$$
\begin{tikzcd}
x_T \arrow[r] & \cdots & x_t \arrow[r] \arrow[r, bend left, "p_\theta(x_{t-1}|x_t)"] & x_{t-1} \arrow[r] \arrow[l, bend left, "q(x_t|x_{t-1})"] & x_0
\end{tikzcd}
$$

pure noise $\qquad$ $q(x_t|x_{t-1})$ $\qquad$ original image

- Forward Pass

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t ; \underbrace{\sqrt{1-\beta_t}\, x_{t-1}}_{\text{mean}}, \underbrace{\beta_t \cdot I}_{\text{variance}})$$

$$= \sqrt{1-\beta_t} \cdot x_{t-1} + \sqrt{\beta_t} \cdot \epsilon \qquad \epsilon \sim \mathcal{N}(0, I)$$

variance $\quad 0 < \beta_1 < \beta_2 \cdots < \beta_T < 1$

Claim: let $\alpha_t = 1-\beta_t \quad \bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$.

then $x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon \qquad \epsilon \sim \mathcal{N}(0, I)$

thus $q(x_t|x_{t-1}) \sim \mathcal{N}(\sqrt{\bar{\alpha}_t}\, x_{t-1}, (1-\bar{\alpha}_t) I)$ this approaches to $\mathcal{N}(0, I)$ as $t \to +\infty$

proof.

1. $q(x_1|x_0) = \sqrt{1-\beta_1}\, x_0 + \sqrt{\beta_1}\, \epsilon = \sqrt{\bar{\alpha}_1}\, x_0 + \sqrt{1-\bar{\alpha}_1}\, \epsilon$

2. if $x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon$

then $x_{t+1} = \sqrt{1-\beta_{t+1}}\, x_t + \sqrt{\beta_{t+1}}\, \hat{\epsilon}$

$$= \sqrt{1-\beta_{t+1}} \left[\sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon \right] + \sqrt{\beta_{t+1}}\, \hat{\epsilon}$$

$$= \sqrt{\bar{\alpha}_{t+1}}\, x_0 + \sqrt{1-\beta_{t+1}} \sqrt{1-\bar{\alpha}_t}\, \epsilon + \sqrt{\beta_{t+1}}\, \hat{\epsilon}$$

$$= \sqrt{\bar{\alpha}_{t+1}}\, x_0 + \sqrt{(1-\beta_{t+1})(1-\bar{\alpha}_t) + \beta_{t+1}} \cdot \epsilon$$

$$= \sqrt{\bar{\alpha}_{t+1}}\, x_0 + \sqrt{1-\bar{\alpha}_{t+1}} \cdot \epsilon$$

convolution of 2 Gaussian

$\mathcal{N}(\mu_1, \sigma_1^2) * \mathcal{N}(\mu_2, \sigma_2^2) = \mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$

Conditional Probability

Claim: $q(x_{t-1}|x_t, x_0) \sim \mathcal{N}\left(\tilde{\mu}_{t-1}(x_t, x_0), \tilde{\Sigma}_{t-1}\right)$

where $\tilde{\mu}_{t-1}(x_t, x_0) = \frac{(1-\bar{\alpha}_{t-1})\beta_t}{1-\bar{\alpha}_t}\left[\frac{\sqrt{\alpha_t}}{\beta_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0\right] = \frac{1}{\sqrt{\alpha_t}}x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\cdot\sqrt{\alpha_t}}\cdot\epsilon$

$\tilde{\Sigma}_{t-1} = \frac{\beta_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} = \frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}$

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

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

$$\propto \exp\left[-\frac{\|x_t-\sqrt{\alpha_t}x_{t-1}\|_2^2}{2\beta_t} - \frac{\|x_{t-1}-\sqrt{\bar{\alpha}_{t-1}}x_0\|_2^2}{2(1-\bar{\alpha}_{t-1})} + \frac{\|x_t-\sqrt{\bar{\alpha}_t}x_0\|_2^2}{2(1-\bar{\alpha}_t)}\right]$$

$$\propto \exp\left[-\frac{\|x_t\|_2^2 + \alpha_t\|x_{t-1}\|_2^2 - 2\sqrt{\alpha_t}\langle x_t, x_{t-1}\rangle}{2\beta_t} - \frac{\|x_{t-1}\|_2^2 + \bar{\alpha}_{t-1}\|x_0\|_2^2 - 2\sqrt{\bar{\alpha}_{t-1}}\langle x_{t-1}, x_0\rangle}{2(1-\bar{\alpha}_{t-1})} + C(x_0, x_t)\right]$$

$$\propto \exp\left[-\frac{1}{2}\underbrace{\left(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right)}_{a}\|x_{t-1}\|_2^2 + \left\langle x_{t-1}, \underbrace{\frac{\sqrt{\alpha_t}}{\beta_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{(1-\bar{\alpha}_{t-1})}x_0}_{b}\right\rangle\right]$$

$$-\frac{1}{2}a x^Tx + x^Tb = -\frac{1}{2\frac{1}{a}}\left(x^Tx - \frac{2x^Tb}{a}\right) \Rightarrow -\frac{1}{2\frac{1}{a}}\left\|x-\frac{b}{a}\right\|_2^2$$

$$\tilde{\Sigma}_{t-1} = \frac{1}{a}I = \frac{\beta_t\cdot(1-\bar{\alpha}_{t-1})}{\alpha_t(1-\bar{\alpha}_{t-1})+\beta_t}\cdot I = \frac{\beta_t\cdot(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}$$

$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$

$x_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left[x_t - \sqrt{1-\bar{\alpha}_t}\epsilon\right]$

$$\tilde{\mu}_{t-1}(x_t, x_0) = \frac{b}{a} = \frac{\beta_t\cdot(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\cdot\left(\frac{\sqrt{\alpha_t}}{\beta_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0\right)$$

$$= \frac{\beta_t\cdot(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\cdot\left[\frac{\sqrt{\alpha_t}}{\beta_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \sqrt{1-\bar{\alpha}_t}\right)\epsilon\right]$$

$$= \frac{(1-\bar{\alpha}_{t-1})\cdot\sqrt{\alpha_t}}{1-\bar{\alpha}_t}\cdot x_t + \frac{\beta_t\cdot\sqrt{\bar{\alpha}_{t-1}}}{(1-\bar{\alpha}_t)\sqrt{\bar{\alpha}_t}}x_t - \frac{\beta_t\cdot\sqrt{\bar{\alpha}_{t-1}}\sqrt{1-\bar{\alpha}_t}}{(1-\bar{\alpha}_t)\sqrt{\bar{\alpha}_t}}\epsilon$$

$$= \frac{(1-\bar{\alpha}_{t-1})\cdot\alpha_t + \beta_t}{(1-\bar{\alpha}_t)\cdot\sqrt{\alpha_t}}x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}\cdot\sqrt{\alpha_t}}\cdot\epsilon$$

$$= \frac{1}{\sqrt{\alpha_t}}x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\cdot\sqrt{\alpha_t}}\cdot\epsilon$$

- Evidence Lower Bound

$$\log p_\theta(x[0]) = \log \int_{x[1:T]} p_\theta\left(x[0], x[1:T]\right) dx[1:T]$$

$$= \log \int_{x[1:T]} q(x[1:T]|x[0]) \cdot \frac{p_\theta\left(x[0], x[1:T]\right)}{q\left(x[1:T]\,|\,x[0]\right)} \, dx[1:T]$$

$$\geq E_{x[1:T]\sim q}\left[\log \frac{p_\theta\left(x[0], x[1:T]\right)}{q\left(x[1:T]\,|\,x[0]\right)} \right]$$

$$= E_{x[1:T]\sim q}\left[\log \frac{p_\theta(x[T]) \cdot \prod_{t=1}^{T} p_\theta(x[t-1]|x[t]) \;\Big\}\ \text{backward pass}}{\prod_{t=1}^{T} q(x[t]|x[t-1]) \;\Big\}\ \text{forward pass}} \right]$$

$$= E_{x[1:T]\sim q}\left[\log p_\theta(x[T]) + \sum_{t=1}^{T} \log \frac{p_\theta(x[t-1]|x[t])}{q(x[t]\,|\,x[t-1])} \right]$$

$$= E_{x[1:T]\sim q}\left[\log p_\theta(x[T]) + \sum_{t=2}^{T} \log \frac{p_\theta(x[t-1]|x[t])}{q(x[t]|x[t-1])} + \log \frac{p_\theta(x[0]|x[1])}{q(x[1]\,|\,x[0])} \right]$$

$$q\left(x[t]\,|\,x[t-1]\right) = q\left(x[t]\,|\,x[t-1], x[0]\right) = \frac{q(x[t-1]|x[t],x[0]) \cdot q(x[t]|x[0])}{q(x[t-1]|x[0])}$$

$$= E_{x[1:T]\sim q}\left[\log p_\theta(x[T]) + \sum_{t=2}^{T} \log \frac{p_\theta(x[t-1]|x[t])}{q(x[t-1]\,|\,x[t],x[0])} + \sum_{t=2}^{T} \log \frac{q(x[t-1]|x[0])}{q(x[t]\,|\,x[0])} + \log \frac{p_\theta(x[0]|x[1])}{q(x[1]\,|\,x[0])} \right]$$

$$= E_{x[1:T]\sim q}\left[\log \frac{p_\theta(x[T])}{q(x[T]|x[0])} + \sum_{t=2}^{T} \log \frac{p_\theta(x[t-1]|x[t])}{q(x[t-1]\,|\,x[t],x[0])} + \log p_\theta(x[0]|x[1]) \right]$$

$$= E_{x[1:T]\sim q}\left[\log \frac{p_\theta\left(x[T]\right)}{q(x[T]|x[0])} \right]$$

$$+ \sum_{t=2}^{T} KL\left(q(x[t-1]|x[t],x[0]) \,\|\, p_\theta(x[t-1]|x[t]) \right)$$

$$+ E_{x[1:T]\sim q}\left[\log p_\theta(x[0]|x[1]) \right]$$

parameterization of $P_\theta(X[t-1]|X[t])$

mean: since $\tilde{u}_{t-1}(x_t, x_0) = \frac{1}{\sqrt{\alpha_t}} x_t - \frac{1-\alpha_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}} \epsilon$ where $x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon$

Instead of training a mean predictor

we parameterize it as $u_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} x_t - \frac{1-\alpha_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}} \cdot \epsilon_\theta(x_t, t)$

Variance: $\Sigma_q(x_t, x_0) = \frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}$ to match the variance of $q(X[t-1]|X[t], X[0])$

$$KL\left(\mathcal{N}(u_1, \mathrm{diag}(\sigma_1)) \,\|\, \mathcal{N}(u_2, \mathrm{diag}(\sigma_2))\right)$$

$$= \sum_i \frac{1}{2}\left[\log\frac{\sigma_{2i}}{\sigma_{1i}} - 1 + \frac{\sigma_{1i}}{\sigma_{2i}} + \frac{(u_{1i}-u_{2i})^2}{\sigma_{2i}^2}\right]$$

$$KL\left(q(X[t-1]|X[t], X[0]) \,\|\, P_\theta(X[t-1]|X[t])\right)$$

$$\Rightarrow \frac{1}{2\sigma_t^2} \cdot \left\| \left(\frac{1}{\sqrt{\alpha_t}} x_t - \frac{1-\alpha_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}} \epsilon\right) - \left(\frac{1}{\sqrt{\alpha_t}} x_t - \frac{1-\alpha_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t, t)\right) \right\|_2^2$$

$$= \frac{1}{2\sigma_t^2} \cdot \frac{(1-\alpha_t)^2}{\alpha_t(1-\bar{\alpha}_t)} \left\| \epsilon - \epsilon_\theta(x_t, t) \right\|_2^2$$

where $\sigma_t^2 = \frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}$

- Full Algorithm

repeat

$t \sim \mathrm{uniform}(1, T)$ training

$\epsilon \sim \mathcal{N}(0, I)$

$\min_\theta \left\| \epsilon - \epsilon_\theta\big(\underbrace{\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon}_{x_t}, t\big) \right\|_2^2$

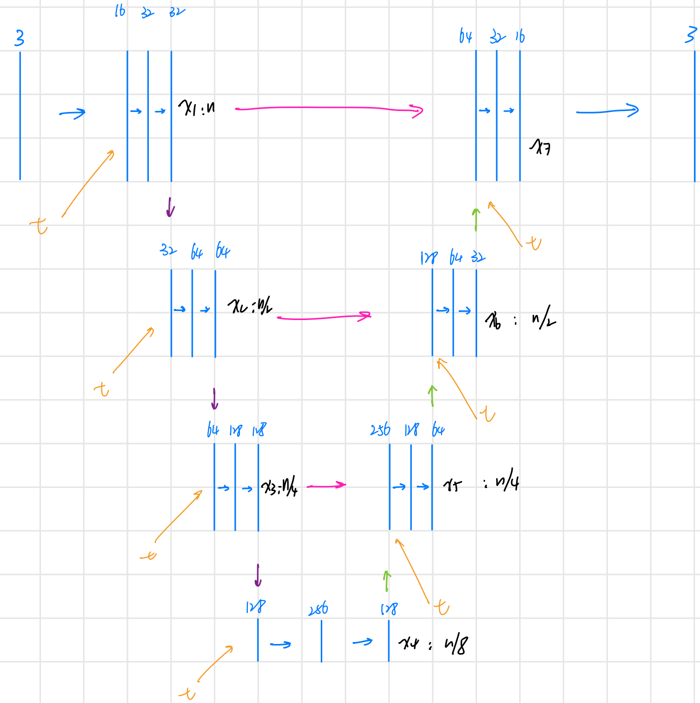
$x_t \sim \mathcal{N}(0, I)$ Sampling

for $t = T \ldots 1$

$z \sim \mathcal{N}(0, I)$

$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} x_t - \frac{1-\alpha_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t, t) + \sigma_t z$ $\sigma_t^2 = \frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}$

· U-net



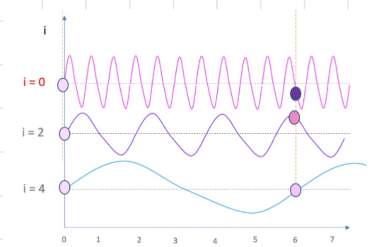
· Positional Encoding

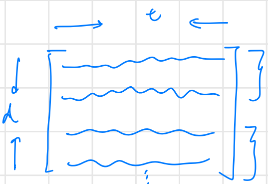
requirement : the encoding is deterministic
each unique encoding for each time step

let $E_t \in \mathbb{R}^d$ be the positional encoding at time-step t

$$E[t]_{2k} = \sin\left(\frac{t}{10000^{2k/d}}\right)$$

$$E[t]_{2k+1} = \cos\left(\frac{t}{10000^{2k/d}}\right)$$





$e^{i\theta t}$ $\quad \theta = \frac{1}{10000^{0/d}}$

$e^{i\theta t}$ $\quad \theta = \frac{1}{10000^{2/d}}$