

## 基于子空间追踪和线性多步方法的模型识别\*

江月梅, 陈浩

(重庆师范大学 数学科学学院, 重庆 401331)

**摘要:**结合传统的数值分析技术,寻求一种高精度的稀疏识别方法,重构非线性动力系统。首先,需要构造一个合适的基函数库,利用该基函数库近似潜在的非线性动力系统。其次,应用线性多步方法离散近似后的非线性动力系统。然后,在状态数据含噪声的情况下,引入广义最小二乘方法的原理,计算噪声残差项的近似协方差矩阵,并利用该矩阵对上述过程得到的优化问题进行加权,从而降低噪声对模型识别结果的影响。最后,通过子空间追踪算法从数据中挑选出系数误差最小的特征集合作为下一次迭代的基函数库,并在迭代终止以后,使用最小二乘方法计算保留下来的特征的对应系数值。得到了稀疏识别非线性动力系统的高精度线性多步子空间追踪算法,且该算法具有较好的鲁棒性。通过数值分析验证了该算法的有效性。

**关键词:**模型识别;子空间追踪;稀疏回归;线性多步方法;广义最小二乘法

**中图分类号:**O241

**文献标志码:**A

**文章编号:**1672-6693(2025)03-0040-12

随着科学技术的快速发展,社会的生产效率和居民的生活质量得到持续提升,越来越多的数据能够被记录并利用起来。因此如何从数量庞大且冗余的数据中挖掘出潜在的且有意义的动态规律,并利用所学到的动态过程进行设计、控制和决策,逐渐发展成了一种令人着迷的科学研究,即模型识别。尤其在近些年,针对常微分方程和偏微分方程,人们做了许多的研究工作来发现隐藏在数据里的关键信息<sup>[1-2]</sup>。

由先前的研究可知,模型识别的过程往往需要足够庞大的数据集和基函数库作为支撑,尤其对于高维复杂的问题而言,模型识别的难度也会随之增加。面对这种情况,人们通常会采用稀疏回归的方法进行处理<sup>[3]</sup>。众所周知,稀疏回归方法的本质是求解一个范数最小化问题,且该问题通常可以用 LASSO 算法<sup>[4]</sup>或 STLS 算法<sup>[5]</sup>进行求解。另外,这类算法中最具有代表性的算法是 SINDy 算法<sup>[6]</sup>,该算法背后的依据是,基函数库中的特征只有一小部分对动态过程是有意义的,因此能否从基函数库中构建出正确的模型框架往往对模式识别起着关键的作用。在过去的研究中,SINDy 算法已经被证实在模型识别问题上具有突出的表现<sup>[6-10]</sup>。并且,与经典的数值方法相结合的 SINDy 衍化算法被发现用于识别潜在动力模型也非常有效。例如,Goyal 等人<sup>[11]</sup>提出的基于龙格库塔方法的 SINDy 衍化算法,该算法不仅具备了龙格库塔方法的高精度的特点,同时还具备了 SINDy 算法在稀疏识别上的优势。还有 Chen<sup>[3]</sup>提出的基于线性多步方法的 SINDy 衍化算法,其中的优化问题利用 STLS 算法进行求解。此外,基于线性多步方法的神经网络算法<sup>[12-13]</sup>也展现出了很好的模型识别能力。

子空间追踪算法<sup>[14]</sup>是求解  $l_1$  范数优化问题的一类优秀算法,目前尚未见到该算法与线性多步方法结合用于模型识别的相关研究。基于此,本文提出了一种基于线性多步方法和子空间追踪算法的稀疏识别方法,即线性多步子空间追踪(linear multistep method subspace pursuit, LMM-SP)算法。该方法首先需要构造一个包含所有可能的合适的基函数库,利用该基函数库近似潜在动力系统。然后应用线性多步方法离散近似后的动力系统,与应用欧拉方法离散动力系统相比,本文的方法具有更高的精度。随后,利用子空间追踪算法<sup>[14]</sup>解决上述过程得到的范数最小化问题,并在子空间追踪算法的每一次迭代过程中,挑选出系数误差最小的特征集合作为下一次迭代的基函数库。在迭代终止以后,根据输出的最小特征集,使用最小二乘方法计算保留下来的特征的对应系数值。与 LASSO 算法和 STLS 算法相比,子空间追踪算法可以将稀疏性作为输入,从而直接控制重构系数

\* 收稿日期:2023-03-13 修回日期:2025-02-25 网络出版时间:2025-05-19T08:43

资助项目:国家自然科学基金面上项目(No. 11971085);重庆市自然科学基金(No. cstc2021jcyj-msxmX0034)

第一作者简介:江月梅,女,研究方向为常微分方程的模型识别,E-mail:jjsuanjym@126.com;通信作者简介:陈浩,男,教授,博士,E-mail:hch@cqnu.edu.cn

网络出版地址:https://link.cnki.net/urlid/50.1165.n.20250515.1658.016

的稀疏性<sup>[15]</sup>。不过,由于现实生活中收集到的数据往往是带有噪声的观测数据,直接从这样的数据中去识别潜在在动力模型,可能会导致识别的结果具有较大的误差。基于此,本文结合广义最小二乘方法求解线性回归问题,目的是使所得估计为最佳线性无偏估计,并提高模型识别的精度和鲁棒性。

## 1 问题描述

一般的非线性动力系统可以表示为:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)), \mathbf{x}(0) = \mathbf{x}_0, 0 \leq t \leq T, \quad (1)$$

其中: $\mathbf{x}$ 表示状态变量向量, $\mathbf{f}$ 是依赖于 $\mathbf{x}$ 的右端函数向量,且 $\mathbf{f}$ 是未知的。在这里,向量 $\mathbf{x}_0 = [x_1(0), x_2(0), \dots, x_d(0)]^T$ 表示动力系统(1)的初始条件, $d$ 表示动力系统的维数。设 $N$ 为正整数,令 $h = T/N$ 表示时间步长, $t_n = nh, n = 0, 1, \dots, N$ 。那么该非线性动力系统的状态数据矩阵可以表示为:

$$\mathbf{X} = \begin{bmatrix} x_1(t_0) & x_2(t_0) & \cdots & x_d(t_0) \\ x_1(t_1) & x_2(t_1) & \cdots & x_d(t_1) \\ \vdots & \vdots & & \vdots \\ x_1(t_N) & x_2(t_N) & \cdots & x_d(t_N) \end{bmatrix},$$

右端函数矩阵可以表示为:

$$\mathbf{F}(\mathbf{X}) = \begin{bmatrix} f_1(\mathbf{x}(t_0)) & f_2(\mathbf{x}(t_0)) & \cdots & f_d(\mathbf{x}(t_0)) \\ f_1(\mathbf{x}(t_1)) & f_2(\mathbf{x}(t_1)) & \cdots & f_d(\mathbf{x}(t_1)) \\ \vdots & \vdots & & \vdots \\ f_1(\mathbf{x}(t_N)) & f_2(\mathbf{x}(t_N)) & \cdots & f_d(\mathbf{x}(t_N)) \end{bmatrix}。$$

接下来,需要挖掘隐藏在状态数据 $\mathbf{x}$ 里的潜在模型 $\mathbf{f}$ 。根据以往的经验,首先需要构建一个包含所有可能候选基函数的基函数库 $\varphi_j(\mathbf{x})$ 。并且,为了找到一个合适的基函数库,一般会对潜在的动力系统进行更多的了解和专业分析,对其中包含的特征种类进行预挑选,使得最终确定下来的基函数库尽可能覆盖精确的动力系统,从而提高模型识别的精度。一般来说,可以假设状态函数 $f_i(\mathbf{x}), i = 1, 2, \dots, d$ 是基函数库 $\varphi_j(\mathbf{x})$ 的一个线性组合:

$$\dot{x}_i = f_i(\mathbf{x}) = \sum_{j=1}^q c_{ji}^* \varphi_j(\mathbf{x}), i = 1, 2, \dots, d, \quad (2)$$

其中: $c_{ji}^*$ 是对应基函数 $\varphi_j(\mathbf{x})$ 的系数, $q$ 表示基函数库中元素的个数。为了简便,将式(2)写为如下矩阵形式:

$$\dot{\mathbf{X}} = \Phi(\mathbf{X})\mathbf{C}^*, \quad (3)$$

$$\Phi(\mathbf{X}) = \begin{bmatrix} \varphi_1(\mathbf{x}(t_0)) & \varphi_2(\mathbf{x}(t_0)) & \cdots & \varphi_q(\mathbf{x}(t_0)) \\ \varphi_1(\mathbf{x}(t_1)) & \varphi_2(\mathbf{x}(t_1)) & \cdots & \varphi_q(\mathbf{x}(t_1)) \\ \vdots & \vdots & & \vdots \\ \varphi_1(\mathbf{x}(t_d)) & \varphi_2(\mathbf{x}(t_d)) & \cdots & \varphi_q(\mathbf{x}(t_d)) \end{bmatrix},$$

$$\dot{\mathbf{X}} = \begin{bmatrix} \dot{x}_1(t_0) & \dot{x}_2(t_0) & \cdots & \dot{x}_d(t_0) \\ \dot{x}_1(t_1) & \dot{x}_2(t_1) & \cdots & \dot{x}_d(t_1) \\ \vdots & \vdots & & \vdots \\ \dot{x}_1(t_N) & \dot{x}_2(t_N) & \cdots & \dot{x}_d(t_N) \end{bmatrix}, \mathbf{C}^* = \begin{bmatrix} c_{11}^* & c_{12}^* & \cdots & c_{1d}^* \\ c_{21}^* & c_{22}^* & \cdots & c_{2d}^* \\ \vdots & \vdots & & \vdots \\ c_{q1}^* & c_{q2}^* & \cdots & c_{qd}^* \end{bmatrix}。$$

其中: $\Phi(\mathbf{X})$ 表示特征矩阵, $\mathbf{C}^*$ 表示未知的系数矩阵。由于大多数动力模型的右端往往由少量的几个特征所刻画,因此矩阵 $\mathbf{C}^*$ 中的元素大多数都为0,只有极少部分元素是非零的,那么只需要寻找方程(3)的稀疏解。

基函数库一般由数据矩阵 $\mathbf{X}$ 的单项式和三角函数组成:

$$\Phi(\mathbf{X}) = [1, \mathbf{X}, \mathbf{X}^2, \dots, \mathbf{X}^k, \dots, \sin(\mathbf{X}), \cos(\mathbf{X}), \dots],$$

其中: $\mathbf{X}^k$ 表示一个单项式矩阵且列向量由 $\mathbf{X}$ 矩阵中列向量的所有可能的 $k$ 次单项式组成。

**注1** 当所给的动力方程不能被上述候选基函数表示时,一种常用的解决方法是引入额外的候选基函数,通过调整基函数库和加入有助于逼近目标函数的函数形式,可以提高模型识别的精度。不过,对于复杂的非线性函数,例如有理函数和复合函数,这种改进方法对于提高模型识别的准确性可能仍然有限。近些年来,随着

机器学习技术的发展和该技术在解决传统数值问题上的优点,可以选用两层神经网络中的线性随机特征函数来近似动力方程的右端项,笔者的后续工作中也证实了使用这种方法构造基函数库用于模型识别具有较好的精度和鲁棒性。

**注 2** 对于不能直接使用基函数表示所给方程的这种情况,可以采用预测误差来评价估计量的质量。首先设置训练集和测试集,在训练集上使用模型识别方法重构潜在的动力模型,再应用得到的动力模型进行预测,将预测值与测试集中的真实值进行比较,得出预测误差,较小的预测误差表示估计量有较好的拟合能力。

**注 3** 基于已有的研究工作可知,绝大多数稀疏识别方法都需要采用确定的基函数库来重构动力方程。在未知原函数的情况下,可以根据动力系统的物理特性和动力学原理,选择与系统行为和结构相关的基函数。例如,多项式函数可以逼近非线性关系,正余弦函数可以表示振动的周期性行为,指数函数可以表示指数衰减或增长。通过这种方式,可以提高模型识别的准确性。

为了呈现所提出的方法从不同质量的数据中挖掘隐藏动力模型的性能,考虑在精确数据中加入不同水平的噪声,并将加入噪声以后得到的数据表示为观测数据:  $\mathbf{Y} = \mathbf{X} + \boldsymbol{\varepsilon}$ , 其中:  $\mathbf{Y}$  是对应精确数据矩阵  $\mathbf{X}$  的观测数据矩阵,  $\boldsymbol{\varepsilon}$  表示一个均值为 0、方差为  $\sigma^2$  且独立同分布的高斯白噪声矩阵,  $\boldsymbol{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_d]$  且  $\boldsymbol{\varepsilon} \in \mathbf{R}^{(N+1) \times d}$ 。那么,不难发现未知函数  $f(\mathbf{x})$  的问题现在被归结为求解一个稀疏系数矩阵  $\mathbf{C} \in \mathbf{R}^{q \times d}$ , 使得  $\mathbf{C}$  满足:

$$\dot{\mathbf{Y}} \approx \boldsymbol{\Phi}(\mathbf{Y})\mathbf{C}, \quad (4)$$

其中:  $\dot{\mathbf{Y}}$  是对应观测矩阵  $\mathbf{Y}$  的数值时间导数矩阵,  $\boldsymbol{\Phi}(\mathbf{Y})$  是与  $\boldsymbol{\Phi}(\mathbf{X})$  类似定义的基函数库。

接下来,需要使用稀疏回归的方法求解问题(4),也就是解决如下的最小化问题:

$$\min_{\mathbf{C}} \|\boldsymbol{\Phi}\mathbf{C} - \dot{\mathbf{Y}}\|_2^2 + \lambda \|\mathbf{C}\|_0, \quad (5)$$

其中: 矩阵  $\mathbf{C}$  的  $l_0$  范数  $\|\mathbf{C}\|_0$  用于描述矩阵  $\mathbf{C}$  的稀疏性, 参数  $\lambda > 0$  用于在式(4)的精度和解的稀疏性之间建立一种平衡。

## 2 LMM-SP 算法

### 2.1 子空间追踪算法

本节将介绍用于稀疏识别动力系统的 LMM-SP 算法,为了更好地理解该算法,下面将从子空间追踪算法、线性多步方法和广义最小二乘法 3 个板块进行详细的理论阐述和公式推导。首先,介绍用于寻找稀疏向量的非零位置的子空间追踪算法<sup>[15]</sup>:

步骤 1: 输入  $\boldsymbol{\Phi} \in \mathbf{R}^{(N+1) \times q}$ ,  $\mathbf{b} \in \mathbf{R}^{(N+1) \times 1}$  和稀疏性  $k \in \mathbf{N}$ 。

步骤 2: 初始化  $j = 0$ ;  $\mathbf{G} \leftarrow \boldsymbol{\Phi}$  的列标准化矩阵;  $I^0 = \{k \text{ 对应向量 } \mathbf{G} * \mathbf{b} \text{ 中最大内积的指标}\}$ ;  $\mathbf{b}_{\text{res}}^0 = \mathbf{b} - \mathbf{G}_{I^0} \mathbf{G}_{I^0}^\dagger \mathbf{b}$ 。

步骤 3: 继续循环

步骤 3.1:  $\tilde{I}^{j+1} = I^j \cup \{k \text{ 对应向量 } \mathbf{G} * \mathbf{b}_{\text{res}}^j \text{ 中最大内积的指标}\}$ ;

步骤 3.2: 令  $\mathbf{c}_l = \mathbf{G}_{\tilde{I}^{j+1}}^\dagger \mathbf{b}$ ;

步骤 3.3:  $I^{j+1} = \{k \text{ 对应向量 } \mathbf{c}_l \text{ 中数值最大的指标}\}$ ;

步骤 3.4: 计算  $\mathbf{b}_{\text{res}}^{j+1} = \mathbf{b} - \mathbf{G}_{I^{j+1}} \mathbf{G}_{I^{j+1}}^\dagger \mathbf{b}$ , 如果  $\|\mathbf{b}_{\text{res}}^{j+1}\|_2 > \|\mathbf{b}_{\text{res}}^j\|_2$ , 让  $I^{j+1} = I^j$ , 终止算法; 否则更新  $j \leftarrow j + 1$ , 继续步骤 3。

步骤 4: 输出  $\hat{\mathbf{c}} \in \mathbf{R}^q$  满足  $\hat{\mathbf{c}}_{I_j} = \mathbf{G}_{I_j}^\dagger \mathbf{b}$ ,  $\hat{\mathbf{c}}_{(I_j)^c} = 0$ 。

为了提高模型识别的精度和算法的鲁棒性,下面将子空间追踪算法与经典的线性多步方法相结合,考虑在离散的时间框架下识别非线性动力系统。

### 2.2 线性多步方法

本文应用 3 种线性多步方法对动力系统进行了离散,包括显式 Adams 方法、隐式 Adams 方法和向后微分公式。并且在每一类线性多步方法下面,又包括了对应不同步数、不同阶数的 5 种线性多步方法。其中,应用  $s$  步  $p$  阶线性多步法对动力系统(1)进行离散的分量结果是:

$$\sum_{j=0}^s \alpha_j x_i(t_{n+j}) = h \sum_{j=0}^s \beta_j f_i(\mathbf{x}(t_{n+j})) + O(h^{p+1}), n = 0, \dots, N - s, \quad (6)$$

在这里  $\alpha_j, \beta_j (j=0, 1, \dots, s)$  表示线性多步方法的离散系数,它依赖于时间步长  $h$ 、样本数量  $n$  和潜在的微分方程。通常情况下,令  $\alpha_s = 1$ ,可以对式(6)实现标准化处理。注意,当  $\beta_s = 0$  时,式(6)为显式 Adams 方法;当  $\beta_s \neq 0$  时,式(6)为隐式 Adams 方法。为了下文更好地开展实验,现将各线性多步方法的离散公式及离散系数列举如下。

1) 显式 Adams 方法的离散公式为:

$$\begin{aligned} s=1: y_{n+1} &= y_n + hf_n; \\ s=2: y_{n+1} &= y_n + h\left(\frac{3}{2}f_n - \frac{1}{2}f_{n-1}\right); \\ s=3: y_{n+1} &= y_n + h\left(\frac{23}{12}f_n - \frac{16}{12}f_{n-1} + \frac{5}{12}f_{n-2}\right); \\ s=4: y_{n+1} &= y_n + h\left(\frac{55}{24}f_n - \frac{59}{24}f_{n-1} + \frac{37}{24}f_{n-2} - \frac{9}{24}f_{n-3}\right); \\ s=5: y_{n+1} &= y_n + h\left(\frac{1901}{720}f_n - \frac{2774}{720}f_{n-1} + \frac{2616}{720}f_{n-2} - \frac{1274}{720}f_{n-3} + \frac{251}{720}f_{n-4}\right). \end{aligned}$$

2) 隐式 Adams 方法的离散公式为:

$$\begin{aligned} s=1: y_{n+1} &= y_n + h\left(\frac{1}{2}f_{n+1} + \frac{1}{2}f_n\right); \\ s=2: y_{n+1} &= y_n + h\left(\frac{5}{12}f_{n+1} + \frac{8}{12}f_n - \frac{1}{12}f_{n-1}\right); \\ s=3: y_{n+1} &= y_n + h\left(\frac{9}{24}f_{n+1} + \frac{19}{24}f_n - \frac{5}{24}f_{n-1} + \frac{1}{24}f_{n-2}\right); \\ s=4: y_{n+1} &= y_n + h\left(\frac{251}{720}f_{n+1} + \frac{646}{720}f_n - \frac{264}{720}f_{n-1} + \frac{106}{720}f_{n-2} - \frac{19}{720}f_{n-3}\right); \\ s=5: y_{n+1} &= y_n + h\left(\frac{95}{288}f_{n+1} + \frac{1427}{1440}f_n - \frac{133}{240}f_{n-1} + \frac{241}{720}f_{n-2} - \frac{173}{1440}f_{n-3} + \frac{3}{160}f_{n-4}\right). \end{aligned}$$

3) 向后微分公式线性多步方法的离散公式为:

$$\begin{aligned} s=1: y_{n+1} - y_n &= hf(t_{n+1}, y_{n+1}); \\ s=2: y_{n+2} - \frac{4}{3}y_{n+1} + \frac{1}{3}y_n &= \frac{2}{3}hf(t_{n+2}, y_{n+2}); \\ s=3: y_{n+3} - \frac{18}{11}y_{n+2} + \frac{9}{11}y_{n+1} - \frac{2}{11}y_n &= \frac{6}{11}hf(t_{n+3}, y_{n+3}); \\ s=4: y_{n+4} - \frac{48}{25}y_{n+3} + \frac{36}{25}y_{n+2} - \frac{16}{25}y_{n+1} + \frac{3}{25}y_n &= \frac{12}{25}hf(t_{n+4}, y_{n+4}); \\ s=5: y_{n+5} - \frac{300}{137}y_{n+4} + \frac{300}{137}y_{n+3} - \frac{200}{137}y_{n+2} + \frac{75}{137}y_{n+1} - \frac{12}{137}y_n &= \frac{60}{137}hf(t_{n+5}, y_{n+5}). \end{aligned}$$

为了简便,定义矩阵  $\mathbf{A}, \mathbf{B} \in \mathbf{R}^{(N-s+1) \times (N+1)}$  为:

$$\mathbf{A} = \begin{bmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_s \\ & \alpha_0 & \alpha_1 & \cdots & \alpha_s \\ & & \ddots & \ddots & \ddots & \ddots \\ & & & \alpha_0 & \alpha_1 & \cdots & \alpha_s \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \beta_0 & \beta_1 & \cdots & \beta_s \\ & \beta_0 & \beta_1 & \cdots & \beta_s \\ & & \ddots & \ddots & \ddots & \ddots \\ & & & \beta_0 & \beta_1 & \cdots & \beta_s \end{bmatrix}.$$

基于此,式(6)的矩阵形式可以表示为:

$$\mathbf{AX} = h\mathbf{BF}(\mathbf{X}) + O(h^{p+1}), \quad (7)$$

其中:  $O(h^{p+1})$  表示  $p$  阶的线性多步方法离散动力系统的局部截断误差,它仅依赖于时间步长  $h$ 。

在应用线性多步方法离散动力系统(1)以后,结合式(3)可以得出  $\dot{\mathbf{X}} = \mathbf{F}(\mathbf{X}) = \Phi(\mathbf{X})\mathbf{C}^*$ ,再与式(7)相结合,有:

$$\mathbf{AX} = h\mathbf{B}\Phi(\mathbf{X})\mathbf{C}^* + O(h^{p+1}). \quad (8)$$

注意,当精确的状态数据不可用,只有噪声轨迹的状态数据可用时,式(8)成为:

$$\mathbf{AY} = h\mathbf{B}\Phi(\mathbf{Y})\mathbf{C} + O(h^{p+1}).$$

相应地,动力系统的识别问题被转化为求解如下线性回归问题:

$$\mathbf{AY} \approx h\mathbf{B}\Phi(\mathbf{Y})\mathbf{C}. \quad (9)$$

由下一节的数值结果可知,在精确状态数据的情况下,应用线性多步方法离散动力系统,可以在很大程度上提高子空间追踪算法的精度。甚至在时间步长较大的情况下,也能较准确地从数据中发现潜在动力模型。但是当面对的状态数据是含噪声的观测数据时,仅仅应用线性多步方法离散动力系统,再使用子空间追踪算法进行稀疏识别,很难得到令人满意的识别结果。在以往的研究中,人们尝试了各种各样的办法对观测数据进行消噪,例如移动最小二乘消噪<sup>[15]</sup>,但是该方法需要确定一个合适的距离参数才能展现出较好的消噪效果。因此,下文通过引入广义最小二乘方法,利用噪声残差项的近似协方差矩阵对求解稀疏系数矩阵的最小化问题进行加权,以此来增强算法对噪声的鲁棒性。

### 2.3 广义最小二乘方法

引入广义最小二乘方法<sup>[9]</sup>,利用噪声残差项的近似协方差矩阵对问题(9)进行加权,将之转化为求解如下加权最小化问题:

$$\min_{\mathbf{C}} \|[h\mathbf{B}\Phi(\mathbf{Y})\mathbf{C} - \mathbf{AY}]^T \boldsymbol{\Sigma}^{-1} [h\mathbf{B}\Phi(\mathbf{Y})\mathbf{C} - \mathbf{AY}]\|_2^2 + \lambda \|\mathbf{C}\|_0, \quad (10)$$

其中: $\boldsymbol{\Sigma}$ 表示广义最小二乘方法中噪声残差项的近似协方差矩阵,目的是使所得估计为最佳线性无偏估计。受文献[3]的启发,可以选择近似协方差矩阵 $\boldsymbol{\Sigma} = (h\mathbf{B} - \mathbf{A})(h\mathbf{B} - \mathbf{A})^T$ 。

这类近似协方差矩阵在文献[3]中被证明对噪声具有很好的鲁棒性,甚至在噪声较大的情况下,也能够从观测数据中得到高精度的模型识别结果。并且,Messenger 等人<sup>[9]</sup>也证实了广义最小二乘方法在提高算法精度上的有效性。

**注 4** Chen<sup>[3]</sup>提出的 LMM-SINDy 算法结合了线性多步方法和 STLS 算法,在该算法中  $l_1$  范数  $\|\mathbf{C}\|_1$  被用于描述矩阵  $\mathbf{C}$  的稀疏性。下文提出的 LMM-SP 算法结合了线性多步方法和子空间追踪算法,在该算法中  $l_0$  范数  $\|\mathbf{C}\|_0$  被用于描述矩阵  $\mathbf{C}$  的稀疏性。

### 2.4 LMM-SP 算法

根据上面的分析,LMM-SP 算法的具体步骤如下。

步骤 1:输入观测数据矩阵  $\mathbf{Y}$ 、特征矩阵  $\Phi(\mathbf{Y})$ 、近似协方差矩阵  $\boldsymbol{\Sigma}$  以及线性多步方法对应的参数矩阵  $\mathbf{A}$  和  $\mathbf{B}$ 。

步骤 2:现在需要求解最小化问题(10),通过简单分析,可以将它转化为求解如下的线性方程组,即寻找矩阵  $\mathbf{C}$  的稀疏解: $\Phi(\mathbf{Y})^T \mathbf{B}^T \boldsymbol{\Sigma}^{-1} h\mathbf{B}\Phi(\mathbf{Y})\mathbf{C} = \Phi(\mathbf{Y})^T \mathbf{B}^T \boldsymbol{\Sigma}^{-1} \mathbf{AY}$ ,改写为分量形式:

$$\Phi(\mathbf{Y})^T \mathbf{B}^T \boldsymbol{\Sigma}^{-1} h\mathbf{B}\Phi(\mathbf{Y})\mathbf{c}_i = \Phi(\mathbf{Y})^T \mathbf{B}^T \boldsymbol{\Sigma}^{-1} \mathbf{A}y_i, i = 1, \dots, d.$$

步骤 3:对近似协方差矩阵  $\boldsymbol{\Sigma}$  做 Cholesky 分解,使得  $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T$ ,其中  $\mathbf{L}$  是上三角矩阵。然后步骤 2 中方程组的分量形式可以写为:

$$\Phi(\mathbf{Y})^T \mathbf{B}^T (\mathbf{L}\mathbf{L}^T)^{-1} h\mathbf{B}\Phi(\mathbf{Y})\mathbf{c}_i = \Phi(\mathbf{Y})^T \mathbf{B}^T (\mathbf{L}\mathbf{L}^T)^{-1} \mathbf{A}y_i, i = 1, \dots, d,$$

进一步可写为:

$$\Phi(\mathbf{Y})^T \mathbf{B}^T (\mathbf{L}^T)^{-1} \mathbf{L}^{-1} h\mathbf{B}\Phi(\mathbf{Y})\mathbf{c}_i = \Phi(\mathbf{Y})^T \mathbf{B}^T (\mathbf{L}^T)^{-1} \mathbf{L}^{-1} \mathbf{A}y_i, i = 1, \dots, d,$$

继续可得:

$$\mathbf{L}^{-1} h\mathbf{B}\Phi(\mathbf{Y})\mathbf{c}_i = \mathbf{L}^{-1} \mathbf{A}y_i, i = 1, \dots, d. \quad (10)$$

步骤 4:下面将基于子空间追踪算法求解问题(10),为便于描述,记  $\Psi(\mathbf{Y}) = \mathbf{L}^{-1} h\mathbf{B}\Phi(\mathbf{Y})$ ,  $\mathbf{b} = \mathbf{L}^{-1} \mathbf{A}y_i, i = 1, \dots, d$ ,并将  $\Psi(\mathbf{Y})$ ,  $\mathbf{b}$  和稀疏性  $K$  作为输入。

步骤 5:对每个稀疏水平  $K = 1, 2, \dots, q$ ,使用 LMM-SP 算法挑选出最佳(系数误差最小)的候选基函数集:

$$A_K = \text{supp}(K; \Psi(\mathbf{Y}), \mathbf{b}), \hat{\mathbf{c}}_{A_K}^{(K)} = \text{LMM-SP}([K]_q; [\Psi(\mathbf{Y})]_{A_K}, \mathbf{b}), \quad (11)$$

其中:supp 表示支撑集。

步骤 6:计算通过算法输出的特征系数的误差  $e$ ,并用  $\tilde{\mathbf{c}}$  表示精确系数:

$$e_{\tilde{\mathbf{c}}^{(K)}}(A_K, q) = \|\hat{\mathbf{c}}_{A_K}^{(K)} - \tilde{\mathbf{c}}\|_2^2, q_K = \underset{K=1, \dots, q}{\text{argmin}} e_{\tilde{\mathbf{c}}^{(K)}}(A_K, q).$$

步骤 7:从对应不同稀疏性的候选基函数集中,选择系数误差最小的一个特征集合作为下一次迭代的基函数

库,并用  $A_1$  表示该集合:

$$A_1 = \text{supp}(\text{LMM-SP}([K]_{q_1}; [\Psi(\mathbf{Y})]_{A_1}, \mathbf{b})), \hat{\mathbf{c}}_{A_1}^{(K)} = \text{LMM-SP}([K]_{q_1}; [\Psi(\mathbf{Y})]_{A_1}, \mathbf{b}).$$

其中:  $q_1$  表示集合  $A_1$  中的元素个数。

步骤 8:继续步骤 5 到步骤 6,根据更新的基函数库  $A_1$ ,对所有的稀疏水平  $K_1 = 1, 2, \dots, q_1$  进行下一次迭代,其中候选基函数集  $A_{K_1}$  同式(11):

$$e_{\hat{\mathbf{c}}_{A_{K_1}}^{(K_1)}}(A_{K_1}, q_1) = \|\hat{\mathbf{c}}_{A_{K_1}}^{(K_1)} - \tilde{\mathbf{c}}\|_2^2, q_{K_1} = \underset{K_1=1, \dots, q_1}{\text{argmin}} e_{\hat{\mathbf{c}}_{A_{K_1}}^{(K_1)}}(A_{K_1}, q_1).$$

步骤 9:继续步骤 7,可得:

$$A_2 = \text{supp}(\text{LMM-SP}([K_1]_{q_2}; [\Psi(\mathbf{Y})]_{A_2}, \mathbf{b})), \hat{\mathbf{c}}_{A_2}^{(K)} = \text{LMM-SP}([K_1]_{q_2}; [\Psi(\mathbf{Y})]_{A_2}, \mathbf{b}).$$

其中:  $q_2$  表示集合  $A_2$  中的元素个数。

步骤 10:如果支撑集满足  $A_2 = A_1$ ,那么算法将终止;否则,继续上面的迭代过程。注意,这个过程需要并行 LMM-SP 算法至少  $q$  次<sup>[15]</sup>。直到索引集满足  $A_i = A_{i+1}$  的时候,算法才能被终止,此时有  $\hat{\mathbf{c}}_{A_i}^{(K)} = 0$ 。

### 3 数值实验

本文使用 Matlab R2021a 内置的函数 ode45 求解动力系统,以此来生成精确的状态数据。并通过在精确的数据中加入均值为 0,方差为  $\sigma^2$  的噪声生成观测数据,且  $\sigma$  的计算公式为  $\sigma = \sigma_{\text{NR}} \frac{\|\mathbf{X}\|_F}{\sqrt{d(N+1)}}$ ,其中:  $\sigma_{\text{NR}}$  为噪声水平,  $\|\mathbf{X}\|_F$  表示矩阵  $\mathbf{X}$  的 Frobenius 范数。并且噪声与数据的比值近似等于  $\sigma$ ,即  $\frac{\|\boldsymbol{\varepsilon}\|_F}{\|\mathbf{X}\|_F} \approx \sigma$ <sup>[3]</sup>。

为了呈现该算法在模型识别上的精度,将识别系数的相对误差定义为  $e(\mathbf{C}) = \frac{\|\mathbf{C} - \mathbf{C}^*\|_F}{\|\mathbf{C}^*\|_F}$ ,其中:  $\mathbf{C}$  表示使用 LMM-SP 算法进行模型识别得到的动力系统的系数矩阵,  $\mathbf{C}^*$  表示先验动力系统的精确系数矩阵。

#### 3.1 洛伦兹方程

首先,考虑洛伦兹方程的识别问题,在此之前确定时间变量  $t$  的范围为  $[0, 10]$  和初始条件为  $\mathbf{x}_0 = [-8, 7, 27]^T$ ,具体的模型为:

$$\begin{cases} \dot{x}_1 = -10x_1 + 10x_2, \\ \dot{x}_2 = 28x_1 - x_2 - x_1x_3, \\ \dot{x}_3 = x_1x_2 - \frac{8}{3}x_3. \end{cases}$$

根据少量先验知识,可以选择特征库为  $\Phi(\mathbf{X}) = [1, \mathbf{X}, \mathbf{X}^2, \mathbf{X}^3]$ ,其中  $\mathbf{X}^2$  矩阵可以表示为:

$$\mathbf{X}^2 = \begin{bmatrix} x_1^2(t_0) & x_1x_2(t_0) & \cdots & x_2^2(t_0) & \cdots & x_3^2(t_0) \\ x_1^2(t_1) & x_1x_2(t_1) & \cdots & x_2^2(t_1) & \cdots & x_3^2(t_1) \\ \vdots & \vdots & & \vdots & & \vdots \\ x_1^2(t_N) & x_1x_2(t_N) & \cdots & x_2^2(t_N) & \cdots & x_3^2(t_N) \end{bmatrix}.$$

本文将 LMM-SP 算法按照不同的离散方式分为 3 大类:显式 Adams 子空间追踪算法(Adams-Bashforth subspace pursuit, AB- $s$ -SP)算法、隐式 Adams 子空间追踪(Adams-Moulton subspace pursuit, AM- $s$ -SP)算法和向后微分公式子空间追踪(backward differentiation formula subspace pursuit, BDF- $s$ -SP)算法,其中  $s$  表示线性多步方法的步数。

考虑在数据无噪声的情况下,选择时间步长为 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, 并计算在不同的时间步长下,使用 LMM-SP 算法进行模型识别的相对系数误差,结果见图 1。为方便对比,在图 1 中标注了不同离散方式的 LMM-SP 算法所对应的局部截断误差。

由图 1 可知,在数据无噪声的情况下,LMM-SP 算法中选用的线性多步方法的阶数越高,那么利用该算法进行模型识别的精度也越高。并且不难发现,在步数  $s$  相同的情况下,AM- $s$ -SP 算法比 AB- $s$ -SP 算法和 BDF- $s$ -SP 算法的精度都更高。

考虑在数据含噪声的情况下,基于时间步长 0.001,在精确数据中加入的噪声水平  $\sigma_{NR}$  分别为:  $10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$ , 计算不同噪声水平下应用 LMM-SP 算法从得到的观测数据中识别模型的相对系数误差(图 2)。由图 2 可以看出,在步数  $s$  相同的情况下,使用 AM- $s$ -SP 算法从观测数据中挖掘潜在模型,比使用其他 2 种算法具有更高的精度和更好的鲁棒性。即使当  $\sigma_{NR}=0.1$  的时候,AM- $s$ -SP 算法被证实仍能得出正确的模型框架和系数估计。

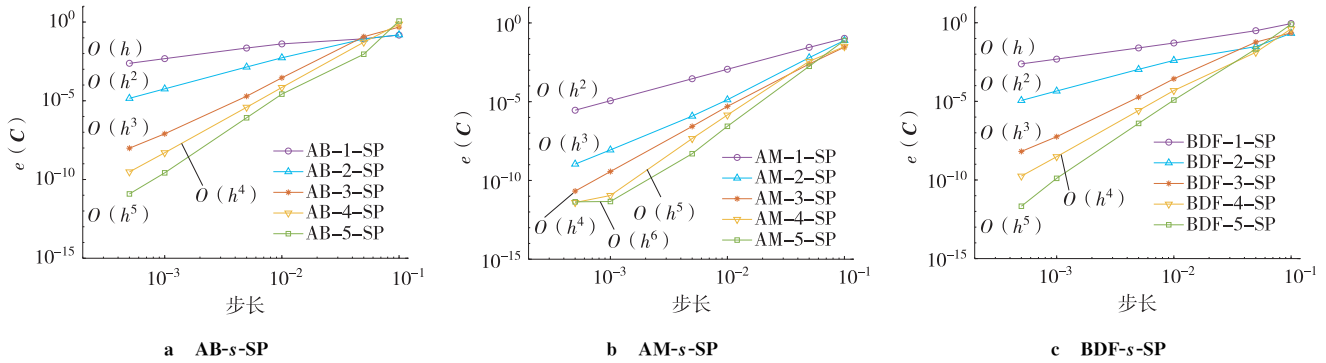


图 1 在数据无噪声的情况下,使用 LMM-SP 算法识别洛伦兹模型的相对系数误差

Fig. 1 The relative coefficient error of identifying Lorenz model using LMM-SP algorithm with noise-free data

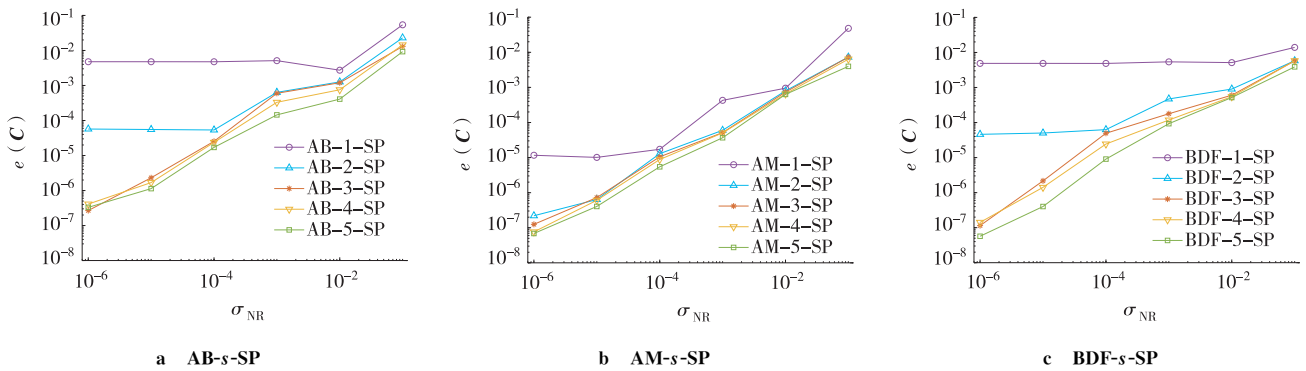


图 2 在数据含噪声的情况下,使用 LMM-SP 算法识别洛伦兹模型的相对系数误差

Fig. 2 The relative coefficient error of identifying Lorenz model using LMM-SP algorithm with noise data

图 3 是应用 AM-5-SP 算法对洛伦兹模型进行识别的预测结果和相图。从图中不难发现,使用 AM-5-SP 算法识别模型得到的预测结果与精确轨迹之间有着很好的一致性,但是由于洛伦兹模型自身的属性,导致学习到的轨迹与正确的轨迹之间还是有一些轻微的振动。

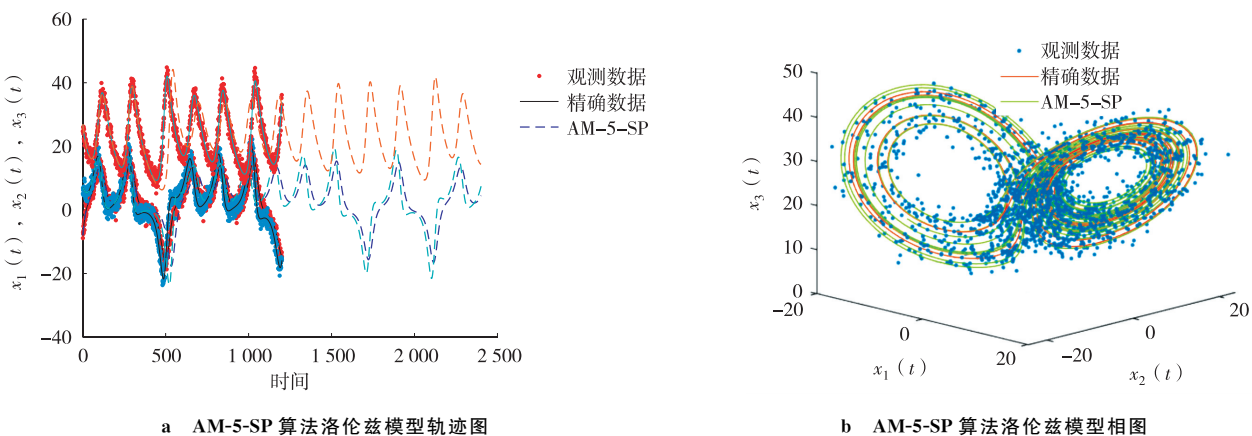


图 3 当  $\sigma_{NR}=0.1$  时,使用 AM-5-SP 算法识别洛伦兹模型的结果

Fig. 3 The results of identifying Lorenz model using AM-5-SP algorithm with  $\sigma_{NR}=0.1$

### 3.2 三次阻尼方程

接下来,考虑三次阻尼方程的识别问题,在这里确定时间变量  $t$  的范围为  $[0,10]$ 、初始条件为  $\mathbf{x}_0 = [-1,2]^T$  以及具体的模型为:

$$\begin{cases} \dot{x}_1 = -x_1^3 + 3x_2^3, \\ \dot{x}_2 = -2x_1^3 + x_2^3. \end{cases}$$

根据少量先验知识,可以选择特征库为  $\Phi(\mathbf{X}) = [1, \mathbf{X}, \mathbf{X}^2, \mathbf{X}^3, \mathbf{X}^4]$ , 其中  $\mathbf{X}^3$  矩阵可以表示为:

$$\mathbf{X}^3 = \begin{bmatrix} x_1^3(t_0) & x_1^2 x_2(t_0) & \cdots & x_2^3(t_0) \\ x_1^3(t_1) & x_1^2 x_2(t_1) & \cdots & x_2^3(t_1) \\ \vdots & \vdots & & \vdots \\ x_1^3(t_N) & x_1^2 x_2(t_N) & \cdots & x_2^3(t_N) \end{bmatrix}.$$

考虑在数据无噪声的情况下,选择与第 1 个例子中相同的时间步长设置。使用 LMM-SP 算法从数据中恢复隐藏的动力系统,并在图 4 中绘制了对应不同时间步长下的模型识别结果。

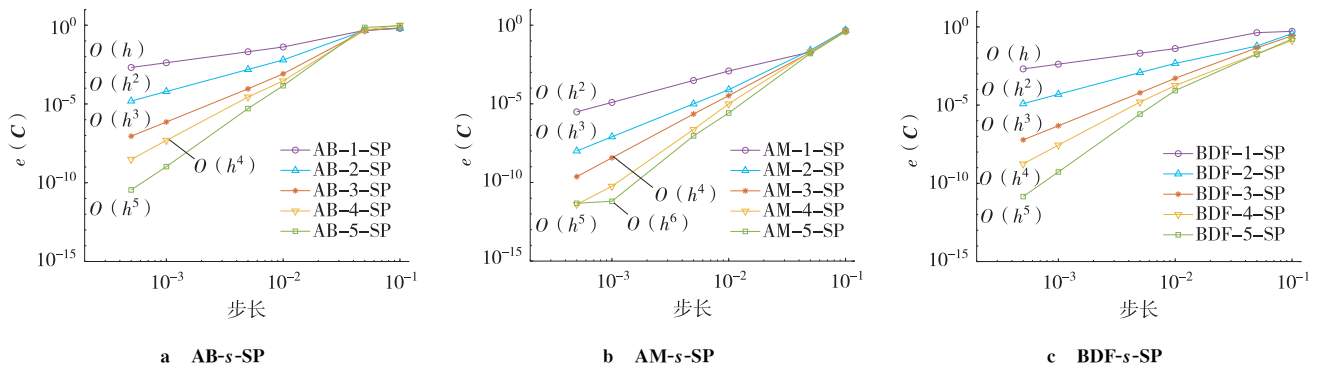


图 4 在数据无噪声的情况下,使用 LMM-SP 算法识别三次阻尼模型的相对系数误差

Fig. 4 The relative coefficient error of identifying cubic damped model using LMM-SP algorithm with noise-free data

由图 4 可以看出,在数据无噪声的情况下,应用同一类的 LMM-SP 算法从数据中识别潜在模型,当用于时间离散的线性多步方法的阶数越高,对应 LMM-SP 算法用于模型识别的精度也越高。而比较不同类的 LMM-SP 算法在模型识别上的性能可以发现,在步数  $s$  相同的情况下,AM- $s$ -SP 算法比 AB- $s$ -SP 算法和 BDF- $s$ -SP 算法的精度都更高。

考虑在数据含噪声的情况下,应用 LMM-SP 算法在不同的噪声水平下识别三次阻尼模型,结果见图 5。并在图 6 中绘制了使用 AM-5-SP 算法对三次阻尼模型进行识别的结果。由图 6 可以看出,即使当  $\sigma_{NR} = 0.1$  时,通过 LMM-SP 算法得到的预测结果与精确轨迹之间依然达到了高度的重合。

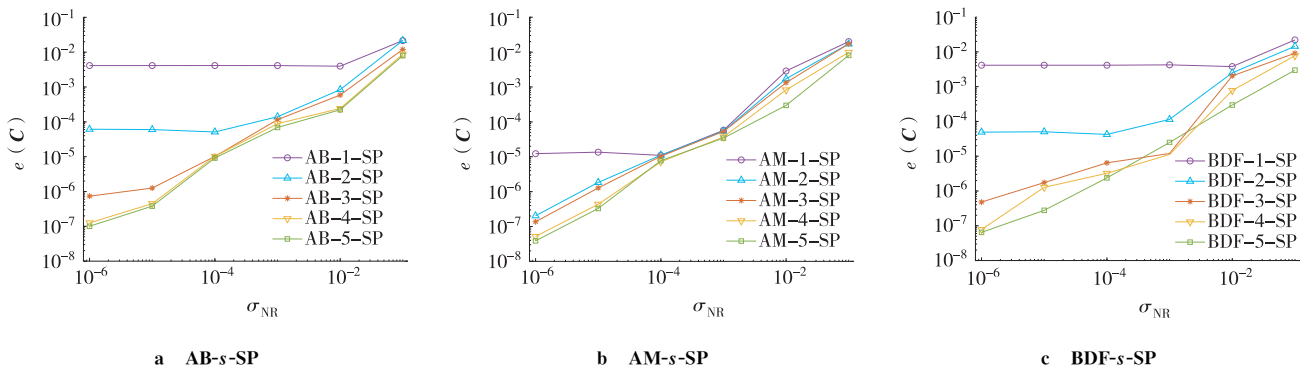
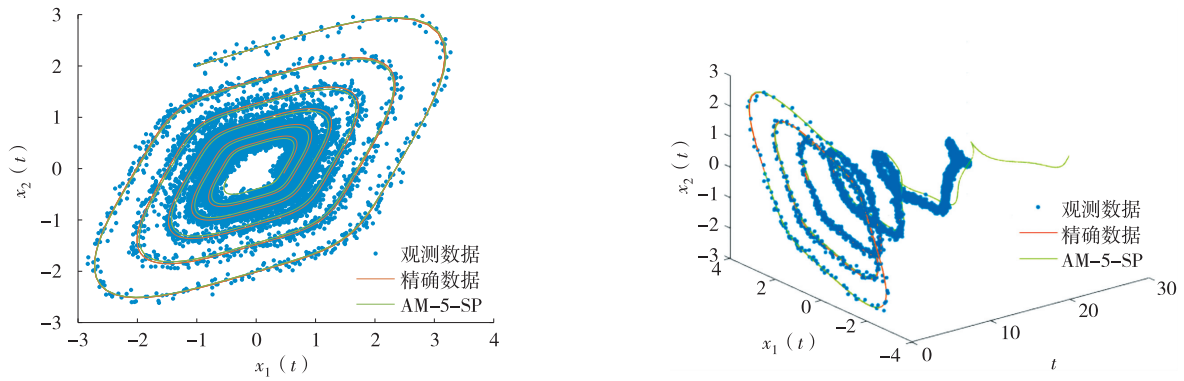


图 5 在数据含噪声的情况下,使用 LMM-SP 算法识别三次阻尼模型的相对系数误差

Fig. 5 The relative coefficient error of identifying cubic damped model using LMM-SP algorithm with noise data



a AM-5-SP 算法三次阻尼模型轨迹图

b AM-5-SP 算法三次阻尼模型相图

图 6 当  $\sigma_{NR} = 0.1$  时,使用 AM-5-SP 算法识别三次阻尼模型的结果

Fig. 6 The results of identifying cubic damped model using AM-5-SP algorithm with  $\sigma_{NR} = 0.1$

### 3.3 阻尼摆方程

下面考虑阻尼摆方程的识别问题。同样地,可以确定时间变量  $t$  的范围为  $[0, 10]$ 、初始条件为  $\mathbf{x}_0 = [-1.193, -3.876]^T$  以及具体的模型为:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = -0.2x_2 - 8.91\sin(x_1). \end{cases}$$

注意,在这个方程中右端函数包含了三角函数  $\sin(x_1)$ ,基于这个先验知识,可以选择特征库为:

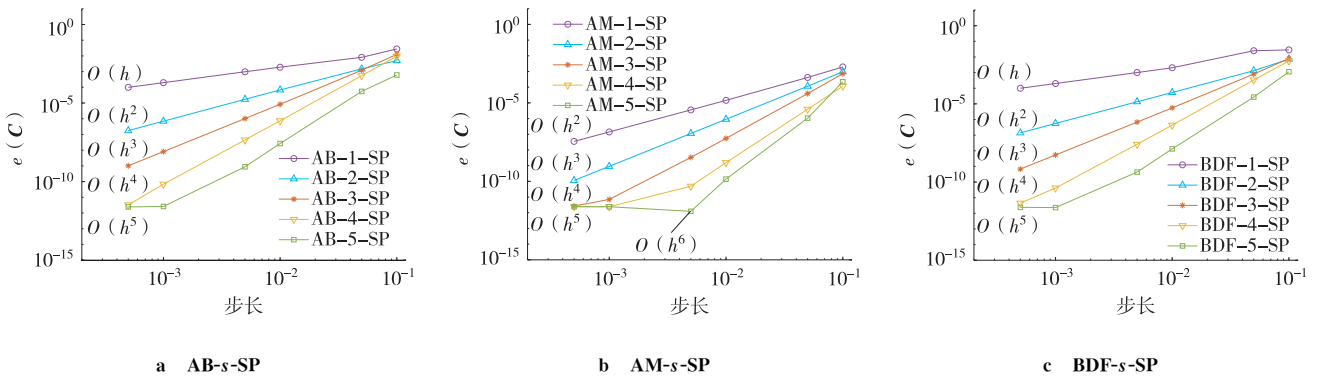
$$\Phi(\mathbf{X}) = [1, \mathbf{X}, \mathbf{X}^2, \sin(\mathbf{X}), \cos(\mathbf{X})],$$

其中:

$$\sin(\mathbf{X}) = \begin{bmatrix} \sin(x_1(t_0)) & \sin(x_2(t_0)) \\ \sin(x_1(t_1)) & \sin(x_2(t_1)) \\ \vdots & \vdots \\ \sin(x_1(t_N)) & \sin(x_2(t_N)) \end{bmatrix},$$

$$\cos(\mathbf{X}) = \begin{bmatrix} \cos(x_1(t_0)) & \cos(x_2(t_0)) \\ \cos(x_1(t_1)) & \cos(x_2(t_1)) \\ \vdots & \vdots \\ \cos(x_1(t_N)) & \cos(x_2(t_N)) \end{bmatrix}.$$

考虑在数据无噪声的情况下,比较 AB- $s$ -SP 算法、AM- $s$ -SP 算法和 BDF- $s$ -SP 算法从数据中重构模型的性能,结果见图 7。



a AB- $s$ -SP

b AM- $s$ -SP

c BDF- $s$ -SP

图 7 在数据无噪声的情况下,使用 LMM-SP 算法识别阻尼摆模型的相对系数误差

Fig. 7 The relative coefficient error of identifying damped pendulum model using LMM-SP algorithm with noise-free data

由图 7 可以看出,在数据无噪声的情况下,用于时间离散的线性多步方法的阶数越高,所对应的 LMM-SP 算法在模型识别的过程中表现越好。同样地,通过比较相同步数  $s$  下 AB- $s$ -SP 算法,AM- $s$ -SP 算法和 BDF- $s$ -SP 算法在模型识别上的相对系数误差,可以发现使用 AM- $s$ -SP 算法重构潜在模型比使用其他 2 种算法的精度都更高,这与线性多步方法的阶是密切相关的。一般来说,在步数  $s$  相同的情况下,算法本身的阶数越高,往往意味着它们识别系数的精度也越高。

考虑在数据含噪声的情况下,应用 LMM-SP 算法识别阻尼摆模型,结果见图 8。图 9 是使用 AM-5-SP 算法对阻尼摆模型进行识别的结果。

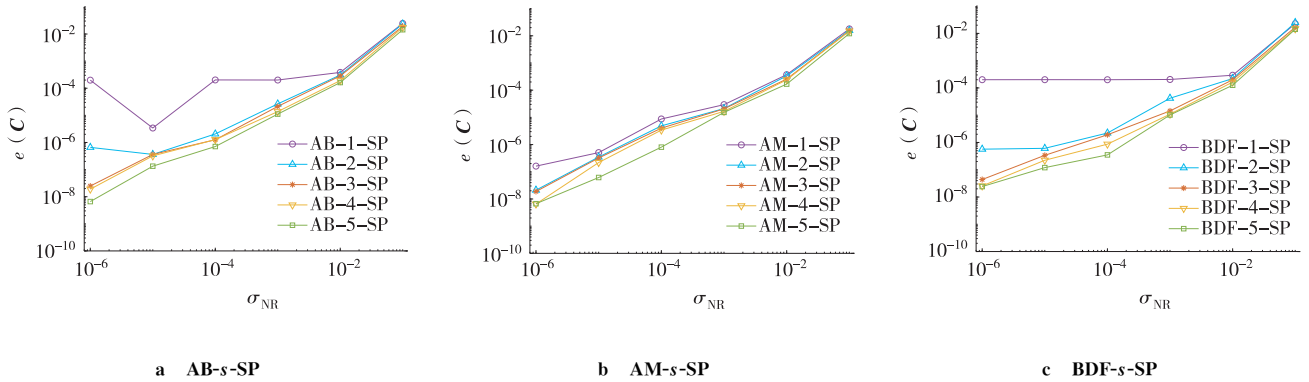


图 8 在数据含噪声的情况下,使用 LMM-SP 算法识别阻尼摆模型的相对系数误差

Fig. 8 The relative coefficient error of identifying damped pendulum model using LMM-SP algorithm with noise data

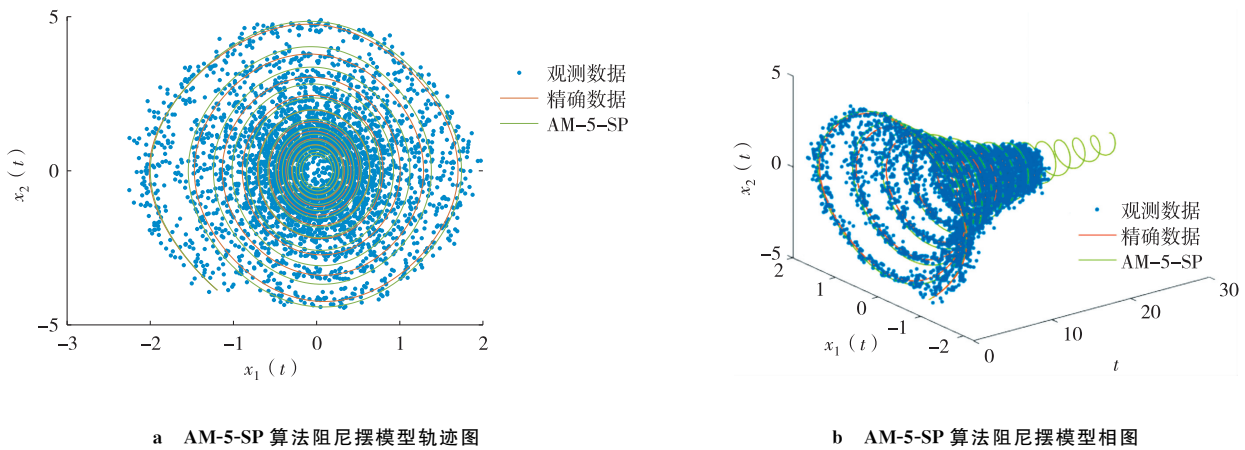


图 9 当  $\sigma_{NR} = 0.1$  时,使用 LMM-SP 算法识别阻尼摆模型的结果

Fig. 9 The results of identifying damped pendulum model using LMM-SP algorithm with  $\sigma_{NR} = 0.1$

### 3.4 对比实验

为了更好地展现本文提出的 LMM-SP 算法在模型识别上的性能,现在将该算法与 MLS-SP 算法<sup>[15]</sup>进行对比,其中 MLS-SP 算法应用欧拉方法离散动力系统,并应用移动最小二乘(moving least squares, MLS)方法对观测数据进行消噪,最后利用子空间追踪算法进行特征选择和系数估计,在这个实验中,选择的时间范围为  $[0, 10]$ ,时间步长为 0.001,具体的实验结果如表 1 所示。

由表 1 可以看出,本文提出的 LMM-SP 算法在识别非线性动力系统上具有比 MLS-SP 算法更高的精度。得到这样结果的原因主要有 2 个:1) 应用线性多步方法离散动力系统,比应用欧拉方法离散动力系统精度更高;2) 利用广义最小二乘加权的方式减少噪声,比利用 MLS 方法减少噪声具有更好的效果。

表 1 对比 MLS-SP 算法和 AM-5-SP 算法  
 Tab. 1 Comparison of MLS-SP algorithm and AM-5-SP algorithm

动力方程	算法	$\sigma_{NR}$					
		$10^{-6}$	$10^{-5}$	$10^{-4}$	$10^{-3}$	$10^{-2}$	$10^{-1}$
洛伦兹方程	MLS-SP	7.39E-03	7.39E-03	7.37E-03	7.67E-03	8.31E-03	3.65E-02
	AM-5-SP	6.88E-08	4.04E-07	5.46E-06	3.67E-05	6.32E-04	3.98E-03
三次阻尼方程	MLS-SP	9.11E-02	9.11E-02	9.11E-02	9.10E-02	9.10E-02	1.02E-01
	AM-5-SP	3.29E-07	3.29E-07	7.74E-06	3.42E-05	2.96E-04	8.04E-03
阻尼摆方程	MLS-SP	1.10E-04	1.10E-04	1.09E-04	2.57E-04	1.36E-02	1.30E-02
	AM-5-SP	6.69E-09	6.16E-08	8.05E-07	1.53E-05	1.68E-04	1.21E-02

## 4 结论

本文介绍了一种基于稀疏回归的模型识别算法:LMM-SP 算法,其中包括 AB- $s$ -SP 算法、AM- $s$ -SP 算法和 BDF- $s$ -SP 算法。该算法结合了子空间追踪算法、经典的线性多步方法和广义最小二乘方法的原理。利用该算法从观测数据中识别未知模型,需要在每一次迭代过程中,从基函数库里挑选出系数误差最小的特征集合作为下一次迭代的基函数库。并在迭代终止以后,根据输出的最小基函数集,使用最小二乘方法计算保留下来的特征的对对应系数值。最后,通过大量的数值实验验证了所提出的 LMM-SP 算法的有效性。

### 参考文献:

- [1] BRUNTON S L, KUTZ J N. Data-driven science and engineering[M]. Cambridge, UK: Cambridge University Press, 2019.
- [2] SCHMIDT M, LIPSON H. Distilling free-form natural laws from experimental data[J]. Science, 2009, 324(5923): 81.
- [3] CHEN H. Data-driven sparse identification of nonlinear dynamical systems using linear multistep methods[J]. Calcolo, 2023, 60(1): 11.
- [4] TIBSHIRANI R. Regression shrinkage and selection via the lasso[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1996, 58(1): 267-288.
- [5] BRUNTON S L, PROCTOR J L, KUTZ J N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems[J]. PNAS, 2016, 113(15): 3932-3937.
- [6] CORTIELLA A, PARK K C, DOOSTAN A. Sparse identification of nonlinear dynamical systems via reweighted  $\ell_1$ -regularized least squares[J]. Computer Methods in Applied Mechanics and Engineering, 2021, 376: 113620.
- [7] 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[J]. Proceedings Mathematical, Physical, and Engineering Sciences, 2022, 478(2260): 20210904.
- [8] MANGAN N M, KUTZ J N, BRUNTON S L, et al. Model selection for dynamical systems via sparse regression and information criteria[J]. Proceedings Mathematical, Physical, and Engineering Sciences, 2017, 473(2204): 20170009.
- [9] MESSENGER D A, BORTZ D M. Weak SINDy: Galerkin-based data-driven model selection [J]. Multiscale Modeling & Simulation, 2021, 19(3): 1474-1497.
- [10] MESSENGER D A, BORTZ D M. Weak SINDy for partial differential equations[J]. Journal of Computational Physics, 2021, 443: 110525.
- [11] GOYAL P, BENNER P. Discovery of nonlinear dynamical systems using a Runge-Kutta inspired dictionary-based sparse regression approach[J]. Proceedings Mathematical, Physical, and Engineering Sciences, 2022, 478(2262): 20210883.
- [12] DU Q, GU Y Q, YANG H Z, et al. The discovery of dynamics via linear multistep methods and deep learning: error estimation [J]. SIAM Journal on Numerical Analysis, 2022, 60(4): 2014-2045.
- [13] KELLER R T, DU Q. Discovery of dynamics using linear multistep methods[J]. SIAM Journal on Numerical Analysis, 2021, 59(1): 429-455.
- [14] LIU L F, DU X P, CHENG L Z. Stable signal recovery via randomly enhanced adaptive subspace pursuit method[J]. IEEE

Signal Processing Letters, 2013, 20(8):823-826.

- [15] HE Y C, KANG S H, LIAO W J, et al. Robust identification of differential equations by numerical techniques from a single set of noisy observation[J]. SIAM Journal on Scientific Computing, 2022, 44(3):1145-1175.

## Operations Research and Cybernetics

### Model Identification via Subspace Pursuit and Linear Multistep Method

JIANG Yuemei, CHEN Hao

(College of Mathematical Science, Chongqing Normal University, Chongqing 401331, China)

**Abstract:** It seeks a highly accurate sparse identification method for the nonlinear dynamical systems, combining with traditional numerical analysis techniques. Firstly, a suitable basis function library must be created in order to approximate the potential nonlinear dynamical systems. Then, the approximated nonlinear dynamical systems are discretized using the linear multistep method. Next, when the state data contains noise, the generalised least squares method is used to calculate the approximate covariance matrix of the noise residual term and use this matrix to weight the optimisation problem obtained from the preceding process, thereby reducing the influence of noise on the model identification results. Finally, the subspace pursuit algorithm selects the set of features with the smallest coefficient error from the data to serve as the basis function library for the next iteration, and after the iteration is completed, the coefficient values of the retained features are computed using the least squares method. The proposed linear multistep subspace pursuit methods for identifying nonlinear dynamic systems possess high accuracy and robustness. Numerical results are presented to demonstrate the effectiveness of the proposed methods.

**Keywords:** model identification; subspace pursuit; sparse regression; linear multistep method; generalized least squares method

(责任编辑 黄 颖)