1st Conference on Applied AI and Scientific Machine Learning (CASML 2024)

Learning algorithms for non-linear dynamical systems, control and autonomy

Lakshman Mahto **lm.optlearning@gmail.com**
Indian Institute of Information Institute of Technology Dharwad and BIT Mesra, India

Abstract

In this work, we study and analyse state-of-the-art learning algorithms of non-linear dynamical systems, control and autonomy with scalable optimization algorithms and additional statistical assumptions in high dimensional framework. This work establish three learning algorithms, namely a. learning of a non-linear dynamical system, b. learning a Lyapunov function of the learned system, and c. learning control & decision making with improved generalization assumptions for generalization bounds and computational complexities. This work particularly involved around the norm, the criterion function & regularized term, the design matrix and the noise model. The advantages of the approach for learning algorithms, stabilization and control are illustrated with few results featuring robust formulations and regularization with kernel embeddings.

1. Introduction

A mathematical & statistical modelling of dynamical systems & control involve two alternative approaches:

(a.) a physics based laws and first principle derive a forward dynamical system and then extrapolating locally.

(b.) a data driven based constrained under physical observation derive an inverse modelling of a unknown dynamical system, and then finally interpolating globally.

Our emphasis is the second one, may be considered as an advancement of system identification (SI). Since, SI cannot incorporate physic based first principle information of the system (e.g. compact representation, multi-modality etc.), and computational intractability due to time dependencies, large sample size, high dimensional estimation & inference. Furthermore, reliance on analytic ability of least-squares (in low or moderate dimensional framework) leaves the system identification process vulnerable to model misspecification, including outliers in the observations. In summary, learning system dynamics with big data involves either high dimensional statistics or large sample theory for parameter estimation & generalization bounds..

Inspired by the emergence learning dynamical system (in the interface of computational mathematics & machine intelligence) and its applicability to autonomous control [1, 7, 8], in this paper, we aim at developing and generalizing learning algorithms of non-linear internal state space model based on statistical learning framework and kernel embedding of original state space to a suitable a RKHS. The main contributions of this paper are as follows:

(a.) mixing-time based IID approximation of dependency in data under one-point convexity and smooth loss function and bounded, Lipschitz, stable flow of the system.

(b.) capturing non-linearity with an adaptive kernel embedding.

(c.) regularized high dimensional parameter estimation.

(c.) generalization bounds and robust learning algorithms via uniform convergence of loss and its gradient.

2. Problem formulation and results

In this work, we learn a non-linear dynamical system $\dot{x} = f(x), f : \mathbb{R}^d \to \mathbb{R}^d$ with unknown vector field $f : \mathbb{R}^d \to \mathbb{R}^d$, from a limited information along a single trajectory (i.e. a noisy measurement or time series data as a realization of a random sample of size n.

$$\mathcal{S} := \left\{ z_i = (x_i, y_i) = (x(t_i), y(t_i)) | x(t_i) = x_i, y(t_i) = dotx(t_i) = y_i \in \mathbb{R}^d \right\}_{i=1}^{n}.$$

Now, in order to capture non-linearity in the dynamical system via a linear model framework, we introduce an explicit feature map embedding $\phi : \mathbb{X} \subset \mathbb{R}^d \to \mathbb{H}_\phi \subset \mathbb{R}^p, p = p(n, d)$ such that $f(x) \approx \Omega^\top \phi(x)$. Mathematically, we have

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_d(x) \end{pmatrix} \approx \begin{pmatrix} \omega_1^\top \phi(x) \\ \omega_2^\top \phi(x) \\ \vdots \\ \omega_d^\top \phi(x) \end{pmatrix} = \begin{pmatrix} -\omega_1^\top - \\ -\omega_2^\top - \\ \vdots \\ -\omega_d^\top - \end{pmatrix} \phi(x) = \Omega^\top \phi(x),\ \Omega \in \mathbb{R}^{p \times d}, \phi(x) \in \mathbb{R}^p.$$

Now,running the model $y = \dot{x} = f(x) \approx \Omega^\top \phi(x)$ over the given data S as

$$y_i \approx y(t_i) = \dot{x}(t_i) = f(x(t_i)) \approx f(x_i) \approx \Omega^\top \phi(x_i), \to y_i^\top = \phi^\top(x_i)\Omega + \epsilon_i^\top, i = 1, \dots, n \implies$$
$$\mathcal{Y} = \mathcal{X}\Omega + \Xi, \mathcal{Y} \in \mathbb{R}^{n \times d},\ \mathcal{X} \in \mathbb{R}^{n \times p},\ \Omega \in \mathbb{R}^{p \times d},\ \Xi \sim N(0, \Sigma)\ \text{a Gaussian noise model.} \quad (1)$$

Towards estimating Ω, we formulate a regularized empirical risk minimization problem (ERM) on a finite noisy observations along a single trajectory as:

$$\hat{\Omega} = \arg\min_{\Omega} \hat{\mathcal{L}}_{\mathcal{S}}(\Omega) + \lambda \|\Omega\|_{\mathbb{H}}^2,\ \hat{\mathcal{L}}_{\mathcal{S}}(\Omega) = \frac{1}{n} \sum_{i=1}^{n} \|y_i - \hat{y}_i\|^2$$

and corresponding error dynamics satisfies the following decomposition:

$$\|\Omega^t - \Omega^*\| \le \kappa^t \|\Omega^0 - \Omega^*\| + \sqrt{\frac{1}{SNR}} \sqrt{\frac{pdH}{n}}$$

Conclusion

Non-linearity in input phase space can be captures well by a class of separable functions such as separable functions, polynomials, linear combination of non-linear feature maps of original state variables etc. After applying any of those class of functions, learning a sparse dynamical system is turned into optimization problem that off-the-shelf our purposed optimization algorithms can handle well.

References

[1] Ahmadi, Amir Ali, and Bachir El Khadir. ”Learning dynamical systems with side information.” SIAM Review 65.1 (2023): 183-223.

[2] Ljung, Lennart. ”System identification.” Signal analysis and prediction. Boston, MA: Birkhäuser Boston, 1998. 163-173.

[3] Romberg, Justin. ”Compressive sensing by random convolution.” SIAM Journal on Imaging Sciences 2.4 (2009): 1098-1128.

[4] Beck, Amir. Introduction to nonlinear optimization: Theory, algorithms, and applications with MATLAB. Society for Industrial and Applied Mathematics, 2014.

[5] 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.

[6] Sanandaji, Borhan M., Tyrone L. Vincent, and Michael B. Wakin. ”Compressive topology identification of interconnected dynamic systems via clustered orthogonal matching pursuit.” 2011 50th IEEE Conference on Decision and Control and European Control Conference. IEEE, 2011.

[7] Sattar, Yahya, and Samet Oymak. ”Non-asymptotic and accurate learning of nonlinear dynamical systems.” Journal of Machine Learning Research 23.140 (2022): 1-49.

[8] Fattahi, Salar, Nikolai Matni, and Somayeh Sojoudi. ”Learning sparse dynamical systems from a single sample trajectory.” 2019 IEEE 58th Conference on Decision and Control (CDC). IEEE, 2019.

[9] Khansari-Zadeh, S. Mohammad, and Aude Billard. ”Learning stable nonlinear dynamical systems with gaussian mixture models.” IEEE Transactions on Robotics 27.5 (2011): 943-957.

[10] A. M. Lyapunov. General problem of the stability of motion. In Kharkov Mathematical Society, (In Russian), 1992.