提交至《统计年鉴》

arXiv:2304.03590v2 [math.ST] 4 Sep 2023

二部图上的图子估计，具有可观测边标签和不可观测节点标签

作者：埃蒂安·多尼耶-梅罗兹[1,a]，阿尔纳克 S. 达拉良[2,b]，弗朗西斯·克拉马尔兹[2,c]，菲利普·肖内[2,d] 和泽维耶·达尔福伊[2,e]

[1]*CREST*，法国国家科学研究中心，巴黎综合理工学院，a*Etienne.Donier-Meroz@ensae.fr*

[2]*CREST*, 法国国立统计与经济管理学校,巴黎综合理工学院,b*arnak.dalalyan@ensae.fr*; c*francis.kramarz@ensae.fr*;d*philippe.chone@ensae.fr*; e*xavier.dhaltfoeuille@ensae.fr*

许多现实世界中的数据集都可以表示为一个矩阵，其元素对应于两种不同性质实体之间的交互（例如，顾客对产品的选择、网络用户访问网页的次数、学生某门课程的成绩、患者对医生的评分等）。本文假设，所述交互由描述各实体的不可观测潜变量决定。我们的目标是估计给定不可观测变量时数据矩阵的条件期望。这可以表述为一个二元函数的估计问题，该函数称为图子。我们研究分段常数图子和 Hölder 连续图子的情形。我们为最小二乘估计量和指数加权聚合估计量建立了有限样本风险界。这些风险界揭示了估计误差与数据集大小、交互的最大强度以及噪声水平之间的关系。由于所研究的最小二乘估计量难以计算，我们提出对劳埃德交替最小化算法进行改造，以计算最小二乘估计量的近似值。最后，我们通过数值实验展示图子估计量在合成数据集上的经验表现。

1 引言 在本文中，我们考虑估计由双变量图子和（未观测到的）潜变量生成的随机矩阵的条件均值问题。更确切地说，设 n 和 m 为两个假定较大的正整数，且 $\mathbf{H}$ 为一个具有实数元素 $H_{i,j}$ 的 $n \times m$ 随机矩阵。我们假设该矩阵 $\mathbf{H}$ 的分布满足以下条件。

假设1. 存在一个函数 $W^*\colon [0,1]^2 \to \mathbb{R}$，称为图子，以及两个随机向量 $\boldsymbol{U} = (U_1, \ldots, U_n)$ 和 $\boldsymbol{V} = (V_1, \ldots, V_m)$，使得
随机变量 $U_1, \ldots, U_n, V_1, \ldots, V_m$ 相互独立，并且服从均匀分布 $\mathcal{U}([0,1])$。

A 1.2 在$(\boldsymbol{U}, \boldsymbol{V})$的条件下，条目 $H_{i,j}$相互独立，并且 $\mathbb{E}[H_{i,j}|\boldsymbol{U}, \boldsymbol{V}] = W^*(U_i, V_j)$。

上述设置对应于这样一种实际情况：其中有 n客户和 m 产品。每位客户都有一个未观测到的潜在特征 U，每个产品都有一个未观测到的潜在特征 V。我们观测到标签 H，该标签刻画了客户与产品之间的交互（通常表示该客户选择了该产品）。函数 W^* 对应于给定潜在特征取值时交互的均值。

MSC2020 学科分类：主要分类 62G05；次要分类62C20。关键词和短语：图函数估计、二部图、风险界、极小极大速率。

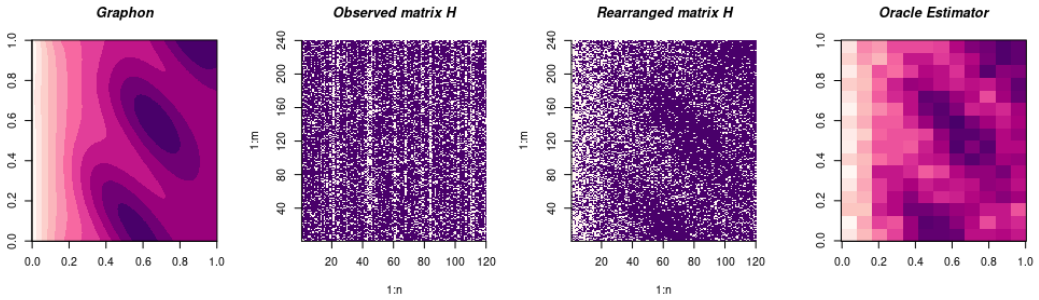


图 1：图子问题示意图。最左侧的图表示未知的图子 W^*。左起第二个图是图中观测到的邻接矩阵，其中的连接是根据伯努利模型生成的。第三个图是如果我们能够获取潜变量，在对行和列重新排列后得到的邻接矩阵。最右侧的图表示根据重新排列后的邻接矩阵得到的直方图估计量。我们的目标是在无法获取潜变量的情况下，设计出一个几乎与预言机一样好的估计量。

上述框架及本文推导结果的适用范围涵盖以下主要示例：

1.条目$H_{i,j}$取值为 0 或1，表示二部图中一条边的存在。在客户和产品的示例中，可以设定 $H_{i,j}=1$，当且仅当客户 i 已经购买了产品 j。在这种情况下，务必考虑大型稀疏图的情形，在这类图中，对于所有节点对，两个节点之间存在一条边的概率都很小。

2.所提供的数据由前一段所概述的顾客-产品图的 N 个实例的聚合组成。每个条目 $H_{i,j}$ 表示顾客 i 购买产品 j的观测频率。例如，在顾客到访商店 $N=10$ 次且 $H_{i,j}=3/10$的情形下，这表示该顾客在十次机会中的三次购买了产品i。在此背景下，$H_{i,j}$ 的取值范围为 0 到 1。假设上述 N 次试验之间具有独立性，$H_{i,j}$的方差随 $1/N$的阶数变化。因此，该框架表现出具有低噪声方差这一独特特征。这个特定实例可以纳入更广泛的次高斯分布背景中，该分布由较小的方差参数 σ^2刻画，其对应于 $1/N$。

3.所提供的数据对应于在时长为$T>0$的时间窗口内收集的观测值。每个 $H_{i,j}$ 表示客户 i 在长度为 $T>0$的时间区间内购买产品 j的平均次数。该情形可用公式 $H_{i,j}=N_{i,j}(T)/T$表示，其中 $N_{i,j}(T)$ 表示一个泊松随机变量，其强度与 T成正比。参数 T 可能很大，从而有助于进行准确估计。需要将此例与前一个例子区分开来，在前一个例子中，$H_{i,j}$ 的值被限制在区间 $[0,1]$内。在当前情形下，$H_{i,j}$ 的值不受此类约束，其大小可以任意变化。

我们的目标是研究估计 W^* 的极小极大风险，并强调其对该问题重要参数的依赖关系。矩阵的大小 n 和 m 是这些参数中的一部分，但我们也关注其对 W^*的光滑性、相互作用的“稀疏性”（记为 ρ）以及噪声水平（记为 σ）的依赖关系。这些参数 ρ 和 σ 是满足以下条件的正实数

$$\|W^*\|_\infty = \sup_{u,v\in[0,1]} |W^*(u,v)| \leqslant \rho \quad \text{and} \quad \mathbf{Var}[H_{i,j}|U_i,V_j] \leqslant \sigma^2 \text{ a.s.}, \quad \forall i\in[n], \forall j\in[m].$$

参数 ρ 和 σ 可能取决于 n 和 m，但为简洁起见，我们选择写作 ρ，m 和 σ_n，m。

在估计图子 W^* 的过程中，一个重要的中间步骤是估计矩阵 $\mathbf{\Theta}^* = W^*(U_i, V_j)$。矩阵估计本身也具有重要价值。我们通过在分块常数矩阵集合上求解最小二乘问题来完成这项任务，其中的块由矩阵 $\mathbf{H}$ 的行集和列集的划分生成。我们还将进一步说明，可以使用指数加权聚合方法来确保对块数的自适应性。在图子为分段常数或按照霍尔德光滑性意义下具有 α-光滑性的条件下，我们为基于 $\mathbf{\Theta}^*$ 的估计量得到的图子估计量建立了风险界。这些风险界是非渐近的，并且对于广泛的参数范围，在极小极大意义下证明了其速率最优性。

1.1 估计量质量的衡量 图 1展示了图子估计问题的一个示例，其中 $\mathbf{H}$ 是二部图的邻接矩阵，也就是说，$\mathbf{H}$ 的元素为 0 或 1。我们从图中可以看到，缺乏潜变量信息会对图子的恢复产生重大影响。事实上，与重新排列后的邻接矩阵（如第三幅图所示）相比，第二幅最靠左的图中所示的邻接矩阵 $\mathbf{H}$ 几乎不包含关于 W^* 的信息。实际上，当 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 未知时，图子 W^* 不可识别。若存在两个保持勒贝格测度的双射，使得 $\tau_1\colon\ [0,1]\to[0,1]$ 和 $\tau_2\colon\ [0,1]\to[0,1]$ 满足[1] $W = W'\circ(\tau_1\otimes\tau_2)$，则称两个图子 W 和 W' 等价。可以验证，由等价图子 W^* 和 W^* 生成的两个矩阵 $\mathbf{H}^-$ 同分布。这意味着我们至多只能估计包含 W^* 的等价类。这就是本文采用以下（伪）距离来衡量估计量 W 对 W^* 的估计质量的原因，即 $\widehat{\ }$

$$\begin{aligned}\delta(\widehat{W}, W^*) &= \inf_{\tau_1,\tau_2\in\mathcal{M}}\left(\iint_{[0,1]^2}\left|\widehat{W}(\tau_1(u),\tau_2(v)) - W^*(u,v)\right|^2 du\,dv\right)^{1/2}\\ &= \inf_{\tau_1,\tau_2\in\mathcal{M}}\|\widehat{W}\circ(\tau_1\otimes\tau_2) - W^*\|_{\mathbb{L}^2}\end{aligned}$$

其中 $\mathcal{M}$ 是所有自同构 $\tau\colon\ [0,1]\to[0,1]$ 的集合，使得 τ 和 τ^{-1} 是可测的，并且 τ 以如下意义保持勒贝格测度：$\lambda(\tau^{-1}(B)) = \lambda(B)$</style}对于每个博雷尔集。如果 和 满足 $(W_1, W_2) = 0$，则称两个图子 W_1 和 W_2 弱同构。

1.2 我们的贡献 本文的主要贡献如下：

• 我们提出了一个基于双变量图子函数和不可观测潜变量的非参数框架，为随机矩阵建模提供了一种灵活的方法，尤其适用于随机二部图的邻接矩阵。

• 我们为在具有给定簇数的分段常数矩阵上最小化平方误差的估计量，以及结合上述最小二乘估计量的指数加权聚合估计量，建立了有限样本风险界。这些结果适用于其元素服从伯努利分布的二部图邻接矩阵，同时也适用于二项分布、缩放泊松分布和次高斯分布。

• 我们将 Lloyd 交替最小化算法（包括一个凸松弛步骤）改编并应用于我们的设定，从而得到最小二乘估计量的计算上可处理的近似。

[1] 我们用记号 $\tau_1\otimes\tau_2$ 表示从 $[0,1]^2$ 到 $[0,1]^2$ 的函数，其定义为 $(\tau_1\otimes\tau_2)(u,\ v) = (\tau_1(u)\ \tau_2(v))$。

• 对于给定潜变量时其条目条件分布为伯努利分布的矩阵，我们证明了对于任意图子估计量，在分段常数图子集合上的最坏情形风险下界。在绝大多数情况下，这些下界与最小二乘估计量所得上界具有相同的阶。

1.3 相关工作 基于与本文所研究模型相似的区块模型，对矩阵与网络数据进行统计分析，至少二十年来一直是一个活跃的研究领域。由 Holland 等人（1983）提出的随机区块模型，或许是网络数据中研究最为广泛的潜在结构之一；早期参考文献还可见（Nowicki 和 Snijders，2001）以及（Govaert 和 Nadif，2003）。社区检测，即检测潜在区块结构的问题，一直是大量研究工作的重点，相关研究包括（Zhao 等人，2012；Chin 等人，2015；Lei 和 Rinaldo，2015；Lei，2016；Zhang 和 Zhou，2016；Wang 和 Bickel，2017；Chen 等人，2018；Xu 等人，2020），相关综述则包括（Gao 等人，2017；Abbe，2018）。需要指出的是，这些研究大多聚焦于具有二元或离散边标签的单部图，其最优性主要与确定社区参数之间、能够实现一致恢复的最小分离率有关。针对二部区块模型的类似问题已在（Feldman 等人，2015；Florescu 和 Perkins，2016；Neumann，2018；Zhou 和 Amini，2019a，2020；Cai 等人，2021；Ndaoud 等人，2022）中进行研究。

与上述论文不同，本文关注的是一个涵盖带有实值边标签的二部图的模型，并从估计误差的角度研究其最优性。图子估计量的一致性已在（Airoldi 等人，2013；Wolfe 和 Olhede，2013；Olhede 和 Wolfe，2014）中得到研究。对于同一问题，最小二乘估计量的极小极大速率最优性已在（Gao 等人，2015，2016；Klopp 等人，2017；Klopp 和 Verzelen，2019）中得到确立，另请参见综述文章（Gao 和 Ma，2021）。为了结合当前的研究现状更好地介绍我们的贡献，简要概述这些论文的内容将很有帮助。Gao 等人（2015）考虑了二元观测 H_i 的情形，重点研究稠密情形 $\rho \asymp 1$，并得到了分段常数图子类和 Hölder连续图函数类上的极小极大估计速率。Gao 等人（2016）将这些结果推广到了具有次高斯元素的矩阵 H，其中部分元素可能是完全随机缺失的。然而，正如第 3.1 节所述，他们的结果在某些情况下关于噪声方差是次优的，例如当边标签服从二项分布时。在具有二元边标签的单部图情形下，（Klopp 等人，2017）确立了稀疏连接图的极小极大最优估计速率，其中 $\rho \ll 1$。虽然（Gao 等人，2015，2016）使用估计矩阵与真实矩阵之差的归一化弗罗贝尼乌斯范数来度量估计误差，但（Klopp 等人，2017）还研究了图子等价类之间的 $\mathbb{L}_2$-距离。在（Klopp 和 Verzelen，2019）中，具有二元观测的单部图的图函数估计在割距离下的极小极大最优速率已得到确立。

本文研究的问题与计量经济学中的一些近期工作有关。事实上，根据（Aldous，1981，定理1.4），如果矩阵 H 是行列交换的，那么存在一个函数 $g^*: [0,1]^4 \to \mathbb{R}$ 以及独立随机变量 $\alpha, \{U_i\}, \{V_j\}, \{\xi_{ij}\}$，它们在 $[0,1]$ 上服从均匀分布，使得随机矩阵 H 和 $\big(g^*(\alpha, U_i, V_j, \xi_{ij});\ i \in [n], j \in [m]\big)$ 同分布。本文所考虑的问题等价于根据观测值 $g^*(\alpha, U_i, V_j, \xi_{ij});\ i \in [n], j \in [m]$，在不假设函数 g^*具有任何参数形式的前提下，估计随机函数 $W^*(u,v) = \int_0^1 g^*(\alpha, u, v, z)\, dz$。如果除了 $H_{i,j}$之外，对于每一对节点 (i,j)，我们还知道一个特征向量$\boldsymbol{X}_{i,j}$，那么可以考虑由 $g^*(\boldsymbol{X}_{i,j}, \alpha, U_i, V_j, \xi_{i,j})$ 定义的扩展模型。例如，（Graham，2017）采用了这种方法，其中考虑了具体的参数形式 $g^*(\boldsymbol{x}, \alpha, u, v, z) = \mathbb{1}\big(\boldsymbol{x}^\top \boldsymbol{\beta} + u + v + \log(z/(1-z))\big)$。

在这种参数化背景下，感兴趣的参数是向量 $\boldsymbol{\beta}$。在（Graham，2020）中，假设回归函数 $\int_{[0,1]^3} g^*(\boldsymbol{x},\ \alpha,\ u,\ v,\ z)\,du\,dv\,dz$ 具有参数形式 $\exp\{a+\boldsymbol{x}^\top\boldsymbol{\beta}\}/(1+\exp\{a+\boldsymbol{x}^\top\boldsymbol{\beta}\})$，并研究了向量 $(a,\ \boldsymbol{\beta})$ 的估计问题。Davezies et al.（2021）证明了可交换数组的渐近结果（大数定律和中心极限定理）。Davezieset al.(2021).

1.4 符号对于整数 $n\geqslant 1$，我们设 $[n]=\{1,\dots,n\}$。在数学公式中，矩阵用粗体大写字母表示，向量用粗体斜体字母表示。实数 x 的整数部分记为 $\lfloor x\rfloor$，而两个实数 $x,\ y$ 的最小值和最大值分别记为 $x\wedge y$ 和 $x\vee y$。对于两个 $n\times m$ 矩阵 $\mathbf{B}$ 和 $\mathbf{B}$，内积定义为

$$\langle \mathbf{B},\bar{\mathbf{B}}\rangle=\operatorname{tr}(\mathbf{B}\bar{\mathbf{B}}^\top)=\sum_{i=1}^{n}\sum_{j=1}^{m}B_{ij}\bar{B}_{ij},$$

我们用 $\|\mathbf{B}\|_{\mathsf{F}}=\sqrt{\langle \mathbf{B},\ \mathbf{B}\rangle}$ 表示矩阵 $\mathbf{B}$的弗罗贝尼乌斯范数。$\mathbf{B}$的上确界范数记为 $\|\mathbf{B}\|_\infty$，定义为 $\mathbf{B}$中绝对值最大的元素。我们分别用 $\boldsymbol{B}_{i,\ \bullet}$ 和$\boldsymbol{B}_{\bullet,\ j}$ 表示$\mathbf{B}$的第 i行和第 j列。区间 $I\subset\mathbb{R}$ 的长度记为 $|I|$。对于 $n\in\mathbb{N}$，$\mathbf{1}_n$ 是所有元素均为一的 n维向量。

2 均值矩阵与图子的估计量 本节定义均值矩阵 $\boldsymbol{\Theta}^*=\mathbf{E}[\mathbf{H}|\boldsymbol{U},\boldsymbol{V}]$ 以及图子 W^* 的估计量，并对本文研究的这些估计量进行介绍。此处我们仅关注数学定义；这些估计量及其易于处理的近似形式在计算和算法方面的性质将在第 4节中讨论。

2.1 $\boldsymbol{\Theta}^*$ 的最小二乘估计量我们先介绍一些记号。对于满足 $Kn_0\leqslant n$ 和 $K\geqslant 2$的正整数 n_0,n,K，定义集合

$$\mathcal{Z}(n,K,n_0)=\left\{\mathbf{Z}\in\{0,1\}^{n\times K}:\mathbf{Z}\mathbf{1}_K=\mathbf{1}_n \text{ and } \min_{k\in[K]}\mathbf{1}_n^\top\boldsymbol{Z}_{\bullet,k}\geqslant n_0\right\}. \tag{1}$$

该集合中的元素可视为指派矩阵：n 个用户中的每一个都被指派给 K 个“社群”中的一个（且仅一个），并且每个社群至少有 n_0 个“成员”。类似地，我们将反复使用集合 $\mathcal{Z}(m,L,m_0)$，其中包含与各项对应的指派矩阵。由于 H 的 n 行对应用户，而 H 的 m 列对应各项，因此 $\mathcal{Z}(n,K,n_0)$中的元素记为 $\mathbf{Z}^{\text{cust}}$，而 $\mathcal{Z}(m,L,m_0)$中的元素记为 $\mathbf{Z}^{\text{prod}}$。矩阵 $\mathbf{Z}^{\text{cust}}$ 和 $\mathbf{Z}^{\text{prod}}$ 对应一种双聚类：用户簇由矩阵 $\mathbf{Z}^{\text{cust}}$指定；同样，$\mathbf{Z}^{\text{prod}}$ 编码了各项的簇。

给定观测到的邻接矩阵 $\mathbf{H}$，最小二乘估计量定义为

$$(\widehat{\mathbf{Q}},\widehat{\mathbf{Z}}^{\text{cust}},\widehat{\mathbf{Z}}^{\text{prod}})^{\mathsf{LS}}\in\arg\min_{\substack{\mathbf{Q}\in\mathbb{R}^{K\times L}\\ \mathbf{Z}^{\text{cust}}\in\mathcal{Z}(n,K,n_0)\\ \mathbf{Z}^{\text{prod}}\in\mathcal{Z}(m,L,m_0)}}\left\|\mathbf{H}-\mathbf{Z}^{\text{cust}}\mathbf{Q}(\mathbf{Z}^{\text{prod}})^\top\right\|_{\mathsf{F}}^2. \tag{2}$$

这里，$\mathbf{Z}^{\text{cust}}\mathbf{Q}(\mathbf{Z}^{\text{prod}})^\top$ 是一个 $n\times m$ 分块常数矩阵。因此，我们的目标是找到一个与 $\mathbf{H}$ 最接近的分块常数矩阵，这种接近程度由弗罗贝尼乌斯范数诱导的度量定义，其中块由矩阵$\mathbf{Z}^{\text{cust}}$ 和$\mathbf{Z}^{\text{prod}}$给出，且块的大小至少为$n_0\times m_0$。

通过（2）计算得到的这些估计量，导出了$\boldsymbol{\Theta}^*$ 的分块常数最小二乘估计量，其定义为 $\widehat{\boldsymbol{\Theta}}^{\mathsf{LS}}=\widehat{\ }\mathbf{Z}^{\text{cust}}\widehat{\ }\mathbf{Q}(\widehat{\ }\mathbf{Z}^{\text{prod}})^\top$。可以将 $\widehat{\boldsymbol{\Theta}}^{\mathsf{LS}}$ 写成以下另一种形式。考虑分块常数矩阵类

$$\mathcal{T}=\mathcal{T}_{n_0,m_0}^{K,L}=\left\{\boldsymbol{\Theta}=\mathbf{Z}^{\text{cust}}\mathbf{Q}(\mathbf{Z}^{\text{prod}})^\top:\ (\mathbf{Q},\mathbf{Z}^{\text{cust}},\mathbf{Z}^{\text{prod}})\in\mathbb{R}^{K\times L}\times\mathcal{Z}_{n,K,n_0}\times\mathcal{Z}_{m,L,m_0}\right\}. \tag{3}$$

最小二乘估计量 $\widehat{\mathbf{\Theta}}^{\mathsf{LS}}$ 是以下问题的一个解：

$$\widehat{\mathbf{\Theta}}^{\mathsf{LS}} = \widehat{\mathbf{\Theta}}^{\mathsf{LS}}_{n_0,m_0}[K,L] \in \arg\min_{\mathbf{\Theta}\in\mathcal{T}^{K,L}_{n_0,m_0}} \|\mathbf{H}-\mathbf{\Theta}\|^2_{\mathsf{F}}. \tag{4}$$

我们的第一项结果将在下一节中给出，为估计量 $\widehat{\mathbf{\Theta}}^{\mathsf{LS}}$ 的风险提供非渐近上界。

2.2 指数加权聚合 最小二乘估计量 $\widehat{\mathbf{\Theta}}^{\mathsf{LS}}$ 在 KL 个块上为常数。这些块的数量预先选定，是该方法的一个超参数。如果真实矩阵 $\mathbf{\Theta}^*$ 远非在 KL 个块上分块常数，那么由于存在较大偏差，基于 $\widehat{\mathbf{\Theta}}^{\mathsf{LS}}_{n_0,m_0}[K,L]$ 的估计质量可能较差。可以针对 K,L的多个取值（也包括 n_0 和 m_0）计算最小二乘估计量，再对这些估计量进行聚合，从而降低偏差。

为此，我们在此考虑扩展框架，其中观测到两个相互独立的副本 $\mathbf{H}$和 $\mathbf{H}'$，二者均满足假设 A 1.2（具有完全相同的 $\boldsymbol{U}$ 和 $\boldsymbol{V}$）。矩阵 $\mathbf{H}$ 用于构造估计量，而 $\mathbf{H}'$ 用于定义“权重”，这些“权重”用于计算指数加权聚合估计量（EWA）。更准确地说，我们将$\widehat{\mathbf{\Theta}}^{\mathsf{LS}}_1$、.....、$\widehat{\mathbf{\Theta}}^{\mathsf{LS}}_s$最小二乘估计量记为通过对 个不同的（$K$， L， n_0， m_0）求解（4）得到。我们定义

$$\widehat{\mathbf{\Theta}}^{\mathsf{EWA}} = \sum_{\ell=1}^{s} w_\ell \widehat{\mathbf{\Theta}}^{\mathsf{LS}}_\ell, \quad \text{with} \quad w_\ell = \frac{\exp\{\|\mathbf{H}'-\widehat{\mathbf{\Theta}}^{\mathsf{LS}}_\ell\|^2_{\mathsf{F}}/\beta\}}{\sum_{r=1}^{s}\exp\{\|\mathbf{H}'-\widehat{\mathbf{\Theta}}^{\mathsf{LS}}_r\|^2_{\mathsf{F}}/\beta\}} \tag{5}$$

其中， $\beta>0$ 是一个通常被称为温度参数的参数。由于 $\mathbf{H}'$ 和 $\widehat{\mathbf{\Theta}}^{\mathsf{LS}}_\ell$是在两个相互独立的数据矩阵上计算得到的，因此它们相互独立；这一点将在证明中发挥重要作用。β 的选择取决于观测值的性质，更确切地说，取决于噪声 $\boldsymbol{\xi}$的分布，详见下一节。

2.3 缺失值情况下的调整 $\mathbf{\Theta}^*$ 在前文中介绍的估计量使用了矩阵H 的所有条目。然而，这些估计量以及下一节中陈述的数学结果，都很容易调整到存在缺失观测的情形。更准确地说，假设我们观测到一些取值为0 和1 的独立同分布随机变量 $M_{i,j}$，并且当且仅当 $M_{i,j}=1$时，值 $H_{i,j}$ 才会被统计学家观测到。记$\mathsf{p}=\mathsf{P}(M_{i,j}=1)$，并假设 $M_{i,j}$ 与$(H_{i,j},U_i,V_j)$ 独立（这种情形通常称为完全随机缺失），我们可以通过其条目 $H_{i,j}=H_{i,j}M_{i,j}/\mathsf{p}$（对于 $i\in[n]$ 和 $j\in[m]$）来定义调整后的$\widetilde{\ }$ $\widetilde{\ }$观测矩阵 $\mathbf{H}$。

然后，为了定义 $\mathbf{\Theta}^*$ 的最小二乘估计量，只需将 $\mathbf{H}$ 替换为 $\widetilde{\mathbf{H}}$ 在（2）中。定义指数加权聚合估计量时也可以进行类似的修改。请注意，该策略已经在（Gao 等人，2016）中成功应用。矩阵$\mathbf{H}$的条目 $\widetilde{\ }$ $\widetilde{\ }$均可观测，给定$(\boldsymbol{U},\boldsymbol{V})$的 $\mathbf{H}$的期望仍为 $\mathbf{\Theta}^*$，且$\mathbf{Var}[\widetilde{H}_{i,j}|U_i,V_j]\leqslant\big(\sigma^2+\rho^2(1-\mathsf{p})\big)/\mathsf{p}$，其中， σ^2 是条件方差$\mathbf{Var}[H_{i,j}|U_i,V_j]$ 的上界，而 ρ^2 是$(\Theta^*_{i,j})^2$的上界。

2.4 估计图子在估计出矩阵 $\mathbf{\Theta}^*$ 后，重点转向开发图子 W^* 的估计量。为此，对于任意 $n\times m$ 矩阵 $\mathbf{\Theta}$，我们将其关联图子 $W_{\mathbf{\Theta}}$： $[0,1]^2\to[0,1]$ 定义为在每个矩形$I_i\times J_j=[(i-1)/n,\ i/n[\times[(j-1)/m,\ j/m[$,上取常数的函数，其中$(i,\ j)\in[n]\times[m]$，定义如下：

$$W_{\mathbf{\Theta}}(u,v)=\Theta_{i,j}, \quad \text{for all} \quad (u,v)\in I_i\times J_j. \tag{6}$$

这一概念背后的理由如下：当 n 和 m 较大时，顺序统计量 $U_{(i)}$ 和 $V_{(j)}$ 以较高概率落在区间 I_i 和 J_j 内。因此，由 $\Theta_{i,\,j} = W_{\boldsymbol{\Theta}}(U_i,\ V_j)$ 定义的矩阵 $\sim\boldsymbol{\Theta}$，经过行置换和列置换后，与 $\boldsymbol{\Theta}^*$ 一致。这意味着由 W^* 和 $W_{\boldsymbol{\Theta}^*}$ 生成的矩阵是等价的。因此，可以预期这两个图子彼此接近。

此外，对于由 (6) 定义的估计图子，估计误差可以很容易地与矩阵 $\boldsymbol{\Theta}^*$ 的估计误差（以弗罗贝尼乌斯范数衡量）联系起来。事实上，只需简单验证即可得到 $\|W_{\widehat{\boldsymbol{\Theta}}} - W_{\boldsymbol{\Theta}^*}\|_{\mathbb{L}_2} = \|\boldsymbol{\Theta} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}/\sqrt{nm}$，由此可得

$$\delta(W_{\widehat{\boldsymbol{\Theta}}}, W^*) \leqslant \|W_{\widehat{\boldsymbol{\Theta}}} - W_{\boldsymbol{\Theta}^*}\|_{\mathbb{L}_2} + \delta(W_{\boldsymbol{\Theta}^*}, W^*) \leqslant \frac{\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}}{\sqrt{nm}} + \delta(W_{\boldsymbol{\Theta}^*}, W^*). \quad (7)$$

我们将对 $\widehat{\boldsymbol{\Theta}} = \widehat{\boldsymbol{\Theta}}^{\mathsf{LS}}$ 和 $\widehat{\boldsymbol{\Theta}} = \widehat{\boldsymbol{\Theta}}^{\mathsf{EWA}}$都使用这一不等式。为简化记号，我们常写 $\widehat{W}_{\mathsf{LS}}$ 和 $\widehat{W}_{\mathsf{EWA}}$，而不写 $W_{\widehat{\boldsymbol{\Theta}}^{\mathsf{LS}}}$和 $W_{\widehat{\boldsymbol{\Theta}}^{\mathsf{EWA}}}$。由（7）给出的分解将图函数估计误差分为两个部分：估计条件均值矩阵 $\boldsymbol{\Theta}^*$ 的误差，以及用分段常数函数 $W_{\boldsymbol{\Theta}^*}$逼近 W^* 所产生的偏差。前者是唯一依赖于估计程序和关于噪声的概率假设的项；下一节将在各种此类假设下对其进行分析。后者仅取决于图子的“光滑性”。下面的结果使我们能够评估这一项。

命题 1 $\Theta^*_{ij} = W^*\ U_i\ V_j \quad i \in [n] \quad j \in [m] \quad W^*\ [0,1]^2 \to$。设 $(,)$ 对于 $[A,B]$ $A\ B$ $-\infty \leqslant A < B \leqslant +\infty$ 和，其中：对于某个 $,$，使得 。1.（分段常数图子）更确切地说，对于 $0 = a_0 < \ldots < a_K = 1$ 和 $0 = b_0 < \ldots < b_L = 1$，函数 W^* 在每个矩形 $[a_k, a_{k+1}[\times[b_\ell, b_{\ell+1}[$上均为常数。若定义 $W_{\boldsymbol{\Theta}^*}\ [0,1]^2 \to [A,B] \quad W_{\boldsymbol{\Theta}^*}\ u\ v\ = \Theta^*_{i\,j} \quad u\ v\ \in [\ i-1\ /n\ i/n[\times[\ j-:$ 为 $(,)$,对所有 $(,)$ 1 $(),()/n, j/n[$, 则

$$\mathbb{E}[\delta(W_{\boldsymbol{\Theta}^*}, W^*)^2]^{1/2} \leqslant \frac{(B-A)}{\sqrt{2}}\left(\sqrt{\frac{K}{n}} + \sqrt{\frac{L}{m}}\right)^{1/2}.$$

2.（赫尔德连续图子）如果图子 W^* 是 α-光滑的，即 W^* 属于赫尔德类[2] $\mathbb{H}_{\alpha,\ \mathcal{L}}$，其中 $\alpha \in (0,1]$ 且 $\mathcal{L} > 0$，那么

$$\mathbb{E}[\delta(W_{\boldsymbol{\Theta}^*}, W^*)^2]^{1/2} \leqslant \frac{2\mathcal{L}}{n^{\alpha/2}} + \frac{2\mathcal{L}}{m^{\alpha/2}}.$$

该结果的证明推迟至第8.1节，其基本沿用了（Klopp 等人，2017，命题 3.2）的步骤；该命题仅讨论对称函数。顺带一提，我们结果中的常数小于文献中已有结果中的常数。

3 有限样本风险界 我们在第2 节介绍了用于估计矩阵 $\boldsymbol{\Theta}^*$ 的最小二乘估计量和指数加权聚合估计量，以及相应的图子估计量。本节给出这些估计量的风险上界。这些风险界的主要目的是突出当 n, m 较大而 ρ, σ 较小时估计量的表现（σ 表示噪声幅度）。

[2] $\mathbb{H}_{\alpha,\ \mathcal{L}}$是满足以下条件的函数集合 W: $[0,1]^2 \to \mathbb{R}$: $|W(x,\ y)\ - W(x',\ y')| \leqslant \mathcal{L}((x-x')^2 + (y-y')^2)^{\alpha/2}$对所有 $x,\ y,\ x',\ y' \in [0,1]$均成立。

3.1 最小二乘估计量的风险界 $\widehat{\boldsymbol{\Theta}}^{\mathsf{LS}}$ 我们首先给出 $\boldsymbol{\Theta}^*$ 估计量的风险界。为此，在陈述和证明中，我们不失一般性地将 $\boldsymbol{\Theta}^*$ 视为确定性矩阵；因此，我们要求 $H_{i,j}$ 相互独立，而不只是要求它们在给定 $\boldsymbol{U}, \boldsymbol{V}$ 的条件下相互独立。

定理*1*。设 n、m、n_0、m_0、L、K 为正整数，使得 $L \geqslant 2$、$K \geqslant 2$ $3 \leqslant$ 和 $3 \leqslant m_0 \leqslant m$。设 $\boldsymbol{H}$为一个 $n \times m$ 随机矩阵，其独立元素满足：对任意 $i \in [n]$、$j \in [m]$ 以及某个 $\rho > 0$，均有 $\mathbb{E}[H_{ij}] \in [0, \rho]$。此外，假设随机变量 $(H_{ij} - \mathbb{E}[H_{ij}])$ 满足 (σ^2, b)-伯恩斯坦条件。那么，由式*(4)* 定义的均值矩阵 $\boldsymbol{\Theta}^* = \mathbb{E}[\boldsymbol{H}]$，的最小二乘估计量 $\widehat{\boldsymbol{\Theta}}^{\mathsf{LS}}$ 满足精确的预言机不等式

$$\frac{\mathbb{E}\big[\|\widehat{\boldsymbol{\Theta}}^{\mathsf{LS}} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}^2\big]^{1/2}}{\sqrt{nm}} \leqslant \inf_{\boldsymbol{\Theta} \in \mathcal{T}_{n_0,m_0}^{K,L}} \frac{\|\boldsymbol{\Theta} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}}{\sqrt{nm}} + \big(25\sigma^2 + 4b\rho\big)^{1/2} r_{n,m}(K,L),$$

其中 $r_{n,m}(K,L)$ 由

$$r_{n,m}(K,L) = \Big(\frac{3KL}{nm} + \frac{\log K}{m} + \frac{\log L}{n}\Big)^{1/2} \tag{8}$$

前提是 $\psi_{n,\,m}(n_0,\ m_0) := \frac{3}{m_0}\log(en/n_0) + \frac{3}{n_0}\log(em/m_0) \leqslant (\sigma/b)^2$。

所有可能估计量的极小极大风险下界将于第5节给出，并表明定理1 中的风险界在不同情形下均达到最优速率。这里提一下，在$\mathbf{E}[H_{i,j}] \in [0, \rho]$ 对应次高斯分布这一特殊情形下，可以去掉该条件，而定理的结论仍然成立。

定理 1 针对一般分布进行陈述，因此，考察其在引言中提到的常见分布$H_{i,\,j}$ 情形下的推论是有帮助的。

推论 1. 假设 n、m、n_0、m_0、L、K 满足定理1中要求的条件。

1. 如果 H_{ij} 彼此独立，服从伯努利分布——或任何其他支撑集为 $[0,1]$ 的分布——且为均值为 $\Theta^*_{i,\,j} \leqslant \rho$ 的随机变量，那么它们满足 $(\rho,\ 1/3)$-伯恩斯坦条件，因此，

$$\frac{\mathbb{E}\big[\|\widehat{\boldsymbol{\Theta}}^{\mathsf{LS}} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}^2\big]^{1/2}}{\sqrt{nm}} \leqslant \inf_{\boldsymbol{\Theta} \in \mathcal{T}_{n_0,m_0}^{K,L}} \frac{\|\boldsymbol{\Theta} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}}{\sqrt{nm}} + 9\sqrt{\rho}\Big(\frac{KL}{nm} + \frac{\log K}{m} + \frac{\log L}{n}\Big)^{1/2} \tag{9}$$

前提是 $\frac{1}{m_0}$ 对数$(en/n_0) + \frac{1}{n_0}$ 对数$(em/m_0) \leqslant 3\rho$。

2. 如果对于某个 $N \in \mathbb{N}$，(NH_{ij}) 是参数为$(N, \Theta^*_{i,j})$的相互独立二项随机变量，使得 $\Theta^*_{i,j} \leqslant \rho$，那么 $H_{i,j}$ 满足$(\rho/N, 1/3N)$-伯恩斯坦条件，因此，

$$\frac{\mathbb{E}\big[\|\widehat{\boldsymbol{\Theta}}^{\mathsf{LS}} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}^2\big]^{1/2}}{\sqrt{nm}} \leqslant \inf_{\boldsymbol{\Theta} \in \mathcal{T}_{n_0,m_0}^{K,L}} \frac{\|\boldsymbol{\Theta} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}}{\sqrt{nm}} + 9\sqrt{\rho}\Big(\frac{KL}{Nnm} + \frac{\log K}{Nm} + \frac{\log L}{Nn}\Big)^{1/2}$$

前提是$\frac{1}{m_0}\log(en/n_0) + \frac{1}{n_0}\log(em/m_0) \leqslant 3N\rho$.

3. 如果 H_{ij} 是均值为 $\Theta^*_{i,\,j}$ 且方差代理为 $\tau^2_{i,\,j} \leqslant \sigma^2$ 的相互独立的次高斯随机变量，那么它们满足$(\sigma^2,\ 0)$伯恩斯坦条件，因此，

$$\frac{\mathbb{E}\big[\|\widehat{\boldsymbol{\Theta}}^{\mathsf{LS}} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}^2\big]^{1/2}}{\sqrt{nm}} \leqslant \inf_{\boldsymbol{\Theta} \in \mathcal{T}_{n_0,m_0}^{K,L}} \frac{\|\boldsymbol{\Theta} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}}{\sqrt{nm}} + 5\sigma\Big(\frac{KL}{nm} + \frac{\log K}{m} + \frac{\log L}{n}\Big)^{1/2}.$$

4. 如果对于某个 $T>0$，(TH_{ij}) 是参数为$T\Theta^*_{i,\ j}\leqslant T\rho$ 的相互独立的泊松随机变量，那么 $H_{i,\ j}$满足$(\rho/T,\ 1/3T)$伯恩斯坦条件，因此，

$$\frac{\mathbb{E}\big[\|\widehat{\mathbf{\Theta}}^{\mathsf{LS}}-\mathbf{\Theta}^*\|_{\mathsf{F}}^2\big]^{1/2}}{\sqrt{nm}}\leqslant\inf_{\mathbf{\Theta}\in\mathcal{T}^{K,L}_{n_0,m_0}}\frac{\|\mathbf{\Theta}-\mathbf{\Theta}^*\|_{\mathsf{F}}}{\sqrt{nm}}+9\sqrt{\rho}\left(\frac{KL}{Tnm}+\frac{\log K}{Tm}+\frac{\log L}{Tn}\right)^{1/2}$$

前提是 $\frac{1}{m_0}\log(\cancel{e}n/n_0)+\frac{1}{n_0}\log(\cancel{e}m/m_0)\leqslant 3T\rho$。

这些风险界中出现的余项表达式可以看作是

$$\text{noise magnitude}\times\frac{\text{size of the parameter space}}{\text{sample size}}.$$

确实，$KL+n$对数 $K+m$对数 L是 $\mathcal{T}^{K,_{n_0}\ L,_{m_0}}$的覆盖数对数的数量级，这是衡量参数空间复杂度的一种常用方法。除了具有启发性之外，这一解释还说明，在某些温和条件下，该上界在乘法常数意义下是最优的。

另请注意，上述风险界所针对的最小二乘估计量不需要知道 ρ、σ 和 b。这解释了为何需要对 ρ 施加条件，要求其不能过小。对于 ρ 的较小取值，我们的证明仍可用于获得最小二乘估计量的风险界。例如，在伯努利模型中，当 $\rho\leqslant\frac{1}{m_0}\log(en/n_0)+\frac{1}{n_0}\log(em/m_0)$ 时，余项与公式 (9) 中的相同，只是将 ρ 替换为 $\frac{1}{m_0}\log(en/n_0)+\frac{1}{n_0}\log(em/m_0)$。然而，对于如此小的 ρ，可以为输出一个所有元素均为零的矩阵的估计量，或为带有约束 $\|\mathbf{\Theta}\|_\infty\leqslant\rho$ 的约束最小二乘估计量，获得更小的风险界（有关此类结果，参见 Gao 等人（2016）；Klopp 等人（2017））。

注释1. 如引言中所述，在$H_{i,\ j}-\mathbb{E}[H_{i,j}]$是 σ-次高斯的条件下，(Gao等人，2016)建立了一个类似于定理1 的上界。其中得到的余项阶数为$(\sigma+\rho)r_{n,m}(K,L)$。由于 σ-次高斯性等价于 (σ^2,0)-伯恩斯坦条件，我们的定理适用于相同的情形，并给出更小的余项，$\sigma r_{n,m}(K,L)$（其与 ρ 无关）。特别值得关注的是，条目为缩放的二项随机变量的情形，如最后一个推论的第二个结论所述。在这种情况下，(Gao等人，2016)的风险界包含一个阶数为$r_{n,m}(K,L)$的余项，因为相互独立的伯努利随机变量的平均值的次高斯范数为常数阶。有趣的是，我们的上界明显更紧，因为其余项包含一个$\sqrt{\rho/N}$的缩减因子。在第 5节中，我们证明了该上界是紧的，至少当 n 和 m 具有相同的数量级时如此。

3.2 EWA 的风险界 $\widehat{\mathbf{\Theta}}^{\mathsf{EWA}}$ 在信号去噪的背景下，Leung 和 Barron(2006) 的开创性工作建立了精确的风险界，但最初仅限于高斯噪声。随后，相关研究扩展到了更广泛的噪声分布，代表性工作包括（Dalalyan 和 Tsybakov，2007，2008，2012；Dalalyan，2020，2022）。本节给出的结果源自（Dalalyan，2022）中的研究成果；据我们所知，这些成果仍是文献中唯一适用于具有非对称噪声分布模型的结果——这类情形类似于伯努利模型和二项模型。

定理 2. 令$\mathbf{H}$*an* $n\times m$矩阵，且其元素相互独立，并令 $\mathbf{\Theta}^*=\mathbf{E}[\mathbf{H}]$。令 $\mathbf{H}'$为 $\boldsymbol{H}$ 的一个独立副本。令 $\mathfrak{P}$ 为四元组的集合 $(K,L,n_0,m_0)=:\ \mathfrak{p}$ 且令 $\widehat{\mathbf{\Theta}}^{\mathsf{LS}}[\mathfrak{p}]$

为由以下定义的最小二乘估计量(4)。令 $\widehat{\boldsymbol{\Theta}}^{\mathsf{EWA}}[\mathfrak{P}]$ 为应用于估计量的指数加权聚合估计量(5)，其应用对象为估计量$\{\widehat{\boldsymbol{\Theta}}^{\mathsf{LS}}[\mathfrak{p}]:\ \mathfrak{p}\in\mathfrak{P}\}$，其中温度参数为 $\beta>0$。令$r_{n,\ m}(K,\ L)$ 如(8)所示。

1.（伯努利/二项模型）假设对于某些 $N\in\mathbb{N}$, $NH_{i,\ j}\sim$二项$(N,\ \Theta^*_{i,\ j})$ 满足$\Theta^*_{i,\ j}\leqslant\rho$ 对每个 $i\in[n]$ 和 $j\in[m]$均成立。对于 $\beta=8/(3N)$，估计量 $\widehat{\boldsymbol{\Theta}}^{\mathsf{EWA}}=\widehat{\boldsymbol{\Theta}}^{\mathsf{EWA}}[\mathfrak{P}]$满足

$$\frac{\mathbb{E}\big[\|\widehat{\boldsymbol{\Theta}}^{\mathsf{EWA}}-\boldsymbol{\Theta}^*\|_{\mathsf{F}}^2\big]^{1/2}}{\sqrt{nm}}\leqslant\min_{\mathfrak{p}\in\mathfrak{P}}\Big\{\inf_{\boldsymbol{\Theta}\in\mathcal{T}[\mathfrak{p}]}\frac{\|\boldsymbol{\Theta}-\boldsymbol{\Theta}^*\|_{\mathsf{F}}}{\sqrt{nm}}+9\sqrt{\frac{\rho}{N}}\,r_{n,m}(K,L)\Big\}+\Big\{\frac{8\log|\mathfrak{P}|}{3Nnm}\Big\}^{1/2}$$

前提是 最大值$_{\mathfrak{p}\in\mathfrak{P}}\big(\frac{1}{m_0}$对数$(en/n_0)+\frac{1}{n_0}$对数$(em/m_0)\big)\leqslant 3N\rho$。

2.（高斯模型）假设 $H_{i,j}\sim\mathcal{N}(\Theta^*_{i,j},\sigma^2_{i,j})$ 其中 $\sigma^2_{i,j}\leqslant\sigma^2$ 对每个 $i\in[n]$ 和$j\in[m]$均成立。对于 $\beta=4\sigma^2$，估计量 $\widehat{\boldsymbol{\Theta}}^{\mathsf{EWA}}=\widehat{\boldsymbol{\Theta}}^{\mathsf{EWA}}[\mathfrak{P}]$ 满足

$$\frac{\mathbb{E}\big[\|\widehat{\boldsymbol{\Theta}}^{\mathsf{EWA}}-\boldsymbol{\Theta}^*\|_{\mathsf{F}}^2\big]^{1/2}}{\sqrt{nm}}\leqslant\min_{\mathfrak{p}\in\mathfrak{P}}\Big\{\inf_{\boldsymbol{\Theta}\in\mathcal{T}[\mathfrak{p}]}\frac{\|\boldsymbol{\Theta}-\boldsymbol{\Theta}^*\|_{\mathsf{F}}}{\sqrt{nm}}+5\sigma r_{n,m}(K,L)\Big\}+\Big\{\frac{\log|\mathfrak{P}|}{nm}\Big\}^{1/2}.$$

要证明该定理的第一个结论，只需将（Dalalyan，2022，推论 4）与定理 1结合起来即可。类似地，第二个结论由（Dalalyan，2022，推论 2）和定理1推出。对于具有有界支集的任意分布以及拉普拉斯分布，分别利用（Dalalyan，2022）中的推论 3 和推论 5，也可以得到类似的结果。遗憾的是，我们目前尚不知道有任何结果能够将这些界扩展到泊松分布以及更广泛的次高斯分布类。

注 2. 定理2中得到的上界表明，当样本量 nm 较大时，由聚合导致的额外误差项并不大。注意，因子 $\log|\mathfrak{P}|$通常并不大。对此集合，一个合理的选择如下：为 $0\leqslant i\leqslant 2\log_2(n/10)$和 $0\leqslant j\leqslant 2\log_2(m/10)$选择几何递增的序列 $K_i=\lfloor 2^{1+i/2}\rfloor$ 和 $L_j=\lfloor 2^{1+j/2}\rfloor$。然后，对于每个 K_i和 L_j，选择 $n_0\leqslant n/K_i$ 和 $m_0\leqslant m/L_j$ 使其具有 $\lfloor 2^{2+\ell/2}\rfloor$的形式。这种选择 $\mathfrak{P}$ 的方法确保 $|\mathfrak{P}|\leqslant 4\log_2^2(n/7)\log_2^2(m/7)$。因此，在几乎所有情形下，项 log $|\mathfrak{P}|$ 的阶都小于 KL；事实上，可以验证，$\log|\mathfrak{P}|\geqslant 12$ 蕴含 $\min(n,m)\geqslant 12\times 10^4$。

3.3 图子估计量的风险界 $\widehat{W}^{\mathsf{LS}}$ 和 $\widehat{W}^{\mathsf{EWA}}$ 将定理 1 中关于均值矩阵估计量的误差 $\boldsymbol{\Theta}^*=\mathbf{E}[\mathbf{H}]$ 以及不等式（7）所建立的风险界进行适当组合，可以得到图子估计量 $\widehat{W}^{\mathsf{LS}}$和 $\widehat{W}_{\mathsf{EWA}}$的风险界。我们将重点讨论两类图子：分段常数图子和赫尔德连续图子，命题 1给出了这两类图子的近似误差。显然，采用这一策略会使项 $\inf_{\boldsymbol{\Theta}\in\mathcal{T}}\|\boldsymbol{\Theta}-\boldsymbol{\Theta}^*\|_{\mathsf{F}}$（即预言机误差）出现在误差界中，其中 $\mathcal{T}$是（3）中定义的按块常数矩阵集合。在分段常数图子的情形下，该预言机误差为零；而在赫尔德连续图子的情形下，则需要对其进行评估，下一命题将完成这一工作。注意，与第 3.1 节和第 3.2节不同，我们现在回到随机 $\boldsymbol{\Theta}^*$ 的原始框架，并且所有期望都包括关于潜变量$\boldsymbol{U}$和 $\boldsymbol{V}$的积分。

命题 2. 令 W^* 为 α-赫尔德连续函数，即 $W^* \in \mathbb{H}_{\alpha,\ \mathcal{L}}$ 对某个 $\alpha \in (0,1], \mathcal{L} > 0$。令 $\Theta^*_{i,\ j} = W^*(U_i,\ V_j)$ 对 $i \in [n]$ 和 $j \in [m]$ 成立。令 $n_0 \geqslant 2$、$m_0 \geqslant 2$、$K \leqslant n/n_0$ 和 $L \leqslant m/m_0$ 为四个整数。则，具有条目 $\Theta_{i,\ j}$ 的 $n \times m$ 矩阵 $\mathbf{\Theta}^*$ 满足

$$\mathbb{E}\left[\inf_{\mathbf{\Theta} \in \mathcal{T}} \frac{\|\mathbf{\Theta} - \mathbf{\Theta}^*\|_{\mathsf{F}}}{\sqrt{nm}}\right]^{1/2} \leqslant \frac{3\mathcal{L}}{2}\left(\frac{1}{K^\alpha} + \frac{1}{L^\alpha}\right),$$

其中 $\mathcal{T} = \mathcal{T}^{K,\ n_0}_{L,\ m_0}$ 是（3）中定义的分块常数矩阵集合。

现在，我们已经具备了陈述本文主要结果所需的全部要素，这些结果量化了估计图子时的误差。我们首先针对最小二乘估计量进行分析，并特别考虑集合 K 和 L 中的参数 $\mathcal{T}$（方程（2）中的最小值在该集合上计算）固定的情形。随后，我们陈述指数加权聚合估计量的结果。

定理 3. 设 $\mathbf{H}$ 为一个 $n \times m$ 随机矩阵，满足假设 *1* 以及某个图子 $W^*\colon\ [0,1]^2 \to [0,\rho]$。假设对于某个常数 $\sigma > 0$，在条件于 $\boldsymbol{U}$，$\boldsymbol{V}$ 的情况下，随机变量 $(H_{ij} - \mathbb{E}[H_{ij}|\boldsymbol{U}, \boldsymbol{V}])$ 满足 $(\sigma^2,\ b)$-伯恩斯坦条件。

1. 假设图子 W^* 是 (K,L)-分段常数，这意味着对于某些整数 $K, L \geqslant 2$ 且对于 $0 = a_0 < \ldots < a_K = 1$, $0 = b_1 < \ldots < b_L = 1$，使得

$$\Delta^{(K)} := \min_{k \in [K]} |a_k - a_{k-1}| \geqslant \frac{8\log(nK)}{n}, \quad \Delta^{(L)} := \min_{\ell \in [L]} |b_\ell - b_{\ell-1}| \geqslant \frac{8\log(mL)}{m}, \tag{10}$$

函数 W^* 在每个矩形上为常数 $[a_{k-1}, a_k[\times[b_{\ell-1}, b_\ell[$。然后，估计量 $\widehat{W}^{\mathsf{LS}} = W_{\widehat{\mathbf{\Theta}}^{\mathsf{LS}}}$ 其中 $\widehat{\mathbf{\Theta}}^{\mathsf{LS}} = \widehat{\mathbf{\Theta}}^{\mathsf{LS}}_{n_0,\ m_0}[K,L]$ 定义为 (4) 满足

$$\mathbb{E}[\delta(\widehat{W}^{\mathsf{LS}}, W^*)^2]^{1/2} \leqslant \left(27\sigma^2 + 4b\rho\right)^{1/2}\left(\frac{3KL}{nm} + \frac{\log K}{m} + \frac{\log L}{n}\right)^{1/2} + \rho\left(\sqrt{\frac{K}{n}} + \sqrt{\frac{L}{m}}\right)^{1/2},$$

前提是 $\psi_{n,m}(\Delta^{(K,L)}) = {}^{6}\log(2e/\Delta^{(K)}) + 6)\log(2e/\Delta^{(L)})_{n\Delta^{(K)}} \leqslant (\sigma/b)^2$.

2. 假设图子 W^* 是 α-赫尔德连续的，即 $W^* \in \mathbb{H}_{\alpha,\ \mathcal{L}}$ 其中某个 $\alpha \in (0,1]$ 以及 $\mathcal{L} > 0$。假设[3] 节点数量 n，m 满足 $n \geqslant m$ 且

$$\frac{m^{2\alpha+1}}{n\log^4(2n)} \geqslant \mathcal{L}^2 \frac{(4b/\sigma)^{4(\alpha+1)} \vee 3}{(25\sigma^2 + 4b\rho)}. \tag{11}$$

令 $\beta = \alpha/(2\alpha+2)$。那么，存在一组选择 K，L，n_0，m_0，使得最小二乘估计量 $\widehat{W}^{\mathsf{LS}} = W_{\widehat{\mathbf{\Theta}}^{\mathsf{LS}}}$ 满足 $\widehat{\mathbf{\Theta}}^{\mathsf{LS}} = \widehat{\mathbf{\Theta}}^{\mathsf{LS}}_{n_0,\ m_0}[K,L]$

$$\mathbb{E}[\delta(\widehat{W}^{\mathsf{LS}}, W^*)^2]^{1/2} \leqslant 6\mathcal{L}^{1-2\beta}\left(\frac{25\sigma^2 + 4b\rho}{3nm}\right)^{\beta} + \left(\frac{(50\sigma^2 + 8b\rho)\log m}{m}\right)^{1/2} + \frac{4\mathcal{L}}{m^{\alpha/2}}. \tag{12}$$

为了便于理解这些结果，我们做一些说明。首先，可以注意到，将定理的第一项结论应用于伯努利随机变量[4] $H_{i,\ j}$（此时伯恩斯坦条件由 $\sigma^2 = \rho$、$b = 1/3$ 满足），我们得到

$$\mathbb{E}[\delta(\widehat{W}^{\mathsf{LS}}, W^*)^2]^{1/2} \leqslant 10\left(\frac{\rho KL}{nm} + \frac{\rho\log K}{3m} + \frac{\rho\log L}{3n}\right)^{1/2} + \rho\left(\sqrt{\frac{K}{n}} + \sqrt{\frac{L}{m}}\right)^{1/2}, \tag{13}$$

[3] 假设 $n \geqslant m$ 不会造成任何一般性的损失，因为在所考虑的框架中，n 和 m 的作用是对称的。[4] 事实上，如果将伯努利分布替换为任意一个以 $[0,1]$ 为支撑集的分布，完全相同的结果仍然成立。

H_{ij}的分布	取值（σ^2，b）	条件（11）	风险界（12）
伯努利（ρ）	（ρ，1/3）	$\rho^5 \geqslant \frac{\mathcal{L}^2 n \log^4(2n)}{m^3}$	$\frac{11\sqrt{\mathcal{L}}\rho^{1/4}}{(nm)^{1/4}} + \frac{8\sqrt{\rho \text{日志 } m}}{\sqrt{m}} + \frac{4\mathcal{L}}{\sqrt{m}}$
二项式$(N,\rho)/N$	$(\rho/N, 1/3N)$	$\rho^5 \geqslant \frac{\mathcal{L}^2 N n \log^4(2n)}{m^3}$	$\frac{11\sqrt{\mathcal{L}}\rho^{1/4}}{(Nnm)^{1/4}} + \frac{8\sqrt{\rho \text{对数 } m}}{\sqrt{Nm}} + \frac{4\mathcal{L}}{\sqrt{m}}$
次高斯$(\sigma)^2$	$(\sigma^2, 0)$	$\sigma^2 \geqslant \frac{3\mathcal{L}^2 n \text{对数}^4(2n)}{25m^3}$	$\frac{11\sqrt{\mathcal{L}\sigma}}{(nm)^{1/4}} + \frac{8\sigma\sqrt{\text{对数 } m}}{\sqrt{m}} + \frac{4\mathcal{L}}{\sqrt{m}}$
泊松$(T\rho)/T$	$(\rho/T, 1/3T)$	$\rho^5 \geqslant \frac{\mathcal{L}^2 T n \text{对数}^4(2n)}{m^3}$	$\frac{11\sqrt{\mathcal{L}}\rho^{1/4}}{(Tnm)^{1/4}} + \frac{8\sqrt{\rho \text{对数 } m}}{\sqrt{Tm}} + \frac{4\mathcal{L}}{\sqrt{m}}$

表格 1*Lipschitz* 连续图子和各种分布的上界。

前提是 $\frac{2}{3m\Delta^{(L)}}\log(2e/\Delta^{(K)}) + \frac{2}{3n\Delta^{(K)}}\log(2e/\Delta^{(L)}) \leqslant \rho$。在平衡情形 $n = m$，$K = L$ 且 $\Delta^{(K)} = \Delta^{(L)} = 1/K$，式（13）可简化为

$$10\sqrt{\rho}\,\frac{K}{n} + 9\Big(\frac{\rho \log K}{n}\Big)^{1/2} + \sqrt{2}\rho\Big(\frac{K}{n}\Big)^{1/4},$$

只要 $\rho \geqslant 2(K/n)\log(2eK)$。最后一个表达式与（Klopp 等人，2017，推论 3.3.i）中为单部图的图子所确立的速率同阶。此外，该结果在关于观测值的更一般条件下仍然成立，并给出了常数的明确取值。

其次，在 $n = m$ 趋于无穷且假设 $\mathcal{L}, \sigma, b$ 以及 ρ 均为一阶量的情况下，可以更仔细地考察式（12）中出现的三个项的数量级。于是，第一项的数量级为 $(n^2)^{-\alpha/(2\alpha+2)}$，已知这是基于 n^2个观测值估计一个 α-赫尔德连续的、$d = 2$ 变量回归函数的极小极大最优速率。第二项的数量级为 $(\log n/n)^{-1/2}$，当$\alpha < 1$时，它被第一项所支配；当 $\alpha = 1$ 时，除一个对数因子外，它具有最优数量级。第三项的数量级为 $n^{-\alpha/2}$，当 $\alpha = 1$ 时具有最优数量级，并且对于所有 $\alpha < 1$，它都是该和式中最大的项。

据我们所知，关于是否存在能够实现比 $n^{-\alpha/2}$ 更快收敛速率的 Hölder 连续图子估计量这一问题，目前仍未解决。普遍观点认为，该项是不可避免的，这是由于无法观测协变量 $\boldsymbol{U}$，$\boldsymbol{V}$所必须付出的代价。请注意，当和具有相同阶时，这种由信息缺失导致的恶化程度，即风险界 (12 $n^{\alpha(1-\alpha)/(2\alpha+2)} \leqslant$) 中第三项与第一项之比，其阶为 $n^{0.086}$。从实际角度来看，这种恶化并不显著，因为即使对于 $n = 10^9$，$n^{0.086} \leqslant 6$。

还可以考察定理第二个结论在 $\boldsymbol{U}$、$\boldsymbol{V}$ 的各种（条件于 $H_{i,\,j}$）分布下的结果。对于 $\alpha = 1$（$\beta = 1/4$），即 Lipschitz 连续图子，条件（11）和不等式（12）列于表 1。

为了结束本节，我们陈述结合定理2和定理3可以得到的风险界。为使陈述简洁，这里只介绍分段常数图子的情形。

推论2。令 $n \leqslant m$，并令 H 为满足假设1的一个 $n \times m$ 随机矩阵，其对应某个 (K, L)-分段常数图子 $W^*\colon [0,1]^2 \to [0, \rho]$。这意味着，对于某些整数 K、$L \geqslant 2$ 以及满足 (10)的 $0 = a_0 < \ldots < a_K = 1$、$0 = b_1 < \ldots < b_L = 1$，函数 W^* 在每个矩形 $[a_{k-1}, a_k[\times [b_{\ell-1}, b_\ell[$ 上均为常数。令 $\mathfrak{P}$ 按照注 2 进行选取，并令 $\widehat{W}^{\mathsf{EWA}}[\mathfrak{P}] = W_{\widehat{\Theta}^{\mathsf{EWA}}[\mathfrak{P}]}$。

1.（伯努利/二项式模型）假设对于某个 $N \in \mathbb{N}$，条件于 $(\boldsymbol{U}, \boldsymbol{V})$$NH_{i,j}$服从带有 $\Theta^*_{i,j} = W^*(U_i, V_j) \leqslant \rho$的 二项式$(N, \Theta^*_{i,j})$ 分布，其中对于每个 $i \in [n]$ 和 $j \in [m]$。对于 $\beta = 8/(3N)$，我们有

$$\mathbb{E}[\delta(\widehat{W}^{\mathsf{EWA}}, W^*)^2]^{1/2} \leqslant 9\Big(\frac{2\rho KL + \log\log_2 m}{nm} + \frac{\rho \log K}{m} + \frac{\rho \log L}{n}\Big)^{1/2} + \rho\left(\sqrt{\frac{K}{n}} + \sqrt{\frac{L}{m}}\right)^{1/2}$$

只要 $\frac{\log(3e/\Delta^{(K)})}{m\Delta^{(L)}} + \frac{\log(3e/\Delta^{(L)})}{n\Delta^{(K)}} \leqslant N\rho$。

2.（高斯模型）假设，在给定 $(\boldsymbol{U}, \boldsymbol{V})$的条件下，条目 $H_{i,j}$服从高斯 $\mathcal{N}(W^*(U_i, V_j)\ \sigma^2(U_i, V_j))$ 分布，其中 $|W^*(u, v)| \leqslant \rho$，且 $\sigma^2(u, v) \leqslant \sigma^2$，对每个 $u, v \in [0, 1]$均成立。对于 $\beta = 4\sigma^2$，我们有

$$\mathbb{E}[\delta(\widehat{W}^{\mathsf{EWA}}, W^*)^2]^{1/2} \leqslant 5\sigma\Big(\frac{2KL + \log\log_2 m}{nm} + \frac{\log K}{m} + \frac{\log L}{n}\Big)^{1/2} + 2\rho\left(\sqrt{\frac{K}{n}} + \sqrt{\frac{L}{m}}\right)^{1/2}.$$

4 最小二乘估计量的可处理近似 第（2）节介绍并在前几节中研究的最小二乘估计量，是组合优化问题的一个解，而该问题在计算上难以处理。无法在多项式时间内计算这一估计量。本节旨在介绍一种可处理的算法，用于计算 $\widehat{\boldsymbol{\Theta}}^{\mathsf{LS}}$ 的近似值。当然，无法保证所介绍的算法所得的估计量始终接近 $\widehat{\boldsymbol{\Theta}}^{\mathsf{LS}}$，但在许多情况下，这一点是合理的。

所提出的近似可以看作是用于 k均值聚类的 Lloyd 算法的一个版本（Lloyd, 1982）。为了描述它，我们回顾一下，最小二乘估计量定义为

$$(\widehat{\mathbf{Q}}, \widehat{\mathbf{Z}}^{\text{cust}}, \widehat{\mathbf{Z}}^{\text{prod}})^{\mathsf{LS}} \in \arg \min_{\substack{\mathbf{Q} \in \mathbb{R}^{K \times L} \\ \mathbf{Z}^{\text{cust}} \in \mathcal{Z}(n,K,n_0) \\ \mathbf{Z}^{\text{prod}} \in \mathcal{Z}(m,L,m_0)}} \left\|\mathbf{H} - \mathbf{Z}^{\text{cust}}\mathbf{Q}(\mathbf{Z}^{\text{prod}})^\top\right\|_{\mathsf{F}}^2. \tag{14}$$

事实证明，当固定 $\mathcal{L}$ 的三个参数 Q、Z^{cust}、Z^{prod} 中的两个时，关于第三个参数的最小化问题便变得易于处理。因此，我们可以使用下面的交替最小化算法，并保证代价函数 $\mathcal{L}(\mathrm{Z}^{\text{cust}}, \mathrm{Q}, \mathrm{Z}^{\text{prod}}) = \|\mathrm{H} - \mathrm{Z}^{\text{cust}}\mathrm{Q}(\mathrm{Z}^{\text{prod}})^\top\|_{\mathsf{F}}^2$ 在每次迭代时都会下降。针对存在潜在结构时的估计与检测，文献中研究了该算法的不同版本（Coja-Oghlan，2010；Lei 和 Rinaldo，2015；Lu 和 Zhou，2016；Giraud 和 Verzelen，2019）。

算法 1用于近似 LSE 的 Lloyd 交替最小化算法（14）

要求：Zcust、Zprod 分别为左侧和右侧簇矩阵，其元素取自 $\{0, 1\}$；H 为数据矩阵。确保：(Zcust, Q, Zprod) 是 $\mathcal{L}(\cdot, \cdot, \cdot)$ 的局部极小值。重复：1. 计算 Q $=$ (Zcustnorm)$^\top$HZprodnorm，其中 Zcustnorm 是将 Zcust 的列按 ℓ^1-范数（列中 1 的数量）归一化后的矩阵，Zprodnorm 同理。2. 更新 Zcust，使其最小化 Z $\mapsto \mathcal{L}$(Z, Q, Zprod)。3. 更新 Zprod，使其最小化 Z $\mapsto \mathcal{L}$(Zcust, Q, Z)

本节其余部分将详细介绍该算法的每一步，以及初始化和停止准则。

第二个参数中的最小化 $\mathbf{Q}$。当簇已知时，即我们知道矩阵 $\mathbf{Z}^{\text{cust}} \in \mathcal{Z}(n,\ K,\ n_0)$ 和 $\mathbf{Z}^{\text{prod}} \in \mathcal{Z}(m,\ L,\ m_0)$，见（1），方程的解为

$$\widehat{\mathbf{Q}} = \arg\min_{\mathbf{Q}\in\mathbb{R}^{K\times L}} \mathcal{L}(\mathbf{Z}^{\text{cust}}, \mathbf{Q}, \mathbf{Z}^{\text{prod}})$$

易于计算：每个条目 Q_{kl} 是属于由 $\mathbf{Z}^{\text{cust}}$ 和 $\mathbf{Z}^{\text{prod}}$ 定义的 (k,l)-块的系数 H_{ij} 的平均值（如果 $Z^{\text{cust}}_{ik} = 1$ 和 $Z^{\text{prod}}_{jl} = 1$，则系数 (i,j) 位于块 (k,l) 中）。形式上，这等价于 $\mathbf{Q} = (\mathbf{Z}^{\text{cust}}_{\text{norm}})^\top \mathbf{H} \mathbf{Z}^{\text{prod}}_{\text{norm}}$，其中 $\mathbf{Z}_{\text{norm}} = \mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}$ 对于 $\mathbf{Z} \in \{\mathbf{Z}^{\text{cust}},\ \mathbf{Z}^{\text{prod}}\}$。

关于 $\mathbf{Z}^{cust}$ 的最小化。现在我们关注关于 $\mathbf{Z}^{\text{cust}} \in \mathcal{Z}(n, K, n_0)$ 最小化代价函数 $\mathcal{L}(\mathbf{Z}^{\text{cust}}, \mathbf{Q}, \mathbf{Z}^{\text{prod}})$ 的问题。首先考虑相对简单的情形 $n_0 = 0$，其中左侧簇的数量没有限制。我们的目标是找到 $\mathbf{Z}^{\text{cust}} \in \mathbb{R}^{n\times K}$，使 $\mathbf{Z} \mapsto \mathcal{L}(\mathbf{Z}, \mathbf{Q}, \mathbf{Z}^{\text{prod}})$ 在约束 $\mathbf{Z} \in \{0,1\}^{n\times K}$ 且 $\mathbf{Z}\mathbf{1}_K = \mathbf{1}_n$（即 $\mathbf{Z}$ 的每一行只有一个等于 1 的元素）下达到最小。定义 $C^{\text{prod}}_\ell = \{j \in [m] : Z^{\text{prod}}_{i\ell} = 1\}$ 为第 ℓ 个右侧簇，并引入记号

$$\bar{H}^{\text{item}}_{i\ell} = \frac{1}{|C^{\text{prod}}_\ell|} \sum_{j\in C^{\text{prod}}_\ell} H_{ij} \quad \text{and} \quad \mathbf{D} = \mathsf{diag}(|C^{\text{prod}}_\ell|)_{\ell\in[L]}.$$

简单代数运算可得

$$\mathcal{L}(\mathbf{Z}, \mathbf{Q}, \mathbf{Z}^{\text{prod}}) = \sum_{i=1}^{n} \|H_{i,\bullet} - \boldsymbol{Z}^\top_{i,\bullet}\mathbf{Q}(\mathbf{Z}^{\text{prod}})^\top\|_2^2,$$

其中 $\boldsymbol{Z}_{i,\bullet}$ 是 $\mathbf{Z}$ 的第 i 行。由于 $\boldsymbol{Z}_{i,\bullet}$ 只能有一个非零元素，且该元素必须等于 1，因此 $\boldsymbol{Z}^\top_{i,\bullet}\mathbf{Q}(\mathbf{Z}^{\text{prod}})^\top$ 实际上等于 $\mathbf{Q}(\mathbf{Z}^{\text{prod}})^\top$ 的一行。这意味着 $Z^{\text{cust}}_{ij} = \mathbb{1}(j = k_i)$，其中

$$\widehat{k}_i = \arg\min_{k\in[K]} \left\|H_{i,\bullet} - \mathbf{Q}_{k,\bullet}(\mathbf{Z}^{\text{prod}})^\top\right\|_2^2 = \arg\min_{k\in[K]} \sum_{\ell=1}^{L} \sum_{j\in C^{\text{prod}}_\ell} (H_{ij} - Q_{kl})^2.$$

我们可以将上面的 k_i 表达式改写为

$$\widehat{k}_i = \arg\min_{k\in[K]} \|\mathbf{D}^{1/2}(\bar{\boldsymbol{H}}^{\text{item}}_{i,\bullet} - \boldsymbol{Q}_{k,\bullet})^\top\|_2^2.$$

因此，为了确定 $\mathbf{Z}^{\text{cust}}$，只需计算矩阵 $\bar{\mathbf{H}}^{\text{item}} \in \mathbb{R}^{n\times L}$，然后为 $\bar{\mathbf{H}}^{\text{item}}$ 的每一行找到 $\mathbf{Q}$ 中与之最接近的一行。当然，对于已知 $\mathbf{Z}^{\text{cust}}$ 和 $\mathbf{Q}$ 的情况，同样的步骤也适用于最小化 $\mathcal{L}$（关于 $\mathbf{Z}^{\text{prod}}$）。

回到一般情形 $n_0 \geqslant 0$。在这种情况下，我们证明，可以通过求解一个线性规划来实现关于 $\mathbf{Z}^{\text{cust}}$ 的最小化 $\mathcal{L}$。事实上，定义

$$\phi(\mathbf{Z}) = -2\,\mathrm{tr}(\mathbf{Z}\mathbf{Q}(\mathbf{Z}^{\text{prod}})^\top\mathbf{H}^\top) + \sum_{k=1}^{K} \mathbf{1}_n^\top \boldsymbol{Z}_{\bullet,k}\boldsymbol{Q}_{k,\bullet}\mathbf{D}\boldsymbol{Q}_{k,\bullet}$$

$$\widetilde{\mathcal{Z}}(n, K, n_0) = \left\{\mathbf{Z} \in [0,1]^{n\times K} : \mathbf{Z}\mathbf{1}_K = \mathbf{1}_n \text{ and } \min_{k\in[K]} \mathbf{1}_n^\top \boldsymbol{Z}_{\bullet,k} \geqslant n_0\right\}.$$

注意到，ϕ 是 $\mathbf{Z}^{\text{cust}}$ 的线性函数，而 $\mathcal{Z}(n,\ K,\ n_0)$ 是一个包含 $\mathcal{Z}(n,\ K,\ n_0)$ 的凸多面体。

命题 3. 以下两个结论成立。

1. 函数 $\mathcal{L}(Z, Q, Z^{prod}) - \phi(Z) = \|H\|_F^2$ 与 Z 无关。~*2.* $\mathcal{Z}(n, K, n_0)$ 的极端点集合为 $\mathcal{Z}(n, K, n_0)$。等价地，$\sim\mathcal{Z}(n, K, n_0)$ 中的一个元素是极端点，当且仅当其所有项均为 *0* 或 *1*。

~集合 $\mathcal{Z}(n、K、n_0)$ 是一个凸多面体，因为它由线性约束定义。一个众所周知的结果（Bertsimas 和 Tsitsiklis，1997，第 65 页，定理 2.7）表明，如果 $\mathcal{L}(\cdot、\mathbf{Q}、\mathbf{Z}^{\text{prod}})$~在多面体 $\mathcal{Z}(n、K、n_0)$ 中有一个极小值点，那么它在其极端点集合 $\mathcal{Z}(n、K、n_0)$ 中至少有一个解。寻找此类解有许多高效的求解器。

初始化。 算法 1 的初始化可能会对最终结果产生重大影响。一种可能的策略是并行运行 N 个具有不同初始值的算法实例，这些初始值随机选取。最终的估计量是在所得的（$\mathcal{L}$）个候选结果中使 N最小的那个。

另一种通常与 Lloyd 算法结合使用的策略基于谱初始化。当图子为分段常数时，所考虑的问题实际上就是二部网络的双随机区块模型。因此，例如，可以通过（Zhou 和 Amini，2019b）中的谱方法获得初始值$(\mathbf{Z}^{\text{cust}},\mathbf{Z}^{\text{prod}})$ 。该方法包括计算矩阵 $\mathbf{H}$ 的正则化版本的 K 截断奇异值分解，然后对 K 截断左奇异向量应用 k- 均值聚类，以获得 $\mathbf{Z}^{\text{cust}}$ 的初始化。对 $\mathbf{Z}^{\text{prod}}$ 的处理过程类似。

停止规则。如前所述，代价函数在迭代过程中非递增，并且其取值属于一个有限集合，因为$(\mathbf{Z}^{\text{cust}},\mathbf{Z}^{\text{prod}})$的配置数量是有限的。这就是为什么从某次迭代开始，代价函数的值会保持不变。然而，该算法可能需要大量迭代才能达到一致性。这表明，如果代价函数的任意两个连续取值相等，或达到最大迭代次数，就应停止迭代。

最后，我们再次强调，无法保证本文所介绍的计算上可行的算法能够提供代价函数的全局最小值 $\mathcal{L}$。然而，正如第6节中的数值示例所示，结果相当令人满意。

5 极小极大风险的下界 本节将证明，在所有可能的估计量中，最小二乘估计量 $\widehat{W}^{\text{LS}}$ 在最坏情况下的收敛速率意义下是最优的。

$$\mathcal{W}_\rho[K,L] = \left\{ W : [0,1]^2 \to [0,\rho] : \exists\, \{I_k\}_{k=1}^K, \{J_\ell\}_{\ell=1}^L \text{ s.t. } W = \sum_{k=0}^{K-1}\sum_{\ell=0}^{L-1} W(a_k, b_\ell)\mathbb{1}_{I_k \times J_\ell} \right\},$$

其中 $I_k = [a_k, a_{k+1})$ 和 $J_\ell = [b_\ell, b_{\ell+1})$ 将 $[0,1)$ 分割成若干区间。下界将在二项式模型中得到证明，但证明中使用的所有技术都可以扩展到引言中介绍的其他模型。

定理 4. 假设在给定$(\boldsymbol{U}, \boldsymbol{V})$的条件下，观测$_{,j}$矩阵 $\mathbf{H}$ 中的元素 H_i相互独立，且服从参数为$(N, W^*(U_i, V_j))$的二项分布。存在通用常数 $c, C > 0$，使得对于任意满足 $KL \geqslant L\log^2 L + K\log^2 K$ 的 $K, L > C$，以及任意 $\rho > 0$，

$$\inf_{\widehat{W}} \sup_{W^*} \mathbb{E}\left[\delta^2(\widehat{W}, W^*)\right]^{1/2} \geqslant c\sqrt{\rho}\left(\frac{KL}{Nnm} \wedge \rho + \frac{1}{N\sqrt{nm}} \wedge \rho\right)^{1/2} + c\rho\left(\sqrt{\frac{K}{n}} + \sqrt{\frac{L}{m}}\right)^{1/2}, \tag{15}$$

^其中，下确界取遍所有可能的估计量 W，上确界取遍所有 $W^* \in \mathcal{W}_\rho[K,L]$。

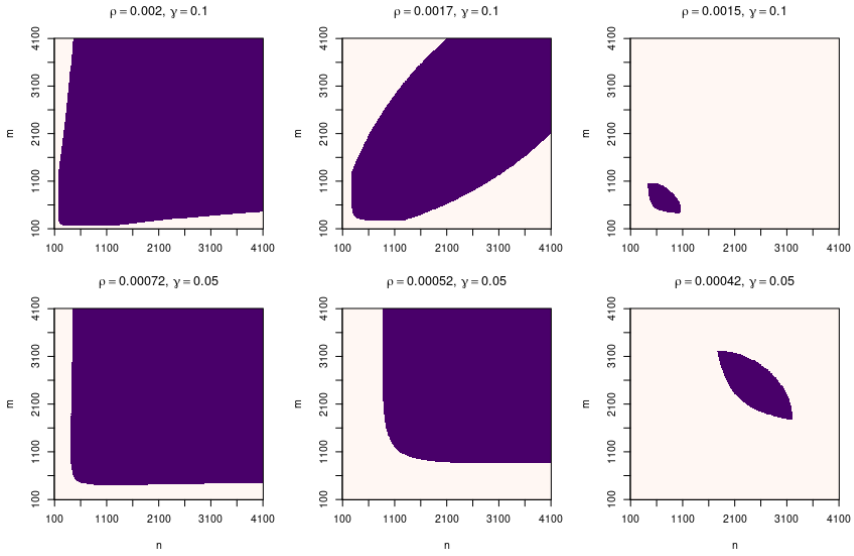


图 2： $N = 1$的最小二乘估计量最优性的示意图。紫色区域对应于在某些固定的 ρ 和 $\gamma = K/n = L/m$取值下， n 和 m的取值范围；在该区域内，下界与上界相差至多一个常数因子。更准确地说，当$\rho\gamma^2 \wedge \rho^2 + \rho(nm)^{-1/2} \wedge \rho^2 + 2\rho^2\sqrt{\gamma}$大于$\rho\gamma^2 + (\rho \log K)/(3m) + (\rho \log L)/(3n) + 2\rho^2\sqrt{\gamma}$的一半时，便是如此。我们注意到，除非 ρ非常小，否则对于 n 和 m的大多数取值，已建立的最小二乘估计量上界与所有估计量的下界相差至多一个常数因子。

(15)的右端应与(13)进行比较。可以观察到，如果 n、m、 K、 L 和 ρ 的取值使得上界中的主导项是$\rho(K/n)^{1/4}$ 和 $\rho(L/m)^{1/4}$中的某一项，那么(15)中的下界与上界具有相同的阶。因此，在这种情况下，图子的最小二乘估计量是极小极大速率最优的。同样地，如果主导项是$(\rho KL/(Nnm))^{1/2}$ 和 $\rho \geqslant KL/(Nnm)$，那么LSE也是极小极大速率最优的。还需注意，如果 ρ 非常小，即小于$KL/(Nnm)$和$(N^2nm)^{-1/2}$，那么(15)中的下界为 ρ阶，这可能远小于针对LSE所得到的上界。这并非证明上的人为结果，$\widehat{\ }$而是反映了这样一个事实：在这种情况下，朴素估计量 $K \equiv 0$ 优于LSE。此外，在对 ρ施加上述约束的条件下，这个朴素估计量事实上是极小极大速率最优的。这基本意味着，对于如此小的 ρ 取值，从统计学角度来看，估计图子的问题并无实际意义。图2描绘了 n 和 m 取值的区域，在这些区域中下界与上界具有相同的阶，从而说明了最小二乘估计量的最优性。同样地，图3展示了在 n 和 m固定时， ρ 和 $\gamma = K/n = L/m$取值的区域，在这些区域中下界大于上界的一半。我们清楚地看到，即使对于非常不平衡的图（m远大于 n），紫色区域也几乎覆盖整个正方形，这意味着在该区域内，最小二乘估计量是极小极大速率最优的。

6 数值实验 本节展示一些数值实验的结果，以说明估计得到的图子 $\widehat{W}^{\mathsf{LS}}$ 的误差表现及其对模型不同参数的依赖关系。我们首先考虑分段常数图子的情形，并研究矩阵 $\boldsymbol{\Theta}^*$的估计误差。我们考察该误差对 n 的依赖关系，其中 (ρ, K, L) 取不同值（假设 $m = n/2$），并考察其对

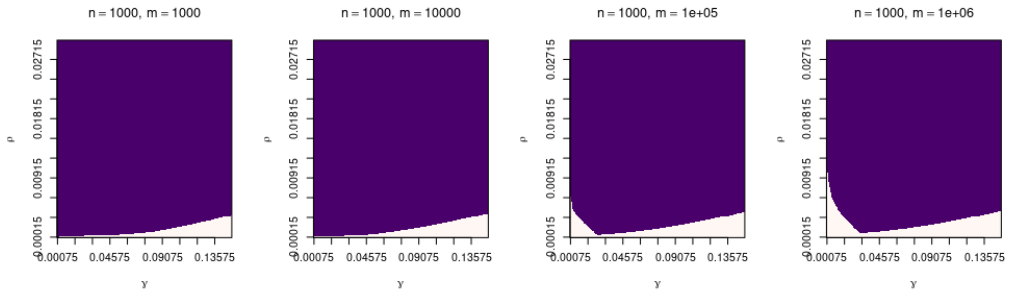


图 3：最小二乘估计量最优性的示意图。对于某些固定的 n 和 m，紫色区域对应 ρ 和 $\gamma = K/n = L/m$的取值，在这些取值下，下界与上界相差一个常数因子。我们观察到，除非 ρ 非常小，否则上界和下界具有相同的阶。

稀疏性参数 ρ 的依赖关系，其中 (n, m, K, L)取不同值。随后，我们展示 Hölder 连续图子的估计结果；此时，按照理论结果的建议，参数 K 和 L 分别选作 n 和 m 的函数。

6.1 分段常数矩阵的估计误差 $\mathbf{\Theta}^*$ 我们报告两种不同实验设定的结果，分别称为随机图子和余弦图子。两种设定的区别仅在于图子的选择不同。在这两种情况下，图子为分段常数所在的划分是由形如$[\,(k-1)\,/K,\ k/K) \times [\,(\ell-1)\,/L,\ \ell/L)$ 的矩形所诱导的规则划分，其中 $k \in [K]$ 和 $\ell \in [L]$。在随机图子设定中，W^*的值在0和 ρ之间随机选取，而在余弦图子设定中，W^*定义为

$$W^*(u,v) = \frac{2\rho}{3} + \frac{\rho}{3}\cos\big(3\pi\lfloor Ku\rfloor\lfloor Lv\rfloor\big), \qquad \forall u, v \in [0,1].$$

这两种设置下获得的结果分别如图 4、图 5和6所示。在每个实验中，我们选择 $m = n/2$，并计算了 50 次独立重复实验所得平方误差 $\frac{1}{nm}\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_{\mathsf{F}}^2$ 的中位数。估计量 $\widehat{\mathbf{\Theta}}$ 由算法 1计算得到。

为提高可读性，图中的误差采用对数尺度呈现。为检验数值结果与理论结果的一致性，我们绘制了定理1上界中出现的余项（绿色）。我们还展示了预言机误差（红色曲线），其对应于利用真实左侧簇矩阵和右侧簇矩阵的知识构建的最佳伪估计量的误差。在这种情况下，我们只计算了块平均值。标签“谱”和“随机”指的是算法所采用的初始化过程。为了展示不确定性，我们绘制了分别对应于 0.1 和0.9阶分位数的彩色区域。（在某些情况下，人们可能会对该区域随n 增大而扩大的事实感到惊讶；这是对数尺度造成的伪象。）在算法 1 中，我们选择了 $\gamma = 10^{-3}$，并将最大迭代次数设为 40。

在这些实验结果中，一个明显的趋势显现出来：“谱”版本的误差随着 n 的增加而持续减小。此外，随着 ρ 取值增大，且 n/K 和 m/L 均较高时，估计量的误差接近预言机的误差，这一模式符合直觉预期。较大的 ρ 意味着更多的连边，从而带来更准确的估计。同样，n/K 和m/L 取值越高，估计准确度也越高。

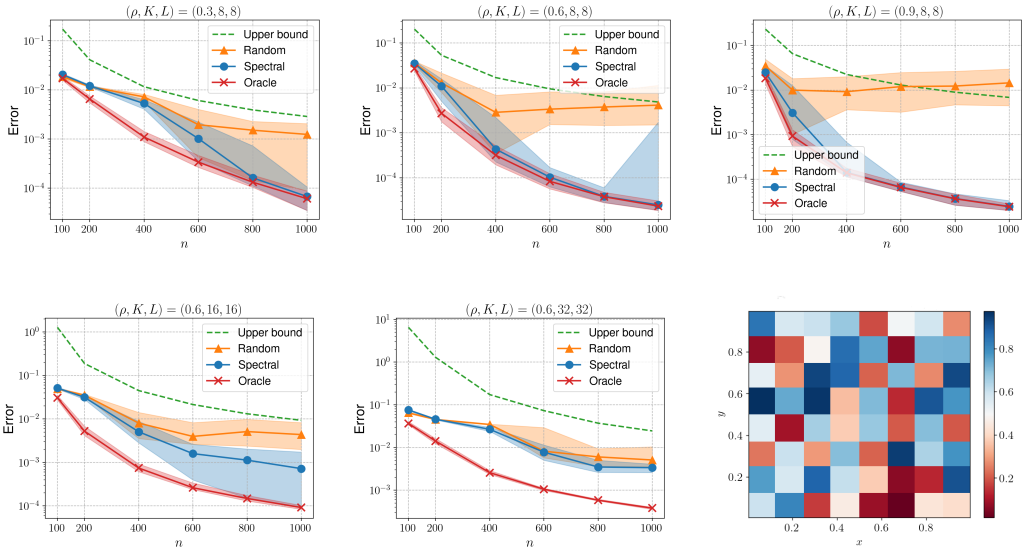


图 4：随机图子设定中的图函数估计。

相反，该算法的“随机”版本表现得更加不稳定。在大多数情况下，它的误差都高于“谱”版本，尤其是在 n/K和 m/L 足够大时。

在余弦图子设定中，我们在图 6 中展示了误差随 ρ 变化的情况。对于较小的 ρ，随机初始化似乎优于谱初始化。此外，对于两种初始化方式，误差都会随 $\rho \in [0, 1/2]$ 增大。出现这种情况的原因是，我们计算的估计量试图模仿预言机估计量；后者知道簇，并通过计算簇内平均值来估计矩阵 $\mathbf{\Theta}^*$。因此，如果 $\mathbf{\Theta}^* = (\mathbf{Z}^{\text{cust}})^* \mathbf{Q}^* (\mathbf{Z}^{\text{prod}})^*$，其中矩阵 $(\mathbf{Z}^{\text{cust}})^*$ 和 $(\mathbf{Z}^{\text{prod}})^*$ 表示簇，那么预言机 $\widehat{\mathbf{\Theta}}_o = (\mathbf{Z}^{\text{cust}})^* \widehat{\mathbf{Q}}_o (\mathbf{Z}^{\text{prod}})^*$ 满足

$$\frac{\mathbb{E}\left[\|\widehat{\mathbf{\Theta}}_o - \mathbf{\Theta}^*\|_{\mathsf{F}}^2\right]}{nm} = \sum_{k,l} \frac{Q^*_{kl}(1 - Q^*_{kl})}{nm} = \sum_{k,l} \frac{\rho \widetilde{Q}_{kl}(1 - \rho \widetilde{Q}_{kl})}{nm} = \frac{\rho(\|\widetilde{\mathbf{Q}}\|_{1,1} - \rho \|\widetilde{\mathbf{Q}}\|_{\mathsf{F}}^2)}{nm} \tag{16}$$

在上述公式中，我们使用了矩阵 $\widetilde{\mathbf{Q}} = \mathbf{Q}^*/\rho$，其所有元素都在 $[0, 1]$，和 $\|\widetilde{\mathbf{Q}}\|_{1,1}$ 中，并记为 $\|\mathbf{Q}\|_{1,1}$ 矩阵 $\mathbf{Q}$ 的元素之和。（16）的右侧是 ρ 的一个函数，在 $\rho \in [0, \|\widetilde{\mathbf{Q}}\|_{1,1}/2\|\widetilde{\mathbf{Q}}\|_{\mathsf{F}}^2]$ 区间内递增，在该区间之外递减。

6.2 赫尔德连续图子的估计误差 为说明图子在赫尔德连续情形下的估计量的行为，我们考虑图 7 中展示并由下式给出的函数

$$W^*(u, v) = \frac{\rho}{2}\Big(1 + \exp\big\{-10\big((u - 1/2)^2 + (v - 1/2)^2\big)\big\}\Big).$$

由于此函数是利普希茨连续的，我们有 $\alpha = 1$。

对 n 和 ρ的不同取值，进行 50 次重复实验所得的估计平均平方误差如图 8所示。由于真实距离 δ 的计算难度极高（因为需要对所有保测双射取最小值），我们计算了它的一个近似值 $\widetilde{\delta}$，记为 δ。粗略地说，δ 是通过将 δ 中的最小值替换为

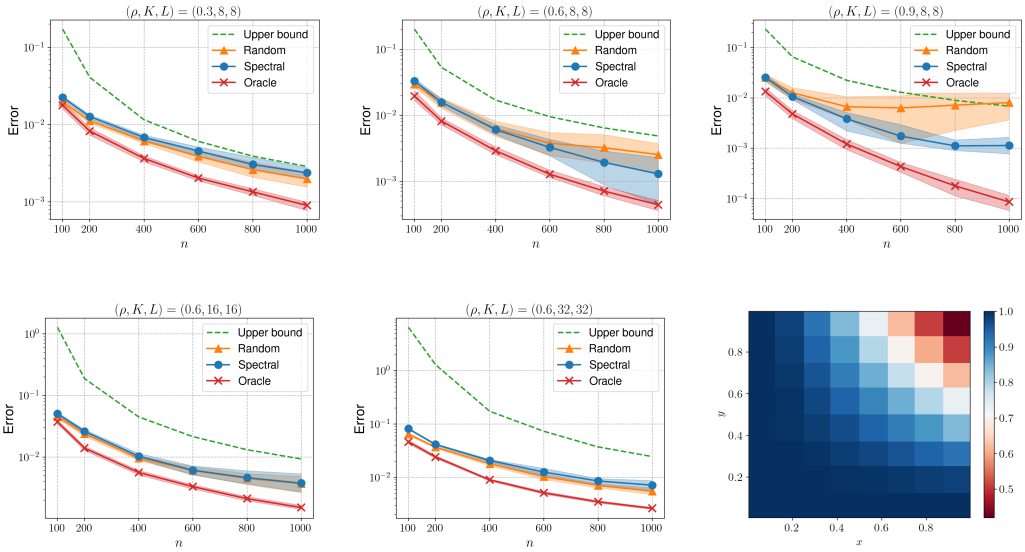


图5：余弦图子设置中的图函数估计，$W^*(x,y) = \frac{2\rho}{3} + \frac{\rho}{3}$余弦$\big(3\pi\lfloor Kx\rfloor\lfloor Ly\rfloor\big)$。

所有双射 τ_1，τ_2 根据双射的特定实例处的代价函数值，τ_{σ_1} 以及 τ_{σ_2}，在命题1的证明中使用。更准确地说，如果 σ_1 和 σ_2 分别是 $[n]$ 和 $[m]$的置换，使得序列 $(U_{\sigma_1^{-1}(i)})_{i\in[n]}$ 和 $(V_{\sigma_2^{-1}(j)})_{j\in[m]}$ 均为非递减序列，那么

$$\tau_{\sigma_1}(u) = \sum_{i=1}^{n} \Big(\frac{\sigma_1(i)-1}{n} - u - \frac{i-1}{n} \Big) \mathbb{1}_{[i-1,i)}(nu), \qquad \tau_{\sigma_1}(1) = \frac{\sigma_1(n)}{n},$$

$$\tau_{\sigma_2}(v) = \sum_{j=1}^{m} \Big(\frac{\sigma_2(j)-1}{m} - v - \frac{j-1}{m} \Big) \mathbb{1}_{[j-1,j)}(nu), \qquad \tau_{\sigma_2}(1) = \frac{\sigma_2(m)}{m}.$$

然后，我们定义

$$\begin{aligned} \widetilde{\delta}^2(\widehat{W}, W^*) &= \big\| W^* \circ (\tau_{\sigma_1} \otimes \tau_{\sigma_2}) - \widehat{W} \big\|_{\mathbb{L}_2}^2 \\ &= \big\| W^* \big\|_{\mathbb{L}_2}^2 - 2 \sum_{i=1}^{n} \sum_{j=1}^{m} \widehat{\Theta}_{ij} \int_{\frac{\sigma_1(i)-1}{n}}^{\frac{\sigma_1(i)}{n}} \int_{\frac{\sigma_2(j)-1}{m}}^{\frac{\sigma_2(j)}{m}} W^*(x,y)\,dx\,dy + \|\widehat{\boldsymbol{\Theta}}\|_{\mathsf{F}}^2. \end{aligned}$$

在数值实验中，最后一个显示公式右侧出现的积分通过黎曼和进行近似。在这种情况下，我们同样观察到，通过蒙特卡洛模拟得到的误差曲线与理论预测的形状相同。有趣且有些出人意料的是，随机初始化的表现与谱初始化一样好。现阶段，我们无法解释这一观察结果。

7 前文所述结果的证明

7.1 定理的证明 *1 (*风险界限适用于 $\widehat{\boldsymbol{\Theta}}^{\mathsf{LS}}$*)* 定义

$$\Pi_{\mathcal{T}}(\boldsymbol{\Theta}^*) = \arg\min_{\boldsymbol{\Theta}\in\mathcal{T}} \|\boldsymbol{\Theta} - \boldsymbol{\Theta}^*\|_{\mathsf{F}},$$

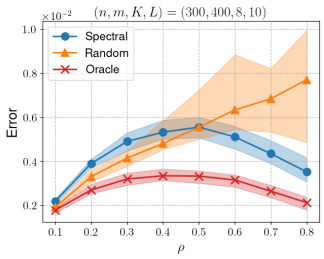


图 6：对于 $W^*(x,y) = \frac{2\rho}{3} + \frac{\rho}{3}\cos(3\pi\lfloor Kx\rfloor\lfloor Ly\rfloor)$， ρ 的估计误差。

$\boldsymbol{\Theta}^*$ 由分块常数矩阵在弗罗贝尼乌斯范数下得到的最佳近似。注意，矩阵 $\Pi_{\mathcal{T}}(\boldsymbol{\Theta}^*)$ 至多有 KL 个不同的元素，其中每个元素都是 $\boldsymbol{\Theta}^*$ 的某个子矩阵中各元素的平均值。由于 $\widehat{\boldsymbol{\Theta}} = \widehat{\mathbf{Z}}^{\text{cust}}\widehat{\mathbf{Q}}(\widehat{\mathbf{Z}}^{\text{prod}})^\top$ 是最小二乘估计量，因此我们有

$$\left\|\mathbf{H} - \widehat{\boldsymbol{\Theta}}\right\|_{\mathsf{F}}^2 \leqslant \left\|\mathbf{H} - \Pi_{\mathcal{T}}(\boldsymbol{\Theta}^*)\right\|_{\mathsf{F}}^2. \tag{17}$$

定义均值为零的“噪声”矩阵 $\mathbf{E} = \mathbf{H} - \mathbb{E}[\mathbf{H}] = \mathbf{H} - \boldsymbol{\Theta}^*$ 并将（17）重写为以下形式

$$\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}^2 \leqslant \|\boldsymbol{\Theta}^* - \Pi_{\mathcal{T}}(\boldsymbol{\Theta}^*)'.\|_{\mathsf{F}}^2 + 2\langle\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*, \mathbf{E}\rangle + 2\langle\boldsymbol{\Theta}^* - \Pi_{\mathcal{T}}(\boldsymbol{\Theta}^*), \mathbf{E}\rangle. \tag{18}$$

由于$\mathbf{E}$ 的期望为零，右侧最后一项也同样如此。我们希望界定 $\langle\boldsymbol{\Theta} - \boldsymbol{\Theta}^*,\mathbf{E}\rangle$的期望。为此，我们定义

$$\widehat{\mathcal{T}} = \left\{\boldsymbol{\Theta} : \exists \mathbf{Q} \in \mathbb{R}^{K\times L} \text{ such that } \boldsymbol{\Theta} = \widehat{\mathbf{Z}}^{\text{cust}}\mathbf{Q}(\widehat{\mathbf{Z}}^{\text{prod}})^\top\right\} \subset \mathcal{T},$$

并令 $\Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*) =$使其最小$_{\boldsymbol{\Theta}\in\widehat{\mathcal{T}}}\|\boldsymbol{\Theta} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}$为 $\boldsymbol{\Theta}^*$ 在 $\mathcal{T}$中的最佳弗罗贝尼乌斯近似。我们使用分解

$$\langle\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*, \mathbf{E}\rangle = \underbrace{\langle\Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*) - \boldsymbol{\Theta}^*, \mathbf{E}\rangle}_{\Xi_1} + \underbrace{\langle\widehat{\boldsymbol{\Theta}} - \Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*), \mathbf{E}\rangle}_{\Xi_2}. \tag{19}$$

引理 1. 在定理1的条件下，我们有

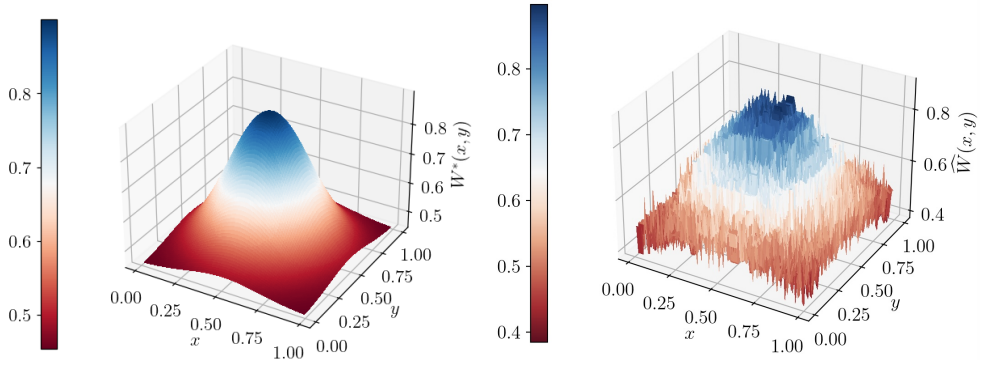
$$\mathbb{E}(\Xi_1) \leqslant \sigma\sqrt{2(n\log K + m\log L + 1)}\,\mathbb{E}[\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}^2]^{1/2} + b\rho(n\log K + m\log L + 1).$$

证明。证明的主要步骤包括：对固定的 $\boldsymbol{\Theta}$，将伯恩斯坦不等式应用于 Ξ_1，而非 $\Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*)$；然后使用并集界，最后对高概率界进行积分。对于第一步，令 $\boldsymbol{\Theta} \in \mathbb{R}^{n\times m}$ 满足 $\Theta_{i,j} \in [0,\rho]$，对每个 $(i,j) \in [n]\times[m]$均成立。根据内积的定义，我们有 $\langle\boldsymbol{\Theta} - \boldsymbol{\Theta}^*, \mathrm{E}\rangle = \sum_{(i,j)\in[n]\times[m]}(\boldsymbol{\Theta} - \boldsymbol{\Theta}^*)_{ij}E_{ij}$。随机变量 E_{ij} 相互独立，并满足 (σ^2, b)-伯恩斯坦条件。由 $(\boldsymbol{\Theta} - \boldsymbol{\Theta}^*)_{ij}$ 组成的 nm-向量的无穷范数有界于 ρ。因此，引理 12 中所述的伯恩斯坦不等式版本表明，对所有 $x > 0$，我们有

$$\mathbb{P}\left(\langle\boldsymbol{\Theta} - \boldsymbol{\Theta}^*, \mathbf{E}\rangle \geqslant \sqrt{2x}\,\sigma\|\boldsymbol{\Theta} - \boldsymbol{\Theta}^*\|_{\mathsf{F}} + b\rho x\right) \leqslant e^{-x}.$$

对于每一对矩阵 $\mathbf{Z} \in \mathcal{Z}_{n,K,n_0}$ 和 $\mathbf{Z}' \in \mathcal{Z}_{m,L,m_0}$，定义 $\Omega_{\mathbf{Z},\mathbf{Z}'} = \{(\widehat{\mathbf{Z}}^{\text{cust}}, \widehat{\mathbf{Z}}^{\text{prod}}) = (\mathbf{Z},\mathbf{Z}')\}$。在事件 $\Omega_{\mathbf{Z},\mathbf{Z}'}$ 上，矩阵 $\Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*)$ 是确定的，其元素是 $\boldsymbol{\Theta}^*$ 中元素的平均值。因此，$0 \leqslant (\Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*))_{ij} \leqslant \|\boldsymbol{\Theta}^*\|_\infty$并且

$$\mathbb{P}\left(\left\{\langle\Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*) - \boldsymbol{\Theta}^*, \mathbf{E}\rangle \geqslant \sqrt{2x}\,\sigma\|\Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*) - \boldsymbol{\Theta}^*\|_{\mathsf{F}} + b\rho x\right\} \cap \Omega_{\mathbf{Z},\mathbf{Z}'}\right) \leq e^{-x}.$$



（a）真实图子 W^* （b）估计的图子 $\widehat{W}$

图 7： $\rho = 0.9$ 的图子表示。

还要注意，$\mathcal{Z}_{n,K,n_0}$ 的基数至多为 K^n。将上一个公式与并集界结合起来，我们得到

$$\begin{aligned}
&\mathbb{P}\Big(\langle \Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*) - \boldsymbol{\Theta}^*, \mathbf{E}\rangle \geqslant \sqrt{2x\rho}\,\sigma \|\Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*) - \boldsymbol{\Theta}^*\|_{\mathsf{F}} + b\rho\sigma^2 x\Big) \\
&\qquad \leqslant \sum_{(\mathbf{Z},\mathbf{Z}')} \mathbb{P}\left(\left\{\langle \Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*) - \boldsymbol{\Theta}^*, \mathbf{E}\rangle \geqslant \sqrt{2x}\,\sigma \|\Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*) - \boldsymbol{\Theta}^*\|_{\mathsf{F}} + b\rho x\right\} \cap \Omega_{\mathbf{Z},\mathbf{Z}'}\right) \\
&\qquad \leqslant K^n L^m e^{-x},
\end{aligned}$$

其中，在上面所示的第一个不等式中，求和遍历集合 $\mathcal{Z}_{n、\ K、\ n_0} \times \mathcal{Z}_{m、\ L、\ m_0}$ 中的 $(\mathbf{Z},\mathbf{Z}')$，因子 $K^n L^m$ 对应于该集合基数的上界。最后，取 $x = n\log K + m\log L + t$，其中 $t > 0$，并使用基本不等式$uv \leqslant \lambda u^2 + v^2/(4\lambda)$，可得

$$\mathbb{P}\left(\Xi_1 \geqslant \lambda \|\Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*) - \boldsymbol{\Theta}^*\|_{\mathsf{F}}^2 + \Big(\frac{\sigma^2}{2\lambda} + b\rho\Big)(n\log K + m\log L + t)\right) \leqslant e^{-t}$$

对任意 $\lambda > 0$。引理 11 下文保证

$$\begin{aligned}
\mathbb{E}(\Xi_1) &\leqslant \lambda \mathbb{E}\Big[\|\Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*) - \boldsymbol{\Theta}^*\|_{\mathsf{F}}^2\Big] + \Big(\frac{\sigma^2}{2\lambda} + b\rho\Big)(n\log K + m\log L + 1) \\
&\leqslant \lambda \mathbb{E}\Big[\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}^2\Big] + \Big(\frac{\sigma^2}{2\lambda} + b\rho\Big)(n\log K + m\log L + 1).
\end{aligned}$$

关于 $\lambda > 0$进行优化，我们得到

$$\mathbb{E}(\Xi_1) \leqslant \sigma\sqrt{2(n\log K + m\log L + 1)}\,\mathbb{E}[\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_{\mathsf{F}}^2]^{1/2} + b\rho(n\log K + m\log L + 1).$$

这就完成了该引理的证明。 □

现在我们转而评估 $\mathbb{E}(\Xi_2)$。为此，我们首先注意到

$$\mathbb{E}(\Xi_2) = \mathbb{E}\big[\langle \Pi_{\widehat{\mathcal{T}}}(\mathbf{H}) - \Pi_{\widehat{\mathcal{T}}}(\boldsymbol{\Theta}^*), \mathbf{E}\rangle\big] = \mathbb{E}\big[\langle \Pi_{\widehat{\mathcal{T}}}(\mathbf{E}), \mathbf{E}\rangle\big] = \mathbb{E}\big[\|\Pi_{\widehat{\mathcal{T}}}(\mathbf{E})\|_{\mathsf{F}}^2\big].$$

引理 2. 在定理*1*的条件下，我们有

$$\mathbb{E}\big[\|\Pi_{\widehat{\mathcal{T}}}(\mathbf{E})\|_{\mathsf{F}}^2\big] \leqslant 4(b+\sigma^2)\big(3KL + n\log K + m\log L\big)\big(2\rho + b\psi_{n,m}(n_0,m_0)\big).$$

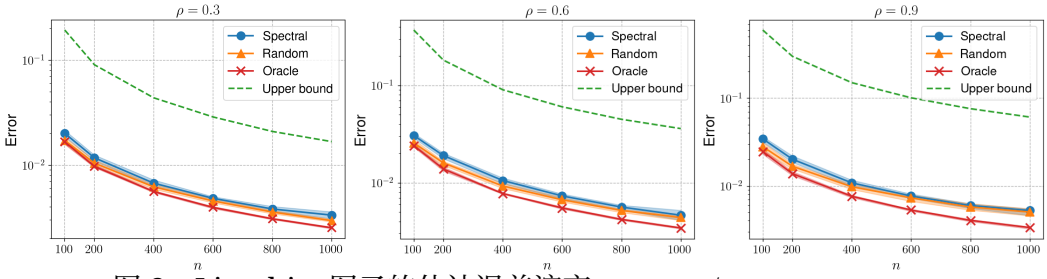


图 8：Lipschitz 图子的估计误差演变，$m = n/2$。

证明。回忆一下 $\psi_{n,\ m}(n_0,\ m_0) = 3\log(en/n_0) + 3\log(em/m_0)\, n_0$。证明思路是依次应用引理 14 和引理15。我们将矩阵向量化，以便只对向量进行处理。为此，考虑一个任意的双射

$$\phi : [n] \times [m] \to [nm]\,.$$

设 $\mathcal{N}$ 和 $\mathcal{M}$ 分别是 $[n]$ 和 $[m]$ 的划分，满足

$$|\mathcal{N}| = K, \quad |\mathcal{M}| = L \qquad \text{and} \qquad \min_{A\in\mathcal{N}} |A| \geqslant n_0 \quad \min_{B\in\mathcal{M}} |B| \geqslant m_0.$$

我们定义 $\mathcal{N}\times\mathcal{M} = \{A\times B\colon\ \ A\in\mathcal{N},\ \ B\in\mathcal{M}\}$ 它是 $[n]\times[m]$ 的一个划分，其基数为 $|\mathcal{N}\times\mathcal{M}| = KL$，并且对每个 $A\times B\in\mathcal{N}\times\mathcal{M}$ 都满足 $|A\times B| \geqslant n_0 m_0$。我们将满足上述条件的所有划分 ϕ（$\mathcal{N}\times\mathcal{M}$）（其中 $\mathcal{N}$ 和 $\mathcal{M}$ 如上所述）构成的族记为 $\mathcal{G}$。

由于 $\Pi_{\widehat{\mathcal{F}}}(\mathbf{E})$ 的元素是 $\mathbf{E}$ 的系数的平均值，如果我们根据映射 ϕ 对 $\mathbf{E}$ 进行向量化，即定义 $(\vec{\mathbf{E}})_i = E_{\phi^{-1}(i)}$ 对所有 $i\in[nm]$ 成立，则有

$$\|\Pi_{\widehat{\mathcal{F}}}(\mathbf{E})\|_{\mathsf{F}}^2 \leqslant \max_{G\in\mathcal{G}} \|\Pi_G \vec{\mathbf{E}}\|_2^2$$

其元素为 $(\vec{\mathbf{E}})_i$，满足引理 14 的假设。因此，根据 $\mathcal{G}$ 上的并集界，我们得到

$$\|\Pi_{\widehat{\mathcal{F}}}(\mathbf{E})\|_{\mathsf{F}}^2 \leqslant 4\big(t + \log(2M|\mathcal{G}|)\big)\big(2\sigma^2 + b\max_{G\in\mathcal{G}} \|\Pi_G\vec{\mathbf{E}}\|_\infty\big) \tag{20}$$

其概率至少为 $1 - 0.5e^{-t}$，其中 $\log\, M \leqslant KL\log\, 12 \leqslant 2.5KL$。

令 $\mathcal{A}$ 为 $\mathcal{G}$中所有划分单元构成的族，并令 $\mathcal{A}_{(s,l)} = \{\phi(A\times B)\in\mathcal{A} : |A| = s \text{ 以及 } |B| = l\}$，其中 $j = (s,l)\in[n]\times[m]$。定义

$$\mathfrak{F} = \frac{\log(4nm)}{n_0 m_0} + \max_{s,l} \frac{\log|\mathcal{A}_{(s,l)}|}{sl} \geqslant \frac{\log(4nm)}{n_0 m_0} + \frac{\log|\mathcal{A}_{(n_0,m_0)}|}{n_0 m_0} \geqslant \frac{7}{n_0 m_0}, \qquad \forall n, m \geqslant 3.$$

根据引理 15，在概率至少为 $1 - 0.5e^{-t}$的事件上，

$$\begin{aligned}
\max_{G\in\mathcal{G}} \|\Pi_G\vec{\mathbf{E}}\|_\infty &\leqslant \max_{A\in\mathcal{A}} \frac{1}{|A|}\Big|\sum_{\ell\in A}(\vec{\mathbf{E}})_\ell\Big| \leqslant \sigma\sqrt{2\Big(\frac{t}{n_0 m_0} + \mathfrak{F}\Big)} + b\Big(\frac{t}{n_0 m_0} + \mathfrak{F}\Big) \\
&\leqslant \sigma\sqrt{2\mathfrak{F}(1 + t/7)} + b\mathfrak{F}(1 + t/7) \\
&\leqslant \sigma\sqrt{2\mathfrak{F}}\,(1 + t/14) + b\mathfrak{F}(1 + t/7).
\end{aligned} \tag{21}$$

最后，结合（20）和（21），我们以至少 $1 - e^{-t}$ 的概率得到，

$$\|\Pi_{\widehat{\mathcal{F}}}(\mathbf{E})\|_{\mathsf{F}}^2 \leqslant 4\big(t + \log(2M|\mathcal{G}|)\big)\big(2\sigma^2 + b\sigma\sqrt{2\mathfrak{F}}\,(1 + t/14) + b^2\mathfrak{F}(1 + t/7)\big).$$

利用引理 11，我们得到以下期望上界

$$\begin{aligned}\mathbb{E}\big[\|\Pi_{\widehat{\mathcal{T}}}(\mathbf{E})\|_{\mathsf{F}}^2\big] &\leqslant 4\big(1+\log(2M|\mathcal{G}|)\big)\big(2\sigma^2+b\sigma\sqrt{2\mathfrak{F}}\,(1+1/7)+b^2\mathfrak{F}(1+2/7)\big)\\ &\leqslant \big(1+\log(2M|\mathcal{G}|)\big)\big(9\sigma^2+16\mathfrak{F}b^2\big).\end{aligned}$$

对于任意 $(s,\ l)$ 使得 $n_0\leqslant s\leqslant n$ 且 $m_0\leqslant l\leqslant m$，我们有 $|\mathcal{A}_{(s,l)}|\leqslant\binom{n}{s}\binom{m}{l}$，这意味着

$$|\mathcal{A}_{(s,l)}|\leqslant\Big(\frac{en}{s}\Big)^s\Big(\frac{em}{l}\Big)^l.$$

因此，对两边取对数，可得

$$\frac{\log|\mathcal{A}_{(s,l)}|}{sl}\leqslant\frac{\log(en/s)}{l}+\frac{\log(em/l)}{s}\leqslant\frac{\log(en/n_0)}{m_0}+\frac{\log(em/m_0)}{n_0}.$$

这意味着[5] $\mathfrak{F}\leqslant(1/2)\psi_{n,\ m}(n_0,\ m_0)$。因此，

$$\mathbb{E}\big[\|\Pi_{\widehat{\mathcal{T}}}(\mathbf{E})\|_{\mathsf{F}}^2\big]\leqslant\big(1+\log(2M|\mathcal{G}|)\big)\big(9\sigma^2+8b^2\psi_{n,m}(n_0,m_0)\big)$$

其中 $|\mathcal{G}|=|\mathcal{F}|\leqslant K^nL^m$ 且 $M\leqslant 12^{KL}$。考虑到 $K\geqslant 2$ 且 $L\geqslant 2$这一事实，可得

$$\begin{aligned}1+\log(2M|\mathcal{G}|)&\leqslant 1+\log 2+KL\log 12+n\log K+m\log L\\ &\leqslant 3KL+n\log K+m\log L.\end{aligned}$$

项 $\mathbb{E}\big[\|\Pi_{\widehat{\mathcal{T}}}(\mathbf{E})\|_{\mathsf{F}}^2\big]$ 最终可如下界定:

$$\mathbb{E}\big[\|\Pi_{\widehat{\mathcal{T}}}(\mathbf{E})\|_{\mathsf{F}}^2\big]\leqslant\big(3KL+n\log K+m\log L\big)\big(9\sigma^2+8b^2\psi_{n,m}(n_0,m_0)\big).$$

至此，引理得证。 □

为使证明其余部分的记号简洁，令 $A=n\log K+m\log L$。最后，我们分别使用引理 1和引理 2中得到的关于 Ξ_1 和 Ξ_2 的界，以及分解式（18）和（19）。由于 $\mathbb{E}[\langle\boldsymbol{\Theta}^*-\Pi_{\mathcal{T}}(\boldsymbol{\Theta}^*),\mathbf{E}\rangle]=\langle\boldsymbol{\Theta}^*-\Pi_{\mathcal{T}}(\boldsymbol{\Theta}^*)\ \mathbb{E}[\mathbf{E}]\rangle=0$，我们得到

$$\begin{aligned}\mathbb{E}[\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}^*\|_{\mathsf{F}}^2] &\leqslant \|\boldsymbol{\Theta}^*-\Pi_{\mathcal{T}}\boldsymbol{\Theta}^*\|_{\mathsf{F}}^2+2\mathbb{E}[\Xi_1]+2\mathbb{E}[\Xi_2]\\ &\leqslant \|\boldsymbol{\Theta}^*-\Pi_{\mathcal{T}}\boldsymbol{\Theta}^*\|_{\mathsf{F}}^2+2\sigma\sqrt{2(A+1)}\,\mathbb{E}[\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}^*\|_{\mathsf{F}}^2]^{1/2}+2b\rho(A+1)+2\mathbb{E}[\Xi_2].\end{aligned}$$

容易验证，最后一个不等式推出

$$\big(\mathbb{E}[\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}^*\|_{\mathsf{F}}^2]^{1/2}-\sigma\sqrt{2(A+1)}\big)^2\leqslant\|\boldsymbol{\Theta}^*-\Pi_{\mathcal{T}}(\boldsymbol{\Theta}^*)\|_{\mathsf{F}}^2+2(A+1)(\sigma^2+b\rho)+2\mathbb{E}[\Xi_2].$$

由此立即得到

$$\begin{aligned}\mathbb{E}[\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}^*\|_{\mathsf{F}}^2]^{1/2} &\leqslant \|\boldsymbol{\Theta}^*-\Pi_{\mathcal{T}}(\boldsymbol{\Theta}^*)\|_{\mathsf{F}}+\sigma\sqrt{2(A+1)}+\big(2(A+1)(\sigma^2+b\rho)+2\mathbb{E}[\Xi_2]\big)^{1/2}\\ &\leqslant \|\boldsymbol{\Theta}^*-\Pi_{\mathcal{T}}(\boldsymbol{\Theta}^*)\|_{\mathsf{F}}+\big((A+1)(8\sigma^2+4b\rho)+4\mathbb{E}[\Xi_2]\big)^{1/2}\\ &\leqslant \|\boldsymbol{\Theta}^*-\Pi_{\mathcal{T}}(\boldsymbol{\Theta}^*)\|_{\mathsf{F}}+\big(17\sigma^2+4b\rho+8b^2\psi_{n,m}(n_0,m_0)\big)^{1/2}(3KL+A)^{1/2},\end{aligned}$$

其中，在第二行中我们使用了不等式 $\sqrt{x}+\sqrt{y}\leqslant\sqrt{2x+2y}$。最后，在条件 $\psi_{n,m}(n_0,m_0)\leqslant(\sigma/b)^2$, 下，我们得到了定理的结论。

[5] This is true since $\frac{\log(2n)}{n_0\log(en/n_0)}=\frac{1}{n_0}+\frac{\log(2n_0/e)}{n_0\log(en/n_0)}\leqslant\frac{\log(4n_0)}{n_0\log(2e)}\leqslant 0.5$ provided that $n_0\geqslant 3$.

7.2 定理 *3* （的风险界）

第一项结论：分段常数图子。鉴于（7），事实是 $W = W_{\widehat{\mathbf{\Theta}}}$ 和命题1，我们有

$$\begin{aligned}\mathbb{E}[\delta(W_{\widehat{\mathbf{\Theta}}}, W^*)^2]^{1/2} &\leqslant \frac{\mathbb{E}[\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_{\mathsf{F}}^2]^{1/2}}{\sqrt{nm}} + \mathbb{E}[\delta(W_{\mathbf{\Theta}^*}, W^*)^2]^{1/2} \\ &\leqslant \frac{\mathbb{E}[\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_{\mathsf{F}}^2]^{1/2}}{\sqrt{nm}} + \frac{\rho}{\sqrt{2}}\left(\sqrt{\frac{K}{n}} + \sqrt{\frac{L}{m}}\right)^{1/2}. \qquad (22)\end{aligned}$$

令 $\mathcal{T}$ 和 $\mathcal{T}^*$ 分别表示所有具有实数元素、且在与 $\mathbf{\Theta}$ 和 $\mathbf{\Theta}^*$相同的块上按块取常数的 $n \times m$ 矩阵的集合。显然，$\mathcal{T}$ 和 $\mathcal{T}^*$ 是赋予标量积 $\langle \mathbf{M}_1, \mathbf{M}_2 \rangle = \mathrm{tr}(\mathbf{M}_1^\top \mathbf{M}_2)$的 $n \times m$ 实矩阵空间的线性子空间。令 $\Pi_{\widehat{\mathcal{T}}}$为到 $\mathcal{T}$上的正交投影。我们有 $\Pi_{\widehat{\mathcal{T}}}\mathbf{H} = \mathbf{\Theta}$。因此，

$$\begin{aligned}\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_{\mathsf{F}} = \|\Pi_{\widehat{\mathcal{T}}}\mathbf{H} - \mathbf{\Theta}^*\|_{\mathsf{F}} &\overset{①}{\leqslant} \|\Pi_{\widehat{\mathcal{T}}}(\mathbf{H} - \mathbf{\Theta}^*)\|_{\mathsf{F}} + \|\Pi_{\widehat{\mathcal{T}}}\mathbf{\Theta}^* - \mathbf{\Theta}^*\|_{\mathsf{F}} \\ &\overset{②}{\leqslant} \|\mathbf{H} - \mathbf{\Theta}^*\|_{\mathsf{F}} + \|(\rho/2)\mathbf{1}_n\mathbf{1}_m^\top - \mathbf{\Theta}^*\|_{\mathsf{F}}.\end{aligned}$$

上面，①是三角不等式的结果，而②则源于$\Pi_{\widehat{\mathcal{T}}}$是一个正交投影（因此是一个压缩映射），且矩阵$(\rho/2)\mathbf{1}_n\mathbf{1}_m^\top$ 属于 $\Pi_{\widehat{\mathcal{T}}}$的像。因此，

$$\frac{1}{nm}\mathbb{E}\big[\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_{\mathsf{F}}^2 \,\big|\, \boldsymbol{U}, \boldsymbol{V}\big] \leqslant (\sigma + 0.5\rho)^2.$$

对于每个 $k \in [K]$ 和 $\ell \in [L]$，我们定义 $n_k = n|a_k - a_{k-1}|$, $N_k = \#\{i : \ U_i \in [a_{k-1}, a_k[\}$, $m_\ell = m|b_\ell - b_{\ell-1}|$ 和 $M_\ell = \#\{j : \ V_j \in [b_{\ell-1}, b_\ell[\}$。我们还定义事件 $\Omega_0 = \{N_k \geqslant n_k/2;$ $M_\ell \geqslant m_\ell/2$对于所有 $k \in [K]$和 $\ell \in [L]\}$。由于事件 Ω_0^c是 （$\boldsymbol{U}$, $\boldsymbol{V}$）-可测的，我们得到

$$\frac{1}{nm}\mathbb{E}\big[\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_{\mathsf{F}}^2 \mathbb{1}_{\Omega_0^c}\big] = \frac{1}{nm}\mathbb{E}\Big(\mathbb{E}\big[\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_{\mathsf{F}}^2 \,\big|\, \boldsymbol{U}, \boldsymbol{V}\big]\mathbb{1}_{\Omega_0^c}\Big) \leqslant (\sigma + 0.5\rho)^2\mathbb{P}(\Omega_0^c).$$

利用并集界和切尔诺夫不等式，可以验证

$$\mathbb{P}(\Omega_0^c) \leqslant \sum_{k=1}^K \mathbb{P}(N_k \leqslant n_k/2) + \sum_{\ell=1}^L \mathbb{P}(M_\ell \leqslant m_\ell/2) \leqslant \sum_{k=1}^K e^{-n_k/8} + \sum_{\ell=1}^L e^{-m_\ell/8}.$$

由于我们假设 $n_k \geqslant 8\log(nK)$ 和 $m_\ell \geqslant 8\log(mL)$，因此得到 $\mathbb{P}(\Omega_0^c) \leqslant n^{-1} + m^{-1}$。如果定义最小二乘估计量 $\widehat{\mathbf{\Theta}}$时所使用的参数 n_0 和 m_0 满足 $n_0 = \min_k n_k/2 = n\Delta^{(K)}/2$和 $m_0 = \min_\ell m_\ell/2 = m\Delta^{(L)}/2$，那么在事件 Ω_0上我们可以应用定理1。可以验证 $\psi_{n,m}(n_0, m_0) = \psi_{n,m}(\Delta^{(K,L)})$。结合前面的不等式，这意味着

$$\begin{aligned}\frac{\mathbb{E}[\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_{\mathsf{F}}^2]}{nm} &= \frac{\mathbb{E}[\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_{\mathsf{F}}^2 \mathbb{1}_{\Omega_0}] + \mathbb{E}[\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_{\mathsf{F}}^2 \mathbb{1}_{\Omega_0^c}]}{nm} \\ &\leqslant \big(25\sigma^2 + 4b\rho\big)\Big(\frac{3KL}{nm} + \frac{\log K}{m} + \frac{\log L}{n}\Big) + \frac{(\sigma + 0.5\rho)^2}{n} + \frac{(\sigma + 0.5\rho)^2}{m} \\ &\leqslant \Big\{\big(27\sigma^2 + 4b\rho\big)^{1/2}\Big(\frac{3KL}{nm} + \frac{\log K}{m} + \frac{\log L}{n}\Big)^{1/2} + \frac{\rho}{2}\sqrt{\frac{1}{n} + \frac{1}{m}}\Big\}^2,\end{aligned}$$

在 $\psi_{n,\,m}(\Delta^{(K,L)}) \leqslant (\sigma/b)^2$ 的条件下。还可以验证，如果 $K,\ L \geqslant 2$ 且 $n,\ m \geqslant$ 5</style}，则成立

$$\frac{1}{n}+\frac{1}{m} \leq \frac{1}{3}\left(\sqrt{\frac{K}{n}}+\sqrt{\frac{L}{m}}\right).$$

这个不等式结合（22），完成了该定理的证明。

第二个结论：赫尔德连续图函数。利用（7）和闵可夫斯基不等式，我们得到

$$\mathbb{E}[\delta(\widehat{W}^{\mathsf{LS}}, W^*)^2]^{1/2} = \mathbb{E}[\delta(W_{\widehat{\mathbf{\Theta}}^{\mathsf{LS}}}, W^*)^2]^{1/2} \leqslant \frac{\mathbb{E}[\|\widehat{\mathbf{\Theta}}^{\mathsf{LS}} - \mathbf{\Theta}^*\|_{\mathsf{F}}^2]^{1/2}}{\sqrt{nm}} + \mathbb{E}[\delta(W_{\mathbf{\Theta}^*}, W^*)^2]^{1/2}.$$

令我们设定

$$K = L = \left\lfloor \left(\frac{3nm\mathcal{L}^2}{25\sigma^2+4b\rho}\right)^{1/2(\alpha+1)} \right\rfloor.$$

鉴于（11），我们有

$$(K/m)^{2(\alpha+1)} \leqslant \frac{3n\mathcal{L}^2}{(25\sigma^2+4b\rho)m^{2\alpha+1}} \leqslant \frac{3(\sigma/2b)^{4(\alpha+1)} \wedge 1}{\log^4(2n)}. \tag{23}$$

我们选择 $n_0 = \lfloor n/K \rfloor$ 和 $m_0 = \lfloor m/K \rfloor$。根据（23），我们有

$$\frac{m}{K} \geqslant \left(\log^4(2n)\right)^{1/2(\alpha+1)} \geqslant \log 8 > 2.$$

This 这意味着 $m_0 \geqslant 2$ 因此，$n_0 \geqslant 2$。再次使用（23），可以验证

$$\frac{6\log(em/2)}{n_0} \leqslant \frac{6\log(en/2)}{m_0} \leqslant (\sigma/b)^2.$$

这意味着

$$\psi_{n,m}(n_0, m_0) = \frac{3\log(en/n_0)}{m_0} + \frac{3\log(em/m_0)}{n_0} \leqslant \frac{3\log(en/2)}{m_0} + \frac{3\log(em/2)}{n_0} \leqslant (\sigma/b)^2.$$

结合定理 1、命题2 以及命题 1的断言 2，我们得到

$$\begin{aligned}
\mathbb{E}\left[\delta(\widehat{W}^{\mathsf{LS}}, W^*)^2\right]^{1/2} &\leqslant \frac{3\mathcal{L}}{2K^\alpha} + \frac{3\mathcal{L}}{2L^\alpha} + \left(25\sigma^2+4b\rho\right)^{1/2}\left(\frac{3KL}{nm} + \frac{\log K}{m} + \frac{\log L}{n}\right)^{1/2} \\
&\quad + \frac{2\mathcal{L}}{n^{\alpha/2}} + \frac{2\mathcal{L}}{m^{\alpha/2}} \\
&\leqslant \frac{3\mathcal{L}}{K^\alpha} + \left(25\sigma^2+4b\rho\right)^{1/2}\left(\frac{3K^2}{nm} + \frac{2\log K}{m}\right)^{1/2} + \frac{4\mathcal{L}}{m^{\alpha/2}} \\
&\leqslant \frac{3\mathcal{L}}{K^\alpha} + 3K\left(\frac{25\sigma^2+4b\rho}{3nm}\right)^{1/2} + \left(\frac{(50\sigma^2+8b\rho)\log K}{m}\right)^{1/2} + \frac{4\mathcal{L}}{m^{\alpha/2}}.
\end{aligned}$$

在上一个公式中，将 K 替换为其表达式（23），即可得到该定理的结论。

资助。本研究部分受 Investissements d'Avenir 项目资助（ANR- 11-IDEX0003/Labex Ecodec/ANR-11-LABX-0047），并获得欧洲研究委员会（ERC）通过欧盟"地平线2020"研究与创新计划资助协议 741467-FIRMNET 提供的资金，以及 Hi! PARIS 中心的支持。

参考文献

Paul W. Holland、Kathryn Blackmond Laskey 和 Samuel Leinhardt。随机区块模型：初探。《Social Networks》，5(2):109–137，1983。Krzysztof Nowicki 和 Tom A. B. Snijders。随机区块结构的估计与预测。《Journal of the American Statistical Association》，96(455):1077–1087，2001。Gérard Govaert 和 Mohamed Nadif。采用块混合模型进行聚类。《Pattern Recognition》，36 473，2003 2 463( ): – 。《Biometrics》。Yunpeng Zhao、Elizaveta Levina 和 Ji Zhu。度修正随机区块模型下网络社区检测的一致性。《Ann. Statist.》，40(4):2266–2292，2012。Peter Chin、Anup Rao 和 Van Vu。随机区块模型与稀疏图中的社区检测：具有最优恢复速率的谱算法。载于《Conference on Learning Theory》，第391–423页。PMLR，2015。Jing Lei 和 Alessandro Rinaldo。随机区块模型中谱聚类的一致性。《Ann. Statist.》，43 (1):215–237，2015。Jing Lei。随机区块模型的拟合优度检验。《Ann. Statist.》，44(1):401–424，2016。Anderson Y. Zhang 和 Harrison H. Zhou。随机区块模型中社区检测的极小极大速率。《Ann. Statist.》，44(5):2252–2280，2016。Y. X. Rachel Wang 和 Peter J. Bickel。基于似然的随机区块模型选择。《Ann. Statist.》，45(2):500–528，2017。Yudong Chen、Xiaodong Li 和 Jiaming Xu。度修正随机区块模型的凸化模块度最大化。《Ann. Statist.》，46(4):1573–1602，2018。Min Xu、Varun Jog 和 Po-Ling Loh。加权随机区块模型中社区估计的最优速率。《Ann. Statist.》，48(1):183–204，2020。Chao Gao、Zongming Ma、Anderson Y. Zhang 和 Harrison H. Zhou。实现随机区块模型中的最优误分类比例。《Journal of Machine Learning Research》，18(60):1–45，2017。Emmanuel Abbe。社区检测与随机区块模型：近期进展。《Journal of Machine Learning Research》，18(177):1–86，2018。网址：http://jmlr.org/papers/v18/16-480.html。Vitaly Feldman、Will Perkins 和 Santosh S. Vempala。子采样幂迭代：一种适用于区块模型和植入式 CSP 的统一算法。载于《Advances in Neurips 2015》，第2836–2844页，2015。Laura Florescu 和 Will Perkins。二部随机区块模型中的谱阈值。载于《Proceedings of COLT 2016》，JMLR Workshop and Conference Proceedings 第49卷，第943–959页。JMLR.org，2016。Stefan Neumann。具有极小簇的二部随机区块模型。载于《Advances in NeurIPS 2018》，第3871–3881页，2018。Zhixin Zhou 和 Arash A. Amini。社区检测谱聚类算法分析：一般二部情形。《J. Mach. Learn. Res.》，20:47:1–47:47，2019a。Zhixin Zhou 和 Arash A. Amini。最优二部网络聚类。《J. Mach. Learn. Res.》，21 68，2020 40 1 40: : – : 。Changxiao Cai、Gen Li、Yuejie Chi、H. Vincent Poor 和 Yuxin Chen。从不平衡且不完整的数据矩阵中估计子空间：$\ell_{2,\infty}$ 统计保证。《The Annals of Statistics》，49(2):944 – 967，2021。Mohamed Ndaoud、Suzanne Sigalla 和 Alexandre B. Tsybakov。二部随机区块模型的改进聚类算法。《IEEE Trans. Inf. Theory》，68(3):1960–1975，2022。Edoardo M. Airoldi、Thiago B. Costa 和 Stanley H. Chan。图子随机区块模型近似：理论与一致估计。载于《Advances in Neurips 2013》，第692–700页，2013。Patrick J. Wolfe 和 Sofia C. Olhede。非参数图子估计，2013。Sofia C Olhede 和 Patrick J Wolfe。网络直方图与区块模型近似的普适性。《Proceedings of the National Academy of Sciences》，111(41):14722–14727，2014。Chao Gao、Yu Lu 和 Harrison H. Zhou。速率最优的图子估计。《The Annals of Statistics》，43(6): 2624–2652，2015。Chao Gao、Yu Lu、Zongming Ma 和 Harrison H. Zhou。具有双聚类结构的矩阵的最优估计与补全。《J. Mach. Learn. Res.》，17:161:1–161:29，2016。Olga Klopp、Alexandre B. Tsybakov 和 Nicolas Verzelen。网络模型与稀疏图子估计的预言机不等式。《The Annals of Statistics》，45(1):316 – 354，2017。Olga Klopp 和 Nicolas Verzelen。割距离下的最优图子估计。《Probability Theory and Related Fields》，174:1033–1090，2019。Chao Gao 和 Zongming Ma。网络分析中的极小极大速率：图子估计、社区检测与假设检验。《Statistical Science》，36(1):16 – 33，2021。David J. Aldous。随机变量部分可交换数组的表示。《J. Multivariate Anal.》，11 (4):581–598，1981。Bryan S. Graham。具有度异质性的网络形成计量经济学模型。《Econometrica》，85(4): 1033–1063，2017。ISSN 0012-9682。

Bryan S. Graham。《稀疏网络的逻辑回归渐近理论》。Journal of Multivariate Analysis，2020年10月。网址：https://arxiv.org/pdf/2010.04703.pdf。Laurent Davezies、Xavier D'Haultfœuille 和 Yannick Guyonvarch。《可交换数组的经验过程结果》。《统计年鉴》，49(2):845–862，2021年。G. Leung 和 A.R. Barron。《信息论与混合最小二乘回归》。IEEE Transactions on Information Theory，52(8):3396–3410，2006年。Arnak S. Dalalyan 和 Alexandre B. Tsybakov。《指数加权聚合与尖锐预言机不等式》。载于《学习理论》，计算机科学讲义第4539卷，第97–111页。柏林：Springer，2007年。Arnak S. Dalalyan 和 Alexandre B. Tsybakov。《指数加权聚合、尖锐 PAC-贝叶斯界与稀疏性》。Machine Learning，72(1-2):39–61，2008年。A. S. Dalalyan 和 A. B. Tsybakov。《通过聚合和 Langevin 蒙特卡洛进行稀疏回归学习》。J. Comput. System Sci.，78(5):1423–1443，2012年。Arnak S. Dalalyan。《多元回归中的指数权重及有利于低秩性的先验》。Annales de l'Institut Henri Poincaré, Probabilités et Statistiques，56(2):1465–1483，2020年。Arnak S. Dalalyan。《非对称噪声分布下指数权重去噪风险界的简单证明》。预印本，arXiv，2022年12月。S. Lloyd。《PCM 中的最小二乘量化》。IEEE Transactions on Information Theory，28(2):129–137，1982年。Amin Coja-Oghlan。《通过自适应谱技术进行图划分》。Combinatorics, Probability and Computing，19(2):227–284，2010年。Yu Lu 和 Harrison H Zhou。《Lloyd 算法及其变体的统计与计算保证》。arXiv 预印本 arXiv:1612.02099，2016年。Christophe Giraud 和 Nicolas Verzelen。《采用松弛 k-均值进行聚类的部分恢复界》。Mathematical Statistics and Learning，1(3):317–374，2019年。Dimitris Bertsimas 和 John N Tsitsiklis。《线性优化导论》。Athena Scientific，1997年。Zhixin Zhou 和 Arash A. Amini。《社区检测谱聚类算法分析：一般二部图情形》。Journal of Machine Learning Research，20:1–47，2019b年2月。网址：https: //jmlr.org/papers/volume20/18-170/18-170.pdf。Gabriel Peyré 和 Marco Cuturi。《计算最优传输：及其在数据科学中的应用》。Foundations and Trends® in Machine Learning，11(5-6):355–607，2019年。Alexandre B Tsybakov。《非参数估计导论》。Springer，2008年。Wassily Hoeffding。《有界随机变量和的概率不等式》。Journal of the American Statistical Association，58(301):13–30，1963年。Phillippe Rigollet 和 Jan-Christian Hütter。《高维统计》。课程 18S997 讲义，813 (814):46，2015年。Ramon van Handel。《高维概率》。APC 550 课程讲义，普林斯顿大学，2016年12月。网址：https://web.math.princeton.edu/~rvan/APC550.pdf。

8 命题的证明

8.1 命题 *1* （图子的近似误差）

第一条结论（分段常数图子）。在下文中， λ 表示 $\mathbb{R}$ 上的勒贝格测度， λ_2 是 $\mathbb{R}^2$上的勒贝格测度。设 W^* 为一个图子，且存在某个$K \times L$矩阵 $\mathbf{Q}^*$以及满足 $a_0 = b_0 = 0$ 和 $a_K = b_L = 1$的某些序列 $a_0 < \ldots < a_K$、 $b_0 < \ldots < b_L$，使得对于每个 $u \in [a_{k-1}, a_k)$ ，都有 $W^*(u,v) = Q^*_{k,\ell}$，且 $v \in [b_{\ell-1}, b_\ell)$。等价地，

$$W^*(u,v) = \sum_{k=1}^{K}\sum_{\ell=1}^{L} \mathbf{Q}^*_{k,\ell} \mathbb{1}_{[a_{k-1},a_k)\times[b_{\ell-1},b_\ell)}(u,v).$$

我们还定义“权重”序列 $w^{(1)}_k = a_k - a_{k-1}$、 $w^{(2)}_\ell = b_\ell - b_{\ell-1}$ 和

$$\widehat{w}^{(1)}_k = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[a_{k-1},a_k[}(U_i) \qquad \text{and} \qquad \widehat{w}^{(2)}_\ell = \frac{1}{m}\sum_{j=1}^{m} \mathbb{1}_{[b_{\ell-1},b_\ell[}(V_j).$$

注意，四个权重序列 $w^{(1)}$、 $w^{(2)}$、 $\widehat{w}^{(1)}$ 和 $\widehat{w}^{(2)}$ 均为正，且总和为一。正如（Klopp 等人, 2017，第16页）中所证明的，存在两个函数 $\psi_1 \colon [0,1] \to [K]$ 和 $\psi_2 \colon [0,1] \to [L]$，使得

1. 对所有 $k \in [K]$ 和 $x \in \big[a_{k-1}, (a_{k-1} + \widehat{w}^{(1)}_k) \wedge a_k\big)$，有 $\psi_1(x) = k$2. 对所有 $\ell \in [L]$ 和 $x \in \big[b_{\ell-1}, (b_{\ell-1} + \widehat{w}^{(2)}_\ell) \wedge b_\ell\big)$，有 $\psi_2(x) = \ell$3. $\lambda(\psi_1^{-1}(k)) = \widehat{w}^{(1)}_k$ 对所有 $k \in [K]$4. $\lambda(\psi_2^{-1}(\ell)) = \widehat{w}^{(2)}_\ell$ 对所有 $\ell \in [L]$。

利用这些映射 ψ_1和 ψ_2，我们构造图子 $W^*_\psi(u,v) = \mathbf{Q}^*_{\psi_1(u)\,\psi_2(v)}$，其满足 $\delta(W_{\mathbf{\Theta}^*}, W^*_\psi) = 0$。这就得到

$$\delta^2(W_{\mathbf{\Theta}^*}, W^*) = \delta^2(W^*_\psi, W^*) \leqslant \|W^*_\psi - W^*\|^2_{\mathbb{L}_2} \leqslant (B-A)^2 \lambda_2\big((u,v) : W^*_\psi(u,v) \neq W^*(u,v)\big).$$

我们对 W^*_ψ的选择确保了 $W^*(u,v) = W^*_\psi(u,v)$，除非$u \in \big[(a_{k-1} + \widehat{w}^{(1)}_k) \wedge a_k, a_k\big)$或 $v \in \big[(b_{\ell-1} + \widehat{w}^{(2)}_\ell) \wedge b_\ell, b_\ell\big)$。这意味着

$$\delta^2(W_{\mathbf{\Theta}^*}, W^*) \leqslant (B-A)^2 \left(\sum_{k=1}^{K} (w^{(1)}_k - \widehat{w}^{(1)}_k)_+ + \sum_{\ell=1}^{L} (w^{(2)}_\ell - \widehat{w}^{(2)}_\ell)_+ \right). \tag{24}$$

由于 U_i 是独立同分布的随机变量，并且在 $[0,1]$ 上均匀分布， $n\widehat{w}^{(1)}_k$ 服从二项分布 $\mathcal{B}(n, w^{(1)}_k)$。这意味着对于每个 k， $\mathbb{E}[\widehat{w}^{(1)}_k] = w^{(1)}_k$。因此，

$$\mathbb{E}[(w^{(1)}_k - \widehat{w}^{(1)}_k)_\pm] = \frac{1}{2}\mathbb{E}[|w^{(1)}_k - \widehat{w}^{(1)}_k|] \leqslant \left(\frac{w^{(1)}_k (1 - w^{(1)}_k)}{4n} \right)^{1/2} \leqslant \frac{(w^{(1)}_k)^{1/2}}{2\sqrt{n}}. \tag{25}$$

对于 $\mathbb{E}[(w^{(2)}_\ell - \widehat{w}^{(2)}_\ell)_+]$也可以得到类似的上界。因此，结合24）、25）、柯西-施瓦茨不等式以及权重 $w^{(1)}$ 之和为 1，可得

$$\begin{aligned}\mathbb{E}[\delta^2(W_{\mathbf{\Theta}^*}, W^*)] &\leqslant (B-A)^2 \left(\sum_{k=1}^{K} \mathbb{E}[(w^{(1)}_k - \widehat{w}^{(1)}_k)_+] + \sum_{\ell=1}^{L} \mathbb{E}[(w^{(2)}_\ell - \widehat{w}^{(2)}_\ell)_+] \right) \\ &\leqslant \frac{(B-A)^2}{2} \left(\sqrt{\frac{K}{n}} + \sqrt{\frac{L}{m}} \right).\end{aligned}$$

因此，对所得不等式两边开平方，即可完成证明。

第二个结论（*Hölder* 连续图子）。为使后续公式更简洁，我们令 $a_n(i)=i/n$，并假设 $\mathcal{L}=1$。我们引入 $I_i=\big[a_n(i-1)\ a_n(i)\big[$ 和 $J_j=\big[a_m(j-1)\ a_m(j)\big[$。对于任意正的 $k\in\mathbb{N}$，令 $\mathfrak{S}_n$ 为 $[k]$ 的置换集合。对于 $\sigma\in\mathfrak{S}_k$，令 τ_σ 为特定的保测映射

$$\tau_\sigma(x)=\sum_{i=1}^{k}\Big(\frac{\sigma(i)-1}{k}+x-\frac{i-1}{k}\Big)\mathbb{1}_{I_i}(x)\qquad\forall x\in[0,1[,\qquad\tau_\sigma(1)=\frac{\sigma(k)}{k}.$$

注意，τ_σ 对应于区间 $\{I_i:\ i\in[k]\}$ 的置换，符合 σ。根据 δ的定义，我们有

$$\begin{aligned}\delta^2(W_{\mathbf{\Theta}^*},W^*)&\leqslant\inf_{\substack{\sigma_1\in\mathfrak{S}_n\\ \sigma_2\in\mathfrak{S}_m}}\sum_{i\in[n],j\in[m]}\iint_{I_i\times J_j}\big(\Theta^*_{ij}-W^*(\tau_{\sigma_1}(x),\tau_{\sigma_2}(y))\big)^2dxdy\\&=\inf_{\substack{\sigma_1\in\mathfrak{S}_n\\ \sigma_2\in\mathfrak{S}_m}}\sum_{i\in[n],j\in[m]}\iint_{I_{\sigma_1(i)}\times J_{\sigma_2(j)}}\big(\Theta^*_{ij}-W^*(x,y)\big)^2dxdy.\end{aligned}\tag{26}$$

令 σ_1 为满足 $U_{(\sigma_1(i))}=U_i$ 的一个随机置换；例如，令 σ_1 为满足 $U_{\sigma_1^{-1}(1)}\leqslant U_{\sigma_1^{-1}(2)}\leqslant\cdots\leqslant U_{\sigma_1^{-1}(n)}$ 的对象。我们以类似的方式选择 σ_2，使得 $V_{(\sigma_2(j))}=V_j$。回顾 $\Theta^*_{ij}=W^*(U_i,V_j)$。令 $i'=\sigma_1(i)$，$j'=\sigma_2(j)$，并应用三角不等式，我们得到

$$\begin{aligned}|\Theta^*_{i,j}-W^*(x,y)|&=|W^*(x,y)-W^*(U_{(i')},V_{(j')})|\\&\leqslant\big|W^*(x,y)-W^*\big(a_{n+1}(i'),a_{m+1}(j')\big)\big|\\&\quad+\big|W^*\big(a_{n+1}(i'),a_{m+1}(j')\big)-W^*(U_{(i')},V_{(j')})\big|.\end{aligned}\tag{27}$$

若 $(x,y)\in I_{\sigma_1(i)}\times J_{\sigma_2(j)}=I_{i'}\times J_{j'}$，由于 $\big(a_{n+1}(i')\ a_{m+1}(j')\big)$ 属于同一集合 $I_{i'}\times J_{j'}$，赫尔德性质给出

$$\Big(W^*(x,y)-W^*\big(a_{n+1}(i'),a_{m+1}(j')\big)\Big)^2\leqslant\Big(\frac{1}{n^2}+\frac{1}{m^2}\Big)^{\alpha}.\tag{28}$$

对于第二项（27），我们再次使用赫尔德性质，从而得到

$$\big\{W^*\big(a_{n+1}(i'),a_{m+1}(j')\big)-W^*(U_{(i')},V_{(j')})\big\}^2\leqslant\big\{\big|a_{n+1}(i')-U_{(i')}\big|^2+\big|a_{m+1}(j')-V_{(j')}\big|^2\big\}^{\alpha}.$$

于是，记 $\sum_{i,j}$ 双重求和为 $\sum_{i\in[n]}\sum_{j\in[m]}$，我们有

$$\begin{aligned}\mathbb{E}\bigg[\sum_{i,j}\iint_{I_{i'}\times J_{j'}}\Big(\frac{i'}{n+1}-U_{(i')}\Big)^2\bigg]&=\frac{1}{n}\mathbb{E}\bigg[\sum_{i=1}^{n}\Big(\frac{i'}{n+1}-U_{(i')}\Big)^2\bigg]=\frac{1}{n}\mathbb{E}\bigg[\sum_{i=1}^{n}\Big(\frac{i}{n+1}-U_{(i)}\Big)^2\bigg]\\&\leqslant\max_{i=1,\dots,n}[\mathbf{Var}(U_{(i)})]\leqslant\frac{1}{4n}\end{aligned}\tag{29}$$

其中我们利用了以下事实：$U_{(i)}$ 服从贝塔分布 $\beta(i,n+1-i)$。含有$V_{(j')}$ 的项也以类似方式处理。结合式（26）、闵可夫斯基不等式、式（28）和式（29），可得

$$\begin{aligned}\mathbb{E}[\delta^2(W_{\mathbf{\Theta}^*},W^*)]^{1/2}&\leqslant\Big(\frac{1}{n^2}+\frac{1}{m^2}\Big)^{\alpha/2}+\mathbb{E}\bigg[\sum_{i,j}\iint_{I_{i'}\times J_{j'}}\Big(\frac{i'}{n+1}-U_{(i')}\Big)^2+\Big(\frac{j'}{m+1}-V_{(j')}\Big)^2\bigg]^{\alpha/2}\\&\leqslant\Big(\frac{1}{n^2}+\frac{1}{m^2}\Big)^{\alpha/2}+\Big(\frac{1}{4n}+\frac{1}{4m}\Big)^{\alpha/2}\leqslant2\Big(\frac{1}{n}+\frac{1}{m}\Big)^{\alpha/2}.\end{aligned}$$

证明至此完成。

8.2 命题的证明 *2*（矩阵的近似误差 $\boldsymbol{\Theta}^*$）不失一般性，我们证明 $K=\lfloor n/n_0\rfloor$ 和 $L=\lfloor m/m_0\rfloor$ 时所需不等式成立的情形。事实上，如果不等式对某个 K 成立，则对任何更小的值也必然成立。对 L 也是如此。此外，我们假设 $\mathcal{L}=1$。

我们构造分块常数矩阵（3）定义如下。设 $n=n_0K+r$ 其中 $0\leqslant r<n_0$ 且 $m=m_0L+s$ 其中 $0\leqslant s<m_0$。对所有 $k\in[K]$ 和 $\ell\in[L]$，令

$$I_k=[(k-1)n_0+1,(kn_0)\wedge n]\qquad\text{and}\qquad J_\ell=[(\ell-1)m_0+1,(\ell m_0)\wedge m].$$

每个形如 $I_k\times J_\ell$ 的集合中所含整数的个数记为 $n_{k,\,\ell}$。令

$$\widetilde{Q}_{k,\ell}=\frac{1}{n_{k,\ell}}\sum_{i\in I_k,j\in J_\ell}W^*(U_{(i)},V_{(j)}).$$

最后，对每个 $(i,\ j)\in[n]\times[m]$，令

$$\widetilde{\Theta}_{i,j}=\widetilde{Q}_{\lceil\sigma_U(i)/n_0\rceil,\lceil\sigma_V(j)/m_0\rceil}=\sum_{k,\ell}\widetilde{Q}_{k,\ell}\mathbb{1}_{\{\sigma_U(i)\in I_k\}}\mathbb{1}_{\{\sigma_V(j)\in J_\ell\}},$$

其中 σ_U（分别是 σ_V）为 $[n]$（分别为 $[m]$）的一个置换，将 $(U_i:\ i\in[n])$（分别将 $(V_j:\ j\in[m])$）变换为非递减序列。换言之，对于所有 i 和 j，$U_{\sigma_U(i)}=U_{(i)}$ 且 $V_{\sigma_V(j)}=V_{(j)}$。为界定我们关注的近似误差，注意到（$\sum_{k,\,\ell}$ 表示 $\sum_{k\in[K]}\sum_{\ell\in[L]}$）

$$\begin{aligned}\mathbb{E}\Big[\frac{1}{nm}\|\widetilde{\boldsymbol{\Theta}}-\boldsymbol{\Theta}^*\|_{\mathsf{F}}^2\Big]&=\frac{1}{nm}\sum_{k,\ell}\sum_{i:\sigma_U(i)\in I_k}\sum_{j:\sigma_V(j)\in J_\ell}\mathbb{E}\big[(\widetilde{Q}_{k,\ell}-\Theta^*_{ij})^2\big]\\&=\frac{1}{nm}\sum_{k,\ell}\sum_{i\in I_k}\sum_{j\in J_\ell}\mathbb{E}\bigg[\bigg(\frac{1}{n_{k,\ell}}\sum_{i'\in I_k}\sum_{j'\in J_\ell}W^*(U_{(i')},V_{(j')})-W^*(U_{(i)},V_{(j)})\bigg)^2\bigg]\\&\overset{\text{(Jensen)}}{\leqslant}\frac{1}{nm}\sum_{k,\ell}\sum_{i\in I_k}\sum_{j\in J_\ell}\bigg(\frac{1}{n_{k,\ell}}\sum_{i'\in I_k}\sum_{j'\in J_\ell}\mathbb{E}\Big[\big(W^*(U_{(i')},V_{(j')})-W^*(U_{(i)},V_{(j)})\big)^2\Big]\bigg).\end{aligned}$$

利用赫尔德性质和詹森不等式，我们得到

$$\mathbb{E}\Big[\big(W^*(U_{(i')},V_{(j')})-W^*(U_{(i)},V_{(j)})\big)^2\Big]\leqslant\mathbb{E}\bigg[\bigg\|\begin{bmatrix}U_{(i')}-U_{(i)}\\V_{(j')}-V_{(j)}\end{bmatrix}\bigg\|^{2\alpha}\bigg]\leqslant\bigg\{\mathbb{E}\bigg[\bigg\|\begin{bmatrix}U_{(i')}-U_{(i)}\\V_{(j')}-V_{(j)}\end{bmatrix}\bigg\|^2\bigg]\bigg\}^{\alpha}.$$

由于 $|i-i'|<n_0+1\leqslant(3/2)\,n_0\leqslant(3n)/(2K)$，（Klopp 等人，2017，引理 4.10，第 27 页）可得

$$\mathbb{E}\big[|U_{(i')}-U_{(i)}|^2\big]\leqslant 9/(2K)^2.$$

同样地，$\mathbb{E}\big[|V_{(j')}-U_{(j)}|^2\big]\leqslant 9/(2L)^2$。因此，

$$\mathbb{E}\Big[\frac{1}{nm}\|\widetilde{\boldsymbol{\Theta}}-\boldsymbol{\Theta}^*\|_{\mathsf{F}}^2\Big]^{1/2}\leqslant\mathcal{L}\Big[\Big(\frac{3}{2K}\Big)^2+\Big(\frac{3}{2L}\Big)^2\Big]^{\alpha/2}\leqslant\frac{3\mathcal{L}}{2}\Big(\frac{1}{K^\alpha}+\frac{1}{L^\alpha}\Big),$$

其中，我们使用了不等式 $(a+b)^c\leqslant a^c+b^c$ 对于 $a,\ b\geqslant 0$ 以及对于 $c\in(0,1]$。

8.3 命题的证明 *3*（放宽为线性规划）为便于后续参考，我们回顾一下，我们所关注的是解决该问题

$$\min_{\mathbf{Z}\in\mathcal{Z}(n,K,n_0)}\mathcal{L}(\mathbf{Z},\mathbf{Q},\mathbf{Z}^{\text{prod}}).\tag{OPT 1}$$

第一项主张（代价函数的线性化）。我们有

$$\mathcal{L}(\mathbf{Z},\mathbf{Q},\mathbf{Z}^{\text{prod}}) - \|\mathbf{H}\|_{\mathsf{F}}^2 = -2\operatorname{tr}(\mathbf{Z}\mathbf{Q}(\mathbf{Z}^{\text{prod}})^\top\mathbf{H}^\top) + \operatorname{tr}(\mathbf{Z}\mathbf{Q}(\mathbf{Z}^{\text{prod}})^\top\mathbf{Z}^{\text{prod}}\mathbf{Q}^\top\mathbf{Z}^\top)$$
$$= -2\operatorname{tr}(\mathbf{Z}\mathbf{Q}(\mathbf{Z}^{\text{prod}})^\top\mathbf{H}^\top) + \operatorname{tr}(\mathbf{Q}\mathbf{D}\mathbf{Q}^\top\mathbf{Z}^\top\mathbf{Z}).$$

我们将证明 $\phi(\mathbf{Z}) = -2\operatorname{tr}(\mathbf{Z}\mathbf{Q}$（$\mathbf{Z}^{\text{prod}})^\top\mathbf{H}^\top) + \operatorname{tr}(\mathbf{Q}\mathbf{D}\mathbf{Q}^\top\mathbf{Z}^\top\mathbf{Z})$。我们注意到，由于 $\mathbf{Z}$ 的行受到约束，其列彼此正交，并且 $\mathbf{Z}^\top\mathbf{Z} = \mathsf{diag}(n_k;\, k\in[K])$，其中 $n_k = \boldsymbol{Z}_{\bullet k}^\top\mathbf{1}_n$ 是 $\mathbf{Z}$ 的第 k 列中的非零项数量。因此我们得到

$$\operatorname{tr}(\mathbf{Q}\mathbf{D}\mathbf{Q}^\top\mathbf{Z}^\top\mathbf{Z}) = \sum_{k=1}^{K} n_k \boldsymbol{Q}_{k,\bullet}\mathbf{D}\boldsymbol{Q}_{k,\bullet}^\top = \sum_{k=1}^{K} \boldsymbol{Z}_{\bullet k}^\top\mathbf{1}_n\, \boldsymbol{Q}_{k,\bullet}\mathbf{D}\boldsymbol{Q}_{k,\bullet}^\top.$$

这完成了第一个论断的证明，并意味着（OPT 1）等价于

$$\arg\min_{\mathbf{Z}\in\mathcal{Z}(n,K,n_0)} \phi(\mathbf{Z}). \tag{OPT 2}$$

第二个论断（极端点的刻画）。凸多面体$\mathcal{P}$ 中的极端点，定义为 $\mathcal{P}$ 中的一个点，且该点不能表示为 $\mathcal{P}$中两个元素的非平凡凸组合。首先，我们证明 $\mathcal{Z}(n,K,n_0)$ 中的任意点都是极端点。令$\widetilde{\mathbf{Z}} \in \mathcal{Z}(n,K,n_0)$、 $\lambda\in(0,1)$、$\mathbf{Z}_1$和$\mathbf{Z}_2 \in \mathcal{Z}(n,K,n_0)$满足如下条件。

$$\mathbf{Z} = \lambda\mathbf{Z}_1 + (1-\lambda)\mathbf{Z}_2\,.$$

Fix 某些 $i\in[n]$，由于对 $\mathbf{Z}$ 的行的约束，存在 $j\in[K]$ 使得 t

$$1 = Z_{ij} = \lambda(Z_1)_{ij} + (1-\lambda)(Z_2)_{ij}\,. \tag{30}$$

满足（30）的唯一方式是令 $(Z_1)_{ij} = (Z_2)_{ij} = 1$，因为 $\lambda\in(0,1)$。于是 $(Z_1)_{i,\bullet} = (Z_2)_{i,\bullet} = Z_{i,\bullet}$，这是由行约束得出的。最后，这对每个 $i\in[n]$ 都成立，因此 $\mathbf{Z} = \mathbf{Z}_1 = \mathbf{Z}_2$，，这确保了 $\mathbf{Z}$ 是一个极点。

现在只需证明，$\mathcal{Z}(n,K,n_0)$ 的所有极点都属于 $\{0,1\}$。设$\widetilde{\mathbf{Z}} \in \mathcal{Z}(n,\ K,\ n_0)$ 至少有一个元素属于$(0,\ 1)$。我们来证明$\mathbf{Z}$ 不可能是一个极点，也就是说，我们可以将$\mathbf{Z}$ 写成 $\mathcal{Z}(n,\ K,\ n_0)$中两个元素的凸组合。证明将用到下面两个引理。

引理 3. 设 $i_0\in[n]$，$k_0\in[K]$ 满足 $Z_{i_0k_0}\in(0,1)$。则1. 存在 $k_0'\in[K]$，使得 $Z_{i_0,k_0'}\in(0,1)$。2. 要么 $\sum_{i=1}^n Z_{i,k_0}\notin\mathbb{N}$，要么存在 $i_0'\in[n]$，使得 $Z_{i_0'k_0}\in(0,1)$。

引理3的证明。证明直接由行和列约束得出。事实上，由于每行元素之和是整数，如果某个系数不为 0 或 1，那么同一行中必有另一个系数属于$(0,1)$，这就证明了(i)。此外，如果 $\sum_{i=1}^n Z_{i,k_0}\in\mathbb{N}$，同样的论证表明，列 k_0 中还有另一个系数属于$(0,1)$。

□

如果我们将矩阵 $\mathbf{Z}$ 视为二部图的双邻接矩阵 $\mathcal{G}$，其加权边由 $\mathbf{Z}$ 的元素给出，那么下一个引理将形式化地说明：要么 $\mathcal{G}$ 包含一个环，要么它包含一条路径，其极端点对应于一列，该列的元素之和严格大于 n_0（见图 9）。

引理 4. 设 $\mathbf{Z}\in\widetilde{\mathcal{Z}}(n,K,n_0)$*{v21}*。存在 $T\geqslant 1$，以及两个（i_t）$_{t=1}^{T}$ 和（k_t）$_{t=0}^{T}$ 不同指标的序列（可能有 $k_0 = k_T$），使得1. $Z_{i_tk_t}\in(0,1)$ 对所有 $1\leqslant t\leqslant T$均成立。

2. $Z_{i_{t+1}k_t} \in (0,1)$ 对所有 $0 \leqslant t \leqslant T-1$。3. 要么 $\sum_{i=1}^{n} Z_{ik_t} > n_0$ 对 $t \in \{0, T\}$（我们称 **Z** 具有一条死胡同路径），要么 $k_T = k_0$（**Z**具有一个环）。

证明 引理 4. 首先假设**Z** 的所有列的和都是某些整数。我们将 $Z_{i_1 k_0}$ 记为 **Z** 中满足 $(0,1)$ 的一个元素。根据引理 (i) 的第 i 部分，存在 $k_1 \neq k_0$ 使得 $Z_{i_1 k_1} \in (0,1)$。按照同一个引理 3，我们可以找到 $i_2 \neq i_1$，以及 $k_2 \neq k_1$，使得 $Z_{i_2 k_1} \in (0,1)$ 和 $Z_{i_2 k_2} \in (0,1)$（因此 $T \geqslant 2$）。我们重复这一过程，直到第 T次迭代，其中将 T定义为首次出现以下情况的时刻：对于某个 $0 \leqslant t_0 \leqslant T-1$，我们有 $i_T = i_{t_0}$ 或 $k_T = k_{t_0}$，这意味着我们遇到了一个在先前迭代中已经出现过的行或列。注意到 $t_0 \leqslant T-2$，因为我们始终有 $i_t \neq i_{t+1}$ 和 $k_t \neq k_{t+1}$。然后我们考虑以下移位序列。

- $i\# = (i_{t+t_0-1})_{t=1}^{T-t_0}$以及 $k\#_t = k_{t+t_0}$对于 $t \in \{0, \ldots, T-t_0-1\}$和 $k\#_{T-t_0} = k_{t_0}$在 $i_T = i_{t_0}$的情况下，这意味着我们首先遇到了一个在前几次迭代中已经出现过的行。在这种情况下，$t_0 \geqslant 1$。

- $i\# = (i_{t+t_0})_{t=1}^{T-t_0}$并且 $k\# = (k_{t+t_0})_{t=0}^{T-t_0}$在 $k_T = k_{t_0}$ 的情况下，这意味着我们首先遇到了一个在之前迭代中已经拥有的列。

这些序列 $i^{\#}$ 和 $k^{\#}$ 根据构造满足该引理的环条件 .

现在我们假设存在一列 k_0，其和不属于 $\mathbb{N}$。那么它有一个坐标$Z_{i_1 k_0} \in (0,1)$。应用引理 3 的第(i) 部分，可得 $k_1 \neq k_0$，使得 $Z_{i_1 k_1} \in (0,1)$。

- 如果 $\sum_{i=1}^{n} Z_{ik_1} > n_0$，那么引理就针对 $T=1$ 和死胡同路径情形得证。 .

- 否则，$\sum_{i=1}^{n} Z_{ik_1} = n_0$ 然后我们可以类似地重复前面的过程，直到迭代 T，将其定义为首次满足以下条件的时刻：$i_T = i_{t_0}$ 或 $k_T = k_{t_0}$，其中某个 $0 \leqslant t_0 \leqslant T-2$ 或 $\sum_{i=1}^{n} Z_{ik_T} > n_0$成立。在前两种情况下，我们像之前一样考虑平移后的序列 $i\#$ 和 $k\#$，它们满足引理的条件。在最后一种情况下，序列 $(i_t)_{t=1}^{T}$ 和 $(k_t)_{t=0}^{T}$ 满足所需的死胡同路径条件。

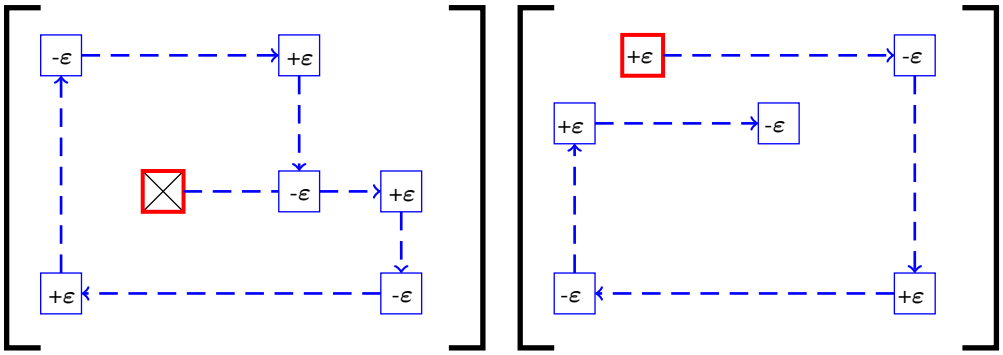
注意，$T < +\infty$，因为行数和列数有限。 □

命题3 的证明结尾与（Peyré和 Cuturi，2019，命题 3.4）中的证明类似，只需将**Z** 视为二部图的双邻接矩阵。下面我们针对本文的目的改写该证明。我们引入

$$\varepsilon_1 = \min_{k \in [K]} \left\{ \sum_{i=1}^{n} \frac{Z_{i,k} - n_0}{2} : \sum_{i=1}^{n} Z_{i,k} > n_0 \right\} \wedge 1, \quad \varepsilon_2 = \min_{i,k} \left\{ \frac{Z_{ik}}{2} : Z_{ik} > 0 \right\}$$

以及 $\varepsilon_3 = (1/2) \min_{i,k} \{1 - Z_{ik} : \ Z_{ik} < 1\}$。因此，$\varepsilon_1$、$\varepsilon_2$ 和 ε_3 都是正实数。按照约定，空集的最小值为 $+\infty$。最后，我们定义 $\varepsilon = \min(\varepsilon_1, \varepsilon_2, \varepsilon_3) \in (0,1)$。因此，当我们将 ε 加到 Z中一个属于 (0, 1)的元素上，或从该元素中减去 ε 时，该元素仍属于(0,1)。此外，如果该元素所在的列的和严格大于 n_0，那么从该元素中减去 ε 后，该列的和仍严格大于 n_0。我们应用引理 4，该引理适用于满足引理条件的序列 $(i_t)_{t=1}^{T}$ 和$(k_t)_{t=0}^{T}$。然后，我们定义 Z_ε，使得（Z_ε 的可视化构造见图 9）

$$(Z_\varepsilon)_{ij} = \begin{cases} 0 & \text{if} \quad i \neq i_t \text{ or } k \neq k_t \\ +\varepsilon & \text{if} \quad i = i_{t+1}, k = k_t \qquad \text{for some } 0 \leqslant t \leqslant T-1 \\ -\varepsilon & \text{if} \quad i = i_t, k = k_t \qquad \text{for some } 1 \leqslant t \leqslant T\,. \end{cases}$$



(a) 环 (b) 死胡同路径

图 9：起点用红色方块表示。在左侧图示中，我们遇到了一行与之前遇到的行相同，这类似于引理 4 中所述的过程。随后，我们忽略前面的步骤，以带叉的方块表示。此过程生成构造 $\mathbf{Z}_\varepsilon$ 所需的序列。在右侧图示中，我们没有遇到任何之前出现过的行或列。尽管如此，由于最后一列的和超过了 n_0，该过程仍然停止。这一情况为构造 $\mathbf{Z}_\varepsilon$ 提供了途径。

这些序列的性质意味着

• $\mathbf{Z}_\varepsilon \mathbf{1}_K = \mathbf{0}_n$。事实上，当 $i \notin \{i_1, \ldots, i_T\}$ 时，$\mathbf{Z}$ 的第 i 行之和为 0。但如果对于某个 $t \in [T]$ 有 $i = i_t$，那么第 k_t 列上有一个 $+\varepsilon$，第 k_{t-1} 列上有一个 $-\varepsilon$，其他位置均为 0。

• 此外，我们有 $\mathbf{Z}_\varepsilon^\top \mathbf{1}_n = \delta_\varepsilon \mathbf{1}_K$，其中 $\delta_\varepsilon \in \{0, \pm\varepsilon\}$。事实上，对于 $k \notin \{k_0, \ldots, k_T\}$，第 k 列的和为 0。对于 $k = k_t$，其中 $t \in [T-1]$，那么第 i_{t+1} 行上有一个 $+\varepsilon$，第 i_t 行上有一个 $-\varepsilon$，其余位置均为 0。对于 $t \in \{0, T\}$，分两种情况。如果 $k_T = k_0$，那么第 k_0 列的和为 0。如果 $k_T \neq k_0$，那么第 k_0 列上只有一个 $+\varepsilon$，且第 k_T 列上只有一个 $-\varepsilon$。

~选择 ε 可确保$\mathbf{Z}_+ = \mathbf{Z} + \mathbf{Z}_\varepsilon$和$\mathbf{Z}_- = \mathbf{Z} - \mathbf{Z}_\varepsilon$属于 $\mathcal{Z}(n, K, n_0)$。最后，$\mathbf{Z} = (\mathbf{Z}_+ + \mathbf{Z}_-)/2$和$\mathbf{Z}$不可能是极点。

9 定理的证明 4 （**下界**）尽管证明的总体结构及其中一些重要部分与Klopp 等人，2017，命题 3.4）的证明相似，但由于图子和观测矩阵并非对称，因此存在一些技术上的差异。此外，这些界还涉及若干依赖于 L 和 m的项，而文献中已证明的下界并不存在这种情况。

为得到所需的下界，我们将问题分解为以下三个极小极大下界

$$\inf_{\widehat{W}} \sup_{W^* \in \mathcal{W}_\rho[K,L]} \mathbb{E}_{W^*}\left[\delta^2(\widehat{W}, W^*)\right] \geqslant c\rho^2\left(\sqrt{\frac{K}{n}} + \sqrt{\frac{L}{m}}\right) \tag{31}$$

$$\inf_{\widehat{W}} \sup_{W^* \in \mathcal{W}_\rho[K,L]} \mathbb{E}_{W^*}\left[\delta^2(\widehat{W}, W^*)\right] \geqslant c\rho\left(\frac{KL}{Nnm} \wedge \rho\right) \tag{32}$$

$$\inf_{\widehat{W}} \sup_{W^* \in \mathcal{W}_\rho[2,2]} \mathbb{E}_{W^*}\left[\delta^2(\widehat{W}, W^*)\right] \geqslant c\rho\left(\frac{1}{N\sqrt{nm}} \wedge \rho\right) \tag{33}$$

若这三个不等式均成立，则所需结果将以常数 $c/3$ 成立。本节其余部分分为三个小节，分别给出其中一个不等式的证明，即（31）、（32）和（33）。

9.1 （(31) 的证明：由未知分割导致的误差 由于 $(K,\ n)$ 和 $(L,\ m)$ 的作用对称，我们只证明下界 $\rho^2\sqrt{K/n}$。同样的论证可得下界 $\rho^2\sqrt{L/m}$。因此，我们将得到下界 $\rho^2(\sqrt{K}/n \vee \sqrt{L/m}) \geqslant \rho^2(\sqrt{K}/n + \sqrt{L/m})/2$.

不失一般性，我们可以假设 K 是 16 的倍数。事实上，选取 $C > 17$，对于任意 $K > 17$，令 $K' = 16\lfloor K/16 \rfloor$，则不等式

$$\inf_{\widehat{W}} \sup_{W^* \in \mathcal{W}_\rho[K',L]} \mathbb{E}_{W^*}\left[\delta^2(\widehat{W}, W^*)\right] \geqslant c\rho^2\sqrt{\frac{K'}{n}}$$

将意味着

$$\inf_{\widehat{W}} \sup_{W^* \in \mathcal{W}_\rho[K,L]} \mathbb{E}_{W^*}\left[\delta^2(\widehat{W}, W^*)\right] \geqslant \inf_{\widehat{W}} \sup_{W^* \in \mathcal{W}_\rho[K',L]} \mathbb{E}_{W^*}\left[\delta^2(\widehat{W}, W^*)\right] \geqslant \frac{c}{4}\rho^2\sqrt{\frac{K}{n}}.$$

为了建立所需的下界，我们遵循标准方法（Tsybakov，2008，定理 2.7）：设计一个有限的图子集合，使其满足以下两个性质：根据度量 δ^2 衡量时，该集合中的图子彼此充分分离；与此同时，这些图子生成的分布彼此接近，因此难以根据数据矩阵 $\mathbf{H}$ 对它们加以区分。

为定义这个集合，我们选择一个 $K \times L$ 矩阵 $\mathbf{Q}$，其元素取自 $\{0, \rho\}$，以及两个正数 $\varepsilon, \varepsilon'$；关于 $\mathbf{Q}$、ε 和 ε' 的条件将在稍后说明。对于任意 $K \in \mathbb{N}$，定义

$$\mathcal{C}^K = \left\{\boldsymbol{s} \in \{-1,+1\}^K : \sum_{k=1}^K s_k = 0\right\}, \quad I^a_{\boldsymbol{s},\varepsilon} = \left[\sum_{k=1}^{a-1}\left(\frac{1}{K} + \varepsilon s_k\right), \sum_{k=1}^{a}\left(\frac{1}{K} + \varepsilon s_k\right)\right)$$

对于 $a \in [K]$ 和 $\boldsymbol{s} \in \mathcal{C}^K$，约定 $\sum_{k=1}^0 = 0$。类似地，对于每个 $\boldsymbol{t} \in \mathcal{C}^L$，我们设$J^b_{\boldsymbol{t},\varepsilon'} = \left[\sum_{\ell=1}^{b-1}\left(\frac{1}{L} + \varepsilon' t_\ell\right), \sum_{\ell=1}^{b}\left(\frac{1}{L} + \varepsilon' t_\ell\right)\right)$。区间 I的长度记为 $|I|$。现在我们可以定义图子类 $W_{\boldsymbol{s},\ \boldsymbol{t},\ \varepsilon}$，对于每个 $\boldsymbol{s} \in \mathcal{C}^K$ 和 $\boldsymbol{t} \in \mathcal{C}^L$，定义为

$$W_{\boldsymbol{s},\boldsymbol{t},\varepsilon}(u,v) = \sum_{k=1}^K \sum_{\ell=1}^L Q_{k,\ell} \mathbb{1}_{I^k_{\boldsymbol{s},\varepsilon}}(u) \mathbb{1}_{J^\ell_{\boldsymbol{t},\varepsilon'}}(v).$$

我们用 $\mathbf{P}^{W_{\boldsymbol{s},\boldsymbol{t},\varepsilon}}$ 表示 $\mathbf{H}=(H_{ij}, i\in[n], j\in[m])$的分布，其中 $\mathbf{H}$ 是根据图子 $W_{\boldsymbol{s},\boldsymbol{t},\varepsilon}$的二项模型抽样得到的。接下来的三个引理（其证明推迟到本小节末尾）提供了建立所需下界所必需的主要技术工具。

引理 5. 若 $4K\varepsilon\leqslant 1$，则对于所有 $\boldsymbol{s}$ 和 $\boldsymbol{s}'$ 来自 $\mathcal{C}^K$ 以及所有 $\boldsymbol{t}\in\mathcal{C}^L$，以下不等式 $D_{\mathsf{KL}}\left(\mathbf{P}^{W_{\boldsymbol{s},\boldsymbol{t},\varepsilon}}\|\mathbf{P}^{W_{\boldsymbol{s}',\boldsymbol{t},\varepsilon}}\right)\leqslant 6n(K\varepsilon)^2$ 成立。

引理 6. 假设 K，L 是足够大的、为 *16*的倍数的整数，并满足$KL\geqslant L$对数^2L+K对数^2K。存在 $\mathbf{B}\in\{-1,1\}^{K\times L}$满足以下两个性质。

i) 对于所有 $(k_1,k_2)\in[K]^2$, $k_1\neq k_2$，以及对于所有 $(\ell_1,\ell_2)\in[L]^2$, $\ell_1\neq\ell_2$，均有

$$|\langle\mathbf{B}_{k_1\bullet},\mathbf{B}_{k_2\bullet}\rangle|\leqslant L/4\qquad \textit{and}\qquad |\langle\mathbf{B}_{\bullet\ell_1},\mathbf{B}_{\bullet\ell_2}\rangle|\leqslant K/4$$

ii) 令 $\pi_i:\ [K/16]\to[K]$ 和 $\nu_i:\ [L/16]\to[L]$、$i=1,2$ 为任意双射，使得 *либо* $\mathrm{Im}(\pi_1)\cap\mathrm{Im}(\pi_2)=\varnothing$，或 $\mathrm{Im}(\nu_1)\cap\mathrm{Im}(\nu_2)=\varnothing$。则

$$\sum_{k=1}^{K/16}\sum_{\ell=1}^{L/16}\left(B_{\pi_1(k),\nu_1(\ell)}-B_{\pi_2(k),\nu_2(\ell)}\right)^2\geqslant\frac{KL}{512}.$$

对于每个 $\boldsymbol{s}\in\mathcal{C}^K$，定义 $\mathcal{A}_{\boldsymbol{s}}=\{k\in[K]:\ s_k=1\}$。注意到 $|\mathcal{A}_{\boldsymbol{s}}|=K/2$。

引理 7。 取 $\mathbf{Q}=\mathbf{B}+\mathbf{1}_K\mathbf{1}_L^\top$，其中 $\mathbf{B}$ 由引理 6给出。令 $\boldsymbol{s}$ 和 $\boldsymbol{s}'$ 为来自 $\mathcal{C}^K$ 的两个不同向量，使得 $|\mathcal{A}_{\boldsymbol{s}}\Delta\mathcal{A}_{\boldsymbol{s}'}|\geqslant K/4$。对于 $\varepsilon'=1/(4L)$，任意 $K\in\mathbb{N}$、$\varepsilon\in[0,1/(4K)]$ 和 $\boldsymbol{t}\in\mathcal{C}^L$，有

$$\delta^2(W_{\boldsymbol{s},\boldsymbol{t},\varepsilon},W_{\boldsymbol{s}',\boldsymbol{t},\varepsilon})\geqslant\frac{\varepsilon K\rho^2}{512}.$$

(Klopp 等人，2017) 中的引理 4.4 表明，存在一个子集 $\widetilde{\mathcal{C}}^K\subset\mathcal{C}^K$，使得 $\log|\widetilde{\mathcal{C}}^K|\geqslant K/16$，并且对于来自 $\widetilde{\mathcal{C}}^K$ 的任意 $\boldsymbol{s}\neq\boldsymbol{s}'$，都有 $|\mathcal{A}_{\boldsymbol{s}}\triangle\mathcal{A}_{\boldsymbol{s}'}|\geqslant K/4$。对于固定的 $\boldsymbol{t}\in\mathcal{C}^L$ 和 $\varepsilon^{-1}=2^4\sqrt{6nK}$，我们考虑集合 $\{W_{\boldsymbol{s},\boldsymbol{t},\varepsilon}:\ \boldsymbol{s}\in\widetilde{\mathcal{C}}^K\}$。根据引理 5 和引理 7，对于任意满足 $\boldsymbol{s}\neq\boldsymbol{s}'$ 的 $\boldsymbol{s}$，$\boldsymbol{s}'\in\widetilde{\mathcal{C}}^K$，有

$$\delta^2(W_{\boldsymbol{s},\varepsilon},W_{\boldsymbol{s}',\varepsilon})\geqslant\frac{\rho^2}{2^{13}}\sqrt{\frac{K}{6n}}\quad\text{and}\quad D_{\mathsf{KL}}(\mathbf{P}_{W_{\boldsymbol{s},\varepsilon}}\|\mathbf{P}_{W_{\boldsymbol{s}',\varepsilon}})\leqslant\frac{1}{16^2}K\leqslant\frac{1}{16}\log|\widetilde{\mathcal{C}}^K|.$$

因此，我们可以应用（Tsybakov，2008，定理 2.7）得到

$$\inf_{\widehat{W}}\sup_{W^*\in\mathcal{W}_\rho[K,L]}\mathbb{E}_{W^*}\left[\delta^2(\widehat{W},W^*)\right]\geqslant c\rho^2\sqrt{\frac{K}{n}}$$

对于某个普适常数 $c>0$。这就完成了（31）的证明，剩下的只是下面附录中三个技术引理的证明。

引理5的证明。可以验证，对于每个矩阵$\mathbf{A}\in[N]^{n\times m}$以及每个$\boldsymbol{s}\in\mathcal{C}^K$，

$$\mathbf{P}_{W_{\boldsymbol{s},\boldsymbol{t},\varepsilon}}(\mathbf{H}=\mathbf{A})=\prod_{i=1}^n\prod_{j=1}^m\int_{[0,1]^2}\binom{N}{A_{ij}}W_{\boldsymbol{s},\boldsymbol{t},\varepsilon}(u_i,v_j)^{A_{ij}}(1-W_{\boldsymbol{s},\boldsymbol{t},\varepsilon}(u_i,v_j))^{N-A_{ij}}\,du_i\,dv_j.$$

利用$W_{\boldsymbol{s},\boldsymbol{t},\varepsilon}$ 是分段常数这一事实，我们得到

$$\begin{aligned}\mathbf{P}_{W_{\boldsymbol{s},\boldsymbol{t},\varepsilon}}(\mathbf{H}=\mathbf{A}) &= \prod_{i,j}\sum_{k=1}^{K}\sum_{\ell=1}^{L}\int_{I^k_{\boldsymbol{s},\varepsilon}}\int_{J^\ell_{\boldsymbol{t},\varepsilon'}}\binom{N}{A_{ij}}Q_{k\ell}^{A_{i,j}}(1-Q_{k\ell})^{N-A_{i,j}}\,du_i\,dv_j \\ &= \prod_{i,j}\sum_{k,\ell}\binom{N}{A_{ij}}Q_{k\ell}^{A_{i,j}}(1-Q_{k\ell})^{N-A_{i,j}}\big|I^k_{\boldsymbol{s},\varepsilon}\big|\cdot\big|J^\ell_{\boldsymbol{t},\varepsilon'}\big| \\ &= \sum_{\mathbf{K},\mathbf{L}}\prod_{i,j}\binom{N}{A_{ij}}Q_{K_{ij}L_{ij}}^{A_{ij}}(1-Q_{K_{ij}L_{ij}})^{N-A_{ij}}\big|I^{K_{ij}}_{\boldsymbol{s},\varepsilon}\big|\cdot\big|J^{L_{ij}}_{\boldsymbol{t},\varepsilon'}\big|,\end{aligned}$$

其中，最后一行外层求和是对所有矩阵 $\mathbf{K}$ 和$\mathbf{L}$ 进行的，它们的元素分别属于$[K]$ 和 $[L]$。定义

$$\Psi(\mathbf{K},\mathbf{L},\mathbf{A})=\prod_{i,j}\binom{N}{A_{ij}}Q_{K_{ij}L_{ij}}^{A_{ij}}(1-Q_{K_{ij}L_{ij}})^{N-A_{ij}},\quad w_{\boldsymbol{s},\boldsymbol{t},\varepsilon}(\mathbf{K},\mathbf{L})=\prod_{i,j}\big|I^{K_{ij}}_{\boldsymbol{s},\varepsilon}\big|\cdot\big|J^{L_{ij}}_{\boldsymbol{t},\varepsilon'}\big|.$$

上述计算表明

$$\begin{aligned}D_{\mathsf{KL}}(\mathbf{P}_{W_{\boldsymbol{s},\boldsymbol{t},\varepsilon}}||\,\mathbf{P}_{W_{\boldsymbol{s}',\boldsymbol{t},\varepsilon}}) &= \sum_{\mathbf{A}\in\{0,1\}^{n\times m}}\mathbf{P}_{W_{\boldsymbol{s},\boldsymbol{t},\varepsilon}}(\mathbf{H}=\mathbf{A})\log\Big(\frac{\mathbf{P}_{W_{\boldsymbol{s},\boldsymbol{t},\varepsilon}}(\mathbf{H}=\mathbf{A})}{\mathbf{P}_{W_{\boldsymbol{s}',\boldsymbol{t},\varepsilon}}(\mathbf{H}=\mathbf{A})}\Big) \\ &= \sum_{\mathbf{A},\mathbf{K},\mathbf{L}}\Psi(\mathbf{K},\mathbf{L},\mathbf{A})w_{\boldsymbol{s},\boldsymbol{t},\varepsilon}(\mathbf{K},\mathbf{L})\log\Big(\frac{\sum_{\mathbf{K}',\mathbf{L}'}\Psi(\mathbf{K}',\mathbf{L}',\mathbf{A})w_{\boldsymbol{s},\boldsymbol{t},\varepsilon}(\mathbf{K}',\mathbf{L}')}{\sum_{\mathbf{K}',\mathbf{L}'}\Psi(\mathbf{K}',\mathbf{L}',\mathbf{A})w_{\boldsymbol{s}',\boldsymbol{t},\varepsilon}(\mathbf{K}',\mathbf{L}')}\Big).\end{aligned}$$

显然，$w_{\boldsymbol{s},\ \boldsymbol{t},\ \varepsilon}(\mathbf{K},\ \mathbf{L})\geqslant 0$以及$\sum_{\mathbf{K},\ \mathbf{L}}w_{\boldsymbol{s},\ \boldsymbol{t},\ \varepsilon}(\mathbf{K},\ \mathbf{L})=1$。由于函数$(x,\ y)\mapsto x\log(x/y)$ 是凸函数，我们应用詹森不等式，得到

$$\begin{aligned}D_{\mathsf{KL}}(\mathbf{P}_{W_{\boldsymbol{s},\boldsymbol{t}\varepsilon}}||\,\mathbf{P}_{W_{\boldsymbol{s}',\boldsymbol{t},\varepsilon}}) &\leqslant \sum_{\mathbf{A},\mathbf{K},\mathbf{L}}\Psi(\mathbf{K},\mathbf{L},\mathbf{A})w_{\boldsymbol{s},\boldsymbol{t},\varepsilon}(\mathbf{K},\mathbf{L})\log\Big(\frac{\Psi(\mathbf{K},\mathbf{L},\mathbf{A})w_{\boldsymbol{s},\boldsymbol{t},\varepsilon}(\mathbf{K},\mathbf{L})}{\Psi(\mathbf{K},\mathbf{L},\mathbf{A})w_{\boldsymbol{s}',\boldsymbol{t},\varepsilon}(\mathbf{K},\mathbf{L})}\Big) \\ &= \sum_{\boldsymbol{K},\mathbf{L}}w_{\boldsymbol{s},\boldsymbol{t},\varepsilon}(\mathbf{K},\mathbf{L})\log\Big(\frac{w_{\boldsymbol{s},\boldsymbol{t},\varepsilon}(\mathbf{K},\mathbf{L})}{w_{\boldsymbol{s}',\boldsymbol{t},\varepsilon}(\mathbf{K},\mathbf{L})}\Big).\end{aligned}$$

最后一个表达式可以看作是 $[K]^{n\times m}\times[L]^{n\times m}$上两个乘积分布之间的 Kullback-Leibler 散度。由于乘积分布之间的 Kullback-Leibler 散度等于各项 Kullback-Leibler 散度之和，我们得到

$$\sum_{\mathbf{K},\mathbf{L}}w_{\boldsymbol{s},\boldsymbol{t},\varepsilon}(\mathbf{K},\mathbf{L})\log\Big(\frac{w_{\boldsymbol{s},\boldsymbol{t},\varepsilon}(\mathbf{K},\mathbf{L})}{w_{\boldsymbol{s}',\boldsymbol{t},\varepsilon}(\mathbf{K},\mathbf{L})}\Big)=n\sum_{k=1}^{K}\big|I^k_{\boldsymbol{s},\varepsilon}\big|\log\Big(\frac{|I^k_{\boldsymbol{s},\varepsilon}|}{|I^k_{\boldsymbol{s}',\varepsilon}|}\Big)\leqslant n\sum_{k=1}^{K}\frac{\big(|I^k_{\boldsymbol{s},\varepsilon}|-|I^k_{\boldsymbol{s}',\varepsilon}|\big)^2}{|I^k_{\boldsymbol{s}',\varepsilon}|},$$

其中，最后一个不等式由 Kullback-Leibler 散度不超过卡方散度这一事实得出。由于$|I^k_{\boldsymbol{s}',\ \varepsilon}|=(1/K)+\varepsilon s'_k\geqslant(1/K)\ -\varepsilon\geqslant 3/(4K)$，我们得到

$$\sum_{\mathbf{K},\mathbf{L}}w_{\boldsymbol{s},\boldsymbol{t},\varepsilon}(\mathbf{K},\mathbf{L})\log\Big(\frac{w_{\boldsymbol{s},\boldsymbol{t},\varepsilon}(\mathbf{K},\mathbf{L})}{w_{\boldsymbol{s}',\boldsymbol{t},\varepsilon}(\mathbf{K},\mathbf{L})}\Big)\leqslant n\sum_{k=1}^{K}\frac{\big(\varepsilon s_k-\varepsilon s'_k\big)^2}{3/(4K)}\leqslant 6nK^2\varepsilon^2.$$

这就完成了该引理的证明。 □

证明: 引理6。令 $\boldsymbol{\Xi}$ 是一个 $K\times L$ 随机矩阵，其元素为独立同分布的拉德马赫变量 $\xi_{k,\ \ell}$即， $\mathbf{P}(\xi_{k,\ \ell}=\pm 1)$ {v29>。于是 $\langle\boldsymbol{\Xi}_{k_1\bullet},\ \boldsymbol{\Xi}_{k_2\bullet}\rangle=\sum_{\ell=1}^{L}\xi_{k_1,\ \ell}\xi_{k_2,\ \ell}$。根据霍夫丁不等式

$$\mathbf{P}(|\langle\boldsymbol{\Xi}_{k_1\bullet},\boldsymbol{\Xi}_{k_2\bullet}\rangle|\geqslant L/4)\leqslant 2e^{-L/32}.$$

由并集界可得，对于所有 $k_1 \neq k_2 \in [K], |\langle \boldsymbol{\Xi}_{k_1 \bullet}, \boldsymbol{\Xi}_{k_2 \bullet} \rangle| \leqslant L/4$，其概率大于 $1 - 2K^2 e^{-L/32}$，而对于 $L \geqslant 480$，该概率大于 $3/4$。类似地，可以验证

$$\mathbf{P}(\max_{\ell_1 \neq \ell_2} |\langle \boldsymbol{\Xi}_{\bullet,\ell_1}, \boldsymbol{\Xi}_{\bullet,\ell_2} \rangle| \leqslant K/4) > 3/4.$$

因此，我们得到
的是

$$\mathbf{P}\Big(\max_{\ell_1 \neq \ell_2} |\langle \boldsymbol{\Xi}_{\bullet,\ell_1}, \boldsymbol{\Xi}_{\bullet,\ell_2} \rangle| \leqslant K/4 \text{ and } \max_{k_1 \neq k_2} |\langle \boldsymbol{\Xi}_{k_1,\bullet}, \boldsymbol{\Xi}_{k_2,\bullet} \rangle| \leqslant L/4 \Big) > 1/2. \tag{34}$$

对于引理中陈述的第二个性质，我们按照陈述中的要求固定一些 $\mathcal{X}_i$，$\mathcal{Y}_i$ 和 π_i，ν_i，并定义

$$T(\pi_{1:2}, \nu_{1:2}, \boldsymbol{\Xi}) = \frac{1}{4} \sum_{k=1}^{K/16} \sum_{\ell=1}^{L/16} \big(\xi_{\pi_1(k),\nu_1(\ell)} - \xi_{\pi_2(k),\nu_2(\ell)} \big)^2.$$

显然，$T[\pi_{1:2}, \nu_{1:2}, \boldsymbol{\Xi}]$ 是 $KL/2^8$ 个参数为 1/2 的独立同分布伯努利随机变量之和。再次应用霍夫丁不等式，我们有

$$\mathbf{P}\Big(T[\pi_{1:2}, \nu_{1:2}, \boldsymbol{\Xi}] \leqslant \frac{KL}{2^{11}} \Big) = \mathbf{P}\Big(\frac{KL}{2^9} - T \geqslant \frac{3KL}{2^{11}} \Big) \leqslant e^{-9KL/2^{13}}.$$

满足 ii) 条件的函数不超过 $(K/16)!^2 (L/16)!^2$ $\pi_1, \pi_2, \nu_1, \nu_2$ 个。因此，并集界表明，以至少 $1 - (K/16)!^2 (L/16)!^2 e^{-9KL/2^{13}}$ 的概率，对所有 $\pi_1, \pi_2, \nu_1, \nu_2$，都有 $T[\pi_{1:2}, \nu_{1:2}, \boldsymbol{\Xi}] \geqslant KL/2^{11}$。取足够大的 K 和 L，并利用条件 $KL \geqslant K \log^2 K + L \log^2 L$，可得。

$$\mathbf{P}\Big(\min_{\pi_1, \pi_2, \nu_1, \nu_2} \sum_{k=1}^{K/16} \sum_{\ell=1}^{L/16} \big(\xi_{\pi_1(k),\nu_1(\ell)} - \xi_{\pi_2(k),c\nu_2(\ell)} \big)^2 \geqslant \frac{KL}{512} \Big) > 1/2. \tag{35}$$

结合（34）和（35），我们得到随机矩阵 $\boldsymbol{\Xi}$ 满足性质 i）和 ii）的概率严格大于零。这意味着此类矩阵的集合非空。

□

引理 7的证明。在不失一般性的情况下，在整个证明过程中，我们假设$\rho = 1$。此外，由于 $\boldsymbol{t}$ 是固定的，我们通常会在记号中省略它，并写作$W_{\boldsymbol{s},\,\varepsilon}$，而不是 $W_{\boldsymbol{s},\,\boldsymbol{t},\,\varepsilon}$。

只需证明，对于所有保测双射 $\tau_1 \colon [0,1] \to [0,1]$ 和 $\tau_2 \colon [0,1] \to [0,1]$,

$$\|W_{\boldsymbol{s},\varepsilon} - W_{\boldsymbol{s}',\varepsilon} \circ (\tau_1 \otimes \tau_2)\|_{\mathbb{L}_2}^2 \geqslant \frac{\varepsilon K}{512}.$$

如果对于某些 $k, k' \in [K]$，$u \in I^k_{\boldsymbol{s},\varepsilon}$ 且 $u' \in I^{k'}_{\boldsymbol{s}',\varepsilon}$，那么

$$\begin{aligned} \Big| \int_0^1 (W_{\boldsymbol{s},\varepsilon}(u,v) - 1/2)(W_{\boldsymbol{s},\varepsilon}(u',v) - 1/2) dv \Big| &= \Big| \sum_{\ell=1}^{L} (1/L + \varepsilon' t_\ell) \Big(Q_{k,\ell} - \frac{1}{2} \Big) \Big(Q_{k',\ell} - \frac{1}{2} \Big) \Big| \\ &\leqslant \frac{1}{4L} |\langle \mathbf{B}_{k,\bullet}, \mathbf{B}_{k',\bullet} \rangle| + \frac{L\varepsilon'}{4} \leqslant \frac{1}{8}. \end{aligned} \tag{36}$$

对于 k，$k' \in [K]$，令 $\omega_{kk'} = \lambda\big\{ I^k_{\boldsymbol{s},\varepsilon} \cap \tau_1^{-1}\big(I^{k'}_{\boldsymbol{s}',\varepsilon} \big) \big\}$，其中 λ 是 $\mathbb{R}$ 上的勒贝格测度。注意，$\sum_{k=1}^K \omega_{kk'} =$ （$1/K$） $+ \varepsilon s'_{k'}$，且 $\sum_{k'=1}^K \omega_{kk'} =$ （$1/K$） $+ \varepsilon s_k$。我们还引入了

$h_{\boldsymbol{s},k}(v)=W_{\boldsymbol{s},\varepsilon}(u^k_{\boldsymbol{s},\varepsilon},v)\ -1/2$，其中 $u^k_{\boldsymbol{s},\varepsilon}$ 是来自 $I^k_{\boldsymbol{s},\varepsilon}$的任意一点。我们有

$$\|W_{\boldsymbol{s},\varepsilon}-W_{\boldsymbol{s}',\varepsilon}(\tau_1\otimes\tau_2)\|^2_{\mathbb{L}_2}=\sum_{k=1}^K\sum_{k'=1}^K\omega_{k,k'}\|h_{\boldsymbol{s},k}-h_{\boldsymbol{s}',k'}\circ\tau_2\|^2_{\mathbb{L}_2}.$$

鉴于$|h_{\boldsymbol{s}',\ k'}(v)|=1/2$ 对所有 $v\in[0,1]$ 以及 (36)，对于任意 $k'\neq k$，我们有

$$\|h_{\boldsymbol{s}',\ k'}\circ\tau_2-h_{\boldsymbol{s}',\ k''}\circ\tau_2\|^2_{\mathbb{L}_2}=\|h_{\boldsymbol{s}',\ k'}\|^2_{\mathbb{L}_2}+\|h_{\boldsymbol{s}',\ k''}\|^2_{\mathbb{L}_2}-2\langle h_{\boldsymbol{s}',\ k'},\ h_{\boldsymbol{s}',\ k''}\rangle\geqslant 1/2-1/4=1/4.$$

根据三角不等式

$$\|h_{\boldsymbol{s},k}-h_{\boldsymbol{s}',k'}\circ\tau_2\|_{\mathbb{L}_2}+\|h_{\boldsymbol{s},k}-h_{\boldsymbol{s}',k''}\circ\tau_2\|_{\mathbb{L}_2}\geqslant\|h_{\boldsymbol{s}',k'}\circ\tau_2-h_{\boldsymbol{s}',k''}\circ\tau_2\|_{\mathbb{L}_2}\geqslant 1/2\,.$$

因此，对于任意 $k\in[K]$，至多存在一个 $k'\in[K]$ 使得$\|h_{\boldsymbol{s},\ k}-h_{\boldsymbol{s}',\ k'}\circ\tau_2\|_{\mathbb{L}_2}<1/4$。如果这样的 k' 存在，我们将其记为 $\pi(k)$。如果不存在，则令$\pi(k)=k$。用同样的论证，对于任意 $k'\in[K]$，至多存在一个 $k\in[K]$ 使得$\|h_{\boldsymbol{s},\ k}-h_{\boldsymbol{s}',\ k'}\circ\tau_2\|_{\mathbb{L}_2}<1/4$。这表明 π 是单射，因此是$[K]$的一个置换。此外，我们得到

$$\|W_{\boldsymbol{s},\varepsilon}-W_{\boldsymbol{s}',\varepsilon}\circ(\tau_1\otimes\tau_2)\|^2_{\mathbb{L}_2}\geqslant\frac{1}{16}\sum_{k=1}^K\sum_{k'\neq\pi(k)}\omega_{k,k'}=\frac{1}{16}\sum_{k=1}^K(1/K+\varepsilon s_k-\omega_{k,\pi(k)})\,. \tag{37}$$

If 若和 $\sum_{k=1}^K(1/K+\varepsilon s_k-\omega_{k,\pi(k)})$ 大于 $K\varepsilon/16$，则引理得证 .

接下来，我们检查如果 $\sum_{k=1}^K(1/K+\varepsilon s_k-\omega_{k,\pi(k)})<K\varepsilon/16$，结论是否也成立。注意，最后一个不等式可以重写为 $\sum_{k=1}^K\omega_{k,\pi(k)}>1-K\varepsilon/16$。我们来证明集合 $A=\{k\in\mathcal{A}_{\boldsymbol{s}}:$ $s'_{\pi(k)}>0$ 且 $\omega_{k,\pi(k)}\geqslant 1/K\}$ 的基数至少为 $7K/16$。事实上，注意到由于 $\omega_{k,}$ $_{k'}\leqslant\big(1/K+\varepsilon s_k\big)\wedge\big(1/K+\varepsilon s_{k'}\big)$, $\omega_{k,\pi(k)}\geqslant 1/K$ 意味着$s_k>0$ 且 $s'_{\pi(k)}>0$。因此，

$$\begin{aligned}1-\frac{K\varepsilon}{16}&\leqslant\sum_{k=1}^K\omega_{k,\pi(k)}=\sum_{k:s_k<0}\omega_{k,\pi(k)}+\sum_{k\in A}\omega_{k,\pi(k)}+\sum_{\substack{k\notin A\\ s_k>0}}\omega_{k,\pi(k)}\\&\leqslant\frac{K}{2}\Big(\frac{1}{K}-\varepsilon\Big)+|A|\Big(\frac{1}{K}+\varepsilon\Big)+\Big(\frac{K}{2}-|A|\Big)\frac{1}{K}=1+\varepsilon\Big(|A|-\frac{K}{2}\Big),\end{aligned}$$

这就得到 $|A|\geqslant 7K/16$。

由$\mathcal{A}^K$ 满足$s'\notin\widetilde{\mathcal{A}}_{\boldsymbol{s}}\triangle\mathcal{A}_{\boldsymbol{s}'}|\geqslant K/4$，我们有 $|\mathcal{A}_{\boldsymbol{s}}\cap\mathcal{A}_{\boldsymbol{s}'}|\leqslant 3K/8$。取基数为 $K/16$ 的 $B\subset A\cap\mathcal{A}^c_{\boldsymbol{s}'}$，并令 $C=\pi(B)$。这样的选择是可行的，因为

$$|A\cap\mathcal{A}^c_{\boldsymbol{s}'}|=|A|-|A\cap\mathcal{A}_{\boldsymbol{s}'}|\geqslant\frac{7K}{16}-|\mathcal{A}_{\boldsymbol{s}}\cap\mathcal{A}_{\boldsymbol{s}'}|\geqslant\frac{7K}{16}-\frac{3K}{8}=\frac{K}{16}.$$

还需注意，$B\cap C=\varnothing$。事实上，若 $k\in B$，则 $\pi(k)\in\mathcal{A}_{\boldsymbol{s}'}$。因此，$\pi(k)\notin B$，这意味着 $k\notin C$。

对于 ℓ，$\ell'\in[L]$，令 $\omega'_{\ell,\ \ell'}=\lambda\big\{J^\ell_{\boldsymbol{t},\varepsilon'}\cap\tau_2^{-1}(J^{\ell'}_{\boldsymbol{t}',\ \varepsilon'})\big\}$。采用与上文相同的论证，我们可以证明存在一个置换 $\nu\colon\ [L]\to[L]$ 使得，与（37）类似，

$$\|W_{\boldsymbol{s},\varepsilon}-W_{\boldsymbol{s}',\varepsilon}\circ(\tau_1\otimes\tau_2)\|^2_{\mathbb{L}_2}\geqslant\frac{1}{16}\sum_{b=1}^L(1/L+\varepsilon' t_\ell-\omega'_{\ell,\nu(\ell)}).$$

定义集合 $A'=\{\ell\in[L]:\ t_\ell>0,\ \omega'_{\ell,\,\nu(\ell)}\geqslant 1/L\}$。如果 $|A'|\leqslant L/16$，那么

$$\begin{aligned}\sum_{\ell=1}^{L}(1/L+\varepsilon' t_\ell-\omega'_{\ell,\nu(\ell)})&\geqslant 1-\sum_{t_\ell<0}\omega'_{\ell,\nu(\ell)}-\sum_{t_\ell>0}\omega'_{\ell,\nu(\ell)}(\mathbb{1}_{\ell\notin A'}+\mathbb{1}_{\ell\in A'})\\&\geqslant 1-\frac{L}{2}\Big(\frac{1}{L}-\varepsilon'\Big)-\Big(\frac{L}{2}-|A'|\Big)\frac{1}{L}-|A'|\Big(\frac{1}{L}+\varepsilon'\Big)=\Big(\frac{L}{2}-|A'|\Big)\varepsilon'\\&\geqslant\frac{7L\varepsilon'}{16}\geqslant 1/16\end{aligned}$$

因此 之前 $\|W_{\boldsymbol{s},\varepsilon}-W_{\boldsymbol{s}',\varepsilon}\circ(\tau_1\otimes\tau_2)\|^2_{\mathbb{L}_2}\geqslant 1/256\geqslant K\varepsilon/64$，其中我们用到了 $4K\varepsilon<1$。现在假设 $|A'|>L/16$。令 B' 为 A' 的任意一个子集，其基数为 $L/16$。我们有

$$\begin{aligned}\|W_{\boldsymbol{s},\varepsilon}-W_{\boldsymbol{s}',\varepsilon}\circ(\tau_1\otimes\tau_2)\|^2_{\mathbb{L}_2}&\geqslant\sum_{k\in B}\sum_{\ell\in B'}\int_{I^k_{\boldsymbol{s},\varepsilon}\times J^\ell_{\boldsymbol{s}',\varepsilon}}\big(W_{\boldsymbol{s},\varepsilon}-W_{\boldsymbol{s}',\varepsilon}(\tau_1\otimes\tau_2)\big)^2(u,v)\,du\,dv\\&\geqslant\sum_{k\in B}\sum_{\ell\in B'}\omega_{k,\pi(k)}\omega'_{\ell,\nu(\ell)}(Q_{k,\ell}-Q_{\pi(k),\nu(\ell)})^2\\&\geqslant\frac{1}{4KL}\sum_{k\in B}\sum_{\ell\in B'}(B_{k,\ell}-B_{\pi(k),\nu(\ell)})^2.\end{aligned}$$

根据引理 6，最后一项大于 $1/(4\times 512)\geqslant K\varepsilon/512$，因此引理的结论成立。

□

9.2 （(32)）的证明：图子未知取值导致的误差 与前面的证明类似，我们将使用（Tsybakov，2008，定理 2.7），这需要一类在距离 δ 上彼此充分分离且在 $n\times m$ 矩阵空间上生成相似分布的图子。在本证明中，集合中的所有图子都具有相同的划分，并且仅在该划分上函数的取值不同。

令 $\mathcal{Q}_0=\{(\rho/2)(1-\varepsilon)、(\rho/2)(1+\varepsilon)\}^{K\times L}$ 表示所有 $K\times L$ 矩阵构成的集合，其元素的取值要么为 $(\rho/2)(1-\varepsilon)$，要么为 $(\rho/2)(1+\varepsilon)$，其中 $\varepsilon\in(0,1/2)$ 将在稍后指定。对于任意$\mathbf{Q}\in\mathcal{Q}_0$，我们定义图子

$$W_{\mathbf{Q}}(u,v)=\sum_{k=1}^{K}\sum_{\ell=1}^{L}Q_{k,\ell}\mathbb{1}_{[(k-1)/K,k/K)}(u)\mathbb{1}_{[(\ell-1)/L,\ell/L)}(v)\,.$$

我们需要两个技术性引理来完成证明。下面陈述这两个引理，其证明则推迟到本小节末尾。对于任意一对置换π：$[K]\to[K]$ 和 ν：$[L]\to[L]$，以及任意矩阵 $\mathbf{Q}$，我们用 $\mathbf{Q}^{\pi,\,\nu}$ 表示将行和列置换后的矩阵 $Q^{\pi,\,k}{}^{\nu},\,\ell=Q_{\pi(k),\,\nu(\ell)}$。

引理 8. 对于 K和 L 足够大且满足 $KL\geqslant L$对数$^2\,L+K$ 对数$^2\,K$的情形，存在一个集合 $\mathcal{Q}\subset\mathcal{Q}_0$ 满足 对数$|\mathcal{Q}|\geqslant KL/32$ 且 最小值$_{\pi,\,\nu}\|\mathbf{Q}_1-\mathbf{Q}_2^{\pi\,\nu}\|^2_{\mathsf{F}}\geqslant\rho^2\varepsilon^2KL/8$，对于每个$\mathbf{Q}_1$,$\mathbf{Q}_2\in\mathcal{Q}$，都有 $\mathbf{Q}_1\neq\mathbf{Q}_2$。

引理 9. 以下断言成立1. 如果$\mathbf{Q}$

$$\begin{matrix}1&&2&&\pi\,\nu\|&1-&\pi\,\nu\\&&&&&2&\end{matrix}\|^2_{\mathsf{F}}\geqslant\rho^2\varepsilon^2KL/8\qquad\delta^2\ W_{\mathbf{Q}_1}\ W_{\mathbf{Q}_2}\ \geqslant\text{和}\mathbf{Q}\text{满足}\ \rho^2\varepsilon^2/8$$

$\min,\mathbf{Q}\mathbf{Q}_,$ 则$(,)$。2. 对于来自 $\mathcal{Q}$ 的任意一对矩阵$\mathbf{Q}_1$和$\mathbf{Q}_2$，有 $D_{\mathsf{KL}}(\mathbf{P}_{W_{\mathbf{Q}_1}}||\mathbf{P}^{W_{\mathbf{Q}_2}})\leqslant 6Nnm\rho\varepsilon^2$。

我们设定 $\varepsilon^2 = \frac{1}{54\times 32}\big(\frac{KL}{Nnm\rho}\wedge 1\big)$，这使我们能够应用（Tsybakov，2008，定理 2.7），因为根据引理 9 和引理 8，

$$D_{\mathsf{KL}}(\mathbf{P}_{W_{\mathbf{Q}_1}}\|\mathbf{P}_{W_{\mathbf{Q}_2}}) \leqslant \frac{1}{9}\log|\mathcal{Q}|, \qquad \delta^2(W_{\mathbf{Q}_1}, W_{\mathbf{Q}_2}) \geqslant \frac{1}{13824}\Big(\frac{\rho KL}{Nnm}\wedge\rho^2\Big),$$

对每个 $\mathbf{Q}_1$,$\mathbf{Q}_2\in\mathcal{Q}$。这就完成了（32）。

引理 8 的证明。在不失一般性的情况下，我们在本证明中假设 $\rho=2$。我们定义伪距离 $\delta(\mathbf{Q}_1,\mathbf{Q}_2)=$最小值$_{\pi,\,\nu}\|\mathbf{Q}_1-\mathbf{Q}_2^{\pi\,\nu}\|_{\mathsf{F}}$，其中最小值取遍 $[K]$ 和 $[L]$的所有置换。令 $\mathcal{Q}$ 为 $\mathcal{Q}$ 中矩阵的一个极大子集，使得这些矩阵关于 δ 是 $r:=\rho\varepsilon\sqrt{KL}/2$-分离的。由 $\mathcal{Q}$ 的极大性，我们有包含关系

$$\mathcal{Q}\subset\bigcup_{\mathbf{Q}\in\mathcal{Q}}\mathbb{B}_\delta(\mathbf{Q},r),$$

其中，$\mathbb{B}_\delta(\mathrm{Q}\ r)$ 是以 Q为中心、半径为 r 的球，其度量为 δ-距离。因此，$|\mathcal{Q}|\cdot|\mathbb{B}_\delta(\mathrm{Q}\ r)|\geqslant 2^{KL}$，因为所有球具有相同的基数。注意，$\mathbb{B}_\delta(\mathrm{Q}\ r)\subset\bigcup_{\pi,\,\nu}\mathbb{B}_{\mathsf{F}}(\mathrm{Q}_2^{\pi\,\nu},r)$ 从而 $|\mathbb{B}_\delta(\mathrm{Q}\ r)|\leqslant K!\ L!\ \ |\mathbb{B}_{\mathsf{F}}(\mathrm{Q}\ r)|$。如果 Q_1, $\mathrm{Q}_2\in\mathcal{Q}$，则有 $\|\mathrm{Q}$ $r^2/\ 2\rho\varepsilon^{\ 2}$ $r^2/\ 2\rho\varepsilon^{\ 2}=KL/8$ 2008,Q $_2\|_{\mathsf{F}}^2=4\rho^2\varepsilon^2 d_{\mathsf{H}}(\mathrm{Q}\ \mathrm{Q}\ )$，其中 d_{H}为汉明距离。然后 $\mathbb{B}_{\mathsf{F}}(\mathrm{Q}\ )\ (\mathrm{Q}\ ())$，其中 $(\ )$。Varshamov-Gilbert 引理（Tsybakov，引理 2.9）给出

$$|\mathcal{Q}'|\cdot|\mathbb{B}_{\mathsf{H}}(\mathbf{Q},KL/8)|\leqslant 2^{KL}$$

其中 $\mathcal{Q}'$ 是一个由矩阵组成的极大子集 $\mathbf{Q}$，这些矩阵彼此 $KL/8$-分离，并且 $|\mathcal{Q}'|\geqslant 2^{KL/8}$。因此我们得到

$$|\mathcal{Q}|\geqslant\frac{2^{KL}}{|\mathbb{B}_\delta(\mathbf{Q},r)|}\geqslant\frac{2^{KL}}{K!L!\,|\mathbb{B}_{\mathsf{F}}(\mathbf{Q},r)|}=\frac{2^{KL}}{K!L!\,|\mathbb{B}_{\mathsf{H}}(\mathbf{Q},KL/8)|}\geqslant\frac{2^{KL}}{K!L!2^{KL}2^{-KL/8}}$$
$$=\exp\Big(\frac{KL}{8}\log 2-\log(K!L!)\Big)\geqslant\exp\Big(\frac{KL}{16}-K\log K-L\log L\Big).$$

最后一项在 K 和 L 大于某些常数时，大于 $e^{KL/32}$。 □

引理 9 的证明。

第一个断言。令 $\tau_1,\tau_2\colon[0,1]\to[0,1]$ 是两个保测双射。我们想证明

$$\|W_{\mathbf{Q}_1}-W_{\mathbf{Q}_2}\circ(\tau_1\otimes\tau_2)\|_{\mathbb{L}_2}^2\geqslant\varepsilon^2/8.$$

对于任意 $k,\ k'\in[K]$，令 $\omega_{k,\ k'}=\lambda\big([(k-1)/K,\ \ k/K]\cap\tau_1^{-1}\big([(k'-1)/K,\ \ k'/K]\big)\big)$ 其中 λ是 $[0,1]$上的勒贝格测度。类似地，对于任意 $\ell,\ \ell'\in[L]$，令 $\omega'_{\ell,\ \ell'}=\lambda\big([(\ell-1)/L,\ \ \ell/L]\cap\tau_2^{-1}\big([(\ell'-1)/L,\ \ \ell'/L]\big)\big)$。

我们有 $\sum_k\omega_{k,\ k'}=1/K=\sum_{k'}\omega_{k,\ k'}$，也就是说，矩阵 $M\omega$ 是双随机的。对于 $[K]$ 的任意置换 π，记 $A(\pi)$ 为相应的置换矩阵。根据伯克霍夫-冯·诺依曼定理，$M\omega$ 是置换矩阵的凸组合，因此存在正数 γ_π，使得 $\omega=\sum_\pi\gamma_\pi A(\pi)$ 和 $\sum_\pi\gamma_\pi=1/K$，其中求和遍历 $[K]$ 的所有置换。因此

$$\|W_{\mathbf{Q}_1}-W_{\mathbf{Q}_2}\circ(\tau_1\otimes\tau_2)\|_{\mathbb{L}_2}^2=\sum_{k,k'\in[K]}\sum_{\ell,\ell'\in[L]}\omega_{k,k'}\omega'_{\ell,\ell'}|Q_{1,k,\ell}-Q_{2,k',\ell'}|^2$$
$$=\sum_{\pi,\nu}\sum_{k,k'\in[K]}\sum_{\ell,\ell'\in[L]}\gamma_\pi\gamma_\nu A(\pi)_{k,k'}A(\nu)_{\ell,\ell'}|Q_{1,k,\ell}-Q_{2,k',\ell'}|^2$$

$$= \sum_{\pi,\nu} \sum_{k\in[K]} \sum_{\ell\in[L]} \gamma_\pi \gamma_\nu |Q_{1,k,\ell} - Q_{2,\pi(k),\nu(\ell)}|^2$$

$$= \sum_{\pi,\nu} \gamma_\pi \gamma_\nu \underbrace{\|\mathbf{Q}_1 - \mathbf{Q}_2^{\pi,\nu}\|_{\mathsf{F}}^2}_{\geqslant \rho^2\varepsilon^2 KL/8} \geqslant \frac{\rho^2\varepsilon^2}{8}$$

因此，引理的结论得证。

第二个断言。令 $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_n)$ 为一个由 n 个独立同分布的随机变量组成的向量，这些随机变量在 $[K]$上均匀分布。我们还记 $\boldsymbol{\chi} = (\chi_1, \ldots, \chi_m)$ 为一个由 m 个独立同分布的随机变量组成的向量，这些随机变量在 $[L]$ 上均匀分布。令 $\boldsymbol{\Theta}_1 \in [0,1]^{n\times m}$ 为其元素为 $(\Theta_1)_{ij} = (Q_1)_{\zeta_i,\chi_j}$ 的矩阵。假设 $\mathbf{H}_1$ 在给定 $\boldsymbol{\zeta}$ 和 $\boldsymbol{\chi}$ 的条件下，是根据参数为 N 且概率矩阵为 $\boldsymbol{\Theta}_1$ 的二项模型抽样得到的矩阵。注意，$\mathbf{H}_1$ 的分布为 $\mathbf{P}^{W_{\mathbf{Q}_1}}$。对于任意 $\mathbf{a} \in [K]^n$、$\mathbf{b} \in [L]^m$ 和 $\mathbf{H} \in \{0,1\}^{n\times m}$，我们引入 $\alpha_{\mathbf{a}} = \mathsf{P}(\boldsymbol{\zeta} = \mathbf{a})$、$\beta_{\mathbf{b}} = \mathsf{P}(\boldsymbol{\chi} = \mathbf{b})$ 以及 $p^{(1)}_{\mathbf{Hab}} = \mathsf{P}(\mathbf{H}_1 = \mathbf{H} | \boldsymbol{\zeta} = \mathbf{a}\ \boldsymbol{\xi} = \mathbf{b})$。用上面的指标替换后，我们有类似的记号 $p^{(2)}_{\mathbf{Hab}}$。然后

$$D_{\mathsf{KL}}(\mathbf{P}_{W_{\mathbf{Q}_1}}, \mathbf{P}_{W_{\mathbf{Q}_2}}) = \sum_{\mathbf{H}} \sum_{\mathbf{a}} \sum_{\mathbf{b}} \alpha_{\mathbf{a}}\beta_{\mathbf{b}} p^{(1)}_{\mathbf{Hab}} \log\left(\frac{\sum_{\mathbf{a}}\sum_{\mathbf{b}} \alpha_{\mathbf{a}}\beta_{\mathbf{b}} p^{(1)}_{\mathbf{Hab}}}{\sum_{\mathbf{a}}\sum_{\mathbf{b}} \alpha_{\mathbf{a}}\beta_{\mathbf{b}} p^{(2)}_{\mathbf{Hab}}}\right)$$

$$\text{(Jensen)} \qquad \leqslant \sum_{\mathbf{a}} \sum_{\mathbf{b}} \alpha_{\mathbf{a}}\beta_{\mathbf{b}} \sum_{\mathbf{H}} p^{(1)}_{\mathbf{Hab}} \log\left(\frac{p^{(1)}_{\mathbf{Hab}}}{p^{(2)}_{\mathbf{Hab}}}\right).$$

当 $\mathbf{a}$ 和$\mathbf{b}$ 固定时，对$\mathbf{H}$ 的求和是两个nm二项分布测度乘积之间的 Kullback-Leibler 散度，其中每个测度的参数为(N,p) 或(N,q)，其中 $p := \rho(1+\varepsilon)/2$，以及 $q := \rho(1-\varepsilon)/2$。这给出

$$D_{\mathsf{KL}}(\mathbf{P}_{W_{\mathbf{Q}_1}} || \mathbf{P}_{W_{\mathbf{Q}_2}}) \leqslant Nnm\kappa(p,q)$$

其中 $\kappa(p,q)$ 是参数分别为 p和 q的两个伯努利测度之间的 Kullback-Leibler 散度。我们有 $\kappa(p,q) \leqslant (p-q)^2(p^{-1}+q^{-1}) = 4\rho\varepsilon^2/(1-\varepsilon^2) \leqslant 16\rho\varepsilon^2/3$。引理得证。 □

9.3 证明 (33) 固定某个 $\varepsilon \in (0,1/4)$，并令 $W_1 \equiv \rho/2$ 为一个常数图子。我们还定义 $W_2(u,v) = \rho(1/2+\varepsilon)$，若 $(u,v) \in [0,1/2)^2 \cup [1/2,1]^2$ 且 $W_2(u,v) = \rho(1/2-\varepsilon)$，否则为其他地方。我们得到

$$\delta(W_1, W_2) = \rho\varepsilon\,.$$

placeholder

$$\inf_{\widehat{W}} \max_{W^*\in\{W_1,W_2\}} \mathbb{E}_{W^*}\left[\delta^2(\widehat{W}, W^*)\right] \geqslant \frac{1}{2}\left(\int \delta^2(\widehat{W}, W_1)\, d\mathbf{P}_{W_1} + \int \delta^2(\widehat{W}, W_2)\, d\mathbf{P}_{W_2}\right)$$

$$\geqslant \frac{1}{2}\int \delta^2(\widehat{W}, W_1) + \delta^2(\widehat{W}, W_2)\, \min(d\mathbf{P}_{W_1}, d\mathbf{P}_{W_2})$$

$$\geqslant \frac{\delta^2(W_2, W_1)}{4} \int \min(d\mathbf{P}_{W_1}, d\mathbf{P}_{W_2})$$

$$\geqslant \frac{\rho^2\varepsilon^2}{8} \exp(-\chi^2(\mathbf{P}_{W_1} || \mathbf{P}_{W_2})),$$

其中 $\chi^2(\mathbf{P}_{W_1} || \mathbf{P}^{W_2})$ 表示 $\mathbf{P}^{W_1}$ 与 $\mathbf{P}^{W_1}$之间的卡方散度。在最后一个不等式中，我们使用了（Tsybakov，2008）中的 (2.24) 和 (2.26)。最后，下一个引理给出了卡方散度的上界，从而使我们能够完成证明，取 $\varepsilon^2 = c_0/12\rho N\sqrt{nm}$。

引理 10. 如果 $14\sqrt{n\overline{m}N}\rho\varepsilon^2 \leqslant 1$，则 $\chi^2(\mathbf{P}_{W_1}||\mathbf{P}_{W_2}) \leqslant 1/4$。

证明。令 $L(\mathbf{H})$ 为 P_{W_2} 关于 P_{W_1} 的 Radon-Nikodym 导数。我们有 $\chi^2(\mathrm{P}_{W_1}||\mathrm{P}_{W_2}) = \mathbb{E}_{W_1}[L(\mathbf{H})^2] - 1$，因此只需证明 $\mathbb{E}_{W_1}[L(\mathbf{H})^2] \leqslant 5/4$。下文中，$\mathbb{E}[\cdot]$ 将指代 $\mathbb{E}_{W_1}[\cdot]$。我们还引入 $p_0 = \rho/2$，$p_1 = \rho(1/2+\varepsilon)$ 和 $p_2 = \rho(1/2-\varepsilon)$。为避免公式过长，我们将用记号 $\bar{p}_j = 1 - p_j$ 表示 $j = 0,1,2$。

作为图子 $W_2 \in \mathcal{W}_\rho[2,2]$，我们可以假设 $\{U_i\}$ 是参数为1/2 的独立同分布伯努利随机变量，$\{V_j\}$ 同理。给定 $\{U_i\}$ 和 $\{V_j\}$，定义集合 $S = \{(a,b) \colon U_a = V_b\}$。对于 $(i,j) \in S$（分别为 S^c），H_{ij}服从参数为(N, p_1)（分别为 (N,p_2)）的二项分布。令 μ 为 S 的分布，则有

$$L(\mathbf{H}) = \int L_S(\mathbf{H})\, d\mu(S)$$

with

$$L_S(\mathbf{H}) = \left(\frac{\bar{p}_1}{\bar{p}_0}\right)^{N|S|} \left(\frac{\bar{p}_2}{\bar{p}_0}\right)^{Nnm - N|S|} \prod_{(a,b)\in S} \left(\frac{p_1\bar{p}_0}{p_0\bar{p}_1}\right)^{H_{a,b}} \prod_{(a,b)\in S^c} \left(\frac{p_2\bar{p}_0}{p_0\bar{p}_2}\right)^{H_{a,b}}.$$

根据 Fubini 定理，我们可以写出 $\mathbb{E}[L(\mathbf{H})^2] = \int \mathbb{E}[L_{S_1}(\mathbf{H}) L_{S_2}(\mathbf{H})]\, d\mu(S_1) d\mu(S_2)$ 其中

$$\begin{aligned}\mathbb{E}[L_{S_1}(\mathbf{H}) L_{S_2}(\mathbf{H})] = &\left(\frac{\bar{p}_1}{\bar{p}_0}\right)^{(|S_1|+|S_2|)N} \left(\frac{\bar{p}_2}{\bar{p}_0}\right)^{(2nm-|S_1|-|S_2|)N} \mathbb{E}\Bigg[\prod_{(a,b)\in S_1\cap S_2} \left(\frac{p_1\bar{p}_0}{p_0\bar{p}_1}\right)^{2H_{a,b}} \\ &\times \prod_{(a,b)\in S_1^c\cap S_2^c} \left(\frac{p_2\bar{p}_0}{p_0\bar{p}_2}\right)^{2H_{a,b}} \times \prod_{(a,b)\in S_1\Delta S_2} \left(\frac{p_1 p_2 \bar{p}_0^2}{p_0^2\bar{p}_1\bar{p}_2}\right)^{H_{a,b}}\Bigg].\end{aligned}$$

我们记作

$$\begin{aligned}\left(\frac{\bar{p}_1}{\bar{p}_0}\right)^{N|S_1|+N|S_2|} &= \left(\frac{\bar{p}_1}{\bar{p}_0}\right)^{2N|S_1\cap S_2| + N|S_1\Delta S_2|} \\ \left(\frac{\bar{p}_2}{\bar{p}_0}\right)^{2Nnm - N|S_1| - N|S_2|} &= \left(\frac{\bar{p}_2}{\bar{p}_0}\right)^{2N|S_1^c\cap S_2^c| + N|S_1\Delta S_2|}.\end{aligned}$$

此外，由于期望是相对于 $\mathbf{P}_{W_1}$ 取的，且变量 H_{ij} 在给定 $\{U_i\}, \{V_j\}$ 的条件下相互独立，我们有

$$\begin{aligned}\mathbb{E}\Bigg[\prod_{(a,b)\in S} \left(\frac{q_1 q_2 \bar{p}_0^2}{p_0^2 \bar{q}_1 \bar{q}_2}\right)^{H_{a,b}}\Bigg] &= \left\{\left(\frac{q_1 q_2 \bar{p}_0^2}{p_0^2\bar{q}_1\bar{q}_2}\right) p_0 + \bar{p}_0\right\}^{N|S|} \\ &= \left\{\frac{\bar{p}_0^2}{\bar{q}_1\bar{q}_2}\left(1 + \frac{(q_1-p_0)(q_2-p_0))}{p_0\bar{p}_0}\right)\right\}^{N|S|}\end{aligned}$$

对于 $[n]\times[m]$ 的任意子集 S以及任意 $q_1,\ q_2 \in [0,1]$。将此公式应用于三元组$(S_1\cap S_2,\ p_1,\ p_1)$、$(S_1^c \cap S_2^c,\ p_2,\ p_2)$ 以及 $S_1 \Delta S_2,\ p_1,\ p_2$，我们得到

$$\mathbb{E}\Bigg[\prod_{(a,b)\in S_1\cap S_2} \left(\frac{p_1\bar{p}_0}{p_0\bar{p}_1}\right)^{2H_{a,b}}\Bigg] = \left\{\frac{\bar{p}_0^2}{\bar{p}_1^2}\left(1 + \frac{(p_1-p_0)^2}{p_0\bar{p}_0}\right)\right\}^{N|S_1\cap S_2|},$$

$$\mathbb{E}\Bigg[\prod_{(a,b)\in S_1^c\cap S_2^c} \left(\frac{p_2\bar{p}_0}{p_0\bar{p}_2}\right)^{2H_{a,b}}\Bigg] = \left\{\frac{\bar{p}_0^2}{\bar{p}_2^2}\left(1 + \frac{(p_2-p_0)^2}{p_0\bar{p}_0}\right)\right\}^{N|S_1^c\cap S_2^c|},$$

$$\mathbb{E}\Big[\prod_{(a,b)\in S_1\Delta S_2}\Big(\frac{p_1p_2\bar{p}_0^2}{p_0^2\bar{p}_1\bar{p}_2}\Big)^{H_{a,b}}\Big]=\Big\{\frac{\bar{p}_0^2}{\bar{p}_1\bar{p}_2}\Big(1+\frac{(p_1-p_0)(p_2-p_0))}{p_0\bar{p}_0}\Big)\Big\}^{N|S_1\Delta S_2|}.$$

综合这些结果，我们得到如下表达式

$$\mathbb{E}[L_{S_1}(\mathbf{H})L_{S_2}(\mathbf{H})]=A^{N|S_1\cap S_2|}B^{N|S_1^c\cap S_2^c|}C^{N|S_1\Delta S_2|}$$

with

$$A=1+\frac{(p_1-p_0)^2}{p_0(1-p_0)}=1+\frac{\rho^2\varepsilon^2}{p_0(1-p_0)}\leqslant 1+4\rho\varepsilon^2$$

$$B=1+\frac{(p_2-p_0)^2}{p_0(1-p_0)}=1+\frac{\rho^2\varepsilon^2}{p_0(1-p_0)}=A$$

$$C=1+\frac{(p_1-p_0)(p_2-p_0)}{p_0(1-p_0)}=1-\frac{\rho^2\varepsilon^2}{p_0(1-p_0)}\leqslant\frac{1}{A}.$$

这导致

$$\begin{aligned}\mathbb{E}[L_{S_1}(\mathbf{H})L_{S_2}(\mathbf{H})]&=A^{N|S_1\cap S_2|+N|S_1^c\cap S_2^c|}C^{N|S_1\Delta S_2|}\leqslant A^{N|S_1\cap S_2|+N|S_1^c\cap S_2^c|-N|S_1\Delta S_2|}\\&=A^{2N|S_1\cap S_2|+2N|S_1^c\cap S_2^c|-Nnm}\\&\leqslant\exp\Big[\Big(2|S_1\cap S_2|+2|S_1^c\cap S_2^c|-nm\Big)4N\rho\varepsilon^2\Big].\end{aligned}$$

因此，剩下的问题是界定 $T=|S_1\cap S_2|+|S_1^c\cap S_2^c|$ 的指数矩，其中 S_1 和 S_2 相互独立，且服从 μ。我们将生成 S_1 和 S_2 的相互独立变量组分别记为 $(\{U_i\},\{V_j\})$ 和$(\{U_i'\},\{V_j'\})$。这意味着 $S_1=\{(i,j)\in[n]\times[m]\colon U_i=V_j\}$ 且 $S_2=\{(i,j)\in[n]\times[m]\colon U_i'=V_j'\}$。

我们还引入随机变量

$$X=\sum_{i=1}^n U_iU_i'+(1-U_i)(1-U_i'),\quad Y=\sum_{j=1}^m V_jV_j'+(1-V_j)(1-V_j').$$

可以验证， X 和 Y 相互独立，分别服从二项分布 $\mathcal{B}$（n，$1/2$）和 $\mathcal{B}$（m，1。2），并且

$$T=XY+(n-X)(m-Y)=2(X-n/2)(Y-m/2)+nm/2.$$

霍夫丁不等式（霍夫丁，1963，式 4.16）意味着

$$\mathbb{E}[e^{\lambda(X-n/2)(Y-m/2)}]\leqslant\mathbb{E}[e^{m\lambda^2(X-n/2)^2/8}].$$

此外， $X-n/2$ 是 $n/8$-次高斯的（再次根据霍夫丁不等式），这意味着 $(X-n/2)^2-n/4$ 是次指数的，其参数为 $2n$（Rigollet和 Hütter，2015，引理 1.12）。因此，

$$\mathbb{E}[e^{m\lambda^2(X-n/2)^2/8}]\leqslant e^{nm\lambda^2/32}e^{n^2m^2\lambda^4/32}\leqslant e^{5nm\lambda^2/32}$$

在满足 $\lambda\leqslant 2/\sqrt{nm}$ 的条件下。综合所有这些结果，我们得到

$$\begin{aligned}\mathbb{E}[L(\mathbf{H})^2]=\mathbb{E}[e^{(2T-nm)4N\rho\varepsilon^2}]=\mathbb{E}[e^{(X-n/2)(Y-m/2)16N\rho\varepsilon^2}]\\\leqslant e^{40nmN^2\rho^2\varepsilon^4}\leqslant 5/4\end{aligned}$$

只要 $14\sqrt{nmN\rho\varepsilon^2}\leqslant 14\sqrt{(1/40)\log(5/4)}\approx 1.04$。这就完成了该引理的证明。

□

由于三个下界均已证明，定理的证明4 至此完成。

10 辅助结果

引理 11. 设 X 为随机变量， $a\in\mathbb{R}$、 b， c， $d\geqslant 0$ 为一些常数。 *If*

$$\mathbb{P}(X\geqslant a+bt+ct^2)\leqslant de^{-t}\qquad \textit{for all } t\geqslant 0,$$

则 $\mathbb{E}[X]\leqslant a+bd+2cd$。

证明。在 $c=0$ 的情况下，这个不等式是众所周知的。因此，我们只考虑 $c>0$ 的情况。不失一般性，假设 $a=0$ 和 $c=1$。事实上，通过考虑随机变量 $X'=(X-a)_+/c$，其中 $b'=b/c$，总可以将问题化归到这种情况。因此，对于每个 $t\geqslant 0$，我们都有 $\mathbb{P}(X\geqslant t^2+bt)\leqslant de^{-t}$。注意，条件 $b\geqslant 0$ 意味着定义在 $[0,+\infty)$ 上的映射 $t\mapsto t^2+bt$ 是双射。令 $z=t^2+bt$，这意味着

$$\mathbb{P}\big(X\geqslant z\big)\leqslant d\exp\big\{(b/2)-\sqrt{z+(b/2)^2}\big\},\qquad \forall z\geq 0.$$

由此不等式可得

$$\begin{aligned}\mathbb{E}[X]&\leqslant d\int_0^\infty \exp\big\{(b/2)-\sqrt{z+(b/2)^2}\big\}\,dz\\&=d\int_0^\infty e^{-t}(2t+b)\,dt\\&=bd+2d.\end{aligned}$$

证明完毕。 □

我们称一个均值为零的随机变量 ζ 满足(a,b)-伯恩斯坦条件，如果有

$$\mathbb{E}[e^{\lambda\zeta}]\leqslant\exp\left\{\frac{\lambda^2 a}{2(1-b|\lambda|)}\right\}\qquad \text{provided that}\quad |\lambda|\leqslant 1/b.$$

可以证明，如果 ξ 满足(a,b)-伯恩斯坦条件，那么 ξ 的方差有上界 a。事实上，由于对于每个 $x\in\mathbb{R}$都有 $x^2\leqslant 2(e^x-1-x-x^3/6)$，将 x 替换为$\lambda\xi$ （其中 λ 足够小）并取期望，可得

$$\begin{aligned}\lambda^2\mathbb{E}[\xi^2]&\leqslant 2(\mathbb{E}[e^{\lambda\xi}]-1-\lambda\mathbb{E}[\xi]-(\lambda^3/6)\mathbb{E}[\xi^3])\\&\leqslant 2\left(\exp\left\{\frac{\lambda^2 a}{2(1-b|\lambda|)}\right\}-1\right)-(\lambda^3/3)\mathbb{E}[\xi^3]=\frac{\lambda^2 a}{1-b|\lambda|}+o(\lambda^2)\end{aligned}$$

由于 $\lambda\to 0$。将上一个不等式的两边除以 λ^2，并令 λ 趋于零，可得$\mathbf{Var}[\xi]=\mathbb{E}[\xi^2]\leqslant a$。

引理 12（伯恩斯坦不等式）。设 $\boldsymbol{\xi}=(\xi_1,\ldots,\xi_n)$ 为一个坐标相互独立的零均值随机向量，设 $\boldsymbol{\alpha}\in\mathbb{R}^n$ 为一个确定性向量。假设对于某些$a,b>0$，所有 ξ_i's 均满足(a,b)-伯恩斯坦条件。那么，对于每个 $\delta\in(0,1)$，我们有

$$\mathbb{P}\Big(\boldsymbol{\alpha}^\top\boldsymbol{\xi}\leqslant\sqrt{2a\log(1/\delta)}\,\|\boldsymbol{\alpha}\|_2+b\|\boldsymbol{\alpha}\|_\infty\log(1/\delta)\Big)\geqslant 1-\delta,$$

$$\mathbb{P}\Big(|\boldsymbol{\alpha}^\top\boldsymbol{\xi}|\leqslant\sqrt{2a\log(2/\delta)}\,\|\boldsymbol{\alpha}\|_2+b\|\boldsymbol{\alpha}\|_\infty\log(2/\delta)\Big)\geqslant 1-\delta.$$

证明。不失一般性，我们假设 $\|\boldsymbol{\alpha}\|_\infty = 1$。马尔可夫不等式表明，对于任意满足 $|\lambda| \leqslant 1/b$ 的 λ，

$$\mathbb{P}(\boldsymbol{\alpha}^\top \boldsymbol{\xi} \geqslant t) \leqslant e^{-\lambda t}\mathbb{E}[e^{\lambda \boldsymbol{\alpha}^\top \boldsymbol{\xi}}] = e^{-\lambda t}\prod_{i\in[n]}\mathbb{E}[e^{\lambda \alpha_i \xi_i}] \leqslant \exp\Big(-\lambda t + \frac{\lambda^2 a\|\boldsymbol{\alpha}\|_2^2}{2(1-b|\lambda|)}\Big).$$

可以验证，对于 $h(u) = 1+u-\sqrt{1+2u},\, u>0$，

$$\sup_{|\lambda|\leqslant 1/b}\Big(\lambda t - \frac{\lambda^2 A}{2(1-b|\lambda|)}\Big) = \frac{A}{b^2}h\Big(\frac{bt}{A}\Big) \geqslant \frac{t^2}{2(A+bt)}, \qquad \forall t>0.$$

因此，我们得到

$$\mathbb{P}(\boldsymbol{\alpha}^\top \boldsymbol{\xi} \geqslant t) \leqslant \exp\Big\{-\frac{a\|\boldsymbol{\alpha}\|_2^2}{b^2}h\Big(\frac{bt}{a\|\boldsymbol{\alpha}\|_2^2}\Big)\Big\}, \qquad \forall t\in\mathbb{R}.$$

因此，若 $bt/a\|\boldsymbol{\alpha}\|_2^2 \geqslant h^{-1}(b^2 z/a\|\boldsymbol{\alpha}\|_2^2)$，则有 $\mathbb{P}(\boldsymbol{\alpha}^\top\boldsymbol{\xi}\geqslant t) \leqslant e^{-z}$。可以验证，$h^{-1}(v) = v+\sqrt{2v}$，这意味着 t 的上述条件等价于

$$t \geqslant \frac{a\|\boldsymbol{\alpha}\|_2^2}{b}\Big(\frac{b^2 z}{a\|\boldsymbol{\alpha}\|_2^2} + \frac{b\sqrt{2z}}{\sqrt{a}\|\boldsymbol{\alpha}\|_2}\Big) = bz + \|\boldsymbol{\alpha}\|_2\sqrt{2az}.$$

将 z 替换为 $\log(1/\delta)$，即可得到引理中的第一个不等式。

对于第二个不等式，我们只需指出，随机变量 $-\xi_i$ 的矩生成函数满足与 ξ_i的矩生成函数相同的假设。因此，我们有

$$\mathbb{P}\Big(\boldsymbol{\alpha}^\top\boldsymbol{\xi} \leqslant \sqrt{2a\log(1/\delta)}\,\|\boldsymbol{\alpha}\|_2 + b\log(1/\delta)\Big) \geqslant 1-\delta,$$

$$\mathbb{P}\Big(-\boldsymbol{\alpha}^\top\boldsymbol{\xi} \leqslant \sqrt{2a\log(1/\delta)}\,\|\boldsymbol{\alpha}\|_2 + b\log(1/\delta)\Big) \geqslant 1-\delta.$$

使用并集界，并将 δ 替换为 $\delta/2$，我们得到该引理的第二个结论。 □

引理 13. 令 $B>0$ 和 $N\in\mathbb{N}$。

1. 如果 ζ 是均值为零的随机变量，且 $\mathbb{P}(|\zeta|\leqslant B) = 1$，那么它满足$(a,b)$-伯恩斯坦条件，其中 $a = \mathbf{Var}[\zeta]$ 且 $b = B/3$。

2. 如果 ζ 是 N 个相互独立的零均值随机变量的平均值，且每个随机变量都取值于 $[-B,B]$，那么它满足 (a,b)-伯恩斯坦条件，其中 $a = \mathbf{方差}[\zeta]$ 和$b = B/(3N)$。

3. 如果从强度为 $N\theta$ 的泊松分布中抽取 ζ'，则 $\zeta = (\zeta' - \mathbb{E}[\zeta'])/N$满足 (a,b)-伯恩斯坦条件，其中 $a = \mathbf{Var}[\zeta]$ 且 $b = 1/(3N)$。

证明。 这些结论的证明相当标准，基于不等式 $e^z - z - 1 \leqslant 3z^2/(3-|z|)$，前提是 $|z|<3$。 □

LEMMA 14. 设 $\xi_1,\dots,\xi_N$为均值为零且相互独立的随机变量，并满足(σ_ξ^2, b_ξ)-伯恩斯坦条件。对于 $[N]$ 的每个分割 $G = \{A_1,\dots,A_k\}$，定义投影矩阵 Π_G如下：

$$(\Pi_G \boldsymbol{v})_i = \frac{1}{|A_j|}\sum_{\ell\in A_j} v_\ell \quad \textit{if} \quad i\in A_j, \quad \forall \boldsymbol{v}\in\mathbb{R}^N.$$

我们有，以至少 $1-e^{-t}$ 的概率，对所有 $t>0$

$$\forall \boldsymbol{v},\ \boldsymbol{v}^\top \Pi_G \boldsymbol{\xi} \leqslant \frac{1}{2}\|\Pi_G \boldsymbol{\xi}\|_2 \|\boldsymbol{v}\|_2 + \sigma_\xi \sqrt{2(t+\log M)}\|\boldsymbol{v}\|_2 + b_\xi (t+\log M)\|\boldsymbol{v}\|_\infty.$$

其中 $\log M \leqslant k \log 12 \leqslant 2.5k$。在同一事件上，有

$$\|\Pi_G \boldsymbol{\xi}\|_2^2 \leqslant 4(t+\log M)\big(2\sigma_\xi^2 + b_\xi \|\Pi_G \boldsymbol{\xi}\|_\infty\big).$$

证明。令 $\boldsymbol{w} \in \mathbb{R}^N$ 为满足 $\|\Pi_G \boldsymbol{w}\|_2 = 1$ 的向量。我们有 $\boldsymbol{w}^\top \Pi_G \boldsymbol{\xi} = \sum_{i\in[N]} (\Pi_G \boldsymbol{w})_i \xi_i$，其中最后一个和式中的各项相互独立。伯恩斯坦不等式给出

$$\mathbb{P}\big(\boldsymbol{w}^\top \Pi_G \boldsymbol{\xi} \leqslant \sigma_\xi \sqrt{2t} + b_\xi t \|\Pi_G \boldsymbol{w}\|_\infty\big) \geqslant 1 - e^{-t}, \qquad t \geqslant 0. \tag{38}$$

令 V_G 为 $\mathbb{R}^N$ 在 Π_G 下单位球的像，并令 $\mathcal{N}_G = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_M\}$ 为 ε 的 V_G 的一个 $\varepsilon = 1/4$。集合 V_G 包含于一个维数为 k 的线性空间的单位球中，（van Handel，2016，引理 5.13）表明 $M \leqslant 12^k$。定义

$$\boldsymbol{u}_m = \operatorname*{arg\,min}_{u \in V_G} \{\|\boldsymbol{u}\|_\infty : \|\boldsymbol{u} - \boldsymbol{w}_m\|_2 \leq 1/4\}, \qquad m = 1, \dots, M.$$

利用（38）和并集界，我们得到

$$\mathbb{P}\Big(\forall m \in [M],\ \boldsymbol{u}_m^\top \Pi_G \boldsymbol{\xi} \leqslant \sigma_\xi \sqrt{2(t+\log M)} + b_\xi (t+\log M)\|\boldsymbol{u}_m\|_\infty\Big) \geqslant 1 - e^{-t}, \qquad t \geqslant 0,$$

其中我们利用了 $\Pi_G \boldsymbol{u}_m = \boldsymbol{u}_m$ 这一事实。令 $\boldsymbol{w}$ 为 V_G 中的任意向量。令 $\boldsymbol{w}_m$ 为网 $\mathcal{N}_G$ 中满足 $\|\boldsymbol{w} - \boldsymbol{w}_m\|_2 \leq 1/4$ 的任意点。我们有

$$\|\boldsymbol{u}_m\|_\infty \leq \|\boldsymbol{w}\|_\infty \qquad \text{and} \qquad \|\boldsymbol{w} - \boldsymbol{u}_m\|_2 \leq 1/2.$$

这意味着，以至少 $1-e^{-t}$ 的概率，对于任意向量 $\boldsymbol{w} \in V_G$,

$$\begin{aligned}
\boldsymbol{w}^\top \Pi_G \boldsymbol{\xi} &= (\boldsymbol{w} - \boldsymbol{u}_m)^\top \Pi_G \boldsymbol{\xi} + \boldsymbol{u}_m^\top \Pi_G \boldsymbol{\xi} \\
&\leqslant \|\boldsymbol{w} - \boldsymbol{u}_m\|_2 \|\Pi_G \boldsymbol{\xi}\|_2 + \boldsymbol{u}_m^\top \Pi_G \boldsymbol{\xi} \\
&\leqslant (1/2)\|\Pi_G \boldsymbol{\xi}\|_2 + \sigma_\xi \sqrt{2(t+\log M)} + b_\xi (t+\log M)\|\boldsymbol{u}_m\|_\infty \\
&\leqslant (1/2)\|\Pi_G \boldsymbol{\xi}\|_2 + \sigma_\xi \sqrt{2(t+\log M)} + b_\xi (t+\log M)\|\boldsymbol{w}\|_\infty.
\end{aligned}$$

由于对于由 Π_G 作用于单位球所得的像中的任意向量 $\boldsymbol{w}$，该不等式都成立，因此对 $\boldsymbol{w} = \boldsymbol{v}/\|\Pi_G \boldsymbol{v}\|_2$ 也成立。将其代入上一个公式，然后将不等式两边乘以 $\|\Pi_G \boldsymbol{v}\|_2$，我们得到引理的第一个结论。

对于第二个断言，我们取 $v = \boldsymbol{\xi}$ 得到

$$\begin{aligned}
\|\Pi_G \boldsymbol{\xi}\|_2^2 &\leqslant 2\sigma_\xi \sqrt{2(t+\log M)}\|\Pi_G \boldsymbol{\xi}\|_2 + 2b_\xi (t+\log M)\|\Pi_G \boldsymbol{\xi}\|_\infty \\
&\leqslant (1/2)\|\Pi_G \boldsymbol{\xi}\|_2^2 + 4\sigma_\xi^2 (t+\log M) + 2b_\xi (t+\log M)\|\Pi_G \boldsymbol{\xi}\|_\infty,
\end{aligned}$$

其中我们使用了不等式 $2uv \leqslant (1/2)u^2 + 2v^2$ 对所有 $u, v \in \mathbb{R}$。重新排列上一个显示式中的各项，我们得到该引理的结论。 □

引理 15. 令 ξ_1，...，ξ_N 为相互独立的均值为零的随机变量，满足（σ_ξ^2，b_ξ）-伯恩斯坦条件，其中 σ_ξ^2，$b_\xi > 0$。令 $\{\mathcal{A}_j : j \in J\}$ 为 $\{1, \dots, N\}$ 的子集族，使得 $N_j =$ 最小值$_{A\in\mathcal{A}_j} |A| \geqslant N_0$。以至少 $1-e^{-t}$ 的概率，对于所有 $j \in J$，均有

$$\max_{A\in\mathcal{A}_j} \frac{1}{|A|}\Big|\sum_{\ell\in A} \xi_\ell\Big| \leqslant \sigma_\xi \sqrt{\frac{2(t+\log(2|J|))}{N_0} + \frac{2\log|\mathcal{A}_j|}{N_j}} + \frac{b_\xi (t+\log(2|J|))}{N_0} + \frac{b_\xi \log|\mathcal{A}_j|}{N_j}.$$

证明。使用伯恩斯坦不等式的引理12版本，我们得到

$$\mathbb{P}\Big(\Big|\sum_{\ell\in A}\xi_\ell\Big| \leqslant \sigma_\xi\sqrt{2|A|t}+b_\xi t\Big)\geqslant 1-2e^{-t},$$

由此可得

$$\mathbb{P}\bigg(\frac{1}{|A|}\Big|\sum_{\ell\in A}\xi_\ell\Big| \leqslant \sigma_\xi\sqrt{\frac{2t}{|A|}}+\frac{b_\xi t}{|A|}\bigg)\geqslant 1-2e^{-t}.$$

由上一个公式和并集界可得

$$\mathbb{P}\bigg(\max_{A\in\mathcal{A}_j}\frac{1}{|A|}\Big|\sum_{\ell\in A}\xi_\ell\Big| \leqslant \sigma_\xi\sqrt{\frac{2(t+\log|\mathcal{A}_j|)}{N_j}}+\frac{b_\xi(t+\log|\mathcal{A}_j|)}{N_j}\bigg)\geqslant 1-2e^{-t}.$$

对 $j\in J$ 应用并集界可得，以至少 $1-e^{-t}$ 的概率，对所有 $j\in J$，均有

$$\max_{A\in\mathcal{A}_j}\frac{1}{|A|}\Big|\sum_{\ell\in A}\xi_\ell\Big| \leqslant \sigma_\xi\sqrt{\frac{2(t+\log(2|J|)+\log|\mathcal{A}_j|)}{N_j}}+\frac{b_\xi(t+\log(2|J|)+\log|\mathcal{A}_j|)}{N_j}.$$

引理证毕。 □