Score-Based Generative Modeling through Stochastic Differential Equations (ICLR'21)

统一了 Score-Based 和 DDPM 的理论

SDE 本质还是 Score-Based

SDE 1

① 为什么要用SDE来描述扩散模型?

② 用SDE描述扩散模型的扩散公式是什么?

① X_t { t固定 → 随机变量 $X_t \sim N(\sqrt{\bar{\alpha}_t}\, X_0, (1-\bar{\alpha}_t) I)$

X固定 → $X_T \; X_{T-1} \; X_{T-2} \cdots X_1 \, X_0$ 轨迹(采样) [应用随机过程]

↓

随机过程 → SDE

② 扩散过程既可以离散的看,又可以连续的看

离散 $X_0 \to X_1 \to X_2 \to \cdots \to X_T$

$X_T \to X_{T-1} \to \cdots \to X_0$ $t \in \{0, 1, 2, \cdots, T\}$

连续 $t \in [0, 1]$ $X_t \to X_{t+\Delta t}$ $X_{t+\Delta t} \to X_t$

上下界不重要 $\Delta t \to 0$ $\Delta t \to 0$

人为规定

③ SDE-based diffusion process

$$\underbrace{dx}_{X_t} = \overbrace{\boxed{f(x,t)\,dt}}^{\text{确定性}} + \overbrace{\boxed{g(t)\,dw}}^{\text{不确定性}}$$

研究对象

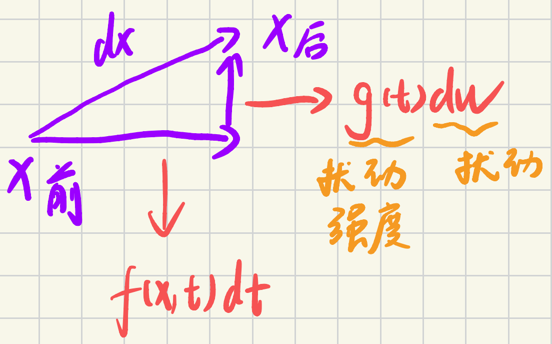
噪声图

$f(x,t)$: drift coefficient 漂移系数

$g(t)$: diffusion coefficient

w: Brownian motion 布朗运动

扩散模型的根源是热力学状态



SDE2

如何推导基于SDE的扩散模型的重建公式

扩散公式：连续：$dx = f(x,t)dt + g(t)dw$ $x_t \to x_{t+\Delta t}$

为什么一阶量变成了阶数更小的量？

记住即可，结论性的

离散：$x_{t+\Delta t} - x_t = f(x_t,t)\Delta t + g(t)\sqrt{\Delta t}\,\varepsilon$ $\varepsilon \sim N(0,1)$

$$x_{t+\Delta t} = x_t + f(x_t,t)\Delta t + g(t)\sqrt{\Delta t}\,\varepsilon$$

$$p(x_{t+\Delta t}|x_t) \sim \mathcal{N}(x_t + f(x_t,t)\Delta t,\ g^2(t)\Delta t\, I)$$

用扩散公式 $\Rightarrow$ 重建公式

$x_{t+\Delta t} \to x_t$ $p(x_t|x_{t+\Delta t}) = \frac{p(x_{t+\Delta t}|x_t)\,p(x_t)}{p(x_{t+\Delta t})}$

$$= p(x_{t+\Delta t}|x_t)\exp\{\log p(x_t) - \log p(x_{t+\Delta t})\}$$

为了构造score

$$\boxed{\log p(x_{t+\Delta t}) \overset{\text{泰勒展开}}{\approx} \log p(x_t) + (x_{t+\Delta t}-x_t)\nabla_{x_t}\log p(x_t) + \Delta t\frac{\partial}{\partial t}\log p(x_t)}$$

$$= p(x_{t+\Delta t}|x_t)\exp\{-(x_{t+\Delta t}-x_t)\nabla_{x_t}\log p(x_t) - \Delta t\frac{\partial}{\partial t}\log p(x_t)\}$$

$$\propto \exp\left\{-\frac{\|x_{t+\Delta t}-x_t-f(x_t,t)\Delta t\|_2^2}{2g^2(t)\Delta t} - (x_{t+\Delta t}-x_t)\nabla_{x_t}\log p(x_t) - \Delta t\frac{\partial}{\partial t}\log p(x_t)\right\}$$

$$= \exp\left\{-\frac{1}{2g^2(t)\Delta t}\left((x_{t+\Delta t}-x_t)^2 - (2f(x_t,t)\Delta t - 2g^2(t)\Delta t\nabla_{x_t}\log p(x_t))(x_{t+\Delta t}-x_t)\right) - \Delta t\frac{\partial}{\partial t}\log p(x_t) - \frac{f^2(x_t,t)\Delta t}{2g^2(t)}\right\}$$

$$= \exp\left\{-\frac{1}{2g^2(t)\Delta t}\|(x_{t+\Delta t}-x_t) - (f(x_t,t) - g^2(t)\nabla_{x_t}\log p(x_t))\Delta t\|_2^2 - \Delta t\frac{\partial}{\partial t}\log p(x_t) - \frac{f^2(x_t,t)\Delta t}{2g^2(t)} + \frac{(f(x_t,t)-g^2(t)\nabla_{x_t}\log p(x_t))^2\Delta t}{2g^2(t)}\right\}$$

$$\underset{x_{t+\Delta t}\to x_t}{\overset{\Delta t\to 0}{\approx}} \exp\left\{-\frac{1}{2g^2(t+\Delta t)\Delta t}\|(x_{t+\Delta t}-x_t) - (f(x_{t+\Delta t},t+\Delta t) - g^2(t+\Delta t)\nabla_{x_{t+\Delta t}}\log p(x_{t+\Delta t}))\Delta t\|_2^2\right\}$$

$$x_t = x_{t+\Delta t} - (f(x_{t+\Delta t},t+\Delta t) - g^2(t+\Delta t)\nabla_{x_{t+\Delta t}}\log p(x_{t+\Delta t})\Delta t)$$

$p(x_t|x_{t+\Delta t})$ 均值: $x_{t+\Delta t}-\left(f(x_{t+\Delta t},t+\Delta t)-g^2(t+\Delta t)\nabla_{t+\Delta t}\log p(x_{t+\Delta t})\right)\Delta t$

方差: $g^2(t+\Delta t)\Delta t$

$$dx=\left[f(x,t)-g^2(t)\nabla_{x_t}\log p(x_t)\right]dt+g(t)\,dw$$

采样

$$x_{t+\Delta t}-x_t=f(x_{t+\Delta t},t+\Delta t)-g^2(t+\Delta t)\nabla_{x_{t+\Delta t}}\log p(x_{t+\Delta t})]\Delta t+g(t+\Delta t)\sqrt{\Delta t}\,\varepsilon$$

score 部分并不清楚

$$x_{t-1}=x_t-\left[f(x_t,t)-g^2(t)\underbrace{\nabla_{x_t}\log p(x_t)}_{\text{score}}\right]+g(t)\varepsilon$$

扩散: $dx=f(x,t)dt+g(t)dw$

重建: $dx=\left[f(x,t)-g^2(t)\nabla_x\log p(x)\right]dt+g(t)dw$

SDE3

VE-SDE ⇔ NCSN

(Variance Exploding)

$$x_t=x_0+\sigma_t\varepsilon$$

$$x_{t+1}=x_t+\sqrt{\sigma_{t+1}^2-\sigma_t^2}\,\varepsilon$$

$$x_T=x_0+\sigma_T\varepsilon$$

大

VP-SDE ⇔ DDPM

(Variance Preserving)

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

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

$$x_T=\sqrt{\bar\alpha_T}\,x_0+\sqrt{1-\bar\alpha_T}\,\varepsilon$$

$\sqrt{\bar\alpha_T}\approx 0$ $\sqrt{1-\bar\alpha_T}\approx 1$

$dx=f(x,t)dt+g(t)dw$ $\qquad x_t\to x_{t+\Delta t}$

$$\boxed{x_{t+\Delta t}=x_t+f(x_t,t)\Delta t+g(t)\sqrt{\Delta t}\,\varepsilon}$$

$$\boxed{X_{t+\Delta t} = X_t + f(X_t, t)\Delta t + g(t)\sqrt{\Delta t}\,\varepsilon}$$

VE

$$X_{t+\Delta t} = X_t + \sqrt{\sigma_{t+\Delta t}^2 - \sigma_t^2}\;\varepsilon$$

$$= X_t + \sqrt{\frac{\sigma_{t+\Delta t}^2 - \sigma_t^2}{\Delta t}} \cdot \sqrt{\Delta t}\,\varepsilon$$

$$= X_t + \sqrt{\frac{\Delta \sigma_t^2}{\Delta t}}\,\sqrt{\Delta t}\,\varepsilon$$

$$f(X_t, t) = 0 \qquad g(t) = \frac{d}{dt}\sigma_t^2$$

VP

$$X_{t+1} = \sqrt{1-\beta_{t+1}}\,X_t + \sqrt{\beta_{t+1}}\,\varepsilon$$

$$\{\beta\}_{i=1}^{T} \quad \{\bar{\beta}_i = T\beta_i\}_{i=1}^{T}$$

$$X_{t+1} = \sqrt{1-\frac{\bar{\beta}_{t+1}}{T}}\,X_t + \sqrt{\frac{\bar{\beta}_{t+1}}{T}}\,\varepsilon$$

$$T\to\infty \quad \{\bar{\beta}_i\}_{i=1}^{T} \to \beta(t) \quad t\in[0,1]$$

$$\beta\left(\frac{i}{T}\right) = \bar{\beta}_i \qquad \Delta t = \frac{1}{T}$$

$$X_{t+\Delta t} = \sqrt{1-\beta(t+\Delta t)\Delta t}\;X_t + \sqrt{\beta(t+\Delta t)\Delta t}\;\varepsilon$$

$(1-x)^a \overset{x\to 0}{\approx} 1-ax$

$$\approx \left(1-\frac{1}{2}\beta(t+\Delta t)\Delta t\right)X_t + \sqrt{\beta(t+\Delta t)}\,\sqrt{\Delta t}\,\varepsilon$$

$X_t \to X_{t+\Delta t}$

$\Delta t \to 0$

$$\approx \left(1-\frac{1}{2}\beta(t)\Delta t\right)X_t + \sqrt{\beta(t)}\,\sqrt{\Delta t}\,\varepsilon$$

$$\boxed{X_{t+\Delta t} = X_t + f(X_t, t)\Delta t + g(t)\sqrt{\Delta t}\,\varepsilon}$$

$$\boxed{f(X_t, t) = -\frac{1}{2}\beta(t)X_t \qquad g(t) = \sqrt{\beta(t)}}$$

补充：

① $\begin{cases} \text{score estimator} & S_\theta(x_t, t) \\ \text{DDPM denoiser} & \varepsilon_\theta(x_t, t) = \dfrac{x_t - \sqrt{\bar\alpha_t}\, x_0}{\sqrt{1-\bar\alpha_t}} = -\dfrac{1}{\sqrt{1-\bar\alpha_t}}\, \varepsilon_\theta(x_t, t) \end{cases}$

$$S_\theta(x_t, t) \approx \nabla_{x_t} \log p(x_t) = -\frac{x_t - \sqrt{\bar\alpha_t}\, x_0}{1-\bar\alpha_t}$$

$$x_t \sim N(\sqrt{\bar\alpha_t}\, x_0, (1-\bar\alpha_t) I)$$

$$p(x_t) \propto \exp\left\{-\frac{\| x_t - \sqrt{\bar\alpha_t}\, x_0 \|_2^2}{2(1-\bar\alpha_t)}\right\}$$

$$S_\theta(x_t, t) = -\frac{1}{\sqrt{1-\bar\alpha_t}}\, \varepsilon_\theta(x_t, t)$$

② VE $\Longleftrightarrow$ VP

VE: $x_t = x_0 + \sigma_t \varepsilon$

$$\frac{x_t}{\sqrt{1+\sigma_t^2}} = \frac{x_0}{\sqrt{1+\sigma_t^2}} + \frac{\sigma_t}{\sqrt{1+\sigma_t^2}}\, \varepsilon$$

$$x_t = \frac{x_t}{\sqrt{1+\sigma_t^2}} \qquad \sqrt{\bar\alpha_t} = \frac{1}{\sqrt{1+\sigma_t^2}} \qquad \sqrt{1-\bar\alpha_t} = \frac{\sigma_t}{\sqrt{1+\sigma_t^2}}$$

VP: $x_t = \sqrt{\bar\alpha_t}\, x_0 + \sqrt{1-\bar\alpha_t}\, \varepsilon$