

几何感知的去中心化 Sinkhorn 方法用于 Wasserstein 重心

Ali Baheri, David Millard, and Alireza Vahid

Abstract— 分布式系统需要将异构的局部概率分布融合成一个全局摘要，通过稀疏且不可靠的通信网络进行传输。传统的共识算法在欧几里得空间中平均分布，忽略了它们内在的几何结构，导致了误导性的结果。Wasserstein 巴氏中心提供了一种基于最小化最优运输成本的几何感知替代方法，但通过 Sinkhorn 算法的熵近似通常需要集中式协调。本文提出了一种完全分散式的 Sinkhorn 算法，将集中的几何平均重新表述为对数域中的算术平均，从而可以通过局部闲聊协议进行近似。代理与邻居交换对数消息，在共识阶段和本地更新之间交替进行，以模仿没有协调器的集中式迭代。为了优化带宽，我们整合了事件触发传输和 b 位量化，提供了在准确性和通信之间的可调折衷，并适应异步和数据包丢失。在温和假设下，我们证明收敛到集中的熵巴氏中心的一个邻域内，偏差与共识容差、触发阈值和量化误差线性相关。复杂度几乎与网络大小呈线性关系。仿真结果证实了在各种拓扑结构和条件下，使用显著减少的消息实现了接近集中式的准确度。

I. 介绍

分布式系统越来越需要在稀疏、不可靠的链路上将异构局部概率分布融合成网络范围内的汇总 [1], [2]。考虑自主机器人共享关于目标位置的信念，传感器网络聚合环境测量数据，或者边缘设备在不共享原始数据的情况下合作机器学习模型的情况。在每种情况下，各个节点都维护着必须结合起来以获得有意义共识的概率分布。

经典的分布式共识算法将概率分布视为欧几里得空间中的向量，并计算算术平均值 [3]。虽然简单且研究充分，但这种方法从根本上忽略了概率测度的几何结构。简单的平均可能会产生误导的结果：当两个机器人认为目标位于相反位置时，算术平均值会表明它位于中点，这可能是一个不可能的位置，比如墙的中心。同样，在不考虑传输机制的情况下对空间区域中的污染分布进行平均会产生误导决策算法的人工模式。

Ali Baheri and David Millard are with the Mechanical Engineering Department, Alireza Vahid is with the Electrical and Microelectronic Engineering Department, all at Rochester Institute of Technology, Rochester, NY 14623. Emails: akbeme@rit.edu, djm3622@rit.edu, arveme@rit.edu.

核心问题是概率分布存在于一个弯曲的流形上，而不是平坦的欧几里得空间中。将一种分布转换为另一种分布的成本取决于底层几何结构，无论是空间上的（传感器网络）、时间上的（时间序列）还是基于特征的（机器学习）。最优传输提供了一个框架来进行这种感知几何的比较，通过使用一个基础成本函数 [4] 测量分布之间的距离来实现。

Wasserstein 度量下的重心扩展了这一概念，能够在保持几何结构的同时对多个分布进行平均处理 [5]。与欧几里得平均不同，Wasserstein 重心能够维持分布形状，尊重空间或特征关系，并且对异常值具有鲁棒性。这些特性推动了它们在计算机视觉 [6]、计算生物学 [7]、机器学习 [8] 和机器人应用 [9] 中的应用，其中分布的几何结构至关重要。

计算挑战在于高效地计算这些重心。精确计算的代价极高，但熵正则化通过 Sinkhorn 算法 [10] 使实际近似成为可能。这种方法迭代交替执行每个代理上的局部缩放操作和全局几何平均值计算。然而，几何平均值步骤需要在每次迭代中收集所有代理的信息，从而形成一个根本性的集中瓶颈，破坏了分布式计算的目的。

现有的分散式方法存在关键限制。ADMM 和原对偶方法 [11], [12] 交换完整的耦合矩阵或对偶变量，导致通信开销随问题维度的平方增长。Push-sum 变体 [13] 处理有向图但忽略了 Sinkhorn 的乘法结构，并且需要精确算术精度。联邦方法 [14], [15] 假设星形拓扑和中央参数服务器，这与点对点需求相矛盾。没有一种提供明确的机制来权衡准确性与带宽或处理不可靠通信。

我们的核心想法是，几何平均运算可以重新表述为对数空间中的算术平均。这种变换改变了通信模式，使其从需要中央协调转变为通过邻居间的共识实现完全分布式计算。每个代理在对数空间中本地计算其贡献，然后使用标准的八卦算法 [3] 来达成关于平均值的一致意见。为了解决带宽限制问题，我们引入了事

件触发传输 [16]，其中只有当代理的本地状态显著变化时才进行通信，并采用自适应量化 [17]，根据可用带宽降低消息精度。这些机制提供了明确的控制手段，在支持异步和数据包丢失的同时权衡准确性和通信成本。

贡献。本文从三个方面推进了去中心化最优传输的研究。算法上，我们开发了一种完全去中心化的 Sinkhorn 算法，通过将几何平均操作重新表述为对数空间中的分布式算术共识来消除中央协调，保持数值稳定性，而传统方法在此失败。通信方面，我们设计了事件触发和量化协议，这些协议在减少带宽几个数量级的同时保留了收敛保证，并提供了明确的参数用于调整精度与通信之间的权衡。理论上，我们建立了与集中版本相同的解决方案的收敛性，并明确了共识容差、触发阈值和量化精度如何影响精度的具体界限。

论文组织。第二节回顾了相关工作。第三节介绍了预备知识和集中式熵中心点。第四节开发了一种基于事件触发、量化通信的分散式对数-闲谈 Sinkhorn 算法。第五节建立了收敛性和复杂度界限。第六节报告了关于准确率与带宽权衡及可扩展性的实验结果。第七节总结了局限性并指出了未来的研究方向。

II. 相关工作

A. 最优传输和 Wasserstein 重心

最优传输是一种用于比较具有深层次几何结构的概率测度的框架。维拉尼的工作涵盖了理论基础 [4]，而佩雷和库图里则概述了计算方面的内容 [8]。Wasserstein barycenters 由 Agueh 和 Carlier 正式提出 [5]，将 Fréchet 均值扩展到概率测度的空间。最优传输在机器学习 [8], [18]、计算机视觉 [6]、自然语言处理 [19]、强化学习 [20]–[23]、计算生物学 [7] 以及经济学 [24] 等领域得到了广泛应用。

B. 集中式重心算法

熵正则化 [10] 允许使用 Sinkhorn 算法及其 Bregman 投影形式进行实际重心计算 [25]。快速变体 [26] 在集中式设置中提供了强大的精度-复杂度权衡。高级技术包括对数域稳定化以提高数值鲁棒性 [27]，几何域的卷积核 [6]，高维的切片变体 [28]，以及随机形式化 [29]。这些方法假设了集中式协调或星形拓扑并行性，在每次迭代中通信完整的直方图或对偶势。

C. 分布式优化与通信效率

经典分布式共识 [30] 和随机闲谈 [3] 为去中心化平均提供了基础的收敛保证。事件触发通信 [16] 和量化策略 [31] 被广泛研究以减少分布式优化中的带宽需求，同时保持收敛特性。

D. 去中心化 Wasserstein 重心计算

现有的分散式方法分为三类，各有其局限性。ADMM/原始-对偶方法 [11], [12], [32] 每次迭代中交换完整的耦合矩阵或对偶变量，导致通信开销随着问题维度的增加而呈二次增长。这些方法实现了亚线性收敛，并在较小的正则化参数下遭受数值不稳定。Push-sum 变体 [13] 能够处理有向通信图，但忽略了 Sinkhorn 的乘法结构，需要精确算术精度并且无法利用对数域稳定技术。联邦方法 [14], [15] 假设星形拓扑结构并使用中心参数服务器，这与点对点分布式计算的需求相矛盾。

III. 预备知识

A. 问题数据和符号约定

我们考虑 N 个代理，每个代理在共同的支持集 $X = \{x_1, \dots, x_d\}$ 上具有离散概率 $\mu_i \in \Delta_d$ 。令 $C \in \mathbb{R}_{\geq 0}^{d \times d}$ 为地面成本，并且对于 $\varepsilon > 0$ ，定义严格正的吉布斯核

$$K = \exp(-C/\varepsilon),$$

其元素为 $\exp(\cdot)$ 。向量是列； $\odot$ 和 $\oslash$ 表示元素-wise 乘法/除法， $\mathbf{1}$ 是全 1，而 $\text{softmax}(z) = \exp(z)/\langle \mathbf{1}, \exp(z) \rangle$ 。我们使用 $\|\cdot\|_1$ 和 $\|\cdot\|_\infty$ 。表 I 概括了本工作中使用的全部符号和参数。

B. 集中熵重心（参考）

令 b_ε^* 表示在 X 上的 $\{\mu_i\}_{i=1}^N$ 的均匀权重熵 Wasserstein 重心。标准 IBP/Sinkhorn 重心保持每个代理的比例 $u_i \in \mathbb{R}_{>0}^d$ 和一个共享比例 $v \in \mathbb{R}_{>0}^d$ 。给定 v ，每个代理使用一个小的 $\eta > 0$ 更新

$$u_i \leftarrow \mu_i \oslash (K v + \eta), \quad i = 1, \dots, N, \quad (1)$$

以保持数值稳定性。共享更新是 $\{K^\top u_i\}$ 的几何平均值：

$$v \leftarrow \left(\prod_{i=1}^N K^\top u_i \right)^{1/N}. \quad (2)$$

原始迭代是 $b = v/\langle \mathbf{1}, v \rangle = \text{softmax}(\log v)$ 。

TABLE I: 符号和参数。

符号	描述
N	Number of agents
d	Support dimension
$C \in \mathbb{R}_{\geq 0}^{d \times d}$	Ground cost matrix
$\varepsilon > 0$	Entropic regularization
$K = \exp(-C/\varepsilon)$	Gibbs kernel
$\eta > 0$	Ridge parameter
$u_i \in \mathbb{R}_{>0}^d$	Agent i scaling vector
$v \in \mathbb{R}_{>0}^d$	Shared scaling; $b = v/\langle \mathbf{1}, v \rangle$
$s_i = \log(K^\top u_i)$	Agent i log-message
z_i	Local estimate of $\frac{1}{N} \sum_j s_j$
δ	Event-trigger threshold
$\tau_{\text{in}}, \tau_{\text{out}}$	Inner/outer tolerances
$[s_{\min}, s_{\max}]$	Message clipping range
$v_{\min}, v_{\max}$	$e^{s_{\min}}, e^{s_{\max}}$
Δ_q	Quantization error: $(s_{\max} - s_{\min})/2(2^b - 1)$
L_{exp}	Lipschitz const. of exp: $e^{s_{\max}}$
L_{norm}	Norm. map Lipschitz: $\leq 2/v_{\min}$
$\rho \in (0, 1)$	Contraction factor

C. 对数域消息和平均恒等式

定义每个代理的日志消息

$$s_i := \log(K^\top u_i) \in \mathbb{R}^d. \quad (3)$$

那么 (2) 等价于算术平均值:

$$\log v = \frac{1}{N} \sum_{i=1}^N s_i \iff v = \exp\left(\frac{1}{N} \sum_{i=1}^N s_i\right).$$

这个恒等式是我们通过局部平均实现去中心化的方式。

IV. 方法论

我们开发了一个完全去中心化的 Sinkhorn 重心迭代变体, 通过以下步骤: (i) 将共享的几何平均更新重写为对数域中的算术平均, (ii) 使用短暂的邻域八卦轮次来近似该平均值, (iii) 通过事件触发的量化传输减少带宽。在整个过程中, 向量均为列; $\odot$ 和 $\oslash$ 表示逐元素乘除; $\mathbf{1}$ 是全 1 向量; $\text{softmax}(z) := \exp(z)/\langle \mathbf{1}, \exp(z) \rangle$ 。

令 $X = \{x_1, \dots, x_d\}$ 是公共支撑集, $C \in \mathbb{R}_{\geq 0}^{d \times d}$ 是基础成本。对于 $\varepsilon > 0$ 定义 Gibbs 核函数 $K := \exp(-C/\varepsilon)$ (逐元素指数)。每个代理 $i \in \{1, \dots, N\}$ 持有一个直方图 $\mu_i \in \Delta^d$ 。熵中心 IBP/Sinkhorn 迭代保持每个代理的缩放因子 $u_i \in \mathbb{R}_{>0}^d$ 和一个共享的缩放因子 $v \in \mathbb{R}_{>0}^d$ 。给定一个小的脊线 $\eta > 0$ 以确保数值

稳定性, 更新为

$$u_i \leftarrow \mu_i \odot (Kv + \eta \mathbf{1}), \quad i = 1, \dots, N, \quad (4)$$

$$v \leftarrow \left(\prod_{i=1}^N K^\top u_i \right)^{1/N}. \quad (5)$$

定义每个代理的日志消息

$$s_i := \log(K^\top u_i) \in \mathbb{R}^d. \quad (6)$$

然后共享更新 (5) 相当于对数域中的算术平均值:

$$\log v = \frac{1}{N} \sum_{i=1}^N s_i \iff v = \exp\left(\frac{1}{N} \sum_{i=1}^N s_i\right). \quad (7)$$

令 $G = (V, E)$ 是一个连接的无向通信图, 具有 $|V| = N$ 和邻居集合 $\mathcal{N}_i$ 。为了避免在 (7) 中出现集中聚合, 每个节点 i 维护一个局部估计器 $z_i \approx \frac{1}{N} \sum_j s_j$ 并使用最近的邻居数据包 $\tilde{z}_k$ 运行加权闲谈:

$$z_i^{(0)} \leftarrow s_i, \quad (8)$$

$$z_i^{(s+1)} \leftarrow \sum_{k \in \mathcal{N}_i \cup \{i\}} w_{ik}^{(s)} \tilde{z}_k^{(s)}, \quad s = 0, 1, \dots \quad (9)$$

采用 (可能随时间变化的) 双随机权重 $w_{ik}^{(s)}$ 。短暂内部共识后 (参见 §IV), 每个节点形成一个共享的迭代

$$v \leftarrow \exp(z_i), \quad b \leftarrow \text{softmax}(\log v). \quad (10)$$

交错 (4), (6), (9) 和 (10) 再现了没有协调器的集中映射。

为了最小化通信成本, 节点 i 只在日志消息有显著变化时才进行传输。设 $s_{i, \text{last}}$ 为最后传输的 s_i 。在第 t 次迭代中, 仅当

$$\|s_i - s_{i, \text{last}}\|_\infty > \delta, \quad (11)$$

满足条件时才传输, 否则邻居节点重用缓存的数据包。我们传输一个 b 位的量化和裁剪日志消息。令 $[s_{\min}, s_{\max}]$ 限制传输条目 (本地保持全精度), 并在该区间上定义一个统一的 b 位量化器。每项最差情况下的量化误差则满足

$$\Delta_q = \frac{s_{\max} - s_{\min}}{2(2^b - 1)}. \quad (12)$$

剪辑将仅应用于打包的日志值; 所有本地计算均使用未剪辑的 s_i 。

我们在对数域中计算, 将 $\eta > 0$ 保留在 (4) 中, 并通过 $b = \text{softmax}(\log v)$ 形成原始迭代。对于内部共识, 节点 i 在

$$\max_{k \in \mathcal{N}_i} \|z_i - \tilde{z}_k\|_\infty < \tau, \quad (13)$$

Algorithm 1 分散的日志-闲谈 Sinkhorn 算法，带有事件触发和量化

Require: 图 G , 核 K , 容差 τ , 触发器 δ , 量化器 $\mathcal{Q}_b(\cdot)$, 裁剪范围 $[s_{\min}, s_{\max}]$, 岭 $\eta > 0$.

- 1: **for** each node $i \in V$ **in parallel** **do**
- 2: $u_i \leftarrow \mathbf{1}$; $s_i \leftarrow \log(K^\top u_i)$; $z_i \leftarrow s_i$;
 $s_{i,\text{last}} \leftarrow s_i$
- 3: **while** not converged **do**
- 4: **for** each node i **in parallel** **do** $\triangleright$ Local scaling
- 5: $v \leftarrow \exp(z_i)$; $u_i \leftarrow \mu_i \odot (Kv + \eta \mathbf{1})$
- 6: $s_i \leftarrow \log(K^\top u_i)$
- 7: **if** $\|s_i - s_{i,\text{last}}\|_\infty > \delta$ **then** $\triangleright$ Event trigger
- 8: 发送 $\mathcal{Q}_b(\text{clip}(s_i; [s_{\min}, s_{\max}]))$ 到 $\mathcal{N}_i$; $s_{i,\text{last}} \leftarrow s_i$
- 9: **(内部八卦)** 每个节点 i 使用 (9) 和最近的邻居数据包更新 z_i 直到达到 $\max_{k \in \mathcal{N}_i} \|z_i - \tilde{z}_k\|_\infty < \tau$ 或步骤限制。
- 10: **for** each node i **in parallel** **do** $\triangleright$ Shared projection
- 11: $v \leftarrow \exp(z_i)$; $b \leftarrow \text{softmax}(\log v)$
- 12: **return** b

或在固定步数限制之后声明局部一致。对于外部循环，当连续共享迭代稳定时，例如

$$\|\log v^{(t+1)} - \log v^{(t)}\|_\infty < \tau. \quad (14)$$

时，我们停止。

A. 去中心化 Sinkhorn: 完整过程

V. 理论结果

我们分析了去中心化、事件触发的量化对数八卦 Sinkhorn 方案。我们的目标是：(i) 在 Hilbert 度量中为集中式 IBP/Sinkhorn 重心映射建立单一的一致收缩；(ii) 通过谱隙量化内循环共识时间；(iii) 使用明确的范围感知常数，将共识/触发/量化扰动传播到 $\exp + \text{normalize}$ 管道中的 ℓ_1 ；以及 (iv) 声明具有一致的 d 依赖和异步处理能力的通信/计算复杂度。

A. 假设、稳定化和范围管理

假设 1 (网络与权重)。通信图 $G = (V, E)$ 是连通的。平均权重 $W = [w_{ik}]$ 是双随机的，对于 $(i, k) \in$

$E \cup \{(i, i)\}$ 有 $w_{ik} \geq \beta > 0$ 。对于时变/异步更新我们在 §V-H 中指定假设。

假设 2 (问题的正则性与稳定化)。Gibbs 核心 $K = \exp(-C/\varepsilon)$ 的所有元素严格为正数对于 $\varepsilon > 0$ 成立。每个代理 i 持有一个直方图 $\mu_i \in \Delta^d$ 。我们运行带有正的脊 $\eta > 0$ 的稳定化 IBP 更新，在 u 更新中（参见 (1)），使得 $K^\top u_i + \eta \mathbf{1} > 0$ 总是成立。这意味着每个代理的日志消息

$$s_i := \log(K^\top u_i)$$

在每次迭代中都是明确定义的，即使 μ_i 是稀疏的（包含零）。

假设 3 (通过裁剪实现有限对数范围)。存在有限的边界 $s_{\min} \leq s_{\max}$ ，使得传输日志消息在量化之前满足 $s_i^{\text{tx}} = \text{clip}(s_i; [s_{\min}, s_{\max}]) \in [s_{\min}, s_{\max}]^d$ 。定义 $v_{\min} := e^{s_{\min}}$, $v_{\max} := e^{s_{\max}}$ 。我们将 s_i^{tx} 分量用一个在 $[s_{\min}, s_{\max}]$ 上具有 b 位的均匀量化器进行量化，得到每项误差为 $\Delta_q = (s_{\max} - s_{\min})/2(2^b - 1)$ 。接收方缓存最新数据包并在新数据包到达前重复使用它。

备注 1 (为何剪枝是算法的一部分): 裁剪使分析自洽并消除了在小 ε 或峰值直方图时的病态范围爆炸。它被应用于仅针对传达的日志消息；局部计算保持全精度。如果使用自适应范围，节点可以捎带当前 $(s_{\min}, s_{\max})$ 并偶尔进行偏移/缩放更新。

B. 巴里中心映射在希尔伯特度量中的收缩

令 $F : \Delta^{d-1} \rightarrow \Delta^{d-1}$ 表示一个完整的心重循环：所有具有共同 v 的 u_i 更新，接着是共享的 v 更新和归一化 $b = v / \langle \mathbf{1}, v \rangle$ 。我们使用希尔伯特射影度量 $d_H(x, y) = \log(\max_j x_j / y_j) - \log(\min_j x_j / y_j)$ 在 $\mathbb{R}_{>0}^d$ 上。

引理 1 (希尔伯特压缩 (集中式)): 设 $\text{osc}(\log K) = \max_{j,j'} \max_{\ell} (\log K_{\ell j} - \log K_{\ell j'})$ 。定义

$$\theta := \tanh\left(\frac{1}{4} \text{osc}(\log K)\right), \quad \rho := \theta^2. \quad (15)$$

那么

$$d_H(F(b), F(b')) \leq \rho d_H(b, b'). \quad (16)$$

此外，

$$\rho \leq \tanh^2\left(\frac{\|C\|_\infty}{2\varepsilon}\right) < 1, \quad (17)$$

因此 $\rho \uparrow 1$ 作为 $\varepsilon \downarrow 0$ 。

草图。每个 u_i -更新和共享的 v -更新都是由正线性映射诱导, 随后进行投影归一化。经典的 Birkhoff – Bushell 理论给出一个每映射系数 $\leq \theta$ [33]。一个完整的 IBP 重心循环由两个这样的正则化映射组成, 因此收缩因子最多为 θ^2 , 与 N 无关。□

C. 在范围界限下连接 d_H 和 ℓ_1

引理 2 (希尔伯特 $\rightarrow \ell_1$ 桥): 如果 $x, y \in \mathbb{R}_{\geq 0}^d$ 满足 $v_{\min} \leq x_j, y_j \leq v_{\max}$ 和 $p = x/\langle \mathbf{1}, x \rangle$, $q = y/\langle \mathbf{1}, y \rangle$, 则

$$\|p - q\|_1 \leq \frac{2}{v_{\min}} \|x - y\|_1. \quad (18)$$

此外, 如果 $d_H(x, y) \leq \gamma$, 则

$$\|p - q\|_1 \leq \frac{2}{v_{\min}} (e^\gamma - 1). \quad (19)$$

D. 一步扰动通过 exp 和归一化

设 $\tilde{F}$ 为 F 的去中心化对应物, 它使用具有容忍度 τ 、事件触发阈值 δ 和在传输对数消息上的 b 位量化, 步长为 Δ_q 的内环共识。

引理 3 (一步扰动): 设 $L_{\text{exp}} := e^{s_{\max}}$ 和

$$L_{\text{norm}} := \sup_{x > 0} \left\| \frac{x}{\langle \mathbf{1}, x \rangle} - \frac{\cdot}{\langle \mathbf{1}, \cdot \rangle} \right\|_{1 \leftarrow 1} \quad (\text{operator } \ell_1 \rightarrow \ell_1 \text{ norm}).$$

然后对于任意的 $b \in \Delta^{d-1}$,

$$\|\tilde{F}(b) - F(b)\|_1 \leq L_{\text{exp}} L_{\text{norm}} (\tau + \delta + \Delta_q).$$

在假设 3 (有限对数范围) 下, $L_{\text{norm}} \leq 2/v_{\min}$, 因此

$$\|\tilde{F}(b) - F(b)\|_1 \leq \frac{2e^{s_{\max}}}{v_{\min}} (\tau + \delta + \Delta_q).$$

讨论。 $(\tau + \delta + \Delta_q)$ 项是每个节点使用的预 exp 对数向量在每维上的最坏情况下的 ℓ_∞ 界。不假设误差源之间的独立性; 我们简单地添加幅度以涵盖缓存中的相关性和陈旧性。

E. 跟踪集中重心和 ε -灵敏度

令 b_ε^* 为集中熵重心不动点。

定理 4 (线性收敛到 b_ε^* 的邻域): 在假设条件 1 至 3 以及引理 1 到 3 下, 分散迭代服从

$$\|\tilde{b}^{(t)} - b_\varepsilon^*\|_1 \leq \rho^t \|\tilde{b}^{(0)} - b_\varepsilon^*\|_1 + \frac{L_{\text{exp}} L_{\text{norm}}}{1 - \rho} (\tau + \delta + \Delta_q), \quad (20)$$

其中 ρ 来自引理 1

备注 2: 因为 $\rho \leq \tanh^2(\|C\|_\infty/(2\varepsilon))$, 速率下降为 $\varepsilon \downarrow 0$ 。此外, 当小范围的 ε 扩大时, $L_{\text{exp}}/v_{\min} = e^{s_{\max} - s_{\min}}$ 也会增长。因此, 我们的分析建议要么 (i) 在 ε 中继续, 要么 (ii) 对小范围的 ε 使用更严格的裁剪范围。

F. 内环共识时间

令 $z_i^{(s)}$ 表示节点 i 在 s 步骤后的内部消息状态, 初始化为 $z_i^{(0)} = s_i^{\text{tx}}$ 。对于同步情况 $W^{(s)} \equiv W$:

引理 5 (通过谱间隙达成共识): 如果 $\sigma_2(W)$ 是第二大奇异值, 则

$$\|z^{(s)} - \bar{z} \mathbf{1}\|_2 \leq \sigma_2(W)^s \|z^{(0)} - \bar{z} \mathbf{1}\|_2. \quad (21)$$

因此, 满足容差 τ 所需的内部步骤为

$$T_{\text{inner}}(\tau) = \mathcal{O}\left(\frac{\log(\sqrt{d} \|z^{(0)} - \bar{z} \mathbf{1}\|_2 / \tau)}{1 - \sigma_2(W)}\right). \quad (22)$$

如果 W 是无向图上的 Metropolis (或类似) 权重, 则 $1 - \sigma_2(W) \gtrsim \lambda_2(G)/d_{\max}$ 。

命题 6 (触发率来自变动预算): 在定理 4 下, 存在常数 c_1, c_2 仅依赖于 $(s_{\min}, s_{\max})$ 和问题数据, 使得

$$V_i(T) \leq c_1 \rho^T + c_2 \frac{1 - \rho^T}{1 - \rho} (\tau + \delta + \Delta_q). \quad (23)$$

因此, 节点 i 在 T 外层迭代中广播的次数满足

$$M_i(T) \leq 1 + \left\lceil V_i(T)/\delta \right\rceil. \quad (24)$$

草图。一步收缩贡献了 ρ^t 衰减; 来自引理 3 的稳态偏差产生了几何级数; 触发器将累积变化转换为事件计数。

G. 通信和计算复杂性

令 T_{outer} 表示外部 IBP 循环的数量, T_{inner} 表示每次外部迭代中的内部八卦步骤数量。选择 (τ, δ, Δ_q) 使得

$$\frac{L_{\text{exp}} L_{\text{norm}}}{1 - \rho} (\tau + \delta + \Delta_q) \leq \frac{1}{2} \varepsilon_{\text{tar}}. \quad (25)$$

为了达到 $\|\tilde{b}^{(T)} - b_\varepsilon^*\|_1 \leq \varepsilon_{\text{tar}}$, 只需取

$$T_{\text{outer}} = \mathcal{O}\left(\frac{\log(1/\varepsilon_{\text{tar}})}{\log(1/\rho)}\right), \quad (26)$$

和

$$T_{\text{inner}} = \mathcal{O}\left(\frac{\log(d/\tau)}{1 - \sigma_2(W)}\right). \quad (27)$$

消息计数满足

$$M_{\text{total}} = \mathcal{O}(|E| T), \quad M_{\text{agent}} = \mathcal{O}(\deg(G) T), \quad (28)$$

$$T = T_{\text{outer}} T_{\text{inner}}.$$

每个外层迭代和每个代理, 计算 $K^\top u_i$ 的成本为 $\mathcal{O}(d^2)$ (或使用快速变换时为 $\mathcal{O}(d \log d)$), 而每个内层步骤的

序列化成本为 $\mathcal{O}(\deg(G) \cdot d)$ 。写作 $\kappa(d) \in \{d^2, d \log d\}$ ，所有代理的总运行时间是

$$\mathcal{O}(T_{\text{outer}}(N \kappa(d) + T_{\text{inner}} |E| d)). \quad (29)$$

收敛每次迭代呈线性（按次数呈指数），因子为 ρ 。

H. 异步、延迟和过时的数据包

我们现在用标准的随机化/异步模型替换同步内循环。设 $\{W^{(s)}\}$ 是具有 $\mathbb{E}[W^{(s)}] = \bar{W}$ 和谱隙 $1 - \sigma_2(\bar{W}) > 0$ 的独立同分布双重随机平均矩阵。假设通信延迟有界，陈旧程度最多为 Δ 内部步骤和缓存的数据包，并且激活过程独立于量化噪声。那么：

定理 7 (期望中的异步共识): 在上述模型下，

$$\mathbb{E}\|z^{(s)} - \bar{z} \mathbf{1}\|_2 \leq \sigma_2(\bar{W})^s \mathbb{E}\|z^{(0)} - \bar{z} \mathbf{1}\|_2, \quad (30)$$

因此引理 5 中的界值期望中成立，其中 $1 - \sigma_2(W)$ 被替换为 $1 - \sigma_2(\bar{W})$ 。有界滞后仅扩大了有效触发/缓存项，这已经在引理 3 和命题 6 的 (δ, τ) 中考虑到了。

备注（符号和对称性）。当 W 是对称时， $\sigma_2(W) = \lambda_2(W)$ （按模数计算的第二大特征值）。我们始终使用 $\|\cdot\|_{1 \leftarrow 1}$ 表示算子 ℓ_1 的范数。

总结。一个统一的希尔伯特收缩用于中心化（引理 1）、谱隙内部共识（引理 5,7）和范围感知扰动传播（引理 3），共同产生线性跟踪，其显式的稳态邻域缩放与 (τ, δ, Δ_q) 呈线性关系（定理 4），以及消息/运行时复杂度界。

VI. 结果

我们评估了所提出的分散式、事件触发的日志-八卦算法。实验使用了一个 4×4 网格 ($N=16$)，支持大小为 $d=64$ ，熵参数为 ε ，并且内部共识容差为 τ 。对于每个条件，我们对多个种子进行平均并报告 95% 的置信区间。准确性是相对于中心化的熵重心来衡量的。

图 1 在对数尺度上绘制了“始终闲聊”和事件触发变体的内部循环共识残差。两条轨迹都表现出预期的锯齿状轮廓：每次外部迭代内的快速内部收缩，接着在下次外部更新时出现小跳跃，这与光谱间隙控制的共识速率一致（参见闲聊收缩引理）。事件触发适度减慢了内部收敛速度，但达到相同的残差底限，这符合引入有界扰动而不改变收缩状态的理论。图 2 在共同支持上叠加了集中式和分散式重心。在主导模式下，这些曲线视觉上难以区分，表明分散迭代保持

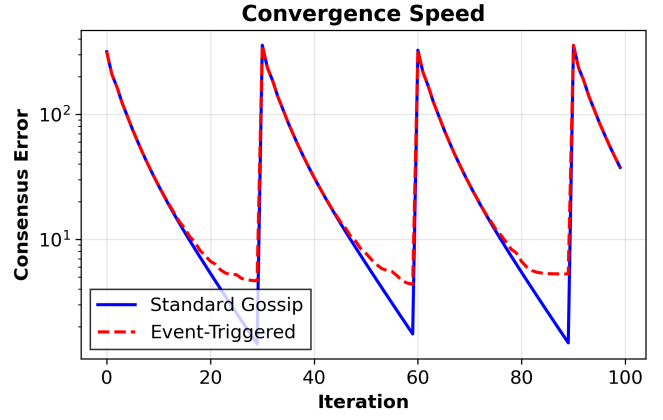


Fig. 1: 收敛轨迹（对数残差）对比始终闲聊与事件触发。

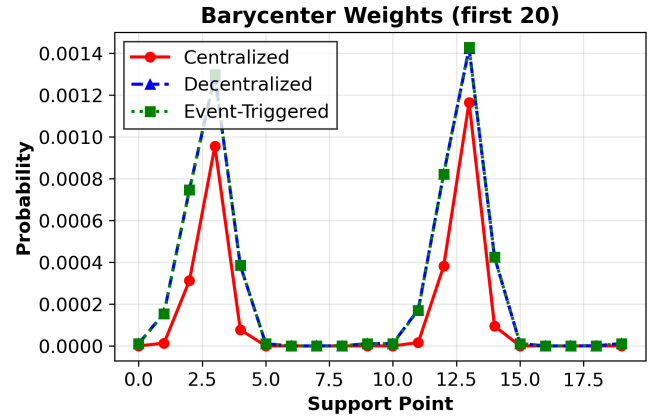


Fig. 2: 集中式与去中心化重心在共同支撑上的重叠。

在集中式解的一个小 ℓ_1 -邻域内。这一观察结果与定理 4 的跟踪界限一致。

图 3 显示了消息总数与网络规模在对数-对数尺度上的关系，并以 $O(N)$ 和 $O(N^2)$ 参考线表示。在网格图中，当 τ 和内部步骤上限固定时，实测斜率接近 $O(N)$ ，这与以下情况相符：(i) 每次外部迭代中的内部消息传递步骤数保持大致恒定；(ii) 消息量随边的数量 $|E| \sim O(N)$ 而变化。图 4 报告了实时时长，随着 N 的增加而单调递增，这与每次迭代的成本分析预期一致。图 5 绘制了精度与支持大小 d 之间的关系。精度随 d 增加然后饱和，符合在固定 ε 下离散化误差和熵偏差之间预期的权衡。

VII. 结论

我们提出了一种几何感知的分散式 Sinkhorn 算法，用于在稀疏、不可靠的网络上计算熵 Wasserstein 重心。通过将中心化的几何平均值重新表述为对数域

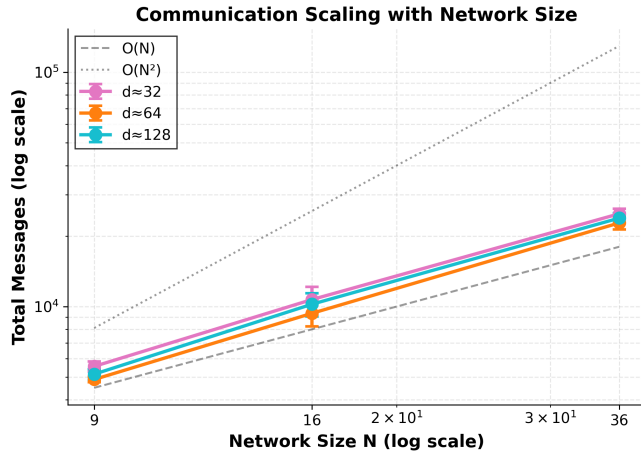


Fig. 3: 总消息数与网络规模（对数-对数）图，带有 $O(N)$ 和 $O(N^2)$ 指南。

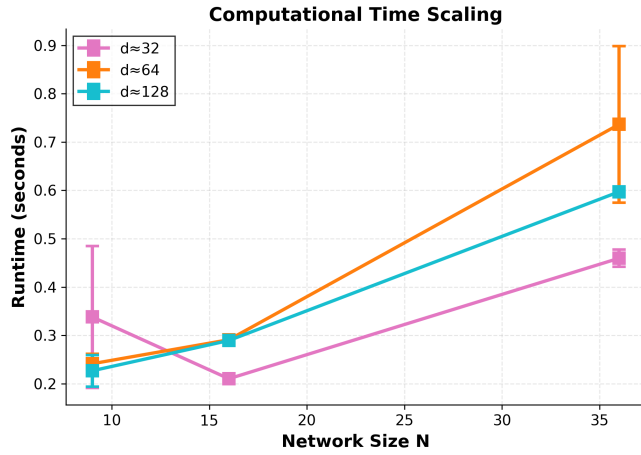


Fig. 4: 运行时与网络规模。

中的算术平均值，并通过局部闲聊共识近似它，我们的方法消除了协调者的需要，同时保持了迭代 Bregman 投影的乘法结构。我们进一步通过事件触发传输和量化增强了通信效率，提供了明确、可调的精度与带宽之间的权衡，能够容忍异步性、数据包丢失和延迟。理论分析建立了收敛到中心化熵重心邻域的结果，偏差界限按共识容差、触发阈值和量化步骤线性扩展。复杂度特征表明在常见拓扑中接近线性的网络规模扩展，并且经验评估确认了在各种图、支撑和压力条件下大幅减少消息量的同时保持中心化级别的精度。尽管我们的方法推进了分散式 OT 求解器的发展，但限制包括对共同支撑和平固定正则化的假设，以及基于模拟的验证仅限于适度规模。未来方向可能探索异构支持、自适应量化方案、与切片或卷积 OT 集成以处理结构化数据，并在分布式学习或传感器融合场景中

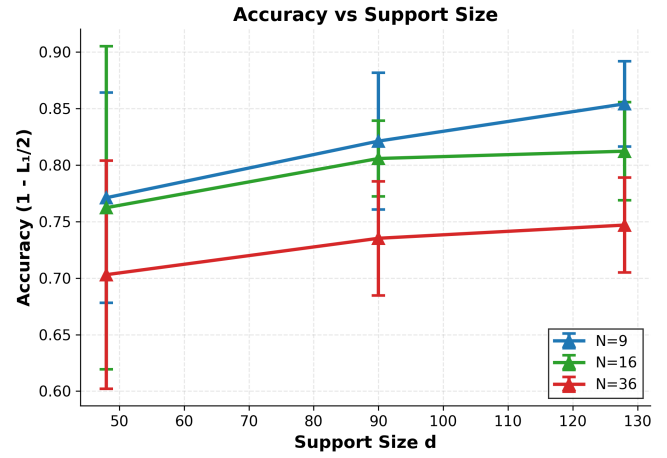


Fig. 5: 准确性与支持规模 d 。

进行实际部署。

REFERENCES

- [1] Reza Olfati-Saber, J. A. Fax, and Richard M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, 2007.
- [2] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguerre y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.
- [3] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2006.
- [4] Cédric Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 2009.
- [5] Martial Agueh and Guillaume Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.
- [6] Justin Solomon, Fernando de Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics*, 34(4):66:1–66:11, 2015. (Proc. SIGGRAPH).
- [7] Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.
- [8] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019.
- [9] Nicolas Papadakis, Edoardo Provenzi, and Vicent Caselles. Optimal transport for diffeomorphic registration. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 212–219, 2014.

- [10] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 26, 2013.
- [11] Xiangfeng Ye, Jian Li, Hongyuan Zhang, and Yinyu Wu. Fast decentralized optimal transport via ADMM. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 4024–4033, 2017.
- [12] Pavel Dvurechensky, Alexander Gasnikov, and Alexey Kroshnin. Decentralized distributed optimization with accelerated ADMM for Wasserstein barycenters. *arXiv preprint arXiv:1806.07396*, 2018.
- [13] Alexander Rogozin, Valentina Bochko, Pavel Dvurechensky, and Alexander Gasnikov. Decentralized optimal transport on general graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 15960–15971, 2021.
- [14] Matthew Staib, Sashank J Reddi, Satyen Kale, Sanjiv Kumar, and Suvrit Sra. Federated learning with Wasserstein barycenters. In *International Conference on Learning Representations (ICLR)*, 2021.
- [15] Ruidi Li and Yin Tat Lee. Federated optimization of Wasserstein distributionally robust models. *arXiv preprint arXiv:2102.05628*, 2021.
- [16] Paulo Tabuada. Event-triggered real-time scheduling of stabilizing control tasks. *IEEE Transactions on Automatic Control*, 52(9):1680–1685, 2007.
- [17] Tuncer Can Aysal and Kenneth E Barner. Compressive wireless sensing. *42nd Annual Conference on Information Sciences and Systems*, pages 134–139, 2008.
- [18] Luis Caicedo Torres, Luiz Manella Pereira, and M Hadi Amini. A survey on optimal transport for machine learning: Theory and applications. *arXiv preprint arXiv:2106.01963*, 2021.
- [19] Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger. From word embeddings to document distances. In *International Conference on Machine Learning*, pages 957–966, 2015.
- [20] Ali Baheri. Wasserstein-barycenter consensus for cooperative multi-agent reinforcement learning. *arXiv preprint arXiv:2506.12497*, 2025.
- [21] Ali Baheri. Understanding reward ambiguity through optimal transport theory in inverse reinforcement learning. *arXiv preprint arXiv:2310.12055*, 2023.
- [22] Ali Baheri, Zahra Sharooei, and Chirayu Salgarkar. WAVE: Wasserstein adaptive value estimation for actor-critic reinforcement learning. *Proceedings of Machine Learning Research* vol, 283:1–12, 2025.
- [23] Ali Baheri. Risk-aware reinforcement learning through optimal transport theory. *arXiv preprint arXiv:2309.06239*, 2023.
- [24] Alfred Galichon. *Optimal transport methods in economics*. Princeton University Press, 2016.
- [25] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.
- [26] Marco Cuturi and Arnaud Doucet. Fast computation of wasserstein barycenters. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, volume 32 of *Proceedings of Machine Learning Research*, pages 685–693, 2014.
- [27] Bernhard Schmitzer. Stabilized sparse scaling algorithms for entropy regularized transport problems. *SIAM Journal on Scientific Computing*, 41(3):A1443–A1481, 2019.
- [28] Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and radon wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015.
- [29] Sebastian Clatici, Edward Chien, and Justin Solomon. Stochastic wasserstein barycenters. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, volume 80 of *Proceedings of Machine Learning Research*, pages 999–1008, 2018.
- [30] Lin Xiao and Stephen Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 53(1):65–78, 2004.
- [31] Anand Kashyap, Tamer Başar, and Rayadurgam Srikant. Quantized consensus. *Automatica*, 43(7):1192–1203, 2007.
- [32] César A Uribe, Soomin Lee, Alexander Gasnikov, and Angelia Nedić. A dual approach for optimal algorithms in distributed optimization over networks. *IEEE Transactions on Automatic Control*, 63(10):3265–3280, 2018.
- [33] Bas Lemmens and Roger Nussbaum. *Nonlinear Perron-Frobenius Theory*, volume 189. Cambridge University Press, 2012.