Sparse optimization assisted hybrid data driven modeling of process systems

Abhishek Raj, [*] **Ankit Ajay Mandpe,** [*] **Santhosh Kumar Varanasi** [*]

[*] *Indian Institute Technology, Jodhpur, Nagaur Road, Karwar, Jodhpur, Rajasthan - 342037, India. (e-mail: m23che002@iitj.ac.in, mandpe.1@iitj.ac.in, skvaranasi@iitj.ac.in).*

Abstract: Digitization, which involves the adoption of various domain-relevant technologies to create a digital equivalent of physical assets, is the main principle of Industry 4.0. As most process plants such as Waster water treatment process (WWTP) and Post-carbon combustion capture (PCC) process exhibit multi-scale dynamics, identification of models in differential equation form (continuous-time) is advantageous. Further, any underlying physical understanding of the system can be easily captured in this modeling strategy, through appropriate choice of functionality resulting the model Gray-box in nature. Since models in differential equation form are considered, the accuracy of modeling depends on the estimation of derivatives from the sampled data. Therefore, the main objective of this paper is to develop an identification methodology in continuous-time (CT) framework that can capture the physical behavior of the system. To address the issue with the derivative information, the data set is fitted using functions like B-splines subjected to a model-based penalty to ensure that the data fit also satisfies the model of the process. For estimation of a parsimonious model, a sparsity constraint in terms of zero norm on the parameter vector of the model is considered. The efficacy of the method is demonstrated on a Van der Pol oscillator and a Continuous stirred tank reactor (CSTR) system and the results are compared with the existing methods.

Keywords: Continuous Time Models, Sparse Optimization, Industry 4.0, Nonlinear systems, B-splines.

1. INTRODUCTION

With the principle of digitization as primary focus, many service industries started providing digital solutions to the process industries. Some of the digital solutions includes, intelligent tools for predictive maintenance, optimization under uncertainty and constraints, Artificial Intelligence (AI) for real time control, digital twin technology for the processes at different levels of hierarchy and platforms for integrating digital solutions to the industries. Utilization of learning techniques in smart modeling and in advanced control strategies such as Model Predictive Control (MPC) will enable this design, control and optimization of complex process systems in an efficient and smart way for maintaining process safety, efficiency and productivity.

As the models with first principles result in a set of highly complex and non-linear differential algebraic equations, computational cost associated with such models in real-time application of smart modeling is too-costly. Further, the first principles modeling require a thorough understanding of the system, which might not always be feasible. The challenge of computational time with detailed first principles model resulting in a prominent development of data-driven system identification models. Objective of system identification is to provide an accurate future predictions through identifying an optimal structure of model from input-output data. Such models can be identified in discrete-time (DT) (difference equation models) or in continuous-time (CT) (ordinary differential equation models). Identification of discrete-time models is more mature (Ljung, 1999) over its counterpart. For example, while modeling a PCC process, a Non-linear Auto-Regressive with Exogenous inputs (NARX) model in discrete form is considered. Various non-linear functional form ranging from Polynomials to Neural networks have been utilized (Jung and Lee, 2023; Wu et al., 2020; Zhang et al., 2018). While control design of a WWTP process, an LSTM model is considered in Liu et al. (2023). Being a black-box model, a strike of balance between the accuracy and complexity of the model has to be made. Further in the case of sparse/insufficient historical data, development of a black-box model be challenging and may also limits the applicability of the model for control and optimization. Therefore, a hybrid/gray-box model which aims to combine a first-principles knowledge with a black-box modeling of ML has attracted a significant amount of interest in the recent years.

Pioneering the work of sparse identification of nonlinear dynamics (SINDy) for data-driven discovery of dynamics (Brunton et al., 2016; Raissi and Karniadakis, 2018; Raissi et al., 2019) through construction of a library by incorporating the priori knowledge such as conservation laws, a significant research on gray-box modeling is being carried out for various areas in the field of systems engineering. Some of the important areas include Model predictive control (Arnold and King, 2021; Kaiser et al., 2018), fault prognosis of chemical processes (Bhadriraju et al., 2021a,b) and digital twin modeling (Wang et al., 2023). Since not all terms in the library plays a prominent role in defining the dynamics of the system, a sparsity constraint is considered while estimating the parameters of the system and methods like LASSO (Tibshirani, 1996), Sequential thresholded least squares (Brunton et al., 2016) will be used for parameter estimation. Recently, the authors in Chakraborty et al. (2020) utilized genetic algorithms for automatic identification of the gray-box models that can capture the underlying physical, chemical and/or biological mechanisms generating the data.

As most of the underlying physical mechanisms are ordinary/partial differential equations form, equations in such form are considered for modeling. In reality, since only data at finite sample instants is available, estimation of derivatives from sampled data is an important challenge to be addressed as requirement of derivative information is essential for modeling. Most of the aforementioned methods relies on finite difference method for estimation of derivatives which is sensitive to noise in the data. Some methods uses robust derivative estimation strategies with inclusion of regularization's like Total-variation, cubic smoothing spline (Knowles and Renka, 2014; Chartrand, 2011). Although these methods provides relatively reliable estimates of derivatives, similar to the works of Poyton et al. (2006); Varanasi and Jampana (2018), the overall efficacy may be improved with inclusion of model based regularization term.

The primary objective of this paper is to propose a method for data-driven identification of models with inclusion of first principle's information that can handle noise in the data effectively. To address the challenges with the derivative information especially with the noise, the data set is fitted using B-splines subjected to a model-based penalty to ensure that the data fit also satisfies the model of the process. For estimation of a parsimonious model, a sparsity constraint in terms of zero norm on the parameter vector of the model is considered.

The rest of the paper is organized as follows: Section 2 presents the problem statement and explains the details of the SINDy method and the proposed methodology. In Section 3, the efficacy of the proposed method is demonstrated on a Van der Pol oscillator and a Continuous stirred tank reactor (CSTR) system and the results are compared with the existing methods. Finally conclusions are drawn in Section 4.

2. SMART LEARNING FROM DATA

2.1 Problem Statement

A general class of nonlinear processes is expressed by the system of non-linear ordinary differential equations as follows:

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t), \beta) \tag{1}$$

where, β is the vector of parameters that defines the system, $\mathbf{x}(t) = [x_1(t)\ x_2(t)\ \cdots\ x_n(t)] \in \mathbb{R}^n$, $\mathbf{u} = [u_1(t)\ u_2(t)\ \cdots\ u_m(t)] \in \mathbb{R}^m$ denote the state and input vectors respectively. A noise corrupted version of state data $\mathbf{x}(t_k)$ is available at discrete time samples as,

$$\mathbf{y}(t_k) = \mathbf{x}(t_k) + \eta(t_k) \tag{2}$$

where, $\eta(t_k)$ are i.i.d Gaussian random variables. Let $\Theta(\mathbf{x}(t), \mathbf{u}(t))$ denote a library of terms that would describe the dynamics of the system and Σ denote a coefficient vector of the library then $f(\mathbf{x}(t), \mathbf{u}(t), \beta) \approx \Theta(\mathbf{x}(t), \mathbf{u}(t))\,\Sigma$. Since it is challenging to explicitly include the exact dynamics into the library, a variety of functions like polynomials, exponential and so on are included to make the library rich in terms for better approximation. Since not all terms in the library may play significant role in defining the dynamics of the system, the main objective of smart learning is to find the parsimonious/sparse parameter vector Σ from the measured input-output data i.e., $\{(\mathbf{u}(t_k), \mathbf{y}(t_k)), k = 1, 2, \cdots, N\}$.

2.2 Methodology

Identification of sparse parameter vector Σ can be performed using regularization techniques such as LASSO in the context of SINDy (Brunton et al., 2016) as

$$\arg\min_{\Sigma} \|\dot{\mathbf{X}} - \mathbf{A}\Sigma\|_2^2 + \lambda\|\Sigma\|_1 \tag{3}$$

where,

$$\dot{\mathbf{X}} = \begin{bmatrix} \dot{x}_1(t_1) \\ \dot{x}_1(t_2) \\ \vdots \\ \dot{x}_1(t_N) \\ \dot{x}_2(t_1) \\ \dot{x}_2(t_2) \\ \vdots \\ \dot{x}_2(t_N) \\ \dot{x}_n(t_1) \\ \dot{x}_n(t_2) \\ \vdots \\ \dot{x}_n(t_N) \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \Theta(\mathbf{x}(t_1), \mathbf{u}(t_1)) & \mathbf{0} & \cdots & \mathbf{0} \\ \Theta(\mathbf{x}(t_2), \mathbf{u}(t_2)) & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \Theta(\mathbf{x}(t_N), \mathbf{u}(t_N)) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Theta(\mathbf{x}(t_1), \mathbf{u}(t_1)) & \cdots & \mathbf{0} \\ \mathbf{0} & \Theta(\mathbf{x}(t_2), \mathbf{u}(t_2)) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \Theta(\mathbf{x}(t_N), \mathbf{u}(t_N)) & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \Theta(\mathbf{x}(t_1), \mathbf{u}(t_1)) \\ \mathbf{0} & \mathbf{0} & \cdots & \Theta(\mathbf{x}(t_2), \mathbf{u}(t_2)) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Theta(\mathbf{x}(t_N), \mathbf{u}(t_N)) \end{bmatrix} \tag{4}$$

and methods like Sequential thresholded least squares (Brunton et al., 2016) can be used for solving the problem of Eq. (3).

Since only noise corrupted values of $\mathbf{x}(t_k)$ are measured, the main challenge of identification is the estimation of derivatives $\dot{\mathbf{x}}(t_k)$. In traditional approaches, methods like finite difference are used for estimation of derivatives. Since these methods are sensitive to noise in the data, the main idea in this paper is to fit the discrete values of $\mathbf{x}(t_k)$ through the measured data of $\mathbf{y}(t_k)$ with smooth curves such as B-splines. Other global functional approximators such as neural networks can also be used. However, an important property of B-splines is that they allow better local control i.e., if few control points are modified, the splines related to those control points might change but the splines corresponding to unchanged data will remain the same thereby providing better control in approximating functions with sharp changes (Boor, 1978) and hence utilized in this paper. Further, the derivatives can then be computed easily by differentiating the B-splines using a recursive relation thereby improving the computational cost. Further details on construction of B-splines and the derivative estimations are given in Varanasi and Jampana (2018).

As the main goal of identification is to obtain the parameter vector Σ, the coefficients of the B-splines enter the optimization problem as nuisance parameters. Since L_0 regularization promotes better sparsity over the relaxed formulation of L_1, the former is considered as regularization and various sparse optimization algorithms detailed in Beck and Eldar (2013); Tropp and Gilbert (2007) may be used to solve the formulated problem.

Defining $\|\Sigma\|_0 = \#\{i : \Sigma_i \neq 0\}$, i.e. the number of non-zero elements of Σ, the optimization problem for computing the B-spline coefficients (γ) and the model parameters (Σ) can be posed as,

$$(P1) \quad \underset{\gamma, \Sigma, \|\Sigma\|_0 \leq s}{\arg\min} \; f_y + \lambda g$$

where,

$$f_y = \sum_{k=1}^{M} \|\mathbf{y}(t_k) - \hat{\mathbf{x}}(t_k)\|_2^2$$

$$g = \|\dot{\hat{X}} - \hat{\mathbf{A}}\Sigma\|_2^2$$

$$\hat{\mathbf{x}}(t) = \sum_{i=0}^{N_y} \gamma_i \phi_i^{l_y}(t)$$

and

$$\dot{\hat{X}} = \begin{bmatrix} \dot{\hat{x}}_1(t_1) \\ \dot{\hat{x}}_1(t_2) \\ \vdots \\ \dot{\hat{x}}_1(t_N) \\ \dot{\hat{x}}_2(t_1) \\ \dot{\hat{x}}_2(t_2) \\ \vdots \\ \dot{\hat{x}}_2(t_N) \\ \dot{\hat{x}}_n(t_1) \\ \dot{\hat{x}}_n(t_2) \\ \vdots \\ \dot{\hat{x}}_n(t_N) \end{bmatrix}, \hat{\mathbf{A}} = \begin{bmatrix} \Theta(\hat{\mathbf{x}}(t_1), \mathbf{u}(t_1)) & \mathbf{0} & \cdots & \mathbf{0} \\ \Theta(\hat{\mathbf{x}}(t_2), \mathbf{u}(t_2)) & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \Theta(\hat{\mathbf{x}}(t_N), \mathbf{u}(t_N)) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Theta(\hat{\mathbf{x}}(t_1), \mathbf{u}(t_1)) & \cdots & \mathbf{0} \\ \mathbf{0} & \Theta(\hat{\mathbf{x}}(t_2), \mathbf{u}(t_2)) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \Theta(\hat{\mathbf{x}}(t_N), \mathbf{u}(t_N)) & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \Theta(\hat{\mathbf{x}}(t_1), \mathbf{u}(t_1)) \\ \mathbf{0} & \mathbf{0} & \cdots & \Theta(\hat{\mathbf{x}}(t_2), \mathbf{u}(t_2)) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Theta(\hat{\mathbf{x}}(t_N), \mathbf{u}(t_N)) \end{bmatrix} \tag{5}$$

with N_y is the number of knot points, l_y is the order of the spline. The term f_y in the objective function denote the data fit error while the term g denote the model fit error and the term $\sum_{i=0}^{N_y} \gamma_i \phi_i^{l_y}(t)$ denotes the spline fit to the ODE solution ($\mathbf{x}$).

The objective here is to minimize the data fit error while maintaining the constraint that the data should satisfy the model. Since the noise $\eta(t_k)$ is assumed to follow i.i.d. Gaussian, the negative log-likelihood function is used to define the data fit term. Other general noise distributions can also be incorporated by an appropriate definition of f_y.

Since the problem ($P1$) involves a joint minimization with respect to γ and Σ along with a sparsity constraint for a fixed value of λ, in the proposed approach, similar to the idea in Varanasi and Jampana (2018), the optimization problem is split into two steps. In the first step, the problem ($P1$) is minimized only with respect to the nuisance parameters γ for a given Σ and in the second step, the minimization is carried out with respect to Σ for a given γ alongside a sparsity constraint. These two steps are iterated for a given value of λ till the error in Σ between two iterations is within a tolerance range.

Mathematically, the first step of optimization can be written as

$$\gamma^* = \underset{\gamma}{\arg\min} \; f_y + \lambda g \tag{6}$$

Since the optimization problem is nonlinear in nature, gradient descent methods or search based methods can be used to solve the problem. In this paper, fminsearch of MATLAB has been used to solve the minimization problem for estimation of γ and to avoid the convergence to local optima, efficient choice of initial guess has to be considered. Details on the selection of initial guess for better accuracy is given in Section 2.3

The second step of optimization can be mathematically represented as

$$\Sigma^* = \underset{\Sigma, \|\Sigma\|_0 \leq s}{\arg\min} \; f_y + \lambda g$$

Since the first term f_y is independent of Σ, this optimization problem can be modified as

$$\Sigma^* = \underset{\Sigma, \|\Sigma\|_0 \leq s}{\arg\min} \; g \tag{7}$$

Since the function g is a linear function of Σ, standard sparse optimization techniques such as the Orthogonal Matching Pursuit (OMP) or Least Angle Regression (LAR) and its variants (Candes and Tao, 2005; Tropp and Gilbert, 2007; Luu et al., 2015) can be used to solve the optimization problem. In this paper, OMP is used to solve the optimization problem owing to its simplicity in implementation and the steps of OMP is given in Algorithm 1. Utilization of various other algorithms and theoretical analysis of convergence of estimates will be pursued in future.

2.3 Selection of hyperparameters

The efficacy of the algorithm depends on the choice of parameter λ and the choice of initial guess for γ. For λ, following the same approach as in Varanasi and Jampana (2018) and Ramsay et al. (2007), λ value is varied starting from a low value (0.001) to a very high value in orders of 5 or 10. For every value of λ, the optimization (Steps-1 and 2) is performed till convergence and the final solution is considered as initial guess for the next choice of λ. These iterations are repeated until the error in Σ between two consecutive choice of λ is small or the objective function starts to increase after reaching a minimum value.

Now, to obtain an efficient initial choice for γ, the idea of principal differential analysis (PDA) as detailed in Poyton et al. (2006) is considered. Unlike the proposed method wherein a model based regularization is considered, a second order gradient term $\left(\int_0^T \frac{d^2\hat{\mathbf{x}}}{dt^2} dt\right)$ is considered as a regularization term i.e., as g in PDA. Since this term is independent of Σ, PDA is not an iterative approach and one can obtain an explicit solution for γ. The solution with the PDA method is considered as an initial choice of γ in the proposed method.

The overall flowchart of the proposed methodology is given in Algorithm 1.

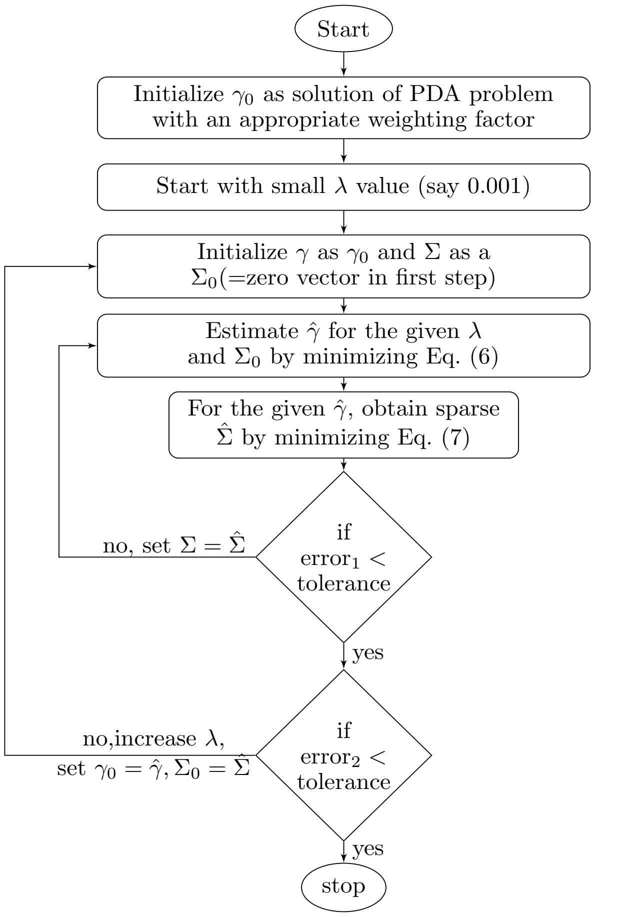

Fig. 1. Overall Algorithm where, error$_1$ denote the root mean squared error of $\hat{\Sigma}$ between two consecutive iterations of step-1 and step-2 and error$_2$ denote the root mean squared error of $\hat{\Sigma}$ between two consecutive iterations of λ.

Algorithm 1 Orthogonal Matching Pursuit

- **Inputs**: Matrix $\hat{\mathbf{A}}$, measurements $\dot{\hat{X}}$ and termination criteria (rms value (ϵ) or the length of the support vector (k)).
- **Initialization:** $\mathbf{r}^0 = \dot{\hat{X}}$ and $\Lambda^0 = \{\}$
- **Main step**: At l^{th} step,
  * Compute error $\xi(j) = \min_{z_j} \|\hat{\mathbf{a}}_j \sigma_j - \mathbf{r}^{l-1}\|, \forall\ j$ using optimal choice, $\sigma_j^* = \hat{\mathbf{a}}_j^T \mathbf{r}^{k-1} / \|\hat{\mathbf{a}}_j\|_2^2$.
  * Identify support as $\hat{\lambda}^l = \min\limits_j \|\xi(j)\|\ \forall\ j \notin \Lambda^{l-1}$.
  * Update support as $\Lambda^{l+1} = \Lambda^l \cup \hat{\lambda}^l$.
  * Estimate solution (Σ_{l+1}) as minimizer of $||\dot{\hat{X}} - \hat{\mathbf{A}}\Sigma||_2^2$ subject to support set, Λ^l.
  * Estimate residual, $\mathbf{r}^{l+1} = \dot{\hat{X}} - \hat{\mathbf{A}}\Sigma_{l+1}$.
  * if $\|\mathbf{r}^l\|_2 < \epsilon$ or length$(\Lambda^l) \leq k$, terminate algorithm or else repeat the steps.

3. NUMERICAL EXPERIMENTS

To demonstrate the efficacy of the proposed method, Van der Pol Oscillator and a Continuous Stirred Tank Reactor systems are considered. The systems are simulated in MATLAB to obtain the noiseless state data i.e., $\{\mathbf{x}(t_k)$, for $k = 1, 2, \cdots, N\}$ and a Gaussian noise with a signal to noise ratio [1] of 20 to 25 is added to the state data to obtain the output data. Monte-carlo simulations are performed with 200 realizations of such noise and the mean and standard deviation of the estimates are reported in this paper. A tolerance value of 1×10^{-3} is considered for both error$_1$ and error$_2$.

3.1 Van der Pol Oscillator

The first system considered is a Van der Pol Oscillator whose dynamics are given below:

$$\frac{dx_1}{dt} = x_2$$
$$\frac{dx_2}{dt} = -x_1 + \mu\left(1 - x_1^2\right)x_2$$

To obtain the measurements of x_1 and x_2, the system is simulated in MATLAB over the time domain of $[0, 15]$ with $\mu = 2$ and initial conditions as $x_1(0) = 1$ and $x_2(0) = 0$. An i.i.d Gaussian noise of 200 realizations with a SNR of 20 is added to the measurements to obtain the output dataset. To model the system using the proposed method, a candidate library with the polynomial terms as given in Eq. (8) are considered.

$$\Theta(x_1, x_2) = \left[1\ x_1\ x_2\ x_1^2\ x_2^2\ x_1^2 x_2\ x_1 x_2^2\right] \tag{8}$$

The estimated parameters (mean and standard deviation) for state x_2 with the proposed method and the SINDy (Brunton et al., 2016) method (with a hyper-parameter value of 0.22) are shown in Table. 1 and the corresponding response with the mean value is shown in Fig. 2.

Table 1. Estimates (mean and standard deviation) of parameters of coefficients of each candidate term in the Library for state x_2 of Van der Pol Oscillator case study.

Candidate term	**True Value**	**Proposed method**	**SINDy**
1	0	0	0
x_1	-1	-1.013±0.046	-0.983±0.037
x_2	2	1.923±0.032	1.8353±0.0711
x_1^2	0	0	0
x_2^2	0	0	0
$x_1^2 x_2$	-2	-1.822±0.076	-1.799±0.068
$x_1 x_2^2$	0	0	0

From the results reported in Table 1 and the corresponding response plot in Fig. 2, it can be concluded that although both the methods were able to identify the non-zero candidate terms accurately, the accuracy of parameter estimation is slightly higher with the proposed method over the SINDy method. This can be attributed to the fact that as finite difference method is employed in SINDy, the derivative estimation might be inaccurate thereby resulting in relatively large error in the estimates.

[1] Defined as $10 \times \log_{10}(\mathrm{P_{signal}}/\mathrm{P_{noise}})$ where $\mathrm{P_{signal}}$, $\mathrm{P_{noise}}$ are the power of the signal and noise respectively; power of a signal is defined as its root mean square value

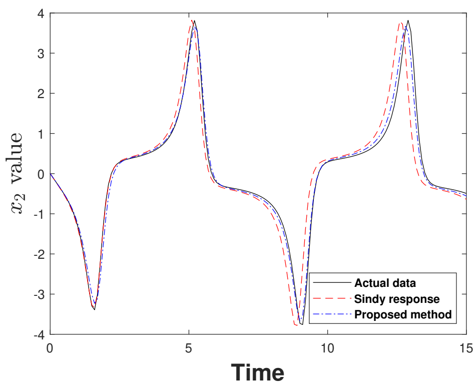

Fig. 2. Mean response of state x_2 with different methods

3.2 CSTR Case Study

In this section, the efficacy of the proposed method is demonstrated using a perfectly mixed non-isothermal CSTR system of volume V in which an irreversible and endothermic chemical reaction of second order dynamics is taking place. The feed with an initial concentration of C_{A_0} enters the reactor with a flow rate F_r and Temperature T_0. The dynamics of such a reactor system is represented using mass and energy balance equations as:

$$\frac{dC_A}{dt} = \frac{F_r}{V}\left(C_{A0} - C_A\right) - K_0 \exp\left(-\frac{E}{RT}\right) C_A^2$$
$$\frac{dT}{dt} = \frac{F_r}{V}\left(T_0 - T\right) + \frac{(-\Delta H_r)}{\rho c_p} K_0 \exp\left(-\frac{E}{RT}\right) C_A^2 + \frac{Q}{\rho c_p V}$$

where, K_0 and E denote the rate constant, activation energy of the reaction respectively and R is the universal gas constant. Q is the heat input to the system and using this variable the reaction mechanism is to be controlled, ΔH_r is the heat of reaction and ρ and c_p are the density and specific heat of the inlet fluid.

To generate the concentration and temperature data, the system is simulated in MATLAB with a random heat input profile in the range of -6×10^4 and 10×10^4. All the other parameters required to simulate the system is given in Table. 2.

Table 2. Parameter values of CSTR

Parameter	Value	Units
F_r	5	m^3/hr
C_{A_0}	4	$kmol/m^3$
T_0	300	K
V	1	m^3
K_0	8.46×10^6	hr^{-1}
E/R	6.014×10^3	k
ΔH_r	1.15×10^4	kJ/kmol
ρ	1×10^3	kg/m^3
c_p	0.231	kJ/kg.K

For the generated concentration and temperature data, an i.i.d Gaussian noise of 200 realizations with a SNR of 20 is added to obtain the final output dataset to be used for identification of model. Further, a candidate library of terms as given in Eq. (9) are considered for modeling the system.

$$\Theta(C_A, T, Q) = \Big[1 \;\; C_A \;\; T \;\; C_A^2 \;\; T^2 \;\; \log(C_A) \;\; \log(T) \;\; \sin(C_A) \;\; \sin(T) \;\; \exp\left(-\frac{E}{RT}\right) C_A^2 \;\; Q\Big]$$

The estimated parameters (mean and standard deviation) for the concentration and temperature with the proposed method and the SINDy are shown in Tables 3 and 4 respectively.

Table 3. Estimates (mean and standard deviation) of parameters of coefficients of each candidate term in the Library for state C_A of CSTR case study

Library term	**True Value**	**Proposed method**	**SINDy**
1	20	20.2693 ± 1.184	-3211.2 ± 4443.8
C_A	-5	-4.8260 ± 0.125	106.72 ± 140.97
T	0	0	-8842.9 ± 6470
C_A^2	0	0	-53501 ± 75501
T^2	0	0	$(6.2456 \pm 8.8136) \times 10^5$
$\log(C_A)$	0	0	-6.9261 ± 9.5584
$\log(T)$	0	0	4884.2 ± 6107.5
$\sin(C_A)$	0	0	0
$\sin(T)$	0	$(-3.3 \pm 0.6) \times 10^{-3}$	0
$\exp\left(-\frac{E}{RT}\right) C_A^2$	-8.46×10^6	$(-8.314 \pm 0.406) \times 10^6$	-3.8083 ± 5.3857
Q	0	0	-1144.3 ± 1623.4

Table 4. Estimates (mean and standard deviation) of parameters of coefficients of each candidate term in the Library for state T of CSTR case study

Library term	**True Value**	**Proposed method**	**SINDy**
1	1500	1590 ± 210.15	1131.8 ± 0.002
C_A	0	0	9681.5 ± 7083.6
T	-5	-4.9598 ± 0.056	$(1.4793 \pm 2.077) \times 10^5$
C_A^2	0	0	$(-4.6799 \pm 6.6182) \times 10^7$
T^2	0	0	$(-8.5092 \pm 12.034) \times 10^6$
$\log(C_A)$	0	0	-2289 ± 191.81
$\log(T)$	0	0	-11287 ± 11169
$\sin(C_A)$	0	0	-2.2569 ± 3.3348
$\sin(T)$	0	-0.3374 ± 0.049	0
$\exp\left(-\frac{E}{RT}\right) C_A^2$	-4.2117×10^8	$(-4.102 \pm 0.0032) \times 10^8$	0
Q	43×10^{-3}	$(4.2 \pm 0.0286) \times 10^{-3}$	-10806 ± 15283

From the results reported in Tables 3 and 4, it can be observed that although very few zero-coefficients are identified by the proposed method, the overall accuracy is high as the estimates of parameters for the non-zero coefficients are very close to the true values. On the other hand, when the SINDy algorithm is applied for the same dataset, a dense parameter vector with values very far from the actual values is obtained. A probable reason for the SINDy algorithm to fail might be due to the inaccurate estimates of derivatives in the presence of noise.

4. CONCLUSIONS

In this paper, an algorithm for data-driven identification of models with inclusion of physical information is proposed. To address the challenges with the derivative information especially in the presence of noise, the data set is fitted using B-splines subjected to a model-based penalty to ensure that the data fit also satisfies the model of the process. For estimation of a parsimonious model, a sparsity constraint in terms of zero norm on the parameter vector of the model is considered. A two step approach is followed to solve the coefficients of B-splines and to estimate the parsimonious model parameters. A systematic approach for selection of initial choice of estimates and the choice of regularization parameter is also presented in the paper. The efficacy of the method is demonstrated on a Van der Pol oscillator and a Continuous stirred tank reactor system and the results are compared with the existing methods. From the results, it can be concluded that the proposed method is providing better and reliable estimates over the existing methods.

REFERENCES

Arnold, F. and King, R. (2021). State–space modeling for control based on physics-informed neural networks. *Engineering Applications of Artificial Intelligence*, 101, 104195.

Beck, A. and Eldar, Y.C. (2013). Sparsity constrained nonlinear optimization: Optimality conditions and algorithms. *SIAM Journal on Optimization*, 23(3), 1480–1509.

Bhadriraju, B., Kwon, J.S.I., and Khan, F. (2021a). Oasis-p: Operable adaptive sparse identification of systems for fault prognosis of chemical processes. *Journal of Process Control*, 107, 114–126.

Bhadriraju, B., Kwon, J.S.I., and Khan, F. (2021b). Risk-based fault prediction of chemical processes using operable adaptive sparse identification of systems (oasis). *Computers & Chemical Engineering*, 152, 107378.

Boor, C.D. (1978). *A practical guide to splines*. Springer Verlag.

Brunton, S.L., Proctor, J.L., and Kutz, J.N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15), 3932–3937.

Candes, E. and Tao, T. (2005). Decoding by linear programming. *IEEE Transactions in Information Theory*, 51, 4203–4215.

Chakraborty, A., Sivaram, A., Samavedham, L., and Venkatasubramanian, V. (2020). Mechanism discovery and model identification using genetic feature extraction and statistical testing. *Computers & Chemical Engineering*, 140, 106900.

Chartrand, R. (2011). Numerical differentiation of noisy, nonsmooth data. *International Scholarly Research Notices*, 2011(1), 164564.

Jung, H. and Lee, J.H. (2023). Flexible operation of post-combustion co2 capture process enabled by narx-mpc using neural network. *Computers & Chemical Engineering*, 179, 108447.

Kaiser, E., Kutz, J.N., and Brunton, S.L. (2018). Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proceedings of the Royal Society A*, 474(2219), 20180335.

Knowles, I. and Renka, R.J. (2014). Methods for numerical differentiation of noisy data. *Electron. J. Differ. Equ*, 21, 235–246.

Liu, Y., Tian, W., Xie, J., Huang, W., and Xin, K. (2023). Lstm-based model-predictive control with rationality verification for bioreactors in wastewater treatment. *Water*, 15(9), 1779.

Ljung, L. (1999). *System identification: Theory for the user*. Prentice Hall, New Jersey.

Luu, M.T., Manaf, N.A., and Abbas, A. (2015). Dynamic modelling and control strategies for flexible operation of amine-based post-combustion co2 capture systems. *International Journal of Greenhouse Gas Control*, 39, 377–389.

Poyton, A., Varziri, M.S., McAuley, K.B., McLellan, P.J., and Ramsay, J.O. (2006). Parameter estimation in continuous-time dynamic models using principal differential analysis. *Computers & chemical engineering*, 30(4), 698–708.

Raissi, M. and Karniadakis, G.E. (2018). Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, 357, 125–141.

Raissi, M., Perdikaris, P., and Karniadakis, G.E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378, 686–707.

Ramsay, J.O., Hooker, G., Campbell, D., and Cao, J. (2007). Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5), 741–796.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of Royal Statistical Society*, 58, 267–288.

Tropp, J. and Gilbert, A.C. (2007). Signal recovery from partial information via orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, 53(12), 4655–4666.

Varanasi, S.K. and Jampana, P. (2018). Identification of parsimonious continuous time lti models with applications. *Journal of Process Control*, 69, 128–137.

Wang, J., Moreira, J., Cao, Y., and Gopaluni, R.B. (2023). Simultaneous digital twin identification and signal-noise decomposition through modified generalized sparse identification of nonlinear dynamics. *Computers & Chemical Engineering*, 177, 108294.

Wu, X., Shen, J., Wang, M., and Lee, K.Y. (2020). Intelligent predictive control of large-scale solvent-based co2 capture plant using artificial neural network and particle swarm optimization. *Energy*, 196, 117070.

Zhang, Q., Turton, R., and Bhattacharyya, D. (2018). Nonlinear model predictive control and h-∞ robust control for a post-combustion co2 capture process. *International Journal of Greenhouse Gas Control*, 70, 105–116.