💯

DPM-Solver

参考文献：80岁的公园大爷问我 DPMsolver 是什么然后我给他写了这个他说哦我看懂了 - 知乎

Diffusion的SDE表示

假设目标分布为$q_0(x_0)$，定义前向随机过程$\{x_t\}_{t\in[0,T]}$，随着时间增加，x_t趋于一个均值为0的正态分布$q_T(x_T)=\mathcal{N}(0,\sigma_{\max}^2\mathbf{I})$

$$x_t=\alpha_t x_0+\sigma_t\epsilon_t,\epsilon_t\sim\mathcal{N}(0,\mathbf{I})$$

$\frac{\alpha_t}{\sigma_t}$直观上表示信噪比，且

$$\alpha_t\to 0,\sigma_t\to\sigma_{\max}$$

该过程可用如下SDE表示

$$\mathrm{d}x_t=f(t)x_t\mathrm{d}t+g(t)\mathrm{d}w_t,x_0\sim q_0(x_0) \tag{1}$$

其中w_t为标准Winner过程

Winner过程$W(t)$：$W(t+\Delta t)-W(t)\sim\mathcal{N}(0,\Delta t\mathbf{I})$，因此有

$$\mathrm{d}w_t=\sqrt{\mathrm{d}t}\epsilon,\epsilon\sim\mathcal{N}(0,\mathbf{I})$$

且

$$f(t)=\frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t}, g^2(t)=\frac{\mathrm{d}\sigma_t^2}{\mathrm{d}t}-2\frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t}\sigma_t^2$$

为了获取服从目标分布的样本，也就是$\mathcal{N}(0,\mathbf{I})\to q_0(x_0)$，我们需要求解(1)的逆过程，可以证明，此逆过程在一定条件下亦为SDE

$$\mathrm{d}x_t=[f(t)x_t-g^2(t)\nabla_x\log q_t(x_t)]\mathrm{d}t+g(t)\mathrm{d}\bar{w}_t, x_T\sim q_T(x_T) \tag{2}$$

其中$\bar{w}_t$亦为标准维纳过程

其中唯一的未知项为$\nabla_x\log q_t(x_t)$，称为score function

Denoising Score Matching(学习Score Function)

假设我们直接用NN $s_\theta(x_t,t)$来拟合score function

$$\mathcal{L}_{\mathrm{SM}}(\theta)=\mathbb{E}_{q_t(x_t)}\left[\|s_\theta(x_t,t)-\nabla_x\log q_t(x_t)\|^2\right] \tag{3}$$

称为Score Matching Loss

由于边际分布$q_t(x_t)$非常难求，因此上述方法并不是很可行，但好在可以证明，最小化(3)在一定条件下等价于最小化下式

$$\mathcal{L}_{\mathrm{DSM}}(\theta)=\mathbb{E}_{q_{0t}(x_t|x_0)}\left[\|s_\theta(x_t,t)-\nabla_x\log q_{0t}(x_t|x_0)\|^2\right] \tag{4}$$

其中

$$\begin{aligned}\nabla_x\log q_{0t}(x_t|x_0)&=\nabla_x\log\left[C\exp\{-\frac{(x_t-\alpha_t x_0)^2}{2\sigma_t^2}\}\right]\\&=-\frac{1}{\sigma_t^2}(x_t-\alpha_t x_0)\\&=-\frac{\epsilon_t}{\sigma_t}\end{aligned}$$

代入(4)可得

$$\mathcal{L}_{\mathrm{DSM}}(\theta)=\mathbb{E}_{x_0\sim q_0(x_0),\epsilon_t\sim\mathcal{N}(0,\mathbf{I})}\left[\|s_\theta(x_t,t)+\frac{\epsilon_t}{\sigma_t}\right] \tag{5}$$

称为Denoising Score Matching Loss

实践中，我们用另一种形式 $\epsilon_\theta = -\sigma_t s_\theta \approx -\sigma_t \nabla_x \log q_t(x_t)$，也就是学噪声，对应有下面的Loss

$$\mathcal{L}_{\text{DSM}}(\theta) = \mathbb{E}_{x_0 \sim q_0(x_0), \epsilon_t \sim \mathcal{N}(0,\mathbf{I}), t \in [0,T]} \left[|| \epsilon_\theta(x_t, t) - \epsilon_t \right] \tag{6}$$

通过最小化上式得到ϵ_θ后，就可以得到Diffusion的逆向SDE

$$\mathrm{d}x_t = [f(t)x_t + \frac{g^2(t)}{\sigma_t}\epsilon_\theta(x_t, t)]\mathrm{d}t + g(t)\mathrm{d}\bar{w}_t, x_T \sim q_T(x_T) \tag{7}$$

于是，我们只需要$x_T \sim q_T(x_T)$，然后逆向求解(7)即可得到$x_0 \sim q_0(x_0)$

SDE—>ODE

对(7)的求解涉及到离散化，一旦步长变大，维纳过程$\mathrm{d}\bar{w}_t$的随机性会让误差急剧增大，导致无法收敛(因此DDPM采用1000/4000步采样)，好在(7)等价于一个ODE

$$\begin{aligned} \frac{\mathrm{d}x_t}{\mathrm{d}t} &= f(t)x_t - \frac{1}{2}g^2(t)\nabla_x \log q_t(x_t) \\ &= f(t)x_t + \frac{g^2(t)}{2\sigma_t}\epsilon_\theta(x_t, t) \end{aligned} \tag{8}$$

ODE的数值求解已有广泛研究，现在只需要10步即可得到逆向SDE的数值解

DPM-Solver

下面来求解(8)，令$u_\tau = e^{-\int_\infty^\tau f(r)\mathrm{d}r}$，记$A_\tau = \frac{g^2(\tau)}{2\sigma_\tau}$，于是

$$
\begin{aligned}
\dot{x}_\tau - f(\tau)x_\tau &= A_\tau \epsilon_\theta(x_\tau, \tau) && (9)\\
u_\tau \dot{x}_\tau - u_\tau f(\tau)x_\tau &= u_\tau A_\tau \epsilon_\theta(x_\tau, \tau) && (10)\\
\frac{\mathrm{d}(u_\tau x_\tau)}{\mathrm{d}\tau} &= u_\tau A_\tau \epsilon_\theta(x_\tau, \tau) && (11)\\
u_t x_t - u_s x_s &= \int_s^t u_\tau A_\tau \epsilon_\theta(x_\tau, \tau)\mathrm{d}\tau && (12)\\
x_t &= \frac{u_s}{u_t} x_s + \int_s^t \frac{u_\tau}{u_t} A_\tau \epsilon_\theta(x_\tau, \tau)\mathrm{d}\tau && (13)\\
x_t &= e^{\int_s^t f(r)\mathrm{d}r} x_s + \int_s^t e^{\int_\tau^t f(r)\mathrm{d}r} \frac{g^2(\tau)}{2\sigma_\tau} \epsilon_\theta(x_\tau, \tau)\mathrm{d}\tau && (14)
\end{aligned}
$$

令$\lambda_t = \log \frac{\alpha_t}{\sigma_t}$为log-SNR，则有

$$
\exp \int_s^t f(r)\mathrm{d}r = \exp \log \frac{\alpha_t}{\alpha_s} = \frac{\alpha_t}{\alpha_s}
$$

进而

$$
e^{\int_\tau^t f(r)\mathrm{d}r} \frac{g^2(\tau)}{2\sigma_\tau} = \frac{\alpha_t}{\alpha_\tau} \sigma_\tau \frac{\mathrm{d}(\log \sigma_\tau - \log \alpha_\tau)}{\mathrm{d}\tau} = -\alpha_t \frac{\sigma_\tau}{\alpha_\tau} \frac{\mathrm{d}\lambda_\tau}{\mathrm{d}\tau}
$$

又考虑到λ_t与t一一对应(注意，下面将不再区分λ和t)，于是，(14)化为

$$
\begin{aligned}
x_t &= \frac{\alpha_t}{\alpha_s} x_s - \alpha_t \int_s^t \frac{\sigma_\tau}{\alpha_\tau} \frac{\mathrm{d}\lambda_\tau}{\mathrm{d}\tau} \epsilon_\theta(x_\tau, \tau)\mathrm{d}\tau && (15)\\
&= \frac{\alpha_t}{\alpha_s} x_s - \alpha_t \int_{\lambda_s}^{\lambda_t} e^{-\lambda} \epsilon_\theta(x_\lambda, \lambda)\mathrm{d}\lambda && (16)
\end{aligned}
$$

后面一项需要依靠泰勒展开进行估计

$$
\epsilon_\theta(x_\lambda, \lambda) = \sum_{i=0}^{k-1} \frac{\epsilon_\theta^{(i)}(x_s, s)}{i!} (\lambda - \lambda_s)^i + \mathcal{O}(|\lambda - \lambda_s|^k)
$$

因此

$$
x_t = \frac{\alpha_t}{\alpha_s} x_s - \alpha_t \sum_{i=0}^{k-1} \int_{\lambda_s}^{\lambda_t} e^{-\lambda} \frac{\epsilon_\theta^{(i)}(x_s, s)}{i!} (\lambda - \lambda_s)^i \mathrm{d}\lambda + \mathcal{O}(h^{k+1})
$$

k=1(零阶近似, DDIM)

$$
\begin{aligned}
x_t &= \frac{\alpha_t}{\alpha_s} x_s - \alpha_t \int_{\lambda_s}^{\lambda_t} e^{-\lambda} \epsilon_\theta(x_\lambda, \lambda) \mathrm{d}\lambda && (17) \\
&= \frac{\alpha_t}{\alpha_s} x_s + \alpha_t (e^{-\lambda_t} - e^{-\lambda_s}) \epsilon_\theta(x_s, s) && (18) \\
&= \frac{\alpha_t}{\alpha_s} x_s - \alpha_t (\frac{\sigma_s}{\alpha_s} - \frac{\sigma_t}{\alpha_t}) \epsilon_\theta(x_s, s) && (19)
\end{aligned}
$$

即

$$
x_t = \frac{\alpha_t}{\alpha_s} x_s - \alpha_t (\frac{\sigma_s}{\alpha_s} - \frac{\sigma_t}{\alpha_t}) \epsilon_\theta(x_s, s)
$$

此即DDIM的采样过程！！！

k=2(一阶近似)

此时涉及到求ϵ_θ的一阶导数，用两个forward过程来近似

$$
\begin{aligned}
&\epsilon_\theta(x_{\lambda_s + rh}, \lambda_s + rh) = \epsilon_\theta(x_{s_1}, s_1) \\
&\epsilon_\theta(x_{\lambda_s}, \lambda_s) = \epsilon_\theta(x_s, s)
\end{aligned}
$$

其中$s_1 = t_\lambda(\lambda_s + rh)$，$t_\lambda$为$\lambda$到$t$的映射，$x_{s_1}$可通过下式估计

$$
\begin{aligned}
x_{s_1} &\approx \frac{\alpha_{s_1}}{\alpha_s} x_s + \alpha_{s_1} (e^{-(\lambda_s + rh)} - e^{-\lambda_s}) \epsilon_\theta(x_s, s) \\
&= \frac{\alpha_{s_1}}{\alpha_s} x_s + \alpha_{s_1} e^{-(\lambda_s + rh)} (1 - e^{rh}) \epsilon_\theta(x_s, s) \\
&= \frac{\alpha_{s_1}}{\alpha_s} x_s + \sigma_{s_1} (1 - e^{rh}) \epsilon_\theta(x_s, s) := \hat{x}_{s_1}
\end{aligned}
$$

于是，采样过程的一阶估计为

$$
\bar{x}_t = \frac{\alpha_t}{\alpha_s} x_s - \sigma_t (e^h - 1) \epsilon_\theta(x_s, s) - \frac{\sigma_t}{2r} (e^h - 1) \left[\epsilon_\theta(\hat{x}_{s_1}, s_1) - \epsilon_\theta(x_s, s) \right]
$$

可以证明

$$x_t - \bar{x}_t = \mathcal{O}(h^3)$$

Implementation Details

Noise Schedule

即α_t, σ_t的设计方案，考虑到在DDPM中，两者分别为

$$\begin{aligned}\alpha_t &= \sqrt{\bar{\alpha}_t} \\ \sigma_t &= \sqrt{1-\bar{\alpha}_t}\end{aligned}$$

也就是满足$\sigma_t = \sqrt{1-\alpha_t^2}$，本文延用，因此只需设计$\alpha_t$即可

- **Linear Noise Schedule**

  DDPM中被提出，要求β_t随着时间线性增加，在DPM-Solver中对应

  $$\alpha_t = \prod_{i=0}^{t}(1-\beta_t)$$

  连续化

  $$\begin{aligned}\alpha_{t+1} &= \alpha_t(1-\beta_{t+1}) \\ \alpha_{t+\Delta t} &= \alpha_t(1-\Delta t\beta_{t+\Delta t}) \\ \alpha_t + \Delta\alpha_t &= \alpha_t - (\beta_t + \Delta\beta_t)\alpha_t\Delta t \\ \frac{1}{\alpha_t}\frac{\mathrm{d}\alpha_t}{\mathrm{d}t} &= -\beta_t - \mathrm{d}\beta_t \\ \mathrm{d}\log\alpha_t &= -\beta_t\mathrm{d}t \\ \log\alpha_t &= -\frac{\beta_1-\beta_0}{4}t^2 - \frac{\beta_0}{2}t\end{aligned}$$

  其中$\beta_0 = 0.1, \beta_1 = 20$为超参数

- **Cosine Noise Schedule**

  $$\log\alpha_t = \log\left[\cos\left(\frac{t/T+s}{1+s}\cdot\frac{\pi}{2}\right)\right] - \log\left[\cos\left(\frac{s}{1+s}\cdot\frac{\pi}{2}\right)\right]$$

Sampling Timestep Schedule

以往方法对t均匀采样

DPM-Solver对

λ均匀采样

$$\lambda_i = \lambda_T + \frac{i}{N}(\lambda_0 - \lambda_T)$$

DPM-Solver++

从估计噪声ϵ_θ转为估计数据x_θ

$$x_\theta = \frac{x_t - \sigma_t \epsilon_\theta}{\alpha_t}$$

代入(8)得到

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \left[f(t) + \frac{g^2(t)}{2\sigma_t^2}\right] x_t - \frac{\alpha_t g^2(t)}{2\sigma_t^2} x_\theta(x_t, t)$$

由常数变异法求得

$$x_t = \frac{\sigma_t}{\sigma_s} x_s + \sigma_t \int_{\lambda_s}^{\lambda_t} e^{\lambda} x_\theta(x_\lambda, \lambda) \, \mathrm{d}\lambda$$

后面同样通过Taylor展开得到高阶解

$$x_t = \frac{\sigma_t}{\sigma_s} e^{-h} x_s + \alpha_t (1 - e^{-2h}) x_\theta(x_s, s) + \sigma_t \sqrt{1 - e^{-2h}} z_s$$