Data-driven model reduction for stochastic Burgers equations

Fei Lu

Department of Mathematics, Johns Hopkins

Joint work with: Alexandre J. Chorin (UC Berkeley)
Kevin K. Lin (U. of Arizona)

2nd Symposium on Machine Learning and Dynamical Systems
September, 2020

Consider a stochastic Burgers equation

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

N-mode Fourier-Galerkin: $k = 1, \ldots, N$

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

Need: $N \gtrsim 1/\nu$, $dt \sim 1/N$ by (CFL)

$\to$ Costly: $\nu = 10^{-4} \to N \sim 10^4$, time steps= $10^4 T$

To simulate 10^4 time units, we need 10^8 time steps!

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

Question: a reduced closure model of $(\widehat{v}_{1:K})$?

Space-time reduction:
reduce spatial dimension + increase time step size

Motivation: data assimilation with ensemble prediction

$$\begin{aligned} x' &= F(x) + U(x,y), && \text{resolved scales} && (\widehat{v}_{1:K}) \\ y' &= G(x,y), && \text{subgrid-scales} && (\widehat{v}_{K+1:N}) \end{aligned}$$

Data assimilation: partial noisy observation $\rightarrow$ prediction

- missing i.c. $\rightarrow$ ensemble prediction
- can only afford to resolve $x' = F(x)$

Motivation: data assimilation with ensemble prediction

$$x' = F(x) + U(x, y), \quad \text{resolved scales} \quad (\widehat{v}_{1:K})$$
$$y' = G(x, y), \quad \text{subgrid-scales} \quad (\widehat{v}_{K+1:N})$$

Data assimilation: partial noisy observation $\rightarrow$ prediction

- missing i.c. $\rightarrow$ ensemble prediction
- can only afford to resolve $x' = F(x)$

Objective: Develop a closure reduced model of x that

- captures key statistical + dynamical properties
- can be used for ensemble simulations

Closure modeling, model error UQ, subgrid parametrization

Direct constructions:

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

Inference/Data-driven ROM

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

Inference-based model reduction

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

KEY: **approx. the distribution of the stochastic process**

Approximate the discrete-time forward map:

$$x_n = F_n(x_{1:n-1})$$

- curse of dimensionality
- parametric inference: use the structure of the map

Discrete-time stochastic parametrization

NARMA(p, q) [Chorin-Lu15]

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

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

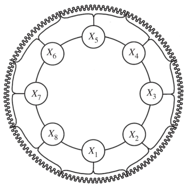
Tasks:

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

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



Model reduction for dissipative PDEs

nonlinear Galerkin

$\downarrow$

parametric inference

- 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 ✓;
- $dim(u) > K$: parametrization with time delay (Lu-Lin-Chorin17)

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

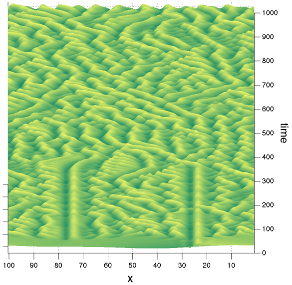
$$\Phi_k^n = \sum_{j=1}^{p} c_{k,j}^v u_k^{n-j} + c_{k,j}^R R^\delta(u_k^{n-j}) + c_{k,j}^w \sum_{\substack{|k-l|\le K, K<|l|\le 2K \\ \text{or } |l|\le K, K<|k-l|\le 2K}} \widetilde{u}_l^{n-1}\widetilde{u}_{k-l}^{n-j}$$

KEY: high-modes = functions of low modes

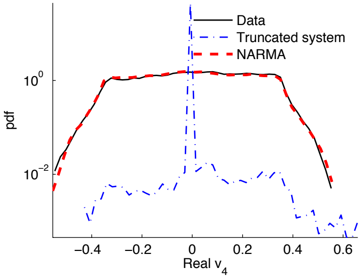
[1] Foias, Constantin, Temam, Sell, Jolly, Kevrekidis, Titi et al (88-94)

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

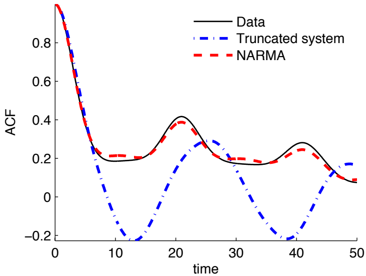
- 3 unstable modes
- 2 stable modes



Long-term statistics:



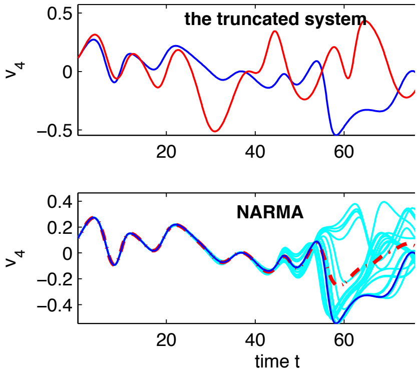
probability density function



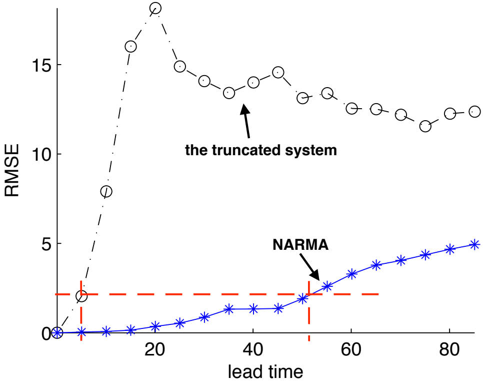
auto-correlation function

Prediction

A typical forecast:



RMSE of many forecasts:



Forecast time:

the truncated system: $T \approx 5$

the NARMA system: $T \approx$ **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

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

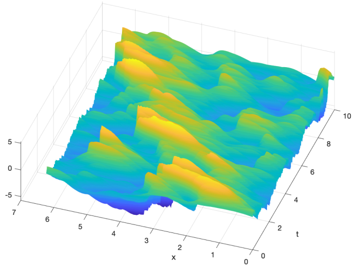
$$u_k^n = R^\delta(u_k^{n-1}) + f_k^n + g_k^n + \Phi_k^n,$$

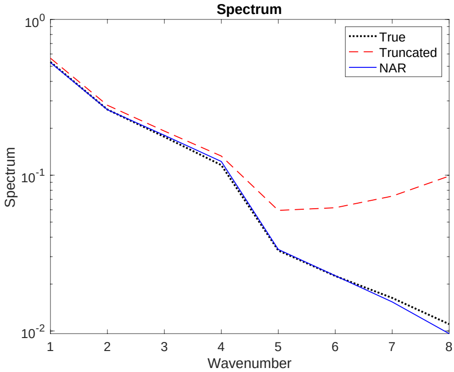
with $\Phi_k^n := \Phi_k^n(u^{n-p:n-1}, f^{n-p:n-1})$ in form of

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

Numerical tests:

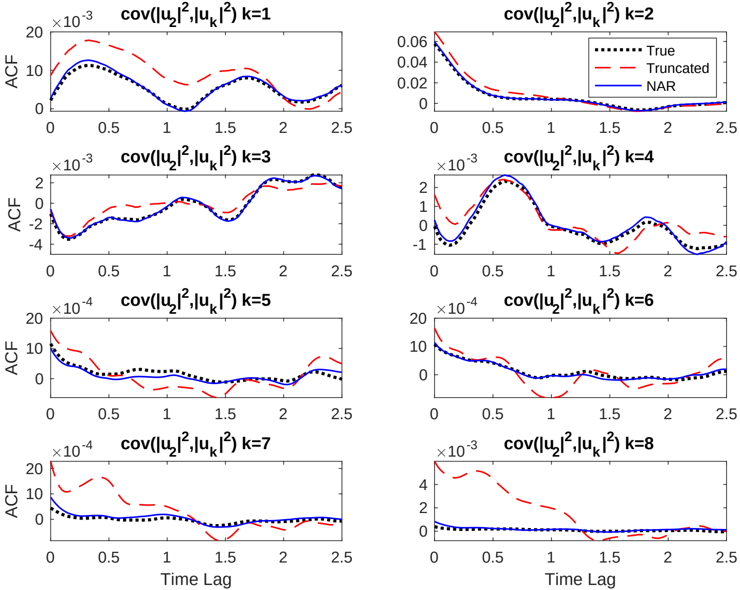
$\nu = 0.05$, $K_0 = 4 \to$ random shocks



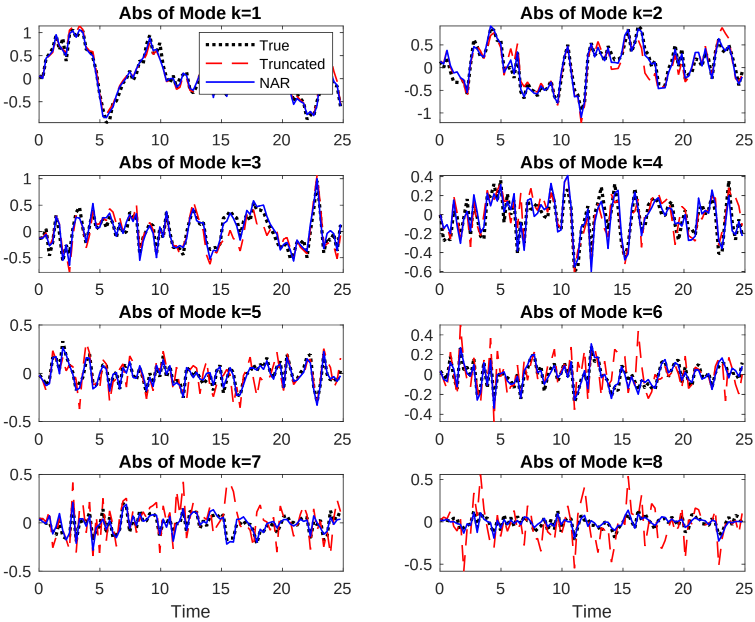


Energy spectrum

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



Cross-ACF of energy (4th moments!)



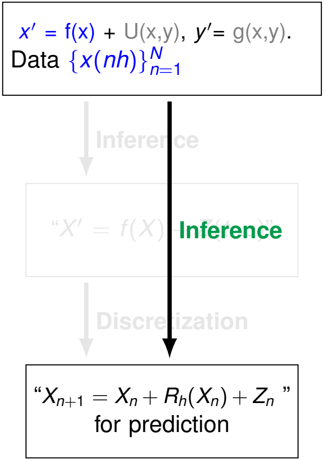
Trajectory prediction in response to force

Spacial-temporal reduction:

- how small can K (spatial dim.) be?
- how large can δ (time-step size) be?

$$\text{CFL number: } |u|\frac{dt}{dx} \sim |u|N dt \sim |u|K\delta$$

Summary and ongoing work



Inference-based stochastic model reduction

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

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

$$x_n = F_n(x_{1:n-1}) \approx \mathbb{E}[x_n | x_{1:n-1}]$$

$\rightarrow$ Effective stochastic reduced model

Open problems:

- general dissipative systems + model selection
- post-processing to predict shocks
- theoretical understanding of the approximation
  - optimal on the basis space in L^2 (Lin-L.19)
  - distance between the two stochastic processes?

References

- Data-driven stochastic model reduction
  - Chorin-Lu: Discrete approach to stochastic parametrization and dimension reduction in nonlinear dynamics. **PNAS 112 (2015)**, no. 32, 9804–9809.
  - Lu-Lin-Chorin: Comparison of continuous and discrete-time data-based modeling for hypoelliptic systems. **CAMCoS, 11 (2016)**, no. 8, 4227–4246.
  - Lu-Lin-Chorin: Data-based stochastic model reduction for the Kuramoto – Sivashinsky equation. **Physica D, 340 (2017)**, 46–57.
  - Lin-Lu: Data-driven model reduction, Wiener projections, and the Mori-Zwanzig formalism. preprint (2019)
  - Lu: Data-driven model reduction for stochastic Burgers equations. In preparation.
- Data assimilation
  - Lu-Tu-Chorin: Accounting for model error from unresolved scales in EnKFs: improving the forecast model. **MWR, 340 (2017)**.

Thank you!

FL acknowledges supports from JHU, LBL, NSF