ORIGINAL PAPER

Joint sparse least squares via generalized fused lasso penalty for identifying nonlinear dynamical systems

Yisha Lu · Wei Xu (iD) **· Lizhi Niu · Wenting Zhang · Minjuan Yuan**

Abstract This paper proposes a joint sparse least-square model that utilizes a generalized fused lasso penalty to jointly identify governing equations of nonlinear dynamical systems from multiple noisy state measurements. The idea to combine a data fusion approach with least squares is that the fusion term added could make use of the more similarity information carried by different state measurements, which can further improve identified accuracy of dynamical system in regression tasks. The threshold joint sparse least-square algorithm is developed, wherein the threshold parameter is picked using the L-curve criterion. The results from the simulation experiment demonstrate our method possesses a higher accuracy in identifying dynamical system compared to traditional sparse least squares, indicating its potential power to understand gradually more complex systems from multiple data sources in the future.

Keywords Nonlinear system identification · Sparse regression · Generalized fused lasso penalty · Threshold joint sparse least-square algorithm

1 Introduction

Data-driven is an emerging and important study field of engineering mathematics and mathematical physics in identifying governing equations of dynamical systems from noisy state measurements. Thanks to the fact that the amount of digitally stored information has explored over the past two decades, facilitating a variety of innovations and growing intersection of data-driven methods and applied optimization. The purpose of this research is expected to discover the structure that explains spatiotemporal activities of the system, thereby further providing more comprehensive information that can hopefully deepen understanding and analysis the structure and operating mechanism of the system. Hence, it is significant and meaningful to investigate how to learn more accurate dynamical systems from abundant observational or simulated time-series data and grasp their physical behaviors and interpretability.

In recent years, many efforts have been made to construct new models or explore computing methods for identifying governing equations of the systems from measured data [1, 2]. The early method of building possible nonlinear chaotic systems from data was discussed [3]. Advanced regression method from statistics, such as sparse regression, is driving new algorithms that identify parsimonious nonlinear dynamical systems from measurements of complex systems. While the expression of the dynamical

Y. Lu · W. Xu (✉) · L. Niu · W. Zhang · M. Yuan
Department of Applied Mathematics, Northwestern
Polytechnical University, Xi'an 710072,
Shaanxi, People's Republic of China
e-mail: weixunpu@nwpu.edu.cn

system is unknown, a special method of approximating dynamical systems based on the overcomplete functional space of state variables is studied, which eliminates the expansion terms that have no contribution to the dynamics and attracts more and more attention [4, 5]. The symbolic regression was used in the past work to find equations of dynamical systems directly from observational data [6–10] and also was employed to predict and identify the dynamical systems [11–13]. However, it may easily cause overfitting.

With the influx of a large number of experimental data, regression and machine learning are becoming important tools to discover dynamical systems from data. The Least Absolute Shrinkage and Selection Operator (LASSO) algorithm was proposed to alleviate the influence of noise when identifying nonlinear dynamical systems from observational experimental data [14]. Based on minimum coefficient error of the polynomial basis expansion, Wang et al. [15] proposed the compressed sensing to discover the nonlinear dynamical system. Brunton et al. [16] proposed a sparse identification of nonlinear dynamics (SINDy) method that employs the Sequentially Thresholded Least Squares (STLS) algorithm to learn governing equations from the overcomplete candidate basis function. The SINDy was extended to identify the system including input and control [17] and was also extended to identify the dynamical system with rational function nonlinearity [18], integral terms [19], and highly damaged and incomplete data [20]. The SINDy method was also used to include partial derivatives, making it possible to identify partial differential equation [21, 22]. Most methods of system identification involve in the form of regression from data to dynamics, and the difference between different technologies is the limitation of this regression.

In all the above methods, a common assumption is that the data are collected from a measurement source or a simulation experiment, and all the data here are discrete points obtained by discretizing the system in time or space. However, in practical dynamical system studies, obtaining data from some complex systems

needs to consume a lot of manpower and financial resources, and at the same time, higher accuracy is required for system identification problems. Therefore, it is very necessary to make full use of all the information collected in as many ways as possible, and the information usually has prominent similarities that are generally ignored. Separate estimates will lack power that uses important shared patterns due to the limited size of each individual data source, but simply merging datasets together may suffer from large errors. Thus, we use the similarity information to improve the performance of dynamical system identification through joint analysis of multiple datasets. The fused lasso penalty is chosen because it has been successfully applied to jointly estimate multiple graphical models to discover differential dependency networks [23–25]. To the best of the authors' knowledge, it is an unexplored topic to apply it to identification task of nonlinear dynamical systems from multiple measurement datasets for achieving sparsity and more accurate estimates simultaneously. Hence, there has been a growing interest in fully mining the information of multiple datasets to fill this gap.

Motivated by this, we propose a novel joint sparse least-square model via generalized fused lasso penalty (JSLSGF) to jointly identify multiple sparse least squares models corresponding to different data sources. As illustrated in Fig. 1 with data from L measurement sources, the main idea is to explore the similarity of multiple sources, thereby improving the identification accuracy of the system. In this way, the generalized fused lasso penalty is imposed on the coefficient values of the basis function, in which the l_1 - norm penalty of the system coefficients for each set of data encourages the sparsity for each individual coefficient vector that takes advantage of the sparsity prior property of system identification, and the l_1 - norm penalty of the difference between the system coefficients presented by every two datasets stimulates them to share the similar physical mechanism, which utilizes the similarity among different measurements. Inspired by the optimization framework [26], we here
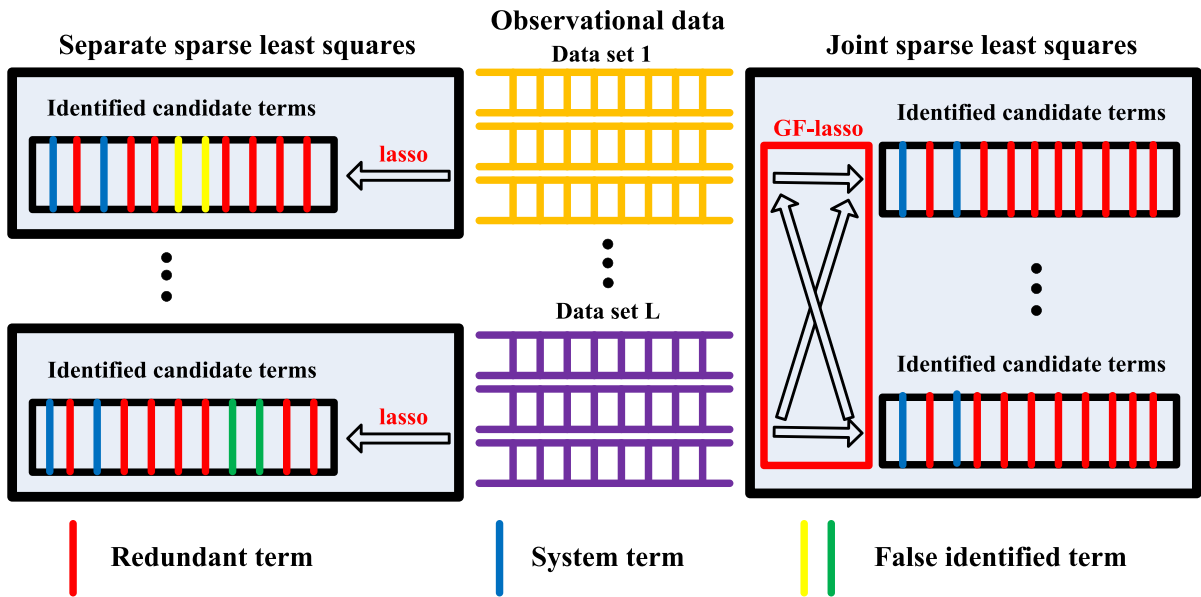
Fig. 1 Comparison of the separate sparse least squares and joint sparse least squares

design an effective two-step algorithm to solve this optimization problem and develop a selected criterion of the threshold parameter in the first step of the algorithm according to the L-curve criterion. Using multiple simulated test datasets and taking the comprehensive comparison, we demonstrate the effectiveness of the proposed method in identifying nonlinear dynamical systems from multiple measurements.

The rest of the paper is arranged as follows: In Section 2, as preliminary, the background of identifying dynamical systems from state measurements by sparse regression techniques is introduced. In Sect. 3, the JSLSGF model and threshold joint sparse least-square algorithm are proposed, and the selection of the threshold parameter and the calculation of time derivatives of state variables are discussed. In Sect. 4, three numerical experiments are presented to illustrate that our method often results in improved accuracy compared with separate analysis on each data source for system identification. Section 5 concludes this paper.

2 Problem and background

Dynamical systems provide physical insight and interpretability into a system's behavior through trajectories and solutions obtained from the governing equations of motion. In most regression problems, only a few terms are important, and thus, the feature selection mechanism is required. Considering the dynamical system [27]

$$\dot{x}(t) = \frac{\mathrm{d}x(t)}{\mathrm{d}t} = f(x(t)), \ x(0) = x_0. \tag{1}$$

The (known or measurable) state variable $x(t)$ and derivative $\dot{x}(t)$ denote the evolution of the system on discrete time points $t_1, t_1, \ldots, t_m$, and m is the number of samples, the function $f(x(t))$ constrains dynamics that describe the motion of the systems. Our goal is to seek a parsimonious system that $f(x(t))$ includes only a few active terms, meaning that it is sparsely expressed in candidate basis function.

We review the SINDy that is used as a sparse regression problem. The expanded library of candidate p functions $\Phi(x) = [\phi_1(x) \cdots \phi_p(x)] \in \mathbb{R}^p$ is built, wherein each $\phi_i(x)$, $i = 1, \ldots, p$ is a candidate function term. The select and use of the basis function usually reflect some knowledge of the system of interest, gathering such information often requires a certain domain knowledge. The common choice of the basis function is constant, polynomial, and trigonometric; we select polynomials as basis function because the Taylor expansion of complex nonlinear systems is usually polynomial. The basis function is used to establish an overdetermined linear system that can be written as:

$$\dot{x}(t) = \Phi(x)\xi, \tag{2}$$

where the unknown vector ξ is estimated sparse coefficient values of the basis function that decide which terms work in $f(x(t))$.

Equation (2) does not actually hold because there is always noise in practical applications. Instead, calculating the coefficient vector ξ that satisfies the least-square constraint

$$\|\Phi(x)\xi - \dot{x}\|_2 \le \delta, \tag{3}$$

where $\delta \ge 0$ is a tolerance parameter to avoid overfitting. To solve Eq. (3) and obtain a sparse solution, one usually relaxes the l_0 - norm via convex l_1 - norm penalty

$$\min_{\xi} \|\xi\|_1 \quad \text{s.t.} \quad \|\Phi(x)\xi - \dot{x}\|_2 \le \delta. \tag{4}$$

The unconstrained expression is given by:

$$\arg\min_{\xi} \frac{1}{2} \|\Phi(x)\xi - \dot{x}\|_2^2 + \lambda\|\xi\|_1, \tag{5}$$

where λ is a nonnegative parameter that balances the trade-off between fitting error and sparsity.

Enabled by unprecedented availability of data and computational resources, datasets from multiple measurement sources can be collected using different devices or collection ways, and those datasets have the prominent similarity that is usually neglected, even though it is helpful to enhance the system identification performance.

3 Joint sparse least squares via generalized fused lasso penalty approach

3.1 Generalized model

To make full use of the similarity information among multiple noisy state measurements and their derivatives to boost the accuracy and robustness of SINDy for dynamical system identification, we add an additional generalized fused lasso penalty to the sparse least-square model. The generalized fused lasso penalty encourages two sparse vectors to share a similar system mechanism, which is crucial to achieve the fusion of solutions. As described in the introduction and shown in Fig. 1, we assume that L state measurements are collected and propose the joint sparse least squares via generalized fused lasso penalty (JSLSGF) model as

$$\arg\min_{[\xi_1,\ldots,\xi_L]} \sum_{l=1}^{L} \frac{1}{2} \|\Phi(x)\xi_l - \dot{x}_l\|_2^2 + \sum_{l=1}^{L} \lambda\|\xi_l\|_1 + \sum_{l<l'} \beta\|\xi_l - \xi_{l'}\|_1, \tag{6}$$

where $[\xi_1, \ldots, \xi_L] \in \mathbb{R}^{p \times L}$ are L sparse coefficient vectors of the system, and each sparse column vector corresponds to the basis function coefficient values of each data resource and determines which terms are contributing in the candidate basis function, λ and β are nonnegative parameters. Note that the penalty terms (the second and third terms) of the object function (6) are convex, so the object function (6) is convex with respect to ξ. Our approach is inspired by the work [28], wherein the generalized fused lasso has been shown to outperform the standard l_1 - norm in promoting sparsity and accuracy from varied data sources. The l_1 - norm on every ξ_l, controlled by λ, encourages the sparsity for individual identified coefficient vector. The l_1 - norm penalty on the differences between every two coefficient vectors from different data resources encourage them to share a similar physical mechanism, which applies the similarity among different measurements. The parameter β plays a vital role to adjust the degree of fusion. It is not fused across multiple sparse vectors when $\beta = 0$. The object

function (6) is achieved only when all the sparse vectors are identical to each other with $\beta = \infty$. Especially for $L = 1$, the proposed model reduces to traditional sparse least squares.

3.2 The proposed optimization algorithm

To solve the JSLSGF model that is a special fused lasso signal approximation [29], the two-step (sparsification and fusion) approach is an efficient algorithm and available as in [23, 26].

In the sparsification step, the preliminary sparse vector is computed by setting $\beta = 0$ and writing

$$\underset{[\xi_1,\dots,\xi_L]}{\arg\min} \sum_{l=1}^{L} \frac{1}{2} \|\Phi_l(x)\xi_l - \dot{x}_l\|_2^2 + \sum_{l=1}^{L} \lambda \|\xi_l\|_1. \qquad (7)$$

Considering that the convergence of the sequential thresholded least-square (STLS) algorithm used in SINDy had been guaranteed [26], we utilize the STLS algorithm to improve the numerical robustness of the system identification in calculation process. Computing the derivative of $\sum_{l=1}^{L} \frac{1}{2} \|\Phi_l(x)\xi_l - \dot{x}_l\|_2^2$ with respect to ξ_l, we enjoy

$$\Phi_l^T(x)(\Phi_l(x)\xi_l - \dot{x}_l) = 0, \, l = 1,\dots,L, \qquad (8)$$

and have the preliminary solution

$$\xi_l = (\Phi_l^T(x)\Phi_l(x))^{-1}\Phi_l^T(x)\dot{x}_l, \quad l = 1,\dots,L. \qquad (9)$$

Introducing the discriminant criteria

$$M = \left\{ j : |\xi_{lj}| \geq \alpha \times \max_{1 \leq i \leq p} |\xi_{li}| \right\}, \quad j = 1,\dots,p, \qquad (10)$$

where the threshold parameter $\alpha \in (0, 1)$. Its selection is very challenging and difficult because the content of ξ_l is unknown and mainly depends on the candidate basis function. To alleviate this problem, we adopt L-curve criterion to guide the selection of the tuning threshold parameter α, yielding a much less sensitive calculation process. Then, the preliminary sparse solution is achieved.

In the fusion step, setting $\lambda = 0$, the object function (6) can be rewritten as:

$$\underset{[\xi_1,\dots,\xi_L]}{\arg\min} \sum_{l=1}^{L} \frac{1}{2} \left\|\Phi_l^M(x)\xi_l^M - \dot{x}_l\right\|_2^2 + \sum_{l<l'} \beta \left\|\xi_l^M - \xi_{l'}^M\right\|_1, \qquad (11)$$

where $\Phi_l^M(x)$ is the updated matrix composed of the columns of the matrix $\Phi_l(x)$, the position of the columns depends on the row of the nonzero terms in the derived sparse vector ξ_l from the sparsification step. ξ_l^M is the vector of all the nonzero terms in ξ_l, the items that are set to zeros not considered in the subsequent iteration process, which greatly reduces the computational complexity. This step fuses the terms in the sparse vectors that do not have significant difference (dependent on β). We say that the j - th polynomial term in $\Phi_l(x)$ and $\Phi_{l'}(x)$ is fused between the l-th and l'-th data sources if $\xi_{lj} = \xi_{l'j}$.

After the above two-step iterations, until the absolute value of the difference between the sparse vectors for the next iteration and the previous iteration reaches a minimum.

To assess the quality of the solution obtained by STLS and JSLSGF, the arithmetic mean vector $\overline{\xi}$ of L sparse solution vectors is introduced as the final solution. The relative error of the solution is reported as:

$$\varepsilon_\xi = \frac{\left\|\overline{\xi} - \xi\right\|_2}{\|\xi\|_2}, \qquad (12)$$

where $\overline{\xi}$ and ξ are the approximate and exact solution vectors of each state variable, respectively. The iteration will terminate when the error of the solution vector satisfies $\left\|\overline{\xi}^{w+1} - \overline{\xi}^w\right\|_2^2 \leq \varepsilon_0 \, (\varepsilon_0 = 10^{-10})$. We summarize the pseudo-code of the proposed threshold joint sparse least-square algorithm as below.

Algorithm 1 The threshold joint sparse least-square algorithm

1. **Input:** Set the matrix $\Phi_l(x)$, vector $\dot{x}_l$, $w=1$, regularization parameters λ, β, the threshold parameter α and termination threshold ε_0.

2. **while** not converged **do**

3.   if $w=1$

4. $\xi_l = (\Phi_l^T(x)\Phi_l(x))^{-1}\Phi_l^T(x)\dot{x}_l$;

5.   else

6. $\xi_l^M = ((\Phi_l^M(x))^T\Phi_l^M(x))^{-1}(\Phi_l^M(x))^T\dot{x}_l$;

7.   Update $M = \left\{ j : \left|\xi_{lj}\right| \geq \alpha \times \max_{1\leq i \leq p}\left|\xi_{li}\right| \right\}$;

8.   Update $\underset{[\xi_1,\cdots,\xi_L]}{\arg\min} \sum_{l=1}^{L} \frac{1}{2}\left\|\Phi_l^M(x)\xi_l^M - \dot{x}_l\right\|_2^2 + \sum_{l<l'}\beta\left\|\xi_l^M - \xi_{l'}^M\right\|_1$;

9.   Update $M = \left\{ j : \left|\xi_{lj}\right| \geq \alpha \times \max_{1\leq i \leq p}\left|\xi_{li}\right| \right\}$;

10. Update $\bar{\xi} = \frac{1}{L}\sum_{l=1}^{L}\xi_l$;

11. Update w: $w=w+1$;

12. Termination criterion: $\left\|\bar{\xi}^{w+1} - \bar{\xi}^w\right\|_2^2 \leq \varepsilon_0$;

13. **end while**

14. **Output:** the solution vector $\bar{\xi}$ is obtained after w iterations.

3.3 Parameter selection

There are three tuning parameters λ, α and β in joint sparse least-square model. The parameter λ plays an unimportant role and will not work on the continuous operation of the algorithm, since we adopt the threshold algorithm in the sparsification step and the l_1 - norm penalty sparse regression problem does not rely on it theoretically. Therefore, the remaining two parameters α and β need to be adjusted. The former controls the number of selected polynomial terms from candidate basis function. The latter determines how similar the derived coefficient values of the selected polynomial terms are. In this work, we utilize a two-step parameter selection method.

In the sparsification step, we note that the model complexity (defined by the number of nonzero coefficients in sparse solution vectors) reduces and the model accuracy (residual norm) increases as the increase of the threshold parameter α, and thus, the results may lose fidelity. In contrast, for a small α, the

model accuracy will be very small, leading to an overfitting solution in practice. Therefore, a proper α should be based on the trade-off between data fidelity and solution sparsity, making both norms small simultaneously. The Pareto method has been shown to outperform cross-validation for sparse approximation [30]. Motivated by the previous study [31], we employ the linear scale instead of log–log to show the solution. The curves of the residual norm and solution norm with change of the threshold parameter α are used to determine the optimal α according to the L-curve criterion, wherein we utilize one state measurement ($L = 1$) and its derivative and set $\beta = 0$. This process will be represented in the following numerical examples.

In the fusion step, the joint sparse least-square algorithm is fitted with fixed α from the sparsification step and a candidate set β. It is extremely time-consuming by blind grid search. We employ some tricks to accelerate the tuning procedure. The major difference between the traditional sparse least squares

and joint sparse least squares is the last penalty term. On the one hand, JSLSGF will reduce to STLS when β is too small, i.e., $\beta < \min_j |(\Phi_l^M(x))^T \dot{x}_l|$. On the other hand, JSLSGF will over-fuse the result if β is very large, i.e., $\beta > \max_j |(\Phi_l^M(x))^T \dot{x}_l|$, which implies all the sparse vectors are identical to each other. Thus, it is reasonable that the parameter range that is neither too large nor too small is used in the simulation experiment. We will conduct a series of experiments to display the identification performance of the system coefficients with proper β.

3.4 Numerical computation of state time derivatives

In many practical applications, the state trajectory x can be available, and its derivative $\dot{x}$ can be usually obtained by the finite difference method for ordinary differential equations (ODEs). To mitigate the error caused by differentiation, we employ the total variation regularization method [32] to compute the derivative of the ODEs, which can de-noise the derivative, and retain more edge information as well as estimate more robust derivative [33].

4 Numerical examples

In the simulation study, to evaluate the performance of the proposed model, we mainly tend to investigate the potential power whether joint sparse least squares by generalized fused lasso penalty can boost the performance of STLS from our motivation or not. Therefore, it is reasonable to use STLS as benchmark for comparison. For STLS, we estimate the coefficient values of the system individually on each state measurement of every case and then compute their arithmetic mean as the final coefficient values, which is named as STLS (Separate) that is the separate sparse least squares in Fig. 1. The other refers to STLS (Combined), which identifies coefficient values of the system when all state measurements with different noise levels are combined into one data set.

We exhibit three specific numerical examples to explore the validity and serviceability of JSLSGF for identifying dynamical system. To study the performance of the proposed method in different scenarios, several cases are conducted to estimate the effect of the fusion for every example. We simulate two

datasets ($L = 2$) with different noise levels as every experimental case in this work to see the effect of the proposed model on identification performance. Supposing that we have no prior knowledge of the governing equation that generates the data, excepting that they can be sparsely expressed by a multivariate polynomial basis of the state variable with the known dimension. The exact state variables (numerical solution) of ODEs are computed by the fourth-order Runge–Kutta method under the tolerance 10^{-12}, and each Δt time unit at discrete time $t_k = k\Delta t\,(\Delta t = 0.001)$ is sampled. The number of simulated samples is considered to catch the fundamental behavior of every dynamical system. The independent zero mean Gaussian white noise with different variance σ^2 is contaminated to the exact state variables to obtain the state measurements of different noise levels, which is used to test the robustness of the proposed model. The noisy formula

$$x_v^*(t_k) = x_v(t_k) + \eta_v(t_k), \tag{13}$$

where $x_v^*(t_k)$, $v = 1, \ldots, n$ and $x_v(t_k)$, $v = 1, \ldots, n$ are the corrupted and exact state variables at times t_k, $k = 1, \ldots, m$, respectively, and v is the state variable as well as $\eta_v(t_k) \sim N(0, \sigma)$. We approximate two ODEs by using the multivariate monomial basis function library $\Phi(x)$ to form candidate nonlinear function [16], where the total degree of the basis is one more than the highest degree in the governing equation.

To alleviate the error of the numerical differentiation, mentioned in Sect. 3.4, may yield wrong results close to the function boundary, and that the time derivatives of the state are unknown at the end of the training time, we therefore compute the time derivatives of the state within a time interval (from both sides) 5% larger than the required time range for experiment, but merely utilize the state measurement and estimated time derivative within the original required time to calculate solution vector. To evaluate the quality of the derivative, we define the relative error

$$\varepsilon_{\dot{x}} = \frac{\|\dot{x}^* - \dot{x}\|_2}{\|\dot{x}\|_2}, \tag{14}$$

where $\dot{x}^*$ and $\dot{x}$ denote the computed and true state derivative values, respectively. To ensure the fairness of the comparison of numerical experimental results,

we separately calculate the derivative of each noisy state measurement with different noise levels, so that the used data are same when STLS (Separate and Combined) and our proposed method are executed.

The nonlinear dynamical systems considered are as follows: the Lorenz 63 system as a model to identify chaotic dynamics, the Duffing oscillator as nonlinear stiffness and damping model and Burgers' equation as a partial differential equation. In the next section, we will explore how the proposed model can further improve the dynamical system identification performance by using the similarity information of two state measurement sources.

4.1 Lorenz 63 system

We consider a well-known nonlinear chaotic Lorenz system [34] that was developed when Lorenz studied atmospheric motion by simplifying the convection model. Its state trajectories are chaotic, aperiodic and confined to a butterfly shaped attractor. The governing equation of Lorenz system is:

$$\begin{cases} \dot{x} = \beta(y - x), \\ \dot{y} = x(\rho - z) - y, \\ \dot{z} = xy - \mu z, \end{cases} \quad (15)$$

where the coefficient values are $\beta = 10$, $\rho = 28$, $\mu = 8/3$, and the initial value is $(x_0, y_0, z_0) = (-8, 7, 27)$. The highest order of the state variables is square on the right-hand side of Eq. (15), and the basis function $\Phi(x) = [1 \ x \ y \ z \ x^2 \ xy \ xz \ y^2 \ yz \ z^2 \ x^3 \ x^2y \ x^2z \ xy^2 \ xyz \ xz^2 \ y^3 \ y^2z \ yz^2 \ z^3]$ is constructed.

The state trajectories of the Lorenz system for each state measurement were obtained from $t = 0.001$ to $t = 2.2$ with $\Delta t = 0.001$, yielding $m = 2200$ samples, and we disturb them with different noise levels σ. Truncating the discrete state trajectory and calculating state measurement derivative by the total variation regularized numerical differentiation, and then generating 2000 samples from $t = 0.1$ to $t = 2.1$. We show in Table 1 the relative errors of state derivatives for each state measurement with different noise levels. Observe that the error of derivative is fast increasing around $\sigma = 10^{-2}$, and the error of derivative is quite large when $\sigma = 10^0$. After conducting a lot of simulation experiments, we found that the Lorenz systems identified by all methods failed seriously, while the noise level is $\sigma = 10^0$. Therefore, it is thrown, we give the combination ($L = 2$) of the state measurement with different noise levels in Table 2 for every case, which means that there are 4000 samples for each simulation case. The first two cases with different noise levels, and the difference between them is relatively large. The difference of the noisy levels in the latter two cases is small. The purpose of this setup is to explore the effect and impact of fusion on data with different gap noisy levels.

We take the state measurement and its derivative with noise level $\sigma = 10^{-1}$ in Case 1 and implement L-curve criterion as an example to investigate how the threshold parameter α impacts the accuracy of the result as well as how to select the proper α. The object function (7) is computed for three state variables with an increment $\Delta\alpha = 0.0001$ ranging $0.0001 - 1$, and the residual norm and solution norm are calculated,

Table 1 The relative errors of the Lorenz system state derivatives at different noise levels

Noise levels	$\sigma = 10^{-3}$	$\sigma = 10^{-2}$	$\sigma = 2 \times 10^{-2}$	$\sigma = 10^{-1}$	$\sigma = 2 \times 10^{-1}$	$\sigma = 10^0$
$\varepsilon_{\dot{x}}$	5.2276×10^{-4}	4.8×10^{-3}	10.5×10^{-3}	1.75×10^{-2}	3.67×10^{-2}	1.468×10^{-1}
$\varepsilon_{\dot{y}}$	4.1811×10^{-4}	3.2×10^{-3}	6.6×10^{-3}	1.33×10^{-2}	2.57×10^{-2}	1.321×10^{-1}
$\varepsilon_{\dot{z}}$	3.0761×10^{-4}	2.7×10^{-3}	5.2×10^{-3}	1.14×10^{-2}	1.99×10^{-2}	1.187×10^{-1}

Table 2 Cases consisting of the Lorenz system state with different noise levels

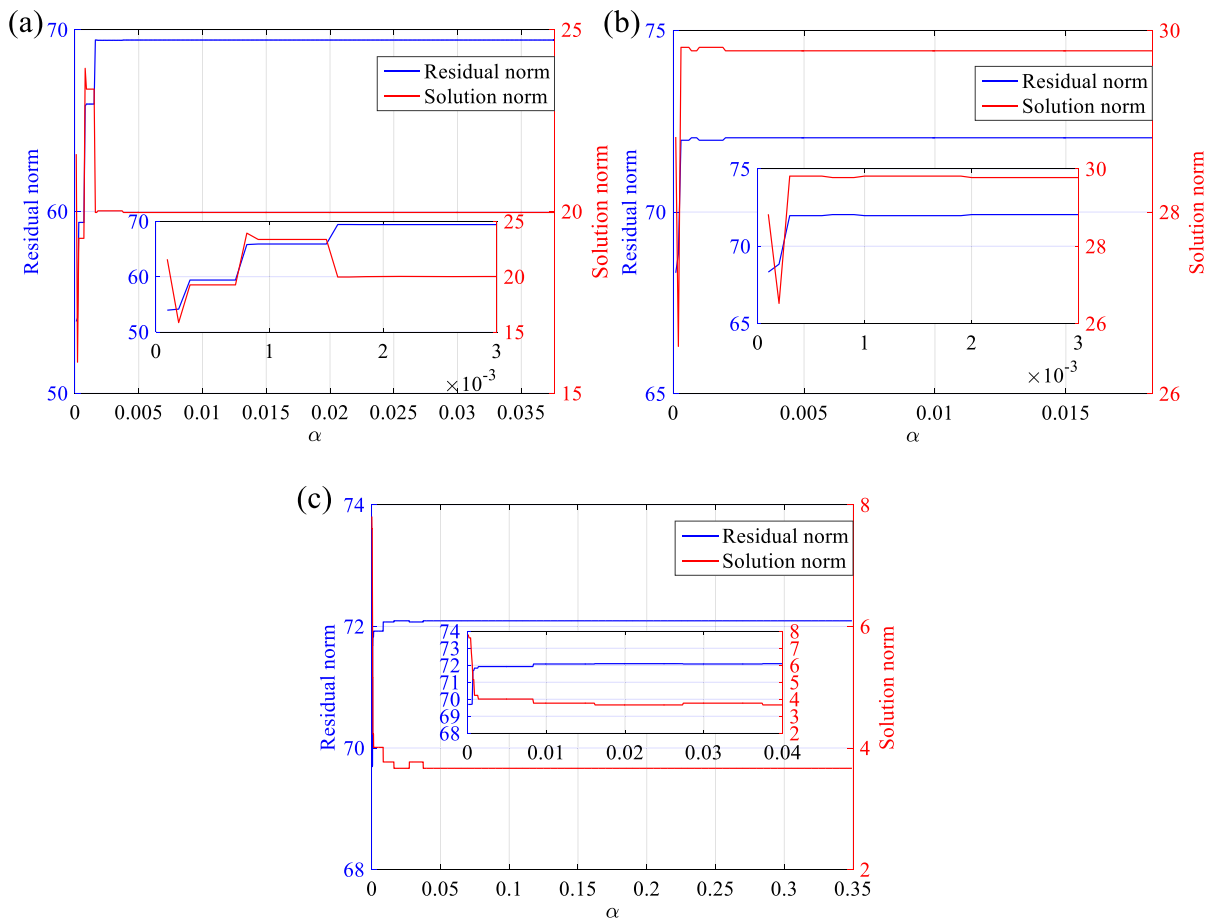| Cases	Case 1	Case 2	Case 3	Case 4
Noise levels	$\sigma = 10^{-3}$	$\sigma = 10^{-2}$	$\sigma = 10^{-2}$	$\sigma = 10^{-1}$
	$\sigma = 10^{-1}$	$\sigma = 10^{-1}$	$\sigma = 2 \times 10^{-2}$	$\sigma = 2 \times 10^{-1}$

Fig. 2 The L-curve for each state of the Lorenz system. **a** State x, **b** state y and **c** state z at noise level $\sigma = 10^{-1}$

respectively. We explain in Fig. 2 the resulting curves versus different α. As mentioned previously, the residual norm increases and solution norm decreases as α increases. According to the L-curve criterion and the results obtained with different α in the sparsification step, for the variables x, y and z, we can observe in Fig. 2a–c that $0.0016 \leq \alpha \leq 0.0376$, $0.002 \leq \alpha \leq 0.0183$ and $0.0375 \leq \alpha \leq 0.3492$ are the approximate parameter value ranges, respectively, meaning that there exist a wide range of α values that produce similar and even same solutions.

Here we empirically analyze the convergence behavior of the algorithm 1 by showing the changes of the absolute sum of the obtained solution vector versus iteration numbers in Fig. 3 for Case 1, clearly indicating that the absolute sum of the solution vector

decreases during iterations and finally converges to a constant value.

To probe and compare whether the fusion item in the proposed method has effect on the identification of system coefficients, we conduct JSLSGF in the presence of the Case 1. The trajectories of exact and learned system up to time-stamp $t = 10$ with three state variables in Fig. 4 are drawn for comparisons in the prediction performance of the identified Lorenz system. Note that the predicted trajectories agree with the exact ones from $t = 0$ to more than $t = 5$. Moreover, in Fig. 5, the short time prediction from $t = 0$ to $t = 3$ (dashed red line) and long-time exact trajectory from $t = 0$ to $t = 16$ (solid blue line) show that the predicted trajectory is consistent with the exact trajectory in a long-time period even the Lorenz system is chaotic.
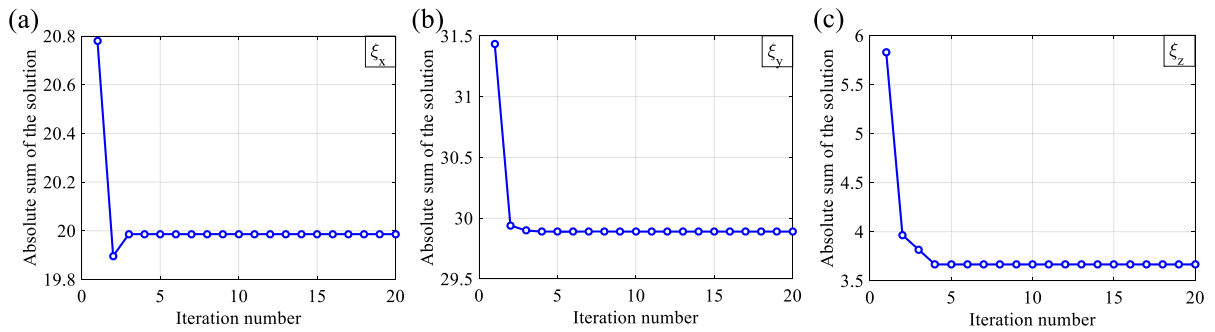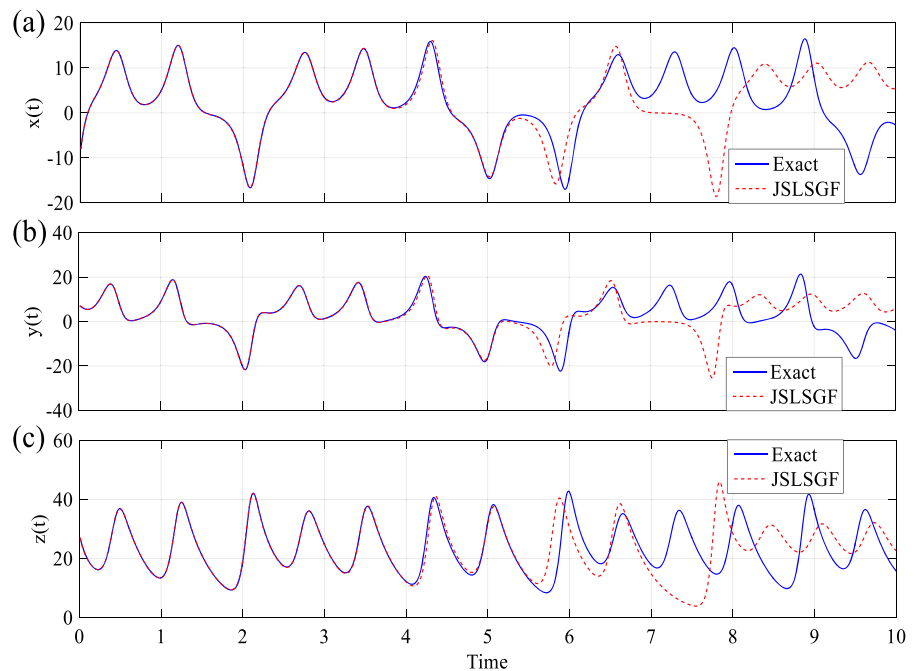
Fig. 3 Convergence curves of our method on **a** the state x, **b** the state y and **c** the state z for Case 1

Fig. 4 Comparison of predicted state trajectories of **a** the $x(t)$, **b** the $y(t)$ and **c** the $z(t)$ of Lorenz system for Case 1

The ability to obtain attractor dynamics of the Lorenz system is more essential than to predict the state trajectory because chaos can cause the system to diverge exponentially while small fluctuations occur in the initial condition or system coefficient values. Since the identified Lorenz system coefficient values of most terms in the basis function are zeros, we only list in Table 3 the nonzero coefficient values of the results obtained from STLS (Separate and Combined)

and JSLSGF at different cases, and the optimal values are shown in boldface.

For the first three cases, we detect that JSLSGF not only recognizes the correct basis function terms, but also the identified coefficient values are closer to the real values. However, under the condition of ensuring that all the terms involved in Lorenz system completely appear in the sparse solution vectors, meaning that the corresponding entries in the sparse vectors are nonzero and the error of the solutions is minimized

simultaneously for Case 4 in Table 2. After conducting many experiments, we derive the identified systems and compute the root-mean-square error (RMSE) of the coefficient values for the identified and true systems as
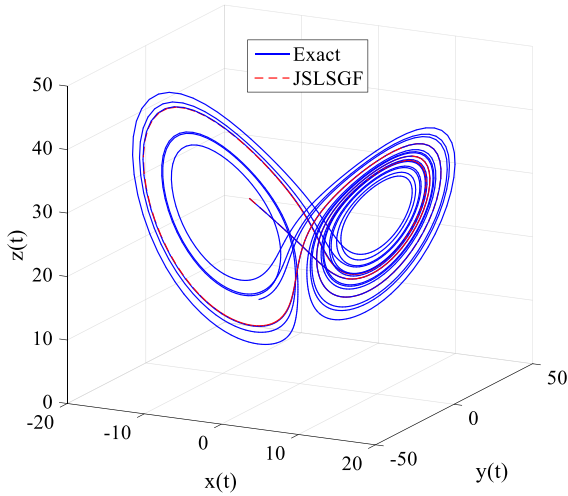

Fig. 5 Exact and identified state trajectories of the Lorenz system for Case 1

STLS (Separate) :
$$\begin{cases} \dot{x} = -9.9766x + 9.976y, \\ \dot{y} = 27.6966x - 0.9117y - 0.993xz, \\ \dot{z} = -2.6649z + 0.999xy, \end{cases}$$

with $\text{RMSE}_{\text{STLS(Separate)}} = 0.1201$;

STLS(Combined):
$$\begin{cases} \dot{x} = -9.9766x + 9.976y, \\ \dot{y} = 27.648x - 0.8939y - 0.9922xz + 0.1205, \\ \dot{z} = -2.6649z + 0.999xy, \end{cases}$$

with $\text{RMSE}_{\text{STLS(Combined)}} = 0.1468$;

and JSLSGF :
$$\begin{cases} \dot{x} = -9.9766x + 9.976y, \\ \dot{y} = 27.7046x - 0.914y - 0.9932xz, \\ \dot{z} = -2.6649z + 0.999xy, \end{cases}$$

with $\text{RMSE}_{\text{JSLSGF}} = 0.117$.

Based on the above discussion and representation, in most cases, JSLSGF identifies items inherent to the

Table 3 The sparse solutions obtained from STLS and JSLSGF for the Lorenz system at different cases that are composed of the state measurement polluted by different noise levels

ξ	Basis	True	Methods	Case 1	Case 2	Case 3
ξ_1	x	-10	STLS (Separate)	-9.9947	-9.9918	-9.9986
			STLS (Combined)	-9.9947	-9.9918	-9.9986
			JSLSGF	$\mathbf{-9.9961}$	$\mathbf{-9.9928}$	$\mathbf{-9.9994}$
	y	10	STLS (Separate)	9.9914	9.9865	9.9956
			STLS (Combined)	9.9913	9.9864	9.9956
			JSLSGF	$\mathbf{9.9918}$	$\mathbf{9.9872}$	$\mathbf{9.9967}$
ξ_2	x	28	STLS (Separate)	27.9114	27.8786	27.955
			STLS (Combined)	27.9112	27.8785	27.955
			JSLSGF	$\mathbf{27.918}$	27.8785	27.955
	y	-1	STLS (Separate)	-0.976	-0.971	-0.9916
			STLS (Combined)	-0.9759	-0.9709	-0.9916
			JSLSGF	$\mathbf{-0.9787}$	-0.971	-0.9916
	xz	-1	STLS (Separate)	-0.9978	-0.9968	-0.9988
			STLS (Combined)	-0.9978	-0.9968	-0.9988
			JSLSGF	$\mathbf{-0.9979}$	-0.9968	-0.9988
ξ_3	z	$-8/3$	STLS (Separate)	-2.6664	-2.6659	-2.6658
			STLS (Combined)	-2.6665	-2.6659	-2.6658
			JSLSGF	-2.6665	$\mathbf{-2.6661}$	$\mathbf{-2.6659}$
	xy	1	STLS (Separate)	0.9998	0.9994	0.9996
			STLS (Combined)	0.9998	0.9994	0.9996
			JSLSGF	0.9998	$\mathbf{0.9995}$	0.9996

Table 4 The relative errors of the Duffing oscillator state derivatives at different noise levels

σ	$\sigma = 10^{-5}$	$\sigma = 10^{-4}$	$\sigma = 2 \times 10^{-4}$	$\sigma = 10^{-3}$	$\sigma = 1.5 \times 10^{-3}$	$\sigma = 10^{-2}$
$e_{\dot{x}}$	1.5263×10^{-4}	6.2234×10^{-4}	1.3×10^{-3}	7.4×10^{-3}	9.9×10^{-3}	3.21×10^{-2}
$e_{\dot{y}}$	1.8614×10^{-4}	7.0878×10^{-4}	1.5×10^{-3}	3.1×10^{-3}	4.5×10^{-3}	2.11×10^{-2}

Table 5 Cases consisting of the Duffing oscillator state with different noise levels

Cases	Case 1	Case 2	Case 3	Case 4
Noise levels	$\sigma = 10^{-5}$	$\sigma = 10^{-4}$	$\sigma = 10^{-4}$	$\sigma = 10^{-3}$
	$\sigma = 10^{-3}$	$\sigma = 10^{-3}$	$\sigma = 2 \times 10^{-4}$	$\sigma = 1.5 \times 10^{-3}$

system from the two datasets, that is, the similarity information of the datasets is mined. For different state measurements, with a small noisy difference between the latter two kinds of cases, the fusion plays a relatively small role. The underlying reason for this condition could be that the difference of measurement data is smaller when the difference of noise level is small. However, for all cases, we obtain that JSLSGF has an essential advantage than STLS, and the STLS (Combined) performs even worse than STLS (Separate), especially for Case 4, the identified system via STLS (Combined) introduces an extra constant term 0.1205. These results demonstrate that JSLSGF with the proper fusion could hold the best combination of multiple state measurements to increase the identification accuracy.

4.2 Duffing oscillator

As the second example, the classical Duffing oscillator describes forced vibration and presents cubic nonlinearity and chaotic behavior. The motion equation of the unforced Duffing oscillator is:

$$\ddot{x} + \beta \dot{x} + \rho x + \mu x^3 = 0, \tag{16}$$

which can be rewritten into the first-order system of two variables by setting $\dot{x} = y$, we have

$$\begin{cases} \dot{x} = y, \\ \dot{y} = -\beta y - \rho x - \mu x^3, \end{cases} \tag{17}$$

where $\beta = 0.1$, $\rho = 1$ and $\mu = 5$, and the initial condition is $(x_0, y_0) = (1, 0)$. The Duffing oscillator

does not show chaotic behaviors for these values. The highest order on the state variables of (17) is cube; we establish the basis function $\Phi(x) = [1\ x\ y\ x^2\ xy\ y^2\ x^3\ x^2y\ xy^2\ y^3\ x^4\ x^3y\ x^2y^2\ xy^3\ y^4]$. The system, in fact, only is expressed by four terms of them. Like the Lorenz system, the number of samples is 2200 for every state measurement with different noise levels σ in each case, ranging from $t = 0.001$ to $t = 2.2$ with $\Delta t = 0.001$. The derivatives of the noisy state measurements are calculated by the total variation regularized numerical differentiation. We truncate the noisy state measurements and compute derivatives, gaining 2000 samples ranging from $t = 0.1$ to $t = 2.1$. The relative errors of state derivatives for each state measurement with different noise levels are given in Table 4. We found that the error of derivative is quickly increasing around $\sigma = 10^{-3}$, and the error is large at $\sigma = 10^{-2}$. After conducting numerous simulation experiments, we discover that all methods fail to recover the equations at the noise level $\sigma = 10^{-2}$. Therefore, we give the combination ($L = 2$) of the state measurement with different noise levels in Table 5 for each case without including $\sigma = 10^{-2}$, signifying that 4000 samples are contained in each case. The noise level in the first two cases is quite different. The difference of noise level in the latter two cases is small.

We adopt the state measurement and its derivative with noise level $\sigma = 10^{-3}$ as input. For two state variables in the range $0.0001 - 1$ with $\Delta \alpha = 0.0001$, we illustrate in Fig. 6 the resulting curves versus different α. Based on the L-curve criterion, for the
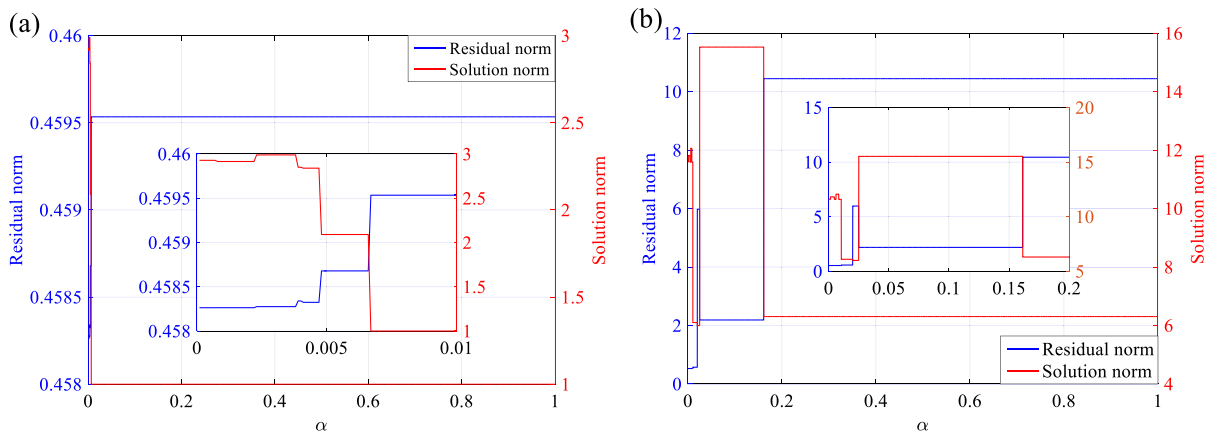
Fig. 6 The L-curves for each state of the Duffing oscillator. **a** State x and **b** state y at noise level $\sigma = 10^{-3}$
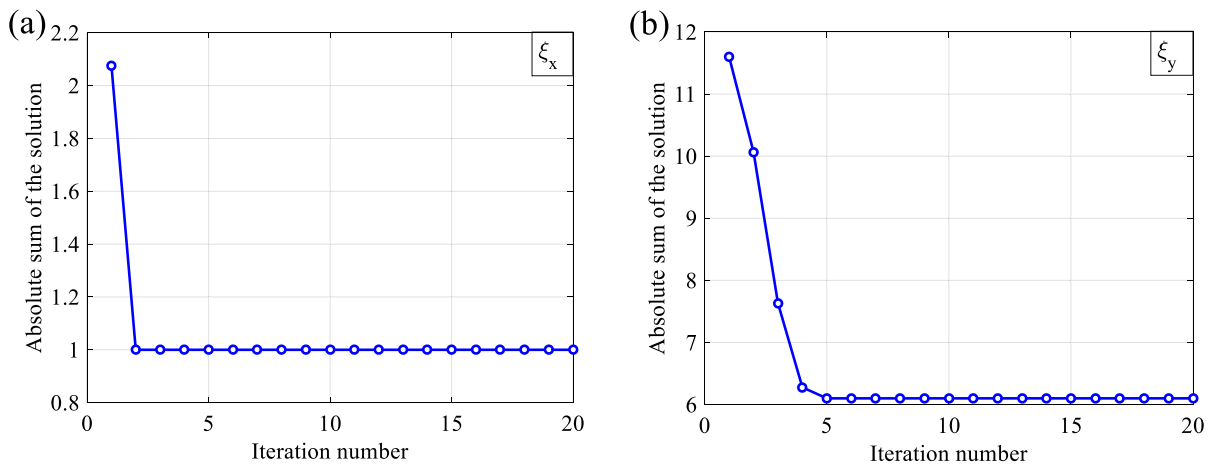

Fig. 7 Convergence curves of our method on **a** the state x and **b** the state y for Case 1

variables x and y, we acquire in Fig. 6a, b that $0.0024 \leq \alpha \leq 0.1613$ and $0.0106 \leq \alpha \leq 0.02$ are the optimal parameter value ranges, respectively, implying that there is a large range of α will generate similar and even identical solutions.

We now investigate the convergence behavior of the algorithm 1. As we can observe from Fig. 7, our proposed optimization algorithm has a good convergence because it basically converges after five iterations.

To probe and compare whether the additional fusion item in the proposed method has effect on the system identification, we conduct JSLSGF in the presence of the Case 1. The trajectories of exact and learned systems up to time-stamp $t = 16$ with two state variables are drawn in Fig. 8a, b. Note that the

predicted trajectories agree with the exact ones for a long time. Moreover, in Fig. 8c, the prediction from $t = 0$ to $t = 20$ (dashed red line) and exact trajectory from $t = 0$ to $t = 50$ (solid blue line) describe that the exact and learned trajectories agree in a relatively long-time period.

We show the results of STLS and JSLSGF on simulated data for every case in Table 6 and just give in Table 6 the approximate coefficient values of the nonzero terms that we derived by using methods STLS and JSLSGF, wherein the best values are highlighted by the boldface.

As expected, in all cases, three methods pick out the correct terms that describe the Duffing oscillator. The estimated values of the proposed JSLSGF are basically consistent with that of the true coefficient values,
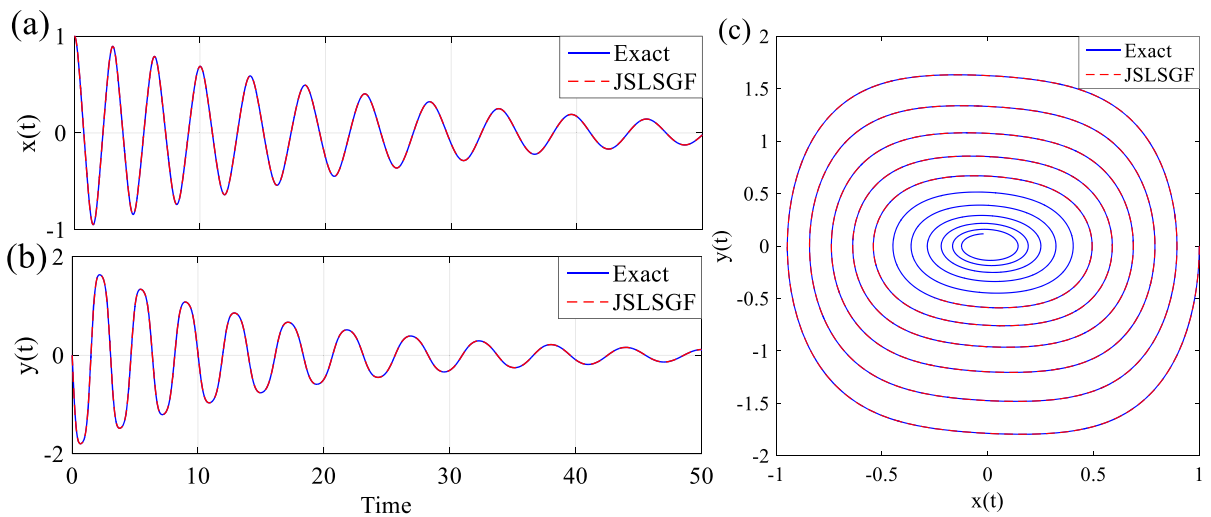
Fig. 8 Comparison of predicted state trajectories of **a** the $x(t)$ and **b** the $y(t)$ of Duffing oscillator for Case 1. **c** Exact and identified trajectories of Duffing oscillator for Case 1

Table 6 The sparse solutions obtained from STLS and JSLSGF for the Duffing oscillator at different cases composed of the state measurement polluted by different noise levels

ξ	Basis	True	Methods	Case 1	Case 2	Case 3	Case 4
ξ_1	y	1	STLS (Separate)	1.00005	1.00005	1	1.0002
			STLS (Combined)	1	1	1	1.0002
			JSLSGF	1	1	1	1.0002
ξ_2	x	-1	STLS (Separate)	-1.00025	-1.00025	-1.0001	-1.00085
			STLS (Combined)	-1.0002	-1.0003	-1.0001	-1.0008
			JSLSGF	$-\mathbf{0.9999}$	$-\mathbf{1}$	-1.0001	$-\mathbf{1.0002}$
	y	-0.1	STLS (Separate)	-0.10005	-0.10005	-0.1	-0.1
			STLS (Combined)	-0.1	-0.1	-0.1	-0.1
			JSLSGF	-0.1	-0.1	-0.1	-0.1
	x^3	-5	STLS (Separate)	-4.9995	4.99945	-4.99965	4.9975
			STLS (Combined)	-4.9995	-4.9994	-4.9996	-4.9975
			JSLSGF	$-\mathbf{5}$	$-\mathbf{4.9998}$	-4.9996	$-\mathbf{4.9984}$

meaning that JSLSGF performs better than competing methods STLS (Separate) and STLS (Combined) in identifying coefficient values of the system. All results indicate that JSLSGF could effectively use more information from multiple state measurements and achieve superior performance in identification accuracy of the dynamical system.

4.3 Burgers' equation

Many systems of interest are governed by partial differential equations (PDEs) [35]. Here we generalize our proposed method to the PDEs and only consider using combinations of monomials to express state

variables and partial derivatives of the PDEs in this work, since these are the common terms in physics. The general form of the PDEs is:

$$u_t = N(1, u, u_x, u_{xx}, \ldots, \xi), \ t \in [0, T], \tag{18}$$

where the subscripts denote the partial differentiation with respect to time or space, and $N(\bullet)$ describes the evolution of the system that is generally a nonlinear function of $u(x, t)$ and coefficients in ξ. As a classic example, the Burgers' equation is derived from the Navier–Stokes equations for the velocity field. The Burgers' equation is given as:

$$u_t = \beta u u_x + \rho u_{xx}, \tag{19}$$

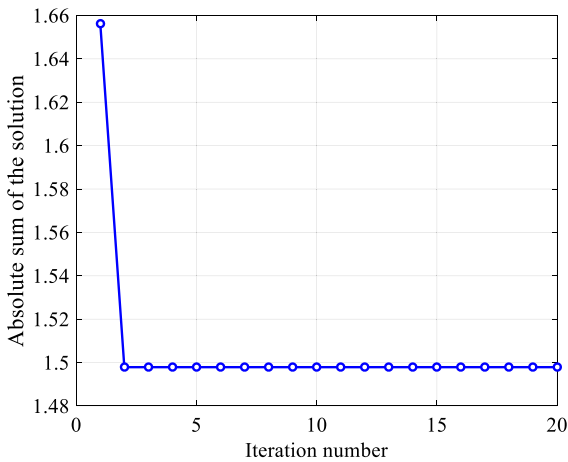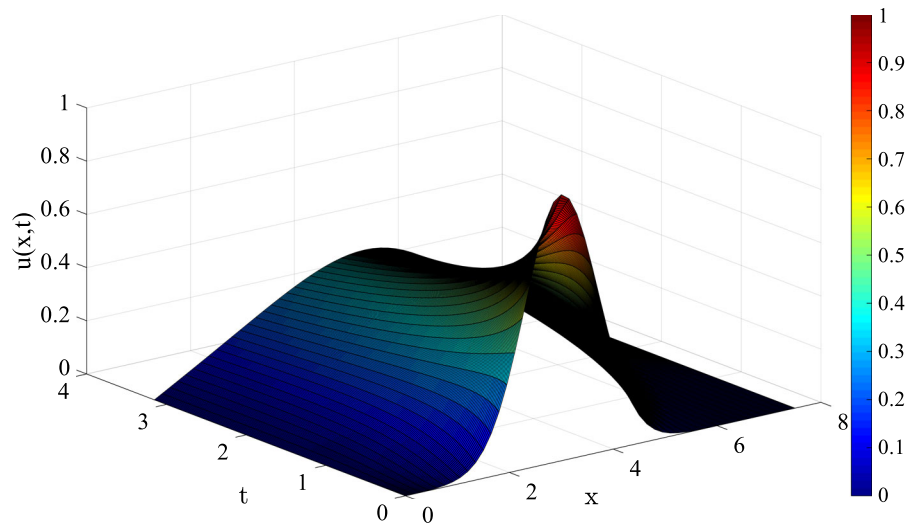**Fig. 9** The numerical solution of the Burgers' equation with $\beta = -1$ and $\rho = 0.5$

Fig. 10 Convergence curve of our method for Case 1

Table 7 Cases for numerical solution of the Burgers' equation with different noise levels

Cases	Case 1	Case 2	Case 3
Noise levels	$\sigma = 10^{-5}$	$\sigma = 10^{-4}$	$\sigma = 10^{-3}$
	$\sigma = 10^{-3}$	$\sigma = 10^{-3}$	$\sigma = 1.13 \times 10^{-3}$

($L = 2$) of the numerical solution for the Burgers' equation with different noise levels in Table 7.

In Fig. 10, we show the convergence curve during the experiment to intuitively illustrate the convergence property of the designed algorithm. One can see that the absolute sum of the obtained solution vector decreases rapidly within three iterations, illustrating the fast convergence of the algorithm 1.

Next, the identified results are presented in Table 8 to investigate the performance of JSLSGF for identifying PDEs. We observe that our proposed method significantly outperforms STLS (Separate) and STLS (Combined), and learn that STLS (Combined) is even worse than STLS (Separate). Note that though all the terms of the Burgers' equation are identified by STLS (Separate), the coefficient errors for Case 1 and Case 2 are relatively high. However, for Case 3, STLS (Separate) and STLS (Combined) introduce more non-contributing terms than JSLSGF, and the coefficient errors of the selected contributing terms are larger.

where coefficients $\beta = -1$, $\rho = 0.5$. Set the initial condition $u(x, 0) = e^{-(x-3)^2}$ and the Dirichlet boundary condition $u(0, t) = u(7.5, t) = 0$. The numerical solution is calculated using finite difference method at the grid points on $x \in [0, 7.5]$ and $t \in [0, 3.15]$ with $\Delta x = 0.15$ and $\Delta t = 0.015$, see Fig. 9. As for the partial derivatives of time, we use the points from two adjacent times and calculate partial derivatives with respect to space using central-difference formula. Assume that 5 data points at both ends of the time and space intervals are removed. The basis function $\Phi(u) = [1 \ u \ u_x \ u^2 \ uu_x \ u_{xx} \ u^3 \ uu_{xx} \ u^2u_x \ u_{xxx}]$ is constructed, there are 10 terms, only two of which are contributing. We provide different combinations

Table 8 The identified PDEs from STLS and JSLSGF at different cases composed of the numerical solution polluted with different noise levels

| Cases	Methods	Identified Burgers' equation (True: $u_t = -uu_x + 0.5u_{xx}$)
Case 1	STLS (Separate)	$u_t = -1.0006uu_x + 0.5022u_{xx}$
	STLS (Combined)	$u_t = -0.2043uu_x + 0.5049u_{xx} - 1.217u^2u_x$
	JSLSGF	$u_t = -uu_x + 0.5007u_{xx}$
Case 2	STLS (Separate)	$u_t = -1.0005uu_x + 0.5022u_{xx}$
	STLS (Combined)	$u_t = -0.204uu_x + 0.5049u_{xx} - 1.2175u^2u_x$
	JSLSGF	$u_t = -uu_x + 0.5009u_{xx}$
Case 3	STLS (Separate)	$u_t = 12.4465 - 32.22705u - 0.218u_x - 0.4953uu_x + 0.272u_{xx}$
	STLS (Combined)	$u_t = 23.3805 - 64.6603u - 0.4125u_x - 0.0114uu_x + 0.0402u_{xx}$
	JSLSGF	$u_t = 8.8834 - 32.7677u - 1.0052uu_x - 0.3329u_{xx}$
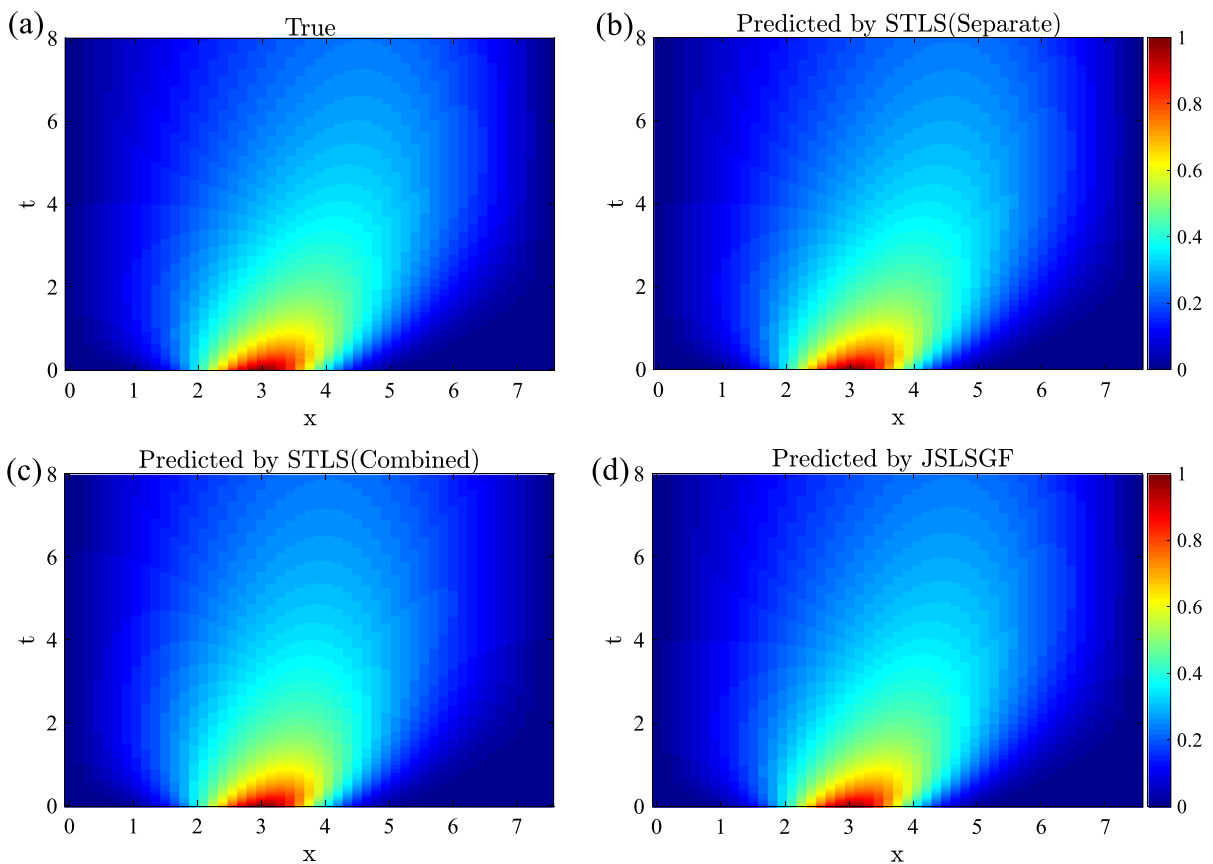
Fig. 11 The solution of the Burgers' equation plotted in space–time: **a** the true numerical solution, **b** predicted by STLS (Separate), **c** predicted by STLS (Combined); and **d** predicted by JSLSGF

Now, we solve the PDEs on $x \in [0, 7.5]$ and $t \in [0, 8]$ using the coefficients identified by all methods for Case 2 to study predictive performance of our method, and the results are shown in Fig. 11. As a comparison, the Burgers' equation is computed by Eq. (19) with $\beta = -1$ and $\rho = 0.5$ as illustrated in Fig. 11a. We know that Fig. 11c deviates greatly from the true equation. Figure 11b, d shows obvious differences on $x \in [5, 6]$ and around $t = 4$, while Fig. 11d is closer to the true Burgers' equation. These highlight the broad applicability of our proposed method and the better performance in identifying PDEs.

5 Conclusion

In this work, we have proposed a joint sparse least-square model with generalized fused lasso penalty, which can make full use of more similarity information among different state measurements, thereby boosting the performance of system identification. We have also presented an efficient threshold joint sparse least-square algorithm, wherein the selection of the threshold parameter in the sparsification step is achieved by using the L-curve criterion. In addition, we have studied the performance of the proposed method on two ODEs and generalized the proposed method to a PDE with a series of cases. The numerical results demonstrate that the appropriate fusion of multiple state measurements can improve the identification accuracy of the systems than traditional sparse least-square model. Thus, the proposed method identifies more accurate governing equations and exploits the richness of the data when explicitly accounting for the similarity information of multiple datasets, which means that it has a greater potential applicability for dynamical system identification problems that require higher accuracy.

As part of future work, we will investigate how to find a more extensive basis function search space to further promote the power for identifying complicated dynamical systems. Furthermore, it is an exciting avenue of future research to extend this work to dynamical system identification problems with multiple data sources.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Aguirre, L.A., Christophe, L.: Modeling nonlinear dynamics and chaos: a review. Math. Probl. Eng. **2009**(1024–123X), 238960 (2014)
2. Ljung, L.: Perspectives on system identification. Annu. Rev. Control. **34**(1), 1–12 (2010)
3. Judd, K., Mees, A.: On selecting models for nonlinear time-series. Physica D **82**(4), 426–444 (1995)
4. Kashyap, R.: System identification. IEEE Trans. Autom. Control **18**(1), 85–86 (1973)
5. Kolassa, S.: Statistical learning with sparsity. The lasso and generalizations. Int. J. Forecast. **33**(3), 743–744 (2017)
6. Abel, M.: Nonparametric modeling and spatiotemporal dynamical systems. Int J Bifurc. Chaos **14**(6), 2027–2039 (2004)
7. Abel, M., Ahnert, K., Kurths, J., Mandelj, S.: Additive nonparametric reconstruction of dynamical systems from time series. Phys. Rev. E **71**(1), 015203 (2004)
8. Schmidt, M., Lipson, H.: Distilling free-form natural laws from experimental data. Science **324**(5923), 81–85 (2009)
9. Rey, D., Eldridge, M., Kostuk, M., Abarbanel, H., Schumann-Bischoff, J., Parlitz, U.: Accurate state and parameter estimation in nonlinear systems with sparse observations. Phys. Lett. A **378**(11–12), 869–873 (2014)
10. Voss, H.U., Kolodner, P., Abel, M., Kurths, J.: Amplitude equations from spatiotemporal binary-fluid convection data. Phys. Rev. Lett. **83**(17), 3422–3425 (1999)
11. Bongard, J., Lipson, H.: From the cover: automated reverse engineering of nonlinear dynamical systems. Proc. Natl. Acad. Sci. **104**(24), 9943–9948 (2007)
12. Schmidt, M., Lipson, H.: Symbolic Regression of Implicit Equations. Springer (2009)
13. Schmidt, M., Vallabhajosyula, R.R., Jenkins, J.W., Hood, J.E., Soni, A.S., Wikswo, J.P., Lipson, H.: Automated refinement and inference of analytical models for metabolic networks. Phys. Biol. **8**(5), 055011 (2011)
14. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B (Methodol.) **58**(1), 267–288 (1996)
15. Wang, W.X., Yang, R., Lai, Y.C., Kovanis, V., Grebogi, C.: Predicting catastrophes in nonlinear dynamical systems by compressive sensing. Phys. Rev. Lett. **106**(15), 154101 (2011)
16. Brunton, S.L., Proctor, J.L., Kutz, N.J.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc. Natl. Acad. Sci. **113**(15), 3932–3937 (2016)
17. Brunton, S.L., Xing, F., Kutz, J.N.: Self-tuning fiber lasers. IEEE J. Sel. Top. Quantum Electron. **20**(5), 464–471 (2016)
18. Mangan, N.M., Brunton, S.L., Proctor, J.L., Kutz, J.N.: Inferring biological networks by sparse identification of nonlinear dynamics. IEEE Trans. Mol. Biol. Multi-Scale Commun. **2**(1), 52–63 (2016)
19. Schaeffer, H., Mccalla, S.G.: Sparse model selection via integral terms. Phys. Rev. E **96**(2), 023302 (2017)
20. Tran, G., Ward, R.: Exact recovery of chaotic systems from highly corrupted data. Multiscale Model. Simul. **15**(3), 1108–1129 (2017)
21. Rudy, S.H., Brunton, S.L., Proctor, J.L., Kutz, J.N.: Data-driven discovery of partial differential equations. Sci. Adv. **3**(4), e1602614 (2017)
22. Schaeffer, H.: Learning partial differential equations via data discovery and sparse optimization. Proc. R. Soc. Lond. **473**(2197), 20160446 (2017)
23. Danaher, P., Wang, P., Witten, D.M.: The joint graphical lasso for inverse covariance estimation across multiple classes. J. R. Stat. Soc. Ser. B Stat. Methodol. **76**(2), 373–397 (2014)

24. Tian, Y., Zhang, B., Hoffman, E.P., Clarke, R., Shih, I.M., Xuan, J.H., Herrington, D.M.: Knowledge-fused differential dependency network models for detecting significant rewiring in biological networks. BMC Syst. Biol. **8**, 87 (2014)

25. Yang, S., Lu, Z., Shen, X., Wonka, P., Ye, J.: Fused multiple graphical lasso. SIAM J. Optim. **25**(2), 916–943 (2012)

26. Zhang, L., Schaeffer, H.: On the convergence of the SINDy algorithm. Multiscale Model. Simul. **17**(3), 948–972 (2018)

27. Guckenheimer, J., Holmes, P.: Nonlinear oscillations, dynamical systems and bifurcations of vector fields/John Guckenheimer, Philip Holmes. SIAM Rev. **26**(4), 600–601 (1984)

28. Fang, J., Lin, D., Schulz, S.C., Xu, Z., Calhoun, V.D., Wang, Y.P.: Joint sparse canonical correlation analysis for detecting differential imaging genetics modules. Bioinformatics **32**(22), 3480–3488 (2016)

29. Hoefling, H.: A path algorithm for the fused lasso signal approximator. J. Comput. Graph. Stat. **19**(4), 984–1006 (2010)

30. Tehrani, J.N., McEwan, A., Jin, C., Van Schaik, A.: L1 regularization method in electrical impedance tomography by using the L1-curve (Pareto Frontier Curve). Appl. Math. Model. **36**(3), 1095–1105 (2012)

31. Yao, H., Gerstoft, P., Shearer, P.M., Mecklenbruker, C.: Compressive sensing of the Tohoku-Oki Mw 9.0 earthquake: frequency-dependent rupture modes. Geophys. Res. Lett. **38**(20), L20310 (2011)

32. Leonid, R., Stanley, O., Emad, F.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1–4), 259–268 (1992)

33. Chartrand, R.: Numerical differentiation of noisy, nonsmooth data. ISRN Appl. Math. **2011**, 1023–1033 (2011)

34. Lorenz, E.N.: Deterministic nonperiodic flow. J. Atmos. Sci. **20**(2), 130–141 (1963)

35. Schaeffer, H., Caflisch, R., Hauck, C.D., Osher, S.: Sparse dynamics for partial differential equations. Proc. Natl. Acad. Sci. **110**(17), 6634–6639 (2013)