arXiv:2302.10787v1 [cs.LG] 4 Feb 2023

以低维混沌为基准测试稀疏系统辨识

Alan A. Kaptanoglu，[1,2,*] Lanyue Zhang，[3] Zachary G. Nicolaou，[3] Urban Fasel，[4] 和 Steven L. Brunton[2]*IREAP*，马里兰大学，美国马里兰州 *College Park*[2]华盛顿大学机械工程系，美国华盛顿州西雅图[3]华盛顿大学应用数学系，美国华盛顿州西雅图[4]帝国理工学院航空系，英国伦敦

稀疏系统辨识是一种数据驱动过程，旨在获得能够描述动力系统演化的简约微分方程，并在模型复杂度与准确性之间取得平衡。尽管系统辨识在各科学领域迅速发展，但文献中仍缺少在多种动力系统上评估的大规模方法比较研究。在本研究中，我们利用 Gilpin [1]提出的 dysts 混沌系统标准化数据库，对稀疏回归的不同变体进行系统基准测试。具体而言，我们展示了如何利用这一开源工具定量比较不同的系统辨识方法。为说明如何使用该基准，我们对求解非线性动力学稀疏辨识（SINDy）优化问题的四种算法进行了大规模比较，发现原始算法和近期提出的混合整数离散算法表现出色。在所有情况下，我们都采用集成来提升 SINDy 的噪声鲁棒性，并进行统计比较。此外，我们还提供了有力证据，表明即使在干净数据上，弱 SINDy 公式也显著优于传统方法。最后，我们探究了 SINDy 算法生成的帕累托最优模型如何依赖方程的性质，发现其性能与一组用于量化混沌程度、尺度分离、非线性程度和句法复杂度的动力学性质并无显著关联。

Key关键词：系统辨识、动力系统、混沌、非线性系统、稀疏回归、SIND y

I. 引言

传统上，动力系统的控制方程通过第一性原理或现象学方法推导，并经实验验证。然而，现代社会中海量数据的可用性正推动数据驱动的模型发现，这种方法如今已广泛应用于数学、物理学、化学、生物学、工程学和经济学等诸多领域。近年来，人们开发了大量此类方法，例如动态模态分解 [2–4], Koopman 理论 [5, 6], 非线性自回归算法 [7], 神经网络 [8–10], 高斯过程回归 [11], 算子推断和降阶建模 [12–14], 符号回归 [15, 16], 分治策略 [17], 以及稀疏回归 [18]。即便在每一种方法内部，方法学创新也层出不穷。

选择数据驱动建模策略时，必须考虑多个因素，包括数据的数量和质量，以及模型的最终用途。所有这些方法都面临一个关键权衡：如何平衡准确性与模型复杂度。因为如果不加考虑地确定大量自由且非零的参数，模型很可能会对训练数据过拟合，尤其是现实世界数据中会包含各种类型的噪声。尽管数据驱动的模型发现方法层出不穷，但在大量非平凡动力系统上对这些方法进行比较的研究相对匮乏。我们旨在弥补这一文献空白，对近期开发的大型精选数据库中的稀疏系统辨识变体进行系统比较，该数据库收录了已知的混沌动力系统，该数据库收录了已知的混沌动力系统 [1]。

稀疏系统辨识是一个总称，指通过促进模型参数化的稀疏性，在模型复杂度与准确性之间实现关键权衡的系统辨识方法。这有助于先验地构建简约描述，用尽可能少的项来拟合数据。非线性动力学稀疏辨识（SINDy） [18]是简约建模的一种领先方法，本文将研究该方法及其变体。它基于稀疏正则化线性回归，因此计算快速且稳健。此外，研究人员可以轻松理解辨识出的模型和优化结果，因为广义线性回归与稀疏优化方面已有大量文献。研究 SINDy 各种变体的另一个重要原因是，自 2016年 引入以来，它已得到迅速应用和扩展。SINDy 已广泛应用于化学反应动力学 [19], 、化学过程 [20, 21], 、空气污染物建模 [22], 、生物运输 [23], 、恒星动力学 [24], 、疾病传播 [25],、对流传热 [26], 、非线性光学 [27], 、电力系统 [28, 29], 、水力学 [30], 、科学计算 [31],、人类行为模型 [32], 、流体动力学 [33–44], 、湍流建模 [45–50], 、等离子体物理学 [51–53], 、结构建模 [54], 等领域的模型识别 [30, 55–61]。此外，它还扩展到了更复杂的建模场景，例如偏微分方程 [62, 63], 、时滞方程 [64], 、随机微分方程 [65–69], 、贝叶斯建模 [70], 、量纲分析 [71], 、带输入或控制的系统 [72–74], 、具有隐式动力学的系统 [75, 76], 、混合系统 [77, 78], 、施加物理约束 [33, 52, 79], 、融入信息论 [80] 、组稀疏性 [81] 或全局稳定性 [82], ， 以及从受污染或有限数据中识别模型 [83–

* 通讯作者（akaptano@uw.edu）。

87], 从状态空间的部分测量中识别模型 [88–91], 采用巧妙的子采样策略识别模型 [92], , 并使用初始条件集成 [93], 结合集成方法进行交叉验证 [94, 95], , 还扩展到相关目标函数 [96], 、弱形式或积分形式 [97–107], 、张量表示 [108, 109],以及随机强迫 [110]。

该领域的大多数研究仅在少数几个典型动力系统上展示性能，例如 Lorenz63系统 [111]。此外，针对不同稀疏系统辨识技术的比较性能，迄今鲜有系统性研究。我们认为，文献中这类研究相对匮乏可能有两个原因：（1）对于较复杂的系统辨识方法，计算数千甚至数十万个动力学模型的计算成本可能很高，而这样做是为了得出本文所展示的统计结论；（2）如果已知真实的控制方程，分析系统辨识的性能会更加直接。显然，真实实验数据并不具备这一条件。借助基准测试，可以对新方法和现有方法进行适当比较，从而了解新算法带来了哪些新颖优势和功能。此外，几乎所有系统辨识方法都需要调整超参数；若要在大量示例上进行恰当比较，就必须只比较每种方法经过最优调参的模型。

乍看之下，影响因素似乎太多，难以对各种方法进行全面比较。原则上，稀疏系统辨识的性能可能取决于方程复杂度、可搜索的函数形式空间、训练数据的规模与多样性、可用于超参数扫描的计算能力、底层动力学的非线性、数据中噪声的数量与类型、用于解决问题的算法质量（及其相关超参数），以及可能还有许多其他因素。

A. 本文的贡献

为控制这些参数，我们从多个方面精心设计了实验。我们使用固定的函数空间来拟合所有系统，其中包含成功进行模型识别所需的所有可能项。此外，通过使用 dysts 数据库中的全部多项式动力系统，对动力学和训练数据的规模进行标准化；同时，采用仅需极少超参数的稀疏回归算法，使计算能力的消耗可以忽略不计。为得出稳健的结论，我们采取了以下步骤：从均值为零的高斯分布 $\mathcal{N}(0, \sigma = d\|\boldsymbol{X}_{\text{train}}\|_2/100)$ 中采样 d% 的噪声，并将其添加到训练数据中的每个数据点 $\boldsymbol{X}_{\text{train}}$ （随后通过计算 $\dot{\boldsymbol{X}}_{\text{train}}$ from $\boldsymbol{X}_{\text{train}}$ 放大噪声）。此外，通过对相同数据进行子采样来生成大量 SINDy 模型 [94], , 执行超参数扫描，最后仅报告统计模型性能指标。本研究首先详细介绍这一标准化过程和方法论，然后展开说明。

本文中的所有方法和结果均可通过 PySINDy 代码 [112, 113] 复现。此外，我们在附录 A 中讨论了如何在未来的研究中有效使用这一工具，以便研究人员设计新的算法、软件库、数据库或其他功能。本文通过开展若干关于 SINDy 优化与性能的大规模研究，展示了这一工具的适用范围和灵活性。

首先，我们在一个大型数据库上，对用于求解 SINDy 模型辨识问题的不同优化算法进行了统计比较。我们比较了 STLSQ [18],SR3 [114], Lasso [115], MIOSR [116], 以及弱形式 [117]算法。据我们所知，这是对这些算法首次开展大规模统计比较。我们发现，原始 STLSQ 算法 [18] 在执行 SINDy 优化时表现稳健，而近期提出的 MIOSR 算法性能出色，尤其适用于低维或受约束问题。此外，我们发现了有力证据表明，SINDy 问题的弱形式能够显著提升性能，即使是在使用干净数据的情况下也是如此。接下来，我们根据真实底层动力学的性质，对帕累托最优模型性能进行分类。我们发现，SINDy 方法与混沌程度、尺度分离、非线性程度，甚至方程的句法复杂度均无显著关联，尽管存在一些有趣的偏差，值得今后进一步研究。

II. 方法

要对系统辨识方法的性能得出可靠结论，就必须确保只进行标准化比较。接下来，我们将介绍数据集、衡量稀疏系统辨识性能的指标、如何通过超参数扫描为每个系统找到帕累托最优模型，并明确界定要研究的动力学性质。本节旨在阐明我们如何在系统辨识过程中谨慎控制混杂变量，从而针对不同优化算法以及数据库的动力学性质得出结论。

A. Dysts 数据库

Gilpin 于 2021年 [1], 推出的 dysts 数据库非常适合用于评估系统辨识性能，因为它收录了大量高度标准化且运动方程已知的混沌系统。许多混沌系统的方程在语法上都较为简单，例如，Lorenz63 是一个三状态系统，其微分方程仅包含二次非线性和七个项。然而，与先前用于对符号回归算法 [119]进行基准测试的 Nguyen 符号回归数据库 [118] 等许多现有数据库相比，该数据库生成的数据在动力学方面要复杂得多。该

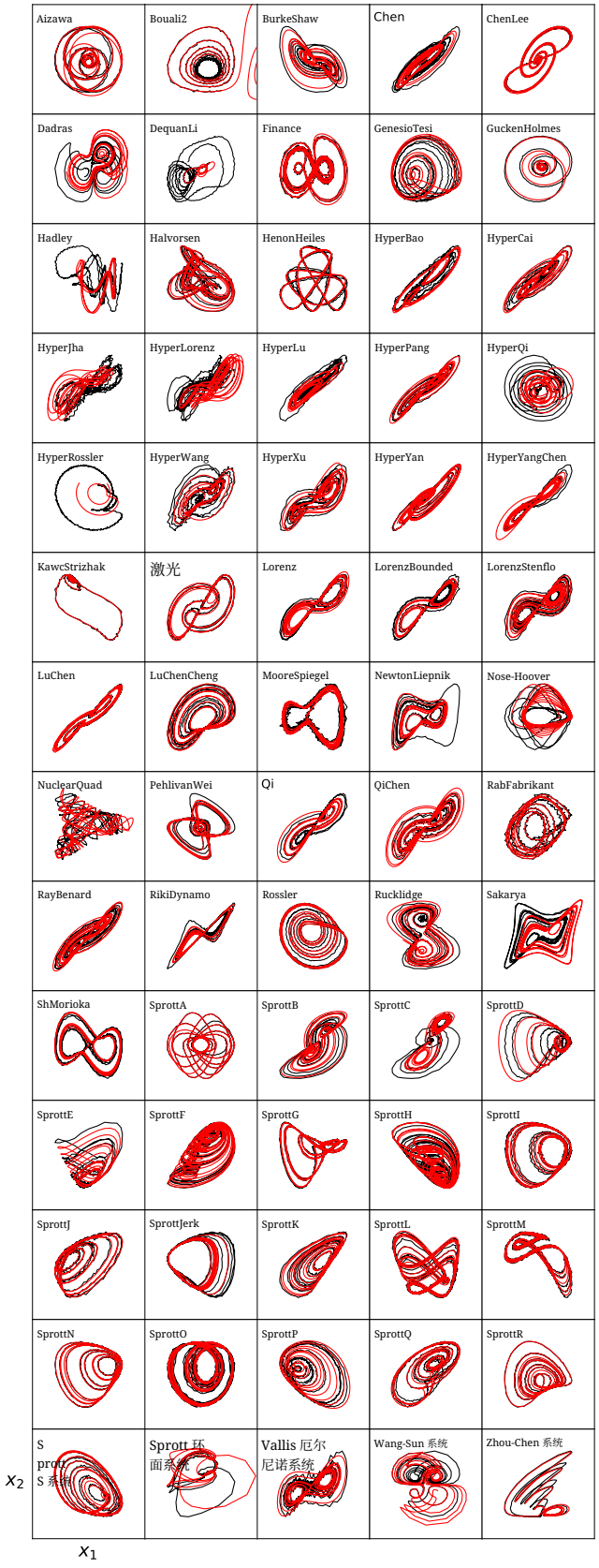


图 1：本研究中各系统的（x_1,x_2）子空间中的训练轨迹，添加了 1% 噪声（黑色）；同时展示了干净的测试轨迹（红色），数据取自 [1]。一些动力系统具有多个吸引子，数据可能不足以完全“覆盖”吸引子。

数据库还提供了来自不同科学领域的 100 多种混沌系统的数据、方程和动力学属性，这些系统都具有奇异吸引子。由于这些系统的动力学属性已预先计算，且系统之间高度标准化，因此该数据库可用于测试稀疏系统辨识。例如，所有系统都采用相位替代显著性检验来选取最优积分时间步长和采样率 [120]，从而使不同系统的动力学在最小和最大时间尺度上保持一致。这是为每个模型生成相似时间序列的关键步骤；该数据集中最小显著时间尺度的跨度超过四个数量级。最重要的是，可通过已知的真实控制方程来评估模型性能。

为缩小比较范围，我们考察数据库中的 70 个系统。这些系统由具有多项式非线性的常微分方程（ODEs）构成，非线性次数不超过四次。我们仅考察这些系统，并为每次模型拟合定义一个统一的“特征库”（可用于拟合数据的函数空间）。对于一般的非线性系统，在吸引集或不动点附近，动力学往往可以合理地展开为泰勒级数，因此多项式模型通常是捕捉其主导阶动力学的自然选择。

与某些其他动力系统不同，混沌系统呈现连续谱的频率，因此其动力学数据尤其难以与噪声进行系统区分，例如使用频率滤波器。我们所选的系统均属于混沌的一个特定子类：它们都有界，并且在数据库中采用默认系统参数时，以奇异吸引子为特征。此外，54 个系统的状态空间维数为三，其余 16 个为四；数据中还有一个哈密顿系统。估计有 51 个系统属于“超混沌”系统，具有两个或更多正李雅普诺夫指数。图 1 展示了各自包含 1000 个数据点的训练轨迹和测试轨迹，以便观察奇异吸引子。

为得出统计结论，我们生成五条训练轨迹和五条测试轨迹，分别代表十个运动周期；每个周期采样 100 个点，因此每条轨迹总计包含 1000 个数据点。每条轨迹均由吸引子上的不同初始条件生成。

总而言之，底层动力学是混沌的，我们假设动力学方程可以表示为最高四阶、具有多项式非线性项的常微分方程组。此外，我们还可以获得全状态测量数据；在接下来的分析中，这些数据会在不同阶段加入零均值高斯噪声。

B. SINDy 方法

SINDy 利用稀疏回归从数据中发现控制方程 [18]。其目标是识别如下形式的微分方程模型

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) \approx \Theta(\boldsymbol{x})\Xi, \tag{1}$$

其中， $\boldsymbol{f}(\boldsymbol{x})$ 是状态 $\boldsymbol{x}$ 的非线性函数， $\boldsymbol{\Theta}$ 是一个非线性函数库， $\boldsymbol{\Xi}$ 是一个常系数矩阵。 $\boldsymbol{\Theta\Xi}$ 共同表示非线性函数的线性组合，这些函数经过选择，能够尽可能准确地逼近 $\boldsymbol{f}(\boldsymbol{x})$。 $\boldsymbol{\Theta}$ 中非线性函数的选择显然对成功逼近至关重要。在从连续状态空间转向离散测量数据之前，请注意，高阶微分方程总是可以化为一阶耦合微分方程，因此这种形式具有相当的普适性。PySINDy [121]支持针对控制输入 [72], 、偏微分方程 [63], 、隐式项 [76], 及其他变化形式对式（1）进行扩展。

为了从数据中辨识方程 (1)，我们需要对状态 $\boldsymbol{x}$ 进行测量。首先，我们在一组训练轨迹上对完整状态数据进行均匀采样，并由状态的时间序列数据构成数据矩阵 $\boldsymbol{X} \in \mathbb{R}^{M\times d}$ ，即$\boldsymbol{x}(t_1)\ \boldsymbol{x}(t_2)...\ \boldsymbol{x}(t_M)$:

$$\boldsymbol{X} = \begin{array}{c} \xrightarrow{\qquad\qquad\text{state}\qquad\qquad} \\ \begin{bmatrix} x_1(t_1) & x_2(t_1) & \cdots & x_d(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_d(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_M) & x_2(t_M) & \cdots & x_d(t_M) \end{bmatrix} \end{array} \Bigg\downarrow . \tag{2}$$

M 表示时间点的数量，d 表示状态空间的维数，即对于我们的数据集而言为三维或四维。类似地，定义时间导数矩阵 $\dot{\boldsymbol{X}}$，并可根据 $\boldsymbol{X}$ 进行数值计算，例如使用有限差分或更复杂的微分 [86, 122, 123]。随后，我们构建一个特征库 $\boldsymbol{\Theta}(\boldsymbol{X})$，其中包含可能描述系统演化的项，并尝试找到一个稀疏的系数向量 $\boldsymbol{\xi}$ ，表示该库中各项的最佳线性组合。一般而言，特征库 $\boldsymbol{\Theta}$ 可以包含广泛的广义函数，而设计该库通常需要领域知识；然而，本文研究的系统全部是多项式系统。因此，我们将 $\boldsymbol{\Theta}$ 固定为关于 $\boldsymbol{X}$ 的多项式库，最高为四阶。这一点非常重要；一般来说，系统辨识方法依赖于特征库中恰当的函数形式，而在这里我们完全绕过了这一问题，将优化性能与库的选择相互隔离并加以区分。我们明确知道，该特征库包含描述本文所研究的全部 70 个混沌系统所必需的项。

系数向量 $\boldsymbol{\xi}$ 通过以下稀疏优化问题确定:

$$\arg\min_{\boldsymbol{\xi}} \left[\frac{1}{2} \|\boldsymbol{\Theta\xi} - \dot{\boldsymbol{X}}\|^2 + \lambda \|\boldsymbol{\xi}\|_0 \right]. \tag{3}$$

式（3）中的 SINDy 优化问题的第一项，是对给定数据 $\dot{\boldsymbol{X}}$ 进行系统常微分方程 $\boldsymbol{\Theta\xi}$ 的最小二乘拟合。l_0 损失 $\|\boldsymbol{\xi}\|_0$ 是一个用于计算 $\boldsymbol{\xi}$中非零元素个数的函数。使用l_0 损失或其他正则化方法来促进已识别模型系数的稀疏性，往往能提升模型对数值噪声或实验噪声的鲁棒性，从而应对准确性与泛化能力之间的权衡，并避免对训练数据过拟合。l_0 损失是非凸且非光滑的，因此人们提出了许多针对该问题的凸松弛方法 [18, 63, 79]。本文介绍的首项重要成果，是比较了多种求解式（3）的算法，包括原始 SINDy 论文中的序贯阈值最小二乘（ STLSQ）算法 [18], 、使用 Lasso 求解式（3）的传统凸松弛方法 [115]、SR3 算法 [79, 114], ，以及一种近期提出的名为 MIOSR 的混合整数离散算法 [116]。

在可行范围内，我们尽量使用 PySINDy 的默认参数，以减少需要扫描的超参数数量。例如，在本研究中， $\dot{\boldsymbol{X}}$ 通过简单的有限差分计算得出。未来的工作可以立即采用更复杂的差分方法 [86, 122, 123], 来提升性能，其中许多方法已在 PySINDy 代码中实现。众所周知，SINDy 的弱形式或积分形式 [97] 也能减轻有限差分固有的噪声放大问题；我们将在第 III B 节说明，PySINDy 中实现的弱形式可显著提升性能。

C. 性能度量指标

方法比较需要适当的度量指标来确定不同方法的相对性能。与现实世界数据不同，我们可以获得所有 70 个待测试系统的真实运动方程。因此，我们能够计算多种有用的度量指标，以表征所得稀疏符号模型在这些示例中的性能。我们定义归一化系数误差和均方根误差

$$\mathrm{E}_{\mathrm{coef}} = \frac{\|\boldsymbol{\xi}_{\mathrm{True}} - \boldsymbol{\xi}_{\mathrm{SINDy}}\|_2}{\|\boldsymbol{\xi}_{\mathrm{True}}\|_2}, \tag{4}$$

$$\mathrm{E}_{\mathrm{RMSE}} = \frac{\|\dot{\boldsymbol{X}}_{\mathrm{True}} - \dot{\boldsymbol{X}}_{\mathrm{SINDy}}\|_{\mathrm{F}}}{\|\dot{\boldsymbol{X}}_{\mathrm{True}}\|_{\mathrm{F}}}. \tag{5}$$

其他研究也探讨了式（4）的 $\|\cdot\|_1$ 和 $\|\cdot\|_0$ 范数形式，以量化模型失配 [76, 119]。对于所选数据库中的干净数据，使用基础的系统辨识算法即可轻松获得非常小的 $\mathrm{E}_{\mathrm{RMSE}}$ 值（或平均绝对误差、常规平均绝对百分比误差等变体）。这是因为，参数众多的模型可以利用许多虽小但非零的项实现过拟合；如果我们只考察奇异吸引子附近的区域，这类模型还可能生成非常准确的测试轨迹。尽管存在这一注意事项，在预测、奇异吸引子重构、李雅普诺夫谱计算，以及其他用于捕捉动力系统行为的重要度量指标方面，$\mathrm{E}_{\mathrm{RMSE}}$ 较低的模型通常已经足够。此外，人们通常关注的情形是动力学方程以及由此产生的系数误差均未知。我们将在附录B中进一步讨论误差和稳定性。

既然我们知道真实的系数，就可以通过考察系数误差来检查是否过拟合；这些误差可能

即使 E_{RMSE} 极小*，也可能不正确。事实上，在奇异吸引子上，与完整状态空间中的动力学相比，动力学可能会显著简化。附录 C 讨论了这与不变流形以及动力系统有效维数和熵的降低之间的联系。

D. 使用集成进行自动超参数扫描

在对系统辨识性能得出任何结论之前，必须调优每种算法的超参数。扫描与稀疏性相关的超参数会生成一个谱系，其范围从准确率有限的高度稀疏模型，到预测准确率较高的稠密模型；这通常被称为帕累托前沿 [124]。数据集规模过大，无法针对每条混沌轨迹手动调优超参数，以获得最佳的系统辨识性能，因此，在进行超参数扫描时，我们采用以下流程。对于所有算法，我们都使用集成 [94]。我们对五条训练轨迹进行子采样，即在不重复抽样的情况下选取总轨迹数据的 50%，从而针对每个阈值超参数值生成 10 个模型。对于 STLSQ，我们随后扫描 300 个逐渐增大的硬阈值，并根据对通过子采样生成的 10 个模型取平均后的、经过有限样本量校正的赤池信息准则（AIC） [80] 的最小值，确定最佳阈值，

$$\mathrm{AIC} = M\log(\|\dot{\boldsymbol{X}}_{\mathrm{True}} - \dot{\boldsymbol{X}}_{\mathrm{SINDy}}\|_2^2) + 2k + \frac{2k(k+1)}{(M-k-1)}, \quad (6)$$

其中 $k = \|\boldsymbol{\xi}_{SINDy}\|_0$ 和 $\dot{\boldsymbol{X}}_{\mathrm{True}}$ 以及 $\dot{\boldsymbol{X}}_{\mathrm{SINDy}}$ 均根据测试轨迹计算得出。AIC（赤池信息准则）是一种广为人知的统计指标，可在平衡稀疏性与准确性的同时比较模型 [80]。这一系统化流程需要辨识 $300 \times 10 \times 70 = 210{,}000$ 个模型。尽管需要生成大量模型，STLSQ 的整个过程在单台 Intel Core i9-7920X 2.90GHz 中央处理器上运行约需 10 分钟。通过帕累托扫描确定最佳阈值后，相应的十个模型使我们能够描述辨识出的系数和误差的分布。其他 SINDy 算法也针对各自的超参数进行类似调优。SR3 算法还包含一个额外参数 ν，我们将其设为 1 或 0.1，因为这些取值通常适用于干净数据，而且无需额外扫描超参数。

使用 SINDy，而非更复杂的系统辨识方法，有助于极快地计算大量模型。例如，先前的研究 [125] 主要考察用于符号回归的遗传编程算法，每种算法针对每个动力系统大约需要 200 个核时，才能从 Pareto 前沿生成最佳模型。相比之下，使用 STLSQ，我们针对每个动力系统大约只需 $\sim 10^{-3}$ 个核时即可生成 Pareto 最优模型，这与其他研究的发现一致 [1, 126]。之所以需要这样的速度，是因为我们所概述的系统化流程要求每种算法生成 $\mathcal{O}(10^5)$ 个 SINDy 模型。

E. 动力学性质

微分方程系统通常可以根据各种动力学性质进行有效分类。现在，本文所使用的动力学系统已经介绍完毕，我们定义了一组用于测试稀疏系统辨识性能的动力学性质：混沌程度、尺度分离程度、基础方程的句法复杂度以及非线性程度。在本文中，混沌程度通过最大李雅普诺夫指数进行测量。尺度分离程度定义为系统的主导时间尺度与其最小显著时间尺度之比，其中最小显著时间尺度由 Gilpin 的方法确定 [1]。句法复杂度通过描述长度进行量化，表示描述方程中每个对象所需的比特数。最后，非线性程度定义为方程中项总数的加权和，其中每一项按其多项式次数加权，因此它综合反映了非线性程度和方程项数。关于更精确的定义以及指标选择的讨论，请参见附录 C。图 2 展示了动力学方程中出现的非线性项概览，表明大多数系统都表现出很强的二次非线性。二次非线性常见于流体的降阶模型以及许多其他动力学系统（例如 Lorenz 或 Rossler 系统）中，这是因为物质导数中的对流项会产生二次非线性。十个系统表现出最高达三次的非线性，只有两个系统具有四次非线性。图 3 总结了其余动力学性质的分布，表明数据集所取的数值范围很广。

III. 结果

现在，我们通过展示一系列基准实验，说明将 dysts 与 PySINDy 结合使用的价值，以便更好地理解不同稀疏系统辨识变体在性能上的权衡。在解释了数据库的特征、稀疏系统辨识算法、误差指标以及我们关注的动力学性质之后，我们将通过标准化的帕累托最优 SINDy 模型展示研究结果。我们强调，所有这些结果都以示例形式在 PySINDy 代码中复现，而以下研究

*另一个常用指标是恢复率 [119],，对于方程中每个被正确识别的项，该指标返回离散值 0 或 1。然而，超参数调优不足、数据中的噪声、优化收敛不充分以及许多其他因素都可能导致小误差。此外，通常还需要对较大的系数误差施加比小误差更重的惩罚。因此，本文省略恢复率这一指标，但将其作为未来研究的一个方向。

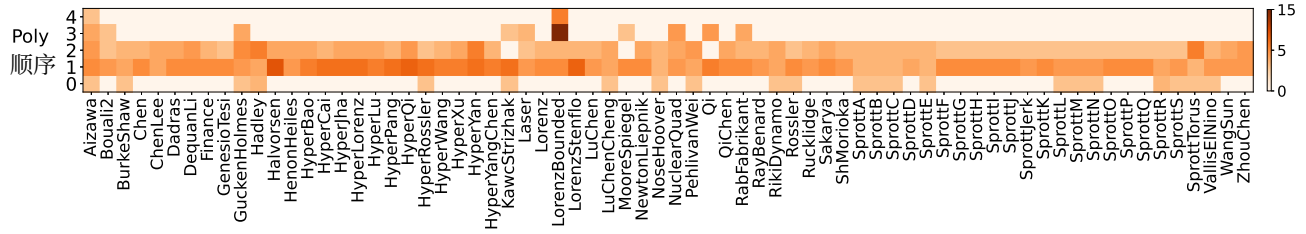


图 2：每个方程中各多项式次数对应的多项式项总数。可以看出，二次非线性项占主导地位。

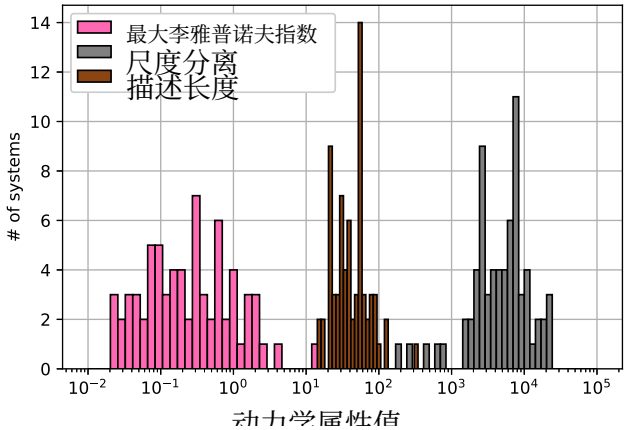


图 3：本研究所用动力系统各项属性的分布。

使用通过超参数扫描选出、赤池信息准则值最小的帕累托最优辨识模型。

A. SINDy 算法比较

为了展示这种方法和数据库，我们比较了一类 SINDy 算法的帕累托最优模型：STLSQ、Lasso、SR3（将 ν 超参数设为 1 或 0.1）以及 MIOSR。此外，我们还使用 PySINDy 代码中采用弱形式实现的 STLSQ 优化器，以便与传统方法进行比较。虽然本文未展示相关结果，但只需进行少量修改，弱形式也可用于其他任何 PySINDy 优化器。我们使用其他算法进行了弱形式实验，观察到算法之间的趋势相似，因此省略了这些结果。弱形式使用 200 个子域，对应 200 个用于优化的点，而传统优化使用 1000 个轨迹点。子域数量可能过多；经验发现，超过 100 个子域后，性能基本不再提升（事实上，仅用 10 个子域，计算速度就快得多，而且效果仍然不错）。MIOSR 优化器的设置是：每个状态变量最多用五秒来求解并证明最优性。使用干净数据时，很少会触发五秒时限；而当数据中加入 1% 的噪声后，则更常触发。

结果通过前文第 II D 节所述的流程生成。图 4a 展示了各 SINDy 算法的帕累托最优模型性能。结果分别针对干净数据和添加 1% 噪声的数据给出。这些结果表明，已知 Lasso 难以生成高性能的系统辨识结果，不过随着噪声增加，Lasso 的性能退化程度小于某些其他优化器。对于干净数据，其余优化器通常能生成满足 $E_{\text{coef}} < 1\%$ 和 $E_{\text{RMSE}} < 10\%$ 的模型。

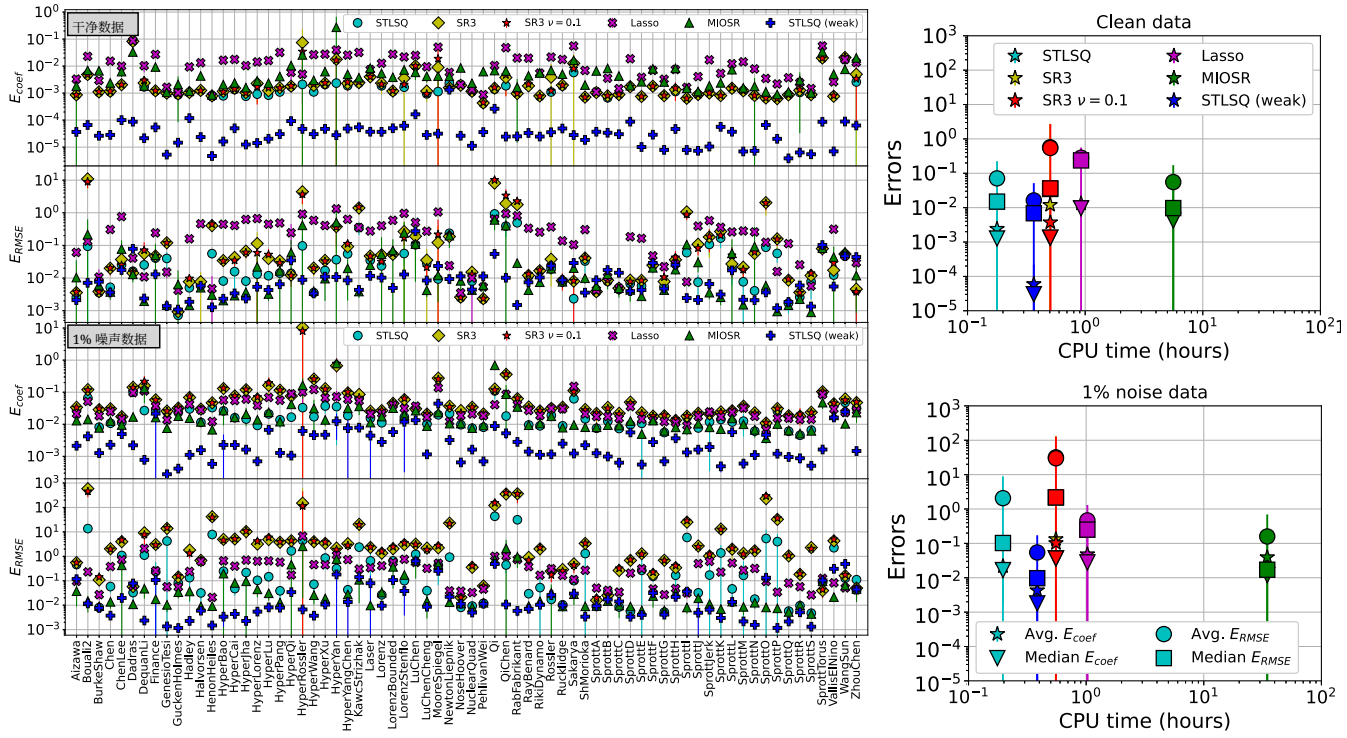
为进一步证明这些动力学模型的准确性，图 5 展示了帕累托最优 STLSQ 模型对十条轨迹的模拟结果。这些轨迹从新的初始条件出发，且初始位置略微偏离各个奇异吸引子。背景灰度逐渐加深，表示在常微分方程求解器中变得不稳定的模拟轨迹数量逐渐增加。总的来说，主要结论是：在无噪声的情况下，帕累托最优 SINDy 模型能够稳定地产生新的轨迹，重现每个混沌系统的奇异吸引子，这意味着可以准确计算李雅普诺夫谱等动力学特征。

B. 弱形式结果

SINDy [97, 98, 100] 的弱形式或积分形式将轨迹的时间跨度划分为若干子域，并考虑常微分方程在这些子域上的积分。该方法通过在适当情况下分部积分，避免了对含噪数据求导及由此导致的噪声放大。

图 4a 的结果表明，弱形式在求得正确的方程系数方面表现极为出色。弱形式 E_{RMSE} 结果看起来稍逊一筹，原因颇为技术性。为了与其他优化器进行比较，我们计算了式 (5) 中定义的逐点 E_{RMSE}。与其他优化器不同，在问题的弱形式表述中，并未直接最小化 E_{RMSE}（训练数据）；弱形式通过最小化在子域上积分的误差度量来求解其优化问题。尽管没有直接最小化这些量，弱形式通常仍能得到最小的 E_{RMSE} 值。

这些结果有力地证明了弱形式的显著优势。有趣的是，稀疏回归的弱形式表现出色，通常被认为是因为它能让用户在计算数据导数时避免放大噪声——



(a) (b)

图 4：（a）在全部 70 个混沌系统上，分别在无噪声（前两行）和含 1% 噪声（后两行）情况下，各优化算法平均误差的汇总。STLSQ 和 MIOSR 通常能够在传统优化器中生成最佳模型。Lasso 的性能较弱，而当存在噪声时，SR3 似乎需要额外的超参数扫描才能避免较大的误差。采用 STLSQ 的弱形式能够极好地识别系数，并展现出使用弱形式的显著优势。（b）汇总所有动力系统后的算法平均性能和中位数性能。标记颜色表示不同的优化器，标记形状表示误差类型。所有运行均在配备 Intel Core i9-7920X 2.90GHz 中央处理器的系统上以串行方式完成。

然而，我们发现，即使没有添加噪声，仅存在有限采样率所固有且不可避免的“噪声”，弱形式仍能带来显著改进。此外，弱形式的结果并非真正的帕累托最优；虽然仍会扫描 STLSQ 的阈值，但弱形式中可用的其他超参数只是固定为一组合理的值。综合来看，我们认为这些结果有力地表明，只要条件允许，就应采用弱形式。为便于灵活使用，弱形式已全面集成到 PySINDy 代码的特征库功能中，因此可与当前或未来的任何优化器，以及任何感兴趣的通用库项搭配使用。

C. 贪心算法与精确算法

MIOSR 是一种精确的混合整数算法，通常能够将l_0 正则化（即非凸）SINDy 问题求解至最优。相比之下，STLSQ 算法采用贪心策略求解该问题，也就是说，它在每次迭代中从系数的一个子集中选出最佳模型，并且无法纠正早期迭代中出现的阈值筛选错误。

对于含噪数据，MIOSR 算法似乎优于其他算法；而对于干净数据，除 Lasso 外，所有表现不弱的算法效果都相近。对于干净数据，SR3 算法表现非常出色。在 $\nu = 1$ 和 $\nu = 0.1$ 两种情况下，SR3 的表现存在一些差异，但看不出明显趋势。加入噪声后，SR3 的误差显著增大，这些结果似乎表明，存在噪声时需要针对 ν 进行超参数扫描，才能获得良好性能。加入噪声后，所有优化器的表现都略有下降。图 4b 进一步总结了这些总体结论。

在 1% 噪声下，STLSQ 和 MIOSR 的表现优于其他优化器。根据平均系数误差衡量，MIOSR 领先于其他方法。STLSQ 在 1% 噪声下出现了少数非常大的均方根误差，这往往会拉低其平均表现。这与这样一种解释相符：随着噪声水平升高，贪心方法越来越容易犯下灾难性的错误。在弱形式结果中也能在一定程度上观察到这一趋势，因为优化仍然是通过 STLSQ 以贪心方式进行的。

一个重要结论是，正如我们对精确的混合整数优化所预期的那样，MIOSR 算法表现良好。

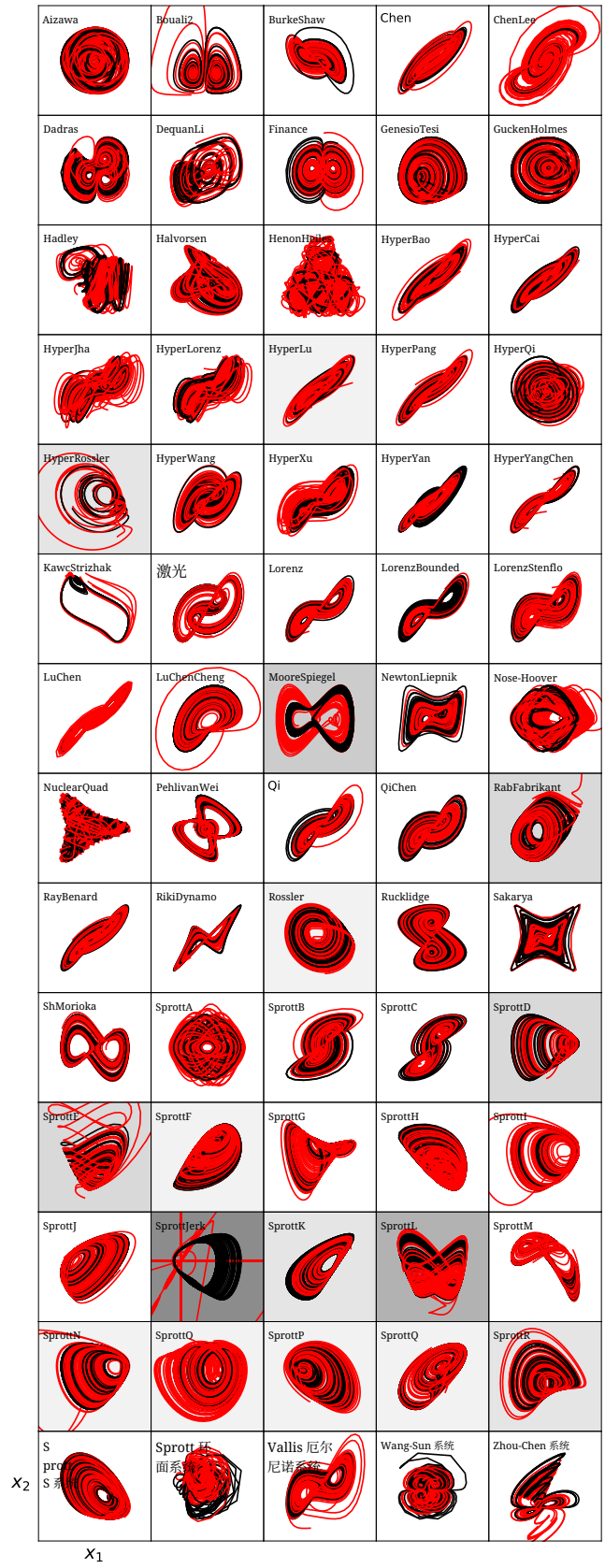


图 5：使用十个新的初始条件，在（x_1,x_2）子空间中，帕累托最优 SINDy STLSQ 模型仿真（红色）与真实轨迹（黑色）的对比。背景颜色越灰，表示轨迹越不稳定；这些轨迹未在图中显示。

然而，图 4b 显示，在噪声水平较高时，MIOSR 算法所需的计算时间远多于其他算法。尽管计算设有时间限制，情况依然如此；不过令人欣慰的是，即便优化器未能找到达到最优性的解，算法仍表现强劲。事实上，正如 Bertsimas 等人 [116]，详细指出的那样，回归问题越难（例如在有噪声的情况下），MIOSR 运行得越慢；但其计算效率随数据量增加而显著提升，而且通过并行化或使用比本研究所用中央处理器更好的中央处理器，还能大幅加速。不过，我们的结果似乎表明，在所选的性能指标上，MIOSR 并未显著优于其他算法。与 Bertsimas 等人 [116]的结果存在这一差异，可能是因为他们主要关注模型系数的真阳性率这一性能指标，而且 MIOSR 在低数据量极限下的结果最为出色。

事实上，这些结果也凸显了 STLSQ 算法的出色表现：STLSQ 对噪声相对稳健，计算速度很快（且不受噪声水平影响），并且在大多数情况下，生成的模型性能与 MIOSR 相当。即使在 1% 的噪声下，STLSQ 贪心特性所累积的误差似乎也可以接受。这些优化器并未在更高噪声水平下进行测试。众所周知，在高噪声情况下，必须采用优化问题的弱形式或积分形式，才能保持较高性能 [94, 97, 99]。图 4a 中弱形式的出色表现清楚地印证了这一点。

D. 动力学性质

对于系统辨识领域的从业者而言，一个亟待解答的问题是：某些类型的动力系统是否比其他类型更难辨识。例如，有些方法专门用于搜索哈密顿系统 [127–129]。另一个例子是，为了避免梯度无界，传统循环神经网络在用于混沌系统建模时应进行修改 [130]。

现在，我们探究动力学性质对各种 SINDy 优化器性能的重要性。本分析所需的数据和计算已在上一节中生成。对于数据集中每个混沌系统，我们按照所考察的动力学性质对帕累托最优模型的误差进行排序，并尝试进行一些简单的数据拟合：线性拟合、对数-线性拟合和双对数拟合。对于所有优化器，$\log(E_{RMSE})$ 与尺度分离之间的线性拟合似乎产生了最大的 R^2 决定系数；同样，对于 $\log(E_{RMSE})$ 与其余动力学性质的对数之间的关系，双对数拟合也有类似表现。我们并不声称这些简单拟合最能反映数据中的相关性；它们的用途是粗略量化动力学性质数值增加与系统辨识性能变差之间是否存在相关性。

对图 6 中 0.1% 噪声下弱形式 STLSQ 优化器的不同动力学性质拟合进行目视检查，似乎确实表明，平均而言，优化器性能

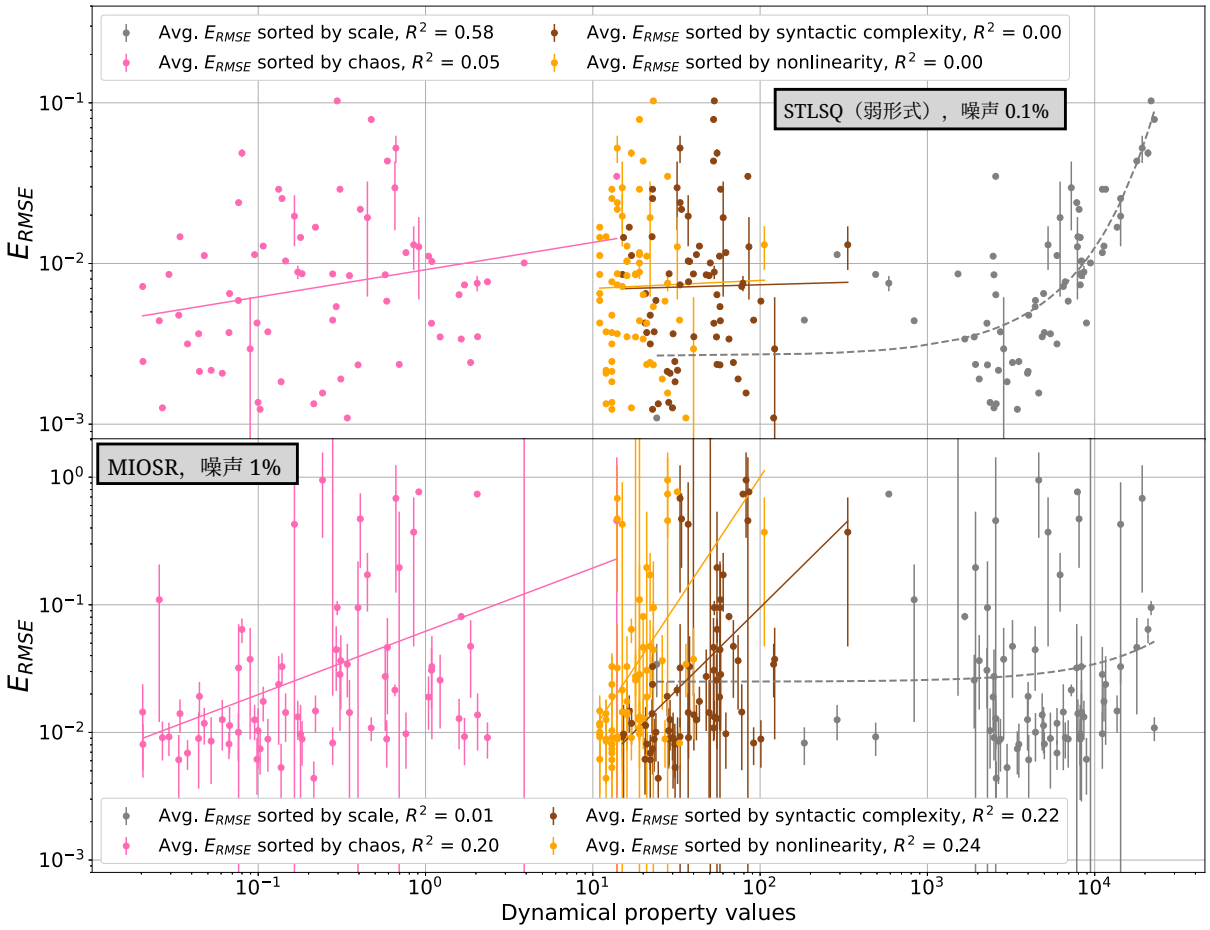


图7：图示了在含 0.1% 噪声的数据上，帕累托最优 STLSQ 优化器（弱形式）的E_{RMSE} 结果，以及在含 1% 噪声的数据上 MIOSR 优化器的E_{RMSE} 结果。STLSQ 结果表明，E_{RMSE} 会随着尺度分离而快速增长。MIOSR 算法与本文测试的另外三种动力学性质呈现出较弱的双对数趋势，尽管随着噪声增加，这些趋势会增强。

随着尺度分离的增大而迅速下降。类似地，对图 6 中 1% 噪声下 MIOSR 优化器的不同动力学性质拟合进行目视检查表明，对于除尺度分离之外的所有动力学性质，数据的双对数拟合都能较弱地反映优化器性能，而尺度分离没有显示出明确的趋势。

图 7 展示了每种优化器、每种动力学性质以及 0、0.1% 和 1% 噪声水平下所有最佳R^2 决定系数。大多数 R^2 值低于 0.25，表明占主导地位的情况是：优化器产生的结果大致独立于动力学性质。然而，仍然可以观察到一些值得关注的趋势；例如，弱形式E_{RMSE} 与尺度分离表现出显著相关性。在较高噪声水平下，MIOSR 似乎与动力学性质的相关性逐渐增强（尺度分离除外）。

这些结果表明，与方程的底层动力学性质存在一些相对持久但相当微弱的相关性。此外，哪些相关性占主导地位似乎取决于数据中的噪声水平以及所选的优化算法。

我们的规模较大的实验有力地印证了 Gilpin 中“任务 4”的结论：SINDy 模型的结果不受混沌程度的影响 [1]。我们衡量系统混沌程度的指标相当粗略，但在 Gilpin 和本研究中，采用其他混沌度量指标时也得到了类似结果。与描述长度缺乏相关性是一个有趣且可能违背直觉的发现。具体而言，如果动力学项位于特征库的子空间内，那么对数据进行稀疏回归所生成的帕累托最优模型，其质量似乎与底层动力学方程的长度或复杂性大致无关。这可以视为一个积极的结果。如果有足够的函数项可用于描述底层动力学，且数据质量很高，那么方程的复杂性便不再产生影响。这为构建大型、内容广泛的特征库提供了依据，尽管这通常会增大特征库的条件数，因此需要额外的

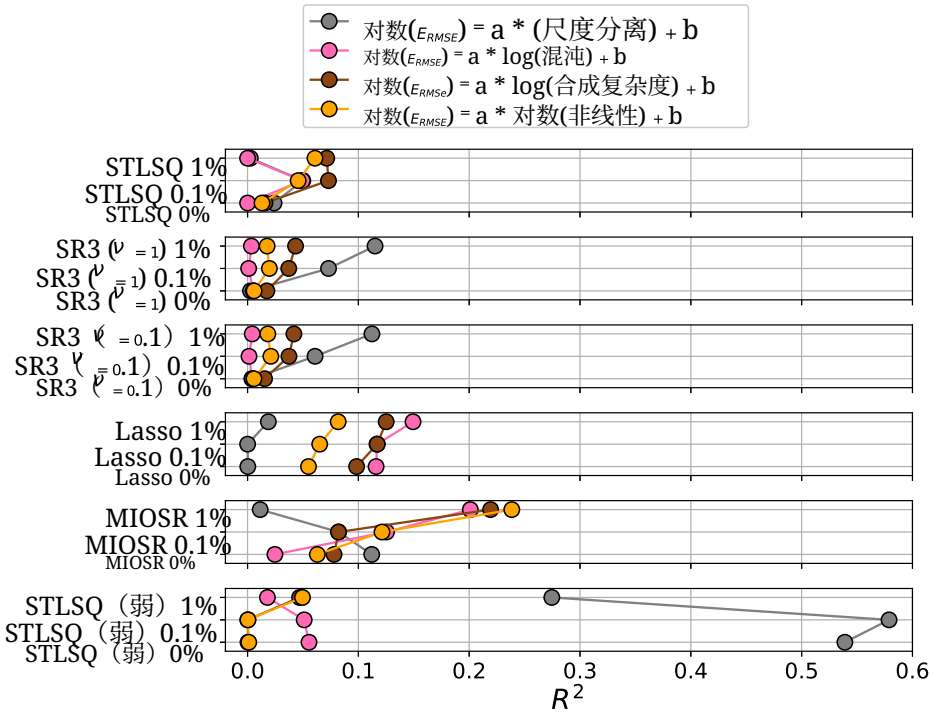


图7：在各优化器的帕累托最优R^2 决定系数与动力学性质之间，选取线性、对数线性和双对数回归中拟合效果最佳者，所得的弱 E_{RMSE} 决定系数汇总。可以观察到各优化器呈现出一些不同的趋势。E_{coef} 与动力学性质之间的相关性甚至更弱（未显示）。

正则化，并使帕累托前沿发生偏移。

在噪声水平为 1% 时，帕累托最优模型的质量通常会下降；例如，STLSQ 得出的一个非典型结果中，只有少数模型的E_{RMSE} 低于 10%，且其中八个模型无法生成任何新轨迹（从略微偏离吸引子的位置开始），使其既能保持有界，又能在定性上近似奇异吸引子的行为。

IV. 结论

本研究中的方法和分析表明，稀疏系统辨识可用于辨识 70 个多项式混沌系统的动力学，重现奇异吸引子，进而重现其基本动力学。我们广泛采用子采样来生成大型模型集，利用赤池信息准则在最优超参数值下辨识集成，并依靠稀疏线性回归的计算速度，在仅两天的中央处理器时间内计算出超过一百万个动力学模型（其中大部分时间耗费在 MIOSR 算法上）。

我们采用这一方法对多种不同的 SINDy 优化器进行了基准测试，结果发现：STLSQ 和 MIOSR 的表现最佳；Lasso 对噪声相对稳健，但除此之外表现较差；而使用 SR3 进行大规模拟合时，需要额外进行超参数扫描才能获得较高性能。随着数据中加入的噪声不断增多，STLSQ 仍能保持快速的计算速度，但开始出现无法弥补的“贪心”错误；相反，MIOSR 的运行速度会显著变慢，但在大多数系统中仍能保持良好表现。最后，我们发现了极具说服力的证据，表明弱形式能显著提升整个数据库上的性能，即使在未添加噪声的情况下也是如此。弱形式可与任何 PySINDy 优化器配合使用，我们建议几乎在所有情况下都采用弱形式。同样，我们强烈建议对数据进行子采样，以构建 SINDy 模型集成 [94]，尤其是在存在噪声的情况下。

我们还开展了首批大规模定量分析之一，研究稀疏回归如何依赖于底层控制方程的动力学性质。总体而言，我们发现动力学性质与帕累托最优 SINDy 模型的性能之间存在非常弱的相关性，但也出现了一些值得注意的偏差，这为未来开展更深入的研究奠定了基础。

未来有许多机会可以在本文所呈现的数据库和分析基础上开展后续研究。例如，可以进一步改进近期关于部分测量的研究成果 [89, 131], , 探讨低数据量、高噪声情形 [94, 116], , 扩展特征库（这些特征库已可在 PySINDy 中构建），以及比较带有凸约束的算法。例如，大多数优化器——STLSQ、SR3、Lasso 和 MIOSR——原则上都可以处理一般的凸约束。不过，未来仍需在 PySINDy 中

实现这些选项。目前，STLSQ [33]和 Lasso 尚不支持约束；而 PySINDy 中 SR3 的约束（线性等式除外）依赖 CVXPY [132] 后端，这会显著降低算法运行速度。有关带二次不等式约束的快得多的 SR3 算法，参见 Kaptanoglu 等人 [133] 。

此外，还有许多其他动力学指标、稀疏系统辨识算法、子采样方法、动力系统类型（即不仅限于由多项式动力学定义的系统），以及其他有待系统测试的参数。尽管这超出了本文的研究范围，dysts 数据库为另一项非常有趣的研究提供了便利：构建数据驱动模型，以预测全局动力学行为中的混沌分岔，例如从呈现奇异吸引子转变为稳定周期解。未来还可以开展定量分析，研究不变流形如何影响系统辨识的性能和实用性；附录 C 讨论了这一主题。最后，许多面向动力学或工程的度量指标（例如方程的描述长度）可以在优化过程中直接最小化；这可能会改变本文基于 SINDy 基准优化问题得出的定性结论，尤其是因为本文发现模型性能与所测试指标之间的相关性较弱。

V. 致谢

AAK 谨此感谢与 William Gilpin、Chris Hansen 和 Nathan Kutz 的宝贵交流。ZGN 是华盛顿研究基金会博士后研究员。作者感谢国家科学基金会动态系统人工智能研究所提供的支持，资助编号为 2112085。

附录A：如何使用此基准

本研究旨在为稀疏系统辨识领域的新旧创新成果提供一种系统化的基准测试方法。为此，我们致力于提供简单直观的脚本，这些脚本基于开源 PySINDy 代码构建。用于执行本文中帕累托最优扫描的脚本可在 PySINDy 的代码示例中找到。该脚本提供多个选项，供用户指定：(1) 数据（包括训练与测试轨迹的点数、采样率等），(2) 优化算法，以及 (3) 特征库（包括是否对特征库进行归一化、如何对数据进行子采样以构建模型集成，以及为得出统计结论而生成的模型数量）。对于 dysts 数据库，该脚本还演示了如何汇总所有方程及其相关的动力学性质。在同一目录中运行 run_all.py 脚本，即可重现本文的全部结果。

PySINDy 代码会定期更新，纳入文献中的新方法，如今还包含许多 SINDy 变体以及高级功能。新的算法、特征库、子采样策略等都可以轻松添加到 PySINDy 代码中，并立即通过上述脚本使用。代码中已预先实现了用于构建通用 SINDy 库和新优化器的大部分后端功能，这应该能显著减少设计和测试新算法所需的额外工作。

附录B：微小模型误差及其相关不稳定性

真正的稀疏回归是一个非凸问题，因为目标函数中包含l_0 损失项。由于在某些条件下l_1 损失项能够以较高概率得到相同的解，因此常用它替代l_0 [134]。当这些条件不满足时，已知 Lasso 有时会选择无关特征 [135]。此外，即使对l_0正则化问题进行了最优求解，当数据受到的噪声污染足够严重时，也可能无法准确预测出正确的特征。在系统辨识中，微小的模型误差可能引发严重问题。

例如，模型识别中的极小误差可能导致从辨识模型生成新轨迹时出现有限时间爆破和非物理不稳定性（不过，如果用户只是想近似这些方程，而不是使用它们生成新轨迹，则不会有这个问题）。对于四阶多项式，我们可以采用以下经验判断：我们预计，三阶和四阶项中的微小系数误差几乎必然会导致某些轨迹不稳定，因为这些项在原点（或吸引子附近）可以忽略不计，但总能选择一个离原点足够远的初始条件，使高阶项占主导地位。只有在受约束的 [33] 和促进稳定性的 [82]设置下，我们才应预期从辨识模型生成的大多数（或全部）轨迹都是稳定的。解的有限时间爆破显然会妨碍对 $\boldsymbol{X}$（而非 $\dot{\boldsymbol{X}}$）均方根误差进行合理比较，而且还会妨碍计算 Lyapunov 谱及其他相关动力学性质。

有两种治标的方法：只使用初始条件与奇异吸引子相当接近的轨迹来测试模型，并避免对 SINDy矩阵重新缩放。第一种方法源于这样的直觉：高阶项上的微小系数误差，只有在状态空间中距离原点足够远的区域才会占据主导地位。如果考虑 STLSQ 算法，后一种方法也很容易理解。此前已有研究指出，例如 Delahunt等人 [85], ，如果$x(t) \sim 10$ 位于奇异吸引子上，那么$x(t)$和$0.1x^2(t)$这两项对时间演化的贡献相近；然而，对于依赖某种硬阈值处理来截去方程中最小系数的算法，后一项会最先被截去。与其将此视为算法的问题，不如将其看作一种算法偏差：倾向于用低阶多项式项描述每个系统；我们认为，这应能间接促成有界性。不幸的是，如果$x(t) \sim 0.1$，情况恰恰相反，因此，对不同的动力系统进行重新缩放可能会有所帮助，

使整个数据库中的这种阈值不匹配方向一致。尽管本文未对此进行探讨，但与传统的单时间步预测不同，最小化与N_t步 SINDy 模型预测相关的目标函数（例如 Kaheman 等人 [87] 或 Bakarji [89], 等人的做法），通常应能提高模型稳定性，但代价是引入非凸性（此时已不再是稀疏线性 回归）。

最后，即使辨识模型几乎完美，并能重现奇异吸引子，报告混沌系统的 $\boldsymbol{X}$均方根误差也往往会产生严重误导，因为积分、辨识出的系数等方面的微小误差，必然会导致轨迹均方根误差出现（起初呈指数级增长，随后有界的）误差。我们发现，更有说服力的做法是直观展示：在新的初始条件下，辨识模型能够近似重现正确的吸引子，因此可以准确计算李雅普诺夫谱及其他动力学性质。

附录 C：动力学性质定义

本节将精确定义本文中计算和分析的动力学性质。

1.混沌程度可以通过多种方式衡量 [1], ，其中一种常见选择是最大李雅普诺夫指数。最大李雅普诺夫指数衡量的是，随时间演化，初始位置无限接近的两点所对应的轨迹彼此呈指数级分离的速率 [136]。该指数在一条轨迹上的变化可能非常大，因此通常将其视为全局平均值。有证据表明 [1] ，深度符号回归 [119] 和循环神经网络在混沌程度较高时性能会下降 [130]。

2. 尺度分离程度可以通过底层方程中最大时间尺度与最小时间尺度之比来定义。然而，即使对于多项式非线性系统，这些时间尺度通常也没有明确定义，因此我们可以改用最大李雅普诺夫指数与最小李雅普诺夫指数之比。和前面一样，李雅普诺夫谱往往会沿着一条轨迹发生巨大变化，因此这只是对尺度分离程度的一种相当粗略的度量。为了更准确地衡量尺度分离，我们采用主导时间尺度与最小显著时间尺度之比，定义见 Gilpin [1]。为理解这种尺度分离度量为何可能与符号回归在动力学上相关，不妨考虑以下简化模型

$$\begin{aligned} \dot{x}_1 &= f(x_1, x_2), \\ \dot{x}_2 &= \frac{1}{\varepsilon} g(x_1, x_2), \end{aligned} \tag{C1}$$

对于 $\varepsilon \ll 1$ ，并假设存在奇异吸引子，如我们的数据集所示。按照不变流形的标准推导，在长时间区间内（相较于 ε 时间尺度）$x_2(t)$ 很快就能找到一个由 $x_2(t) \sim \eta(x_1(t))$ 参数化的平衡态，并且在函数上受$x_1(t)$ 的演化控制。结论是存在一个不变流形，其定义为

$$\begin{aligned} \dot{x}_1 &= f(x_1, \eta(x_1)), \\ \dot{x}_2 &= \partial_{x_1}\eta(x_1)\dot{x}_1 = \partial_{x_1}\eta(x_1) f(x_1, \eta(x_1)), \end{aligned} \tag{C2}$$

它是 (x_1, x_2) 状态空间中的一个流形。如果在 $t \sim \varepsilon$ 期间（或者等价地，如果初始值 $x_2(0)$ 从平衡态或奇异吸引子附近开始）可用的数据点很少，我们就无法合理地期望区分方程 (C1) 和 (C2)。换句话说，为了获得最佳的符号回归结果（即正确捕捉从远离吸引子的初始条件收敛到奇异吸引子所必需的动力学项），我们可能需要使用远离平衡态的初始条件，并以最小时间尺度的分辨率采样。Bucci 等人 [138] 为这一直觉提供了进一步的实践和理论依据。研究明确表明，来自线性不稳定不动点的轨迹所含的熵较少，即信息量较少；而从吸引子上开始的轨迹所含信息更多，因为在不动点附近只有线性项起作用，无法从中获取有关非线性项的信息。类似地，在吸引子上，各项之间存在一种平衡，这会减少关于方程各项的可用信息；相比之下，相空间中的一般轨迹所提供的信息更多。有些混沌系统在吸引子上的有效维数，可能显著小于状态空间维数。

这里还有一个重要说明。我们知道，使用预先定义的四阶多项式特征库进行符号回归，可以发现方程（C1）。但定义不变流形的方程并不一定能被发现，因此我们仍可能只是因为回归无法获取更具体的不变流形方程，才找到了正确的方程。沿着这一思路继续分析颇为复杂，而且往往多个多项式项彼此非常接近线性相关，因此多个稀疏模型都可能很好地拟合数据 [85]。

为了在轨迹离开吸引子时对其进行采样，即当$t \sim \varepsilon$ 时，我们需要以每个周期生成 $\sim \varepsilon^{-1}$ 个点的方式进行采样，这会大幅提高所需的采样分辨率。或者，也可以采用更复杂的加权策略。近期有一些符号回归研究 [139–142] 将系统辨识与不变流形联系起来。未来一项有趣的研究可以尝试通过在搜索 Eq. (C2) 时施加约束，要求用于表示 $\dot{x}_2$ 的库项必须与 $\dot{x}_1$成正比，从而识别映射 $\eta(x_1(t))$。这类

约束已在 PySINDy 中实现。

3. 模型的句法复杂度 在早期研究中已被纳入考虑，既可用于直接优化，也可作为拟合后的度量 [143]。句法复杂度有许多可能的定义，但我们采用近似的描述长度度量 [144]，因为描述长度已在多种情境中用于直接优化，而且可以自然地理解为对描述方程中每个对象所需比特数的近似。

4. 非线性的程度通常也没有明确定义。非线性的最高阶可以定义为支配各个系统的方程中出现的最高阶多项式；以往的研究曾在稀疏回归中将此定义用作复杂度度量 [145]。然而，有些系统包含许多二次非线性项，尽管缺少高阶多项式，我们仍可能认为这类系统具有很强的非线性。为了在这些考量之间进行权衡，我们采用一个权重向量 $w_i =$ $[1,2,3,4,5]$ 并且对每个系统只需报告

$$N_{\text{nonlinearity}} = \sum_{j=1}^{d}\sum_{i=1}^{5} w_i \|x^{i-1} \text{ terms in equation j}\|_0. \quad \text{(C3)}$$

换言之，我们将耦合常微分方程组中出现的所有项相加，并按每一项的多项式次数加权。这包括线性项和常数项，因此它兼顾了非线性程度与方程项的数量。由于该定义考虑了项的数量，因此它可能与描述长度度量所捕捉的信息有所重叠。

我们并不声称这些度量指标中的任何一个都是捕捉特定动力学信息的最佳选择，但每个指标似乎都捕捉到了一个重要方面。Murdoch 等人 [146] 对其中一些问题进行了深入讨论。未来的工作可以研究更复杂的度量指标，例如已辨识 ODE 系统的刚性或可积性，或者探究模型相应的稳定性特征。对于一般的非线性 ODE 系统，研究稳定性似乎难以实现，但关于为多项式非线性系统寻找李雅普诺夫函数的文献十分丰富 [147, 148]。优化刚性 [149]和稳定性，也有助于找到可用于混沌控制 [150] 的模型。

[1]W. Gilpin，《将混沌作为可解释的预测与数据驱动建模基准》，神经信息处理系统进展（NeurIPS），arXiv:2110.05266（2021年）。[2] P. J. Schmid，数值数据与实验数据的动态模态分解，《流体力学杂志》**656**，5（2010）。[3] C. W. Rowley、I. Mezi´c、S. Bagheri、P. Schlatter 和 D. Henningson，非线性流动的谱分析，《流体力学杂志》**641**，115（2009年）。[4] J. N. Kutz、S. L. Brunton、B. W. Brunton 和 J. L. Proctor，动态模态分解：复杂系统的数据驱动建模（工业与应用数学学会，2016年）。[5] I. Mezi´c，动力系统的谱性质、模型降阶与分解，《非线性动力学》**41**，309（2005）。[6] S. L. Brunton、M. Budiši´c、E. Kaiser 和 J. N. Kutz，动力系统的现代 Koopman 理论，《SIAM评论》**64**，229 2022( )。[7] S. A. Billings，非线性系统辨识：时域、频域和时空域中的 *NARMAX* 方法（John Wiley & Sons，2013）。[8] J. Pathak、B. Hunt、M. Girvan、Z. Lu 和 E. Ott，基于数据的大规模时空混沌系统无模型预测：一种储备池计算方法，《物理评论快报》**120**，024102（2018）。[9] P. R. Vlachas、W. Byeon、Z. Y. Wan、T. P. Sapsis 和 P. Koumoutsakos，利用长短期记忆网络对高维混沌系统进行数据驱动预测，《英国皇家学会会刊 A》**474**，20170844（2018）。

[10]M. Raissi、P. Perdikaris 和 G. Karniadakis，物理信息神经网络：用于求解涉及非线性偏微分方程的正问题和反问题的深度学习框架，《计算物理学杂志》**378**, 686 (2019年). [11] M. Raissi、P. Perdikaris 和 G. E. Karniadakis，使用高斯过程学习线性微分方程，《计算物理学杂志》**348**, 683 (2017年). [12] P. Benner、S. Gugercin 和 K. Willcox，参数化动力系统基于投影的模型降阶方法综述，SIAM评论**57**, 483 (2015年). [13] B. Peherstorfer 和 K. Willcox，面向非侵入式基于投影的模型降阶的数据驱动算子推断，《应用力学与工程中的计算方法》**306**, 196 (2016年). [14] E. Qian、B. Kramer、B. Peherstorfer 和 K. Willcox，Lift & Learn：面向大规模非线性动力系统的物理信息机器学习，Physica D: 非线性现象**406**, 132401 (2020年). [15] J. Bongard 和 H. Lipson，非线性动力系统的自动化逆向工程，美国国家科学院院刊**104**, 9943 (2007年). [16] M. Schmidt 和 H. Lipson，从实验数据中提炼自由形式自然定律，《科学》**324**, 81 (2009年). [17] S.-M. Udrescu 和 M. Tegmark，AI Feynman：一种受物理学启发的符号回归方法，《科学进展》**6**, eaay2631 (2020年). [18] S. L. Brunton、J. L. Proctor 和 J. N. Kutz，通过非线性动力系统稀疏辨识从数据中发现控制方程，

《美国国家科学院院刊》**113**, 3932 (2016年). [19] M. Hoffmann、C. Fröhner 和 F. Noé，反应式 SINDy：从浓度数据中发现控制反应，《化学物理学杂志》**150**, 025101 (2019年). [20] B. Bhadriraju、M. S. F. Bangi、A. Narasingam 和 J. S.-I. Kwon，可操作的自适应稀疏系统辨识：在化学过程中的应用，《AIChE期刊》**66**, e16980 (2020年). [21] L. Scheffold、T. Finkler 和 U. Piechottka，使用符号回归对半间歇聚合进行灰箱系统建模与非线性模型预测控制，《计算机与化学工程》**146**, 107204 (2021年). [22] J. Rubio-Herrero、C. O. Marrero 和 W.-T. L. Fan，大气数据建模与动力学辨识：空气污染物的时间数据驱动建模，《清洁生产杂志》**333**, 129863 (2022). [23] J. H. Lagergren、J. T. Nardini、G. Michael Lavigne、E. M. Rutter 和 K. B. Flores，从含噪时空数据中学习生物运输模型的偏微分方程，《皇家学会会刊A辑》**476**, 20190800 (2020年). [24] M. Pasquato、M. Abbas、A. A. Trani、M. Nori、J. A. Kwiecinski、P. Trevisan、V. F. Braga、G. Bono 和 A. V. Macciò，变星动力学的稀疏辨识，《天体物理学杂志》**930**, 161 (2022). [25] Y.-X. Jiang、X. Xiong、S. Zhang、J.-X. Wang、J.-C. Li 和 L. Du，基于 SINDy-LM方法的新冠肺炎传播动力学建模与预测，《非线性动力学》**105**, 2775 (2021年). [26] V. Zucatti、H. F. Lui、D. B. Pitz 和 W. R. Wolf，对流传热降阶建模策略评估，《数值传热学A辑：应用》**77**, 702 (2020年). [27] M. Sorokina、S. Sygletos 和 S. Turitsyn，非线性光通信系统的稀疏辨识：SINO方法，《光学快报》**24**, 30433 (2016年). [28] A. M. Stankovi´c、A. A. Sari´c、A. T. Sari´c 和 M. K. Transtrum，电力系统非线性动力学辨识的数据驱动符号回归，载于*2020*年 *IEEE*电力与能源协会大会 *(PESGM)* (IEEE, 2020) 第 1–5 页。 [29] Y. Cai、X. Wang、G. Joos 和 I. Kamwa，一种基于非线性动力学稀疏辨识（SINDy）的在线数据驱动方法，用于定位发电厂中的强迫振荡源，《IEEE电力系统汇刊》(2022). [30] A. Narasingam 和 J. S.-I. Kwon，使用稀疏回归进行可解释降阶模型的数据驱动辨识，《计算机与化学工程》**119**, 101 (2018年). [31] S. Thaler、L. Paehler 和 N. A. Adams，截断误差的稀疏辨识，《计算物理学杂志》**397** (2019年). [32] R. Dale 和 H. S. Bhat，心智方程：用于推断社会认知系统非线性动力学的数据科学，《认知系统研究》**52**, 275 (2018年). [33] J.-C. Loiseau 和 S. L. Brunton，约束稀疏伽辽金回归，《流体力学杂志》**838**, 42 (2018年). [34] J.-C. Loiseau、B. R. Noack 和 S. L. Brunton，稀疏降阶建模：从基于传感器的动力学到全状态估计，《流体力学杂志》**844**, 459 (2018年). [35] J.-C. Loiseau，环形热虹吸管中混沌热对流的数据驱动建模，《理论与计算流体动力学》**34**, 339 (2020年). [36] Y. El Sayed M、R. Semaan 和 R. Radespiel，周期性柯恩达吹气条件下高升力构型升力增益的稀疏建模，载于*2018*年 *AIAA*航空航天科学会议 (2018) 第 1054 页。

[37] H. Chang 和 D. Zhang，从数据中机器学习地下流动方程，计算地球科学 **23**，895（2019年）。[38] N. Deng、B. R. Noack、M. Morzy´nski 和 L. R. Pastur，流体弹球连续分岔的降阶模型，《流体力学杂志》**884**，A37（2020年）。[39] K. Fukami、T. Murata、K. Zhang 和 K. Fukagata，利用低维流动表征进行稀疏非线性动力学辨识，《流体力学杂志》**926**（2021年）。[40] J. L. Callaham、S. L. Brunton 和 J.-C. Loiseau，论非线性相关性在降阶建模中的作用，《流体力学杂志》**938**（2022）。[41] Z. C. Khoo、C. H. Chan 和 Y. Hwang，壁面湍流降阶模型的稀疏最优闭合，《流体力学杂志》**939**（2022）。[42] N. Deng、B. R. Noack、M. Morzy´nski 和 L. R. Pastur，流体弹球瞬态及瞬态后动力学的 Galerkin 力模型，《流体力学杂志》**918**（2021年）。[43] Q. Xiao、J. Wang、X. Yang 和 B. Jiang，电渗微混合器降阶模型的构建及花瓣结构吸引子的发现，流体物理（2023年）。[44] J. A. Foster、J. Decuyper、T. De Troyer 和 M. Runacres，使用 SINDy 估计流体流动中振荡圆柱绕流的稀疏非线性动力学模型，载于噪声与振动工程会议 *ISMA 2022*（ISMA 2022，2022）第 tba 页。[45] M. Schmelzer、R. P. Dwight 和 P. Cinnella，利用稀疏符号回归发现代数雷诺应力模型，流动、湍流与燃烧**104**，579（2020年）。[46] S. Beetham 和 J. Capecelatro，利用嵌入形式不变性的稀疏回归构建湍流闭合模型，《物理评论：流体》**5**，084611（2020年）。[47] S. Beetham 和 J. Capecelatro，利用稀疏回归和基因表达式编程进行多相湍流建模，arXiv 预印本 arXiv:2106.10397（2021年）。[48] S. Beetham、R. O. Fox 和 J. Capecelatro，耦合流-颗粒流中多相湍流闭合模型的稀疏辨识，《流体力学杂志》**914**（2021年）。[49] J. L. Callaham、G. Rigas、J.-C. Loiseau 和 S. L. Brunton，湍流尾流中对称性破缺的经验平均场模型，《科学进展》**8**, eabm4786（2022）。[50] A. Sansica、J.-C. Loiseau、M. Kanamori、A. Hashimoto 和 J.-C. Robinet，二维跨声速抖振的系统辨识，《美国航空航天学会杂志》**60**，3090（2022）。[51] M. Dam、M. Brøns、J. Juul Rasmussen、V. Naulin 和 J. S. Hesthaven，基于对流模型模拟数据的捕食者-猎物系统稀疏辨识，等离子体物理学**24**，022310（2017年）。[52] A. A. Kaptanoglu、K. D. Morgan、C. J. Hansen 和 S. L. Brunton，磁流体力学的物理约束低维模型：第一性原理与数据驱动方法，《物理评论 E》**104**，015206（2021年）。[53] E. P. Alves 和 F. Fiuza，从全动理学模拟中数据驱动地发现降阶等离子体物理模型，《物理评论研究》**4**，033192（2022）。[54] Z. Lai 和 S. Nagarajaiah，具有迟滞/非弹性行为的非线性动力系统稀疏结构系统辨识方法，机械系统与信号处理**117**，813（2019年）。[55] B. M. de Silva、D. M. Higdon、S. L. Brunton 和 J. N. Kutz，从数据中发现物理规律：普适定律与差异，《人工智能前沿》**3**，25（2020年）。

[56] S. Pan、N. Arnold-Medabalimi 和 K. Duraisamy，发现具有信息性的 Koopman不变子空间的稀疏促进算法，《流体力学杂志》**917**（2021年）。[57] R. Subramanian、R. R. Moar 和 S. Singh，用于识别制造系统整体动力学控制方程的白盒机器学习方法：以蒸馏塔为例，《机器学习与应用》**3**，100014（2021年）。[58] M. Brenner、F. Hess、J. M. Mikhaeil、L. F. Bereska、Z. Mon- fared、P.-C. Kuo 和 D. Durstewitz，用于重构非线性动力系统的可处理树突状循环神经网络，载于国际机器学习会议（PMLR，2022）第2292–2320页。[59] S. Zhang、F. Ahamed 和 H.-S. Song，知识引导的数据驱动建模：稀疏辨识食品微生物灭活过程的控制方程，《食品科学与技术前沿》，29（2022年）。[60] M. Golden、R. Grigoriev、J. Nambisan 和 A. Fernandez-Nieves，活性向列相的物理引导数据驱动建模，arXiv 预印本 arXiv:2202.12853（2022年）。[61] C. Joshi、S. Ray、L. M. Lemma、M. Varghese、G. Sharp、Z. Dogic、A. Baskaran 和 M. F. Hagan，活性向列相流体动力学的数据驱动发现，《物理评论快报》**129**，258001（2022年）。[62] H. Schaeffer，通过数据发现和稀疏优化学习偏微分方程，载于《皇家学会会刊*A*辑》，第473卷（皇家学会，2017年），第20160446页。[63] S. H. Rudy、S. L. Brunton、J. L. Proctor 和 J. N. Kutz，偏微分方程的数据驱动发现，《科学进展》**3**，e1602614（2017年）。[64] A. Sandoz、V. Ducret、G. A. Gottwald、G. Vilmart 和 K. Perron，延迟微分方程的 SINDy：应用于细菌锌响应建模，《皇家学会会刊A辑》**479**，20220556（2023年）。[65] A. Klimovskaia、S. Ganscha 和 M. Claassen，基于稀疏回归的结构学习：从单细胞快照时间序列中学习随机反应网络，《PLoS 计算生物学》**12**，e1005234（2016年）。[66] D. B. Brückner、P. Ronceray 和 C. P. Broedersz，推断欠阻尼随机系统的动力学，《物理评论快报》**125**，058103（2020年）。[67] M. Dai、T. Gao、Y. Lu、Y. Zheng 和 J. Duan，从随机动力系统数据中检测最大似然转移路径，《混沌：非线性科学跨学科期刊》**30**，113124（2020年）。[68] J. L. Callaham、J.-C. Loiseau、G. Rigas 和 S. L. Brunton，利用朗之万回归进行非线性随机建模，《皇家学会会刊A辑》**477**，20210092（2021年）。[69] Y. Huang、Y. Mabrouk、G. Gompper 和 B. Sabass，从数据中对随机微分方程进行稀疏推断与主动学习，《科学报告》**12**，21691（2022）。[70] S. M. Hirsh、D. A. Barajas-Solano 和 J. N. Kutz，模型发现中贝叶斯不确定性量化的稀疏化先验，《皇家学会开放科学》**9**，211823（2022）。[71] J. Bakarji、J. Callaham、S. L. Brunton 和 J. N. Kutz，具有维度一致性的 Buckingham Pi定理学习，《自然计算科学》，1（2022）。[72] S. L. Brunton、J. L. Proctor 和 J. N. Kutz，带控制的稀疏非线性动力学辨识（SINDYc），IFAC 在线论文**49**，710（2016年）。[73] E. Kaiser、J. N. Kutz 和 S. L. Brunton，在低数据量条件下用于模型预测控制的稀疏非线性动力学辨识

，《伦敦皇家学会会刊 A》**474**

(2018年)。[74] E. Kaiser、J. N. Kutz 和 S. L. Brunton，《从数据中发现守恒定律以用于控制》，载于*2018 IEEE* 决策与控制会议（*CDC*）（IEEE，2018年），第 6415–6421 页。[75] N. M. Mangan、S. L. Brunton、J. L. Proctor 和 J. N. Kutz，《通过稀疏非线性动力学辨识推断生物网络》，IEEE《分子、生物与多尺度通信汇刊》**2**，52（2016年）。[76] K. Kaheman、J. N. Kutz 和 S. L. Brunton，《SINDy-PI：一种用于并行隐式稀疏非线性动力学辨识的稳健算法》，《皇家学会会刊A》**476**，20200279（2020年）。[77] N. M. Mangan、T. Askham、S. L. Brunton、J. N. Kutz 和 J. L. Proctor，《通过稀疏回归对混合动力系统进行模型选择》，《皇家学会会刊A》**475**，20180534（2019年）。[78] G. Thiele、A. Fey、D. Sommer 和 J. Krüger，《使用 SINDy 对滞回控制泵系统进行系统辨识》，载于*2020*年第*24*届国际系统理论、控制与计算会议（*ICSTCC*）（IEEE，2020年），第 457–464 页。[79] K. Champion、P. Zheng、A. Y. Aravkin、S. L. Brunton 和 J. N. Kutz，《一种从数据中学习简约物理信息模型的统一稀疏优化框架》，《IEEE Access》**8**，169259（2020年）。[80] N. M. Mangan、J. N. Kutz、S. L. Brunton 和 J. L. Proctor，《通过稀疏回归和信息准则对动力系统进行模型选择》，《皇家学会会刊A：数学、物理与工程科学》**473**，20170009（2017年）。[81] X. Dong、Y.-L. Bai、Y. Lu 和 M. Fan，《结合赤池信息准则和组稀疏性的改进型稀疏非线性动力学辨识》，《非线性动力学》，1（2022）。[82] A. A. Kaptanoglu、J. L. Callaham、A. Aravkin、C. J. Hansen 和 S. L. Brunton，《提升数据驱动二次非线性动力学模型的全局稳定性》，《物理评论流体》**6**，094401（2021年）。[83] G. Tran 和 R. Ward，《从高度损坏的数据中精确恢复混沌系统》，《多尺度建模与仿真》**15**，1108（2017年）。[84] H. Schaeffer、G. Tran 和 R. Ward，《从有限数据中提取稀疏高维动力学》，工业与应用数学学会《应用数学杂志》**78**，3279（2018年）。[85] C. B. Delahunt 和 J. N. Kutz，《一种用于在高噪声状态下数据驱动发现控制方程的工具包》，《IEEE Access》**10**，31210（2022）。[86] J. Wentz 和 A. Doostan，《基于导数的 SINDy（DSINDy）：应对从含噪数据中发现控制方程的挑战》，arXiv 预印本 arXiv:2211.05918（2022）。[87] K. Kaheman、S. L. Brunton 和 J. N. Kutz，《通过自动微分同时辨识非线性动力学并从数据中提取噪声概率分布》，《机器学习：科学与技术》**3**，015031（2022）。[88] A. Somacal、Y. Barrera、L. Boechi、M. Jonckheere、V. Lefieux、D. Picard 和 E. Smucler，《利用隐藏变量从数据中揭示微分方程》，《物理评论 E》**105**，054209（2022）。[89] J. Bakarji、K. Champion、J. N. Kutz 和 S. L. Brunton，《利用深度延迟自编码器从部分测量中发现控制方程》，arXiv 预印本 arXiv:2201.05136（2022）。[90] P. Conti、G. Gobat、S. Fresca、A. Manzoni 和 A. Frangi，《通过自编码器和 SINDy 方法对参数化系统进行降阶建模：周期解的延拓》

，arXiv 预印本 arXiv:2211.06786（2022）。

[91] L. Gao 和 J. N. Kutz，贝叶斯自编码器：用于数据驱动地发现坐标、控制方程和基本常数，arXiv 预印本 arXiv:2211.10575（2022）。[92] Z. Zhao 和 Q. Li，学习动力系统的自适应采样方法，载于数学与科学机器学习（PMLR，2022），第 335–350 页。[93] K. Wu 和 D. Xiu，使用数据近似控制方程的数值方面，《计算物理学杂志》**384**，200（2019）。[94] U. Fasel、J. N. Kutz、B. W. Brunton 和 S. L. Brunton，集成 SINDy：在低数据、高噪声极限下，结合主动学习和控制实现稳健的稀疏模型发现，《英国皇家学会学报 A》**478**，20210904（2022）。[95] L. Gao、U. Fasel、S. L. Brunton 和 J. N. Kutz，集成与贝叶斯稀疏模型发现中不确定性估计的收敛性，arXiv 预印本 arXiv:2301.12649（2023）。[96] F. Jiang 、 L. Du 、 F. Yang 和 Z.-C. Deng，基于正则化最小绝对偏差的动力系统稀疏辨识，《混沌：非线性科学跨学科期刊》**33**，013103（2023）。[97] H. Schaeffer 和 S. G. McCalla，通过积分项进行稀疏模型选择，《物理评论 E》**96**，023302（2017）。[98] D. R. Gurevich、P. A. Reinbold 和 R. O. Grigoriev，非线性偏微分方程模型的稳健且最优稀疏回归，《混沌：非线性科学跨学科期刊》**29**，103113（2019）。[99] P. A. Reinbold 、 D. R. Gurevich 和 R. O. Grigoriev，利用含噪或不完整数据发现时空动力学模型，《物理评论 E》**101**，010203（2020）。[100] D. A. Messenger 和 D. M. Bortz，用于偏微分方程的弱 SINDy，《计算物理学杂志》，110525（2021）。[101] L. M. Kageorge、R. O. Grigoriev 和 M. F. Schatz，数据驱动检测漂移的系统参数，arXiv 预印本 arXiv:2111.12114（2021）。[102] P. A. Reinbold、L. M. Kageorge、M. F. Schatz 和 R. O. Grigoriev，通过物理约束符号回归，从含噪、不完整的高维实验数据中稳健学习，《自然·通讯》**12**，1（2021）。[103] D. R. Gurevich、P. A. Reinbold 和 R. O. Grigoriev，利用稀疏物理信息发现经验关系（SPIDER），从高度湍流数据中学习流体物理，arXiv 预印本 arXiv:2105.00048（2021）。[104] B. Russo 和 M. P. Laiu，弱 SINDy 代理模型的收敛性，arXiv 预印本 arXiv:2209.15573（2022）。[105] D. A. Messenger 和 D. M. Bortz，在连续数据极限下 WSINDy 算法的渐近一致性，arXiv 预印本 arXiv:2211.16000（2022）。[106] D. A. Messenger 和 D. M. Bortz，利用 WSINDy 从粒子数据中学习平均场方程，《物理学D：非线性现象》**439**，133406（2022）。[107] D. A. Messenger 、 E. Dall’Anese 和 D. Bortz，偏微分方程的在线弱形式稀疏辨识，载于数学与科学机器学习（PMLR，2022），第 241–256 页。[108] P. Gelß、S. Klus、J. Eisert 和 C. Schütte，非线性动力系统的多维近似，《计算与非线性动力学杂志》**14**（2019）。[109] A. Goeßmann、M. Götte、I. Roth、R. Sweke、G. Kutyniok 和 J. Eisert，用于学习非线性动力学定律的张量网络方法，arXiv 预印本 arXiv:2002.12388（2020）。[110] L. Boninsegna、F. Nüske 和 C. Clementi，随机动力方程的稀疏学习，《化学物理学杂志》**148**

, 241723（2018）。

[111] E. N. Lorenz，确定性非周期流，《大气科学杂志》**20**，130（1963）。[112] B. de Silva、K. Champion、M. Quade、J.-C. Loiseau、J. N. Kutz 和 S. Brunton，PySINDy：用于从数据中稀疏辨识非线性动力系统的 Python 软件包，《开放源码软件杂志》**5**，1（2020）。[113] A. A. Kaptanoglu、B. M. de Silva、U. Fasel、K. Kaheman、A. J. Goldschmidt、J. Callaham、C. B. Delahunt、Z. G. Nicolaou、K. Champion、J.-C. Loiseau、J. N. Kutz 和 S. L. Brunton，PySINDy：用于稳健稀疏系统辨识的综合 Python 软件包，《开放源码软件杂志》**7**，3994（2022）。[114] P. Zheng、T. Askham、S. L. Brunton、J. N. Kutz 和 A. Y. Aravkin，稀疏松弛正则化回归的统一框架：SR3，《IEEE Access》**7**，1404（2019）。[115] R. Tibshirani、M. Wainwright 和 T. Hastie，稀疏性统计学习：*Lasso* 及其推广（Chapman and Hall/CRC，2015）。[116] D. Bertsimas 和 W. Gurnee，通过混合整数优化学习稀疏非线性动力学，《非线性动力学》，1（2023）。[117] H. Schaeffer、G. Tran 和 R. Ward，通过组稀疏性学习动力系统与分岔，arXiv 预印本 arXiv:1709.01558（2017）。[118] N. Q. Uy、N. X. Hoai、M. O’Neill、R. I. McKay 和 E. Galván-López，遗传编程中的语义交叉：在实值符号回归中的应用，《遗传编程与可进化机器》**12**，91（2011）。[119] B. K. Petersen、M. L. Larma、T. N. Mundhenk、C. P. Santiago、S. K. Kim 和 J. T. Kim，深度符号回归：通过风险偏好型策略梯度从数据中恢复数学表达式，载于国际学习表征会议（2021）。[120] H. Kantz 和 T. Schreiber，非线性时间序列分析，第7卷（剑桥大学出版社，2004）。[121] A. Kaptanoglu，等离子体物理学中的数据驱动系统辨识与机器学习探索（华盛顿大学，2021）。[122] F. Van Breugel、J. N. Kutz 和 B. W. Brunton，含噪数据的数值微分：统一的多目标优化框架，《IEEE Access》**8**，196865（2020）。[123] F. Van Breugel、Y. Liu、B. W. Brunton 和 J. N. Kutz，Pynumdiff：用于含噪时间序列数据数值微分的 Python 软件包，《开放源码软件杂志》**7**，4078（2022）。[124] X. Blasco、J. M. Herrero、J. Sanchis 和 M. Martínez，用于多目标优化决策的新型 n 维 Pareto 前沿图形可视化，《信息科学》**178**，3908（2008）。[125] W. La Cava、P. Orzechowski、B. Burlacu、F. O. de Franca、M. Virgolin、Y. Jin、M. Kommenda 和 J. H. Moore，当代符号回归方法及其相对性能，载于第三十五届神经信息处理系统大会：数据集与基准测试专题（第一轮）。[126] P. Orzechowski、W. La Cava 和 J. H. Moore，我们目前处于什么阶段？近期符号回归方法的大规模基准测试研究，载于遗传与进化计算会议论文集（2018），第1183–1190页。[127] H. S. Bhat，经典哈密顿系统势函数的学习与解释，载于欧洲机器学习与数据库知识发现联合会议（Springer，2019），第217–228页。[128] H. K. Chu 和 M. Hayashibe，通过促进能量与拉格朗日量的稀疏性来发现可解释的动力学

,

IEEE机器人与自动化快报**5**，2154（2020）。[129] T. Bertalan、F. Dietrich、I. Mezi´c 和 I. G. Kevrekidis，关于从数据中学习哈密顿系统，《混沌：非线性科学跨学科期刊》**29**，121107（2019）。[130] J. M. Mikhaeil、Z. Monfared 和 D. Durstewitz，利用循环神经网络学习混沌动力学的难点，arXiv 预印本 arXiv:2110.07238（2021）。[131] S. Ouala、S. L. Brunton、B. Chapron、A. Pascual、F. Collard、L. Gaultier 和 R. Fablet，利用物理约束深度学习对部分观测的地球物理系统进行有界非线性预测，《物理学D：非线性现象》，133630（2023）。[132] S. Diamond 和 S. Boyd，CVXPY：用于凸优化的 Python 嵌入式建模语言，《机器学习研究杂志》**17**，2909（2016）。[133] A. A. Kaptanoglu、T. Qian、F. Wechsung 和 M. Landreman，将永磁体优化作为稀疏回归用于仿星器，《物理评论应用》**18**，044006（2022）。[134] M. J. Wainwright，使用$l-1$-约束二次规划（Lasso）进行高维含噪稀疏性恢复的精确阈值，《IEEE信息论汇刊》**55**，2183（2009）。[135] D. Bertsimas、J. Pauphilet 和 B. V. Parys，稀疏回归：可扩展算法与经验性能，《统计科学》**35**，555（2020）。[136] J. C. Sommerer 和 E. Ott，漂浮在运动流体上的粒子：一种动力学上可理解的物理分形，《科学》**259**，335（1993）。[137] G. Pavliotis 和 A. Stuart，多尺度方法：平均化与均匀化（Springer Science & Business Media，2008）。[138] A. Bucci、O. Semeraro、A. Allauzen、S. Chibbaro 和 L. Mathelin，用于动力系统数据驱动建模的课程学习（2021）。[139] J. J. Bramburger、D. Dylewsky 和 J. N. Kutz，慢时间尺度动力学的稀疏辨识，《物理评论 E》**102**，022204（2020）。[140] M. Cenedese、J. Axås、B. Bäuerlein、K. Avila 和 G. Haller，基于谱子流形对不可线性化动力学进行数据驱动建模与预测，《自然·通讯》**13**，1 2022( )。[141] R. Szalai，利用不变叶状结构、流形和自编码器进行数据驱动的降阶建模，arXiv预印本 arXiv:2206.12269（2022）。[142] J. Axås、M. Cenedese 和 G. Haller，非线性动力系统的快速数据驱动模型降阶，《非线性动力学》，1（2022）。[143] S.-M. Udrescu、A. Tan、J. Feng、O. Neto、T. Wu 和 M. Tegmark，AI Feynman 2.0：利用图模块度的帕累托最优符号回归，《神经信息处理系统进展》**33**，4860（2020）。[144] P. D. Grünwald、I. J. Myung 和 M. A. Pitt，最小描述长度进展：理论与应用（麻省理工学院出版社，2005）。[145] E. J. Vladislavleva、G. F. Smits 和 D. Den Hertog，作为通过帕累托遗传编程进行符号回归所生成模型之复杂度度量的非线性阶数，《IEEE进化计算汇刊》**13**，333（2008）。[146] W. J. Murdoch、C. Singh、K. Kumbier、R. Abbasi-Asl 和 B. Yu，可解释机器学习的定义、方法与应用，《美国国家科学院院刊》**116**，22071（2019）。[147] A. Tesi、F. Villoresi 和 R. Genesio，关于利用二次李雅普诺夫函数估计稳定域：多项式系统的凸性与最优性性质，《IEEE自动控制汇刊》**41**，1650（1996）。[148] A. A. Ahmadi、A. Majumdar 和 R. Tedrake，连续时间动力系统中十个判定问题的复杂度，载于*2013*年美国控制会议（IEEE，2013），第6376–6381页。[149] H. E. Dikeman、H. Zhang 和 S. Yang，用于燃烧化学动力学数据驱动降阶建模的刚性降低神经常微分方程模型，载于*AIAA SCITECH 2022*论坛（2022），第0226页。[150] E. Ott、C. Grebogi 和 J. A. Yorke，控制混沌，《物理评论快报》**64**，1196（1990）。