图数据中 Gromov-Wasserstein 松弛问题的收敛单循环算法

Jiajin Li斯坦福大学
jiajinli@stanford.edu

Jianheng Tang香港科技大学（广州） jtangbf@connect.ust.hk

Lemin Kong香港中文大学
lkong@se.cuhk.edu.hk

Huikang Liu上海财经大学
liuhuikang@sufe.edu.cn

Jia Li香港科技大学（广州）
jialee@ust.hk

Anthony Man-Cho So香港中文大学 manchoso@se.cuhk.edu.hk

Jose Blanchet斯坦福大学
jose.blanchet@stanford.edu

arXiv:2303.06595v1 [cs.CG] 12 Mar 2023

摘要

在这项工作中，我们提出了 Bregman交替投影梯度（BAPG）方法，这是一种能够近似求解 Gromov-Wasserstein（GW）距离的单循环算法。我们引入了一种新颖的松弛技术，在准确率与计算效率之间取得平衡，但在耦合映射的可行性方面有所折衷。我们的分析基于以下观察：GW 问题满足 Luo-Tseng 误差界条件，该条件将点到 GW 问题临界点集的距离与最优性残差的估计联系起来。基于这一观察，我们给出了 BAPG 不动点集与 GW 临界点集之间距离的近似界。此外，在一个温和的技术性假设下，我们证明了 BAPG 收敛到其不动点集。通过在图对齐和划分任务中开展全面的数值实验，我们验证了 BAPG 的有效性；与现有方法相比，它在解的质量和实际运行时间方面均表现更优。

1 引言

GW距离为比较和耦合支撑于不同度量空间上的概率分布提供了一种灵活的方法。这推动了大量文献将GW距离应用于各种结构化数据分析任务，包括二维/三维形状匹配（Peyré 等人，2016；Mémoli & Sapiro，2004；Mémoli，2009）、分子分析（Vayer 等人，2018；2019年a）、图对齐与划分（Chowdhury & Mémoli，2019；Xu等人，2019年b；a；Chowdhury & Needham，2021；Gao 等人，2021）、图嵌入与分类（Vincent-Cuaz 等人，2021年b；Xu等人，2022年）以及生成建模（Bunne 等人，2019；Xu等人，2021）。

尽管GW距离在机器学习和数据科学界获得了广泛关注，但现有的大多数GW距离计算算法都是双循环算法，需要将另一个迭代算法作为子程序，因此并不适合实际应用。最近，Solomon等人（2016年）提出了一种称为eBPG的熵正则化迭代Sinkhorn投影算法，该算法已被证明在Kurdyka-Łojasiewicz框架下能够收敛。然而，eBPG存在一些局限性。首先，它处理的是熵正则化GW目标函数，其正则化参数会对模型性能产生重大影响。其次，它要求在每次迭代中求解一个熵正则化最优传输问题，这既计算成本高，又不具备实际可行性。为了直接求解GW问题，Xu等人（2019b）提出了Bregman投影梯度（BPG），但这仍然是一种双循环算法，

依赖于将另一个迭代算法作为子程序。此外，由于缺乏熵正则化器，它还会受到数值不稳定性的影响。尽管Vayer 等人（2019a）；Mémoli（2011年）引入了Frank-Wolfe 方法来求解GW问题，但他们仍然依赖线性规划求解器和线搜索方案，因此即使对于中等规模的任务也不适用。最近，Xu等人（2019b）基于BPG开发了一种简单的启发式单循环方法BPG-S，并在节点对应任务上展现出良好的实证性能。然而，由于缺乏理论支持，其在存在噪声时的性能尚不明确。

主要挑战在于如何高效处理耦合矩阵的 Birkhoff 多面体约束（即双随机矩阵构成的多面体）。关键问题是，其 Bregman 投影不存在闭式更新，这迫使当前算法依赖计算成本高昂或对超参数敏感的迭代方法。为解决这一困难，我们提出了一种近似求解 GW距离的单循环算法（BAPG）。我们的方法引入了一种新颖的松弛技术，以牺牲耦合映射的一部分可行性为代价，实现计算效率的提升。对于某些图学习任务（如图对齐和划分），这种违反是可以接受的，因为耦合的质量并非首要关注点。我们发现，在一些图学习任务中，BAPG 能够取得理想的性能，因为这些任务的性能指标是匹配准确率，而不是概率对应关系的精确程度。总之，BAPG 提供了一种牺牲可行性、同时提升计算效率和匹配准确率的方法。

在我们的方法中，我们将 Birkhoff多面体约束解耦为针对行和列的独立单纯形约束。随后，采用交替方式，在构造的惩罚函数上执行投影梯度下降。通过利用以相对熵为基函数的单纯形约束闭式 Bregman 投影，BAPG 在每次迭代中仅需进行向量-矩阵/矩阵-矩阵乘法和逐元素矩阵运算，因此成为一种计算效率较高的算法。因此，BAPG 具有多项便利特性，例如兼容 GPU 实现、对步长（唯一的超参数）具有鲁棒性，以及较低的内存需求。

接下来，我们研究 BAPG 的近似界和收敛行为。我们意外地发现，GW 问题满足 Luo-Tseng 误差界条件（Luo& Tseng，1992）。这一事实使我们能够界定 BAPG 的不动点集与 GW 问题临界点集之间的距离，这明显不同于通常利用 Luo-Tseng 误差界条件来建立结构化凸问题线性收敛速率的方法（Zhou & So，2017）。基于这一发现，我们能够明确量化 BAPG 不动点集的近似界。此外，当 Bregman距离的累积非对称误差有界时，我们建立了子序列收敛结果。

最后，我们通过大量实验结果验证 BAPG 在图对齐和图划分方面的有效性。结果表明，在计算效率和匹配准确率方面，BAPG 均优于其他启发式单循环方法和理论上严谨的双循环方法。我们还对 BAPG 进行了敏感性分析，并通过在合成数据和真实世界数据集上的实验展示了其 GPU 加速的优势。所有理论洞见和结果都在实验中得到了充分验证。

2 提出的算法

本节首先将 GW距离表示为一个带有 Birkhoff 多面体约束的非凸二次问题。随后，我们深入探讨所提出算法 BAPG 的理论洞见和计算特性。

Gromov-Wasserstein距离最早由（Mémoli，2011年；2014年；Peyré 等人，2019年）提出，用于量化支撑于不同度量空间上的两个概率测度之间的距离。更准确地说：

定义 2.1（GW距离）。 设给定两个未配准的紧致度量空间（$\mathcal{X}$，d_X）和（$\mathcal{Y}$，d_Y），并分别配备博雷尔概率测度 μ，ν。μ与 ν 之间的*GW*距离定义为

$$\inf_{\pi\in\Pi(\mu,\nu)} \iint |d_X(x,x') - d_Y(y,y')|^2 d\pi(x,y)d\pi(x',y'),$$

其中，$\Pi(\mu,\nu)$ 是定义在 $\mathcal{X}\times\mathcal{Y}$ 上、以 μ 和 ν 为边缘分布的所有概率测度的集合。

直观地说，GW距离旨在通过最优传输保留两个概率测度之间的等距结构。如果存在一个映射将 $x \to y$ 与 $x' \to y'$ 配对，那么 x 与 x' 之间的距离应当与 y 与 y' 之间的距离相近。由于这些理想特性，GW距离是结构化数据分析，尤其是图学习中的一种强大工具。其应用示例包括（Vayer 等人，2019年b；Xu等人，2019年b；a；Solomon等人，2016；Peyré 等人，2016）以及相关参考文献。

首先，为了便于算法开发，我们考虑离散情形，以兼顾简洁性和实用性，其中 μ 和 ν 是两个经验分布，即 $\mu = \sum_{i=1}^{n} \mu_i \delta_{x_i}$ 和 $\nu = \sum_{j=1}^{m} \nu_j \delta_{y_j}$。因此，GW距离可以重新表述如下：

$$\begin{aligned} \min_{\pi \in \mathbb{R}^{n \times m}} \; & -\mathrm{Tr}(D_X \pi D_Y \pi^T) \\ \text{s.t.} \quad & \pi 1_m = \mu,\ \pi^T 1_n = \nu,\ \pi \geq 0, \end{aligned} \tag{1}$$

其中， D_X 和 D_Y 是两个对称距离矩阵。

2.1 GW距离松弛

现在，我们将介绍 GW距离松弛。带有多面体约束的非凸二次规划（1）通常采用（Bregman）投影梯度下降类算法求解。然而，现有算法需要在每次迭代中使用内部迭代算法，例如 Sinkhorn（Cuturi，2013）或半光滑牛顿法（Cuturi，2013），以求解正则化最优传输问题。这可能导致计算开销较大的双重循环方案，不利于 GPU 友好的计算。为解决这一问题，我们旨在采用基于算子分裂的松弛技术，分别处理行约束和列约束。

For 为简洁起见，我们采用紧致形式，以便更好地利用问题特定结构：

$$\min_{\pi} f(\pi) + g_1(\pi) + g_2(\pi). \tag{2}$$

这里， $f(\pi) = -\mathrm{Tr}(D_X \pi D_Y \pi^T)$ 是一个非凸二次函数； $g_1(\pi) = \mathbb{I}_{\{\pi \in C_1\}}$ 以及 $g_2(\pi) = \mathbb{I}_{\{\pi \in C_2\}}$ 是定义在闭凸多面体集合上的两个指示函数。这里， $C_1 = \{\pi \geq 0 : \pi 1_m = \mu\}$和 $C_2 = \{\pi \geq 0 : \pi^T 1_n = \nu\}$。为解耦 Birkhoff 多面体约束，我们采用算子分裂策略，将（2）重写为

$$\min_{\pi = w} f(\pi, w) + g_1(\pi) + g_2(w) \tag{3}$$

其中 $f(\pi, w) = -\mathrm{Tr}(D_X \pi D_Y w^T)$。随后，我们对等式约束进行惩罚，并针对构造的惩罚函数执行交替最小化方案，即

$$F_\rho(\pi, w) = f(\pi, w) + g_1(\pi) + g_2(w) + \rho D_h(\pi, w).$$

这里， $D_h(\cdot, \cdot)$ 是所谓的 Bregman 散度，即 $D_h(x, y) := h(x) - h(y) - \nabla h(y), x - y$，其中 $h(\cdot)$ 是 Legendre 函数，例如 $\frac{1}{2}\|x\|^2$、相对熵 $x \log x$ 等。

2.2 Bregman交替投影梯度（BAPG）

接下来，我们将介绍所提出的单循环 Bregman交替投影梯度（BAPG）方法。BAPG 的核心是在 C_1 和 C_2 之间执行交替投影梯度下降步骤。对于第 k 次迭代，BAPG 的更新形式为

$$\begin{aligned} \pi^{k+1} &= \arg\min_{\pi \in C_1} \left\{ f(\pi, w^k) + \rho D_h(\pi, w^k) \right\}, \\ w^{k+1} &= \arg\min_{w \in C_2} \left\{ f(\pi^{k+1}, w) + \rho D_h(w, \pi^{k+1}) \right\}. \end{aligned} \tag{4}$$

选择相对熵作为 h 还带来了高效计算单纯形约束下 Bregman 投影的优势，例如 C_1 和 C_2，如（Krichene 等人，2015）所述。这些观察使得（4）中 BAPG 的每次迭代都可以进行闭式更新。我们将这一特例称为 KL-BAPG。

KL-BAPG

$$\pi \leftarrow \pi \odot \exp(D_X \pi D_Y / \rho), \quad \pi \leftarrow \mathrm{diag}(\mu./\pi 1_m)\pi, \\ \pi \leftarrow \pi \odot \exp(D_X \pi D_Y / \rho), \quad \pi \leftarrow \pi \mathrm{diag}(\nu./\pi^T 1_n), \tag{5}$$

其中 ρ 是步长，$\odot$ 表示逐元素（Hadamard）矩阵乘法。KL-BAPG 具有多项优势，使其非常适用于中大型图学习任务。首先，它是一种单循环算法，仅需要矩阵-向量/矩阵-矩阵乘法以及逐元素矩阵运算，而这些运算在 GPU 上都得到了高度优化。其次，与 eBPG 中的熵正则化参数不同，KL-BAPG 对步长 ρ的选择不太敏感。第三，对于大小为 nm的矩阵，KL-BAPG 只需要进行一次内存操作；在大规模最优传输问题中，这才是主要瓶颈，而不是浮点运算。（Mai 等人，2021）。

与二次罚函数法（Nocedal &Wright，2006）类似，BAPG 是一种不可行方法，只能在渐近意义上收敛到1的临界点，这意味着如果 ρ 被选为常数，就始终会存在不可行性间隙。尽管如此，对于优先考虑效率和匹配准确率的学习任务，例如图对齐和图划分，BAPG 仍是合适的选择。近期研究也进一步支持了这种以牺牲部分可行性换取其他优势的思路，例如（Vincent-Cuaz et al.，2021a）提出了用于图划分的 GW 距离松弛版本。此外，Séjourné et al.（2021）提出了一种密切相关的边缘松弛，但没有开发出具有收敛保证的高效算法。也就是说，我们令 $\pi = w$，并且 F_ρ（π，π）是（Séjourné et al.，2021）中提出的目标函数。第4.2 节和第4.3 节的实验表明，在图对齐和图划分任务中，KL-BAPG 优于现有基线方法。

3 理论结果

本节将介绍本文所取得的理论结果，包括 BAPG 不动点集的近似界及其收敛性分析。我们分析的基石是方程（1）中 GW 问题所满足的以下正则性条件

命题 3.1（方程（1）的 Luo-Tseng 误差界条件）。 存在标量 $\epsilon > 0$ 和 $\tau > 0$，使得

$$\mathrm{dist}(\pi, \mathcal{X}) \le \tau \left\| \pi - \mathrm{proj}_{C_1 \cap C_2}(\pi + D_X \pi D_Y) \right\|, \tag{6}$$

当 $\|\pi -$投影 $_{C_1 \cap C_2}$ $(\pi + D_X \pi D_Y)\| \le \epsilon$，其中 $\mathcal{X}$是方程 （2）定义的临界点集

$$\mathcal{X} = \{\pi \in C_1 \cap C_2 : 0 \in \nabla f(\pi) + \mathcal{N}_{C_1}(\pi) + \mathcal{N}_{C_2}(\pi)\} \tag{7}$$

并且 $\mathcal{N}_C$（π）表示在 π 处的 C 的法锥。

由于 GW 问题是一个带多面体约束的非凸二次规划，我们可以调用（Luo & Tseng，1992）中的定理2.3，从而得出误差界条件（6）在整个可行集 $C_1 \cap C_2$上成立。命题 3.1 将（6）扩展到了整个空间 $\mathbb{R}^{n \times m}$。该正则性条件旨在通过最优性残差来界定任意耦合矩阵到 GW 问题临界点集的距离，而最优性残差由执行一步投影梯度下降后的差异刻画。事实证明，该误差界条件在明确量化 BAPG 不动点集的近似界方面发挥着重要作用。

3.1 BAPG 不动点集的近似界

首先，我们给出一个将在研究 BAPG 的近似界时使用的关键引理。**引理 3.2**。 设 C_1 和 C_2 为凸多面体集合。存在一个常数 $M > 0$ 使得

$$\left\| \mathrm{proj}_{C_1}(x) + \mathrm{proj}_{C_2}(y) - 2\,\mathrm{proj}_{C_1 \cap C_2}\left(\frac{x+y}{2}\right) \right\| \le M \left\| \mathrm{proj}_{C_1}(x) - \mathrm{proj}_{C_2}(y) \right\|, \quad \forall x \in C_1, y \in C_2.$$

Th 该证明的本质上源于以下观察 f 该不等式可以视为线性二次问题的最优解的稳定性，即h

$$(p(r), q(r)) = \begin{array}{ll} \arg\min\limits_{p,q} & \frac{1}{2}\|x-p\|^2 + \frac{1}{2}\|y-q\|^2 \\ \text{s.t.} & p - q = r,\ p \in C_1, q \in C_2. \end{array}$$

参数 r 确实是扰动量。如果 $r = 0$，则有 $p(0) = q(0) = \text{proj}_{C_1 \cap C_2}(\frac{x+y}{2})$；通过取 $r = \text{proj}_{C_1}(x) - \text{proj}_{C_2}(y)$，不难看出 $(p(r), q(r)) = (\text{proj}_{C_1}(x), \text{proj}_{C_2}(y))$。结合 (Zhang & Luo, 2022) 中的定理 4.1，即可得到所需结果。所有证明细节见附录。

结合引理 3.2 和命题 3.1，不难得到该逼近结果。

命题 3.3（BAPG 不动点集的逼近界）。若点 $(\pi^\star, w^\star)$ 满足以下条件，则其属于 *BAPG* 的不动点集 $\mathcal{X}_{BAPG}$

$$\begin{aligned} \nabla f(w^\star) + \rho(\nabla h(\pi^\star) - \nabla h(w^\star)) + p = 0, \\ \nabla f(\pi^\star) + \rho(\nabla h(w^\star) - \nabla h(\pi^\star)) + q = 0, \end{aligned} \tag{8}$$

其中，$p \in \mathcal{N}_{C_1}(\pi^\star)$ 以及 $q \in \mathcal{N}_{C_2}(w^\star)$。则，不可行性误差满足 $\|\pi^\star - w^\star\| \le \frac{\tau_1}{\rho}$ 并且 $\mathcal{X}_{BAPG}$ 与 $\mathcal{X}$ 之间的间隙满足

$$\text{dist}\left(\frac{\pi^\star + w^\star}{2}, \mathcal{X}\right) \le \frac{\tau_2}{\rho},$$

其中，τ_1 和 τ_2 是两个常数。

注 *3.4.* 若 $\pi^\star = w^\star$，则 X_{BAPG}（（*8*））与 *X* 相同，并且 *BAPG* 可以达到 *GW* 问题（*1*）的一个临界点。命题 *3.3* 表明，随着 $\rho \to +\infty$，不可行性误差项 $\|\pi^\star - w^\star\|$收缩至零，因此 *BAPG* 以渐近方式收敛到（*1*）的一个临界点。此外，当我们将参数 ρ 选为常数时，该命题还明确量化了逼近间隙。证明见附录。τ_1 和 τ_2 的显式形式仅取决于问题本身，包括 $\sigma_{\max}(D_X)$ $\sigma_{\max}(D_Y)$、命题 *3.1* 中 *Luo-Tseng* 误差界条件的常数等。

3.2 BAPG 的收敛性分析

一个自然的后续问题是 BAPG 是否收敛。我们的回答是肯定的。在若干标准假设下，我们证明了 BAPG 的任意极限点都是 X_{BAPG} 中的元素。基于这一目标，我们首先建立势函数 $F_\rho(\cdot)$ 的充分下降性质，

命题 3.5。设 $\{(\pi^k, w^k)\}_{k\ge 0}$ 是由 *BAPG* 生成的序列。假设 $\sum_{k=0}^{\infty} \left(D_h(\pi^{k+1}, w^k) - D_h(w^k, \pi^{k+1})\right)$ 有界。则有

$$F_\rho(\pi^{k+1}, w^{k+1}) - F_\rho(\pi^k, w^k) \le -\rho D_h(\pi^k, \pi^{k+1}) - \rho D_h(w^k, w^{k+1}). \tag{9}$$

由于 $F_\rho(\cdot)$ 是强制的，我们有 $\sum_{k=0}^{\infty} D_h(\pi^k, \pi^{k+1}) + D_h(w^k, w^{k+1}) < +\infty$。$\{D_h(\pi^k, \pi^{k+1})\}_{k\ge 0}$ 和 $\{D_h(w^k, w^{k+1})\}_{k\ge 0}$ 均收敛于零。因此，得到如下收敛性结果。

定理 3.6（BAPG 子序列收敛）。*BAPG* 生成的序列 $\{(\pi^k, w^k)\}_{k\ge 0}$ 的任何极限点都属于 $\mathcal{X}_{BAPG}$。

注 3.7。验证累积非对称误差的有界性是一项具有挑战性的任务，除非 h 是二次函数。为此，我们在二维玩具示例上进行了实证验证，如第 *4.1* 节所述。对于各种步长，该验证的结果见附录 *3* 中的 *C.1*。此外，当 h 是二次函数时，我们可以采用 *Kurdyka-Lojasiewicz* 分析框架，该框架由（*Attouch* 等人，*2010*；*2013*）提出，以证明全局收敛。

据我们所知，交替投影梯度下降方法的收敛性分析仅在凸情形下给出，详见（Wang 与 Bertsekas，2016年；Nedi´c，2011年）以了解详情。在本文中，通过充分利用 GW 问题的误差界条件，我们迈出了第一步，为非凸问题中的交替投影下降方法的分析提供了一条新路径，这一方法本身也可能具有独立的研究价值。

4 实验结果

在本节中，我们通过大量实验结果验证所提出的 KL-BAPG 在各种具有代表性的图学习任务中的有效性，包括图对齐（Tanget al.，2023）和图划分（Li et al.，2021）。所有模拟均使用 Python 3.9，在一台

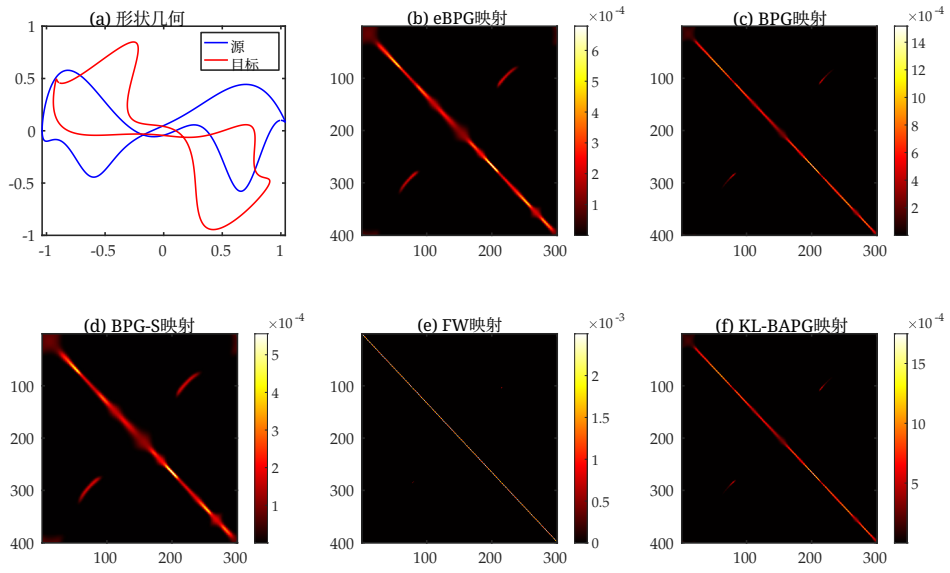


图 1：（a）：源与目标的二维形状几何；（b）–（f）：耦合矩阵的可视化。

运行 Ubuntu 20.04、配备 Intel(R) Xeon(R) Gold 6226R CPU 和 NVIDIA GeForce RTX 3090 GPU 的高性能计算服务器上完成。对于实验部分采用的所有方法，我们使用相对误差 $\|\pi^{k+1} - \pi^k\|_2 / \|\pi^k\|_2 \leq 1e^{-6}$ 和最大迭代次数作为停止准则，即 $\min\{k \in \mathbb{Z} : \|\pi^{k+1} - \pi^k\|_2 / \|\pi^k\|_2 \leq 1e^{-6} \text{ 和 } k \leq 2000\}$。我们的代码可在以下地址获取：https://github.com/squareRoot3/Gromov-Wasserstein-for-Graph。

4.1 二维玩具匹配问题

在本小节中，我们研究二维中的一个玩具匹配问题，以验证我们的理论见解和结果，详见第 2 节和第 3节。图 1 （a）展示了一个示例：将一个没有任何对称性的二维形状映射到其旋转后的同一形状。此处，我们从源形状中采样 300 个点，从目标形状中采样 400 个点，并使用欧氏距离构造距离矩阵 D_X 和 D_Y。

图1 （b）–（f）给出了耦合矩阵的全部颜色图，以直观展示匹配结果。在这里，越稀疏的耦合矩阵表示映射越清晰。所有实验结果均与表 4中的结果一致。我们可以观察到，BPG 和 FW算法都能取得令人满意的求解性能，因为它们旨在精确求解 GW 问题。然而，FW算法和 BPG 会承受显著的计算负担，这一点将在第 4.2 节和4.3节中进一步论证。另一方面，BPG-S 和 eBPG 的性能显然受到不精确问题的损害。尽管 KL-BAPG 的耦合矩阵比 BPG 和 FW算法的耦合矩阵更稠密，但其耦合矩阵的清晰度基本未受到不可行性问题的影响。正如我们稍后将看到的，当惩罚参数 ρ 不太小时，不可行性问题的影响很小，甚至不会给图对齐和划分任务带来任何实际代价，因为这些任务只关注匹配准确率，而不关注耦合的清晰度。

4.2 图对齐

图对齐旨在确定两个图之间的节点对应关系，即使它们可能具有不同的拓扑结构（Zhang等人，2021年；Chen等人，2020年）。不同于求解受限二次分配问题（Lawler，1963年；Lacoste-Julien等人，2006年），GW距离通过保持等距性质提供了最优的概率对应关系。在此，我们将所提出的KL-BAPG与所有现有基线方法进行比较：FW算法（Vayer 等人，2019年a）、BPG（Xu等人，2019年b）、BPG-S（Xu等人，2019年b）（即，BPG与BPG-S之间唯一的区别在于，BPG-S的内部迭代次数仅为一次）、ScalaGW（Xu 等人，2019年a）、SpecGW（Chowdhury & Needham，2021年）和eBPG（Solomon等人，2016年）。除BPG和eBPG外，其他方法都是纯启发式方法，没有任何理论保证。除基于GW的方法外，我们还

表1：图对齐任务上的匹配准确率（%）和实际运行时间（秒）对比。对于 KL-BAPG，我们还报告了 GPU 实现的时间。

方法	合成数据 Acc	合成数据 Time	蛋白质 Raw	蛋白质 含噪数据	蛋白质 Time	酶 Raw	酶 含噪数据	酶 Time	Reddit Raw	Reddit 含噪数据	Reddit Time
IPFP	-	-	43.84	29.89	87.0	40.37	27.39	23.7	-	-	-
RRWM	-	-	71.79	33.92	239.3	60.56	30.51	114.1	-	-	-
SpecMethod	-	-	72.40	22.92	40.5	71.43	21.39	9.6	-	-	-
FW	24.50	8182	29.96	20.24	54.2	32.17	22.80	10.8	21.51	17.17	1121
ScalaGW	17.93	12002	16.37	16.05	372.2	12.72	11.46	213.0	0.54	0.70	1109
SpecGW	13.27	1462	78.11	19.31	**30.7**	79.07	21.14	**6.7**	50.71	19.66	1074
eBPG	34.33	9502	67.48	45.85	208.2	78.25	60.46	499.7	3.76	3.34	1234
BPG	57.56	22600	71.99	52.46	130.4	79.19	62.32	73.1	39.04	36.68	1907
BPG-S	61.48	18587	71.74	52.74	40.4	79.25	62.21	13.4	39.04	36.68	1431
KL-BAPG	**99.79**	9024	**78.18**	**57.16**	59.1	**79.66**	**62.85**	6.9	**50.93**	**49.45**	780
KL-BAPG-GPU	-	**1253**	-	-	75.4	-	-	21.8	-	-	**115**

考虑三种广泛使用的非 GW 图匹配基线算法，包括 IPFP（Leordeanu 等人，2009年）、RRWM（Cho 等人，2010年）以及 SpecMethod（Leordeanu 和 Hebert，2005年）。

参数设置 我们使用无权对称邻接矩阵作为输入距离矩阵，即 D_X 和 D_Y。或者，SpecGW使用热核 $\exp(-L)$ 其中 L 是归一化图拉普拉斯矩阵。我们将 μ 和 ν 都设置为均匀分布。对于三种启发式方法——BPG-S、ScalaGW和SpecGW，我们采用其论文中报告的相同设置。如前所述，eBPG对熵正则化参数非常敏感。为获得可比结果，我们报告正则化参数集合 $\{0.1, 0.01, 0.001\}$ 中的最佳结果。对于BPG和KL-BAPG，我们分别使用常数步长5 和 $\rho = 0.1$。对于FW算法，我们使用PythonOT 软件包中的默认实现（见附录4）。此处报告的所有实验结果均为在不同随机种子下进行5次独立试验的平均值，标准差收集于附录 4。

数据库统计信息 我们在合成数据库和现实世界数据库上测试所有方法。合成数据库的设置与（Xu等人，2019年b）中的设置相同。源图 $G_s = \{V_s, E_s\}$ 由两种理想的随机模型生成，即高斯随机划分模型和 Barabási–Albert 模型，并采用不同的规模，即 $|V_s| \in \{500, 1000, 1500, 2000, 2500\}$。随后，我们首先向源图中添加 $q\%$ 个噪声节点，然后在 V_t 中的节点之间生成 $q\%$ 条噪声边，从而生成目标图 $G_t = \{V_t, E_t\}$，即 $|V_t| = (1+q\%)|V_s|$，$|E_t| = (1+q\%)|E_s|$，其中 $q \in \{0, 10, 20, 30, 40, 50\}$。对于每种设置，我们使用不同的随机种子生成五对合成图。总而言之，合成数据库包含 300 对不同的图。我们还在来自（Chowdhury & Needham，2021年）的另外三个现实世界数据库上验证了所提出的方法，包括两个生物图数据库 Proteins 和 Enzymes，以及一个社交网络数据库 Reddit。此外，为了展示我们的方法在噪声水平方面的鲁棒性，我们遵循合成案例所采用的噪声生成过程（即 $q = 10\%$），在这三个现实世界数据库的基础上创建新的数据库。为此，用于图对齐任务的所有数据库的统计信息已汇总于附录 4 中。我们根据优化后的 $\pi^\star$，将 G_s 中的每个节点与 G_t 中最可能的节点进行匹配。给定预测对应集合 S_{pred} 和真实对应集合 S_{gt}，我们通过 $\text{Acc} = |S_{gt} \cap S_{pred}|/|S_{gt}| \times 100\%$ 计算匹配准确率。

我们方法的结果 表1 展示了四个数据库上的匹配准确率和实际运行时间对比。我们观察到，KL-BAPG 在计算时间和准确率方面都表现出色，尤其是在两个大规模含噪图数据库 合成数据和 *Reddit* 上。值得注意的是，KL-BAPG 具有足够的鲁棒性，因此无需进行参数调优。正如我们在第 2 节所提到的，在 合成数据 和 *Reddit* 上，GPU 加速对 KL-BAPG 有效性的提升也得到了充分验证。由于这两个数据库中的图规模较小，GPU 无法进一步缩短 蛋白质 和*Reddit* 的训练时间。我们将在第 4.4 节给出进一步的实验结果，以证明 KL-BAPG 及其 GPU 加速的鲁棒性。

与其他方法的比较 传统的非GW图对齐方法（IPFP、RRWM和SpecMethod）在节点数超过500的图（例如合成数据和Reddit）上会出现内存不足问题，并且对噪声敏感。eBPG和ScalaGW的性能分别受到

表 2：使用邻接矩阵和热核矩阵在图划分数据集上对 AMI 分数的比较。

类别	方法	Wikipedia 原始值	Wikipedia 含噪数据	EU-email 原始值	EU-email 含噪数据	Amazon 原始值	Amazon 含噪数据	Village 原始值	Village 含噪数据
非 GW	FastGreedy	0.382	0.341	0.312	0.251	0.637	0.573	**0.881**	0.778
	Louvain	0.377	0.329	0.447	0.382	0.622	**0.584**	**0.881**	**0.827**
	Infomap	0.332	0.329	0.374	0.379	**0.940**	0.463	**0.881**	0.190
邻接	FW	0.341	0.323	0.440	0.409	0.374	0.338	0.684	0.539
	eBPG	0.461	**0.413**	**0.517**	0.422	0.429	**0.387**	0.703	0.658
	BPG	0.367	0.333	0.478	0.414	0.412	0.368	0.642	0.575
	BPG-S	0.357	0.285	0.451	0.404	0.443	0.352	0.606	0.560
	KL-BAPG	**0.469**	0.396	0.508	**0.428**	**0.457**	0.362	**0.736**	**0.681**
热核	SpecGW	0.442	**0.395**	0.487	0.425	0.565	0.487	0.758	0.707
	eBPG	0.100	0.082	0.011	0.188	0.604	0.031	0.002	0.003
	BPG	0.418	0.373	0.473	0.253	0.492	0.436	0.705	0.619
	BPG-S	0.411	0.373	0.475	0.253	0.483	0.425	0.642	0.619
	KL-BAPG	**0.533**	0.365	**0.533**	**0.436**	**0.630**	**0.502**	**0.797**	**0.711**

熵正则化参数和近似误差的影响，这解释了它们性能不佳的原因。此外，很容易观察到，SpecGW在小数据集上表现相当不错，但在大数据集上性能急剧下降，例如，合成数据。原因在于，SpecGW依赖线性规划求解器作为其子程序，而该求解器并不适用于大规模场景。此外，尽管ScalaGW具有最低的单次迭代计算复杂度，但（ KXu等人（Xu等人，2019年a）提出的递归划分机制不利于并行计算。因此，ScalaGW在多核处理器上未能展现出具有吸引力的性能。

4.3 图划分

GW距离也可以应用于图划分任务。也就是说，我们试图将源图与一个具有 K 孤立且自连接的超节点的不连通目标图进行匹配，其中 K 表示簇的数量（Abrishami 等，2020年；Li 等，2020年）。类似地，我们在四个真实世界图划分数据集上，将所提出的 KL-BAPG 与第 4.2 节中描述的其他基线方法进行比较。遵循（Chowdhury &Needham，2021），我们还加入了三种专门用于图对齐的非 GW 方法，包括 FastGreedy（Clauset 等，2004）、Louvain（Blondel 等，2008）和 Infomap（Rosvall & Bergstrom，2008）。

参数设置 对于输入距离矩阵 D_X 和 D_Y，我们在邻接矩阵和（Chowdhury &Needham，2020年）提出的热核矩阵上测试我们的方法。对于 KL-BAPG，我们从 $\rho \in \{0.1, 0.05, 0.01\}$ 中选择邻接矩阵对应的收敛函数值最低者，并从 $\rho \in \{0.001, 0.0005, 0.0001\}$ 中选择热核矩阵对应的收敛函数值最低者。通过计算相对于真实划分的调整互信息（AMI）分数（Vinh 等，2010）来量化图划分结果的质量。

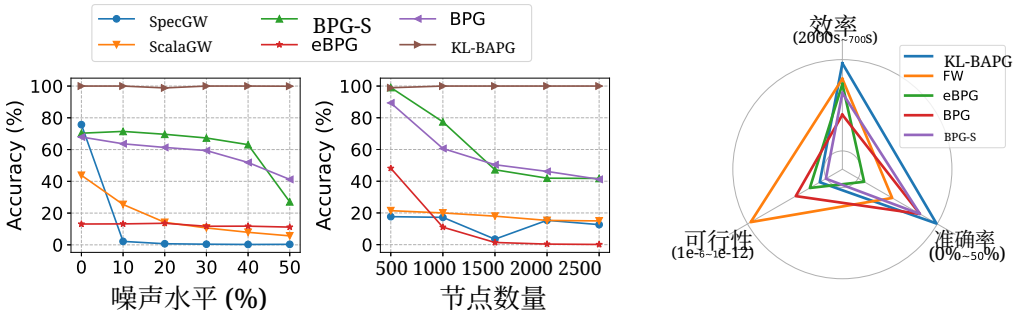
所有方法的结果 表 2 展示了所有方法在图划分任务上的 AMI 分数对比。在大多数情况下，KL-BAPG 优于其他基于GW的方法，并且在噪声条件下更加稳健。具体而言，KL-BAPG 始终优于 FW算法 和 SpecGW，后两者依赖 Frank-Wolfe 方法求解该问题。在使用邻接矩阵时，eBPG 的结果与之相当，但在处理谱矩阵时较为敏感。可能的原因是邻接矩阵和热核矩阵的结构存在显著差异，例如前者是稀疏的，而后者是稠密的。在大多数情况下，BPG 和 BPG-S 的性能相近，但在所有数据集上都不如我们提出的 KL-BAPG。与专门用于非 GW 图划分的图划分方法相比，KL-BAPG 也展现出具有竞争力的性能。例如，在 8 个数据集中，KL-BAPG 分别在其中 6 个和 4 个数据集上优于 Infomap 和 Louvain

4.4 KL-BAPG 的有效性和鲁棒性

首先，我们旨在展示 KL-BAPG 在图对齐上的鲁棒性，因为在数据库（例如图对齐）上测试一种方法的鲁棒性，比在单个点（例如图划分）上进行测试更为合理。

表 3：KL-BAPG、BPG 和 eBPG 在图对齐任务中的 GPU 和 CPU 实际运行时间对比。

	Reddit 数据集			合成数据集		
	KL-BAPG	BPG	eBPG	KL-BAPG	BPG	eBPG
CPU 时间(s)	780	1907	1234	9024	22600	9502
GPU 时间(s)	115	1013	2274	1253	4458	2709
加速比	**6.78**	1.88	0.54	**7.20**	5.07	3.51



(a) (b)

图 2：（a）噪声水平和图规模对合成图对齐数据库的敏感性。（b）Reddit 数据库中效率、准确率和可行性之间权衡的可视化。不可行性误差通过 $\|\pi^T 1_n - \nu\| + \|\pi 1_m - \mu\|$ 计算。“通常而言，越接近外圈边界，表示准确率越高、速度越快，且不可行性误差越低。

噪声水平和图规模首先，我们使用合成数据库，在图 $q\%$ 和图规模 $|\mathcal{V}_s|$ 方面，展示 KL-BAPG 的敏感性分析，结果见图 2 。令人惊讶的是，KL-BAPG 的解性能对噪声水平和图规模均具有鲁棒性。相比之下，随着噪声水平或图规模的增加，其他方法的准确率会显著下降。

效率、准确率与可行性之间的权衡 我们在图 2 (b) 中对 Reddit 数据库上所有基于 GW 的算法在效率、准确率和可行性之间的权衡进行了统一视角的分析。如图所示，我们提出的 KL-BAPG 能够在这三个因素之间实现理想的平衡。表 7对四个数据库进行了详细比较，而表 8 展示了 KL-BAPG 对步长的鲁棒性 ρ。该实验验证了命题 3.3 的有效性，并为选择最优步长提供了实践指导。请注意，较大的 ρ 会带来更低的不可行性误差，但收敛速度会更慢。

KL-BAPG 的 GPU 加速。 我们进行了实验，以进一步证明 KL-BAPG 对 GPU 友好。在表 3中，我们在同一台计算服务器上，使用两个大规模图对齐数据集，对 KL-BAPG、eBPG 和 BPG 的加速比（即 CPU 实际运行时间除以 GPU 实际运行时间）进行了比较。对于 eBPG，我们使用 PythonOT 软件包中的官方实现，该实现支持在 GPU 上运行。对于 BPG，我们使用 Pytorch 自行实现了 GPU 版本。可以看出，与 BPG 和 eBPG 相比，KL-BAPG 在 GPU 上具有高得多的加速比。

5 结束语

在本研究中，我们探索了用于 GW 距离计算松弛的单循环算法的开发。通过利用此前在 GW 文献中尚未研究的误差界条件，我们成功完成了对 BAPG 的收敛性分析。然而，所提出的算法仍面临每次迭代的三次计算复杂度这一挑战，这限制了其在大规模实际问题中的适用性。一个潜在的未来方向是在匹配矩阵中引入稀疏和低秩结构，以降低每次迭代的成本并提升算法性能。此外，由于 Luo-Tseng 误差界条件仍然有效，我们的方法也可以应用于非对称距离矩阵。另一方面，我们的工作还为研究一般非凸问题的交替投影下降方法提供了新的视角，而这在优化领域仍是一个有待探索的方向。

致谢

本文的研究工作得到了美国空军科学研究办公室项目资助，项目编号为 FA9550-20-1-0397。在此基础上，我们还衷心感谢美国国家科学基金会（NSF）项目 1915967 和 2118199 提供的额外支持。Jianheng Tang 和 Jia Li 受到国家自然科学基金委员会（NSFC）项目 62206067 以及广州—香港科技大学（广州）联合资助计划的支持。Anthony Man-Cho So 部分受到香港研究资助局优配研究金项目 CUHK 14203920 的支持。

参考文献

Tara Abrishami、Nestor Guillen、Parker Rule、Zachary Schutzman、JustinSolomon、ThomasWeighill 和 Si Wu。通过最优传输研究图划分的几何。《*SIAM*科学计算杂志》，42(5):A3340–A3366，2020年。

Hédy Attouch、Jérôme Bolte、Patrick Redont 和 Antoine Soubeyran。用于非凸问题的近端交替最小化和投影方法：一种基于 Kurdyka–Łojasiewicz 不等式的方法。《运筹学数学》，35(2):438–457，2010年。

Hedy Attouch、Jérôme Bolte 和 Benar Fux Svaiter。半代数问题和驯良问题的下降方法的收敛性：近端算法、前向–后向分裂和正则化 Gauss–Seidel 方法。*Mathematical Programming*，137(1):91–129，2013年。

Heinz H. Bauschke. 投影算法与单调算子。博士论文，西蒙弗雷泽大学，1996年。

Vincent D Blondel、Jean-Loup Guillaume、Renaud Lambiotte 和 Etienne Lefebvre。大型网络中的社区快速展开。统计力学杂志：理论与实验，2008 (10):P10008，2008年。

Charlotte Bunne、David Alvarez-Melis、Andreas Krause 和 Stefanie Jegelka。学习跨不可比空间的生成模型。发表于国际机器学习会议，第 851–861 页。PMLR，2019年。

Kai Cao、Xiangqi Bai、Yiguang Hong 和 Lin Wan。用于单细胞多组学整合的无监督拓扑对齐。《生物信息学》，36（增刊_1）：i48–i56，2020年。

Kai Cao、Yiguang Hong 和 Lin Wan。使用 Pamona 对异质单细胞多组学数据进行整合的流形对齐。《生物信息学》，38(1)：211–219，2022年。

Xiyuan Chen、Mark Heimann、Fatemeh Vahedian 和 Danai Koutra。通过保持邻近性的节点嵌入实现一致的网络对齐。*arXiv*预印本 *arXiv:2005.04725*，2020年。

Minsu Cho、Jungmin Lee 和 Kyoung Mu Lee。用于图匹配的重加权随机游走。载于欧洲计算机视觉会议，第492–505页。Springer，2010年。

Samir Chowdhury 和 Facundo Mémoli。网络之间的 Gromov–Wasserstein距离与稳定网络不变量。*Information and Inference*：*IMA*期刊，8(4)：757–787，2019年。

Samir Chowdhury 和 Tom Needham。通过 Gromov–Wasserstein 学习实现广义谱聚类。载于国际人工智能与统计会议，第712–720页。PMLR，2021年。

Aaron Clauset、Mark EJ Newman 和 Cristopher Moore。在非常大的网络中寻找社区结构。《网络*alreviewE*》，70(6):066111，2004。

Marco Cuturi。Sinkhorn距离：最优传输的高速计算。《神经信息处理系统进展》，26:2292–2300，2013。

Rémi Flamary、Nicolas Courty、Alexandre Gramfort、Mokhtar Z. Alaya、Aurélie Boisbunon、Stanislas Chambon、Laetitia Chapel、Adrien Corenflos、Kilian Fatras、Nemo Fournier、Léo Gautheron、Nathalie T.H. Gayraud、Hicham Janati、Alain Rakotomamonjy、Ievgen Redko、Antoine Rolet、Antony Schutz、Vivien Seguy、Danica J. Sutherland、Romain Tavenard、Alexander Tong 和 Titouan Vayer。Pot：Python最优传输。机器学习研究杂志，22(78):1–8，2021年。URLhttp://jmlr.org/papers/v22/20-451.html。

Ji Gao、Xiao Huang 和 Jundong Li。基于 Wasserstein距离判别器的无监督图对齐。载于 第*27*届*ACM SIGKDD*知识发现与数据挖掘会议论文集，第426–435页，2021年。Walid Krichene、Syrine Krichene 和 Alexandre Bayen。单纯形上的高效 Bregman 投影。载于 *2015*年第*54*届*IEEE*决策与控制会议（*CDC*），第3291–3298页。IEEE，2015年。Simon Lacoste-Julien、Ben Taskar、Dan Klein 和 Michael Jordan。通过二次分配实现词对齐。2006年。Eugene L Lawler。二次分配问题。管理科学，9(4):586–599，1963年。Marius Leordeanu 和 Martial Hebert。利用成对约束解决对应问题的谱技术。载于 国际计算机视觉会议，第1482–1489页。IEEE，2005年。Marius Leordeanu、Martial Hebert 和 Rahul Sukthankar。用于图匹配和映射推断的整数投影不动点方法。神经信息处理系统进展，22，2009年。Jia Li、Jianwei Yu、Jiajin Li、Honglei Zhang、Kangfei Zhao、Yu Rong、Hong Cheng 和 Junzhou Huang。Dirichlet图变分自编码器。神经信息处理系统进展，33:5274–5283，2020年。Jia Li、Mengzhou Liu、Honglei Zhang、Pengyun Wang、Yong Wen、Lujia Pan 和 Hong Cheng。Mask-GVAE：通过划分对图进行盲去噪。载于 *2021*年网络会议论文集，第3688–3698页，2021年。Zhi-Quan Luo 和 Paul Tseng。仿射变分不等式问题的矩阵分裂算法的误差界与收敛性分析。*SIAM* 优化杂志，2(1):43–54，1992年。Vien V Mai、Jacob Lindbäck 和 Mikael Johansson。用于最优传输的快速准确分裂方法：分析与实现。*arXiv*预印本 *arXiv:2110.11738*，2021年。Facundo Mémoli。用于形状匹配的谱Gromov-Wasserstein距离。载于 *2009*年*IEEE*第*12*届国际计算机视觉会议研讨会，*ICCV* 研讨会，第256–263页。IEEE，2009年。Facundo Mémoli。Gromov–Wasserstein距离与对象匹配的度量方法。计算数学基础，11(4):417–487，2011年。Facundo Mémoli。Gromov–Wasserstein距离：简要概述。公理，3(3):335–341，2014年。Facundo Mémoli 和 Guillermo Sapiro。比较点云。载于 *2004*年*Eurographics/ACM SIGGRAPH* 几何处理研讨会论文集，第32–40页，2004年。Angelia Nedi´c。凸最小化问题的随机算法。*Mathematicalprogramming*，129(2):225–253，2011年。Jorge Nocedal 和 Stephen Wright。数值优化。Springer Science & Business Media，2006年。Gabriel Peyré、Marco Cuturi 和 Justin Solomon。核矩阵与距离矩阵的 Gromov-Wasserstein平均。载于 国际机器学习会议，第2664–2672页。PMLR，2016年。Gabriel Peyré、Marco Cuturi 等。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(5-6):355–607，2019年。Martin Rosvall 和 Carl T Bergstrom。复杂网络上的随机游走图揭示社区结构。美国国家科学院院刊，105(4):1118–1123，2008年。

Thibault Séjourné、François-Xavier Vialard 和 Gabriel Peyré。不平衡 Gromov-Wasserstein 距离：锥体 formulation 与松弛。神经信息处理系统进展，34:8766–8779，2021。Justin Solomon、Gabriel Peyré、Vladimir G Kim 和 Suvrit Sra。对应问题的熵度量对齐。*ACM* 图形学汇刊 *(TOG)*，35(4):1–13，2016。Jianheng Tang、Weiqi Zhang、Jiajin Li、Kangfei Zhao、Fugee Tsung 和 Jia Li。通过联合结构学习与最优传输实现稳健的属性图对齐。*ICDE*，2023。Sarah Teichmann 和 Mirjana Efremova。2019年度方法：单细胞多模态组学。*NatureMethods*，17(1):2020，2020。Titouan Vayer、Laetita Chapel、Rémi Flamary、Romain Tavenard 和 Nicolas Courty。结构化对象的融合 Gromov-Wasserstein 距离：理论基础与数学性质。*arXiv* 预印本 *arXiv:1811.02834*，2018。Titouan Vayer、Nicolas Courty、Romain Tavenard 和 Rémi Flamary。面向结构化数据的最优传输及其在图上的应用。载于 国际机器学习会议，第6275–6284页。PMLR，2019a。Titouan Vayer、Rémi Flamary、Romain Tavenard、Laetitia Chapel 和 Nicolas Courty。切片 Gromov-Wasserstein。载于 *NeurIPS 2019*-第三十三届神经信息处理系统会议，第32卷，2019b。Cédric Vincent-Cuaz、Rémi Flamary、Marco Corneli、Titouan Vayer 和 Nicolas Courty。半松弛 Gromov-Wasserstein 散度及其在图上的应用。*arXiv* 预印本 *arXiv:2110.02753*，2021a。Cédric Vincent-Cuaz、Titouan Vayer、Rémi Flamary、Marco Corneli 和 Nicolas Courty。在线图字典学习。*arXiv* 预印本 *arXiv:2102.06555*，2021b。Nguyen Xuan Vinh、Julien Epps 和 James Bailey。用于聚类比较的信息论度量：变体、性质、归一化与随机校正。机器学习研究杂志，11:2837–2854，2010。Mengdi Wang 和 Dimitri P Bertsekas。带随机约束投影的随机一阶方法。*SIAM* 优化杂志，26(1):681–717，2016。Hongzhi Wen、Jiayuan Ding、Wei Jin、Yiqi Wang、Yuying Xie 和 Jiliang Tang。用于多模态单细胞数据整合的图神经网络。载于 第*28*届*ACM SIGKDD*知识发现与数据挖掘会议论文集，第4153–4163页，2022年。Hongteng Xu、Dixin Luo 和 Lawrence Carin。面向图划分与图匹配的可扩展 Gromov-Wasserstein 学习。神经信息处理系统进展，32:3052–3062，2019a。Hongteng Xu、Dixin Luo、Hongyuan Zha 和 Lawrence Carin。用于图匹配与节点嵌入的 Gromov-Wasserstein 学习。载于 国际机器学习会议，第6932–6941页。PMLR，2019b。Hongteng Xu、Dixin Luo、Lawrence Carin 和 Hongyuan Zha。通过结构化 Gromov-Wasserstein 重心学习图子。载于 *AAAI* 人工智能会议论文集，第35卷，第10505–10513页，2021年。Hongteng Xu、Jiachang Liu、Dixin Luo 和 Lawrence Carin。通过 Gromov-Wasserstein 分解表示图。*IEEE* 模式分析与机器智能汇刊，2022年。Jiawei Zhang 和 Zhi-Quan Luo。全局对偶误差界及其在线性约束非凸优化分析中的应用。*SIAM* 优化杂志，32(3):2319–2346，2022年。

Si Zhang、Hanghang Tong、Long Jin、Yinglong Xia 和 Yunsong Guo。平衡网络对齐中的一致性与差异性。载于 第*27*届*ACM SIGKDD*知识发现与数据挖掘会议论文集，第2212–2222页，2021年。

Zirui Zhou 和 Anthony Man-Cho So。《结构化凸优化问题误差界的统一方法》。*Mathematical Programming*，165(2):689–728，2017。

总体而言，范数 $\|\cdot\|$ 表示 ℓ_2 范数。dist 和 proj 均相对于 ℓ_2 范数定义，即对于任意凸集 C 和点 x,

$$\operatorname{dist}(x, C) = \min_{y \in C} \|x - y\|_2, \tag{10}$$

$$\operatorname{proj}_C(x) = \arg\min_{y \in C} \|x - y\|_2. \tag{11}$$

问题 性质(1)

本文考虑离散情形，其中仅从实分布中抽取 m、n 样本，即 $\mu = \sum_{i=1}^{n} \mu_i \delta_{x_i}$ 和 $\nu = \sum_{j=1}^{m} \nu_j \delta_{y_j}$。$\mu$ 与 ν 之间的 Gromov-Wasserstein距离可改写为以下最小化问题：

$$\begin{aligned} &\min_{\pi \in \Pi(\mu,\nu)} \iint |d_X(x, x') - d_Y(y, y')|^2 d\pi(x, y) d\pi(x', y') \\ \Rightarrow &\min_{\pi \in \Pi(\mu,\nu)} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{m} |d_X(x_i, x_j) - d_Y(y_k, y_l)|^2 \pi_{ik}\pi_{jl}, \end{aligned}$$

其中，$\Pi(\mu,\nu)$ 是广义Birkhoff多面体，即 $\Pi(\mu,\nu) = \{\pi \in \mathbb{R}_+^{n\times m} | \pi 1_m = \mu, \pi^T 1 = \nu\}$。为进一步简化问题并检测隐藏结构，我们将其重新表述为紧致形式，

$$\begin{aligned} &\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{m} (D_X(i,j) - D_Y(k,l))^2 \pi_{ik}\pi_{jl} \\ =&\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{m} (D_X^2(i,j) + D_Y^2(k,l))\pi_{ik}\pi_{jl} - 2\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{m} D_X(i,j) D_Y(k,l) \pi_{ik}\pi_{jl} \\ =&\underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n} D_X^2(i,j)\mu_i\mu_j + \sum_{k=1}^{m}\sum_{l=1}^{m} D_Y^2(k,l)\nu_k\nu_l}_{\text{constant}} - 2\operatorname{Tr}(D_X \pi D_Y \pi^T). \end{aligned}$$

其中 $D_X(i, j) = d_X(x_i, x_j)$ 以及 $D_Y(k, l) = d_Y(y_k, y_l)$ 对于 $i, j \in [n]$ 和 $k, l \in [m]$。因此，得到（1）。

命题 3.1 — GW 问题的 Luo-Tseng 误差界条件（1 ）

证明。 D_X 和 D_Y 是两个对称矩阵，且 $C_1 \cap C_2$ 是一个凸多面体集合。根据 [24], 定理2.3, Luo-Tseng局部误差界条件（6）仅对可行集$C_1 \cap C_2$成立。也就是说，

$$\operatorname{dist}(\pi, \mathcal{X}) \le \tau \left\| \pi - \operatorname{proj}_{C_1 \cap C_2}(\pi + D_X \pi D_Y) \right\|, \tag{12}$$

其中 $\pi \in C_1 \cap C_2$。接下来，我们旨在将（12）推广到整个空间。定义 $\tilde{\pi}$ = 投影$_{C_1 \cap C_2}$（π），则有

$$\operatorname{dist}(\pi, \mathcal{X}) \le \|\pi - \tilde{\pi}\| + d(\tilde{\pi}, \mathcal{X}) \le \|\pi - \tilde{\pi}\| + \tau \|\tilde{\pi} - \operatorname{proj}_{C1\cap C_2}(\tilde{\pi} + D_X \tilde{\pi} D_Y)\| \tag{13}$$

其中，第一个不等式由三角不等式得到，第二个不等式则由以下事实得到：$\tilde{\pi} \in C_1 \cap C_2$ 以及（12）。再次应用三角不等式，可得

$$\begin{aligned} &\|\tilde{\pi} - \operatorname{proj}_{C1\cap C_2}(\tilde{\pi} + D_X \tilde{\pi} D_Y)\| \\ \le &\|\tilde{\pi} - \operatorname{proj}_{C1\cap C_2}(\pi + D_X \pi D_Y)\| + \|\operatorname{proj}_{C1\cap C_2}(\pi + D_X \pi D_Y) - \operatorname{proj}_{C1\cap C_2}(\tilde{\pi} + D_X \tilde{\pi} D_Y)\| \end{aligned} \tag{14}$$

由于 $\tilde{\pi} = \operatorname{proj}_{C_1\cap C_2}(\pi)$ 以及凸集上的投影算子 $\operatorname{proj}_{C_1\cap C_2}(\cdot)$ 是非扩张的，即对于任意 x 和 y, i都有 $\|\operatorname{proj}_{C_1\cap C_2}(x) - \operatorname{proj}_{C_1\cap C_2}(y)\| \le \|x - y\|$，因此我们有

$$\begin{aligned} \|\tilde{\pi} - \operatorname{proj}_{C1\cap C_2}(\pi + D_X \pi D_Y)\| &= \|\operatorname{proj}_{C_1\cap C_2}(\pi) - \operatorname{proj}_{C1\cap C_2}(\operatorname{proj}_{C1\cap C_2}(\pi + D_X \pi D_Y))\| \\ &\le \|\pi - \operatorname{proj}_{C1\cap C_2}(\pi + D_X \pi D_Y)\|, \end{aligned} \tag{15}$$

and

$$\begin{aligned}\|\operatorname{proj}_{C1\cap C_2}(\pi + D_X\pi D_Y) - \operatorname{proj}_{C1\cap C_2}(\tilde{\pi} + D_X\tilde{\pi}D_Y)\| &\le \|(\pi + D_X\pi D_Y) - (\tilde{\pi} + D_X\tilde{\pi}D_Y)\| \\ &\le (\sigma_{\max}(D_X)\sigma_{\max}(D_Y) + 1)\|\pi - \tilde{\pi}\|\end{aligned} \tag{16}$$

其中 $\sigma_{\max}(D_X)$ 和 $\sigma_{\max}(D_Y)$ 分别表示 D_X 和 D_Y 的最大奇异值。将（14）和（16）代入（13），可得

$$\begin{aligned}\operatorname{dist}(\pi, \mathcal{X}) &\le (\tau\sigma_{\max}(D_X)\sigma_{\max}(D_Y) + \tau + 1)\|\pi - \tilde{\pi}\| + \tau\|\pi - \operatorname{proj}_{C_1\cap C_2}(\pi + D_X\pi D_Y)\| \\ &\le (\tau\sigma_{\max}(D_X)\sigma_{\max}(D_Y) + 2\tau + 1)\|\pi - \operatorname{proj}_{C_1\cap C_2}(\pi + D_X\pi D_Y)\|\end{aligned} \tag{17}$$

其中，最后一个不等式成立是因为

$$\|\pi - \tilde{\pi}\| = \|\pi - \operatorname{proj}_{C_1\cap C_2}(\pi)\| = \min_{x\in C_1\cap C_2}\|\pi - x\| \le \|\pi - \operatorname{proj}_{C_1\cap C_2}(\pi + D_X\pi D_Y)\|.$$

令 $\tau = \tau\sigma_{\max}(D_X)\sigma_{\max}(D_Y) + 2\tau + 1$，即可得到所需结果。 □

B BAPG 的收敛性分析

假设 B.1。 临界点集$\mathcal{X}$非空。

定义 B.2 （Bregman 散度）。 我们定义邻近度量 D_h ：定义域（h） ×内部（定义域（h） $\to \mathbb{R}_+$

$$D_h(x, y) := h(x) - h(y) - \langle\nabla h(y), x - y\rangle$$

邻近度量 D_h 即所谓的 ***Bregman***距离。它度量 x与 y之间的邻近程度。事实上，由梯度不等式可知， h 是凸的当且仅当 D_h（x， y） ≥ 0， $\forall x \in$定义域 h， $y \in$内部定义域 h。

具体示例

- 相对熵，即 Kullback–Leibler散度， h（x） $= \sum_i x_i \log x_i$，其 Bregman 散度为

$$D_h(x, y) = \sum_i x_i \log\frac{x_i}{y_i}.$$

- 二次函数： h（x） $= \frac{1}{2}\|x\|^2$。因此，Bregman 散度 D_h（x,y）将化为欧氏距离。

为完整呈现我们的算法开发，我们还给出一般 BAPG（4）更新在选择相对熵作为 Legendre 函数时如何推出（5）的详细过程。

$$\begin{aligned}\pi^{k+1} &= \arg\min_{\pi\in C_1}\left\{-\langle\pi, D_Y w^k D_Y\rangle + \rho D_h(\pi, w^k)\right\} \\ &= \arg\min_{\pi\in C_1} D_h\left(\pi, w^k \odot \exp\left(\frac{D_X w^k D_Y}{\rho}\right)\right), \\ &= \operatorname{diag}(\mu./P^k 1_m)P^k,\end{aligned}$$

其中 $P^k = w^k \odot \exp\left(\frac{D_X w^k D_Y}{\rho}\right)$。

$$\begin{aligned}w^{k+1} &= \arg\min_{w\in C_2}\left\{-\langle w, D_X\pi^{k+1}D_Y\rangle + \rho D_h(w, \pi^{k+1})\right\} \\ &= \arg\min_{w\in C_2} D_h\left(w, \pi^{k+1} \odot \exp\left(\frac{D_X\pi^{k+1}D_Y}{\rho}\right)\right), \\ &= Z^k\operatorname{diag}(\nu./{Z^k}^T 1_n),\end{aligned}$$

其中 $Z^k = \pi^{k+1} \odot \exp\left(\frac{D_X\pi^{k+1}D_Y}{\rho}\right)$。每次迭代的最优性条件如下，

$$\begin{cases}0 \in -D_X w^k D_Y + \rho(\nabla h(\pi^{k+1}) - \nabla h(w^k)) + \mathcal{N}_{C_1}(\pi^{k+1}), \\ 0 \in -D_X\pi^{k+1}D_Y + \rho(\nabla h(w^{k+1}) - \nabla h(\pi^{k+1})) + \mathcal{N}_{C_2}(w^{k+1}).\end{cases} \tag{18}$$

引理 3.2 的证明

证明。为证明

$$\left\|\mathrm{proj}_{C_1}(x)+\mathrm{proj}_{C_2}(y)-2\,\mathrm{proj}_{C_1\cap C_2}\left(\frac{x+y}{2}\right)\right\|\leqslant M\left\|\mathrm{proj}_{C_1}(x)-\mathrm{proj}_{C_2}(y)\right\|, \tag{19}$$

我们首先将（19）的左端转换为

$$\begin{aligned}&\left\|\mathrm{proj}_{C_1}(x)-\mathrm{proj}_{C_1\cap C_2}\left(\frac{x+y}{2}\right)+\mathrm{proj}_{C_2}(y)-\mathrm{proj}_{C_1\cap C_2}\left(\frac{x+y}{2}\right)\right\|\\ \leq&\sqrt{2}\left\|\left(\mathrm{proj}_{C_1\cap C_2}\left(\frac{x+y}{2}\right),\mathrm{proj}_{C_1\cap C_2}\left(\frac{x+y}{2}\right)\right)-\left(\mathrm{proj}_{C_1}(x),\mathrm{proj}_{C_2}(y)\right)\right\|.\end{aligned}$$

然后，该不等式可以视为线性二次问题的最优解的稳定性，即

$$(p(r),q(r))=\begin{array}{ll}\arg\min\limits_{p,q} & \frac{1}{2}\|x-p\|^2+\frac{1}{2}\|y-q\|^2\\ \text{s.t.} & p-q=r,\\ & p\in C_1,q\in C_2.\end{array} \tag{20}$$

当 $r=0$时，二元组 $(p(0),\ q(0))$ 满足 $p(0)=q(0)=\mathrm{proj}_{C_1\cap C_2}\left(\frac{x+y}{2}\right)$。此外，参数$r$ 本身可以视为扰动量，而它确实就是（19）的右端。根据（Zhang & Luo，2022年）中的定理 4.1，我们可以用扰动量 r界定两个最优解之间的距离，即

$$\|(p(0),q(0))-(p(r),q(r))\|\leq M\|r\|.$$

备注 *B.3.* 对于任意向量 r，令（*p(r)*，*q(r)*）为（*20*）的最优解。一方面，取 $r=0$，容易看出 $p(0)=q(0)=\mathrm{proj}_{C_1\cap C_2}(\frac{x+y}{2})$。另一方面，通过取$r=\mathrm{proj}_{C_1}(x)-\mathrm{proj}_{C_2}(y)$，容易看出（*p(r)*，*q(r)*）$=(\mathrm{proj}_{C_1}(x),\mathrm{proj}_{C_2}(y))$。由于（*Zhang & Luo*，*2022*年）中的定理 *4.1* 意味着对于任意 r，我们有

$$\|(p(0),q(0))-(p(r),q(r))\|_2\leq M\|r\|_2$$

其中 M仅取决于 C_1和 C_2。通过选取 $r=$投影$_{C_1}$（x）−投影$_{C_2}$（y），我们有

$$\|(\mathrm{proj}_{C_1\cap C_2}(\frac{x+y}{2}),\mathrm{proj}_{C_1\cap C_2}(\frac{x+y}{2}))-(\mathrm{proj}_{C_1}(x),\mathrm{proj}_{C_2}(y))\leq M\|\,\mathrm{proj}_{C_1}(x)-\mathrm{proj}_{C_2}(y)\|_2.$$

□

命题 3.3 的证明——BAPG 不动点的近似界

定义 B.4（**有界线性正则性**（参见 Bauschke （1996））。设C_1、……、C_N是 $\mathbb{R}^n$ 的闭凸子集，且具有 C非空交集$\{C_1$、……、$C_N\}$。若对于每个有界子集 $\mathbb{B}$ 中的 $\mathbb{R}^n$，均存在一个常数 $\kappa>0$ 使得

$$d(x,C)\leq\kappa\max_{i\in\{1,\ldots,N\}}d(x,C_i),\ \textit{for all}\ x\in\mathbb{B}.$$

如果 C_i 全部是多面体集合，则 BLR 条件将自动成立。

证明。 回顾 BAPG 的不动点集：

$$\mathcal{X}_{\mathrm{BAPG}}=\left\{(\pi^\star,w^\star):\begin{array}{l}\nabla f(w^\star)+\rho(\nabla h(\pi^\star)-\nabla h(w^\star))+p=0,p\in\mathcal{N}_{C_1}(\pi^\star)\\ \nabla f(\pi^\star)+\rho(\nabla h(w^\star)-\nabla h(\pi^\star))+q=0,q\in\mathcal{N}_{C_2}(w^\star)\end{array}\right\}. \tag{21}$$

定义 $\hat{}\pi=\mathrm{proj}_{C_1\cap C_2}(\pi^\star)$，我们首先要说明以下不等式成立，

$$\|\hat{\pi}-\pi^\star\|+\|\hat{\pi}-w^\star\|\leq(2\kappa+1)\|\pi^\star-w^\star\|.$$

由于多面体约束满足 BLR 条件，详见定义 B.4，我们有

$$\begin{aligned}\|\hat{\pi}-\pi^\star\|+\|\hat{\pi}-w^\star\| &\le 2\|\hat{\pi}-\pi^\star\|+\|\pi^\star-w^\star\| \\ &= 2\,\mathrm{dist}(\pi^\star, C_1\cap C_2)+\|\pi^\star-w^\star\| \\ &\le (2\kappa+1)\|\pi^\star-w^\star\|.\end{aligned} \tag{22}$$

基于（8）中定义的驻点，我们有：

$$\nabla f(w^\star)^T(\hat{\pi}-\pi^\star)+\rho(\nabla h(\pi^\star)-\nabla h(w^\star))^T(\hat{\pi}-\pi^\star)+p^T(\hat{\pi}-\pi^\star)=0, p\in\mathcal{N}_{C_1}(\pi^\star),$$
$$\nabla f(\pi^\star)^T(\hat{\pi}-w^\star)+\rho(\nabla h(w^\star)-\nabla h(\pi^\star))^T(\hat{\pi}-w^\star)+q^T(\hat{\pi}-w^\star)=0, q\in\mathcal{N}_{C_2}(w^\star).$$

将上述两个方程相加，

$$\nabla f(w^\star)^T(\hat{\pi}-\pi^\star)+\nabla f(\pi^\star)^T(\hat{\pi}-w^\star)+\rho(\nabla h(\pi^\star)-\nabla h(w^\star))^T(w^\star-\pi^\star)+p^T(\hat{\pi}-\pi^\star)+q^T(\hat{\pi}-w^\star)=0$$
$$\overset{(\clubsuit)}{\Rightarrow}\nabla f(w^\star)^T(\hat{\pi}-\pi^\star)+\nabla f(\pi^\star)^T(\hat{\pi}-w^\star)-\rho(D_h(\pi^\star,w^\star)+D_h(w^\star,\pi^\star))+p^T(\hat{\pi}-\pi^\star)+q^T(\hat{\pi}-w^\star)=0$$
$$\overset{(\spadesuit)}{\Rightarrow}\nabla f(w^\star)^T(\hat{\pi}-\pi^\star)+\nabla f(\pi^\star)^T(\hat{\pi}-w^\star)-\rho(D_h(\pi^\star,w^\star)+D_h(w^\star,\pi^\star))\ge 0,$$

其中（♣）成立，因为$(\nabla h(x)-\nabla h(y))^T(x-y)=D_h(x,y)+D_h(y,x)$，而（♠）由法锥的性质可得。例如，由于 $q\in\mathcal{N}_{C_1}(\pi^\star)$，我们有 $q^T(\hat{\pi}-\pi^\star)\le 0$（即，$\hat{\pi}\in C_1$）。因此，

$$\begin{aligned}\rho(D_h(\pi^\star,w^\star)+D_h(w^\star,\pi^\star)) &\le \nabla f(w^\star)^T(\hat{\pi}-\pi^\star)+\nabla f(\pi^\star)^T(\hat{\pi}-w^\star) \\ &\le \|\nabla f(w^\star)\|\|\hat{\pi}-\pi^\star\|+\|\nabla f(\pi^\star)\|\|\hat{\pi}-w^\star\| \\ &\le L_f(\|\hat{\pi}-\pi^\star\|+\|\|\hat{\pi}-w^\star\|) \\ &\overset{(22)}{\le} (2\kappa+1)L_f\|\pi^\star-w^\star\|.\end{aligned}$$

由于 $f(\cdot)$ 是二次函数，且有效定义域 $C_1\cap C_2$有界，因此倒数第二个不等式成立。由此，其梯度的范数自然具有一个常数上界，即 $L_f=\sigma_{\max}(D_X)\sigma_{\max}(D_Y)$。由于 h 是 σ-强凸函数，我们有

$$D_h(\pi^\star,w^\star)\ge\frac{\sigma}{2}\|\pi^\star-w^\star\|^2.$$

结合这一性质，我们可以量化不可行性误差 $\|\pi^\star-w^\star\|$，

$$\|\pi^\star-w^\star\|\le\frac{(2\kappa+1)L_f}{\sigma\rho}. \tag{23}$$

当 $\rho\to+\infty$时，很容易观察到不可行性误差项 $\|\pi^\star-w^\star\|$ 将收缩至零。更重要的是，如果 $\pi^\star=w^\star$，那么 X_{BAPG} 将与 $\mathcal{X}$ 完全相同。接下来，我们旨在量化 BAPG 的不动点集与原问题的临界点集之间的逼近间隙（1）。根据（21），我们有

$$\nabla f\left(\frac{\pi^\star+w^\star}{2}\right)+\frac{p+q}{2}=0,$$

where $\nabla f(\cdot)$ 是一个线性算子。通过对（1）应用 Luo-Tseng 局部误差界条件，即

即命题 述3.1，我们有

$$\begin{aligned}\mathrm{dist}\left(\frac{\pi^\star+w^\star}{2},\mathcal{X}\right) &\le \tau\left\|\frac{\pi^\star+w^\star}{2}-\mathrm{proj}_{C_1\cap C_2}\left(\frac{\pi^\star+w^\star}{2}-\nabla f\left(\frac{\pi^\star+w^\star}{2}\right)\right)\right\| \\ &= \tau\left\|\frac{\pi^\star+w^\star}{2}-\mathrm{proj}_{C_1\cap C_2}\left(\frac{\pi^\star+p+w^\star+q}{2}\right)\right\| \\ &\overset{(\clubsuit)}{=} \frac{\tau}{2}\left\|\mathrm{proj}_{C_1}(\pi^\star+p)+\mathrm{proj}_{C_1}(w^\star+q)-2\,\mathrm{proj}_{C_1\cap C_2}\left(\frac{\pi^\star+p+w^\star+q}{2}\right)\right\| \\ &\overset{(\spadesuit)}{\le} \frac{M\tau}{2}\|\pi^\star-w^\star\|,\end{aligned}$$

其中 (♣) 由法锥性质成立，即 投影$_C(x+z)=x$ 对于任意 $x\in C$ 以及$z\in\mathcal{N}_C(x)$，且 (♠) 由引理 3.2得出。结合（23），BAPG 的近似界已得到定量刻画，即

$$\mathrm{dist}\left(\frac{\pi^\star+w^\star}{2},\mathcal{X}\right)\le\frac{(2\kappa+1)L_fM}{2\sigma\rho}.$$

□

命题 3.5 — BAPG 的充分下降性质

证明。 我们首先由主要更新的最优性条件注意到，即

$$0 \in \nabla_\pi f(\pi^k, w^k) + \rho(\nabla h(\pi^{k+1}) - \nabla h(w^k)) + \partial g_1(\pi^{k+1}) \tag{24}$$

$$0 \in \nabla_w f(\pi^{k+1}, w^k) + \rho(\nabla h(w^{k+1}) - \nabla h(\pi^{k+1})) + \partial g_2(w^{k+1}) \tag{25}$$

其中 $g_1(\pi) = I_{\{\pi \in C_1\}}$ 和 $g_2(w) = I_{\{w \in C_2\}}$。 由于 $g_1(\cdot)$ 的凸性，容易得到：

$$\begin{aligned} g_1(\pi^k) - g_1(\pi^{k+1}) &\geq -\langle \nabla_\pi f(\pi^k, w^k) + \rho(\nabla h(\pi^{k+1}) - \nabla h(w^k)), \pi^k - \pi^{k+1} \rangle \\ &= -\langle \nabla_\pi f(\pi^k, w^k), \pi^k - \pi^{k+1} \rangle + \langle \rho(\nabla h(\pi^{k+1}) - \nabla h(w^k)), \pi^{k+1} - \pi^k \rangle \end{aligned}$$

由于 $f(\pi, w) = -\mathrm{tr}(D_X \pi D_Y w^T)$ 是双线性函数，因此有：

$$f(\pi^k, w^k) - f(\pi^{k+1}, w^k) - \langle \nabla_\pi f(\pi^k, w^k), \pi^k - \pi^{k+1} \rangle = 0.$$

因此，我们得到

$$f(\pi^k, w^k) + g_1(\pi^k) - f(\pi^{k+1}, w^k) - g_1(\pi^{k+1}) \geq \rho \langle \nabla h(\pi^{k+1}) - \nabla h(w^k), \pi^{k+1} - \pi^k \rangle. \tag{26}$$

类似地，基于 w-更新，我们得到

$$f(\pi^{k+1}, w^k) + g_2(w^k) - f(\pi^{k+1}, w^{k+1}) - g_2(w^{k+1}) \geq \rho \langle \nabla h(w^{k+1}) - \nabla h(\pi^{k+1}), w^{k+1} - w^k \rangle. \tag{27}$$

Combine with (26) and (27), we obtain

$$\begin{aligned} &f(\pi^k, w^k) + g_1(\pi^k) + g_2(w^k) - f(\pi^{k+1}, w^{k+1}) - g_1(\pi^{k+1}) - g_2(w^{k+1}) \\ &\geq \rho \langle \nabla h(\pi^{k+1}) - \nabla h(w^k), \pi^{k+1} - \pi^k \rangle + \rho \langle \nabla h(w^{k+1}) - \nabla h(\pi^{k+1}), w^{k+1} - w^k \rangle. \end{aligned} \tag{28}$$

右侧可以进一步简化，

$$\begin{aligned} &\rho \langle \nabla h(\pi^{k+1}) - \nabla h(w^k), \pi^{k+1} - \pi^k \rangle + \rho \langle \nabla h(w^{k+1}) - \nabla h(\pi^{k+1}), w^{k+1} - w^k \rangle \\ &\overset{(\spadesuit)}{=} \rho(D_h(\pi^k, \pi^{k+1}) + D_h(\pi^{k+1}, w^k) - D_h(\pi^k, w^k)) + \rho(D_h(w^k, w^{k+1}) + \\ &\quad D_h(w^{k+1}, \pi^{k+1}) - D_h(w^k, \pi^{k+1})) \\ &\overset{(\spadesuit)}{=} \rho D_h(\pi^k, \pi^{k+1}) + \rho D_h(w^k, w^{k+1}) - \rho D_h(\pi^k, w^k) + \rho D_h(\pi^{k+1}, w^{k+1}). \end{aligned}$$

这里，$(\clubsuit)$ 使用了 Bregman 散度的三点性质成立这一事实，即对于任意 y、$z \in \mathrm{int\,dom}\, h$ 和 $x \in \mathrm{dom}\, h$,

$$D_h(x, z) - D_h(x, y) - D_h(y, z) = \langle \nabla h(y) - \nabla h(z), x - y \rangle.$$

结合(26)和(27)，我们得到(28)。(28)的右侧可以进一步简化，

$$\begin{aligned} &\rho \langle \nabla h(\pi^{k+1}) - \nabla h(w^k), \pi^{k+1} - \pi^k \rangle + \rho \langle \nabla h(w^{k+1}) - \nabla h(\pi^{k+1}), w^{k+1} - w^k \rangle \\ &\overset{(\spadesuit)}{=} \rho(D_h(\pi^k, \pi^{k+1}) + D_h(\pi^{k+1}, w^k) - D_h(\pi^k, w^k)) + \rho(D_h(w^k, w^{k+1}) + \\ &\quad D_h(w^{k+1}, \pi^{k+1}) - D_h(w^k, \pi^{k+1})) \end{aligned}$$

这里，$(\clubsuit)$ 使用了 Bregman 散度的三点性质成立这一事实，即对于任意 y、$z \in \mathrm{int\,dom}\, h$ 和 $x \in \mathrm{dom}\, h$,

$$D_h(x, z) - D_h(x, y) - D_h(y, z) = \langle \nabla h(y) - \nabla h(z), x - y \rangle.$$

这与(28)一起推出

$$F_\rho(\pi^{k+1}, w^{k+1}) - F_\rho(\pi^k, w^k) \leq -\rho D_h(\pi^k, \pi^{k+1}) - \rho D_h(w^k, w^{k+1}) - \rho \left(D_h(\pi^{k+1}, w^k) - D_h(w^k, \pi^{k+1}) \right). \tag{29}$$

对(29)从 $k = 0$ 到 $+\infty$求和，得到

$$F_\rho(\pi^\infty, w^\infty) - F_\rho(\pi^0, w^0) + \rho \sum_{k=0}^{\infty} \left(D_h(\pi^{k+1}, w^k) - D_h(w^k, \pi^{k+1}) \right) \leq -\rho \sum_{k=0}^{\infty} \left(D_h(\pi^k, \pi^{k+1}) + D_h(w^k, w^{k+1}) \right).$$

由于势函数 $F_\rho(\cdot, \cdot)$ 是强制的，且 $\{(\pi^k, w^k)\}_{k \geq 0}$ 是有界序列，因此，假设 $\sum_{k=0}^{\infty} (D_h(\pi^{k+1}, w^k) - D_h(w^k, \pi^{k+1}))$ 有界，就意味着左端有界，这意味着

$$\sum_{k=0}^{\infty} \left(D_h(\pi^k, \pi^{k+1}) + D_h(w^k, w^{k+1}) \right) < +\infty.$$

□

定理 3.6 — BAPG 的子序列收敛结果

证明。设 (π^∞, w^∞) 是序列 $\{(\pi^k, w^k)\}_{k\geq 0}$ 的一个极限点。那么，存在一个序列 $\{n_k\}_{k\geq 0}$，使得 $\{(\pi^{n_k}, w^{n_k})\}_{k\geq 0}$ 收敛于 (π^∞, w^∞)。在式 (24) 和 (25) 中用 n_k 替换 k，并令 $k \to \infty$ 时两边取极限

$$0 \in \nabla_\pi f(\pi^\infty, w^\infty) + \rho(\nabla h(\pi^\infty) - \nabla h(w^\infty)) + \partial g_1(\pi^\infty)$$
$$0 \in \nabla_w f(\pi^\infty, w^\infty) + \rho(\nabla h(w^\infty) - \nabla h(\pi^\infty)) + \partial g_2(w^\infty).$$

基于 $\nabla_\pi f(\pi^\infty, w^\infty) = \nabla f(w^\infty)$ 且 $\nabla_w f(\pi^\infty, w^\infty) = \nabla f(\pi^\infty)$ 这一事实，可以很容易地得出 $(\pi^\infty, w^\infty) \in \mathrm{X}_{\mathrm{BAPG}}$。 □

Mo更重要的是，上述子序列收敛结果可以扩展到全局收敛 r-
gen性，只要 Legendre 函数 $h(\cdot)$ 是二次函数。

定理 B.5 (BAPG 的全局收敛 — 二次情形)。 序列 $\{(\pi^k, w^k)\}_{k\geq 0}$收敛到 $F_\rho(\pi, w)$ 的临界点。

证明。 为了调用（Attouch 等人，2010；2013）提出的 Kurdyka-Lojasiewicz 分析框架，我们必须建立两个关键性质——充分下降和保护条件。第一个性质已在命题 3.5 中得到证明。接下来，我们希望证明存在一个常数 κ_2 使得，

$$\mathrm{dist}(0, \partial F_\rho(\pi^{k+1}, w^{k+1})) \leq \kappa_2(\|\pi^{k+1} - \pi^k\| + \|w^{k+1} - w^k\|).$$

首先，注意到

$$\partial F_\rho(\pi^{k+1}, w^{k+1}) = \begin{bmatrix} \nabla_\pi f(\pi^{k+1}, w^{k+1}) + \rho(\pi^{k+1} - w^{k+1}) + \partial g_1(\pi^{k+1}) \\ \nabla_w f(\pi^{k+1}, w^{k+1}) + \rho(w^{k+1} - \pi^{k+1}) + \partial g_2(w^{k+1}) \end{bmatrix}.$$

再次结合更新式（24）和（25），可得

$$\begin{aligned} \mathrm{dist}^2(0, \partial F_\rho(\pi^{k+1}, w^{k+1})) &\leq \|\nabla_\pi f(\pi^{k+1}, w^{k+1}) - \nabla_\pi f(\pi^k, w^k)\|^2 + \rho\|w^k - w^{k+1}\|^2 \\ &\quad + \|\nabla_w f(\pi^{k+1}, w^k) - \nabla_w f(\pi^{k+1}, w^{k+1})\|^2 \\ &\leq (\sigma_{\max}^2(D_X)\sigma_{\max}^2(D_Y) + \rho)\|w^k - w^{k+1}\|^2, \end{aligned}$$

其中，$\sigma_{\max}(\cdot)$ 表示最大奇异值。令 $\kappa_2 = \sqrt{\sigma_{\max}^2(D_X)\sigma_{\max}^2(D_Y) + \rho}$，即可得到所需结果。 □

C 其他实验细节

表 4：各种算法的比较。“精确求解”意味着算法能够达到原始 GW 问题的临界点，而不是近似问题的临界点。

	单循环?	可证明?	精确求解（1）?
eBPG（Solomon等人，2016）	No	Yes	No
BPG（Xu等人，2019b）	No	Yes	Yes
BPG-S（Xu等人，2019a）	Yes	No	No
FW算法（Vayer 等人，2019年a）	No	Yes	Yes
BAPG（**我们的方法**）	Yes	Yes	No

表5：图对齐数据库的统计信息。

数据集	# 样本	平均节点数	平均边数
合成数据	300	1500	56579
蛋白质	1113	39.06	72.82
酶	600	32.63	62.14
Reddit	500	375.9	449.3

Tab图对齐结果（5 个独立实验中各基于GW的方法的均值 ± 标准差） t
试验 在噪声生成过程的不同随机种子下。

方法	合成数据	蛋白质-噪声	酶-噪声	Reddit-噪声
FW	24.50±0.95	20.24±0.38	22.80±1.15	18.34±0.22
SpecGW	13.27±0.49	19.31±0.70	21.14±0.45	19.66±0.16
ScalaGW	17.93±0.27	16.05±0.06	11.46±0.22	0.70±0.05
eBPG	34.33±0.61	45.85±0.73	60.46±0.57	3.34±0.03
BPG	57.56±0.45	52.46±0.25	62.32±0.51	36.68±0.02
BPG-S	61.48±0.34	52.74±0.13	62.21±0.55	36.68±0.02
KL-BAPG	**99.79±0.21**	**57.16±0.23**	**62.85±0.48**	**49.45±0.02**

C.1 验证 有界性累积对称误差在二维玩具示例中的验证 见第4.1

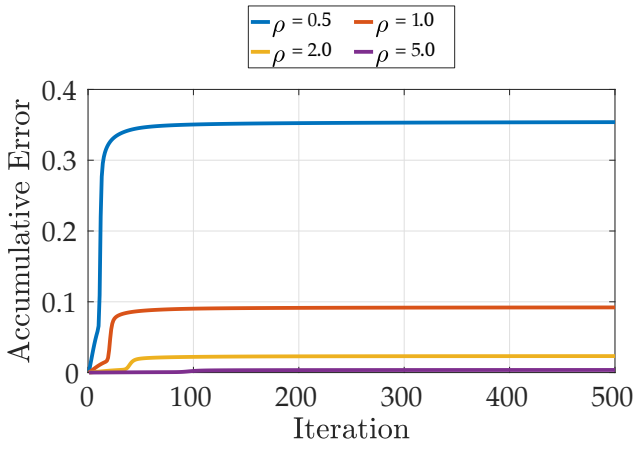


图 3：累积非对称误差

C.2 效率、准确率与可行性权衡的补充实验

在本小节中，我们通过全面的实验来证明 KL-BAPG 的有效性。表 7 显示，FW算法和 SpecGW 使用线性规划作为内层求解器，因此不可行性误差几乎为零；eBPG 和 BPG 使用 Sinkhorn 求解器，因此产生中等程度的不可行性误差；BPG-S 和所提出的 KL-BAPG 属于不可行方法，因此产生相对较大的不可行性误差，但具有更快的收敛速度。需要注意的是，BPG-S 是一种启发式方法，缺乏理论支持，在自然含噪数据观测下可能表现不佳（甚至无法收敛）。总体而言，我们可以看到，所提出的 KL-BAPG 能够牺牲一定的可行性，以同时获得更高的计算效率和匹配准确率。

表 7：在四个图对齐数据库的原始值版本上，各比较基线的平均边际误差 $\|\pi^T 1_n - \nu\| + \|\pi 1_m - \mu\|$

方法	合成数据			蛋白质			酶			Reddit		
	Acc	Time	误差	Raw	Time	误差	Raw	Time	误差	Raw	Time	误差
FW	24.50	8182	4.6e-9	29.96	54.2	4.5e-9	32.17	10.8	<1e-10	21.51	1121	<1e-10
SpecGW	13.27	1462	<1e-10	78.11	30.7	<1e-10	79.07	6.7	<1e-10	50.71	1074	<1e-10
eBPG	34.33	9502	3.0e-10	67.48	208.2	9.2e-6	78.25	499.7	1.2e-6	3.76	1234	2.9e-6
BPG	57.56	22600	2.5e-7	71.99	130.4	2.2e-7	79.19	73.1	1.2e-7	39.04	1907	1.3e-6
BPG-S	61.48	18587	9.9e-6	71.74	40.4	5.5e-6	79.25	13.4	1.2e-6	39.04	1431	7.8e-6

我们还针对步长 ρ对 KL-BAPG 开展了敏感性实验。表 8中的结果从实验上验证了命题 3.3，并指导我们如何选择步长。较大的 ρ意味着较小的不可行性误差，但收敛速度较慢。如果所得不可行性误差不是太大（即步长 ρ 不是太小），最终匹配准确率对步长具有鲁棒性，步长只会影响收敛速度。为进一步说明适度的不可行性误差对于图对齐任务并非严重问题，我们执行了后处理步骤（即使用较大的步长额外运行 100 步），以减小不可行性误差。匹配准确率没有变化。

表8：KL-BAPG 在四个图对齐数据库上使用不同 ρ 时的平均边际误差。* 表示在收敛后，使用更大的 ρ （即 $\rho = 100$ ）继续运行 KL-BAPG 100epochs。

KL-BAPG	合成数据			蛋白质			酶			Reddit		
	Acc	Time	误差	Raw	Time	误差	Raw	Time	误差	Raw	Time	误差
ρ=0.5	94.58	4610	9.4e-4	77.21	88.7	8.8e-6	79.78	18.2	8.6e-7	50.19	264	3.6e-6
ρ=0.5*	94.57	4703	1.4e-5	77.22	98.1	8.7e-8	79.79	22.2	4.8e-8	50.19	280	8.7e-8
ρ=0.2	99.50	2760	2.0e-3	78.27	63.5	9.4e-5	79.39	10.4	6.2e-6	50.92	197	1.9e-6
ρ=0.2*	99.50	2854	1.5e-5	78.27	72.9	2.1e-7	79.39	14.3	5.6e-8	50.92	213	9.1e-8
ρ=0.1	99.79	1253	3.3e-3	78.18	59.1	1.4e-3	79.67	6.9	2.1e-5	50.93	115	5.1e-6
ρ=0.1*	99.79	1349	9.8e-6	78.16	68.6	7.8e-6	79.67	10.8	6.1e-8	50.93	130	9.5e-8
ρ=0.05	99.20	956	5.2e-3	77.59	36.2	1.7e-2	79.27	5.0	2.9e-3	50.96	102	1.5e-5
ρ=0.05*	99.20	1051	8.1e-6	77.48	45.7	2.1e-5	79.30	8.9	3.1e-6	50.96	118	9.7e-8
ρ=0.01	93.63	616	4.1e-2	60.08	14.6	1.3e-1	66.86	4.0	8.4e-2	-	-	-
ρ=0.01*	95.02	712	4.9e-6	60.92	24.0	9.3e-3	67.20	7.8	4.4e-5	-	-	-

C.3 多组学单细胞整合

为验证 KL-BAPG 在图对齐和划分之外的应用能力，我们进一步评估了所提出的 BAPG 在另一项任务——多组学单细胞整合中的表现。该任务旨在整合细胞不同模态下的多种分子特征，例如基因表达和染色质可及性（Teichmann & Efremova，2020年；Wen 等，2022年；Cao 等，2020年；2022年）。这为全面了解细胞提供了机会，并被《自然》评选为“2019年度方法”（Teichmann & Efremova，2020年）。

任务定义 假设我们有两个单细胞多组学数据集，$X = [x_1; \cdots; x_{n_x}] \in \mathbb{R}^{n_x \times d_x}$和 $Y = [y_1; \cdots; y_{n_y}] \in \mathbb{R}^{n_x \times d_x}$，涵盖两种模态，其中 n_x/n_y和 d_x/d_y分别表示 X/Y的细胞样本数和特征维度。给定 X和 Y，多组学单细胞整合的任务旨在计算两种模态的联合表示。在联合表示中，X和 Y中相同的细胞类型应聚类在一起。因此，预先对齐 X和 Y中的相同细胞类型，可以提升联合表示的质量。

表 9：两个真实世界单细胞多组学数据集的统计信息。

	scGEM		scMNT		
模态名称	Gene 表达	DNA 甲基化	Gene 表达	DNA 甲基化	染色质可及性
样本 n	177	177	612	709	1940
特征 d	34	27	300	300	300

数据集与评估 根据Cao 等 （2020年；2022年），我们在两个真实的单细胞多组学数据集上开展实验：(1) 用于单细胞基因型、表达和甲基化数据分析的 scGEM 数据集，以及 (2) 用于单细胞核小体、甲基组和转录组数据分析的 scNMT 数据集。数据集统计信息列于表 9中。我们使用标签转移准确率（即两个模态之间具有相同细胞类型的已对齐细胞对所占的比例）来衡量将共享细胞的标签从一个模态转移到另一个模态的能力。

基线方法 在本文中，我们将所提出的 KL-BAPG 与其他 GW 方法进行比较，具体做法是直接将距离矩阵 D_X 和 D_Y设置为特征相似度矩阵。此外，我们加入了两个针对该任务的基线方法，包括 UnionCom（Cao 等，2020年）和 Pamona（Cao 等，2022年）。Pamona 同样依赖于 GW距离方法的计算，并使用 eBPG 作为 GW 求解器。我们将 Panona 中的 eBPG 替换为 KL-BAPG，从而得到一种名为 Pamona-BAPG 的新方法。

表10：单细胞多组学整合中，scGEM 和 scNMT 数据集上的标签迁移准确率（%）和实际运行时间（秒）比较。D.、G. 和 Ch. 分别表示 DNA甲基化、基因表达和染色质可及性模态。

	scGEM			scMNT				
方法	D.→G.	G.→D.	Time	D.→G.	G.→D.	Ch.→G.	G→Ch.	Time
FW	40.7	40.7	0.7	74.2	**74.3**	44.1	45.4	19.9
eBPG	7.9	10.7	7.4	63.4	68.0	46.6	44.8	64.2
specBPG	21.5	21.5	0.4	61.4	57.2	36.4	39.7	2.2
BPG-S	9.0	14.1	1.2	69.0	73.2	60.1	60.7	2.3
BPG	11.9	11.3	11.5	**75.1**	73.2	60.8	60.0	22.3
KL-BAPG	**52.0**	**50.8**	**0.3**	74.0	**74.3**	**68.0**	**67.0**	**0.8**
UnionCom	48.0	50.8	6.4	85.6	76.3	80.1	78.8	147.4
Pamona	61.6	**58.2**	4.4	71.6	56.6	71.9	66.8	26.1
Pamona-BAPG	**67.2**	**58.2**	0.7	**85.7**	**91.3**	**80.7**	**82.9**	6.5

所有方法的结果 表 10 汇总了 scGWM 和 scNMT 数据集上所有比较方法的结果。在所有 GW 方法中，KL-BAPG 在 6 项指标中的 4 项上取得了最佳性能，并且运行时间最短。UnionCom 和 Pamona 的表现优于大多数纯 GW 方法，因为它们采用了针对该任务的额外预处理和后处理技术。此外，Pamona-BAPG 在 6 项指标中的 5 项上优于 Pamona，并将运行时间缩短了 16%–25%。这表明，GW距离的计算是该任务的瓶颈，而更优的 GW 求解器能够极大促进相关下游应用。由于这只是一项旨在展示将 KL-BAPG 应用于其他任务潜力的初步研究，因此我们将其他潜在的、由应用驱动的任务留待未来研究。

C.4 本文所有基线方法的源代码及其许可证

- BPG-S（GPL-3.0 许可证）（Xu等人，2019b）：
  https://github.com/HongtengXu/gwl
- ScalaGW（无许可证）（Xu等人，2019a）：
  https://github.com/HongtengXu/s-gwl
- SpecGW（MIT 许可证）（Chowdhury 和 Needham，2021）：
  https://github.com/trneedham/Spectral-Gromov-Wasserstein
- eBPG（MIT 许可证）（Flamary 等人，2021）：https://github.com/PythonOT/POT
- FW算法（MIT 许可证）（Flamary 等人，2021）：https://github.com/PythonOT/POT
- IPFP、RRWM、SpecMethod（Mulan PSL v2 许可证）：
  https://github.com/Thinklab-SJTU/pygmtools