

基于线性多步方法的二阶动力系统的模型识别

江月梅, 陈浩

重庆师范大学 数学科学学院, 重庆 401331

摘要:目的 针对二阶动力系统的识别问题, 提出一种基于线性多步方法的稀疏识别方法。方法 首先, 构造一个包含几乎所有可能基函数的基函数库, 用于近似潜在的二阶动力系统; 然后, 利用线性多步方法离散近似后的二阶动力系统; 接着, 在广义最小二乘原理的指导下, 选取一个合适的噪声残差项近似协方差矩阵, 再利用该矩阵对上述过程得到的最小化问题进行加权, 从而降低噪声对模型识别过程的影响; 最后, 使用稀疏回归算法从基函数库中挑选出最有意义的最小特征项, 并通过稀疏迭代求解其对应系数。结果 比较了不同时间步长和不同噪声水平下, 使用提出的线性多步稀疏识别方法挖掘潜在二阶动力系统的数值实验, 通过这些实验可以得出: 所提出的方法用于识别未知的二阶动力系统具有较高的精度和较好的鲁棒性。结论 通过数值实验, 验证了算法的有效性。

关键词: 模型识别; 线性多步法; 稀疏回归; 广义最小二乘法

中图分类号: O241 文献标识码: A doi: 10.16055/j.issn.1672-058X.2024.0006.010

Model Identification of Second-order Dynamical Systems Based on Linear Multistep Methods

JIANG Yuemei, CHEN Hao

School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China

Abstract: Objective A sparse identification method based on linear multistep methods was proposed for the identification of second-order dynamical systems. **Methods** Firstly, a basis function library containing nearly all possible basis functions was constructed to approximate the right-hand function of the potential second-order dynamical system. Then, the second-order dynamical system was discretized using linear multistep methods. Next, under the guidance of the generalized least squares principle, a suitable approximate covariance matrix of the noise residual term was selected and this matrix was used to weigh the minimization issues obtained by the above process to reduce the influence of noise on the model identification process. Finally, a sparse regression algorithm was used to select the most meaningful minimal feature terms from the basis function library and solve their corresponding coefficients through sparse iteration. **Results** Numerical experiments were done to explore potential second-order dynamical systems using the linear multistep sparse identification method with different time steps and noise levels. These experiments showed that the proposed method had higher accuracy and better robustness in identifying unknown second-order dynamical systems. **Conclusion** The effectiveness of the algorithm is verified by numerical experiments.

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

1 引言

近年来, 由于自然或人为活动的增加, 越来越多的数据涌现在现实生活中, 并且随着科学技术的发展和

生产生活的进步, 人们能够从海量数据中挖掘出潜在的动力模型, 并利用所学到的动力模型进行预测、控制和决策。在识别潜在模型的过程中, 由于需要处理的

收稿日期: 2023-06-12 修回日期: 2023-07-15 文章编号: 1672-058X(2024)06-0080-07

基金项目: 重庆市自然科学基金(CSTC2021JCYJ-MSXMX0034)。

作者简介: 江月梅(1999—), 女, 重庆丰都人, 硕士研究生, 从事微分方程数值解研究。

通讯作者: 陈浩(1986—), 男, 湖北潜江人, 教授, 博士, 从事微分方程数值解研究。Email: hch@cqu.edu.cn。

引用格式: 江月梅, 陈浩. 基于线性多步方法的二阶动力系统的模型识别[J]. 重庆工商大学学报(自然科学版), 2024, 41(6): 80-86.

JIANG Yuemei, CHEN Hao. Model identification of second-order dynamical systems based on linear multistep methods[J].

Journal of Chongqing Technology and Business University (Natural Science Edition), 2024, 41(6): 80-86.

数据十分庞大,且可能的模型形式也非常多,这使得模型识别问题成为一个具有挑战性的难题。特别地,对于高维高阶微分方程,要想找到其正确的模型框架和对应系数并不容易,因此研究者们做了大量的工作来解决这个问题。

Brunton 等^[1]结合稀疏回归算法和机器学习技术,提出一种从观测数据中发现动力系统的稀疏识别方法,该方法假设构建的基函数库中只有极少数特征对动力系统是有意义的,并使用稀疏回归算法从中挑选出能够描述状态数据的特征最小项;Cortiella 等^[2]在 SINDy 算法的基础上,提出一种迭代稀疏正则化回归算法,该算法提高了一般 SINDy 算法的精度和鲁棒性,降低了噪声对模型识别的影响;Mangan 等^[3]结合 SINDy 算法开发了一种新型的模型选择算法,该算法有效地考虑了动力系统的大量候选模型;除此之外, Messenger 等^[4]在 SINDy 算法的基础上,通过引入数值积分,利用其自然降噪的作用,提出一种弱离散形式的弱 SINDy 算法,该算法甚至能够从含较大噪声的数据里精准地识别出潜在的动力系统;Tibshirani 等^[5]提出一种基于 LASSO 算法的模型识别方法,该方法能够得到可解释的动力模型;Goyal 等^[6]将龙格库塔方法与 SINDy 算法相结合得到了一种精度更高的 SINDy 衍化算法。稀疏回归算法及其衍化算法在识别未知模型上具有以下的一般规律和优缺点:通常与线性多步方法或龙格库塔方法结合使用;需要对可能含噪声的状态数据进行消噪处理;需要先构建一个包含几乎所有可能基函数的确定的基函数库,且该基函数库中只有极少的特征对动力系统有意义;稀疏回归方法的本质是求解一个最小化问题,该问题通常可以用 LASSO 算法和 STLS 算法求解,能够给出具体的模型形式,可以进一步做理论分析。

另外,随着人工智能技术的快速发展,近来也有很多人将它们应用于从状态数据中挖掘未知的动力模型。例如,Harlim 等^[7]使用机器学习技术从状态数据中发现隐藏的动力系统,并通过数值实验证实了它是一种非常有效的模型识别方法,即使需要重构的模型中包含高维函数,也能被准确地识别出来;Qin 等^[8]也使用深度神经网络从观测数据中发现了潜在的动力系统;Du 等^[9]将经典的线性多步方法与深度网络方法相结合,大大提高了使用传统深度网络方法识别动力系统的精度。神经网络算法及其衍化算法用于识别未知模型具有以下的一般规律和优缺点:不需要构造确定的基函数库,使用神经网络中的随机特征基函数库代替;能够重构高维复杂的动力系统;一般使用神经网络算法进行模型识别,由于算法需要训练,所以需要的时

间较久;很难给出具体的模型形式,不容易做进一步的理论分析。

除了上述提到的方法,一些基于高斯过程^[10-12]、贝叶斯推断^[13-14]和统计学习^[15]的模型识别方法也备受研究者的青睐,并在过去的研究中取得了较大的成功。

本文的目的是从可能含噪声的观测数据中识别出潜在的二阶动力模型,由以往的研究结果可知,动力方程中数值时间导数的阶数越高,对数据中噪声的放大程度也越大。结合上述分析,本文提出一种基于线性多步方法和稀疏回归算法的模型识别方法,即线性多步稀疏识别方法。该方法首先应用线性多步方法离散近似后的二阶动力系统,然后使用稀疏回归算法通过稀疏迭代运算从基函数库中找出最有意义的最少特征项,并求解其对应系数。为降低噪声对识别结果的影响,在广义最小二乘原理的指导下,计算噪声残差项的近似协方差矩阵^[16],并利用该矩阵对时间离散后得到的最小化问题进行加权,从而降低噪声对模型识别结果的影响。综上所述,提出的方法具有以下创新点:使用线性多步方法离散二阶动力系统,提高了算法的精度;使用广义最小二乘方法对数据进行消噪,降低了噪声对模型识别结果的影响;将线性多步方法和稀疏回归算法相结合,通过数值实验证实了线性多步稀疏识别算法在识别模型上的优越性。

2 理论基础

考虑如下形式的二阶动力系统:

$$\begin{cases} \ddot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) \\ \dot{\mathbf{x}}(0) = \dot{\mathbf{x}}_0, \mathbf{x}(0) = \mathbf{x}_0 \end{cases} \quad (1)$$

其中, \mathbf{x} 表示状态变量向量,维数为 d ; \mathbf{f} 是依赖于 \mathbf{x} 的右端函数向量,在这里 \mathbf{f} 是未知的; $\mathbf{x}_0 = [x_1(0), x_2(0), \dots, x_d(0)]^T$ 和 $\dot{\mathbf{x}}_0 = [\dot{x}_1(0), \dot{x}_2(0), \dots, \dot{x}_d(0)]^T$ 表示二阶动力系统式(1)的初始条件;设时间 t 的范围为 $0 \leq t \leq T$,收集到的状态数据个数为 N ,则时间步长可以表示为 $h = T/N, t_n = nh, n = 0, 1, \dots, N$ 。状态数据矩阵 \mathbf{X} 可以表示为

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

及右端函数矩阵 $\mathbf{F}(\mathbf{X})$ 可以表示为

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

现在,需要从状态数据 \mathbf{x} 中恢复出未知的模型 f 。由过去的研究可知,首先需要构造一个涵盖几乎所有可能特征的基函数库 $\varphi_j(\mathbf{x})$,并假设右端函数 $f_i(\mathbf{x})$ 是基函数库 $\varphi_j(\mathbf{x})$ 中元素的一个线性组合:

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

其中, q 表示基函数库中元素的个数, c_{ji}^* 是对应基函数 $\varphi_j(\mathbf{x})$ 的系数,则式(2)的矩阵形式为

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

其中,特征矩阵 $\Phi(\mathbf{X})$ 可以表示为

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

由先前的研究经验和专业知识可知,基函数库一般可选取下面这样的矩阵:

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

其中, \mathbf{X}^k 表示状态数据矩阵 \mathbf{X} 的 k 次幂, $\sin(\mathbf{X})$ 表示对矩阵 \mathbf{X} 中的所有元素作用一个三角函数 \sin 。

同样地,二阶数值时间导数状态数据矩阵 $\bar{\mathbf{X}}$ 和系数矩阵 \mathbf{C}^* 可以表示为

$$\bar{\mathbf{X}} = \begin{bmatrix} \bar{x}_1(t_0) & \bar{x}_2(t_0) & \cdots & \bar{x}_d(t_0) \\ \bar{x}_1(t_1) & \bar{x}_2(t_1) & \cdots & \bar{x}_d(t_1) \\ \vdots & \vdots & & \vdots \\ \bar{x}_1(t_N) & \bar{x}_2(t_N) & \cdots & \bar{x}_d(t_N) \end{bmatrix}$$

$$\mathbf{C}^* = \begin{bmatrix} c_{11}^* & c_{12}^* & \cdots & c_{1d}^* \\ c_{21}^* & c_{22}^* & \cdots & c_{2d}^* \\ \vdots & \vdots & & \vdots \\ c_{q1}^* & c_{q2}^* & \cdots & c_{qd}^* \end{bmatrix}$$

由动力系统的稀疏性可知,矩阵 \mathbf{C}^* 中的元素大部分都为零,只有极少数元素是不为零的。

由于在现实生活中收集到的数据大多数带有噪声,从这些含噪的数据里挖掘潜在的动力模型往往具有挑战性,但这通常也是检验一个算法稳定性和鲁棒性的重要措施。基于此,本文通过在精确的状态数据矩阵中加入不同水平的噪声,以此检验提出的线性多步稀疏识别算法的性能。那么,加入噪声以后得到的观测数据矩阵 \mathbf{Y} 可以表示如下:

$$\mathbf{Y} = \mathbf{X} + \boldsymbol{\varepsilon} \quad (4)$$

其中, $\boldsymbol{\varepsilon} \in \mathbf{R}^{(N+1) \times d}$ 表示均值为 0,方差为 σ^2 且独立同分布的高斯白噪声矩阵。

那么,稀疏系数矩阵 \mathbf{C}^* 的求解问题式(3)被转化

为求解以下问题:

$$\bar{\mathbf{Y}} \approx \Phi(\mathbf{Y}) \mathbf{C} \quad (5)$$

在这里,基函数库 $\Phi(\mathbf{Y})$ 是与 $\Phi(\mathbf{X})$ 定义类似的矩阵形式,其中矩阵 \mathbf{X} 被矩阵 \mathbf{Y} 代替。

接下来,使用稀疏回归的方法解决问题式(5),即求解以下最小化问题:

$$\min_c \|\Phi(\mathbf{Y}) \mathbf{C} - \bar{\mathbf{Y}}\|_{2+\lambda} + \|\mathbf{C}\|_1 \quad (6)$$

在这里,参数 $\lambda > 0$ 用于平衡式(6)的精度和解的稀疏性。

3 线性多步稀疏识别算法

3.1 线性多步方法

本文应用线性多步方法对二阶动力系统式(1)进行时间离散的分量结果是

$$\sum_{j=0}^s \alpha_j x_i(t_{n+j}) = h^2 \sum_{j=0}^s \beta_j f_i(x(t_{n+j})) + \Delta(h^{p+2})$$

$$n = 0, \dots, N - s \quad (7)$$

其中, s 表示线性多步方法的步数, $\Delta(h^{p+2})$ 是应用该方法离散二阶动力系统的局部截断误差, p 是对应步数为 s 的线性多步方法的阶数。

本文应用步数 $s=2, 3, 4$ 的线性多步方法离散二阶方程(1),具体的离散公式如下:

当 $s=2, p=2$ 时,线性多步方法的离散公式为

$$x_{n+1} - 2x_n + x_{n-1} = h^2 f_n$$

当 $s=3, p=3$ 时,线性多步方法的离散公式为

$$x_{n+1} - 2x_n + x_{n-1} = h^2 \left(\frac{13}{12} f_n - \frac{1}{6} f_{n-1} + \frac{1}{12} f_{n-2} \right)$$

当 $s=4, p=4$ 时,线性多步方法的离散公式为

$$x_{n+1} - 2x_n + x_{n-1} = h^2 \left(\frac{7}{6} f_n - \frac{5}{12} f_{n-1} + \frac{1}{3} f_{n-2} - \frac{1}{12} f_{n-3} \right)$$

结合上述分析,式(7)的矩阵形式可以转化为

$$\mathbf{A} \mathbf{X} = h^2 \mathbf{B} \mathbf{F}(\mathbf{X}) + \Delta(h^{p+2}) \quad (8)$$

其中,矩阵 $\mathbf{A}, \mathbf{B} \in \mathbf{R}^{(N-s+1) \times (N+1)}$ 可以表示为

$$\mathbf{A} = \begin{bmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_s \\ \alpha_0 & \alpha_1 & \cdots & \alpha_s \\ \vdots & \vdots & & \vdots \\ \alpha_0 & \alpha_1 & \cdots & \alpha_s \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \beta_0 & \beta_1 & \cdots & \beta_s \\ \beta_0 & \beta_1 & \cdots & \beta_s \\ \vdots & \vdots & & \vdots \\ \beta_0 & \beta_1 & \cdots & \beta_s \end{bmatrix}$$

接着,结合式(1)与式(3),可以得出:

$$\mathbf{F}(\mathbf{X}) = \Phi(\mathbf{X}) \mathbf{C}^*$$

进一步与式(8)结合,可得:

$$\mathbf{A} \mathbf{X} = h^2 \mathbf{B} \Phi(\mathbf{X}) \mathbf{C}^* + \Delta(h^{p+2}) \quad (9)$$

在这里,当面对的数据是含噪声的观测数据时,式(9)即为

$$\mathbf{A} \mathbf{Y} = h^2 \mathbf{B} \Phi(\mathbf{Y}) \mathbf{C} + \Delta(h^{p+2})$$

那么,识别潜在的二阶动力模型式(1)的问题被归结为

求解

$$AY \approx h^2 B \Phi(Y) C \quad (10)$$

接下来,使用稀疏回归方法解决问题式(10),即求解下面这样一个最小化问题:

$$\min_c \| h^2 B \Phi(Y) C - AY \|_2^2 + \lambda \| C \|_1 \quad (11)$$

由过去的研究结果可知,应用线性多步方法离散动力系统,能够在很大程度上提高算法的精度和鲁棒性。但是当面对的状态数据是含噪声的观测数据时,数值时间导数将放大这种噪声,特别地,当数值时间导数的阶数越大,对噪声的放大程度也越大。在过去,研究者们尝试了各种各样的方法对数据做消噪处理,比较经典的有移动最小二乘消噪^[17],该方法通常与一些模型识别算法一起出现,用于算法准备阶段的数据消噪。接下来,在文中引入广义最小二乘法,并将它用于数据消噪。

3.2 广义最小二乘法和近似协方差矩阵 Σ

在广义最小二乘方法原理^[4]的指导下,将最小化问题式(11)转化为求解以下问题:

$$\min_c D^T \Sigma^{-1} D + \lambda \| C \|_1 \quad (12)$$

其中, $D = h^2 B \Phi(Y) C - AY$ 。矩阵 Σ 是使用广义最小二乘法计算出的噪声残差项的近似协方差矩阵。通过对最小化问题式(11)进行加权,可以使线性多步方法的每一个时间步对计算稀疏矩阵 C 都有近乎相同的重要性。根据先前的研究经验^[4,16], Σ 可以选择为 $\Sigma = (hB - A)(hB - A)^T$ 。

3.3 线性多步稀疏识别算法

输入: 观测数据矩阵 $Y \in \mathbb{R}^{(N+1) \times d}$, 基函数库 $\Phi(Y) \in \mathbb{R}^{(N+1) \times q}$, s 步线性多步方法的离散系数矩阵 $A, B \in \mathbb{R}^{(N-s+1) \times (N+1)}$, 时间步长 h , 近似协方差矩阵 Σ , 阈值参数 λ 。

步骤 1 对 $j=1, 2, \dots, d$, 初始化:

$$c_j^0 = \operatorname{argmin}_c \{ D^T \Sigma^{-1} D + \lambda \| c \|_2^2 \}$$

其中, $D = h^2 B \Phi(Y) C - Ay_j$, y_j 表示第 j 维的状态数据向量。设 I_j^0 是对应集合 c_j^0 中元素小于 λ 的所有元素的集合, 那么设 $I_j^{0*} = (I_j^0)^c$ 中元素的个数为 r 。

步骤 2 开始循环 $k=1, 2, \dots, r$, 求解:

$$c_j^k = \operatorname{argmin}_c \{ D^T \Sigma^{-1} D + \lambda \| c \|_2^2 \}$$

在这里, 有 $c(I_j^{k-1}) = 0$, 同样地, 设 I_j^k 是对应集合 c_j^k 中元素小于 λ 的所有元素的集合。直到集合 c_j^k 中的元素均大于 λ , 结束循环。

输出: 稀疏系数矩阵:

$$C^k = [c_1^k, \dots, c_d^k] \in \mathbb{R}^{q \times d}$$

4 数值实验

为获得精确的状态数据,需先将二阶动力系统转化为一阶动力系统,再利用 MATLAB 内置的函数 ode45 求解该一阶动力系统,从而得到精确的状态数据。为检测所提出的算法在识别未知模型上的性能和鲁棒性,在精确的数据中加入均值为 0, 方差为 σ^2 的噪声生成观测数据,在这里,计算 σ 的具体公式为

$$\sigma = \sigma_{NR} \frac{\| X \|_F}{\sqrt{d(N+1)}}$$

其中, $\| X \|_F$ 是矩阵 X 的 Frobenius 范数。为了更好地呈现线性多步稀疏识别算法在重构潜在动力模型上的精度,定义模型重构的相对系数误差为

$$e_c = \frac{\| C - C^* \|_F}{\| C^* \|_F}$$

4.1 实验 1

首先,考虑使用提出的线性多步稀疏识别算法重构以下二阶动力系统,为便于下文描述,将其简记为模型 1。在此之前确定时间变量的范围为 $t \in [0, 10]$, 初始条件为 $x_0 = 1$ 和 $\dot{x}_0 = 2$ 。模型 1 的具体形式为

$$\ddot{x} = -0.2 - \sin(x)$$

根据少量先验知识,在这个实验中所选择的基函数库应包含数据 x 的 k 次单项式和 $\sin(x)$ 等三角函数。

考虑在数据无噪声的情况下,基于不同的时间步长 h ,使用线性多步稀疏识别算法从数据中找到隐藏的二阶动力系统。在这里选择时间步长为 0.005、0.008、0.01、0.025、0.04、0.08、0.1、0.2, λ 的取值为 0.15。使用提出的算法可以得出不同时间步长下模型识别结果的相对误差,参见图 1。

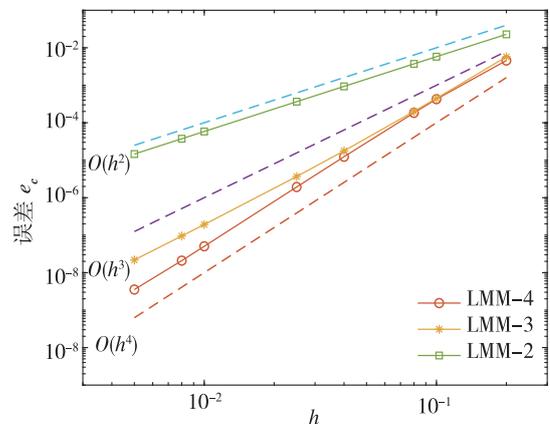


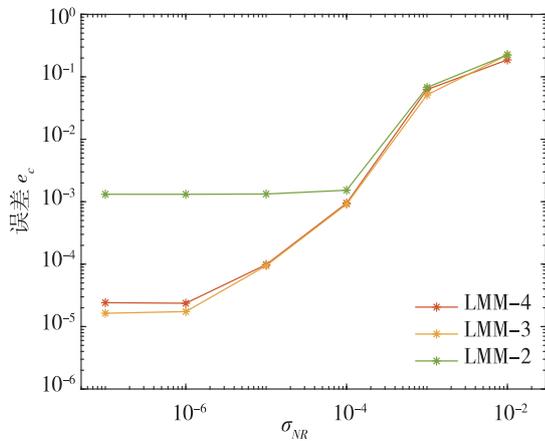
图 1 使用线性多步稀疏识别算法识别模型 1 的误差 (数据无噪声)

Fig. 1 The error in identifying model 1 using LMM sparse identification algorithm (noise-free data)

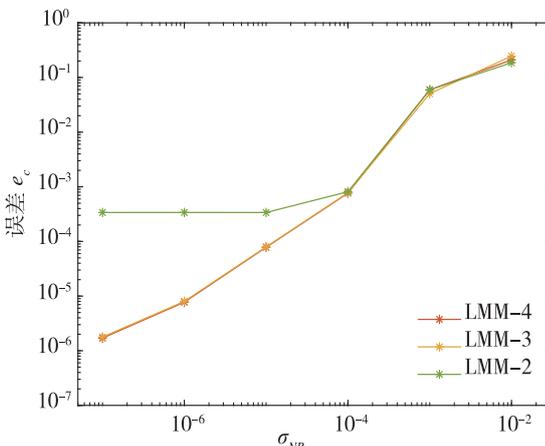
图 1 呈现了数据无噪声情况下,基于不同时间步长 h ,使用线性多步稀疏识别算法重构模型 1 的相对系数误差结果。从图中可以看到:时间步长越小,模型识别结果的精度越高;在相同的时间步长下,所使用的线性多步方法的阶数越高,其对应的线性多步稀疏识别算法的精度也越高。

考虑在数据含噪声的情况下,基于时间步长 $h = 0.05, h = 0.025$ 和 $h = 0.001$,使用所提出的算法从数据中恢复潜在的二阶动力系统。在这个实验中,选择加入的噪声水平为 $10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}$, λ 的取值为 0.15。使用线性多步稀疏识别算法重构潜在模型的相对系数误差见图 2。

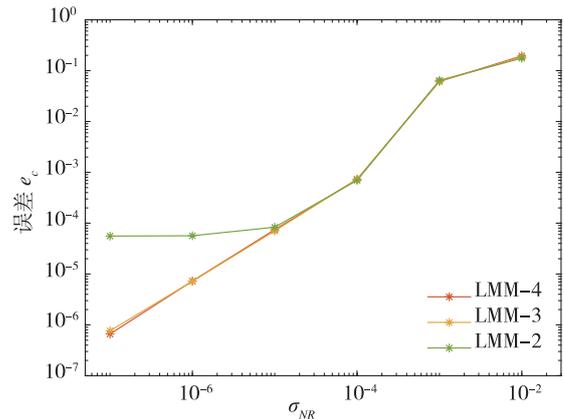
由图 2 可以看出:在时间步长相同的情况下,数据中含有的噪声越大,使用线性多步稀疏识别算法重构潜在模型结果的相对误差也越大。从图 2 中,还可以看到:当 $s = 3$ 和 4 的时候,模型识别的结果相差不大,当 $s = 2$ 的时候,识别结果的相对误差比前两个 $s = 3, 4$ 时更大;且在噪声水平相同的情况下,时间步长越小,使用提出的算法重构模型的精度也越高。



(a) $h = 0.05$



(b) $h = 0.025$



(c) $h = 0.001$

图 2 不同时间步长情况下,使用线性多步稀疏识别算法识别模型 1 的误差(数据含噪声)

Fig. 2 Errors in identifying model 1 using LMM sparse identification algorithm at different timesteps (noise data)

图 3 是使用线性多步稀疏识别算法对模型 1 进行识别的预测结果。由图 3 可知,使用线性多步稀疏识别算法进行模型重构得到的预测结果与参考结果之间重合度很高。

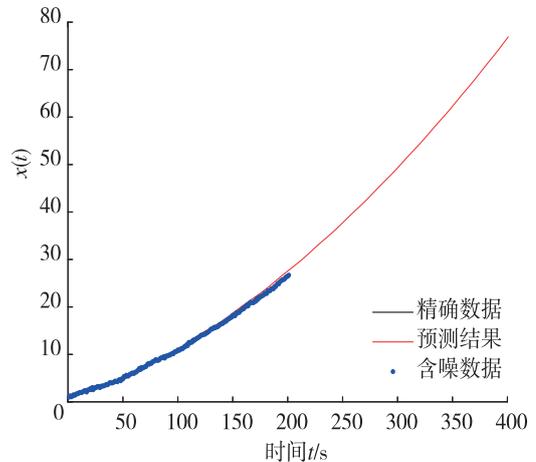


图 3 线性多步稀疏识别算法识别模型 1 的结果

($\sigma_{NR} = 10^{-2}, h = 0.05, s = 4$)

Fig. 3 Errors in identifying model 1 using LMM sparse identification algorithm with $\sigma_{NR} = 10^{-2}, h = 0.05, s = 4$

4.2 实验 2

现在,考虑使用提出的线性多步稀疏识别算法重构以下二维二阶动力系统,为便于下文描述,将其简记为模型 2。在这里确定时间变量的范围为 $t \in [0, 10]$,初始条件为 $x_0 = [0.2, -0.25]^T, \dot{x}_0 = [0.3, -0.25]^T$,模型 2 的具体形式为

$$\begin{cases} \ddot{x}_1 = -x_1 - 2x_1x_2 \\ \ddot{x}_2 = -x_2 - x_1^2 + x_2^2 \end{cases}$$

根据少量先验知识,在这个实验中所选择的基函数数据库应包含数据 x 的 k 次单项式。

考虑在数据无噪声的情况下,基于不同的时间步长 h ,使用本文提出的算法从状态数据中挖掘潜在的二阶动力系统。在这个实验中,用于检验算法性能的时间步长为 0.005、0.008、0.01、0.025、0.04、0.08、0.1、0.2, λ 的取值为 0.25。使用线性多步稀疏识别算法,可以得到不同时间步长下模型识别结果的相对系数误差,并将结果呈现在图 4 中。

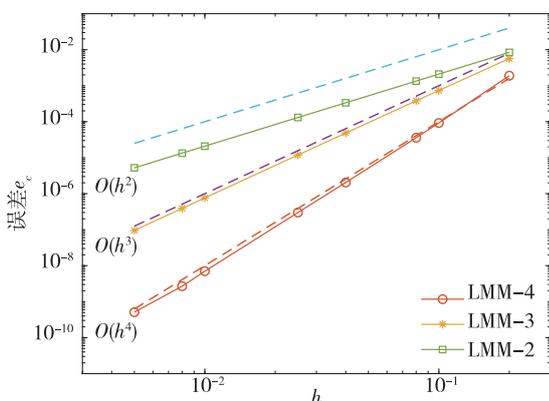


图 4 使用线性多步稀疏识别算法识别模型 2 的误差 (数据无噪声)

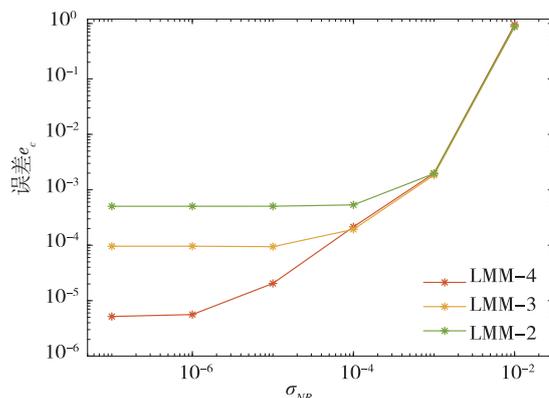
Fig. 4 Errors in identifying model 2 using LMM sparse identification algorithm (noise-free data)

图 4 呈现了数据无噪声情况下,基于不同的时间步长 h ,使用线性多步稀疏识别算法重构模型 2 的相对系数误差结果。从图中可以看出:在数据不含噪声的情况下,时间步长越小,使用提出的算法识别二阶动力模型的精度也越高。除此之外,还发现,用于时间离散的线性多步方法的阶数越高,对应的模型识别结果精度也越高。

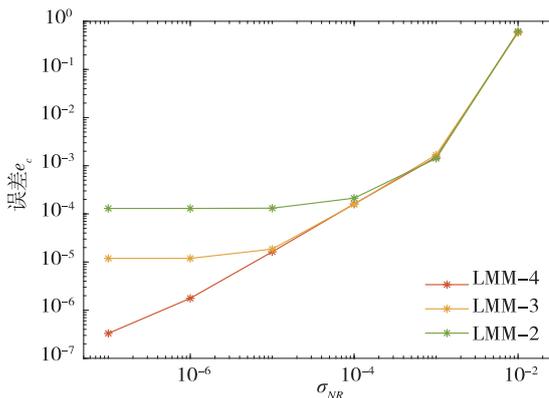
考虑在数据含噪声的情况下,基于时间步长 $h = 0.05, h = 0.025$ 和 $h = 0.001$,使用所提出的算法从数据中挖掘潜在的二阶动力系统。在这个实验中,选择加入的噪声水平为 $10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$, λ 的取值为 0.25。使用线性多步稀疏识别算法重构模型 2 的相对系数误差结果见图 5。

观察图 5,不难看出:选择相同时间步长的情况下,观测数据中含有的噪声越大,使用提出算法识别模型 2 的结果误差也越大;除此之外,在使用线性多步方法进行时间离散的过程中,线性多步方法的步数越多,对应

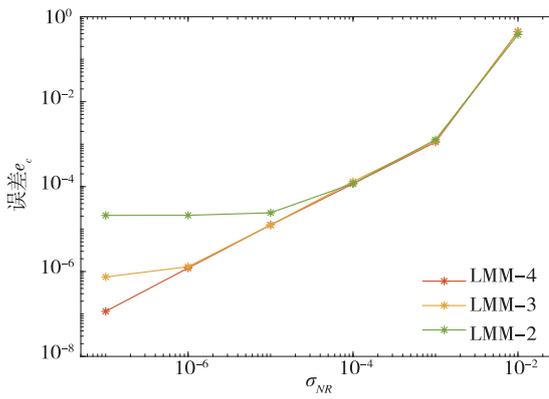
模型识别算法的精度也越高。步数 $s = 4$ 的时候,模型识别结果相对误差最小; $s = 3$ 的时候,将图 5 中的 3 个子图结合起来观察可以发现,在噪声水平相同的时候,状态数据的时间步长越小,使用线性多步稀疏识别算法进行模型识别的精度也越高。



(a) $h = 0.05$



(b) $h = 0.025$



(c) $h = 0.001$

图 5 不同时间步长情况下,使用线性多步稀疏识别算法识别模型 2 的误差 (数据含噪声)

Fig. 5 Errors in identifying model 2 using LMM sparse identification algorithm at different timesteps (noise data)

图 6 是使用线性多步稀疏识别算法对模型 2 进行识别的预测结果。由图 6 可知,使用线性多步稀疏识别算法从数据中挖掘潜在动力模型表现很好。

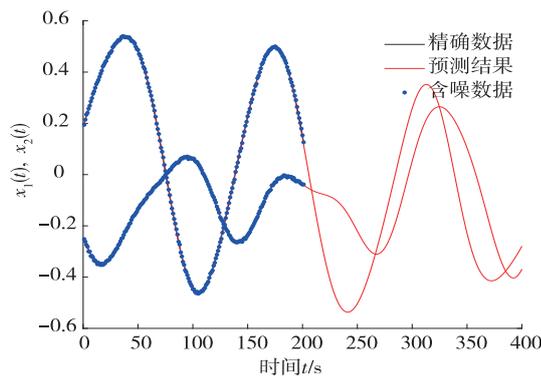


图 6 线性多步稀疏识别算法识别模型 2 的结果
($\sigma_{NR} = 10^{-2}, h = 0.05, s = 4$)

Fig. 6 Errors in identifying model 2 using LMM sparse identification algorithm $\sigma_{NR} = 10^{-2}, h = 0.05, s = 4$

5 结论

针对二阶动力系统的模型识别问题,提出了一种基于线性多步方法的稀疏识别方法,即线性多步稀疏识别方法。首先,通过构造一个确定的基函数库近似未知的二阶动力系统;其次,应用线性多步方法离散近似后的二阶动力系统;然后,在广义最小二乘原理的指导下,计算噪声残差项的近似协方差矩阵,再利用该矩阵对时间离散后得到的最小化问题进行加权,使得线性多步方法离散动力系统的每一个时间步具有近乎相等的重要性,从而达到数据消噪的目的;最后,使用稀疏回归算法从数据中挖掘出潜在的二阶动力模型。数值实验结果表明,使用提出的算法从观测数据中重构潜在动力模型具有较高的精度和鲁棒性。

参考文献(References):

- [1] BRUNTON S L, PROCTOR J L, KUTZ J N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems[J]. *Proceedings of the National Academy of Sciences*, 2016, 113(15): 3932–3937.
- [2] CORTIELLA A, PARK K C, DOOSTAN A. Sparse identification of nonlinear dynamical systems via reweighted ℓ_1 -regularized least squares[J]. *Computer Methods in Applied Mechanics and Engineering*, 2021, 376: 113620–113627.
- [3] MANGAN N M, KUTZ J N, BRUNTON S L, et al. Model selection for dynamical systems via sparse regression and information criteria[J]. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2017, 473(2204): 20170009–2170018.
- [4] MESSENGER D A, BORTZ D M. Weak SINDy: Galerkin-based data-driven model selection[J]. *Multiscale Modeling & Simulation*, 2021, 19(3): 1474–1497.
- [5] TIBSHIRANI R. Regression shrinkage and selection via the lasso[J]. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1996, 58(1): 267–288.
- [6] GOYAL P, BENNER P. Discovery of nonlinear dynamical systems using a Runge-Kutta inspired dictionary-based sparse regression approach[J]. *Proceedings of the Royal Society A*, 2022, 478(2262): 20210883–20210888.
- [7] HARLIM J, JIANG S W, LIANG S, et al. Machine learning for prediction with missing dynamics[J]. *Journal of Computational Physics*, 2021, 428: 109922–109929.
- [8] QIN T, WU K, XIU D. Data driven governing equations approximation using deep neural networks[J]. *Journal of Computational Physics*, 2019, 395: 620–635.
- [9] DU Q, GU Y, YANG H, 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.
- [10] RAISSI M, KARNIADAKIS G E. Hidden physics models: Machine learning of nonlinear partial differential equations[J]. *Journal of Computational Physics*, 2018, 357: 125–141.
- [11] RAISSI M, PERDIKARIS P, KARNIADAKIS G E. Inferring solutions of differential equations using noisy multi-fidelity data [J]. *Journal of Computational Physics*, 2017, 335: 736–746.
- [12] RAISSI M, PERDIKARIS P, KARNIADAKIS G E. Machine learning of linear differential equations using Gaussian processes[J]. *Journal of Computational Physics*, 2017, 348: 683–693.
- [13] ZHANG S, LIN G. Robust data-driven discovery of governing physical laws with error bars[J]. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2018, 474(2217): 20180305–20180311.
- [14] ZHANG S, LIN G. SubTSBR to tackle high noise and outliers for data-driven discovery of differential equations[J]. *Journal of Computational Physics*, 2021, 428: 109962–109969.
- [15] LU F, ZHONG M, TANG S, et al. Nonparametric inference of interaction laws in systems of agents from trajectory data[J]. *Proceedings of the National Academy of Sciences*, 2019, 116(29): 14424–14433.
- [16] CHEN H. Data-driven sparse identification of nonlinear dynamical systems using linear multistep methods[J]. *Calcolo*, 2023, 60(1): 11–18.
- [17] HE Y, KANG S H, LIAO W, 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.

责任编辑:李翠薇