DATA-DRIVEN SPATIALLY DEPENDENT PDE IDENTIFICATION

Ruixian Liu, Michael J. Bianco, Peter Gerstoft, Bhaskar D. Rao

University of California, San Diego

ABSTRACT

We propose a data-driven partial differential equation (PDE) identification scheme based on ℓ_1-norm minimization which can identify spatially-dependent PDEs from measurements. Spatially-dependent PDEs refers to that the terms in the PDEs vary across space. In reality a physical system is often governed by spatially-dependent PDEs because the properties of the medium can be various across space, and the proposed method is the first data-driven spatially-dependent PDEs identification scheme. In addition, our method is efficient owing to its non-iterative nature and efficient implementation by coordinate descent.[1]

***Index Terms*—** data-driven, ℓ_1-norm minimization, lasso, spatially-dependent PDEs, efficient identification

1. INTRODUCTION

Physical systems are often described by partial differential equations (PDEs). Given the measurements $\mathbf{U} \in \mathbb{R}^{N_x \times M_t}$ at discrete spatial-temporal coordinates of a physical system $U(x,t)$, the properties of the system are recovered by identifying its governing PDEs from the measurements. Recently, there are many data-driven developments focusing on identifying PDEs [1–11]. As detailed below, they have two limitations: the requirement of assumed active PDE terms and the incapability of recovering spatially-dependent parameters.

First, most papers assume an active PDE-term, as the 1st-order time-derivative U_t [1–7], the 2nd-order time-derivative U_{tt} [8], or multiple PDE terms [9]. They then derive other active PDE terms with coefficients (or learn behaviors contributed by other terms [7]). Thus only parts of the PDE is inferred, this is problematic when the assumed active term is not obvious. E.g., identifying the governing PDE for a wave which can either be an inviscid Burgers' equation ($U_t + UU_x = 0$) or a non-attenuating wave equation ($U_{tt} - c^2\nabla^2 U = 0$) with no PDE terms in common. The two papers not requiring assumed active terms use either sparse Bayesian learning (SBL) [10] or cross-validation (CV) [11]. They both iteratively assume one active term from a library of terms, identify the PDE for each assumption using SBL [10] or sparsity penalized CV [11], and select the best assumption by comparing the posterior confidence [10] or minimal fitting error [11]. They are time consuming as the identification process is repeated for every assumption.

Second, the PDEs governing the observed system can be spatially-dependent. The properties in a system can vary spatially, leading to a PDE with varying PDE coefficients across space. The above methods [1–11] can only identify PDEs with spatially constant terms. The current spatially dependent coefficients recovery schemes are limited for a few specific PDEs [12, 13] and cannot be used for PDE identification since they require the kind of PDE known. We find no methods capable to identify unknown PDEs which are potentially spatially-dependent.

[1]Codes available at: `https://tinyurl.com/2p94zbfw`

This paper aims at solving these two issues above. With no information about the spatial distribution of the PDEs, we must identify the PDE for every location. The challenges to tackle this include: (a) fewer measurements available for one location compared to the whole field; (b) longer CPU time because the process is repeated for all locations. A viable method should be robust to identify the PDE from limited measurements and be computational efficient.

We propose an ℓ_1-norm minimization based data-driven method that identifies unknown spatially-dependent PDEs without assuming active terms. An auxiliary vector robustifies the identification from limited measurements and enables the method to recover all active PDE terms without iterative assumptions. If the PDE is spatially-independent, our method can identify it faster than others since it does not require repeated operations for various assumptions like [10, 11] and has an efficient implementation scheme [14].

Notation: For the measurements $\mathbf{U} \in \mathbb{R}^{N_x \times M_t}$, $\mathbf{U}(i_x, i_t)$ is $\mathbf{U}$ sampled at the coordinate $(i_x\Delta x, i_t\Delta t)$ of U, where $0 \le i_x \le N_x - 1$, $0 \le i_t \le M_t - 1$, and $\Delta x, \Delta t$ are sampling intervals. Sets I_x, I_t contain N spatial and M temporal coordinates within the region of interest (ROI). Use $n \in [1, N]$ as the index of the elements in I_x. The temporal coordinates in I_t are indexed by $m \in [1, M]$. For any matrix $\mathbf{A}$ other than $\mathbf{U}$, its entry $\mathbf{A}(i,j)$ is at ith row and jth column where $i \ge 1,\ j \ge 1$. $\mathbf{A}(i,:)$ denotes the ith row, and $\mathbf{A}(:,j)$ the jth column. For a vector $\mathbf{a}$, its ith entry ($i \ge 1$) can be denoted by either $\mathbf{a}(i)$ or a_i. We start from spatially 1D case in the theory and include a 2D example for experiments.

2. THEORY

Given measurements $\mathbf{U} \in \mathbb{R}^{N_x \times M_t}$, we choose N spatial locations with M time steps as the ROI, and identify the PDE for every spatial point within it. Suppose we know a range of PDEs that potentially govern the observed system, e.g., in this paper, we assume they include the heat equation, the (attenuating) wave equation and its variant WijngGaarden's equation [15], the (viscous) Burgers' equation with a non-linear term UU_x, and the sine-Gordon equation with a non-derivative term $\sin(U)$ [16]. They can model the widely-existing heat diffusion process and various fluid dynamics.

Within the ROI, for the nth spatial location, use $\mathbf{u} \in \mathbb{R}^M$ to denote its measurements at all M time steps, i.e., from $\mathbf{U}(I_x(n), I_t(1))$ to $\mathbf{U}(I_x(n), I_t(M))$. We build a dictionary $\mathbf{\Phi}_n$ containing all $D = 7$ PDE terms potentially appearing in the PDEs mentioned above as:

$$\mathbf{\Phi}_n = \left[\overset{1}{\mathbf{u}_t} \quad \overset{2}{\mathbf{u}_{tt}} \quad \overset{3}{\mathbf{u} \circ \mathbf{u}_x} \quad \overset{4}{\mathbf{u}_{xx}} \quad \overset{5}{\mathbf{u}_{tx}} \quad \overset{6}{\mathbf{u}_{ttxx}} \quad \overset{7}{\sin(\mathbf{u})}\right] \in \mathbb{R}^{M \times D} \tag{1}$$

where each term is an M-length vector evaluated at all the M time steps and the derivatives are computed numerically by finite difference [17], e.g., the mth entry of $\mathbf{u}_t$ is calculated by $[\mathbf{U}(I_x(n), I_t(m)+1) - \mathbf{U}(I_x(n), I_t(m)-1)]/(2\Delta t)$. The $\circ$ denotes element-wise production, e.g., the mth entry of $\mathbf{u} \circ \mathbf{u}_x$ is $\mathbf{U}(I_x(n), I_t(m)) \times \{[\mathbf{U}(I_x(n)+1, I_t(m)) - \mathbf{U}(I_x(n) -$

$1, I_t(m))]/(2\Delta x)\}$. Measurements outside the boundaries of the ROI should exist so that the spatial derivatives for $n=1$ or N and the temporal derivatives for $m=1$ or M can be computed by finite difference, which requires $N<N_x-1$ and $M<M_t-1$.

Originally, the problem can be formed as finding the coefficient $\mathbf{a}_n=[\mathbf{a}_n(1)\ \dots\ \mathbf{a}_n(D)]^{\mathrm{T}}\in\mathbb{R}^D$ such that

$$\mathbf{\Phi}_n\mathbf{a}_n\approx\mathbf{0},\ \|\mathbf{a}_n\|_1>0 \tag{2}$$

where $\|\mathbf{a}_n\|_1>0$ is to avoid the trivial solution $\mathbf{a}_n=\mathbf{0}$, and "$\approx$" is due to $M>D$ (assumed) and the noise introduced by numerical differentiation. To implement (2), we might consider finding $\mathbf{a}_n$ by

$$\mathbf{a}_n=\mathbf{W}_n^{-1}\bar{\mathbf{a}}_n\quad\text{s.t. }\|\mathbf{\Phi}_n\mathbf{W}_n^{-1}\bar{\mathbf{a}}_n\|_2^2\le\mathrm{tol}_n\ \text{ and }\ \|\bar{\mathbf{a}}_n\|_1=1 \tag{3}$$

where the normalization matrix $\mathbf{W}_n\in\mathbb{R}^{D\times D}$ is diagonal with $\mathbf{W}_n(i,i)=\|\mathbf{\Phi}_n(:,i)\|_2$, $\bar{\mathbf{a}}_n=[\bar{\mathbf{a}}_n(1)\dots\bar{\mathbf{a}}_n(D)]^{\mathrm{T}}$ is the coefficient corresponding to the normalized dictionary, and tol_n is the tolerance of fitting error. However, $\|\bar{\mathbf{a}}_n\|_1=1$ is not a convex set. To make (3) solvable using convex optimization, we specify the positive/negative signs in the $\|\bar{\mathbf{a}}_n\|_1$ and thus reduce it to an affine constraint.

We incorporate physical information to reduce $\|\bar{\mathbf{a}}_n\|_1=1$ to $\sum_{i=1}^D s_i\bar{\mathbf{a}}_n(i)=1$, where $s_i\in\{-1,1\}$ is determined by the knowledge of the potential PDE forms, e.g., s_1, s_2 are the same since coefficients for U_t and U_{tt} are always of the same sign for the wave equations we consider [15]. Although there are various PDEs we consider, the relationship of the coefficient signs for the shared terms among different PDEs do not conflict, e.g., the signs for U_t and U_{xx} are always different for the heat equation, the viscous Burgers' equation, or the attenuating wave equation (which indicates s_1 and s_4 are opposite). Setting $s_1=1$, we obtain $\mathbf{s}=[s_1\ \dots\ s_D]\in\mathbb{R}^D$ for the dictionary (1) that is consistent with all the potential PDEs as

$$\mathbf{s}=[1,1,1,-1,-1,-1,1]^{\mathrm{T}} \tag{4}$$

and $\|\bar{\mathbf{a}}_n\|_1=1$ is reduced to $\mathbf{s}^{\mathrm{T}}\bar{\mathbf{a}}_n=1$.

The terms in (1) are for all potential PDEs and the true governing PDE is only consist of a few of them. So it is preferable for the identified PDE to have as fewer active terms as possible under the data fitting constraint. Thus we use the ℓ_1-norm minimization to find $\mathbf{a}_n$ due to its promotion of sparsity [18]:

$$\bar{\mathbf{\Phi}}_n=\mathbf{\Phi}_n\mathbf{W}_n^{-1},\quad \bar{\mathbf{\Phi}}_n^s=\begin{bmatrix}\bar{\mathbf{\Phi}}_n\\ \mathbf{s}^{\mathrm{T}}\end{bmatrix}\in\mathbb{R}^{(M+1)\times D} \tag{5a}$$

$$\widehat{\bar{\mathbf{a}}}_n=\arg\min\|\bar{\mathbf{a}}_n\|_1\ \text{ s.t. }\|\bar{\mathbf{\Phi}}_n^s\bar{\mathbf{a}}_n-\mathbf{e}\|_2^2\le\tau_n \tag{5b}$$

$$\widehat{\mathbf{a}}_n=\mathbf{W}_n^{-1}\widehat{\bar{\mathbf{a}}}_n \tag{5c}$$

where all entries in $\mathbf{e}\in\mathbb{R}^{M+1}$ are zero except $\mathbf{e}(M+1)=1$, $\tau_n>\|\bar{\mathbf{\Phi}}_n^s\bar{\mathbf{\Phi}}_n^{s\dagger}\mathbf{e}-\mathbf{e}\|_2^2$ is a constant larger than the minimal fitting error ($\dagger$ for pseudo-inverse). $\bar{\mathbf{\Phi}}_n$ is normalized. The (5) enables identifying all active PDE terms simultaneously, which is faster than first iteratively assuming one active term in the dictionary and finding others to fit it, and then selecting the correct assumption as in [10, 11].

The physical information in the auxiliary vector $\mathbf{s}$ robustifies the method on identifying PDEs from limited data, because it encourages (5b) to select $\bar{\mathbf{a}}_n$ whose non-zero entries have the same signs as their corresponding entries in $\mathbf{s}$, and thus filter out the potential combinations of columns in $\bar{\mathbf{\Phi}}_n$ which are better fitted but meaningless physically. To see this, suppose there are 2 vectors $\mathbf{p},\mathbf{q}\in\mathbb{R}^D$ satisfying $\bar{\mathbf{\Phi}}_n^s\mathbf{p}\approx\bar{\mathbf{\Phi}}_n^s\mathbf{q}\approx\mathbf{e}$, which indicates $\sum_{i=1}^D s_ip_i\approx\sum_{i=1}^D s_iq_i\approx 1$. If the signs of non-zero entries in $\mathbf{p}$ are the same as corresponding entries in $\mathbf{s}$, while for $\mathbf{q}$ one entry q_{i_0} has the opposite sign of s_{i_0} resulting in $s_{i_0}q_{i_0}<0$, then when $\sum_{i=1}^D s_ip_i$ and $\sum_{i=1}^D s_iq_i$ are sufficiently close to 1, $\|\mathbf{q}\|_1=\sum_{i=1}^D s_iq_i-2s_{i_0}q_{i_0}>1\approx\sum_{i=1}^D s_ip_i=\|\mathbf{p}\|_1$.

We repeat (5) for every location in the ROI (in total N points), and thus recover the physical properties described by spatially dependent PDEs. To solve (5), a convex optimization tool is required. But to our knowledge, there is no such tool that can support solving (5) for all n in a short time when N is large. To accelerate it, we solve (5b) using its Lagrangian (i.e., lasso [19]):

$$\begin{aligned}\widehat{\bar{\mathbf{a}}}_n&=\arg\min_{\bar{\mathbf{a}}_n}\|\bar{\mathbf{a}}_n\|_1+\lambda(\|\bar{\mathbf{\Phi}}_n^s\bar{\mathbf{a}}_n-\mathbf{e}\|_2^2-\tau_n)\\&=\arg\min_{\bar{\mathbf{a}}_n}\|\bar{\mathbf{\Phi}}_n^s\bar{\mathbf{a}}_n-\mathbf{e}\|_2^2+\lambda_n\|\bar{\mathbf{a}}_n\|_1\ .\end{aligned} \tag{6}$$

where $\lambda_n=\frac{1}{\lambda}=0.2\lambda_0$ is chosen empirically with $\lambda_0=2\|\bar{\mathbf{\Phi}}_n^{s\mathrm{T}}\mathbf{e}\|_\infty=2$ the boundary parameter above which the output of (6) is $\mathbf{0}$ according to the lasso path [20]. The advantage of (6) is that the lasso objective can be efficiently solved by coordinate descent [14], where a complete iteration of updating all D entries in $\bar{\mathbf{a}}_n$ costs $O((M+1)D)$ operations, and the number of iterations to reach convergence is often small. Our experiments show that a same $\bar{\mathbf{a}}_n$ can be achieved by solving (6) at least 40x faster than by solving (5b) via CVX [21].

The $\bar{\mathbf{a}}_n$ minimizing (6) may not be sparse enough due to the noise from numerical computation. To further promote sparsity, we threshold entries of $\bar{\mathbf{a}}_n$ using an adaptive threshold proportional to $\|\bar{\mathbf{a}}_n\|_\infty$, and use least squares regression to solve the coefficients only in the T kept entries (denoted by $\tilde{\mathbf{a}}_n(\Omega)$, where Ω with cardinality T is the set of indices for the T kept non-zero entries) and assign 0 to other entries. Thus (5c) is replaced by:

$$\Omega=\{\forall i,|\bar{\mathbf{a}}_n(i)|\ge\epsilon\|\bar{\mathbf{a}}_n\|_\infty\} \tag{7a}$$

$$\tilde{\mathbf{\Phi}}_n^s=[\mathbf{\Phi}_n(:,\Omega)^{\mathrm{T}}\ \mathbf{s}(\Omega)]^{\mathrm{T}}\in\mathbb{R}^{(M+1)\times T},\quad \widehat{\tilde{\mathbf{a}}}_n(\Omega)=\tilde{\mathbf{\Phi}}_n^{s\dagger}\mathbf{e} \tag{7b}$$

where the cut-off parameter $\epsilon=10^{-4}$ is chosen empirically.

If we know in prior that the PDE to be identified is spatially-independent, i.e., the $\mathbf{a}_n$ for $\forall n$ are the same, we can identify the PDE from measurements for all locations simultaneously to increase the robustness of identification, instead of using the result for one location as the PDE for the whole ROI. That is, build $\bar{\mathbf{\Phi}}_{\mathrm{all}}^s$ as:

$$\begin{aligned}\mathbf{\Phi}_{\mathrm{all}}&=[\mathbf{\Phi}_1^{\mathrm{T}}\ \mathbf{\Phi}_2^{\mathrm{T}}\ \dots\ \mathbf{\Phi}_N^{\mathrm{T}}]^{\mathrm{T}}\in\mathbb{R}^{NM\times D}\\ \bar{\mathbf{\Phi}}_{\mathrm{all}}^s&=[(\mathbf{\Phi}_{\mathrm{all}}\mathbf{W}_{\mathrm{all}}^{-1})^{\mathrm{T}}\ \mathbf{s}]^{\mathrm{T}}\in\mathbb{R}^{(NM+1)\times D}\end{aligned} \tag{8}$$

where the normalization matrix $\mathbf{W}_{\mathrm{all}}\in\mathbb{R}^{D\times D}$ is diagonal with $\mathbf{W}_{\mathrm{all}}(i,i)=\|\mathbf{\Phi}_{\mathrm{all}}(:,i)\|_2$. Let $\mathbf{e}_{\mathrm{all}}\in\mathbb{R}^{NM+1}$ be an all-zero vector except its last entry being one, we solve the coefficient $\bar{\mathbf{a}}\in\mathbb{R}^D$ by

$$\widehat{\bar{\mathbf{a}}}=\arg\min_{\bar{\mathbf{a}}}\|\bar{\mathbf{\Phi}}_{\mathrm{all}}^s\bar{\mathbf{a}}-\mathbf{e}_{\mathrm{all}}\|_2^2+\lambda_{\mathrm{all}}\|\bar{\mathbf{a}}\|_1 \tag{9}$$

where $\lambda_{\mathrm{all}}=0.2\lambda_0$ is chosen empirically with $\lambda_0=2\|\bar{\mathbf{\Phi}}_{\mathrm{all}}^{s\mathrm{T}}\mathbf{e}_{\mathrm{all}}\|_\infty=2$ the boundary parameter above which (9) will output $\mathbf{0}$. Finally, the coefficients in the correct scale are obtained by implementing (7) with $\bar{\mathbf{a}}_n,\ \mathbf{\Phi}_n,\ \mathbf{e}$ replaced by $\bar{\mathbf{a}},\ \mathbf{\Phi}_{\mathrm{all}}$ and $\mathbf{e}_{\mathrm{all}}$ respectively.

3. NUMERICAL EXPERIMENTS

This part includes 1 spatially independent (which is not known in prior) and 2 spatially dependent PDE identification experiments, and an efficiency comparison between the proposed, the SBL based [10] and the cross-validation based [11] methods for spatially independent (known in prior) PDE identification. All datasets are generated by finite difference modeling.

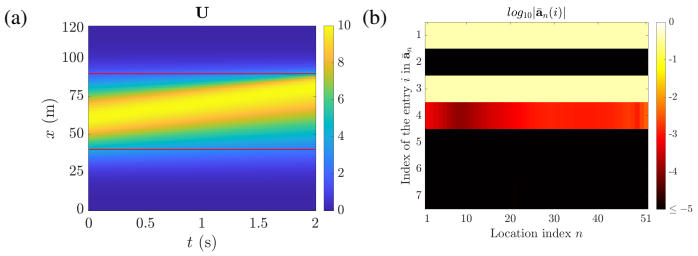

Fig. 1: For the Burgers' equation (10), (a) $\mathbf{U}$ with spatially independent $\nu = 0.15$, $\Delta x = 1$ m, $\Delta t = 0.01$ s. So $0 \le i_x \le 120$, $0 \le i_t \le 200$. The ROI is $\{i_x | 40 \le i_x \le 90\}$ between the red lines where obvious dynamics can be observed. (b) $log_{10}|\bar{\mathbf{a}}_n(i)|$ where i corresponds to the superscripts in (1) for all $n = 1 \dots 51$, magnitudes less than 10^{-5} are not shown.

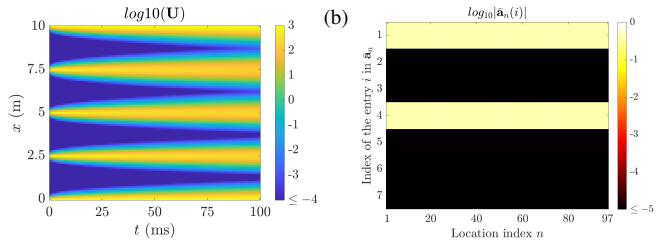

Fig. 2: For the heat equation (11), (a) $log_{10}(\mathbf{U})$ with spatially dependent α, $\Delta x = 0.1$ m, $\Delta t = 5 \times 10^{-2}$ ms. So $0 \le i_x \le 100$, $0 \le i_t \le 2000$. Magnitudes less than 10^{-4} are not shown. (b) $log_{10}|\bar{\mathbf{a}}_n(i)|$ where i corresponds to the superscripts in (1) for all $n = 1 \dots 97$, magnitudes less than 10^{-5} are not shown.

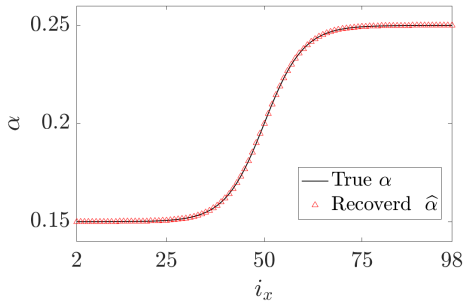

Fig. 3: For the heat equation, the true α and the recovered $\widehat{\alpha}$ in the ROI. They are basically equivalent with $\text{RMSE} = 1.4 \times 10^{-16}$.

3.1. Spatially-independent Burgers' equation

The Burgers' equation

$$U_t + UU_x - \nu U_{xx} = 0 \tag{10}$$

is non-linear and can model the formation of shock waves, where $\nu \ge 0$ is the viscosity being a constant for a same liquid.

A dataset $\mathbf{U} \in \mathbb{R}^{121 \times 201}$ with $\Delta t = 0.01$ s and $\Delta x = 1$ m modeling the process governed by (10) where $\nu = 0.15$ is generated as Fig.1a. The initial state is a scaled normal distribution probability density function (PDF), and as time goes by, the peak is tilting to the positive direction of x-axis, which forms a shock.

We choose the spatial region where the dynamics can be easily observed as the ROI, to be specific, choose $I_x = \{i_x | 40 \le i_x \le 90\}$ which is bounded by the red lines in Fig.1a, so $N = 51$ and $n = 1$ corresponds to $i_x = 40$. We do not consider the temporal boundaries where the derivatives are not well defined and thus use $I_t = \{i_t | 1 \le i_x \le 199\}$ for the ROI, i.e., $M = 199$.

Build $\bar{\mathbf{\Phi}}_n^s$ according to (1),(4) and (5a) for every $1 \le n \le N$. From (6), the coefficients are distributed as Fig.1b. After thresholding, $\{\mathbf{u}_t, \mathbf{u} \circ \mathbf{u}_x, \mathbf{u}_{xx}\}$ appearing in the Burgers' equation are selected for all locations in the ROI.

For every location, we build $\tilde{\mathbf{\Phi}}_n^s$ and compute $\tilde{\mathbf{a}}_n$ as (7). The coefficients for U_t and UU_x are always nearly identical since $\sum_{n=1}^{N} |\tilde{\mathbf{a}}_n(3) - \tilde{\mathbf{a}}_n(1)| = 3.0 \times 10^{-14}$. The estimated viscosity is $\widehat{\nu} = -\frac{\tilde{\mathbf{a}}_n(4)}{\tilde{\mathbf{a}}_n(1)}$, which equals to 0.15 for every n. This example indicates the method can work when the PDE to be identified is in fact spatially-independent but not known in prior.

3.2. Spatially-dependent heat equation

The heat equation is the PDE that models the diffusion of heat within medium. The equation is

$$U_t - \alpha U_{xx} = 0 \tag{11}$$

where $\alpha > 0$ is the potentially spatially-dependent thermal diffusivity of the medium.

We generate a dataset $\mathbf{U} \in \mathbb{R}^{101 \times 2001}$ depicting the process of the heat diffusion from 3 sources at the beginning time $\mathbf{U}(25, 0) = \mathbf{U}(50, 0) = \mathbf{U}(75, 0) = 1000$ and the boundaries for all the time $\mathbf{U}(0, :) = \mathbf{U}(100, :) = 1000$ to other places with zero heat originally, as shown in Fig.2a with $\Delta x = 0.1$ m and $\Delta t = 5 \times 10^{-5}$ s.

We choose the ROI to be $I_x = \{i_x | 2 \le i_x \le 98\}$ so that the boundaries and their immediate neighboring locations influenced by them when calculating derivatives by finite difference are excluded, and choose $I_t = \{i_t | 1 \le i_t \le 1999\}$, so $N = 97$, $M = 1999$ and $n = 1$ corresponds to $i_x = 2$. The diffusivity α is distributed as a $\tanh$ function offset to positive values:

$$\alpha(x) = 0.05 \tanh(x - 50\Delta x) + 0.2,\ 0 \le x \le 100\Delta x. \tag{12}$$

Build $\bar{\mathbf{\Phi}}_n^s$ according to (1),(4) and (5a) for every $1 \le n \le N$. From (6), the coefficients are distributed as Fig.2b. After thresholding, $\{\mathbf{u}_t, \mathbf{u}_{xx}\}$ appearing in the heat equation are selected for all locations within the ROI.

For every location in the ROI, we build $\tilde{\mathbf{\Phi}}_n^s$ and compute $\tilde{\mathbf{a}}_n$ as (7), the diffusivity is recovered by $\widehat{\alpha} = -\frac{\tilde{\mathbf{a}}_n(4)}{\tilde{\mathbf{a}}_n(1)}$, shown by the red triangles in Fig.3. The recovery is nearly perfect with the root-mean-square error $\text{RMSE} = 1.4 \times 10^{-16}$.

3.3. 2D spatially-dependent wave equation

A 2D wavefield $\mathbf{U} \in \mathbb{R}^{32 \times 32 \times 200}$ in which $\Delta x = \Delta y = 0.1$ m and $\Delta t = 0.01$ s describing waves excited by an initial perturbation and propagating through various media is used for experiments. The PDE governing it is the wave equation

$$U_{tt} + \alpha U_t - c^2 \nabla^2 U = 0 \tag{13}$$

where $\alpha \ge 0$ the attenuating factor, $c > 0$ the phase speed and ∇^2 the Laplacian, i.e., $U_{xx} + U_{yy}$. The initial perturbation is shaped as a scaled 2D normal distribution PDF, and the phase speeds $2 \le c \le 3$ m/s and attenuation $\alpha \in \{0, 0.025, 0.05\}$ are varying across the

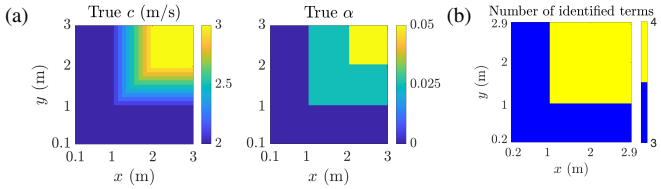

Fig. 4: (a) The true attenuating factors α and phase speeds c for i_x, i_y in $[1, 30]$ ($\Delta x = \Delta y = 0.1$ m). Waves can not arrive the places where either i_x or i_y being 0 or 31 because of the boundary condition. (b) Number of identified active PDE terms within the ROI.

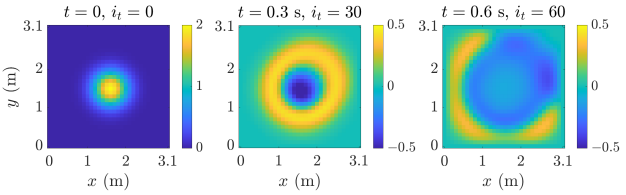

Fig. 5: The wavefield governed by (13) with spatially dependent c and α at 3 selected time points. $\Delta x = \Delta y = 0.1$ m, $\Delta t = 0.01$ s. So i_x, i_y both ranged from 0 to 31.

domain as shown in Fig.4a. Some frames are shown in Fig.5. The spatial boundaries are kept zero. We choose the ROI to be all the spatial regions without the boundaries and its immediate neighboring points (i.e., $2 \le i_x \le 29$, $2 \le i_y \le 29$) and the time steps $1 \le i_t \le 198$. So $N = 28^2 = 784$, $M = 198$ and $n = 1$ corresponds to $i_x = i_y = 2$. The 2D locations within ROI are indexed from 1 to N in the row-major manner.

For this 2D case, we extend the dictionary (1) and $\mathbf{s}$ (4) to

$$\begin{aligned} \mathbf{\Phi}_n = [&\overset{1}{\mathbf{u}}_t \ \overset{2}{\mathbf{u}}_{tt} \ \overset{3}{\mathbf{u}} \circ \mathbf{u}_x \ \overset{4}{\mathbf{u}} \circ \mathbf{u}_y \ \overset{5}{\mathbf{u}}_{xx} \ \overset{6}{\mathbf{u}}_{yy} \\ &\overset{7}{\mathbf{u}}_{tx} \ \overset{8}{\mathbf{u}}_{ttxx} \ \overset{9}{\mathbf{u}}_{ty} \ \overset{10}{\mathbf{u}}_{ttyy} \ \overset{11}{\sin(\mathbf{u})}] \\ \mathbf{s} = [&1, 1, 1, 1, -1, -1, -1, -1, -1, -1, 1]^{\mathrm{T}} . \end{aligned} \tag{14}$$

Build dictionaries using (14) according to (5a), and from (6), the $\bar{\mathbf{a}}_1$ to $\bar{\mathbf{a}}_N$ are acquired as Fig.6(a). After thresholding, the kept non-zero entries are indicated in Fig.6(b).

Comparing Fig.6(b) to (13), the method successfully identifies the PDEs for all 784 locations. Fig.4b shows the number of identified active PDE terms for each location in the ROI, where the 3

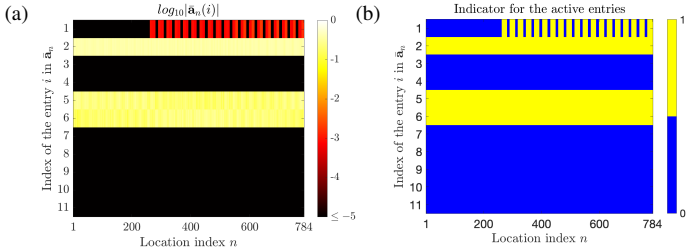

Fig. 6: For 2D wave equation, (a) $log_{10}|\bar{\mathbf{a}}_n(i)|$ where i corresponds to the superscripts of $\mathbf{\Phi}_n$ in (14) for all $n = 1 \ldots 28^2$; (b) the locations of their active entries after thresholding. The 2D 28×28 locations are indexed from 1 to 784 in the row-major manner, i.e., $n = 1, \ldots, 28$ corresponds to $i_x = 2, \ldots, 29$ with $i_y = 2$, and $n = 29, \ldots, 56$ corresponds to $i_x = 2, \ldots, 29$ with $i_y = 3$, etc.

	SBL [10]	CV [11]	Proposed
Burgers' eq.	0.98 s	0.30 s	0.01 s
Heat eq.	15.87 s	0.51 s	0.02 s
Non-attenuating wave eq.	30.48 s	16.92 s	0.03 s
Attenuating wave eq.	30.16 s	16.74 s	0.03 s

Table 1: The average time used for 10 trials to correctly identify spatially independent PDEs on 4 datasets using the SBL based (SBL) [10], the cross-validation based (CV) [11] and the proposed method. All tests are run on a same computer.

terms contain $\{\mathbf{u}_{tt}, \mathbf{u}_{xx}, \mathbf{u}_{yy}\}$, and for 4 terms $\mathbf{u}_t$ is also included. Build $\tilde{\mathbf{\Phi}}_n^s$ and compute $\tilde{\mathbf{a}}_n$ as (7), the coefficients for U_{xx} and U_{yy} are nearly identical as $\sum_{n=1}^{N} |\tilde{\mathbf{a}}_n(5) - \tilde{\mathbf{a}}_n(6)| = 3.4 \times 10^{-15}$. The recovered $\widehat{c} = \sqrt{-\frac{\tilde{\mathbf{a}}_n(5)+\tilde{\mathbf{a}}_n(6)}{2\tilde{\mathbf{a}}_n(2)}}$ and $\widehat{\alpha} = \frac{\tilde{\mathbf{a}}_n(1)}{\tilde{\mathbf{a}}_n(2)}$, which are nearly perfect as the RMSE $= 5.1 \times 10^{-15}$ for phase speeds and 9.9×10^{-15} for attenuating factors with respect to the ground truth in Fig.4a across the whole ROI.

3.4. Compare efficiency: identify spatially independent PDEs

We use 4 datasets governed by spatially independent PDEs for efficiency comparison between the proposed method and [10, 11] that do not require assumed active term. One is the data used in Sec.3.1. Three other datasets describe fields governed by: (a) heat equation (11) with $\alpha = 0.2$; (b) non-attenuating 2D wave equation (13) with $\alpha = 0, c = 2.5$ m/s; (c) attenuating 2D wave equation (13) with $\alpha = 0.025, c = 2.5$ m/s. All other settings (dataset size, sampling interval, ROI, etc.) are the same as in Sec. 3.2 and 3.3. All experiments are performed on one MacBook Pro.

For the SBL based method [10], the dictionary

$$\mathbf{\Phi} = [\mathbf{\Phi}_1^{\mathrm{T}} \ \mathbf{\Phi}_2^{\mathrm{T}} \ \ldots \ \mathbf{\Phi}_N^{\mathrm{T}}]^{\mathrm{T}} \tag{15}$$

is built as required in [10]. It correctly identified the PDEs for all cases. For the cross-validation based method, also use (15) as the input and it successfully identified all PDEs. We implement each of the methods 10 times, and record the average time used in Table.1.

For the proposed method, the PDEs are also correctly identified from (8) and (9), and Table.1 shows the average time used for 10 implementations. Comparing the time spent, the proposed method is significantly more efficient. This owes to its non-iterative nature regarding to assumed active PDE terms and the efficient implementation of lasso (i.e., coordinate descent [14]).

4. CONCLUSION

We proposed an ℓ_1-norm minimization based PDE identification method that can identify spatially-dependent PDEs, and validated it with multiple numerical experiments. The method can also identify spatially-independent PDEs without assumed active terms more efficiently than previous developments.

5. REFERENCES

[1] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proc. Natl. Acad. Sci.*, vol. 113, no. 15, pp. 3932–3937, 2016.

[2] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Data-driven discovery of partial differential equations," *Sci. Adv.*, vol. 3, no. 4, pp. e1602614, 2017.

[3] H. Schaeffer, G. Tran, and R. Ward, "Extracting sparse high-dimensional dynamics from limited data," *SIAM J. Appl. Math.*, vol. 78, no. 6, pp. 3279–3295, 2018.

[4] Z. Long, Y. Lu, and B. Dong, "Pde-net 2.0: Learning pdes from data with a numeric-symbolic hybrid deep network," *J. Comput. Phys*, vol. 399, pp. 108925, 2019.

[5] S. L. Brunton and J. N. Kutz, *Data-driven science and engineering: Machine learning, dynamical systems, and control*, Cambridge University Press, 2019.

[6] P. A.K. Reinbold, D. R. Gurevich, and R. O. Grigoriev, "Using noisy or incomplete data to discover models of spatiotemporal dynamics," *Phys. Rev. E*, vol. 101, no. 1, pp. 010203, 2020.

[7] K. Wu and D. Xiu, "Data-driven deep learning of partial differential equations in modal space," *J. Comput. Phys.*, vol. 408, pp. 109307, 2020.

[8] R. Liu, M. J. Bianco, and P. Gerstoft, "Wave equation extraction from a video using sparse modeling," in *Proc.53th Asilomar Conf. on Circuits, Systems and Computers.* IEEE, 2019, pp. 2160–2165.

[9] 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," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019.

[10] S. Zhang and G. Lin, "Robust data-driven discovery of governing physical laws with error bars," *Proc. Math. Phys. Eng. Sci.*, vol. 474, no. 2217, pp. 20180305, 2018.

[11] R. Liu, M. Bianco, and P. Gerstoft, "Automated partial differential equation identification," *J. Acoust. Soc. Am.*, vol. 150, no. 4, pp. 2364–2374, 2021.

[12] M. van Berkel, G. Vandersteen, E. Geerardyn, R. Pintelon, H. Zwart, and M. de Baar, "Frequency domain sample maximum likelihood estimation for spatially dependent parameter estimation in pdes," *Automatica*, vol. 50, no. 8, pp. 2113–2119, 2014.

[13] S. Kramer and E. M. Bollt, "Spatially dependent parameter estimation and nonlinear data assimilation by autosynchronization of a system of partial differential equations," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 23, no. 3, pp. 033101, 2013.

[14] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *J. Stat. Softw.*, vol. 33, no. 1, pp. 1, 2010.

[15] M. J. Buckingham, "On the transient solutions of three acoustic wave equations: van wijngaarden's equation, stokes' equation and the time-dependent diffusion equation," *J. Acoust. Soc. Am*, vol. 124, no. 4, pp. 1909–1920, 2008.

[16] J. Rubinstein, "Sine-gordon equation," *J. Math. Phys.*, vol. 11, no. 1, pp. 258–266, 1970.

[17] W. F. Ames, *Numerical methods for partial differential equations*, Academic press, 2014.

[18] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell 1$ minimization," *Proc. Natl. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, 2003.

[19] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[20] P. Gerstoft, A. Xenaki, and C. F. Mecklenbräuker, "Multiple and single snapshot compressive beamforming," *J. Acoust. Soc. Am.*, vol. 138, no. 4, pp. 2003–2014, 2015.

[21] M. Grant and S. Boyd, "Cvx: Matlab software for disciplined convex programming, version 2.1," 2014.