arXiv:2602.03954v1 [stat.ML] 3 Feb 2026

学习 多类型异质相互作用粒子系统 S

QuanjunLang*[1]，Xiong Wang†[2]，Fei Lu‡[3]和 Mauro Maggioni§[3,4]

[1]美国达勒姆杜克大学数学系。[2]中国广州中山大学数学学院。[3]美国巴尔的摩约翰斯·霍普金斯大学数学系。[4]美国巴尔的摩约翰斯·霍普金斯大学应用数学与统计学系。

摘要

我们提出了一个框架，旨在利用多轨迹数据，对异质相互作用粒子系统中的网络拓扑、多类型相互作用核和潜在类型分配进行联合推断。这项学习任务是一个具有挑战性的非凸混合整数优化问题，我们通过一种新颖的三阶段方法加以解决。首先，我们利用智能体相互作用之间的共享结构，通过矩阵感知恢复系统参数的低秩嵌入。其次，我们在学习得到的嵌入空间中进行聚类，以识别离散的相互作用类型。第三，我们通过矩阵分解和后处理精化，恢复网络权重矩阵和核系数。在受限等距性质（RIP）假设下，我们给出了带有估计误差界的理论保证，并基于簇可分性建立了精确恢复相互作用类型的条件。在合成数据集（包括异质捕食者-猎物系统）上的数值实验表明，我们的方法能够准确重建底层动力学，并且对噪声具有鲁棒性。

目录

1 引言 2 学习框架 3 算法 4 理 **2**
论保证 5 数值实验 6 结论 A
完全网络 **7**
11
18
28
32
33
B 证明 **34**

*quanjun.lang@duke.edu；共同第一作者
†xiongwang@ualberta.ca；共同第一作者
‡feilu@math.jhu.edu§mauromaggionijhu@icloud.com

1 引言

相互作用粒子系统（IPS）广泛用于建模通过成对作用力相互作用的智能体的集体动力学。这类模型应用于生物学等多个领域，例如用于描述细胞间信号传导和集体迁移 [8, 15]；用于研究捕食者—猎物相互作用的生态学[40, 42]；用于建模颗粒和分子系统的物理学 [3, 4, 22]；以及用于刻画意见动力学和人群行为的社会科学 [20], 仅举几例。

在许多这类系统中，相互作用具有异质性：不同类型的智能体可能以不同方式彼此响应，而且不同智能体对之间的相互作用强度也可能有所不同。这种异质性可能源于物种或角色等智能体类型，也可能源于领导者与追随者之间的非对称关系，或依赖情境的行为。为捕捉这些效应，必须扩展相互作用粒子系统模型，使其纳入多个相互作用核，并根据离散类型矩阵选择相应的核；该矩阵编码了智能体类型或成对类型之间的关系。

然而，尽管从轨迹数据中学习相互作用粒子系统已取得显著进展，先前的研究仅探讨了齐次系统或异质系统的特殊情形。关于齐次相互作用粒子系统的相互作用核估计，已有研究分别针对 [33, 32, 17] 微分方程和 [35, 36, 24] 平均场方程展开。关于异质系统的推断，已有研究分别涉及 [31]核估计，以及 [25] 在类型已知时对网络和相互作用核的联合估计。

本研究探讨网络上具有多种相互作用类型的异质系统学习这一颇具挑战性的问题。我们考虑一般情形，其中网络拓扑、相互作用核以及决定成对相互作用的具体类型分配均未知。我们的目标是根据观测轨迹同时恢复这些潜在结构。由此形成一个具有挑战性的非凸混合整数优化问题，我们通过一种新颖的三阶段方法加以解决，该方法包括低秩矩阵感知、聚类和矩阵分解。

1.1 问题陈述

考虑一个异质　　网络上由 N 个智能体组成的系统，它们按照控制方程相互作用　　n

$$dX_t^i = \sum_{j\neq i} a_{ij}\Phi_{\kappa_{ij}}(X_t^j - X_t^i)dt + \sigma dW_t^i, \quad i \in [N]\,, \tag{1.1}$$

其中，$X_t^i \in \mathbb{R}^d$ 表示时刻 t 第 i 个智能体的状态，$[N] := \{1, \cdots, N\}$，并且我们将 $\sum_{j\neq i}$ 记作 $\sum_{j\in[N]\setminus\{i\}}$。我们假设图矩阵 $\mathsf{a} \in [0,1]^{N\times N}$ 是行归一化的，并确定了相互作用网络。共有 Q 种相互作用核 $\{\Phi_1, \ldots, \Phi_Q\}$，类型矩阵 $\boldsymbol{\kappa} = (\kappa_{ij})_{1\leqslant i,j\leqslant N} \in [Q]^{N\times N}$ 中取整数值的元素 $\kappa_{ij} \in [Q]$ 指定了智能体 i 与智能体 j 之间相互作用核的类型。随机力项 σdW_t^i（其中 $\{W_t^i\}$ 为 $\mathbb{R}^d$ 上的标准布朗运动）表示环境噪声；当 $\sigma = 0$ 时，系统是确定性的。在本文中，我们假设给定了一组基函数 $\{\psi_k\}$，使得每个相互作用核 $\Phi_q \in \mathcal{H} =$ 都张成 $\{\psi_k\}_{k=1}^K$，因此 Φ_q 是由系数 $\mathsf{c}^{(q)} = [c_1^{(q)}, \ldots, c_K^{(q)}]^{\mathrm{T}} \in \mathbb{R}^K$ 构成的 ψ_k 的线性组合，所以这组相互作用核 $(\Phi_1, \ldots, \Phi_Q)$ 对应于系数矩阵 c:

$$\Phi_q = \sum_{k=1}^K c_k^{(q)}\psi_k\,, \quad q \in [Q], \qquad \mathbf{c} = \left[\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \cdots, \mathbf{c}^{(Q)}\right] \in \mathbb{R}^{K\times Q}\,. \tag{1.2}$$

在本研究中，我们假设基函数 $\{\psi_k\}$是已知的，因此系统（1.1）依赖于$\mathbf{a}$，$\mathbf{c}$ 和 $\boldsymbol{\kappa}$。

我们的目标是根据由系统（ (**1.1***) **的多条轨迹组成的观测数据，估计未知的参数三元组** *。$\boldsymbol{\kappa}$*) 观测数据的形式如下:

$$\textbf{Data:} \quad \{\mathbf{X}_{t_0:t_L}^m, \dot{\mathbf{X}}_{t_0:t_L}^m\}_{m=1}^M\,, \tag{1.3}$$

其中 t_0： t_L 表示 $(t_0,\ t_1,\ t_2,\ \ldots\ldots,\ t_L)$ 且 $t_l = l\Delta t$。这里，$\mathbf{X}_t$ 和 $\mathbf{X}_{t_l}$ 分别表示系统的状态以及时刻 t_l 的速度有限差分近似：

$$\mathbf{X}_t := [X_t^1, \cdots, X_t^N]^{\mathrm{T}} \in \mathbb{R}^{dN}, \quad \dot{\mathbf{X}}_{t_l} := (\Delta t)^{-1}(\mathbf{X}_{t_{l+1}} - \mathbf{X}_{t_l}) \in \mathbb{R}^{dN}. \tag{1.4}$$

上标 m 表示独立轨迹样本，其初始条件 $\{\mathbf{X}_{t_0}^m\}$ 从 $\mu^{\otimes N}$ 中独立抽取，其中 μ 是定义在 $\mathbb{R}^d$ 上的概率分布。本文关注的情形是：智能体数量 N 适中且较大，相互作用核类型总数 Q 相对于 N 仍然较小，而假设空间维度 K 可能很大，在非参数设定下甚至可能随观测轨迹的样本量 M 增长。

例1.1（异质相互作用核系统）考虑一个包含 $N = 5$ 个智能体和 $Q = 3$ 种交互类型的一维系统，即 $\boldsymbol{\Phi} = \{\Phi_1、\Phi_2、\Phi_3\}$。假设类型矩阵为

$$\boldsymbol{\kappa}^* = \begin{bmatrix} * & 2 & 3 & 1 & 1 \\ 1 & * & 1 & 3 & 2 \\ 2 & 1 & * & 2 & 3 \\ 1 & 3 & 3 & * & 3 \\ 2 & 1 & 1 & 1 & * \end{bmatrix} \in \{1,2,3\}^{N\times N}. \tag{1.5}$$

由于排除自相互作用，对角元素无关紧要。记 $\mathbf{r}_t^{i\,j}$，：$= X_t^j - X_t^i$。控制方程 (1.1) 可明确写为

$$\begin{cases} dX_t^1 = \left[a_{12}\Phi_2(\mathbf{r}_t^{1,2}) + a_{13}\Phi_3(\mathbf{r}_t^{1,3}) + a_{14}\Phi_1(\mathbf{r}_t^{1,4}) + a_{15}\Phi_1(\mathbf{r}_t^{1,5})\right]dt + \sigma dW_t^1, \\ dX_t^2 = \left[a_{21}\Phi_1(\mathbf{r}_t^{2,1}) + a_{23}\Phi_1(\mathbf{r}_t^{2,3}) + a_{24}\Phi_3(\mathbf{r}_t^{2,4}) + a_{25}\Phi_2(\mathbf{r}_t^{2,5})\right]dt + \sigma dW_t^2, \\ dX_t^3 = \left[a_{31}\Phi_2(\mathbf{r}_t^{3,1}) + a_{32}\Phi_1(\mathbf{r}_t^{3,2}) + a_{34}\Phi_2(\mathbf{r}_t^{3,4}) + a_{35}\Phi_3(\mathbf{r}_t^{3,5})\right]dt + \sigma dW_t^3, \\ dX_t^4 = \left[a_{41}\Phi_1(\mathbf{r}_t^{4,1}) + a_{42}\Phi_3(\mathbf{r}_t^{4,2}) + a_{43}\Phi_3(\mathbf{r}_t^{4,3}) + a_{45}\Phi_3(\mathbf{r}_t^{4,5})\right]dt + \sigma dW_t^4, \\ dX_t^5 = \left[a_{51}\Phi_2(\mathbf{r}_t^{5,1}) + a_{52}\Phi_1(\mathbf{r}_t^{5,2}) + a_{53}\Phi_1(\mathbf{r}_t^{5,3}) + a_{54}\Phi_1(\mathbf{r}_t^{5,4})\right]dt + \sigma dW_t^5, \end{cases}$$

其中，$\mathbf{a}$ 表示连接各智能体的网络结构。

注1.2（完全网络）若 $\mathbf{a} = \frac{1}{N-1}\left(\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}} - \mathbf{I}_N\right)$，即对于 $i \neq j$ 和 $a_{ii} = 0$，$a_{ij} = \frac{1}{N-1}$，则该系统对应于一个具有 Q 种不同核的完全网络上的相互作用粒子系统，其中所有粒子对的网络相互作用强度相同。若 $Q = 1$，则可得到具有单一相互作用核的图上的相互作用粒子系统，如 *[25]*所研究的情形。该图可理解为编码了由粒子特定参数决定的相互作用强度。例如，若粒子的质量各不相同，网络矩阵可写为 $\mathbf{a} = \mathbf{m}\mathbf{m}^{\mathrm{T}}$，其中 $\mathbf{m} \in \mathbb{R}^N$ 包含粒子的质量。不同类型的交互在实践中很常见。

备注 1.3（核异质性）相互作用粒子系统中的相互作用核可根据不同的结构规则进行指定。在基于物种的异质性中，核仅取决于相互作用智能体所属的物种。更一般的有序对结构会为每个有序类型对 (i, j) *[33,Sec. 4]*, *[31]*指定一个不同的核。或者，核也可以仅取决于接受作用的粒子（按行依赖），或仅取决于施加作用的粒子（按列依赖） *[25,* 第 *4* 节。*3]*。我们的研究考虑最一般的情形，即类型矩阵任意且未知 $\boldsymbol{\kappa}$，并且未知 Q，从而允许相互作用模式完全异质，且不受先验结构约束。

1.2 学习框架概述

我们考虑一种常用的估计量，它通过最小化粒子预测速度经验均方误差的损失函数得到

$$
\begin{aligned}
(\widehat{\mathbf{a}}, \widehat{\mathbf{c}}, \widehat{\boldsymbol{\kappa}}) = \operatorname{argmin} \mathcal{L}(\mathbf{a}, \mathbf{c}, \boldsymbol{\kappa}), \quad \text{with } \mathcal{L}(\mathbf{a}, \mathbf{c}, \boldsymbol{\kappa}) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_i(\mathbf{a}_i, \mathbf{c}, \boldsymbol{\kappa}), \quad \text{where} \\
\mathcal{L}_i(\mathbf{a}_i, \mathbf{c}, \boldsymbol{\kappa}_i) = \frac{1}{LM} \sum_{l=1}^{L} \sum_{m=1}^{M} \left\| \dot{X}_{t_l}^{i,m} - \sum_{j \neq i} a_{ij} \sum_{k=1}^{K} c_k^{(\kappa_{ij})} \psi_k(X_{t_l}^{j,m} - X_{t_l}^{i,m}) \right\|_{\mathbb{R}^{dN}}^2 .
\end{aligned} \tag{1.6}
$$

一个核心挑战在于，(1.6) 是一个非凸混合整数优化问题。损失函数是非凸的，因为漂移项 $\sum_{j \neq i} a_{ij} \sum_{k=1}^{K} c_k^{(\kappa_{ij})} \psi_k(X_t^j - X_t^i)$ 对参数 $(\mathbf{a}、\mathbf{c}\ \boldsymbol{\kappa})$的依赖是非线性的，而整数值类型矩阵 $\boldsymbol{\kappa}$ 引入了组合结构，使得直接优化成为 NP难问题。因此，统计学和机器学习领域深入研究的经典线性回归方法，无论是参数化方法还是非参数化方法（例如 [12, 13, 33, 31, 36]），通常都不适用于此处。因此，要解决 (1.6)，就需要专门的非凸优化方法，明确处理整数值类型函数。

为解决联合估计问题 (1.6)，我们提出一种三阶段方法，将连续参数 $(\mathbf{a}、\mathbf{c})$ 的估计与离散参数 $\boldsymbol{\kappa}$的估计分开。这里仅对三阶段框架作高层次概述；损失函数、矩阵感知问题和聚类过程的正式定义见第 2 节，算法细节见第 3节。核心思想是使用矩阵变量重新参数化每个智能体 i 的交互结构

$$
\mathbf{Z}_i = \operatorname{Diag}(\mathbf{a}_i) \boldsymbol{\mathcal{K}}_i \mathbf{c}^{\mathrm{T}} \in \mathbb{R}^{N \times K}, \quad i \in [N], \tag{1.7}
$$

它编码了智能体 i 与系统其余部分的交互方式。这里 $\boldsymbol{\mathcal{K}}_i := \left(\mathbb{1}_{\{\kappa_{ij}=q\}}\right) \in \{0,1\}^{(N-1)\times Q}$是分配矩阵；更多详情请参见示例 2.1 和第 2.1 节。利用该矩阵，式（ $\mathcal{L}_i(\mathbf{a}_i,\ \mathbf{c}\ \boldsymbol{\kappa})$）中的损失函数 1.6可改写为

$$
\mathcal{L}_i(\mathbf{Z}_i) = \frac{1}{LM} \sum_{l,m=1}^{L,M} \left\| \dot{X}_{t_l}^{i,m} - \langle \mathbf{B}(\mathbf{X}_{t_l}^m)_i, \mathbf{Z}_i \rangle_F \right\|_{\mathbb{R}^d}^2 , \tag{1.8}
$$

其中 $\{\mathbf{B}_i(\mathbf{X}_{t_l}^m)\}_{m=1}^M$ 是与轨迹观测和基函数相关的感知矩阵，将在式（2.4）中定义。三个阶段如下：

- 阶段 1：矩阵感知。我们通过使用交替最小二乘（ALS）算法求解矩阵感知问题（1.8），估计集合 $\{\mathbf{Z}_i\}_{i=1}^N$ 。该算法利用了每个 $\mathbf{Z}_i$ 都是低秩的，并通过 $\mathbf{c}$ 隐式共享相同的全局核结构这一事实。本阶段从含噪轨迹数据中恢复动力学中的隐藏低维结构。

- 阶段 2：聚类。我们对 $\{\widehat{\ }\mathbf{Z}_i\}_{i=1}^N$ 的归一化非零行应用 K-均值聚类，以估计类型矩阵 $\widehat{\boldsymbol{\kappa}}$；此方法利用了$\mathbf{c}$各行之间的几何分离度，这些行代表相互作用。如果类型数量 Q未知，则通过一种基于分离度的准则来确定，该准则比较簇内角度和簇间角度。

- 阶段 3：矩阵分解与后处理。我们对估计数组 $\{\widehat{\ }\mathbf{Z}_i\}_{i=1}^N$ 进行因式分解，通过求解一个满足 $\widehat{\ }\mathbf{a}$ 行归一化约束的线性系统，得到估计的图矩阵 $\widehat{\ }\mathbf{a}$ 和估计的核系数矩阵 $\widehat{\ }\mathbf{c}$；这项操作的计算成本可以忽略不计。最后，执行后处理交替最小二乘精炼，以增强对噪声的鲁棒性。

由此得到的估计值 $(\hat{\mathbf{a}}\ \hat{\mathbf{c}}\ \hat{\boldsymbol{\kappa}})$ 给出了恢复出的网络拓扑、相互作用核和支配动力学的类型分配。

我们还为所提出的框架建立了恢复保证。在第一阶段，在感知算子满足受限等距性质（RIP）假设的条件下，我们证明矩阵感知问题（1.8）的解在弗罗贝尼乌斯范数下达到近最优精度，其界随样本量和噪声水平以最优方式缩放。在第二阶段，我们表明，当$\mathbf{Z}$ 中对应于不同相互作用类型的行在几何上充分分离时， K-均值聚类能够精确恢复类型分配 $\hat{\boldsymbol{\kappa}}$，即使 Q未知亦然。详细内容见第4节。

主要贡献。本文做出以下贡献：

1. 一种求解非凸混合整数优化问题的新颖三阶段框架。我们提出一种表述，将非凸混合整数优化问题归结为低秩矩阵感知问题，并继以几何聚类，从而解耦混合整数优化的连续与离散方面，提供一种可扩展的求解方案。

2. 恢复保证。 我们证明，在感知算子满足受限等距性质（RIP）的假设下， $\{\mathbf{Z}_i\}_{i=1}^N$ 的交替最小二乘估计量在两种弗罗贝尼乌斯范数下均达到最优速率，从而确保图和核的准确恢复。我们还给出了充分条件，以簇间和簇内角度间隔以及最小行范数来表述。在这些条件下，对估计得到的核图嵌入 $\{\mathbf{Z}_i\}_{i=1}^N$ 进行 K-均值聚类，可精确恢复 $\boldsymbol{\kappa}$，即使相互作用核类型的数量 Q未知也是如此。

3. 可扩展且鲁棒的算法。 我们设计了可扩展至中等规模 N 和 K的算法，其中包括用于增强噪声鲁棒性的后处理交替最小二乘精炼，并且同样适用于稀疏网络和完全网络。我们在合成数据上展示了这些算法，包括异质捕食者-猎物系统，并说明了相变如何随分离参数和样本量发生急剧变化。

1.3 相关工作

异质相互作用粒子系统（IPS）中的结构化相互作用学习，结合了动力系统的逆问题、网络推断以及无监督类型（或角色）发现，并应用于生物学、生态学、物理学和社会科学。

相互作用核学习。已有大量文献研究如何从均质相互作用粒子系统的轨迹中恢复相互作用核。文献 [32, 33] 中采用了强形式非参数回归。WSINDy 等弱形式方法能够抵抗噪声和不规则采样，并已扩展至多力相互作用粒子系统，例如生物系统中的吸引力—排斥力、对齐力和阻力 [36]。其他方法包括先进行密度估计，再通过稀疏回归恢复相互作用定律 [35, 37]。文献 [23, 24] 也研究了平均场相互作用粒子系统学习。

图推断与潜在类型发现。 当相互作用核取决于特定的粒子对时，由此产生的“多样物种”系统已在 [14, 21], 受到 Stein变分梯度下降模型 [27, 28]的启发并得到研究。用于恢复底层网络的方法包括根据时间序列进行邻接矩阵估计 [18]。研究者还通过对轨迹进行与模型无关的降维聚类 [43], 、动物群体中异质性的非参数检验 [41], 以及社区结构网络上的生态进化博弈动力学[5], 开展了潜在类型或角色识别。然而，这些方法通常分别处理图恢复、核估计和类型分类这三个任务，很少提供能够将三者整合起来的统一框架。

IPS中的核异质性 在假设IPS存在异质性的情况下，相互作用核的分配方式可能由不同的结构所支配。一种常见情形是 基于物种的 异质性，即核仅取决于相互作用智能体所属的物种；更一般的变体是 有序对异质性，即核取决于有序类型对 $(i,\ j)$，从而产生 $Q = P^2$ 种可能的核。在 [33,第 4] 节中， [31], 作者考虑了在具有多种智能体类型的系统中学习相互作用定律这一逆问题，其中有序对结构已预先确定。

或者，相互作用核也可能仅取决于受力粒子（按行依赖），或仅取决于施力粒子（按列依赖）；这两种情况均产生 $Q = P$ 核，并将在 [25, 第 43] 节讨论。我们的工作针对具有任意未知类型矩阵 κ且未知 Q 的一般情形，因此无需预先作出结构假设，即可允许完全异质的相互作用模式。

完全网络与一般网络 作为一种特殊情形，当交互图为完全图（见注释 1.2）或已知时，已有大量研究致力于学习相互作用核[32, 35, 36]。在这些简化情形下，同时估计类型矩阵和核系数通常更为容易，这并不令人意外。我们还针对完全图情形提供了一种专门算法（见附录 A），在这种情形下，无需先区分零相互作用与非零相互作用，便可直接对未归一化行进行聚类。相比之下，对于一般网络，必须先识别零元素，再进行聚类。我们的框架旨在兼容这两种情形。

低秩结构与混合整数难题。 在我们的设定中，针对 1.7 式中 $\mathbf{Z}_i$的分解会在堆叠后的 Q结构中诱导出秩为 $\mathbf{Z}$ 的结构，从而能够将问题简化为满足 RIP 条件 [25] 的低秩矩阵感知 问题 [6, 7] 。这与低秩优化方法 [2, 10, 19, 26, 38]以及满足 RIP 的矩阵感知理论相关。

总结。 以往研究分别推进了核估计、图恢复和异质性检测，且往往依赖简化假设（类型已知、图为稠密图）。我们提出了一个统一且可扩展的框架，用于从含噪轨迹数据中联合恢复网络拓扑、多类型相互作用核和潜在类型分配，并提供基于 RIP 和簇可分性的性能保证。

1.4 符号约定

本文在此汇总全文使用的符号约定。

标量、向量、矩阵与范数。 标量用普通字母表示（例如 a），向量和矩阵用粗体字母表示（例如 $\mathbf{x}$、$\mathbf{A}$）。对于向量 $\mathbf{x} \in \mathbb{R}^d$，我们用 $\|\mathbf{x}\|$ 表示其欧几里得范数，有时也写作 $\|\mathbf{x}\|_{\mathbb{R}^d}$ 以强调维数。对于标量 $a \in \mathbb{R}$，其绝对值记为 $|a|$ 。对于矩阵 $\mathbf{A}$，我们用 $\|\mathbf{A}\|_2$ 表示算子2-范数，用 $\|\mathbf{A}\|_{2,\ \infty}$ 表示 $(2,\ \infty)$-范数，并用 $\|\mathbf{A}\|_F$ 表示弗罗贝尼乌斯范数。

块结构。 对于具有 $N(N-1)$行的矩阵，例如$\mathbf{Z} \in \mathbb{R}^{N(N-1)\times K}$，我们用$\mathbf{Z}_i \in \mathbb{R}^{(N-1)\times K}$表示第 i 个行块，用$\mathbf{Z}_{i,\ j} \in \mathbb{R}^{1\times K}$ 表示$\mathbf{Z}_i$ 的第 j 行（因其为向量，故用粗体表示）。对于方阵$\mathbf{a} \in \mathbb{R}^{N\times N}$，我们用$\mathbf{a}_i$表示其第 i 行，用 a_{ij} 表示其第 (i,j) 个元素（因其为标量，故用普通字体表示）。因此，标量（例如 a_{ij}）与行块（例如$\mathbf{Z}_{i,j}$）通过字体和索引加以区分。

估计值与真实值。 真实量用上标星号表示（例如$\mathbf{Z}^*$、$\mathbf{a}^*$），而 $\hat{}$其估计量则用帽号表示（例如$\mathbf{Z}$、 $\hat{}\mathbf{a}$）。

本文结构安排如下。我们首先在第 2 节介绍学习框架。然后在第 3 节给出详细算法，并在第 4 节证明理论保证。我们在第 5 节给出数值示例，并在第 6 节总结全文。

表 1：索引与矩阵的符号约定

$[N]$ 等：索引集 $\{1、\ldots\ldots、N\}$ 等	$\mathbf{a} \in \mathbb{R}^{N\times N}$：图矩阵
$i,\ j \in [N]$：智能体的索引	$\boldsymbol{\kappa} \in [Q]^{N\times N}$：类型矩阵
$k \in [K]$：核的基的索引	$\mathbf{c} \in \mathbb{R}^{K\times Q}$：系数矩阵
$m \in [M]$：样本索引	$\boldsymbol{\mathcal{K}} \in \{0,1\}^{N(N-1)\times Q}$：赋值矩阵
$l \in [L]$：时刻索引	$\mathbf{u} \in \mathbb{R}^{N(N-1)\times Q}$：图类型特征矩阵
$q \in [Q]$：核类型索引	$\mathbf{Z} \in \mathbb{R}^{N(N-1)\times K}$：核图嵌入矩阵
$\mathbb{1}_{\{\ldots\}}$：表示函数	$\mathbf{1}_Q \in \mathbb{R}^Q$：长度为 Q 的全 1 向量

2 学习框架

参数三元组 $(\mathbf{a}^*,\mathbf{c}^*,\boldsymbol{\kappa}^*)$ 的联合估计会产生一个非凸混合整数优化问题，在最坏情况下属于 NP难问题。为解决这一问题，我们将由（1.6）定义的问题拆分为矩阵感知子问题，随后是聚类子问题。这种重新表述揭示了问题的结构，并支持构造易于处理的松弛问题，为下一节的算法奠定了基础。

2.1 网络、相互作用核和类型分配的约束

We 首先假设行归一化的 图矩阵 $\mathbf{a}$属于可行集合 $\mathcal{M}$ ：

$$\mathcal{M} := \left\{ \mathbf{a} = [a_{ij}] \in [0,1]^{N\times N} :\ a_{ii} = 0\ ,\ |\mathbf{a}_i|^2 = \sum_{j\neq i} a_{ij}^2 = 1\ ,\ \forall i \in [N] \right\}. \tag{2.1}$$

这种归一化消除了因缩放$\mathbf{a}$ 而产生的平凡非唯一性；其他行归一化方案也同样可行（参见 [25, 第 1.1] 节的讨论）。

现在，我们提出一种关键的重新表述，将核系数$\mathbf{c}$与类型矩阵 $\boldsymbol{\kappa}$解耦。注意到 $c_k^{(\kappa_{ij})} = \sum_{q=1}^{Q} c_k^{(q)}\, \mathbb{1}_{\{\kappa_{ij}=q\}}$，这启发我们引入 N 二元分配矩阵$\{\boldsymbol{\mathcal{K}}_i\}_{i=1}^N$：

$$\boldsymbol{\mathcal{K}} = [\boldsymbol{\mathcal{K}}_1, \cdots, \boldsymbol{\mathcal{K}}_N]^{\mathrm{T}}, \quad \text{where } \boldsymbol{\mathcal{K}}_i = \big(\boldsymbol{\mathcal{K}}_i\big)_{jq} = \big(\mathbb{1}_{\{\kappa_{ij}=q\}}\big) \in \{0,1\}^{(N-1)\times Q}, \tag{2.2}$$

其中各行按省略对角线（即当 $i = j$时）元素的顺序排列。这样，我们便可以将核—类型系数简洁地表示为 $\big(c_k^{(\kappa_{ij})}\big)_{1\leqslant k\leqslant K} = \boldsymbol{\mathcal{K}}_i \mathbf{c}^{\mathrm{T}}$。这种重新表述自然引出了一个矩阵感知问题，我们将在下一节详细讨论。

每个分配矩阵 $\{\boldsymbol{\mathcal{K}}_i\}$ 都具有若干优良性质：它是稀疏的，每一行恰有一个元素为 1，其余元素均为 0。我们在示例2.1中展示分配矩阵的构造。引入集合

$$\mathcal{B} = \left\{ B \in \{0,1\}^{(N-1)\times Q} \text{ and } B\mathbf{1}_Q = \mathbf{1}_{N-1} \right\}, \tag{2.3}$$

其中，$\mathbf{1}_Q$ 和 $\mathbf{1}_{N-1}$ 分别表示长度为 Q 和 $N-1$的全 1 向量。上述约束意味着$\boldsymbol{\mathcal{K}}$ 属于可行集合 $\mathcal{B}^N$。

例 2.1 考虑例 *1.1* 中相同的设定，其中有 $N = 5$ 个智能体和 $Q = 3$ 种交互类型。回顾 $\boldsymbol{\kappa}$ 中的类型矩阵 (1.5)。因此，补偿分配矩阵 $\boldsymbol{\mathcal{K}}_1$ 和 $\boldsymbol{\mathcal{K}}_2$如图 *1*所示。

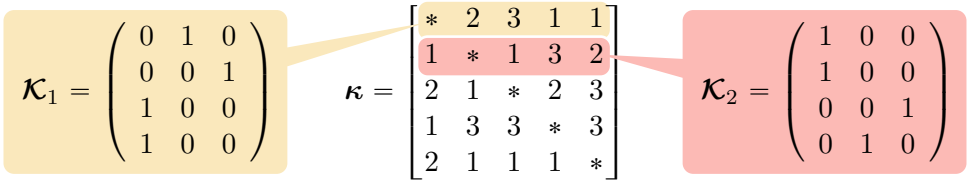

图1：从类型矩阵 $\boldsymbol{\kappa}$ 到分配矩阵$\boldsymbol{\mathcal{K}}_1$ 和 $\boldsymbol{\mathcal{K}}_2$。 $\boldsymbol{\kappa}$ 的对角元素并不重要，因为不允许自相互作用。

2.2 矩阵感知阶段

为了将学习问题（1.6）重新表述为矩阵感知问题，我们首先将漂移项转换为矩阵感知表示。对于堆叠状态向量 $\mathbf{X}_t \in \mathbb{R}^{dN}$ 在（1.4）中，我们引入三维感知张量 $\mathbf{B}_i(\mathbf{X}_t)$ ，其由基函数 $\{\psi_k\}_{k=1}^K$构造:

$$\mathbf{B}_i(\mathbf{X}_t) := (\mathbf{B}_i^{jk}(\mathbf{X}_t))_{jk} \in \mathbb{R}^{(N-1)\times K\times d}, \quad \mathbf{B}_i^{jk}(\mathbf{X}_t) = \psi_k(X_t^j - X_t^i) \in \mathbb{R}^d, \tag{2.4}$$

其中，按照惯例，我们省略自相互作用项 $j = i$。

我们引入图类型特征矩阵

$$\mathbf{u}_i = \mathrm{Diag}(\mathbf{a}_i)\boldsymbol{\mathcal{K}}_i \in \mathbb{R}^{(N-1)\times Q}, \quad \forall\, i \in [N] \tag{2.5}$$

以及核图嵌入矩阵

$$\mathbf{Z}_i := \mathrm{Diag}(\mathbf{a}_i)\boldsymbol{\mathcal{K}}_i\mathbf{c}^{\mathrm{T}} = \mathbf{u}_i\mathbf{c}^{\mathrm{T}} \in \mathbb{R}^{(N-1)\times K}, \quad \forall\, i \in [N], \tag{2.6}$$

其中，$\mathrm{Diag}(\mathbf{a}_i) \in \mathbb{R}^{(N-1)\times(N-1)}$ 是邻接向量$\mathbf{a}_i$ 的对角矩阵，并按惯例去掉 $j = i$。利用第 i个分配矩阵 $\boldsymbol{\mathcal{K}}_i^{jq} = \mathbb{1}_{\{\boldsymbol{\kappa}_{ij}=q\}}$，我们可以写出

$$a_{ij}\sum_{q=1}^{Q} c_k^{(q)}\mathbb{1}_{\{\boldsymbol{\kappa}_{ij}=q\}} = a_{ij}\sum_{q=1}^{Q}\boldsymbol{\mathcal{K}}_i^{jq}\mathbf{c}_k^{(q)} = \mathrm{Diag}(\mathbf{a}_i)\boldsymbol{\mathcal{K}}_i\mathbf{c}^{\mathrm{T}} = \mathbf{Z}_i\,.$$

结合上述记号，我们将模型（1.1）中的漂移项转换如下:

$$\begin{aligned}\sum_{j\neq i} a_{ij}\sum_{k=1}^{K} c_k^{(\boldsymbol{\kappa}_{ij})}\psi_k(X_t^i - X_t^j) &= \sum_{j\neq i}\sum_{k=1}^{K}\left(\sum_{q=1}^{Q} a_{ij}c_k^{(q)}\mathbb{1}_{\{\boldsymbol{\kappa}_{ij}=q\}}\right)\psi_k(X_t^i - X_t^j)\\ &= \langle \mathbf{B}_i(\mathbf{X}_t), \mathbf{Z}_i\rangle_F, \quad \forall i \in [N],\end{aligned}$$

其中 $\langle\cdot,\ \cdot\rangle_F$表示 $\mathbb{R}^{(N-1)\times K}$ 中作用于三维张量 $\mathbf{B}_i(\mathbf{X}_t) \in \mathbb{R}^{(N-1)\times K\times d}$ 前两个维度与矩阵 $\mathbf{Z}$ $_i \in \mathbb{R}^{(N-1)\times K}$的弗罗贝尼乌斯内积，所得结果为 $\mathbb{R}^d$中的一个向量。

注意，每个矩阵 $\mathbf{Z}_i$ 都是低秩的：秩为 $(\mathbf{Z}_i) \leqslant Q(< N)$ ，因为 $\mathbf{Z}_i = \mathbf{u}_i\mathbf{c}^{\mathrm{T}}$ 且 $\mathbf{u}_i \in \mathbb{R}^{(N-1)\times Q}$。将矩阵堆叠为 $\mathbf{Z}$:= $[\mathbf{Z}_1,\ \cdots,\ \mathbf{Z}_N]^{\mathrm{T}} \in \mathbb{R}^{N(N-1)\times K}$，并将 $\mathbf{u}$ 堆叠为 = $[\mathbf{u}_1,\ \cdots,\ \mathbf{u}_N]^{\mathrm{T}} \in \mathbb{R}^{N(N-1)\times Q}$，我们注意到

$$\mathbf{Z} = \mathbf{u}\mathbf{c}^{\mathrm{T}},$$

这意味着矩阵$\mathbf{Z}$ 是低秩的： $rank(\mathbf{Z}) \leqslant Q$。

因此，我们可以将（1.6）中的学习问题拆分为矩阵感知子问题，即估计$\mathbf{Z} = [\mathbf{Z}_1, \cdots, \mathbf{Z}_N]^{\mathrm{T}}$，然后再将$\mathbf{Z}_i$因子分解为 $(\mathbf{a}_i, \boldsymbol{\mathcal{K}}_i, \mathbf{c})$ 中的（2.6）。矩阵感知子问题为

$$\widehat{\mathbf{Z}} = \underset{rank(\mathbf{Z})\leqslant Q}{\operatorname{argmin}} \mathcal{L}(\mathbf{Z}), \quad \mathcal{L}(\mathbf{Z}) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_i(\mathbf{Z}_i), \tag{2.7}$$

$$\mathcal{L}_i(\mathbf{Z}_i) = \frac{1}{LM}\sum_{l,m=1}^{L,M}\left\|\dot{X}_{t_l}^{i,m} - \langle \mathbf{B}_i(\mathbf{X}_{t_l}^m), \mathbf{Z}_i\rangle_F\right\|_{\mathbb{R}^d}^2, \tag{2.8}$$

其中感知张量$\mathbf{B}_i(\mathbf{X}_{t_l}^m)$ 的定义见（2.4）。

我们可以借鉴矩阵感知文献中丰富的理论与算法。在理论方面，受限等距性质（RIP）条件保证，即使数据有限，也能通过可证明收敛的优化程序恢复低秩矩阵 [2, 10, 19, 26, 38]。

多种求解器可用于矩阵感知问题（2.7），包括核范数最小化、交替最小二乘和奇异值阈值法。特别地，已知目标秩会使交替最小二乘算法在实践中尤为有效。或者，也可以放弃显式的低秩约束，直接采用普通最小二乘，这一方法在 [25]中有深入探讨。我们将这些算法的完整描述留至第3.1节。

2.3 聚类阶段

在第二阶段，我们恢复核类型$\boldsymbol{\mathcal{K}}$，其取自估计张量$\mathbf{Z}$ （2.7），方法是将其转化为聚类问题。由（2.5）可知，每个$\mathbf{u}_i$ 都是稀疏矩阵，至多每行含一个非零元素。因此，$\mathbf{Z}_i$ 的各行至多指向 Q 个不同方向（参见示例2.2 如下）。这一观察促使我们对$\mathbf{Z}_i$ 的各行按其归一化方向向量进行聚类。

示例 2.2 考虑示例 *1.1* 中的相同设定，其中有 $N = 5$ 个智能体和 $Q = 3$ 种交互类型。回忆第一个粒子的类型矩阵，如图 *1*所示；我们可以将其与第一个粒子的邻接行向量相乘$\mathrm{Diag}(\mathbf{a}_1)$，从而得到

$$\boldsymbol{\mathcal{K}}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{u}_1 = \mathrm{Diag}(\mathbf{a}_1)\boldsymbol{\mathcal{K}}_1 = \begin{bmatrix} 0 & a_{12} & 0 \\ 0 & 0 & a_{13} \\ a_{14} & 0 & 0 \\ a_{15} & 0 & 0 \end{bmatrix}, \tag{2.9}$$

这是一个稀疏矩阵。因此，真实的$\mathbf{Z}_1$ 矩阵为

$$\mathbf{Z}_1 = \mathbf{u}_1\mathbf{c}^{\mathrm{T}} = \begin{bmatrix} a_{12}(\mathbf{c}^{(2)})^{\mathrm{T}} \\ a_{13}(\mathbf{c}^{(3)})^{\mathrm{T}} \\ a_{14}(\mathbf{c}^{(1)})^{\mathrm{T}} \\ a_{15}(\mathbf{c}^{(1)})^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} a_{12}c_1^{(2)} & a_{12}c_2^{(2)} & \cdots & a_{12}c_K^{(2)} \\ a_{13}c_1^{(3)} & a_{13}c_2^{(3)} & \cdots & a_{13}c_K^{(3)} \\ a_{14}c_1^{(1)} & a_{14}c_2^{(1)} & \cdots & a_{14}c_K^{(1)} \\ a_{15}c_1^{(1)} & a_{15}c_2^{(1)} & \cdots & a_{15}c_K^{(1)} \end{bmatrix}. \tag{2.10}$$

理想情况下，估计矩阵$\mathbf{Z}_1$ 将接近真实矩阵$\mathbf{Z}_1$，，后者由沿着 $\{\mathbf{c}^{(i)}\}_{i=1}^3$方向的四行组成。因此，我们预计 $\mathbf{Z}_1$ 的最后两行可以归为同一个簇。

然而，当两个相互作用核过于相似，或相关的网络权重较小时，它们便难以区分，尤其是在存在噪声的情况下。为实现可靠的聚类，某些支配分离度、且源于核与权重矩阵的参数必须

足够大。我们首先将真实核系数$\mathbf{c}^* = [\mathbf{c}^{*(1)}, \cdots, \mathbf{c}^{*(Q)}]$的预言机最小类间角 定义为

$$\theta_{min} := \min_{1 \leqslant p < q \leqslant Q} \arccos \frac{\langle \mathbf{c}^{*(p)}, \mathbf{c}^{*(q)} \rangle}{\|\mathbf{c}^{*(p)}\| \|\mathbf{c}^{*(q)}\|}. \tag{2.11}$$

接下来，我们定义真实矩阵 $\mathbf{Z}^*$ 的预言机最小行2范数，

$$z_{min} = \min_{1 \leqslant i,j \leqslant N, a_{ij} \neq 0} \left\| \mathbf{Z}^*_{i,j} \right\|, \tag{2.12}$$

其中，$\mathbf{Z}^*_{i,j}$ 表示第 i 个块 $\mathbf{Z}^*_i$ 的第 j 行。注意， z_{min} 将图矩阵中最小的非零元素与最小核系数范数联系起来:

$$z_{min} \geqslant \left(\min_{1 \leqslant i,j \leqslant N, a^*_{ij} \neq 0} a^*_{ij} \right) \cdot \left(\min_{1 \leqslant q \leqslant Q} \|\mathbf{c}^{*(q)}\| \right). \tag{2.13}$$

这些参数的影响见图 2 。保证不同核类型可分的 θ_{min} 和 z_{min} 条件见第4.3节。

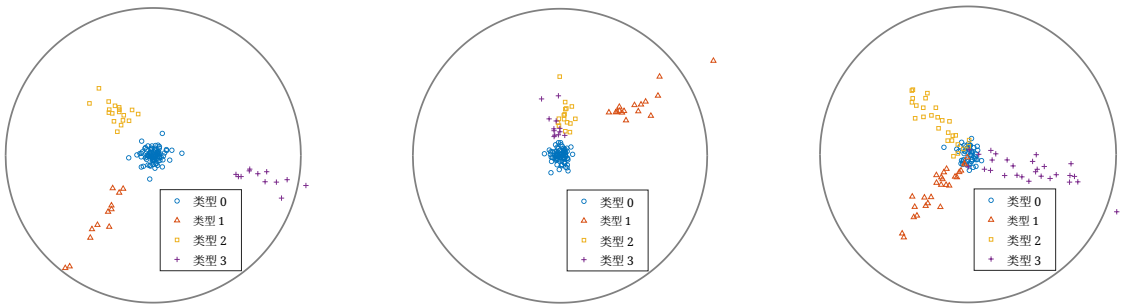

图 2：在二维假设空间上，根据含噪数据估计得到的矩阵 $\mathbf{Z}$ 中的行向量示例（$K = 2$），其中非零行有 $Q = 3$ 种类型。类型 0 对应图上的一个 0 元素。**左：** 由于角度 θ_{min} 和行范数 z_{min}足够大，所有类型都能清晰区分。**中：** 由于 $\theta_{\min}$ 较小，不同类型的非零行变得无法区分。**右：** 由于 $z_{\min}$ 较小，零行与非零行变得无法区分。

我们将对 K的归一化行应用 - 均值聚类$\mathbf{Z}$，以估计分配矩阵$\{\boldsymbol{\mathcal{K}}_i\}$，因此也就等价地估计类型矩阵 $\boldsymbol{\kappa}$。完整的聚类算法详见第3.2节。给定参数 θ_{min} 和 z_{min}的先验信息后，算法的性能在定量上取决于噪声水平和可用数据量，详见第4.3节。

2.4 矩阵分解阶段

在第三阶段，我们通过采用包含约束（**2.1**）的线性回归，将$\mathbf{Z}$ 与核类型矩阵 $\hat{\boldsymbol{\kappa}}$ 进行分解，从而恢复权重矩阵 a 和系数矩阵$\mathbf{c}$。

首先，根据$\mathbf{Z}_i$ 中的（2.6），回顾一下，$\mathbf{Z}_i = \mathrm{Diag}(\mathbf{a}_i)\boldsymbol{\mathcal{K}}_i\mathbf{c}^{\mathrm{T}} = \mathbf{u}_i\mathbf{c}^{\mathrm{T}}$ 的每一行都是$\mathbf{c}$对应列的缩放版本。设核系数向量的范数为 $v_q = \|\mathbf{c}^{(q)}\|$，则$\mathbf{c}$ 可以通过归一化核系数向量$\mathbf{v}^q$ 重建为

$$\mathbf{c} = [v_1\mathbf{v}_1^{\mathrm{T}}, \ldots, v_Q\mathbf{v}_Q^{\mathrm{T}}] \in \mathbb{R}^{Q \times K}. \tag{2.14}$$

此外，范数 $\{v_q\}_{q=1}^{Q}$ 由线性系统确定

$$\mathbf{A}\boldsymbol{\eta} = \mathbf{1}_N, \tag{2.15}$$

其中 $\mathbf{1}_N$ 是 N 维全 1 向量，且

$$\boldsymbol{\eta} = \begin{bmatrix} 1/v_1^2 \\ \vdots \\ 1/v_Q^2 \end{bmatrix} \in \mathbb{R}^Q, \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_N \end{bmatrix} \in \mathbb{R}^{(N-1)\times Q}, \quad \mathbf{A}_i = [z_{i,1}^2, \ldots, z_{i,N}^2]\boldsymbol{\mathcal{K}}_i \in \mathbb{R}^{1\times Q}. \tag{2.16}$$

公式（2.15）推导如下。注意

$$z_{i,j} = a_{ij} v_{\kappa_{ij}} \tag{2.17}$$

是$\mathbf{Z}$ 的行范数，其中 $i \neq j$，而满足 $i = j$ 的索引可以省略，因为 $a_{ii} = 0$。对$\mathbf{a}$ 的行归一化约束意味着

$$1 = \sum_{j=1, j\neq i}^{N} a_{ij}^2 = \sum_{j=1, j\neq i}^{N} \frac{z_{i,j}^2}{v_{\kappa_{ij}}^2}, \quad i = 1, \ldots, N,$$

由此得到（2.15）。

因此，给定估计得到的分配矩阵 $\boldsymbol{\mathcal{K}}$ （进而得到类型矩阵 $\hat{\boldsymbol{\kappa}}$）和聚类中心 $\{\hat{\mathbf{v}}_q\}$，我们可以先通过线性回归求解 $\{v_q\}$，根据（2.15），然后根据（**2.14**）求得$\mathbf{a}$，根据（2.17）求得$\mathbf{c}$。注意，将估计矩阵 $\mathbf{Z}$ 分解为图矩阵 $\mathbf{a}$ 和核系数$\mathbf{c}$ 的计算成本可以忽略不计。最后，通过后处理交替最小二乘精炼来增强对噪声的鲁棒性。分解和后处理的细节见第3.3节。

注释 2.3（完全网络）许多研究致力于研究异质相互作用粒子系统（*IPS*），但未考虑其底层相互作用网络（参见 *[31, 35]* 及其中的参考文献）。这些研究通常假设网络是完全网络，即 $a_{ij} \equiv 1/N$。在这种设定下，无需预先知道粒子类型，即可使用本文提出的同一框架估计多类型相互作用核。为说明这一点，考虑（1号粒子$\boldsymbol{\mathcal{K}}_1$ 的类型矩阵，与（2.9）中给出的相同。则（（2.9）和（（2.10）中的相应矩阵$\mathbf{u}_1$ 和 $\mathbf{Z}_1$ 分别变为

$$\mathbf{u}_1 = \frac{1}{N}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad and \quad \mathbf{Z}_1 = \begin{bmatrix} (\mathbf{c}^{(2)})^{\mathrm{T}} \\ (\mathbf{c}^{(3)})^{\mathrm{T}} \\ (\mathbf{c}^{(1)})^{\mathrm{T}} \\ (\mathbf{c}^{(1)})^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} c_1^{(2)} & c_2^{(2)} & \cdots & c_K^{(2)} \\ c_1^{(3)} & c_2^{(3)} & \cdots & c_K^{(3)} \\ c_1^{(1)} & c_2^{(1)} & \cdots & c_K^{(1)} \\ c_1^{(1)} & c_2^{(1)} & \cdots & c_K^{(1)} \end{bmatrix}.$$

我们的三阶段优化框架为这一类问题带来了更优且更鲁棒的聚类性能。我们可以在框架的第二阶段，对 $\mathbf{Z}_1$ 的行应用位置聚类。其关键优势在于灵活性：我们的框架放宽了在完全网络设定中$\mathbf{u}_1$ 必须符合固定矩阵结构这一约束。这种额外的灵活性可以提升鲁棒性。

3 算法

我们的主算法与上文所述的三个阶段相对应：前两个阶段将混合整数优化问题松弛为矩阵感知问题和聚类问题，第三阶段通过分解与后处理，以鲁棒的方式得到最终估计量。

1. **矩阵感知：** 估计核图嵌入矩阵 $\{\mathbf{Z}_i\}$，其定义见（2.6）。我们采用交替最小二乘（ALS）算法，这种方法类似于低秩矩阵感知中常用的分解技术（例如，[6, 7]）。2. **聚类：** 通过对 $\{\mathbf{Z}_i\}$的归一化行进行聚类，估计类型矩阵 $\boldsymbol{\kappa}$。簇的数量可以预先给定，也可以在聚类过程中估计。

3. **分解与后处理：** 使用学习得到的类型矩阵 $\hat{\boldsymbol{\kappa}}$ 估计 $\{\mathbf{a}_i\}$ 和 $\mathbf{c}$。可以直接通过对 $\{\mathbf{a}_i\}$ 施加行归一化约束来求解它们。为进一步提升性能，我们再使用 ALS，基于 $\hat{\boldsymbol{\kappa}}$ 估计 $\{\mathbf{a}_i\}$ 和 $\mathbf{c}$。

此外，第3.4节讨论了我们主要算法的变体实现和扩展，包括完全网络的情形，以及用于联合学习 ($\mathbf{a}$ $\boldsymbol{\kappa}$、$\mathbf{c}$)的三向交替最小二乘算法。

3.1 矩阵感知

在第一阶段，我们处理秩为Q 的矩阵感知问题（2.7）。如果已知 Q 的真实值，我们就直接使用它；否则，我们以一个上界进行初始化，然后再对其进行迭代细化。

回顾一下，对于每个 $i \in [N]$，我们有 $\mathbf{Z}_i = \mathbf{u}_i\mathbf{c}^{\mathrm{T}} \in$（2.6）。在这一表示中，图类型特征矩阵 $\mathbf{u}_i$ 编码了第 i个粒子的邻接和类型信息，而系数矩阵 $\mathbf{c}$ 表示 Q 个不同的相互作用核。此外，请注意，当 $\mathbf{c}$ 固定时，损失函数 $\mathcal{L}_i(\mathbf{Z}_i) = \mathcal{L}_i(\mathbf{u}_i、\mathbf{c}) \in$（2.8）关于 $\mathbf{u}_i$ 是二次的，反之亦然。因此，我们可以使用交替最小二乘（ALS）方案来估计 $\mathbf{Z}_i$：从初始估计值 $\hat{}\mathbf{u}$ 开始，交替使用最小二乘法更新 $\hat{}\mathbf{c}$ 和 $\hat{}\mathbf{u}$，直到收敛。

- **更新 $\mathbf{c}$：** 给定 估计值 $\hat{}\mathbf{u}$ 固定后，我们通过求解以下最小二乘问题来更新$\mathbf{c}$ ：

$$\min_{\mathbf{c}\in\mathbb{R}^{K\times Q}} \mathcal{L}(\hat{\mathbf{u}}, \mathbf{c}) = \frac{1}{LMN} \sum_{l,m,i=1}^{L,M,N} \left\| \dot{X}^i_{t_l,m} - \langle \mathbf{B}(\mathbf{X}^m_{t_l})_i, \hat{\mathbf{u}}_i\mathbf{c}^{\mathrm{T}} \rangle_F \right\|^2_{\mathbb{R}^d} . \tag{3.1}$$

- **更新$\mathbf{u}$：** 给定新的估计值 $\hat{}\mathbf{c}$ 固定后，我们通过求解相应的最小二乘问题，更新每个$\mathbf{u}_i$ （ $i \in [N]$ ）：

$$\min_{\mathbf{u}_i\in\mathbb{R}^{(N-1)\times Q}} \mathcal{L}_i(\mathbf{u}_i, \hat{\mathbf{c}}) = \frac{1}{LM} \sum_{l,m}^{L,M} \left\| \dot{X}^i_{t_l,m} - \langle \mathbf{B}(\mathbf{X}^m_{t_l})_i, \mathbf{u}_i\hat{\mathbf{c}}^{\mathrm{T}} \rangle_F \right\|^2_{\mathbb{R}^d} . \tag{3.2}$$

重复上述步骤，直至参数的相对变化低于预设阈值 ε ，或达到最大迭代次数 $n_{maxiter}$ 。请注意，使用相同的估计值 $\hat{}\mathbf{c}$ 并行更新所有 $\{\mathbf{u}_i\}_{i=1}^N$ ，随后使用这些 $\{\hat{}\mathbf{u}_i\}_{i=1}^N$ 更新$\mathbf{c}$。完整过程总结于算法1中。矩阵感知问题的适定性，以及 ALS 等相关算法有效性的保证，都关键依赖于受限等距性质（RIP），如假设4.3所述。关于 ALS 收敛性的分析，以及可辨识性和鲁棒估计所需临界数据量的讨论，将推迟到第4.2节。

注 *3.1*（变换不变性）通常，矩阵感知阶段无法恢复分解 $Z_i = u_i c^{\mathrm{T}}$ （见式 *(2.6)*）中的因子 u_i （因此也无法恢复 a_i、K_i）和 c。事实上，对于任意可逆矩阵 $S \in \mathbb{R}^{Q\times Q}$，都有 $\hat{Z}_i = \hat{u}_i\hat{c}^{\mathrm{T}} = (\hat{u}_i S)(S^{-1}\hat{c})^{\mathrm{T}}$，这表明该分解在此类变换下具有不变性。通过施加整数约束 $K_i \in \mathcal{B}$即可消除这种歧义；我们将在下一阶段通过 K-均值聚类处理这一约束。

ALS 的计算复杂度 ALS 的计算复杂度由每次迭代中求解最小二乘问题的代价以及收敛所需的迭代次数决定。在每次迭代中，需要 $\mathcal{O}(LMN \cdot d^2QK)$次运算来更新 $\mathbf{c}$，并需要 $\mathcal{O}(LM \cdot d^2QN)$次运算来更新每个 $\mathbf{u}_i$，因此总复杂度为 $\mathcal{O}(LMN \cdot d^2Q(K+N))$，假设 Q和 d相较于 N 和 K较小。迭代次数取决于矩阵感知问题的 RIP 条件。对于 $Q = 1$， [26]在适当的 RIP条件下，经过 ε次 $\mathcal{O}\left(^{\log N^2+}_{\log N^2}{}^{(1/\varepsilon)}\right)$ $\log\log$迭代即可收敛至精度 $Q > 1$。关于一般 $Q > 1$的收敛行为~~，似乎鲜~~有研究。我们在表2中总结了复杂度结果。

表2：交替最小二乘的计算复杂度

组成部分	复杂度	备注
更新 $\mathbf{c}$	$\mathcal{O}(LMN \cdot d^2QK)$	对所有样本求解最小二乘
更新每个 $\mathbf{u}_i$	$\mathcal{O}(LM \cdot d^2QN)$	并行更新所有 $\mathbf{u}_i$
每次迭代的总数	$\mathcal{O}(LMN \cdot d^2Q(K+N))$	假设 Q和 d相对于.....较小 N 和 K
迭代次数（$Q=1$）	$\mathcal{O}\left(\frac{\text{对数}\ N^2 + \text{对数}\ (1/\varepsilon)}{N^2\ \text{对数}\ \text{对数}}\right)$	在 [26],RIP 条件下
迭代次数（$Q>1$）	未知	当 $Q>1$时，收敛性仍是一个悬而未决的问题

1：过程 交替最小二乘（ALS）($\{\mathrm{X}_{t_l}^m\}^L$，$_{l=1}^{M}$，$_{m=1}$，$\{\psi_k\}_{k=1}^p$，ε，$n_{maxiter}$) 2：根据基函数计算导数 $\{\Delta\mathrm{X}_{t_l}^m\}$ 和感知张量 $\{\mathrm{B}(\mathrm{X}_{t_l}^m)\}_{l}$、$_m$。3：如果 True Q 未知，则、4：选择相对较大的 $Q = Q$。5：随机选取初始值 $\widehat{U}_0 \in \mathbb{R}^{N(N-1)\times Q}$。6：对于 $\tau = 1$，....., $n_{maxiter}$ 执行 7：使用带有 $\hat{}\mathrm{u} = \hat{}\mathrm{u}_{\tau-1}$ 的最小二乘法，在（3.1）中求解 $\hat{}\mathrm{c}_\tau = \mathrm{argmin}_{\mathbf{c}}\,\mathcal{L}(\hat{}\mathrm{u}, \mathrm{c})$。8：对于 $i = 1$，.....， N执行 9：使用带有 $\hat{}\mathrm{c} = \hat{}\mathrm{c}_\tau$ 的最小二乘法，在（3.2）中求解 $\hat{}\mathrm{u}_{\tau、i} = \mathrm{argmin}_{\mathbf{u}_i}\,\mathcal{L}(\mathrm{u}_{i、}\ \hat{}\mathrm{c})$。10：构造矩阵 $\hat{}\mathrm{u}_\tau = [\hat{}\mathrm{u}_{\tau、1、.....、}^{\mathrm{T}}\ \hat{}\mathrm{u}_{\tau、N}^{\mathrm{T}}]^{\mathrm{T}}$。11：如果 $\|\hat{}\mathrm{u}_\tau - \hat{}\mathrm{u}_{\tau-1}\|_F \leqslant \varepsilon\|\hat{}\mathrm{u}_{\tau-1}\|_F$和 $\|\hat{}\mathrm{c}_\tau - \hat{}\mathrm{c}_{\tau-1}\|_F \leqslant \varepsilon\|\hat{}\mathrm{c}_{\tau-1}\|_F$，则退出循环。12：返回 $\mathrm{Z} = \hat{}\mathrm{u}_\tau\hat{}\mathrm{c}_\tau$。

算法 1：ALS：交替最小二乘

3.2 聚类

在第二阶段，我们通过对 $\mathbf{Z}$ 应用 $\hat{}\mathrm{K}$ 均值聚类来估计类型矩阵 $\boldsymbol{\kappa}$ （等价地，即分配矩阵 $\boldsymbol{\mathcal{K}}$）。该聚类过程还会得到系数矩阵 $\mathbf{c}$ 的归一化行。为说明核心聚类过程，我们首先考虑 Q的真实值已知的情况。随后，对于 Q 未知的更一般情形，我们将介绍一种迭代搜索算法。

3.2.1 已知类型数量的 K均值聚类

如第 2.3 节所述，矩阵 $\mathbf{Z} \in \mathbb{R}^{N(N-1)\times K}$ 预计具有 Q 个在缩放意义下互不相同的行。$\hat{}$这一结构促使我们对行归一化矩阵 $\mathbf{Z}$进行聚类，以恢复分配矩阵。

回顾一下，$\mathbf{Z}_i \in \mathbb{R}^{(N-1)\times K}$ 是 $\mathbf{Z}$ 的第 i 个块。令 $\mathbf{Z}_{i,j}$ 表示$\mathbf{Z}_i$ 的第 j 行，并将行范数定义为

$$z_{i,j} = \|\mathbf{Z}_{i,j}\|_{\mathbb{R}^{1\times K}}. \tag{3.3}$$

此外，设

$$\overline{\mathbf{Z}}_{i,j} = \frac{1}{z_{i,j}}\mathbf{Z}_{i,j} \in \mathbb{R}^{1\times K}, \quad \overline{\mathbf{Z}}_i = \begin{bmatrix} \overline{\mathbf{Z}}_{i,1} \\ \vdots \\ \overline{\mathbf{Z}}_{i,N} \end{bmatrix} \in \mathbb{R}^{(N-1)\times K}, \quad \overline{\mathbf{Z}} = \begin{bmatrix} \overline{\mathbf{Z}}_1 \\ \vdots \\ \overline{\mathbf{Z}}_N \end{bmatrix} \in \mathbb{R}^{N(N-1)\times K}, \tag{3.4}$$

其中，$(i,\ j)$ 遍历所有可能的配对，约定省略 $i=j$。我们类似地定义 $\widehat{z}_{i,\ j}$，$\mathbf{Z}_{i,\ j}$、$\widehat{\mathbf{Z}}_i$ 以及用于估计$\mathbf{Z}$的$\overline{\mathbf{Z}}$。归一化后，$\mathbf{Z}$ 理想情况下应包含 Q 个在缩放意义下互不相同的行模式。然而，噪声和估计误差可能会扭曲这一结构。特别是，如果 $a_{ij}=0$，那么$\widehat{\mathbf{Z}}$中的对应行应恰好为零。然而，由于噪声，$\mathbf{Z}$的这些行中可能会出现较小的非零值，而归一化可能会放大这些值，造成显著失真并导致聚类结果不可靠。因此，在聚类之前将零行与非零行分开至关重要。

我们首先识别与图中实际为零的条目相对应的索引对 $(i,\ j)$ 并将它们收集到集合中

$$I_0=\{(i,j):\widehat{z}_{i,j}<z_0\},\tag{3.5}$$

其中 z_0 表示预设阈值。当存在式（ z_{min} 2.12）中定义的先验信息时，z_0 可根据噪声水平和数据规模进行最优选择；详情见第4.3节。若没有此类先验信息，则 z_0 可作为可调参数，并通过经验选取。

接着，我们使用余弦距离 [1, 30] 对$\mathbf{Z}$ 中剩余的行进行 K 均值聚类 $\overline{d}(\mathbf{x},\ \mathbf{y}):=1-\frac{\langle\mathbf{x},\ \mathbf{y}\rangle}{\|\mathbf{x}\|\|\mathbf{y}\|}$，其中任意 $\mathbf{x}$、$\mathbf{y}\in\mathbb{R}^K$均满足此关系。在这种角度聚类设定中，簇中心$\mathbf{v}$ 通过对簇内的点 $\{\mathbf{x}_i\}_{i=1}^n$ 取归一化平均值来确定：

$$\mathbf{v}=\mathbf{u}/\|\mathbf{u}\|,\quad \mathbf{u}=\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i.\tag{3.6}$$

该过程得到 Q 个簇及其对应的索引集 I_1、……、I_Q，

$$I_q=\big\{(i,j):\overline{\widehat{\mathbf{Z}}}_{i,j}\text{ belongs to the }q\text{-th cluster}\big\},$$

以及簇中心$\mathbf{v}_1$、……、$\mathbf{v}_Q\in\mathbb{R}^{1\times K}$，它们表示估计得到的归一化核系数。若 $(i,\ j)\in I_q$，则我们将第 i个粒子与第 j个粒子之间的相互作用识别为第 q种核类型，并设定

$$\widehat{\boldsymbol{\mathcal{K}}}_{i,j}=\mathbf{e}_q,\quad \widehat{\kappa}_{ij}=q,\tag{3.7}$$

对于 $q=0,\ 1,\ \ldots\ldots,\ Q$，其中 $\boldsymbol{\mathcal{K}}_{i,\ j}$表示$\boldsymbol{\mathcal{K}}_i$的第 j 行，$\mathbf{e}_q$是$\mathbb{R}^Q$中的第 q 个标准基行向量，而$\mathbf{e}_0$ 是零向量。在 $\widehat{\boldsymbol{\mathcal{K}}}$ 中包含$\mathbf{e}_0$ 反映了类型 κ_{ij} 在$a_{ij}=0$时的不可辨识性。

注 3.2（簇排序）为了解决聚类中的标签歧义，我们按照簇在索引集 $(1、\ 2)、\ (1、\ 3)、\ \ldots\ldots、\ (N、\ N-1)$ 上按字典序排列时的首次出现顺序对簇进行排序。

注 3.3（活跃核与非活跃核）即使已知相互作用核的真实数量 Q，观测到的动力学中也不一定所有核都是活跃的。如果图矩阵$\mathbf{a}$ 包含零元素，即 $a_{ij}=0$ 对于所有满足 (i,j) 的$\kappa_{ij}=q$ 第 q个核而言，因而该核处于非活跃状态。因此，最好根据数据自适应地确定活跃核的数量。相比之下，在注2.3中所述的完全网络假设下，所有相互作用均存在，且权重矩阵$\mathbf{a}$ 不含零元素。在这种情况下，动力学中使用的核的数量与真实值 Q一致，因此可以直接应用算法2。

~1: 过程 球面 K 均值, Q 已知（估计 Z, z_0, Q）2: 对 $\widehat{\mathbf{Z}} \in \mathbb{R}^{N(N-1)\times K}$ 进行行归一化，得到 $\widehat{\mathbf{Z}}$ 和行范数 $\{\widehat{z}_{i,j}\}$。3: 构造 $I_0 = \{(i,j), \widehat{z}_{i,j} < z_0\}$。~4: 使用 K 均值，根据 $z_{i,j} \geqslant \delta z_0$ 将 Z 的各行划分为 Q 个簇，采用余弦距离，得到索引集 $\{I_q\}_{q=1}^{Q}$ 和簇中心 $\{\mathbf{v}_q\}_{q=1}^{Q}$。5: 通过 $\widehat{\mathbf{K}}_{i,j} = \mathbf{e}^q$ 以及 $\widehat{\kappa}_{ij} = q$ ，若 $(i,j) \in I_q$ 对~$q = 0$（其中 1, ..., Q），构造 $\widehat{\mathbf{K}} \in \{0,1\}^{N(N-1)\times Q}$ 和 $\widehat{\boldsymbol{\kappa}} \in [Q]^{N\times N}$ 。6: 返回 $\widehat{\mathbf{K}}$ $\widehat{\boldsymbol{\kappa}}, \{I_q\}_{q=0}^{Q}, \{\widehat{\mathbf{v}}_q\}_{q=1}^{Q}$

算法 2: 已知 Q 时对**Z**进行聚类

3.2.2 未知类型数量时的 K 均值聚类

当核类型的真实数量未知时，必须从数据中推断。令 Q 表示当前估计的簇数，其对应的聚类结果为 $\{I_q, \widehat{\mathbf{v}}_q\}_{q=1}^{Q}$. 我们引入经验最大簇内角，定义如下:

$$\Theta = \max_{1\leqslant q\leqslant \widehat{Q}} \max_{(i,j),(i',j')\in I_q} \arccos\langle \overline{\widehat{\mathbf{Z}}}_{i,j}, \overline{\widehat{\mathbf{Z}}}_{i',j'}\rangle, \tag{3.8}$$

并将其与预设阈值 Θ_0进行比较。若 $\Theta > \Theta_0$，则表明至少有一个簇包含角度离散度较大的元素，这意味着需要增加簇的数量。此外，我们根据当前簇中心~$\{\widehat{\mathbf{v}}_q\}_{q=1}^{Q}$定义经验最小簇间角:

$$\theta := \min_{1\leqslant p<q\leqslant \widehat{Q}} \arccos \frac{\langle \widehat{\mathbf{v}}_p, \widehat{\mathbf{v}}_q\rangle}{\|\widehat{\mathbf{v}}_p\|\|\widehat{\mathbf{v}}_q\|}, \tag{3.9}$$

并将其与阈值 θ_0进行比较。若 $\theta < \theta_0$，则说明簇中心之间的间隔不足，意味着簇的数量被高估，因此需要减少 Q。

该过程能够在迭代过程中自适应地调整簇的数量。完整方法总结于算法 3中。与算法2相比，当理想阈值 θ_0 和 Θ_0 可用时，这种自适应方法通常更加稳健。先验信息 θ_{min}（定义见（2.11））可用于指导推荐阈值 θ_0 和 Θ_0的选择。这些阈值的适当取值还取决于噪声水平和可用数据量；第 4.3节提供了详细分析。若无先验知识，则可将 θ_0 和 Θ_0 视为可调参数，并通过经验选取。

注 3.4 *K*均值聚类通过最小化簇内方差将数据划分为 K 个簇，通常采用 *Lloyd* 算法 [30], ，在该算法中，交替执行将数据点分配给最近的簇中心，以及将簇中心更新为其所分配数据点的均值。*Lloyd* 算法虽然有效，但对初始化十分敏感，且可能收敛到次优解。

为解决这一问题，*K* 均值$^{++}$ [1] 采用概率播种策略，选择分离度较高的初始中心，从而显著提升聚类性能。受此理念启发，我们在算法中用于确定簇数量的迭代过程 Q 在算法 *3* 中发挥了类似作用: 利用先验信息自适应地增加簇的数量，从而提高分离度和鲁棒性。

K 均值的复杂度取决于每一步的聚类情况以及迭代次数。每次迭代中，需要进行 $\mathcal{O}(N^2KQ)$ 步比较，将 $N(N-1)$行与 Q 个簇中心进行比较。

ㅅ1：过程 球面 K 均值， Q Unknown(Z, z_0, θ_0, Θ_0, Q_0, $n_{\max}$) 2：对 $\hat{}$Z $\in \mathbb{R}^{N(N-1)\times K}$ 进行行归一化，得到 $\hat{}$Z 和行范数 $\{\hat{z}_{i,j}\}$。3：构造 $I_0 = \{(i,j), \hat{z}_{i,j} < \delta z_0\}$ㅅ4：初始化 $Q = Q_0$， success = false 5：对于 $\tau = 0, \ldots, n_{\max}$，执行ㅅ ㅅ6：在 Z 的各行上运行 K 均值，使用余弦距离和 $\hat{z}_{i,j} \geqslant z_0$ ，聚为 Q 个簇，得到索引集ㅅ ㅅ$\{I_q\}_{q=1}^{Q}$ 和簇中心 $\{\hat{}\mathrm{v}_q\}_{q=1}^{Q}$。7：计算最小簇间角度 θ 和最大簇内角 Θ8：如果 $\theta > \theta_0$ 且 $\Theta < \Theta_{\max}$ ，则 9：success = true，跳出循环 10：否则如果 $\theta < \theta_0$ ，则ㅅ ㅅ11：$Q \leftarrow Q - 1$12：否则如果 $\Theta > \Theta_0$ ，则ㅅ ㅅ13：$Q \leftarrow Q + 1$14：如果 success =false，则 15：返回分类失败 16：否则 17：根据聚类结果构造 $\hat{}$K $\in \{0,1\}^{N(N-1)\times\widehat{Q}}$ 和 $\widehat{\boldsymbol{\kappa}} \in [\widehat{Q}]^{N\times N}$，方法是设置 $\hat{}\mathrm{K}_{i,j} = \mathrm{e}_{\widehat{q}}$，并且对于 $q = 0, 1, \ldots, Q$，如果 $(i,j) \in I_q$，则设置 $\widehat{\boldsymbol{\kappa}}_{ij} = q$。18：返回 $\hat{}$K $\widehat{\boldsymbol{\kappa}}, \widehat{Q}, \{I_q\}_{q=1}^{Q}, \{\hat{}\mathrm{v}_q\}_{q=1}^{Q}\hat{}$ $\hat{}$算法 3：对未知 Q的 Z 进行聚类

K 均值迭代次数在最坏情况下可能呈指数级增大。然而，使用 K 均值$^{++}$初始化可以改善收敛并减少迭代次数。在实践中，迭代次数通常较少。

3.3 分解与后处理

$\hat{}$3.3.1 将 Z 分解为 $\hat{}$a 与 $\hat{}$c。

$\hat{}$给定估计的分配矩阵 $\boldsymbol{\mathcal{K}}$ （因此类型矩阵 $\widehat{\boldsymbol{\kappa}}$）和聚类中心 $\{\hat{}\mathbf{v}_q\}$，我们使用$\hat{}$简单线性回归（2.15）将估计矩阵 $\mathbf{Z}$ 分解为图矩阵 $\mathbf{a}$ 和核系数$\mathbf{c}$ ，计算成本可忽略不计。

应当强调，并不能保证 $\mathbf{A}$ （由（2.16）定义）可逆，因此正则化是必要的。事实上，$\mathbf{A}$ 的秩和谱承载着关于底层系统的重要结构信息，尤其是哪些核处于活跃状态，如注记 3.3中所讨论。参见第4.4节，以详细讨论 $\mathbf{A}$ 的可逆性及相关误差分析。

$\hat{}$将 $\mathbf{Z}$ 和 $\boldsymbol{\mathcal{K}}$ 替换为其估计值 $\mathbf{Z}$ 和 $\boldsymbol{\mathcal{K}}$，我们求解方程（2.15），得到 $\{\hat{v}_q\}$ 并恢复 $\hat{}\mathbf{c}$ （见2.14）。随后逐元素得到图矩阵：

$$\widehat{a}_{ij} = \frac{\widehat{z}_{i,j}}{\widehat{v}_{\kappa_{ij}}}. \tag{3.10}$$

整个恢复过程总结于算法 4。求解小型线性系统的计算成本与前两个阶段的成本相比可以忽略不计。

$\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$1: 过程 后处理分解(Z, K, $\hat{\kappa}$, $\{\widehat{\ }\mathrm{v}_q\}$) 2: 求解 $\hat{\boldsymbol{\eta}}$ 由 $\widehat{\ }\mathrm{A}\hat{\boldsymbol{\eta}} = \mathbf{1}_N$ 以及 $\widehat{\ }\mathrm{A}_i = [\hat{z}_i^2, 1, \ldots, \hat{z}_i^2, N]$ $\widehat{\ }\mathrm{K}_i \in \mathbb{R}^{1\times Q}$ 和 $\widehat{\ }\mathrm{A} = [\widehat{\ }\mathrm{A}_i]_{1\leqslant i\leqslant N} \in \mathbb{R}^{N\times Q}$。 3: 令 $\hat{v}_q = \sqrt{1/\hat{\eta}_q}$ 并令 $\widehat{\ }\mathrm{c} = [\hat{v}_1 \widehat{\ }\mathrm{v}_1^{\mathrm{T}}, \cdots, \hat{v}_Q \widehat{\ }\mathrm{v}_Q^{\mathrm{T}}]$。 4: 构造 $\widehat{\ }\mathrm{a} = [\hat{a}_{ij}]$ 通过 $\hat{a}_{ij} = \hat{z}_i, j/\hat{v}_{\kappa_{ij}}$。 5: 构造 $\hat{\Phi}_q(x) = \sum_{k=1}^{K} \widehat{\ }\mathrm{c}_k^{(q)} \psi_k(x)$ 对于 $q = 1, \ldots, \hat{Q}$。$\widehat{\ }$6: 返回 $\widehat{\ }\mathrm{a}$ $\widehat{\ }\mathrm{c}$ $\{\Phi_q\}$。

$\widehat{\ }$算法 4：使用$\mathbf{Z}$ 、 $\widehat{\ }\mathbf{a}$ 和 $\widehat{\ }\mathbf{c}$ 以及聚类中心进行因子分解。

1：过程 后处理交替最小二乘($\{\mathrm{X}_{t_l}^m\}_{l=1}^{L,M}$, m=1, $\{\psi_k\}_{k=1}^K$, $\widehat{\ }\mathrm{K}$, ε, $n_{maxiter}$) 2：根据基函数计算导数 $\{\Delta \mathrm{X}_{t_l}^m\}$ 和感知张量 $\{\mathrm{B}(\mathrm{X}_{t_l}^m)\}_{l,m}$。 3：随机选取初始值 $\widehat{\ }\mathrm{a}_0 \in \mathbb{R}^{N\times(N-1)}$，或将算法 4 的结果作为 $\widehat{\ }\mathrm{a}_0$。 4：对于 $\tau = 1, \ldots, n_{maxiter}$，执行$\widehat{\ }$5：使用 $\widehat{\ }\mathrm{a} = \widehat{\ }\mathrm{a}_{\tau-1}$, 通过最小二乘法求解 (3.11) 中的 $\widehat{\ }\mathrm{c}_\tau = \mathrm{argmin}_{\mathbf{c}}\, \mathcal{L}(\widehat{\ }\mathrm{a}, \mathrm{K}, \mathrm{c})$。 6：对于 $i = 1, \ldots, N$, 执行$\widehat{\ }$7：使用 $\widehat{\ }\mathrm{c} = \widehat{\ }\mathrm{c}_\tau$，通过最小二乘法求解 (3.12) 中的 $\widehat{\ }\mathrm{a}_{\tau,i} = \mathrm{argmin}_{\mathbf{a}_i}\, \mathcal{L}(\mathrm{a}_i, \mathrm{K}\ \widehat{\ }\mathrm{c})$。 8：对 a_τ 进行行归一化。 9：构造矩阵 $\widehat{\ }\mathrm{a}_\tau = [\widehat{\ }\mathrm{a}_\tau^{\mathrm{T}}, 1, \ldots, \widehat{\ }\mathrm{a}_{\tau,N}^{\mathrm{T}}]^{\mathrm{T}}$。 10：如果 $\|\widehat{\ }\mathrm{a}_\tau - \widehat{\ }\mathrm{a}_{\tau-1}\|_F \leqslant \varepsilon \|\widehat{\ }\mathrm{a}_{\tau-1}\|_F$ 且 $\|\widehat{\ }\mathrm{c}_\tau - \widehat{\ }\mathrm{c}_{\tau-1}\|_F \leqslant \varepsilon \|\widehat{\ }\mathrm{c}_{\tau-1}\|_F$，则退出循环。 11：为 $q = 1, \ldots, \hat{Q}$ 构造 $\hat{\Phi}_q(x) = \sum_{k=1}^{K} \widehat{\ }\mathrm{c}_k^{(q)} \psi_k(x)$。$\widehat{\ }$12：返回 $\widehat{\ }\mathrm{a}$ $\widehat{\ }\mathrm{c}$ $\{\Phi_q\}$。

算法5：通过交替最小二乘进行后处理

3.3.2 通过交替最小二乘进行后处理

$\widehat{\ }$为进一步提高估计精度，我们固定估计得到的分配矩阵$\boldsymbol{\mathcal{K}}$ ，并执行额外的交替最小二乘步骤，以细化$\mathbf{a}$ 和$\mathbf{c}$的估计。具体而言，在给定固定的 $\widehat{\ }\mathbf{a}$ 后，我们通过以下方式估计$\mathbf{c}$ ：

$$\min_{\mathbf{c}\in\mathbb{R}^{K\times Q}} \mathcal{L}(\hat{\mathbf{a}}, \hat{\boldsymbol{\mathcal{K}}}, \mathbf{c}) = \frac{1}{LMN} \sum_{l,m,i=1}^{L,M,N} \left\| \dot{X}_{t_l,m}^i - \langle \mathbf{B}(\mathbf{X}_{t_l}^m)_i, \mathrm{Diag}(\hat{\mathbf{a}}) \hat{\boldsymbol{\mathcal{K}}}_i \mathbf{c}^{\mathrm{T}} \rangle_F \right\|_{\mathbb{R}^d}^2 . \tag{3.11}$$

$\widehat{\ }$然后，在给定$\boldsymbol{\mathcal{K}}$ $\widehat{\ }\mathbf{c}$的情况下，我们通过以下方式估计图$\mathbf{a}_i$ 的每一行:

$$\min_{\mathbf{a}_i\in[0,1]^{1\times(N-1)}} \mathcal{L}(\mathbf{a}_i, \hat{\mathcal{K}}, \hat{\mathbf{c}}) = \frac{1}{LM} \sum_{l,m=1}^{L,M} \left\| \dot{X}_{t_l,m}^i - \langle \mathbf{B}(\mathbf{X}_{t_l}^m)_i, \mathrm{Diag}(\mathbf{a}_i) \hat{\boldsymbol{\mathcal{K}}}_i \hat{\mathbf{c}}^{\mathrm{T}} \rangle_F \right\|_{\mathbb{R}^d}^2 . \tag{3.12}$$

这些更新迭代进行，直至收敛，其中 $\mathbf{a}$ 需满足行归一化条件。该后处理步骤提高了对环境噪声或模型扰动的鲁棒性；其性能还取决于感知算子的RIP性质，不过具体细节在此省略。完整过程总结于算法5中。

3.4 联系与扩展

3.4.1 完全网络

对于完全网络，可以采用类似的步骤来恢复分配矩阵$\boldsymbol{\mathcal{K}}$$\widehat{\ }$和核系数$\mathbf{c}$。我们首先应用算法1 ，得到估计值$\mathbf{Z}$。在第二阶段，如果已知相互作用类型的数量 Q ，我们直接对未归一化的行 K应用 K 均值聚类

$\widehat{\mathbf{Z}}$的，其中得到的簇中心对应于估计的核系数。如果 Q 未知，$\widehat{\mathbf{Z}}$我们通过将分离度量 Q 和 θ 与各自的阈值进行比较，迭代更新 Θ 。随后通过（分配矩阵$\boldsymbol{\mathcal{K}}$ ）3.7构造。

与一般图设定不同，在一般图设定中，$\widehat{\mathbf{Z}}$ 沿着 Q 个不同的方向，而在完全网络设定中，它们恰好占据 Q 个不同的位置。因此，无需区分零行与非零行，即可直接基于欧几里得距离进行聚类。实践中，这种结构使聚类结果具有更强的鲁棒性和更高的准确性。该算法总结于附录A。

更广泛地说，我们的框架不仅适用于完全未知的设定，也适用于具有部分信息的情形。给定交互图，上述完全网络流程说明了如何绕过图恢复，专注于聚类和核估计。若已知类型分配，则可直接应用后处理交替最小二乘细化来恢复图和核系数。最后，若已知核，则问题可简化为相对容易的图和类型分配恢复。这些特殊情形凸显了我们方法的灵活性，以及它在广泛学习场景中的适用性。

3.4.2 整数约束

$\widehat{\mathbf{a}}$算法5 的一个自然扩展是考虑三向交替最小二乘方法。当$\boldsymbol{\mathcal{K}}$ 未知时，我们可以增加一个步骤，根据给定的$\boldsymbol{\mathcal{K}}$ ，从 $\widehat{\mathbf{a}}$ 和 $\widehat{\mathbf{c}}$中估计$\boldsymbol{\mathcal{K}}$ $\in \mathcal{B}^N$ ，并满足式（2.3）所定义的约束。然而，这是一个非凸整数约束，使得直接优化十分困难。为实现高效的最小二乘估计，可以对式（2.3）进行凸松弛。但在实践中，这往往会导致较差的局部极小值和不具备所需结构的次优解。鉴于这一挑战，我们省略对$\boldsymbol{\mathcal{K}}$ 的约束，转而将问题分为两个阶段，随后进行后处理。

4 理论保证

4.1 相互作用核的探索度量与空间

为函数空间奠定基础，我们首先按照 [33, 31, 25]引入探索度量和假设空间。

我们定义一个函数空间 $L^2(\rho_L)$ ，用于学习相互作用核 $\boldsymbol{\Phi}$，其中 ρ_L 是一个概率测度，用于量化对相互作用核的数据探索。记成对差分

$$\mathbf{r}_{ij}(t_l) := X^j_{t_l} - X^i_{t_l} \quad \text{and} \quad \mathbf{r}^m_{ij}(t_l) := X^{j,m}_{t_l} - X^{i,m}_{t_l}, \tag{4.1}$$

w它们是相互作用核的变量。因此，我们定义 ρ_L 如下。

定义 4.1（相互作用核的探索度量）给定在离散时间 $\{t_l\}_{l=0}^{L-1}$ 下对 M 轨迹的观测值，我们在 $\mathbb{R}^d$ 上定义经验测度及其大样本极限如下：

$$\rho_{L,M}(d\mathbf{r}) := \frac{1}{LMN(N-1)} \sum_{l=0}^{L-1} \sum_{m=1}^{M} \sum_{1\leqslant i\neq j\leqslant N} \delta_{\mathbf{r}^m_{ij}(t_l)}(d\mathbf{r}), \tag{4.2}$$

$$\rho_L(d\mathbf{r}) := \frac{1}{LN(N-1)} \sum_{l=0}^{L-1} \sum_{1\leqslant i\neq j\leqslant N} \mathbb{E}[\delta_{\mathbf{r}_{ij}(t_l)}(d\mathbf{r})], \tag{4.3}$$

其中 $\delta_{\mathbf{r}}$ 表示在$\mathbf{r}$处的狄拉克测度，且符号约定 $\sum_{1\leqslant i\neq j\leqslant N}$ 表示$\sum_{i=1}^{N}\sum_{j=1,\ j\neq i}^{N}$。

经验测度依赖于样本轨迹，但 ρ_L 是大样本极限，由随机过程$\mathbf{X}_{t_0:\ t_{L-1}}$的分布唯一确定，因而与数据无关。

我们con 考虑一个假设空间，其基函数满足以下条件 .

假设 4.2（一致有界基函数）假设空间的基函数$\mathcal{H}:=\text{span}\{\psi_k:\mathbb{R}^d\to\mathbb{R}^d\}_{k=1}^K$ 在 $L^2(\rho_L)$ 中正交归一。此外，它们一致有界，即上确界$_{k\in[K]}\|\psi_k\|_\infty\leqslant L_{\mathcal{H}}$。

假设空间的规模 K 会随维度呈指数级增长 d，这是非参数回归中的维数灾难所致。为避免这一问题，常见做法是在假设空间中考虑径向相互作用核 $\mathcal{H}:=\text{span}\{\psi_k:\ \mathbb{R}^+\to\mathbb{R}\}_{k=1}^K$；例如参见 [32, 33]。我们也认为，让相互作用仅依赖于少量变量（这些变量可能未知）是很自然的做法（例如参见 [16]）。

4.2 矩阵感知阶段的误差界

本节基于矩阵感知理论，为第一阶段提供理论保证 [19,38, 44]。为便于阐述，我们聚焦于 $d=1$ 和 $L=1$均满足的简化情形。我们分析固定 $\mathbf{Z}_i$ 时单个向量 $i\in[N]$的估计，并注意到式（2.8）中的整体矩阵感知问题可分解为 N 个独立子问题，每个子问题对应一个 $\mathbf{Z}_i$，$i\in[N]$。

在 $d=1$ 和 $L=1$的情况下，**(2.4)** 中的感知算子 B_i 退化为感知矩阵 $\mathbf{B}_i(\mathbf{X}_0^m)\in\mathbb{R}^{(N-1)\times K}$，该矩阵编码了以第 N 个粒子为中心的相对相互作用结构。观测量是第 i 个粒子在时间 $t=0$ 的速度，并受到加性高斯噪声的污染，即:

$$b_m=\dot{X}_0^{i,m}+\epsilon^{i,m}=\langle\mathbf{B}_i(\mathbf{X}_0^m),\mathbf{Z}_i^\star\rangle_F,\quad \epsilon^{i,m}\sim\mathcal{N}(0,\sigma_\epsilon^2). \tag{4.4}$$

当系统受到随机驱动，且使用有限差分近似来估计导数时，这一观测模型自然产生。在此情况下，$\sigma_\epsilon=\sigma\sqrt{\Delta t}$ 其中 σ 是 (1.1) 的扩散系数，而 $\sqrt{\Delta t}$ 则来自 Euler–Maruyama 离散化。特别地，损失函数$\mathcal{L}_i(\mathbf{Z}_i)$ 在 (2.8) 中可化为

$$\mathcal{L}_i(\mathbf{Z}_i)=\frac{1}{M}\sum_{m=1}^M\|b_m-\langle\mathbf{B}_i(\mathbf{X}_0^m),\mathbf{Z}_i\rangle_F\|^2. \tag{4.5}$$

我们对感知矩阵作出如下受限等距性质（RIP）条件的假设。

假设 4.3（RIP 条件）固定 $0<\delta<1$。对于任意 $1\leqslant Q\leqslant((N-1)\wedge K)$，存在仅依赖于 δ 的正常数 c_0、c_1，其中 $c_0=O(1/\delta^2)$,，使得当 $M\geqslant c_0Q(N+K)\log(NK)$ 时，以至少 $1-exp(-c_1M)$ 的概率，满足下式的最小常数为 δ_Q

$$(1-\delta_Q)\|Z\|_F^2\leqslant\frac{1}{CM}\sum_{m=1}^M\langle\mathbf{B}_i(\mathbf{X}_0^m),Z\rangle_F^2\leqslant(1+\delta_Q)\|Z\|_F^2 \tag{4.6}$$

对于所有秩为 Q 的矩阵 $Z\in\mathbb{R}^{(N-1)\times K}$ 和 $i\in[N]$，均小于 δ。

对于近等距随机矩阵，上述 RIP 条件成立。这类矩阵包括具有独立同分布的高斯条目的矩阵 $\mathcal{N}(0,\frac{1}{M})$，如 [38, 定理 4.2] 所示。对于一般的相互作用粒子系统，由于感知矩阵的条目共同依赖于同一个参考粒子和基函数，因此呈现出非平凡的相关性。虽然 [29] 中的某些技术或许能提供部分见解，但在这种结构依赖下验证 RIP 类型条件仍是一项重大挑战。此外，将分析扩展到包含更长时间步和更高维度的情形，会使问题更加复杂。因此，我们提出上述条件，作为理解这一结构化感知设置样本复杂度的一个合理基础，同时也承认，对该条件进行严格验证仍是一个尚未解决的问题。

该 RIP 条件为无噪声数据和含噪数据均提供恢复保证。下一个命题说明了$\mathbf{Z}_i$ 从无噪声数据中的可辨识性，以及含噪数据下的估计误差界。

命题 4.4 设 $d = 1$, $L = 1$, 并假设 4.3 成立。那么，对于任意 $1 \leqslant Q \leqslant (N-1) \wedge K$，存在一个常数 $C > 0$ 使得以下结论以高概率对每个$i \in [N]$成立。

- **(无噪声情形)**：若 $\sigma_\epsilon = 0$ 且 $M \geqslant CQ(N+K)\log(NK)$，则 $\mathbf{Z}_i^*$是目标函数在 (4.5)中的唯一极小值。

- **(有噪声情形)**：若 $\sigma_\epsilon > 0$ 且 $M \geqslant CQ(N+K)\,log(NK)$，则任意局部极小值 $\hat{}Z_i = \hat{}u_i\hat{}cT\,off(u_i, c) := \frac{1}{M}\sum_{m=1}^{M} \|b_m - \langle B_i(X_0^m), u_i cT\rangle_F\|_{\mathbb{R}^d}^2 + \|uT_i\,u_i - ccT\|^2$ 满足 **(4.4)** 中的 b_m，并且

$$\|\mathbf{Z}_i^* - \widehat{\mathbf{Z}}_i\|_F \lesssim \sigma_\epsilon \sqrt{\frac{(N+K)Q\log M}{M}} \tag{4.7}$$

对于每个 $i \in [N]$ 均成立。这里，$A \lesssim B$ 表示对于某个通用常数 $C_0 > 0$，有 $A \leqslant C_0 B$。

上述结果可直接由 [38, 定理 3。2] 以及 [19, 定理 32], 推出，我们将在第 B.1节给出该定理。需要注意的是，观测时间步 L 和粒子维度 d 并未出现在给出的界中，因为这里的分析针对的是简化情形 $d = L = 1$。

由（2.8）中的表述可知，整体矩阵感知问题可分解为N个子问题，每个子问题对应于对 $\mathbf{Z}_i$的估计，其中 $i = 1, \ldots\ldots, N$。因此，相同的样本复杂度与误差尺度分别适用于每个 $\mathbf{Z}_i$ 本身。对所有 i求和，我们得到如下全局误差界：

$$\|\mathbf{Z}^* - \widehat{\mathbf{Z}}\|_F \lesssim \sigma_\epsilon \sqrt{\frac{N(N+K)Q\log M}{M}}. \tag{4.8}$$

注 4.5 注意，我们的算法并非对矩阵 $\mathbf{Z}_i$ 进行独立估计：它们共享一个共同的系数矩阵 $\mathbf{c}$。然而，图 5和图6中的数值结果表明，估计量的全局误差量级分别为 $\mathcal{O}(M^{-1/2})$和 $\mathcal{O}(N)$，与上述界一致。

样本复杂度为 RIP 条件及前述命题提供了另一种视角。第 5.4 节中的数值实验表明，恢复类型矩阵 κ所需的样本量关于 $N+K$近似呈线性缩放。从启发式角度看，缩放 $M = O\big(Q(N+K)\log(NK)\big)$ 是自然的：一个秩为Q的矩阵具有 $Q(N+K-Q)$ 个自由度，因此 $Q(N+K)$ 这一因子在至多相差对数项的意义下与该维数计数相符。

4.3 可分离性

$\hat{}$我们已经证明，在满足 RIP 条件时，估计量$\mathbf{Z}$ 在弗罗贝尼乌斯范数下的误差很小。现在，我们将探讨这一估计误差如何影响聚类步骤；该步骤对以 $(2,\infty)$-范数衡量的误差十分敏感。对于任意矩阵 $A \in \mathbb{R}^{m\times n}$，我们将其 $(2,\infty)$-范数定义为最大行范数：

$$\|A\|_{2,\infty} := \max_{1\leqslant i\leqslant m} \|A_i\|_{\mathbb{R}^n} = \max_{1\leqslant i\leqslant m} \left(\sum_{j=1}^{n} A_{ij}^2\right)^{1/2}, \tag{4.9}$$

其中，A_i 表示 A 的第 i 行。显然，$(2,\infty)$-范数与弗罗贝尼乌斯范数等价：

$$\|A\|_{2,\infty} \leqslant \|A\|_F \leqslant \sqrt{m}\,\|A\|_{2,\infty}. \tag{4.10}$$

$\hat{}$回顾一下，对于每个 $i \in [N]$，(4.7) 都成立。然后，将 (4.10) 应用于 $\|\mathbf{Z}_i^* - \mathbf{Z}_i\|_{2,\infty}$，可得

$$\|\mathbf{Z}^* - \widehat{\mathbf{Z}}\|_{2,\infty} = \max_{i=1,\ldots,N} \|\mathbf{Z}_i^* - \widehat{\mathbf{Z}}_i\|_{2,\infty} \lesssim \sigma_\epsilon \sqrt{\frac{(N+K)Q\log M}{M}}, \tag{4.11}$$

其增长阶为 $\sqrt{N}$，优于 N 中弗罗贝尼乌斯范数 $\|\mathbf{Z}^*-\widehat{\mathbf{Z}}\|_F$在 (4.8) 中的线性标度。图 6 中的数值实验表明，当图权重矩阵稀疏时，阶数 $\|\mathbf{Z}^*-\widehat{\mathbf{Z}}\|_{2,\infty}$可能远低于 $\sqrt{N}$。

我们首先证明，如果 $\|\mathbf{Z}^*-\mathbf{Z}\|_{2,\infty}$ 足够小，那么通过对（$\mathbf{Z}$ 进行阈值化，可以正确识别其零行与非零行（3.5）。

命题 4.6（支撑恢复）回忆 $z_{min}=\min_{1\leqslant i,\ j\leqslant N,\ a_{ij}\neq 0}\left\|\mathbf{Z}^*_{i,\ j}\right\|$ 见（2.12）。假设

$$\|\mathbf{Z}^*-\widehat{\mathbf{Z}}\|_{2,\infty}<\varepsilon<\frac{z_{min}}{2}. \tag{4.12}$$

那么，对于每个 $z_0\in(\varepsilon、\ z_{min}-\varepsilon)$，都有 $\|\mathbf{Z}_{i_1,\ j_1}\|<z_0<\|\mathbf{Z}_{i_2,\ j_2}\|$ 对于满足 $a^*_{i_1j_1}=0$、$a^*_{i_2j_2}\neq 0$的任意索引对 $(i_1、\ j_1)$、$(i_2、\ j_2)$。

证明。 由于 $a^*_{i_1,\ j_1}=0$，真实矩阵 $\mathbf{Z}^*$ 满足 $\|\mathbf{Z}^*_{i_1,\ j_1}\|=0$，因此

$$\|\widehat{\mathbf{Z}}_{i_1,j_1}\|\leqslant\|\mathbf{Z}^*_{i_1,j_1}\|+\|\mathbf{Z}^*_{i_1,j_1}-\widehat{\mathbf{Z}}_{i_1,j_1}\|<\varepsilon.$$

其中第二个不等式由（4.12）以及 $(2,\infty)$-范数的定义得到。此外，由于 $a^*_{i_2,j_2}>0$，我们有 $\|\mathbf{Z}^*_{i_2,\ j_2}\|\geqslant z_{min}$ 以及

$$\|\widehat{\mathbf{Z}}_{i_2,j_2}\|\geqslant\|\mathbf{Z}^*_{i_2,j_2}\|-\|\mathbf{Z}^*_{i_2,j_2}-\widehat{\mathbf{Z}}_{i_2,j_2}\|>z_{min}-\varepsilon.$$

因此，在 $\varepsilon<\frac{z_{min}}{2}$ 下，对任意 $z_0\in(\varepsilon,z_{min}-\varepsilon)$，我们有 $\|\mathbf{Z}_{i_1,j_1}\|<z_0<\|\mathbf{Z}_{i_2,j_2}\|$。 ■

实践中，当 z_{min} 不可用时，我们将参数 z_0 调整到区间 $(\varepsilon,\ z_{min}-\varepsilon)$，并取$\varepsilon=O(\sigma_\epsilon\sqrt{(N+K)Q\log M})$，依据命题4.4。

接下来，我们证明，若 ε 足够小，则成功执行的算法3 的结果保证能够提供$\widehat{}$对$\mathbf{Z}$的行的正确分类。正确分类取决于两个准则：（i）来自不同真实簇的元素必须被分配到不同的估计簇；（ii）属于同一真实簇的元素必须被分配到同一估计簇。特别地，这意味着$\widehat{}$估计簇数 Q与真实数目一致。

回顾一下，如果 $\{I_q,\mathbf{v}_q\}_{q=1}^{Q}$ 是算法3的成功输出，则它必须满足

$$\Theta\leqslant\Theta_0,\quad \theta\geqslant\theta_0, \tag{4.13}$$

其中，$\theta:=\min_{1\leqslant p<q\leqslant\hat{Q}}\arccos\frac{\langle\widehat{}_p,\widehat{}_q\rangle}{\|\widehat{}_p\|\|\widehat{}_p\|}$v ,v v v$_q\|$ 和 $\Theta=\max_{1\leqslant q\leqslant\hat{Q}}\max_{(i,j),(i',j')\in I_q}\arccos\langle\widehat{\mathbf{Z}}_{i,j},\widehat{\mathbf{Z}}_{i',j'}\rangle$ 分别定义于（3.8）和（3.9）中。这里，θ_0 和 Θ_0 是预设的分离阈值。直观而言，$\Theta\leqslant\Theta_0$ 控制簇内离散度，而 $\theta\geqslant\theta_0$ 保证了足够的簇间分离度。

命题 4.7（簇可分性）设 $\{I_q,\mathbf{v}_q\}_{q=1}^{Q}$ 为算法3 的聚类结果，且满足-$\widehat{}$足(4.13)。假设 $\|\mathbf{Z}^*-\mathbf{Z}\|_{2,\infty}<\varepsilon$，其中 ε 满足

$$\frac{\pi\varepsilon}{(z_{min}-\varepsilon)}\leqslant\frac{\theta_{min}}{2}, \tag{4.14}$$

其中 θ_{min} 和 z_{min} 定义见 (2.11) 和 (2.12)。那么对于任意

$$\Theta_0\in\left(\frac{\pi\varepsilon}{(z_{min}-\varepsilon)},\theta_{min}-\frac{\pi\varepsilon}{(z_{min}-\varepsilon)}\right),\quad \theta_0\in\left(\Theta_0+\frac{\pi\varepsilon}{(z_{min}-\varepsilon)},\theta_{min}\right), \tag{4.15}$$

对于任意索引对 $(i_2,j_2),(i_3,j_3)$，若满足 $a^*_{i_2,\ j_2}a^*_{i_3,\ j_3}\neq 0$，则下式成立。

(i) 簇间一致性：

$$\kappa^*_{i_2,j_2} \neq \kappa^*_{i_3,j_3} \quad \Longrightarrow \quad (i_2,j_2) \in I_{q_2},\ (i_3,j_3) \in I_{q_3},\ \ for\ \ q_2 \neq q_3. \tag{4.16}$$

（*ii*）簇内一致性：

$$\kappa^*_{i_2,j_2} = \kappa^*_{i_3,j_3} \quad \Longrightarrow \quad (i_2,j_2),(i_3,j_3) \in I_q,\ for\ some\ q = 1,\ldots,\widehat{Q}. \tag{4.17}$$

证明。 我们首先用反证法证明簇间一致性。假设 $\kappa^*_{i_2,\ j_2} \neq \kappa^*_{i_3,\ j_3}$ 和 $(i_2,\ \ j_2)$，$(i_3,\ \ j_3)$被归入同一个簇，其中心为$\mathbf{v}_q$。那么，簇内角满足

$$\arccos\langle \overline{\widehat{\mathbf{Z}}}_{i_2,j_2}, \overline{\widehat{\mathbf{Z}}}_{i_3,j_3} \rangle \leqslant \Theta_0.$$

由引理B.4，可得

$$\|\overline{\mathbf{Z}}^*_{i_2,j_2} - \overline{\widehat{\mathbf{Z}}}_{i_2,j_2}\| \leqslant \frac{1}{z_{min}-\varepsilon}\|\mathbf{Z}^*_{i_2,j_2} - \widehat{\mathbf{Z}}_{i_2,j_2}\| < \frac{\varepsilon}{z_{min}-\varepsilon}, \tag{4.18}$$

因为 $\|\mathbf{Z}^*_{i_2,\ j_2}\| \geqslant z_{min}$ 且 $\|\mathbf{Z}_{i_2,\ j_2}\| \geqslant \|\mathbf{Z}^*_{i_2,\ j_2}\| - \|\mathbf{Z}^*_{i_2,\ j_2} - \mathbf{Z}_{i_2,\ j_2}\| > z_{min} - \varepsilon$。结合（B.7），可得 (4.18)

$$\arccos\langle \overline{\mathbf{Z}}^*_{i_2,j_2}, \overline{\widehat{\mathbf{Z}}}_{i_2,j_2} \rangle = 2\arcsin\frac{\|\overline{\mathbf{Z}}^*_{i_2,j_2} - \overline{\widehat{\mathbf{Z}}}_{i_2,j_2}\|}{2} \leqslant \frac{\pi\varepsilon}{2(z_{min}-\varepsilon)}. \tag{4.19}$$

注意，这个估计对指标对 $(i_3,\ \ j_3)$ 同样成立。那么，

$$\arccos\langle \overline{\mathbf{Z}}^*_{i_2,j_2}, \overline{\mathbf{Z}}^*_{i_3,j_3} \rangle \leqslant \arccos\langle \overline{\mathbf{Z}}^*_{i_2,j_2}, \overline{\widehat{\mathbf{Z}}}_{i_2,j_2} \rangle + \arccos\langle \overline{\widehat{\mathbf{Z}}}_{i_2,j_2}, \overline{\widehat{\mathbf{Z}}}_{i_3,j_3} \rangle + \arccos\langle \overline{\mathbf{Z}}^*_{i_3,j_3}, \overline{\widehat{\mathbf{Z}}}_{i_3,j_3} \rangle \leqslant \Theta_0 + \frac{\pi\varepsilon}{(z_{min}-\varepsilon)}.$$

然而，根据我们在（ Θ_0 （4.15）中的选取上界以及条件（4.14），可得

$$\arccos\langle \overline{\mathbf{Z}}^*_{i_2,j_2}, \overline{\mathbf{Z}}^*_{i_3,j_3} \rangle < \theta_{min}, \tag{4.20}$$

这与 θ_{min} 在（2.11）中的定义矛盾。因此，$(i_2、\ j_2)$、$(i_3、\ j_3)$ 必须被归入不同的簇。

接下来，我们通过反证法证明簇内一致性。假设$\kappa^*_{i_2、\ j_2} = \kappa^*_{i_3、\ j_3}$ 但$\mathbf{Z}_{i_2、\ j_2}$ 和$\mathbf{Z}_{i_3、\ j_3}$ 分别被归入聚类中心$\mathbf{v}_{q_2}$ 和$\mathbf{v}_{q_3}$。由簇内角最大值 Θ_0，可得

$$\arccos\langle \overline{\widehat{\mathbf{Z}}}_{i_2,j_2}, \overline{\mathbf{v}}_{q_2} \rangle \leqslant \Theta_0/2, \quad \arccos\langle \overline{\widehat{\mathbf{Z}}}_{i_3,j_3}, \overline{\mathbf{v}}_{q_2} \rangle \leqslant \Theta_0/2, \tag{4.21}$$

因为$\mathbf{v}_{q_2}$ 和$\mathbf{v}_{q_3}$ 是聚类中心。注意到（4.18）对 $(i_2、\ j_2)$ 以及 $(i_3、\ j_3)$成立。此外，$\mathbf{Z}^*_{i_2、\ j_2} = \mathbf{Z}^*_{i_3、\ j_3}$，因为$\kappa^*_{i_2、\ j_2} = \kappa^*_{i_3、\ j_3}$。因此，

$$\|\overline{\widehat{\mathbf{Z}}}_{i_2,j_2} - \overline{\widehat{\mathbf{Z}}}_{i_3,j_3}\| \leqslant \|\overline{\widehat{\mathbf{Z}}}_{i_2,j_2} - \overline{\mathbf{Z}}^*_{i_2,j_2}\| + \|\overline{\mathbf{Z}}^*_{i_3,j_3} - \overline{\widehat{\mathbf{Z}}}_{i_3,j_3}\| < \frac{2\varepsilon}{z_{min}-\varepsilon}. \tag{4.22}$$

然后再次应用（B.7），我们得到

$$\arccos\langle \overline{\widehat{\mathbf{Z}}}_{i_2,j_2}, \overline{\widehat{\mathbf{Z}}}_{i_3,j_3} \rangle = 2\arcsin\frac{\|\overline{\widehat{\mathbf{Z}}}_{i_2,j_2} - \overline{\widehat{\mathbf{Z}}}_{i_3,j_3}\|}{2} \leqslant \frac{\pi\varepsilon}{z_{min}-\varepsilon}.$$

结合（4.21），可得

$$\arccos\langle \overline{\mathbf{v}}_{q_2}, \overline{\mathbf{v}}_{q_3} \rangle \leqslant \arccos\langle \overline{\widehat{\mathbf{Z}}}_{i_2,j_2}, \overline{\mathbf{v}}_{q_2} \rangle + \arccos\langle \overline{\widehat{\mathbf{Z}}}_{i_3,j_3}, \overline{\mathbf{v}}_{q_2} \rangle + \arccos\langle \overline{\widehat{\mathbf{Z}}}_{i_2,j_2}, \overline{\widehat{\mathbf{Z}}}_{i_3,j_3} \rangle \leqslant \Theta_0 + \frac{\pi\varepsilon}{z_{min} - \varepsilon}. \quad (4.23)$$

然而，由于我们要求簇间角度 θ 大于 θ_0，因此

$$\arccos\langle \overline{\mathbf{v}}_{q_2}, \overline{\mathbf{v}}_{q_3} \rangle \geqslant \theta_0. \quad (4.24)$$

由 (4.15) 中 θ_0 的下界可得矛盾。因此，$\widehat{\mathbf{Z}}_{i_2,\,j_2}$ 和 $\widehat{\mathbf{Z}}_{i_3,\,j_3}$ 必须被归入同一中心。 ■

我们在 (4.15) 中对 Θ_0 下界的选择保证了条件 (4.13) 成立，因为当 $\|\mathbf{Z}^* - \mathbf{Z}\|_{2,\,\infty} < \varepsilon$ 时，簇内角度的上界为 $\frac{\pi\varepsilon}{z_{min}-\varepsilon}$。我们令 θ_0 小于 θ_{min}，以便算法能够分离中心之间达到最小角度 θ_{min}的簇。还要注意，Θ_0 的取值范围使得能够选择 θ_0。

注释 4.8（选择 Θ_0 和 θ_0） Θ_0 和 θ_0 的允许取值范围取决于$\mathbf{Z}$ 估计误差 ε 以及预言机参数 z_{min} 和 θ_{min}。尤其是，较小的 ε 和较大的 z_{min}、θ_{min} 有助于更有利地选择这些阈值

条件（4.14）在 $\varepsilon \leqslant \frac{z_{min}\theta_{min}}{2\pi}$ 时成立，而条件（4.12）要求 $\varepsilon \leqslant \frac{z_{min}}{2}$。因此，命题 4.6 和命题 4.7 中的条件仅在以下情况下才满足

$$\|\mathbf{Z}^* - \widehat{\mathbf{Z}}\|_{2,\infty} < \min\left(\frac{z_{min}\theta_{min}}{2\pi}, \frac{z_{min}}{2}\right) := \varepsilon_0. \quad (4.25)$$

根据（4.11）和命题 4.4中的条件，通过将样本量增加到 $M = \mathcal{O}\left(\frac{(N+K)Q}{\log_{(\varepsilon_0/\sigma_\epsilon)^2}(NK)}\right)$ 的量级，可以得到（4.25）。

注释 4.9 条件 $\|\mathbf{Z}^* - \mathbf{Z}\|_{2,\infty} < \varepsilon_0$ 并不能保证通过 K 均值成功聚类，尤其是在某些簇较小时。不过，它能保证簇内和簇间的一致性。算法 *3* 采用 K 均值进行迭代细化，从而提高了收敛概率，详见注释 *3.4*。

尽管已有多种关于 K 均值成功概率的分析（例如，*[11, 34, 46]*），但这些分析通常依赖特定的分布假设。相比之下，我们不对图或核系数的分布作任何先验假设，而且矩阵$\mathbf{Z}$ 的分布似乎难以处理。因此，我们采用一种确定性条件；尽管该条件较为保守，但切实可行，且数值实验似乎也证实了其合理性。

注意，命题 4.7 并不一定意味着 $\widehat{\boldsymbol{\kappa}} = \boldsymbol{\kappa}^*$，因为$\kappa_{ij}$ 在 $a_{ij} = 0$ 时存在不可辨识性（回顾注 3.3）。接下来我们将证明，可以识别出满足 $a_{ij} > 0$ 的那些 κ_{ij}，并将它们称为活跃核。

定义 4.10（活跃核索引）对于给定的图矩阵 $\mathbf{a}^*$ 和类型矩阵 $\boldsymbol{\kappa}^*$，我们将活跃核索引集合定义为

$$\mathcal{Q}^*_{\text{act}} := \{q = 1, \ldots, Q : \exists\,(i,j),\, a^*_{ij} \neq 0 \ \textit{and}\ \kappa^*_{ij} = q\}. \quad (4.26)$$

我们得到以下结果，它刻画了类型矩阵 $\boldsymbol{\kappa}$的估计误差。

命题 4.11 设 $\|\mathbf{Z}^* - \mathbf{Z}\|_{2,\infty} < \varepsilon_0$ 其中 ε_0 定义见 (4.25)。那么算法 3 给出 $\widehat{\boldsymbol{\kappa}}$ 以及 Q使得

$$\widehat{Q} = |\mathcal{Q}^*_{act}|, \quad \max_{(i,j):a_{ij}\neq 0} |\kappa^*_{ij} - \widehat{\kappa}_{ij}| = 0. \tag{4.27}$$

证明。 注意， $\|\mathbf{Z}^* - \mathbf{Z}\|_{2,\infty} < \varepsilon_0$ 由命题 4.6可知支撑恢复正确；此外，根据命题 4.7，簇间一致性（4.16）和簇内一致性（4.17）也成立。

令 $E^*_{\text{act}} := \{(i, j) : a^*_{ij} \neq 0\}$ 表示活跃索引对的集合。对于 $q \in [Q]$，定义真实类型类

$$E^*_q := \{(i,j) \in E^*_{\text{act}} : \kappa^*_{ij} = q\}.$$

令 $\{I_1、\ldots\ldots、I_{\widehat{Q}}\}$ 为算法 3生成的（非空）簇，使得 $\bigsqcup_{r=1}^{Q} I_r = E^*_{\text{act}}$。

任取 q 满足 $E^*_q \neq \varnothing$ （即 $q \in \mathcal{Q}^*_{\text{act}}$）。根据簇内一致性（4.17）， E^*_q 中的所有索引对都属于同一个估计簇；将其记为 $I_{r(q)}$，其中 $r(q) \in \{1、\ldots\ldots、Q\}$。根据簇间一致性（4.16），不同真实类型的索引对都不属于 $I_{r(q)}$，因此

$$I_{r(q)} = E^*_q.$$

因此，映射 $r : \mathcal{Q}_{\text{act}} \to \{1, \ldots, Q\}$ 定义良好。

（单射性）。若 $q_1 \neq q_2$ 均为活跃的，则 $E^*_{q_1}$ 与$E^*_{q_2}$ 互不相交，并且根据簇间条件，它们必须位于不同的估计簇中；因此 $r(q_1) \neq r(q_2)$。由此可知， r 是单射。

（满射性）。任取一个估计簇 I_r。根据构造， $I_r \subset E^*_{\text{act}}$ and 非空，因此其中的元素都具有某个真实类型 $q \in [Q]$。簇内一致性要求所有 I_r 中的元素都具有同一个真实类型 q，即 ，因此 $I_r = E^*_q$且 $q \in \mathcal{Q}^*_{\text{act}}$。所以， r满射到 $\{1,\ldots,Q\}$。

组合 结合单射性与满射性， r是 $\mathcal{Q}^*_{\text{act}}$与 $\{1,\ldots,Q\}$之间的双射，因此

$$\widehat{Q} = |\mathcal{Q}^*_{\text{act}}|.$$

最后，在活跃对上定义估计类型： $\widehat{\kappa}_{ij} = r(\kappa^*_{ij})$ 对于 $(i, j) \in E^*_{\text{act}}$。根据上述结果，$(i, j) \in E^*_q$当且仅当 $(i, j) \in I_{r(q)}$，因此 $\widehat{\kappa}_{ij}$与 κ^*_{ij}仅相差标签的一个排列 r。采用备注 3.2中引入的排序约定，可得

$$\widehat{\kappa}_{ij} = \kappa^*_{ij} \quad \text{for all } (i,j) \in E^*_{\text{act}},$$

这给出 $\max_{(i,\ j):a_{ij}\neq 0} |\kappa^*_{ij} - \widehat{\kappa}_{ij}| = 0.$ ■

4.4 分解

现在，我们基于估计误差 $\|\mathbf{Z}^* - \mathbf{Z}\|_{2,\infty}$，刻画算法 4的因子分解步骤中的误差。我们需要以下几何事实，它与余弦距离下的簇均值有关，其证明推迟至附录B.2。

引理 4.12（余弦 k-均值下的球面均值） 设 $\mathbf{v}^*$、$\mathbf{z}_1$、……、$\mathbf{z}_n \in \mathbb{R}^d$满足 $\|\mathbf{v}^*\| = \|\mathbf{z}_i\| = 1$ 且对所有 $i = 1$、……、n，均有 $\|\mathbf{v}^* - \mathbf{z}_i\| \leqslant \varepsilon$，其中 $\varepsilon < \sqrt{2}$。定义（欧几里得）均值 $\mathbf{u} := \frac{1}{n}\sum_{i=1}^n \mathbf{z}_i$ 以及余弦-k-均值质心 $\widehat{\mathbf{v}} := \mathbf{u}/\|\mathbf{u}\|$。则

$$\|\mathbf{v}^* - \widehat{\mathbf{v}}\| \leqslant \varepsilon.$$

下面的定理表明，从 $\mathbf{Z}$ 到 $(\mathbf{a}, \mathbf{c})$ 的分解是稳定的。

定理 4.13（从 Z 到 (a, c) 的分解的稳定性）设 $\|Z^* - \widehat{Z}\|_{2,\infty} < \varepsilon$ 且 $\varepsilon < \sqrt{2}\widehat{}$回顾一下，$z^*_{i,j}$ 和 $\widehat{z}_{i,j}$ 分别表示式 (3.3) 中 Z^* 和 Z 的行范数。对于每种核类型 $q \in [Q]$，令$v^*_q := \|c^{*(q)}\| > 0$ 和 $\eta^*_q := 1/v^{*2}_q$。令 $v_{\min} := \min_q\{v^*_q, \widehat{v}_q\}$, $v_{\max} := \max_q\{v^*_q, \widehat{v}_q\}$。

假设正确分配成立，即 (4.27) 成立，并且由 (2.16) 结合$\widehat{}$ $\widehat{}$真实参数定义的矩阵 $\mathbf{A}^* \in \mathbb{R}^{N\times Q}$ 满足 $\sigma_{\min}(\mathbf{A}^*) > 0$。令$\mathbf{A}$ 根据$\mathbf{Z}$ 以与$\mathbf{A}^*$类似的方式构造，并令 $\widehat{\boldsymbol{\eta}}$ 为$\widehat{A}$ $\widehat{\boldsymbol{\eta}} = \mathbf{1}_N$. 的最小二乘解。那么，只要

$$\varepsilon < \min\left(\frac{\sigma_{min}(\mathbf{A}^*)}{3Nv_{max}},\ v_{max},\ (2-\sqrt{2})z_{min}\right), \tag{4.28}$$

以下界成立：

$$\|\boldsymbol{\eta}^* - \widehat{\boldsymbol{\eta}}\| \leqslant \frac{\sqrt{Q}N(2v_{max}+\varepsilon)\varepsilon}{v^2_{min}[\sigma_{\min}(\mathbf{A}^*) - N(2v_{max}+\varepsilon)\varepsilon]}, \tag{4.29}$$

$$\max_{q=1,\dots,Q}\|\Phi^*_q - \widehat{\Phi}_q\|_{L^2(\rho)} \leqslant \frac{v^3_{\max}}{2v^2_{min}}\frac{\sqrt{Q}N(2v_{max}+\varepsilon)\varepsilon}{\sigma_{\min}(\mathbf{A}^*) - N(2v_{max}+\varepsilon)\varepsilon} + \frac{v_{max}\,\varepsilon}{z_{min}-\varepsilon} \tag{4.30}$$

$$\|\mathbf{a}^* - \widehat{\mathbf{a}}\|_F \leqslant \frac{\varepsilon}{v_{min}} + \frac{v^4_{max}}{2v^2_{min}}\frac{\sqrt{Q}N^{\frac{3}{2}}(2v_{max}+\varepsilon)\varepsilon}{\sigma_{\min}(\mathbf{A}^*) - N(2v_{max}+\varepsilon)\varepsilon}. \tag{4.31}$$

特别地，当 ε 较小时（因此 ε^2 可以忽略），定理 4.13 意味着

$$\max_{q=1,\dots,Q}\|\Phi^*_q - \widehat{\Phi}_q\|_{L^2(\rho)} \lesssim \left(\frac{N\sqrt{Q}v^4_{max}}{\sigma_{min}(\mathbf{A}^*)v^2_{min}} + \frac{v_{max}}{z_{min}}\right)\varepsilon, \qquad \|\mathbf{a}^* - \widehat{\mathbf{a}}\|_F \lesssim \left(\frac{N\sqrt{NQ}v^5_{\max}}{\sigma_{\min}(\mathbf{A}^*)v^4_{\min}} + \frac{1}{v_{\min}}\right)\varepsilon. \tag{4.32}$$

证明。(i)。我们首先控制对 $\boldsymbol{\eta}$ $^*\boldsymbol{\eta}^* = \widehat{}\,\widehat{\boldsymbol{\eta}} = \mathbf{1}_N$ *† $\widehat{}^{\,\dagger}$ 的估计误差。回顾 $\mathbf{AA}$。令$\mathbf{A}$和$\mathbf{A}$表示$\mathbf{A}^*$和$\mathbf{A}$的伪逆。根据假设 $\sigma_{min}(\mathbf{A}^*) > 0$，若 $\|\mathbf{A}^* - \widehat{\mathbf{A}}\|_2 < \sigma_{min}(\mathbf{A}^*)$，则有$\mathbf{A}^{*\dagger}\mathbf{A}^* = \widehat{}\,\mathbf{A}^\dagger\widehat{}\,\mathbf{A} = \mathrm{Id}_Q$，其中 Id_Q 是 $\mathbb{R}^{Q\times Q}$ 中的单位矩阵。因此

$$\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* = (\widehat{\mathbf{A}}^\dagger\mathbf{A}^* - \widehat{\mathbf{A}}^\dagger\widehat{\mathbf{A}})\boldsymbol{\eta}^* = \widehat{\mathbf{A}}^\dagger(\mathbf{A}^* - \widehat{\mathbf{A}})\boldsymbol{\eta}^*.$$

因此，我们有

$$\|\boldsymbol{\eta}^* - \widehat{\boldsymbol{\eta}}\| \leqslant \|\widehat{\mathbf{A}}^\dagger\|_2\|\mathbf{A}^* - \widehat{\mathbf{A}}\|_2\|\boldsymbol{\eta}^*\| \leqslant \frac{\|\mathbf{A}^* - \widehat{\mathbf{A}}\|_2\,\|\boldsymbol{\eta}^*\|}{\sigma_{\min}(\mathbf{A}^*) - \|\mathbf{A}^* - \widehat{\mathbf{A}}\|_2}, \tag{4.33}$$

其中，第二个不等式是根据以下事实得出的：$\|\widehat{}\,\mathbf{A}^\dagger\|_2 = 1/\sigma_{min}(\widehat{}\,\mathbf{A}$ $)$ $) \geqslant \sigma_{min}(\ ^*) -$且 $\sigma_{min}(\widehat{}\,\mathbf{AA}\|\mathbf{A}^* - \mathbf{A}\|_2$，因为奇异值关于算子范数是 1-Lipschitz 连续的。显然，$\|\boldsymbol{\eta}^*\| \leqslant \sqrt{Q}/v^2_{min}$ 是由构造得到的，我们接下来需要控制 $\|\mathbf{A}^* - \widehat{}\,\mathbf{A}\|_2$。

$\widehat{}$由于 $\widehat{\boldsymbol{\kappa}} = \boldsymbol{\kappa}^*$，有 $\mathcal{K} = \mathcal{K}^*$，因此第 i 行满足

$$\mathbf{A}^*_i - \widehat{\mathbf{A}}_i = [z^{*2}_{i,1} - \widehat{z}^2_{i,1}, \dots, z^{*2}_{i,N} - \widehat{z}^2_{i,N}]\mathcal{K}^*_i. \tag{4.34}$$

对 $\mathbb{R}^{1\times Q}$ 中的 $\|\cdot\|$ 取欧几里得范数，对于固定的 i，我们有

$$\|\mathbf{A}^*_i - \widehat{\mathbf{A}}_i\| \leqslant \|\mathcal{K}^*_i\|_2(\max_j z^*_{i,j} + \max_j \widehat{z}_{i,j})\sqrt{\sum_{j\neq i, j=1}^{N}\Delta z^2_{i,j}}, \tag{4.35}$$

其中 $\|\mathcal{K}^*_i\|_2$ 表示算子2-范数。现在，对于每个 (i, j),

$$|\Delta z_{i,j}| = \left|\|\mathbf{Z}^*_{i,j}\| - \|\widehat{\mathbf{Z}}_{i,j}\|\right| \leqslant \|\mathbf{Z}^*_{i,j} - \widehat{\mathbf{Z}}_{i,j}\| \leqslant \|\mathbf{Z}^*_{i,j} - \widehat{\mathbf{Z}}_{i,j}\|_{2,\infty} < \varepsilon,$$

并且由于 $\mathbf{a}^*$ 上的行归一化约束，

$$\max_{i,j} z^*_{i,j} = \max_{i,j} a^*_{ij} v^*_{\kappa^*_{ij}} \leqslant v_{max}.$$

因此$\max_{i,\ j} \widehat{z}_{i,\ j} + \max_{i,\ j} z^*_{i,\ j} < 2v_{max} + \varepsilon.$ 最后，根据$\boldsymbol{\mathcal{K}}^*_i \in \{0,\ \ 1\}^{(N-1)\times Q}$ 在（2.2）中的构造，每行恰好有一个 1，因此

$$\|\boldsymbol{\mathcal{K}}^*_i\|_2 \leqslant \|\boldsymbol{\mathcal{K}}^*_i\|_F = \sqrt{N-1}. \tag{4.36}$$

由此我们得出

$$\|\mathbf{A}^*_i - \widehat{\mathbf{A}}_i\| \leqslant \sqrt{N-1}(2v_{max} + \varepsilon)\varepsilon, \tag{4.37}$$

因此，有

$$\|\mathbf{A}^* - \widehat{\mathbf{A}}\|_2 \leqslant N(2v_{max} + \varepsilon)\varepsilon, \tag{4.38}$$

$\widehat{\ }$将估计值代入（4.33）可得（4.29）。通过 $\|\mathbf{A}^* - \mathbf{A}\|_2 < \sigma_{min}(\mathbf{A}^*)$ 的上界满足要求 ε 在（4.28）中。

（ii）。接下来我们考虑核的估计误差。由于我们选择了 $\{\psi_k\}$中的正交归一基 $L^2(\rho)$，只需考虑系数的上界。设 $(i,j) \in I_q$，我们有 $\mathbf{Z}^*_{i,j} = \mathbf{v}^*_q$，其中 $\mathbf{v}^*_q$是对 $\mathbf{c}^{*(q)}$ 进行归一化后的结果，使得 $\mathbf{c}^{*(q)} = v^*_q \mathbf{v}^*_q$，其中 v^*_q表示 $\mathbf{c}^{*(q)}$的 2-范数。我们首先估计 $\mathbf{v}_q$中的误差。根据（4.18），对于任意 $(i,j) \notin I_0$, ‖，我们有 $\mathbf{Z}^*_{i,j} - \overline{\mathbf{Z}}_{i,j}\| < \frac{\varepsilon}{z_{min}-\varepsilon}$。此外，由（3.6）可知，通过 K-means算法得到的估计簇中心满足

$$\widehat{\mathbf{u}}_q = \frac{1}{I_q} \sum_{(i,j)\in I_q} \overline{\widehat{\mathbf{Z}}}_{i,j}, \quad \widehat{\mathbf{v}}_q = \frac{\widehat{\mathbf{u}}_q}{\|\widehat{\mathbf{u}}_q\|}, \quad q = 1, \ldots, Q. \tag{4.39}$$

于是由引理 4.12，我们有

$$\|\mathbf{v}^*_q - \widehat{\mathbf{v}}_q\| < \frac{\varepsilon}{z_{min} - \varepsilon}, \tag{4.40}$$

只要 $\frac{\varepsilon}{z_{min}-\varepsilon} \leq \sqrt{2}$, 而这在 ε 满足（4.28）时成立。随后我们估计 v_q 中的误差。标量映射 $g(t) = 1/t^2$ $|g'(t)| = 2t^{-3}$ $|\eta^*_q - \widehat{\eta}_q| = |1/v^{*2}_q - 1/\widehat{v}^2_q|$ = 满足 。因此，由中值定理，$2t^{-3}|v^*_q - \widehat{v}_q|$, t $\widehat{v}_q$ v^*_q q 对介于 与 之间的某个值。因此，对每个 ，

$$|v^*_q - \widehat{v}_q| = \frac{t^3}{2}|\eta^*_q - \widehat{\eta}_q| \leqslant \frac{v^3_{max}}{2}|\eta^*_q - \widehat{\eta}_q| \leqslant \frac{v^3_{max}}{2}\|\boldsymbol{\eta}^* - \widehat{\boldsymbol{\eta}}\|. \tag{4.41}$$

最后，由于 $\|\mathbf{v}^*\| = 1$，故有，

$$\|\mathbf{c}^{*(q)} - \widehat{\mathbf{c}}^{(q)}\| = |v^*_q - \widehat{v}_q|\|\mathbf{v}^*\| + |\widehat{v}_q|\|\mathbf{v}^*_q - \widehat{\mathbf{v}}_q\| \leqslant \frac{v^3_{max}}{2}\|\boldsymbol{\eta}^* - \widehat{\boldsymbol{\eta}}\| + \frac{v_{max}\,\varepsilon}{z_{min} - \varepsilon}.$$

事实上，（4.30）可通过将（4.29）代入上述方程得到。

（iii）最后，我们估计推断 $\mathbf{a}$ 的误差。如果 $a^*_{ij} = 0$，我们可以恢复 $\widehat{a}_{ij} = 0$，这是因为命题 4.6。那么，如果 $a^*_{ij} \neq 0$，回想一下，$\mathbf{a}$ 是通过（3.10）逐元素恢复的，因此有

$$\begin{aligned}\|\mathbf{a}^* - \widehat{\mathbf{a}}\|_F &= \sqrt{\sum_{i\neq j}\left|\frac{z^*_{i,j}}{v^*_{\kappa^*_{ij}}} - \frac{\widehat{z}_{i,j}}{\widehat{v}_{\widehat{\kappa}_{ij}}}\right|^2} \leqslant \left(\max_{i\neq j}\frac{1}{\widehat{v}_{\widehat{\kappa}_{ij}}}\right)\sqrt{\sum_{i\neq j}\left|z^*_{i,j} - \widehat{z}_{i,j}\right|^2} + \left(\max_{i\neq j}\left|\frac{1}{v^*_{\kappa^*_{ij}}} - \frac{1}{\widehat{v}_{\widehat{\kappa}_{ij}}}\right|\right)\sqrt{\sum_{i\neq j}\left|z^*_{i,j}\right|^2} \\ &\leqslant \frac{1}{v_{min}}\|\mathbf{Z}^* - \widehat{\mathbf{Z}}\|_F + \frac{1}{v^2_{min}}\left(\max_q\left|v^*_q - \widehat{v}_q\right|\right)\|\mathbf{Z}^*\|_F,\end{aligned} \tag{4.42}$$

其中第一步使用了三角不等式，第二步则由定义以及关于 $f(t)=1/t^2$ 的类似中值定理得到。此外，

$$\|\mathbf{Z}^*\|_F=\sqrt{\sum_{i\neq j}\left|a^*_{ij}v^*_{\kappa^*_{ij}}\right|^2}\leqslant v_{max}\sqrt{\sum_{i\neq j}\left|a^*_{ij}\right|^2}=v_{max}\sqrt{\sum_{i=1}^{N}1}=v_{max}\sqrt{N},\tag{4.43}$$

其中最后一步是由$\mathbf{a}^*$的行归一化条件推出的。因此结合（4.42）和（4.41），我们得到

$$\|\mathbf{a}^*-\hat{\mathbf{a}}\|_F\leqslant\frac{\varepsilon}{v_{min}}+\frac{\sqrt{N}v_{max}^4}{2v_{min}^2}\|\boldsymbol{\eta}^*-\hat{\boldsymbol{\eta}}\|$$

而（4.31）可由（4.29）和上面的表达式推出。 ■

恢复的核和图矩阵的误差随 ε 和 $Q^{1/2}$, 线性增长，其常数主要由核系数的差异（即相对核尺度）决定。核误差随 N线性增长，而图误差随 $N^{3/2}$, 以更高的速率增长，因为图矩阵$\mathbf{a}$ 本身的弗罗贝尼乌斯范数随 $N^{1/2}$增长。

现在我们讨论线性系统$\mathbf{A}\boldsymbol{\eta}=\mathbf{1}_N$的适定性。如第3.3节所述，$\mathbf{A}$可能不可逆。事实上，我们将看到秩$(\mathbf{A})$ 与活跃核的数量密切相关。

命题4.14（A 的秩与活跃核的数量）设$\mathbf{A}$定义见（2.16），并回顾$\mathcal{Q}_{act}$ 定义见（4.26）。假设 $N\geqslant|\mathcal{Q}_{act}|$，则

$$\operatorname{rank}(\mathbf{A})\leqslant|\mathcal{Q}_{act}|.\tag{4.44}$$

证明。 令 $v_q:=\|\mathbf{c}^{(q)}\|>0$ 对于 $q\in[Q]$。对于每个 $i\in[N]$ 和 $q\in[Q]$，令

$$S_{iq}:=\sum_{j\neq i:\,\kappa^*_{ij}=q}a_{ij}^2\quad\text{and}\quad\mathbf{S}:=\left[S_{iq}\right]\in\mathbb{R}^{N\times Q}.$$

根据$\boldsymbol{\mathcal{K}}_i$ 在（2.2）中的定义，

$$A_{iq}=\sum_{j\neq i}z_{ij}^2\,(\boldsymbol{\mathcal{K}}_i)_{jq}=\sum_{j\neq i:\,\kappa_{ij}=q}z_{ij}^2.$$

利用 $z_{ij}=a_{ij}v_{\kappa_{ij}}$，我们得到

$$A_{iq}=v_q^2\sum_{j\neq i:\,\kappa_{ij}=q}a_{ij}^2=v_q^2\,s_{iq},$$

即 $\mathbf{A}=\mathbf{S}\,\mathrm{Diag}(v_1^2,\ \ldots\ldots,\ v_Q^2)$。由于 $v_q>0$，在右侧乘以对角矩阵不会改变活跃列上的列相关性，因此秩$(\mathbf{A})$ =秩$(\mathbf{S})$。现在假设存在 $q\in[Q]$使得 $q\notin\mathcal{Q}_{act}$，则有 $S_{iq}=0$，$i\in[N]$，即 $\mathbf{S}$ 的第 i 行全为零，因此秩$(\mathbf{A})$ =秩$(\mathbf{S})\leqslant|\mathcal{Q}_{act}|$. ■

注4.15（秩亏与核范数可辨识性） 如果 秩$(\mathbf{S})=|\mathcal{Q}_{\mathrm{act}}|<Q$，则某些核是非活跃 的，不会出现在系统中；这些核可以忽略，估计仅针对活跃核进行。

若秩$(\mathbf{S})<|\mathcal{Q}_{\mathrm{act}}|$，则$\mathbf{S}$ 的活跃核列线性相关，从而无法唯一识别核范数。例如，当两种核类型在所有行中始终以固定比例出现时，就会出现这种情况。令 $N=3$ 与 $Q=|\mathcal{Q}_{\mathrm{act}}|=2$。设

$$\mathbf{a}=\begin{bmatrix}0&\sqrt{1/3}&\sqrt{2/3}\\\sqrt{1/3}&0&\sqrt{2/3}\\\sqrt{1/3}&\sqrt{2/3}&0\end{bmatrix},\quad\boldsymbol{\kappa}=\begin{bmatrix}*&1&2\\1&*&2\\1&2&*\end{bmatrix}\Longrightarrow\mathbf{S}=\begin{bmatrix}1/3&2/3\\1/3&2/3\\1/3&2/3\end{bmatrix},$$

其列成比例，因此秩$(\mathbf{S}) = 1 < 2$。由于$\mathbf{A} = \mathbf{S}\mathrm{Diag}(v_1^2, v_2^2)$，该依赖关系会传递到$\mathbf{A}$，且系统(2.15)对 $\boldsymbol{\eta}$存在无穷多解。

虽然$\mathbf{S}$在 $\mathcal{Q}_{\mathrm{act}}$ 上的满列秩对于唯一的核范数恢复是必要的，但此处的非唯一性并不关键：对$\mathbf{A}$ 的归一化约束多少有些人为，本质的交互结构编码在$\mathbf{Z}$中。人们仍可准确地恢复核至多相差一个常数缩放因子，这足以实现结构辨识。

5 数值实验

在本节中，我们给出数值示例，说明我们算法的性能，并采用以下误差指标。对于矩阵感知步骤，我们测度

$$\textbf{Z error: } \|\mathbf{Z}^* - \widehat{\mathbf{Z}}\|_{2,\infty}, \tag{5.1}$$

$\widehat{\ }$其中$\mathbf{Z}^*$ 与$\mathbf{Z}$ 分别表示真实的与估计的核图嵌入矩阵。我们使用绝对误差，以便该指标可直接与阈值 ε_0 （见(4.25)）进行比较。在聚类步骤中，我们使用来量化误分类

$$\textbf{Type error: } \frac{\#\{(i,j) : a_{ij}^* \neq 0, \kappa_{ij}^* \neq \widehat{\kappa}_{ij}\}}{N(N-1)}, \tag{5.2}$$

其中 κ_{ij}^*是索引对 $(i,\ j)$的真实核类型。索引对 $(i,\ j)$ 满足 $a_{ij}^* = 0$者被排除，因为相应的核是非活跃的（见注3.3）。为消除标签置换歧义，核按照其在系统中首次出现的顺序重新标记（见注3.2）。我们通过衡量图结构与相互作用核的恢复来评估因子分解步骤。相对图误差为

$$\textbf{Graph error: } \frac{\|\mathbf{a}^* - \widehat{\mathbf{a}}\|_F}{\|\mathbf{a}^*\|_F}, \tag{5.3}$$

其中$\mathbf{a}^*$ 是真实图矩阵，$\|\mathbf{a}^*\|_F = \sqrt{N}$ 则是由于行归一化所致。对于核，类型 q 的相对误差为

$$\textbf{Kernel error: } \frac{\|\Phi_q^* - \widehat{\Phi}_q\|_{L^2(\rho)}}{\|\Phi_q^*\|_{L^2(\rho)}}, \quad q = 1, \dots, Q, \tag{5.4}$$

其中 Φ_q^* 是第 q 个真实核。在展示收敛结果时，我们报告平均核误差，即 Q个单项误差的均值。最后，我们通过$\widehat{\ }$比较根据估计参数（记为$\mathbf{X}$）生成的轨迹与根据真实参数生成的轨迹来评估预测准确率，并使用相同的随机强迫和初始条件。为减轻潜在的$\widehat{\ }$类型误分类的影响，轨迹直接根据估计的嵌入矩阵$\mathbf{Z}$ $\widehat{\ }$生成，而不是根据 ($\widehat{\ }\mathbf{a}$ $\{\Phi_q\}$，$\widehat{\boldsymbol{\kappa}}$)生成。轨迹预测误差定义为

$$\textbf{Trajectory error: } \frac{1}{M'T}\sum_{m=1}^{M'} \|\mathbf{X}_t^{(m)} - \widehat{\mathbf{X}}_t^{(m)}\|_{L^2([0,T];\mathbb{R}^{N\times d})} \approx \frac{1}{M'L}\sum_{m=1}^{M'}\sum_{l=1}^{L} \|\mathbf{X}_{t_l}^{(m)} - \widehat{\mathbf{X}}_{t_l}^{(m)}\|_{\mathbb{R}^{N\times d}}. \tag{5.5}$$

在我们的数值 实验中，除非另有说明，我们假设相互作用核为径向核：

$$\Phi_{\kappa_{ij}}(X_t^j - X_t^i) = \phi_{\kappa_{ij}}\big(|X_t^j - X_t^i|\big)\frac{X_t^j - X_t^i}{|X_t^j - X_t^i|}, \quad X_t^i, X_t^j \in \mathbb{R}^d. \tag{5.6}$$

这一假设简化了计算，但并非必要；我们的方法同样适用于一般的非径向核。

我们先介绍一个典型示例，展示如何在图上联合推断多类型相互作用核。随后，我们研究估计误差随样本量 M 的收敛情况。

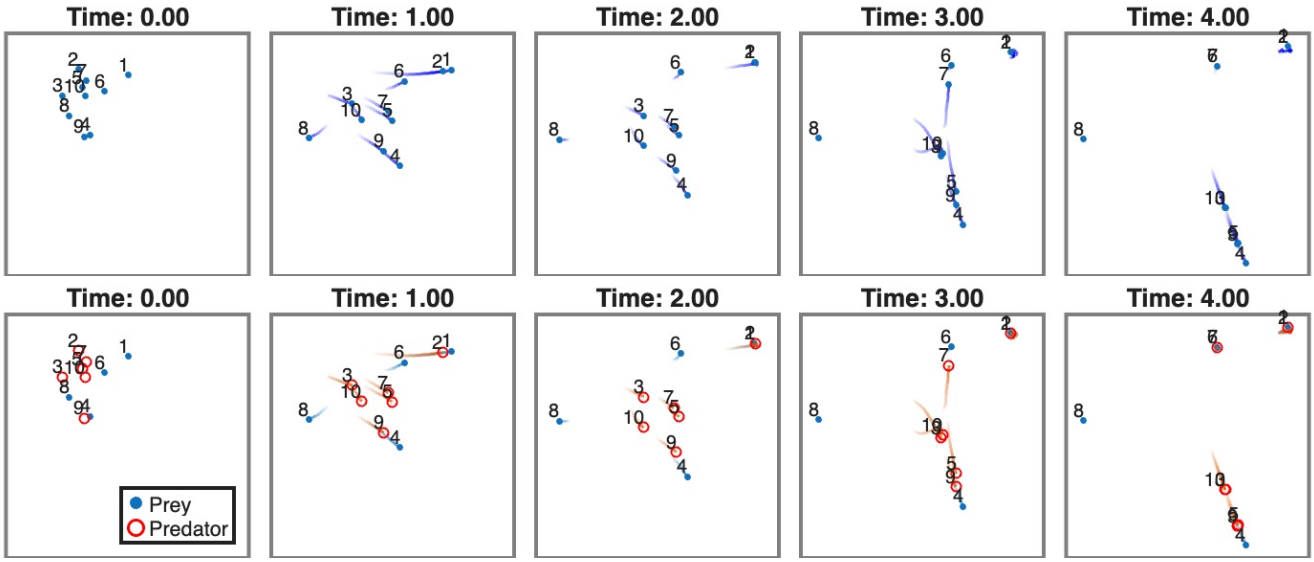

图 3：捕食者-猎物系统的典型轨迹。**上图：** 6 个捕食者和 4 个猎物的观测轨迹数据，单个粒子的角色未知。**下图：** 捕食者和猎物的真实分类。

5.1 典型示例：捕食者-猎物

我们首先考虑一个包含 $N = 10$ 个粒子的捕食者-猎物系统，其中有 6 个捕食者和 4 个猎物。系统中存在四种相互作用类型：捕食者-猎物（ϕ_1）、猎物-猎物（ϕ_2）、猎物-捕食者（ϕ_3）和捕食者-捕食者（ϕ_4）。核采用三次样条表示，在 $[0,5]$上设置 10 个节点，得到维度为 $K = 13$的假设空间。核值为正表示吸引，为负则表示排斥。捕食者会远距离吸引猎物，以此模拟敏锐视觉；而猎物仅会短距离排斥捕食者，反映其较低的警觉性。捕食者-捕食者之间的相互作用表现为轻微吸引，以促进合作；猎物-猎物之间则表现为轻微排斥，以促进分散。真实核见图 4，典型轨迹见图 3。

我们在 $\mathbb{R}^2$ 中生成 $M = 20$ 条独立轨迹，每条轨迹包含 $L = 20$ 个时间步（$\Delta t = 0.1$），初始位置从 $[0,\ 4]^2$ 中均匀采样。动力学采用欧拉法积分，导数通过有限差分近似。施加黏度为 $\sigma = 10^{-3}$ 的随机强迫和 $\sigma_o = 10^{-3}$ 的观测噪声。

联合估计误差见表 3，图结构和核函数的结果见图 4。类型矩阵得到了精确恢复。在此示例中，$z_{min} = 0.5043$，$\theta_{min} = 0.5967$ 得出$\varepsilon_0 = 0.0437$。请注意，条件（4.25）是成功聚类的充分条件，但非必要条件。图误差较小，核函数估计与真实曲线高度吻合；当探索度量较小、 ρ数据较为稀缺时，仅存在细微差异。如表中所示，后处理交替最小二乘步骤进一步降低了所有误差指标。

错误类型	Z	Type	图	核				轨迹
				ϕ_1	ϕ_2	ϕ_3	ϕ_4	
分解	0.173	0	0.036	0.046	0.298	0.207	0.060	1.378
后处理	0.115	0	0.033	0.023	0.106	0.050	0.022	0.549

表3：捕食者-猎物系统的误差指标。

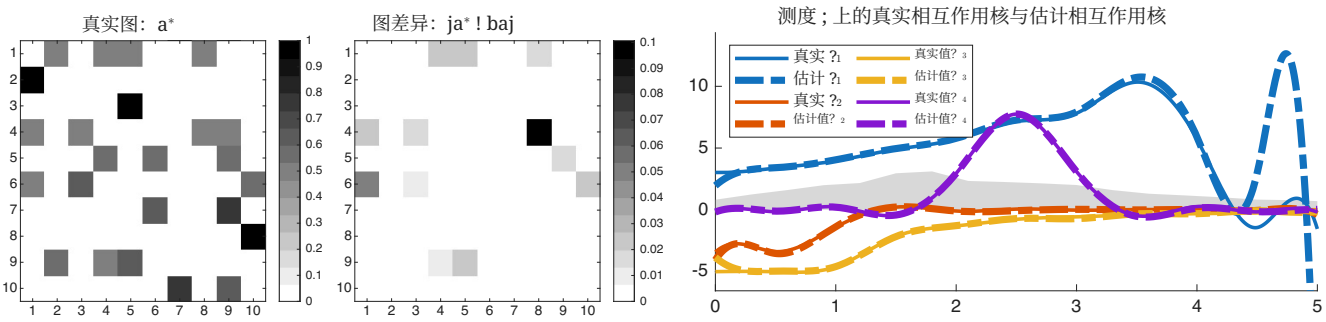

图 4：捕食者-猎物系统中的图、核及其估计值：捕食者-猎物（ϕ_1）、猎物-猎物（ϕ_2）、猎物-捕食者（ϕ_3）和捕食者-捕食者（ϕ_4）。

5.2 样本量为 M 时的收敛性

我们研究估计误差随样本量 M 增加时的收敛情况。我们考虑一个包含 $N = 12$ 个智能体的系统，其相互作用核由具有随机系数的傅里叶基生成，假设空间维数为 $K = 10$，核类型数固定为 $Q = 3$。类型矩阵κ 随机选取，并在维数为 $d = 2$ 的空间中模拟长度为 $L = 1$ 的轨迹，时间步长为$\Delta t = 10^{-1}$，黏度为 $\sigma = 10^{-4}$。样本量 M 在 2^4 和 2^{10}之间按对数均匀分布变化。对于每个M，实验重复 10 次，并报告平均误差及 90% 分位数。Z、图、核以及轨迹预测的误差均采用对数刻度显示。类型分类误差以 0 到 1 之间的比值表示；若未能正确恢复簇数，则将核误差设为默认值 10^3。我们考虑两种图结构：一种相对稀疏，另一种较稠密。二者的差异由邻接矩阵 a 的最小非零元素 a_0决定，稠密图取 0.05，较稀疏图取 0.25。由于行归一化约束，较大的 a_0 对应更高的稀疏度。所选核设置得到 $\theta_{\min} = 0.8272$，且核系数范数的最小值为 $v_{\min} = 0.9273$。由于成功恢复的阈值取决于 $z_{\min}$，而其本身由 a_0 和 $v_{\min}$ 通过（2.13）确定，因此该设置能够直接比较图稀疏度对估计性能的影响。相应结果见图 5。

在矩阵感知阶段，命题 4.4 证明，在 RIP 假设下，Z 的估计误差以 $O(M^{-1/2})$ 的速率衰减。对于稀疏图，从经验上看，我们拟合得到的衰减率为 0.58，与理论预测一致。在聚类阶段，命题 4.6 和 4.7 保证一旦 $\|\mathbf{Z}^* - \mathbf{Z}\|_2, \infty < \varepsilon_0$，我们的算法就能在式 (4.27) 的意义下恢复类型矩阵 $\boldsymbol{\kappa}$。实验中 Z 误差与类型误差之间的高度对应关系证实了这一现象。在我们研究的两个图中，相应的分离阈值为 $\varepsilon_0 = 0.0147$ 和 $\varepsilon_0 = 0.0835$，因此，稠密图中的类型恢复需要更大的样本量。最后，在因子分解阶段，定理 4.13 表明，一旦 $\boldsymbol{\kappa}$ 被精确恢复，图误差和核误差都会以与 Z 误差相同的速率衰减，这一点也得到了数值实验的验证。总体而言，实验结果印证了第4节中建立的理论保证。

我们强调，即使未能成功恢复类型矩阵，轨迹误差仍会下降，因为估计轨迹是直接由嵌入矩阵 $\mathbf{Z}$ 生成的。这突显出，在实践中，准确恢复图和核可能没那么重要，尤其是在其含义可能模糊不清的真实世界数据中；而低秩嵌入 $\mathbf{Z}$ 则能充分刻画系统动力学。此外，我们观察到，在稀疏图和稠密图设定中，$\mathbf{Z}$ 误差的下降行为完全相同。这是意料之中的，因为只应用了一个时间步长，因此矩阵感知阶段中的感知矩阵相同。这一观察也进一步印证了

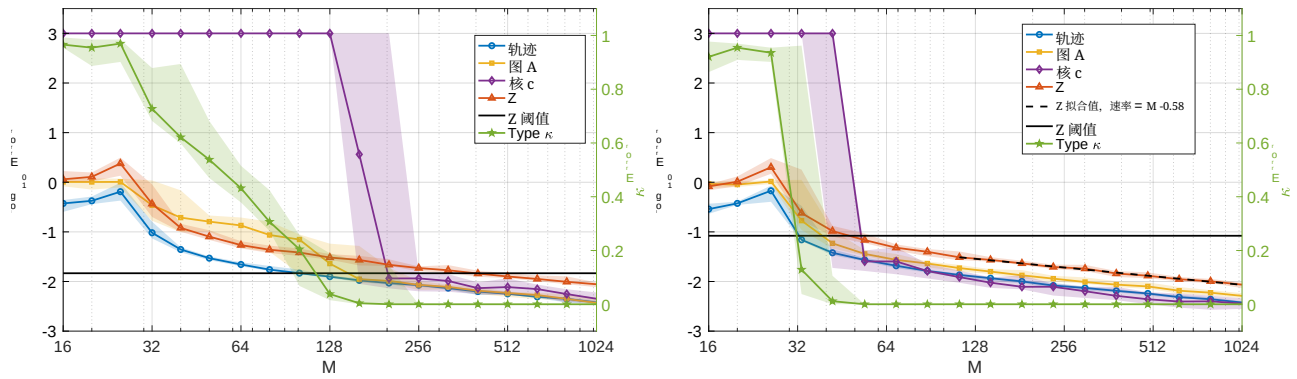

图 5：样本量 M增加时估计误差的收敛情况。**左侧：**稠密图。**右侧：**稀疏图。当 $M > 32$时，**Z** 和轨迹预测的误差开始下降，其中**Z** 误差以速率 $M^{-1/2}$衰减。当**Z** 误差低于分离阈值（约为 $M = 300$ 和 $M = 54$）时，类型矩阵 κ 成功恢复，之后核误差也开始下降。与稠密图相比，稀疏图恢复所需的样本更少。

我们对矩阵感知第一阶段的理论保证。

5.3 智能体数量 N变化时的收敛

我们研究估计误差随 N 从 4 增加到 64 时的收敛情况，系统设置与前一个示例相同。也就是说，我们考虑来自随机傅里叶基的 $Q = 3$ 个核，其参数为 $K = 10$，并在 $d = 2$ 中生成轨迹，时间步长为 $\Delta t = 0.1$，同时施加黏度为 $\sigma = 10^{-4}$的随机强迫。我们取 $L = 10$、$M = 100$，使其足够大，从而聚类估计精确无误，且类型误差为零。

特别地，我们采用网络连通参数 $a_0 = 0.25$；由于行归一化后每行至多有 16 个非零项，随着 N 增大，图会变得更加稀疏。此设定与前一个示例 $\varepsilon_0 = 0.0835$ 得到相同的 ε_0。为突出图稀疏性的影响，我们还考虑一个已知的完全图设定，其中所有智能体以相同强度彼此交互。

图 6 左侧的图展示了图未知且稀疏时，误差和计算时间随 N 的变化规律。计算时间随 N 呈二次增长，这与第3.2.1节的分析一致。轨迹误差随 $N^{0.77}$ 增长，反映了智能体数量的增加；由于图的稀疏性，其增长率略低于线性。对于矩阵误差，Z 的弗罗贝尼乌斯误差呈线性增长，与 (4.8) 一致；而 $(2,\infty)$-误差仅按 $N^{0.08}$ 增长。后者反映了交互图的稀疏性：当 N 增大且下界 a_0 固定时，图会逐渐变得更加稀疏，从而使增长速度减缓。值得注意的是，(4.8) 和 (4.11) 中对 M 的依赖相同，但对 N 的依赖不同，而这一点恰好由此例体现出来。由于 $(2,\infty)$-误差较小，归一化图误差按 $N^{0.20}$ 增长，低于 (4.32) 预测的线性阶，这同样是由稀疏性造成的。核误差实际上按 $N^{-0.59}$ 衰减。这种衰减可由 A 的条件性改善来解释：随着 N 增大，最小特征值增大，误差因而减小。直观而言，如注4.15所述，只有当 A 具有强对称性时才会出现秩亏，进而导致病态性；而在随机生成图中，这种情况不太可能发生，尤其是在 N 较大时。尽管如此，增长率仍受定理 4.13 给出的理论速率控制。

图6的右图展示了已知图且图完整时的结果，所用算法见附录A。轨迹误差、计算时间以及**Z** 的弗罗贝尼乌斯误差与图未知时的情况类似。然而，核误差会随着 N 增大而增大。这是因为，即使

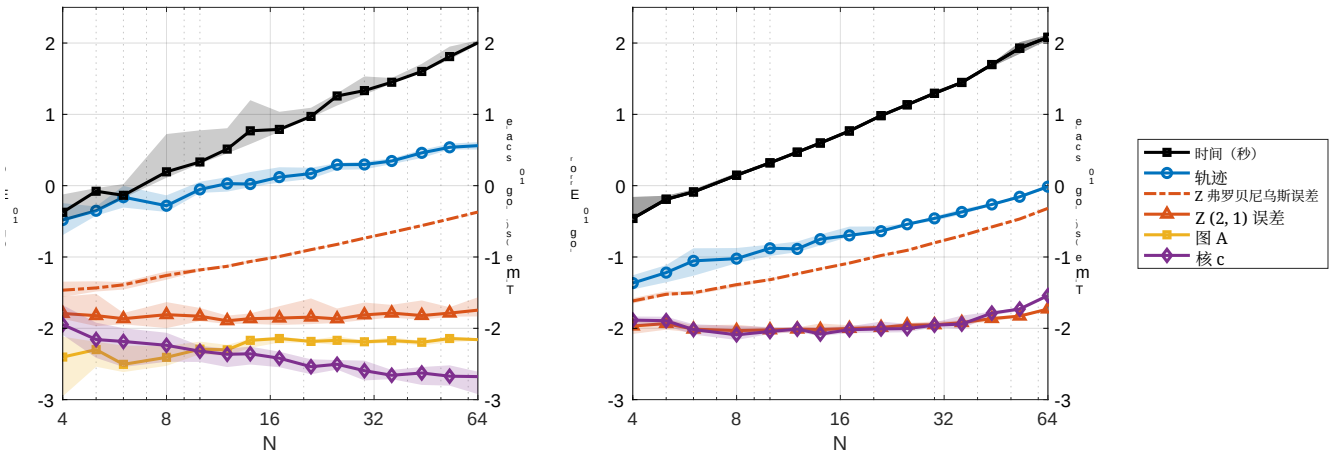

图 6：误差相对于 N 的缩放。**左：** 未知稀疏图。轨迹误差随 N 近似线性增长，而计算时间则呈二次增长。弗罗贝尼乌斯**Z**误差的增长速率与理论值接近，而 $(2,\ \infty)$ 误差的增长则慢得多。出于相同原因，图误差始终较小；随着矩阵**A** 的条件变得更好，核误差也随之减小。**右：** 已知完全图。轨迹误差、计算时间和 Z 弗罗贝尼乌斯误差的表现与图未知时相似。然而，随着 N 增大，核误差也随之增大，因为图的规模不断扩大，妨碍了对 Z 的估计。

已知图，图规模不断增大仍会妨碍对**Z**的估计，从而导致核误差增大。因此，当图较为稠密时，已知图所带来的益处有限。

5.4 必要样本量

现在，我们展示当粒子数 N 和假设空间维数 K 增大时，成功恢复类型分配 κ 所需的样本量 M。

我们令 N 的取值范围为 4 到 16，K 的取值范围为 2 到 10，并设定 $Q = 2$。真实核由随机生成的傅里叶函数构成，并用 K 个三次样条基函数进行近似。这确保了关于 z_{min} 和 θ_{min} 的结构先验信息变化不大。我们将图固定为完全图，并使用附录 A 中的算法，因为在这种设置下也能获得稳定的结构先验。对于每一对 N 和 K，我们在 $d = 2$ 中生成 M 条轨迹，每条轨迹包含 $L = 1$ 个时间步。我们令 M 的取值范围为 2 到 22；对于每个固定的 M，重复进行 $B = 3$ 次实验。我们记录成功恢复所需的最小 M，其中成功恢复意味着 κ 得到了正确恢复。我们在图 7 中展示最小必要样本量 M 随 $N + K$ 增长的变化率。显然，M 大致随 $N + K$ 线性增长，这验证了假设 4.3。

6 结论

我们建立了一个用于学习网络上异质相互作用粒子系统（相互作用粒子系统）的统一框架，其中网络拓扑与相互作用核均未知，且智能体通过潜在类型进行相互作用。通过将联合估计问题重构为三阶段过程——低秩矩阵感知，随后是几何聚类与矩阵分解——我们将优化的连续与离散方面解耦。这一重构得到了一种高效且稳健的算法，并具有强有力的理论保证：矩阵感知阶段的基于 RIP 的恢复界，以及聚类阶段在分离条件下的精确类型恢复。

我们的分析表明，即使类型数量未知，该框架仍能以可证明的精度恢复网络结构、相互作用核和类型分配。数值实验证实了理论所预测的尖锐恢复阈值，并展示了该方法的

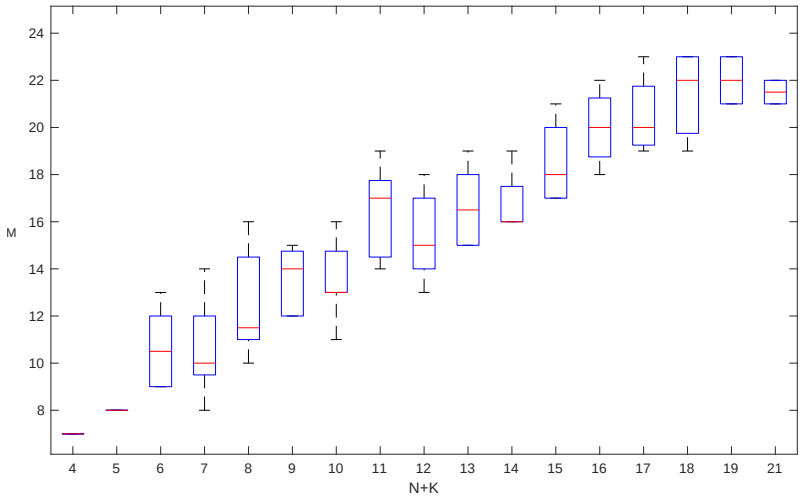

图 7：使用矩阵感知阶段估计得到的 $\mathbf{Z}$ 时，恢复类型函数所需样本量 M 随 $N+K$ 的缩放关系。

可扩展性。

除了本文研究的具体示例外，所提出的方法还广泛适用于生物学、物理学和社会科学领域的异质相互作用粒子系统。未来工作的自然方向包括扩展到更复杂的模型，例如高阶动力学、不完整观测或时变网络。总的来说，本研究为在多智能体系统中实现结构化交互的数据驱动发现提供了一条有原则且可扩展的途径。

A 完全网络

A.1 算法

我们介绍在完全交互图的情形下，学习类型矩阵 $\boldsymbol{\kappa}$ 和系数矩阵$\mathbf{c}$ 的具体步骤，详见注释 2.3 和第 3.4.1 节。第一步是通过算法 $\mathbf{Z}$ 1 估计嵌入矩阵 。在后续阶段，如果已知核的数量Q ，则按照算法 $\mathbf{Z}$6 所述，对 的各行进行位置聚类。当 Q 未知时，我们按照算法 θ 中的步骤，通过将量 Θ 和 与适当的阈值进行比较来确定其值。算法 7。两种情形下都不需要后处理。

$\frown$1: **过程**位置 K-means， Q 已知（估计的 $\mathbf{Z}$， z_0， Q）$\frown$2: 使用 K 均值根据欧氏距离对$\mathbf{Z}$ 的行进行分类，分为 Q 个簇，得到索引集$\{I_q\}_{q=1}^{Q}$ 和簇中心 $\{\mathbf{v}_q\}_{q=1}^{Q}$。 3: 通过设置 $\widehat{\kappa}_{ij}=q$（若 $(i,\ j)\in I_q$对于 $q=1$，……，Q）来构造 $\widehat{\boldsymbol{\kappa}}\in[Q]^{N\times N}$。4: 构造 $\widehat{\ }\mathbf{c}=[\mathbf{v}1$，……，$\mathbf{v}_Q]$。5: **返回** $\widehat{\boldsymbol{\kappa}}$， $\widehat{\ }\mathbf{c}$。

$\widehat{\ }$算法 6：对$\mathbf{Z}$ 进行聚类，已知 Q，适用于完全网络

A.2 数值实验

我们给出完全图情形下的收敛结果。除网络取为完全图，且使用算法6进行估计外，所有实验设置均与第5.2节相同。相应的误差如图8所示。

$\widehat{}$1：过程 位置 K-means， Q 未知(Z, θ_0, Θ_{max}, Q_0, $n_{\max}$) $\widehat{}$2：初始化 $Q = Q_0$，成功 $=$ false 3：对于 $\tau = 0, \ldots, n_{\max}$，执行$\widehat{}$4：使用欧几里得距离，对 $\widehat{}$Z 的各行运行含有 $\widehat{Q}$个簇的 K 均值，得到索引集 $\{I_q\}_{q=1}^{Q}$$\widehat{}$和簇中心 $\{\mathrm{v}_q\}_{q=1}^{Q}$。5：计算最小簇间角度 θ 和最大簇内角 Θ6：如果 $\theta > \theta_0$ 且 $\Theta < \Theta_{\max}$，则 7：成功 $=$true，退出循环 8：否则如果 $\theta < \theta_0$，则$\widehat{}$ $\widehat{}$9：$Q \leftarrow Q - 1$10：否则如果 $\Theta > \Theta_{\max}$，则$\widehat{}$ $\widehat{}$11：$Q \leftarrow Q + 1$12：如果成功 $=$false，则 13：返回分类失败 14：否则 15：通过设置 $\widehat{\kappa}_{ij} = q$（如果 $(i$，则对 $q = 1, \ldots, Q$ 设置 $j) \in I_q$）来构造 $\widehat{\boldsymbol{\kappa}} \in [Q]^{N \times N}$。16：构造 $\widehat{}$c $=$ [v1, . . . , v$_Q$]。17：返回 $\widehat{\boldsymbol{\kappa}}$, $\widehat{}$c。

$\widehat{}$ $\widehat{}$算法 7：在未知 Q的情况下对完全网络聚类**Z**

B 证明

B.1 矩阵感知与受限等距性质预备

矩阵感知问题旨在寻找一个低秩矩阵 $Z^* \in \mathbb{R}^{n_1 \times n_2}$，数据为 $b_m = \langle A_m, Z^* \rangle_F$，其中 $A_1, \cdots, A_M \in \mathbb{R}^{n_1 \times n_2}$为感知矩阵[1]。为了寻找 Z^*，其秩为 $r \ll n_1 \wedge n_2$，需要求解如下非凸优化问题

$$\min_{Z \in \mathbb{R}^{n_1 \times n_2}, \mathrm{rank}(Z) = r} F(Z) = \frac{1}{M} \sum_{m=1}^{M} |\langle A_m, Z \rangle_F - b_m|^2 = \frac{1}{M} \sum_{m=1}^{M} |\mathrm{Tr}(A_m^{\mathrm{T}} Z) - b_m|^2 . \tag{B.1}$$

众所周知，约束优化问题（B.1）是NP难的。一种常见的分解方法由Burer和Monteiro提出 [6, 7]。具体而言，我们将 $Z = UV^{\mathrm{T}}$ 表示出来，其中 $U \in \mathbb{R}^{n_1 \times r}$且 $V \in \mathbb{R}^{n_2 \times r}$。于是（B.1）可转化为一个无约束问题

$$\min_{U \in \mathbb{R}^{n_1 \times r}, V \in \mathbb{R}^{n_2 \times r}} F(U, V) = \frac{1}{M} \sum_{m=1}^{M} |\langle A_m, UV^{\mathrm{T}} \rangle_F - b_m|^2. \tag{B.2}$$

为了 si 为简化记号，我们定义一个线性感知算子 $\mathscr{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^M$ b y

$$\mathscr{A}(Z) = \left(\frac{1}{\sqrt{M}} \langle A_1, Z \rangle_F, \cdots, \frac{1}{\sqrt{M}} \langle A_M, Z \rangle_F \right). \tag{B.3}$$

定义 B.1（受限等距性质（RIP）） 线性映射 $\mathscr{A}$ 满足 (r, δ_r)-*RIP* 条件，*RIP* 常数为 $\delta_r = \delta_r(\mathscr{A}) \in [0, 1)$，当且仅当存在一个正数常数 C，使得

$$(1 - \delta_r) \|Z\|_F^2 \leqslant \frac{1}{C} \|\mathscr{A}(Z)\|_{\mathbb{R}^m}^2 = \frac{1}{CM} \sum_{m=1}^{M} \langle A_m, Z \rangle_F^2 \leqslant (1 + \delta_r) \|Z\|_F^2 \tag{B.4}$$

[1]我们在第 $A_m := \mathbf{B}_i(\mathbf{X}_0^m) \in \mathbb{R}^{(N-1) \times K}$ 节中4.2。

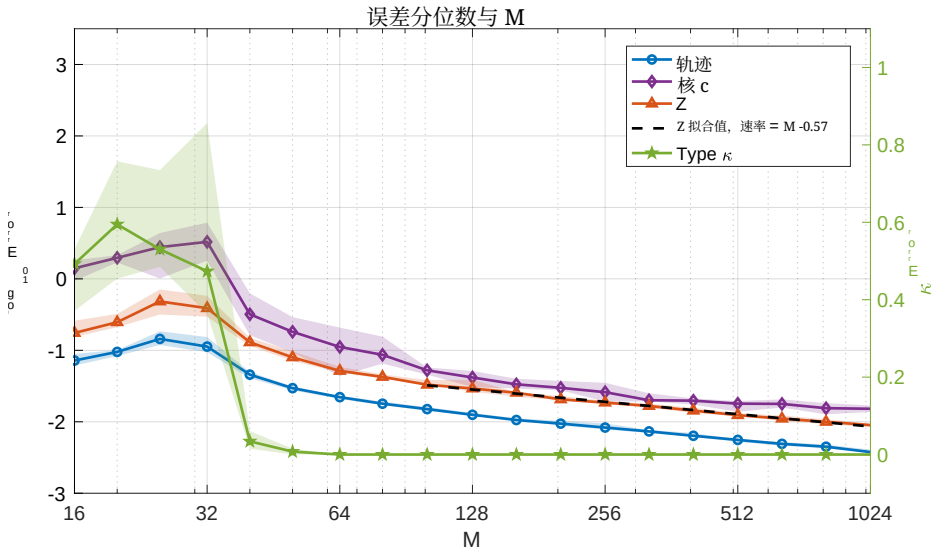

图 8：已知图时，估计误差随样本量 M 的收敛情况。在此设置下，无需进行图恢复，因此我们仅报告核误差和类型误差。由于所有核均处于活跃状态，且已知核类型的数量 Q，即使轨迹数量很少，也可以评估核估计结果。结果表明，核误差随 M 的增加而降低，其速率基本与**Z**误差相同。拟合得到的衰减率 -0.57 证实了这一表现，与理论矩阵感知速率一致。与一般的图未知情形相比，所需样本量大幅减少。

适用于　所有秩至多为 r 的 Z，且 δ_r 是上述不等式的最小常数　　$s.$

归一化常数 C 在条件 (B.4) 中由 [39], 引入，使得 RIP 可应用于一大类能够通过缩放达到近似等距变换的感知矩阵。特别是在我们的设定中，感知矩阵通常与等距变换相差甚远，因此这种缩放尤为重要。

受限等距性质在矩阵感知理论中起着关键作用 [38], ，矩阵感知是压缩感知的一种推广 [9]。当数据 $\{b_m\}$ 无噪声时，它确保矩阵感知问题具有可辨识性。

定理 B.2（[38]中的定理 3.2）假设 $\mathscr{A}$ 满足具有 $\delta_{2r} < 1$和整数 $r \geqslant 1$的 $(2r, \delta_{2r})$-*RIP* 条件。那么， Z^* 是唯一一个秩至多为 r 且满足 $\mathscr{A}(Z) = b = [b_1, \cdots, b_M]^{\mathrm{T}}$及 $b_m = \langle A_m, Z^* \rangle_F$的矩阵。

当感测数据 $\{b_m\}$ 存在噪声时，非对称问题 $n_1 \neq n_2$中一种广泛使用的正则化项（例如参见 [45, 19]）为

$$\min_{U \in \mathbb{R}^{n_1 \times r}, V \in \mathbb{R}^{n_2 \times r}} F(U, V) = \frac{1}{M} \sum_{m=1}^{M} |\langle A_m, UV^{\mathrm{T}} \rangle_F - b_m|^2 + \frac{1}{4} \left\| U^{\mathrm{T}} U - V^{\mathrm{T}} V \right\|_F^2, \tag{B.5}$$

下一个定理给出了正则化估计量的误差界。

定理 B.3（[19]中的定理 32）假设感知算子 $\mathscr{A}$ 满足 $(2r,\ 1/20)$-*RIP* 条件，且观测值为 $b_m = \langle A_m,$ $Z^* \rangle_F + \varepsilon_m,\ m = 1,\ \ldots M,$ 其中噪声项 $\{\varepsilon_m\}$是分布为 $\mathcal{N}(0,\ \sigma_\epsilon^2)$ 的独立同分布高斯随机变量。那么，以高概率，非凸目标函数 $(U,\ V)$ 的每个局部极小值都满足 (B.5) 如下条件

$$\left\| UV^{\mathrm{T}} - Z^* \right\|_F \lesssim \sigma_\epsilon \sqrt{\frac{(n_1 + n_2) r \log M}{M}}. \tag{B.6}$$

这里，若点 $x=(U,V)$ 满足 $\nabla F(x)=0$ 和 $\nabla^2 F(x)\succeq 0$，则称其为 F 的局部极小值。

B.2 几何事实

The 以下几何事实有助于给出第4节中的可分性。 3.

引理 B.4 设 x、$y\in\mathbb{R}^d$ 为非零向量，并将其归一化向量定义为 $\bar{x}=\frac{x}{\|x\|}$、$\bar{y}=\frac{y}{\|y\|}$。则，

$$\|\bar{x}-\bar{y}\|\leqslant\frac{1}{\min(\|x\|,\|y\|)}\|x-y\|$$

and

$$\arccos\langle\bar{x},\bar{y}\rangle=2\arcsin\frac{\|\bar{x}-\bar{y}\|}{2}\leqslant\frac{\pi}{2}\|\bar{x}-\bar{y}\|. \tag{B.7}$$

引理 B.4的证明。不失一般性，假设 $\|x\|\leqslant\|y\|$。我们旨在证明 $\|\bar{x}-\bar{y}\|\leqslant\frac{1}{\|x\|}\|x-y\|$。将两边平方，并用右边减去左边，只需证明

$$\frac{1}{\|x\|^2}\|x-y\|^2-\|\bar{x}-\bar{y}\|^2\geqslant 0.$$

由于 $\|\bar{x}-\bar{y}\|^2=2-2\frac{\langle x,y\rangle}{\|x\|\|y\|}$ 且 $\frac{1}{\|x\|^2}\|x-y\|^2=1+\frac{\|y\|^2}{\|x\|^2}-2\frac{\langle x,y\rangle}{\|x\|^2}$，上述不等式等价于

$$\frac{\|y\|^2}{\|x\|^2}-1+2\frac{\langle x,y\rangle}{\|x\|}\left(\frac{1}{\|y\|}-\frac{1}{\|x\|}\right)\geqslant 0. \tag{B.8}$$

由柯西–施瓦茨不等式 $|\langle x,\ y\rangle|\leqslant\|x\|\|y\|$，我们有

$$\left|2\frac{\langle x,y\rangle}{\|x\|}\left(\frac{1}{\|y\|}-\frac{1}{\|x\|}\right)\right|\leqslant 2\|y\|\left(\frac{1}{\|x\|}-\frac{1}{\|y\|}\right)=\frac{2(\|y\|-\|x\|)}{\|x\|}.$$

同时，由于 $\|y\|\geqslant\|x\|$，我们有

$$\frac{\|y\|^2}{\|x\|^2}-1=\frac{(\|y\|+\|x\|)(\|y\|-\|x\|)}{\|x\|^2}\geqslant\frac{2(\|y\|-\|x\|)}{\|x\|}.$$

因此，（B.8）成立，等号仅当 x 与 y 共线且 $\|x\|=\|y\|$时成立。

To p证明（B.7），令 $\theta=\arccos\langle\bar{x},\ \bar{y}\rangle\in[0,\ \pi]$。由于 $\bar{x}$ 与 $\bar{y}$ 均为单位向量，我们有

$$\|\bar{x}-\bar{y}\|^2=\|\bar{x}\|^2+\|\bar{y}\|^2-2\langle\bar{x},\bar{y}\rangle=2-2\cos\theta=4\sin^2\frac{\theta}{2},$$

即 $\|\bar{x}-\bar{y}\|=2\sin(\theta/2)$。因此，

$$2\arcsin\left(\frac{\|\bar{x}-\bar{y}\|}{2}\right)=2\arcsin(\sin(\theta/2))=\theta=\arccos\langle\bar{x},\bar{y}\rangle,$$

其中我们用到了 $\theta/2\in[0,\pi/2]$，因此 $\arcsin(\sin(\theta/2))=\theta/2$。最后一个不等式由 $\arcsin(x)\leqslant\frac{\pi x}{2}$ 对所有 $x\in[0,1]$成立这一事实得到。 ■

证明 o f 引理 4.12。由于 $\|\mathbf{z}_i\|=\|\mathbf{v}^\star\|=1$，我们有 $\|\mathbf{z}_i-\mathbf{v}^\star\|^2=2-2\mathbf{z}_i^{\mathrm{T}}\mathbf{v}^\star$，因此 e

$$\mathbf{z}_i^{\mathrm{T}}\mathbf{v}^\star=1-\tfrac{1}{2}\|\mathbf{z}_i-\mathbf{v}^\star\|^2\ \geqslant\ 1-\frac{\varepsilon^2}{2}.$$

对 i gives 取平均可得

$$\mathbf{v}^{\star\mathrm{T}}\mathbf{u} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i^{\mathrm{T}}\mathbf{v}^{\star} \;\geqslant\; 1 - \frac{\varepsilon^2}{2}.$$

若 $\varepsilon < \sqrt{2}$ 则 $\mathbf{v}^{\star\mathrm{T}}\mathbf{u} > 0$，因此 $\mathbf{u} \neq \mathbf{0}$ 和 $\mathbf{v}$ 是良定义的。利用 $\|\mathbf{u}\| \leqslant 1$,

$$\mathbf{v}^{\star\mathrm{T}}\mathbf{v} = \frac{\mathbf{v}^{\star\mathrm{T}}\mathbf{u}}{\|\mathbf{u}\|} \;\geqslant\; \mathbf{v}^{\star\mathrm{T}}\mathbf{u} \;\geqslant\; 1 - \frac{\varepsilon^2}{2}.$$

最后,

$$\|\mathbf{v} - \mathbf{v}^{\star}\|^2 = 2\bigl(1 - \mathbf{v}^{\star\mathrm{T}}\mathbf{v}\bigr) \;\leqslant\; 2\left(1 - \left(1 - \frac{\varepsilon^2}{2}\right)\right) = \varepsilon^2,$$

由此得到 $\|\mathbf{v} - \mathbf{v}^{\star}\| \leqslant \varepsilon$。 ∎

致谢

本研究得到美国国家科学基金会资助，资助项目编号为 DMS-2238486 和 DMS-2511283；并得到空军科学研究办公室资助，资助项目编号为 AFOSR-FA9550-20-1-0288、FA9550-21-1-0317 和 FA9550-23- 1-0445。

参考文献

[1] David Arthur 和 Sergei Vassilvitskii。*K* 均值++：谨慎初始化的优势。技术报告。斯坦福大学，2006年。[2] Sohail Bahmani 和 Justin Romberg。“相位恢复与统计学习理论相遇：一种灵活的凸松弛方法”。载于：人工智能与统计国际会议论文集。第54卷。机器学习研究会议论文集。2017年，第252–260页。[3] Aaron S. Baumgarten 和 Ken Kamrin。“一种适用于致密细颗粒悬浮体系的通用本构模型，并在多种几何构型中得到验证”。载于：美国国家科学院院刊 116.42（2019年10月），第20828–20836页。[4] Nathan Bell、Yizhou Yu 和 Peter J. Mucha。“颗粒材料的基于粒子的模拟”。载于：*2005*年*ACM SIGGRAPH/Eurographics*计算机动画研讨会论文集。SCA ’05。2005年7月，第77–86页。[5] Katherine Betz、Feng Fu 和 Naoki Masuda。“具有两个社群的网络中存在环境反馈的演化博弈动力学”。载于：数学生物学通报 86.7（2024年），论文编号84，共45页。[6] Samuel Burer 和 Renato D. C. Monteiro。“通过低秩分解求解半定规划的非线性规划算法”。载于：《数学规划》95.2（2003年），第329–357页。[7] Samuel Burer 和 Renato D. C. Monteiro。“低秩半定规划中的局部极小值与收敛”。载于：《数学规划》103.3（2005年），第427–444页。[8] Brian A. Camley 和 Wouter-Jan Rappel。“细胞集体运动的物理模型：从细胞到组织”。载于：《物理学杂志*D*：应用物理》50.11（2017年3月），第113002页。[9] Emmanuel J. Candès 和 Terence Tao。“基于线性规划的解码”。载于：*IEEE*信息论汇刊51.12（2005年），第4203–4215页。[10] Kabir Aladin Chandrasekher、Mengqi Lou 和 Ashwin Pananjady。“广义秩一矩阵感知的交替最小化：随机初始化下的精确预测”。载于：《信息与推断：*IMA*期刊》13.3（2024年），iaae025。

[11] Xiaohui Chen 和 Yun Yang。“高斯混合模型精确恢复的截断界限”。载于：*IEEETransactions on Information Theory* 67.6（2021年），第4223–4238页。[12] Felipe Cucker 和 Steve Smale。“论学习的数学基础”。载于：美国数学会通报39.1（2002年），第1–49页。[13] Felipe Cucker 和 Steve Smale。学习理论：逼近理论视角。第24卷。2007年。[14] Esther S. Daus、Markus Fellner 和 Ansgar Jüngel。“多物种随机相互作用粒子系统的随机批次方法”。载于：*Journal of Computational Physics* 463（2022年），论文编号111220，共27页。[15] Andreas Deutsch、Josué Manik Nava-Sedeño、Simon Syga 和 Haralampos Hatzikirou。“BIO-LGCA：用于分析集体细胞迁移的细胞自动机建模类别”。载于：*PLOS*计算生物学17.6（2021年），e1009066。[16] Jinchao Feng、Mauro Maggioni、Patrick Martin 和 Ming Zhong。“从基于智能体的系统观测值中学习相互作用变量与核”。载于：*IFAC 2022*论文集（2022年）。[17] Jinchao Feng、Yunxiang Ren 和 Sui Tang。“利用高斯过程数据驱动发现相互作用粒子系统”。载于：*arXiv*预印本 *arXiv:2106.02735*（2021年）。[18] Thomas Gaskin、Grigorios A Pavliotis 和 Mark Girolami。“从时间序列推断网络：一种神经网络方法”。载于：*PNAS Nexus* 3.4（2024年），第63页。[19] Rong Ge、Chi Jin 和 Yi Zheng。“非凸低秩问题不存在伪局部极小值：统一的几何分析”。载于：*International Conference on Machine Learning*。PMLR。2017年，第1233–1242页。[20] Rainer Hegselmann。“意见动力学与有限信任：模型、分析和模拟”。载于：*The Journal of Artificial Societies and Social Simulation*（2015年）。[21] Shi Jin、Lei Li 和 Jian-Guo Liu。“不同物种与权重的相互作用粒子的随机批次方法收敛性”。载于：*SIAM Journal on Numerical Analysis* 59.2（2021年），第746–768页。[22] Ken Kamrin、Kimberly M. Hill、Daniel I. Goldman 和 Jose E. Andrade。“致密颗粒介质建模进展”。载于：*Annual Review of Fluid Mechanics* 56.1（2024年1月），第215–240页。[23] Quanjun Lang 和 Fei Lu。“相互作用粒子平均场方程中相互作用核的可辨识性”。载于：*Foundations of Data Science* 5.4（2023年），第480–502页。[24] Quanjun Lang 和 Fei Lu。“一阶相互作用粒子系统平均场方程中的相互作用核学习”。载于：*SIAM Journal on Scientific Computing* 44.1（2022年），A260–A285。[25] Quanjun Lang、Xiong Wang、Fei Lu 和 Mauro Maggioni。“网络上的相互作用粒子系统：联合推断网络与相互作用核”。载于：*arXiv*预印本 *arXiv:2402.08412*（2024年）。[26] Kiryung Lee 和 Dominik Stöger。“随机初始化的交替最小二乘：矩阵感知的快速收敛”。载于：*SIAM*数据科学数学期刊5.3（2023年），第774–799页。[27] Lei Li、Yingzhou Li、Jian-Guo Liu、Zibu Liu 和 Jianfeng Lu。“用于高效采样的 Stein 变分梯度下降随机版本”。载于：*Communications in Applied Mathematics andComputational Science* 15.1（2020年），第37–63页。[28] Qiang Liu 和 Dilin Wang。“Stein变分梯度下降：一种通用的贝叶斯推断算法”。载于：*Advances in Neural Information Processing Systems*。第29卷。2016年。

[29] Yi-kai Liu. “基于 Pauli 测量的通用低秩矩阵恢复”。载于：*Advances in NeuralInformation Processing Systems*。第 24 卷。2011 年。[30] Stuart Lloyd. “PCM 中的最小二乘量化”。载于：*IEEE transactions on information theory*28.2 (1982)，第 129–137 页。[31] Fei Lu、Mauro Maggioni 和 Sui Tang. “从多条轨迹中学习异质主体系统的相互作用核”。载于：机器学习研究杂志 22.32 (2021)，第 1–67 页。[32] Fei Lu 、 Mauro Maggioni 和 Sui Tang. “ 从多条轨迹中学习随机粒子相互作用系统的相互作用核”。载于：计算数学基础 (2021)，第 1–55 页。[33] Fei Lu、Ming Zhong、Sui Tang 和 Mauro Maggioni. “基于轨迹数据推断主体系统中的相互作用定律”。载于：*Proc. Natl. Acad. Sci. USA* 116.29 (2019)，第 14424–14433 页。[34] Yu Lu 和 Harrison H Zhou. “Lloyd 算法及其变体的统计与计算保证”。载于：*arXiv* 预印本 *arXiv:1612.02099* (2016)。[35] Daniel A. Messenger 和 David M. Bortz. “使用 WSINDy 从粒子数据中学习平均场方程”。载于：*Physica D*：非线性现象 439 (2022)，第 133406 页。[36] Daniel A. Messenger、Graycen E. Wheeler、Xuedong Liu 和 David M. Bortz. “从异质细胞群体的个体轨迹中学习各向异性相互作用规则”。载于：英国皇家学会界面期刊 19.195 (2022)。[37] Tiago Pereira、Edmilson Roque dos Santos 和 Sebastian van Strien. “稀疏网络动力学的稳健重建”。载于：*arXiv* 预印本 *arXiv:2308.06433* (2023)。[38] Benjamin Recht、Maryam Fazel 和 Pablo A. Parrilo. “通过核范数最小化求解线性矩阵方程并保证得到最小秩解”。载于：*SIAM* 评论 52.3 (2010)，第 471–501 页。[39] Angelika Rohde 和 Alexandre B. Tsybakov. “高维低秩矩阵的估计”。载于：统计年鉴 39.2 (2011)，第 887–930 页。[40] Javier E. Satulovsky 和 Tânia Tomé. “捕食者-猎物系统的随机格气模型”。载于：物理评论 *E* 49.6 (1994)，第 5073–5079 页。[41] Timothy M. Schaerf、James E. Herbert-Read 和 Ashley J. W. Ward. “一种识别移动动物群体中个体间不同相互作用规则的统计方法”。载于：英国皇家学会界面期刊 18.176 (2021)，第 20200925–20200925 页。[42] Rinaldo B. Schinazi. “捕食者-猎物与宿主-寄生虫空间随机模型”。载于：应用概率年鉴 7.1 (1997 年 2 月)。[43] Pei Tan 和 Christopher E. Miles. “通过降维实现异质集体运动中亚群体的内在统计分离”。载于：物理评论 *E* 109.1 (2024)，论文编号 014403，共 9 页。[44] Richard Y. Zhang、Somayeh Sojoudi 和 Javad Lavaei. “非凸矩阵恢复中伪局部极小值不存在的尖锐受限等距界。”载于：机器学习研究杂志 20.114 (2019)，第 1–34 页。[45] Qinqing Zheng 和 John Lafferty. “使用 Burer–Monteiro 分解和梯度下降进行矩形矩阵补全的收敛性分析”。载于：*arXiv* 预印本 *arXiv:1605.07051* (2016)。[46] Yubo Zhuang、Xiaohui Chen、Yun Yang 和 Richard Y. Zhang. “通过非负低秩半定规划实现统计最优 K 均值聚类”。载于：*arXiv* 预印本 *arXiv:2305.18436*(2023)。