

参考: 生成扩散模型漫谈(六): 一般框架之ODE篇 - 科学空间|Scientific Spaces

本文建立从SDE到ODE的过程

Review SDE

Fokker-Planck Equation

**Dirac Function** 

F-P Equation

**Equivalent Transform** 

ODE

Review DDIM

Conclusion

### **Review SDE**

一个前向扩散过程由如下SDE给出

$$d\boldsymbol{x} = f_t(\boldsymbol{x}_t)dt + g_t d\boldsymbol{w} \tag{1}$$

其对应逆向SDE为

$$\mathrm{d}oldsymbol{x} = [f_t(oldsymbol{x}_t) - g_t^2 
abla_{oldsymbol{x}_t} \log p(oldsymbol{x}_t)] \mathrm{d}t + g_t \mathrm{d}oldsymbol{w}$$

现在的问题是,是否存在一个SDE,使得该SDE对应的前向边缘分布与上式相同?描述SDE前向边缘分布的方程为Fokker-Planck方程,因此我们先推导F-P方程再利用其来解决上述问题,进而得到ODE形式

# **Fokker-Planck Equation**

### **Dirac Function**

Dirac函数为一类广义函数,满足

$$p(oldsymbol{x}) = \int \delta(oldsymbol{x} - oldsymbol{y}) p(oldsymbol{y}) \mathrm{d}oldsymbol{y} = \mathbb{E}_{oldsymbol{y}}[\delta(oldsymbol{x} - oldsymbol{y})]$$

进一步

$$p(oldsymbol{x})f(oldsymbol{x}) = \int \delta(oldsymbol{x} - oldsymbol{y})p(oldsymbol{y})f(oldsymbol{y})f(oldsymbol{y})doldsymbol{y} = \mathbb{E}_{oldsymbol{y}}[\delta(oldsymbol{x} - oldsymbol{y})f(oldsymbol{y})]$$

两边对x求梯度得

$$abla_{m{x}}[p(m{x})f(m{x})] = \mathbb{E}_{m{y}}[f(m{y})
abla_{m{x}}\delta(m{x}-m{y})]$$

再求梯度有

$$[
abla_{m{x}} \cdot 
abla_{m{x}}[p(m{x})f(m{x})] = \mathbb{E}_{m{y}}[f(m{y})
abla_{m{x}} \cdot 
abla_{m{x}}\delta(m{x}-m{y})]$$

#### F-P Equation

SDE(1)对应的离散形式为

$$oldsymbol{x}_{t+\Delta t} = oldsymbol{x}_t + f_t(oldsymbol{x}_t)\Delta t + g_t\sqrt{\Delta t}oldsymbol{\epsilon}$$

于是有

$$egin{aligned} \delta(oldsymbol{x} - oldsymbol{x}_{t+\Delta t}) &= \delta(oldsymbol{x} - oldsymbol{x}_t - f_t(oldsymbol{x}_t) \Delta t - g_t \sqrt{\Delta t} oldsymbol{\epsilon}) \ &= \delta(oldsymbol{x} - oldsymbol{x}_t) - (f_t(oldsymbol{x}_t) \Delta t + g_t \sqrt{\Delta t} oldsymbol{\epsilon}) \cdot 
abla_{oldsymbol{x}} \delta(oldsymbol{x} - oldsymbol{x}_t) + rac{1}{2} (g_t \sqrt{\Delta t} oldsymbol{\epsilon} \cdot 
abla_{oldsymbol{x}})^2 \delta(oldsymbol{x} - oldsymbol{x}_t) \end{aligned}$$

这里用了二阶Taylor展开并舍去一阶以上的余项

两边求期望得

$$egin{aligned} p_{t+\Delta t}(oldsymbol{x}) &= \mathbb{E}_{oldsymbol{x}_{t+\Delta t}}[\delta(oldsymbol{x} - oldsymbol{x}_{t+\Delta t})] \ &= \mathbb{E}_{oldsymbol{x}_{t},oldsymbol{\epsilon}} \left[ \delta(oldsymbol{x} - oldsymbol{x}_{t}) - (f_t(oldsymbol{x}_t)\Delta t + g_t\sqrt{\Delta t}oldsymbol{\epsilon}) \cdot 
abla_{oldsymbol{x}} \delta(oldsymbol{x} - oldsymbol{x}_t) + rac{1}{2}(g_t\sqrt{\Delta t}oldsymbol{\epsilon} \cdot 
abla_{oldsymbol{x}})^2\delta(oldsymbol{x} - oldsymbol{x}_t) 
ight] \ &= \mathbb{E}_{oldsymbol{x}_t} \left[ \delta(oldsymbol{x} - oldsymbol{x}_t) - f_t(oldsymbol{x}_t)\Delta t \cdot 
abla_{oldsymbol{x}} \delta(oldsymbol{x} - oldsymbol{x}_t) + rac{1}{2}g_t^2\Delta t 
abla_{oldsymbol{x}} \cdot 
abla_{oldsymbol{x}} \delta(oldsymbol{x} - oldsymbol{x}_t) 
ight] \ &= p_t(oldsymbol{x}) - \Delta t 
abla_{oldsymbol{x}} \cdot [f_t(oldsymbol{x})p_t(oldsymbol{x})] + rac{1}{2}g_t^2\Delta t 
abla_{oldsymbol{x}} \cdot 
abla_{oldsymbol{x}} p_t(oldsymbol{x}) \end{aligned}$$

$$rac{\partial}{\partial t} p_t(m{x}) = -
abla_{m{x}} \cdot [f_t(m{x}) p_t(m{x})] + rac{1}{2} g_t^2 
abla_{m{x}} \cdot 
abla_{m{x}} p_t(m{x})$$

此即Fokker-Planck Equation

## **Equivalent Transform**

上面的F-P方程等价于

$$egin{aligned} rac{\partial}{\partial t} p_t(oldsymbol{x}) &= -
abla_{oldsymbol{x}} \cdot [f_t(oldsymbol{x}) p_t(oldsymbol{x})] + rac{1}{2} g_t^2 
abla_{oldsymbol{x}} \cdot 
abla_{oldsymbol{x}} p_t(oldsymbol{x}) &= -
abla_{oldsymbol{x}} \cdot \left[ f_t(oldsymbol{x}) p_t(oldsymbol{x}) - rac{1}{2} g_t^2 
abla_{oldsymbol{x}} p_t(oldsymbol{x}) 
ight] &= -
abla_{oldsymbol{x}} \cdot \left[ f_t(oldsymbol{x}) p_t(oldsymbol{x}) - rac{1}{2} (g_t^2 - \sigma_t^2) 
abla_{oldsymbol{x}} p_t(oldsymbol{x}) 
ight] + rac{1}{2} \sigma_t^2 
abla_{oldsymbol{x}} \cdot 
abla_{oldsymbol{x}} p_t(oldsymbol{x}) &= -
abla_{oldsymbol{x}} \cdot \left[ p_t(oldsymbol{x}) \left( f_t(oldsymbol{x}) - rac{1}{2} (g_t^2 - \sigma_t^2) 
abla_{oldsymbol{x}} \log p_t(oldsymbol{x}) 
ight) 
ight] + rac{1}{2} \sigma_t^2 
abla_{oldsymbol{x}} \cdot 
abla_{oldsymbol{x}} p_t(oldsymbol{x}) &= -
abla_{oldsymbol{x}} \cdot \left[ p_t(oldsymbol{x}) \left( f_t(oldsymbol{x}) - rac{1}{2} (g_t^2 - \sigma_t^2) 
abla_{oldsymbol{x}} \log p_t(oldsymbol{x}) 
ight) 
ight] + rac{1}{2} \sigma_t^2 
abla_{oldsymbol{x}} \cdot 
abla_{oldsymbol{x}} p_t(oldsymbol{x}) &= -
abla_{oldsymbol{x}} \cdot \left[ p_t(oldsymbol{x}) \left( f_t(oldsymbol{x}) - rac{1}{2} (g_t^2 - \sigma_t^2) 
abla_{oldsymbol{x}} \log p_t(oldsymbol{x}) 
ight) 
ight] + rac{1}{2} \sigma_t^2 
abla_{oldsymbol{x}} \cdot 
abla_{oldsymbol{x}} p_t(oldsymbol{x}) &= -
abla_{oldsymbol{x}} \cdot \left[ p_t(oldsymbol{x}) \left( f_t(oldsymbol{x}) - rac{1}{2} (g_t^2 - \sigma_t^2) 
abla_{oldsymbol{x}} \log p_t(oldsymbol{x}) 
ight) 
ight] + rac{1}{2} \sigma_t^2 
abla_{oldsymbol{x}} \cdot 
abla_{oldsymbol{x}} \cdot$$

其中 $\sigma_t^2 \leq g_t^2$ 

将 $f_t(\boldsymbol{x}) - \frac{1}{2}(g_t^2 - \sigma_t^2)\nabla_{\boldsymbol{x}}\log p_t(\boldsymbol{x})$ 看成新的 $f_t(\boldsymbol{x})$ , $\sigma_t$ 看成新的 $g_t$ ,则上述F-P方程亦对应如下SDE

$$\mathrm{d} oldsymbol{x} = \left[ f_t(oldsymbol{x}) - rac{1}{2} (g_t^2 - \sigma_t^2) 
abla_{oldsymbol{x}} \log p_t(oldsymbol{x}) 
ight] \mathrm{d} t + \sigma_t \mathrm{d} oldsymbol{w}$$
 (2)

也就是说,上述SDE对应的边缘分布 $p_t(\boldsymbol{x})$ 与(1)相同

对应的逆向SDE为

$$\mathrm{d}oldsymbol{x} = \left[f_t(oldsymbol{x}) - rac{1}{2}(g_t^2 + \sigma_t^2)
abla_{oldsymbol{x}} \log p_t(oldsymbol{x})
ight]\mathrm{d}t + \sigma_t \mathrm{d}oldsymbol{w}$$

### **ODE**

考虑 $\sigma_t \equiv 0$ ,则上述SDE(2)退化为ODE

$$d\mathbf{x} = \left[ f_t(\mathbf{x}) - \frac{1}{2} g_t^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) \right] dt$$
 (3)

称为Probability flow ODE(概率流ODE)

在拟合得到 $\nabla_{\boldsymbol{x}}\log p_t(\boldsymbol{x})$ 的估计值之后,便可以用ODE数值求解法来求解逆向过程的ODE,相比SDE更加高效和 简洁

### **Review DDIM**

✓ DDIM

 $\underline{\text{SDE}}$ 中有如下关系式(即线性解 $f_t(\boldsymbol{x}) = f_t \boldsymbol{x}$ )

$$egin{aligned} f_t &= rac{1}{ar{lpha}_t} rac{\mathrm{d}ar{lpha}_t}{\mathrm{d}t} \ g_t^2 &= 2ar{lpha}_tar{eta}_t rac{\mathrm{d}}{\mathrm{d}t} \left(rac{ar{eta}_t}{ar{lpha}_t}
ight) \ s_ heta(oldsymbol{x}_t,t) &= -rac{oldsymbol{\epsilon}_ heta(oldsymbol{x}_t,t)}{ar{eta}_t} \end{aligned}$$

将其代入(3)后即可得到DDIM对应ODE

$$rac{\mathrm{d}}{\mathrm{d}t}rac{oldsymbol{x}_t}{ar{lpha}(t)} = oldsymbol{\epsilon}_{ heta}(oldsymbol{x}(t),t)rac{\mathrm{d}}{\mathrm{d}t}rac{ar{eta}(t)}{ar{lpha}(t)}$$

### **Conclusion**

对DDIM, SDE, ODE作一个总结

- 1. SDE首先给出了扩散模型(DDPM)的连续形式(由SDE)描述
- 2. ODE利用F-P方程对SDE作了进一步推广(保持边缘概率密度不变),得到了变方差采样形式。取零方差采样、即 可得到对应的ODE形式
- 3. DDIM实际上利用非Markov过程推导得到了ODE的特殊情形(SDE线性解+ODE零方差)

这些推导中最核心的量便是

$$oldsymbol{x}_t = ar{lpha}_t oldsymbol{x}_0 + ar{eta}_t oldsymbol{\epsilon}$$

# 即 $\bar{\alpha}_t, \bar{\beta}_t$ 的选择

从上述方程出发,我们可以推导得到对应的SDE,进而用F-P方程一般化;也可以直接推出线性解情形的ODE(DDIM实际上隐含假设了线性解: $m{x}_{t-1} = \kappa_t m{x}_t + \lambda_t m{x}_0 + \sigma_t m{\epsilon}$ )