A New Objective Function for Identification of Partially Observed LTI Dynamical Systems from Input-Output Data

Nicholas Galioto (ngalioto@umich.edu) and Alex Gorodetsky

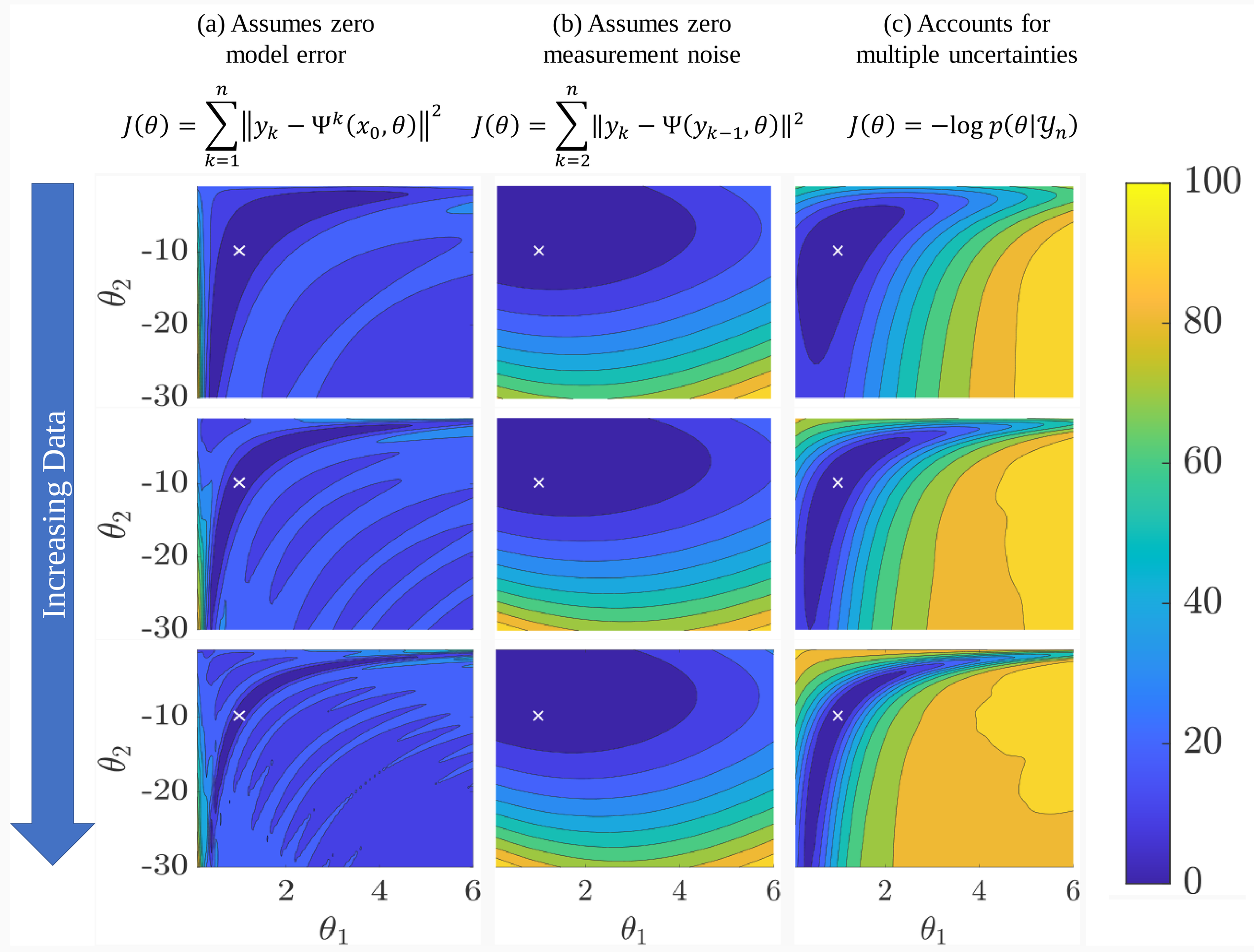
University of Michigan

Motivation

Many existing objective functions for system identification face the following challenges:

- They do not consider the existence and/or interaction of the three primary sources of uncertainty: (1) parameter, (2) model, and (3) measurement [GG20]
- As a result, they struggle to find good estimates when the data are noisy and/or sparse

The challenges and benefits of neglecting and including all three sources of uncertainty in the objective function are shown below



The table below displays key observations.

	(a)	(b)	(c)
Steep optimization surfaces without plateaus	✓	✗	✓
Suppresses local minima	✗	✓	✓
Increased confidence with data	✓	✗	✓

Our contributions include:

- A new objective function for system ID that differs from existing Bayesian approaches by using a stochastic dynamics to account for model error
- Empirical evidence that our method yields greater accuracy and precision compared to a least squares-based method, even when paired with the more robust ERA [JP85]

Least Squares Estimation of Markov Parameters + ERA

Given system (2), the input-output equation is written as

$$\mathbf{y}_k = \mathbf{C}\mathbf{A}^k x_0 + \sum_{i=1}^{k} \mathbf{C}\mathbf{A}^{i-1}\mathbf{B}u_{k-i} + \sum_{i=1}^{k} \mathbf{C}\mathbf{A}\xi_{k-i} + \eta_k$$

Denote the Markov parameter at time k as $\mathbf{g}_k = \mathbf{C}\mathbf{A}^{k-1}\mathbf{B}$. Many works [OO19; SRD19; Fat20] that consider $\mathbf{x}_0 = \mathbf{0}$ and zero-mean inputs use the following objective function

$$\hat{\mathbf{G}} = \arg\min_{\mathbf{G}} \sum_{i=0}^{n} \|\mathbf{y}_i - \mathbf{G}\bar{\mathbf{u}}_i\|_2^2, \tag{1}$$

where $\bar{\mathbf{u}}_i = \left[u_{i-1}\ u_{i-2} \cdots u_{i-T}\right]^*$ and $\mathbf{G} = \left[\mathbf{g}_1\ \mathbf{g}_2 \cdots \mathbf{g}_n\right]$.

Two important observations:

- The variance of $\mathbf{y}_k$ grows with k due to the term $\sum_{i=1}^{k} \mathbf{C}\mathbf{A}\xi_{k-i}$.
- Objective (1) effectively assumes constant covariance and is therefore suboptimal.

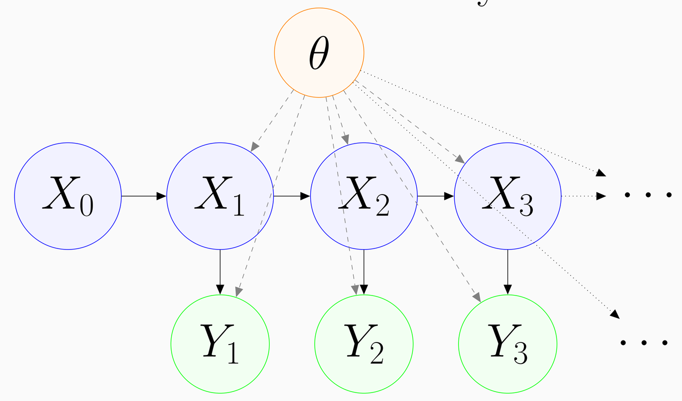
The estimated Markov parameters are then fed into the ERA

Probabilistic Formulation and Resulting Objective

We consider linear time-invariant models and assume the uncertainty is additive Gaussian

$$\begin{aligned} X_{k+1} &= \mathbf{A}X_k + \mathbf{B}u_k + \xi_k, & \xi_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \\ Y_k &= \mathbf{C}X_k + \eta_k, & \eta_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}), \end{aligned} \tag{2}$$

The interaction of the uncertainty can be visualized through the Bayesian network



1. Measurement uncertainty
2. Model uncertainty
3. Measurement uncertainty

Goal: Evaluate the posterior distribution $p(\Theta \mid \mathcal{Y}_n)$

$$\text{Bayes' rule: } p(\Theta \mid \mathcal{Y}_n) = \frac{\mathcal{L}(\Theta; \mathcal{Y}_n) p(\Theta)}{p(\mathcal{Y}_n)}, \quad \text{where } \mathcal{L}(\Theta; \mathcal{Y}_n) := p(\mathcal{Y}_n \mid \Theta)$$

In order to obtain the marginal likelihood $\mathcal{L}(\Theta; \mathcal{Y}_n)$, we must marginalize out the states

$$\mathcal{L}(\Theta; \mathcal{Y}_n) = \int \mathcal{L}(\Theta, \mathcal{X}_n \mid \mathcal{Y}_n) d\mathcal{X}_n. \tag{3}$$

We use the MAP optimization objective to obtain our model estimate

$$\Theta^{MAP} = \arg\max_{\Theta} \log \mathcal{L}(\Theta; \mathcal{Y}_n) + \log p(\theta).$$

Algorithm

Theorem 1 (Marginal likelihood (Th. 12.1 [Sär13])) *Let $\mathcal{Y}_k \equiv \{y_i; i \leq k\}$ denote the set of all observations up to time k. Let the initial condition be uncertain with distribution $p(X_0 \mid \Theta)$. Then the marginal likelihood (3) is defined as $\mathcal{L}(\Theta \mid \mathcal{Y}_n) = \prod_{k=1}^{n} \mathcal{L}_k(\Theta \mid \mathcal{Y}_k)$, where $\mathcal{L}_k(\Theta \mid \mathcal{Y}_k)$ is computed recursively in three stages for $k = 1, 2, \ldots$: prediction*

$$p(X_{k+1} \mid \Theta, \mathcal{Y}_k) = \int \frac{\exp\left(-\frac{1}{2}\|X_{k+1} - \mathbf{A}X_k - \mathbf{B}u_k\|_{\mathbf{\Sigma}}^2\right)}{\sqrt{2\pi}^{d_x} |\mathbf{\Sigma}|^{\frac{1}{2}}} p(X_k \mid \Theta, \mathcal{Y}_k) dX_k \tag{4}$$

update,

$$p\left(X_{k+1} \mid \Theta, \mathcal{Y}_{k+1}\right) = p(X_{k+1} \mid \Theta, \mathcal{Y}_k) \frac{\exp\left(-\frac{1}{2}\|\mathbf{y}_{k+1} - \mathbf{C}X_{k+1}\|_{\mathbf{\Gamma}}^2\right)}{\sqrt{2\pi}^{d_y} |\mathbf{\Gamma}|^{\frac{1}{2}} p(Y_{k+1} \mid \Theta, \mathcal{Y}_k)} \tag{5}$$

and marginalization,

$$\mathcal{L}_{k+1}(\Theta \mid \mathcal{Y}_{k+1}) = \int p(X_{k+1} \mid \Theta, \mathcal{Y}_k) \frac{\exp\left(-\frac{1}{2}\|\mathbf{y}_{k+1} - \mathbf{C}X_{k+1}\|_{\mathbf{\Gamma}}^2\right)}{\sqrt{2\pi}^{d_y} |\mathbf{\Gamma}|^{\frac{1}{2}}} dX_{k+1}. \tag{6}$$

For linear-Gaussian systems such as (2), the Kalman filter is used to compute the above three distributions.

Funding and References

AFOSR Program in Computational Mathematics

References

[Fat20] Salar Fattahi. "Learning Partially Observed Linear Dynamical Systems from Logarithmic Number of Samples". In: *arXiv preprint arXiv:2010.04015* (2020).

[GG20] Nicholas Galioto and Alex Arkady Gorodetsky. "Bayesian system ID: optimal management of parameter, model, and measurement uncertainty". In: *Nonlinear Dynamics* 102 (2020), pp. 241–267. ISSN: 1573-269X. DOI: 10.1007/s11071-020-05925-8.

[JP85] Jer-Nan Juang and Richard S Pappa. "An eigensystem realization algorithm for modal parameter identification and model reduction". In: *Journal of guidance, control, and dynamics* 8.5 (1985), pp. 620–627.

[OO19] Samet Oymak and Necmiye Ozay. "Non-asymptotic identification of lti systems from a single trajectory". In: *2019 American Control Conference (ACC)*. IEEE. 2019, pp. 5655–5661.

[Sär13] Simo Särkkä. *Bayesian filtering and smoothing*. Vol. 3. Cambridge University Press, 2013.

[SRD19] Tuhin Sarkar, Alexander Rakhlin, and Munther A Dahleh. "Finite-time system identification for partially observed lti systems of unknown order". In: *arXiv preprint arXiv:1902.01848* (2019).

Results

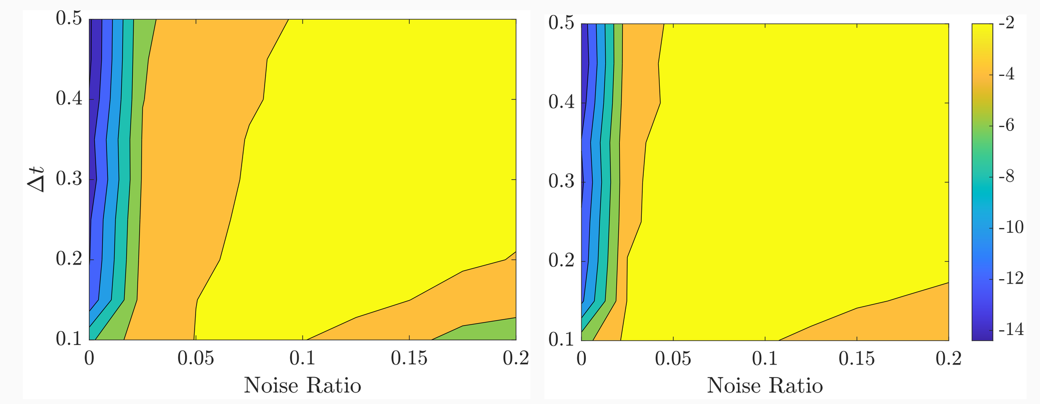
Linear pendulum with periodic forcing

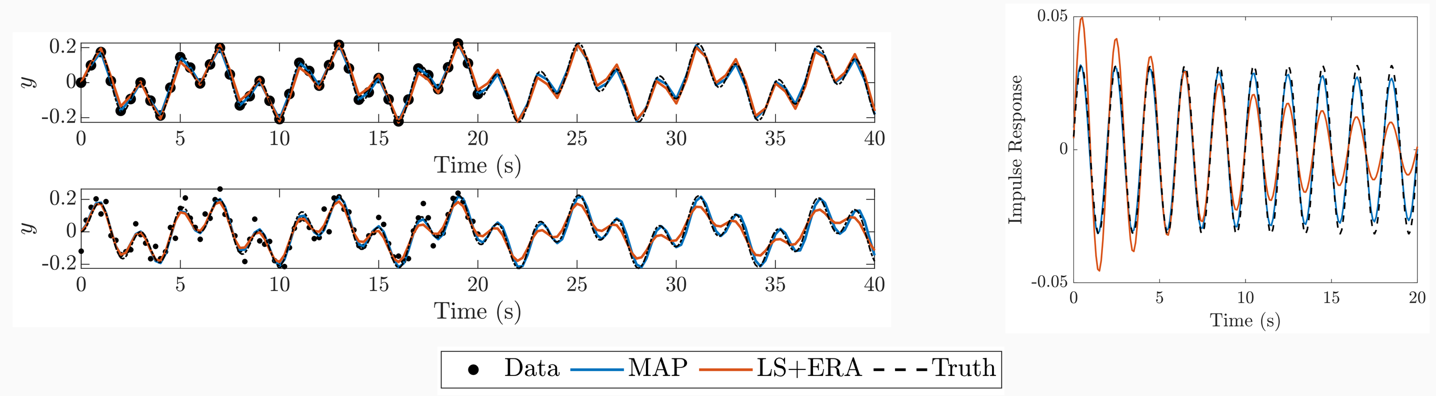
- Our method achieves orders of magnitude gains in performance
- The MAP estimate remains robust when data are noisy and/or sparse
- Due to inclusion of the process noise term, the estimate is not overfit

$$\begin{bmatrix} \dot{\phi} \\ \ddot{\phi} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -g/L & 0 \end{bmatrix} \begin{bmatrix} \phi \\ \dot{\phi} \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \cos(t), \qquad y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \phi_k \\ \dot{\phi}_k \end{bmatrix},$$

The contour plots metric is defined below

$$\log\left(\frac{\sum_{i=1}^{100} MSE_i^{MAP}}{\sum_{i=1}^{100} MSE_i^{LS+ERA}}\right), \quad \text{where } MSE_i = \frac{1}{n}\sum_{k=1}^{n}(\phi_k - \hat{\mathbf{y}}_k)^2,$$

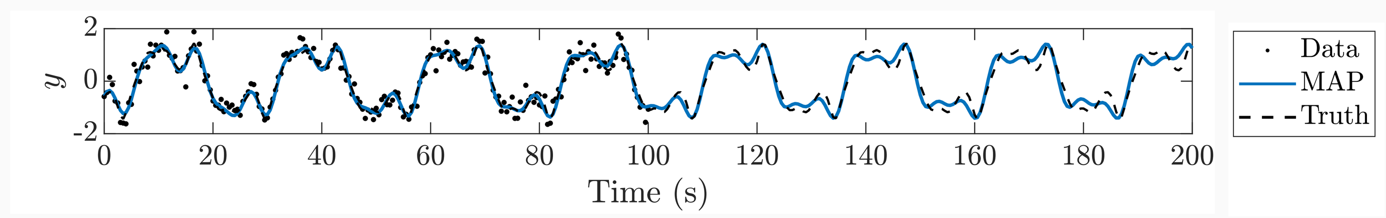




Periodic limit cycle of forced Duffing oscillator

- We are able to learn non-zero initial conditions
- Our method can be applied to nonlinear systems

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \alpha & \delta \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} + \beta \begin{bmatrix} 0 \\ x^3 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \gamma \cos(\omega t), \qquad y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ \dot{x}_k \end{bmatrix}.$$



Conclusion

The main takeaways of our research are

- Accounting for parameter, model, and measurement uncertainties and their interactions in the objective function yields more accurate estimates that are more robust to sparse/noisy data
- Using a stochastic dynamics model, even for deterministic systems, has a regularizing effect that can prevent overfitting

LaTeX TikZposter