DDPM reporter

watney0225

更新：2025 年 1 月 20 日

DDPM 主要使用了 Diffusion Model，主要分为前向过程（扩散过程）和反向过程（重建过程）。

在实际训练中，模型通过预测噪声，并去除预测的噪声来输出降噪后的图片。

1 前向过程

给定输入图片 x_0，其经过扩散过程的第 t 步表示为 x_t,，往后 $x_{t-1} \to x_t$ 是一个不断加噪声的过程。相邻两个变量之间是一个线性关系，我们可以建模成

$$x_t = a_t x_{t-1} + b_t \epsilon_t, \quad \epsilon_t \sim N(\mathbf{0}, I)$$

我们先看一下这两个系数，因为 x_{t-1} 具有的信息更多，因此 a_t 是一个衰减系数，值应该满足 $0 < a_t < 1$；同样的噪声系数也满足 $0 < b_t < 1$。

当我们用 $x_{t-1} = a_{t-1}x_{t-2} + b_{t-1}\epsilon_{t-1}$ 不断展开这个式子，可以得到

$$\begin{aligned}
x_t &= a_t x_{t-1} + b_t \epsilon_t \\
&= a_t(a_{t-1}x_{t-2} + b_{t-1}\epsilon_{t-1}) + b_t\epsilon_t \\
&= a_t a_{t-1} x_{t-2} + a_t b_{t-1}\epsilon_{t-1} + b_t\epsilon_t \\
&= \dots \\
&= (a_t \dots a_1)x_0 + (a_t \dots a_2)b_1\epsilon_1 + (a_t \dots a_3)b_2\epsilon_2 + \cdots + a_t b_{t-1}\epsilon_{t-1} + b_t\epsilon_t
\end{aligned}$$

假设 $X \sim N(\mu_1, \sigma_1^2)$ 和 $Y \sim N(\mu_2, \sigma_2^2)$ 是两个独立的正态随机变量，那么 $X+Y$ 也服从正态分布，其均值为 $\mu_1 + \mu_2$，方差为 $\sigma_1^2 + \sigma_2^2$。即 $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
$(a_t \dots a_2)^2 b_1^2 + (a_t \dots a_3)^2 b_2^2 + \cdots + a_t^2 b_{t-1}^2 + b_t^2$

这样原式就可以写成

$$x_t = (a_t \dots a_1)x_0 + \sqrt{(a_t \dots a_2)^2 b_1^2 + (a_t \dots a_3)^2 b_2^2 + \cdots + a_t^2 b_{t-1}^2 + b_t^2}\,\bar{\epsilon}_t, \quad \bar{\epsilon}_t \sim N(\mathbf{0}, I)$$

这里还有一个细节，如果我们把系数的平方和都加起来

$$
\begin{aligned}
&(a_t \dots a_1)^2 + (a_t \dots a_2)^2 b_1^2 + (a_t \dots a_3)^2 b_2^2 + \dots + a_t^2 b_{t-1}^2 + b_t^2 \\
=&(a_t \dots a_2)^2 a_1^2 + (a_t \dots a_2)^2 b_1^2 + (a_t \dots a_3)^2 b_2^2 + \dots + a_t^2 b_{t-1}^2 + b_t^2 \\
=&(a_t \dots a_2)^2 (a_1^2 + b_1^2) + (a_t \dots a_3)^2 b_2^2 + \dots + a_t^2 b_{t-1}^2 + b_t^2 \\
=&(a_t \dots a_3)^2 \left(a_2^2 (a_1^2 + b_1^2) + b_2^2\right) + \dots + a_t^2 b_{t-1}^2 + b_t^2 \\
=&a_t^2 \left(a_{t-1}^2 \left(\dots \left(a_2^2 (a_1^2 + b_1^2) + b_2^2\right) + \dots\right) + b_{t-1}^2\right) + b_t^2
\end{aligned}
$$

可以发现，如果我们加一个约束将会极大简化这个式子，即要求 $a_t^2 + b_t^2 = 1$，这样上面的平方和就等于 1 了。同时如果我们记 $\bar{a}_t = (a_t \dots a_1)^2$，平方和的后面部分就可以表示为 $1 - \bar{a}_t$ 。这样我们就可以将递推式写成

$x_t = \sqrt{\bar{a}_t} x_0 + \sqrt{1 - \bar{a}_t} \bar{\epsilon}_t, \quad \bar{\epsilon}_t \sim N(\mathbf{0}, I)$

同时把 a_t 用 $\sqrt{\alpha_t}$ 代替，把 b_t 用 $\sqrt{1 - \alpha_t}$ 代替，也就和原论文一致了。

$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1 - \alpha_t} \epsilon_t, \quad \epsilon_t \sim N(\mathbf{0}, I)$

并且记 $\bar{\alpha}_t = \alpha_t \dots \alpha_1$，得到递推式 $x_t = \sqrt{\bar{\alpha}_t} x_{t-1} + \sqrt{1 - \bar{\alpha}_t} \epsilon_t, \quad \bar{\epsilon}_t \sim N(\mathbf{0}, I)$

这就是整个前向过程，我们可以通过递推式经过一步就从 x_0 得到 x_t。

并且 $\lim_{t \to \infty} x_t \sim N(\mathbf{0}, I)$，这是由于 $0 < \alpha_t < 1$, 所以 $\lim_{t \to \infty} \bar{\alpha}_t = 0$

2 反向过程

一个比较自然的想法就是把刚刚的式子移项后就可以从 x_t 得到 x_0，

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

但是由于实际训练的效果不好，所以我们选择通过 x_t 得到 x_{t-1}, 从而逐渐推出 x_0。

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

在论文中，作者给出了这样的递推公式 $x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t, t)\right) + \sigma_t z$ ，其

中 z 是一个服从标准正态分布的噪声，我们接下来会给出这个公式的来源。

根据贝叶斯公式 $q(x_{t-1} \mid x_t) = \frac{q(x_t|x_{t-1})q(x_{t-1})}{q(x_t)}$

$q(x_{t-1}), q(x_t)$ 未知，但是 $q(x_{t-1} \mid x_0), q(x_t \mid x_0)$ 已知。

由马尔科夫性,$q(x_{t-1} \mid x_t) = q(x_{t-1} \mid x_t, x_0) = \frac{q(x_t|x_{t-1},x_0)q(x_{t-1}|x_0)}{q(x_t|x_0)} = \frac{q(x_t|x_{t-1})q(x_{t-1}|x_0)}{q(x_t|x_0)}$
由前向传播的递推式可知，这上面的每一个分布都是高斯分布，并且均值和方差如下。

$$
\begin{aligned}
q(x_t|x_{t-1}) &= N(x_t; \sqrt{\alpha_t} * x_{t-1}, (1-\alpha_t)I) \\
q(x_t|x_0) &= N(x_t; \sqrt{\overline{\alpha_t}} * x_0, (1-\overline{\alpha_t})I) \\
q(x_{t-1}|x_0) &= N(x_{t-1}; \sqrt{\overline{\alpha_{t-1}}} * x_0, (1-\overline{\alpha}_{t-1})I)
\end{aligned}
$$

一个高斯分布 $N(\mu, \sigma^2) \propto exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2})$

$$
\begin{aligned}
&\text{于是有} \quad \frac{q(x_t \mid x_{t-1})q(x_{t-1} \mid x_0)}{q(x_t \mid x_0)} \\
&\propto \exp\left(-\frac{1}{2}\left(\frac{(x_t - \sqrt{\alpha_t}x_{t-1})^2}{1-\alpha_t} + \frac{(x_{t-1} - \sqrt{\bar{\alpha}_{t-1}}x_0)^2}{1-\bar{\alpha}_{t-1}} - \frac{(x_t - \sqrt{\bar{\alpha}_t}x_0)^2}{1-\bar{\alpha}_t}\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
& q\left(x_{t-1} \mid x_t, x_0\right) \\
= & \frac{q\left(x_t \mid x_{t-1}, x_0\right) q\left(x_{t-1} \mid x_0\right)}{q\left(x_t \mid x_0\right)} \\
= & \frac{q\left(x_t \mid x_{t-1}\right) q\left(x_{t-1} \mid x_0\right)}{q\left(x_t \mid x_0\right)} \\
= & \frac{N\left(x_t ; \sqrt{\alpha_t} x_{t-1},\left(1-\alpha_t\right) I\right) N\left(x_{t-1} ; \sqrt{\bar{\alpha}_{t-1}} x_0,\left(1-\bar{\alpha}_{t-1}\right) I\right)}{N\left(x_t ; \sqrt{\bar{\alpha}_t} x_0,\left(1-\bar{\alpha}_t\right) I\right)} \\
\propto & \exp \left\{-\frac{1}{2}\left[\frac{\left(x_t-\sqrt{\alpha_t} x_{t-1}\right)^2}{\left(1-\alpha_t\right)}+\frac{\left(x_{t-1}-\sqrt{\bar{\alpha}_{t-1}} x_0\right)^2}{\left(1-\bar{\alpha}_{t-1}\right)}-\frac{\left(x_t-\sqrt{\bar{\alpha}_t} x_0\right)^2}{\left(1-\bar{\alpha}_t\right)}\right]\right\} \\
= & \exp \left\{-\frac{1}{2}\left[\frac{\left(x_t-\sqrt{\alpha_t} x_{t-1}\right)^2}{1-\alpha_t}+\frac{\left(x_{t-1}-\sqrt{\bar{\alpha}_{t-1}} x_0\right)^2}{1-\bar{\alpha}_{t-1}}-\frac{\left(x_t-\sqrt{\bar{\alpha}_t} x_0\right)^2}{1-\bar{\alpha}_t}\right]\right\} \\
= & \exp \left\{-\frac{1}{2}\left[\frac{-2 \sqrt{\alpha_t} x_t x_{t-1}+\alpha_t x_{t-1}^2}{1-\alpha_t}+\frac{x_{t-1}^2-2 \sqrt{\bar{\alpha}_{t-1}} x_{t-1} x_0}{1-\bar{\alpha}_{t-1}}+C\left(x_t, x_0\right)\right]\right\} \\
= & \exp \left\{-\frac{1}{2}\left[-\frac{2 \sqrt{\alpha_t} x_t x_{t-1}}{1-\alpha_t}+\frac{\alpha_t x_{t-1}^2}{1-\alpha_t}+\frac{x_{t-1}^2}{1-\bar{\alpha}_{t-1}}-\frac{2 \sqrt{\bar{\alpha}_{t-1}} x_{t-1} x_0}{1-\bar{\alpha}_{t-1}}+C\left(x_t, x_0\right)\right]\right\} \\
= & \exp \left\{-\frac{1}{2}\left[\left(\frac{\alpha_t}{1-\alpha_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right) x_{t-1}^2-2\left(\frac{\sqrt{\alpha_t} x_t}{1-\alpha_t}+\frac{\sqrt{\bar{\alpha}_{t-1}} x_0}{1-\bar{\alpha}_{t-1}}\right) x_{t-1}+C\left(x_t, x_0\right)\right]\right\} \\
= & \exp \left\{-\frac{1}{2}\left[\frac{\alpha_t\left(1-\bar{\alpha}_{t-1}\right)+1-\alpha_t}{\left(1-\alpha_t\right)\left(1-\bar{\alpha}_{t-1}\right)} x_{t-1}^2-2\left(\frac{\sqrt{\alpha_t} x_t}{1-\alpha_t}+\frac{\sqrt{\bar{\alpha}_{t-1}} x_0}{1-\bar{\alpha}_{t-1}}\right) x_{t-1}+C\left(x_t, x_0\right)\right]\right\} \\
= & \exp \left\{-\frac{1}{2}\left(A x_{t-1}^2+B x_{t-1}+C\right)\right\} \\
= & \exp \left\{-\frac{1}{2} A\left(x_{t-1}+\frac{B}{2 A}\right)^2+\bar{C}\right\}
\end{aligned}
$$

这里的 x_t 可以看做是常量，x_{t-1} 是变量，我们需要求出这个式子的均值和方差，所以一些无关的东西我都用 C 和 $\bar{C}$ 表示了。因此均值是 $-\frac{B}{2A}$ 方差是 A^{-1}

其中 $A=\frac{1-\bar{\alpha}_t}{\left(1-\alpha_t\right)\left(1-\bar{\alpha}_{t-1}\right)}, B=-2\left(\frac{\sqrt{\alpha_t} x_t}{1-\alpha_t}+\frac{\sqrt{\bar{\alpha}_{t-1}} x_0}{1-\bar{\alpha}_{t-1}}\right)$

因此 $\mu_q\left(x_t, x_0\right)=\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right) x_t+\sqrt{\bar{\alpha}_{t-1}}\left(1-\alpha_t\right) x_0}{1-\bar{\alpha}_t}, \quad \Sigma_q=\frac{\left(1-\alpha_t\right)\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t} I$

借助 $x_t=\sqrt{\bar{a}_t} x_0+\sqrt{1-\bar{a}_t} \bar{\epsilon}_t, \quad \bar{\epsilon}_t$ 消去 x_0 可得

$$
\begin{aligned}
\boldsymbol{\mu}_q &= \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\boldsymbol{x}_0 + \frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}\boldsymbol{x}_t \\
&= \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\frac{\boldsymbol{x}_t-\sqrt{1-\bar{\alpha}_t}\boldsymbol{\varepsilon}_t}{\sqrt{\bar{\alpha}_t}} + \frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}\boldsymbol{x}_t \\
&= \left[\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{(1-\bar{\alpha}_t)\sqrt{\bar{\alpha}_t}} + \frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}\right]\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}}\boldsymbol{\varepsilon}_t \\
&= \left[\frac{\beta_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}} + \frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}\right]\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}}\boldsymbol{\varepsilon}_t \\
&= \frac{(1-\alpha_t)+\alpha_t(1-\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}}\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}}\boldsymbol{\varepsilon}_t \\
&= \frac{1-\alpha_t+\alpha_t-\bar{\alpha}_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}}\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}}\boldsymbol{\varepsilon}_t \\
&= \frac{1}{\sqrt{\alpha_t}}\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}}\boldsymbol{\varepsilon}_t
\end{aligned}
$$

和论文中的式子一致

3 优化目标

由 VAE 的理论，这一类模型可以使用极大似然估计来求解，我们来给出这个模型的 KL 散度来进行极大似然估计。

$p_\theta(x_0) = \int p_\theta(x_0, x_1, \ldots, x_T)dx_1 dx_2 \ldots dx_T$

$p_\theta(x_0) = \int p_\theta(x_{0:T})dx_{1:T}$

$p_\theta(x_{0:T}) = p(x_T) * p_\theta(x_{T-1}|x_T) * p_\theta(x_{T-2}|x_{T-1}) * \cdots * p_\theta(x_0|x_1)$

$p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$ 是一个高斯分布

取对数似然, 并且由全概率公式 $P(A) = \int P(A \mid B)P(B)\,dB$，我们可以得到

$$
\begin{aligned}
&\log p_\theta(x_0) \\
&= \log \int p_\theta(x_0, x_1, \ldots, x_T) dx_1 dx_2 \ldots dx_T \\
&= \log \int p_\theta\left(x_{0:T}\right) dx_{1:T} \\
&= \log \int \frac{p_\theta(x_{0:T}) q(x_{1:T}|x_0)}{q(x_{1:T}|x_0)} dx_{1:T} \\
&= \log \mathbb{E}_{q(x_{1:T}|x_0)} \left[\frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)} \right] \text{(期望的定义)} \\
&\geq \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)} \right] \text{(Jensen 不等式)} \\
&= \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}{\prod_{t=1}^{T} q(x_t|x_{t-1})} \right] \text{(之前的式子)} \\
&= \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{p(x_T) p_\theta(x_0|x_1) \prod_{t=2}^{T} p_\theta(x_{t-1}|x_t)}{q(x_1|x_0) \prod_{t=2}^{T} q(x_t|x_{t-1})} \right] \\
&= \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{p(x_T) p_\theta(x_0|x_1) \prod_{t=2}^{T} p_\theta(x_{t-1}|x_t)}{q(x_1|x_0) \prod_{t=2}^{T} q(x_t|x_{t-1}, x_0)} \right] \\
&= \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{p_\theta(x_T) p_\theta(x_0|x_1)}{q(x_1|x_0)} + \log \prod_{t=2}^{T} \frac{p_\theta(x_{t-1}|x_t)}{q(x_t|x_{t-1}, x_0)} \right] \text{(拆分)} \\
&= \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{p(x_T) p_\theta(x_0|x_1)}{q(x_1|x_0)} + \log \prod_{t=2}^{T} \frac{p_\theta(x_{t-1}|x_t)}{\frac{q(x_{t-1}|x_t, x_0) q(x_t|x_0)}{q(x_{t-1}|x_0)}} \right] \\
&= \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{p(x_T) p_\theta(x_0|x_1)}{q(x_1|x_0)} + \log \prod_{t=2}^{T} \frac{q(x_{t-1}|x_0)}{q(x_t|x_0)} + \log \prod_{t=2}^{T} \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)} \right] \\
&= \mathbb{E}_{q(x_{1:T}|x_0)} [\log \frac{p(x_T) p_\theta(x_0|x_1)}{q(x_1|x_0)} + \log \frac{q(x_1|x_0)}{q(x_2|x_0)} \frac{q(x_2|x_0)}{q(x_3|x_0)} \cdots \frac{q(x_{T-1}|x_0)}{q(x_T|x_0)} \\
&+ \log \prod_{t=2}^{T} \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)}] \\
&= \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{p(x_T) p_\theta(x_0|x_1)}{q(x_1|x_0)} + \log \frac{q(x_1|x_0)}{q(x_T|x_0)} + \log \prod_{t=2}^{T} \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)} \right] \\
&= \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{p(x_T) p_\theta(x_0|x_1)}{q(x_1|x_0)} \frac{q(x_1|x_0)}{q(x_T|x_0)} + \log \prod_{t=2}^{T} \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)} \right] \\
&= \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{p(x_T) p_\theta(x_0|x_1)}{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)} \right]
\end{aligned}
$$

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log p_\theta(x_0|x_1)\right] + \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p(x_T)}{q(x_T|x_0)}\right] + \sum_{t=2}^{T} \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)}\right]$$

$$= \mathbb{E}_{q(x_1|x_0)}\left[\log p_\theta(x_0|x_1)\right] + \mathbb{E}_{q(x_T|x_0)}\left[\log \frac{p(x_T)}{q(x_T|x_0)}\right] + \sum_{t=2}^{T} \mathbb{E}_{q(x_{t-1},x_t|x_0)}\left[\log \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)}\right]$$

这里的下标变化是因为期望的内容只和其中几个x_i 有关。

例如，对于$\mathbb{E}_{q(x_{1:T}|x_0)}\left[\log p_\theta(x_0|x_1)\right]$ 由于 $\log p_\theta(x_0|x_1)$ 仅依赖于x_1

所以可将对$q(x_{1:T}|x_0)$ 的期望简化为对$q(x_1|x_0)$ 的期望，

同理适用于其他两项。

$$= \mathbb{E}_{q(x_1|x_0)}\left[\log p_\theta(x_0|x_1)\right] - \mathbb{E}_{q(x_T|x_0)}\left[\log \frac{q(x_T|x_0)}{p(x_T)}\right]$$

$$- \sum_{t=2}^{T} \mathbb{E}_{q(x_t|x_0)}\mathbb{E}_{q(x_{t-1}|x_t,x_0)}\left[\log \frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)}\right] \text{(条件期望)}$$

$$= \underbrace{\mathbb{E}_{q(x_1|x_0)}\left[\log p_\theta(x_0|x_1)\right]}_{\text{reconstruction term}} - \underbrace{D_{\mathrm{KL}}(q(x_T|x_0)\|p(x_T))}_{\text{prior matching term}}$$

$$- \sum_{t=2}^{T} \underbrace{\mathbb{E}_{q(x_t|x_0)}\left[D_{\mathrm{KL}}(q(x_{t-1}|x_t, x_0)\|p_\theta(x_{t-1}|x_t))\right]}_{\text{denoising matching term}}$$

这里将各项分别标记为不同的项，$\mathbb{E}_{q(x_1|x_0)}\left[\log p_\theta(x_0|x_1)\right]$ 被称为重构项，因为它与从x_1 重构

$D_{\mathrm{KL}}(q(x_T|x_0)\|p(x_T))$ 是先验匹配项，用于衡量$q(x_T|x_0)$ 与$p(x_T)$ 的差异；

$\sum_{t=2}^{T} \mathbb{E}_{q(x_t|x_0)}\left[D_{\mathrm{KL}}(q(x_{t-1}|x_t, x_0)\|p_\theta(x_{t-1}|x_t))\right]$ 是去噪匹配项，反映了在不同时间步的去噪过

之后我们还需要求解 $p_\theta(x_{t-1}|x_t)$

作者令 $\Sigma_\theta(x_t, t) = \sigma_t^2$, 并且经过实验，$\sigma_t^2 = \beta_t$ 或者 $\sigma_t^2 = \widetilde{\beta}_t$， $\widetilde{\beta}_t = \Sigma_q$

4 损失函数

根据$D_{\mathrm{KL}}\left(\mathcal{N}\left(x;\mu_x,\Sigma_x\right)\|\mathcal{N}\left(y;\mu_y,\Sigma_y\right)\right)$
$=\frac{1}{2}\left[\log\frac{|\Sigma_y|}{|\Sigma_x|}-d+\operatorname{tr}\left(\Sigma_y^{-1}\Sigma_x\right)+\left(\mu_y-\mu_x\right)^T\Sigma_y^{-1}\left(\mu_y-\mu_x\right)\right]$
可以得到

$$
\begin{aligned}
&\arg\min_\theta D_{\mathrm{KL}}\left(q\left(x_{t-1}|x_t,x_0\right)\|p_\theta\left(x_{t-1}|x_t\right)\right)\\
=&\arg\min_\theta D_{\mathrm{KL}}\left(\mathcal{N}\left(x_{t-1};\mu_q,\Sigma_q(t)\right)\|\mathcal{N}\left(x_{t-1};\mu_\theta,\Sigma_q(t)\right)\right)\\
=&\arg\min_\theta\frac{1}{2}\left[\log\frac{|\Sigma_q(t)|}{|\Sigma_q(t)|}-d+\operatorname{tr}\left(\Sigma_q(t)^{-1}\Sigma_q(t)\right)+\left(\mu_\theta-\mu_q\right)^T\Sigma_q(t)^{-1}\left(\mu_\theta-\mu_q\right)\right]\\
=&\arg\min_\theta\frac{1}{2}\left[\log 1-d+d+\left(\mu_\theta-\mu_q\right)^T\Sigma_q(t)^{-1}\left(\mu_\theta-\mu_q\right)\right]\\
=&\arg\min_\theta\frac{1}{2}\left[\left(\mu_\theta-\mu_q\right)^T\Sigma_q(t)^{-1}\left(\mu_\theta-\mu_q\right)\right]\\
=&\arg\min_\theta\frac{1}{2}\left[\left(\mu_\theta-\mu_q\right)^T\left(\sigma_q^2(t)\mathbf{I}\right)^{-1}\left(\mu_\theta-\mu_q\right)\right]\\
=&\arg\min_\theta\frac{1}{2\sigma_q^2(t)}\left[\|\mu_\theta-\mu_q\|_2^2\right]
\end{aligned}
$$

由于 μ_q,σ_q 已知, 因此我们需要比较的是两个模型之间的均值。
由于 $\mu_q=\frac{\sqrt{\bar\alpha_{t-1}}\beta_t}{1-\bar\alpha_t}x_0+\frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})}{1-\bar\alpha_t}x_t$
我们可以令 $\mu_\theta=\frac{\sqrt{\bar\alpha_{t-1}}\beta_t}{1-\bar\alpha_t}f_\theta(x_t,t)+\frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})}{1-\bar\alpha_t}x_t$

可以计算出

$$
\begin{aligned}
&\arg\min_{\boldsymbol{\theta}} D_{\mathrm{KL}}\left(q\left(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0\right)\|p_{\boldsymbol{\theta}}\left(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t\right)\right)\\
=&\arg\min_{\boldsymbol{\theta}}\frac{1}{2\sigma_q^2(t)}\left[\left\|\boldsymbol{\mu}_{\boldsymbol{\theta}}-\boldsymbol{\mu}_q\right\|_2^2\right]\\
=&\arg\min_{\boldsymbol{\theta}}\frac{1}{2\sigma_q^2(t)}\left[\left\|\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t f_{\boldsymbol{\theta}(\boldsymbol{x}_t,t)}}{1-\bar{\alpha}_t}-\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t\boldsymbol{x}_0}{1-\bar{\alpha}_t}\right\|_2^2\right]\\
=&\arg\min_{\boldsymbol{\theta}}\frac{1}{2\sigma_q^2(t)}\left[\left\|\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\left(f_{\boldsymbol{\theta}(\boldsymbol{x}_t,t)-\boldsymbol{x}_0}\right)\right\|_2^2\right]\\
=&\arg\min_{\boldsymbol{\theta}}\frac{1}{2\sigma_q^2(t)}\frac{\bar{\alpha}_{t-1}\beta_t^2}{\left(1-\bar{\alpha}_t\right)^2}\left[\left\|f_{\boldsymbol{\theta}(\boldsymbol{x}_t,t)-\boldsymbol{x}_0}\right\|_2^2\right]
\end{aligned}
$$

相比于输出图片，作者的实验表明输出噪声的效果更好，因此有以下式子。

$$
\begin{aligned}
&\arg\min_{\boldsymbol{\theta}} D_{\mathrm{KL}}\left(q\left(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0\right)\|p_{\boldsymbol{\theta}}\left(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t\right)\right)\\
=&\arg\min_{\boldsymbol{\theta}}\frac{1}{2\sigma_q^2(t)}\frac{\left(1-\alpha_t\right)^2}{\left(1-\bar{\alpha}_t\right)\alpha_t}\left[\left\|f_{\boldsymbol{\theta}(\boldsymbol{x}_t,t)-\boldsymbol{\varepsilon}_t}\right\|_2^2\right]
\end{aligned}
$$