

Runge-Kutta Gill与Euler框架下SINDy动力学建模

云飞阳

北方工业大学理学院数学系, 北京

收稿日期: 2026年6月28日; 录用日期: 2026年7月20日; 发布日期: 2026年7月29日

摘要

近年来, 龙格-库塔方法与深度学习已被验证可从观测数据中高效学习动力系统控制方程, 为数据驱动动力学建模提供了有效途径。在此背景下, 本文将基于字典的稀疏学习策略与两种经典的数值离散方法相结合, 针对稀疏采样、含噪数据开展微分方程的高精度还原研究。工作在Brunton等提出的非线性动力学稀疏辨识(Sparse Identification of Nonlinear Dynamics, SINDy)框架上进行改进与扩展, 分别构造了Runge-Kutta Gill SINDy (以下简称RK-Gill SINDy)与Euler SINDy两类辨识算法。通过采用稳定的数值积分格式替代传统有限差分近似, 所提方法能够显著削弱损毁数据和噪声对辨识精度的影响, 表现出优于经典SINDy的模型重构能力, 具备良好的稳定性与噪声鲁棒性。

关键词

Runge-Kutta Gill, Euler方法, 稀疏回归, SINDy

SINDy Dynamic Modeling under the Runge-Kutta Gill and Euler Frameworks

Feiyang Yun

Department of Mathematics, College of Science, North China University of Technology, Beijing

Received: June 28, 2026; accepted: July 20, 2026; published: July 29, 2026

Abstract

In recent years, Runge-Kutta methods and deep learning have been proven to efficiently learn the governing equations of dynamic systems from observational data, providing an effective approach for data-driven dynamic modeling. Against this background, this paper combines dictionary-based sparse learning strategies with two classic numerical discretization methods to conduct high-precision

recovery research on differential equations under sparse sampling and noisy data. Based on the Sparse Identification of Nonlinear Dynamics (SINDy) framework proposed by Brunton *et al.*, this work makes improvements and extensions, and constructs two identification algorithms: Runge-Kutta Gill SINDy (hereinafter referred to as RK-Gill SINDy) and Euler SINDy. By adopting stable numerical integration schemes to replace the traditional finite difference approximation, the proposed methods can significantly reduce the impact of corrupted data and noise on identification accuracy, exhibit better model reconstruction capability than the classic SINDy, and possess satisfactory stability and noise robustness.

Keywords

Runge-Kutta Gill, Euler Method, Sparse Regression, SINDy

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着应用数学与深度学习的迅速发展[1]-[3]，数据驱动的动力系统发现方法受到了广泛关注[4]-[7]。

在数据科学与动力学辨识的交叉研究中，基于稀疏约束的非线性动力学辨识方法为从观测数据中自动识别微分方程模型提供了有效途径。SINDy 借助稀疏正则化手段，在预设的非线性基函数库中筛选出对系统起主导作用的动力学项，从而实现动力系统的可解释性建模，有效弥补了传统深度学习等黑箱模型难以揭示物理机理的不足[8]。

当前，SINDy 相关研究正从基础算法设计向面向复杂场景的实用化方向不断推进。经典 SINDy 方法依靠稀疏正则化实现方程结构辨识，在低维、低噪声环境下表现良好[5]，但在强噪声、数据稀疏或强非线性系统中，辨识精度与鲁棒性往往难以满足要求。为改善这一问题，弱形式 SINDy 被相继提出[9]。弱形式 SINDy (Weak SINDy)是 SINDy 研究分支中最具代表性的改进方向之一。该方法通过引入测试函数空间与广义矩匹配思想，将原 SINDy 中对状态导数的直接估计问题转化为对系统积分形式的稀疏回归，从而规避了有限差分法在含噪数据中导数估计误差被放大的缺陷，显著提升了算法对噪声的鲁棒性与辨识精度。然而，弱形式 SINDy 也存在固有局限：一方面，其测试函数空间的选择对辨识结果影响较大，缺乏统一的最优设计准则；另一方面，该方法需要对积分项进行大规模数值计算，带来较高的计算开销。

除弱形式框架外，将高阶数值积分格式与 SINDy 相结合的改进思路也逐渐成为研究热点。已有研究表明，这类方法通过利用数值积分过程中的多阶段中间信息，能够同时优化导数估计精度与动力学结构辨识效果，在非线性动力学建模中展现出显著优势[10]。但现有基于龙格-库塔(Runge-Kutta, RK)格式的 SINDy 方法，多采用传统的显式 RK 格式(如 RK4)，其系数构造未充分考虑稀疏辨识场景下的数值稳定性与误差累积特性，在长时间积分或低采样率场景下，仍可能出现辨识精度下降、参数估计偏差增大的问题。

与上述方法相比，本文提出的 RK-Gill SINDy 方法在理论与实践层面均具有明确优势：其一，RK-Gill 格式是专门针对长时间数值模拟优化的积分方案，其在保持高阶精度的同时，具备更优的数值稳定性，能够有效抑制低采样率下的误差累积，这是传统 RK 格式与弱形式 SINDy 难以兼顾的；其二，相较于弱形式 SINDy 的复杂积分与测试函数构造，RK-Gill SINDy 的计算流程更为简洁，无需额外的测试函数空间设计，在保证精度的同时降低了计算复杂度，更易拓展至大规模系统；其三，该方法直接在数值积分

框架内实现导数信息的高效利用,克服了标准 SINDy 对噪声敏感的缺陷,在噪声、数据稀疏的双重挑战下展现出更强的鲁棒性与辨识精度。

本文在不同数据采样频率与噪声干扰下开展系统仿真实验,验证了 RK-Gill [11]在辨识精度与数值稳定性上的优越性。同时引入基于欧拉法[12]的 Euler SINDy 方法,将其与 RK-Gill SINDy 及标准 SINDy 进行全面对比。实验结果表明,在多个典型动力系统及噪声干扰条件下, RK-Gill SINDy 均能实现更为精准的动力学方程识别。

本研究基于 RK-Gill 方法与 Euler 方法,分别构建了对应的 SINDy 辨识框架。

2. Runge-Kutta Gill 和 Euler 方法简介

简要回顾这两种数值离散格式的基本原理,考虑如下形式的微分方程初值问题:

$$\dot{x}(t) = f(x(t)), x(t_0) = x_0, \quad (1)$$

式中 $x(t) := [x_1(t), x_2(t), \dots, x_n(t)]$, 其中 $x_j(t)$ 是向量 $x(t)$ 的第 j 个元素, 函数 $f: R^n \rightarrow R^n$ 为定义在状态空间上的向量场。

设时间节点 $t_{k+1} = t_k + h_k$, 若从已知状态 $x(t_k)$ 预测未来时刻 $x(t_{k+1})$, 其中 $k \in \{0, 1, \dots, N\}$ 。RK-Gill 方法和 Euler 方法可表示出 $x(t_{k+1})$, 涉及时间步长 h_k 和向量场信息 $f(\cdot)$ 的乘积。

RK-Gill 方法的数值迭代格式表示如下

$$x(t_{k+1}) \approx x(t_k) + \frac{1}{6} h_k (k_1 + (2 - \sqrt{2})k_2 + (2 + \sqrt{2})k_3 + k_4), h_k = t_{k+1} - t_k, \quad (2)$$

其中,

$$\begin{aligned} k_1 &= f(x(t_k)), k_2 = f\left(x(t_k) + \frac{1}{2} h_k k_1\right), \\ k_3 &= f\left(x(t_k) + \left(\frac{1}{\sqrt{2}} - \frac{1}{2}\right) h_k k_1 + \left(1 - \frac{1}{\sqrt{2}}\right) h_k k_2\right), \\ k_4 &= f\left(x(t_k) - \frac{1}{\sqrt{2}} h_k k_2 + \left(1 + \frac{1}{\sqrt{2}}\right) h_k k_3\right). \end{aligned} \quad (3)$$

同样也可使用欧拉法由 $x(t_k)$ 预测 $x(t_{k+1})$, 欧拉法的数值格式为

$$x(t_{k+1}) \approx x(t_k) + h_k k_1, h_k = t_{k+1} - t_k, \quad (4)$$

其中,

$$k_1 = f(x(t_k)). \quad (5)$$

若从 t_0 到 t_f 对方程(1)进行积分,可采用一系列时间步长 h_k , $k \in \{1, \dots, N\}$, 满足 $t_f = t_0 + \sum_{i=0}^N h_k$ 。后文为表述简洁,引入以下记号: $F_{RK-Gill}(f, x(t_k), h_k)$ 和 $F_{Euler}(f, x(t_k), h_k)$, 即对于(2)和(4)中的步骤,分别有

$$x(t_{k+1}) = x(t_k + h_k) \approx F_{RK-Gill}(f, x(t_k), h_k), \quad (6)$$

以及

$$x(t_{k+1}) = x(t_k + h_k) \approx F_{Euler}(f, x(t_k), h_k). \quad (7)$$

分别表示采用 RK-Gill 格式和欧拉格式由当前状态 $x(t_k)$ 及步长 h_k 计算预测下一步状态 $x(t_{k+1})$ 。

3. 非线性动力系统的识别

我们的目标是用状态变量 $x(t)$ 的时间序列数据来揭示式(1)中向量场 $f(x(t))$ 的最简洁表达式。假设数据在时间点 $\{t_0, t_1, \dots, t_N\}$ 处采集, 定义时间步长 $h_k = t_{k+1} - t_k$ 。基于此, 构造两个数据矩阵:

$$X := \begin{bmatrix} x(t_1) \\ x(t_2) \\ \vdots \\ x(t_N) \end{bmatrix} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \cdots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_N) & x_2(t_N) & \cdots & x_n(t_N) \end{bmatrix}, X_F(f) := \begin{bmatrix} F_{RK-Gill}(f, x(t_0), h_1) \\ F_{RK-Gill}(f, x(t_1), h_2) \\ \vdots \\ F_{RK-Gill}(f, x(t_{N-1}), h_N) \end{bmatrix}. \quad (8)$$

$$X := \begin{bmatrix} x(t_1) \\ x(t_2) \\ \vdots \\ x(t_N) \end{bmatrix} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \cdots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_N) & x_2(t_N) & \cdots & x_n(t_N) \end{bmatrix}, X_F(f) := \begin{bmatrix} F_{Euler}(f, x(t_0), h_1) \\ F_{Euler}(f, x(t_1), h_2) \\ \vdots \\ F_{Euler}(f, x(t_{N-1}), h_N) \end{bmatrix}. \quad (9)$$

下一步是要构建一个包含潜在的非线性特征的符号字典 Θ 。假设函数 $f(\cdot)$ 可由字典中少量项的线性组合表达。

例如, 构造一个多项式函数的字典, 对于任意给定的向量 $u := [u_1, \dots, u_n]$, 该字典可以设计成如下形式

$$\Theta(u) = [1, u, u^{q_2}, u^{q_3}, \dots], \quad (10)$$

其中 u^{q_i} ($i \in \{2, 3\}$) 为高阶多项式, 例如 u^{q_3} 包含 u 元素的所有可能的 3 次多项式

$$u^{q_3} = [u_1^3, u_1^2 u_2, u_1 u_2^2, u_2^3, u_2^2 u_3, \dots, u_n^3]. \quad (11)$$

字典 Θ 中的每一项都可作为描述函数 f 的潜在候选基函数, 我们的目标是从中选取尽量少的候选特征, 来准确刻画(1)中的非线性函数 f 。因此建立一个促进稀疏性的优化问题, 引入稀疏向量 $\chi_k \in R^p$ (其中 p 是字典 Θ 中特征的总数), 使得

$$f_k(x(t)) = \Theta(x(t)) \chi_k, \quad (12)$$

其中 $f_k: R^n \rightarrow R$ 是 f 的第 k 个元素, 因此, 从字典中选择合适的候选特征决定了控制方程。进一步, 可将式(1)中的函数 $f(\cdot)$ 写成如下:

$$f(x) = [f_1(x), f_2(x), \dots, f_n(x)] = [\Theta(x) \chi_1, \Theta(x) \chi_2, \dots, \Theta(x) \chi_n] = \Theta(x) \Psi, \quad (13)$$

式中 $\Psi = [\chi_1, \dots, \chi_n]$ 。基于该表示, 则识别控制方程的问题就转化为稀疏优化问题。即寻找最稀疏的系数矩阵 Ψ , 满足

$$X = X_F(f), \quad (14)$$

这里 $f(x) = \Theta(x) \Psi$, 其中 X 为状态矩阵, $X_F(f)$ 为基于所选数值格式的进一步预测矩阵, 如(8)和(9)所示。

一旦通过优化确定或者得到稀疏向量, 则动力学系统可显式表示为

$$[x_1(t), x_2(t), \dots, x_n(t)] = [\Theta(x) \chi_1, \Theta(x) \chi_2, \dots, \Theta(x) \chi_n].$$

本文将提出的基于 RK-Gill 格式的方法称为 RK-Gill SINDy, 将基于欧拉格式的方法称为 Euler SINDy。

4. 实验结果分析

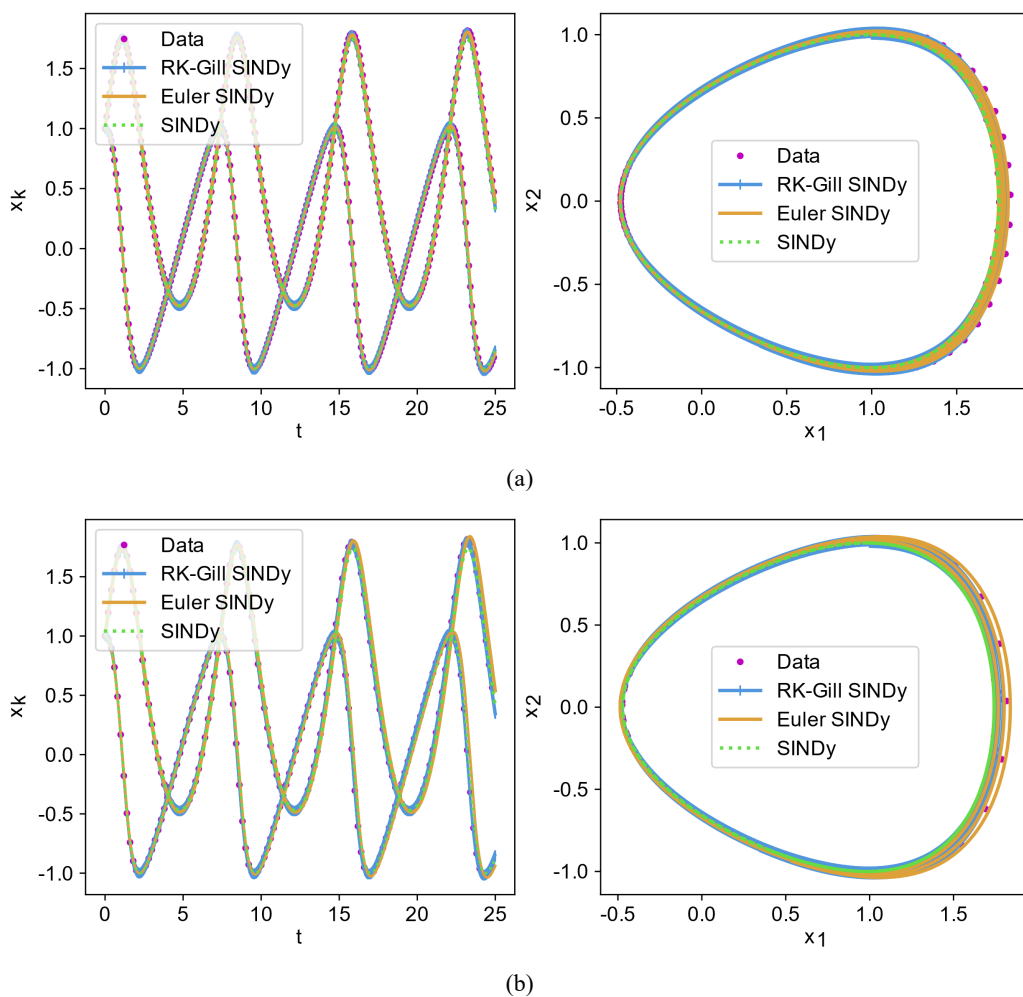
4.1. Sprott 多项式系统

其基本形式可由以下方程组描述[13]

$$\begin{cases} \dot{x}(t) = y(t), \\ \dot{y}(t) = -x(t) + y(t)^2. \end{cases} \quad (15)$$

为探究数据稀疏性对模型辨识过程的影响机制,本研究设置多组不同时间步长($\Delta t = 0.1, 0.2, 0.3, 0.4$),构建了系列稀疏采样数据集,并分别采用 RK-Gill SINDy、Euler SINDy 与标准 SINDy 方法进行控制方程辨识。结果表明,在所有稀疏采样场景下, RK-Gill SINDy 方法的辨识参数均与真实模型参数高度吻合,能够较准确重构目标系统的动力学方程。此外,两种基于数值积分的 SINDy 变体(RK-Gill SINDy 与 Euler SINDy)在动力学方程恢复效果与结构一致性方面,均显著优于传统标准 SINDy 方法,这也验证了数值积分策略对稀疏采样数据动力学建模的有效性与优越性。

图 1 直观呈现了不同采样步长下, RK-Gill SINDy、Euler SINDy 与标准 SINDy 辨识轨迹与真实轨迹的对比结果。随着时间步长增大(数据稀疏性增强),标准 SINDy 方法的轨迹偏差逐渐扩大,而 RK-Gill SINDy 仍能保持与真实轨迹高度重合,展现出极强的轨迹重构稳定性。为进一步量化三种方法在结构



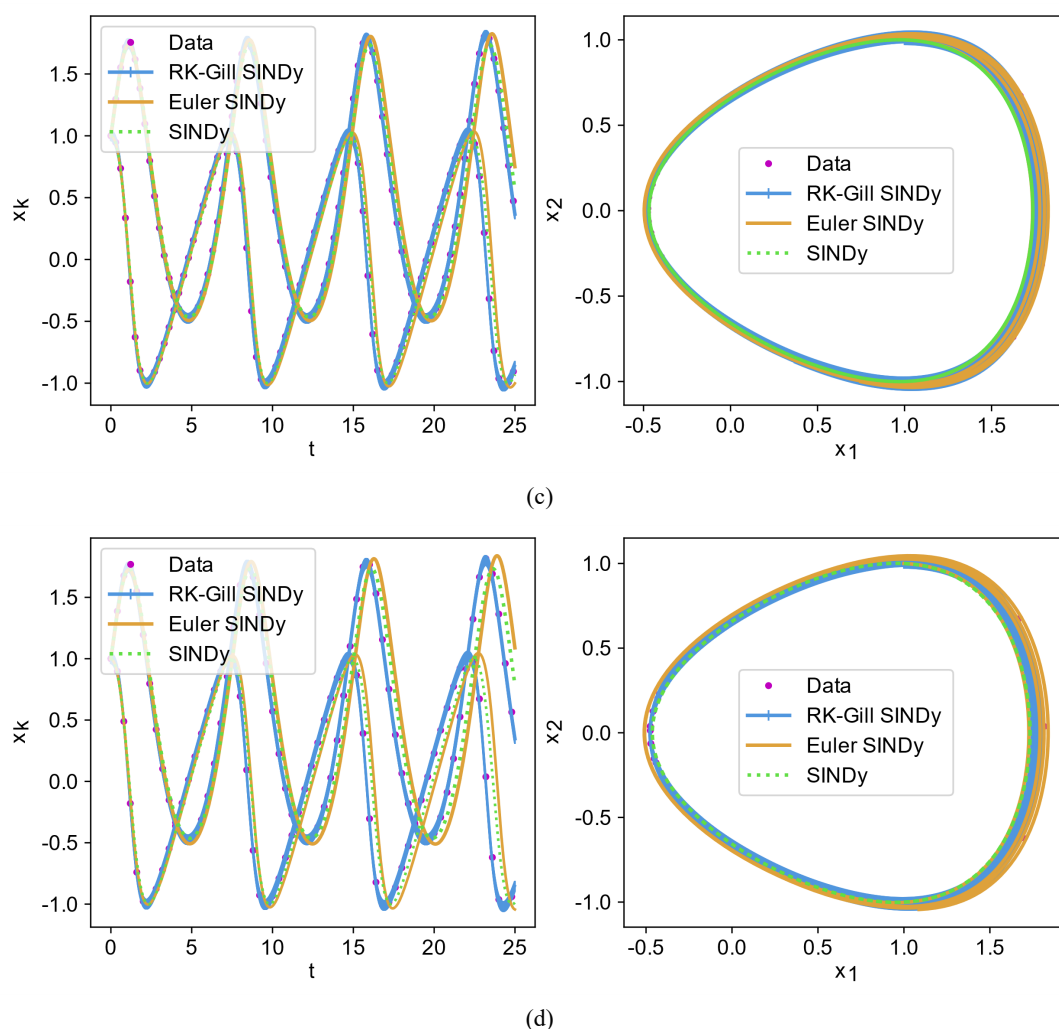


Figure 1. Sprott polynomial system: Comparison of model trajectories identified by RK-Gill SINDy, Euler SINDy and standard SINDy with the true trajectories under different time step sizes: (a) $\Delta t = 1 \times 10^{-1}$; (b) $\Delta t = 2 \times 10^{-1}$; (c) $\Delta t = 3 \times 10^{-1}$; (d) $\Delta t = 4 \times 10^{-1}$

图 1. Sprott 多项式系统: 不同时间步长下 RK-Gill SINDy、Euler SINDy 与标准 SINDy 方法辨识的模型轨迹与真实轨迹对比: (a) $\Delta t = 1 \times 10^{-1}$; (b) $\Delta t = 2 \times 10^{-1}$; (c) $\Delta t = 3 \times 10^{-1}$; (d) $\Delta t = 4 \times 10^{-1}$

辨识与参数估计上的性能差异, 表 1 和图 2 给出了不同时间步长下, 各方法重构的控制方程与真实模型的对比结果, 从结构一致性与参数精度两个维度验证了 RK-Gill SINDy 在稀疏采样场景下的优越性。

由表 1 的重构方程对比可见, 随着数据稀疏性增强, RK-Gill SINDy 方法始终能够完整还原系统的动力学结构, 且参数估计值与真实模型高度吻合; 相比之下, Euler SINDy 在步长增大后参数偏差明显增大, 而标准 SINDy 则出现了结构误判, 引入了无关常数项与交叉项, 模型一致性遭到破坏。表 2 的定量误差结果进一步验证了这一趋势: RK-Gill SINDy 在所有步长下均保持了极低数量级的误差, 且误差随步长变化的增幅很小, 展现出非常好的数值稳定性; Euler SINDy 的误差随步长呈明显上升趋势, 在 $\Delta t = 4 \times 10^{-1}$ 时误差已达到 10^{-2} 量级; 标准 SINDy 的误差增幅最为显著, 在大时间步长下几乎失去辨识能力。上述结果表明, RK-Gill 格式对积分误差的抑制作用, 使其在稀疏采样场景下仍能保持高精度的结构辨识与参数重构能力, 显著优于另外两种提到的方法。

Table 1. Sprott polynomial system: Comparison of governing equations reconstructed by RK-Gill SINDy, Euler SINDy and standard SINDy under different time step sizes

表 1. Sprott 多项式系统：不同时间步长下 RK-Gill SINDy, Euler SINDy 与标准 SINDy 方法还原的控制方程对比

Δt	RK-Gill SINDy	Euler SINDy	SINDy
1×10^{-1}	$\dot{x}(t) = 1.000y(t)$	$\dot{x}(t) = 0.998y(t)$	$\dot{x}(t) = 0.998y(t)$
	$\dot{y}(t) = -1.000x(t)$	$\dot{y}(t) = -0.996x(t)$	$\dot{y}(t) = -0.997x(t)$
	$+0.997y(t)^2$	$+0.991y(t)^2$	$+0.991y(t)^2$
2×10^{-1}	$\dot{x}(t) = 1.000y(t)$	$\dot{x}(t) = 0.993y(t)$	$\dot{x}(t) = 0.993y(t)$
	$\dot{y}(t) = -1.000x(t)$	$\dot{y}(t) = -0.986x(t)$	$\dot{y}(t) = 0.044 - 0.919x(t)$
	$+0.997y(t)^2$	$+0.971y(t)^2$	$-0.046x(t)^2 + 0.903y(t)^2$
3×10^{-1}	$\dot{x}(t) = 1.000y(t)$	$\dot{x}(t) = 0.984y(t)$	$\dot{x}(t) = 0.986y(t)$
	$\dot{y}(t) = -1.000x(t)$	$\dot{y}(t) = -0.968x(t)$	$\dot{y}(t) = 0.093 - 0.831x(t)$
	$+0.997y(t)^2$	$+0.941y(t)^2$	$-0.094x(t)^2 + 0.800y(t)^2$
4×10^{-1}	$\dot{x}(t) = 1.000y(t)$	$\dot{x}(t) = 0.972y(t)$	$\dot{x}(t) = 0.977y(t)$
	$\dot{y}(t) = -1.001x(t)$	$\dot{y}(t) = -0.944x(t)$	$\dot{y}(t) = 0.155 - 0.719x(t)$
	$+0.997y(t)^2$	$+0.898y(t)^2$	$-0.155x(t)^2 + 0.670y(t)^2$

Table 2. Sprott polynomial system: Errors of governing equations reconstructed by RK-Gill SINDy, Euler SINDy and standard SINDy under different time step sizes

表 2. Sprott 多项式系统：不同时间步长下 RK-Gill SINDy, Euler SINDy 和标准 SINDy 还原控制方程的误差

Δt	RK-Gill SINDy	Euler SINDy	SINDy
1×10^{-1}	9.00×10^{-6}	1.01×10^{-4}	9.40×10^{-5}
2×10^{-1}	9.00×10^{-6}	1.09×10^{-3}	2.01×10^{-2}
3×10^{-1}	9.00×10^{-6}	4.76×10^{-3}	8.62×10^{-2}
4×10^{-1}	1.00×10^{-5}	1.43×10^{-2}	2.36×10^{-1}

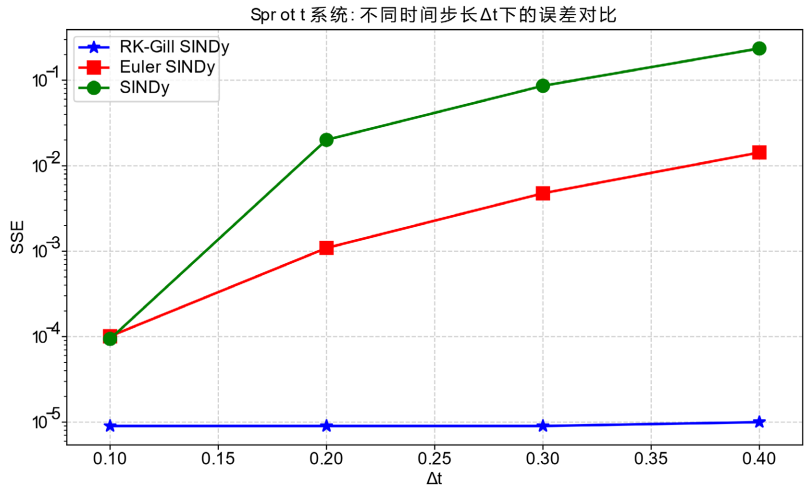


Figure 2. Sprott polynomial system: Error plots of the reconstructed models by RK-Gill SINDy, Euler SINDy and standard SINDy under different time step sizes

图 2. Sprott 多项式系统：不同时间步长下 RK-Gill SINDy、Euler SINDy 和标准 SINDy 方法还原模型的误差图

4.2. 非线性振子模型

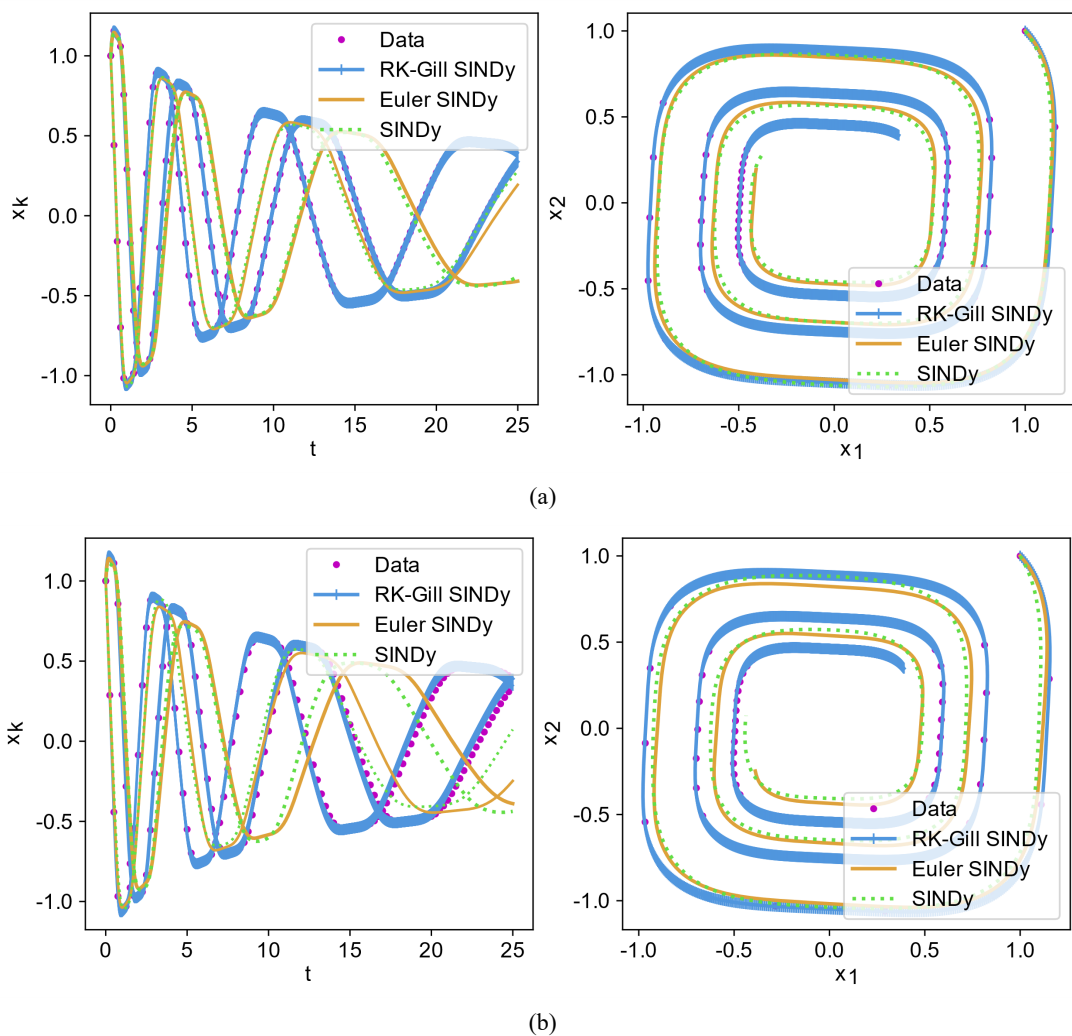
考虑如下模型[5]

$$\begin{cases} \dot{x}(t) = -0.1x(t)^3 + 2y(t)^3, \\ \dot{y}(t) = -2x(t)^3 - 0.1y(t)^3. \end{cases} \quad (16)$$

本研究构建了包含三次多项式项的非线性符号字典，并基于 RK-Gill SINDy 算法实现了系统控制方程的稀疏辨识。为全面评估该方法的鲁棒性，在多种时间步长条件下，对其辨识精度进行了系统测试，并与 Euler SINDy 及标准 SINDy 方法展开对比分析。

从理论层面分析，这一显著差异的根源在于 RK-Gill 积分格式对离散误差的控制能力，通过多阶段的斜率估计与加权平均，将局部截断误差提升至四阶，能够在稀疏采样时更精准地捕捉系统的非线性演化，大幅抑制了由离散化引入的偏差，从而保证了后续稀疏回归步骤中字典矩阵的可靠性，最终实现了控制方程的高精度辨识。

因此综合考虑图3和图4的轨迹稳定性，以及表3和表4的重构方程和误差，在较大时间步长下鲁棒性更好的 RK-Gill SINDy，比 Euler SINDy 和标准 SINDy 更适用于描述非线性振子类型的非线性动力学系统。



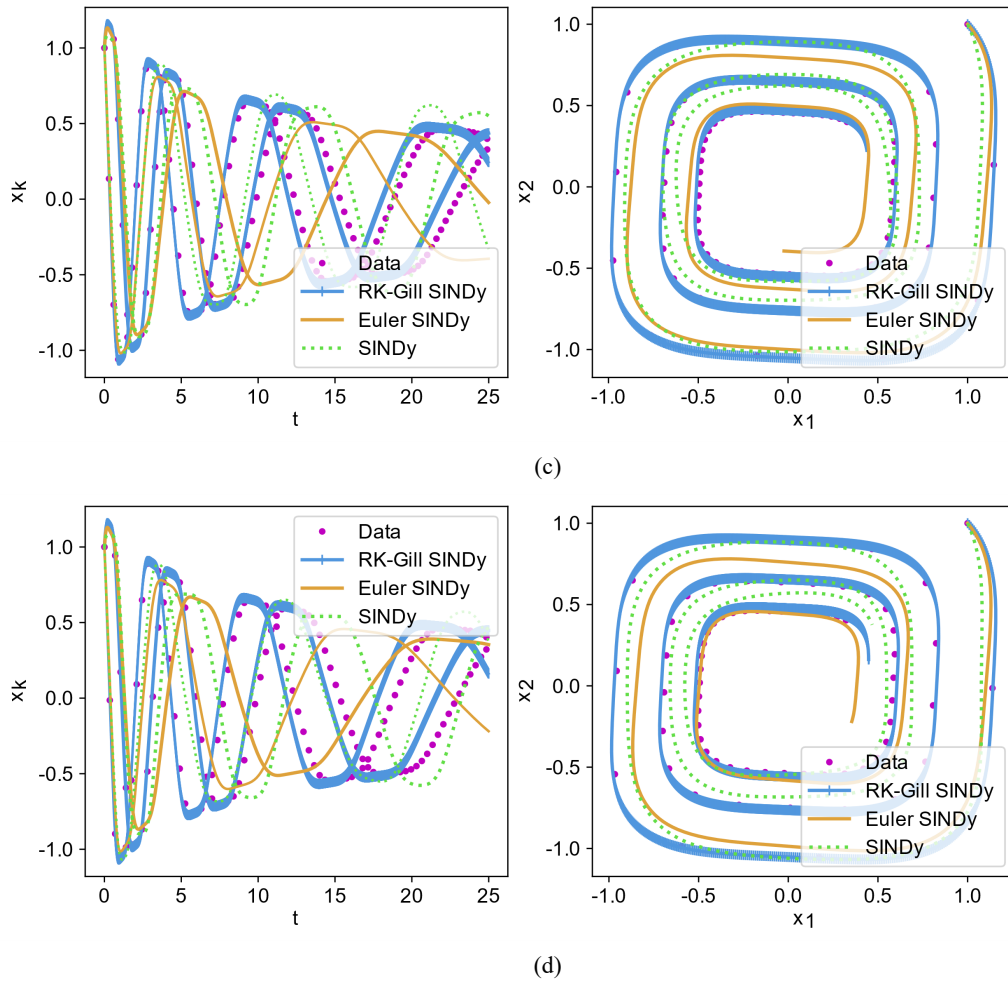


Figure 3. Nonlinear oscillator model: Comparison of model trajectories identified by RK-Gill SINDy, Euler SINDy and standard SINDy with the real trajectories under different time step sizes(Δt): (a) $\Delta t = 2 \times 10^{-1}$; (b) $\Delta t = 2.5 \times 10^{-1}$; (c) $\Delta t = 3 \times 10^{-1}$; (d) $\Delta t = 3.5 \times 10^{-1}$

图 3. 非线性振子模型：不同时间步长(Δt)下由 RK-Gill SINDy、 Euler SINDy 与标准 SINDy 方法辨识的模型轨迹与真实轨迹对比：(a) $\Delta t = 2 \times 10^{-1}$ ；(b) $\Delta t = 2.5 \times 10^{-1}$ ；(c) $\Delta t = 3 \times 10^{-1}$ ；(d) $\Delta t = 3.5 \times 10^{-1}$

Table 3. Nonlinear oscillator model: Comparison of governing equations reconstructed by RK-Gill SINDy, Euler SINDy and standard SINDy under different time step sizes

表 3. 非线性振子模型：在不同时间步长下，RK-Gill SINDy， Euler SINDy 与标准 SINDy 方法还原的控制方程对比

Δt	RK-Gill SINDy	Euler SINDy	SINDy
2×10^{-1}	$\dot{x}(t) = -0.099x(t)^3$	$\dot{x}(t) = -0.113x(t)^3$	$\dot{x}(t) = 0.081y(t) - 0.075x(t)^3$
	$+2.000y(t)^3$	$+1.900y(t)^3$	$-0.122x(t)y(t)^2 + 1.778y(t)^3$
	$\dot{y}(t) = -2.001x(t)^3$	$\dot{y}(t) = -1.965x(t)^3$	$\dot{y}(t) = -1.903x(t)^3 - 0.090x(t)^2 y(t)$
	$-0.101y(t)^3$	$-0.115y(t)^3$	$-0.224x(t)y(t)^2 - 0.098y(t)^3$
2.5×10^{-1}	$\dot{x}(t) = -0.097x(t)^3$	$\dot{x}(t) = -0.123x(t)^3$	$\dot{x}(t) = 0.122y(t) - 0.082x(t)y(t) - 0.058y(t)^2$
	$+2.005y(t)^3$	$+1.853y(t)^3$	$-0.066x(t)^3 - 0.177x(t)y(t)^2 + 1.661y(t)^3$

续表

	$\dot{y}(t) = -2.004x(t)^3$ $-0.101y(t)^3$	$\dot{y}(t) = -1.954x(t)^3$ $-0.120y(t)^3$	$\dot{y}(t) = -0.082x(t)y(t) - 1.866x(t)^3$ $-0.124x(t)^2y(t) - 0.326x(t)y(t)^2 - 0.088y(t)^3$
3×10^{-1}	$\dot{x}(t) = -0.094x(t)^3$ $+2.008y(t)^3$	$\dot{x}(t) = -0.146x(t)^3$ $+1.805y(t)^3$	$\dot{x}(t) = 0.095x(t) + 0.141y(t) - 0.101x(t)y(t)$ $-0.060y(t)^2 - 0.164x(t)^3 - 0.315x(t)y(t)^2 + 1.601y(t)^3$
	$\dot{y}(t) = -2.008x(t)^3$ $-0.100y(t)^3$	$\dot{y}(t) = -1.933x(t)^3$ $-0.115y(t)^3$	$\dot{y}(t) = -0.110x(t)y(t) - 1.827x(t)^3$ $-0.274x(t)^2y(t) - 0.410x(t)y(t)^2$
3.5×10^{-1}	$\dot{x}(t) = -0.096x(t)^3$ $+2.014y(t)^3$	$\dot{x}(t) = -0.144x(t)^3$ $+1.764y(t)^3$	$\dot{x}(t) = 0.149y(t) - 0.127x(t)y(t)$ $-0.158x(t)^2y(t) - 0.312x(t)y(t)^2 + 1.636y(t)^3$
	$\dot{y}(t) = -2.007x(t)^3$ $-0.095y(t)^3$	$\dot{y}(t) = -1.932x(t)^3$ $-0.148y(t)^3$	$\dot{y}(t) = 0.091y(t) - 0.120x(t)y(t) - 1.820x(t)^3$ $-0.318x(t)^2y(t) - 0.443x(t)y(t)^2 - 0.175y(t)^3$

Table 4. Nonlinear oscillator model: SSE of governing equations reconstructed by RK-Gill SINDy, Euler SINDy and standard SINDy under different time step sizes

表 4. 非线性振子模型：不同时间步长下，RK-Gill SINDy, Euler SINDy 与标准 SINDy 方法还原的控制方程的 SSE

Δt	RK-Gill SINDy	Euler SINDy	SINDy
2×10^{-1}	3.00×10^{-6}	1.16×10^{-2}	1.39×10^{-1}
2.5×10^{-1}	5.10×10^{-5}	2.47×10^{-2}	3.19×10^{-1}
3×10^{-1}	1.64×10^{-4}	4.49×10^{-2}	6.00×10^{-1}
3.5×10^{-1}	2.86×10^{-4}	6.46×10^{-2}	6.61×10^{-1}

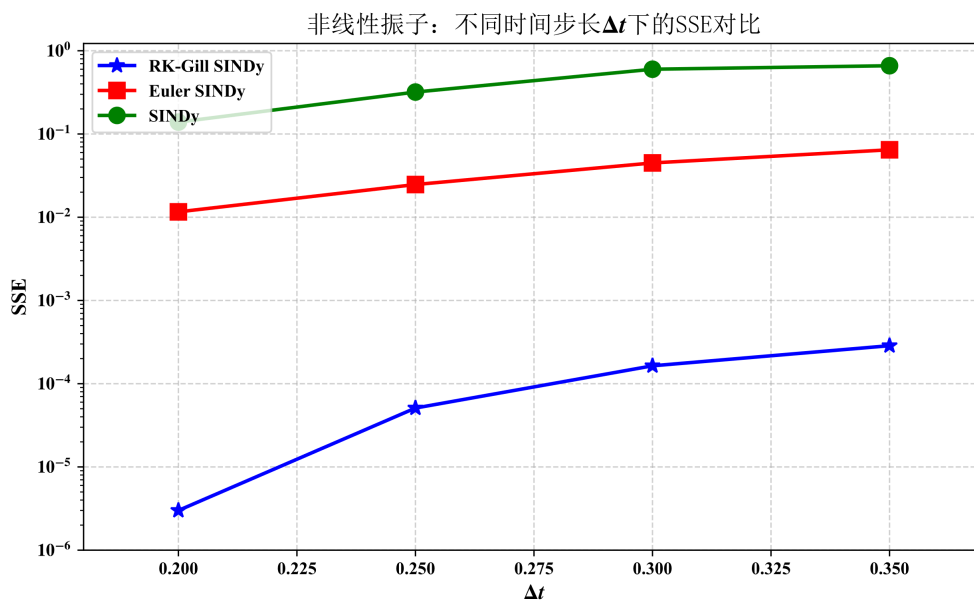


Figure 4. Nonlinear oscillator model: Error comparison of reconstructed models obtained by RK-Gill SINDy, Euler SINDy and standard SINDy under different time step sizes

图 4. 非线性振子模型：不同时间步长下 RK-Gill SINDy、Euler SINDy 与标准 SINDy 方法还原模型的误差对比图

4.3. Halvorsen 模型

模型形式如下[14]

$$\begin{cases} \dot{x}(t) = -1.4x - 4y - 4z - y^2, \\ \dot{y}(t) = -1.4y - 4z - 4x - z^2, \\ \dot{z}(t) = -1.4z - 4x - 4y - x^2. \end{cases} \quad (17)$$

在噪声水平为 0.2 的情况下, 本部分对这三种方法进行比较研究, 其识别结果以及相应的误差如下图表所示。可以看出, RK-Gill SINDy 方法得到的模型误差明显小于 Euler SINDy 法和标准 SINDy 法, 这也符合前面两种模型学习后得到的结论, 进一步说明了 RK-Gill SINDy 方法在含有噪声的数据上也具有很好的鲁棒性和准确性。

Table 5. Halvorsen model: Comparison of governing equations reconstructed by RK-Gill SINDy, Euler SINDy and standard SINDy under the noise level of 0.2

表 5. Halvorsen 模型: 噪声水平为 0.2 时, RK-Gill SINDy, Euler SINDy 和标准 SINDy 方法还原的控制方程对比

σ	<i>RK-Gill SINDy</i>	<i>Euler SINDy</i>	<i>SINDy</i>
0.2	$\dot{x}(t) = -1.395x(t) - 3.977y(t) - 3.974z(t) - 0.992y(t)^2$	$\dot{x}(t) = -1.395x(t) - 3.972y(t) - 3.965z(t) - 0.990y(t)^2$	$\dot{x}(t) = -1.398x(t) - 3.966y(t) - 3.963z(t) - 0.991y(t)^2$
	$\dot{y}(t) = -3.975x(t) - 1.387y(t) - 3.917z(t) - 0.987z(t)^2$	$\dot{y}(t) = -3.967x(t) - 1.386y(t) - 3.912z(t) - 0.984z(t)^2$	$\dot{y}(t) = -3.971x(t) - 1.388y(t) - 3.921z(t) - 0.986z(t)^2$
	$\dot{z}(t) = -3.974x(t) - 3.997y(t) - 1.401z(t) - 0.995x(t)^2$	$\dot{z}(t) = -3.967x(t) - 3.986y(t) - 1.400z(t) - 0.992x(t)^2$	$\dot{z}(t) = -0.220 - 3.978x(t) - 3.995y(t) - 1.417z(t) - 0.991x(t)^2$

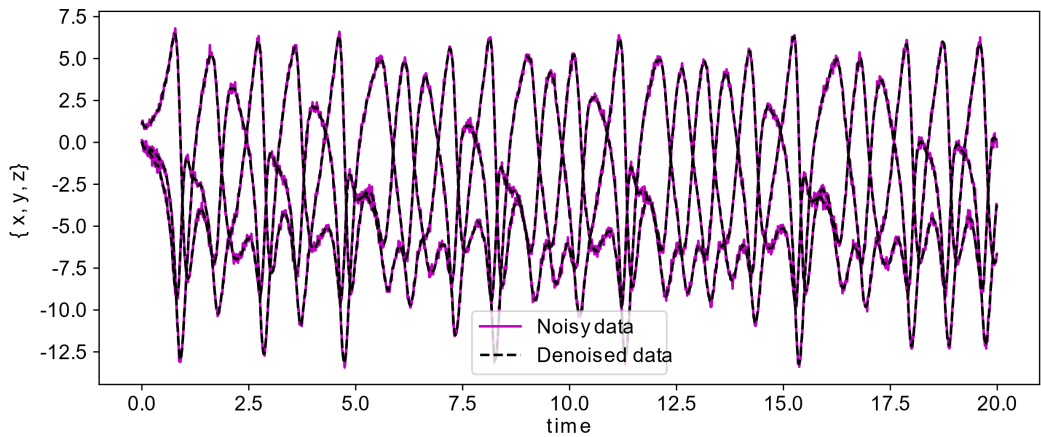


Figure 5. Halvorsen model: The observed samples generated by adding Gaussian white noise with a standard deviation of 0.2 to the original data are represented by dashed lines, which are compared with the denoised results processed by the Savitzky-Golay filtering algorithm shown by solid lines. The time series of the state variables $\{x, y, z\}$ are presented respectively in the figure

图 5. Halvorsen 模型: 将原始数据叠加 0.2 标准差的高斯白噪声生成的观测样本(以虚线表示), 与采用 Savitzky-Golay 滤波算法处理过的去噪结果(以实线呈现)进行对比分析, 图中分别展示了状态变量 $\{x, y, z\}$ 的时间序列

图 5 为添加 0.2 高斯白噪声后的时间序列, 表 5 为该噪声下三种方法学习到的控制方程结构, 进一步观察发现, RK-Gill SINDy 方法重构的相轨迹表现出显著优势。从图 6 中可以清晰看到, 在高斯白噪声

干扰下, RK-Gill SINDy 重构的轨道则与真实轨迹高度重合, 重构出的混沌轨道与实际轨道最为接近, 在大噪声的情况下, 其对正确轨迹的偏离最小, 对于混沌吸引子形状的复现也是最好的, 能够完整复现 Halvorsen 吸引子的复杂拓扑结构。这一现象与表 6 中的误差数据完全吻合: RK-Gill SINDy 的方程重构误差仅为 9.86×10^{-3} , 远低于 Euler SINDy 的 1.28×10^{-2} 和标准 SINDy 的 5.93×10^{-2} 。这充分说明 RK-Gill 积分框架带来的高阶数值稳定性, 能够有效抑制噪声在辨识过程中的放大效应, 使方法在强噪声环境下依然保持优异的动力学特征捕捉能力, 展现出远优于对比方法的鲁棒性。

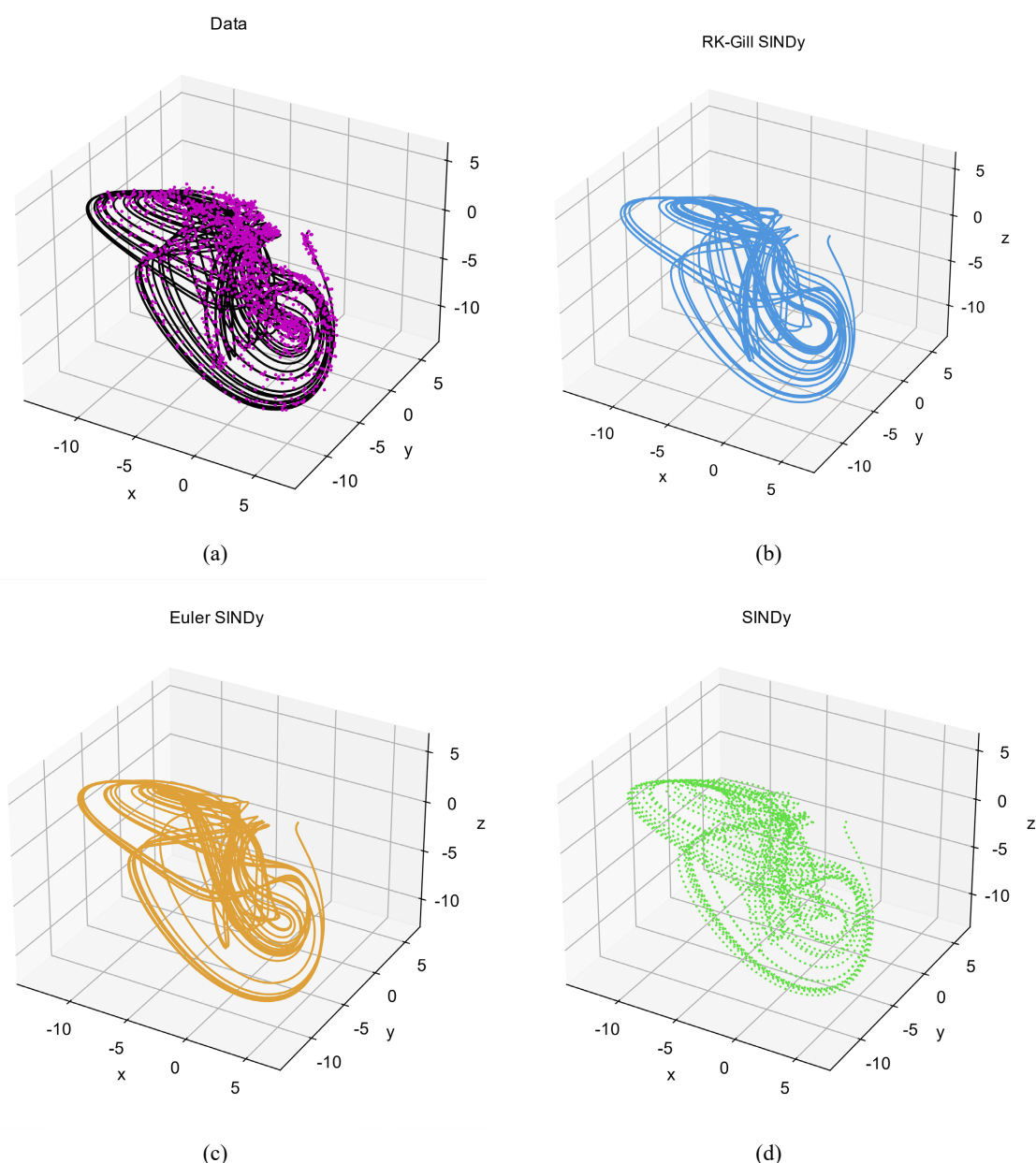


Figure 6. Halvorsen model: (a) Comparison between the observed values contaminated by Gaussian white noise with a standard deviation of 0.2 and the real continuous trajectory; (b)-(d) show the trajectories reconstructed by RK-Gill SINDy, Euler SINDy and standard SINDy for this model, respectively

图 6. Halvorsen 模型: (a) 为加入大小为 0.2 的高斯白噪声之后所得观测值与真实连续轨迹比较; (b)-(d) 分别为用 RK-Gill SINDy、Euler SINDy 和 SINDy 重构此模型所得轨迹

Table 6. Halvorsen model: Errors of governing equations reconstructed by RK-Gill SINDy, Euler SINDy and standard SINDy at the noise level of 0.2

表 6. Halvorsen 模型：噪声水平为 0.2 时 RK-Gill SINDy, Euler SINDy 和标准 SINDy 还原控制方程的误差

σ	RK-Gill SINDy	Euler SINDy	SINDy
0.2	9.86×10^{-3}	1.28×10^{-2}	5.93×10^{-2}

5. 结论

针对稀疏采样与噪声干扰双重挑战下的非线性动力系统辨识难题，本文提出了一种基于 RK-Gill 格式的 SINDy 方法。该方法在上述复杂场景下展现出全面的性能优势，表现优于对比方法。与 Euler SINDy 及标准 SINDy 方法相比，RK-Gill SINDy 更能在严重的数据稀疏性和噪声扰动的约束下实现对目标系统控制方程的高精度结构辨识与高鲁棒性参数重构。值得注意的是，在本研究涉及的所有测试动力学模型中，该方法在参数估计精度这一量化指标上始终保持最优性能。

进一步的定量分析结果表明，相对系数误差与数值积分时间步长正相关，这一规律不仅揭示了影响模型辨识精度的关键因素，也为 RK-Gill SINDy 方法在实际应用中的参数设置策略与适用场景界定提供了重要的理论依据。

6. 局限性与展望

尽管本文提出的 RK-Gill SINDy 方法在低维常微分方程系统的辨识任务中展现出优异的性能与鲁棒性，但当前研究仍存在一定的局限性，有待在未来工作中进一步探索与完善。

首先，该方法的有效性仅在低维非线性常微分方程(ODE)系统上得到了验证，其在高维复杂系统、偏微分方程(PDEs)以及时空动力学系统中的适用性仍需进一步研究。更高维场景下，状态空间维度的增加会导致候选函数库规模呈指数级增长，这不仅会显著提升计算成本，还可能降低稀疏回归算法的稳定性，从而影响辨识精度与效率。其次，本研究主要基于等距采样数据开展实验，而实际工程场景中常存在非等距采样、数据缺失或传感器异步等问题，如何将 RK-Gill 格式与非均匀采样框架相结合，是未来需要解决的关键问题之一。此外，相较于基于标准 SINDy 方法，RK-Gill 格式和 Euler 格式因引入了中间阶段的计算步骤，单次迭代的计算成本有所增加，如何在保证辨识精度的前提下优化算法实现、降低计算开销，也是后续研究的重要方向。

针对上述局限性，未来的工作将围绕以下几个方面展开：一是将 RK-Gill SINDy 方法拓展至高维系统与偏微分方程辨识，并结合模型降阶或稀疏采样技术以控制计算复杂度；二是研究非等距、含缺失值数据下的鲁棒辨识框架，提升方法在真实场景中的泛化能力；三是对算法进行工程化优化，探索并行计算或低复杂度实现方案，平衡辨识精度与计算效率。这些后续探索将进一步完善该方法的理论体系，并推动其在更广泛的实际问题中得到应用。

参考文献

- [1] Lu, L., Meng, X., Mao, Z., *et al.* (2021) DeepXDE: A Deep Learning Library for Solving Differential Equations. *SIAM Review*, **63**, 208-228.
- [2] Higham, C.F. and Higham, D.J. (2019) Deep Learning: An Introduction for Applied Mathematicians. *SIAM Review*, **61**, 860-891. <https://doi.org/10.1137/18m1165748>
- [3] 肖冉, 安新磊, 祁慧敏, 等. 电场作用下 HR 神经元的分岔分析及参数辨识[J]. 浙江大学学报(理学版), 2022, 49(6): 691-697+705.
- [4] Raissi, M., Perdikaris, P. and Karniadakis, G.E. (2019) Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations. *Journal of Computational*

-
- Physics*, **378**, 686-707. <https://doi.org/10.1016/j.jcp.2018.10.045>
- [5] Brunton, S.L., Proctor, J.L. and Kutz, J.N. (2016) Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems. *Proceedings of the National Academy of Sciences*, **113**, 3932-3937. <https://doi.org/10.1073/pnas.1517384113>
 - [6] Wang, Y., Yuan, Y., Fang, H. and Ding, H. (2024) Data-Driven Discovery of Linear Dynamical Systems from Noisy Data. *Science China Technological Sciences*, **67**, 121-129. <https://doi.org/10.1007/s11431-023-2520-6>
 - [7] Basiri, M.A. and Khanmohammadi, S. (2025) SINDyG: Sparse Identification of Nonlinear Dynamical Systems from Graph-Structured Data, with Applications to Stuart-Landau Oscillator Networks. *Journal of Complex Networks*, **13**, cnaf029. <https://doi.org/10.1093/comnet/cnaf029>
 - [8] Rico-Martinez, R., Anderson, J.S. and Kevrekidis, I.G. (1994) Continuous-Time Nonlinear Signal Processing: A Neural Network Based Approach for Gray Box Identification. *Proceedings of IEEE Workshop on Neural Networks for Signal Processing*, Ermioni, 6-8 September 1994, 596-605. <https://doi.org/10.1109/nnsdp.1994.366006>
 - [9] Messenger, D.A. and Bortz, D.M. (2021) Weak SINDy: Galerkin-Based Data-Driven Model Selection. *Multiscale Modeling & Simulation*, **19**, 1474-1497. <https://doi.org/10.1137/20m1343166>
 - [10] Goyal, P. and Benner, P. (2022) Discovery of Nonlinear Dynamical Systems Using a Runge-Kutta Inspired Dictionary-Based Sparse Regression Approach. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **478**, Article 20210883. <https://doi.org/10.1098/rspa.2021.0883>
 - [11] Ben Amma, B., Melliani, S. and Chadli, L.S. (2018) A Fourth Order Runge-Kutta Gill Method for the Numerical Solution of Intuitionistic Fuzzy Differential Equations. In: *Studies in Fuzziness and Soft Computing*, Springer, 55-68. https://doi.org/10.1007/978-3-030-02155-9_5
 - [12] Euler, L. (1770) *Institutionum Calculi Integralis Volumen Tertium* [EB/OL]. <https://scholarlycommons.pacific.edu/euler-works/385/>
 - [13] Sprott, J.C. (1994) Some Simple Chaotic Flows. *Physical Review E*, **50**, R647-R650. <https://doi.org/10.1103/physreve.50.r647>
 - [14] Vaidyanathan, S. and Azar, A.T. (2016) Adaptive Control and Synchronization of Halvorsen Circulant Chaotic Systems. In: *Studies in Fuzziness and Soft Computing*, Springer, 225-247. https://doi.org/10.1007/978-3-319-30340-6_10