Robust Low-Rank Discovery of Data-Driven Partial Differential Equations

Jun Li,[1,2] **Gan Sun,**[2] **Guoshuai Zhao,**[2] **Li-wei H. Lehman**[2,3]

[1]School of Computer Science & Engineering, Nanjing University of Science and Technology, Nanjing, China
[2]Institute for Medical Engineering & Science, MIT, Cambridge, MA, USA
[3]MIT-IBM Watson AI Lab, Cambridge, MA, USA
junl.mldl@gmail.com, {sungan, gshzhao, lilehman}@mit.edu

Abstract

Partial differential equations (PDEs) are essential foundations to model dynamic processes in natural sciences. Discovering the underlying PDEs of complex data collected from real world is key to understanding the dynamic processes of natural laws or behaviors. However, both the collected data and their partial derivatives are often corrupted by noise, especially from sparse outlying entries, due to measurement/process noise in the real-world applications. Our work is motivated by the observation that the underlying data modeled by PDEs are in fact often low rank. We thus develop a robust low-rank discovery framework to recover both the low-rank data and the sparse outlying entries by integrating double low-rank and sparse recoveries with a (group) sparse regression method, which is implemented as a minimization problem using mixed nuclear norms with ℓ_1 and ℓ_0 norms. We propose a low-rank sequential (grouped) threshold ridge regression algorithm to solve the minimization problem. Results from several experiments on seven canonical models (i.e., four PDEs and three parametric PDEs) verify that our framework outperforms the state-of-art sparse and group sparse regression methods. Code is available at https://github.com/junli2019/Robust-Discovery-of-PDEs

Machine learning plays a transformative role in analyzing and understanding dynamic processes from complex data in the natural sciences (e.g., physics, chemistry, biology and neuroscience) (Jordan and Mitchell 2015; Butler et al. 2018). Many real-world complex data can be modeled by a function $u(x, t)$ with spatial locations x and/or time points t, and its underlying partial differential equations (PDEs) provide essential foundations to govern the dynamic processes, such as, Schröinger's equations for quantum physics and chemistry, Navier-Stokes equations for fluid and gas dynamics, and FitzHugh-Nagumo models for neural excitement (Schaeffer 2017). A key problem in machine learning is to identify the governing PDEs from the complex data. However, the complex data are often corrupted by noise, especially from sparse outlying entries, due to measurement/process noise, limitations in sensor technologies and ad-hoc data collection techniques (Elhamifar and Vidal 2013). Therefore, discovering the governing PDEs from the
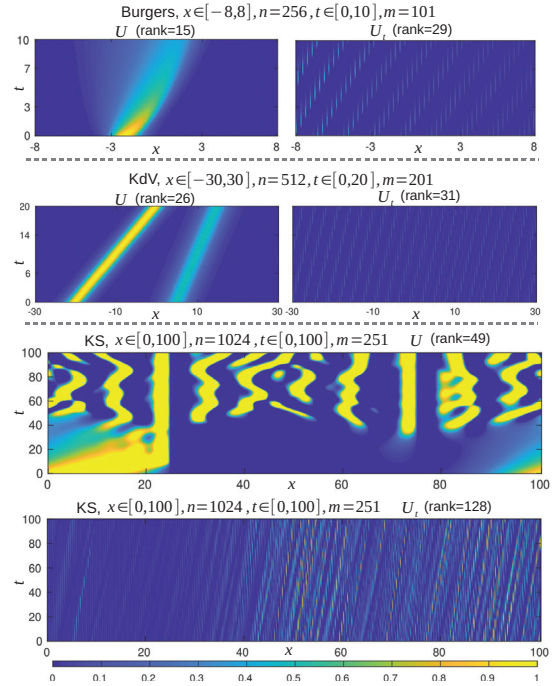
Figure 1: Low-rank numerical solutions $U \in \mathbb{C}^{n \times m}$ and their partial derivatives $U_t \in \mathbb{C}^{n \times m}$ of PDEs (e.g., Burgers, Korteweg-de Vries (KdV) and Kuramoto-Sivashinsky (KS)). U and U_t are collected from all the spatial time series data $u(x, t)$ and $u_t(x, t)$ in the intervals $[a, b]$ and $[c, d]$ with n spatial locations and m time points.

complex data with sparse outlying entries becomes a central challenge in many diverse areas of the natural sciences.

Our work is based on the observation that the underlying data modeled by the PDE function $u(x, t)$ are often low rank since there exist similar patterns and/or linear combinations in its underlying PDEs. For example, in the filtering Gaussian noise process via singular value decomposition, low-rank solutions of Navier Stokes and reaction diffusion have been confirmed in the literature (Rudy et al. 2017). Moreover, Fig. 1 also shows that the ranks of the numeri-

cal solutions U of Burgers, Korteweg-de Vries (KdV) and Kuramoto-Sivashinsky (KS) equations are 15, 26 and 49, respectively, with ranks computed using MatLab functions, $rank(U, 0.01)$ or $rank(U_t, 0.01)$. In addition, we observe that a data matrix U_t discretely sampled from the partial differentiation $u_t(x, t)$ is also low rank, that is, the ranks of the corresponding U_t are respectively 29, 32 and 128 in Fig. 1.

By considering both the complex data U and U_t with the sparse outlying entries, they can be regarded as the superposition of the low-rank component and the sparse component (Candés et al. 2011). Therefore, this leads us to recover both the complex data U and the partial differentiations U_t by using double low-rank and sparse decompositions for robustly discovering the underlying PDEs. In this paper, we introduce a robust approach to discover the governing PDEs by separating the low-rank data and the sparse outlying entries. Overall, our contributions are three folds.

- To address the PDEs discovery challenge with the sparse outlying entries, we develop a robust PDEs discovery framework by integrating double low-rank and sparse recoveries with the (group) sparse regression method. This framework is implemented as a minimization problem with mixed nuclear norms using ℓ_1 and ℓ_0 norms.

- To solve the optimization problem, we introduce a double low-rank sparse regression (DLrSR) algorithm that includes three parts: (i) derive a clean low-rank data matrix by separating the outliers from the noisy data U, (ii) build a library using Z from the clean data matrix, and (iii) use a low-rank sequential (grouped) threshold ridge regression to robustly find the PDEs.

- Several experiments are demonstrated on seven canonical models (i.e., four PDEs and three parametric PDEs) with the introduction of the sparse outliers. The results verify that our framework outperforms the state-of-the-art group sparse and sparse regression methods.

Related Work

We mainly review sparse representation, low-rank representation and sparse identification of nonlinear dynamics (SINDy) in this section. We also discuss our robust low-rank PDEs discovery with SINDy and its variants.

Sparse representation. Suppose an observation $x \in \mathbb{C}^n$ can be represented by a linear combination of a given overcomplete dictionary $\mathcal{D} \in \mathbb{C}^{n \times d}(n > d)$, sparse representation problem is to seek the possible coefficients $\alpha \in \mathbb{C}^d$ with the fewest non-zeros satisfying an equation $x = \mathcal{D}\alpha$ (Olshausen and Field 1997). To solve this problem, many different norm minimizations are usually considered as sparsity constraints, e.g., ℓ_0, $\ell_p (0 < p < 1)$, ℓ_1 and $\ell_{2,1}$-norms (Donoho and Elad 2003; Hoyer 2004; Bruckstein, Donoho, and Elad 2009; Kim and Xing 2012). Moreover, it also has been successfully applied into many areas, especially in signal processing, image processing, machine learning, and computer vision, such as compressive sensing, image denoising/debluring/inpainting/super-resolution/classification/clustering, and visual tracking (Beck and Teboulle 2009; Yang et al. 2012;

Li, Kong, and Fu 2017; Sun, Cong, and Xu 2018; Sun et al. 2018; Luo et al. 2018).

Low-rank representation (LRR). Suppose a given data observation matrix $U \in \mathbb{C}^{n \times m}$ can be decomposed as a low-rank matrix $Z \in \mathbb{C}^{n \times m}$ and sparse (or Gaussian) noise matrix $E \in \mathbb{C}^{n \times m}$, $U = Z + E$, low-rank problem is to recover the low-rank and sparse (or Gaussian) components both accurately (perhaps even exactly) and efficiently. Robust principal component analysis (RPCA) (Candés et al. 2011) can recover both the low-rank and the sparse components exactly under some suitable assumptions. RPCA is extended into multiple subspaces, called low-rank representation (LRR) (Liu et al. 2013). RPCA and LRR have been used in many important applications, such as, video surveillance, face recognition, subspace clustering, ranking and collaborative filtering (Candés et al. 2011; Liu et al. 2013; Li et al. 2016; 2017; He, Tan, and Wang 2014).

Sparse identification of nonlinear dynamics (SINDy). Sparse representation with underlying dynamical systems is used to discover governing nonlinear equations from data observations (Bruntona, Proctor, and Kutz 2016). In SINDy, the spatial time series data $u(x, t)$ with location $x \in [a, b]$ and time $t \in [c, d]$ are collected into a matrix $U \in \mathbb{C}^{n \times m}$, where the intervals $[a, b]$ and $[c, d]$ are divided into n spatial locations and m time points. Next, the collected data is used to construct an overcomplete library $\mathcal{D} \in \mathbb{C}^{n \times m \times d}$, which includes d linear, nonlinear and partial derivatives terms. Finally, the ℓ_0 or ℓ_1 sparse coding methods are employed to find the nonlinear and partial derivative terms of the governing PDEs that most accurately represent partial derivative data U_t of the collected data U. In the final step, they employ the ℓ_0 or ℓ_1 sparse coding methods to find the nonlinear and partial derivative terms of the governing PDEs that most accurately represent partial derivative data U_t of the collected data U. In this method, many SINDy variants have been proposed, for example, exhibiting multiscale physical phenomenon by discovering nonlinear multiscale systems (Champion, Brunton, and Kutz 2019), and characterizing hybrid (switching) behaviors by using Hybrid-SINDy (Mangan et al. 2019). In addition, SINDy has been widely applied to discover nonlinear equations for biological network systems (Mangan et al. 2016), fluid flows (Loiseau and Brunton 2018), model predictive control (Kaiser, Kutz, and Brunton 2018), convection in a plasma (Dam et al. 2017) and chemical reaction dynamics (Hoffmann, Frohner, and Noé 2019).

More importantly, some variants of SINDy, such as sequential threshold ridge regression (STR) (Rudy et al. 2017; Schaeffer 2017), have been successfully applied to understand the underlying physical laws by identifying nonlinear PDEs from time series measurements with Gaussian noise. Moreover, group sparse coding is extended to discover parametric PDEs as the time-series measurements usually obey unknown PDEs with time-evolving parameters, called sequential grouped threshold ridge regression (SGTR) (Rudy et al. 2019). In addition, deep learning has been applied into the data-driven discovery of PDEs (Sirignano and Spiliopoulos 2018; Xu et al. 2019).

Discussion. In the process of collecting and analyzing the data, however, both the time series data U and the partial

derivative data U_t are often corrupted by noise (e.g., Gaussian noise and/or outliers). The noise, especially those from the outliers, make PDEs discovery a challenge. Although STR and SGTR are robust to the Gaussian noise (Rudy et al. 2017; 2019; Schaeffer 2017), the outliers can lead to failed or limited discovery of the governing PDEs. Here, we develop a robust low-rank PDEs discovery framework to effectively handle the outliers with high magnitude Gaussian noise. Moreover, our robust low-rank discovery framework can also be extended into SINDy and its variants.

Robust Low-rank PDEs Discovery

In this section, we develop a robust low-rank PDEs discovery approach by first formalizing it as an optimization problem, and proposing a double low-rank sparse regression (DLrSR) algorithm to solve this optimization problem.

Problem Formalization

We consider a general partial differential equation in the following form (Rudy et al. 2017; 2019)

$$u_t = \mathcal{N}(u, u_x, u_{xx}, \cdots, x, \overline{\mu}) \\ = \sum_{i=1}^{d} \mathcal{N}_i(u, u_x, u_{xx}, \cdots)\overline{\xi}_i, \quad (1)$$

where $\mathcal{N}(\cdot)$ is a nonlinear function of $u(x,t)$ and its partial differentiations, $\overline{\mu}$ is a constant μ or time-evolving parameters $\mu(t) : [0, T] \to \mathbb{R}$, and $\overline{\xi}$ is a constant coefficient ξ or time-evolving coefficients $\xi(t)$. Actually, $\mathcal{N}(\cdot)$ has polynomial nonlinearities, which are common in many of the canonical models of the natural sciences. For example, we consider $d = 7$ candidate terms $[\mathcal{N}_1, \cdots, \mathcal{N}_7] = [u, u_x, u_{xx}, uu_x, u^2u_x, uu_{xx}, u^2u_{xx}]$, Burgers' equation is $\mathcal{N} = \sum_{i=1}^{7} \mathcal{N}_i\xi_i(t) = -uu_x + \mu(t)u_{xx}$, where $[\xi_1(t), \cdots, \xi_7(t)] = [0, 0, \mu(t), -1, 0, 0, 0]$. (Parametric) PDE-FIND (Rudy et al. 2017; 2019) is to find the parameters (i.e., $\mu(t)$ and -1) from data measurements by using (group) sparse coding.

Mathematically, a data matrix $U \in \mathbb{C}^{n \times m}$ with location $x \in [a, b]$ and time $t \in [0, T]$ is discretely and corruptly collected from the natural function $u(x, t)$ that we assume satisfies the PDE in Eq. (1). Based on our low-rank observations of U and its differentiation U_t in Figure 1, we suppose that both U and U_t may be decomposed as

$$U = Z + E_1, \quad U_t = \mathcal{D}(Z, Q)\overline{\xi} + E_2, \quad (2)$$

where Z and $\mathcal{D}(Z, Q)\overline{\xi}$ have low-rank, E_1 and E_2 are sparse; here, all components are of arbitrary magnitude. $\mathcal{D}(Z, Q) \in \mathbb{C}^{n \times m \times d}$ is a large library of candidate terms that may appear in $\mathcal{N}$, and Q is denoted as a matrix containing additional information that may be relevant, such as dependencies on the absolute value of U. Similar to PDE-FIND, $\mathcal{D}(Z, Q) = [1, U, U^2, \cdots, Q, \cdots, U_x, UU_x, \cdots, Q^2U^3U_{xxx}]$.

In this paper, we seek to robustly discover the parameters from the data measurements with sparse noise. It is addressed by solving the optimization problem as follows:

$$\min_{\overline{\xi}, Z, E_1, E_2} \|Z\|_* + \|\mathcal{D}(Z, Q)\overline{\xi}\|_* + \lambda_1\mathcal{R}(\overline{\xi}) \\ + \lambda_2\|E_1\|_1 + \lambda_3\|E_2\|_1, \quad (3)$$

$$\text{s.t.} \quad U = Z + E_1, \quad U_t = \mathcal{D}(Z, Q)\overline{\xi} + E_2,$$

where $\mathcal{R}(\overline{\xi})$ is a sparse regularization of the parameters $\overline{\xi}$, such as, ℓ_0 norm $\|\overline{\xi}\|_0$, ℓ_1 norm $\|\overline{\xi}\|_1$, and group sparse norm $\sum_{j=1}^{m} \|\xi(j)\|_2^2$.

Optimization

Since both the nuclear norm $\|\mathcal{D}(Z, Q)\overline{\xi}\|_*$ and the sparse regularization term $\mathcal{R}(\overline{\xi})$ include the variable $\overline{\xi}$, we introduce an auxiliary variable X to separate them and the problem (3) is written as

$$\min_{\xi(t), Z, X, E_1, E_2} \|Z\|_* + \|X\|_* + \lambda_1\mathcal{R}(\overline{\xi}) + \lambda_2\|E_1\|_1$$

$$+ \lambda_3\|E_2\|_1 + \frac{\lambda_4}{2}\|X - \mathcal{D}(Z, Q)\overline{\xi}\|_F^2, \quad (4)$$

$$\text{s.t.} \quad U = Z + E_1, \quad U_t = X + E_2,$$

Clearly, the problem (4) is nonconvex due to the dependence between $\mathcal{D}(Z, Q)$ and Z. To solve this challenging problem, we consider the following augmented Lagrangian function:

$$\mathcal{L} = \|Z\|_* + \|X\|_* + \lambda_1\mathcal{R}(\xi(t)) + \lambda_2\|E_1\|_1 + \lambda_3\|E_2\|_1 +$$

$$\frac{\lambda_4}{2}\|X - \mathcal{D}(Z, Q)\overline{\xi}\|_F^2 + \frac{\eta_1}{2}\|U - Z - E_1 + Y_1/\eta_1\|_F^2$$

$$+ \frac{\eta_2}{2}\|U_t - X - E_2 + Y_2/\eta_2\|_F^2), \quad (5)$$

where Y_1 and Y_2 are the Lagrange multipliers with penalty parameters η_1 and η_2. We develop a double low-rank sparse regression (DLrSR) framework shown in Figure 2 to solve the problem in Eq. (5), which includes the following three steps. Step 1 is to separate the clean data matrix Z and the sparse noise matrix E_1 by using robust principal component analysis (Candés et al. 2011). In step 2, Z is used to build the the library $\mathcal{D}(Z, Q)$. In step 3, we use the proposed low-rank sequential (grouped) threshold ridge regression (LrSTR) to robustly find the PDE equations.

Step 1 is to ignore the relationship between $\mathcal{D}(Z, Q)$ and Z, and consider the following subproblem:

$$\mathcal{L}_1 = \|Z\|_* + \lambda_2\|E_1\|_1 \\ + \frac{\eta_1}{2}\|U - Z - E_1 + Y_1/\eta_1\|_F^2, \quad (6)$$

which is often called robust Principal Component Analysis (rPCA). To handle the $\mathcal{L}_1$ subproblem (6), we alternately update the variables $\{Z, E_1, Y_1\}$ as follows:

Update the low-rank variable Z by

$$Z = \underset{Z}{\arg\min} \|Z\|_* + \frac{\eta_1}{2}\|U - Z - E_1 + Y_1/\eta_1\|_F^2, \\ = \mathbf{J}_{\frac{1}{\eta_1}}(U - E_1 + Y_1/\eta_1), \quad (7)$$

where $\mathbf{J}_{\frac{1}{\eta_1}}(A) = U_A\mathbf{J}_{\frac{1}{\eta_1}}(\Sigma)V_A$, $\mathbf{J}_{\frac{1}{\eta_1}}(\Sigma) = \text{diag}(\{\sigma_i - \frac{1}{\eta_1}\}_+)$, the singular value decomposition (SVD) of the matrix A of rank r is $A = U_A\Sigma V_A$, $\Sigma = \text{diag}(\{\sigma_i\}_{1 \le i \le r})$ and $a_+ = \max(0, a)$ (Cai, Candés, and Shen 2010).

Update the sparse outliers variable E_1 by

$$E = \underset{E}{\arg\min} \lambda_2\|E_1\|_1 + \frac{\eta_1}{2}\|E_1 - (U - Z - Y_1/\eta_1)\|_F^2, \\ = \mathbf{S}_{\frac{\lambda_2}{\eta_1}}(U - Z - Y_1/\eta_1), \quad (8)$$
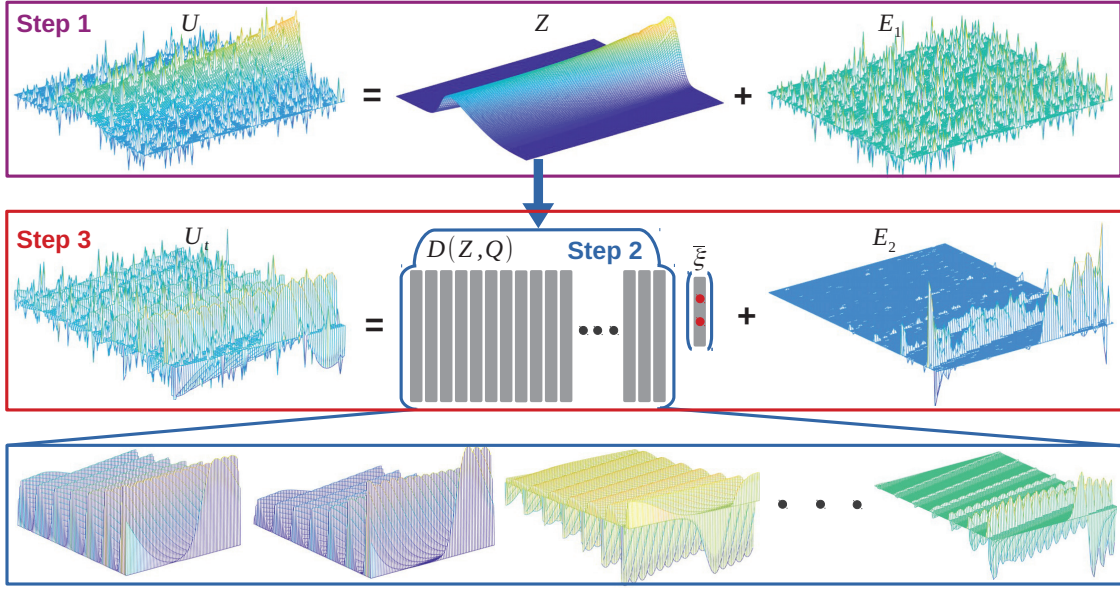
Figure 2: **Steps in the robust PDEs discovery of double low-rank sparse regression (DLrSR)** algorithm on Burgers' equation. **Step 1** separates the outliers E_1 from the collected data U; **Step 2** builds a library $\mathcal{D}(Z, Q)$ using the clean data Z; **Step 3** finds the sparse coefficients $\overline{\xi}$ by separating the outliers E_2 from the partial derivative U_t and mixing the nuclear with ℓ_0 norms.

Algorithm 1 rPCA

1: **Input:** U and λ_2
2: **Initialize:** $Y_1 = U/J(U)$, $E_1 = 0$, $Z = 0$, $\eta_1 > 0$, $\rho > 0$, and $\eta_{\max}$.
3: **while** not converged **do**
4:    update $\{Z, E_1, Y_1\}$ by (7), (8) and (10), respectively,
5:    $\eta_1 = \min\{\rho\eta_1, \eta_{\max}\}$,
6: **end while**
7: **return** Z and Y_1.

Algorithm 2 Low-rank S(G)TRidge (LrSTR).

1: **Input:** U_t, $\mathcal{D}(Z, Q)$, λ_1, λ_3 and λ_4
2: **Initialize:** $Y_2 = U_t/J(U_t)$, $E_1 = 0$, $Z = 0$, $\eta_2 > 0$, $\overline{\rho} > 0$, and $\overline{\eta}_{\max}$.
3: **while** not converged **do**
4:    update $\{X, \overline{\xi}, E_2, Y_2\}$ by (12)-(15), respectively,
5:    $\eta_2 = \min\{\overline{\rho}\eta_2, \overline{\eta}_{\max}\}$,
6: **end while**
7: **return** X, $\xi(t)$, and E_2.

---

where $\mathbf{S}_{\frac{\lambda_2}{\eta_1}}(\cdot)$ is the shrinkage-thresholding operator acting on each element of the given matrix, and is defined as

$$\mathbf{S}_{\frac{\lambda_2}{\eta_1}}(a) = \begin{cases} a - \frac{\lambda_2}{\eta_1}, & \text{if } a > \frac{\lambda_2}{\eta_1} \\ a + \frac{\lambda_2}{\eta_1}, & \text{if } a < \frac{\lambda_2}{\eta_1} \\ 0, & \text{otherwise} \end{cases}. \quad (9)$$

Update the Lagrange multiplier Y_1 by

$$Y_1 = Y_1 + \eta_1(U - Z - E_1). \quad (10)$$

The procedure of rPCA is outlined in **Algorithm 1**.

Step 2 is to build the library $\mathcal{D}(Z, Q)$. Based on the clean data Z solved in (6), the second part is to create the library $\mathcal{D}(Z, Q)$ by using the PDE-FIND algorithm (Rudy et al. 2017).

Step 3 is to propose a low-rank sequential (grouped) threshold ridge regression (LrSTR) algorithm. To robustly find the PDE equations, the third part is to consider the fol-

lowing subproblem:

$$\mathcal{L}_2 = \|X\|_* + \lambda_1 \mathcal{R}(\overline{\xi}) + \frac{\lambda_4}{2}\|X - \mathcal{D}(Z, Q)\overline{\xi}\|_F^2 + \\ \lambda_3\|E_2\|_1 + \frac{\eta_2}{2}\|U_t - X - E_2 + Y_2/\eta\|_F^2. \quad (11)$$

To solve the $\mathcal{L}_2$ subproblem (11), we propose a LrSTR algorithm, which alternately updates the variables $\{X, \overline{\xi}, E_2, Y_2\}$ as follows:

Update the low-rank variable X by

$$X = \underset{X}{\operatorname{argmin}} \|X\|_* + \frac{\eta_2 + \lambda_4}{2}\|Z - (\frac{\eta_2}{\eta_2 + \lambda_4}(U_t \\ - E_2 + Y_2/\eta_2) + \frac{\lambda_4}{\eta_2 + \lambda_4}\mathcal{D}(Z, Q)\overline{\xi}\|_F^2, \\ = \mathbf{J}_{\frac{1}{\eta_2 + \lambda_4}}\left(\frac{\eta_2(U - E_2 + Y_2/\eta_2) + \lambda_4\mathcal{D}(Z, Q)\overline{\xi}}{\eta_2 + \lambda_4}\right), \quad (12)$$

where $\mathbf{J}_{\frac{1}{\eta_2 + \lambda_4}}(\cdot)$ is defined in the Eq. (7).

Table 1: Summary of robust discovery results for PDEs on four canonical modes. † denotes a failed discovery.

modes	methods	no noise	U+noise	$\{U, U_t\}$+noise	form and discretization
Burgers	STR	0.159%±0.061%	$u_t = -0.744uu_x^\dagger$	$u_t = -0.740uu_x^\dagger$	$u_t = -uu_x + \epsilon u_{xx}, \epsilon = 0.1,$
	rSTR	0.124%±0.065%	3.20%±2.198%	$u_t = -0.864uu_x^\dagger$	$x \in [-8, 8], n = 256, t \in [0, 10], m = 101$
	DLrSR	**0.076%±0.013%**	**1.271%±0.960%**	**3.084%±1.285%**	
KdV	STR	0.957%±0.232%	①†	$u_t = -2.932uu_x^\dagger$	$u_t = -6uu_x - u_{xxx}$
	rSTR	0.957%±0.232%	9.171%±5.595%	9.840%±5.883%	$x \in [-30, 30], n = 512, t \in [0, 20], m = 201$
	DLrSR	0.957%±0.232%	**9.161%±5.583%**	**9.203%±5.819%**	
NLS	STR	0.047%±0.015%	②†	③†	$iu_t = -\frac{1}{2}u_{xx} - |u|^2u$
	rSTR	0.047%±0.015%	0.347%±0.029%	④†	$x \in [-5, 5], n = 512, t \in [0, \pi], m = 501$
	DLrSR	0.047%±0.015%	0.347%±0.029%	**0.356%±0.030%**	
KS	STR	1.275%±1.314%	36.144%±2.939%	⑤†	$u_t = -uu_x - u_{xx} - u_{xxxx}$
	rSTR	1.221%±1.296%	19.498%±4.650%	⑤†	$x \in [0, 100], n = 1024, t \in [0, 100], m = 251$
	DLrSR	**1.218%±1.295%**	19.498%±4.650%	**39.012%±2.715%**	

① $u_t = -0.056u - 0.341u_x - 0.070u_{xxx}$; ③ $iu_t = (-0.003 + 0.000i)u^2|u|u_{xx}$;
④ $iu_t = (0.032 + 0.032i)u|u|u_{xx} + (0.006 - 0.000i)u^2|u|u_{xx}$; ⑤ $u_t = -0.174uu_x - 0.085u_{xxxx}$;
② $iu_t = -(0.487 + 0.490i)uu_{xx} + (0.345 + 0.349i)u|u|u_{xx} - (0.052 + 0.053i)u|u|^2u_{xx}$.

Algorithm 3 Double Low-rank Sparse Regression (DLrSR)

1: **Input:** $U, \lambda_1, \lambda_2, \lambda_3$ and λ_4.
2: obtain $\{Z, E_1\}$ by rPCA(U, λ_2) in **Algorithm 1**,
3: build $\mathcal{D}(Z, Q)$ by PDE-FIND (Rudy et al. 2017),
4: obtain $\{X, \overline{\xi}, E_2\}$ by LrSTR$(U_t, \mathcal{D}(Z, Q)), \lambda_1, \lambda_3, \lambda_4$) in **Algorithm 2**,
5: **return** $\{Z, X, \overline{\xi}, E_1, E_2\}$.

Update the group sparse variable $\overline{\xi}$ by

$$\overline{\xi} = \underset{\overline{\xi}}{\operatorname{argmin}} \; \lambda_1 \mathcal{R}(\overline{\xi}) + \frac{\lambda_4}{2}\|Z - \mathcal{D}(Z, Q)\overline{\xi}\|_{\mathrm{F}}^2, \quad (13)$$

where $\overline{\xi}$ is solved by sequential threshold ridge regression (STR) (Rudy et al. 2017) or sequential grouped threshold ridge regression (SGTR) (Rudy et al. 2019).

Update the sparse outliers variable E_2 by

$$
\begin{aligned}
E_2 &= \underset{E_2}{\operatorname{argmin}} \; \lambda_3\|E_2\|_1 + \frac{\eta_2}{2}\|E - (U - X - Y_2/\eta_2)\|_{\mathrm{F}}^2, \\
&= \mathbf{S}_{\frac{\lambda_3}{\eta_2}}(U - X - Y_2/\eta_2), \quad (14)
\end{aligned}
$$

where $\mathbf{S}_{\frac{\lambda_3}{\eta_2}}(\cdot)$ is defined in the Eq. (9).

Update the Lagrange multiplier Y_2 by

$$Y_2 = Y_2 + \eta_2(U_t - X - E_2). \quad (15)$$

The procedure of LrSTR is outlined in **Algorithm 2**. By collecting the three parts, the completed optimization procedure of the double low-rank sparse regression (DLrSR) is summarized in **Algorithm 3**. Overall, the convergence of DLrSR is key to LrSTR as the convergence of rPCA has been proved in the literature (Candés et al. 2011). Our experiments verify that LrSTR is convergent in Figure 3.

Experiments

In this section, we conduct several experiments to verify the robust PDEs and parametric PDE discoveries by using our proposed DLrSR method. Due to the limited space, the hyper-parameter settings (i.e., λ_1-λ_4) are provided in the supplementary material.
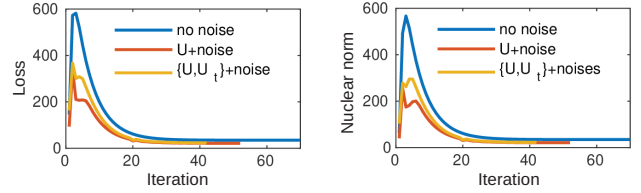

Figure 3: Convergence of LrSTR on Burgers' equation with no noise, U+sparse noise, and $\{U, U_t\}$+sparse noise. Left is the loss, and right is the nuclear norm of $\mathcal{D}(Z, Q)\overline{\xi}$.

Robust Discovery of PDEs

In this subsection, we evaluate our DLrSR method to robustly find the PDEs on four canonical models; Burgers', Korteweg-de Vries (KdV), Nonlinear Schrödinger (NLS) and Kuramoto-Sivashinsky (KS) equations. For fair comparison, we follow the settings in (Rudy et al. 2017) and numerically solve all parametric PDEs by employing the discrete Fourier transform (DFT) to evaluate spatial derivatives and using the SciPy function *odeint* (Jones et al. 2001) for temporal integration with n and m. To replicate the effects of sparse sensor noise, outliers with 90% sparsity (meaning that 90% of its cells are zeros) are added directly to both U and U_t (Schaeffer 2017). In particular, the magnitudes of the outliers E_1 of U are equal to 100% (Burgers' and KdV) and 2% (NLS and KS) of their standard deviations, and for the outliers E_2 of U_t, the magnitudes are 200% (Burgers', KdV and KS) and 100% (NLS), respectively. We also follow the error metric (Rudy et al. 2017) between the discovered sparse coefficients $\overline{\xi}$ and the ground truth $\widehat{\xi}$ to fairly evaluate the found results, which is defined as: for $\overline{\xi} = \xi$,

$$\text{error} = \text{mean}(\alpha) \pm \text{std}(\alpha), \quad (16)$$

where $\alpha = \left\{ \frac{|\xi_j - \widehat{\xi}_j|}{|\widehat{\xi}_j|} \times 100\% \right\}$ and j is the effective coefficient in ξ. Note that a lower value shows a better discovery. Moreover, we compare DLrSR with two methods. One is the
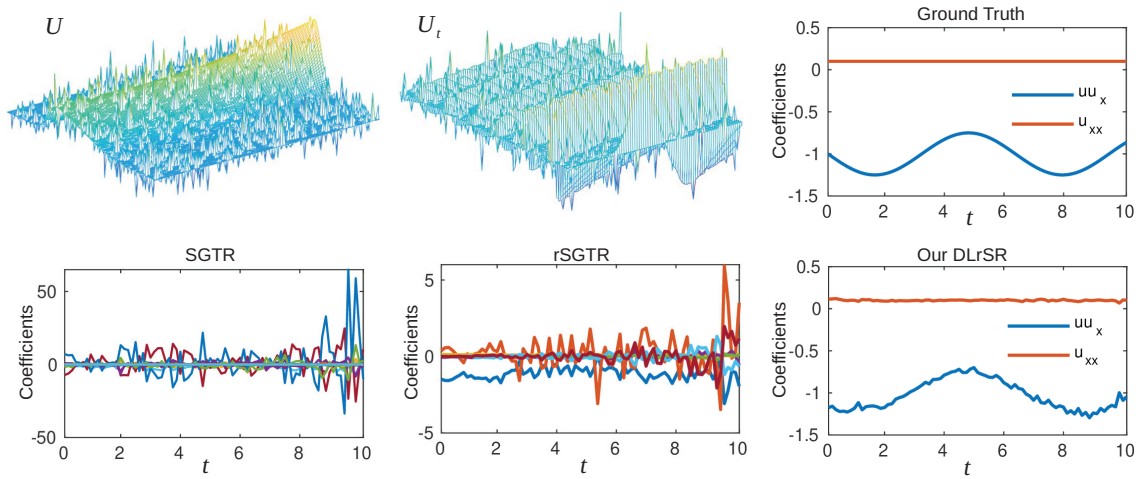
Figure 4: Robust coefficients discovery of the Burgers' equation with a nonlinear term $a(t) = -(1 + \sin(t)/4)$. Both U and U_t are corrupted by the outliers with 90% sparsity, and their magnitudes are 100% and 150% of their standard deviations. Compared to the ground truth, our DLrSR is much better than SGTR and rSGTR, which fail to discover the equation.

Table 2: Summary of robust discovery results for parametric PDEs on three canonical modes. † denotes a failed discovery.

modes	methods	no noise	U+noise	$\{U, U_t\}$+noise	form and discretization
Burgers	SGTR	1.0274531%	6592.7198%†	9088.3600%†	$u_t = a(t)uu_x + \epsilon u_{xx}, \epsilon = 0.1,$
	rSGTR	0.6218406%	6.3265683%	832.45845%†	$x \in [-8, 8], n = 256, t \in [0, 10], m = 101$
	DLrSR	**0.6082473**%	6.3265683%	**13.335198**%	$a(t) = -(1 + \sin(t)/4)$
advection diffusion	SGTR	0.1726398%	2091.9652%†	2441.1882%†	$u_t = c(x)u_x + c'(x)u_x + \epsilon u_{xx}$
	rSGTR	0.1726290%	0.76343382%	45.833947%	$x \in [-5, 5], n = 256, t \in [0, 5], m = 256$
	DLrSR	0.1726290%	0.76343382%	**5.2214745**%	$c(x) = -1.5 + \cos(2\pi x/5)$
KS	SGTR	1.3595882%	56.255372%	58.559030%	$u_t = a(x)uu_x + b(x)u_{xx} + c(x)u_{xxxx}$
	rSGTR	1.3595882%	13.972810%	16.428652%	$x \in [-20, 20], n = 512, t \in [0, 200], m = 1024$
	DLrSR	1.3595882%	13.972810%	**13.961776**%	$a(x) = 1 + \sin(2\pi x/20)/4, b(x) = -1 + e^{-(x-2)^2/5}/4$
					$c(x) = -1 - e^{-(x-2)^2/5}/4$

state-of-the-art STR (Rudy et al. 2017); another is to combine rPCA with STR, rPCA+STR (rSTR).

Results. The discovery results of the PDEs are summarized in Table 1. Depending on the different sparse noise, we have the following three observations. First, DLrSR is better than STR and rSTR on Burgers' and KS equations without sparse noise although all the methods have same results on KdV and NLS. For example, compared to the ground truth $u_t = -uu_x + 0.1u_{xx}$ of the Burgers' equation, $u_t = -0.999367uu_x + 0.100089u_{xx}$ discovered by DLrSR is better estimate than $u_t = -1.000987uu_x + 0.100220u_{xx}$ by STR. Second, since U is corrupted by the sparse noise, STR is worse than DLrSR and rSTR, which often fail to identify the PDEs equations except KS. Third, to handle the challenging case that U and U_t are corrupted by the sparse noise, both STR and rSTR cannot discover the correct terms in all the PDEs equations expect KdV using rSTR. Our DLrSR not only identifies the correct terms, but also shows good discovery results, for example, $u_t = -0.956320uu_x + 0.101799u_{xx}$ (Burgers'), $u_t = -6.23746uu_x - 1.149225u_{xxx}$ (KdV),

and $u_t = (-0.000268 + 0.498145i)u_{xx} + (-0.000132 + 0.996659i)u_{xxx}$ (NLS). In fact, the KS equation is particularly challenging to identify with low error on the coefficients. Although its correct terms are discovered, there is a high coefficient error. The discovered model of KS is given by $u_t = -0.571916uu_x - 0.623884u_{xx} - 0.633841u_{xxxx}$.

Robust Discovery of Parametric PDE

In this subsection, we evaluate our DLrSR method to robustly find the parametric PDEs on three canonical models; Burgers' equation with a time varying nonlinear term, spatially dependent advection-diffusion (AD) and KS equations. Similar to (Rudy et al. 2017; 2019), we numerically solve all parametric PDEs by employing the discrete Fourier transform (DFT) to evaluate spatial derivatives and using the SciPy function *odeint* (Jones et al. 2001) for temporal integration with n and m. To replicate the effects of sparse sensor noise, outliers with 90% sparsity (meaning that 90% of its cells are zeros) are added directly to both U and U_t (Schaeffer 2017; Rudy et al. 2019). In particular, the magnitudes of the outliers of U and U_t are equal to 100% and 150% of
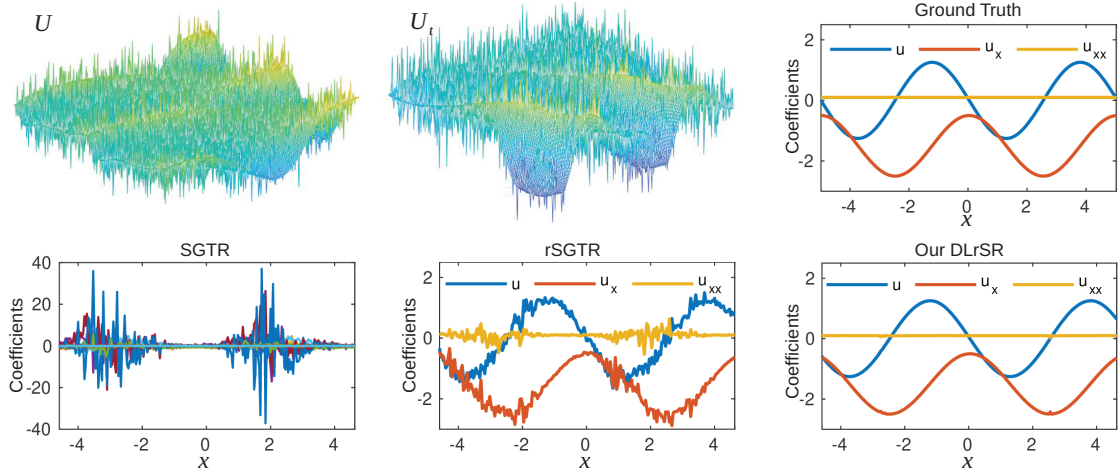
Figure 5: Robust coefficients discovery of the spatially dependent advection-diffusion equation. Both U and U_t are corrupted by the outliers with 90% sparsity, and their magnitudes are 100% and 150% of their standard deviations. Compared to the ground truth, our DLrSR is much better than rSGTR, and SGTR fails to discover the equation.
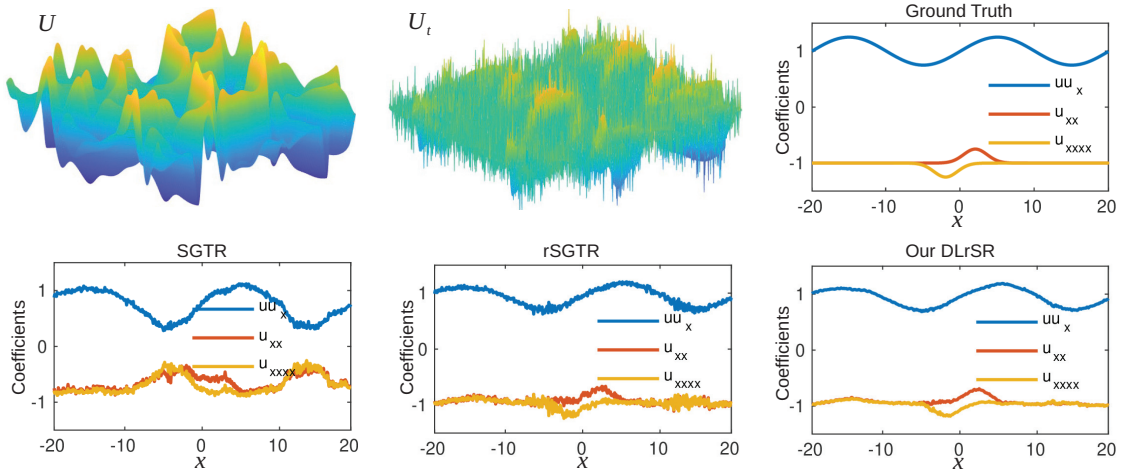

Figure 6: Robust coefficients discovery of the spatially dependent Kuramoto-Sivashinsky (KS) equation. Both U and U_t are corrupted by the outliers with 90% sparsity, and their magnitudes are 0.05% and 150% of their standard deviations. Compared to the ground truth, our DLrSR is better than SGTR and rSGTR.

their standard deviations, respectively.

Before showing the results, we first consider an error metric between the discovered sparse coefficients $\bar{\xi}$ and the ground truth $\hat{\xi}$ to evaluate the found results of the parametric PDEs discovery, which is defined as: for $\bar{\xi} = \xi(t)$,

$$\text{error} = \frac{\|\xi(t) - \hat{\xi}(t)\|_1}{\|\hat{\xi}(t)\|_1} \times 100\%, \qquad (17)$$

Note that a lower value shows a better equation discovery. We compare our DLrSR with two methods. One is the state-of-the-art SGTR (Rudy et al. 2019); another is to combine rPCA with SGTR, rPCA+SGTR (rSGTR).

Results. The discovery results of the parametric PDEs are summarized in Table 2, and the parametric identification of the Burgers', AD and KS modes are shown in Figures 4, 5 and 6, respectively. Depending on the different sparse noise, we have the following three observations. First, SGTR and rSGTR are worse than our DLrSR on the Burgers' equation without sparse noise although they have the same results on AD and KS. Second, since U is corrupted by the sparse noise, DLrSR and rSGTR are better than SGTR, which often fails to identify the modes except KS. Third, our DLrSR can effectively address the challenging case that U and U_t are corrupted by the sparse noise, and achieve lower error than both SGTR and rSGTR. Moreover, Figure 4 shows that SGTR and rSGTR cannot discover the correct terms in the Burgers' modes, and DLrSR identifies the correct terms. Figures 5 and 6 also show that DLrSR outperforms SGTR and rSGTR on both AD and KS modes.

Conclusion

In this paper, we have presented a robust method for identifying the governing laws for physical systems which are often corrupted by noise, especially outliers. To the best of our knowledge, our method is the first approach for deriving the challenging PDEs discovery with outliers. Specifically, we can separate the clean low-rank data and the outliers. The DLrSR algorithm exhibits equal or superior performance to the state-of-the-art STR and SGTR in determining correct active terms on all the examples with or without outliers. Even in the cases when both STR and SGTR fail to identify the correct terms, our DLrSR can still effectively find them.

Acknowledgments

References

Beck, A., and Teboulle, M. 2009. A fast iterative shrinkage thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences* 2(1):183–202.

Bruckstein, A.; Donoho, D.; and Elad, M. 2009. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review* 51:34–81.

Bruntona, S.; Proctor, J.; and Kutz, J. 2016. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* 113(15):3932–3937.

Butler, K.; Davies, D.; Cartwright, H.; Isayev, O.; and walsh, A. 2018. Machine learning for molecular and materials science. *Nature* 559:547–555.

Cai, J.; Candés, E.; and Shen, Z. 2010. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* 20(4):1956–1982.

Candés, E.; Li, X.; Ma, Y.; and Wright, J. 2011. Robust principal component analysis? *Journal of the ACM* 58(3):1–37.

Champion, K.; Brunton, S.; and Kutz, J. 2019. Discovery of nonlinear multiscale systems: Sampling strategies and embeddings. *SIAM J. Applied Dynamical Systems* 18(1):312–333.

Dam, M.; Br∅ns, M.; Rasmussen, J.; Naulin, V.; and Hesthaven, J. 2017. Sparse identification of a predator-prey system from simulation data of a convection model. *Physics of Plasmas* 24(2):022310.

Donoho, D., and Elad, M. 2003. Optimally sparse representation in general (non-orthogonal) dictionaries via l1 minimization. *Proc. Natl. Acad. Sci.* 100:2197–2202.

Elhamifar, E., and Vidal, R. 2013. Sparse subspace clustering: algorithm, theory, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* 35(11):2765–2781.

He, R.; Tan, T.; and Wang, L. 2014. Robust recovery of corrupted low-rank matrix by implicit regularizers. *IEEE Trans. Pattern Anal. Mach. Intell.* 36(4):770–783.

Hoffmann, M.; Frohner, C.; and Noé, F. 2019. Reactive sindy: Discovering governing reactions from concentration data. *Journal of Chemical Physics* 150(2):025101.

Hoyer, P. 2004. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research* 5:1457–1469.

Jones, E.; Oliphant, T.; Peterson, P.; et al. 2001–. SciPy: Open source scientific tools for Python.

Jordan, M., and Mitchell, T. 2015. Machine learning: Trends, perspectives, and prospects. *Science* 349(6245):255–260.

Kaiser, E.; Kutz, J. N.; and Brunton, S. 2018. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proc. R. Soc. A* 474(2204):20180335.

Kim, S., and Xing, E. 2012. Tree-guided group lasso for multi-response regression with structured sparsity, with an application to eqtl mapping. *Annals of Applied Statistics* 6:1095–1117.

Li, J.; Kong, Y.; Zhao, H.; Yang, J.; and Fu, Y. 2016. Learning fast low-rank projection for image classification. *IEEE Trans. Image Process.* 25(10):4803–4814.

Li, J.; Liu, H.; Zhao, H.; and Fu, Y. 2017. Projective low-rank subspace clustering via learning deep encoder. In *Proc. of of Int. Joint Conf. on Artif. Intell.*, 2145–2151.

Li, J.; Kong, Y.; and Fu, Y. 2017. Sparse subspace clustering by learning approximation ℓ_0 codes. In *Proc. of the AAAI Conf. on Artif. Intell.*, 2189–2195.

Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2013. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 35(1):171–184.

Loiseau, J., and Brunton, S. 2018. Constrained sparse galerkin regression. *Journal of Fluid Mechanics* 838:42–67.

Luo, W.; Li, J.; Yang, J.; Xu, W.; and Zhang, J. 2018. Convolutional sparse autoencoders for image classification. *IEEE Trans. Neural Netw. Learn. Syst.* 29(7):3289–3294.

Mangan, N.; Bruntona, S.; Proctor, J.; and Kutz, J. 2016. Inferring biological networks by sparse identification of nonlinear dynamics. *IEEE Transactions on Molecular, Biological, and MultiScale Communications* 2(1):52–63.

Mangan, N.; Askham, T.; Brunton, S.; J, N, K.; and Proctor, J. 2019. Model selection for hybrid dynamical systems via sparse regression. *Proc. R. Soc. A* 475(2223):20180534.

Olshausen, B., and Field, D. 1997. Sparse coding with an overcomplete basis set: a strategy employed by v1? *Vision Research* 37(23):3311–3325.

Rudy, S.; Brunton, S.; Proctor, J.; and Kutz, J. 2017. Data-driven discovery of partial differential equations. *Science Advances* 3(4):e160261.

Rudy, S.; Alla, A.; Brunton, S.; Proctor, J.; and Kutz, J. 2019. Data-driven discovery of partial differential equations. *SIAM J. Applied Dynamical Systems* 18(2):643–660.

Schaeffer, H. 2017. Learning partial differential equations via data discovery and sparse optimization. *Proc. R. Soc. A* 473(2197):20160446.

Sirignano, J., and Spiliopoulos, K. 2018. Dgm: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics* 375:1339–1364.

Sun, G.; Cong, Y.; Li, J.; and Fu, Y. 2018. Robust lifelong multi-task multi-view representation learning. In *IEEE Int. Conf. on Big Knowledge (ICBK)*, 91–98.

Sun, G.; Cong, Y.; and Xu, X. 2018. Active lifelong learning with" watchdog". In *Proc. of the AAAI Conf. on Artif. Intell.*, 4107–4114.

Xu, H.; Chang, H.; ; and Zhang, D. 2019. Dl-pde: Deep-learning based data-driven discovery of partial differential equations from discrete and noisy data. *arXiv:1908.04463* 1–24.

Yang, J.; Wang, Z.; Lin, Z.; Cohen, S.; and Huang, T. 2012. Couple dictionary training for image superresolution. *IEEE Trans. Image Process.* 21(8):3467–3478.