LA-UR-23-31772
Approved for public release; distribution is unlimited.

Title: Regression-based projection for learning Mori-Zwanzig operators

Author(s): Lin, Yen Ting

Intended for: Materials Research Society (MRS) Spring Meeting and Exhibit, 2023-04-10/2023-10-14 (San Francisco, California, United States)

Issued: 2023-10-16

Regression-based projection for learning Mori–Zwanzig operators

2023 MRS Spring Meeting

4/12, 2023

Yen Ting Lin[1], Yifeng Tian[1], Danny Perez[2], & Daniel Livescu[3]

[1] *Information Sciences Group (CCS-3)*

[2] *Physics and Chemistry of Materials (T-1)*

[3] *Computational Physics and Methods Group (CCS-2)*

Los Alamos National Laboratory (LANL)

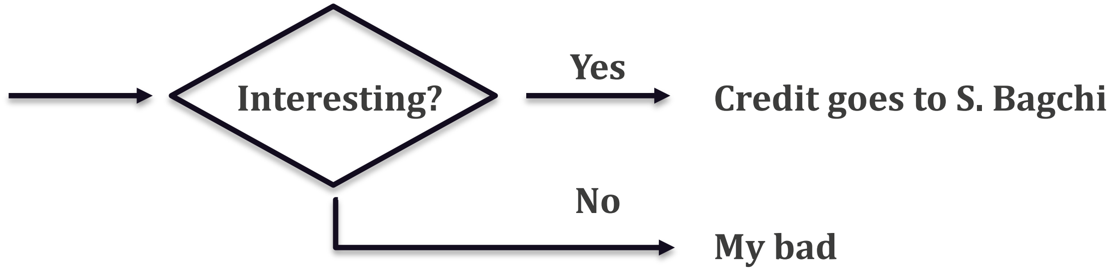
Working example on data-driven closure of dislocation density evolution

Soumendu Bagchi[2] & **YTL**

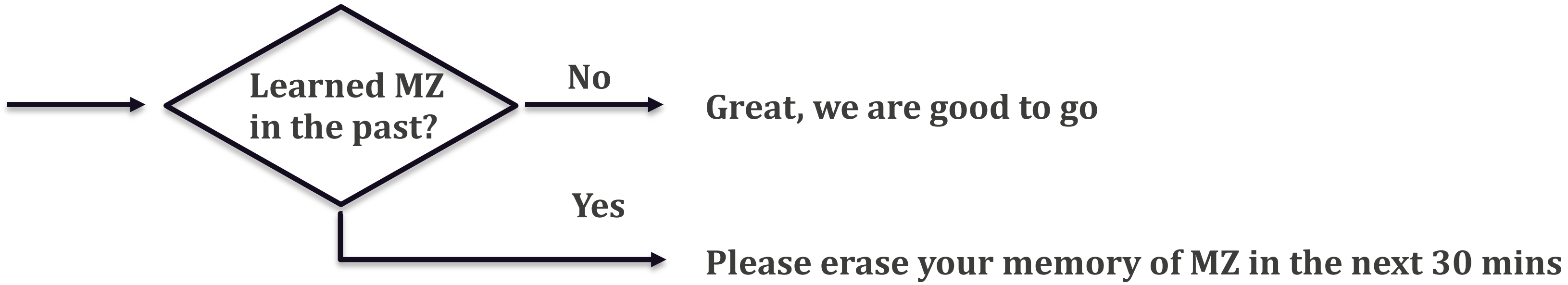
Based on the work arXiv: 2205.05135, V2 coming soon

"Disclaimers"

- **Intention: start an inter-disciplinary dialogue**



- **On Mori-Zwanzig theory/formalism**



Outline

- **Introduction**
  - Mori-Zwanzig (MZ) projection-operator formalism for building reduced-order/coarse-grained dynamical models
- **Regression analysis**
  - Regression is a projection
  - A principled way of extracting MZ operators for regression-based projection operators
- **Numerical experiments**
  - **Work in progress on data-driven closure of dislocation density evolution**: Our attempt to make connection to modeling in materials science
  - **Lorenz ’63 model**: Progressive improvements from the linear Mori’s projector, over nonlinear and spline regression, to neural networks
  - **Kuramoto–Sivashinsky model**
    - Important difference between MZ memory and Delay Embedding
- **Summary**

Introduction to Mori–Zwanzig (MZ) formalism

Context

Non-equilibrium statistical physics, for coarse graining/model reduction/reduced-order modeling

Problem

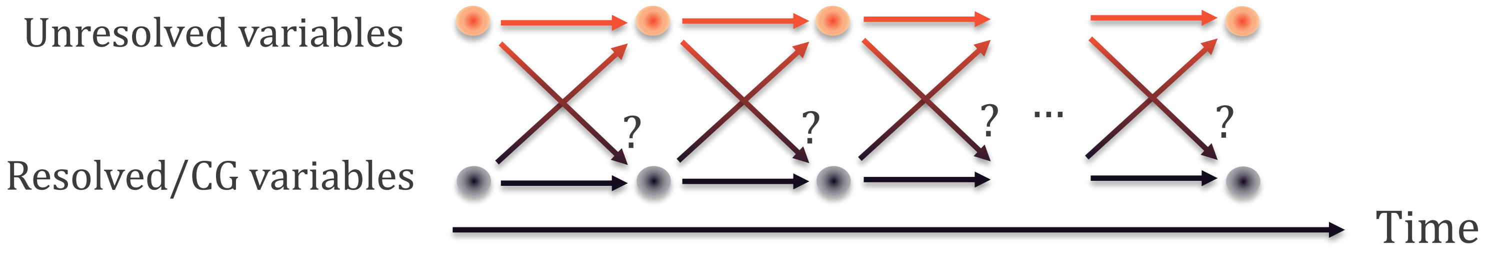
- A dynamical system with a many degrees of freedom D
- One only cares about the evolution of a small set ($M \ll D$) of *resolved/coarse-grained* variables (observables, descriptors, dynamic variables: functions of the system's state)
- Also for partially-observed dynamical system

Example

- Models describing biomolecules with many atoms: the dynamics of the "rough" molecular conformation
- Models describing materials: meso/macroscopic features, e.g. dislocation density
- Complex fluid-dynamical models: the dynamics of the mesoscopic features: e.g., large eddies/coherent structures

Key challenge

How to evolve M variables under the influence of unresolved degrees of freedom?



Projection

Projection operator

Suppose the system's state $\boldsymbol{\phi} \in \mathbb{R}^D$ is fully characterized by

- Resolved/CG observables $\mathbf{g}_r: \mathbb{R}^D \to \mathbb{R}^M$ and
- Unresolved observables $\mathbf{g}_u: \mathbb{R}^D \to \mathbb{R}^{D-M}$

Given an state $\boldsymbol{\phi} \in \mathbb{R}^D$, the CG and unresolved observations are $\mathbf{g}_r(\boldsymbol{\phi})$ and $\mathbf{g}_u(\boldsymbol{\phi})$.

The projection operator $\mathcal{P}$ maps any function f of the resolved *and* under-resolved observations to another function $\mathcal{P}f$ that depends only on the resolved observation:

$$f \overset{\mathcal{P}}{\to} (\mathcal{P}f),$$
$$f\big(\mathbf{g}_r(\boldsymbol{\phi}), \mathbf{g}_u(\boldsymbol{\phi})\big) \approx (\mathcal{P}f)\big(\mathbf{g}_r(\boldsymbol{\phi})\big) \ \forall \boldsymbol{\phi} \in \mathbb{R}^D$$

This allows us to write a *closed reduced-order dynamics* in terms of the resolved observables only.

We will denote $\mathbf{g} = \mathbf{g}_r$ when appropriate.

Projection operators

Existing projection operators include

- Mori's [Mori65, Lin21b] linear functional projection operator: $(\mathcal{P}f)(\mathbf{g}_r) \coloneqq \langle f, \mathbf{g}_r \rangle_\rho \cdot \langle \mathbf{g}_r, \mathbf{g}_r^T \rangle_\rho^{-1} \cdot \mathbf{g}_r$ with inner product space $\langle f, g \rangle_\rho \coloneqq \int_\Omega f(\phi) g(\phi)\, \rho(\phi)\, d\phi$, with a density ρ induced by the dynamics
- Finite-rank projection: orthonormal components of $\mathbf{g}_r$ under the induced density ρ, $(\mathcal{P}f)(\mathbf{g}_r) \coloneqq \langle f, \mathbf{g}_r \rangle_\rho \cdot \mathbf{g}_r$
- Zwanzig's [Zwanzig73] conditional expectation projection:

$$(\mathcal{P}f)(\boldsymbol{h}) = \mathbb{E}_\rho\left[f\left(\mathbf{g}_r(\boldsymbol{\phi}), \mathbf{g}_u(\boldsymbol{\phi})\right) | \mathbf{g}_r(\boldsymbol{\phi}) = \boldsymbol{h}\right] = \int_{\mathbf{g}_r^{-1}(\boldsymbol{h})} f\left(\mathbf{g}_r(\phi), \mathbf{g}_u(\phi)\right) \rho(\phi)\, d\phi$$

- Truncations [Durasaimi,Stinis19,Stinis21]: sending $\mathbf{g}_u(\boldsymbol{\phi}) \to \mathbf{0}$
- Wiener projection [Lin21a]: delay embedding but with infinite delay to augment state space; no MZ memory kernel

Mori's linear $\mathcal{P}$
Computationally OK
but with unsatisfactory
predictions [Lin21b]

→

Zwanzig's nonlinear $\mathcal{P}$
Optimal yet computationally infeasible

Question: can we gradationally fill the gap?

Mori-Zwanzig (MZ) formalism

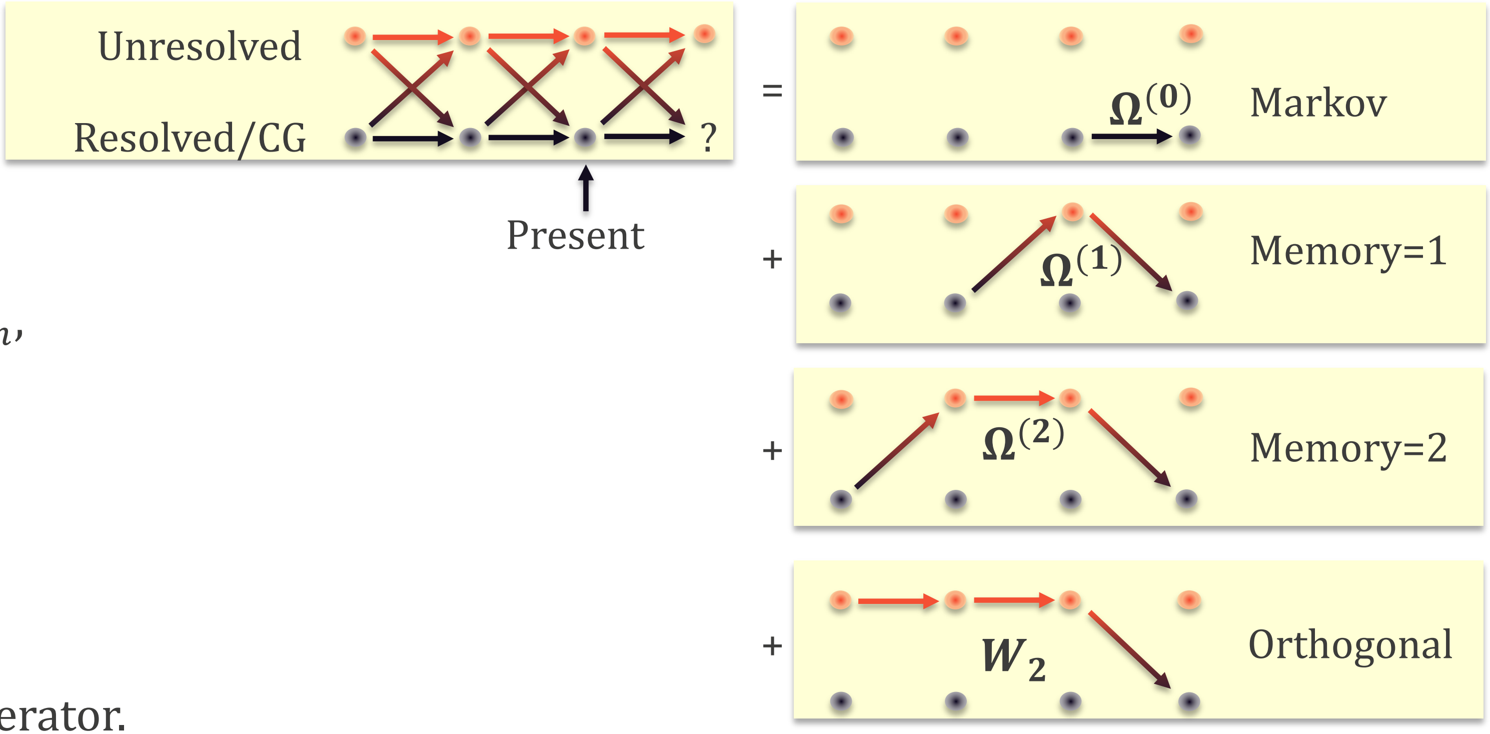
We consider an autonomous and deterministic dynamical system:

$$\dot{\boldsymbol{\phi}} = \boldsymbol{R}(\boldsymbol{\phi}), \qquad \boldsymbol{\phi}(t=0) = \boldsymbol{\phi}_0,$$

where $\boldsymbol{R}: \mathbb{R}^D \to \mathbb{R}^D$ is the locally Lipschitz vector field.

Suppose we always observe at discrete tim

$\boldsymbol{\phi}$



The discrete-time Mori–Zwanzig formalism
the *Generalized Langevin Equation*:

$$\mathbf{g}_{n+1} \triangleq \mathcal{K}_\Delta^n \mathbf{g} = \sum_{\ell=0}^{n} \boldsymbol{\Omega}^{(\ell)} \circ \mathbf{g}_{n-\ell} + \boldsymbol{W}_n,$$

$$\boldsymbol{\Omega}^{(\ell)} \coloneqq \mathcal{P}\mathcal{K}_\Delta[(1-\mathcal{P})\mathcal{K}_\Delta]^n$$

$$\boldsymbol{W}_n \coloneqq [(1-\mathcal{P})\mathcal{K}_\Delta]^{n+1}\mathbf{g},$$

where $\mathcal{K}_\Delta$ is the finite-time (Δ) Koopman operator.

a

Generalized Fluctuation-Dissipation Relation

$$\mathbf{g}_{n+1} \triangleq \mathcal{K}_\Delta^n \mathbf{g} = \sum_{\ell=0}^{n} \boldsymbol{\Omega}^{(\ell)} \circ \mathbf{g}_{n-\ell} + \boldsymbol{W}_n \quad \text{(Generalized Langevin Equation)}$$

$$\boldsymbol{\Omega}^{(\ell)} \coloneqq \mathcal{P}\mathcal{K}_\Delta[(1-\mathcal{P})\mathcal{K}_\Delta]^n \quad (\ell = 0\text{: Markov, } \ell > 0\text{: memory kernel})$$

$$\boldsymbol{W}_n \coloneqq \boldsymbol{W}_n \coloneqq [(1-\mathcal{P})\mathcal{K}_\Delta]^{n+1}\mathbf{g} \quad (\textit{orthogonal dynamics}, \mathcal{P}\boldsymbol{W}_n = 0)$$

Importantly, the operators are related by the ***Generalized Fluctuation-Dissipation Relation***:

$$\boldsymbol{\Omega}^{(\boldsymbol{n})} = \mathcal{P}(\boldsymbol{W}_{n-1} \circ \boldsymbol{F}), \qquad n \geq 1$$

which relates the nth *memory kernel* to the $(n-1)$th orthogonal dynamics.

It is challenging to compute $\boldsymbol{\Omega}^{(\boldsymbol{n})}$ and $\boldsymbol{W}_n$ analytically.

➢ **Research question**: Can we learn the *operators* $\boldsymbol{\Omega}^{(n)}$ and *observables* $\mathbf{W}_n$ from snapshots (time series) of $\mathbf{g}_n(\phi_0)$ out of exact simulations of the full system, with sufficiently many samples of ϕ_0?

Regression as a projection operator

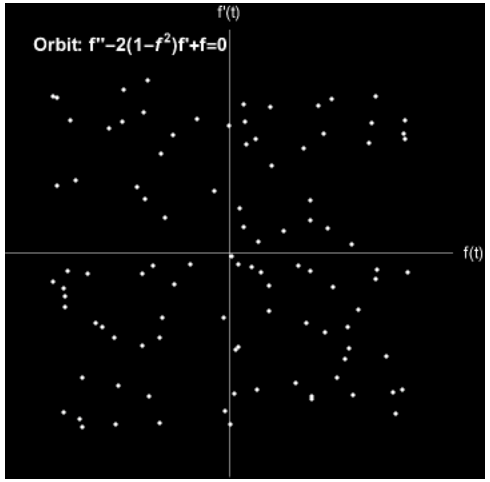
We propose to use statistical regression as a projection operator

$f_{\boldsymbol{\theta}}: \mathbb{R}^M \to \mathbb{R}$ a family of functions parametrized by $\boldsymbol{\theta}$ to approximate $f\left(\mathbf{g}_r(\boldsymbol{\phi}^{[i]}), \mathbf{g}_u(\boldsymbol{\phi}^{[i]})\right)$

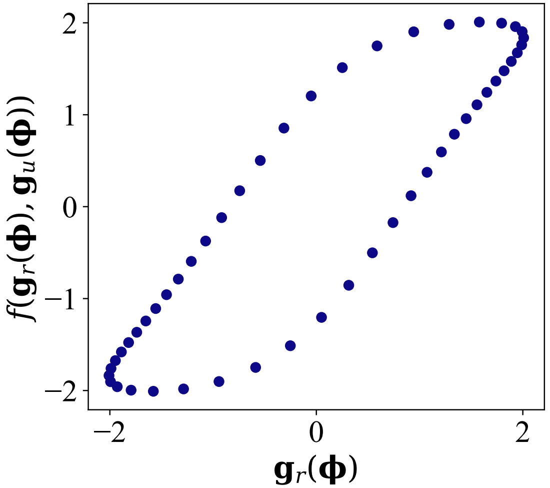
Cost/Risk/loss/Negative log-likelihood $C\left(\boldsymbol{\theta}; \text{observed data} = \left\{\mathbf{g}_r(\boldsymbol{\phi}^{[i]}),\ f\left(\mathbf{g}_r(\boldsymbol{\phi}^{[i]}), \mathbf{g}_u(\boldsymbol{\phi}^{[i]})\right)\right\}_i\right)$

Best-fit parameter: $\boldsymbol{\theta}_* = \operatorname{argmin}_\theta C(\boldsymbol{\theta}; \text{observed data})$

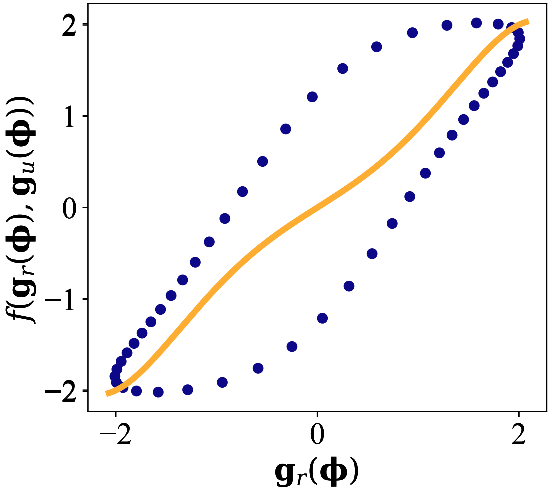
In dynamics, **f** is just resolved part of the dynamics in the future!



Van der Pol oscillator



$f \xrightarrow{\mathcal{P}=\text{regression}} f_{\boldsymbol{\theta}_*}$



Advantages:

1. → Zwanzig's
2. NN-based ML
3. Modeling

Learning the memory kernels and orthogonal dynamics

Generalized Langevin Equation (**GLE**): $\boldsymbol{W}_n \equiv \; \mathbf{g}_{n+1} - \sum_{\ell=0}^{n} \boldsymbol{\Omega}^{(\ell)}(\mathbf{g}_{n-\ell})$

Key idea: $\boldsymbol{W}_n$ is the *residual* of the regression model, which can be computed

Generalized Fluctuation-Dissipation relation (**GFD**): $\boldsymbol{\Omega}^{(\boldsymbol{n})} = \mathcal{P}(\boldsymbol{W}_{n-1} \circ \boldsymbol{F})$

$\boldsymbol{\Omega}^{(0)} \coloneqq \mathcal{P}\mathcal{K}_\Delta$ is just a regression of $\mathbf{g}_1$ on $\mathbf{g}_0$: Many talks are about this Markov operator!

We can learn $\boldsymbol{\Omega}^{(n+1)}$ and $\boldsymbol{W}_n$ if $\boldsymbol{\Omega}^{(0)}$,… $\boldsymbol{\Omega}^{(n)}$ and $\mathbf{g}_{n+1}$ are given.

Operationally an intuitive iterative procedure (statistical boosting) :

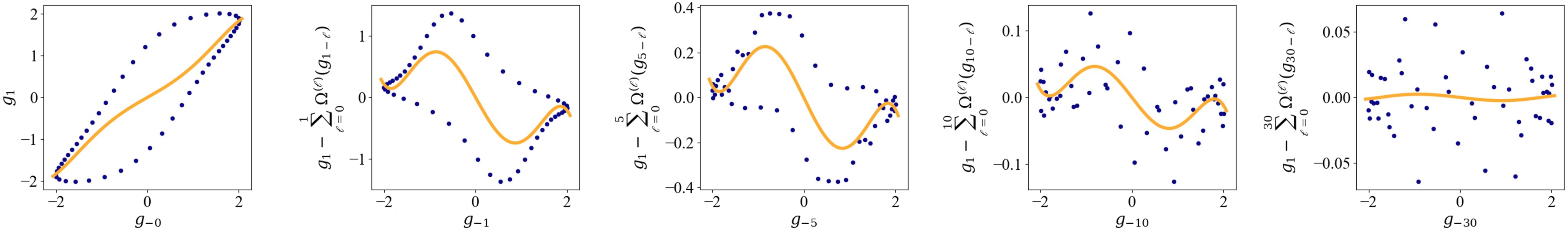
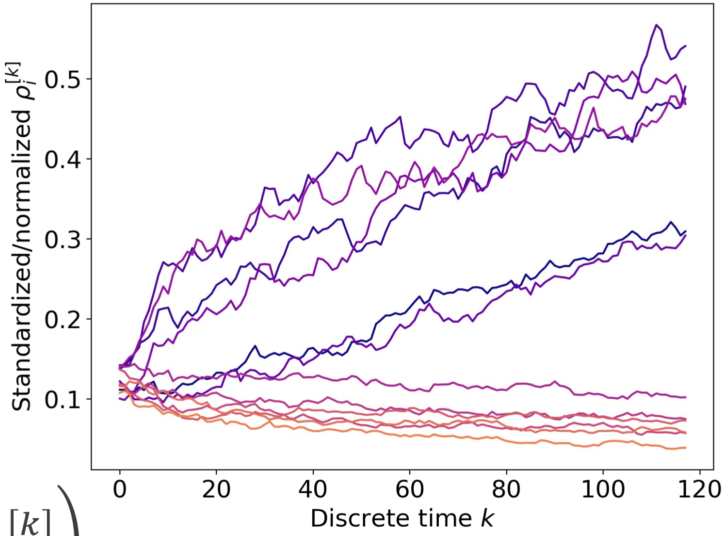


Illustration: data-driven closure of dislocation density



Modeling dislocation density: Akhondzadeh et al. *J. Mech. Phys. Solids* (2020)

Fully resolved: Discrete Dislocation Dynamics simulation
Target: predicting future $\rho_i(t)$, with slip system index $i = 1 \dots 12$

1. CG variables: Dislocation density ρ_i and strain γ_i
2. "Regression-based projection operator"
   a) Model: Kocks–Mecking model structure $\dot{\rho}_i = \dot{\gamma}_i \left(\sqrt{\alpha_{ij}\rho_j} - \beta\rho_i \right)$

      ➡️ Modification 1: discretization of time: $\hat{\rho}_i^{[k+1]} = \rho_i^{[k+1]} + \dot{\gamma}_i^{[k]} \left(\sqrt{\alpha_{ij}\rho_j^{[k]}} - \beta\rho_i^{[k]} \right)$

      ➡️ Modification 2: use delay-embedding to estimate $\dot{\gamma}_i^{[k]} \approx \sum_{\ell=1}^{3} \theta_\ell \, \gamma_i^{[k-\ell]}$

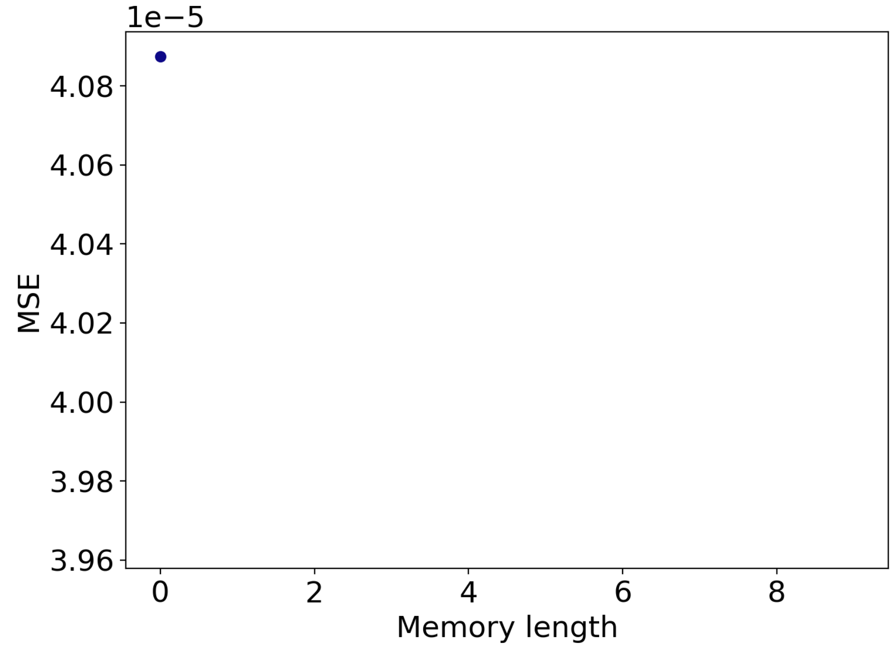
$$\hat{\rho}_i^{[k+1]} = \theta_0 \, \rho_i^{[k]} + \left(\sum_{\ell=1}^{3} \theta_\ell \gamma_i^{[k-\ell]} \right) \left(\sqrt{\alpha_{ij}\rho_j^{[k]}} - \theta_4 \rho_i^{[k]} \right)$$

   b) Risk function: mean square error of *one-step* prediction.

$$C\left(\theta; \rho_i^{[k+1]}, \rho_i^{[k]}, \gamma_i^{[k-2]} \dots \gamma_i^{[k]}\right) \propto \sum_{k,i} \left\| \rho_i^{[k+1]} - \hat{\rho}_i^{[k+1]} \right\|_2^2$$

Illustration: data-driven closure of dislocation density

$$\hat{\rho}_i^{[k+1]} = 1.01\rho_i^{[k]} + \left(13.34\gamma_i^{[k]} + 0.01\gamma_i^{[k-1]} - 13.45\gamma_i^{[k-2]}\right)\left(\sqrt{\alpha_{ij}\rho_i^{[k]}} - 2.23\rho_i^{[k]}\right)$$



Clarification

Our proposition:

1. [Crucial decision!] Define a set of resolved/CG variables
2. [Crucial decision!] Define a projection operator: a model and a regression-based parametrization scheme
3. Use ***Generalized Fluctuation Dissipation*** to recursively ***extract*** the memory kernels and the orthogonal dynamics (The kernels and the orthogonal dynamics depends on the choices of CG variables and projection operators)
   a. Solve for the the best-fit function
   b. Compute the residual
   c. Assign the residual as the dependent variable and an earlier snapshots as the independent variables
   d. Repeat

What our proposition is not:

1. Motivated by Mori-Zwanzig's memory-dependent dynamics
2. Postulate a memory-dependent dynamics (e.g., delay-embedded dynamics; Recurrent Neural Network with Long Short-Term Memory; time-embedded Transformer)
3. Use the data to fit a memory kernel without enforcing or checking ***Generalized Fluctuation Dissipation***

➢ Logical fallacy: MZ is memory dependent, but not all memory-dependent dynamics is MZ.

Numerical experiment 1: Lorenz `63

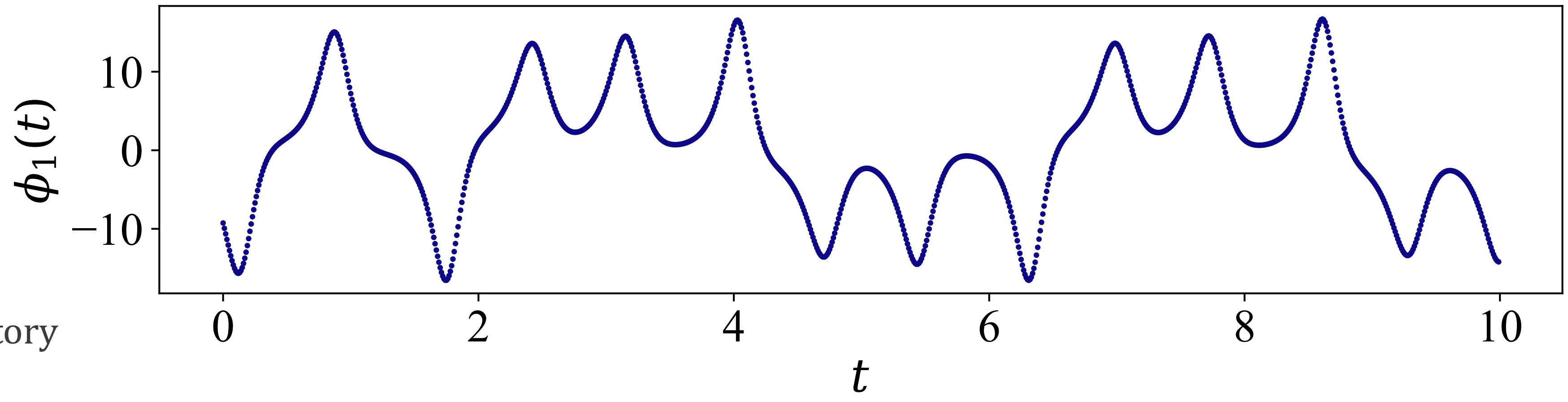
$\dot{\phi_1} = 10\,(\phi_2 - \phi_1)$

$\dot{\phi_2} = \phi_1(28 - \phi_3) - \phi_2$

$\dot{\phi_3} = \phi_1\phi_2 - \frac{8}{3}\phi_3$

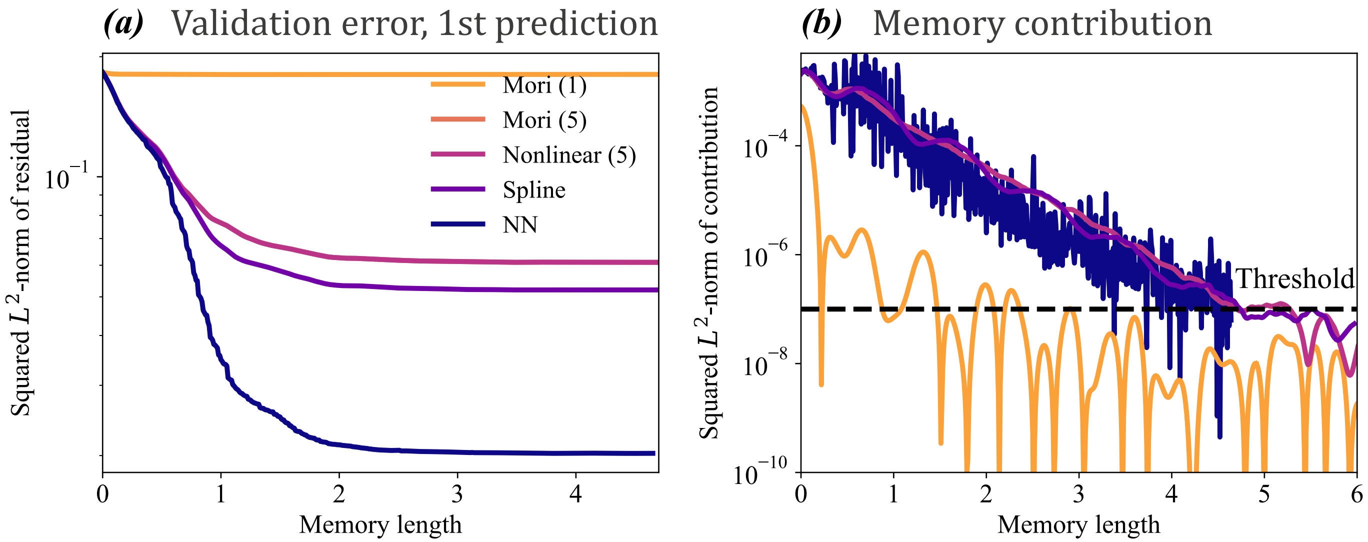
Only $\phi_1(n\Delta)$ is observed, $\Delta = 0.01$

$N = 10^6$ data points along a long trajectory



Regression models:

- Mori (1): linear regression on ϕ_1^0, ϕ_1^1 (DMD)
- Mori (5): linear regression on $\phi_1^{0\ldots5}$ (EDMD)
- Nonlinear: 5^{th} order polynomial regression on ϕ_1
  - ➢ Identical to Mori (5) only in the first step of prediction
- Spline regression
- Fully-connected Feedforward Neural Network



Numerical experiment 1: Lorenz `63

How well do these "trained" model (truncated H) predict?

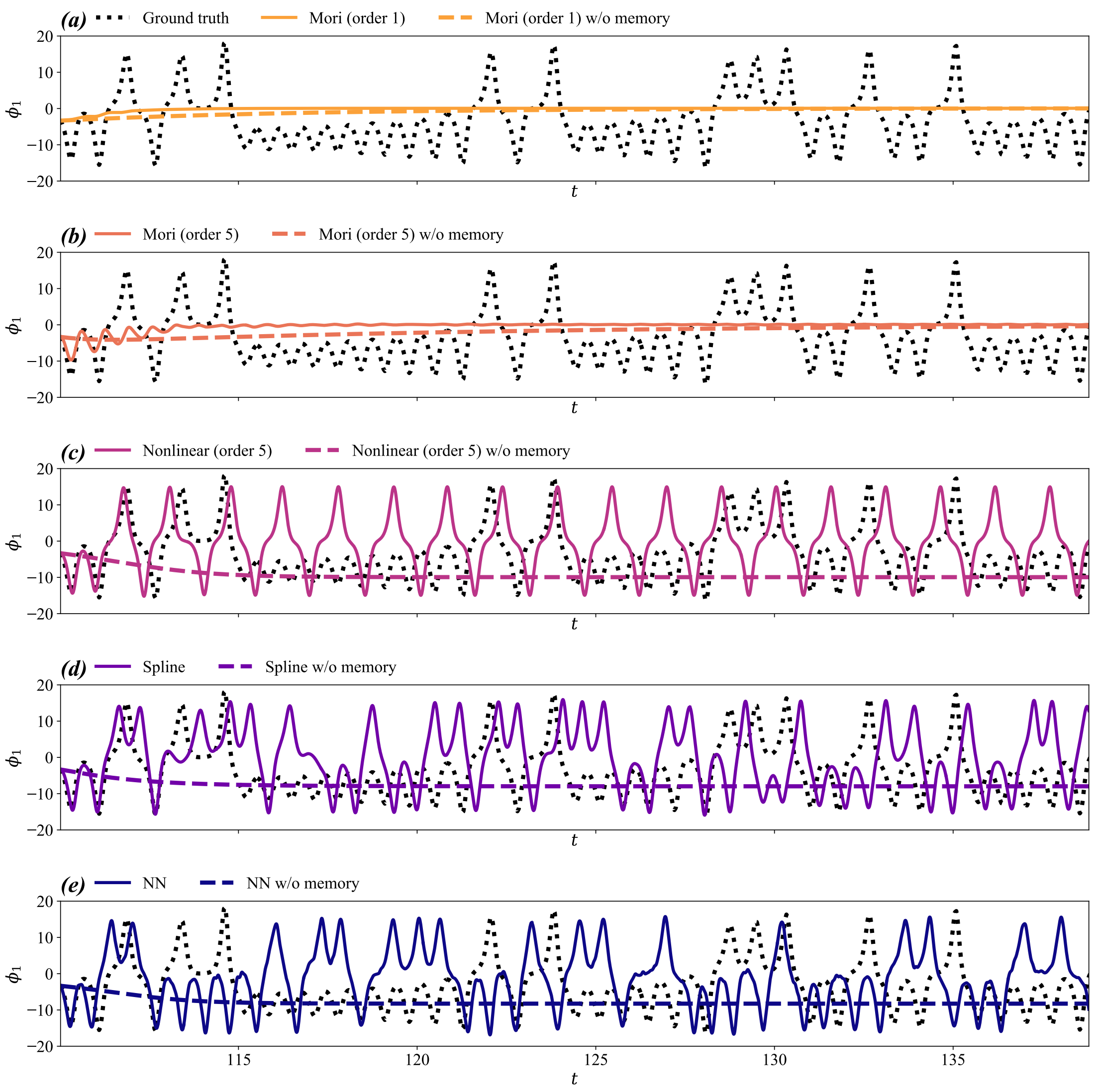
In prediction, we truncated the memory by the threshold and provided a sufficiently long history for the GLE to propagate:

$$\mathbf{g}_1 \leftarrow \sum_{\ell=0}^{H} \boldsymbol{\Omega}^{(\ell)} \circ \mathbf{g}_{-\ell} + \boldsymbol{W}_H$$

The noise is modelled as negligible, $\boldsymbol{W}_H \approx 0$. We iteratively used the predicted observables, e.g.,

$$\mathbf{g}_2 \leftarrow \sum_{\ell=0}^{L} \boldsymbol{\Omega}^{(\ell)} \circ \mathbf{g}_{1-\ell} + \boldsymbol{W}_{H+1}$$

again assuming $\boldsymbol{W}_{H+1} = 0$.



Numerical experiment 2: 1D Kuramoto–Sivashinsky Equation

$$\partial_t u(x,t) + \partial_x[\lambda(x)\partial_x u(x,t)] + \partial_{xxxx} u(x,t) + \frac{1}{2}[\partial_x u(x,t)]^2$$

$x \in [0, 16\pi]$, periodic boundary condition.

Ground truth: spatially discretized as 128 points

Integration step $\delta = 0.001$, observe every 1,000 steps ($\Delta = 1$)

Integrator: Exponential Time-Derivative 4th order Runge–Kutta

Reduced-order observables: observation **every 4 points**

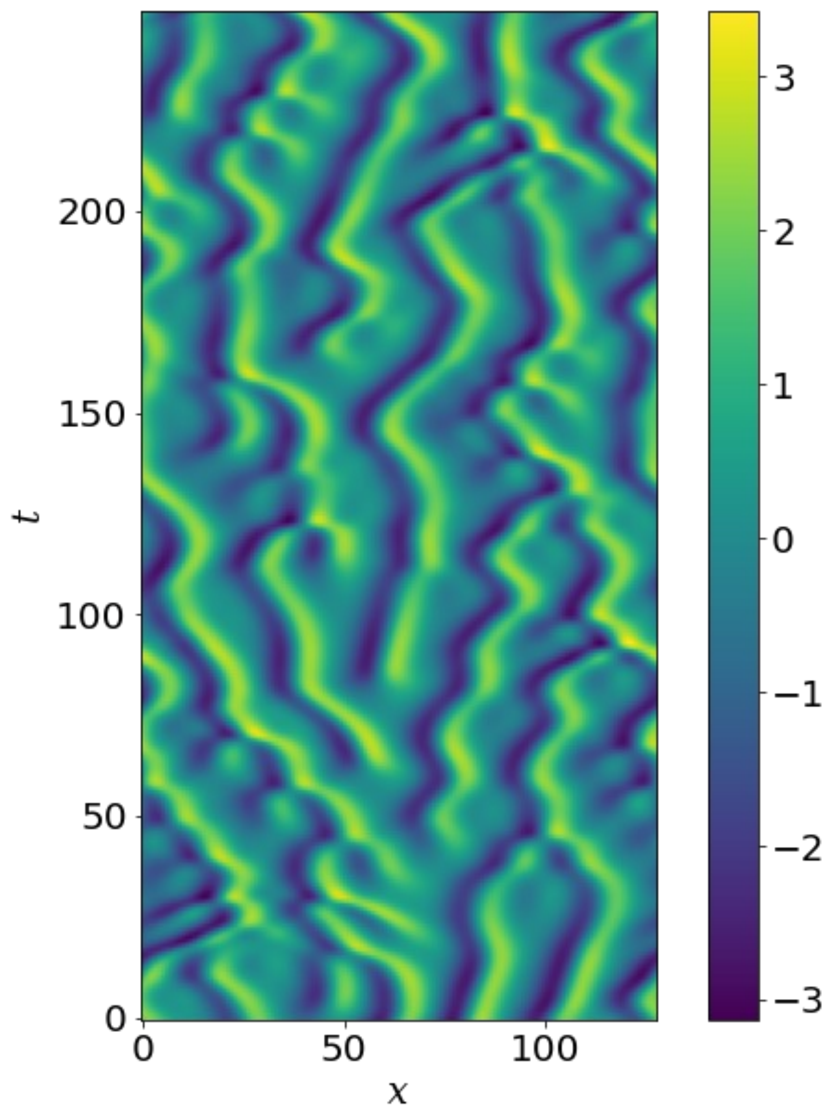
Total number of observations: 10^5

Data-augmentation by translational symmetry + PBC

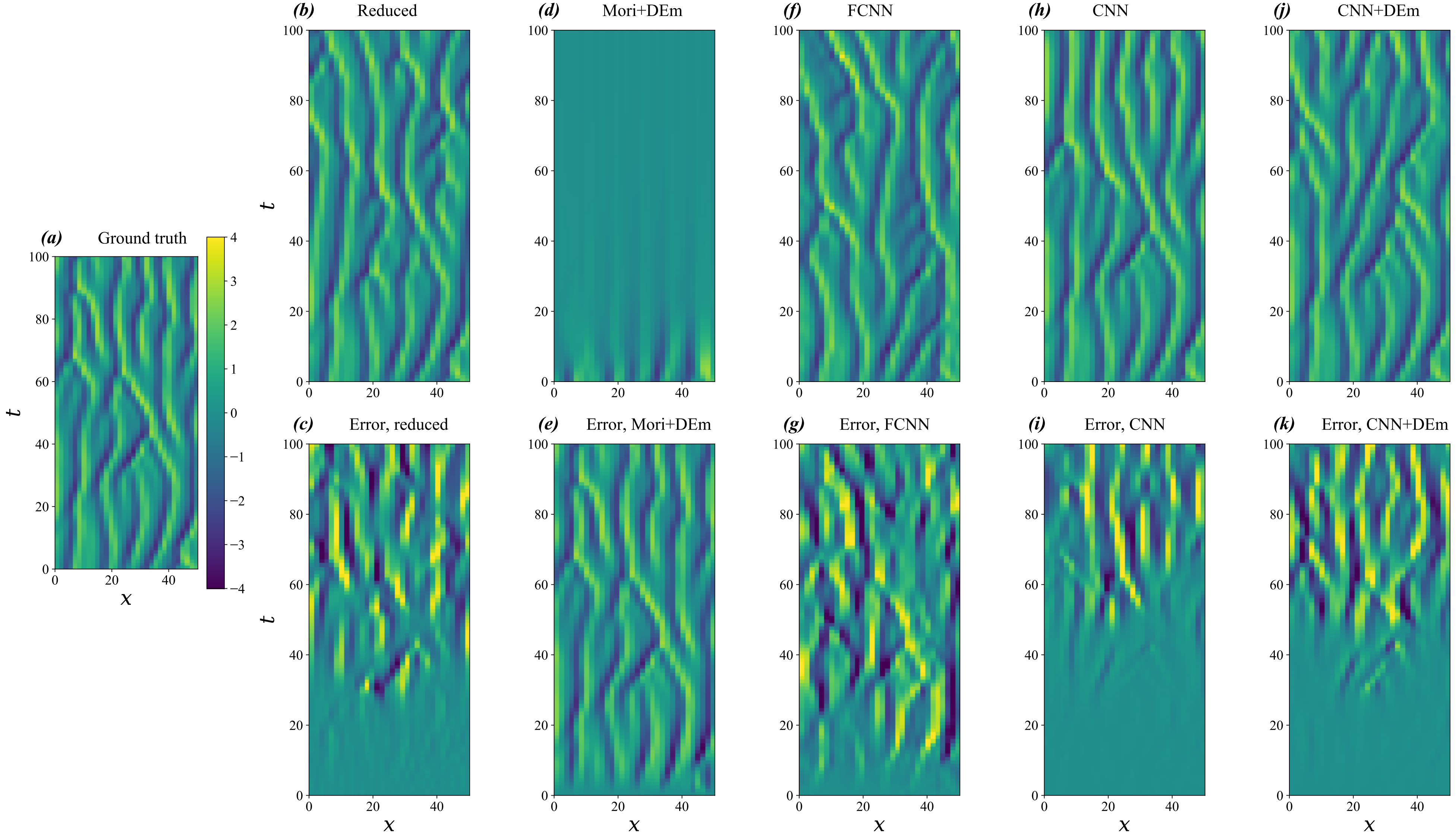
Baseline: Reduced-order simulation (as if a 32-point system)

Regression models:

- Mori+Delay Embedding (DEm) = Hankel DMD [Arbabi17]
- FCNN
- CNN
- CNN+DEm



Numerical experiment 2: 1D Kuramoto–Sivashinsky Equation



Prediction

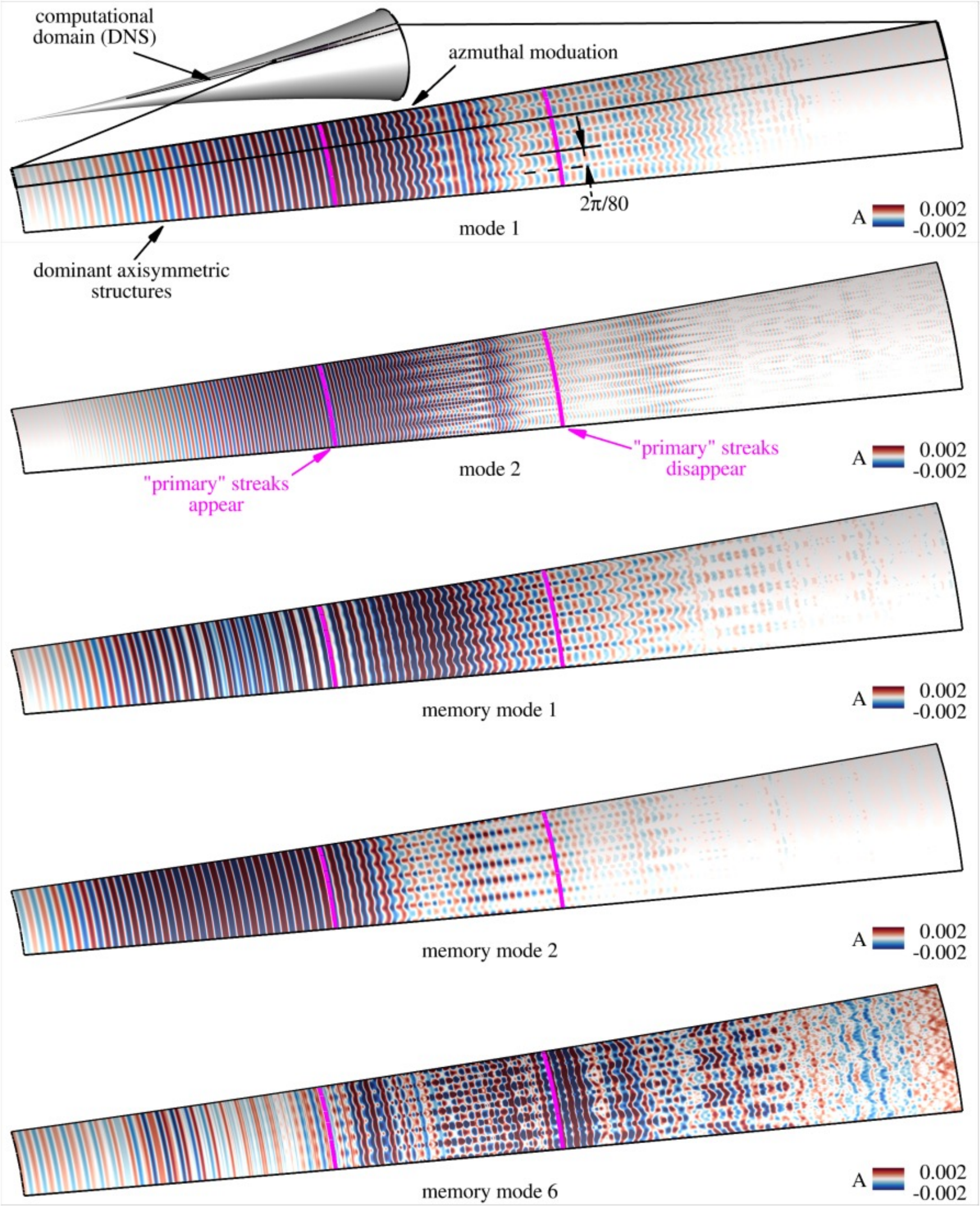
Error:=Prediction− GT

Summary

- **Theoretical contribution:** Regression is a projection operator for learning Mori–Zwanzig formalism
  - Generalize Mori's projection operator, which is already a higher-order generalization of approximate Koopman [Lin21b]
  - Consistent to nonlinear closure schemes, and also other learning frameworks such as SINDy [Brunton16] and Koopman + NN [Li17, Yeung 17, Lusch18]
  - Bridging the gap between Mori's [Mori65] and Zwanzig's [Zwanzig73] projection operators
  - Makes connection to mechanistic models that are parametrized by data

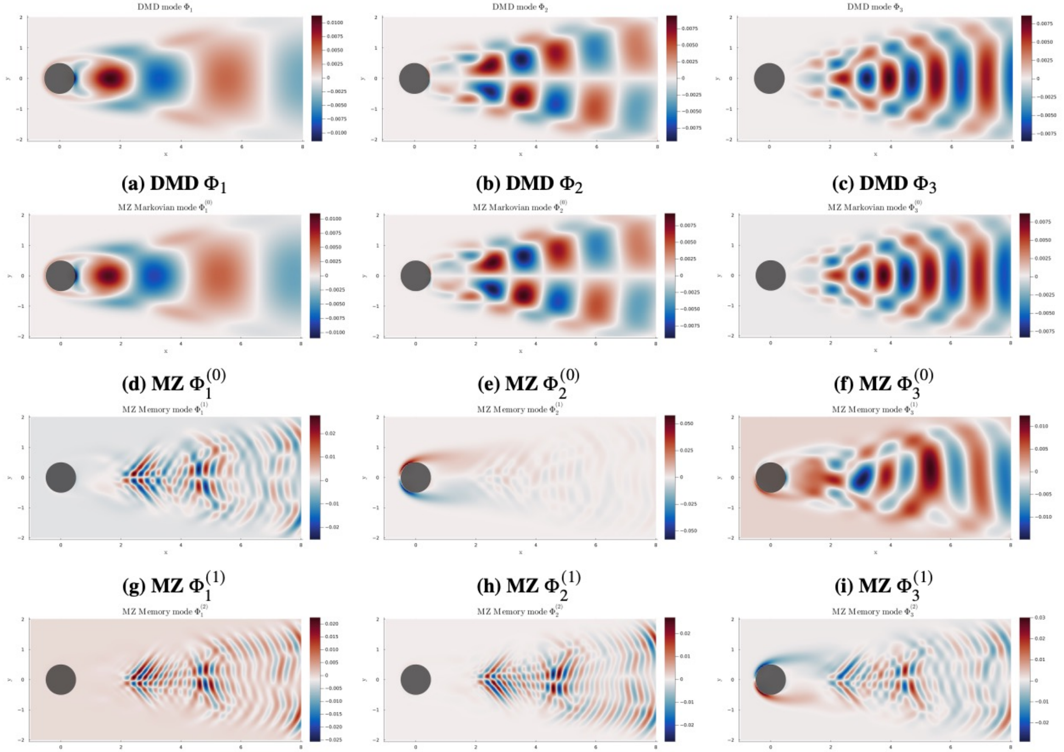
- **Computational Contribution:** A principled way of extracting MZ operators
  - … that are applicable to regression models with adjustable complexities:
    - ➢ Linear regression on nonlinear observables
    - ➢ Nonlinear regression on linear observables
    - ➢ Non-parametric (spline regression)
    - ➢ Neural architectures
  - The reduced-order/coarse-grained model can again be a nonlinear dynamical system (with MZ memory)
  - Finite memory truncation and zero-orthogonal-dynamics seemed to work relatively well

- **Future directions**
  - Applications: isotropic turbulence [Tian21], hypersonic boundary layer transition [Woodward22] and dislocation density evolution
  - Stochastic systems
  - Beyond zero-orthogonal-dynamics model; modeling by correlated noise



Data-Driven Mori-Zwanzig: Approaching a Reduced Order Model for Hypersonic Boundary Layer Transition

Michael Woodward[1,2]*, Yifeng Tian[2], Arvind Mohan[2], Yen Ting Lin[2], Christoph Hader[3], Hermann Fasel[3], Misha Chertkov[1], Daniel Livescu[2]



(a) DMD Φ_1 (b) DMD Φ_2 (c) DMD Φ_3

(d) MZ $\Phi_1^{(0)}$ (e) MZ $\Phi_2^{(0)}$ (f) MZ $\Phi_3^{(0)}$

(g) MZ $\Phi_1^{(1)}$ (h) MZ $\Phi_2^{(1)}$ (i) MZ $\Phi_3^{(1)}$

(j) MZ $\Phi_1^{(2)}$ (k) MZ $\Phi_2^{(2)}$ (l) MZ $\Phi_3^{(2)}$

Selected References

[Mori65] Mori H. *Transport, Collective Motion, and Brownian Motion*. Progress of theoretical physics. 1965

[Zwanzig73] Zwanzig R. *Nonlinear Generalized Langevin Equations*. Journal of Statistical Physics. 1973

[Darve05] Darve E, Solomon J, Kia A. Computing *Generalized Langevin Equations and Generalized Fokker-Planck Equations*. PNAS. 2009

[Schmid10] Schmid PJ. Dynamic Mode Decomposition of Numerical and Experimental Data. Journal of Fluid Mechanics. 2010

[Williams15] Williams MO, Rowley CW, Kevrekidis IG. *A Data–Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition*. Journal of Nonlinear Science. 2015

[Brunton16] Brunton SL, Proctor JL, Kutz JN. *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*. PNAS 2016

[Arbabi17] Arbabi H, Mezi´c I. *Ergodic Theory, Dynamic Mode Decomposition, and Computation of Spectral Properties of the Koopman Operator*. SIAM Journal on Applied Dynamical Systems. 2017

[Li17] Li Q, Dietrich F, Bollt EM, Kevrekidis IG. *Extended Dynamic Mode Decomposition with Dictionary Learning: A Data-Driven Adaptive Spectral Decomposition of the Koopman Operator*. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2017

[Yeung17] Yeung E, Kundu S, Hodas NO. *Learning Deep Neural Network Representations for Koopman Operators of Nonlinear Dynamical Systems*. American Control Conference (ACC). 2019

[Lusch18] Lusch B, Kutz JN, Brunton SL. *Deep Learning for Universal Linear Embeddings of Nonlinear Dynamics*. Nature Communications. 2018

[Lin21a] Lin KK, Lu F. *Data-Driven Model Reduction, Wiener Projections, and the Koopman-Mori-Zwanzig Formalism*. J. Comp. Phys. 2021

[Lin21b] Lin YT, Tian Y, Livescu D, Anghel M. *Data-Driven Learning for the Mori–Zwanzig Formalism: A Generalization of the Koopman Learning Framework*. SIAM Journal on Applied Dynamical Systems. 2021

[Tian21] Tian Y, Lin YT, Anghel M, Livescu D. *Data-Driven Learning of Mori–Zwanzig Operators for Isotropic Turbulence*. Physics of Fluids. 2021