arXiv:2603.08945v1 [math.ST] 9 Mar 2026

基于通用最不利子模型的核去偏插件估计

Haiyi Chen[1], *, Yang Liu[2], *, Ivana Malenica[1], **[1]生物统计学系，北卡罗来纳大学教堂山分校，美国[2]统计学与运筹学系，北卡罗来纳大学教堂山分校，美国

Abstract. 我们提出ULFS-KDPE，一种基于通用最不利子模型的核去偏插件估计量，用于估计非参数模型中的路径可微参数。该方法在再生核希尔伯特空间（RKHS）中构造数据自适应去偏流，从而得到一个插件估计量，使其在无需显式推导或评估有效影响函数的情况下达到半参数效率。我们通过将通用最不利更新表述为概率密度上的非线性常微分方程，为ULFS-KDPE奠定严格的泛函分析基础。我们建立了经验得分沿诱导流的存在性、唯一性、稳定性与有限时间收敛。在标准正则性条件下，所得估计量是正则且渐近线性的，并同时对一大类路径可微参数达到半参数效率界。该方法可采用基于有限维核表示和有原则的停止准则的计算可行实现。在有限样本中，求解大量得分方程，并结合基于RKHS的平滑以及避免直接进行影响函数评估，可提高数值稳定性。模拟研究阐释了该方法，并支持理论结果。

1. Introduction

半参数效率理论为构造估计量提供了一个有原则的框架，使其在丰富的非参数统计模型中达到最优渐近方差。该理论的核心是路径可微性的概念及其对应的典范梯度，或有效影响函数（EIF）；它刻画了正则估计量的一阶行为以及半参数效率界限 [3, 29, 20]。因果推断中广泛使用的典范估计量，例如一步估计量和目标最大似然估计量（TMLE），利用这一结构，确保估计量具有渐近线性展开，且其影响函数（IF）等于有效影响函数（EIF） [3, 25, 27, 28]。这些方法在因果推断和缺失数据问题中取得了巨大成功，但通常需要明确掌握有效影响函数（EIF）的解析形式，而且往往针对单一目标参数量身定制 [27, 28]。

目标最大似然估计在这方面的文献中尤为突出。它通过沿着某个数据生成分布的初始估计进行更新，构造一个插件估计量，更新过程沿着

*第一作者
**通讯作者： imalenic@unc.edu

局部最不利参数子模型（LLFS）进行，直至有效影响函数估计方程近似得到求解 [25]。在适当的正则性条件下，由此得到的估计量符合统计模型，并实现半参数效率。近来，Cho 等人提出了核去偏插件估计量（KDPE），从不同角度逼近效率 [9]。KDPE 将去偏问题嵌入再生核希尔伯特空间（RKHS），并构造数据自适应扰动，在无需显式表达有效影响函数的情况下，近似求解经验得分方程。KDPE 利用再生核希尔伯特空间几何和有限维核表示，为复杂模型中的有效估计提供了一种灵活且计算上易于处理的途径。

然而，典范目标最大似然估计和 KDPE 都依赖于局部的最不利性概念；所构造的更新仅在当前分布的无穷小邻域内达到最优。这促使我们考虑能够全局实现最不利性的分布路径。通用最不利子模型（ULFS）定义了一条分布路径，其得分函数在路径上的每一点都与有效影响函数一致，而不仅仅是在路径起点处一致 [23]。从几何角度看，这样的路径沿着目标参数单位信息变化量最大的方向，在整个轨迹中保持前进。因此，沿通用最不利路径进行定向，可以一步解出有效影响函数估计方程，同时避免不必要的似然扰动。不过，仍然需要推导和评估有效影响函数 [23]。

本文提出一种新估计量，将这些思路融为一体——*ULFS–KDPE*，亦称*一步 KDPE*——结合通用最不利路径的全局最优性，以及基于再生核希尔伯特空间的去偏方法在计算和理论上的优势。我们构造了通用最不利路径的核限制替代路径，将其定义为密度上的非线性常微分方程的解，其速度场由经验矩偏差的再生核希尔伯特空间值 Riesz 表示元给出。重要的是，这种构造不需要显式知晓任何目标参数的有效影响函数。相反，它产生一个单一的数据自适应流，可同时对所有典范梯度属于再生核希尔伯特空间 $L^2(P_0)$-闭包的路径可微参数进行去偏。

由此得到的估计量是一种插件估计量，通过在通用最不利流上的最终分布处计算目标参数而获得。按照构造，该流会单调提高经验对数似然，并在相关经验得分方程近似满足时终止，从而避免迭代局部定向所伴随的收敛病态。该路径的全局最不利性确保在将似然扰动降至最低的同时实现所需的偏差降低，因此与迭代目标最大似然估计和局部靶向 KDPE 相比，有限样本稳定性更高；在重叠不足的情形下尤其如此。

从理论角度看，我们为 ULFS–KDPE 奠定了严谨的泛函分析基础。我们将通用最不利更新表述为一个取值于密度空间的常微分方程，并证明其解在适当的 H¨older 空间中存在、唯一且稳定，同时证明算法保持归一化和正性，并能在有限时间内收敛。从统计角度看，我们证明，对于满足标准余项条件的所有路径可微目标参数，由此得到的插件估计量都是正则、渐近线性的，并且在半参数意义下有效。值得注意的是，这种效率可以同时实现

于多个参数，而无需修改算法或指定特定参数的有效影响函数。

最后，我们证明 ULFS–KDPE 在计算上切实可行。尽管其底层构造是无限维的，但每次更新都可表示为有限维形式，且只涉及观测数据点处的核函数取值。由此得到的算法类似于再生核希尔伯特空间中的稳定化梯度流，并配有与经验得分方程直接相关、具有理论依据的停止准则；在模拟研究中，该算法还表现出良好的有限样本表现，包括在迭代靶向方法出现不稳定性的挑战性情形下。

1.1. 贡献。 本文的贡献如下。首先，我们提出*ULFS–KDPE*，这是一种基于核的一步去偏插件估计量，可在再生核希尔伯特空间内实现通用最不利路径。该方法生成单一的数据自适应分布流，无需显式计算有效影响函数，即可同时对一大类路径可微目标参数进行去偏。其次，我们将通用最不利更新表述为概率密度上的非线性常微分方程。我们在适当的 H¨older 空间中证明了其解的存在性、唯一性、稳定性和有限时间收敛，从而为所提出的估计量奠定严格的泛函分析基础。第三，我们证明，在标准条件下，所得估计量是正则的、渐近线性的，并且具有半参数有效性。重要的是，对于所有典范梯度属于 RKHS 的 $L^2(P_0)$-闭包的目标参数，均可同时实现有效性，其中包括多变量目标。最后，我们开发了一种计算上易于处理的实现方法，该方法基于有限维核表示，并采用源自通用最不利流几何结构、具有理论依据的停止准则。

1.2. 相关工作。 我们的工作与两大类半参数去偏方法相关：基于影响函数的方法和无影响函数计算方法。

基于影响函数的经典方法包括估计方程、一步估计量、双重机器学习（DML）和目标最大似然估计（TMLE）。估计方程方法通过求解由有效影响函数导出的得分方程来识别目标参数 [17, 14, 3, 24, 20]。一步估计量通过加入有效影响函数的经验均值来校正初始估计量，从而在适当的正则性条件下获得渐近效率 [3]。双重机器学习基于有效影响函数构造内曼正交估计方程，并通过样本拆分将其与灵活的机器学习得到的干扰参数估计量相结合 [7]。TMLE通过对数据生成分布的初始估计沿局部最不利子模型进行扰动，直至有效影响函数的估计方程被近似求解，从而得到插件估计量 [25, 27, 28]。尽管这些方法在适当条件下具有渐近有效性，但最不利性通常仅在局部得到保证，实际实现往往需要迭代靶向步骤。通用最不利子模型通过沿整条分布路径在全局上保证最不利性来克服这一局限，从而以最小的似然扰动实现所需的偏差降低 [23]。然而，这些方法的一个共同特点是依赖于对有效影响函数的显式推导与求值，这在复杂的半参数模型中可能在分析上要求很高，并且具有参数特异性。

不依赖影响函数的方法旨在绕过对有效影响函数的解析推导。现有方法包括通过有限差分扰动来近似有效影响函数的方法 [5, 12] 或蒙特卡洛近似 [1], ，以及自动去偏机器学习（AutoDML）方法 [8, 21]。有限差分与蒙特卡洛方法以数值方式估计有效影响函数，这可能引入额外的近似误差。AutoDML则依赖于正交再参数化，从而能够在某些结构化设定中实现自动去偏 [8]。Cho等人的核去偏插件估计量（KDPE）通过将去偏问题嵌入再生核希尔伯特空间，提供了一种灵活的替代方案[9]。KDPE构造数据自适应扰动，在通用再生核希尔伯特空间中近似求解经验得分方程，而无需有效影响函数的显式表达式。然而，KDPE依赖于局部定义更新，并且通常通过迭代靶向步骤来实现。

我们的方法也与欠平滑高度自适应套索极大似然估计量（HAL-MLE） [22]相关。HAL-MLE在一类丰富的、具有有界 ℓ_1 范数的càdlàg函数上求解得分方程。与HAL类似，基于KDPE的方法求解大量得分方程，而不是针对单一参数特定的得分函数，从而实现 $\sqrt{n}$速率的偏差校正。然而，HAL在高维情形下可能在计算上变得困难，因为基函数的数量会随协变量维数和样本量迅速增长。相比之下，基于KDPE的方法依赖于以观测数据点为中心的 n核基函数，从而显著提高了计算可扩展性。近期工作将原始HAL估计量重新表述为高度自适应岭（HAR）和PCA-HAR估计量，用 ℓ_2 岭惩罚替换 ℓ_1 惩罚，并利用设计矩阵的核化表示 [18, 30]。这些方法也允许再生核希尔伯特空间表述，并实现类似的与维数无关的收敛速率（至多相差对数因子），但需要更强的光滑性假设，例如平方可积截面导数 [18]。

2. Problem Setup and Notation

令 $d, n \in \mathbb{N}$ 为固定的正整数，其中 d 视为固定值且不随 n 变化。为便于阐述，我们定义协变量空间 $\mathcal{X} = [0,1]^d$、处理空间 $\mathcal{A} = \{0,1\}$ 和结局空间 $\mathcal{Y} = \{0,1\}$。样本空间为 $\mathcal{O} = \mathcal{X} \times \mathcal{A} \times \mathcal{Y}$，配备乘积 σ-代数 $\mathcal{B}(\mathcal{O}) = \mathcal{B}([0,1]^d) \otimes 2^{\mathcal{A}} \otimes 2^{\mathcal{Y}}$。令 $O = (X, A, Y)$ 为定义在某个底层概率空间 $(\Omega\ \mathcal{F}, \mathbb{P})$ 上的 $\mathcal{O}$ 随机元素。对于大小为 n 的样本，我们观测到独立同分布的随机变量 $O_1, \ldots, O_n$，其中 $O_i = (X_i, A_i, Y_i)$。该理论可直接推广至有界可测样本空间 $\mathcal{O} \subset \mathbb{R}^k$。

我们考虑一个非参数统计模型 $\mathcal{M}$，它由定义在 $(\mathcal{O},\ \mathcal{B}(\mathcal{O}))$ 上、受 σ-有限乘积测度 $\mu = \lambda_d \otimes \mu_A \otimes \mu_Y$ 支配的概率测度构成，其中 λ_d 表示 $\mathcal{X} = [0,1]^d$ 上的勒贝格测度，而 μ_A，μ_Y 分别表示 $\mathcal{A}$ 和 $\mathcal{Y}$ 上的计数测度。令 P^* 表示 O 的真实数据生成分布，并假设 $P^* \in \mathcal{M}$。全文将使用上标“$*$”表示数据生成机制中真实但未知的特征。对于 $P \in \mathcal{M}$，将其相对于 μ 的拉东–尼科迪姆导数记为 $p = dP/d\mu$ 。令 $\bar{\mathcal{M}} := \{p \geq 0 : \int p d\mu = 1\}$ 表示与 $\mathcal{M}$ 中分布相对应的密度集合。

O 的实现值记为 $o \in \mathcal{O}$。令 P_X^* 表示 P^* 下 X 的边缘分布，并将其相对于 $\mathcal{X} = [0,1]^d$ 上勒贝格测度的密度记为 $p_X^* = dP_X^*/d\lambda_d$。那么联合密度 $p^* = dP^*/d\mu$ 可作如下分解

$$p^*(x,a,y) = p_X^*(x)e^*(a \mid x)q^*(y \mid a,x), \tag{1}$$

对于 μ-几乎处处的 $(x,\ a,\ y) \in \mathcal{O}$，其中 $e^*(\cdot \mid x)$ 和 $q^*(\cdot \mid a,\ x)$ 分别表示 A 给定 $X = x$ 以及 Y 给定 $(A = a,\ X = x)$ 的真实条件概率质量函数，将其视为相对于计数测度的密度。此外，设 $t \mapsto p_t$ 是一条相对于 μ 的概率密度路径，由开区间 $I \subset \mathbb{R}$ 中的 t 索引，并满足对所有 $t \in I,\ p_t \in \mathcal{M}$。

记号： 对于可测函数 $f\colon \mathcal{O} \to \mathbb{R}$ 且满足 $P|f| < \infty$，记作$Pf = P[f]$：$= \int f\,dP = \mathbb{E}_P[f]$。经验测度记为 P_n，定义为$P_n f := \frac{1}{n}\sum_{i=1}^n f(O_i)$。令 $L^2(P)$ 表示由满足 $P[f^2] < \infty$ 的可测函数$f\colon \mathcal{O} \to \mathbb{R}$ 构成的希尔伯特空间，这些函数在 P-几乎处处相等的意义下视为相同，并配备内积 $\langle f,\ g\rangle_{L^2(P)} = P[fg]$。我们将 $L_0^2(P) := \{f \in L^2(P) :\ P[f] = 0\}$ 定义为由均值为零的函数组成的 $L^2(P)$ 子空间。

2.1. 目标参数。 令 $\Psi\colon \mathcal{M} \to \mathbb{R}^m$ 表示定义在统计模型 $\mathcal{M}$ 上的一个 m 维参数映射。为简便起见，我们聚焦于标量情形 $m = 1$，并用 $\psi^* = \Psi(P^*)$ 表示感兴趣的目标参数。我们的目标是基于 n 个独立同分布观测，有效地估计 ψ^*。我们聚焦于形式为 $\psi_n = \Psi(P_n)$ 的插件估计量，其中 P_n 是 $\mathcal{M}$ 中 P^* 的数据依赖估计。这种表述在保留插件结构的同时，将一大类估计量纳入统一框架。

3. Preliminaries

我们将讨论范围限定于关于 $\mathcal{M}$ 路径可微的参数，从而确保其对底层分布的平滑扰动具有一阶敏感性。路径可微性是正则估计量存在的必要条件，并通过典范梯度构成半参数效率理论的基础。

定义 1（路径可微性）。称参数 $\Psi\colon \mathcal{M} \to \mathbb{R}$ 为路径可微于 $P \in \mathcal{M}$，如果存在 $\phi_P^* \in L_0^2(P)$，使得对每一个正则参数子模型 $\{P_\epsilon :\ \epsilon \in (-l, l)\} \subset \mathcal{M}$，其满足 $P_{\epsilon=0} = P$，且在 $\epsilon = 0$ 处具有二次均值可微性，并带有得分函数

$$h = \left.\frac{d}{d\epsilon}\log p_\epsilon\right|_{\epsilon=0} \in L_0^2(P),$$

映射 $\epsilon \mapsto \Psi(P_\epsilon)$ 在 $\epsilon = 0$ 处可微，且

$$\left.\frac{d}{d\epsilon}\Psi(P_\epsilon)\right|_{\epsilon=0} = P[\phi_P^* h].$$

函数 ϕ_P^* 称为 Ψ 在 P处相对于模型 $\mathcal{M}$的典范梯度，或有效影响函数（*EIF*）。

我们关注正则渐近线性（RAL）估计量，这是一类范围广且易于处理的估计量，其中包括在标准正则性条件 [29]下针对路径可微参数的有效估计量。渐近线性通过影响函数给出一阶行为和渐近正态性，而正则性则确保估计量在局部扰动下保持稳定。接下来，我们将这一概念形式化。

定义 2（渐近线性）。一个 $\psi_n = \Psi(P_n)$ 估计量 $\psi^* = \Psi(P^*)$ 在 P^* 处渐近线性，如果存在一个函数 $\phi_{P^*} \in L_0^2(P^*)$ 使得

$$\sqrt{n}\Big[(\psi_n - \psi^*) - P_n\phi_{P^*}\Big] \xrightarrow{P} 0.$$

函数 ϕ_{P^*} 称为该估计量的影响函数。

定义 3（正则估计量）。若对每个正则参数子模型 $\{P_\epsilon : \epsilon \in (-l, l)\} \subset \mathcal{M}$（满足 $P_{\epsilon=0} = P^*$，且在 $\epsilon = 0$ 处具有二次均值可微性，得分函数为$h \in L_0^2(P^*)$），估计量 ψ_n *of* $\psi^* = \Psi(P^*)$ 在 $P^* \in \mathcal{M}$ 处是正则的，则数列

$$\sqrt{n}\big(\psi_n - \Psi(P_{\epsilon_n})\big)$$

依分布收敛到一个不依赖于方向 h 的非退化极限，对于满足 $t \in \mathbb{R}$ 的局部备择 $\epsilon_n = t/\sqrt{n}$。

路径可微性、渐近线性和正则性共同构成半参数效率框架的特征。若 Ψ 在P^*处路径可微，且其典范梯度为 $\phi^*_{P^*}$，而估计量 ψ_n正则且渐近线性，影响函数为 ϕ_{P^*}，那么 ψ_n的渐近方差为 $P^*[\phi^2_{P^*}]$。半参数效率理论表明，$P^*[\phi^2_{P^*}] \geq P^*[(\phi^*_{P^*})^2]$,，当且仅当$\phi_{P^*} = \phi^*_{P^*}$ 几乎处处成立时取等号。在这种情况下，该估计量称为模型 $\mathcal{M}$ [3, 29]中的有效估计量。

3.1. 得分函数。令$h = \frac{d}{d\epsilon}\log p_\epsilon\big|_{\epsilon=0}$ 表示正则一维参数子模型 $\{P_\epsilon$的得分函数：$\epsilon \in (-l, l)\} \subset \mathcal{M}$经过 P。二次均值可微性意味着 $h \in L_0^2(P)$ [29]。所有此类得分函数的线性张成空间在 $L_0^2(P)$中的闭包，定义为 $\mathcal{M}$ 在 P处的切空间；在全非参数模型中，该切空间与 $L_0^2(P)$重合。给定$h \in L_0^2(P)$，实现这一得分函数的常见子模型包括线性路径$p_{\epsilon,\ h} = (1 + \epsilon h)p$，其对足够小的 $|\epsilon|$定义，使得 $1 + \epsilon h \geq 0$；以及指数倾斜路径$p_{\epsilon,\ h} = \exp(\epsilon h)p/P[\exp(\epsilon h)]$,，二者均满足 $\frac{d}{d\epsilon}\log p_{\epsilon,\ h}\big|_{\epsilon=0} = h$。尽管这些构造足以用于局部效率理论，但并不适合用来定义通用最不利路径。

3.2. 影响函数。Ψ 在 P 处的路径可微性意味着，其沿得分函数为 $h \in L_0^2(P)$的任意子模型的方向导数都是 h的连续线性泛函。根据里斯表示定理，存在唯一的 $\phi^*_P \in L_0^2(P)$，使得$\frac{d}{d\epsilon}\Psi(P_{\epsilon,\ h})\big|_{\epsilon=0} = P[\phi^*_P h]$，对切空间中的所有 h均成立（在非参数模型中，该切空间等于 $L_0^2(P)$）。估计量通过渐近线性性具有影响函数 ϕ_{P^*}。如果该估计量是正则且有效的，则 $\phi_{P^*} = \phi^*_{P^*}$ 几乎处处成立。

我们使用冯·米塞斯型展开来分析估计误差 $\Psi(P_n) - \Psi(P^*)$，这为正则渐近线性估计量的一阶渐近分析奠定了基础，[3], 如定理 3.1 所示。我们不围绕 P^* 展开 $\Psi(\cdot)$，而是围绕数据自适应估计 P_n进行展开，从而以便于分析靶向估计量的形式，将经验过程波动与高阶余项分离出来。

命题 3.1 (冯·米塞斯展开 [3])。假设 Ψ 在 P^* 和 $\widehat{P}_n$ 处均路径可微，其典范梯度分别为 $\phi^*_{P^*} \in L_0^2(P^*)$ 和 $\phi^*_{\widehat{P}_n} \in L_0^2(\widehat{P}_n)$。

假设 Ψ 在 P_n的某个邻域内存在二阶余项 $R_2(\cdot,\cdot)$，因此

$$\Psi(P^*) - \Psi(\widehat{P}_n) = (P^* - \widehat{P}_n)\phi^*_{\widehat{P}_n} + R_2(P^*, \widehat{P}_n).$$

那么，估计误差 $\Psi(P_n) - \Psi(P^*)$ 可分解为

$$\mathbb{P}_n\phi^*_{P^*} - \underbrace{\mathbb{P}_n\phi^*_{\widehat{P}_n}}_{Term\ 1} + \underbrace{(\mathbb{P}_n - P^*)\big(\phi^*_{\widehat{P}_n} - \phi^*_{P^*}\big)}_{Term\ 2} + \underbrace{R_2(P^*, \widehat{P}_n).}_{Term\ 3} \tag{2}$$

命题 3.1 中的展开式表明，若能控制经验过程（$Term\ 2$）、二阶余项（$Term\ 3$）和代入偏差（$Term\ 1$），则可实现渐近线性。前两项可通过下文陈述的标准经验过程条件和余项条件加以处理，这些条件通常假设在因果推断文献中 [27, 28]。所提方法的目标是构造 P_n，使得 $\phi^*_{\widehat{P}_n}$ 的经验均值较小，从而消除上述展开式中的主要代入偏差项。

假设 1. 假设存在一个确定性函数类 $\mathcal{F} \subset L_0^2(P^*)$，使得 $P^*\{\phi^*_{\widehat{P}_n} \in \mathcal{F}\} \to 1$，函数类 $\mathcal{F}$是 P^*-*Donsker* 类，且 $\|\phi^*_{\widehat{P}_n} - \phi^*_{P^*}\|_{L^2(P^*)} \to_{P^*} 0$。那么经验过程项满足 $(\mathbb{P}_n - P^*)\big(\phi^*_{\widehat{P}_n} - \phi^*_{P^*}\big) = o_{P^*}(n^{-1/2})$。

假设 2. 假设 P_n 收敛得足够快，且 Ψ足够规则，从而命题 *3.1* 中的二阶余项满足 $R_2(P_n, P^*) = o_{P^*}(n^{-1/2})$。

3.3. **最不利子模型**。经典半参数方法通过局部最不利子模型 （LLFS）引入分布路径，其在初始分布处的得分函数与参数 [27, 28] 的典范梯度相一致。LLFS 是经过 P 的一维参数子模型 $\{P_\varepsilon\}$，其在 $\varepsilon = 0$ 处的得分函数等于 Ψ 在 P 处的典范梯度 ϕ^*_P。这一构造纯粹是局部的：最不利性仅在单一分布 P 处无穷小地成立。

定义 4 （局部最不利子模型）。设 $P \in \mathcal{M}$ 的密度为 $p = dP/d\mu$，并设 $\Psi : \mathcal{M} \to \mathbb{R}$ 在 P 处路径可微，其典范梯度为 $\phi^*_P \in L_0^2(P)$。经过 $\delta > 0$ 的一维参数子模型 $\{P_\varepsilon : \varepsilon \in (-\delta, \delta)\} \subset \mathcal{M}$ 称为 Ψ 在 P 处的局部最不利子模型，若：

(1) $P_{\varepsilon=0} = P$；*(2)* 子模型在 $\varepsilon = 0$ 处关于二次均值可微；

(3) 子模型在 $\varepsilon = 0$ 处的得分函数满足

$$\frac{d}{d\varepsilon}\log p_\varepsilon(o)\bigg|_{\varepsilon=0} = \phi^*_P(o), \qquad P\text{-}a.e.\ o \in \mathcal{O}. \tag{3}$$

相比之下，通用最不利子模型 （ULFS）沿整个路径 [23] 一致地确保最不利性。经过 P 的 ULFS 是一个参数子模型 $\{P^{(t)} : t \in I\}$，其沿路径各点的得分函数均与当前分布处的典范梯度一致。

定义5 (通用最不利子模型)。设 $P \in \mathcal{M}$ 的密度为 $p = dP/d\mu$，并设 Ψ: $\mathcal{M} \to \mathbb{R}$ 在 $\mathcal{M}$ 上路径可微，其典范梯度为

ϕ_P^*。参数子模型 通用最不利子模型$(P) = \{P_t : \ t \in I\} \subset \mathcal{M}$，其中 $I \subset \mathbb{R}$ 是开集，称为经过 P 的通用最不利子模型，如果：

(1) $P_{t=0} = P$；*(2)* 对每个 $t \in I$，映射 $t \mapsto p_t$ 可微；*(3)* 对每个 $t \in I$，

$$\frac{d}{dt} \log p_t(o) = \phi^*_{P_t}(o), \qquad \mu\text{-a.e. } o \in \mathcal{O}, \tag{4}$$

其中 $\phi^*_{P_t}$ 表示在 P_t处求得的 Ψ的典范梯度。

方程 (4) 是定义 ULFS 的微分方程，其初始条件为$p_0 = p$。假设路径定义良好且可归一化，积分可得归一化表示

$$p_t(o) = \frac{p(o) \exp\Big(\int_0^t \phi^*_{P_x}(o)\, dx\Big)}{P\Big[\exp\Big(\int_0^t \phi^*_{P_x}\, dx\Big)\Big]}. \tag{5}$$

反过来，任何具有形式 (5) 且关于 t 可微的路径都满足 (4)。因此，ULFS 既可以看作 ODE (4) 的解，也可以看作由沿路径求得的典范梯度所生成的指数流。

3.4. 再生核希尔伯特空间（RKHS）。设 $\mathcal{V}$ 为一个非空集合。函数 $K : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ 称为正定（PD），如果对任意有限集合 $\mathbf{v} = (v_1, \ldots, v_n) \in \mathcal{V}^n$ 以及任意 $\alpha \in \mathbb{R}^n \setminus \{0\}$，$\sum_{i,\ j=1}^{n} \alpha_i \alpha_j K(v_i, v_j) > 0$[3]。对于 $v \in \mathcal{V}$，定义核截面 $k_v(\cdot) = K(v, \cdot)$。给定有限集合 $\mathbf{v} \in \mathcal{V}^n$ 和 $\mathbf{v}' \in \mathcal{V}^m$，相应的核矩阵为 $K_{\mathbf{vv}'} = [K(v_i, v'_j)] \in \mathbb{R}^{n \times m}$；特别地，$K_{\mathbf{v}} := K_{\mathbf{vv}}$ 是对称且正定的。等价地，将核截面收集为 $k_{\mathbf{v}}(\cdot) = (k_{v_1}(\cdot) \ldots, k_{v_n}(\cdot))^{\mathsf{T}}$，可得 $k_{\mathbf{v}}(\mathbf{v}') = K_{\mathbf{vv}'}$。

给定如上定义的正定核 K，一个自然的问题是：是否存在一个定义在 $\mathcal{V}$ 上的规范函数希尔伯特空间，使得核截面 $\{k_v : v \in \mathcal{V}\}$ 表示点赋值。Moore-Aronszajn[2]，在定理 3.2 中给出了答案，该定理表明，与 K相关联的再生核希尔伯特空间（RKHS）具有存在唯一性。

命题 3.2 (Moore–Aronszajn [2])。设 $\mathcal{V}$ 为一个非空集合，且 $K : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ 为一个正定核。则存在唯一的函数希尔伯特空间$\mathcal{H}_K$，其中的函数为 $f : \mathcal{V} \to \mathbb{R}$，满足：

(1) 对于每个 $v \in \mathcal{V}$，核截面 $k_v(\cdot) = K(v, \cdot)$ 属于 $\mathcal{H}_K$；*(2)* 再生性质成立，即，

$$\langle f, k_v \rangle_{\mathcal{H}_K} = f(v) \qquad \textit{for all } f \in \mathcal{H}_K \textit{ and } v \in \mathcal{V}.$$

此外，$\mathcal{H}_K$ 是在 K所诱导的内积下，由 $\{k_v : \ v \in \mathcal{V}\}$ 的线性张成空间的完备化。

我们将 $\mathcal{H}_K$ 称为与K相关联的再生核希尔伯特空间 (RKHS)，并将其内积和范数分别记为 $\langle \cdot, \ \cdot \rangle_{\mathcal{H}_K}$和 $\| \cdot \|_{\mathcal{H}_K}$。根据 RKHS 内积的定义，核截面满足 $\langle k_v, \ k_{v'} \rangle_{\mathcal{H}_K} = K(v, \ v')$。RKHS 的两项结构性质将在下文反复用到。首先，点赋值是连续线性泛函：对于每个 $v \in \mathcal{V}$, $f(v) = \langle f, \ k_v \rangle_{\mathcal{H}_K}$, $f \in \mathcal{H}_K$。其次，$\mathcal{H}_K$ 是张成空间$\{k_v$ 的希尔伯特空间完备化：$v \in \mathcal{V}\}$，因此每个 $f \in \mathcal{H}_K$ 都可以

[3]对于半正定的情形，这通常已足够，$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j K(v_i, \ v_j) \geq 0$。

通过核截面的有限线性组合在 $\|\cdot\|_{\mathcal{H}_K}$ 中逼近。如果核是有界的，即 $\sup_{v\in\mathcal{V}} K(v,\ v) < \infty$，那么对于每个 $f \in \mathcal{H}_K$，

$$\|f\|_\infty = \sup_{v\in\mathcal{V}} |f(v)| \le \|f\|_{\mathcal{H}_K} \left(\sup_{v\in\mathcal{V}} K(v,v)\right)^{1/2}.$$

3.4.1. 均值为零的再生核希尔伯特空间。 设 K 为 $\mathcal{V}$ 上的正定核，其关联的再生核希尔伯特空间为 $\mathcal{H}_K$，并设 P 为 $\mathcal{V}$上的概率测度。当核均值嵌入$m_P := \int k_v\, dP(v)$作为 $\mathcal{H}_K$中的元素存在时，期望可表示为 $\mathcal{H}_K$上的内积，从而得到一个自然的 P-均值为零子空间。

命题 3.3 （均值为零的再生核希尔伯特空间）。 设 K 为 $\mathcal{V}$ 上的正定核，其关联的再生核希尔伯特空间为 $\mathcal{H}_K$，并设 P 为 $\mathcal{V}$ 上的概率测度，使得核均值嵌入 $m_P := \int k_v\, dP(v)$ 存在于 $\mathcal{H}_K$中。定义 P-均值为零子空间

$$\mathcal{H}_{K,P} := \{g \in \mathcal{H}_K : P[g] = 0\}.$$

则$\mathcal{H}_{K,\ P}$ 是 $\mathcal{H}_K$的闭线性子空间，因而在所继承的内积下是一个希尔伯特空间。

设 Π_P :$\mathcal{H}_K \to \mathcal{H}_{K,\ P}$ 表示正交投影。对每个 $v \in \mathcal{V}$，$\mathcal{H}_{K,\ P}$ 中在 v 处的求值表示元为

$$k_{P,v} := \Pi_P k_v = k_v - \frac{\langle k_v, m_P\rangle_{\mathcal{H}_K}}{\|m_P\|^2_{\mathcal{H}_K}}\, m_P = k_v - \frac{m_P(v)}{\|m_P\|^2_{\mathcal{H}_K}}\, m_P,$$

并且，中心化核有一个等价表示

$$K_P(v,v') = \langle \Pi_P k_v, \Pi_P k_{v'}\rangle_{\mathcal{H}_K} = K(v,v') - \frac{m_P(v)\, m_P(v')}{\|m_P\|^2_{\mathcal{H}_K}}.$$

K_P 的正定性源于它是$\mathcal{H}_{K,P}$的再生核。

3.4.2. 通用核。 某些核所生成的再生核希尔伯特空间能够近似大类函数，而不受底层概率测度的影响。令$C_0(\mathcal{V})$表示定义在 $\mathcal{V}$ 上的实值连续函数所构成的巴拿赫空间，其中的函数在无穷远处趋于零，并配备一致范数 $\|\cdot\|_\infty$。

定义 6 (通用核)。 定义在 $\mathcal{V}$ 上的连续正定核 K 称为通用的，若其对应的再生核希尔伯特空间关于一致范数在 $C_0(\mathcal{V})$ 中稠密。

通用性意味着，对于每个定义在 $\mathcal{V}$ 上的概率测度 P 以及每个 $1 \le p < \infty$ [13], 通用核 K的再生核希尔伯特空间 $\mathcal{H}_K$在 $L^p(P)$中稠密。特别地， $\mathcal{H}_K$在 $L^2_0(P)$中稠密。一个典型例子是定义在 $\mathcal{V} = \mathbb{R}^d$上的高斯核。对于固定的 $\sigma > 0$,

$$K(v,v') = \exp\left(-\frac{\|v - v'\|_2^2}{2\sigma^2}\right), \qquad v, v' \in \mathbb{R}^d.$$

该核是有界的，且 $K(v,v) = 1$，其对应的再生核希尔伯特空间 $\mathcal{H}_K$ 由有界连续函数构成。由于该核有界，对于定义在 $\mathbb{R}^d$ 上的每个概率测度 P，核均值嵌入 $m_P = \int k_v\, dP(v)$ 都存在于 $\mathcal{H}_K$ 中。高斯核是通用的，因此 $\mathcal{H}_K$ 在 $L^2_0(P)$ 中稠密。因此，将扰动限制在 $\mathcal{H}_K$ 中，可以使我们在 $L^2_0(P)$ 中逼近影响函数。

3.5. **不动点分析**。 设 (X, d) 是一个完备度量空间，且 $T : X \to X$ 是该度量空间上的映射。下面给出巴拿赫压缩原理所依据的一些基本定义和定理。

定义 7 （压缩映射）。 设 X 是配有距离 d的度量空间。映射 $T: X \to X$ 称为利普希茨连续，若存在常数$\lambda \in$ （0，1）使得

$$d\big(Tx, Ty\big) \leq \lambda\, d(x, y) \qquad for\ all\ x, y \in X.$$

对于 $\lambda < 1$， T 是压缩映射。若 $\lambda \leq 1$，则 T 是非扩张的。

经典的Banach不动点定理保证了完备度量空间上压缩映射的不动点的存在唯一性，如定理 3.4所述。其证明可见于例如 [15]的推论 1.3。

定理 3.4 （Banach不动点）。 设 T 是完备度量空间（X， d）上的压缩映射。则 T 有唯一不动点 $x^* \in X$。

以下推广有时称为最终压缩原理，它允许映射本身不是压缩映射，只要其某个有限次迭代是压缩映射即可。对于 T: $X \to X$和 $m \in \mathbb{N}$，我们用 T^m表示 T的 m^{th} 次迭代，即迭代 $T \circ T \cdots \circ T$ m 次，其中 T^0是恒等映射。

推论 3.5 (通过迭代压缩求不动点)。 设(X, d) 为完备度量空间，且 $T : X \to X$ 为一个映射。假设存在整数$m \geq 1$ ，使得迭代 T^m 在 X上是压缩映射。那么 T 存在唯一不动点 $x^* \in X$，即 $T(x^*) = x^*$。此外，对于任意初始点 $x_0 \in X$，迭代序列 $\{T^k(x_0)\}_{k\geq 0}$ 收敛到 x^*。

注意，推论 3.5 可直接由巴拿赫不动点定理应用于 T^m 得出；这是因为 T^m的任何不动点必然也是 T的不动点，而 T^m的不动点的唯一性意味着 T的不动点也唯一。

定理 3.6 (巴拿赫空间中的皮卡-林德勒¨定理 [19])。 设$(\mathcal{X}, \|\cdot\|)$ 是一个巴拿赫空间，且 $U \subset \mathcal{X}$ 是开集。假设 $F : U \to \mathcal{X}$ 是局部利普希茨的。那么，对于每个 $x_0 \in U$，都存在 $\tau > 0$ 和唯一的解 $x \in C^1((-\tau, \tau); \mathcal{X})$，使其满足初值问题

$$\frac{d}{dt}x(t) = F(x(t)), \qquad x(0) = x_0.$$

此外，该解满足积分方程

$$x(t) = x_0 + \int_0^t F(x(s))\, ds, \qquad t \in (-\tau, \tau).$$

4. Debiasing with RKHS

设 $t \mapsto p_t$ 表示相对于固定的σ-有限测度 μ、定义在 $\mathcal{O}$ 上的概率密度路径，并将其对应的概率测度记为 $P^{(t)} \in \mathcal{M}$ 。对于通用最不利子模型（ULFS），路径 $\{P^{(t)} : t \in I\}$ 并未被指定为有限维参数族 [23]。相反，密度的演化通过得分型方向 $h(P_{(t)}) \in L_0^2(P^{(t)})$由微分方程定义

$$\frac{d}{dt}\log p_t(o) = h(P^{(t)})(o). \tag{6}$$

因此，ULFS 在概率测度空间上定义了一个流，其瞬时方向取决于当前分布。下文将 $h(P^{(t)})$ 简称为（得分型）方向 。

受此构造启发，我们引入一种受再生核希尔伯特空间限制的替代流，使由再生核希尔伯特空间索引的经验估计方程趋于零。设 $\mathcal{H}_K$ 是定义在 $\mathcal{O}$ 上的实值函数构成的再生核希尔伯特空间，其再生核为 $K\colon \mathcal{O}\times\mathcal{O}\to\mathbb{R}$。我们考虑高斯核

$$K(o,o')=\exp\left(-\frac{\|o-o'\|_2^2}{2\sigma^2}\right),\qquad o,o'\in\mathcal{O},$$

其中带宽 $\sigma>0$为某个固定值。该核连续、正定且有界，并且 $K(o,\ o)\ =\ 1$。对于每个 $o\in\mathcal{O}$，我们用 $k_o(\cdot)\ =\ K(o,\ \cdot)$表示相应的核截面。以下结果记录了将期望视为定义在 $\mathcal{H}_K$上的泛函时的一些基本性质。第4节的所有证明和推导均置于附录 A（第10节）。

命题 4.1（作为 RKHS 泛函的期望）。 设 P 是 $\mathcal{O}$上的任意概率测度， $\mathcal{H}_K$ 是与高斯核 K相关联的再生核希尔伯特空间。则核均值嵌入 $m_P\colon\ =\int k_o\,dP$（o）存在于 $\mathcal{H}_K$中。此外，对于每个 $f\in\mathcal{H}_K$，

$$P[f]=\langle f,m_P\rangle_{\mathcal{H}_K}.$$

因此，映射 $f\mapsto P[f]$ 是 $\mathcal{H}_K$上的连续线性泛函，满足

$$|P[f]|\le\|f\|_{\mathcal{H}_K}.$$

这一表示表明，期望可以表示为再生核希尔伯特空间中的内积。特别地，再生核希尔伯特空间范数控制着期望的大小。在我们的设定中，再生核希尔伯特空间范数较小的函数在 P下的期望也较小。

注 1. 在许多应用中，核可以定义在协变量空间$\mathcal{X}$ 上，而非完整的样本空间 $\mathcal{O}$上，因为统计模型的无限维部分由 X索引。将 $\mathcal{O}$替换为 $\mathcal{X}$，并将 $P^{(t)}$ 替换为 X的边缘分布后，所有构造仍然有效。

4.1. 基于工作再生核希尔伯特空间的切空间。 对于每个 $t\in I$，我们定义相应的均值为零的再生核希尔伯特空间

$$\mathcal{H}_K^{(t)}:=\mathcal{H}_{K,P^{(t)}}:=\{f\in\mathcal{H}_K:P^{(t)}[f]=0\}.$$

该空间由 $\mathcal{H}_K$ 中在 $P^{(t)}$下期望为零的函数构成。如下文所示， $\mathcal{H}_K^{(t)}$ 是 $\mathcal{H}_K$ 的闭线性子空间，因此其自身也是一个希尔伯特空间。此外，由于点赋值在 $\mathcal{H}_K^{(t)}$ 上仍然连续，它也是一个再生核希尔伯特空间，其再生核记为 $K^{(t)}$。对于 $o\in\mathcal{O}$，我们将对应的核截面记为 $k_o^{(t)}(\cdot)\ =\ K^{(t)}(o,\ \cdot)$。

命题4.2 (均值为零的再生核希尔伯特空间作为闭子空间**)。** 对于每个 $t\in I$，空间 $\mathcal{H}_K^{(t)}$ 是 $\mathcal{H}_K$的闭线性子空间。此外，

$$\mathcal{H}_K^{(t)}=\{f\in\mathcal{H}_K:\langle f,m_{P^{(t)}}\rangle_{\mathcal{H}_K}=0\}=(\operatorname{span}\{m_{P^{(t)}}\})^{\perp}.$$

特别地， $\mathcal{H}_K^{(t)}$ 在继承自 $\mathcal{H}_K$的内积下是一个希尔伯特空间。

设 Π_t ： $\mathcal{H}_K \to \mathcal{H}_K^{(t)}$ 表示到 $\mathcal{H}_K^{(t)}$ 上的正交投影，该投影存在，因为 $\mathcal{H}_K^{(t)}$ 是 $\mathcal{H}_K$的闭子空间。由于 $\mathcal{H}_K^{(t)} = (\text{张成空间}\{m_{P^{(t)}}\})^{\perp}$，该投影去除了 $\mathcal{H}_K$ 中函数沿方向 $m_{P^{(t)}}$的分量。

命题 4.3（投影到均值为零的再生核希尔伯特空间上的正交投影）。 若 $m_{P^{(t)}} \neq 0$, 正交投影 Π_t : $\mathcal{H}_K \to \mathcal{H}_K^{(t)}$ 由下式给出

$$\Pi_t f = f - \frac{\langle f, m_{P^{(t)}} \rangle_{\mathcal{H}_K}}{\|m_{P^{(t)}}\|^2_{\mathcal{H}_K}} m_{P^{(t)}} = f - \frac{P^{(t)}[f]}{\|m_{P^{(t)}}\|^2_{\mathcal{H}_K}} m_{P^{(t)}}, \qquad f \in \mathcal{H}_K.$$

该投影还决定了 $\mathcal{H}_K^{(t)}$ 的再生核。对于每个 $o \in \mathcal{O}$,

$$\Pi_t k_o = k_o - \frac{m_{P^{(t)}}(o)}{\|m_{P^{(t)}}\|^2_{\mathcal{H}_K}} m_{P^{(t)}},$$

其中我们用到了再生性质 $\langle k_o, m_{P^{(t)}} \rangle_{\mathcal{H}_K} = m_{P^{(t)}}(o)$。因此， $\mathcal{H}_K^{(t)}$ 的再生核为

$$K^{(t)}(o, o') = \langle \Pi_t k_o, \Pi_t k_{o'} \rangle_{\mathcal{H}_K} = K(o, o') - \frac{m_{P^{(t)}}(o)\, m_{P^{(t)}}(o')}{\|m_{P^{(t)}}\|^2_{\mathcal{H}_K}}, \qquad o, o' \in \mathcal{O}.$$

若 $m_{P^{(t)}} = 0$，则 $\mathcal{H}_K^{(t)} = \mathcal{H}_K$ 和 $K^{(t)} = K$。

4.2. 受再生核希尔伯特空间约束的去偏方向与诱导流。 将允许的方向限制为 $\mathcal{H}_K^{(t)}$ ，可确保任意方向 $h(P^{(t)})$ 自动满足零均值约束 $P_{(t)}[h(P_{(t)})] = 0$。因此，路径 $t \mapsto P^{(t)}$ 按照一个方向场演化，该方向场在每个 $P^{(t)}$ 处的取值都属于 $\mathcal{H}_K^{(t)}$ 。核心问题是：对于每个当前状态 $P^{(t)}$，我们应如何选择一个方向$h(P^{(t)}) \in \mathcal{H}_K^{(t)}$ ，使路径朝着去偏的方向前进?

在固定的 $P^{(t)}$ 下， $\mathcal{H}_K^{(t)}$ 内的经验偏差由定义在 $\mathcal{H}_K^{(t)}$ 上的泛函 $f \mapsto P_n[f]$ 刻画，因为对于所有 $f \in \mathcal{H}_K^{(t)}$ ，有 $(P_n - P_{(t)})[f] = P_n[f]$。由于高斯核是有界的，映射 $f \mapsto P_n[f]$ 关于 $\mathcal{H}_K$ 范数连续，因此在 $\mathcal{H}_K$ 上定义了一个有界线性泛函。其对闭子空间 $\mathcal{H}_K^{(t)} \subset \mathcal{H}_K$ 的限制同样有界；我们将这一受限泛函记为 L_n : $\mathcal{H}_K^{(t)} \to \mathbb{R}$，其中 $L_n(f) = P_n[f] = \frac{1}{n}\sum_{i=1}^n f(O_i)$。 _

命题 4.4（均值为零的再生核希尔伯特空间中的经验黎斯表示元）。 存在唯一的元素 $m_n^{(t)} \in \mathcal{H}_K^{(t)}$ 使得

$$P_n[f] = \langle f, m_n^{(t)} \rangle_{\mathcal{H}_K^{(t)}} \qquad for\ all\ f \in \mathcal{H}_K^{(t)}.$$

此外,

$$m_n^{(t)} = \frac{1}{n}\sum_{i=1}^n k_{O_i}^{(t)}.$$

根据 $\mathcal{H}_K^{(t)}$ 的构造，总体均值嵌入满足 $\int k_o^{(t)}\, dP^{(t)}(o) = 0$。因此， $m_n^{(t)}$ 表示经验偏差泛函 $f \mapsto (P_n - P_{(t)})[f]$ 在 $\mathcal{H}_K^{(t)}$ 上的 Riesz 表示元。因此，对于任意 $f \in \mathcal{H}_K^{(t)}$,

$$|P_n[f]| = |\langle f, m_n^{(t)} \rangle_{\mathcal{H}_K^{(t)}}| \le \|f\|_{\mathcal{H}_K^{(t)}} \|m_n^{(t)}\|_{\mathcal{H}_K^{(t)}},$$

当且仅当 $m_n^{(t)} \neq 0$ 时， $f = m_n^{(t)}/\|m_n^{(t)}\|_{\mathcal{H}_K^{(t)}}$ 处取等号。因此

$$\sup_{\substack{f\in\mathcal{H}_K^{(t)}\\ \|f\|_{\mathcal{H}_K^{(t)}}\le 1}} |P_n[f]| = \sup_{\substack{f\in\mathcal{H}_K^{(t)}\\ \|f\|_{\mathcal{H}_K^{(t)}}\le 1}} |\langle f, m_n^{(t)}\rangle_{\mathcal{H}_K^{(t)}}| = \|m_n^{(t)}\|_{\mathcal{H}_K^{(t)}}.$$

我们注意到，这种表示有几个重要推论。首先， $\|m_n^{(t)}\|_{\mathcal{H}_K^{(t)}}$ 量化了 $\mathcal{H}_K^{(t)}$ 单位球上经验矩偏差的大小，且这种量化是均匀的。具体而言， $\|m_n^{(t)}\|_{\mathcal{H}_K^{(t)}}$ 是 $\mathcal{H}_K^{(t)}$ 单位球上的最坏情况经验偏差。其次，使 $\|m_n^{(t)}\|_{\mathcal{H}_K^{(t)}}$ 趋近于零，会迫使 $P_n[f]$ 在 $\{f \in \mathcal{H}_K^{(t)}$ 上均匀地趋于较小： $\|f\|_{\mathcal{H}_K^{(t)}} \le 1\}$。这是 RKHS 中近似求解由 $f \in \mathcal{H}_K^{(t)}$ 索引的一系列连续估计方程的对应形式。

5. One-Step KDPE (ULFS–KDPE)

固定 $t \in I$ ，并回顾经验均值嵌入 $m_n^{(t)} \in \mathcal{H}_K^{(t)}$ 的定义为 $m_n^{(t)} = P_n[k_O^{(t)}] = \frac{1}{n}\sum_{i=1}^n k_{O_i}^{(t)}$。根据构造， $m_n^{(t)}$是定义在 $\mathcal{H}_K^{(t)}$ 上的有界线性泛函 $f \mapsto P_n[f]$ 的 Riesz 表示元。尽管 $\mathcal{H}_K^{(t)}$ 是无限维的，经验均值嵌入却位于有限维子空间 $\mathrm{span}\{k_{O_1}^{(t)}, \dots, k_{O_n}^{(t)}\} \subset \mathcal{H}_K^{(t)}$ 中。因此，我们可以通过经验均值嵌入 $m_n^{(t)}$在观测样本点处的取值来表示它。定义坐标向量 $\boldsymbol{\alpha}^{(t)} \in \mathbb{R}^n$ 如下:

$$\alpha_j^{(t)} := P_n[k_{O_j}^{(t)}] = \frac{1}{n}\sum_{i=1}^n K^{(t)}(O_i, O_j), \qquad 1 \le j \le n. \tag{7}$$

等价地，根据 $\mathcal{H}_K^{(t)}$ 中的再生性质，

$$\alpha_j^{(t)} = \langle k_{O_j}^{(t)}, m_n^{(t)}\rangle_{\mathcal{H}_K^{(t)}} = m_n^{(t)}(O_j),$$

因此 $\boldsymbol{\alpha}^{(t)}$ 记录了经验均值嵌入在观测样本点处的取值。

经验均值嵌入 $m_n^{(t)}$完整表示了 $\mathcal{H}_K^{(t)}$ 上的经验矩偏差：当且仅当对所有 $f \in \mathcal{H}_K^{(t)}$ 有 $P_n[f] = 0$ 时， $m_n^{(t)} = 0$ 成立。因此，为了沿最不利路径对经验分布去偏差，自然应定义一个（得分函数型）方向，使 $m_n^{(t)}$趋向于零，同时仍保持在 $\mathcal{H}_K^{(t)}$ 中。我们并不直接沿 $m_n^{(t)}$的方向更新，而是采用经验预条件，使更新与由观测样本点处核函数取值所诱导的几何结构相一致。具体而言，定义

$$D_t = D(p_t) := \frac{1}{n}\sum_{j=1}^n \alpha_j^{(t)} k_{O_j}^{(t)} = \frac{1}{n}\sum_{j=1}^n \langle k_{O_j}^{(t)}, m_n^{(t)}\rangle_{\mathcal{H}_K^{(t)}} k_{O_j}^{(t)} \in \mathcal{H}_K^{(t)}. \tag{8}$$

等价地，

$$D_t = D(p_t) = \left(\frac{1}{n}\sum_{j=1}^n k_{O_j}^{(t)} \otimes k_{O_j}^{(t)}\right) m_n^{(t)} = \widehat{C}_t\, m_n^{(t)},$$

其中 $\widehat{C}_t := \frac{1}{n}\sum_{j=1}^{n} k_{O_j}^{(t)} \otimes k_{O_j}^{(t)}$ 是 $\mathcal{H}_K^{(t)}$ 上的经验协方差算子。沿此方向，通用最不利子模型方程 (6) 变为

$$\frac{d}{dt}\log p_t(o) = D(p_t)(o). \tag{9}$$

由于 $D(p_t) \in \mathcal{H}_K^{(t)}$，我们有 $P^{(t)}[D(p_t)] = 0$，这确保了 p_t 的归一化沿该流得以保持。

直观而言，经验均值嵌入 $m_n^{(t)}$度量 $\mathcal{H}_K^{(t)}$ 中当前的经验偏差。它是再生核希尔伯特空间范数下的最速下降方向，并对应于最小化 $\|m_n^{(t)}\|^2_{\mathcal{H}_K^{(t)}}$。经验协方差算子 $\widehat{C}_t$ 描述了该偏差如何通过观测数据点处的核函数取值被感知。因此，更新方向 $D(p_t) = \widehat{C}_t m_n^{(t)}$是一种核自然梯度流：经验矩几何中的最速下降。因此，该更新强调那些最能减少观测数据处经验矩违背的方向。当经验偏差消失时，该流恰好消失，这一性质我们将在后文定理 6.2 中建立。

5.1. 通用最不利子模型–核密度概率估计算法。 下文介绍 ULFS–KDPE 的一种实用实现。该方法利用观测数据处的核函数取值所诱导的有限维离散化，近似常微分方程 (9) 中的连续时间流。完整算法见附录 B（第 11 节）。

5.1.1. 离散化通用最不利子模型更新。 固定步长 $\Delta > 0$ 和初始密度估计 $\widehat{p}_0$。给定当前迭代 $\widehat{p}_t$ 和受再生核希尔伯特空间约束的方向 $D(\widehat{p}_t) \in \mathcal{H}_K^{(t)}$，我们对对数密度采用显式欧拉步来离散化该流

$$\log \widehat{p}_{t+\Delta}(o) = \log \widehat{p}_t(o) + \Delta\, D(\widehat{p}_t)(o). \tag{10}$$

等价地，

$$\widehat{p}_{t+\Delta}(o) = \widehat{p}_t(o)\exp\big(\Delta\, D(\widehat{p}_t)(o)\big). \tag{11}$$

这种乘法形式保持了密度估计的正性。

在连续时间流中，由于 $D(p_t) \in \mathcal{H}_K^{(t)}$ 蕴含 $P_t[D(p_t)] = 0$，归一化得以保持。然而，显式欧拉离散化无法精确保持积分，因此实际应用中可在更新后执行重新归一化步骤：

$$\widehat{p}_{t+\Delta}(o) \leftarrow \frac{\widehat{p}_{t+\Delta}(o)}{\int \widehat{p}_{t+\Delta}(u)\, d\mu(u)}. \tag{12}$$

因此，每次迭代都包括一个保持正性的指数倾斜，随后进行归一化。

5.1.2. 每次迭代计算什么？ 在迭代时刻 t，更新方向是方程（8）中定义的再生核希尔伯特空间元素。因此，方程（11）中的离散化更新仅需要系数 $\boldsymbol{\alpha}^{(t)}$ 以及对核截面 $k_{O_j}^{(t)}(\cdot)$求值的能力。在操作上，这些计算使用与零均值核 $K^{(t)}$相关联的中心化格拉姆矩阵来完成，

$$G^{(t)} = \big[K^{(t)}(O_i, O_j)\big]_{i,j=1}^{n}. \tag{13}$$

利用公式（7），可得到系数向量

$$\boldsymbol{\alpha}^{(t)} = \frac{1}{n} G^{(t)} \mathbb{1}, \tag{14}$$

其中，1 表示由 1 组成的 n 向量。在观测样本点处评估的方向为

$$\Big(D(\widehat{p}_t)(O_i)\Big)_{i=1}^{n} = \frac{1}{n} G^{(t)} \boldsymbol{\alpha}^{(t)}, \tag{15}$$

因为 $D(\widehat{p}_t)(O_i) = \frac{1}{n}\sum_{j=1}^{n} \alpha_j^{(t)} K^{(t)}(O_i, O_j)$。因此，每次迭代都可归结为涉及格拉姆矩阵 $G^{(t)}$的简单矩阵–向量运算。

从初始密度估计 $\widehat{p}_0$ （或分布估计 P_0）开始，该算法迭代执行以下步骤:

i 更新核 $K^{(t)}$ 以及由此诱导的均值为零的再生核希尔伯特空间 $\mathcal{H}_K^{(t)}$，使其对应于当前分布 P_t；ii 计算经验均值嵌入 $m_n^{(t)} = P_n[k_O^{(t)}]$，并通过其在样本点处的取值以系数向量 $\boldsymbol{\alpha}^{(t)}$表示；iii 构造再生核希尔伯特空间方向 $D(\widehat{p}_t)$；iv 使用指数倾斜步骤 (11) 更新密度，随后进行重新归一化 (12)。

该过程持续进行，直到经验均值嵌入足够小，或满足等价的判据。最终迭代结果 $\widehat{p}_T$ 定义了去偏插入式分布，并将 $\Psi(P_T)$ 作为 ULFS-KDPE估计量报告。

5.1.3. 停止准则。 我们描述离散化 ULFS–KDPE 流的实用停止规则。这些准则旨在作为互补性诊断；理论有效性并不要求任何单一规则。令 $\{P_t : t = 0、\Delta、2\Delta、\ldots\}$ 表示使用欧拉步长 Δ生成的迭代序列，其对应的密度为$\widehat{p}_t$、经验均值嵌入为 $m_n^{(t)}$、系数向量为 $\boldsymbol{\alpha}^{(t)} \in \mathbb{R}^n$。

（SC1）密度平台：拟合密度是否基本停止变化？该准则监测拟合密度本身的稳定情况。 定义对数密度的平方增量变化对数密度的增量变化

$$\Delta_t^{(p)} := P_n\Big[\big(\log \widehat{p}_t(O) - \log \widehat{p}_{t-\Delta}(O)\big)^2\Big].$$

算法在以下情况下停止

$$\Delta_t^{(p)} \le \delta_p \qquad \text{or} \qquad \big|\Delta_t^{(p)} - \Delta_{t-\Delta}^{(p)}\big| \le 0.1\,\delta_p.$$

因此，当连续的欧拉更新对拟合分布只产生可忽略的变化时，SC1 停止迭代。

(SC2) 得分平台：流是否已停止以有意义的方式提升目标？由于流方向为 $D(\widehat{p}_t)$，平稳性对应于方向得分的消失。定义 $s_t := P_n[D(\widehat{p}_t)] = \frac{1}{n}\sum_{j=1}^{n}\big(\alpha_j^{(t)}\big)^2$。算法在以下条件成立时停止

$$|s_t| \le \delta_s \qquad \text{or} \qquad |s_t - s_{t-\Delta}| \le 0.1\,\delta_s.$$

该判据与李雅普诺夫导数 $\frac{d}{dt} P_n[\log\, p_t]$ 直接相关，因此具有明确的理论解释：一旦 s_t 足够小，流就不再取得显著进展。

（SC3）再生核希尔伯特空间更新方向趋于消失：实际的再生核希尔伯特空间更新方-向是否接近于零？由于 $D(\widehat{p}_t)$ 是实际的更新方向，再生核希尔伯特空间范数较小表明流接近平衡态。若满足以下条件，算法即停止

$$\frac{1}{n}\|D(\widehat{p}_t)\|^2_{\mathcal{H}_K^{(t)}} = \frac{1}{n}\|\widehat{C}_t m_n^{(t)}\|^2_{\mathcal{H}_K^{(t)}} \le \delta_\alpha.$$

该准则直接衡量预条件偏差的大小。实际中，它与系数向量 $\boldsymbol{\alpha}^{(t)}$ 的大小密切相关，且计算成本低。

(SC4) 方差主导的更新：我们是否主要是在累积变异性，而没有取得实质性进展？为防止数值过拟合或不稳定性，我们比较更新的变异性与其平均改进幅度。定义

$$\Delta_t^{(v)} := P_n\Big[\big(\log \widehat{p}_t(O) - \log \widehat{p}_{t-\Delta}(O)\big)^2\Big]$$

and

$$\Delta_t^{(\ell)} := \Big|P_n[\log \widehat{p}_t(O) - \log \widehat{p}_{t-\Delta}(O)]\Big|.$$

如果满足以下任一条件，算法就会停止

$$\Delta_t^{(v)} \le \delta_v$$

or

$$P_n\Big[\big(\log \widehat{p}_t(O) - \log \widehat{p}_0(O)\big)^2\Big] \gtrsim n^{-1} \qquad \text{and} \qquad \Delta_t^{(\ell)} \le \delta_\ell.$$

第一个条件用于检测增量移动是否可以忽略不计；第二个条件则在累积变异性已变得不可忽略、但平均改进幅度进一步减小且可以忽略时，停止该流。

（SC5）有效影响函数近似求解：若有效影响函数可用，通常的目标最大似然估计得分方程是否已基本求解？当有效影响函数 $\phi^*_{\widehat{P}_t}$ 对目标参数可用，或可被稳定近似时，我们还可额外监测其经验均值。算法停止，如果

$$\big|P_n[\phi^*_{\widehat{P}_t}]\big| \le c\, n^{-1/2} \qquad \text{and} \qquad \big|P_n[\phi^*_{\widehat{P}_t}]\big| \ge \big|P_n[\phi^*_{\widehat{P}_{t-\Delta}}]\big|.$$

该准则对应于经典的目标最大似然估计停止规则：一旦有效影响函数估计方程被求解至一阶，且未观察到进一步改进，目标化步骤即终止。此处该准则为可选项，因为ULFS–KDPE在构造上不依赖影响函数。

6. Theoretical Results

我们不直接处理方程（9）中的对数密度表述，而是分析等价的常微分方程

$$\frac{d}{dt}p_t(o) = p_t(o)\, D(p_t)(o), \tag{16}$$

这由恒等式 $\frac{d}{dt}\log p_t = D(p_t)$得出,_只要 $p_t(o) > 0$ 且映射$t \mapsto p_t(o)$可微。在本节中，我们构造解 p_t，使其具有充分的正则性，从而使两种表述均定义良好且彼此等价。

为将方程（16）置于泛函分析框架中，我们将动力学写成按状态索引的形式。设 p 为满足 $\int_{\mathcal{O}} p\, d\mu = 1$的非负函数，并定义相应的概率测度 $P_p(A) := \int_A p(o)\, d\mu(o)$，$A \subset \mathcal{O}$。当 $p = p_t$ 时，我们便得到 $P_p = P_t$。对于每个 p，令 $\Pi_p : \mathcal{H}_K \to \mathcal{H}_{K,\,p}$ 表示

关于均值为零的再生核希尔伯特空间 $H_{K,\,p}$的正交投影。我们定义投影核截面 $k_o^{(p)} = \Pi_p k_o$，其中 $K^{(p)}(o,\ o') = \langle k_o^{(p)},\ k_{o'}^{(p)} \rangle_{\mathcal{H}_K}$，对于 $o \in \mathcal{O}$。给定 $O_1,\ \ldots, O_n$，定义系数向量 $\boldsymbol{\alpha}^{(p)} \in \mathbb{R}^n$，使其满足 $\alpha_j^{(p)} = \frac{1}{n} \sum_{i=1}^n K^{(p)}(O_i,\ O_j)$，并定义相应的流方向 $D(p) := \frac{1}{n} \sum_{j=1}^n \alpha_j^{(p)}\, k_{O_j}^{(p)}$。采用这一记号，方程（16）中的通用最不利子模型演化可简洁地写成自治常微分方程

$$\frac{d}{dt} l = F(p), \qquad F(p) := p\, D(p), \tag{17}$$

我们将其作为适当函数空间上的动力系统进行研究。向量场 $F(p) = p\, D(p)$ 的结构类似于复制子型动力学；密度按照方向 $D(p)$ 成比例地乘法演化。由于 $D(p) \in \mathcal{H}_{K,\,p}$因此满足 $P_p[D(p)] = 0$，p 的总质量沿流保持不变。直观而言，演化过程依据再生核希尔伯特空间方向 $D(p)$，在 $\mathcal{O}$ 之间重新分配概率质量，同时不改变归一化。

6.0.1. *H¨older structure and notation.* 样本空间为 $\mathcal{O} = [0,1]^d \times \{0,1\} \times \{0,1\}$，并配备积拓扑。因此，$\mathcal{O}$ 可分解为四个互不相交的紧切片之并

$$\mathcal{O}_{a,y} = [0,1]^d \times \{a\} \times \{y\}, \qquad (a,y) \in \{0,1\}^2,$$

每个切片都同胚于 $[0,1]^d$。因此，定义在 $\mathcal{O}$ 上的函数可视为由四个函数组成的集合，每个函数分别定义在一个切片上。令 $O_k = (X_k, A_k, Y_k)$ 表示一个观测。对于 $o = (x, A_k, Y_k) \in \mathcal{O}_{A_k,Y_k}$ 我们将连续坐标上的欧几里得距离定义为 $\|o - O_k\|_2 := \|x - X_k\|_2$。

固定 $\alpha \in (0,1]$。我们定义巴拿赫空间 $C^{1,\,\alpha}(\mathcal{O}) \cong \bigoplus_{(a,\,y)\in\{0,1\}^2} C^{1,\,\alpha}([0,1]^d)$，它由满足以下条件的函数构成：函数在每个切片 $\mathcal{O}_{a,\,y}$ 上的限制均属于 $C^{1,\,\alpha}([0,1]^d)$。$C^{1,\,\alpha}(\mathcal{O})$上的范数定义为

$$\|u\|_{C^{1,\alpha}(\mathcal{O})} = \max_{(a,y)\in\{0,1\}^2} \|u(\cdot, a, y)\|_{C^{1,\alpha}},$$

其中，对于定义在 $[0,1]^d$ 上的函数 v，

$$\|v\|_{C^{1,\alpha}} = \|v\|_{L^\infty} + \sum_{i=1}^d \|\nabla_i v\|_{L^\infty} + \sum_{i=1}^d [\nabla_i v]_\alpha,$$

而 H¨older 半范数为

$$[\nabla_i v]_\alpha = \sup_{x \neq y} \frac{|\nabla_i v(x) - \nabla_i v(y)|}{\|x - y\|_2^\alpha}.$$

本节中，我们用 $\|\cdot\|_{C^{1,\,\alpha}}$ 表示 H¨older 范数，用 $\|\cdot\|_\infty$ 表示上确界范数，用 $\|\cdot\|_\alpha$ 表示 H¨older 半范数，用 $\|\cdot\|_2$ 表示欧几里得范数；当上下文明确时，省略定义域。

6.1. **常微分方程解的存在唯一性。** 我们证明，对于经验再生核希尔伯特空间得分方程 $P_n D(p_t) \le \delta_n$ 的任意目标精度 $\delta_n > 0$，存在一个有限区间 $I = [0,T]$，使得式 (16) 中的常微分方程在该区间上存在唯一解。此外，该解在某个时刻 $t \in [0,T]$ 达到精度水平 δ_n。这保证了算法流定义良好，并始终属于概率密度类。

主要的技术挑战在于，向量场 $p \mapsto p\, D(p)$ 依赖于数据，并作用于无限维巴拿赫空间 $C^{1,\,\alpha}(\mathcal{O})$。我们的策略是将

方程（16）视为定义在 $C^{1,\,\alpha}(\mathcal{O})$ 的一个有界子集上的自治常微分方程（如方程（17）所示），并验证由此导出的向量场在该集合上是局部利普希茨的。直观而言，核的有界性和平滑性确保基于再生核希尔伯特空间的更新不会在有限时间范围内产生任意粗糙或不稳定的密度。令 p_0 表示初始密度，并设 $M_0 := \|p_0\|_{C^{1,\,\alpha}(\mathcal{O})}$。我们考虑闭集

$$\mathcal{B}_M := \Big\{ p \in C^{1,\alpha}(\mathcal{O}) : p \geq 0,\ \int_{\mathcal{O}} p\, d\mu = 1,\ \|p\|_{C^{1,\alpha}(\mathcal{O})} \leq M \Big\},$$

其中，半径 M 将选取得足够大，以包含截至时间 T 的整个解轨迹。M 和 T 对 δ_n、p_0 及核常数的精确依赖关系见证明（附录 D，第 13.1 节）。

我们在假设 3 中对核施加正则性条件。这些条件确保更新 $D(p)$ 中出现的 RKHS表示元沿轨迹在 $C^{1,\,\alpha}(\mathcal{O})$ 中保持一致有界。这些条件较为宽松，紧域上的高斯核均满足这些条件；相关常数仅取决于带宽 σ、H¨older 指数 α 和 $\mathrm{diam}(\mathcal{O})$。

假设 3。存在有限常数 C、M_K、$M_{1,\,\alpha}$，使得

$$C := \sup_{o} K(o,o) < \infty, \tag{18}$$

$$M_K := \sup_{y \in \mathcal{O}} \|K(\cdot, y)\|_{C^{1,\alpha}} < \infty, \tag{19}$$

$$M_{1,\alpha} := \sup_{p \in \mathcal{B}_M} \sup_{1 \leq j \leq n} \|k_{O_j}^{(p)}\|_{C^{1,\alpha}} < \infty. \tag{20}$$

以下定理确立了 ULFS–KDPE 流的存在性、唯一性和稳定性。具体而言，定理 6.1 表明，方程（16）中的密度值常微分方程存在一个以 p_0 为初值的唯一解；并且在有限时间区间 $I = [0, T]$ 内，该解始终是有效的概率密度，且其 $C^{1,\,\alpha}$ 范数受到控制。证明见附录 D，第 13.1 节。

定理 6.1（常微分方程解的存在唯一性）。设 $p_0 \in C^{1,\,\alpha}(\mathcal{O})$ 为初始密度。假定假设 *3* 成立，并令 $p_0 \in \mathcal{B}_M \subset C^{1,\,\alpha}(\mathcal{O})$。令$F\colon\ \mathcal{B}_M \to C^{1,\,\alpha}(\mathcal{O})$ 由 $F(p) := p\, D(p)$ 定义。则（16）中的密度值常微分方程，在 $p_{t=0} = p_0$ 下，

$$\frac{d}{dt} p_t = F(p_t),$$

存在唯一解 $t \mapsto p_t \in C^1([0,T];\ C^{1,\,\alpha}(\mathcal{O}))$。对所有 $t \in [0,T]$，$p_t \in \mathcal{B}_M$。

6.2. **经验得分的有限时间收敛**。以下结果确立了RKHS受限流的一个关键结构性质。再生核希尔伯特空间的更新方向 $D(p_t)$ 为该流诱导出李雅普诺夫结构。特别地，沿任意充分正则的解，经验对数似然 $P_n[\log p_t]$ 单调不减，且平稳性出现当且仅当经验均值嵌入消失。该结果既证明了选择 $D(p_t)$的合理性，又提供了一个有原则的停止准则，将常微分方程的平衡点与 $\mathcal{H}_K^{(t)}$ 上经验估计方程的解联系起来。证明见附录 D，第 13.2 节。

引理 6.2（经验对数似然的单调性和平稳性）。设$t \mapsto p_t$ 是微分方程的一个充分正则解

$$\frac{d}{dt} \log p_t(o) = D(p_t)(o).$$

满足 $D(p_t) \coloneqq \frac{1}{n}\sum_{j=1}^{n}\alpha_j^{(t)}k_{O_j}^{(t)}$。则有以下结论：

（*1*）映射 $t \mapsto P_n[\log p_t]$ 在 I 上非递减。（*2*）对任意 $t_1 \in I$,

$$\frac{d}{dt}P_n[\log p_t]\Big|_{t=t_1} = 0 \quad \Longleftrightarrow \quad m_n^{(t_1)} = 0 \ in\ \mathcal{H}_K^{(t_1)}.$$

定理 6.3 保证 ULFS–KDPE 流在有限时间内达到算法停止准则。特别地，经验得分 $P_nD(p_t)$ 不能在整个区间 $[0,T]$ 上始终一致地高于目标容差 δ_n。这确保 ULFS–KDP 估计量是良定义的，且停止规则并非仅仅是启发式的。证明安排在附录 D，第 13.2 节。

定理 6.3. 存在 $t \in [0,T]$，使得 $P_nD(p_t) \le \delta_n$。

6.3. **渐近线性与效率**。最后，我们建立 ULFS–KDPE 估计量的渐近线性与效率。核心要求是再生核希尔伯特空间足够丰富，以逼近典范梯度，同时在经验上保持良好性质。假设 4 将这一条件形式化。在假设 4 下，定理 6.4 表明，求解经验得分方程可得到具有影响函数 $\phi_{P^*}^*$ 的渐近线性估计量，从而具有半参数效率。证明安排在附录 D，第 13.3 节。

假设 4. 存在一个事件 Ω' 满足 $P^*(\Omega') = 1$，使得对于每个$\omega \in \Omega'$，都存在一个序列 $\{h_j(\omega)\}_{j\ge1} \subset \mathcal{H}_K^{(t_n(\omega))}(\omega)$ 满足：

(i) $\|h_j(\omega) - \phi^*_{P_{t_n(\omega)}}(\omega)\|_{L^2(P^*)} \to 0$ 当 $j \to \infty$ 时；*(ii)* $\sup_{j\ge1}\|h_j(\omega)\|_{\mathcal{H}_K^{(t_n(\omega))}} < \infty$；*(iii)* 存在 $j_0(\omega) < \infty$，使得类 $\{h_j(\omega) : w \in \Omega', j \ge j_0(\omega)\}$ 是P^*-*Donsker*。

我们选取 $\delta_n = o(n^{-2})$，并令 $t_n \in [0,T]$ 满足 $P_nD(p_{t_n}) \le \delta_n$。于是，得到以下结果。

定理 6.4。设 $\Psi : \mathcal{M} \to \mathbb{R}$ 为分布 P的路径可微泛函，其典范梯度为 $\phi_P^* \in L_0^2(P)$。设 $t_n \in [0,T]$满足 $P_nD(p_{t_n}) \le \delta_n$，且 $P_n := P_{t_n}$。在假设 *1*、*2*、*3*、*4*、$P_n\phi^*_{P_{t_n}} = o_{P^*}(n^{-1/2})$ 下，*ULFS–KDPE* 估计量满足

$$\Psi(\widehat{P}_n) - \Psi(P^*) = P_n\phi^*_{P^*} + o_{P^*}(n^{-1/2}).$$

7. Simulation Results

我们开展了一项模拟研究，以展示 ULFS–KDPE 估计量的实现方法、验证理论结果，并评估其在两种数据生成过程（DGP）下的有限样本表现。我们评估偏差、方差和去偏程序的稳定性，并比较第 5.1.3 节所述的五种停止准则。所有实验均进行 500 次蒙特卡洛重复，样本量为 $n = 300$，使用高斯零均值核、步长 $\Delta = 0.01$，最多迭代 100 次。估计时采用密度稳定化，下界为 $c = 10^{-3}$，容差为$\lambda = 10^{-8}$。

7.1. 数据生成过程与目标参数。

7.1.1. 数据生成过程 *1*：具有二元结局的观察性研究。我们定义 $\mathcal{O} = \mathcal{X} \times \mathcal{A} \times \mathcal{Y}$，其中包含基线协变量 $X \in \mathcal{X} \equiv [0,1]$、二元处理 $A \in \mathcal{A} \equiv \{0,1\}$和二元结局 $Y \in \mathcal{Y} \equiv \{0,1\}$。真实的数据生成过程如下：

$$X \sim \mathrm{Unif}(0,1), \quad A|X \sim \mathrm{Bern}(0.5 + \frac{1}{3}\sin(50X/\pi))$$
$$Y|A,X \sim \mathrm{Bern}(0.4 + A(X-0.3)2 + \frac{1}{4}\sin(40X/\pi))$$

7.1.2. 数据生成过程 *2*：存在正值性问题和二元结局的观察性研究。我们定义 $\mathcal{O} = \mathcal{X} \times \mathcal{A} \times \mathcal{Y}$，其中包含基线协变量 $X \in \mathcal{X} \equiv [-1,1]$、二元处理 $A \in \mathcal{A} \equiv \{0,1\}$和二元结局 $Y \in \mathcal{Y} \equiv \{0,1\}$。真实的数据生成过程如下：

$$\begin{aligned} &X \sim 0.9 \cdot \mathrm{Unif}(-1,1) + 0.1 \cdot \mathrm{Unif}(-2,2), \\ &A|X \sim \mathrm{Bern}(\mathrm{expit}(4X)), \\ &Y|A,X \sim \mathrm{Bern}(\mathrm{expit}(-0.5 + A + 0.5X)). \end{aligned}$$

我们注意到，所提出的数据生成过程中存在正值性问题。具体而言，X 接近±2会导致 e^* 非常接近 0 或 1。

7.1.3. 目标参数。我们关注几个典型的因果参数。我们将处理水平 $a \in \{0,1\}$ 下的平均潜在结局定义为

$$\begin{aligned} \mu_a(P) := \mathbb{E}_P\big[Y^a\big] &= P_X\big[\mathbb{E}_P\big[Y \mid A=a, X\big]\big] \\ &= \int \bar{Q}_P(a,x)\, dP_X(x), \end{aligned}$$

¯其中 $Q_P(a,\ x) := \mathbb{E}_P[Y \mid A=a, X=x]$。我们考虑以下参数：(1) 平均处理效应（ATE），$\psi_{\mathrm{ATE}}(P) := \mu_1(P) - \mu_0(P)$；(2) 风险比（RR），$\psi_{\mathrm{RR}}(P) := \mu_1(P)/\mu_0(P)$；以及 (3) 优势比（OR），$\psi_{\mathrm{OR}}(P) := {}^{\mu_1}(P)/(1-\mu_1(P))\,\mu_0(P)/(1-\mu_0(P))$。

7.1.4. 模拟设置。对于所有数据生成过程（DGPs），我们将基线协变量 X 的分布初始化为 $P_n(X)$，并在定向步骤中将 X 的经验分布视为固定。其余条件分布使用 Super Learner 框架 [26]，进行估计，该框架在 R 中通过 sl3 软件包 [11] 实现。候选学习器库包括样本均值、广义线性模型（GLM）、随机森林 [4],，以及通过 XGBoost [6] 实现的梯度提升。

对于两个数据生成过程，我们将提出的 ULFS-KDPE 与以下方法进行比较：(i) 原始迭代 KDPE，该方法依赖于局部最不利子模型 [9]；(ii) 目标最大似然估计（TMLE） [25] ；以及 (iii) 一步 TMLE [23]。TMLE 和一步 TMLE 使用 R 中的 tmle3 软件包 [10] 实现。由于当前的 tmle3 实现不支持风险比 ψ_{RR} 或优势比 ψ_{OR} 的单步估计，因此一步 TMLE 仅用于估计平均处理效应 ψ_{ATE}。TMLE 和一步 TMLE 均显式使用目标参数的有效影响函数，并且已知在标准正则性条件下能够达到半参数效率界。

7.2. **经验性能**。 在所有数据生成过程中，模拟结果均有力支持 ULFS–KDPE 的理论性质，并展现出明显的实际优势。在结局为二元变量且表现良好的观察性环境（DGP1）中，ULFS–KDPE 的表现符合一阶效率，其经验$\widehat{}$分布 $\psi_{\mathrm{ATE}}($ $P_{\text{ULFS-KDPE}})$与渐近正态极限高度吻合。与 TMLE 和一步 TMLE 相比，ULFS–KDPE 在表1所列的所有数据生成过程中均取得更小的均方根误差，表明正则化流带来了更有利的偏差–方差权衡。与（迭代）KDPE 相比，ULFS–KDPE 在几乎所有情况下都显著降低了偏差，同时保持了相近的方差。其一项关键优势是，单个 ULFS–KDPE 分布可用于同时估计多个路径可微参数。这一点在非线性目标参数（如 ψ_{RR} 和 ψ_{OR}）中尤为明显：尽管 TMLE 需要分别进行目标化步骤，ULFS–KDPE 得到的偏差和均方根误差仍低于 TMLE。

ULFS–KDPE 的优势在正性违背（DGP2）情况下尤为明显。在这一更具挑战性的设置下，ULFS–KDPE 在所有目标参数上均持续优于 TMLE 和一步 TMLE，表现出显著更低的方差，并且与极限分布更加一致。对于 ψ_{RR} 和 ψ_{OR}等非线性目标参数，ULFS–KDPE 实现了最小的偏差和均方根误差，凸显了 RKHS约束的通用最不利流和密度正则化在基于EIF的方法遭受方差膨胀时所具有的稳定化作用。

除统计性能外，与原始的（迭代式）KDPE 相比，ULFS–KDPE 还表现出显著改善的数值稳定性。在更多模拟中，ULFS–KDPE 能够在迭代上限内收敛，体现了其微步更新方案的稳健性。该方案将全局定义的最不利流离散化，并避免了可能阻碍局部定义的 KDPE 更新收敛的过冲行为。如表3所示，对于两个数据生成过程，当将迭代上限适度提高到 150 时，ULFS–KDPE 在所有模拟中均实现收敛。

停止准则的表现进一步印证了 ULFS–KDPE 的算法稳定性。如表 2 所示，与 ULFP 流几何结构相契合的停止规则在所有数据生成过程中都最为可靠。尤其是，密度稳定化规则利用经验对数似然的单调性，在后续更新不再显著降低经验偏差时终止程序。相比之下，基于局部得分条件或显式 EIF 求解的停止规则对有限样本不稳定性更为敏感，尤其是在处理非线性目标参数以及存在正性违背的情况下。基于方差的停止规则较为保守，往往会较晚终止。

总体而言，这些结果表明，ULFS–KDPE 在正则情形下实现了一阶效率，在具有挑战性的正性条件情形下显著降低了有限样本方差，能够基于单个分布同时对多个目标参数进行去偏，并且在配合通用最不利流内在的停止规则时，可提升数值稳定性。

TABLE 1. 不同估计量（ULFS-KDPE、KDPE、目标最大似然估计、一步TMLE）下平均处理效应、风险比和优势比的偏差、100步收敛、方差和均方根误差（均方根误差）。

	参数	方法	协变量数	偏差(×100)	方差[a]	均方根误差
DGP1	ψ_{ATE}	ULFS-KDPE	475	-0.824	0.0561	0.0567
		KDPE	343	-1.438	0.0546	0.0565
		TMLE	500	-0.004	0.0618	0.0618
		一步 TMLE	500	-0.247	0.0605	0.0606
	ψ_{RR}	ULFS-KDPE	475	-1.082	0.1724	0.1728
		KDPE	343	-3.213	0.1584	0.1616
		TMLE	500	1.694	0.1919	0.1928
	ψ_{OR}	ULFS-KDPE	475	-0.499	0.3951	0.3954
		KDPE	343	-5.010	0.3683	0.3717
		TMLE	500	6.193	0.4464	0.4510
DGP2	ψ_{ATE}	ULFS-KDPE	500	-0.779	0.0772	0.0777
		KDPE	389	0.034	0.0750	0.0751
		TMLE	500	0.128	0.1203	0.1204
		一步 TMLE	500	0.009	0.1176	0.1177
	ψ_{RR}	ULFS-KDPE	500	0.427	0.2777	0.2779
		KDPE	389	3.123	0.2766	0.2784
		TMLE	500	8.663	0.4599	0.4683
	ψ_{OR}	ULFS-KDPE	500	8.729	0.9505	0.9552
		KDPE	389	17.575	0.9621	0.9784
		TMLE	500	49.752	1.6948	1.7674

[a] 方差表示估计量在 B 个模拟数据集上的蒙特卡洛方差，计算方式为 $\frac{1}{B}\sum_{b=1}^{B}(\widehat{\psi}_b - \bar{\psi})^2$，，其中 $\bar{\psi} = \frac{1}{B}\sum_{b=1}^{B}\widehat{\psi}_b$。

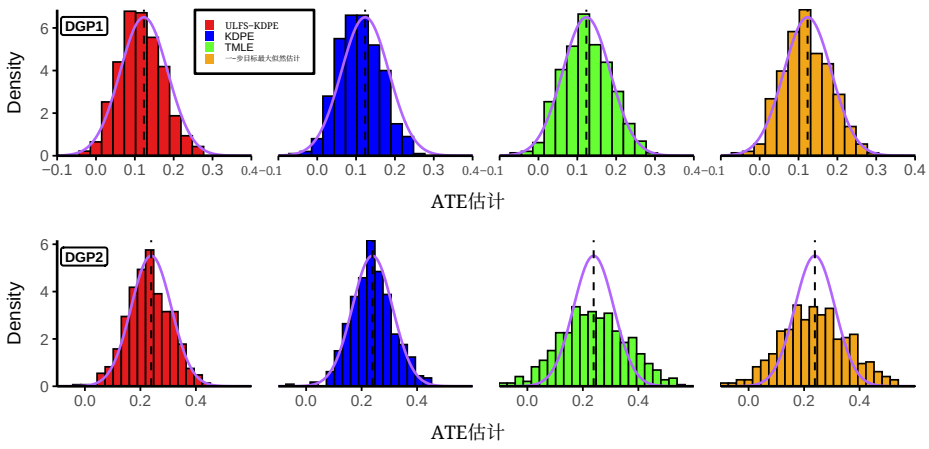

FIGURE 1. 不同平均处理效应估计量（ULFS-KDPE、KDPE、目标最大似然估计、One-stepTMLE）的有限样本表现。紫色表示真实的渐近分布。

TABLE 2. 在不同停止准则下，ULFS-KDPE的ATE、RR和OR的偏差、200步时的收敛情况、方差及均方根误差。

	参数	停止准则	协变量数	偏差（×100）	方差[a]	均方根误差
DGP1	ψ_{ATE}	密度稳定化	500	-0.840	0.0553	0.0560
		经验得分稳定化	493	-0.806	0.0566	0.0572
		消失的更新方向	500	-1.357	0.0547	0.0564
		方差主导的更新	500	-1.573	0.0550	0.0573
		有效影响函数近似求解	447	-1.156	0.0561	0.0573
	ψ_{RR}	密度稳定化	500	-1.142	0.1700	0.1705
		经验得分稳定化	493	-0.985	0.1738	0.1742
		消失的更新方向	500	-2.690	0.1647	0.1670
		方差主导的更新	500	-3.274	0.1651	0.1684
		有效影响函数近似求解	447	-2.070	0.1698	0.1712
	ψ_{OR}	密度稳定化	500	-0.727	0.3893	0.3896
		经验得分稳定化	493	-0.274	0.4002	0.4005
		消失的更新方向	500	-4.301	0.3759	0.3786
		方差主导的更新	500	-5.648	0.3758	0.3803
		有效影响函数近似求解	447	-2.750	0.3897	0.3909
DGP2	ψ_{ATE}	密度稳定化	500	-0.779	0.0772	0.0777
		经验得分稳定化	499	-0.511	0.0770	0.0772
		消失的更新方向	497	-0.981	0.0779	0.0785
		方差主导的更新	500	-1.240	0.0780	0.0790
		有效影响函数近似求解	306	-1.269	0.0786	0.0797
	ψ_{RR}	密度稳定化	500	0.427	0.2777	0.2779
		经验得分稳定化	499	1.374	0.2788	0.2794
		消失的更新方向	497	-0.238	0.2784	0.2786
		方差主导的更新	500	-1.114	0.2779	0.2783
		有效影响函数近似求解	306	-1.069	0.2797	0.2799
	ψ_{OR}	密度稳定化	500	8.729	0.9505	0.9552
		经验得分稳定化	499	11.805	0.9572	0.9650
		消失的更新方向	497	6.629	0.9504	0.9533
		方差主导的更新	500	3.712	0.9441	0.9455
		有效影响函数近似求解	306	3.620	0.9419	0.9427

[a] $\frac{1}{B}\sum_{b=1}^{B}(\hat{\psi}_b - \bar{\psi})^2$, 其中 $\bar{\psi} = \frac{1}{B}\sum_{b=1}^{B}\hat{\psi}_b$。

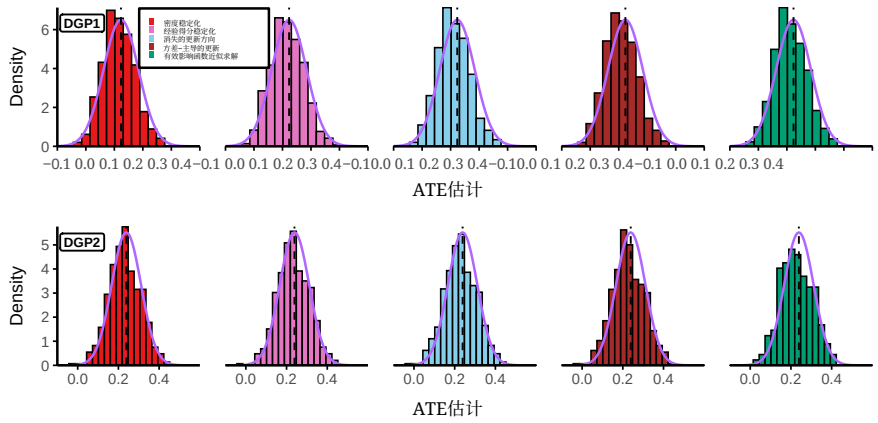

FIGURE 2. 不同停止准则下的 ULFS-KDPE。真实渐近分布以紫色表示。

TABLE 3. 不同迭代上限（允许的步数）下，ULFS-KDPE 对平均处理效应、风险比和优势比的收敛情况及均方根误差。

	迭代上限	停止准则	协变量数	ψ_{ATE}	ψ_{RR}	ψ_{OR}
DGP1	100	密度平台	475	0.0567	0.1728	0.3975
		得分平台	404	0.0573	0.1676	0.3792
		消失方向	498	0.0563	0.1669	0.3784
		方差占主导	500	0.0567	0.1684	0.3803
		EIF 已求解	407	0.0547	0.1721	0.3943
	150	密度平台	500	0.0560	0.1702	0.3896
		得分平台	472	0.0569	0.1691	0.3900
		消失方向	500	0.0564	0.1650	0.3786
		方差占主导	500	0.0573	0.1654	0.3803
		EIF 已求解	435	0.0576	0.1711	0.3930
	200	密度平台	500	0.0560	0.1702	0.3896
		得分平台	493	0.0572	0.1741	0.4005
		消失方向	500	0.0564	0.1650	0.3786
		方差占主导	500	0.0573	0.1654	0.3803
		EIF 已求解	447	0.0573	0.1701	0.3909
DGP2	100	密度平台	500	0.0777	0.2779	0.9552
		得分平台	197	0.0838	0.2942	1.0551
		消失方向	484	0.0786	0.2775	0.9501
		方差占主导	500	0.0790	0.2783	0.9455
		EIF 已求解	251	0.0818	0.2844	0.9553
	150	密度平台	500	0.0777	0.2779	0.9552
		得分平台	415	0.0782	0.2757	0.9476
		消失方向	492	0.0783	0.2769	0.9458
		方差占主导	500	0.0790	0.2783	0.9455
		EIF 已求解	286	0.0806	0.2820	0.9509
	200	密度平台	500	0.0777	0.2779	0.9552
		得分平台	499	0.0772	0.2794	0.9650
		消失方向	497	0.0785	0.2786	0.9533
		方差占主导	500	0.0790	0.2783	0.9455
		EIF 已求解	306	0.0797	0.2799	0.9427

8. Discussion and Future Work

本文提出了一种基于通用最不利子模型的核去偏插件估计量（ULFS–KDPE），为有效半参数估计提供了一个统一且具有坚实理论基础的框架，既无需影响函数，又具有良好的计算可行性。该方法将去偏嵌入再生核希尔伯特空间（RKHS），并通过通用最不利流定义更新过程；它由此摆脱了经典的局部定向方法，转而沿单一分布路径在全局范围内实现最不利性。

从理论角度看，主要贡献在于为 ULFS–KDPE 奠定了严格的泛函分析基础。我们将通用最不利更新表述为定义在密度上的非线性常微分方程，并证明其解在适当的 H¨older 空间中具有存在性、唯一性和稳定性，且能在有限时间内收敛。这些结果保证了所提出的流是适定的，能够保持正性和归一化，并在有限时间内到达经验再生核希尔伯特空间

得分函数充分小的点。重要的是，该构造产生的插件估计量在标准条件下是正则、渐近线性且半参数有效的，无需显式推导有效影响函数。此外，对于典范梯度属于 RKHS 在 $L^2(P_0)$下的闭包的所有路径可微参数（包括多变量目标），该估计量均能同时达到效率。

从实践角度来看，ULFS–KDPE 兼具通用最不利子模型和基于核的去偏方法的优势。路径的全局最不利性避免了迭代局部定向可能导致的收敛病态，而再生核希尔伯特空间（RKHS）表示则将更新简化为有限维计算，只需计算观测数据点处的核函数取值。该流的李雅普诺夫结构为更新提供了几何解释：更新可视为一种稳定化梯度流，它能单调提高经验对数似然，并随着经验得分趋于零而接近平稳状态。模拟结果支持这些理论性质，并表明与局部定向方法相比，该方法在有限样本下具有更好的稳定性和准确性，尤其是在重叠不足等具有挑战性的情形中。

这项工作引出了若干未来研究方向。首先，需要对停止准则进行更深入的理论分析。尽管当前准则的依据是李雅普诺夫结构和经验得分方程，但若能通过形式化结果刻画这些准则对渐近线性、效率和有限样本偏差的影响，将有助于增强 ULFS–KDPE 的推断保证。尤其是，开发具有明确收敛速率且能保证推断有效的数据自适应停止规则，是下一步的重要工作。其次，有必要进一步研究离散化方案。我们的实现采用 ULFS 常微分方程的显式欧拉离散化，这种方法简单且稳定，但未必是最优的。离散化误差、核函数平滑性和收敛速率之间的相互作用仍是一个有待解决的问题。

将该框架扩展至高阶推断是另一个颇具前景的方向。尽管 ULFS–KDPE 达到了一阶半参数效率，但基于通用再生核希尔伯特空间的得分方程表明，该方法有望捕捉影响函数 [16] 的高阶分量。沿通用最不利流，系统地发展二阶或更高阶展开理论，有望改善中等样本下的推断，并为有效估计的几何结构提供新的见解。最后，可扩展性与近似方法也值得进一步研究。随机特征近似、低秩核方法或局部化核，或许能使 ULFS–KDPE 扩展至更大型的数据集，同时保留其理论性质。探索这些方向将拓宽该方法的实际适用范围，并进一步弥合坚实的半参数理论与现代数据密集型应用之间的差距。

9. 致谢

感谢 Patrick Lopatto 就这一主题提出的诸多意见并参与讨论。

10. Appendix A

命题 10.1（期望作为 RKHS 泛函）。设 P 是 $\mathcal{O}$ 上的任意概率测度， $\mathcal{H}_K$ 是与高斯核 K 相关联的再生核希尔伯特空间。那么，核均值嵌入 $m_P := \int k_o \, dP$（o）存在于 $\mathcal{H}_K$ 中。此外，对每个 $f \in \mathcal{H}_K$，

$$P[f] = \langle f, m_P \rangle_{\mathcal{H}_K}.$$

因此，映射 $f \mapsto P[f]$ 是 $\mathcal{H}_K$上的连续线性泛函，满足

$$|P[f]| \le \|f\|_{\mathcal{H}_K}.$$

证明。 由于对所有 $o \in \mathcal{O}$， $K(o,o) \le 1$ ，我们有

$$\|k_o\|^2_{\mathcal{H}_K} = \langle k_o, k_o \rangle_{\mathcal{H}_K} = K(o,o) \le 1,$$

因此，对所有 $o \in \mathcal{O}$都有 $\|k_o\|_{\mathcal{H}_K} \le 1$。由此可得 $\int \|k_o\|_{\mathcal{H}_K} \, dP(o) \le 1 < \infty$。因此， $\mathcal{H}_K$值 Bochner 积分 $m_P := \int k_o \, dP(o)$ 定义良好。

设 $f \in \mathcal{H}_K$。根据再生性质，

$$f(o) = \langle f, k_o \rangle_{\mathcal{H}_K} \qquad \text{for all } o \in \mathcal{O}.$$

因此，

$$P[f] = \int f(o) \, dP(o) = \int \langle f, k_o \rangle_{\mathcal{H}_K} \, dP(o).$$

由于映射 $g \mapsto \langle f,g \rangle_{\mathcal{H}_K}$ 是 $\mathcal{H}_K$ 上的有界线性泛函，因此它与 Bochner 积分可交换，从而得到

$$\int \langle f, k_o \rangle_{\mathcal{H}_K} \, dP(o) = \left\langle f, \int k_o \, dP(o) \right\rangle_{\mathcal{H}_K} = \langle f, m_P \rangle_{\mathcal{H}_K}.$$

Thus

$$P[f] = \langle f, m_P \rangle_{\mathcal{H}_K}.$$

最后，根据柯西–施瓦茨不等式，

$$|P[f]| = |\langle f, m_P \rangle_{\mathcal{H}_K}| \le \|f\|_{\mathcal{H}_K} \, \|m_P\|_{\mathcal{H}_K}.$$

此外，

$$\|m_P\|_{\mathcal{H}_K} = \left\| \int k_o \, dP(o) \right\|_{\mathcal{H}_K} \le \int \|k_o\|_{\mathcal{H}_K} \, dP(o) \le 1.$$

因此

$$|P[f]| \le \|f\|_{\mathcal{H}_K}.$$

这证明了 $f \mapsto P[f]$ 是 $\mathcal{H}_K$上的连续线性泛函。 □

命题 10.2 (均值为零的再生核希尔伯特空间作为闭子空间)。 对于每个 $t \in I$，空间 $\mathcal{H}_K^{(t)}$ 是 $\mathcal{H}_K$的闭线性子空间。此外，

$$\mathcal{H}_K^{(t)} = \{ f \in \mathcal{H}_K : \langle f, m_{P^{(t)}} \rangle_{\mathcal{H}_K} = 0 \} = (\operatorname{span}\{ m_{P^{(t)}} \})^{\perp}.$$

特别地， $\mathcal{H}_K^{(t)}$ 在继承自 $\mathcal{H}_K$的内积下是一个希尔伯特空间。

证明。 根据命题 4.1，对于每个 $f \in \mathcal{H}_K$ ，我们有

$$P^{(t)}[f] = \langle f, m_{P^{(t)}} \rangle_{\mathcal{H}_K}.$$

因此，约束 $P_{(t)}[f] = 0$等价于与 $m_{P^{(t)}}$ 正交。由于映射 $f \mapsto \langle f, m_{P^{(t)}} \rangle_{\mathcal{H}_K}$ 是 $\mathcal{H}_K$ 上的连续线性泛函，其零空间

$$\{ f \in \mathcal{H}_K : \langle f, m_{P^{(t)}} \rangle_{\mathcal{H}_K} = 0 \}$$

是 $\mathcal{H}_K$ 的闭线性子空间。这就证明了该论断。 □

命题 10.3（投影到均值为零的再生核希尔伯特空间上的正交投影）。 若 $m_{P^{(t)}} \neq 0$，正交投影 Π_t：$\mathcal{H}_K \to \mathcal{H}_K^{(t)}$ 由下式给出

$$\Pi_t f = f - \frac{\langle f, m_{P^{(t)}} \rangle_{\mathcal{H}_K}}{\|m_{P^{(t)}}\|^2_{\mathcal{H}_K}} m_{P^{(t)}} = f - \frac{P^{(t)}[f]}{\|m_{P^{(t)}}\|^2_{\mathcal{H}_K}} m_{P^{(t)}}, \qquad f \in \mathcal{H}_K.$$

证明。 由于 $\mathcal{H}_K^{(t)} = (\text{span}\{m_{P^{(t)}}\})^{\perp}$，将 $f \in \mathcal{H}_K$ 正交投影到$\mathcal{H}_K^{(t)}$ 上，可通过减去其沿 $m_{P^{(t)}}$方向的分量得到。记

$$\Pi_t f = f - c\, m_{P^{(t)}}$$

并施加 $\Pi_t f \in \mathcal{H}_K^{(t)}$ ，得到

$$0 = \langle f - c\, m_{P^{(t)}}, m_{P^{(t)}} \rangle_{\mathcal{H}_K} = \langle f, m_{P^{(t)}} \rangle_{\mathcal{H}_K} - c\|m_{P^{(t)}}\|^2_{\mathcal{H}_K}.$$

解出 c ，得到

$$c = \frac{\langle f, m_{P^{(t)}} \rangle_{\mathcal{H}_K}}{\|m_{P^{(t)}}\|^2_{\mathcal{H}_K}},$$

由此得到所述公式。 □

命题 10.4（均值为零的再生核希尔伯特空间中的经验黎斯表示元）。 存在唯一一个元素 $m_n^{(t)} \in \mathcal{H}_K^{(t)}$ 使得

$$P_n[f] = \langle f, m_n^{(t)} \rangle_{\mathcal{H}_K^{(t)}} \qquad \textit{for all } f \in \mathcal{H}_K^{(t)}.$$

此外，

$$m_n^{(t)} = \frac{1}{n} \sum_{i=1}^{n} k_{O_i}^{(t)}.$$

证明。 L_n 的线性由求和的线性立即可得。为验证有界性，注意到对于任意 $f \in \mathcal{H}_K^{(t)} \subset \mathcal{H}_K$ 和任意 $o \in \mathcal{O}$，再生性质给出 $f(o) = \langle f, k_o \rangle_{\mathcal{H}_K}$。由于高斯核满足 $K(o,o) \leq 1$，因此有 $\|k_o\|_{\mathcal{H}_K} \leq 1$，从而

$$|f(o)| \leq \|f\|_{\mathcal{H}_K}.$$

将此界应用于每个观测 O_i ，得到

$$|P_n[f]| \leq \frac{1}{n} \sum_{i=1}^{n} |f(O_i)| \leq \|f\|_{\mathcal{H}_K}.$$

因此， L_n 是希尔伯特空间 $\mathcal{H}_K^{(t)}$ 上的有界线性泛函，而里斯表示定理保证存在唯一的 $m_n^{(t)} \in \mathcal{H}_K^{(t)}$ 满足

$$P_n[f] = \langle f, m_n^{(t)} \rangle_{\mathcal{H}_K^{(t)}}.$$

该显式表达式由 $\mathcal{H}_K^{(t)}$ 中的再生性质得出。 □

11. Appendix B

算法 1 ULFS–KDPE（ $\mathcal{O}$上的离散化RKHS受限流） 要求：数据 $O_1, \dots, O_n$；初始密度估计 $\widehat{p}_0$； $\mathcal{O}$上的核 K；步长$\Delta > 0$；容差 $\delta_p, \delta_s, \delta_\alpha, \delta_v, \delta_\ell$；最大迭代次数 M.^ ^ ^确保：最终迭代 P_T（密度 $\widehat{p}_T$）和插件估计 $\psi_T := \Psi(P_T)$。 1：设置 $t \leftarrow 0$ 和 $\widehat{p}_t \leftarrow \widehat{p}_0$。 2：初始化 $\Delta_t^{(p)} \leftarrow +\infty$, $\Delta_t^{(\ell)} \leftarrow +\infty$, $s_t \leftarrow +\infty$。 3：对于 $m = 0, 1, \dots, M-1$，执行⌢4：构造当前迭代 P_t （密度为 $\widehat{p}_t$）。 5：以 $\widehat{P}_t$ 为中心化核，得到 $\mathcal{O}$上的零均值 RKHS核 $K^{(t)}$ （等价地，通过均值嵌入 $m_{\widehat{P}_t} := \int k_o \, d\widehat{P}_t(o)$）。 6：构造中心化格拉姆矩阵 $G^{(t)} := \left[K^{(t)}(O_i, O_j)\right]_{i,j=1}^n$。 7：计算 $\boldsymbol{\alpha}^{(t)} \leftarrow \frac{1}{n} G^{(t)} \mathbb{1}$，因此 $\alpha_j^{(t)} = \mathbb{P}_n[k_{O_j}^{(t)}] = m_n^{(t)}(O_j)$。 8：定义RKHS方向函数 $D_t(o) := D(\widehat{p}_t)(o) := \frac{1}{n} \sum_{j=1}^n \alpha_j^{(t)} k_{O_j}^{(t)}(o)$，以及其在样本上的取值 $\mathbf{d}^{(t)} \leftarrow \left(D_t(O_i)\right)_{i=1}^n = \frac{1}{n} G^{(t)} \boldsymbol{\alpha}^{(t)}$。 9：计算李雅普诺夫得分⌣ $s_t \leftarrow \mathbb{P}_n[D_t] = \frac{1}{n} \sum_{j=1}^n \left(\alpha_j^{(t)}\right)^2 = \frac{1}{n} \|\boldsymbol{\alpha}^{(t)}\|_2^2$。 10：如果 $m \geq 1$，则 11：计算平台诊断： $\Delta_t^{(p)} \leftarrow \mathbb{P}_n\left[\left(\log \widehat{p}_t(O) - \log \widehat{p}_{t-\Delta}(O)\right)^2\right]$， 12：$\Delta_t^{(\ell)} \leftarrow \left|\mathbb{P}_n[\log \widehat{p}_t(O) - \log \widehat{p}_{t-\Delta}(O)]\right|$。 13：结束如果 14：如果满足§5.1.3中的停止准则，则 15：跳出 16：结束如果 17：通过对数密度上的显式欧拉步更新密度： 18：$\widehat{p}_{t+\Delta}(o) \leftarrow \widehat{p}_t(o) \exp\left(\Delta \, D_t(o)\right)$。 19：归一化： $\widehat{p}_{t+\Delta}(o) \leftarrow \widehat{p}_{t+\Delta}(o) \Big/ \int \widehat{p}_{t+\Delta}(u) \, d\mu(u)$。 20：设置 $t \leftarrow t + \Delta$ 并覆盖 $\widehat{p}_t \leftarrow \widehat{p}_{t+\Delta}$。 21：结束循环 ^ ^ ⌢22：设置 $T \leftarrow t$，并输出密度为 $\widehat{p}_T$ 的 $P_T := \widehat{p}_t$ 以及 $\psi_T := \Psi(P_T)$。

12. Appendix C

12.1. H 的乘积估计与巴拿赫代数结构¨older 空间。为了在巴拿赫空间框架下证明方程（16）中的非线性常微分方程具有适定性，必须验证 H¨older 空间 $C^{1,\alpha}(\mathcal{O})$ 在乘法下保持稳定，并且相应的范数可以进行定量控制。下面的第一个结果——引理 12.1——给出了函数乘积的基本 H¨older 半范数估计，表明乘法不会使振荡程度超过各个因子本身已有的振荡程度。在此基础上，引理 12.2 证明了 $C^{1,\alpha}(\mathcal{O})$ 是一个巴拿赫代数：两个 $C^{1,\alpha}$ 函数的乘积仍属于 $C^{1,\alpha}$，且其范数界显式地依赖于定义域和 H¨older 指数。这些估计共同构成了控制非线性向量场 $p \mapsto pD(p)$，并应用巴拿赫空间中常微分方程标准存在唯一性结果所需的分析基础。

引理 12.1. 对于 $g, h \in C^{0,\alpha}(\mathcal{O})$,

$$[gh]_\alpha \le \|g\|_\infty [h]_\alpha + \|h\|_\infty [g]_\alpha$$

证明。 对于任意不同的 $x, y \in [0,1]^d$ 以及固定的 $b \in \{0,1\}^2$,

$$g(x,b)h(x,b) - g(y,b)h(y,b) = g(x,b)\big(h(x,b) - h(y,b)\big) + h(y,b)\big(g(x,b) - g(y,b)\big).$$

对于切片 b，取绝对值并除以 $\|x-y\|_2^\alpha$，得到

$$\frac{|g(x)h(x) - g(y)h(y)|}{\|x-y\|_2^\alpha} \le |g(x)| \frac{|h(x)-h(y)|}{\|x-y\|_2^\alpha} + |h(y)| \frac{|g(x)-g(y)|}{\|x-y\|_2^\alpha}.$$

现在对所有 $x \neq y \in [0,1]^d$ 取上确界，可得

$$[gh]_\alpha \le \|g\|_\infty [h]_\alpha + \|h\|_\infty [g]_\alpha$$

对于任意切片 b。 □

下一个引理可应用于任意固定切片 $b \in \{0,1\}^2$。

引理 12.2. 设 $\mathcal{O} \subset \mathbb{R}^d$ 为凸且有界，且设 $\alpha \in (0,1]$。则存在常数 $C_{\mathrm{alg}} < \infty$ 使得

$$\|fg\|_{C^{1,\alpha}(\mathcal{O})} \le C_{\mathrm{alg}} \|f\|_{C^{1,\alpha}(\mathcal{O})} \|g\|_{C^{1,\alpha}(\mathcal{O})} \qquad \forall f, g \in C^{1,\alpha}(\mathcal{O}).$$

此外，可以采用显式界

$$C_{\mathrm{alg}} = 1 + 2\,\mathrm{diam}(\mathcal{O})^{\,1-\alpha}, \quad \mathrm{diam}(\mathcal{O}) := \sup_{x,y\in\mathcal{O}} \|x-y\|_2.$$

特别地，对于 $\mathcal{O} = [0,1]^d$,

$$\mathrm{diam}(\mathcal{O}) = \sqrt{d} \quad \textit{and thus} \quad C_{\mathrm{alg}} = 1 + 2\,d^{\frac{1-\alpha}{2}}.$$

证明。 固定 $f, g \in C^{1,\alpha}(\mathcal{O})$，记

$$A := \|f\|_\infty, B := \|\nabla f\|_\infty, C := [\nabla f]_\alpha, A' := \|g\|_\infty, B' := \|\nabla g\|_\infty, C' := [\nabla g]_\alpha.$$

根据 $\|\cdot\|_\infty$ 的定义,

$$\|fg\|_\infty \le \|f\|_\infty \|g\|_\infty = AA'.$$

利用乘积法则 $\nabla(fg) = f\,\nabla g + g\,\nabla f$,

$$\|\nabla(fg)\|_\infty \le \|f\|_\infty \|\nabla g\|_\infty + \|g\|_\infty \|\nabla f\|_\infty = AB' + A'B.$$

对于 $x,\ y \in [0,1]^d, x \neq y$，我们记为

$$\begin{aligned}\nabla(fg)(x) - \nabla(fg)(y) &= f(x)\nabla g(x) + g(x)\nabla f(x) - f(y)\nabla g(y) - g(y)\nabla f(y)\\ &= f(x)\big(\nabla g(x) - \nabla g(y)\big) + \big(f(x) - f(y)\big)\nabla g(y)\\ &\quad + g(x)\big(\nabla f(x) - \nabla f(y)\big) + \big(g(x) - g(y)\big)\nabla f(y).\end{aligned}$$

取欧几里得范数并应用三角不等式可得

$$\begin{aligned}\|\nabla(fg)(x) - \nabla(fg)(y)\|_2 \le{}& \|f\|_\infty \|\nabla g(x) - \nabla g(y)\|_2 + |f(x) - f(y)|\,\|\nabla g\|_\infty\\ &+ \|g\|_\infty \|\nabla f(x) - \nabla f(y)\|_2 + |g(x) - g(y)|\,\|\nabla f\|_\infty.\end{aligned}$$

除以 $\|x-y\|_2^\alpha$，并对 $x \neq y$ 取上确界:

$$[\nabla(fg)]_\alpha \le A\,C' + A'\,C + \Big(\sup_{x\neq y} \frac{|f(x)-f(y)|}{\|x-y\|_2^\alpha}\Big) B' + \Big(\sup_{x\neq y} \frac{|g(x)-g(y)|}{\|x-y\|_2^\alpha}\Big) B.$$

剩下需要通过 f 和 g 的梯度来控制它们的 H¨lder 半范数。根据中值定理，对于任意 $x,\ y \in [0,1]^d$,

$$|f(x)-f(y)| \le \|\nabla f\|_\infty \|x-y\|_2 = B\,\|x-y\|_2,$$

因此

$$\frac{|f(x)-f(y)|}{\|x-y\|_2^\alpha} \le B\,\|x-y\|_2^{1-\alpha} \le B\,\mathrm{diam}(\mathcal{O})^{1-\alpha}.$$

对 $x \neq y$ 取上确界可得

$$[f]_\alpha \le \mathrm{diam}(\mathcal{O})^{1-\alpha}\,B.$$

类似地，$[g]_\alpha \le \mathrm{diam}(\mathcal{O})^{1-\alpha}\,B'$。将这些结果代入前面的公式可得

$$[\nabla(fg)]_\alpha \le A\,C' + A'\,C + 2\,\mathrm{diam}(\mathcal{O})^{1-\alpha}\,B\,B'.$$

将上述界相加，我们得到

$$\|fg\|_{C^{1,\alpha}(\mathcal{O})} \le AA' + (AB'+A'B) + (AC'+A'C) + 2\,\mathrm{diam}(\mathcal{O})^{1-\alpha}\,BB'.$$

现在注意到，右侧的每一项都有界于

$$(1+2\,\mathrm{diam}(\mathcal{O})^{1-\alpha})(A+B+C)(A'+B'+C'),$$

由于 $AA',\ AB',\ A'B,\ AC',\ A'C \le (A+B+C)(A'+B'+C')$ 且 $2\mathrm{diam}(\mathcal{O})^{1-\alpha}BB' \le (1+2\mathrm{diam}(\mathcal{O})^{1-\alpha})(A+B+C)(A'+B'+C')$。因此，

$$\|fg\|_{C^{1,\alpha}(\mathcal{O})} \le \big(1+2\,\mathrm{diam}(\mathcal{O})^{1-\alpha}\big)\|f\|_{C^{1,\alpha}(\mathcal{O})}\,\|g\|_{C^{1,\alpha}(\mathcal{O})}.$$

这就以 $C_{\mathrm{alg}} = 1+2\,\mathrm{diam}(\mathcal{O})^{1-\alpha}$ 证明了该结论。最后，当 $\mathcal{O} = [0,1]^d$ 时，欧几里得直径为

$$\mathrm{diam}(\mathcal{O}) = \sup_{x,y\in[0,1]^d} \|x-y\|_2 = \|(1,\dots,1)-(0,\dots,0)\|_2 = \sqrt{d},$$

因此 $C_{\mathrm{alg}} = 1+2\,d^{(1-\alpha)/2}$。 □

13. Appendix D

回顾一下，观测空间为 $\mathcal{O} = [0,1]^d \times \{0,1\} \times \{0,1\}$，并记 $O_k = (X_k, A_k, Y_k)$。对于 $(a,y) \in \{0,1\}^2$，定义切片 $\mathcal{O}_{a,y} := [0,1]^d \times \{a\} \times \{y\}$。对于任意 $o = (x, A_k, Y_k) \in \mathcal{O}_{A_k,Y_k}$，沿连续坐标到 O_k 的欧几里得距离为 $\|o - O_k\|_2 := \|x - X_k\|_2$，该距离在切片 $\mathcal{O}_{A_k,Y_k}$ 上定义良好。固定 $\delta_n > 0$。令 C 和 $M_{1,\alpha}$ 为假设 3 中的常数，并定义半径 $w(\delta_n) := \frac{\delta_n}{2CM_{1,\alpha}}$。对于每个 $k = 1,\dots,n$，定义切片限制球

$$B(O_k, w(\delta_n)) := \big\{ o \in \mathcal{O}_{A_k,Y_k} :\ \|o - O_k\|_2 < w(\delta_n) = \frac{\delta_n}{2CM_{1,\alpha}} \big\}.$$

假设 p_0 为每个邻域 $B(O_k, w(\delta_n))$ 赋予正质量，并令

$$\mathfrak{J}(\delta_n) := \max_{1\le k\le n} \left(\int_{B(O_k, w(\delta_n))} p_0(o)\,d\mu(o) \right)^{-1}.$$

为避免反证论证中出现边界/等式情形，我们引入一个固定松弛因子（任意常数 > 1 均可）并定义 $T := \frac{2}{\delta_n}\log(2\,\mathfrak{J}(\delta_n))$。令 $M_0 := \|p_0\|_{C^{1,\alpha}(\mathcal{O})}$ 并定义

$$M := M_0 \exp\big(C_{\mathrm{alg}} C M_{1,\alpha} T\big) = M_0 \big(2\,\mathfrak{J}(\delta_n)\big)^{\frac{2C_{\mathrm{alg}}CM_{1,\alpha}}{\delta_n}},$$

其中 C_{alg} 是 $C^{1,\,\alpha}(\mathcal{O})$ 的巴拿赫代数性质中的一个常数，该性质定义于定理 12.2。接着我们引入（闭）约束集

$$\mathcal{B}_M := \Big\{ p \in C^{1,\alpha}(\mathcal{O}) :\ p \geq 0,\ \int_{\mathcal{O}} p \, d\mu = 1,\ \|p\|_{C^{1,\alpha}(\mathcal{O})} \leq M \Big\}.$$

所有存在性/唯一性论证都将在 C^1中的p_0 在$C^{1,\,\alpha}(\mathcal{O})$的一个开邻域内进行，随后我们将验证所得的解在 $[0,T]$上仍然属于 $\mathcal{B}_M$。

引理 13.1（ D_t的一致 L^∞ 界）。 回顾 $D_t = D(p_t) := \frac{1}{n}\sum_{j=1}^n \alpha_j^{(t)}\, k_{O_j}^{(t)}$ 其中

$$\alpha_j^{(t)} := P_n\Big[k_{O_j}^{(t)}\Big] = \frac{1}{n}\sum_{i=1}^n K^{(t)}(O_i, O_j), \qquad j = 1, \ldots, n.$$

假设基准正定核 K 满足$\sup_{o\in\mathcal{O}} K(o,o) \leq C$，其中 C为某个常数。那么对每个 $t \in I$，我们有

$$\|D_t\|_\infty \leq C^2.$$

特别地，该界关于 t一致成立，并且对任意固定样本逐路径成立。

证明。 由于 $\mathcal{H}_K^{(t)}$ 是 $\mathcal{H}_K$的闭子空间，正交投影存在。每个核截面都满足 $k_o^{(t)} = \Pi_t k_o$，其中 Π_t 表示到 $\mathcal{H}_K^{(t)}$ 上的正交投影。因此，根据假设 3，投影核的对角线值会减小，

$$K^{(t)}(o,o) = \|k_o^{(t)}\|^2_{\mathcal{H}_K^{(t)}} = \|\Pi_t k_o\|^2_{\mathcal{H}_K} \leq \|k_o\|^2_{\mathcal{H}_K} = K(o,o) \leq C.$$

根据 $\mathcal{H}_K^{(t)}$ 中的柯西–施瓦茨不等式，

$$|K^{(t)}(o,o')| = \big|\langle k_o^{(t)}, k_{o'}^{(t)}\rangle_{\mathcal{H}_K^{(t)}}\big| \leq \|k_o^{(t)}\|_{\mathcal{H}_K^{(t)}}\, \|k_{o'}^{(t)}\|_{\mathcal{H}_K^{(t)}} = \sqrt{K^{(t)}(o,o)\, K^{(t)}(o',o')} \leq C.$$

因此，对每个 j，我们都可以对系数作如下界定，

$$|\alpha_j^{(t)}| = \left|\frac{1}{n}\sum_{i=1}^n K^{(t)}(O_i, O_j)\right| \leq \frac{1}{n}\sum_{i=1}^n |K^{(t)}(O_i, O_j)| \leq C.$$

最后，利用 $k_{O_j}^{(t)}(o) = K^{(t)}(o,\ O_j)$，对于每个 $o \in \mathcal{O}$，我们都可以得到 D_t的逐点上界：

$$|D_t(o)| = \left|\frac{1}{n}\sum_{j=1}^n \alpha_j^{(t)}\, K^{(t)}(o, O_j)\right| \leq \frac{1}{n}\sum_{j=1}^n |\alpha_j^{(t)}|\, |K^{(t)}(o, O_j)| \leq \frac{1}{n}\sum_{j=1}^n C \cdot C = C^2.$$

对 $o \in \mathcal{O}$ 取上确界可得 $\|D_t\|_\infty \leq C^2$。 □

定理 13.1 给出了 D_t的一致 L^∞ 界；只要核本身（在对角线上）有界，得分型方向 D_t就不会在逐点意义上发散。与前面一样，对于高斯核，我们有 $C = 1$。接下来的定理 13.2 表明，如果核截面沿流一致 $C^{1,\,\alpha}$-有界，那么 D_t 也具有相同的正则性。

引理 13.2. 令 $\mathcal{O} = [0,1]^d \times \{0,1\}^2$，并固定 $\alpha \in (0,1]$。假设对于某个常数 $M_{1,\,\alpha} < \infty$，

$$\sup_{t\in I}\ \sup_{1\leq j\leq n} \big\|k_{O_j}^{(t)}\big\|_{C^{1,\alpha}} \leq M_{1,\alpha}. \tag{21}$$

那么，对于每个 $t \in I$、 $D_t \in C^{1,\,\alpha}(\mathcal{O})$ 以及 上确界$_{t\in I}$ $\|D_t\|_{C^{1,\,\alpha}} \leq C M_{1,\,\alpha}$。

证明。根据定义，我们有 $D_t = \frac{1}{n}\sum_{j=1}^{n} \alpha_{\mathcal{F}}^{(t)}\, k_{O_j}^{(t)}$。由于 $C^{1,\alpha}(\mathcal{O})$ 是一个赋范向量空间，由三角不等式可得

$$\|D_t\|_{C^{1,\alpha}(\mathcal{O})} \le \frac{1}{n}\sum_{j=1}^{n} |\alpha_j^{(t)}|\, \|k_{O_j}^{(t)}\|_{C^{1,\alpha}(\mathcal{O})}.$$

根据定理 13.1，我们有 $|\alpha_j^{(t)}| \le C$ 在 j 和 t 上一致成立。结合公式 (21)，可得

$$\|D_t\|_{C^{1,\alpha}(\mathcal{O})} \le \frac{1}{n}\sum_{j=1}^{n} C\, M_{1,\alpha} = C\, M_{1,\alpha}.$$

对 $t \in I$ 取上确界，即完成证明。 □

下一个引理证明了 $p \mapsto D(p)$ 在 $\mathcal{B}_M$ 上关于 $C^{1,\alpha}$ 范数是利普希茨连续的，这对于证明密度值常微分方程解的存在唯一性至关重要。

引理 13.3。存在一个常数 $K_D < \infty$（仅依赖于 C、M_K、$M_{1,\alpha}$），使得对所有 p、$q \in \mathcal{B}_M$，

$$\|D(p) - D(q)\|_{C^{1,\alpha}} \le K_D\, \|p - q\|_{C^{1,\alpha}}.$$

证明。固定 $p,\ q \in \mathcal{B}_M$ 并记 $\eta := p - q$。对于任意密度 $r \in \mathcal{B}_M$，按前文定义核均值嵌入

$$m_r(\cdot) := \int_{\mathcal{O}} K(\cdot, s) r(s)\, d\mu(s) \in C^{1,\alpha}(\mathcal{O}),$$
$$\kappa_r := \|m_r\|_{\mathcal{H}_K}^2 = \iint_{\mathcal{O}\times\mathcal{O}} K(s,t)\, r(s) r(t)\, d\mu(s) d\mu(t).$$

我们使用正交投影截面

$$k_y^{(r)} := \Pi_r k_y = k_y - \frac{P_r[k_y]}{\|m_r\|_{\mathcal{H}_K}^2}\, m_r = k_y - \frac{m_r(y)}{\kappa_r}\, m_r,$$

因此，对于每个 $x \in \mathcal{O}$，

$$k_y^{(r)}(x) = K(x,y) - \frac{m_r(y)}{\kappa_r}\, m_r(x).$$

假设 $\kappa_0 := \inf_{r\in\mathcal{B}_M} \kappa_r > 0$（例如，对于紧致 $\mathcal{O}$ 上的高斯核，$\kappa_0 \ge \inf_{s,\, t\in\mathcal{O}} K(s,\ t) > 0$）。

由于 $\mathcal{O}$ 具有有限测度且 $\|\cdot\|_{C^{1,\alpha}}$ 逐切片支配 $\|\cdot\|_\infty$，$\|\eta\|_{L^1} \le \|\eta\|_\infty \le \|\eta\|_{C^{1,\alpha}}$ 利用假设 3 中的 (19)，

$$\|m_p - m_q\|_{C^{1,\alpha}} = \left\| \int_{\mathcal{O}} K(\cdot, s)\eta(s)\, d\mu(s) \right\|_{C^{1,\alpha}} \tag{22}$$
$$\le \int_{\mathcal{O}} \|K(\cdot, s)\|_{C^{1,\alpha}} |\eta(s)|\, d\mu(s)$$
$$\le M_K \|\eta\|_{L^1} \le M_K \|\eta\|_{C^{1,\alpha}}. \tag{23}$$

此外，对于任意固定的 y，

$$|m_p(y) - m_q(y)| = \left| \int_{\mathcal{O}} K(y, s)\eta(s)\, d\mu(s) \right| \le C\|\eta\|_{L^1} \le C\|\eta\|_{C^{1,\alpha}}. \tag{24}$$

此外,

$$\kappa_p - \kappa_q = \iint_{\mathcal{O}\times\mathcal{O}} K(s,t)\big(p(s)p(t) - q(s)q(t)\big)\, d\mu(s)d\mu(t).$$

注意， $p(s)p(t) - q(s)q(t) = \eta(s)p(t) + q(s)\eta(t)$，因此

$$\big|p(s)p(t) - q(s)q(t)\big| \le |\eta(s)|\,|p(t)| + |q(s)|\,|\eta(t)|.$$

由于 p， q 是密度， $\int_{\mathcal{O}} p\, d\mu = \int_{\mathcal{O}} q\, d\mu = 1$ 和 p， $q \ge 0$；因此,

$$\begin{aligned}\left|\iint_{\mathcal{O}\times\mathcal{O}} K(s,t)\big(p(s)p(t) - q(s)q(t)\big)d\mu(s)d\mu(t)\right| &\le C\iint_{\mathcal{O}\times\mathcal{O}} \big(|\eta(s)|p(t) + q(s)|\eta(t)|\big)d\mu(s)d\mu(t)\\ &= C\left(\int_{\mathcal{O}} |\eta(s)|\, d\mu(s)\right)\left(\int_{\mathcal{O}} p(t)\, d\mu(t)\right)\\ &\quad + C\left(\int_{\mathcal{O}} q(s)d\mu(s)\right)\left(\int_{\mathcal{O}} |\eta(t)|d\mu(t)\right)\\ &= 2C\|\eta\|_{L^1} \le 2C\|\eta\|_{C^{1,\alpha}}. \qquad (25)\end{aligned}$$

因此,

$$|\kappa_p - \kappa_q| \le 2C\|\eta\|_{L^1} \le 2C\|\eta\|_{C^{1,\alpha}}. \qquad (26)$$

定义 $\beta_r(y) := m_r(y)/\kappa_r$。那么 $|\beta_q(y)| \le C/\kappa_0$ 并且

$$\begin{aligned}|\beta_p(y) - \beta_q(y)| &\le \frac{|m_p(y) - m_q(y)|}{\kappa_0} + |m_q(y)|\left|\frac{1}{\kappa_p} - \frac{1}{\kappa_q}\right|\\ &\le \frac{C}{\kappa_0}\|\eta\|_{C^{1,\alpha}} + \frac{C}{\kappa_0^2}|\kappa_p - \kappa_q| \le \left(\frac{C}{\kappa_0} + \frac{2C^2}{\kappa_0^2}\right)\|\eta\|_{C^{1,\alpha}}, \qquad (27)\end{aligned}$$

利用 (24) 和 (26)。

Now,

$$k_y^{(p)} - k_y^{(q)} = -\big(\beta_p(y)m_p - \beta_q(y)m_q\big) = -(\beta_p(y) - \beta_q(y))m_p - \beta_q(y)(m_p - m_q).$$

由于根据 (19)， $\|m_p\|_{C^{1,\alpha}} \le \int \|K(\cdot, s)\|_{C^{1,\alpha}} p(s)\, d\mu(s) \le M_K$，结合 (23) 和 (27) 可得

$$\begin{aligned}\|k_y^{(p)} - k_y^{(q)}\|_{C^{1,\alpha}} &\le |\beta_p(y) - \beta_q(y)|\,\|m_p\|_{C^{1,\alpha}} + |\beta_q(y)|\,\|m_p - m_q\|_{C^{1,\alpha}}\\ &\le M_K\left(\frac{C}{\kappa_0} + \frac{2C^2}{\kappa_0^2}\right)\|\eta\|_{C^{1,\alpha}} + \frac{C}{\kappa_0}M_K\|\eta\|_{C^{1,\alpha}}\\ &= M_K\left(\frac{2C}{\kappa_0} + \frac{2C^2}{\kappa_0^2}\right)\|p - q\|_{C^{1,\alpha}}. \qquad (28)\end{aligned}$$

对 $y \in \mathcal{O}$ 取上确界，可得

$$\sup_{y\in\mathcal{O}} \|k_y^{(p)} - k_y^{(q)}\|_{C^{1,\alpha}} \le L_K\|p - q\|_{C^{1,\alpha}}, \quad L_K := M_K\left(\frac{2C}{\kappa_0} + \frac{2C^2}{\kappa_0^2}\right).$$

特别是，对于每个 j,

$$\|k_{O_j}^{(p)} - k_{O_j}^{(q)}\|_{C^{1,\alpha}} \le L_K\|p - q\|_{C^{1,\alpha}}.$$

回顾一下 $\alpha_j(p) = \frac{1}{n}\sum_{i=1}^n K^{(p)}(O_i, O_j)$。因此，

$$\begin{aligned}|\alpha_j(p) - \alpha_j(q)| &\le \frac{1}{n}\sum_{i=1}^n |K^{(p)}(O_i, O_j) - K^{(q)}(O_i, O_j)| \\ &\le \sup_{x,y\in\mathcal{O}} |K^{(p)}(x,y) - K^{(q)}(x,y)|. \end{aligned} \tag{29}$$

但对于任意固定的 y，量 $|K^{(p)}(x, y) - K^{(q)}(x, y)| = |k_y^{(p)}(x) - k_y^{(q)}(x)|$ 由 $\|k_y^{(p)} - k_y^{(q)}\|_\infty \le \|k_y^{(p)} - k_y^{(q)}\|_{C^{1,\alpha}}$ 有界。因此，

$$\max_{1\le j\le n} |\alpha_j(p) - \alpha_j(q)| \le L_K \|p - q\|_{C^{1,\alpha}}. \tag{30}$$

记为

$$\begin{aligned}D(p) - D(q) &= \frac{1}{n}\sum_{j=1}^n \left(\alpha_j(p) k_{O_j}^{(p)} - \alpha_j(q) k_{O_j}^{(q)}\right) \\ &= \frac{1}{n}\sum_{j=1}^n \left((\alpha_j(p) - \alpha_j(q)) k_{O_j}^{(p)} + \alpha_j(q)(k_{O_j}^{(p)} - k_{O_j}^{(q)})\right).\end{aligned}$$

取 $C^{1,\alpha}$ 范数并使用三角不等式，

$$\|D(p) - D(q)\|_{C^{1,\alpha}} \le \frac{1}{n}\sum_{j=1}^n \left(|\alpha_j(p) - \alpha_j(q)|\, \|k_{O_j}^{(p)}\|_{C^{1,\alpha}} + |\alpha_j(q)|\, \|k_{O_j}^{(p)} - k_{O_j}^{(q)}\|_{C^{1,\alpha}}\right).$$

由于 $|\alpha_j(q)| \le C$ 一致成立，且根据假设 3 中的式 (20)，$\|k_{O_j}^{(p)}\|_{C^{1,\alpha}} \le M_{1,\alpha}$ 在 $p \in \mathcal{B}_M$ 上一致成立。利用式 (28) 和 (30)，得到

$$\|D(p) - D(q)\|_{C^{1,\alpha}} \le (M_{1,\alpha} + C) L_K \|p - q\|_{C^{1,\alpha}}.$$

因此，该结论在 $K_D := (M_{1,\alpha} + C)L_K$ 下成立。 □

13.1. **常微分方程解的存在唯一性的证明。**

定理13.4 （常微分方程解的存在唯一性）。 设 $p_0 \in C^{1,\alpha}(\mathcal{O})$为初始密度。假定假设 *3* 成立，并设 $p_0 \in \mathcal{B}_M \subset C^{1,\alpha}(\mathcal{O})$。令$F\colon \mathcal{B}_M \to C^{1,\alpha}(\mathcal{O})$ 由 $F(p) := p\,D(p)$ 定义。则(16)中的密度值常微分方程，连同 $p_{t=0} = p_0$，

$$\frac{d}{dt} p_t = F(p_t),$$

存在唯一解 $t \mapsto p_t \in C^1([0,T]; C^{1,\alpha}(\mathcal{O}))$。对所有 $t \in [0,T]$, $p_t \in \mathcal{B}_M$。

证明。 首先，我们需要证明 $\frac{d}{dt}p_t = p_t D(p_t)$ 的解在某个区间 $I = (-\tau, \tau)$ 上的 $C^1(I, C^{1,\alpha}(\mathcal{O}))$ 中存在且唯一，其中 τ为一个常数。将算子 F定义为 $F(p) = pD(p)$。方程变为： $\frac{d}{dt}p_t = F(p_t)$， 初始条件为 p_0。

由于 $p \in C^{1,\alpha}(\mathcal{O})$与 $D(p) \in C^{1,\alpha}(\mathcal{O})$对 $p \in \mathcal{B}_M$ 成立（由 $D(\cdot)$的定义，并结合假设 3与定理 13.2），我们验证 $F(p) = pD(p) \in C^{1,\alpha}(\mathcal{O})$对 $p \in \mathcal{B}_M$成立。事实上，由定理12.2中 $C^{1,\alpha}(\mathcal{O})$的巴拿赫代数性质，

$$\|F(p)\|_{C^{1,\alpha}} = \|pD(p)\|_{C^{1,\alpha}} \le C_{\text{alg}}\, \|p\|_{C^{1,\alpha}}\, \|D(p)\|_{C^{1,\alpha}}, \tag{31}$$

因此 $F(p) \in C^{1,\alpha}(\mathcal{O})$。

接下来，我们证明 F 在 $C^{1,\alpha}$ 范数下的局部利普希茨连续性。令 $\mathcal{B}_R(p_0) = \{q \in C^{1,\alpha}(\mathcal{O}) : \|q - p_0\|_{C^{1,\alpha}} \le R\}$，其中任意固定的 $R > 0$。设

$$U := \{p \in C^{1,\alpha}(\mathcal{O}) : \ \|p\|_{C^{1,\alpha}} < 2M\},$$

它是 $C^{1,\alpha}(\mathcal{O})$ 的一个开子集，包含 $\mathcal{B}_M$ 和 p_0。根据定理 13.3，映射 $D(\cdot)$ 在 $\mathcal{B}_M$ 上关于 $C^{1,\alpha}$ 范数满足利普希茨条件，其常数为 K_D；根据假设 3 以及（20），我们有一致界

$$\sup_{p \in \mathcal{B}_M} \|D(p)\|_{C^{1,\alpha}} \le C\, M_{1,\alpha}. \tag{32}$$

对于 $p,\ q \in \mathcal{B}_M \cap \mathcal{B}_R(p_0)$，我们估计

$$\|F(p) - F(q)\|_{C^{1,\alpha}} = \|pD(p) - qD(q)\|_{C^{1,\alpha}}.$$

记

$$F(p) - F(q) = p(D(p) - D(q)) + (p - q)D(q).$$

根据定理 12.2,

$$\|p(D(p) - D(q))\|_{C^{1,\alpha}} \le C_{\mathrm{alg}}\, \|p\|_{C^{1,\alpha}}\, \|D(p) - D(q)\|_{C^{1,\alpha}},$$

and

$$\|(p - q)D(q)\|_{C^{1,\alpha}} \le C_{\mathrm{alg}}\, \|p - q\|_{C^{1,\alpha}}\, \|D(q)\|_{C^{1,\alpha}}.$$

利用 $\|p\|_{C^{1,\alpha}} \le M$ 对于 $p \in \mathcal{B}_M$，定理 13.3 和式 (32)，我们得到

$$\begin{aligned} \|F(p) - F(q)\|_{C^{1,\alpha}} &\le C_{\mathrm{alg}}\, M\, K_D\, \|p - q\|_{C^{1,\alpha}} + C_{\mathrm{alg}}\, \|p - q\|_{C^{1,\alpha}}\, (C\, M_{1,\alpha}) \\ &= C_{\mathrm{alg}}\big(MK_D + CM_{1,\alpha}\big)\, \|p - q\|_{C^{1,\alpha}}. \end{aligned} \tag{33}$$

因此，F 在 $\mathcal{B}_M$ 上是利普希茨的（因而在开集 U 上局部利普希茨），其利普希茨常数为 $L_F : = C_{\mathrm{alg}}\big(MK_D + CM_{1,\alpha}\big)$。

根据巴拿赫空间中的皮卡–林德洛夫定理（定理 3.6），在开集 U 上应用该定理，可知存在 $\tau > 0$，且存在唯一的解 $p \in C^1((-\tau,\ \tau);\, C^{1,\alpha}(\mathcal{O}))$，满足 $p_{t=0} = p_0$ 的 $\frac{d}{dt}p_t = F(p_t)$。该解满足积分方程

$$p_t = p_0 + \int_0^t p_s D(p_s)\, ds, \qquad t \in (-\tau, \tau). \tag{34}$$

由于对每个 t 都有 $p_t \in C^{1,\alpha}(\mathcal{O})$，并且 $C^{1,\alpha}(\mathcal{O})$ 连续嵌入到 $C(\mathcal{O})$ 中，因此该常微分方程在 $\mathcal{O}$ 上逐点成立。

接下来，我们验证 $p_t \in \mathcal{B}_M$ 在 $[0, T]$ 上成立。首先，考虑连续线性泛函 $\Phi :\ C^{1,\alpha}(\mathcal{O}) \to \mathbb{R}$，其定义为 $\Phi(q) = \int_{\mathcal{O}} q(o)\, d\mu(o)$。对于任意足够小的 $\delta \ne 0$，由 Φ 的线性和连续性可得，

$$\frac{\Phi(p(t + \delta)) - \Phi(p(t))}{\delta} = \Phi\left(\frac{p(t + \delta) - p(t)}{\delta}\right) \xrightarrow[\delta \to 0]{\|\cdot\|_{C^{1,\alpha}}} \Phi(p^{'}(t))$$

因此,

$$\frac{d}{dt}\int_{\mathcal{O}} p_t(o)\, d\mu(o) = \int_{\mathcal{O}} \frac{d}{dt} p_t(o)\, d\mu(o) = \int_{\mathcal{O}} p_t(o) D(p_t)(o)\, d\mu(o) = P_t[D(p_t)].$$

根据正交投影的定义，$D(p_t) \in \mathcal{H}_{K,p_t} := \{f \in \mathcal{H}_K :\ P_t[f] = 0\}$对每个 t，按构造均成立，因此 $P_t[D(p_t)] = 0$。由于 $\int p_0(o)\, d\mu(o) = 1$，我们得出，对于所有 $t \in (-\tau, \tau)$，$\int p_t(o)\, d\mu(o) = 1$。正性成立，因为对每个固定的 $o \in \mathcal{O}$，逐点常微分方程 $\frac{d}{dt} p_t(o) = p_t(o) D(p_t)(o)$ 蕴含

$$p_t(o) = p_0(o) \exp\left(\int_0^t D(p_s)(o)\, ds\right), \qquad t \in (-\tau, \tau), \tag{35}$$

因此，只要 $p_0(o) \ge 0$，就有 $p_t(o) \ge 0$。

接下来，我们验证$\|p_t\|_{C^{1,\alpha}} \le M$对 $t \in [0,T]$成立。由 (34) 和定理 12.2,

$$\begin{aligned}\|p_t\|_{C^{1,\alpha}} &\le \|p_0\|_{C^{1,\alpha}} + \int_0^t \|p_s D(p_s)\|_{C^{1,\alpha}}\, ds \\ &\le \|p_0\|_{C^{1,\alpha}} + C_{\mathrm{alg}} \int_0^t \|p_s\|_{C^{1,\alpha}} \|D(p_s)\|_{C^{1,\alpha}}\, ds \\ &\le \|p_0\|_{C^{1,\alpha}} + C_{\mathrm{alg}}\, C\, M_{1,\alpha} \int_0^t \|p_s\|_{C^{1,\alpha}}\, ds,\end{aligned}$$

其中最后一行使用了 (32)（只要 $p_s \in \mathcal{B}_M$即成立）。Gr¨onwall 不等式给出

$$\|p_t\|_{C^{1,\alpha}} \le \|p_0\|_{C^{1,\alpha}} \exp\bigl(C_{\mathrm{alg}}\, C\, M_{1,\alpha}\, t\bigr), \qquad t \in [0,\tau). \tag{36}$$

根据上文给出的 M和 T的定义（即 $M := M_0 \exp(C_{\mathrm{alg}} C M_{1,\alpha} T)$，其中 $M_0 := \|p_0\|_{C^{1,\alpha}}$，以及$T = \frac{2}{\delta_n} \log(2\mathfrak{J}(\delta_n))$），对所有 $t \in [0,T]$，我们有

$$\|p_t\|_{C^{1,\alpha}} \le M_0 \exp(C_{\mathrm{alg}} C M_{1,\alpha} t) \le M_0 \exp(C_{\mathrm{alg}} C M_{1,\alpha} T) = M,$$

因此$\|p_t\|_{C^{1,\alpha}} \le M$在 $[0,T]$上成立。

最后，由于 F 在开集 U 上局部利普希茨，且界（36）表明解在时间 T 之前不会离开 U （事实上，$\|p_t\|_{C^{1,\alpha}} \le M < 2M$ 在$[0,T]$上），因此根据巴拿赫空间中常微分方程的标准延拓论证（在中间时刻重新应用 Picard–Lindel¨of 定理），我们可以将局部解唯一地延拓至 $[0,T]$。因此，存在唯一的解

$$p \in C^1([0,T]; C^{1,\alpha}(\mathcal{O}))$$

并且对所有 $t \in [0,T]$，都有 $p_t \ge 0$、$\int_{\mathcal{O}} p_t\, d\mu = 1$ 和 $\|p_t\|_{C^{1,\alpha}} \le M$，即 $p_t \in \mathcal{B}_M$。

到目前为止，我们已经证明，对所有 $t \in [0,T]$，$\frac{d}{dt} p_t(o) = p_t(o) D(p_t)(o)$ 对所有 $o \in \mathcal{O}$ 均逐点成立，且 $p_t \in \mathcal{B}_M$。注意，这就解决了（16）。此外，在集合 $\{o \in \mathcal{O} : p_t(o) > 0\}$ 上，它还意味着 $\frac{d}{dt} \log p_t(o) = \frac{d}{dt} \overline{p_t(o)}\ p_t(o)$。 □

13.2. **经验得分函数收敛的证明。**

引理 13.5. 我们有

$$\sup_{p \in \mathcal{B}_M} \sup_{o \in \mathcal{O}} \|\nabla D(p)(o)\|_2 \le C M_{1,\alpha} < \infty,$$

其中 $C, M_{1,\alpha}$ 在假设 *3* 中定义。

证明。任取一个切片 $b \in \{0,1\}^2$，并在 $[0,1]^d$ 上写作 $v(x) := D(p)(x,b)$。根据切片上 $C^{1,\alpha}$ 范数的定义，

$$\|v\|_{C^{1,\alpha}} = \|v\|_\infty + \sum_{i=1}^d \|\nabla_i v\|_\infty + \sum_{i=1}^d [\nabla_i v]_\alpha.$$

因此，由于所有项均非负，我们得到

$$\sum_{i=1}^d \|\nabla_i v\|_\infty \ \le\ \|v\|_{C^{1,\alpha}}.$$

对于任意 $x \in [0,1]^d$,

$$\|\nabla_x v(x)\|_2 = \Big(\sum_{i=1}^{d} |\nabla_i v(x)|^2\Big)^{1/2} \le \Big(\sum_{i=1}^{d} \|\nabla_i v\|_\infty^2\Big)^{1/2} \le \sum_{i=1}^{d} \|\nabla_i v\|_\infty \le \|v\|_{C^{1,\alpha}}.$$

对各切片取最大值，并使用

$$\|D(p)\|_{C^{1,\alpha}(\mathcal{O})} = \max_b \ \|D(p)(\cdot, b)\|_{C^{1,\alpha}([0,1]^d)},$$

我们得到，对任意 $o = (x,\ b) \in \mathcal{O}$ 和 $p \in \mathcal{B}_M$,

$$\|\nabla_x D(p)(o)\|_2 \le \|D(p)(\cdot, b)\|_{C^{1,\alpha}([0,1]^d)} \le \|D(p)\|_{C^{1,\alpha}(\mathcal{O})}.$$

最后，对 $o \in \mathcal{O}$ 和 $p \in \mathcal{B}_M$ 取上确界，得到

$$\sup_{p\in\mathcal{B}_M} \sup_{o\in\mathcal{O}} \|\nabla_x D(p)(o)\|_2 \le \sup_{p\in\mathcal{B}_M} \|D(p)\|_{C^{1,\alpha}(\mathcal{O})} \le C M_{1,\alpha} < \infty,$$

由定理 13.2。 □

定理 13.5 证明了一致范数 $C^{1,\alpha}$ 对$D(p)$ 在 $\mathcal{B}_M$ 上的控制所导出的一个有用推论。它给出了 $D(p)$ 关于连续协变量坐标的梯度欧几里得范数的一致界，该界将用于控制 $D(p)$ 在观测样本点周围小邻域内的变化。

引理 13.6 (经验对数似然的单调性和平稳性)。 设$t \mapsto p_t$ 是微分方程的一个足够正则的解

$$\frac{d}{dt} \log p_t(o) = D(p_t)(o).$$

其中 $D(p_t) := \frac{1}{n} \sum_{j=1}^{n} \alpha_j^{(t)} k_{O_j}^{(t)}$。则有以下结论:

(1) 映射 $t \mapsto \mathbb{P}_n[\log p_t]$ 在 I 上单调不减。

(2) 对任意 $t_1 \in I$,

$$\frac{d}{dt}\mathbb{P}_n[\log p_t]\Big|_{t=t_1} = 0 \quad \Longleftrightarrow \quad m_n^{(t_1)} = 0 \ \text{in}\ \mathcal{H}_K^{(t_1)}.$$

证明。 沿着流对 $\mathbb{P}_n[\log p_t]$ 求导，并结合式 (9) 使用 $D(p_t)$ 的定义，可得

$$\begin{aligned}
\frac{d}{dt}\mathbb{P}_n[\log p_t] = \mathbb{P}_n[D(p_t)] &= \frac{1}{n}\sum_{i=1}^{n} D(p_t)(O_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n} \alpha_j^{(t)}\, K^{(t)}(O_j, O_i) \\
&= \frac{1}{n}\sum_{j=1}^{n} \alpha_j^{(t)}\, \mathbb{P}_n[k_{O_j}^{(t)}] \\
&= \frac{1}{n}\sum_{j=1}^{n} \big(\alpha_j^{(t)}\big)^2 \ge 0.
\end{aligned}$$

这就证明了单调性。如果导数在某个 $t_1 \in I$处为零，那么$\alpha^{(t_1)}{}_j = 0$对所有 $j = 1, \ldots, n$均成立。由于 $\alpha^{(t_1)}{}_j = \big\langle k_{O_j}^{(t)}, m^{(t_1)}{}_n \big\rangle_{\mathcal{H}_K^{(t_1)}}$，这意味着 $m^{(t_1)}{}_n$与张成空间$\{k_{O_1}^{(t)}, \ldots, k_{O_n}^{(t)}\}$正交。然而，根据定义， $m^{(t_1)}{}_n = \frac{1}{n}\sum_{i=1}^{n} k_{O_i}^{(t)} \in$

张成空间$\{k_{O_1}^{(t)},\ldots,k_{O_n}^{(t)}\}$，因此 $m^{(t_1)}{}_n = 0$ 属于 $\mathcal{H}^{(t_1)}{}_K$。反过来，如果 $m^{(t_1)}{}_n = 0$，那么所有 $\alpha^{(t_1)}{}_j = 0$，且导数 $\frac{d}{dt}\mathbb{P}_n[\log p_t]$ 在 t_1处为零。□

定理 13.7。 存在 $t \in [0,T]$，使得 $\mathbb{P}_n D(p_t) \leq \delta_n$。

证明。 根据定理 6.1，解 p_t 在 $[0,T]$上存在且唯一。为导出矛盾，假设对于所有 $t \in [0,T]$，都有 $\mathbb{P}_n D(p_t) > \delta_n$。然后，将（16）从 0 积分到 T，由于 $P_n \frac{d}{dt}\log p_t = \mathbb{P}_n D(p_t)$，我们得到

$$P_n[\log p_T] - P_n[\log p_0] > \delta_n T.$$

利用表示式

$$p_T = p_0 \exp\left(\int_0^T D(p_s)\,ds\right), \tag{37}$$

这给出

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^T D(p_s)(O_i)\,ds > \delta_n T.$$

因此，必然存在一个指标 k，使得 $\int_0^T D(p_s)(O_k)\,ds > \delta_n T$。根据定理 13.5，$D$ 在欧几里得范数下是一致有界的，并且对所有 $o \in \mathcal{O}_k$，我们有

$$D(p_s)(o) \geq D^*(p_s)(O_k) - CM_{1,\alpha}\|o - O_k\|_2.$$

因此，

$$\int_0^T D(p_s)(o)\,ds \geq \int_0^T D(p_s)(O_k)\,ds - CM_{1,\alpha}T\|o - O_k\|_2.$$

由（37）可知，这意味着

$$\begin{aligned} p_T(o) &\geq p_0(o)\exp\left(\int_0^T D(p_s)(O_k)\,ds - CM_{1,\alpha}T\|o - O_k\|_2\right) \\ &> p_0(o)\exp\left(\delta_n T - CM_{1,\alpha}T\|o - O_k\|_2\right). \end{aligned}$$

最后，我们有

$$\begin{aligned} \int_{\mathcal{O}_{A_k,Y_k}} p_T(o)\,d\mu(o) &\geq \int_{B(O_k,w)} p_T(o)\,d\mu(o) \\ &> \int_{B(O_k,w)} p_0(o)\exp\left(\frac{\delta_n T}{2}\right)\,d\mu(o) \\ &= \exp\left(\frac{\delta_n T}{2}\right)\int_{B(O_k,w)} p_0(o)\,d\mu(o). \end{aligned}$$

由于 p_T 是概率密度，当

$$\exp\left(\frac{\delta_n T}{2}\right) > \left(\int_{B(O_k,w(\delta_n))} p_0(o)\,d\mu(o)\right)^{-1}.$$

证明至此完成。 □

13.3. **渐近线性与效率**。

定理 13.8。设 $\Psi:\ \mathcal{M}\to\mathbb{R}$ 是分布 P的一个路径可微泛函，其典范梯度为 $\phi_P^*\in L_0^2(P)$。设 $t_n\in[0,T]$满足 $\mathbb{P}_nD(p_{t_n})\ \le\delta_n$，且 $P_n:=P_{t_n}$。在假设 *1*、*2*、*3*、*4*、 $\mathbb{P}_n\phi^*_{P_{t_n}}=o_{P^*}(n^{-1/2})$ 下，*ULFS–KDPE* 估计量满足

$$\Psi(\widehat{P}_n)-\Psi(P^*)=\mathbb{P}_n\phi^*_{P^*}+o_{P^*}(n^{-1/2}).$$

证明。根据冯·米塞斯展开（定理 3.1），我们有

$$\Psi(\widehat{P}_n)-\Psi(P^*)=\mathbb{P}_n\phi^*_{P^*}-\mathbb{P}_n\phi^*_{\widehat{P}_n}+(\mathbb{P}_n-P^*)(\phi^*_{\widehat{P}_n}-\phi^*_{P^*})+R_2(\widehat{P}_n,P^*).\quad(38)$$

在假设1和假设2下，只需证明 $\mathbb{P}_n\phi^*_{\widehat{P}_n}=o_{P^*}(n^{-1/2})$。

固定 $t\in[0,T]$，并令 $m_n^{(t)}\in\mathcal{H}_{K,\ p_t}$为经验均值嵌入，即$P_n[f]=\langle f,\ m_n^{(t)}\rangle_{\mathcal{H}_K^{(t)}}$对所有 $f\in\mathcal{H}_K^{(t)}$均成立。回顾 $\alpha_j^{(t)}:=\mathbb{P}_n[k_{O_j}^{(t)}]$以及 $\mathbb{P}_nD(p_t)=\frac{1}{n}\sum_{j=1}^n(\alpha_j^{(t)})^2\ge0$。取 $f=m_n^{(t)}$可得

$$\mathbb{P}_n[m_n^{(t)}]=\langle m_n^{(t)},m_n^{(t)}\rangle_{\mathcal{H}_K^{(t)}}=\|m_n^{(t)}\|^2_{\mathcal{H}_K^{(t)}}.$$

由于 $\alpha_j^{(t)}=m_n^{(t)}(O_j)$，我们有

$$\mathbb{P}_n[m_n^{(t)}]=\frac{1}{n}\sum_{j=1}^n m_n^{(t)}(O_j)=\frac{1}{n}\sum_{j=1}^n\alpha_j^{(t)}.$$

由柯西–施瓦茨不等式,

$$\Big(\frac{1}{n}\sum_{j=1}^n\alpha_j^{(t)}\Big)^2\le\frac{1}{n}\sum_{j=1}^n(\alpha_j^{(t)})^2.$$

最后，由 ULFP 恒等式 $\mathbb{P}_nD(p_t)=\frac{1}{n}\sum_{j=1}^n(\alpha_j^{(t)})^2$, 我们得到

$$\|m_n^{(t)}\|^4_{\mathcal{H}_K^{(t)}}=\Big(\mathbb{P}_n[m_n^{(t)}]\Big)^2=\Big(\frac{1}{n}\sum_{j=1}^n\alpha_j^{(t)}\Big)^2\le\mathbb{P}_nD(p_t),$$

这意味着 $\|m_n^{(t)}\|_{\mathcal{H}_K^{(t)}}\le\{\mathbb{P}_nD(p_t)\}^{1/4}$。我们知道 $\mathbb{P}_nD(p_{t_n})\ \le\delta_n$ 且 $\delta_n=o(n^{-2})$, 因此我们得到

$$\|m_n^{(t_n)}\|_{\mathcal{H}_K^{(t_n)}}\le\delta_n^{1/4}=o(n^{-1/2}).\quad(39)$$

由假设 4(i)，对每条样本路径 ω 我们可以选取一个下标 $j_n(\omega)\uparrow\infty$使得

$$\|h_{j_n}-\phi^*_{\widehat{P}_n}\|_{L^2(P^*)}=o_{P^*}(n^{-1/2}).\quad(40)$$

分解

$$\mathbb{P}_n\phi^*_{\widehat{P}_n}=\mathbb{P}_nh_{j_n}+\mathbb{P}_n(\phi^*_{\widehat{P}_n}-h_{j_n}).\quad(41)$$

由于 $h_{j_n}\in\mathcal{H}_K^{(t_n)}$，里斯恒等式给出 $\mathbb{P}_nh_{j_n}=\langle h_{j_n},\ m^{(t_n)}{}_n\ \rangle_{\mathcal{H}_K^{(t_n)}}$，因此

$$|\mathbb{P}_nh_{j_n}|\le\|h_{j_n}\|_{\mathcal{H}_K^{(t_n)}}\|m_n^{(t_n)}\|_{\mathcal{H}_K^{(t_n)}}.$$

根据假设4(ii)和（39），可得

$$\mathbb{P}_nh_{j_n}=o_{P^*}(n^{-1/2}).\quad(42)$$

接下来，我们写出

$$\mathbb{P}_n(\phi^*_{\widehat{P}_n} - h_{j_n}) = (\mathbb{P}_n - P^*)(\phi^*_{\widehat{P}_n} - h_{j_n}) + P^*(\phi^*_{\widehat{P}_n} - h_{j_n}).$$

根据柯西–施瓦茨不等式和 (40),

$$|P^*(\phi^*_{\widehat{P}_n} - h_{j_n})| \le \|\phi^*_{\widehat{P}_n} - h_{j_n}\|_{L^2(P^*)} = o_{P^*}(n^{-1/2}). \tag{43}$$

此外，根据假设4(iii)，尾类是 P^*-Donsker。由于$\|\phi^*_{\widehat{P}_n} - h_{j_n}\|_{L^2(P^*)} \to 0$，经验过程的渐近等度连续性意味着

$$(\mathbb{P}_n - P^*)(\phi^*_{\widehat{P}_n} - h_{j_n}) = o_{P^*}(n^{-1/2}). \tag{44}$$

结合 (43) 和 (44) 得到

$$\mathbb{P}_n(\phi^*_{\widehat{P}_n} - h_{j_n}) = o_{P^*}(n^{-1/2}). \tag{45}$$

最后，由 (41) 以及 (42) 和 (45) 可得

$$\mathbb{P}_n \phi^*_{\widehat{P}_n} = o_{P^*}(n^{-1/2}). \tag{46}$$

将（46）代入（38），并利用假设1和假设2，我们得到

$$\Psi(\widehat{P}_n) - \Psi(P^*) = \mathbb{P}_n \phi^*_{P^*} + o(n^{-1/2}),$$

至此，证明完成。 □

References

[1] Raj Agrawal、Sam Witty、Andy Zane和Eli Bingham。使用蒙特卡洛有效影响函数的自动有效估计。*arXiv*预印本 *arXiv:2403.00158*，2024年。[2] N. Aronszajn。再生核理论。《美国数学学会汇刊》，68(3):337–404，1950年。[3] Peter J. Bickel、Chris Klaassen、Yaacov Ritov和Jon Wellner。半参数模型的有效与自适应估计。Springer，1998年。[4] Leo Breiman。随机森林。《机器学习》，45(1):5–32，2001年。[5] Marco Carone、Alexander R Luedtke和Mark J van der Laan。迈向无限维模型中的计算机化有效估计。《美国统计学会会刊》，2018年。[6] Tianqi Chen。Xgboost：可扩展的树提升系统。康奈尔大学，2016年。[7] Victor Chernozhukov、Denis Chetverikov、Mert Demirer、Esther Duflo、Christian Hansen和Whitney Newey。处理效应的双重/去偏/Neyman机器学习。《美国经济评论》，107(5):261–265，2017年。[8]Victor Chernozhukov、Whitney K Newey和Rahul Singh。因果效应和结构效应的自动去偏机器学习。《计量经济学》，90(3):967–1027，2022年。[9] Brian Cho、Yaroslav Mukhin、Kyra Gan和Ivana Malenica。核去偏插件估计：面向多个目标参数、无需影响函数的同时自动去偏。《机器学习研究论文集》，235:8534，2024年。[10] Jeremy Coyle。tmle3：可扩展的TMLE框架。*Zenodo*，2021年。[11] Jeremy R Coyle、Nima S Hejazi、Ivana Malenica、Rachael V Phillips和Oleg Sofrygin。sl3：机器学习与超级学习的现代流程。*R*软件包版本，1(2)，2021年。

[12] Michael Jordan、Yixin Wang和Angela Zhou。用于因果推断的经验加托导数。神经信息处理系统进展，35:8512–8525，2022年。[13] Charles A Micchelli、Yuesheng Xu和Haizhang Zhang。通用核。机器学习研究期刊，7(12)，2006年。[14] Whitney K. Newey。半参数效率界。应用计量经济学期刊，5(2):99–135，1990年。[15] Vittorino Pata。不动点定理及其应用，第116卷。Springer，2019年。[16] Noel Pimentel、Alejandro Schuler和Mark van der Laan。得分保持型定向最大似然估计，2025年。[17] J. Robins。持续暴露期死亡率研究中因果推断的新方法——应用于控制健康工人幸存者效应。数学建模，7(9):1393–1512，1986年。[18] Alejandro Schuler、Alexander Hagemeister和Mark van der Laan。高度自适应岭，2024年。[19] Gerald Teschl。常微分方程与动力系统，第140卷。美国数学学会，2012年。[20] A.A. Tsiatis。半参数理论与缺失数据。施普林格纽约，2006年。[21] Lars van der Laan、Aurelien Bibaut、Nathan Kallus和Alex Luedtke。非参数M估计目标之光滑泛函的自动去偏机器学习，2025年。[22] M. J. van der Laan、D. Benkeser和W. Cai。使用欠平滑高度自适应套索对路径可微目标参数进行有效估计。国际生物统计学期刊，2022年。[23] Mark van der Laan和Susan Gruber。基于通用最不利一维子模型的一步定向最小损失估计。国际生物统计学期刊，12:351–378，2016年5月。[24] Mark J. van der Laan和James M. Robins。因果推断与删失数据的统一方法。载于删失纵向数据与因果关系的统一方法，施普林格统计学丛书，第311–370页。Springer，纽约，2003年。[25] Mark J. van der Laan和Daniel Rubin。目标最大似然学习。工作论文213，加州大学伯克利分校生物统计学部，2006年10月。加州大学伯克利分校生物统计学部工作论文系列。[26] M.J. van der Laan、E.C. Polley和A.E. Hubbard。超级学习器。技术报告，工作论文222，加州大学伯克利分校生物统计学部工作论文系列，2007年7月。[27] M.J. van der Laan和S. Rose。定向学习：观察性与实验性数据的因果推断（施普林格统计学丛书）。Springer，2011年。[28] M.J. van der Laan和S. Rose。数据科学中的定向学习：复杂纵向研究的因果推断。施普林格科学与商业媒体，2018年。[29] A.W van der Vaart。渐近统计学，第3卷。剑桥大学出版社，2000年。

[30] Mingxun Wang、Alejandro Schuler、Mark van der Laan 和 Carlos Garc´ıa Meixide。高度自适应主成分回归，2026 年。