A Causality-Based Learning Approach for Underlying Dynamics of Complex Dynamical Systems

Yinling Zhang
Joint work with Prof. Nan Chen

April 26, 2023

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Content

- Introduction
- Overview of the Causality-Based Learning Algorithm
    - Conditional Sampling
    - Causal Inference
    - Parameter Estimation
- A simple Example for Proof-of-Concept
- A High-Dimensional System with Stochastic Parameterizations
- Summary

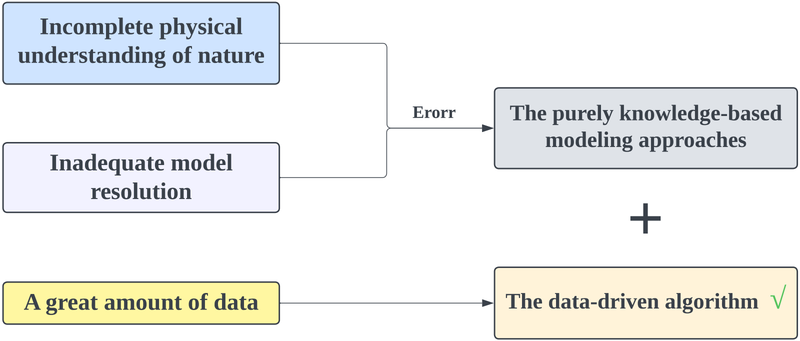
Content

- Introduction
- Overview of the Causality-Based Learning Algorithm
  - Conditional Sampling
  - Causal Inference
  - Parameter Estimation
- A simple Example for Proof-of-Concept
- A High-Dimensional System with Stochastic Parameterizations
- Summary

Introduction

1. Complex turbulent dynamical systems appear in many areas.
2. Key features:
    - Strong nonlinearity
    - High dimensionality
    - Multiscale structures
3. Important Task:
    - State estimation
    - Uncertainty quantification
    - Prediction
4. A suitable model is important.

Introduction



Introduction

Existing Methods:

1. Based on constrained optimizations
2. Determine regularization parameter in advance
3. A certain sparse identification technique is used in the optimization procedure
4. Sensitive to noises
5. Hard to identify the physical meaning

Introduction

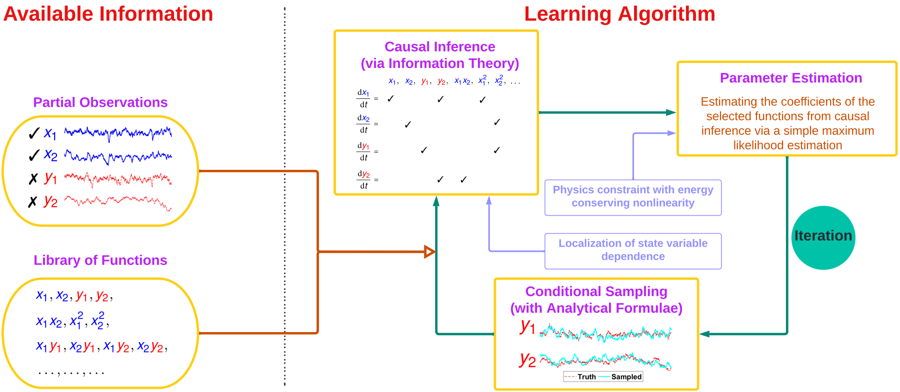
Causality-Based Methods(Chen and Zhang [2022]):

1. Use information theory to clarify physical meaning
2. Estimating parameters via quadratic optimization formula
3. Robust to stochastic noises
4. Applicable to the situation with partial observations

Content

- Introduction
- Overview of the Causality-Based Learning Algorithm
    - Conditional Sampling
    - Causal Inference
    - Parameter Estimation
- A simple Example for Proof-of-Concept
- A High-Dimensional System with Stochastic Parameterizations
- Summary

Overview of the Causality-Based Learning Algorithm



The algorithm has three components, we use the iterative method to learn the underlying dynamics and stochastic parameterization.

Content

Conditional Sampling

The general stochastic parameterization structure:

$$\frac{\mathrm{d}\mathbf{X}}{\mathrm{d}t} = \Big[\mathbf{A_0}(\mathbf{X},t) + \mathbf{A_1}(\mathbf{X},t)\mathbf{Y}(t)\Big] + \mathbf{B_1}(\mathbf{X},t)\dot{\mathbf{W}}_\mathbf{1}(t), \tag{1a}$$

$$\frac{\mathrm{d}\mathbf{Y}}{\mathrm{d}t} = \Big[\mathbf{a_0}(\mathbf{X},t) + \mathbf{a_1}(\mathbf{X},t)\mathbf{Y}(t)\Big] + \mathbf{b_2}(\mathbf{X},t)\dot{\mathbf{W}}_\mathbf{2}(t), \tag{1b}$$

where $\mathbf{X}$ is the observed state variable, and $\mathbf{Y}$ is the stochastic parameterization. $\mathbf{X}$ could have arbitrary nonlinearity while $\mathbf{Y}$ is conditionally linear once $\mathbf{X}$ is given. But (1b) is overall highly nonlinear so $\mathbf{Y}$ could create non-Gaussian features.

- Given this conditionally Gaussian structure, we can use the closed analytic formula to sample the trajectory of $\mathbf{Y}$ that could significantly reduce the computational cost.
- The conditional sampling is based on the Bayesian framework.

Conditional Sampling

Many complex nonlinear systems already fit into this framework of (1), Chen Majda 2018 Entropy, Chen, Li and Liu, 2022 Chaos.

- Physics-constrained nonlinear stochastic models.

  Examples: the noisy versions of Lorez models, Charney-DeVore flows, and the paradigm model for topographic mean flow interactions.

- Stochastically coupled reaction-diffusion models in neuroscience and ecology.

  Examples: the FitzHugh-Nagumo models and the SIR epidemic models.

- Multi-scale models in turbulence and geophysical flows.

  Examples: the Boussinesq equations and rotating shallow water equation.

Conditional Sampling

The purpose of stochastic parameterization:

- NOT to recover the exact dynamics of the unobserved variables. ✗
- Recover the statistic feedback from the unobserved variable $\mathbf{Y}$ to the observed variable $\mathbf{X}$. ✔

Content

$$\mathbf{Z} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$$

$$\begin{aligned} \begin{bmatrix} \frac{\mathrm{d}z_1}{\mathrm{d}t} \\ \frac{\mathrm{d}z_2}{\mathrm{d}t} \\ \vdots \\ \frac{\mathrm{d}z_N}{\mathrm{d}t} \end{bmatrix} &= \begin{bmatrix} \xi_{1,1} & \cdots & \xi_{1,M} \\ \xi_{2,1} & \cdots & \xi_{2,M} \\ \vdots & \ddots & \vdots \\ \xi_{N,1} & \cdots & \xi_{N,M} \end{bmatrix} \begin{bmatrix} f_1\left(z_1(t), \ldots, z_N(t), t\right) \\ f_2\left(z_1(t), \ldots, z_N(t), t\right) \\ \vdots \\ f_M\left(z_1(t), \ldots, z_N(t), t\right) \end{bmatrix} \\ &= \mathbf{\Xi} \times \mathbf{F}\left(\mathbf{Z}(t), t\right), \end{aligned} \tag{2}$$

where $\mathbf{\Xi}$ is the coefficient matrix to be estimated, $\mathbf{Z}$ is the state variable and f_i is the candidate function in the function library.

GOAL: Determine the non-zero entries of this matrix $\mathbf{\Xi}$.

METHOD: A new concept introduced here is causation entropy.

The causation entropy $C_{f_m \to z_n | [\mathbf{F} \backslash f_m]}$ is defined as follows,

$$\begin{aligned} C_{f_m \to z_n | [\mathbf{F} \backslash f_m]} &= H(z_n | \left[\mathbf{F} \backslash f_m\right]) - H(z_n | \left[\mathbf{F} \backslash f_m\right], f_m) \\ &= H(z_n | \left[\mathbf{F} \backslash f_m\right]) - H(z_n | \mathbf{F}). \end{aligned} \tag{3}$$

where $H(\cdot|\cdot)$ is the conditional entropy.

If such a causation entropy is zero (or practically nearly zero), then $f_m(t)$ does not contribute any information to $\frac{\mathrm{d}z_n}{\mathrm{d}t}$ and the associated parameter $\xi_{n,m}$ is set to be zero.

Causal Inference

Practical calculation of the causation entropy:

$$
\begin{aligned}
C_{Z\to X|Y} &= H(X|Y) - H(X|Y,Z) \\
&= H(X,Y) - H(Y) - H(X,Y,Z) + H(Y,Z) \\
&= \frac{1}{2}\ln(\det(\mathbf{R}_{XY})) - \frac{1}{2}\ln(\det(\mathbf{R}_{Y})) \\
&\qquad - \frac{1}{2}\ln(\det(\mathbf{R}_{XYZ})) + \frac{1}{2}\ln(\det(\mathbf{R}_{YZ})),
\end{aligned} \tag{4}
$$

where $\mathbf{R}_{XYZ}$ denotes the covariance matrix of the state variables $(X,Y,Z)^{\mathtt{T}}$ and similar for other covariances.

Problem: How about non-Gaussian distribution?

- The primary goal is not to obtain the exact value of causation entropies, we want to find if the causation entropy is zero or not.
- Usually if a significant causal relationship is detected in the higher order moments, then very likely it exists in the Gaussian approximation as well.

Content

- Introduction
- Overview of the Causality-Based Learning Algorithm
    - Conditional Sampling
    - Causal Inference
    - Parameter Estimation
- A simple Example for Proof-of-Concept
- A High-Dimensional System with Stochastic Parameterizations

Parameter Estimation

Parameter estimation via a simple maximum likelihood estimation

$$\hat{\mathbf{\Theta}} = \arg\min \mathcal{L}(\mathbf{\Theta}). \tag{5}$$

Physics constraints, together with other constraints, can in general be represented in the following way:

$$\mathbf{H}\mathbf{\Theta} = \mathbf{g}, \tag{6}$$

where $\mathbf{H}$ and $\mathbf{g}$ are constant matrices.

To incorporate these constraints, the Lagrangian multiplier method is applied,

$$\mathcal{L} = \mathcal{L}(\mathbf{\Theta}; \lambda), \quad \hat{\mathbf{\Theta}} = \arg\min \mathcal{L}(\mathbf{\Theta}; \lambda). \tag{7}$$

Content

- Introduction
- Overview of the Causality-Based Learning Algorithm
  - Conditional Sampling
  - Causal Inference
  - Parameter Estimation
- A simple Example for Proof-of-Concept
- A High-Dimensional System with Stochastic Parameterizations
- Summary

A Simple Example for Proof-of-Concept

The first test example is a low-dimensional chaotic system, known as the Lorenz 1984 (L-84) model,

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= -(y^2+z^2) - a(x-f) + \sigma_x \dot{W}_x,\\ \frac{\mathrm{d}y}{\mathrm{d}t} &= -bxz + xy - y + g + \sigma_y \dot{W}_y,\\ \frac{\mathrm{d}z}{\mathrm{d}t} &= bxy + xz - z + \sigma_z \dot{W}_z.\end{aligned} \tag{8}$$

In (8), the zonal flow x represents the intensity of the mid-latitude westerly wind current, and a wave component exists with y and z representing the cosine and sine phases of a chain of vortices superimposed on the zonal flow.

Here y and z are observed variables, and x is the unobserved variable.

A Simple Example for Proof-of-Concept

The library of functions:

$$
\begin{gathered}
y,\ z,\ y^2,\ z^2,\ yz,\ 1,\ x,\ xy,\ xz, \\
xy^2,\ xz^2,\ xyz.
\end{gathered}
\tag{9}
$$

A random and complicated initial model structure is utilized to start the iterative algorithm,

$$
\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= y^2 - z^2 + 2 + (y^2 - z^2)x + \sigma_x \dot{W}_x, \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= -y - 2y^2 + z^2 + 1 + (-y - 8z - yz)x + \sigma_y \dot{W}_y, \\
\frac{\mathrm{d}z}{\mathrm{d}t} &= -z + z^2 - yz + (8y + z + z^2)x + \sigma_z \dot{W}_z.
\end{aligned}
\tag{10}
$$

A Simple Example for Proof-of-Concept

The identified model has the same model structure as the truth,

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= \theta^x_{yy}y^2 + \theta^x_{zz}z^2 + \theta^x_x x + \theta^x_1 + \sigma_x \dot{W}_x,\\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \theta^y_{xz}xz + \theta^y_{xy}xy + \theta^y_y y + \theta^y_1 + \sigma_y \dot{W}_y,\\ \frac{\mathrm{d}z}{\mathrm{d}t} &= \theta^z_{xy}xy + \theta^z_{xz}xz + \theta^z_z z + \theta^z_1 + \sigma_z \dot{W}_z.\end{aligned} \tag{11}$$

	θ^x_x	θ^y_y	θ^z_z	θ^x_{yy}	θ^x_{zz}	θ^y_{xz}
Truth	-0.2500	-1.0000	-1.0000	-1.0000	-1.0000	-4.0000
Identified	-0.2680	-0.9987	-1.0076	-0.9993	-1.0061	-3.9956
	θ^y_{xy}	θ^z_{xy}	θ^z_{xz}	θ^x_1	θ^y_1	θ^z_1
Truth	1.0000	4.0000	1.0000	2.0000	1.0000	0.0000
Identified	0.9993	3.9956	1.0061	2.0223	0.9939	0.0053

Table: Comparison of the parameters in the true system (8) and those in the identified model.

A Simple Example for Proof-of-Concept

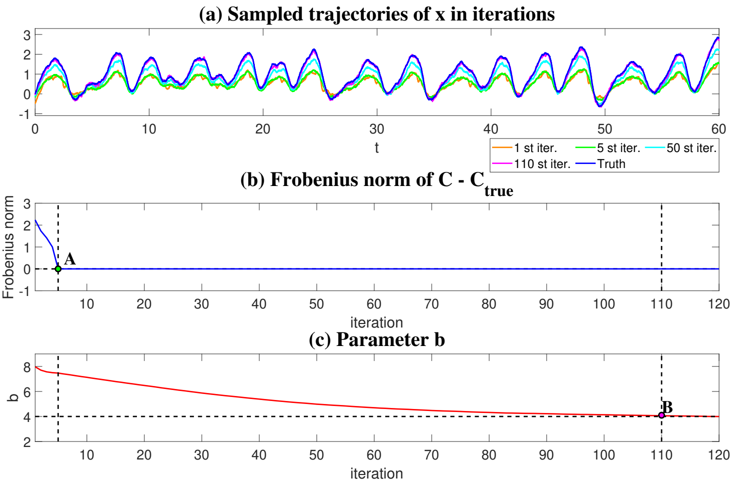


Figure: Iterative procedure of learning the L-84 model with partial observations $(y, z)^{\mathtt{T}}$.

Content

- Introduction
- Overview of the Causality-Based Learning Algorithm
  - Conditional Sampling
  - Causal Inference
  - Parameter Estimation
- A simple Example for Proof-of-Concept
- A High-Dimensional System with Stochastic Parameterizations
- Summary

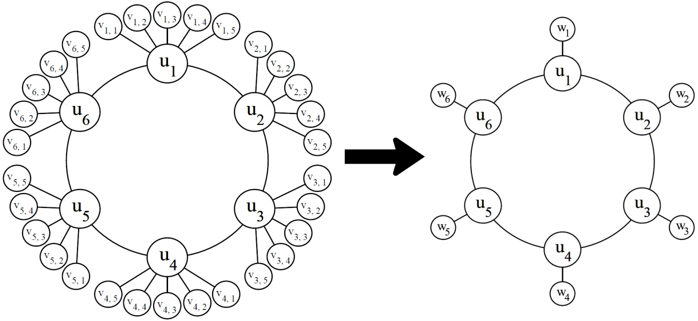
A High-Dimensional System with Stoch Parameterizations

The two-layer Lorenz 1996 (L-96) model is a conceptual representation of geophysical turbulence that is commonly used as a test for data assimilation and parameterization in numerical weather forecasting.

$$\begin{aligned}\frac{\mathrm{d}u_i}{\mathrm{d}t} &= -u_{i-1}\left(u_{i-2}-u_{i+1}\right) - u_i + f - \frac{hc_i}{J}\sum_{j=1}^{J} v_{i,j} \\ &\qquad + \sigma_{u_i}\dot{W}_{u_i}, \quad i = 1,\ldots,I, \qquad \text{(12a)}\\ \frac{\mathrm{d}v_{i,j}}{\mathrm{d}t} &= -bc_i v_{i,j+1}\left(v_{i,j+2}-v_{i,j-1}\right) - c_i v_{i,j} + \frac{hc_i}{J}u_i \\ &\qquad + \sigma_{v_{i,j}}\dot{W}_{v_{i,j}}, \quad j = 1,\ldots,J, \qquad \text{(12b)}\end{aligned}$$

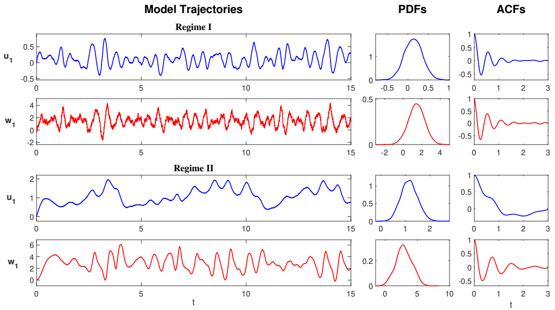
where I denotes the total number of large-scale variables and J is the number of small-scale variables corresponding to each large-scale variable.

A High-Dimensional System with Stoch Parameterizations



For the convenience of discussing the behavior of the two layers, a new single variable $w_i = \sum_{j=1}^{J} v_{i,j}$ is introduced, which describes the total variabilities in the second layer.

A High-Dimensional System with Stoch Parameterizations



Two dynamical regimes are considered here as the truth. They share most of the parameters:

$$I = 20, \quad J = 4, \quad c_i = 2 + 0.7\cos(2\pi i/I), \quad b = 2, \quad f = 4, \quad \sigma_{u_i} = 0.05, \tag{13}$$

but they are differed by h and $\sigma_{v_{i,j}}$:

$$\begin{aligned} &\text{Regime I:} && h = 4.0 && \text{and} && \sigma_{v_{i,j}} = 1.00 \\ &\text{Regime II:} && h = 1.5 && \text{and} && \sigma_{v_{i,j}} = 0.05. \end{aligned} \tag{14}$$

The set of the candidate functions for u_i is then given by a vector $\mathbf{F}_{u_i}$, which includes 23 terms:

$$\begin{gathered} u_i,\ u_{i-1},\ u_{i-2},\ u_{i+1},\ u_{i+2},\ u_i^2,\ u_{i-1}^2,\ u_{i-2}^2,\ u_{i+1}^2, \\ u_{i+2}^2,\ u_iu_{i-1},\ u_iu_{i-2},\ u_iu_{i+1},\ u_iu_{i+2},\ u_{i-1}u_{i-2},\ u_{i-1}u_{i+1}, \\ u_{i-1}u_{i+2},\ u_{i-2}u_{i+1},\ u_{i-2}u_{i+2},\ u_{i+1}u_{i+2},\ 1,\ w_i,\ u_iw_i. \end{gathered} \tag{15}$$

Only 4 terms are included in the library for each w_i, given by another vector $\mathbf{F}_{w_i}$,

$$u_i,\ u_i^2,\ 1,\ w_i. \tag{16}$$

It's obvious that

$$\mathbf{F}_{w_i} \subset \mathbf{F}_{u_i} := \mathbf{T}$$

A High-Dimensional System with Stoch Parameterizations

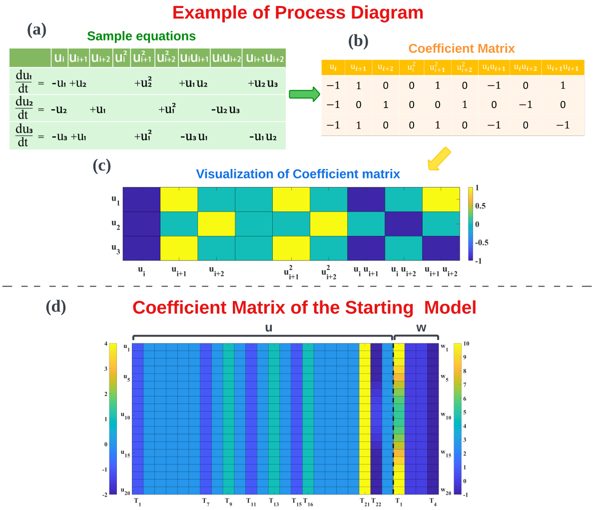


Figure: Using a visualization diagram to represent the identified model structure and parameters.

A High-Dimensional System with Stoch Parameterizations

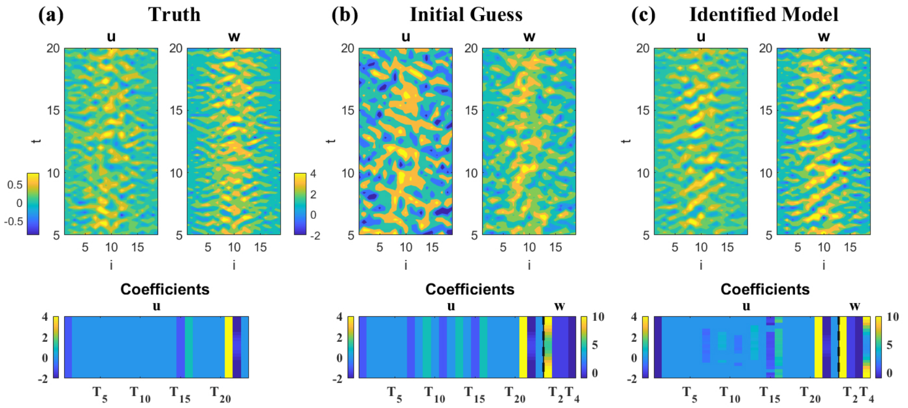
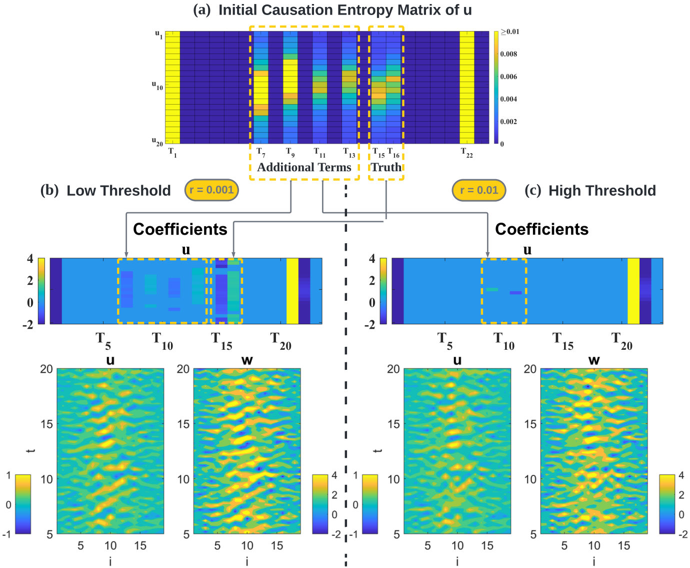


Figure: Identifying the two-layer L-96 model in Regime I. Different columns show the truth, the initial guess, and the identified model.

A High-Dimensional System with Stoch Parameterizations



A High-Dimensional System with Stoch Parameterizations

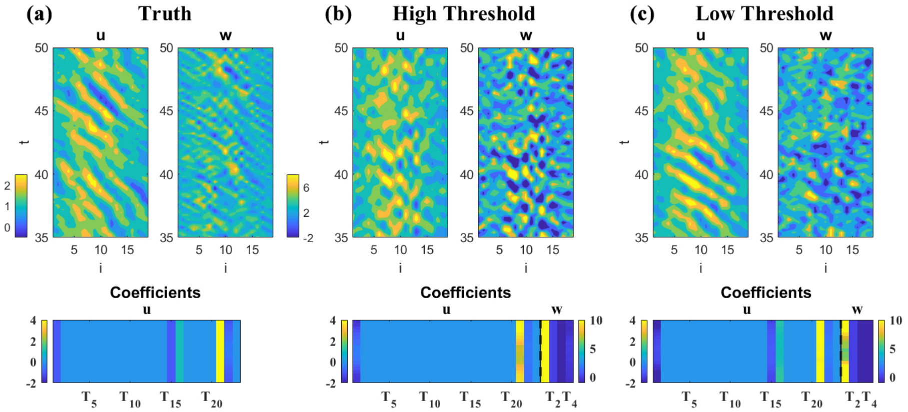


Figure: Similar to Figure 4 but for Regime II. Note that instead of repeating the column for the initial guess, the column for the high threshold case $r = 0.01$ is shown instead.

Content

- Introduction
- Overview of the Causality-Based Learning Algorithm
  - Conditional Sampling
  - Causal Inference
  - Parameter Estimation
- A simple Example for Proof-of-Concept
- A High-Dimensional System with Stochastic Parameterizations
- Summary

Summary

- A causality-based learning algorithm is developed.
- The method exploits the causation entropy to pre-determine the candidate functions.
- The closed analytic formula of conditional sampling is used.
- A quadratic optimization problem is solved for parameter estimation via maximum likelihood estimates
- Physics constraints and localization techniques are further incorporated into the learning algorithm

References

Nan Chen and Yinling Zhang. A causality-based learning approach for discovering the underlying dynamics of complex systems from partial observations with stochastic parameterization. *arXiv preprint arXiv:2208.09104*, 2022.

Q&A

Thank you for listening!

Email: zhang2447@wisc.edu