28 Sep

生成扩散模型漫谈（十二）：“硬刚”扩散ODE

By 苏剑林 | 2022-09-28 | 67079位读者 引用

在《生成扩散模型漫谈（五）：一般框架之SDE篇》中，我们从SDE的角度理解了生成扩散模型，然后在《生成扩散模型漫谈（六）：一般框架之ODE篇》中，我们知道SDE对应的扩散模型中，实际上隐含了一个ODE模型。无独有偶，在《生成扩散模型漫谈（四）：DDIM = 高观点DDPM》中我们也知道原本随机采样的DDPM模型中，也隐含了一个确定性的采样过程DDIM，它的连续极限也是一个ODE。

细想上述过程，可以发现不管是“DDPM→DDIM”还是“SDE→ODE”，都是从随机采样模型过渡到确定性模型，而如果我们一开始的目标就是ODE，那么该过程未免显得有点“迂回”了。在本文中，笔者尝试给出ODE扩散模型的直接推导，并揭示了它与雅可比行列式、热传导方程等内容的联系。

微分方程

像GAN这样的生成模型，它本质上是希望找到一个确定性变换，能将从简单分布（如标准正态分布）采样出来的随机变量，变换为特定数据分布的样本。flow模型也是生成模型之一，它的思路是反过来，先找到一个能将数据分布变换简单分布的可逆变换，再求解相应的逆变换来得到一个生成模型。

传统的flow模型是通过设计精巧的耦合层（参考“细水长flow”系列）来实现这个可逆变换，但后来大家就意识到，其实通过微分方程也能实现这个变换，并且理论上还很优雅。基于“神经网络 + 微分方程”做生成模型等一系列研究，构成了被称为“神经ODE”的一个子领域。

考虑$\boldsymbol{x}_t \in \mathbb{R}^d$上的一阶（常）微分方程（组）

$$\frac{d\boldsymbol{x}_t}{dt} = \boldsymbol{f}_t(\boldsymbol{x}_t) \tag{1}$$

假设$t \in [0, T]$，那么给定$\boldsymbol{x}_0$，（在比较容易实现的条件下）我们可以确定地求解出

$\boldsymbol{x}_T$，也就是说该微分方程描述了从$\boldsymbol{x}_0$到$\boldsymbol{x}_T$的一个变换。特别地，该变换还是可逆的，即可以逆向求解该微分方程，得到从$\boldsymbol{x}_T$到$\boldsymbol{x}_0$的变换。所以说，微分方程本身就是构建可逆变换的一个理论优雅的方案。

雅可比行列式

跟之前的扩散模型一样，在这篇文章中，我们将$\boldsymbol{x}_0$视为一个数据样本，而将$\boldsymbol{x}_T$视为简单分布的样本，我们希望通过微分方程，来实现从数据分布到简单分布的变换。

首先，我们从离散化的角度来理解微分方程(1)：

$$\boldsymbol{x}_{t+\Delta t} - \boldsymbol{x}_t = \boldsymbol{f}_t(\boldsymbol{x}_t)\Delta t \tag{2}$$

由于是确定性变换，所以我们有

$$p_t(\boldsymbol{x}_t)d\boldsymbol{x}_t = p_{t+\Delta t}(\boldsymbol{x}_{t+\Delta t})d\boldsymbol{x}_{t+\Delta t} = p_{t+\Delta t}(\boldsymbol{x}_{t+\Delta t})\left|\frac{\partial \boldsymbol{x}_{t+\Delta t}}{\partial \boldsymbol{x}_t}\right| d\boldsymbol{x}_t \tag{3}$$

这里的$\frac{\partial \boldsymbol{x}_{t+\Delta t}}{\partial \boldsymbol{x}_t}$表示变换的雅可比矩阵，$|\cdot|$代表行列式的绝对值。直接对式(2)两边求偏导，我们就得到

$$\frac{\partial \boldsymbol{x}_{t+\Delta t}}{\partial \boldsymbol{x}_t} = \boldsymbol{I} + \frac{\partial \boldsymbol{f}_t(\boldsymbol{x}_t)}{\partial \boldsymbol{x}_t}\Delta t \tag{4}$$

根据《行列式的导数》一文，我们就有

$$\left|\frac{\partial \boldsymbol{x}_{t+\Delta t}}{\partial \boldsymbol{x}_t}\right| \approx 1 + \text{Tr}\,\frac{\partial \boldsymbol{f}_t(\boldsymbol{x}_t)}{\partial \boldsymbol{x}_t}\Delta t = 1 + \nabla_{\boldsymbol{x}_t}\cdot \boldsymbol{f}_t(\boldsymbol{x}_t)\Delta t \approx e^{\nabla_{\boldsymbol{x}_t}\cdot \boldsymbol{f}_t(\boldsymbol{x}_t)\Delta t} \tag{5}$$

于是我们可以写出

$$\log p_{t+\Delta t}(\boldsymbol{x}_{t+\Delta t}) - \log p_t(\boldsymbol{x}_t) \approx -\nabla_{\boldsymbol{x}_t}\cdot \boldsymbol{f}_t(\boldsymbol{x}_t)\Delta t \tag{6}$$

泰勒近似

假设$p_t(\boldsymbol{x}_t)$是一簇随着参数t连续变化的分布的概率密度函数，其中$p_0(\boldsymbol{x}_0)$是数据分布，$p_T(\boldsymbol{x}_T)$则是简单分布，当Δt和$\boldsymbol{x}_{t+\Delta t}-\boldsymbol{x}_t$都较小时，我们有一阶泰勒近似

$$\log p_{t+\Delta t}(\boldsymbol{x}_{t+\Delta t}) - \log p_t(\boldsymbol{x}_t) \approx (\boldsymbol{x}_{t+\Delta t}-\boldsymbol{x}_t)\cdot\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t) + \Delta t\frac{\partial}{\partial t}\log p_t(\boldsymbol{x}_t)$$

代入式(2)的$\boldsymbol{x}_{t+\Delta t}-\boldsymbol{x}_t$，然后对照式(6)，可以得到$\boldsymbol{f}_t(\boldsymbol{x}_t)$所满足的方程

$$-\nabla_{\boldsymbol{x}_t}\cdot\boldsymbol{f}_t(\boldsymbol{x}_t) = \boldsymbol{f}_t(\boldsymbol{x}_t)\cdot\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t) + \frac{\partial}{\partial t}\log p_t(\boldsymbol{x}_t) \tag{8}$$

换句话说，满足该方程的任意$\boldsymbol{f}_t(\boldsymbol{x}_t)$，都可以用来构造一个微分方程(1)，通过求解它来实现数据分布和简单分布之间的变换。我们也可以将它整理得

$$\frac{\partial}{\partial t}p_t(\boldsymbol{x}_t) = -\nabla_{\boldsymbol{x}_t}\cdot\Big(\boldsymbol{f}_t(\boldsymbol{x}_t)p_t(\boldsymbol{x}_t)\Big) \tag{9}$$

它其实就是《生成扩散模型漫谈（六）：一般框架之ODE篇》介绍的“Fokker-Planck方程”在$g_t=0$时的特例。

热传导方程

我们考虑如下格式的解

$$\boldsymbol{f}_t(\boldsymbol{x}_t) = -\boldsymbol{D}_t(\boldsymbol{x}_t)\,\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t) \tag{10}$$

其中$\boldsymbol{D}_t(\boldsymbol{x}_t)$可以是一个矩阵，也可能是一个标量，视具体考虑的复杂度而定。为什么要考虑这种形式的解？说实话，笔者一开始就是往DDIM格式去凑的，后来就是发现一般化后能跟下面的扩散方程联系起来，所以就直接设为式(10)了。事后来看，如果假设$\boldsymbol{D}_t(\boldsymbol{x}_t)$是非负标量函数，那么将它代入式(2)后，就会发现其格式跟梯度下降有点相似，即从$\boldsymbol{x}_0$到$\boldsymbol{x}_T$是逐渐寻找低概率区域，反之从$\boldsymbol{x}_T$到$\boldsymbol{x}_0$就是逐渐寻找高概率区域，跟直觉相符，这也算是式(10)的一个启发式引导吧。

将式(10)代入方程(9)后，我们可以得到

$$\frac{\partial}{\partial t}p_t(\boldsymbol{x}_t) = \nabla_{\boldsymbol{x}_t}\cdot\Big(\boldsymbol{D}_t(\boldsymbol{x}_t)\,\nabla_{\boldsymbol{x}_t}p_t(\boldsymbol{x}_t)\Big) \tag{11}$$

这就是偏微分方程中的“扩散方程”。这里我们只考虑一个极简单的情形——$\boldsymbol{D}_t(\boldsymbol{x}_t)$是跟$\boldsymbol{x}_t$无关的标量函数$D_t$，此时扩散方程简化为

$$\frac{\partial}{\partial t}p_t(\boldsymbol{x}_t) = D_t\nabla_{\boldsymbol{x}_t}^2 p_t(\boldsymbol{x}_t) \tag{12}$$

这就是“热传导方程”，是我们接下来要重点求解和分析的对象。

求解分布

利用傅里叶变换，可以将热传导方程转为常微分方程，继而完成分布$p_t(\boldsymbol{x}_t)$的求解，结果是：

$$\begin{aligned}p_t(\boldsymbol{x}_t) =&\, \int \frac{1}{(2\pi\sigma_t^2)^{d/2}}\exp\left(-\frac{\Vert\boldsymbol{x}_t - \boldsymbol{x}_0\Vert^2}{2\sigma_t^2}\right)p_0(\boldsymbol{x}_0)d\boldsymbol{x}_0 \\ =&\, \int \mathcal{N}(\boldsymbol{x}_t;\boldsymbol{x}_0,\sigma_t^2\boldsymbol{I})\,p_0(\boldsymbol{x}_0)d\boldsymbol{x}_0\end{aligned} \tag{13}$$

其中$\sigma_t^2 = 2\int_0^t D_s ds$，或者$D_t = \dot{\sigma}_t\sigma_t$（其中$\sigma_0 = 0$）。可以看到，热传导方程的解正好是以$p_0(\boldsymbol{x}_0)$为初始分布的高斯混合模型。

过程：这里简单介绍一下热传导方程的求解思路。对于不关心求解过程的读者，或者已经熟悉热传导方程的读者，可以跳过这部分内容。

用傅里叶变换求热传导方程(12)其实很简单，对两边的$\boldsymbol{x}_t$变量做傅里叶变换，根据$\nabla_{\boldsymbol{x}_t}\to i\boldsymbol{\omega}$的原则，结果是

$$\frac{\partial}{\partial t}\mathcal{F}_t(\boldsymbol{\omega}) = -D_t\boldsymbol{\omega}^2\mathcal{F}_t(\boldsymbol{\omega}) \tag{14}$$

这只是关于t的常微分方程，可以解得

$$\mathcal{F}_t(\boldsymbol{\omega}) = \mathcal{F}_0(\boldsymbol{\omega})\exp\left(-\frac{1}{2}\sigma_t^2\boldsymbol{\omega}^2\right) \tag{15}$$

其中$\sigma_t^2 = 2\int_0^t D_s ds$，而$\mathcal{F}_0(\boldsymbol{\omega})$则是$p_0(\boldsymbol{x}_0)$的傅里叶变换。现在对两边做傅里叶逆变换，$\mathcal{F}_t(\boldsymbol{\omega})$自然变回$p_t(\boldsymbol{x}_t)$，$\mathcal{F}_0(\boldsymbol{\omega})$变回$p_0(\boldsymbol{x}_0)$，$\exp\left(-\frac{1}{2}\sigma_t^2\boldsymbol{\omega}^2\right)$则对应正态分布$\mathcal{N}(\boldsymbol{x}_t;\boldsymbol{0},\sigma_t^2\boldsymbol{I})$，最后利用傅里叶变换的卷积性质，就得到解(13)。

完成设计

现在我们汇总一下我们的结果：通过求解热传导方程，我们确定了

$$p_t(\boldsymbol{x}_t) = \int \mathcal{N}(\boldsymbol{x}_t;\boldsymbol{x}_0,\sigma_t^2\boldsymbol{I})\, p_0(\boldsymbol{x}_0)d\boldsymbol{x}_0 \tag{16}$$

此时对应的微分方程

$$\frac{d\boldsymbol{x}_t}{dt} = -\dot{\sigma}_t\sigma_t\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t) \tag{17}$$

给出了从$p_0(\boldsymbol{x}_0)$到$p_T(\boldsymbol{x}_T)$的一个确定性变换。如果$p_T(\boldsymbol{x}_T)$易于采样，并且$\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t)$已知，那么我们就可以随机采样$\boldsymbol{x}_T\sim p_T(\boldsymbol{x}_T)$，然后逆向求解该微分方程，来生成$\boldsymbol{x}_0\sim p_0(\boldsymbol{x}_0)$的样本。

第一个问题，什么时候$p_T(\boldsymbol{x}_T)$是易于采样的？根据结果(16)，我们知道

$$\boldsymbol{x}_T\sim p_T(\boldsymbol{x}_T) \quad\Leftrightarrow\quad \boldsymbol{x}_T = \boldsymbol{x}_0 + \sigma_T\boldsymbol{\varepsilon},\quad \boldsymbol{x}_0\sim p_0(\boldsymbol{x}_0),\ \boldsymbol{\varepsilon}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}) \tag{18}$$

当σ_T足够大时，$\boldsymbol{x}_0$对$\boldsymbol{x}_T$的影响就很微弱了，此时可以认为

$$\boldsymbol{x}_T\sim p_T(\boldsymbol{x}_T) \quad\Leftrightarrow\quad \boldsymbol{x}_T = \sigma_T\boldsymbol{\varepsilon},\quad \boldsymbol{\varepsilon}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}) \tag{19}$$

这就实现了$p_T(\boldsymbol{x}_T)$易于采样的目的。因此，选择σ_t的一般要求是：满足$\sigma_0 = 0$和$\sigma_T \gg 1$的光滑单调递增函数。

第二个问题，就是如何计算$\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t)$？这其实跟《生成扩散模型漫谈（五）：一般框架之SDE篇》中的"得分匹配"一节是一样的，我们用一个神经网络$\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$去拟合

它，训练目标是

$$\mathbb{E}_{\boldsymbol{x}_0,\boldsymbol{x}_t\sim\mathcal{N}(\boldsymbol{x}_t;\boldsymbol{x}_0,\sigma_t^2\boldsymbol{I})p_0(\boldsymbol{x}_0)}\left[\left\Vert\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t) - \nabla_{\boldsymbol{x}_t}\log\mathcal{N}(\boldsymbol{x}_t;\boldsymbol{x}_0,\sigma_t^2\boldsymbol{I})\right\Vert^2\right] \tag{20}$$

这叫做“条件得分匹配”，其推导我们在SDE篇已经给出了，这里就不重复了。

文章小结

在这篇文章中，我们对ODE式扩散模型做了一个“自上而下”的推导：首先从ODE出发，结合雅可比行列式得到了概率变化的一阶近似，然后对比直接泰勒展开的一阶近似，得到了ODE应该要满足的方程，继而转化为扩散方程、热传导方程来求解。相对来说，整个过程比较一步到位，不需要通过SDE、FP方程等结果来做过渡。

如果您需要引用本文，请参考：

苏剑林. (Sep. 28, 2022). 《生成扩散模型漫谈（十二）：“硬刚”扩散ODE 》[Blog post]. Retrieved from https://spaces.ac.cn/archives/9280

```
@online{kexuefm-9280,
    title={生成扩散模型漫谈（十二）：“硬刚”扩散ODE},
    author={苏剑林},
    year={2022},
    month={Sep},
    url={\url{https://spaces.ac.cn/archives/9280}},
}
```