arXiv:2304.06821v1 [stat.ML] 13 Apr 2023

基于成对比较的排序：一般图与具有局部性的图

Yanxi Chen*

2023年4月17日

摘要

本技术报告研究了经典 Bradley-Terry-Luce（BTL）模型中基于成对比较的排序问题，重点关注分数估计。对于一般图，我们证明，当样本数量足够多时，最大似然估计（MLE）能够实现与克拉美-罗下界相匹配的逐元素估计误差，该下界可以用有效电阻来表示；我们分析的关键在于建立了统计估计与通过预条件梯度下降进行的迭代优化之间的联系。我们还特别关注具有局部性的图，在这类图中，只有彼此邻近的项目才能通过边连接；我们的分析确定了局部性不会造成性能损失的条件，即比较图中相距较远的一对项目的得分，与比较一对相邻项目的得分几乎同样容易。我们进一步探讨了分治算法，这些算法即使在样本最为稀疏的情况下，也能以可证明的方式实现类似的性能保证，同时具备一定的计算优势。数值结果验证了我们的理论，并证实了所提出算法的有效性。

关键词： B　布拉德利–特里–卢斯模型、成对比较、逐项误差、具有局部性的图

目录

1 引言 **2**

1.1 主要贡献.3 1.2 相关工作.4 1.3 记号.4

2 预备知识 **4**

2.1 BTL模型的经典算法 .4 2.2 图拉普拉斯矩阵与有效电阻 .5 2.3 误差度量 .7

3 经典算法 **7**

3.1 一般图中的最大似然估计 .7 3.2 具有局部性的图中的最大似然估计 .9 3.3 定理 1 的证明 .9 3.4 推论1的证明 .12 3.5 谱方法是否有效？ .14

*电气与计算机工程系，普林斯顿大学，美国新泽西州普林斯顿市 08544；电子邮箱：yxchen0@gmail.com。

4 分治算法 **15**
4.1 DC-overlap：采用重叠子图的分治方法 .154.2 DC-overlap 的分析 .164.2.1 误差分解 .164.2.2 对具有局部性的图的影响 .174.3 通过投影梯度下降求解最大似然估计 .204.4 DC-community：采用社区的分治方法 .21

5 数值实验 **23**
5.1 最大似然估计与谱方法.235.2 最大似然估计与 DC-overlap.235.3 用各种方法求解最大似然估计.24

6 讨论 **26**

A 延后证明 **26**A.1 引理1的证明 .26A.1.1 第 1 部分 .26A.1.2 第 2 部分 .26

B 其他结果 **28**

1 引言

我们研究 Bradley-Terry-Luce（BTL）模型中基于成对比较的排序问题 [Zer29，BT52，Luc12]。考虑 n 个具有真实分数的项目 $\boldsymbol{\theta}^\star \in \mathbb{R}^n$，以及一个一般的连通图 $G = (V, E)$，它有 $n = |V|$ 个节点；在本技术报告中，我们始终考虑无向图，并不失一般性地假设 $\langle \boldsymbol{\theta}^\star, \mathbf{1}_n \rangle = \sum_{1 \le i \le n} \theta_i^\star = 0$。假设 $\boldsymbol{\theta}^\star$ 在全局和局部都具有有限的动态范围，即存在某些 $\kappa \ge \kappa_E \ge 1$ 使得

$$\max_{1 \le i < j \le n} |\theta_i^\star - \theta_j^\star| \le \log \kappa \quad \text{and} \quad \max_{(i,j) \in E} |\theta_i^\star - \theta_j^\star| \le \log \kappa_E. \tag{1}$$

将 Sigmoid 函数 σ 记为 $\sigma(x) = 1/(1 + e^{-x})$。对于每条边 $(i, j) \in E, i < j$，我们收集 $L_{i,j}$个独立的伯努利随机样本，从而得到数据集

$$\{y_{i,j}^{(\ell)}, 1 \le \ell \le L_{i,j}\}_{(i,j) \in E, i<j}, \quad \text{where} \quad \mathbb{P}(y_{i,j}^{(\ell)} = 1) = 1 - \mathbb{P}(y_{i,j}^{(\ell)} = 0) = \sigma(\theta_i^\star - \theta_j^\star).$$

假设 $y_{i,j}^{(\ell)} = 1 - y_{j,i}^{(\ell)}$对所有 $(i, j) \in E$ 均成立。我们还将充分统计量记为

$$y_{i,j} := \frac{1}{L_{i,j}} \sum_{1 \le l \le L_{ij}} y_{i,j}^{(\ell)} = 1 - y_{j,i}, \quad (i, j) \in E.$$

我们的目标是基于数据 $\{y_{i,j}^{(\ell)}\}$ 或 $\{y_{i,j}\}$，准确估计真实分数 $\boldsymbol{\theta}^\star$。

关于 BTL 模型的文献十分丰富；尤其是 Erdős-Rényi 图这一特殊情形的统计保证已得到广泛研究。相比之下，关于逐元素 估计误差的一般图 文献相对稀少，且未能通过紧致的误差界正确刻画问题的内在结构；有关先前工作的更详细讨论，请参见第 1.2 节。此外，现实世界中的图通常具有局部性 [CKST16]，，即只能对物理空间中彼此邻近的一对项目进行测量；BTL 模型的相关文献尚未涉及这种情形。具体而言，我们考虑具有 n 个节点、半径为 r 且边采样概率为p的一维/二维网格，形式化定义为 一维网格/二维网格，见定义 1，并在图 1中示意。在本文中，我们将一维/二维网格与一维网格/二维网格 这两种说法互换使用。值得注意的是，本文中的许多分析都可以扩展到其他具有局部性的图，例如随机几何图或具有社区结构的图。

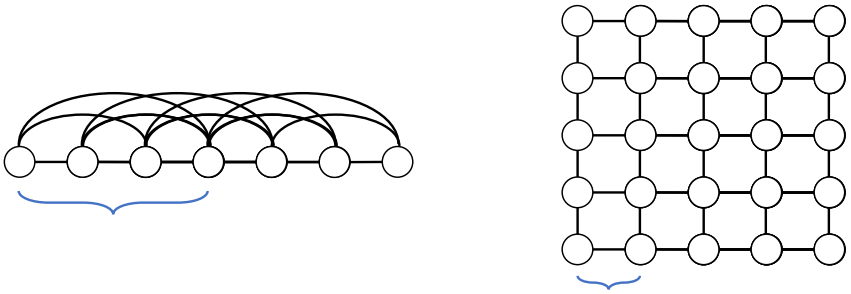


图 1：具有局部性的图的可视化。左：一维网格$(n = 7,\ r = 3,\ p = 1)$。右：二维网格$(n = 25,\ r = 1,\ p = 1)$。

定义 1. 我们定义一维网格和二维网格如下。

- 一维网格$(n,\ r,\ p)$：节点可由 $\{1, 2, \ldots, n\}$ 编号，使得 $(i,\ j) \in E$ 的概率为 p（当 $|i-j| \le r$ 时），而 $(i,\ j) \notin E$ 则在其他情况下成立。
- 二维网格$(n,\ r,\ p)$：节点可由 $\{\boldsymbol{i} = (i_1,\ i_2) : 1 \le i_1,\ i_2 \le \sqrt{n}\}$ 编号，使得 $(\boldsymbol{i},\ \boldsymbol{j}) \in E$ 的概率为 p（当 $|i_1 - j_1| + |i_2 - j_2| \le r$ 时），而 $(i,\ j) \notin E$ 则在其他情况下成立。

1.1 主要贡献

本研究的主要贡献有两方面。

1. 对于一般图，我们证明，当样本数量足够多时，经典的最大似然估计（MLE）方法能够实现逐元素估计误差，该误差与克拉美-罗下界相匹配，差异至多为一个 $\widetilde{O}(\sqrt{\kappa_E})$ 因子以及若干残差项。更具体地，记 $\boldsymbol{\theta}^{\mathsf{MLE}}$ 为最大似然估计解，则对于所有 $1 \le k < \ell \le n$,

$$\left|(\theta_k^{\mathsf{MLE}} - \theta_\ell^{\mathsf{MLE}}) - (\theta_k^\star - \theta_\ell^\star)\right| \le \widetilde{O}\left(\sqrt{\kappa_E \Omega_{k,\ell}(\boldsymbol{L}_z)}\right) + \text{residual terms},$$

其中 $\boldsymbol{L}_z = \nabla^2 \mathcal{L}(\boldsymbol{\theta}^\star)$ 是图拉普拉斯矩阵，而 $\Omega_{k,\ell}(\boldsymbol{L}_z) = (\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_k - \boldsymbol{e}_\ell)$ 可解释为加权图中一对节点之间的有效电阻 (k, ℓ)；形式化表述见定理 1。注意，我们只要求 κ_E 有限，而不对式 (1) 中定义的全局动态范围作任何假设；我们将看到，这意味着谱方法（BTL模型的另一种经典算法）可能存在缺点。

2. 对于具有局部性的图，尤其是一维网格/二维网格$(n,\ r,\ p)$，通过在前述结果中指定有效电阻，我们可以轻松得到最大似然估计的保证；然而，证明最大似然估计在最稀疏区域也能达到此类保证仍是一个开放问题，即 $L_{i,\ j} \ge 1$，且对于一维网格为 $\Omega(1)\ rp \ge$，或对于二维网格为 $\Omega(1)\ r^2 p \ge \widetilde{}$。我们进一步探讨分治算法，它们不仅具有一定的计算优势，还能在最稀疏区域实现近乎最优的逐元素估计误差；具体而言，假设 $\kappa_E \lesssim 1$ 且 $L_{i,\ j} = L$ 对所有 $(i,\ j) \in E$均成立，则输出解$\boldsymbol{\theta}$ 对所有 $1 \le k < \ell \le n$均满足，

$$\left|(\theta_k - \theta_\ell) - (\theta_k^\star - \theta_\ell^\star)\right| \le \begin{cases} \widetilde{O}\left(\sqrt{\frac{n}{r^2} + 1}\sqrt{\frac{1}{rpL}}\right) + \text{residual terms} & \text{for } \mathsf{Grid1D}(n, r, p), \\ \widetilde{O}\left(\sqrt{\frac{\log n}{r^2} + 1}\sqrt{\frac{1}{r^2 pL}}\right) + \text{residual terms} & \text{for } \mathsf{Grid2D}(n, r, p). \end{cases}$$

一个令人意外的结论是，局部性并不会带来负面影响，只要半径 r 满足 $r \gtrsim \sqrt{n}$（对于一维网格），或满足 $r \gtrsim \sqrt{\log n}$（对于二维网格）即可；换句话说，在图中比较两个相距较远的项目的分数，几乎与比较两个相邻项目的分数一样容易。

作为我们分析的副产品，我们设计了两种用于求解最大似然估计优化问题的迭代方法，即预条件梯度下降（注 3）以及结合重新参数化的投影梯度下降（算法 4），这两种方法均具有快速收敛的特性。相比之下，用于求解最大似然估计解的标准方法（包括梯度下降、坐标下降、下界最大化等）本质上都是去中心化的，因此不可避免地会面临收敛缓慢的问题，尤其是在具有局部性的图中。最后，我们通过数值实验验证了理论研究结果。

1.2 相关工作

过去几十年来，关于基于成对比较的排序已有大量文献；例如参见 [FJ57、YYX12、CS15、CFMW19、HYTC20、GSZ21、CGZ22a、CGZ22b、BR22、HOS19、HOS20、Hun04、LSR22、LFL22、NOS12、SBB+16、SY99、VYZ22、WSR20] 及其中的参考文献。下文将重点讨论与我们工作最相关的论文。

针对 Bradley-Terry-Luce（BTL）模型的两种经典算法，即最大似然估计（MLE）和谱方法（即 Rank Centrality [NOS12]），已经得到了广泛研究。例如，[CFMW19] 研究了 Erdős-Rényi 图，并证明正则化最大似然估计和谱方法均能在样本数最少的情况下实现最优的逐元素估计误差。特别地，他们基于迭代优化算法的动力学分析了 MLE 的统计误差；这一技术将为我们分析一般图中的 MLE 提供启发。[GSZ21]也研究了 Erdős-Rényi 图，并研究了最稀疏区域下 MLE 和谱方法的不确定性量化；更具体地，他们证明误差向量 $\boldsymbol{\theta} - \boldsymbol{\theta}^\star$（其中 $\boldsymbol{\theta}$ 是任一算法的解）可以分解为一个主项与一个低阶项之和，其中主项可以显式地表示为独立噪声项 $\{y_{i,j} - \sigma(\theta_i^\star - \theta_j^\star)\}_{(i,j)\in E}$ 的线性组合$_{i<j}$。文献中还研究了 MLE 解的其他方面，例如其存在性与唯一性 [FJ57, SY99, YYX12, HYTC20, BR22], 、偏差 [WSR20], 、渐近正态性 [WJN22, LFL22], 等。

关于一般图中的 BTL 模型的 ℓ_2 和 ℓ_∞ 估计误差，现有研究的结果（例如 [SBB+16、LSR22]）大多依赖于某些图拉普拉斯矩阵的谱间隙，而这些谱间隙无法准确刻画误差率。两个例外是 [HOS19、HOS20], ，它们将有效电阻[DS84、CRR+96、ESVM+11] 确定为 BTL 模型中估计误差的相关参数。这些研究考察了（加权）最小二乘算法，并针对某种形式的 ℓ_2 误差给出了理论保证，尽管其分析可以很容易地调整为给出 ℓ_∞ 误差界。不足之处在于，（加权）最小二乘算法无法处理稀疏样本（不仅在理论上如此，在实践中也是如此），因为这类算法需要针对每条边 (i, j) 显式计算 $\sigma^{-1}(y_{i,j}) \in E$，而这只有在每条边上都有足够多的样本时才可行（即 $y_{i,j} \notin \{0, 1\}$）。

1.3 记号

给定一个图 $G = (V, E)$，令 $\mathcal{N}(i) = \{j \in V : (i, j) \in E\}$ 表示节点 i 的邻居集合。令 $\boldsymbol{e}_i \in \mathbb{R}^n$ 表示 n 维独热向量，其第 i 个分量为 1，其余分量为 0。对于向量 $\boldsymbol{x}$ 和对称矩阵 $\boldsymbol{M} \succeq \boldsymbol{0}$，定义范数 $\|\boldsymbol{x}\|_{\boldsymbol{M}} := \sqrt{\boldsymbol{x}^\top \boldsymbol{M} \boldsymbol{x}}$。对于具有紧凑特征分解 $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$的对称矩阵 $\boldsymbol{M}$，其伪逆定义为 $\boldsymbol{M}^\dagger = \boldsymbol{U}\boldsymbol{\Lambda}^{-1}\boldsymbol{U}^\top$。对于向量 $\boldsymbol{x}$，$\exp(\boldsymbol{x})$、$\log(\boldsymbol{x})$ 等表示与其大小相同的向量，其中运算以逐元素方式作用于 $\boldsymbol{x}$ 。 $f_n \lesssim g_n$ 或 $f_n = O(g_n)$ 的含义是，存在一个通用常数 $C > 0$，使得对于任意 $n \geq 1$，都有 $f_n \leq C g_n$； $f_n \gtrsim g_n$ 或 $f_n = \Omega(g_n)$ 等价于 $g_n \lesssim f_n$； $f_n \asymp g_n$ 或 $f_n = \Theta(g_n)$ ~ ~ ~表示 $f_n \gtrsim g_n$ 和 $g_n \gtrsim f_n$ 均成立。 O、Ω Θ 的含义与 O、Ω Θ相同，只是其中隐藏了对数项。

2 预备知识

2.1 BTL模型的经典算法

BTL模型中分数估计的黄金标准方法是最大似然估计（MLE），其总结见算法1。该方法旨在找到一个解 $\boldsymbol{\theta}$，使负对数似然最小化

算法 1： 最大似然估计

输入： 图 $G=(V,\ E)$，数据 $\{y_{i,j}, L_{i,j}\}_{(i,j)\in E,\ i<j}$。

求解以下优化问题：

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^n:\boldsymbol{\theta}^\top \mathbf{1}_n=0} \quad \mathcal{L}(\boldsymbol{\theta}) = \sum_{(i,j)\in E, i<j} L_{i,j}\Big(-y_{i,j}(\theta_i-\theta_j) + \log(1+e^{\theta_i-\theta_j})\Big).$$

输出： 解 $\boldsymbol{\theta}\in\mathbb{R}^n$。

函数 $\mathcal{L}(\boldsymbol{\theta})$，这是一个凸优化问题 [CFMW19]:

$$\begin{aligned}\min_{\boldsymbol{\theta}\in\mathbb{R}^n:\boldsymbol{\theta}^\top \mathbf{1}_n=0} \quad \mathcal{L}(\boldsymbol{\theta}) &\coloneqq \sum_{(i,j)\in E, i<j}\ \sum_{1\le l\le L_{ij}} \Big(-y_{i,j}^{(\ell)}(\theta_i-\theta_j) + \log(1+e^{\theta_i-\theta_j})\Big) \\ &= \sum_{(i,j)\in E, i<j} L_{i,j}\Big(-y_{i,j}(\theta_i-\theta_j) + \log(1+e^{\theta_i-\theta_j})\Big). \end{aligned} \tag{2}$$

$\mathcal{L}(\boldsymbol{\theta})$ 的第一种形式将有助于我们的理论分析，而第二种形式对于最大似然估计的实际实现效率更高。 $\mathcal{L}(\boldsymbol{\theta})$ 的梯度和海森矩阵为

$$\nabla\mathcal{L}(\boldsymbol{\theta}) = \sum_{(i,j)\in E, i<j}\ \sum_{1\le l\le L_{ij}} \Big(\sigma(\theta_i-\theta_j) - y_{i,j}^{(\ell)}\Big)(\boldsymbol{e}_i-\boldsymbol{e}_j) = \sum_{(i,j)\in E, i<j} L_{i,j}\Big(\sigma(\theta_i-\theta_j) - y_{i,j}\Big)(\boldsymbol{e}_i-\boldsymbol{e}_j), \tag{3a}$$

$$\nabla^2\mathcal{L}(\boldsymbol{\theta}) = \sum_{(i,j)\in E, i<j} L_{i,j}\sigma'(\theta_i-\theta_j)(\boldsymbol{e}_i-\boldsymbol{e}_j)(\boldsymbol{e}_i-\boldsymbol{e}_j)^\top. \tag{3b}$$

下面给出一个经典条件，确保损失函数（2）存在唯一有限最小化器。

命题1 ([FJ57], 使用我们的记号表述)。 优化问题（2）存在唯一有限的极小值点，当且仅当满足以下条件：对于任意不相交划分 $\Omega_1\cup\Omega_2=\{1,\dots,n\}$，其中$|\Omega_1|,\ |\Omega_2|\ge 1$，存在某个$(i,j)\in(\Omega_1\times\Omega_2)\cap E$，使得 $y_{i,\ j}^{(\ell)}=1$ 对于某个$1\le\ell\le L_{i,\ j}$。

除最大似然估计外，另一种流行的算法是谱方法，也称为 Rank Centrality[NOS12]。它计算随机矩阵 $\boldsymbol{\pi}$ 的平稳分布 $\boldsymbol{P}\in\mathbb{R}^{n\times n}$ （即 $\boldsymbol{\pi}^\top=\boldsymbol{\pi}^\top\boldsymbol{P}$），该矩阵定义如下：

$$\text{for each } i,\quad P_{i,j}=\begin{cases}\frac{1}{d}y_{j,i} & \text{if } j\in\mathcal{N}(i),\\ 1-\frac{1}{d}\sum_{k\in\mathcal{N}(i)} y_{k,i} & \text{if } j=i,\\ 0 & \text{otherwise;}\end{cases}$$

此处，参数 $d>0$ 的选取使得 $\boldsymbol{P}$具有非负元素。样本数量趋于无穷时，可以验证 $\boldsymbol{\pi}^\star=\exp(\boldsymbol{\theta}^\star)$ 恰好是 $\boldsymbol{P}$ 的平稳分布。因此，样本数量足够多时，$\boldsymbol{\pi}$ 是 $\boldsymbol{\pi}^\star$的良好估计，因此$\log\boldsymbol{\pi}$ 可用作 $\boldsymbol{\theta}^\star$的估计。该方法总结于算法2。

2.2 图拉普拉斯矩阵与有效电阻

考虑一个连通无向图 $G=(V,\ E)$，其加权图拉普拉斯矩阵为

$$\boldsymbol{L}\coloneqq\boldsymbol{D}-\boldsymbol{A}=\sum_{(i,j)\in E, i<j} C_{i,j}(\boldsymbol{e}_i-\boldsymbol{e}_j)(\boldsymbol{e}_i-\boldsymbol{e}_j)^\top,$$

其中，$\boldsymbol{A}$ 是加权邻接矩阵，$\boldsymbol{D}$ 是包含节点度的对角矩阵，$C_{i,\ j}>0$ 是边（$i,\ j$）的权重。众所周知，$\boldsymbol{L}$ 的秩为 $n-1$，满足 $\boldsymbol{L}\cdot\mathbf{1}_n=\mathbf{0}$ 和$\boldsymbol{L}^\dagger\cdot\mathbf{1}_n=\mathbf{0}$，其中 $\boldsymbol{L}^\dagger$ 表示 $\boldsymbol{L}$的伪逆。任意一对节点（$k,\ \ell$）之间相对于 $\boldsymbol{L}$的有效电阻定义为

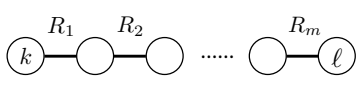
$$\Omega_{k,\ell}(\boldsymbol{L})\coloneqq(\boldsymbol{e}_k-\boldsymbol{e}_\ell)^\top\boldsymbol{L}^\dagger(\boldsymbol{e}_k-\boldsymbol{e}_\ell). \tag{4}$$

算法 2: 谱方法

输入: 图 $G = (V, E)$，数据$\{y_{i,j}\}_{(i,j)\in E,\ i<j}$，参数 d。按如下方式构造矩阵 $\boldsymbol{P} \in \mathbb{R}^{n\times n}$：

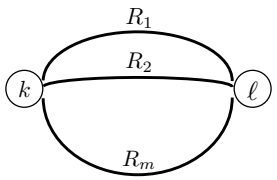
$$\text{for each } 1 \le i \le n, \quad P_{i,j} = \begin{cases} \frac{1}{d} y_{j,i} & \text{if } j \in \mathcal{N}(i), \\ 1 - \frac{1}{d}\sum_{k\in\mathcal{N}(i)} y_{k,i} & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

令 $\boldsymbol{\pi} \in \mathbb{R}^n$ 为 $\boldsymbol{P}$ 的平稳分布，并令 $\boldsymbol{\theta} =$ 为 $\log \boldsymbol{\pi}$。**输出:** $\boldsymbol{\theta} \in \mathbb{R}^n$。



串联定律:

$$\Omega_{k,\ell} = R_1 + R_2 + \cdots + R_m$$



并联定律:

$$_{k,\ \ell} = \frac{1}{C_{k,\ell}}, \quad \text{其中}$$

$$C_{k,\ell} = \frac{1}{R_1} + \frac{1}{R_2} + \cdots + \frac{1}{R_m}$$

图2：图电阻的串并联定律。

物理解释如下。图可以视为一个电网络，而权重$C_{i,j}$ 表示边 $(i, j) \in E$上的电导，即电阻$R_{i,j}$的倒数。根据欧姆定律，对于任意一对连通的节点 $(i, j) \in E$，这些节点上的电压 v_i, v_j与该边上的电流（或电流流）$I_{i,j}$ 满足$v_i - v_j = I_{i,j}R_{i,j} = I_{i,j}/C_{i,j}$。用矩阵形式表示为

$$\boldsymbol{c} = \boldsymbol{L}\boldsymbol{v}, \quad \text{or} \quad \boldsymbol{v} = \boldsymbol{L}^{\dagger}\boldsymbol{c},$$

其中，$\boldsymbol{v} \in \mathbb{R}^n$ 表示电压，$\boldsymbol{c} \in \mathbb{R}^n$ 表示外部电流。因此，有效电阻 $\Omega_{k,\ \ell}(\boldsymbol{L})$ 就是当单位外部电流流入节点 k 并从节点 ℓ流出时，两个节点之间的电压差 $v_k - v_\ell$，即 $\boldsymbol{c} = \boldsymbol{e}_k - \boldsymbol{e}_\ell$。

我们来汇总一些有效电阻的著名性质，这些性质是图论和电路理论中的标准内容 [DS84，ESVM+11，LP17，Spi19]。

事实 1。 有效电阻（4）满足以下性质：

- (串并联定律) 考虑 m 电阻为 $R_1, \ldots, R_m$ 且电导为 $C_1, \ldots, C_m$的电阻器，其中 $C_i = 1/R_i$。如果它们串联连接，则两个端点之间的有效电阻为$\Omega = \sum_{i=1}^m R_i$。如果它们并联连接，则有效电阻为 $\Omega = 1/C$，其中 $C = \sum_{i=1}^m C_i$。（可视化示意见图 2。）

- (瑞利单调性定律) 增加（分别地，减小）电阻网络中任意边上的电阻，绝不会减小（分别地，增大）任意一对节点之间的有效电阻。

- （**三角不等式**）对于三个节点 $a, b, c \in V$，有效电阻满足 $\Omega_{a,b} \le \Omega_{a,c} + \Omega_{c,b}$。

作为热身练习，我们利用其中的一些性质来证明完全图的有效电阻上界。

事实2。 设 $G = (V, E)$ 是一个含有 n 个节点的完全图，且 $\boldsymbol{L} = \sum_{1\le i<j\le n} C_{i,j}(\boldsymbol{e}_i - \boldsymbol{e}_j(\boldsymbol{e}_i - \boldsymbol{e}_j)^\top$ 是其加权图拉普拉斯矩阵，其中 $C_{i,j} \asymp 1$。那么 $\Omega_{k,\ell}(\boldsymbol{L}) \lesssim 1/n$ 对所有 $1 \le k < \ell \le n$均成立。

证明。给定 $1 \le k < \ell \le n$，考虑一个图 $G' = (V,\ E')$，其中 $E' = \{(i,\ j)\colon\ 1 \le i < j \le n,\ i = k$ 或 $j = \ell\}$，并将其图拉普拉斯矩阵记为 $\boldsymbol{L}'$。由于可以通过将边 $E \backslash E'$ 上的电阻增加至无穷大，从 G 得到 G'，根据瑞利单调性定律，我们有 $\Omega_{k,\ \ell}(\boldsymbol{L}) \le \Omega_{k,\ \ell}(\boldsymbol{L}')$。注意，在 G'中，节点 k 和 ℓ 通过 $\Theta(n)$条并行路径相连，每条路径的电阻均为 $O(1)$；因此，根据并联定律，我们有 $\Omega_{k,\ \ell}(\boldsymbol{L}') \lesssim 1/n$。 □

2.3 误差度量

将误差向量记为 $\boldsymbol{\delta} = \boldsymbol{\theta} - \boldsymbol{\theta}^\star$，其中 $\boldsymbol{\theta} \in \mathbb{R}^n$ 是真实分数 $\boldsymbol{\theta}^\star$ 的估计值。为了衡量估计误差，文献中常用的度量包括 ℓ_∞ 误差 $\|\boldsymbol{\delta}\|_\infty$ 和 ℓ_2误差 $\|\boldsymbol{\delta}\|_2$。需要注意的是，在 BTL 模型中，分数只能在全局平移的意义下确定；换言之，对于任意标量值 $c \in \mathbb{R}$，分数向量 $\boldsymbol{\theta}$ 本质上等同于 $\boldsymbol{\theta} + c\mathbf{1}_n$。因此，只有在额外假设（例如 $\boldsymbol{\delta}^\top \mathbf{1}_n = \sum_{1 \le i \le n} \delta_i = 0$）下，这些误差度量才有意义。

在这项工作中，我们提出以成对方式衡量估计误差，使用

$$\Big\{|\delta_k - \delta_\ell|\Big\}_{1 \le k < \ell \le n} \quad \text{and} \quad \max_{1 \le k < \ell \le n} |\delta_k - \delta_\ell|.$$

这些度量具有多种优点。(1) 它们不受全局平移影响，即 $\boldsymbol{\theta}$和 $\boldsymbol{\theta} + c\mathbf{1}_n$具有完全相同的误差。(2) 对每一对项目（$k$、$\ell$）测量成对误差 $|\delta_k - \delta_\ell|$，可以更全面地了解估计误差；这有助于（例如）区分图中相距较近或较远的两项，其分数比较难度。(3) 正如我们将在分析中看到的，成对误差 $\{|\delta_k - \delta_\ell|\}$ 与有效电阻 $\{\Omega_{k,\ \ell}\}$之间存在自然的对应关系。(4) 有了这些成对误差，就可以轻松推导出标准的 ℓ_∞ 和 ℓ_2 误差：不失一般性地假设 $\boldsymbol{\delta}^\top \mathbf{1}_n = 0$，则有

$$\|\boldsymbol{\delta}\|_\infty \le \max_{1 \le k < \ell \le n} |\delta_k - \delta_\ell| \le 2\|\boldsymbol{\delta}\|_\infty, \quad \text{and}$$

$$\|\boldsymbol{\delta}\|_2^2 = \frac{1}{n}\boldsymbol{\delta}^\top (n\boldsymbol{I}_n - \mathbf{1}_n \mathbf{1}_n^\top)\boldsymbol{\delta} = \frac{1}{n}\boldsymbol{\delta}^\top \sum_{k<\ell} (\boldsymbol{e}_k - \boldsymbol{e}_\ell)(\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{\delta} = \frac{1}{n}\sum_{k<\ell} (\delta_k - \delta_\ell)^2. \tag{5}$$

3 经典算法

本节将介绍关于 ℓ_∞ 一般图中最大似然估计的估计误差（第3.1节）和具有局部性的图中的估计误差（第3.2节）的新颖理论分析，并给出详细证明（第3.3节和第3.4节）。我们还将在第3.5 节讨论为何在某些最大似然估计能够成功的情形下，谱方法可能无法可靠地估计 $\boldsymbol{\theta}^\star$。

3.1 一般图中的最大似然估计

我们的主要结果和分析关键依赖于以下图拉普拉斯矩阵 $\boldsymbol{L}_z$，它由3b式中 $\mathcal{L}(\boldsymbol{\theta})$ 在真实值 $\boldsymbol{\theta}^\star$处的海森矩阵定义：

$$\boldsymbol{L}_z := \nabla^2 \mathcal{L}(\boldsymbol{\theta}^\star) = \sum_{(i,j)\in E, i<j} L_{i,j} z_{i,j} (\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^\top, \quad \text{where} \quad z_{i,j} := \sigma'(\theta_i^\star - \theta_j^\star) > 0. \tag{6}$$

以下定理为最大似然估计的 ℓ_∞ 误差提供高概率保证。请注意，随机性仅来自样本 $\{y_{i,j}^{(\ell)}\}$。

定理 1。考虑第 *1* 节中的模型和假设。存在某个通用常数 $C_0 > 0$，使得对于任意固定的 $0 < \delta < 0.5$，以下结论成立。设 $\{B_{k,\ell}, Q_{k,\ell}\}_{1 \le k < \ell \le n}$ 是以下各项的有效上界：

$$B_{k,\ell} \ge C_0 \sqrt{\Omega_{k,\ell}(\boldsymbol{L}_z) \cdot \kappa_E \log\big(\frac{n}{\delta}\big)}, \quad Q_{k,\ell} \ge \sum_{(i,j)\in E, i<j} L_{i,j} B_{i,j}^2 \big|(\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_i - \boldsymbol{e}_j)\big|, \quad 1 \le k, \ell \le n. \tag{7}$$

如果 $\{B_{k,\ell}, Q_{k,\ell}\}$ 满足：

$$Q_{k,\ell} \le 4B_{k,\ell} \quad \textit{for all } (k,\ell) \in E, \tag{8}$$

那么，以至少 $1-\delta$的概率，最大似然估计优化问题（2）存在唯一有限的极小值点$\boldsymbol{\theta}^{\mathsf{MLE}}$，并且满足以下条件：对所有 $1 \le k < \ell \le n$，

$$\left|(\theta_k^{\mathsf{MLE}} - \theta_\ell^{\mathsf{MLE}}) - (\theta_k^\star - \theta_\ell^\star)\right| \le \begin{cases} B_{k,\ell}, & (k,\ell) \in E, \\ \frac{1}{2}B_{k,\ell} + \frac{1}{8}Q_{k,\ell}, & (k,\ell) \notin E. \end{cases} \tag{9}$$

证明概要。 关键思路是考虑预条件梯度下降（PrecondGD）迭代

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \eta \boldsymbol{L}_z^\dagger \nabla \mathcal{L}(\boldsymbol{\theta}^t), \quad t = 0, 1, 2, \ldots,$$

，其初始化为 $\boldsymbol{\theta}^0 = \boldsymbol{\theta}^\star$。我们的分析包含两个主要部分。（1）以高概率，PrecondGD迭代 $\{\boldsymbol{\theta}^t\}$ 始终接近 $\boldsymbol{\theta}^\star$ ，对所有 $t \ge 0$，前提是步长 η 足够小。（2）在此条件下，可以保证损失函数 $\mathcal{L}$ 存在唯一有限的极小值点$\boldsymbol{\theta}^{\mathsf{MLE}}$，且 $\{\boldsymbol{\theta}^t\}$ 由于 $\mathcal{L}$ （当约束在子空间 $\mathbf{1}_n$内的有界凸集上时）的强凸性和光滑性，会线性收敛到该极小值点。完整证明见第 3.3节。 □

定理 1 为原始最大似然估计（未经正则化 [CFMW19]）的逐元素误差提供了一个简单统一的保证，它具有以下优点：

- 它适用于一般的连通图和一般的样本数量 $\{L_{i,j}\}$。
- 当样本量足够大时，式（9）中的主要误差项$B_{k,\ell}$ 的平方与克拉美-罗下界（$\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_k - \boldsymbol{e}_\ell) = \Omega_{k,\ell}$（$\boldsymbol{L}_z$）相符，二者相差至多一个 $\widetilde{O}$（κ_E）因子。与 [HOS19 类似，HOS20]，定理 1 指出，在一般图上的 BTL 模型中，与估计误差相关的参数是有效电阻，而不是谱间隙 [SBB+16、LSR22] 或其他参数。
- 定理 1 指出，只要 κ_E 在（1）中定义的量有界，最大似然估计就能奏效；该量可能远小于动态范围 κ。 [BR22]也有类似发现。这表明，当 κ_E 较小而 κ较大时，最大似然估计可能优于谱方法；我们将在第 3.5节对此作进一步阐述。

注 1. 对于足够大的样本量，条件 (8) 必然成立；此外，在误差界 (9) 中，$B_{k,\ell}$ 是主项，而$Q_{k,\ell}$ 是“低阶”项。为了解释这一点，我们先回顾 (6) 中 $\boldsymbol{L}_z$ 的定义，它关于 $\{L_{i,j}\}$ 因子（即各边上的样本数）呈线性依赖。为简单起见，假设对于所有 $(i,j) \in E$，都有 $L_{i,j} = L$ ；那么，在其他条件不变时，$B_{k,\ell} \asymp \sqrt{1/L}$，而$Q_{k,\ell} \asymp 1/L$，，由此便可得到上述解释。注意，如果对某些感兴趣的图能够得到 $\{Q_{k,\ell}\}$ 的紧上界，那么即使样本量达到最小值 $L = 1$，条件 (8) 也可能成立。关于定理 1 对一些特殊图的推论，参见附录中的例 1。

备注 2. 标准 ℓ_∞ 和 ℓ_2 误差的上界可根据第 1节中的定理轻松推导得出，参见第 2.3节。对于 ℓ_∞ 误差，由于 $\boldsymbol{\delta} := \boldsymbol{\theta}^{\mathsf{MLE}} - \boldsymbol{\theta}^\star$ 满足 $\boldsymbol{\delta}^\top \mathbf{1}_n = 0$，因此有

$$\|\boldsymbol{\delta}\|_\infty \le \max_{k \ne \ell} |\delta_k - \delta_\ell| = \max_{k \ne \ell} \left|(\theta_k^{\mathsf{MLE}} - \theta_\ell^{\mathsf{MLE}}) - (\theta_k^\star - \theta_\ell^\star)\right|.$$

对于 ℓ_2 错误，请回顾(5)，有 $\|\boldsymbol{\delta}\|_2^2 = \frac{1}{n}\sum_{k<\ell}(\delta_k - \delta_\ell)^2$。为简便起见，忽略低阶项 $\{Q_{k,\ell}\}$ 和对数项，我们有$(\delta_k - \delta_\ell)^2 \lesssim B_{k,\ell}^2 \lesssim \Omega_{k,\ell}(\boldsymbol{L}_z)\kappa_E$，因此

$$\|\boldsymbol{\delta}\|_2^2 = \frac{1}{n}\sum_{k<\ell}(\delta_k - \delta_\ell)^2 \lesssim \frac{1}{n}\sum_{k<\ell} B_{k,\ell}^2 \lesssim \frac{1}{n}\sum_{k<\ell} \Omega_{k,\ell}(\boldsymbol{L}_z)\kappa_E.$$

让我们将总电阻 [GBS08]表示为

$$\sum_{k<\ell} \Omega_{k,\ell}(\boldsymbol{L}_z) = \sum_{k<\ell} (\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_k - \boldsymbol{e}_\ell) = \mathsf{Tr}\Big(\boldsymbol{L}_z^\dagger \sum_{k<\ell} (\boldsymbol{e}_k - \boldsymbol{e}_\ell)(\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top\Big) = n\mathsf{Tr}(\boldsymbol{L}_z^\dagger).$$

总之，我们得到了关于 ℓ_2 误差的简洁（尽管是非正式的）界： $\|\boldsymbol{\delta}\|_2 \lesssim \sqrt{\mathsf{Tr}(\boldsymbol{L}_z^\dagger)\kappa_E}$

将事物 .

注3。我们的理论分析表明，可以采用实用的 PrecondGD 方法求解最大似然估计。虽然计算预条件器 $\boldsymbol{L}_z^\dagger$ 需要知道 $\boldsymbol{\theta}^\star$（这当然是不现实的），但在实践中可以使用某个替代矩阵作为预条件器。例如，可以直接忽略 $\{z_{i,j}\}$ 在 $\boldsymbol{L}_z$的定义（6）中的各项，从而得到可实现的预条件器 $\boldsymbol{L}_G^\dagger$，其满足 $\boldsymbol{L}_G := \sum_{(i,j)\in E,\ i<j} L_{i,j}(\boldsymbol{e}_i-\boldsymbol{e}_j)(\boldsymbol{e}_i-\boldsymbol{e}_j)^\top$。注意，$\boldsymbol{L}_G$ 满足 $\boldsymbol{L}_z \preccurlyeq \boldsymbol{L}_G \preccurlyeq 4\kappa_E \cdot \boldsymbol{L}_z$。假设 $\kappa_E \lesssim 1$ 较小且样本量 $\{L_{i,j}\}$ 较大，使得 $\boldsymbol{\theta}^{\mathsf{MLE}}$ 和朴素初始化（例如 $\boldsymbol{\theta}^0 = \mathbf{0}$）在逐对逐元素误差意义下都接近 $\boldsymbol{\theta}^\star$（见第 2.3 节）。那么，命题 3（非正式地）表明，采用步长 $\eta \asymp 1$ 的 PrecondGD 将以线性速率收敛至 $\boldsymbol{\theta}^{\mathsf{MLE}}$，其收敛速度可能显著快于通常的一阶方法（例如梯度下降、坐标下降、下界最大化 [Hun04、VYZ22], 等），尤其是在图拉普拉斯矩阵病态时；这一点将在第5.3节通过数值实验得到验证。至于计算复杂度，如果利用最先进的拉普拉斯线性系统求解器 [ST04、V$^+$13]，则 PrecondGD 更新 $\boldsymbol{L}_G^\dagger \nabla \mathcal{L}(\boldsymbol{\theta}^t)$（至少在理论上）可以在 $\widetilde{O}(|E|)$时间内近似计算。

3.2 具有局部性的图中的最大似然估计

我们将定理 1 应用于 一维网格$(n,\ r,\ p)$ 和 二维网格$(n,\ r,\ p)$，其中 $p = 1$，由此得到以下结果。

推论*1*。考虑第*1*节中的模型和假设。令 $G = (V, E)$ 为具有 $p = 1$ 的 *Grid1D*(n, r, p) 或 *Grid2D*(n, r, p)，并为简单起见，假设对于所有 $(i, j) \in E$，均有 $L_{i,j} = L$。如果对于一维网格，$rL \gtrsim \kappa_E^5 n \log(n/\delta)$，或者对于二维网格，$r^2 L \gtrsim \kappa_E^5 n \log(n/\delta)$，那么以至少 $1-\delta$ 的概率，最大似然估计优化问题 *(2)* 存在唯一有限极小值点 $\boldsymbol{\theta}^{\mathsf{MLE}}$，其满足以下条件：对于所有 $1 \le k < \ell \le n$,

$$\left|(\theta_k^{\mathsf{MLE}} - \theta_\ell^{\mathsf{MLE}}) - (\theta_k^\star - \theta_\ell^\star)\right| \lesssim \begin{cases} \kappa_E \sqrt{(\frac{n}{r^2}+1)\frac{1}{rL}\log\left(\frac{n}{\delta}\right)} & \text{for 1D grids,} \\ \kappa_E \sqrt{(\frac{\log n}{r^2}+1)\frac{1}{r^2 L}\log\left(\frac{n}{\delta}\right)} & \text{for 2D grids.} \end{cases}$$

证明概要。证明这一推论的关键步骤是，在一维/二维网格中为 $\{\Omega_{k,\ell},\ V_{k,\ \ell}\}_{1\le k<\ell\le n}$ 找到有效的上界，其中 $\Omega_{k,\ \ell} = \Omega_{k,\ \ell}(\boldsymbol{L}_z)$ 和

$$V_{k,\ell} := \sum_{(i,j)\in E, i<j} L_{i,j}\left|(\boldsymbol{e}_i - \boldsymbol{e}_j)^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_k - \boldsymbol{e}_\ell)\right|. \tag{10}$$

完成这一步后，我们便可得到 $\{B_{k,\ell}, Q_{k,\ell}\}$ 在（7）中定义的有效上界，推论随即由定理 1得出。完整证明见第 3.4 节。 □

接下来有两点说明。

- 推论 1 中的误差界有一个令人惊讶的含义：只要半径满足 $r \gtrsim \sqrt{n}$（对于 一维网格）或 $r \gtrsim \sqrt{\text{对数}\ n}$（对于 二维网格），局部性并不会带来影响（相差一个（κ_E）因子）；换言之，估计图中相距较远的两个项目之间的分数差，几乎与估计相邻两个项目之间的分数差一样容易。

- 推论1的主要不足在于对样本量的要求较高，这是因为我们的证明中对 $\{V_{k,\ell}\}$ 采用了较宽松的界。尽管我们推测，通过针对许多感兴趣的图进行专门分析，可以对相关参数 $\{\Omega_{k,\ell}, V_{k,\ell}\}$ 进行精确控制，但遗憾的是，我们未能对一维网格和二维网格做到这一点。理想情况下，我们希望证明最大似然估计在最稀疏区域的误差界，即对于一维网格(n, r, p)，有 $rpL \ge \widetilde{\Omega}(1)$；对于二维网格$(n, r, p)$，有 $r^2 pL \ge \widetilde{\Omega}(1)$；此外，对于一般采样概率 p 的情形，逐项误差界中还应多出一个 $1/\sqrt{p}$ 因子。证明最大似然估计能够达到这些保证，是未来工作中一个有价值的方向。在第 4 节中，我们提出了能够实现此类保证的替代算法。

3.3 定理1

在继续之前，我们先回顾一下 $\boldsymbol{L}_z$ 和 $\{z_{i,j}\}$ 在（6）中的定义，并引入一些有用的额外记号。在整个分析过程中，我们将用 C_0，C_1，C_2，.....表示正的通用常数。下面定义关联矩阵

$$\boldsymbol{B} := \left[\ldots, \sqrt{z_{i,j}}(\boldsymbol{e}_i - \boldsymbol{e}_j), \ldots\right]_{1\le \ell \le L_{i,j};(i,j)\in E, i<j} \in \mathbb{R}^{n\times L_{\text{total}}},$$

其中，下标中的“$1 \le \ell \le L_{i,\,j}$”表示列$\sqrt{z_{i,\,j}}(\boldsymbol{e}_i - \boldsymbol{e}_j)$重复出现$L_{i,\,j}$次，而$L_{\mathsf{total}} := \sum_{(i,\,j)\in E,\,i<j} L_{i,\,j}$是比较的总数。注意，

$$\boldsymbol{B}\boldsymbol{B}^\top = \sum_{(i,j)\in E, i<j} L_{i,j} z_{i,j} (\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^\top = \boldsymbol{L}_z.$$

此外，我们定义噪声项

$$\epsilon_{i,j}^{(\ell)} := y_{i,j}^{(\ell)} - \sigma(\theta_i^\star - \theta_j^\star), \quad 1 \le \ell \le L_{i,j}, (i,j) \in E, i < j,$$

每个噪声项的均值均为零，方差为$\mathsf{var}(\epsilon_i^{(\ell)}{}_{,\,j}) = z_{i,\,j}$，且在 $[-1,1]$ 上有界。此外，令 $\widehat{\boldsymbol{\epsilon}}$ 为归一化独立噪声项组成的向量，即

$$\widehat{\boldsymbol{\epsilon}} := \Big[\ldots, \epsilon_{i,j}^{(\ell)}/\sqrt{z_{i,j}}, \ldots\Big]^\top_{1\le\ell\le L_{i,j};(i,j)\in E, i<j} \in \mathbb{R}^{L_{\mathsf{total}}};$$

每个条目 $\widehat{\boldsymbol{\epsilon}}$ 的均值为零、方差为一，且次高斯范数为 [Ver18，第 2 节。5] $\|\cdot\|_{\psi_2} \lesssim 1/\sqrt{z_{i,\,j}} \lesssim \sqrt{\kappa_E}$(因为 $|\overline{\theta_i^\star} - \theta_j^\star| \le$对数 κ_E，因此 $z_{i,\,j} = \sigma'(\theta_i^\star - \theta_j^\star) \ge 1/(4\kappa_E)$ 根据事实 3）。

步骤 1：PrecondGD 动力学 考虑损失函数 $\{\boldsymbol{\theta}^t\}$ 对应的预条件梯度下降动力学 $\mathcal{L}(\boldsymbol{\theta})$，其定义见 (2，初始化为 $\boldsymbol{\theta}^0 = \boldsymbol{\theta}^\star$。在第 t 次 PrecondGD 迭代中，有

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \eta \boldsymbol{L}_z^\dagger \nabla \mathcal{L}(\boldsymbol{\theta}^t),$$

其中，梯度可以改写为

$$\begin{aligned}
\nabla \mathcal{L}(\boldsymbol{\theta}^t) &= \sum_{(i,j)\in E, i<j} \sum_{1\le l\le L_{ij}} \Big(\sigma(\theta_i^t - \theta_j^t) - y_{i,j}^{(\ell)}\Big)(\boldsymbol{e}_i - \boldsymbol{e}_j) \\
&= \sum_{(i,j)\in E, i<j} \sum_{1\le l\le L_{ij}} \Big(\sigma(\theta_i^t - \theta_j^t) - \sigma(\theta_i^\star - \theta_j^\star) - \epsilon_{i,j}^{(\ell)}\Big)(\boldsymbol{e}_i - \boldsymbol{e}_j) \\
&\overset{\text{(i)}}{=} \sum_{(i,j)\in E, i<j} \sum_{1\le l\le L_{ij}} \Big(\sigma'(\theta_i^\star - \theta_j^\star)(\delta_i^t - \delta_j^t) + \frac{1}{2}\sigma''(\zeta_{i,j}^t)(\delta_i^t - \delta_j^t)^2 - \epsilon_{i,j}^{(\ell)}\Big)(\boldsymbol{e}_i - \boldsymbol{e}_j) \\
&= \boldsymbol{L}_z \boldsymbol{\delta}^t - \boldsymbol{B}\widehat{\boldsymbol{\epsilon}} + \underbrace{\frac{1}{2}\sum_{(i,j)\in E, i<j} L_{i,j}\sigma''(\zeta_{i,j}^t)(\delta_i^t - \delta_j^t)^2(\boldsymbol{e}_i - \boldsymbol{e}_j)}_{=:\boldsymbol{r}^t} \\
&= \boldsymbol{L}_z \boldsymbol{\delta}^t - \boldsymbol{B}\widehat{\boldsymbol{\epsilon}} + \boldsymbol{r}^t;
\end{aligned} \tag{11}$$

这里，(i) 是由于 根据泰勒定理，对于每个 $(i,\ j)$，$\zeta_{i,\,j}^t$ 位于 $\theta_i^\star - \theta_j^\star$ 和 $\theta_i^t - \theta_j^t$ 之间 E.

回顾这一假设：不失一般性，$\boldsymbol{\theta}^\star \perp \mathbf{1}_n$。容易验证，$\nabla\mathcal{L}(\boldsymbol{\theta}^t)^\top \mathbf{1}_n = 0$ 和 $\boldsymbol{L}_z^\dagger \mathbf{1}_n = 0$，这意味着对于所有 $t \ge 0$，都有 $\boldsymbol{\theta}^t$，$\boldsymbol{\theta}^t - \boldsymbol{\theta}^\star \perp \mathbf{1}_n$。因此，PrecondGD 动力学的一切都发生在与 $\mathbf{1}_n$ 正交的子空间中。

将第 t 次迭代时的优化误差记为 $\boldsymbol{\delta}^t := \boldsymbol{\theta}^t - \boldsymbol{\theta}^\star$，可得

$$\boldsymbol{\delta}^{t+1} = \boldsymbol{\delta}^t - \eta \boldsymbol{L}_z^\dagger(\boldsymbol{L}_z\boldsymbol{\delta}^t - \boldsymbol{B}\widehat{\boldsymbol{\epsilon}} + \boldsymbol{r}^t) = \boldsymbol{\delta}^t - \eta(\boldsymbol{\delta}^t - \boldsymbol{L}_z^\dagger \boldsymbol{B}\widehat{\boldsymbol{\epsilon}} + \boldsymbol{L}_z^\dagger \boldsymbol{r}^t) = (1-\eta)\boldsymbol{\delta}^t + \eta\boldsymbol{\xi} - \eta\boldsymbol{L}_z^\dagger \boldsymbol{r}^t,$$

其中我们定义

$$\boldsymbol{\xi} := \boldsymbol{L}_z^\dagger \boldsymbol{B}\widehat{\boldsymbol{\epsilon}}.$$

直观而言，我们希望证明 $\{\boldsymbol{\delta}^t\}$ 收敛到主要误差项 $\boldsymbol{\xi}$，其误差仅为由残差项 $\boldsymbol{L}_z^\dagger \boldsymbol{r}^t$ 引起的低阶误差。在接下来的步骤中，我们将进一步研究这两个误差项。

步骤 1.1：主要误差项 $\boldsymbol{\xi}$。对于任意 $k \neq \ell$，有

$$\xi_k - \xi_\ell = (\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{\xi} = (\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger \boldsymbol{B}\widehat{\boldsymbol{\epsilon}} = \langle \boldsymbol{B}^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_k - \boldsymbol{e}_\ell), \widehat{\boldsymbol{\epsilon}} \rangle,$$

这是一个均值为零的次高斯随机变量，且满足

$$\begin{aligned}
\|\xi_k - \xi_\ell\|_{\psi_2}^2 &\lesssim \left\|\boldsymbol{B}^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_k - \boldsymbol{e}_\ell)\right\|_2^2 \cdot \kappa_E \\
&= (\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger \boldsymbol{B}\boldsymbol{B}^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_k - \boldsymbol{e}_\ell) \cdot \kappa_E \\
&= (\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_k - \boldsymbol{e}_\ell) \cdot \kappa_E \\
&= \Omega_{k,\ell}(\boldsymbol{L}_z) \cdot \kappa_E.
\end{aligned}$$

利用次高斯随机变量的集中性 [Ver18，第 2 节。5] 并对所有$(k,\ \ell)$使用并集界，我们得到以下结果：以至少 $1-\delta$ 的概率，有

$$|\xi_k - \xi_\ell| \leq C_1 \sqrt{\Omega_{k,\ell}(\boldsymbol{L}_z) \cdot \kappa_E \log\left(\frac{n}{\delta}\right)} \leq \frac{1}{2} B_{k,\ell} \quad \text{for all } 1 \leq k < \ell \leq n, \tag{12}$$

这里我们回顾7中$B_{k\text{的定义},\ \ell} \geq C_0 \sqrt{\Omega_{k,\ \ell}(\boldsymbol{L}_z) \cdot \kappa_E \log(\frac{n}{\delta})}$，并假设 $C_0 > 2C_1$。值得注意的是，在证明定理1时，这一步是唯一涉及数据随机性的部分；证明中的其他内容都是确定性的。

步骤 1.2：残差项 $\boldsymbol{L}_z^\dagger \boldsymbol{r}^t$。回顾11中 $\boldsymbol{r}^t$的定义，我们有

$$\boldsymbol{L}_z^\dagger \boldsymbol{r}^t = \frac{1}{2} \sum_{(i,j)\in E, i<j} L_{i,j} \sigma''(\zeta_{i,j}^t)(\delta_i^t - \delta_j^t)^2 \boldsymbol{L}_z^\dagger (\boldsymbol{e}_i - \boldsymbol{e}_j).$$

因此，对于任意 $k \neq \ell$，

$$\begin{aligned}
\left|(\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger \boldsymbol{r}^t\right| &= \left|\frac{1}{2} \sum_{(i,j)\in E, i<j} L_{i,j} \sigma''(\zeta_{i,j}^t)(\delta_i^t - \delta_j^t)^2 (\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_i - \boldsymbol{e}_j)\right| \\
&\leq \frac{1}{2} \sum_{(i,j)\in E, i<j} L_{i,j} \left|\sigma''(\zeta_{i,j}^t)\right| \cdot (\delta_i^t - \delta_j^t)^2 \cdot \left|(\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_i - \boldsymbol{e}_j)\right| \\
&\leq \frac{1}{8} \sum_{(i,j)\in E, i<j} L_{i,j} (\delta_i^t - \delta_j^t)^2 \left|(\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_i - \boldsymbol{e}_j)\right|,
\end{aligned}$$

其中我们使用了因　对于所有 $x \in \mathbb{R}$，t 满足 $|\sigma''(x)| \leq 1/4$。这对 $\{\delta_i^t - \delta_j^t\}_{(i,j)\in E}$ 呈现二次依赖。　.

步骤 2：PrecondGD 与真实值保持接近 步骤 1 中的分析得出以下结论：对于任意 $k \neq \ell$，

$$\begin{aligned}
|\delta_k^{t+1} - \delta_\ell^{t+1}| &\leq (1-\eta)|\delta_k^t - \delta_\ell^t| + \eta|\xi_k - \xi_\ell| + \eta\left|(\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger \boldsymbol{r}^t\right| \\
&\leq (1-\eta)|\delta_k^t - \delta_\ell^t| + \eta\Big(\frac{1}{2} B_{k,\ell} + \frac{1}{8} \sum_{(i,j)\in E, i<j} L_{i,j} (\delta_i^t - \delta_j^t)^2 \left|(\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_i - \boldsymbol{e}_j)\right|\Big).
\end{aligned} \tag{13}$$

考虑以下归纳假设（显然在 $t=0$ 时对 $\boldsymbol{\delta}^0 = \boldsymbol{\theta}^0 - \boldsymbol{\theta}^\star = \boldsymbol{0}$ 成立）：在第 t 次迭代时，

$$|\delta_k^t - \delta_\ell^t| \leq B_{k,\ell} \quad \text{for all} \quad (k,\ell) \in E. \tag{14}$$

然后，利用假设（8），对于所有 （k，ℓ） $\in E$，

$$|\delta_k^{t+1} - \delta_\ell^{t+1}| \leq (1-\eta)|\delta_k^t - \delta_\ell^t| + \eta\Big(\frac{1}{2} B_{k,\ell} + \frac{1}{8} \sum_{(i,j)\in E, i<j} L_{i,j} B_{i,j}^2 \left|(\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_i - \boldsymbol{e}_j)\right|\Big)$$

$$\le (1-\eta)B_{k,\ell} + \eta\Big(\frac{1}{2}B_{k,\ell} + \frac{1}{8}Q_{k,\ell}\Big)$$
$$\le (1-\eta)B_{k,\ell} + \eta\Big(\frac{1}{2}B_{k,\ell} + \frac{1}{8}\cdot 4B_{k,\ell}\Big) = B_{k,\ell}.$$

因此，我们证明了第（$t+1$）次迭代的归纳假设。请注意，此分析仅涉及（$k,\ \ell$）$\in E$ 对。

接下来，考虑剩余的 $(k,\ \ell) \notin E$ 对。从 (13) 出发并利用归纳假设 (14)，可得

$$|\delta_k^{t+1} - \delta_\ell^{t+1}| \le (1-\eta)|\delta_k^t - \delta_\ell^t| + \eta\Big(\frac{1}{2}B_{k,\ell} + \frac{1}{8}\sum_{(i,j)\in E, i<j} L_{i,j}B_{i,j}^2\big|(\boldsymbol{e}_k-\boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger(\boldsymbol{e}_i-\boldsymbol{e}_j)\big|\Big)$$
$$\le (1-\eta)|\delta_k^t - \delta_\ell^t| + \eta\Big(\frac{1}{2}B_{k,\ell} + \frac{1}{8}Q_{k,\ell}\Big).$$

因此，$|\delta_k^t - \delta_\ell^t| \le \frac{1}{2}B_{k,\ \ell} + \frac{1}{8}Q_{k,-\ell}$ 意味着 $|\delta_k^{t+1} - \delta_\ell^{t+1}|$ 具有相同的上界。

<escape>**步骤 3：PrecondGD 收敛至最大似然估计解** 我们将此步骤的结果，即 $\boldsymbol{\theta}^{\mathsf{MLE}}$ 的存在性以及 PrecondGD 的收敛性，总结在以下引理中。其证明推迟至附录 A.1，该证明基于凸优化中的标准结果以及命题 1。</escape>

引理1。C　考虑定理*1*的设定。在事件（*12*）发生的条件下，有如下结论　：

1. 该优化问题（2）具有唯一有限最小化器 $\boldsymbol{\theta}^{\mathsf{MLE}}$；2. 存在 $0<\alpha_1\le\alpha_2$ 使得 PrecondGD 迭代满足 $\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^{\mathsf{MLE}}\|_{\boldsymbol{L}_z} \le (1-\eta\alpha_1)^t\|\boldsymbol{\theta}^0 - \boldsymbol{\theta}^{\mathsf{MLE}}\|_{\boldsymbol{L}_z}$ 对所有 $t\ge 0$均成立，前提是 $0<\eta\le 1/\alpha_2$。

总结 迄今为止的分析可概括如下。存在一个事件 $\mathcal{E}$ （即 (12) 成立），其概率为 $\mathbb{P}(\mathcal{E}) \ge 1-\delta$，使得在条件 $\mathcal{E}$ 下，PrecondGD 迭代 $\{\boldsymbol{\theta}^t\}$ 满足以下性质：首先，对于 $(k,\ell)\in E$，$|(\theta_k^t - \theta_\ell^t) - (\theta_k^\star - \theta_\ell^\star)| \le B_{k,\ \ell}$；并且，对于 $(k,\ell)\notin E$，$B_{k,\ell}/2 + Q_{k,\ell}/8$（步骤 2）；此外，存在唯一且有限的最大似然估计解 $\boldsymbol{\theta}^{\mathsf{MLE}}$，且 $\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^{\mathsf{MLE}}\|_{\boldsymbol{L}_z} \le (1-\eta\alpha_1)^t\|\boldsymbol{\theta}^0 - \boldsymbol{\theta}^{\mathsf{MLE}}\|_{\boldsymbol{L}_z}$（步骤 3）。换言之，PrecondGD 迭代 $\{\boldsymbol{\theta}^t\}$ 逐元素地保持接近 $\boldsymbol{\theta}^\star$，同时收敛到 $\boldsymbol{\theta}^{\mathsf{MLE}}$，误差可任意小（经过足够多次迭代后）。至此，我们完成了定理 1 的证明。

3.4 推论1 的证明

我们对推论 1 的证明依赖于以下关键引理。回顾记号

$$\Omega_{k,\ell} = \Omega_{k,\ell}(\boldsymbol{L}_z) = (\boldsymbol{e}_k-\boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^\dagger(\boldsymbol{e}_k-\boldsymbol{e}_\ell), \quad V_{k,\ell} = \sum_{(i,j)\in E, i<j} L_{i,j}\big|(\boldsymbol{e}_i-\boldsymbol{e}_j)^\top \boldsymbol{L}_z^\dagger(\boldsymbol{e}_k-\boldsymbol{e}_\ell)\big|.$$

引理 2。 对于一维网格$(n, r, p=1)$，我们有

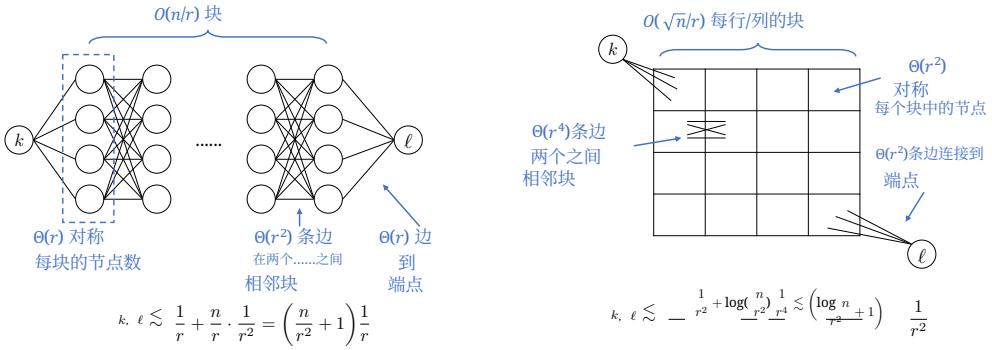
$$\Omega_{k,\ell} \lesssim \begin{cases} \frac{1}{r}\cdot\frac{\kappa_E}{L}, & (k,\ell)\in E,\\ \frac{1}{r}\cdot\frac{\kappa_E}{L}\cdot(\frac{n}{r^2}+1) & (k,\ell)\notin E,\end{cases} \qquad V_{k,\ell} \lesssim \begin{cases} \kappa_E^{1.5}\sqrt{n}, & (k,\ell)\in E,\\ \kappa_E^{1.5}\sqrt{n}\cdot\sqrt{\frac{n}{r^2}+1}, & (k,\ell)\notin E;\end{cases}$$

类似地，对于二维网格$(n, r, p=1)$，我们有

$$\Omega_{k,\ell} \lesssim \begin{cases} \frac{1}{r^2}\cdot\frac{\kappa_E}{L}, & (k,\ell)\in E,\\ \frac{1}{r^2}\cdot\frac{\kappa_E}{L}\cdot(\frac{\log n}{r^2}+1) & (k,\ell)\notin E,\end{cases} \qquad V_{k,\ell} \lesssim \begin{cases} \kappa_E^{1.5}\sqrt{n}, & (k,\ell)\in E,\\ \kappa_E^{1.5}\sqrt{n}\cdot\sqrt{\frac{\log n}{r^2}+1}, & (k,\ell)\notin E.\end{cases}$$

有了这个引理，对于一维网格，我们定义

$$B_E := \sqrt{\frac{\kappa_E^2}{rL}\log(\frac{n}{\delta})}, \quad V_E := \kappa_E^{1.5}\sqrt{n},$$



(a) 一维网格$(n, r, p = 1)$　　(b) 二维网格$(n, r, p = 1)$

图 3：具有局部性的图的有效电阻。

并令

$$B_{k,\ell} \asymp \begin{cases} B_E, & (k,\ell) \in E, \\ B_E\sqrt{\frac{n}{r^2}+1}, & (k,\ell) \notin E, \end{cases} \quad Q_{k,\ell} \asymp \begin{cases} B_E^2 V_E, & (k,\ell) \in E, \\ B_E^2 V_E\sqrt{\frac{n}{r^2}+1}, & (k,\ell) \notin E. \end{cases}$$

那么，很容易验证 $\{B_{k,\ell}, Q_{k,\ell}\}$ 满足定理 7) 和 (8) 中的两个条件1，只要$B_E V_E \lesssim 1$，即 $rL \gtrsim \kappa_E^5 n\log(n/\delta)$。至此，我们完成了针对一维网格 的推论 1的证明。关于二维网格 的分析非常相似；为简洁起见，我们略去细节。

本节的其余部分将用于给出引理 2 的证明。

一维网格的有效电阻Grid1D。对于一个 Grid1D$(n, r, p = 1)$ 图，我们可以通过将其转换为另一个图来得到 $\Omega_{k,\ell}(\boldsymbol{L}_z)$ 的上界，其中 $\boldsymbol{L}_z = \sum_{(i,j)\in E,\ i<j} L_{i,j} z_{i,j} (\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^\top$，该图的有效电阻可以很容易地得到控制。证明遵循以下步骤。

1. 让我们将原始一维网格中每条边 $1/(L_{i,j}z_{i,j}) = 1/(Lz_{i,j})$ 上的电阻 (i,j) 统一增大至相同的值 $4\kappa_E/L$。根据瑞利单调性定律，这一操作绝不会降低任意一对节点之间的有效电阻。接下来，为了简化记号，我们假设所有边的电阻均为 1；完成下面的分析后，只需将所得结果乘以因子 κ_E/L，即可得到关于 $\{\Omega_{k\text{的最终结果},\ \ell}\}$ 即引理 2中的结论。

2. 对于 $(k,\ \ell) \in E$，$\Omega_{k,\ell}$ 可以（同样根据单调性）由一个完全子图（包含 $k,\ \ell$）的有效电阻上界，该完全子图有 $\Theta(r)$ 个节点，根据事实 2。

3. 对于一般的 $k \neq \ell$，我们按如下方式构造一个包含 $O(n)$ 个节点（包含 $k,\ \ell$）的子图：(1) 将$O(n)$ 个节点划分为 $O(n/r)$ 个块组成的序列，每块包含 $\Theta(r)$ 个节点；(2) 保留任意一对相邻块之间由 $\Theta(r^2)$ 条边构成的完整连接，以及节点 k （分别为 ℓ）与第一个（分别为最后一个）块之间的 $\Theta(r)$ 条边。可视化结果见图3(a)。根据对称性，同一块中的所有节点彼此等价，因此可将它们视为一个单独的节点，而不影响电阻 $\Omega_{k,\ell}$。现在，我们可以利用串并联定律轻松控制 $\Omega_{k,\ell}$：

$$\Omega_{k,\ell} \lesssim \frac{1}{r} + \frac{n}{r}\cdot\frac{1}{r^2} = \left(\frac{n}{r^2}+1\right)\frac{1}{r}.$$

二维网格的有效电阻。我们对二维网格的证明与之前对一维网格的分析类似，只是在步骤 2 和 3 中做了一些修改。对于 $(k,\ell) \in E$，我们可以找到一个包含 k 和 ℓ 、具有 $\Theta(r^2)$ 个节点的完全子图，因此 $\Omega_{k,\ell} \lesssim 1/r^2$。对于一般的 $k \neq \ell$，我们按如下方式构造一个具有 $O(n)$ 个节点_(包含 k,ℓ) 的子图：(1) 将 $O(\sqrt{n}) \times O(\sqrt{n})$ 二维网格划分为大小为 $\Theta(r) \times \Theta(r)$ 的块；(2) 保留任意一对相邻块之间由 $\Theta(r^4)$ 条边构成的完整连接，以及 k/ℓ 与其最近块之间的 $\Theta(r^2)$ 条边。图 3(b) 展示了示意图。由对称性，同一块内的所有节点

都等价，因此可以将它们视为一个单独的节点，而不影响电阻$\Omega_{k,\ell}$。现在，利用有效电阻的三角不等式以及 二维网格$(m,\ r=1,\ p=1)$满足 $\Omega_{i,\ j} \lesssim \log m$ 对所有 $1 \le i < j \le m$ [CRR+96成立这一事实，根据定理 6，1], 我们有

$$\Omega_{k,\ell} \lesssim \frac{1}{r^2} + \log\left(\frac{n}{r^2}\right) \cdot \frac{1}{r^4} \lesssim \left(\frac{\log n}{r^2} + 1\right)\frac{1}{r^2}.$$

控制残差项。考虑与 $\boldsymbol{L}_z$相对应的电网络，并将电压 $\boldsymbol{v} = \boldsymbol{L}_z^{\dagger}(\boldsymbol{e}_k - \boldsymbol{e}_\ell)$ 记为某个固定的 $k \neq \ell$。回顾V_{k}的定义，ℓ (10)，可得

$$\begin{aligned}
V_{k,\ell} &= \sum_{(i,j)\in E, i<j} L_{i,j}\left|(\boldsymbol{e}_i - \boldsymbol{e}_j)^\top \boldsymbol{L}_z^{\dagger}(\boldsymbol{e}_k - \boldsymbol{e}_\ell)\right| = \sum_{(i,j)\in E, i<j} L_{i,j}|v_i - v_j| \\
&= \sum_{(i,j)\in E, i<j} \frac{L_{i,j}z_{i,j}|v_i - v_j|}{z_{i,j}} \le 4\kappa_E \sum_{(i,j)\in E, i<j} L_{i,j}z_{i,j}|v_i - v_j| \\
&= 4\kappa_E \sum_{(i,j)\in E, i<j} \sqrt{L_{i,j}z_{i,j}} \cdot \sqrt{L_{i,j}z_{i,j}}|v_i - v_j| \\
&\overset{\text{(i)}}{\le} 4\kappa_E \sqrt{\sum_{(i,j)\in E, i<j} L_{i,j}z_{i,j}} \cdot \sqrt{\sum_{(i,j)\in E, i<j} L_{i,j}z_{i,j}(v_i - v_j)^2} \\
&\le 4\kappa_E \sqrt{L\cdot|E|} \cdot \sqrt{\boldsymbol{v}^\top \boldsymbol{L}_z \boldsymbol{v}} \\
&= 4\kappa_E \sqrt{L\cdot|E|} \cdot \sqrt{(\boldsymbol{e}_k - \boldsymbol{e}_\ell)^\top \boldsymbol{L}_z^{\dagger}(\boldsymbol{e}_k - \boldsymbol{e}_\ell)} = 4\kappa_E \sqrt{L\cdot|E|} \cdot \sqrt{\Omega_{k,\ell}(\boldsymbol{L}_z)},
\end{aligned}$$

其中，(i) 根据柯西-施瓦茨不等式得出。这里，$|E| \lesssim nr$ 对于一维网格，以及 $|E| \lesssim nr^2$ 对于二维网格；此外，关于$\Omega_{k,\ell}(\boldsymbol{L}_z)$ 的上界已在我们此前的分析中得到。综合以上结果，即完成了引理 2的证明。

3.5 谱方法是否有效?

我们认为，谱方法（算法 2）在对（1）中定义的 κ_E 和 κ 的依赖性方面可能劣于最大似然估计；具体而言，尽管只要 κ_E 有界，最大似然估计就能发挥作用（无论 κ 如何），谱方法却不具备这种良好性质，尤其是在具有局部性的图中。为说明这一点，考虑一个 Grid1D(n, r, p) 图，其中 n 较大、 $r \lesssim 1$ 和 $p = 1$ 较小；此外，令 $\boldsymbol{\theta}^\star = \log \boldsymbol{\pi}^\star \in \mathbb{R}^n$ 由 $\theta_i^\star = i/r$、$1 \le i \le n$ 定义。这意味着 $\kappa_E \lesssim 1$，但对于某个通用常数 $c_0 > 0$， $\kappa \ge \exp(c_0 n)$；另外，对于所有 $1 \le i \le n/2$，都有 $\pi_i^\star \le \exp(-c_1 n)$，其中 $c_1 > 0$ 是一个通用常数。这会带来以下两个潜在问题。

1. 即使允许谱方法使用无限数量的样本来构造随机矩阵 $\boldsymbol{P}$，计算平稳分布 $\boldsymbol{\pi}^\star$ 时仍会遇到数值问题，尤其是对于 $\boldsymbol{\pi}^\star$ 中指数级小的元素；因此，估计分数 $\boldsymbol{\theta} = \log \boldsymbol{\pi}$ 可能不准确。事实上，任何试图先估计 $\boldsymbol{\pi} \approx \boldsymbol{\pi}^\star$、再计算 $\boldsymbol{\theta} = \log \boldsymbol{\pi}$ 的算法都普遍存在这个问题。[1]

2. 在有限样本这一现实情形下，为了能够对低分数对应的元素进行有意义的估计，必须确保 $\boldsymbol{\pi}$ 的逐元素误差呈指数级小；即使不考虑数值问题，能否通过多项式数量的样本实现这一点仍值得怀疑。

谱方法与最大似然估计的经验比较见第 5.1 节。

[1] 不难验证，准确的 $\boldsymbol{\theta} \approx \boldsymbol{\theta}^\star$ （就加性误差而言）意味着 $\boldsymbol{\pi} \approx \boldsymbol{\pi}^\star$ 准确，但在我们的例子中，反过来未必成立。在这项工作中，我们主要关注如何准确估计 $\boldsymbol{\theta}^\star$。

4 分治算法

在本节中，我们将设计具有理论依据的分治算法，用于估计 $\boldsymbol{\theta}^\star$ 在 一维网格 或 二维网格图中的相关量，方法是将此类图视为由若干小型 Erdős-Rényi 子图组合而成 [CKST16]。我们将证明，即使在样本最为稀疏的情况下，所提出的方法仍能保证逐元素估计误差的性能，尽管这些方法需要额外的信息（即适当的图划分）。本节中的一些思路也可能有助于对具有局部性的图中的最大似然估计进行更精细的理论分析。

补充记号。 假设给定一个图 $G = (V,\ E)$，它有 $|V| = n$ 个节点，以及对其节点 $V = V_{(1)} \cup V_{(2)} \cup \cdots \cup V_{(m)}$的一个划分，该划分可以重叠，也可以不相交。在本节中，我们在下标中使用“(a)”来索引子图，其中 $1 \le a \le m$。对于 $\boldsymbol{\theta} \in \mathbb{R}^n$，令 $\boldsymbol{\theta}|_{V_{(a)}} \in \mathbb{R}^{|V_{(a)}|}$ 表示由 $\boldsymbol{\theta}$ 在 $V_{(a)}$中的条目组成的子向量，其索引与原向量 $\boldsymbol{\theta}$中的相同；例如，如果 $V_{(a)} = \{15, 20, 25\}$，则 $\boldsymbol{\theta}|_{V_{(a)}} \in \mathbb{R}^3$ 包含 3 个条目，索引为 $\{15, 20, 25\}$，而不是 $\{1, 2, 3\}$。类似地，我们使用 $\boldsymbol{\theta}_{(a)} \in \mathbb{R}^{|V_{(a)}|}$ 和全一向量 $\mathbf{1}_{(a)} \in \mathbb{R}^{|V_{(a)}|}$的概念，它们的索引为 $V_{(a)}$，而不是$\{1, 2, \ldots, |V_{(a)}|\}$。

4.1 DC-overlap：采用重叠子图的分治法

我们提出了购彩平台 使用 DC-overlap，这是一种由局部估计和全局对齐组成的两步算法。

思路。 假设我们给定了节点的一个划分 $V = V_{(1)} \cup V_{(2)} \cup \cdots \cup V_{(m)}$；此外，每个子图$G_{(a)} = (V_{(a)},\ E_{(a)})$（其中 $E_{(a)}$ 包含 E 中端点属于 $V_{(a)}$的边）都具有某些良性~性质，例如它是一个（近似）Erdős-Rényi 图。定义一个超图 $G = (V,\ E)$，其中 $|V| = m$；~每个超节点代表一个子图，并且对于任意 $1 \le a < b \le m$，$(a,\ b) \in E$ 当且仅当 $V_{(a)} \cap V_{(b)} \neq \emptyset$。~我们假设 G 是一个连通图。

对于任意真实得分向量 $\boldsymbol{\theta}^\star$，定义 $\boldsymbol{c}^\star = [c^\star_{(a)}]_{1 \le a \le m}$，其中

$$c^\star_{(a)} := \frac{1}{|V_{(a)}|} \sum_{i \in V_{(a)}} \theta^\star_i, \quad 1 \le a \le m$$

表示 $V_{(a)}$ 的平均分数。此外，令

$$\boldsymbol{\theta}^\star_{(a)} := \boldsymbol{\theta}^\star|_{V_{(a)}} - c^\star_{(a)} \mathbf{1}_{(a)}, \quad 1 \le a \le m,$$

其满足 $\boldsymbol{\theta}^\star_{(a)} \perp \mathbf{1}_{(a)}$。因此，要估计 $\boldsymbol{\theta}^\star$，我们可以转而分别估计每个 $\boldsymbol{\theta}^\star_{(a)}$，然后通过对它们分别进行适当的平移来对齐这些估计值，从而得到对 $\boldsymbol{\theta}^\star$ 的估计值。

算法。 首先，对于每个子图 $G_{(a)} = (V_{(a)},\ E_{(a)})$，我们使用相应的数据子集 $\{y_{i,j}\}$ 来估计得分 $\boldsymbol{\theta}_{(a)} \in \mathbb{R}^{|V_{(a)}|}$；这可以在各个子图上并行完成。由于每个 $\boldsymbol{\theta}_{(a)}$ 对 $\boldsymbol{\theta}^\star_{(a)}$ 的估计值都只能精确到一个常数平移，在第二步中，我们尝试对齐 $\{\boldsymbol{\theta}_{(a)}\}_{1 \le a \le m}$，即为每个 $\boldsymbol{\theta}_{(a)}$找到适当的常数平移 $c_{(a)} \in \mathbb{R}$。其思路是确保每个项目$1 \le i \le n$ 在其所属的所有子图中具有近似相同的分数；更形式化地，令$\boldsymbol{c} = [c_{(a)}]_{1 \le a \le m}$ 为以下最小二乘问题的解：

$$\min_{\boldsymbol{c} \in \mathbb{R}^m} \sum_{(a,b) \in \widetilde{E}, a<b} \sum_{i \in V_{(a)} \cap V_{(b)}} \Big((\theta_{(a)i} + c_{(a)}) - (\theta_{(b)i} + c_{(b)}) \Big)^2.$$

通过考察其最优性条件，可以验证该解可通过求解一个拉普拉斯线性系统得到。求解完成后，我们对每个 $\boldsymbol{\theta}_{(a)} \leftarrow \boldsymbol{\theta}_{(a)} + c_{(a)} \mathbf{1}_{(a)}$ $1 \le a \le m$进行更新。对于最终解$\boldsymbol{\theta} \in \mathbb{R}^n$，我们令 θ_i 等于所有包含节点 $\{\theta_{(a)i}\}$ 的子图 $1 \le a \le m$ 中 i的平均值。DCDC-overlap 的完整过程总结于算法3。

算法 3： DC-overlap：采用重叠子图的分治

1 输入：图 $G=(V,E)$ 和子图 $\{G_{(a)}=(V_{(a)},E_{(a)})\}_{1\le a\le m}$，数据 $\{y_{i,j},L_{i,j}\}_{(i,j)\in E,i<j}$.

~ ~ ~ ~2 定义 $G=(V,E)$，使得 $|V|=m$，且 $(a,b)\in E$ 当且仅当 $V_{(a)}\cap V_{(b)}\neq\emptyset$。// 步骤 1：局部估计 3 对于每个 $1\le a\le m$，令 $\boldsymbol{\theta}_{(a)}$ 为算法 1 或 2 在输入 $\{y_{i,j},L_{i,j}\}_{(i,j)\in E_{(a)},i<j}$ 时得到的解。// 步骤 2：全局对齐 4 计算 $\boldsymbol{c}=\widetilde{\boldsymbol{L}}^{\dagger}\boldsymbol{x}\in\mathbb{R}^m$，其中

$$\widetilde{\boldsymbol{L}}=\sum_{(a,b)\in\widetilde{E},a<b}|V_{(a)}\cap V_{(b)}|(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)})(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)})^{\top},\tag{15}$$

$$\boldsymbol{x}=\sum_{(a,b)\in\widetilde{E},a<b}\sum_{i\in V_{(a)}\cap V_{(b)}}(\theta_{(b)i}-\theta_{(a)i})(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)}).\tag{16}$$

5 对于所有 $1\le i\le n$，令

$$\theta_i\leftarrow\frac{1}{s_i}\sum_{1\le a\le m:i\in V_{(a)}}(\theta_{(a)i}+c_{(a)}),\quad\text{where}\quad s_i=\big|\{1\le a\le m:i\in V_{(a)}\}\big|.$$

6 **输出：** $\boldsymbol{\theta}\in\mathbb{R}^n$。

4.2 DC-overlap 的分析

4.2.1 误差分解

我们首先给出关于 DC-overlap 误差分解的一个基本结果，该结果适用于一般图和节点划分。首先介绍一些有用的记号。记误差向量

$$\boldsymbol{\delta}=\boldsymbol{\theta}-\boldsymbol{\theta}^{\star},\quad\text{and}\quad\boldsymbol{\delta}_{(a)}=\boldsymbol{\theta}_{(a)}-\boldsymbol{\theta}^{\star}_{(a)},\quad 1\le a\le m.$$

将 $\boldsymbol{c}^{\star}$ 的平均值记为

$$\overline{c}^{\star}:=\frac{1}{m}\langle\boldsymbol{c}^{\star},\mathbf{1}_m\rangle.$$

定义一个关联矩阵

$$\widetilde{\boldsymbol{B}}:=\Big[\ldots,\sqrt{n_{(a,b)}}(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)}),\ldots\Big]_{(a,b)\in\widetilde{E},a<b},\quad\text{where}\quad n_{(a,b)}:=|V_{(a)}\cap V_{(b)}|;$$

回顾 拉普拉斯矩阵 $\widetilde{\boldsymbol{L}}$ 在（15）中定义，可以验证 $\widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{B}}^{\top}=\widetilde{\boldsymbol{L}}$。最后，定义

$$\widetilde{\boldsymbol{\epsilon}}:=\Big[\ldots,\sqrt{n_{(a,b)}}\big(\overline{\delta}^{a,b}_{(b)}-\overline{\delta}^{a,b}_{(a)}\big),\ldots\Big]^{\top}_{(a,b)\in\widetilde{E},a<b},\tag{17}$$

其中

$$\overline{\delta}^{a,b}_{(b)}=\frac{1}{n_{(a,b)}}\sum_{i\in V_{(a)}\cap V_{(b)}}\delta_{(b)i},\quad\overline{\delta}^{a,b}_{(a)}=\frac{1}{n_{(a,b)}}\sum_{i\in V_{(a)}\cap V_{(b)}}\delta_{(a)i},\quad(a,b)\in\widetilde{E}.$$

以下关于 DC-overlap 的结果刻画了对 $\boldsymbol{c}^{\star}$ 的估计误差，并将最终的 ℓ_∞ 估计误差分解为局部估计和全局对齐误差之和。

命题 2。 在算法 3中，估计 $\boldsymbol{c}\in\mathbb{R}^m$ 满足

$$\boldsymbol{c}-\boldsymbol{c}^{\star}=-\overline{c}^{\star}\mathbf{1}_m+\widetilde{\boldsymbol{L}}^{\dagger}\sum_{(a,b)\in\widetilde{E},a<b}\sum_{i\in V_{(a)}\cap V_{(b)}}(\delta_{(b)i}-\delta_{(a)i})(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)})=-\overline{c}^{\star}\mathbf{1}_m+\widetilde{\boldsymbol{L}}^{\dagger}\widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{\epsilon}}.\tag{18}$$

*Mo*此外，$\boldsymbol{\theta}\in\mathbb{R}^n$ 的估计误差可按如下方式控制：对于所有 $1\le k$、$\ell\le n$ ，

$$\big|\delta_k-\delta_\ell\big|\le\max_{a:k\in V_{(a)}}\big|\delta_{(a),k}\big|+\max_{a:\ell\in V_{(a)}}\big|\delta_{(a),\ell}\big|+\max_{a,b:k\in V_{(a)},\ell\in V_{(b)}}\big|(c_{(a)}-c^{\star}_{(a)})-(c_{(b)}-c^{\star}_{(b)})\big|.\tag{19}$$

证明。 我们首先计算 $\boldsymbol{c}-\boldsymbol{c}^\star$。回顾算法 3 可知， $\boldsymbol{c}=\widetilde{\boldsymbol{L}}^\dagger\boldsymbol{x}$，其中

$$\widetilde{\boldsymbol{L}}=\sum_{(a,b)\in\widetilde{E},a<b} n_{(a,b)}(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)})(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)})^\top,$$
$$\boldsymbol{x}=\sum_{(a,b)\in\widetilde{E},a<b}\sum_{i\in V_{(a)}\cap V_{(b)}}(\theta_{(b)i}-\theta_{(a)i})(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)}).$$

代入

$$\theta_{(a)i}=\theta^\star_{(a)i}+\delta_{(a)i}=\theta^\star_i-c^\star_{(a)}+\delta_{(a)i},$$
$$\theta_{(b)i}=\theta^\star_{(b)i}+\delta_{(b)i}=\theta^\star_i-c^\star_{(b)}+\delta_{(b)i},$$

可得

$$\begin{aligned}\boldsymbol{x}&=\sum_{(a,b)\in\widetilde{E},a<b}\sum_{i\in V_{(a)}\cap V_{(b)}}(\theta_{(b)i}-\theta_{(a)i})(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)})\\&=\sum_{(a,b)\in\widetilde{E},a<b}\sum_{i\in V_{(a)}\cap V_{(b)}}(c^\star_{(a)}-c^\star_{(b)}+\delta_{(b)i}-\delta_{(a)i})(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)})\\&=\widetilde{\boldsymbol{L}}\boldsymbol{c}^\star+\sum_{(a,b)\in\widetilde{E},a<b}\sum_{i\in V_{(a)}\cap V_{(b)}}(\delta_{(b)i}-\delta_{(a)i})(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)}).\end{aligned}$$

因此，

$$\begin{aligned}\boldsymbol{c}-\boldsymbol{c}^\star&=\widetilde{\boldsymbol{L}}^\dagger\boldsymbol{x}-\boldsymbol{c}^\star\\&=\widetilde{\boldsymbol{L}}^\dagger\widetilde{\boldsymbol{L}}\boldsymbol{c}^\star-\boldsymbol{c}^\star+\widetilde{\boldsymbol{L}}^\dagger\sum_{(a,b)\in\widetilde{E},a<b}\sum_{i\in V_{(a)}\cap V_{(b)}}(\delta_{(b)i}-\delta_{(a)i})(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)})\\&\overset{\text{(i)}}{=}-\frac{1}{m}\mathbf{1}_m\mathbf{1}_m^\top\boldsymbol{c}^\star+\widetilde{\boldsymbol{L}}^\dagger\sum_{(a,b)\in\widetilde{E},a<b}\sum_{i\in V_{(a)}\cap V_{(b)}}(\delta_{(b)i}-\delta_{(a)i})(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)})\\&=-\bar{c}^\star\mathbf{1}_m+\widetilde{\boldsymbol{L}}^\dagger\widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{\epsilon}},\end{aligned}$$

其中，(i) 利用了 $\widetilde{\boldsymbol{L}}^\dagger\widetilde{\boldsymbol{L}}=\boldsymbol{I}_m-\mathbf{1}_m\mathbf{1}_m^\top/m$ 只要 $\widetilde{G}$ 是连通的这一事实。这就证明了我们的第一个结果（18）。关于最终估计误差 $\boldsymbol{\delta}=\boldsymbol{\theta}-\boldsymbol{\theta}^\star$，回顾算法3 可知

$$\theta_k=\frac{1}{s_k}\sum_{1\le a\le m:k\in V_{(a)}}\Big(\theta_{(a),k}+c_{(a)}\Big),\quad 1\le k\le n,$$

这意味着

$$\delta_k=\theta_k-\theta^\star_k=\frac{1}{s_k}\sum_{1\le a\le m:k\in V_{(a)}}\Big(\theta_{(a),k}-\theta^\star_{(a),k}+c_{(a)}-c^\star_{(a)}\Big)=\frac{1}{s_k}\sum_{1\le a\le m:k\in V_{(a)}}\Big(\delta_{(a),k}+c_{(a)}-c^\star_{(a)}\Big).$$

该 d 所需结果（19）由对 $|\delta_k-\delta_\ell$应用三角不等式立即得到 |. □

4.2.2 局部性图的推论

命题 2将 $\boldsymbol{\theta}$ 的逐项误差分解为局部估计误差之和 $\{\delta_{(a),k}\}$，以及形式为 $\{(c_{(a)}-c^\star_{(a)})-(c_{(b)}-c^\star_{(b)})\}$的全局对齐误差。我们来具体说明一维网格（$n$， r， p）或二维网格（n， r， p）图中的这些误差项，不过以下分析的大部分内容也适用于更一般的情形。更具体地说，我们旨在证明 $\boldsymbol{\delta}=\boldsymbol{\theta}-\boldsymbol{\theta}^\star$ 对所有 $1\le k<\ell\le n$，

$$|\delta_k-\delta_\ell|\le\begin{cases}\widetilde{O}\left(\sqrt{\frac{n}{r^2}+1}\sqrt{\frac{1}{rpL}}\right)+\text{residual terms} & \text{for Grid1D}(n,r,p),\\ \widetilde{O}\left(\sqrt{\frac{\log(n)}{r^2}+1}\sqrt{\frac{1}{r^2pL}}\right)+\text{residual terms} & \text{for Grid2D}(n,r,p),\end{cases}$$

其中，“残差项”的含义将在我们的分析中变得清晰。

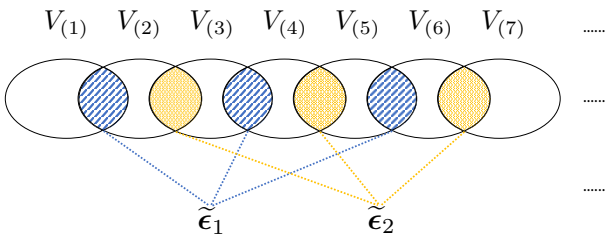


图4：一维网格的分解 $\widetilde{\boldsymbol{\epsilon}}_{>} = \widetilde{\boldsymbol{\epsilon}}_1 + \widetilde{\boldsymbol{\epsilon}}_2$，使每一项都包含相互独立的条目。

准备工作。 假设划分 $V = V_{(1)} \cup \cdots \cup V_{(m)}$ 和诱导出的超图 $G = (V, E)$ 满足以下条件：

1. 每个子图都是 Erdős-Rényi 图；对于一维网格，包含 $\Theta(r)$ 个节点；对于二维网格，包含 $\Theta(r^2)$ 个节点；
2. For each $(a, b) \in \widetilde{E}$, it holds that $n_{(a,b)} = |V_{(a)} \cap V_{(b)}| \asymp r$ for Grid1D, or $n_{(a,b)} \asymp r^2$ for Grid2D;
3. G 的每个节点出现在 $O(1)$ 个子图中。

只要按照定义 1 中所述使用正确的节点索引，就很容易实现这样的划分：对于一维网格，每个子图由 $\Theta(r)$ 个连续节点组成；对于二维网格，每个子图由二维格子上的一个 $\Theta(r) \times \Theta(r)$ 块组成。我们进一步假设 $\kappa_E \lesssim 1$, $L_{i,j} = L \geq 1$，且对于一维网格，$rp \geq \Omega(1)$；对于二维网格，$r^2 p \geq \Omega(1)$，这对应于最稀疏区域。

我们分析的一个关键要素是以下关于 Erdős-Rényi 图中最大似然估计不确定性量化的结果 [GSZ21，定理 2。2]；下文中，我们回顾一下 $\boldsymbol{A}$ 表示一个（随机）图的邻接矩阵。

引理3 ([GSZ21])。考虑一个具有 N 个节点、边采样概率为 p，且每条边上有 L 次测量的 *Erdős-Rényi* 图。假设 $\kappa \lesssim 1$ 且 $Np \gg \log^{1.5} N$。那么，式（2）的最大似然估计解 $\boldsymbol{\theta}^{\mathsf{MLE}}$ 满足

$$\boldsymbol{\theta}^{\mathsf{MLE}} - \boldsymbol{\theta}^{\star} = \boldsymbol{\delta}^{\mathsf{main}} + \boldsymbol{\delta}^{\mathsf{res}}, \tag{20}$$

其中，主要误差项 $\boldsymbol{\delta}^{\mathsf{main}}$ 定义为

$$\boldsymbol{\delta}_i^{\mathsf{main}} = \frac{\sum_{j \neq i} A_{i,j} \big(y_{i,j} - \sigma(\theta_i^{\star} - \theta_j^{\star})\big)}{\sum_{j \neq i} A_{i,j} \sigma'(\theta_i^{\star} - \theta_j^{\star})}, \quad 1 \leq i \leq N,$$

且残差项 $\boldsymbol{\delta}^{\mathsf{res}}$ 以概率 $1 - O(N^{-10})$ 满足

$$\|\boldsymbol{\delta}^{\mathsf{res}}\|_{\infty} = o\bigg(\frac{1}{\sqrt{NpL}}\bigg).$$

注释 4. 对于算法3中的局部估计，可以使用谱方法代替最大似然估计（MLE）。在这种情况下进行理论分析时，只需将引理 [GSZ21，定理 2。5], 即谱方法的不确定性量化结果（其效果略逊于最大似然估计）。

全局对齐的误差。（19）中的局部误差项可以通过引理3轻松控制，而全局对齐的误差则需要进一步分析。回顾（18），可知 $\boldsymbol{c} - \boldsymbol{c}^{\star} = -c^{\star} \mathbf{1}_m + \widetilde{\boldsymbol{L}}^{\dagger} \widetilde{\boldsymbol{B}} \widetilde{\boldsymbol{\epsilon}}$，其中$\widetilde{\boldsymbol{\epsilon}}$ 定义于（17）中。尽管 $\widetilde{\boldsymbol{\epsilon}}$ 的各个元素彼此相关，但很容易进行如下分解

$$\widetilde{\boldsymbol{\epsilon}} = \sum_{f=1}^{F} \widetilde{\boldsymbol{\epsilon}}_f,$$

s使得 $F \lesssim 1$，且每个 $\widetilde{\boldsymbol{\epsilon}}_f$ 都具有相互独立的条目。为此，我们首先改写

$$\widetilde{\boldsymbol{\epsilon}} = \widetilde{\boldsymbol{\epsilon}}_{>} - \widetilde{\boldsymbol{\epsilon}}_{<}, \quad \text{where} \quad \widetilde{\boldsymbol{\epsilon}}_{>} \coloneqq \Big[\ldots, \sqrt{n_{(a,b)}}\, \overline{\delta}_{(b)}^{a,b}, \ldots\Big]^{\top}_{(a,b) \in \widetilde{E}, a<b}, \quad \widetilde{\boldsymbol{\epsilon}}_{<} = \widetilde{\boldsymbol{\epsilon}} \coloneqq \Big[\ldots, \sqrt{n_{(a,b)}}\, \overline{\delta}_{(a)}^{a,b}, \ldots\Big]^{\top}_{(a,b) \in \widetilde{E}, a<b}.$$

接着，我们进一步将每个 $\widetilde{\boldsymbol{\epsilon}}_{>}$ 和 $\widetilde{\boldsymbol{\epsilon}}_{<}$ 分解为 $O(1)$ 项之和：将 $\widetilde{\boldsymbol{\epsilon}}_{>}$(或 $\widetilde{\boldsymbol{\epsilon}}_{<}$) 的每个条目分配给其中一项，并将这些项中的其余条目设为零。此方法基于一个关键观察：每个 $\delta^{a}_{,\ b\ (b)}$ 在统计上仅依赖于数据子集 $\{A_{i,j}, y_{i,j}\}_{i,\ j\in V_{(b)}}$。例如，一维网格的这种分解如图4所示。

有了这一分解之后，我们可以将其重写为

$$\boldsymbol{c} - \boldsymbol{c}^{\star} = -\overline{c}^{\star}\mathbf{1}_m + \sum_{f=1}^{F} \widetilde{\boldsymbol{L}}^{\dagger}\widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{\epsilon}}_f.$$

请注意，

$$(c_{(a)} - c^{\star}_{(a)}) - (c_{(b)} - c^{\star}_{(b)}) = \big\langle \boldsymbol{e}_{(a)} - \boldsymbol{e}_{(b)}, \sum_{f=1}^{F} \widetilde{\boldsymbol{L}}^{\dagger}\widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{\epsilon}}_f \big\rangle, \quad 1 \le a < b \le m.$$

在接下来的分析中，我们将重点关注“主要”误差项。首先从分解开始。

$$\boldsymbol{\delta}_{(a)} = \boldsymbol{\delta}^{\mathsf{main}}_{(a)} + \boldsymbol{\delta}^{\mathsf{res}}_{(a)}, \quad 1 \le a \le m,$$

这对应于引理20中的分解（3）。类似地，我们改写

$$\widetilde{\boldsymbol{\epsilon}} = \widetilde{\boldsymbol{\epsilon}}^{\mathsf{main}} + \widetilde{\boldsymbol{\epsilon}}^{\mathsf{res}}, \quad \text{and} \quad \widetilde{\boldsymbol{\epsilon}}_f = \widetilde{\boldsymbol{\epsilon}}^{\mathsf{main}}_f + \widetilde{\boldsymbol{\epsilon}}^{\mathsf{res}}_f, \quad 1 \le f \le F.$$

假设 $\widetilde{\boldsymbol{\epsilon}}^{\mathsf{main}}_f$ 的各个元素相互独立、均值为零，且次高斯范数为 $\|\cdot\|_{\psi_2} \le \sigma_{\mathsf{main}}$，则有

$$\big\|\langle \boldsymbol{e}_{(a)} - \boldsymbol{e}_{(b)}, \widetilde{\boldsymbol{L}}^{\dagger}\widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{\epsilon}}^{\mathsf{main}}_f\rangle\big\|_{\psi_2} \le \sigma_{\mathsf{main}}\big\|\widetilde{\boldsymbol{B}}^{\top}\widetilde{\boldsymbol{L}}^{\dagger}(\boldsymbol{e}_{(a)} - \boldsymbol{e}_{(b)})\big\|_2 = \sigma_{\mathsf{main}}\sqrt{\Omega_{a,b}(\widetilde{\boldsymbol{L}})}, \quad 1 \le a < b \le m.$$

根据标准的集中性论证，以高概率有

$$\big|(c_{(a)} - c^{\star}_{(a)}) - (c_{(b)} - c^{\star}_{(b)})\big| \le \widetilde{O}\big(\sigma_{\mathsf{main}}\sqrt{\Omega_{a,b}(\widetilde{\boldsymbol{L}})}\big) + \text{residual terms}, \quad 1 \le a < b \le m. \tag{21}$$

接下来需要为 $\widetilde{\boldsymbol{\epsilon}}^{\mathsf{main}}_f$ 的次高斯范数寻找一个有效的上界 σ_{main}，并控制与拉普拉斯矩阵 $\widetilde{\boldsymbol{L}} = \sum_{(a,b)\in\widetilde{E},\ a<b} n_{(a,b)}(\boldsymbol{e}_{(a)} - \boldsymbol{e}_{(b)})(\boldsymbol{e}_{(a)} - \boldsymbol{e}_{(b)})^{\top}$ 相对应的有效电阻。

特殊情况 se: 一维网格. 聚焦于主要误差项，我们得到以下近似：

$$\overline{\delta}^{a,b}_{(b)} = \frac{1}{n_{(a,b)}} \sum_{i\in V_{(a)}\cap V_{(b)}} \delta_{(b)i} \approx \frac{1}{n_{(a,b)}} \sum_{i\in V_{(a)}\cap V_{(b)}} \delta^{\mathsf{main}}_{(b)i} = \frac{1}{n_{(a,b)}} \sum_{i\in V_{(a)}\cap V_{(b)}} \frac{\sum_{j\neq i, j\in V_{(b)}} A_{i,j}\epsilon_{i,j}}{\sum_{j\neq i, j\in V_{(b)}} A_{i,j}z_{i,j}} =: \overline{\delta}^{a,b,\mathsf{main}}_{(b)},$$

其中，$z_{i,\ j} = \sigma'(\theta^{\star}_i - \theta^{\star}_j) \asymp 1$，而 $\epsilon_{i,\ j} = y_{i,\ j} - \sigma(\theta^{\star}_i - \theta^{\star}_j)$ 是一个满足 $\|\epsilon_{i,\ j}\|_{\psi_2} \lesssim \sqrt{1/L}$ 的零均值噪声项。此外，我们关于 $rp \ge \Omega(1)$ 的假设意味着，以高概率有 $\sum_{j\neq i,\ j\in V_{(b)}} A_{i,\ j} \asymp rp$。在此条件下，容易验证 $\widetilde{\boldsymbol{\epsilon}}^{\mathsf{main}}_f$ 的每个元素的次高斯范数均有界于 $\sigma_{\mathsf{main}} \lesssim 1/\sqrt{rpL}$。关于式 (21) 中的 $\Omega_{a,\ b}(\boldsymbol{L})$ 项，注意到超图 G 是一个长度为 $O(n/r)$ 的线图；此外，回顾假设 $n_{(a,b)} \asymp r$。因此，根据串并联定律，

$$\Omega_{a,b}(\widetilde{\boldsymbol{L}}) \lesssim \frac{n}{r}\cdot\frac{1}{r} \lesssim \frac{n}{r^2}.$$

将这些结果以及局部误差界 $O(1/\sqrt{rpL})$ 代入（19），我们得到

$$|\delta_k - \delta_\ell| \le \widetilde{O}\left(\sqrt{\frac{n}{r^2}+1}\sqrt{\frac{1}{rpL}}\right) + \text{residual terms}, \quad 1 \le k < \ell \le n.$$

特殊情况：二维网格。对二维网格的分析遵循相同步骤。首先，可以证明 $\widetilde{\epsilon}_f^{\mathsf{main}}$的每个条目都具有有界的次高斯范数 $\sigma_{\mathsf{main}} \lesssim 1/\sqrt{r^2pL}$。此外，注意到超图 $\widetilde{G}$ 是一个大小为 $O(\sqrt[r]{n} \times \sqrt[r]{n})$的二维格子；另外，回想一下假设 $n_{(a,b)} \asymp r^2$。因此，根据并联定律和二维格子的有效电阻 [CRR+96，定理 6。1], 可得

$$\Omega_{a,b}(\widetilde{\boldsymbol{L}}) \lesssim \log\left(\frac{\sqrt{n}}{r}\right) \cdot \frac{1}{r^2} \lesssim \frac{\log(n)}{r^2}.$$

将这些结果以及局部误差界 $\widetilde{O}(1/\sqrt{r^2pL})$代入（19），我们得到 at

$$|\delta_k - \delta_\ell| \le \widetilde{O}\left(\sqrt{\frac{\log(n)}{r^2}} + 1\sqrt{\frac{1}{r^2pL}}\right) + \text{residual terms}, \quad 1 \le k < \ell \le n.$$

关于“主要”误差项和“残差”误差项的注意事项。在 Erdős-Rényi 图中，最大似然估计解的主要误差项占主导地位，而残差项则相对较小。然而，对于具有强局部性的图，情况并非总是如此。为此，考虑一个最简单的例子：线图，即一维网格$(n, r = 1, p = 1)$。最大似然估计解 $\boldsymbol{\theta}^{\mathsf{MLE}}$ 有一个闭式表达：

$$\theta_{k+1}^{\mathsf{MLE}} - \theta_k^{\mathsf{MLE}} = \sigma^{-1}(y_{k+1,k}), \quad 1 \le k \le n-1,$$

其中 σ^{-1} 是 Sigmoid 函数的反函数。照常将误差记作 $\boldsymbol{\delta} = \boldsymbol{\theta}^{\mathsf{MLE}} - \boldsymbol{\theta}^\star$。回想一下，$y_{k+1,k} = \sigma(\theta_{k+1}^\star - \theta_k^\star) + \epsilon_{k+1,k}$，其中 $\epsilon_{k+1,k}$ 是一个零均值噪声项，满足 $\|\epsilon_{k+1,k}\|_{\psi_2} \asymp 1$。此外，$\sigma^{-1}$ 在 $(0, 0.5]$ 上是凹函数，在 $[0.5, 1$ 上是凸函数。考虑 $\theta_k^\star = k, 1 \le k \le n$ 的情况；根据泰勒级数展开，$\delta_{k+1} - \delta_k$ 可以分解为一个 $O(1/\sqrt{L})$ 零均值噪声项（主要项），以及一个 $O(1/L)$ 偏差项（残差）。通过线图，零均值项以 $\sqrt{n}$ 的速率累积（这是由于独立性和集中性），而偏差项则随 n 线性累积。因此，有】【。

$$|\delta_n - \delta_1| \asymp \sqrt{\frac{n}{L}} + \frac{n}{L}.$$

这表明，尽管残差项的量级小于主项，但它们的和最终仍可能主导线图中的整体 ℓ_∞ 估计误差，除非 $L \gtrsim n$。虽然这一结果可由我们的定理 1推出，但对于“残差”项在更一般的具有局部性的图中的影响，尚缺乏正式分析。

4.3 通过投影梯度下降求解最大似然估计

受 DC-overlap 启发，我们设计了一种在具有局部性的图中求解最大似然估计的新方法，该方法在数值上具有快速收敛的特点，并有望启发对最大似然估计开展进一步的理论分析。

要素 1：重新参数化。给定节点的一个划分 $V = V_{(1)} \cup \cdots \cup V_{(m)}$，按照 DC-overlap 中相同的方式定义子图 $G_{(a)} = (V_{(a)}, E_{(a)})$ $1 \le a \le m$ 和超图 $G = (V, E)$。假设这些子图覆盖原图中的所有边，即 $E_{(1)} \cup \cdots \cup E_{(m)} = E$。我们提出使用重新参数化 $\boldsymbol{\phi} = \{\boldsymbol{\theta}_{(a)} \in \mathbb{R}^{|V_{(a)}|}\}_{1 \le a \le m}$ 来求解最大似然估计。定义约束集

$$\mathcal{C} = \left\{\boldsymbol{\phi} = \{\boldsymbol{\theta}_{(a)} \in \mathbb{R}^{|V_{(a)}|}\}_{1 \le a \le m} : \theta_{(a)i} = \theta_{(b)i} \text{ for all } (a,b) \in \widetilde{E} \text{ and } i \in V_{(a)} \cap V_{(b)}\right\}.$$

Wi将这些值固定后，最小化最大似然估计损失函数 $\mathcal{L}(\boldsymbol{\theta})$ 定义于（2）等价于 o

$$\min_{\boldsymbol{\phi} = \{\boldsymbol{\theta}_{(a)}\} \in \mathcal{C}} \quad \sum_{1 \le a \le m} \mathcal{L}_{(a)}(\boldsymbol{\theta}_{(a)}),$$

其中

$$\mathcal{L}_{(a)}(\boldsymbol{\theta}_{(a)}) = \sum_{(i,j) \in E_{(a)}, i<j} w_{i,j} L_{i,j}\Big(- y_{i,j}(\theta_{(a)i} - \theta_{(a)j}) + \log(1 + e^{\theta_{(a)i} - \theta_{(a)j}})\Big).$$

请注意，$\mathcal{L}_{(a)}(\boldsymbol{\theta}_{(a)})$与定义在子图 $G_{(a)}$上的最大似然估计损失相同，只是增加了额外的权重 $\{w_{i,j}\}$；这些权重经过选择，以确保对于任意 $\boldsymbol{\theta} \in \mathbb{R}^n$，$\mathcal{L}(\boldsymbol{\theta}) = \sum_{1 \le a \le m} \mathcal{L}_{(a)}(\boldsymbol{\theta}|_{V_{(a)}})$。

成分 2：投影梯度下降（投影梯度下降）。现在，给定任意 $\boldsymbol{\theta} \in \mathbb{R}^n$，我们进行重新参数化 $\boldsymbol{\phi} = \{\boldsymbol{\theta}_{(a)} = \boldsymbol{\theta}|_{V_{(a)}}\}_{1\le a\le m}$，并在 $\boldsymbol{\phi}$ 上执行一步梯度下降，以获得 $\{\boldsymbol{\theta}'_{(a)}\}$；这可以通过分别并行地对每个损失函数 $\mathcal{L}_{(a)}(\boldsymbol{\theta}_{(a)})$ 执行一步梯度下降来完成。下一步是将 $\{\boldsymbol{\theta}'_{(a)}\}$ 投影回约束集 $\mathcal{C}$。如果我们直接使用通常的欧几里得范数作为距离度量，那么投影到 $\mathcal{C}$ 上就等价于求解以下问题

$$\min_{\boldsymbol{\theta}''\in\mathbb{R}^n} \quad \sum_{1\le i\le n}\sum_{1\le a\le m: i\in V_{(a)}} (\theta'_{(a)i} - \theta''_i)^2.$$

解 $\boldsymbol{\theta}''$ 满足每个 θ''_i 都是 $\{\theta'_{(a)i}\}_{1\le a\le m: i\in V_{(a)}}$ 的简单平均值；这样产生的损失最小。

$$\sum_{1\le i\le n}\sum_{1\le a\le m: i\in V_{(a)}} \left(\theta'_{(a)i} - \frac{1}{s_i}\sum_{1\le b\le m: i\in V_{(b)}} \theta'_{(b)i}\right)^2,$$

其中 $s_i = |\{1 \le a \le m :\ i \in V_{(a)}\}|$，$1 \le i \le n$。

为了加速收敛，我们提出以不同的方式进行投影。关键观察是，对于损失函数 $\mathcal{L}_{(a)}$，$\boldsymbol{\theta}_{(a)}$ 对于任意 $c_{(a)} \in \mathbb{R}$都等价于 $\boldsymbol{\theta}_{(a)} + c_{(a)}\mathbf{1}_{(a)}$。因此，更好的投影方式是在$\boldsymbol{\theta}'' \in \mathbb{R}^n$ 和 $\boldsymbol{c} \in \mathbb{R}^m$ 上求解联合最小化问题。给定任意 $\boldsymbol{c}$，通过对每个条目取简单平均即可得到最优的 $\boldsymbol{\theta}''$；因此，问题归结为如下优化 $\boldsymbol{c}$：

$$\min_{\boldsymbol{c}\in\mathbb{R}^m} \quad f(\boldsymbol{c}) = \sum_{1\le i\le n}\sum_{1\le a\le m: i\in V_{(a)}} \left(\theta'_{(a)i} + c_{(a)} - \frac{1}{s_i}\sum_{1\le b\le m: i\in V_{(b)}} (\theta'_{(b)i} + c_{(b)})\right)^2.$$

利用 $\sum_{1\le i\le n}(x_i - x)^2 = \frac{1}{n}\sum_{1\le i<j\le n}(x_i - x_j)^2$（其中 $x = \frac{1}{n}\sum_{1\le i\le n} x_i$）这一事实，此损失函数可重写为

$$\begin{aligned} f(\boldsymbol{c}) &= \sum_{1\le i\le n}\frac{1}{s_i}\sum_{1\le a<b\le m: i\in V_{(a)}\cap V_{(b)}} \left((\theta'_{(a)i} + c_{(a)}) - (\theta'_{(b)i} + c_{(b)})\right)^2 \\ &= \sum_{(a,b)\in\widetilde{E}, a<b}\ \sum_{i\in V_{(a)}\cap V_{(b)}} \frac{1}{s_i}\left((\theta'_{(a)i} + c_{(a)}) - (\theta'_{(b)i} + c_{(b)})\right)^2. \end{aligned}$$

因此，这就转化为一个加权最小二乘问题；同样，最优的 $\boldsymbol{c}$ 可以通过求解一个拉普拉斯线性系统来求得，这与我们之前分析 DC-overlap 情形时的方法类似。

算法4 总结了通过重新参数化和投影梯度下降求解最大似然估计的所提方法。

注释 5. 求解最大似然估计（MLE）的常见一阶迭代方法（例如梯度下降、坐标下降、下界最大化 [Hun04, VYZ22]）在具有局部性的图中注定会遭遇收敛缓慢的问题，因为它们在本质上是去中心化的。更具体地说，在每次迭代中，估计值 $\boldsymbol{\theta}$ 的第 i 个分量仅使用其相邻分量 $\{\theta_j\}_{j\in\mathcal{N}(i)}$ 进行更新；因此，如果图具有局部性，一个分量的变化需要经过许多次迭代才能传播到整个图中。PrecondGD 方法（参见注释 3）通过一种能够适应 $\mathcal{L}(\boldsymbol{\theta})$ 优化景观的预条件器来加速收敛。相比之下，本节提出的 PGD 方法专为具有局部性的图而设计，通过在子图之间进行显式全局对齐来加速收敛；与 PrecondGD 相比，该方法需要求解的拉普拉斯线性系统规模更小。关于求解最大似然估计的各种方法的实证比较，请参见第 5.3 节。

4.4 DC-community：利用社区进行分治

算法。DC-community 是另一种分治算法，适用于由多个社区组成的图（包括一维网格和二维网格）。它遵循类似的两步流程，即局部估计和全局对齐。

假设给定一个图 $G = (V, E)$ 和节点的不相交划分 $V = V_{(1)} \cup \cdots \cup V_{(m)}$；每个 $V_{(a)}$ 都可以视为一个社区。定义超图 $\widetilde{G} = (V, \widetilde{E})$，其中 $|V| = m$，且 $(a, b) \in \widetilde{E}$当且仅当存在连接 $V_{(a)}$ 和 $V_{(b)}$ 的边，即 $(V_{(a)} \times V_{(b)}) \cap E \neq \emptyset$。与 DC-overlap 类似，

算法4： 用于求解最大似然估计的投影梯度下降（PGD）

输入： 图 $G=(V,\ E)$ 和子图 $\{G_{(a)}=(V_{(a)},\ E_{(a)})\}_{1\le a\le m}$，数据 $\{y_{i,j},L_{i,j}\}_{(i,j)\in E,\ i<j}$，初始化 $\boldsymbol{\theta}^0\in\mathbb{R}^n$，步长 η，迭代次数 T。~ ~ ~ ~ 定义 $G=(V,E)$，使得 $|V|=m$，且 $(a,\ b)\in E$ 当且仅当 $V_{(a)}\cap V_{(b)}\neq\emptyset$。按如下方式计算 $\{s_i\}$，$\{w_{i,j}\}$：

$$s_i\leftarrow\big|\{1\le a\le m: i\in V_{(a)}\}\big|,\quad 1\le i\le n;\quad w_{i,j}\leftarrow\frac{1}{\big|\{1\le a\le m:(i,j)\in E_{(a)}\}\big|},\quad (i,j)\in E.$$

定义图拉普拉斯矩阵

$$\widetilde{\boldsymbol{L}}=\sum_{(a,b)\in\widetilde{E},a<b}\Big(\sum_{i\in V_{(a)}\cap V_{(b)}}\frac{1}{s_i}\Big)(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)})(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)})^\top.$$

对于每个 $1\le a\le m$，定义损失函数

$$\mathcal{L}_{(a)}(\boldsymbol{\theta}_{(a)})=\sum_{(i,j)\in E_{(a)},i<j}w_{i,j}L_{i,j}\Big(-y_{i,j}(\theta_{(a)i}-\theta_{(a)j})+\log(1+e^{\theta_{(a)i}-\theta_{(a)j}})\Big).$$

// 投影梯度下降迭代

对于 $t=0,\ 1,\ \ldots,\ T-1$ **执行**

- 对于每个 $1\le a\le m$，令 $\boldsymbol{\theta}_{(a)}\leftarrow\boldsymbol{\theta}^t|_{V_{(a)}}$ 和 $\boldsymbol{\theta}'_{(a)}\leftarrow\boldsymbol{\theta}_{(a)}-\eta\nabla\mathcal{L}_{(a)}(\boldsymbol{\theta}_{(a)})$。
- 计算 $\boldsymbol{c}=\quad\widetilde{\boldsymbol{L}}^\dagger\boldsymbol{x}\in\mathbb{R}^m$，其中

$$\boldsymbol{x}=\sum_{(a,b)\in\widetilde{E},a<b}\sum_{i\in V_{(a)}\cap V_{(b)}}\frac{1}{s_i}(\theta'_{(b)i}-\theta'_{(a)i})(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)}).$$

- 对于所有 $1\le i\le n$，令 $\theta_i^{t+1}\leftarrow\frac{1}{s_i}\sum_{1\le a\le m:i\in V_{(a)}}(\theta'_{(a)i}+c_{(a)})$。

输出： $\boldsymbol{\theta}^T\in\mathbb{R}^n$。

我们首先为每个子图获得局部估计 $\{\boldsymbol{\theta}_{(a)}\}_{1\le a\le m}$。接下来需要通过一些常数偏移 $\boldsymbol{c}=[c_{(a)}]_{1\le a\le m}$ 实现全局对齐。为此，我们先通过求解下式，估计一对子图 $(a,b)\in E$ 之间的偏移差 $\Delta_{(a,b)}$：

$$\sum_{(i,j)\in(V_{(a)}\times V_{(b)})\cap E}L_{i,j}\Big(\sigma(\theta_{(a)i}-\theta_{(b)j}+\Delta_{(a,b)})-y_{i,j}\Big)=0;$$

这可以通过多种方法求解，例如牛顿法或二分查找。等价地，选择 $\Delta_{(a,b)}$ 以最大化边 $V_{(a)}$ 和 $V_{(b)}$ 之间的样本 $\{y_{i,j}\}$ 的似然。现在，将 $\{\Delta_{(a,b)}\}$ 作为 $\{c_{(a)}-c_{(b)}\}$ 的替代，我们通过求解以下带有给定权重 $\{w_{(a,b)}\}$ 的加权最小二乘问题来求得 $\boldsymbol{c}$（例如 $w_{(a,b)}=1$，或 $w_{(a,b)}=|(V_{(a)}\times V_{(b)})\cap E|$）：

$$\min_{\boldsymbol{c}\in\mathbb{R}^m}\ \sum_{(a,b)\in\widetilde{E},a<b}w_{(a,b)}\big(c_{(a)}-c_{(b)}-\Delta_{(a,b)}\big)^2.$$

再次，最优 $\boldsymbol{c}$ 是一个拉普拉斯线性系统的解。最后，对于每个 $1\le a\le m$ 和 $i\in V_{(a)}$，令 $\theta_i=\theta_{(a)i}+c_{(a)}$；由于 $\{V_{(a)}\}$ 两两不相交，这已经给出了一个解 $\boldsymbol{\theta}\in\mathbb{R}^n$。DCDC-社区 的完整过程总结在算法5中。

分析。 对 DC-community 进行理论分析的关键步骤，是证明每个 $\Delta_{(a,b)}\in\mathbb{R}$ 都近似线性地依赖于 $\boldsymbol{\theta}_{(a)}$ 和 $\boldsymbol{\theta}_{(b)}$ 的局部估计误差。其余分析所需的要素包括

算法 5： DC-community：按社区进行分治

1 输入：图 $G=(V,E)$ 和子图 $\{G_{(a)}=(V_{(a)},E_{(a)})\}_{1\le a\le m}$，权重 $\{w_{(a,b)}\}$，数据$\{y_{i,j},L_{i,j}\}_{(i,j)\in E,i<j}$.

2 定义 $G=(V,E)$，使得 $|V|=m$，并且 $(a,b)\in E$ 当且仅当 $(V_{(a)}\times V_{(b)})\cap E\neq\emptyset$。// 步骤 1：局部估计 3 对于每个 $1\le a\le m$，令 $\boldsymbol{\theta}_{(a)}$ 为算法 1 或 2 在输入 $\{y_{i,j},L_{i,j}\}_{(i,j)\in E_{(a)},i<j}$ 下得到的解。// 步骤 2：全局对齐4 对于每个 $(a,b)\in E,a<b$，令 $\Delta_{(a,b)}$ 为下列方程的解：

$$\sum_{(i,j)\in(V_{(a)}\times V_{(b)})\cap E} L_{i,j}\Big(\sigma(\theta_{(a)i}-\theta_{(b)j}+\Delta_{(a,b)})-y_{i,j}\Big)=0.$$

5 令 $\boldsymbol{c}=\widetilde{\boldsymbol{L}}^{\dagger}\boldsymbol{x}\in\mathbb{R}^m$，其中

$$\widetilde{\boldsymbol{L}}=\sum_{(a,b)\in\widetilde{E},a<b} w_{(a,b)}(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)})(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)})^{\top},$$
$$\boldsymbol{x}=\sum_{(a,b)\in\widetilde{E},a<b} w_{(a,b)}\Delta_{(a,b)}(\boldsymbol{e}_{(a)}-\boldsymbol{e}_{(b)}).$$

6 定义 $\boldsymbol{\theta}\in\mathbb{R}^n$ 如下：对于每个 $1\le a\le m$，令 $\theta_i\leftarrow\theta_{(a)i}+c_{(a)}$ 对所有 $i\in V_{(a)}$成立。

7 **输出：** $\boldsymbol{\theta}\in\mathbb{R}^n$。

与 DC-overlap 的内容相同，即 Erdős-Rényi 图中最大似然估计的不确定性量化，以及拉普拉斯系统与有效电阻之间的联系。为简洁起见，我们略去细节。

5 数值实验

在本节中，我们进行数值实验，以验证我们的理论并展示所提出算法的有效性。

5.1 最大似然估计与谱方法

我们比较经典算法，即最大似然估计和谱方法，在一维网格$(n,\ r,\ p)$图中每条边具有L 次测量的情形。固定 $r=10,\ p=0.8,\ L=100$，并令 n 在 60 到 780 之间变化。对于真实分数 $\boldsymbol{\theta}^\star$，考虑以下两种设置：

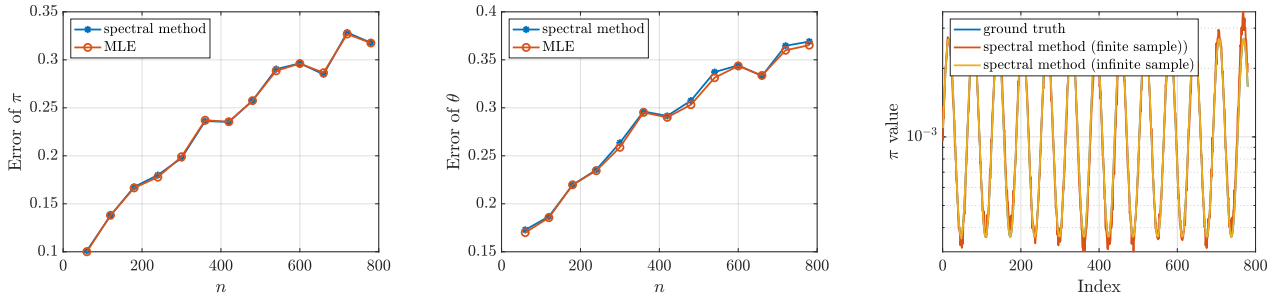
- “正弦”设置，其中 $\theta_i^\star=\sin(i/r)$ $1\le i\le n$；这意味着 $\kappa_E,\ \kappa\lesssim 1$。
- “线性”设置，其中 $\theta_i^\star=i/r$；这意味着 $\kappa_E\lesssim 1$，而 κ 随 n呈指数增长。

数值结果汇总于图5中。在“正弦”设置下，两种算法的表现相近。然而，在“线性”设置下，谱方法能够准确估计 $\boldsymbol{\pi}^\star$，但在估计 ℓ_∞ 时会产生随之线性增长的 $\boldsymbol{\theta}^\star$误差，如图5(e)所示。图5(f)表明，这是因为即使在样本数无限的情况下，谱方法也无法通过数值计算得到某些节点上极小的平稳概率。这证实了在 κ_E 较小而 κ 可能非常大的情况下，最大似然估计优于谱方法。

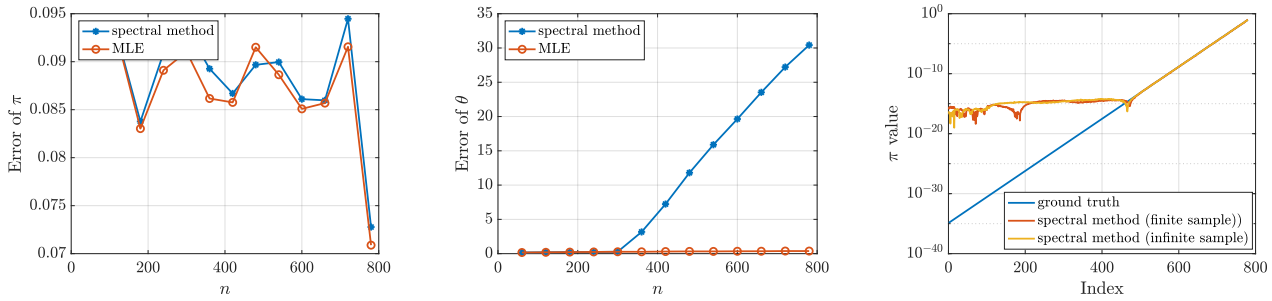
5.2 最大似然估计和 DC-overlap

We 比较最大似然估计和 DC-overlap 所达到的 ℓ_∞ 误差。考虑两种设置：

- 对于。Grid1D(n,r,p), we set $\theta_i^\star=i/r, 1\le i\le n$, and the theory bound $5\sqrt{n/r^2+1}\sqrt{1/rpL}$



(a) “正弦”， $\|\boldsymbol{\pi}-\boldsymbol{\pi}^\star\|_\infty/\|\boldsymbol{\pi}^\star\|_\infty$ (b) “正弦”， $\|\boldsymbol{\theta}-\boldsymbol{\theta}^\star\|_\infty$ (c) “正弦”， $\boldsymbol{\pi}^\star$ 的估计



(d) “线性”， $\|\boldsymbol{\pi}-\boldsymbol{\pi}^\star\|_\infty/\|\boldsymbol{\pi}^\star\|_\infty$ (e) “线性”， $\|\boldsymbol{\theta}-\boldsymbol{\theta}^\star\|_\infty$ (f) “线性”， $\boldsymbol{\pi}^\star$ 的估计

图 5：最大似然估计与谱方法的数值比较。第一行采用“正弦”设置，第二行采用“线性”设置。测量误差时，我们确保 $\boldsymbol{\theta}$ 和 $\boldsymbol{\theta}^\star$ 均值为零，且 $\boldsymbol{\pi}$ 和 $\boldsymbol{\pi}^\star$ 均为概率向量。前两列中的每条曲线都是 30 次独立试验的平均值。

- 对于二维网格$(n,\ r,\ p)$，我们设定 $\theta_{\boldsymbol{i}}^\star=(i_1+i_2)/r,\ 1\le i_1,\ i_2\le\sqrt{n}$（其中 $\boldsymbol{i}=(i_1,\ i_2)$ 是每个节点的二维索引，如定义 1所述），理论界为 $6\sqrt{\log(n)/r^2+1}\sqrt{1/r^2pL}$。

图 6 总结了 ℓ_∞ 误差随各种参数的变化情况。结果证实，最大似然估计和DC-重叠 的性能非常接近，并且与理论界相当吻合。

5.3 使用各种方法求解最大似然估计

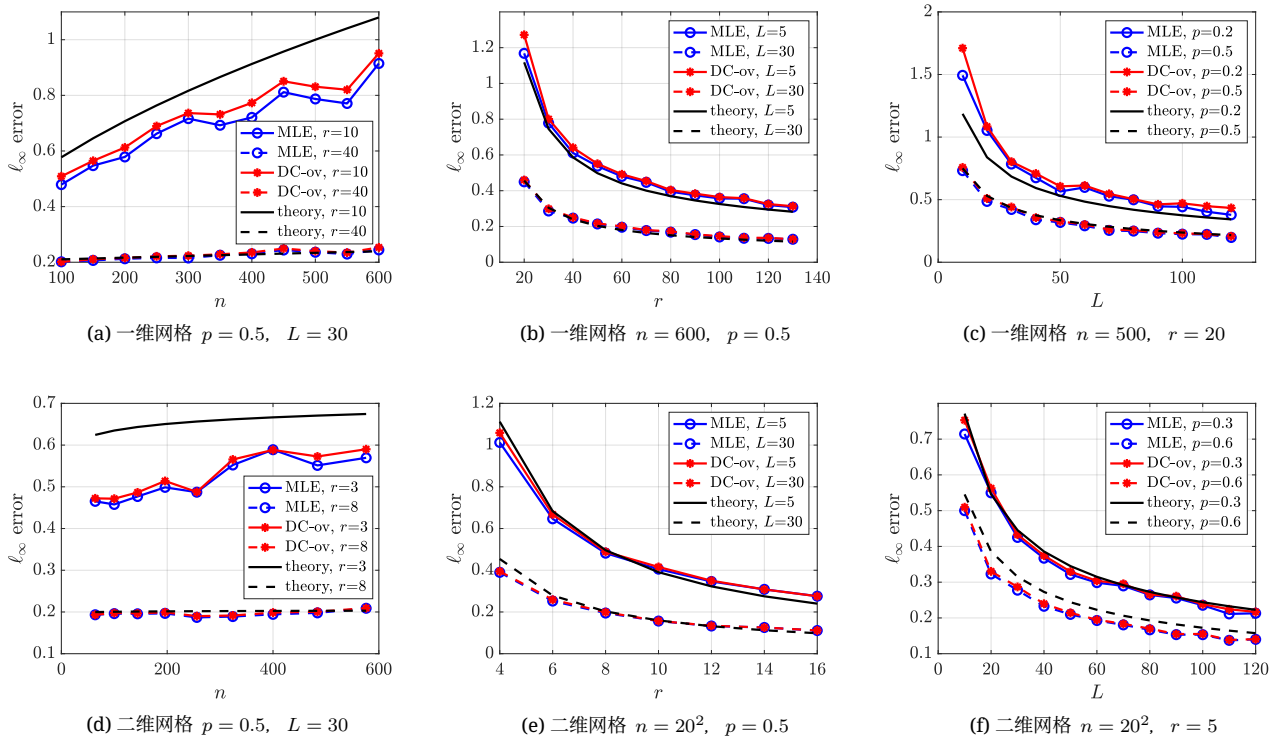
我们比较了用于求解最大似然估计（2）的各种算法的收敛情况。我们考虑两种设定：

- 考虑一个具有 $n=400, r=10, p=0.8, L=100$ 的一维网格图。真实值 $\boldsymbol{\theta}^\star$ 定义为 $\theta_i^\star=i/r$, $1\le i\le n$。
- 考虑一个具有 $n=30^2=900, r=5, p=0.8, L=100$ 的二维网格图。真实值 $\boldsymbol{\theta}^\star$ 定义为 $\theta_{\boldsymbol{i}}^\star=(i_1+i_2)/r$。

感兴趣的算法（及其配置）包括以下内容：

- 预条件梯度下降（PrecondGD）：我们固定步长 $\eta=1$；对于预条件器，我们考虑“预言机”选择 $\boldsymbol{L}_z^\dagger$，其中 $\boldsymbol{L}_z$ 在（6）中定义，以及可实现的选择 $\boldsymbol{L}^\dagger$，其中$\boldsymbol{L}=(1/4)\sum_{(i,j)\in E,i<j}L_{i,j}(\boldsymbol{e}_i-\boldsymbol{e}_j)(\boldsymbol{e}_i-\boldsymbol{e}_j)^\top$（即，定义$\boldsymbol{L}_z$ 中的 $\{z_{i,j}=\sigma'(\theta_i^\star-\theta_j^\star)\}$ 项被替换为 $\sigma'(0)=1/4$）。

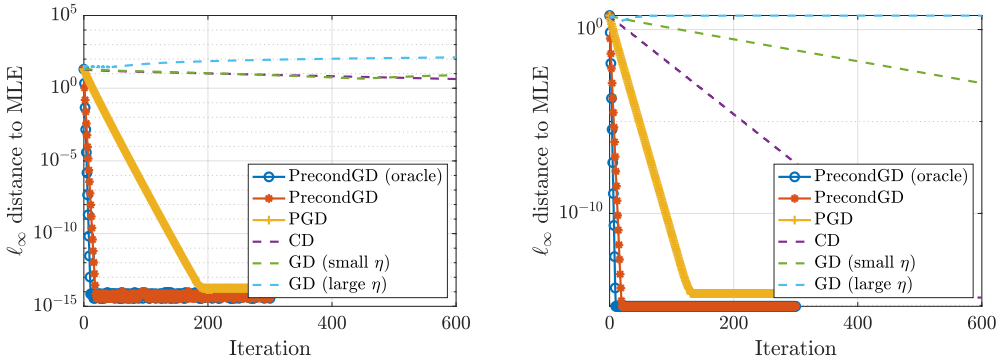
- 投影梯度下降（PGD）：将图划分为大小为 $\Theta(r)$ 的重叠 Erdős-Rényi 子图（适用于一维网格），或大小为 $\Theta(r^2)$ 的重叠 Erdős-Rényi 子图（适用于二维网格）；一维网格的步长设为 $\eta\asymp 1/rpL$，二维网格的步长设为 $\eta\asymp 1/r^2pL$。



(a) 一维网格 $p = 0.5,\ \ L = 30$　(b) 一维网格 $n = 600,\ \ p = 0.5$　(c) 一维网格 $n = 500,\ \ r = 20$

(d) 二维网格 $p = 0.5,\ \ L = 30$　(e) 二维网格 $n = 20^2,\ \ p = 0.5$　(f) 二维网格 $n = 20^2,\ \ r = 5$

图 6：MLE 与 DC-overlap 的经验比较，误差度量为 $\|\boldsymbol{\theta} - \boldsymbol{\theta}^\star\|_\infty$。每条曲线均为 40 次独立试验的平均值。



(a) 一维网格　(b) 二维网格

图 7：求解最大似然估计的各种方法的收敛情况。

- 坐标下降（CD）：每一步更新 $\boldsymbol{\theta}$ 的一个分量（通过最小化式（2）中与该分量相关的项之和），并以无环方式进行；我们将一系列 n 更新视为一次迭代。

- 梯度下降（GD）：对于每种设置，我们考虑两种步长 $\eta_{\text{较小的}}$和 $\eta_{\text{较大的}} = 5\eta_{\text{较小步长}}$，其中$\eta_{\text{较小步长}} \asymp 1/rpL$ 用于一维网格，或 $\eta_{\text{较小步长}} \asymp 1/r^2pL$ 用于二维网格；这些步长选在收敛边界处，即采用$\eta_{\text{较小步长}}$的 GD 会收敛，而采用 $\eta_{\text{较大步长}}$的 GD 则不会收敛。

很容易验证，CD 和 GD 的每次迭代计算复杂度为 $O(|E|)$。对于 PrecondGD 和投影梯度下降（PGD）也是如此（至少在理论上），前提是它们采用最先进的拉普拉斯线性系统求解器实现。

图 7 展示了两种设置下各算法的实测比较结果。结果证实，PrecondGD 的收敛速度最快，其次是 PGD；另一方面，CD 和 GD 等经典一阶方法则慢得多，尤其是在规模较大且半径 n 较小的一维网格中 r。

6 讨论

我们研究了具有一般图或局部性图的 BTL 模型中的基于成对比较的排序问题。展望未来，亟需证明 BTL 模型中具有理论依据的黄金标准算法最大似然估计，能够在具有局部性的图中的最稀疏区域 达到最优误差率；我们希望本研究能有助于实现这一最终目标。在实践方面，值得探索计算效率更高的最大似然估计求解方法（尤其是在损失函数景观条件不佳时），以及更实用的分治算法（例如，无需将图划分作为输入）。

致谢

我们谨此感谢陈宇新在讨论中提供的诸多宝贵帮助。

A 延后证明

A.1 引理证明 1

A.1.1 第 1 部分

我们通过反证法证明引理 1 的第一部分。假设不存在唯一且有限的最大似然估计解。那么，根据命题 1，存在节点的一个不相交划分 Ω_1, Ω_2，使得$y_i^{(\ell)}{}_{,j} = 1$ 对所有 $(i,j) \in (\Omega_1 \times \Omega_2) \cap E$ 以及 $1 \le \ell \le L_{i,j}$均成立。因此，在任何包含 $\boldsymbol{\theta}^\star$ 的有界集合内，对于该集合中的任意$\boldsymbol{\theta}$，都存在一个严格下降方向（使 Ω_1 上的分数增加，并使 Ω_2 上的分数降低），这意味着 $\|\nabla \mathcal{L}(\boldsymbol{\theta})\|_2$ 有一个严格大于零的下界。

然而，我们已在第 3.3 节的步骤 2 中证明，在以（12）为条件时，PrecondGD迭代$\{\boldsymbol{\theta}^t\}$ 始终处于一个包含 $\boldsymbol{\theta}^\star$的有界集合内；此外，由于损失函数 $\mathcal{L}$（$\boldsymbol{\theta}$）是凸且光滑的，经典下降引理（经过调整以考虑预条件器）告诉我们，当步长 η足够小时，有 $\min_{1\le t\le T} \|\nabla \mathcal{L}$（$\boldsymbol{\theta}^t$）$\|_2 \lesssim 1/T \to 0$ 随着 T 趋于无穷大。因此，我们遇到了矛盾，这意味着必定存在唯一且有限的最大似然估计解。

A.1.2 第 2 部分

考虑最小化一般损失函数 $\mathcal{L}(\boldsymbol{\theta})$，其中 $\boldsymbol{\theta} \in \mathbb{R}^d$。PrecondGD 的第 t 次迭代为

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \eta \boldsymbol{M} \cdot \nabla \mathcal{L}(\boldsymbol{\theta}^t), \tag{22}$$

其中 $\boldsymbol{M} \in \mathbb{R}^{d\times d}$ 是一个对称预条件器。以下结果保证了 Pre- condGD 的收敛，其证明稍后给出。

命题 3. 考虑 *PrecondGD* 的第 t 次迭代（22）。设 $\boldsymbol{U} \in \mathbb{R}^{d\times m}$ 是一个正交归一矩阵，表示 m 维子空间 $\mathbb{R}^d$。假设满足以下条件：

- 优化问题 $\min_{\boldsymbol{\theta}\in\mathsf{col}(\boldsymbol{U})}\mathcal{L}(\boldsymbol{\theta})$，其中 $\mathcal{L}$ 是光滑且凸的，具有唯一极小值点 $\widehat{\boldsymbol{\theta}}$；
- 存在一个凸集 $\mathcal{C} \subseteq \mathsf{col}(\boldsymbol{U})$，使得 $\boldsymbol{U}^\top\boldsymbol{\theta}^t, \boldsymbol{U}^\top\widehat{\boldsymbol{\theta}} \in$ 是$\boldsymbol{U}^\top\mathcal{C})$的内部(；
- 海森矩阵和预条件器满足：$\mathsf{col}(\nabla^2\mathcal{L}(\boldsymbol{\theta}))$、$\mathsf{col}(\boldsymbol{M}) \subseteq \mathsf{col}(\boldsymbol{U})$、 $\boldsymbol{M} \succcurlyeq \epsilon\cdot\boldsymbol{U}\boldsymbol{U}^\top$ （对于某个$\epsilon > 0$）是对称的，并且对于所有 $\boldsymbol{\theta} \in \mathcal{C}$， $\alpha_1\boldsymbol{M}^\dagger \preccurlyeq \nabla^2 L(\boldsymbol{\theta}) \preccurlyeq \alpha_2\boldsymbol{M}^\dagger$，其中 $0 < \alpha_1 \le \alpha_2$。

在这些条件下，我们有 $\|\boldsymbol{\theta}^{t+1} - \widehat{\boldsymbol{\theta}}\|_{\boldsymbol{M}^\dagger} \le (1-\eta\alpha_1)\|\boldsymbol{\theta}^t - \widehat{\boldsymbol{\theta}}\|_{\boldsymbol{M}^\dagger}$，前提是 $0 < \eta \le 1/\alpha_2$。

接下来需要说明引理 1 的第 2 部分如何由此命题推出。首先，我们将指定预条件器 $\boldsymbol{M} = \boldsymbol{L}_z^\dagger$ 和子空间 $\boldsymbol{U}$，其中 $\mathsf{col}(\boldsymbol{U}) = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x}^\top\mathbf{1}_n = 0\}$。此外，令 $\mathcal{C}$ 为一个足够大的有界凸集，使得 $\boldsymbol{\theta}^{\mathsf{MLE}}$ 和

$$\mathcal{C}_B := \left\{\boldsymbol{\theta} \in \mathbb{R}^n : \boldsymbol{\theta}^\top\mathbf{1}_n = 0; \left|(\theta_k - \theta_\ell) - (\theta_k^\star - \theta_\ell^\star)\right| \le B_{k,\ell}, (k,\ell) \in E\right\}$$

位于的内部 $\mathcal{C}$ （投影到子空间 col（$\boldsymbol{U}$））后。最后，回顾表达式（3b）中的 $\nabla^2\mathcal{L}$（$\boldsymbol{\theta}$）；由于 $\mathcal{C}$的有界性，存在一些 $0 < \alpha_1 \le \alpha_2$ 使得 $\alpha_1\boldsymbol{L}_z \preccurlyeq\nabla^2\mathcal{L}$（$\boldsymbol{\theta}$） $\preccurlyeq \alpha_2\boldsymbol{L}_z$ 对于所有 $\boldsymbol{\theta} \in \mathcal{C}$。现在，很容易检查出，在引理 1的设置下，对于所有 $t \ge 0$，命题 3 中的必要条件均得到满足，这就完成了引理的证明。

命题的证明 *3*。我们首先证明 $\boldsymbol{U} = \boldsymbol{I}_d$ 这一较简单情形。记 $\boldsymbol{\delta}^t := \boldsymbol{\theta}^t - \widehat{\boldsymbol{\theta}}$。根据 PrecondGD 更新以及关于 $\boldsymbol{\theta}$ 的最优性条件，即 $\nabla\mathcal{L}(\boldsymbol{\theta}) = \mathbf{0}$，我们有

$$\boldsymbol{\delta}^{t+1} = \boldsymbol{\delta}^t - \eta\boldsymbol{M}\nabla\mathcal{L}(\boldsymbol{\theta}^t) = \boldsymbol{\delta}^t - \eta\boldsymbol{M}\big(\nabla\mathcal{L}(\boldsymbol{\theta}^t) - \nabla\mathcal{L}(\widehat{\boldsymbol{\theta}})\big) = \boldsymbol{\delta}^t - \eta\boldsymbol{M}\int_0^1 \nabla^2\mathcal{L}\big(\boldsymbol{\theta}^t(\tau)\big)\mathrm{d}\tau\cdot\boldsymbol{\delta}^t, \tag{23}$$

其中 $\boldsymbol{\theta}^t(\tau) := \widehat{\boldsymbol{\theta}} + \tau\cdot(\boldsymbol{\theta}^t - \widehat{\boldsymbol{\theta}})$ $0 \le \tau \le 1$。因此，

$$\begin{aligned}\boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{\delta}^{t+1} &= \boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{\delta}^t - \eta\boldsymbol{M}^{\frac{1}{2}}\int_0^1 \nabla^2\mathcal{L}\big(\boldsymbol{\theta}^t(\tau)\big)\mathrm{d}\tau\cdot\boldsymbol{M}^{\frac{1}{2}}\cdot\boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{\delta}^t \\ &= \Big(\boldsymbol{I}_n - \eta\boldsymbol{M}^{\frac{1}{2}}\int_0^1 \nabla^2\mathcal{L}\big(\boldsymbol{\theta}^t(\tau)\big)\mathrm{d}\tau\cdot\boldsymbol{M}^{\frac{1}{2}}\Big)\boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{\delta}^t.\end{aligned}$$

根据 $\mathcal{C}$ 的凸性，对于所有 $0 \le \tau \le 1$，我们有 $\boldsymbol{\theta}^t(\tau) \in \mathcal{C}$，因此根据假设，$\alpha_1\boldsymbol{I}_n \preccurlyeq \boldsymbol{M}^{\frac{1}{2}}\int_0^1\nabla^2\mathcal{L}(\boldsymbol{\theta}^t(\tau))\mathrm{d}\tau\cdot\boldsymbol{M}^{\frac{1}{2}} \preccurlyeq \alpha_2\boldsymbol{I}_n$。因此，若 $0 < \eta \le 1/\alpha_2$，则 $\mathbf{0} \preccurlyeq \boldsymbol{I}_n - \eta\boldsymbol{M}^{\frac{1}{2}}\int_0^1\nabla^2\mathcal{L}(\boldsymbol{\theta}^t(\tau))\mathrm{d}\tau\cdot\boldsymbol{M}^{\frac{1}{2}} \preccurlyeq (1-\eta\alpha_1)\boldsymbol{I}_n$，这意味着 $\|\boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{\delta}^{t+1}\|_2 = \|\boldsymbol{\delta}^{t+1}\|_{\boldsymbol{M}^{-1}} \le (1-\eta\alpha_1)\|\boldsymbol{\delta}^t\|_{\boldsymbol{M}^{-1}}$。

一般 $\boldsymbol{U}$ 情形的证明非常相似，只需做一些细微修改。注意，根据假设，我们有 $\boldsymbol{\delta}^t = \boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{\delta}^t$， $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{M}$，以及 $\boldsymbol{U}\boldsymbol{U}^\top\nabla\mathcal{L}(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$。因此 $\boldsymbol{M}\nabla\mathcal{L}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{M}\boldsymbol{U}\boldsymbol{U}^\top\nabla\mathcal{L}(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$，因此不难验证 (23) 仍然成立。因此，

$$\begin{aligned}(\boldsymbol{M}^\dagger)^{\frac{1}{2}}\boldsymbol{\delta}^{t+1} &= (\boldsymbol{M}^\dagger)^{\frac{1}{2}}\boldsymbol{\delta}^t - \eta\boldsymbol{M}^{\frac{1}{2}}\int_0^1 \nabla^2\mathcal{L}\big(\boldsymbol{\theta}^t(\tau)\big)\mathrm{d}\tau\cdot\boldsymbol{M}^{\frac{1}{2}}\cdot(\boldsymbol{M}^\dagger)^{\frac{1}{2}}\boldsymbol{\delta}^t \\ &= \Big(\boldsymbol{U}\boldsymbol{U}^\top - \eta\boldsymbol{M}^{\frac{1}{2}}\int_0^1 \nabla^2\mathcal{L}\big(\boldsymbol{\theta}^t(\tau)\big)\mathrm{d}\tau\cdot\boldsymbol{M}^{\frac{1}{2}}\Big)(\boldsymbol{M}^\dagger)^{\frac{1}{2}}\boldsymbol{\delta}^t.\end{aligned}$$

若 $0 < \eta \le 1/\alpha_2$，则 $\mathbf{0} \preccurlyeq \boldsymbol{U}\boldsymbol{U}^\top - \eta\boldsymbol{M}^{\frac{1}{2}}\int_0^1\nabla^2\mathcal{L}(\boldsymbol{\theta}^t(\tau))\mathrm{d}\tau\cdot\boldsymbol{M}^{\frac{1}{2}} \preccurlyeq (1-\eta\alpha_1)\boldsymbol{U}\boldsymbol{U}^\top$，这意味着$\|(\boldsymbol{M}^\dagger)^{\frac{1}{2}}\boldsymbol{\delta}^{t+1}\|_2 = \|\boldsymbol{\delta}^{t+1}\|_{\boldsymbol{M}^\dagger} \le (1-\eta\alpha_1)\|\boldsymbol{\delta}^t\|_{\boldsymbol{M}^\dagger}$。

□

B 其他结果

事实 3. *Sigmoid* 函数 $\sigma(x) = 1/(1+e^{-x})$ 满足：对于所有 $x \in \mathbb{R}$, $\sigma'(x) \geq 1/(4e^{|x|})$。因此，对于任意 $\alpha \geq 1$, $|x| \leq \log \alpha$ 意味着 $\sigma'(x) \geq 1/(4\alpha)$。

证明。 $\sigma'(x) = \sigma(x) \cdot (1-\sigma(x)) = \frac{1}{(1+e^{-x})(1+e^{x})} = \frac{1}{(1+e^{-|x|})(1+e^{|x|})} \geq \frac{1}{2(1+e^{|x|})} \geq \frac{1}{4e^{|x|}}$。 □

例 1。 我们考虑一些特殊图，其中参数 $\{B_{k,\ell}\}$ (回顾一下 $B_{k,\ell} \lesssim \sqrt{\Omega_{k,\ell}(\boldsymbol{L}_z)\kappa_E}$) 和/或 $\{V_{k,\ell}\}$ 在第 3 节中定义的参数，可以借助事实 1 显式计算或给出界。我们关于 $\Omega_{k,\ell}$ 的一些界已见于文献，例如 [HOS19，表 1]。下文中，我们始终令 n 表示图中的节点数。

- 树。任取 $k \neq \ell$，并令 $P_{k,\ell}$ 为连接节点 k 和 ℓ的路径；则根据串联定律，$\Omega_{k,\ell}(\boldsymbol{L}_z) = \sum_{(i,j)\in P_{k,\ell}} 1/L_{i,j}z_{i,j}$。此外，对于任意$(k,\ell) \in E$，$(\boldsymbol{e}_i - \boldsymbol{e}_j)^\top \boldsymbol{L}_z^\dagger (\boldsymbol{e}_k - \boldsymbol{e}_\ell)$ 仅当 $(i,j) = (k,\ell)$ 时非零，这意味着$V_{k,\ell} = L_{k,\ell}\Omega_{k,\ell}(\boldsymbol{L}_z) = L_{k,\ell} \cdot (1/L_{k,\ell}z_{k,\ell}) = 1/z_{k,\ell} \lesssim \kappa_E$。
- 环。为简单起见，假设 $L_{i,j} = L$ 。考虑任意 $k \neq \ell$，并令 $P_{k,\ell}$ 为连接节点 k 和 ℓ 的两条路径中的任意一条；则 $\Omega_{k,\ell}(\boldsymbol{L}_z) \leq \sum_{(i,j)\in P_{k,\ell}} 1/L_{i,j}z_{i,j} \lesssim \kappa_E n/L$。至于 $V_{k,\ell}$，其中 $(k,\ell) \in E$，回想一下，$V_{k,\ell} = L\sum_{(i,j)\in E, i<j} |v_i - v_j|$，其中 $\boldsymbol{v} = \boldsymbol{L}_z^\dagger(\boldsymbol{e}_k - \boldsymbol{e}_\ell)$；由于沿着连接节点 k 和 ℓ 的路径，有许多 $|v_i - v_j|$ 项相互抵消，因此得到 $V_{k,\ell} = 2L(v_k - v_\ell) = 2L \cdot \Omega_{k,\ell}(\boldsymbol{L}_z) \leq 2/z_{k,\ell} \lesssim \kappa_E$。
- **完全图**。 为简单起见，假设 $L_{i,j} = L$ 对于任意 $k \neq \ell$，我们都可以找到 $\Theta(n)$ 条连接节点 k 和 ℓ 的并行路径，每条路径的长度至多为 2，电阻为 $O(\kappa_E/L)$。因此，根据并联定律，$\Omega_{k,\ell}(\boldsymbol{L}_z) \lesssim \kappa_E/nL$ 。
- 哑铃图。假设两个完全子图（各有 $\Theta(n)$个节点）通过一条单条边 (s,t) 相连，其中 s 属于一个子图，而 t 属于另一个子图。假设对于所有 $(i,j) \neq (s,t)$，$L_{i,j} = L$ 。如果节点 k 和 ℓ 属于同一个完全子图，那么根据先前关于完全图的结果，$\Omega_{k,\ell}(\boldsymbol{L}_z) \lesssim \kappa_E/nL$。否则，若节点 k,s 属于一个子图，而 ℓ,t 属于另一个子图，则根据有效电阻的三角不等式，我们有 $\Omega_{k,\ell}(\boldsymbol{L}_z) \leq \Omega_{k,s}(\boldsymbol{L}_z) + \Omega_{s,t}(\boldsymbol{L}_z) + \Omega_{t,\ell}(\boldsymbol{L}_z) \lesssim \kappa_E(\frac{1}{nL} + \frac{1}{L_{s,t}})$；这说明，增加图的瓶颈（即边 (s,t)）上的样本量$L_{s,t}$，对于实现较小的估计误差至关重要。

参考文献

[BR22] H. Bong 和 A. Rinaldo。Bradley-Terry-Luce 模型中最大似然估计的存在性与一致性的广义结果。发表于国际机器学习会议，第 2160–2177 页。PMLR，2022 年。[BT52] R. A. Bradley 和 M. E. Terry。不完全区组设计的秩分析：I. 成对比较方法。*Biometrika*，39(3/4)：324–345，1952 年。[CFMW19] Y. Chen、J. Fan、C. Ma 和 K. Wang。谱方法和正则化最大似然估计在前 k 名排序中均具有最优性。《统计年鉴》，47(4)：2204，2019 年。[CGZ22a] P. Chen、C. Gao 和 A. Y. Zhang。基于成对比较的最优完整排序。《统计年鉴》，50(3)：1775–1805，2022 年。[CGZ22b] P. Chen、C. Gao 和 A. Y. Zhang。前 k 名排序的部分恢复：最大似然估计的最优性与谱方法的次优性。《统计年鉴》，50(3)：1618–1652，2022 年。[CKST16] Y. Chen、G. Kamath、C. Suh 和 D. Tse。具有局部性的图中的社群恢复。发表于国际机器学习会议，第 689–698 页。PMLR，2016 年。[CRR+96] A. K. Chandra、P. Raghavan、W. L. Ruzzo、R. Smolensky 和 P. Tiwari。图的电阻刻画了其往返时间和覆盖时间。计算复杂度，6(4)：312–340，1996 年。

[CS15] Y. Chen 和 C. Suh。谱最大似然估计：基于成对比较的 Top-k 排名聚合。载于《国际机器学习会议》论文集，第 371–380 页。PMLR，2015 年。[DS84] P. G. Doyle 和 J. L. Snell。《随机游走与电网络》，第 22 卷。美国数学学会，1984 年。[ESVM+11] W. Ellens、F. M. Spieksma、P. Van Mieghem、A. Jamakovic 和 R. E. Kooij。有效图电阻。《线性代数及其应用》，435(10):2491–2506，2011 年。[FJ57] L. R. Ford Jr. 基于二元比较的排序问题的求解。《美国数学月刊》，64(8P2):28–33，1957 年。[GBS08] A. Ghosh、S. Boyd 和 A. Saberi。最小化图的有效电阻。《SIAM评论》，50(1):37–66，2008 年。[GSZ21] C. Gao、Y. Shen 和 A. Y. Zhang。Bradley-Terry-Luce 模型中的不确定性量化。arXiv 预印本 arXiv:2110.03874，2021 年。[HOS19] J. Hendrickx、A. Olshevsky 和 V. Saligrama。图电阻与基于成对比较的学习。载于《国际机器学习会议》论文集，第 2702–2711 页。PMLR，2019 年。[HOS20] J. Hendrickx、A. Olshevsky 和 V. Saligrama。BTL 模型中基于成对比较的学习的极小极大速率。载于《第37届国际机器学习会议论文集》，《机器学习研究论文集》第 119 卷，第 4193–4202 页。PMLR，2020 年 7 月 13–18 日。[Hun04] D. R. Hunter。广义 Bradley-Terry 模型的 MM 算法。《统计年鉴》，32(1):384–406，2004 年。[HYTC20] R. Han、R. Ye、C. Tan 和 K. Chen。稀疏 Bradley-Terry 模型的渐近理论。《应用概率年鉴》，30(5):2491–2515，2020 年。[LFL22] Y. Liu、E. X. Fang 和 J. Lu。排序问题的拉格朗日推断。《运筹学》，2022 年。[LP17] R. Lyons 和 Y. Peres。《树与网络上的概率》，第 42 卷。剑桥大学出版社，2017 年。[LSR22] W. Li、S. Shrotriya 和 A. Rinaldo。一般比较图下 BTL 模型中最大似然估计的 ℓ_∞-界。载于《第三十八届人工智能不确定性会议论文集》，《机器学习研究论文集》第 180 卷，第 1178–1187 页。PMLR，2022 年 8 月 1–5 日。[Luc12] R. D. Luce。《个体选择行为：理论分析》。Courier Corporation，2012 年。[NOS12] S. Negahban、S. Oh 和 D. Shah。基于成对比较的迭代排序。《神经信息处理系统进展》，第 25 卷，2012 年。[SBB+16] N. B. Shah、S. Balakrishnan、J. Bradley、A. Parekh、K. Ramch, ran 和 M. J. Wainwright。基于成对比较的估计：具有拓扑依赖性的精确极小极大界。《机器学习研究杂志》，17(58):1–47，2016 年。[Spi19] D. Spielman。《谱图论与代数图论》。耶鲁大学讲义，2019 年 12 月草稿，第 4 卷：47 页。[ST04] D. A. Spielman 和 S.-H. Teng。用于图划分、图稀疏化和求解线性系统的近线性时间算法。载于《第三十六届 ACM 计算理论年度研讨会论文集》，第 81–90 页，2004 年。[SY99] G. Simons 和 Y.-C. Yao。成对比较 Bradley-Terry 模型中参数数量趋于无穷时的渐近性质。《统计年鉴》，27(3):1041–1060，1999 年。

[V+13] N. K. Vishnoi 等。Lx=b. 理论计算机科学基础与前沿㊣ 理论计算机科学，8(1–2):1–141，2013。[Ver18] R. Vershynin。《高维概率：数据科学中的应用导论》，第 47 卷。剑桥大学出版社，2018。[VYZ22] M. Vojnovic、S. Yun 和 K. Zhou。利用比较数据推断排序分数的加速MM算法。《运筹学》，2022。[WJN22] W. Wu、B. W. Junker 和 N. Niezink。Bradley-Terry 模型识别约束的渐近比较。*arXiv* 预印本 *arXiv:2205.04341*，2022。[WSR20] J. Wang、N. Shah 和 R. Ravi。将最大似然估计在成对比较中的效用从准确性拓展至偏差。载于人工智能与统计国际会议，第 66–76 页。PMLR，2020。[YYX12] T. Yan、Y. Yang 和 J. Xu。Bradley-Terry 模型中的稀疏成对比较。《统计学报》，第 1305–1318 页，2012。[Zer29] E. Zermelo。将比赛结果计算为概率论中的一个极大值问题。《数学杂志》，29(1):436–460，1929。