15
Dec

生成扩散模型漫谈（十四）：构建ODE的一般步骤（上）

By 苏剑林 | 2022-12-15 | 54619位读者 引用

书接上文，在《生成扩散模型漫谈（十三）：从万有引力到扩散模型》中，我们介绍了一个由万有引力启发的、几何意义非常清晰的ODE式生成扩散模型。有的读者看了之后就疑问：似乎“万有引力”并不是唯一的选择，其他形式的力是否可以由同样的物理绘景构建扩散模型？另一方面，该模型在物理上确实很直观，但还欠缺从数学上证明最后确实能学习到数据分布。

本文就尝试从数学角度比较精确地回答“什么样的力场适合构建ODE式生成扩散模型”这个问题。

基础结论

要回答这个问题，需要用到在《生成扩散模型漫谈（十二）：“硬刚”扩散ODE》中我们推导过的一个关于常微分方程对应的分布变化的结论。

考虑$\boldsymbol{x}_t \in \mathbb{R}^d, t \in [0, T]$的一阶（常）微分方程（组）

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

它描述了从$\boldsymbol{x}_0$到$\boldsymbol{x}_T$的一个（可逆）变换，如果$\boldsymbol{x}_0$是一个随机变量，那么整个过程中的$\boldsymbol{x}_t$也都是随机变量，它的分布变化规律，可以由如下方程描述

$$\frac{\partial}{\partial t} p_t(\boldsymbol{x}_t) = -\nabla_{\boldsymbol{x}_t} \cdot \Big(\boldsymbol{f}_t(\boldsymbol{x}_t) p_t(\boldsymbol{x}_t)\Big) \tag{2}$$

该结果可以按照《生成扩散模型漫谈（十二）：“硬刚”扩散ODE》的格式用“雅可比行列式+泰勒近似”的方式推导，也可以像《生成扩散模型漫谈（六）：一般框架之ODE篇》一样先推导完整的“Fokker-Planck方程”，然后让$g_t = 0$。顺便一提，方程(2)在物理上非常出名，它被称为“连续性方程”，是各种守恒定律的体现之一。

回到扩散模型，扩散模型想要做的事情，是构造一个变换，能够将简单分布的样本变换成目标分布的样本。而利用式(2)，理论上我们可以通过给定的$p_t(\boldsymbol{x}_t)$来可以求出可行的$\boldsymbol{f}_t(\boldsymbol{x}_t)$，继而利用式$(1)$完成生成过程。注意，式$(2)$只是一个方程，但是要求解的$\boldsymbol{f}_t(\boldsymbol{x}_t)$有$d$个分量，所以这是一个不定方程，原则上来说我们可以任意指定完整的$p_t(\boldsymbol{x}_t)$（而不单单是$t=0,T$两个边界）来求解$\boldsymbol{f}_t(\boldsymbol{x}_t)$。

所以从理论上来说，构建ODE式扩散模型只是求解一个非常轻松的几乎没约束的不定方程。确实如此，但问题是这样求出来的解在实践上会有困难，说白了就是代码上不好实现。因此，问题的准确提法是如何从式(2)中求出更实用的解。

简化方程

留意到，式(2)可以改写成

$$\underbrace{\left(\frac{\partial}{\partial t},\nabla_{\boldsymbol{x}_t}\right)}_{\nabla_{(t,\,\boldsymbol{x}_t)}}\cdot\underbrace{\Big(p_t(\boldsymbol{x}_t),\boldsymbol{f}_t(\boldsymbol{x}_t)p_t(\boldsymbol{x}_t)\Big)}_{\boldsymbol{u}\in\mathbb{R}^{d+1}}=0 \tag{3}$$

如上式所示，$\left(\frac{\partial}{\partial t},\nabla_{\boldsymbol{x}_t}\right)$我们刚好可以当成$d+1$维的梯度$\nabla_{(t,\,\boldsymbol{x}_t)}$，$\Big(p_t(\boldsymbol{x}_t),\boldsymbol{f}_t(\boldsymbol{x}_t)p_t(\boldsymbol{x}_t)\Big)$正好可以组成了一个$d+1$的向量$\boldsymbol{u}(t,\boldsymbol{x}_t)$，所以$(2)$可以写成简单的散度方程

$$\nabla_{(t,\,\boldsymbol{x}_t)}\cdot\boldsymbol{u}(t,\boldsymbol{x}_t)=0 \tag{4}$$

在此形式之下有

$$\frac{d\boldsymbol{x}_t}{dt}=\boldsymbol{f}_t(\boldsymbol{x}_t)=\frac{\boldsymbol{u}_{>1}(t,\boldsymbol{x}_t)}{\boldsymbol{u}_1(t,\boldsymbol{x}_t)} \tag{5}$$

其中$\boldsymbol{u}_1$、$\boldsymbol{u}_{>1}$分别代表$\boldsymbol{u}$的第一维分量和后d维分量。当然，不能忘了约束条件

$$\left\{\begin{aligned}
&\boldsymbol{u}_1(0,\boldsymbol{x}_0)=p_0(\boldsymbol{x}_0) \qquad &(\text{初值条件})\\
&\int\boldsymbol{u}_1(t,\boldsymbol{x}_t)d\boldsymbol{x}_t=1 \qquad &(\text{积分条件})
\end{aligned}\right. \tag{6}$$

其中$p_0(\boldsymbol{x}_0)$是数据分布，即要生成的目标样本分布。对于$t=T$时的终值分布，我们对它的要求只是尽可能简单，方便采样，除此之外没有定量要求，因此这里暂时不用写出。

格林函数

经过这样的形式变换后，我们可以将$\boldsymbol{u}(t,\boldsymbol{x}_t)$看成一个$d+1$维的向量场，而微分方程$(5)$正好描述的是质点沿着场线运动的轨迹，这样就跟《生成扩散模型漫谈（十三）：从万有引力到扩散模型》所给出的物理图景同出一辙了。

为了求出$\boldsymbol{u}(t,\boldsymbol{x}_t)$的一般解，我们可以用格林函数的思想。首先尝试求解如下问题：

$$\left\{\begin{aligned}&\nabla_{(t,\,\boldsymbol{x}_t)}\cdot\boldsymbol{G}(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)=0\\ &\boldsymbol{G}_1(0,0;\boldsymbol{x}_t,\boldsymbol{x}_0)=\delta(\boldsymbol{x}_t-\boldsymbol{x}_0),\int\boldsymbol{G}_1(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)d\boldsymbol{x}_t=1\end{aligned}\right.\tag{7}$$

容易证明，如果上式成立，那么

$$\boldsymbol{u}(t,\boldsymbol{x}_t)=\int\boldsymbol{G}(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)p_0(\boldsymbol{x}_0)d\boldsymbol{x}_0=\mathbb{E}_{\boldsymbol{x}_0\sim p_0(\boldsymbol{x}_0)}[\boldsymbol{G}(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)]\tag{8}$$

将是方程(4)满足相应约束的解。这样一来，我们就将$\boldsymbol{u}(t,\boldsymbol{x}_t)$表示为了训练样本的期望形式，这有利于模型的训练。不难看出，这里的$\boldsymbol{G}_1(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)$实际上就是扩散模型中的条件概率$p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)$。

事实上，式(7)所定义的$\boldsymbol{G}(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)$，并非通常意义下的格林函数。一般的格林函数指的是点源下的解，而这里的格林函数的“点源”放到了边界处。但即便如此，所定义的$\boldsymbol{G}(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)$依然具有常规格林函数类似的性质，它本身也相当于点源产生的“力场”，而式(8)也正好是对点源的场进行积分，求出了连续分布源的场。

万有引力

现在我们根据上述框架，求解一些具体的结果。前面已经提到，方程(4)或(7)，都是“$d+1$个未知数、一个方程”的不定方程，理论上具有无穷多的各式各样的解，我们要对

它进行求解，反而要引入一些额外的假设，使得它的解更为明确一些。第一个解是基于各向同性假设，它正好对应《生成扩散模型漫谈（十三）：从万有引力到扩散模型》中的结果。

假设求解

注意，这里的“各向同性”，指的是在$(t, \boldsymbol{x}_t)$组成的$d+1$维空间中的各向同性，这意味着$\boldsymbol{G}(t, 0; \boldsymbol{x}_t, \boldsymbol{x}_0)$是指向源点$(0,\ \boldsymbol{x}_0)$的，且模长只依赖于$R = \sqrt{(t-0)^2 + \Vert\boldsymbol{x}_t - \boldsymbol{x}_0\Vert^2}$，因此可以设

$$\boldsymbol{G}(t, 0; \boldsymbol{x}_t, \boldsymbol{x}_0) = \varphi(R)(t, \boldsymbol{x}_t - \boldsymbol{x}_0) \tag{9}$$

于是

$$\begin{aligned}
0 =&\, \nabla_{(t,\, \boldsymbol{x}_t)}\cdot\boldsymbol{G}(t, 0; \boldsymbol{x}_t, \boldsymbol{x}_0) \\
=&\, \nabla_{(t,\, \boldsymbol{x}_t)}\varphi(R)\cdot(t, \boldsymbol{x}_t - \boldsymbol{x}_0) + \varphi(R)\nabla_{(t,\, \boldsymbol{x}_t)}\cdot(t, \boldsymbol{x}_t - \boldsymbol{x}_0) \\
=&\, \varphi'(R)\frac{(t, \boldsymbol{x}_t - \boldsymbol{x}_0)}{R}\cdot(t, \boldsymbol{x}_t - \boldsymbol{x}_0) + (d+1)\varphi(R) \\
=&\, \varphi'(R)R + (d+1)\varphi(R) \\
=&\, \frac{[\varphi(R)R^{d+1}]'}{R^d}
\end{aligned} \tag{10}$$

也即$[\varphi(R)R^{d+1}]' = 0$，或$\varphi(R)R^{d+1} = C$，即$\varphi(R) = C\times R^{-(d+1)}$，因此一个候选解是

$$\boldsymbol{G}(t, 0; \boldsymbol{x}_t, \boldsymbol{x}_0) = C\times\frac{(t, \boldsymbol{x}_t - \boldsymbol{x}_0)}{(t^2 + \Vert\boldsymbol{x}_t - \boldsymbol{x}_0\Vert^2)^{(d+1)/2}} \tag{11}$$

约束条件

可以看到，在各向同性假设下，万有引力解是唯一解了。为了证明是可行解，还要检验约束条件，其中关键一条是

$$\int \boldsymbol{G}_1(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)d\boldsymbol{x}_t = C\times\int\frac{t}{(t^2+\|\boldsymbol{x}_t-\boldsymbol{x}_0\|^2)^{(d+1)/2}}d\boldsymbol{x}_t \tag{12}$$

其实我们只需要检验积分结果跟t和$\boldsymbol{x}_0$都没关系，那么就可以选择适当的常数C让积分结果为1。而对于$t>0$，可以检验做变量代换$\boldsymbol{z}=(\boldsymbol{x}_t-\boldsymbol{x}_0)/t$，由于$\boldsymbol{x}_t$的范围是全空间的，所以$\boldsymbol{z}$也是全空间的，代入上式得到

$$\int \boldsymbol{G}_1(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)d\boldsymbol{x}_t = C\times\int\frac{1}{(1+\|\boldsymbol{z}\|^2)^{(d+1)/2}}d\boldsymbol{z} \tag{13}$$

现在可以看出积分结果跟t和$\boldsymbol{x}_0$都无关了。因此只要选择适当的C，积分为1这一条检验可以通过。下面都假设已经选择了让积分为1的C。

至于初值，我们需要验证$\lim\limits_{t\to 0^+}\boldsymbol{G}_1(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)=\delta(\boldsymbol{x}_t-\boldsymbol{x}_0)$，这只需要按照狄拉克函数的定义进行检验就行了：

> 1、当$\boldsymbol{x}_t\neq\boldsymbol{x}_0$时，极限显然为0；
>
> 2、当$\boldsymbol{x}_t=\boldsymbol{x}_0$时，极限显然为$\infty$；
>
> 3、刚才我们已经检验了，$\boldsymbol{G}_1(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)$关于$\boldsymbol{x}_t$的积分恒为1。

这三点正好是狄拉克函数的基本性质，甚至可以说是狄拉克函数的定义，因此初值检验也可以通过。

结果分析

现在，根据式(8)我们就有

$$\boldsymbol{u}(t,\boldsymbol{x}_t) = C\times\mathbb{E}_{\boldsymbol{x}_0\sim p_0(\boldsymbol{x}_0)}\left[\frac{(t,\boldsymbol{x}_t-\boldsymbol{x}_0)}{(t^2+\|\boldsymbol{x}_t-\boldsymbol{x}_0\|^2)^{(d+1)/2}}\right] \tag{14}$$

接下来利用$\mathbb{E}_{\boldsymbol{x}}[\boldsymbol{x}] = \mathop{\text{argmin}}_{\boldsymbol{\mu}} \mathbb{E}_{\boldsymbol{x}}\left[\Vert\boldsymbol{x} - \boldsymbol{\mu}\Vert^2\right]$构建一个类似得分匹配的目标进行学习就行了，这个过程已经说过多次，不再重复展开。

前面提到过，$\boldsymbol{G}_1(t, 0; \boldsymbol{x}_t, \boldsymbol{x}_0)$实际上就是$p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)$，现在我们已经知道它的具体形式为

$$p_t(\boldsymbol{x}_t|\boldsymbol{x}_0) \propto \frac{t}{(t^2 + \Vert\boldsymbol{x}_t - \boldsymbol{x}_0\Vert^2)^{(d+1)/2}} \tag{15}$$

当$t=T$足够大的时候，$\boldsymbol{x}_0$的影响就微乎其微，即$p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)$退化为跟$\boldsymbol{x}_0$无关的先验分布

$$p_{prior}(\boldsymbol{x}_T) \propto \frac{T}{(T^2 + \Vert\boldsymbol{x}_T\Vert^2)^{(d+1)/2}} \tag{16}$$

之前我们在《生成扩散模型漫谈（十三）：从万有引力到扩散模型》中推导这一结果还颇费周折，而在这个框架下这一结果可谓是“水到渠成”了。不仅如此，现在我们$p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)$也有了，那么理论上就可以完成$\boldsymbol{x}_t\sim p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)$的采样了。从式$(13)$的推导我们知道，如果做代换$\boldsymbol{z} = (\boldsymbol{x}_t - \boldsymbol{x}_0)/t$，就有

$$p(\boldsymbol{z}) \propto \frac{1}{(1 + \Vert\boldsymbol{z}\Vert^2)^{(d+1)/2}} \tag{17}$$

于是我们可以先从$p(\boldsymbol{z})$中采样，然后通过$\boldsymbol{x}_t = \boldsymbol{x}_0 + t\,\boldsymbol{z}$来得到相应的$\boldsymbol{x}_t$。至于从$p(\boldsymbol{z})$的采样，它只依赖于模长，所以我们可以通过逆累积函数法先采样模长，然后随机采样一个方向来构成采样结果，这跟先验分布的采样是完全一样的。不过，笔者在进一步研究下面的遗留问题时，发现了一个让人意外的“惊喜”！

问题重拾

在《生成扩散模型漫谈（十三）：从万有引力到扩散模型》中，我们曾指出原论文给出的采样方案是：

$$\boldsymbol{x}_t = \boldsymbol{x}_0 + \Vert\boldsymbol{\varepsilon}_{\boldsymbol{x}}\Vert(1+\tau)^m\boldsymbol{u},\quad t = |\varepsilon_t|(1+\tau)^m \tag{18}$$

其中$(\boldsymbol{\varepsilon}_{\boldsymbol{x}},\varepsilon_t)\sim\mathcal{N}(\boldsymbol{0},\sigma^2\boldsymbol{I}_{(d+1)\times(d+1)})$，$m\sim U[0,M]$，$\boldsymbol{u}$是$d$维单位球面上均匀分布的单位向量，而$\tau,\sigma,M$则都是常数。当时对这个采样的评价是“有颇多的主观性”，也就是觉得是原作者主观设计的，没太多的理由。然而，不知道作者有意还是无意，笔者发现了一个神奇的“巧合”：这个采样正好是式(17)的一个实现！

接下来我们证明这一点。首先，我们将上式后半部分代入前半部分，得到

$$\boldsymbol{x}_t=\boldsymbol{x}_0+t\times\frac{\Vert\boldsymbol{\varepsilon}_{\boldsymbol{x}}\Vert}{|\varepsilon_t|}\boldsymbol{u}\tag{19}$$

形式上已经跟上一节说的$\boldsymbol{x}_t=\boldsymbol{x}_0+t\,\boldsymbol{z}$一样了，并且$\boldsymbol{u}$也是各向同性的单位随机向量，所以问题变为$\frac{\Vert\boldsymbol{\varepsilon}_{\boldsymbol{x}}\Vert}{|\varepsilon_t|}$是否跟$\Vert\boldsymbol{z}\Vert$同分布，答案是肯定的！注意，概率密度从笛卡尔坐标变为球坐标，要多乘以一个半径$^{d-1}$，所以根据式(17)有

$$p(\Vert\boldsymbol{z}\Vert)\propto\frac{\Vert\boldsymbol{z}\Vert^{d-1}}{(1+\Vert\boldsymbol{z}\Vert^2)^{(d+1)/2}}\tag{20}$$

而根据$(\boldsymbol{\varepsilon}_{\boldsymbol{x}},\varepsilon_t)\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_{(d+1)\times(d+1)})$（由于研究的是比值，方差可以约掉，因此简单起见取$\sigma=1$）有

$$p(\Vert\boldsymbol{\varepsilon}_{\boldsymbol{x}}\Vert)\propto\Vert\boldsymbol{\varepsilon}_{\boldsymbol{x}}\Vert^{d-1}e^{-\Vert\boldsymbol{\varepsilon}_{\boldsymbol{x}}\Vert^2/2},\quad p(|\varepsilon_t|)\propto e^{-|\varepsilon_t|^2/2}\tag{21}$$

记$r=\frac{\Vert\boldsymbol{\varepsilon}_{\boldsymbol{x}}\Vert}{|\varepsilon_t|}$，则$\Vert\boldsymbol{\varepsilon}_{\boldsymbol{x}}\Vert=r|\varepsilon_t|$，然后根据概率的相等性，有

$$\begin{aligned}
p(r)dr =&\, \mathbb{E}_{|\varepsilon_t|\sim p(|\varepsilon_t|)}\left[p(\Vert\boldsymbol{\varepsilon}_{\boldsymbol{x}}\Vert = r|\varepsilon_t|)d(r|\varepsilon_t|)\right] \\
\propto&\, \mathbb{E}_{|\varepsilon_t|\sim p(|\varepsilon_t|)}\left[r^{d-1}|\varepsilon_t|^d e^{-r^2|\varepsilon_t|^2/2}dr\right] \\
\propto&\, \int_0^{\infty} r^{d-1}|\varepsilon_t|^d e^{-r^2|\varepsilon_t|^2/2}e^{-|\varepsilon_t|^2/2}d|\varepsilon_t|dr \\
=&\, \int_0^{\infty} r^{d-1}|\varepsilon_t|^d e^{-(r^2+1)|\varepsilon_t|^2/2}d|\varepsilon_t|dr \\
=&\, \frac{r^{d-1}}{(1+r^2)^{(d+1)/2}}\int_0^{\infty} s^d e^{-s^2/2}dsdr\quad\left(\text{设}s = |\varepsilon_t|\sqrt{r^2+1}\right) \\
\propto&\, \frac{r^{d-1}}{(1+r^2)^{(d+1)/2}}dr
\end{aligned}\tag{22}$$

因此$p(r)\propto\frac{r^{d-1}}{(1+r^2)^{(d+1)/2}}$，跟$(20)$完全一致。所以，$\frac{\Vert\boldsymbol{\varepsilon}_{\boldsymbol{x}}\Vert}{|\varepsilon_t|}\boldsymbol{u}$确实提供了$\boldsymbol{z}$的一种有效采样方式，这在实现上要比逆累积函数法简单得多，但原论文并没有提及这一点。

时空分离

刚才我们求解了$(t,\boldsymbol{x}_t)$组成的$d+1$维空间中的各向同性解，其实某种意义上来说，这算是最简单的一个解。可能这种说明有些读者难以接受，毕竟这个万有引力扩散模型在数学上看上去明显复杂得多。但事实上，在求解数学物理方程时，很多时候各向同性解确实是作为最简单的解来试探求解的。

当然，将$(t,\boldsymbol{x}_t)$看成“时-空”整体的各向同性，在理解上确实没那么直观，我们更习惯的是理解空间上的各向同性，将时间维度独立开来，这一节就在这个假设下求解。

假设求解

也就是说，这部分的“各向同性”，指的是在$\boldsymbol{x}_t$的d维空间中的各向同性，$\boldsymbol{G}(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)$被分解为$(G_1(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0),\boldsymbol{G}_{>1}(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0))$两部分来理解。其中$G_1(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)$只是一个标量，各向同性意味着它只依赖于$r=\Vert\boldsymbol{x}_t-\boldsymbol{x}_0\Vert$，我们将它记为$\phi_t(r)$；$\boldsymbol{G}_{>1}(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)$是一个$d$维向量，各向同性意味着$\boldsymbol{G}(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)$指向源点$\boldsymbol{x}_0$，且模长只依赖于$r=\Vert\boldsymbol{x}_t-\boldsymbol{x}_0\Vert$，因此可以设

$$\boldsymbol{G}_{>1}(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0) = \varphi_t(r)(\boldsymbol{x}_t - \boldsymbol{x}_0) \tag{23}$$

于是

$$\begin{aligned}
0 =&\, \frac{\partial}{\partial t}\phi_t(r) + \nabla_{\boldsymbol{x}_t}\cdot\big(\varphi_t(r)(\boldsymbol{x}_t - \boldsymbol{x}_0)\big) \\
=&\, \frac{\partial}{\partial t}\phi_t(r) + r\frac{\partial}{\partial r}\varphi_t(r) + d\,\varphi_t(r) \\
=&\, \frac{\partial}{\partial t}\phi_t(r) + \frac{1}{r^{d-1}}\frac{\partial}{\partial r}\big(\varphi_t(r) r^d\big)
\end{aligned} \tag{24}$$

这里有两个待定函数$\phi_t(r)$、$\varphi_t(r)$，但只有一个方程，所以求解就更简单了。由于约束条件约束的是$\boldsymbol{G}_1(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0)$，也就是$\phi_t(r)$而不是$\varphi_t(r)$，所以简单起见通常是给定满足条件的$\phi_t(r)$来求解$\varphi_t(r)$，结果是

$$\varphi_t(r) = -\frac{1}{r^d}\int\frac{\partial}{\partial t}\phi_t(r) r^{d-1} dr = -\frac{1}{r^d}\frac{\partial}{\partial t}\int\phi_t(r) r^{d-1} dr \tag{25}$$

高斯扩散

这部分我们来表明，常见的基于高斯分布假设的ODE扩散模型，也是式(25)的一个特例。对于高斯分布假设，有

$$\boldsymbol{G}_1(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0) = p_t(\boldsymbol{x}_t|\boldsymbol{x}_0) = \frac{1}{(2\pi\sigma_t^2)^{d/2}}e^{-\Vert\boldsymbol{x}_t - \boldsymbol{x}_0\Vert^2/2\sigma_t^2} \tag{26}$$

即$\phi_t(r) = \frac{1}{(2\pi\sigma_t^2)^{d/2}}e^{-r^2/2\sigma_t^2}$，其中$\sigma_t$是关于$t$的单调递增函数，满足$\sigma_0 = 0$且$\sigma_T$足够大，$\sigma_0 = 0$是为了成立初值条件，$\sigma_T$足够大是为了先验分布与数据无关，至于积分等于1的约束，这是高斯分布的基本性质，自然满足。

代入式(25)后解得：

$$\varphi_t(r) = \frac{\dot{\sigma}_t}{(2\pi\sigma_t^2)^{d/2}\sigma_t}e^{-r^2/2\sigma_t^2} = \frac{\dot{\sigma}_t}{\sigma_t}\phi_t(r) \tag{27}$$

其中r的积分涉及到不完全伽马函数，比较复杂，笔者是直接用Mathematica算的。有了这个结果后，我们有

$$\begin{aligned}
\boldsymbol{u}_1(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0) =&\, \int p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)p_0(\boldsymbol{x}_0)d\boldsymbol{x}_0 = p_t(\boldsymbol{x}_t) \\
\boldsymbol{u}_{>1}(t,0;\boldsymbol{x}_t,\boldsymbol{x}_0) =&\, \int \frac{\dot{\sigma}_t}{\sigma_t}(\boldsymbol{x}_t - \boldsymbol{x}_0)p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)p_0(\boldsymbol{x}_0)d\boldsymbol{x}_0 \\
=&\, -\dot{\sigma}_t\sigma_t\int \nabla_{\boldsymbol{x}_t}p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)p_0(\boldsymbol{x}_0)d\boldsymbol{x}_0 \\
=&\, -\dot{\sigma}_t\sigma_t\nabla_{\boldsymbol{x}_t}p_t(\boldsymbol{x}_t)
\end{aligned} \tag{28}$$

从而根据式(5)有

$$\boldsymbol{f}_t(\boldsymbol{x}_t) = \frac{\boldsymbol{u}_{>1}(t,\boldsymbol{x}_t)}{\boldsymbol{u}_1(t,\boldsymbol{x}_t)} = -\dot{\sigma}_t\sigma_t\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t) \tag{29}$$

这些结果跟《生成扩散模型漫谈（十二）：“硬刚”扩散ODE》的完全一致，剩下的处理细节，也可以参考该文章。

逆向构造

像刚才那样给定$\phi_t(r)$来求解$\varphi_t(r)$的做法在理论上很简单，但在实践上会有两个困难：1、$\phi_t(r)$既要满足初值条件，又要满足积分条件，不是那么容易构造的；2、对r的积分也不一定有简单的初等形式。既然如此，我们可以想一个逆向构造的方法。

我们知道，$\phi_t(r)$是在笛卡尔坐标下的概率密度，换到球坐标下要乘以$C_d r^{d-1}$，C_d是某个常数（跟d有关），根据式(5)，最终结果是一个比值，不受常数影响，所以简单起见我们忽略这个常数，而忽略常数后正好是式(25)的被积函数，所以式(25)中的积分

$$\int \phi_t(r)r^{d-1}dr \tag{30}$$

正好是一个累积概率函数（更准确说，是累积概率函数的$1/C_d$再加上一个常数，但我们已经忽略掉无关紧要的常数），而从概率密度算累积概率不一定容易，但从累积概率

算概率密度很简单（求导），所以我们可以先构造累积概率函数，然后再去求相应的$\phi_t(r), \varphi_t(r)$，这样就免去了积分的困难。

具体来说，构造累积概率函数$\psi_t(r)$，满足如下条件：

> 1、$\psi_t(0)=0$，$\psi_t(\infty)=1$；
>
> 2、$\psi_t(r)$关于r单调递增；
>
> 3、$\forall r>0, \lim\limits_{t\to 0^+}\psi_t(r)=1$。

稍微研究过激活函数的同学，应该不难构造满足上述条件的函数，它其实这就是“阶跃函数”的光滑近似，比如$\tanh\left(\frac{r}{t}\right)$、$1-e^{-r/t}$等。有了$\psi_t(r)$后，根据式$(25)$，我们就有

$$\phi_t(r)=\frac{1}{r^{d-1}}\frac{\partial}{\partial r}\psi_t(r),\quad \varphi_t(r)=-\frac{1}{r^d}\frac{\partial}{\partial t}(\psi_t(r)+\lambda_t) \tag{31}$$

其中λ_t是t的任意函数，一般情况下可以直接设为0。当然，这些各向同性解本质上都是等价的，包括前一节推导的“万有引力扩散”也是如此，它们都可以纳入上式之中，也可以通过坐标变换相互推导，这是因为上式只依赖于一个一元的累积概率函数$\psi_t(r)$，不同分布之间的累积概率函数一般都可以相互变换（它们都是形态良好的单调递增函数）。

文章小结

本文构建了一个ODE式扩散的一般框架，理论上来说，所有的ODE式扩散模型可以纳入到该框架之中，我们也可以从中推导出各种新奇的、奇葩的ODE式扩散模型，比如目前的推导都是基于各向同性假设的，其实也可以将各向同性的$\varphi(R)$换成更一般的$\varphi(t;\boldsymbol{x}_t,\boldsymbol{x}_0)$，这可以利用《一阶偏微分方程的特征线法》的方法来完成求解，得到一簇新的模型。总的来说，这是一个名副其实的ODE式扩散模型的“生产车间”。

可能有读者想问，我不就想要一个可用的生成扩散模型而已，你搞那么多花里花俏的变体又有什么价值？事实上，跟之前《f-GAN简介：GAN模型的生产车间》、《Designing GANs：又一个GAN生产车间》一样，我们希望发现、掌握生成模型的构建规律，以便进一步理解生成模型的关键，从而发现更有效的生成模型，这是一个追求完美的永无止境的过程。

之前“万有引力扩散”论文中的实验结果已经表明，作为一个ODE式扩散模型，它要比高斯扩散的效果要好些。这就说明，即便是基于各向同性假设，这些数学本质等价的扩散模型在实践上依然会有效果差异。所以，如何更好地结合实验细节来回答“什么样的设计才是更好的扩散模型”，将会是未来的一个非常有意义的研究问题。

如果您需要引用本文，请参考：

苏剑林. (Dec. 15, 2022). 《生成扩散模型漫谈（十四）：构建ODE的一般步骤（上）》[Blog post]. Retrieved from https://spaces.ac.cn/archives/9370

```
@online{kexuefm-9370,
        title={生成扩散模型漫谈（十四）：构建ODE的一般步骤（上）},
        author={苏剑林},
        year={2022},
        month={Dec},
        url={\url{https://spaces.ac.cn/archives/9370}},
}
```