

Adaptive Input Estimation in Linear Dynamical Systems with Applications to Learning-from-Observations

Sebastian Curi, Kfir. Y. Levy and Andreas Krause

Abstract—We address the problem of estimating the inputs of a dynamical system from measurements of the system’s outputs. To this end, we introduce a novel estimation algorithm that explicitly trades off bias and variance to optimally reduce the overall estimation error. This optimal trade-off is done efficiently and adaptively in every time step. Experimentally, we show that our method often produces estimates with substantially lower error compared to the state-of-the-art. Finally, we consider the more complex *Learning-from-Observations* framework, where an agent should learn a controller from the outputs of an expert’s demonstration. We incorporate our estimation algorithm as a building block inside this framework and show that it enables learning controllers successfully.

I. INTRODUCTION

Input estimation consists of finding a sequence of inputs that produced a sequence of outputs for a given dynamical system. It has found many application in practice, such as fault detection and isolation [1], vehicle tracking [2], sensor bias estimation [3], economics [4], and geophysics [5].

Recently, a similar problem received attention in the Imitation Learning community. The *Learning-from-Observations* (LfO) problem, introduced in [6], consists of learning a controller from the outputs of a system when an expert controls it. It is more challenging than standard imitation learning methods that allow access to input-output pairs [7].

The motivation of the LfO framework is to learn controllers from everyday data instead of controlled experiments. For example, imitation learning of self-driving cars requires an expert to drive a car where both observations and driver inputs are recorded. On the other hand, there exists a lot of video footage of cars driven by drivers in daily life (e.g., from surveillance cameras) where no direct input is available. The LfO framework is a natural way of modeling this problem.

As one can imagine, an approach to solving the LfO problem uses *input estimation* methods to yield estimates of the actual inputs [8], [9]. Then, it uses (recorded observations, *estimated* inputs) pairs inside standard imitation learning methods. The quality of the controllers learned this way crucially depends on the quality of the input estimates. Thus, in this work, we focus on the input estimation problem. Concretely, we consider the setting where a teacher controls a given dynamical system. A learner has access to measurements of the system’s outputs, and her goal is to estimate the teacher’s inputs from these measurements.

Our Contributions: We design a novel and efficient algorithm called Adaptive Linear Input Estimator (AdaL-IE) for the input estimation problem. Our algorithm predicts the input at each time step in a manner that optimally trades off bias and variance, to yield low error estimates. Moreover, our method provides bounds on the estimation error. In contrast to other approaches, our method does not require any prior knowledge about the process or measurement noise, such as Gaussianity, but only requires a bound on the magnitude (or the variance) of these noise terms. Instead, we exploit the structure on the input signal, which we assume to be Lipschitz continuous. Although stringent at first, the low-pass characteristics of actuators make this assumption to hold in practice. Furthermore, we do not need to know the Lipschitz constant nor the noise bound, but just the Lipschitz-to-Noise ratio that has the same interpretation as the conventional Signal-to-Noise ratio. In the case that this ratio is unknown, a simple cross-validation scheme can be used to find this ratio. We also quantify the effect of system uncertainty on our predictions, and we show how to extend our approach to address non-linear systems. Finally, we apply our method within the LfO framework to an inverted pendulum. We show that it enables learning stabilizing controllers while the state-of-the-art input estimation often fails to stabilize the system.

A. Related Work

Input observability, as well as conditions on the system that ensure this property, were introduced in [10]. However, the authors did not provide any practical method to estimate the inputs. Later, practical methods were designed based on Kalman filtering techniques using joint or two-step state-input estimation. One such method, which is considered to be the *state-of-the-art* input estimator, is the Unbiased Minimum Variance Input Estimator (UMV-IE) [11]. As the name suggests, this method ensures unbiased input estimates. However, UMV-IE often suffers from high variance, is only optimal under particular cases, such as Gaussian noise and linear dynamical systems, and may become unstable if not tuned appropriately [12]. These claims also apply to other methods based on Kalman filters [13], [14].

An alternative to Kalman filtering was presented in [15], where authors invert the system and truncate the impulse response to the first M components to find the “inverse” dynamical system by making the strong assumption that the unknown input sequence is generated by a random walk and the common assumptions that the system and initial state distribution, as well as the mean and covariances of the measurement and process noise, are known. The resulting

Sebastian Curi and Andreas Krause are with the Learning & Adaptive Systems Group, Department of Computer Science, ETH Zurich, Switzerland. Kfir. Y. Levy is with the Electrical Engineering Department, Technion-Israel Institute of Technology. Emails: scuri@inf.ethz.ch, kfirylevy@technion.ac.il and krausea@ethz.ch.

estimator is a *fixed* weighted linear combination of the past M measurements. As shown in this paper, the choice of M is crucial for the performance of their estimator. Nevertheless, the authors do not discuss how to choose M appropriately.

In [6], authors introduced the LfO framework and solved it using an off-the-shelf reinforcement learning algorithm where the per-step cost function is the distance between the observations of the expert and the learned trajectories. This work suffers from high sample complexity, hence in [8] and [9], authors extended this framework to behavioral cloning imitation learning [16]. In these works, authors consider discrete states and actions on a Markov Decision Process (MDP). The algorithm they propose iterates between learning the inverse model of the MDP around the expert trajectory and filtering the observed states with the learned (inverse) model to estimate the inputs. First, the learner applies the estimated input to the real system and updates the model. Second, the learner uses the model to re-estimate the input.

In contrast to Kalman based input estimators, our approach does not enforce unbiased estimates. Thus, we obtain a lower estimation error, which is crucial for the LfO problem. Furthermore, while the Kalman based estimators optimality crucially relies on the Gaussianity assumption, our method enables us to handle much more general noise terms. Compared to the truncated inverse method, we do not assume that the input is generated with a random walk, but rather that the input is Lipschitz continuous. Moreover, we explicitly select the window width with a principled optimization procedure that trades off bias with variance. For the LfO problem, our method may be used as a building block that allows generalizing discrete to continuous state and actions.

Finally, we comment that all of the existing works on input estimation assume the knowledge of the system dynamics. This assumption is necessary since the case where both the system and inputs are unknown is an ill-posed problem. In this context, we show how a mismatch between our model and the *true* system introduces extra bias in our estimation.

II. BACKGROUND AND PROBLEM STATEMENT

A. Mathematical Notation

We denote a vector at time t by $v_t \in \mathbb{R}^n$, a sequence of vectors by $\{v_{t_1}, v_{t_1+1}, \dots, v_{t_2-1}, v_{t_2}\} = \{v_\tau\}_{\tau=t_1}^{t_2}$, and its i -th entry by $v_t[i] \in \mathbb{R}$. The Moore-Penrose pseudo-inverse of a matrix $A \in \mathbb{R}^{n \times m}$ is $A^\dagger = (A^\top A)^{-1} A^\top$. We use $\|\cdot\|$ for the 2-norm of a vector or the induced 2-norm of a matrix. We denote the identity matrix of size n by I_n and the delta dirac function by $\delta(\cdot)$.

B. Input Estimation in Linear Dynamical Systems

A time-varying discrete-time linear dynamical system [17] can be described with a state $x_t \in \mathbb{R}^{n_x}$, and the input $u_t \in \mathbb{R}^{n_u}$ to output $y_t \in \mathbb{R}^{n_y}$ relationship evolves as:

$$\begin{cases} x_{t+1} = A_t x_t + B_t u_t + \epsilon_t \\ y_t = C_t x_t + \nu_t, \end{cases} \quad (1)$$

where $\epsilon_t \in \mathbb{R}^{n_x}$ is process noise and $\nu_t \in \mathbb{R}^{n_y}$ is measurement noise. The state transition matrix is defined as:

$\Phi_{(t,i)} \equiv A_{t-1} \Phi_{(t-1,i)}$, for $0 \leq i < t$, and $\Phi_{(t,t)} = I$. The output of the system at time t is expressed as:

$$y_t = C_t \left(\Phi_{(t,0)} x_0 + \sum_{i=0}^{t-1} \Phi_{(t,i+1)} (B_i u_i + \epsilon_i) \right) + \nu_t. \quad (2)$$

Definition 1 (Strong Observability [14]). The linear system (1) is considered *strongly observable*, or equivalently state and input observable, if the initial condition x_0 and the unknown input sequence $\{u_\tau\}_{\tau=0}^{r-1}$ can be uniquely determined from the measured output sequence $\{y_\tau\}_{\tau=0}^r$, for a large enough number of observations.

Assumption 1 (System Knowledge). The system dynamics is known and it is strongly observable. Furthermore, we have an unbiased estimate \hat{x}_0 of the initial state, $x_0 \equiv \hat{x}_0 + \epsilon_0$.

Remark 1. System knowledge could be acquired with physical modeling or system identification. Input observability is needed for well-possessedness of the input estimation problem. By Definition 1, the initial state x_0 can be reconstructed after r measurements hence the initial state knowledge is a reasonable assumption in a strongly observable system.

Assumption 2 (Input Functions). We will consider the inputs $u_t = f(t, y_t)$ to be an unknown function of the time step t and the output y_t . This models both open-loop and closed-loop behavior. Furthermore, the input function is Lipschitz continuous: For any $(t_1, t_2) \in \mathbb{R}$ it holds that $\|u_{t_1} - u_{t_2}\| = \|f(t_1, y_{t_1}) - f(t_2, y_{t_2})\| \leq Ld(t_1, y_{t_1}; t_2, y_{t_2})$, where the distance function $d(\cdot; \cdot)$ is known in advance.

Remark 2. Assumption 2 is stringent but the low-pass characteristic of actuators makes this assumption to hold in practice. It is also less restrictive than random-walk assumptions done in previous work. The distance function is problem dependent and key for the success of this method. In this work, we use $d(t_1, y_{t_1}; t_2, y_{t_2}) = |t_2 - t_1| + \|y_{t_2} - y_{t_1}\|_2$.

Assumption 3 (System Noise). The process ϵ_t and measurement ν_t noise terms are zero-mean and bounded by b . The noise sequence is i.i.d. and uncorrelated, $\mathbb{E}[\epsilon_t \epsilon_{t'}] = \mathbb{E}[\nu_t \nu_{t'}] = 0$, for all $t \neq t'$, and $\mathbb{E}[\epsilon_t \nu_{t'}] = 0$, for all t, t' .

Remark 3. We need Assumption 3 to apply Hoeffding's inequality when deriving an upper bound of the estimation error. Alternatively, we could assume that the variance of the noise is bounded and apply Bernstein inequalities – the result and analysis remain qualitatively similar. If the noise is not zero-mean, then the mean is included in the input estimation.

Remark 4. Without loss of generality one can consider that $B_t = I_{n_u}$ and solve the input estimation problem for the input $\tilde{u}_t = B_t u_t$. The input is estimated as $\hat{u}_t = B_t^\dagger \hat{y}_t$. Furthermore, if the original system is observable, the modified system is strongly observable and \tilde{u}_t can be estimated in any observable system.

III. ADAPTIVE LINEAR INPUT ESTIMATOR

A. Estimator Architecture

The unbiased, minimum variance input estimator (UMV-IE) has a recursive and linear architecture that ensures unbiasedness [11, Theorem 1]. To have a lower estimation error, we allow our estimates to break the unbiasedness constraint and propose for each timestep t , the following anti-causal finite impulse response estimation architecture:

$$\hat{u}_t = \sum_{\tau=1}^T \alpha_t[\tau] (C_\tau H_\tau)^\dagger (y_\tau - C_\tau \Phi_{(\tau,0)} \hat{x}_0), \quad (3)$$

where $C_\tau H_\tau \equiv \sum_{i=0}^{\tau-1} C_\tau \Phi_{(\tau,i+1)}$ is the τ -row sum of the invertibility matrix \mathcal{I}_T , and the filter parameters α_t are defined over the simplex, $\Delta_T = \{\alpha_t \in \mathbb{R}^T \mid \sum_{\tau=1}^T \alpha_t[\tau] = 1 \text{ and } \alpha_t[\tau] \geq 0, \forall \tau \in \{1, 2, \dots, T\}\}$. By Assumption 1, the product $(C_\tau H_\tau)^\dagger (C_\tau H_\tau) = I$, else \mathcal{I}_T would be rank deficient and the system not strongly observable. Furthermore, the inverse of H_τ exists and is unique.

Next, we give some intuition behind the structure of the estimators in (3). For a constant input u and when no noise is present, $u = (C_\tau H_\tau)^\dagger (y_\tau - C_\tau \Phi_{(\tau,0)} \hat{x}_0)$, $\forall \tau$. When the input is constant, but noise is present, then averaging the different τ -indexes with weights $\alpha_t \in \Delta_T$ concentrates the different noise components to average it out. However, when the input is not constant, and noise is present, then averaging the different τ -indexes with weights $\alpha_t \in \Delta_T$ will average out the noise, but will also introduce a bias in the estimation, as discussed in the next subsection. The challenge resides in choosing the weights $\alpha_t \in \Delta_T$ to optimally trade-off the noise variance with the estimation bias.

B. Optimization Algorithm

Given the estimator architecture, our goal is to find the best coefficients α_t , that minimize the squared 2-norm of the estimation error

$$\min_{\alpha_t \in \Delta_T} \left\| \sum_{\tau=1}^T \alpha_t[\tau] (C_\tau H_\tau)^\dagger (y_\tau - C_\tau \Phi_{(\tau,0)} \hat{x}_0) - u_t \right\|^2. \quad (4)$$

Although this problem is convex on α_t and looks like a typical regression problem, we do not have the targets u_t . To bypass this problem, we build on the analysis by [18] and decompose our objective in (4) into bias and variance terms to find a relaxed objective that is tractable to optimize.

Proposition 1. *The optimization problem in (4) can be upper bounded with probability at least $1 - \beta$ by*

$$2V(\beta) \min_{\alpha_t \in \Delta_T} \alpha_t^\top \left(Q + \frac{L^2}{V(\beta)} q_t q_t^\top \right) \alpha_t, \quad (5)$$

where $Q \in \mathbb{R}^{T \times T}$ is a positive definite matrix that depends on the system dynamics, $q_t \in \mathbb{R}^T$ depends on the system dynamics and the distance metric, $V(\beta) \in \mathbb{R}_{>0}$ depends on the noise magnitude and the probability level β , and $L \in \mathbb{R}_{\geq 0}$ is the input function Lipschitz constant.

Proof. The error can be upper bounded using in bias and variance terms:

$$\begin{aligned} \text{error}^2 &= \left\| \sum_{\tau=1}^T \alpha_t[\tau] (C_\tau H_\tau)^\dagger (y_\tau - C_\tau \Phi_{(\tau,0)} \hat{x}_0) - u_t \right\|^2 \\ &\leq 2 \underbrace{\left\| \sum_{\tau=1}^T \alpha_t[\tau] H_\tau^{-1} \left(\sum_{i=0}^{\tau-1} \Phi_{(\tau,i)} \epsilon_i + \nu_\tau \right) \right\|^2}_{\text{variance}} \\ &\quad + 2 \underbrace{\left\| \sum_{\tau=1}^T \alpha_t[\tau] H_\tau^{-1} \left(\sum_{i=0}^{\tau-1} \Phi_{(\tau,i+1)} (u_i - u_t) \right) \right\|^2}_{\text{bias}^2}. \end{aligned}$$

The variance term only depends on the process and measurement noise, while the bias is present even in the noiseless case. Using Assumption 2, we bound the bias term:

$$\begin{aligned} \text{bias} &= \left\| \sum_{\tau=1}^T \alpha_t[\tau] H_\tau^{-1} \left(\sum_{i=0}^{\tau-1} \Phi_{(\tau,i+1)} (u_i - u_t) \right) \right\| \\ &\leq \sum_{\tau=1}^T |\alpha_t[\tau]| \left(\sum_{i=0}^{\tau-1} \|H_\tau^{-1} \Phi_{(\tau,i+1)}\| \|u_i - u_t\| \right) \\ &\leq L \sum_{\tau=1}^T \alpha_t[\tau] \left(\sum_{i=0}^{\tau-1} d(i, t) \|H_\tau^{-1} \Phi_{(\tau,i+1)}\| \right) = L \alpha_t^\top q_t. \end{aligned}$$

where $q_t[\tau] = \left(\sum_{i=0}^{\tau-1} d(i, t) \|H_\tau^{-1} \Phi_{(\tau,i+1)}\| \right)$. Using Hoeffding's inequality for matrix concentrations [19], [20], we bound the variance term with probability at least $(1 - \beta)$:

$$\begin{aligned} \text{var} &= \left\| \sum_{\tau=1}^T \alpha_t[\tau] H_\tau^{-1} \left(\sum_{i=0}^{\tau-1} \Phi_{(\tau,i)} \epsilon_i + \nu_\tau \right) \right\|^2 \\ &= \left\| \sum_{i=0}^T \sum_{\tau=\max\{1, i\}}^T \alpha_t[\tau] H_\tau^{-1} \Phi_{(\tau,i)} \epsilon_i + \sum_{\tau=1}^T \alpha_t[\tau] H_\tau^{-1} \nu_\tau \right\|^2 \\ &\leq 36b^2 \left(1 + \sqrt{\log(\beta^{-1})} \right)^2 \sum_{i=0}^{2T} M_i^2 \leq V(\beta) \alpha_t^\top Q \alpha_t, \end{aligned}$$

where $M_i \geq \left\| \sum_{\tau=\max\{1, i\}}^T \alpha_t[\tau] H_\tau^{-1} \Phi_{(\tau,i)} \right\|$ for $i \in \{0, \dots, T\}$ and $M_i \geq \|\alpha_t[i - T] H_{i-T}^{-1}\|$ for $i \in \{T + 1, \dots, 2T\}$. Defining $V \equiv 36b^2 \left(1 + \sqrt{\log(\beta^{-1})} \right)^2$ the quadratic form with $Q[\tau, \tau'] = \sum_{i=0}^{\min\{\tau, \tau'\}} \|H_\tau^{-1} \Phi_{(\tau,i)}\| \|H_{\tau'}^{-1} \Phi_{(\tau',i)}\| + \|H_\tau^{-1}\|^2 \delta(\tau - \tau')$ appears. The rest of the proof is simple algebra. \square

The relaxed optimization problem (5) is a quadratic program over the simplex, which can be solved efficiently via projected gradient descent [21], [22]. The solution of (5) are the T parameters of the estimator. Furthermore, the estimator has only one hyperparameter, the Lipschitz-to-Noise ratio $\frac{L^2}{V(\beta)}$, which can be tuned by cross-validation. Finally, the projection onto the simplex of the T parameters implicitly regularize them.

C. Extensions to Uncertain and Non-Linear Systems

Uncertain Systems: For simplicity, we consider $B_t = C_t = I_n$. If the state transition matrix estimate has an error as $A_t = \hat{A}_t + \Delta_{A_t}$, the dynamics can be expressed as:

$$\begin{cases} x_{t+1} = A_t x_t + u_t + \epsilon_t = \hat{A}_t x_t + (\Delta_{A_t} x_t + u_t) + \epsilon_t \\ y_t = x_t + \nu_t, \end{cases}$$

We call $\tilde{u}_t = \Delta_{A_t} x_t + u_t$ and use AdaL-IE to recover \tilde{u}_t . The Lipschitz constant of \tilde{u} is $\tilde{L} = \|\Delta_{A_t}\| + L$. If the mismatch is large, then $\|\Delta_{A_t}\|$ dominates the signal information of u , the upper bound of the bias is loose, and the estimate poor.

Initial State Knowledge: If the initial state of the system is not known, then the estimation will have an extra bias of $\sum_{\tau=1}^T \alpha_t[\tau] (C_\tau H_\tau)^\dagger \Phi_{(\tau,0)} \hat{x}_0$. For stable systems, this term is significant at times t close to 0. For unstable systems, this term can be the prominent source of error for larger times.

Non-Linear Dynamical Systems: We consider a linearization of a class of non-linear systems that include most robotics systems [23]:

$$\begin{cases} x_{t+1} = g(x_t) + h(x_t)u_t + \epsilon_t \\ y_t = C_t x_t + \nu_t, \end{cases} \approx \begin{cases} x_{t+1} = A_t x_t + B_t u_t + \epsilon_t \\ y_t = C_t x_t + \nu_t, \end{cases}$$

where $g(x_t)$ and $h(x_t)$ are non-linearities that depend only on the current state. The inputs affect the next state linearly, and the observations are linear functions of the states. Up to small errors, the linearization around the state trajectory yields $A_t = \frac{\partial g(x)}{\partial x} \Big|_{x=x_t}$, $B_t = h(x_t) + \frac{\partial h(x)}{\partial x} x \Big|_{x=x_t}$.

Of course, this is not a trivial task because we do not have access to the state. In the non-linear systems used in this work, we assumed that C_t is full rank and thus it could be inverted to obtain a *noisy* estimate of the state and we use this estimate to linearize the system. Although stringent, this assumption is used in the literature of non-linear observers with unknown inputs [24], [25]. In fact, for the problem of non-linear state estimation with unknown inputs, few results exist for particular classes of systems.

IV. EXPERIMENTS

A. Input Estimation

We perform a set of experiments to compare AdaL-IE with UMV-IE on the RMS estimation error. We apply a set of 8 different input signals to 5 different systems for 20 episodes of 100 time-steps. Even though the input signals were a priori known, we fit the SNR ratio for setting up the optimization problem (5) via cross-validation. To simulate modeling mismatch, we perturb the entries of the state transition matrix of each system with Gaussian noise. The signals are (a) a unit step, (b) a unit sine, (c) a unit ramp, (d) a unit triangle, (e) a ramp followed by a unit step, (f) a ramp followed by a zero step, (g) a unit step followed by a zero step, and (h) a unit sine superposed to a unit step. The systems were (i) a spring-mass system, (ii) a double integrator, (iii) a email server [26, Section 2.6.1], (iv) a HTTP server [26, Section 7.8.1], (v) a stable system with random coefficients in the state transition matrix, (vi) a first order

non-linear system [27, Example 4.27], and (vii) a second order non-linear system [27, Example 4.61].

We show the input estimation of system (i) in Fig. 1. We plot the input signal in blue, the UMV-IE estimates mean plus/minus one standard deviation in shaded green, and AdaL-IE estimates in shaded red. AdaL-IE has lower error than UMV-IE due to a much lower variance, but at the cost of some bias. The rest of the systems are summarized in with the experimental mean RMS estimation error in Table I. The top sub-row shows the AdaL-IE RMS error and the bottom one the UMV-IE RMS error. From the 56 experiments, in 50 AdaL-IE outperforms UMV-IE, while it underperforms in 3. In the remaining 3, there is no statistical difference.

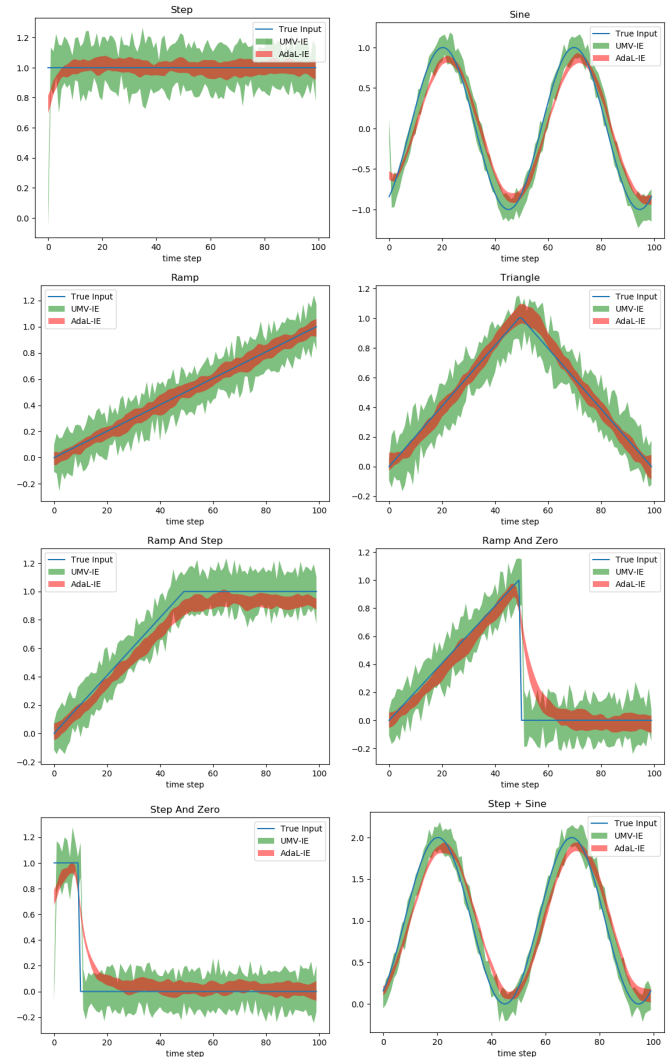


Fig. 1: Input Signal Estimation in the Spring-Mass System. We can observe that the variance of the AdaL-IE estimates is lower than those of UMV-IE. The trade-off comes at cost of some bias, which can also be seen in the plots. Finally, the RMS error of AdaL-IE is smaller than that of UMV-IE.

When the input signal changes abruptly, such as (f) and (g), Assumption 2 is violated. In these cases, the bias around discontinuity points is large, the total error is larger than with UMV-IE, and a delay appears in the estimated input signal.

TABLE I: Mean RMS Estimation Error. In each row a different input signal is applied to each system in the columns. The top sub-row has the error of the AdaL-IE estimates while the bottom sub-row the UMV-IE estimates.

	Spring -mass	Double int.	Email server	HTTP server	Rand sys.	NonLin 4.27	NonLin 4.61
(a)	0.061 0.191	0.049 0.187	0.112 0.325	0.234 0.885	0.049 0.188	0.237 1.673	0.237 1.673
(b)	0.544 1.626	0.194 0.202	0.139 0.334	0.318 0.887	0.375 0.228	0.500 1.678	0.500 1.677
(c)	0.584 1.666	0.044 0.162	0.123 0.317	0.241 0.898	0.085 0.166	0.257 1.674	0.263 1.677
(d)	0.063 0.162	0.055 0.162	0.113 0.310	0.255 0.869	0.056 0.161	0.285 1.677	0.259 1.681
(e)	0.073 0.021	0.049 0.161	0.111 0.312	0.239 0.859	0.062 0.156	0.261 1.681	0.266 1.645
(f)	0.584 1.640	0.102 0.187	0.131 0.320	0.267 0.889	0.163 0.221	0.297 1.653	0.307 1.650
(g)	0.123 0.217	0.120 0.217	0.137 0.345	0.274 0.910	0.259 0.277	0.290 1.651	0.294 1.676
(h)	0.579 1.638	0.176 0.183	0.140 0.324	0.305 0.897	0.533 0.222	0.533 1.676	0.515 1.701

Sparsity of estimator parameters: In Fig. 2 we show the coefficients of some of 1st, 50th, and 98th input filters and observe that only a small number of indexes around t to have non-zero entries. In Fig. 2d we show the count of how many non-zero parameters each estimator has for all t .

The bias term and the simplex geometry induces the sparsity pattern. The further away the indexes τ are from the estimation point t , the larger $q_t[\tau]$ is leading to larger values of the objective in (5). The simplex geometry is the dual of l_1 regularization, which is known to be sparsity inducing. For example, for estimating u_{50} , only $\alpha_{50}[49 : 53]$ are non-zero. This coincides with the intuition that u_{50} only affects current and future outputs, but later outputs have exponentially less information. However, $\alpha_{50}[49]$ is also non-zero. Inputs u_{49} and u_{50} are close due to Lipschitz continuity (Assumption 2), hence y_{49} also contains information about u_{50} .

The sparsity pattern also suggests that this estimator can be used (almost) online. Rather than waiting for T measurements, we only have to wait for 4 or 5 to estimate the input.

Noise Magnitude Effect: To understand the trade-off between bias and variance of the estimator, we analyze the spring-mass system with different noise magnitudes. We show the results in Fig. 3. When the noise is small compared to the magnitude of the signal as in Fig. 3a, then the bias is larger than the variance reduction and UMV-IE performs better. On the other hand, when the noise has the same magnitude as the signal, then the bias is meaningless as shown in Fig. 3b, and AdaL-IE outperforms UMV-IE. Even for cases in which the noise is smaller than the signal, the variance reduction is considerable, as shown in Fig. 1.

B. Learning-from-Observations

After analyzing the quality of the data that the estimator recovers, the question that we answer in this section is if this data is *good enough* to learn a controller that successfully executes a given task. As a benchmark, we consider the swing-up and stabilization of the inverted pendulum task of the openAI gym [28] with randomly perturbed parameters (10 % of the value). We generate the *expert* trajectories

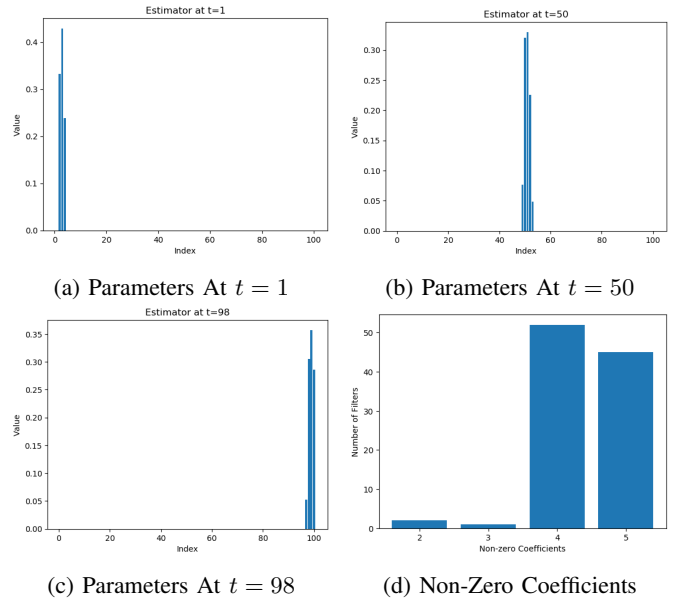


Fig. 2: Sparsity Pattern Of Estimator Parameters. At every time step, the estimator parameters have all non-zero coefficients around this time step. The inherent locality of the solution is due to the bias-variance trade-off.

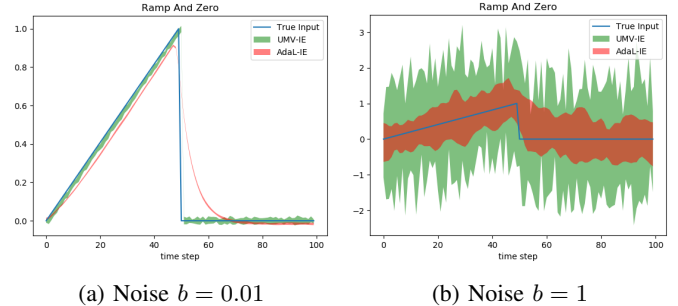


Fig. 3: For settings with high-noise to signal ratio, the variance reduction obtained with AdaL-IE overcomes the bias in the estimator. On the other hand, when the noise is low, UMV-IE estimates have lower errors; the variance reduction does not compensate for the bias increase.

with feedback linearizing controller, $f(\phi, \omega) = 15 \sin(\phi) + 30\phi + 8\omega$, where ϕ is the angular position and ω the angular velocity of the pendulum. Additively to the input created by the expert, we simulate Uniform process and measurement noise. From the demonstrations, only noisy measurements of the angular position and velocity are available. A task is successful if the resulting controller stabilizes the pendulum in the upright position after 100 seconds. We compare three different controller architectures, a linear function, a neural network with 2 hidden layers of 5 neurons with ReLU activation layers, and a neural network with the same architecture but sigmoid activation layers.

We split this LfO problem into two steps: input estimation and imitation learning with the estimated inputs. For the former, we use both UMV-IE and AdaL-IE. For the latter we use Behavioral Cloning [16]. More advanced algorithms, such as DAGGER [29] or IRL [30], are left for future work.

In Table II we show how many stabilizing controllers

we learned this way. The original controller can stabilize 930 out of 1000 demonstrations because it is not good enough for the swing-up task from *all* initial conditions due to actuator saturation. The linear controller performs even worse because it does not account for non-linearities. The sigmoid activations architecture performs better than the linear, but it can not replicate the original controller exactly due to a mixture between noisy targets, insufficient data, and approximation error. More work is needed to understand how having more data affects the quality of the input estimation targets and its interaction with LfO. The ReLU activations architecture performs worst than the linear because it has very sharp boundaries around the equilibrium, and large noise signals drive the system quickly out of stability.

To understand the difference between AdaL-IE and UMV-IE targets, we look at the closed-loop eigenvalues of the linearized system around the unstable equilibrium for the linear controller. When we use AdaL-IE targets, the closed-loop system is stable, while when we use UMV-IE targets, it is closed-loop unstable. The intuition for this is that after the swing-up phase, most of the expert trajectories are around the equilibrium. In this phase, the noise that drives the system out of the equilibrium and the controller input that drives the system back have about the same magnitude. The situation is that of Fig. 3b and the estimates of AdaL-IE have smaller error than those of UMV-IE.

TABLE II: Number of stabilizing controllers out of 1000 demonstrations. The inputs are estimated using AdaL-IE and UMV-IE respectively and used as Behavioral Cloning targets.

Controller	AdaL-IE	UMV-IE
Linear Architecture	496	0
ReLU Activations	332	0
Sigmoid Activations	754	224
Original Controller	930	

V. CONCLUSION

In this paper, we presented a new input estimator for a general class of dynamical systems by optimizing a high-probability upper bound of the estimation error. This upper bound comes from a natural decomposition in bias and variance of the estimation. We showed in experiments that this estimator has lower errors than previous state-of-the-art (UMV-IE) methods. Finally, we tested this estimator on the LfO framework and successfully learned stabilizing controllers using AdaL-IE as input estimation block.

REFERENCES

- [1] J. Chen and R. J. Patton, "Optimal filtering and robust fault diagnosis of stochastic systems with unknown disturbances," *IEE Proceedings-Control Theory and Applications*, vol. 143, no. 1, pp. 31–36, 1996.
- [2] Y. Chan, A. Hu, and J. Plant, "A kalman filter based tracking scheme with input estimation," *IEEE transactions on Aerospace and Electronic Systems*, no. 2, pp. 237–244, 1979.
- [3] S. S. Park and Y. Altintas, "Dynamic compensation of spindle integrated force sensors with kalman filter," *Journal of Dynamic Systems, Measurement, and Control*, vol. 126, no. 3, pp. 443–452, 2004.
- [4] B. Li, "State estimation with partially observed inputs: A unified kalman filtering approach," *Automatica*, vol. 49, no. 3, pp. 816–820, 2013.

- [5] P. K. Kitanidis, "Unbiased minimum-variance linear state estimation," *Automatica*, vol. 23, no. 6, pp. 775–778, 1987.
- [6] Y. Liu, A. Gupta, P. Abbeel, and S. Levine, "Imitation from observation: Learning to imitate behaviors from raw video via context translation," *arXiv preprint arXiv:1707.03374*, 2017.
- [7] A. Hussein, M. M. Gaber, E. Elyan, and C. Jayne, "Imitation learning: A survey of learning methods," *ACM Computing Surveys (CSUR)*, vol. 50, no. 2, p. 21, 2017.
- [8] F. Torabi, G. Warnell, and P. Stone, "Behavioral cloning from observation," *arXiv preprint arXiv:1805.01954*, 2018.
- [9] A. D. Edwards, H. Sahni, Y. Schroecker, and C. L. Isbell, "Imitating latent policies from observation," *arXiv preprint arXiv:1805.07914*, 2018.
- [10] M. Hou and R. J. Patton, "Input observability and input reconstruction," *Automatica*, vol. 34, no. 6, pp. 789–794, 1998.
- [11] S. Gillijns and B. De Moor, "Unbiased minimum-variance input and state estimation for linear discrete-time systems," *Automatica*, vol. 43, no. 1, pp. 111–116, 2007.
- [12] C.-S. Hsieh and M. A. Majidi, "Implementation issues of unbiased minimum-variance state estimation for systems with unknown inputs," in *Automatic Control Conference (CAC), 2014 CACS International*. IEEE, 2014, pp. 323–328.
- [13] C.-S. Hsieh, "On the optimality of two-stage kalman filtering for systems with unknown inputs," *Asian Journal of Control*, vol. 12, no. 4, pp. 510–523, 2010.
- [14] S. Z. Yong, M. Zhu, and E. Frazzoli, "A unified filter for simultaneous input and state estimation of linear discrete-time stochastic systems," *arXiv preprint arXiv:1309.6627*, 2013.
- [15] S. H. Park, P. S. Kim, O.-K. Kwon, and W. H. Kwon, "Estimation and detection of unknown inputs using optimal fir filter," *Automatica*, vol. 36, no. 10, pp. 1481–1488, 2000.
- [16] M. Bain and C. Sommut, "A framework for behavioural cloning," *Machine intelligence*, vol. 15, no. 15, p. 103, 1999.
- [17] M. Dahleh, M. A. Dahleh, and G. Verghese, "Lectures on dynamic systems and control. chapter 10," *MIT OCW*, 2004.
- [18] O. Anava and K. Levy, "k*-nearest neighbors: From global to local," in *Advances in Neural Information Processing Systems*, 2016, pp. 4916–4924.
- [19] A. Juditsky and A. S. Nemirovski, "Large deviations of vector-valued martingales in 2-smooth normed spaces," *arXiv preprint arXiv:0809.0813*, 2008.
- [20] S. Kakade, "Lecture notes in multivariate analysis, dimensionality reduction, and spectral methods." http://stat.wharton.upenn.edu/~skakade/courses/stat991_mult/lectures/MatrixConcen_derivations.pdf, 2010.
- [21] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [22] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the l_1 -ball for learning in high dimensions," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 272–279.
- [23] E. Theodorou, J. Buchli, and S. Schaal, "A generalized path integral control approach to reinforcement learning," *Journal of Machine Learning Research*, vol. 11, no. Nov, pp. 3137–3181, 2010.
- [24] Q. P. Ha and H. Trinh, "State and input simultaneous estimation for a class of nonlinear systems," *Automatica*, vol. 40, no. 10, pp. 1779–1785, 2004.
- [25] M. Corless and J. Tu, "State and input estimation for a class of uncertain systems," *Automatica*, vol. 34, no. 6, pp. 757–764, 1998.
- [26] J. L. Hellerstein, Y. Diao, S. Parekh, and D. M. Tilbury, *Feedback control of computing systems*. John Wiley & Sons, 2004.
- [27] H. K. Khalil, "Nonlinear systems, 3rd," *New Jersey, Prentice Hall*, vol. 9, no. 4.2, 2002.
- [28] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "Openai gym," *arXiv preprint arXiv:1606.01540*, 2016.
- [29] S. Ross, G. Gordon, and D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 2011, pp. 627–635.
- [30] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 1.