arXiv:2301.01430v3 [eess.SY] 9 Sep 2023

相似线性时不变动力系统的多任务系统辨识

Yiting Chen、Ana M. Ospina、Fabio Pasqualetti 和 Emiliano Dall'Anese

摘要——本文提出一种受多任务学习启发的系统辨识框架，通过利用系统之间的结构相似性，联合估计给定数量的线性时不变（LTI）系统的动态特性。具体而言，我们考虑用于建模网络化系统的 LTI 系统，这些系统具有相似的连通性，或者其矩阵之间仅存在微小差异。系统辨识任务包括最小化各个系统的最小二乘（LS）拟合误差，并加入一个用于强制实现结构相似性的正则化函数。当一个或多个 LTI 系统的记录轨迹不够丰富，导致最小二乘法病态时，所提出的方法尤为适用。我们分析了所提出的方法在 LTI 系统矩阵具有共同稀疏模式（即连通性相似）时的性能，并基于真实数据提供了用于估计脑动力学的仿真结果。我们表明，与最小二乘法达到相似的误差水平相比，所提出的方法所需的功能性磁共振成像扫描数量显著更少。

I. 引言

系统辨识是一项核心任务，其目标是根据观测到的输入和状态来估计动力系统的模型 [1], [2]。尤其是，线性时不变（LTI）系统的辨识已得到广泛研究，近年来，由于数据驱动控制与优化领域的相关研究，这一问题再次受到关注（例如，参见 [3]–[8] 中具有代表性的研究成果）。

当状态观测不含噪声时，只要记录轨迹满足持续激励（PE）条件（例如，参见 [9], [10]中的讨论），就可以利用 Willems 基本引理估计线性时不变系统的系统矩阵。另一方面，当过程噪声或扰动进入线性时不变系统时，已有多项研究关注最小二乘（LS）估计量的渐近和有限时间估计误差及样本复杂度；代表性工作可参见 [11]–[16]，以及其中的相关参考文献。此外，正则化系统辨识方法也已有研究，例如 [2]，[17], [18]； [19]则研究了通过正则化回归进行低阶线性系统辨识。这些正则化辨识方法允许对系统矩阵加入先验，并在最小二乘拟合与模型复杂度之间取得平衡 [20], [21]。

Y. Chen、A. M. Ospina 和 E. Dall'Anese 任职于科罗拉多大学博尔德分校电气、计算机与能源工程系；电子邮箱：{yiting.chen-1、ana.ospina、emiliano.dallanese}@colorado.edu。F. Pasqualetti 任职于加州大学河滨分校机械工程系；电子邮箱：fabiopas@engr.ucr.edu。本研究部分由美国国家科学基金会（NSF）通过项目资助编号 1941896 和 1926829，以及美国陆军研究办公室（ARO）通过项目资助编号 W911NF1910360 提供支持。

最小二乘估计量的性能取决于是否有足够丰富的记录轨迹，从而使最小二乘问题具有良好的条件数。本文关注最小二乘法病态的情形。具体而言，我们考虑估计 $N > 1$ 线性时不变系统的系统矩阵这一任务，其中至少一个系统（或某些系统）的记录轨迹不够长（也不够丰富）。因此，本文提出的问题如下：能否利用这些 N 系统之间的“相似性”，即使最小二乘问题病态，也能准确估计系统矩阵？具体来说，如果某个 i系统只有少量测量数据，能否利用其他线性时不变系统的记录数据来降低估计误差？

沿着这一思路， [22] 研究了如何根据由“相似”线性系统生成的样本来估计其矩阵；在 [22], 中，如果一个线性时不变系统的矩阵是给定矩阵经过扰动后得到的版本，就认为它与该矩阵“相似”。最近， [23]研究了这样一种情形：线性时不变系统的矩阵之间的差的矩阵范数较小。本文扩展了“结构相似性”的概念，以涵盖 N 系统可能具有的其他共同属性，并提出一种新的系统辨识方法，利用并交叉融合多任务学习 [24]–[26], 、统计学习 [20], 和正则化辨识方法 [2]，[27]领域所研究的核心工具。

我们首先考虑这样一种情形： N线性时不变系统对网络化系统建模，且具有相似的连通（子）图；这意味着这些线性时不变系统的矩阵具有共同稀疏性（即系统矩阵在一组相同的元素位置上为零）。基于这一模型，我们将多任务系统辨识任务表述为一个正则化最小二乘问题：对每个线性时不变系统，最小化其最小二乘拟合误差，并加上一个能施加共同稀疏模式的正则化函数 [28], [29]。通过适当调整（通常通过交叉验证 [21]）赋予正则化函数的权重，可以在记录数据拟合与模型复杂度之间取得平衡。我们分析了所提出的多任务系统辨识方法相对于线性时不变系统真实矩阵以及相对于“预言机”的估计误差；后者表示在给定模型相容性条件 [30]下，采用（组）稀疏模型时可达到的最佳估计效果。

接下来，我们将说明如何调整所提出的多任务系统辨识方法，以考虑额外的结构相似性。

具体而言，我们提出了处理以下情况的方法：矩阵具有小异质性（即矩阵差较小，如 [23] 中所示），以及某些线性时不变系统的矩阵可以表示为彼此的线性组合。在这种情况下，我们采用惩罚较大矩阵偏差的正则化函数，以及受核范数最小化启发的函数 [31]–[33]。

我们通过以下两类数据展示所提出的多任务系统辨识方法的有效性：(i) 具有结构相似性的合成线性时不变系统；(ii) 来自人类连接组计划（HCP）的真实数据，其中血氧水平依赖（BOLD）信号取自静息态功能磁共振成像（fMRI）[34], [35]。对于后一类数据，我们表明，只需假设不同受试者的脑区分割具有相似的潜在功能或结构连接，所提出的方法便能以显著更少的功能性磁共振成像扫描次数达到与最小二乘法相同的误差。我们还考虑了某位受试者仅有少量功能性磁共振成像读数的情况，并展示了从其他受试者的动态中“传递信息”的能力。

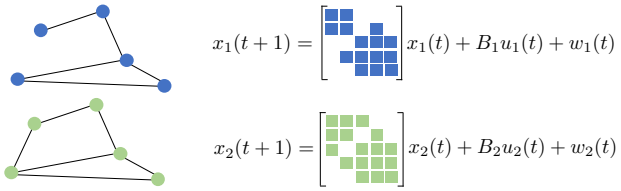


图 1. 具有相似连通性的网络系统示例。两个矩阵 A_1 和 A_2 在一组共同的元素位置上均为零。

II. 预备知识与系统辨识设置

考虑 N 线性时不变系统 S[1]

$$x_i(t+1) = A_i x_i(t) + B_i u_i(t) + w_i(t), \quad x_i(0) \in \mathbb{R}^n, \quad (1)$$

其中，$i \in [N]$ 为系统索引，$t \in \mathbb{N}$ 为时间索引，$A_i \in \mathbb{R}^{n\times n}$ 和 $B_i \in \mathbb{R}^{n\times p}$，且 $x_i(t) \in \mathbb{R}^n$、$u_i(t) \in \mathbb{R}^p$和 $w_i(t) \in \mathbb{R}^n$ 分别为第 i个系统的状态、输入和过程噪声。假设对于每个系统，输入 $u_i(t)$ 和状态 $x_i(t)$ 均可测量，且 $B_i \in \mathbb{R}^{n\times p}$ 已知；另一方面，系统矩阵未知，扰动 $w_i(t)$ 无法测量[2]。

对于第 i 个系统，假设可以获取一条轨迹 $\{x_i(\tau), u_i(\tau)\}_{\tau=1}^{P_i+1}$，其中 $P_i \in \mathbb{N}$ 为某个值，记录了状态和输入。利用这些测量数据，可以通过以下最小二乘准则估计系统矩阵：

$$\min_{A_i \in \mathbb{R}^{n\times n}} \mathcal{L}_i(A_i), \quad (2)$$

其中 $\mathcal{L}_i(A_i) := \sum_{\tau=1}^{P_i} \|x_i(\tau+1) - A_i x_i(\tau) - B_i u_i(\tau)\|_2^2$；对每个 N 个系统，分别求解最小二乘问题 (2)

[1]**记号**：我们分别用 $\mathbb{N}$ 和 $\mathbb{R}$ 表示自然数集和实数集，并定义 $[n] = \{1, 2, \ldots, n\}$。给定 $S \subset$ ，$|S|$ 是 S 的基数。我们用 $^\top$ 表示转置。对于给定的列向量 $x \in \mathbb{R}^n$，$\|x\|_2$ 是欧几里得范数，$\|x\|_1$ 表示ℓ_1 范数；对于一个矩阵 $X \in \mathbb{R}^{n\times m}$, $\|X\|_F$表示 Frobenius 范数，$\|X\|_*$ 表示核范数。此外，$(x)_i$ 表示向量 x的第 i 个元素，$(X)_{ij}$ 表示矩阵 X的第 (i, j) 个元素，而 vec(X) 是一个按列堆叠 X各列的 $mn \times 1$ 向量。给定可微函数 $f\colon \mathbb{R}^n \to \mathbb{R}$,$\nabla f$ x f x ( ) 表示其在处的梯度（视为列向量）。给定闭凸集 $C \subseteq \mathbb{R}^n$，proj$_C\colon \mathbb{R}^n \to \mathbb{R}^n$ 表示 y在 C上的欧几里得投影，即 proj$_C(y) := \arg\min_{v\in C} \|y - v\|$。给定下半连续凸函数 $g\colon \mathbb{R}^n \to \mathbb{R}$，近端算子定义为 prox$^\lambda_g(y) := \arg\min_{x\in\mathbb{R}^n} g(x) + \frac{1}{2\lambda}\|x - y\|_2^2$。

[2]需要指出的是，尽管为简化符号并使论述更加简洁，我们重点关注矩阵 A_i 的估计，但本文提出的技术也可直接应用于根据数据同时估计 A_i 和 B_i 的情形。

时的情形。最小二乘估计量 (2) 已得到文献的广泛研究，尤其是在记录的数据使最小二乘 (2) 具有良好条件数的情况下 [11]–[14], [19]。

本文关注这样一种情况：对于某些 N 线性时不变系统，最小二乘问题 (2) 是病态的；这可能是由于记录轨迹不够丰富，或仅仅是长度不足。在这种情况下，本文要探讨的问题是：即使某些系统的最小二乘问题是病态的，是否仍能利用这些 N 系统之间的“相似性”，准确估计系统矩阵。我们考虑线性时不变系统之间的以下结构相似性：

(s1) 矩阵 $A_1, \ldots A_N$在相同元素位置上为零；即，对于某些元素位置 (i, j)，$(A_1)_{ij} = (A_2)_{ij} = \cdots = (A_N)_{ij} = 0$。(s2) 对于任意一对 A_i, A_j、$i, j \in [N]$，存在 $\epsilon > 0$，使得 $\|A_i - A_j\|_F^2 \le \epsilon$。(s3) 对于系统子集 $i \in \mathcal{C}$，$\mathcal{C} \subseteq [N]$，存在 $\{\alpha_{i,j} \in \mathbb{R}\}$，使得 $A_i = \sum_{j=1,\ j\neq i}^{N} \alpha_{ij} A_j$。

我们注意到，(s1) 自然地对描述具有相似连通性的网络化系统动态的线性时不变系统进行建模；一个具体例子是，不同个体的脑网络具有相似的功能连接或结构连接 [34], [35]，由此便得到 (s1)。类似地，(s2) 对矩阵差 $A_i - A_j$ 的范数较小的情形进行建模；这与 [22], [23]中考虑的模型一致。最后，(s3) 对这样的情形进行建模：第 i 个系统的矩阵 A_i 可以表示为其他若干矩阵 $\{A_j\}_{j=1,\ j\neq i}^N$的线性组合；例如，这一模型可能适用于交通流和按需出行服务（参见例如 [36]），其中线性时不变系统 (1) 对给定时段内、特定地理区域的车辆密度变化进行建模。

给定模型 (s1)-(s3)，我们考虑通过求解以下系统辨识问题，联合估计矩阵 $\{A_i\}_{i\in[N]}$：

$$\min_{\{A_k\}_{k\in[N]}} \sum_{i=1}^{N} \mathcal{L}_i(A_i) + \lambda \mathcal{R}(A_1, \ldots, A_N), \quad (3)$$

其中我们回顾一下，$\mathcal{L}_i(A_i)$是第 i个系统的最小二乘拟合；按照正则化回归方法的思路，[2],[20], [27], $(A_1, \ldots\ldots, A_N) \mapsto \mathcal{R}(A_1, \ldots\ldots, A_N)$是一个下半连续凸函数，用于促进由(s1)–(s3)指定的先验，而 $\lambda > 0$是一个调节参数。问题(3)受到多任务学习方法 [24]–[26],[37],的启发，其中各项学习任务同时进行（在

我们的情形中，即最小二乘拟合），同时利用所学习参数之间的共性 [26, 定义 1]。在接下来的章节中，我们将说明正则化函数 $\mathcal{R}(A_1, \ldots, A_N)$的具体选择；我们首先讨论线性时不变系统具有共同连通性的情况。

III. 具有相似连通性的系统

A. 多任务系统辨识

本节考虑系统矩阵 $A_1, \ldots A_N$在一组共同位置上为零的情形；即，对于索引子集 $i \in [n]$ 和 $j \in [n]$，有 $(A_1)_{ij} = (A_2)_{ij} = \cdots = (A_N)_{ij} = 0$。在统计学习文献中，这种结构相似性带来了一种设定，其中未知量 $\{A_1, \ldots, A_N\}$ 具有组稀疏性 [28],[29]。借鉴 [28], [29], 的技术方法，我们进而将具有相似连通性的线性时不变系统的多任务系统辨识问题表述如下：

$$\min_{\{A_k\}_{k\in[N]}} \sum_{i=1}^{N}\sum_{\tau=1}^{P_i} \|x_i(\tau+1) - A_i x_i(\tau) - B_i u_i(\tau)\|_2^2 + \lambda \sum_{i=1}^{N}\sum_{j=1}^{N} \|[(A_1)_{ij}, (A_2)_{ij}, \ldots, (A_N)_{ij}]\|_2, \quad (4)$$

其中 $\lambda \geq 0$，函数 $\mathcal{R}(A_1, \cdots, A_N) = \sum_{i=1}^{N}\sum_{j=1}^{N} \|[A_{1\ ij}\ \ A_{2\ ij}\ \cdots\ \ A_{N\ ij}]\|_2$ 用于实现组稀疏性；也就是说，式 (4) 的解使得整个向量 $[(A_1)_{ij}, (A_2)_{ij}, \ldots, (A_N)_{ij}]$ 要么为零，要么不为零。式 (3) 中的参数 λ 在最小二乘拟合（在本例中，即对各个线性时不变系统进行最小二乘拟合）与设为零的向量数量 $[(A_1)_{ij}, (A_2)_{ij}, \ldots, (A_N)_{ij}]$（稍后将说明）之间进行权衡。

问题（4）是一个无约束凸规划；鉴于该复合代价函数，我们考虑采用近端梯度法（结合线搜索）求解（4）（例如参见 [38], [39]）。近端梯度法列于算法 1中。

算法1适用于连通性相似系统的近端梯度法

给定：$\hat{A}_1^{(0)}, \cdots, \hat{A}_N^{(0)}, \eta^{(0)}, \beta \in (0,1)$, 和 λ。

重复：$m = 0, 1, 2, \ldots$ 直至收敛

[S1] $\alpha \leftarrow \eta^{(m)}$。

[S2] 带线搜索的近端梯度：

[S2。1] $Z_i = \hat{A}_i^{(m)} - \alpha \nabla \mathcal{L}_i(\hat{A}_i^{(m)})$, $i \in [N]$

[S2.2] 令 $z_{jk} := [(Z_1)_{jk}, (Z_2)_{jk}, \ldots, (Z_N)_{jk}]$。

对于所有 $j, k \in [n]$，计算：

$$y_{jk} = \frac{z_{jk}}{\|z_{jk}\|_2} \max(\|z_{jk}\|_2 - \alpha\lambda,\ 0)。 \quad (5)$$

[S2.3] Form matrices $\{Y_\ell\}_{\ell\in[N]}$ as $(Y_\ell)_{jk} = (y_{jk})_\ell$.

[S2.4] Break if: $\sum_{i=1}^{N} \mathcal{L}_i(Y_i) \leq \frac{1}{2\lambda}\|Y_i - \hat{A}_i^{(m)}\|_F^2 + \sum_{i=1}^{N}\left(\mathcal{L}_i(\hat{A}_i^{(m)}) + \nabla\mathcal{L}_i(\hat{A}_i^{(m)})^\top (Y_i - \hat{A}_i^{(m)})\right)$

[S2.5] Update $\alpha \leftarrow \beta\alpha$.

[S3] $\eta^{(m+1)} \leftarrow \alpha$, $\hat{A}_i^{(m+1)} \leftarrow Y_i$, $i \in [N]$.

我们注意到，步骤 [S2.3] 实际上是近端映射的闭式表达式：

$$\{Y_i\}_{i\in[N]} = \text{prox}_{\alpha\lambda\mathcal{R}}(\{Z_i\}_{i\in[N]}), \quad (6)$$

并且它涉及与（5）中相同的 n^2 并行计算。利用 [38,第 2]章中的结果，可以轻松分析算法1的收敛行为。此外，如果不执行线搜索 [40]，算法1可以转换为“标准”的近端梯度法。

根据阈值操作（5），显然，增大 λ 会使更多系统矩阵中的元素变为零。遗憾的是，调整 λ 并非易事，通常会采用交叉验证程序来寻找 λ 的取值，使估计矩阵在测试数据上的误差最低；例如参见 [20], [21]。

B. 分析

本节分析多任务系统辨识方法的性能。对于带有稀疏性正则化项的回归问题，其性能通常会与一个“预言机”进行比较；在给定模型相容性条件 [30]的情况下，该预言机代表采用（组）稀疏模型时可达到的最佳估计。我们将分别给出相对于预言机和真实系统矩阵的误差界。

为简化符号表示，我们概述 $P_i = P$ 对所有 $i \in [N]$ 均成立的情形下的结果（当 $\{P_i\}$ 不同时，也有类似结果）。令 $A^S \in \mathbb{R}^{n\times n}$ 为一个矩阵，它通过选取 $A \in \mathbb{R}^{n\times n}$ 中由 $S \subset [n]\times[n]$ 索引的元素构成，其余元素置零。对于 N 矩阵 $A_i \in \mathbb{R}^{n\times n}$、$i \in [N]$，为简便起见，定义 $\|\{A_i\}_{i\in[N]}\|_{2,1} = \sum_{i=1}^{n}\sum_{j=1}^{n} \|[A_{1\ ij}\ \ A_{2\ ij}\ \cdots\ \ A_{N\ ij}]\|_2$。有了这一记号后，我们首先陈述主要假设，并概述与组稀疏模型 [30,Chapter 8]相关的相容性条件。

假设*1*：扰动 $\{w_i(k)\}$ 是独立同分布的高斯随机变量 $\mathcal{N}(0, \sigma^2)$。□

定义*1*（*[30]*）：设 $S \subset [n]\times[n]$ 和 $S^c := ([n]\times[n]) \setminus S$。我们称索引集 S 满足相容性条件，如果对于任意满足 $\|\{A_i^{S^c}\}_{i\in[N]}\|_{2,1} \leq 3\|\{A_i^S\}_{i\in[N]}\|_{2,1}$ 的 $\{A_i \in \mathbb{R}^{n\times n}\}_{i\in[N]}$，都有

$$\|\{A_i^S\}_{i\in[N]}\|_{2,1}^2 \leq \frac{|S|\sum_{i=1}^{N}\sum_{\tau=1}^{P}\|A_i x_i(\tau)\|_2^2}{P\phi(S)} \quad (7)$$

对于某个常数 $\phi(S) > 0$成立。此外，我们将满足相容性条件的集合 S 的全体定义为 $\mathcal{S}$。□

注意，当 S 与矩阵 $\{A_i\}_{i\in[n]}$的支撑集重合时，这一技术条件为矩阵的非零元素取值提供了界。

下文中，令 $\{A_i^{MT}\}_{i\in[N]}$ 为系统辨识问题 (4) 的一个最优解，并将 (1) 中线性时不变系统的真实矩阵记为$\{A_i^\star\}_{i\in[N]}$。以下定理给出了所提出的多任务系统辨识方法 (4) 相对于（真实）系统矩阵 $\{A_i^\star\}_{i\in[N]}$的误差界。

定理 *1*：令 $\{A_i^\star\}_{i\in[N]}$ 为线性时不变系统的真实矩阵，并定义 $S^\star = \{(i,j)\,[(A_1^\star)_{ij},\ (A_2^\star)_{ij},\ \ldots\ldots,\ (A_N^\star)_{ij}] \neq 0\}$。设假设1成立，并假设 $S^\star \in \mathcal{S}$。定义

$$\lambda_0 := \frac{2\sqrt{M}}{nP}\left(1+\sqrt{\frac{4\gamma+8\log n}{N}}+\frac{4\gamma+8\log n}{N}\right)^{\frac{1}{2}} \tag{8}$$

对于某个 $\gamma > 0$，$M = \max_{1\le i\le N,\ 1\le j\le n} \sum_{k=1}^{P} [x_k^{(j)}]^2$，$\lambda \ge 4nNP\lambda_0$ 和 σ^2 的下界成立

$$\sum_{i=1}^{N}\sum_{\tau=1}^{P}\|(A_i^\star - A_i^{MT})x_i(\tau)\|_2^2 \le \frac{24\lambda^2|S^\star|}{PN\phi(S^\star)} \tag{9}$$

以至少 $1-\mathrm{e}^{-\gamma}$ 的概率。 □

根据定理 1，可以看出平均误差 $E(N,\ P) := \frac{1}{PN}\sum_{i=1}^{N}\sum_{\tau=1}^{P}\|(A_i^\star - A_i^{MT})x_i(\tau)\|_2^2$ 的阶为

$$E(N,P) \approx O\left(\frac{|S^\star|}{\phi(S^\star)}P^{-1}(1+N^{-\frac{1}{2}}+N^{-1})\right),$$

当取 $\lambda = O(\sqrt{P}\sqrt{N}+\sqrt{N}+1)$ 时。有趣的是，随着多任务系统辨识中的线性时不变系统数量增加（即 N 增加），平均误差 $E(N,\ P)$ 会减小；然而，当 $N\to\infty$ 时，误差不会趋于 0。这可以理解为系统之间“传递信息”能力的一个平台期。

定理 1 可能存在的一个局限是，假设用于描述矩阵 $\{A_i^\star\}_{i\in[N]}$（的共同）支撑集的集合 $S^\star$ 满足相容性条件。下文将针对 $\{A_i^\star\}_{i\in[N]}$ 不一定满足相容性条件的情形，给出一个额外的误差界；该误差界利用了预言机 [30]的概念。

定理 *2*：考虑定义如下的预言机 $\{A_i^\dagger\}_{i\in[N]}$：

$$\{A_i^\dagger\}_{i\in[N]} \in \arg\min_{\{A_i\}:S_{\{A_i\}}\in\mathcal{S}}\left\{\sum_{i=1}^{N}\sum_{\tau=1}^{P}\|(A_i^\star - A_i)x_i(\tau)\|_2^2 + \frac{4\lambda^2|S_{\{A_i\}}|}{PN\phi\left(S_{\{A_i\}}\right)}\right\}, \tag{10}$$

其中 $S_{\{A_i\}} = \{(i,j)\ |\ [(A_1)_{ij},\ (A_2)_{ij},\ \ldots\ldots,\ (A_N)_{ij}] \neq 0\}$，并设假设1成立。令 λ_0 和 λ 与定理 1 中的定义相同。则对于给定的 $\gamma>0$，以下界以至少 $1-\mathrm{e}^{-\gamma}$ 的概率成立：

$$\sum_{i=1}^{N}\sum_{\tau=1}^{P}\|(A_i^\star - A_i^{MT})x_i(\tau)\|_2^2 \le 6\sum_{i=1}^{N}\sum_{\tau=1}^{P}\|(A_i^\star - A_i^\dagger)x_i(\tau)\|_2^2 + \frac{24\lambda^2|S^\dagger|}{PN\phi(S^\dagger)}, \tag{11}$$

其中 $S^\dagger = \{(i,j)\ |\ [(A_1^\dagger)_{ij}, (A_2^\dagger)_{ij}, \ldots, (A_N^\dagger)_{ij}] \neq 0\}$。 □

定理 2 表明，多任务系统辨识 (4) 所产生的误差有界于预言机对应的估计误差，再加上一个用于刻画估计矩阵与预言机之间误差的额外项。这里，预言机表示在考虑支撑集索引集合满足相容性条件的矩阵时所能达到的最佳估计。

IV. 处理具有其他相似性的系统

本节将说明如何将所提出的多任务系统辨识方法应用于具有其他相似性的线性时不变系统，正如第 II 节中所述。

A. 小异质性

考虑这样一种情况：对于任意一对 $A_i,\ A_j,\ i,\ j\in[N]$，都存在 $\epsilon>0$ 使得 $\|A_i - A_j\|_F^2 \le \epsilon$。在 [23] 中，这种情况被称为“小异质性”。基于这一先验，多任务系统辨识问题可表述为

$$\min_{\{A_k\}_{k\in[N]}}\sum_{i=1}^{N}\mathcal{L}_i(A_i) + \lambda\sum_{i=1}^{N}\sum_{j=i}^{N}\|A_i - A_j\|_F^2, \tag{12}$$

这里回顾一下，$\mathcal{L}_i(A_i)$ 是第 i 个系统的最小二乘拟合，而正则化项用于惩罚估计矩阵 [20] 之间的较大偏差。

问题 (12) 是凸问题，可以通过闭式解求解。不过，为避免计算量大的矩阵求逆（尤其是在维度 b 较大且系统辨识过程中考虑多个系统时），可以使用算法1求解 (12)，方法是将 [S2.2] 替换为以下并行计算：

$$y_{ij} = \frac{z_{ij} + 2\alpha\lambda s_{ij}[1,1,\cdots,1]}{2\alpha\lambda N+1}, \quad i,j\in[N], \tag{13}$$

其中 $s_{ij} = \sum_{\ell=1}^{N}(Z_\ell)_{ij}$。

从更新式 (13) 可以看出，$y_{ij} \to (s_{ij}/N)[1,1,\cdots,1]$ 当 $\lambda\to\infty$ 时成立，因此将所有系统矩阵设为相同。另一方面，设置$\lambda = 0$ 即可得到最小二乘法。即使在这种情况下，也可以利用交叉验证过程找到 λ 的取值，使得估计矩阵在测试数据 [21]上的误差最低。

可以通过推导 (12) 的最优解与最小二乘法解之间的界，来分析 (12) 的估计性能。由于该界的推导很直接，且受篇幅限制，我们在论文中省略这一结果。

B. 线性组合

最后，我们讨论一种额外的“相似性”：系统矩阵是（近似）线性相关的。具体而言，对于一部分系统，我们考虑存在系数 $\{\alpha_{i,j}\in\mathbb{R}\}$，满足 $\alpha_{i,\ j}\neq 0$，其中某个 $j\in[N]$ 使得 $A_i \approx \sum_{j=1,\ j\neq i}^{N}\alpha_{ij}A_j$。

为使该设置形式化，假设 $q \ll N$ 个矩阵 $\{A_i\}_{i\in[N]}$ 满足：其余的 $N-q$ 可以表示为这些q 矩阵的线性组合。于是考虑构造 $n^2\times N$ 矩阵[vec(A_1), vec(A_2).....，vec(A_N)]；由此可知，该矩阵的秩为 $q\ll N$。基于这一观察，我们提出

针对这种情况，将多任务系统辨识问题表述为

$$\min_{\{A_k\}_{k\in[N]}} \sum_{i=1}^{N} \mathcal{L}_i(A_i) + \lambda \left\| [\mathrm{vec}(A_1), \mathrm{vec}(A_2), \ldots, \mathrm{vec}(A_N)] \right\|_*,$$

其中，正则化函数会促使矩阵[vec(A_1)、vec(A_2).....vec(A_N)]的奇异值具有稀疏性（例如，参见 [31], [32]）。

问题（14）是凸问题，其代价函数为复合代价函数，且正则化函数不可微 [31]‚[32]。不过，可以对算法1中列出的近端梯度算法进行修改，以求解（14）。具体而言，可以将算法1中的步骤 [S2.2]替换为 prox$_{\alpha\lambda\|[\mathrm{vec}(A_1),\ \mathrm{vec}(A_2).....,\ \mathrm{vec}(A_N)]\|_*}(\{Z_i\}_{i\in[N]})$；该近端映射可得到如下闭式解：

$$\bar{Y} = U \mathrm{diag}(\{\max\{\sigma_i - \alpha\lambda, 0\}\}) V^*, \qquad (14)$$

其中，[vec(A_1)、vec(A_2).....vec(A_N)]的奇异值分解为 $U\mathrm{diag}(\{\sigma_i\})V$。随后，从 $\bar{Y}$的各列中提取矩阵 $\{Y_\ell\}_{\ell\in[N]}$。从 Y的计算过程可以看出， λ 的值越大， Y中被置为零的奇异值数量就越多。

虽然文献中已通过数值实验验证了核范数最小化方法的有效性，但只有在对回归量 [31], [32] 作出特定假设的情况下，才有可用的可辨识性结果和估计误差解析界，而这些假设可能不适用于 (14)；另见 [33]。推导 (14) 的解析误差界是我们正在进行的研究工作。

注 *1*： 所提出的多任务系统辨识问题可扩展到系统矩阵 $\{A_i\}_{i\in[N]}$ 根据先验 (s1)–(s2) 中不止一种先验而彼此相似的情形。例如，若矩阵具有共同稀疏模式，且非零元素之间的差异较小，则可采用复合正则化函数 $\lambda_1 \sum_{i=1}^{N}\sum_{j=1}^{N} \|[\ A_1\ _{ij}\ \ A_2\ _{ij}\ \ldots\ \ A_N\ _{ij}]^\top\|_2 + \lambda_2 \sum_{i=1}^{N}\sum_{j=1}^{N} \|[\ A_1\ _{ij}\ \ A_2\ _{ij}\ \ldots\ \ A_N\ _{ij}]^\top\|_2$ ( ),( ),,( )，其中λ_1, $\lambda_2 \geq 0$ 为调节参数。□

V. 数值实验

A. 脑网络实验

我们使用来自人类连接组计划（HCP）的静息态功能磁共振成像数据来测试所提出的方法，以估计脑网络的动态特性[3] [34], [35], [41]。在这里， $x_i(t)$ 是第 i 个受试者大脑的 116 个脑区分割所对应的 116 维血氧水平依赖（BOLD）时间序列。我们的目标是估计形如 $x_i(t+1) = A_i x_i(t) + w_i(t)$ 的 $N = 5$ 动力系统，用于描述个体处于静息状态时 BOLD信号的演化，其中 $w_i(t)$ 表示过程噪声（由于静息状态条件，该模型不包含外部输入 u_i）。

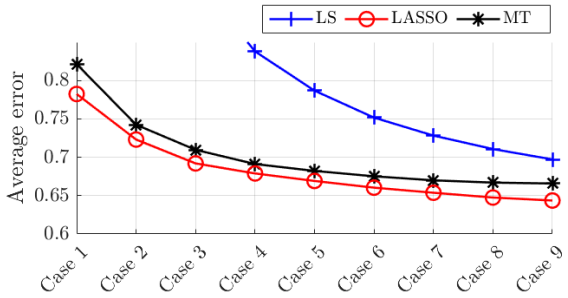
[3]数据可在 https://wiki.humanconnectome.org/ 获取

由于矩阵 $\{A_i\}_{i\in[5]}$ 未知，我们针对每个系统考虑以下误差：

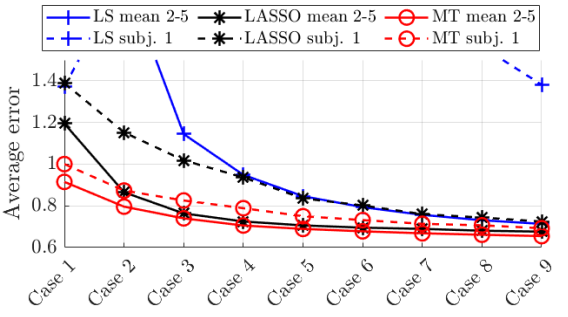
$$\mathcal{E}(A) := \frac{1}{n}\sum_{k=1}^{n} \frac{\sum_{i=1}^{p}(x_i(k) - [Ax_i](k))^2}{\sum_{i=1}^{p}(x_i(k) - \bar{x}(k))^2},$$

其中， n 是测试向量的长度， p 是测试数据的数量，而 $\bar{x}(k) := \frac{1}{p}\sum_{i=1}^{p} x_i(k)$。注意， $1 - \mathcal{E}(A)$ 正是 [35] 的平均 R^2 指标。

我们考虑三种不同的方法：（i）最小二乘估计器（2），对每个个体分别使用；（ii）最小绝对收缩与选择算子（LASSO），同样按照 [35]中的方法对每个个体分别使用；以及（iii）采用组稀疏正则化函数（4）的所提方法（3）。采用组稀疏性的依据是，脑动力学在不同脑区分割之间应呈现相同的有效连接，而其余条目则体现不同个体间交互作用强度的差异。我们注意到，最小二乘和 LASSO 的有效性已通过实验验证；在 [35],中，它们的估计精度与若干种识别方法进行了比较。此外，我们还进行了交叉验证，以优化 LASSO 的性能。



(a)



(b)

图 2。（a）LS、LASSO 和多任务（MT）系统辨识的平均误差；“情况 k”表示每个受试者有 $100k$ 训练数据点可用（$k = 1, 2, \cdots, 9$）。（b）受试者 2–5 的平均误差和受试者 1 的误差。“情况 k”表示受试者 1 使用 $25k$ 功能性磁共振成像扫描（虚线），而受试者 2–5 使用 $100k$（实线）扫描。

在图 2 中，我们比较了最小二乘、LASSO 和我们的方法（标记为“MT”）在两种情况下的表现：(a) 五名受试者使用相同数量的训练数据；(b) 对于受试者 1，我们使用的训练数据点仅为其他受试者 2–5 所用数量的 25%。我们使用了 100 个测试

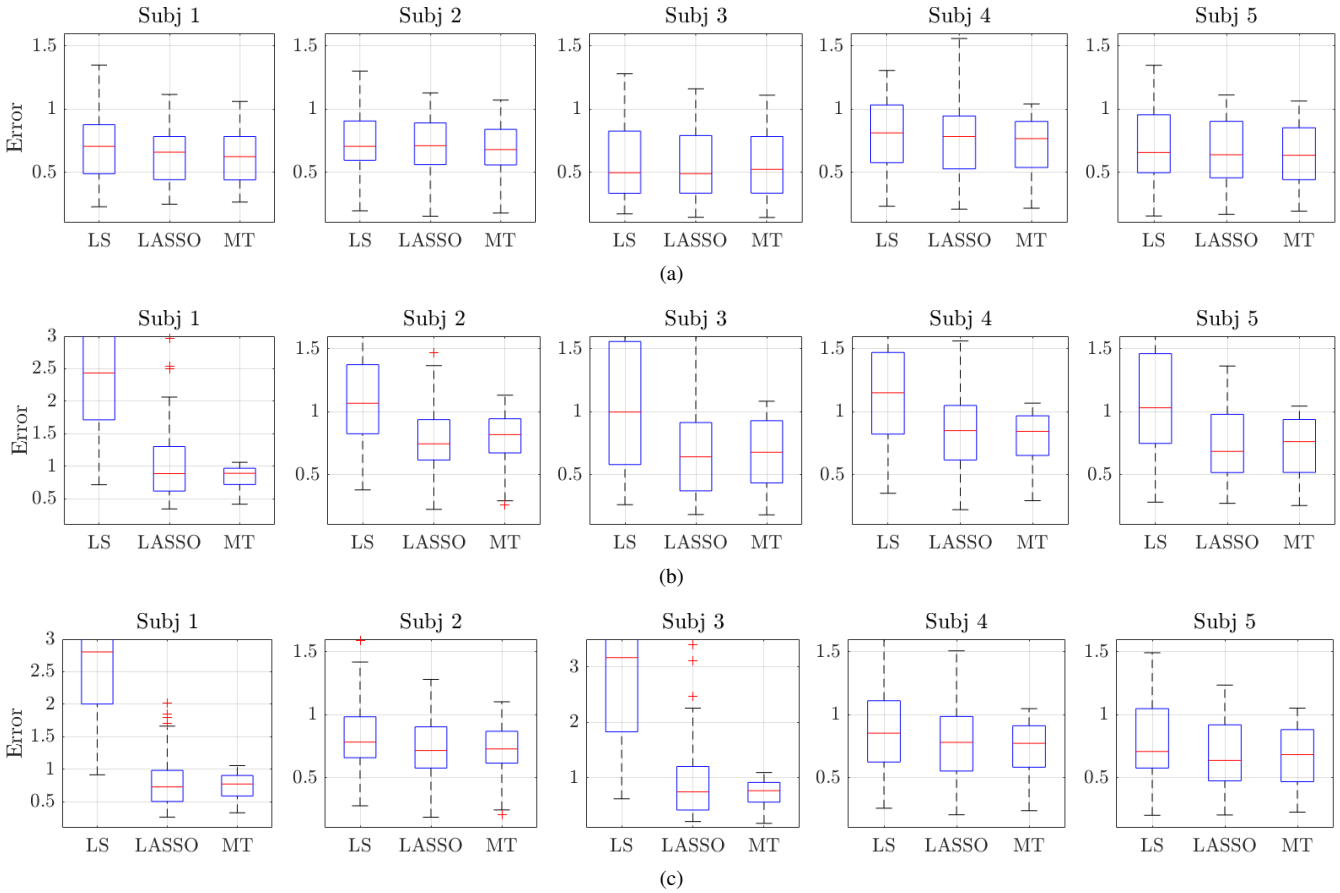


图3. 最小二乘、LASSO 和多任务（MT）系统辨识的比较 (a) 案例1：每位受试者有 900 个训练数据点。(b) 案例2：受试者1有 75 个训练数据点，受试者2–5各有 300 个。(c) 案例3：受试者1和3各有 150 个训练数据点，受试者2、4和5各有 600 个。在箱线图中，红色中心线、箱体边界和须线分别表示中位数、上四分位数和下四分位数，以及最小和最大样本。红色十字表示离群值。

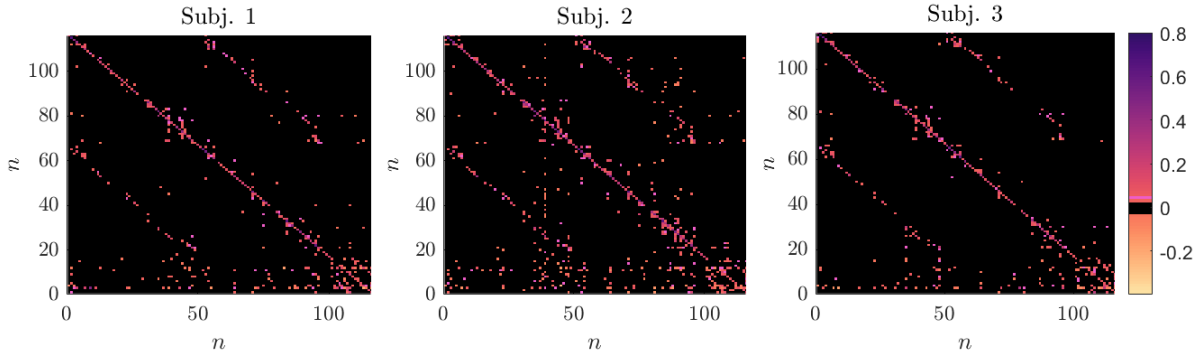


图4. 案例3中个体1、2和3在 $n = 116$ 脑分区下的估计矩阵 $\hat{} A_i$ 。

点。在图 2(a) 中，我们绘制了受试者 1–5 的平均误差；在图 2(b) 中，我们绘制了受试者 2–5 的平均误差，以及受试者 1 的误差；受试者 1 可用的功能性磁共振成像读数较少。在功能性磁共振成像扫描数量相当的情况下，所提出的方法在两种情形下均优于最小二乘和 LASSO。图 2(b) 尤其凸显了所提出方法的优势：对于受试者 1，该方法显著优于 LASSO；另一方面，最小二乘法存在病态问题，无法得出有意义的估计值。这表明该方法能够利用信息和数据（本例中为功能性磁共振成像读数）利用受试者2–5的动态特性，辅助估计受试者1的动态特性。

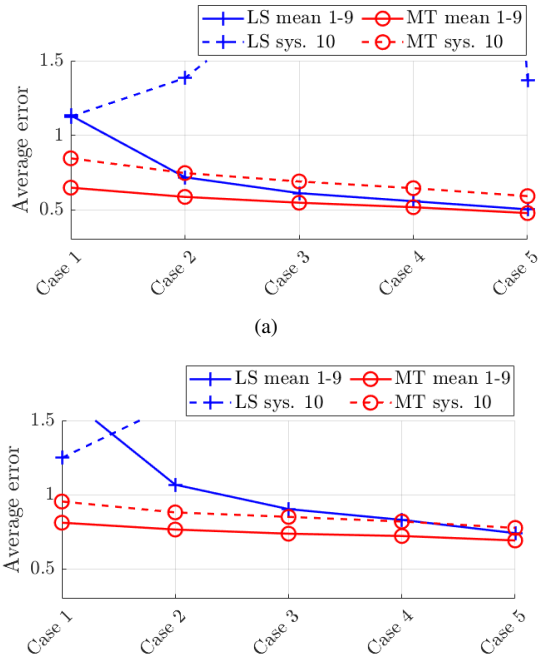
除平均误差外，为了提供更多比较，图3展示了三种不同情形下最小二乘、LASSO 和所提出方法的箱线图。具体而言，图3(a)表明，当每位受试者使用900个时间步的轨迹（且这些训练轨迹足够丰富）时，所提出的多任务辨识方法能够达到比最小二乘或LASSO更小或相当的误差（平均而言）。图3(b)考虑的是受试者1有75个训练数据点可用的情形

而受试者2至5各有300个。此时，由于病态，最小二乘的表现不佳。就中位数而言，LASSO的表现与所提出的方法相当；不过，所提出方法的上四分位数和下四分位数更小。此外，图3(c)考虑的是受试者1和3可用的功能性磁共振成像读数更少的情形；所提出的方法在四分位数方面优于LASSO，并且在不同脑区分割之间的误差偏差显著更小。

最后，图4给出了受试者1–3的估计矩阵$\hat{}A_i$的一个代表性示例。这些估计矩阵是在图3(c)所考虑的情形下获得的，其中受试者1和3的训练点较少。可以看到，这三个矩阵有许多相同位置的元素为零。基于这一结果，我们将探索其他正则化方法，将组稀疏性与（逐元素）稀疏性结合起来。

B. 合成数据上的实验

我们给出合成数据上的其他结果。我们考虑式(1)中的10个系统，其中$\{A_i\}_{i\in[10]} \in \mathbb{R}^{50\times 50}$、$\{B_i\}_{i\in[10]} \in \mathbb{R}^{50\times 4}$ u_i t $\mathbb{R}^4$ 、( )是 中全1向量，即$u_i(t)$是常向量，且 $w_i(t) \sim \mathcal{N}(0,\ 0.1_2)$。我们考虑两种不同的情形：共同稀疏性和线性组合。我们比较最小二乘估计量(2)、所提出的方法(3)、组稀疏正则化(4)以及核范数正则化(14)。



）图 5. 比较最小二乘和多任务（MT）系统辨识方法的平均误差曲线。(a) 线性组合。(b) 共同稀疏性。“情况k”表示，使用轨迹中的 $10k + 10$ 个样本进行系统 10 的辨识（虚线），而使用 $40k + 40k$ 1 9 k = 个样本进行系统 - 的辨识（实线）1 2 3 4 5 75 10 ，，，，。在“情况”中，使用数据点进行系统的辨识（虚线），而使用 $300k$ 个样本进行系统 1-9 的辨识（实线）。

图5比较了最小二乘方法与我们的方法在两种情况下的表现：(a) 所有10个系统都可以表示为3个系统的线性组合，并且与其他系统1–9相比，第10个系统仅有25%的训练数据点可用；(b) 所有10个系统具有相同的稀疏模式，并且与其他系统1–9相比，第10个系统仅有25%的训练数据点可用。测试采用60个数据点。在图5(a)中，我们绘制了系统1–9的平均误差以及系统10的误差。所提出的方法在平均误差和系统10的误差方面均优于最小二乘方法，尤其是在可用数据量较少时。在图5(b)中，我们观察到类似的结果。

参考文献

[1] L. Ljung，《系统辨识：理论与实践》。普伦蒂斯-霍尔，1987年。[2] G. Pillonetto、T. Chen、A. Chiuso、G. De Nicolao 和 L. Ljung，“正则化系统辨识：从数据中学习动态模型”，2022年。[3] C. De Persis 和 P. Tesi，“数据驱动控制公式：稳定化、最优性与鲁棒性”，《IEEE 自动控制汇刊》，第65卷，第3期，第909–924页，2019年。[4] J. Coulson、J. Lygeros 和 F. D¨orfler，“数据赋能预测控制：探索 DeePC 的浅层领域”，载于《欧洲控制会议》，2019年，第307–312页。[5] L. Hewing、K. P. Wabersich、M. Menner 和 M. N. Zeilinger，“基于学习的模型预测控制：迈向控制领域的安全学习”，《控制、机器人与自主系统年度评论》，第3卷，第269–296页，2020年。[6] J. Berberich、J. K¨ohler、M. A. M¨uller 和 F. Allg¨ower，“具有稳定性与鲁棒性保证的数据驱动模型预测控制”，《IEEE 自动控制汇刊》，第66卷，第4期，第1702–1717页，2020年。

[7] V. Krishnan 和 F. Pasqualetti，《论直接与间接数据驱动预测控制》，载于*IEEE* 决策与控制会议，2021年，第 736–741 页。[8] L. Li、C. De Persis、P. Tesi 和 N. Monshizadeh，《线性系统中的基于数据的迁移稳定化》，*arXiv* 预印本 *arXiv:2211 05536*，2022。。[9] J. C. Willems、P. Rapisarda、I. Markovsky 和 B. L. De Moor，《关于持续激励的一则短札》，《系统与控制快报》，第 54 卷，第 4 期，第 325–329 页，2005年。[10] C. De Persis 和 P. Tesi，《论持续激励与数据驱动控制公式》，载于*IEEE* 决策与控制会议，2019年，第 873–878 页。[11] T. Sarkar 和 A. Rakhlin，《任意线性动力系统的近最优有限时间辨识》，载于国际机器学习会议，2019年，第 5610–5618 页。[12] M. Simchowitz、H. Mania、S. Tu、M. I. Jordan 和 B. Recht，《无需混合：迈向线性系统辨识的精确分析》，载于学习理论会议，2018年，第 439–473 页。[13] M. K. S. Faradonbeh、A. Tewari 和 G. Michailidis，《不稳定线性系统的有限时间辨识》，《自动化》，第 96 卷，第 342–353 页，2018年。[14] S. Oymak 和 N. Ozay，《从单条轨迹对线性时不变系统进行非渐近辨识》，载于美国控制会议，2019年，第 5655–5661 页。[15] Y. Zheng 和 N. Li，《利用多条轨迹对线性动力系统进行非渐近辨识》，《*IEEE* 控制系统快报》，第 5 卷，第 5 期，第 1693–1698 页，2020年。[16] L. Xin、G. Chiu 和 S. Sundaram，《从多条轨迹中学习自主线性系统的动态》，载于美国控制会议，2022年。[17] T. Chen、M. S. Andersen、L. Ljung、A. Chiuso 和 G. Pillonetto，《通过基于稀疏多核正则化的序贯凸优化技术进行系统辨识》，《*IEEE* 自动控制汇刊》，第 59 卷，第 11 期，第 2933–2945 页，2014年。[18] A. Chiuso 和 G. Pillonetto，《系统辨识：机器学习视角》，《控制、机器人与自主系统年度评论》，第 2 卷，第 281–304 页，2019年。[19] Y. Sun、S. Oymak 和 M. Fazel，《有限样本系统辨识：

最优速率与正则化的作用》，载于《动力学与控制学习》，2020年，第 16–25 页。[20] T. Hastie、R. Tibshirani、J. H. Friedman 和 J. H. Friedman，《统计学习基础：数据挖掘、推断与预测》。Springer，2009年，第 2 卷。[21] T. Hastie、R. Tibshirani 和 M. Wainwright，《稀疏性统计学习》，《统计与应用概率专著》，第 143 卷，第 143 页，2015年。[22] L. Xin、L. Ye、G. Chiu 和 S. Sundaram，《利用相似系统的数据识别系统动态》，载于《美国控制会议》，2022年。[23] H. Wang、L. F. Toso 和 J. Anderson，《Fedsysid：一种样本高效系统辨识的联邦方法》，arXiv 预印本 arXiv:2211.14393，2022年。[24] T. Evgeniou 和 M. Pontil，《正则化多任务学习》，载于第十届 ACM SIGKDD 国际知识发现与数据挖掘会议论文集，2004年，第 109–117 页。[25] O. Sener 和 V. Koltun，《将多任务学习视为多目标优化》，《神经信息处理系统进展》，第 31 卷，2018年。[26] Y. Zhang 和 Q. Yang，《多任务学习综述》，《IEEE知识与数据工程汇刊》，2021年。[27] S. L. Brunton、J. L. Proctor 和 J. N. Kutz，《通过稀疏辨识非线性动力系统，从数据中发现控制方程》，《美国国家科学院院刊》，第 113 卷，第 15 期，第 3932–3937 页，2016年。[28] M. Yuan 和 Y. Lin，《含分组变量的回归模型选择与估计》，《皇家统计学会会刊：B辑（统计方法学）》，第 68 卷，第 1 期，第 49–67 页，2006年。[29] J. Huang 和 T. Zhang，《组稀疏性的益处》，《统计年刊》，第 38 卷，第 4 期，第 1978–2004 页，2010年。[30] P. Bhlmann 和 S. van de Geer，《高维数据统计：方法、理论与应用》。Springer 出版公司，2011年。[31] V. Chandrasekaran、S. Sanghavi、P. A. Parrilo 和 A. S. Willsky，《稀疏与低秩矩阵分解》，《国际自动控制联合会论文集》，第 42 卷，第 10 期，第 1493–1498 页，2009年。[32] M. Mardani、G. Mateos 和 G. B. Giannakis，《流式大数据矩阵与张量的子空间学习和插补》，《IEEE 信号处理汇刊》，第 63 卷，2677，2015年 10，第 2663 页起。[33] K. Mohan 和 M. Fazel，《重加权核范数最小化及其在系统辨识中的应用》，载于《2010年美国控制会议论文集》。IEEE，2010年，第 2953–2959 页。[34] P. Srivastava、E. Nozari、J. Z. Kim、H. Ju、D. Zhou、C. Becker、F. Pasqualetti、G. J. Pappas 和 D. S. Bassett，《脑网络的通信与控制模型：区别、趋同与未来展望》，《网络神经科学》，第 4 卷，1159，2020年 4，第 1122 页起。[35] E. Nozari、M. A. Bertolero、J. Stiso、L. Caciagli、E. J. Cornblath、X. He、A. S. Mahadevan、G. J. Pappas 和 D. S. Bassett，《大脑在宏观尺度上是线性的吗？静息态动态的系统辨识》，arXiv 预印本 arXiv:2012.12351，2020年。[36] B. Turan 和 M. Alizadeh，《按需电动自主出行系统中的竞争》，IEEE网络系统控制汇刊，2021年。[37] M. Crawshaw，《深度神经网络的多任务学习：综述》，arXiv 预印本 arXiv:2009.09796，2020年。[38] A. Beck 和 M. Teboulle，《基于梯度的算法及其在信号恢复中的应用》，《信号处理与通信中的凸优化》，第 42–88 页，2009年。[39] P. L. Combettes 和 J.-C. Pesquet，《信号处理中的近端分裂方法》，载于《科学与工程逆问题的不动点算法》。Springer，2011年，第 185–212 页。[40] N. Parikh、S. Boyd 等，《近端算法》，《优化基础与趋势®》，第 1 卷，第 3 期，第 127–239 页，2014年。[41] S. Gu、F. Pasqualetti、M. Cieslak、Q. K. Telesford、A. B. Yu、A. E. Kahn、J. D. Medaglia、J. M. Vettel、M. B. Miller、S. T. Grafton 等，《结构性脑网络的可控性》，《自然·通讯》，第 6 卷，第 1 期，第 1–10 页，2015年。

致谢

作者谨此感谢 Erfan Nozari 博士（加州大学河滨分校）协助提供仿真中使用的数据，并感谢 Stephen Becker 博士（科罗拉多大学博尔德分校）就组套索进行讨论。

附录

定理*1*和定理*2*的证明。我们简要给出定理1和定理2的证明。为此，我们引入一些额外的记号。对于第 i个系统，将记录轨迹收集到矩阵 X_i 中：$= [x_i(1) \ldots, x_i(P_i)]$；那么，第 i个系统的最小二乘拟合 $\mathcal{L}_i(A_i)$可以等价地表示为

$$\mathcal{L}_i(A_i) = \|\tilde{\mathbf{y}}_i - \tilde{X}_i\tilde{\mathbf{a}}_i\|_2^2,$$

其中 $\tilde{}\mathbf{y}_i := \text{vec}([x_i(2) - B_iu_i(1) \ldots, x_i(P_i + 1) - B_iu_i(P_i)])$，$\tilde{}\mathbf{a}_i = \text{vec}(A_i)$，且 $\tilde{X}_i = X_i^\top \otimes I_n$。此外，令$\tilde{}\mathbf{A} := [\tilde{}\mathbf{a}1, ..., \tilde{}\mathbf{a}_N]$ 收集所有向量化系统矩阵 $\{A_i\}_{i\in[N]}$，$\mathbf{x}_i^{(j)}$ 表示 $\tilde{X}_i$的第 j列，并将 $\|\tilde{}\mathbf{A}\|_2$记为矩阵A每一行的 l_2-范数之和。最后，回顾一下，对于任意矩阵 $\tilde{}\mathbf{A} \in \mathbb{R}^{n^2\times N}$，$\tilde{}\mathbf{A}_S$ 是通过选取A中由 S 索引的行并将其他行置零而形成的矩阵。

采用上述记号，并回顾一下，$\{A_i^{MT}\}_{i\in[N]}$表示式 (4) 的一个最优解，我们有如下结果。

引理*1*：设假设1成立。令 λ_0 如定理 1，$\lambda \geq 4nNP\lambda_0$ $\in$ 中的$\tilde{}$所定义，并取 。$\mathbb{R}^{n^2\times N}$ $S \in \mathcal{S}$ 那么，对于任意$\mathbf{A}$以及任意满足相容性条件的，以下界以至少 $1-\mathrm{e}^{-\gamma}$ 的概率成立：

$$\sum_{i=1}^N \|\tilde{X}_i\tilde{\mathbf{a}}_i^\star - \tilde{X}_i\tilde{\mathbf{a}}_i^{MT}\|_2^2 + \frac{\lambda}{\sqrt{N}}\|\tilde{\mathbf{A}}^{MT} - \tilde{\mathbf{A}}_S\|_{2,1} \leq 6\sum_{i=1}^N \left\|\tilde{X}_i\tilde{\mathbf{a}}_i^\star - \tilde{X}_i\tilde{\mathbf{a}}_{S,i}\right\|_2^2 + \frac{24\lambda^2|S|}{PN\phi^2(S)} \quad (15)$$

对于任意给定的 $\gamma > 0$。 □

由于篇幅所限，引理1的证明从略。简而言之，引理1的证明借鉴了文献 [30] 第 6 章和第 8 章中的一些论证。首先，我们仿照 [30, 第 6], 章，为我们的问题建立基本不等式，并利用 [30,引理 8.5] 对经验过程进行界定。接着，我们为组稀疏模型推导出一个类似于 [30,引理 6.3] 的不等式。最后，结合这些不等式，证明引理1中的不等式（类似于[30,定理 6.2]）。完整证明将在在线扩展版本中提供。

根据引理1的结果，通过在式 (15) 中令 $\mathbf{A}_S = \mathbf{A}^\star$，即可得到定理 1$\tilde{}$ $\tilde{}$ 中的界，从而得到

$$\sum_{i=1}^N \|\tilde{X}_i\tilde{\mathbf{a}}_i^\star - \tilde{X}_i\tilde{\mathbf{a}}_i^{MT}\|_2^2 + \frac{\lambda}{\sqrt{N}}\|\tilde{\mathbf{A}}^{MT} - \tilde{\mathbf{A}}^\star\|_{2,1} \leq \frac{24\lambda^2|S^\star|}{PN\phi^2(S^\star)} \quad (16)$$

并注意到 $\sum_{i=1}^N \|\tilde{X}_i\tilde{}_i^\star - \tilde{X}_i\tilde{}_i^{MT}\|_2^2 = \mathbf{aa}()()$。另一方面，通过在式 (15) 中令 $\sum_{i=1}^N\sum_{\tau=1}^P \| A_i^\star - A_i^{MT} x_i \tau \|_2^2$ $\tilde{}\mathbf{A}_S = \tilde{}\mathbf{A}^\dagger$，即可得到定理 2 中的界；这里回顾一下，$\tilde{}\mathbf{A}^\dagger$ 是预言机解。