ODE

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

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

其对应逆向SDE为

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

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

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

进一步

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

两边对$\boldsymbol{x}$求梯度得

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

再求梯度有

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

F-P Equation

SDE(1)对应的离散形式为

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

于是有

$$\begin{aligned}\delta(\boldsymbol{x}-\boldsymbol{x}_{t+\Delta t})&=\delta(\boldsymbol{x}-\boldsymbol{x}_t-f_t(\boldsymbol{x}_t)\Delta t-g_t\sqrt{\Delta t}\boldsymbol{\epsilon})\\&=\delta(\boldsymbol{x}-\boldsymbol{x}_t)-(f_t(\boldsymbol{x}_t)\Delta t+g_t\sqrt{\Delta t}\boldsymbol{\epsilon})\cdot\nabla_{\boldsymbol{x}}\delta(\boldsymbol{x}-\boldsymbol{x}_t)+\frac{1}{2}(g_t\sqrt{\Delta t}\boldsymbol{\epsilon}\cdot\nabla_{\boldsymbol{x}})^2\delta(\boldsymbol{x}-\boldsymbol{x}_t)\end{aligned}$$

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

$$\begin{aligned}p_{t+\Delta t}(\boldsymbol{x})&=\mathbb{E}_{\boldsymbol{x}_{t+\Delta t}}[\delta(\boldsymbol{x}-\boldsymbol{x}_{t+\Delta t})]\\&=\mathbb{E}_{\boldsymbol{x}_t,\boldsymbol{\epsilon}}\left[\delta(\boldsymbol{x}-\boldsymbol{x}_t)-(f_t(\boldsymbol{x}_t)\Delta t+g_t\sqrt{\Delta t}\boldsymbol{\epsilon})\cdot\nabla_{\boldsymbol{x}}\delta(\boldsymbol{x}-\boldsymbol{x}_t)+\frac{1}{2}(g_t\sqrt{\Delta t}\boldsymbol{\epsilon}\cdot\nabla_{\boldsymbol{x}})^2\delta(\boldsymbol{x}-\boldsymbol{x}_t)\right]\\&=\mathbb{E}_{\boldsymbol{x}_t}\left[\delta(\boldsymbol{x}-\boldsymbol{x}_t)-f_t(\boldsymbol{x}_t)\Delta t\cdot\nabla_{\boldsymbol{x}}\delta(\boldsymbol{x}-\boldsymbol{x}_t)+\frac{1}{2}g_t^2\Delta t\nabla_{\boldsymbol{x}}\cdot\nabla_{\boldsymbol{x}}\delta(\boldsymbol{x}-\boldsymbol{x}_t)\right]\\&=p_t(\boldsymbol{x})-\Delta t\nabla_{\boldsymbol{x}}\cdot[f_t(\boldsymbol{x})p_t(\boldsymbol{x})]+\frac{1}{2}g_t^2\Delta t\nabla_{\boldsymbol{x}}\cdot\nabla_{\boldsymbol{x}}p_t(\boldsymbol{x})\end{aligned}$$

令$\Delta t\to 0$得到

$$\frac{\partial}{\partial t}p_t(\boldsymbol{x})=-\nabla_{\boldsymbol{x}}\cdot[f_t(\boldsymbol{x})p_t(\boldsymbol{x})]+\frac{1}{2}g_t^2\nabla_{\boldsymbol{x}}\cdot\nabla_{\boldsymbol{x}}p_t(\boldsymbol{x})$$

此即Fokker-Planck Equation

Equivalent Transform

上面的F-P方程等价于

$$\begin{aligned}\frac{\partial}{\partial t}p_t(\boldsymbol{x})&=-\nabla_{\boldsymbol{x}}\cdot[f_t(\boldsymbol{x})p_t(\boldsymbol{x})]+\frac{1}{2}g_t^2\nabla_{\boldsymbol{x}}\cdot\nabla_{\boldsymbol{x}}p_t(\boldsymbol{x})\\&=-\nabla_{\boldsymbol{x}}\cdot\left[f_t(\boldsymbol{x})p_t(\boldsymbol{x})-\frac{1}{2}g_t^2\nabla_{\boldsymbol{x}}p_t(\boldsymbol{x})\right]\\&=-\nabla_{\boldsymbol{x}}\cdot\left[f_t(\boldsymbol{x})p_t(\boldsymbol{x})-\frac{1}{2}(g_t^2-\sigma_t^2)\nabla_{\boldsymbol{x}}p_t(\boldsymbol{x})\right]+\frac{1}{2}\sigma_t^2\nabla_{\boldsymbol{x}}\cdot\nabla_{\boldsymbol{x}}p_t(\boldsymbol{x})\\&=-\nabla_{\boldsymbol{x}}\cdot\left[p_t(\boldsymbol{x})\left(f_t(\boldsymbol{x})-\frac{1}{2}(g_t^2-\sigma_t^2)\nabla_{\boldsymbol{x}}\log p_t(\boldsymbol{x})\right)\right]+\frac{1}{2}\sigma_t^2\nabla_{\boldsymbol{x}}\cdot\nabla_{\boldsymbol{x}}p_t(\boldsymbol{x})\end{aligned}$$

其中$\sigma_t^2\le g_t^2$

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

$$\mathrm{d}\boldsymbol{x}=\left[f_t(\boldsymbol{x})-\frac{1}{2}(g_t^2-\sigma_t^2)\nabla_{\boldsymbol{x}}\log p_t(\boldsymbol{x})\right]\mathrm{d}t+\sigma_t\mathrm{d}\boldsymbol{w}\tag{2}$$

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

$$\mathrm{d}\boldsymbol{x}=\left[f_t(\boldsymbol{x})-\frac{1}{2}(g_t^2+\sigma_t^2)\nabla_{\boldsymbol{x}}\log p_t(\boldsymbol{x})\right]\mathrm{d}t+\sigma_t\mathrm{d}\boldsymbol{w}$$

ODE

考虑$\sigma_t\equiv 0$，则上述SDE(2)退化为ODE

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

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

- [x] ~~DDIM~~

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

$$\begin{aligned}f_t&=\frac{1}{\bar{\alpha}_t}\frac{\mathrm{d}\bar{\alpha}_t}{\mathrm{d}t}\\ g_t^2&=2\bar{\alpha}_t\bar{\beta}_t\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\bar{\beta}_t}{\bar{\alpha}_t}\right)\\ s_\theta(\boldsymbol{x}_t,t)&=-\frac{\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t,t)}{\bar{\beta}_t}\end{aligned}$$

将其代入(3)后即可得到DDIM对应ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\boldsymbol{x}_t}{\bar{\alpha}(t)}=\boldsymbol{\epsilon}_\theta(\boldsymbol{x}(t),t)\frac{\mathrm{d}}{\mathrm{d}t}\frac{\bar{\beta}(t)}{\bar{\alpha}(t)}$$

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

$$\boldsymbol{x}_t = \bar{\alpha}_t \boldsymbol{x}_0 + \bar{\beta}_t \boldsymbol{\epsilon}$$

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：$\boldsymbol{x}_{t-1} = \kappa_t \boldsymbol{x}_t + \lambda_t \boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon}$)