Introduction

Flow Matching 是一种用于机器学习生成模型训练的技术，主要出现在扩散模型（Diffusion Models）和流模型（Flow-based Models）的背景中。通俗易懂地理解它，可以把它比喻为一个“路径校正器”，通过优化模型生成的路径，让其逐步接近目标分布。

先不做技术细节讲解，说说Flow Matching是什么

类比解释：

假如你学骑自行车，目标是从家里骑到学校（终点）。刚开始你会晃来晃去（随机噪声），但是你不断调整方向和速度（流动的校正），最终找到一条平稳的路径抵达学校。Flow Matching 就是在训练“骑车路径校正器”，让你的骑行过程更加流畅，尽量避免偏离目标路线。

为什么我们需要 Flow Matching?

- 在 **扩散模型 (Diffusion Models)** 中，我们是通过逐步加噪声，再逐步去噪声的方式，从数据分布走向噪声，再反向回来。
- 而在 **流模型 (Flow-based Models)** 中，我们通常需要设计一系列可逆变换，把噪声变成数据。

Flow Matching 的贡献在于，它提供了一种 **更直接的方式**：

- 不用完全依赖复杂的噪声注入和去噪过程；

- 也不用设计繁琐的可逆变换；
- 而是通过学习一个合适的速度场，让整个生成路径变得光滑、确定，并且更容易优化。

Change of Variable for Probability Density Function

设随机变量 z 及其概率密度函数 $z \sim \pi(z)$，通过一个一一对应的映射函数 f 构造一个新的随机变量 $x = f(z)$。如果存在逆函数 f^{-1}，那么新变量 x 的概率密度函数 $p(x)$ 计算如下：

(1) 当 z 为随机变量：

$$p(x) = \pi(z) \left| \frac{dz}{dx} \right| = \pi\left(f^{-1}(x)\right) \left| \frac{df^{-1}}{dx} \right| = \pi\left(f^{-1}(x)\right) \left| \left(f^{-1}\right)'(x) \right|$$

(2) 当 z 为随机向量：

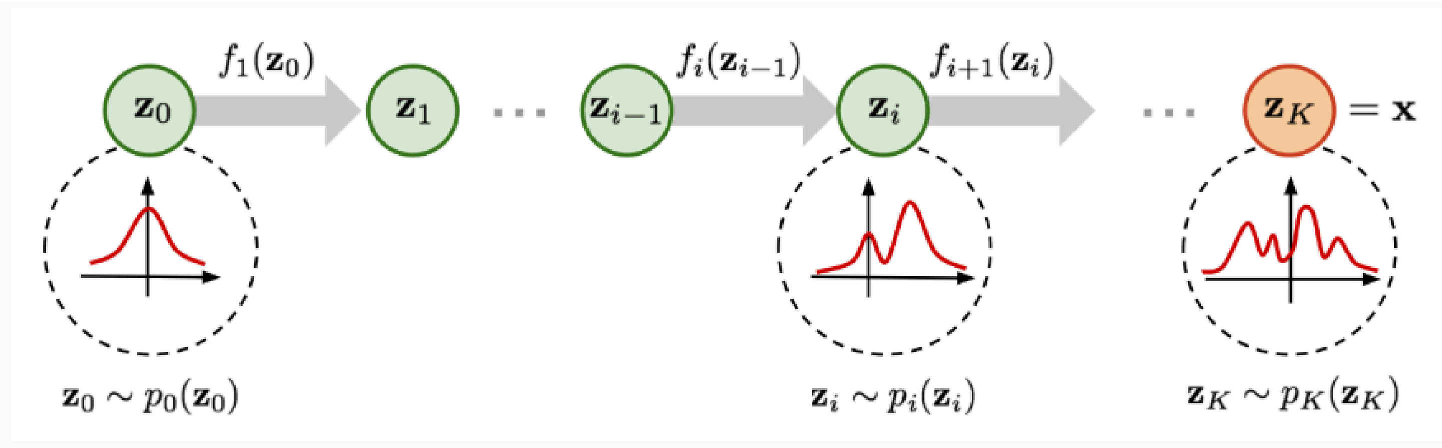
$$p(\mathbf{x}) = \pi(\mathbf{z}) \left| \det \frac{d\mathbf{z}}{d\mathbf{x}} \right| = \pi\left(f^{-1}(\mathbf{x})\right) \left| \det \frac{df^{-1}}{d\mathbf{x}} \right|$$

其中，$\det$ 是行列式，$\dfrac{df^{-1}}{dx}$ 是雅可比矩阵。

特例： 如果 $x \sim \mathcal{N}(\mu, \sigma^2)$，当 a, b 为实数时，则有

$$z = f(x) = ax + b \sim \mathcal{N}\left(a\mu + b,\ (a\sigma)^2\right)$$

Normalization flow



$$\mathbf{x} = \mathbf{z}_K = f_K \circ f_{K-1} \circ \cdots \circ f_1(\mathbf{z}_0)$$

对于其中第 i 步，有：

$$\mathbf{z}_{i-1} \sim p_{i-1}(\mathbf{z}_{i-1})$$

$$\mathbf{z}_i = f_i(\mathbf{z}_{i-1}), \quad \text{thus } \mathbf{z}_{i-1} = f_i^{-1}(\mathbf{z}_i)$$

根据概率密度函数的变量变换关系可得：

$$p_i(\mathbf{z}_i) = p_{i-1}\left(f_i^{-1}(\mathbf{z}_i)\right) \left|\det \frac{df_i^{-1}}{d\mathbf{z}_i}\right|$$

$$= p_{i-1}(\mathbf{z}_{i-1}) \left|\det \left(\frac{df_i}{d\mathbf{z}_{i-1}}\right)^{-1}\right|$$

$$= p_{i-1}(\mathbf{z}_{i-1}) \left|\det \frac{df_i}{d\mathbf{z}_{i-1}}\right|^{-1}$$

由

$$\log p_i(\mathbf{z}_i) = \log p_{i-1}(\mathbf{z}_{i-1}) - \log \left|\det \frac{df_i}{d\mathbf{z}_{i-1}}\right|$$

给定这样一连串的概率密度函数和变换关系，可以逐步展开直至追溯到初始分布，可得：

$$
\begin{aligned}
\log p(\mathbf{x}) &= \log \pi_K(\mathbf{z}_K) \\
&= \log \pi_{K-1}(\mathbf{z}_{K-1}) - \log \left| \det \frac{d f_K}{d\mathbf{z}_{K-1}} \right| \\
&= \log \pi_{K-2}(\mathbf{z}_{K-2}) - \log \left| \det \frac{d f_{K-1}}{d\mathbf{z}_{K-2}} \right| - \log \left| \det \frac{d f_K}{d\mathbf{z}_{K-1}} \right| \\
&\vdots \\
&= \log \pi_0(\mathbf{z}_0) - \sum_{i=1}^{K} \log \left| \det \frac{d f_i}{d\mathbf{z}_{i-1}} \right|
\end{aligned}
$$

当这一系列变换函数 f_i 可逆，且雅可比矩阵易于计算时，模型训练时的优化目标为负对数似然：

$$\mathcal{L}(\mathcal{D}) = -\frac{1}{|\mathcal{D}|} \sum_{\mathbf{x} \in \mathcal{D}} \log p(\mathbf{x})$$

Continuous Normalization Flow

Continuous Normalizing Flows (CNFs) 是 Normalizing Flows 的一种扩展，它可以更好地建模复杂的概率分布。在传统的Normalizing Flows中，变换通常是通过一系列可逆的离散函数来定义的，而在CNFs中，这种变换是连续的。

$$\frac{d}{dt}\,\phi_t(x) = v_t\big(\phi_t(x)\big)$$

$$\phi_0(x) = x$$

$$z_{t+\Delta t} = z_t + \Delta t \cdot v(z_t, t)$$

Flow($\phi_t | \varphi_t$), velocity field($v_t | u_t$), probability density function(p_t):

$$\phi_t(x) = x + \int_0^t v_s\big(\phi_s(x)\big)\, ds$$

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \mathbf{v}) = 0$$

- ρ 是流体的密度。
- $\mathbf{v}$ 是流体的速度矢量。
- $\frac{\partial \rho}{\partial t}$ 是密度随时间的变化率。
- $\operatorname{div}(\rho \mathbf{v})$ 是质量通量密度的散度，表示单位时间内通过单位面积的净质量流量。

$$\frac{\partial}{\partial t} p_t(x) + div\big(p_t(x) v_t(x)\big) = 0$$

其中，$p_t(x)$ 是 t 时刻的概率密度函数，$v_t(x)$ 是与 $p_t(x)$ 相关的向量场，它描述了概率密度随位置和时间的变化，
$\operatorname{div}\big(p_t(x)v_t(x)\big)$ 是向量场与概率密度的乘积的**散度**，表示概率流通过某个区域的净变化率。
因此 CE 方程保证：
概率质量不会凭空产生或消失，而是通过流动在不同位置之间转移。
这正是 Flow Matching 的理论基础：

- **flow** 定义了样本如何移动；
- **velocity** 描述了每个点的运动方向；
- **PDF (概率密度函数)** 则通过 CE 方程来保证整个分布的一致演化。

Flow Matching

给定一个目标概率密度路径 $p_t(x)$ 及其对应的向量场 $u_t(x)$，这里的概率密度路径 $p_t(x)$ 是由这个向量场 $u_t(x)$ 生成的， $v_t(x)$ 是待学习的向量场，那么Flow Matching的优化目标可以定义为：

$$\mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}_{t,p_t(x)} \left\|v_t(x) - u_t(x)\right\|^2$$

缺乏合适的先验知识来确定u_t, 无法直接使用.

1. $u_t(x)$ **是分布的函数**
   - 它依赖于整个概率路径 p_t，而不是单个采样点。
   - 即便我们可以从噪声分布 p_0 和数据分布 p_1 各自采样，也没法直接得到中间时间 t 的分布 p_t。
   - 没有 p_t，我们无法直接计算 $u_t(x)$。
2. **条件速度场不可观测**
   - 从定义上，$u_t(x)$ 需要满足连续性方程：

$$\partial_t p_t + \nabla \cdot (p_t u_t) = 0.$$

   - 这相当于在整个空间和时间上知道分布的演化规律。
   - 但我们只能采样两端点分布，没法直接观测 p_t 的动态演化。

所以我们希望通过$u_t(x|x_1)$来实现
这与diffusion model也有相似性，即反向过程中，$q(x_{t-1} \,\big|\, x_t, x_0) = q(x_t \,\big|\, x_{t-1}, x_0)\frac{(q(x_(t-1)|x_0))}{(q(x_t|x_0))}$

为什么我们能够用预测条件向量场来替代无条件向量场

Theorem 1. *Given vector fields $u_t(x|x_1)$ that generate conditional probability paths $p_t(x|x_1)$, for any distribution $q(x_1)$, the marginal vector field u_t in equation 8 generates the marginal probability path p_t in equation 6, i.e., u_t and p_t satisfy the continuity equation (equation 26).*

从定理中我们可以得到，$p_t(x|x_1)$可以由$u_t(x|x_1)$得到(满足条件概率的CE方程)。

证明目标：$u_t(x|x_1) \to p_t(x|x_1)$ 能否得到$u_t(x) \to p_t(x)$

Proof

已知条件：$\frac{\partial}{\partial t} p_t(x|x_1) + div\big(p_t(x|x_1)u_t(x|x_1)\big) = 0$

$p_t(x) = \int p_t(x|x_1)q(x_1)dx_1$

$$\begin{aligned} \frac{d}{dt} p_t(x) &= \frac{d}{dt} \int \left(p_t(x|x_1)\right) q(x_1)\, dx_1 \\ &= \int \left(\frac{d}{dt} p_t(x|x_1) \right) q(x_1)\, dx_1 \\ &= -\int \mathrm{div}(u_t(x|x_1)p_t(x|x_1))\, q(x_1)\, dx_1 \\ &= -\mathrm{div}\left(\int u_t(x|x_1)p_t(x|x_1)q(x_1)\, dx_1 \right) \\ &= -\mathrm{div}\big(u_t(x)p_t(x)\big). \end{aligned}$$

而$u_t(x) = \int u_t(x|x_1)\, \frac{p_t(x|x_1)}{p_t(x)}\, q(x_1)\, dx_1$就是原文中假设的条件.

hint: $q(x_1)$是目标分布的probability density function

有了条件向量场，我们要怎么设计条件向量场的优化目标呢

Theorem 2. *Assuming that $p_t(x) > 0$ for all $x \in \mathbb{R}^d$ and $t \in [0,1]$, then, up to a constant independent of θ, $\mathcal{L}_{CFM}$ and $\mathcal{L}_{FM}$ are equal. Hence, $\nabla_\theta \mathcal{L}_{FM}(\theta) = \nabla_\theta \mathcal{L}_{CFM}(\theta)$.*

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{t,q(x_1),p_t(x|x_1)} \left\| v_t(x) - u_t(x|x_1) \right\|^2$$

只需要证明这个loss和原先的等价就行

$$\|v_t(x) - u_t(x)\|^2 = \|v_t(x)\|^2 - 2\langle v_t(x), u_t(x)\rangle + \|u_t(x)\|^2$$

$$\|v_t(x) - u_t(x|x_1)\|^2 = \|v_t(x)\|^2 - 2\langle v_t(x), u_t(x|x_1)\rangle + \|u_t(x|x_1)\|^2$$

我们需要证明 $\mathbb{E}_{p_t(x)} \|v_t(x)\|^2 = \mathbb{E}_{q(x_1),p_t(x|x_1)} \|v_t(x)\|^2$

$$\mathbb{E}_{p_t(x)} \|v_t(x)\|^2 = \int \|v_t(x)\|^2 p_t(x)\, dx = \iint \|v_t(x)\|^2 p_t(x|x_1)q(x_1)\, dx_1 dx = \mathbb{E}_{q(x_1),p_t(x|x_1)} \|v_t(x)\|^2$$

$$
\begin{aligned}
\mathbb{E}_{p_t(x)}\langle v_t(x), u_t(x)\rangle &= \int \left\langle v_t(x), \frac{\int u_t(x|x_1)p_t(x|x_1)q(x_1)\,dx_1}{p_t(x)} \right\rangle p_t(x)\,dx \\
&= \int \left\langle v_t(x), \int u_t(x|x_1)p_t(x|x_1)q(x_1)\,dx_1 \right\rangle dx \\
&= \iint \langle v_t(x), u_t(x|x_1)\rangle p_t(x|x_1)q(x_1)\,dx_1 dx \\
&= \mathbb{E}_{q(x_1),p_t(x|x_1)}\langle v_t(x), u_t(x|x_1)\rangle\,.
\end{aligned}
$$

hint: $E[x] = \int xf(x)dx$，如果这个看着不顺眼，也可以用$p(x, x_1) = p(x|x_1)q(x_1)$做代换，就变成了单个前置条件的期望公式

那么为什么不需要去证明$\|u_t(x)\|^2 = \|u_t(x|x_1)\|^2$呢?

因为这个是ground truth常数项，与定理2中期望最后通过两个loss得到的模型梯度无关

条件向量具体形式是什么

> **Theorem 3.** *Let $p_t(x|x_1)$ be a Gaussian probability path as in equation 10, and ψ_t its corresponding flow map as in equation 11. Then, the unique vector field that defines ψ_t has the form:*
>
> $$u_t(x|x_1) = \frac{\sigma_t'(x_1)}{\sigma_t(x_1)}(x - \mu_t(x_1)) + \mu_t'(x_1). \tag{15}$$
>
> *Consequently, $u_t(x|x_1)$ generates the Gaussian path $p_t(x|x_1)$.*

首先，我们希望$p_t(x|x_1)$满足以下分布

$$p_t(x \mid x_1) = \mathcal{N}\big(x \,\big|\, \mu_t(x_1),\, \sigma_t(x_1)^2 I\big),$$

其次，

$$\psi_t(x) = \sigma_t(x_1)\,x + \mu_t(x_1).$$

$$
\begin{gathered}
\frac{d\psi_t(x)}{dt} = u_t(\psi_t(x)|x_1) = \sigma_t'(x_1)\,x + \mu_t'(x_1), \\
\psi_t(x) = y, x = \psi^{-1}(y) \\
x \sim N(0,1), x = \psi^{-1}(y) = \frac{y - \mu_t(x_1)}{\sigma_t(x_1)} \\
u_t(y|x_1) = \frac{\sigma_t'(x_1)(y - \mu_t(x_1))}{\sigma_t(x_1)} + \mu_t'(x_1)
\end{gathered}
$$

另一种proof，$x \sim N(0,1)$等价于$x = x_0$,

$$x_t = \mu_t(x_1) + \sigma_t(x_1)\,x_0,$$

$$\frac{d}{dt}x_t = \mu_t'(x_1) + \sigma_t'(x_1)\, x_0.$$
$$x_0 = \frac{x_t - \mu_t(x_1)}{\sigma_t(x_1)},$$
$$\frac{d}{dt}x_t = \mu_t'(x_1) + \frac{\sigma_t'(x_1)}{\sigma_t(x_1)}\big(x_t - \mu_t(x_1)\big).$$
$$u_t(x|x_1) = \mu_t'(x_1) \ + \ \frac{\sigma_t'(x_1)}{\sigma_t(x_1)}\big(x - \mu_t(x_1)\big).$$

线性插值公式:

$$x(t) = (1-t)\cdot x_0 + t\cdot x_1$$

与diffusion的关系

$$p_t(x \mid x_1) = \mathcal{N}\big(x \,\big|\, \alpha_{1-t}x_1,\ (1-\alpha_{1-t}^2)I\big)$$

$$\psi_t(x) = x_t = \sqrt{\alpha_{1-t}}x_1 + \sqrt{1-\alpha_{1-t}^2}x_0$$

为什么是1-t呢，因为flow matching和DDPM是反过来的， x_0代表噪声，x_1代表真实图像

$x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}z_{t-1}$

与最优传输的关系

$$\mu_t(x) = tx_1, \quad \text{and} \quad \sigma_t(x) = 1-(1-\sigma_{\min})t.$$

当$\sigma_{min} \to 0$时，$\psi_t(x) = x_t = (1-t)x + tx_1$
即我们找到了一条从x_0到x_1之间的最优传输路径，所有的中间值都可以由线性插值表示

Result

	CIFAR-10			ImageNet 32×32			ImageNet 64×64		
Model	NLL↓	FID↓	NFE↓	NLL↓	FID↓	NFE↓	NLL↓	FID↓	NFE↓
Ablations									
DDPM	3.12	7.48	274	3.54	6.99	262	3.32	17.36	264
Score Matching	3.16	19.94	242	3.56	5.68	178	3.40	19.74	441
ScoreFlow	3.09	20.78	428	3.55	14.14	195	3.36	24.95	601
Ours									
FM ʷ/ Diffusion	3.10	8.06	183	3.54	6.37	193	3.33	16.88	187
FM ʷ/ OT	**2.99**	**6.35**	**142**	**3.53**	**5.02**	**122**	**3.31**	**14.45**	**138**

	ImageNet 128×128	
Model	NLL↓	FID↓
MGAN (Hoang et al., 2018)	–	58.9
PacGAN2 (Lin et al., 2018)	–	57.5
Logo-GAN-AE (Sage et al., 2018)	–	50.9
Self-cond. GAN (Lučić et al., 2019)	–	41.7
Uncond. BigGAN (Lučić et al., 2019)	–	25.3
PGMGAN (Armandpour et al., 2021)	–	21.7
FM ʷ/ OT	**2.90**	**20.9**

Table 1: Likelihood (BPD), quality of generated samples (FID), and evaluation time (NFE) for the same model trained with different methods.

NLL↓（负对数似然，BPD）：越低越好，表示模型对数据的拟合程度。
FID↓（Frechet Inception Distance）：越低越好，衡量生成样本与真实样本的相似度。
NFE↓（评估时间）：越低越好，表示生成或评估所需计算成本。