



学习笔记

作者：不梦秋深

组织：电子科技大学

时间：May 24, 2025

邮箱：zhh.aca@foxmail.com

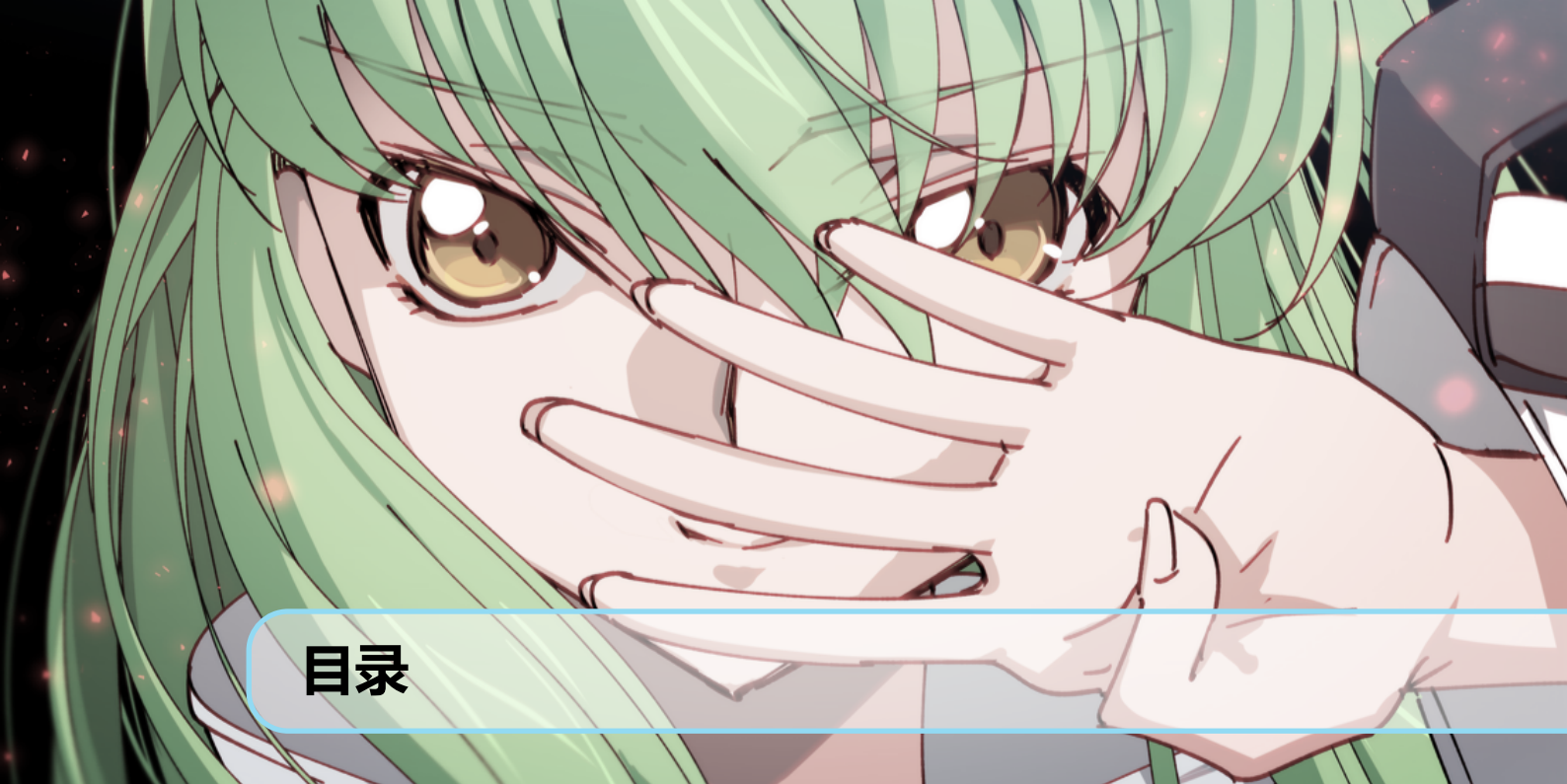


如果觉得写的不错可以打个赏呗



图 1: 赞赏码

分享链接: <https://pan.baidu.com/s/1DH18hVEFkuB3ZcdUX9Gcpg?pwd=maf>
小红书号:9599156808



目录

第 1 章 MIT 6.S184 学习笔记一: Flow Matching	1
1.1 Flow Model.	1
1.2 构造训练目标.	3
1.3 训练	5
第 2 章 Mean Flows 阅读笔记	8
2.1 Flow Matching.	8
2.2 MeanFlow Models	9
2.3 一些其他细节.	9
第 3 章 MIT 6.S184 学习笔记二: SDE	11
3.1 伊藤过程 (Itô Diffusion Process)	11
3.2 扩散模型 (Diffusion Model)	12
3.3 构造训练目标: Conditional and Marginal Score Functions	14
3.4 定理 3.2 解释	14
3.5 训练: 匹配得分 (Score Matching).	16
3.6 时间反演 vs 求解 FP 方程	19
3.7 小结	20
第 4 章 冷冻电镜学习笔记	21
4.1 TEM(Transmission Electron Microscope) 简介	21
4.2 问题解答	23
第 5 章 傅里叶变换解析	29
5.1 向量空间	29
5.2 连续 Parseval 框架 (Continuous Parseval Frame).	37
5.3 装备希尔伯特空间 (rigged Hilbert space)	37
5.4 谱定理 (Spectral Theorem)	39
5.5 总结	42
第 6 章 高斯泼溅算法简介	44
6.1 高斯泼溅简介.	44
6.2 高斯椭球: 定义与性质.	44
6.3 高斯椭球的渲染技术.	46



6.4 高斯泼溅示例.....	48
-----------------	----



第 1 章 MIT 6.S184 学习笔记一: Flow Matching

May 7, 2025

最近几年, 我们见证了生成式 AI 的快速崛起, 诸如 Stable Diffusion 等图像生成式模型可以从噪声生成出及其丰富多彩的图片. 从形似实拍的照片, 到动漫风的美术, 图片风格一应俱全. 这场 AI 革命的核心是: 生成事物的能力. 与过往用于预测的 AI 不同, 生成式 AI 具有更多的创造性: 生成式 AI 允许用户自定义输入, 并根据这些输入创造新的事物.

MIT Class 6.S184 课程 (<https://diffusion.csail.mit.edu/>) 主要教学目的是让同学们学会两个著名的生成式 AI 算法: flow matching 和 denoising diffusion models. 这两个模型是当前最好的生成式模型的骨架 (应用如 Stable Diffusion 3 和 Movie Gen Video), 而且在蛋白质结构预测上也有极好的应用 (如 AlphaFold3 是 diffusion model). 毫无疑问, 理解这些模型对我们来说是必要的.

所有的生成式模型都是逐渐从噪声转变为数据. 这个演化主要是由模拟 ODE 或 SDE 实现的. 我们利用神经网络来构造, 训练和模拟这些 ODE/SDE 的过程. 这些模型很容易实现, 但是其背后的数学机理却不那么容易理解. 这份笔记主要记录了理解 flow matching 和 denoising diffusion models 所必要的数学原理.

一些数学符号的意义:

- u : 向量场
- ψ : flow, ODE 的解
- p : 概率分布
- p_{data} : 目标分布
- p_{init} : 初始分布
- $\text{Unif} = \text{Unif}_{[0,1]}$

1.1 Flow Model

1.1.1 ODE 的定义

ODE 的解被 trajectory 所定义, 例如如下形式:

$$X : [0, 1] \mapsto \mathbb{R}^d, t \mapsto X_t,$$

从时间 t 映射到 \mathbb{R}^d 空间的某个位置. 每一个 ODE 被一个向量场 (vector field) u 所定义, 例如如下形式:

$$u : \mathbb{R}^d \times [0, 1] \mapsto \mathbb{R}^d, (x, t) \mapsto u_t(x)$$

一个 ODE 定义了一条轨迹 (trajectory): 一条轨迹 X , 从 x_0 出发, 沿着向量场 u_t 前进. 我们利用如下方程的解来定义一条轨迹:

$$\frac{d}{dx} X_t = u_t(X_t), \quad \text{ODE} \quad (1.1a)$$

$$X_0 = x_0, \quad \text{initial conditions} \quad (1.1b)$$

公式 3.7a 要求 X_t 的导数直接由 u_t 提供, 公式 1.1b 要求我们在 $t = 0$ 时刻从 x_0 出发. 在 t 时刻, 我们所在的位置 X_t 由被称为流 (flow) 的函数定义, 也就是我们 ODE 得到解:

$$\psi : \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}^d, (x_0, t) \mapsto \psi_t(x_0) \quad (1.2a)$$

$$\frac{d}{dt} \psi_t(x_0) = u_t(\psi_t(x_0)), \quad \text{flow ODE} \quad (1.2b)$$

$$\psi_0(x_0) = x_0, \quad \text{flow initial conditions} \quad (1.2c)$$

给定的初始条件 $X_0 = x_0$, 我们可以通过 $X_t = \psi_0(X_0)$ 恢复出一条 ODE 的轨迹. 因此, 向量场 (vector fields), ODEs, 和流 (flows) 的关系是:

- 向量场 (vector fields) 定义了 ODEs
- 流 (flows) 是 ODEs 的解

同时, 对于每一个 ODE, 都有解的存在性和唯一性问题, 若我们对 u_t 做一个弱的假设, 则有:

定理 1.1 (Flow existence and uniqueness)

If $u : \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}^d$ is continuously differentiable with a bounded derivative, then the ODE in 1.2 has a unique solution given by a flow ψ_t . Moreover, ψ_t is a **diffeomorphism** for all t , meaning ψ_t is continuously differentiable with a continuously differentiable inverse ψ_t^{-1} .

若 $u : \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}^d$ 连续可微且导数有界, 则方程 1.2 的 ODE 存在唯一解, 即流 ψ_t . 此时, ψ_t 对所有 t 都是微分同胚 (diffeomorphism), 即 ψ_t 连续可微, 且其逆映射 ψ_t^{-1} 也连续可微。

1.1.2 模拟 ODE/数值解

若 ψ_t 的解析解很难得出, 可以数值方法模拟 ODEs, 即求 ODEs 的数值解. 其中一个最简单也最直观的方法是欧拉法 (Euler method): 我们初始化 $X_0 = x_0$, 并通过下式更新:

$$X_{t+h} = X_t + hu_t(X_t), t = 0, h, 2h, 3h, \dots, 1 - h \quad (1.3)$$

其中 $h = 1/n, n \in \mathbb{N}$ 是一个超参数. 若想尝试更复杂的方法, 可以考虑 Heun's method:

$$X'_{t+h} = X_t + hu_t(X_t) \quad (\text{初始猜测的新状态}) \quad (1.4a)$$

$$X_{t+h} = X_t + \frac{h}{2} (u_t(X_t) + u_{t+h}(X'_{t+h})) \quad (\text{用当前和猜测状态的平均值更新}) \quad (1.4b)$$

1.1.3 Flow models

现在我们可以通过 ODE 来构建一个生成式模型. 我们的目标是将一个简单的分布 p_{init} 转变成一个复杂分布 p_{data} . 因此, 使用 ODE 来模拟这个转变的过程是一个很自然的选择. 一个 flow model 可以通过如下的 ODE 来描述:

$$X_0 \sim p_{init} \quad \text{random initialization} \quad (1.5a)$$

$$\frac{d}{dt} X_t = u_t^\theta(X_t) \quad \text{ODE} \quad (1.5b)$$

向量场 u_t^θ 是一个以 θ 为参数的神经网络. 对于现在来说, u_t^θ 就是一个通用的神经网络, 例如一个用有参数 θ 的连续的方程 $u_t^\theta: \mathbb{R}^d \times [0, 1] \mapsto \mathbb{R}^d$. 具体的网络结构以后再讨论. 我们的目标是使得轨迹 (trajectory) 的终点 X_1 服从 p_{data} 的分布:

$$X_1 \sim p_{data} \Leftrightarrow \psi_1^\theta(X_0) \sim p_{data} \tag{1.6}$$

$\psi_t^\theta(X_0)$ 是由 u_t^θ 描述的流 (flow).

尽管我们现在叫做 flow model, 实际上神经网络表征的是向量场, 而不是流 (flow). 为了计算流 (flow), 我们需要模拟 ODE 的过程. 在算法 1 中, 我们总结了 flow model 的采样方法.

Algorithm 1 Sampling from a Flow Model with Euler method

Require: Neural network vector field u_t^θ , number of steps n

- 1: Set $t = 0$
 - 2: Set step size $h = \frac{1}{n}$
 - 3: Draw a sample $X_0 \sim p_{init}$
 - 4: **for** $i = 1, \dots, n - 1$ **do**
 - 5: $X_{t+h} = X_t + hu_t^\theta(X_t)$
 - 6: Update $t \leftarrow t + h$
 - 7: **end for**
 - 8: **return** X_1
-

构造训练目标

在上一节, 我们构造了服从轨迹 $(X_t)_{0 \leq t \leq 1}$ 的 flow model, 其模拟的 ODE 为:

$$X_0 \sim p_{init}, \quad dX_t = u_t^\theta(X_t)dt \tag{1.7}$$

u_t^θ 是一个神经网络, 通过训练可以使得这个网络正确的模拟 ODE. 我们通过最小化损失函数 $\mathcal{L}(\theta)$ 来训练参数 θ , 例如 MSE(mean-squared error):

$$\mathcal{L}(\theta) = \|u_t^\theta(x) - \underbrace{u_t^{\text{target}}(x)}_{\text{training target}}\|^2 \tag{1.8}$$

1.2.1 条件与边际概率路径 (Conditional and Marginal Probability Path)

构造 u_t^{target} 的第一步是确定一个概率路径 (probability path). 一个概率路径确定了一条从噪声 p_{init} 到数据 p_{data} 的内插路径. 接下来, 对于数据点 $z \in \mathbb{R}^d$, 我们令其服从 δ_z Dirac delta 分布. 一个 conditional (interpolating) probability path 是指一组在 \mathbb{R}^d 上, 服从分布 $p_t(x|z)$ 的集合:

$$p_0(\cdot|z) = p_{init}, \quad p_1(\cdot|z) = \delta_z \quad \text{for all } z \in \mathbb{R}^d \tag{1.9}$$

换言之, 条件概率路径 (conditional probability path) 逐渐将一个数据点 z 转变进分布 p_{init} . 由每个条件概率路径 (conditional probability path) $p_t(x|z)$ 可以推导出边际概率路径 (marginal probability path) $p_t(x)$:

$$z \sim p_{data}, \quad x \sim p_t(\cdot|z) \Rightarrow x \sim p_t, \quad \text{sampling from marginal path} \tag{1.10}$$

$$p_t(x) = \int p_t(x|z)p_{data}(z) dz, \quad \text{density of marginal path} \tag{1.11}$$

边际概率路径 (marginal probability path) p_t 是 p_{init} 到数据 p_{data} 的内插路径:

$$p_0 = p_{init} \quad \text{and} \quad p_1 = p_{data} \quad (\text{noise-data interpolation}) \tag{1.12}$$

示例 1.1. 高斯概率路径 (Gaussian Conditional Probability Path)

高斯概率路径 (Gaussian probability path) 是一个特别流行的概率路径, 这也是 **denoising diffusion models** 所使用的概率路径. 设 α_t, β_t 是噪声调度器 (**noise schedulers**): 它们是两个连续可微的、单调的函数, 满足 $\alpha_0 = \beta_1 = 0$ 和 $\alpha_1 = \beta_0 = 1$. 我们定义条件概率路径为:

$$p_t(\cdot|z) = \mathcal{N}(\alpha_t z, \beta_t^2 I) \quad (\text{高斯条件路径}) \tag{1.13}$$

根据我们对 α_t 和 β_t 的约束条件, 这个路径满足:

$$p_0(\cdot|z) = \mathcal{N}(\alpha_0 z, \beta_0^2 I) = \mathcal{N}(0, I), \quad p_1(\cdot|z) = \mathcal{N}(\alpha_1 z, \beta_1^2 I) = \delta_z, \tag{1.14}$$

从边缘概率路径 p_t 中采样可得:

$$z \sim p_{\text{data}}, \quad x \sim p_t(\cdot|z) \Rightarrow x = \alpha_t z + \beta_t \epsilon \sim p_t \tag{1.15}$$

1.2.2 条件与边际向量场 (Conditional and Marginal Vector Fields)

基于前文定义的概率路径 p_t , 我们现通过解析方法构建流模型的训练目标 u_t^{target} . 该方法的核心在于: 利用可解析推导 (derive analytically by hand) 的简单分量来构造该向量场。

定理 1.2 (Marginalization trick)

对于每个数据点 $z \in \mathbb{R}^d$, 设 $u_t^{\text{target}}(\cdot|z)$ 表示一个条件向量场, 使得对应的 ODE 产生条件概率路径 $p_t(\cdot|z)$, 即:

$$X_0 \sim p_{\text{init}}, \quad \frac{d}{dt} X_t = u_t^{\text{target}}(X_t|z) \Rightarrow X_t \sim p_t(\cdot|z) \quad (0 \leq t \leq 1) \tag{1.16}$$

那么由下式定义的边缘向量场 $u_t^{\text{target}}(x)$:

$$u_t^{\text{target}}(x) = \int u_t^{\text{target}}(x|z) \frac{p_t(x|z) p_{\text{data}}(z)}{p_t(x)} dz \tag{1.17}$$

将遵循边缘概率路径, 即:

$$X_0 \sim p_{\text{init}}, \quad \frac{d}{dt} X_t = u_t^{\text{target}}(X_t) \Rightarrow X_t \sim p_t \quad (0 \leq t \leq 1) \tag{1.18}$$

特别地, 对于该 ODE 有 $X_1 \sim p_{\text{data}}$, 因此可以说“ u_t^{target} 将噪声 p_{init} 转换为数据 p_{data} ”。

示例 1.2. 高斯概率路径的 target ODE

如前所述, 设 $p_t(\cdot|z) = \mathcal{N}(\alpha_t z, \beta_t^2 I_d)$, 其中 α_t, β_t 是噪声调度器 (参见式 (1.13)). 令 $\dot{\alpha}_t = \partial_t \alpha_t$ 和 $\dot{\beta}_t = \partial_t \beta_t$ 分别表示 α_t 和 β_t 的时间导数. 这里, 我们想要证明由下式给出的 *conditional Gaussian vector field*:

$$u_t^{\text{target}}(x|z) = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t} \alpha_t \right) z + \frac{\dot{\beta}_t}{\beta_t} x \tag{1.19}$$

是定理 1.2 意义下的有效条件向量场: 如果 $X_0 \sim \mathcal{N}(0, I_d)$, 则其 ODE 轨迹 X_t 满足 $X_t \sim p_t(\cdot|z) = \mathcal{N}(\alpha_t z, \beta_t^2 I_d)$.

证明 首先我们通过定义构造一个条件流模型 $\psi_t^{\text{target}}(x|z)$:

$$\psi_t^{\text{target}}(x|z) = \alpha_t z + \beta_t x \tag{1.20}$$

如果 X_t 是 $\psi_t^{\text{target}}(\cdot|z)$ 的 ODE 轨迹且 $X_0 \sim p_{\text{init}} = \mathcal{N}(0, I_d)$, 那么根据定义:

$$X_t = \psi_t^{\text{target}}(X_0|z) = \alpha_t z + \beta_t X_0 \sim \mathcal{N}(\alpha_t z, \beta_t^2 I_d) = p_t(\cdot|z) \quad (1.21)$$

我们得出结论: 轨迹的分布与条件概率路径相同 (即满足式 (18)). 剩下的工作是从 $\psi_t^{\text{target}}(x|z)$ 中提取向量场 $u_t^{\text{target}}(x|z)$. 根据流的定义 (式 1.2), 对于所有 $x, z \in \mathbb{R}^d$ 有:

$$\frac{d}{dt} \psi_t^{\text{target}}(x|z) = u_t^{\text{target}}(\psi_t^{\text{target}}(x|z)|z) \quad \text{对所有 } x, z \in \mathbb{R}^d \quad (1.22)$$

$$\Leftrightarrow \dot{\alpha}_t z + \dot{\beta}_t x = u_t^{\text{target}}(\alpha_t z + \beta_t x|z) \quad \text{对所有 } x, z \in \mathbb{R}^d \quad (1.23)$$

$$\Leftrightarrow \dot{\alpha}_t z + \dot{\beta}_t \left(\frac{x - \alpha_t z}{\beta_t} \right) = u_t^{\text{target}}(x|z) \quad \text{对所有 } x, z \in \mathbb{R}^d \quad (1.24)$$

$$\Leftrightarrow \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t} \alpha_t \right) z + \frac{\dot{\beta}_t}{\beta_t} x = u_t^{\text{target}}(x|z) \quad \text{对所有 } x, z \in \mathbb{R}^d \quad (1.25)$$

其中在 (i) 中我们使用了 $\psi_t^{\text{target}}(x|z)$ 的定义 (式 (1.20)), 在 (ii) 中我们进行了变量替换 $x \rightarrow (x - \alpha_t z)/\beta_t$, 在 (iii) 中我们只是做了一些代数运算. 注意最后一个等式正是我们在式 (1.19) 中定义的条件高斯向量场. 证毕.

定理 1.3 (连续性方程)

考虑一个具有向量场 u_t^{target} 且 $X_0 \sim p_{\text{init}}$ 的流模型. 那么对于所有 $0 \leq t \leq 1$ 有 $X_t \sim p_t$ 当且仅当满足连续性方程:

$$\partial_t p_t(x) = -\text{div}(p_t u_t^{\text{target}})(x) \quad \text{对所有 } x \in \mathbb{R}^d, 0 \leq t \leq 1 \quad (1.26)$$

其中 $\partial_t p_t(x) = \frac{d}{dt} p_t(x)$ 表示 $p_t(x)$ 的时间导数. 方程 1.26 被称为连续性方程.

1.3 训练

前两节我们如何通过神经网络给定的向量场 u_t^θ 构造生成式模型, 还展示了训练目标 u_t^{target} 的形式, 这一节我们将展示如何使用 u_t^θ 来逼近 u_t^{target} .

Flow matching loss 被定义为:

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{t \sim \text{Unif}, x \sim p_t} [\|u_t^\theta(x) - u_t^{\text{target}}(x)\|^2] \quad (1.27)$$

通过使用式 1.10 的采样过程, 我们可以将公式 1.27 重写为:

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, x \sim p_t(\cdot|z)} [\|u_t^\theta(x) - u_t^{\text{target}}(x)\|^2] \quad (1.28)$$

这两个形式是等价的, 因为边际概率路径 $p_t(x)$ 可以表示为: $p_t(x) = \int p_t(x|z) p_{\text{data}}(z) dz$. 通过定理 1.2 可知 u_t^{target} :

$$u_t^{\text{target}}(x) = \int u_t^{\text{target}}(x|z) \frac{p_t(x|z) p_{\text{data}}(z)}{p_t(x)} dz \quad (1.29)$$

因为上述积分不可处理, 我们转而使用好处理的条件向量场 $u_t^{\text{target}}(x|z)$. 首先, 我们定义 **Conditional Flow Matching loss**:

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t \sim \mathcal{U}(0,1), z \sim p_{\text{data}}, x \sim p_t(\cdot|z)} [\|u_t^\theta(x) - u_t^{\text{target}}(x|z)\|^2] \quad (1.30)$$

定理 1.4

Marginal Flow Matching loss 等于 Conditional Flow Matching loss 加一个常数:

$$\mathcal{L}_{\text{FM}}(\theta) = \mathcal{L}_{\text{CFM}}(\theta) + C \quad (1.31)$$

C 与 θ 无关, 也就是说:

$$\nabla_{\theta} \mathcal{L}_{\text{FM}}(\theta) = \nabla_{\theta} \mathcal{L}_{\text{CFM}}(\theta) \quad (1.32)$$

因此, 最小化 $\mathcal{L}_{\text{FM}}(\theta)$ 就是最小化 $\mathcal{L}_{\text{CFM}}(\theta)$

证明

$$\begin{aligned} \mathcal{L}_{\text{FM}}(\theta) &\stackrel{(i)}{=} \mathbb{E}_{t \sim \text{Unif}, x \sim p_t} [\|u_t^{\theta}(x) - u_t^{\text{target}}(x)\|^2] \\ &\stackrel{(ii)}{=} \mathbb{E}_{t \sim \text{Unif}, x \sim p_t} [\|u_t^{\theta}(x)\|^2 - 2u_t^{\theta}(x)^T u_t^{\text{target}}(x) + \|u_t^{\text{target}}(x)\|^2] \\ &\stackrel{(iii)}{=} \mathbb{E}_{t \sim \text{Unif}, x \sim p_t} [\|u_t^{\theta}(x)\|^2] - 2\mathbb{E}_{t \sim \text{Unif}, x \sim p_t} [u_t^{\theta}(x)^T u_t^{\text{target}}(x)] + \underbrace{\mathbb{E}_{t \sim \text{Unif}, x \sim p_t} [\|u_t^{\text{target}}(x)\|^2]}_{=: C_1} \\ &\stackrel{(iv)}{=} \mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, x \sim p_t(\cdot|z)} [\|u_t^{\theta}(x)\|^2] - 2\mathbb{E}_{t \sim \text{Unif}, x \sim p_t} [u_t^{\theta}(x)^T u_t^{\text{target}}(x)] + C_1 \end{aligned}$$

其中 (i) 就是定义, (ii) 就是 $\|a - b\|^2 = \|a\|^2 - 2ab + \|b\|^2$, (iii) 我们定义了一个常数 C_1 , 最后在 (iv) 我们使用了式 1.10 的采样过程得到 p_t . 对第二项继续展开:

$$\begin{aligned} \mathbb{E}_{t \sim \text{Unif}, x \sim p_t} [u_t^{\theta}(x)^T u_t^{\text{target}}(x)] &\stackrel{(i)}{=} \int_0^1 \int p_t(x) u_t^{\theta}(x)^T u_t^{\text{target}}(x) dx dt \\ &\stackrel{(ii)}{=} \int_0^1 \int p_t(x) u_t^{\theta}(x)^T \left(\int u_t^{\text{target}}(x|z) \frac{p_t(x|z) p_{\text{data}}(z)}{p_t(x)} dz \right) dx dt \\ &\stackrel{(iii)}{=} \int_0^1 \int \int u_t^{\theta}(x)^T u_t^{\text{target}}(x|z) p_t(x|z) p_{\text{data}}(z) dz dx dt \\ &\stackrel{(iv)}{=} \mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, x \sim p_t(\cdot|z)} [u_t^{\theta}(x)^T u_t^{\text{target}}(x|z)] \end{aligned}$$

(i) 我们使用积分展开, (ii) 我们使用了公式 1.17, (iii) 普通乘法, 最后用 (iv) 我们将公式表示为期望的形式. 在证明一开始我们使用了 marginal vector field $u_t^{\text{target}}(x)$, 而现在我们使用了 conditional vector field $u_t^{\text{target}}(x|z)$, 我们将其代入 \mathcal{L}_{FM} :

$$\begin{aligned} \mathcal{L}_{\text{FM}}(\theta) &\stackrel{(i)}{=} \mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, x \sim p_t(\cdot|z)} [\|u_t^{\theta}(x)\|^2] - 2\mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, x \sim p_t(\cdot|z)} [u_t^{\theta}(x)^T u_t^{\text{target}}(x|z)] + C_1 \\ &\stackrel{(ii)}{=} \mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, x \sim p_t(\cdot|z)} [\|u_t^{\theta}(x)\|^2 - 2u_t^{\theta}(x)^T u_t^{\text{target}}(x|z) + \|u_t^{\text{target}}(x|z)\|^2 - \|u_t^{\text{target}}(x|z)\|^2] + C_1 \\ &\stackrel{(iii)}{=} \mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, x \sim p_t(\cdot|z)} [\|u_t^{\theta}(x) - u_t^{\text{target}}(x|z)\|^2] + \underbrace{\mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, x \sim p_t(\cdot|z)} [-\|u_t^{\text{target}}(x|z)\|^2]}_{C_2} + C_1 \\ &\stackrel{(iv)}{=} \mathcal{L}_{\text{CFM}}(\theta) + \underbrace{C_2 + C_1}_{=: C} \end{aligned}$$

(i) 利用上式替换, (ii) 从 0 中裂项, (iii) 再次使用公式 $\|a - b\|^2 = \|a\|^2 - 2a^T b + \|b\|^2$, 最后在 (iv) 我们定义了一个与 θ 无关的常数.

证毕.

一旦我们得到训练好的 u_t^{θ} , 我们可以模拟流模型 (flow model):

$$dX_t = u_t^{\theta}(X_t) dt, \quad X_0 \sim p_{\text{init}} \quad (1.33)$$

通过算法 1 我们可以采样得出 $X_1 \sim p_{\text{data}}$, 通过算法 2 我们可以训练出 u_t^{θ}

Algorithm 2 Flow Matching Training Procedure (for Gaussian CondOT path $p_t(x|z) = \mathcal{N}(tz, (1-t)^2)$)**Require:** A dataset of samples $z \sim p_{\text{data}}$, neural network u_t^θ

- 1: **for** each mini-batch of data **do**
- 2: Sample a data example z from the dataset.
- 3: Sample a random time $t \sim \text{Unif}[0, 1]$.
- 4: Sample noise $\epsilon \sim \mathcal{N}(0, I_d)$.
- 5: Set $x = tz + (1-t)\epsilon$ (General case: $x \sim p_t(\cdot|z)$)
- 6: Compute loss

$$\mathcal{L}(\theta) = \|u_t^\theta(x) - (z - \epsilon)\|, \quad (\text{General case := } \|u_t^\theta(x) - u_t^{\text{target}}(x|z)\|^2)$$

- 7: Update the model parameters θ via gradient descent on $L(\theta)$.
- 8: **end for**

示例 1.3. Flow Matching for Gaussian Conditional Probability Paths

让我们回到高斯条件概率路径 $p_t(\cdot|z) = \mathcal{N}(\alpha_t z; \beta_t^2 I_d)$ 的例子，其中我们可以通过以下方式从条件路径中采样：

$$\epsilon \sim \mathcal{N}(0, I_d) \quad \Rightarrow \quad x_t = \alpha_t z + \beta_t \epsilon \sim \mathcal{N}(\alpha_t z, \beta_t^2 I_d) = p_t(\cdot|z). \quad (1.34)$$

正如我们在公式 1.25 中推导的那样，条件向量场 $u_t^{\text{target}}(x|z)$ 由下式给出：

$$u_t^{\text{target}}(x|z) = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t} \alpha_t \right) z + \frac{\dot{\beta}_t}{\beta_t} x, \quad (1.35)$$

其中 $\dot{\alpha}_t = \frac{d\alpha_t}{dt}$ 和 $\dot{\beta}_t = \frac{d\beta_t}{dt}$ 分别是 α_t 和 β_t 的时间导数。将这个公式代入，条件流匹配损失为：

$$\begin{aligned} \mathcal{L}_{\text{CFM}}(\theta) &= \mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, x \sim \mathcal{N}(\alpha_t z, \beta_t^2 I_d)} \left[\|u_t^\theta(x) - \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t} \alpha_t \right) z - \frac{\dot{\beta}_t}{\beta_t} x\|^2 \right] \\ &\stackrel{(i)}{=} \mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, \epsilon \sim \mathcal{N}(0, I_d)} \left[\|u_t^\theta(\alpha_t z + \beta_t \epsilon) - (\dot{\alpha}_t z + \dot{\beta}_t \epsilon)\|^2 \right] \end{aligned}$$

在 (i) 中，利用公式 1.34，替换 x 为 $\alpha_t z + \beta_t \epsilon$ 。 \mathcal{L}_{CFM} 非常简洁：我们只需要采样一个数据点 z ，再采样一些噪声 ϵ ，然后计算均方误差即可。在 $\alpha_t = t$ 和 $\beta_t = 1-t$ 的特殊情况下，对应的概率 $p_t(x|z) = \mathcal{N}(tz, (1-t)^2)$ 有时被称为 (高斯)CondOT 概率路径。那么我们有 $\dot{\alpha}_t = 1$ 和 $\dot{\beta}_t = -1$ ，因此：

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, \epsilon \sim \mathcal{N}(0, I_d)} \left[\|u_t^\theta(tz + (1-t)\epsilon) - (z - \epsilon)\|^2 \right].$$

许多著名的最先进的模型都是通过这种简单而有效的程序训练的，例如 Stable Diffusion 3、Meta 的 Movie Gen Video，以及许多其他专有模型。我们在算法 2 中总结了训练过程。

第 2 章 Mean Flows 阅读笔记

May 24, 2025

5 月 19 日, Kaiming 团队在 arxiv 上提交了论文 “Mean Flows for One-step Generative Modeling”, 提出了 Mean Flows 方法. Flow Matching 构建了两个分布的一个微分过程, 并通过逐步采样的方式实现分布之间的转移, 但有着推断时间过长的问题. 为了解决这个问题, Kaiming 团队构建了这样一个微分过程的平均过程, 名为 Mean Flows.

本章的符号表示与第 1 章不同, 采用 Mean Flows 论文中相同的符号表示.

2.1 Flow Matching

取 $x \sim p_{\text{data}}(x)$, $\epsilon \sim p_{\text{prior}}(\epsilon)$ (p_{prior} 就是 p_{init} , 例如噪声), 在时间步骤 t 下, 流过程可以表示为 $z_t = a_t x + b_t \epsilon$, a_t, b_t 是 schedules. 其速度 v_t 可以表示为 $v_t = z'_t = a'_t x + b'_t \epsilon$ ($'$ 代表对时间的导数). 一个常用的 schedule 是 $a_t = 1 - t$, $b_t = t$, 在这个条件下, $v_t = \epsilon - x$.

构建微分过程:

$$\frac{d}{dt} z_t = v(z_t, t) \quad (2.1)$$

从 $z_1 = \epsilon \sim p_{\text{prior}}$ 开始, 其解可以被写成 $z_r = z_t - \int_r^t v(z_\tau, \tau) d\tau$, 其中 r 是另一个时间步.

Flow Matching loss $\mathcal{L}_{\text{FM}}(\theta)$ 被定义为:

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{t, p_t(z_t)} \|v_\theta(z_t, t) - v(z_t, t)\|^2 \quad (2.2)$$

Conditional Flow Matching loss $\mathcal{L}_{\text{CFM}}(\theta)$ 被定义为:

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t, x, \epsilon} \|v_\theta(z_t, t) - v(z_t | x)\|^2 \quad (2.3)$$

\mathcal{L}_{FM} 与 \mathcal{L}_{CFM} 只差一个常数.

更详细的内容见第 1 章.

2.2 MeanFlow Models

2.2.1 MeanFlow 定义

定义时间步 t 和 r 之间速度场的平均值 u 为:

$$u(z_t, r, t) \triangleq \frac{1}{t-r} \int_r^t v(z_\tau, \tau) d\tau \quad (2.4)$$

即:

$$(t-r)u(z_t, r, t) = \int_r^t v(z_\tau, \tau) d\tau \quad (2.5)$$

接下来先做一些数学变换. 两边同时求导得:

$$\frac{d}{dt}((t-r)u(z_t, r, t)) = \frac{d}{dt} \int_r^t v(z_\tau, \tau) d\tau \Rightarrow u(z_t, r, t) + (t-r) \frac{d}{dt} u(z_t, r, t) = v(z_t, t), \quad (2.6)$$

将上式重排得:

$$\boxed{\underbrace{u(z_t, r, t)}_{\text{Average velocity}} = \underbrace{v(z_t, t)}_{\text{Instant velocity}} - (t-r) \underbrace{\frac{d}{dt} u(z_t, r, t)}_{\text{Time derivative of } u}} \quad (2.7)$$

公式2.7即为 Mean Flows 的定义. 对公式2.7两边继续求导:

$$\frac{d}{dt} u(z_t, r, t) = \underbrace{\frac{dz_t}{dt} \frac{\partial u}{\partial z}}_{\text{JVP with } v} + \underbrace{\frac{dr}{dt} \frac{\partial u}{\partial r}}_{=0} + \underbrace{\frac{dt}{dt} \frac{\partial u}{\partial t}}_{=\partial_t u} \quad (2.8)$$

也就是说:

$$\boxed{\frac{d}{dt} u(z_t, r, t) = v(z_t, t) \frac{\partial u}{\partial z} + \frac{\partial u}{\partial t}} \quad (2.9)$$

2.2.2 MeanFlow 损失函数

对于参数化网络 u_θ 来说, 损失函数定义为:

$$\mathcal{L}(\theta) = \mathbb{E} \|u_\theta(z_t, r, t) - \text{sg}(u_{\text{tgt}})\|_2^2, \quad (2.10)$$

$$\text{其中 } u_{\text{tgt}} = v_t - (t-r) \left(v_t \frac{\partial u_\theta}{\partial z} + \frac{\partial u_\theta}{\partial t} \right) \quad (2.11)$$

公式2.9带入公式2.7得到公式2.11. sg 代表 stop-gradient.

2.2.3 Sampling

将公式2.1的解写成平均值的形式:

$$z_r = z_t - (t-r)u(z_t, r, t) \quad (2.12)$$

一步采样 (1-step sampling) 为: $z_0 = z_1 - u(z_1, 0, 1)$. $z_1 = \epsilon \sim p_{\text{prior}}(\epsilon)$

2.3 一些其他细节

本节内容主要由 DeepSeek 生成.

2.3.1 Stop-Gradient

2.3.1.1 Stop-Gradient 的数学定义

符号表示: 通常记为 $\text{sg}(\cdot)$ 或 $\text{detach}()$ (在 PyTorch 中)。

数学行为：对任意变量 x ，有：

$$\text{sg}(x) = x \quad (\text{前向传播时}) \tag{2.13}$$

$$\frac{\partial \text{sg}(x)}{\partial \theta} = 0 \quad (\text{反向传播时}) \tag{2.14}$$

即，前向传播时保留原值，反向传播时梯度为零。

2.3.1.2 在 MeanFlow 中的具体作用

在论文的式 2.10 和式 2.11 中，Stop-Gradient 用于训练目标 u_{tgt} 。

为什么需要 Stop-Gradient?

1. **避免高阶导数**： u_{tgt} 中已经包含 u_θ 的导数项（如 $\frac{\partial u_\theta}{\partial z}$ ）。如果不对 u_{tgt} 停梯度，损失函数会需要对 u_θ 求二阶导数（即 $\frac{\partial^2 u_\theta}{\partial \theta \partial z}$ ），计算复杂且不稳定。
2. **固定目标值**：停梯度将 u_{tgt} 视为一个常数目标，仅让 u_θ 去拟合它，而不是动态调整 u_{tgt} 本身。这类似于强化学习中的“固定目标网络”技术。
3. **理论一致性**：MeanFlow 的推导（式 2.7）要求 u_{tgt} 是理论计算值，而非可学习的变量。停梯度强制模型遵守这一约束。

2.3.2 JVP

在本论文中，给定一个函数 $u_\theta(z, r, t)$ ，其雅可比矩阵 J 包含所有偏导数：

$$J = \begin{pmatrix} \frac{\partial u}{\partial z} & \frac{\partial u}{\partial r} & \frac{\partial u}{\partial t} \end{pmatrix} \tag{2.15}$$

JVP 计算的是雅可比矩阵 J 与一个向量 $\mathbf{v} = (v, 0, 1)$ 的乘积：

$$J \cdot \mathbf{v} = \frac{\partial u}{\partial z} v + \frac{\partial u}{\partial r} \cdot 0 + \frac{\partial u}{\partial t} \cdot 1 = v \frac{\partial u}{\partial z} + \frac{\partial u}{\partial t} \tag{2.16}$$


这正是论文中式 2.8 和式 2.11 的推导结果。

为什么使用 JVP?

1. **避免显式构造雅可比矩阵**：雅可比矩阵的显式存储需要 $O(n^2)$ 内存（ n 是输出维度），而 JVP 通过一次前向和反向传播直接计算乘积，内存效率为 $O(n)$ 。
2. **支持停梯度 (Stop-Gradient)**：在式 (9) 的损失函数中， u_{tgt} 的导数项（通过 JVP 计算）被设为常数（detach()），避免高阶导数计算。
3. **与 Flow Matching 的兼容性**：传统 Flow Matching 只需计算 $v(z, t)$ ，而 MeanFlow 通过 JVP 引入对 u_θ 的导数约束，实现单步生成。

与其他自动微分操作的对比：

操作	计算模式	适用场景	内存开销
JVP	前向模式	计算“函数输出对某一输入的导数”	$O(n)$
反向传播 (VJP)	反向模式	计算“标量损失对大量参数的梯度”	$O(m)$
全雅可比矩阵	显式构造	需要完整导数矩阵（如二阶优化）	$O(n^2)$



第 3 章 MIT 6.S184 学习笔记二: SDE

May 25, 2025

「選べない……どっちも欲しい」¹

——《School Days》

近年来,生成式模型在人工智能领域掀起了一场静默的革命.从 GAN 的对抗博弈到 Flow 模型的精确可逆,研究者们不断探索数据生成的本质规律.而在这条技术演进的路径上,扩散模型 (Diffusion Model) 名声大噪,DDPM (Denoising Diffusion Probabilistic Model) 为 LDM(Latent Diffusion Models) 以及后续 Stable Diffusion 模型的出现打下了坚实的基础,但其数学理论未触及本质. Song 提出的 SDE(Stochastic Differential Equation) 为扩散过程提供了坚实的数学框架,将数据的逐步噪声化与生成视为连续时间的随机过程. Song 的理论不仅解释了扩散模型的本质,更在图像生成、音频合成等领域实现了超越原有方法的逼真效果.

在 DDPM 原文中, Ho 构建了一个马尔科夫过程 (Forward process 和 Reverse process) 来实现两个数据分布之间转换.而在 SDE 的视角下, DDPM 的前向过程 (Forward process) 仅为 SDE 的高斯概率路径演化的一种特殊情况.在 Song 的原文中,使用了 SDE 时间反演的技巧,虽然可以这么做,但略显麻烦.在本文中,我们构造 SDE,使其与 ODE 有相同的概率路径演化,形式上与 ODE(Flow Matching) 的过程保持一致 (Flow Matching 详见第 1 章).

本文将侧重 SDE 的数学理论,主要参考资料为 MIT Class 6.S184 课程,网址为 <https://diffusion.csail.mit.edu/>.

3.1 伊藤过程 (Itô Diffusion Process)

伊藤过程的 SDE(Stochastic Differential Equations) 的构建源于对物理、金融等领域中随机动力学系统的数学建模需求.其核心思想是将确定性微分方程 (ODE) 与随机噪声 (维纳过程) 结合,以描述受随机扰动影响的系统演化.以下是其来源与推导的逐步解析:

¹“无法选择……我全都要”.在随机微分方程的世界里,路径的选择往往和某著名伊藤氏的人生一样,充满不可逆的纠结

3.1.1 物理背景: 从确定性到随机性

3.1.1.1 经典动力学方程

确定性系统通常用常微分方程 (ODE) 描述:

$$\frac{dX_t}{dt} = \mu(X_t, t) \quad (3.1)$$

其中 $\mu(X_t, t)$ 为确定性漂移项 (如速度场).

3.1.1.2 引入随机扰动

在实际系统中, 噪声和不确定性是普遍存在的. 为了更好地描述这些现象, 我们引入了随机扰动项: 当系统受微观随机扰动 (如分子碰撞、市场波动) 时, 需在方程中加入噪声项:

$$\frac{dX_t}{dt} = \mu(X_t, t) + \sigma(X_t, t) \cdot \text{噪声}, \quad (3.2)$$

其中 $\sigma(X_t, t)$ 控制噪声强度.

关键问题: 如何数学上严格定义“噪声”?

3.1.2 噪声的数学建模: 维纳过程

3.1.2.1 白噪声与维纳过程

理想白噪声 $\xi(t)$: 均值为零, 瞬时相关 $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$, 但数学上难以直接处理.

维纳过程 W_t : 白噪声的积分 $W_t = \int_0^t \xi(s)ds$ 是数学上良好的定义, 满足:

- $W_0 = 0$, 初始值为零.
- 增量 $\Delta W_t = W_{t+\Delta t} - W_t$ 服从正态分布: $\Delta W_t \sim \mathcal{N}(0, \Delta t)$, 均值为零, 方差为 Δt .
- 增量独立性: 对于任意不重叠的时间区间, 增量相互独立.

3.1.2.2 随机微分方程的启发式形式

将噪声替换为维纳过程的微分 dW_t :

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t \quad (3.3)$$

注意: dW_t 并非经典微分, 需通过积分严格定义.

3.2 扩散模型 (Diffusion Model)

在 Diffusion Model 中, 我们可以使用如下 SDE 来表示:

$$dX_t = u_t(X_t)dt + \sigma_t dW_t, \quad \text{SDE} \quad (3.4a)$$

$$X_0 = x_0, \quad \text{initial conditions} \quad (3.4b)$$

与 ODE 不同的是, 不存在一个流映射 (flow map) 来表示 SDE 的解². 由于 SDE 的演化过程是随机的, 所以 X_t 不再是由 $X_0 \sim p_{\text{init}}$ 所唯一决定的.

²ODE 中的流映射是确定性映射, 而 SDE 的解受随机噪声影响, 无法用单一映射描述

对于 SDE 的轨迹 (trajectory) $(X_t)_{0 \leq t \leq 1}$ 来说³, 一个小的 step 的方向来自于布朗运动加上 $u_t(X_t)$ 方向:

$$X_{t+h} = X_t + \underbrace{hu_t(X_t)}_{\text{deterministic}} + \underbrace{\sigma_t(W_{t+h} - W_t)}_{\text{stochastic}} + \underbrace{hR_t(h)}_{\text{error term}} \tag{3.5}$$

其中 $\sigma_t \geq 0$ 是扩散系数 (diffusion coefficient), $R_t(h)$ 描述的随机误差项满足 $\mathbb{E}[\|R_t(h)\|^2]^{1/2} \rightarrow 0$ (当 $h \rightarrow 0$).

当然, 类似于 ODE, 我们有:

定理 3.1 (SDE 解的存在唯一性定理 (SDE Solution Existence and Uniqueness))

If $u : \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}^d$ is continuously differentiable with a bounded derivative and σ_t is continuous, then the SDE in (7) has a solution given by the unique stochastic process $(X_t)_{0 \leq t \leq 1}$ satisfying eq. 3.5.

如果 $u : \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}^d$ 连续可微且有界导数, 且 σ_t 连续, 那么方程 (7) 中的 SDE 存在解, 且由满足方程 3.5 的唯一随机过程 $(X_t)_{0 \leq t \leq 1}$ 给出.

3.2.1 模拟 SDE/数值解

寻求 SDE 数值解最简单的方法是欧拉-丸山法 (Euler-Maruyama method), 其重要性等同于 ODE 的欧拉法. 使用欧拉-丸山法, 从 $X_0 = x_0$ 出发, 逐步迭代:

$$X_{t+h} = X_t + hu_t(X_t) + \sqrt{h}\sigma_t\epsilon_t, \quad \epsilon_t \sim N(0, I_d) \tag{3.6}$$

其中 $h = n^{-1} > 0$ ($n \in \mathbb{E}$) 是一个超参数. 换句话说, 使用欧拉-丸山法进行模拟时, 我们会在 $u_t(X_t)$ 的方向上迈出一小步, 并且加上一些按 $\sqrt{h}\sigma_t$ 缩放的高斯噪声.

3.2.2 Diffusion Models

与 ODE 相同, 我们可以通过 SDE 构造一个生成式模型. 我们的目标依然是将简单分布 p_{init} 转化为复杂分布 p_{data} . 自然地, 用 $X_0 \sim p_{\text{init}}$ 随机初始化的 SDE, 然后模拟这种转换. 为了参数化这个 SDE, 我们只需简单地参数化其核心成分——一个神经网络向量场 u_t^θ . 因此, 一个扩散模型可以表示为:

$$dX_t = u_t(X_t)dt + \sigma_t dW_t, \tag{3.7a} \text{SDE}$$

$$X_0 \sim p_{\text{init}}, \tag{3.7b} \text{random conditions}$$

算法 3 描述了使用欧拉-丸山法采样 (sample) 扩散模型的方法.

Algorithm 3 Sampling from a Diffusion Model (Euler-Maruyama method)

Require: Neural network vector field u_t^θ , number of steps n , diffusion coefficient σ_t

- 1: Set $t = 0$
 - 2: Set step size $h = \frac{1}{n}$
 - 3: Draw a sample $X_0 \sim p_{\text{init}}$
 - 4: **for** $i = 1, \dots, n - 1$ **do**
 - 5: Draw a sample $\epsilon \sim N(0, I_d)$
 - 6: $X_{t+h} = X_t + hu_t^\theta(X_t) + \sqrt{h}\sigma_t\epsilon$
 - 7: Update $t \leftarrow t + h$
 - 8: **end for**
 - 9: **return** X_1
-

³Trajectory(轨迹): 数学表示为 $t \mapsto X_t$, 是单个初始条件 X_0 随时间 t 演化的具体路径 (时间函数); Flow Map(流映射) 数学表示为 $\phi_t : X_0 \mapsto X_t$ 所有初始条件 X_0 在时间 t 后的全局映射规则 (函数族).

3.3 构造训练目标: Conditional and Marginal Score Functions

在第 1 章中, 我们介绍了 Flow Model 的训练目标. 现在, 我们要构造 Diffusion Model 的训练目标. 不过在此之前, 我们先来定义 p_t 的 **边际匹配得分 (marginal score function)** 为 $\nabla \log p_t(x)$.

定理 3.2 (SDE 扩展技巧)

条件向量场 $u_t^{\text{target}}(x|z)$ 和边际向量场 $u_t^{\text{target}}(x)$ 的定义如公式 1.17. 那么, 对于扩散系数 $\sigma_t \geq 0$, 我们可以构造一个 SDE, 使其遵循相同的概率路径:

$$X_0 \sim p_{\text{init}}, \quad dX_t = \left(u_t^{\text{target}}(X_t) + \frac{\sigma_t^2}{2} \nabla \log p_t(X_t) \right) dt + \sigma_t dW_t \quad (3.8)$$

$$\Rightarrow X_t \sim p_t \quad (0 \leq t \leq 1) \quad (3.9)$$

特别地, $X_1 \sim p_{\text{data}}$ 对于这个 SDE. 如果我们将边际概率 $p_t(x)$ 和向量场 $u_t^{\text{target}}(x)$ 替换为条件概率路径 $p_t(x|z)$ 和向量场 $u_t^{\text{target}}(x|z)$, 等式也同样成立.

公式 3.8 很有用, 因为我们可以将边际匹配得分通过条件匹配得分 (**conditional score function**) $\nabla \log p_t(x|z)$ 得到:

$$\nabla \log p_t(x) = \frac{\nabla p_t(x)}{p_t(x)} = \frac{\nabla \int p_t(x|z) p_{\text{data}}(z) dz}{p_t(x)} = \frac{\int \nabla p_t(x|z) p_{\text{data}}(z) dz}{p_t(x)} = \int \nabla \log p_t(x|z) \frac{p_t(x|z) p_{\text{data}}(z)}{p_t(x)} dz \quad (3.10)$$

大多数条件下, 条件匹配得分 $\nabla \log p_t(x|z)$ 的解析式是已知的.

示例 3.1. 高斯概率路径的得分函数 (Score Function)

对于高斯概率路径 $p_t(x|z) = \mathcal{N}(x; \alpha_t z, \beta_t^2 I_d)$, 我们可以利用高斯概率密度函数的形式 (高斯分布的函数) 得到:

$$\nabla \log p_t(x|z) = \nabla \log \mathcal{N}(x; \alpha_t z, \beta_t^2 I_d) = -\frac{x - \alpha_t z}{\beta_t^2}. \quad (3.11)$$

注意到, 匹配得分是 x 的线性函数, 这是高斯分布的性质.

定理 3.3 (Fokker-Planck 方程)

设 p_t 是一个概率路径, 考虑随机微分方程 (SDE):

$$X_0 \sim p_{\text{init}}, \quad dX_t = u_t(X_t) dt + \sigma_t dW_t. \quad (3.12)$$

则 $X_t (0 \leq t \leq 1)$ 的分布为 p_t , 当且仅当其满足 **Fokker-Planck** 方程:

$$\partial_t p_t(x) = -\text{div}(p_t u_t)(x) + \frac{\sigma_t^2}{2} \Delta p_t(x), \quad \forall x \in \mathbb{R}^d, 0 \leq t \leq 1, \quad (3.13)$$

其中:

- $\partial_t p_t(x) = \frac{d}{dt} p_t(x)$ 是概率密度的随时间变化率;
- $\text{div}(p_t u_t)$ 是概率流 $p_t u_t$ 的散度 (描述确定性漂移项的影响);
- Δp_t 是拉普拉斯算子^a作用到概率密度 p_t 上 (描述扩散项的效应).

^a $\Delta = \nabla \cdot \nabla$

3.4 定理 3.2 解释

为什么在定理 3.2 中我们要构造公式 3.8?

我们希望通过调整 SDE 的漂移项 (drift) 和扩散项 (diffusion), 使其概率演化与 ODE 保持一致. 具体来说, SDE 的漂移项在 ODE 的向量场基础上增加了一个修正项 $\frac{\sigma_t^2}{2} \nabla \log p_t(x)$, 以抵消噪声引入的分布偏移.

3.4.1 ODE 与 SDE 的概率演化对比

ODE 的演化 (流模型) ODE 的形式为:

$$\frac{dX_t}{dt} = u_t^{\text{target}}(X_t), \quad (3.14)$$

其概率密度 $p_t(x)$ 的演化由连续性方程 (Continuity Equation) 描述:

$$\partial_t p_t(x) = -\text{div}(p_t u_t^{\text{target}})(x). \quad (3.15)$$

这个方程表示概率密度的变化完全由向量场 u_t^{target} 的流动决定.

SDE 的演化 (扩散模型) SDE 的形式为:

$$dX_t = u_t^{\text{target}}(X_t)dt + \sigma_t dW_t, \quad (3.16)$$

其概率密度 $p_t(x)$ 的演化由 Fokker-Planck 方程描述:

$$\partial_t p_t(x) = -\text{div}(p_t u_t^{\text{target}})(x) + \frac{\sigma_t^2}{2} \Delta p_t(x). \quad (3.17)$$

这里多了一项 $\frac{\sigma_t^2}{2} \Delta p_t(x)$, 表示噪声对概率密度的扩散效应.

3.4.2 如何让 SDE 的概率演化与 ODE 一致?

为了让 SDE 的 $p_t(x)$ 与 ODE 的 $p_t(x)$ 相同, 我们需要抵消扩散项的影响. 具体方法是在 SDE 的漂移项中增加一个修正项:

$$dX_t = \left(u_t^{\text{target}}(X_t) + \frac{\sigma_t^2}{2} \nabla \log p_t(X_t) \right) dt + \sigma_t dW_t. \quad (3.18)$$

修正项的数学作用

扩散项 $\frac{\sigma_t^2}{2} \Delta p_t(x)$ 会使得概率密度扩散 (变得更平坦). 修正项 $\frac{\sigma_t^2}{2} \nabla \log p_t(x)$ 的加入会生成一个反向的漂移力, 将扩散的概率质量拉回原始分布.

通过计算修正后的 Fokker-Planck 方程:

$$\partial_t p_t(x) = -\text{div} \left(p_t \left(u_t^{\text{target}} + \frac{\sigma_t^2}{2} \nabla \log p_t \right) \right) + \frac{\sigma_t^2}{2} \Delta p_t, \quad (3.19)$$

展开后可以发现:

$$\partial_t p_t(x) = -\text{div}(p_t u_t^{\text{target}}) - \frac{\sigma_t^2}{2} \text{div}(p_t \nabla \log p_t) + \frac{\sigma_t^2}{2} \Delta p_t. \quad (3.20)$$

由于 $\text{div}(p_t \nabla \log p_t) = \Delta p_t$ (因为 $\nabla \log p_t = \frac{\nabla p_t}{p_t}$), 两项相消, 最终得到:

$$\partial_t p_t(x) = -\text{div}(p_t u_t^{\text{target}}), \quad (3.21)$$

这与 ODE 的连续性方程完全一致! 因此, 修正后的 SDE 和原始 ODE 具有**相同的概率演化**.

3.4.3 直观理解

ODE: 确定性流动, 概率质量沿着向量场 u_t^{target} 移动.

SDE: 加入噪声后, 概率质量会扩散 (变得更分散). 为了保持与 ODE 相同的分布, 需要增加一个得分函数驱动的漂移 $\frac{\sigma_t^2}{2} \nabla \log p_t$, 将扩散的部分拉回来.

这类似于朗之万动力学 (Langevin Dynamics) 中的修正力, 确保系统收敛到目标分布.

3.4.4 为什么得分函数 $\nabla \log p_t$ 能修正噪声?

得分函数 $\nabla \log p_t(x)$ 指向概率密度增长最快的方向, 因此:

- 在低密度区域 ($p_t(x)$ 小), $\nabla \log p_t(x)$ 较大, 修正力更强, 防止概率质量扩散到这些区域.
- 在高密度区域 ($p_t(x)$ 大), 修正力较弱, 允许概率质量保持在主流分布附近.

这样,SDE 的轨迹会围绕 ODE 的确定性路径波动,但整体分布保持一致.

3.4.5 修正后的 SDE 与原始 ODE 具有相同概率演化的意义与优势

如果修正后的 SDE(加入得分函数调整项后)和原始 ODE 具有完全相同的概率演化 $p_t(x)$, 这意味着:

- 生成样本分布相同:ODE 和 SDE 的终点 X_1 均服从目标分布 p_{data}
- 路径性质不同:
 - ODE: 确定性轨迹 ($X_0 \rightarrow X_1$ 唯一)
 - SDE: 随机性轨迹 (X_1 随噪声 dW_t 变化)

3.5 训练: 匹配得分 (Score Matching)

若使得 ODE 与 SDE 有相同的边际概率分布 (概率演化均为 $p_t(x)$), 需有:

$$dX_t = \left(u_t^{\text{target}}(X_t) + \frac{\sigma_t^2}{2} \nabla \log p_t(X_t) \right) dt + \sigma_t dW_t \quad (\text{SDE 形式}) \quad (3.22)$$

$$X_0 \sim p_{\text{init}} \quad (\text{初始条件}) \quad (3.23)$$

$$\Rightarrow X_t \sim p_t \quad \text{对所有 } 0 \leq t \leq 1 \quad (\text{概率路径匹配}) \quad (3.24)$$

其中 u_t^{target} 是边际向量场, $\nabla \log p_t$ 是边际得分函数 (marginal score function), $\nabla \log p_t$ 通过下式得到:

$$\nabla \log p_t(x) = \int \nabla \log p_t(x|z) \underbrace{\frac{p_t(x|z)p_{\text{data}}(z)}{p_t(x)}}_{\text{后验分布 } p_t(z|x)} dz \quad (3.25)$$

为估计边际得分函数 $\nabla \log p_t$, 我们可以使用一个得分网络 (score network) 的神经网络 $s_t^\theta: \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}^d$. 类似于之前的方法, 我们可以设计两种损失函数, **score matching loss** 和 **conditional score matching loss**:

$$L_{\text{SM}}(\theta) = \mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, x \sim p_t(\cdot|z)} [\|s_t^\theta(x) - \nabla \log p_t(x)\|^2] \quad (\text{score matching loss}) \quad (3.26)$$

$$L_{\text{CSM}}(\theta) = \mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, x \sim p_t(\cdot|z)} [\|s_t^\theta(x) - \nabla \log p_t(x|z)\|^2] \quad (\text{conditional score matching loss}) \quad (3.27)$$

关键区别在于:

- 边际匹配得分使用 $\nabla \log p_t(x)$ (全局边际分布得分)
- 条件匹配得分使用 $\nabla \log p_t(x|z)$ (已知数据点 z 的条件得分)

虽然理想情况下我们希望最小化 L_{SM} (边际得分匹配), 但由于 $\nabla \log p_t(x)$ 未知, 实际上采用可处理的替代方案:

定理 3.4 (条件得分匹配的等价性)

Score matching loss 和 Conditional score matching loss 之间之差一个常数:

$$\mathcal{L}_{\text{SM}}(\theta) = \mathcal{L}_{\text{CSM}}(\theta) + C, \quad (3.28)$$

其中 C 与参数 θ 无关. 因此, 他们的梯度相等:

$$\nabla \mathcal{L}_{\text{SM}}(\theta) = \nabla \mathcal{L}_{\text{CSM}}(\theta) \quad (3.29)$$

特别地, 对于最小化的参数 θ^* , 我们有 $\theta^* = \nabla \log p_t$

证明 参考 1.3, 略.

上述过程描述了扩散模型的标准训练流程. 训练完成后, 我们可以选择任意的扩散系数 $\sigma_t \geq 0$ 来模拟随机微分方程 (SDE):

$$X_0 \sim p_{\text{init}}, \quad dX_t = \left(u_t^\theta(X_t) + \frac{\sigma_t^2}{2} s_t^\theta(X_t) \right) dt + \sigma_t dW_t \quad (3.30)$$

从而生成样本 $X_1 \sim p_{\text{data}}$. 理论上, 当训练完美时, 任意 σ_t 都应产生符合目标分布的样本. 但实际上会遇到两类误差:

1. 数值误差: SDE 数值模拟的不精确性
 - 离散化误差 (如 Euler-Maruyama 方法的 $O(\Delta t^{1/2})$ 阶误差)
 - 累积误差随模拟步数增加而放大
2. 训练误差: 得分网络 s_t^θ 的近似不完美

乍一看, 似乎使用扩散模型而不是流模型的一个缺点是我们需要同时学习 s_t^θ 和 u_t^θ . 然而, 需要注意的是, 我们通常可以在一个有双输出的单一网络中直接学习 s_t^θ 和 u_t^θ , 因此额外的计算工作量通常非常小. 此外, 在高斯概率路径的特殊情况下, s_t^θ 和 u_t^θ 可以相互转换, 因此我们不需要分别训练它们.

注 Denoising Diffusion Models

如果你对扩散模型有所了解, 你可能已经听说过去噪扩散模型 (denoising diffusion model) 这个术语. 这类模型变得如此流行, 以至于现在大多数人干脆省略了“去噪 (denoising)”这个词, 直接用“扩散模型 (diffusion model)”来称呼它. 在本文档的语境中, 这些就是具有高斯概率路径 $p_t(\cdot|z) = \mathcal{N}(\alpha_t z, \beta_t^2 I)$ 的扩散模型. 然而, 需要注意的是, 在一些最初的扩散模型论文中, 这可能并不那么显而易见. 这些论文使用了不同的时间约定 (时间是倒置的), 因此你需要应用适当的时间重新缩放. 此外, 它们通过所谓的前向过程 (forward process) 来构建概率路径.

示例 3.2. Denoising Diffusion Models: Score Matching for Gaussian Probability Paths

首先, 我们实例化去噪分数匹配损失函数在 $p_t(x|z) = \mathcal{N}(\alpha_t z, \beta_t^2 I_d)$ 情况下的表达式. 如公式 3.11 所推导, 条件分数 $\nabla \log p_t(x|z)$ 的解析式为:

$$\nabla \log p_t(x|z) = -\frac{x - \alpha_t z}{\beta_t^2}. \tag{3.31}$$

将其代入 conditional score matching loss 可得:

$$L_{\text{CSM}}(\theta) = \mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, x \sim p_t(\cdot|z)} \left[\left\| s_t^\theta(x) + \frac{x - \alpha_t z}{\beta_t^2} \right\|^2 \right] \tag{3.32}$$

$$\stackrel{(i)}{=} \mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, \epsilon \sim \mathcal{N}(0, I_d)} \left[\left\| s_t^\theta(\alpha_t z + \beta_t \epsilon) + \frac{\epsilon}{\beta_t} \right\|^2 \right] \tag{3.33}$$

$$= \mathbb{E}_{t \sim \text{Unif}, z \sim p_{\text{data}}, \epsilon \sim \mathcal{N}(0, I_d)} \left[\frac{1}{\beta_t^2} \left\| \beta_t s_t^\theta(\alpha_t z + \beta_t \epsilon) + \epsilon \right\|^2 \right] \tag{3.34}$$

其中步骤 (i) 通过公式 1.34 将 x 替换为 $\alpha_t z + \beta_t \epsilon$. 值得注意的是, 网络 s_t^θ 本质上学习的是用于破坏数据样本 z 的噪声预测. 因此, 上述训练损失也被称为去噪匹配得分 (denoising score matching), 这是学习扩散模型最早采用的方法之一.

研究发现当 $\beta_t \approx 0$ 接近零时 (即仅在添加足够噪声时有效), 上述损失函数存在数值不稳定性. 在早期去噪扩散概率模型 (如 DDPM) 中, 提出通过舍弃常数 $\frac{1}{\beta_t^2}$ 并将 s_t^θ 重参数化为噪声预测 (noise predictor) 网络 $\epsilon_t^\theta: \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}^d$:

$$-\beta_t s_t^\theta(x) = \epsilon_t^\theta(x) \implies L_{\text{DDPM}}(\theta) = \mathbb{E}_{t, z, \epsilon} \left[\left\| \epsilon_t^\theta(\alpha_t z + \beta_t \epsilon) - \epsilon \right\|^2 \right] \tag{3.35}$$

如上所述, 网络 ϵ_t^θ 本质上是在学习预测用来破坏数据样本 z 的噪声. 算法 4 总结了该训练流程.

除了其简洁性外, 高斯概率路径还具有另一个实用特性: 通过学习得分函数 s_t^θ 或噪声预测器 ϵ_t^θ , 我们可以自动获得对应的向量场 u_t^θ , 反之亦然:

Algorithm 4 Score Matching Training Procedure for Gaussian probability path

Require: A dataset of samples $z \sim p_{\text{data}}$, score network s_t^θ or noise predictor ϵ_t^θ

- 1: **for** each mini-batch of data **do**
- 2: Sample a data example z from the dataset.
- 3: Sample a random time $t \sim \text{Unif}[0, 1]$.
- 4: Sample noise $\epsilon \sim \mathcal{N}(0, I_d)$.
- 5: Set $x_t = \alpha_t z + \beta_t \epsilon$ (General case: $x_t \sim p_t(\cdot|z)$).
- 6: Compute loss

$$\mathcal{L}(\theta) = \left\| s_t^\theta(x_t) + \frac{\epsilon}{\beta_t} \right\|^2 \quad (\text{General case: } = \left\| s_t^\theta(x_t) - \nabla \log p_t(x_t|z) \right\|^2)$$

Alternatively : $\mathcal{L}(\theta) = \left\| \epsilon_t^\theta(x_t) - \epsilon \right\|^2$

- 7: Update the model parameters θ via gradient descent on $\mathcal{L}(\theta)$.
- 8: **end for**

命题 3.1 (得分函数与向量场的双向转换 (Conversion formula for Gaussian probability path))

对于任意高斯概率路径 $p_t(x|z) = \mathcal{N}(\alpha_t z, \beta_t^2 I_d)$, 其条件/边际得分与向量场满足以下等价关系:

$$u_t^{\text{target}}(x|z) = \left(\beta_t^2 \frac{\dot{\alpha}_t}{\alpha_t} - \dot{\beta}_t \beta_t \right) \nabla \log p_t(x|z) + \frac{\dot{\alpha}_t}{\alpha_t} x \quad (3.36)$$

$$u_t^{\text{target}}(x) = \left(\beta_t^2 \frac{\dot{\alpha}_t}{\alpha_t} - \dot{\beta}_t \beta_t \right) \nabla \log p_t(x) + \frac{\dot{\alpha}_t}{\alpha_t} x \quad (3.37)$$

其中 $\dot{\alpha}_t, \dot{\beta}_t$ 表示时间导数. 其中, 上述边际向量场 u_t^{target} 的公式在文献中被称为 **概率流 ODE (probability flow ODE)** (更准确地说, 是指其对应的常微分方程). 该性质表明二者可通过线性变换相互导出.

证明 对于条件向量场和条件分数函数, 我们可以推导:

$$u_t^{\text{target}}(x|z) = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t} \alpha_t \right) z + \frac{\dot{\beta}_t}{\beta_t} x \quad (3.38)$$

$$\stackrel{(i)}{=} \left(\beta_t^2 \frac{\dot{\alpha}_t}{\alpha_t} - \dot{\beta}_t \beta_t \right) \left(\frac{\alpha_t z - x}{\beta_t^2} \right) + \frac{\dot{\alpha}_t}{\alpha_t} x \quad (3.39)$$

$$= \left(\beta_t^2 \frac{\dot{\alpha}_t}{\alpha_t} - \dot{\beta}_t \beta_t \right) \nabla \log p_t(x|z) + \frac{\dot{\alpha}_t}{\alpha_t} x \quad (3.40)$$

其中步骤 (i) 通过代数运算完成. 通过积分运算, 相同恒等式对边际流向量场和边际分数函数同样成立:

$$u_t^{\text{target}}(x) = \int u_t^{\text{target}}(x|z) \frac{p_t(x|z) p_{\text{data}}(z)}{p_t(x)} dz \quad (3.41)$$

$$= \int \left[\left(\beta_t^2 \frac{\dot{\alpha}_t}{\alpha_t} - \dot{\beta}_t \beta_t \right) \nabla \log p_t(x|z) + \frac{\dot{\alpha}_t}{\alpha_t} x \right] \frac{p_t(x|z) p_{\text{data}}(z)}{p_t(x)} dz \quad (3.42)$$

$$\stackrel{(i)}{=} \left(\beta_t^2 \frac{\dot{\alpha}_t}{\alpha_t} - \dot{\beta}_t \beta_t \right) \nabla \log p_t(x) + \frac{\dot{\alpha}_t}{\alpha_t} x \quad (3.43)$$

其中步骤 (i) 应用了公式 3.25 的分数函数边际化性质.

我们可以利用转换公式, 通过以下方式在得分网络 s_t^θ 和向量场网络 u_t^θ 之间进行参数化转换:

$$u_t^\theta = \left(\beta_t^2 \frac{\dot{\alpha}_t}{\alpha_t} - \dot{\beta}_t \beta_t \right) s_t^\theta(x) + \frac{\dot{\alpha}_t}{\alpha_t} x \quad (3.44)$$

类似地, 只要满足 $\beta_t^2 \dot{\alpha}_t - \alpha_t \dot{\beta}_t \beta_t \neq 0$ (对于 $t \in [0, 1]$ 恒成立), 则有:

$$s_t^\theta(x) = \frac{\alpha_t u_t^\theta(x) - \dot{\alpha}_t x}{\beta_t^2 \dot{\alpha}_t - \alpha_t \dot{\beta}_t \beta_t} \quad (3.45)$$

通过这种参数化方式可以证明,denoising score matching loss 与 conditional flow matching losses 仅相差一个常数项. 由此我们得出结论: 对于高斯概率路径, 无需单独训练边际分数 (marginal score) 和边际向量场 (marginal vector field), 因为已知其中一个就足以计算另一个. 特别地, 我们可以自由选择使用流匹配或匹配得分进行训练.

若已训练好得分网络 s_t^θ , 根据公式 3.30 可采用任意 $\sigma_t \geq 0$ 从以下 SDE 采样:

$$X_0 \sim p_{\text{init}}, \quad dX_t = \left[\left(\beta_t^2 \frac{\dot{\alpha}_t}{\alpha_t} - \dot{\beta}_t \beta_t + \frac{\sigma_t^2}{2} \right) s_t^\theta(x) + \frac{\dot{\alpha}_t}{\alpha_t} x \right] dt + \sigma_t dW_t \quad (3.46)$$

从而获得服从 p_{data} 分布的样本 X_1 (在训练和模拟误差范围内). 这对应于去噪扩散模型的随机采样过程 (stochastic sampling from a denoising diffusion model).

3.6 时间反演 vs 求解 FP 方程

(Time-Reversals vs Solving the Fokker-Planck equation)

扩散模型的原始描述并未通过 Fokker-Planck 方程 (或连续性方程) 构造训练目标 u_t^θ 或 $\nabla \log p_t$, 而是通过前向过程的时间反演实现. 时间反演 $(X_t)_{0 \leq t \leq T}$ 是一个 SDE, 其轨迹分布与原始过程相同但时间倒置, 即:

$$\mathbb{P}[\bar{X}_{t_1} \in A_1, \dots, \bar{X}_{t_n} \in A_n] = \mathbb{P}[\bar{X}_{T-t_1} \in A_1, \dots, \bar{X}_{T-t_n} \in A_n] \quad (3.47)$$

$$\forall 0 \leq t_1, \dots, t_n \leq T, \text{ and } A_1, \dots, A_n \subset \mathcal{S} \quad (3.48)$$

根据 Anderson 的研究, 满足上述条件的时间反演可以通过以下 SDE 实现:

$$dX_t = \left(-u_t(X_t) + \sigma_t^2 \nabla \log p_t(X_t) \right) dt + \sigma_t dW_t, \quad u_t(x) = u_{T-t}^{\text{forw}}(x), \quad \sigma_t = \bar{\sigma}_{T-t} \quad (3.49)$$

若 $u_t(X_t) = a_t X_t$, 上述 SDE 对应命题 3.1 中训练目标的一个特例 (这一点并不显然, 因为使用了不同的时间约定). 然而, 对于生成建模的许多应用, 我们通常只使用马尔可夫过程的终点 X_1 (例如生成的图像), 而丢弃中间时间点. 因此, 马尔可夫过程是否是“真正”的时间反演, 或者是否沿着概率路径演化, 对许多应用并不重要. 即使用时间反演并非必要, 且通常会导导致次优结果 (例如概率流 ODE 通常效果更好).

3.6.1 时间反演

扩散模型的逆向过程并非直接反转原 SDE 的路径, 而是构造一个新的 SDE, 使得其生成的边缘分布 $p_t(x_t)$ 与原过程的边缘分布反向对齐, 这与伊藤过程不可逆不矛盾. 这一构造依赖于以下关键理论: 前向扩散通过 SDE 逐步加噪:

$$dx_t = u(x_t, t)dt + \sigma_t dw_t \quad (3.50)$$

对应的 Fokker-Planck 方程 (描述概率密度 $p_t(x)$ 的演化):

$$\frac{\partial p_t(x)}{\partial t} = -\nabla_x \cdot [u(x, t)p_t(x)] + \frac{\sigma_t^2}{2} \nabla_x^2 p_t(x) \quad (3.51)$$

我们希望找到一个逆向过程 $dx_t = \bar{u}(x_t, t)dt + \sigma_t d\bar{w}_t$, 使得: 从噪声分布 $p_T(x)$ 出发, 逆向演化回数据分布 $p_0(x)$; 逆向过程的 Fokker-Planck 方程应与前向过程“对称”.

逆向过程的时间是反向的 ($t \rightarrow T - t$), 其 Fokker-Planck 方程为:

$$\frac{\partial p_t(x)}{\partial t} = -\nabla_x \cdot [\bar{u}(x, t)p_t(x)] - \frac{\sigma_t^2}{2} \nabla_x^2 p_t(x) \quad (3.52)$$

(注意扩散项前的负号, 因为时间方向反转.)

前向 Fokker-Planck 方程可以重写为概率流形式:

$$\frac{\partial p_t(x)}{\partial t} = -\nabla_x \cdot \left(\underbrace{\left(u(x, t) - \frac{\sigma_t^2}{2} \nabla_x \log p_t(x) \right)}_{\text{概率流速度}} p_t(x) \right) \quad (3.53)$$

概率流速度 $v(x, t) = u(x, t) - \frac{\sigma_t^2}{2} \nabla_x \log p_t(x)$ 描述了概率密度的整体运动趋势.

为了使逆向过程“回放”前向过程,我们需要:

- 反转漂移方向: 概率流速度反向, 即 $\tilde{v}(x, t) = -v(x, t)$;
- 保持扩散强度: 噪声项 σ_t 不变。

因此, 逆向过程的漂移项为:

$$\tilde{f}(x, t) = \tilde{v}(x, t) + \frac{\sigma_t^2}{2} \nabla_x \log p_t(x) = -u(x, t) + \sigma_t^2 \nabla_x \log p_t(x) \quad (3.54)$$

最终得到逆向 SDE:

$$dx_t = [u(x_t, t) - \sigma_t^2 \nabla_x \log p_t(x_t)] dt + \sigma_t d\tilde{w}_t \quad (3.55)$$

3.6.2 SDE 的 ODE 概率流 (SDE 的 ODE 概率演化)

在 Song 的原文中, 给出了与 SDE 方程 3.50 演化路径相同的 ODE 方程:

$$dx = \left[u_t(x) - \frac{\sigma_t^2}{2} \nabla_x \log p_t(x) \right] dt \quad (3.56)$$

公式 3.56 的概率流为:

$$\partial_t p_t(x) = -\nabla \left[p_t \left(u_t(x) - \frac{\sigma_t^2}{2} \nabla_x \log p_t(x) \right) \right] (x) \quad (3.57)$$

$$= -\nabla (p_t u_t)(x) + \frac{\sigma_t^2}{2} \Delta_x \log p_t(x) \quad (3.58)$$

这就是 SDE 的概率流方程 (Fokker-Planck 方程).

利用公式 3.56 即可在 ODE 中使用 SDE 的概率演化 (联想 DDIM).

3.7 小结

本章主要讲述 SDE 的基本原理, 并利用 Fokker-Planck 方程构造出一个 SDE, 使其拥有与 ODE (Flow Matching) 相同的概率演化, 形式上与 ODE (Flow Matching) 保持一致, 而非利用 Song 所提出的时间反演方法. 同时, 在本文语境中, DDPM 只是 SDE 的一种高斯概率路径演化的具体情况.



第 4 章 ❖ 冷冻电镜学习笔记

October 21, 2025

❖ 4.1 TEM(Transmission Electron Microscope) 简介

4.1.1 核心物理原理

- 电子与物质的相互作用
 - 弹性散射: 无能量损失
 - 是衍射衬度成像和电子衍射的基础
 - 用于分析晶体结构、取向和缺陷
 - 非弹性散射: 有能量损失
 - 是分析电子显微学 (如 EELS, XEDS) 的基础
 - 用于分析元素组成、化学价态和电子结构
- 波动性
 - 电子的波动性 (德布罗意波长) 是衍射和高分辨相位衬度成像的物理基础
- 关键原理
 - 惠更斯-菲涅耳原理: 解释了衍射现象, 即波前上每一点都是次级子波源, 它们相干叠加形成新的波前
 - 夫琅禾费衍射: 远场衍射, 是理解 TEM 中衍射花样的模型

4.1.2 仪器构成与功能

- 电子源: 产生照明电子束.
 - 热离子源 (如 LaB_6): 亮度较高, 性价比高, 适用于常规 TEM.
 - 场发射源 (FEG): 亮度极高、相干性好, 是高分辨成像和微区分析的首选.
- 电磁透镜: 聚焦和放大电子束. 其不完美性 (像差) 是限制分辨率的主要因素.
 - 球差 (C_s): 限制分辨率的关键像差, 可通过球差校正器克服.
 - 色差 (C_c): 由电子能量分散引起, 对厚样品影响显著, 可通过能量过滤或单色器改善.
 - 像散: 导致图像定向模糊, 可通过消像散器校正.

- **真空系统:** 为电子束提供无干扰的路径, 并保护电子源.
- **样品杆:** 在真空中承载和操控样品, 实现平移、倾斜、加热、冷却等多种功能.
- **探测系统:**
 - 荧光屏/CCD 相机: 用于接收图像和衍射花样.
 - X 射线能谱仪 (XEDS/EDS): 用于元素成分分析.
 - 电子能量损失谱仪 (EELS): 用于轻元素分析、化学态和电子结构研究.

4.1.3 主要操作模式与信息获取

- **成像模式:** 观察样品的实空间形貌和结构
 - **明场像 (BF):** 使用直射束成像, 质量/厚度大的区域显暗
 - **暗场像 (DF):** 使用衍射束成像, 用于凸显特定取向的晶体或相
 - **高分辨像 (HRTEM):** 利用透射束与衍射束的干涉, 直接观察晶格条纹甚至原子柱
 - 其解释依赖于衬度传递函数 (CTF) 理论
 - 最佳条件是 **Scherzer 欠焦**
- **衍射模式:** 获取样品的倒易空间信息, 用于晶体结构分析
 - **选区电子衍射 (SAED):** 分析样品特定微区的晶体结构
 - **会聚束电子衍射 (CBED):** 提供更丰富的晶体学信息, 如厚度、应变等
- **扫描透射模式 (STEM)**
 - 用聚焦的细电子束在样品上扫描, 同步收集各种信号
 - **高角环形暗场像 (HAADF):** 衬度近似与原子序数的平方 (Z^2) 成正比, 又称“Z 衬度像”
- **分析模式**
 - **XEDS:** 适用于重元素的成分分析
 - **EELS:** 适用于轻元素, 并提供化学键合信息

4.1.4 样品制备

- **核心要求:** 样品必须非常薄 (通常 $< 100 \text{ nm}$) 以保证电子束透明
- **常用方法**
 - **电解双喷:** 适用于金属和合金
 - **离子减薄:** 适用于陶瓷、半导体、岩石等硬脆材料
 - **超薄切片:** 适用于聚合物、生物等软材料
 - **聚焦离子束 (FIB):** 可进行定点、截面样品制备, 精度极高
- **特殊分支: 冷冻电镜 (Cryo-EM)**
 - **原理:** 将生物样品在液态乙烷中快速冷冻, 使其包埋在玻璃态冰中, 从而在低温下用 TEM 观察其近天然状态的结构
 - **关键技术:** 直接电子探测器、强大的三维重构软件
 - **应用:** 主要用于解析生物大分子 (如蛋白质、病毒) 的高分辨率三维结构, 无需结晶

4.1.5 关键数学与公式

1. 电子波长 (相对论修正):

$$\lambda = \frac{h}{\sqrt{2m_0eV(1 + \frac{eV}{2m_0c^2})}}$$

2. 布拉格定律 (衍射几何基础):

$$n\lambda = 2d \sin \theta$$

3. 散射截面: 描述散射概率, 与原子序数 (Z) 和电子能量 (E) 密切相关 (例如, 卢瑟福截面 $\sigma \propto Z^2/E^2$).

4. 平均自由程:

$$\lambda_{mfp} = 1/\sigma_{total}$$

决定样品理想厚度.

5. 分辨率:

- 衍射极限: $r_{th} \approx 0.61\lambda/\beta$
- 球差限制 (Scherzer 分辨率): $r_{min} \approx 0.66C_s^{1/4} \lambda^{3/4}$

6. 衬度传递函数 (CTF): 是理解 HRTEM 图像的核心.

$$CTF(k) = \sin(\chi(k)), \quad \chi(k) = \frac{\pi}{2}C_s\lambda^3k^4 + \pi\Delta f\lambda k^2$$

7. Scherzer 欠焦:

$$\Delta f_{Sch} = -\sqrt{\frac{4}{3}C_s\lambda}$$

4.2 问题解答

4.2.1 冷冻电镜图像噪声来源

- **非弹性散射电子:** 当高能电子穿过样品时, 会与样品中的原子发生相互作用, 主要产生弹性散射与非弹性散射. 弹性散射电子携带了样品内部结构 (原子位置) 的相位信息, 是构成高分辨率图像有用信号的核心; 非弹性散射电子主要携带关于样品元素组成、化学键 (电子结构) 的信息. 但对于成像来说, 它们不直接携带高分辨率的空间结构信息. 非弹性散射电子由于能量损失和方向随机改变, 会形成一个弥漫的、非相干背景. 这个背景会叠加在由弹性散射形成的清晰信号之上, 降低了图像的信噪比和衬度, 使得微弱的细节 (尤其是高分辨率信息) 更难被探测.
- **CMOS 噪声:** CMOS 探测器的噪声可以按其产生环节分为以下几类:
 - **读出噪声:** 这是当相机将每个像素收集的电荷 (模拟信号) 转换为数字值时引入的随机噪声. 主要来自读出电路中的热噪声和晶体管开关噪声.
 - **暗电流噪声:** 由于热效应, 探测器像素在完全没有光或电子照射的情况下, 也会自发地产生电荷.
 - **散粒噪声:** 虽然散粒噪声的本质是电子数量的量子涨落 (物理噪声), 但它最终体现在探测器接收到的信号上.
 - **像素响应非均匀性:** 制造工艺导致探测器的数百万个像素之间存在微小差异. 有的像素可能对电子更“敏感” (增益高), 有的则更“迟钝” (增益低).
- **样品相关噪声:** 这是由样品本身及其制备过程引入的:
 - **冰层噪声/背景:** 样品被包裹在无定形冰中. 冰层本身并非绝对均匀, 其密度起伏会在图像中产生背景信号, 掩盖微弱的样品信号.
 - **样品异质性:** 样品的结构和状态差异会引起重构模糊, 主要包括:
 1. **构象异质性:** 蛋白质分子处于不同的功能状态 (如“开放”态、“关闭”态).
 2. **化学组成异质性:** 复合物中组分的结合与否、化学修饰的不同等.
 3. **取向异质性:** 颗粒在冰层中以随机取向存在, 但如果某些取向缺失, 会导致重构缺失信息, 产生“伪影”.
 - **污染物和杂质:** 品中存在的盐晶、去垢剂胶束、降解的蛋白片段等, 都会增加图像的复杂性.
- **电子束引起的噪声/损伤:** 高能电子束在轰击样品时, 会破坏化学键, 导致样品结构发生损伤. 这种损伤是累积性的, 随着电子剂量的增加而加剧.

4.2.2 CTF(Contrast Transfer Function), 对比度传递函数

在理想的光学透镜中, 一个物点经过透镜后会完美地汇聚成一个像点. 但在现实中, 由于像差的存在, 物点会变成模糊的斑 (艾里斑). 在电子光学中, 这个问题尤为严重.

当一束平行的电子波穿过一个极薄的晶体样品时, 它会与样品中的原子发生相互作用, 发生衍射. 出射的电子波包含两种信息:

- **振幅衬度:** 由于电子被样品吸收或散射到光阑外而造成的强度差异. 这在较厚的样品中占主导.
- **相位衬度:** 对于 HRTEM 观察的极薄样品, 电子波主要通过改变相位来携带信息. 我们的眼睛和探测器无法直接感知相位, 只能感知强度 (振幅的平方).

这些携带了样品结构信息的衍射波和直接透射波, 在穿过 TEM 的电磁透镜系统时, 会因为透镜的不完美性 (像差) 和欠焦操作而发生额外的相位改变. 最终, 它们在像平面上干涉, 形成了我们看到的灰度图像.

CTF 的核心作用就是描述: 一个来自样品的空间频率信息 (可以想象成不同间距的晶格), 在经过整个显微镜系统后, 其相位和振幅是如何被改变的.

CTF 公式:

$$\text{CTF}(k) = \sin(\chi(k)) = \sin\left(\frac{\pi}{2}C_s\lambda^3k^4 + \pi\Delta f\lambda k^2\right) \quad (4.1)$$

这里, $\chi(k)$ 是相位, Δf 是欠焦值 (Defocus), λ 是电子束的在样品中的波长, k 是空间频率 (Spatial Frequency), C_s 是透镜的球差 (Spherical Aberration).

在透射电镜中, 当 Electron Gun 的电压确定了, 那么电子束的波长 λ 也就确定了. Δf 和 C_s 是可以人为控制的参量. 所以, 在 CTF 中, 实际上只有一个自变量 k .

成像系统决定的参数: $C_s, \lambda, \Delta f$

4.2.3 Thon 环

Thon 环 (Thon rings) 是透射电子显微镜 (TEM) 成像中出现的一种周期性明暗交替的条纹图案, 通常出现在欠焦或过焦条件下的离焦像中. 其产生原因与电子波的离焦衬度和衍射效应有关.

物理背景: 电子波的波动性: 在 TEM 中, 电子束穿过样品时, 与样品内部的电势场发生相互作用, 导致电子波的相位发生变化. 当电子波穿过样品后, 其波前会发生畸变, 形成携带样品信息的相位衬度.

离焦引入的衬度: 在 TEM 中, 电子束穿过样品时, 与样品内部的电势场发生相互作用, 导致电子波的相位发生变化. 当电子波穿过样品后, 其波前会发生畸变, 形成携带样品信息的相位衬度. 为了将相位衬度转换为可见的强度衬度, 通常需要让成像平面偏离样品的准确焦平面 (即离焦). 离焦会引入一个相位衬度传递函数 (CTF, 如公式 4.1)

Thon 环的形成机制: $\text{CTF}(\sin(\chi(k)))$ 随空间频率 k 的变化而振荡, 导致某些空间频率 (对应样品的特定结构周期) 被增强, 而另一些被抑制. 在特定的离焦条件下, CTF 的振荡会在傅里叶变换 (或衍射图) 中表现为明暗相间的同心圆环, 即 Thon 环. 环的间距与离焦量 (Δf) 和电子波长 (λ) 相关, 通过分析 Thon 环可以反推离焦量或校正透镜参数. 在收集完图片之后, 我们要对图片进行 CTF 修正, 还原它在各个分辨率的衬度.

Thon 环完整: 能够看到清晰、规整的 Thon 环, 说明了以下几点:

- **仪器状态极佳:** 这表明电镜的相干性非常好 (电子束像一道完美的激光), 机械稳定性高 (没有漂移和振动), 像散被完美校正 (否则环会变成椭圆). 这是获得高分辨图像的前提.
- **样品质量合格:** 样品必须非常薄且均匀, 才能产生强烈的相位衬度干涉, 从而形成清晰的 Thon 环. 如果样品太厚、污染严重或不均匀, 环会变得模糊或不完整.
- **图像信息量足:** 清晰的 Thon 环意味着图像包含了从低频到高频的丰富信息. 环能延伸到的空间频率越高, 说明图像的分辨率潜力越大.

4.2.4 冷冻电镜处理过程

Workflow: 单颗粒冷冻电镜 (cryo-EM SPA) 拍摄 Workflow 如图 4.1:

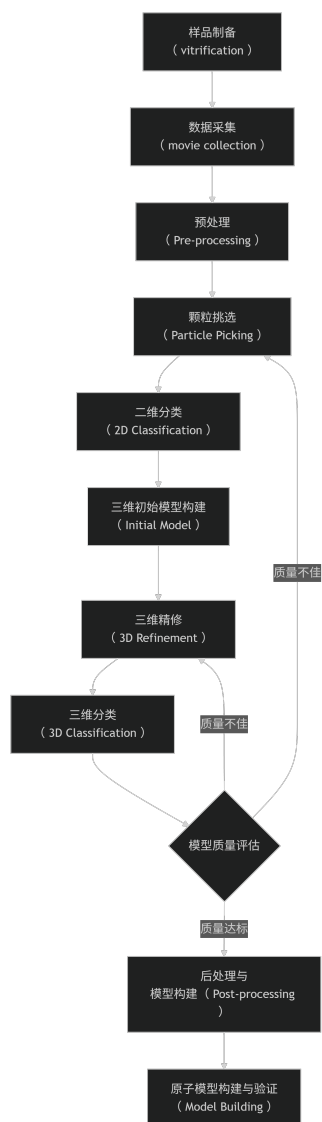


图 4.1: 冷冻电镜 workflow

拍摄角度: 在冷冻电镜中: 研究者将样品快速冷冻成一层无定形冰 (vitrified ice). 在这层冰中, 每个分子 (粒子) 会被随机嵌入、随机取向. 电镜在拍摄时, 对冰层进行垂直照射 (电子束几乎固定方向), 因此:

- 每个粒子的朝向不同;
- 每张 micrograph 中, 其实包含了成千上万个不同角度的投影图像.

在标准的单颗粒冷冻电镜拍摄中, 电子束方向固定不变, 样品角度也不是动态变化的. 我们依靠样品中粒子的随机取向分布来获得各个方向的投影图像, 从而通过计算重建出三维结构.

4.2.5 同质和异质

同质: 指的是样品中所有蛋白质颗粒在结构上都是完全相同的. **定义:** 样品中的所有蛋白质分子都具有完全相同的三维结构、构象和化学组成. **侧重点:** 同质样品数据分析的核心目标是提高信噪比和分辨率.

异质: 指的是样品中存在结构上不同的多种颗粒. **定义:** 样品中的蛋白质颗粒在结构上存在差异. 这种差异可以是:

- 组成型异质性: 样品中本身就存在不同的分子. 例如: 蛋白质与不同配体的结合/未结合状态; 蛋白质复合物中不同亚基的组成比例不同; 样品中混杂了完全不同的蛋白质.
- 构象型异质性: 同一个蛋白质分子可以动态地在不同三维形状之间转换. 例如: 核糖体在翻译过程中的不同

状态;膜通道的“开放”与“关闭”状态;酶在催化循环中的不同中间态.

侧重点: 异质样品数据分析的核心目标是“分离”与“分类”,即从混合的颗粒集合中,识别出不同的结构状态,并分别重构出它们各自的三维结构.

4.2.6 傅里叶切片定理

傅里叶切片定理 (Fourier Slice Theorem), 也称为中心切片定理 (Central Slice Theorem) 或投影切片定理, 是计算机断层成像 (CT) 和医学影像领域中的一个基本数学原理. 该定理建立了图像的投影数据与其傅里叶变换之间的关系, 是理解 CT 图像重建算法的核心. 傅里叶切片定理可以表述为: 一个二维函数 (例如图像) 在某个角度下的投影的一维傅里叶变换, 等于该二维函数的二维傅里叶变换在相同角度下通过原点的切片.

数学表达: 设 $f(x, y)$ 是一个二维函数, 其二维傅里叶变换为:

$$F(u, v) = \iint_{-\infty}^{\infty} f(x, y) e^{-i2\pi(ux+vy)} dx dy \quad (4.2)$$

在角度 θ 下的投影 $p_{\theta}(t)$ 定义为:

$$p_{\theta}(t) = \int_{-\infty}^{\infty} f(t \cos \theta - s \sin \theta, t \sin \theta + s \cos \theta) ds \quad (4.3)$$

其中, t 是投影的坐标轴. 投影 $p_{\theta}(t)$ 的一维傅里叶变换为:

$$P_{\theta}(\omega) = \int_{-\infty}^{\infty} p_{\theta}(t) e^{-i2\pi\omega t} dt \quad (4.4)$$

傅里叶切片定理指出:

$$P_{\theta}(\omega) = F(\omega \cos \theta, \omega \sin \theta) \quad (4.5)$$

即, 投影的一维傅里叶变换等于二维傅里叶变换在频率平面中沿角度 θ 的径向切片. 公式4.5即为傅里叶切片定理

证明 在角度 θ 下的投影 $p_{\theta}(t)$ 定义为沿直线 $L(t, \theta) : x \cos \theta + y \sin \theta = t$ 的线积分:

$$p_{\theta}(t) = \int_{L(t, \theta)} f(x, y) ds \quad (4.6)$$

其中 s 是沿该直线的坐标. 易得:

$$\begin{cases} t = x \cos \theta + y \sin \theta \\ s = -x \sin \theta + y \cos \theta \end{cases} \quad (4.7)$$

注意到公式4.7同时可以表示坐标 (x, y) 旋转 θ 得到坐标系 (t, s) , 则其逆变换为:

$$\begin{cases} x = t \cos \theta - s \sin \theta \\ y = t \sin \theta + s \cos \theta \end{cases} \quad (4.8)$$

该坐标变换的雅可比行列式为:

$$\left| \frac{\partial(x, y)}{\partial(t, s)} \right| = \begin{vmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{vmatrix} = \cos^2 \theta + \sin^2 \theta = 1 \quad (4.9)$$

因此, $dx dy = dt ds$. 于是投影可写为:

$$p_{\theta}(t) = \int_{-\infty}^{\infty} f(t \cos \theta - s \sin \theta, t \sin \theta + s \cos \theta) ds \quad (4.10)$$

投影 $p_{\theta}(t)$ 的一维傅里叶变换为:

$$P_{\theta}(\omega) = \int_{-\infty}^{\infty} p_{\theta}(t) e^{-i2\pi\omega t} dt \quad (4.11)$$

将 $p_\theta(t)$ 的表达式代入:

$$\begin{aligned} P_\theta(\omega) &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(t \cos \theta - s \sin \theta, t \sin \theta + s \cos \theta) ds \right] e^{-i2\pi\omega t} dt \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t \cos \theta - s \sin \theta, t \sin \theta + s \cos \theta) e^{-i2\pi\omega t} dt ds \\ &= \iint_{-\infty}^{\infty} f(x, y) e^{-i2\pi\omega(x \cos \theta + y \sin \theta)} dx dy \end{aligned} \quad (4.12)$$

函数 $f(x, y)$ 的二维傅里叶变换定义为:

$$F(u, v) = \iint_{-\infty}^{\infty} f(x, y) e^{-i2\pi(ux+vy)} dx dy \quad (4.13)$$

现在, 比较 $P_\theta(\omega)$ 的表达式与 $F(u, v)$ 的定义. 如果我们令:

$$u = \omega \cos \theta, \quad v = \omega \sin \theta \quad (4.14)$$

那么:

$$ux + vy = \omega \cos \theta \cdot x + \omega \sin \theta \cdot y = \omega(x \cos \theta + y \sin \theta) \quad (4.15)$$

因此:

$$P_\theta(\omega) = \iint_{-\infty}^{\infty} f(x, y) e^{-i2\pi(ux+vy)} dx dy = F(u, v) \Big|_{u=\omega \cos \theta, v=\omega \sin \theta} \quad (4.16)$$

即:

$$P_\theta(\omega) = F(\omega \cos \theta, \omega \sin \theta) \quad (4.17)$$

证毕.

4.2.7 图像的相位操作与平移

我们有一个离散图像 $f[m, n]$, 尺寸为 $M \times N$ (行 \times 列). 目标: 证明在频域乘以相位因子 $e^{-j2\pi(\frac{dx l}{N} + \frac{dy k}{M})}$ 对应空域的循环平移.

证明 图像的 2D-DFT:

$$F[k, l] = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f[m, n] e^{-j2\pi(\frac{km}{M} + \frac{ln}{N})}, \quad k = 0, 1, \dots, M-1, l = 0, 1, \dots, N-1 \quad (4.18)$$

在频域乘以相位因子:

$$G[k, l] = F[k, l] \cdot e^{-j2\pi(\frac{dx l}{N} + \frac{dy k}{M})} \quad (4.19)$$

其中 d_x, d_y 是平移参数 (整数).

对 $G[k, l]$ 进行逆 DFT:

$$g[m, n] = \frac{1}{MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} G[k, l] e^{j2\pi(\frac{km}{M} + \frac{ln}{N})} \quad (4.20)$$

代入 $G[k, l]$:

$$g[m, n] = \frac{1}{MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} F[k, l] e^{-j2\pi(\frac{dx l}{N} + \frac{dy k}{M})} e^{j2\pi(\frac{km}{M} + \frac{ln}{N})} \quad (4.21)$$

将 $F[k, l]$ 的 DFT 定义代入:

$$g[m, n] = \frac{1}{MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} \left[\sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} f[m', n'] e^{-j2\pi(\frac{km'}{M} + \frac{ln'}{N})} \right] e^{j2\pi(\frac{k(m-dy)}{M} + \frac{l(n-dx)}{N})} \quad (4.22)$$

重新排列求和顺序:

$$g[m, n] = \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} f[m', n'] \left[\frac{1}{MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} e^{j2\pi(\frac{k(m-dy-m')}{M} + \frac{l(n-dx-n')}{N})} \right] \quad (4.23)$$

离散复指数的正交性, 对于行方向:

$$\frac{1}{M} \sum_{k=0}^{M-1} e^{j2\pi \frac{k(m-m')}{M}} = \delta[(m-m') \bmod M] \quad (4.24)$$

对于列方向:

$$\frac{1}{N} \sum_{l=0}^{N-1} e^{j2\pi \frac{l(n-n')}{N}} = \delta[(n-n') \bmod N] \quad (4.25)$$

因此:

$$\frac{1}{MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} e^{j2\pi \left(\frac{k(m-d_y-m')}{M} + \frac{l(n-d_x-n')}{N} \right)} = \delta[(m-d_y-m') \bmod M] \cdot \delta[(n-d_x-n') \bmod N] \quad (4.26)$$

代入正交性关系:

$$g[m, n] = \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} f[m', n'] \cdot \delta[(m-d_y-m') \bmod M] \cdot \delta[(n-d_x-n') \bmod N] \quad (4.27)$$

因此:

$$g[m, n] = f[(m-d_y) \bmod M, (n-d_x) \bmod N] \quad (4.28)$$

最终结果:

$$\boxed{g[m, n] = f[(m-d_y) \bmod M, (n-d_x) \bmod N]} \quad (4.29)$$

这正好是循环平移的定义:

- 垂直方向循环平移 d_y 像素
- 水平方向循环平移 d_x 像素

特殊情况: 当 $d_x \equiv 0 \pmod{N}$ 且 $d_y \equiv 0 \pmod{M}$ 时, 图像保持不变.

证毕.



第 5 章 傅里叶变换解析

February 20, 2026

数学最吸引人的地方,就是它能够揭示那些看似不同事物之间的深刻联系.

—Joseph Fourier

自第一次学习傅里叶变换 (Fourier Transform, FT) 时,我便对教科书上“时域到频域的变换”这一解释感到不满足. 这个看似直观的说法背后,隐藏着三个未被澄清的问题:

1. 什么是时域? 它仅仅是一堆随时间变化的函数吗?
2. 什么是频域? 为什么时域可以通过变换的到频域?
3. 什么是变换? 它真的只是把函数从时间变量改写为频率变量吗?

这些疑问促使我寻找一个更深刻、更统一的数学框架来理解傅里叶变换. 最终,我在 Parseval 框架理论与谱分析理论中找到了答案—傅里叶变换本质上是一种广义的正交基变换,更是自伴算子谱分解的具体实现. 在装备希尔伯特空间 (rigged Hilbert space) $\Phi \subset \mathcal{H} \subset \Phi'$ 的框架下:

1. 时域是由 Dirac δ 函数族 $\{\delta(t - \tau)\}_{\tau \in \mathbb{R}}$ 构成的广义连续 Parseval 框架,对应着位置算子的广义特征函数;
2. 频域是由复指数函数族 $\{e^{i\omega t}\}_{\omega \in \mathbb{R}}$ 构成的广义连续 Parseval 框架,对应着动量算子的广义特征函数;
3. 傅里叶变换则是这两个框架之间的幺正变换 (unitary transformation),它实现了信号在不同“广义基”下的表示转换,本质上是对微分算子进行谱分解的数学工具.

本文旨在从 Parseval 框架与谱分析的视角,为读者重新解读傅里叶变换的数学本质. 我们将看到,傅里叶变换不仅是信号分析的利器,更是理解算子谱理论、量子力学对易关系以及无穷维空间几何结构的窗口. 我们假设读者具备微积分和线性代数的基础知识,文中会自然地引入实分析、泛函分析等概念,但力求直观清晰、深入浅出. 通过这一框架,读者不仅能看到 FT 的“形”,更能理解其背后的“神”—它如何统一地描述信号、算子与空间,如何将复杂的微分运算转化为简单的乘法运算,以及如何揭示物理系统的基本对称性.

5.1 向量空间

本节内容主要参考 Axler 所著书籍 *Linear Algebra Done Right*.

5.1.1 向量空间

首先是对加法和标量乘法的定义:

定义 5.1 (加法, 标量乘法)

- 集合 V 上的加法是指这样一个函数: 对任意 $u, v \in V$, 它都指定一个对应的元素 $u + v \in V$.
- 集合 V 上的标量乘法是指这样一个函数: 对任意 $\lambda \in \mathbb{F}$ 和 $v \in V$, 它都指定一个对应的元素 $\lambda v \in V$.

\mathbb{F} 表示实数域 \mathbb{R} 或复数域 \mathbb{C} , 下同

有了加法和标量乘法的定义之后, 我们可以定义向量空间:

定义 5.2 (向量空间 (vector space))

设 V 是一个集合, 其上定义了加法和标量乘法运算. 如果这些运算满足以下条件, 则称 V 是一个向量空间:

- **加法交换律 (commutativity)**
对任意 $u, v \in V$, 有 $u + v = v + u$;
- **结合律 (associativity)**
对任意 $u, v, w \in V$ 和 $a, b \in \mathbb{F}$, 有 $(u + v) + w = u + (v + w)$ 且 $(ab)v = a(bv)$;
- **零元存在性 (additive identity)**
存在元素 $0 \in V$, 使得对任意 $v \in V$, 有 $v + 0 = v$;
- **负元存在性 (additive inverse)**
对任意 $v \in V$, 都存在 $w \in V$, 使得 $v + w = 0$;
- **单位标量作用 (multiplicative identity)**
对任意 $v \in V$, 有 $1v = v$;
- **分配律 (distributive properties)**
对任意 $a, b \in \mathbb{F}$ 和 $u, v \in V$, 有 $a(u + v) = au + av$ 且 $(a + b)v = av + bv$.

同时, 以下几何语言符合我们的直觉:

定义 5.3 (向量, 点)

向量空间中的元素称为向量, 有时也称为点.

以下给出示例:

示例 5.1. 一个由三角函数构成的向量空间

考虑所有形如 $a \sin x + b \cos x$ 的实值函数组成的集合, 其中 $a, b \in \mathbb{R}$:

$$V = \{f : \mathbb{R} \rightarrow \mathbb{R} \mid f(x) = a \sin x + b \cos x, a, b \in \mathbb{R}\} \quad (5.1)$$

在 V 上定义加法和标量乘法如下:

- **加法:** 对任意 $f, g \in V$, 设 $f(x) = a_1 \sin x + b_1 \cos x, g(x) = a_2 \sin x + b_2 \cos x$, 定义

$$(f + g)(x) = f(x) + g(x) = (a_1 + a_2) \sin x + (b_1 + b_2) \cos x \quad (5.2)$$

- **标量乘法:** 对任意 $\lambda \in \mathbb{R}$ 和 $f \in V$, 设 $f(x) = a \sin x + b \cos x$, 定义

$$(\lambda f)(x) = \lambda f(x) = (\lambda a) \sin x + (\lambda b) \cos x \quad (5.3)$$

验证向量空间各项性质:

1. 封闭性: $f + g$ 和 λf 仍然是 $a \sin x + b \cos x$ 的形式, 所以属于 V .
2. 加法交换律: $(f + g)(x) = f(x) + g(x) = g(x) + f(x) = (g + f)(x)$.

3. 零函数: 取 $a = 0, b = 0$, 得到零函数 $0(x) = 0$, 满足 $f + 0 = f$.
4. 负函数: 对 $f(x) = a \sin x + b \cos x$, 取 $-f(x) = (-a) \sin x + (-b) \cos x$.
5. 标量乘法单位元: $1 \cdot f = f$.
6. 加法结合律: 对 $f, g, h \in V, [(f+g)+h](x) = (f+g)(x) + h(x) = f(x) + g(x) + h(x) = f(x) + (g+h)(x) = [f + (g+h)](x)$
7. 第一分配律 (标量对向量加法的分配): 对 $\lambda \in \mathbb{R}$ 和 $f, g \in V, [\lambda(f+g)](x) = \lambda[f(x) + g(x)] = \lambda f(x) + \lambda g(x) = (\lambda f)(x) + (\lambda g)(x) = [\lambda f + \lambda g](x)$
8. 第二分配律 (向量对标量加法的分配): 对 $\lambda, \mu \in \mathbb{R}$ 和 $f \in V, [(\lambda + \mu)f](x) = (\lambda + \mu)f(x) = \lambda f(x) + \mu f(x) = (\lambda f)(x) + (\mu f)(x) = [\lambda f + \mu f](x)$

具体例子:

- $f(x) = 2 \sin x + 3 \cos x, g(x) = -\sin x + \cos x$
- 加法: $(f+g)(x) = (2-1) \sin x + (3+1) \cos x = \sin x + 4 \cos x$
- 标量乘法: $3f(x) = 6 \sin x + 9 \cos x$
- 零函数: $0(x) = 0 \sin x + 0 \cos x = 0$

示例5.1展示了函数也可以构成向量空间, 函数之间的加法和数乘运算满足向量空间的所有公理.

5.1.2 内积与内积空间

接下来是关于内积和内积空间的定义.

定义 5.4 (内积)

设 V 是 \mathbb{F} 上的向量空间. V 上的一个内积是一个函数, 它将 V 中每个有序对 (u, v) 对应到 \mathbb{F} 中的一个数 $\langle u, v \rangle$, 并且满足以下性质:

1. 正定性 (positivity): 对所有 $v \in V$, 有 $\langle v, v \rangle \geq 0$;
2. 确定性 (definiteness): $\langle v, v \rangle = 0$ 当且仅当 $v = 0$;
3. 第一个变量的可加性 (additivity in first slot): 对所有 $u, v, w \in V$, 有 $\langle u+v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$;
4. 第一个变量的齐次性 (homogeneity in first slot)^a: 对所有 $\lambda \in \mathbb{F}$ 和所有 $u, v \in V$, 有 $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle$;
5. 共轭对称性 (conjugate symmetry): 对所有 $u, v \in V$, 有 $\langle u, v \rangle = \overline{\langle v, u \rangle}$.

^a大多数数学家采用上述方式定义内积, 而许多物理学家则采用另一种定义, 要求齐次性体现在第二个变量而非第一个变量.

定义 5.5 (内积空间)

一个内积空间是一个向量空间 V 连同 V 上的一个内积.

以及关于正交 (orthogonal) 的定义:

定义 5.6 (正交)

若向量 $u, v \in V$ 满足 $\langle u, v \rangle = 0$, 则称 u 与 v 是正交的.

以下给出一个示例:

示例 5.2. 傅里叶级数正交基

考虑定义在区间 $[-\pi, \pi]$ 上的连续实值函数空间 $C_{\mathbb{R}}[-\pi, \pi]^a$, 其上定义内积为:

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)g(x) dx \quad (5.4)$$

考虑三角函数集合:

$$\left\{ \frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \dots, \frac{\cos nx}{\sqrt{\pi}}, \frac{\sin nx}{\sqrt{\pi}}, \dots \right\} \quad (5.5)$$

这些函数在该内积下构成一个正交集合, 因为:

$$\begin{aligned} \langle \cos mx, \cos nx \rangle &= \int_{-\pi}^{\pi} \cos(mx) \cos(nx) dx = \begin{cases} 0 & \text{若 } m \neq n \\ \pi & \text{若 } m = n \neq 0 \\ 2\pi & \text{若 } m = n = 0 \end{cases} \\ \langle \sin mx, \sin nx \rangle &= \int_{-\pi}^{\pi} \sin(mx) \sin(nx) dx = \begin{cases} 0 & \text{若 } m \neq n \\ \pi & \text{若 } m = n \neq 0 \end{cases} \\ \langle \cos mx, \sin nx \rangle &= \int_{-\pi}^{\pi} \cos(mx) \sin(nx) dx = 0 \quad \text{对所有 } m, n \end{aligned}$$

特别地, 对于任意不同的正整数 $m \neq n$:

- $\cos mx$ 与 $\cos nx$ 正交
- $\sin mx$ 与 $\sin nx$ 正交
- $\cos mx$ 与 $\sin nx$ 正交 (对所有 m, n)

将每个函数除以其范数 (即 $\sqrt{\langle f, f \rangle}$) 后, 我们得到标准正交函数集:

$$\left\{ \frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \dots \right\} \quad (5.6)$$

这个正交函数系构成了 $L^2[-\pi, \pi]^b$ 空间的一组完备正交基, 是傅里叶级数展开的基础.

^a $C_{\mathbb{R}}[-\pi, \pi] = \{f : [-\pi, \pi] \rightarrow \mathbb{R} \mid f \text{ 在 } [-\pi, \pi] \text{ 上连续}\}$

^b希尔伯特空间 (Hilbert space), 一种内积空间, 下面会提到

示例5.2显示, 可以由式5.5构成一个向量空间. 同时, 公式5.5中元素的正交性, 也是正交频分复用 (Orthogonal Frequency Division Multiplexing, OFDM) 技术的数学基础.

5.1.3 希尔伯特空间 (Hilbert space)

定义 5.7 (希尔伯特空间)

希尔伯特空间就是完备的内积空间. 具体来说:

设 $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ 是一个内积空间, 定义范数 $\|x\| = \sqrt{\langle x, x \rangle}$. 如果 \mathcal{H} 在这个范数诱导的度量下是完备的 (即每个柯西序列都收敛), 那么 \mathcal{H} 就是一个希尔伯特空间.

示例 5.3. 经典的希尔伯特空间

1. 有限维空间:

$$\mathbb{R}^n \text{ 和 } \mathbb{C}^n \text{ 装备标准内积 } \langle x, y \rangle = \sum_{i=1}^n x_i \bar{y}_i \quad (5.7)$$

都是希尔伯特空间 (有限维空间自动完备).

2. 序列空间 ℓ^2 :

$$\ell^2 = \left\{ (x_n)_{n=1}^{\infty} : x_n \in \mathbb{C}, \sum_{n=1}^{\infty} |x_n|^2 < \infty \right\} \quad (5.8)$$

装备内积 $\langle x, y \rangle = \sum_{n=1}^{\infty} x_n \bar{y}_n$, 这是无限维希尔伯特空间的典型例子.

3. 函数空间 $L^2[a, b]$ (平方可积空间):

$$L^2[a, b] = \left\{ f : [a, b] \rightarrow \mathbb{C} : \int_a^b |f(x)|^2 dx < \infty \right\} \quad (5.9)$$

装备内积 $\langle f, g \rangle = \int_a^b f(x) \overline{g(x)} dx$, 这是信号处理和量子力学中最重要的希尔伯特空间.

5.1.4 对偶 (Duality)

定义 5.8 (线性映射 (linear map))

设 V 和 W 是域 \mathbb{F} 上的向量空间. 从 V 到 W 的线性映射是指满足以下性质的函数 $T : V \rightarrow W$:

- 可加性: 对任意 $u, v \in V$, 有 $T(u + v) = Tu + Tv$;
- 齐次性: 对任意 $\lambda \in \mathbb{F}$ 和 $v \in V$, 有 $T(\lambda v) = \lambda(Tv)$.

注 有些数学家使用“线性变换 (linear transformation)”这一术语, 其含义与“线性映射”相同.

注 设 V 和 W 是域 \mathbb{F} 上的向量空间. 所有从 V 到 W 的线性映射构成的集合记作 $\mathcal{L}(V, W)$.

示例 5.4.

考虑函数空间 $L^1(\mathbb{R})$. 傅里叶变换 $\mathcal{F} : L^1(\mathbb{R}) \rightarrow C_0(\mathbb{R})$ 定义为:

$$(\mathcal{F}f)(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt. \quad (5.10)$$

对任意 $f, g \in L^1(\mathbb{R})$ 和 $\alpha, \beta \in \mathbb{C}$, 有:

$$\mathcal{F}(\alpha f + \beta g) = \alpha \mathcal{F}f + \beta \mathcal{F}g, \quad (5.11)$$

因此 \mathcal{F} 是线性映射.

这里解答了我们抛出的问题 3).

定义 5.9 (线性泛函 (linear functional))

设 V 是域 \mathbb{F} 上的向量空间. V 上的线性泛函是指从 V 到 \mathbb{F} 的线性映射. 也就是说, 线性泛函是 $\mathcal{L}(V, \mathbb{F})$ 中的元素..

定义 5.10 (对偶空间 (dual space, V'))

设 V 是域 \mathbb{F} 上的向量空间, V 的对偶空间, 记作 V' , 是指由 V 上所有线性泛函构成的向量空间. 换言之, $V' = \mathcal{L}(V, \mathbb{F})$.

示例 5.5. 对偶空间

1. 有限维情形: 设 $V = \mathbb{R}^n$, 其标准基为 $\{e_1, \dots, e_n\}$. 对偶空间 V' 由所有形如

$$\varphi(x) = a_1x_1 + \dots + a_nx_n, \quad a_i \in \mathbb{R} \quad (5.12)$$

的线性泛函组成. 事实上, V' 与 \mathbb{R}^n 同构: 每个 $\varphi_a(x) = a \cdot x$ 对应向量 $a \in \mathbb{R}^n$.

2. 对偶基: 对标准基 $\{e_1, \dots, e_n\}$, 其对偶基 $\{\varphi_1, \dots, \varphi_n\} \subset V'$ 定义为

$$\varphi_i(e_j) = \delta_{ij} = \begin{cases} 1 & \text{若 } i = j, \\ 0 & \text{若 } i \neq j. \end{cases} \quad (5.13)$$

具体地, $\varphi_i(x_1, \dots, x_n) = x_i$, 即第 i 个坐标函数.

3. 函数空间情形: 设 $V = \mathcal{P}_n(\mathbb{R})$ 为次数不超过 n 的多项式空间. 考虑线性泛函

$$\varphi_k(p) = p^{(k)}(0) \quad (k \text{ 阶导数在 } 0 \text{ 点的值}). \quad (5.14)$$

对基 $\{1, x, x^2, \dots, x^n\}$, 其对偶基为 $\{\varphi_0, \varphi_1, \frac{1}{2!}\varphi_2, \dots, \frac{1}{n!}\varphi_n\}$, 其中

$$\frac{1}{k!}\varphi_k(x^j) = \delta_{jk}. \quad (5.15)$$

示例 5.6. 对偶空间的计算

设 $V = \mathbb{R}^3$, 考虑线性泛函 $\varphi: \mathbb{R}^3 \rightarrow \mathbb{R}$ 定义为

$$\varphi(x, y, z) = 2x - 3y + 5z. \quad (5.16)$$

1. 矩阵表示: φ 可写为 $\varphi(v) = a^T v$, 其中 $a = (2, -3, 5)^T$.
 2. 在基下的坐标: 对标准基 $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$, 有

$$\varphi(e_1) = 2, \quad \varphi(e_2) = -3, \quad \varphi(e_3) = 5. \quad (5.17)$$

因此 φ 在对偶基下的坐标为 $(2, -3, 5)^a$.

3. 对偶空间的维数: $\dim V' = \dim V = 3$, 且 V' 的基可取为坐标函数:

$$\varphi_1(x, y, z) = x, \quad \varphi_2(x, y, z) = y, \quad \varphi_3(x, y, z) = z. \quad (5.18)$$

^a线性泛函提供了一种不依赖于具体基的向量描述方式

5.1.5 算子, 特征值与特征向量

定义 5.11 (算子, $\mathcal{L}(V)$)

- 从一个向量空间到其自身的线性映射称为算子.
- 记号 $\mathcal{L}(V)$ 表示 V 上所有算子的集合. 换言之, $\mathcal{L}(V) = \mathcal{L}(V, V)$.

定义 5.12 (特征值)

设 $T \in \mathcal{L}(V)$. 若存在非零向量 $v \in V$ 使得 $Tv = \lambda v$, 则称数 $\lambda \in \mathbb{F}$ 为 T 的特征值.

定义 5.13 (特征向量)

设 $T \in \mathcal{L}(V)$, 且 $\lambda \in \mathbb{F}$ 是 T 的特征值. 若向量 $v \in V$ 满足 $v \neq 0$ 且 $Tv = \lambda v$, 则称 v 为对应于特征值 λ 的特征向量.

例如, 对于微分算子 $\frac{d}{dx}$ 而言, 其特征向量为 $e^{\lambda x}$ ($\lambda \in \mathbb{F}$), 特征值为 λ .

示例 5.7. 位置算子与动量算子的特征值问题

1. 位置算子 \hat{x}

定义: 在一维空间中, 位置算子 \hat{x} 作用在波函数 $\psi(x)$ 上就是乘以坐标 x :

$$\hat{x} \psi(x) = x \psi(x) \quad (5.19)$$

特征方程: 要找特征值 x_0 和特征函数 $\psi_{x_0}(x)$ 使得:

$$\hat{x} \psi_{x_0}(x) = x_0 \psi_{x_0}(x) \quad (5.20)$$

即:

$$x \psi_{x_0}(x) = x_0 \psi_{x_0}(x) \implies (x - x_0) \psi_{x_0}(x) = 0 \quad (5.21)$$

解: 这个方程意味着 $\psi_{x_0}(x)$ 在 $x \neq x_0$ 时必须为零. 唯一满足此条件 (且在平方可积函数空间中非零) 的广义函数是狄拉克 δ 函数:

$$\psi_{x_0}(x) = \delta(x - x_0) \quad (5.22)$$

验证:

$$\hat{x} \delta(x - x_0) = x \delta(x - x_0) = x_0 \delta(x - x_0) \quad (5.23)$$

(利用 δ 函数的筛选性质)

2. 动量算子 \hat{p}

定义: 在一维空间中 (取 $\hbar = 1$):

$$\hat{p} = -i \frac{d}{dx} \quad (5.24)$$

作用为:

$$\hat{p} \psi(x) = -i \frac{d\psi}{dx} \quad (5.25)$$

特征方程: 找特征值 p 和特征函数 $\psi_p(x)$ 使得:

$$-i \frac{d\psi_p}{dx} = p \psi_p(x) \quad (5.26)$$

这是一个一阶常微分方程:

$$\frac{d\psi_p}{dx} = ip\psi_p(x) \quad (5.27)$$

解: 解为平面波:

$$\psi_p(x) = Ce^{ipx} \quad (5.28)$$

其中 C 是归一化常数.

验证:

$$\hat{p}e^{ipx} = -i\frac{d}{dx}e^{ipx} = -i(ip)e^{ipx} = pe^{ipx} \quad (5.29)$$

关键对易关系: 位置和动量算子满足著名的对易关系:

$$[\hat{x}, \hat{p}] = \hat{x}\hat{p} - \hat{p}\hat{x} = i\hbar \quad (5.30)$$

对于任意两个不对易的算子 \hat{A} 和 \hat{B} , 它们对应的物理量 A 和 B 的测量精度存在根本限制:

$$\Delta A \cdot \Delta B \geq \frac{1}{2} |[\hat{A}, \hat{B}]| \quad (5.31)$$

对于位置和动量, 这个对易关系直接导致了海森堡不确定性原理:

$$\Delta x \Delta p \geq \frac{\hbar}{2} \quad (5.32)$$

意味着位置和动量不能同时被精确确定.

定义 5.14 (谱, Spectral)

设 $T \in \mathcal{L}(V)$, 其中 V 是数域 \mathbb{F} (通常为实数域或复数域) 上的向量空间 (可以是有限维或无限维). T 的谱, 记作 $\sigma(T)$, 是指所有满足以下条件的标量 $\lambda \in \mathbb{F}$ 的集合: 算子

$$T - \lambda I \quad (5.33)$$

在 V 上不可逆 (即不存在有界逆算子).

在有限维情形下, 谱 $\sigma(T)$ 恰好是 T 的所有特征值的集合.

在无限维情形下, 谱 $\sigma(T)$ 包含了特征值, 但也可能包含其他使得 $T - \lambda I$ 不是双射 (即使得该算子不可逆) 的非特征值标量. 谱通常可分为以下部分:

- 点谱 (特征值谱): 所有使得 $T - \lambda I$ 不是单射的 λ , 即存在非零向量 $v \in V$ 使得 $Tv = \lambda v$.
- 连续谱: 所有使得 $T - \lambda I$ 是单射、值域稠密但不可逆的 λ .
- 剩余谱: 所有使得 $T - \lambda I$ 是单射但其值域不稠密的 λ .

注 在例 5.7 中:

- 位置算子 \hat{x} 的谱是连续的, 由所有实数构成 (在物理中通常取实数), 其特征函数是狄拉克 δ 函数, 它们不属于平方可积函数空间 $L^2(\mathbb{R})$, 而是广义特征函数.
- 动量算子 \hat{p} 的谱也是连续的, 由所有实数构成 (若取 $\hbar = 1$), 其特征函数是平面波 e^{ipx} , 同样不属于 $L^2(\mathbb{R})$, 而是广义特征函数.

这两个算子的谱都是纯连续谱, 没有特征值 (在通常的平方可积函数空间中).

5.1.6 对 FT 的一些观察与猜想

到目前为止, 本文都在介绍空间的定义. 学习内积空间有利于我们完成对时域和频域的定义.

观察下面两个公式:

$$f(t) = \int_{-\infty}^{\infty} f(\tau)\delta(t-\tau) d\tau \quad (5.34)$$

$$f(t) = \int_{-\infty}^{\infty} F(\omega)e^{i\omega t} d\frac{\omega}{2\pi} \quad (5.35)$$

如果可以将 $\delta(t-\tau)$ 和 $e^{i\omega t}$ 视作时域和频域的“正交基”, 我们即可完成对时域和频域的完整定义, 也就是说, 可以很好的回答问题 1) 和问题 2).

5.2 连续 Parseval 框架 (Continuous Parseval Frame)

定义 5.15 (连续 Parseval 框架 (Continuous Parseval Frame))

假设 \mathcal{H} 是一个 Hilbert 空间, (Ω, μ) 是一个测度空间. 如果一个族向量 $\{\phi_\omega\}_{\omega \in \Omega} \subset \mathcal{H}$ 满足对于任何 $f \in \mathcal{H}$, 映射 $\omega \mapsto \langle f, \phi_\omega \rangle$ 是可测的, 且满足以下等式:

$$\|f\|^2 = \int_{\Omega} |\langle f, \phi_\omega \rangle|^2 d\mu(\omega) \quad (5.36)$$

那么这个向量族 $\{\phi_\omega\}$ 就被称为一个连续 Parseval 框架.

定理 5.1 (连续 Parseval 框架重建公式)

设 $\{\phi_\omega\}_{\omega \in \Omega}$ 是 \mathcal{H} 上的一个连续 Parseval 框架. 那么对于任意 $f \in \mathcal{H}$, 有以下弱收敛意义下的重建公式:

$$f = \int_{\Omega} \langle f, \phi_\omega \rangle \phi_\omega d\mu(\omega) \quad (5.37)$$

即, 对任意 $g \in \mathcal{H}$, 有 $\langle f, g \rangle = \int_{\Omega} \langle f, \phi_\omega \rangle \langle \phi_\omega, g \rangle d\mu(\omega)$.

我们可以对 $f(t)$ 做如下两种变形, 第一种是利用函数族 $\delta(\tau), \tau \in \mathbb{R}$ 来进行重构:

$$f(t) = \int_{-\infty}^{\infty} f(\tau)\delta(t-\tau) d\tau \quad (5.38)$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(t)\delta(\tau-t)\delta(t-\tau) dt \right) d\tau \quad (5.39)$$

$$= \int_{-\infty}^{\infty} \langle f(t), \delta(\tau-t) \rangle \delta(t-\tau) d\tau \quad (5.40)$$

$$= \int_{-\infty}^{\infty} \langle f(t), \delta(t-\tau) \rangle \delta(t-\tau) d\tau \quad (5.41)$$

第二种是利用函数族 $e^{i\omega t}, \omega \in \mathbb{R}$ 来进行重构:

$$f(t) = \int_{-\infty}^{\infty} F(\omega)e^{i\omega t} d\frac{\omega}{2\pi} \quad (5.42)$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \right) e^{i\omega t} d\frac{\omega}{2\pi} \quad (5.43)$$

$$= \int_{-\infty}^{\infty} \langle f(t), e^{i\omega t} \rangle e^{i\omega t} d\frac{\omega}{2\pi} \quad (5.44)$$

注意到 $\delta(t-\tau)$ 和 $e^{i\omega t}$ 都不在希尔伯特空间 \mathcal{H} 中 (不是平方可积函数). 因此, 我们需要更加完备的数学框架.

5.3 装备希尔伯特空间 (rigged Hilbert space)

为解决 δ 函数不在希尔伯特空间中的问题, 由量子力学发展出了装备希尔伯特空间的理论.

5.3.1 装备希尔伯特空间 (rigged Hilbert space)

Gelfand 三元组 (Gelfand triple), 也称为装备 Hilbert 空间, 是泛函分析中一个重要的结构, 用于将 Hilbert 空间理论与广义函数 (分布) 理论统一起来.

定义 5.16 (装备希尔伯特空间)

装备希尔伯特空间 (Rigged Hilbert Space), 也称为 **Gelfand** 三元组 (Gelfand triple), 是由三个嵌套的拓扑向量空间构成的结构:

$$\Phi \subset \mathcal{H} \subset \Phi^* \quad (5.45)$$

其中:

1. Hilbert 空间 \mathcal{H} : 通常是我们主要工作的空间, 例如 $L^2(\mathbb{R}^n)$.
2. 测试函数空间 Φ :
 - Φ 是 \mathcal{H} 的一个稠密子空间.
 - Φ 上赋予一个比 \mathcal{H} 范数更强的拓扑 (通常由一族半范数定义), 使得 Φ 成为一个完备的、可分的、局部凸的拓扑向量空间 (通常是 Fréchet 空间或可数 Hilbert 空间).
 - 典型例子: Schwartz 空间 $\mathcal{S}(\mathbb{R}^n)$.
3. 对偶空间 Φ' :
 - Φ' 是 Φ 的连续对偶空间, 即所有连续线性泛函 $T: \Phi \rightarrow \mathbb{C}$ 构成的空间.
 - Φ' 中的元素称为广义函数 (分布).
 - 赋予 Φ' 弱拓扑 (即逐点收敛拓扑).

示例 5.8. 量子力学应用

量子力学中最常见的装备希尔伯特空间是:

$$\mathcal{S}(\mathbb{R}^n) \subset L^2(\mathbb{R}^n) \subset \mathcal{S}'(\mathbb{R}^n) \quad (5.46)$$

其中:

- $\mathcal{S}(\mathbb{R}^n)$: 施瓦茨空间 (Schwartz space), 由所有速降光滑函数组成

$$\mathcal{S}(\mathbb{R}^n) = \{f \in C^\infty(\mathbb{R}^n) : \sup_{x \in \mathbb{R}^n} |x^\alpha \partial^\beta f(x)| < \infty \forall \alpha, \beta\} \quad (5.47)$$

- $L^2(\mathbb{R}^n)$: 平方可积函数空间, 标准量子力学态空间
- $\mathcal{S}'(\mathbb{R}^n)$: 缓增分布空间 (tempered distributions), $\mathcal{S}(\mathbb{R}^n)$ 的连续对偶

在量子力学中, 装备希尔伯特空间解决了以下问题:

1. 连续谱算子的本征矢: 位置算符 \hat{x} 和动量算符 \hat{p} 在 $L^2(\mathbb{R})$ 中没有真正的本征矢, 但在 $\mathcal{S}'(\mathbb{R})$ 中有:

$$\hat{x}\delta(x-x_0) = x_0\delta(x-x_0), \quad \hat{p}e^{ip_0x/\hbar} = p_0e^{ip_0x/\hbar} \quad (5.48)$$

其中 $\delta(x-x_0) \in \mathcal{S}'(\mathbb{R}), e^{ip_0x/\hbar} \in \mathcal{S}'(\mathbb{R})$. (因为 $\delta(x-x_0)$ 和 $e^{ip_0x/\hbar}$ 都是 $\mathcal{S}(\mathbb{R})$ 上的线性泛函, 证明略. 同时注意 5.1.6 小节中我们想要的“正交基”)

2. 傅里叶变换: 在 $\mathcal{S}(\mathbb{R}) \subset L^2(\mathbb{R}) \subset \mathcal{S}'(\mathbb{R})$ 上, 傅里叶变换是良定义的同构. 特别地, 傅里叶变换将位置表象与动量表象联系起来:

$$\mathcal{F}: \mathcal{S}(\mathbb{R}) \rightarrow \mathcal{S}(\mathbb{R}), \quad (\mathcal{F}\psi)(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} e^{-ipx/\hbar} \psi(x) dx \quad (5.49)$$

在此变换下, 位置算符和动量算符的角色互换。

5.3.2 Gelfand-Naimark-Maurin 定理

定理 5.2 (Gelfand-Naimark-Maurin 定理)

设 A 是装备希尔伯特空间 $(\Phi, \mathcal{H}, \Phi^*)$ 上的自伴算子, 则存在测度空间 (Λ, μ) 和一系列广义特征向量 $e_\lambda \in \Phi^*$, 使得对于任意 $f \in \Phi$, 其谱分解可以写为:

$$f = \int_{\Lambda} \hat{f}(\lambda) e_\lambda d\mu(\lambda) \quad (5.50)$$

其中, $\hat{f}(\lambda) = \langle f, e_\lambda \rangle$

对于量子力学中的位置算子和动量算子, Gelfand-Naimark-Maurin 定理的具体形式如下:

示例 5.9. 位置算子 \hat{x}

设装备希尔伯特空间为 $(\mathcal{S}(\mathbb{R}), L^2(\mathbb{R}), \mathcal{S}'(\mathbb{R}))$, 其中 $\mathcal{S}(\mathbb{R})$ 是 Schwartz 空间, $\mathcal{S}'(\mathbb{R})$ 是缓增分布空间. 自伴算子 $\hat{x}: f(x) \mapsto xf(x)$. 则存在测度空间 (\mathbb{R}, dx) 和广义特征向量 $e_\lambda \in \mathcal{S}'(\mathbb{R}), e_\lambda(x) = \delta(x - \lambda)$, 使得对于任意 $f \in \mathcal{S}(\mathbb{R})$, 其谱分解为:

$$f = \int_{\mathbb{R}} \hat{f}(\lambda) e_\lambda d\lambda, \quad \hat{f}(\lambda) = \langle f, e_\lambda \rangle = f(\lambda). \quad (5.51)$$

示例 5.10. 动量算子 \hat{p}

设装备希尔伯特空间为 $(\mathcal{S}(\mathbb{R}), L^2(\mathbb{R}), \mathcal{S}'(\mathbb{R}))$, 自伴算子 $\hat{p} = -i\hbar \frac{d}{dx}$ (取 $\hbar = 1$). 则存在测度空间 (\mathbb{R}, dp) 和广义特征向量 $e_p \in \mathcal{S}'(\mathbb{R}), e_p(x) = \frac{1}{\sqrt{2\pi}} e^{ipx}$, 使得对于任意 $f \in \mathcal{S}(\mathbb{R})$, 其谱分解为:

$$f = \int_{\mathbb{R}} \hat{f}(p) e_p dp, \quad \hat{f}(p) = \langle f, e_p \rangle = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-ipx} dx. \quad (5.52)$$

这里 $\langle \cdot, \cdot \rangle$ 表示 \mathcal{S}' 与 \mathcal{S} 之间的对偶配对, 在计算时需理解为广义函数意义下的积分。

5.4 谱定理 (Spectral Theorem)

谱定理 (Spectral Theorem) 是线性代数中“对角化”概念在一般算子空间上的推广. 它断言: 对于一类重要的算子 (自伴算子), 我们总能找到一个合适的基 (或测度), 使得算子的作用表现为简单的标量乘法。

5.4.1 有限维情形: 自伴算子的谱分解

在有限维希尔伯特空间 \mathcal{H} 中, 谱定理即是对称 (或厄米特) 矩阵的对角化:

定理 5.3 (有限维谱定理)

设 $T \in \mathcal{L}(V)$ 是自伴算子 ($T = T^*$), 则存在 V 的一组标准正交基 $\{e_1, \dots, e_n\}$, 由 T 的特征向量组成. 算子 T 可以表示为:

$$T = \sum_{i=1}^n \lambda_i |e_i\rangle \langle e_i| \quad (5.53)$$

其中 $\lambda_i \in \sigma(T)$ 是特征值, $|e_i\rangle$ 是列向量, $\langle e_i|$ 是行向量, $|e_i\rangle\langle e_i|$ 是向对应特征子空间的投影算子.

5.4.2 无限维情形: 连续谱与谱测度

当算子具有连续谱时, 求和符号 \sum 必须演变为积分符号 \int .

定理 5.4 (无限维谱定理 - 算子形式)

设 \mathcal{H} 为希尔伯特空间, T 为 H 上的有界自伴算子. 则存在唯一的谱测度 (Spectral Measure) E , 使得:

$$T = \int_{\sigma(T)} \lambda dE(\lambda) \quad (5.54)$$

注 只有自伴算子的谱 $\sigma(T)$ 才是实数轴上的子集, 且其剩余谱 (Residual Spectrum) 一定为空. 这意味着自伴算子的谱结构最清晰, 物理意义最明确

5.4.3 谱展开

定理 5.1 即为希尔伯特空间中的谱展开.

定理 5.2 即为装备希尔伯特空间中的谱展开.

示例 5.11. 自伴算子 $A = -i\frac{d}{dx}$ 的谱展开

考虑希尔伯特空间 $H = L^2(\mathbb{R})$ 上的自伴算子 $A = -i\frac{d}{dx}$ (定义在适当域上, 如 Sobolev 空间 $H^1(\mathbb{R})$). 根据谱定理, 对于自伴算子 A , 存在唯一的谱测度 E 使得

$$A = \int_{\mathbb{R}} \lambda dE(\lambda). \quad (5.55)$$

谱定理的另一形式指出, 存在酉算子 $U: H \rightarrow L^2(\mathbb{R}, d\mu)$ 将 A 转换为乘法算子, 即

$$(UAU^{-1}\phi)(\lambda) = \lambda\phi(\lambda), \quad \forall \phi \in L^2(\mathbb{R}, d\mu), \quad (5.56)$$

其中 μ 是 A 的谱测度. 对于 $A = -id/dx$, 其谱为连续谱 $\sigma(A) = \mathbb{R}$, 且谱测度等价于勒贝格测度. 为了得到 U 的具体形式, 求解 A 的广义特征值问题:

$$A\phi_\lambda = \lambda\phi_\lambda \quad \Rightarrow \quad -i\frac{d}{dx}\phi_\lambda(x) = \lambda\phi_\lambda(x). \quad (5.57)$$

解此微分方程得 $\phi_\lambda(x) = Ce^{i\lambda x}$, 其中 C 为归一化常数. 广义特征函数不在 $L^2(\mathbb{R})$ 中, 但可在分布意义下使用. 为使广义特征函数满足 δ 归一化条件:

$$\int_{\mathbb{R}} \phi_\lambda(x) \overline{\phi_{\lambda'}(x)} dx = \delta(\lambda - \lambda'), \quad (5.58)$$

计算得

$$\int_{\mathbb{R}} Ce^{i\lambda x} \overline{Ce^{i\lambda' x}} dx = |C|^2 \int_{\mathbb{R}} e^{i(\lambda - \lambda')x} dx = 2\pi |C|^2 \delta(\lambda - \lambda'), \quad (5.59)$$

因此取 $C = \frac{1}{\sqrt{2\pi}}$, 即

$$\phi_\lambda(x) = \frac{1}{\sqrt{2\pi}} e^{i\lambda x}. \quad (5.60)$$

根据谱展开的思想, 任意函数 $f \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$ 可表示为广义特征函数的连续叠加:

$$f(x) = \int_{\mathbb{R}} c(\lambda) \phi_{\lambda}(x) d\lambda, \quad (5.61)$$

其中系数 $c(\lambda)$ 由 f 与 ϕ_{λ} 的内积给出:

$$c(\lambda) = \langle f, \phi_{\lambda} \rangle = \int_{\mathbb{R}} f(x) \overline{\phi_{\lambda}(x)} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-i\lambda x} dx. \quad (5.62)$$

定义 $\hat{f}(\lambda) = c(\lambda)$, 即得傅里叶变换:

$$\hat{f}(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-i\lambda x} dx. \quad (5.63)$$

代入展开式即得傅里叶逆变换:

$$f(x) = \int_{\mathbb{R}} \hat{f}(\lambda) \phi_{\lambda}(x) d\lambda = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{f}(\lambda) e^{i\lambda x} d\lambda. \quad (5.64)$$

上述变换 $U: f \mapsto \hat{f}$ 是酉算子, 它将 A 转化为乘法算子:

$$(UAU^{-1}\hat{f})(\lambda) = \lambda\hat{f}(\lambda), \quad (5.65)$$

这正是谱定理所保证的. 因此, 傅里叶变换与逆变换可视为自伴算子 $A = -id/dx$ 的谱表示的具体实现.

^a取 $C = \frac{1}{\sqrt{2\pi}}$ 即为常见形式的 FT

示例 5.12. 拉普拉斯算子的谱展开

考虑一维拉普拉斯算子 $A = -\frac{d^2}{dx^2}$, 定义在希尔伯特空间 $H = L^2(\mathbb{R})$ 的适当子集上 (例如 Sobolev 空间 $H^2(\mathbb{R})$). 该算子是自伴的, 其谱为连续谱 $\sigma(A) = [0, \infty)$. 下面通过谱定理推导相应的变换—傅里叶变换及其逆变换.

根据谱定理, 存在谱测度 E 使得

$$A = \int_0^{\infty} \lambda dE(\lambda), \quad (5.66)$$

且存在酉算子 $U: H \rightarrow L^2(\mathbb{R}, d\mu)$ 将 A 转换为乘法算子. 为了显式构造 U , 求解广义特征值问题:

$$A\phi = \lambda\phi \quad \Rightarrow \quad -\frac{d^2}{dx^2}\phi(x) = \lambda\phi(x). \quad (5.67)$$

令 $\lambda = k^2 (k \in \mathbb{R})$, 则方程化为

$$\phi''(x) = -k^2\phi(x), \quad (5.68)$$

其基本解为 e^{ikx} 和 e^{-ikx} . 选择参数 $k \in \mathbb{R}$, 广义特征函数可取为

$$\phi_k(x) = Ce^{ikx}, \quad (5.69)$$

对应特征值 $\lambda = k^2$. 注意 k 与 $-k$ 对应相同的 λ , 因此谱具有二重性.

广义特征函数不在 $L^2(\mathbb{R})$ 中, 但可在分布意义下使用. 为使其满足 δ 归一化:

$$\int_{\mathbb{R}} \phi_k(x) \overline{\phi_{k'}(x)} dx = \delta(k - k'), \quad (5.70)$$

计算左式:

$$\int_{\mathbb{R}} C e^{ikx} \overline{C e^{ik'x}} dx = |C|^2 \int_{\mathbb{R}} e^{i(k'-k)x} dx = 2\pi |C|^2 \delta(k' - k) = 2\pi |C|^2 \delta(k - k'). \quad (5.71)$$

因此取 $C = \frac{1}{\sqrt{2\pi}}$ ^a, 即

$$\phi_k(x) = \frac{1}{\sqrt{2\pi}} e^{ikx}. \quad (5.72)$$

对于 $f \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$, 按广义特征函数展开:

$$f(x) = \int_{\mathbb{R}} c(k) \phi_k(x) dk, \quad (5.73)$$

其中系数 $c(k)$ 由 f 与 ϕ_k 的内积给出:

$$c(k) = \langle f, \phi_k \rangle = \int_{\mathbb{R}} f(x) \overline{\phi_k(x)} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-ikx} dx. \quad (5.74)$$

定义 $\hat{f}(k) = c(k)$, 即得傅里叶变换:

$$\hat{f}(k) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-ikx} dx. \quad (5.75)$$

代入展开式即得傅里叶逆变换:

$$f(x) = \int_{\mathbb{R}} \hat{f}(k) \phi_k(x) dk = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{f}(k) e^{ikx} dk. \quad (5.76)$$

变换 $U: f \mapsto \hat{f}$ 是酉算子, 且将拉普拉斯算子转化为乘法算子:

$$(UAU^{-1}\hat{f})(k) = k^2 \hat{f}(k). \quad (5.77)$$

这验证了谱定理: 在变换后的空间 $L^2(\mathbb{R}, dk)$ 中, A 表现为乘以 k^2 的算子.

通过一维拉普拉斯算子的谱展开, 我们同样导出了傅里叶变换与逆变换. 这再次表明傅里叶变换是实现自伴算子对角化的具体酉变换, 其中广义特征函数 $e^{ikx}/\sqrt{2\pi}$ 构成连续谱下的“基”. 该推导揭示了傅里叶变换在分析微分算子谱理论中的核心作用.

^a取 $C = \frac{1}{\sqrt{2\pi}}$ 即为常见形式的 FT

5.5 总结

现在, 我们可以对时域、频域和傅里叶变换给出一个统一而清晰的解释:

时域是由广义函数簇 $\{\delta(t - \tau)\}_{\tau \in \mathbb{R}}$ 作为“连续基”所张成的空间. 直观上, 时域就是我们通常所说的信号 $f(t)$ 所在的函数空间 (例如 $L^2(\mathbb{R})$). 狄拉克 δ 函数 $\delta(t - \tau)$ 是广义函数 (分布), 它满足筛选性质:

$$\int_{-\infty}^{\infty} f(t) \delta(\tau - t) dt = f(\tau). \quad (5.78)$$

因此,任意充分正则的函数 f 都可以写成

$$f(t) = \int_{\mathbb{R}} f(\tau)\delta(t-\tau) d\tau, \quad (5.79)$$

这相当于将 f 按“时间点” τ 进行分解。 $\delta(t-\tau)$ 可以看作是时间平移算子(或位置算子)的广义特征函数,对应的特征值就是 τ 。虽然它们不属于平方可积空间 L^2 ,但在装备希尔伯特空间(如 $\mathcal{S}'(\mathbb{R})$)的框架下,它们构成了时域的连接“广义基”。

频域是由复指数函数簇 $\{e^{i\omega t}\}_{\omega \in \mathbb{R}}$ 作为“连续正交基”所张成的空间。频域是傅里叶变换后函数 $F(\omega)$ 所在的空间。函数 $e^{i\omega t}$ 是微分算子 $-i\frac{d}{dt}$ (动量算子)的广义特征函数,满足

$$-i\frac{d}{dt}e^{i\omega t} = \omega e^{i\omega t}. \quad (5.80)$$

它们在分布意义下具有正交性:

$$\int_{-\infty}^{\infty} e^{i(\omega-\omega')t} dt = 2\pi\delta(\omega-\omega'), \quad (5.81)$$

因此构成频域的连接“正交基”。同时,它们也是拉普拉斯算子 $\frac{d^2}{dt^2}$ 的广义特征函数,对应的特征值为 $-\omega^2$ 。

傅里叶变换是时域与频域之间的正交基变换(么正变换)。傅里叶正变换

$$\mathcal{F}: f(t) \mapsto F(\omega) = \int_{\mathbb{R}} f(t)e^{-i\omega t} dt. \quad (5.82)$$

将信号 f 从时域的 δ 基表示(即 $f(t)$ 本身)转换到频域的指数基表示(系数 $F(\omega)$)。由于两种广义基在适当的内积下是正交的,该变换保持内积不变,即满足 Parseval 恒等式:

$$\int |f(t)|^2 dt = \int |F(\omega)|^2 \frac{d\omega}{2\pi}, \quad (5.83)$$

因此,傅里叶变换实际上是无穷维希尔伯特空间中的一个正交变换(更准确地说,是酉算子),类似于有限维空间中从一个正交基到另一个正交基的变换。

傅里叶逆变换是时域信号关于拉普拉斯算子(或动量算子)的谱展开。逆变换

$$f(t) = \frac{1}{2\pi} \int_{\mathbb{R}} F(\omega)e^{i\omega t} d\omega \quad (5.84)$$

将时域信号 f 表示为复指数函数 $e^{i\omega t}$ 的连续线性组合。由于 $e^{i\omega t}$ 是微分算子 $-i\frac{d}{dt}$ 或拉普拉斯算子 $\frac{d^2}{dt^2}$ 的广义特征函数,这一展开正是这些自伴算子的谱展开。傅里叶变换实际上实现了所有平移不变线性算子(特别是微分算子)的对角化:在频域中,这些算子的作用简化为乘法运算。

这些观点统一于调和分析与算子理论的框架:傅里叶变换是平移算子群表示论中的谱分解工具,同时实现了函数在时域和频域两种对偶广义基之间的转换,从而为求解微分方程、分析信号以及理解物理系统的频谱结构提供了核心的数学基础。

第 6 章 高斯泼溅算法简介

March 06, 2026

6.1 高斯泼溅简介

高斯泼溅 (Gaussian Splatting) 是一种基于三维点云的直接渲染技术, 其核心思想是通过高斯核函数模型将离散的空间点转换为连续的视觉表示. 该方法避免了传统三维重建中必需的表面网格化步骤, 直接对原始点数据进行可视化处理, 在科学计算可视化、医学成像和地理信息系统等领域具有重要应用价值.

从数学本质上讲, 高斯泼溅为每个三维空间点赋予一个三维高斯函数. 该函数在空间中表现为一个椭球状的密度场, 中心密度最高, 沿径向按指数规律衰减. 在渲染过程中, 每个三维高斯椭球通过相机投影变换为二维高斯椭圆, 随后按深度排序并通过前向透明度合成 (alpha blending) 进行融合, 最终生成连续平滑的渲染图像.

该技术的优势主要体现在三个方面: 首先, 它保持了原始数据的完整性, 避免了网格重建过程中的信息损失和人为误差; 其次, 高斯函数的自然衰减特性使得渲染结果具有连续性, 特别适合表现模糊边界和渐变特征; 最后, 算法支持各向异性调整, 能够通过协方差矩阵控制椭球的形状和方向, 从而准确表达材料的空间各向异性特征.

在技术实现层面, 高斯泼溅建立了一套完整的渲染管线. 输入的点云数据经过参数化处理, 为每个点分配位置、协方差矩阵、颜色和不透明度等属性. 随后, 系统执行三维到二维的投影变换, 将三维高斯椭球映射为屏幕空间的二维高斯分布. 在光栅化阶段, 通过深度排序并执行前向透明度合成处理椭球间的遮挡关系, 最终生成输出图像.

本文将简单介绍高斯泼溅算法的数学原理.

6.2 高斯椭球: 定义与性质

6.2.1 数学定义

三维高斯函数构成了高斯泼溅技术的数学基础. 其标准形式定义为:

$$G(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad (6.1)$$

其中 $\mathbf{x} \in \mathbb{R}^3$ 表示空间中的任意点, $\boldsymbol{\mu} \in \mathbb{R}^3$ 是高斯椭球的中心位置, $\boldsymbol{\Sigma} \in \mathbb{R}^{3 \times 3}$ 是协方差矩阵. 该矩阵必须是实对称正定矩阵, 这一性质保证了高斯函数在三维空间中形成一个椭球状的密度分布. 协方差矩阵 $\boldsymbol{\Sigma}$ 的几何意义至关重

要, 它完整地定义了椭球的形状、大小和空间取向. 通过矩阵分解, 可以将 Σ 表示为:

$$\Sigma = \mathbf{R} \mathbf{S} \mathbf{S}^T \mathbf{R}^T \quad (6.2)$$

其中 $\mathbf{S} = \text{diag}(s_x, s_y, s_z)$ 是对角缩放矩阵, 其对角线元素 s_x, s_y, s_z 分别表示沿三个主轴方向的缩放因子. $\mathbf{R} \in \mathbb{R}^{3 \times 3}$ 是旋转矩阵, 描述了椭球主轴在空间中的方向.

在实际应用中, 旋转矩阵 \mathbf{R} 通常采用四元数 (quaternion) 表示. 四元数 $\mathbf{q} = (q_w, q_x, q_y, q_z)$ 是一个四维超复数, 满足 $q_w^2 + q_x^2 + q_y^2 + q_z^2 = 1$. 与欧拉角相比, 四元数表示具有以下优势: 避免了万向节锁问题, 插值计算更加平滑 (通过球面线性插值 SLERP), 且存储效率更高 (仅需 4 个浮点数). 从四元数到旋转矩阵的转换公式为:

$$\mathbf{R} = \begin{pmatrix} 1 - 2(q_y^2 + q_z^2) & 2(q_x q_y - q_w q_z) & 2(q_x q_z + q_w q_y) \\ 2(q_x q_y + q_w q_z) & 1 - 2(q_x^2 + q_z^2) & 2(q_y q_z - q_w q_x) \\ 2(q_x q_z - q_w q_y) & 2(q_y q_z + q_w q_x) & 1 - 2(q_x^2 + q_y^2) \end{pmatrix} \quad (6.3)$$

6.2.2 光学属性建模

6.2.2.1 基础颜色表示

每个高斯椭球除了几何参数外, 还需要定义其光学特性. 最简单的表示方式是赋予每个椭球一个 RGB 颜色值 $\mathbf{c} \in [0, 1]^3$, 该颜色在传统实现中被视为与视角无关的常量. 然而, 这种简化表示无法捕捉真实世界中材质外观随观察角度变化的现象, 特别是对于具有镜面反射或复杂表面散射特性的材料.

6.2.2.2 球谐函数表示法

为更精确地建模视角相关的外观变化, 现代高斯泼溅系统普遍采用球谐函数 (Spherical Harmonics)¹ 作为颜色表示的基础数学工具.

球谐函数 $Y_l^m(\theta, \phi)$ 构成了一组定义在单位球面上的完备正交基函数, 其中 $l \geq 0$ 表示函数的阶数, $m \in [-l, l]$ 是阶内索引, (θ, ϕ) 为球坐标系下的方向角. 在计算机图形学中, 球谐函数因其良好的数学性质而被广泛使用: 它们具有旋转不变性, 能够通过有限阶截断有效近似低频信号, 且计算相对高效.

将视角相关的颜色函数 $c(\mathbf{d})$ 表示为球谐基函数的线性组合:

$$c(\mathbf{d}) = \sum_{l=0}^L \sum_{m=-l}^l c_l^m Y_l^m(\mathbf{d}) \quad (6.4)$$

这里 \mathbf{d} 是观察方向 (单位向量), c_l^m 是对应的球谐系数 (通常为三维 RGB 向量), L 是使用的最大阶数. 实际应用中, $L=3$ (对应 16 个系数) 或 $L=4$ (对应 25 个系数) 通常能在表达能力和存储成本之间取得良好平衡.

各阶球谐系数具有明确的物理意义: 零阶系数 ($l=0$) 表示环境光或平均颜色; 一阶系数 ($l=1$) 对应漫反射光照的基本方向性; 二阶及以上系数则捕获更复杂的照明效应, 包括镜面高光和材质细节. 这种层次化表示使得系统能够以可控的复杂度模拟从粗糙到精细的各种材质外观.

6.2.2.3 不透明度控制

不透明度参数 $\alpha \in [0, 1]$ 是高斯椭球的另一个关键光学属性, 它控制着该椭球对最终像素颜色的贡献程度. 在混合计算中, 空间位置 \mathbf{x} 处的有效不透明度权重由下式给出:

$$w(\mathbf{x}) = \alpha \cdot G(\mathbf{x}) \quad (6.5)$$

这一设计使得椭球中心区域贡献最大, 边缘区域贡献逐渐衰减为零, 从而产生自然的羽化效果.

¹ 球谐函数是拉普拉斯算子在球坐标下的特征函数, 特征函数之间互相正交. 球坐标下性质良好的函数可由球谐函数进行谱展开, 对应直角坐标系的谱展开可参考示例 5.12. 一般而言, 无界区域的谱是连续谱, 而球坐标系中两个角度坐标都是有界的, 在相应的边界条件下, 使得拉普拉斯算子在两个角度方向上的谱是离散谱.

6.2.3 空间变换与投影

高斯椭球在仿射变换 $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{b}$ 下具有明确的变换规律. 中心位置按 $\boldsymbol{\mu}' = \mathbf{A}\boldsymbol{\mu} + \mathbf{b}$ 变换, 而协方差矩阵遵循 $\boldsymbol{\Sigma}' = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T$ 的变换规则. 这一性质使得高斯表示能够自然地适应场景中的刚体运动、非均匀缩放等变换, 为动态场景的渲染奠定了数学基础.

将三维高斯椭球投影到二维屏幕空间是渲染流程中的关键步骤. 这一过程涉及三个主要计算: 首先, 椭球中心通过相机模型投影到图像平面, 得到二维中心坐标. 接着, 计算透视投影变换在当前中心位置处的雅可比矩阵 \mathbf{J} , 该矩阵编码了投影引起的局部线性变换. 最后, 三维协方差矩阵通过 $\boldsymbol{\Sigma}'_{2D} = \mathbf{J}\boldsymbol{\Sigma}_{3D}\mathbf{J}^T$ 转换为二维形式, 得到屏幕空间中的椭圆表示.

6.2.4 混合合成与渲染优化

6.2.4.1 深度排序策略

由于透明度混合的顺序依赖性, 高斯椭球必须按深度进行排序. 实践中根据性能和质量的不同需求, 可采用多种排序策略: 完全排序对所有椭球进行精确的深度排序, 虽计算成本较高但能保证正确的混合结果; 分块排序将屏幕划分为多个区域, 在各区域内独立排序, 平衡了效率与准确性; 近似排序则采用启发式方法快速获得基本正确的顺序, 适用于对实时性要求极高的应用场景.

6.2.4.2 Alpha 混合公式

对于按深度排序后的椭球序列, 最终像素颜色通过经典的 **over** 操作符计算:

$$C_{\text{final}} = \sum_{i=1}^N c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) \quad (6.6)$$

其中 c_i 是第 i 个椭球的颜色 (可能通过球谐函数计算得到), α_i 是该椭球在当前像素位置的不透明度权重, 乘积项 $\prod_{j=1}^{i-1} (1 - \alpha_j)$ 表示前面所有椭球的累积透光率. 这一公式确保了物理正确的透明度混合效果.

6.2.4.3 参数优化框架

在实际应用中, 高斯椭球的各项参数通常通过可微渲染框架进行端到端优化. 初始化阶段根据输入点云的局部密度分布设置椭球位置和初始尺度; 随后采用基于梯度的优化算法 (如随机梯度下降) 同时优化所有几何和光学参数; 优化过程中还会根据重建质量动态调整椭球分布——在重建不足的区域增加椭球密度, 对冗余椭球进行剪枝, 并根据视角变化自适应调整椭球形状, 从而实现高效且高质量的场景表示.

6.3 高斯椭球的渲染技术

6.3.1 投影变换的数学推导

6.3.1.1 相机模型与坐标变换

设世界坐标系中的点 $\mathbf{x}_w = (x_w, y_w, z_w)^T$, 通过相机外参变换到相机坐标系. 相机外参包含旋转矩阵 $\mathbf{R}_c \in \mathbb{R}^{3 \times 3}$ 和平移向量 $\mathbf{t}_c \in \mathbb{R}^3$, 变换公式为:

$$\mathbf{x}_c = \mathbf{R}_c \mathbf{x}_w + \mathbf{t}_c \quad (6.7)$$

其中 $\mathbf{x}_c = (x_c, y_c, z_c)^T$ 是点在相机坐标系中的坐标. 这一线性变换保持了三维空间的度量关系, 为后续投影计算奠定了基础.

6.3.1.2 协方差矩阵的坐标系变换

高斯椭球的协方差矩阵 Σ_w 定义在世界坐标系中. 当椭球从世界坐标系变换到相机坐标系时, 其协方差矩阵遵循仿射变换规则:

$$\Sigma_c = \mathbf{R}_c \Sigma_w \mathbf{R}_c^\top \quad (6.8)$$

这一变换仅包含旋转部分, 因为协方差矩阵描述的是椭球的形状和方向, 不受平移影响. 变换后的 Σ_c 表示椭球在相机坐标系中的几何特性.

6.3.1.3 透视投影与雅可比矩阵

不失一般性, 假设相机内参已包含在坐标变换中, 或采用归一化相机坐标系进行讨论. 透视投影将相机坐标系中的点 $\mathbf{x}_c = (x_c, y_c, z_c)^\top$ 映射到归一化设备坐标 (x_n, y_n) :

$$x_n = \frac{x_c}{z_c}, \quad y_n = \frac{y_c}{z_c} \quad (6.9)$$

对于高斯椭球中心 μ_c 在相机坐标系中的投影, 需要计算投影变换的雅可比矩阵 $\mathbf{J} \in \mathbb{R}^{2 \times 3}$. 该矩阵描述了三维空间到二维图像平面的局部线性近似. 设投影函数 $\pi: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ 定义为:

$$\pi(\mathbf{x}_c) = \begin{pmatrix} x_c/z_c \\ y_c/z_c \end{pmatrix} \quad (6.10)$$

其雅可比矩阵为:

$$\mathbf{J} = \frac{\partial \pi}{\partial \mathbf{x}_c} = \begin{pmatrix} \frac{1}{z_c} & 0 & -\frac{x_c}{z_c^2} \\ 0 & \frac{1}{z_c} & -\frac{y_c}{z_c^2} \end{pmatrix} \quad (6.11)$$

在椭球中心 μ_c 处求值:

$$\mathbf{J}(\mu_c) = \begin{pmatrix} \frac{1}{\mu_{c,z}} & 0 & -\frac{\mu_{c,x}}{\mu_{c,z}^2} \\ 0 & \frac{1}{\mu_{c,z}} & -\frac{\mu_{c,y}}{\mu_{c,z}^2} \end{pmatrix} \quad (6.12)$$

6.3.1.4 完整的协方差投影变换

三维协方差矩阵 Σ_c 经过投影变换后, 得到二维协方差矩阵 Σ_{2D} . 完整的变换链为:

$$\Sigma_{2D} = \mathbf{J} \Sigma_c \mathbf{J}^\top = \mathbf{J} \mathbf{R}_c \Sigma_w \mathbf{R}_c^\top \mathbf{J}^\top \quad (6.13)$$

这是一个 2×2 的实对称正定矩阵, 描述了投影后椭圆在图像平面上的形状、大小和方向. 矩阵 Σ_{2D} 的特征值的平方根决定了椭圆的长短轴半径, 特征向量决定了椭圆的方向.

6.3.2 二维高斯分布的参数化

6.3.2.1 椭圆的标准形式

投影得到的二维高斯函数可以表示为:

$$G_{2D}(\mathbf{u}) = \exp\left(-\frac{1}{2}(\mathbf{u} - \mu_{2D})^\top \Sigma_{2D}^{-1}(\mathbf{u} - \mu_{2D})\right) \quad (6.14)$$

其中 $\mathbf{u} = (u, v)^\top$ 是图像平面坐标, $\mu_{2D} = \pi(\mu_c)$ 是投影中心.

6.3.2.2 椭圆几何参数提取

从 2×2 协方差矩阵 Σ_{2D} 可以提取椭圆的几何参数. 对 Σ_{2D} 进行特征值分解:

$$\Sigma_{2D} = \mathbf{V} \Lambda \mathbf{V}^\top \quad (6.15)$$

其中 $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2]$ 是特征向量矩阵, $\Lambda = \text{diag}(\lambda_1, \lambda_2)$ 是特征值矩阵, 且 $\lambda_1 \geq \lambda_2 > 0$. 椭圆参数为:

长轴半径: $a = \sqrt{\lambda_1}$ 短轴半径: $b = \sqrt{\lambda_2}$ 旋转角度: $\theta = \arctan 2(v_{1,y}, v_{1,x})$, 其中 $\mathbf{v}_1 = (v_{1,x}, v_{1,y})^\top$

6.3.2.3 边界框计算

为加速光栅化, 需要计算椭圆在图像空间中的轴对齐边界框 (AABB). 对于旋转椭圆, 其边界框可以通过以下方式计算: 设椭圆方程为:

$$\frac{((u - \mu_u) \cos \theta + (v - \mu_v) \sin \theta)^2}{a^2} + \frac{(-(u - \mu_u) \sin \theta + (v - \mu_v) \cos \theta)^2}{b^2} = 1 \quad (6.16)$$

边界框的极值点为:

$$u_{\min} = \mu_u - \sqrt{a^2 \cos^2 \theta + b^2 \sin^2 \theta} \quad (6.17)$$

$$u_{\max} = \mu_u + \sqrt{a^2 \cos^2 \theta + b^2 \sin^2 \theta} \quad (6.18)$$

$$v_{\min} = \mu_v - \sqrt{a^2 \sin^2 \theta + b^2 \cos^2 \theta} \quad (6.19)$$

$$v_{\max} = \mu_v + \sqrt{a^2 \sin^2 \theta + b^2 \cos^2 \theta} \quad (6.20)$$

在工程实践中, 为减少计算开销, 常利用 Σ_{2D} 的对角线元素结合置信区间直接估计 AABB 范围. 设 $\Sigma_{2D} = \begin{pmatrix} s_{uu} & s_{uv} \\ s_{uv} & s_{vv} \end{pmatrix}$, 则边界框的半宽可直接由 $\Delta u = 3\sqrt{s_{uu}}$ 和 $\Delta v = 3\sqrt{s_{vv}}$ 估计, 其中 3 是常用的置信度因子, 对应约 99.7% 的置信区间.

6.4 高斯泼溅示例

6.4.1 问题设定

考虑一个简化场景: 单个高斯椭球在特定视角下的渲染计算. 为聚焦核心数学推导, 我们做如下简化假设:

1. 几何简化: 椭球位于世界坐标系原点, 且已对齐相机坐标系
 - 中心位置: $\mu_w = (0, 0, 1)^\top$
 - 旋转矩阵: $\mathbf{R}_c = \mathbf{I}$ (单位矩阵)
 - 缩放矩阵: $\mathbf{S} = \text{diag}(0.1, 0.1, 0.1)$
 - 协方差矩阵: $\Sigma_w = \mathbf{R}\mathbf{S}\mathbf{S}^\top\mathbf{R}^\top = \mathbf{S}^2 = \text{diag}(0.01, 0.01, 0.01)$
2. 光学简化: 使用一阶球谐函数 (4 个系数) 表示颜色
 - 球谐系数: $\mathbf{c}_{\text{sh}} = [c_0^0, c_1^{-1}, c_1^0, c_1^1]^\top$
 - 不透明度: $\alpha = 0.8$
3. 相机简化: 相机位于 $(0, 0, 0)$, 看向 $+z$ 方向
 - 观察方向: $\mathbf{d} = (0, 0, 1)^\top$
 - 像素位置: $\mathbf{p} = (0.02, 0.01)^\top$ (归一化设备坐标)

6.4.2 坐标变换计算

考虑世界坐标到相机坐标, 由于 $\mathbf{R}_c = \mathbf{I}$ 且 $\mathbf{t}_c = \mathbf{0}$, 椭球中心在相机坐标系中为:

$$\mu_c = \mathbf{R}_c \mu_w + \mathbf{t}_c = \mu_w = (0, 0, 1)^\top \quad (6.21)$$

协方差矩阵变换:

$$\Sigma_c = \mathbf{R}_c \Sigma_w \mathbf{R}_c^\top = \Sigma_w = \text{diag}(0.01, 0.01, 0.01)$$

计算投影中心:

$$\mu_{2D} = \pi(\mu_c) = \begin{pmatrix} 0/1 \\ 0/1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (6.22)$$

计算雅可比矩阵 \mathbf{J} 在 $\boldsymbol{\mu}_c$ 处的值:

$$\mathbf{J} = \begin{pmatrix} \frac{1}{1} & 0 & -\frac{0}{1^2} \\ 0 & \frac{1}{1} & -\frac{0}{1^2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad (6.23)$$

接下来计算二维协方差矩阵:

$$\boldsymbol{\Sigma}_{2D} = \mathbf{J}\boldsymbol{\Sigma}_c\mathbf{J}^\top = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0.01 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (6.24)$$

逐项计算:

$$\mathbf{J}\boldsymbol{\Sigma}_c = \begin{pmatrix} 0.01 & 0 & 0 \\ 0 & 0.01 & 0 \end{pmatrix} (\mathbf{J}\boldsymbol{\Sigma}_c)\mathbf{J}^\top = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}$$

因此:

$$\boldsymbol{\Sigma}_{2D} = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix} \quad (6.25)$$

这是一个各向同性的协方差矩阵, 对应圆形高斯分布.

6.4.3 权重计算

二次型计算, 对于像素 $\mathbf{p} = (0.02, 0.01)^\top$:

$$\mathbf{p} - \boldsymbol{\mu}_{2D} = \begin{pmatrix} 0.02 \\ 0.01 \end{pmatrix} \quad (6.26)$$

计算二次型:

$$Q(\mathbf{p}) = (\mathbf{p} - \boldsymbol{\mu}_{2D})^\top \boldsymbol{\Sigma}_{2D}^{-1} (\mathbf{p} - \boldsymbol{\mu}_{2D}) \quad (6.27)$$

展开计算:

$$Q(\mathbf{p}) = \begin{pmatrix} 0.02 & 0.01 \end{pmatrix} \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix} \begin{pmatrix} 0.02 \\ 0.01 \end{pmatrix} \quad (6.28)$$

$$= \begin{pmatrix} 0.02 & 0.01 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 0.02 \times 2 + 0.01 \times 1 = 0.04 + 0.01 = 0.05 \quad (6.29)$$

高斯权重:

$$w(\mathbf{p}) = \exp\left(-\frac{1}{2}Q(\mathbf{p})\right) = \exp\left(-\frac{0.05}{2}\right) = \exp(-0.025) \quad (6.30)$$

计算数值:

$$w(\mathbf{p}) \approx 0.9753 \quad (6.31)$$

有效不透明度:

$$\alpha_{\text{eff}}(\mathbf{p}) = \alpha \cdot w(\mathbf{p}) = 0.8 \times 0.9753 \approx 0.7802 \quad (6.32)$$

6.4.4 颜色计算 (一阶球谐函数)

6.4.4.1 一阶球谐基函数

对于观察方向 $\mathbf{d} = (0, 0, 1)^\top$, 在球坐标系中对应 $\theta = 0, \phi$ 任意. 一阶球谐基函数值为:

$$Y_0^0(\mathbf{d}) = \frac{1}{2\sqrt{\pi}} \approx 0.2821 \quad (6.33)$$

$$Y_1^{-1}(\mathbf{d}) = \frac{\sqrt{3}}{2\sqrt{\pi}} \cdot \sin\theta \cdot \sin\phi = 0 \quad (\text{因 } \sin\theta = 0) \quad (6.34)$$

$$Y_1^0(\mathbf{d}) = \frac{\sqrt{3}}{2\sqrt{\pi}} \cdot \cos \theta = \frac{\sqrt{3}}{2\sqrt{\pi}} \cdot 1 \approx 0.4886 \quad (6.35)$$

$$Y_1^1(\mathbf{d}) = \frac{\sqrt{3}}{2\sqrt{\pi}} \cdot \sin \theta \cdot \cos \phi = 0 \quad (\text{因 } \sin \theta = 0) \quad (6.36)$$

因此, 基函数向量为:

$$\mathbf{y}(\mathbf{d}) = \begin{pmatrix} 0.2821 \\ 0 \\ 0.4886 \\ 0 \end{pmatrix} \quad (6.37)$$

6.4.4.2 球谐系数设定

假设球谐系数为:

$$\mathbf{c}_{\text{sh}} = \begin{pmatrix} c_0^0 \\ c_1^{-1} \\ c_1^0 \\ c_1^1 \end{pmatrix} = \begin{pmatrix} 1.0 & 0.0 & 0.0 \\ (\text{红色})0.2 & 0.0 & 0.0 \\ 0.5 & 0.0 & 0.0 \\ 0.1 & 0.0 & 0.0 \end{pmatrix} \quad (6.38)$$

这里每个系数是 RGB 向量, 为简化只考虑红色通道.

6.4.4.3 颜色计算

颜色是基函数与系数的线性组合:

$$c(\mathbf{d}) = \sum_{l=0}^1 \sum_{m=-l}^l c_l^m Y_l^m(\mathbf{d}) \quad (6.39)$$

对于红色通道:

$$c_r(\mathbf{d}) = c_0^0 Y_0^0 + c_1^{-1} Y_1^{-1} + c_1^0 Y_1^0 + c_1^1 Y_1^1 \quad (6.40)$$

代入数值:

$$c_r(\mathbf{d}) = 1.0 \times 0.2821 + 0.2 \times 0 + 0.5 \times 0.4886 + 0.1 \times 0 \quad (6.41)$$

$$= 0.2821 + 0.2443 = 0.5264 \quad (6.42)$$

因此, 该像素处的颜色为 (0.5264, 0, 0)(深红色).

6.4.5 最终像素颜色

由于只有一个椭球, 混合计算简化为:

$$C_{\text{final}} = \alpha_{\text{eff}}(\mathbf{p}) \cdot c(\mathbf{d}) \quad (6.43)$$

对于红色通道:

$$C_r = 0.7802 \times 0.5264 \approx 0.4107 \quad (6.44)$$

最终像素颜色为 (0.4107, 0, 0).

如果混合第二个高斯椭球: 假设场景中增加第二个绿色高斯椭球, 其在该像素处的有效不透明度 $\alpha_{\text{eff},2} = 0.5$, 颜色 $c_2 = (0, 0.5, 0)$. 我们需要按照深度顺序进行混合. 假设红色椭球离相机较近 (在前), 绿色椭球较远 (在后), 则从前往后混合: 初始化: $C = (0, 0, 0)$, $T = 1$ 第一步 (红色椭球):

$$C \leftarrow C + T \cdot \alpha_{\text{eff},1} \cdot c_1 = (0, 0, 0) + 1 \times 0.7802 \times (0.5264, 0, 0) \approx (0.4107, 0, 0) \quad (6.45)$$

$$T \leftarrow T \cdot (1 - \alpha_{\text{eff},1}) = 1 \times (1 - 0.7802) = 0.2198 \quad (6.46)$$

第二步 (绿色椭球):

$$C \leftarrow C + T \cdot \alpha_{\text{eff},2} \cdot c_2 = (0.4107, 0, 0) + 0.2198 \times 0.5 \times (0, 0.5, 0) \approx (0.4107, 0.05495, 0) \quad (6.47)$$

最终颜色约为 $(0.4107, 0.05495, 0)$, 呈现以红色为主, 略带绿色的混合效果. 如果两个椭球顺序颠倒, 最终颜色会明显不同, 这正是 *alpha* 混合顺序依赖性的体现.