# Deriving the Reverse Process (Inference) Equations

So we know how to corrupt an image

$$q(x_t \mid x_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\, x_{t-1}, (1-\bar{\alpha}_t) I) \qquad \begin{cases} \mu = \bar{\alpha}\, x_{t-1} \\ \sigma^2 = (1-\bar{\alpha}_t) I \end{cases}$$

$\Rightarrow$ to get $x_t$ all we need to do is

$$x_t = \sqrt{\bar{\alpha}_t}\, x_{t-1} + \sqrt{1-\bar{\alpha}_t}\, \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, I)$$

But to actually generate an image wnt go from pure noise to our image: $x_T \longrightarrow x_0$

This is the whole point of SDXL, the inference where we generate an image

Steps:

1) Show Marginal forward posterior intractable
2) Condition on $x_0$: Exact forward posterior
3) we don't have $x_0$: $\rightarrow \hat{x}_0$
4) use $\hat{x}_0$ to get mean
5) Get $\hat{\beta}_t$ for var

1) Show Marginal forward posterior intractable

To perform inference, we need the posterior distribution

$$q(x_{t-1} \mid x_t) = \frac{q(x_{t-1}, x_t)}{q(x_t)}$$

$$= \frac{q(x_t \mid x_{t-1})\, q(x_{t-1})}{\cancel{q(x_t)}}$$

$$= q(x_t \mid x_{t-1})\, q(x_{t-1})$$

we know $q(x_t \mid x_{t-1})$ and $q(x_{t-1})$ are both Gaussian by modeling assumption $\therefore$ the normalization is irrelevant by $\mathcal{N}(a,b) \times \mathcal{N}(a,b) = \mathcal{N}(g,f)$

So now, what is $q(x_{t-1})$?

$$q(x_{t-1}) = \int q(x_{t-1}, x_0)\, dx_0 = \int p_{Data}(x_0)\, q(x_{t-1} \mid x_0)\, dx_0$$

Technically, we have closed form

So $q(x_{t-1})$ is a marginal distribution and an unconditional distribution where we must average over all possible $x_0$'s in order to infer the next step in the reverse process.

Problem: $P_{data}(x_0)$ is unknown

$\therefore q(x_{t-1}|x_t)$ intractable

2) Condition on $x_0$ : Exact forward posterior

$$q(x_{t-1}|x_t, x_0) \propto q(x_t|x_{t-1})\, q(x_{t-1}|x_0)$$

$$= N(\sqrt{\alpha_t}\, x_{t-1}, \beta_t I) \times N(\sqrt{\bar{\alpha}_{t-1}}\, x_0, (1-\bar{\alpha}_{t-1}) I)$$

we already know from forward process

How $q(x_{t-1}|x_t, x_0) \propto q(x_t|x_{t-1}, x_0)\, q(x_{t-1}|x_0)$

$$= q(x_t|x_{t-1})\, q(x_{t-1}|x_0)$$

$$q(x_{t-1}|x_t, x_0) = \frac{q(x_{t-1}, x_t, x_0)}{q(x_t, x_0)}$$ Bayes Rule

by Markov assumption

conditional probabilities

$$= \frac{q(x_t|x_{t-1}, \cancel{x_0})\, q(x_{t-1}|x_0)\, \cancel{q(x_0)}}{q(x_t|x_0)\, \cancel{q(x_0)}}$$

$$= \frac{\underbrace{q(x_t|x_{t-1})}_{\text{known Gaussian}}\, \underbrace{q(x_{t-1}|x_0)}_{\text{known Gaussian}}}{\cancel{q(x_t|x_0)}}$$

drop because posterior form known to be gaussian

$$\propto q(x_t|x_{t-1})\, q(x_{t-1}|x_0)$$

$$\boxed{\begin{aligned} q(x_{t-1}|x_t, x_0) &\propto q(x_t|x_{t-1})\, q(x_{t-1}|x_0) \\ &= N(\sqrt{\alpha_t}\, x_{t-1}, \beta_t I) \times N(\sqrt{\bar{\alpha}_{t-1}}\, x_0, (1-\bar{\alpha}_{t-1}) I) \end{aligned}}$$

So now to get the parameters of $q(x_{t-1} | x_t, x_0)$ we will multiply the log densities of the two Gaussians and simplify to find $\{\mu, \sigma^2\}$

why log?

Since log is monatic and additive for products the quadratic form in $x$ is identical whether you start from the product or go straight to logs

$$\boxed{\begin{array}{l}\text{log pdf Multivariate Gaussian} \\ \log \mathcal{N}(x; m, \Sigma) = \\ -\frac{1}{2}\left[d \log(2\pi) + \log|\Sigma| + (x-m)^T \Sigma^{-1}(x-m)\right]\end{array}}$$

(A)

Drop constants!

$$\mathcal{N}(\sqrt{\alpha_t}\, x_{t-1}, \beta_t I) \rightarrow \Sigma^{-1} = \frac{1}{\beta_t} I$$

$$= \frac{1}{2}(x_t - \sqrt{\alpha_t}\, x_{t-1})^T \frac{1}{\beta_t} I (x_{t-1} - \sqrt{\alpha_t}\, x_{t-1}) = \boxed{\frac{1}{2\beta_t} \|x_t - \sqrt{\alpha_t}\, x_{t-1}\|^2}$$

(B)

$$\mathcal{N}(\sqrt{\bar{\alpha}_{t-1}}\, x_0, (1-\bar{\alpha}_{t-1}) I) = \frac{1}{2}(x_{t-1} - \sqrt{\bar{\alpha}_{t-1}}\, x_0)^T \frac{1}{(1-\bar{\alpha}_{t-1})} I^{-1} (x_{t-1} - \sqrt{\bar{\alpha}_{t-1}}\, x_0)$$

$$= \boxed{\frac{1}{2(1-\bar{\alpha}_{t-1})} \|x_{t-1} - \sqrt{\bar{\alpha}_{t-1}}\, x_0\|^2}$$

$$\lg A + \lg B = \frac{1}{2\beta_t}\|x_t - \sqrt{\alpha_t}\, x_{t-1}\|^2 + \frac{1}{2(1-\bar\alpha_{t-1})}\|x_{t-1} - \sqrt{\bar\alpha_{t-1}}\, x_0\|^2$$

$$= \frac{1}{2\beta_t}\left(x_t^T x_t - 2\sqrt{\alpha_t}\, x_t^T x_{t-1} + \alpha_t x_{t-1}^T x_{t-1}\right)$$

$$+ \frac{1}{2(1-\bar\alpha_{t-1})}\left(x_{t-1}^T x_{t-1} - 2\sqrt{\bar\alpha_{t-1}}\, x_{t-1}^T x_0 + \bar\alpha_{t-1} x_0^T x_0\right)$$

$(xy)^T = y^T x^T$

$(a+b)^T = a^T + b^T$

$M(c+d) = Mc + Md$

$M$ symmetric $\rightarrow M^T = M$

$a, b$ scalars

$a^T M b = (a^T M b)^T = b^T M^T a = b^T M a$

To get the form of the new Gaussian we collect terms and complete the square

...

$$\text{covariance: } \frac{\beta_t(1-\bar\alpha_{t-1})}{1-\bar\alpha_t} I \equiv \tilde\beta_t I \qquad (a.1)$$

$$\text{mean } \mu_{post}(x_t, x_0) = \frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})}{1-\bar\alpha_t} x_t + \frac{\sqrt{\bar\alpha_{t-1}}\,\beta_t}{1-\bar\alpha_t} x_0$$

using $x_t = \sqrt{\bar\alpha_t}\, x_0 + \sqrt{1-\bar\alpha_t}\, \epsilon$, we can eliminate $x_0$ and show

$$\boxed{\mu_{post}(x_t, \epsilon) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar\alpha_t}}\epsilon\right)} \qquad (a.1)$$

replace with $\epsilon_\theta$

we need a predictor for $x_0$ called $\hat{x}_0$

$x_t = \sqrt{\bar\alpha_t}\, x_0 + \sqrt{1-\bar\alpha_t}\, \epsilon$

$\Rightarrow$ solve for $x_0$

$$x_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon\right)$$

Swap $\epsilon$ for $\epsilon_\theta$, the models prediction

$$\Rightarrow x_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon(x_t, t)\right)$$

plug $\hat{x}_0$ into (Q.1) and simplify

$$\Rightarrow \mu_t(x_t, \hat{x}_0) = \frac{1}{\sqrt{\alpha_t}}\left[x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right] \quad \text{(Im.1)}$$

Now we have mean $\mu_t(x_t)$ and var $\tilde{\beta}_t$ to get $x_{t-1}$

$$x_{t-1} = \mu_t(x_t, \hat{x}_0) + \sigma_t z \qquad z \sim N(0, I)$$

$$\sigma_t^2 = \tilde{\beta}_t$$