Filtering without a model or with a partial model

Tyrus Berry
George Mason University

July 2, 2016

Joint work with John Harlim, PSU and Dimitris Giannakis, NYU
Model Error

- Partially known model $\dot{x} = f(x, \theta)$
- Dynamics $d\theta = a(\theta) \, dt + b(\theta) \, dW_t$ are unknown
- Stochastic parameterizations:

\[
(x^k(t), \theta^k(t)) \xrightarrow{\dot{x}=f(x,\theta)} (x^k(t+\tau), \theta^k(t+\tau))
\]

\[
p(\theta, t) \xrightarrow{\text{Diffusion Forecast}} p(\theta, t + \tau)
\]
Goal of the Diffusion Forecast

- **Stochastic Dynamics:** Autonomous, Markov, Ergodic
- **Data:** Generic observable
The Shift Map

- Consider a dynamical system $d\theta = a(\theta) \, dt + b(\theta) \, dW_t$
- Consider an uncertain initial state $\theta(0)$ with density $p(\theta, 0)$
- Given data samples $\theta_i = \theta(t_i)$ with $\tau = t_{i+1} - t_i$
- Using the Itô lemma we can show:

$$Sf(\theta_i) = f(\theta_{i+1}) = e^{\tau L} f(\theta_i) + \int_{t_i}^{t_{i+1}} \nabla f^\top b \, dW_s + \int_{t_i}^{t_{i+1}} Bf \, ds$$

- Feynman-Kac implies unbiased estimator: $\mathbb{E}[S(f)] = e^{\tau L} f$
- Project on smooth basis to minimize stochastic integrand $\nabla f^\top b$
Representing the Shift Map

- Choose a basis \( \{\varphi_j\} \) orthonormal with respect to \( \langle \cdot, \cdot \rangle_{p_{eq}} \)

- The coefficients \( c_l(t) = \langle p(\theta, t), \varphi_l \rangle \) have evolution:

\[
\begin{align*}
c_l(t + \tau) &= \langle p(\theta, t + \tau), \varphi_l \rangle = \langle e^{\tau L^*} p(\theta, t), \varphi_l \rangle = \langle p(\theta, t), e^{\tau L} \varphi_l \rangle \\
&= \sum_j c_j(t) \langle \varphi_j, e^{\tau L} \varphi_l \rangle_{p_{eq}} = \sum_j A_{lj} c_j(t)
\end{align*}
\]

- So \( \vec{c}(t + \tau) = A \vec{c}(t) \)

- Where \( A_{lj} = \langle \varphi_j, e^{\tau L} \varphi_l \rangle_{p_{eq}} \approx \frac{1}{N} \sum_{i=1}^{N} \varphi_j(\theta_i) \varphi_l(\theta_{i+1}) \)
Diffusion Maps → Custom Fourier Basis
Forecast Operator is Linear in the Diffusion Basis

\[ p(\theta, t) \xrightarrow{\text{Diffusion Forecast}} p(\theta, t + \tau) \]

\[ \langle p, \varphi_j \rangle \]

\[ \sum_j c_j \varphi_j p_{eq} \]

\[ \vec{c}(t) \xrightarrow{A_{lj} \equiv \mathbb{E}[\langle \varphi_j, S \varphi_l \rangle_{peq}]} \vec{c}(t + \tau) = A\vec{c}(t) \]

- Shift Map: \( S(\varphi_l)(\theta_i) = \varphi_l(\theta_{i+1}) \)
- Forecast Operator: \( A_{lj} = \varphi_j(\theta_i) \varphi_l(\theta_{i+1}) \)
- Diffusion maps: Eigenfunctions of \( \Delta + \frac{\nabla p_{eq}}{p_{eq}} \cdot \nabla \) are optimal basis
Forecasting with the Shift Map

\[ p(\theta, t) \xrightarrow{\text{Nonparametric Forecast}} p(\theta, t + \tau) \]

\[ \langle p, \varphi_j \rangle \xrightarrow{\text{Filtering: No model/Partial model}} \sum_j c_j \varphi_j p_{eq} \]

\[ \bar{c}(t) \xrightarrow{A_{lj} = \mathbb{E}[\langle \varphi_j, S \varphi_l \rangle_{peq}]} \bar{c}(t + \tau) = A\bar{c}(t) \]

- We approximate \( c_l(t) \approx \frac{1}{N} \sum_{i=1}^{N} \varphi_l(\theta_i)p(\theta_i, t)/p_{eq}(\theta_i) \)
- We approximate \( A_{lj} \) with \( \hat{A}_{lj} = \frac{1}{N} \sum_{i=1}^{N} \varphi_j(\theta_i)\varphi_l(\theta_{i+1}) \)
- \( \mathbb{E}[\hat{A}_{lj}] = A_{lj} \) with error \( \mathcal{O}(||\nabla \varphi_l||_{p_{eq}} \sqrt{\tau/N}) \)
Filtering with the Shift Map

Introduce an observable $y = h(\theta) + \nu$ with $\nu \sim q$

\[
p^a(\theta, t - \tau) \xrightarrow{\text{Diffusion Forecast}} p^f(\theta, t) \xrightarrow{p^f(\theta)p(y|\theta)} p^a(\theta, t)
\]

\[
\langle p^a, \varphi_j \rangle \xrightarrow{\sum_j c^f_j \varphi_j p_{eq}} \langle p^a, \varphi_j \rangle
\]

\[
\tilde{c}^a(t - \tau) \xrightarrow{A_{lj}c^a(t - \tau)} \tilde{c}^f(t) \xrightarrow{\text{Bayesian Update}} \tilde{c}^a(t)
\]

- Likelihood is $p(y|\theta) = q(y - h(\theta))$
- Evaluate on the training data $p^a(\theta_i) = p^f(\theta_i)q(y - h(\theta_i))$
- Continuing work: Learn the conditional density from data
Recovering the Kalman Filter for Linear Systems

- **Linear Dynamics:** $dx = ax\ dt + b\ dW_t$
- **Linear Observation:** $dy = x\ dt + R\ dW_t$
Filtering in a Double Well Potential

- **Double Well Potential:** $dx = x(1 - x^2) \, dt + b \, dW_t$
- **Linear Observation:** $dy = x \, dt + R \, dW_t$
Non-observatibility

- **Double Well Potential:** \( dx = x(1 - x^2) \, dt + b \, dW_t \)
- **Absolute Value Observation:** \( dy = |x| \, dt + R \, dW_t \)
Restricted Observability

- **Double Well Potential:** \( dx = x(1 - x^2) \, dt + b \, dW_t \)
- **Tough Observation:** \( dy = (x - 0.05)^2 \, dt + R \, dW_t \)
Double Well with Observability Restrictions

\[ dy = |x| \, dt + R \, dW_t \]

\[ dy = (x - 0.05)^2 \, dt + R \, dW_t \]
Model Error and the Curse of Dimensionality

- Nonparametric model *interpolates* from the training data
- Data required grows exponentially in the dimension of the manifold
- For high-dimensional systems we usually have an approximate model
- High-dimensional models are subject to model error
- Idea: Use the nonparametric model for the model error
Semiparametric Model

- Partially known model: $\dot{x} = f(x, \theta)$
- Dynamics: $d\theta = a(\theta) \, dt + b(\theta) \, dW_t$ are unknown
- Build a Diffusion Forecast model for $p(\theta, t)$
- Sample $\theta^k(t) \sim p(\theta, t)$ to use with ensemble forecast $(x^k, \theta^k)$

\[
\begin{align*}
(x^k(t), \theta^k(t)) & \xrightarrow{\dot{x}=f(x,\theta)} (x^k(t+\tau), \theta^k(t+\tau)) \\
p(\theta, t) & \rightarrow \text{Diffusion Forecast} \rightarrow p(\theta, t+\tau)
\end{align*}
\]
Semiparametric Filter: It’s a bit complicated...

\[
\begin{align*}
\left( x^{k,a}(t - \tau) \right) & \xrightarrow{\dot{x} = f(x, \theta)} \left( x^{k,f}(t) \right) \\
\left( \theta^{k,a}(t - \tau) \right) & \xrightarrow{\text{Diffusion Forecast}} \left( \theta^{k,f}(t) \right)
\end{align*}
\]

\[
\begin{align*}
p^a(\theta, t - \tau) & \xrightarrow{\text{Bayesian Update}} \bar{c}^a(t - \tau) \\
\langle p^a, \varphi_j \rangle & \xrightarrow{\text{Filtering: No model/Partial model}} \sum_j c^f_j \varphi_j p_{eq} \langle p^a, \varphi_j \rangle \\
\bar{c}^a(t - \tau) & \xrightarrow{A_{lj}c^a(t-\tau)} \bar{c}^f(t)
\end{align*}
\]
Example: 40-dimensional Lorenz-96 system

\[ \dot{x}_i = f(x_i, \theta) = \theta x_{i-1}x_{i+1} - x_{i-1}x_{i-2} - x_i + 8 \]
Example: 40-dimensional Lorenz-96 system

\[ \dot{x}_i = f(x_i, \theta) = \theta x_{i-1} x_{i+1} - x_{i-1} x_{i-2} - x_i + 8 \]
Additional challenges of semiparametric modeling

- Need a training data set for $\theta$
- Need initial condition $p(\theta, t)$ for nonparametric forecast
- We developed semiparametric filtering to address these
- Still require that evolution of $\theta$ does not depend on $x$
Building the basis

▶ Coifman and Lafon, *Diffusion maps*.
▶ Berry and Harlim, *Variable Bandwidth Diffusion Kernels*.
▶ Berry and Sauer, *Local Kernels and the Geometric Structure of Data*.

**Nonparametric forecast**

▶ Berry, Giannakis, and Harlim, *Nonparametric forecasting of low-dimensional dynamical systems*.
▶ Berry and Harlim, *Forecasting Turbulent Modes with Nonparametric Diffusion Models*.

**Semiparametric forecast**

▶ Berry and Harlim, *Semiparametric forecasting and filtering: correcting low-dimensional model error in parametric models*.
Franz Hamilton (North Carolina State University), et al.

**Ensemble Kalman Filtering without a Model** in PRX

- No model, but observation function is known
- Only care about the mean (non-probabilistic)
- Takens embedding + local linear forecast $\Rightarrow$ Short term forecast
- Apply EnKF using the local linear forecast for the ensemble
- High-dimensional systems: assimilate one variable at a time
Filtering Lorenz-96 with no model (one variable at a time):