Bayesian Approaches for Data-Driven Learning of Dynamical Systems

Engineering Mechanics Institute Conference and Probabilistic Mechanics & Reliability Conference (Virtual)
May 25-28, 2021

Alex Gorodetsky and Nicholas Galioto
Aerospace Engineering, UM
Motivation

- We seek to learn physically relevant models from time-series data in a way that
  - Handles sparse and noisy data
  - Scales well with dimension
  - Captures uncertainty in predictions
- Two primary design choices must be made
  - Model structure: linear, linear subspace, neural networks, etc.
  - Objective functions: least squares, regularization, Bayesian posterior
Main takeaway

- A majority of existing approaches minimize least-squares objectives
  
  1. Assume perfect model
     \[ \dot{x} = f(t, x; \theta) \]
     \[ J(\theta) = \sum_{i=1}^{n} \| y_i - x(t_i) \|^2 \]
  
  2. Effectively assume noiseless measurements (DMD)
     \[ J(\theta) = \sum_{i=1}^{n} \| y_i - \Psi(y_{i-1}; \theta) \|^2 \]

- Our approach: assume noisy measurements + **stochastic model**
  
  - Provides optimal combination of (1) and (2)
  - Provides natural dynamics-based regularization
Probabilistic modeling framework
Joint parameter-state inference problem

Parameterized Hidden Markov Model
Generic formulation: linear, physical model, dictionary of bases, NN, etc.

\[ X_{k+1} = \Psi(X_k; \theta) + \eta_k \]
\[ Y_k = h(X_k) + \nu_k, \]
\[ \eta_k \sim \mathcal{N}(0, \Sigma) \]
\[ \nu_k \sim \mathcal{N}(0, \Gamma) \]

1. Accounts for model parameter uncertainty through \( \theta \)
2. Accounts for model form uncertainty through \( \eta_k \)
3. Accounts for measurement noise through \( \nu_k \)
Bayesian inference
Updating models with data

Prior $p(\theta, X_n | I)$

Bayes’ Rule

Posterior $p(\theta, X_n | Y_n, I)$

For our Markovian system

$$p(\theta, X_n | Y_n, I) = \frac{1}{Z} \left[ \prod_{i=1}^{n} p(y_i | X_i, \theta) \right] \left[ \prod_{i=1}^{n} p(X_i | X_{i-1}, \theta) \right] p(\theta | I)$$
Log posterior is a generalized objective function
Many existing approaches recovered under simplifying assumptions

1. Identity observations: Hills et. al. 2015, Raissi, 2018, Qin et.al 2019
3. Many of these works look at parameterizing models (e.g., via NN architectures)
Log posterior is a generalized objective function
Many existing approaches recovered under simplifying assumptions

1. Identity observations: Hills et al. 2015, Raissi, 2018, Qin et al. 2019
3. Many of these works look at parameterizing models (e.g., via NN architectures)

Accounting for all uncertainties
Inference with the marginal likelihood

**Goals: learn and predict**

- Learn the model: \( p(\theta \mid \mathcal{Y}_n, I) \)
- Make (future) predictions: \( p(X_k \mid \mathcal{Y}_n, I) = \int p(X_k \mid \theta)p(\theta \mid \mathcal{Y}_n, I)d\theta \)

Marginal posterior — required by Optimization/Markov Chain Monte Carlo

\[
p(\theta \mid \mathcal{Y}_n, I) = \int p(\theta, \mathcal{X}_n \mid \mathcal{Y}_n, I)d\mathcal{X}_n = \frac{1}{Z} p(\mathcal{Y}_n \mid \theta) p(\theta)
\]

Marginal likelihood

Evaluating the marginal posterior

### Recursive evaluation $p(\theta \mid \mathcal{Y}_n)$

1. for $k = 1$ to $n$ do
2. Predict $p(X_k \mid \theta, \mathcal{Y}_{k-1}) = \int p(X_k \mid \theta, X_{k-1})p(X_{k-1} \mid \theta, \mathcal{Y}_{k-1})dX_{k-1}$
3. Compute the evidence $p(y_k \mid \theta, \mathcal{Y}_{k-1}) = \int p(y_k \mid \theta, X_k)p(X_k \mid \theta, \mathcal{Y}_{k-1})dX_k$
4. Update filter $p(X_k \mid \theta, \mathcal{Y}_k) = \frac{p(y_k \mid \theta, X_k)p(X_k \mid \theta, \mathcal{Y}_{k-1})}{p(y_k \mid \theta, \mathcal{Y}_{k-1})}$
5. Update posterior $p(\theta \mid \mathcal{Y}_k) = \frac{p(y_k \mid \theta, \mathcal{Y}_{k-1})p(\theta \mid \mathcal{Y}_{k-1})}{p(y_k \mid \mathcal{Y}_{k-1})}$
6. end for
Linear pendulum: DMD model is “correct”
Correct DMD model still does not recover system

Reconstruction

(a) $x_1$, $\sigma = 10^{-2}$, $n = 8$
(b) $x_2$, $\sigma = 10^{-2}$, $n = 8$
(c) $x_1$, $\sigma = 10^{-1}$, $n = 40$
(d) $x_2$, $\sigma = 10^{-1}$, $n = 40$
Linear pendulum: DMD model is “correct”
Correct DMD model still does not recover system

Reconstruction

<table>
<thead>
<tr>
<th>Posterior</th>
<th>Samples</th>
<th>Data</th>
<th>Mean</th>
<th>DMD</th>
<th>TDMD</th>
<th>Truth</th>
</tr>
</thead>
</table>

(a) $x_1, \sigma = 10^{-2}, n = 8$
(b) $x_2, \sigma = 10^{-2}, n = 8$
(c) $x_1, \sigma = 10^{-1}, n = 40$
(d) $x_2, \sigma = 10^{-1}, n = 40$

Prediction

<table>
<thead>
<tr>
<th>Posterior</th>
<th>Samples</th>
<th>Mean</th>
<th>DMD</th>
<th>TDMD</th>
<th>Truth</th>
</tr>
</thead>
</table>

(a) $x_2, \sigma = 10^{-2}, n = 8$
(b) $x_2, \sigma = 10^{-1}, n = 40$
Linear pendulum: DMD model is “correct”
Correct DMD model still does not recover system

Reconstruction

<table>
<thead>
<tr>
<th>Posterior</th>
<th>Samples</th>
<th>Data</th>
<th>Mean</th>
<th>DMD</th>
<th>TDMD</th>
<th>Truth</th>
</tr>
</thead>
</table>

(a) $x_1, \sigma = 10^{-2}, n = 8$
(b) $x_2, \sigma = 10^{-2}, n = 8$
(c) $x_1, \sigma = 10^{-1}, n = 40$
(d) $x_2, \sigma = 10^{-1}, n = 40$

Prediction

<table>
<thead>
<tr>
<th>Posterior</th>
<th>Samples</th>
<th>Mean</th>
<th>DMD</th>
<th>TDMD</th>
<th>Truth</th>
</tr>
</thead>
</table>

(a) $x_2, \sigma = 10^{-2}, n = 8$
(b) $x_2, \sigma = 10^{-1}, n = 40$

Spectrum

<table>
<thead>
<tr>
<th>Posterior</th>
<th>Mean</th>
<th>DMD</th>
<th>TDMD</th>
<th>True</th>
</tr>
</thead>
</table>

(a) Lower $n, \sigma$
(b) Higher $n, \sigma$
Lorenz 63
With only 300 noisy data points, we recover the attractor

\[ \dot{x} = \sigma (y - x) \]
\[ \dot{y} = x (\rho - z) - y \]
\[ \dot{z} = xy - \beta z \]

Posterior Lyapunov exp.

(a) \( \lambda_1 \) Estimate  
(b) \( \lambda_2 \) Estimate  
(c) \( \lambda_3 \) Estimate
Specializing to Hamiltonian systems

- Hamiltonian systems are reversible and preserve certain invariants (energy)
  \[
  \dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q}
  \]

- Dynamics $\Psi$ become a mixture of a leap frog and Hamiltonian parameterization:
  \[
  H = \frac{1}{2} p^T M^{-1}(q) p + U(q, p)
  \]
  \[
  \Psi(q_k, p_k; \theta_\Psi) = \begin{bmatrix}
  q_k + \Delta t p_k - \frac{\Delta t^2}{2} \left. \frac{\partial U(q, p, \theta_\Psi)}{\partial q} \right|_{q_k} \\
  p_k - \frac{\Delta t}{2} \left( \left. \frac{\partial U(q, p, \theta_\Psi)}{\partial q} \right|_{q_k} + \left. \frac{\partial U(q, p, \theta_\Psi)}{\partial q} \right|_{q_{k+1}} \right)
  \end{bmatrix},
  \]

- We parameterize the potential energy $U$
- We both learn the Hamiltonian and assume the data is from a symplective process
Hénon Heiles Potential

\[ U(q_1, q_2) = \frac{1}{2}q_1^2 + \frac{1}{2}q_2^2 + q_1^2 q_2 - \frac{1}{3}q_2^3 \]

Posterior estimates of \( q_1 \) trajectory

Time (s)

Process noise marginal posteriors

Symplectic approach learns a model with an order of magnitude greater certainty
Conclusions and Acknowledgements

1 Conclusions
   - Learning stochastic models is beneficial for system ID to handle
     - Parameter uncertainty
     - Model uncertainty
     - Measurement uncertainty
   - MCMC approaches are feasible for moderately large problems
   - How do we know the size of the latent space?

2 Papers on my website
   www.alexgorodetsky.com/publications.html

3 Funding: DARPA PAI and AIRA, AFOSR Computational Mathematics
Computational science for autonomy

Compression enabled control and estimation

Real-time autonomy

Source code: github.com/goroda, papers: alexgorodetsky.com

Thanks!