arXiv:2304.06616v2 [math.OC] 7 Oct 2024

Tsallis 熵正则化最优传输的收敛速率

Takeshi Suguro 和 Toshiaki Yachimura

摘要

本文研究连续情形下的 Tsallis 熵正则化最优传输，并建立了若干基本结果，例如当正则化参数趋于零时，Tsallis 正则化最优传输收敛至 Monge–Kantorovich 问题的Γ-收敛。此外，我们利用 Eckstein–Nutz 提出的量化和影子论证，推导出 Tsallis 熵正则化的收敛速率，并给出显式常数。进一步地，我们将这些结果与众所周知的 Kullback–Leibler（KL）散度正则化情形进行比较，并表明在 Tsallis 框架下，KL 正则化可达到最快的收敛速率。

关键词：最优传输、熵正则化、Tsallis 相对熵、Kullback–Leibler 散度、 Γ-收敛、量化

AMS 学科分类：49Q22、90C25、94A17。

1 引言

最优传输理论广泛用于衡量概率测度之间的距离并对其进行比较。近年来，最优传输发展迅速。尤其是对最优传输进行正则化后，即使面对高维数据，也能够进行计算。

设 (X_1, μ_1) 和 (X_2, μ_2) 为 Polish 概率空间，且 $c\colon X_1 \times X_2 \to [0, +\infty)$ 为连续代价函数。以下正则化最优传输是

最常用的:

$$\mathrm{OT}_\varepsilon = \inf_{\pi \in \Pi(\mu_1,\mu_2)} \left\{ \int_{X_1 \times X_2} c\, d\pi + \varepsilon\, \mathrm{KL}(\pi, \mu_1 \otimes \mu_2) \right\}, \tag{1.1}$$

其中， $\Pi(\mu_1, \mu_2)$是给定边缘分布 μ_1 和 μ_2的耦合集合。此外， $\varepsilon > 0$是正则化参数, KL 表示传输计划 π 相对于乘积测度 $\mu_1 \otimes \mu_2$的 Kullback–Leibler 散度 [21]; 对于一般的概率测度，其定义如下:

$$\mathrm{KL}(\mu, \nu) = \begin{cases} \displaystyle\int \frac{d\mu}{d\nu} \log \frac{d\mu}{d\nu}\, d\nu & \text{if } \mu \ll \nu, \\ \infty & \text{otherwise.} \end{cases} \tag{1.2}$$

当 $\varepsilon = 0$时，我们注意到，问题 (1.1) 与以下 Monge–Kantorovich 问题一致（标准教材可参见 [1, 17, 31, 41, 42] 及其中的参考文献）:

$$\mathrm{OT} = \inf_{\pi \in \Pi(\mu_1,\mu_2)} \int_{X_1 \times X_2} c\, d\pi. \tag{1.3}$$

最早将熵正则化引入最优传输问题的开创性工作，出自最著名的物理学家之一 E. Schr¨odinger，他从统计物理的角度研究了这一问题 [32]。Schr¨odinger 研究了以下最小化问题:

$$\inf_{\pi \in \Pi(\mu_1,\mu_2)} \varepsilon \mathrm{KL}(\pi, \mathcal{K}), \tag{1.4}$$

其中， $\mathcal{K} = e^{-\frac{c}{\varepsilon}} \mu_1 \otimes \mu_2$是吉布斯核。由于熵正则化最优传输 (1.1) 等价于 (1.4)，它通常被称为 Schr¨odinger 桥问题（综述论文可参见 [9, 22, 26] 及其中的参考文献）。

通过熵对线性规划进行凸松弛的问题，已在文献中得到广泛研究 [6, 11, 16, 37, 44]。Cuturi [12]，作出了一项重要贡献，他证明熵正则化最优传输 (1.1) 可以通过 Sinkhorn–Knopp算法 [33, 34] 高效计算。这一突破推动了熵正则化最优传输在图像处理、计算机图形学和机器学习等多个领域的广泛应用（例如，参见 [2, 27, 35]）。若要全面了解这一主题，感兴趣的读者可参阅 Peyre–Cuturi [28] 的著作及其中的参考文献。

(1.1) 关于 $\varepsilon \to 0$ 的收敛速率，从数学和应用两个角度都引起了广泛关注，并一直是活跃的研究课题。在离散情形下，Cominetti– San Mart´ın [11] 研究了收敛速率问题。他们证明了收敛速率呈指数级。Weed [43] 对收敛速率的定量估计进行了研究。

在连续情形下，Carlier–Duval–Peyré–Schmitzer [7]证明了在二次成本函数的情形下（即，$c(x,y) = |x-y|^2$），(1.1) 向 Monge–Kantorovich 问题 (1.3) 的 Γ-收敛。因此，他们证明了 OT$_\varepsilon$ 收敛于 OT。本文引入的分块近似方法，是估计正则化最优传输中收敛速率的一种标准技术。Carlier–Pegon–Tamanini [8] 的论文确立了 (1.1) 的以下收敛速率：

$$\mathrm{OT}_\varepsilon - \mathrm{OT} \le D\varepsilon \log \frac{1}{\varepsilon} + O(\varepsilon), \tag{1.5}$$

其中常数 D 是边缘分布的 Entropy维数的最小值。他们的方法基于分块近似法，并且还证明了估计式(1.5) 中的首阶项 $\varepsilon \log 1/\varepsilon$ 是尖锐的。最近，Eckstein–Nutz [15, 14] 提出了一种基于量化和影子论证的获取收敛速率的新方法。这种方法能够推导出与 (1.5)中相同的结-果，并且可以应用于广泛的正则化项，因此是一种非常有用的工具。这些方法在本文中具有重要意义。

因此，采用 KL 散度进行正则化的传输问题历史悠久，且已得到广泛研究。近年来，从计算最优传输的角度出发，对其他正则项的研究也逐渐受到关注。特别是，使用 Tsallis 相对熵进行正则化的方法作为 KL 散度的一参数扩展，已从熵传输计划的稀疏性和计算复杂度方面得到研究 [3, 24]。

然而，据我们所知，除多元正态分布的情形 [38] 外，鲜有研究关注连续情形下 Tsallis 正则化的性质及其与 KL 正则化的差异。本文考虑以下采用 Tsallis 相对熵正则化的最优传输问题：

$$\mathrm{OT}_{q,\varepsilon} = \inf_{\pi \in \Pi(\mu_1,\mu_2)} \left\{ \int_{X_1 \times X_2} c\, d\pi + \varepsilon D_q(\pi, \mu_1 \otimes \mu_2) \right\}, \tag{1.6}$$

其中， D_q 表示传输计划 π 相对于乘积测度 $\mu_1 \otimes \mu_2$ 的 Tsallis 相对熵，对一般概率测度，其定义如下：

$$D_q(\mu,\nu) = \begin{cases} \dfrac{1}{q-1}\displaystyle\int \left[\left(\frac{d\mu}{d\nu}\right)^q - \frac{d\mu}{d\nu}\right] d\nu & \text{if } \mu \ll \nu \text{ and } q > 1, \\ \infty & \text{otherwise.} \end{cases} \tag{1.7}$$

相对熵 (1.7) 是 KL 散度的一种推广，出现在 Tsallis [39] 提出的所谓 Tsallis 统计力学中。它不仅在非线性动力学和非平衡物理学领域得到了广泛研究，也在信息论领域受到广泛关注（参见 [5, 19, 18, 25, 30, 36, 40] 及其中的参考文献）。

我们引入 q-对数函数和 q-指数函数，它们在 Tsallis 统计力学领域已得到充分研究。对于满足 $q \neq 1$ 的参数 $q \in \mathbb{R}$ ， q-对数函数和 q-指数函数定义如下：

$$\log_q y = \frac{y^{1-q}-1}{1-q} \quad (y>0), \quad \exp_q y = [1+(1-q)y]_+^{\frac{1}{1-q}} \quad (y \in \mathbb{R}),$$

其中，我们定义 $f_+(x) = \max\{f(x)\ 0\}$。当 $q \to 1$ 时，它们还原为对数函数和指数函数的自然定义。 D_q 可以用 q-对数函数表示如下：

$$D_q(\mu,\nu) = \int \frac{d\mu}{d\nu} \log_{2-q}\left(\frac{d\mu}{d\nu}\right) d\nu$$

对于 $\mu \ll \nu$ 和 $q > 1$。本文旨在运用 Eckstein–Nutz [15, 14] 提出的方法，研究 Tsallis 熵正则化最优传输 (1.6) 的基本性质。具体而言，我们采用他们的技巧，证明当正则化参数 ε 趋于零时的 Γ-收敛，并展示极小化子的收敛性。此外，我们还建立了 Tsallis 熵正则化最优传输的收敛速率，并将其与 KL 散度的收敛速率进行比较。通过聚焦这些方面，我们旨在为 Tsallis 相对熵在最优传输中的应用奠定理论基础。

令 $p \in [1,\infty)$ 和 $\mathcal{P}_p(X)$ 表示波兰空间 X 上具有有限 p 阶矩的概率测度集合。令 $\mathcal{P}^n(X) \subset \mathcal{P}(X)$ 表示至多由 n 个点支撑的概率测度集合。我们还假设代价函数 c

满足($A_{L,C}$) 条件，其中某些常数 $L, C \geq 0$ 在 $[15, \text{定理 } 3.2]$, 中引入，即对于任意 $\mu_i, \tilde{\mu}_i \in \mathcal{P}_p(X_i)$ $(i = 1,2)$，满足 $W_p(\tilde{\mu}_i, \mu_i) \leq C$，有

$$\left| \int_{X_1 \times X_2} c \, d(\pi - \tilde{\pi}) \right| \leq L W_p(\pi, \tilde{\pi}) \tag{1.8}$$

对于所有的 $\pi \in \Pi(\mu_1, \mu_2)$ 和 $\tilde{\pi} \in \Pi(\tilde{\mu}_1, \tilde{\mu}_2)$，其中 $W_p(\pi, \tilde{\pi})$ 是 π 与 $\tilde{\pi}$ 之间的第 p 个 Wasserstein 距离。例如，如果代价函数 c 是一个 L-Lipschitz 函数，或 $c(x, y)$ 在 $X_1 \times X_2$ 上对于 $1 \leq p < \infty$ 满足 $= |x - y|^p$，那么 c 满足（$A_{L,c}$）条件。在后一种情况下，常数 L 取决于 μ_i 的矩以及 C。

下面我们介绍本文的主要结果。首先，我们考虑 Tsallis熵正则最优传输（1.6）的 Γ-收敛性。让我们回顾 Γ-收敛的定义。

定义1.1（Γ-收敛）。设一列由 $F_k : \mathcal{P}_p(X_1 \times X_2) \to \mathbb{R} \cup \{\infty\}$ 定义的泛函 $\{F_k\}_{k \in \mathbb{N}}$ 为

$$F_k(\pi) = \begin{cases} \displaystyle\int_{X_1 \times X_2} c \, d\pi + \varepsilon_k D_q(\pi, \mu_1 \otimes \mu_2) & \text{if } \pi \in \Pi(\mu_1, \mu_2) \\ +\infty & \text{otherwise,} \end{cases}$$

其中，当 $\varepsilon_k \to 0$ 趋于 $k \to \infty$时，并且还将 F: $\mathcal{P}_p(X_1 \times X_2) \to \mathbb{R} \cup \{\infty\}$ 定义为

$$F(\pi) = \begin{cases} \displaystyle\int_{X_1 \times X_2} c \, d\pi & \text{if } \pi \in \Pi(\mu_1, \mu_2) \\ +\infty & \text{otherwise.} \end{cases}$$

若满足以下两个条件，我们称序列 $\{F_k\}_{k \in \mathbb{N}}$ Γ关于窄拓扑 F 收敛:

（下极限不等式）对于任意序列 $\{\pi_k\}_{k \in \mathbb{N}} \subset \mathcal{P}_p(X_1 \times X_2)$，若 $\pi_k \to \pi$在窄拓扑下收敛,

$$F(\pi) \leq \liminf_{k \to \infty} F_k(\pi_k). \tag{1.9}$$

(上极限不等式) 对于任意 $\pi \in \mathcal{P}_p(X_1 \times X_2)$，存在一个序列 $\{\pi_k\}_{k \in \mathbb{N}} \subset \mathcal{P}_p(X_1 \times X_2)$ 窄收敛于 $\pi_k \to \pi$，使得

$$F(\pi) \geq \limsup_{k \to \infty} F_k(\pi_k). \tag{1.10}$$

根据 Γ-收敛定义的直接推论，我们得到以下定理。

定理 1.2. 设 $\mu_i \in \mathcal{P}_p(X_i)$ $(i = 1, 2)$ 且 $1 < q \leq 2$。则，序列$\{F_k\}_{k\in\mathbb{N}}$ Γ在窄拓扑意义下收敛于 F。此外， F_k的唯一极小化子 $\pi_k^* \in \Pi(\mu_1, \mu_2)$ 窄收敛于 F的一个极小化子 $\pi^* \in \Pi(\mu_1, \mu_2)$。

此外，利用 Eckstein–Nutz [14, 15], 提出的量化和影子论证，可以得到 Tsallis 熵正则化最优传输 (1.6) 的收敛速率。为陈述该定理，我们回顾概率测度的量化，这一概念源自信号处理（见 [20]）。概率测度的量化是一个与最优意义下用给定数量 n 个离散概率逼近一个 d 维概率测度 μ 相关的主题。

定义 1.3（量化）。对于 $\mu \in \mathcal{P}_p(X)$，我们称 μ 满足量化条件 $\text{quant}_p(C, \beta)$，其中常数为 $C \geq 0$ 和 $\beta > 0$，如果存在 $\{\mu^n\}_{n\in\mathbb{N}} \subset \mathcal{P}^n(X)$ 使得

$$W_p(\mu^n, \mu) \leq Cn^{-\frac{1}{\beta}}. \tag{1.11}$$

注 1.4。令 $X = \mathbb{R}^d$ 和 $\mu \in \mathcal{P}_{p+\delta}(\mathbb{R}^d)$ 对某个 $\delta > 0$成立。于是可以取$\beta = d$，且 μ 满足$\text{quant}_p(C, d)$（参见 *[20,* 定理 6.*2]*）。

于是，得到以下定理:

定理 1.5. 设 $\mu_i \in \mathcal{P}_p(X_i)$ $(i = 1, 2)$。假设 μ_2 和代价函数 c分别满足 $\text{quant}_p(C, \beta)$ 和 $(\mathbf{A}_{L,C})$ 条件，其中常数分别为 $\beta \geq 1$和 $L, C \geq 0$。那么对于 $1 < q \leq 2$,

$$\mathrm{OT}_{q,\varepsilon} - \mathrm{OT} \leq K_{1,q}\frac{\beta}{(q-1)\beta+1}\varepsilon \log_{\frac{1}{(q-1)\beta+1}} \frac{1}{\varepsilon} + \frac{K_{1,q}-1}{q-1}\varepsilon + K_{2,q}\varepsilon^{\frac{1}{(q-1)\beta+1}}, \tag{1.12}$$

其中

$$K_{1,q} = \left(\frac{2LC}{\beta}\right)^{\frac{(q-1)\beta}{(q-1)\beta+1}} \quad \text{and} \quad K_{2,q} = (2LC)^{\frac{(q-1)\beta}{(q-1)\beta+1}}\beta^{\frac{1}{(q-1)\beta+1}}.$$

注 1.6. 我们指出，根据 *[15,* 定理 3.*3],*

$$\mathrm{OT}_\varepsilon - \mathrm{OT} \leq \beta\varepsilon \log \frac{1}{\varepsilon} + 4LC\varepsilon. \tag{1.13}$$

由于 $q>1$，我们注意到 $\tilde{q}:=\frac{1}{(q-1)\beta+1}<1$。因此我们得到

$$\lim_{x\to\infty}\frac{\log x}{\log_{\tilde{q}} x}=\lim_{x\to\infty}\frac{(1-\tilde{q})\log x}{x^{1-\tilde{q}}-1}=\lim_{x\to\infty}\frac{\frac{1}{x}}{x^{-\tilde{q}}}=\lim_{x\to\infty}x^{\tilde{q}-1}=0.$$

因此，在考虑 $\varepsilon\to 0$时，收敛速率(1.13) 比 (1.12) 更快。由于 KL散度 (1.2) 是Tsallis 相对熵 (1.7) 的极限情形，我们可以得出结论：就 Tsallis 相对熵而言，KL 具有最快的收敛速率。

备注1.7。虽然 *[15],* 中已有针对 $\rho>1$ 的最优传输 L^ρ 正则化的类似于 (1.12) 的结果（见 *[15]* 中的例3.1），但我们计算了 (1.12) 中的显式常数。我们指出，由于 $\rho\to 1$, L^ρ 正则化所得的估计与 (1.13) 并不一致；而在我们的设定中，当 $q\to 1$时，可以由 (1.12) 推导出 (1.13)。事实上，我们观察到 $K_{1,\,q}\to 1$ 以及 $K_{2,\,q}\to\beta$，因为 $q\to 1$。此外，当 $q\to 1$ 时， q-对数函数的对数函数收敛，这意味着 Tsallis 熵正则化收敛速率 (1.12) 右侧的首项收敛到 $\beta\varepsilon\log(\frac{1}{\varepsilon})$。因此，当 $q\to 1$时，Tsallis 熵正则化 (1.12) 的首项与 KL 散度 (1.13) 的首项具有相同的阶。

此外，可以证明 Tsallis 熵正则化最优传输(1.12)的收敛速率具有尖锐性。

定理1.8（收敛速率的尖锐性）。设 $X_1=X_2=\mathbb{R}^d$，其中 $\mu_1=\mu_2$是 $[0,1]^d$上的均匀分布，且 $c(x,y)=\sum_{i=1}^d|x_i-y_i|$。那么，对于 $q>1$,

$$\mathrm{OT}_{q,\varepsilon}-\mathrm{OT}\geq\tilde{K}_{1,q}\varepsilon^{\frac{1}{(q-1)d+1}}-\tilde{K}_{2,q}\varepsilon,$$

其中

$$\tilde{K}_{1,q}=\frac{(q-1)d+1}{(q-1)d+q}\left[\frac{q-1}{((q-1)d+q)2^{\frac{(q-1)d+1}{q-1}}}\left(\frac{q}{q-1}\right)^{\frac{q}{q-1}}\right]^{\frac{q-1}{(q-1)d+1}},$$
$$\tilde{K}_{2,q}=\left(1+2^{\frac{(q-1)d+1}{q-1}}\right)q^{-\frac{q}{q-1}}.$$

本文结构如下。第 2 节回顾数据处理不等式，以及 Eckstein–Nutz [14, 15] 提出的耦合的影子的定义和性质。第 3 节通过量化和影子论证，证明 Tsallis 熵正则化最优传输 (1.6) 的 Γ-收敛。第 4 节给出 Tsallis 熵正则化最优传输 (1.12) 的收敛速率。可以看到，采用 KL 正则化的最优传输代价 (1.13) 的收敛速度快于采用 Tsallis 熵正则化的最优传输代价 (1.12)。第 5 节证明第 4 节所得收敛速率的尖锐性。

2 预备知识

接下来，我们将介绍数据处理不等式，并解释 Eckstein–Nutz [14, 15] 引入的耦合影子的论证。

为简便起见，我们记作

$$f_q(x) = \begin{cases} \dfrac{x^q - x}{q-1} & \text{if } q > 1, \\ x \log x & \text{if } q = 1 \end{cases} \quad \text{and} \quad \phi_q(x) = \frac{f_q(x)}{x} = \begin{cases} \dfrac{x^{q-1} - 1}{q-1} & \text{if } q > 1, \\ \log x & \text{if } q = 1. \end{cases}$$

于是，Tsallis 相对熵 (1.7) 可表示为

$$D_q(\mu, \nu) = \int f_q\left(\frac{d\mu}{d\nu}\right) d\nu = \int \frac{d\mu}{d\nu} \phi_q\left(\frac{d\mu}{d\nu}\right) d\nu = \int \phi_q\left(\frac{d\mu}{d\nu}\right) d\mu.$$

注意，当 $1 \le q \le 2$ 时，ϕ_q 是凹函数；而对于任意 $q \ge 1$，f_q 都是凸函数。

2.1 马尔可夫核与数据处理不等式

我们解释马尔可夫核（也称为随机核）以及数据处理不等式。这些概念和不等式是信息论中的重要工具，并且在控制本文中的 Tsallis 相对熵 (1.7) 方面发挥着重要作用。感兴趣的读者可参阅 [10, 29]。

定义 2.1（马尔可夫核）。设 X_1 和 X_2 为波兰空间。映射 K称为马尔可夫核，若满足以下两个条件：

1. 对每个可测集 B，映射 $x \mapsto K(x, B)$ 是 X_1 上的可测函数。

2. 对于每个 $x \in X_1$，映射 $B \mapsto K(x, B)$ 是 X_2 上的概率测度。

马尔可夫核可以解释为随机操作或变换。概率测度 $\pi \in \mathcal{P}(X_1 \times X_2)$可以通过一个马尔可夫核进行分解。根据 [10, 第四章定理2.18], 存在一个概率测度 $\mu_1 \in \mathcal{P}(X_1)$ 和一个马尔可夫核 $K: X_1 \to \mathcal{P}(X_2)$，使得

$$\pi(A \times B) = \int_A \mu_1(dx) K(x, B) \quad A \subset X_1,\ B \subset X_2. \tag{2.14}$$

我们将 $\pi = \mu_1 \otimes K \in \mathcal{P}(X_1 \times X_2)$ 的第二边缘记为 $\mu_1 K \in \mathcal{P}(X_2)$。利用耦合的解体，可以得到 Tsallis 相对熵 (1.7) 的界。

引理 2.2.设 $\pi \in \Pi(\mu_1, \mu_2)$。如果 $\mu_2 \in \mathcal{P}^n(X_2)$，那么对于 $1 < q \leq 2$,

$$D_q(\pi, \mu_1 \otimes \mu_2) \leq \phi_q(n). \tag{2.15}$$

引理 2.2 的证明 令 σ 为 μ_2 支撑集上的计数测度，并令 $\{x_i\}_{i=1}^n$ 为 μ_2 的支撑集。由于 $\mu_2 \in \mathcal{P}^n(X_2)$，不妨设 $\sigma = \Sigma_{i=1}^n \delta_{x_i}$ 且 $\mu_2 = \Sigma_{i=1}^n \alpha_i \delta_{x_i}$，其中 $\Sigma_{i=1}^n \alpha_i = 1$。考虑 π 的解体: $\mu_1 \otimes K$。由于 $K: X_1 \to \mathcal{P}^n(X_2)$，我们可以写成 $K = \Sigma_{i=1}^n \beta_i \delta_{x_i}$，其中 $\Sigma_{i=1}^n \beta_i = 1$。于是有

$$\frac{d\pi}{d(\mu_1 \otimes \mu_2)} = \frac{d\mu_1}{d\mu_1} \frac{dK}{d\mu_2} = \frac{dK}{d\mu_2} = \frac{\beta_i}{\alpha_i} \leq \frac{1}{\alpha_i} = \frac{d\sigma}{d\mu_2}. \tag{2.16}$$

由于函数 ϕ_q 单调不减，由 (2.16) 可得

$$D_q(\pi, \mu_1 \otimes \mu_2) = \int_{X_1 \times X_2} \phi_q \left(\frac{d\pi}{d(\mu_1 \otimes \mu_2)} \right) d\pi \leq \int_{X_1 \times X_2} \phi_q \left(\frac{d\sigma}{d\mu_2} \right) d\pi.$$

由 ϕ_q 的凹性，Jensen 不等式给出

$$\begin{aligned} D_q(\pi, \mu_1 \otimes \mu_2) &\leq \int_{X_1 \times X_2} \phi_q \left(\frac{d\sigma}{d\mu_2} \right) d\pi = \int_{X_2} \phi_q \left(\frac{d\sigma}{d\mu_2} \right) d\mu_2 \\ &\leq \phi_q \left(\int_{X_2} d\sigma \right) = \phi_q(n), \end{aligned}$$

由此可知，不等式 (2.15) 成立。

引理 2.3（数据处理不等式）。设 μ、$\nu \in \mathcal{P}(X_1)$，且令 $K\colon X_1 \to \mathcal{P}(X_2)$ 为一个马尔可夫核。则对于 Tsallis 相对熵 (1.7)，以下不等式成立：

$$D_q(\mu K, \nu K) \leq D_q(\mu, \nu). \tag{2.17}$$

引理 2.3 的证明 此证明与 [14, 引理 4.1] 中关于 Tsallis 相对熵 (1.7) 的证明方式相同。我们可以假设 $\mu \ll \nu$。对于任意马尔可夫核 $K_1 \ll K_2\colon X_1 \to \mathcal{P}(X_2)$，有

$$\frac{d(\mu \otimes K_1)}{d(\nu \otimes K_2)} = \frac{d\mu}{d\nu}\frac{dK_1}{dK_2} \quad \nu \otimes K_2\text{-a.s.}$$

注意到

$$\begin{aligned} D_q(\mu \otimes K, \nu \otimes K) &= \int_{X_1 \times X_2} f_q\left(\frac{d(\mu \otimes K)}{d(\nu \otimes K)}\right) d(\nu \otimes K) \\ &= \int_{X_1} f_q\left(\frac{d\mu}{d\nu}\right) d\nu = D_q(\mu, \nu), \end{aligned} \tag{2.18}$$

其中用到了 $\int_{X_2} dK = 1$。由于当 $q > 1$ 时，f_q 是凸函数，因此根据 Jensen 不等式，有

$$\begin{aligned} D_q(\mu \otimes K_1, \nu \otimes K_2) &= \int_{X_1 \times X_2} f_q\left(\frac{d\mu}{d\nu}\frac{dK_1}{dK_2}\right) d(\nu \otimes K_2) \\ &\geq \int_{X_1} f_q\left(\frac{d\mu}{d\nu}\right) d\nu = D_q(\mu, \nu). \end{aligned} \tag{2.19}$$

$\tilde{}$ $\tilde{}$分别将 $\mu \otimes K$和 $\nu \otimes K$从第二边缘到第一边缘解体为 $(\mu K) \otimes K_1$和 $(\nu K) \otimes K_2$。应用 (2.19)，有

$$D_q(\mu \otimes K, \nu \otimes K) = D_q\big((\mu K) \otimes \tilde{K}_1, (\nu K) \otimes \tilde{K}_2\big) \geq D_q(\mu K, \nu K) \tag{2.20}$$

结合不等式 (2.18) 和 (2.19)，得到不等式 (2.17)。 □

2.2 影子

影子的理念是考虑将一个耦合进行适当的 W_p 投影 π 到其他边缘分布的耦合集合上。

定义 2.4（[14, 定义 3.1]）。设 $1 \le p < \infty$ 和 $\mu_i, \tilde{\mu}_i \in \mathcal{P}_p(X_i)$ $(i = 1, 2)$。设 $\kappa_i \in \Pi(\mu_i, \tilde{\mu}_i)$ 是 $W_p(\mu_i, \tilde{\mu}_i)$ 的一个最优耦合，并设 $\kappa_i = \mu_i \otimes K_i$ 是一个解体。对于 $\pi \in \Pi(\mu_1, \mu_2)$，其影子 $\tilde{\pi} \in \Pi(\tilde{\mu}_1, \tilde{\mu}_2)$ 定义为 $\pi \otimes K \in \mathcal{P}_p((X_1 \times X_2)^2)$ 的第二边缘，其中核 $K\colon X_1 \times X_2 \to \mathcal{P}_p(X_1 \times X_2)$ 定义为 $K(x) = K_1(x_1) \otimes K_2(x_2)$。

回顾影子的性质。利用影子的论证和数据处理不等式 (2.17)，可以控制 Tsallis 相对熵D_q。这些性质在本文中始终发挥着重要作用。

引理 2.5 ([14, 引理 3.2])。设 $1 \le p < \infty$ 和 $\mu_i, \tilde{\mu}_i \in \mathcal{P}_p(X_i)$ $(i = 1, 2)$。对于任意 $\pi \in \Pi(\mu_1, \mu_2)$，其影子 $\tilde{\pi} \in \Pi(\tilde{\mu}_1, \tilde{\mu}_2)$ 满足以下性质：

$$W_p(\pi, \tilde{\pi})^p = W_p(\mu_1, \tilde{\mu}_1)^p + W_p(\mu_2, \tilde{\mu}_2)^p, \tag{2.21}$$

$$D_q(\tilde{\pi}, \tilde{\mu}_1 \otimes \tilde{\mu}_2) \le D_q(\pi, \mu_1 \otimes \mu_2). \tag{2.22}$$

3 Γ-收敛：Tsallis 熵正则化最优传输

本节证明 Tsallis 熵正则化最优传输 (1.6) 的 Γ-收敛。定理 1.2 的证明可分为以下两个命题（下极限不等式 (1.9) 和上极限不等式 (1.10)）：

命题3.1。对于任意序列 $\{\pi_k\}_{k\in\mathbb{N}} \subset \mathcal{P}_p(X_1 \times X_2)$，使得 $\pi_k \to \pi$ 窄收敛，我们有

$$F(\pi) \le \liminf_{k\to\infty} F_k(\pi_k).$$

命题 3.1 的证明 Jensen 不等式表明，

$$\begin{aligned} D_q(\pi_k, \mu_1 \otimes \mu_2) &= \int_{X_1\times X_2} f_q\left(\frac{d\pi_k}{d(\mu_1 \otimes \mu_2)}\right) d(\mu_1 \otimes \mu_2) \\ &\ge f_q\left(\int_{X_1\times X_2} d\pi_k\right) = f_q(1) = 0. \end{aligned}$$

因此，我们得到

$$\liminf_{k\to\infty} \varepsilon_k D_q(\pi_k, \mu_1 \otimes \mu_2) \ge 0. \tag{3.23}$$

此外，由耦合的下半连续性（见 [41, 引理 4.3]）可得，

$$\int_{X_1\times X_2} c\,d\pi \le \liminf_{k\to\infty} \int_{X_1\times X_2} c\,d\pi_k. \tag{3.24}$$

结合 (3.23) 和 (3.24)，下极限不等式 (1.9) 成立。 □

命题 3.2. 设 $\mu_i \in \mathcal{P}_p(X_i)$ $(i = 1, 2)$，且 $\{\varepsilon_k\}_{k\in\mathbb{N}}$ 是一个收敛到零并满足 $\varepsilon_k\phi_q(k) \to 0$ 的正数列。假设 $1 < q \le 2$。对于任意 $\pi \in \mathcal{P}_p(X_1 \times X_2)$，存在一个序列 $\{\pi_k\}_{k\in\mathbb{N}} \subset \mathcal{P}_p(X_1 \times X_2)$ 弱收敛到 $\pi_k \to \pi$ ，使得

$$F(\pi) \ge \limsup_{k\to\infty} F_k(\pi_k).$$

证明（命题 3.2）由假设及量化论证可知，存在 $\{\mu_2^k\}_{k\in\mathbb{N}} \subset \mathcal{P}_p^k$ 使得

$$\lim_{k\to\infty} W_p(\mu_2^k, \mu_2) = 0$$

（见 [41, 定理 6.18]）。对于任意 $\pi \in \Pi(\mu_1, \mu_2)$，我们取 π 的影子 $\tilde{\pi} \in \Pi(\mu_1, \mu_2^k)$。再次取 $\tilde{\pi}$ 的影子 $\pi_k \in \Pi(\mu_1, \mu_2)$，由关于 Wasserstein 距离的三角不等式（见 [42, 定理 7.3]）和引理 2.5 可得

$$W_p(\pi_k, \pi) \le W_p(\pi_k, \tilde{\pi}) + W_p(\tilde{\pi}, \pi) \le 2W_p(\mu_2^k, \mu_2) \to 0 \text{ as } k \to \infty.$$

因此，由Wasserstein 距离的一个性质（见 [42, 定理 7。12]）可知

$$\pi_k \to \pi \text{ narrowly as } k \to \infty.$$

此外，由于代价函数 c 满足 $(\mathbf{A}_{L,C})$ 条件，我们有

$$\left|\int_{X_1\times X_2} c\,d(\pi_k - \pi)\right| \le LW_p(\pi_k, \pi) \to 0 \text{ as } k \to \infty.$$

因此，我们得到

$$\int_{X_1\times X_2} c\,d\pi_k \to \int_{X_1\times X_2} c\,d\pi \text{ as } k \to \infty. \tag{3.25}$$

另一方面，引理 2.2 和引理 2.5 表明

$$D_q(\pi_k, \mu_1 \otimes \mu_2) \leq D_q(\tilde{\pi}, \mu_1 \otimes \mu_2^k) \leq \phi_q(k).$$

于是，我们得到

$$\limsup_{k\to\infty} \left(\varepsilon_k D_q(\pi_k, \mu_1 \otimes \mu_2)\right) \leq \limsup_{k\to\infty} \left(\varepsilon_k \phi_q(k)\right) = 0. \tag{3.26}$$

结合（3.25）和（3.26），上极限不等式（1.10）成立。 □

定理 1.2 的证明 命题 3.1、3.2 表明，$\{F_k\}_{k\in\mathbb{N}}$ Γ关于窄拓扑F 收敛于 π_k^* 。此外，由于 Tsallis 相对熵（1.7）的严格凸性，F_k是 Π的唯一极小化子。由于μ_1 (μ_2, [41,)是紧的（见引理 4.4]），序列 $\{\pi_k^*\}_{k\in\mathbb{N}} \subset$ Π(μ_1, μ_2)关于窄拓扑收敛于耦合$\pi^* \in$ Π(μ_1, μ_2)。利用 Γ-收敛的性质，π^* 是 F的一个极小化子。这就是所需的结论。

□

4 Tsallis 熵正则化最优传输的收敛速率

在本节中，我们证明定理 1.5。在定理 1.5 的证明中，我们采用文献 [15, 14] 提出的耦合的双重影子论证。

定理 1.5 的证明 令 $\pi^* \in$ Π(μ_1, μ_2) 为 OT 的最优耦合。由于 μ_2 满足 quant$_p$ (C, β)，因此存在 $\{\mu_2^n\}_{n\in\mathbb{N}} \subset \mathcal{P}^n(X_2)$ 使得

$$W_p(\mu_2^n, \mu_2) \leq Cn^{-\frac{1}{\beta}}. \tag{4.27}$$

对于 $n \in \mathbb{N}$，我们取 π^* 的一个耦合 $\pi \in$ Π(μ_1, μ_2) 作为双重影子：令 $\tilde{\pi} \in$ Π(μ_1, μ_2^n) 为 π^* 的一个影子，并将 $\pi \in$ Π(μ_1, μ_2) 定义为 $\tilde{\pi}$ 的一个影子。根据三角不等式和引理 2.5，我们有

$$W_p(\pi, \pi^*) \leq W_p(\pi, \tilde{\pi}) + W_p(\tilde{\pi}, \pi^*) \leq 2W_p(\mu_2^n, \mu_2). \tag{4.28}$$

由于代价函数 c 满足 (A$_{L,C}$) 条件，由 (4.27) 和 (4.28) 可得，

$$\int_{X_1\times X_2} c\, d\pi - \int_{X_1\times X_2} c\, d\pi^* \leq 2LW_p(\mu_2^n, \mu_2) \leq 2LCn^{-\frac{1}{\beta}}. \tag{4.29}$$

另一方面，由于 π 是 $\tilde{\pi}$ 的影子，引理 2.5 再次给出

$$D_q(\pi, \mu_1 \otimes \mu_2) \leq D_q(\tilde{\pi}, \mu_1 \otimes \mu_2^n). \tag{4.30}$$

因此，结合 (4.29) 和 (4.30)，可得

$$\mathrm{OT}_{q,\varepsilon} - \mathrm{OT} \leq 2LCn^{-\frac{1}{\beta}} + \varepsilon D_q(\tilde{\pi}, \mu_1 \otimes \mu_2^n).$$

由引理 2.2 可知

$$D_q(\tilde{\pi}, \mu_1 \otimes \mu_2^n) \leq \phi_q(n),$$

因此，我们得到

$$\mathrm{OT}_{q,\varepsilon} - \mathrm{OT} \leq 2LCn^{-\frac{1}{\beta}} + \varepsilon\phi_q(n). \tag{4.31}$$

那么，不等式 (4.31) 意味着

$$\mathrm{OT}_{q,\varepsilon} - \mathrm{OT} \leq 2LCn^{-\frac{1}{\beta}} + \frac{\varepsilon}{q-1}(n^{q-1} - 1).$$

若取自然数 $n \in \mathbb{N}$ 为

$$n \sim \left(\frac{2LC}{\beta}\right)^{\frac{\beta}{(q-1)\beta+1}} \varepsilon^{-\frac{\beta}{(q-1)\beta+1}},$$

则可得到 Tsallis 熵正则化最优传输 (1.6) 的收敛速率：

$$\mathrm{OT}_{q,\varepsilon} - \mathrm{OT} \leq \frac{1}{q-1}\left(K_{1,q}\varepsilon^{\frac{1}{(q-1)\beta+1}} - \varepsilon\right) + K_{2,q}\varepsilon^{\frac{1}{(q-1)\beta+1}}, \tag{4.32}$$

其中常数$K_{1,q}$ 和 $K_{2,q}$ 定义为

$$K_{1,q} = \left(\frac{2LC}{\beta}\right)^{\frac{(q-1)\beta}{(q-1)\beta+1}} \quad \text{and} \quad K_{2,q} = (2LC)^{\frac{(q-1)\beta}{(q-1)\beta+1}} \beta^{\frac{1}{(q-1)\beta+1}}.$$

当我们在收敛速率 (4.32) 右侧的第一项中使用 q-对数函数时，可以将其改写如下：

$$\begin{aligned}
&\frac{1}{q-1}\left(K_{1,q}\varepsilon^{\frac{1}{(q-1)\beta+1}} - \varepsilon\right) \\
&\quad = K_{1,q}\frac{\varepsilon}{q-1}\left(\varepsilon^{-\left(1-\frac{1}{(q-1)\beta+1}\right)} - 1\right) + \frac{\varepsilon}{q-1}(K_{1,q} - 1) \\
&\quad = K_{1,q}\frac{1-\frac{1}{(q-1)\beta+1}}{q-1}\frac{\varepsilon}{1-\frac{1}{(q-1)\beta+1}}\left(\varepsilon^{-\left(1-\frac{1}{(q-1)\beta+1}\right)} - 1\right) + \frac{\varepsilon}{q-1}(K_{1,q} - 1) \\
&\quad = K_{1,q}\frac{\beta}{(q-1)\beta+1}\varepsilon\log_{\frac{1}{(q-1)\beta+1}}\frac{1}{\varepsilon} + \frac{\varepsilon}{q-1}(K_{1,q} - 1).
\end{aligned}$$

这就是所需的结论。 □

5 Tsallis 熵正则化最优传输收敛率的尖锐性

本节中，我们证明 Tsallis 熵正则化最优传输（1.12）的收敛速率具有尖锐性。为考察收敛速率的尖锐性，我们引入熵正则化最优传输问题的对偶问题（见[13,定理3.6]）。

对于凸函数 f， f 的勒让德变换定义为

$$f^*(y) = \sup_{x\geq 0} \left\{xy - f(x)\right\},\ y \in \mathbb{R}.$$

通过考虑 f_q 的勒让德变换，可以将 $\mathrm{OT}_{q,\ \varepsilon}$ 的对偶问题表示为

$$\begin{aligned}\mathrm{OT}_{q,\varepsilon} = \sup_{(h_1,h_2)\in L^1(\mu_1)\times L^1(\mu_2)} \Bigg\{ &\int_{X_1} h_1(x)\, d\mu_1(x) + \int_{X_2} h_2(y)\, d\mu_2(y) \\ &-\varepsilon \int_{X_1\times X_2} f_q^*\left(\frac{h_1(x)+h_2(y)-c(x,y)}{\varepsilon}\right) d\mu_1(x)\, d\mu_2(y)\Bigg\},\end{aligned} \tag{5.33}$$

其中 f_q^* 定义为

$$f_q^*(y) = \left(\frac{q-1}{q}y + \frac{1}{q}\right)^{\frac{q}{q-1}} = q^{-\frac{q}{q-1}}\left[1+(q-1)y\right]_+^{\frac{q}{q-1}} = q^{-\frac{q}{q-1}}\exp_{2-q}(y)^q$$

对于 $y \in \mathbb{R}$。此外，OT 的对偶问题为

$$\mathrm{OT} = \sup_{\substack{(h_1,h_2)\in L^1(\mu_1)\times L^1(\mu_2),\\ h_1+h_2\leq c}} \left\{\int_{X_1} h_1(x)\, d\mu_1(x) + \int_{X_2} h_2(y)\, d\mu_2(y)\right\}.$$

定理1.8的证明 由于 $X_1 = X_2 = \mathbb{R}^d$ 和 $\mu_1 = \mu_2$ 是 $[0,1]^d$上的均匀分布，我们看到$\mathrm{OT} = 0$。如果对于对偶问题（5.33）中的任意 $a \in \mathbb{R}$，我们设定 $h_1 = a$ 和 $h_2 = 0$，则有

$$\begin{aligned}\mathrm{OT}_{q,\varepsilon} - \mathrm{OT} &\geq \int_{X_1} h_1(x)\, d\mu_1(x) + \int_{X_2} h_2(y)\, d\mu_2(y) \\ &\quad - \varepsilon \int_{X_1\times X_2} f_q^*\left(\frac{h_1(x)+h_2(y)-c(x,y)}{\varepsilon}\right) d\mu_1(x)\, d\mu_2(y) \\ &= a - \varepsilon \int_{X_1\times X_2} f_q^*\left(\frac{a-c(x,y)}{\varepsilon}\right) d\mu_1(x)\, d\mu_2(y).\end{aligned}$$

因此，我们有

$$\mathrm{OT}_{q,\varepsilon} - \mathrm{OT} \geq \sup_{a\in\mathbb{R}} \left\{ a - \varepsilon \int_{X_1\times X_2} f_q^* \left(\frac{a - c(x,y)}{\varepsilon} \right) d\mu_1(x)\, d\mu_2(y) \right\}. \tag{5.34}$$

由于 f_q^*不是递减的，且 $c(x,y) \geq 0$，我们有

$$\begin{aligned}
&\varepsilon \int_{X_1\times X_2} f_q^* \left(\frac{a - c(x,y)}{\varepsilon} \right) d\mu_1(x)\, d\mu_2(y) \\
&= \varepsilon \int_{X_1\times X_2} f_q^* \left(\frac{a - c(x,y)}{\varepsilon} \right) \chi_{\{a\geq c(x,y)\}}\, d\mu_1(x)\, d\mu_2(y) \\
&\quad + \varepsilon \int_{X_1\times X_2} f_q^* \left(\frac{a - c(x,y)}{\varepsilon} \right) \chi_{\{a< c(x,y)\}}\, d\mu_1(x)\, d\mu_2(y) \\
&\leq \varepsilon f_q^* \left(\frac{a}{\varepsilon} \right) \int_{X_1\times X_2} \chi_{\{a\geq c(x,y)\}}\, d\mu_1(x)\, d\mu_2(y) + \varepsilon f_q^*(0),
\end{aligned} \tag{5.35}$$

其中， χ_A 是集合 A的特征函数。因此，将 (5.35) 代入 (5.34)，可得

$$\mathrm{OT}_{q,\varepsilon} - \mathrm{OT} \geq \sup_{a\in\mathbb{R}} \left\{ a - \varepsilon f_q^* \left(\frac{a}{\varepsilon} \right) \int \chi_{\{a\geq c(x,y)\}}\, d\mu_1(x)\, d\mu_2(y) \right\} - q^{-\frac{q}{q-1}}\varepsilon.$$

根据代价函数 c 的定义，可以看出

$$\chi_{\{a\geq c(x,y)\}} \leq \prod_{i=1}^{d} \chi_{\{|x_i - y_i|\leq a\}}.$$

因此，有:

$$\begin{aligned}
\int_{X_1\times X_2} \chi_{\{a\geq c(x,y)\}}\, d\mu_1(x) d\mu_2(y) &\leq \prod_{i=1}^{d} \int_0^1 \int_0^1 \chi_{\{|x_i - y_i|\leq a\}}\, dx_i\, dy_i \\
&= (2a - a^2)^d \leq (2a)^d
\end{aligned} \tag{5.36}$$

对于任意 $0 \leq a \leq 1$。另一方面，

$$\varepsilon f_q^* \left(\frac{a}{\varepsilon} \right) \leq \varepsilon 2^{\frac{q}{q-1}-1} \left[q^{-\frac{q}{q-1}} + \left(\frac{a(q-1)}{\varepsilon q} \right)^{\frac{q}{q-1}} \right] \leq C_1 \varepsilon^{-\frac{1}{q-1}} a^{\frac{q}{q-1}} + C_2 \varepsilon, \tag{5.37}$$

其中

$$C_1 = 2^{\frac{1}{q-1}} \left(\frac{q-1}{q} \right)^{\frac{q}{q-1}} \quad \text{and} \quad C_2 = 2^{\frac{1}{q-1}} q^{-\frac{q}{q-1}}.$$

因此，由 (5.36) 和 (5.37) 可知,

$$\begin{aligned} a - \varepsilon f^*\left(\frac{a}{\varepsilon}\right) \int_{X_1 \times X_2} \chi_{\{a \geq c(x,y)\}} \, d\mu_1(x) \, d\mu_2(y) \\ \geq a - 2^d C_1 \varepsilon^{-\frac{1}{q-1}} a^{\frac{(q-1)d+q}{q-1}} - 2^d C_2 \varepsilon a^d. \end{aligned}$$

选取

$$a = k\varepsilon^{\frac{1}{(q-1)d+1}} \quad \text{with } k = \left[\frac{q-1}{((q-1)d+q)2^d C_1}\right]^{\frac{q-1}{(q-1)d+1}},$$

于是我们得到

$$\mathrm{OT}_{q,\varepsilon} - \mathrm{OT} \geq \tilde{K}_{1,q} \varepsilon^{\frac{1}{(q-1)d+1}} - \tilde{K}_{2,q} \varepsilon,$$

其中

$$\tilde{K}_{1,q} = \frac{(q-1)d+1}{(q-1)d+q} \left[\frac{q-1}{((q-1)d+q)2^d C_1}\right]^{\frac{q-1}{(q-1)d+1}} \text{ and } \tilde{K}_{2,q} = q^{-\frac{q}{q-1}} + 2^d C_2.$$

这就是所需的结论。 □

6 结论

本文研究了连续 Tsallis 熵正则化最优传输问题 (1.6)。我们讨论了当正则化参数 ε 趋于零时，该问题向 Monge–Kantorovich 问题 (1.3) 的 Γ-收敛（定理 1.2），并确立了其收敛速率（定理 1.5）。特别地，我们将其收敛速率与 KL 熵正则化最优传输问题 (1.1) 的收敛速率进行了比较，并证明在 $1 < q \leq 2$ 的情况下，前者的收敛速率较慢。此外，我们还证明了该收敛速率是锐利的（定理1.8）。

本文是对连续 Tsallis 熵正则化最优传输问题 (1.6) 进行理论分析的初步探索，仍有许多尚待解答的问题。例如，对于 $0 < q < 1$，正如注 1.6 所讨论的，Tsallis 熵正则化最优传输问题 (1.6) 的收敛速率预计会快于 KL 熵正则化最优传输问题 (1.1)。然而，由于技术上依赖于 ϕ_q 的凹性，$0 < q < 1$ 的收敛速率仍未得到解决。在离散情形下，相对熵的凸性会显著影响正则化最优传输问题的收敛速率，其衡量方式为 Bregman 散度

[23]。预计连续情形也存在类似情况，这将留待未来工作探讨。

此外，从计算最优传输的角度定量估计 Tsallis 相对熵正则化传输计划支撑集的扩散范围，也将是一个有趣的研究方向。KL 熵的情形已在 [4] 中讨论。

作为 Tsallis 相对熵背景下 Schr¨odinger 桥问题 (1.4) 的自然推广，我们考虑以下极小化问题：

$$\inf_{\pi\in\Pi(\mu_1,\mu_2)} \varepsilon D_q(\pi, \mathcal{K}_q), \tag{6.38}$$

其中， $\mathcal{K}_q$ 定义为

$$\mathcal{K}_q = \exp_q\left(-\frac{c(x,y)}{\varepsilon}\right)\mu_1\otimes\mu_2 = \left[1+(q-1)\left(\frac{c(x,y)}{\varepsilon}\right)\right]^{-\frac{1}{q-1}}\mu_1\otimes\mu_2.$$

利用 q-对数函数的非可加性

$$\log_{2-q}(xy) = \log_{2-q}x + \log_{2-q}y + (q-1)\log_{2-q}x\log_{2-q}y$$

以及 $\log_{2-q}\left(\exp_q(-x)^{-1}\right) = x$，极小化问题 (6.38) 可对应为以下 Tsallis 熵正则化最优传输问题 (1.6)：

$$\mathrm{OT}_{q,\varepsilon} = \inf_{\pi\in\Pi(\mu_1,\mu_2)}\left\{\varepsilon D_q(\pi,\mathcal{K}_q) - (q-1)\int_{X_1\times X_2} c\log_{2-q}\left(\frac{d\pi}{d(\mu_1\otimes\mu_2)}\right)d\pi\right\}.$$

我们注意到，Tsallis 熵正则化最优传输问题 (1.6) 与极小化问题 (6.38) 的等价关系涉及一个相互作用项，而 KL 熵正则化最优传输问题 (1.1) 等价于 Schr¨odinger 桥问题 (1.4)。对 (1.6) 和 (6.38) 之间的关系建立定量估计，将是未来研究的一个有趣方向。

数据可用性声明

本文不适用数据共享，因为本研究期间未生成或分析任何数据集。

致谢

作者感谢 Jun Okamoto 博士（京都大学）提供的宝贵讨论。第一作者获得了JSPS研究活动启动资助（22K20336）的部分资助。第二作者获得了 JSPS青年科学家研究资助（21K13822）、JSPS变革性研究领域（A）资助（22A201）以及京都大学 WPI-ASHBi 的部分资助。

参考文献

[1] L. Ambrosio, E. Brué, and D. Semola,《最优传输讲义》，Springer, 2021。[2] M. Arjovsky, S. Chintala, and L. Bottou, Wasserstein 生成对抗网络，（2017），第214–223页。[3] H. Bao and S. Sakaue, 具有变形q熵的稀疏正则化最优传输，Entropy，24（2022）。[4] E. Bernton, P. Ghosal, and M. Nutz, 熵正则化最优传输：几何与大偏差，《杜克数学杂志》，171（2022），第3363–3400页。[5] L. Borland, A. R. Plastino, and C. Tsallis, 非广延热力学统计中的信息增益，《数学物理学杂志》，39（1998），第6490–6501页。[6] L. M. Bregman, 寻找凸集公共点的松弛法及其在求解凸规划问题中的应用，《苏联计算数学与数学物理》，7（1967），第200–217页。[7] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer, 最优传输与梯度流的熵正则化格式的收敛性，《SIAM数学分析杂志》，49（2017），第1385–1418页。

[8] G. Carlier, P. Pegon, and L. Tamanini, 一般熵正则化最优传输代价的收敛速率, 《变分与偏微分方程》, 62（2023）, 第116页。[9] Y. Chen, T. T. Georgiou, and M. Pavon, 随机控制的联结：Richard Sinkhorn 在 Schrödinger 桥问题中遇见 Gaspard Monge, 《SIAM评论》, 63（2021）, 第249–313页。[10] E. Cinlar,《概率与随机过程》, 《数学研究生教材》第261卷, Springer, 纽约, 2011。[11] R. Cominetti and J. S. Martín, 线性规划中指数惩罚轨迹的渐近分析, 《数学规划》, 67（1994）, 第169–187页。[12] M. Cuturi, Sinkhorn 距离：快速计算最优传输, 《神经信息处理系统进展》, 26（2013）, 第2292–2300页。[13] S. Di Marino and A. Gerolin, 一般凸正则化下的最优传输损失与 Sinkhorn 算法, arXiv 预印本 arXiv:2007.00976,（2020）。[14] S. Eckstein and M. Nutz, 正则化最优传输的定量稳定性与 Sinkhorn 算法的收敛性, 《SIAM数学分析杂志》, 54（2022）, 第5922–5948页。[15] ——, 通过量化研究正则化最优传输的收敛速率, 《运筹学数学》, 49（2024）, 第1223–1240页。[16] S. Erlander,《最优空间相互作用与引力模型》, 第173卷, Springer-Verlag, 1980。[17] A. Figalli and F. Glaudo, 《最优传输、Wasserstein 距离与梯度流导论》, 《EMS数学教材》, EMS Press, 柏林, 2021。[18] S. Furuichi, 关于 Tsallis 熵与 Tsallis 相对熵的唯一性定理, IEEE信息论汇刊, 51（2005）, 第3638–3645页。

[19] S. Furuichi, K. Yanagi, and K. Kuriyama, Tsallis 相对熵的基本性质，《数学物理学杂志》，45 (2004)，第 4868–4877 页。
[20] S. Graf and H. Luschgy, 概率分布量化基础，Springer，2007。
[21] S. Kullback and R. A. Leibler, 关于信息与充分性，数学统计年鉴，22 (1951)，第 79–86 页。[22] C. Léonard, Schr ¨odinger 问题及其与最优传输某些联系的综述，《离散与连续动力系统——A辑》，34 (2014)，第 1533–1574 页。[23] K. Morikuni, K. Sakakibara, and A. Takatsu, 基于 Bregman 散度的正则化最优传输问题误差估计，arXiv 预印本 arXiv:2309.11666，(2023)。[24] B. Muzellec, R. Nock, G. Patrini, and F. Nielsen, Tsallis 正则化最优传输与生态推断，AAAI 人工智能会议论文集，31 (2017)。[25] J. Naudts, 广义热力学统计学，Springer London，2011。
[26] M. Nutz, 熵正则化最优传输导论，讲义，哥伦比亚大学，(2022)。
[27] G. Patrini, R. Van den Berg, P. Forre, M. Carioni, S. Bhargav, M. Welling, T. Genewein, and F. Nielsen, Sinkhorn 自动编码器，《人工智能中的不确定性》，(2020)，第 733–743 页。
[28] G. Peyré and M. Cuturi, 计算最优传输：及其在数据科学中的应用，《机器学习基础与趋势》，11 (2019)，第 355–607 页。
[29] Y. Polyanskiy and Y. Wu, 信息论：从编码到学习，2022。

[30] A. Robledo and C. Velarde, Tsallis统计力学何时、为何以及如何对自然现象作出精确描述，Entropy，24（2022）。[31] F. Santambrogio，应用数学家的最优传输，第55卷，Springer，2015。.[32] E. Schrödinger，论自然定律的逆转，普鲁士科学院会议报告·数学与物理学部，（1931），第144–153页。[33] R. Sinkhorn，任意正矩阵与双随机矩阵之间的关系，数学统计年鉴，35（1964），第876–879页。[34] R. Sinkhorn and P. Knopp，关于非负矩阵与双随机矩阵，太平洋数学杂志，21（1967），第343–348页。
[35] J. Solomon, F. De Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas，卷积Wasserstein距离：几何域上的高效最优传输，ACM图形学汇刊（ToG），34（2015），第1–11页。
[36] H. Suyari，将Shannon-Khinchin公理推广至非广延系统及非广延熵的唯一性定理，IEEE信息论汇刊，50（2004），第1783–1787页。[37] A. Takatsu，通过严格凸函数松弛最优传输问题，arXiv预印本 arXiv:2102.07336，（2021）。[38] Q. Tong and K. Kobayashi，多元正态分布与q正态分布上的熵正则化最优传输，Entropy，23（2021）。[39] C. Tsallis，玻尔兹曼-吉布斯统计的可能推广，统计物理学杂志，52（1988），第479–487页。[40] ——, 非广延统计力学导论：探索复杂世界，第1卷，Springer，2009。

[41] C. Villani, 《最优传输：新旧观点》，第338卷，Springer，2009年。

[42] ——, 《最优传输专题》，第58卷，美国数学学会，2021年。

[43] J. Weed, 《线性规划中熵惩罚项的显式分析》，学习理论会议，（2018），第1841–1855页。

[44] A. G. Wilson, 《熵最大化模型在出行分布、交通方式划分和路线划分理论中的应用》，《交通经济学与政策杂志》，（1969），第108–126页。

Faculty of Advanced Science and Technology,
Kumamoto Univer-sity, Kumamoto 860-8555, Japan电子邮箱:
suguro@kumamoto-u.ac.jp

Mathematical Science Center for Co-creative Society,
Tohoku Uni-versity, Sendai 980-0845, Japan电子邮箱:
toshiaki.yachimura.a4@tohoku.ac.jp