Bayesian Approaches for Data-Driven Learning of Dynamical Systems

Nick Galioto and Alex Gorodetsky

Department of Aerospace Engineering, University of Michigan

Funding by DARPA program on the physics of AI

Problem Statement

Data-driven methods are popular for system identification, but existing algorithms carry the following problems:

- Imprecise modeling accounting for model form error
- Lack of robustness to noisy data

We seek a probabilistic modeling framework that

- Enables a Bayesian inference approach to account for uncertainty
- Recovers existing common approaches under idealized conditions
  - Dynamic Mode Decomposition (DMD) [1]
  - Sparse Identification of Nonlinear Dynamics (SINDy) [2].
- Enables deviation from idealized conditions

Approach: Bayesian methods use a flexible model to inform the negative log posterior with guaranteed optimality for derived estimators (mean, MAP, mode) -- whereas many existing system ID methods solve heuristic optimization problems, yielding sub-optimal estimators.

Probabilistic Formulation

1. Define a dynamical system in its most general form:

$$x_k = \Psi(x_{k-1}, \theta_\Psi) + \xi_k, \qquad \xi_k \sim \mathcal{N}(0, \Sigma(\theta_\Sigma));$$
$$y_k = h(x_k, \theta_h) + \eta_k, \qquad \eta_k \sim \mathcal{N}(0, \Gamma(\theta_\Gamma)).$$

2. Utilize Bayes' rule to derive parameter-state joint posterior

$$p(\theta, x|y) \propto p(y|\theta, x)p(x|\theta)p(\theta),$$

where the right-hand probabilities are defined as

$$p(y|\theta, x) = \prod_{k=1}^N \mathcal{N}(y_k;\ h(x_k, \theta_h), \Gamma(\theta_\Gamma)),$$
$$p(x|\theta) = \prod_{k=1}^N \mathcal{N}(x_k; \Psi(x_{k-1}, \theta_\Psi), \Sigma(\theta_\Sigma)).$$

Taking the negative log yields the result given in the top middle panel.

The marginal posterior is derived similarly

$$p(\theta|y) \propto p(y|\theta)p(\theta).$$

Algorithm

We require a Markov chain Monte Carlo (MCMC) algorithm to sample from the parameter-state joint posterior.

- Challenge: The states are high dimensional making sampling costly and inefficient
- Solution: Pseudo-marginal MCMC [3]

To sample from the parameter marginal posterior, the following algorithm is used:

for $j = 1$ to M **do**

1. Sample from proposal $\theta^* \sim \pi(\mathcal{X}_n, \theta|\mathcal{Y}_n) = \pi(\theta)p(\mathcal{X}_n|\theta, \mathcal{Y}_n)$

for $k = 1$ to N **do**

2. Predict $p(x_k|\theta^*, \mathcal{Y}_{k-1}) = \int p(x_k|\theta^*, x_{k-1})p(x_{k-1}|\theta^*, \mathcal{Y}_{k-1})dx_{k-1}$
3. Compute the evidence $p(y_k|\theta^*, \mathcal{Y}_{k-1}) = \int p(y_k|\theta,^* x_k)p(x_k|\theta^*, \mathcal{Y}_{k-1})dx_k$
4. Update $p(x_k|\theta^*, \mathcal{Y}_k) = \frac{p(y_k|\theta^*, \mathcal{Y}_{k-1})p(\theta^*|\mathcal{Y}_{k-1})}{p(y_k|\mathcal{Y}_{k-1})}$
5. Update $p(\theta^*|\mathcal{Y}_k) = \frac{p(y_k|\theta^*, \mathcal{Y}_{k-1})p(\theta^*|\mathcal{Y}_{k-1})}{p(y_k|\mathcal{Y}_{k-1})}$

end for

6. Accept θ^* with probability $\alpha(\theta, \theta^*) = \frac{p(\theta^*|\mathcal{Y}_n)\pi(\theta)}{p(\theta|\mathcal{Y}_n)\pi(\theta^*)}$

end for

Steps 2-4 are computed in the

- Linear case with a Kalman filter
- Nonlinear case with an unscented Kalman filter

The objective functions of DMD and SINDy are special cases of the negative log posterior under certain modeling assumptions

Full Negative Log Posterior

$$-\operatorname{L}p(\theta, x|y) \propto \sum_{i=1}^N ||y_i - h(x_i, \theta_h)||_\Gamma^2 + \sum_{i=1}^N ||x_i - \Psi(x_{i-1}, \theta_\Psi)||_\Sigma^2 - \operatorname{L}p(\theta)$$

DMD Optimization and Assumptions

$$A = \underset{\tilde{A}}{\operatorname{argmin}} \sum_{i=1}^N \left|y_i - \tilde{A}y_{i-1}\right|^2$$

i. noiseless measurements
ii. identity observations
iii. linear dynamics
iv. maximum likelihood estimator
v. identity process covariance

SINDy Optimization and Assumptions

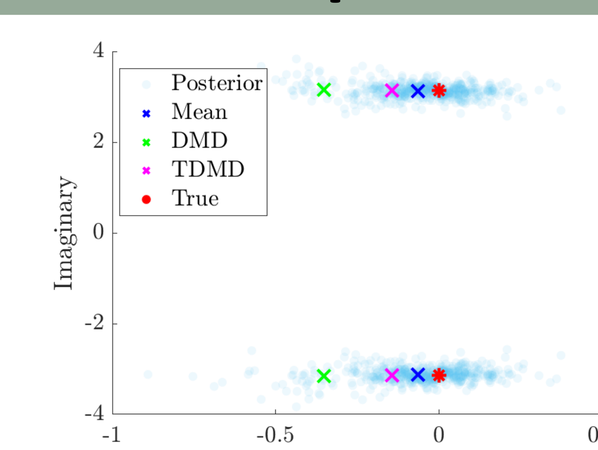
$$\theta_\Psi = \underset{\tilde{\theta}}{\operatorname{argmin}} \sum_{i=1}^N \left|\frac{y_i - y_{i-1}}{\Delta t} - \Xi(y_{i-1})\tilde{\theta}\right|^2 + \lambda||\tilde{\theta}||$$

i. noiseless measurements
ii. identity observations
iii. assumed time stepping scheme for derivatives
iv. a sparsity promoting prior
v. identity process covariance

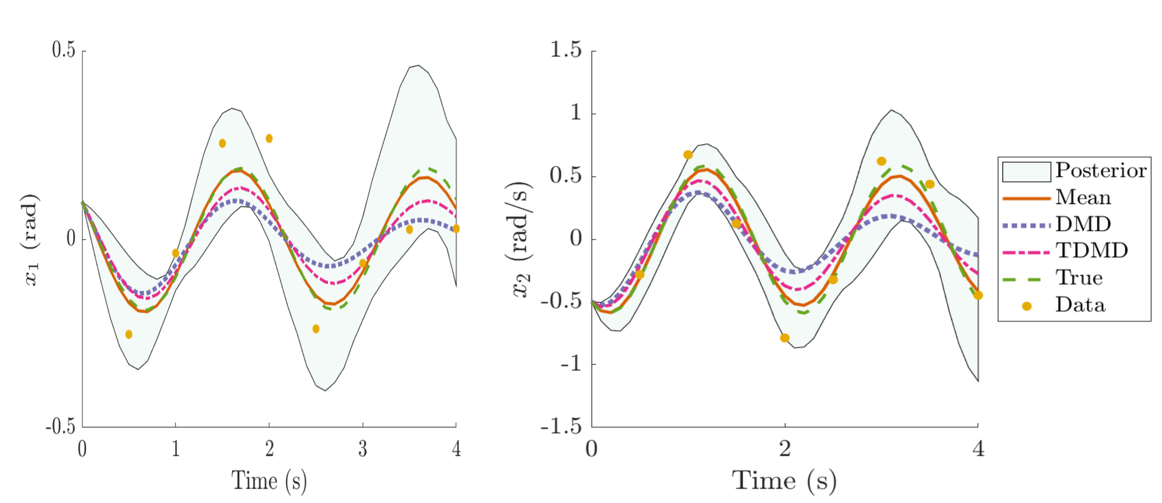
Linear Pendulum: Bayes and DMD Comparison

$$x_k = A(\theta_\Psi)x_{k-1} + \xi_k, \qquad \xi_k \sim \mathcal{N}(0, \Sigma(\theta_\Sigma))$$
$$y_k = x_k + \eta_k, \qquad \eta_k \sim \mathcal{N}(0, \Gamma(\theta_\Gamma))$$

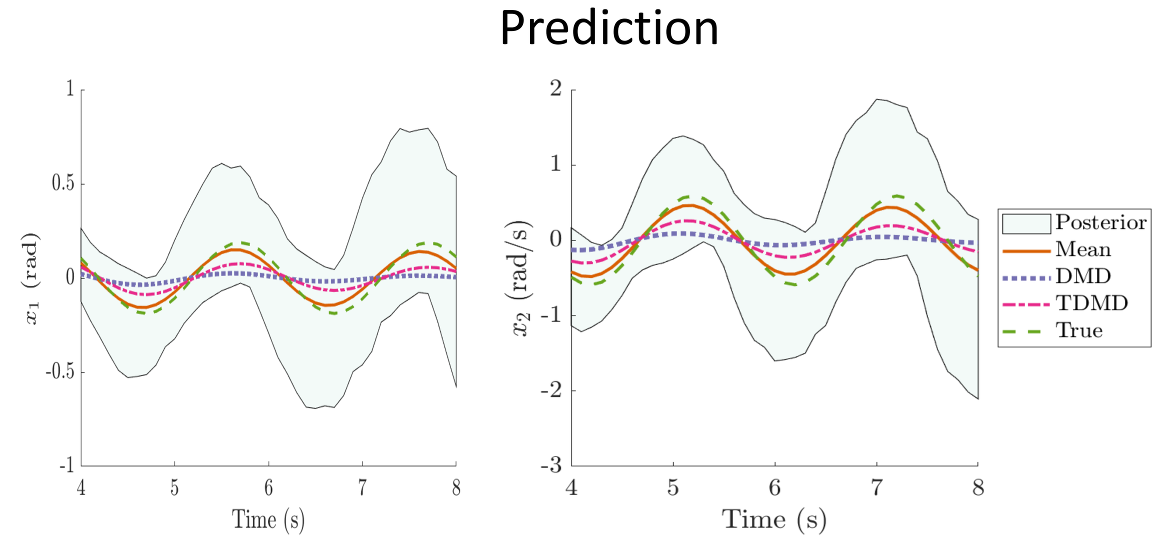
$$A(\theta_\Psi) = \begin{bmatrix}\theta_1 & \theta_2 \\ \theta_3 & \theta_4\end{bmatrix}, \Sigma(\theta_\Sigma) = \begin{bmatrix}\theta_5 & 0 \\ 0 & \theta_5\end{bmatrix}, \Gamma(\theta_\Gamma) = \begin{bmatrix}\theta_6 & 0 \\ 0 & \theta_6\end{bmatrix}$$

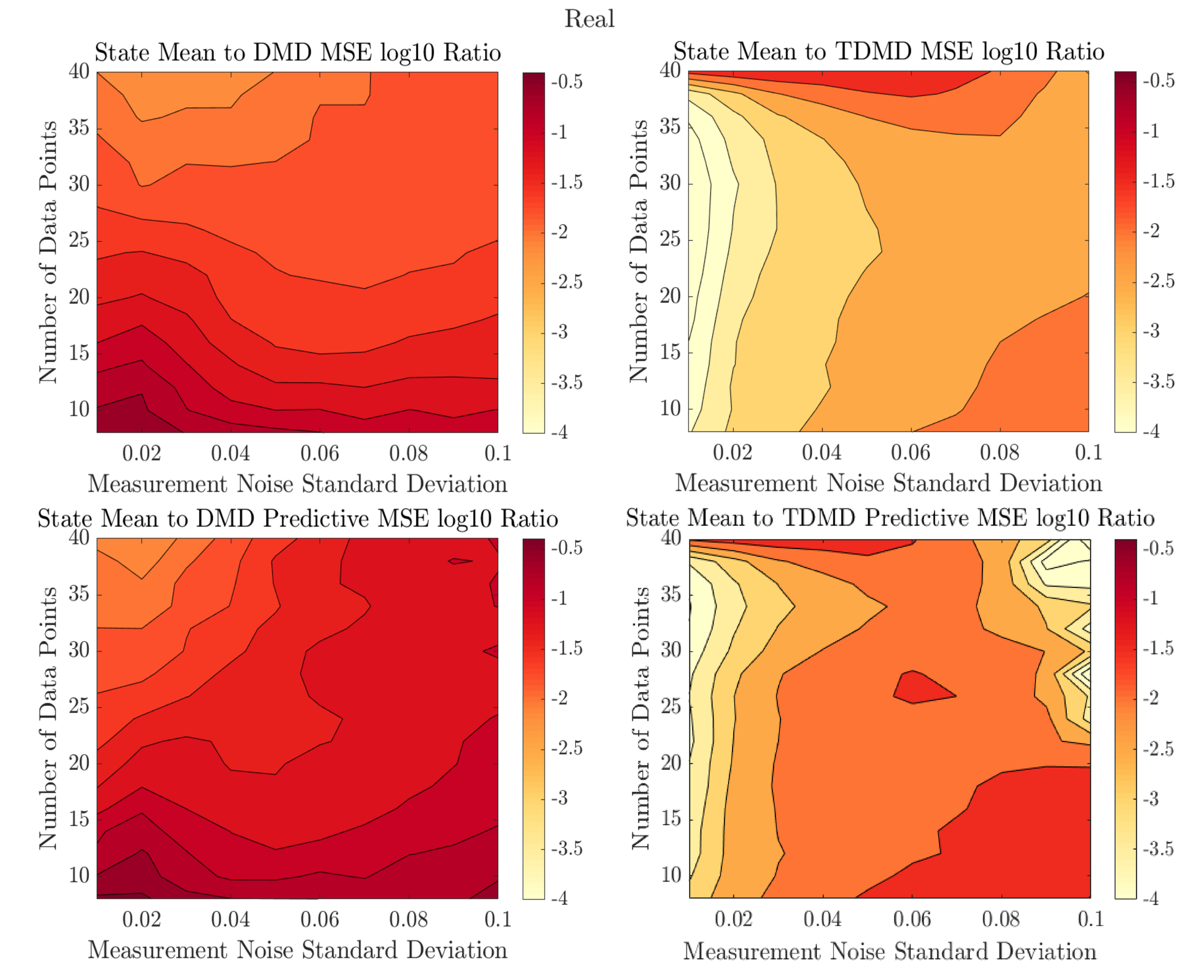


Reconstruction



Prediction

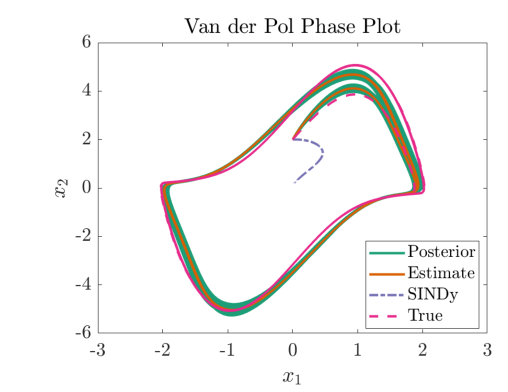


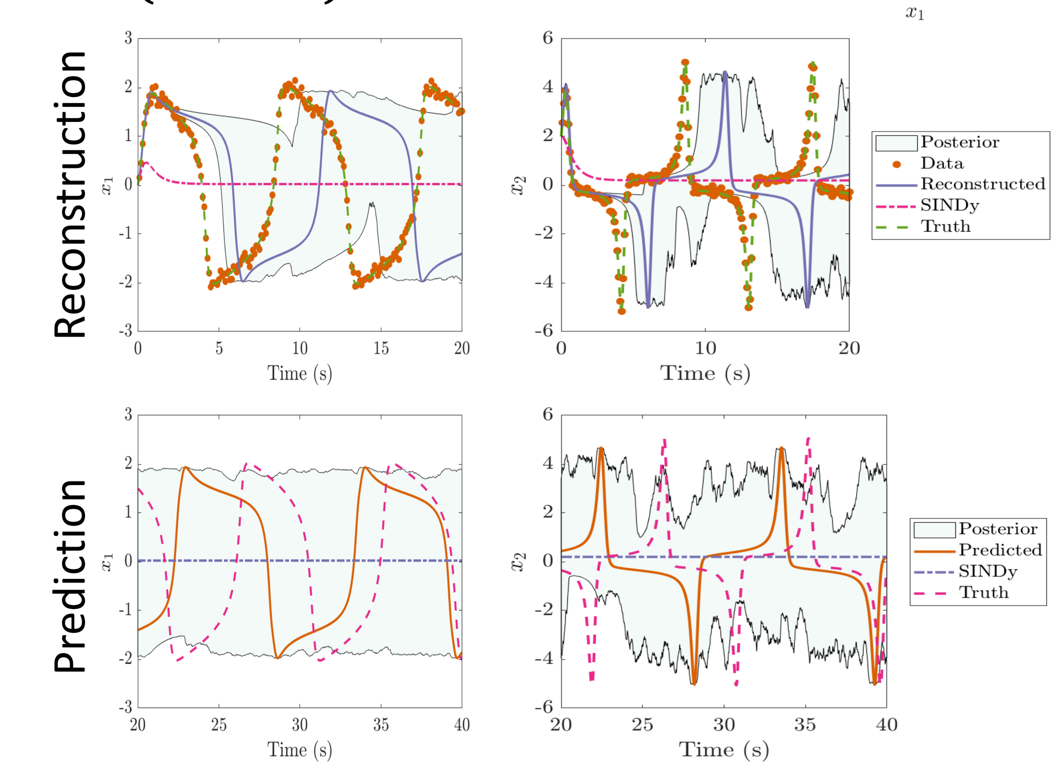


Nonlinear Examples

Van der Pol Oscillator

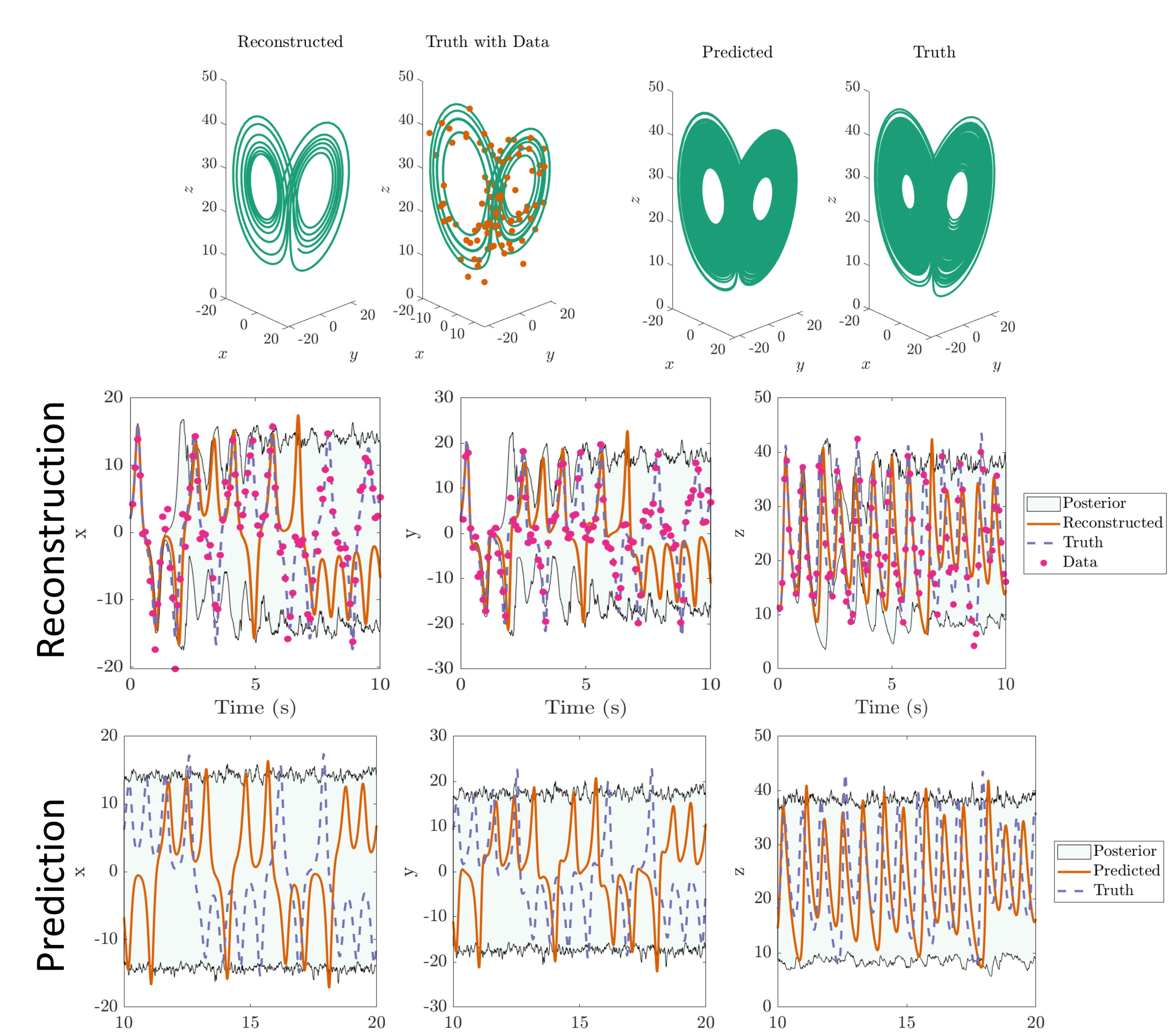
$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = \mu(1 - x_1^2)x_2 - x_1$$





Lorenz 63

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



References

[1] Schmid, Peter J. "Dynamic Mode Decomposition of Numerical and Experimental Data." *Journal of Fluid Mechanics*, vol. 656, 2010, pp. 5–28.

[2] Brunton, Steven L., Joshua L. Proctor, and J. Nathan Kutz. "Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems." Proceedings of the National Academy of Sciences 113.15 (2016): 3932–3937.

[3] Andrieu, Christophe, and Gareth O. Roberts. "The Pseudo-Marginal Approach for Efficient Monte Carlo Computations." The Annals of Statistics 37.2 (2009): 697–725.

[4] Galioto, Nicholas and Gorodetsky, Alex. "Bayesian Inference for Data-Driven System Identification." [In preparation]