arXiv:2301.06741v1 [cs.DS] 17 Jan 2023

小树宽下更快的 Sinkhorn 算法

Zhao Song* Tianyi Zhou†

摘要

计算最优传输（OT）距离（例如地球移动者距离）是机器学习、统计学和计算机视觉中的一个基本问题。本文研究近似两个大小为 n 的离散分布之间一般最优传输距离的问题。给定成本矩阵 $C = AA^\top$，其中 $A \in \mathbb{R}^{n \times d}$，当矩阵 A 的树宽为 τ 时，我们提出了一种更快的 Sinkhorn 算法来近似最优传输距离。为了近似最优传输距离，我们的算法将当前最先进的结果 [Dvurechensky、Gasnikov 和 Kroshnin（ICML 2018] ）的运行时间从 $\widetilde{O}(\epsilon^{-2}n^2)$ 提升至 $\widetilde{O}(\epsilon^{-2}n\tau)$。

*zsong@adobe.com。Adobe Research。

†t8zhou@ucsd.edu。加州大学圣迭戈分校。

目录

1 引言 **2**

1.1 我们的结果.3 1.2 相关工作.3 1.3 技术概述.4

2 预备知识 **5**

2.1 问题表述.5 2.2 不等式.6 2.3 树宽预备知识.7

3 Sinkhorn 算法分析 8

3.1 定义.8 3.2 有界 max – min.9 ̃3.3 势函数 ψ .10 3.4 势函数上界.11 3.5 迭代复杂度界.12 3.6 归纳法.13

4 小树宽设置下的运行时间 14

4.1 K 的隐式形式.14 4.2 小树宽下 Sinkhorn 的运行时间.14 4.3 舍入算法的正确性.16 4.4 舍入算法的运行时间.18 4.5 Sinkhorn 计算最优传输距离的运行时间.19

5 对称 **21**

1 引言

最优传输是一种数学理论，研究如何以最高效的方式将货物或材料从一个地方运送到另一个地方。其目标是尽量降低运输成本，通常以运输距离或所用资源量来衡量。计算科学中的许多问题都需要使用最优传输来比较概率测度或直方图，包括 Wasserstein 距离或地球移动者距离[WPR85, RTG00，Vil09]。最优传输有广泛的应用，例如用于自然语言处理的词袋[KSKW15], 多标签分类 [FZM+15]、无监督学习 [ACB17，BGKL17], 半监督学习 [SRGB14], 统计学 [ESS17，PZ16], 以及其他应用[KPT+17]。特别是由于其在图像处理中的应用，如今高效计算或近似计算两个测度之间的最优传输距离或 Wasserstein 距离已变得至关重要。

关于最优传输问题的研究由来已久。 [Cut13] 采用 Sinkhorn 算法求解熵正则化最优传输优化问题。正如近期研究所示， [ANWR17], 这种方法可以在 ϵ时间内求得最优传输距离的 $\widetilde{O}(\epsilon^{-3}n^2)$-近似。就对 n的依赖而言，这一结果改进了网络单纯形法或内点法直接求解最优传输问题（一个线性规划）所达到的复杂度 $\widetilde{O}(n^3$[PW09], 。对 [Kan42]的三次依赖使得难以高精度地近似最优传输距离。随后， ϵ 在 [DGK18], 中，他们基于 Sinkhorn 算法提出了一种算法，其复杂度上界为$\widetilde{O}($ $\epsilon^{-2}n^2$)。

矩阵的树宽衡量其结构的复杂程度，在设计和分析矩阵操作与处理算法时发挥着关键作用。特别是，矩阵的树宽可用于确定依赖树分解的算法（例如动态规划和分治技术）的效率。在树宽较小的情形下，矩阵操作与处理算法通常能够实现近线性运行时间，因此效率高且可扩展。这对广泛的应用领域具有重要意义，包括内点法 [GS22、DLY21], 计算 John 椭球[SYYZ22], 流式算法 [LSZ+22]。树宽在图结构理论中也很重要，尤其是在 Robertson 和 Seymour 对图子式的研究中 [RS10]。许多研究结果 [Bod94]表明，在树宽有界的图类上，NP困难问题可以在多项式时间内求解。

此前解决这个问题的最佳成果需要 n^2。自然会有人提出这样一个问题

是否可以在某些温和假设下（例如树宽为 o）于 n^2 内解决?

在本文中，我们对这个问题给出了肯定的答案。我们的结果与先前工作的比较见表 1。

参考文献	方法	时间复杂度
[PW09]	网络单纯形法	n^3
[ANWR17]	Sinkhorn 算法	$\epsilon^{-3}n^2$
[DGK18]	Sinkhorn 算法	$\epsilon^{-2}n^2$
定理 4.5	Sinkhorn 算法	$\epsilon^{-2}n\tau$

表1：给定成本矩阵 $C = AA^\top \in \mathbb{R}^{n\times n}$，令 τ 表示矩阵 A的树宽。令 ϵ表示精度参数。由于 $\tau \leq n$，我们的算法（定理 4.5，算法 3）总是优于 [DGK18]。

1.1 我们的结果

我们正式陈述主要定理

定理 1.1。给定成本矩阵 $C = AA^\top$，其中 A 的树宽为 τ，我们可以在

$$O(\epsilon^{-2} n\tau \|C\|_\infty^2 \ln n)$$

时间。

与 [DGK18], 在 $O(\epsilon^{-2} n^2 \|C\|_\infty^2 \ln n)$ 时间内解决该问题的方法相比，我们提出了一种利用隐式形式构造矩阵的算法。借助低树宽的性质，我们的运行时间不依赖于 n^2。

1.2 相关工作

最优传输问题 最优传输距离也称为土方搬运距离 [RTG00], ，正逐渐被广泛应用于各种场景，成为一种有效工具，从计算机图形学[BPC16] 到监督学习 [FZM+15]、无监督密度拟合 [BBR06] 和生成模型学习（[MMC16, ACB17, SZRM18, GPC18, SBRL18]）。针对降低求解最优传输问题时间复杂度的研究由来已久。在 [ACB17], 中，作者证明，对于正则化最优传输，Sinkhorn 和 Greenkhorn 算法均可实现近线性时间复杂度。他们指出，这两种算法的复杂度均为 $\widetilde{O}(\epsilon^{-3} n^2)$, 其中 n 表示所考虑概率测度中原子的数量（或维数），而 ϵ 表示所需的容差水平。在 [DGK18], 中，Sinkhorn 算法的复杂度被改进为 $\widetilde{O}(\epsilon^{-3} n^2)$。此外，文中还提出了一种自适应原始-对偶加速梯度下降（APDAGD）算法，其复杂度为 $\widetilde{O}(\min\{\epsilon^{-1} n^{9/4}, \epsilon^{-2} n^2\})$。借助精心设计的牛顿型算法，[AZLOW17, CMTV17] 通过利用与矩阵缩放问题之间的联系来求解最优传输问题。 [BJKS18, Qua 18] 给出了牛顿型算法的复杂度上界 $\widetilde{O}(\epsilon^{-1} n^2)$。

树宽问题 树宽是结构图论中的一个概念，已在多个领域中与固定参数可处理算法相关联地得到研究，这些领域包括组合学、整数线性规划和数值分析。 [FLS+18] 表明，若干问题可以高效地归约为矩阵分解，包括计算行列式、计算秩和寻找最大匹配；由此可得到运行时间为 $O(\tau^{O(1)} \cdot n)$的算法，其中 τ 是给定图的树分解的宽度。[Bod94] 表明，诸如Independent Set、Hamiltonian Circuit、 Steiner Tree、 and Travelling Salesman等多个NP困难问题可通过动态规划求解，其运行时间仅在线性依赖于问题规模，并随树宽呈指数增长。利用小树宽情形，[DLY21] 提出了一种求解线性规划问题的算法，其运行时间几乎与求解子问题 $Ax = b$的最快运行时间相当。 [LSZ+22] 在流式模型中提出了一种空间高效内点法~(IPM)。对于树宽为 τ的线性规划，他们可在 $O(n\tau)$空间内求解，其中 n 是特征空间的维数。 [SYYZ22] 表明，当约束矩阵的树宽为 τ 时，约翰椭球问题可在 $O(n\tau^2)$时间内求解。小树宽情形也被应用于求解半正定规划。在 [GS22], 中，他们给出了首个在这种情形下运行时间与变量数量呈线性关系的半正定规划求解器。

1.3 技术概述

分析 我们首先给出 u_k、 v_k 以及方程（u^*、 v^*）的一个最优解（5）。然后，我们将如下引入（$\widehat{u}$、 $\widehat{v}$）的凸函数:

$$\langle \mathbf{1}_n, B(\widehat{u},\widehat{v})\mathbf{1}_n\rangle - \langle \widehat{u}, B(u_k,v_k)\mathbf{1}_n\rangle - \langle \widehat{v}, B(u_k,v_k)^\top \mathbf{1}_n\rangle.$$

当 $(u^*, v^*) = (u_k, v_k)$ 时，上述函数的梯度为零，因此点 (u_k, v_k) 是该函数的最小化点。

因此，我们可以证明

$$\widetilde{\psi}(u_k,v_k) \le \langle u_k - u_*, B_k\mathbf{1}_n - r\rangle + \langle v_k - v_*, B_k^\top \mathbf{1}_n - c\rangle$$

接着，对于算法的每次迭代，我们对右端项进行上界估计，并得 t

$$\widetilde{\psi}(u_k,v_k) \le R\cdot(\|B_k\mathbf{1}_n - r\|_1 + \|B_k^\top \mathbf{1}_n - c\|_1).$$

其中，该不等式由迭代点的界 u_k、 v_k 以及一个最优解（u^*， v^*）得到。 ~接下来，利用 ψ 的这个上界以及引理 2.8 可得:

$$\begin{aligned}&\widetilde{\psi}(u_k,v_k) - \widetilde{\psi}(u_{k+1},v_{k+1})\\ \ge\ &\max\{\frac{\widetilde{\psi}(u_k,v_k)^2}{2R^2}, \frac{\epsilon_0^2}{2}\},\end{aligned}$$

通过归纳法，我们证明势函数 $\widetilde{\psi}$ 也有上界 $\frac{2R^2}{k+\ell-1}$，其中$^{\ell} = {}^{2R^2}_{\widetilde{\psi}\ u_1\ v_1(,\ )}$。最后，通过切换策略，我们给出 ~~Si~~nkhorn 算法总迭代次数 k 的上界如下

$$k \le 2 + \frac{4R}{\epsilon_0}.$$

运行时间 给定成本矩阵 $C = MM^\top$，其中 $M \in \mathbb{R}^{n\times d}$ 的树宽为 τ，我们利用它存在简洁的 Cholesky 分解这一事实，且 nnz(C) = $O(n\tau)$。

对于 Sinkhorn 算法（算法 1）的每次迭代，我们需要计算 $B(u,v)$ =diag(e^u)K diag(e^v), 其中 $K_{i,j}$:= exp($-C_{i,j}/\gamma$)。事实上，显式写出 K 需要 $O(n^2)$。为绕过这一问题，我们首先将 K 写成隐式形式 $K_{i,j} := A_{i,j} - D_{i,j}$，其中 $A_{i,j} = e^{-C_{i,j}/\gamma} - 1$ 且 $D_{i,j} = 1$，使得矩阵 A 与矩阵 C 一样稀疏。此外，我们用 $ww^\top$ 表示矩阵 D ，其中 $w = \mathbf{1}_n$。利用 nnz(A) = $O(n\tau)$ 且矩阵 D 是秩一矩阵这一事实，我们将 Sinkhorn 算法的每次迭代运行时间从 $O(n^2)$ 降至 $O(n\tau)$。

对于运输方案的舍入算法（算法 4），我们也以隐式方式记下运输方案，并在 $O(n\tau)$ 时间内完成计算。请注意，我们从不显式写出 B， B_0， B_1 ，也不显式输出 G 。在计算 $B\mathbf{1}_n$ 时，我们利用 B 的隐式形式，并按如下方式进行计算:

$$\mathrm{diag}(e^{u_k})A\mathbf{1}_n\,\mathrm{diag}(e^{v_k}) + \mathrm{diag}(e^{u_k})(ww^\top)\mathbf{1}_n\,\mathrm{diag}(e^{v_k}).$$

由于 nnz(A) = $O(n\tau)$，计算 $A\mathbf{1}_n$ 需要 $O(n\tau)$ 时间。同样，在计算 XB时，其中 X是一个对角矩阵，我们利用 B 的隐式形式，并按以下方式进行计算:

$$\mathrm{diag}(e^{u_k})AX\,\mathrm{diag}(e^{v_k}) + \mathrm{diag}(e^{u_k})(ww^\top)X\,\mathrm{diag}(e^{v_k}).$$

当 nnz(A) $= O(n\tau)$，计算 $A\widetilde{X}$ 需要 $O(n\tau)$ 时间，而且 AX 也具有 $O(n\tau)$ 稀疏性 e.

最后，我们指出，使用 $O(\epsilon^{-2}n\tau)$ 时间，我们可以近似求解最优传输距离问题的运输方案。

路线图。

我们首先在第 2 节介绍所需的全部预备知识。然后，我们在第 3 节分析 Sinkhorn 算法。在第 4 节中，我们针对小树宽情形给出更快的 Sinkhorn 算法，并应用该算法求解最优传输距离。

2 预备知识

对于正整数 n，我们记作 $[n] = \{1, 2, \cdots, n\}$。我们用 $\mathbf{1}_n$ 表示长度为 n、所有元素均为 1 的向量。

对于向量 a，我们记 e^a 和 ln a 分别为其逐元素指数和自然对数。我们定义$a_{k,i}$ 为 a的第 i 次迭代的第 k 个坐标。

对于矩阵 $A \in \mathbb{R}^{n\times n}$，我们定义 $\|A\|_\infty := \max_{i,j\in[n]} |A_{i,j}|$。我们定义 $A_{i,j}$ 为矩阵 A 的第 i 行、第 j 列上的元素。我们用 e^A 和 ln A 分别表示其逐元素指数和自然对数。我们用 vec(A) 表示 $\mathbb{R}^{n^2}$ 中的向量，它由 A 的各列依次向下排列得到。对于两个矩阵 A、B，我们用 $\langle A$、$B\rangle$ 表示它们的内积。我们定义 n 维单纯形为 $\triangle_n := \{x \in \mathbb{R}^n_+ : \sum_{i=1}^n x_i = 1\}$。

对于向量 $x \in \mathbb{R}^n$，我们将其 ℓ_p 范数定义为 $\|x\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$。对于两个向量 x、y，我们将内积定义为 $\langle x$、$y\rangle = \sum_{i=1}^n x_i y_i$。

熵的定义如下：

定义 2.1（熵）。我们将向量 p 的熵 $H(p)$ 定义为

$$H(p) := \sum_{i=1}^n p_i \log(\frac{1}{p_i}).$$

类似地，对于矩阵 $P \in \mathbb{R}^{n\times n}_+$，我们逐项定义其熵 $H(P)$ 为

$$\sum_{i=1}^n \sum_{j=1}^n \log \frac{1}{P_{i,j}}.$$

2.1 问题表述

我们首先介绍最优传输问题的定义。

定义 2.2. 给定一个树宽较小的矩阵 C（例如 $C = AA^\top$，其中 $A \in \mathbb{R}^{n\times d}$），最优传输问题定义如下：

$$\begin{aligned} \min_X \ & \langle C, X\rangle \\ \text{s.t.} \ & X \in \mathbb{R}^{n\times n}_+ \\ & X\mathbf{1}_n = r \\ & X^\top \mathbf{1}_n = c \end{aligned}$$

其中 $\mathbf{1}_n \in \mathbb{R}^n$ 表示一个每个元素均为 1 的向量。

接下来，我们给出正则化最优传输问题的定义。

定义 2.3。给定一个强凸正则项 $\mathcal{R}(X)$，例如负熵或欧几里得范数的平方，正则化最优传输问题定义如下:

$$\begin{aligned} \min_{X} \ & \langle C, X\rangle + \gamma\mathcal{R}(X) \\ \text{s.t.} \ & X \in \mathbb{R}_{+}^{n\times n} \\ & X\mathbf{1}_n = r \\ & X^{\top}\mathbf{1}_n = c \end{aligned} \tag{1}$$

其中 $\gamma > 0$ 表示正则化参数。

本文旨在求出运输方案 X 的近似，如下所示:

定义 2.4（ϵ-近似）。最优传输距离的 ϵ-近似定义为

$$\begin{aligned} \langle C, \widehat{X}\rangle \leq \ & \min_{X}\langle C, X\rangle + \epsilon \\ \text{s.t.} \ & X \in \mathbb{R}_{+}^{n\times n} \\ & X\mathbf{1}_n = r \\ & X^{\top}\mathbf{1}_n = c \end{aligned} \tag{2}$$

其中 X 表示运输方案的近似。

为简单起见，我们引入 $\mathcal{U}_{r,\ c} \subset \mathbb{R}_{+}^{n\times n}$的定义

定义 2.5. 给定最优传输问题 $\arg\min_{X\in\mathcal{U}_{r,c}}\langle X, C\rangle$，我们定义

$$\mathcal{U}_{r,c} := \{X \in \mathbb{R}_{+}^{n\times n} : X\mathbf{1}_n = r, X^{\top}\mathbf{1}_n = c\}$$

其中， $\mathbf{1}_n$ 是 $\mathbb{R}^n$ 中的全一向量， $C \in \mathbb{R}_{+}^{n\times n}$是给定的成本矩阵，而 $r \in \mathbb{R}^n$、 $c \in \mathbb{R}^n$ 是给定的正元素向量，且其元素之和为一。

接下来，我们给出一个关于运输方案 X的引理。

引理 2.6 ([截断13])。对于任意成本矩阵 $C \in \mathbb{R}^{n\times n}$、 $\mathcal{U}_{r,c} \subset \mathbb{R}_{+}^{n\times n}$以及 r、 $c \in \triangle_n$，最小化问题

$$X_\gamma := \arg\min_{X\in\mathcal{U}_{r,c}} \langle X, C\rangle + \gamma \cdot \mathcal{R}(X),$$

其中， $\gamma > 0$ 是正则化参数， $\mathcal{R}(X)$是强凸正则项，其在$X_\gamma \in \mathcal{U}_{r,c}$处取得唯一最小值，形式为 $X_\gamma = MAN$，其中 A:=exp$(-\frac{1}{\gamma}C)$，且 M、 $N \in \mathbb{R}_{+}^{n\times n}$均为对角矩阵。矩阵$(M, N)$在相差一个常数因子的意义下唯一。

2.2 不等式

我们引入如下 H¨older 不等式:

引理 2.7（H¨older不等式）。如果 $p > 1$ 和 $q > 1$ 满足

$$\frac{1}{p} + \frac{1}{q} = 1$$

then

$$\|ab\|_1 \leq \|a\|_p \|b\|_q.$$

我们还给出了Pinsker不等式。

引理 2.8（Pinsker不等式）。设 P 和 Q 是定义在全集 U上的两个分布。那么，

$$\mathrm{KL}(P\|Q) \geq \frac{1}{2\ln 2} \cdot \|P - Q\|_1^2.$$

其中 KL($P\|Q$) 是 P 与 Q之间的 KL散度。

2.3 树宽预备知识

我们首先介绍给定矩阵的树宽定义。

定义 2.9（树宽 τ）。给定矩阵 $A \in \mathbb{R}^{n\times d}$，我们按如下方式构造其图 $G = (V, E)$：顶点集为列 $[d]$；边 $(i, j) \in E$ 当且仅当存在 $k \in [n]$ 使得 $A_{k,\ i} \neq 0$,$A_{k,\ j} \neq 0$。那么，矩阵 A 的树宽就是所构造图的树宽。特别地， A 的每一列都是 τ-稀疏的。

接下来，我们介绍 Cholesky 分解的定义。

定义 2.10（Cholesky 分解）。给定一个正定矩阵 P，存在唯一的 Cholesky 分解 $P = LL^\top \in \mathbb{R}^{d\times d}$，其中 $L \in \mathbb{R}^{d\times d}$ 是一个对角元素为实数且为正的下三角矩阵。

我们还给出了计算 Cholesky 分解的运行时间。

引理 2.11 ([GLN94, DLY21])。给定一个正定矩阵 $M \in \mathbb{R}^{d\times d}$，我们可以通过 Cholesky 分解 $M = LL^\top$ 在以下时间内

$$\Theta(\sum_{j=1}^{d} |\mathcal{L}_j|^2),$$

其中 $|\mathcal{L}_j|$ 是 L 的第 j 列中的非零元素个数。

接下来，我们介绍一些基于给定树宽为 τ 的矩阵的 Cholesky 分解的结果：

引理 2.12 ([BGHK95, Dav06, DLY21])。对于任意矩阵 $A \in \mathbb{R}^{n\times n}$，其树宽为 τ，我们可以在 $AA^\top = LL^\top \in \mathbb{R}^{n\times n}$ $O(n\tau^2)$ 时间内计算其 Cholesky 分解，其中 $L \in \mathbb{R}^{n\times n}$ 是一个对角元素为实数且为正的下三角矩阵。 L 满足每一列均为τ稀疏的性质。

命题 2.13 ([DLY21、GS22、SYYZ22、LSZ+22])。给定 $L = MM^\top$，其中 M的树宽为 τ且$M \in \mathbb{R}^{m\times n}$，我们有$\mathrm{nnz}(L) = O(n\tau)$。

证明。我们首先证明 nnz(L) $= O(m)$。令 $M \in \mathbb{R}^{m\times n}$ 表示图$G = (V, E)$的邻接矩阵，其中 $|E(G)| = m$， $|V(G)| = n$。图 G的拉普拉斯矩阵为 $L = MM^\top$，也定义为 $D - A$，其中 D 是度矩阵，A 是图G的邻接矩阵。由于 nnz(A) $= O(m)$， nnz(D) $= O(n)$ 且 $m \geq n$，我们有

$$\mathrm{nnz}(L) = O(m) + O(n) = O(m). \tag{3}$$

接下来，我们证明图 G 的边数 m 有界于 $O(n\tau)$。树宽为 τ的极大图是 τ-树，它们通过从一个 $(\tau+1)$-团开始，并迭代添加度为 τ 的顶点构造而成，使得这些顶点的邻居构成一个 τ-团。通过计算 $(\tau+1)$-团中的边数，以及迭代添加到τ-树中的 $n-\tau-1$ 个顶点所关联的边数，具有 n个顶点的 τ-树中的边总数为

$$\binom{\tau+1}{2}+\tau(n-\tau-1)=O(n\tau). \tag{4}$$

由于任何树宽为 τ 的图 G 都是某个 τ-树的子图，因此 $O(n\tau)$ 是 $|E(G)|=m$ 的上界。结合公式 (3) 和公式 (4)，可得 $\mathrm{nnz}(L)=O(n\tau)$。

因此，证明完毕。 □

3 Sinkhorn 算法分析

算法 1 Sinkhorn 算法

过程 SinkhornAlgorithm(c, r, ϵ_0) ▷ 定理3.8

▷精度 ϵ_0

$k \leftarrow 0$

$u_0 \leftarrow 0$

$v_0 \leftarrow 0$

当 $\|B(u_k, v_k)\mathbf{1}_n - r\|_1 + \|B(u_k, v_k)^\top \mathbf{1}_n - c\|_1 \geq \epsilon_0$ 时

如果 k 模 $2=0$，则

$u_{k+1} \leftarrow u_k + \ln\, r - \ln(B(u_k, v_k)\mathbf{1}_n)$

$v_{k+1} \leftarrow v_k$

else

$v_{k+1} \leftarrow v_k + \ln\, c - \ln(B(u_k, v_k)^\top \mathbf{1}_n)$

$u_{k+1} \leftarrow u_k$

结束如果

$k \leftarrow k+1$

结束 while 循环

返回 $B(u_k, v_k)$.

结束过程

在第 3.1 节中，我们给出 Sinkhorn 算法中使用的一些定义。在第 3.2 节中，我们~给出与 $u \in \mathbb{R}^n$ 相关的界，$v \in \mathbb{R}^n$。在第 3.3 节中，我们定义势函数 ψ。~在第 3.4 节中，我们给出 ψ 的上界。在第 3.5 节中，我们给出 Sinkhorn 算法的迭代复杂度界。在第 3.6 节中，我们给出势函数上界的归纳证明。

3.1 定义

我们首先介绍一些定义，以简化推导。

定义 3.1。我们将矩阵函数 B：$\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{n\times n}$ 定义如下：对于任意给定的向量u、$v \in \mathbb{R}^n$

$$B(u,v) := \mathrm{diag}(e^u) K \,\mathrm{diag}(e^v)$$

其中，$\mathrm{diag}(a) \in \mathbb{R}^{n\times n}$ 是以向量 $a \in \mathbb{R}^n$ 为对角线元素的对角矩阵，而 $K \in \mathbb{R}^{n\times n}$是定义如下的矩阵

$$K_{i,j} := \exp(-C_{i,j}/\gamma).$$

定义 3.2。我们定义函数 ψ： $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ 如下：对于任意给定的向量 u、 $v \in \mathbb{R}^n$

$$\psi(u, v) := \mathbf{1}_n^\top B(u, v)\mathbf{1}_n - \langle u, r\rangle - \langle v, c\rangle,$$

其中 B 在定义*3.1*中定义。

我们考虑 Sinkhorn–Knopp算法（算法1），该算法求解引理2 中引入的以下最小化问题[Cut13]:

$$\min_{u,v\in\mathbb{R}^n} \psi(u, v), \tag{5}$$

其中 ψ 在定义3.2中定义。

Pr问题方程（5）是方程（1）的对偶形式，因为我们选择了 $\mathcal{R}$（X） $= -H$（X）.

这里，我们展示证明 Sinkhorn 算法复杂度的总体思路。

我们首先展示如何求出 u_k、 v_k 以及方程（u_*、 v_*）的一个最优解（5）的界。

接下来，我们证明对于每次迭代， $\psi(u_k, v_k)$ 有如下上界：

$$\|B(u_k, v_k)\mathbf{1}_n - r\|_1 + \|B(u_k, v_k)^\top \mathbf{1}_n - c\|_1.$$

最后，利用 $\psi(u_k, v_k)$ 的界，我们给出 Sinkhorn 算法的复杂度结果。

定义 3.3。我们将 R 定义为

$$R := \ -\ln(K_{\min} \min_{i,j\in[n]}\{r_i, c_j\}),.$$

其中

$$K_{\min} := \min_{i,j\in[n]} K_{i,j} = e^{-\|C\|_\infty/\gamma}$$

3.2 有界 max –min

我们首先介绍一个与 $u_k \in \mathbb{R}^n$、 $v_k \in \mathbb{R}^n$、 $u_* \in \mathbb{R}^n$ 和 $v_* \in \mathbb{R}^n$ 的界有关的工具。

引理 3.4. 设 $k \geq 0$ 和 $u_k \in \mathbb{R}^n$， $v_k \in \mathbb{R}^n$ 由算法 *1* 和 $(u_*,\ v_*) \in \mathbb{R}^n \times \mathbb{R}^n$是方程(5)的一个解。则

$$\begin{aligned}
&\max_{i\in[n]} u_{k,i} - \min_{i\in[n]} u_{k,i} \leq R,\ \max_{j\in[n]} v_{k,j} - \min_{j\in[n]} v_{k,j} \leq R, \\
&\max_{i\in[n]} u_{*,i} - \min_{i\in[n]} u_{*,i} \leq R,\ \max_{j\in[n]} v_{*,j} - \min_{j\in[n]} v_{*,j} \leq R,
\end{aligned} \tag{6}$$

其中 R 在定义*3.3*中给出。

证明。首先，我们证明 $u_k \in \mathbb{R}^n$ 的界。由于 u, v 初始化为 $\mathbf{0}_n$，因此对于$k = 0$，不等式成立。鉴于 $k-1$ 为偶数，变量 u 在迭代 $k-1$ 和 $B(u_k, v_k)\mathbf{1}_n = r$ 中由算法构造进行更新。

因此，对于每个 $i \in [n]$，我们有

$$\begin{aligned} e^{u_{k,i}} K_{\min} \langle \mathbf{1}_n, e^{v_k} \rangle &\leq \sum_{j=1}^{n} e^{e_{k,i}} K_{i,j} e^{v_{k,j}} \\ &= [B(u_k, v_k)(\mathbf{1}_n)_i] \\ &= r_i \\ &\leq 1 \end{aligned} \tag{7}$$

其中，第一步由 $K_{\min}$ 的定义得出，第二步由 B 的定义得出，第三步由 $B(u_k, v_k)\mathbf{1}_n = r$ 得出，最后一步由概率单纯形 r的定义得出。

因此，通过重组式（7），我们有

$$\max_{i \in [n]} u_{k,i} \leq -\ln(K_{\min} \langle \mathbf{1}_n, e^{v_k} \rangle). \tag{8}$$

另一方面，由于 $0 \leq K_{i,\ j} \leq 1$，对于每个 $i \in [n]$,

$$\begin{aligned} & e^{u_{k,i}} \langle \mathbf{1}_n, e^{v_k} \rangle \\ \geq & \sum_{j=1}^{n} e^{u_{k,i}} K_{i,j} e^{v_{k,j}} \\ = & [B(u_k, v_k)\mathbf{1}_n]_i \\ = & r_i \end{aligned}$$

其中，第一步由 $K_{i,\ j} \leq 1$ 得出，第二步由 B 的定义得出，最后一步由 $B(u_k, v_k)\mathbf{1}_n = r$ 得出。

我们还有

$$\min_{i \in [n]} u_{k,i} \geq \min_{i \in [n]} \ln(\frac{r_i}{\langle \mathbf{1}_n, e^{v_k} \rangle}) = \ln(\frac{\min_{i \in [n]} r_i}{\langle \mathbf{1}_n, e^{v_k} \rangle}).$$

后一个等式和式（8）给出

$$\max_{i \in [n]} u_{k,i} - \min_{i \in [n]} u_{k,i} \leq -\ln(K_{\min} \min_{i \in [n]} r_i) \leq R$$

□

˜3.3 势函数 ψ ψ

˜为简化推导，我们定义 ψ 如下:

˜定义 3.5。我们将 ψ 定义为

$$\widetilde{\psi}(u, v) := \psi(u, v) - \psi(u_*, v_*)$$

其中，最后一步由 ψ 的定义得出。

命题 3.6. 我们有

$$\widetilde{\psi}(u,v) = \langle \mathbf{1}_n, B(u,v)\mathbf{1}_n\rangle - \langle \mathbf{1}_n, B(u_*,v_*)\mathbf{1}_n\rangle + \langle u_* - u, r\rangle + \langle v_* - v, c\rangle.$$

证明。我们可以得到

$$\begin{aligned}\widetilde{\psi}(u,v) &= \psi(u,v) - \psi(u_*,v_*)\\ &= \langle \mathbf{1}_n, B(u,v)\mathbf{1}_n\rangle - \langle \mathbf{1}_n, B(u_*,v_*)\mathbf{1}_n\rangle + \langle u_* - u, r\rangle + \langle v_* - v, c\rangle.\end{aligned}$$

~其中，第一步根据 ψ的定义得出，第二步根据ψ的定义得出。

□

3.4 势函数上界

这里，我们给出一个引理，稍后将用它来界定迭代复杂度。

引理 3.7。 设 $k \geq 1$和 u_k， $v_k \in \mathbb{R}^n$ 为算法 1的输出。记 $B_k :\, = B(u_k,\ v_k)$。则有

$$\widetilde{\psi}(u_k, v_k) \leq R \cdot (\|B_k\mathbf{1}_n - r\|_1 + \|B_k^\top \mathbf{1}_n - c\|_1).$$

证明。给定一个固定的 $k \geq 1$，对于以下关于$(\widehat{u}, \widehat{v})$的凸函数

$$\langle \mathbf{1}_n, B(\widehat{u},\widehat{v})\mathbf{1}_n\rangle - \langle \widehat{u}, B(u_k,v_k)\mathbf{1}_n\rangle - \langle \widehat{v}, B(u_k,v_k)^\top \mathbf{1}_n\rangle.$$

凸函数的梯度在 $(\widehat{u}, \widehat{v})$ $=(u_k, v_k)$ 处为零，因此点 (u_k, v_k) 是其极小点。

因此，

$$\begin{aligned}\widetilde{\psi}(u_k,v_k) &= [\langle \mathbf{1}_n, B_k\mathbf{1}_n\rangle - \langle u_k, B_k\mathbf{1}_n\rangle - \langle v_k, B_k^\top\mathbf{1}_n\rangle]\\ &\quad - [\langle \mathbf{1}_n, B(u_*,v_*)\mathbf{1}_n\rangle - \langle u_*, B_k\mathbf{1}_n\rangle - \langle v_*, B_k^\top\mathbf{1}_n\rangle]\\ &\quad + \langle u_k - u_*, B_k\mathbf{1}_n - r\rangle + \langle v_k - v_*, B_k^\top\mathbf{1}_n - c\rangle\\ &\leq \langle u_k - u_*, B_k\mathbf{1}_n - r\rangle + \langle v_k - v_*, B_k^\top\mathbf{1}_n - c\rangle. \qquad (9)\end{aligned}$$

~其中第一步根据 ψ 的定义得出。接下来，我们对不等式右侧进行界定。对于每次迭代，我们知道 $B_k\mathbf{1}_n = r$ 或 $B_k^\top\mathbf{1}_n = c$，因此有 $\langle \mathbf{1}_n, B_k\mathbf{1}_n\rangle = 1$和 $\langle \mathbf{1}_n, B_k\mathbf{1}_n - r\rangle = 0$。

取 $a = 0.5 \cdot (\max_{i\in[n]} u_{k,\ i} + \min_{i\in[n]} u_{k,\ i})$。于是，我们有

$$\begin{aligned}\langle u_k, B_k\mathbf{1}_n - r\rangle &= \langle u_k - a\mathbf{1}_n, B_k\mathbf{1}_n - r\rangle\\ &\leq \|u_k - a\mathbf{1}_n\|_\infty \|B_k\mathbf{1}_n - r\|_1\\ &= 0.5 \cdot (\max_{i\in[n]} u_{k,i} - \min_{i\in[n]} u_{k,i})\|B_k\mathbf{1}_n - r\|_1\\ &\leq \frac{R}{2}\|B_k\mathbf{1}_n - r\|_1.\end{aligned}$$

其中第一步由 $\langle \mathbf{1}_n$ 推出， $B_k\mathbf{1}_n - r\rangle = 0$ ，第二步由 H¨older 不等式推出，第三步根据 a 的定义得出，最后一步由引理 3.4得出。

类似地，我们在式 $\langle -u_*$、 $B_k\mathbf{1}_n - r\rangle$、 $\langle v_k$、 $B_k^\top\mathbf{1}_n - c\rangle$ 以及 $\langle -v_*$、 $B_k^\top\mathbf{1}_n - c\rangle$ 中给出界，并完成9的证明。 □

3.5 迭代复杂度界

本节将给出算法1的迭代复杂度界。

定理 3.8. 给定成本矩阵 $C \in \mathbb{R}^{n\times n}$ 和两个单纯形 r, $c \in \mathbb{R}^n_+$，存在一个算法（算法 *1*）输出 B（u_k, v_k）（定义 *3.1*），满足

$$\|B(u_k, v_k)\mathbf{1}_n - r\|_1 + \|B(u_k, v_k)^\top \mathbf{1}_n - c\|_1 \le \epsilon_0$$

在迭代次数 k 内满足

$$k \le 2 + \frac{4R}{\epsilon_0}$$

证明。我们首先考虑 $k \ge 1$ 为偶数，并定义 $B_k := B(u_k, v_k)$。我们有 e

$$\begin{aligned}
&\psi(u_k, v_k) - \psi(u_{k+1}, v_{k+1}) \\
=\ &\langle \mathbf{1}_n, B_k\mathbf{1}_n\rangle - \langle \mathbf{1}_n, B_{k+1}\mathbf{1}_n\rangle + \langle u_{k+1} - u_k, r\rangle + \langle v_{k+1} - v_k, c\rangle \\
=\ &\langle r, u_{k+1} - u_k\rangle \\
=\ &\langle r, \ln r - \ln(B_k\mathbf{1}_n)\rangle \\
=\ &\mathrm{KL}(r\|B_k\mathbf{1}_n)
\end{aligned} \tag{10}$$

于是，我们得到

$$\begin{aligned}
&\widetilde{\psi}(u_k, v_k) - \widetilde{\psi}(u_{k+1}, v_{k+1}) \\
=\ &\psi(u_k, v_k) - \psi(u_{k+1}, v_{k+1}) \\
=\ &\mathrm{KL}(r\|B_k\mathbf{1}_n) \\
\ge\ &\frac{1}{2}\|B_k\mathbf{1}_n - r\|_1^2 \\
\ge\ &\max\{\frac{\widetilde{\psi}(u_k, v_k)^2}{2R^2}, \frac{\epsilon_0^2}{2}\},
\end{aligned} \tag{11}$$

其中，第一步根据 ψ 的定义得到，第二步根据式（10）得到，第三步根据 Pinsker 不等式得到，最后一步根据引理 3.7 和 $B_k^\top\mathbf{1}_n = c$得到。对于最后一步，我们还用到了：只要尚未满足停止准则且 $B_k^\top\mathbf{1}_n = c$, $\|B_k\mathbf{1}_n - r\|_1^2 \ge \epsilon_0^2$。类似地，当 k 为奇数时，我们也可以证明相同的不等式。

给定 $\ell = \frac{2R^2}{\widetilde{\psi}(u_1, v_1)}$，利用引理 3.9，对于任意 $k \ge 1$

$$\frac{\widetilde{\psi}(u_k, v_k)}{2R^2} \le \frac{1}{k + \ell - 1}$$

因此，

$$k \le 1 + \frac{2R^2}{\widetilde{\psi}(u_k, v_k)} - \frac{2R^2}{\widetilde{\psi}(u_1, v_1)} \tag{12}$$

另一方面，

$$\widetilde{\psi}(u_{k+m}, v_{k+m}) \le \widetilde{\psi}(u_k, v_k) - \frac{\epsilon_0^2 m}{2}, \quad k, m \ge 0 \tag{13}$$

接下来，我们采用一种切换策略，以数字 $s \in (0, \widetilde{\psi}(u_1, v_1)]$, 为参数，将式（15）与式（13）结合起来。

首先，利用式 (15)，我们计算将 $\widetilde{\psi}(u, v)$ 从其初始值 $\widetilde{\psi}(u_1, v_1)$ 降至某个值 s 所需的迭代次数。然后，根据式 (13)，并根据定义给出 $\widetilde{\psi}(u, v) \geq 0$(,)，我们计算进一步将 $\widetilde{\psi}(u, v)$ 从 降至零所需的迭代次数。通过在 $s \in (0, \widetilde{\psi}(u_1, v_1)]$, 中最小化这两个估计值之和， k 总迭代次数满足以下条件

$$
\begin{aligned}
k &\leq \min_{0<s\leq\widetilde{\psi}(u_1,v_1)} (2 + \frac{2R^2}{s} - \frac{2R^2}{\widetilde{\psi}(u_1, v_1)} + \frac{2s}{\epsilon_0^2}) \\
&= \begin{cases} 2 + \frac{4R}{\epsilon_0} - \frac{2R^2}{\widetilde{\psi}(u_1,v_1)}, & \widetilde{\psi}(u_1, v_1) \geq R\epsilon_0, \\ 2 + \frac{2\widetilde{\psi}(u_1,v_1)}{\epsilon_0^2}, & \widetilde{\psi}(u_1, v_1) < R\epsilon_0. \end{cases}
\end{aligned}
$$

其中，第一步由式 (12) 得出，最后一步的前半部分由 $a + b \geq 2\sqrt{ab}$ 对于 $a \geq 0$, $b \geq 0$ 而后半部分由 $s = \widetilde{\psi}(u_1, v_1)$ 得出。两种情况下，我们都有 $k \leq 2 + \frac{4R}{\epsilon_0}$. □

3.6 归纳法

这里，我们给出势函数上界的归纳证明。

引理 3.9. 对所有 $k \geq 1$,

$$
\frac{\widetilde{\psi}(u_k, v_k)}{2R^2} \leq \frac{1}{k + \ell - 1},
$$

其中 $\ell := \frac{2R^2}{\widetilde{\psi}(u_1, v_1)}$(,) 并且 $\widetilde{\psi}$ 定义见定义 3.2。

证明。我们的证明分为两部分。首先，我们考察上述不等式在 $k = 1$ 下的正确性。然后，对 $k > 1$ 进行归纳，即可完成证明。

基础情形。对于 $k = 1$。

$$
\begin{aligned}
&\frac{\widetilde{\psi}(u_1, v_1)}{2R^2} \\
&= \frac{1}{\ell} \\
&= \frac{1}{k + \ell - 1},
\end{aligned}
$$

其中，第一步由 ℓ 的定义得出，最后一步由 $k - 1 = 0$ 得出。因此，我们有 $\psi(\frac{u_k}{2R^2}, \frac{v_k}{}) \leq \frac{1}{k+\ell-1}$(,) 当 k=1 时。

一般情形。假设，

$$
\frac{\widetilde{\psi}(u_k, v_k)}{2R^2} \leq \frac{1}{k + \ell - 1} \tag{14}
$$

然后我们可以证明

$$
\begin{aligned}
&\frac{\widetilde{\psi}(u_{k+1}, v_{k+1})}{2R^2} \\
&\leq \frac{\widetilde{\psi}(u_k, v_k)}{2R^2} - (\frac{\widetilde{\psi}(u_k, v_k)}{2R^2})^2
\end{aligned}
$$

$$\leq \frac{1}{k+\ell-1} - (\frac{1}{k+\ell-1})^2$$
$$\leq \frac{1}{k+\ell}, \tag{15}$$

其中第一步由式（11）得出，第二步由式（14）以及函数 $f(x) = x - x^2$（即 $f(y) \leq f(z)$，若 $y \leq z \leq 1/2$）的性质得出，最后一步由$\frac{1}{A} - \frac{1}{A^2} \leq \frac{1}{A+1}$对任意整数 $A \geq 2$得出。根据归纳法，证明完成。□

4 小树宽情形下的运行时间

在第 4.1 节中，我们介绍 K 的隐式形式。在第 4.2 节中，我们给出了适用于小树宽情形的更快 Sinkhorn 算法。在第 4.3 节中，我们证明了舍入算法的正确性。在第 4.4 节中，我们给出了舍入算法所需的运行时间。在第 4.5 节中，我们给出了使用更快的 Sinkhorn 算法近似最优传输距离所需的运行时间。

4.1 K的隐式形式

Here 我们介绍 K 的隐式形式，以利用小树宽情形 .

引理 4.1。 假设 $C = MM^\top \in \mathbb{R}^{n\times n}$，其中 $M \in \mathbb{R}^{n\times d}$的树宽为 τ。给定 $A :=K - D$，其中 $D_{i,\ j} := 1$ 对于 $i,\ j \in [n]$ 且 K 定义见定义 *3.1*，乔列斯基因子 L_A 对于$A = L_A L_A^\top$ 是 τ列稀疏的。

证明。 给定 $C = MM^\top$和 M的树宽为 τ，根据引理 L_C， $C = MM^\top = L_M L_M^\top$的乔列斯基因子 τ 按列是稀疏的。2.12。由于

$$A_{i,j} = e^{-C_{i,j}/\gamma} - 1,$$

我们有 $A_{i,\ j} = 0$ 当 $C_{i,\ j} = 0$。因此，矩阵 A 与矩阵 C同样稀疏。乔列斯基因子 L_A对于 $A = L_A L_A^\top$与 L_M同样稀疏。由于 L_M是 τ-稀疏的，证明完毕。□

4.2 小树宽下 Sinkhorn 的运行时间

本节旨在证明算法 2的运行时间。

定理 4.2 (算法 2 的运行时间）。 给定成本矩阵 $C \in \mathbb{R}^{n\times n}$，其树宽较小$\tau$以及两个单纯形 r，$c \in \mathbb{R}^n_+$，存在一种算法（算法 *2*），每次迭代耗时 $O(n\tau)$，初始化耗时 $O(n\tau^2)$，并输出

- 下三角矩阵 L_A
- 向量 u、 v、 $w \in \mathbb{R}^n$

使得 $B(u_k, v_k) \in \mathbb{R}^{n\times n}$ 可以通过以下方式（隐式地）构造

$$B(u_k, v_k) = \mathrm{diag}(e^{u_k})(L_A L_A^\top)\,\mathrm{diag}(e^{v_k}) + \mathrm{diag}(e^{u_k})(ww^\top)\,\mathrm{diag}(e^{v_k})$$

满足

$$\|B(u_k, v_k)\mathbf{1}_n - r\|_1 + \|B(u_k, v_k)^\top \mathbf{1}_n - c\|_1 \leq \epsilon_0.$$

算法 2 树宽较小时的 Sinkhorn 算法

过程 SinkhornAlgorithm($r \in \mathbb{R}^n, c \in \mathbb{R}^n, \epsilon_0 \in (0,1)$) ▷ 定理 4.2 ▷精度 ϵ_0 $k \leftarrow 0$, $u_0 \leftarrow 0$ $v_0 \leftarrow 0$ $w \leftarrow \mathbf{1}_n$ $x_0 \leftarrow e^{-C_{i,j}/\gamma}\mathbf{1}_n$ $y_0 \leftarrow (e^{-C_{i,j}/\gamma})^\top \mathbf{1}_n$ 隐式构造 $D = ww^\top$ 隐式构造 $A \in \mathbb{R}^{n \times n}$，其中 $A_{i,j} = e^{-C_{i,j}/\gamma} - 1$ ▷显式写出 A 需要 n^2，然而，我们从不需要显式写出A。知道 A 的确切形式就足以进行 Cholesky 分解 $L \leftarrow$ A 的 Cholesky 分解矩阵，即 $A = L_A L_A^\top$ ▷ $O(n\tau^2)$，引理 2.11 当 $\|x_k - r\|_1 + \|y_k - c\|_1 \geq \epsilon_0$ 时执行 如果 k 模 $2 = 0$ 则 $u_{k+1} \leftarrow u_k + \ln r - \ln(x_k)$ $v_{k+1} \leftarrow v_k$ 否则 $v_{k+1} \leftarrow v_k + \ln c - \ln(y_k)$ $u_{k+1} \leftarrow u_k$ 结束如果 $x_k \leftarrow (\mathrm{diag}(e^{u_k})(L_A L_A^\top)\,\mathrm{diag}(e^{v_k}) + \mathrm{diag}(e^{u_k}) D\,\mathrm{diag}(e^{v_k}))\mathbf{1}_n$ $y_k \leftarrow (\mathrm{diag}(e^{u_k})(L_A L_A^\top)\,\mathrm{diag}(e^{v_k}) + \mathrm{diag}(e^{u_k}) D\,\mathrm{diag}(e^{v_k}))^\top \mathbf{1}_n$ $k \leftarrow k + 1$ 结束循环 返回 u_k, v_k, L_A, w ▷我们以隐式方式返回 $B(u_k, v_k)$，即$B(u_k, v_k) = \mathrm{diag}(e^{u_k})(L_A L_A^\top)\,\mathrm{diag}(e^{v_k}) + \mathrm{diag}(e^{u_k})(ww^\top)\,\mathrm{diag}(e^{v_k})$。 结束过程

证明。每一步的运行时间如下所示：

- 写出成本矩阵 $C \in \mathbb{R}^{n \times n}$ 需要 $O(n\tau)$ 时间，因为 $\mathrm{nnz}(C) = n\tau$，根据断言2.13。
- 隐式写出矩阵 $D \in \mathbb{R}^{n \times n}$，需要 $O(n)$ 时间，因为 $D \in \mathbb{R}^{n \times n}$ 是一个秩一矩阵。
- 初始化 x_0和 y_0需要 $O(n\tau)$，因为 $\mathrm{nnz}(C) = n\tau$。
- 根据引理4.1，我们知道 L_A 是 τ-稀疏列矩阵。然后，使用引理2.11，计算 A 的 Cholesky 分解需要 $O(n\tau^2)$ 时间。。
- 计算 $\mathrm{diag}(e^{u_k})(L_A L_A^\top)\,\mathrm{diag}(e^{v_k})$ 需要 $O(n\tau)$ 时间，因为 L_A在列上是 τ-稀疏的。
- 计算 $\mathrm{diag}(e^{u_k}) D\,\mathrm{diag}(e^{v_k})$ 需要 $O(n)$ 时间，因为矩阵 D 是秩一矩阵。
- 更新 $u \in \mathbb{R}^n, v \in \mathbb{R}^n$ 需要 $O(n)$ 时间。

因此，算法 2 的初始化时间为 $O(n\tau^2)$，每次迭代的运行时间为$O(n\tau)$。

□

算法 3 Sinkhorn 近似最优传输

过程 ApproxOT(ϵ) ▷ 定理4.5

▷精度 ϵ

$\gamma \leftarrow \frac{\epsilon}{4 \ln n}$

$\epsilon_0 \leftarrow \frac{\epsilon}{8\|C\|_\infty}$

▷求取 $\widetilde{r}, \widetilde{c} \in \Delta^n$ 使得 $\|\widetilde{r} - r\|_1 \le \epsilon_0/4, \|\widetilde{c} - c\|_1 \le \epsilon_0/4$ 且 $\min_{i\in[n]} \widetilde{r}_i \ge \epsilon_0/(8n)$, $\min_{j\in[n]} \widetilde{c}_j \ge \epsilon_0/(8n)$。

$(\widetilde{r}, \widetilde{c}) \leftarrow (1 - \frac{\epsilon_0}{8})((r, c) + \frac{\epsilon_0}{n(8-\epsilon_0)}(\mathbf{1}_n, \mathbf{1}_n))$

$(u, v, L, w) \leftarrow$ SinkhornAlgorithm$(\widetilde{r}, \widetilde{c}, \epsilon_0/2)$ ▷算法 2

▷注意，u, v, L, w 是 B的隐式表示，即 $\mathrm{diag}(e^{u_k})(L_A L_A^\top)\,\mathrm{diag}(e^{v_k}) + \mathrm{diag}(e^{u_k})(ww^\top)\,\mathrm{diag}(e^{v_k})$

$(p, q, X, Y, w, u, v) \leftarrow$ Round(u, v, L, w, r, c) ▷算法 4

return (p, q, X, Y, L, w, u, v) ▷我们返回 $\widehat{X}$ 以隐式方式，即，$\widehat{X} := XBY + pq^\top/\|p\|_1$

结束过程

算法 4 对 B 在 $\mathcal{U}$ 上的投影进行舍入

过程 Round$(u \in \mathbb{R}^n, v \in \mathbb{R}^n, L \in \mathbb{R}^{n\times n}, w \in \mathbb{R}^n, r \in \mathbb{R}^n, c \in \mathbb{R}^n)$ ▷引理 4.4

▷ L 是一个仅有 $O(n\tau)$ 个非零元素的下三角矩阵

▷我们从不显式写出 B。 B 可以隐式表示为 $\mathrm{diag}(e^u)(L_A L_A^\top)\,\mathrm{diag}(e^v) + \mathrm{diag}(e^u)(ww^\top)\,\mathrm{diag}(e^v)$

$X \leftarrow \mathrm{diag}(x)$，其中 $x_i = \min\{\frac{r_i}{r_i(B)}, 1\}$ ▷ $r(B) := B\mathbf{1}_n$, $X \in \mathbb{R}^{n\times n}$

$B_0 \leftarrow XB$ ▷我们只隐式构造 B_0

$Y \leftarrow \mathrm{diag}(y)$ 与 $y_j = \min\{\frac{c_j}{c_j(B_0)}, 1\}$ ▷ $c(B_0) := B_0^\top \mathbf{1}_n$

$B_1 \leftarrow B_0 Y$ ▷我们仅以隐式方式构造 B_1

$p \leftarrow r - B_1\mathbf{1}_n$, $q \leftarrow c - B_1^\top \mathbf{1}_n$

返回 p, q, X, Y, w, u, v ▷我们以隐式方式返回 G，即 $G := XBY + pq^\top/\|p\|_1$

结束过程

4.3 舍入算法的正确性

我们首先证明舍入算法（算法 4）的正确性。

引理 4.3 (引理 7 在 [ANWR17]中的改进版本。 给定 r, $c \in \triangle_n$, $B \in \mathbb{R}_+^{n\times n}$, u, $v \in \mathbb{R}^n$和 r, $c \in \mathbb{R}^n$, 存在一个算法（算法 4）输出

- 一个对角矩阵 $X \in \mathbb{R}^{n\times n}$
- 一个对角矩阵 $Y \in \mathbb{R}^{n\times n}$
- 一个下三角矩阵 L_A
- 向量 $u, v, w \in \mathbb{R}^n$
- 向量 $p \in \mathbb{R}^n$, $q \in \mathbb{R}^n$

使得 $G \in \mathcal{U}_{r,\ c}$ 可以通过以下方式（隐式地）构造

$$\widehat{X} = X(\mathrm{diag}(e^u) L_A L_A^\top \mathrm{diag}(e^v) + \mathrm{diag}(e^u)(ww^\top)\,\mathrm{diag}(e^v))Y + pq^\top/\|p\|_1$$

满足以下条件

$$\|G - B\|_1 \leq 2(\|B\mathbf{1}_n - r\|_1 + \|B^\top \mathbf{1}_n - c\|_1).$$

证明。令 G 为算法 4的输出。由于矩阵 B_1 均为非负，且输出 q 和p 均为负，其中 $\|p\|_1 = \|q\|_1 = 1 - \|B_1\|_1$，矩阵 G 均为非负，并且

$$\begin{aligned} r(G) &= r(B_1) + r(pq^\top / \|p\|_1) \\ &= r(B_1) + p \\ &= r, \end{aligned} \tag{16}$$

其中，我们记 $r(A) := A\mathbf{1}_n, c(A) := A^\top \mathbf{1}_n$ ，前两个步骤根据r 的定义得出，最后一步根据 $p = r - B_1\mathbf{1}_n$得出。同理，我们有 $c(G) = c$。因此，我们有$G \in \mathcal{U}_{r,\ c}$。

接下来，我们记 $\Delta := \|B\|_1 - \|B_1\|_1$，并证明矩阵 B 与矩阵G之间的 ℓ_1 界。首先，当 $r_i(B) \geq r_i$ 时，我们从 B 的一行中移除质量；然后，当 $c_j(B_0) \geq c_j$时，我们从一列中移除质量。现在，我们有

$$\Delta = \sum_{i=1}^{n} (r_i(B) - r_i)_+ + \sum_{j=1}^{n} (c_j(B_0) - c_j)_+. \tag{17}$$

接下来，我们分析等式 (17)。首先，对于等式 (17) 左侧的求和，我们有

$$\sum_{i=1}^{n} (r_i(B) - r_i)_+ = \frac{1}{2}(\|r(B) - r\|_1 + \|B\| - 1).$$

对于式（17）中的第二个求和项。

$$\sum_{j=1}^{n} (c_j(B_0) - c_j)_+ \leq \sum_{j=1}^{n} (c_j(B) - c_j)_+ \leq \|c(B) - c\|_1$$

其中，第一步是因为向量 $c(B)$ 逐项大于 $c(B_0)$，最后一步则来自 c 的定义。

因此我们得出

$$\begin{aligned} \|G - B\|_1 &\leq \Delta + \|pq^\top\|_1 / \|p\|_1 \\ &= \Delta + 1 - \|B_1\|_1 \\ &= 2\Delta + 1 - \|B\|_1 \\ &\leq \|r(B) - r\|_1 + 2\|c(B) - c\|_1 \\ &\leq 2(\|r(B) - r\|_1 + \|c(B) - c\|_1) \end{aligned} \tag{18}$$

其中，第一步来自 Δ 的定义，第二步来自以下事实：$\|p\|_1 = \|q\|_1 = 1 - \|B_1\|_1$，第三步来自 Δ 的定义，第四步根据式（17），最后一步通过重新整理得到。至此，证明完成。

□

4.4 舍入算法的运行时间

Ne接下来，我们说明舍入算法（算法 4）所需的运行时间 .

引理 4.4 (引理 7 的改进版本，见 [ANWR17])。 给定 r， $c \in \triangle_n$， $B \in \mathbb{R}_+^{n \times n}$， u， $v \in \mathbb{R}^n$ 和 r，$c \in \mathbb{R}^n$，存在一个算法（算法 4）输出

- 一个对角矩阵 $X \in \mathbb{R}^{n \times n}$
- 一个对角矩阵 $Y \in \mathbb{R}^{n \times n}$
- 一个下三角矩阵 L_A
- 向量 $u, v, w \in \mathbb{R}^n$
- 向量 $p \in \mathbb{R}^n$, $q \in \mathbb{R}^n$

使得 $G \in \mathcal{U}_{r,c}$ 可以通过以下方式（隐式地）构造

$$\widehat{X} = X(\operatorname{diag}(e^u) L_A L_A^\top \operatorname{diag}(e^v) + \operatorname{diag}(e^u)(ww^\top) \operatorname{diag}(e^v))Y + pq^\top / \|p\|_1$$

满足

$$\|G - B\|_1 \leq 2(\|B\mathbf{1}_n - r\|_1 + \|B^\top \mathbf{1}_n - c\|_1)$$

用时 $O(n\tau)$。

证明。每个步骤的运行时间如下所示：

- 计算 $r(B)$ 需要 $O(n\tau)$ 时间。给定

  $$r(B) = B\mathbf{1}_n = \operatorname{diag}(e^{u_k})(L_A L_A^\top)\mathbf{1}_n \operatorname{diag}(e^{v_k}) + \operatorname{diag}(e^{u_k})(ww^\top)\mathbf{1}_n \operatorname{diag}(e^{v_k}),$$

  计算 $L_A(L_A^\top \mathbf{1}_n)$ 需要 $O(n\tau)$，因为 $\mathsf{nnz}(L_A) = n\tau$。由于 $w = \mathbf{1}_n$，计算 $(ww^\top)\mathbf{1}_n$ 需要$O(n)$。

- 计算 $X = \mathsf{diag}(x)$，其中 $x_i = \min\{\frac{r_i}{r_i(B)}, 1\}$，需要 $O(n)$ 时间。
- 对于 $B_0 = XB$，我们指出， B_0 并未显式写出。它由 L_A, w, u, v, X 隐式表示。
- 类似地，我们可以在 $O(n)$ 时间内计算 Y，并隐式写出 B_1。
- 我们有

  $$B_1\mathbf{1}_n = XBY = \operatorname{diag}(e^{u_k})X(L_A L_A^\top)Y\mathbf{1}_n \operatorname{diag}(e^{v_k}) + \operatorname{diag}(e^{u_k})X(ww^\top)Y\mathbf{1}_n \operatorname{diag}(e^{v_k}).$$

  对于任意对角矩阵 M， $M \cdot L_A$ 与 L_A 一样稀疏，且计算它需要 $O(n\tau)$。因此，计算 $P = \mathsf{diag}(e^{u_k})X(L_A L_A^\top)Y\mathsf{diag}(e^{v_k})$ 需要 $O(n\tau)$ 时间，并且 P 是 $n\tau$-稀疏的。然后，我们计算 $P\mathbf{1}_n$，这需要 $O(n\tau)$ 时间。因此，更新 p 需要 $O(n\tau)$ 时间。

- 类似地，更新 q 需要 $O(n\tau)$ 时间。
- 对于矩阵 阵 G，它以隐式方式返回。我们用 p, q, X, Y, w, u, v 表示 it.

因此，总运行时间为 $O(n\tau)$。 □

4.5 Sinkhorn 算法求解最优传输距离的运行时间

我们的最优传输算法的核心是熵正则项

$$X_\gamma := \arg\min_{X \in \mathcal{U}_{r,c}} \langle X, C \rangle + \gamma \cdot \mathcal{R}(X). \tag{19}$$

通过分析其最优性的一阶条件，可以明确刻画方程（19）的解。

现在，我们应用上一小节的结果，推导出求解$\backsim X \in \mathcal{U}(r,c)$并满足方程（2）的复杂度估计。使用 Sinkhorn 算法近似最优传输距离的过程列于算法3。

定理 4.5. 存在一个算法（算法 3），以成本矩阵 $C \in \mathbb{R}^{n\times n}$、两个 n 维单纯形 r，c 作为输入，并输出

- 一个对角矩阵 $X \in \mathbb{R}^{n\times n}$
- 一个对角矩阵 $Y \in \mathbb{R}^{n\times n}$
- 下三角矩阵 L_A
- 向量 u、v、$w \in \mathbb{R}^n$
- 向量 $p \in \mathbb{R}^n$、$q \in \mathbb{R}^n$

$\widehat{}$使得 $X \in \mathcal{U}(r$、$c)$ 可以通过以下方式（隐式地）构造

$$\widehat{X} = X(\operatorname{diag}(e^u) L_A L_A^\top \operatorname{diag}(e^v) + \operatorname{diag}(e^u)(ww^\top) \operatorname{diag}(e^v))Y + pq^\top / \|p\|_1$$

满足公式(2) 的

$$O(n\tau^2 + \epsilon^{-2} n\tau \|C\|_\infty^2 \ln n)$$

时间。

备注 4.6。如果不要求输出格式为下三角矩阵，则可以去掉加法项 $n\tau^2$ 。

证明。令 $X_* \in \arg\min_{X\in\mathcal{U}_{r,c}} \langle P, C \rangle$ 为原始最优传输规划的一个最优解。

我们首先证明，$\langle B, C \rangle$ 不会比 $\langle X_*, C \rangle$ 大很多。

由于 $B = MAN \in \mathbb{R}^{n\times n}$ 对于正对角矩阵 $M, N \in \mathbb{R}_+^{n\times n}$，引理 2.6 表明 B 是以下问题的最优解

$$\arg\min_{X \in \mathcal{U}_{r,c}} \langle X, C \rangle + \gamma \mathcal{R}(X). \tag{20}$$

根据引理 4.4，存在一个矩阵 $X_0 \in \mathcal{U}_{B\mathbf{1}_n, B^\top \mathbf{1}_n}$（定义 2.5），使得

$$\|X_0 - X_*\|_1 \le 2(\|B\mathbf{1}_n - r\|_1 + \|B^\top \mathbf{1}_n - c\|_1). \tag{21}$$

此外，由于 $B \in \mathbb{R}^{n\times n}$ 是方程（20）的最优解，因此我们有

$$\langle B, C \rangle + \gamma \mathcal{R}(B) \le \langle X_0, C \rangle + \gamma \mathcal{R}(X_0). \tag{22}$$

因此，我们有

$$\begin{aligned}\langle B, C\rangle - \langle X_*, C\rangle &= \langle B, C\rangle - \langle X_0, C\rangle + \langle X_0, C\rangle - \langle X_*, C\rangle \\ &= \langle B, C\rangle - \langle X_0, C\rangle + \|X_0 - X_*\|_1 \|C\|_\infty \\ &\leq \gamma(H(B) - H(X_0)) + \|X_0 - X_*\|_1 \|C\|_\infty \\ &\leq \gamma(H(B) - H(X_0)) + 2(\|B\mathbf{1}_n - r\|_1 + \|B^\top \mathbf{1}_n - c\|_1)\|C\|_\infty \\ &\leq 2\gamma \ln n + 2(\|B\mathbf{1}_n - r\|_1 + \|B^\top \mathbf{1}_n - c\|_1)\|C\|_\infty \end{aligned} \tag{23}$$

其中，第一步由重新整理得到，第二步由 H¨older 不等式（引理 2.7）得到，第三步由等式（22）和 $\mathcal{R}(X)$ 得到，$= -H(X)$ 得到第四步，等式（21）得到最后一步；这是因为 $0 < H(B)\ H(X_0) \leq 2\ln n$。

^引理 4.4 表明，算法 X 的输出满足以下不等式：4

$$\|B - \widehat{X}\|_1 \leq 2(\|B\mathbf{1}_n - r\|_1 + \|B^\top \mathbf{1}_n - c\|_1). \tag{24}$$

^回顾一下，X 是算法 3的输出，X_* 是等式（2）中最优传输问题的一个解，而 B是算法 3 第 7 行中得到的矩阵。我们有

$$\begin{aligned}\langle \widehat{X}, C\rangle &= \langle \widehat{X} - B, C\rangle + \langle B, C\rangle \\ &\leq \|\widehat{X} - B\|_1 \|C\|_\infty + \langle B, C\rangle \\ &\leq 2(\|B\mathbf{1}_n - r\|_1 + \|B^\top \mathbf{1}_n - c\|_1)\|C\|_\infty + \langle B, C\rangle \\ &\leq \langle X_*, C\rangle + 2\gamma \ln n + 4(\|B\mathbf{1}_n - r\|_1 + \|B^\top \mathbf{1}_n - c\|_1)\|C\|_\infty. \end{aligned} \tag{25}$$

其中，第一步由重新整理得到，第二步由 H¨older 不等式得到，第三步由等式（24）得到，最后一步由等式（23）得到。

同时，我们有

$$\begin{aligned}&\|B\mathbf{1}_n - r\|_1 + \|B^\top \mathbf{1}_n - c\|_1 \\ \leq\ &\|B\mathbf{1}_n - \widetilde{r}\|_1 + \|\widetilde{r} - r\|_1 + \|B^\top \mathbf{1}_n - \widetilde{c}\|_1 + \|\widetilde{c} - c\|_1 \\ \leq\ &\epsilon_0 \end{aligned}$$

其中，第一步根据 ℓ_1-范数的定义得到，最后一步根据 $\|B\mathbf{1}_n - r\|_1 + \|B^\top \mathbf{1}_n - c\|_1 \leq \epsilon_0$（算法 2的输出）以及 $\widetilde{r}$和 $\widetilde{c}$的定义得到。

令 $\gamma = \frac{\epsilon}{4\ln n}$和 $\epsilon_0 = \frac{\epsilon}{8\|C\|_\infty}$，由上述不等式和式（25）可得 $\widehat{X}$满足不等式（2）。

接下来，我们分析算法 3 的复杂度。当 ϵ_0 足够小时，根据定理 3.8，算法 3 第 7 行中 Sinkhorn 算法的迭代次数为 $O(R/\epsilon_0)$。根据定义 3.3，我们有

$$\begin{aligned}R &= -\ln(K_{\min} \min_{i,j\in[n]}\{\widetilde{r}_i, \widetilde{c}_j\}) \\ &= -\ln(e^{-\|C\|_\infty/\gamma} \min_{i,j\in[n]}\{\widetilde{r}_i, \widetilde{c}_j\}) \\ &\leq \frac{\|C\|_\infty}{\gamma} - \ln(\frac{\epsilon_0}{8n}), \end{aligned}$$

其中第一步根据 R的定义得出，第二步根据$K_{\min}$的定义得出，最后一步根据 $\widetilde{r}_i$的条件得出，$\widetilde{c}_j$ 见算法3 第 6行。

由于 $\gamma = \frac{\epsilon}{4 \ln n}$ 和 $\epsilon_0 = \frac{\epsilon}{8\|C\|_\infty}$，因此我们有

$$R = O(\epsilon^{-1}\|C\|_\infty \ln n).$$

由于算法 3 的迭代次数为 $O(R/\epsilon_0)$，我们得出 Sinkhorn 算法的总迭代次数上界为 $O(\epsilon^{-2}\|C\|_\infty^2 \ln n)$。

显然，$\widetilde{r} \in \mathbb{R}_+^n$ 和 $\widetilde{c} \in \mathbb{R}_+^n$ 可在算法 3 第 6 行中以 $O(n)$ 时间求得。

由于 Sinkhorn 算法的每次迭代需要 $O(n\tau)$ 时间，且如定理 $O(n\tau^2)$ 所示，初始化需要 4.2 时间，因此算法 3 的总复杂度为

$$O(n\tau^2 + \epsilon^{-2}n\tau\|C\|_\infty^2 \ln n).$$

□

5 对称

本节给出一种算法（算法 6），在两个分布相同（即 $O(\epsilon^{-2}n\tau\|C\|_\infty^2 \ln n)$ 的条件下，可在 $c = r$ 时间内求解最优传输问题。

定义 5.1. 给定对称最优传输问题 $\arg\min_{X \in \mathcal{U}_r} \langle X \rangle$，定义

$$\mathcal{U}_r = \{X \in \mathbb{R}_+^{n \times n} : X\mathbf{1}_n = r, X^\top \mathbf{1}_n = r\}$$

其中 $\mathbf{1}_n$ 是 $\mathbb{R}^n$ 中的全一向量，$C \in \mathbb{R}_+^{n \times n}$ 是给定的成本矩阵，而 $r \in \mathbb{R}^n$ 是给定的正元素向量，且其元素之和为一。

我们首先给出对称情形下 Sinkhorn 算法（算法 5）的运行时间。

定理 5.2 （算法 5 的运行时间）。给定具有小树宽的成本矩阵 $C \in \mathbb{R}^{n \times n}$ 和τ 单纯形 $r \in \mathbb{R}_+^n$，存在一种算法（算法 5），每次迭代耗时 $O(n\tau)$，初始化耗时$O(n\tau^2)$，并输出

- 一个下三角矩阵 L_A
- 向量 $u, w \in \mathbb{R}^n$

使得 $B(u_k) \in \mathbb{R}^{n \times n}$ 可以通过以下方式（隐式地）构造

$$B(u_k) = \mathrm{diag}(e^{u_k})(L_A L_A^\top)\,\mathrm{diag}(e^{u_k}) + \mathrm{diag}(e^{u_k})(ww^\top)\,\mathrm{diag}(e^{u_k})$$

满足

$$\|B(u_k)\mathbf{1}_n - r\|_1 \leq \epsilon_0.$$

证明。类似于定理 4.2 的证明，此处两个分布相同，即 $c = r$. □

接下来，我们展示对称情形下舍入算法（算法 7）的运行时间。

引理 5.3 （引理 7 的改进版本，见[ANWR17]）。给定 $r \in \triangle_n$，$B \in \mathbb{R}_+^{n \times n}$，$u \in \mathbb{R}^n$，存在一个算法（算法 7）输出

- 一个对角矩阵 $X \in \mathbb{R}^{n \times n}$
- 一个对角矩阵 $Y \in \mathbb{R}^{n \times n}$

Algo法 5 适用于小树宽对称分布的 Sinkhorn 算法

过程 SinkhornAlgorithmSym($r, \epsilon_0 \in (0,1)$) ▷ 定理5.2

▷精度 ϵ_0

$k \leftarrow 0$,

$u_0 \leftarrow 0$

$v_0 \leftarrow 0$

$w \leftarrow \mathbf{1}_n$

$x_0 \leftarrow e^{-C_{i,j}/\gamma}\mathbf{1}_n$

$y_0 \leftarrow (e^{-C_{i,\ j}/\gamma})^\top \mathbf{1}_n$

隐式构造 $D = ww^\top$

隐式构造 $A \in \mathbb{R}^{n\times n}$，其中 $A_{i,\ j} = e^{-C_{i,j}/\gamma} - 1$

▷显式写出 A 需要 n^2，不过，我们从不需要显式写出 A。知道 A 的精确形式就足以进行 Cholesky 分解

$L \leftarrow$ A 的 Cholesky 分解矩阵，即 $A = L_A L_A^\top$ ▷ $O(n\tau^2)$，引理 2.11

当 $\|x_k - r\|_1 \geq \epsilon_0$ 时执行

$u_{k+1} \leftarrow u_k + \ln r - \ln(x_k)$

$x_k \leftarrow (\mathrm{diag}(e^{u_k})(L_A L_A^\top)\,\mathrm{diag}(e^{u_k}) + \mathrm{diag}(e^{u_k}) D\,\mathrm{diag}(e^{u_k}))\mathbf{1}_n$

$k \leftarrow k + 1$

结束循环

返回 u_k、L_A、w ▷我们以隐式方式返回 $B(u_k)$，即，$B(u_k) = \mathrm{diag}(e^{u_k})(L_A L_A^\top)\,\mathrm{diag}(e^{u_k}) + \mathrm{diag}(e^{u_k})(ww^\top)\,\mathrm{diag}(e^{u_k})$。

结束过程

算法 6 对称分布的 Sinkhorn 近似最优传输

过程 ApproxOTSym(ϵ) ▷ 定理5.4

▷精度 ϵ

$\gamma \leftarrow \frac{\epsilon}{4\ln n}$

$\epsilon_0 \leftarrow \frac{\epsilon}{8\|C\|_\infty}$

▷求解 $\widetilde{r} \in \Delta^n$，使得 $\|\widetilde{r} - r\|_1 \leq \epsilon_0/4$，并使$\sum_{i\in[n]} \widetilde{r}_i \geq \epsilon_0/(8n)$最小。

$\widetilde{r} \leftarrow (1 - \frac{\epsilon_0}{8})(r + \frac{\epsilon_0}{n(8-\epsilon_0)}\mathbf{1}_n)$

$(u, L, w) \leftarrow$ SinkhornAlgorithm($\widetilde{r}$,$\epsilon_0/2$) ▷算法 5

▷注意，u, v, L, w 是 B的一种隐式表示，即 $\mathrm{diag}(e^u)(L_A L_A^\top)\,\mathrm{diag}(e^u) + \mathrm{diag}(e^u)(ww^\top)\,\mathrm{diag}(e^u)$

$(p,\ X, Y, w, u) \leftarrow$ Round(u, L, w, r) ▷算法 7

return $(p,\ X, Y, L, w, u)$ ▷我们返回 $\widehat{X}$ 以隐式方式，即，$\widehat{X} := XBY + pp^\top/\|p\|_1$

结束过程

- 一个下三角矩阵 L_A
- 向量 u、w、$p \in \mathbb{R}^n$

使得可以通过以下方式（隐式地）构造 $G \in \mathcal{U}_r$

$$\widehat{X} = X(\mathrm{diag}(e^u)L_A L_A^\top \mathrm{diag}(e^u) + \mathrm{diag}(e^u)(ww^\top)\,\mathrm{diag}(e^u))Y + pp^\top/\|p\|_1$$

满足该条件

$$\|G - B\|_1 \leq 2(\|B\mathbf{1}_n - r\|_1)$$

对称分布下 B 在 $\mathcal{U}$ 上的投影的舍入算法 7

过程 RoundSym($u \in \mathbb{R}^n, L \in \mathbb{R}^{n \times n}, w \in \mathbb{R}^n, r \in \mathbb{R}^n$) ▷引理 5.3

▷ L 是一个仅含有 $O(n\tau)$ 个非零元素的下三角矩阵

▷ 我们从不显式地写出 B。 B 可以通过以下方式隐式表示 $\operatorname{diag}(e^u)(L_A L_A^\top) \operatorname{diag}(e^u) + \operatorname{diag}(e^u)(ww^\top) \operatorname{diag}(e^u)$

$X \leftarrow \operatorname{diag}(x)$，其中 $x_i = \min\{ \frac{r_i}{(B\mathbf{1}_n)_i}, 1\}$

$B_0 \leftarrow XB$ ▷ 我们仅隐式地构造 B_0

$Y \leftarrow \operatorname{diag}(y)$，其中 $y_j = \min\{ \frac{r_j}{(B_0^\top \mathbf{1}_n)_j}, 1\}$

$B_1 \leftarrow B_0 Y$ ▷ 我们仅隐式地构造 B_1

$p \leftarrow r - B_1 \mathbf{1}_n$

返回 p、X、Y、w、u ▷ 我们以隐式方式返回 G，即 $G := B_1 + pp^\top / \|p\|_1$

结束过程

时间复杂度为 $O(n\tau)$。

证明。 与引理 4.3 和引理 4.4的证明类似，此处两个分布相同，即 $c = r$。

□

总而言之，我们给出了用于对称情形下近似最优传输的算法（算法 6）的运行时间。

定理 5.4。 存在一个算法（算法 6），它以成本矩阵 $C = MM^\top = \in \mathbb{R}^{n \times n}$、一个$n$维单纯形 r 作为输入，并输出

- 一个对角矩阵 $X \in \mathbb{R}^{n \times n}$
- 一个对角矩阵 $Y \in \mathbb{R}^{n \times n}$
- 一个下三角矩阵 L_A
- 向量 $u, w, p \in \mathbb{R}^n$

使得 $X \in \mathcal{U}(r)$ 可以通过以下方式（隐式地）构造

$$\widehat{X} = X(\operatorname{diag}(e^u) L_A L_A^\top \operatorname{diag}(e^u) + \operatorname{diag}(e^u)(ww^\top) \operatorname{diag}(e^u))Y + pp^\top / \|p\|_1$$

满足式(2) 中的

$$O(n\tau^2 + \epsilon^{-2} n\tau \|C\|_\infty^2 \ln n)$$

时间。

备注 5.5。如果我们不在意输出格式是否为下三角矩阵，那么可以移除加法项 $n\tau^2$ 。

证明。 根据定理 5.2，我们得到第 7 行的运行时间为 $O(n\tau \cdot T)$，其中 T 是 Sinkhorn 算法的迭代总次数。根据引理 5.3，第 9行的运行时间为 $O(n\tau$。证明的其余部分与定理 4.5的证明类似。

□

参考文献

[ACB17] Martin Arjovsky、Soumith Chintala 和 L´eon Bottou。Wasserstein生成对抗网络。载于《国际机器学习会议》，第214–223页。PMLR，2017年。[ANWR17] Jason Altschuler、Jonathan Niles-Weed 和 Philippe Rigollet。通过 Sinkhorn 迭代实现最优传输的近线性时间近似算法。载于《神经信息处理系统进展》，第30卷，2017年。[AZLOW17] Zeyuan Allen-Zhu、Yuanzhi Li、Rafael Oliveira 和 Avi Wigderson。更快速的矩阵缩放算法。载于2017年IEEE第58届计算机科学基础年度研讨会（FOCS），第890–901页。IEEE，2017年。[BBR06] Federico Bassetti、Antonella Bodini 和 Eugenio Regazzini。关于最小Kantorovich距离估计量。《统计与概率快报》，76(12):1298–1302，2006年。[BGHK95] Hans L Bodlaender、John R Gilbert、Hj´almtyr Hafsteinsson 和 Ton Kloks。树宽、路径宽度、前沿大小和最短消去树的近似。《算法杂志》，18(2):238–255，1995年。[BGKL17] J´er´emie Bigot、Ra´ul Gouet、Thierry Klein 和 Alfredo L´opez。通过凸主成分分析，在Wasserstein空间中进行测地线主成分分析。载于《亨利·庞加莱研究所概率与统计年刊》，第53卷，第1–26页。亨利·庞加莱研究所，2017年。[BJKS18] Jose Blanchet、Arun Jambulapati、Carson Kent 和 Aaron Sidford。迈向最优传输的最优运行时间。arXiv预印本arXiv:1810.07717，2018年。[Bod94] Hans L Bodlaender。树宽导览。《控制论学报》，11(1-2):1，1994年。[BPC16] Nicolas Bonneel、Gabriel Peyr´e 和 Marco Cuturi。Wasserstein重心坐标：利用最优传输进行直方图回归。《ACM图形学汇刊》，35(4):71–1，2016年。[CMTV17] Michael B Cohen、Aleksander Madry、Dimitris Tsipras 和 Adrian Vladu。通过盒约束牛顿法和内点法实现矩阵缩放与平衡。载于2017年IEEE第58届计算机科学基础年度研讨会（FOCS），第902–913页。IEEE，2017年。[Cut13] Marco Cuturi。Sinkhorn距离：以闪电般的速度计算最优传输。《神经信息处理系统进展》，第26卷，2013年。[Dav06] Timothy A Davis。稀疏线性系统的直接法。SIAM，2006年。[DGK18]Pavel Dvurechensky、Alexander Gasnikov 和 Alexey Kroshnin。计算最优传输：加速梯度下降的复杂度优于 Sinkhorn 算法。载于《国际机器学习会议》（ICML），第1367–1376页。PMLR，2018年。[DLY21] Sally Dong、Yin Tat Lee 和 Guanghao Ye。一种用于小树宽线性规划的近线性时间算法：鲁棒中心路径的多尺度表示。载于 STOC，2021年。

[ESS17] Johannes Ebert、Vladimir Spokoiny 和 Alexandra Suvorikova。《在 2-Wasserstein 空间中构造非渐近置信集》。arXiv 预印本 arXiv:1703.03658，2017 年。[FLS+18] Fedor V Fomin、Daniel Lokshtanov、Saket Saurabh、Micha l Pilipczyk 和 Marcin Wrochna。《针对低树宽图与矩阵的全多项式时间参数化计算》。ACM算法汇刊（TALG），14(3):1–45，2018 年。[FZM+15] Charlie Frogner、Chiyuan Zhang、Hossein Mobahi、Mauricio Araya 和 Tomaso A Poggio。《使用Wasserstein损失进行学习》。神经信息处理系统进展，28，2015 年。[GLN94] Alan George、Joseph Liu 和 Esmond Ng。《稀疏线性系统的计算机求解》。橡树岭国家实验室，1994 年。[GPC18] Aude Genevay、Gabriel Peyr´e 和 Marco Cuturi。《使用Sinkhorn散度学习生成模型》。载于《人工智能与统计国际会议》，第 1608–1617 页。PMLR，2018 年。[GS22] Yuzhou Gu 和 Zhao Song。《一种更快的小树宽半正定规划求解器》。arXiv 预印本 arXiv:2211.06033，2022 年。[Kan42] Leonid V Kantorovich。《论质量搬运》。载于 Dokl. Akad. Nauk. USSR (NS)，第 37 卷，第 199–201 页，1942 年。[KPT+17] Soheil Kolouri、Se Rim Park、Matthew Thorpe、Dejan Slepcev 和 Gustavo K Rohde。《最优质量输运：信号处理与机器学习应用》。IEEE信号处理杂志，34(4):43–59，2017 年。[KSKW15] Matt Kusner、Yu Sun、Nicholas Kolkin 和 Kilian Weinberger。《从词嵌入到文档距离》。载于《国际机器学习会议》，第 957–966 页。PMLR，2015 年。[LSZ+22] Sixue Liu、Zhao Song、Hengjie Zhang、Lichen Zhang 和 Tianyi Zhou。《空间高效内点法及其在线性规划和最大权二分匹配中的应用》。arXiv 预印本 arXiv:2009.06106，2022 年。[MMC16] Gr´egoire Montavon、Klaus-Robert M¨uller 和 Marco Cuturi。《受限玻尔兹曼机的Wasserstein训练》。神经信息处理系统进展，29，2016 年。[PW09] Ofir Pele 和 Michael Werman。《快速且鲁棒的地球移动者距离》。载于 2009 IEEE第12届国际计算机视觉会议，第 460–467 页。IEEE，2009 年。[PZ16] Victor M Panaretos 和 Yoav Zemel。《点过程的振幅与相位变化》。统计年刊，44(2):771–812，2016 年。[Qua18]Kent Quanrud。《用线性规划近似最优传输》。arXiv 预印本 arXiv:1810.05957，2018 年。[RS10] Neil Robertson 和 Paul Seymour。《图子式 XXIII：Nash-Williams浸入猜想》。组合理论杂志B辑，100(2):181–205，2010 年。

[RTG00] Yossi Rubner、Carlo Tomasi 和 Leonidas J Guibas。《地球移动者距离作为图像检索的距离度量》。《国际计算机视觉杂志》，40(2):99–121，2000 年。[SBRL18] Maziar Sanjabi、Jimmy Ba、Meisam Razaviyayn 和 Jason D Lee。《正则化最优传输下生成对抗网络训练的收敛性与鲁棒性》。神经信息处理系统进展，31，2018 年。[SRGB14] Justin Solomon、Raif Rustamov、Leonidas Guibas 和 Adrian Butscher。《用于半监督学习的Wasserstein传播》。载于《国际机器学习会议》，第 306–314 页。PMLR，2014 年。[SYYZ22] Zhao Song、Xin Yang、Yuanyuan Yang 和 Tianyi Zhou。《一种更快的结构化John椭球计算算法》。arXiv 预印本 arXiv:2211.14407，2022 年。[SZRM18] Tim Salimans、Han Zhang、Alec Radford 和 Dimitris Metaxas。《使用最优传输改进生成对抗网络》。arXiv 预印本 arXiv:1803.05573，2018 年。[Vil09] C´edric Villani。《最优传输：旧与新》，第 338 卷。施普林格，2009 年。[WPR85] Michael Werman、Shmuel Peleg 和 Azriel Rosenfeld。《多维直方图的距离度量》。《计算机视觉、图形学与图像处理》，32(3):328–336，1985 年。