



基于高阶积分的控制方程稀疏系数分布辨识

曹瑞^{1,2*}, 裔扬¹

1. 扬州大学信息与人工智能学院, 扬州 225101

2. 北京理工大学自主智能无人系统全国重点实验室, 北京 100081

* 通信作者. E-mail: stdio@yzu.edu.cn

收稿日期: 2025-10-30; 修回日期: 2025-11-25; 接受日期: 2025-12-02; 网络出版日期: 2025-04-03

江苏省自然科学基金项目 (批准号: BK20230560)、国家自然科学基金青年项目 (批准号: 62303400)、国家自然科学基金重大研究计划项目 (批准号: 92371116)、国家自然科学基金面上项目 (批准号: 52272369)、自主智能无人系统全国重点实验室开放课题 (批准号: ZZKF2025-3-1)、江苏省高等学校面上项目 (批准号: 23KJB590003) 和 2022 年度扬州市“绿扬金凤计划”优秀博士项目 (批准号: YZLYJFJH2022YXBS109) 资助

摘要 从数据集中辨识物理系统控制方程一直是各科学领域的长期挑战. 稀疏辨识非线性动力学方法通常依赖精确的导数估计, 导致其在数据稀缺和噪声环境中容易失效. 为提升辨识的数据鲁棒性, 本文提出了一种新的数据驱动方法 DIRK-BSINDy, 它将高阶积分 (高斯隐式龙格-库塔方法) 与贝叶斯稀疏识别相结合, 并使用深度神经网络辅助求解高阶积分阶段值以提高计算效率. 该方法通过贝叶斯方法获得控制方程稀疏系数的后验分布, 并结合前向与后向预测构建损失函数, 增强稳定性. 凭借高阶积分对步长约束较弱的特性, DIRK-BSINDy 方法在数据稀缺和强噪声干扰的场景下仍展现出显著的鲁棒性, 相较于传统方法具有明显优势.

关键词 数据驱动学习, 系统辨识, 数据鲁棒性, 隐式龙格-库塔

1 引言

在过去的几十年里, 人们付出了大量努力来开发系统中用于建模和预测的数据驱动方法^[1,2], 并在控制工程和湍流建模等领域取得了显著成功^[2]. 尽管这些方法能够捕捉复杂系统的潜在结构^[3,4], 但主流方法 (如神经网络^[5]、随机森林^[6]、谱方法^[7]等) 生成的模型缺乏可解释性. 为应对这一挑战, 创新性的混合方法应运而生, 例如物理信息机器学习^[8]和基于深度神经网络的通用微分方程^[9].

在控制工程、航空航天和生物等领域应用科学机器学习方法的一个主要困难在于, 这些领域的测量数据通常含噪且稀疏. 这一限制使得许多数据密集型机器学习技术不可行. 此外, 在控制等实时应用中, 学习速度也是一项主要要求. 目前流行的稀疏非线性动力学识别 (sparse identification of nonlinear dynamics, SINDy) 框架^[10]提供了一种有前景的方法, 其可直接从数据中提取简约方程. SINDy 通过稀疏回归推动简约性, 能够识别可解释的模型, 平衡准确性和泛化性, 同时其简单性使其与其他机器学习技术相比具有较高的学习效率. 尽管 SINDy 比大多数机器学习技术的数据密集度低且速度更快, 但在数据集稀疏时仍易受噪声影响^[11]. 在某些情况下, 如果数据过于稀疏和含噪, SINDy 可能难以恢复准确且简约的模型. North 等^[12]系统综述了数据驱动系统建模的 3 类主要方法 (稀疏回归、符号回归和深度学习), 并指出传统 SINDy 类方法面临噪声鲁棒性与数据效率的平衡难题, 尤其在高阶系统辨识中显式求导会放大噪声. 为应对这一挑战, Schaeffer 和 McCalla^[13]提出了 SINDy

引用格式: 曹瑞, 裔扬. 基于高阶积分的控制方程稀疏系数分布辨识. 中国科学: 信息科学, 2026, 56: 1847–1864, doi: 10.1360/SSI-2025-0277

Cao R, Yi Y. Sparse coefficient distribution identification of control equations based on high-order integration. Sci Sin Inform, 2026, 56: 1847–1864, doi: 10.1360/SSI-2025-0277

的积分公式. Messenger 和 Bortz^[14] 通过提出弱 SINDy (weak SINDy, WSINDy) 以提供更好的抗噪声鲁棒性. Raissi 等^[15] 提出一种融合物理信息的神经网络框架, 其通过物理方程约束神经网络的训练方向, 但依赖于已知方程部分结构. 翁若昊等^[16] 提出了一种基于 Koopman 特征核的推理网络, 建立了变量之间的强非线性关系, 但该方法较为依赖因果关系的准确性, 并且当变量维数较高时会非常耗时. Tipping 和 Faul^[17] 提出了 SparseBayes 方法将贝叶斯与 SINDy 相结合以提高抗噪性能和量化不确定性. 但是该方法依然需要使用导数信息, 会导致噪声被显著放大, 导致学习结果中的不确定性过大. Goyal 和 Benner^[18] 将稀疏识别与经典四阶龙格-库塔方法 (RK4)^[19] 相结合, 提出了 RK4-SINDy 框架, 与 Chen^[20] 对线性多步法的研究类似, 减轻了对导数近似的依赖, 为动态系统的稀疏识别提供了新框架. RK4-SINDy 相比传统 SINDy 的效能提升归因于 RK4 的高阶收敛性. 然而, 微分方程数值求解中的显式方法 (如 RK4) 受限于有限的稳定区域, 不可避免地存在实际采样时间步长限制^[19].

文献 [21] 指出对于任意指定的阶数, 均存在与之对应的 A 稳定隐式龙格-库塔方法 (implicit Runge-Kutta, IRK). IRK 方法因步长限制更少且精度高, 尤其适用于动态系统的辨识^[11]. 隐式高斯方法属于 A 稳定的全隐式龙格-库塔方法, 具有最高精度^[22], 因此成为解决刚性问题的理想选择. 在本文中, 我们将 IRK 与贝叶斯稀疏识别相结合, 并使用深度神经网络辅助求解 IRK 阶段值提高计算效率, 提出了深度 IRK 贝叶斯稀疏非线性动力学辨识方法, 简称 DIRK-BSINDy 方法. 该方法继承 WSINDy/RK4-SINDy 的积分思想, 不直接依赖导数计算, 对噪声具有更强鲁棒性; 延续 SparseBayes 的概率建模, 引入双向预测约束提升抗过拟合能力; 融合物理约束理念^[15], 通过 IRK 数值约束引导神经网络学习. 此外, 将深度神经网络用于 IRK 阶段值预测, 解决隐式方法计算复杂性问题, 提高计算效率.

2 问题描述

在此以式 (1) 所示的具有三阶动力学的二维阻尼谐波振荡器为例

$$\begin{cases} \dot{x}_1(t) = -0.1x_1^3(t) + 2.0x_2^3(t), \\ \dot{x}_2(t) = -2.0x_1^3(t) - 0.1x_2^3(t), \end{cases} \quad (1)$$

在初始状态 $[x_1(0), x_2(0)]^T = [2.0, 0.0]^T$ 下获得一组离线采样数据. 其仿真总时间为 20 s, 样本量 m 分别取 500, 100 和 50, 以分析数据量对算法的影响. 图 1 展示了不同样本量下传统 SINDy^[10], RK4-SINDy^[18] 和 SparseBayes^[17] 方法的辨识结果.

为了进一步分析噪声对辨识结果的影响, 我们对洁净采样数据加入噪声. 测量噪声均采用零均值高斯分布, 且不同时刻、不同状态维度的噪声相互独立 (无时间相关性). 第 i 维噪声其概率密度函数为

$$p(\epsilon_i) = \frac{1}{\sigma_{x,i}\sqrt{2\pi}} \exp\left(-\frac{\epsilon_i^2}{2\sigma_{x,i}^2}\right), \quad (2)$$

其中 $i = 1, \dots, d$ 代表系统状态维度, ϵ 为噪声随机变量, σ_x 为噪声标准差, 均值是 0.

考虑测量噪声的最大幅值为系统真实数据方差的 5%. 对应噪声标准差为 $\sigma_{x,i} = 0.02 \times \text{std}(x_{i,\text{ideal}})$, $\text{std}(x_{i,\text{ideal}})$ 为无噪声时第 i 维洁净数据 $x_{i,\text{ideal}}$ 的标准差. 则含噪数据为 $x_{i,\text{noisy}}(t_k) = x_{i,\text{ideal}}(t_k) + \epsilon_{i,k}$, $k = 1, 2, \dots, m$. 图 2 展示了存在测量噪声下传统 SINDy, RK4-SINDy 及 SparseBayes 方法的辨识结果, 其中蓝色实线 (ideal state) 为理想无噪声系统状态, 蓝色点号 (sampling point) 为实际含噪声采样数据.

根据仿真结果可以得出, 当采样数据集数量较小或者存在噪声影响时, 辨识结果会偏离正确信息. 这是因为 SINDy 和 SparseBayes 方法均需要显式计算导数信息, 而对噪声数据求导会进一步加剧噪声作用, 进而影响辨识结果. RK4-SINDy 方法是通过 4 次显式函数评估近似导数, 虽然未显式构建导数估计, 但是当数据含噪声时, 函数评估误差会累积, 也导致积分精度下降. 此外, RK4-SINDy 和 SparseBayes 方法仅采用前向预测构建稀疏回归

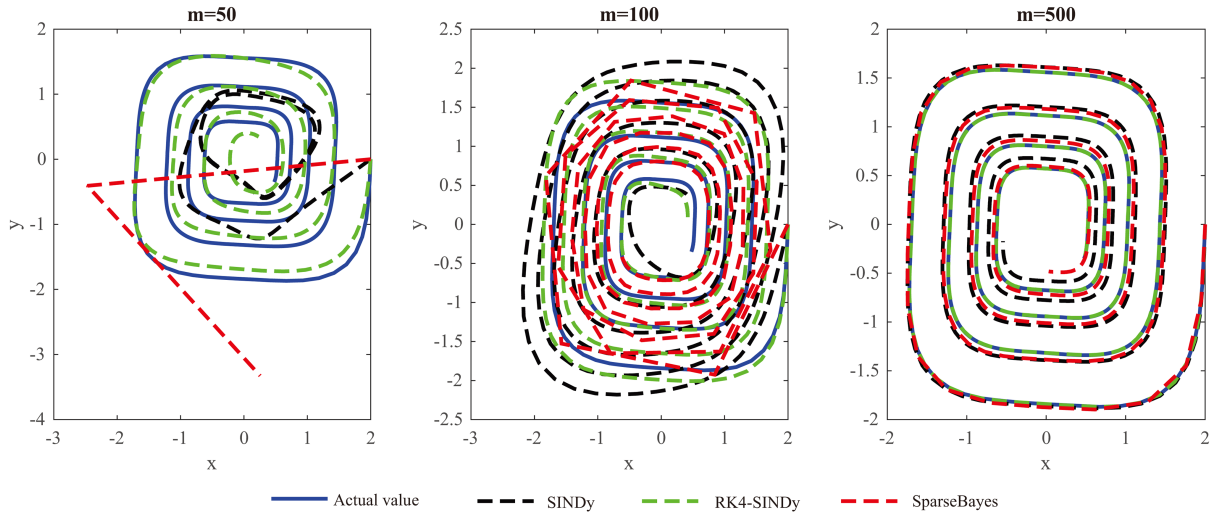
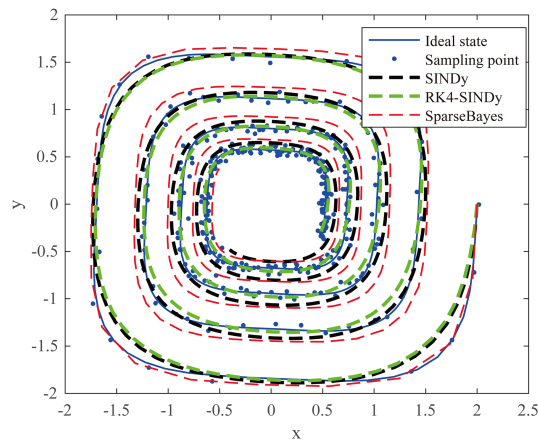


图 1 (网络版彩图) 采样数据集大小对辨识结果影响.

Figure 1 (Color online) The influence of sampling dataset size on identification results.

图 2 (网络版彩图) 噪声对辨识结果影响 ($m = 500$).Figure 2 (Color online) The impact of noise on identification results ($m = 500$).

问题, 缺乏后向约束, 因此在噪声环境下可能产生过拟合; 当数据量较少时, 上述方法也会难以准确识别模型中的关键项.

为了解决上述噪声和训练数据量的问题, 我们提出了 DIRK-BSINDy 算法. 该方法首先使用高斯 IRK 以获得高精度积分. 然后利用深度神经网络辅助预测高斯 IRK 阶段值进一步降低计算量. 最后通过贝叶斯稀疏性惩罚复杂模型, 并且同时使用前向预测和后向预测改进损失函数, 利用双向一致性增强稀疏识别的稳定性. DIRK-BSINDy 方法能够同时优化系数矩阵概率分布和神经网络参数, 并有效整合深度神经网络、隐式龙格-库塔方法和贝叶斯稀疏识别技术, 提高在噪声和数据稀缺场景下的辨识鲁棒性.

3 DIRK-BSINDy 算法

本节结合高斯隐式龙格-库塔方法, 提出从稀疏采样观测数据中发现控制方程的方法, 其核心是将具有强数值稳定性的高阶数值积分技术, 与非线性函数空间中控制方程的内在稀疏性相结合. 与 SparseBayes 方法^[17]和 RK4-SINDy 框架^[18]不同, 该方法利用高斯 IRK 方法获得高精度积分, 并使用前向预测和后向预测构建损失函数, 通过双向一致性增强稀疏识别的稳定性. 图 3 展示了本文所提出的 DIRK-BSINDy 算法框架.

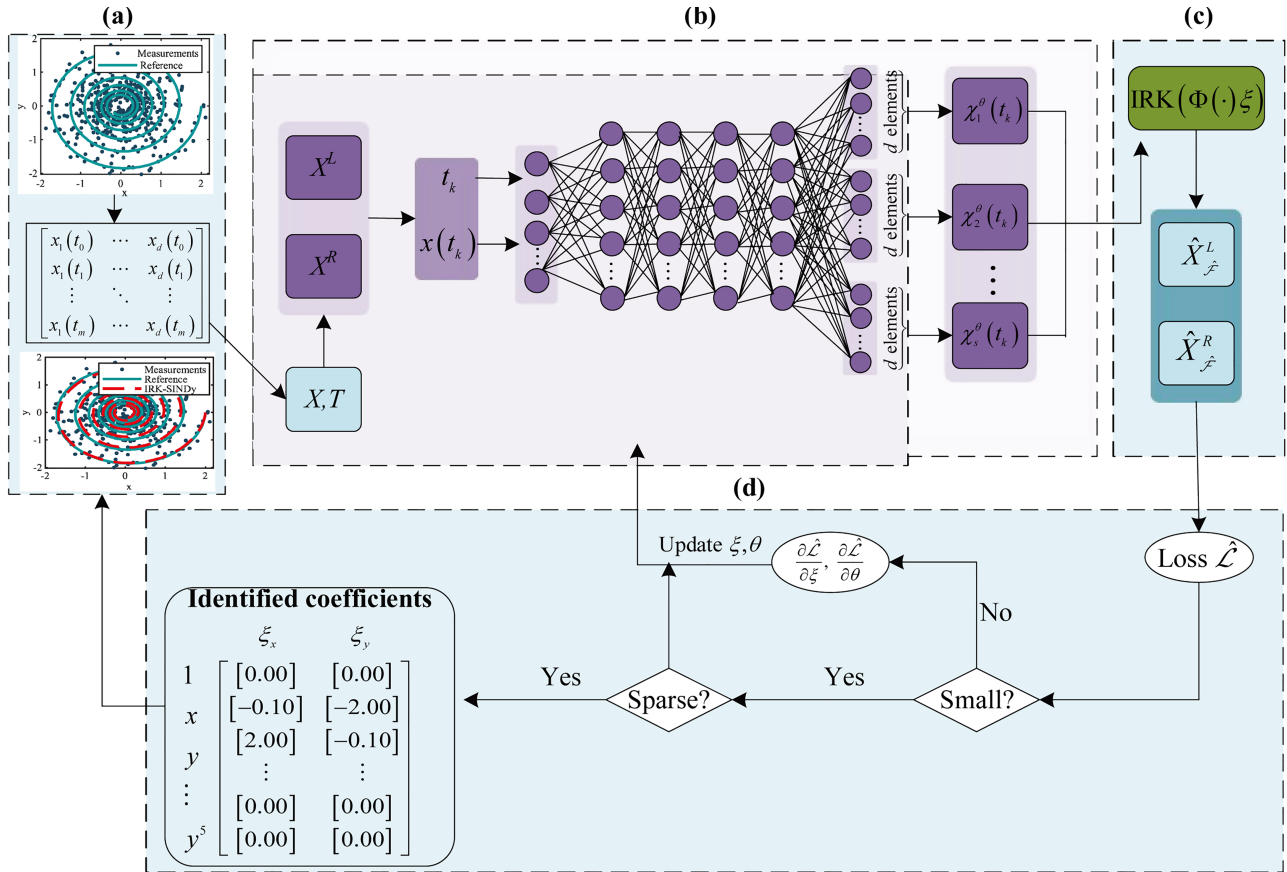


图3 (网络版彩图) DIRK-BSINDy 框架. (a) 训练数据集; (b) 神经网络; (c) 前向和后向预测; (d) 迭代优化.

Figure 3 (Color online) DIRK-BSINDy framework. (a) Training dataset; (b) neural network; (c) forward and backward prediction; (d) iterative optimization.

该方法包括 4 个部分: (a) 采集用于训练神经网络的数据集; (b) 建立用于预测 IRK 阶段值 χ_i 的深度神经网络 (deep neural network, DNN); (c) 将 DNN 前向传播过程预测阶段值应用于 IRK 步骤, 实现状态的前向和后向预测; (d) 利用贝叶斯理论构建损失函数并执行优化步骤. 当达到指定的训练轮次且损失充分小时, 仅对系数矩阵 ξ 应用稀疏促进算法. 最终, ξ 中的非零系数表示非线性特征库中的关键项. 所提出的 DIRK-BSINDy 算法具有以下优势:

- (1) 用神经网络替代传统迭代, 解决 IRK 随阶段值阶次和系统状态维数呈指数级增长的计算瓶颈;
- (2) 利用双向预测和贝叶斯稀疏性, 避免过拟合, 提升模型的物理可解释性与泛化能力, 并提供不确定性量化 (稀疏系数的均值和方差);
- (3) 为复杂动力系统 (如生物振荡、混沌系统) 的无导数建模提供普适性方案.

3.1 隐式龙格 – 库塔方法

龙格 – 库塔方法由于能够根据任何指定的阶数构造, 因此被广泛用于求解初值问题^[19,21,23]. 隐式龙格 – 库塔方法被广泛认为是解决与刚性相关问题的非常合适的候选方法^[19,21]. 高阶隐式龙格 – 库塔方法对步长的限制较少, 因此在数据可用性受限的动态系统的稀疏识别中可以发挥关键作用.

考虑以下初值问题:

$$\begin{cases} \dot{\mathbf{x}}(t) = f(\mathbf{x}(t)), \\ \mathbf{x}(t_0) = \mathbf{x}_0, \end{cases} \quad (3)$$

其中向量 $\mathbf{x}(t) = [x_1(t) \cdots x_d(t)]^T$ 表示状态变量, $f: \mathbb{R}^d \rightarrow \mathbb{R}^d$ 代表待从含噪声数据中识别的未知向量场. 为了

识别方程 (1) 的非线性动力学, 即 $f(\cdot)$, 我们在时刻 $\{t_0, t_1, \dots, t_m\}$ 采样, 记采样数据为 $\{\mathbf{x}(t_k)\}_{k=0}^m$, 其中采样步长为 $h_k = t_{k+1} - t_k$.

本文提出的方法利用 IRK 方法, 比较的是数据 $\mathbf{x}(t_{k+1})$ 与根据 $\mathbf{x}(t_k)$ 预测的 t_{k+1} 时刻的 $\mathbf{x}$ 值 (参考文献 [18, 20] 的思路). 因此, 对方程 (1) 采用具有 s 个阶段的通用形式龙格-库塔方法 [21]. 式 (4) 和 (5) 是 IRKs 用于求解常微分方程初值问题的核心递推公式

$$\mathbf{x}(t_{k+1}) \approx \mathbf{x}(t_k) + h_k \sum_{j=1}^s b_j f(\chi_j^f(t_k)), \quad (4)$$

$$\chi_i^f(t_k) = \mathbf{x}(t_k) + h_k \sum_{j=1}^s a_{ij} f(\chi_j^f(t_k)), \quad i = 1, \dots, s, \quad (5)$$

其中, b_j 为权重系数, a_{ij} 是 IRK 方法求解的矩阵系数, $f(\chi_j^f(t_k))$ 代表第 j 个阶段的向量场值. $\chi_j^f(t_k) \approx \mathbf{x}(t_k + c_j h_k)$ 是第 j 个阶段值, 表示 $t_k + c_j h_k$ 时刻的状态近似值 (c_j 为阶段位置参数). 将阶段值 χ 和向量场 $f(\chi)$ 向量化

$$\chi = \begin{pmatrix} \chi_1^f \\ \chi_2^f \\ \vdots \\ \chi_s^f \end{pmatrix}, \quad F(\chi) = \begin{pmatrix} f(\chi_1^f) \\ f(\chi_2^f) \\ \vdots \\ f(\chi_s^f) \end{pmatrix}, \quad (6)$$

则, 基于式 (6), 式 (4) 和 (5) 可表示为矩阵形式:

$$\begin{bmatrix} \chi(t_k) \\ \mathbf{x}(t_{k+1}) \end{bmatrix} \approx \begin{bmatrix} \mathbf{A} \otimes \mathbb{I}_d & \mathbf{1}_s \otimes \mathbb{I}_d \\ \mathbf{b}^T \otimes \mathbb{I}_d & \mathbf{1}_d \end{bmatrix} \begin{bmatrix} h_k F(\chi(t_k)) \\ \mathbf{x}(t_k) \end{bmatrix}, \quad a_{ij} \in \mathbf{A}, \quad (7)$$

其中, $\mathbf{A}$ 是由元素 a_{ij} 组成的系数矩阵, $\mathbf{b}$ 由 b_j 构成的权重矩阵, $\mathbb{I}_d$ 表示 $d \times d$ 单位矩阵, d 为未知系统的状态维数, $\mathbf{1}_s$ 是 s 维单位向量, $\otimes$ 表示克罗内克积.

上式通过阶段值 χ 构成中间状态, 通过 IRK 系数矩阵 $\mathbf{A}$ 和权重 $\mathbf{b}$ 连接前后时刻的状态, 形成多阶段的数值积分过程. 这一通用公式的灵活性允许根据系数矩阵 $\mathbf{A}$ 的选择, 实现隐式或显式时间步进方案.

由于本文用于发现系统控制方程的方法论聚焦于利用高阶 IRK 方法, 因此本方法的核心思想是通过神经网络预测阶段值 χ , 向量场 f 由非线性特征库 $\Phi(\mathbf{x})$ 的稀疏线性组合表示 ($f \approx \Phi(\mathbf{x})\xi$). 最后通过最小化损失函数, 同时优化系数矩阵 ξ 和 IRK 参数, 实现稀疏模型发现.

3.2 基于 IRK 的向量场贝叶斯稀疏表示

本文采用的高斯隐龙格-库塔方法, 对于 s 个阶段, 其单步误差随步长 h 呈 $2s$ 次方衰减 [19, 21, 23]. 因此该方法能够用较少的阶段实现高阶精度, 尤其适合对精度要求极高的问题. 对于足够小的步长 h_k , 该方法在从 $\mathbf{x}(t_k)$ 计算 $\mathbf{x}(t_{k+1})$ 时具有高精度. 若将方程 (4) 的右侧视为关于 f , $\mathbf{x}(t_k)$ 和 h_k 的函数, 记为 $\mathcal{F}_{irk}$, 则 $\mathbf{x}(t_{k+1}) \approx \mathcal{F}_{irk}(f, \mathbf{x}(t_k), h_k)$ 的合理性直接成立:

$$\mathcal{F}_{irk}(f, \mathbf{x}(t_k), h_k) := \mathbf{x}(t_k) + h_k \sum_{j=1}^s b_j f(\chi_j^f(t_k)), \quad (8)$$

这种表示将 IRK 方法视为“黑箱”函数, 为后续章节用神经网络近似 $\mathcal{F}_{irk}$ 奠定理论基础.

接下来, 我们对待识别向量场 f 执行贝叶斯稀疏表示. 本节将基于先前假设的时间点 $\{t_0, \dots, t_m\}$ 上的 $\mathbf{x}(t)$

时间序列数据, 推导向量场 f 的最简约表示. 按以下方式考虑数据矩阵:

$$\mathbf{X}^L := \begin{bmatrix} \mathbf{x}^T(t_0) \\ \mathbf{x}^T(t_1) \\ \vdots \\ \mathbf{x}^T(t_{m-1}) \end{bmatrix}, \quad \mathbf{X}^R := \begin{bmatrix} \mathbf{x}^T(t_1) \\ \mathbf{x}^T(t_2) \\ \vdots \\ \mathbf{x}^T(t_m) \end{bmatrix}, \quad (9)$$

并将预测表示为

$$\mathbf{X}_{\mathcal{F}}^R(f) = \begin{bmatrix} \mathcal{F}_{irk}(f, \mathbf{x}(t_0), h_0) \\ \mathcal{F}_{irk}(f, \mathbf{x}(t_1), h_1) \\ \vdots \\ \mathcal{F}_{irk}(f, \mathbf{x}(t_{m-1}), h_{m-1}) \end{bmatrix}, \quad \mathbf{X}_{\mathcal{F}}^L(f) = \begin{bmatrix} \mathcal{F}_{irk}(f, \mathbf{x}(t_0), -h_0) \\ \mathcal{F}_{irk}(f, \mathbf{x}(t_1), -h_1) \\ \vdots \\ \mathcal{F}_{irk}(f, \mathbf{x}(t_{m-1}), -h_{m-1}) \end{bmatrix}, \quad (10)$$

上式定义了基于 IRK 方法的前向预测矩阵 $\mathbf{X}_{\mathcal{F}}^R(f)$ 和后向预测矩阵 $\mathbf{X}_{\mathcal{F}}^L(f)$. 式中 $h_k, k = 0, 1, \dots, m-1$ 是时间步长, $-h_k$ 表示反向积分 (从未来状态反推当前状态).

稀疏识别的下一个关键步骤是生成一个庞大的非线性特征库 (记为 Φ), 其中包含可能在动力学中起作用的潜在非线性函数. 假设函数 $f(\cdot)$ 可以表示为从该特征库中选择的少数项的线性组合^[10]. 例如, 特征库中可以选择一组多项式、指数函数以及三角函数. 对于向量 $\mathbf{x} = [x_1, \dots, x_d]^T$, 一个特征库可以表示为

$$\Phi(\mathbf{x}) := [1, \mathbf{x}, \mathbf{x}^{P_2}, \dots, \mathbf{x}^{P_d}, \dots, \sin(\mathbf{x}), \cos(\mathbf{x}), \dots, \exp(-\mathbf{x}), \exp(-2\mathbf{x}), \dots], \quad (11)$$

其中, $\mathbf{x}^{P_i}$ 表示 i 次多项式. 例如, 当维度 $d = 2$ 时, 二次多项式 $\mathbf{x}^{P_2}$ 可表示为

$$\mathbf{x}^{P_2} = [x_1^2(t), x_1(t)x_2(t), x_2^2(t)]. \quad (12)$$

特征库 Φ 中的每个元素均为表示向量场 f 的潜在候选函数. 在传统 SINDy 中是通过最小二乘优化计算出稀疏系数矩阵 ξ 的点估计值, 而贝叶斯稀疏模型是通过似然函数和 ξ 的先验分布来获取其后验分布. 因此, 我们为获得最终的 ξ 分布情况, 需要设置 ξ 的先验分布以及计算似然函数.

为采用估计的后验分布, 本文首先设置 ξ 先验分布为正则化马蹄先验 (具有增强稀疏性与抗噪声能力) 形式, 其数学表达式为

$$\xi_{jk} \sim \mathcal{N}(0, \tau^2 \lambda_{jk}^2), \quad \lambda_{jk} \sim \text{Half-Cauchy}(0, c), \quad \tau \sim \text{Half-Cauchy}(0, \tau_0), \quad (13)$$

其中, τ 控制全局稀疏性, λ_{jk} 用于调整单个系数 ξ_{jk} 的收缩强度, Half-Cauchy 为半柯西分布, 其概率密度在 0 附近集中、尾部平缓, 既能使冗余项趋近于 0, 又能让关键项保持较大值^[24]. 根据半柯西分布特性, τ 取较小值可增强整体系数收缩, 推动更多冗余项趋于 0; 并且允许 λ_{jk} 取较大值, 使得 ξ_{jk} 的先验方差足够大, 避免先验信息过度主导数据信息. 基于上述特性, 本工作中默认取值为 $c = 1, \tau_0 = 0.1$.

在原方法中通过下式计算的后验分布:

$$p(\xi, \mathbf{x}_0, \sigma | \mathbf{X}) \propto \underbrace{p(\mathbf{X} | \xi, \mathbf{x}_0, \sigma)}_{\text{似然函数}} \cdot \underbrace{p(\xi)p(\mathbf{x}_0)p(\sigma)}_{\text{先验分布}}, \quad (14)$$

其中 σ 为概率模型的辅助变量, 如噪声标准差. 似然函数为 $p(\mathbf{X} | \xi, \mathbf{x}_0, \sigma) = \prod_{i=1}^n \prod_{j=1}^d \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2\sigma^2} |y_{i,j} - \hat{x}_j(t_i; \xi, \mathbf{x}_0)|^2)$, 用于描述观测数据与模型预测的匹配程度. 贝叶斯方法不追求最大化似然函数, 而是通过马尔可夫链蒙特卡罗方法 (Markov chain Monte Carlo method, MCMC)^[25] 采样来探索后验分布中概率较高的区域.

本文为了降低贝叶斯方法使用 MCMC 采样获取系数后验分布的计算复杂性, 以及便于后续将概率建模与神经网络相结合, 考虑采用最大化证据下界 (evidence lower bound, ELBO) [26] 的方式估算 ξ 的后验分布. 为此, 我们假设 ξ 的后验分布满足高斯变分分布, 即

$$q(\xi) = \mathcal{N}(\xi_\mu, \xi_\sigma^2), \quad (15)$$

其中 $\xi_\mu \in \mathbb{R}^{M \times d}$ 为系数均值矩阵, 表征各项对动力学的贡献强度. $\xi_\sigma \in \mathbb{R}_+^{M \times d}$ 为系数标准差矩阵, 量化噪声和数据稀缺导致的不确定性. M 代表基函数数目, d 为系统维度. ξ_σ 本质反映动力学结构对数据扰动的敏感程度, 大小代表对系数估计的置信度. 则向量场 f 的稀疏表示为

$$f(x) = \Phi(x)\xi, \quad (16)$$

其中 ξ 为满足变分分布 $q(\xi) = \mathcal{N}(\xi_\mu, \xi_\sigma^2)$ 的期望值. 需特别说明的是, 通过上述变分优化得到的 ξ , 其非零系数具有明确数学与物理含义. 数学上, ξ_μ 非零且 ξ_σ 较小代表了统计显著项, 此现象源于正则化马蹄先验的双重稀疏约束. 物理上, 这些非零项对应系统主导动力学机制, 如三阶阻尼振荡器中的三阶非线性项, 确保模型稀疏性与物理真实性统一. 在后续优化过程中, 我们通过重参数化技巧将 ξ 表示为

$$\xi = \xi_\mu + \xi_\sigma \odot \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \mathbf{I}_M), \quad (17)$$

其中 ε 为标准正态随机向量, $\odot$ 为逐元素乘法运算符. 通过重参数化技巧, 构建式 (17), 以实现 ξ 的梯度可导. 这一方式能够将 ξ 随机性隔离至 ε , 确保损失函数沿待优化参数 ξ_μ 和 ξ_σ 的梯度稳定回传 [27].

然后, 将 $f(x) = \Phi(x)\xi$ 代入 $\mathcal{F}_{irk}$, 得到 $\mathcal{F}_{irk}(\Phi(x)\xi, x(t_k), h_k)$, 并且记相对应的预测值为 $\mathbf{X}_{\mathcal{F}}^R(\Phi(\cdot)\xi)$ 和 $\mathbf{X}_{\mathcal{F}}^L(\Phi(\cdot)\xi)$. 总损失函数 ELBO 可分解为重构损失项和散度正则项, 即

$$\mathcal{L} = \mathcal{L}_{rec} + \beta \mathcal{L}_{KL}, \quad (18)$$

其中

$$\mathcal{L}_{rec} = \alpha \|\mathbf{X}^L - \mathbf{X}_{\mathcal{F}}^L(\Phi(\cdot)\xi)\|_2^2 + (1 - \alpha) \|\mathbf{X}^R - \mathbf{X}_{\mathcal{F}}^R(\Phi(\cdot)\xi)\|_2^2, \quad \xi = \xi_\mu + \xi_\sigma \odot \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \mathbf{I}_M). \quad (19)$$

重构损失项 (19) 同时使用前向和后向预测矩阵是达到保证物理一致性、提升抗噪性和稳定优化问题的目的. 其中 α ($0 \leq \alpha \leq 1$) 是前向和后向预测的加权融合项, 作用是强制模型满足动力学演化的双向一致性. $\alpha > 0.5$ 时, 增强后向预测权重, 优先保障物理可逆性; $\alpha < 0.5$ 时, 增强前向预测权重, 优先保障数据拟合精度; $\alpha = 0.5$ 时, 前向与后向预测权重相等, 强制模型在“正向演化拟合”与“反向物理一致”间达到平衡, 避免单一方向约束导致的偏差. 本算法中设置默认值为 0.5.

$$\mathcal{L}_{KL} = \sum_{j=1}^M \sum_{k=1}^d \left[\log \frac{\xi_{\sigma jk}}{\tau \lambda_{jk}} + \frac{\tau^2 \lambda_{jk}^2 + \xi_{\mu jk}^2}{2\xi_{\sigma jk}} - \frac{1}{2} \right], \quad (20)$$

正则项 (20) 用于衡量变分后验分布 $q(\xi)$ 与先验分布 $p(\xi)$ 的差异, 目标是通过正则化约束系数稀疏性, 反映“物理约束”的简约性需求. 其中权衡参数 β 控制重构损失与稀疏正则的平衡. 当 $\beta > 1$ 时, 增强稀疏正则项的权重, 优先保障模型简约性; $\beta < 1$ 时, 减弱正则项权重, 优先保障数据拟合精度; $\beta = 1$ 时, 默认两类需求权重相等, 是“准确性 - 简约性”的平衡基准. 本文方法的核心目标是在数据稀缺与噪声环境下实现鲁棒辨识, 故取 β 默认值为 1, 以适配该目标. 研究者可根据实际场景需求进行调整.

然而, 仅通过最小化损失函数 (18) 仍无法完全保证稀疏性的严格性. 这是由于变分后验分布 $q(\xi)$ 是连续高斯分布, 即使 ξ_μ (系数均值) 趋近于 0, 仍可能存在“近零系数”, 导致冗余基函数未被剔除, 影响模型可解释性. 因此本文考虑引入正则化项 $\mathcal{R}(\xi_\mu, \xi_\sigma, \lambda)$, 其定义为

$$\mathcal{R}(\xi_\mu, \xi_\sigma, \lambda) = \begin{cases} \infty, & |\xi_{\mu jk}| \leq \lambda, \\ 0, & |\xi_{\mu jk}| > \lambda, \end{cases}$$

其中, 阈值 $\lambda \in (0, 1)$. 该正则项的物理含义是, 对 $|\xi_{\mu_{jk}}| \leq \lambda$ 的非零系数施加无穷大惩罚, 强制其在优化过程中设为 0; 对 $|\xi_{\mu_{jk}}| > \lambda$ 的系数无惩罚, 保留其贡献.

将 $\mathcal{R}(\xi_{\mu}, \xi_{\delta}, \lambda)$ 与基础损失函数 $\mathcal{L}$ (式 (18)) 融合, 最终形成式 (21) 所示的正则优化问题:

$$[\xi_{\mu}^*, \xi_{\delta}^*] = \arg \min_{\xi_{\mu}, \xi_{\delta}} \{ \mathcal{L}(\xi_{\mu}, \xi_{\delta}) + \mathcal{R}(\xi_{\mu}, \xi_{\delta}, \lambda) \}. \quad (21)$$

其优化逻辑为, 在最小化拟合误差和 KL 散度的同时, 通过 λ 约束强制剔除非显著项, 最终得到稀疏性严格、物理意义明确的系数矩阵 ξ_{μ} (关键项系数) 与 ξ_{δ} (量化关键项的不确定性). 该转换通过损失函数保精度和正则化项强稀疏的双重约束, 避免数据稀疏/噪声下的过拟合, 原理细节可参考文献 [14, 28].

3.3 阶段值神经网络预测

求解优化问题 (21) 至关重要的一点是获取 IRKs 中的阶段值 $\chi_i(t_k)$, $i = 1, \dots, s$. 在经典 IRK 实现中, 阶段值通过迭代方法 (如定点迭代和牛顿法) 求解非线性代数方程组 (5) 来计算 [21]. 但是, 鉴于定点迭代法在处理非线性和刚性问题时的局限性 [22, 23], 它们在非线性动力系统的稀疏识别中效率较低. 并且, 对于复杂系统, 为了提高预测精度, 必须使用更高阶的 IRKs, 这意味着需要更多的阶段值. 随着阶段数 s 的增加, 与定点迭代计算相关的计算成本呈指数级增长 [21].

因此, 为了解决这一问题, 我们通过引入辅助深度神经网络 [29] 对阶段值 χ_i^f 进行预测, 如图 3(b) 所示, 以确保计算成本随所使用的阶段值数量呈线性增长. 建立一个辅助 DNN $\chi^{\theta} = [\chi_1^{\theta}, \dots, \chi_s^{\theta}]$, 其输入为时间 t_k 和状态 $\mathbf{x}(t_k)$, 输出为 IRKs 阶段值, 其中 θ 为神经网络参数. 辅助 DNN 采用全连接网络结构, 其激活函数根据系统特性适配, 即处理周期性动力学系统时, 选用 SIREN (周期性激活函数), 通用非线性场景可选 Swish 函数. 此外, 辅助 DNN 利用 Adam 优化器对参数 θ 进行优化, 其中神经网络参数学习率默认 10^{-3} , 以避免神经网络参数更新过快导致的阶段值预测震荡. 如图 3 所示, 通过训练辅助 DNN, 期望 χ^{θ} 能生成近似系统控制方程中 IRK 阶段值 χ^f 的预测值

$$\chi_i^{\theta}(t_k) \approx \chi_i^f(t_k), \quad i = 1, \dots, s, \quad k = 0, 1, \dots, m-1. \quad (22)$$

注. 针对不同维度与复杂度的系统, 建议的辅助 DNN 全连接网络结构如下. 对于低维系统 ($d \leq 3$), 隐藏层数量 1~4 层, 每层神经元数 32~256 个 (系统复杂度越高, 神经元数越多); 高维系统 ($d > 3$) 可增至 5~6 个隐藏层, 每层 64~128 个神经元, 以适配更高维度的阶段值预测需求.

为了在优化过程框架内实现辅助深度神经网络与隐式龙格-库塔方法的结合, 必须对式 (8) 进行重构. 因此, 通过从式 (5) 中减去式 (4), 推导出式 (23) 和 (24):

$$\hat{\mathcal{F}}_{irk}^L(f, \mathbf{x}(t_k), h_k, i) := \chi_i^{\theta}(t_k) - h_k \sum_{j=1}^s a_{ij} f(\chi_j^{\theta}(t_k)), \quad i = 1, \dots, s, \quad (23)$$

$$\hat{\mathcal{F}}_{irk}^R(f, \mathbf{x}(t_k), h_k, i) := \chi_i^{\theta}(t_k) + h_k \sum_{j=1}^s (b_j - a_{ij}) f(\chi_j^{\theta}(t_k)), \quad i = 1, \dots, s, \quad (24)$$

基于式 (4) 和 (5), 式 (23) 和 (24), 可以得出

$$\hat{\mathcal{F}}_{irk}^L(f, \mathbf{x}(t_k), h_k, i) \approx \mathbf{x}(t_k), \quad i = 1, \dots, s, \quad (25)$$

$$\hat{\mathcal{F}}_{irk}^R(f, \mathbf{x}(t_k), h_k, i) \approx \mathbf{x}(t_{k+1}) = \mathbf{x}(t_k + h_k), \quad i = 1, \dots, s, \quad (26)$$

二者源于 IRK 阶段值定义 (式 (5)) 与状态更新 (式 (5)) 的联立推导, 体现状态守恒 ($\hat{\mathcal{F}}_{irk}^L$) 与动力学演化 ($\hat{\mathcal{F}}_{irk}^R$) 的强约束条件.

基于式 (25) 和 (26), 根据式 (9) 的定义, IRK 网络可系统地定义为

$$\hat{\mathbf{X}}_{\mathcal{F}}^i(f) = \begin{bmatrix} \hat{\mathcal{F}}_{irk}^i(f, \mathbf{x}(t_0), h_0, 0) & \dots & \hat{\mathcal{F}}_{irk}^i(f, \mathbf{x}(t_0), h_0, s) \\ \hat{\mathcal{F}}_{irk}^i(f, \mathbf{x}(t_1), h_1, 0) & \dots & \hat{\mathcal{F}}_{irk}^i(f, \mathbf{x}(t_1), h_1, s) \\ \vdots & & \vdots \\ \hat{\mathcal{F}}_{irk}^i(f, \mathbf{x}(t_{m-1}), h_{m-1}, 0) & \dots & \hat{\mathcal{F}}_{irk}^i(f, \mathbf{x}(t_{m-1}), h_{m-1}, s) \end{bmatrix}, \quad i = L, R. \quad (27)$$

进而式 (18) 中的重构损失项被更新为

$$\hat{\mathcal{L}}_{rec} = \alpha \left\| \mathbf{X}^L - \hat{\mathbf{X}}_{\mathcal{F}}^L(\Phi(\cdot)\boldsymbol{\xi}) \right\|_2^2 + (1 - \alpha) \left\| \mathbf{X}^R - \hat{\mathbf{X}}_{\mathcal{F}}^R(\Phi(\cdot)\boldsymbol{\xi}) \right\|_2^2. \quad (28)$$

故总损失函数 (18) 被更新为 $\hat{\mathcal{L}} = \hat{\mathcal{L}}_{rec} + \beta \mathcal{L}_{KL}$. 则正则化优化问题 (21) 重写为如下表达式:

$$[\boldsymbol{\xi}_{\mu}^*, \boldsymbol{\xi}_{\delta}^*, \boldsymbol{\theta}^*] = \arg \min_{\boldsymbol{\xi}_{\mu}, \boldsymbol{\xi}_{\delta}, \boldsymbol{\theta}} \{ \hat{\mathcal{L}}(\boldsymbol{\xi}_{\mu}, \boldsymbol{\xi}_{\delta}) + \mathcal{R}(\boldsymbol{\xi}_{\mu}, \boldsymbol{\xi}_{\delta}, \lambda) \}. \quad (29)$$

通过这种方法能够同时确定系数矩阵均值 $\boldsymbol{\xi}_{\mu}$ 、方差 $\boldsymbol{\xi}_{\delta}^2$ 和神经网络参数 $\boldsymbol{\theta}$, 有效整合辅助深度神经网络、隐式龙格-库塔方法和贝叶斯稀疏识别技术, 形成 DIRK-BSINDy 框架.

3.4 算法执行流程

本文提出的 DIRK-BSINDy 方法使用 IRK 的 A 稳定性支持大步长与噪声数据, 通过贝叶斯稀疏性惩罚复杂模型, 避免过拟合, 提供不确定性量化 (稀疏系数的均值和方差), 并利用深度网络预测复杂 IRK 阶段值, 提高计算效率. 该方法无需导数信息即可执行优化过程, 发现系统的控制微分方程. DIRK-BSINDy 方法包括数据准备、模型构建与训练优化 3 个主要阶段, 详细执行流程如下.

步骤 1: 数据收集. 收集时间序列数据 $\{\mathbf{x}_k\}_{k=1}^N$, 其中 $\mathbf{x}_k \in \mathbb{R}^d$ 为系统状态, $t_k = t_0 + kh$ 为时间点.

步骤 2: 数据矩阵构建. 基于步骤 1 采样的数据集, 按式 (9) 构建 $\mathbf{X}^L, \mathbf{X}^R$.

步骤 3: 深度网络架构设计. 设计全连接神经网络以预测阶段值, 其中激活函数选用 SIREN (处理周期动力学) 或 Swish (通用非线性). 网络输入为当前状态 $\mathbf{x}_k$ 和步长 h . 输出层为对应时刻的阶段值 $\chi_k \in \mathbb{R}^d$, 总输出维度 $s \times d$.

步骤 4: 符号特征库构建. 参考式 (11) 和 (12), 建立非线性函数库 $\Phi(\mathbf{x}) = [\phi_1(\mathbf{x}), \dots, \phi_m(\mathbf{x})]$, 为稀疏回归提供候选特征.

步骤 5: 贝叶斯稀疏建模. 首先根据式 (13) 设置稀疏系数矩阵 $\boldsymbol{\xi}$ 的先验分布, 获取 $\boldsymbol{\xi}_{\mu}$ 和 $\boldsymbol{\xi}_{\delta}$ 的初值. 然后进行变分后验近似, 即 $q(\boldsymbol{\xi}) = \mathcal{N}(\boldsymbol{\xi}_{\mu}, \boldsymbol{\xi}_{\delta}^2)$ (高斯变分分布), 其中 $\boldsymbol{\xi}_{\mu} \in \mathbb{R}^{M \times d}$ 和 $\boldsymbol{\xi}_{\delta} \in \mathbb{R}_+^{M \times d}$ 为优化参数.

步骤 6: 训练初始化. 设置当前迭代步数 $k = 0$, 总迭代步数 N (默认 $N = 20000$).

步骤 7: 阶段值预测. 根据步骤 1 构建的数据集, 利用步骤 3 建立的神经网络预测阶段值.

步骤 8: 前后向预测状态矩阵计算. 基于步骤 7 获取的阶段值, 根据式 (23) 计算局部守恒项, 按式 (24) 计算演化项, 并按式 (27) 计算前向和后向状态预测矩阵.

步骤 9: 稀疏系数更新. 基于当前 $\boldsymbol{\xi}_{\mu}^{(k)}$ 和 $\boldsymbol{\xi}_{\delta}^{(k)}$, 使用式 (17) 更新 $\boldsymbol{\xi}^{(k)}$.

步骤 10: 损失函数计算. 结合步骤 2 构建的 $\mathbf{X}^L, \mathbf{X}^R$ 和步骤 5 建立的预测矩阵 (27), 基于式 (28) 计算 $\hat{\mathcal{L}}_{rec}$. 并获得总损失值 $\hat{\mathcal{L}} = \hat{\mathcal{L}}_{rec} + \beta \mathcal{L}_{KL}$.

步骤 11: 损失值优化. 使用 Adam 优化器最小化损失值, 对系数均值、方差和神经网络参数进行联合优化, 获得优化后的参数 $\boldsymbol{\xi}_{\mu}^{(k+1)'}$, $\boldsymbol{\xi}_{\delta}^{(k+1)'}$, $\boldsymbol{\theta}^{(k+1)}$.

步骤 12: 参数更新. 当系数均值满足此不等式 $|\xi_{\mu i}^{(k+1)'}| \leq \lambda$ 时, 强制该系数均值和方差为零. 然后, 更新稀疏处理后的参数值 $\boldsymbol{\xi}_{\mu}^{(k+1)}$, $\boldsymbol{\xi}_{\delta}^{(k+1)}$.

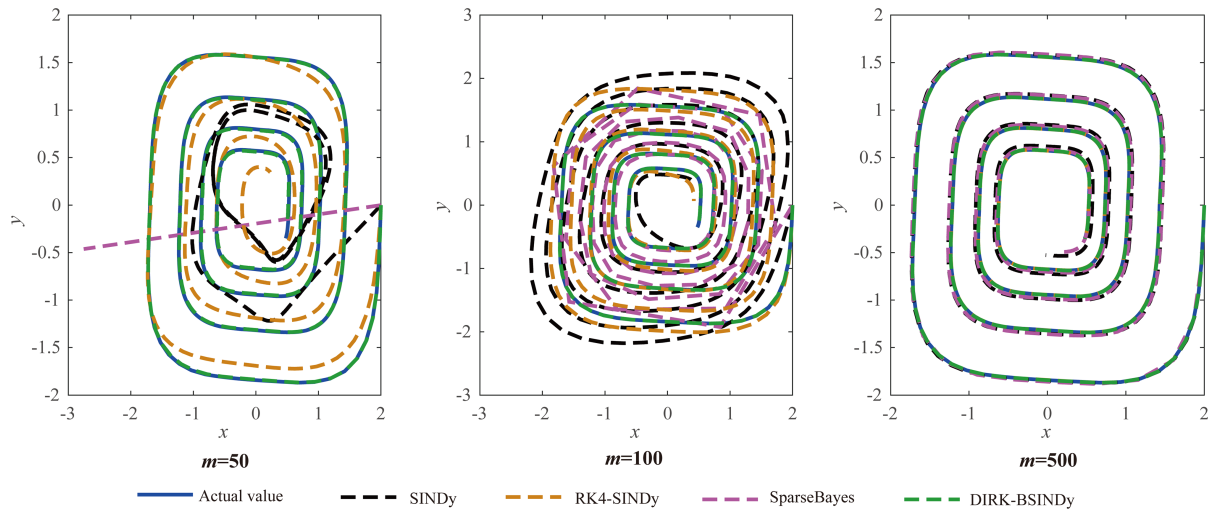


图4 (网络版彩图) 样本量对辨识结果的影响.

Figure 4 (Color online) The influence of sample size on identification results.

步骤 13: 迭代与输出. 更新迭代步数 $k = k + 1$. 若 $k > N$, 则退出训练, 输出系数矩阵均值 ξ_μ 、方差 ξ_δ ; 否则返回步骤 7 继续循环.

说明. 本文直接利用文献 [30~32] 中所述的用于模型选择的信息准则, 通过惩罚参数数量自动选择最优 λ .

4 数值实验

本节将采用不同复杂程度问题进行数值实验, 验证所提出的 DIRK-BSINDy 方法的有效性. 并通过与传统 SINDy [10], SparseBayes [17] 和 RK4-SINDy [18] 对比, 验证该方法对噪声和数据量的鲁棒性. 传统 SINDy 是稀疏辨识经典基准, RK4-SINDy 代表积分型改进方向, SparseBayes 是贝叶斯稀疏辨识前沿方法. 三者覆盖不同技术路线, 能全面对标 DIRK-BSINDy, 凸显其在数据稀缺、噪声环境下的鲁棒性优势. 在 DIRK-BSINDy 方法实现中, 采用 Adam 优化器 [33] 迭代更新 ξ_μ , ξ_δ 和神经网络参数 θ .

4.1 三阶阻尼振荡器

考虑由式 (1) 给出的具有三阶动力学的二维阻尼谐波振荡器, 在初始状态 $[1.0, 0.0]^T$ 下获得一组离线采样数据, 仿真总时间为 20 s. 样本量 m 分别取 500, 100 和 50, 以分析数据量对算法的影响. 在 DIRK-BSINDy 方法中, 设置 DNN 架构为 4 个隐藏层, 每层有 32 个神经元. 此外, 设定阈值 $\lambda = 0.05$, DNN 参数的学习率为 $l_r = 10^{-3}$, 同时系数矩阵 ξ_μ , ξ_δ 的学习率为 10^{-2} . ξ 先验分布中参数设置为 $c = 1$, $\tau_0 = 0.1$. 激活函数采用 SIREN [34]. 为了进行对比分析, 我们使用了与 RK4 算法同阶的 IRK 方法.

图 4 展示了样本量分别为 $m = 50$, $m = 100$, $m = 500$ 时 DIRK-BSINDy, SparseBayes, RK4-SINDy 和 SINDy 的对比结果. 图 5 为不同样本量下各算法估算的基函数对应系数均值分布, 各色块颜色和数值相对应 (见图 5 中色条), 其中稀疏系数均值越接近零色块颜色越接近白色, 系数均值越趋近负值色块越接近深蓝, 越趋近正值色块越接近橙色.

此外, 为了能够量化数据稀缺性对各算法的影响, 在此经过 100 次重复实验设计, 获取各算法在不同样本量下对关键项的辨识准确率. 表 1 中准确率指正确辨识出系统方程关键项的比率, 适配稀疏辨识精准锁定核心动力学项的核心目标.

图 4 和表 1 仿真结果表明, DIRK-BSINDy 方法能够准确识别系统 (1) 的动力学, 而 SINDy 和 RK4-SINDy 方法在面对较小的 m 值时表现出明显困难. 由图 5 可得, 当 $m = 50$ 时, 除 DIRK-BSINDy 外, 其余算法所对应的区域全部色块均有颜色, 表明其构成两个微分方程的所有基函数的系数均有数值, 且颜色越深对应系数数值

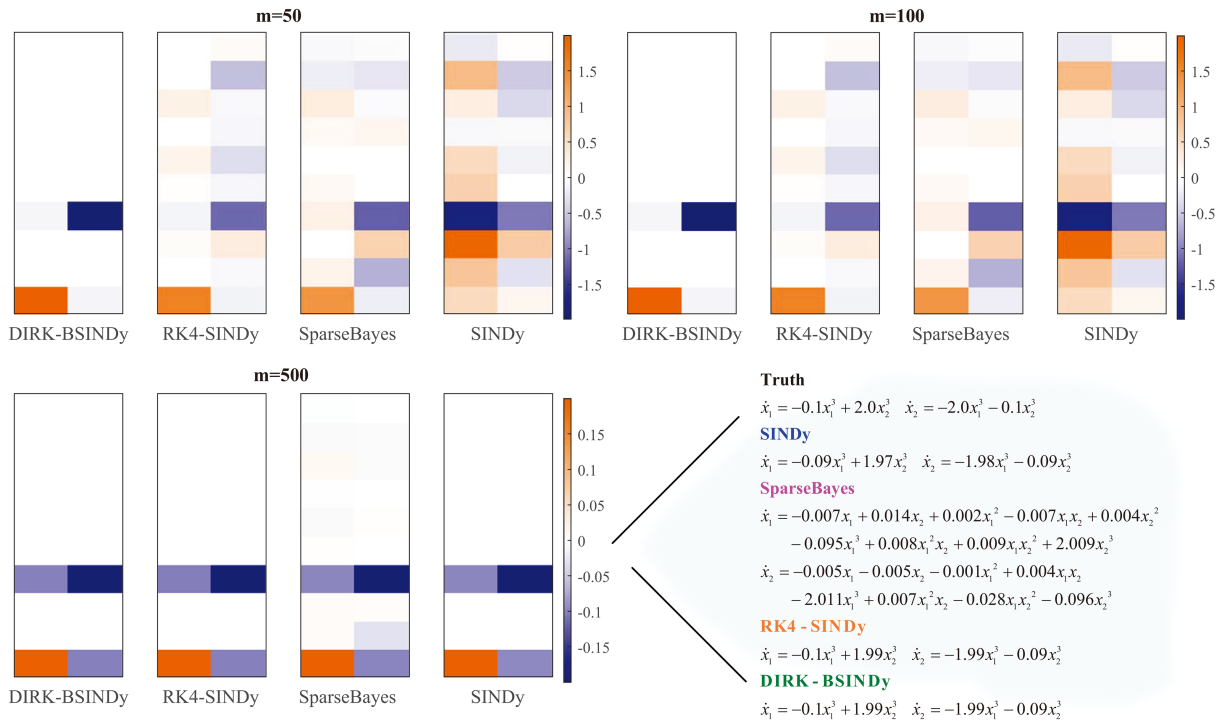


图 5 (网络版彩图) 不同样本量下各算法的辨识系数分布。

Figure 5 (Color online) Distribution of identification coefficients for various algorithms under different sample sizes.

表 1 不同样本量下算法辨识准确率 (%) 分析。

Table 1 Algorithm identification accuracy (%) analysis under different sample sizes.

Algorithm	Accuracy ($m = 50$)	Accuracy ($m = 100$)	Accuracy ($m = 500$)
DIRK-BSINDy	95.7	98.2	100
SparseBayes	72.3	85.1	93.5
RK4-SINDy	Failed	53.2	89.6
SINDy	Failed	Failed	78.2

越大. 这说明在小样本下, 除 DIRK-BSINDy 外, 其余算法均无法有效获取关键项. 随着样本量增加, 其余算法所对应区域中带颜色色块数目减小, 说明了这些算法对高样本信息的依赖性.

基于贝叶斯概率建模, 系数后验方差 ξ_δ 量化系数估计不确定性. 图 6 为 $m = 500$ 时 SparseBayes 和 DIRK-BSINDy 算法获取的稀疏系数分布的图示. 图中每一行对应基函数的系数分布, 其中无分布图的行代表此基函数无系数, 即微分方程中不考虑此基函数. 对于有分布图的行, 则是微分方程中包含此基函数. 并且, 分布图峰值所对应的横坐标代表系数均值 ξ_μ , 分布越集中说明 ξ_δ 越趋近于零, 表明获取的稀疏系数的确定性越强; 分布越分散代表 ξ_δ 越大, 表明不确定性升高.

根据图 5 和 6 可以看出, SparseBayes 算法在任何采样量级下均存在除关键项外的非零项. 这是因为该算法并无稀疏促进结构, 其完全依赖参数后验分布情况进行处理, 即非关键项的后验分布在非零值处达到峰值也会被保留下来. 因此, 相对于 SparseBayes, DIRK-BSINDy 方法使用正则化马蹄先验与变分后验近似设计, 使得关键项系数分布的‘尖峰集中’特征, 证明了双重稀疏约束可精准锁定真实系数, 同时抑制非关键项冗余. 这些对比结果也体现了传统贝叶斯方法缺乏显式稀疏促进结构的局限, 凸显本文贝叶斯建模改进的必要性. 此外, DIRK-BSINDy 在发现可解释方程方面优于 RK4-SINDy, 尤其是在数据稀缺的情况下. 这种能力归因于高斯方法的 A 稳定性特性^[21], 其稳定性区域包含坐标系的整个左半平面, 而 RK4 等显式方法的稳定性区域有限.

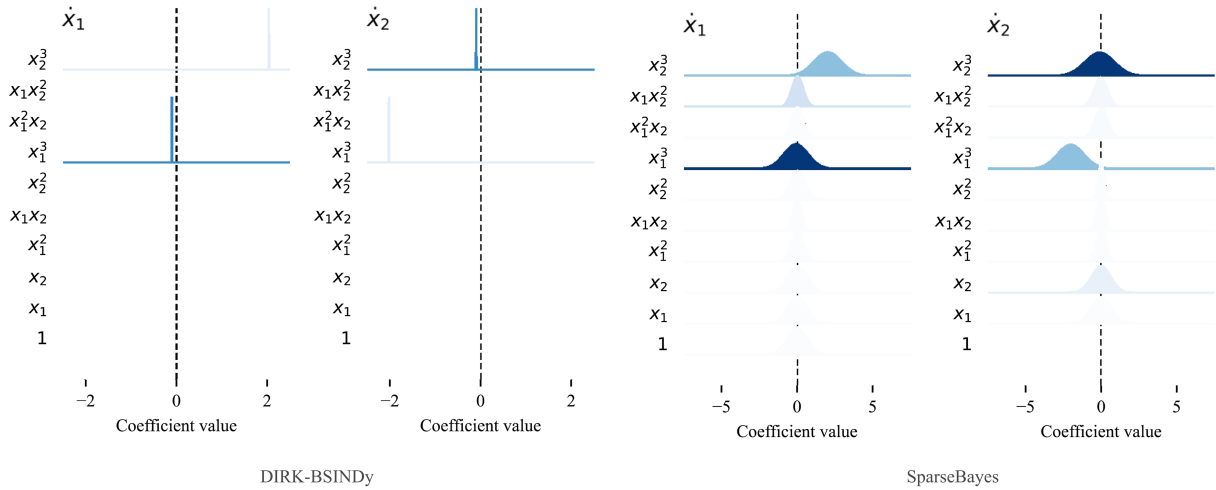


图 6 (网络版彩图) 无噪声下 DIRK-BSINDy 和 SparseBayes 系数后验概率分布 ($m = 500$).

Figure 6 (Color online) Posterior probability distribution of DIRK-BSINDy and SparseBayes coefficients under noiseless conditions ($m = 500$).

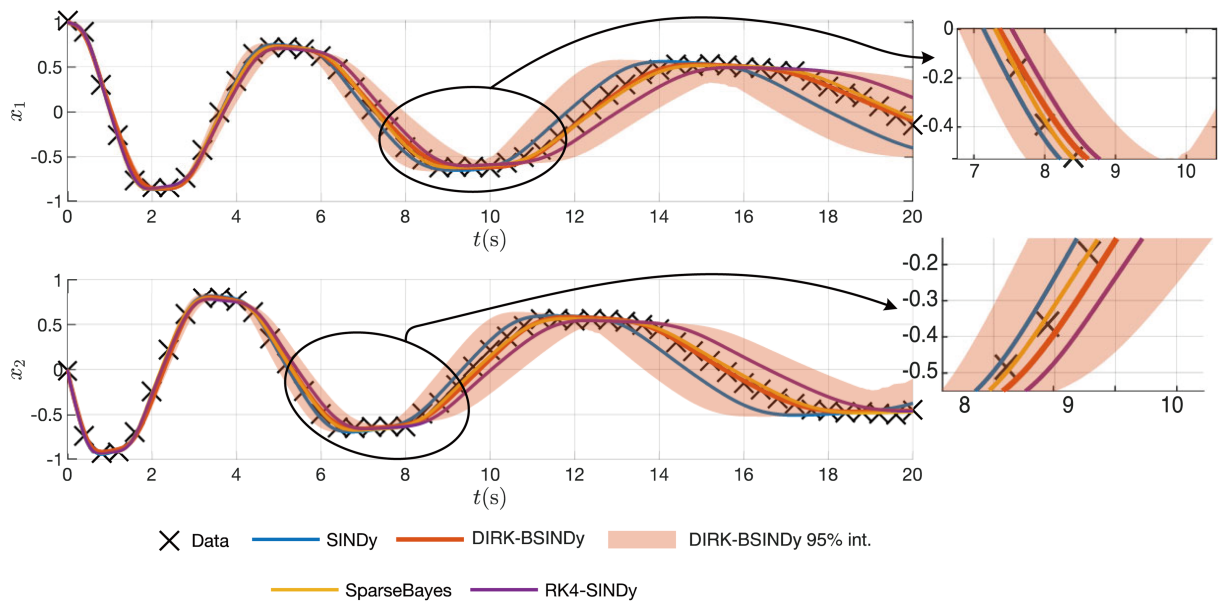


图 7 (网络版彩图) $2\% \cdot \max(|x_{i,\text{ideal}}|)$ 噪声下各算法状态预测曲线.

Figure 7 (Color online) Prediction curves of algorithm states under $2\% \cdot \max(|x_{i,\text{ideal}}|)$ noise.

随后, 我们向具有 500 个样本点的理想采样数据中添加概率密度函数满足式 (2) 的观测噪声, 其均值为 0, 标准差为洁净数据绝对值最大值的 2%, 即 $\sigma_{x,i} = 2\% \cdot \max(|x_{i,\text{ideal}}|)$, 以生成训练数据. 对于噪声水平和初始条件, 我们采用先验 $c = 1$, $\tau_0 = 0.1$. 然后将 SINDy, RK4-SINDy, SparseBayes 和 DIRK-BSINDy 算法应用于数据, 得到的估计系数如表 2 所示, 响应曲线如图 7 所示.

表 2 中第一行为每一维状态微分方程的关键项, SINDy 和 RK4-SINDy 所对应行的数值是关键项的系数值. 由于 SparseBayes 和 DIRK-BSINDy 算法不仅能够获取关键项系数均值, 还可以得出相应方差. 因此 SparseBayes 和 DIRK-BSINDy 所对应行中“ $\pm$ ”之前的数值为稀疏系数均值, 之后的数值是稀疏系数标准差. 图 7 中, 实线表示各算法通过估计参数重建的模型, 阴影区域为 DIRK-BSINDy 利用参数不确定性通过集成重建得到的 95% 置信区间.

表 2 $2\% \cdot \max(|x_{i,\text{ideal}}|)$ 噪声下各算法估算的关键项系数.Table 2 The key term coefficients estimated by various algorithms under $2\% \cdot \max(|x_{i,\text{ideal}}|)$ noise.

	$\dot{x}_1 : x_1^3$	$\dot{x}_1 : x_2^3$	$\dot{x}_2 : x_1^3$	$\dot{x}_2 : x_2^3$
SINDy	-0.09	1.78	-1.94	-0.133
RK4-SINDy	-0.12	1.97	-2.00	-0.11
SparseBayes	-0.09 ± 3.86	1.99 ± 5.18	-2.04 ± 5.17	-0.09 ± 10.41
DIRK-BSINDy	-0.12 ± 0.001	2.00 ± 0.001	-2.02 ± 0.001	$-0.09 \pm 9.78 \times 10^{-4}$

通过表 2 和图 7 所示结果可以得出, SINDy 未能正确估计模型参数值. 事实上, SINDy 模型中没有一个项为零. 由于数据采样稀疏、库的规模较大以及非零系数的幅度范围较大, 这个示例对 SINDy 来说尤其具有挑战性. 虽然 RK4-SINDy 在充足采样数据下获得的系数值接近真实系统, 但是观察图 7 预测曲线, 其与真实系统状态依然存在较大误差. 在表 2 中 DIRK-BSINDy 的 ξ_μ 接近真值, $\xi_\sigma \leq 0.001$ 代表获取的系数确定性强, 反观 SparseBayes 的 ξ_σ 较大, 表明其预测结果的不确定性更高.

为了进一步对比, 我们绘制了图 8. 图 8(a) 中展示了 SparseBayes 和 DIRK-BSINDy 算法估计的系数边际后验分布. 可以看出, DIRK-BSINDy 从含 $2\% \cdot \max(|x|)$ 噪声的测量数据中正确识别了控制方程, 这些结果也反映在表 2 中. 对于 DIRK-BSINDy, 其关键项系数分布呈现窄峰特性, 源于本文高斯 IRK 与辅助 DNN 的融合创新, 即 IRK 避免显式求导放大噪声, DNN 预测阶段值减少积分误差, 二者共同保障噪声下系数估计的确定性. 反观 SparseBayes 虽然识别出了控制方程中的 4 个项, 但是依然存在不以零为中心的其他项分布, 并且所有项的分布平缓, 说明其辨识结果包含的不确定性较强.

另外在图 8(b) 和 (c) 中, 我们展示了 4 种算法在观测噪声为 $2\% \cdot \max(|x|)$ 情况下的误差琴箱图, 以及 DIRK-BSINDy 算法在不同噪声幅值 ($0\%, 5\%, 10\%, 20\%, 30\% \cdot \max(|x|)$) 下的误差琴箱图. 在图 8(b) 和 (c) 的误差琴箱图中, 中位线表示误差绝对值 $|x_i - \hat{x}_i|$ (为 x_i 真实系统的第 i 维状态, $\hat{x}_i$ 为算法估计值) 的均值, 琴箱长度表示误差分布程度 (方差大小). 若误差琴箱图中中位线位于 0 附近且琴箱较短, 则说明预测状态越接近实际系统状态, 且具有较强的抗扰能力. 这一对比结果进一步证明, 本算法的多技术协同框架的抗噪性远超传统算法, 尤其在 20% 高噪声下仍保持窄幅误差, 验证了创新方法对极端场景的适配能力.

4.2 Lorenz 模型

为进一步探究 DIRK-BSINDy 在识别混沌动力学方面的有效性, 我们将非线性三维洛伦兹系统作为下一个示例进行研究. Lorenz 模型为三维混沌动力系统的典型代表, 包含 $x(t)z(t)$, $x(t)y(t)$ 等非线性交叉项, 且具有对条件敏感的混沌特征, 微小初始误差会随时间指数放大, 这与本文面向复杂动力系统无导数建模的核心目标一致, 可验证 DIRK-BSINDy 对强非线性、高动态复杂性系统的辨识能力. 该系统的控制方程如下所示:

$$\begin{aligned}
 \dot{x}(t) &= 10(y(t) - x(t)), \\
 \dot{y}(t) &= x(t)(28 - z(t)) - y(t), \\
 \dot{z}(t) &= x(t)y(t) - \frac{8}{3}z(t).
 \end{aligned} \tag{30}$$

我们使用初始条件 $[x(0), y(0), z(0)]^T = [-1, 6, 15]^T$, 在时间区间内 $t \in [0, 2.5]$ 进行不同样本量 $m \in [30, 260]$ 的测量. DIRK-BSINDy 方法设置参数如下: 采用具有周期性激活函数 SIREN^[34] 的深度神经网络, 该网络包含一个有 256 个神经元的隐藏层, 用于表示 IRK 阶段值. 阈值参数设为 $\lambda = 0.5$. 在训练阶段, 我们为 DNN 参数采用 10^{-3} 的学习率, 为系数矩阵采用 10^{-2} 的学习率, 同时在二次多项式空间中搜索关键项. 另外, 稀疏系数 ξ 先验分布中参数设置为 $c = 1$, $\tau_0 = 0.1$.

接着, 向上述无噪声数据中添加均值为 0、标准差 σ_x 介于 0.025~0.2 之间的高斯测量噪声. 在带噪声测量数

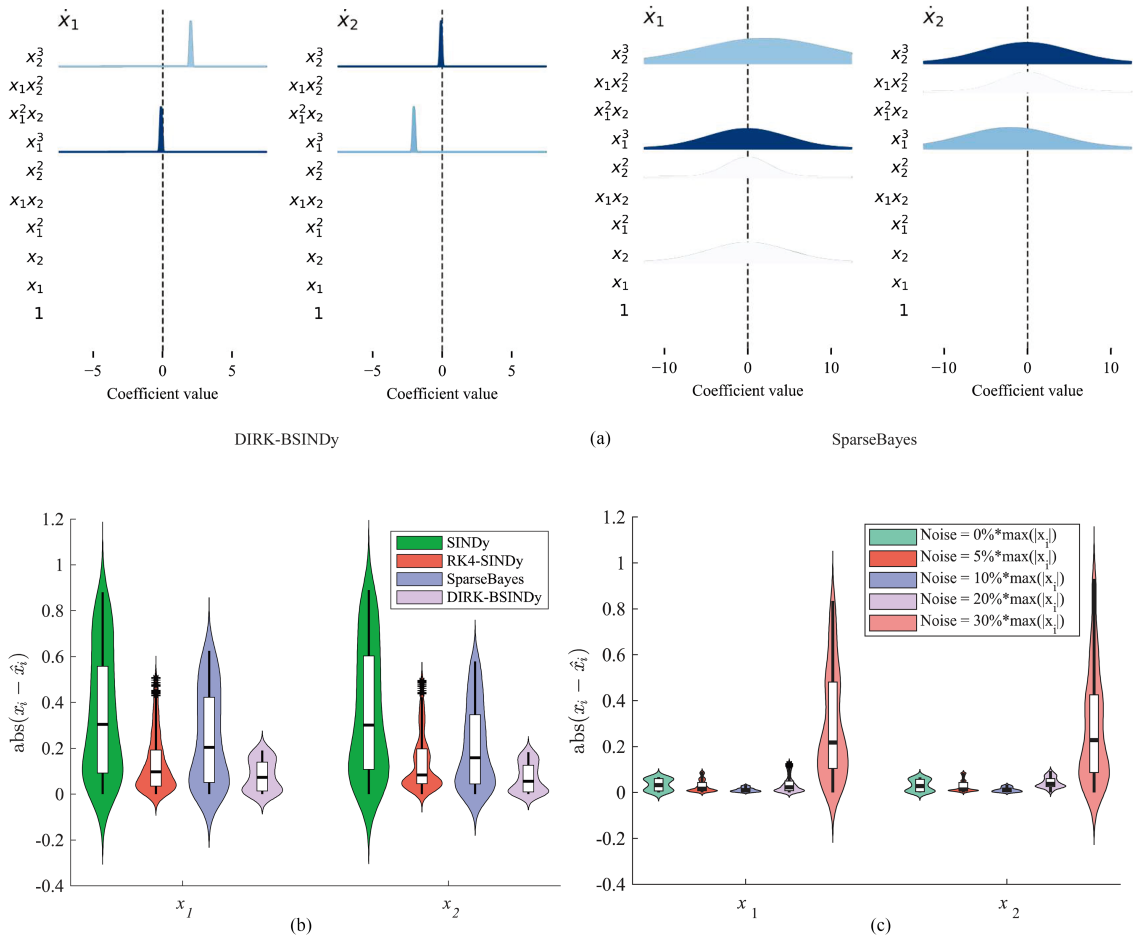


图 8 (网络版彩图) 噪声对各算法影响分析. (a) 系数后验分布; (b) $2\% \cdot \max(|x|)$ 噪声下误差琴箱图; (c) 不同噪声下误差琴箱图.

Figure 8 (Color online) Analysis of the impact of noise on various algorithms. (a) Coefficient posterior distribution; (b) box plot of errors under $2\% \cdot \max(|x|)$; (c) box plot of errors under different noise levels.

据基础上, 计算了 4 种算法在 100 次重复估算下与实际系统的平均相对误差 $\bar{e}$, 其表达式如下所示:

$$\bar{e} = \frac{1}{100} \sum_{j=1}^{100} (e_{\text{single},j}), \quad e_{\text{single}} = \max \left(\frac{1}{m} \sum_{k=1}^m \left| \frac{x_i(t_k) - \hat{x}_i(t_k)}{x_i(t_k)} \right| \middle| i = 1, 2, \dots, d \right),$$

其中, m 为样本量, d 为系统维数, x 是无噪真实系统状态, $\hat{x}$ 为辨识模型预测状态. 该指标通过相对误差消除不同系统状态量级差异的影响, 通过最大值统计确保模型在所有维度均满足精度要求, 相较于绝对误差更具通用性. 为了显示更加清楚, 我们最终展示的曲线值为 $1 - \bar{e}$, 即误差越小越接近 1, 越大越接近 0, 结果如图 9 所示. 其中图 9(d) 将训练数据的时间区间延长至 $t \in [0, 10]$.

根据图 9 可以得出, SINDy 方法随着噪声增加, 其性能迅速下降; RK4-SINDy 虽然辨识误差随噪声幅值变化不大, 但是整体性能不佳. 相比之下, DIRK-BSINDy 和 SparseBayes 在噪声增大时表现出更强的鲁棒性. 在图 9(c) 中, 当噪声水平达到 $\sigma_x = 0.2$ 时, 4 种算法对模型恢复的性能都较差. 虽然随着训练数据量增加, 各算法预测性能有所提升 (图 9(d)), 但是相比之下 DIRK-BSINDy 的性能更好. 对比图 9(c) 和 (d) 也验证了 DIRK-BSINDy 能够在低数据量情况下保持较好性能.

此外, 为了进一步分析噪声影响辨识结果的原因, 我们绘制了图 10. 在图 10 中, 标识 “x” 表示 $28x - xz - \dot{y}$ 值, 其应与 y (黄线) 匹配. 然而, x (红线) 与 y 高度共线, 并且不明显哪些函数更适合数据. 这表明在短时间区间 $t \in [0, 2.5]$ 内, 噪声严重掩盖了信号, 无法充分采样这些区域. 实际上, 当数据量过少或噪声过高时, 数据中的

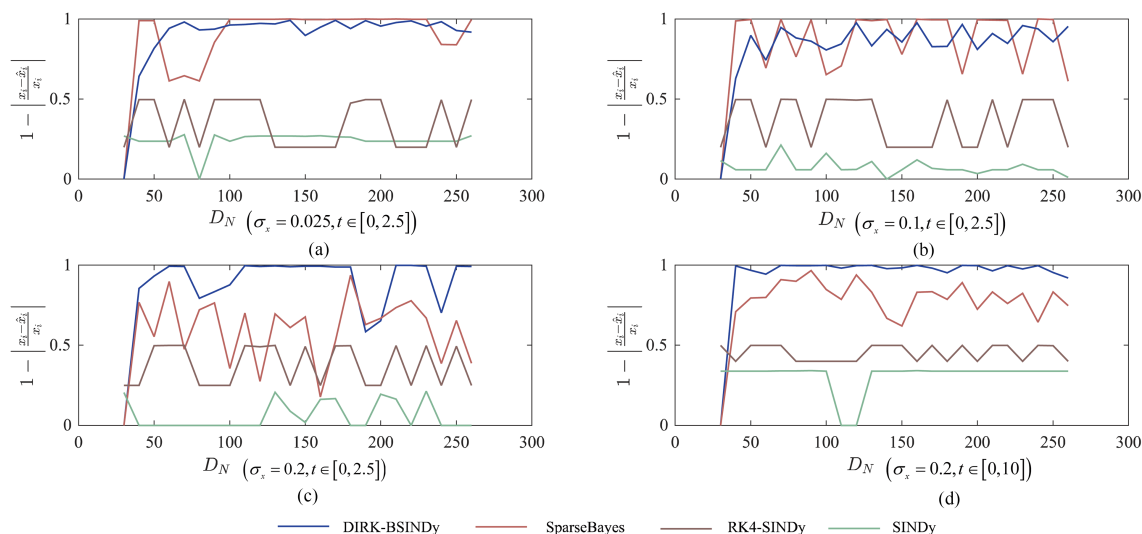


图 9 (网络版彩图) 性能分析. (a) $\sigma_x = 0.025$ 噪声标准差下的系统相对误差; (b) $\sigma_x = 0.1$ 噪声标准差下的系统相对误差; (c) $\sigma_x = 0.2$ 噪声标准差下的系统相对误差; (d) 延长训练时域后的系统相对误差.

Figure 9 (Color online) Performance analysis. (a) System relative error at $\sigma_x = 0.025$ noise; (b) system relative error at $\sigma_x = 0.1$ noise; (c) system relative error at $\sigma_x = 0.2$ noise; (d) system relative error after extending the training horizon.

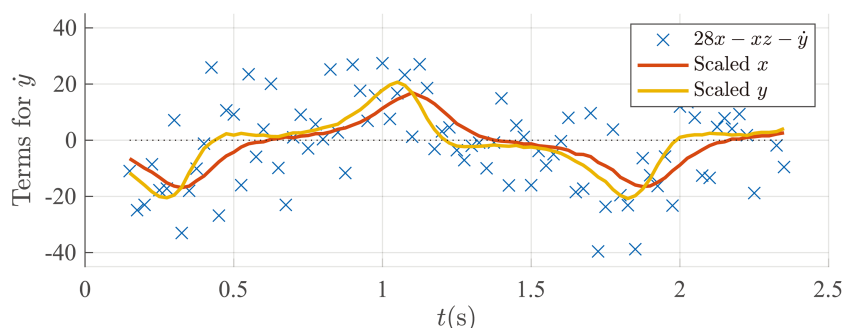


图 10 (网络版彩图) 噪声影响 (x 和 y 都被缩放以匹配数据 $28x - xz - \dot{y}$, $m = 100$, $\sigma_x = 0.2$).

Figure 10 (Color online) Noise impact (both x and y are scaled to match the data $28x - xz - \dot{y}$, $m = 100$, $\sigma_x = 0.2$).

信息可能不足以推断完整的动力学. 此时, 必须通过延长时间区间 (如 $t \in [0, 10]$) 来增加可用数据量. 如图 9(d) 所示, DIRK-BSINDy 在延长时间区间后大多能恢复完整模型, 而传统 SINDy 在高噪声情况下仍表现较差. 这一结果也表明, 若能在信息最丰富的状态下采样数据, 可进一步提高数据效率.

最后, 为验证 DIRK-BSINDy 算法的计算效率优势, 表 3 对各辨识算法的计算量进行了对比分析. 表 3 中 m 是样本数据量, d 是系统状态维数, M 代表基函数数目, E 是模型训练的总迭代次数, n_h 是神经元数目, L 是隐藏层层数, d_{in} 是 DNN 输入层维数, d_{out} 是 DNN 输出层维数, T 为 MCMC 采样次数.

根据上述仿真结果和表 3 中各算法计算复杂度分析可以得出, 若场景为无噪声且数据充足, SINDy 可作为优先选择; 若场景为含噪声或数据稀缺, DIRK-BSINDy 优于 RK4-SINDy (鲁棒性弱) 与 SparseBayes (计算量高). DIRK-BSINDy 通过适度增加计算量 (引入 DNN 与 IRK) 换取抗噪声、数据高效的核心性能, 避免了盲目追求低计算量导致鲁棒性缺失或过度强调鲁棒性导致计算量激增的极端情况, 符合工程应用中复杂度-性能平衡的核心需求. 此外, 当系统维度从 $d = 2$ (三阶振荡器) 增至 $d = 3$ (Lorenz 系统) 时, DIRK-BSINDy 的计算量仅增幅 32 s, 而 SparseBayes 增幅达 187 s (基函数数目 M 随系统维度增加而增加), 验证了 DIRK-BSINDy 在高维系统中的可扩展性优势.

表 3 各算法计算量对比分析.

Table 3 Comparative analysis of computational load of each algorithm.

	SINDy ^[10]	RK4-SINDy ^[18]	SparseBayes ^[17]	DIRK-BSINDy ^[33, 35]
Core computational modules	1. Derivative estimation 2. Sparse regression	1. Explicit RK4 2. Sparse regression	1. MCMC sampling 2. Likelihood function calculation	1. DNN training 2. Gaussian IRK integration 3. Bayesian variational optimization
Theoretical complexity	1. Derivative: $O(m \times d^2)$ 2. Sparse: $O(E \times m \times d \times M)$	1. RK4: $O(m \times d \times 4)$ 2. Sparse: $O(E \times m \times d \times M)$	1. MCMC: $O(T \times m \times d \times M)$ 2. Likelihood: $O(m \times d \times M)$	1. DNN: $O(E \times m \times 2[(d_{in} + d_{out})n_h + (L - 1)n_h^2])$ 2. IRK: $O(m \times s \times d)$ 3. Bayesian: $O(E \times (m \times d + M \times d))$
Computational order	$O(10^7)$	$O(10^7)$	$O(10^8)$	$O(10^8)$
Third-order damped oscillator $m = 500$	15 s	41 s	165 s	178 s
Lorenz system $m = 500$	32 s	57 s	352 s	210 s
Optimizability	Difficult to optimize	Parallel evaluation available	Reducible sampling frequency	GPU parallel acceleration supported

5 结论

本文提出了一种新的数据驱动框架, 将高斯隐式龙格-库塔方法与贝叶斯稀疏识别相结合, 并使用深度神经网络辅助求解 IRKs 阶段值提高计算效率. 该框架为融合求解微分方程的数值方法、稀疏回归和深度学习这 3 项关键技术开辟了新路径, 可直接从数据集对物理现象进行建模. 此外, 由于该算法所获得的模型能够估算模型不确定性, 因此可以将该模型用于传统或强化学习控制中, 提供除状态预测值之外的不确定信息, 使得控制在优化跟踪精度的同时对方差信息进行优化.

这项研究不仅获得了较好的结果, 而且也引发了诸多值得进一步探究的问题. 例如, 哪种神经网络架构最适合特定数据集? 如何使用 DIRK-BSINDy 框架有效识别含参数的动力系统? 均方误差是否是损失函数的最佳选择? 鉴于神经网络架构的最新进展, 在数据稀缺且噪声水平较高的场景中, 如何开发保持鲁棒性的算法? 在后续的研究中, 我们计划通过采用计算成本更低且实现过程与数据无关的替代高阶隐式方法来扩展所提出的框架, 促进其与稀疏识别的融合; 并探索自适应基函数库 (如物理引导的稀疏约束) 和轻量化 DNN 架构以优化计算效率, 提高算法对更高维系统及时变系统的适应性. 尽管 DIRK-BSINDy 通过贝叶斯不确定性量化和隐式积分对扰动具有内在鲁棒性, 但是针对强外部扰动 (如脉冲干扰) 和时变系统误差依然需要设计专用策略进行解决. 此外, 本工作的加噪实验遵循系统辨识领域对算法鲁棒性的标准评估流程, 未来还可以探索非高斯、非平稳噪声场景下的扩展, 并融合实际工业数据验证泛化性. 因此, 在未来的研究中还需要进一步拓展其在复杂实际场景中的适用性.

参考文献

- 1 Brunton S L, Zolman N, Kutz J N, et al. Machine learning for sparse nonlinear modeling and control. *Ann Rev Control Robot Autonom Syst*, 2025, 8: 1–24
- 2 Zhao F R, You K Y. Survey of recent progress in data-driven policy optimization for controller design. *Sci Sin Inform*, 2023, 53: 1027–1049 [赵斐然, 游科友, 数据驱动的策略优化控制律设计最新研究综述. *中国科学: 信息科学*, 2023, 53: 1027–1049]
- 3 Cortés-Ciriano I, Gulhan D C, Lee J J K, et al. Computational analysis of cancer genome sequencing data. *Nat Rev Genet*, 2022, 23: 298–314

- 4 Li Q K, Gao X F, Peng C, et al. Data-driven change control design for product and supply chain synchronous evolution systems under cyber-attacks. *Sci Sin Inform*, 2023, 53: 325–343 [李庆奎, 高雪峰, 彭晨, 等. 网络攻击下产品与供应链协同演进系统数据驱动变更控制设计. *中国科学: 信息科学*, 2023, 53: 325–343]
- 5 Zhao M, He W, Tang J, et al. A hybrid deep learning framework for gene regulatory network inference from single-cell transcriptomic data. *Brief BioInf*, 2022, 23: bbab568
- 6 Jiao S, Zou Q, Guo H, et al. iTTCA-RF: a random forest predictor for tumor T cell antigens. *J Transl Med*, 2021, 19: 1
- 7 Park S H, Ha S, Kim J K. A general model-based causal inference method overcomes the curse of synchrony and indirect effect. *Nat Commun*, 2023, 14: 4287
- 8 Wang H J, Zhao H T, Ren B Q, et al. Cyber-physical framework for UAV intelligent communications. *Sci Sin Inform*, 2022, 52: 2141–2154 [王海军, 赵海涛, 任保全, 等. 一种面向无人机智能通信的信息物理融合框架. *中国科学: 信息科学*, 2022, 52: 2141–2154]
- 9 Rackauckas C, Ma Y, Martensen J, et al. Universal differential equations for scientific machine learning. 2020. ArXiv:2001.04385
- 10 Brunton S L, Proctor J L, Kutz J N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc Nat Acad Sci*, 2016, 113: 3932–3937
- 11 Fasel U, Kutz J N, Brunton B W, et al. Ensemble-SINDy: robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proc R Soc A*, 2022, 478: 20210904
- 12 North J S, Wikle C K, Schliep E M. A review of data-driven discovery for dynamic systems. *Int Stat Rev*, 2023, 91: 464–492
- 13 Schaeffer H, McCalla S G. Sparse model selection via integral terms. *Phys Rev E*, 2017, 96: 023302
- 14 Messenger D A, Bortz D M. Weak sindy: Galerkin-based data-driven model selection. *Multi Model Simul*, 2021, 19: 1474–1497
- 15 Raissi M, Perdikaris P, Karniadakis G E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys*, 2019, 378: 686–707
- 16 Weng R H, Hao K R, Chen L, et al. Koopman feature kernel-based time-frequency causal and delay inference network for industrial systems. *Acta Autom Sin*, 2025, 51: 1829–1841 [翁若昊, 郝矿荣, 陈磊, 等. 基于 Koopman 特征核的工业时频因果时延推理网络. *自动化学报*, 2025, 51: 1829–1841]
- 17 Tipping M E, Faul A C. Fast marginal likelihood maximisation for sparse bayesian models. In: *Proceedings the 9th International Workshop on Artificial Intelligence and Statistics*, 2003. 276–283
- 18 Goyal P, Benner P. Discovery of nonlinear dynamical systems using a Runge-Kutta inspired dictionary-based sparse regression approach. *Proc R Soc A*, 2022, 478: 20210883
- 19 Hairer E, Wanner G, Nsett S P. *Runge-kutta and Extrapolation Methods*. Berlin: Heidelberg, 1993. 129–353
- 20 Chen H. Data-driven sparse identification of nonlinear dynamical systems using linear multistep methods. *Calcolo*, 2023, 60: 11
- 21 Butcher J C. *Numerical Methods for Ordinary Differential Equations*. London: Springer, 2016. 1615–2085
- 22 Sato S, Miyatake Y, Butcher J C. High-order linearly implicit schemes conserving quadratic invariants. *Appl Numer Math*, 2023, 187: 71–88
- 23 Butcher J C. Implicit Runge-Kutta processes. *Math Comp*, 1964, 18: 50–64
- 24 Lal B S T, Murari K, Vijay K. Half cauchy exponential geometric distribution: model, properties and application. *Interd J Manage Soc Sci*, 2022, 3: 96–110
- 25 Wu J J, Zhao J Y, Sun C J, et al. Predictive resource allocation: unsupervised learning of Markov decision processes. *Sci Sin Inform*, 2024, 54: 1983–2000 [吴佳骏, 赵剑羽, 孙乘坚, 等. 预测资源分配: 马尔可夫决策过程的无监督学习. *中国科学: 信息科学*, 2024, 54: 1983–2000]
- 26 Nihat A, Jesse V O. On the fisher-rao gradient of the evidence lower bound. 2023. ArXiv:2307.11249
- 27 Kingma D P, Welling M. Auto-encoding variational Bayes. 2014. ArXiv:1312.6114
- 28 Xu H, Zhang D, Wang N. Deep-learning based discovery of partial differential equations in integral form from sparse and noisy data. *J Comput Phys*, 2021, 445: 110592
- 29 Liu X, Wang D. Universal consistency of deep ReLU neural networks. *Sci Sin Inform*, 2024, 54: 638–652 [刘霞, 王迪. 深度 ReLU 神经网络的万有一致性. *中国科学: 信息科学*, 2024, 54: 638–652]
- 30 Mangan N M, Kutz J N, Brunton S L, et al. Model selection for dynamical systems via sparse regression and information criteria. *Proc Royal Soc A Math Phys Eng Sci*, 2017, 473: 20170009
- 31 Dong X, Bai Y L, Lu Y, et al. An improved sparse identification of nonlinear dynamics with Akaike information criterion and group sparsity. *Nonlinear Dyn*, 2023, 111: 1485–1510
- 32 Kaptanoglu A A, Zhang L, Nicolaou Z G, et al. Benchmarking sparse system identification with low-dimensional chaos. *Nonlinear Dyn*, 2023, 111: 13143–13164
- 33 Kingma D P, Ba J. Adam: a method for stochastic optimization. 2014. ArXiv:1412.6980
- 34 Sitzmann V, Martel J, Bergman A, et al. Implicit neural representations with periodic activation functions. In: *Proceedings of*

Advances in Neural Information Processing Systems, 2020. 7462–7473

35 Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge: MIT Press, 2016. 200–215

Sparse coefficient distribution identification of control equations based on high-order integration

Rui CAO^{1,2*} & Yang YI¹

1. School of Information and Artificial Intelligence, Yangzhou University, Yangzhou 225101, China

2. National Key Laboratory of Autonomous Intelligent Unmanned Systems, Beijing Institute of Technology, Beijing 100081, China

* Corresponding author. E-mail: stdio@yzu.edu.cn

Abstract Identifying the governing equations of physical systems from datasets has long been a persistent challenge across various scientific disciplines. The sparse identification of nonlinear dynamics method typically relies on accurate derivative estimation, which makes it prone to failure in data-scarce and noisy environments. To enhance the robustness of identification, this paper proposes a novel data-driven approach, DIRK-BSINDy, that combines high-order integration (Gauss implicit Runge-Kutta method) with Bayesian sparse identification, and leverages deep neural networks to assist in solving the stage values of high-order integration, thereby improving computational efficiency. The proposed method obtains the posterior distribution of sparse coefficients in the governing equations via Bayesian inference and constructs a loss function by integrating forward and backward predictions to enhance stability. Furthermore, owing to the weak step-size constraint inherent in high-order integration, DIRK-BSINDy exhibits remarkable robustness even in scenarios with scarce data and strong noise interference, demonstrating distinct advantages over traditional methods.

Keywords data-driven learning, system identification, data robustness, implicit Runge-Kutta