

# GN-SINDy: Greedy Sampling Neural Network in Sparse Identification of Nonlinear Partial Differential Equations

Ali Forootani\* Peter Benner\*

\**Max Planck Institute for Dynamics of Complex Technical Systems, 39106 Magdeburg, Germany.*

Email: [forootani@mpi-magdeburg.mpg.de](mailto:forootani@mpi-magdeburg.mpg.de), ORCID: 0000-0001-7612-4016

<sup>2</sup> Email: [benner@mpi-magdeburg.mpg.de](mailto:benner@mpi-magdeburg.mpg.de), ORCID: 0000-0003-3362-4103

**Abstract:** The sparse identification of nonlinear dynamical systems (**SINDy**) is a data-driven technique employed for uncovering and representing the fundamental dynamics of intricate systems based on observational data. However, a primary obstacle in the discovery of models for nonlinear partial differential equations (**PDEs**) lies in addressing the challenges posed by the curse of dimensionality and large dataset. Consequently, the strategic selection of the most informative samples within a given dataset plays a crucial role in reducing computational costs and enhancing the effectiveness of **SINDy**-based algorithms. To this aim, we employ a greedy sampling approach to the snapshot matrix of a **PDE** to obtain its valuable samples, which are suitable to train a deep neural network (**DNN**) in a **SINDy** framework. **SINDy** based algorithms often consist of data collection unit, constructing a dictionary of basis functions, computing the time derivative, and solving a sparse identification problem which ends to regularised least squares minimization. In this paper, we extend the results of a **SINDy** based deep learning model discovery (**DeePyMoD**) approach by integrating greedy sampling technique in its data collection unit and new sparsity promoting algorithms in the least squares minimization unit. In this regard we introduce the greedy sampling neural network in sparse identification of nonlinear partial differential equations (**GN-SINDy**) which blends a greedy sampling method, the **DNN**, and the **SINDy** algorithm. In the implementation phase, to show the effectiveness of **GN-SINDy** we compare its results with **DeePyMoD** by using a **Python** package that is prepared for this purpose on numerous **PDE** discovery.

**Keywords:** Discrete Empirical Interpolation Method (**DEIM**), Deep Neural Network (**DNN**), Sparse Identification of Nonlinear Dynamical Systems (**SINDy**)

**Mathematics subject classification:** MSC1, MSC2, MSC3

**Novelty statement:** We introduce the Greedy Sampling Neural Network in Sparse Identification of Nonlinear Partial Differential Equations (**GN-SINDy**), a pioneering approach that seamlessly integrates a novel greedy sampling technique, deep neural networks, and advanced sparsity-promoting algorithms. Our method not only addresses the formidable challenges posed by the curse of dimensionality and large datasets in discovering models for nonlinear **PDEs** but also sets a new standard for efficiency and accuracy by redefining the data collection and minimization units within the **SINDy** framework. By combining the strengths of these diverse techniques, **GN-SINDy** represents a leap forward in the realm of model discovery, promising unprecedented insights into the intricate dynamics of complex systems.

## 1 Introduction

Nonlinear dynamical systems are often encountered in various scientific fields, ranging from physics and biology to economics and engineering. Such systems can be extremely complex and difficult to understand, especially when they involve a large number of variables. Sparse identification of nonlinear dynamics (**SINDy**) is a powerful approach that can help unravel the mysteries of these systems. Using a combination of machine learning and optimisation techniques, **SINDy** can identify the governing equations of a nonlinear dynamical system from noisy and scarce data [1, 2].

When dealing with partial differential equations (PDEs), **SINDy** can be adapted to identify sparse representations of the nonlinear terms in the PDE [3, 4]. Traditional methods for PDE identification often depend on a combination of domain specific knowledge, mathematical derivations, and experimental data. However, these approaches can be labor-intensive and may encounter limitations when applied to complex or poorly understood systems. Overcoming these challenges requires innovative techniques that can efficiently identify PDEs, provide a more comprehensive understanding of intricate physical phenomena, and enable improved predictions in diverse applications [5, 6].

In the existing body of literature, the predominant focus has been on solving PDEs through either analytical or numerical means. Analytical approaches involve techniques like variable transformations to render the equation amenable or the derivation of an integral form of the solution [7]. While these methods find applicability in handling straightforward PDEs, their efficacy diminishes when faced with more intricate equations. On the other hand, numerical methods aim to approximate the solution of a PDE by discretizing its domain and solving a set of algebraic equations. Widely used numerical techniques include the finite difference method (FDM) [8] and the finite element method (FEM) [9].

Besides solving the PDEs with conventional methods, recent advances in machine learning techniques have proved their potential to address PDE problems in scenarios with limited dataset. This implies having access solely to the PDE problem data, rather than an extensive set of value pairs for the independent and dependent variables [10]. Taking advantage of modern machine learning software environments have provided automatic differentiation capabilities for functions realized by deep neural networks (DNN) which is a mesh-free approach and can break the curse of dimensionality of the conventional methods [11].

This approach was introduced in [12], where the term physics-informed neural networks (PINNs) was coined. With the emergence of PINN, employing a neural network has become a prominent method to construct a surrogate for data and subsequently conduct sparse regression on the network's predictions [13–15]. Alternatively, Neural ODEs have been introduced to unveil unknown governing equations [16] from physical datasets. Diverse optimization strategies, employing the method of alternating direction, are explored in [17], and graph-based approaches are formulated in [18]. Symmetry incorporation into neural networks is addressed directly by [19], utilizing both the Hamiltonian and Lagrangian frameworks. Furthermore, auto-encoders have been utilized to model PDEs and uncover latent variables [20]. However, this approach does not yield an explicit equation and demands substantial amounts of data. It is worth to highlight that unlike the traditional PDE solvers that focus more on methods such as the FDM [8] and FEM [9], the DNN based approaches (such as PINN) are mesh free and therefore highly flexible.

The technique that we focus in this paper for PDE identification is based on **SINDy** algorithm, i.e. a method able to select, from a large dictionary, the correct linear, nonlinear, and spatial derivative terms, resulting in the identification associated PDEs from data [21]. In **SINDy** only those dictionary terms that are most informative about the dynamics are selected as part of the discovered PDE. Previous sparse identification algorithms faced a number of challenges [21]: They were not able to handle sub-sampled spatial data, and the algorithm did not scale well to high-dimensional measurements. Standard model reduction techniques such as proper orthogonal decomposition (POD) were used to overcome the high-dimensional measurements, allowing for a lower-order ODE model to be constructed on energetic POD modes. This procedure resembles standard Galerkin projection onto POD modes [3]. **SINDy** already applied on various model discovery applications, including for reduced-order models of fluid dynamics [22] and plasma dynamics [23], turbulence closures [24], mesoscale ocean closures [25], nonlinear optics [26], computational chemistry [27], and numerical integration schemes [28]. However, **SINDy** algorithm in its original form has drawbacks such as sensitivity to accurate derivative information, lack of performance in the scarce data, sensitivity

to the noisy dataset [3].

Enhancements to the SINDy framework have aimed at improvement its resilience to noise, providing uncertainty quantification, and adapting it for the modeling of stochastic dynamics [27, 29–32]. Nevertheless, these extensions have typically depended on computationally intensive methods for acquiring the necessary knowledge about probability distributions.

## 1.1 Contribution

Expressing the unknown differential equation as  $\partial_t \mathbf{u} = \mathbf{f}(\mathbf{u}, \mathbf{u}_x, \dots)$  and assuming that the right-hand side is a linear combination of predefined terms, i.e.,  $\mathbf{f}(\mathbf{u}, \mathbf{u}_x, \dots) = a\mathbf{u} + b\mathbf{u}_x + \dots = \Theta\xi$ , simplifies model discovery to identifying a sparse coefficient vector  $\xi$ . The challenge lies in computing the time derivative  $\mathbf{u}_t$  and the function dictionary  $\Theta$ , especially when dealing with large dataset. The associated error in these terms tends to be high due to the utilization of numerical differentiation techniques like finite difference or spline interpolation. This limitation confines classical model discovery to datasets characterized by low noise and dense sampling. In contrast, deep learning-based methods overcome this challenge by constructing a surrogate from the data and determining the feature library  $\Theta$ , along with the time derivative  $\mathbf{u}_t$ , through automatic differentiation. The DNN can be integrated seamlessly into SINDy algorithm and be employed to effectively model the data, therefore facilitating the construction of a comprehensive function dictionary.

A pivotal aspect of this approach involves the dynamic application of a mask during training loop of the DNN, selectively activating terms in the function dictionary. In this regard the DNN is constrained to conform to solutions derived from these active terms. To determine this mask, any non-differentiable sparsity-promoting algorithm can be employed, such as STRidge [21]. This sophisticated approach not only allows for the use of a constrained neural network to precisely model the data and construct a robust function library but also employs an advanced sparsity-promoting algorithm to dynamically unveil the underlying equation based on the network's output. However, as the number of training samples increase for a DNN, the longer it takes to train the network. To alleviate this situation, it becomes crucial to choose a set of informative samples for training the network together with sparsity promoting algorithm (SINDy), and later, for PDE model discovery.

In this study, the Discrete Empirical Interpolation Method (DEIM [33]) is employed to strategically select informative samples from a snapshot matrix associated to a PDE. The objective is to reduce the dimensionality of the high-dimensional system while retaining essential features, ultimately enhancing the efficiency of subsequent data-driven methodologies. DEIM is a discrete variant of the empirical interpolation method (EIM) designed for constructing approximations of non-affine parameterized functions in continuous bounded domains, offering an associated error bound on the quality of approximation [34]. In particular, we utilize Q-DEIM method which improves upon the original DEIM algorithm by providing a superior upper bound error and exhibiting numerically robust, high-performance procedures available in widely-used software packages such as Python, LAPACK, ScaLAPACK, and MATLAB [33].

The selected Q-DEIM samples are employed to train a DNN within the context of a SINDy-based algorithm. This integrated approach leverages the strengths of both Q-DEIM and neural networks, where Q-DEIM efficiently captures the most relevant information from the PDE solution snapshots, and the neural network learns the underlying dynamics. Therefore, we name our proposed approach greedy sampling neural network for sparse identification of nonlinear dynamical systems (GN-SINDy). In a detailed exploration, we systematically assess the influence of most informative samples obtained through Q-DEIM on the snapshot matrix of a partial differential equation (PDE) for the purpose of discovering its governing equation. Additionally, we conduct a comparative analysis between the outcomes of GN-SINDy and DeePyMoD[4] which uses random sampling approach. Our results underscore the substantial effect of leveraging highly informative samples on the training process of the DNN architecture for sparse identification of nonlinear PDE. Moreover, a Python package is prepared to support different implementation phases of GN-SINDy and its comparison with DeePyMoD PDE model discovery corresponding to the Allen-Cahn equation, Burgers' equation and Korteweg-de Vries equation.

This article is organised as follows:

In [Section 2](#), we provide an overview of the SINDy algorithm, focusing on its application for identifying PDEs. We elaborate on our approach to greedy sampling and its utilization in selecting the most informative samples from the dataset associated with PDEs in [Section 3](#). The GN-SINDy algorithm, introduced in [Section 4](#), is dedicated to the discovery of PDEs. In [Section 5](#), we delve into the simulation and comparative analysis of GS-PINN with DeePyMoD. The paper concludes with a summary in [Section 6](#).

## 2 An Overview of SINDy for PDE identification

Consider a nonlinear partial differential equations of the form

$$\mathbf{u}_t = \mathbf{f}(\mathbf{u}, \mathbf{u}_x, \mathbf{u}_{xx}, \dots, x), \quad (1)$$

where the subscripts denote partial differentiation in either time  $t \in [0, t_{\max}]$ , or space  $x \in [x_{\min}, x_{\max}]$ , and  $\mathbf{f}$  is an unknown right-hand side that is generally a non-linear function of  $\mathbf{u}(x, t)$ , and its derivatives. The goal in SINDy algorithm is to discover  $\mathbf{f}(\cdot)$  from the time series measurements of the dynamical system at a set of spatial points  $x$ . The main assumption behind SINDy is that the un-known function  $\mathbf{f}(\cdot)$  has only a few terms which makes the functional form sparse with respect to large space of possible contributing terms. For instance we can name Allen-Chan equation ( $\mathbf{f} = 0.0001\mathbf{u}_{xx} - 5\mathbf{u}^3 + 5\mathbf{u}$ ) or Korteweg-De Vries equation ( $\mathbf{f} = -6\mathbf{u}\mathbf{u}_x - \mathbf{u}_{xxx}$ ).

SINDy algorithm is summarized in the following four steps.

### 1. Data collection:

To do so we first measure  $\mathbf{u}$  at  $m$  different time instances and  $n$  spatial locations, and we construct a single column vector  $\mathbf{U} \in \mathbb{R}^{n \cdot m}$ .

### 2. Constructing the dictionary of basis functions:

At the second step we have to construct a dictionary consist of  $D$  candidate liner, nonlinear, and partial derivatives for the PDE:

$$\Theta(\mathbf{U}) = [\mathbf{1}, \mathbf{U}, \mathbf{U}^2, \dots, \mathbf{U}_x, \mathbf{U}\mathbf{U}_x, \dots], \quad \Theta \in \mathbb{R}^{nm \times D}, \quad (2)$$

where each single column of the matrix  $\Theta$  contains all of the values of a particular candidate function across all of the  $n \cdot m$  space-time grid points that data are collected. For example if we measure our process at 200 spacial locations, at 300 time instances and we include 40 candidate terms in the PDE, then  $\Theta \in \mathbb{R}^{200 \cdot 300 \times 40}$ . These basis functions represent the possible dynamics that the system can exhibit. The choice of basis functions depends on the specific problem and can range from simple polynomials to more complex functions such as trigonometric functions or exponential functions and it is quite essential for the success of the algorithm.

### 3. Computing the time derivative:

The third step in the SINDy algorithm is to compute the time derivative of  $\mathbf{U}$ , which is often implemented numerically. Having  $\mathbf{U}_t$  and other ingredients we can write the system of equation (1) in the following form:

$$\mathbf{U}_t = \Theta(\mathbf{U})\xi, \quad (3)$$

where  $\xi \in \mathbb{R}^D$  is an unknown vector that has to be computed by proper algorithm and its elements are coefficients corresponding to the active terms in the dictionary  $\Theta(\cdot)$  that describe the evolution of the dynamic system in time. We show by  $\xi_i$ , the  $i$ th element of this column vector.

### 4. Solving the sparse identification problem:

The last and fourth step in the SINDy algorithm is to compute the coefficient vector  $\xi$  through a least squares optimization formulation which is a Nondeterministic Polynomial time (NP) mathematical hard problem. Therefore, there is a need to introduce a regularization technique that promotes sparsity. In the realm of regression analysis, different techniques address the challenges associated with model complexity and

overfitting. Ordinary least squares (**OLS**) optimization, a classical method, often leads to complex models vulnerable to overfitting. Recognizing this limitation, sparse optimization techniques, such as **LASSO** and **Ridge** regression, introduce regularization to the **OLS** objective function [35, 36]. **LASSO** employs an  $l_1$  penalty, promoting sparsity in the coefficient vector and facilitating effective feature selection, while **Ridge** regression, utilizing an  $l_2$  penalty, enhances stability by addressing multicollinearity and preventing overfitting.

The sequential threshold ridge regression (**STRidge**) algorithm builds on these principles, introducing a regularized variant of **OLS** that effectively deals with challenges in discovering physical laws within highly correlated and high-dimensional datasets. **STRidge** addresses the limitations of standard regression methods, making it particularly useful in scenarios involving spatio-temporal data or complex, correlated features. Each of these approaches represents a nuanced trade-off between model complexity and predictive accuracy, allowing researchers and practitioners to tailor their choices based on the specific characteristics of their datasets and the goals of their modeling endeavors.

In the original **SINDy** algorithm authors used **STRidge** to unveil elusive governing equations, typically expressed as **ODEs** [21]. This approach laid the foundation for a broader framework known as PDE-functional identification of nonlinear dynamics (PDE-FIND)[3], facilitating the discovery of unknown relationships within the system [37]. In some other approaches as discussed in [38, 39], the authors considered to have some prior knowledge regarding non-zero terms in the coefficient vector  $\xi$ , which contradicts with the main assumption of **SINDy** based approaches. The hyperparameters in **STRidge** include the regularization weight  $\lambda_{\text{STR}}$  tolerance level  $\text{tol}_{\text{STR}}$ , and maximum iteration  $\text{max-iter}_{\text{STR}}$  which are to be tuned to identify appropriate physical models. The convergence of analysis of sequential thresholding can be found in [40].

### 3 Importance of the sampling methods for PDE discovery

By strategically opting the most informative samples, we can ensure that the identified PDE faithfully represents the intricate dynamics of the system. This careful selection not only enhances the predictive capacity of the model but also provides valuable insights into the underlying physical processes. Moreover, choosing the most informative samples expedite the training time of the DNN, and hence reduces the associated computational cost. Conversely, an inadequate sampling strategy may lead to inaccuracies in PDE identification, hindering scientific advancements in diverse fields. Therefore, the judicious combination of an effective sampling strategy with the consideration of informative samples is pivotal for the successful and reliable identification of PDEs.

Within the existing body of literature, several residual point sampling methodologies have predominantly found application. Noteworthy among these are non-adaptive uniform sampling techniques, including (i) the equispaced uniform grid, (ii) uniformly random sampling, (iii) Latin hypercube sampling [12], (iv) the Halton sequence [41], (v) Hammersley sequence [42], and (vi) the Sobol sequence [43]. Additionally, adaptive non-uniform sampling approaches, such as Residual-based Adaptive Distribution (**RAD**) and Residual-based Adaptive Refinement with Distribution (**RAR-D**) [44], have been explored. Despite the potential offered by these methodologies, they exhibit a strong dependence on specific problem characteristics and commonly entail laborious and time-intensive processes. Notably, adaptive sampling methods like **RAD** or **RAR-D** necessitate dataset re-sampling within the training loop of the **NN**, introducing considerable computational overhead to the associated algorithm. For a comprehensive exploration of non-adaptive and residual-based adaptive sampling strategies and their comparisons in PINN training, the interested reader is directed to the detailed investigation presented in [44].

In numerous practical scenarios, like phase-field modeling or fluid dynamics, sensors are commonly employed to measure state variables. The quantity of sensors is often restricted by physical or financial constraints, and strategically situating these sensors is vital for attaining accurate estimations. Unfortunately, identifying the optimal locations for sensors to deduce the parameters of the PDE poses an inherent combinatorial challenge. Existing approximation algorithms may not consistently yield efficient solutions across all pertinent cases. The matter of optimal sensor placement captivates research attention, extending its significance even into domains such as control

theory and signal processing [15].

Literature have documented five categories of strategies for positioning sensors, as follows: (i) techniques relying on proper orthogonal decomposition (POD) [45] or sparse sensing like compressed sensing [46], (ii) methodologies involving convex optimization [47], (iii) algorithms guided by a greedy approach, exemplified by Frame-Sens [48], (iv) heuristic methods, including population-based search [49], and (v) application of machine learning methodologies [50]. The applicability of these methods is limited to particular cases due to conservative assumptions, complexity in implementation for large scale problems.

An interesting numerical technique that can be employed as a sampling method is Q-DEIM algorithm which is originated in the context of model order reduction for high-dimensional dynamical systems [51]. It is particularly useful when dealing with large-scale problems, such as those arising in computational physics or engineering simulations. Q-DEIM aims to identify a reduced set of basis functions that capture the essential dynamics of the system by exploiting the empirical interpolation idea. The method strategically selects a sparse set of interpolation points from the high-dimensional state space, enabling an efficient representation of the system's behavior. Q-DEIM has proven to be effective in reducing the computational cost associated with solving complex systems by constructing a low-dimensional surrogate model while preserving key system features.

In the case of PDE discovery, and specifically DNN framework, with an increase in the number of training samples, the training duration also extends. To address this issue, selecting a set of informative samples for network training becomes crucial. Our subsequent goal is to identify the governing PDE underlying the data set within SINDy procedure. To be more precise, we methodically investigate the impact of the most informative samples acquired via Q-DEIM on PDE snapshot matrix.

### 3.1 Notes on Q-DEIM algorithm

Herein, we provide a concise introduction to Q-DEIM algorithm, emphasizing its association with a well-known model order reduction technique known as POD. POD has been used widely to select measurements in the state space that are informative for feature space reconstruction [52]. Consider the set of snapshots  $\{u_1, \dots, u_m\} \in \mathbb{R}^n$  and an associated snapshot matrix  $\mathcal{U} = [u_1, \dots, u_m] \in \mathbb{R}^{n \times m}$  that is constructed by measuring the solution at  $m$  different time instances and  $n$  different spatial locations of a PDE. In the conventional POD, we construct an orthogonal basis that can represent the dominant characteristics of the space of expected solutions that is defined as Range  $\mathcal{U}$ , i.e., the span of the snapshots. We compute the singular value decomposition (SVD) of the snapshot matrix  $\mathcal{U}$ ,

$$\mathcal{U} = \mathbf{Z}\Sigma\mathbf{Y}^\top, \quad (4)$$

where  $\mathbf{Z} \in \mathbb{R}^{n \times n}$ ,  $\Sigma \in \mathbb{R}^{n \times m}$ , and  $\mathbf{Y} \in \mathbb{R}^{m \times m}$  with  $\mathbf{Z}^\top\mathbf{Z} = \mathbf{I}_n$ ,  $\mathbf{Y}^\top\mathbf{Y} = \mathbf{I}_m$ , and  $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_z)$  with  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_z \geq 0$  and  $z = \min\{m, n\}$ . The POD will select  $\mathbf{V}$  as the leading  $r$  left singular vectors of  $\mathbf{U}$  corresponding to the  $r$  largest singular values. Using Python-Numpy array notation, we denote this as  $\mathbf{V} = \mathbf{Z}[:, :r]$ . The basis selection via POD minimizes  $\mathbf{V} := \min_{\Phi \in \mathbb{R}^{n \times r}} \|\mathcal{U} - \Phi\Phi^\top\mathcal{U}\|_F^2$ , where  $\|\cdot\|_F$  is the Frobenius norm, over all  $\Phi \in \mathbb{R}^{n \times r}$  with orthonormal columns. In this regard, we can say  $\mathcal{U} = \mathbf{Z}\Sigma\mathbf{Y}^\top \approx \mathbf{Z}_r\Sigma_r\mathbf{Y}_r^\top$ , where matrices  $\mathbf{Z}_r$  and  $\mathbf{Y}_r^\top$  contain the first  $r$  columns of  $\mathbf{Z}$  and  $\mathbf{Y}^\top$ , and  $\Sigma_r$  contains the first  $r \times r$  block of  $\Sigma$ . More details regarding POD can be found in [53].

While the reduced-order model resides within the  $r$ -dimensional subspace, the conventional POD encounters a challenge when transitioning back to the original space. To address this issue, various approaches have been proposed in the literature, with DEIM being one such solution [51]. Notably, DEIM offers the distinct advantage of flexibility, allowing its outcomes to extend beyond the realm of model order reduction, especially in the context of nonlinear function approximation. Additionally, the performance of the original DEIM algorithm has been enhanced by incorporating QR-factorization, resulting in two notable improvements: (i) a reduction in upper bound error and (ii) increased simplicity and robustness in implementation. Q-DEIM leverages the pivoted factorization of QR factorization and the SVD, providing a robust sampling method. Specifically, we employ QR factorization with column pivoting on  $\mathbf{Z}_r^\top$  and  $\mathbf{Y}_r^\top$  to identify the most informative samples in the snapshot matrix  $\mathcal{U}$ . This pivoting technique offers an approximate greedy solution for feature selection, termed submatrix volume maximization, as the matrix volume is defined by the absolute value of the determinant. Note that QR-factorization has been implemented and

optimized in most scientific computing packages and libraries, such as MATLAB, and Python. Further details about Q-DEIM, its theoretical analysis, and applications can be found in [51, 52].

### 3.2 Applying Q-DEIM algorithm on PDE dataset

We employ the Q-DEIM algorithm to acquire the most valuable samples on the spatio-temporal grid, utilizing the snapshot matrix  $\mathcal{U}$ . To accomplish this, initially SVD is performed on the snapshot matrix  $\mathcal{U}$ , yielding matrices  $\mathbf{Z}$ ,  $\Sigma$ , and  $\mathbf{Y}^\top$ . Choosing  $r$  leading singular values is determined based on a precision value  $\epsilon_{\text{thr}}$ , which is associated with the underlying dynamical system and is subject to heuristic selection. In the literature,  $\epsilon_{\text{thr}}$  is commonly known as the energy criterion [54]. Specifically, the  $r$  leading singular values are chosen to satisfy the following criterion:

$$1 - \frac{\sum_{j=1}^r \sigma_j}{\sum_{k=1}^z \sigma_k} < \epsilon_{\text{thr}}, \quad r < z, \quad (5)$$

once the desired level of precision is attained, we create a reduced approximation of the snapshot matrix  $\mathcal{U}$  by extracting the first  $r$  columns of matrix  $\mathbf{Z}$ , the first  $r$  singular values from the diagonal matrix  $\Sigma$ , and the initial  $r$  rows of  $\mathbf{Y}^\top$ . In Python notation, this process is expressed as  $\mathbf{Z}_r = \mathbf{Z}[:, :r]$ ,  $\Sigma_r = \Sigma[:r, :r]$ , and  $\mathbf{Y}^\top r = \mathbf{Y}^\top[:r, :]$ . To identify significant time and space indices, we implement QR decomposition with column pivoting on the reduced-order matrices  $\mathbf{Y}_r^\top$  and  $\mathbf{Z}_r^\top$ , housing the foremost  $r$  left and right singular vectors. The indices corresponding to the most informative spatio-temporal points in the snapshot matrix  $\mathcal{U}$  are denoted as `indx` and `indt`. For simplicity, we represent pairs of space-time points as  $(t_i, x_i)$  and their associated solutions as  $u_i$ . [Algorithm 1](#) succinctly outlines the essential steps of the Q-DEIM sampling approach, taking the snapshot matrix  $\mathcal{U}$ , spatio-temporal domains  $x$ ,  $t$ , and the precision value  $\epsilon_{\text{thr}}$  as inputs and yielding sampled pairs  $(t_i, x_i)$  along with the corresponding measured values  $u(t_i, x_i)$ . The cardinality of the sampled dataset is denoted as  $\mathcal{N}$ .

#### 3.2.1 Importance of domain division

To analyze the local dynamics in the dataset and pinpoint optimal points in the spatio-temporal domain, we utilize a method that involves partitioning the time domain into uniform intervals. This approach employs the Q-DEIM on each sub-domain to enhance efficiency. The parameter representing the number of divisions in the time domain is denoted as  $t_{\text{div}}$ . In this particular setup, the subdomains are non-overlapping, resulting in the total number of selected points being a combination of samples chosen from each subdomain.

In the context of Python notation, if we opt to divide the time domain into three parts, the representation will be as follows: the first part as  $\mathcal{U}[:, :m/3]$ , the second part as  $\mathcal{U}[:, m/3 : 2m/3]$ , and the third part as  $\mathcal{U}[:, 2m/3 :]$ . This decomposition strategy is designed to capture the local dynamics of the partial differential equation (PDE) in each subdomain, focusing on sampling the most informative segments of the snapshot matrix. The rationale behind this lies in recognizing that the system's behavior might showcase variations across different domains, with specific physical attributes exhibiting notable distinctions. Noteworthy examples include issues related to abrupt features like shock waves, which may manifest these differences. Furthermore, dividing a large domain into smaller sub-domains and independently applying Q-DEIM to each sub-domain serves to alleviate the necessity for intricate neural network structures in PDE discovery. This streamlined approach enhances computational efficiency and facilitates a more targeted analysis of local variations within the dataset.

## 4 GN-SINDy: Greedy Sampling Neural Network for Sparse Identification of PDEs

DNN has been widely used in literature for solving PDEs due to their strength as the universal function approximators that can represent and learn complex relationships between input data and output solutions. This property is especially valuable in the context of PDEs, where finding analytical solutions may be challenging or impossible for certain complex systems [55]. However,

**Algorithm 1:** Sample selection based on a two-way Q-DEIM procedure

---

**Data:**  $\mathcal{U}$ ,  $\{x_k\}_{k=1}^n$ ,  $\{t_k\}_{k=1}^m$ ,  $\epsilon_{\text{thr}}$ .  
**Result:**  $\text{ind}_t$ ,  $\text{ind}_x$ , domain sampled pairs  $(t_i, x_i)$  and  $u(t_i, x_i)$ .

- 1  $r = 1$ ;
- 2  $\mathbf{Z}, \Sigma, \mathbf{Y}^\top \leftarrow \text{SVD}(\mathcal{U})$ , ▷ computing SVD on snapshot matrix  $\mathcal{U}$ ;
- 3 Find the lowest  $r$ , such that  $1 - \frac{\sum_{j=1}^r \sigma_j}{\sum_{j=1}^m \sigma_j} \geq \epsilon_{\text{thr}}$ ;
- 4  $\mathbf{Z}_r \leftarrow \mathbf{Z}[:, :r]$ ,  $\mathbf{Y}_r^\top \leftarrow \mathbf{Y}^\top[:, r:]$ , ▷ selecting  $r$  dominant left and right singular vectors;
- 5  $\text{ind}_x \leftarrow \text{QR}(\mathbf{Z}_r^\top, \text{pivoting} = \text{True})$ , ▷ storing pivots from pivoted QR factorization of  $\mathbf{Z}_r^\top$ ;
- 6  $\text{ind}_t \leftarrow \text{QR}(\mathbf{Y}_r, \text{pivoting} = \text{True})$ , ▷ storing pivots from pivoted QR factorization of  $\mathbf{Y}_r^\top$ ;
- 7  $x_i \leftarrow \text{from } \text{ind}_x$ ,  $t_i \leftarrow \text{from } \text{ind}_t$ ,  $u(t_i, x_i)$ ;

---

for the case of PDE model discovering from a given snapshot matrix there exist a few works that addressed such problem. Our DNN structure is similar to the work reported in [4] with mainly difference in the Q-DEIM algorithm that we employ to select the most valuable samples in the training loop. Several other works extended the results of [4] for the case of noisy and scarce dataset by using integration scheme in the DNN training loop [1, 2].

The process of model discovery through deep learning involves the utilization of a neural network to generate a surrogate model, denoted as  $\hat{u}$ , for the given data  $u$ . A collection of potential terms, represented by the dictionary  $\Theta$ , is established through automatic differentiation from  $\hat{u}$ . The neural network is then restricted to solutions permissible within this term dictionary. The network's loss function encompasses two key components: (i) a mean square error aimed at learning the mapping  $(t, x) \rightarrow \hat{u}$  and (ii) a term designed to impose constraints on the network's solutions,

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \left( \mathbf{u}(t_i, x_i) - \hat{\mathbf{u}}(t_i, x_i) \right)^2 + \frac{1}{N} \sum_{i=1}^N \left( \frac{\partial \hat{\mathbf{u}}(t_i, x_i)}{\partial t_i} - \Theta(\hat{\mathbf{u}}(t_i, x_i)) \xi \right)^2, \quad (6)$$

where the domain sampled pairs  $(t_i, x_i)$ , and  $u(t_i, x_i)$  are computed by employing Q-DEIM algorithm on the snapshot matrix  $\mathcal{U}$ .

The sparse vector  $\xi$  is being learned during the training loop, and has two roles: (i) identifying the active elements, or those with non-zero values, in the PDE, and (ii) imposing constraints on the network based on these active terms. In our implementation we dissociate these two objectives by uncoupling the constraint from the actual process of selecting sparsity. Initially, we compute a sparsity mask, denoted as  $\mathbf{g}$ , and subsequently restrict the network solely based on the active terms identified within this mask. In the sense that rather than imposing constraints on the neural network using  $\xi$  alone, we opt to constrain it with the element-wise multiplication of  $\xi$  and  $\mathbf{g}$ , thereby replacing equation (6) with:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \left( \mathbf{u}(t_i, x_i) - \hat{\mathbf{u}}(t_i, x_i) \right)^2 + \frac{1}{N} \sum_{i=1}^N \left( \frac{\partial \hat{\mathbf{u}}(t_i, x_i)}{\partial t_i} - \Theta(\hat{\mathbf{u}}(t_i, x_i)) (\xi \cdot \mathbf{g}) \right)^2. \quad (7)$$

The training process according to equation (7) involves a two-step procedure. Initially, we compute the sparsity mask, denoted as  $\mathbf{g}$ , utilizing a sparse estimator. Subsequently, we minimize it with respect to network parameters, employing the masked coefficient vector. Importantly, the sparsity mask  $\mathbf{g}$  does not necessitate a differentiable calculation, allowing for the utilization of any traditional, non-differentiable sparse estimator.

This approach has the following advantages: i) it furnishes an impartial estimate of the coefficient vector by abstaining from applying  $l_1$  or  $l_2$  regularization on  $\xi$  ii) the sparsity pattern is derived from the complete dictionary  $\Theta$  rather than solely from the persisting active terms, enabling the dynamic inclusion and exclusion of active terms throughout the training process. In this regard our training loop consist of a (i)function approximator that creates a surrogate model of the dataset, (ii) a dictionary  $\Theta$  of possible terms and time derivatives  $\frac{\partial \hat{u}}{\partial t}$ , (iii) a sparsity estimator that creates a mask to choose the active columns in the dictionary using sparse regression technique, and (iv) a constraint that imposes the function approximator to the solutions allowed by the active terms obtained by sparsity estimator.

**Notes on sparsity mask and sparsity estimator** The sparsity mask  $\mathbf{g}$  is computed using an estimator which is not involved in the training loop of the DNN structure and it is not differentiable. Therefore, the right procedure to update the sparsity mask  $\mathbf{g}$  within the training loop is essential for the success of the algorithm. In particular, it is important to allow the function approximator unit learns the solution  $\mathbf{u}$  for some iterations, then the sparsity mask be updated. We control the update of the sparsity mask  $\mathbf{g}$  with parameters `patience`, `periodicity`. We define the `patience` as the number of iterations that we allow the DNN to learn the solution  $u$  before applying the sparsity mask. Moreover, the `periodicity` refers to the consistent intervals at which we check whether sparsity mask needs to be updated or not within the training iterations. Sparsity threshold  $\delta_{\text{spr}}$  is a value that we impose on the solution that is computed via sparsity estimator to update the sparsity mask  $\mathbf{g}$ . Note that the sparse estimator solves  $\frac{\partial \hat{\mathbf{u}}(t_i, x_i)}{\partial t_i} - \Theta(\hat{\mathbf{u}}(t_i, x_i))\xi_{\text{est}} = 0$ , where  $\xi_{\text{est}}$  is its solution. The elements of the vector  $\xi_{\text{est}}$  that have absolute values less than sparsity threshold  $\delta_{\text{spr}}$  will be used to update the sparsity mask  $\mathbf{g}$ . It is worth to highlight that an appropriate algorithm that promotes sparsity can be used to compute  $\xi_{\text{est}}$  such as STRidge [3], LASSO [56].

**Notes on constraint** The sparse coefficient vector  $\xi$  is computed concurrently through the DNN training loop. When the sparsity mask  $\mathbf{g}$  is updated, the dictionary columns corresponding to inactive terms are omitted in the solution of  $\frac{\partial \hat{\mathbf{u}}(\cdot)}{\partial t} - \Theta(\hat{\mathbf{u}}(\cdot))\xi = 0$ . Note that by default it is assumed that all the dictionary terms are active (sparsity mask  $\mathbf{g}$ ). Omitting inactive terms in the dictionary  $\Theta$  in parallel to decreasing mean square error which decreases the complexity of finding the coefficient vector  $\xi$ . Moreover, any suitable algorithm that promotes sparsity can be used to compute  $\xi_{\text{est}}$  such as STRidge [3], LASSO [56]. [Algorithm 2](#) summarizes the procedure that has been explained in this section.

---

**Algorithm 2:** GN-SINDy: greedy sampling neural network for sparse identification of nonlinear PDEs

---

**Data:**  $\mathcal{U}, x, t, \epsilon_{\text{thr}}$  for the Q-DEIM algorithm, a neural network  $\mathcal{G}_\theta$  (parameterized by  $\theta$ ), maximum iterations `max-iter`, `patience`, `periodicity`, and sparsity threshold  $\delta_{\text{spr}}$

**Result:** Estimated coefficient vector  $\xi$

- 1  $(t_i, x_i), u(t_i, x_i) \leftarrow$  Apply Q-DEIM( $\mathcal{U}$ ) based on [Algorithm 1](#)  $\triangleright$  selecting most informative samples ;
- 2 Initialize the DNN module parameters;
- 3  $k = 1$ ;
- 4 Initialize the sparsity mask  $\mathbf{g}$ ;
- 5 **while**  $k < \text{max-iter}$  **do**
  - 6   - Feed the domain pairs  $(t_i, x_i)$  as an input to the DNN ( $\mathcal{G}_\theta$ ) and predict output  $\hat{\mathbf{u}}(t_i, x_i)$  ;
  - 7   - Construct the library  $\Theta(\hat{\mathbf{u}}(t_i, x_i))$ ;
  - 8   - Compute the derivative information  $\frac{\partial \hat{\mathbf{u}}(t_i, x_i)}{\partial t_i}$  using automatic differentiation ;
  - 9   - Compute the coefficient vector  $\xi$  by sparsity promoting algorithm such as STRidge, subject to the sparsity mask  $\mathbf{g}$  to solve  $\frac{\partial \hat{\mathbf{u}}(t_i, x_i)}{\partial t_i} - \Theta(\hat{\mathbf{u}}(t_i, x_i))(\xi \cdot \mathbf{g}) = 0$ ;
  - 10   - Compute the cost function (7) ;
  - 11   - Update the parameters of  $\mathcal{G}_\theta$  and the coefficient vector  $\xi$  using gradient descent ;
  - 12   **if**  $(k - \text{patience}) \% \text{periodicity} == 0$  **then**
    - 13       - Update the sparsity mask  $\mathbf{g}$  using Lasso or STRidge algorithm subject to the solution  $\frac{\partial \hat{\mathbf{u}}(t_i, x_i)}{\partial t_i} - \Theta(\hat{\mathbf{u}}(t_i, x_i))\xi_{\text{est}} = 0$  and sparsity threshold  $\delta_{\text{spr}}$  ;
      - 14            $\triangleright$  note that updating the sparsity mask is done independent from estimated coefficients

---

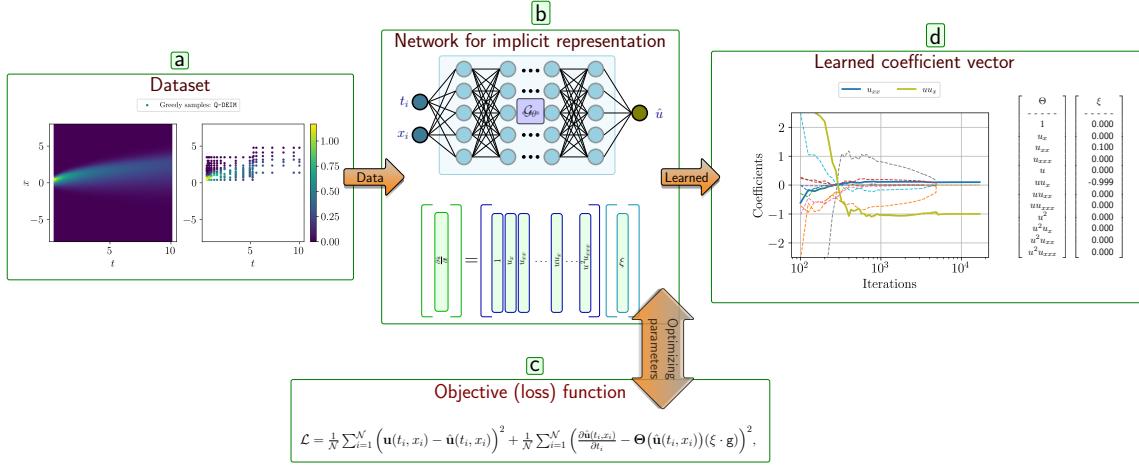


Figure 1: A schematic diagram of GN-SINDy for PDE discovery. (a) sampling PDE dataset with Q-DEIM algorithm to choose the most informative samples, (b) feeding the valuable sample pairs  $(t_i, x_i)$  computed by Q-DEIM into DNNand utilizing the output of the DNN as the function approximator  $\hat{u}$  to construct the dictionary  $\Theta$  and to compute the  $\frac{\partial \hat{u}}{\partial t}$  via automatic differentiation (c) estimating the coefficient vector  $\xi$  via stochastic gradient descent subject to the loss function  $\mathcal{L}$ .

## 5 Simulation Results

In this section we employ GN-SINDy algorithm on several PDEs including, Burgers' equation, Allen-Cahn equation, and Korteweg-de Vries equation. We also compare the results of the GN-SINDy with DeePyMoD [4] where samples are selected randomly for PDE discovery. The simulation examples have different level of complexity and non-linearity. We emphasize that, within our framework, we do not take into account the boundary conditions of PDEs. This is because our primary objective is not to solve these equations but rather to discover the governing equation. Each example undergoes a two-part simulation process. Initially, we apply Q-DEIM to the provided snapshot matrix of a PDE to pinpoint the most informative samples in the first part.

Since we have no prior knowledge of the various hyperparameters required for GN-SINDy, we do some trial and error to find the right configuration setup. The Q-DEIM algorithm, which filters out informative samples from unimportant ones, is the main step in our procedure. A suitable trade-off must be considered for the time division  $t_{\text{div}}$  and the precision value  $\epsilon_{\text{thr}}$ , since the cardinality of the sample set can be varied by setting these parameters. We first fix the time division parameter  $t_{\text{div}}$  for each example and vary the precision value  $\epsilon_{\text{thr}}$  to obtain the most informative samples corresponding to each example. Then we randomly select a part of these informative samples to train the DNN structure for PDE discovery. It is worth highlighting that the cardinality of the candidate informative samples used in the GN-SINDy algorithm is considered as the minimum possible value needed to recover the PDE. We identify this minimum number by testing different setups for Q-DEIM algorithm, sparse estimator, constraint, sparsity scheduler and DNN structure. We show that using the Q-DEIM algorithm on the PDE snapshot matrix  $\mathcal{U}$  provides us with valuable insight to capture the local dynamics, so that even randomly selecting the fraction of resulting subsamples will increase the success rate of discovering the governing equation compared to the approaches that directly randomly sample the PDE snapshot matrix, such as DeePyMoD.

**Data generation.** The dataset in this article are taken from repositories reported in [3, 4]. Our data preprocessing step only includes selecting the most informative samples by employing Q-DEIM algorithm on the snapshot matrix of PDE dataset. Our procedure to select the samples has two stages: (i) we employ Q-DEIM with the appropriate setting for the precision threshold  $\epsilon_{\text{thr}}$  and the time division  $t_{\text{div}}$ , (ii) we choose randomly portion of the Q-DEIM samples for the training loop of the GN-SINDy. For each PDE we also mention the original range of the space-time domain dataset.

**Architecture.** Our method utilizes multi-layer perceptron networks equipped with periodic activation functions, particularly adopting the **SIREN** architecture introduced by [57]. This approach facilitates the extraction of implicit representations from measurement data, allowing us to customize the number of hidden layers and neurons for each specific example. We specify number of hidden layers and associated number of neurons for each PDE.

**Hardware.** Our endeavor to uncover governing equations via neural network training and parameter estimation necessitated the utilization of cutting-edge computational resources. To this end, we employed an NVIDIA® RTX A4000 GPU, with its robust 16 GB of RAM, to handle the computationally demanding aspects of our research. For CPU-intensive tasks, such as data generation, we harnessed the power of a 12th Gen Intel® Core™ i5-12600K processor, equipped with a remarkable 32 GB of memory.

**Training set-up.** We use the Adam optimizer which is a popular optimization algorithm in deep learning. The optimizer is configured with `learning_rate = 10-3`, signifying the step size for updating the model's parameters during training. Notably, the `beta` values for the exponential moving averages were set to 0.99, placing increased emphasis on past gradients in the optimization process. Additionally, the `amsgrad` variant was enabled, ensuring the stability of the optimizer's update rule for the moving average of squared gradients. These hyperparameter choices were made to enhance the convergence and performance of the neural network model, demonstrating a thoughtful consideration of optimization strategies in the pursuit of improved computational results. Specification of the dictionary  $\Theta$ , the sparsity estimator to update the sparsity mask  $g$ , and the constraint on the coefficient vector  $\xi$  for the PDE discovery together with their related parameters will be given accordingly. Furthermore, number of time domain division ( $t_{div}$ ) to apply Q-DEIM for each PDE dataset, precision value ( $\epsilon_{thr}$ ), and maximum number of iterations (`max-iter`) will be mentioned separately for each example. Lastly, to ensure reproducibility in our experiments, random number generators were seeded using specific values. The NumPy library was initialized with the seed 42, and the PyTorch library with the seed 50 in all the simulation examples. This deliberate seeding allows for the precise replication of random processes, facilitating the verification and validation of our results by other researchers.

## 5.1 Burgers' equation

Burgers' equation, named after the Dutch mathematician and physicist Jan Martinus Burgers, is a fundamental PDE that arises in the study of fluid dynamics and nonlinear waves. Introduced in 1948, this equation represents a simplified model for one-dimensional, inviscid fluid flow with small amplitude and shallow water conditions. Burgers' equation combines elements of both the linear advection equation and the nonlinear conservation law, making it a versatile tool in the analysis of various physical phenomena. Burgers' equation is often expressed as:

$$\frac{\partial u}{\partial t} + \gamma u \frac{\partial u}{\partial x} + \nu \frac{\partial^2 u}{\partial x^2} = 0, \quad (8)$$

where  $u(x, t)$  represents the velocity field of the fluid at a spatial location  $x$  and time  $t$  and  $\gamma$  and  $\nu$  are constant with nominal values  $\gamma = -1$  and  $\nu = 0.1$ , respectively.

The physical interpretation of Burgers' equation can be understood by considering the propagation of a wave in a fluid. As the wave propagates, the fluid particles experience shearing forces due to the presence of the wave. These shearing forces cause a change in the velocity of the fluid particles, which is governed by Burgers' equation.

It is worth to highlight that Burgers' equation has been applied to model a wide range of physical phenomena, including: fluid flows, traffic flow, and population dynamics [58]. Moreover, it exhibits several important properties that make it a versatile tool for modeling physical phenomena. In particular, the nonlinear term  $u \frac{\partial u}{\partial x}$  gives rise to the formation of shock waves, which are characterized by sharp discontinuities in the velocity field. The presence of the diffusion term  $\frac{\partial^2 u}{\partial x^2}$  imparts a dissipative nature to the equation. This means that the amplitude of the waves will decay over time, reflecting the energy dissipation due to viscosity. The equation also exhibits dispersion,

meaning that the speed of the waves depends on their wavelength. This is a consequence of the linear advection term  $u \frac{\partial u}{\partial x}$ . These properties make Burgers' equation a powerful tool for analyzing a wide range of physical phenomena, from fluid flows to traffic patterns to population dynamics.

**Analysis** To generate the dataset for our simulation we use the solution of Burgers' equation<sup>1</sup>. The spatial domain  $x \in [-8, 8]$  and the time domain  $t \in [0.5, 10]$  are considered each having 100 samples, therefore the associate snapshot matrix  $\mathcal{U} \in \mathbb{R}^{100 \times 100}$ . For the simulation setup of GN-SINDy we consider a 4 layer DNN structure each layer having 64 neurons, a dictionary consist of the combination of polynomial and derivative terms up to order 2, a sparsity scheduler with `periodicity = 100`, `patience = 500`, a sparse estimator `STRidge` with the hyperparameters maximum iteration 100 and sparsity threshold  $\delta_{\text{spr}} = 0.05$ , a constraint `STRidge` with maximum iteration `max-iterSTR = 100` and tolerance `tolSTR = 0.05`. For both sparse estimator and constraint the regularization weight set to  $\lambda_{\text{STR}} = 0$ .

Moreover, the maximum iteration of the training loop is considered `max-iter = 25000`. With the mentioned settings our dictionary has the following terms

$$[1, \mathbf{u}, \mathbf{u}_{xx}, \mathbf{u}, \mathbf{u}\mathbf{u}_x, \mathbf{u}\mathbf{u}_{xx}, \mathbf{u}^2, \mathbf{u}^2\mathbf{u}_x, \mathbf{u}^2\mathbf{u}_{xx}].$$

In a set of experiments, we consider to choose different values for the precision  $\epsilon_{\text{thr}}$  of the Q-DEIM algorithm with time division  $t_{\text{div}} = 2$  to see how it affects GN-SINDy performance. Set of different values are  $\epsilon_{\text{thr}} = \{10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$  which results 121, 180, 245 and 1313 samples respectively.

With a few times trial and error we identify Burgers' equation with 50 random samples among total of 245 samples corresponding to  $\epsilon_{\text{thr}} = 10^{-5}$ . Applying the Q-DEIM algorithm with the mentioned setup on the snapshot matrix reveals that the valuable parts of the Burgers' equation lies on the interval  $x \in [0, 5]$ . In [Figure 2](#) we show the result of Q-DEIM algorithm corresponding to  $t_{\text{div}} = 2$  and  $\epsilon_{\text{thr}} = 10^{-5}$ .

To evaluate the GN-SINDy performance corresponding to each precision value  $\epsilon_{\text{thr}}$  we choose 50 random samples among the samples that Q-DEIM provided, for the training of our DNN structure. The results of the simulations is shown in the [Table 1](#). This result proves the importance of choosing right precision value  $\epsilon_{\text{thr}}$  to select the most informative samples. Decreasing precision value  $\epsilon_{\text{thr}}$  forces Q-DEIM algorithm to choose more samples form the snapshot matrix that these samples may not carry valuable information, therefore increases the failure rate of the GN-SINDy. In addition, with the results of [Table 1](#), we identify the proper setting for the Q-DEIM algorithm hyperparameters which are  $\epsilon_{\text{thr}} = 10^{-5}$  and  $t_{\text{div}} = 2$ .

Table 1: GN-SINDy performance with different precision value  $\epsilon_{\text{thr}}$ , fixed  $t_{\text{div}} = 2$  and fixed sample size 50 for recovering Burgers' equation

Precision value	Estimated PDE
$\epsilon_{\text{thr}} = 10^{-3}$	$\mathbf{u}_t + 0.0993\mathbf{u}_{xx} - 1.0037\mathbf{u}\mathbf{u}_{xx} = 0$
$\epsilon_{\text{thr}} = 10^{-4}$	$\mathbf{u}_t + 0.1019\mathbf{u}_{xx} - 1.0236\mathbf{u}\mathbf{u}_{xx} = 0$
$\epsilon_{\text{thr}} = 10^{-5}$	$\mathbf{u}_t + 0.0985\mathbf{u}_{xx} - 0.9857\mathbf{u}\mathbf{u}_{xx} = 0$
$\epsilon_{\text{thr}} = 10^{-6}$	$\mathbf{u}_t + 0.0983\mathbf{u}_{xx} - 0.9969\mathbf{u}\mathbf{u}_{xx} = 0$

In a more advanced analysis, we contemplate exploring how the performance of the GN-SINDy algorithm is influenced by varying sparse estimators and constraints. To do so, we maintain the simulation setup as previously specified and evaluate diverse combinations of sparse estimators and constraints, including `STRidge`, `LASSO`, and `OLS`. The sparsity threshold for sparse estimator of both types `STRidge` and `LASSO` is set to  $\delta_{\text{thr}} = 0.05$ . From the results reported in [Table 2](#) we can conclude that choosing `STRidge` as the sparsity estimator has a significant impact on recovering right PDE model for Burgers' equation. Moreover, these results show that choosing most informative samples has to be combined with the suitable choice of sparse estimator and constraint so that GN-SINDy can recover the model correctly.

<sup>1</sup>[https://www.iist.ac.in/sites/default/files/people/IN08026/Burgers\\_equation\\_viscous.pdf](https://www.iist.ac.in/sites/default/files/people/IN08026/Burgers_equation_viscous.pdf)

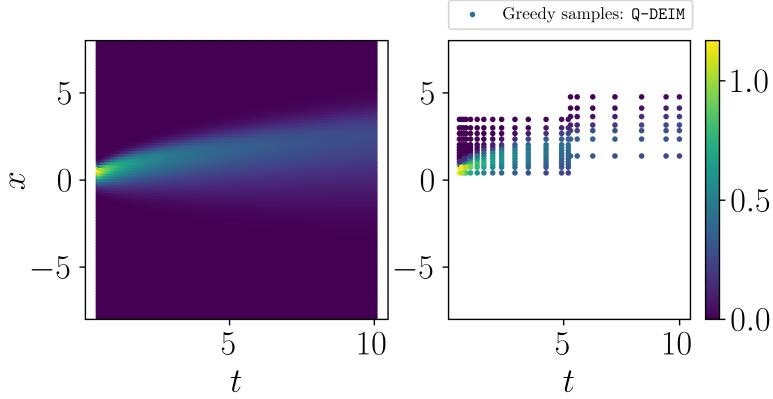


Figure 2: (left) Entire dataset ; (right) Greedy samples resulted by Q-DEIM algorithm for Burgers' equation with  $t_{\text{div}} = 2$  and  $\epsilon_{\text{thr}} = 10^{-5}$ .

Table 2: GN-SINDy performance with different sparse estimator and different constraint under fixed Q-DEIM setting  $\epsilon_{\text{thr}} = 10^{-5}$ ,  $t_{\text{div}} = 2$  and fixed sample size 50 for recovering Burgers' equation

Sparse estimator	Constraint	Estimated PDE
STRidge	STRidge	$\mathbf{u}_t + 0.0985\mathbf{u}_{xx} - 0.9857\mathbf{u}\mathbf{u}_{xx} = 0$
LASSO	STRidge	$\mathbf{u}_t + 0.1014\mathbf{u}_{xx} - 0.7531\mathbf{u}\mathbf{u}_x - 0.2754\mathbf{u}^2\mathbf{u}_x = 0$
STRidge	OLS	$\mathbf{u}_t + 0.1014\mathbf{u}_{xx} - 1.0018\mathbf{u}\mathbf{u}_x = 0$
LASSO	OLS	$\mathbf{u}_t + 0.1011\mathbf{u}_{xx} - 0.7481\mathbf{u}\mathbf{u}_x - 0.2798\mathbf{u}^2\mathbf{u}_x = 0$

In the next phase of simulation we consider to evaluate the performance of GN-SINDy under variation of the DNN structure. In this regard we fixed the previous setup, i.e. precision value  $\epsilon_{\text{thr}} = 10^{-5}$ ,  $t_{\text{div}} = 2$ , selecting 50 samples from samples earned by Q-DEIM algorithm, sparse estimator and constraint both of the type STRidge. Moreover, the number of hidden layers is fixed to 4 and we vary the number of neurons in each layer based on a geometric sequence with a common ratio of 2 with initial value 8 neurons. As reported in Table 3, the results show that all the configurations have a good performance and the combination of STRidgeas the sparse estimator and as the constraint is robust under alteration of DNN structure except for the case of 8 neurons in each layer.

Table 3: GN-SINDy performance with different DNN structure, sparse estimator STRidge, constraint STRidge, fixed precision value  $\epsilon_{\text{thr}} = 10^{-5}$ , fixed  $t_{\text{div}} = 2$  and fixed sample size 50 for recovering Burgers' equation

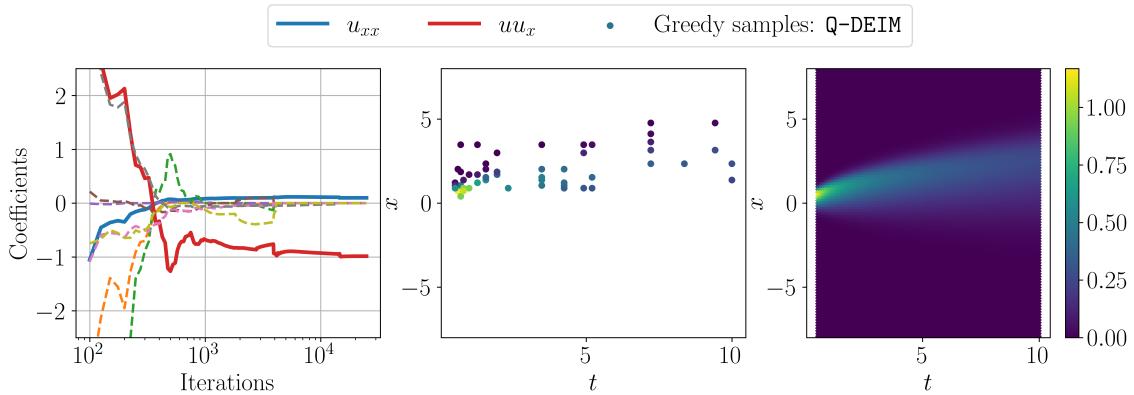
DNN structure	Estimated PDE
[2, 8, 8, 8, 8, 1]	$\mathbf{u}_t - 0.1014\mathbf{u}_{xx} - 0.9054\mathbf{u}\mathbf{u}_x - 0.2230\mathbf{u}^2\mathbf{u}_{xx} = 0$
[2, 16, 16, 16, 16, 1]	$\mathbf{u}_t + 0.1006\mathbf{u}_{xx} - 1.0086\mathbf{u}\mathbf{u}_{xx} = 0$
[2, 32, 32, 32, 32, 1]	$\mathbf{u}_t + 0.1004\mathbf{u}_{xx} - 1.0091\mathbf{u}\mathbf{u}_{xx} = 0$
[2, 64, 64, 64, 64, 1]	$\mathbf{u}_t + 0.0985\mathbf{u}_{xx} - 0.9857\mathbf{u}\mathbf{u}_{xx} = 0$

Finally, we compare the performance of GN-SINDy and DeePyMoD([4]). The same setup is assumed for both algorithms, with the knowledge that DeePyMoD uses random sampling to choose samples from snapshot matrix, sparse estimator of the type LASSO and constraint of the type OLS. In Figure 3a, and Figure 3b the results of comparison between GN-SINDy and DeePyMoD are shown. It is quite obvious how Q-DEIM algorithm assist GN-SINDy with 0.5%(50/10000) of the entire dataset

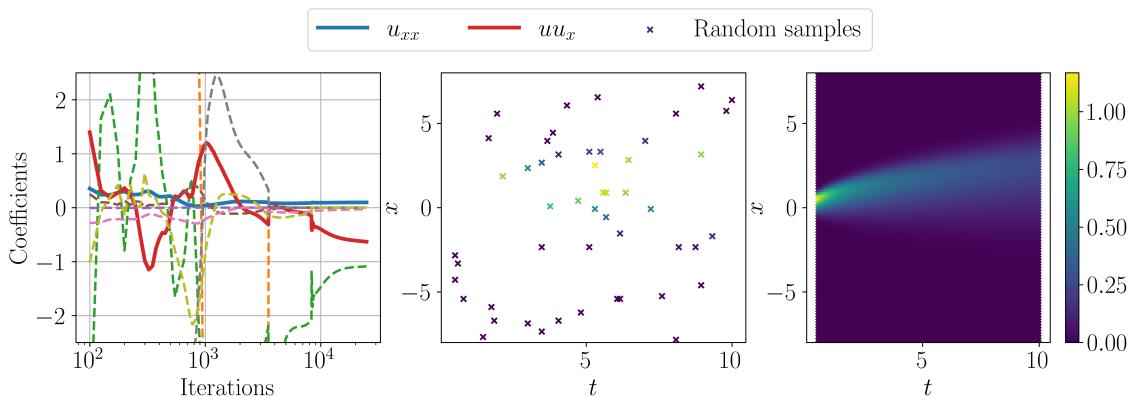
to successfully recover the right model. In particular, from Figure 3a (left) as the evolution of the coefficient demonstrates the  $\mathbf{u}_{xx}$  term is identified earlier while  $\mathbf{u}\mathbf{u}_x$  has more fluctuations. Precise value of the estimated coefficients for Burgers' equation with sampling size 50 and different methods are reported in Table 4. From this result we see that GN-SINDy outperforms DeePyMoD and can recover the Burgers' equation with acceptable precision.

Table 4: Comparing GN-SINDy and DeePyMoD for Burgers' PDE discovery

Algorithm	Estimated PDE
GN-SINDy	$\mathbf{u}_t + 0.0985\mathbf{u}_{xx} - 0.9857\mathbf{u}\mathbf{u}_x = 0$
DeePyMoD[4]	$\mathbf{u}_t - 0.0287\mathbf{u}_x + 0.0971\mathbf{u}_{xx} - 0.6326\mathbf{u}\mathbf{u}_x - 1.0872\mathbf{u}^2\mathbf{u}_x = 0$



(a) GN-SINDy performance: (right) Entire dataset; (middle) selection of 50 Greedy samples resulted by Q-DEIM algorithm for Burgers' equation with  $t_{div} = 2$  and  $\epsilon_{thr} = 10^{-5}$ ; (left) estimated coefficients with GN-SINDy through the iterations.



(b) DeePyMoD performance: (right) Entire dataset; (middle) selection of 50 random samples for Burgers' equation; (left) estimated coefficients with DeePyMoD through the iterations.

Figure 3: Comparison of the GN-SINDy performance with DeePyMoD in Burgers' equation model discovery

## 5.2 Allen-Cahn equation

The Allen-Cahn equation stands as a cornerstone of phase transition modeling, capturing the transformation of physical quantities, commonly referred to as order parameters, as a material undergoes transitions from one phase to another. Its elegance encompasses phenomena such as solidification, crystal growth, and the emergence of domain patterns in magnetic materials.

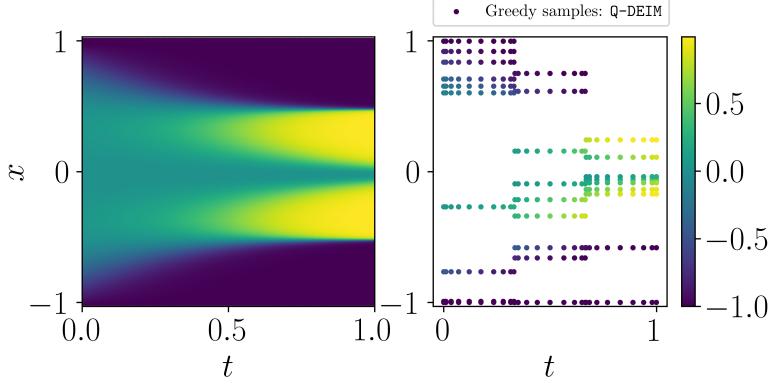


Figure 4: (left) Entire dataset ; (right) Greedy samples resulted by Q-DEIM algorithm for Allen-Cahn equation with  $t_{\text{div}} = 3$  and  $\epsilon_{\text{thr}} = 10^{-7}$ .

The Allen-Cahn equation is typically expressed in one dimension as

$$\frac{\partial u}{\partial t} + \gamma_1 u_{xx} + \gamma_2(u - u^3) = 0, \quad (9)$$

where  $u$  represents the order parameter, quantifying the level of order in the system, with nominal values  $\gamma_1 = 0.0001$  and  $\gamma_2 = 5$ . For example in a binary alloy,  $u = 1$  corresponds to a pure A-phase,  $u = -1$  corresponds to a pure B-phase, and 0 represents a mixture of A and B. The Allen-Cahn equation extends its reach beyond phase transitions, finding applications in various domains, such as magnetic Domain Formation [59], phase Separation in Binary Alloys [60], wetting of Surfaces [61], and pattern Formation in Biological Systems[62].

**Analysis** For Allen-Cahn equation we use the dataset that is reported in [4]. The PDE is discretized in 512 spatial points and 201 time instances, therefore the snapshot matrix  $\mathcal{U} \in \mathbb{R}^{512 \times 201}$  which demonstrates the curse of dimensionality regarding PDE discovery. The setup of GN-SINDy algorithm for Allen-Cahn equation is as follows: a 4 layer DNN structure each having 64 neurons, a sparsity scheduler with `patience` = 1000 and `periodicity` = 100, both sparsity estimator and the constraint are of the type STRidge with `max-iterSTR` = 100 and tolerance `tolSTR` = 0.1, a dictionary with polynomial and derivative terms up to order 3. The maximum iteration for training loop is set to `max-iter` = 25000. We note that the dictionary terms are

$$[1, \mathbf{u}_x, \mathbf{u}_{xx}, \mathbf{u}_{xxx}, \mathbf{u}, \mathbf{uu}_x, \mathbf{uu}_{xx}, \mathbf{uu}_{xxx}, \mathbf{u}^2, \mathbf{u}^2\mathbf{u}_x, \mathbf{u}^2\mathbf{u}_{xx}, \mathbf{u}^2\mathbf{u}_{xxx}, \mathbf{u}^3, \mathbf{u}^3\mathbf{u}_x, \mathbf{u}^3\mathbf{u}_{xx}, \mathbf{u}^3\mathbf{u}_{xxx}].$$

To find out the minimum number of samples that is required to discover the Allen-Cahn equation we investigate the impact of precision value  $\epsilon_{\text{thr}}$  on the performance of GN-SINDy. In this regard we consider to evaluate GN-SINDy with the following set of precision values  $\epsilon_{\text{thr}} = \{10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}\}$ . Q-DEIM algorithm based on these precision values will result 147, 209, 262, and 386 informative samples, respectively. With a few times trial and error we discover Allen-Cahn PDE with 120 samples out of total 262 samples corresponding to the precision value  $\epsilon_{\text{thr}} = 10^{-7}$ . With this setting, These 262 valuable samples are shown in Figure 4. From this figure we see how Q-DEIM algorithm selects informative samples to capture the local dynamics at each part of the domain. Moreover, it shows the importance of the time domain division for the sample selection of the Q-DEIM algorithm.

To see how the other GN-SINDy configurations perform under different precision values we choose 120 samples randomly out of 147, 209, and 386 and we evaluate the performance of the algorithm again. The results are reported in Table 5 which demonstrates that  $\epsilon_{\text{thr}} = 10^{-7}$  has a better performance with respect to the other cases. Our purpose to do comparison for different precision value  $\epsilon_{\text{thr}}$  is to find the best setting among all the possible combinations for the PDE discovery.

In the more quantitative analysis, we evaluate the performance of GN-SINDy under different choices of the sparse estimator and the constraint. For the Q-DEIM settings we consider  $\epsilon_{\text{thr}} = 10^{-7}$ ,

Table 5: GN-SINDy performance with different precision value  $\epsilon_{\text{thr}}$ ,  $t_{\text{div}} = 3$  and fixed sample size 120 for recovering Allen-Cahn equation

Precision value	Estimated PDE
$\epsilon_{\text{thr}} = 10^{-5}$	$\mathbf{u}_t + 0.0000\mathbf{u}_{xx} + 4.8759\mathbf{u} - 4.8517\mathbf{u}^3 = 0$
$\epsilon_{\text{thr}} = 10^{-6}$	$\mathbf{u}_t + 0.0000\mathbf{u}_{xx} + 4.9379\mathbf{u} - 4.9026\mathbf{u}^3 = 0$
$\epsilon_{\text{thr}} = 10^{-7}$	$\mathbf{u}_t + 0.0000\mathbf{u}_{xx} + 4.9643\mathbf{u} - 4.9328\mathbf{u}^3 = 0$
$\epsilon_{\text{thr}} = 10^{-8}$	$\mathbf{u}_t + 0.0000\mathbf{u}_{xx} + 4.8733\mathbf{u} - 4.8171\mathbf{u}^3 = 0$

$t_{\text{div}} = 3$  and we randomly select 120 samples for the DNN training. In addition, the DNN structure is considered fixed as mentioned earlier, i.e. a 4 layer DNN with 64 neurons in each layer. In [Table 6](#) we report the results of these simulations. Almost all the configurations have the same performance, however the combination of LASSO as the sparse estimator and OLS as the constraint is slightly better than the others while the combination of LASSO and STRidge has .

Table 6: GN-SINDy performance with different sparse estimator and different constraint under fixed Q-DEIM setting  $\epsilon_{\text{thr}} = 10^{-7}$ ,  $t_{\text{div}} = 3$  and fixed sample size 120 for recovering Allen-Cahn equation

Sparse estimator	Constraint	Estimated PDE
STRidge	STRidge	$\mathbf{u}_t + 0.0000\mathbf{u}_{xx} + 4.9643\mathbf{u} - 4.9328\mathbf{u}^3 = 0$
LASSO	STRidge	$\mathbf{u}_t + 0.0000\mathbf{u}_{xx} + 4.8477\mathbf{u} - 4.7882\mathbf{u}^3 = 0$
STRidge	OLS	$\mathbf{u}_t + 0.0000\mathbf{u}_{xx} + 4.9790\mathbf{u} - 4.9709\mathbf{u}^3 = 0$
LASSO	OLS	$\mathbf{u}_t + 0.0000\mathbf{u}_{xx} + 5.0175\mathbf{u} - 4.9724\mathbf{u}^3 = 0$

We also consider to evaluate the GN-SINDy algorithm when we vary the structure of the DNN. To do so, we fix the number of hidden layers to 4 and we vary the number of neurons in each layer based on a geometric sequence with a common ration of 2 with initial value 8 neurons. The results of this experiments are reported in [Table 7](#) where we GN-SINDy is robust to the DNN alteration.

Table 7: GN-SINDy performance with different DNN structure, sparse estimator STRidge, constraint STRidge, fixed precision value  $\epsilon_{\text{thr}} = 10^{-7}$ , fixed  $t_{\text{div}} = 3$  and fixed sample size 120 for recovering Allen-Cahn equation

DNN structure	Estimated PDE
[2, 8, 8, 8, 8, 1]	$\mathbf{u}_t + 0.0000\mathbf{u}_{xx} + 5.0275\mathbf{u} - 4.9655\mathbf{u}^3 = 0$
[2, 16, 16, 16, 16, 1]	$\mathbf{u}_t + 0.0000\mathbf{u}_{xx} + 4.9901\mathbf{u} - 4.9428\mathbf{u}^3 = 0$
[2, 32, 32, 32, 32, 1]	$\mathbf{u}_t + 0.0000\mathbf{u}_{xx} + 4.9925\mathbf{u} - 4.9481\mathbf{u}^3 = 0$
[2, 64, 64, 64, 64, 1]	$\mathbf{u}_t + 0.0000\mathbf{u}_{xx} + 4.9643\mathbf{u} - 4.9328\mathbf{u}^3 = 0$

Moreover, we consider to do a comparison between GN-SINDy and DeePyMoD ([\[4\]](#)) to see how they perform in recovering the PDE model with the same sample size. The same setup is considered for DeePyMoD with its default sparse estimator as LASSO and constraint of the type OLS. The results of the simulation are presented in [Table 8](#), that we see clearly GN-SINDy outperforms DeePyMoD to recover the Allen-Cahn PDE with acceptable precision. It is worth to highlight that the coefficient corresponding to the term  $\mathbf{u}_{xx}$  is considerably less than the coefficient corresponding the terms  $\mathbf{u}$  and  $\mathbf{u}^3$  in the ground truth equation, hence recovering such a small coefficient it is quite difficult for the DNN. Moreover, in [Figure 5a](#) and [Figure 5b](#) the selected greedy samples for GN-SINDy, random samples for DeePyMoD and evolution of different coefficients through the iteration are shown. The impact of Q-DEIM algorithm in the success of recovering correct coefficients is quite clear. Vice versa DeePyMoD has a weak performance due to the usage of randomly selected samples in the

training loop of the DNN structure. Since **DeePyMoD** does not have a comparable results versus **GN-SINDy** therefore we do not consider it for the rest of this section.

Table 8: **GN-SINDy** performance with different precision value  $\epsilon_{\text{thr}}$ ,  $t_{\text{div}} = 3$  and fixed sample size 120 for recovering Allen-Cahn equation

Algorithm	Estimated PDE
<b>GN-SINDy</b>	$\mathbf{u}_t + 0.0000\mathbf{u}_{xx} + 4.9643\mathbf{u} - 4.9328\mathbf{u}^3 = 0$
<b>DeePyMoD[4]</b>	$\mathbf{u}_t + 0.8913 + 14.6374\mathbf{u}_x - 3.2769\mathbf{u}_{xx} + 1.5190\mathbf{u}_{xxx} - 7.7121\mathbf{u} + 26.0523\mathbf{u}\mathbf{u}_x - 10.0496\mathbf{u}\mathbf{u}_{xx} + 1.4176\mathbf{u}\mathbf{u}_{xxx} - 43.0762\mathbf{u} + 26.6733\mathbf{u}^2 + 1.9071\mathbf{u}^2\mathbf{u}_x - 0.3155\mathbf{u}^2\mathbf{u}_{xx} - 32.1551\mathbf{u}^3 - 12.7150\mathbf{u}^3\mathbf{u}_x + 0.3304\mathbf{u}^3\mathbf{u}_{xx} - 0.8281\mathbf{u}^3\mathbf{u}_{xxx} = 0$

### 5.3 Korteweg-de Vries (KdV) equation

The Korteweg-de Vries (KdV) equation is a nonlinear, dispersive partial differential equation that plays a central role in mathematical physics and nonlinear science. It was first introduced in 1895 by Dutch physicists Diederik Korteweg and Gustav de Vries to describe the propagation of shallow water waves. Since then, it has found applications in a wide range of physical systems, including plasma physics, optics, and condensed matter physics. The KdV equation is remarkable for its integrability, meaning that it can be solved exactly using various mathematical techniques. This property has led to the discovery of solitons, which are stable, localized waves that propagate without changing shape. Solitons have been observed in various physical systems, including shallow water waves, optical fibers, and Bose-Einstein condensates. The KdV equation has also been used to develop a deeper understanding of nonlinear dynamics. It has been shown that the KdV equation can exhibit a variety of complex behaviors, such as chaos and self-organization. These behaviors have relevance to a wide range of phenomena in physics, biology, and social sciences. The KdV equation remains a cornerstone of mathematical physics and nonlinear science, providing a rich and insightful framework for understanding nonlinear waves and their diverse applications [63]. It can be written as a partial differential equation in the form:

$$\frac{\partial u}{\partial t} + c u u_x + \alpha u_{xxx} = 0, \quad (10)$$

where  $c$  and  $\alpha$  are constant variables.

**Analysis** The discretization of the KdV equation is outlined in [3], which involves 512 spatial points and 201 temporal points within the original spatial range of  $x \in [-30, 30]$  and a time span of  $t \in [0, 20]$ . The snapshot matrix  $\mathcal{U} \in \mathbb{R}^{512 \times 201}$  illustrates the challenge of dimensionality in training the DNN for PDE discovery. In the first set of experiment to identify the most promising configuration of the **GN-SINDy** to recover the KdV equation, we fix the time division  $t_{\text{div}} = 2$ , and the following set of precision values are considered  $\epsilon_{\text{thr}} = \{5 \times 10^{-5}, 10^{-5}, 10^{-6}, 1 \times 10^{-7}\}$ . Now we need to set the **GN-SINDy** hyperparameters which include the following settings: to choose the polynomial order and the derivative orders for dictionary, to opt the type of sparse estimator and the constraint, and finally the setting corresponding to the DNN structure. A 4 layer DNN structure is considered where each layer have 32 neurons, sparse estimator and constraint of the type **STRidge** with **max-iter<sub>STR</sub>** = 100 and tolerance **tol<sub>STR</sub>** = 0.1, sparse scheduler with **patience** = 1000 and **periodicity** = 50, a dictionary with polynomial terms up to order 2 and derivative terms up to order 3. The maximum iteration of the DNN training loop is considered **max-iter** = 25000. The **Q-DEIM** algorithm provides 5725, 14450, 18432 and 19801 samples corresponding to each precision value. This set of filtered samples shows the difficulty of choosing proper number of informative samples for PDE discovery. With the proposed settings for the order of polynomial as well as derivative terms of the dictionary the associated columns are

$$[1, \mathbf{u}_x, \mathbf{u}_{xx}, \mathbf{u}_{xxx}, \mathbf{u}, \mathbf{u}\mathbf{u}_x, \mathbf{u}\mathbf{u}_{xx}, \mathbf{u}\mathbf{u}_{xxx}, \mathbf{u}^2, \mathbf{u}^2\mathbf{u}_x, \mathbf{u}^2\mathbf{u}_{xx}, \mathbf{u}^2\mathbf{u}_{xxx}].$$

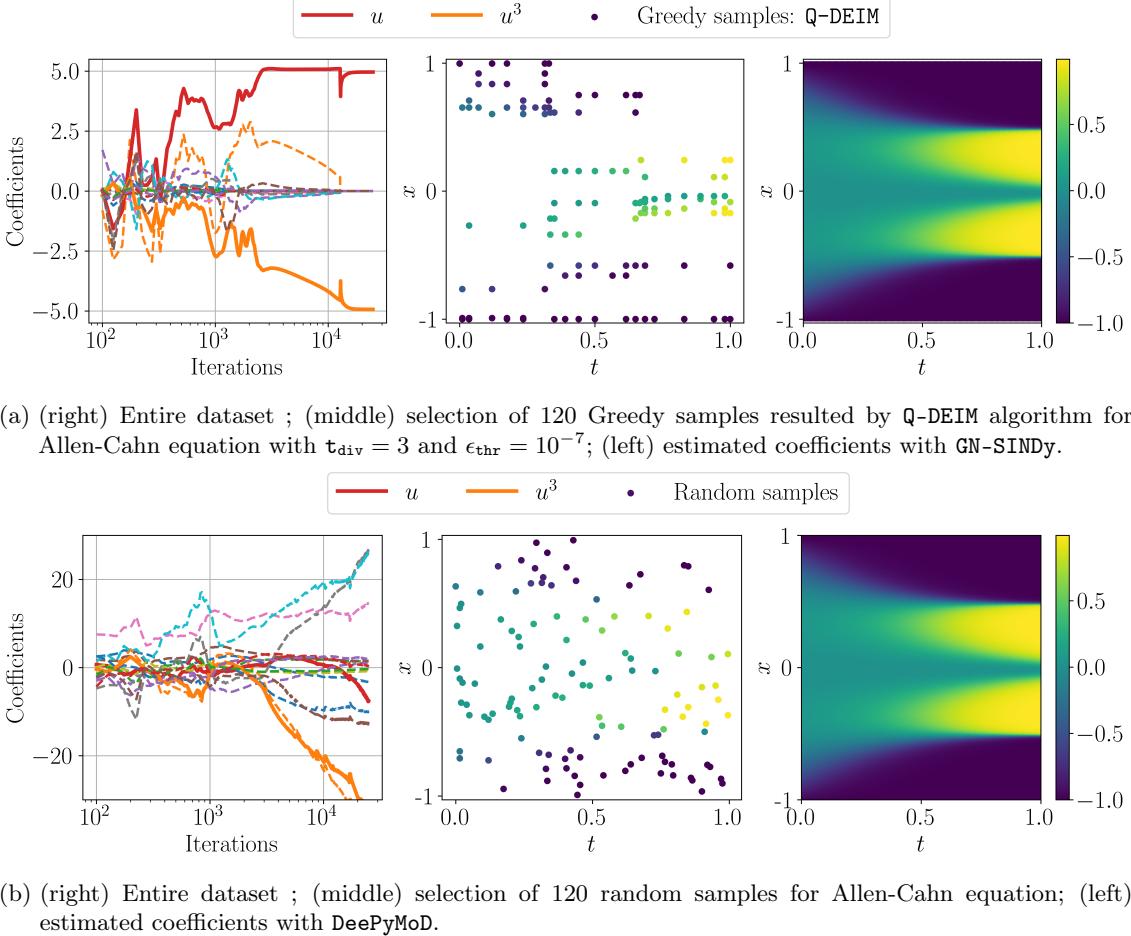


Figure 5: Comparison of the GN-SINDy performance with DeePyMoD in Allen-Cahn equation model discovery

With the mentioned GN-SINDy settings and among these set of informative samples corresponding to each precision value we randomly select portion of each dataset to identify the smallest possible number that can successfully recover the KdV equation. After a few times trial and error we realize to take 900 samples from total number of informative samples corresponding to  $\epsilon_{\text{thr}} = 10^{-5}$  that results correct KdV equation. Note that we only use 0.87% of the entire dataset approximately which shows significant reduction in computational cost. The outcomes of this experiment is shown in the Table 9, which demonstrate that choosing the precision value  $\epsilon_{\text{thr}} = 10^{-5}$  has the better performance with respect to the other choices. Moreover, these results prove the importance of choosing right precision value  $\epsilon_{\text{thr}}$  in the success rate of GN-SINDy algorithm. Indeed decreasing  $\epsilon_{\text{thr}}$  selects the samples that may not contribute significantly into the model discovery. Moreover, Figure 6 shows the entire dataset and selected samples by Q-DEIM algorithm corresponding to  $t_{\text{div}} = 2$  and  $\epsilon_{\text{thr}} = 10^{-5}$ .

In a more sophisticated analysis we consider to investigate the impact of sparse estimator and constraint on the KdV model discovery. To do so, we utilize the knowledge that we already acquired regarding the settings of the Q-DEIM algorithm to filter out the informative samples as well as the minimum number of cardinality that is required to recover the model. These settings are  $t_{\text{div}} = 2$ ,  $\epsilon_{\text{thr}} = 10^{-5}$  and 900 randomly selected samples out of total 14450 informative samples. In this regard different combinations of sparse estimator and constraint are considered. The results of these experiments are reported in Table 10 where we see all of the combinations have good performance.

To see how GN-SINDy performs under DNN structural variations we consider to do set of experiments. The GN-SINDy hyperparameters are set as mentioned earlier with the only alteration in

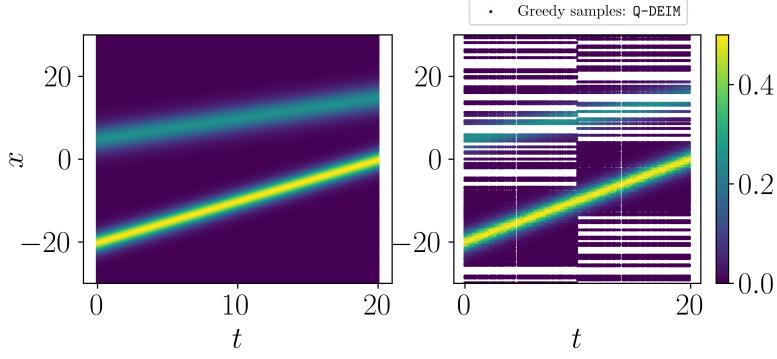


Figure 6: (left) Entire dataset ; (right) Greedy samples resulted by Q-DEIM algorithm for KdV equation with  $t_{\text{div}} = 2$  and  $\epsilon_{\text{thr}} = 10^{-5}$ .

Table 9: GN-SINDy performance with different precision value  $\epsilon_{\text{thr}}$ , time division  $t_{\text{div}} = 2$  and fixed sample size 900 for recovering KdV equation

Precision value	Estimated PDE
$\epsilon_{\text{thr}} = 5 \times 10^{-5}$	$\mathbf{u}_t - 5.9994\mathbf{u}\mathbf{u}_x - 0.9995\mathbf{u}_{xxx} = 0$
$\epsilon_{\text{thr}} = 1 \times 10^{-5}$	$\mathbf{u}_t - 5.9955\mathbf{u}\mathbf{u}_x - 0.9975\mathbf{u}_{xxx} = 0$
$\epsilon_{\text{thr}} = 1 \times 10^{-6}$	$\mathbf{u}_t - 5.9639\mathbf{u}\mathbf{u}_x - 0.9850\mathbf{u}_{xxx} = 0$
$\epsilon_{\text{thr}} = 1 \times 10^{-7}$	$\mathbf{u}_t - 5.9860\mathbf{u}\mathbf{u}_x - 0.9959\mathbf{u}_{xxx} = 0$

Table 10: GN-SINDy performance with different sparse estimator and different constraint under fixed Q-DEIM setting  $\epsilon_{\text{thr}} = 10^{-5}$ ,  $t_{\text{div}} = 2$  and fixed sample size 900 for recovering KdV equation

Sparse estimator	Constraint	Estimated PDE
STRidge	STRidge	$\mathbf{u}_t - 5.9955\mathbf{u}\mathbf{u}_x - 0.9975\mathbf{u}_{xxx} = 0$
LASSO	STRidge	$\mathbf{u}_t - 5.9899\mathbf{u}\mathbf{u}_x - 0.9952\mathbf{u}_{xxx} = 0$
STRidge	OLS	$\mathbf{u}_t - 5.9579\mathbf{u}\mathbf{u}_x - 0.9827\mathbf{u}_{xxx} = 0$
LASSO	OLS	$\mathbf{u}_t - 5.9799\mathbf{u}\mathbf{u}_x - 0.9933\mathbf{u}_{xxx} = 0$

the number of neurons in each layer. We vary the number of neurons in each layer based on a geometric sequence with a common ratio of 2 with initial value 8 neurons. The results of these experiments are reported in Table 11 where we see GN-SINDy can not recover the KdV when the number of neurons in hidden layers are 8 and 16. This reveals the challenge of selecting right DNNstructure for PDE discovery.

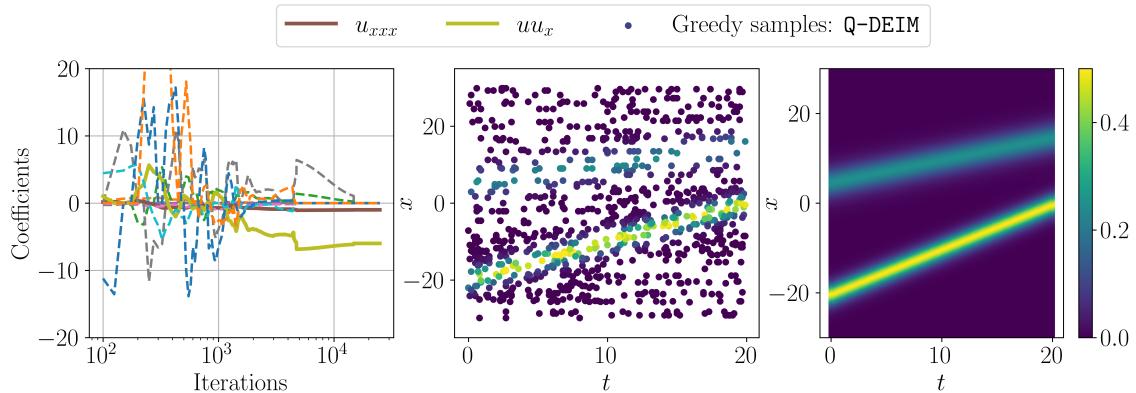
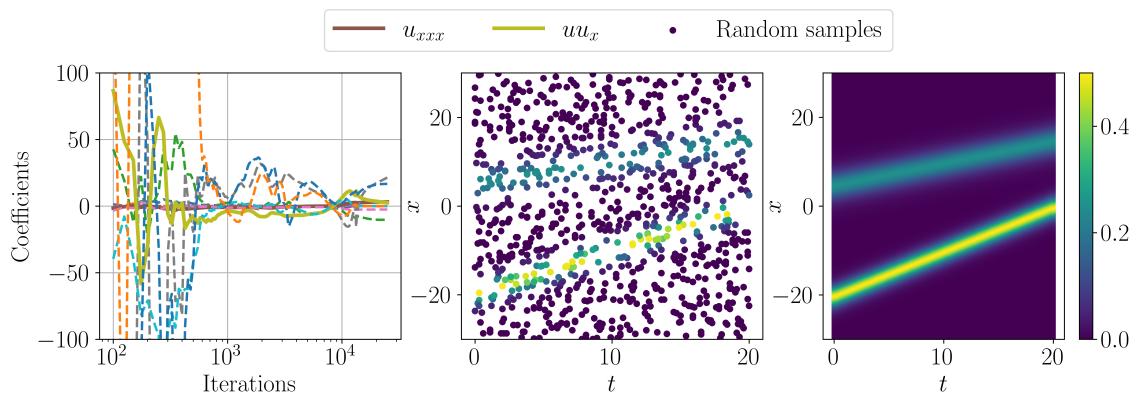
Table 11: GN-SINDy performance with different DNN structure, fixed precision value  $\epsilon_{\text{thr}} = 10^{-5}$ , fixed time division  $t_{\text{div}} = 2$ , fixed sample size 900 for recovering KdV equation, sparse estimator and constraint of type STRidge

DNN structure	Estimated PDE
[2, 8, 8, 8, 8, 1]	$\mathbf{u}_t - 0.5534\mathbf{u}_{xxx} - 8.4820\mathbf{u}\mathbf{u}_x + 11.1437\mathbf{u}^2\mathbf{u}_x = 0$
[2, 16, 16, 16, 16, 1]	$\mathbf{u}_t - 1.1290\mathbf{u}_{xxx} - 5.7801\mathbf{u}\mathbf{u}_x + 2.3874\mathbf{u}\mathbf{u}_{xxx} - 4.7362\mathbf{u}^2\mathbf{u}_{xxx} = 0$
[2, 32, 32, 32, 32, 1]	$\mathbf{u}_t - 5.9955\mathbf{u}\mathbf{u}_x - 0.9975\mathbf{u}_{xxx} = 0$
[2, 64, 64, 64, 64, 1]	$\mathbf{u}_t - 5.9812\mathbf{u}\mathbf{u}_x - 0.9923\mathbf{u}_{xxx} = 0$

Lastly we consider to compare the GN-SINDy performance with DeePyMoD([4]). The same DNN structure is considered for DeePyMoD with its default sparse estimator and constraint of the type LASSO and OLS respectively. The results of this simulation is shown in the Table 12 where we clearly see that GN-SINDy outperforms DeePyMoD and can recover the PDE with 900 greedy samples. It is worth to highlight that DeePyMoD selects the samples randomly to train its DNN structure. In Figure 7a and Figure 7b the candidate samples that are used in the training loop of each algorithm are shown in the middle graphs, while the left graphs depict the evolution of different coefficients thorough the training loop iterations corresponding to each algorithm. In particular from coefficient evolution of GN-SINDy algorithm we see that the coefficient corresponding to the term  $\mathbf{u}_{xxx}$  has a faster convergence rate respect to the coefficient corresponding to the term  $\mathbf{uu}_x$ .

Table 12: Comparing GN-SINDy and DeePyMoD for KdV PDE discovery

Algorithm	Estimated PDE
GN-SINDy	$\mathbf{u}_t - 5.9955\mathbf{uu}_x - 0.9975\mathbf{u}_{xxx} = 0$
DeePyMoD[4]	$\mathbf{u}_t - 2.5393\mathbf{u}_x + 3.1320\mathbf{u}_{xxx} + 2.5640\mathbf{uu}_x - 10.1184\mathbf{uu}_{xxx} + 21.4131\mathbf{u}^2\mathbf{u}_x + 16.6569\mathbf{u}^2\mathbf{u}_{xxx} = 0$

(a) (right) Entire dataset ; (middle) selection of 900 Greedy samples resulted by Q-DEIM algorithm for KdV equation with  $t_{div} = 2$  and  $\epsilon_{thr} = 10^{-5}$ ; (left) estimated coefficients with GN-SINDy.

(b) (right) Entire dataset ; (middle) selection of 900 random samples for KdV equation; (left) estimated coefficients with DeePyMoD.

Figure 7: Comparison of the GN-SINDy performance with DeePyMoD in KdV equation model discovery

## 6 Conclusion

A greedy sampling approach is considered in the framework of sparse identification of non linear dynamical systems (**SINDy**) which takes advantage of deep neural network (**DNN**) and its strength for the **PDE** model discovery. In particular our proposed methodology is the extension of **DeePyMoD** with integration of greedy sampling approach as the data collection and new sparse estimator and constraint functions in the training loops of the **DNN**. In particular discrete empirical interpolation method (**DEIM**) has been employed to extract the most informative samples of a snapshot matrix associated to a **PDE**. Due to the usage of greedy samples in combination of **DNN** and **SINDy** algorithm the proposed approach has been named **GN-SINDy**. Our comprehensive study on Burgers' equation, Allen-Cahn equation, and Korteweg-de Vries equation revealed that usage of greedy samples and sequential threshold ridge regression (**STRidge**) significantly increase the success rate of the model discovery algorithm. In the comparison phase, **GN-SINDy** outperformed **DeePyMoD** in all the simulation settings and we could discover Burgers' equation, Allen-Cahn equation, and Korteweg-de Vries equation with with 0.5%, 0.1%, and 0.874% of the dataset respectively. After conducting a comprehensive analysis with various simulation settings, we have uncovered a fundamental challenge associated with the discovery of **PDE**and the selection of appropriate hyperparameters.

## References

- [1] A. Forootani, P. Goyal, and P. Benner, “A robust **SINDy** approach by combining neural networks and an integral form,” *arXiv preprint arXiv:2309.07193*, 2023.
- [2] P. Goyal and P. Benner, “Neural ordinary differential equations with irregular and noisy data,” *Roy. Soc. Open Sci.*, vol. 10, no. 7, p. 221475, 2023.
- [3] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, “Data-driven discovery of partial differential equations,” *Science Advances*, vol. 3, no. 4, p. e1602614, 2017.
- [4] G.-J. Both, S. Choudhury, P. Sens, and R. Kusters, “Deepmod: Deep learning for model discovery in noisy data,” *Journal of Computational Physics*, vol. 428, p. 109985, 2021.
- [5] S. Rudy, A. Alla, S. L. Brunton, and J. N. Kutz, “Data-driven identification of parametric partial differential equations,” *SIAM Journal on Applied Dynamical Systems*, vol. 18, no. 2, pp. 643–660, 2019.
- [6] D. N. Tanyu, J. Ning, T. Freudenberg, N. Heilenkötter, A. Rademacher, U. Iben, and P. Maass, “Deep learning methods for partial differential equations and related parameter identification problems,” *Inverse Problems*, vol. 39, no. 10, p. 103001, 2023.
- [7] F. De Barros, W. Mills, and R. Cotta, “Integral transform solution of a two-dimensional model for contaminant dispersion in rivers and channels with spatially variable coefficients,” *Environmental Modelling & Software*, vol. 21, no. 5, pp. 699–709, 2006.
- [8] M. Zeneli, A. Nikolopoulos, S. Karella, and N. Nikolopoulos, “Numerical methods for solid-liquid phase-change problems,” in *Ultra-high Temperature Thermal Energy Storage, Transfer and Conversion*. Elsevier, 2021, pp. 165–199.
- [9] B. Bai, H. Ci, H. Lei, and Y. Cui, “A local integral-generalized finite difference method with mesh-meshless duality and its application,” *Engineering Analysis with Boundary Elements*, vol. 139, pp. 14–31, 2022.
- [10] J. Blechschmidt and O. G. Ernst, “Three ways to solve partial differential equations with neural networks – a review,” *GAMM-Mitteilungen*, vol. 44, no. 2, p. e202100006, 2021.
- [11] L. Lu, X. Meng, Z. Mao, and G. E. Karniadakis, “Deepxde: A deep learning library for solving differential equations,” *SIAM Review*, vol. 63, no. 1, pp. 208–228, 2021.

- [12] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.
- [13] H. Schaeffer, “Learning partial differential equations via data discovery and sparse optimization,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 473, no. 2197, p. 20160446, 2017.
- [14] J. Berg and K. Nyström, “Data-driven discovery of pdes in complex datasets,” *Journal of Computational Physics*, vol. 384, pp. 239–252, 2019.
- [15] G.-J. Both, G. Tod, and R. Kusters, “Model discovery in the sparse sampling regime,” *arXiv preprint arXiv:2105.00400*, 2021.
- [16] C. Rackauckas, Y. Ma, J. Martensen, C. Warner, K. Zubov, R. Supekar, D. Skinner, A. Ramadhan, and A. Edelman, “Universal differential equations for scientific machine learning,” *arXiv preprint arXiv:2001.04385*, 2020.
- [17] Z. Chen, Y. Liu, and H. Sun, “Physics-informed learning of governing equations from scarce data,” *Nature communications*, vol. 12, no. 1, p. 6136, 2021.
- [18] S. Seo and Y. Liu, “Differentiable physics-informed graph networks,” *arXiv preprint arXiv:1902.02950*, 2019.
- [19] M. Cranmer, S. Greydanus, S. Hoyer, P. Battaglia, D. Spergel, and S. Ho, “Lagrangian neural networks,” *arXiv preprint arXiv:2003.04630*, 2020.
- [20] R. Iten, T. Metger, H. Wilming, L. Del Rio, and R. Renner, “Discovering physical concepts with neural networks,” *Physical review letters*, vol. 124, no. 1, p. 010508, 2020.
- [21] S. L. Brunton, J. L. Proctor, and J. N. Kutz, “Discovering governing equations from data by sparse identification of nonlinear dynamical systems,” *Proc. Nat. Acad. Sci. U.S.A.*, vol. 113, no. 15, pp. 3932–3937, 2016.
- [22] J.-C. Loiseau and S. L. Brunton, “Constrained sparse Galerkin regression,” *J. Fluid Mechanics*, vol. 838, pp. 42–67, 2018.
- [23] A. A. Kaptanoglu, K. D. Morgan, C. J. Hansen, and S. L. Brunton, “Physics-constrained, low-dimensional models for magnetohydrodynamics: First-principles and data-driven approaches,” *Physical Review E*, vol. 104, no. 1, p. 015206, 2021.
- [24] S. Beetham and J. Capecelatro, “Formulating turbulence closures using sparse regression with embedded form invariance,” *Physical Review Fluids*, vol. 5, no. 8, p. 084611, 2020.
- [25] L. Zanna and T. Bolton, “Data-driven equation discovery of ocean mesoscale closures,” *Geophysical Research Letters*, vol. 47, no. 17, p. e2020GL088376, 2020.
- [26] M. Sorokina, S. Sygletos, and S. Turitsyn, “Sparse identification for nonlinear optical communication systems: Sino method,” *Optics express*, vol. 24, no. 26, pp. 30 433–30 443, 2016.
- [27] L. Boninsegna, F. Nüske, and C. Clementi, “Sparse learning of stochastic dynamical equations,” *The Journal of chemical physics*, vol. 148, no. 24, 2018.
- [28] S. Thaler, L. Paehler, and N. A. Adams, “Sparse identification of truncation errors,” *Journal of Computational Physics*, vol. 397, p. 108851, 2019.
- [29] R. K. Niven, A. Mohammad-Djafari, L. Cordier, M. Abel, and M. Quade, “Bayesian identification of dynamical systems,” *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 33, no. 1, p. 33, 2020.
- [30] D. A. Messenger and D. M. Bortz, “Weak sindy: Galerkin-based data-driven model selection,” *Multiscale Modeling & Simulation*, vol. 19, no. 3, pp. 1474–1497, 2021.

- [31] S. M. Hirsh, D. A. Barajas-Solano, and J. N. Kutz, “Sparsifying priors for bayesian uncertainty quantification in model discovery,” *Royal Society Open Science*, vol. 9, no. 2, p. 211823, 2022.
- [32] Y. Wang, H. Fang, J. Jin, G. Ma, X. He, X. Dai, Z. Yue, C. Cheng, H.-T. Zhang, D. Pu *et al.*, “Data-driven discovery of stochastic differential equations,” *Engineering*, vol. 17, pp. 244–252, 2022.
- [33] Z. Drmac and S. Gugercin, “A new selection operator for the discrete empirical interpolation method—improved a priori error bound and extensions,” *SIAM Journal on Scientific Computing*, vol. 38, no. 2, pp. A631–A648, 2016.
- [34] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera, “An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations,” *Comptes Rendus Mathematique*, vol. 339, no. 9, pp. 667–672, 2004.
- [35] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009, vol. 2.
- [36] R. Tibshirani, “Regression shrinkage and selection via the Lasso,” *J. Roy. Statist. Soc.: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [37] H. Vaddireddy, A. Rasheed, A. E. Staples, and O. San, “Feature engineering and symbolic regression methods for detecting hidden physics from sparse sensor observation data,” *Physics of Fluids*, vol. 32, no. 1, 2020.
- [38] A. Beck and Y. C. Eldar, “Sparsity constrained nonlinear optimization: Optimality conditions and algorithms,” *SIAM J. Optim.*, vol. 23, no. 3, pp. 1480–1509, 2013.
- [39] Z. Yang, Z. Wang, H. Liu, Y. Eldar, and T. Zhang, “Sparse nonlinear regression: Parameter estimation under nonconvexity,” in *Intern. Conf. on Mach. Learn.* PMLR, 2016, pp. 2472–2481.
- [40] L. Zhang and H. Schaeffer, “On the convergence of the SINDy algorithm,” *Multiscale Model. Simul.*, vol. 17, no. 3, pp. 948–972, 2019.
- [41] J. H. Halton, “On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals,” *Numerische Mathematik*, vol. 2, pp. 84–90, 1960.
- [42] W.-S. L. Tien-Tsin Wong and P.-A. Heng, “Sampling with hammersley and halton points,” *Journal of Graphics Tools*, vol. 2, no. 2, pp. 9–24, 1997.
- [43] I. M. Sobol, “On the distribution of points in a cube and the approximate evaluation of integrals,” *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 4, pp. 86–112, 1967.
- [44] C. Wu, M. Zhu, Q. Tan, Y. Kartha, and L. Lu, “A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks,” *Computer Methods in Applied Mechanics and Engineering*, vol. 403, p. 115671, 2023.
- [45] Z. Zhang, X. Yang, and G. Lin, “POD-based constrained sensor placement and field reconstruction from noisy wind measurements: A perturbation study,” *Mathematics*, vol. 4, no. 2, p. 26, 2016.
- [46] A. Moslemi, “Sparse representation learning using  $\ell_1$ -  $\ell_2$  compressed sensing and rank-revealing qr factorization,” *Engineering Applications of Artificial Intelligence*, vol. 125, p. 106663, 2023.
- [47] S. Joshi and S. Boyd, “Sensor selection via convex optimization,” *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 451–462, 2008.
- [48] J. Ranieri, A. Chebira, and M. Vetterli, “Near-optimal sensor placement for linear inverse problems,” *IEEE Transactions on signal processing*, vol. 62, no. 5, pp. 1135–1146, 2014.

- [49] S. Lau, R. Eichardt, L. Di Renzo, and J. Haueisen, “Tabu search optimization of magnetic sensor systems for magnetocardiography,” *IEEE Transactions on Magnetics*, vol. 44, no. 6, pp. 1442–1445, 2008.
- [50] Z. Wang, H.-X. Li, and C. Chen, “Reinforcement learning-based optimal sensor placement for spatiotemporal modeling,” *IEEE Transactions on Cybernetics*, vol. 50, no. 6, pp. 2861–2871, 2019.
- [51] S. Chaturantabut and D. C. Sorensen, “Nonlinear model reduction via discrete empirical interpolation,” *SIAM Journal on Scientific Computing*, vol. 32, no. 5, pp. 2737–2764, 2010.
- [52] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, “Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns,” *IEEE Control Systems Magazine*, vol. 38, no. 3, pp. 63–86, 2018.
- [53] K. Kunisch and S. Volkwein, “Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics,” *SIAM J. Numer. Anal.*, vol. 40, no. 2, pp. 492–515, 2002.
- [54] M. Gavish and D. L. Donoho, “The optimal hard threshold for singular values is  $\frac{4}{\sqrt{3}}$ ,” *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 5040–5053, 2014.
- [55] S. Huang, W. Feng, C. Tang, and J. Lv, “Partial differential equations meet deep neural networks: A survey,” *arXiv preprint arXiv:2211.05567*, 2022.
- [56] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent,” *Journal of Statistical Software*, vol. 33, no. 1, p. 1, 2010.
- [57] V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, “Implicit neural representations with periodic activation functions,” in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, ser. NIPS’20. Red Hook, NY, USA: Curran Associates Inc., 2020.
- [58] P. Kachroo, K. M. Özbay, P. Kachroo, and K. M. Özbay, “Traffic flow theory,” *Feedback Control Theory for Dynamic Traffic Assignment*, pp. 57–87, 2018.
- [59] L. Shen and G.-W. Chern, “Cell dynamics simulations of coupled charge and magnetic phase transformation in correlated oxides,” *Physical Review E*, vol. 103, no. 3, p. 032134, 2021.
- [60] J. Cahn and A. Novick-Cohen, “Evolution equations for phase separation and ordering in binary alloys,” *Journal of statistical physics*, vol. 76, pp. 877–909, 1994.
- [61] H. Zhang, Y. Wu, F. Wang, and B. Nestler, “Effect of wall free energy formulation on the wetting phenomenon: Conservative allen-cahn model,” *The Journal of Chemical Physics*, vol. 159, no. 16, 2023.
- [62] H. Zhao, B. D. Storey, R. D. Braatz, and M. Z. Bazant, “Learning the physics of pattern formation from images,” *Physical review letters*, vol. 124, no. 6, p. 060201, 2020.
- [63] X.-J. Yang, J. Tenreiro Machado, D. Baleanu, and C. Cattani, “On exact traveling-wave solutions for local fractional korteweg-de vries equation,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 26, no. 8, 2016.