圆形测度的Wasserstein主成分分析

Mario Beraha[1] 和 Matteo Pegoraro[2]

[1]都灵大学经济与统计系[2]奥尔堡大学数学科学系

arXiv:2304.02402v1 [stat.ME] 5 Apr 2023

2023年4月6日

摘要

我们考虑支撑在单位圆上的概率测度的2-Wasserstein空间，并提出了针对生活在该空间中的数据进行主成分分析（PCA）的框架。我们基于对单位圆上测度的最优传输问题的详细研究，这可能具有独立的兴趣。特别地，我们推导出（几乎）闭形式的最优传输映射表达式，并提出了绝对连续概率测度处切空间的替代定义，以及相关的指数映射和对数映射。PCA通过将数据映射到Wasserstein重心处的切空间来执行，我们通过迭代方案对其进行近似，并为其建立了评估其收敛性的充分后验条件。我们的方法在若干模拟场景以及视神经厚度测量的真实数据分析中得到了说明。

关键词：最优传输、圆形测度、主成分分析、弱黎曼结构、分布数据分析

1 引言

对复杂数据（例如高维数据、函数数据、成分数据或流形值数据）的分析，正成为统计学文献中的一种新兴趋势。事实上，医学成像、基因组分析、地球科学等领域都会常规收集此类复杂数据。要获得有意义的分析结果，关键在于为数据取值所在的空间赋予恰当的数学结构，以捕捉所研究现象的变异性。例如，分析函数数据时，人们往往希望将定义域的重新参数化下等价的函数视为相同。这便引出了“对齐”问题 [参见 Sangalli 等人， 2009]。同样，分析成分数据时，将分析嵌入通常的欧几里得空间并不合适，因为数据受限于

单位维单纯形，因此若不谨慎处理，加法和数乘等运算便毫无意义 [Hron 等人，2014]。与此同时，某些度量可能比其他度量更适合比较两个数据点，例如，对于函数，可以考虑上确界范数或 L_2 范数。

本文聚焦于分布数据分析，即复杂数据分析中的一种特定情形，其中数据点是概率测度。具体而言，我们考虑支撑于单位圆 $\mathbb{S}_1 := \{(x, y) \in \text{实数集}^2 : x^2 + y^2 = 1\}$ 上的概率测度所构成的 2-Wasserstein空间，并提出一个框架，用于对这类空间中的数据进行主成分分析（PCA）。主成分分析在实践中广受欢迎，因为它既能生成一组正交方向（通常解释为数据集中变异性的主要方向），也能将数据所在的空间映射到这些方向所张成的空间。因此，可以通过在得分上投影数据，利用主成分分析直观地解读数据集中的变异性，并降低数据的维度。特别是，线性回归或聚类等经典多元统计工具可以基于得分进行分析。在数据位于非线性空间的情境下，后一特点尤其具有吸引力，因为它使我们能够直接将现成工具用于主成分分析得分。

我们的研究源于 Ali 等人 [2021], 的分析，其中通过光学相干断层扫描（OCT）获得的视神经头测量值被研究，以探讨其与青光眼等视神经病变的发展和进展的关系。OCT 对眼睛进行圆形扫描，测量神经视网膜缘（NRR）厚度，因此每个数据点可被视为支撑在 $\mathbb{S}_1$ 上的函数。然后对这些数据进行归一化，以消除由不同量级引入的不期望的变异性，从而使数据可被视为 $\mathbb{S}_1$ 上的概率密度。随后开发了一种基于密度傅里叶级数展开系数的聚类流程，从而考虑了支撑的圆形性质，但忽略了数据的组成性质。

近年来，人们开始质疑使用函数数据分析的方法来分析概率密度，因为这忽略了此类对象受约束的特性。参见 Hron 等人 [2014] 及其中的参考文献。同样，将圆映射到实数线的一个子集，以处理定义在 $\mathbb{S}_1$ 上的数据，可能会产生误导性结果，因为结果取决于所选的映射。例如，参见图 1：我们比较了 Ali 等人 [2021] 的数据集中两个数据点之间的 Wasserstein距离，分别将它们视为实数线上的测度，并采用两种不同的映射将圆“展开”到 $[0, 2\pi]$上。在第 6 节中，我们讨论本文开发的工具如何用于分析 OCT测量值数据。具体而言，通过对 $\mathbb{S}_1$上的测度进行 Wasserstein主成分分析，我们展示了主方向如何帮助解读数据中的主要变异来源。此外，我们通过展示基于得分拟合的层次聚类算法如何有效地将数据划分为视神经形态各异的组，来评估降维流程的有效性。

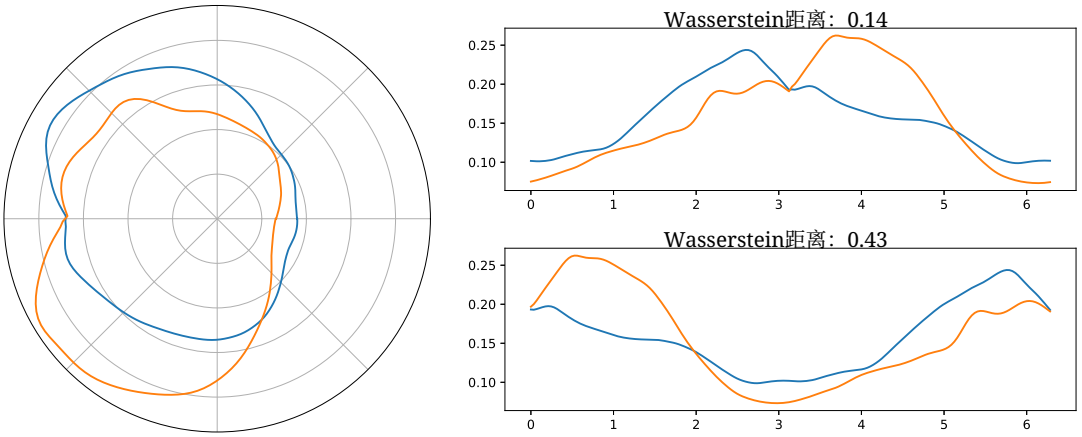


图1：在 $\mathbb{S}_1$ （左）上的两个OCT样本，以及从 0（上）或从 π （下）开始在 $[0, 2\pi]$ （左）上展开时的两个 OCT 样本，以及在 $[0, 2\pi]$ 上的概率测度之间计算得到的相应 Wasserstein 距离。

1.1 相关工作

据我们所知，概率测度的主成分分析虽已在不同背景下提出，但研究重点要么是分析直方图（或离散测度），要么是分析定义在 $\mathbb{R}$中的测度。

针对Wasserstein度量下分布的主成分分析（及相关算法）的不同定义已由Bigot等人 [2017], Cazelles等人 [2018]和Pegoraro和Beraha [2022]提出。在这些工作中，实数线上平方可积概率测度的空间，赋予2-Wasserstein度量（也称为Wasserstein空间），被紧密类比为“黎曼”流形，并利用绝对连续概率测度处切空间的特征 [Ambrosio等人, 2008]来进行统计分析。

当统计单位并非嵌入在线性空间中时，就需要推广多元统计学中的经典工具，以考虑空间的非线性。例如，可以想想 Fréchet 均值如何推广了样本均值的概念。对于支撑在流形上的数据，统计工具可分为外在 或内在 [Bhattacharya 等，2012；Pennec，2006、2008；Huckemann 等，2010；Patrangenaru 和 Ellingson，2015；Fletcher，2013；Banerjee 等， 2015]。外在方法是寻找一个线性空间（通常是某个合适中心点处的切空间）来近似流形（或数据所在的流形区域），并对数据在该线性空间上的投影进行标准的（欧几里得）主成分分析。相反，内在方法利用流形的测地结构，根据数据点与流形中（测地）凸子集之间的距离来定义主成分分析，并将凸子集视为线性子空间的自然推广。需要注意的是，若流形无法得到良好近似，外在技术引入的近似可能会显著影响结果；而内在技术通常计算量较大，不适合分析大型数据集。

聚焦于支撑在

实数集上的测度所构成的 2-Wasserstein空间中的数据情形，我们可以将 Bigot等人 [2017]提出的测地线-主成分分析视为一种内在方法，而 Cazelles 等人 [2018] 提出的对数 主成分分析以及 Pegoraro和Beraha [2022]提出的投影 方法则属于外在工具。这些方法都依赖于明确知道从一个绝对连续测度到任意其他测度的最优传输映射，而这是这一特定情形的特点。此外，Bigot等人[2017], 、Cazelles 等人 [2018], 以及 Pegoraro和Beraha [2022]利用了 2-Wasserstein空间与在适当定义的 L_2 空间中的闭凸锥之间众所周知的等距同构。因此，通过考虑“锥约束”，将概率测度空间的“流形”性质纳入了考量。原则上，Cazelles 等人 [2018]提出的对数-主成分分析可以应用于定义域更复杂的分布。然而，正如 Pegoraro和Beraha [2022], 所讨论的，对数-主成分分析得到的分量难以解释，也无法对得分进行操作，而这通常是主成分分析的一项标准要求。

1.2 我们的贡献与概要

将先前提出的 Wasserstein 主成分分析方法扩展到 $\mathbb{S}_1$ 上的测度时，我们面临若干非平凡的问题。这些问题源于 $\mathbb{S}_1$的非欧几里得性质，而这一点不容忽视。事实上，考虑以下例子：固定一点 $\theta \in \mathbb{S}_1$ ，从 θ开始将圆“展开”，由此得到 $\mathbb{S}_1$ 与 $[0, 2\pi]$之间的一个双射。我们或许会想把 $\mathbb{S}_1$ 上的分布视为实数线某个区间上的分布。然而，如图 1 所示，此时 Wasserstein 度量会依赖于所选的 θ。

黎曼流形上测度的最优传输是一个活跃的研究领域 [McCann, 2001, Gigli, 2011, Kim 和 Pass, 2017]。具体而言，McCann [2001] 给出了最优传输映射的刻画，而 Gigli[2011] 则基于传输映射与传输计划提出了不同的切空间定义。由于其框架具有一般性，所得表达式并不便于计算。

本文的第一个主要贡献是对支撑在 $\mathbb{S}_1$上的测度的最优传输进行详细研究。特别地，我们推导出（几乎）闭形式的最优传输映射表达式，并提出在任意绝对连续概率测度处切空间的另一种定义。与Gigli [2011], 中的一般定义相反，可以显式刻画“对数映射”（即从Wasserstein空间到切空间的映射）的像。此外，我们建立了Wasserstein空间与对数映射的像之间的同胚。

随后，在这些重要结果的基础上，我们提出了一个针对 $\mathbb{S}_1$ 上测度进行主成分分析的框架。我们的方法是在点 $\bar{\mu}$ 处选择一个合适的切空间，并通过对数映射将观测值映射到切空间，从而分析变换后的数据。切空间是希尔伯特空间，因此可以对变换后的数据进行标准主成分分析。然而，对数映射的像是切空间内部的一个凸锥。我们认为，在进行主成分分析以获得可解释的结果时，应考虑这一约束。事实上，正如 Cazelles 等人 [2018],Pegoraro和Beraha [2022], 所讨论的，忽略这一点会导致方向的可解释性较差，也使得无法

在得分上进行操作。从本质上说，这两个问题都是由于主方向在 Wasserstein 空间中作为曲线来看时，可能并不正交（甚至可能不是测地线）。沿着 Bigot等人 [2017], 的思路，我们提出一种嵌套主成分分析，需要在概率测度空间上求解一个变分问题，以找到主方向。通过引入合适的 B 样条近似，我们说明了如何将这类优化问题转化为有限维约束优化问题，并可使用标准约束优化软件对其解进行数值近似。

最后，我们讨论一种近似 Wasserstein 重心的算法，并提议将该算法的输出用作主成分分析的中心点 $\bar{\mu}$ 。我们的算法借鉴了 Zemel 和 Panaretos [2019], 提出的方法，该方法需要显式了解最优传输映射。我们推导出一个充分的事后条件，用于评估算法是否收敛到重心，并通过若干模拟验证该条件；理论分析留待未来工作。

本文结构如下。第 2 节介绍最优传输所需的背景知识。第 3 节给出关于 $\mathbb{S}_1$ 上测度的最优传输主要结果，第 4 节则讨论我们的主成分分析框架以及 Wasserestin 重心的数值近似。第 5 节展示数值示例，其中讨论了对位置-尺度族分布进行主成分分析的模拟研究，并着重比较实数集 R 上测度与 $\mathbb{S}_1$ 上测度的情形。在第 6 节中，我们分析 OCT 测量数据。最后，我们在第 7 节讨论尚未解决的问题和未来工作，并总结全文。证明、更多背景知识及补充结果见附录。

2 最优传输与流形值数据分析的背景

本节简要介绍紧流形上测度的最优传输与 Wasserstein距离。详细论述请参见 Ambrosio等人 [2008] 。技术细节将在附录 A中给出。

2.1 黎曼流形。

非正式地，可以将一个 n 维光滑流形 M 理解为一个局部表现得像欧几里得空间的集合：它可以被一族开集 $(U_i)_{i\geq 1}$ 覆盖，对于这些开集，存在同胚 $\varphi : U_i \to \varphi(U_i) \subset \mathbb{R}^n$，称为坐标图，并满足某些相容性条件。我们可以将 $(U_i, \varphi(U_i))$ 称为流形的局部参数化 。维数为 n 的黎曼流形 (M, g) 是赋有（光滑变化的一族）内积 $g = (g_x)_{x\in M}$ 的光滑流形 M ，这些内积定义在每个点 $x \in M$ 处的切空间 T_xM 上。其切丛 TM 定义为

$$TM := \coprod_{x\in M} T_xM = \bigcup_{x\in M} \{x\} \times T_xM. \tag{1}$$

每个 T_xM 都是一个维数为 n 的向量空间。切丛本身是一个维数为 $2n$ 的光滑流形，并带有标准光滑结构。有关黎曼流形的入门介绍，请参阅 Lee [2013]。

$z \in M$ 处的指数映射记为 $\exp_z : TM \to M$，它可以将切向量 $v \in T_xM$ 映射到流形本身。直观地说，$\exp_z(v)$ 是从 z 出发、方向为 v、经过单位时间后到达的测地线终点。对数映射 $\log_z : M \to TM$ 在其有定义处满足 $\exp_z \circ \log_z(x) = x$。内积 g 诱导出体积测度 ω，它在局部（即在坐标图 (U, φ) 上）由下式给出

$$\mathcal{L}_M(A) = \int_{\varphi(A)} |\det(g(\varphi^{-1}(x)))|^{1/2} \, d\mathcal{L}(x) \tag{2}$$

对于任意可测的 $A \subset U$，且 $\mathcal{L}$ 为勒贝格测度。关于测度论的细节，请参见附录 A。

2.2 Wasserstein空间。

为定义 Wasserstein度量，记 $\mathcal{P}(M)$ 为 M 上概率测度的空间，并令 $c : M \times M \to$ 实数集$\mathbb{R}_+$ 为代价函数。 p-Wasserstein距离是 M 上两个概率测度（例如 μ 和 ν）之间的距离，其定义为

$$W_p(\mu, \nu)^p = \min_{\gamma \in \Gamma(\mu,\nu)} \int_{M \times M} c(x, y)^p d\gamma(x, y), \qquad \mu, \nu \in \mathcal{P}(M) \tag{3}$$

其中 $\Gamma(\mu, \nu)$ 是 $M \times M$ 上所有具有边际 μ 和 ν 的概率测度的集合。如果 c 是下半连续的 [Ambrosio等人， 2008]，则保证存在（至少一个）达到 (3) 中最小值的最优计划 γ^o。定义 (3) 归功于 Kantorovich，可视为 Monge最优传输问题的弱形式，即

$$W_p(\mu, \nu)^p = \inf_{T : T\#\mu = \nu} \int_M c(x, T(x))^p d\mu(x)$$

其中 # 表示前推算子： $T\#\mu(A) = \mu(T^{-1}(A))$ 对所有可测 A。可以证明，当最优映射存在时，它会诱导出最优传输计划 $\gamma^o = (\mathrm{Id}_M, T)\#\mu$，并且这两种形式是等价的。然而，存在几种情况，Monge 的问题没有解。

在下文中，我们将始终将黎曼距离 $\mathrm{d}_R(\cdot, \cdot)$ 作为代价函数，并设置 $p = 2$。我们将重点限制在 2-Wasserstein空间中的测度上，即概率测度的子集

$$\mathcal{W}_2(M) = \left\{ \mu \in \mathcal{P}(M) : \int_M \mathrm{d}_R(x, x_0)^2 d\mu(x) < \infty \text{ for every } x_0 \in M \right\}.$$

这确保 Wasserstein距离始终有限。

2.3 Wasserstein空间的几何。

Wasserstein空间（$\mathcal{W}_2$， W_2）可以赋予一种弱黎曼结构，该结构由 $\mathcal{W}_2$ 在相对于体积测度绝对连续的任意测度处的切空间诱导而来

（2）。与支撑在实数集^{n}中的测度情形一样，切空间是 L^2 上定义的向量值函数空间的子集，其中底空间为 M。其定义还需要一些背景知识。

考虑一个向量场 $v\colon M \to TM$，使得对于每个 $z \in M$, $v_z := v(z) \in T_zM$。更准确地说，记典范投影映射 $\pi\colon TM \to M$ 为 π，即 $\pi(z,\ v) = z \in M$，那么 v 必须满足

$$\pi \circ v = \mathrm{Id}_M$$

其中 Id_M 是 M 上的恒等映射。令 $S(M)$ 表示所有此类向量场的集合。于是，对于一个测度 $\mu \in \mathcal{P}(M)$，我们可以将 L^2_μ 定义为

$$L^2_\mu(M) = \Big\{ v \in S(M) : \int g(v_z, v_z)^2 d\mu(z) < \infty \Big\}. \tag{4}$$

更多细节参见附录 A。对于 $v \in S(M)$，我们可以定义映射 $\exp(v) : M \to M$，使得 $\exp(v)(z) := \exp_z(v_z)$，其中 $z \in M$。采用这一记号，我们可以陈述 McCann 提出的最优传输基本定理 [2001]。

定理 1（最优传输计划的刻画）。设 $\mu, \nu \in \mathcal{W}_2$（$M$）。若 μ 关于体积测度（2）绝对连续，则存在唯一的最优传输计划 γ^o，其形式为 $\gamma^o = (Id_M, T)\#\mu$，其中 $T\colon\ M \to M$。此外，存在一个 d^2_R-凹函数 ϕ，使得 $T = \exp(-\nabla\phi)$。

d^2_R-凹性条件相当技术性，且在下文中并非必需，因此我们仅在附录A的附录中报告该条件；有关更多细节，请参见 Gigli [2011] 。为明确传输映射对源测度和目标测度的依赖关系，我们将使用记号 T^ν_μ 来指代从 μ 到 ν的最优传输映射（最优传输映射）。

最优传输映射的存在性和唯一性启发了以下切空间的定义 [Gigli 的推论 6.4， 2011]

$$\mathrm{Tan}_\mu(\mathcal{W}_2(M)) = \overline{\{v \in L^2_\mu(M) \mid \exists \varepsilon > 0 : (\mathrm{Id}_M, \exp(tv))\#\mu \text{ is optimal for } t \le \varepsilon\}}^{L^2_\mu} \tag{5}$$

正如在黎曼流形的情形中，我们可以定义指数映射和对数映射，使我们能够从切空间 $\mathrm{Tan}_\mu(\mathcal{W}_2(M))$ 移动到 Wasserstein空间，反之亦然。

$$\begin{aligned} \exp_\mu &: L^2_\mu(M) \to \mathcal{W}_2(M), \qquad \exp_\mu(v) = \exp(v)\#\mu \\ \log_\mu &: \mathcal{W}_2(M) \to L^2_\mu(M), \qquad \log_\mu(\nu) = v \text{ s.t. } \exp(v) = T^\nu_\mu \end{aligned} \tag{6}$$

这种结构通常被称为弱黎曼结构 $\mathcal{W}_2(M)$的。

3 圆上的最优传输

本节将第 2 节概述的一般理论专门应用于支撑在单位圆上的测度。

3.1 $\mathbb{S}_1$的几何

为便于讨论，我们将单位圆定义为 $\mathbb{S}_1 := \{z \in \mathbb{C} : |z| = 1\}$，其中 $|\cdot|$ 表示复数的模。我们先介绍 $\mathbb{S}_1$ 的光滑（群）结构，再描述其黎曼结构。

为了赋予 $\mathbb{S}_1$ 群结构，我们首先考虑映射 $\exp_c : \mathbb{R} \to \mathbb{S}_1$，其定义为 $\exp_c(x) = e^{ix}$；以及映射 $\log_c : \mathbb{S}_1 \to \mathbb{R}$，其定义为 $\log_c(z) = x \in [0, 2\pi)$，使得 $z = e^{ix}$。注意，log_c是 $\exp_c$的右逆，即 $\exp_c \circ \log_c = \mathrm{Id}_{\mathbb{S}_1}$。指数映射 $\exp_c$ 通常被称为 $\mathbb{S}_1$的万有覆盖（Munkres, 2000）。显然，我们将复数的乘法作为群运算：$\cdot : \mathbb{S}_1 \times \mathbb{S}_1 \to \mathbb{S}_1$，定义为 $z \cdot w = \exp_c(\log_c(z) + \log_c(w))$。直观地说，$\log_c(z)$是与 z 的极坐标表示相关联的“角”，而 $\cdot$ 是这些角的和。显然，$(\mathbb{S}_1, \cdot)$是一个群，且 $\exp_c : (\mathbb{R}, +) \to (\mathbb{S}_1, \cdot)$ 是一个群同态。

通过 $\exp_c$ 和 $\log_c$，考虑每个 $z \in \mathbb{S}_1$ 处的映射 $\exp_z(x) := \exp_c(x + \log_c(z))$，即指数映射的平移版本，以及满足 $y \in [-\pi/2, \pi/2)$ 的 $\log_z(w) = y$ 和 $\exp_z(\log_z(w)) = w$，我们可以定义 $\mathbb{S}_1$ 的光滑结构。令 $V_z := \mathbb{S}_1 \setminus \{-z\}$，则对每个 $z \in \mathbb{S}_1$，二元组 $(V_z, \log_z)$ 是一个坐标图。借助这一微分结构，$\mathbb{S}_1$ 是一个李群，其切丛为 $T\mathbb{S}_1 = \{(x, v) \mid x \in \mathbb{S}_1 \text{和 } v \in T_x\mathbb{S}_1\} \simeq \mathbb{S}_1 \times \mathbb{R}$。我们将点 $(1, 0)$ 记为 1，它给出了 $\mathbb{S}_1$中的单位元。

我们考虑由嵌入 $\mathbb{S}_1 \hookrightarrow \mathbb{C} \simeq \mathbb{R}^2$诱导的黎曼度量 g，即对于 $x, y \in T_z\mathbb{S}_1 \simeq \mathbb{R}$，有 $g_z(x, y) = xy$。这诱导出的弧长距离为 $d_R(z, w) = |\log_c(z) - \log_c(w)|$。注意，$det(g) \equiv 1$，因此 $\mathcal{L}_{\mathbb{S}_1} = \exp_c \# \mathcal{L}$，或者等价地，$\log_c \# \mathcal{L}_{\mathbb{S}_1} = \mathcal{L}$。于是对于任意 $f : \mathbb{S}_1 \to \mathbb{R}$

$$\int_{\mathbb{S}_1} f(z) d\mathcal{L}_{\mathbb{S}_1}(z) = \int_{-\pi/2}^{\pi/2} f(\exp_c(x)) d\mathcal{L}(x). \tag{7}$$

详见附录 A。

3.2 最优传输映射

借助上一节引入的记号，我们现在关注赋予其黎曼距离 d_R 的 $M = \mathbb{S}_1$ 上的最优传输问题。

基本观察是，$\mathbb{S}_1$ 上的测度 μ 可以等价地表示为 $\mathbb{R}$ 上的周期测度，定义为对任意可测 A，$\widetilde{\mu}(A) := \mu(\exp_c(A))$，这蕴含对任意 $p \in 2\pi\mathbb{Z}$，$\widetilde{\mu}(A) = \widetilde{\mu}(A + p)$，其中 $A + p$相当于将 A 中的所有点平移量 p。然后我们定义与 $\widetilde{\mu}$ 相关联的“周期累积分布函数”为对 $x \in [0, 2\pi]$，$F_{\widetilde{\mu}}(x) = \widetilde{\mu}([0, x))$，并通过规则 $F_{\widetilde{\mu}}(x + 2\pi) = F_{\widetilde{\mu}}(x) + 1$ 将其扩展到实数集上。对于 $\theta \in$ 实数集，令 $F^{\theta}_{\widetilde{\mu}}(x) = F_{\widetilde{\mu}}(x) + \theta$ 表示累积分布函数的垂直平移。注意由 $F^{\theta}_{\widetilde{\mu}}$ 诱导的测度独立于 θ 且始终为 $\widetilde{\mu}$。这很容易从例如 $\widetilde{\mu}([a, b]) = F^{\theta}_{\widetilde{\mu}}(b) - F^{\theta}_{\widetilde{\mu}}(a) = F_{\widetilde{\mu}}(b) - F_{\widetilde{\mu}}(a)$ 得出。

用 $F^{-}_{\widetilde{\mu}}$ 表示相应的分位数函数，即 $F_{\widetilde{\mu}}$ 的（广义）逆函数。我们有 $(F^{\theta}_{\widetilde{\mu}})^{-}(x) = F^{-}_{\widetilde{\mu}}(x - \theta)$。因此，$\theta$ 相当于将分位数旋转

绕圆旋转，旋转因子为 $z_\theta^{-1} = \exp_c(-\theta)$。因此，第 0 分位数 $(F_{\widetilde{\mu}}^{\theta})^-(0)$ 并非 0，而是 z_θ^{-1}。等价地，$F_{\widetilde{\mu}}^{\theta}(y) = \widetilde{\mu}([z_\theta^{-1}, y))$。

利用 Delon 等人 [2010], 的结果，下面的定理给出了两个定义在 $\mathbb{S}_1$上的测度之间最优传输映射的显式刻画。

定理 2. 定义 θ^* 为以下最小化问题的解：

$$\theta^* = \underset{\theta\in\mathbb{R}}{\operatorname{argmin}} \int_0^1 \left(F_{\widetilde{\mu}}^-(u) - (F_{\widetilde{\nu}}^{\theta})^-(u)\right)^2 \mathrm{d}u \tag{8}$$

那么，μ 与 ν 之间的最优传输映射为

$$T_\mu^\nu := \exp_c \circ \left((F_{\widetilde{\nu}}^{\theta^*})^- \circ F_{\widetilde{\mu}}\right) \circ \log_c . \tag{9}$$

注意，(9) 与实数集 R 上测度的最优传输映射表达式密切相关。在这种情况下，令 $\exp_c = \log_c = \mathrm{Id}$ 且 $\theta^* = 0$，即可恢复实数线上的测度的最优传输映射的经典表述。下文中，我们将用 $\widetilde{T}_{\widetilde{\mu}}^{\widetilde{\nu}} := (F_{\widetilde{\nu}}^{\theta^*})^- \circ F_{\widetilde{\mu}}$ 表示与 (8) 中的最优 θ^* 相关联的、$\widetilde{\mu}$ 与 $\widetilde{\nu}$之间的映射。尽管 $\widetilde{T}_{\widetilde{\mu}}^{\widetilde{\nu}}$ 并非“最优”（因为周期测度传输所对应的代价要么为零，要么无界），但鉴于它与 T_μ^ν之间的联系，我们仍将其称为 $\widetilde{\mu}$与 $\widetilde{\nu}$之间的最优传输映射。

下面我们直观地解释最优传输映射 T_μ^ν。注意，对 $(F_{\widetilde{\nu}}^{\theta^*})^-$ 预先复合 $(F_{\widetilde{\mu}})_{|[0,2\pi]}$，得到 $\widetilde{T}_{\widetilde{\mu}}^{\widetilde{\nu}}$，意味着将由 $F_{\widetilde{\mu}}^-$ 确定的分位数传输到 $(F_{\widetilde{\nu}}^{\theta^*})_{|[0,1]}$中相应的平移后分位数，方向为逆时针（这是由 $\exp_c$ 的定义所致）。注意，$T_{\widetilde{\mu}}^{\widetilde{\nu}}((F_{\widetilde{\mu}})^-(0)) = T_{\widetilde{\mu}}^{\widetilde{\nu}}(0) = F_{\widetilde{\nu}}^-(-\theta^*) =: x_{-\theta^*}$ 以及

$$T_{\widetilde{\nu}}^{\widetilde{\mu}}((F_{\widetilde{\mu}})^-(1)) \le T_{\widetilde{\nu}}^{\widetilde{\mu}}(2\pi) = (F_{\widetilde{\nu}}^{\theta^*})^-(1) = F_{\widetilde{\nu}}^-(1-\theta^*) = 2\pi + F_{\widetilde{\nu}}^-(-\theta^*) = 2\pi + x_{-\theta^*}, \tag{10}$$

这意味着最优传输映射会将 $[0, 2\pi)$ 映射到 $[x_{-\theta^*}, 2\pi + x_{-\theta^*})$。因此，我们可以把这种情况看作在两个不同的点将圆“展开”，即对 ν 而言从 $z_{\theta^*}^{-1} = \exp_c(-\theta^*)$ 展开，对 μ 而言从 $1 = \exp_c(0)$ 展开，然后匹配在实数集 R 上诱导的测度。例如，假设 μ 和 ν 关于 $\mathbb{S}_1$ 上的勒贝格测度分别具有密度 f_μ 和 f_ν，那么 $(F_{\widetilde{\nu}}^{\theta})_{|[0,1]}$ 就是与密度 $f_\nu(\exp_c(x))$ 相对应的分位数函数，其支撑集为 $[x_{-\theta}, 2\pi + x_{-\theta}]$。显然，$\mu$ 上不进行任何操作，因此我们将支撑集为 $[0, 2\pi]$ 的 $f_\mu(\exp_c(x))$ 传输到支撑集为 $[x_{-\theta}, 2\pi + x_{-\theta}]$的 $f_\nu(\exp_c(x))$。参数 θ^* 随后选出使 ν 的圆展开过程最优的起始点。

最优传输映射是我们在后续章节中发展的静态方法的基础：从参考分布到第 i 个数据点的最优传输映射 T_i 将发挥“ 切向量 ”的作用，使我们能够用函数空间来近似 Wasserstein空间。因此，刻画 $\mathbb{S}_1$ 上的最优传输映射、了解其性质，并从相关映射 T（即实数集 R 上周期测度之间的映射）的视角对其进行考察，是至关重要的。~

以下定理证明了最优传输映射的一项基本性质。

定理 3. 给定 μ 绝对连续测度和 $\nu \in \mathcal{W}_2(\mathbb{S}_1)$，$\widetilde{T} := (F_{\widetilde{\nu}}^{\theta^*})^- \circ F_{\widetilde{\mu}}$ 是一个最优传输映射，当且仅当：

$$\int_0^{2\pi} \widetilde{T}(u) - u \, \mathrm{d}u = 0. \tag{11}$$

关于定理 3 的评论将在全文中陆续给出，因为它会影响许多后续定义和结果。这里我们只指出，方程（11）与测度 μ 无关，是关于 T 的纯分析/几何条件 $\widetilde{}$。

3.3 弱黎曼结构

如前所述，我们的目标是利用 $\mathcal{W}_2(\mathbb{S}_1)$ 的弱黎曼结构，以获得概率测度数据集更易处理的表示，从而能够使用统计工具。因此，我们现在针对 $M \equiv \mathbb{S}_1$，专门化 $\mathrm{Tan}_\mu(\mathcal{W}_2(M))$ 的定义以及相关的指数映射和对数映射，并将原来的向量场定义转化为更易处理的函数。此外，我们还将确立对数映射的性质，这些性质对于建立一个连贯的统计框架，以分析 $\mathcal{W}_2(\mathbb{S}_1)$ 中的概率测度至关重要。

就我们的目的而言，将 $L^2_\mu(\mathbb{S}_1)$ 定义为如下形式较为方便：

$$\begin{aligned} L^2_\mu(\mathbb{S}_1) :&= \left\{ v : \mathbb{S}_1 \to \mathbb{R} \text{ such that } \int_{\mathbb{S}_1} v^2(x) d\mu(x) < +\infty \right\} \\ &= \left\{ v : [0, 2\pi) \to \mathbb{R} \text{ such that } \int_0^{2\pi} v^2(x) d\widetilde{\mu}(x) < +\infty \right\} \end{aligned}$$

其中第二个等式成立，通过轻微滥用记号，考虑$v \mapsto v \circ \log_c$。注意到通过将 $v(x)$ 识别为 $T_x\mathbb{S}_1$ 的一个元素，我们恢复了 (4) 中的空间。然后，如果 μ 是一个绝对连续测度，我们有

$$\mathrm{Tan}_\mu(\mathcal{W}_2(\mathbb{S}_1)) = \overline{\{v : L^2_\mu(\mathbb{S}_1) \mid \exists \varepsilon > 0 : (\mathrm{Id}_{\mathbb{S}_1}, \exp(tv))\#\mu \text{ is optimal for } t \leq \varepsilon\}}^{L^2_\mu} \tag{12}$$

其中我们可以根据需要将 v 解释为定义在 $\mathbb{S}_1$ 或 $[0, 2\pi)$ 上的函数。现在我们想重写这个定义，使其更易于处理。

首先，注意到 (12) 中的最优性条件等价于存在 ν，使得 $\exp(tv)$ 是 μ 与 ν 之间的一个最优传输映射。然后，根据定理 2 以及 $\exp_z(v_z) = \exp_c(\log_c(z) + v_z)$ 这一事实，(12) 中的向量场 $v\widetilde{}$ $\widetilde{}$可写为 $tv(\log_c(x)) = T(x) - x$，其中 T 如定理 2 所述，因此$\widetilde{}$ $\widetilde{}$最优传输映射为 $\exp_c(x + (T(x) - x)) \equiv \exp_c(T(x))$。因此，我们可以用映射 T 来重述$\widetilde{}$切空间的定义：

$$\mathrm{Tan}_\mu(\mathcal{W}_2(\mathbb{S}_1)) = \overline{\{\widetilde{T} : L^2_{\widetilde{\mu}}([0, 2\pi]) \mid \exists \varepsilon > 0 : \exp_c(\mathrm{Id} + t(\widetilde{T} - \mathrm{Id})) \text{ is OTM for } t \leq \varepsilon\}}^{L^2_{\widetilde{\mu}}} \tag{13}$$

指数映射和对数映射的定义十分自然:

$$\begin{aligned} &\exp_\mu : L^2_\mu(\mathbb{S}_1) \to \mathcal{W}_2(\mathbb{S}_1), \qquad \exp_\mu(\widetilde{T}) = \left(\exp_c \circ \widetilde{T} \circ \log_c\right) \#\mu \\ &\log_\mu : \mathcal{W}_2(\mathbb{S}_1) \to L^2_\mu(\mathbb{S}_1), \qquad \log_\mu(\nu) = \widetilde{T} \text{ s.t. } \widetilde{T}(x) = F^-_{\widetilde{\nu}}(F_{\widetilde{\mu}}(x) - \theta^*) \end{aligned} \tag{14}$$

~其中，对数μ 映射定义中的 θ^* 与定理 2 中相同。注意，$\exp_c \circ T \circ \log_c$ 是 μ 与 ν之间的最优传输映射。此外，根据定理 3，我们注意到，由最优传输映射 T（即 $v(u) = \widetilde{T}(u) - u$）诱导的~向量场 $v : [0, 2\pi) \to$ 实数集，满足式 (11)，沿 $\mathbb{S}_1$ 相对于 $\mathcal{L}_{\mathbb{S}_1}$积分时均值为零。特别地，请注意，该条件不依赖于 μ ，并给出了最优传输映射的一种纯几何刻画。这与其他常用的最优性条件一致，例如传输计划支撑集的循环单调性，以及 Ambrosio等人对实数集上测度的最优传输映射的 Brenier 刻画^n [Ambrosio等人， 2008]。

现在，结合我们刚刚引入的符号，我们进一步给出最优传输映射的若干刻画。这些刻画有助于研究映射 $\log_\mu$ 以及数值算法的实现。

~**定理 4.** 给定 μ 绝对连续测度， T： 实数集 $\to$ 实数集 诱导出一个最优传输映射~，其定义于 μ 与 ν 之间：$= \exp_c \circ T \circ \log_c \#\mu$ 当且仅当

- $\widetilde{T}$单调不减，且对所有 $p \in 2\pi\mathbb{Z}_\sim$，有 $\widetilde{T}(x + p) = \widetilde{T}(x) + p$
- T 满足$(11)_\sim$
- $|T(x) - x| < \pi$ μ-几乎处处。

由前面的结果，可以直接证明以下结论。

推论 1. 设 μ 是 $\mathbb{S}_1$上的绝对连续测度。那么，式(14)中定义的$\log_\mu$ 的像是一个凸集。

此外，以下命题确立了 $\exp_\mu$ 和 $\log_\mu$ 的连续性。

定理 5。 设 μ 是 $\mathbb{S}_1$上的绝对连续测度。那么:

对于任意 $\nu_1, \nu_2 \in \mathcal{W}(\mathbb{S}_1)$

$$W_2^2(\nu_1, \nu_2) \leq \int_{\mathbb{S}_1} d_R^2(T_\mu^{\nu_1}, T_\mu^{\nu_2}) d\mu \leq \| \log_\mu(\nu_1) - log_\mu(\nu_2) \|^2_{L^2_\mu} .$$

特别地，$\exp_\mu$ 映射是连续的;

2. 若 $W_2(\nu, \nu_n) \to 0$ 属于 $\mathcal{W}_2(\mathbb{S}_1)$ 则

$$\| \log_\mu(\nu_n) - \log_\mu(\nu) \|_{L^2_\mu} \to 0,$$

也就是说，对数μ 映射是序列连续的。因此，由于在度量空间中，序列连续性等价于连续性， $\mathcal{W}_2(\mathbb{S}_1)$ 与对数$\mu(\mathcal{W}_2(\mathbb{S}_1))$可通过对数$\mu$ 和指数μ映射而同胚。

我们还给出了定理 5 第 2 项的另一种证明。。更准确地说，这是对一个较弱结果的证明，但我们认为该证明可以推广到其他紧黎曼流形；而定理 5 第 2 项的证明则不能。。

命题 6. 设 σ 是一个绝对连续测度，$\{\mu_t\}_t$ 是一个绝对连续测度序列，且满足 $\mu_t \to \mu_0$ （在 *Wasserstein*度量下），当 $t \to 0$时。此外，假设 σ 和 μ_t 的支撑集是（测地）凸的，且其密度有上界$\tilde{}$ $\tilde{}$并严格大于零。则 $\|T_\sigma^{\mu_t} - T_\sigma^{\mu_0}\| \to 0.$

我们强调，定理 5 确保了 $\mathcal{W}_2(\mathbb{S}_1)$中的测度与其通过 $\log_\mu(\mathcal{W}_2(\mathbb{S}_1))$得到的表示之间具有高度一致性。这种表示并不像 $\mathcal{W}_2$(R)的情形那样是等距表示（参见Pegoraro和Beraha[2022]），但指数映射和对数映射的连续性意味着其具有很强的可解释性。

4 面向 $\mathbb{S}_1$上测度的主成分分析

本节将展示如何利用第 3 节的结果，以外在方式为 $\mathbb{S}_1$ 上的测度构建主成分分析框架，具体做法是考虑 $\mu_1, \ldots, \mu_n \in \mathcal{W}_2(\mathbb{S}_1)$，并将其类比为黎曼流形上的点，参见第 3.3 节。Bigot等人 [2017]、Cazelles等人 [2018],以及Pegoraro和Beraha [2022] 首次利用这种对应关系，在Wasserstein空间中进行推断，并为实数线上的概率测度构建主成分分析；Chen等人 [2021] 和 Zhang等人 [2020] 则分别提出了针对实数集上测度的线性回归模型和自回归模型。

如引言中所述，对于实数线上的测度，Wasserstein空间的弱黎曼结构使得我们能够定义内在和外在方法 [Bigot等人，2017，Cazelles等人，2018，Chen等人，2021，Zhang等人，2020，Pegoraro和Beraha， 2022]。具体而言，由于 $\mathcal{W}_2$(R) 可以通过将每个测度与其对应的最优传输映射相对应，视为适当定义的 L_2 空间中的凸锥，因此内在方法只需考虑“锥约束” [Pegoraro和Beraha， 2022]。如上所述，这种锥表示不适用于 $\mathcal{W}(\mathbb{S}_1)$的情形。因此，开发内在方法需要处理概率测度曲线。我们认为，第 3 节中确立的结果可以成为此类内在方法的首个构建模块。不过，基于定理 5 的第 (3.) 项中的连续性结果，我们提出一种对数主成分分析，它通过将所有数据点映射到适当的切空间后计算得出。事实上，连续性结果表明，我们将数据映射到切空间时所做的近似并不过于粗糙，或者至少应始终能够产生可解释的结果。第 5 节所展示的数值示例似乎验证了这一主张。

4.1 对数凸主成分分析在 $\mathcal{W}_2(\mathbb{S}_1)$ 上

如 Gigli [2011], 的推论 6.6 所示，对于绝对连续测度，切空间是希尔伯特空间，因此我们可以对

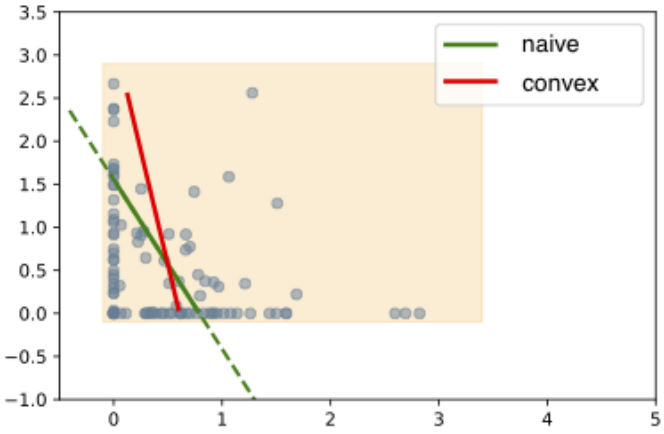


图2：当空间 $H = \mathrm{R}^2$ 且 X 为黄色矩形时，朴素 L_2 方法与凸主成分分析得到的第一主方向。蓝点表示观测值。

$\log_{\bar{\mu}}(\mu_1) \ldots,\ \log_{\bar{\mu}}(\mu_n)$ 应用标准主成分分析技术，其中某个测度 $\bar{\mu}$ 固定不变。我们将这种方法称为“朴素”对数主成分分析（log-PCA）。然而，正如 Pegoraro和 Beraha [2022], 所指出的，忽略对数映射的像 $\log_{\bar{\mu}}$ 并非整个 $\mathrm{Tan}_{\bar{\mu}}(\mathcal{W}_2(\mathbb{S}_1))$ 切空间，而只是其一个凸子集这一事实，可能会导致误导性结果。具体而言，当切空间中的两个元素位于对数映射的像 $\log_{\bar{\mu}}$ 之外时，先返回 Wasserstein空间，再通过 $\log_{\bar{\mu}} \circ \exp_{\bar{\mu}}$ 回到切空间，可能会导致距离和角方面出现不理想的行为。更一般地说，只有当主方向位于该凸子集内时，它才具有可解释性并能捕捉有意义的变异性。例如，当得分和主方向位于 $\log_{\bar{\mu}}(\mathcal{W}_2(\mathbb{S}_1))$ 之外时，其可解释性就会受到影响：这些方向可能并不正交，而且 $\mathcal{W}_2(\mathbb{S}_1)$ 内的方差可能无法恰当地分解。

为避免“朴素”对数主成分分析所带来的问题，我们提出如下对数凸主成分分析的定义：该方法相当于在切空间中进行凸主成分分析 [Bigot等人,2017] ，从而将对数映射像所施加的约束纳入考量。首先引入一些记号。令 $X := \log_{\bar{\mu}}(\mathcal{W}_2(\mathbb{S}_1)), H := \mathrm{Tan}_{\bar{\mu}}(\mathcal{W}_2(\mathbb{S}_1))$。对于一个闭凸集 $C \subset X$ 和一点 $x \in X$ ，令 $d(x, C) = \operatorname{argmin}_{y \in C} \|x - y\|_{L^2_{\bar{\mu}}}$。令 Sp 表示一组向量的张成空间，并令 $\mathcal{C}_{x_0}(U) := (x_0 + Sp(U)) \cap X$，其中 $x_0 \in X$ 且 $U \subset H$。

与 Pegoraro和Beraha [2022], 一样，我们还作出以下技术性假设：给定一组概率测度 $\bar{\mu},\ \mu_0,\ \mu_1,\ \ldots,\ \mu_n \in \mathcal{W}(\mathbb{S}_1)$，我们假设 $\log_{\bar{\mu}}(\mu_0)$ 位于 $\{\log_{\bar{\mu}}(\mu_i)\}$ 的凸包的相对内部。对于 μ_0 ，最常见的选择是将其选为 $\{\log_{\bar{\mu}}(\mu_i)\}$ 的“均值”；由于它位于希尔伯特空间中，在某些病理情况下可能会违反我们的假设。然而，在实际应用中，我们总是采用$L^2_{\bar{\mu}}$的有限维近似，在这种情况下该假设总是成立。更多细节请参见 Pegoraro和Beraha [2022]的附录 A。

定义 1. 考虑一组概率测度 $\bar{\mu}, \mu_0, \mu_1, \ldots, \mu_n \in \mathcal{W}(\mathbb{S}_1)$. 设 $T_i = \text{对数}_{\bar{\mu}}(\mu_i) = T^{\mu_i}_{\bar{\mu}},\ i = 0, \ldots, n$。对于 $\mu_1, \ldots, \mu_n$，一个 $(k, \bar{\mu}, \mu_0)$ 对数凸主成分是满足以下条件的子集 $C_k :\ = \mathcal{C}_{\widetilde{T}_0}(\{w^*_1, \ldots, w^*_k\})$

1. 对于 $k = 1$,

$$w_1^* = \underset{w \in H, \|w\|=1}{\operatorname{argmin}} \sum_{i=1}^{n} d\left(\widetilde{T}_i, \mathcal{C}_{\widetilde{T}_0}(\{w\})\right)$$

2. 对于 $k > 1$,

$$w_k^* = \underset{w \in H, \|w\|=1, w \perp Sp(\{w_1^*, \ldots, w_{k-1}^*\}}{\operatorname{argmin}} \sum_{i=1}^{n} d\left(\widetilde{T}_i, \mathcal{C}_{\widetilde{T}_0}(\{w\})\right)$$

图 2 用一个更简单的例子说明了朴素 L_2 与凸主成分分析的区别：当 $H = \mathbb{R}^2$ 且 X 是凸子集时，若数据接近 X的边界，数据与主成分之间的 L_2 度量会捕捉到凸集之外的变异性。另见 Pegoraro和Beraha [2022]，其中给出了一些指标，用于量化非凸 L_2 主成分分析相较于凸主成分分析的信息损失。

4.2 通过 B 样条逼近计算对数凸主成分分析

凸主成分分析的定义可转化为一个约束优化问题，以求得方向 $\{w_1^*, \ldots, w_k^*\}$。在 Cazelles 等人 [2018], 的研究中，作者对传输映射进行离散化，并通过前向—后向算法求解优化问题。正如 Pegoraro和Beraha [2022], 所讨论的，更高效的方法是用二次 B 样条逼近传输映射，并通过内点法求解约束优化问题。这里我们采用第二种方法。

令 $\{\psi_1, \ldots, \psi_J\}$ 为定义在 $[0, 2\pi]$ 等距节点上的 B 样条基。我们令 $\widetilde{T}_i(x) \approx \sum_{j=1}^{J} a_{ij}\psi_j(x)$，令 $\widetilde{w}_k(x) \approx \sum_{j=1}^{J} a_j\psi_j(x)$。注意，若样条为二次样条，则当且仅当系数 $a_1, \ldots, a_J$ 为非减时，函数 $\widetilde{w}_k$ 单调不减（例如，参见 Pegoraro和Beraha 的命题 4, 2022）。因此，以下我们将 ψ_j 视为定义在 $[0, 2\pi]$ 上的二次样条基函数。样条基展开还可加快 L_2 内积的计算：令 E 为一个矩阵，其元素为 $E_{i,j} = \int_0^{2\pi} \psi_i(x)\psi_j(x)\mathrm{d}x$，且 $\boldsymbol{a}_i = (a_{i,1}, \ldots, a_{i,J})$，则有 $\langle \widetilde{T}_i, \widetilde{T}_j \rangle = \langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle_E := \boldsymbol{a}_i^T E \boldsymbol{a}_j$。我们将 $\|\cdot\|_E$ 记为相应的范数。

与 Pegoraro和Beraha 的命题 6 类似，我们得到，对数凸主成分分析的第k个方向 $\boldsymbol{w}_k$及其相关得分 $\lambda_{1:n}^k = \lambda_1, \ldots, \lambda_n$（观测值的第 k个方向）可以通过求解一个约束优化问题来计算。目标函数为：

$$\lambda_{1:n}^k, \boldsymbol{w}_k = \underset{\lambda_{1:n}, \boldsymbol{w}}{\operatorname{argmin}} \sum_{i=1}^{n} \|\boldsymbol{a}_i - \boldsymbol{a}_0 - \sum_{j=1}^{k} \lambda_i^k \boldsymbol{w}_k\| \tag{15}$$

其中，$\lambda_i \in \mathbb{R}$ 是沿第 k个方向的第 i个数据点的得分。此外，还必须满足通常的正交约束和单位范数约束：

$$\|\boldsymbol{w}\|_E = 1, \quad \langle \boldsymbol{w}_h, \boldsymbol{w} \rangle_E = 0, \quad h = 1, \ldots, k-1.$$

除此之外，我们还必须要求 $\sum w_j\psi_j$ 属于 $H := \mathrm{Tan}_{\bar{\mu}}(\mathcal{W}_2(\mathbb{S}_1))$。单调性约束等价于

$$\lambda_i w_j + a_{0,j} - \lambda_i w_{j-1} - a_{0,j-1} \geq 0, \quad j = 2 \dots J$$

也就是说，样条系数具有单调性（因为样条是二次的。参见 Pegoraro和 Beraha [2022]中的命题 4）。此外，“周期性”约束在设计上已得到满足。为施加 (11)，令 $M_j = \int \psi_j(u)\mathrm{d}u$，则 (11) 等价于

$$\sum w_j M_j = 2\pi^2.$$

最后，由于 (11)，只需控制函数 w 在初始点处的取值，即 $w_0 \in (-\pi/2,\ \pi/2)$。

我们使用 Python 包 pyomo 实现由此得到的约束优化问题，并使用 Ipopt 求解器通过内点法近似求解。

4.3 Wasserstein重心

接下来需要讨论主成分分析中基点 μ_0 的选择，以及考虑切空间时所对应的测度 $\bar{\mu}$。在非欧几里得空间中进行主成分分析时，一种标准做法是将 μ_0 和 $\bar{\mu}$都设为重心，即弗雷歇均值。在我们的情形中，重心使下列弗雷歇泛函最小：

$$F(\nu; \mu_1, \dots, \mu_n) = \frac{1}{2n} \sum_{i=1}^{n} W_2^2(\nu, \mu_i). \tag{16}$$

原则上，log-PCA 可以通过在任意绝对连续测度的切空间中进行计算。但更推荐将 PCA 嵌入重心处的切空间中，因为直观上，这样会使重心处切空间中数据点之间的距离更接近 Wasserstein空间中的距离。随着与切点距离的增加，切空间所提供的近似质量会下降，因此选择数据集的重心作为切点，有助于尽量减小近似所产生的平均误差。因此，主成分的投影可以解释为数据集“平均值”的偏离。请注意，在我们的情形中，以重心为中心进行 PCA 在概念上没有问题，因为只要测度 μ_j 中至少有一个是绝对连续的，Wasserstein重心就是唯一的。参见 Kim 和 Pass[2017] 中的定理 3.1。Agueh 和 Carlier [2011] 对支撑在 $\mathbb{R}^d$ 上的测度也得出了类似结果。

用于计算式 (16) 解的数值算法已由 Carlier 等人 [2015], Srivastava 等人 [2015] 针对原子测度的情形提出，此时优化问题可化简为线性规划。Zemel和Panaretos [2019]则提出了一种基于梯度下降的 Procrustes 算法，适用于 $\mathbb{R}^d$ 上的一般测度（其中至少有一个测度是绝对连续的）。简而言之，Zemel和Panaretos [2019] 中的梯度下降算法从重心的一个初始估计开始，并通过以下方式对当前估计进行前推 ν_r

算法1。Procrustes 重心

输入 测度 μ_1，……，μ_n，起始点 ν，阈值 ε。**重复**~按照定理 2 计算最优传输映射 $T_\nu^{\mu_i}$。设定$\widetilde{\nu}' := \left(\frac{1}{n}\sum_{i=1}^{n} \widetilde{T}_{\tilde{\mu}}^{\mu_i}\right)_{\#\widetilde{\nu}}$ **直到** $W_2(\nu, \nu') < \varepsilon$ 输出 $\bar{\mu} = \exp_c \circ(\widetilde{\nu}')$。**结束**

传输映射的平均值，其中映射从 ν_r 指向所有测度。在测度 μ_i' 上满足某些技术条件时，可以保证该过程收敛到重心。特别是，当测度兼容 [时，该过程在一次迭代内即可收敛，参见 Panaretos 和 Zemel， 2020] 的第 2.3.2 节。该方法的缺点是每次迭代都需要求解 n 最优传输问题；对于支撑在实数集 R 上的测度或位置-散布族之外的情形，这可能颇具挑战性，因为后两种情形存在显式解 [Alvarez-Esteban 等人， 2018]。Cuturi and Doucet [2014] 采用了不同思路，通过在式 (16) 中引入“熵正则化”项，提出了弗雷歇均值的近似解，从而使优化更容易。

本文提出使用 Zemel 和 Panaretos [2019] 开发的梯度下降算法。事实上，我们的定理 2 使得最优传输问题几乎可以得到显式解。此外，正如 Delon 等人 [2010]，所示，式 (8) 中的优化问题关于 θ 是凸的，因此求得 θ^* 很简单。我们在算法1中给出伪代码。

我们想指出，目前我们尚未能够证明该算法在一般情况下收敛到重心，也尚未证明在我们的框架下，这种 Procrustes 算法是否等同于梯度下降。从技术角度来看，Zemel和Panaretos [2019]中的证明并不适用于我们的情形，因为这些证明基于 Ambrosio等人[2008] 在定理10.2.2和10.2.6中给出的 Wasserstein距离次可微性和超可微性结果，而这些结果针对的是可分希尔伯特空间上的测度。尽管如此，以下结果给出了算法1收敛的一个充分条件。

命题7. 设 μ^* 是 $\mathcal{W}(\mathbb{S}_1)$中的一个绝对连续测度，且 $\mu_1, \ldots, \mu_n$是 $\mathcal{W}(\mathbb{S}_1)$中的测度。若对于任意 $i, j = 1, \ldots, n$

$$\|\log_{\mu^*}(\mu_i) - \log_{\mu^*}(\mu_j)\|_{L^2_{\mu^*}} = W_2(\mu_i, \mu_j),$$

然后令 $\bar{T} := n^{-1}\sum_{i=1}^{n} T_{\mu^*}^{\mu_i}$ 为这些对数μ^*（μ_i）的重心，我们有 $\bar{T}\#\mu^*$ 是 $\mu_1, \ldots, \mu_n$ 的 *Wasserstein* 重心。

定理 7 中的条件有一个实际优势：算法1终止后，可以轻松检验该条件。事实上，如果 $\|\log_{\bar{\mu}}(\mu_i) - \log_{\bar{\mu}}(\mu_j)\|_{L^2_{\mu^*}} =$

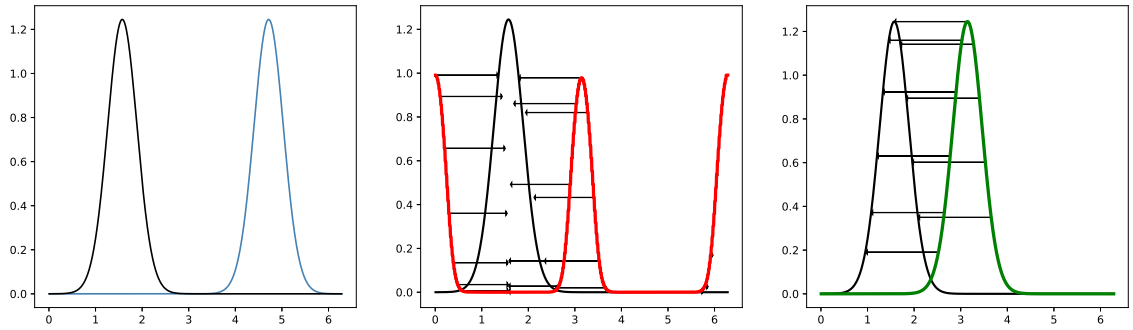


图 3：从左到右： $\mathbb{S}_1$ 上的两个测度（在 $[0,1]$上展开）、 $\mathbb{S}_1$ 上的重心（红色）及其向最左侧测度的传输、 $\mathbb{R}$ 上的重心及其向最左侧测度的传输

$W_2(\mu_i,\ \mu_j)$，其中 $\bar{\mu}$ 是算法1的输出，我们可以确定 $\bar{\mu}$ 是重心。直观而言，如果 Wasserstein 距离与切空间中的距离相近，这意味着连接各数据点的测地线上曲率较小。因此，求解 Wasserstein 重心的问题就简化为对分位数求平均。由此，算法1的输出应当是准确的。在下一节中，我们将通过检验定理 7 中的条件，并将算法1的输出与 Cuturi and Doucet [2014] 提出的 Sinkhorn 算法的结果进行比较，提供算法收敛性的实证依据。

备注 1. 虽然定理 *7* 的表述针对的是 $\mathbb{S}_1$上的测度，但对于定义在指数映射非扩张的一般连通紧致有限维黎曼流形上的测度，该定理同样成立。例如，具有正曲率的流形便属于这种情形。我们在附录 *B.6* 中证明了这一更一般情形下的结果。

5 数值示例

本节介绍关于第 4 节中定义的 Wasserstein重心和主成分分析的数值模拟。

5.1 重心的数值模拟

下面举例说明在考虑 $\mathbb{S}_1$上的分布时可能出现的一些特殊现象。考虑图 3 最左侧面板中的两个测度。当传输代价采用欧几里得度量时，得到的重心如最右侧面板所示：它具有单峰密度，尺度与两个测度相同，且中心恰好位于二者正中间。若改为在 $\mathbb{S}_1$上计算代价，重心则会变为双峰，如图 3 中间面板所示。在这个具体例子中，将“正确”重心传输到两个测度上的代价（在 $\mathbb{S}_1$上）比传输“欧几里得”重心的代价低 30%。

下面给出一些重心的示例。在下文中，我们用 $\bar{\mu}$ 表示从算法1返回的、定义在 $\mathbb{S}_1$ 上的测度，用 $\tilde{\mu}$ 表示其对应的周期测度

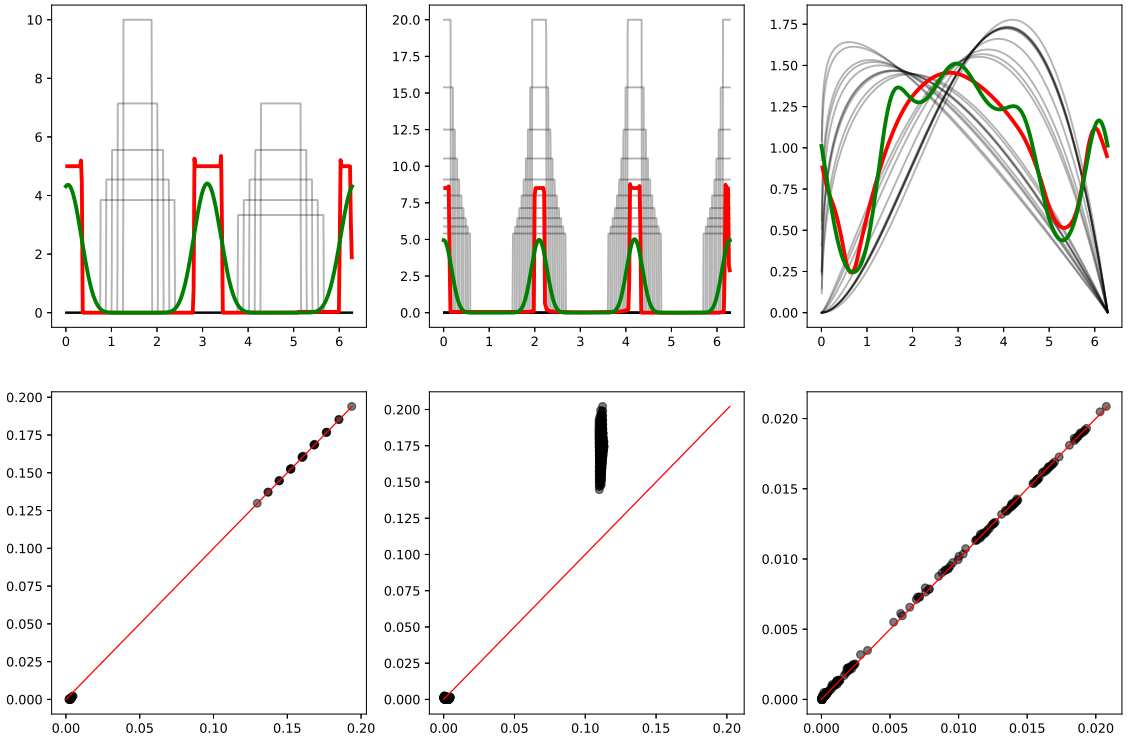


图4：上排： $\widetilde{\mu}_j$在 $[0, 2\pi]$上的密度，以及Wasserstein重心和Sinkhorn重心（分别以红线和绿线表示）。下排：每一对可能的测度的Wasserstein距离与 $d_{\log}$ 的关系。

定义在实数集上。在某些情况下，重心的结果直观可知，我们也展示了算法能够正确收敛到该结果。在另一些情况下，直觉并不可靠，但我们仍可通过比较Wasserstein距离 $W_2(\mu_i, \mu_j)$ 与定理 7 中切空间内的距离，来了解重心近似的效果。此外，我们还将算法1的输出与所谓的 Sinkhorn 重心进行比较。后者由 Cuturi and Doucet（2014）和 Janati et al. 提出，并在 Python 软件包 ott-jax 中实现。为了计算 Sinkhorn 重心，我们将每个测度近似为一个原子测度，在 $[0, 2\pi$ 上设置 1，000 个等间距支撑点，并配备 $\mathbb{S}_1$ 上的测地距离，再为每个点 x_i 赋予与 $\mu(\mathrm{d}x_i)$ 成比例的权重。简单来说，我们预期 Wasserstein 重心与 Sinkhorn 重心相似，但由于 Sinkhorn散度中包含正则化项，Sinkhorn 重心应当更平滑。

我们考虑如下三个模拟数据集。令 $\mathcal{U}(c, w)$ 表示以 c 为中心、宽度为 w 的均匀测度，即在 $(c - w/2, c + w/2)$ 上的均匀测度。在第一个示例中，测度是

$$\widetilde{\mu}_i = \mathcal{U}\left(0.25, 0.1 + 0.05i\right), \quad i = 1, \ldots, 5$$
$$\widetilde{\mu}_i = \mathcal{U}\left(0.75, 0.1 + 0.05(i - 5)\right), \quad i = 5, \ldots, 10$$

并在整个 $\mathbb{R}$上进行周期性延拓。相反，在第二种情形中

$$\widetilde{\mu}_i = \mathcal{U}\left(0, 0.05 + 0.015i\right), \quad i = 1, \ldots, 10$$

$$\widetilde{\mu}_i = \mathcal{U}\left(1/3, 0.05 + 0.015(i-10)\right), \quad i = 11, \ldots, 20$$
$$\widetilde{\mu}_i = \mathcal{U}\left(2/3, 0.05 + 0.015(i-20)\right), \quad i = 21, \ldots, 30$$

相反，在第三种情形中，我们先考虑定义在 (0，2π) 上、参数为 (a_i，2) 的 Beta 分布，再将其进行周期性延拓，由此生成 $\widetilde{\mu}_i$。具体而言，对于 $i = 1$，……，10，取 $a_i \sim \mathcal{U}$(1.3，0.2)；对于 $i = 11$，……，20，取 $a_i \sim \mathcal{U}$(2.6，0.4)。图 4 展示了算法1和 Sinkhorn 方法针对三个不同模拟数据集得到的 Wasserstein重心。可以看出，Wasserstein重心与 Sinkhorn 重心一致，而 Sinkhorn 方法得到的重心通常更平滑。此外，在第一个和第三个例子中，对数距离与 Wasserstein距离无法区分，这表明算法1收敛；而在第二个例子中，两者存在一些差异。第三次模拟使我们得以深入了解 $\mathcal{W}_2(\mathbb{S}_1)$ 的几何结构。事实上，值得注意的是，尽管所有测度$\widetilde{\mu}_i$ 都是单峰的，重心却是双峰的。这显然源于 $\mathbb{S}_1$ 的流形结构，具体而言，是因为质量沿着连接某些测度的测地线经过 0。

5.2 主成分分析的模拟

本节分析一些模拟数据集，用以展示并解读前几节所定义的主成分分析的一些特性。附录C中还提供了另一项模拟及更多细节和比较。为了理解主成分分析所找到的主方向，我们绘制 $\exp_{\bar{\mu}}(\log_{\bar{\mu}}(\mu_0) + \lambda w_k^*)$ 的密度图，其中 w_k^* 是第 k 个主方向，λ 在针对具体情形指定的某个范围内变化。除非另有说明，$\bar{\mu}$ 和 μ_0 均为使用算法1近似得到的 Wasserstein重心。特别要注意，得分 λ 表示沿着方向由第 k 个主方向指定的测地线，从基点出发所经过的距离。因此，可以比较各项模拟中 λ 的不同取值，以了解测度在距离重心多远处开始出现某些变化（例如，测度可能在距离重心达到某一距离时从单峰变为双峰）。

首先，我们考虑一个来自vonMises分布的样本，其位置为 π、尺度为 α，其在 $[0, 2\pi]$ 上的密度函数为

$$f(x; \alpha) = \frac{\exp\left(3\cos\left(\frac{x-\pi}{\alpha}\right)\right)}{2\pi I_0(3)}, \tag{17}$$

其中，I_0 是零阶修正贝塞尔函数。我们通过分别取 $\alpha \sim \mathcal{U}$(0.8, 1.5)和 $\alpha \sim \mathcal{U}$(2, 3.5)，模拟由（17）生成的两个数据集，每个数据集包含 $n = 100$ 个测度。数据和第一主方向如图5所示。第一种情况下，测度足够集中，因此在与 λ取值网格相关的重心邻域内，$\mathbb{S}_1$ 的周期性实际上无关紧要，第一主方向反映了分布尺度的变化。另一方面，在第二种情况下，数据集中所有分布在0附近都有相当多的质量，且这些分布的方差范围比第一个数据集更大。因此

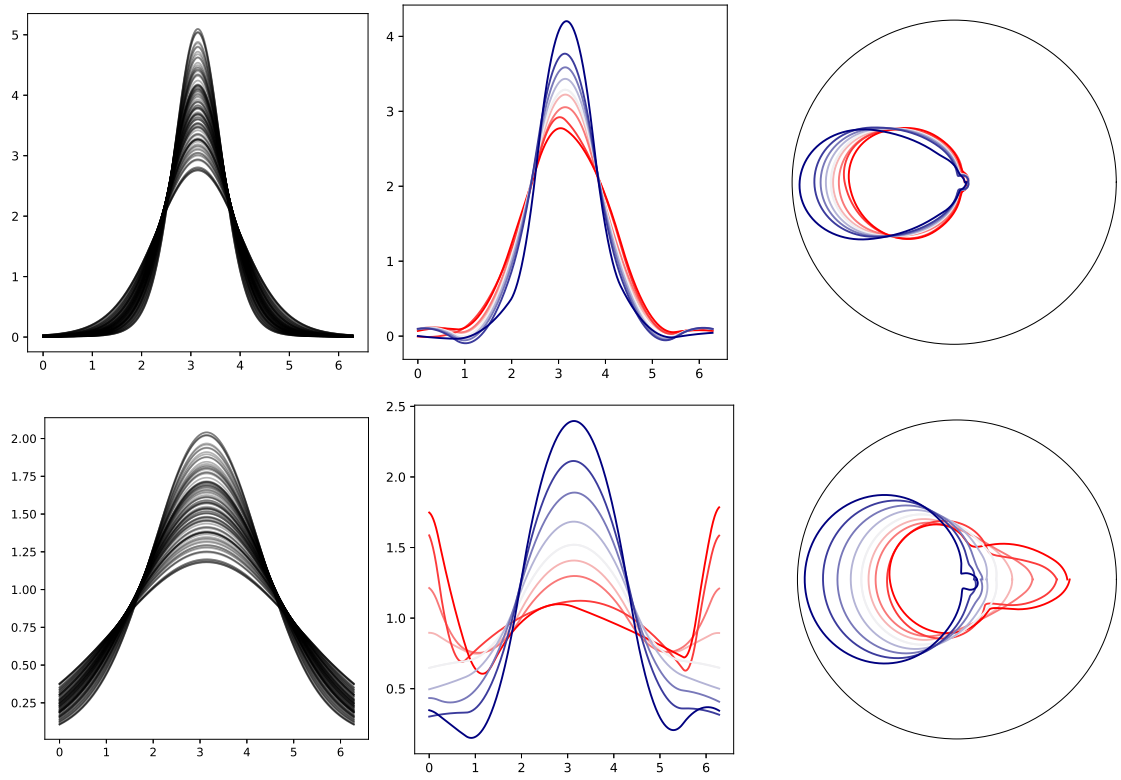


图5：冯·米塞斯模拟的数据及第一主方向。第二列和第三列表示沿第一主方向的密度； λ在 -0.1 （最深蓝色）至0.1（最深红色）之间变化时，分别将其绘制为 $[0, 2\pi)$ 上的分布和 $\mathbb{S}_1$ 上的分布。

沿第一主方向移动（尺度与前一个示例相同）时，我们会以更快的速度持续将质量推向“两侧”，使其进一步集中到0附近，并从单峰密度变为双峰密度。

此处未展示：将相同的测度视为$\mathcal{W}_2$(实数集)中的点时，两种情况下，第一主方向都对应于测度尺度的变化，而位置保持不变。

接下来，我们考虑与第5.1节第三次模拟相同的数据集。图6展示了前两个主方向。第一主方向主要对应于位置的偏移，同时也反映了位于0处的第二个众数附近密度的下降（参见图4中的重心）。从重心（白色）出发，若朝红色密度方向移动，可以看到零点处的众数逐渐并入主众数；若朝蓝色密度方向移动，0处的众数则穿过圆，并与主众数合并，但位于图的右侧。这符合 $\mathcal{W}(\mathbb{S}_1)$的测地结构。相比之下，第二主方向更清晰地体现了两类分布的区别：一类是在0附近有相当多质量的分布（蓝色），另一类则是全部质量都远离0的测度（红色）。

总之，这些模拟有助于我们理解 $\mathcal{W}_2(\mathbb{S}_1)$的几何结构，尤其是它与 $\mathcal{W}_2$(实数集)之间的差异。事实上，众所周知，对于测度

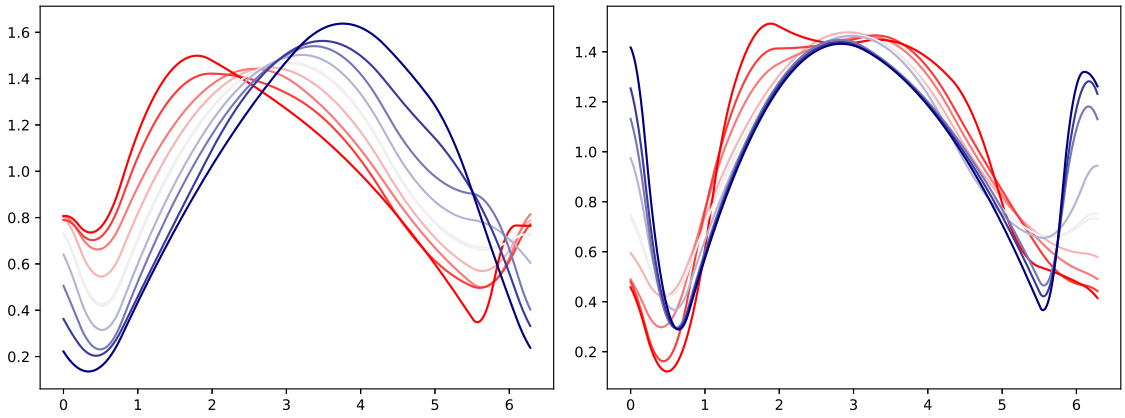


图6：如第5.1节所示，Beta分布沿前两个主方向的密度，其中 λ 从 -0.05 （最深蓝色）变化到 0.05（最深红色），并绘制为 $[0, 2\pi$ 上的分布）。

定义在实数集上的位置-尺度族，其 Wasserstein测地线可通过将位置-尺度平面中的欧氏测地线提升到Wasserstein空间得到。因此，Wasserstein主成分分析能够区分位置效应和尺度效应。相反，正如我们的模拟所示，当测度支撑在 $\mathbb{S}_1$ 上时，无法完全分离位置和尺度的效应。此外，即使数据点是单峰的，重心也常常是多峰的。多峰性会遗传到沿主方向变化的测度中，这可能使其解释变得复杂。在附录C中，我们报告了一个关于主成分分析的补充模拟，讨论点 $\bar{\mu}$ （切空间附着于该点）的选择及其对方向可解释性的影响。具体而言，我们考虑一个由截断高斯分布构成的数据集，其重心有三个众数。相反，如果选择 $\bar{\mu}$ 等于某个数据点，那么沿主方向移动所得到的密度就是单峰的，因此易于解释。当然，这带来了一个概念上的问题：主方向本身并非“主要变异方向”，而是从某个特定 $\bar{\mu}$出发的主要变异方向。

6 案例研究：眼睛数据集

这里展示了将主成分分析应用于 Ali 等人提供的 NRR OCT测量值数据集后得到的结果。 [2021], 该数据集可在其补充材料中找到，包含 3973 位患者的 OCT测量值，并按年龄组分层。具体而言，我们通过解读主方向并对得分进行聚类来评估Wasserstein主成分分析的适用性，结果表明这些簇能够有意义地捕捉数据中的形状模式。图 1 展示了这些数据，以及通过算法1求得的 Wasserstein重心。在最右侧的图中，我们展示了重心处切空间中的 Wasserstein 距离与 L_2 距离几乎对所有数据点对都一致，从而验证了在图 1 中使用红色测度作为主成分分析中心点的合理性。

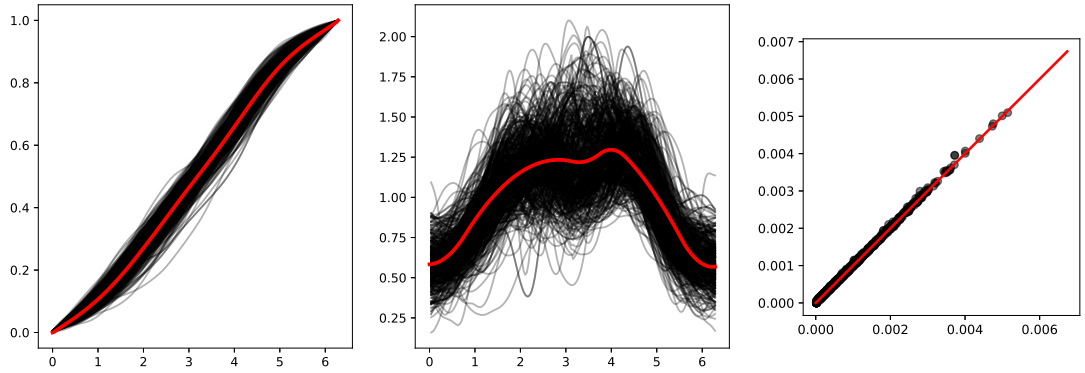


图7：从左到右：（部分子样本的）眼睛的数据集测度的累积分布函数（红线表示重心）、眼睛的数据集测度的概率密度函数（红线表示重心）、重心处切空间中相对于 $d_{\log}$ 的Wasserstein距离。

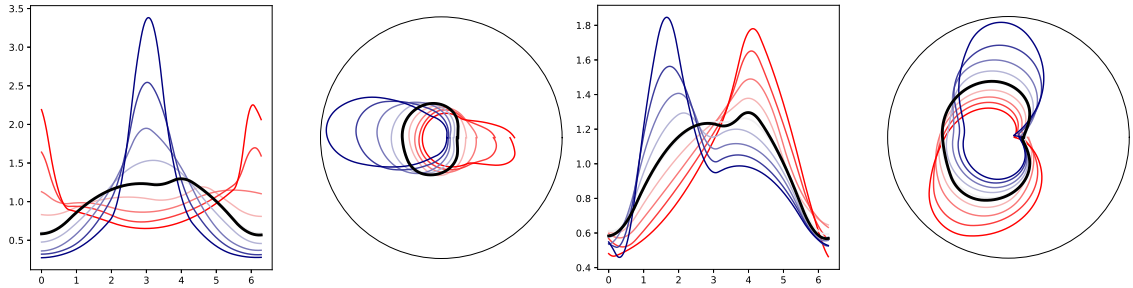
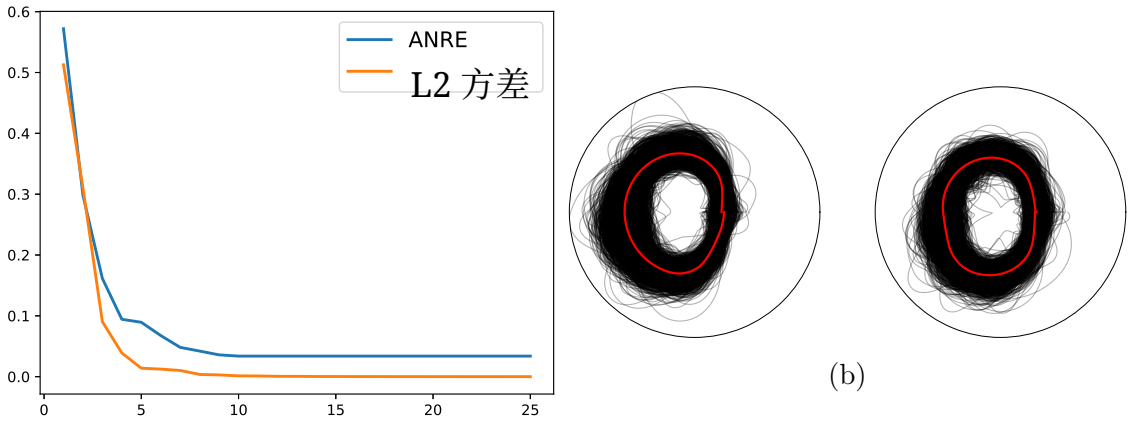


图 8：第一（左侧图）和第二（右侧图）主方向：我们展示了 $[0,1]$ 上的概率密度函数（第一和第三个面板）以及极坐标图（第二和第四个面板）。黑线表示重心。

前两个主方向按构造能够捕捉最多变异性，如图 8 所示。可以清楚地看到，它们将沿水平轴和垂直轴的形状变异性分离开来。具体而言，这意味着数据中的大部分变异性来自沿水平轴的视神经厚度分布变化。为评估 Wasserstein主成分分析在该数据集上的适用性，我们计算平均归一化重构误差随用于主成分分析的方向数量 k 变化的情况。

$$\mathrm{ANRE}_k := \frac{1}{n}\sum_{i=1}^{n}\frac{W_2^2(\mu_i,\mu_i^k)}{W_2^2(\bar{\mu},\mu_i)},$$

其中 μ_i^k 是测度 μ_i 在前 k 个主成分上的投影。平均归一化重构误差指数衡量近似误差，通过数据点相对于主成分分析中心的偏差进行归一化，这与欧几里得空间中主成分分析的方差分解非常类似。图 9（左图）报告了平均归一化重构误差指数作为 k 的函数，以及切空间中 L_2 主成分分析的（归一化）特征值。这两种度量都表明前 $k=5$ 个方向足以捕捉数据集的变异性。此外，L_2 方差比 ANRE 下降得更快。这是预期的，因为 L_2 主成分分析忽略了数据被约束在对数映射的像上，并且也“捕捉”了该集合之外的变异性。最后，我们



(a) (b)

图 9：平均归一化重构误差指数随维度变化的情况，以及各分量解释的方差比例 (L_2)（左图）；数据被分为两个簇及其对应的 Wasserstein重心（红线）（右图）。

认为 ANRE 稳定到一个正的（小）数是由于数值误差。在附录 D 中，我们报告了沿前两个方向的得分散点图，按年龄组分层。从图中可以清楚地看出，在前两个分量上，年龄单独对视神经形状没有明显影响。

我们使用层次聚类算法，根据沿前 $k = 5$ 个主方向的得分对数据点进行聚类，采用 Ward 链接方法。附录 D 展示了树状图，而图 9 展示了找到的两个主要簇。图 12 展示了将树状图截断以获得 7 个簇后得到的精细化聚类。我们用红色标出了各簇的重心，这或许有助于解读这些簇，尽管我们的聚类流程并不像 K-means 算法那样以重心为基础。观察图 9(b) 中的两个簇，可以清楚地看出，它们对应着视神经的两种不同形状；左侧的簇在左边有一个明显的隆起。附录中图 12 所示的精细化簇也呈现出有趣的模式，详情请参阅附录。

我们通过强调以下这一点来结束本节：正如引言中所述，主成分分析的一个非常重要的副产品是，将数据投影到主成分上后，可以对我们的数据集应用多元统计学中的经典工具。我们将补充无监督分析的工作留待后续研究，并进一步考察原始数据集中包含的协变量。

7 讨论

本文探讨了如何分析支撑在圆上的分布型数据。顺应统计学和机器学习领域的近期趋势，我们尝试使用Wasserstein距离比较概率分布。为此，我们研究了 $\mathbb{S}_1$ 上的最优传输问题，并建立了若干新的理论结果，这些结果本身也具有独立的研究价值。具体而言，

我们给出了最优传输映射的显式刻画。鉴于黎曼流形上的最优传输理论尚未成熟，而目前只有实数线上的测度才有此类显式公式，这一结果颇为出人意料。我们还进一步探究了Wasserstein空间的弱黎曼结构，并证明了指数映射和对数映射具有强连续性，同时显式刻画了对数映射的像。

基于我们的理论发现，我们提出了 Bigot等人 [2017] 针对 $\mathbb{S}_1$上测度的凸主成分分析的对应方法。遵循 Pegoraro和Beraha [2022], 的方法，我们提出了一种通过 B样条展开计算主方向的数值方法，由此得到一种易于实现的数值算法。

我们的主成分分析定义需要一个“中心点”，通常将其设为重心。我们使用 Zemel和Panaretos [2019] 的算法近似计算 Wasserstein重心。然而，我们尚未能证明该算法在我们的设定下收敛。尽管数值模拟似乎确实验证了算法1的适用性，但其理论分析仍是一个未解决的问题。

我们的研究为若干有趣的拓展铺平了道路。首先，自然可以考虑 Wasserstein回归问题。借助最优传输映射的表达式，Fletcher [2013] 中的测地线回归可以类似地定义为测度值因变量。类似地，我们的切空间定义适用于为 $\mathbb{S}_1$ 上的测度定义对数回归。对于实数集R上的测度，Pegoraro和Beraha [2022] 提出将因变量和自变量都映射到同一个切空间，因为 Wasserstein空间与任意切空间同构。在这里，更合适的做法是考虑两个切平面：一个用于自变量，另一个用于因变量，并分别以各自的重心为中心，这与 Chen等人[2021]的做法类似。

更广泛地说，我们认为，最优传输与分布数据分析之间的相互作用能够推动这两个领域的进一步发展。针对具有特定统计意义的情形（例如球面），专门研究最优传输理论，一方面有助于更好地理解切空间的性质与底流形之间的关系；另一方面，也可能促成数据分析框架，从地球相关分布及如今收集的其他相关数据中提取洞见。

参考文献

M. Agueh 和 G. Carlier。Wasserstein空间中的重心。《*SIAM* 数学分析杂志》，43(2):904–924，2011。ISSN 0036-1410。doi: 10.1137/100805741。URL https://doi.org/10.1137/100805741。

M.Ali, B.Wainwright, A.Petersen, G. B. Jonnadula, M. Desai, H. L. Rao, M. Srini- vas, S. R. Jammalamadaka, S. Senthil, S. Pyne 等。用于精准识别眼部结构表型的 OCT 数据循环函数分析。*Scientific Reports*，11(1):1–13，2021。

P. C. Alvarez-Esteban, E. del Barrio, J. A. Cuesta-Albertos 和 C. Matran。Wasserstein空间中的广泛共识聚合：在位置-散布族中的应用。《伯努利》，24(4A):3147–3179，2018。L. Ambrosio, N. Gigli 和 G. Savaré。梯度流：度量空间与概率测度空间中的梯度流。Springer Science & Business Media，2008。L. Ambrosio, F. Glaudo 和 D. Trevisan。关于二维随机匹配问题中的最优映射。*arXiv* 预印本 *arXiv:1903.12153*，2019。M. Banerjee, R. Chakraborty, E. Ofori, D. Vaillancourt 和 B. C. Vemuri。黎曼流形上的非线性回归及其在神经影像分析中的应用。载于国际医学图像计算与计算机辅助干预会议，第719–727页。Springer，2015。R. N. Bhattacharya, L. Ellingson, X. Liu, V. Patrangenaru 和 M. Crane。流形上的外在分析比内在分析计算更快，并应用于机器视觉质量控制。应用随机模型：商业与工业，28 (3):222–235，2012。J. Bigot, R. Gouet, T. Klein 和 A. López。通过凸主成分分析在Wasserstein空间中进行测地线主成分分析。载于亨利·庞加莱研究所年刊（*B*辑）：概率与统计，第53卷，第1–26页。亨利·庞加莱研究所，2017。G. Carlier, A. Oberman 和 E. Oudet。团队匹配与Wasserstein重心的数值方法。*ESAIM*数学建模与数值分析，49(6):1621–1642，2015。E. Cazelles, V. Seguy, J. Bigot, M. Cuturi 和 N. Papadakis。Wasserstein空间中直方图的测地线主成分分析与对数主成分分析之比较。*SIAM*科学计算杂志，40(2): B429–B456，2018。Y. Chen, Z. Lin 和 H.-G. Müller。Wasserstein回归。美国统计学会杂志，0 (ja):1–40，2021。M. Cuturi 和 A. Doucet。Wasserstein重心的快速计算。载于国际机器学习会议，第685–693页。PMLR，2014。M. Cuturi, L. Meng-Papaxanthos, Y. Tian, C. Bunne, G. Davis 和 O. Teboul。最优传输工具（ott）：用于各种Wasserstein问题的JAX工具箱。*arXiv* 预印本 *arXiv:2201.12324*，2022。J. Delon, J. Salomon 和 A. Sobolevski。圆上Monge代价的快速传输优化。*SIAM*应用数学杂志，70(7):2239–2258，2010。P. Fletcher。黎曼流形上的测地线回归与最小二乘理论。国际计算机视觉杂志，105，2013年11月。N. Gigli。关于Brenier-McCann定理逆命题及$(P_2(M),\ W_2)$的结构。分析方法与应用，18(2):127–158，2011。

K. Hron、A. Menafoglio、M. Templ、K. Hrůzová 和 P. Filzmoser。贝叶斯空间中密度函数的单纯形主成分分析。计算统计与数据分析，94：330–350，2014。S. Huckemann、T. Hotzand 和 A. Munk。内在形状分析：对模去等距李群作用的黎曼流形进行测地线主成分分析。统计辛，20：1–58，2010。H. Janati、M. Cuturi 和 A. Gramfort。去偏 Sinkhorn 重心。在国际机器学习会议，第 4692–4701 页。PMLR，2020。Y.-H. Kim 和 B. Pass。黎曼流形上的 Wasserstein 重心。数学进展，307：640–683，2017。ISSN 0001-8708。doi：10.1016/j.aim.2016.11.026。URL https://doi.org/10.1016/j.aim.2016.11.026。J. M. Lee。光滑流形导论，数学研究生教材第 218 卷。Springer，纽约，第二版，2013。ISBN 978-1-4419-9981-8。R. J. McCann。黎曼流形上映射的极分解。几何与泛函分析，11(3)：589–608，2001。J. R. Munkres。拓扑学。Prentice Hall, Inc.，新泽西州上萨德尔河，2000。ISBN 0-13-181629-2。 [MR0464128]的第二版。V. M. Panaretos 和 Y. Zemel。*Wasserstein* 空间统计学导论。Springer Nature，2020。V. Patrangenaru 和 L. Ellingson。流形上的非参数统计及其在对象数据分析中的应用。CRC Press，2015。M. Pegoraro 和 M. Beraha。基于 Wasserstein 度量的实数线上分布型数据投影统计方法。机器学习研究杂志，23(37)：1–59，2022。X. Pennec。黎曼流形上的内在统计学：几何测量的基本工具。数学成像与视觉杂志，25：127–154，2006 年 7 月。X. Pennec。流形上的统计计算：从黎曼几何到计算解剖学。载于*LIX* 秋季新兴视觉计算趋势研讨会，第 347–386 页。Springer，2008。L. M. Sangalli、P. Secchi、S. Vantini 和 A. Veneziani。探索性函数数据分析案例研究：颈内动脉的几何特征。美国统计学会杂志，104(485)：37–48，2009。S. Srivastava、V. Cevher、Q. Dinh 和 D. Dunson。WASP：通过子集后验分布的重心实现可扩展贝叶斯方法。载于 G. Lebanon 和 S. V. N. Vishwanathan（编），第十八届国际人工智能与统计会议论文集，机器学习研究论文集第 38 卷，机器学习研究论文集，第 912–920 页，美国加利福尼亚州圣迭戈，2015 年 5 月 9–12 日。PMLR。

C. Villani。《最优传输专题》。2003年。URL https://bookstore.ams.org/gsm-58。Y. Zemel 和 V. M. Panaretos。Wasserstein空间中的Fréchet 均值与普罗克鲁斯特分析。《伯努利》，25(2)：932–976，2019年。C. Zhang、P. Kokoszka 和 A. Petersen。用于密度时间序列的瓦瑟斯坦自回归模型。*arXiv* 预印本 *arXiv:2006.12640*，2020。

A 技术预备知识

A.1 测度论预备知识

设 $(M,\ g)$ 是一个维数为 n 的黎曼流形，其中 TM 是其切丛，TM^* 是其余切丛。根据定义，g 是一个截面$g\colon\ M \to (TM \otimes TM)^*$，体积形式 $\omega\colon\ M \to \wedge^n(TM)^*$ 在局部定义为 $\omega = |\det(g)|^{1/2}\, dx_1 \wedge \ldots \wedge dx_n$。

设 $\mathcal{L}$ 为实数集 R^n 上的勒贝格测度。我们考虑由所有满足 $\varphi(A \cup U)$ 属于某个坐标图 $(U,\ \varphi)$ 所对应的实数集 R^n 上的勒贝格 σ代数的集合 A 生成的 σ代数。接着，我们用 $\mathcal{L}_M$ 表示黎曼-勒贝格体积测度，即定义在 M 上的测度，使得对于每个坐标图 $(U,\ \varphi)$ 以及包含于刚定义的 σ代数中的 $A \subset U$：

$$\mathcal{L}_M(A) = \int_{\varphi(A)} |\det(g(\varphi^{-1}))|^{1/2}\, d\mathcal{L} \tag{18}$$

注意，一般而言，$\varphi\#\mathcal{L}_M \neq \mathcal{L}$。

考虑 $h\colon\ M \to$实数集，使得 $\mathrm{supp}(h) \subset U$，其中 (U, φ) 是一个坐标图，我们可以按如下方式对 h 积分：

$$\int_M h d\mu = \int_U h d\mu = \int_U h f_\mu d\mathcal{L}_M = \int_{z(U)} |\det(g(\varphi^{-1}))|^{1/2}\, h(\varphi^{-1}) f_\mu(z^{-1}) d\mathcal{L}. \tag{19}$$

一般情形通过单位分解以一种自然的方式定义。

现在我们可以考虑一个测度 μ，关于 M，其密度函数 f_μ 相对于 $\mathcal{L}_M$，即：

$$\mu(A) = \int_A f_\mu d(\mathcal{L}_m) = \int_{\varphi(A)} |\det(g(\varphi^{-1}))|^{1/2}\, f_\mu(\varphi^{-1}) d\mathcal{L}. \tag{20}$$

最后，如果 μ 相对于 $\mathcal{L}_M$ 没有密度函数，为了对某个函数关于 μ 积分，我们选取一个弱收敛序列 $\mu_n \rightharpoonup \mu$，使得对于每个 n，μ_n 都有密度函数，并通过取积分的极限来扩展定义。

A.2 麦肯的结果

让我们回顾一下 c-凹性的定义。设 $c\colon\ M \times M \to$实数集$\cup +\infty$。对于函数$\psi\colon\ M \to$实数集$\cup \{-\infty\}$，将其 c-变换 $\psi^{c+}\colon\ M \to$实数集$\cup \{-\infty\}$ 定义为

$$\psi^{c+}(x) = \inf_{y \in M} c(x, y) - \psi(y).$$

注意，这推广了勒让德变换；当 $M=$ 实数集^{d}且 $c(x,\ y)=\langle x,y\rangle$ 时，可得到勒让德变换。

定义 2. 若函数 $\phi:\ M\to$ 实数集 $\cup\{-\infty\}$ 不恒等于$-\infty$，且存在 $\psi:\ M\to$ 实数集 $\cup\{-\infty\}$ 使得

$$\phi=\psi^{c+}$$

给定 $\mu\in\mathcal{W}_2(M)$ 和开集 $U\subset M$，定义 $S(U)=\{v\colon\ U\to TM\mid\pi\circ v=\mathrm{Id}_U\}$为 M的切丛的局部截面的层，也就是 U上的切向量场构成的向量空间。每当 U 是切丛的一个局部平凡化时，对于 $v\in S(U)$，我们可以使用记号 $v_z:=v(z)\in T_zM$。现在可以定义以下函数层:

$$L^2_\mu(U)=\{v\in S(U)\mid\int\|v_z\|^2d\mu(z)<\infty\},\tag{21}$$

其中 $\|v_z\|^2$ 表示 $g(v_z,\ v_z)$。

对于任意 $v\in L^2_\mu(U)$，我们可以考虑映射 $\exp(v)$，其定义为 $\exp(v)(z):=\exp_z(v_z),\quad z\in U$。McCann [2001] 证明，若 μ 相对于 M上的体积测度绝对连续，则 μ 与 ν之间唯一的最优输运计划 γ^o 由一个映射诱导，即存在 T: $M\to M$，诱导出 $(\mathrm{Id}\ T):\ M\to M\times M$，使得 $\gamma^o=(\mathrm{Id}\ T)\#\nu$。此外，映射 T 具有形式 $T=\exp(-\nabla\phi)$，其中ϕ 是一个 d^2-凹函数 [Gigli, 2011]。

A.3 关于 $\mathbb{S}_1$ 的更多细节

我们将 $\exp_c:(\mathbb{R},+)\to(\mathbb{S}_1,\cdot)$ 称为由 $\exp_c(x)=e^{ix}$ 定义的映射，并将 $\mathbb{S}_1\subset\mathbb{C}$ 视为配备乘法运算 $e^{ix}\cdot e^{iy}=e^{i(x+y)}$ 的对象。因此，$\exp_c$ 是一个群同态：$\exp_c(x+y)=\exp_c(x)\cdot\exp_c(y)$。类似地，$log_c:\ \mathbb{S}_1\to\mathbb{R}$ 是由 $\log_c(z)=x\in[0,2\pi)$ 给出的映射，满足 $z=e^{ix}$。显然，log_c 是 $\exp_c$ 的右逆，即 $\exp_c\circ\log_c=\mathrm{Id}_{\mathbb{S}_1}$。

类似地，我们使用商空间上的投影 $\pi_c:\mathbb{R}\to\mathbb{R}/2\pi\mathbb{Z}$，它是一个群映射。这是对 $\mathbb{S}_1$ 的另一种表示，虽然等价，具体来说：我们有 $\exp_c(x)=\exp_c(\pi_c(x))$，因此 $\log_c=(\exp_c\circ\pi_c)^{-1}$ 是群同构。证明中我们采用由圆的这种替代表示诱导的度量，见公式（24）。

最后，对于任意 $z\in\mathbb{S}_1$，我们定义映射 $\exp_c$ 和 $\log_c$ 以 z 为中心的平移版本：$\exp_z(x)=\exp_c(x+\log_c(z))$，以及 $\log_z(z')=x'$ with $x'\in[-\pi/2,\pi/2)$，使得 $\exp_z(x')=z'$。若记 $W:=(-\pi/2,\ \pi/2)$ 和 $V_z:=\mathbb{S}_1-\{-z\}$，则对每个 $z\in\mathbb{S}_1$，有序对 $(V_z,\ \log_z)$ 是一个局部坐标图，它在 W 上给出一个同胚。在这一微分结构下，$\mathbb{S}_1$ 是一个李群，其切丛为 $T\mathbb{S}_1=\{(x,\ v)\mid x\in\mathbb{S}_1\text{ 和 }v\in T_x\mathbb{S}_1\}\simeq\mathbb{S}_1\times$ 实数集。我们将点 $(1,\ 0)$ 记作 1，它是 $\mathbb{S}_1$ 中的单位元。

我们可以作如下观察：黎曼度量 $g\colon\ \mathbb{S}_1\to(T\mathbb{S}_1\otimes T\mathbb{S}_1)^*$ 由嵌入 $\mathbb{S}_1\hookrightarrow\mathbb{C}\simeq\mathbb{R}^2$诱导。在以 $1\in\mathbb{S}_1$ 为中心的局部坐标 (V_1,log_1) 中，嵌入为 $x\in(-\pi/2,\pi/2)\ \mapsto e^{ix}=(\cos(x),\sin(x))$。因此，切空间之间的映射为 $d/dx\mapsto-\sin(x)d/dx_1+\cos(x)d/dx_2$。令$g_{\mathbb{E}}$ 为实数集 $\mathbb{R}^2$上的欧几里得度量。在 $\mathbb{R}^2$的典范坐标 x_1,x_2中，该度量

显然由二次型 $\mathrm{Id}_{2\times2}$给出。因此，$g(d/dx) = g_E(-\sin(x), \cos(x)) = \sin^2(x) + \cos^2(x) = 1$。换言之，在局部坐标 $(V_1, \log_1)$ 中，对于 $v,\ w \in T_1\mathbb{S}_1 \simeq \mathbb{R},\ g(v,w) = v \cdot w$。对于此度量，$\exp_z : T_z\mathbb{S}_1 \to \mathbb{S}_1$ 和 $\log_z : \mathbb{S}_1 \to T_z\mathbb{S}_1$ 分别是黎曼指数映射和黎曼对数映射。

A.4 $\mathbb{S}_1$ 上的向量场

现在考虑 $\mu \in \mathcal{P}(\mathbb{S}_1)$ 以及一个实值函数 $f : \mathbb{S}_1 \to$ 实数集，我们希望将其对 μ 进行积分。我们看到 $det(g) \equiv 1$ 且体积形式 ω 局部为 $\omega = dx$。这立即意味着 $\mathcal{L}_{\mathbb{S}_1} = \exp_c\#\mathcal{L}$，或者等价地，$\log_c\#\mathcal{L}_{\mathbb{S}_1} = \mathcal{L}$。因此：

$$\int_{\mathbb{S}_1} f(z)d\mathcal{L}_{\mathbb{S}_1}(z) = \int_{[-\pi/2,\pi/2)} f(\exp_c(x))d\mathcal{L}(x) \tag{22}$$

沿着这条线，切向量场层 L^2_μ 可以通过变量替换轻松传输到 $[0, 2\pi$ 上：

$$\int_{\mathbb{S}_1} |\ v_z\ |^2\ d\mu(z) = \int_{[0,2\pi)} |\ v(\exp_c(x))\ |^2\ d\log_c\#\mu(x) \tag{23}$$

实际上，从方程 (20) 以及观察 $\mathcal{L}_{\mathbb{S}_1} = \exp_c\#\mathcal{L}$，我们得到 $\mathbb{S}_1$ 上的测度 μ 等价地由测度 $\log_c\#\mu$ 表示，该测度扩展为实数集上的周期测度 $\widetilde{\mu}$，如下所示：对于任何可测的 A 和 $p \in \mathbb{Z}$，$\mu(\exp_c(A)) =: \widetilde{\mu}(A) = \widetilde{\mu}(A + p)$

B 证明

我们引入一些记号，其灵感来自附录 A.3。

$$d_{\mathbb{Z}}(x,y)^2 := \inf_{p\in 2\pi\mathbb{Z}}(x - y - p)^2 \le (x - y)^2 \tag{24}$$

注意，$d_R(z, z') = d_{\mathbb{Z}}(\log_c(z), \log_c(z'))$

B.1 定理 2 的证明

证明。 证明可由 Delon 等人[2010]提出的局部最优输运计划概念直接得出。设 γ_θ 为传输计划，它将一份质量从位置$F^-_{\widetilde{\nu}}(u)$传输到位置 $(F^\theta_{\widetilde{\mu}})^-(u)$。那么 γ_θ是局部最优的，且相关代价为

$$C_{[\mu,\nu]}(\theta) = \int_0^1 \left(F^-_{\widetilde{\mu}}(u) - (F^\theta_{\widetilde{\nu}})^-(u)\right)^2 \mathrm{d}u$$

（全局）最优计划与 $\theta^* = \operatorname{argmin}\ C(\theta) = W_2^2(\mu,\nu)$ 相关联。为了恢复最优传输映射，我们进行变量替换 $x = F^-_{\widetilde{\mu}}(u)$，从而得到：

$$W_2^2(\mu,\nu) = \int_0^{2\pi} \left(T_{\widetilde{\mu}}^{\widetilde{\nu}}(x) - x\right)^2 \mathrm{d}\widetilde{\mu}(x) \geq \tag{25}$$

$$\int_0^{2\pi} d_{\mathbb{Z}}^2(T_{\widetilde{\mu}}^{\widetilde{\nu}}(x), x)\mathrm{d}\widetilde{\mu}(x) = \tag{26}$$

$$\int_{\mathbb{S}_1} d_R^2(\exp_c(T_{\widetilde{\mu}}^{\widetilde{\nu}}(\log_c(z))), z)^2 \mathrm{d}\mu(z) \geq W_2^2(\mu,\nu) \tag{27}$$

其中第一个等式通过定义 $T_{\widetilde{\mu}}^{\widetilde{\nu}} := (F_{\widetilde{\nu}}^{\theta^*})^- \circ F_{\widetilde{\mu}}$ 得出，而最后一个等式通过 $z = \exp_c(x)$ 以及 $d_{\mathbb{Z}}$ 的性质得到。 □

B.2 定理 3 的证明

证明。 首先我们观察到：

$$(F_{\widetilde{\nu}}^{\theta^*})^-(F_{\widetilde{\mu}}(u+2\pi p)) = (F_{\widetilde{\nu}}^{\theta^*})^-(F_{\widetilde{\mu}}(u)+p) = (F_{\widetilde{\nu}})^-(F_{\widetilde{\mu}}(u)+p-\theta^*) = (F_{\widetilde{\nu}}^{\theta^*})^-(F_{\widetilde{\mu}}(u))+2\pi p$$

这意味着 $T(u+2\pi p) = T(u) + 2\pi p$ 对于每一个 $p \in \mathbb{Z}$.

根据定理 2，我们知道当且仅当 $\theta = \theta^*$ 如公式（8）所示。定义：

$$C_{[\mu,\nu]}(\theta) = \int_0^1 \left(F_{\widetilde{\mu}}^-(u) - (F_{\widetilde{\nu}}^{\theta})^-(u)\right)^2 \mathrm{d}u$$

Delon 等人 [2010] 证明，当 μ 绝对连续时，映射 $\theta \mapsto C_{[\mu,\nu]}(\theta)$ 是严格凸的。因此，θ^* 是该函数的唯一驻点。基于此，我们计算 $C_{[\mu,\nu]}$ 关于 θ 的导数，并注意到 θ^* 是满足 $(C_{[\mu,\nu]})' = 0$ 的唯一值。根据莱布尼茨法则，我们可以写出：

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\theta} C_{[\mu,\nu]}(\theta) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\theta}\left(F_{\widetilde{\mu}}^-(u) - (F_{\widetilde{\nu}})^-(u-\theta)\right)^2 \mathrm{d}u \\
&= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\theta}\left(F_{\widetilde{\mu}}^-(u+\theta) - (F_{\widetilde{\nu}})^-(u)\right)^2 \mathrm{d}u \\
&= \int_0^1 2\left(F_{\widetilde{\mu}}^-(u+\theta) - (F_{\widetilde{\nu}})^-(u)\right) \frac{1}{f_{\widetilde{\mu}}(F_{\widetilde{\mu}}^-(u+\theta))} \mathrm{d}u
\end{aligned}$$

采用变量替换 $v = F_{\widetilde{\mu}}^-(u+\theta)$，这意味着 $F_{\widetilde{\mu}}(v) - \theta = u$ 并且 $\mathrm{d}u = \mathrm{d}\widetilde{\mu}(v) = f_{\widetilde{\mu}}(v)\mathrm{d}v$，我们得到

$$= \int_{v_0}^{2\pi+v_0} 2\left(v - \widetilde{T}(v)\right) \frac{f_{\widetilde{\mu}}(v)}{f_{\widetilde{\mu}}(v)} \mathrm{d}v = -2\int_{v_0}^{2\pi} \left(\widetilde{T}(v) - v\right)\mathrm{d}v - 2\int_{2\pi}^{2\pi+v_0} \left(\widetilde{T}(v) - v\right)\mathrm{d}v$$

其中 $v_0 = F_{\widetilde{\mu}}^-(\theta)$。

通过变量替换 $u = v - 2\pi$ 我们得到：

$$\frac{\mathrm{d}}{\mathrm{d}\theta} C_{[\mu,\nu]}(\theta) = -2\int_0^{2\pi} \left(\widetilde{T}(u) - 2\pi - (u - 2\pi)\right)\mathrm{d}v = -2\int_0^{2\pi} \left(\widetilde{T}(u) - u\right)\mathrm{d}u.$$

当且仅当该量等于零时，最优性成立，因此：

$$0 = \int_0^{2\pi} \left(\widetilde{T}(u) - u\right)\mathrm{d}u.$$

□

B.3 定理4的证明

证明。 注意到 T 是一个最优传输映射，然后定义

$$\widetilde{T}(x) := \log_c \circ T \circ \exp_c,\ x \in (0, 2\pi), \qquad \widetilde{T}(x+p) = \widetilde{T}(x) + p,\ p \in 2\pi\mathbb{Z}$$

显然满足单调性和“周期性”要求。此外，根据定理 3，(11) 得以满足。为证明 $|T(x) - x| < \pi/2$，注意到这等价于 $\inf_{p\in 2\pi\mathbb{Z}} |T(x) - x - p| = |T(x) - x|$。由于 T： $\mathbb{S}_1 \to \mathbb{S}_1$ 是从 μ 到 ν 的最优传输映射，因此我们有：

$$\begin{aligned} W_2(\mu,\nu) &= \int_{\mathbb{S}_1} d_R(T(z), z)^2 d\mu \\ &= \int_{[0,2\pi]} d_R(T(\exp_c(x)), \exp_c(x))^2 d(\log_c \#\mu)(x) \\ &= \int_{[0,2\pi]} d_{\mathbb{Z}}(\widetilde{T}(x), x)^2 d\widetilde{\mu}(x) \end{aligned}$$

其中，第一个等号由最优传输映射的定义得到，第二个等号通过变量替换 $z = \exp_{c|[0,2\pi]}(x)$ 得到，而在最后一个等号中我们使用了 T、 $\widetilde{\mu}$ 的定义以及 $d_{\mathbb{Z}}$ 的性质。如前所述，我们有 $\inf_{p\in\mathbb{Z}} |T(x) - x - p| \le |T(x) - x|$。如果对某个 $A \subset [0, 2\pi]$，满足 $\widetilde{\mu}(A) > 0$ 时严格不等式成立，那么 $[0, 2\pi]$ 上的积分也必然不同，由此结论得证。

要证明逆命题，只需证明 $F_{\widetilde{\nu}}^{-}(v) = \widetilde{T}(F_{\widetilde{\mu}}^{-}(v) + \theta)$，这等价于说

$$F_{\widetilde{\nu}}^{-}(v) = \widetilde{T}\left((F_{\widetilde{\mu}}(\cdot) + \theta)^{-1}(v)\right) = \widetilde{T}(F_{\widetilde{\mu}}^{-}(v - \theta)).$$

定义 $G_{\widetilde{\nu}} := T \circ F_{\widetilde{\mu}}^{-}$，则显然 $T = G_{\widetilde{\nu}} \circ F_{\widetilde{\mu}}$。我们证明 $G_{\widetilde{\nu}}(u) \equiv F_{\widetilde{\nu}}^{-}(u + \theta)$。对于 $x \in [0, 2\pi)$，有

$$F_{\widetilde{\nu}}(x) = \widetilde{\nu}([0, x]) = \widetilde{\mu}(\widetilde{T}^{-1}([0, x])) = \widetilde{\mu}([\widetilde{T}^{-1}(0), \widetilde{T}^{-1}(x)]) = F_{\widetilde{\mu}}(\widetilde{T}^{-1}(x)) - F_{\widetilde{\mu}}(\widetilde{T}^{-1}(0))$$

并注意到，由于 $|T(x) - x| < \pi/2$，有 $F_{\widetilde{\nu}}(x) \le 1$。因此，$\widetilde{\mu}$ 在 $\mathbb{S}_1$ 上的推前测度是一个有效的概率测度。取 $F_{\widetilde{\nu}}$ 的逆

$$F_{\widetilde{\nu}}^{-}(u) = \left(F_{\widetilde{\mu}}(\widetilde{T}^{-1}(\cdot)) - F_{\widetilde{\mu}}(\widetilde{T}^{-1}(0))\right)^{-}(u) = \widetilde{T} \circ F_{\widetilde{\mu}}^{-}(u + F_{\widetilde{\mu}}(\widetilde{T}^{-1}(0)))$$

并令 $-\theta = F_{\widetilde{\mu}}(T^{-1}(0))$，即可得到所需结果。 □

B.4 定理5的证明

为了证明（ii）项，我们需要以下两个预备引理。

引理 1. 假设我们有 $W_2(\nu, \nu_n) \to 0$ 属于 $\mathcal{W}_2(\mathbb{S}_1)$，其中 ν、 ν_n 关于 $\mathcal{L}_{\mathbb{S}_1}$（对每个 n）绝对连续。则 $W_2(\log_c \#\nu,\ \log_c \#\nu_n) \to 0$ 属于 $\mathcal{W}_2$(实数集)。

证明。 根据 Villani 中的定理 7.12 [2003], , Wasserstein度量下的收敛等价于弱收敛加上紧性条件：存在 x_0 使得

$$\lim_{R\to+\infty}\limsup_{k\to+\infty}\int_{d(x,x_0)>R} d(x,x_0)^p d\log_c \#\nu_k(x)$$

注意，每个测度 $\log_c \#\nu_k$ 都支撑在 $[0,2\pi]$ 上，因此始终满足该条件。因此，我们只需证明序列 $\log_c\#\nu_k$ 弱收敛。对于实数线上的测度，弱收敛等价于相应分布函数在连续点处逐点收敛。也就是说，令 $F_k(x) := \log_c \#\nu_k([0,x))$和 $F(x) := \log_c \#\nu([0,x))$，则必须有

$$F_k(x) \to F(x), \qquad \text{all } x \text{ such that } F(x) \text{ is continuous} \tag{28}$$

根据定义， $F_k(x) = \nu_k(\exp_c([0,x)))$。根据Portmanteau定理，对于任意满足 $\nu(\{\exp_c(x)\}) = 0$ 的 x ，有 $\nu_k(\exp_c([0,x))) \to \nu_k(\exp_c([0,x)))$，由此很容易推出（28）

□

引理 2. 设有 $W_2(\nu,\nu_n) \to 0$，其中 μ、 ν、 ν_n 相对于 $\mathcal{L}_{\mathbb{S}_1}$ （对每个 n）均绝对连续。则 $\| \log_\mu(\nu_n) - \log_\mu(\nu) \|_{L^2_\mu} \to 0$。

证明。 根据引理 1，我们有 $F_{\widetilde{\nu}_n}(x) \to F_{\widetilde{\nu}}(x)$，分位数函数也同样如此。因此， $C_{[\nu_n,\mu]}(\theta) \to C_{[\nu,\mu]}(\theta)$。

因此，考虑 $\theta_n = \arg\min C_{[\nu_n,\mu]}$。根据第3.2节的讨论，尤其是方程 (10)，我们有 $C_{[\nu_n,\mu]}$ 的最小化定义域可以限制在一个足够大的紧区间 K 内。由于 $\{\theta_n\} \subset K$ 紧，我们可以考虑一个收敛子序列，出于记号上的方便，仍将其记为 $\{\theta_n\}$ 。令 $\theta_n \to \theta^*$。回顾一下， $C_{[\nu_n,\mu]}$ [Delon 等人， 2010] 是严格凸的。因此，根据标准论证，我们得出 $\theta^* = \arg\min C_{[\nu,\mu]}$。

现在考虑:

$$\left| \left(F_{\widetilde{\mu}}(F^-_{\widetilde{\nu}_n}(u+\theta_n))\right) - \left(F_{\widetilde{\mu}}(F^-_{\widetilde{\nu}}(u+\theta^*))\right) \right| \leq \tag{29}$$

$$\left| \left(F_{\widetilde{\mu}}(F^-_{\widetilde{\nu}_n}(u+\theta_n))\right) - \left(F_{\widetilde{\mu}}(F^-_{\widetilde{\nu}}(u+\theta_n))\right) \right| + \left| \left(F_{\widetilde{\mu}}(F^-_{\widetilde{\nu}}(u+\theta_n))\right) - \left(F_{\widetilde{\mu}}(F^-_{\widetilde{\nu}}(u+\theta^*))\right) \right| \tag{30}$$

~ ~这意味着逐点收敛 $T^{\mu}_{\widetilde{\nu}_n}(u) \to T^{\mu}_{\widetilde{\nu}}(u)$：最后一个求和中的两项都趋于 0。由于这些映射在 $[0,2\pi]$ 上连续且有界，我们得到一致收敛和强收敛。对数映射的像μ 中的强收敛由此得证。

□

现在我们可以证明定理 5 了。

证明。1. 为检验 $\exp^\mu$ 的连续性，考虑 $T^{\nu_1}_\mu \times T^{\nu_2}_\mu$: $\mathbb{S}_1 \to \mathbb{S}_1 \times \mathbb{S}_1$ 和

诱导出传输计划 $\gamma = (T_\mu^{\nu_1},\ T_\mu^{\nu_2})\#\mu$。于是有：

$$\begin{aligned} W_2^2(\nu_1,\nu_2) &\le \int_{\mathbb{S}_1\times\mathbb{S}_1} d_R(z,w)^2 d\gamma(dzdw) \\ &= \int_{\mathbb{S}_1} d_R(T_\mu^{\nu_1}(z), T_\mu^{\nu_2}(z))^2 d\mu(dz) \\ &\le \int_{[0,2\pi]} d_{\mathbb{Z}}(T_{\widetilde{\mu}}^{\widetilde{\nu}_1}(x), T_{\widetilde{\mu}}^{\widetilde{\nu}_2}(x))^2 d\widetilde{\mu}(dx) \\ &\le \| T_{\widetilde{\mu}}^{\widetilde{\nu}_1} - T_{\widetilde{\mu}}^{\widetilde{\nu}_2} \|^2_{L^2_{\widetilde{\mu}}([0,2\pi])} = \| \log_\mu(\nu_1) - \log_\mu(\nu_2) \|^2_{L^2_\mu}, \end{aligned}$$

其中最后一个恒等式是由于 $\log_c \#\mu = \widetilde{\mu}$ 在 $[0,2\pi]$ 上得到的。

2. 为了检验 $\log_\mu$ 的连续性，通过逼近论证，我们得到 $\log_\mu$ 在任意测度 $\nu \in \mathcal{W}_2(\mathbb{S}_1)$ 处的序列连续性：考虑$\nu_n \to \nu$，其中 ν_n 为绝对连续测度。则 $\{\log_\mu(\nu_n)\}$ 是 L^2_μ 中的柯西序列，而 L^2_μ 是完备度量空间，因此它收敛到一个向量场 v。考虑 $\exp_\mu(v)$。由 $\exp_\mu$ 的连续性，我们有 $\nu_n \to \exp_\mu(v)$，这进而蕴含 $\exp_\mu(v) = \nu$。

最后，度量空间中的序列连续性蕴含连续性。

□

B.5 定理6的证明

Ambrosio等人中的定理3.2 [2019] 确保，在命题的假设下，$\int_{\mathbb{S}_1} d_R(T_\mu^{\nu_n}(x)\ T_\mu^{\nu}(x))^2 d\mu(x) \to 0$。

现在我们证明以下引理，以完成证明。

引理 3. 假设 μ、ν、ν_n 关于 $\mathcal{L}_{\mathbb{S}_1}$（对每个 n）几乎处处成立。并且假设以下条件成立。

$$\int_{\mathbb{S}_1} d_R(T_\mu^{\nu_n}(x), T_\mu^{\nu}(x))^2 d\mu(x) \to 0.$$

那么 $\| \log_\mu(\nu_n) - \log_\mu(\nu) \|_{L^2_\mu} \to 0$。

证明。为简化记号，记 $G := \log_\mu(\nu)$，并记 $G_n := G := \log_\mu(\nu_n)$。于是$G(x) = \log_x(T_\mu^{\nu}(x))$，且 $G_n(x) = \log_x(T_\mu^{\nu_n}(x))$。此外，定义 f_n：$\mathbb{S}_1 \to$实数集，使得 $f_n(x) = G(x) - G_n(x)$。注意，f_n 有界。

我们可以写成：

$$\int_{\mathbb{S}_1} d_R(T_\mu^{\nu_n}(x), T_\mu^{\nu}(x))^2 d\mu(x) = \int_{A_n^+} |\ f_n(x)\ |^2\ d\mu(x) + \int_{A_n^-} |\ f_n(x) + 2\pi\ |^2\ d\mu(x).$$

其中 $A_n^+ = f_n^{-1}([-\pi,\pi])$ 且 $A_n^- = f_n^{-1}([\pi,2\pi]) \cup f_n^{-1}([-2\pi,-\pi])$。

我们希望证明：

$$\int_{\mathbb{S}_1} |\ f_n(x)\ |^2\ d\mu(x) = \int_{A_n^+} |\ f_n(x)\ |^2\ d\mu(x) + \int_{A_n^-} |\ f_n(x)\ |^2\ d\mu(x) \to 0,$$

因此，我们需要处理积分 $\int_{A_n^-} \mid f_n(x) \mid^2 d\mu(x)$，正如我们已经知道的 $\int_{A_n^+} \mid f_n(x) \mid^2 d\mu(x) \to 0$。

我们要证明 $\mu(A_n^-) \to 0$。

用反证法，假设存在 $\varepsilon > 0$，使得对每个$N > 0$，都存在 $n > N$ 满足 $\mu(A_n^-) > \varepsilon$。为简化记号，我们不取满足 $\mu(A_{n_k}^-) > \varepsilon$ 的子序列 $\{n_k\}$，而直接假设该结论对每个 n 都成立。

考虑 $c > 0, c \in$ 实数集。集合 $B_{n,c}^+ := f_n^{-1}([c, +\infty)) \cap A_n^-$ 必须满足 $\mu(B_{n,c}^+) \to 0$。事实上：

$$\int_{A_n^-} \mid f_n(x) + 2\pi \mid^2 d\mu(x) = \int_{B_{n,c}^+} \mid f_n(x) + 2\pi \mid^2 d\mu(x) + \int_{B_{n,c}^-} \mid f_n(x) + 2\pi \mid^2 d\mu(x),$$

和 $\int_{B_{n,c}^+} \mid f_n(x) + 2\pi \mid^2 d\mu(x) \geq c \cdot \mu(B_{n,c}^+)$。

若 $x \in B_{n,c}^+$，则以下两种情况必有其一：1) $G(x) \in [-\pi, -\pi + c]$ 且 $G_n(x) \in [\pi - c, \pi]$；或 2)$G_n(x) \in [-\pi, -\pi + c]$ 且 $G(x) \in [\pi - c, \pi]$。1) 或 2) 至少有一种必须无限多次成立。不失一般性，假设 1) 成立可数多次。同样，我们不取使 1) 始终成立的子序列 $\{n_k\}$，而直接假设对每个 n，1) 都成立。

换言之，对于任意 $c > 0, c \in \mathbb{R}$，我们有：对于任意 n，对于每个 $x \in B_{n,c}^+$, $G(x) \in [-\pi, -\pi + c]$。并且 $B_{n,c}^+$ 始终是一个具有正测度的集合。由 $\mathbb{S}_1$ 的紧致性，我们可以找到一个具有正测度的集合 B，使得对于所有 $x \in B$，$G(x) = -\pi$。因此，T_μ^ν 是一个最优传输映射，它将一个满测度集中的每个点映射到圆上的对径点。这与循环单调性条件相矛盾。因此，$\mu(A_n^-) \to 0$。

至此，证明完成。

$$\begin{aligned}\int_{\mathbb{S}_1} \mid f_n(x) \mid^2 d\mu(x) &= \int_{A_n^+} \mid f_n(x) \mid^2 d\mu(x) + \int_{A_n^-} \mid f_n(x) \mid^2 d\mu(x) \\ &\leq \int_{A_n^+} \mid f_n(x) \mid^2 d\mu(x) + 4\pi^2 \mu(A_n^-) \to 0.\end{aligned}$$

□

B.6 定理7的证明

我们的证明依赖于 Agueh 和 Carlier [2011]，Panaretos 和 Zemel [2020]提出的 Wasserstein重心多边缘表述。下面，考虑 $\mu_1, \ldots, \mu_n$ 属于 $\mathcal{W}(M)$，其中 (M, d) 是一个连通紧致流形。设 $\Pi(\mu_1, \ldots, \mu_n)$ 为定义在 M^n 上的概率测度集合：$= M \times \cdots \times M$ 具有边缘测度 $\mu_1, \ldots, \mu_n$。多边缘问题的目标是最小化

$$G(\pi) = \frac{1}{2n^2} \int_{M^n} \sum_{i<j} d(x_i, x_j)^2 \mathrm{d}\pi(x_1, \ldots, x), \quad \pi \in \Pi(\mu_1, \ldots, \mu_n).$$

如 Kim 和 Pass [2017]（定理 2.4）所示，最小化 $G(\pi)$ 等价于在式 (16) 中最小化 $F(\nu)$。事实上，令 $\bar{x}: M^n \to M$ 使得

$$x_1, \ldots, x_n \mapsto \operatorname*{argmin}_{z \in M} \sum_{i=1}^n d^2(x_i, z)$$

则最优多重耦合 $\pi^o = \operatorname{argmin}_{\pi\in\Pi} G(\pi)$ 根据以下规则给出 Fréchet 泛函的极小值点

$$\bar{\mu} := \bar{x}\#\pi^o. \tag{31}$$

引理 4. 设(M, d) 为一个连通紧致黎曼流形，其指数映射$\exp_M$ 非扩张。记$\log_M$ 为相应的对数映射。设 μ^*是 $\mathcal{W}_2(M)$上的一个绝对连续测度，且 $\mu_1, \ldots, \mu_n \in \mathcal{W}_2(M)$。假设对于任意 $i, j = 1, \ldots, n$,

$$\|\log_{\mu^*}(\mu_i) - \log_{\mu^*}(\mu_j)\|^2_{L^2_{\mu^*}} = W_2^2(\mu_i, \mu_j),$$

其中 $T^{\mu_i}_{\mu^*} = \exp_M \circ \log_{\mu^*}(\mu_i)$。令 $\bar{T} = n^{-1}\sum_{i=1}^n \log_{\mu^*}(\mu_i)$，则 $\mu_1, \ldots, \mu_n$ 的 *Wasserstein* 重心为 $T\#\mu^*$。

证明。 考虑多重耦合 $\pi^* = (T^{\mu_1}_{\mu^*}, \ldots, T^{\mu_n}_{\mu^*})\#\mu^*$。显然，根据构造， $\pi^* \in \Pi(\mu_1, \ldots, \mu_n)$, 且 $G(\pi^*) \geq G(\pi^o)$。我们注意到，由于指数映射非扩张,

$$\begin{aligned}\int_{M^n} d^2(x_i, x_j)\mathrm{d}\pi^*(x_1, \ldots, x_n) &= \int_M d^2(T^{\mu_i}_{\mu^*}(x), T^{\mu_j}_{\mu^*}(x))\mathrm{d}\mu^*(x)\\ &\leq \|\log_{\mu^*}(\mu_i) - \log_{\mu^*}(\mu_j)\|^2_{L^2_{\mu^*}}\\ &= W_2^2(\mu_i, \mu_j),\end{aligned}$$

其中，第一个不等式由定理 5 的第一点推出，最后一个等式则由假设得出。因此， π^* 使 $G(\pi)$达到最小值，结论由（31）得出。

□

定理7的证明由引理 4以及 $\mathbb{S}_1$ 的指数映射确实具有非扩张性这一事实得出，见（24）。

C 附加模拟

这里报告一项关于主成分分析的补充模拟研究，旨在展示基点 $\bar{\mu}$的影响。我们考虑定义在 $[0, 2\pi)$上（并在实数集上周期性延拓）的截断高斯测度，其参数为均值参数 m 和尺度参数（方差的平方根） s。我们通过对 $m \sim \mathcal{U}(0.3\pi,\ 1.7\pi)$和 $s \mid m \sim \mathcal{U}(\pi/20,\ \pi/10)I_{(0,2\pi)}(m \pm 3s)$进行抽样，模拟 $n = 100$ 个数据点。图 10（上排）展示了数据以及 $\mathcal{W}_2$(实数集)和 $\mathcal{W}_2(\mathbb{S}_1)$中的重心。特别地， $\mathcal{W}_2$(实数集)中的重心是单峰的，且以定义域中心为中心。 $\mathcal{W}_2(\mathbb{S}_1)$中的重心最高的众数也位于定义域中心，但可以看出它呈三峰分布，尤其是在原点附近有一个众数。正如第 5 节所指出的，这是因为某些测度之间的测地线会迫使质量沿圆周穿过 0。

$\mathcal{W}_2$(实数集)中的前两个主方向如图 10（底排）所示，正如预期，它们清楚地区分了位置和尺度的影响。在 $\mathcal{W}_2(\mathbb{S}_1)$ 中进行主成分分析时，尺度和位置的影响并未清楚地区分，如图 11（顶排）所示：重心中位于次要

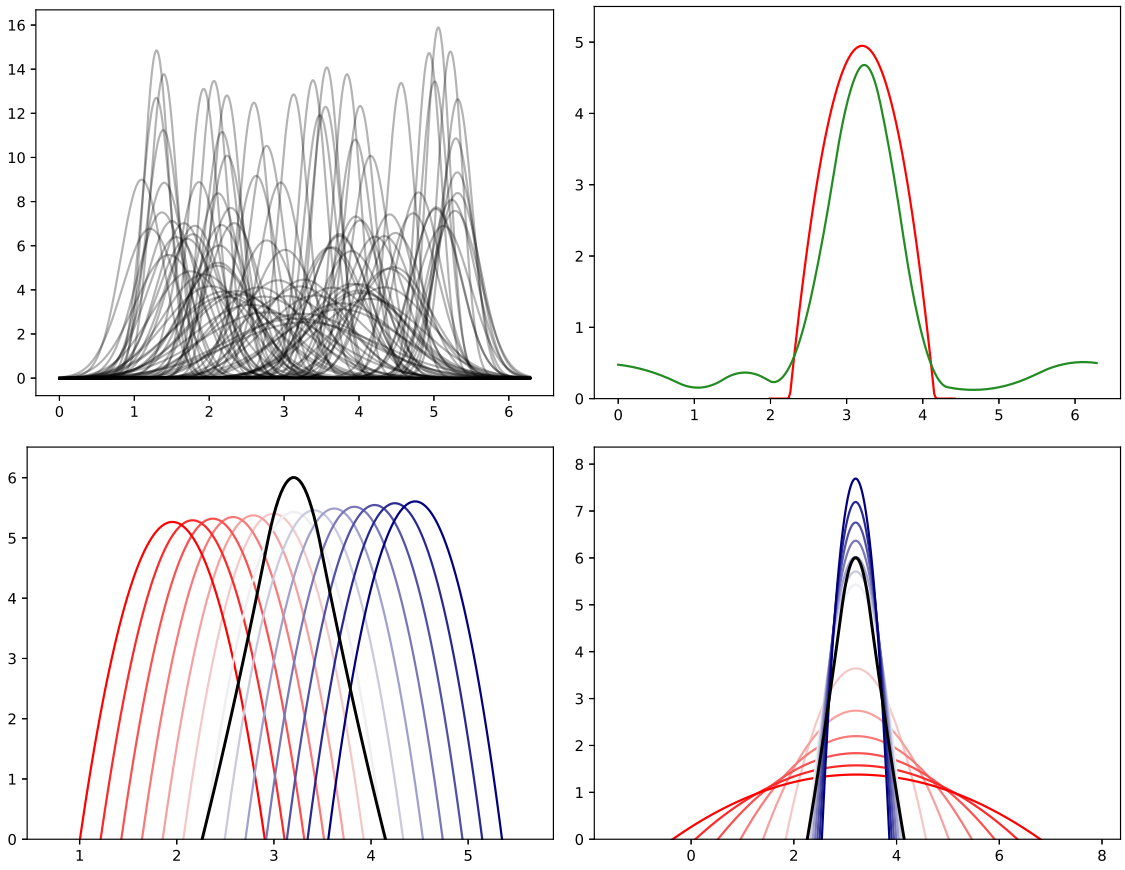


图 10：上排：模拟数据（左）以及 Wasserstein 重心的比较（右）：红线是 $\mathcal{W}_2$(R) 中的重心，绿线是 $\mathcal{W}_2(\mathbb{S}_1)$ 中的重心。下排：高斯测度在 $\mathcal{W}_2$(R) 中的前两个主方向，黑色实线为重心。

众数处的质量需要移动，才能与我们生成的单峰测度相匹配。具体而言，沿第一主方向移动所导致的位置变化，比在实数集上进行主成分分析时更小，且密度在原点附近呈现出更明显的众数。请注意，接近 0 的众数有逐渐靠近主众数的趋势。这两种影响与第二主方向（图中未显示）共同作用，从而逼近数据集中的单峰分布。

我们还考虑了另一个用于对 PCA 进行中心化的点 $\bar{\mu}$，即其中一个观测值，如图 11（下图）中的黑色实线所示。在这种情况下，我们令 $\log_{\bar{\mu}}(\mu_0)$ 等于 $\log_{\bar{\mu}}(\mu_i)$ 的经验均值。该结果显然是单峰的，并且得到的第一主方向显著不同。事实上，沿此方向移动时，密度呈单峰，其位置变化相当大，尽管略有偏斜（见图 11（下图）中的红色密度曲线）。有人可能会认为，与以重心为中心进行 PCA 时得到的方向相比，这个方向更容易解释。然而，使用该方向的重构误差是使用重心所得“原始”方向时的两倍。我们认为，当 PCA 以重心以外的点为中心时，

将主方向解释为主要“变异来源”可能会产生误导，如本例所示。不过，这一过程确实可能为所分析的数据集提供具有实际意义的见解：在本例中，测度的位置确实发生了显著变化。

因此，如果进行主成分分析的最终目的是深入了解数据集，而重心呈现出数据点中没有体现的某些特征（如本例中，重心呈三峰分布，而所有数据点均呈单峰分布），那么可以考虑选择其他候选点作为主成分分析的中心。另一方面，如果目的是降维，我们认为重心是唯一合理的主成分分析中心候选点，因为它必然会带来更小的重构误差，这意味着降维后的数据所含的信息损失更少。

D 眼睛数据集的其他图

图 12 展示了在高度 0.3 处切割树状图后得到的眼睛数据集精细化簇。观察图 12 中的精细化簇可以发现，所有簇的形状都略有不同：从左到右、从第一行到第二行，第一个簇左侧有一个凸起，右侧相对平坦；第二个簇左侧有一个凸起，右侧还有一个较小的凸起；第三个簇沿垂直轴方向更加细长；第四个簇右侧的凸起更大，而且与第一个簇相比，右侧更平坦；第五个簇的形状与第二个簇相似，但左下区域的质量更多；第六个簇的上部比其他所有簇都厚得多；最后一个簇的上部也相当厚，但右侧有一个小凸起，并且与第六个簇相比，左侧轮廓更平坦。

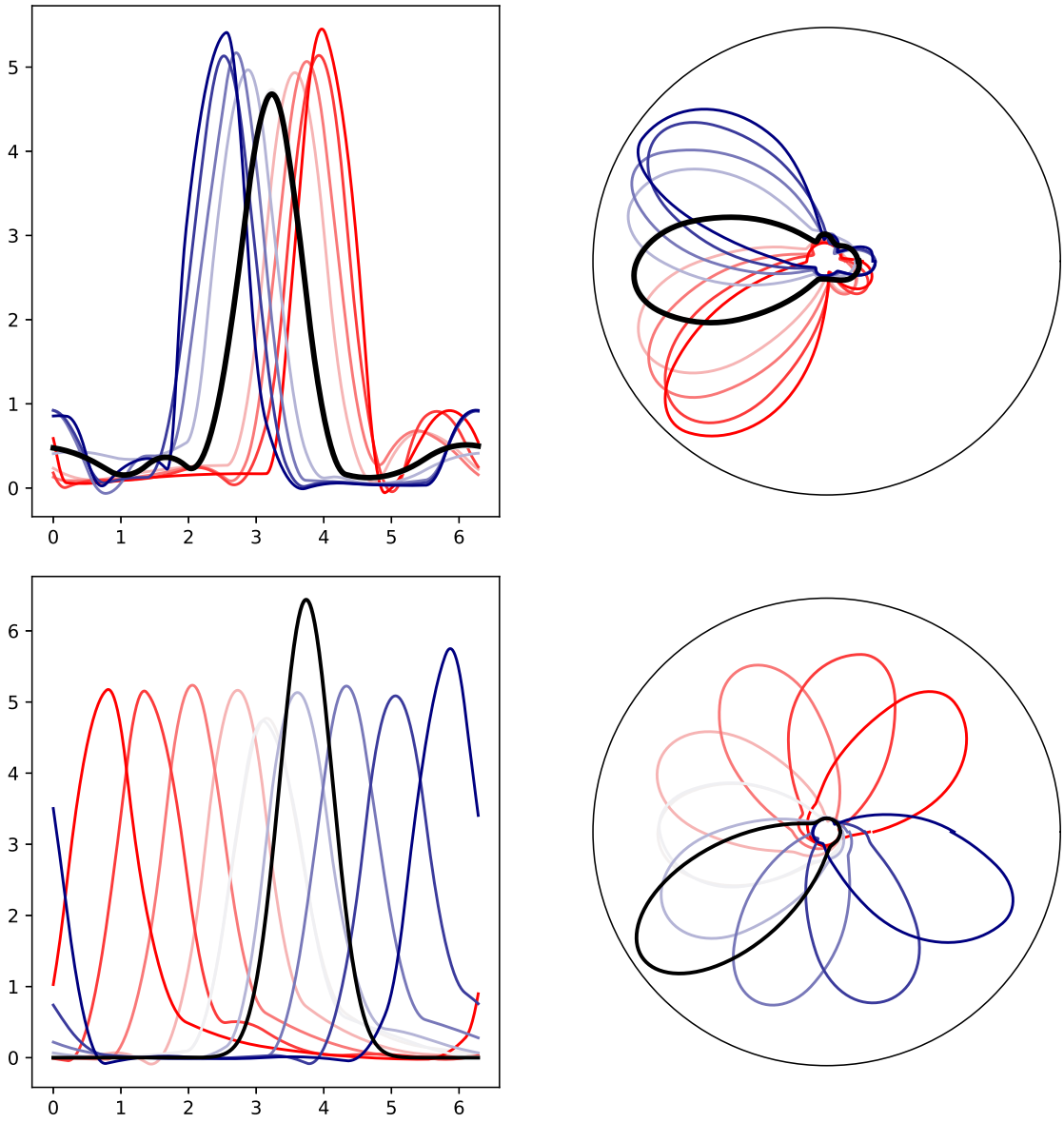


图 11：上排：下排：

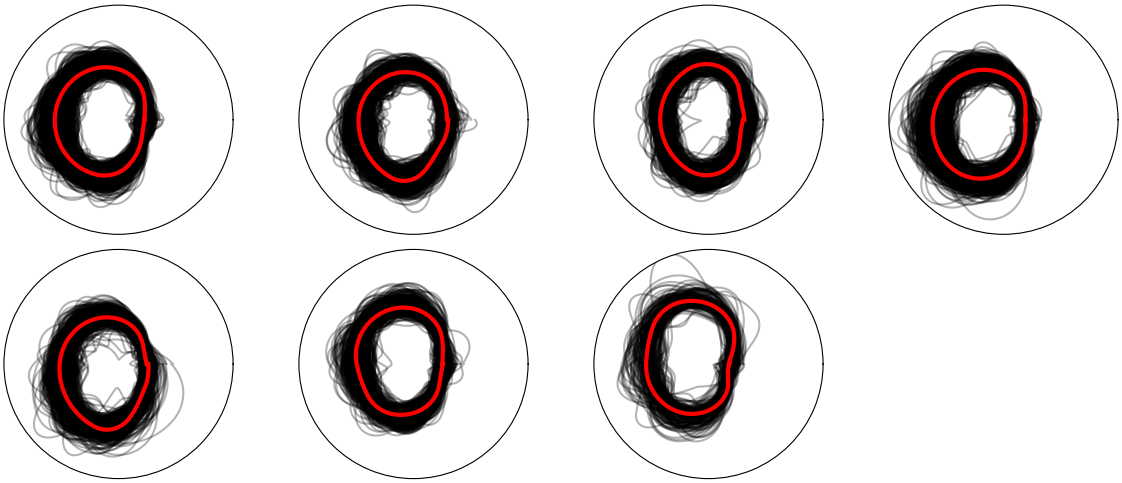


图 12：眼睛数据集细分为七个簇。

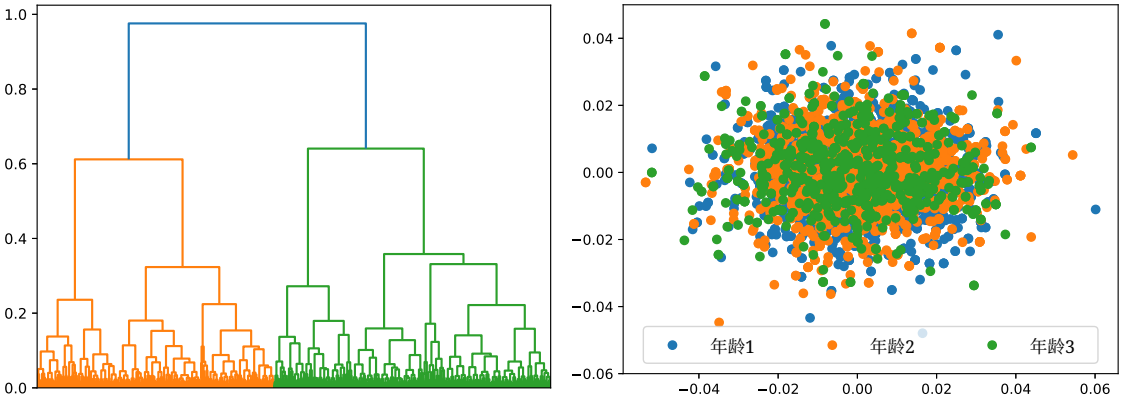


图13：层次聚类树状图（左）以及按年龄组（颜色）分层显示的沿前两个主方向的得分散点图，用于第6节所分析的OCT测量数据集的分析。