Available online at www.sciencedirect.com

ScienceDirect

《数学与计算机模拟》00（2026年）1–21

Procedia Computer Science

arXiv:2601.01072v1 [math.NA] 3 Jan 2026

F用于准蒙特卡洛积分的低匹配传输

曾志军[a]，陈建龙[a,*]

[a]清华大学数学科学系，中国北京 *100084*

摘要

针对复杂目标测度的高维积分仍是计算科学中的一项根本挑战。流匹配（FM）为构建连续时间传输映射提供了一种强大范式，但其在高精度积分中的应用受到数值常微分方程求解器固有离散化偏差的严重限制；与准蒙特卡罗（QMC）方法结合时，也缺乏严格的收敛保证。本文通过提出流匹配重要性抽样准蒙特卡洛（FM-ISQMC），弥补了这些关键缺口。该框架旨在将有偏生成流转化为无偏的高阶积分方案。在方法上，我们将逻辑斯蒂基变换与欧拉离散化神经常微分方程场复合，以构建传输映射，并采用重要性抽样校正残余传输误差。我们的核心贡献有两方面。首先，我们针对任意传输映射建立了准蒙特卡罗重要性抽样的一般收敛分析，并确定了使均方根误差达到 $O(N^{-1+\varepsilon})$ 阶的充分增长条件。其次，我们严格证明了流匹配的特定传输架构满足这些条件。因此，我们证明无偏 FM-ISQMC 估计量的均方根误差为 $O(N^{-1+\varepsilon})$，将经典准蒙特卡罗理论拓展至生成模型领域。数值实验表明，FM-ISQMC 持续突破直接传输方法中观察到的误差下限，实现了更高的精度。因此，本研究架起了深度生成建模与数值积分之间的桥梁。

© 2011 爱思唯尔有限公司出版。

关键词： 准蒙特卡罗；流匹配；生成模型；传输映射。

1. 引言

高维数值积分是科学与工程众多领域中的一项基本挑战，包括贝叶斯统计、计算物理和金融数学。通常，目标是计算函数f相对于复杂高维概率分布 π 的期望，即 $\mathbb{E}_{X\sim\pi}[f(X)]$。标准蒙特卡洛（MC）方法具有稳健性，并且在很大程度上与维数无关，但其收敛速率受限于$O(N^{-1/2})$，其中N表示样本数量。为缓解这一根本性限制，人们引入了准蒙特卡罗（QMC）和随机化准蒙特卡罗（RQMC）方法。在满足充分正则性的条件下，这些变体能够实现显著更快的收敛，通常可达到接近$O(N^{-1})$ [1, 2]的速率。

然而，准蒙特卡罗方法的效率受被积函数的光滑性和定义域几何形状的显著影响。当目标分布 π 高度不均匀或呈多峰时，直接应用准蒙特卡罗方法（例如，通过对均匀测度进行简单的逆变换）可能效果不佳；如果由此得到的

*通讯作者。电子邮箱： E-mail address： zengzj22@mails.tsinghua.edu.cn (Z. Zeng).chen-jl22@mails.tsinghua.edu.cn (J. Chen).（JianlongChen）

被积函数变得奇异或高度振荡，则在理论上可能无效。解决这一问题的有效策略是测度传输或传输映射 [3]。其思路是构造一个微分同胚 $\boldsymbol{\tau} : (0,1)^d \to \mathbb{R}^d$，将简单的参考测度（例如，均匀分布 $\mathcal{U}(0,1)^d$）推前至复杂的目标分布 π。如果能得到这样的映射，就可以将积分转化为超立方体上的期望，从而有效利用准蒙特卡罗点（例如 Sobol 序列或格点规则）。

近年来，深度生成模型已成为从数据或未归一化密度中学习传输映射的主流范式。正规化流 (NFs) [4, 5]将映射参数化为可逆层的复合，并已成功应用于重要性抽样和变分推断。近来，由常微分方程 (ODEs) [5] 建模的连续正规化流 (CNFs) 也受到关注。其中，流匹配(FM) [5] 已发展成为一种最先进的框架。与传统扩散模型或连续正规化流的最大似然训练不同，流匹配直接回归一个向量场，以生成连接源分布与目标分布的概率路径。这种无需模拟的训练目标能够带来更稳定的训练和更直的传输路径，因此流匹配在设计高效采样器方面颇具吸引力。

尽管基于流的采样器在经验上表现出色，但关于其与准蒙特卡罗方法结合的理论分析直到最近才开始受到关注。例如，Andral [6]通过经验结果证明，将正规化流与随机化准蒙特卡罗结合能够缩减方差。在理论方面，Liu [7]提出了采用三角映射的传输准蒙特卡罗框架，而 Du 和 He [8] 则针对被积函数无界的随机化准蒙特卡罗自归一化重要性抽样，建立了L^p-误差率。然而，将流匹配（连续时间流）的架构特性与准蒙特卡罗边界增长条件联系起来的具体分析仍然有限。核心问题在于，准蒙特卡罗收敛理论要求被积函数具有严格的正则性。具体而言，若要使变换后的被积函数$f(\boldsymbol{\tau}(\mathbf{u}))$受益于随机化准蒙特卡罗的$O(N^{-1+\epsilon})$收敛率，就必须满足*Owen* 边界增长条件 [9]。该条件限制了当输入$\mathbf{u}$趋近单位超立方体边界时，函数的混合偏导数增长的速度。生成模型虽然针对全局分布匹配进行优化（例如最小化 Kullback-Leibler 散度），却不会显式控制边界处逐点导数的增长。此外，近期理论结果表明，流匹配中的最优速度场在边界处可能出现奇点 [10], ，从而可能违反准蒙特卡罗所要求的光滑性假设。

本文提出了一个针对流匹配准蒙特卡罗（FM-QMC）的严谨框架，并对其收敛性质进行了全面的理论分析。我们重点研究由流匹配生成的传输映射的结构性质，采用的架构选择为：将逻辑斯蒂基映射与由有界向量场驱动的常微分方程流复合。

我们的主要贡献总结如下。首先，我们针对使用任意传输映射的重要性抽样准蒙特卡罗（ISQMC）估计量，建立了一般性的收敛分析。我们为一般微分同胚保留最优准蒙特卡罗收敛率提出了一组充分条件，并区分了值增长条件 （控制点被映射到目标分布尾部的速度）和导数增长条件（控制雅可比矩阵及高阶导数的剧增）。其次，我们证明，通过流匹配构造的传输映射满足这些严格的增长条件。具体而言，我们证明，如果学习得到的向量场及其导数一致有界（标准神经网络在采用有界激活函数时满足这一条件），且常微分方程通过固定步长欧拉法求解，那么所得映射 $\boldsymbol{\tau}$在边界附近具有足够良好的性质。据我们所知，这是首批明确将神经常微分方程的架构性质与准蒙特卡罗理论所要求的边界增长条件联系起来的研究之一。第三，基于增长分析，我们证明，对于任意 $\epsilon > 0$，只要目标函数f 具有多项式增长，FM-QMC 估计量就能达到$O(N^{-1+\epsilon})$的均方根误差（RMSE）。这为在高精度积分任务中使用流匹配奠定了坚实的理论基础。最后，我们通过在各种高维基准测试上开展数值实验验证了理论结果，观察到 FM-QMC 的表现始终优于标准蒙特卡罗方法，且其收敛率与理论预测一致。

本文其余部分的组织如下。第 2 节回顾随机化准蒙特卡罗积分和流匹配框架的背景。第 3 节介绍传输映射边界增长条件的一般理论。第 4 节构造流匹配传输映射，并给出证明其满足所需增长条件的理论证明（定理15和定理16）。第 5 节介绍数值实验，展示我们方法的效率和收敛率。最后，第 6 节总结全文并讨论未来的研究方向。

2. 背景

我们介绍准蒙特卡罗（QMC）、随机化准蒙特卡罗（RQMC）和流匹配的背景

2.1. 准蒙特卡罗与随机化准蒙特卡罗

准蒙特卡罗（QMC）积分通过以下方式估计形如下式的积分

$$\mu = \int_{[0,1]^d} f(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

用确定性的低差异点集 $\{\mathbf{u}_i\}_{i=1}^n \subset [0,1]^d$ 取代独立随机抽样。准蒙特卡罗估计量为

$$\widehat{\mu}_n \;=\; \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{u}_i)\,.$$

著名的 Koksma–Hlawka不等式 [11, 12] 表明

$$\left|\widehat{\mu}_n - \mu\right| \;\le\; V_{\mathrm{HK}}(f)\;\; D^*(\mathbf{u}_1,\dots,\mathbf{u}_n), \tag{1}$$

其中 $V_{\mathrm{HK}}(f)$ 是 f 的 Hardy–Krause变差，而星差异为

$$D^*(\mathbf{u}_1,\dots,\mathbf{u}_n) \;=\; \sup_{\mathbf{a}\in[0,1]^d}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{\mathbf{u}_i\in[0,\mathbf{a})\} \;-\; \prod_{j=1}^{d} a_j\right|.$$

星差异为 $O(n^{-1}(\log n)^d)$ 的点集称为低差异点集 [13]。对于低差异点集，当 $V_{\mathrm{HK}}(f) < \infty$ 时，对任意 $\varepsilon > 0$，都可以得到阶为 $O(n^{-1+\varepsilon})$ 的积分误差。

在实践中，$V_{\mathrm{HK}}(f)$ 往往是无穷大，尤其是当 f 无界或存在奇点时。随机化准蒙特卡罗（RQMC）对低差异点集 $\{\mathbf{u}_i\}_{i=1}^n \subset [0,1]^d$ 施加随机扰动，使随机化后的点 $\{\mathbf{u}'_i\}$ 的每个边缘分布均为均匀分布$([0,1]^d)$，同时几乎必然保持低差异性。RQMC估计量

$$\hat{\mu}'_n \;=\; \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{u}'_i)$$

is u无偏（即 $\mathbb{E}[\hat{\mu}'_n] = \mu$），其方差可根据$R$ 个独立重复样本 $\{\hat{\mu}_n^{(r)}\}_{r=1}^R$ 估计为

$$\widehat{\mathrm{Var}}(\hat{\mu}'_n) = \frac{1}{R-1}\sum_{r=1}^{R}(\hat{\mu}_n^{(r)} - \bar{\mu}_n)^2, \quad \bar{\mu}_n = \frac{1}{R}\sum_{r=1}^{R}\hat{\mu}_n^{(r)}.$$

常见的随机化方案包括:

1. 模 *1* 随机平移。 抽取 $\mathbf{\Delta} \sim$ 均匀分布$([0,1]^d)$，并令

$$\mathbf{u}'_i = (\mathbf{u}_i + \mathbf{\Delta}) \bmod 1.$$

在 f 满足适当的光滑性和周期性的条件下，对于任意 $\varepsilon > 0$，$\mathrm{Var}[\hat{\mu}'_n] = O(n^{-2+\varepsilon})$。

2. 数字平移。 对于一个以 b 为基数的数字网，抽取 $\mathbf{\Delta} \sim$ 均匀分布$(\{0,\dots,b-1\}^d)$，并定义

$$\mathbf{u}'_i = \mathbf{u}_i \oplus \mathbf{\Delta},$$

其中，$\oplus$ 表示逐位模 b 加法。只要 $\mathbb{E}[f(\mathbf{u})^2] < \infty$，就有 $\mathrm{Var}[\hat{\mu}'_n] = o(n^{-1})$。3. *Owen* 扰乱。对一个 (*t*,*m*, *s*)-网进行嵌套均匀扰乱，可得

$$\mathrm{Var}[\hat{\mu}'_n] = O(n^{-3+\varepsilon}) \quad \text{for any } \varepsilon > 0,$$

若 f 具有平方可积混合一阶导数 [14, 15]。

关于准蒙特卡罗的全面介绍，请参阅 [16, 17] ； 关于随机化准蒙特卡罗的综述，请参阅 [18] 。

2.2. 增长条件

我们说明一下偏导数所用的记号。令 $1:d = \{1,\dots,d\}$。对于集合 $v \subseteq \{1,\dots,d\}$，符号$\partial^v$ 表示关于由 v索引的变量的混合偏导数，定义为 $\partial^v = \prod_{j\in v} \frac{\partial}{\partial u_j}$。对于多重指标 $\lambda = (\lambda_1,\dots,\lambda_d) \in \mathbb{N}_0^d$， ∂^λ 表示微分算子 $\frac{\partial^{|\lambda|}}{\partial x_1^{\lambda_1}\dots x_d^{\lambda_d}}$，其中 $|\lambda| = \sum_{i=1}^d \lambda_i$。随后，我们回顾由 [9]提出的被积函数$h$： $[0,1]^d \to \mathbb{R}$的边界增长条件。

假设1（边界增长条件）。 对于任意小的 $B > 0$，存在常数 $C > 0$，使得

$$|\partial^v h(\mathbf{u})| \le C \prod_{j=1}^{d} [\min(u_j, 1-u_j)]^{-B-\mathbf{1}_{\{j\in v\}}} \tag{2}$$

对于任意 $v \subseteq 1{:}\ d$ 且 $\mathbf{u} \in (0,\ 1)^d$。

假设1限制了被积函数及其一阶混合偏导数在边界附近的增长率。当坐标 u_j 趋近于 0 或 1 时，导数 $\partial^v h$（$\mathbf{u}$）受边界距离倒数的某次幂所界定。

以下定理改编自 [9] 的定理5.7。

定理2 (改编自 [9] 的定理5.7)。 若被积函数 h 满足边界增长条件 *1*，则扰乱网估计量 $\hat{\mu}_n$ 的均方根误差（*RMSE*）为 $O(n^{-1+\epsilon})$ 阶，其中 $\epsilon > 0$ 可任意小。

关于 复合函数导数增长的分析依赖于多元 Fa`a di Bruno 公式 .

定理3 (多元 Fa`a di Bruno 公式 [19])。 设 $f:\ \mathbb{R}^d \to \mathbb{R}$ 和 $\boldsymbol{\tau}:\ \mathbb{R}^d \to \mathbb{R}^d$ 为光滑函数。对于 $h = f \circ \boldsymbol{\tau}$ 以及任意非空集合 $v \subseteq \{1,\dots,d\}$，有

$$\partial^v h(\mathbf{u}) = \sum_{\lambda\in\mathbb{N}_0^d:\, 1\le|\lambda|\le|v|} (\partial^\lambda f)(\boldsymbol{\tau}(\mathbf{u})) \sum_{s=1}^{|v|} \sum_{(\mathbf{k},\boldsymbol{\ell})\in\mathcal{A}(\lambda,s)} \prod_{r=1}^{s} \partial^{\boldsymbol{\ell}_r} \boldsymbol{\tau}_{k_r}(\mathbf{u}),$$

其中

$$\mathcal{A}(\lambda,s) = \left\{ (\mathbf{k},\boldsymbol{\ell}) = (k_1,\dots,k_s,\ell_1,\dots,\ell_s) : k_r \in \{1,\dots,d\}, \ell_r \subseteq \{1,\dots,d\}, |\{j\in\{1,\dots,s\} : k_j = i\}| = \lambda_i, \bigsqcup_{r=1}^{s} \ell_r = v \right\}.$$

这里，记号 $\bigsqcup_{r=1}^{s} \mathrm{l}_r = v$ 表示集合 $\mathrm{l}_1,\dots,\mathrm{l}_s$ 构成 v 的一个划分，也就是说，集合 $\mathrm{l}_1,\dots,\mathrm{l}_s$ 两两不相交，且其并集为 v。

2.3. 传输映射与重要性抽样

准蒙特卡罗及其随机化变体（随机化准蒙特卡罗）生成的点集能够均匀覆盖单位立方体 $[0,1]^d$。在许多应用中，我们必须从定义在 $\mathbb{R}^d$ 上、难以处理的目标分布 π 中抽样，以估计 $E_{X\sim\pi}[f(X)]$。因此，我们希望找到一个可测映射

$$\boldsymbol{\tau}:\ [0,1]^d \longrightarrow \mathbb{R}^d \quad \text{such that} \quad \mathbf{x} = \boldsymbol{\tau}(\mathbf{u}) \sim \pi \quad \text{for} \quad \mathbf{u} \sim \mathrm{Uniform}([0,1]^d).$$

在这种情形下， $\boldsymbol{\tau}$ 将均匀分布推前 到定义在 $[0,1]^d$ 上的目标分布 π，并称为传输映射。我们将密度 q 经映射 $\boldsymbol{\tau}$ 推前后的结果记作 $\boldsymbol{\tau}_{\#}q$。虽然只有在特殊情形下才存在闭式传输映射，但我们可以在灵活的函数类中近似 $\boldsymbol{\tau}$，使得诱导的密度

$$q_{\boldsymbol{\tau}}(\mathbf{x}) = (\boldsymbol{\tau}_{\#}\, q_u)(\mathbf{x}) \quad \text{with} \quad q_u(\mathbf{u}) = \mathbf{1}_{[0,1]^d}(\mathbf{u})$$

与$p(\mathbf{x})$非常接近。此类近似传输映射使得准蒙特卡罗和随机化准蒙特卡罗抽样能够直接应用于广泛的目标分布。若 $\boldsymbol{\tau}:(0,\ 1)^d \to \mathbb{R}^d$ 是一个C^1 微分同胚，则密度$q_{\boldsymbol{\tau}}$ 由变量变换公式给出：

$$q_{\boldsymbol{\tau}}(\mathbf{x}) = |\det J_{\boldsymbol{\tau}}(\boldsymbol{\tau}^{-1}(x))|^{-1} \tag{3}$$

其中J_τ是变换 $\boldsymbol{\tau}$ 的雅可比矩阵，定义为

$$J_\tau(\mathbf{u}) = \left(\frac{\partial \boldsymbol{\tau}_i}{\partial u_j}\right)_{1\le i,j\le d}.$$

因此，为了估计期望 $\mu = \mathbb{E}_{X\sim\pi}[f(X)],$，我们考虑寻找一个 C^1 微分同胚 $\boldsymbol{\tau}: (0,1)^d \to \mathbb{R}^d$，将均匀分布 $\mathcal{U}(0,1)^d$ 变换为提议分布 q_τ。如果我们有一个良好的传输映射 $\boldsymbol{\tau}$，使得 $q_\tau \approx \pi$，那么期望 $\mathbb{E}_\pi[f(X)]$ 可近似为 $\mathbb{E}_{\mathbf{U}\sim\mathcal{U}(0,1)^d}[f\circ\boldsymbol{\tau}(\mathbf{U})],$，其可通过以下方式估计

$$\frac{1}{n}\sum_{i=1}^{n}(f\circ\boldsymbol{\tau})(\mathbf{u}_i) \tag{4}$$

其中点集为 $\{\mathbf{u}_i\}_{i=1}^n \subset (0,1)^d$。

然而，如果$q_\tau \neq \pi$，估计量会产生偏差。我们可以通过重要性抽样校正这一偏差。注意，

$$E_{X\sim\pi}[f(X)] = E_{Y\sim q_\tau}[\frac{f\pi}{q_\tau}(Y)] = E_{\mathbf{U}\sim\mathcal{U}(0,1)^d}[\frac{f\pi}{q_\tau}\circ\boldsymbol{\tau}(\mathbf{U})].$$

然后利用 (3)，所关注的被积函数可写为

$$h(\mathbf{u}) \coloneqq \left(\frac{f\pi}{q_\tau}\circ\boldsymbol{\tau}\right)(\mathbf{u}) = (f\circ\boldsymbol{\tau})(\mathbf{u})\cdot(\pi\circ\boldsymbol{\tau})(\mathbf{u})\cdot\left|\det J_\tau(\mathbf{u})\right|.$$

因此，期望 μ 可估计为

$$\hat{I}_n \coloneqq \frac{1}{n}\sum_{i=1}^{n}h(\mathbf{u}_i) = \frac{1}{n}\sum_{i=1}^{n}(f\circ\boldsymbol{\tau})(\mathbf{u}_i)\cdot(\pi\circ\boldsymbol{\tau})(\mathbf{u}_i)\cdot\left|\det J_\tau(\mathbf{u}_i)\right| \tag{5}$$

点集为 $\{\mathbf{u}_i\}_{i=1}^n \subset (0,1)^d$。如果我们使用点 $\mathbf{u}_i \sim$均匀分布$([0,1]^d)$，例如独立同分布（i.i.d.）样本或随机化准蒙特卡罗点，那么 $\hat{}I_n$ 是 μ 的**无偏**估计量。

在第 4 节中，我们将证明，在对f、π 和 $\boldsymbol{\tau}$ 作出某些假设时，$h(\mathbf{u})$ 满足边界增长条件，从而确保使用扰乱网 $\{\mathbf{u}_i\}_{i=1}^n$ 的估计量 $\hat{}I_n$ 具有 $O(n^{-1+\epsilon})$ 均方根误差收敛率。

2.4. 流匹配与条件流匹配

令p_0 表示一个简单的先验分布（例如高斯分布或逻辑斯谛分布），令$p_1 = \pi$ 表示一个定义在$\mathbb{R}^d$上的复杂目标分布。基于流的生成建模旨在构造一个连续映射，或称流，将p_0 传输到p_1。为构造这样的流，我们考虑一族条件概率路径。给定一个数据样本z，对于$t \in [0,1]$，条件概率路径 $\rho(x, t \mid z)$ 是满足以下条件的概率流

$$\begin{cases}\rho(x,0\mid z) = p_0(x),\\ \rho(x,1\mid z) = \delta(x-z).\end{cases} \tag{6}$$

这样的条件概率路径可以由 $\mathbb{R}^d$ 上的流诱导。设 $v:\ [0,1]\times\mathbb{R}^d \to \mathbb{R}^d$ 是一个随时间变化的向量场，它通过以下常微分方程生成一个微分同胚映射 $\phi:\ [0,1]\times\mathbb{R}^d \to \mathbb{R}^d$

$$\begin{cases}\dfrac{d\phi_t(x|z)}{dt} = v(\phi_t(x|z), t\mid z),\\ \phi_0(x|z) = x.\end{cases} \tag{7}$$

直线流是速度场中特别简单且可行的一种选择

$$v(x,t\mid z) = \frac{z-x}{1-t}. \tag{8}$$

相应的条件概率路径满足连续性方程

$$\frac{\partial \rho(x,t \mid z)}{\partial t} + \nabla_x \cdot \left(\rho(x,t \mid z)\, v(x,t \mid z)\right) = 0, \tag{9}$$

以分布的意义。

To 得到一个将 p_0 传输至 p_1, 的流。将等式两边乘以 $p_1(z)$，并对z积分，得到

$$\int_{\mathbb{R}^d} p_1(z) \left(\frac{\partial \rho(x,t \mid z)}{\partial t} + \nabla_x \cdot \left(\rho(x,t \mid z) v(x,t \mid z)\right) \right) dz = 0, \qquad \forall x \in \mathbb{R}^d. \tag{10}$$

定义

$$\hat{\rho}(x,t) := \int_{\mathbb{R}^d} p_1(z)\, \rho(x,t \mid z)\, dz,$$

其满足

$$\begin{aligned} \hat{\rho}(x,0) &= \int_{\mathbb{R}^d} p_1(z)\, p_0(x)\, dz = p_0(x), \\ \hat{\rho}(x,1) &= \int_{\mathbb{R}^d} p_1(z)\, \delta(x-z)\, dz = p_1(x). \end{aligned} \tag{11}$$

因此，$\hat{\rho}(x,t)$定义了一条连接 p_0 与 p_1的概率路径。

由此得到的流满足连续性方程

$$\frac{\partial \hat{\rho}(x,t)}{\partial t} + \nabla_x \cdot \left(\hat{\rho}(x,t)\, v(x,t)\right) = 0, \tag{12}$$

其中，平均速度场u定义为

$$v(x,t) := \frac{\int_{\mathbb{R}^d} p_1(z)\, \rho(x,t \mid z)\, v(x,t \mid z)\, dz}{\hat{\rho}(x,t)}. \tag{13}$$

流匹配*Flow Matching* (FM) 的目标是通过最小化以下函数，直接学习速度场$v(x,\, t)$ 的神经网络近似$v(x,\, t;\, \theta)$

$$\mathcal{L}_{\mathrm{FM}}(\theta) := \mathbb{E}_{t\sim\mathcal{U}[0,1],\, x_t\sim\hat{\rho}(\cdot,t)} \left\| v(x_t,t;\theta) - u(x_t,t) \right\|_2^2 = \int_0^1 \int_{\mathbb{R}^d} \left\| v(x,t;\theta) - \frac{\int_{\mathbb{R}^d} p_1(z) \rho(x,t \mid z) v(x,t \mid z)\, dz}{\hat{\rho}(x,t)} \right\|_2^2 \hat{\rho}(x,t)\, dx\, dt. \tag{14}$$

然而，这种表述通常并不实用，因为计算平均速度场$v(x,t)$ 的计算开销通常很大。

为解决这一问题，可以交换期望与条件化的顺序，从而得到条件流匹配 (CFM) 目标

$$\mathcal{L}_{\mathrm{CFM}}(\theta) := \mathbb{E}_{t\sim\mathcal{U}[0,1],\, z\sim p_1,\, x_t\sim\rho(\cdot,t\mid z)} \left\| v(x_t,t;\theta) - v(x_t,t \mid z) \right\|_2^2 = \int_0^1 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \| v(x,t;\theta) - v(x,t \mid z) \|_2^2 \rho(x,t \mid z)\, p_1(z)\, dz\, dx\, dt. \tag{15}$$

可以证明，流匹配目标 (14) 与条件流匹配目标 (15) 的梯度相同[20]

$$\nabla_\theta \mathcal{L}_{\mathrm{FM}}(\theta) = \nabla_\theta \mathcal{L}_{\mathrm{CFM}}(\theta),$$

而后者在实践中则更易于处理。

当先验分布 p_0 为高斯分布时，条件路径$\rho(x,t \mid z)$ 的另一种常见且便于解析处理的选择是高斯族

$$\rho(x,t \mid z) = \mathcal{N}\left(x \,\middle|\, \mu_t(z),\, \sigma_t^2(z) I\right), \tag{16}$$

其中 $\mu_t(z)$ 和 $\sigma_t(z)$ 分别是随时间变化的均值和标准差。相应的条件流映射和速度场为

$$\begin{cases} \phi_t(x \mid z) &= \sigma_t(z)\, x + \mu_t(z), \\ v(x,t \mid z) &= \dfrac{\dot{\sigma}_t(z)}{\sigma_t(z)}\,(x - \mu_t(z)) + \dot{\mu}_t(z), \end{cases} \tag{17}$$

其中，上方的点表示对t求导。

高斯构造中一个尤为重要的情形是，均值和方差在端点之间线性插值。令

$$\mu_t(z) = t\, z, \qquad \sigma_t(z) = 1 - (1-\sigma)\, t, \tag{18}$$

对于某个给定的终端标准差 $\sigma \in [0,1]$。在这种情况下，粒子沿着$\mathbb{R}^d$中的直线轨迹运动，相应的流 $\phi_t(\cdot \mid z)$ 与初始高斯分布 $p_0 = \mathcal{N}(0, I)$ 和终端高斯分布 $p_1(\cdot \mid z) = \mathcal{N}(z, \sigma^2 I)$ 之间的 Wasserstein-2 最优传输位移插值一致，从而定义了 Wasserstein 度量下的一条测地线。然而，在$t = 1$处，得到的边缘密度变为 $\hat{\rho}(x,1) = p_1 * \mathcal{N}(\mathbf{0}, \sigma^2 I)$，也就是说，它是$p_1$的高斯平滑版本。取极限 $\sigma \to 0$ 即可恢复上文所述的直线流。关于速度场的其他变体和设计选择，请参阅 [21, 22, 23] 。

3. 使用流匹配构造准蒙特卡罗和随机化准蒙特卡罗的传输映射

本节说明基于流的生成模型如何构造适用于准蒙特卡罗和随机化准蒙特卡罗积分的传输映射。

流匹配的概率流常微分方程表述描述了一种确定性流，它将简单的先验分布p_0传输到 $\mathbb{R}^d$上的复杂目标分布p_1。令$\boldsymbol{v}$：$\mathbb{R}^d \times [0,1] \to \mathbb{R}^d$ 表示流匹配表述中的（平均）速度场（见 (13)）。相应的流 Ψ：$\mathbb{R}^d \times [0,1] \to \mathbb{R}^d$ 定义为下列初值问题的解

$$\partial_t \Psi(\mathbf{x}, t) = \boldsymbol{v}(\Psi(\mathbf{x}, t), t), \qquad \Psi(\mathbf{x}, 0) = \mathbf{x}, \quad \mathbf{x} \sim p_0. \tag{19}$$

在终止时刻，所诱导的传输映射

$$\tilde{\boldsymbol{\tau}}(\mathbf{x}) := \Psi(\mathbf{x}, 1)$$

将p_0 推送至p_1，，也就是说，$(\tilde{\boldsymbol{\tau}})_{\#} p_0 \approx p_1$。

为了将这一构造与单位立方体上的准蒙特卡罗和随机化准蒙特卡罗方法相结合，我们将学习得到的流与基变换G：$[0,1]^d \to \mathbb{R}^d$结合起来，该基变换将 $[0,1]^d$ 上的均匀分布映射为先验分布p_0。由此得到的复合映射

$$\boldsymbol{\tau} := \tilde{\boldsymbol{\tau}} \circ G : [0,1]^d \longrightarrow \mathbb{R}^d$$

是上一小节所考虑形式的传输映射：它将 $[0,1]^d$ 上的均匀分布推送至数据分布p_1。

我们还回顾一下，如何将流与沿轨迹演化的对数密度联系起来。在连续归一化流（CNF）文献中（例如参见 [23, 24, 25]），与速度场$\boldsymbol{v}$相关的连续性方程蕴含以下瞬时变量变换恒等式：

$$\frac{d}{dt} \log \hat{\rho}(\Psi(\mathbf{x}, t), t) = -\nabla_{\mathbf{x}} \cdot \boldsymbol{v}(\Psi(\mathbf{x}, t), t),$$

其中，$\hat{\rho}(\mathbf{x}, t)$ 表示 $\Psi(\cdot, t)_{\#} p_0$ 的密度。令 $l(t) = \log \hat{\rho}(\Psi(\mathbf{x}, t), t)$，状态 $\Psi(\mathbf{x}, t)$ 及其对数密度可通过增广常微分方程联合演化

$$\frac{d}{dt} \begin{bmatrix} \Psi(\mathbf{x}, t) \\ \ell(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}(\Psi(\mathbf{x}, t), t) \\ -\nabla_{\mathbf{x}} \cdot \boldsymbol{v}(\Psi(\mathbf{x}, t), t) \end{bmatrix}, \qquad \begin{bmatrix} \ell(0) \\ \Psi(\mathbf{x}, 0) \end{bmatrix} = \begin{bmatrix} \log p_0(\mathbf{x}) \\ \mathbf{x} \end{bmatrix} \tag{20}$$

从而 $\mathrm{l}(1) = \log p_1(\Psi(\mathbf{x}, 1))$。

虽然连续形式提供了理论基础，但实际实现需要进行两项关键近似：离散化时间积分，以及对速度场进行参数化。在实践中，真实的速度场$\boldsymbol{v}$ 未知，因此用参数为 θ的神经网络$\boldsymbol{v}_\theta(\mathbf{x},t)$来近似。因此，式（19）中的常微分方程被替换为

$$\partial_t \Psi_\theta(\mathbf{x}, t) = \boldsymbol{v}_\theta(\Psi_\theta(\mathbf{x}, t), t), \qquad \Psi_\theta(\mathbf{x}, 0) = \mathbf{x}. \tag{21}$$

为计算映射 $\tilde{\boldsymbol{\tau}}(\mathbf{x})$，我们必须数值求解常微分方程。我们考虑采用前向欧拉离散化，将时间划分为N 步，每步大小为$h = 1/N$。令$t_k = kh$。步映射 $\boldsymbol{\tau}^k\colon \mathbb{R}^d \to \mathbb{R}^d$ 定义为

$$\boldsymbol{\tau}^k(\mathbf{x}) = \mathbf{x} + h\boldsymbol{v}_\theta(\mathbf{x}, t_k).$$

因此，总传输映射 $\tilde{\boldsymbol{\tau}}$ 可近似为这些步映射的复合，具体如下：

$$\boldsymbol{\tau}^{N-1} \circ \cdots \circ \boldsymbol{\tau}^0(\mathbf{x}).$$

为将这一构造应用于单位立方体上的准蒙特卡罗和随机化准蒙特卡罗，我们将学习到的流与基变换G：$[0,1]^d \to \mathbb{R}^d$结合起来。该变换将 $[0,1]^d$ 上的均匀分布映射为先验分布p_0。一种选择是逐分量的逆逻辑函数（logit 函数），其定义为

$$G(u) = \log\left(\frac{u}{1-u}\right).$$

由于G 可微且单调递增，其逆函数对应于逻辑斯蒂分布的累积分布函数。因此，最终的复合传输映射为

$$\boldsymbol{\tau}(\mathbf{u}) := \boldsymbol{\tau}^{N-1} \circ \cdots \circ \boldsymbol{\tau}^0 \circ G(\mathbf{u}).$$

该映射 $\boldsymbol{\tau}\colon (0,1)^d \to \mathbb{R}^d$ 将 $[0,1]^d$ 上的均匀分布推送到数据分布 p_1。

通过流匹配构造传输映射 $\boldsymbol{\tau}$ 后，我们可以使用第 2 节介绍的内容，通过（随机化）准蒙特卡罗点估计期望。我们将所得方法称为流匹配重要性抽样准蒙特卡罗（FM-ISQMC）。下面我们将在第 4 节和第 5 节中进行详细的收敛阶分析。具体而言，第 4 节针对从 $(0,1)^d$ 到 $\mathbb{R}^d$ 的一般 $\boldsymbol{\tau}$ 提出假设并进行分析；第 5 节验证通过流匹配构造的$\boldsymbol{\tau}$ 满足第 4 节的假设，因此第 4 节的结论对于此类 $\boldsymbol{\tau}$成立。

4. 使用准蒙特卡罗的一般传输映射的收敛分析

本节将分析使用带有重要性抽样的一般传输映射 $\boldsymbol{\tau}\colon (0,1)^d \to \mathbb{R}^d$，通过随机化准蒙特卡罗估计量估计期望 $\mu = \mathbb{E}_{X\sim\pi}[f(X)]$ 时的收敛性。回顾第 2.3 小节，我们将所关注的被积函数定义为

$$h(\mathbf{u}) = (f \circ \boldsymbol{\tau})(\mathbf{u}) \cdot (\pi \circ \boldsymbol{\tau})(\mathbf{u}) \cdot |\det J_{\boldsymbol{\tau}}(\mathbf{u})|\,.$$

我们将证明，在对f、 π和 $\boldsymbol{\tau}$满足某些假设时，被积函数$h(\mathbf{u})$满足边界增长条件（假设1）。这将确保使用扰乱网时，随机化准蒙特卡罗估计量 $\hat{I}_n = \frac{1}{n}\sum_{i=1}^n h(\mathbf{u}_i)$ 对任意$\epsilon > 0$ 均达到$O(n^{-1+\epsilon})$ 的均方根误差收敛率。

4.1. 假设

我们的分析基于以下一组假设，涉及被积函数、目标密度以及传输映射导数的增长与衰减。

假设 4（被积函数 f 增长）。假设存在一个单变量累积分布函数 F（x），使得对于所有满足 $|\lambda| \le d$ 的多重指标$\lambda \in \mathbb{N}_0^d$ 以及任意小的 $B > 0$，都存在 $C_1 > 0$，使得对于所有 $\mathbf{x} \in \mathbb{R}^d$，以下界成立：

$$|\partial^\lambda f(\mathbf{x})| \le C_1 \prod_{k=1}^{d} [\min(F(x_k), 1 - F(x_k))]^{-B}$$

假设5（目标分布 π衰减）。假设 $\pi(\mathbf{x})$ 及其导数呈指数衰减。具体而言，存在常数 $C_2 > 0$ 和 $\alpha > 0$，使得对于所有满足 $|\lambda| \le d$ 的多重指标 $\lambda \in \mathbb{N}_0^d$，我们有

$$|\partial^\lambda \pi(\mathbf{x})| \le C_2 \prod_{k=1}^{d} e^{-\alpha |x_k|}$$

假设6 (传输增长)。令 $F(x)$ 为与假设 *4* 中相同的累积分布函数。假设对于任意 $v \subseteq \{1,\dots,d\}$、$m \in \{1,\dots,d\}$ 和 $j \in \{1,\dots,d\}$，以及任意小的 $B > 0$，存在 $C_3 > 0$，使得对于所有 $\mathbf{u} \in (0,1)^d$，

$$|\partial^v(\partial^{\{m\}}\boldsymbol{\tau}_j)(\mathbf{u})| \le C_3 \cdot \min(u_m, 1-u_m)^{-1} \cdot \prod_{k=1}^{d} \min(u_k, 1-u_k)^{-B-\mathbf{1}_{\{k\in v\}}} \tag{22}$$

此外，存在常数 $B_0 > 0, C_4 > 0$，使得对于任意 $k \in \{1,\dots,d\}$，

$$\min(F(\tau(\mathbf{u})_k), 1-F(\tau(\mathbf{u})_k))^{-1} \le C_4 \prod_{l=1}^{d} \min(u_l, 1-u_l)^{-B_0}. \tag{23}$$

另外，我们假设传输映射在边界附近增长得足够快。具体而言，存在常数 $C_5 > 1/\alpha$ 和常数 $C' \in \mathbb{R}$，使得

$$|\tau(\mathbf{u})_j| \ge C_5 |\ln(\min(u_j, 1-u_j))| - C'. \tag{24}$$

注7。假设*6*中的不等式 (22) 意味着

$$|\partial^v(\boldsymbol{\tau}_j)(\mathbf{u})| \le C_3 \cdot \prod_{k=1}^{d} \min(u_k, 1-u_k)^{-B-\mathbf{1}_{\{k\in v\}}}.$$

4.2. 被积函数各项的增长分析

我们将被积函数 $h(\mathbf{u}) = (f \circ \boldsymbol{\tau})(\mathbf{u}) \cdot (\pi \circ \boldsymbol{\tau})(\mathbf{u}) \cdot |\det J_\tau(\mathbf{u})|$ 分解，并对每一项偏导数的增长进行估计。

命题 8 ($f \circ \boldsymbol{\tau}$的增长)。在假设 *4* 和假设 *6* 下，对于任意小的 $B > 0$，存在常数 $C > 0$，使得 $f \circ \boldsymbol{\tau}$ 满足不等式

$$|\partial^v(f \circ \boldsymbol{\tau}(\mathbf{u}))| \le C \prod_{k=1}^{d} [\min(u_k, 1-u_k)]^{-B-\mathbf{1}_{\{k\in v\}}}.$$

证明。W 我们对 $\partial^v(f \circ \boldsymbol{\tau})$ 应用多元 Faà di Bruno 公式。展开中的一般项为 by

$$T = (\partial^\lambda f)(\boldsymbol{\tau}(\mathbf{u})) \cdot \prod_{l=1}^{|\lambda|} \partial^{v_l} \boldsymbol{\tau}_{j_l}(\mathbf{u})$$

其中 λ 是一个多重指标，$|\lambda|$等于划分中的块数，而 $\{v_l\}$ 划分 v。根据假设 4，在 $\mathbf{x} = \boldsymbol{\tau}(\mathbf{u})$ 处取值，可得

$$|(\partial^\lambda f)(\boldsymbol{\tau}(\mathbf{u}))| \le C_1 \prod_{k=1}^{d} [\min(F(\tau(\mathbf{u})_k), 1-F(\tau(\mathbf{u})_k))]^{-B}. \tag{25}$$

利用假设6中的价值增长条件，即

$$[\min(F(\tau(\mathbf{u})_k), 1-F(\tau(\mathbf{u})_k))]^{-1} \le C_4 \prod_{l=1}^{d} \min(u_l, 1-u_l)^{-B_0},$$

˜将此界代入式 (25)，可得存在一个常数$C > 0$ 使得

$$|(\partial^\lambda f)(\boldsymbol{\tau}(\mathbf{u}))| \le \tilde{C} \prod_{l=1}^{d} \min(u_l, 1-u_l)^{-dB_0 B}.$$

对于导数的乘积，假设6意味着

$$\left| \prod_{l=1}^{|\lambda|} \partial^{\nu_l} \boldsymbol{\tau}_{j_l} \right| \le C_3 \prod_{k=1}^{d} \min(u_k, 1-u_k)^{-1-B-\mathbf{1}_{\{k\in\nu\}}}.$$

将这些界相乘，得到

$$|T| \le C \prod_{k=1}^{d} \min(u_k, 1-u_k)^{-B_0-1-B-\mathbf{1}_{\{k\in\nu\}}}.$$

由于B_0和B可以任意小，它们的组合$B_0 + 1 + B$可写为 $1 + B'$，其中B'可以任意小。将B' 重新记为B后，指数可简化为 $-B - \mathbf{1}_{\{k\in\nu\}}$，这与所需形式一致。 □

命题 9 （$\pi \circ \boldsymbol{\tau}$ 的增长）。 在假设*5*和假设*6*下，对于任意小的 $B > 0$，存在常数 $C > 0$，使得 $\pi \circ \boldsymbol{\tau}$ 满足以下界

$$|\partial^\nu(\pi \circ \boldsymbol{\tau}(\mathbf{u}))| \le C \prod_{k=1}^{d} [\min(u_k, 1-u_k)]^{1-B-\mathbf{1}_{\{k\in\nu\}}}.$$

证明。 我们将多元 Faà di Bruno 公式应用于 $\partial^\nu(\pi \circ \boldsymbol{\tau})$。一个一般项为

$$T = (\partial^\lambda \pi)(\boldsymbol{\tau}(\mathbf{u})) \cdot \prod_{l=1}^{|\lambda|} \partial^{\nu_l} \boldsymbol{\tau}_{j_l}(\mathbf{u})$$

其中， λ 是一个多重指标， $|\lambda|$等于划分中的块数，而 $\{\nu_l\}$ 是ν的划分。根据假设5，第一个因子的上界为

$$|(\partial^\lambda \pi)(\boldsymbol{\tau}(\mathbf{u}))| \le C_2 \prod_{k=1}^{d} e^{-\alpha|\tau(\mathbf{u})_k|}.$$

利用假设6中的增长条件（24），可得 $|\tau(\mathbf{u})_k| \ge C_5 |\ln(\min(u_k, 1-u_k))| - C'$。因此，我们有

$$e^{-\alpha|\tau(\mathbf{u})_k|} \le e^{-\alpha(C_5|\ln(\min(u_k,1-u_k))|-C')} = e^{\alpha C'} \min(u_k, 1-u_k)^{\alpha C_5}.$$

代回后，我们得到

$$|(\partial^\lambda \pi)(\boldsymbol{\tau}(\mathbf{u}))| \le \tilde{C} \prod_{k=1}^{d} \min(u_k, 1-u_k)^{\alpha C_5}.$$

对于导数的乘积，假设6给出的上界为

$$\left| \prod_{l=1}^{|\lambda|} \partial^{\nu_l} \boldsymbol{\tau}_{j_l} \right| \le C_3 \prod_{k=1}^{d} \min(u_k, 1-u_k)^{-B-\mathbf{1}_{\{k\in\nu_l\}}}.$$

将这些上界相乘，可得一般项T 的上界为

$$|T| \le C \prod_{k=1}^{d} \min(u_k, 1-u_k)^{\alpha C_5 - B - \mathbf{1}_{\{k\in\nu\}}} \le C \prod_{k=1}^{d} \min(u_k, 1-u_k)^{\alpha 1 - B - \mathbf{1}_{\{k\in\nu\}}},$$

其中在最后一个不等式中，我们使用了 $\alpha C_5 > 1$ 和 $\min(u_k, 1-u_k) \le 1$。 □

命题 10（|行列式J_τ|的增长）。在假设6下，对于任意小的 $B > 0$，存在常数 $C > 0$，使得一般稠密雅可比矩阵的行列式满足以下增长界

$$|\partial^{v}(|\det J_{\boldsymbol{\tau}}(\mathbf{u})|)| \le C \prod_{k=1}^{d} [\min(u_k, 1-u_k)]^{-B-1-\mathbf{1}_{\{k\in v\}}} .$$

证明。由于 $\boldsymbol{\tau}$ 是一个微分同胚，$\det J_\tau(\mathbf{u})$在所有$\mathbf{u} \in (0,1)^d$处均可微且非零。由行列式的连续性可知，$\det J_\tau(\mathbf{u})$保持符号恒定。因此，我们有 $|\partial^{v}(|\det J_\tau(\mathbf{u})|) = |\partial^{v}(\det J_\tau(\mathbf{u}))|$。

一般雅可比矩阵$J_\tau(\mathbf{u})$的行列式由莱布尼茨公式给出

$$\det J_{\boldsymbol{\tau}}(\mathbf{u}) = \sum_{\sigma\in S_d} \operatorname{sgn}(\sigma) \underbrace{\prod_{i=1}^{d} \partial^{\{\sigma(i)\}}\boldsymbol{\tau}_i(\mathbf{u})}_{P_\sigma(\mathbf{u})}$$

其中S_d表示 $\{1,...、d\}$的所有置换构成的集合。根据三角不等式，$|\partial^{v}(\det J_\tau(\mathbf{u}))| \le \sum_{\sigma\in S_d} |\partial^{v}(P_\sigma(\mathbf{u}))|$。我们只需界定单个项 $\partial^{v}(P_\sigma)$的增长。

我们对 $\partial^{v}(P_\sigma)$应用一般莱布尼茨法则。该展开中的一个一般项T 具有如下形式

$$T = \prod_{i=1}^{d} \partial^{v_i}\left(\partial^{\{\sigma(i)\}}\boldsymbol{\tau}_i(\mathbf{u})\right)$$

其中，$\{v_1, \ldots, v_d\}$ 是v的一个划分，即 $\bigsqcup_{i=1}^{d} v_i = v$。

现在，我们对T中的每个d 因子应用假设6。对于第i个因子，我们令$j = i$、$m = \sigma(i)$，以及$v' = v_i$。由此得到以下界

$$\left|\partial^{v_i}(\partial^{\{\sigma(i)\}}\boldsymbol{\tau}_i)\right| \le C_3 \cdot \min(u_{\sigma(i)}, 1-u_{\sigma(i)})^{-1} \cdot \prod_{k=1}^{d} \min(u_k, 1-u_k)^{-B-\mathbf{1}_{\{k\in v_i\}}}$$

为了给 $|T|$取界，我们将这d 个界相乘，得到

$$\begin{aligned}
|T| &\le \prod_{i=1}^{d}\left[C_3 \cdot \min(u_{\sigma(i)}, 1-u_{\sigma(i)})^{-1} \cdot \left(\prod_{k=1}^{d} \min(u_k, 1-u_k)^{-B-\mathbf{1}_{\{k\in v_i\}}}\right)\right] \\
&\le C_3^d \prod_{k=1}^{d} \min(u_k, 1-u_k)^{-1-dB-\sum_{i=1}^{d}\mathbf{1}_{\{k\in v_i\}}} \\
&= C_3^d \prod_{k=1}^{d} \min(u_k, 1-u_k)^{-1-B'-\mathbf{1}_{\{k\in v\}}} \\
&\le C \prod_{k=1}^{d} \min(u_k, 1-u_k)^{-B'-1-\mathbf{1}_{\{k\in v\}}}
\end{aligned}$$

其中，在第二个不等式中，我们合并了每个 $\min(u_k, 1-u_k)$ 指数中的各项；在第三个等式中，我们利用了 $\{v_1, \ldots, v_d\}$ 是v的一个划分；在最后一个不等式中，我们令$C = C_3^d$ 且$B' = dB$。

由于B 可以任意小，$B' = dB$也可以任意小。因此，我们得到了所需的界。 □

4.3. 主收敛定理

现在，我们将这些界结合起来，以证明主要结果。

定理 11（被积函数h（$\mathbf{u}$）满足边界增长条件）。在假设 *4*、*5* 和 *6* 下，被积函数 h（$\mathbf{u}$）=（$f \circ \boldsymbol{\tau}$）（$\mathbf{u}$）·（$\pi \circ \boldsymbol{\tau}$）（$\mathbf{u}$）·|det J_τ（$\mathbf{u}$）| 满足假设 *1* 中所述的边界增长条件。

证明。 我们对$h(\mathbf{u})$应用一般莱布尼茨法则。展开式 $\partial^{v}h(\mathbf{u})$中的一般项$T$ 为

$$T = \partial^{v_1}((f \circ \boldsymbol{\tau})(\mathbf{u})) \cdot \partial^{v_2}((\pi \circ \boldsymbol{\tau})(\mathbf{u})) \cdot \partial^{v_3}(|\det J_{\boldsymbol{\tau}}(\mathbf{u})|)$$

其中 $\{v_1, v_2, v_3\}$是v的一个划分。我们将命题 8、9 和 10 中的界相乘，从而得到 $|T|$ 的界 ：

$$\begin{aligned}
|T| &\le \left(C \prod_{k=1}^{d} \min(u_k, 1-u_k)^{-B-\mathbf{1}_{\{k\in v_1\}}} \right) \\
&\quad \times \left(C \prod_{k=1}^{d} \min(u_k, 1-u_k)^{1-B-\mathbf{1}_{\{k\in v_2\}}} \right) \\
&\quad \times \left(C \prod_{k=1}^{d} \min(u_k, 1-u_k)^{-1-B-\mathbf{1}_{\{k\in v_3\}}} \right) \\
&= C^3 \prod_{k=1}^{d} \min(u_k, 1-u_k)^{-3B-(\mathbf{1}_{\{k\in v_1\}}+\mathbf{1}_{\{k\in v_2\}}+\mathbf{1}_{\{k\in v_3\}})} \\
&= \tilde{C} \prod_{k=1}^{d} \min(u_k, 1-u_k)^{-B'-\mathbf{1}_{\{k\in v\}}},
\end{aligned}$$

其中 $\tilde{C}$ 是一个常数，且$B' = 3B$。最后一个等式成立，是因为 $\{v_1,v_2,v_3\}$ 是v的一个划分，因此对于任意k，$\mathbf{1}_{\{k\in v_1\}}+\mathbf{1}_{\{k\in v_2\}}+\mathbf{1}_{\{k\in v_3\}} = \mathbf{1}_{\{k\in v\}}$。 这个最终界符合边界增长条件（假设1）所要求的形式，且常数B'可以任意小。由于莱布尼茨展开中 h(u) 的每一项都满足此界，函数 h(u) 本身也满足该条件。 □

推论 12. 根据定理 *11* 及扰乱网的性质，用于估计期望$\mu = \mathbb{E}_{X\sim\pi}[f(X)]$ 的随机化准蒙特卡罗估计量 $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} h(\mathbf{u}_i)$ 的均方根误差收敛率为 $O(n^{-1+\epsilon})$，其中任意 $\epsilon > 0$ 均成立。

5. 流匹配输运映射的增长条件验证

本节将对通过流匹配构造的传输映射给出详细的理论论证。我们证明它满足第 4 节中的假设 6。因此，当它与满足假设 4 的被积函数以及满足假设5的目标密度 π 结合时，所得复合函数满足边界增长条件，从而确保我们的 FM-ISQMC 估计量达到所需的收敛速率。

5.1. 设置与定义

我们首先回顾第 3 节中的设定。考虑传输映射 $\boldsymbol{\tau} : (0,1)^d \to \mathbb{R}^d$，它由基变换$G$ 与离散化流映射复合而成。

基变换。 基变换G 是从 $(0,1)^d$ 到 $\mathbb{R}^d$的双射。它对所有d 个分量应用同一个一元函数$G : (0,1) \to \mathbb{R}$。为简化记号，我们定义

$$G(\mathbf{u}) = (G(u_1), \ldots, G(u_d))^{\top}.$$

在本文中，我们选择G 为逆逻辑函数，也称为 logit 函数，其定义为 $G(u) = \log\left(\frac{u}{1-u}\right)$。由于$G$ 可微且单调递增，其反函数有良好定义，对应于支撑集为 $\mathbb{R}$ 的某个分布的累积分布函数。因此，我们记$G = F^{-1}$，其中F 是定义在 $\mathbb{R}$ 上的逻辑斯蒂分布的累积分布函数，给定为 $F(x) = (1+e^{-x})^{-1}$。G 的导数满足以下增长条件

$$|G^{(k)}(u)| \le C_k \cdot [\min(u, 1-u)]^{-k}, \quad \forall k \ge 1. \tag{26}$$

流匹配离散化。流匹配常微分方程为 $\dot{\mathbf{x}} = \boldsymbol{v}_\theta(\mathbf{x},t)$。我们考虑采用$N$个步长为 $h = 1/N$ 的前向欧拉离散化。令$t_k = kh$。步映射 $\boldsymbol{\tau}^k\colon \mathbb{R}^d \to \mathbb{R}^d$ 定义为 $\boldsymbol{\tau}^k(\mathbf{x}) = \mathbf{x}+h\boldsymbol{v}_\theta(\mathbf{x},t_k)$。总传输映射 $\boldsymbol{\tau}$ 是复合映射 $\boldsymbol{\tau}(\mathbf{u}) = \boldsymbol{\tau}^{N-1}\circ\cdots\circ\boldsymbol{\tau}^0\circ G(\mathbf{u})$。我们将基映射与前$k$个步骤的复合记为 $\boldsymbol{\tau}^{0:k} = \boldsymbol{\tau}^{k-1}\circ\cdots\circ\boldsymbol{\tau}^0\circ G$，其中 $\boldsymbol{\tau}^{0:0} = G$。

关于向量场的假设。我们假设所学习的向量场 $\boldsymbol{v}_\theta(\mathbf{x},t)$ 是光滑的，且具有分析所需各阶有界导数。

假设 13。对于任意多重指标 $\alpha \in \mathbb{N}_0^{d+1}$，包括 $\alpha = 0$的情形，都存在常数 M_α，使得

$$\sup_{\mathbf{x}\in\mathbb{R}^d,t\in[0,1]} |\partial^\alpha \boldsymbol{v}_\theta(\mathbf{x},t)| \le M_\alpha. \tag{27}$$

该假设要求向量场 $\boldsymbol{v}_\theta$ 本身及其导数均一致有界 d.

注释14。采用常见激活函数的标准多层感知机（*MLP*）可以满足假设*13*。考虑一个 *MLP* $f(\mathbf{x}) = W_L\sigma(W_{L-1}\dots\sigma(W_1\mathbf{x})\dots)$。如果激活函数 σ 有界且各阶导数均有界，例如 *Sigmoid* 或 *Tanh*，那么整个网络及其导数均有界，从而满足该假设。如果 σ 无界但导数有界，例如 *Softplus*、*SiLU* 或 *GELU*，则网络输出可能呈线性增长。然而，对于 $\alpha \ge 1$的导数，链式法则涉及 σ'，而其有界。因此，网络的导数有界。请注意，尽管假设*13*严格要求$\boldsymbol{v}_\theta$ 对于$\alpha = 0$有界，但对于下文的增长条件证明，我们主要依赖于 $\alpha \ge 1$的导数有界。$\boldsymbol{v}_\theta$ 本身有界有助于确保轨迹不会在有限时间内逃逸至无穷远，但奇异性分析的关键在于对导数的控制。

5.2. 增长条件的证明

我们首先通过证明以下定理，来验证假设6中关于值增长的第二部分。

定理15 (流匹配映射的值增长)。设 $\boldsymbol{\tau}(\mathbf{u}) = \tilde{\boldsymbol{\tau}}(G(\mathbf{u}))$ 为传输映射，其中 G 是逐分量逻辑斯谛分布逆映射，$\tilde{\boldsymbol{\tau}}$是由满足假设*13*的向量场$\boldsymbol{v}_\theta$ 生成的流映射。则以下性质成立。

1. 存在一个常数 $C_4 > 0$，使得对于所有 $j \in \{1,\dots,d\}$ 以及$\mathbf{u} \in (0,1)^d$，

$$\min((F\circ\boldsymbol{\tau})(\mathbf{u})_j, 1-(F\circ\boldsymbol{\tau})(\mathbf{u})_j)^{-1} \le C_4\min(u_j, 1-u_j)^{-1}. \tag{28}$$

这意味着假设*6*的不等式(23) 在参数 $B_0 = 1$下成立。

2. 该映射满足对数下界增长条件。具体而言，存在一个常数 C'，使得

$$|\boldsymbol{\tau}(\mathbf{u})_j| \ge |\ln(\min(u_j, 1-u_j))| - C'. \tag{29}$$

这意味着假设*6*的不等式(24) 在参数 $C_5 = 1$下成立。

证明。由于$\boldsymbol{v}_\theta$有界，因此存在$M > 0$，使得 $|\tilde{\boldsymbol{\tau}}(\mathbf{x}) - \mathbf{x}|_\infty \le M$。令$x_j = G(\mathbf{u})_j$。于是$x_j - M \le \boldsymbol{\tau}_j(\mathbf{u}) \le x_j + M$。

我们首先建立逆边界控制。利用F的单调性以及恒等式$e^{-x_j} = (1-u_j)/u_j$，可得

$$F(\boldsymbol{\tau}_j) \ge F(x_j - M) = \frac{1}{1+e^{-x_j}e^{M}} = \frac{u_j}{u_j + (1-u_j)e^{M}} \ge e^{-M}u_j.$$

类似地，对于上尾，我们得到

$$1 - F(\boldsymbol{\tau}_j) \ge 1 - F(x_j + M) = \frac{e^{-x_j}e^{-M}}{1+e^{-x_j}e^{-M}} = \frac{(1-u_j)e^{-M}}{u_j + (1-u_j)e^{-M}} \ge e^{-M}(1-u_j).$$

结合这些结果可得 $\min(F(\boldsymbol{\tau}_j),\ 1-F(\boldsymbol{\tau}_j)) \ge e^{-M}\min(u_j,\ 1-u_j)$，这意味着结论成立，其中 $C_4 = e^M$且 $B = 1$。

接下来，我们推导对数下界增长。我们有 $|\tau_j(\mathbf{u})| \geq |x_j| - M = |G(\mathbf{u})_j| - M$。回忆一下，$G(\mathbf{u})_j = \log\left(\frac{u_j}{1-u_j}\right)$。我们注意到 $\left|\log\left(\frac{u_j}{1-u_j}\right)\right| \geq |\ln(\min(u_j, 1-u_j))| - \ln 2$。为此，考虑 $u_j \leq 1/2$。此时 $\min(u_j, 1-u_j) = u_j$。$\log\left(\frac{u_j}{1-u_j}\right) = \ln u_j - \ln(1-u_j)$。由于 $1/2 \leq 1-u_j < 1$，我们有 $-\ln 2 \leq \ln(1-u_j) < 0$。因此 $\log\left(\frac{u_j}{1-u_j}\right) \in [\ln u_j, \ln u_j + \ln 2]$。由于 $\ln u_j$ 是负且绝对值很大，$\left|\log\left(\frac{u_j}{1-u_j}\right)\right| \geq |\ln u_j| - \ln 2$。$u_j > 1/2$ 的情况同理。因此，

$$|\tau_j(\mathbf{u})| \geq |\ln(\min(u_j, 1-u_j))| - (M + \ln 2).$$

这直接验证了该条件，其中 $C_5 = 1$ 和 $C' = M + \ln 2$。 □

接下来，我们证明 τ 的导数增长条件。

定理16（流匹配映射的导数增长）。 设 τ 是由基映射 G 与采用满足假设*13*的向量场 v_θ 的 N 个欧拉步复合而构造的传输映射。那么，对于任意 $m \in \{1, \ldots, d\}$ 以及任意指标集合 $\nu \subseteq \{1, \ldots, d\}$，对于任意小的 $B > 0$，存在常数 $C_3 > 0$，使得 τ 的偏导数满足不等式

$$\left|\partial^\nu(\partial^{\{m\}}\tau_j)(\mathbf{u})\right| \leq C_3 \cdot \min(u_m, 1-u_m)^{-1} \cdot \prod_{k=1}^{d} \min(u_k, 1-u_k)^{-B-\mathbf{1}_{\{k\in\nu\}}}. \tag{30}$$

证明。我们证明对于任意 $0 \leq k \leq N$，映射 $\tau^{0:k}$ 满足增长条件（30）。证明过程如下：
对k进行归纳法。

我们从基础情形 $k = 0$ 开始。令 $\tau^{0:0} = G$。我们检查 $\tau^{0:0}$ 是否满足该条件。由于 G 是对角矩阵，$(\tau^{0:0})_j(u) = G(u_j)$。导数 $\partial^{\{m\}}(\tau^{0:0})_j$ 仅当 $j = m$ 时非零，此时其值为 $G'(u_m)$。对于一般的多重指标 ν，项 $\partial^\nu(\partial^{\{m\}}(\tau^{0:0})_m)(u)$ 仅当 $\nu \subseteq \{m\}$ 时非零。如果 $\nu = \emptyset$，该项为 $G'(u_m)$，其上界为 $C_3 \min(u_m, 1-u_m)^{-1}$。如果 $\nu = \{m\}$，该项为 $G''(u_m)$，其上界为 $C_3 \min(u_m, 1-u_m)^{-2}$。两种情况下，界 $|\partial^\nu\partial^{\{m\}}(\tau^{0:0})_j| \leq C_3 \min(u_m, 1-u_m)^{-1} \prod_{k\in\nu} \min(u_k, 1-u_k)^{-1}$ 均成立，因此在 $B = 0$ 时满足该条件。

对于归纳步骤，假设映射 $\tau^{0:k}$ 满足增长条件 (30)。考虑下一个步映射 $\tau^k(x) = x + h v_\theta(x, t_k)$。我们要证明 $\tau^{0:k+1} = \tau^k \circ \tau^{0:k}$ 也满足该条件。第 j 个分量为 $(\tau^{0:k+1})_j(u) = (\tau^{0:k})_j(u) + h v_{\theta,j}(\tau^{0:k}(u), t_k)$。对 u_m 求导，然后应用 ∂^ν，得到 $\partial^\nu\partial^{\{m\}}(\tau^{0:k+1})_j = \partial^\nu\partial^{\{m\}}(\tau^{0:k})_j + h\partial^\nu\partial^{\{m\}}[v_{\theta,j}(\tau^{0:k}(u), t_k)]$。第一项 $\partial^\nu\partial^{\{m\}}(\tau^{0:k})_j$ 根据归纳假设满足相应的界。我们需要分析第二项。

我们分 析项 $\partial^\nu\partial^{\{m\}}(v_{\theta,j} \circ \tau^{0:k})$，首先对 $\partial^{\{m\}}$ 应用链式法则，然后对 ∂^ν 应用莱布尼茨法则。

$$\partial^{\{m\}}(v_{\theta,j} \circ \tau^{0:k}) = \sum_{p=1}^{d} (\partial^{\{p\}} v_{\theta,j}) \circ \tau^{0:k} \cdot \partial^{\{m\}}(\tau^{0:k})_p.$$

将 ∂^ν 应用于该乘积，可得一个关于 p 以及划分 $\nu_A \uplus \nu_B = \nu$ 的求和：

$$\partial^\nu\partial^{\{m\}}(v_{\theta,j} \circ \tau^{0:k}) = \sum_{p=1}^{d} \sum_{\nu_A \subseteq \nu} \partial^{\nu_A}[(\partial^{\{p\}} v_{\theta,j}) \circ \tau^{0:k}] \cdot \partial^{\nu_B}[\partial^{\{m\}}(\tau^{0:k})_p].$$

对于第一个因子 $\partial^{\nu_A}[(\partial^{\{p\}} v_{\theta,j}) \circ \tau^{0:k}]$，我们应用多元 Fa`a di Bruno 公式。由于 v_θ 及其导数有界，并利用归纳假设（将 ν_A 中的任意指标视为主指标 m，可得到一个上界 $\prod_{k\in\nu_A} \min(u_k, 1-u_k)^{-1-B}$），该因子有界于 $C \prod_{k\in\nu_A} \min(u_k, 1-u_k)^{-1-B}$。对于第二个因子 $\partial^{\nu_B}[\partial^{\{m\}}(\tau^{0:k})_p]$，归纳假设直接给出上界 $C \min(u_m, 1-u_m)^{-1} \prod_{k\in\nu_B} \min(u_k, 1-u_k)^{-1-B}$。将这些上界相乘，并利用 $\nu_A \uplus \nu_B = \nu$，我们得到

$$\left|\partial^\nu\partial^{\{m\}}(v_{\theta,j} \circ \tau^{0:k})\right| \leq \tilde{C} \min(u_m, 1-u_m)^{-1} \prod_{k\in\nu} \min(u_k, 1-u_k)^{-1-B}.$$

这符合所需的增长条件。由于 $v_{\theta,j}$ 及其导数均有界，因此不会改变增长率。因此，$\tau^{0:k+1}$ 满足与 $\tau^{0:k}$ 相同的增长条件。

通过归纳法，我们得出 $\tau^{0:k}$ 对所有 $0 \leq k \leq N$ 都满足增长条件。由于 $\tau = \tau^{0:N}$，最终的传输映射满足假设6。

□

5.3. FM-ISQMC 估计量的收敛分析

我们分析使用所构造的传输映射 $\boldsymbol{\tau}$ 进行数值积分的收敛性。目标是估计期望 $\mu = \mathbb{E}_{X\sim\pi}[f(X)]$。我们采用如下定义的 FM-ISQMC 估计量:

$$\hat{I}_n = \frac{1}{n}\sum_{i=1}^{n} h(\mathbf{u}_i) = \frac{1}{n}\sum_{i=1}^{n} (f\circ\boldsymbol{\tau})(\mathbf{u}_i)\cdot(\pi\circ\boldsymbol{\tau})(\mathbf{u}_i)\cdot|\det J_{\tau}(\mathbf{u}_i)|\,, \tag{31}$$

其中 $\{\mathbf{u}_i\}_{i=1}^{n}$ 是扰乱网点。

定理15（FM-ISQMC的收敛速率）。假设向量场 v_θ 满足假设*13*，被积函数 f 满足假设 *4*，且目标密度 π 满足假设*5*。那么，被积函数 h（$\mathbf{u}$）满足边界增长条件（假设*1*）。因此，对于任意 $\epsilon > 0$，*FM-ISQMC* 估计量对期望 $\mu = \mathbb{E}_{X\sim\pi}[f(X)]$的均方根误差收敛速率为 $O(n^{-1+\epsilon})$。

证明。首先，根据定理15和定理16，v_θ 满足假设13这一条件意味着传输映射 $\boldsymbol{\tau}$ 满足假设 6。接下来，由于 f 满足假设 4，π 满足假设5，且 $\boldsymbol{\tau}$ 满足假设 6，我们可以使用第 4 节中的定理 11。该定理保证重要性采样被积函数$h(\mathbf{u})$满足边界增长条件（假设1）。

最后，对满足边界增长条件的函数应用定理2关于扰乱网的结论，可得估计量 $\hat{}I_n$ 的均方根误差为$O(n^{-1+\epsilon})$。□

6. 实验结果

在数值实验中，我们考察基于流的传输映射与随机化准蒙特卡罗（RQMC）抽样相结合时的性能。整个过程中，我们使用 $[0,1)^d$ 中的扰乱 Sobol' 点集作为底层 RQMC 设计；通过逆累积分布函数，这些点被映射到 $\mathbb{R}^d$ 上简单的因子化逻辑斯蒂基分布 p_0，然后由拟合得到的流传输至目标分布 π。速度场 $v_\theta(x, t)$ 由一个残差全连接网络参数化，该网络结合了状态 x 的线性嵌入、时间 t 的傅里叶特征嵌入、若干带有层归一化和 Mish 激活函数的残差多层感知机模块，以及一个映射至 $\mathbb{R}^d$ 的最终线性投影。给定学习得到的传输映射 τ（从 t = 0 到 t = 1 对概率流常微分方程进行数值积分，并以 p_0 为初始分布得到），对于每个坐标 j = 1, . . . , d，我们估计一阶矩 $\mathbb{E}_{X\sim\pi}[X_j]$ 和二阶矩 $\mathbb{E}_{X\sim\pi}[X_j^2]$。令 $\{\xi^{(S)}\,\mathrm{i}\,\}^{N}\mathrm{i}{=}1$ 表示从 p_0 独立抽取的基分布样本（S =蒙特卡洛），或通过逻辑斯谛逆累积分布函数将扰乱 Sobol' 序列映射到 $\mathbb{R}^d$ 后得到的基分布样本（S = 准蒙特卡罗），并令 $x^{(S)}\,\mathrm{i} = \tau(\xi^{(S)}\,\mathrm{i}\,)$。流匹配蒙特卡洛和随机化准蒙特卡罗估计量（FM-MC 和 FM-QMC）定义如下:

$$\widehat{I}_N^{\text{FM-}S}(f) \;=\; \frac{1}{N}\sum_{i=1}^{N} f(x_i^{(S)}), \qquad S \in \{\mathrm{MC}, \mathrm{QMC}\},$$

而相应的重要性抽样变体（FM-ISMC 和 FM-ISQMC）则通过自归一化估计量，将提议分布q_τ 校正回目标分布

$$\widehat{I}_N^{\text{FM-IS}S}(f) \;=\; \frac{\sum_{i=1}^{N} w_i^{(S)} f(x_i^{(S)})}{\sum_{i=1}^{N} w_i^{(S)}}, \qquad w_i^{(S)} \;=\; \frac{\pi(x_i^{(S)})}{q_\tau(x_i^{(S)})}, \quad S \in \{\mathrm{MC}, \mathrm{QMC}\}.$$

我们在三种具有代表性的后验分布上评估这四种估计量对各坐标一阶和二阶矩的估计准确性：二维高斯混合分布、二维香蕉形分布和30维高斯混合分布。我们的代码可在 https://github.com/yl602019618/QMC_flow_matching 获取

6.1. 高斯混合

6.1.1. 二维情形

我们首先考虑二维高斯分布混合，将其作为一个简单但并非平凡的多模态目标。目标密度 π 是一个具有四个分量且权重相等的高斯混合，

$$\pi(x) \;=\; \frac{1}{4}\sum_{k=1}^{4} \mathcal{N}(x\,|\,\mu_k,\Sigma_k), \qquad x\in\mathbb{R}^2,$$

其中，分量均值（写作列向量）为

$$\mu_1=(1,1)^\top,\quad \mu_2=(2,3.6)^\top,\quad \mu_3=(3.3,2.8)^\top,\quad \mu_4=(1.1,2.9)^\top,$$

相应的协方差矩阵由统一缩放 $1/4^2$, 得到

$$\Sigma_1=\frac{1}{40^2}\begin{pmatrix}2 & 0.6\\ 0.6 & 1\end{pmatrix},\quad \Sigma_2=\frac{1}{4^2}\begin{pmatrix}2 & -0.4\\ -0.4 & 2\end{pmatrix},\quad \Sigma_3=\frac{1}{4^2}\begin{pmatrix}3 & 0.8\\ 0.8 & 2\end{pmatrix},\quad \Sigma_4=\frac{1}{4^2}\begin{pmatrix}3 & 0\\ 0 & 0.5\end{pmatrix}.$$

图 1 总结了所学习传输映射的质量。如图 1(a)–(b) 所示，学习到的传输映射准确地恢复了目标分布的多模态结构。图 1(c)–(d) 比较了真实混合分布密度与流匹配模型估计的密度。在关注区域内，估计密度与解析高斯混合高度吻合，表明该模型在样本空间和底层密度层面均能忠实地近似 π 。

接下来，我们考察基于流的传输和 RQMC 抽样对矩估计准确性的影响。使用第 6 节定义的四种估计量（FM-MC、FM-QMC、FM-ISMC 和 FM-ISQMC），我们估计二维混合分布的一阶矩和二阶矩，并针对每种方法和每个样本量 N，计算 10 次独立重复实验的均方根误差（RMSE）。两种基于蒙特卡罗的方法（FM-MC 和 FM-ISMC）呈现出预期的 $O(N^{-1/2})$ 收敛速度。FM-QMC 估计量起初呈现出近乎 $O(N^{-1})$ 的收敛速度，这与 RQMC 带来的更强方差缩减效果一致；但随着 N 增大，其误差曲线最终趋于饱和，反映出学习得到的传输映射中的残余近似误差占据主导地位。相比之下，FM-ISQMC 的收敛速度显著高于基于蒙特卡罗的方案，并且在所考察的整个样本量范围内，对数均方根误差持续稳步下降。这表明，在此情形下，将基于流的重要性抽样与 RQMC 相结合，能够有效利用方差缩减和偏差校正的双重优势。

6.1.2. 三十维情形

接下来，我们考察一个更高维的例子，即由四个权重相等的分量构成的三十维高斯分布混合。目标密度 π 为

$$\pi(x) \;=\; \frac{1}{4}\sum_{k=1}^{4} \mathcal{N}(x\,|\,m_k,\Sigma), \qquad x\in\mathbb{R}^{30},$$

其中，分量均值 $m_k\in\mathbb{R}^{30}$ 定义为

$$m_1=(-2,-2,0,\dots,0)^\top,\quad m_2=(2,-2,0,\dots,0)^\top,\quad m_3=(-2,2,0,\dots,0)^\top,\quad m_4=(2,2,0,\dots,0)^\top,$$

所有分量共用同一个对角协方差矩阵

$$\Sigma=0.5\,I_{30}.$$

因此，多模态性集中在前两个坐标上，其余 28 个坐标则是相互独立且边缘方差相同的高斯方向。流匹配模型使用来自该混合分布的独立同分布样本进行训练，采用与二维情形相同的架构，但输入维度为 $d = 30$ 且

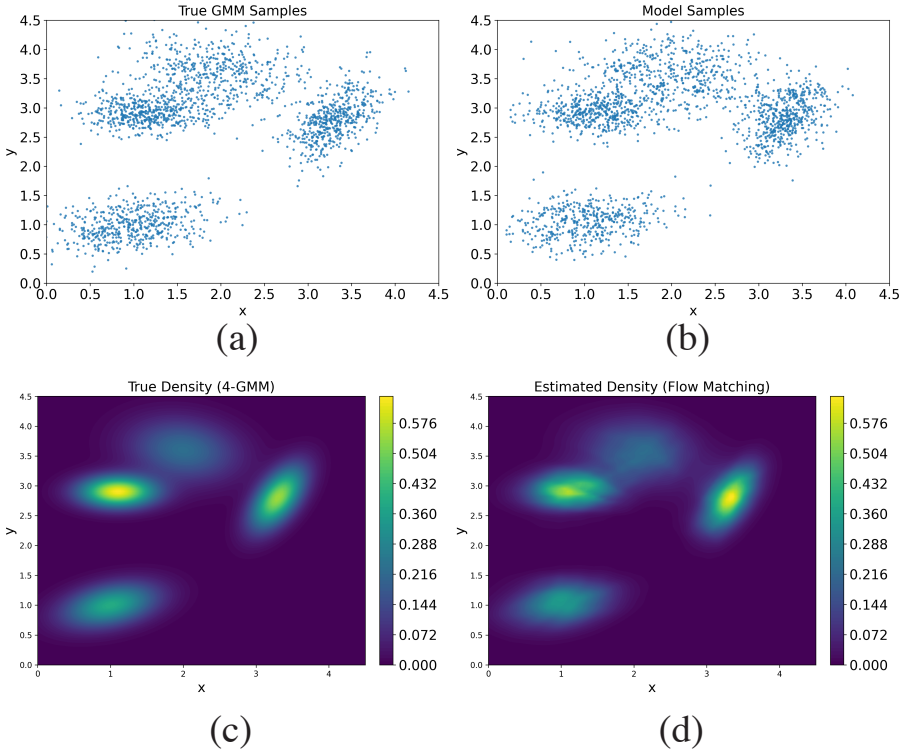


图1。（a）二维高斯混合分布的参考样本。（b）流匹配模型生成的样本。（c）二维高斯混合分布的真实密度。（d）流匹配模型估计的密度。

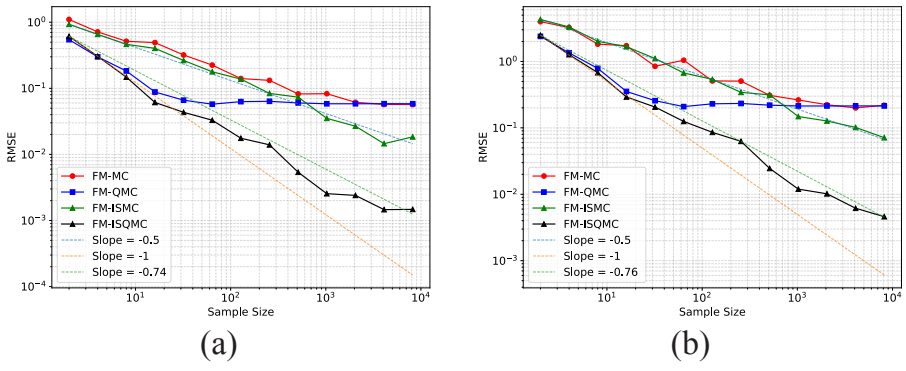


图2。估计二维高斯混合分布的（a）一阶矩和（b）二阶矩时的对数均方根误差。每个均方根误差均基于10次独立重复计算。

稍深一些的残差网络；同样，采用位置参数为零、尺度参数为单位值的独立逻辑斯谛分布作为基准测度，并使用Heun格式对轨迹进行积分，积分区间为$t = 0$至$t = 1$。

由于无法直接可视化30维数据，我们通过主成分分析（PCA）评估所学习传输的生成质量。图3(a)展示了从直接取自30维混合分布的参考样本计算得到的前两个主成分散点图，而图3(b)则显示了流匹配模型生成的样本在同一PCA基底上的投影。两组点云几乎无法区分，其中四个簇清晰分离，并与投影后的分量均值相对应，这表明所学习的传输映射捕获了高维目标分布的主要低维结构，并在线性投影下保留了其多模态几何结构。

我们还研究了在这一更高维设置中，基于流的传输和随机化准蒙特卡罗采样对矩估计的影响。与二维示例一样，我们采用第6节中的四种估计量（FM-MC、FM-QMC、FM-ISMC和FM-ISQMC），估计30维混合分布的一阶矩和二阶矩，并针对每种方法和每个样本量N，计算10次独立重复实验的均方根误差。所得对数均方根误差曲线见图4。两种基于蒙特卡洛的方案（FM-MC和FM-ISMC）再次呈现出典型的$O(N^{-1/2})$衰减。对于中等样本量，流匹配准蒙特卡罗估计量的误差下降明显更快，这反映了扰乱后的Sobol'设计带来的方差缩减；但随着N增大，其收敛速率最终趋于平缓，这与残余近似误差对所学习的高维传输的影响日益增大相一致。相比之下，在整个样本范围内，FM-ISQMC的均方根误差始终明显低于基于蒙特卡洛的方法，并且在对数—对数图中呈现出显著更陡的下降趋势。这表明，即使在这个30维多模态示例中，结合基于流的重要性抽样与随机化准蒙特卡罗，仍能有效提高精度。

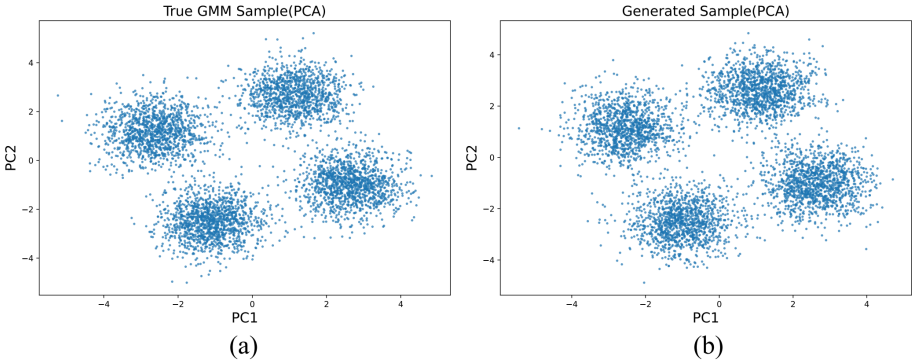


图3。（a）来自 30 维高斯混合分布的样本的前两个主成分散点图。（b）流匹配模型生成的样本的前两个主成分散点图。

6.2. 香蕉形分布

现在我们考虑一种二维香蕉形分布，它提供了一个非线性且强烈非高斯的基准分布。目标密度 π 定义为标准二元高斯分布经非线性变换后的推前

$$z = (z_1, z_2)^\top \sim \mathcal{N}(0, I_2), \qquad x_1 = z_1, \qquad x_2 = az_1^2 + c + bz_2,$$

其参数为 $a = 0.3$、$b = 1/\sqrt{2}$ 和 $c = -1$。等价地，$x = (x_1, x_2)^\top$ 的密度可写为闭式表达式

$$\pi(x_1, x_2) = \frac{1}{|b|}\,\varphi(x_1)\,\varphi\!\left(\frac{x_2 - ax_1^2 - c}{b}\right),$$

其中，φ 表示 $\mathbb{R}$上的标准正态密度。该构造产生了一种弯曲的香蕉形分布，在半径上呈单峰，但x_1 与x_2之间存在很强的非线性依赖。

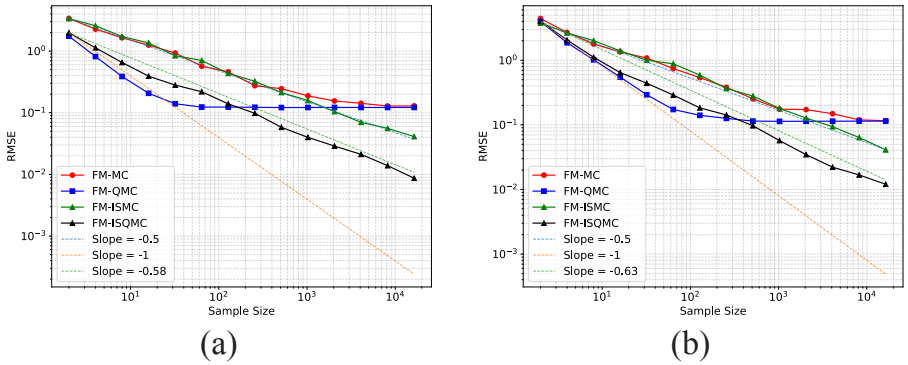


图 4. 估计 30 维高斯混合分布的 (a) 一阶矩和 (b) 二阶矩时的对数均方根误差。每个均方根误差均基于 10 次独立重复计算。

图5评估了针对这一非线性目标分布所学习到的传输的质量。如图5(a)–(b)所示，流匹配模型生成的样本很好地再现了真实分布特有的香蕉形支撑集，包括其弯曲形状和非高斯尾部。图5(c)–(d)比较了解析香蕉分布密度与根据学习到的流估算出的密度。两个等高线图在关注区域内几乎重合，表明该模型无论在样本层面还是在底层密度方面，都能忠实地近似 π 。

接下来，我们使用学得的流，研究基于流的传输和随机化准蒙特卡罗采样对该香蕉形目标分布矩估计精度的影响。图 6 中的对数均方根误差曲线显示，两种基于蒙特卡罗的方法（FM-MC 和 FM-ISMC）呈现出预期的$O(N^{-1/2})$ 收敛速度。FM-QMC 估计量起初达到明显更高的收敛速度，反映了随机化准蒙特卡罗带来的方差缩减；然而，随着N 增大，其误差最终趋于饱和，这是由于学得的传输映射存在残余近似误差，从而引入了不可忽略的偏差。相比之下，在所考察的整个样本量范围内，FM-ISQMC 的收敛速度始终显著优于$O(N^{-1/2})$，对数均方根误差持续稳步下降。这一现象表明，将基于流的重要性抽样与随机化准蒙特卡罗相结合，能够有效利用方差缩减和偏差校正，从而实现高效的矩估计。

7. 结论

本研究探讨了如何将流匹配模型与随机化准蒙特卡罗（RQMC）方法相结合，为高维积分构造高效的传输映射。在一般的重要性抽样框架下，我们分析了基于从单位立方体到提议分布密度q_τ 的微分同胚传输映射 τ的估计量，并证明，在被积函数、目标密度以及 τ的导数满足适当的增长与正则性假设时，所得的 RQMC 估计量满足边界增长条件。因此，对于任意 $\varepsilon > 0$，相应的 RQMC 重要性抽样估计量的均方根误差（RMSE）可达到近乎最优的收敛速率$O(N^{-1+\varepsilon})$。我们进一步验证了，当学习到的速度场具有一致有界且阶数足够高的导数时，采用逻辑斯蒂基分布和概率流常微分方程的欧拉离散化进行流匹配所得到的传输映射满足这些假设。

我们通过一系列针对具有挑战性的目标分布的数值实验，补充了理论分析。这些分布包括二维高斯混合、二维香蕉形分布和三十维高斯混合。实验一致表明，直接使用流匹配的蒙特卡洛和 RQMC 估计量（FM–MC 和 FM–QMC）会受到学习所得传输映射的近似偏差限制：蒙特卡洛变体呈现预期的$O(N^{-1/2})$收敛，而 FM–QMC 起初受益于 RQMC 更强的方差缩减，但随着N 增大，最终会趋于饱和。相比之下，我们的 FM–ISQMC 达到显著优于$O(N^{-1/2})$的收敛速率，并且在所考察的整个样本量范围内，对数均方根误差持续下降。这表明，在实践中，将流匹配模型与基于重要性抽样的 RQMC 相结合，能够有效兼顾方差缩减和偏差校正。

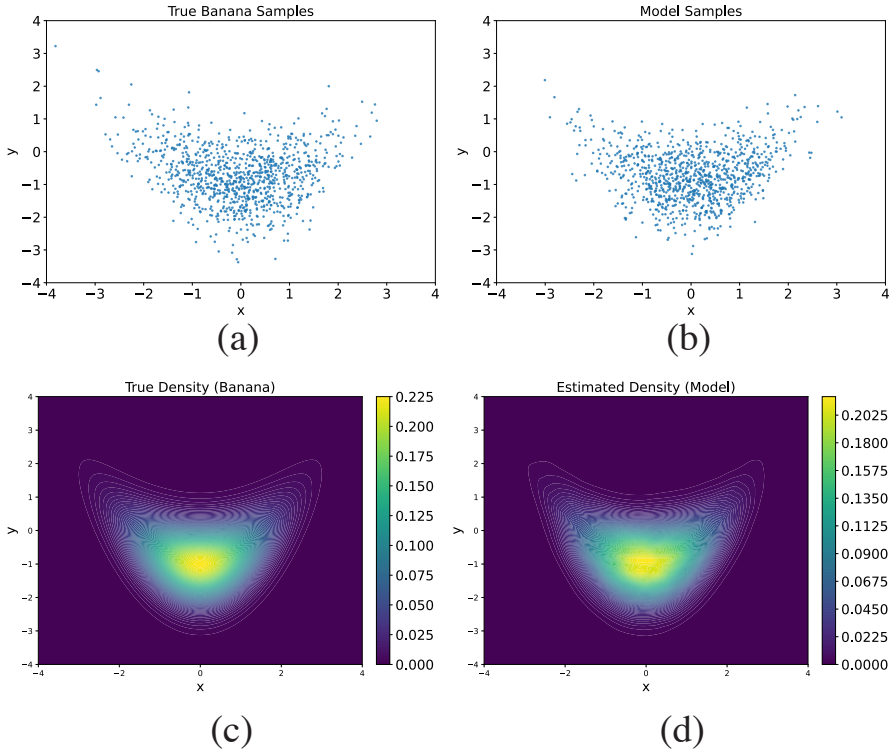


图5. (a) 二维香蕉分布的参考样本。(b) 流匹配模型生成的样本。(c) 二维香蕉分布的真实密度。(d) 流匹配模型估计的密度。

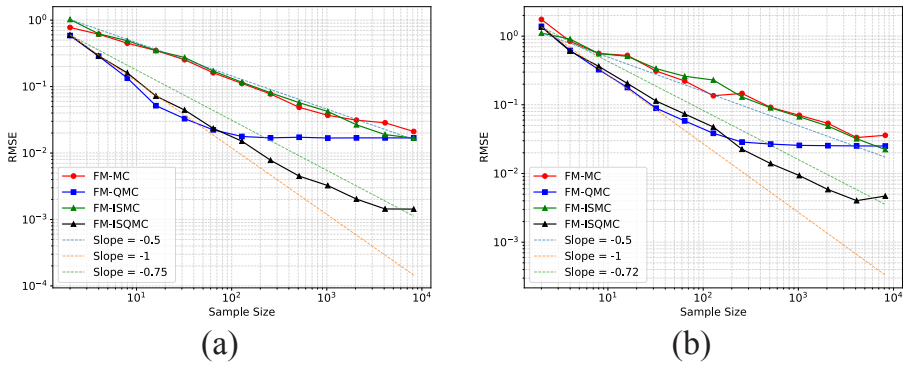


图6. 估计二维香蕉分布 (a) 一阶矩和 (b) 二阶矩的对数均方根误差。每个均方根误差均基于10次独立重复计算。

我们的理论分析假设采用固定步长欧拉离散化，并且速度场有界且界限一致。实践中通常使用自适应常微分方程求解器，而最优速度场可能会在边界处出现奇点[10]。弥合理论与实践之间的这一差距，是未来研究的一个重要方向。此外，更严格地追踪导数增长界中的常数，尤其是这些常数对维数和步数的依赖关系，将进一步巩固理论基础。我们将在未来的工作中考察这些方面。

参考文献

[1] H. Niederreiter，《随机数生成与拟蒙特卡罗方法》，第63卷，工业与应用数学学会，1992年。[2] A. B. Owen，随机置换的(t,m,s)-网与(t,s)-序列，载于：H. Niederreiter、P. J.-S. Shiue（编），《科学计算中的蒙特卡罗与拟蒙特卡罗方法》，施普林格纽约，纽约州纽约市，1995年，第299–317页。[3] B. Moskowitz，R. E. Caflisch，拟蒙特卡罗方法中的光滑性与降维，《数学与计算机建模》23（8-9）（1996年）37–54。[4] Y. Song，J. Sohl-Dickstein，D. P. Kingma，A. Kumar，S. Ermon，B. Poole，通过随机微分方程进行基于评分的生成建模，arXiv预印本 arXiv:2011.13456。[5] X. Liu，C. Gong，Q. Liu，流动又直又快：利用校正流学习生成和传输数据，arXiv预印本 arXiv:2209.03003。[6] C. Andral，结合正规化流与准蒙特卡罗，arXiv预印本 arXiv:2401.05934。[7] S. Liu，传输拟蒙特卡罗，arXiv预印本 arXiv:2412.16416。[8] J. Du，Z. He，《style id='9'>l$_p$-误差率：用于无界被积函数的随机化拟蒙特卡罗自归一化重要性抽样，arXiv预印本 arXiv:2511.10599。[9] A. B. Owen，哈尔顿序列避开原点，《SIAM评论》48（3）（2006年）487–503。[10] Y. Gao，J. Huang，Y. Jiao，S. Zheng，用于学习概率分布的连续归一化流的收敛性，arXiv预印本 arXiv:2404.00551。[11] J. F. Koksma，模1均匀分布理论中的一个一般定理，《Mathematica B》（聚特芬）11（1942/1943年）7–11。[12] E. Hlawka，均匀分布理论中的有界变差函数，《纯粹与应用数学年刊》54（1）（1961年）324–334。¨H. Niederreiter，《随机数生成与拟蒙特卡罗方法》，SIAM，宾夕法尼亚州费城，1992年。[13] A. B. Owen，光滑函数积分的扰乱网方差，《统计年刊》25（4）（1997年）1541–1562。[14] A. B. Owen，使用扰乱网的局部反变量抽样，《统计年刊》36（5）（2008年）2319–2343。[15] J. Dick，F. Pillichshammer，《数字序列、差异度与拟蒙特卡罗积分》，剑桥大学出版社，剑桥，2010年。[16] J. Dick，F. Y. Kuo，I. H. Sloan，高维积分：拟蒙特卡罗方法，《数值学报》22（2013年）133–288。[17] P. L’Ecuyer，C. Lemieux，随机化拟蒙特卡罗方法综述，载于：《不确定性建模：随机理论、方法与应用考察》，克吕沃学术出版社，2002年，第419–474页。[18] G. Constantine，T. Savits，带应用的多元 Faà di Bruno 公式，《美国数学会会刊》348（2）（1996年）503–520。[19] Y. Lipman，R. T. Chen，H. Ben-Hamu，M. Nickel，M. Le，生成建模的流匹配，载于：第十一届国际学习表征会议。[20] A. Tong，K. FATRAS，N. Malkin，G. Huguet，Y. Zhang，J. Rector-Brooks，G. Wolf，Y. Bengio，利用小批量最优传输改进并推广基于流的生成模型，《机器学习研究汇刊》。[21] N. Kornilov，P. Mokrov，A. Gasnikov，A. Korotin，一步学习直线轨迹的最优流匹配，《神经信息处理系统进展》37（2024年）104180–104204。[22] R. T. Chen，Y. Rubanova，J. Bettencourt，D. K. Duvenaud，神经常微分方程，《神经信息处理系统进展》31。[23] J. Behrmann，W. Grathwohl，R. T. Chen，D. Duvenaud，J.-H. Jacobsen，可逆残差网络，载于：国际机器学习会议，PMLR，2019年，第573–582页。[24] H. Ben-Hamu，S. Cohen，J. Bose，B. Amos，M. Nickel，A. Grover，R. T. Chen，Y. Lipman，流形上的归一化流与概率路径匹配，载于：国际机器学习会议，PMLR，2022年，第1749–1763页。