arXiv:2611.32773v1 [math.PR] 24 Nov 2026

椭圆分布的二次瓦瑟斯坦输运

Jonathan Ansari[1] Ludger Rüschendorf[2]

[1] 萨尔茨堡巴黎洛德龙大学，奥地利
jonathan.ansari@plus.ac.at[2] 弗赖堡阿尔贝特-路德维希大学，德国
ruschen@stochastik.uni-freiburg.de

2026年11月24日

摘要

高斯分布之间的二次 Wasserstein 距离由经典的布雷斯–瓦瑟斯坦公式给出，该公式仅取决于均值、协方差矩阵 Σ_0 和 Σ_1，以及它们之间的相互作用 $K = (\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2})^{1/2}$。对于具有共同径向分布的椭圆分布，同一公式也成立。本文通过识别一个额外的输运代价 Γ，将这一表示推广到一般椭圆分布；该代价取决于径向分布以及 K 的谱。我们将 Γ 描述为一个涉及狄利克雷分布的加权二次输运问题，并证明最优耦合在 K 的特征子空间内保持方向不变。当 K 只有一个特征值时，我们得到显式的代价公式和最优蒙日映射。对于一般的谱，我们利用嵌套谱投影和 Knothe–Rosenblatt 耦合，推导出非常紧的 Γ 下界和上界。下界基于一维积分，可以快速计算。对于高斯、Student-t、均匀球分布和均匀球面分布，上界也可通过动态规划得到计算效率高的表示。数值实验表明，在许多情形下，我们将此前已知的界之间的差距缩小了 80% 以上，在某些情况下甚至超过 99。9%，即使是在维度为 10^6 时也是如此。这些改进在高维或重尾情形下尤其有价值，因为此时精确计算二次 Wasserstein 距离颇具挑战性。

关键词。 二次 Wasserstein 距离；椭圆分布；Bures–Wasserstein 距离；Gelbrich界；谱约化；径向分布；Knothe–Rosenblatt 耦合；动态规划。
2020年数学主题分类。 主要分类：49Q22；次要分类：60E05、62H05、90C39。

1 引言

对于概率测度 $P_0,\ P_1 \in \mathcal{P}_2(\mathbb{R}^d)$，二次 Wasserstein 距离定义为

$$W_2^2(P_0,P_1) := \inf_{\pi\in\Pi(P_0,P_1)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \|x-y\|^2 \,\pi(\mathrm{d}x,\mathrm{d}y), \tag{1}$$

其中，$\Pi(P_0,\ P_1)$ 表示 P_0 与 P_1的耦合集合。对于高斯分布 $P_i = N(m_i,\ \Sigma_i)$，$i = 0,\ 1$，有经典公式

$$W_2^2(P_0,P_1) = \|m_0-m_1\|^2 + d_{\mathrm{BW}}^2(\Sigma_0,\Sigma_1), \tag{2}$$

关于*Bures-Wasserstein* 距离

$$d_{\mathrm{BW}}^2(\Sigma_0,\Sigma_1) := \mathrm{tr}(\Sigma_0)+\mathrm{tr}(\Sigma_1) - 2\,\mathrm{tr}\left[\left(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2}\right)^{1/2}\right]. \tag{3}$$

高斯分布的经典公式在早期的最优传输和矩阵文献中以几种等价形式得到；其中包括 [10, 24, 17, 26]。其几何学随后由 Takatsu [29] 发展。关于 Bures–Wasserstein度量的矩阵分析论述，参见 [2]；关于近期将其扩展到高斯过程之间的适应传输，参见 Gunasingam 和 Wong[18]。

高斯公式的两个基本扩展归功于 Gelbrich [13]。首先，对于任意概率分布 P_0, $P_1 \in \mathcal{P}_2(\mathbb{R}^d)$，其均值为 m_0, m_1，协方差矩阵为 Σ_0, Σ_1，*Gelbrich*界 指出

$$W_2^2(P_0, P_1) \geq \|m_0 - m_1\|^2 + d_{\mathrm{BW}}^2(\Sigma_0, \Sigma_1). \tag{4}$$

其次，Gelbrich [13] 证明，对于具有相同标准化径向分布律的椭圆分布，等号成立，即

$$W_2^2(P_0, P_1) = \|m_0 - m_1\|^2 + d_{\mathrm{BW}}^2(\Sigma_0, \Sigma_1) \qquad \text{for } P_i = \mathcal{E}_2(\mu_i, \Sigma_i, \nu), \tag{5}$$

其中，$\mathcal{E}_2(\mu, \Sigma, \nu)$表示均值为 $\mu \in \mathbb{R}^d$、正定协方差矩阵为 $\Sigma \in \mathbb{R}^{d\times d}$且标准化径向分布为 ν的平方可积椭圆分布，该分布定义在 $[0, \infty$上，并满足 $\int_0^\infty r^2 \nu(\mathrm{d}r) = d$。回顾一下，任何此类分布都具有如下随机表示

$$X = \mu + RAU, \qquad R \sim \nu, \quad U \sim \mathcal{U}(\mathbb{S}^{d-1}), \quad R \perp U, \quad AA^\top = \Sigma, \tag{6}$$

其中，$\mathcal{U}(\mathbb{S}^{d-1})$表示 R^d的单位球面上的均匀分布。注意，由于 $\mathbb{E}[UU^\top] = I_d/d$，归一化保证Cov$(X) = \Sigma$。

当径向分布律不同时，情况会发生显著变化。球面对等分布之间的径向传输是经典研究课题；参见 Cuesta-Albertos等人 [8]。近年来，Ghaffari和Walker [16]将此类传输用于Wasserstein重心。此外，Chen和Zhang [6] 研究了径向轮廓分布，并强调，对于生成元不同的一般椭圆分布，目前缺乏显式公式。Ghaffari和Walker [15] 为由公共耦合和正交变换联系起来的若干类分布（包括正态尺度混合分布）推导出了显式映射。Hamm和Korzeniowski [19]则针对随机向量的仿射变换给出了相关的下界和上界。一个关键难点在于，将最优径向传输与最优正定线性传输复合，通常并不能得到最优耦合，如 [8, 命题 3.14]所示。

Wong 在 [28, 第 8 节问题 2] 的开放问题汇编中，也强调了确定径向分布律不同的椭圆分布之间 Wasserstein距离这一难题。更准确地说，Wong 指出，对于径向分布律不同的情形，目前尚无通用的闭式公式；他给出了径向上界，并提出寻找一种同样取决于径向分布律的 Gelbrich下界改进。

谱约化

本文旨在确定椭圆分布的径向分布律不同时产生的额外输运代价。本文的首个主要结果是，根据协方差矩阵的谱结构，对这一代价进行精确分解。为此，对于$P_i = \mathcal{E}_2(m_i, \Sigma_i, \nu_i)$, $i = 0,1$，我们考虑正定的 *Bures*矩阵

$$K := \left(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2}\right)^{1/2}. \tag{7}$$

该矩阵 K 已出现在 Bures-Wasserstein 距离 (3) 中，它由 Bures 对齐的协方差因子 $A_0 = \Sigma_0^{1/2}$ 和 $A_1 = \Sigma_0^{-1/2}K$ 得到，这些因子满足 $A_iA_i^\mathsf{T} = \Sigma_i$ 和 $A_0^\mathsf{T}A_1 = K$。因此，变换 $A_i^{-1}(X_i - m_i)$ 得到标准化球面分布 $\eta_{\nu_i} :=$ law(R_iU_i)，而 K则刻画了这两个椭圆分布律的相对协方差几何；另见图1。

现在，令 $\boldsymbol{\kappa} = (\kappa_1, \ldots\ldots, \kappa_k)$ 表示 K 的互异特征值，其对应的特征子空间维数为 $n_1, \ldots\ldots, n_k$。我们的第一个主要结果，即定理 3.1，给出了如下表示

$$W_2^2(P_0, P_1) = \|m_0 - m_1\|^2 + d_{\mathrm{BW}}^2(\Sigma_0, \Sigma_1) + \mathcal{W}_{2,\boldsymbol{\kappa}}^2(\beta_0, \beta_1), \tag{8}$$

其中，$\mathcal{W}_{2,\ \boldsymbol{\kappa}}$ 表示式（32）中 k 维分布 β_0 与 β_1 之间的加权二次 Wasserstein 距离。后两者是块半径的联合分布；块半径是两两正交特征子空间中的径向部分，即

$$\beta_i = \mathrm{law}\left(R_i\sqrt{D_1}, \ldots, R_i\sqrt{D_k}\right), \quad R_i \sim \nu_i \text{ independent of } (D_1, \ldots, D_k) \sim \mathrm{Dirichlet}\left(\tfrac{n_1}{2}, \ldots, \tfrac{n_k}{2}\right);$$

参见命题 2.5。因此，原来的 d维输运问题可化简为一个加权输运问题，其维数等于 K的互异特征值的数量 k 。

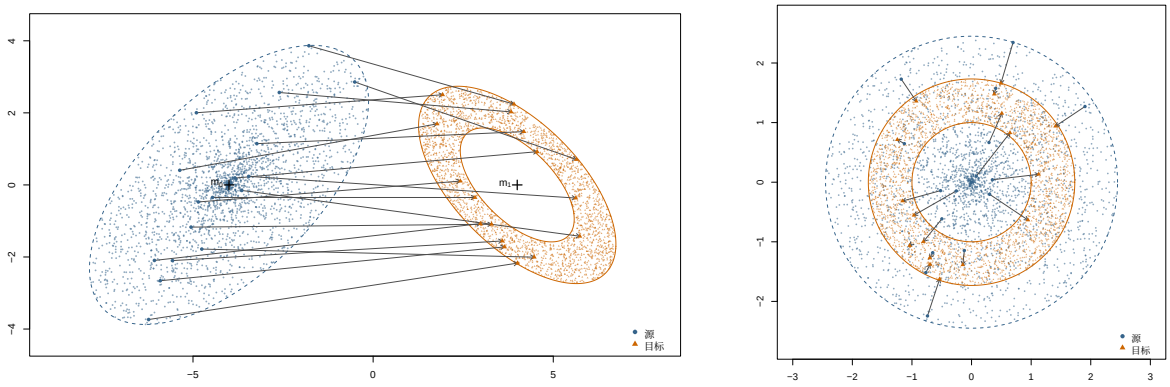

图1：椭圆分布 $P_i = \mathrm{E}_2(m_i, \Sigma_i, \nu_i)$、$i = 0$、1 之间的最优蒙日输运，其均值为 $m_0 = (-4, 0)^\top$ 和 $m_1 = (4, 0)^\top$，协方差因子为 $A = \left(\begin{smallmatrix} 3/2 & 1/2 \\ 1/2 & 3/2 \end{smallmatrix}\right)$ 和 $B = \left(\begin{smallmatrix} 3/2 & -1/2 \\ -1/2 & 3/2 \end{smallmatrix}\right)$，径向变量为 $R_0 \sim \nu_0 = \mathrm{U}(0, \sqrt{6})$ 和 $R_1 \sim \nu_1 = \sqrt{\mathrm{U}(1, 3)}$。协方差矩阵满足 $\Sigma_0 = AA^\top$ 和 $\Sigma_1 = BB^\top$，而 Bures矩阵为 $K = A^\top B = 2I_2$。左：原始椭圆坐标中的最优传输。右：标准化球面坐标中的相应传输，其中最优径向映射由 $\tau(r) = F_{\nu_1}^{-1}(F_{\nu_0}(r)) = \sqrt{1 + 2r/\sqrt{6}}$ 给出。箭头连接由推论3.3中的最优蒙日映射 $\mathcal{T}$ 匹配的点，其代价为 $W_2^2(P_0, P_1) = 64 + 2 + \frac{120 - 72\sqrt{2} - 4\sqrt{6}}{15} \approx 66.56$。

生成元效应

式 (8) 中的最后一项描述了由不同标准化径向分布律引起的额外输运代价。我们将这一量称为生成元效应，定义如下：

$$\Gamma(P_0, P_1) := \mathcal{W}_{2,\boldsymbol{\kappa}}^2(\beta_0, \beta_1) = W_2^2(P_0, P_1) - \|m_0 - m_1\|^2 - d_{\mathrm{BW}}^2(\Sigma_0, \Sigma_1). \tag{9}$$

由于 K 是正定的，当且仅当 $\nu_0 \neq \nu_1$时，生成元效应为正。需要注意的是，该量不仅取决于径向分布，还取决于由 K编码的协方差几何。

在各向同性情形下，$K = \lambda I_d$ 只有一个特征值 λ（即 $k = 1$），我们可以显式确定输运代价 $W_2(P_0, P_1)$以及从 P_0 到 P_1 的最优蒙日映射；参见推论3.3和图1。然而，在各向异性情形下（即 $k \geq 2$），通常不存在类似的闭式表达式，因为谱约化会导出多维最优输运问题$\mathcal{W}_{2,\boldsymbol{\kappa}}(\beta_0, \beta_1)$，其中涉及块半径的联合分布。

改进界

由于$\mathcal{W}_{2,\boldsymbol{\kappa}}(\beta_0, \beta_1)$没有显式公式，可计算的下界和上界便尤其值得关注。Wong也在 [28, 第8节问题 2]，中提出了这一问题，并给出了初等上界

$$U_{\mathrm{Wong}} := \frac{\operatorname{tr} K}{d} W_2^2(\nu_0, \nu_1), \tag{10}$$

这可以通过对径向变量进行共单调耦合，并选取一个与半径独立的共同均匀球面方向来实现。Wong 还要求对 Gelbrich 下界 $L_{\mathrm{Gelbrich}} := 0$ 作出对径向变量敏感的改进，该改进由 (4) 和 (5) 得到。在推论 3.5 中，我们根据 K 的最小特征值，以及标准化球面分布之间的最优径向传输，推导出一个初等下界 $\lambda_{\min}(K)\delta_R^2$。然而，一般而言，这些初等界过于宽泛，无法准确估计生成元效应。

作为本文的第二项主要贡献，我们通过 $\Gamma(P_0, P_1)$显著收紧了关于生成元效应的基本下界和上界

$$0 = L_{\mathrm{Gelbrich}} \leq \lambda_{\min}(K)\delta_R^2 \leq L_K \leq \Gamma(P_0, P_1) \leq U_K \leq U_{\mathrm{Wong}}. \tag{11}$$

d	径向对	Bures 矩阵	k	$L_{\min}$	L_K	$\Gamma(P_0,\ P_1)$	U_K	U_{Wong}	改进
2	N–t_3	K_{ar}	2	0.5325	0.773	≈ 0.79	0.7987	0.7987	96。78%
2	N–t_3	K_{pow}	2	0.3195	0.7524	≈ 0.77	0.784	0.7987	96.04%
2	N–t_3	K_2	2	0.3994	0.7601	≈ 0.78	0.7987	0.7987	95。17%
10	N–t_3	K_2	2	13.2747	23.2587	≈ 24	24.8822	26.5494	93.89%
10	N–t_3	K_{ar}	10	4.8272	24.6431	–	26.1541	26.5494	94。31%
10	N–t_3	K_{pow}	10	1.7131	20.2948	–	21.9639	26.5494	93.71%
100	N–t_3	K_2	2	1 738.367	3 212。9691	$\approx 3\,300$	3 339.4969	3 476.7339	96。36%
100	N–t_3	K_{ar}	100	68.8462	3 329.8786	–	3 437.772	3 476.7339	96.90%
100	N–t_3	K_{pow}	100	21.2646	2 098.3883	–	2 280.9218	3 476.7339	94.75%
10^6	N–t_3	K_2	2	2.0180×10^{11}	4.032050×10^{11}	$\approx 4.0334 \times 10^{11}$	$4.033801 \times 10^{11\,\dagger}$	4.0360×10^{11}	99.956610%
10^6	N–t_3	K_{ar}	10^6	807 191.6524	$4.033847 \times 10^{11\,\dagger}$	–	$4.035401 \times 10^{11\,\dagger}$	4.0360×10^{11}	99.961480%
10^6	N–t_3	K_{pow}	10^6	245 357。2355	$2.1089 \times 10^{11\,\dagger}$	–	$2.2914 \times 10^{11\,\dagger}$	4.0360×10^{11}	95。48%

表 1：高斯分布与 Student-t_3 分布之间的生成元效应 $\Gamma(P_0,\ P_1)$ 在式 (11) 中的界。谱 K_2，K_{ar}，K_{pow} 列于表 3 中，并满足 $\operatorname{tr} K = d^2$。匕首符号表示 L^- 或U^+ 在 $d = 10^6$处的取值；参见式 (58)。短横线表示未报告 $\Gamma(P_0,\ P_1)$ 的数值近似；详情请参见第5节。

我们在定理 3.6 中给出的下界 L_K 基于嵌套谱投影。更具体地说，我们考虑投影到由 K 对应于越来越大特征值的特征子空间的嵌套和上。这一构造利用了 K 的谱结构，并得到了一个关于投影径向分布之间一维 Wasserstein距离的显式下界。

定理 3.8 中的上界 U_K 通过条件共单调 Knothe–Rosenblatt 耦合对分块半径向量进行耦合而得到。一般而言，计算相应的传输代价需要进行多维积分。然而，如定理 4.1 所示，对于任意成对的高斯分布、Student-t分布、均匀球分布和均匀球面分布，这些代价都可以表示为一维积分。此外，U_K 的构造基于对矩阵 K的所有正交特征子空间投影组合进行优化。在定理 4.1 中，我们通过动态规划加速了这一优化，并证明它可以利用 $O(d^2)$ 次转移高效完成。

我们的数值实验表明，即使在维度高达 $d = 10^6$ 的情况下，所得界仍然非常紧。在许多情形下，与经典的 Gelbrich–Wong 界相比，我们的界将相应区间的宽度缩小了 80% 以上；在某些情况下，缩小幅度甚至超过 99.9%；详情请参见表 1 和第5节。这类紧界在高维或重尾情形下尤其有价值，因为此时精确计算二次最优传输代价颇具挑战。基于样本的方法会受到维数灾难的影响，其收敛速率还取决于分布的尾部行为 [12, 31], ，而 Sinkhorn 等熵正则化方法会引入额外的近似偏差 [9, 14]。通过利用椭圆分布的谱结构，我们的方法能够直接给出未经正则化的 Wasserstein 传输代价的下界和上界，无需进行高维离散化或熵正则化。

这类计算上易于处理的界与 Wasserstein 几何的应用相关，包括椭圆分布嵌入 [23], 、仿射变形统计模型 [1], 、以及径向等高分布的 Wasserstein重心和混合分布 [16, 6, 7]。在这些情形下，我们的结果有助于比较具有不同尾部行为的模型。此外，它们还可用于敏感性分析，以及分布鲁棒优化中 Wasserstein 歧义集的校准 [21]。

论文结构

本文其余部分安排如下。第2节介绍数学框架，包括 Bures 分解，以及将问题谱约化为块半径分布。第3节给出主要结果：二次 Wasserstein 距离的精确表示、各向同性情形下显式的最优蒙日映射，以及生成元效应更优的下界和上界。第4节利用一维积分和动态规划，高效计算上界。第5节通过数值实验展示这些界的准确性和计算可行性，包括高维情形。部分辅助结果和所有证明汇集于附录中。

2 数学框架

全文中， $d \geq 1$ 固定，且 $\mathbb{R}_+ = [0, \infty)$ 。此外，$\mathcal{P}_2(\mathbb{R}^d)$表示$\mathbb{R}^d$上具有有限二阶矩的Borel概率测度。对于 $P, Q \in \mathcal{P}_2(\mathbb{R}^d)$， $\Pi(P, Q)$表示$\mathbb{R}^d \times \mathbb{R}^d$ 上边缘分布为 P和 Q的Borel概率测度集合。二次 Wasserstein 距离定义见式（1）。对于最优耦合的存在性，我们采用标准结果而不再赘述；例如，参见 Rachev 和 Rüschendorf[25]以及 Villani [30]。

对于具有二阶矩的椭圆分布，采用归一化参数化方式较为方便。因此，令

$$\mathcal{R}_d := \left\{ \nu \in \mathcal{P}_2(\mathbb{R}_+) : \int_0^\infty r^2 \, \nu(\mathrm{d}r) = d \right\} \tag{12}$$

表示标准化径向分布的类，并令 $\mathbb{S}^d_{++}$ 表示取值于 $\mathbb{R}^{d\times d}$ 的正定矩阵集合。对于 $m \in \mathbb{R}^d$、 $\Sigma \in \mathbb{S}^d_{++}$ 和生成元 $\nu \in \mathcal{R}_d$，回顾

$$\mathcal{E}_2(m, \Sigma, \nu) := \mathrm{law}(m + ARU), \qquad AA^\mathsf{T} = \Sigma, \tag{13}$$

其中 $R \sim \nu$ 在欧氏单位球面 $\mathbb{S}^{d-1}$ 上服从均匀分布，且 R 与 U 相互独立。(13) 中的分布律不依赖于特定的因子 A。此外，(12) 中的平方可积性和归一化条件确保 m 和 Σ 分别为均值和协方差矩阵；参见引理 A.1。对于可测映射 T，推前测度 $T_\# P$ 定义为 $(T_\# P)(B) = P(T^{-1}(B))$。记号 $\mathrm{law}(X)$ 表示 X 的分布，且 $X \overset{\mathrm{d}}{=} Y$ 表示分布相等。对于有限区间 $J \subset \mathbb{R}$，我们用 $\mathcal{U}(J)$ 表示在 J 上的均匀分布。对于子空间 $E \subseteq \mathbb{R}^d$， I_E 表示 E 上的恒等算子，而 Id 表示所考虑环境空间上的恒等映射。

在本节的其余部分中，我们将介绍第3节主要结果所需的概念和预备结果。

2.1 二次最优传输预备知识

首先，我们回顾二次最优传输中的一个基本定理，该定理将最优蒙日映射刻画为某个凸函数的次梯度。

定理2.1 （二次最优性与凸次梯度， [27, 4]）。 设 $P, Q \in P_2$（$\mathbb{R}^d$）。若耦合$\pi \in \Pi$（P, Q）关于二次 *Wasserstein* 距离是最优的，当且仅当存在一个适当的下半连续凸函数 ϕ ，使得

$$y \in \partial\phi(x) \quad \textit{for } \pi\textit{-almost every } (x, y), \tag{14}$$

其中， $\partial\phi$（x）表示 ϕ 在 x处的凸次微分。特别地，如果可测映射 T 满足$T_\# P = Q$和 T（x） $\in \partial\phi$（x）对 P-几乎处处的 x成立，那么 （Id T）$_\# P$是一个最优 *Monge*耦合。如果 P是绝对连续的，则最优耦合是唯一的，并且由凸梯度映射 $T = \nabla\phi$诱导。

这一存在性及次梯度刻画由 Rüschendorf 和 Rachev [27],通过对偶性和凸分析得到。在 Lebesgue连续的情形下，这一刻画尤其意味着存在最优凸梯度传输映射。关于源分布 Lebesgue 绝对连续时凸梯度传输映射的存在性和唯一性，参见 Brenier [3, 4]。

下面这个关于球对称分布的经典恒等式，是定理2.1中凸梯度刻画的一个应用。该恒等式通过将半径进行共单调耦合，并使用一个与半径独立的共同均匀球面方向得到；示例见图1右侧的图。回想一下，对于$\nu \in \mathcal{R}_d$，相应的标准化球面分布记作

$$\eta_\nu := \mathrm{law}(RU), \qquad R \sim \nu, \; U \sim \mathcal{U}(\mathbb{S}^{d-1}), \; R \perp U. \tag{15}$$

还要注意，球对称分布是定义在 $\mathbb{R}^d$ 上的概率分布，且在正交变换下保持不变。

引理 2.2 （球对称分布之间的最优传输， [8]）。 设 ν_0， $\nu_1 \in \mathcal{R}_d$。则有

$$W_2^2(\eta_{\nu_0}, \eta_{\nu_1}) = W_2^2(\nu_0, \nu_1) = \int_0^1 \left| F_{\nu_0}^{-1}(u) - F_{\nu_1}^{-1}(u) \right|^2 \, \mathrm{d}u. \tag{16}$$

若 ν_0 是无原子的，则 W_2（$\eta_{\nu_0}, \eta_{\nu_1}$）的最优蒙日映射为

$$S(z) = \begin{cases} \dfrac{\tau(\|z\|)}{\|z\|} z, & z \neq 0, \\ 0, & z = 0, \end{cases} \tag{17}$$

其中 $\tau = F_{\nu_1}^{-1} \circ F_{\nu_0}$ 将 ν_0 推送至 ν_1。

对于一般的椭圆分布，根据 (6)，它们只是球对称分布的线性变换，且最优耦合存在。如果源分布是绝对连续的，定理2.1进一步保证存在唯一的最优蒙日映射。其显式形式通常未知。

接下来，我们利用 K 的谱分解来降低底层最优传输问题的维数。当 K 只有一个特征值时，我们仍能得到输运代价的显式公式和一个最优蒙日映射。更一般地，我们利用 K 的谱结构，在各种情形下推导出最优传输成本的极为精确的下界和上界。

2.2 Bures 对齐与精确代价分解

现在，令 $\nu_i \in \mathcal{R}_d$ 为任意对象，并考虑 $P_i = \mathcal{E}_2(m_i, \Sigma_i, \nu_i)$，其中 $m_i \in \mathbb{R}^d$ 且 $\Sigma_i \in \mathbb{S}_{++}^d$, $i = 0, 1$。将 $\mathcal{E}_2(m_i, \Sigma_i, \nu_i)$ 分解为位置、协方差几何和径向部分，并非只是对椭圆分布的一种参数化。对于二次代价，它还与一个精确的逐耦合恒等式密切相关。为说明这一点，令 $X_0 \sim P_0$ 且 $X_1 \sim P_1$。由于 $\mathbb{E}(X_0 - m_0) = \mathbb{E}(X_1 - m_1) = 0$，每个耦合都满足

$$\mathbb{E}\,\|X_0 - X_1\|^2 = \|m_0 - m_1\|^2 + \mathbb{E}\,\|(X_0 - m_0) - (X_1 - m_1)\|^2 . \tag{18}$$

因此，在进行任何优化之前，位置项就已分离。

接下来，我们分离协方差几何。因此，回顾一下，式（7）中的正定布雷斯矩阵 K 满足 $A_0 = \Sigma_0^{1/2}$ 和 $A_1 := \Sigma_0^{-1/2} K$，以及 $A_0 A_0^{\mathsf{T}} = \Sigma_0$、$A_1 A_1^{\mathsf{T}} = \Sigma_1$ 和 $A_0^{\mathsf{T}} A_1 = K$。于是，中心化后，我们可以写成

$$X_0 - m_0 = A_0 Z_0, \qquad X_1 - m_1 = A_1 Z_1, \tag{19}$$

其中 $Z_i = R_i U_i \sim \eta_{\nu_i}$ 是标准化球对称随机向量。回顾式（15），η_{ν_i} =law($R_i U_i$) 是中心化的，且协方差为 I_d。

选择 A_0 和 A_1 后，协方差贡献便与其余输运代价分离。事实上，对于 Z_0 和 Z_1 的任意耦合，以下两个二次展开具有相同的依赖耦合的交叉项：

$$\begin{aligned} \mathbb{E}\,\|A_0 Z_0 - A_1 Z_1\|^2 &= \operatorname{tr}(\Sigma_0) + \operatorname{tr}(\Sigma_1) - 2\mathbb{E}[Z_0^{\mathsf{T}} K Z_1], \\ \mathbb{E}[(Z_0 - Z_1)^{\mathsf{T}} K (Z_0 - Z_1)] &= 2\operatorname{tr}(K) - 2\mathbb{E}[Z_0^{\mathsf{T}} K Z_1]. \end{aligned}$$

比较这两个恒等式，并利用式（3）中 Bures–Wasserstein 距离的定义，可得

$$\mathbb{E}\,\|A_0 Z_0 - A_1 Z_1\|^2 = d_{\mathrm{BW}}^2(\Sigma_0, \Sigma_1) + \mathbb{E}\big[(Z_0 - Z_1)^{\mathsf{T}} K (Z_0 - Z_1)\big]. \tag{20}$$

该恒等式在取 下确界之前成立。结合式（18）可知，均值贡献与 Bures 协方差贡献彼此分离，而其余所有自由度都包含在标准化球对称分布之间的加权输运问题中。

对于 $M \in \mathbb{S}_{++}^d$ 和 $P,\ Q \in \mathcal{P}_2(\mathbb{R}^d)$，我们将加权二次 Wasserstein 距离定义为

$$W_{2,M}^2(P, Q) := \inf_{\gamma \in \Pi(P,Q)} \int_{\mathbb{R}^d \times \mathbb{R}^d} (z - w)^{\mathsf{T}} M (z - w)\, \gamma(\mathrm{d}z, \mathrm{d}w). \tag{21}$$

由此可得

$$W_{2,M}^2(P, Q) = W_2^2\big((M^{1/2})_{\#} P, (M^{1/2})_{\#} Q\big),$$

并且，由于 M 是正定的，$W_{2,\ M}$ 是$\mathcal{P}_2(\mathbb{R}^d)$ 上的一个度量。我们将前述输运代价的分解总结如下。

命题 2.3（运输成本分解）。对于 $P_i = \mathcal{E}_2(m_i,\ \Sigma_i,\ \nu_i)$，$i = 0,\ 1$，令 K 由 (7)给出。则

$$W_2^2(P_0, P_1) = \|m_0 - m_1\|^2 + d_{\mathrm{BW}}^2(\Sigma_0, \Sigma_1) + W_{2,K}^2(\eta_{\nu_0}, \eta_{\nu_1}), \tag{22}$$

其中

$$W_{2,K}^2(\eta_{\nu_0}, \eta_{\nu_1}) = W_2^2(\mathcal{E}_2(0, K, \nu_0), \mathcal{E}_2(0, K, \nu_1))\,. \tag{23}$$

注 2.4（生成元效应）。（a）Gelbrich 不等式 (4) 表明，以下量

$$W_{2,K}^2(\eta_{\nu_0}, \eta_{\nu_1}) = W_2^2(P_0, P_1) - \|m_0 - m_1\|^2 - d_{\mathrm{BW}}^2(\Sigma_0, \Sigma_1)$$

是非负的，并且当且仅当 $\nu_0 = \nu_1$ 时为零。事实上，$W_{2,\ K}$ 定义了一个度量，因为 K 是正定的。反之，当径向分布不同时，通过 $W_{2,\ K}^2(\eta_{\nu_0},\ \eta_{\nu_1})$ 得到的径向部分贡献为正。在下文的命题 2.5 中，我们证明，$W_{2,\ K}^2(\eta_{\nu_0},\ \eta_{\nu_1})$ 经过谱约化后，与最优传输成本 $\mathcal{W}_{22,\ \boldsymbol{\kappa}}(\beta_0,\ \beta_1)$ 一致。因此，我们得到

$$\Gamma(P_0, P_1) = W_{2,K}^2(\eta_{\nu_0}, \eta_{\nu_1}) \tag{24}$$

对于 (9) 中定义的生成元效应。

（b）若 $\widetilde{R}$ 是一个非标准化径向变量，且满足 $0 < \mathbb{E}[\widetilde{R}^2] < \infty$，令 $s = \sqrt{\mathbb{E}[\widetilde{R}^2]/d}$、$R = \widetilde{R}/s$，$A_* = sA$。则 $m + \widetilde{R}AU = m + RA_*U$，其中 $\mathbb{E}[R^2] = d$，协方差矩阵为 $A_*A_*^{\mathsf{T}}$。

（c）对于固定的径向分布 ν_0，ν_1，考虑

$$G(K) := W_{2,K}^2(\eta_{\nu_0}, \eta_{\nu_1}). \tag{25}$$

不难看出，函数 G 是正齐次的、凹的，并且关于 Loewner 序递增˜。后者意味着，如果 $K - K$ 是半正定的，那么$G(K) \leq G(K)$。

2.3 谱约化：特征子空间与狄利克雷分块半径

（23）中的各向异性完全由 K的谱编码。为此，设 κ_1、.....、κ_k是 K的互异特征值，设 E_j 是相应的两两正交的特征子空间，设 π_j 表示到 E_j上的正交投影，并令 $n_j := \dim E_j$。则

$$K = \sum_{j=1}^{k} \kappa_j \pi_j, \qquad n_1 + \cdots + n_k = d, \tag{26}$$

因此，加权球面传输问题的代价为

$$\begin{aligned}(z-w)^{\mathsf{T}} K (z-w) &= \sum_{j=1}^{k} \kappa_j (z-w)^{\mathsf{T}} \pi_j (z-w) \\ &= \sum_{j=1}^{k} \kappa_j (z-w)^{\mathsf{T}} \pi_j^{\mathsf{T}} \pi_j (z-w) = \sum_{j=1}^{k} \kappa_j \left\|\pi_j z - \pi_j w\right\|^2,\end{aligned}$$

其中，第二个等式利用了 π_j 是正交投影这一事实。上述恒等式说明了为什么特征子空间是需要考察的正确对象。在每个特征子空间 E_j上，矩阵 K 的作用等同于乘以相应的特征值 κ_j，即对所有 $v \in E_j$都有 $Kv = \kappa_j v$。因此，E_j 内的所有方向都具有相同的二次权重，而对 $\pi_j z$ 和 $\pi_j w$ 同时施加正交变换不会改变相应的代价。另一方面，属于不同特征子空间的方向必须保持可区分，因为它们的权重不同。

现在，对于每个特征子空间 E_j，我们考虑如下块长度，我们也将其记为 块半径：

$$r_j := \|\pi_j z\|, \qquad s_j := \|\pi_j w\|.$$

于是，柯西–施瓦茨不等式给出

$$\|\pi_j z - \pi_j w\|^2 = r_j^2 + s_j^2 - 2\langle \pi_j z, \pi_j w\rangle \geq r_j^2 + s_j^2 - 2r_j s_j = (\|\pi_j z\| - \|\pi_j w\|)^2. \tag{27}$$

当且仅当 $\pi_j z$ 和 $\pi_j w$ 正向共线时等号成立，也就是说，只要二者均非零，它们就指向同一方向。因此，对于固定的块半径，在每个特征子空间内，使两个向量沿同一方向对齐，可以使输运代价最小。由此，剩余的优化问题仅涉及块半径的联合耦合。

因此， Z 和 W 的每个耦合都会诱导出<styლე id='3'>分块半径向量的一个耦合

$$p_K(Z_0) := \big(\|\pi_1 Z_0\|, \dots, \|\pi_k Z_0\|\big), \qquad p_K(Z_1) := \big(\|\pi_1 Z_1\|, \dots, \|\pi_k Z_1\|\big), \tag{28}$$

从而（27）给出下界

$$(Z_0 - Z_1)^{\mathsf{T}} K (Z_0 - Z_1) \geq \sum_{j=1}^{k} \kappa_j \big(\|\pi_j Z_0\| - \|\pi_j Z_1\|\big)^2. \tag{29}$$

现在，关键在于（29）中的下界是可以达到的。其思路是：首先为两个分块半径向量 p_K（Z_0）和 p_K（Z_1）选择一个最优耦合。尽管其依赖结构通常无法显式获知，但最优传输理论保证了这种耦合的存在。然后，在每个特征子空间 E_j 中选择独立的均匀方向 V_j，使其与耦合的块半径相互独立，并在同一特征子空间内的两个向量之间使用相同的方向。

为使这一构造严谨明确，定义

$$\beta_i := \operatorname{law}\Big(R_i\sqrt{D_1}, \dots, R_i\sqrt{D_k}\Big), \qquad \text{for } (D_1, \dots, D_k) \sim \operatorname{Dirichlet}\Big(\frac{n_1}{2}, \dots, \frac{n_k}{2}\Big) \tag{30}$$

与 $R_i \sim \nu_i$ 独立。令 $(r,\ s) = ((r_1,\ \dots\dots,\ r_k),\ (s_1,\ \dots\dots,\ s_k)) \sim \gamma^*$ 为 β_0 与 β_1 关于加权二次代价 $\sum_{j=1}^k \kappa_j (r_j - s_j)^2$ 的最优耦合。独立于 $(r,\ s)$，选取相互独立的随机向量 $V_1,\ \dots\dots,\ V_k$，其中 V_j 在 E_j 的单位球面上服从均匀分布，并定义

$$Z_0^* := \sum_{j=1}^{k} r_j V_j, \qquad Z_1^* := \sum_{j=1}^{k} s_j V_j. \tag{31}$$

作为球形分块分解的结果（详见引理 B.1），我们有 $Z_0^* \sim \eta_{\nu_0}$ 和 $Z_1^* \sim \eta_{\nu_1}$。此外，由于两个向量在每个特征子空间内共享共同的方向 V_j，我们得到

$$\|\pi_j Z_0^* - \pi_j Z_1^*\|^2 = \|(r_j - s_j)V_j\|^2 = (r_j - s_j)^2,$$

其中，最后一个等式由 $\|V_j\| = 1$ 得出。因此，

$$\mathbb{E}[(Z_0^* - Z_1^*)^\top K (Z_0^* - Z_1^*)] = \mathbb{E}_{\gamma^*}\left[\sum_{j=1}^{k} \kappa_j (r_j - s_j)^2\right].$$

因此，(29) 中的下界得以达到，剩余的优化仅涉及分块半径向量的联合耦合。上述论证得到以下谱约化。

命题 2.5（谱约化与球对称分布之间的加权传输）。令 K 为式（7）中具有互异特征值 $\boldsymbol{\kappa} = (\kappa_1,\ \dots\dots,\ \kappa_k)$ 的 ***Bures***矩阵，其对应的特征子空间维数为 $n_1,\ \dots\dots,\ n_k$。对于 $i \in \{0,\ 1\}$，考虑式（30）中的分布 β_i。那么，我们有

$$W_{2,K}^2(\eta_{\nu_0}, \eta_{\nu_1}) = \mathcal{W}_{2,\boldsymbol{\kappa}}^2(\beta_0, \beta_1) := \inf_{\gamma \in \Pi(\beta_0, \beta_1)} \int_{\mathbb{R}_+^k \times \mathbb{R}_+^k} \sum_{j=1}^{k} \kappa_j (r_j - s_j)^2 \, \gamma(\mathrm{d}r, \mathrm{d}s). \tag{32}$$

注 2.6：(a) 尽管 η_{ν_i} 是定义在 $\mathbb{R}^d$上的球对称分布，但 (30) 中的 β_i 是定义在正正交象限 $\mathbb{R}_+^k$上的分布。它们描述了相对于 K的特征子空间分解，球向量的块半径的联合分布。特别地， β_i 通常并不是球对称分布。因此，命题 2.5 将两个球面分布之间的 d维各向异性二次最优传输问题，化简为正正交象限上的 k维输运问题。记住， $k \leq d$ 是 K的互异特征值的数量。这一化简去除了特征子空间内的所有角变量；剩下的是相应块半径之间的最优传输问题。

(b) (25) 中的函数 $G(K)$ 是正交不变且凹的，因此关于 K的特征值是舒尔凹的；参见 [22, 命题 3.C.2]。后者意味着，对于满足 $\operatorname{tr} K = \operatorname{tr} K$的正定矩阵 K 和$\sim$ $\sim K$，有

$$\sum_{j=1}^{q} \lambda_j^{\downarrow}(K) \le \sum_{j=1}^{q} \lambda_j^{\downarrow}(\tilde{K}), \quad q = 1, \dots, d-1, \quad \Longrightarrow \quad W_{2,K}^2(\eta_{\nu_0}, \eta_{\nu_1}) \ge W_{2,\tilde{K}}^2(\eta_{\nu_0}, \eta_{\nu_1}), \tag{33}$$

其中，$\lambda_j^{\downarrow}(K)$ 是 K 的特征值，按降序排列；$\tilde{K}$ 同理。因此，对于固定的 $\operatorname{tr} K$，增加谱离散度不会增大生成元效应。特别地，各向同性情形 $K = (\operatorname{tr} K/d) I_d$ 使 G 达到最大值，并达到 Wong 上界。

3 主要结果

本节介绍我们的主要结果。首先，我们根据 Bures–Wasserstein 距离以及其分块半径分布之间的低维输运问题，对椭圆分布之间的二次 Wasserstein 距离建立了一个精确分解。我们还通过提升约化问题的最优解，构造了最优耦合。

随后，我们考虑 Bures矩阵 K 仅具有一个特征值的特殊情形，并在此情形下得到显式的输运代价和一个最优蒙日映射。最后，我们分别利用嵌套谱投影以及分块半径向量的条件共单调耦合，为一般二次输运问题推导出改进的下界和上界。

3.1 通过谱约化表示 $W_2^2(P_0, P_1)$

现在，我们将命题 2.3 中的输运代价分解与命题 2.5 中确立的谱约化结合起来，从而得到椭圆分布之间二次 Wasserstein 距离的精确表示。此外，我们证明，可以利用各特征子空间内共同的均匀方向，将块半径分布的每个最优耦合提升为原椭圆分布的最优耦合。

定理 3.1 (Bures 分解与谱约化)。对于 $i = 0, 1$，令 $P_i = E_2(m_i, \Sigma_i, \nu_i)$，其中$\Sigma_i \in \mathbb{S}_{++}^d$且 $\nu_i \in R_d$。令 K 为 ***Bures***矩阵，其特征值为 $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_k)$，并令 β_i 由(30)给出。则

$$W_2^2(P_0, P_1) = \|m_0 - m_1\|^2 + d_{\mathrm{BW}}^2(\Sigma_0, \Sigma_1) + \mathcal{W}_{2,\boldsymbol{\kappa}}^2(\beta_0, \beta_1). \tag{34}$$

此外，每个最优耦合 $\gamma^* \in \Pi(\beta_0, \beta_1)$ 都可以如下提升为 P_0 与 P_1 的最优耦合。对于(31)中的 Z_0^* 和 Z_1^*，定义

$$X_0 := m_0 + \Sigma_0^{1/2} Z_0^*, \quad X_1 := m_1 + \Sigma_0^{-1/2} K Z_1^*. \tag{35}$$

那么，$X_0 \sim P_0$、$X_1 \sim P_1$和（X_0、X_1）是二次 ***Wasserstein*** 距离下的最优耦合。

注 3.2：（a）式（34）中的最后一项是式（9）中定义的生成元效应 Γ（P_0，P_1）。块半径分布 β_i 同时捕捉了全局半径大小及其在 K的互异特征子空间之间的分配。因此，生成元效应反映了径向分布与 K的谱几何之间的相互作用。

（b） K 的互异特征值数量 k 决定了约化输运问题的维数。如果 $K = \lambda I_d$，那么 $k = 1$，只剩下全局半径。如果 K 具有单谱，那么$k = d$，块半径就是特征基中的绝对坐标；符号可以进行最优匹配，而 d 绝对坐标仍然必须联合耦合。

(c) 约化问题 $\mathcal{W}_{2,\boldsymbol{\kappa}}(\beta_0, \beta_1)$ 不会分解为 k 个相互独立的一维输运问题。事实上，这些块半径满足 $\sum_{j=1}^{k} \big(R_i \sqrt{D_j}\big)^2 = R_i^2$，并共享同一个狄利克雷向量（$\overline{D_1}$，……，$D_k$）。因此，必须保留它们之间的依赖关系，按坐标构造的最优径向耦合未必能组合成具有给定边缘分布律 β_i 的耦合。一般而言，对于这个 k 维输运问题，不存在闭式解。因此，定理 3.1 提供的是精确的结构约化，而不是关于生成元效应的闭式公式。

在各向同性情形下，当 Bures矩阵 K 是单位矩阵的倍数 I_d 时，我们可以通过定理 2.1 和定理 3.1 确定最优蒙日映射及其代价，具体如下。相关示意请参见图1。

推论3.3 (最优蒙日映射**)**。 假设对于某个 $\lambda > 0$， $K = \lambda I_d$ 。则

(i) $W_2^2(P_0, P_1) = \|m_0 - m_1\|^2 + d_{\mathrm{BW}}^2(\Sigma_0, \Sigma_1) + \lambda W_2^2(\nu_0, \nu_1)$.

(ii) 如果 ν_0 是无原子的，则最优蒙日映射 T： $\mathbb{R}^d \to \mathbb{R}^d$ 由下式给出

$$T(x) = m_1 + \lambda \frac{\tau(t(x))}{t(x)} \Sigma_0^{-1}(x - m_0), \qquad t(x) = \left\| \Sigma_0^{-1/2}(x - m_0) \right\|, \tag{36}$$

P_0-几乎处处成立，其中 $\tau = F_{\nu_1}^{-1} \circ F_{\nu_0}$。

(iii) T 是适当凸势函数的梯度

$$\Phi(x) = m_1^{\mathsf{T}} x + \lambda \int_0^{t(x)} \tau(s)\, \mathrm{d}s, \tag{37}$$

for 对 P_0-几乎所有 $x \in \mathbb{R}^d$，约定 $F_{\nu_1}^{-1}(0) := \inf \mathrm{supp}(\nu_1)$ 以及 $F_{\nu_1}^{-1}(1) := \sup \mathrm{supp}(\nu_1)$.

备注 3.4。 推论3.3中的显式最优蒙日映射要求对于某个 $\lambda > 0$ 有 $K = \lambda I_d$ ，或者等价地，

$$\Sigma_1 = \lambda^2 \Sigma_0^{-1}.$$

因此，协方差矩阵不必是各向同性的，也不必彼此相同。反之，如果 $\Sigma_0 = \Sigma_1 = \Sigma$，那么 $K = \Sigma$；仅当 Σ 是单位矩阵的倍数时，它才是各向同性的。因此，协方差矩阵相同的情形通常不在推论3.3的适用范围内。尽管根据定理2.1中的凸梯度刻画，最优蒙日映射确实存在，但在此情形下没有可用的一般显式公式。

3.2 W_2（P_0， P_1）的改进界

如上所述，当径向分布不同时，且 K 至少有两个互异特征值（$k \geq 2$）时，生成元效应 $\Gamma(P_0, P_1) = \mathcal{W}_{22,\ \boldsymbol{\kappa}}(\beta_0, \beta_1)$ 通常没有显式公式，因为加权输运问题（32）的最优耦合目前尚无闭式表达。这促使我们寻找关于生成元效应的显式且在计算上易处理的下界和上界。

作为定理 3.1的第一个推论，我们得到生成元效应的一个简单的径向敏感下界：它补充了（39）中的初等上界 U_{Wong} ，并且当 $\nu_0 \neq \nu_1$ 和 $k \geq 2$ 时改进了经典的 Gelbrich界 $L_{\mathrm{Gelbrich}} = 0$ 。因此，记

$$\delta_R^2 := W_2^2(\nu_0, \nu_1) = \int_0^1 \left| F_{\nu_0}^{-1}(u) - F_{\nu_1}^{-1}(u) \right|^2 \, \mathrm{d}u \tag{38}$$

径向分布之间的二次 Wasserstein 距离。我们用 $\lambda_{\min}(K)$ 表示 K 的最小特征值，用 $\operatorname{tr} K$ 表示 K 的迹。

推论 3.5 （初等界）。 在定理 *3.1* 的假设下，生成元效应可通过以下方式界定：

$$\lambda_{\min}(K)\, \delta_R^2 \leq \Gamma(P_0, P_1) \leq \frac{\operatorname{tr} K}{d}\, \delta_R^2 = U_{\mathrm{Wong}}. \tag{39}$$

下文将显著改进（39）中的初等界。首先，我们利用 Bures矩阵 K 的谱结构，推导出更精确的下界。然后，通过使块半径条件共单调地耦合，建立改进的上界。

现在，为了确定改进后的下界，令 $q \in \{1, \ldots, d\}$。回顾一下，服从狄利克雷$(q/2,(d-q)/2)$分布的二元随机向量，其第一个分量服从贝塔分布$(q/2, (d-q)/2)$；见附录 A.2。因此，球对称随机向量 RU在任意 q维子空间上的正交投影的欧几里得范数，与 $R\sqrt{B_{q,d}}$具有相同的分布，其中 $B_{q,d} \sim$服从贝塔分布$(q/2,(d-q)/2)$，且与 R独立，约定 $B_{d,d} = 1$。现在，将两个此类投影球对称分布之间的平方 Wasserstein距离记为

$$\Delta_q := W_2^2\big(\mathrm{law}(R_0\sqrt{B_{q,d}}), \mathrm{law}(R_1\sqrt{B_{q,d}})\big), \qquad B_{q,d} \sim \mathrm{Beta}(q/2, (d-q)/2), \tag{40}$$

其中 $B_{q,d}$ 与相应的半径 R_i独立。令 $\Delta_0 = 0$ ，并且对于 $q = d$，有 $B_{d,d} = 1$。特别地，我们有 $\Delta_d = \delta_R^2$。

对于改进的下界，关键思路是考虑与递增特征值相对应的嵌套投影。因此，我们将互异特征值排序如下

$$0 < \kappa_1 < \cdots < \kappa_k \tag{41}$$

并考虑投影 $Q_j := \pi_j + \cdots + \pi_k$，它们作用于嵌套子空间 $E_j \oplus \cdots \oplus E_k$，相应的维数为

$$d_j := n_j + \cdots + n_k \qquad j = 1, \ldots, k. \tag{42}$$

利用望远镜求和论证，矩阵 K 可分解为

$$K = \sum_{\ell=1}^{k} \kappa_\ell \pi_\ell = \kappa_1 \sum_{\ell=1}^{k} \pi_\ell + \sum_{j=2}^{k} (\kappa_j - \kappa_{j-1}) \sum_{\ell=j}^{k} \pi_\ell = \kappa_1 I_d + \sum_{j=2}^{k} (\kappa_j - \kappa_{j-1}) Q_j. \tag{43}$$

第一项 $\kappa_1 I_d$ 作用于整个空间，并给出式 (39) 中的最小特征值下界，即 $\kappa_1 \Delta_d = \lambda_{\min}(K)\delta_R^2$。作用于嵌套子空间的非负权重 $\kappa_j - \kappa_{j-1}$ 会进一步提高此下界，如下所示。

定理 3.6 (通过嵌套投影得到的改进下界)**.** 假设 K 的特征值按(41) 中的方式排序，且 d_j 由(42)给出， $j \in \{1, \ldots, k\}$。那么，式(24) 中的生成元效应 Γ 满足

$$L_K := \kappa_1 \Delta_d + \sum_{j=2}^{k} (\kappa_j - \kappa_{j-1}) \Delta_{d_j} \le \Gamma(P_0, P_1). \tag{44}$$

如果 K 至少有两个互异特征值，且 $\nu_0 \neq \nu_1$，则相对于下界 $\lambda_{\min}（K）\ \delta_R^2 = \kappa_1 \Delta_d$ 的改进是严格的。

注3.7（与分块界的比较）。嵌套投影下界 L_K 优于分别考虑各特征子空间所得到的界。事实上，二次代价在正交子空间上的可加性意味着

$$\Delta_{d_j} \ge \sum_{\ell=j}^{k} \Delta_{n_\ell} \qquad \text{and thus} \qquad L_K \ge \sum_{j=1}^{k} \kappa_j \Delta_{n_j}.$$

因此，嵌套投影通过考虑块半径之间的联合依赖关系，给出了更强的界。该界只需计算 k 个一维 Wasserstein 距离，因此可以高效求得。对于 $k = 1$，有 $L_K = \kappa_1 \Delta_d = \kappa_1 W_2^2(\nu_0, \nu_1) = \Gamma(P_0, P_1)$。

为了改进式（39）中的初等基于迹的上界 U_{Wong} ，我们将 $\mathbb{R}^d$ 分解为 $m \le d$个两两正交的子空间，

$$\mathbb{R}^d = E_1 \oplus \cdots \oplus E_m. \tag{45}$$

随后，我们按照 Knothe–Rosenblatt 型构造，依次以条件共单调的方式耦合相应的块半径。正如第5节中的数值研究所示，即使子空间数量 m 较少，这种方法也能显著改进上界 U_{Wong} 。一般而言，以下式（50）中的改进上界 U_K 可表示为一个 m维积分；当 m较大时，其数值计算可能颇具挑战性。然而，对于某些径向族（包括多元 Student-t 分布），无论环境维数 d 或组数 m如何，该界都可化简为若干一维积分的组合。

现在，为了构造改进的上界，我们将 K 的特征值（按重数计并按降序排列）记为 $\lambda_1^{\downarrow} \ge \cdots \ge \lambda_d^{\downarrow} > 0$。将前 q 个最大特征值之和记为

$$\Lambda_q := \sum_{j=1}^{q} \lambda_j^{\downarrow}, \qquad q \in \{1, \ldots, d\}, \qquad \Lambda_0 := 0. \tag{46}$$

设 $\mathbf{b} = (b_1, \ldots, b_m)$ 是 d 的一个有序合成，我们将其解释为式（45）中正交子空间的维数，即 $b_h = \dim(E_h) \in \{1, 2, \ldots\}$ 和 $\sum_{h=1}^{m} b_h = d$。类似于（30），考虑一个 m 维随机向量

$$(D_1, \ldots, D_m) \sim \mathrm{Dirichlet}\left(\frac{b_1}{2}, \ldots, \frac{b_m}{2}\right),$$

其中 $D_1 = 1$ 当 $m = 1$时。对于与 $(D_1, \ldots, D_m)$ 独立的径向随机变量 $R_i \sim \nu_i$，考虑由块半径构成的随机向量，其定义为

$$Z_i = (Z_{i,1}, \ldots, Z_{i,m}) := (R_i\sqrt{D_1}, \ldots, R_i\sqrt{D_m}), \qquad i = 0, 1. \tag{47}$$

然后，我们递归地将 Z^0 和 Z^1 的*Knothe-Rosenblatt*耦合 定义为 Z_0^{KR} 和 Z_1^{KR} 的联合分布：

$$\begin{aligned} Z_{i,1}^{\mathrm{KR}} &:= F_{Z_{i,1}}^{-1}(U_1), \\ Z_{i,h}^{\mathrm{KR}} &:= F^{-1}_{Z_{i,h}\mid Z_{i,h-1}=Z_{i,h-1}^{\mathrm{KR}},\ldots,Z_{i,1}=Z_{i,1}^{\mathrm{KR}}}(U_h), \quad h = 2, \ldots, m, \end{aligned} \tag{48}$$

对于 $i \in \{0, 1\}$，其中 $U_1, \ldots, U_m \sim \mathcal{U}(0, 1)$ 相互独立。注意，根据构造，$Z_{0\,h}^{\mathrm{KR}}$和 $Z_{1\,h}^{\mathrm{KR}}$是给定 $U_1, \ldots, U_{h-1}$ 后条件共单调的 m 维随机向量。我们将 Knothe-Rosenblatt 耦合中第 h 个分量之间的二次代价记为

$$C_h^{\mathbf{b}} := \mathbb{E}(Z_{0,h}^{\mathrm{KR}} - Z_{1,h}^{\mathrm{KR}})^2. \tag{49}$$

将此代价乘以平均特征值，并对所有子空间求和，我们定义如下量：

$$U_K := \min_{\substack{m \in \{1,\ldots,d\} \\ \mathbf{b}=(b_1,\ldots,b_m)\in\mathbb{N}^m \\ b_1+\cdots+b_m=d}} U_{K,\mathbf{b}}, \qquad U_{K,\mathbf{b}} := \sum_{h=1}^{m} \frac{\Lambda_{a_h+b_h} - \Lambda_{a_h}}{b_h} C_h^{\mathbf{b}}, \qquad a_h := \sum_{\ell=1}^{h-1} b_\ell. \tag{50}$$

The 下述定理表明， U_K 是生成器效应的一个改进上界 ct.

定理 3.8（通过 Knothe–Rosenblatt 耦合改进上界）。令 $P_i = E_2(m_i、\ \Sigma_i、\ \nu_i)$ 、 $i = 0、1$，其中 $\nu_0、\ \nu_1 \in R_d$，并令 K 为相应的 *Bures*矩阵。令 Λ_q 在 （46） 中表示 K的 q个最大特征值之和，则生成元效应满足

$$\Gamma(P_0, P_1) \le U_K \le \frac{\operatorname{tr} K}{d}\delta_R^2. \tag{51}$$

注 3.9。一般而言，对于有序分拆 $\mathbf{b} =$（b_1、.....、b_m），计算 $U_{K、\,\mathbf{b}}$需要进行涉及最多 m 个变量的多维积分。这通常比在（32）中对 k维最优传输问题进行数值求解简单得多。此外，计算 U_K 通常需要考虑 d的所有 2^{d-1} 个有序分拆。然而，将最小化限制在适当选取的分拆上，仍可得到显著改进的上界。

4 通过动态规划快速计算 U_K

本节将介绍一种计算效率较高的方法，用于计算（50）中的条件共单调上界 U_K。因此，回顾其构造基于（48）中的 Knothe–Rosenblatt 耦合（Z_0^{KR}、Z_1^{KR}），以及在对应于按降序排列的 K 特征值的所有有序分拆 $\mathbf{b}$ 上，对 $U_{K、\,\mathbf{b}}$ 进行最小化，其中这些分拆由连续分块组成。

对于某些球对称分布，Knothe–Rosenblatt 耦合具有乘积表示

$$Z_{i,h}^{\mathrm{KR}} = \sqrt{c_i}\, g_{i;a_h,b_h}(U_h) \prod_{\ell<h} \sqrt{1 + \varepsilon_i g_{i;a_\ell,b_\ell}(U_\ell)^2}, \qquad i = 0, 1, \tag{52}$$

如下文定理 4.1 的证明中公式（84）所确立的那样。这一表示使我们能够将 U_K 的计算归约为至多 d^2 个一维积分。此外，动态规划将对 2^{d-1} 的所有有序分拆的优化归约为 $d(d+1)/2$ 次转移，从而得到一个复杂度为 $O(d^2)$ 的优化算法。

球面对称分布 η_ν 于 $\mathbb{R}^d$ ($\nu \in \mathcal{R}_d$)	c	ε	$Q_{a,b}$
高斯 $N(0,\ I_d)$	1	0	χ_b
Student-$t_\theta(0,\ I_d)$，协方差标准化，$\theta > 2$ 自由度	$\theta - 2$	1	$\sqrt{\dfrac{\chi_b^2}{\chi_{\theta+a}^2}}$
半径为 $\sqrt{d+2}$ 的球内均匀分布	$d+2$	-1	$\sqrt{\text{贝塔分布}\left(\frac{b}{2}, \frac{d-a-b}{2}+1\right)}$
半径为 $\sqrt{d}$ 的球面上的均匀分布	d	-1	$\sqrt{\text{贝塔分布}\left(\frac{b}{2}, \frac{d-a-b}{2}\right)}$

表 2：式 (47) 中分块半径向量的常数和条件分布 Q_{a_h,b_h} =分布律($Z_h/\alpha_h \mid Z_{1:h-1} = r$)，其中归一化为 α_h （见式 (75)）。对于上述球对称分布，这些条件分布与 r 独立。

我们为高斯分布和 Student-t 分布建立了这一表示，同时也为球上的均匀分布和球面上的均匀分布建立了这一表示，其中包括这四个分布族之间的任意配对。回顾（47）中的分块半径向量 $Z_i = (Z_{i,\ 1}$、.....、$Z_{i,\ m})$。式（52）中的函数 $g_{i,\ a_h,\ b_h} = F^{-1}_{Q_i\ ,\ a_h,\ b_h}$ 是给定 $Z_{i,\ 1}$、.....、$Z_{i,\ h-1}$ 时第 h 个分块半径 $Z_{i,\ h}$ 的标准化条件分位数函数；详情请参见引理 B.4。相应的分布 $Q_{i;a_h,\ b_h}$ 以及各分布对应的常数 c_i 和 ε_i 列于表 2 中。

定理 4.1 (快速计算 U_K)。 对于 $i = 0,1$，令 $P_i = E_2$<style id='(0, Σ_i, ν_i)，其中 Σ_i 为正定矩阵，η_{ν_i} 为属于表 2 中四个族之一的球面对称分布。对于满足 $a + b \le d$ 的整数 $a \ge 0$ 和 $b \ge 1$，定义一维积分

$$I_{a,b} := \int_0^1 g_{0;a,b}(u) g_{1;a,b}(u)\,\mathrm{d}u, \qquad J_{a,b} := \int_0^1 \sqrt{1+\varepsilon_0\, g_{0;a,b}(u)^2}\,\sqrt{1+\varepsilon_1\, g_{1;a,b}(u)^2}\,\mathrm{d}u.$$

其中 ε_i 为常数，$(0,\ 1) \ni u \mapsto g_{i;a,b}(u) := F^{-1}_{Q_{i;a,b}}(u)$ 是表 2 中指定的条件分块径向分布 $Q_{i;a,b}$ 的广义逆分布函数。对于 $w_{a,b} := (\Lambda_{a+b} - \Lambda_a)/b$，递归定义

$$B_d := 0 \qquad \textit{and} \qquad B_a := \max_{1 \le b \le d-a} \{w_{a,b} I_{a,b} + J_{a,b} B_{a+b}\}, \quad a = d-1, \ldots, 0. \tag{53}$$

于是，我们有

$$U_K = 2\,\mathrm{tr}\,K - 2\sqrt{c_0 c_1}\, B_0,$$

其中，相应的常数 c_i 见表 2。

注 4.2. (a) 式 (53) 中的最大化确定了每个连续块的最优维数。特别地，最优块不一定是一维的，也可以具有不同的维数。例如，若 $K = \lambda I_d$，单块划分 $\mathbf{b} = (d)$ 是最优的；而各向异性谱则可能更适合更细划分。

(b) 根据定理 4.1，计算 U_K 至多需要 d^2 个一维积分，即 $d(d+1)/2$ 个积分 $I_{a,\ b}$ 以及 $d(d-1)/2$ 个积分 $J_{a,\ b}$，随后还需为动态规划 (53) 执行 $O(d^2)$ 次算术运算。

(c) 对于固定的径向分布和维度 d，一维积分 $I_{a,\ b}$ 和 $J_{a,\ b}$ 与协方差矩阵无关。因此，可以预先计算这些积分，并在不同的协方差矩阵中重复使用；只需更新谱权重 $w_{a,\ b}$ 和动态规划递推即可。

Bures矩阵	特征值 $\lambda_j^{\downarrow}(K)$	不同层级 k
K_2	$3d$ ($d/5$ 次)， $d/2(4d/5$ 次)	2
K_5	$2d$， $3d/2$， d， $d/3$， $d/6$，各 $d/5$次	5
K_{ar}	$\dfrac{2d(d-j+1)}{d+1}$， $j=1,\dots,d$	d
K_{pow}	$\dfrac{d^2 j^{-2}}{\sum_{\ell=1}^{d}\ell^{-2}}$， $j=1,\dots,d$	d

表 3：谱模型，其中 tr $K=d^2$ 且对于 $d=2$，我们设 $K_2=\mathrm{diag}(3,1)$。特征值按降序排列，并按重数计数。

5 数值实验

本节将最小特征值下界 $L_{\min}$、嵌套投影下界 L_K 和条件共单调上界 U_K 与基本的 Gelbrich–Wong 界进行比较。对于选定的设置，我们通过求解约化最优传输问题 (32) 的离散化版本，给出生成元效应 $\Gamma(P_0,\ P_1)$ 的数值近似，前提是结果在网格细化下表现出足够的稳定性。

5.1 设置

所有报告的数值均针对生成元效应 $\Gamma(P_0,\ P_1)$，它仅取决于 Bures矩阵 K 的谱以及径向分布 ν_0 和 ν_1。因此，不失一般性，我们可以选取 w.l.o.g. $m_0=m_1=0$ 和 $\Sigma_0=\Sigma_1=K$。相应的 Bures矩阵为 K，且 $W_2^2(P_0,\ P_1)=\Gamma(P_0,\ P_1)$。$K$ 的谱模型列于表 3，并采用共同的归一化

$$\operatorname{tr} K = d^2. \tag{54}$$

缩放 K 会作用于输运代价和界，但不会改变最优块组成或改进百分比。在 (54) 的条件下，初等界为

$$L_{\mathrm{Gelbrich}}=0, \qquad L_{\min}=\lambda_{\min}(K)\delta_R^2, \qquad U_{\mathrm{Wong}}=d\,\delta_R^2, \qquad \delta_R^2=W_2^2(\nu_0,\nu_1).$$

表 3 中的模型 K_2 和 K_5 分别有 2 个和 5 个互异特征值。算术谱 K_{ar} 呈线性递减，而功率谱 K_{pow} 则赋予其主导方向大得多的权重。具体而言，环境维数 d 与约化输运问题的维数 k 可能相差甚远。

对于 P_i，我们考虑高斯分布（N）、自由度为 3 和 5 的 Student-t 分布（t_3 和 t_5）、半径为 $\sqrt{d+2}$ 的球内均匀分布（Ball），以及半径为 $\sqrt{d}$ 的球面均匀分布（Sphere）。它们的径向变量可表示为

$$R_N=\sqrt{H_d}, \qquad R_{t_\theta}=\sqrt{\frac{(\theta-2)H_d}{V_\theta}}, \quad \theta\in\{3,5\},$$
$$R_{\mathrm{Ball}}=\sqrt{d+2}\,W^{1/d}, \qquad R_{\mathrm{Sphere}}=\sqrt{d},$$

其中，$H_d\sim\chi^2_d$、$V_\theta\sim\chi^2_\theta$ 和 $W\sim\mathcal{U}(0,\ 1)$ 相互独立。每个径向分布都属于 $\mathcal{R}_d$，因此相应的径向变量 R 满足 $\mathbb{E}[R^2]=d$。

第 1 节中的表 1 给出了正态分布与 Student-t_3分布之间生成元效应的界。表 4 通过补充 N–球面、球面–球 和 t_5–t_3 这些配对，完善了引言。我们考虑维度为 $d\in\{2,10,100,10^6\}$ 的小维度、中等维度和高维情形。

在两张表中，“改进”量化了我们的新界对经典 Gelbrich–Wong 区间所实现的缩减：

$$\text{Improvement} := 1-\frac{U_K-L_K}{U_{\mathrm{Wong}}-L_{\mathrm{Gelbrich}}}=1-\frac{U_K-L_K}{d\,\delta_R^2}. \tag{55}$$

界的评估。 我们使用一维求积直接计算（44）中的下界 L_K 。对于上界 U_K，我们应用定理 4.1；该定理将所需的 KR 代价表示为一维积分，并通过动态规划实现对有序块组合的优化。

在高维情况下，我们仅针对选定维度计算投影代价，并将上界的优化限制在有序块分解的一个子集中。以下引理使我们能够控制由此得到的下界近似。

引理 5.1（归一化投影代价的单调性）。 设 $\nu_0, \nu_1 \in R_d$，并令 Δ_q, $q = 1, \dots, d$ 表示式 （40）中定义的平方运输代价。则有

$$\frac{\Delta_q}{q} \le \frac{\Delta_r}{r} \qquad for\ 1 \le q \le r \le d. \tag{56}$$

对于较大的 d，我们仅在选定维度 Δ_q $1 = q_1 < \cdots < q_s = d$ 上计算投影代价。随后定义

$$L^- := \omega_d \Delta_d + \sum_{j=1}^{s-1} \frac{\Delta_{q_j}}{q_j} \sum_{q=q_j}^{q_{j+1}-1} q\omega_q, \tag{57}$$

其中 $\omega_q = \lambda_q^{\downarrow}(K) - \lambda_{q+1}^{\downarrow}(K)$，且 $\lambda_{d+1}^{\downarrow}(K) = 0$。于是，根据引理 5.1 可得 $L^- \le L_K$。

为了在高维情况下对上界进行保守近似，我们考虑子集$\mathcal{A} \subseteq \{0, \dots, d\}$，该子集包含 0 和 d。然后，将式（53）中的动态规划限制为 $\mathcal{A}$ 中索引之间的转移，可得 U^+。因此，

$$L^- \le L_K \le \Gamma(P_0, P_1) \le U_K \le U^+ \le U_{\mathrm{Wong}}. \tag{58}$$

对于 $d = 10^6$，我们使用 3663 个选定的投影维度近似下界，并将上界优化限制在 169 个块边界上，从而得到 14 196 种可能的转移。表中的匕首符号表示 L^- 或 U^+。

5.2 数值结果讨论

表 1 展示了我们的界对正态分布和 Student-t_3 分布的有效性。在维度 2 中，改进已经相当显著；对于 K_2 和 K_{ar}，在维度 10^6 中改进幅度超过 99.95%。在这些示例中，嵌套投影下界 L_K 捕捉了径向差异的大部分，而 KR 上界 U_K 则带来了进一步的改进，尤其是在各向异性谱的情况下。

表 4 将比较扩展到了其他径向分布律。当 $\Gamma(P_0, P_1)$ 在网格细化下足够稳定时，报告其数值近似值。短横线表示未获得可靠的近似值；某些数值运输耦合甚至产生了超过 U_K 的成本。

低维情形

对于 $d = 2$， t_5–t_3 的界尤其紧致，改进幅度超过 97%。对于其他径向对，降幅则不太明显。在若干情况下， $U_K = U_{\mathrm{Wong}}$，因此改进完全源于下界。

在 $d = 10$，谱结构的影响变得更加显著。对于球–球，在 K_2 下，改进幅度接近 88%；而在 K_{ar} 下，则低于 29%。因此，界的有效性不仅取决于径向分布，还取决于其差异与 K 的谱之间的相互作用。

中等维度

在 $d = 100$ 处， t_5–t_3 的界变得更加紧致，三个谱的改进均超过 98%。相比之下，对于 K_{ar} 下的球–球，改进降至 8% 以下。在这里，基于投影的下界仍然相对较弱，而 KR 上界接近 Wong界。因此，维度增加并不一定会使我们的界更加紧致。

d 径向对	Bures 矩阵	k	$L_{\min}$	L_K	$\Gamma(P_0,\ P_1)$	U_K	U_{Wong}	改进
2 N–球面	K_2	2	0.4551	0.6614	≈ 0.77	0.8727	0.9102	76.78%
2 N–球面	K_{ar}	2	0.6068	0.7443	≈ 0.85	0.9102	0.9102	81.78%
2 N–球面	K_{pow}	2	0.3641	0.6116	≈ 0.72	0.7807	0.9102	81.42%
2 球–球 K_2		2	0.2288	0.2905	≈ 0.36	0.4227	0.4575	71.09%
2 球–球 K_{ar}		2	0.305	0.3461	≈ 0.41	0.4575	0.4575	75.66%
2 球–球 K_{pow}		2	0.183	0.257	≈ 0.32	0.3629	0.4575	76.87%
2 t_5–t_3	K_2	2	0.1398	0.2716	–	0.2795	0.2795	97.16%
2 t_5–t_3	K_{ar}	2	0.1863	0.2742	–	0.2795	0.2795	98.11%
2 t_5–t_3	K_{pow}	2	0.1118	0.27	–	0.2772	0.2795	97。43%
10 N–球面	K_2	2	2.465	2.7411	≈ 3.17	3.2666	4.93	89。34%
10 N–球面	K_5	5	0.8217	1.9781	–	3.2228	4.93	74.75%
10 N–球面	K_{ar}	10	0.8964	2.4005	–	4.3512	4.93	60.43%
10 球–球 K_2		2	0.4141	0.4225	≈ 0.50	0.522	0.8282	87。98%
10 球–球 K_5		5	0.138	0.1897	–	0.5064	0.8282	61.76%
10 球–球 K_{ar}		10	0.1506	0.2376	–	0.8282	0.8282	28.69%
10 t_5–t_3	K_2	2	3.9601	7.454	–	7.7978	7.9201	95.66%
10 t_5–t_3	K_5	5	1.32	7.6142	–	7.8022	7.9201	97.63%
10 t_5–t_3	K_{ar}	10	1.44	7.6856	–	7.8794	7.9201	97.55%
100 N–球面	K_2	2	24.9684	26.501	≈ 30.9	31.782	49.9367	89.42%
100 N–球面	K_5	5	8.3228	17.2394	–	30.3787	49.9367	73.69%
100 N–球面	K_{ar}	100	0.9888	17.5728	–	35.285	49.9367	64。53%
100 球–球 K_2		2	0.4902	0.4908	≈ 0.59	0.6129	0.9803	87.55%
100 球–球 K_5		5	0.1634	0.1697	–	0.6101	0.9803	55.07%
100 球–球 K_{ar}		100	0.0194	0.0492	–	0.9515	0.9803	7.97%
100 t_5–t_3	K_2	2	419.9412	827.9996	–	839.3492	839.8823	98.65%
100 t_5–t_3	K_5	5	139.9804	833.67833	–	837.75444	839.8823	99.51%
100 t_5–t_3	K_{ar}	100	16.6313	834.19902	–	838.82535	839.8823	99.45%
10^6 N–球面	K_5	5	83 333.3229	169 315.7591	–	$301\,316.9529^{\dagger}$	499 999.9375	73。60%
10^6 N–球面	K_{ar}	10^6	1	$166\,397.7161^{\dagger}$	–	$328\,483.2623^{\dagger}$	499 999.9375	67。58%
10^6 N–球面	K_{pow}	10^6	0.304	$2.6728^{\dagger}$	–	$4.761^{\dagger}$	499 999.9375	99.999582%
10^6 球–球 K_5		5	0.1667	0.1667	–	$0.625^{\dagger}$	1	54.17%
10^6 球–球 K_{ar}		10^6	2.0000×10^{-6}	$1.4168 \times 10^{-5\,\dagger}$	–	$0.9635^{\dagger}$	1	3.65%
10^6 球–球 K_{pow}		10^6	6.0793×10^{-7}	$6.0794 \times 10^{-7\,\dagger}$	–	$2.4317 \times 10^{-6\,\dagger}$	1	99.999818%
10^6 t_5–t_3	K_5	5	1.4110×10^{10}	$8.46604524 \times 10^{10}$	–	$8.46604984 \times 10^{10\,\dagger}$	8.4661×10^{10}	99.999946%
10^6 t_5–t_3	K_{ar}	10^6	169 320.8734	$8.46604522 \times 10^{10\,\dagger}$	–	$8.46605095 \times 10^{10\,\dagger}$	8.4661×10^{10}	99.999932%
10^6 t_5–t_3	K_{pow}	10^6	51 467.4567	$7.0736 \times 10^{10\,\dagger}$	–	$7.4540 \times 10^{10\,\dagger}$	8.4661×10^{10}	95。51%

表4：表 3 中所选径向对和谱的界及最优传输数值近似。 Γ（P_0，P_1）列中的数值是生成元效应的数值估计。所有矩阵的迹均为d^2。匕首符号（†）表示对于 $d = 10^6$，L^- 或 U^+；短横线（–）表示未报告足够稳定的最优传输近似。改进指 Gelbrich–Wong 区间宽度的缩减；见（55）。

高维

在 $d = 10^6$ 下，这些对比趋势变得更加明显。对于 t_5–t_3，在 K_5 和 K_{ar} 下，保守改进超过 99。9999 %，而在 K_{ar} 下，球–球仅实现了 3。65% 的改进。相比之下，高度集中的功率谱使 N–球面和球–球的改进均超过 99。9995%。这说明谱各向异性对界的质量具有显著影响。然而，Gelbrich–Wong 区间几乎完全缩减，并不一定意味着生成元效应中的相对误差同样很小。

渐近尺度。传输代价的不同量级可以通过径向分布的渐近行为来解释。由于 $d \to \infty$，我们有

$$\delta_R^2 = \begin{cases} \Theta(d), & N\text{–}t_3,\quad t_5\text{–}t_3, \\ \Theta(1), & N\text{–Sphere},\quad N\text{–Ball}, \\ \Theta(d^{-1}), & \text{Sphere–Ball}. \end{cases}$$

由于 $\operatorname{tr} K = d^2$，Wong 界满足 $U_{\mathrm{Wong}} = d\,\delta_R^2$，因此相应地其阶为 d^2、d 或 1。对于 K_2 和 K_5，最小特征值的阶为 d，因此最小特征值下界表明，$\Gamma(P_0,\ P_1)$ 具有相同的渐近阶。对于 Student 配对和 N–球面，在 K_{ar} 下也成立相同的阶。然而，对于强各向异性的功率谱，传输代价可能显著更小，N–球面和球面–球的比较说明了这一点。因此，生成元效应的大小取决于 K 的径向差异和谱结构。

计算时间。 包括求积检验在内，计算 $d=2$ 约耗时 0.35 秒， $d=10$ 约耗时 0.86 秒， $d=100$ 约耗时 7.35 秒，而在 $d=10^6$ 处进行受限计算耗时 113 秒。R 实现为 $d=2$，10 使用一个工作进程，在更高维度下使用八个工作进程。报告的计算时间包括全部五个径向对，其中包括 N–球；由于其与 N–球面的相似性，后者未列入表格。

A 辅助结果

A.1 椭圆分布及其径向分布

关于椭圆分布的一般背景，请参见 [5, 11]。本文考虑具有有限二阶矩的椭圆分布。由类$\mathcal{R}_d$ 在 (12) 中对径向变量进行归一化 $\mathbb{E}[R^2]=d$，使我们能够将矩阵 Σ 识别为 $X=\mu+RAU$的协方差矩阵， $AA^\top=\Sigma$, 如下所示。

引理 A.1 （标准化半径的归一化与唯一性）。 设 $X\sim\mathcal{E}_2(m,\Sigma,\nu)$ 满足$\Sigma\in\mathbb{S}_{++}^d$ 且 $\nu\in\mathcal{R}_d$。则

$$\mathbb{E}[X]=m,\qquad \mathrm{Cov}(X)=\Sigma,\qquad and\qquad \left\|\Sigma^{-1/2}(X-m)\right\|\overset{\mathrm{d}}{=}R,\quad for\ R\sim\nu. \tag{59}$$

标准化径向律 ν 由 X 的分布以及(m,Σ)唯一确定。此外，(13) 中的分布律与所选取的满足 $AA^{\mathsf{T}}=\Sigma$的因子 A 无关。

证明。 由于 U是中心化的且 $\mathbb{E}[UU^{\mathsf{T}}]=I_d/d$，独立性和 $\mathbb{E}[R^2]=d$给出

$$\mathbb{E}[ARU]=0,\qquad \mathrm{Cov}(ARU)=A\mathbb{E}[R^2UU^{\mathsf{T}}]A^{\mathsf{T}}=AA^{\mathsf{T}}=\Sigma.$$

令 A 为满足 $AA^{\mathsf{T}}=\Sigma$的任意非奇异因子。则 $Q:=\Sigma^{-1/2}A$ 满足 $QQ^{\mathsf{T}}=I_d$，因此是正交的。于是

$$\left\|\Sigma^{-1/2}(X-m)\right\|=\|QRU\|=R\quad \text{a.s.},$$

这就证明了（59）。如果 $A_1A_1^{\mathsf{T}}=A_2A_2^{\mathsf{T}}=\Sigma$，那么对于正交矩阵 Q，有 $A_2=A_1Q$。由于 $QU\overset{\mathrm{d}}{=}U$，这两个因子生成相同的分布律。 □

A.2 狄利克雷分布

对于参数 $\alpha_1,\dots,\alpha_k>0$， k维随机向量 $D=(D_1,\dots,D_k)$ 服从狄利克雷分布 Dirichlet$(\alpha_1,\dots,\alpha_k)$，当且仅当其支撑集为单纯形

$$\Delta_{k-1}:=\left\{d\in[0,1]^k:\sum_{j=1}^k d_j=1\right\}$$

且其密度为

$$f_D(d_1,\dots,d_k)=\frac{\Gamma(\alpha_1+\cdots+\alpha_k)}{\prod_{j=1}^k\Gamma(\alpha_j)}\prod_{j=1}^k d_j^{\alpha_j-1}$$

相对于前 $k-1$ 个坐标上的勒贝格测度，其中 $d_k=1-\sum_{j=1}^{k-1}d_j$。等价地，如果$\{Y_j\}_{j=1,\dots,k}$ 是具有 $2\alpha_j$ 自由度的独立卡方随机变量，即 $Y_j\sim\chi^2_{2\alpha_j}$,$j=1,\dots,k$，那么

$$D_j=\frac{Y_j}{Y_1+\cdots+Y_k},\qquad j=1,\dots,k, \tag{60}$$

定义一个随机向量 $D=(D_1,\ .....,\ D_k)\sim$狄利克雷$(\alpha_1,\ .....,\ \alpha_k)$；例如参见 [20, 第 49] 章。特别地，其边缘分布为

$$D_j\sim\mathrm{Beta}\left(\alpha_j,\sum_{\ell\neq j}\alpha_\ell\right).$$

对于 $k=1$，我们采用几乎处处成立的约定 $D_1=1$。此外，贝塔分布(α，0) 是集中于 1 的狄拉克测度。(60) 中的表示将在引理 B.1 中与 $\alpha_j=n_j/2$ 一起应用。我们还利用

$$\frac{X}{X+Y}\sim \mathrm{Beta}\left(\frac{m}{2},\frac{n}{2}\right) \qquad \text{and} \qquad \frac{X}{X+Y}\perp (X+Y) \tag{61}$$

对于独立的随机变量 $X\sim\chi^2_m$ 和 $Y\sim\chi^2_n$。

A.3 Bures 对齐与 Procrustes 因子

对于两个矩阵 C，$D\in\mathbb{R}^{d\times d}$，正交 *Procrustes* 问题要求找到一个正交矩阵 $Q\in O(d)$，使下式最小

$$\|C-DQ\|_{\mathrm{F}}^2 . \tag{62}$$

这里，$\|\cdot\|_{\mathrm{F}}$ 表示 Frobenius 范数，$\|A\|_{\mathrm{F}}:=\left(\sum_{i,j=1}^d A_{ij}^2\right)^{1/2}=\sqrt{\mathrm{tr}(A^{\mathsf{T}}A)}$。~~这个~~问题与椭圆分布的协方差矩阵之间的最优传输尤其相关，因为协方差矩阵的矩阵因子并不唯一。如果 $DD^{\mathsf{T}}=\Sigma$，那么，对于每个 $Q\in O(d)$，

$$(DQ)(DQ)^{\mathsf{T}}=DQQ^{\mathsf{T}}D^{\mathsf{T}}=DD^{\mathsf{T}}=\Sigma.$$

因此，右乘一个正交矩阵会改变因子 D，但不会改变由其生成的正定矩阵。Procrustes 问题从同一矩阵的所有因子 DQ 中选出一个与 C 在 Frobenius 范数意义下距离最近的因子。此外，

$$\|C-DQ\|_{\mathrm{F}}^2=\mathrm{tr}\big((C-DQ)^{\mathsf{T}}(C-DQ)\big)=\mathrm{tr}(CC^{\mathsf{T}})+\mathrm{tr}(DD^{\mathsf{T}})-2\,\mathrm{tr}(C^{\mathsf{T}}DQ).$$

因此，最小化 (62) 等价于在 $Q\in O(d)$ 上最大化 $\mathrm{tr}(C^{\mathsf{T}}DQ)$。对于协方差因子$C=\Sigma_0^{1/2}$ 和 $D=\Sigma_1^{1/2}$，经典的 Procrustes 恒等式给出

$$\max_{Q\in O(d)}\mathrm{tr}\Big(\Sigma_0^{1/2}\Sigma_1^{1/2}Q\Big)=\mathrm{tr}\left[\left(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2}\right)^{1/2}\right];$$

例如，参见 [2, 定理 1]。因此，

$$\min_{Q\in O(d)}\left\|\Sigma_0^{1/2}-\Sigma_1^{1/2}Q\right\|_{\mathrm{F}}^2=d_{\mathrm{BW}}^2(\Sigma_0,\Sigma_1). \tag{63}$$

因此，平方 Bures–Wasserstein 距离是协方差因子经过最优正交对齐后的最小平方 Frobenius 距离。

对于 $\Sigma_0,\Sigma_1\in\mathbb{S}^d_{++}$，回顾 (7) 中给出的 Bures矩阵

$$K=\left(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2}\right)^{1/2}$$

下列 引理给出了最优对齐的协方差因子，这些因子在 th e 分解 输运代价。

引理 A.2（Bures 对齐的协方差因子）。设 $A:=\Sigma_0^{1/2}$ 且 $B:=\Sigma_0^{-1/2}K$。则

$$AA^{\mathsf{T}}=\Sigma_0,\qquad BB^{\mathsf{T}}=\Sigma_1,\qquad A^{\mathsf{T}}B=K. \tag{64}$$

此外，矩阵 $Q^*:=\Sigma_1^{-1/2}\Sigma_0^{-1/2}K$ 是正交的，并且是正交 *Procrustes* 问题的优化器。特别地，

$$B=\Sigma_1^{1/2}Q^* \qquad and \qquad \|A-B\|_{\mathrm{F}}^2=d_{\mathrm{BW}}^2(\Sigma_0,\Sigma_1)=\min_{Q\in O(d)}\left\|\Sigma_0^{1/2}-\Sigma_1^{1/2}Q\right\|_{\mathrm{F}}^2. \tag{65}$$

证明。(64) 中的等式是显然的。此外，Q^* 是正交的，因为

$$Q^*(Q^*)^{\mathsf{T}}=\Sigma_1^{-1/2}\Sigma_0^{-1/2}K^2\Sigma_0^{-1/2}\Sigma_1^{-1/2}=I_d.$$

此外，$\Sigma_1^{1/2}Q^*=\Sigma_0^{-1/2}K=B$，这证明了 (65) 的第一部分。最后，

$$\|A-B\|_{\mathrm{F}}^2=\mathrm{tr}(AA^{\mathsf{T}})+\mathrm{tr}(BB^{\mathsf{T}})-2\,\mathrm{tr}(A^{\mathsf{T}}B)=\mathrm{tr}(\Sigma_0)+\mathrm{tr}(\Sigma_1)-2\,\mathrm{tr}(K)=d_{\mathrm{BW}}^2(\Sigma_0,\Sigma_1),$$

其中，最后一个等式根据式 (3) 中 Bures-Wasserstein 距离的定义以及 K 得出。式 (65) 中的最后一个等式由式 (63) 中的 Procrustes 恒等式得出。 □

B 证明

B.1 第 2 节的证明

命题 *2.3* 的证明。 根据引理 A.2，我们有 $AA^{\mathsf{T}} = \Sigma_0$， $BB^{\mathsf{T}} = \Sigma_1$，以及 $A^{\mathsf{T}}B = K$. 由于 A 和 B 可逆， P_0 与 P_1 的耦合通过以下方式与 η_{ν_0} 与 η_{ν_1} 的耦合一一对应

$$X_0 = m_0 + AZ_0, \qquad X_1 = m_1 + BZ_1.$$

对于每一个这样的耦合，由（18）和（20）可得

$$\mathbb{E}\,\|X_0 - X_1\|^2 = \|m_0 - m_1\|^2 + d_{\mathrm{BW}}^2(\Sigma_0, \Sigma_1) + \mathbb{E}\big[(Z_0 - Z_1)^{\mathsf{T}} K (Z_0 - Z_1)\big].$$

因此，对 X_0 和 X_1 的所有耦合取下确界可得

$$W_2^2(P_0, P_1) = \|m_0 - m_1\|^2 + d_{\mathrm{BW}}^2(\Sigma_0, \Sigma_1) + W_{2,K}^2(\eta_{\nu_0}, \eta_{\nu_1}),$$

这就证明了（22）。

最后，通过线性变量变换 $z \mapsto K^{1/2} z$,

$$W_{2,K}^2(\eta_{\nu_0}, \eta_{\nu_1}) = W_2^2\big((K^{1/2})_{\#}\eta_{\nu_0}, (K^{1/2})_{\#}\eta_{\nu_1}\big) = W_2^2(\mathcal{E}_2(0, K, \nu_0), \mathcal{E}_2(0, K, \nu_1)),$$

其中，第二个等式中我们使用了 (13) 中椭圆分布的定义 $K^{1/2}(K^{1/2})^{\mathsf{T}} = K$ 。由此证明了 (23)。

□

为证明命题 2.5，我们将用到以下球面分解。回顾一下，$n_j = \dim(E_j)$ 是与 Bures矩阵 K 相关的两两正交特征子空间 $E_1, \ldots, E_k$ 的维数。此外，π_j 是到 K 的第 j 个特征子空间 E_j 的正交投影，即 $K = \sum_j \kappa_j \pi_j$，其中 κ_j 是 E_j 的特征值。

引理 B.1 (球形分块分解)。 设 U 在 $\mathbb{S}^{d-1}$ 上均匀分布。对于 $j = 1, \ldots, k$，令 $D_j := \|\pi_j U\|^2$，且 $V_j := \pi_j U / \|\pi_j U\|$。则

(i) $(D_1, \ldots, D_k) \sim \mathrm{Dirichlet}\big(\frac{n_1}{2}, \ldots, \frac{n_k}{2}\big)$,

(ii) V_j 在 E_j 的单位球面上均匀分布，

(iii) V_1，……， V_k 相互独立，

(iv) $(V_1$，……， $V_k)$ 独立于$(D_1$，……， $D_k)$。

证明。 对每个 $j \in \{1$，……， $k\}$，选取 E_j的一个标准正交基。由于这些特征子空间两两正交，并且

$$\mathbb{R}^d = E_1 \oplus \cdots \oplus E_k,$$

这些基的并构成 $\mathbb{R}^d$的一个标准正交基。设 $G \sim N(0,\ I_d)$为具有独立分量的标准高斯随机向量。则 $U := G/\|G\|$ 在 $\mathbb{S}^{d-1}$上均匀分布。此外，对于$G_j := \pi_j G$，我们有

$$\|G_j\|^2 \sim \chi^2_{n_j}, \qquad j = 1, \ldots, k,$$

独立地。由于 $D_j = \|\pi_j U\|^2 = \|G_j\|^2 / (\sum_{\ell=1}^k \|G_\ell\|^2)$，含 $\alpha_j = n_j/2$ 的表示(60)给出

$$(D_1, \ldots, D_k) \sim \mathrm{Dirichlet}\Big(\frac{n_1}{2}, \ldots, \frac{n_k}{2}\Big).$$

这就证明了(i)。

对于每个 j，令 $R_j := \|G_j\|$ 和 $V_j := \frac{G_j}{\|G_j\|}$，其中 V_j 在零概率事件$\{G_j = 0\}$上可任意定义。由于 G_j 是 E_j上的标准高斯向量，其分布在 E_j的正交变换下保持不变。因此， V_j 在 E_j的单位球面上服从均匀分布，并且 V_j 与 R_j相互独立。这就证明了 (ii)。

由于随机向量 G_1、.....、 G_k 相互独立，因此这些向量对也相互独立

$$(R_1, V_1), \ldots, (R_k, V_k) \quad \text{are independent.} \tag{66}$$

特别地， $V_1, \ldots, V_k$ 相互独立，这就证明了 (iii)。

此外，由 (66) 以及对于每个 j， R_j 与 V_j 独立，可得

$$\operatorname{law}\big((R_1, V_1), \ldots, (R_k, V_k)\big) = \bigotimes_{j=1}^{k} \big(\operatorname{law}(R_j) \otimes \operatorname{law}(V_j)\big).$$

这给出 分布律($R_1, \ldots, R_k, V_1, \ldots, V_k$) = 分布律($R_1, \ldots, R_k$) $\otimes$ 分布律($V_1, \ldots, V_k$) 从而

$$(V_1, \ldots, V_k) \perp (R_1, \ldots, R_k).$$

最后，由于 $D_j = R_j^2/(R_1^2 + \cdots + R_k^2)$ 是 $(R_1, \ldots, R_k)$ 的函数，我们得到

$$(V_1, \ldots, V_k) \perp (D_1, \ldots, D_k),$$

这就证明了 (iv)。 □

在下面的引理中，我们确定分块半径向量的分布

$$p_K(Z) = \big(\|\pi_1 Z\|, \ldots, \|\pi_k Z\|\big)$$

与 K 的谱分解相关的式 (28) 中。回顾 $K = \sum_{j=1}^{k} \kappa_j \pi_j$，其中 π_j 是到 K 的特征子空间 E_j 上的正交投影。因此， $p_K(z) = \big(\|\pi_1 z\|$、.....、$\|\pi_k z\|\big)$ 记录了 z 在 K 的各个不同特征子空间中的分量长度。对于球面对称随机向量，其分布仅通过特征子空间维数 $n_j = \dim(E_j)$ 取决于 K 。

引理 B.2 （分块半径分布）。 对于 $Z = RU \sim \eta_\nu$， 式 (28) 中的分块半径向量满足

$$p_K(Z) \overset{\mathrm{d}}{=} \big(R\sqrt{D_1}, \ldots, R\sqrt{D_k}\big), \tag{67}$$

其中 （D_1、.....、 D_k） $\sim$狄利克雷$\big(\frac{n_1}{2}$、.....、 $\frac{n_k}{2}\big)$ 与 R 独立。

证明。 定义 $D_j := \|\pi_j U\|^2$。于是，由于 $Z = RU$，对于 $j = 1$、.....、 k，有 $\|\pi_j Z\| = R\|\pi_j U\| = R\sqrt{D_j}$。由于 R 和 U 独立，因此 (D_1, ...、 D_k) 与 R 独立，且其分布由引理 B.1 给出。

□

命题 *2.5* 的证明。 由引理 B.2 可知，对于 $i = 0,1$，有 $\beta_i = (p_K)_{\#}\eta_{\nu_i}$。因此， η_{ν_0}和 η_{ν_1} 的每一个耦合都会诱导出 β_0 和 β_1的一个耦合。因此，(29) 中的不等式给出

$$W_{2,K}^2(\eta_{\nu_0}, \eta_{\nu_1}) \geq \mathcal{W}_{2,\boldsymbol{\kappa}}^2(\beta_0, \beta_1).$$

反向不等式方面，令 $\gamma^* \in \Pi(\beta_0, \beta_1)$ 为最优耦合，并考虑 (31) 中构造的耦合 (Z_0^*, Z_1^*)。由引理 B.1 以及方向 V_j 与块半径相互独立可知，

$$Z_i^* \sim \eta_{\nu_i}, \qquad i = 0, 1.$$

此外，由于两个向量在每个特征子空间内共享相同的单位方向，

$$(Z_0^* - Z_1^*)^{\mathsf{T}} K (Z_0^* - Z_1^*) = \sum_{j=1}^{k} \kappa_j (r_j - s_j)^2.$$

因此，

$$W_{2,K}^2(\eta_{\nu_0}, \eta_{\nu_1}) \leq \mathbb{E}[(Z_0^* - Z_1^*)^{\mathsf{T}} K (Z_0^* - Z_1^*)] = \mathcal{W}_{2,\boldsymbol{\kappa}}^2(\beta_0, \beta_1).$$

结合这两个不等式即可完成证明。 □

B.2 第 3 节的证明

定理 *3.1* 的证明。 由命题 2.3 和命题 2.5 可得

$$W_2^2(P_0,P_1) = \|m_0-m_1\|^2 + d_{\mathrm{BW}}^2(\Sigma_0,\Sigma_1) + \mathcal{W}_{2,\boldsymbol{\kappa}}^2(\beta_0,\beta_1),$$

这就证明了（34）。

现在令 $\gamma^* \in \Pi(\beta_0,\ \beta_1)$ 为最优解。然后按照（31）由 γ^* 构造 $(Z_0^*,\ Z_1^*)$。根据引理 B.1，提升后的变量满足 $Z_0^* \sim \eta_{\nu_0}$ 和 $Z_1^* \sim \eta_{\nu_1}$。此外，由于 γ^* 是约化问题的最优解，且（31）中的提升构造在（29）中达到等号，因此我们得到

$$\mathbb{E}[(Z_0^*-Z_1^*)^{\mathsf{T}} K (Z_0^*-Z_1^*)] = \mathcal{W}_{2,\boldsymbol{\kappa}}^2(\beta_0,\beta_1) = W_{2,K}^2(\eta_{\nu_0},\eta_{\nu_1}), \tag{68}$$

其中，最后一个等式由（32）得出。

根据引理 A.2，$A_0 = \Sigma_0^{1/2}$ 和 $A_1 = \Sigma_0^{-1/2} K$ 分别是 Σ_0 和 Σ_1 的协方差因子。然后 ce

$$X_0 = m_0 + A_0 Z_0^*, \qquad X_1 = m_1 + A_1 Z_1^*$$

其分布律为 P_0 和 P_1。最后，由 (18)、(20) 和 (68) 得到

$$\mathbb{E}\,\|X_0-X_1\|^2 = \|m_0-m_1\|^2 + d_{\mathrm{BW}}^2(\Sigma_0,\Sigma_1) + W_{2,K}^2(\eta_{\nu_0},\eta_{\nu_1}) = W_2^2(P_0,P_1).$$

因此，$(X_0,\ X_1)$ 是二次代价下的最优耦合。 □

推论*3.3*的证明。 由于 $K=\lambda I_d$，由命题 2.3 和引理 2.2 可得

$$W_2^2(P_0,P_1) = \|m_0-m_1\|^2 + d_{\mathrm{BW}}^2(\Sigma_0,\Sigma_1) + \lambda W_2^2(\nu_0,\nu_1),$$

这就证明了 (i)。

现在假设 ν_0 是无原子的。由引理 2.2，映射

$$S(z) = \frac{\tau(\|z\|)}{\|z\|} z, \qquad \tau = F_{\nu_1}^{-1} \circ F_{\nu_0},$$

是从 η_{ν_0} 到 η_{ν_1} 的最优传输，其中该公式几乎处处按 η_{ν_0} 理解。结合 $A_0 = \Sigma_0^{1/2}$ 和 $A_1 = \Sigma_0^{-1/2} K = \lambda \Sigma_0^{-1/2}$，变换回原坐标可得

$$T(x) = m_1 + A_1 S\big(A_0^{-1}(x-m_0)\big),$$

这与（36）一致。该映射将 P_0 推送到 P_1，其最优性由（20）以及 S 的最优性推出。这证明了（ii）。

最后，定义

$$H(r) := \int_0^r \tau(s)\ \mathrm{d}s, \qquad r \ge 0,$$

根据陈述中的端点约定。由于 τ 非负且非递减，H 是适当的、凸的且非递减的。由于 $t(x) = \left\|\Sigma_0^{-1/2}(x-m_0)\right\|$ 是凸的，势函数

$$\Phi(x) = m_1^{\mathsf{T}} x + \lambda H(t(x))$$

也是适当且凸的。在 $t(x) > 0$ 且 τ 连续的点上，链式法则给出

$$\nabla\Phi(x) = m_1 + \lambda \frac{\tau(t(x))}{t(x)} \Sigma_0^{-1}(x-m_0) = T(x).$$

由于 τ 只有至多可数个不连续点，且 ν_0 是无原子的，该恒等式 P_0-几乎处处成立 st
当然，证明 (iii)。 □

推论 *3.5* 的证明。 下界是定理 3.6 的一个特例，而上界是定理 3.8 的一个特例。

□

为证明定理 3.6，我们使用以下引理。

引理 B.3（投影球面对称律）。设 Q 是一个秩为 $q \geq 1$ 的正交投影。对于球面分布 $\eta_i = \eta_{\nu_i}$，有

$$\inf_{\pi \in \Pi(\eta_0, \eta_1)} \int \|Q(z-w)\|^2 \, \mathrm{d}\pi = W_2^2(Q_{\#}\eta_0, Q_{\#}\eta_1) = \Delta_q. \tag{69}$$

证明。令 $E := \mathrm{ran}(Q) := \{Qx \colon x \in \mathbb{R}^d\}$，于是 $\dim E = q$，并定义 $\mu_i^Q := Q_{\#}\eta_i$，其中 $i = 0,1$。我们首先确定投影后的分布律。令 $Z_i = R_i U_i \sim \eta_i$，其中 $R_i \sim \nu_i$ 与 $U_i \sim \mathrm{Unif}(\mathbb{S}^{d-1})$ 独立。将引理 B.1 应用于正交分解 $\mathbb{R}^d = E \oplus E^{\perp}$，可得

$$\|QU_i\|^2 \sim \mathrm{Beta}\left(\frac{q}{2}, \frac{d-q}{2}\right)$$

当 $q < d$ 时，而当 $q = d$ 时 $\|QU_i\| = 1$。此外，方向 $QU_i/\|QU_i\|$ 在 E 的单位球面上均匀分布，且与 $\|QU_i\|$ 独立。因此，μ_i^Q 是 E 上的球面分布，其径向分布为

$$\rho_{i,q} := \mathrm{law}(R_i \sqrt{B_{q,d}}).$$

注意 $\mathbb{E}\big[R_i^2 B_{q,d}\big] = d\,\frac{q}{d} = q$，因此 $\rho_{i,\,q} \in \mathcal{R}_q$。于是，在 q 维空间 E 中应用引理 2.2，可得

$$W_2^2(\mu_0^Q, \mu_1^Q) = W_2^2(\rho_{0,q}, \rho_{1,q}) = \Delta_q, \tag{70}$$

这就证明了（69）中的第二个等式。

为证明（69）中的第一个等式，令 $\pi \in \Pi(\eta_0, \eta_1)$。那么，$\gamma := (Q, Q)_{\#}\pi$ 属于 $\Pi(\mu_0^Q, \mu_1^Q)$，因此

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \|Q(z-w)\|^2 \, \pi(\mathrm{d}z, \mathrm{d}w) = \int_{E \times E} \|x-y\|^2 \, \gamma(\mathrm{d}x, \mathrm{d}y) \geq W_2^2(\mu_0^Q, \mu_1^Q).$$

对 π 取下确界可得

$$\inf_{\pi \in \Pi(\eta_0, \eta_1)} \int \|Q(z-w)\|^2 \, \mathrm{d}\pi \geq W_2^2(\mu_0^Q, \mu_1^Q). \tag{71}$$

对于反向不等式，我们相对于 Q对 η_i 进行解积分。由于所有空间都是波兰空间，存在概率核 $K_i(x, \mathrm{d}z)$、$x \in E$，使得

$$\eta_i(\mathrm{d}z) = \int_E K_i(x, \mathrm{d}z) \, \mu_i^Q(\mathrm{d}x),$$

并且 $K_i\big(x, \{z \in \mathbb{R}^d \colon Qz = x\}\big) = 1$ 对于 μ_i^Q-几乎处处的 x成立。任取 $\gamma \in \Pi(\mu_0^Q, \mu_1^Q)$，并定义

$$\pi_\gamma(\mathrm{d}z, \mathrm{d}w) := \int_{E \times E} K_0(x, \mathrm{d}z) K_1(y, \mathrm{d}w) \, \gamma(\mathrm{d}x, \mathrm{d}y).$$

那么 $\pi_\gamma \in \Pi(\eta_0, \eta_1)$。此外，由于 $K_0\big(x, \{z \colon Qz = x\}\big) = 1$ 以及 $K_1\big(y, \{w \colon Qw = y\}\big) = 1$ 对几乎处处的 x 和 y成立，根据托内利定理可得

$$\begin{aligned} \int \|Q(z-w)\|^2 \, \pi_\gamma(\mathrm{d}z, \mathrm{d}w) &= \int_{E \times E} \int \|Qz - Qw\|^2 K_0(x, \mathrm{d}z) K_1(y, \mathrm{d}w) \, \gamma(\mathrm{d}x, \mathrm{d}y) \\ &= \int_{E \times E} \|x-y\|^2 \, \gamma(\mathrm{d}x, \mathrm{d}y). \end{aligned}$$

因此，对于每个 $\gamma \in \Pi(\mu_0^Q, \mu_1^Q)$，

$$\inf_{\pi \in \Pi(\eta_0, \eta_1)} \int \|Q(z-w)\|^2 \, \pi(\mathrm{d}z, \mathrm{d}w) \leq \int \|x-y\|^2 \, \gamma(\mathrm{d}x, \mathrm{d}y).$$

因此，对 $\gamma \in \Pi(\mu_0^Q, \mu_1^Q)$ 取下确界即可得到反向不等式。结合（71），这就证明了（69）中的第一个等式。 □

定理 3.6 的证明。 令 $Q_j = \pi_j + \cdots + \pi_k$ 为投影到特征值至少为 κ_j 的特征子空间上的投影。根据（43），该矩阵可以分解为 $K = \kappa_1 I_d + \sum_{j=2}^{k}(\kappa_j - \kappa_{j-1})Q_j$。对于 $\pi \in \Pi(\eta_{\nu_0}, \eta_{\nu_1})$，令 $(Z_0, Z_1) \sim \pi$。于是得到

$$\begin{aligned}\mathbb{E}_\pi\left[(Z_0 - Z_1)^\mathsf{T} K (Z_0 - Z_1)\right] &= \kappa_1 \mathbb{E}_\pi \|Z_0 - Z_1\|^2 + \sum_{j=2}^{k}(\kappa_j - \kappa_{j-1})\mathbb{E}_\pi\left[(Z_0 - Z_1)^\mathsf{T} Q_j (Z_0 - Z_1)\right] \\ &= \kappa_1 \mathbb{E}_\pi \|Z_0 - Z_1\|^2 + \sum_{j=2}^{k}(\kappa_j - \kappa_{j-1})\mathbb{E}_\pi \|Q_j(Z_0 - Z_1)\|^2 ,\end{aligned}$$

其中我们使用第二个等式，这是因为 $(Z_0 - Z_1)^\mathsf{T} Q_j (Z_0 - Z_1) = \|Q_j(Z_0 - Z_1)\|^2$，由于每个 Q_j 都是一个正交的投影。

现在，I_d 是一个秩为 d 的正交投影。因此，由引理 B.3 可得 $\mathbb{E}_\pi \|Z_0 - Z_1\|^2 \geq \Delta_d$.此外，$Q_j$ 是到 $E_j \oplus \cdots \oplus E_k$ 上的正交投影，因此其秩为 $d_j = n_j + \cdots + n_k$。将引理 B.3 应用于 Q_j 可得

$$\mathbb{E}_\pi \|Q_j(Z_0 - Z_1)\|^2 \geq \Delta_{d_j}, \qquad j \in \{2, \ldots, k\}.$$

由于对于 j，$\kappa_1 > 0$ 和 $\kappa_j - \kappa_{j-1} > 0$，我们得到

$$\mathbb{E}_\pi\left[(Z_0 - Z_1)^\mathsf{T} K (Z_0 - Z_1)\right] \geq \kappa_1 \Delta_d + \sum_{j=2}^{k}(\kappa_j - \kappa_{j-1})\Delta_{d_j} = L_K.$$

取对所有此类耦合 π 取下确界，并应用 (24)，得到 $\Gamma(P_0, P_1) = W_{2,K}^2(\eta_{\nu_0}, \eta_{\nu_1}) \geq L_K$，这证明了 (44)。

如果 $k \geq 2$ 且 $\nu_0 \neq \nu_1$，那么对每个 $q = 1$、……、d都有 $\Delta_q > 0$。事实上，假设对于某个 $q \geq 1$ 有 $\Delta_q = 0$。那么，根据引理 B.3，将 η_{ν_0} 和 η_{ν_1} 正交投影到任意固定的 q 维子空间上所得的正交投影具有相同的径向分布。由于这些正交投影是球对称的，因此它们的分布相同。特别地，对于投影子空间中的任意单位向量 v，都有 $\langle v, Z_0\rangle \overset{d}{=} \langle v, Z_1\rangle$，其中 $Z_i \sim \eta_{\nu_i}$，$i = 0, 1$。由 η_{ν_0} 和 η_{ν_1}的球对称性，它们所有的一维投影因此都相同。Cramér–Wold 定理推出原始球对称分布相等，并且取范数可得 $\nu_0 = \nu_1$，矛盾。因此，由于对于 $j = 2$、……、k有 $\kappa_j - \kappa_{j-1} > 0$，我们得到 $L_K = \kappa_1 \Delta_d + \sum_{j=2}^{k}(\kappa_j - \kappa_{j-1})\Delta_{d_j} > \kappa_1 \Delta_d$. □

定理 3.8 的证明。 固定$\mathbf{b} = (b_1, \ldots, b_m)$。令 e_1、……、e_d为 K的一个正交归一特征基，其中 e_h是与特征值 $\lambda_h^{\downarrow}$ 对应的特征向量。这些子空间

$$E_h := \operatorname{span}\{e_{a_h+1}, \ldots, e_{a_h+b_h}\}, \qquad h = 1, \ldots, m.$$

两两正交，且维数为 $\dim E_h = b_h$。记 $\pi_{E_h} = \sum_{j=a_h+1}^{a_h+b_h} e_j e_j^\top$为到 E_h上的正交投影。于是有

$$\operatorname{tr}(K\pi_{E_h}) = \operatorname{tr}\left(\sum_{i=1}^{d}\sum_{j=a_h+1}^{a_h+b_h} \lambda_i^{\downarrow} e_i (e_i^\top e_j) e_j^\top\right) = \sum_{j=a_h+1}^{a_h+b_h} \lambda_j^{\downarrow} = \Lambda_{a_h+b_h} - \Lambda_{a_h}, \tag{72}$$

其中，第三个等式利用了 e_i 和 e_j 正交这一事实。

考虑（47）中的随机向量 $Z_i = (Z_{i,1}, \ldots, Z_{i,m})$，其中 $Z_{i,h}$ 表示 ν_i在子空间 E_h内的径向部分。由于（48）中的 Knothe–Rosenblatt 耦合（Z_0^{KR}、Z_1^{KR}）保持边缘分布不变，因此有

$$Z_i^{\mathrm{KR}} \overset{\mathrm{d}}{=} Z_i, \qquad i = 0, 1.$$

令 V_1、……、V_m 为相互独立的 $\mathbb{R}^d$值随机向量，且与 $(Z_0^{\mathrm{KR}}, Z_1^{\mathrm{KR}})$ 独立，其中 V_h 在 E_h的单位球面上均匀分布。定义

$$Y_i := \sum_{h=1}^{m} Z_{i,h}^{\mathrm{KR}} V_h, \qquad i = 0, 1.$$

注 th　其中 Y_i 是 d维的，而 Z_i^{KR}是一个 m维随机向量。我们的目标是证明　　t

$$\Gamma(P_0,P_1)\le \mathbb{E}\left[(Y_0-Y_1)^\top K(Y_0-Y_1)\right]=U_{K,\mathbf{b}}, \tag{73}$$

这得到了（51）中的第一个不等式。由于 Knothe–Rosenblatt 耦合保持块半径分布不变，引理 B.1 中的球形分块分解意味着

$$Y_i\sim \eta_{\nu_i},\qquad i=0,1.$$

因此，(Y_0，Y_1) 是标准化球面分布的一个可容许耦合。

现在，为了计算 $U_{K,\ \mathbf{b}}$ 作为生成元效应的上界，我们计算

$$\mathbb{E}\left[(Y_0-Y_1)^\top K(Y_0-Y_1)\right]=\sum_{h=1}^{m}\mathbb{E}\left[(Z_{0,h}^{\mathrm{KR}}-Z_{1,h}^{\mathrm{KR}})^2\right]\mathbb{E}[V_h^\top K V_h], \tag{74}$$

其中我们使用了 $Y_0-Y_1=\sum_{h=1}^{m}(Z_{0}^{\mathrm{KR}}{}_{,\ h}-Z_{1\,h,}^{\mathrm{KR}})V_h$、$Z_{i\,h,}^{\mathrm{KR}}$ 与方向 V_h 的独立性，以及由于子空间 E_h 两两正交且在 K 下保持不变，所有混合项均为零。

对于（74）中的右侧期望，我们得到

$$\mathbb{E}[V_h^\top K V_h]=\operatorname{tr}\left(K\mathbb{E}[V_hV_h^\top]\right)=\frac{\operatorname{tr}(K\pi_{E_h})}{b_h}=\frac{\Lambda_{a_h+b_h}-\Lambda_{a_h}}{b_h},$$

其中，第一个等式利用了迹的循环性。对于第二个等式，我们利用了 V_h 在 b_h 维子空间 E_h 的单位球面上服从均匀分布，因此 $\mathbb{E}[V_hV_h^\top]=\frac{1}{b_h}\pi_{E_h}$。第三个等式由（72）得出。由此可得　　—

$$\mathbb{E}\left[(Y_0-Y_1)^\top K(Y_0-Y_1)\right]=\sum_{h=1}^{m}\frac{\Lambda_{a_h+b_h}-\Lambda_{a_h}}{b_h}\mathbb{E}\left[(Z_{0,h}^{\mathrm{KR}}-Z_{1,h}^{\mathrm{KR}})^2\right]=\sum_{h=1}^{m}\frac{\Lambda_{a_h+b_h}-\Lambda_{a_h}}{b_h}C_h^{\mathbf{b}}=U_{K,\mathbf{b}}.$$

由于（Y_0，Y_1）是 η_{ν_0} 与 η_{ν_1}的一个可容许耦合，我们证明了（73）。

最后，取 $m=1$ 和 $\mathbf{b}=$（d），Knothe–Rosenblatt 耦合便退化为径向分布 ν_0 和 ν_1的共单调耦合。因此，$C_1^{(d)}=W_2^2$（ν_0，ν_1）$=\Delta_d=\delta_R^2$。由于根据（46）的定义，$\Lambda_d=\operatorname{tr}K$，因此可得

$$U_K\le U_{K,(d)}=\frac{\operatorname{tr}K}{d}\delta_R^2,$$

这证明了（51）中的第二个不等式，并完成了证明。　　□

B.3 第 4 节的证明

为证明定理 4.1，我们确定（47）中分块半径向量 $Z_i=$（$Z_{i,\ 1}$，…，$Z_{i,\ m}$）的表 2 中的条件分布 $Q_{a,\ b}$。

引理 B.4 (块半径的条件分布)。对于$\mathbf{b}=(b_1,\dots,b_m)$，其中整数 $b_h\ge 1$ 满足 $\sum_h b_h=d$，令 $Z=(Z_1,\dots,Z_m)$ 为式 (47) 中的分块半径向量。假设 Z 是表 2 中四个族之一所对应的球面对称随机向量的分块半径向量，其相关常数为 $c>0$ 和 $\varepsilon\in\{-1,0,1\}$。令 $a_h=\sum_{\ell<h}b_\ell$。那么，对于 $h=1,\dots,m$，在给定先前块半径 $Z_{1:h-1}=r$ 的条件下，Z_h 的条件分布满足

$$(Z_h/\alpha_h\mid Z_{1:h-1}=r)\sim Q_{a_h,b_h},\qquad \alpha_h:=\sqrt{c+\varepsilon\sum_{\ell<h}r_\ell^2} \tag{75}$$

表 2 中指定的分布 Q_{a_h,b_h} 不依赖于 r。

证明。我们区分表 2 中考虑的四种分布。

高斯分布。令 $G\sim N(0,\ I_d)$，并根据有序合成 $\mathbf{b}=(b_1,\dots,\ b_m)$ 将 $G=(G_1,\ \dots,\ G_m)$ 分解，其中 $G_h\in\mathbb{R}^{b_h}$。由于协方差矩阵为 I_d，随机向量$G_1,\ \dots,\ G_m$相互独立，且服从 $G_h\sim N(0,\ I_{b_h})$。因此，块半径满足

$$Z_h=\|G_h\|=\sqrt{H_h},\qquad H_h:=\|G_h\|^2\sim\chi^2_{b_h},$$

where $H_1, \ldots, H_m$ 相互独立。特别地，对于几乎处处的 $r = (r_1, \ldots, r_{h-1})$，我们得到

$$(Z_h \mid Z_{1:h-1} = r) \stackrel{\mathrm{d}}{=} Z_h = \sqrt{H_h} \sim Q_{a_h, b_h}.$$

由于 $c = 1$ 和 $\varepsilon = 0$，这就得到了 (75)。

斯图登-t 分布。回忆一下，具有$\theta > 2$ 自由度的协方差标准化球形斯图登-t 分布可表示为 $\sqrt{c/V}\, G$，其中 $c = \theta - 2$、$G \sim N(0, I_d)$ 和 $V \sim \chi^2_\theta$相互独立。将 G 分解为相互独立的高斯分块 $G_h \sim N(0, I_{b_h})$、$h = 1, \ldots, m$，则 (47) 中的分块半径向量 $Z = (Z_1, \ldots, Z_m)$ 满足

$$Z \stackrel{\mathrm{d}}{=} \sqrt{\frac{c}{V}}\, (\|G_1\|, \ldots, \|G_m\|) \stackrel{\mathrm{d}}{=} \left(\sqrt{\frac{cH_1}{V}}, \ldots, \sqrt{\frac{cH_m}{V}} \right), \tag{76}$$

其中 $H_h := \|G_h\|^2 \sim \chi^2_{b_h}$。那么，在给定 $V = v$ 的条件下，块半径 $Z_1, \ldots, Z_m$ 相互独立。特别地，前 $h-1$ 个块半径的联合条件密度在 $r = (r_1, \ldots, r_{h-1})$ 处的取值为

$$f_{Z_{1:h-1} \mid V = v}(r) = \prod_{\ell < h} f_{Z_\ell \mid V = v}(r_\ell) \propto v^{a_h/2} \exp\left(-\frac{v}{2c} \sum_{\ell < h} r_\ell^2 \right),$$

其中 $a_h := \sum_{\ell < h} b_\ell$。这里，$\propto$ 表示作为 v 的函数成比例（即省略了与 v 无关的因子）。将上述似然与 $V \sim \chi^2_\theta$ 的密度相乘，并应用贝叶斯公式，得到

$$\begin{aligned} f_{V \mid Z_{1:h-1} = r}(v) &\propto v^{a_h/2} \exp\left(-\frac{v}{2c} \sum_{\ell < h} r_\ell^2 \right) \frac{1}{2^{\theta/2}\Gamma(\theta/2)} v^{\theta/2 - 1} e^{-v/2} \\ &\propto v^{(\theta + a_h)/2 - 1} \exp\left[-\frac{v}{2} \left(1 + \frac{\sum_{\ell < h} r_\ell^2}{c} \right) \right]. \end{aligned}$$

由此得到

$$(V \mid Z_{1:h-1} = r) \stackrel{\mathrm{d}}{=} \frac{c}{c + \sum_{\ell < h} r_\ell^2} \chi^2_{\theta + a_h}.$$

由于根据 (76)，$Z_h = \sqrt{cH_h/V}$，其中 $H_h \sim \chi^2_{b_h}$ 与 V 及前面的块半径 $Z_{1:h-1}$ 相互独立，因此我们得到

$$(Z_h \mid Z_{1:h-1} = r) \stackrel{\mathrm{d}}{=} \sqrt{c + \sum_{\ell < h} r_\ell^2}\, \sqrt{\frac{H_h}{V_{\theta + a_h}}},$$

其中 $V_{\theta + a_h} \sim \chi^2_{\theta + a_h}$ 与 H_h 相互独立。取 $\varepsilon = 1$，即可证明 (75)。

球上的均匀分布。假设 $Y = (Y_1, \ldots, Y_d)$ 服从半径为 $\sqrt{d+2}$ 的球上的均匀球面分布。令 $c = d + 2$。令 $G \sim N(0, I_d)$，且 $E \sim \chi^2_2$ 相互独立。我们首先注意到

$$Y \stackrel{\mathrm{d}}{=} \sqrt{\frac{c}{\|G\|^2 + E}}\, G. \tag{77}$$

事实上，由于 $\|G\|^2 \sim \chi^2_d$，我们有 $\|G\|^2/(\|G\|^2 + E) \sim$ 贝塔分布 $(d/2, 1)$。因此，(77) 右侧随机向量的半径满足

$$\mathbb{P}\left(\sqrt{\frac{c\|G\|^2}{\|G\|^2 + E}} \le s \right) = \left(\frac{s}{\sqrt{c}} \right)^d, \qquad 0 \le s \le \sqrt{c}. \tag{78}$$

此外，其方向 $G/\|G\|$ 在 $\mathbb{S}^{d-1}$ 上服从均匀分布，并且与其半径相互独立。这就证明了（77）。现在，根据有序合成 $\mathbf{b} = (b_1, \ldots, b_m)$，将 $G = (G_1, \ldots, G_m)$ 分解，其中$G_h \sim N(0, I_{b_h})$ 相互独立。令

$$H_h := \|G_h\|^2 \sim \chi^2_{b_h}, \qquad h = 1, \ldots, m,$$

由（77）可得，按照（47）定义的分块半径向量 $Z = (Z_1, \dots, Z_m)$ 具有如下表示

$$Z \overset{\mathrm{d}}{=} \sqrt{\frac{c}{H_1 + \cdots + H_m + E}} \left(\sqrt{H_1}, \dots, \sqrt{H_m} \right).$$

特别地，

$$\frac{Z_h^2}{c} \overset{\mathrm{d}}{=} \frac{H_h}{H_1 + \cdots + H_m + E},\ h = 1, \dots, m, \qquad \text{and} \qquad 1 - \frac{\sum_{\ell=1}^m Z_\ell^2}{c} \overset{\mathrm{d}}{=} \frac{E}{H_1 + \cdots + H_m + E}. \tag{79}$$

由于 H_1，…，H_m，E 分别是自由度为 b_1，…，b_m，2 的独立卡方随机变量，因此我们有

$$\left(\frac{Z_1^2}{c}, \dots, \frac{Z_m^2}{c}, 1 - \frac{\sum_{\ell=1}^m Z_\ell^2}{c} \right) \sim \text{Dirichlet} \left(\frac{b_1}{2}, \dots, \frac{b_m}{2}, 1 \right),$$

参见（60）。为确定 $Z_h \mid Z_{1:h-1} = r$ 的条件分布，令

$$W_h := \sum_{\ell=h}^m H_\ell + E, \qquad h = 1, \dots, m, \qquad W_{m+1} := E.$$

利用表示式（79），我们得到

$$Z_h^2 \overset{\mathrm{d}}{=} c \frac{H_h}{W_1}, \qquad c - \sum_{\ell<h} Z_\ell^2 \overset{\mathrm{d}}{=} c \frac{W_h}{W_1}, \qquad \text{and thus} \quad \frac{Z_h^2}{c - \sum_{\ell<h} Z_\ell^2} \overset{\mathrm{d}}{=} \frac{H_h}{W_h} = \frac{H_h}{H_h + W_{h+1}}.$$

由于 $H_h \sim \chi^2_{b_h}$ 和 $W_{h+1} \sim \chi^2_{d-a_h-b_h+2}$ 相互独立，（61）中的贝塔分布与卡方分布之间的关系给出

$$\frac{H_h}{H_h + W_{h+1}} \sim \text{Beta} \left(\frac{b_h}{2}, \frac{d - a_h - b_h}{2} + 1 \right) \qquad \text{and} \qquad \frac{H_h}{H_h + W_{h+1}} \perp W_h,$$

这里我们利用了 $W_h = H_h + W_{h+1}$ 服从卡方分布，因为它是若干独立且服从卡方分布的随机变量之和。由于 H_1，…，H_m，E 相互独立，随机向量（H_1，…，H_{h-1}）与（H_h，W_{h+1}）相互独立，因而也与 H_h/W_h 独立。因此，对于几乎所有的 $r =$（r_1，…，r_{h-1}），我们得到

$$\left(\frac{Z_h^2}{c - \sum_{\ell<h} Z_\ell^2} \,\middle|\, Z_{1:h-1} = r \right) \sim \text{Beta} \left(\frac{b_h}{2}, \frac{d - a_h - b_h}{2} + 1 \right)$$

与 r 独立。取 $\varepsilon = -1$，即可证明球上的均匀分布满足 (75)。

球面上的均匀分布。假设 $Y = (Y_1, \dots, Y_d)$ 服从半径为 $\sqrt{d}$ 的球面均匀分布，并令 $c = d$。令 $G \sim N(0, I_d)$。由于 $G/\|G\|$ 在单位球面 $\mathbb{S}^{d-1}$ 上服从均匀分布，设 $E = 0$，则如 (77) 中所示，我们有

$$Y \overset{\mathrm{d}}{=} \sqrt{\frac{d}{\|G\|^2 + E}}\, G \qquad \text{and} \qquad \left\| \sqrt{\frac{c}{\|G\|^2 + E}}\, G \right\|^2 = \frac{c\,\|G\|^2}{\|G\|^2 + E} = c. \tag{80}$$

现在，仿照球内均匀分布的情形进行推导，可得只要 $a_h + b_h < d$（而非 $W_{h+1} \sim \chi^2_{d-a_h-b_h+2}$），就有 $W_{h+1} \sim \chi^2_{d-a_h-b_h}$。于是我们得到

$$\left(\frac{Z_h^2}{c - \sum_{\ell<h} Z_\ell^2} \,\middle|\, Z_{1:h-1} = r \right) \overset{\mathrm{d}}{=} \frac{H_h}{H_h + W_{h+1}} \sim \text{Beta} \left(\frac{b_h}{2}, \frac{d - a_h - b_h}{2} \right)$$

对几乎所有的 $r = (r_1, \dots, r_{h-1})$ 都成立。若 $a_h + b_h = d$，则 $W_{h+1} = 0$，且上述条件分布是在 1 处的狄拉克分布。取 $\varepsilon = -1$，即可证明球面上的均匀分布满足 (75)。 □

定理 *4.1* 的证明。 令 $\mathbf{b} = (b_1, \dots, b_m)$ 为 d 的一种有序合成。我们首先证明，对于表 2 中的球对称分布，式 (48) 中的 Knothe–Rosenblatt 耦合具有交叉矩恒等式表示

$$\mathbb{E}\left[Z_{0,h}^{\mathrm{KR}} Z_{1,h}^{\mathrm{KR}} \right] = \sqrt{c_0 c_1}\, I_{a_h, b_h} \prod_{\ell<h} J_{a_\ell, b_\ell}, \qquad h = 1, \dots, m. \tag{81}$$

为证明该恒等式，我们使用引理 B.4，得到式 (47) 中块半径随机向量 $Z_i = (Z_{i,1}, \dots, Z_{i,m})$ 的条件分布

$$\left(\frac{Z_{i,h}}{\sqrt{c_i + \varepsilon_i \sum_{\ell<h} r_\ell^2}} \mid Z_{i,1:h-1} = r \right) \sim Q_{i;a_h,b_h},$$

T因此，式（48）中 Knothe–Rosenblatt 耦合的条件分位数构造给出

$$Z_{i,h}^{\mathrm{KR}} = \sqrt{c_i + \varepsilon_i \sum_{\ell<h} (Z_{i,\ell}^{\mathrm{KR}})^2}\, g_{i;a_h,b_h}(U_h). \tag{82}$$

这意味着

$$c_i + \varepsilon_i \sum_{\ell\le h} (Z_{i,\ell}^{\mathrm{KR}})^2 = \left(c_i + \varepsilon_i \sum_{\ell<h} (Z_{i,\ell}^{\mathrm{KR}})^2 \right) \left(1 + \varepsilon_i\, g_{i;a_h,b_h}(U_h)^2 \right).$$

对上述恒等式进行迭代可得

$$\sqrt{c_i + \varepsilon_i \sum_{\ell<h} (Z_{i,\ell}^{\mathrm{KR}})^2} = \sqrt{c_i} \prod_{\ell<h} \sqrt{1 + \varepsilon_i\, g_{i;a_\ell,b_\ell}(U_\ell)^2}. \tag{83}$$

因此，结合（82）和（83）可得

$$Z_{i,h}^{\mathrm{KR}} = \sqrt{c_i}\, g_{i;a_h,b_h}(U_h) \prod_{\ell<h} \sqrt{1 + \varepsilon_i\, g_{i;a_\ell,b_\ell}(U_\ell)^2}. \tag{84}$$

由于 U_1、……、U_m 相互独立，我们得到

$$\begin{aligned} \mathbb{E}\left[Z_{0,h}^{\mathrm{KR}} Z_{1,h}^{\mathrm{KR}} \right] &= \sqrt{c_0 c_1}\, \mathbb{E}\left[g_{0;a_h,b_h}(U_h) g_{1;a_h,b_h}(U_h) \right] \prod_{\ell<h} \mathbb{E}\left[\sqrt{1 + \varepsilon_0\, g_{0;a_\ell,b_\ell}(U_\ell)^2} \sqrt{1 + \varepsilon_1\, g_{1;a_\ell,b_\ell}(U_\ell)^2} \right] \\ &= \sqrt{c_0 c_1}\, I_{a_h,b_h} \prod_{\ell<h} J_{a_\ell,b_\ell}, \end{aligned} \tag{85}$$

这就证明了（81）。

接下来，我们证明下文的表示式 (86)。由于对于与 R_i 独立的 Beta$(b_h/2, (d-b_h)/2)$，有 $Z_i^{\mathrm{KR}} \overset{\mathrm{d}}{=} Z_i$ 和 $Z_{i,h} = R_i \sqrt{D_h}$，因此我们得到 $\mathbb{E}[(Z_{i,h}^{\mathrm{KR}})^2] = \mathbb{E}[Z_{i,h}^2] = \mathbb{E}[R_i^2]\mathbb{E}[D_h] = d\frac{b_h}{d} = b_h$。于是，根据 (85)，我们得到第 h 个块的二次代价为

$$C_h^{\mathbf{b}} = \mathbb{E}\left[(Z_{0,h}^{\mathrm{KR}} - Z_{1,h}^{\mathrm{KR}})^2 \right] = 2b_h - 2\sqrt{c_0 c_1}\, I_{a_h,b_h} \prod_{\ell<h} J_{a_\ell,b_\ell}.$$

利用 $w_{a,b} := (\Lambda_{a+b} - \Lambda_a)/b$，我们得到

$$\sum_{h=1}^{m} b_h w_{a_h,b_h} = \sum_{h=1}^{m} (\Lambda_{a_h+b_h} - \Lambda_{a_h}) = \operatorname{tr} K.$$

因此，根据 (50) 中 $U_{K,\mathbf{b}}$ 的定义，我们有

$$U_{K,\mathbf{b}} = \sum_{h=1}^{m} w_{a_h,b_h} C_h^{\mathbf{b}} = 2 \operatorname{tr} K - 2\sqrt{c_0 c_1} \sum_{h=1}^{m} w_{a_h,b_h} I_{a_h,b_h} \prod_{\ell<h} J_{a_\ell,b_\ell}. \tag{86}$$

仍需在 d 的所有有序分拆 $\mathbf{b} = (b_1, \dots, b_m)$ 上对 (86) 式进行最小化。由于 c_0，$c_1 > 0$，这等价于最大化

$$S(\mathbf{b}) := \sum_{h=1}^{m} w_{a_h,b_h} I_{a_h,b_h} \prod_{\ell<h} J_{a_\ell,b_\ell}.$$

我们通过反向归纳证明，递推式 (53) 给出了所需的最大值。对于 $a = d$，我们令$B_d = 0$，无需进行优化。对于归纳步骤 $a \mapsto a-1$，假设对于所有满足 $j \geq a$ 的 B_j，该断言均成立。考虑 $d-a+1$ 的任意有序分拆 $\mathbf{b} = (b_1, \ldots, b_m)$，并令

$$a_h := a - 1 + \sum_{\ell<h} b_\ell.$$

S分离出索引为 $h = 1$ 的块的贡献，并提取出 $J_{a-1,\ b_1}$，我们得到 n

$$\sum_{h=1}^{m} w_{a_h,b_h} I_{a_h,b_h} \prod_{\ell<h} J_{a_\ell,b_\ell} = w_{a-1,b_1} I_{a-1,b_1} + J_{a-1,b_1} \sum_{h=2}^{m} w_{a_h,b_h} I_{a_h,b_h} \prod_{\ell=2}^{h-1} J_{a_\ell,b_\ell},$$

采用空乘积 $\prod_{\ell<1} J_{a_\ell,\ b_\ell}$ 等于 1 的约定。根据归纳假设，右侧求和在所有有序分拆 $(b_2, \ldots\ldots, b_m)$ 上的最大值为 B_{a-1+b_1}，因为 $a-1+b_1 \geq a$。因此，我们最大化

$$w_{a-1,b_1} I_{a-1,b_1} + J_{a-1,b_1} B_{a-1+b_1}$$

所有可能的首个块大小 $1 \leq b_1 \leq d-a+1$。这给出

$$B_{a-1} = \max_{1 \leq b \leq d-a+1} \left\{ w_{a-1,b} I_{a-1,b} + J_{a-1,b} B_{a-1+b} \right\}.$$

这就完成了归纳步骤。因此， B_0 是 $S(\mathbf{b})$ 在所有 d 的有序分拆上的最大值。因此，

$$U_K = \min_{\mathbf{b}} U_{K,\mathbf{b}} = 2 \operatorname{tr} K - 2\sqrt{c_0 c_1}\, B_0,$$

从而完成证明。 □

B.4 第5节的证明

引理 *5.1* 的证明。设 $E \subset \mathbb{R}^d$ 是一个 r 维子空间，并选取其标准正交基$e_1, \ldots, e_r$。对于每个满足 $|A| = q$ 的子集 $A \subset \{1, \ldots, r\}$，令 Q_A 表示到 E_A 的正交投影：=张成空间$\{e_a : a \in A\}$。令 μ_i^E 为 η_{ν_i} 在 E 上的投影， $i = 0,1$，并令 $\gamma \in \Pi(\mu_0^E, \mu_1^E)$。由 (69) 可得，

$$\int \|Q_A(x-y)\|^2 \, \gamma(\mathrm{d}x, \mathrm{d}y) \geq \Delta_q$$

对每个这样的 A 都成立。对所有 $\binom{r}{q}$ 个子集 A 求和，得到

$$\binom{r}{q} \Delta_q \leq \sum_{\substack{A \subset \{1,\ldots,r\} \\ |A|=q}} \int \|Q_A(x-y)\|^2 \, \gamma(\mathrm{d}x, \mathrm{d}y) = \binom{r-1}{q-1} \int \|x-y\|^2 \, \gamma(\mathrm{d}x, \mathrm{d}y), \tag{87}$$

第二个等号中，我们利用了

$$\sum_{\substack{A \subset \{1,\ldots,r\} \\ |A|=q}} \|Q_A v\|^2 = \sum_{\substack{A \subset \{1,\ldots,r\} \\ |A|=q}} \sum_{\ell \in A} \langle v, e_\ell \rangle^2 = \binom{r-1}{q-1} \sum_{\ell=1}^{r} \langle v, e_\ell \rangle^2 = \binom{r-1}{q-1} \|v\|^2.$$

因此，由 (87) 得

$$\int \|x-y\|^2 \, \mathrm{d}\gamma \geq \frac{\binom{r}{q}}{\binom{r-1}{q-1}} \Delta_q = \frac{r}{q} \Delta_q.$$

对 γ 取下确界可得 $\Delta_r \geq \frac{r}{q} \Delta_q$，从而得到 $\Delta_q/q \leq \Delta_r/r$。 □

致谢

第一作者感谢奥地利科学基金（FWF） [10.55776/PAT1669224]通过项目 函数依赖的随机序 提供的资助。在准备本手稿期间，作者使用了 ChatGPT（OpenAI），以协助进行数学论证、技术检查、文献检索、数值计算和语言编辑。作者对本工作的数学内容和结论承担全部责任。

参考文献

[1] S.-I. Amari 和 T. Matsuda。形状与仿射变形的 Wasserstein统计量信息几何。信息几何，7(2):285–309，2024。doi: 10.1007/s41884-024-00139-y。[2] R. Bhatia、T. Jain 和 Y. Lim。正定矩阵之间的 Bures–Wasserstein 距离。数学论述，37(2):165–191，2019。doi: 10.1016/j.exmath.2018.01.002。[3] Y. Brenier。向量场的极分解与单调重排。巴黎科学院学报，*I*辑，305(19):805–808，1987。[4] Y. Brenier。向量值函数的极分解与单调重排。纯粹与应用数学通信，44(4):375–417，1991。doi: 10.1002/cpa.3160440402。[5] S. Cambanis、S. Huang 和 G. Simons。关于椭圆等高分布理论。《多元分析杂志》，11(3):368–385，1981。doi: 10.1016/0047-259X(81)90082-8。[6] K. Chen 和 Y. Zhang。径向等高分布的最优传输与Wasserstein重心。数学分析，51(3):775–792，2025。doi: 10.1007/s10476-025-00105-5。[7] K. Chen、Z. Wang 和 Y. Zhang。径向等高分布混合分布的松弛Wasserstein距离表述。应用数学与计算通信，2026。doi: 10.1007/s42967-026-00579-6。[8] J. A. Cuesta-Albertos、L. Rüschendorf 和 A. Tuero-Díaz。多元分布与随机过程的最优耦合。《多元分析杂志》，46(2):335–361，1993。doi: 10.1006/jmva.1993.1064。[9] M. Cuturi。Sinkhorn距离：最优传输的高速计算。载于《神经信息处理系统进展》，第26卷，2013。[10] D. C. Dowson 和 B. V. Landau。多元正态分布之间的 Fréchet距离。《多元分析杂志》，12(3):450–455，1982。doi: 10.1016/0047-259X(82)90077-X。[11] K.-T. Fang、S. Kotz 和 K. W. Ng。对称多元分布及相关分布，统计学与应用概率专著第36卷。Chapman and Hall，伦敦，1990。[12] N. Fournier 和 A. Guillin。经验测度在 Wasserstein距离下的收敛速率。《概率理论及其相关领域》，162:707–738，2015。doi: 10.1007/s00440-014-0583-7。[13] M. Gelbrich。关于欧氏空间与希尔伯特空间中测度之间的 L^2 Wasserstein度量公式。《数学新闻》，147(1):185–203，1990。doi: 10.1002/mana.19901470121。[14] A. Genevay、L. Chizat、F. Bach、M. Cuturi 和 G. Peyré。Sinkhorn散度的样本复杂度。载于第*22*届国际人工智能与统计学会议论文集，机器学习研究论文集第89卷，第1574–1583页，2019。[15] N. Ghaffari 和 S. G. Walker。Parseval恒等式与最优传输映射。《统计学与概率快报》，170:108989，2021。doi: 10.1016/j.spl.2020.108989。[16] N. Ghaffari 和 S. G. Walker。径向相关分布的 W_2 重心。《统计学与概率快报》，195:109788，2023。doi: 10.1016/j.spl.2023.109788。[17] C. R. Givens 和 R. M. Shortt。概率分布的一类Wasserstein度量。《密歇根数学杂志》，31(2):231–240，1984。doi: 10.1307/MMJ/1029003026。

[18] M. Gunasingam 和 T.-K. L. Wong。离散时间高斯过程之间的适应性最优传输。《概率电子通信》，30:1–14，2025。doi: 10.1214/25-ECP654。[19] K. Hamm 和 A. Korzeniowski。关于随机向量仿射变换的 Wasserstein距离。《数据科学基础》，6(4):468–491，2024。doi: 10.3934/fods.2024023。[20] S. Kotz、N. Balakrishnan 和 N. L. Johnson。连续多元分布，第*1*卷：模型与应用。Wiley，第2版，2000。doi: 10.1002/0471722065。[21] D. Kuhn、P. Mohajerin Esfahani、V. A. Nguyen 和 S. Shafieezadeh-Abadeh。Wasserstein分布鲁棒优化：机器学习中的理论与应用。载于《分析时代的运筹学与管理科学》，《INFORMS运筹学教程》，第130–166页。INFORMS，2019。doi: 10.1287/educ.2019.0198。[22] A. W. Marshall、I. Olkin 和 B. C. Arnold。不等式：主序理论及其应用。Springer统计学丛书。Springer，纽约，第2版，2011。doi: 10.1007/978-0-387-68276-1。[23] B. Muzellec 和 M. Cuturi。利用椭圆分布的 Wasserstein空间推广点嵌入。载于《神经信息处理系统进展》，第31卷，第10258–10269页，2018。[24] I. Olkin 和 F. Pukelsheim。具有给定离散矩阵的两个随机向量之间的距离。《线性代数及其应用》，48:257–263，1982。doi: 10.1016/0024-3795(82)90112-4。[25] S. T. Rachev 和 L. Rüschendorf。质量输运问题，第*I*卷：理论。《概率及其应用》。Springer，纽约，1998。[26] L. Rüschendorf。Wasserstein距离与逼近定理。《概率论与相关领域杂志》，70:117–129，1985。doi: 10.1007/BF00532240。[27] L. Rüschendorf 和 S. T. Rachev。具有最小 L^2 距离的随机变量的刻画。《多元分析杂志》，32(1):48–54，1990。doi: 10.1016/0047-259X(90)90070-X。[28] T. Sei 和 H. Matsuzoe。信息几何中的开放问题：FDIG 2025讨论。arXiv预印本 arXiv:2509.06989，2025。URL `https://arxiv.org/abs/2509.06989`。[29] A. Takatsu。高斯测度的 Wasserstein 几何。《大阪数学杂志》，48(4):1005–1026，2011。doi: 10.18910/4973。[30] C. Villani。最优传输：旧与新，数学科学基础丛书第338卷。Springer，柏林，2009。doi: 10.1007/978-3-540-71050-9。[31] J. Weed 和 F. Bach。经验测度在 Wasserstein距离下收敛的精确渐近速率与有限样本收敛速率。《伯努利》，25(4A):2620–2648，2019。doi: 10.3150/18-BEJ1065。