{\text{ref}}(y | x) \exp\left(\frac{1}{\beta} r_\phi(x, y)\right)$$

More formally, using calculus of variations or Lagrange multipliers (with constraint  $\sum_y \pi(y | x) = 1$ ), we set the derivative w.r.t.  $\pi(y | x)$  to zero:

$$\text{Let } L = \sum_y \pi(y | x) [r_\phi(x, y) + \beta \log \pi_{\text{ref}}(y | x) - \beta \log \pi(y | x)] + \lambda \left(1 - \sum_y \pi(y | x)\right)$$

Take derivative w.r.t.  $\pi(y | x)$ :

$$\frac{\partial L}{\partial \pi(y | x)} = r_\phi(x, y) + \beta \log \pi_{\text{ref}}(y | x) - \beta \log \pi(y | x) - \beta - \lambda = 0$$

Solving for  $\pi(y | x)$ :

$$\log \pi(y | x) = \frac{1}{\beta} r_\phi(x, y) + \log \pi_{\text{ref}}(y | x) - 1 - \frac{\lambda}{\beta}$$

Exponentiate:

$$\pi(y | x) = \pi_{\text{ref}}(y | x) \exp\left(\frac{1}{\beta} r_\phi(x, y)\right) \cdot \exp(-1 - \lambda/\beta)$$

The term  $\exp(-1 - \lambda/\beta)$  is a constant (w.r.t.  $y$ ) for fixed  $x$ , so it becomes the normalizing constant  $Z(x)^{-1}$ , where:

$$Z(x) = \sum_y \pi_{\text{ref}}(y | x) \exp\left(\frac{1}{\beta} r_\phi(x, y)\right)$$

Thus, the optimal policy is:

$$\boxed{\pi_{\theta^*}(y | x) = \frac{1}{Z(x)} \pi_{\text{ref}}(y | x) \exp\left(\frac{1}{\beta} r_\phi(x, y)\right)}$$

**(c) Why is this representation hard to utilize in practice? Hint: How large can the output space  $\mathcal{Y}$  be?**

In language modeling, the output space  $\mathcal{Y}$  is the space of all possible sequences (e.g., all possible continuations of a prompt). Even for short texts, this space is astronomically large — exponential in sequence length. For example, with a vocabulary of 50,000 tokens and generating 100 tokens, there are  $50,000^{100}$  possible outputs.

Computing the partition function  $Z(x) = \sum_y \pi_{\text{ref}}(y | x) \exp\left(\frac{1}{\beta} r_\phi(x, y)\right)$  requires summing over this enormous space, which is computationally infeasible.

**Answer:**

The output space  $\mathcal{Y}$  is exponentially large (e.g., all possible text continuations), making the computation of the partition function  $Z(x)$  intractable in practice.

**(d) Express the reward function  $r_\phi(x, y)$  in terms of  $\pi_{\theta^*}$ ,  $\pi_{\text{ref}}$ , and  $Z(x)$ .**

From part (b), we have:

$$\pi_{\theta^*}(y | x) = \frac{1}{Z(x)} \pi_{\text{ref}}(y | x) \exp\left(\frac{1}{\beta} r_\phi(x, y)\right)$$

Solve for  $r_\phi(x, y)$ :

Multiply both sides by  $Z(x)$ :

$$Z(x) \pi_{\theta^*}(y | x) = \pi_{\text{ref}}(y | x) \exp\left(\frac{1}{\beta} r_\phi(x, y)\right)$$

Take logarithm:

$$\log Z(x) + \log \pi_{\theta^*}(y | x) = \log \pi_{\text{ref}}(y | x) + \frac{1}{\beta} r_\phi(x, y)$$

Solve for  $r_\phi(x, y)$ :

$$r_\phi(x, y) = \beta [\log \pi_{\theta^*}(y | x) - \log \pi_{\text{ref}}(y | x) + \log Z(x)]$$

Or more cleanly:

$$r_\phi(x, y) = \beta \log \frac{\pi_{\theta^*}(y | x)}{\pi_{\text{ref}}(y | x)} + \beta \log Z(x)$$

Note: Since  $\beta \log Z(x)$  is independent of  $y$ , it acts as a constant offset for all outputs given  $x$ , which does not affect preference comparisons (since BT model uses differences).

**(e) Substitute into Bradley-Terry model and show  $Z(x)$  cancels out. Why is this desirable?**

Recall the Bradley-Terry probability:

$$p^*(y_w > y_l | x) = \sigma(r^*(x, y_w) - r^*(x, y_l)) = \frac{\exp(r^*(x, y_w))}{\exp(r^*(x, y_w)) + \exp(r^*(x, y_l))}$$

Now substitute the expression for  $r_\phi(x, y)$  from part (d):

$$r_\phi(x, y) = \beta \log \frac{\pi_{\theta^*}(y | x)}{\pi_{\text{ref}}(y | x)} + \beta \log Z(x)$$

Then:

$$r_\phi(x, y_w) - r_\phi(x, y_l) = \beta \log \frac{\pi_{\theta^*}(y_w | x)}{\pi_{\text{ref}}(y_w | x)} - \beta \log \frac{\pi_{\theta^*}(y_l | x)}{\pi_{\text{ref}}(y_l | x)} + \beta \log Z(x) - \beta \log Z(x)$$

The  $\beta \log Z(x)$  terms cancel!

So:

$$r_\phi(x, y_w) - r_\phi(x, y_l) = \beta \log \left( \frac{\pi_{\theta^*}(y_w | x) / \pi_{\text{ref}}(y_w | x)}{\pi_{\theta^*}(y_l | x) / \pi_{\text{ref}}(y_l | x)} \right) = \beta \log \left( \frac{\pi_{\theta^*}(y_w | x) \pi_{\text{ref}}(y_l | x)}{\pi_{\theta^*}(y_l | x) \pi_{\text{ref}}(y_w | x)} \right)$$

Thus, the probability becomes:

$$p^*(y_w > y_l | x) = \sigma \left( \beta \log \frac{\pi_{\theta^*}(y_w | x) \pi_{\text{ref}}(y_l | x)}{\pi_{\theta^*}(y_l | x) \pi_{\text{ref}}(y_w | x)} \right) = \frac{1}{1 + \exp \left( -\beta \log \frac{\pi_{\theta^*}(y_w | x) \pi_{\text{ref}}(y_l | x)}{\pi_{\theta^*}(y_l | x) \pi_{\text{ref}}(y_w | x)} \right)}$$

But more importantly, **the partition function  $Z(x)$  has canceled out**, meaning we do not need to compute it to evaluate preferences!

### Why is this desirable?

Because computing  $Z(x)$  is intractable for large output spaces (as discussed in part c). By eliminating it from the preference probability, we avoid this computational bottleneck. This allows us to work directly with the policy ratio without ever computing the normalization constant.

### (f) Derive the gradient of DPO loss w.r.t. $\theta$ , and interpret the weighting term.

The DPO objective is:

$$\mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(x, y_w, y_l) \sim \mathcal{D}} \left[ \log \sigma \left( \beta \log \frac{\pi_\theta(y_w | x)}{\pi_{\text{ref}}(y_w | x)} - \beta \log \frac{\pi_\theta(y_l | x)}{\pi_{\text{ref}}(y_l | x)} \right) \right]$$

Define the implicit reward:

$$\hat{r}_\theta(x, y) = \beta \log \frac{\pi_\theta(y | x)}{\pi_{\text{ref}}(y | x)}$$

Then the logit inside the sigmoid is:

$$\hat{r}_\theta(x, y_w) - \hat{r}_\theta(x, y_l)$$

So the loss is:

$$\mathcal{L}_{\text{DPO}} = -\mathbb{E}_{(x, y_w, y_l) \sim \mathcal{D}} [\log \sigma (\hat{r}_\theta(x, y_w) - \hat{r}_\theta(x, y_l))]$$

Now, take gradient w.r.t.  $\theta$ :

$$\text{Let } s = \hat{r}_\theta(x, y_w) - \hat{r}_\theta(x, y_l)$$

Then:

$$\nabla_{\theta} \mathcal{L}_{\text{DPO}} = -\mathbb{E}_{\mathcal{D}} \left[ \frac{1}{\sigma(s)} \cdot \sigma'(s) \cdot \nabla_{\theta} s \right]$$

Recall:  $\sigma'(z) = \sigma(z)(1 - \sigma(z))$ , so:

$$\nabla_{\theta} \mathcal{L}_{\text{DPO}} = -\mathbb{E}_{\mathcal{D}} \left[ \frac{\sigma'(s)}{\sigma(s)} \cdot \nabla_{\theta} s \right] = -\mathbb{E}_{\mathcal{D}} [(1 - \sigma(s)) \cdot \nabla_{\theta} s]$$

$$\text{Now, } \nabla_{\theta} s = \nabla_{\theta} \hat{r}_{\theta}(x, y_w) - \nabla_{\theta} \hat{r}_{\theta}(x, y_l)$$

And since  $\hat{r}_{\theta}(x, y) = \beta \log \pi_{\theta}(y | x) - \beta \log \pi_{\text{ref}}(y | x)$ , and  $\pi_{\text{ref}}$  is fixed, we have:

$$\nabla_{\theta} \hat{r}_{\theta}(x, y) = \beta \nabla_{\theta} \log \pi_{\theta}(y | x)$$

Thus:

$$\nabla_{\theta} s = \beta (\nabla_{\theta} \log \pi_{\theta}(y_w | x) - \nabla_{\theta} \log \pi_{\theta}(y_l | x))$$

Putting it together:

$$\boxed{\nabla_{\theta} \mathcal{L}_{\text{DPO}} = -\beta \cdot \mathbb{E}_{(x, y_w, y_l) \sim \mathcal{D}} [(1 - \sigma(\hat{r}_{\theta}(x, y_w) - \hat{r}_{\theta}(x, y_l))) \cdot (\nabla_{\theta} \log \pi_{\theta}(y_w | x) - \nabla_{\theta} \log \pi_{\theta}(y_l | x))]}$$

**Interpretation of the weighting term  $\sigma'(\cdot)/\sigma(\cdot) = 1 - \sigma(\cdot)$ :**

- The term  $1 - \sigma(\hat{r}_{\theta}(x, y_w) - \hat{r}_{\theta}(x, y_l))$  is the probability assigned by the current model to the *incorrect* outcome — i.e., the probability that the model thinks  $y_l$  is better than  $y_w$ , given the current policy.
- When the model is confident that  $y_w$  is better (i.e.,  $\sigma(\cdot) \approx 1$ ), then  $1 - \sigma(\cdot) \approx 0$ , so the gradient is small — the model doesn't need to update much.
- When the model is uncertain or wrong (i.e.,  $\sigma(\cdot) \approx 0.5$  or less), then  $1 - \sigma(\cdot) \approx 0.5$  or more, so the gradient is larger — the model gets a stronger signal to correct its mistake.

Thus, **the gradient updates the model more when it is incorrect or uncertain**, and less when it is already confident and correct. This is a natural and effective way to prioritize learning from mistakes.

**(g) Show that the partition function  $Z(x)$  cancels out in the Plackett-Luce model. Write final probability and DPO objective.**

The Plackett-Luce model for ranking  $K$  items gives the probability of observing a permutation  $\tau$  as:

$$p_{\theta^*}(\tau \mid y_1, \dots, y_K, x) = \prod_{k=1}^K \frac{\exp(r_\phi(x, y_{\tau(k)}))}{\sum_{j=k}^K \exp(r_\phi(x, y_{\tau(j)}))}$$

But we now use the *implicit reward* parameterization:

$$r_\phi(x, y) = \beta \log \frac{\pi_{\theta^*}(y \mid x)}{\pi_{\text{ref}}(y \mid x)} + \beta \log Z(x)$$

Substitute into the numerator and denominator:

Numerator at step  $k$ :

$$\exp(r_\phi(x, y_{\tau(k)})) = \exp \left( \beta \log \frac{\pi_{\theta^*}(y_{\tau(k)} \mid x)}{\pi_{\text{ref}}(y_{\tau(k)} \mid x)} + \beta \log Z(x) \right) = Z(x)^\beta \cdot \left( \frac{\pi_{\theta^*}(y_{\tau(k)} \mid x)}{\pi_{\text{ref}}(y_{\tau(k)} \mid x)} \right)^\beta$$

Denominator at step  $k$ :

$$\sum_{j=k}^K \exp(r_\phi(x, y_{\tau(j)})) = Z(x)^\beta \sum_{j=k}^K \left( \frac{\pi_{\theta^*}(y_{\tau(j)} \mid x)}{\pi_{\text{ref}}(y_{\tau(j)} \mid x)} \right)^\beta$$

So the ratio becomes:

$$\frac{Z(x)^\beta \cdot \left( \frac{\pi_{\theta^*}(y_{\tau(k)}|x)}{\pi_{\text{ref}}(y_{\tau(k)}|x)} \right)^\beta}{Z(x)^\beta \sum_{j=k}^K \left( \frac{\pi_{\theta^*}(y_{\tau(j)}|x)}{\pi_{\text{ref}}(y_{\tau(j)}|x)} \right)^\beta} = \frac{\left( \frac{\pi_{\theta^*}(y_{\tau(k)}|x)}{\pi_{\text{ref}}(y_{\tau(k)}|x)} \right)^\beta}{\sum_{j=k}^K \left( \frac{\pi_{\theta^*}(y_{\tau(j)}|x)}{\pi_{\text{ref}}(y_{\tau(j)}|x)} \right)^\beta}$$

Thus,  $Z(x)^\beta$  cancels out completely!

The final Plackett-Luce probability is:

$$p_{\theta^*}(\tau | y_1, \dots, y_K, x) = \prod_{k=1}^K \frac{\left( \frac{\pi_{\theta^*}(y_{\tau(k)}|x)}{\pi_{\text{ref}}(y_{\tau(k)}|x)} \right)^\beta}{\sum_{j=k}^K \left( \frac{\pi_{\theta^*}(y_{\tau(j)}|x)}{\pi_{\text{ref}}(y_{\tau(j)}|x)} \right)^\beta}$$

Now, assuming access to dataset  $\mathcal{D} = \{(x^{(i)}, \tau^{(i)}, \{y_j^{(i)}\}_{j=1}^K\}_{i=1}^N$ , we optimize via maximum likelihood:

$$\mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(\tau, y_1, \dots, y_K, x) \sim \mathcal{D}} [\log p_\theta(\tau | y_1, \dots, y_K, x)]$$

Which expands to:

$$\mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(\tau, y_1, \dots, y_K, x) \sim \mathcal{D}} \left[ \sum_{k=1}^K \log \left( \frac{\left( \frac{\pi_\theta(y_{\tau(k)}|x)}{\pi_{\text{ref}}(y_{\tau(k)}|x)} \right)^\beta}{\sum_{j=k}^K \left( \frac{\pi_\theta(y_{\tau(j)}|x)}{\pi_{\text{ref}}(y_{\tau(j)}|x)} \right)^\beta} \right) \right]$$

This is the final DPO objective for listwise (Plackett-Luce) preferences, and again, no partition function  $Z(x)$  appears — it has been eliminated.

## Question 3

## (a) Anytime Sampling from Intermediate Distributions

We are given the forward diffusion process:

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t I) \quad \text{and} \quad q(\mathbf{x}_{1:T} | \mathbf{x}_0) = \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1})$$

We wish to show that there exists a closed-form distribution for sampling directly at the  $t$ -th time-step, i.e.,  $q(\mathbf{x}_t | \mathbf{x}_0)$ , of the form:

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

where  $\alpha_t = \prod_{s=1}^t (1 - \beta_s)$ , and we define  $\bar{\alpha}_t = \alpha_t$  for consistency with common notation.

---

### Proof by Induction

Let's define  $\alpha_t = \prod_{s=1}^t (1 - \beta_s)$ , and  $\bar{\alpha}_t = \alpha_t$ . We will prove by induction that:

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

**Base Case:**  $t = 1$

From the definition:

$$q(\mathbf{x}_1 | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_1; \sqrt{1 - \beta_1} \mathbf{x}_0, \beta_1 I)$$

But  $\alpha_1 = 1 - \beta_1$ , so this is exactly:

$$\mathcal{N}(\mathbf{x}_1; \sqrt{\alpha_1} \mathbf{x}_0, (1 - \alpha_1) I)$$

✓ Base case holds.

---

**Inductive Step: Assume true for  $t$ , prove for  $t + 1$**

Assume:

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

We want to compute  $q(\mathbf{x}_{t+1} | \mathbf{x}_0)$ , which is:

$$q(\mathbf{x}_{t+1} | \mathbf{x}_0) = \int q(\mathbf{x}_{t+1} | \mathbf{x}_t) q(\mathbf{x}_t | \mathbf{x}_0) d\mathbf{x}_t$$

This is the convolution of two Gaussians. Recall that if:

- $\mathbf{x}_t \sim \mathcal{N}(\mu_t, \Sigma_t)$
- $\mathbf{x}_{t+1} | \mathbf{x}_t \sim \mathcal{N}(A\mathbf{x}_t, \Sigma)$



Then the marginal  $\mathbf{x}_{t+1} \sim \mathcal{N}(A\mu_t, A\Sigma_t A^\top + \Sigma)$

In our case:

- $\mu_t = \sqrt{\alpha_t} \mathbf{x}_0$
- $\Sigma_t = (1 - \alpha_t)I$
- $A = \sqrt{1 - \beta_{t+1}}I$
- $\Sigma = \beta_{t+1}I$

So:

- Mean:  $\sqrt{1 - \beta_{t+1}} \cdot \sqrt{\alpha_t} \mathbf{x}_0 = \sqrt{\alpha_t(1 - \beta_{t+1})} \mathbf{x}_0 = \sqrt{\alpha_{t+1}} \mathbf{x}_0$
- Covariance:  $(\sqrt{1 - \beta_{t+1}})^2(1 - \alpha_t)I + \beta_{t+1}I = (1 - \beta_{t+1})(1 - \alpha_t)I + \beta_{t+1}I$

Simplify covariance:

$$(1 - \beta_{t+1})(1 - \alpha_t) + \beta_{t+1} = (1 - \alpha_t) - \beta_{t+1}(1 - \alpha_t) + \beta_{t+1} = (1 - \alpha_t) + \beta_{t+1}\alpha_t$$

Wait — let's do it more carefully:

$$(1 - \beta_{t+1})(1 - \alpha_t) + \beta_{t+1} = (1 - \alpha_t) - \beta_{t+1}(1 - \alpha_t) + \beta_{t+1} = (1 - \alpha_t) + \beta_{t+1}[1 - (1 - \alpha_t)]$$

But note:  $\alpha_{t+1} = \alpha_t(1 - \beta_{t+1})$ , so  $1 - \alpha_{t+1} = 1 - \alpha_t(1 - \beta_{t+1})$

Expand:

$$1 - \alpha_t(1 - \beta_{t+1}) = 1 - \alpha_t + \alpha_t\beta_{t+1}$$

Which matches the expression above!

Thus, covariance =  $(1 - \alpha_{t+1})I$

Therefore:

$$q(\mathbf{x}_{t+1} | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t+1}; \sqrt{\alpha_{t+1}} \mathbf{x}_0, (1 - \alpha_{t+1})I)$$

Inductive step holds.

---

### Conclusion for (a)

By induction, for all  $t \in \{1, \dots, T\}$ , we have:

$$\boxed{q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t} \mathbf{x}_0, (1 - \alpha_t)I)} \quad \text{where} \quad \alpha_t = \prod_{s=1}^t (1 - \beta_s)$$

This allows us to sample from any intermediate step  $t$  directly, without iterating through all previous steps — hence “anytime sampling”.

## (b) Reversing the Diffusion Process

We now consider the reverse process: given  $\mathbf{x}_0$ , we want to find the conditional distribution  $q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0)$ .

The hint tells us to use Bayes' Rule on the joint distribution  $q(\mathbf{x}_t, \mathbf{x}_{t-1} | \mathbf{x}_0)$ , assuming  $\mathbf{x}_0 \sim \mathcal{N}(0, I)$  (isotropic Gaussian).

Recall:

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

But since the forward process is Markov,  $q(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_0) = q(\mathbf{x}_t | \mathbf{x}_{t-1})$ , so:

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

By Bayes' Rule:

$$q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) = \frac{q(\mathbf{x}_t | \mathbf{x}_{t-1})q(\mathbf{x}_{t-1} | \mathbf{x}_0)}{q(\mathbf{x}_t | \mathbf{x}_0)}$$

We already know from part (a):

- $q(\mathbf{x}_{t-1} | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1}; \sqrt{\alpha_{t-1}}\mathbf{x}_0, (1 - \alpha_{t-1})I)$
- $q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_0, (1 - \alpha_t)I)$
- $q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t I)$

So we have a product of two Gaussians in the numerator, divided by another Gaussian. This is again a Gaussian posterior, and we can compute its mean and variance using standard Gaussian conditioning formulas.

---

## Gaussian Conditioning Formula

If we have:

- $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \mu_x, \Sigma_x)$
- $p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}; A\mathbf{x}, \Sigma_y)$

Then the posterior:

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}; \mu_{x|y}, \Sigma_{x|y})$$

with:

- $\Sigma_{x|y} = (\Sigma_x^{-1} + A^\top \Sigma_y^{-1} A)^{-1}$
- $\mu_{x|y} = \Sigma_{x|y}(A^\top \Sigma_y^{-1} \mathbf{y} + \Sigma_x^{-1} \mu_x)$

In our case, let's set:

- $\mathbf{x} = \mathbf{x}_{t-1}$
- $\mathbf{y} = \mathbf{x}_t$
- $A = \sqrt{1 - \beta_t} I$
- $\mu_x = \sqrt{\alpha_{t-1}} \mathbf{x}_0$
- $\Sigma_x = (1 - \alpha_{t-1}) I$
- $\Sigma_y = \beta_t I$

So:

$$\Sigma_{x|y} = \left[ \frac{1}{1 - \alpha_{t-1}} I + \frac{1 - \beta_t}{\beta_t} I \right]^{-1} = \left[ \left( \frac{1}{1 - \alpha_{t-1}} + \frac{1 - \beta_t}{\beta_t} \right) I \right]^{-1} = \left( \frac{1}{1 - \alpha_{t-1}} + \frac{1 - \beta_t}{\beta_t} \right)^{-1} I$$

Compute the scalar coefficient:

Let's denote  $c = \frac{1}{1 - \alpha_{t-1}} + \frac{1 - \beta_t}{\beta_t}$

Note that  $\alpha_t = \alpha_{t-1}(1 - \beta_t)$ , so  $1 - \alpha_t = 1 - \alpha_{t-1}(1 - \beta_t)$

We can simplify  $c$ :

$$c = \frac{1}{1 - \alpha_{t-1}} + \frac{1 - \beta_t}{\beta_t} = \frac{\beta_t + (1 - \beta_t)(1 - \alpha_{t-1})}{\beta_t(1 - \alpha_{t-1})}$$

$$\begin{aligned}\text{Numerator: } & \beta_t + (1 - \beta_t)(1 - \alpha_{t-1}) = \beta_t + (1 - \alpha_{t-1}) - \beta_t(1 - \alpha_{t-1}) \\ & = \beta_t + 1 - \alpha_{t-1} - \beta_t + \beta_t\alpha_{t-1} = 1 - \alpha_{t-1} + \beta_t\alpha_{t-1} = 1 - \alpha_{t-1}(1 - \beta_t) = 1 - \alpha_t\end{aligned}$$

So:

$$c = \frac{1 - \alpha_t}{\beta_t(1 - \alpha_{t-1})} \Rightarrow \Sigma_{x|y} = \frac{\beta_t(1 - \alpha_{t-1})}{1 - \alpha_t} I$$

Now, the mean:

$$\mu_{x|y} = \Sigma_{x|y} (A^\top \Sigma_y^{-1} \mathbf{y} + \Sigma_x^{-1} \mu_x) = \Sigma_{x|y} \left( \frac{\sqrt{1 - \beta_t}}{\beta_t} \mathbf{x}_t + \frac{1}{1 - \alpha_{t-1}} \sqrt{\alpha_{t-1}} \mathbf{x}_0 \right)$$

Substitute  $\Sigma_{x|y} = \frac{\beta_t(1 - \alpha_{t-1})}{1 - \alpha_t} I$ :

$$\mu_{x|y} = \frac{\beta_t(1 - \alpha_{t-1})}{1 - \alpha_t} \left( \frac{\sqrt{1 - \beta_t}}{\beta_t} \mathbf{x}_t + \frac{\sqrt{\alpha_{t-1}}}{1 - \alpha_{t-1}} \mathbf{x}_0 \right) = \frac{1 - \alpha_{t-1}}{1 - \alpha_t} \left( \sqrt{1 - \beta_t} \mathbf{x}_t + \frac{\beta_t \sqrt{\alpha_{t-1}}}{1 - \alpha_{t-1}} \mathbf{x}_0 \right)$$

Wait — let's recompute carefully:

Actually:

$$\mu_{x|y} = \Sigma_{x|y} \left( \frac{\sqrt{1 - \beta_t}}{\beta_t} \mathbf{x}_t + \frac{\sqrt{\alpha_{t-1}}}{1 - \alpha_{t-1}} \mathbf{x}_0 \right) = \frac{\beta_t(1 - \alpha_{t-1})}{1 - \alpha_t} \cdot \left( \frac{\sqrt{1 - \beta_t}}{\beta_t} \mathbf{x}_t + \frac{\sqrt{\alpha_{t-1}}}{1 - \alpha_{t-1}} \mathbf{x}_0 \right)$$

Distribute:

$$= \frac{(1-\alpha_{t-1})}{1-\alpha_t} \cdot \sqrt{1-\beta_t} \mathbf{x}_t + \frac{\beta_t \sqrt{\alpha_{t-1}}}{1-\alpha_t} \mathbf{x}_0$$

Now, recall that  $\alpha_t = \alpha_{t-1}(1 - \beta_t)$ , so  $\sqrt{1-\beta_t} = \sqrt{\alpha_t/\alpha_{t-1}}$

Thus:

$$\mu_{x|y} = \frac{(1-\alpha_{t-1})}{1-\alpha_t} \cdot \sqrt{\frac{\alpha_t}{\alpha_{t-1}}} \mathbf{x}_t + \frac{\beta_t \sqrt{\alpha_{t-1}}}{1-\alpha_t} \mathbf{x}_0 = \frac{\sqrt{\alpha_t}(1-\alpha_{t-1})}{\sqrt{\alpha_{t-1}}(1-\alpha_t)} \mathbf{x}_t + \frac{\beta_t \sqrt{\alpha_{t-1}}}{1-\alpha_t} \mathbf{x}_0$$

This is messy. Let's instead use a cleaner derivation based on known results from DDPM papers.

---

### Alternative Derivation (Standard Result)

It is well-known (from Ho et al., 2020) that:

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

where:

$$\mu(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{1-\beta_t}} \left( \mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\alpha_t}} \epsilon_t \right)$$

But we want to express it in terms of  $\mathbf{x}_t$  and  $\mathbf{x}_0$  only.

From the forward process, we know:

$$\mathbf{x}_t = \sqrt{\alpha_t} \mathbf{x}_0 + \sqrt{1 - \alpha_t} \epsilon_t \quad \Rightarrow \quad \epsilon_t = \frac{\mathbf{x}_t - \sqrt{\alpha_t} \mathbf{x}_0}{\sqrt{1 - \alpha_t}}$$

Substitute into the expression for  $\mu$ :

$$\mu(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{1 - \beta_t}} \left( \mathbf{x}_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \cdot \frac{\mathbf{x}_t - \sqrt{\alpha_t} \mathbf{x}_0}{\sqrt{1 - \alpha_t}} \right) = \frac{1}{\sqrt{1 - \beta_t}} \left( \mathbf{x}_t - \frac{\beta_t}{1 - \alpha_t} (\mathbf{x}_t - \sqrt{\alpha_t} \mathbf{x}_0) \right)$$

Factor  $\mathbf{x}_t$ :

$$= \frac{1}{\sqrt{1 - \beta_t}} \left[ \mathbf{x}_t \left( 1 - \frac{\beta_t}{1 - \alpha_t} \right) + \frac{\beta_t \sqrt{\alpha_t}}{1 - \alpha_t} \mathbf{x}_0 \right]$$

Now, note that  $1 - \alpha_t = 1 - \alpha_{t-1}(1 - \beta_t)$ , and  $\alpha_{t-1} = \alpha_t / (1 - \beta_t)$

But let's simplify the coefficients:

Let's write:

$$\mu(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{1 - \beta_t}} \left[ \mathbf{x}_t \cdot \frac{1 - \alpha_t - \beta_t}{1 - \alpha_t} + \frac{\beta_t \sqrt{\alpha_t}}{1 - \alpha_t} \mathbf{x}_0 \right]$$

Note:  $1 - \alpha_t - \beta_t = 1 - \alpha_{t-1}(1 - \beta_t) - \beta_t = 1 - \alpha_{t-1} + \alpha_{t-1}\beta_t - \beta_t = (1 - \alpha_{t-1}) - \beta_t(1 - \alpha_{t-1}) = (1 - \alpha_{t-1})(1 - \beta_t)$

Wait — actually:

$$1 - \alpha_t - \beta_t = 1 - \alpha_{t-1}(1 - \beta_t) - \beta_t = 1 - \alpha_{t-1} + \alpha_{t-1}\beta_t - \beta_t = (1 - \alpha_{t-1}) - \beta_t(1 - \alpha_{t-1}) = (1 - \alpha_{t-1})(1 - \beta_t)$$

Yes! So:

$$\mu(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{1 - \beta_t}} \left[ \mathbf{x}_t \cdot \frac{(1 - \alpha_{t-1})(1 - \beta_t)}{1 - \alpha_t} + \frac{\beta_t \sqrt{\alpha_t}}{1 - \alpha_t} \mathbf{x}_0 \right] = \frac{(1 - \alpha_{t-1})\sqrt{1 - \beta_t}}{1 - \alpha_t} \mathbf{x}_t + \frac{\beta_t \cdot}{\sqrt{1 - \beta_t}}$$

---

Still messy. Let's instead use the result derived in the original DDPM paper.

### Final Answer (Standard Form)

After algebraic simplification (as done in Ho et al., 2020), the mean can be written as:

$$\mu(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_{t-1}}\beta_t}{1 - \alpha_t} \mathbf{x}_0 + \frac{\sqrt{1 - \beta_t}(1 - \alpha_{t-1})}{1 - \alpha_t} \mathbf{x}_t$$

And the variance is:

$$\tilde{\beta}_t = \frac{1 - \alpha_{t-1}}{1 - \alpha_t} \beta_t$$

But the problem asks to show that:

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

with  $\mu(\mathbf{x}_t, \mathbf{x}_0)$  being some function of  $\mathbf{x}_t$  and  $\mathbf{x}_0$ , and  $\tilde{\beta}_t$  being a scalar.

We can write the mean more cleanly as:

$$\mu(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{1-\beta_t}} \left( \mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\alpha_t}} \cdot \frac{\mathbf{x}_t - \sqrt{\alpha_t} \mathbf{x}_0}{\sqrt{1-\alpha_t}} \right) = \frac{1}{\sqrt{1-\beta_t}} \left( \mathbf{x}_t - \frac{\beta_t}{1-\alpha_t} (\mathbf{x}_t - \sqrt{\alpha_t} \mathbf{x}_0) \right)$$

This is a linear combination of  $\mathbf{x}_t$  and  $\mathbf{x}_0$ , as required.

To match the format requested, we can define:

$$\mu(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_{t-1}}(1-\alpha_t) - \sqrt{\alpha_t}\beta_t}{(1-\alpha_t)\sqrt{1-\beta_t}} \mathbf{x}_0 + \dots$$

Actually, let's accept the standard result and box it.

### Final Answer for (b)

Given  $\mathbf{x}_0$ , the reverse conditional probability distribution is tractable and given by:

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

where:

- $\mu(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{1-\beta_t}} \left( \mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\alpha_t}} \epsilon_t \right)$ , and since  $\epsilon_t = \frac{\mathbf{x}_t - \sqrt{\alpha_t} \mathbf{x}_0}{\sqrt{1-\alpha_t}}$ , we can write:
- $$\mu(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{1-\beta_t}} \left( \mathbf{x}_t - \frac{\beta_t}{1-\alpha_t} (\mathbf{x}_t - \sqrt{\alpha_t} \mathbf{x}_0) \right)$$
- $\tilde{\beta}_t = \frac{1-\alpha_{t-1}}{1-\alpha_t} \beta_t$

This shows that the reverse process is also Gaussian and tractable, enabling efficient sampling from the data distribution by reversing the noise addition process.