Bayesian Approaches for Data-Driven Learning of Dynamical Systems

Physics Informed Machine Learning Workshop
Santa Fe, NM, USA
January 12-17, 2020

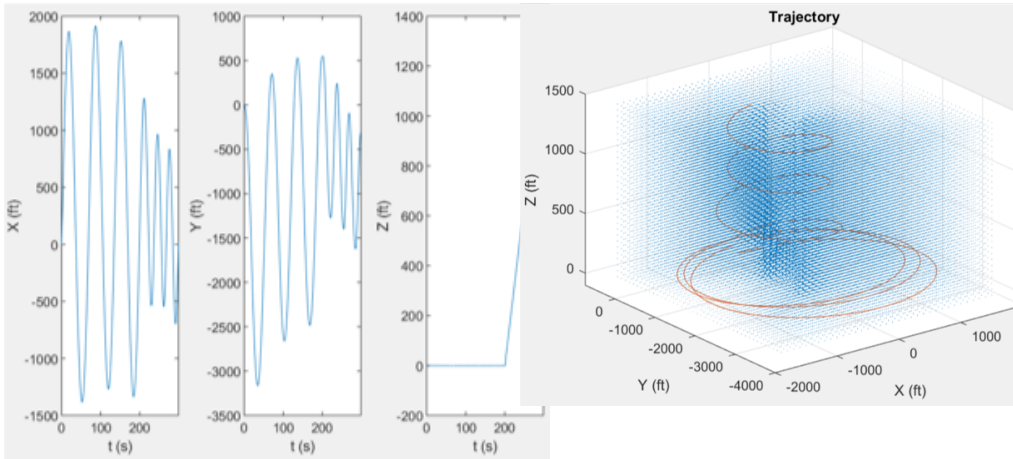
Nicholas Galioto and Alex Gorodetsky,
Aerospace Engineering, UM

UNIVERSITY OF MICHIGAN

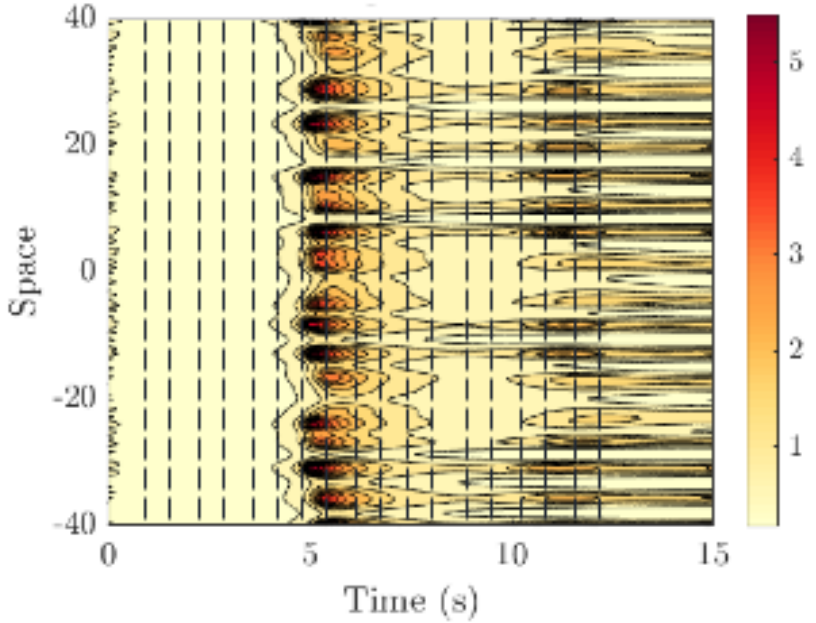
Motivation 1: autonomous dynamical systems
Noisy data and fast dynamics

- Goal: control and plan in systems with unknown system dynamics
- Problem: system identification must happen from limited onboard sensor measurements in real-time
  - Handle dramatic changes in the dynamics
  - Avoid extensive system characterization time
  - Enable real-time replanning

AAI RQ-2 Pioneer, courtesy Wikimedia

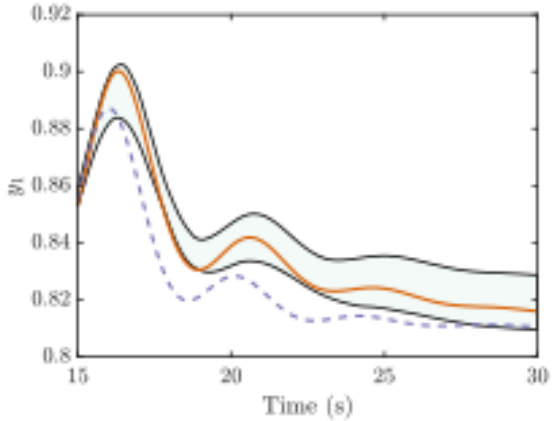


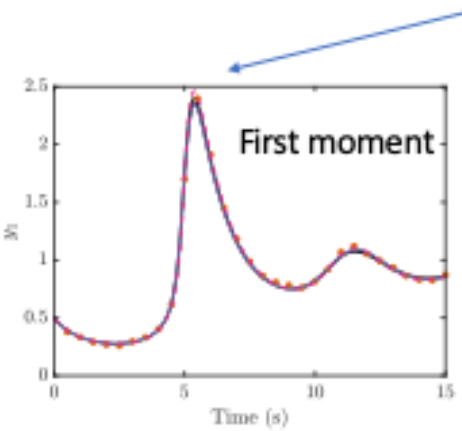
Motivation 2: indirectly observed PDE systems
Time-varying Reaction-diffusion PDE

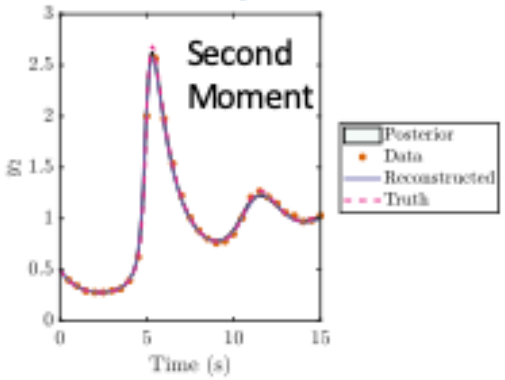


$$\frac{\partial C_1}{\partial t} = \theta_1 \frac{\partial^2 C_1}{\partial x^2} + 0.1 - C_1 + \theta_3 C_1^2 C_2$$
$$\frac{\partial C_2}{\partial t} = \theta_2 \frac{\partial^2 C_2}{\partial x^2} C_2 + 0.9 - C_1^2 C_2$$

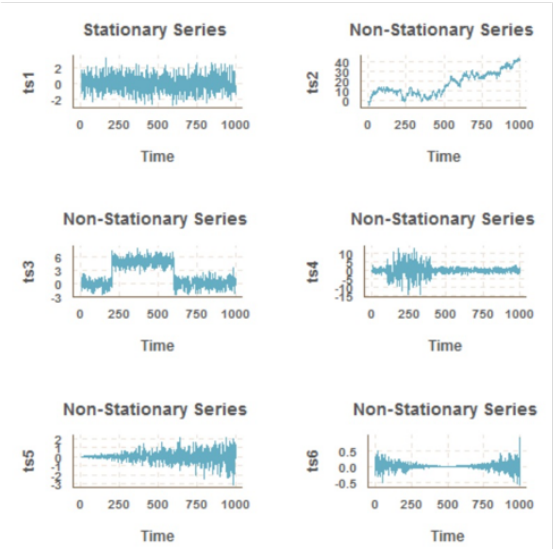
Predict moments for alternate boundary/initial conditions

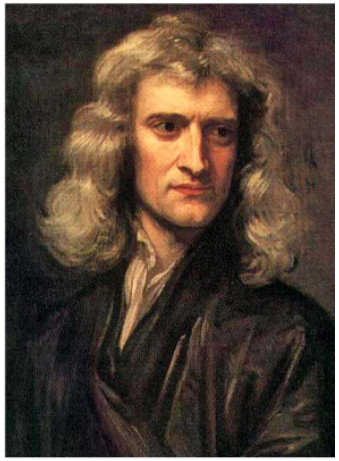






Characterization of the problem





Problem statement

1. Time series $\mathcal{Y}_n$: noisy, irregularly timed, partially informative data
2. Learn model parameters θ

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

3. Make probabilistic predictions about the future or for different conditions

$$p(x(t)|\mathcal{Y}_n) = \int p(x(t) \mid \theta) p(\theta \mid \mathcal{Y}_n) d\theta$$

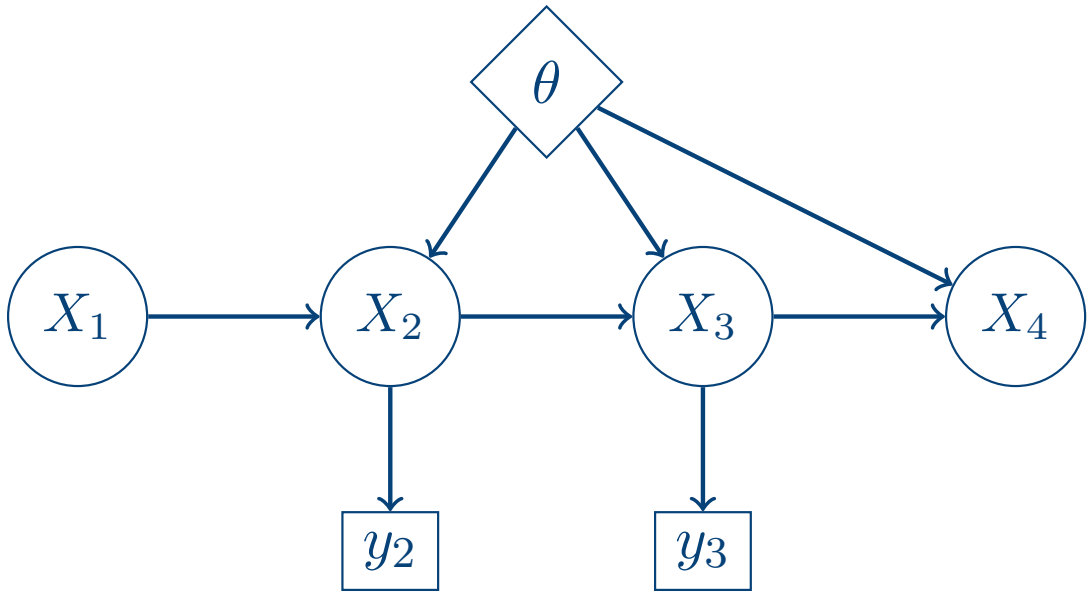
Probabilistic modeling framework
Joint parameter-state inference problem

UNIVERSITY OF MICHIGAN

Parameterized model Ψ

Generic formulation: linear, physical model, dictionary of bases, NN, etc.

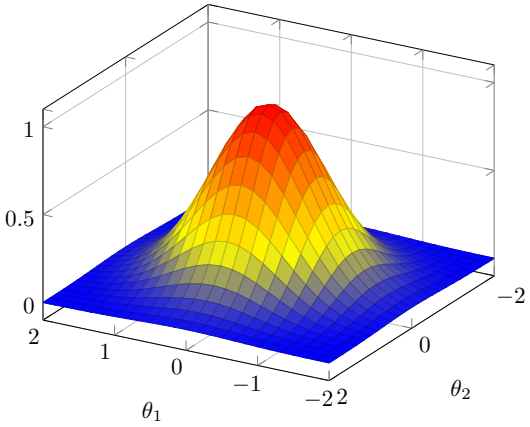
$$X_{k+1} = \Psi(X_k; \theta) + \eta_k$$
$$Y_k = h(X_k) + \nu_k,$$
$$\eta_k \sim \mathcal{N}(0, \Sigma)$$
$$\nu_k n \sim \mathcal{N}(0, \Gamma)$$



1. Accounts for model parameter uncertainty through θ
2. Accounts for model form uncertainty through η_k
3. Accounts for measurement noise through ν_k

Bayesian inference
Updating models with data

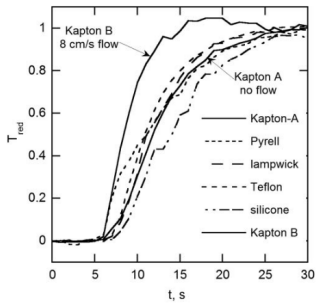
Prior $p(\theta, \mathcal{X}_n|I)$



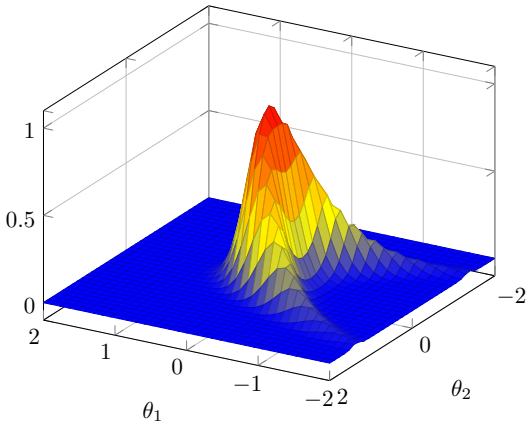
Bayes' Rule

$$p(\theta, \mathcal{X}_n|\mathcal{Y}_n, I) = p(\theta, \mathcal{X}_n \mid I)\frac{p(\mathcal{Y}_n|\theta, \mathcal{X}_n, I)}{p(\mathcal{Y}_n|I)}$$

Data



Posterior $p(\theta, \mathcal{X}_n|\mathcal{Y}_n, I)$



For our Markovian system

$$\underbrace{p(\theta, \mathcal{X}_n \mid \mathcal{Y}_n, I)}_{\text{posterior}} = \frac{1}{Z} \underbrace{\left[\prod_{i=1}^{n} p(y_i \mid X_i, \theta)\right]}_{\text{likelihood}} \underbrace{\left[\prod_{i=1}^{n} p(X_i \mid X_{i-1}, \theta)\right] p(\theta \mid I)}_{\text{prior}}$$

Marginalizing out the states
Pseudomarginal MCMC

UNIVERSITY OF MICHIGAN

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 Markov Chain Monte Carlo (DRAM)

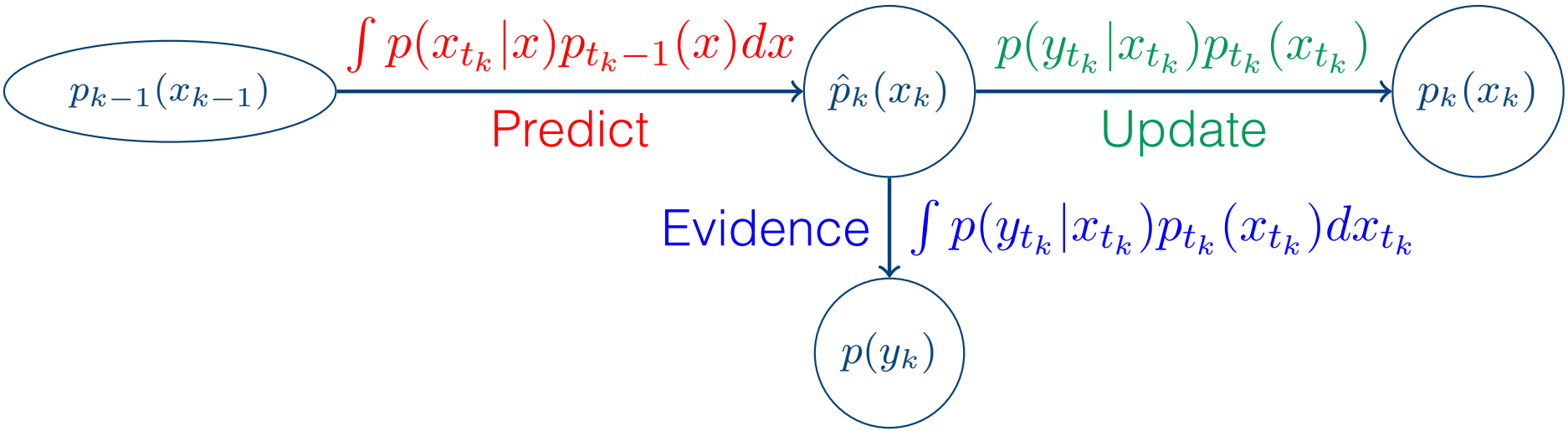
$$p(\theta \mid \mathcal{Y}_n, I) = \int p(\theta, \mathcal{X}_n \mid \mathcal{Y}_n, I) d\mathcal{X}_n = \frac{1}{Z} \underbrace{p(\mathcal{Y}_n \mid \theta)}_{\text{marginal likelihood}} p(\theta)$$

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

Evaluating the marginal posterior

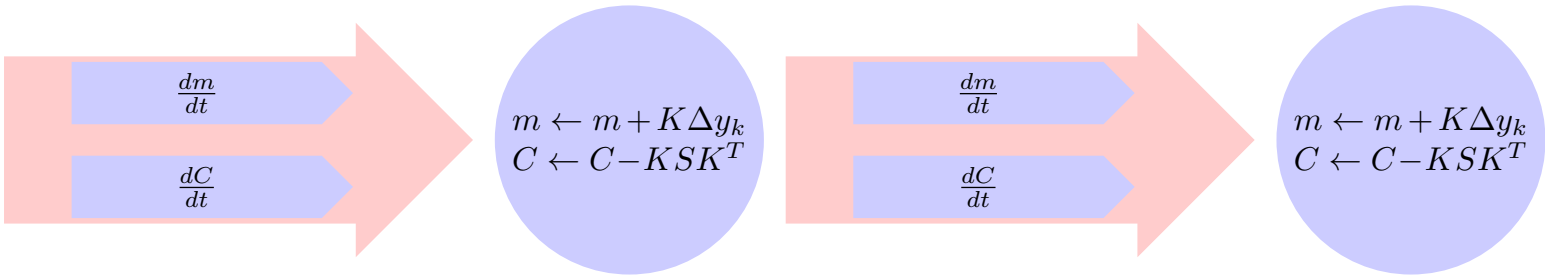
Recursive evaluation $p(\theta \mid \mathcal{Y}_n)$

1: **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 filter $p(X_k|\theta, \mathcal{Y}_k) = \frac{p(y_k|\theta, X_k)p(X_k|\theta, \mathcal{Y}_{k-1})}{p(y_k|\theta, \mathcal{Y}_{k-1})}$
5: Update posterior $p(\theta|\mathcal{Y}_k) = \frac{p(y_k|\theta, \mathcal{Y}_{k-1})p(\theta|\mathcal{Y}_{k-1})}{p(y_k|\mathcal{Y}_{k-1})}$
6: **end for**



State estimation and Bayesian Filtering

Gaussian filters propagate the mean and variance



- Exact filters
  - Kalman filter - only optimal for linear problems
  - Fokker-Planck + Bayes rule
- Approximate filters
  - Extended Kalman filter - accuracy suffers for nonlinear problems
  - Nonlinear Gaussian filters (ensemble, unscented, cubature)
  - Particle filters

Dynamic mode decomposition DMD
Linear mappings in observable space (Koopman)

1. Shift observations

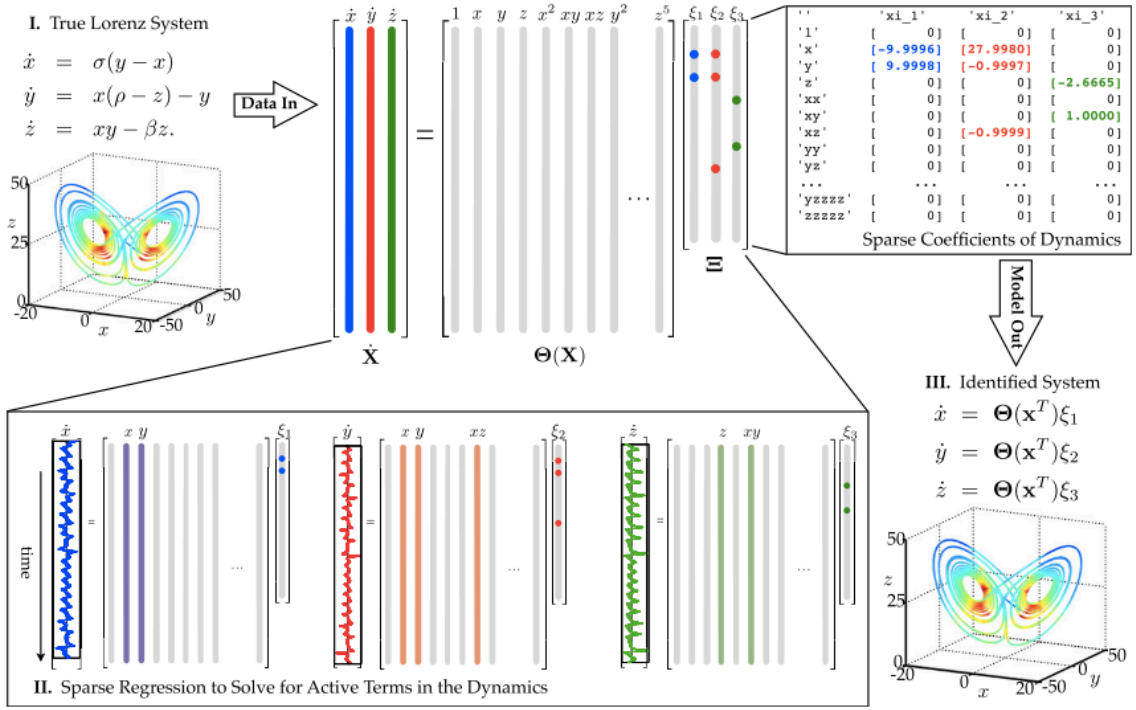
$$Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_{m-1} \end{bmatrix}$$
$$Y' = \begin{bmatrix} y_2 & y_3 & \cdots & y_m \end{bmatrix}.$$

2. Seek $Y' = AY$
3. Solve the optimization problem

$$A = \arg\min_{\tilde{A}} \sum_{k=2}^{N} \|y_k - \tilde{A} y_{k-1}\|^2. \tag{1}$$

Schmid, P. J. “Dynamic Mode Decomposition of Numerical and Experimental Data.” Journal of Fluid Mechanics, vol. 656, 2010, pp. 5–28.

Sparse identification of nonlinear systems (SINDy)
Seeking parsimonious models



Approximate the time derivative via finite differences

$$\theta = \arg\min_{\tilde{\theta}} \sum_{i=2}^{N} \| \frac{y_i - y_{i-1}}{dt} - \phi(y_{i-1})\tilde{\theta}) \|^2 + \lambda \|\tilde{\theta}\|_1$$

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

Special case: noiseless, invertible measurements
$\rightarrow$ DMD and SINDy

UNIVERSITY OF MICHIGAN

Full Negative Log Posterior

$$-Lp(\theta, \mathcal{X}_n \mid \mathcal{Y}_n) \propto \sum_{i=1}^{n} \|y_i - h(X_i)\|_{\Gamma}^2 + \sum_{i=2}^{N} \|X_i - \Psi(X_{i-1}, \theta)\|_{\Sigma}^2 - \log p(\theta)$$

MLE $\leftrightarrow$ DMD estimator

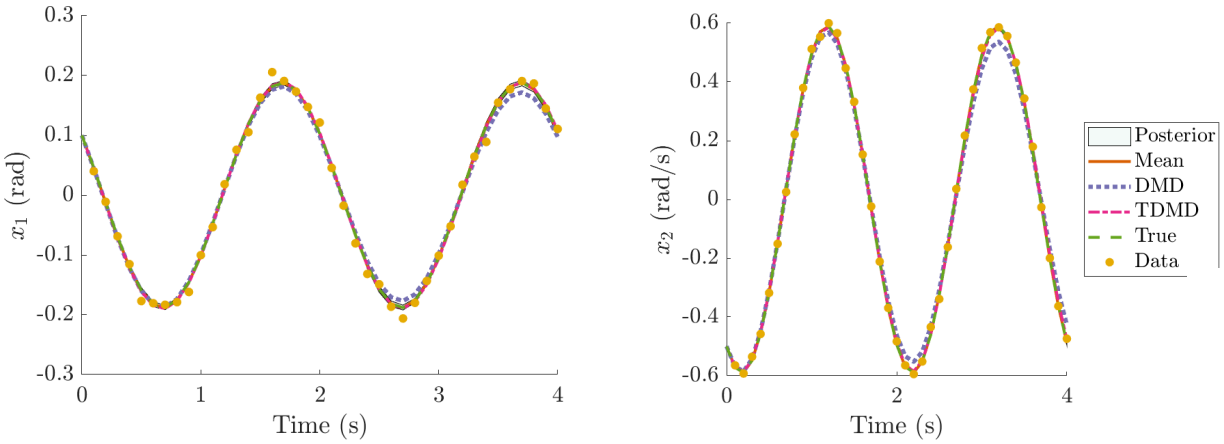
- linear model $\Psi(X_k, \theta) = \theta X_k$
- identity observation $h = I$
- noiseless measurements $\Gamma = 0$
- identity process noise $\Sigma = I$

MAP $\leftrightarrow$ SINDy

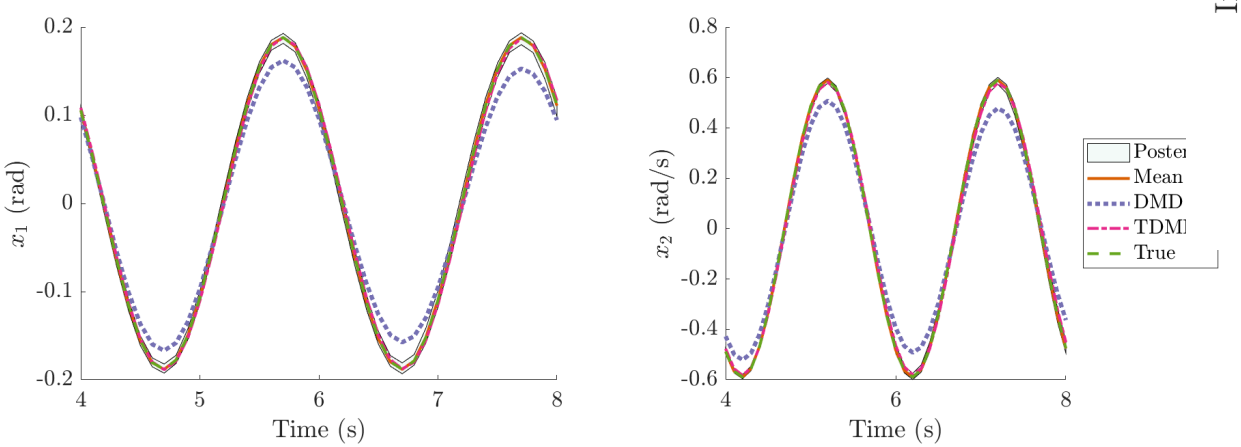
- linear-subspace model $\phi(x)\theta$
- identity observation $h = I$
- noiseless measurements $\Gamma = 0$
- identity process noise $\Sigma = I$
- a time-stepping scheme

Pendulum
Frequent, low-noise measurements

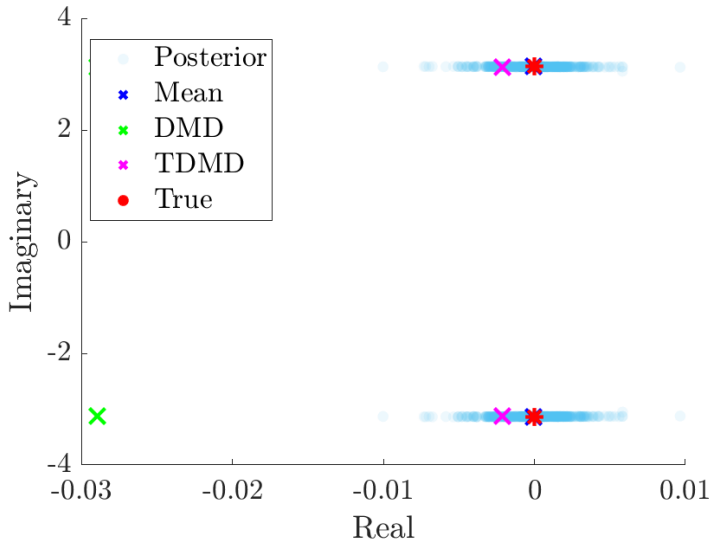
Reconstruction



Prediction

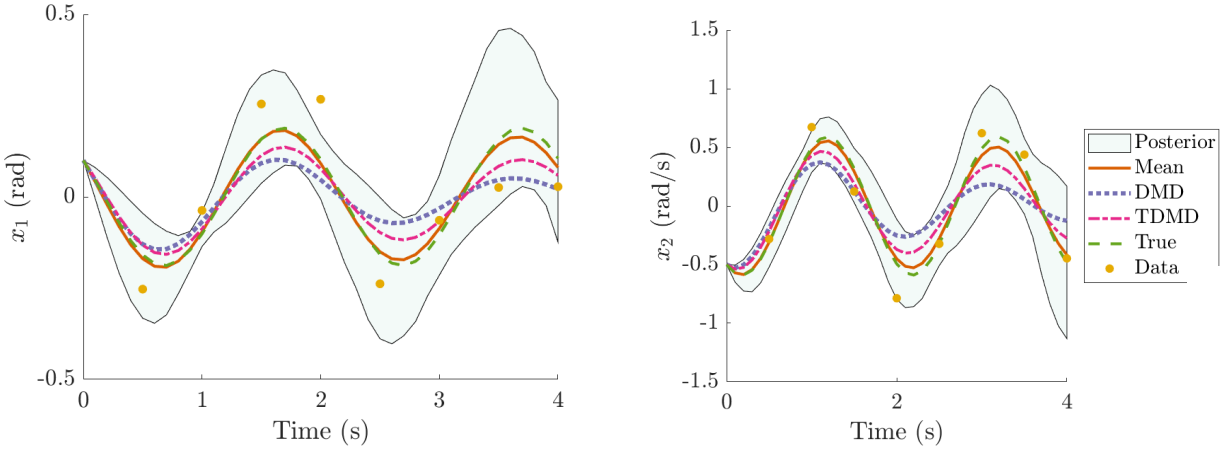


Eigenvalues

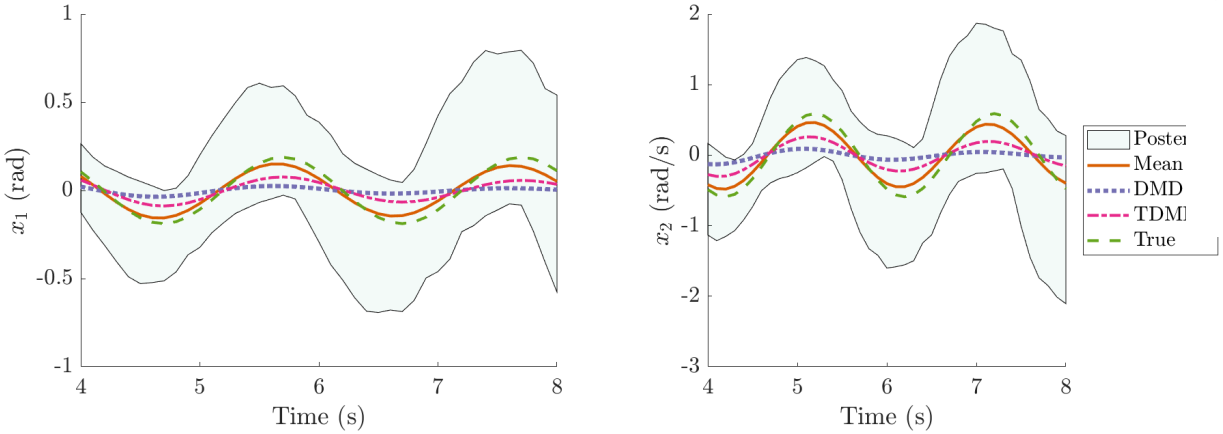


Pendulum
Sparse, higher-noise measurements

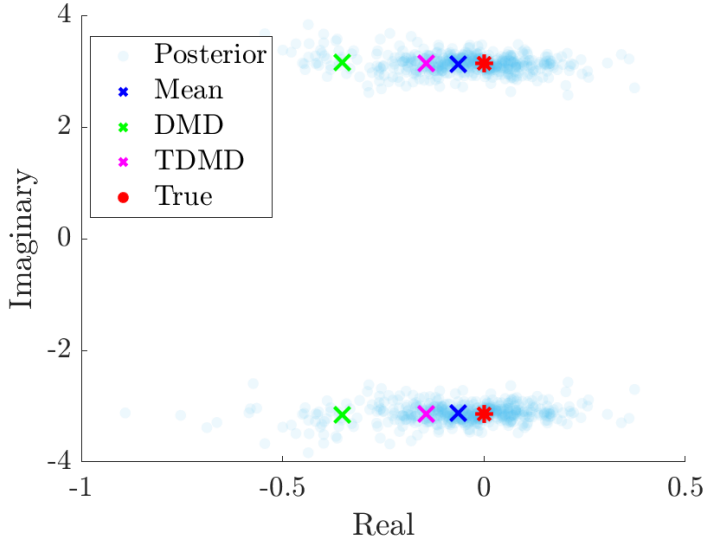
Reconstruction



Prediction



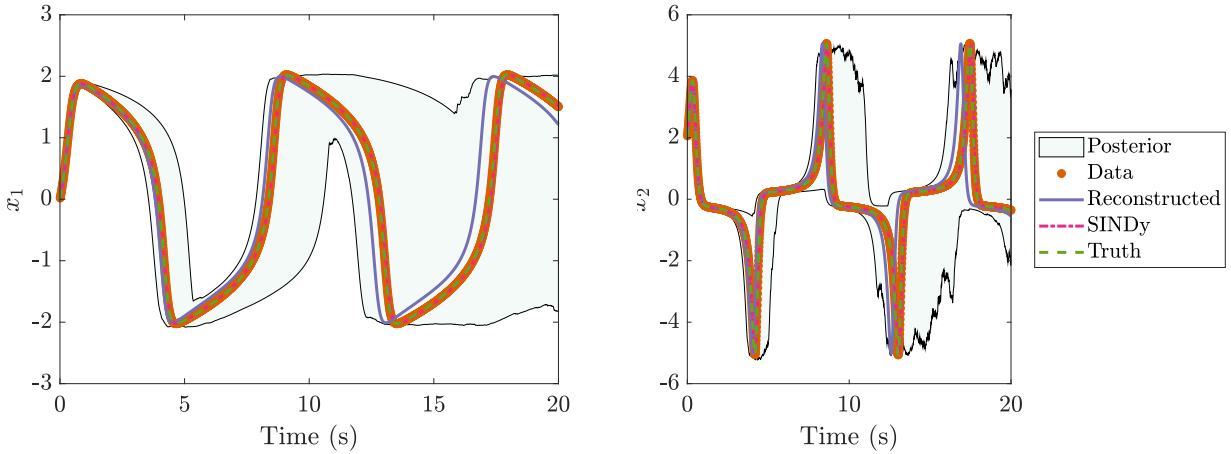
Eigenvalues



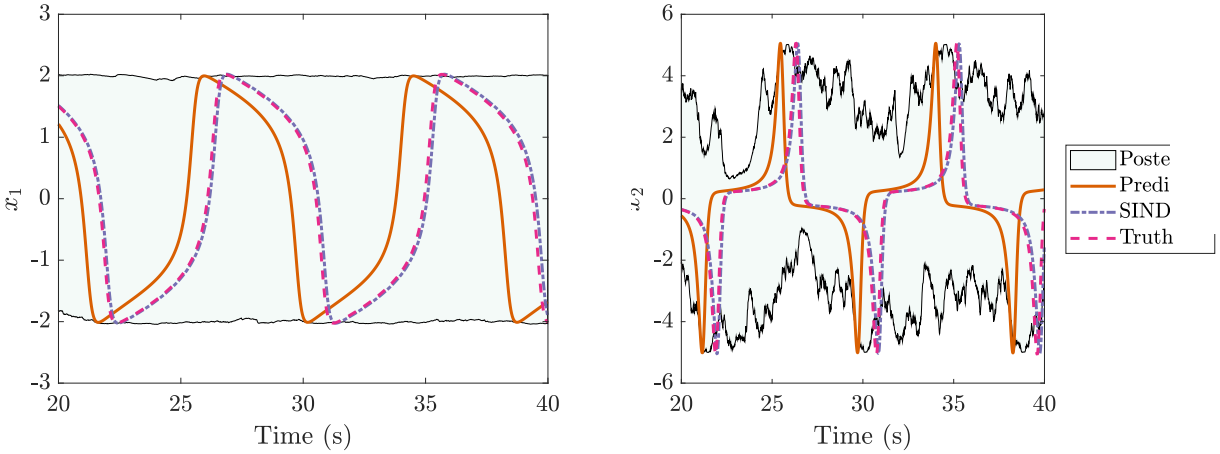
Vanderpol system: 2D nonlinear oscillator
Frequent, low-noise measurements

UNIVERSITY OF MICHIGAN

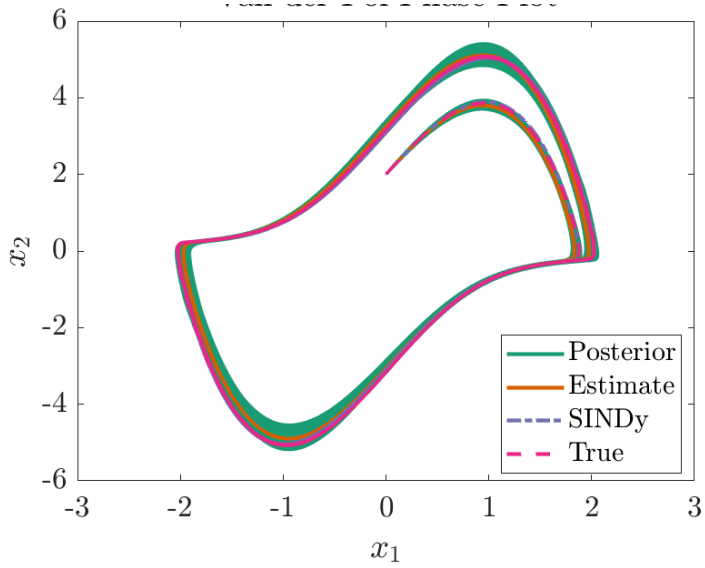
Reconstruction



Prediction



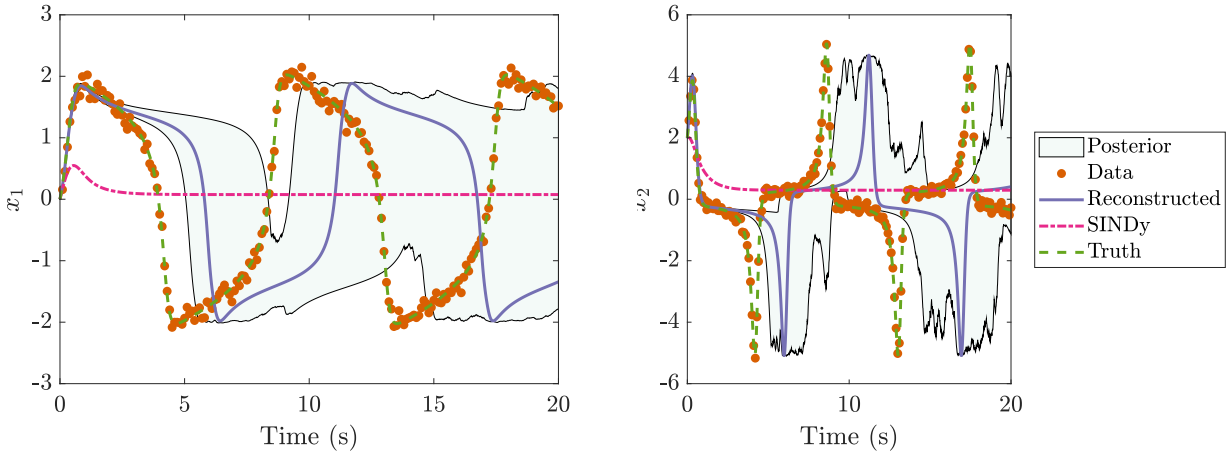
Phase



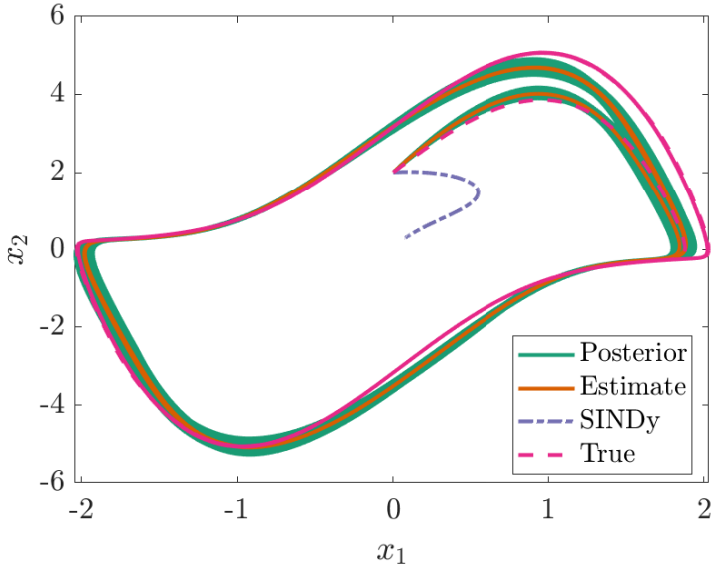
Vanderpol system: 2D nonlinear oscillator
Sparser, higher-noise measurements

UNIVERSITY OF MICHIGAN

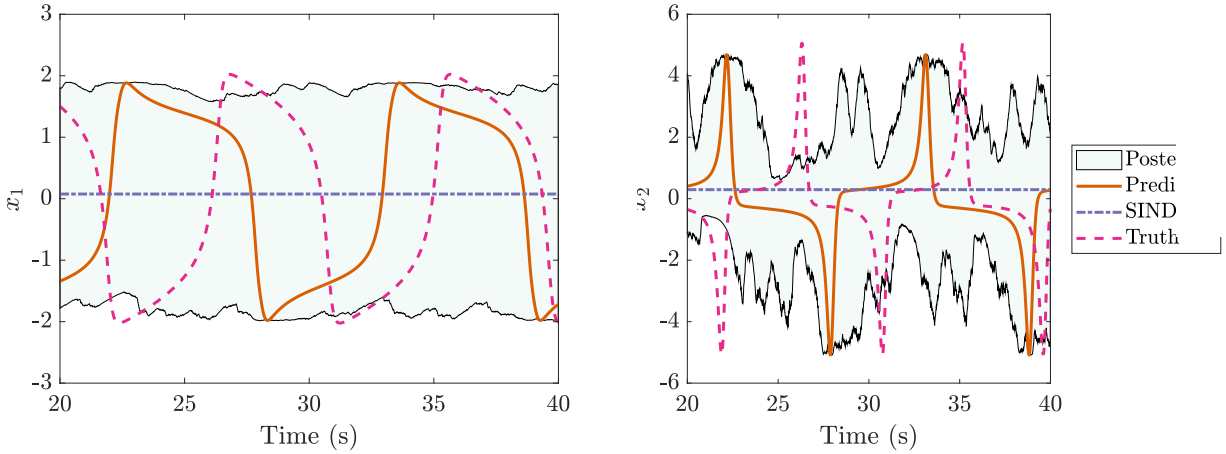
Reconstruction



Phase

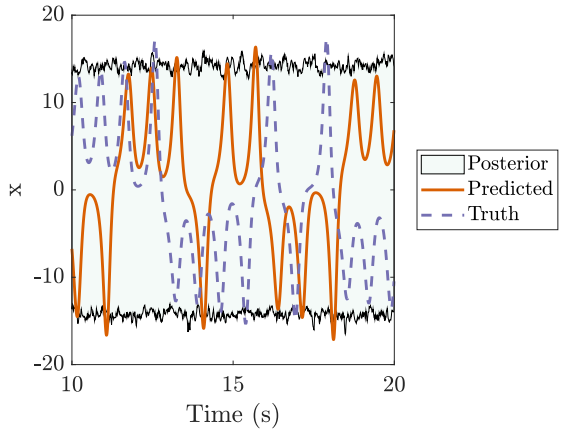


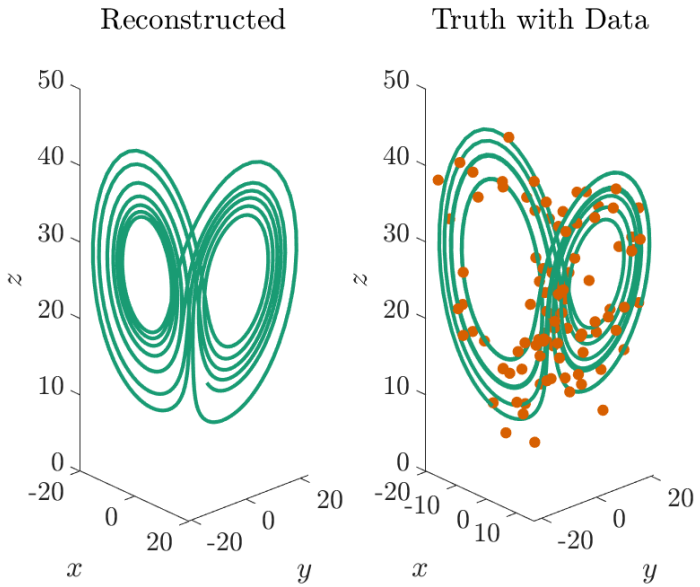
Prediction

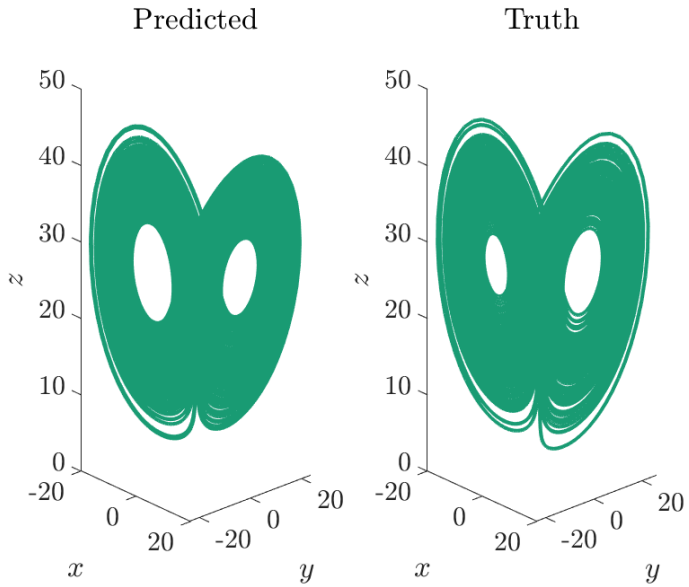


Lorenz 63
With only 300 data points, we recover the attractor

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







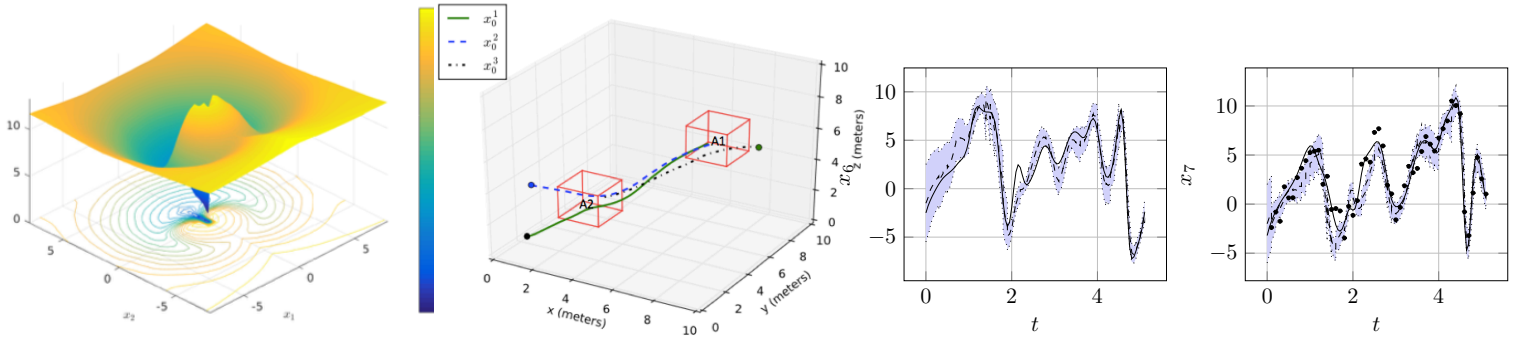
Conclusions and Acknowledgements

UNIVERSITY OF MICHIGAN

1. Conclusions
   - Probabilistic models provide flexible and capable frameworks for system ID
     - Parameter uncertainty
     - Model uncertainty
     - Measurement uncertainty
   - MCMC approaches are feasible for moderately large problems
   - Significant gains can be achieved under non-optimal conditions over DMD and SINDy
   - Please see Nick Galioto's poster for more details
2. Funding: DARPA PAI, AIRA

Computational science for autonomy

Compression enabled control and estimation



Real-time autonomy

- Source code: `github.com/goroda`, papers: `alexgorodetsky.com`
- Thanks!