Stochastic ROM closure

Fei Lu

Department of Mathematics, Johns Hopkins

Join work with **Traian Illiescu, Honghu Liu, and Changhong Mou**

07/08, 2022, ARIA@VT

What is the objective of ROM?

supports from JHU, NSF

POD-Galerkin ROM

Data $>>$ POD basis $>>$ Galerkin ROM.

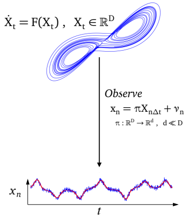
- adaptive/augmented basis
- weighted/scaled function space
- closure: quadratic/DMS

Pro: accurate as FOM data (single trajectory)
Con: Sensitive to data. Generalizability/Robustness/Stability

NN methods: AE, NeuralODE
Pro: flexible: high-D, high nonlinearity
Con: tuning, not using the physics insight

Low-D structure/manifold: interpolation, characteristic, conditional Gaussian

Prediction with Uncertainty Quantification

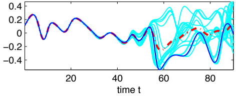


(courtesy of Kevin Lin)

$$x' = F(x) + U(x, y), \qquad \text{resolved scales}$$

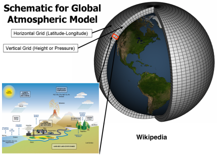
$$y' = G(x, y), \qquad \text{subgrid-scales}$$

Data: $\{x(nh)\}$



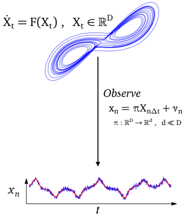
Motivation: Data assimilation:

- ensemble forecasting
- can only afford to resolve $x' = F(x)$



Problem: ensemble prediction of $x(t)$

$$x' = F(x) + U(x,y), \qquad \text{resolved scales}$$
$$y' = G(x,y), \qquad \text{subgrid-scales}$$
$$\text{Data:}\{x(nh)\}$$



courtesy of Kevin Lin

Objective: model the flow map: $x_{1:n-1} \to x_n$

- captures key statistical + dynamical properties
- ensemble simulations (with a larger time-step)

Space-time reduction: spatial dimension ↓; time-step size ↑

Closure modeling, model error UQ, subgrid parametrization

Direct constructions:

- nonlinear Galerkin [Fioas, Jolly, Kevrekidis, Titi...]
- moment closure [Levermore, Morokoff...]
- Mori-Zwanzig formalism
  memory $\rightarrow$ non-Markov process
  [Chorin, Hald, Kupferman, Stinis, Li, Darve, E, Karniadarkis, Venturi, Duraisamy ...]

Data-driven RM

- PCA/POD, DMD, Kooperman [Holmes, Lumley, Marsden, Wilcox, Kutz, Rowley ...]
- ROM closure [Farhat, Carlberg, Iliescu, Wang...]
- stochastic models: SDEs/GLEs, time series models [Chorin/Majda/Gil groups]
- machine learning (...)

- Why and when a data-driven ROM work?
- What does a ROM approximate?

a statistical learning perspective of model reduction

Flow map approximation

$x' = F(x) + U(x,y), \; y' = G(x,y).$
Data $\{x(nh)\}_{n=1}^{N}$

Classical numerical schemes
$$\begin{pmatrix} x_n \\ y_n \end{pmatrix} = \mathbf{F} \begin{pmatrix} x_{n-1} \\ y_{n-1} \end{pmatrix}$$

- trajectory-wise Approx.
- fine resolution
- Closure flow map (Mori-Zwanzig): $x_n = F_n(x_{1:n-1})$

Flow map approximation

$x' = F(x) + U(x,y)$, $y' = G(x,y)$.
Data $\{x(nh)\}_{n=1}^{N}$

Classical numerical schemes
$$\begin{pmatrix} x_n \\ y_n \end{pmatrix} = \mathbf{F} \begin{pmatrix} x_{n-1} \\ y_{n-1} \end{pmatrix}$$

- trajectory-wise Approx.
- fine resolution
- Closure flow map (Mori-Zwanzig): $x_n = F_n(x_{1:n-1})$

Data-driven methods:
$F_n(x_{1:n-1}) \approx \widehat{F}_n(x_{n-p:n-1})$

- average the subgrid-scales
  approximate in distribution
- Learning: curse of dimensionality
  - machine learning: great success
  - parametric inference
    use the structure of the map

NARMA: a numerical time series model

$$(X_n - X_{n-1})/h = R_h(X_{n-1}) + \sum_i c_i \phi_i(X_{n-p:n-1}, \xi_{n-p:n-1}) + \xi_i$$

NARMA(p,q) [Chorin-Lu (15)]

$$(X_n - X_{n-1})/h = R_h(X_{n-1}) + Z_n,$$
$$Z_n = \Phi_n + \xi_n,$$
$$\Phi_n = \underbrace{\sum_{j=1}^{p} a_j X_{n-j} + \sum_{j=1}^{r}\sum_{i=1}^{s} b_{i,j} P_i(X_{n-j})}_{\text{Auto-Regression}} + \underbrace{\sum_{j=1}^{q} c_j \xi_{n-j}}_{\text{Moving Average}}$$

- $R_h(X_{n-1})$ from a numerical scheme for $x' \approx F(x)$
- Φ_n depends on the past
- NARMAX in system identification $Z_n = \Phi(Z, X) + \xi_n,$

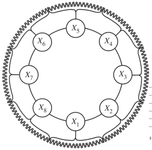
Tasks:

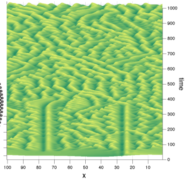
Structure derivation: terms and orders (p, r, s, q) in Φ_n;

Parameter estimation: $a_j, b_{i,j}, c_j$, and σ. Conditional MLE

Chaotic or stochastic systems

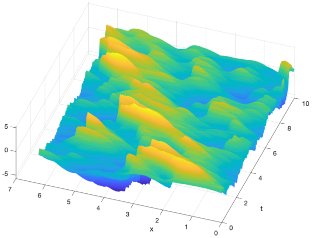
- the two-layer Lorenz96 [Chorin-Lu15]
- Kuramoto-Sivashisky [Lu-Lin-Chorin17]
- stochastic Burgers [Lu20]





The NARMA model can

- tolerate large time-steps
- reproduces statistics: ACF, PDF
- improves Data Assimilation [Lu-Tu-Chorin17]
- predict shock trace [Chen-Liu-Lu22]



1. Motivation and objective

2. Inference-based Model reduction

3. Stochastic ROM closure

POD + Quadratic closure

Randomly parametrized Eq: $\dot{\mathbf{u}} = \mathbf{f}(\mathbf{u})$, $\Rightarrow$

Multiple trajectory Data: $\{\mathbf{u}(t)^{(m)}, t \in [0, T]\}_{m=1}^{M}$

POD basis functions $\{\phi_1, \ldots, \phi_r\}$

$$\mathbf{u}(t,x) = \left(\sum_{i=1}^{r} + \sum_{i=r+1}^{\infty}\right) a_i(t)\phi_i(x)$$

Galerkin projection with closure, $\mathbf{a} = (a_1, \ldots, a_r)$,

$$\begin{aligned}\dot{\mathbf{a}} &= \boldsymbol{F}(\mathbf{a}) + \text{ Closure}(\mathbf{a}) \\ &= \boldsymbol{F}(\mathbf{a}) + \widetilde{A}\mathbf{a} + \mathbf{a}^{\top}\widetilde{B}\mathbf{a} + error\end{aligned}$$

$$(\widehat{A}, \widehat{B}) = \underset{(\widetilde{A},\widetilde{B})}{\arg\min}\ \frac{1}{MT}\sum_{m=1}^{M} \|\dot{\mathbf{a}}^{(m)} - \boldsymbol{F}(\mathbf{a}^{(m)}) - \widetilde{A}\mathbf{a}^{(m)} - (\mathbf{a}^{(m)})^{\top}\widetilde{B}\mathbf{a}^{(m)}\|^2_{L^2([0,T])}$$

Stochastic ROM closure (bad practice)

$$\dot{\mathbf{a}} = \boldsymbol{F}(\mathbf{a}) + \widetilde{A}\mathbf{a} + \mathbf{a}^\top \widetilde{B}\mathbf{a} + \text{~~error~~}\Sigma \dot{W}$$

Maximizing the likelihood:

$$(\widehat{A}, \widehat{B}) = \underset{(\widetilde{A},\widetilde{B})}{\arg\min} \frac{1}{MT} \sum_{m=1}^{M} \|\dot{\mathbf{a}}^{(m)} - \boldsymbol{F}(\mathbf{a}^{(m)}) - \widetilde{A}\mathbf{a}^{(m)} - (\mathbf{a}^{(m)})^\top \widetilde{B}\mathbf{a}^{(m)}\|^2_{L^2([0,T])}$$

Quiz: what wrong with the following procedue?

1. Estimate using FD for $\dot{\mathbf{a}}$ (Euler-Maruyama),

$$\mathbf{a}(t_{l+1}) - \mathbf{a}(t_l) \approx \left[\boldsymbol{F}(\mathbf{a}) + \widetilde{A}\mathbf{a} + \mathbf{a}^\top \widetilde{B}\mathbf{a}\right](t_l)\delta + \sqrt{\delta}\Sigma\boldsymbol{\xi}_l, \quad t_l = l\delta$$

2. Another scheme (e.g. stochastic RK4) for integration

Long-term blow up!

$1.001^{1e5} = 2e+43$

Stochastic ROM closure

Discrete-time stochastic model:

$$\mathbf{a}(t_{l+1}) - \mathbf{a}(t_l) \approx \left[\boldsymbol{F}(\mathbf{a}) + \widetilde{A}\mathbf{a} + \mathbf{a}^\top \widetilde{B}\mathbf{a}\right](t_l)\delta + \sqrt{\delta}\Sigma\boldsymbol{\xi}_l, \quad t_l = l\delta$$

$$(\widehat{A}, \widehat{B}) = \underset{(\widetilde{A},\widetilde{B})}{\arg\min} \frac{1}{ML} \sum_{m=1}^{M} \sum_{l=1}^{L} \| \frac{\Delta \mathbf{a}}{\delta}(t_l) - \boldsymbol{F}(\mathbf{a}^{(m)}) - \widetilde{A}\mathbf{a}^{(m)} - (\mathbf{a}^{(m)})^\top \widetilde{B}\mathbf{a}^{(m)} \|^2$$

- Faithful modeling: no additional discretization error.
  Estimation error 0.0001 $\Rightarrow 1.0001^{1e5} = 2e4$
- Large time-stepping: $\delta = 100\Delta t \Rightarrow 1.0001^{1e3} = 1.1$

Convergence of POD basis and estimators

Under minor conditions on boundedness of $\mathbf{u}$:
Theorem 1: The POD basis converges in $L^2(D)$ at rate $M^{-1/2}$.

Theorem 2: The estimators converges at rate $M^{-1/2}$.

- Not requiring stationary distribution/equilibrium.
- Invertibility of the normal matrix in LS: symmetry

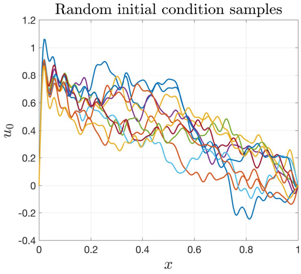
Viscous Burgers with random IC

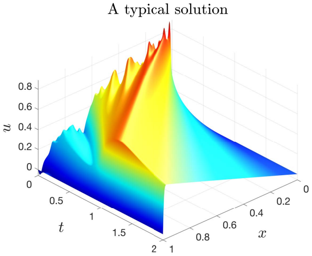
1D Burgers $\nu = 0.002$.

$$u_t = \nu u_{xx} - u u_x\,, \quad 0 < x < 1,\ t > 0,$$
$$u(0,t) = u(1,t) = 0, \quad t \geq 0,$$
$$u(\cdot,0) = u_0(\cdot,\omega) \sim \mu.$$

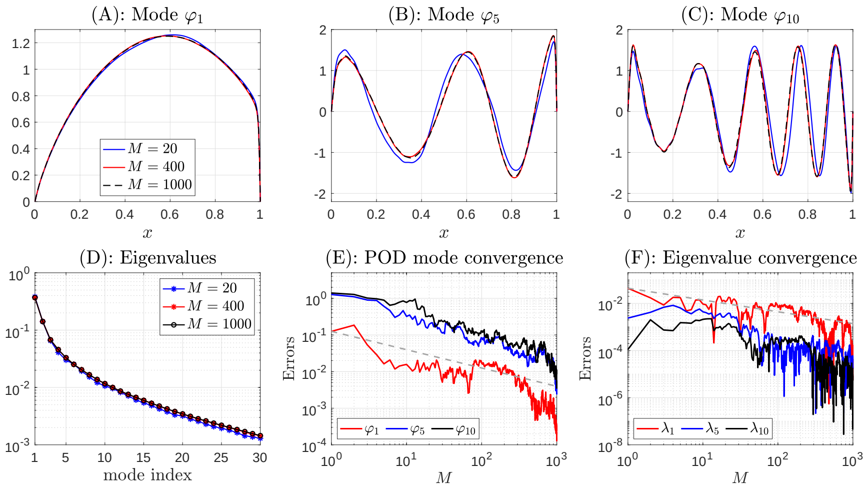
$$u_0(x,\omega) = \sum_{k=1}^{K} \frac{w_k(\omega)}{k} \sin(\pi k x),$$

$K = 50$, $w_k \sim N(0.5, 0.2)$.
$\Delta = 5e-3$, $T = 2$.



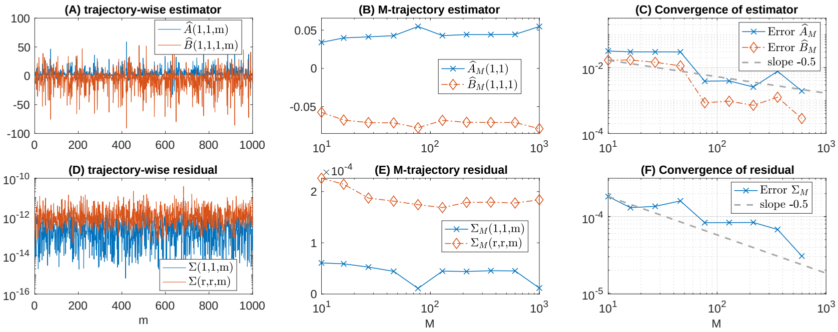


Convergence of POD basis



Convergence of Estimators

ROM with $r = 10$

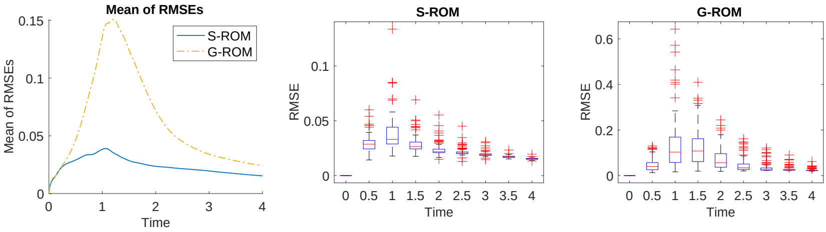


Tikhonov regularization with L-curve.

- single trajectory (400 snapshots): overfitting
- multiple trajectory: convergent

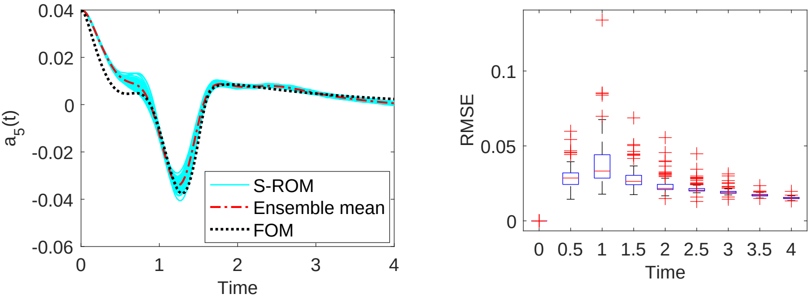
Prediction by ROM

Trajectory-wise prediction (100 new ICs, turn noise-off):



Prediction by ROM

Ensemble prediction (100 realizations):



Open question: Optimal space-time reduction?

Space-time reduction

- dimension reduction r
- large time-stepping $\delta = Gap * \Delta t$

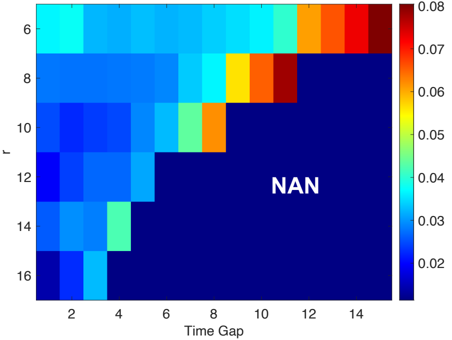
The POD+Quadratic closure:

- 90 ROMs (r, δ)
- RMSE on of 100 trajs on $[0, 4]$

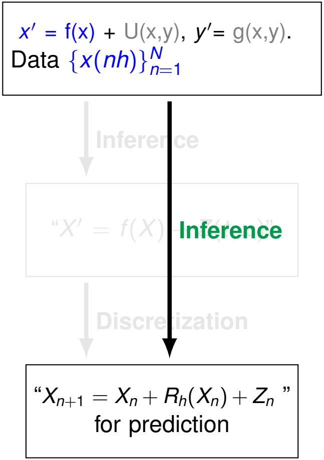
Observations:

- As $r \uparrow$: accuracy $\uparrow$, tolerate $\delta \downarrow$
- Each r: "sweet spot" medium δ

Trade-off (r, δ) for an "optimal" ROM?



Summary



Numerical + inferential model reduction

- non-intrusive time series (**NARMA**)

$\to$ ROM closure with space-time reduction
$\to$ Efficient prediction with UQ

What is the objective of ROM?

This talk: Efficient prediction with uncertainty quantification

- Focus on QoI (dimension reduction)
- Large time-stepping (time reduction)
- Prediction (new random ICs)

- Accurate as FOM
- DiffEq for ROM
- Single-trajectory

Discrete-time stochastic ROM closure

- flow map approximation
  $x_n = F_n(x_{[0,t_{n-1}]}) \approx \widehat{F}_n(x_{1:n-1}) = \sum_k c_k \Phi^k_{n-p:n-1}$
- physical insight + statistical/machine learning

References

- Data-driven stochastic model reduction
  - Chorin-Lu: Discrete approach to stochastic parametrization and dimension reduction in nonlinear dynamics. PNAS 112 (2015).
  - Lu-Lin-Chorin: Comparison of continuous and discrete-time data-based modeling for hypoelliptic systems. CAMCoS, 11 (2016).
  - Lu-Lin-Chorin: Data-based stochastic model reduction for the Kuramoto – Sivashinsky equation. Physica D, 340 (2017).
  - Lin-Lu: Data-driven model reduction, Wiener projections, and the Mori-Zwanzig formalism. JCP (2021).
  - Lu: Data-driven model reduction for stochastic Burgers equations. Entropy 2020.
  - Li-Lu-Ye: ISALT: Inference-based schemes adaptive to large time-stepping for locally Lipschitz ergodic systems, DCDS-S (2021).
- Data assimilation
  - Lu-Tu-Chorin: Accounting for model error from unresolved scales in EnKFs: improving the forecast model. MWR, 340 (2017).
  - Nan Chen, Honghu Liu and F. Lu. Shock trace prediction by reduced models for a viscous stochastic Burgers equation. Chaos (2022).

Thank you!