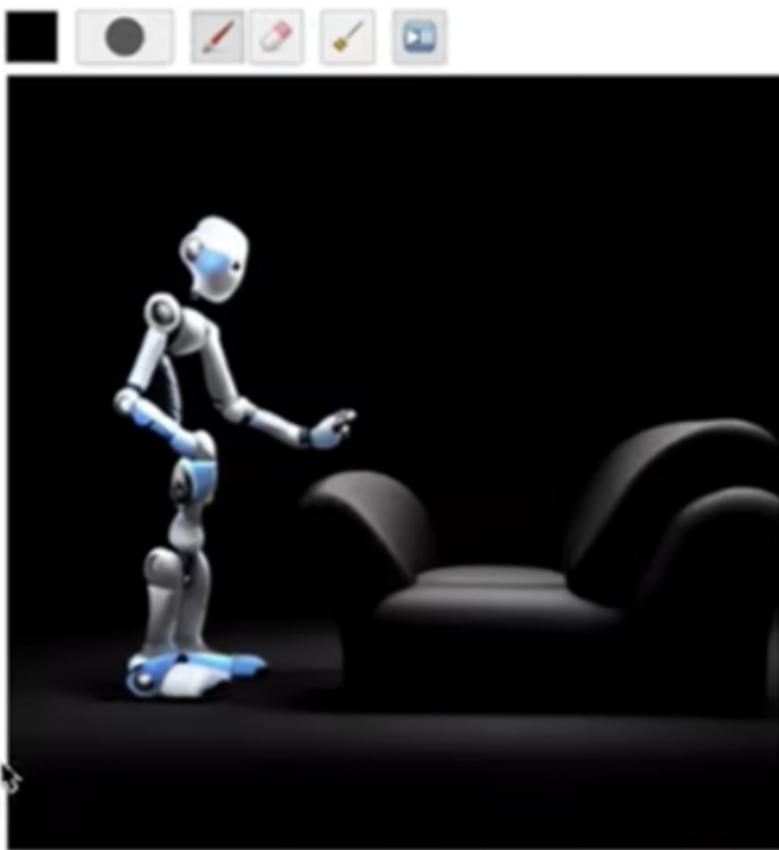


# Score-based Diffusion Model

BY ZHIJUN ZENG

2023年4月9日

# 1 What can diffusion model do



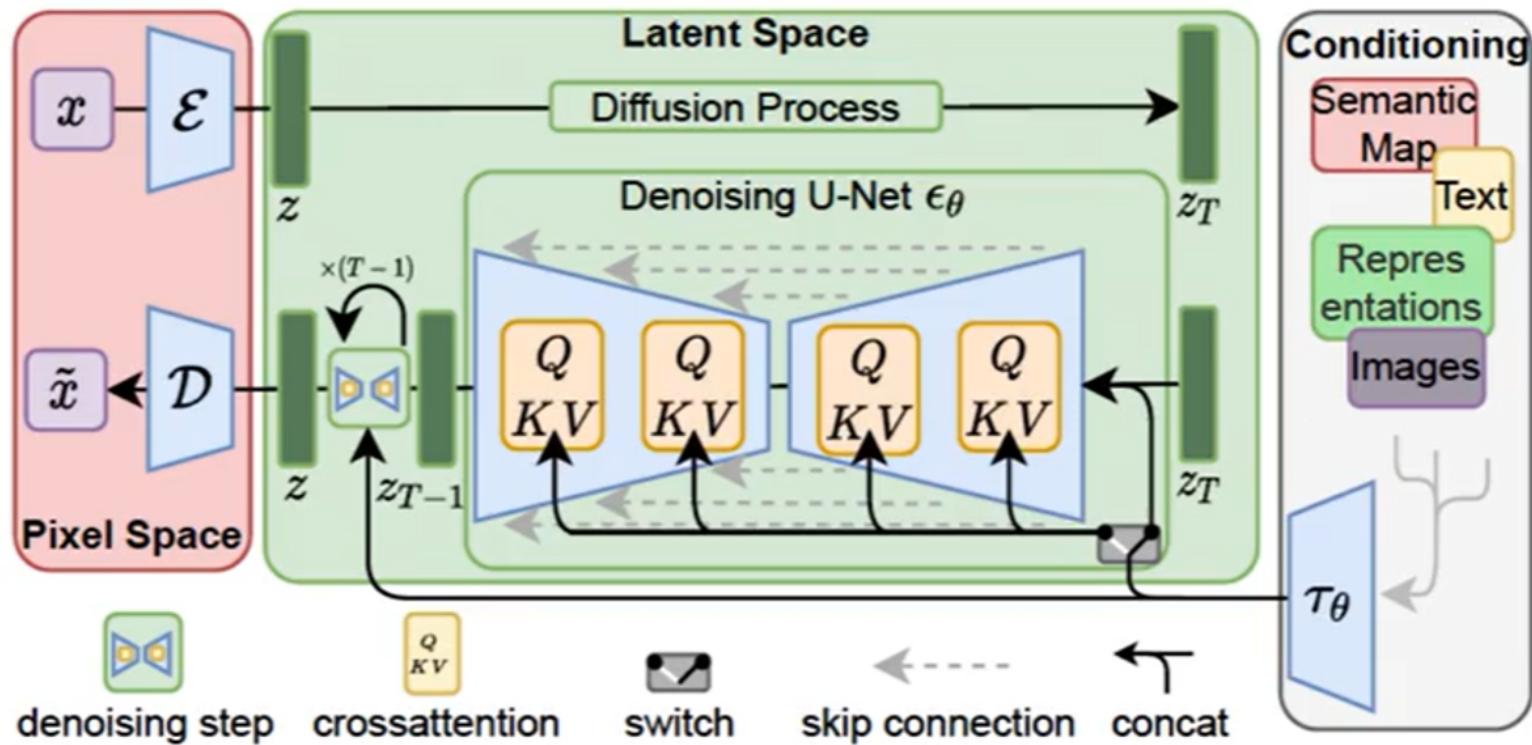
AI robot is thinking on th

diffuse the f rest

pro tip: upload img by dropping on the canvas

## Stable Diffusion原理

模型结构图：



## 2 分数匹配与朗之万采样

假设数据为  $\{\mathbf{x}_i \in \mathbb{R}^D\}_{i=1}^N$  来自分布  $p_{\text{data}}(\mathbf{x})$ , 定义一个概率密度函数的分数为  $\nabla_{\mathbf{x}} \log p(\mathbf{x})$ 。 分数匹配生成模型通过分数网络  $s_\theta: \mathbb{R}^D \rightarrow \mathbb{R}^D$  逼近  $p_{\text{data}}(\mathbf{x})$  的分数, 利用 Langevin 动力学进行采样。

Langevin dynamics 可以通过分数产生给定分布  $p(\mathbf{x})$  的样本

$$\tilde{\mathbf{x}}_t = \tilde{\mathbf{x}}_{t-1} + \frac{\varepsilon}{2} \nabla_{\mathbf{x}} \log p(\tilde{\mathbf{x}}_{t-1}) + \sqrt{\varepsilon} z_t$$

其中  $z(t) \sim \mathcal{N}(0, I)$ , 分布  $\tilde{\mathbf{x}}_T$  收敛于  $p(\mathbf{x})$  当  $\varepsilon \rightarrow 0, T \rightarrow \infty$ .

## 2.1 分数匹配

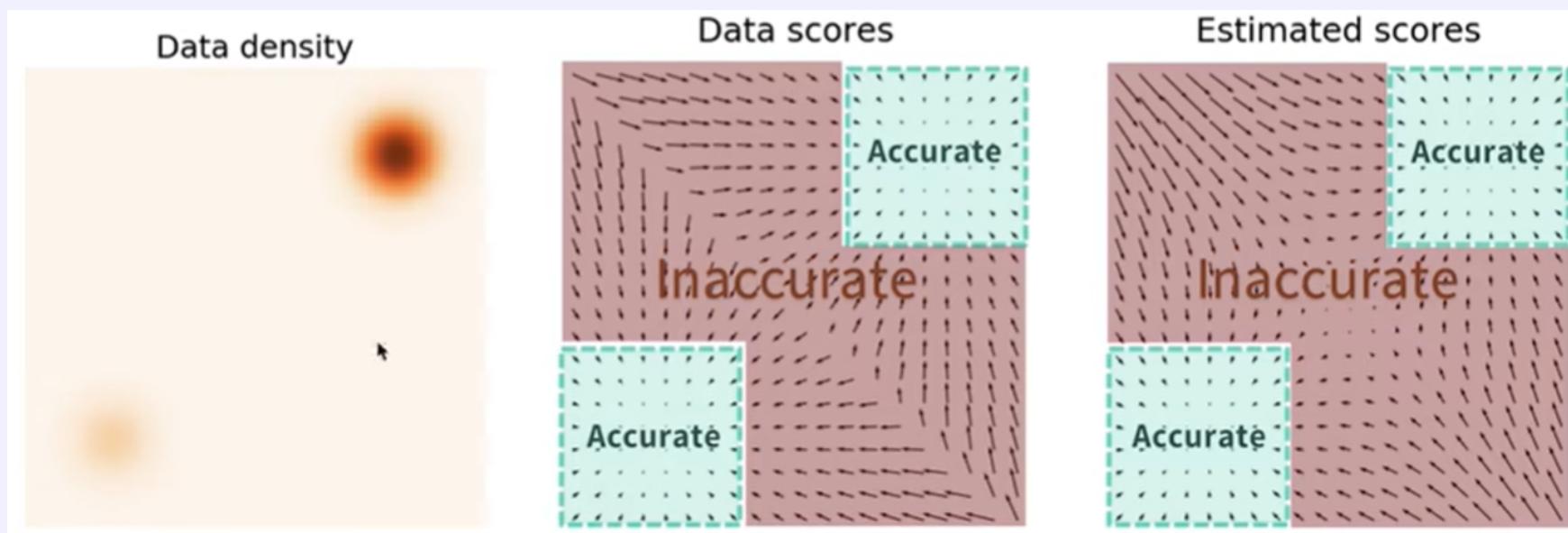
分数匹配的目标就是分数网络  $s_\theta: \mathbb{R}^D \rightarrow \mathbb{R}^D$  逼近  $p_{\text{data}}(\mathbf{x})$  的分数  $\nabla_{\mathbf{x}} \log p_{\text{data}}(\mathbf{x})$ . 最简单的方法就是

$$\min \frac{1}{2} \mathbb{E}_{p_{\text{data}}} [\| s_\theta(\mathbf{x}) - \nabla_{\mathbf{x}} \log p_{\text{data}}(\mathbf{x}) \|_2^2]$$

它等价于

$$\min \mathbb{E}_{p_{\text{data}}} \left[ \text{tr}(\nabla_{\mathbf{x}} s_\theta(\mathbf{x})) + \frac{1}{2} \| s_\theta(\mathbf{x}) \|_2^2 \right]$$

这个方法需要求 Jacobian, 计算代价非常大。

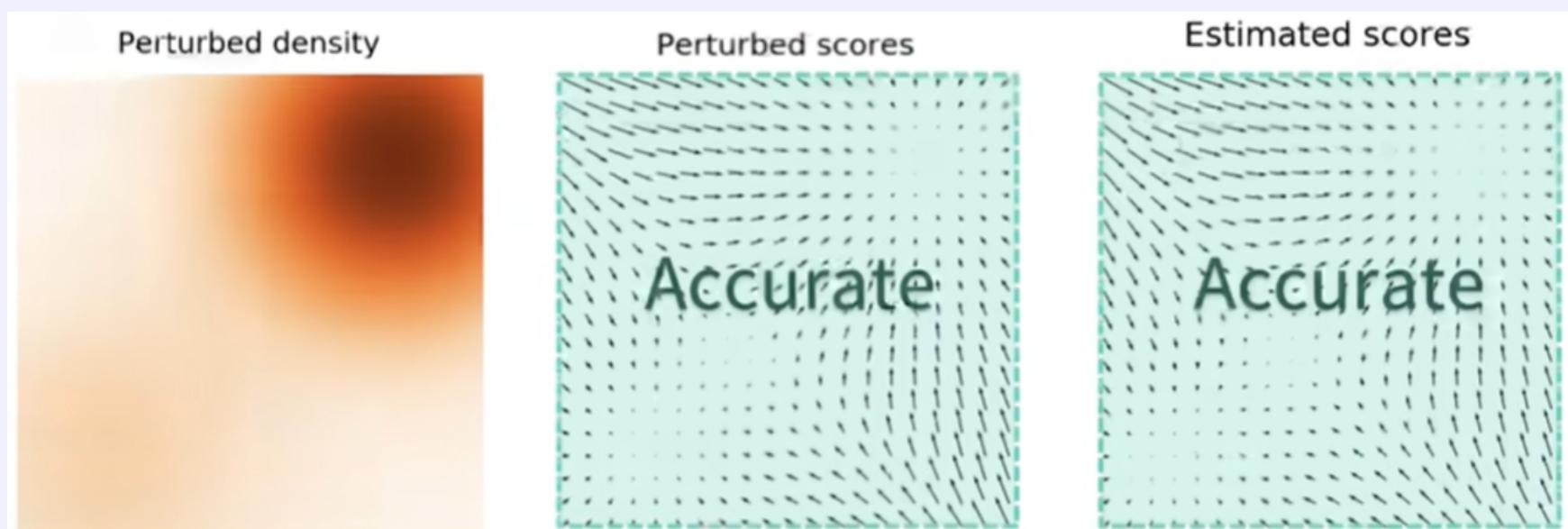


## 2.2 降噪分数匹配

降噪分数匹配用于解决计算代价大的问题，考虑对数据进行一个特定的扰动  $q_\sigma(\tilde{\mathbf{x}}|\mathbf{x})$ ，然后用分数匹配对扰动后的数据  $q_\sigma(\tilde{\mathbf{x}}) = \int q_\sigma(\tilde{\mathbf{x}}|\mathbf{x}) p_{\text{data}}(\mathbf{x}) d\mathbf{x}$  的分数进行匹配。它的目标函数可以证明等价于

$$\frac{1}{2} \mathbb{E}_{q_\sigma(\tilde{\mathbf{x}}|\mathbf{x}) p_{\text{data}}(\mathbf{x})} [\| s_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} q_\sigma(\tilde{\mathbf{x}}|\mathbf{x}) \|_2^2]$$

目标函数的最小元满足  $s_\theta(\mathbf{x}) = \nabla_{\mathbf{x}} \log q_\sigma(\mathbf{x})$ ，当  $\sigma \rightarrow 0$  时收敛于真实解。



## 2.3 降噪分数匹配的问题

小噪声：保持数据形态，但不利于学习分数，导致朗之万采样不准确

大噪声：利于学习分数，但采样图像是加噪后的

解决方式：逐渐加噪，从大噪声逐渐降噪为小噪声.

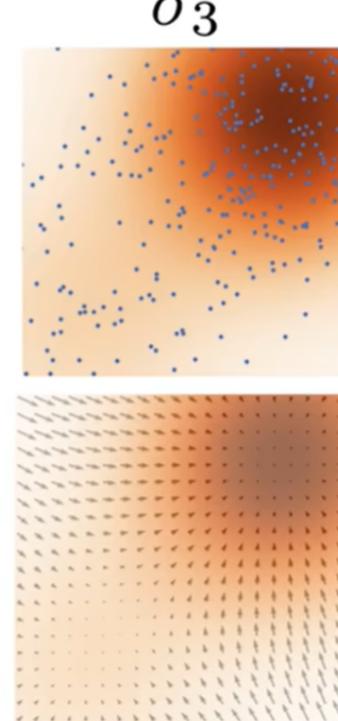
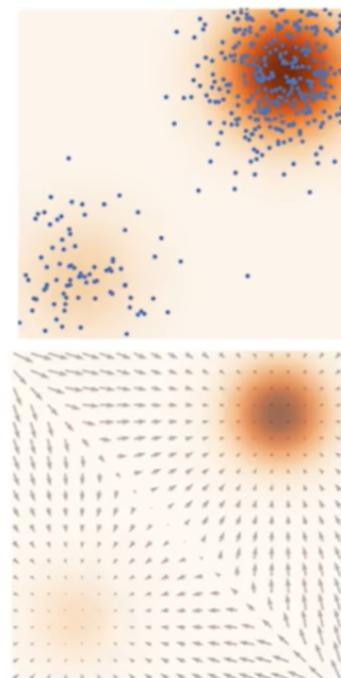
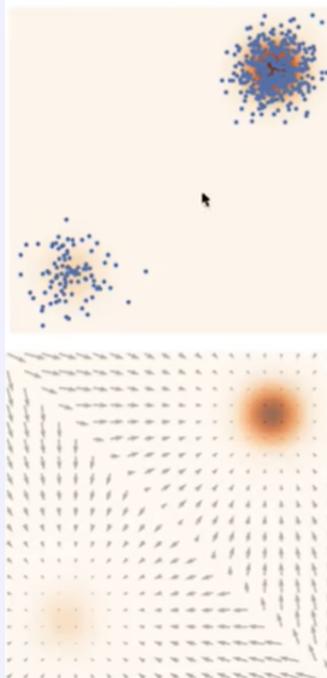
$$\sigma_1$$

<

$$\sigma_2$$

<

$$\sigma_3$$



## 2.4 Noise Conditional Score Network(NCSN)

令 $\{\sigma_i\}_{i=1}^L$ 为几何级数列 $\frac{\sigma_1}{\sigma_2} = \dots = \frac{\sigma_{L-1}}{\sigma_L} > 1$ ,  $q_\sigma(\mathbf{x}) = \int p_{\text{data}}(\mathbf{t}) \mathcal{N}(\mathbf{x} | \mathbf{t}, \sigma^2 I) d\mathbf{t}$ , 希望用一个 Noise Conditional Score Network  $\mathbf{s}_\theta(\mathbf{x}, \sigma)$ , 使得  $\mathbf{s}_\theta(\mathbf{x}, \sigma) \approx \nabla_{\mathbf{x}} \log q_\sigma(\mathbf{x})$ ,  $\forall \sigma_i$ . 此时

$$l(\theta; \sigma) = \frac{1}{2} \mathbb{E}_{p_{\text{data}}(\mathbf{x})} \mathbb{E}_{\tilde{\mathbf{x}} \sim \mathcal{N}(\mathbf{x}, \sigma^2 I)} \left[ \left\| \mathbf{s}_\theta(\tilde{\mathbf{x}}) + \frac{\tilde{\mathbf{x}} - \mathbf{x}}{\sigma^2} \right\|_2^2 \right]$$

$$\mathbf{L}(\theta; \{\sigma_i\}^L) = \frac{1}{L} \sum_{i=1}^L \lambda(\sigma_i) l(\theta; \sigma)$$

---

**Algorithm 1** Annealed Langevin dynamics.

---

**Require:**  $\{\sigma_i\}_{i=1}^L, \epsilon, T$ .

- 1: Initialize  $\tilde{\mathbf{x}}_0$
- 2: **for**  $i \leftarrow 1$  to  $L$  **do**
- 3:      $\alpha_i \leftarrow \epsilon \cdot \sigma_i^2 / \sigma_L^2$      ▷  $\alpha_i$  is the step size.
- 4:     **for**  $t \leftarrow 1$  to  $T$  **do**
- 5:         Draw  $\mathbf{z}_t \sim \mathcal{N}(0, I)$
- 6:          $\tilde{\mathbf{x}}_t \leftarrow \tilde{\mathbf{x}}_{t-1} + \frac{\alpha_i}{2} \mathbf{s}_\theta(\tilde{\mathbf{x}}_{t-1}, \sigma_i) + \sqrt{\alpha_i} \mathbf{z}_t$
- 7:     **end for**
- 8:      $\tilde{\mathbf{x}}_0 \leftarrow \tilde{\mathbf{x}}_T$
- 9: **end for**
- return**  $\tilde{\mathbf{x}}_T$

---

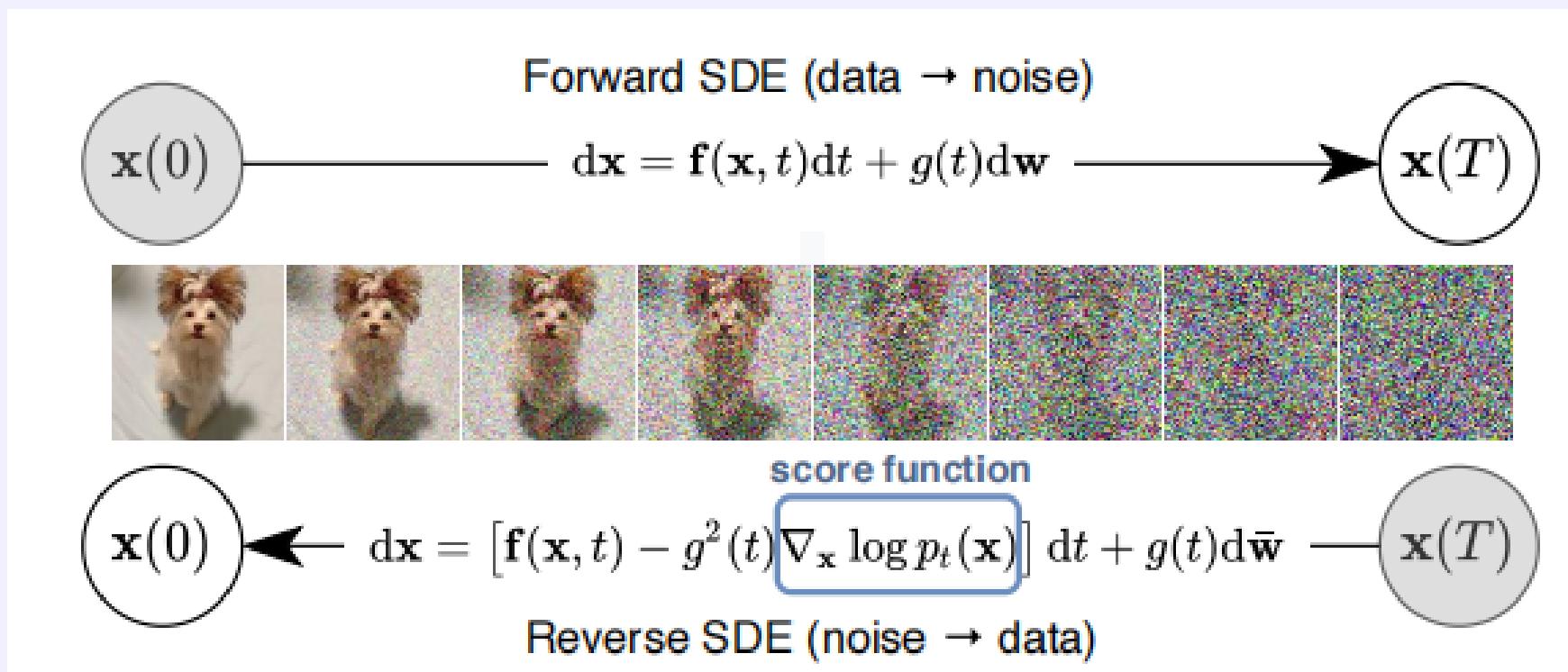
### 3 随机微分方程

连续时间下的加噪过程可以用SDE来描述

$$d\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_t)dt + \mathbf{g}_t(\mathbf{x}_t)dW_t$$

Euler Maruyama离散格式为

$$\mathbf{x}_{t+\Delta t} - \mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_t)\Delta t + \mathbf{g}_t(\mathbf{x}_t)\sqrt{\Delta t}\varepsilon, \varepsilon \sim \mathcal{N}(0, I)$$



### 3.1 Backward Equation

利用离散形式的Euler Maruyama格式形式上的推导SDE的逆向方程，其条件概率为

$$p(\mathbf{x}_{t+\Delta t} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t+\Delta t}; \mathbf{x}_t + \mathbf{f}_t(\mathbf{x}_t) \Delta t, g_t^2 \Delta t \mathbf{I})$$
$$\propto \exp\left(-\frac{|\mathbf{x}_{t+\Delta t} - \mathbf{x}_t - \mathbf{f}_t(\mathbf{x}_t) \Delta t|^2}{2 g_t^2 \Delta t}\right)$$

利用Bayesian Formula

$$p(\mathbf{x}_t | \mathbf{x}_{t+\Delta t}) = \frac{p(\mathbf{x}_{t+\Delta t} | \mathbf{x}_t) p(\mathbf{x}_t)}{p(\mathbf{x}_{t+\Delta t})} = p(\mathbf{x}_{t+\Delta t} | \mathbf{x}_t) \exp(\log p(\mathbf{x}_t) - \log p(\mathbf{x}_{t+\Delta t}))$$
$$\propto \exp\left(-\frac{|\mathbf{x}_{t+\Delta t} - \mathbf{x}_t - \mathbf{f}_t(\mathbf{x}_t) \Delta t|^2}{2 g_t^2 \Delta t} + \log p(\mathbf{x}_t) - \log p(\mathbf{x}_{t+\Delta t})\right)$$

当 $\Delta t$ 足够小，只有当 $\mathbf{x}_{t+\Delta t} - \mathbf{x}_t$ 足够接近，该式才会非0，故

$$\log p(\mathbf{x}_{t+\Delta t}) \approx \log p(\mathbf{x}_t) + (\mathbf{x}_{t+\Delta t} - \mathbf{x}_t) \nabla_{\mathbf{x}_t} \log(p(\mathbf{x}_t)) + \Delta t \frac{\partial}{\partial t} \log p(\mathbf{x}_t) + \dots$$

代入上式得到 $p(\mathbf{x}_t | \mathbf{x}_{t+\Delta t}) \approx$

$$\approx \exp \left( - \frac{|\mathbf{x}_{t+\Delta t} - \mathbf{x}_t - [\mathbf{f}_t(\mathbf{x}_t) - g_t^2 \nabla_{\mathbf{x}_t} \log(p(\mathbf{x}_t))] \Delta t|^2}{2 g_t^2 \Delta t} + o(\Delta t) \right)$$

$$\approx \exp \left( - \frac{|\mathbf{x}_t - \mathbf{x}_{t+\Delta t} - [\mathbf{f}_{t+\Delta t}(\mathbf{x}_{t+\Delta t}) - g_{t+\Delta t}^2 \nabla_{\mathbf{x}_{t+\Delta t}} \log(p(\mathbf{x}_{t+\Delta t}))] \Delta t|^2}{2 g_{t+\Delta t}^2 \Delta t} \right)$$

近似于一个

$\mathcal{N}(\mathbf{x}_t - \mathbf{x}_{t+\Delta t} - [\mathbf{f}_{t+\Delta t}(\mathbf{x}_{t+\Delta t}) - g_{t+\Delta t}^2 \nabla_{\mathbf{x}_{t+\Delta t}} \log(p(\mathbf{x}_{t+\Delta t}))] \Delta t, g_{t+\Delta t}^2 \Delta t \mathbf{I})$ ，  
故对应SDE为

$$d\mathbf{x}_t = [\mathbf{f}_t(\mathbf{x}_t) - g_t(\mathbf{x}_t)^2 \nabla_{\mathbf{x}} \log(p_t(\mathbf{x}_t))] dt + \mathbf{g}_t dW_t$$

## 4 得分匹配

如果知道得分函数  $\nabla_{\mathbf{x}} \text{logp}_t(\mathbf{x})$ , 那么就可以通过离散格式

$$\begin{aligned} \mathbf{x}_t - \mathbf{x}_{t+\Delta t} &= -[\mathbf{f}_{t+\Delta t}(\mathbf{x}_{t+\Delta t}) - \mathbf{g}_{t+\Delta t}^2(\mathbf{x}_{t+\Delta t}) \nabla_{\mathbf{x}_{t+\Delta t}} \text{log}(p(\mathbf{x}_{t+\Delta t}))] \Delta t + \\ &\quad \mathbf{g}_{t+\Delta t}(\mathbf{x}_{t+\Delta t}) \sqrt{\Delta t} \varepsilon \end{aligned}$$

逐步反演，推理生成扩散模型。

由

$$p(\mathbf{x}_t) = \int p(\mathbf{x}_t | \mathbf{x}_0) \tilde{p}(\mathbf{x}_0) d\mathbf{x}_0 = \mathbb{E}_{\mathbf{x}_0}[p(\mathbf{x}_t | \mathbf{x}_0)]$$

$$\nabla_{\mathbf{x}_t} \text{log}(p(\mathbf{x}_t)) = \frac{\mathbb{E}_{\mathbf{x}_0}[\nabla_{\mathbf{x}_t} p(\mathbf{x}_t | \mathbf{x}_0)]}{\mathbb{E}_{\mathbf{x}_0}[p(\mathbf{x}_t | \mathbf{x}_0)]} = \frac{\mathbb{E}_{\mathbf{x}_0}[p(\mathbf{x}_t | \mathbf{x}_0) \nabla_{\mathbf{x}_t} \text{log}(p(\mathbf{x}_t | \mathbf{x}_0))]}{\mathbb{E}_{\mathbf{x}_0}[p(\mathbf{x}_t | \mathbf{x}_0)]}$$

最后式子为  $\nabla_{\mathbf{x}_t} p(\mathbf{x}_t | \mathbf{x}_0)$  的加权平均形式

我们期望利用NN学习一个接近 $\nabla_{\mathbf{x}_t} \log(p(\mathbf{x}_t))$ 的 $s_\theta(\mathbf{x}_t, t)$ , 可以利用最小化其加权平均

$$\frac{\mathbb{E}_{\mathbf{x}_0}[p(\mathbf{x}_t|\mathbf{x}_0) \|\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{x}_0) - s_\theta(\mathbf{x}_t, t)\|^2]}{\mathbb{E}_{\mathbf{x}_0}[p(\mathbf{x}_t|\mathbf{x}_0)]}$$

最后我们去掉调节损失权重的分母, 对 $\mathbf{x}_t$ 积分, 得到最终的损失函数

$$\begin{aligned} & \int \mathbb{E}_{\mathbf{x}_0} [p(\mathbf{x}_t|\mathbf{x}_0) | s_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{x}_0)|^2] d\mathbf{x}_t \\ &= \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_t \sim p(\mathbf{x}_t|\mathbf{x}_0), \bar{p}(\mathbf{x}_0)} [|s_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{x}_0)|^2] \end{aligned}$$

## 4.1 Example

The data distribution  $p_0(\mathbf{x})$ , the prior distribution  $p_T(\mathbf{x})$ , we choose the following SDE

$$d\mathbf{x} = \sigma_t dW_t, t \in [0, 1]$$

In this case

$$p(\mathbf{x}(t) | \mathbf{x}(0)) = \mathcal{N}\left(\mathbf{x}(t); \mathbf{x}(0), \frac{1}{2\log\sigma}(\sigma^{2t} - 1)I\right)$$

Then the gradient  $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t | \mathbf{x}_0)$  is easily calculated.



## 5 SDE diffusion model 一般流程

通过SDE定义前向过程

求 $p(x_t | x_0)$ 表达式

通过损失函数训练 $s_\theta(x_t, t)$

用训练好的得分函数 $s_\theta(x_t, t)$ 替换 $\nabla_{x_t} \log(p_t(x))$  进行反向生成。

## 5.1 统一SDE diffusion model 与 DDPM

定义

$$p(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \bar{\alpha}_t \mathbf{x}_0, \bar{\beta}_t^2 \mathbf{I})$$

假设起点为  $t = 0$ , 终点为  $t = 1$ , 那么

$$\begin{cases} \bar{\alpha}_0 = 1, \bar{\alpha}_1 = 0 \\ \bar{\beta}_0 = 0, \bar{\beta}_1 = 1 \end{cases}$$

有了  $p(\mathbf{x}_t | \mathbf{x}_0)$ , 我们反推SDE格式, 即要求  $p(\mathbf{x}_{t+\Delta t} | \mathbf{x}_t)$  满足

$$p(\mathbf{x}_{t+\Delta t} | \mathbf{x}_0) = \int p(\mathbf{x}_{t+\Delta t} | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{x}_0) d\mathbf{x}_t$$

我们考虑线性形式的解

$$d\mathbf{x}_t = \mathbf{f}_t \mathbf{x}_t dt + \mathbf{g}_t dW_t$$

记号	含义	采样
$p(\mathbf{x}_{t+\Delta t} \mathbf{x}_0)$	$\mathcal{N}(\mathbf{x}_t; \bar{\alpha}_{t+\Delta t}\mathbf{x}_0, \bar{\beta}_{t+\Delta t}^2 \mathbf{I})$	$\mathbf{x}_{t+\Delta t} = \bar{\alpha}_{t+\Delta t}\mathbf{x}_0 + \bar{\beta}_{t+\Delta t}\boldsymbol{\varepsilon}$
$p(\mathbf{x}_t \mathbf{x}_0)$	$\mathcal{N}(\mathbf{x}_t; \bar{\alpha}_t\mathbf{x}_0, \bar{\beta}_t^2 \mathbf{I})$	$\mathbf{x}_t = \bar{\alpha}_t\mathbf{x}_0 + \bar{\beta}_t\boldsymbol{\varepsilon}_1$
$p(\mathbf{x}_{t+\Delta t} \mathbf{x}_t)$	$\mathcal{N}(\mathbf{x}_{t+\Delta t}; (1 + f_t \Delta t)\mathbf{x}_t, g_t^2 \Delta t \mathbf{I})$	$\mathbf{x}_{t+\Delta t} = (1 + f_t \Delta t)\mathbf{x}_t + g_t \Delta t \boldsymbol{\varepsilon}_2$
$\int p(\mathbf{x}_{t+\Delta t} \mathbf{x}_t) p(\mathbf{x}_t \mathbf{x}_0) d\mathbf{x}_t$		$\begin{aligned} & \mathbf{x}_{t+\Delta t} \\ &= (1 + f_t \Delta t)\mathbf{x}_t + g_t \sqrt{\Delta t} \boldsymbol{\varepsilon}_2 \\ &= (1 + f_t \Delta t)(\bar{\alpha}_t \mathbf{x}_0 + \bar{\beta}_t \boldsymbol{\varepsilon}_1) + g_t \sqrt{\Delta t} \boldsymbol{\varepsilon}_2 \\ &= (1 + f_t \Delta t)\bar{\alpha}_t \mathbf{x}_0 + ((1 + f_t \Delta t)\bar{\beta}_t \boldsymbol{\varepsilon}_1 + g_t \sqrt{\Delta t} \boldsymbol{\varepsilon}_2) \end{aligned}$

由此可得,

$$\bar{\alpha}_{t+\Delta t} = (1 + f_t \Delta t)\bar{\alpha}_t$$

$$\bar{\beta}_{t+\Delta t} = (1 + f_t \Delta t)\bar{\beta}_t^2 + g_t^2 \Delta t$$

令  $\Delta t \rightarrow 0$  得到微分方程

$$f_t = \frac{d}{dt} \ln(\bar{\alpha}_t) = \frac{1}{\bar{\alpha}_t} \frac{d\bar{\alpha}_t}{dt}, g_t^2 = 2\bar{\alpha}_t \bar{\beta}_t \frac{d}{dt} \left( \frac{\bar{\beta}_t}{\bar{\alpha}_t} \right)$$

此时

$$\nabla_{\mathbf{x}_t} \log(p(\mathbf{x}_t | \mathbf{x}_0)) = -\frac{\mathbf{x}_t - \bar{\alpha}_t \mathbf{x}_0}{\bar{\beta}_t} = -\frac{\boldsymbol{\varepsilon}}{\bar{\beta}_t}$$

第二个等号因为  $\mathbf{x}_t = \bar{\alpha}_t \mathbf{x}_0 + \bar{\beta}_t \boldsymbol{\varepsilon}$ , 为了对其以往的结果可以设  $s_\theta(\mathbf{x}_t, t) = -\frac{\boldsymbol{\varepsilon}_\theta(\mathbf{x}_t, t)}{\bar{\beta}_t}$ , 此时

$$\begin{aligned} & \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_t \sim p(\mathbf{x}_t | \mathbf{x}_0), \bar{p}(\mathbf{x}_0)} [ \| s_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t | \mathbf{x}_0) \|^2 ] = \\ & \frac{1}{\bar{\beta}_t} \mathbb{E}_{\mathbf{x}_0 \sim p(\mathbf{x}_0), \boldsymbol{\varepsilon} \sim \mathcal{N}(0, I)} [ \| \boldsymbol{\varepsilon}_\theta(\bar{\alpha}_t \mathbf{x}_0 + \bar{\beta}_t \boldsymbol{\varepsilon}, t) - \boldsymbol{\varepsilon} \|^2 ] \end{aligned}$$

是DDPM 的损失函数。

## 6 ODE 加速采样

对前向方程

$$d\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_t)dt + \mathbf{g}_t(\mathbf{x}_t)dW_t$$

其概率密度函数满足FP方程

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

对于任意一个满足  $\sigma_t^2 \leq g_t^2$  的函数  $\sigma_t$ , FP 方程完全等价于

$$\frac{\partial}{\partial t} p_t(\mathbf{x}) = -\nabla_{\mathbf{x}} \left[ \left( \mathbf{f}_t(\mathbf{x}) - \frac{1}{2} (\mathbf{g}_t^2 - \sigma_t^2) \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) \right) p_t(\mathbf{x}) \right] + \frac{1}{2} \sigma_t^2 \nabla_{\mathbf{x}} \cdot \nabla_{\mathbf{x}} p_t(\mathbf{x})$$

于是此时对应的正向方程为

$$d\boldsymbol{x}_t = \left( f_t(\boldsymbol{x}) - \frac{1}{2}(\boldsymbol{g}_t^2 - \sigma_t^2) \nabla_{\boldsymbol{x}} \text{logp}_t(\boldsymbol{x}) \right) dt + \sigma_t dW_t$$

它的反向方程为

$$d\boldsymbol{x}_t = \left( f_t(\boldsymbol{x}) - \frac{1}{2}(\boldsymbol{g}_t^2 + \sigma_t^2) \nabla_{\boldsymbol{x}} \text{logp}_t(\boldsymbol{x}) \right) dt + \sigma_t dW_t$$

当我们取极端情况  $\sigma_t = 0$  时，上述方程退化为ODE方程

$$d\boldsymbol{x}_t = \left( f_t(\boldsymbol{x}) - \frac{1}{2}\boldsymbol{g}_t^2 \nabla_{\boldsymbol{x}} \text{logp}_t(\boldsymbol{x}) \right) dt$$

上式其实对应一个Neural ODE，此时方程定义了一个  $\boldsymbol{x}_T$  到  $\boldsymbol{x}_0$  的随机变量的变换，我们可以通过高阶的ODE 数值算法加速这个变换过程。

## 7 Poisson Flow Model

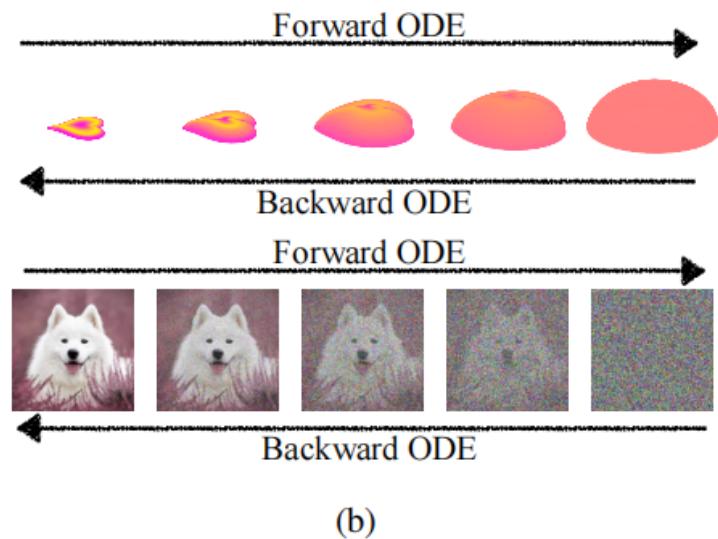
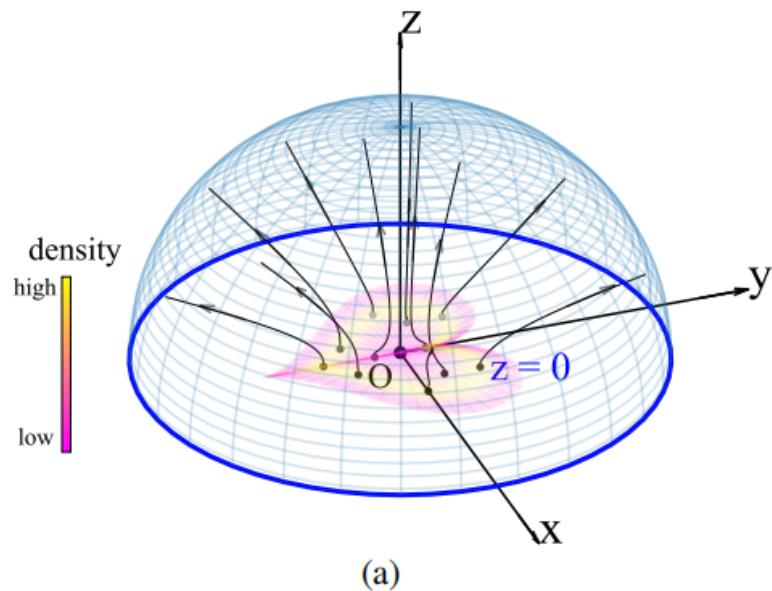


Figure 1: (a) 3D Poisson field trajectories for a heart-shaped distribution (b) The evolutions of a distribution (**top**) or an (augmented) sample (**bottom**) by the forward/backward ODEs pertained to the Poisson field.

## 7.1 万有引力

假设引力源位于  $\mathbf{y}$ , 那么位于 $\mathbf{x}$ 的物体收到的引力为

$$\mathbf{F}(\mathbf{x}) = \frac{-1}{4\pi} \frac{\mathbf{x} - \mathbf{y}}{\|\mathbf{x} - \mathbf{y}\|^3}$$

对于d维空间, 引力场形式为

$$\mathbf{F}(\mathbf{x}) = \frac{-1}{S_d(1)} \frac{\mathbf{x} - \mathbf{y}}{\|\mathbf{x} - \mathbf{y}\|^d}$$

$S_d(1)$ 为d维球面的表面积, 这个式子是d维Poisson方程Green 函数的梯度。

考虑 $\mathbf{x} \in \mathbb{R}^N$ ,  $\rho(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$ , 泊松方程为

$$\nabla^2 \varphi(\mathbf{x}) = -\rho(\mathbf{x})$$

其中 $\varphi : \mathbb{R}^N \rightarrow \mathbb{R}$  为势函数, 记梯度场 $E(\mathbf{x}) = -\nabla \varphi(\mathbf{x})$ , 泊松方程也记作 $\nabla \cdot E = \rho$ 。再无穷远0边界的条件下泊松方程有唯一积分形式解

$$\varphi(\mathbf{x}) = \int G(\mathbf{x}, \mathbf{y}) \rho(\mathbf{y}) d\mathbf{y}, \quad G(\mathbf{x}, \mathbf{y}) = \frac{1}{(N-2)S_{N-1}(1)} \frac{1}{\|\mathbf{x} - \mathbf{y}\|^{N-2}}$$

此时梯度场为

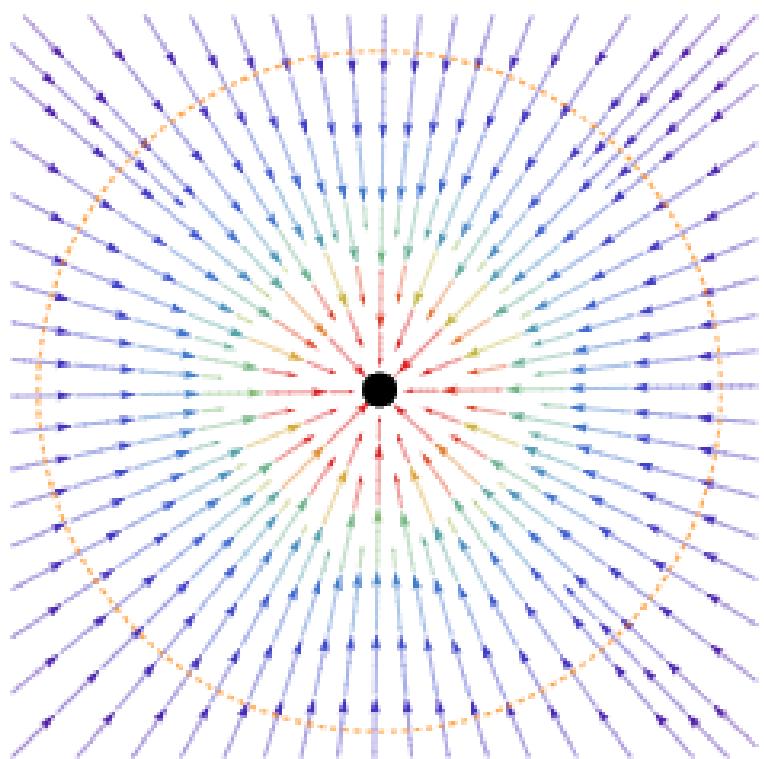
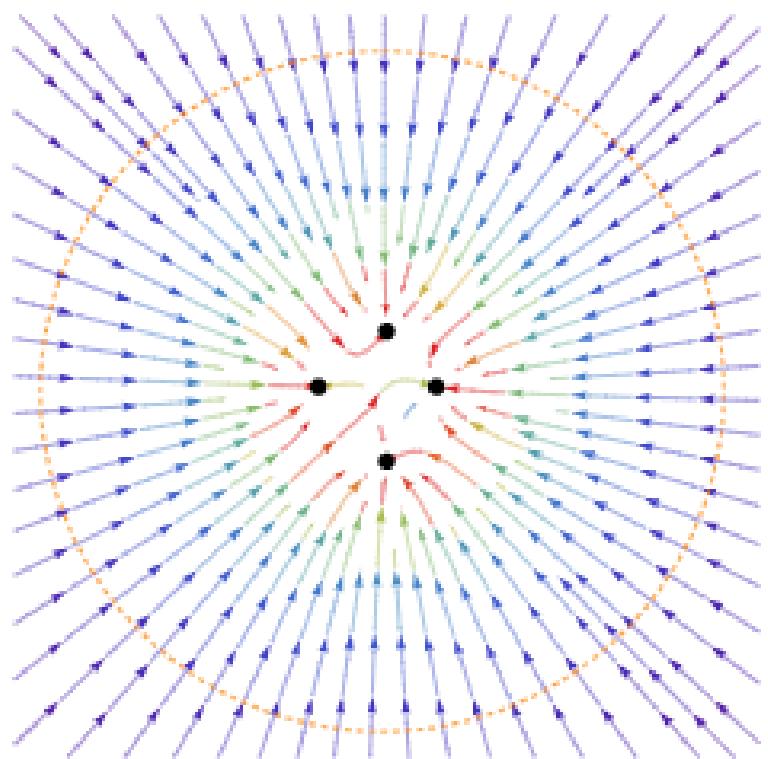
$$E(\mathbf{x}) = -\nabla \varphi(\mathbf{x}) = -\int \nabla_x G(\mathbf{x}, \mathbf{y}) \rho(\mathbf{y}) d\mathbf{y}, \quad \nabla_x G(\mathbf{x}, \mathbf{y}) = -\frac{1}{S_{N-1}(1)} \frac{\mathbf{x} - \mathbf{y}}{\|\mathbf{x} - \mathbf{y}\|^N}$$

Poisson场直接定义了一个flow model: 将概率分布由梯度流演化

$$\begin{aligned} \frac{\partial p_t(\mathbf{x})}{\partial t} &= -\nabla \cdot (p_t(\mathbf{x}) E(\mathbf{x})) \\ \frac{d\mathbf{x}}{dt} &= E(\mathbf{x}) \end{aligned}$$

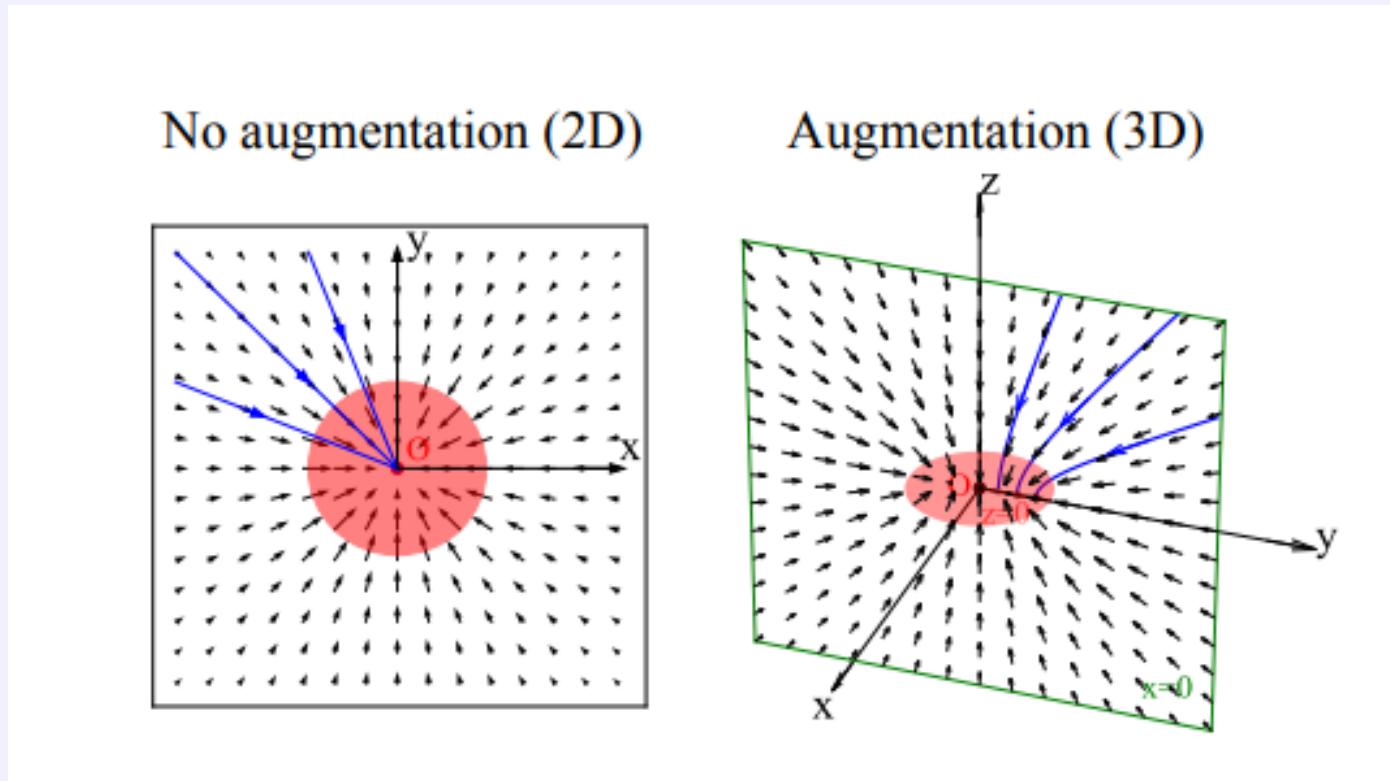
## 7.2 等效质心

无穷远处的多源引力场，等价于位于质心、质量叠加的质点引力场。意味着在足够大半径时，可以认为场线是均匀地穿过以质点为球心的球面的，所以我们在一个半径足够大的球面上进行均匀采样就可以。



## 7.3 模式坍缩

引力场的各向同性使得对应的初始分布易于采样，然而也会造成引力源的相互抵消现象，从而出现“模式坍缩（Mode Collapse）”。



如果数据分布在2D的均匀球面上，那么所有轨道几乎都会来到中心点O

## 7.4 解决方式

为了避免这一点，我们增加了一个时间维度T：

假设真实样本 $\boldsymbol{x} \in \mathbb{R}^d$ , 增强数据点为 $(\boldsymbol{x}, t) \in \mathbb{R}^{d+1}$ , 那么原来的分布 $\boldsymbol{x} \sim \tilde{p}(\boldsymbol{x}) \rightarrow (\boldsymbol{x}, t) \sim \delta(t)\tilde{p}(\boldsymbol{x})$ , 这样就能避免各向同性的现象。此时的引力场为

$$\begin{aligned}\boldsymbol{F}(\boldsymbol{x}, t) &= -\frac{1}{S_{d+1}(1)} \int \frac{(\boldsymbol{x} - \boldsymbol{x}_0, t - t_0)}{(|\boldsymbol{x} - \boldsymbol{x}_0|^2 + (t - t_0)^2)^{(d+1)/2}} \delta(t_0) \tilde{p}(\boldsymbol{x}_0) d\boldsymbol{x}_0 dt_0 \\ &= -\frac{1}{S_{d+1}(1)} \int \frac{(\boldsymbol{x} - \boldsymbol{x}_0, t)}{(|\boldsymbol{x} - \boldsymbol{x}_0|^2 + t^2)^{(d+1)/2}} \tilde{p}(\boldsymbol{x}_0) d\boldsymbol{x}_0 \\ &\triangleq (\boldsymbol{F}_{\boldsymbol{x}}, \boldsymbol{F}_t)\end{aligned}$$

其中 $\boldsymbol{F}_{\boldsymbol{x}}$ 为 $\boldsymbol{F}(\boldsymbol{x}, t)$ 的前d个分量,  $\boldsymbol{F}_t$ 为 $\boldsymbol{F}(\boldsymbol{x}, t)$ 的最后一个分量。

如果已知  $\mathbf{F}(x, t)$ , 那么沿场线运动

$$(dx, dt) = (\mathbf{F}_x, \mathbf{F}_t) d\tau \Rightarrow \frac{dx}{dt} = \frac{\mathbf{F}_x}{\mathbf{F}_t}$$

就是所需要的生成过程的微分方程, 此时真实样本在  $t = 0$  平面上, 这是终止条件。

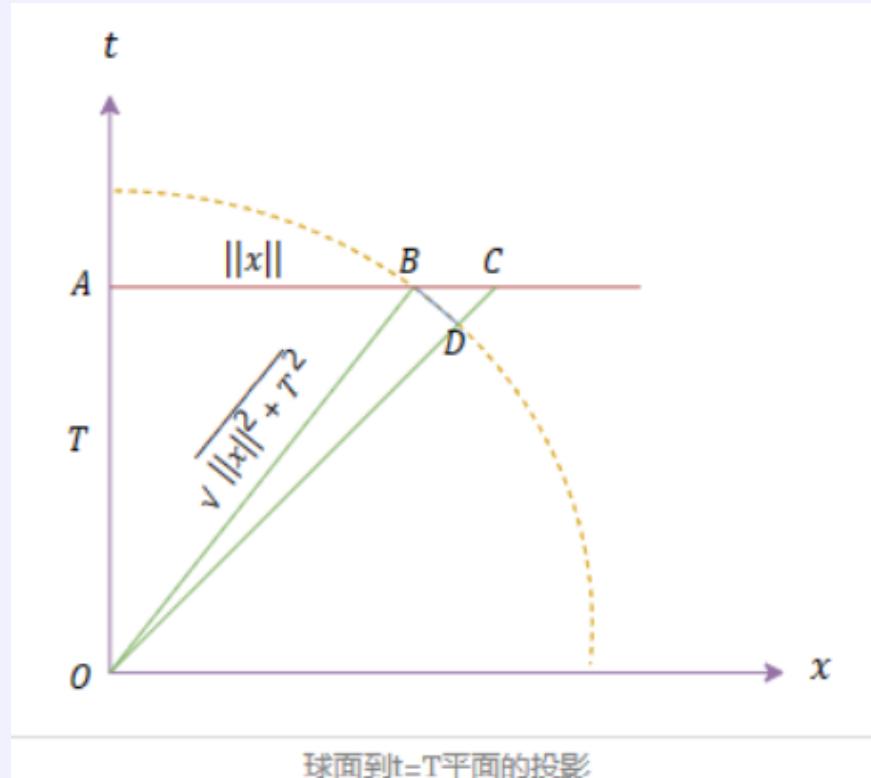
我们在一个足够大的平面  $t = T$  上采样, 这样生成过程就变成从  $t = T$  到  $t = 0$  的ODE 过程了。但在固定平面上采样, 先验分布就不是均匀的:

$$p_{\text{prior}}(\mathbf{x}) \propto \frac{1}{(\|\mathbf{x}\|^2 + T^2)^{(d+1)/2}}$$

证明. 因为场线是均匀穿过  $d + 1$  维球面, 故  $(x, T)$  处密度反比于  $S_{d+1}(x, T)$ , 即  $\propto \frac{1}{(\|x\|^2 + T^2)^{\frac{d}{2}}}$

$$p_{\text{prior}}(x) \propto \frac{1}{(\|x\|^2 + T^2)^{\frac{d}{2}}} \times \frac{T}{(\|x\|^2 + T^2)^{\frac{1}{2}}}$$

□



## 7.5 场的训练

在初始分布为  $\mathbf{x}_0 \sim \tilde{p}(\mathbf{x}_0)$ , 向量场可以写为

$$\mathbb{E}_{\mathbf{x}_0 \sim \tilde{p}(\mathbf{x}_0)} \left[ -\frac{(\mathbf{x} - \mathbf{x}_0, t)}{(\|\mathbf{x} - \mathbf{x}_0\|^2 + t^2)^{\frac{d+1}{2}}} \right]$$

我们引入函数  $s_\theta(\mathbf{x}, t)$  来学习力  $\mathbf{F}(\mathbf{x}, t)$ , 训练目标为

$$\mathbb{E}_{\mathbf{x}_0 \sim \tilde{p}(\mathbf{x}_0)} \left[ \left| s_\theta(\mathbf{x}, t) + \frac{(\mathbf{x} - \mathbf{x}_0, t)}{(|\mathbf{x} - \mathbf{x}_0|^2 + t^2)^{(d+1)/2}} \right|^2 \right]$$

加噪过程为

$$\mathbf{x} = \mathbf{x}_0 + \|\boldsymbol{\varepsilon}_{\mathbf{x}}\|(1 + \tau)^m \mathbf{u}, t = |\varepsilon_t|(1 + \tau)^m$$

其中  $(\boldsymbol{\varepsilon}_x, \varepsilon_t) \sim \mathcal{N}(0, \sigma^2 I_{(d+1) \times (d+1)})$ ,  $m \sim U[0, M]$ ,  $\mathbf{u}$  为  $d$  维单位球面上均匀分布的单位向量,  $r, \sigma, M$  维常数, 这个设计特别主观...

Table 1: CIFAR-10 sample quality (FID, Inception) and number of function evaluation (NFE).

	Invertible?	Inception $\uparrow$	FID $\downarrow$	NFE $\downarrow$
PixelCNN [36]	✗	4.60	65.9	1024
IGEBM [8]	✗	6.02	40.6	60
ViTGAN [24]	✗	9.30	6.66	1
StyleGAN2-ADA [17]	✗	9.83	2.92	1
StyleGAN2-ADA (cond.) [17]	✗	10.14	2.42	1
NCSN [31]	✗	8.87	25.32	1001
NCSNv2 [32]	✗	8.40	10.87	1161
DDPM [16]	✗	9.46	3.17	1000
NCSN++ VE-SDE [33]	✗	9.83	2.38	2000
NCSN++ deep VE-SDE [33]	✗	9.89	2.20	2000
Glow [19]	✓	3.92	48.9	1
DDIM, T=50 [30]	✓	-	4.67	50
DDIM, T=100 [30]	✓	-	4.16	100
NCSN++ VE-ODE [33]	✓	9.34	5.29	194
NCSN++ deep VE-ODE [33]	✓	9.17	7.66	194
<b><i>DDPM++ backbone</i></b>				
VP-SDE [33]	✗	9.58	2.55	1000
sub-VP-SDE [33]	✗	9.56	2.61	1000
<hr/>				
VP-ODE [33]	✓	9.46	2.97	134
sub-VP-ODE [33]	✓	9.30	3.16	146
PFGM (ours)	✓	<b>9.65</b>	<b>2.48</b>	<b>104</b>
<b><i>DDPM++ deep backbone</i></b>				
VP-SDE [33]	✗	9.68	2.41	1000
sub-VP-SDE [33]	✗	9.57	2.41	1000
<hr/>				
VP-ODE [33]	✓	9.47	2.86	134
sub-VP-ODE [33]	✓	9.40	3.05	146
PFGM (ours)	✓	<b>9.68</b>	<b>2.35</b>	<b>110</b>

---

**Algorithm 1** Learning the normalized Poisson Field

---

**Input:** Training iteration  $T$ , Initial model  $f_\theta$ , dataset  $\mathcal{D}$ , constant  $\gamma$ , learning rate  $\eta$ .

**for**  $t = 1 \dots T$  **do**

- Sample a large batch  $\mathcal{B}_L$  from  $\mathcal{D}$  and subsample a batch of datapoints  $\mathcal{B} = \{\mathbf{x}_i\}_{i=1}^{|\mathcal{B}|}$  from  $\mathcal{B}_L$
- Simulate the ODE:  $\{\tilde{\mathbf{y}}_i = \text{perturb}(\mathbf{x}_i)\}_{i=1}^{|\mathcal{B}|}$
- Calculate the normalized field by  $\mathcal{B}_L$ :  $\mathbf{v}_{\mathcal{B}_L}(\tilde{\mathbf{y}}_i) = -\sqrt{N}\hat{\mathbf{E}}_{\mathcal{B}_L}(\tilde{\mathbf{y}}_i)/(\|\hat{\mathbf{E}}_{\mathcal{B}_L}(\tilde{\mathbf{y}}_i)\|_2 + \gamma)$ ,  $\forall i$
- Calculate the loss:  $\mathcal{L}(\theta) = \frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} \|f_\theta(\tilde{\mathbf{y}}_i) - \mathbf{v}_{\mathcal{B}_L}(\tilde{\mathbf{y}}_i)\|_2^2$
- Update the model parameter:  $\theta = \theta - \eta \nabla \mathcal{L}(\theta)$

**end for**

**return**  $f_\theta$

---

---

**Algorithm 2**  $\text{perturb}(\mathbf{x})$ 

---

Sample the power  $m \sim \mathcal{U}[0, M]$

Sample the initial noise  $(\epsilon_x, \epsilon_z) \sim \mathcal{N}(0, \sigma^2 I_{(N+1) \times (N+1)})$

Uniformly sample the vector from the unit ball  $\mathbf{u} \sim \mathcal{U}(S_N(1))$

Construct training point  $\mathbf{y} = \mathbf{x} + \|\epsilon_x\| (1 + \tau)^m \mathbf{u}$ ,  $z = |\epsilon_z| (1 + \tau)^m$

**return**  $\tilde{\mathbf{y}} = (\mathbf{y}, z)$

---

# 8 构建ODE扩散模型的一般方法

## 8.1 直观结果

ODE式扩散模型那个就是需要设计一个source  $f_t(\mathbf{x}_t)$ ,使得它对应的演化轨迹构成给定分布 $p_T(\mathbf{x}_T) \leftrightarrow p_0(\mathbf{x}_0)$ 的一个变换。

如果给定 $\mathbf{x}_0 \sim p_0(\mathbf{x}_0), \mathbf{x}_T \sim p_T(\mathbf{x}_T)$ ,假设轨迹为

$$\mathbf{x}_t = \varphi_t(\mathbf{x}_0, \mathbf{x}_T)$$

只需要求

$$\mathbf{x}_0 = \varphi_0(\mathbf{x}_0, \mathbf{x}_T), \mathbf{x}_T = \varphi_T(\mathbf{x}_0, \mathbf{x}_T)$$

此时的微分方程为

$$\frac{d\mathbf{x}_t}{dt} = \frac{\partial \varphi_t(\mathbf{x}_0, \mathbf{x}_T)}{\partial t}$$

但这个是不实用的，右端的形式需要是一个  $\frac{d\mathbf{x}_t}{dt} = \mathbf{f}_t(\mathbf{x}_t)$ ，不能显含  $\mathbf{x}_0$ 。于是目标改为学习一个  $s_\theta(\mathbf{x}_t, t)$

$$\min_{\theta} \mathbb{E}_{\mathbf{x}_0 \sim p_0(\mathbf{x}_0), \mathbf{x}_T \sim p_T(\mathbf{x}_T)} \left[ \left\| s_\theta(\mathbf{x}_t, t) - \frac{\partial \varphi_t(\mathbf{x}_0, \mathbf{x}_T)}{\partial t} \right\|^2 \right]$$

此时  $s_\theta(\mathbf{x}_t, t)$  逼近了  $\frac{\partial \varphi_t(\mathbf{x}_0, \mathbf{x}_T)}{\partial t}$ ，扩散ODE为

$$\frac{d\mathbf{x}_t}{dt} = s_\theta(\mathbf{x}_t, t)$$

## 8.2 简单例子

设  $T = 1$ , 设变化轨迹是直线

$$\mathbf{x}_t = \varphi_t(\mathbf{x}_0, \mathbf{x}_1) = (\mathbf{x}_1 - \mathbf{x}_0)t + \mathbf{x}_0$$

那么右端项

$$\frac{\partial \varphi_t(\mathbf{x}_0, \mathbf{x}_T)}{\partial t} = \mathbf{x}_1 - \mathbf{x}_0$$

训练目标为

$$\begin{aligned} \mathbb{E}_{\mathbf{x}_0 \sim p_0(\mathbf{x}_0), \mathbf{x}_T \sim p_T(\mathbf{x}_T)} [\|\mathbf{s}_{\theta}((\mathbf{x}_1 - \mathbf{x}_0)t + \mathbf{x}_0, t) - (\mathbf{x}_1 - \mathbf{x}_0)\|^2] &= \\ \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_t \sim p_0(\mathbf{x}_0)p_t(\mathbf{x}_t | \mathbf{x}_0)} \left[ \left\| \mathbf{s}_{\theta}(\mathbf{x}_t, t) - \frac{\mathbf{x}_t - \mathbf{x}_0}{t} \right\|^2 \right] \end{aligned}$$

这个模型被称为“Rectified Flow”.

## 8.3 验证原理

我们需要从理论上证明优化目标

$$\min_{\theta} \mathbb{E}_{x_0 \sim p_0(x_0), x_T \sim p_T(x_T)} \left[ \left\| s_{\theta}(x_t, t) - \frac{\partial \varphi_t(x_0, x_T)}{\partial t} \right\|^2 \right]$$

所得到的方程确实实现了分布  $p_T(x_T)$  到  $p_0(x_0)$  的转换，这里有两种思路：

第一种是证明优化目标的最优解满足连续性方程

$$\frac{\partial p_t(x_t)}{\partial t} = -\nabla_{x_t} \cdot (p_t(x_t) s_{\theta}(x_t, t))$$

那么根据连续性方程与ODE的对应关系得证。

第二种是证明它的微分版本

$$\mathbb{E}_{\mathbf{x}_{t+\Delta t}}[\phi(\mathbf{x}_{t+\Delta t})] = \mathbb{E}_{\mathbf{x}_t} [\phi(\mathbf{x}_t + \mathbf{f}_t(\mathbf{x}_t) \Delta t)]$$

对所有测试函数  $\phi(x)$  成立。即我们要证明最优解满足如上方程，也就是需按照一个  $\mathbf{f}_t(\mathbf{x}_t)$  满足上式，且正好是优化目标的最优解。

于是记  $\boldsymbol{\varphi}_t(\mathbf{x}_0, \mathbf{x}_T)$

$$\begin{aligned}\mathbb{E}_{\mathbf{x}_{t+\Delta t}}[\phi(\mathbf{x}_{t+\Delta t})] &= \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_T}[\phi(\boldsymbol{\varphi}_{t+\Delta t})] \\ &= \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_T} \left[ \phi(\boldsymbol{\varphi}_t) + \Delta t \frac{\partial \boldsymbol{\varphi}_t}{\partial t} \cdot \nabla_{\boldsymbol{\varphi}_t} \phi(\boldsymbol{\varphi}_t) \right] \\ &= \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_T}[\phi(\mathbf{x}_t)] + \Delta t \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_T} \left[ \frac{\partial \boldsymbol{\varphi}_t}{\partial t} \cdot \nabla_{\mathbf{x}_t} \phi(\mathbf{x}_t) \right] \\ &= \mathbb{E}_{\mathbf{x}_t}[\phi(\mathbf{x}_t)] + \Delta t \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_T} \left[ \frac{\partial \boldsymbol{\varphi}_t}{\partial t} \cdot \nabla_{\mathbf{x}_t} \phi(\mathbf{x}_t) \right]\end{aligned}$$

此时  $\frac{\partial \varphi_t}{\partial t}$  是  $\mathbf{x}_0, \mathbf{x}_T$  的函数，我们假设(8.1)关于  $\mathbf{x}_T$  可逆，则可以反解出  $\mathbf{x}_T = \psi_t(\mathbf{x}_0, \mathbf{x}_t)$ ，代入  $\frac{\partial \varphi_t}{\partial t}$  使得它变为  $\mathbf{x}_0, \mathbf{x}_t$  的函数，

$$\begin{aligned}
 \mathbb{E}_{\mathbf{x}_{t+\Delta t}}[\phi(\mathbf{x}_{t+\Delta t})] &= \mathbb{E}_{\mathbf{x}_t}[\phi(\mathbf{x}_t)] + \Delta t \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_T} \left[ \frac{\partial \varphi_t}{\partial t} \cdot \nabla_{\mathbf{x}_t} \phi(\mathbf{x}_t) \right] \\
 &= \mathbb{E}_{\mathbf{x}_t}[\phi(\mathbf{x}_t)] + \Delta t \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_t} \left[ \frac{\partial \varphi_t}{\partial t} \cdot \nabla_{\mathbf{x}_t} \phi(\mathbf{x}_t) \right] \\
 &= \mathbb{E}_{\mathbf{x}_t}[\phi(\mathbf{x}_t)] + \Delta t \mathbb{E}_{\mathbf{x}_t} \left[ \underbrace{\mathbb{E}_{\mathbf{x}_0 | \mathbf{x}_t} \left[ \frac{\partial \varphi_t}{\partial t} \right]}_{\mathbf{x}_t \text{ 的函数}} \cdot \nabla_{\mathbf{x}_t} \phi(\mathbf{x}_t) \right] \\
 &= \mathbb{E}_{\mathbf{x}_t} \left[ \phi \left( \mathbf{x}_t + \Delta t \mathbb{E}_{\mathbf{x}_0 | \mathbf{x}_t} \left[ \frac{\partial \varphi_t}{\partial t} \right] \right) \right]
 \end{aligned}$$

现在有

$$\mathbb{E}_{\mathbf{x}_{t+\Delta t}}[\phi(\mathbf{x}_{t+\Delta t})] = \mathbb{E}_{\mathbf{x}_t} \left[ \phi \left( \mathbf{x}_t + \Delta t \mathbb{E}_{\mathbf{x}_0 | \mathbf{x}_t} \left[ \frac{\partial \varphi_t}{\partial t} \right] \right) \right]$$

对于任意测试函数  $\phi$  成立，这标明

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \Delta t \mathbb{E}_{\mathbf{x}_0 | \mathbf{x}_t} \left[ \frac{\partial \varphi_t}{\partial t} \right] \Rightarrow \frac{d \mathbf{x}_t}{d t} = \mathbb{E}_{\mathbf{x}_0 | \mathbf{x}_t} \left[ \frac{\partial \varphi_t}{\partial t} \right]$$

就是我们寻找的ODE，

$$\mathbb{E}_{\mathbf{x}}[\mathbf{x}] = \arg \min_{\boldsymbol{\mu}} \mathbb{E}_{\mathbf{x}}[\|\mathbf{x} - \boldsymbol{\mu}\|^2]$$

该式的右端正好是优化问题训练目标的最优解，证明了优化目标得到的方程实现了分布  $p_{T(\mathbf{x}_T)}$  到  $p_0(\mathbf{x}_0)$  的转换。