A statistical learning perspective of data-driven model reduction

Fei Lu

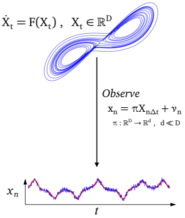
Department of Mathematics, Johns Hopkins

06/06, 2022, CAS-AMSS

supports from JHU, NSF

Prediction with Uncertainty Quantification

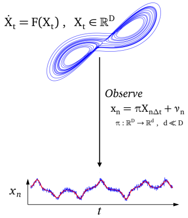
$x' = F(x) + U(x, y)$,	resolved scales
$y' = G(x, y)$,	subgrid-scales
Data: $\{x(nh)\}$	partial observation



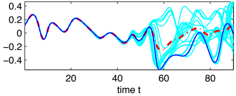
(courtesy of Kevin Lin)

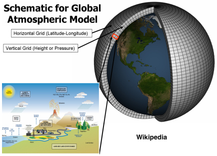
Prediction with Uncertainty Quantification

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



(courtesy of Kevin Lin)



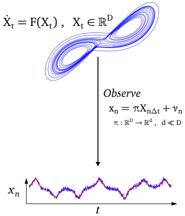


Motivation: Data assimilation:

- 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 large time-step)

<u>Space-time reduction</u>: spatial dimension $\downarrow$; time-step size $\uparrow$

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

Data-driven Model reduction

Computational reduction:

- space: dimension reduction
- time: large time stepping
- space-time

Data (time series):

- full observation: dominating coordinates/basis
- partial observation

Goal: time series model for quantities of interest

1. Motivation and objective
2. Inference-based Model reduction
3. From nonlinear Galerkin to inference

Flow map approximation

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

Two Examples of **Flow map:** $x_{1:n-1} \to x_n$

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

Two Examples of **Flow map:** $x_{1:n-1} \to x_n$

Example 1 (deterministic):

$$x' = \lambda x; \quad \Rightarrow \quad x(h) = x(0)e^{\lambda h}, \quad \forall h > 0$$

Numerical (Euler):

$$x_n = x_{n-1} + h\lambda x_{n-1}, \qquad \text{stability: } |1 + h\lambda| < 1$$

Data $\{x(h), x(2h)\}$, infer $x_n = \theta x_{n-1}$:

$$\theta = x(h)^{-1} x(2h) = e^{\lambda h} \quad \Rightarrow \quad x_n = e^{\lambda h} x_{n-1}, \quad \forall h > 0$$

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

Flow map: $x_{1:n-1} \to x_n$

Example 2 (stochastic, Ornstein-Uhlenbeck):

$$dx_t = \lambda x_t dt + dW_t; \quad \Rightarrow \quad x(h) = x(0)e^{\lambda h} + \int_0^h e^{\lambda s} dW_s$$

Numerical solution (Euler-Maruyama)

$$x_n = x_{n-1} + h\lambda x_{n-1} + N(0,h), \qquad \text{stability: } |1 + h\lambda| < 1$$

Data $\{x(nh)\}_n$, infer $x_n = \theta x_{n-1} + N(0,\sigma)$:

$$\theta = \mathbb{E}[x(h)^2]^{-1}\mathbb{E}[x(2h)x(h)] = e^{\lambda h}$$
$$\sigma = \mathbb{E}[|x_n - \theta x_{n-1}|^2] = (1 - e^{2\lambda h})/(2\lambda)$$
$$\Rightarrow \quad x_n = e^{\lambda h} x_{n-1} + N(0,\sigma), \quad \forall h > 0$$

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:

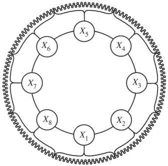
Structure derivation: terms and orders (p, r, s, q) in Φ_n;

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

Example: the two-layer Lorenz 96 model

A NARMA model for the *X* variables

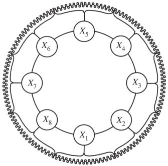
- no scale-separation
- Ansatz: polynomials with time lag 2



Example: the two-layer Lorenz 96 model

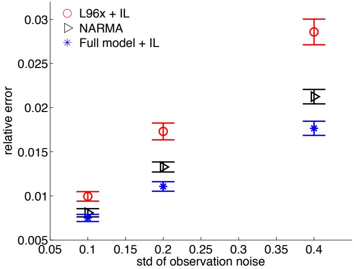
A NARMA model for the X variables

- no scale-separation
- Ansatz: polynomials with time lag 2



The NARMA model can

- tolerate large time-step
- reproduces statistics: ACF, PDF
  [Chorin-Lu15]
- improves Data Assimilation [Lu-Tu-Chorin17]



- Kuramoto-Sivashinsky: $v_t = -v_{xx} - \nu v_{xxxx} - v v_x$
- Burgers: $v_t = \nu v_{xx} - v v_x + f(x,t)$,

Goal: a closed model for $(\widehat{v}_{1:K})$, $K << N$.

$$\frac{d}{dt}\widehat{v}_k = -q_k^\nu \widehat{v}_k + \frac{ik}{2} \sum_{|l|\le K, |k-l|\le K} \widehat{v}_l \widehat{v}_{k-l} + \widehat{f}_k(t),$$
$$+ \frac{ik}{2} \sum_{|l|>K \text{ or } |k-l|>K} \widehat{v}_l \widehat{v}_{k-l}$$

View $(\widehat{v}_{1:K}) \sim x$, $(\widehat{v}_{k>K}) \sim y$:

$$x' = F(x) + U(x,y), \quad y' = G(x,y).$$

TODO: represent the effects of high modes to the low modes

Derivation of a parametric form (KSE): $v_t = -v_{xx} - \nu v_{xxxx} - vv_x$

Let $v = u + w$. In operator form: $v_t = Av + B(v)$,

$$\frac{du}{dt} = PAu + PB(u) + [PB(u+w) - PB(u)]$$
$$\frac{dw}{dt} = QAw + QB(u+w)$$

Nonlinear Galerkin: approximate inertial manifold (IM)[1]

- $\frac{dw}{dt} \approx 0 \Rightarrow w \approx A^{-1}QB(u+w) \Rightarrow w \approx \psi(u)$
- Need: spectral gap condition ✓;
- $dim(u) >> K$ $(u \leftrightarrow \widehat{v}_{1:K})$:

[1] Foias, Constantin, Temam, Sell, Jolly, Kevrekidis, Titi et al (88-94)

Derivation of a parametric form (KSE): $v_t = -v_{xx} - \nu v_{xxxx} - v v_x$

Let $v = u + w$. In operator form: $v_t = Av + B(v)$,

$$\frac{du}{dt} = PAu + PB(u) + [PB(u+w) - PB(u)]$$
$$\frac{dw}{dt} = QAw + QB(u+w)$$

Nonlinear Galerkin: approximate inertial manifold (IM)[1]

- $\frac{dw}{dt} \approx 0 \Rightarrow w \approx A^{-1}QB(u+w) \Rightarrow w \approx \psi(u)$
- Need: spectral gap condition ✓;
- $dim(u) >> K$ ($u \leftrightarrow \widehat{v}_{1:K}$): parametrization with time delay (Lu-Lin17)

A time series (NARMA) model of the form

$$u_k^n = R^\delta(u_k^{n-1}) + \Phi_k^n + g_k^n,$$

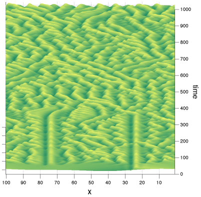
KEY: high-modes = functions of low modes

[1] Foias, Constantin, Temam, Sell, Jolly, Kevrekidis, Titi et al (88-94)

Kuramoto-Sivashinsky Equation

Test setting: $\nu = 3.43$
$N = 128$, $dt = 0.001$
Reduced model: $K = 5, \delta = 100dt$

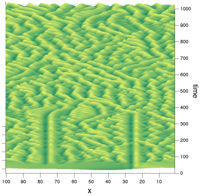
- 3 unstable modes
- 2 stable modes



Kuramoto-Sivashinsky Equation

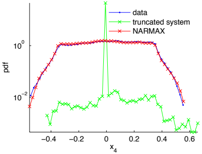
Test setting: $\nu = 3.43$
$N = 128$, $dt = 0.001$
Reduced model: $K = 5, \delta = 100dt$

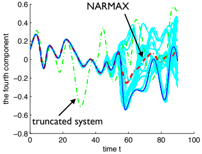
- 3 unstable modes
- 2 stable modes



Long-term statistics:

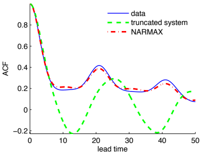
- reproduce PDF /ACF

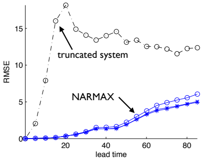




Prediction: Forecast time:

- truncated sys.: $T \approx 5$
- NARMA: $T \approx$ **50**
  ($\approx$ 2 Lyapunov time)





Derivation of parametric form: stochastic Burgers

$$v_t = \nu v_{xx} - v v_x + f(x,t)$$

Let $v = u + w$. In operator form:

$$\frac{du}{dt} = PAu + PB(u) + Pf + [PB(u+w) - PB(u)]$$
$$\frac{dw}{dt} = QAw + QB(u+w) + Qf$$

- no spectral gap
  $w(t)$ is not function of $u(t)$, but a functional of its path

Derivation of parametric form: stochastic Burgers

$$v_t = \nu v_{xx} - v v_x + f(x,t)$$

Let $v = u + w$. In operator form:

$$\frac{du}{dt} = PAu + PB(u) + Pf + [PB(u+w) - PB(u)]$$
$$\frac{dw}{dt} = QAw + QB(u+w) + Qf$$

- no spectral gap
  $w(t)$ is not function of $u(t)$, but a functional of its path

Integration instead:

$$w(t) = e^{-QAt} w(0) + \int_0^t e^{-QA(t-s)} [QB(u(s) + w(s))] ds$$
$$w^n \approx c_0 QB(u^n) + c_1 QB(u^{n-1}) + \cdots + c_p QB(u^{n-p})$$

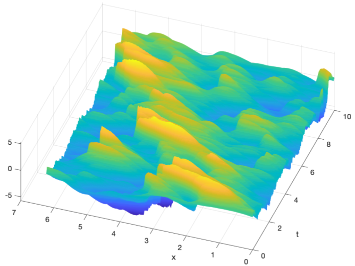
Linear in parameter: $PB(u+w) - PB(u) \approx \sum_{j=0}^{p} c_j P[(u^n QB(u^{n-j}))_x] + noise$

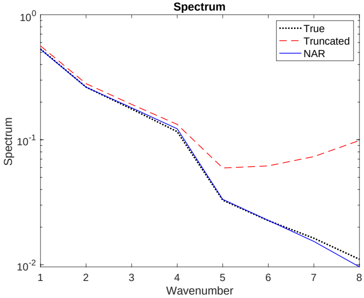
$$u_k^n = R^\delta(u_k^{n-1}) + f_k^n + g_k^n + \Phi_k^n,$$

KEY: high-modes = functionals of paths of low modes

Numerical tests:

$\nu = 0.05$, $K_0 = 4 \rightarrow$ random shocks



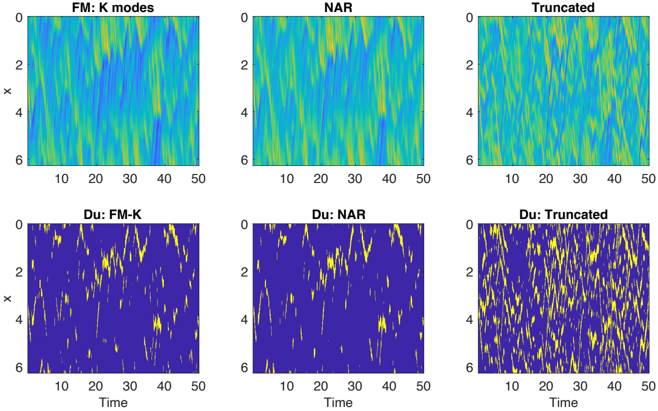


- Full model: $N = 128, dt = 0.005$
- Reduced model: $K = 8$, $\delta = 20dt$

Energy spectrum

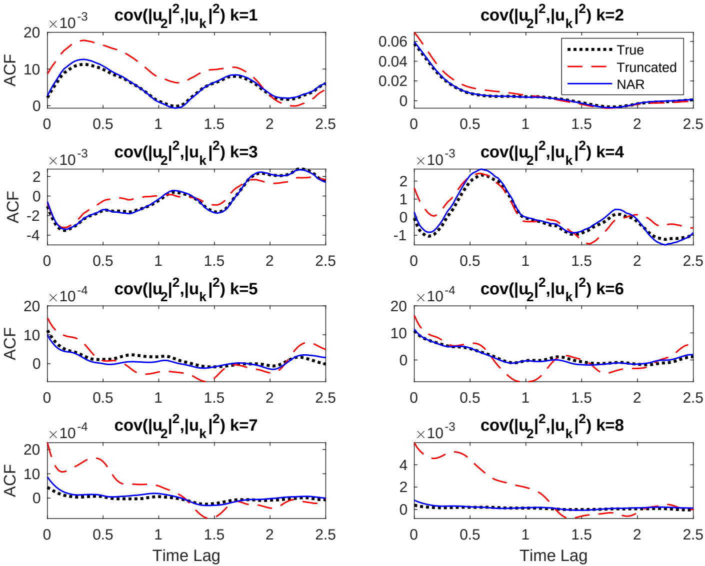
- ▲ Temporal correlation ✓
- ▲ Trajectory prediction ✓
- ▲ Shock trace prediction ✓

Shock trace prediction:



Binary shock trace based on a threshold for u_x

Stochastic Burgers equation



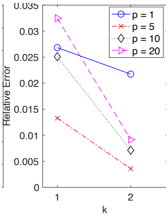
Cross-ACF of energy (4th moments!)

Open questions in space-time reduction

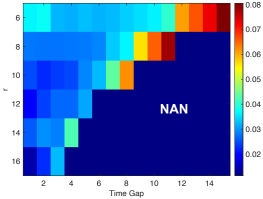
$$(X_n - X_{n-1})/h = R_h(X_{n-1}) + \sum_i c_i \phi_i(x_{n-p:n-1}, \xi_{n-p:n-1}) + \xi_i$$

Observed from numerical tests:

- Memory length: best at medium
- Space reduction: arbitrary $K = 2$

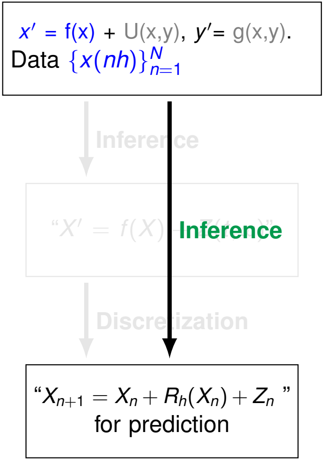


- Time reduction: stability
  h limited by R_h: medium
  CFL (truncated-G) = CFL(FM).



Optimal space-time reduction?

Summary



Numerical + inferential model reduction

- non-intrusive time series (**NARMA**)
- flow map approximation

$$x_n = F_n(x_{[0,t_{n-1}]})$$
$$\approx \widehat{F}_n(x_{1:n-1}) = \sum_k c_k \Phi^k_{n-p:n-1}$$

$\rightarrow$ space-time reduction

Data-driven modeling of dynamics

- Large time stepping for stiff ODEs/SDEs:
  - Approx. the discrete-time flow map
  - Parametric inference: improves but limited (Li-Lu-Ye21)
    - Dependent on the parametric form
    - Nyström: $(0.50, 0.40)$, not the Störmer-Verlet $(0.5, 0.5)$
  - Machine learning: promising
- Space-time reduction for PDEs/SPDEs
  - Data-based coordinates
  - Optimal space-time reduction
  - Optimal memory length

Data-driven modeling of dynamics

- Large time stepping for stiff ODEs/SDEs:
  - Approx. the discrete-time flow map
  - Parametric inference: improves but limited (Li-Lu-Ye21)
    - Dependent on the parametric form
    - Nyström: $(0.50, 0.40)$, not the Störmer-Verlet $(0.5, 0.5)$
  - Machine learning: promising
- Space-time reduction for PDEs/SPDEs
  - Data-based coordinates
  - Optimal space-time reduction
  - Optimal memory length

Probabilistic/statistical numerical integrators adaptive to

- time-step
- space-basis
- parameter distribution

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!