Stochastic model reduction of nonlinear dynamics by inference

Fei Lu

Department of Mathematics, Johns Hopkins

Joint work with: Alexandre J. Chorin (UC Berkeley), Xuemin Tu (KU)
Kevin K. Lin (U. of Arizona) Xingjie Li (UNC) and Felix Ye (JHU)

March 31, 2021
Mathematics Department, U. of Kansas

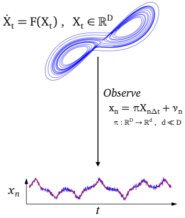
FL acknowledges supports from JHU, LBL, NSF.

1. Motivation and objective
   - Problem and motivation
   - Example
   - Review
2. Inference-based Model reduction
   - Flow map approximation
   - NARMA: a numerical time series model
   - Example: a chaotic system
3. From nonlinear Galerkin to inference
   - Kuramoto-Sivashinsky Equation
   - Stochastic Burgers equation
   - Optimal space-time reduction

Problem: ensemble prediction of $x(t)$

$$\begin{aligned} x' &= F(x) + U(x,y), && \text{resolved scales} \\ y' &= G(x,y), && \text{subgrid-scales} \end{aligned}$$

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



Motivation:

- arise from numerical ODE/PDE/SDEs
- Data assimilation: partial noisy observation
  - ensemble prediction
  - can only afford to resolve $x' = F(x)$

Objective: a closed numerical model of x that

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

Example

Numerical stochastic Burgers equation

$$v_t = \nu v_{xx} - v v_x + f(x,t), x \in [0, 2\pi], \text{ periodic BC}$$

- Fourier-Galerkin: N Fourier modes
- Need small space-grid & time-step:
  $N \gtrsim 10/\nu$, $dt \sim 1/N$ by (CFL)
  $\rightarrow$ Costly: e.g. $\nu = 10^{-3} \rightarrow N \sim 10^4$, time steps= $10^4 T$

Interested in: efficient predictions of low modes $(\widehat{v}_{1:K})$, $K << N$.

Question: a reduced model of low modes?

Space-time reduction:
Reduce spatial dimension + Increase time-step size

Closure modeling, model error UQ, subgrid parametrization

Direct constructions:

- non-linear 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 (...)

Question:

- when and why data-driven approach work?
- Best space-time time reduction?

This talk: a stochastic modeling perspective — statistical inference

1. Motivation and objective
   - Problem and motivation
   - Example
   - Review
2. Inference-based Model reduction
   - Flow map approximation
   - NARMA: a numerical time series model
   - Example: a chaotic system
3. From nonlinear Galerkin to inference
   - Kuramoto-Sivashinsky Equation
   - Stochastic Burgers equation
   - Optimal space-time reduction

Flow map approximation

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

Classical numerical schemes

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

- trajectory-wise Approx.
- fine resolution (stability and accuracy)
- Closure flow map: $x_n = F_n(x_{1:n-1})$:
  - Taylor expansion no longer work
  - depend on subsgrid scale trajectory

Flow map approximation

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

Classical numerical schemes

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

- trajectory-wise Approx.
- fine resolution (stability and accuracy)
- Closure flow map: $x_n = F_n(x_{1:n-1})$:
  - Taylor expansion no longer work
  - depend on subsgrid scale trajectory

Data-driven methods: approximate the flow map

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

- average the subgrid-scales
  KEY: approximate in distribution
- Learning: curse of dimensionality
  - machine learning: blackbox, many parameters
  - 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)]

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

- $R_h(X_{n-1})$ 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

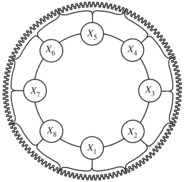
Example: the two-layer Lorenz 96 model

A NARMA model for the X variables

$$\frac{d}{dt}x_k = x_{k-1}(x_{k+1} - x_{k-2}) - x_k + 10 - \frac{1}{J}\sum_j y_{k,j},$$

$$\frac{d}{dt}y_{k,j} = \frac{1}{\varepsilon}[y_{k,j+1}(y_{k,j-1} - y_{k,j+2}) - y_{k,j} + x_k],$$

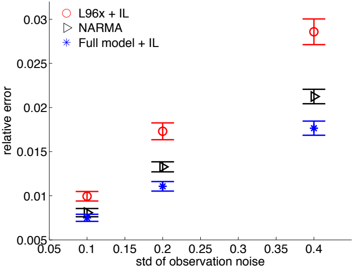
where $x \in \mathbb{R}^{18}$, $y \in \mathbb{R}^{360}$.



- no scale-separation
- Ansatz: polynomial with 2-time lag
- tolerate to large time-step

The NARMA model can

- reproduces statistics: ACF, PDF
  [Chorin-Lu15PNAS]
- improves Data Assimilation
  [Lu-Tu-Chorin17MWR]



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

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): $v_t = -v_{xx} - \nu v_{xxxx} - v v_x$

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 ✓;
- $dim(u) > K$:

[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 ✓;
- $dim(u) > K$: parametrization with time delay (Lu-Lin17)

A time series (NARMA) model of the form

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

with $\Phi_k^n := \Phi_k^n(u^{n-p:n-1}, g^{n-p:n-1})$ in form of

$$\Phi_k^n = \sum_{j=1}^{p} c_{k,j}^v u_k^{n-j} + c_{k,j}^R R^\delta(u_k^{n-j}) + c_{k,j}^w \sum_{\substack{|k-l|\le K, K<|l|\le 2K \\ \text{or } |l|\le K, K<|k-l|\le 2K}} \widetilde{u}_l^{n-1} \widetilde{u}_{k-l}^{n-j}$$

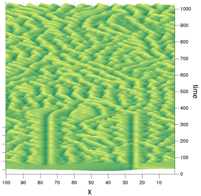
KEY: high-modes = functions of low modes

[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: $K = 5, \delta = 100dt$

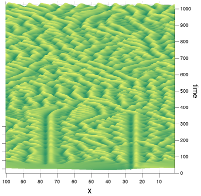
- 3 unstable modes
- 2 stable modes



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

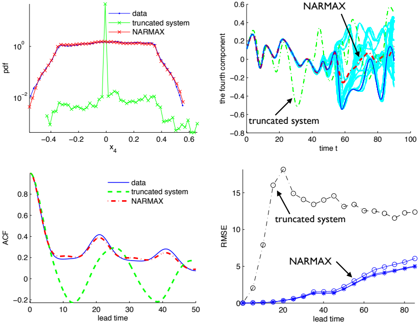


Long-term statistics:

- 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:

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

- spectral gap: Burgers ? (likely not)
  $w(t)$ 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:

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

- spectral gap: Burgers ? (likely not)
  $w(t)$ 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 approximation:

$$PB(u+w) - PB(u) = P[(uw)_x + (u^2)_x]/2 \approx P[(uw)_x]/2 + noise$$
$$\approx \sum_{j=0}^{p} c_j P[(u^n QB(u^{n-j}))_x] + noise$$

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

A time series (NARMA) model of the form

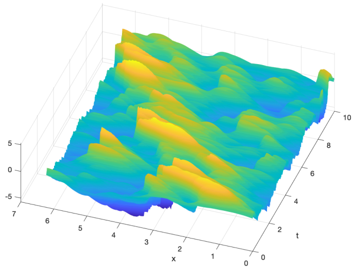
$$u_k^n = R^\delta(u_k^{n-1}) + f_k^n + {\color{red} g_k^n} + \Phi_k^n,$$

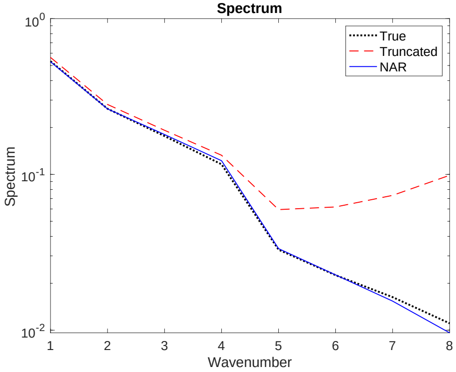
with $\Phi_k^n := \Phi_k^n(u^{n-p:n-1}, f^{n-p:n-1})$ in form of

$$\Phi_k^n = \sum_{j=1}^{p} c_{k,j}^v u_k^{n-j} + c_{k,j}^R R^\delta(u_k^{n-j}) + c_{k,j}^w \sum_{\substack{|k-l|\le K, K<|l|\le 2K \\ \text{or } |l|\le K, K<|k-l|\le 2K}} \widetilde{u}_l^{n-1} \widetilde{u}_{k-l}^{n-j}$$

Numerical tests:

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

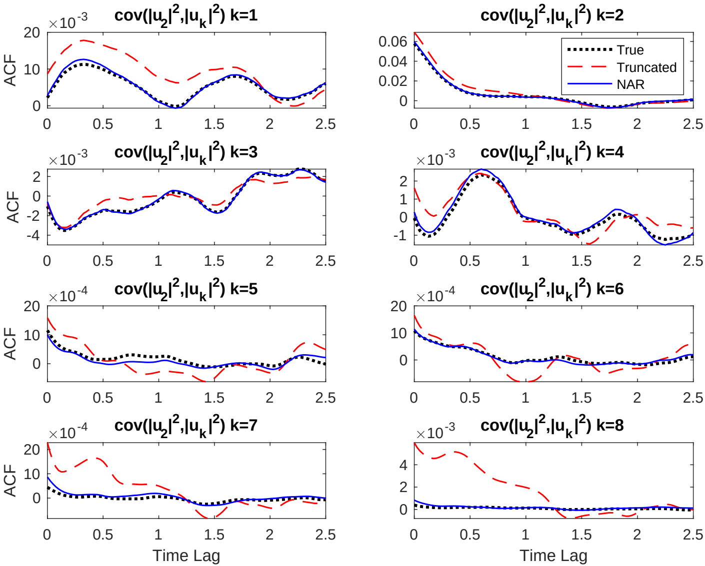




Energy spectrum

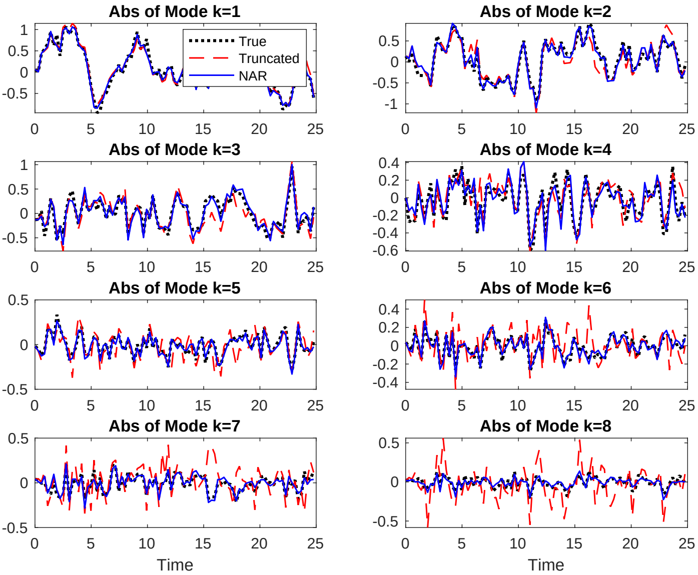
- Full model: $N = 128$, $dt = 0.005$
- Reduced model: $K = 8$, $\delta = 20dt$

Stochastic Burgers equation



Cross-ACF of energy (4th moments!)

Stochastic Burgers equation



Trajectory prediction in response to force

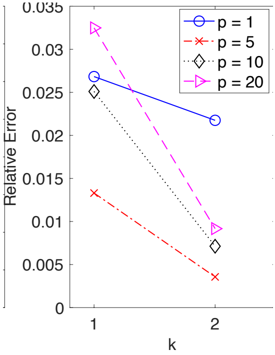
Optimal memory length

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

Best performance at medium memory length?

Relative error of energy spectrum

- first decrease, then increase

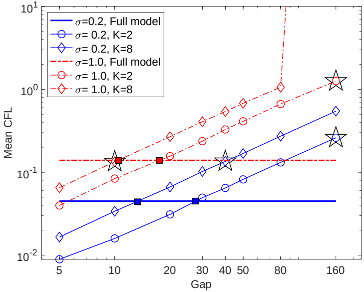


Optimal space-time reduction

- How small can K be? (Space reduction) arbitrary
- How large can δ be? (Time reduction) numerical stability
- What is the optimal space-time reduction ratio?

Best performance when: CFL (truncated Galerkin) $=$ CFL(full model).

- CFL numbers
- NARMA
  - stable up-to large gap
  - best at intersections (squares)



a priori estimate on optimal space-time reduction?

Non-global Lipschitz SDE

Ergodic with non-global Lipschitz drift:

$$dX_t = f(X_t)dt + \sigma dB_t$$

- Explicit scheme: unstable/inaccurate [Mattingly-Stuart-Highm02]
- Implicit scheme: costly (implicit, small Δt)

Infer from data an explicit scheme [Li-Lu-Ye21]

$$(X_n - X_{n-1})/h = \sum_i c_i \phi_i(X_{n-1}, \Delta B_n) + \xi_i$$

- Data from an implicit scheme
- ϕ_i from parametrizing numerical schemes

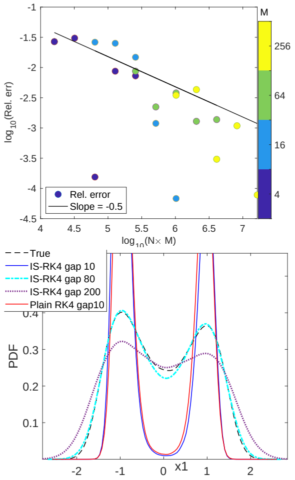
Non-global Lipschitz SDE

Non-global Lipschitz drift:

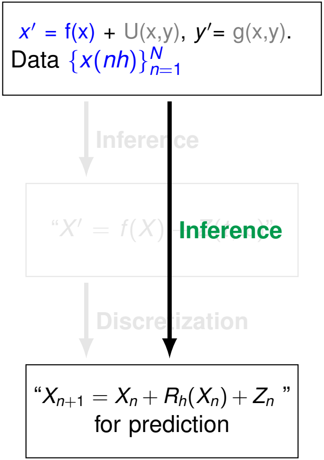
$$dX_t = f(X_t)dt + \sigma dB_t$$

The inferred explicit scheme [Li-Lu-Ye21]

- tolerate large time-step & keep order
- convergent estimators (MLE)
- Insights on optimal time-step
  - medium time-step is the best
  - trade-off: approx. error v.s. sampling /numerical error



Summary



Inference-based stochastic model reduction

- non-intrusive time series (**NARMA**)
- parametrize projections on path space

$$x_n = F_n(x_{1:n-1}) \approx \sum_k c_k \Phi^k_{n-p:n-1}$$

$\rightarrow$ space-time model reduction

Outlook

a bright future for **Numerical + inferential**

- general dissipative systems + model selection
- post-processing to predict shocks
- optimal space-time/time reduction
  - bias-variance tradeoff: “the best in the medium (Zhongyong) ”