Sparse Bayesian System Identification for Dynamical Systems with Neuronized Priors

Prem R. Mohan Ram * **Ulrich Römer** ** **Richard Semaan** ***

* *Institut für Dynamik und Schwingungen, Technische Universität Braunschweig, Germany (e-mail: p.mohan-ram@tu-braunschweig.de).*
** *Institut für Dynamik und Schwingungen, Technische Universität Braunschweig, Germany (e-mail: u.roemer@tu-braunschweig.de).*
*** *Institut für Strömungsmechanik, Technische Universität Braunschweig, Germany (e-mail: r.semaan@tu-braunschweig.de).*

Abstract: This work is concerned with learning the dynamics of technical systems from data within a sparse Bayesian framework. The approach employs a basis representation of the unknown dynamics function, similar to the sparse identification of nonlinear dynamics (SINDy) approach, which is combined with a Bayesian procedure for parameter estimation. We propose to use the recently introduced neuronized priors as a unified approach to enforce sparsity in a dynamical systems context, and illustrate the method with an academic example.

Keywords: Uncertainty, dynamic systems, parameter estimation, parameter identification, probabilistic simulation.

1. INTRODUCTION

Learning the dynamics of a system from data is receiving considerable attention at present. Here, we focus on autonomous systems of the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)), \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{1}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$, $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n, t \in (0, T]$ and $\dot{\mathbf{x}} = \mathrm{d}\mathbf{x}/\mathrm{d}t$. One particularly popular paradigm, the sparse identification of nonlinear dynamics (SINDy) Brunton et al. (2016a), represents the unkonwn dynamics $\mathbf{f}$ as a linear combination of library basis functions as

$$f_i(\mathbf{x}) \approx \boldsymbol{\Theta}(\mathbf{x})\boldsymbol{\xi}_i, \tag{2}$$

where $\boldsymbol{\xi}_i \in \mathbb{R}^p$, $i = 1, \ldots, n$ are the parameters to be identified and $\boldsymbol{\Theta}$ represents the library of basis functions. For notational convenience, all parameter vectors are collected in a matrix as $\boldsymbol{\Xi} = [\boldsymbol{\xi}_1 \cdots \boldsymbol{\xi}_n]$. Popular choices of basis functions are polynomials or splines, which are able to accurately approximate large classes of functions. In the original paper Brunton et al. (2016a), regression with iterative thresholding was used to estimate the parameters from data of $\dot{\mathbf{x}}(t)$. Since then, extensions in many different directions have been proposed. Control scenarios were addressed in Brunton et al. (2016b), whereas uncertainty was included through a Bayesian extension in Fuentes et al. (2021). Another Bayesian approach, also including noisy observations of the state directly, was introduced in Galioto and Gorodetsky (2020), which contained the original SINDy approach as a special case. Also, there exist deep learning based approaches, such as Goyal and Benner (2021) to learn the model as a black box. In contrast, we aim for a white-box model for which a bases library approach is better suited.

Despite these contributions, several challenges remain. Enforcing sparsity in a Bayesian framework is challenging and may require to work with complicated prior formulations. Also, jointly handling all sources of uncertainty goes beyond the linear regression setting and the computational complexity will grow quickly. Here, we report and extend on our work pre-published in Ram et al. (2021).

2. BAYESIAN SYSTEM IDENTIFICATION

In this section we present our approach to estimate the unknown parameters $\boldsymbol{\Xi}$, while simultaneously accounting for observation, process and model uncertainty. Therefore, we apply a suitable approximation method in time, an explicit Runge Kutta method for the sake of simplicity, which yields

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \Psi_h(\boldsymbol{\Theta}(\mathbf{x}_i)\boldsymbol{\Xi}), \quad \mathbf{x}_0 = \mathbf{x}_0, \tag{3}$$

where $i = 0, \ldots, m-1$, $\mathbf{x}_i \approx \mathbf{x}(t_i)$ and Ψ_h specifies the discrete Runge Kutta time propagator on the time grid with uniform grid size h. Then, we introduce the stochastic state space model

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \Psi_h(\boldsymbol{\Theta}(\mathbf{x}_i)\boldsymbol{\Xi}) + \boldsymbol{\eta}_i, \tag{4}$$
$$\mathbf{y}_j = \mathbf{x}_j + \boldsymbol{\varepsilon}_j, \tag{5}$$

where $j = 1, \ldots, k$ indicates the observation time. Note that the considered example employs the full state, which is why we employ an observation of the full state in (5). However, the method can easily be extended to cover more general state-to-observation maps. Also, the process and observation noise are assumed to be distributed as $\boldsymbol{\eta}_i \sim \mathcal{N}(0, \sigma_\eta^2 \mathbb{I})$ and $\boldsymbol{\varepsilon}_j \sim \mathcal{N}(0, \sigma_\varepsilon^2 \mathbb{I})$, respectively. This setting is close to the one considered in Galioto and Gorodetsky

⋆ The authors acknowledge internal funding by the faculty of mechanical engineering at Technische Universität Braunschweig.

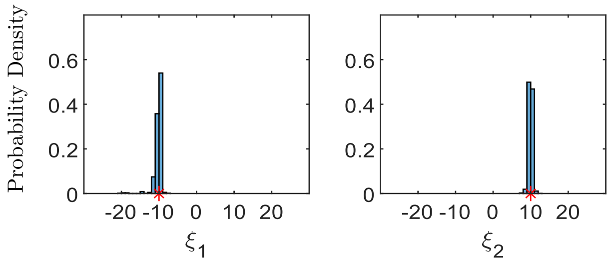

Fig. 1. Posterior histograms of ξ_i and true values (red stars).

(2020), where the joint distribution $p(\mathbf{\Xi}, \mathbf{X})$, with $\boldsymbol{X}$ collecting all states $\boldsymbol{x}_i$ into a matrix, was inferred with a Kalman-type filter. Because we are mainly interested in the marginal distribution

$$p(\mathbf{\Xi}) = \int p(\mathbf{\Xi}, \mathbf{X})\ \mathrm{d}\mathbf{X}, \tag{6}$$

we pursue a different path and focus on updating the parameters directly. We employ an ensemble of N states, propagated in time according to $p(\mathbf{x}_{i+1}|\mathbf{x}_i)$ and base inference on the average Likelihood

$$\hat{L}(\mathbf{Y}|\mathbf{\Xi}) = \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{k}\frac{1}{\sqrt{(2\pi)^k\sigma_\varepsilon^2}}\exp\left(-\frac{|\mathbf{y}_j - \mathbf{x}_j^{(i)}|^2}{2\sigma_\varepsilon^2}\right), \tag{7}$$

which ensures robustness against process noise, see also Conrad et al. (2017). We then update the posterior distribution with Bayes' law as

$$p(\mathbf{\Xi}|\mathbf{Y}) \propto \hat{L}(\mathbf{Y}|\mathbf{\Xi})p(\mathbf{\Xi}), \tag{8}$$

which we sample with Markov Chain Monte Carlo methods. In addition to employing the average Likelihood, another original contribution is the use of a generalized formulation of sparsity priors given as

$$\xi_i = T(\alpha_i - \alpha_0)w_i, \tag{9}$$

where $w_i \sim \mathcal{N}(0, \tau_w)$, $\alpha_i \sim \mathcal{N}(0, 1)$ and T is an activation function from neural network methods, which motivates the name neuronized prior, see Shin and Liu (2021). Through different choices for T we can recover various priors, such as Lasso, Horseshoe and Spike and Slab priors. The case of spike and slab prior is of particular interest, because it allows to obtain zero inclusion probabilities of individual basis functions in the library and hence, model selection can be carried out as well.

3. NUMERICAL EXAMPLE

Here, we present an application to estimate the coefficients of the first equation of the Lorenz system

$$\begin{aligned}\dot{x}_1 &= c_1(x_2 - x_1),\\ \dot{x}_2 &= 28x_1 - x_1x_3 - x_2,\\ \dot{x}_3 &= x_1x_2 - 2.67x_3.\end{aligned} \tag{10}$$

First, c_1 is set to a value of 10, and data $\boldsymbol{y}_j$ is simulated by propagating the initial state $[x_1^0, x_2^0, x_3^0]^\top = [-8, 8, 27]^\top$ using MATLAB's `ODE23` solver. Observation noise with $\sigma_\varepsilon = 0.01$ is added to generate the data. A total simulation time of 1.5 units, with $h = 10^{-4}$ is considered.

For the above application, a basis library $\Theta = [x_1,\ x_2,\ x_3,\ x_1x_2,\ x_1x_3,\ x_2x_3]$ is employed, and corresponding coefficients $\xi_1, \xi_2, \ldots, \xi_6$ are to be estimated. The standard deviation of the process noise σ_η is calibrated with an empirical Bayes approach, as outlined in Conrad et al. (2017), which leads to $\sigma_\eta = 256$ in the current case.

To obtain the posterior given by (8), an Affine Invariant Ensemble MCMC sampler (AIES) is employed. To facilitate straightforward model selection, the ReLu activation function $T(\alpha_i - \alpha_0) = \max(0, \alpha_i - \alpha_0)$ is chosen. With the help of a grid search, the neuronised prior's hyperparameters α_0 and τ_w are assigned values -0.25 and 0.1, respectively. A burn-in of 50% is considered and 75% of the AIES walkers are discarded as bad chains.

The resulting posterior histograms for the first two coefficients are plotted in Figure 1. It can be seen that the posteriors for both the coefficients are centered around their true values denoted by red stars. Also, the resulting posteriors for the coefficients $\xi_3, \xi_4, \ldots, \xi_6$ yield $P(\xi_i = 0) > 0.5$. Hence, a median model selection would remove those ξ_i for which $P(\xi_i = 0) > 0.5$ and successfully recover the original Lorenz system. The histograms for the coefficients $\xi_3, \xi_4, \ldots, \xi_6$ haven't been presented here as they resemble concentrated spikes at zero, with negligible spread.

We shall investigate the performance of the method with different, more complex, examples in the future.

REFERENCES

Brunton, S.L., Proctor, J.L., and Kutz, J.N. (2016a). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15), 3932–3937.

Brunton, S.L., Proctor, J.L., and Kutz, J.N. (2016b). Sparse identification of nonlinear dynamics with control (sindyc). *IFAC-PapersOnLine*, 49(18), 710–715.

Conrad, P.R., Girolami, M., Särkkä, S., Stuart, A., and Zygalakis, K. (2017). Statistical analysis of differential equations: introducing probability measures on numerical solutions. *Statistics and Computing*, 27(4), 1065–1082.

Fuentes, R., Nayek, R., Gardner, P., Dervilis, N., Rogers, T., Worden, K., and Cross, E. (2021). Equation discovery for nonlinear dynamical systems: A Bayesian viewpoint. *Mechanical Systems and Signal Processing*, 154, 107528.

Galioto, N. and Gorodetsky, A. (2020). Bayesian system id: optimal management of parameter, model, and measurement uncertainty. *Nonlinear Dynamics*, 102, 241–267.

Goyal, P. and Benner, P. (2021). Learning dynamics from noisy measurements using deep learning with a runge-kutta constraint. *arXiv preprint arXiv:2109.11446.*

Ram, P.R.M., Römer, U., and Semaan, R. (2021). Bayesian dynamical system identification with unified sparsity priors and model uncertainty. *arXiv preprint arXiv:2103.05090.*

Shin, M. and Liu, J.S. (2021). Neuronized priors for Bayesian sparse linear regression. *Journal of the American Statistical Association*, 1–43.