① Recap

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

$\pi(x_0)$ $\qquad\qquad$ $\mathcal{N}\{0; I\}$

1. $x_t = \sqrt{1-\beta_t}\, x_{t-1} + \sqrt{\beta_t}\, \varepsilon, \quad \varepsilon \sim \mathcal{N}\{0; I\}$

   $q(x_t|x_{t-1}) \sim \mathcal{N}\{x_t | \sqrt{1-\beta_t}\, x_{t-1}; \beta_t I\}$

2. $\alpha_t = 1-\beta_t, \quad \bar{\alpha}_t = \prod_{i=1}^{t} \underbrace{\alpha_i}_{<1}$

   $x_t = \sqrt{\alpha_t}\, x_{t-1} + \sqrt{1-\alpha_t}\, \varepsilon$

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

   $q(x_t|x_0) = \mathcal{N}\{x_t | \underbrace{\sqrt{\bar{\alpha}_t}}_{\xrightarrow{t\to\infty} 0}\, x_0; \underbrace{(1-\bar{\alpha}_t)}_{\xrightarrow{t\to\infty} 1} I\}$

3. $q(x_T) \simeq \mathcal{N}\{0; I\}$

   $q(x_T|x_0) \simeq \mathcal{N}\{0; I\}$

   $\hookrightarrow \; q(x_T) = \int \underbrace{q(x_T|x_0)}_{\simeq \mathcal{N}\{0;I\}} \pi(x_0)\, dx_0 \simeq \mathcal{N}\{0; I\} \int \pi(x_0)\, dx = \mathcal{N}\{0; I\}$

4. In practice (e.g. CIFAR) $\qquad \beta_1 = 10^{-4}, \quad \beta_T = 2 \cdot 10^{-2}, \quad T = 10^3$

   betas = np.linspace(β_1; β_T; T)

   NB: If β_t small enough, $q(x_{t-1}|x_t) \sim \mathcal{N}$

   $T' \mapsto s = \frac{10^3}{T'} \mapsto \beta'_1 = s\,\beta_1; \; \beta'_T = s \cdot \beta_T$

② $\log p_\theta(x_0) \geq L(\theta) = \mathbb{E}_{q(x_1|x_0)} \log p_\theta(x_0|x_1)$ # recon. loss

$- KL\{q(x_T|x_0) \| p(x_T)\}$ # ≥ 0, const(θ)

$- \underbrace{\sum_{t=2}^{T} \mathbb{E}_{q(x_t|x_0)} KL\{q(x_{t-1}|x_t;x_0) \| p_\theta(x_{t-1}|x_t)\}}_{3}$

1. $p_\theta(x_{t-1}|x_t) = N\{x_{t-1}|\mu_\theta(x_t;t); \Sigma_\theta(x_t;t)\}$

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

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

$q(x_t|x_{t-1}) = N\{x_t|\sqrt{\alpha_t}\, x_{t-1}; (1-\alpha_t)I\}$

$q(x_t|x_0) = N\{x_t|\sqrt{\bar\alpha_t}\, x_0; (1-\bar\alpha_t)I\}$

$= C\cdot \exp\left\{-\frac{1}{2}\left[\frac{(x_t-\sqrt{\alpha_t}\,x_{t-1})^2}{1-\alpha_t} + \frac{(x_{t-1}-\sqrt{\bar\alpha_{t-1}}\,x_0)^2}{1-\bar\alpha_{t-1}} - \underbrace{\frac{(x_t-\sqrt{\bar\alpha_t}\,x_0)^2}{1-\bar\alpha_t}}_{C(x_0;x_t)}\right]\right\}$

$= C(x_0;x_t)\exp\left\{-\frac{1}{2}\left[x_{t-1}^2 \underbrace{\left(\frac{\alpha_t}{1-\alpha_t}+\frac{1}{1-\bar\alpha_{t-1}}\right)}_{\tilde\beta_t^{-1}} + x_{t-1}\left(\frac{-2\sqrt{\alpha_t}\,x_t}{1-\alpha_t}+\frac{-2\sqrt{\bar\alpha_{t-1}}\,x_0}{1-\bar\alpha_{t-1}}\right) + C(x_0;x_t)\right]\right\}$

$\Delta\ q(x_{t-1}|x_t;x_0) \sim N\{x_{t-1}|\tilde\mu_t;\tilde\beta_t\}$

$c\cdot\exp\{-\frac{1}{2}[ax^2 - 2bx]\} = c\exp\{-\frac{a}{2}[x^2 - 2\frac{b}{a}x \pm (\frac{b}{a})^2]\}$

$= c\exp\left\{-\dfrac{(x-\frac{b}{a})^2}{2\cdot\frac{1}{a}}\right\}$

3. $\tilde\beta_t^{-1} = \left(\frac{\alpha_t}{1-\alpha_t}+\frac{1}{1-\bar\alpha_{t-1}}\right) = \frac{\cancel{\alpha_t} - \bar\alpha_t + 1 - \cancel{\alpha_t}}{(1-\alpha_t)(1-\bar\alpha_{t-1})} \mapsto \boxed{\tilde\beta_t = \frac{1-\bar\alpha_{t-1}}{1-\bar\alpha_t}(1-\alpha_t)}$

4. $\boxed{\tilde\mu_t = \left(\frac{\sqrt{\alpha_t}\,x_t}{1-\alpha_t}+\frac{\sqrt{\bar\alpha_{t-1}}\,x_0}{1-\bar\alpha_{t-1}}\right)\left(\frac{1-\bar\alpha_{t-1}}{1-\bar\alpha_t}(1-\alpha_t)\right)}$

$= x_0\,\dfrac{\sqrt{\bar\alpha_{t-1}}\,(1-\alpha_t)}{(1-\bar\alpha_t)} + x_t\cdot\dfrac{\sqrt{\alpha_t}\,(1-\alpha_t)}{(1-\bar\alpha_t)}$

③ $L_t = KL\{\underbrace{q(x_{t-1}|x_t;x_0)}_{N\{\tilde\mu_t;\tilde\beta_t\}} \| p_\theta(x_{t-1}|x_0)\}$

$\Delta\ p_\theta(x_{t-1}|x_t) = N\{x_{t-1}|\mu_\theta(x_t;t);\tilde\beta_t\}$

$\hookrightarrow L_t = \frac{1}{2}\cdot\frac{1}{\tilde\beta_t}\|\mu_\theta(x_t;t) - \tilde\mu_t\|^2$

④ Recon Loss: $-\log p_\theta(x_0|x_1)$

x_0: $[0 \dots 255] \xrightarrow{\text{rescaled}} \underline{\underline{[-1 \dots 1]}}$

$$p_\theta(x_0|x_1) = \prod_i \int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)} \mathcal{N}\{x \mid \mu_\theta^i(x_1;1); \sigma\} \, dx$$

$$\delta_+(x_0^i) = \begin{cases} \infty, & x_0^i = 1 \\ x_0^i + \frac{1}{255}, & x_0^i < 1 \end{cases} \qquad \delta_-(x_0^i) = \begin{cases} -\infty, & x_0^i = -1 \\ x_0^i - \frac{1}{255}, & x_0^i > -1 \end{cases}$$

Though, In practice $\|\mu_\theta(x_1;1) - x_0\|_2^2$

⑤ $$L(\theta) = \underline{-\log p_\theta(x_0|x_1)} + \frac{1}{2}\sum_{t=2}^{T} \frac{1}{\beta_t} \|\mu_\theta(x_t;t) - \tilde{\mu}_t\|_2^2 + \text{const}(\theta)$$

$L_1 = \|\mu_\theta(x_1;1) - x_0\|$

Training procedure:

1. $x_0 \sim \pi(x_0)$, $t \sim U[1;T]$
2. $x_t \sim q(x_t|x_0)$
3. compute $\tilde{\mu}_t$, $\tilde{\beta}_t$
4. compute $L_t = \frac{1}{2\tilde{\beta}_t}\|\mu_\theta(x_t;t) - \tilde{\mu}_t\|_2^2$
5. $\theta \leftarrow \theta + \nabla_\theta L_t(\theta)$