Data-driven stochastic model reduction

Fei Lu
joint with Alexandre J. Chorin and Kevin Lin

UC Berkeley
Lawrence Berkeley National Laboratory
U of Arizona

July 18, 2016
ICERM, Brown University
Outline

Part I: Motivation for stochastic model reduction

Part II: A discrete-time approach

Part III: Examples
Climate modeling, numerical weather prediction:

\[
\frac{dx}{dt} = f(x) + U(x, y),
\]
\[
\frac{dy}{dt} = g(x, y).
\]

Observe only \(\{x(n\delta)\}_{n=1}^{N}\).

Forecast \(x(t), t \geq N\delta\).

Possible approaches:

- Estimate \(y\), and solve the full system
  - The full system is expensive to solve
  - The full system may be unknown
- Construct a reduced model for \(x\)
Why stochastic models

Data \( \{ x(n\delta) \} \Rightarrow \) properties of the full system:

- long-term statistics (ergodicity)

\[
\frac{1}{N} \sum_{i=1}^{N} F(x(n_i\delta)) \rightarrow E[F(x)]
\]

- positive Lyapunov exponents
- the effects of \( y \) act like random forces

**Goal:** develop a stochastic reduced system for \( x \) that can

- reproduce long-term statistics;
- make medium range predictions.
Part I: Motivation for stochastic model reduction

Part II: A discrete-time approach
- Previous work: continuous-time approach
- Discrete-time stochastic parametrization

Part III: Examples
The Mori-Zwanzig Formalism (Mori (65) and Zwanzig (73))

Rewrite the deterministic system in a form which resembles a generalized Langevin equation (Chorin-Hald (06)):

\[
\frac{dX_t}{dt} = f(X_t) + \int_0^t \Gamma(t - s, X_s)ds + W_t
\]

Message: memory exists in a closed system for \( x \).

Multi-level regression

Atmospheric sciences: Kondrashov et al (05-15), Wilks (05)

Hypoelliptic SDEs (Majda-Harlim (13,14))

\[
\begin{align*}
    dX_t &= f(X_t)dt + V_t dt, \\
    dV_t &= a(X_t, V_t)dt + \sigma dW_t.
\end{align*}
\]
Continuous-time approach:

\[ dX_t = f(X_t)dt + V_t dt, \]
\[ dV_t = a(X_t, V_t)dt + \sigma dW_t. \]
Continuous-time approach:

\[\begin{align*}
    dX_t &= f(X_t)dt + V_t dt, \\
    dV_t &= a(X_t, V_t)dt + \sigma dW_t.
\end{align*}\]

Discrete-time approach:

\[X_n = \Phi(X_{n-p:n-1}, \xi_{n-q:n-1}) + \xi_n\]
Identify a discrete-time stochastic system:

\[ X_n = X_{n-1} + \delta R_\delta(X_{n-1}) + \delta Z_n, \]
\[ Z_n = \Phi_n + \xi_n, \]
\[ \Phi_n = \sum_{j=1}^{p} a_j Z_{n-j} + \sum_{j=1}^{r} \sum_{i=1}^{s} b_{i,j} P_i(X_{n-j}) + \sum_{j=1}^{q} c_j \xi_{n-j}; \]

- \( R_\delta(X_{n-1}) \) from a discrete scheme for \( x' \approx f(x); \)
- \( \Phi_n = \Phi(Z_{n-p:n-1}, X_{n-r:n-1}, \xi_{n-q:n-1}) \) depends on the past;
- \( \xi_n \) are i.i.d \( N(0, \sigma^2); \)
- a Nonlinear AutoRegression Moving Average model (NARMA).

Structure: terms in \( \Phi_n; \)
Parameters: \( a_j, b_j, c_j, \) and \( \sigma. \)
$X_n = X_{n-1} + \delta R_\delta(X_{n-1}) + \delta Z_n,$

$Z_n = \Phi_n + \xi_n,$

$\Phi_n = \sum_{j=1}^{p} a_j Z_{n-j} + \sum_{j=1}^{r} \sum_{i=1}^{s} b_{i,j} P_i(X_{n-j}) + \sum_{j=1}^{q} c_j \xi_{n-j};$

Parameter estimation: conditional maximum likelihood estimator (MLE)
Conditional on $\xi_1, \ldots, \xi_q$, the log-likelihood of $X_{q+1:N}$ is

$$I(\theta|\xi_1, \ldots, \xi_q) = -\sum_{n=q+1}^{N} \frac{|Z_n-\Phi_n|^2}{2\sigma^2} - \frac{N-q}{2} \ln \sigma^2.$$

- when $q = 0$, MLE = least squares estimator (LSE)
- when $q > 0$, set $\xi_1 = \cdots = \xi_q = 0$
  - ergodicity $\Rightarrow$ MLE is consistent $\Rightarrow$ independent of $\xi_1, \ldots, \xi_q$

Structure derivation/selection:
- stability, long-term statistics
- numerical schemes
- analytical properties of the full system
Part I: Motivation for stochastic model reduction

Part II: A discrete-time approach

Part III: Examples:

- The Lorenz 96 system
- The Kuramoto-Sivashinsky equation
Example I: the Lorenz 96 system

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

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

where \( x \in \mathbb{R}^K, \ y \in \mathbb{R}^{JK} \).

- a chaotic system
- \( K = 18, \ J = 20: \ y \in \mathbb{R}^{360} \)

Find a reduced system for \( x \in \mathbb{R}^{18} \) based on

- Data \( \{x(n\delta)\}_{n=1}^{N} \)
- \( \frac{d}{dt}x_k \approx x_{k-1}(x_{k+1} - x_{k-2}) - x_k + 10. \)

Wilks, 2005
NARMA [Chorin-Lu (15)]:

\[ x^n = x^{n-1} + \delta R_\delta(x^{n-1}) + \delta z^n; \quad z^n = \Phi^n + \xi^n, \]

\[ \Phi^n = \mu + \sum_{j=1}^{p} a_j z^{n-j} + \sum_{j=1}^{r} \sum_{l=1}^{d_x} b_{j,l}(x^{n-j})^l + \sum_{j=1}^{s} \sum_{l=1}^{d_R} c_{j,l}(R_\delta(x^{n-j}))^l + \sum_{j=1}^{q} d_j \xi^{n-j}. \]

Polynomial autoregression (POLYAR) [Wilks (05)]:

\[ \frac{d}{dt} x_k = x_{k-1} (x_{k+1} - x_{k-2}) - x_k + 10 + U, \]

\[ U = P(x_k) + \eta_k, \quad \text{with} \quad \eta_k(t+\delta) = \phi \eta_k(t) + \xi(t) \]

where \( P(x) = \sum_{j=0}^{d} a_j x^j; \ \xi(t) \sim N(0, \sigma^2). \)
Long-term statistics

- run the reduced systems for a long time;
- compare with the data:

empirical probability density function

autocorrelation function
Short-term prediction

A typical ensemble forecast:

Forecast time:
   POLYAR: $t \approx 1$
   NARMA: $t \approx 2.5$

RMSE of many forecasts:
Example II: the Kuramoto-Sivashinsky equation

The KSE with periodic BC: \( v(x, t) = v(x + 2\pi \nu, t) \)

\[
\frac{\partial v}{\partial t} + \frac{\partial^2 v}{\partial x^2} + \frac{\partial^4 v}{\partial x^4} + v \frac{\partial v}{\partial x} = 0, \quad x \in \mathbb{R}, \quad t > 0.
\]

- spatio-temporally chaotic:
- Fourier: \( v(x, t) = \sum_{k=-\infty}^{+\infty} \hat{v}_k e^{iq_k x}, \)
  (here \( q_k = \frac{k}{\nu} \))

\[
\frac{d}{dt} \hat{v}_k = (q_k^2 - q_k^4) \hat{v}_k - \frac{i q_k}{2} \sum_{l=-\infty}^{\infty} \hat{v}_l \hat{v}_{k-l}.
\]

**Problem setting:** \( \nu = 3.43 \)
- observe only \( K = 5 \) modes, \((\hat{v}_k, k = 1, \ldots, K)\)
- predict their evolution
The truncated system is not accurate:

\[
\frac{d}{dt} \hat{v}_k = (q_k^2 - q_k^4) \hat{v}_k - \frac{iq_k}{2} \sum_{|l| \leq K, |k-l| \leq K} \hat{v}_l \hat{v}_{k-l}, \quad |k| \leq K.
\]

**NARMA**

\[
\begin{align*}
X_n &= X_{n-1} + \delta R_\delta(X_{n-1}) + \delta Z_n, \\
Z_n &= \Phi_n + \xi_n, \\
\Phi_n &= \sum_{j=1}^{p} a_j Z_{n-j} + \sum_{j=1}^{q} c_j \xi_{n-j} \\
&\quad + \sum_{j=1}^{r} \sum_{i=1}^{s} b_{i,j} P_i(X_{n-j}).
\end{align*}
\]

- **compute** \( R_\delta(X_{n-1}) \) (from the \( K \)-mode truncated system)
- **structure for** \( \Phi_n \):
  - different modes have different dynamics
  - nonlinear interaction between modes
Large time behavior of the KSE

(Constantin, Jolly, Kevrekidis, Titi et al (88-94))

Inertial manifold $\mathcal{M}$

Let $v = u + w$. Rewrite the KSE:

$$\frac{du}{dt} = Au + Pf(u + w)$$

$$\frac{dw}{dt} = Aw + Qf(u + w)$$

if $\mathcal{M}$ is the graph of a function $w = \psi(u)$,

$$\frac{du}{dt} = Au + Pf(u + \psi(u)).$$

Approximate inertial manifolds (AIMs)

Approximate the function $w = \psi(u)$:

- $\frac{dw}{dt} \approx 0 \Rightarrow w \approx A^{-1}Qf(u + w)$,
- Fixed point: $\psi_0 = 0$; $\psi_{n+1} = A^{-1}Qf(u + \psi_n)$.

An accurate AIM requires $m = \text{dim}(u)$ to be large!

(distance to the attractor $\sim |q_m|^{-\gamma}$)
NARMA with AILMs (Lu-Lin-Chorin (15))

The AILMs hint at how the high modes depend on the low modes:

\[ |k| > K : \hat{v}_k \approx \psi_{1,k} = (A^{-1}Qf(u))_k \Rightarrow \hat{v}_k \approx c_k \sum_{1 \leq |l|, |k-l| \leq K} \hat{v}_l \hat{v}_{k-l}. \]

Important: keep the terms, and estimate the coefficients

\[ \widetilde{u}_j^n = \begin{cases} u_j^n, & 1 \leq j \leq K; \\ i \sum_{l=j-K}^{K} u_l^n u_{j-l}^n, & K < j \leq 2K. \end{cases} \]

A discrete-time stochastic system:

\[ u_k^n = u_k^{n-1} + \delta R_k^\delta(u^{n-1}) + \delta z_k^n, \]
\[ z_k^n = \Phi_k^n + \xi_k^n, \]
\[ \Phi_k^n = \mu_k + \sum_{j=1}^{p} a_{k,j} z_k^{n-j} + \sum_{j=0}^{r} b_{k,j} u_k^{n-j} + \sum_{j=1}^{K} c_{k,j} \widetilde{u}_{j+k}^n \widetilde{u}_{j+k-k}^n \]
\[ + c_{k,(K+1)} R_k^\delta(u^n) + \sum_{j=1}^{q} d_{k,j} \xi_k^{n-j}. \]
Long-term statistics:

- Probability density function
- Auto-correlation function
Short-term prediction

A typical forecast:

![Graph showing the truncated system and the NARMA system with time t and RMSE on the y-axis.]

Forecast time:
- the truncated system: $t \approx 15$
- the NARMA system: $t \approx 50$

RMSE of many forecasts:

![Graph showing the RMSE for different ensemble sizes.]

- ensemble size = 1
- ensemble size = 20
Data-driven stochastic model reduction:

- High-dimensional Full system
- Discrete partial data

\[ X_n = \Phi(X_{n-p:n-1}, \xi_{n-q:n-1}) + \Psi(X_{n-p:n-1}, \xi_{n-q:n-1})\xi_n \]

- integrate numerical scheme into statistical inference
- easy parameter estimation
- flexible structure selection

Thanks for your attention!