$$x_0 \longrightarrow x_1 \cdots \longrightarrow x_{t-1} \xrightarrow{q(x_t|x_{t-1})} x_t \longrightarrow \cdots \longrightarrow x_T$$

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

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

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

$$q(x_T) = \int q(x_0)\, q(x_1|x_0)\, q(x_2|x_1) \cdots q(x_T|x_{T-1})\, dx_{0:T}$$

如果用真实 $q(x_{t-1}|x_t) = \dfrac{q(x_t|x_{t-1})\, q(x_{t-1})}{q(x_t)}$ 进行采样

则逆过程的

$$p(x_0) = \int q(x_T) \cdot \prod_{t=T}^{1} q(x_{t-1}|x_t)\, dx_{1:T}$$

$$= \int q(x_T) \cdot \prod_{t=T}^{1} \frac{q(x_t|x_{t-1})\, q(x_{t-1})}{q(x_t)}\, dx_{1:T}$$

$$= \int q(x_0) \cdot \prod_{t=1}^{T} q(x_t|x_{t-1})\, dx_{1:T} = q(x_0)$$

能成功还原 $q(x_0)$

~~为了完成这个目标，有两种方式. ①有~~

~~方法一：$p_\theta(x_{t-1}|x_t)$ 拟合~~

也就是说 如果有网络能够拟合 ~~$q(x_t|x_{t-1})$~~ $q(x_{t-1}|x_t)$

我们能完成完美采样

定义网络 $p_\theta(x_{t-1}|x_t)$. 如何优化？

极大似然估计. $p_\theta(x_0) = \int p_\theta(x_{0:T})\, dx_{1:T}$

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

$$= \frac{1}{N}\sum_{i=1}^{N} p_\theta(x_0|x_{1:T}^i)$$

① 采样 非常困难.

② 概率输入, x_0 与 $x_{1:T}$ 无关联.

采用重要性采样

$$p_\theta(x_0)=\int p_\theta(x_{0:T})/q(x_{1:T}|x_0)\cdot q(x_{1:T}|x_0)\,dx_{1:T}$$

好处：① 引入 $x_{1:T}$ 与 x_0 的关联，提高网络取采样到有意义区域

② 提高采取样本速度，正向过程中的 $x_{1:T}$ 计算方便

上式利用蒙特卡洛方法可以计算为

$$p_\theta(x_0)\approx\frac{1}{N}\sum_{i=1}^{N}\frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}$$

问题：采样估计存在偏差，上述目标可能不稳定。

利用不等式进行放缩，最大化下界

$$\log p_\theta(x_0)=\log\int\frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\cdot q(x_{1:T}|x_0)\,dx_{1:T}$$

$$\geq\int q(x_{1:T}|x_0)\log\frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\,dx_{1:T}$$

$$=E_{x_{1:T}\sim q}\left[\log\frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right]$$

$$p_\theta(x_{0:T})=\cancel{p_\theta(x_0)}\;p_\theta(x_T)\cdot\prod_{t=T}p_\theta(x_{t-1}|x_t)$$

$$q(x_{1:T}|x_0)=\underbrace{\prod_{t=1}^{T}q(x_t|x_{t-1})}_{①}=\underbrace{q(x_1|x_0)\prod_{T=2}^{T}q(x_t|x_{t-1},x_0)}_{②}$$

如果按式①展开

$$\log p_\theta(x_0)\geq E_{x_{1:T}\sim q}\left[\log\frac{p_\theta(x_T)\prod_{t=T}^{1}p_\theta(x_{t-1}|x_t)}{\prod_{t=1}^{T}q(x_t|x_{t-1})}\right]$$

$$=E_{x_{1:T}\sim q}\left[\log\frac{p_\theta(x_T)\prod_{t=T}^{1}p_\theta(x_{t-1}|x_t)}{\prod_{t=1}^{T}\frac{q(x_{t-1}|x_t)q(x_t)}{q(x_{t-1})}}\right]$$

$$=E_{x_{1:T}\sim q}\left[\log\frac{p_\theta(x_T)}{q(x_T)}+\underbrace{\sum_{t=1}^{T}\log\frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t)}}_{}+\log q(x_0)\right]$$

负的KL

与E结合成E[KL]

不确定性

令 $p_\theta(x_{t-1}|x_0)$ 拟合 $q(x_{t-1}|x_0)$

②式展开:

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

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

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

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

$$\cancel{= \mathbb{E}_{x_{1:T}\sim q}[}$$

$$\mathbb{E}_{x_0}[\log p_\theta(x_0)] \geq \mathbb{E}_q\Big[-KL(q(x_T|x_0)\,||\,p_\theta(x_T)) - \sum_{t=2}^{T} KL(q(x_{t-1}|x_t,x_0)\,||\,p_\theta(x_{t-1}|x_t)) + \log p_\theta(x_0|x_1)\Big].$$

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

$$\propto \exp\left[-\frac{(x_t-\sqrt{\alpha_t}\,x_{t-1})^2}{2(1-\alpha_t)} - \frac{(x_{t-1}-\sqrt{\bar\alpha_{t-1}}\,x_0)^2}{2(1-\bar\alpha_{t-1})} + \frac{(x_t-\sqrt{\bar\alpha_t}\,x_0)^2}{2(1-\bar\alpha_t)}\right]$$

快速凑平方

$$\frac{\partial \log q(x_{t-1}|x_t,x_0)}{\partial x_{t-1}} = \frac{\sqrt{\alpha_t}(x_t-\sqrt{\alpha_t}\,x_{t-1})}{1-\alpha_t} - \frac{x_{t-1}-\sqrt{\bar\alpha_{t-1}}\,x_0}{1-\bar\alpha_{t-1}} = 0$$

$$x_T = \frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})x_t + (1-\alpha_t)\sqrt{\bar\alpha_{t-1}}\,x_0}{1-\bar\alpha_t} \rightarrow x_0 = \frac{x_t-\sqrt{1-\bar\alpha_t}\,\varepsilon}{\sqrt{\bar\alpha_t}}$$

$$= \frac{\alpha_t(1-\bar\alpha_{t-1})x_t + (1-\alpha_t)(x_t-\sqrt{1-\bar\alpha_t}\,\varepsilon)}{\sqrt{\alpha_t}\,(1-\bar\alpha_t)}$$

$$= \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{(1-\alpha_t)}{\sqrt{1-\bar\alpha_t}}\varepsilon\right)$$

~~求~~ 求方差

$$\frac{\partial^2 \log}{\partial x_{t-1}} = -\frac{\alpha_t}{1-\alpha_t} - \frac{1}{1-\bar{\alpha}_{t-1}} = \frac{-\alpha_t + \bar{\alpha}_t - 1 + \alpha_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})} = 2a$$

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

$$-\frac{(x-u)^2}{2\cdot(-\frac{1}{2a})}$$

求得 $u = \frac{1}{\sqrt{\alpha_t}}\left[x_t - \frac{(1-\alpha_t)}{\sqrt{1-\bar{\alpha}_t}}\varepsilon\right]$ $\left(ps: u = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\varepsilon)\right)$

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

定 $p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; u_\theta(x_t,t), \sigma(x_t,t))$

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

$$\sigma_\theta(x_t,t) = \sigma$$

则 ~~KL(p(~~ $KL(q(x_{t-1}|x_t,x_0) \| p_\theta(x_{t-1}|x_t))$

$$= \frac{1}{2\sigma}\|u_\theta - u\|^2$$

$$= \frac{\beta_t^2}{2\sigma\alpha_t(1-\bar{\alpha}_t)}\|\varepsilon_\theta - \varepsilon\|^2$$ 权重一般是递减的

简化 loss $= E_{x_0,\varepsilon,t}\|\varepsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\varepsilon, t) - \varepsilon\|^2$

↓

简化loss在 $t\to 0$ 处约束变小

$t\to 1$ 处约束变大.

随机过程 $\{X(t), t\in T\}$ 可以看作是时间序列上的一系列随机变量：

$B(t)$ or $W(t)$ 布朗运动 或维纳过程

① $B(0)=0$

② ~~独立~~独立增量：$t_1>s_1$，$t_2>s_2$ 且 $[t_1,s_1]$，$[t_2,s_2]$ 没有交集

则 $B(t_1)-B(s_1)$ 与 $B(t_2)-B(s_2)$ 独立

③ 正态分布增量 $t>s$，$B(t)-B(s)\sim N(0,t-s)$

④ 无记忆性：增量 $\Delta B=B(t)-B(s)$ 只与 $t-s$ 有关 与 $B(s)$ $B(t)$ 值无关

随机微分方程（SDE）

$$dx=\underbrace{f(x,t)dt}_{\text{确定性}}+\underbrace{g(x,t)dB}_{\text{随机噪声}}$$

$dB\sim N(0,dt)$，$(dB)^2=dt$（ps：$E(dB^2)=dt$）

看一个例子，$dx=g(t)dB$

易解得 $x=x_0+\underbrace{\int_0^t g(u)dB}_{\text{有什么性质？}}$

$I(t)=\int_0^t g(u)dB$

① 由于 dB 每个时间 t 都相互独立，$I(t)$ 可以看作相互独立的（正态）高斯分布的叠加 可知 $I(t)\sim N(u,\sigma^2)$

$$u=E(I(t))=E\left(\int_0^t g(u)dB\right)=\int_0^t E(g(u)dB)=\int_0^t 0=0$$

$$\sigma^2=D(I(t))=E(I^2(t))-[E(I(t))]^2=E(I^2(t))=E\left[\int_0^t g(u)dB\int_0^t g(v)dB\right]$$

$$E(I^2(t)) = E\left(\int_0^t g(u)dB(u) \int_0^t g(v)dB(v)\right) = E\left(\int_0^t \int_0^t g(u)g(v)dB(u)dB(v)\right)$$

$$= \int_0^t \int_{u=v} E(g(u)g(v)dB(u)dB(v)) + \int_{u\neq v} E(g(u)g(v)dB(u)dB(v))$$

$$= \int_0^t g^2(u)dt.$$

$\sigma^2 = \int_0^t g^2(u)dt.$ (u时刻 $dB \sim N(0, dt)$.
则 $g(u)dB \sim N(0, g^2(u)dt)$.
方差相加则 $D(I(t)) = \int_0^t g^2(u)dt$)

再看一个ODE：

$$dx = f(t)xdt.$$

不难解得： $x = x_0 \exp(\int_0^t f(u)du)$.

如何理解 $\exp(\int_0^t f(u)du)$ 这一项？

$$dx = x_{t+dt} - x_t = f(t)x_t dt$$

$$x_{t+dt} = (1 + f(t)dt)x_t.$$

~~$x_t = (1+f(t)$~~

$$x_{dt} = (1 + f(0)dt)x_0$$

$$x_{2dt} = (1 + f(dt)dt)x_{dt} = (1 + f(dt)dt)(1 + f(0)dt)x_0.$$

$$x_t = (1 + f(t-dt)dt) \cdots\cdots (1 + f(0)dt)x_0.$$

$$\overset{\text{taylor}}{=} \exp(f(t-dt)dt) \cdots \exp(f(0)dt)$$

$$= x_0 \exp\left(\sum_k f(kdt)dt\right) \overset{k\to\infty}{=} x_0 \exp\left(\int_0^t f(u)du\right)$$

再看一个 SDE： $dx = f(t)x\,dt + g(t)dB$.

ODE：$dx = f(t)x\,dt$ 解为 $x = x_0\exp(\int_0^t f(u)du)$

构造 $\mu(t) = \exp(\int_0^t -f(u)du)$

$$\mu(t)dx - \mu(t)f(t)x\,dt = g(t)\mu(t)dB$$

$$\Updownarrow$$

$$d(\mu(t)x) = g(t)\mu(t)dB$$

$$d(\mu(t)x) = d\mu(t)\cdot x + \mu(t)dx$$

$$= \mu(t)\cdot(-f(t))x\cdot dt + \mu(t)dx.$$

$$\mu(t)x = \int_0^t g(u)\mu(u)dB + \mu(0)x_0.$$

$$\cancel{x = \frac{\mu(u)x}{\mu(t)} = \frac{\exp(\int_0^u -f(s)ds)}{\exp(\int_0^t -f(s)ds)}}$$

$$\frac{\mu(s)}{\mu(t)} = \frac{\exp(\int_0^s -f(u)du)}{\exp(\int_0^t -f(u)du)} = \exp(\int_s^t f(u)du).$$

$$x = \overbrace{\underbrace{x_0\exp(\int_0^t f(u)du)}_{\text{确定性项}}}^{\text{衰减项}} + \underbrace{\int_0^t \overbrace{\exp(\int_u^t f(s)ds)}^{\text{衰减项}}g(u)dB}_{\text{噪声项}}$$

显然将 $dx = f(t)x\,dt + g(t)dB$ 按照前面 ODE 的写法

可以写成 ~~$x = x_0\exp$~~

$$x = x_0\prod_{u=0}^t(1+f(u)du) + \sum_{u=0}^t\left(\prod_{s=u}^t(1+f(s)ds)\right)g(u)dB.$$

taylor

$$= x_0\exp(\sum_{u=0}^t f(u)du) + \sum_{u=0}^t \exp(\sum_{s=u}^t f(s)ds)g(u)dB$$

$$= x_0\exp(\int_0^t f(u)du) + \int_0^t \exp(\int_u^t f(s)ds)g(u)dB$$

回归正题.

$$dx = f(x,t)dt + g(t)dB.$$

则 $X_{t+dt} = X_t + f(X_t,t)dt + g(t)dB.$

$$p(X_{t+dt}|X_t) = \mathcal{N}(X_{t+dt};\ X_t + f(X_t,t)dt,\ g^2(t)dt)$$

按照~~前面~~反向加Bayes, 用真实的 $p(X_t|X_{t+dt})$ 逆采样可得 $p(X_0)$

$$p(X_t|X_{t+dt}) = \frac{p(X_{t+dt}|X_t)\,p(X_t)}{p(X_{t+dt})}$$

$$\propto \exp\left(-\frac{[X_{t+dt}-(X_t+f(X_t,t)dt)]^2}{2g^2(t)dt} + \log p(X_t) - \log p(X_{t+dt})\right)$$

$$\log p(X_t) - \log p(X_{t+dt}) = -d\log p(X_t)$$

$$= -\left(\nabla_x \log p(X_t)\,dx + \frac{\partial \log p(x)}{\partial t}dt\right)$$

$$= (X_t - X_{t+dt})\nabla_x \log p(X_t) + O(t).$$

$$\propto \exp\left(-\frac{(X_{t+dt}-(X_t+f(X_t,t)dt))^2}{2g^2(t)dt} + (X_t - X_{t+dt})\nabla_x \log p(X_t)\right)$$

快速配方法: $\frac{\partial \log p(X_t|X_{t+dt})}{\partial X_t} = \frac{X_{t+dt}-(X_t+f(X_t,t)dt)}{g^2(t)dt} + \nabla_x \log p(X_t) = 0$

$$\Rightarrow X_t = X_{t+dt} - [f(X_t,t) - g^2(t)\nabla_x \log p(x)]dt.$$

取 $\Delta t = -dt$, ~~一阶~~近似 ~~…~~ $h(t+dt)dt = h(t)dt$;
一阶

$\frac{\partial^2 \log p}{\partial x} = -\frac{1}{g^2(t)dt} = 2a$

$\sigma^2 = -\frac{1}{2a} = g^2(t)dB$

有 $dx = [f(x,t) - g^2(t)\nabla_x \log p(x)]dt + g(t)dB.$

即 前向 $dx = f(x,t)dt + g(t)dB.$

对应逆向 $dx = [f(x,t) - g^2(t)\nabla_x \log p(x)]dt + g(t)dB.$

NCSN:

$$x_{i+1} = x_i + \sqrt{\sigma^2(i+1) - \sigma^2(i)}\,\varepsilon \quad , i = 1, \cdots, N$$

取 ~~$\Delta t = \frac{1}{N}$~~, $t = \frac{i}{N}$, $\Delta t = \frac{1}{N}$.

则 $x_{t+\Delta t} = x_t + \sqrt{\sigma^2(t+\Delta t) - \sigma^2(t)}\,\varepsilon$

$$= x_t + \sqrt{\frac{d[\sigma^2(t)]}{dt}\Delta t}\,\varepsilon$$

$$= x_t + \sqrt{\frac{d[\sigma^2(t)]}{dt}}\,\Delta B$$

$$\Rightarrow dx = \sqrt{\frac{d[\sigma^2(t)]}{dt}}\,dB.$$

解得. $x = x_0 + \int_0^t \frac{d[\sigma^2(u)]}{du} \cdot du$

$= x_0 + \sigma^2(t)$. 符合原来离散形式.

DDPM: $x_i = \sqrt{1-\beta_i}\,x_{i-1} + \sqrt{\beta_i}\,\varepsilon$, $i = 1, \cdots, N$

$dt = \frac{1}{N}$

~~$\sqrt{\beta(t)} = \sqrt{\beta_i}$~~ $\boxed{\beta_i = \beta(t)dt}$

取 $\beta(t) = N\beta_{i/N}$, $\Delta t = \frac{1}{N}$

则 $x_t = \sqrt{1-\beta(t)\Delta t}\,x_{t-\Delta t} + \sqrt{\beta(t)\Delta t}\,\varepsilon$

$h(x) = \sqrt{1-x}$

$h(x) = h(0) + \frac{h'(0)}{1!}(x-0)$

$= 1 - \frac{1}{2}x$

$$\approx \left[1 - \frac{1}{2}\beta(t)\Delta t\right] x_{t-\Delta t} + \sqrt{\beta(t)}\,\Delta B$$

$$\therefore dx = -\frac{1}{2}\beta(t)x\,dt + \sqrt{\beta(t)}\,dB$$

解得 $x = x_0 \exp\left(\int_0^t -\frac{1}{2}\beta(u)du\right) + \int_0^t \exp\left(\int_u^t -\frac{1}{2}\beta(s)ds\right)\sqrt{\beta(u)}\,dB$

$= x_0 \exp\left(\sum_{u=1}^{t} -\frac{1}{2}\beta(u)du\right) + \sum_{u=1}^{t} \exp\left(\sum_{s=u+1}^{t} -\frac{1}{2}\beta(s)ds\right)\sqrt{\beta(u)}\,dB$

$= x_0 \prod_{u=1}^{t} \exp\left(-\frac{1}{2}\beta(u)du\right) + \sum_{u=1}^{t}\left(\prod_{s=u+1}^{t} \exp\left(-\frac{1}{2}\beta(s)ds\right)\right)\sqrt{\beta(u)}\,dB$

$= x_0 \prod_{u=1}^{t} \left(1 - \frac{1}{2}\beta(u)du\right) + \sum_{u=1}^{t}\left(\prod_{s=u+1}^{t} \left(1 - \frac{1}{2}\beta(s)ds\right)\right)\sqrt{\beta(u)}\,dB$

$= x_0 \prod_{u=1}^{t} \sqrt{1-\beta(u)du} + \sum_{u=1}^{t}\left(\prod_{s=u+1}^{t} \sqrt{1-\beta(s)ds}\right)\sqrt{\beta(u)}\,dB$

$$= x_0 \sqrt{\bar{\alpha}_t} + \sum_{u=1}^{t} \left(\prod_{s=u+1}^{t} \sqrt{\alpha_s} \right) \sqrt{\beta(u)}\, dB$$

$$= x_0 \sqrt{\bar{\alpha}_t} + \underbrace{\sum_{u=1}^{t} \left(\prod_{s=u+1}^{t} \sqrt{\alpha_t} \right) \sqrt{\beta_u}\, \varepsilon_u}$$ (坐标稍微对齐一下)

合并高斯噪声

$$u=0. \quad \sigma^2 = \alpha_2 \cdots \alpha_t (1-\alpha_1) + \alpha_3 \cdots \alpha_t (1-\alpha_2) + \cdots + 1 - \alpha_t.$$

$$= \cancel{\alpha_2 \cdots \alpha_t} - \alpha_1 \cdots \alpha_t + \cancel{\alpha_3 \cdots \alpha_t} - \cancel{\alpha_2 \cdots \alpha_t} \;\cancel{+ \cdots - \alpha_t}\; + \cdots + 1 - \cancel{\alpha_t}.$$

$$= 1 - \alpha_1 \cdots \alpha_t = 1 - \bar{\alpha}_t$$

$$= \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \bar{\varepsilon}_t$$ 符合原式.

$dx = -\frac{1}{2}\beta(t) x\, dt + \sqrt{\beta(t)}\, dB \longrightarrow$ 逆向为 $dx = \left[-\frac{1}{2}\beta(t) x - \beta(t) \nabla \log p(x)\right] dt + \sqrt{\beta(t)}\, dB$

$$x_{t-dt} = x_t + \frac{1}{2}\beta(t)\, dt\, x_t + \beta(t)\, dt\, \nabla \log p(x) + \sqrt{\beta(t)}\, dB$$

$$= \left(1 + \frac{1}{2}\beta(t)\, dt\right) x_t + \beta(t)\, dt\, \nabla \log p(x) + \sqrt{\beta(t)}\, dB.$$

$$= \frac{1}{\sqrt{1-\beta(t)\, dt}} \left(x_t + \beta(t)\, dt \cdot \sqrt{1-\beta(t)\, dt}\, \nabla \log p(x) \right] + \text{noise}.$$

$$= \frac{1}{\sqrt{1-\beta(t)\, dt}} \left(x_t + \beta(t)\, dt \left(1 - \frac{1}{2}\beta(t)\, dt\right) \text{Score} \right] + \text{noise}.$$

$$= \frac{1}{\sqrt{1-\beta(t)\, dt}} \left(x_t + \beta(t)\, dt\, \text{Score} \right) + \text{noise}.$$

$$= \frac{1}{\sqrt{\alpha_t}} \left(x_t + \beta_t\, \text{Score} \right) + \sqrt{\beta(t)}\, \varepsilon.$$

Score：$\nabla_{x_t} \log p(x_t)$

$$= \frac{1}{p(x_t)} \nabla p(x_t) = \frac{1}{p(x_t)} \nabla_{x_t} \int p(x_t|x_0) \cdot p(x_0) dx_0$$

$$= \frac{1}{p(x_t)} \int \nabla_{x_t} p(x_t|x_0) \cdot p(x_0) dx_0$$

$$= \frac{1}{p(x_t)} \int [\nabla_{x_t} \log p(x_t|x_0)] \cdot p(x_t|x_0) p(x_0) dx_0$$

$$= \frac{1}{p(x_t)} \int [\nabla_{x_t} \left(\frac{-(x_t-u)^2}{2\sigma^2} \right) p(x_t, x_0) dx_0$$

$$= \frac{1}{p(x_t)} \int -\frac{\varepsilon}{\sigma} p(x_t, x_0) dx_0$$

$$= -\frac{\varepsilon}{\sigma} = -\frac{\bar{\varepsilon}}{\sqrt{1-\bar{\alpha}_t}}$$

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

~~prop~~ ~~pro~~ flo

概率流 ODE：

$dx = f(x,t)dt + g(t)dB$ → 逆向 $dx = [f - g^2 \text{score}]dt + \text{noise}$

F-K 方程

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial (f(x,t)p(x,t))}{\partial x} + \frac{1}{2} g^2 \frac{\partial^2 p(x,t)}{\partial x^2}$$

等价于 $dx = (f(x,t) - \frac{1}{2}(g^2 - \sigma^2) \nabla \log p(x)]dt + \sigma dB$

$\forall\, \sigma(t) < g(t)$

反向：$dx = [f(x,t) - \frac{1}{2}(g^2 + \sigma^2) \nabla \log p(x)]dt + \sigma dB$

$\sigma \to 0$.

得 ODE：$dx = [f(x,t) - \frac{1}{2} g^2 \nabla \log p(x)]dt$

DDIM.

从 DDPM LOSS 看 $L = E_{x_0,\epsilon} \| \epsilon_\theta(\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon) - \epsilon \|^2$

~~变换~~ 简单变换易知

$$L = E_{x_t, x_0} \| \underbrace{s_\theta(x_t)}_{= k\epsilon_\theta} - \underbrace{\nabla \log p(x_t|x_0)}_{= -\frac{\epsilon}{\sigma}} \|^2$$

$$\nabla_{x_t} \log p(x_t) = \frac{1}{p(x_t)} \nabla_{x_t} \int p(x_t|x_0) p(x_0) dx_0$$

$$= \frac{1}{p(x_t)} \int [\nabla_{x_t} p(x_t|x_0)] p(x_0) dx_0$$

$$= \frac{1}{p(x_t)} \int \nabla_{x_t} \log p(x_t|x_0)\, p(x_t, x_0) dx_0$$

$$= E_{x_0|x_t} [\nabla_{x_t} \log p(x_t|x_0)]$$

$$\to L = E_{x_t, x_0} = E_{x_t} E_{x_0|x_0}$$

也就是 DDPM 实际上在学 $\nabla_{x_t} \log p(x_t)$

有 $\nabla_{x_t} \log p(x_t) = -\frac{\epsilon}{\sigma} = -\frac{\epsilon_\theta(x_t)}{\sqrt{1-\bar{\alpha}_t}}$

而这个过程中只需用到 $p(x_t|x_0)$.

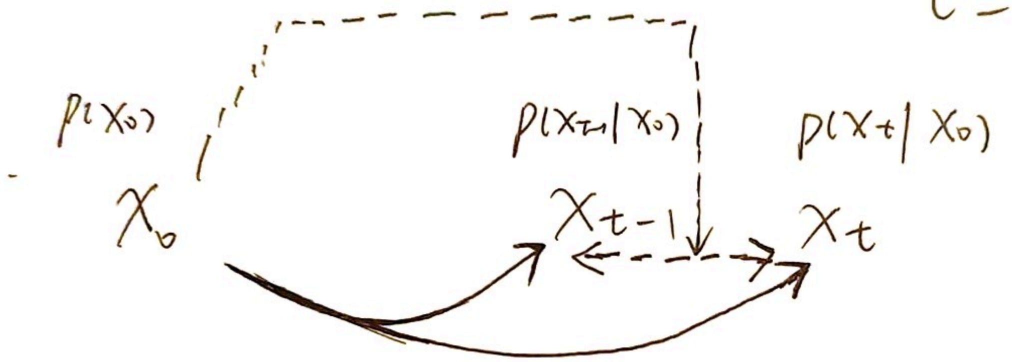
没有用到 $p(x_t|x_{t-1})$ 这个Markov前向

也就是说我们只要保持边缘分布 $p(x_t|x_0)$ 就行而不用管 $p(x_{t-1}|x_t)$ or $p(x_t|x_{t-1})$ 是否 markov.

定义 $\bar{\alpha}_t$, $t=1,\dots,T$

边缘分布 $p(x_t|x_0)=\mathcal{N}(\sqrt{\bar{\alpha}_t}\,x_0,(1-\bar{\alpha}_t)I)$

$t=1,\dots,T$



显然只有边缘分布是没有办法进行采样的

我们仍需要 $x_t \to x_{t-1}$ 的变换

$$\int \underbrace{p(x_{t-1}|x_t,x_0)}\, p(x_t|x_0)\,dx_0 = p(x_{t-1}|x_0)$$

$p(x_{t-1}|x_t,x_0)$ 需满足边缘分布固定.

$$p(x_t|x_0):\quad x_t=\sqrt{\bar{\alpha}_t}\,x_0+\sqrt{1-\bar{\alpha}_t}\,\varepsilon_t$$

$$p(x_{t-1}|x_t,x_0):\quad x_{t-1}=k_1x_t+k_2x_0+\sigma\varepsilon$$

消去 x_t

$$\begin{aligned}&=k_1(\sqrt{\bar{\alpha}_t}\,x_0+\sqrt{1-\bar{\alpha}_t}\,\varepsilon_t)+k_2x_0+\sigma\varepsilon\\&=(k_1\sqrt{\bar{\alpha}_t}+k_2)x_0+k_1\sqrt{1-\bar{\alpha}_t}\,\varepsilon_t+\sigma\varepsilon\end{aligned}$$

$p(x_{t-1}|x_0)$: $x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\, x_0 + \sqrt{1-\bar{\alpha}_{t-1}}\, \varepsilon_{t-1}$

$$\begin{cases} k_1 \sqrt{\bar{\alpha}_t} + k_2 = \sqrt{\bar{\alpha}_{t-1}} \\ \left(k_1 \sqrt{1-\bar{\alpha}_t}\right)^2 + \sigma^2 = \left(\sqrt{1-\bar{\alpha}_{t-1}}\right)^2 \end{cases}$$

$$\Rightarrow \begin{cases} k_1 = \dfrac{\sqrt{1-\bar{\alpha}_{t-1}-\sigma^2}}{\sqrt{1-\bar{\alpha}_t}} \\ k_2 = \sqrt{\bar{\alpha}_{t-1}} - \sqrt{\bar{\alpha}_t}\, \dfrac{\sqrt{1-\bar{\alpha}_{t-1}-\sigma^2}}{\sqrt{1-\bar{\alpha}_t}} \end{cases}$$

$$\Rightarrow p(x_{t-1}|x_t, x_0) = \mathcal{N}(\mu, \sigma I)$$

$$= \mathcal{N}(k_1 x_t + k_2 x_0, \sigma I)$$

$$= \mathcal{N}\left(\sqrt{\bar{\alpha}_{t-1}}\, x_0 + \sqrt{1-\bar{\alpha}_{t-1}-\sigma^2}\, \frac{x_t - \sqrt{\bar{\alpha}_t}\, x_0}{\sqrt{1-\bar{\alpha}_t}}, \sigma I\right)$$

$$\cancel{x_0 = \frac{x_t}{\sqrt{1-\bar{\alpha}_t}}} \qquad \cancel{x_0 = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\, \epsilon_\theta}{\sqrt{\bar{\alpha}_t}}}$$

$$x_0 = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\, \epsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}}$$

$$\Rightarrow x_{t-1} = \underbrace{\sqrt{\bar{\alpha}_{t-1}}\, \frac{x_t - \sqrt{1-\bar{\alpha}_t}\, \epsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}}}_{\text{predict } x_0} + \underbrace{\sqrt{1-\bar{\alpha}_{t-1}-\sigma^2}\, \epsilon_\theta(x_t, t)}_{\text{point to } x_t} + \sigma \epsilon$$

加速采样：

在推导边际分布的过程中，没有用到 $x_t \to x_{t-1}$。

~~$p(x_t|x_0)=\sqrt{\bar\alpha_t}\,x_0+\sqrt{1-}$~~

$p(x_t|x_0)$：$x_t=\sqrt{\bar\alpha_t}\,x_0+\sqrt{1-\bar\alpha_t}\,\epsilon_t$

考虑 $\tau_1<\tau_2$。

~~$p(\tau_1|\tau_1$~~

$$\int p(x_{\tau_1}|x_{\tau_2},x_0)\,p(x_{\tau_2}|x_0)\,dx_0=p(x_{\tau_1}|x_0)$$

$$\Rightarrow x_{\tau_1}=\sqrt{\bar\alpha_{\tau_1}}\,x_0+\sqrt{1-\bar\alpha_{\tau_1}-\sigma^2}\,\frac{x_{\tau_2}-\sqrt{\bar\alpha_{\tau_2}}\,x_0}{\sqrt{1-\bar\alpha_{\tau_2}}}+\sigma\epsilon$$

$$=\sqrt{\bar\alpha_{\tau_1}}\,\frac{x_{\tau_2}-\sqrt{1-\bar\alpha_{\tau_2}}\,\epsilon_\theta(x_{\tau_2},\tau_2)}{\sqrt{\bar\alpha_{\tau_2}}}+\sqrt{1-\bar\alpha_{\tau_1}-\sigma^2}\,\epsilon_\theta(x_{\tau_2},\tau_2)+\sigma\epsilon$$

即可选 $(1,2,\dots,T)$ 的子序列 $(\tau_1,\tau_2,\dots,\tau_S)$ 进行采样。

考虑 $\sigma_t=0$，即

$$x_{t-dt}=\sqrt{\bar\alpha_{t-dt}}\,\frac{x_t-\sqrt{1-\bar\alpha_t}\,\epsilon_\theta}{\sqrt{\bar\alpha_t}}+\sqrt{1-\bar\alpha_{t-dt}}\,\epsilon_\theta.$$

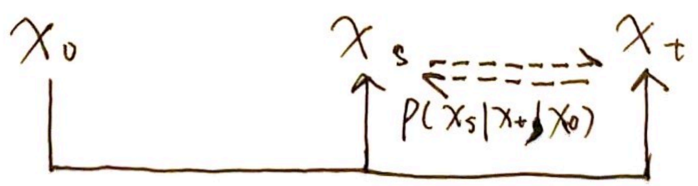
整理有

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

$$\Rightarrow d\left(\frac{x_t}{\sqrt{\bar\alpha_t}}\right) = \epsilon_\theta \, d\left(\sqrt{\frac{1-\bar\alpha_t}{\bar\alpha_t}}\right)$$

即 DDIM 在 $\sigma=0$ 时变成 ODE

OK. 接着DDIM的思路，我们看一种~~特殊~~一般情况：$p(x_t|x_0)$: $x_t = \alpha(t)x_0 + \sigma(t)\varepsilon_t$



$$\int p(x_s|x_t, x_0)\, p(x_t|x_0)\, dx_0 = p(x_s|x_0)$$

$$\to x_s = k_1 x_t + k_2 x_0 + \cancel{\sigma\varepsilon} + \cancel{\gamma\varepsilon}$$
$$+ \gamma\varepsilon$$

$$= k_1(\alpha(t)x_0 + \sigma(t)\varepsilon_t) + k_2 x_0 + \gamma\varepsilon$$

$$= (k_1\alpha(t) + k_2)x_0 + k_1\sigma(t)\varepsilon_t + \gamma\varepsilon$$

$p(x_s|x_0): \quad x_s = \alpha(s) x_0 + \sigma(s) \varepsilon,$

$$\Rightarrow \begin{cases} k_1 \alpha(t) + k_2 = \alpha(s) \\ [k_1 \sigma(t)]^2 + r^2 = \sigma^2(s) \end{cases}$$

$$\Rightarrow \begin{cases} k_1 = \dfrac{\sqrt{\sigma^2(s) - r^2}}{\sigma(t)} \\ k_2 = \alpha(s) - \dfrac{\alpha(t)}{\sigma(t)} \sqrt{\sigma^2(s) - r^2} \end{cases}$$

$p(x_s|x_t, x_0): \quad x_s = k_1 x_t + k_2 x_0 + r\varepsilon$

$$= \alpha(s) x_0 + \sqrt{\sigma^2(s) - r^2} \, \frac{x_t - \alpha(t) x_0}{\sigma(t)} + r\varepsilon$$

令 $\cancel{x_0^\theta}$ $x_0^\theta(x_t, t) = \dfrac{x_t - \sigma(t) \epsilon_\theta(x_t, t)}{\alpha(t)}$

$p_\theta(x_s|x_t) = p(x_s|x_t, x_0^\theta(x_t, t))$

$$KL(p(x_s|x_t, x_0) \| p_\theta(x_s|x_t)) = KL(p(x_s|x_t, x_0) \| p(x_s|x_t, x_0^\theta))$$

$$= \frac{k_2^2 \| x_0 - x_0^\theta(x_t, t) \|}{2r^2}$$

代入 x_0, 和 x_0^θ

$$E_{x_0, x_t}(KL) = E_{x_0, x_t} \frac{k^2 \| x_0 - x_0^\theta(x_t, t) \|}{2r^2} = E_{x_0, \epsilon}\left[\frac{k^2 \sigma^2(t) \| \epsilon - \epsilon_\theta \|^2}{2r^2 \alpha^2(t)} \right]$$

也就说 $p(x_t|x_0): \quad x_t = \alpha(t) x_0 + \sigma(t) \varepsilon$

不用保证 $\alpha^2(t) + \sigma^2(t)$ 也能用 Denosing Score Matching 训一个 DDPM(DDIM)

~~在~~再看看 SDE

$$dx = f(t)\,x\,dt + g(t)\,dB$$

解得 $x(t) = x_0 \exp\left(\int_0^t f(s)\,ds\right) + \int_0^t \exp\left(\int_u^t f(s)\,ds\right) g(t)\,dB$

而我们有 $x(t) = \alpha(t)\,x_0 + \sigma(t)\,\varepsilon$

$$\Rightarrow \alpha(t) = \exp\left(\int_0^t f(s)\,ds\right) \quad \Rightarrow f(t) = \frac{d\ln\alpha(t)}{dt}$$

~~PS~~ (PS: 如果有~~数~~ $f(s) = -\frac{1}{2}\beta(s)$, 然后离散化+taylor.
可有 ~~$\alpha(t) = \sqrt{\ }$~~
$\alpha(t) = \prod \sqrt{1-\beta_i} = \sqrt{\bar{\alpha}_t}$)

$$\sigma^2(t) = \int_0^t \left[\frac{\alpha(t)}{\alpha(u)}\right]^2 g^2(u)\,du$$

$$\frac{d\sigma^2(t)}{dt} = \frac{d\int_0^t \left(\frac{\alpha(t)}{\alpha_u}\right)^2 g^2(t)\,dt}{dt} = \frac{d\left[\alpha^2(t)\int_0^t \left[\frac{g(u)}{\alpha(u)}\right]^2 du\right]}{dt}$$

$$= 2\alpha(t)\frac{d\alpha(t)}{dt}\int_0^t \left[\frac{g(u)}{\alpha(u)}\right]^2 du + \alpha^2(t)\left(\frac{g(t)}{\alpha(t)}\right)^2$$

$$= 2\frac{1}{\alpha(t)}\frac{d\alpha(t)}{dt}\int_0^t \alpha^2(t)\frac{g^2(u)}{\alpha^2(u)}\,du + g^2(t)$$

$$= 2\frac{d\ln\alpha(t)}{dt}\sigma^2(t) + g^2(t)$$

$$\Rightarrow \begin{cases} f(t) = \dfrac{d\ln\alpha(t)}{dt} \\ g^2(t) = \dfrac{d\sigma^2(t)}{dt} - 2\dfrac{d\ln\alpha(t)}{dt}\sigma^2(t) \end{cases}$$

我们已知 $dx = f(x,t)dt + g(t)dB$

的逆向 ~~为~~ SDE 为

$$dx = [f(x,t) - g^2(t)\,\text{Score}]dt + g(t)dB$$

逆向 probability flow SDE 为

$$dx = [f(x,t) - \frac{1}{2}g^2(t)\,\text{Score}]dt$$

由 $\begin{cases} f(x,t) = x \cdot \dfrac{d\ln\alpha(t)}{dt} \\ g^2(t) = \dfrac{d\sigma^2(t)}{dt} - 2\dfrac{d\ln\alpha(t)}{dt}\sigma^2(t) \end{cases}$

$$dx = [x \cdot \frac{d\ln\alpha(t)}{dt} - \frac{1}{2}(\frac{d\sigma^2(t)}{dt} - 2\frac{d\ln\alpha(t)}{dt}\sigma^2(t))\,\text{Score}]dt$$

$$\frac{dx}{dt} - x \cdot \frac{d\ln\alpha(t)}{d(t)} = -\frac{1}{2}[\frac{d\sigma^2(t)}{dt} - 2\frac{d\ln\alpha(t)}{dt}\sigma^2(t)]\,\text{score}$$

先看左边 $\dfrac{dx}{dt} - x \cdot \dfrac{d\ln\alpha(t)}{dt} = \dfrac{dx}{dt} - x \cdot \dfrac{1}{\alpha(t)}\dfrac{d\alpha(t)}{dt}$

$$= \alpha(t)[\frac{d(x)}{dt} \cdot \frac{1}{\alpha(t)} + x \cdot (-\frac{1}{\alpha^2(t)}) \cdot \frac{d\alpha(t)}{dt})$$

$$= \alpha(t) \cdot \frac{d(\frac{x}{\alpha(t)})}{dt}$$

再看右边 $-\dfrac{1}{2}[\dfrac{d\sigma^2(t)}{dt} - 2\dfrac{d\ln\alpha(t)}{dt}\sigma^2(t)]\,\text{Score}$

$$= -\frac{1}{2}[2\sigma(t)\frac{d\sigma(t)}{dt} - 2\frac{1}{\alpha(t)}\frac{d\alpha(t)}{dt}\sigma^2(t)]\,\text{Score}$$

$$= -\alpha(t)\sigma(t)\left[\frac{d\sigma(t)}{dt}\cdot\frac{1}{\alpha(t)} + \sigma(t)\cdot\left(-\frac{1}{\alpha^2(t)}\right)\frac{d\alpha(t)}{dt}\right]\text{score}$$

$$= -\alpha(t)\sigma(t)\frac{d\left[\frac{\sigma(t)}{\alpha(t)}\right]}{dt}\text{Score}$$

$$\Rightarrow \quad \alpha(t)\cdot d\left[\frac{x}{\alpha(t)}\right] = -\alpha(t)\sigma(t)\, d\left[\frac{\sigma(t)}{\alpha(t)}\right]\text{score}.$$

$$\text{Score} = \nabla_{x_t}\log p(x_t) = -\frac{\epsilon}{\sigma} = -\frac{\epsilon_\theta(x_t,t)}{\sigma(t)}$$

$$\Rightarrow \quad d\left[\frac{x}{\alpha(t)}\right] = \epsilon_\theta\, d\left[\frac{\sigma(t)}{\alpha(t)}\right]$$

再看 $\cancel{x_s = \sqrt{\alpha_s}\frac{x_t}{\sqrt{\alpha_t}}}$

$$x_s = \alpha(s)x_0 + \sqrt{\sigma^2(s)-\gamma^2}\,\frac{x_t-\alpha(t)x_0}{\sigma(t)} + \gamma\epsilon$$

令 $\gamma=0$，并代入 $x_0 = \frac{x_t-\sigma(t)\epsilon_\theta}{\alpha(t)}$

$$x_s = \alpha(s)\frac{x_t-\sigma(t)\epsilon_\theta}{\alpha(t)} + \cancel{\frac{\alpha(s)}{\sigma(t)}\epsilon_\theta}\ \sigma(s)\epsilon_\theta$$

$$\Rightarrow \quad \frac{x_s}{\alpha(s)} - \frac{x_t}{\alpha(t)} = \left[\frac{\sigma(s)}{\alpha(s)} - \frac{\sigma(t)}{\alpha(t)}\right]\epsilon_\theta$$

$$\Rightarrow \quad d\left[\frac{x}{\alpha(t)}\right] = \epsilon_\theta\, d\left[\frac{\sigma(t)}{\alpha(t)}\right]$$

如果令 $\alpha^2(t)+\sigma^2(t)=1$ ~~$\Rightarrow$ DDPM 以及~~ ~~ODE~~

可以看到 DDIM 的 ODE 就是 DDPM 的 ODE

当然再迈~~近~~进一步，

令 $z_t = \frac{x}{\alpha(t)} \Rightarrow z_t \sim \mathcal{N}\left(z; \frac{x_0}{\alpha(0)}, \left[\frac{\sigma(t)}{\alpha(t)}\right]^2 I\right)$

$$= \mathcal{N}(z; z_0, \sigma_z^2(t) I)$$

保持 $\sigma_z(t)$ 单调递增.

我们就得到了一个 ~~VP-SDE~~ VE-SDE.

回到了NCSN的情况.