2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Event-triggered learning of Euler-Lagrange systems: A Koopman operator framework

Kaikai Zheng, Dawei Shi, Shilei Li, and Yang Shi

***Abstract*—Euler-Lagrange (EL) systems represent a crucial and large class of dynamical systems, and a precise model of the true system would be beneficial in planning and tracking problems. This work aims to learn an unknown EL system using noisy measurement data to achieve improved data utilization efficiency. Specifically, for the considered EL system, a linear representation of the system is constructed using the Koopman operator, which is further characterized by sample data using Willems' fundamental lemma. Moreover, an event-triggered learning mechanism is proposed to improve data utilization efficiency, and it is designed based on the analysis of the learning error bounds. The effectiveness of the proposed event-triggered learning approach is validated through a manipulator example.**

I. Introduction

With the increasing promotion of intelligence, types of mechanical equipment are increasingly used, including vehicles, robots, manipulators, and so on. However, there is still a bridge needed between the difficulty of accurately modeling equipment using the first principle and the extensive reliance on system models for controller design [1], [2], which has promoted the development of learning-based control.

Among others, the problem of inferring knowledge of system models from sampled data is the core focus of this work. The dynamic models of many systems can be described using Euler-Lagrange (EL) equations [3], [4], many of which exhibit strong nonlinearity. Thus, nonlinear function estimation approaches can be used to learn the system dynamics. The Koopman operator shows the ability to model nonlinear systems [5], [6]. Owning to the lifting function used in the Koopman method, the nonlinear dynamics of systems can be transformed into linear dynamics in a Koopman-invariant subspace, making it possible to characterize the nonlinear dynamics with linear models [7]. Using the linearity obtained in the Koopman framework, some learning and data-driven control approaches proposed for linear systems demonstrate potential for addressing nonlinear systems, e.g., Willems' fundamental lemma [8], [9], S-lemma [10], [11]. In the approaches above, the accuracy of the learned model usually depends on the amount of data [12], [13]. This makes it challenging to deploy the above methods in some resource-constrained scenarios (such as embedded systems, energy or computationally limited distributed systems, etc.).

To improve the data efficiency and resource utilization, event-triggered learning techniques are proposed to achieve learning goals only when necessary or only using sampled data that contains relatively more information [14], [15]. In existing literature, event-triggering conditions are typically designed based on model uncertainty [16], [17] or mismatch [18]. Specifically, model uncertainty can be assessed based on specific model characteristics such as covariance, while model mismatch describes the difference between the model and the actual system. For model uncertainty, [19] proposed an online learning method based on Gaussian processes, where model updates and data forgetting are triggered by model uncertainty. Model mismatches are also used to design learning triggers. The triggering mechanism in [20] is designed by comparing the distribution of communication times with the expected value, which is an indirect evaluation of the difference between the model and the system. Although current event-triggered learning methods have made significant progress in improving learning and data efficiency, more potential solutions for specific systems (such as EL systems) remain to be further explored.

This work considers the problem of inferring the dynamic characteristics of EL systems from noisy measurements. Specifically, we consider a discrete-time EL system and construct a high-dimensional state using a lifting function, which can represent the nonlinear dynamics of the EL system using linear form in a high-dimensional space. However, there are several challenges that need to be addressed to enable the design of event-triggered learning for EL systems. First, the system's state cannot be accurately obtained; instead, it is obtained through a disturbed measurement process. Second, it needs to be clarified how to select data from noisy measurements to enhance learning efficiency. The contributions of this work are summarized as follows:

1) A model learning approach is proposed for nonlinear discrete-time EL systems, with training data obtained through a perturbed measurement process. Specifically, the nonlinear dynamics of the system are transformed into a linear form using the Koopman operator, and Willems' fundamental lemma is adopted to construct a linear non-parametric model. Furthermore, the learning error of the proposed method is proven to be bounded in the sense of high probability (Theorem 1).

This work is supported by the National Science Foundation of China under Grant 62333001. *(Corresponding author: Dawei Shi.)*

K. Zheng, D. Shi, and Shilei Li are now with the MIIT Key Laboratory of Servo Motion System Drive and Control, School of Automation, Beijing Institute of Technology, Beijing 100081, China (e-mail: kaikai.zheng@bit.edu.cn, daweishi@bit.edu.cn, shileili@bit.edu.cn).

Y. Shi is with the Department of Mechanical Engineering, Faculty of Engineering, University of Victoria, Victoria, BC V8N 3P6, Canada (e-mail: yshi@uvic.ca).

2) An event-triggered learning scheme is proposed to improve data efficiency. By analyzing the aforementioned upper bound of the learning error, a data selection strategy is determined to select a subset of available data that reduces the upper bound of the learning error. At the same probability level, the upper bound of the learning error is proved to converge exponentially (Theorem 2).
3) The proposed method is validated for modeling an unknown manipulator, which consists of a resolute joint and a prismatic joint. The implementation details are provided, and the simulation results validate the effectiveness of the proposed learning method and event-triggering mechanism.

The remainder of this paper is organized as follows. The considered problem is introduced in Section II. The Main results on event-triggered learning and learning convergence are proposed in Section III. Moreover, implementation issues and numerical verification of the proposed approaches are presented in Section IV, followed by the conclusion discussed in Section V.

Notation. Throughout this paper, the real space is denoted as $\mathbb{R}$, and positive natural numbers are denoted as $\mathbb{N}_+$. Let $\mathcal{C}^1$ be a function space, where the function $f \in \mathcal{C}^1$ has a continuous first partial derivative. For a vector $\boldsymbol{x}$, $\|\boldsymbol{x}\|_2$ denotes the 2-norm of the vector, which is defined as $\|\boldsymbol{x}\|_2 := \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$, with $\cdot^{\mathrm{T}}$ being the transposition. For a matrix $A \in \mathbb{R}^{n\times m}$, let $\sigma_1(A) \leq \ldots \leq \sigma_n(A)$ be the nonzero singular values of the matrix A. Moreover, we note that $\sigma_n(A) = \|A\|_2$ with $\|A\|_2$ denoting the 2-norm of matrix A, and the Frobenius norm of matrix A is denoted as $\|A\|_F$. In the cases of $m = n$, $|A|$ and $\mathrm{Tr}(A)$ are used to denote the determinant and trace of the matrix. For two matrices, $A \succ B$, $A \succeq B$, $A \prec B$, and $A \preceq B$ are used to denote the positive definite, positive semi-definite, negative definite, negative semi-definite of the matrix $A-B$. Moreover, $A \not\preceq B$ denotes that the matrix $B-A$ is not a positive semi-definite matrix. For a random variable, $\mathbb{P}[\cdot]$ and $\mathbb{E}[\cdot]$ are used to denote the probability and expectation, respectively.

II. Problem Setup

Consider a discrete-time EL system with sampling time interval being T as [21]

$$\begin{aligned} &M(\boldsymbol{q}(k+1))\boldsymbol{v}(k+1) - M(\boldsymbol{q}(k))\boldsymbol{v}(k) \\ &\quad - Tf(\boldsymbol{q}(k),\boldsymbol{v}(k)) = T(\boldsymbol{u}(k) + \tilde{\boldsymbol{w}}(k)), \end{aligned} \tag{1}$$

where $\boldsymbol{q}(k), \boldsymbol{v}(k), \boldsymbol{u}(k), \tilde{\boldsymbol{w}}(k) \in \mathbb{R}^n$ are position vector, velocity vector, input signal, and unknown zeros mean noise, respectively. Moreover, $M(\boldsymbol{q}(k)) \in \mathbb{R}^{n\times n}$ and $f(\boldsymbol{q}(k),\boldsymbol{v}(k)) \in \mathbb{R}^n$ are the symmetric positive definite inertia matrix and the centrifugal, Coriolis and Gravitational torques, respectively. Using the Euler's first-order derivative estimation method as $\boldsymbol{v}(k) = \frac{\boldsymbol{q}(k+1)-\boldsymbol{q}(k)}{T}$, system (1) can be equivalently rewritten as

$$\begin{cases} \boldsymbol{q}(k+1) = \boldsymbol{q}(k) + T\boldsymbol{v}(k), \\ \boldsymbol{v}(k+1) = \boldsymbol{v}(k) + h(k) + g(k)\boldsymbol{u}(k) + \boldsymbol{w}(k), \end{cases} \tag{2}$$

where

$$h(k) = -\left[M(\boldsymbol{q}(k)+T\boldsymbol{v}(k))\right]^{-1} h_1(k), \tag{3}$$

$$h_1(k) = M(\boldsymbol{q}(k)+T\boldsymbol{v}(k)) - (\boldsymbol{q}(k)))\boldsymbol{v}(k) - Tf(\boldsymbol{q}(k),\boldsymbol{v}(k)), \tag{4}$$

$$g(k) = T[M(\boldsymbol{q}(k)+T\boldsymbol{v}(k))]^{-1}, \tag{5}$$

$$\boldsymbol{w}(k) = T\left[M(\boldsymbol{q}(k)+T\boldsymbol{v}(k))\right]^{-1} \bar{\boldsymbol{w}}(k). \tag{6}$$

Furthermore, a nominal state of system (1) is written as $\bar{\boldsymbol{q}}(k) = \boldsymbol{q}(k)$ and $\bar{\boldsymbol{v}}(k) = \boldsymbol{v}(k) - \boldsymbol{w}(k-1)$, which satisfy

$$\begin{cases} \bar{\boldsymbol{q}}(k+1) = \boldsymbol{q}(k) + T\boldsymbol{v}(k), \\ \bar{\boldsymbol{v}}(k+1) = \boldsymbol{v}(k) + h(k) + g(k)\boldsymbol{u}(k). \end{cases} \tag{7}$$

Consider a noisy measurement process

$$\boldsymbol{y}(k) = \begin{bmatrix} \boldsymbol{q}(k) \\ \boldsymbol{v}(k) \end{bmatrix} + \boldsymbol{d}(k), \tag{8}$$

where $\boldsymbol{d}(k) \in \mathbb{R}^{2n}$ is a Gaussian measurement noise with zero mean and covariance matrix $\Sigma_{\boldsymbol{d}}$. The matrix $\Sigma_{\boldsymbol{d}}$ is diagonal with diagonal elements being $\{\Sigma^2_{\boldsymbol{d},1},\ldots,\Sigma^2_{\boldsymbol{d},2n}\}$. For convenience, denote $\bar{\boldsymbol{y}}(k) = \boldsymbol{y}(k) - \boldsymbol{d}(k)$.

Let $\Psi(\cdot) = [\psi_1(\cdot),\ldots,\psi_N(\cdot)]^{\mathrm{T}}$ be a lifting function of $\boldsymbol{q},\boldsymbol{v},\boldsymbol{u}$, and denote the lifted state as

$$\boldsymbol{X}(k) = \Psi(\bar{\boldsymbol{y}}(k),\boldsymbol{u}(k)) = \begin{bmatrix} \psi_1(\bar{\boldsymbol{y}}(k),\boldsymbol{u}(k)) \\ \cdots \\ \psi_N(\bar{\boldsymbol{y}}(k),\boldsymbol{u}(k)) \end{bmatrix}. \tag{9}$$

Similarly, the nominal lifted state is defined as

$$\bar{\boldsymbol{X}}(k) = \Psi\left(\left[\bar{\boldsymbol{q}}^{\mathrm{T}}(k), \bar{\boldsymbol{v}}^{\mathrm{T}}(k)\right]^{\mathrm{T}}, \boldsymbol{u}(k)\right). \tag{10}$$

In (9), N is the dimension of the lifted state, which is usually much larger than n or even infinite. Moreover, let $C \in \mathbb{R}^{2n\times N}$ such that $\bar{\boldsymbol{y}}(k) = C\boldsymbol{X}(k)$ [7]. The properties of the lifting function $\Psi(\cdot)$ are shown in the following assumptions and lemmas.

Assumption 1. *It is assumed that the input $\boldsymbol{u}(k)$ and states $\boldsymbol{q}(k),\boldsymbol{v}(k)$ satisfy*

$$\boldsymbol{u}(k) \in \mathbb{U},\ \boldsymbol{q}(k) \in \mathbb{Q},\ \boldsymbol{v}(k) \in \mathbb{V}, \tag{11}$$

where $\mathbb{Q}\times\mathbb{V}$ is compact and forward invariant, i.e.,

$$\boldsymbol{q}(k) \in \mathbb{Q}, \boldsymbol{v}(k) \in \mathbb{V} \tag{12}$$

$$\overset{(7)}{\Longrightarrow} \exists \boldsymbol{u}(k) \in \mathbb{U}, \bar{\boldsymbol{q}}(k+1) \in \mathbb{Q}, \bar{\boldsymbol{v}}(k+1) \in \mathbb{V}. \tag{13}$$

Assumption 2. *The function $\Psi(\cdot)$ is Lipschitz continuous in $\mathbb{Q}\times\mathbb{V}\times\mathbb{U}$ as*

$$\|\Psi(\boldsymbol{y}'(k),\boldsymbol{u}(k)) - \Psi(\boldsymbol{y}''(k),\boldsymbol{u}(k))\|_2 \leq L_\Psi \|\boldsymbol{y}'(k) - \boldsymbol{y}''(k)\|_2.$$

Lemma 1. *Given a discrete-time EL system in the form of (7), together with a lifting function $\Psi(\cdot): \mathbb{Q}\times\mathbb{V}\times\mathbb{U} \to \mathbb{R}^N$ in $\mathcal{C}^1$ with Assumption 1 satisfied, then there exists an exact finite-dimensional lifting ($N < \infty$) as $\bar{\boldsymbol{X}}(k+1) = \mathcal{K}\boldsymbol{X}(k)$, where $\mathcal{K} \in \mathbb{R}^{N\times N}$ is called a Koopman operator.*

Proof. Lemma 1 can be easily obtained using Theorem 2 in

[22] and thus is omitted.

In this work, we aim to design a learning approach using available lifting function $\Psi(\cdot)$ and noisy sampled dataset $\{(\boldsymbol{u}(i), \boldsymbol{y}(i))\}_{i=0}^{k}$ for system (1) with improved data efficiency, and the following questions are discussed in this work:

1) How to predict future states using the available dataset?
2) How to actively choose sampled data to improve the learning efficiency with bounded prediction error?

III. Main Results

A. Fundamental Lemma and Learning Error Analysis

This work aims to learn a prediction model for the nonlinear EL system (1). We denote the sampled data at time k as $\mathcal{D}(k) = \{D(r_i)|r_i \in \mathcal{R}(k)\}$, where $N_s(k) \in \mathbb{N}_+$, and $\mathcal{R}(k) = \{r_1, \ldots, r_{N_s(k)}\}$ is the index set with $D(r_i)$ being a data sample defined as

$$D(r_i) = (\boldsymbol{y}(r_i), \boldsymbol{u}(r_i), \boldsymbol{y}(r_i + 1)). \tag{14}$$

In the dataset $\mathcal{D}(k)$, the data are sampled intermittently, for which the fundamental lemma can be proved as follows, where $\bar{\mathcal{D}}(k)$ is a dataset defined as

$$\bar{\mathcal{D}}(k) = \{\bar{D}(r_i)|r_i \in \mathcal{R}(k)\}, \tag{15}$$
$$\bar{D}(r_i) = (\bar{\boldsymbol{y}}(r_i), \boldsymbol{u}(r_i), \bar{\boldsymbol{q}}(r_i + 1), \bar{\boldsymbol{v}}(r_i + 1)). \tag{16}$$

Lemma 2. *For a data set $\bar{\mathcal{D}}(k)$, if the matrix of lifted sate $\left[\boldsymbol{X}(r_1), \ldots, \boldsymbol{X}(r_{N_s(k)})\right]$ has full row rank, then the following equation is satisfied*

$$\begin{cases} \boldsymbol{q}_f = \boldsymbol{q}_p + T\boldsymbol{v}_p, \\ \boldsymbol{v}_f = \boldsymbol{v}_p + h_p + g_p\boldsymbol{u}_p \end{cases} \tag{17}$$

if and only if $\exists g \in \mathbb{R}^{N_s(k)}$ such that

$$\begin{bmatrix} \Psi\left(\left[\boldsymbol{q}_p^{\mathrm{T}}, \boldsymbol{v}_p^{\mathrm{T}}\right]^{\mathrm{T}}, \boldsymbol{u}_p\right) \\ \Psi\left(\left[\boldsymbol{q}_f^{\mathrm{T}}, \boldsymbol{v}_f^{\mathrm{T}}\right]^{\mathrm{T}}, \boldsymbol{u}_f\right) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}(r_1) & \cdots & \boldsymbol{X}(r_{N_s(k)}) \\ \bar{\boldsymbol{X}}(r_1 + 1) & \cdots & \bar{\boldsymbol{X}}(r_{N_s(k)+1}) \end{bmatrix} g,$$

where

$$h_p = -\left[M(\boldsymbol{q}_p + T\boldsymbol{v}_p)\right]^{-1} h_{1p},$$
$$h_{1p} = M(\boldsymbol{q}_p + T\boldsymbol{v}_p) - (\boldsymbol{q}_p))\boldsymbol{v}_p - Tf(\boldsymbol{q}_p, \boldsymbol{v}_p),$$
$$g_p = TM(\boldsymbol{q}_p + T\boldsymbol{v}_p).$$

Proof. According to (7)-(9), equation (17) can be equivalently written as

$$\Psi\left(\left[\boldsymbol{q}_f^{\mathrm{T}}, \boldsymbol{v}_f^{\mathrm{T}}\right]^{\mathrm{T}}, \boldsymbol{u}_f\right) = \mathcal{K}\Psi\left(\left[\boldsymbol{q}_p^{\mathrm{T}}, \boldsymbol{v}_p^{\mathrm{T}}\right]^{\mathrm{T}}, \boldsymbol{u}_p\right), \tag{18}$$

for some $\boldsymbol{u}_f \in \mathbb{R}^n$.

Noting that Lemma 1 revealed the linear relationship between $\boldsymbol{X}(r_i)$ and $\bar{\boldsymbol{X}}(r_i + 1)$ under the linear operator $\mathcal{K}$, which is a matrix in the considered problem. Thus, the claim can be proved using Willems' fundamental lemma. □

In Lemma 2, $(\boldsymbol{q}_p, \boldsymbol{v}_p, \boldsymbol{u}_p)$, $\boldsymbol{X}(r_i)$, and $\bar{\boldsymbol{X}}(r_i + 1)$ are available, and $\Psi(\boldsymbol{q}_f, \boldsymbol{v}_f, \boldsymbol{u}_f)$ is an unknown vector to be calculated, which can be obtained through a two-step calculation as

$$\Psi\left(\left[\boldsymbol{q}_p^{\mathrm{T}}, \boldsymbol{v}_p^{\mathrm{T}}\right]^{\mathrm{T}}, \boldsymbol{u}_p\right) = \left[\boldsymbol{X}(r_1), \cdots, \boldsymbol{X}(r_{N_s(k)})\right] g_k \tag{19}$$
$$\Rightarrow g_k = \left[\boldsymbol{X}(r_1), \cdots, \boldsymbol{X}(r_{N_s(k)})\right]^{\dagger} \Psi\left(\left[\boldsymbol{q}_p^{\mathrm{T}}, \boldsymbol{v}_p^{\mathrm{T}}\right]^{\mathrm{T}}, \boldsymbol{u}_p\right),$$

and

$$\Psi\left(\left[\boldsymbol{q}_f^{\mathrm{T}}, \boldsymbol{v}_f^{\mathrm{T}}\right]^{\mathrm{T}}, \boldsymbol{u}_f\right) = \left[\bar{\boldsymbol{X}}(r_1 + 1), \cdots, \bar{\boldsymbol{X}}(r_{N_s(k)+1})\right] g_k.$$

It can be observed that the data used to calculate vector g_k is $\bar{\boldsymbol{y}}(r_i)$, which is unavailable in practice. Thus, a reluctant compromise is to compute an estimation $\hat{g}_k$ of vector g_k using noisy measurements as

$$\hat{g}_k = \left[\tilde{\boldsymbol{X}}(r_1), \cdots, \tilde{\boldsymbol{X}}(r_{N_s(k)})\right]^{\dagger} \Psi\left(\left[\boldsymbol{q}_p^{\mathrm{T}}, \boldsymbol{v}_p^{\mathrm{T}}\right]^{\mathrm{T}}, \boldsymbol{u}_p\right), \tag{20}$$

where $\tilde{\boldsymbol{X}}(r_1)$ is the lifted state of noisy measurements defined as

$$\tilde{\boldsymbol{X}}(k) = \Psi(\boldsymbol{y}(k), \boldsymbol{u}(k)) = \begin{bmatrix} \psi_1(\boldsymbol{y}(k), \boldsymbol{u}(k)) \\ \cdots \\ \psi_N(\boldsymbol{y}(k), \boldsymbol{u}(k)) \end{bmatrix}. \tag{21}$$

The learning error of $\hat{g}_k$ is defined as

$$E_g(k) := \|g_k - \hat{g}_k\|_2. \tag{22}$$

Before analyzing the property of the learning error E_g, two useful lemmas are first introduced as follows.

Lemma 3. *(Theorem 3.3 in [23]) For any matrices A, E and B with $B = A + E$, the following inequality holds:*

$$\|B^{\dagger} - A^{\dagger}\|_F \leq \sqrt{2}\|E\|_F \max\left\{\|A^{\dagger}\|_2^2, \|B^{\dagger}\|_2^2\right\}. \tag{23}$$

Lemma 4. *(Theorem 12 in [24]) Let $\Phi \succ 0$ be a matrix parameter, and let X be a random matrix such that $X \succ 0$ almost surely. Then the following inequality holds:*

$$\mathbb{P}(X \not\preceq \Phi) \leq \mathrm{Tr}(\mathbb{E}[X]\Phi^{-1}). \tag{24}$$

Using Lemma 4, the following lemma can be obtained as a probabilistic bound for random matrices.

Lemma 5. *For the noisy sampled data recorded in dataset $\mathcal{D}(k)$, let $D_{\boldsymbol{d}}(k) = [\boldsymbol{d}(r_1), \ldots, \boldsymbol{d}(r_{N_s(k)})], r_i \in \mathcal{R}(k)$, then a probabilistic upper bound for $\|D_{\boldsymbol{d}}(k)\|_2$ can be obtained as*

$$\mathbb{P}\left[\|D_{\boldsymbol{d}}(k)\|_2^2 \leq \frac{2nN_s(k)}{1-\delta}\|\Sigma_{\boldsymbol{d}}\|_{\infty}\right] \geq \delta, \tag{25}$$

where $\delta \in (0, 1)$ is a constant.

Proof. The proof can be performed using Lemma 4 and thus is omitted. □

Utilizing the Lemmas aforementioned, an upper bound of the learning error E_g is proposed in the following theorem, where we denote that

$$
\begin{aligned}
E_{\boldsymbol{d}}(k) &= \mathcal{X}(k) - \bar{\mathcal{X}}(k). \\
\bar{\mathcal{X}}(k) &= \left[\boldsymbol{X}(r_1),\cdots,\boldsymbol{X}(r_{N_s(k)})\right], \qquad (26)\\
\mathcal{X}(k) &= \left[\tilde{\boldsymbol{X}}(r_1),\cdots,\tilde{\boldsymbol{X}}(r_{N_s(k)})\right], \qquad (27)\\
U_{\boldsymbol{d}}(k) &= \sqrt{\frac{2nN_s(k)}{1-\delta}\|\Sigma_{\boldsymbol{d}}\|_\infty} \qquad (28)\\
\bar{E}_g(k) &= \frac{2\sqrt{n}L_\Psi \left\|\Psi\left(\left[\boldsymbol{q}_p^{\mathrm{T}},\boldsymbol{v}_p^{\mathrm{T}}\right]^{\mathrm{T}},\boldsymbol{u}_p\right)\right\|_2 U_{\boldsymbol{d}}(k)}{\left(\sigma_1(\mathcal{X}(k)) - \sqrt{2n}L_\Psi U_{\boldsymbol{d}}(k)\right)^2}. \qquad (29)
\end{aligned}
$$

Theorem 1. *For a discrete time EL system with the form of (1) and a dataset $\mathcal{D}(k)$ with Assumptions 1-2 and Lemma 1 holds, the learning error $E_g(k) = \|g_k - \hat{g}_k\|_2$ satisfies:*

$$
\mathbb{P}\left[E_g(k) \leq \bar{E}_g(k)\right] \geq \delta, \tag{30}
$$

if $\mathcal{X}(k)$ has full row rank and

$$
\sigma_1(\mathcal{X}(k)) \geq \sqrt{2n}L_\Psi U_{\boldsymbol{d}}(k). \tag{31}
$$

Proof. Using the definitions of g_k, $\hat{g}_k$, and $E_g(k)$, we have

$$
E_g(k) \leq \|\bar{\mathcal{X}}^\dagger(k) - \mathcal{X}^\dagger(k)\|_2 \left\|\Psi\left(\left[\boldsymbol{q}_p^{\mathrm{T}},\boldsymbol{v}_p^{\mathrm{T}}\right]^{\mathrm{T}},\boldsymbol{u}_p\right)\right\|_2. \tag{32}
$$

According to Lemma 3, we obtain

$$
\begin{aligned}
&\|\bar{\mathcal{X}}^\dagger(k) - \mathcal{X}^\dagger(k)\|_2 \\
\leq&\sqrt{2}\|E_{\boldsymbol{d}}(k)\|_F \max\{\sigma_N^2(\mathcal{X}^\dagger(k)), \sigma_N^2(\bar{\mathcal{X}}^\dagger(k))\}.
\end{aligned}
$$

According to Assumption 2, we have

$$
\|\boldsymbol{X}(r_i) - \tilde{\boldsymbol{X}}(r_i)\|_2 \leq L_\Psi\|\bar{\boldsymbol{y}}(r_i) - \boldsymbol{y}(r_i)\|_2 = L_\Psi\|\boldsymbol{d}(k_i)\|_2.
$$

Then by noting that

$$
\begin{aligned}
\|E_{\boldsymbol{d}}(k)\|_F &= \sum_{r_i\in\mathcal{R}(k)} \|\boldsymbol{X}(r_i) - \tilde{\boldsymbol{X}}(r_i)\|_2, \\
\|D_{\boldsymbol{d}}(k)\|_F &= \sum_{r_i\in\mathcal{R}(k)} \|\boldsymbol{d}(r_i) - \tilde{\boldsymbol{d}}(r_i)\|_2,
\end{aligned}
$$

we obtain inequalities as follows, which hold with probability larger than δ according to Lemma 5:

$$
\begin{aligned}
\|E_{\boldsymbol{d}}(k)\|_F &\leq L_\Psi\|D_{\boldsymbol{d}}(k)\|_F \\
&\leq \sqrt{2n}L_\Psi\|D_{\boldsymbol{d}}(k)\|_2 \leq \sqrt{2n}L_\Psi U_{\boldsymbol{d}}(k).
\end{aligned} \tag{33}
$$

Furthermore, we have

$$
\begin{aligned}
&\max\{\sigma_N^2(\mathcal{X}^\dagger(k)), \sigma_N^2(\bar{\mathcal{X}}^\dagger(k)) \\
=&\frac{1}{\min\{\sigma_1^2(\mathcal{X}(k)), \sigma_1^2(\bar{\mathcal{X}}(k))\}}.
\end{aligned} \tag{34}
$$

Then, the following inequality can be obtained,

$$
\begin{aligned}
\sigma_1^2(\bar{\mathcal{X}}(k)) &= \sigma_1^2(\mathcal{X}(k) - E_{\boldsymbol{d}}(k)) \\
&\geq \left(\sigma_1(\mathcal{X}(k)) - \sigma_N(E_{\boldsymbol{d}}(k))\right)^2.
\end{aligned}
$$

Thus (34) can be further derived as

$$
\begin{aligned}
&\max\{\sigma_N^2(\mathcal{X}^\dagger(k)), \sigma_N^2(\bar{\mathcal{X}}^\dagger(k)) \\
\leq&\frac{1}{\left(\sigma_1(\mathcal{X}(k)) - \sqrt{2n}L_\Psi U_{\boldsymbol{d}}(k)\right)^2}.
\end{aligned} \tag{35}
$$

As a result, we can claim the following equation by combining (32) and (35):

$$
\mathbb{P}\left[E_g(k) \leq \frac{2\sqrt{n}L_\Psi \left\|\Psi\left(\left[\boldsymbol{q}_p^{\mathrm{T}},\boldsymbol{v}_p^{\mathrm{T}}\right]^{\mathrm{T}},\boldsymbol{u}_p\right)\right\|_2 U_{\boldsymbol{d}}(k)}{\left(\sigma_1(\mathcal{X}(k)) - \sqrt{2n}L_\Psi U_{\boldsymbol{d}}(k)\right)^2}\right] > \delta,
$$

which complete the proof. □

B. Event-triggered Learning and Prediction Error Analysis

Relative to (30), a more accurate error bound can be expressed as

$$
\tilde{E}_g(k) = \frac{2\left\|\Psi\left(\left[\boldsymbol{q}_p^{\mathrm{T}},\boldsymbol{v}_p^{\mathrm{T}}\right]^{\mathrm{T}},\boldsymbol{u}_p\right)\right\|_2 \|E_{\boldsymbol{d}}(k)\|_F}{\left(\sigma_1(\mathcal{X}(k)) - \|E_{\boldsymbol{d}}(k)\|_2\right)^2}. \tag{36}
$$

which ensures the following probability inequality

$$
\mathbb{P}\left[E_g(k) \leq \tilde{E}_g(k)\right] \geq \delta. \tag{37}
$$

Higher $\sigma_1(\mathcal{X}(k))$ would lead to a smaller estimation error $\|g_k - \hat{g}_k\|_2$, motivating the selection of sample data to construct $\mathcal{D}(k)$. This selection is expected to significantly increase the minimum singular value of the matrix $\mathcal{X}(k)$.

For a full row rank matrix $\mathcal{X}(k)$ and a new sampled data $D(k) = \{\boldsymbol{y}(k), \boldsymbol{u}(k), \boldsymbol{y}(k+1)\}$, an event-triggered learning mechanism is designed as

$$
\mathcal{R}(k+1) = \begin{cases} \mathcal{R}(k) \cup \{k\}, & \text{(39) holds,} \\ \mathcal{R}(k), & \text{otherwise,} \end{cases} \tag{38}
$$

$$
\begin{aligned}
&\left(\frac{N_s(k)+1}{N_s(k)}\right)^{\frac{1}{4}} \qquad (39)\\
<&\frac{\sigma_1(\check{\mathcal{X}}(k+1)\check{\mathcal{X}}^{\mathrm{T}}(k+1)) - \sqrt{2n}L_\Psi\check{U}_{\boldsymbol{d}}(k)}{\sigma_1(\mathcal{X}(k)\mathcal{X}^{\mathrm{T}}(k)) - \sqrt{2n}L_\Psi U_{\boldsymbol{d}}(k)}\sqrt{\delta_l},
\end{aligned}
$$

where $\delta_l < 1$ is a user-defined constant, and

$$
\begin{aligned}
\check{\mathcal{X}}(k+1) &:= \left[\mathcal{X}(k)\ \tilde{\boldsymbol{X}}(k)\right] \\
\check{U}_{\boldsymbol{d}}(k) &:= U_{\boldsymbol{d}}(k) + \sqrt{\frac{2n}{1-\delta}\|\Sigma_{\boldsymbol{d}}\|_\infty}.
\end{aligned} \tag{40}
$$

For the training dataset $\mathcal{D}(k)$ obtained from the even-triggered learning scheme (38), the following theorem is proposed for the convergence of learning error.

Theorem 2. *Consider a discrete time EL system with the form of (2) and a given lifting function $\Psi(\cdot)$ defined on $\mathbb{Q}\times\mathbb{V}\times\mathbb{U}$ such that Assumptions 1-2 and (31) hold. If the dataset $\mathcal{D}(k)$ is obtained according to the event-triggered learning scheme (38), then the following inequality hold*

$$
\bar{E}_g(k) \leq \delta_l^{N_s(k)-N_s(k_0)}\bar{E}_g^*, \tag{41}
$$

where $N_s(k_0)$ if a constant such that $\mathcal{X}(k_0)$ has full row rank and $\bar{E}_g^* := \bar{E}_g(k_0)$.

Proof. In the proposed event-triggered learning scheme (38), a new sampled data $D(k) = \{\boldsymbol{y}(k), \boldsymbol{u}(k), \boldsymbol{y}(k+1)\}$ is included in the training dataset $\mathcal{D}(k+1)$ if and only if (39) holds. Otherwise, the sets $\mathcal{R}(k+1), \mathcal{D}(k+1)$, constants $N_s(k+1), \bar{E}_g(k+1)$ remain invariant comparing to time instant k. Thus, the main discussion of this proof concentrates on the condition that (39) is satisfied, which can be equivalently rewritten as

$$\frac{1}{\sigma_1(\mathcal{X}(k+1)\mathcal{X}^{\mathrm{T}}(k+1)) - \sqrt{2n}L_\Psi U_{\boldsymbol{d}}(k)} \tag{42}$$

$$<\frac{1}{\sigma_1(\mathcal{X}(k)\mathcal{X}^{\mathrm{T}}(k)) - \sqrt{2n}L_\Psi U_{\boldsymbol{d}}(k)}\left(\frac{N_s(k)\delta_l^2}{N_s(k+1)}\right)^{\frac{1}{4}}. \tag{43}$$

Moreover, by squaring both sides of the equation and multiplying the same positive constant, an inequality can be obtained as

$$\frac{2\sqrt{n}L_\Psi\left\|\Psi\left(\left[\boldsymbol{q}_p^{\mathrm{T}}, \boldsymbol{v}_p^{\mathrm{T}}\right]^{\mathrm{T}}, \boldsymbol{u}_p\right)\right\|_2\sqrt{\frac{2n}{1-\delta}\|\Sigma_{\boldsymbol{d}}\|_\infty}}{\left(\sigma_1(\mathcal{X}(k+1)\mathcal{X}^{\mathrm{T}}(k+1)) - \sqrt{2n}L_\Psi U_{\boldsymbol{d}}(k)\right)^2} \tag{44}$$

$$<\frac{2\sqrt{n}L_\Psi\left\|\Psi\left(\left[\boldsymbol{q}_p^{\mathrm{T}}, \boldsymbol{v}_p^{\mathrm{T}}\right]^{\mathrm{T}}, \boldsymbol{u}_p\right)\right\|_2\sqrt{\frac{2n}{1-\delta}\|\Sigma_{\boldsymbol{d}}\|_\infty}}{\left(\sigma_1(\mathcal{X}(k)\mathcal{X}^{\mathrm{T}}(k)) - \sqrt{2n}L_\Psi U_{\boldsymbol{d}}(k)\right)^2}\frac{\sqrt{N_s(k)}\delta_l}{\sqrt{N_s(k+1)}}.$$

Furthermore, by recalling the definitions of $U_{\boldsymbol{d}}(k)$ and $\bar{E}_g(k)$ in (28) and (29), an inequality can be verified from (44) as

$$\bar{E}_g(k+1) < \delta_l \bar{E}_g(k). \tag{45}$$

By denoting the first time instant that the available data matrix $\mathcal{X}(\cdot)$ has full row rank as k_0, Inequality (45) further leads to

$$\bar{E}_g(k) \le \delta_l^{N_s(k)-N_s(k_0)}\bar{E}_g(k_0), \tag{46}$$

which completes the proof. □

IV. Experimental Results

In this section, the proposed method is applied to a simple manipulator with a resolute joint and a prismatic joint, i.e., an RP manipulator. The schematic diagram of the manipulator is shown in Fig. 1, where $m_1 = 15, m_2 = 5$ are the mass, $l = 0.3$ is the distance from the rotation center to the center of mass, $L \in (0.5, 1)$ is the state of the prismatic joint, and $\theta \in (0, \frac{1}{2}\pi)$ is the state of the resolute joint. The EL model of the manipulator is given by

$$\begin{bmatrix} m_1 l^2 + m_2 L^2(k) & 0 \\ 0 & m_2 \end{bmatrix}\begin{bmatrix} \ddot{\theta}(k) \\ \ddot{L}(k) \end{bmatrix} + \begin{bmatrix} 2m_2 L(k)\dot{\theta}(k)\dot{L}(k) \\ -m_2 L(k)\dot{\theta}(k) \end{bmatrix}$$
$$+\begin{bmatrix} (m_1 l + m_2 L(k))g\cos(\theta(k)) \\ m_2 g \sin(\theta(k)) \end{bmatrix} = \begin{bmatrix} u_1(k) \\ u_2(k) \end{bmatrix} + \tilde{\boldsymbol{w}}(k) \tag{47}$$

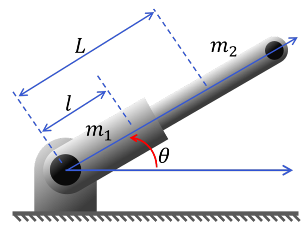

Fig. 1: A schematic diagram of the considered manipulator, which consists of a resolute joint and a prismatic joint.

with $g = 9.8$ being the gravitational acceleration. The position vector, velocity vector, and input signal of the considered system are defined as

$$\boldsymbol{q}(k) = \begin{bmatrix} \theta(k) \\ L(k) \end{bmatrix}, \ \boldsymbol{v}(k) = \begin{bmatrix} \dot{\theta}(k) \\ \dot{L}(k) \end{bmatrix}, \ \boldsymbol{u}(k) = \begin{bmatrix} u_1(k) \\ u_2(k) \end{bmatrix}.$$

Using Euler's first-order derivative estimation method, the discrete-time form of system (47) can be written as

$$\begin{cases} \boldsymbol{q}(k+1) = \boldsymbol{q}(k) + T\boldsymbol{v}(k) \\ \boldsymbol{v}(k+1) = \boldsymbol{v}(k) + h(k) + TM^{-1}\boldsymbol{u}(k) + \boldsymbol{w}(k), \end{cases} \tag{48}$$

where $\boldsymbol{w}(k) = TM^{-1}\tilde{\boldsymbol{w}}(k)$ with $h(k)$ and M being

$$M = \begin{bmatrix} m_1 l^2 + m_2 L^2(k) & 0 \\ 0 & m_2 \end{bmatrix} \tag{49}$$

$$h(k) = -M^{-1}h_1(k) \tag{50}$$

$$h_1(k) = \begin{bmatrix} Td(k)\dot{\theta}(k)\dot{L}(k) + Tm_1 l + Tm_2 g L(k)\cos(\theta(k)) \\ -Tm_2 L(k)\dot{\theta}(k) + Tm_2 g \sin\theta(k) \end{bmatrix}.$$

To validate the effectiveness of the proposed method, N_l available data samples are denoted as $\bar{\mathcal{D}} = \{D(k) | k \in \{1, 2, \ldots, N_l\}\}$, and the parameters of noises are set as $\tilde{\boldsymbol{w}} \sim \mathcal{N}(0, \mathrm{diag}([0.3, \ 0.3]))$, $\Sigma_{\boldsymbol{d},i} = 10^{-2}, i \in \{1, 2, 3, 4\}$. Then, the proposed event-triggered learning method is applied to the set $\bar{\mathcal{D}}$. Specifically, a matrix that makes the rows of $\mathcal{X}(0)$ full rank is first randomly selected, and then the other data in set $\bar{\mathcal{D}}$. All data that satisfy (39) are included in the dataset $\mathcal{D}$. Meanwhile, another dataset $\tilde{\mathcal{D}}$ is constructed without an event-triggering mechanism, which is initialized as same as $\mathcal{X}(0)$. Finally, to validate the effectiveness of the proposed event-triggered learning method, we predict the system states $\boldsymbol{q}_f := [\theta_f, \ L_f]^{\mathrm{T}}$ and $\boldsymbol{v}_f := [\dot{\theta}_f, \ \dot{L}_f]^{\mathrm{T}}$ based on datasets $\bar{\mathcal{D}}$ and $\mathcal{D}$ respectively.

The state prediction obtained based on dataset $\bar{\mathcal{D}}$ is denoted as $\bar{\boldsymbol{q}}_f, \bar{\boldsymbol{v}}_f$, and its error is denoted as

$$\hat{E}_f := \left\|\begin{bmatrix} \bar{\boldsymbol{q}}_f \\ \bar{\boldsymbol{v}}_f \end{bmatrix} - \begin{bmatrix} \mathbb{E}[\boldsymbol{q}_f] \\ \mathbb{E}[\boldsymbol{v}_f] \end{bmatrix}\right\|_2. \tag{51}$$

Moreover, $\bar{E}_f$ is defined similar to $\hat{E}_f$ with the state predictions obtained using dataset $\tilde{\mathcal{D}}$. The obtained prediction errors are shown in Fig. 2 for $\boldsymbol{q}_p = [\theta_p, \ L_p]^{\mathrm{T}}, \theta_p \in (0, \frac{1}{3}\pi), L_p \in (0.5, 1)$.

In the numerical example shown in Fig. 2, 98.7% of the data samples were removed from the dataset $\bar{\mathcal{D}}$ (from 2000 to

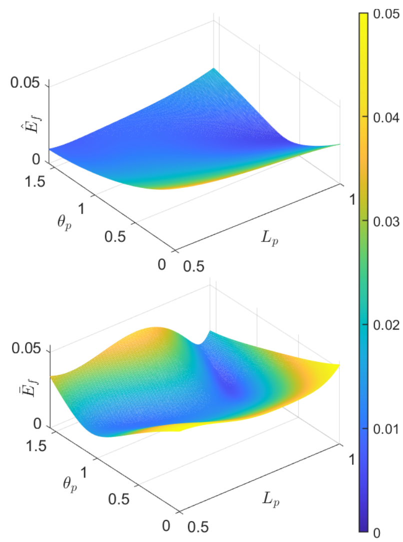

Fig. 2: Prediction error within the state space $\theta_p \in (0, \frac{1}{3}\pi), L_p \in (0.5, 1)$, and the hyper-parameters are chosen as $\delta = 0.1, \delta_l = 0.9$. The number of training data in data set $\bar{\mathcal{D}}$ is $N_l = 2000$, and $\mathcal{D}$ is $N_s = 26$.

26), while the prediction error remained within an acceptable range $\hat{E}_f \leq 0.041$. The simulation results above verify the validity of the proposed upper bound of learning error (Theorems 1-2) and the event-triggered learning mechanism (38), which means that a large number of redundant data are eliminated, and the accuracy of the resulting data-driven prediction model is kept within the acceptable range.

V. Conclusion

This work proposed an event-triggered learning approach for EL systems. By constructing a linear representation of the EL system using the Koopman operator in a high-dimensional space, a linear form of the considered system is proposed. Furthermore, the obtained linear form is learned from sampled data obtained from a noisy measurement process. By analyzing the learning error bound, an event-triggered learning scheme is designed to select part of the available data to improve learning efficiency. Future works will focus on further enhancing the learning performance and implementing data-driven control strategies that guarantee performance and safety.

References

[1] Z.-S. Hou and Z. Wang, "From model-based control to data-driven control: Survey, classification and perspective," *Information Sciences*, vol. 235, pp. 3–35, 2013.

[2] J. Berberich, J. Köhler, M. A. Müller, and F. Allgöwer, "Data-driven model predictive control with stability and robustness guarantees," *IEEE Transactions on Automatic Control*, vol. 66, no. 4, pp. 1702–1717, 2020.

[3] C. Wei, J. Luo, H. Dai, and J. Yuan, "Adaptive model-free constrained control of postcapture flexible spacecraft: a Euler–Lagrange approach," *Journal of Vibration and Control*, vol. 24, no. 20, pp. 4885–4903, 2018.

[4] T. Beckers, D. Kulić, and S. Hirche, "Stable Gaussian process based tracking control of Euler–Lagrange systems," *Automatica*, vol. 103, pp. 390–397, 2019.

[5] P. Bevanda, S. Sosnowski, and S. Hirche, "Koopman operator dynamical models: Learning, analysis and control," *Annual Reviews in Control*, vol. 52, pp. 197–212, 2021.

[6] B. Yi and I. R. Manchester, "On the equivalence of contraction and Koopman approaches for nonlinear stability and control," *IEEE Transactions on Automatic Control*, 2023.

[7] M. Korda and I. Mezić, "Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control," *Automatica*, vol. 93, pp. 149–160, 2018.

[8] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. De Moor, "A note on persistency of excitation," *Systems & Control Letters*, vol. 54, no. 4, pp. 325–329, 2005.

[9] T. Gholaminejad and A. Khaki-Sedigh, "Stable data-driven Koopman predictive control: Concentrated solar collector field case study," *IET Control Theory & Applications*, vol. 17, no. 9, pp. 1116–1131, 2023.

[10] H. J. van Waarde, M. K. Camlibel, and M. Mesbahi, "From noisy data to feedback controllers: Nonconservative design via a matrix S-lemma," *IEEE Transactions on Automatic Control*, vol. 67, no. 1, pp. 162–175, 2020.

[11] H. J. van Waarde, J. Eising, M. K. Camlibel, and H. L. Trentelman, "A behavioral approach to data-driven control with noisy input-output data," *IEEE Transactions on Automatic Control*, 2023.

[12] Z. Allen-Zhu, Y. Li, and Z. Song, "On the convergence rate of training recurrent neural networks," *Advances in neural information processing systems*, vol. 32, 2019.

[13] M. Korda and I. Mezić, "On convergence of extended dynamic mode decomposition to the Koopman operator," *Journal of Nonlinear Science*, vol. 28, pp. 687–710, 2018.

[14] F. Solowjow, D. Baumann, J. Garcke, and S. Trimpe, "Event-triggered learning for resource-efficient networked control," in *2018 Annual American Control Conference (ACC)*, (Milwaukee, WI, USA), pp. 6506–6512, IEEE, 2018.

[15] T. Lew, A. Sharma, J. Harrison, A. Bylard, and M. Pavone, "Safe active dynamics learning and control: A sequential exploration–exploitation framework," *IEEE Transactions on Robotics*, vol. 38, no. 5, pp. 2888–2907, 2022.

[16] G. Chowdhary, H. A. Kingravi, J. P. How, and P. A. Vela, "Bayesian nonparametric adaptive control using Gaussian processes," *IEEE Transactions on Neural Betworks and Learning Systems*, vol. 26, no. 3, pp. 537–550, 2014.

[17] K. Zheng, D. Shi, Y. Shi, and J. Wang, "Nonparameteric event-triggered learning with applications to adaptive model predictive control," *IEEE Transactions on Automatic Control*, vol. 68, no. 6, pp. 3469–3484, 2022.

[18] F. Solowjow and S. Trimpe, "Event-triggered learning," *Automatica*, vol. 117, p. 109009, 2020.

[19] J. Umlauft and S. Hirche, "Feedback linearization based on Gaussian processes with event-triggered online learning," *IEEE Transactions on Automatic Control*, vol. 65, no. 10, pp. 4154–4169, 2019.

[20] H. Schluter, F. Solowjow, and S. Trimpe, "Event-triggered learning for linear quadratic control," *IEEE Transactions on Automatic Control*, pp. 4485–4498, 2020.

[21] X. Guo and G. Wei, "Distributed sliding mode consensus control for multiple discrete-time Euler-Lagrange systems," *Applied Mathematics and Computation*, vol. 446, p. 127878, 2023.

[22] L. C. Iacob, R. Tóth, and M. Schoukens, "Koopman form of nonlinear systems with inputs," *Automatica*, vol. 162, p. 111525, 2024.

[23] G. W. Stewart, "On the perturbation of pseudo-inverses, projections and linear least squares problems," *SIAM review*, vol. 19, no. 4, pp. 634–662, 1977.

[24] R. Ahlswede and A. Winter, "Strong converse for identification via quantum channels," *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp. 569–579, 2002.