Model Identification Adaptive Control with ρ-POMDP Planning

Michelle Ho,[1] Arec Jamgochian,[1,2] and Mykel J. Kochenderfer[1]

Abstract— **Accurate system modeling is crucial for safe, effective control, as misidentification can lead to accumulated errors, especially under partial observability. We address this problem by formulating informative input design and model identification adaptive control (MIAC) as belief space planning problems, modeled as partially observable Markov decision processes with belief-dependent rewards (ρ-POMDPs). We treat system parameters as hidden state variables that must be localized while simultaneously controlling the system. We solve this problem with an adapted belief-space iterative Linear Quadratic Regulator (BiLQR). We demonstrate it on fully and partially observable tasks for cart-pole and steady aircraft flight domains. Our method outperforms baselines such as regression, filtering, and local optimal control methods, even under instantaneous disturbances to system parameters.**

I. Introduction

Accurate system dynamics models are essential for safe, effective control since incomplete knowledge or simplifying assumptions can lead to misspecifications. System identification uses observed data to estimate model parameters for better predictions, fault detection, and robust control [1]. For example, an aircraft approximated as a linear model can be rendered inaccurate due to continuous wind disturbances, but its parameters can be updated online to compensate [2]. This approach is vital in many fields, including robotics [3], biomedical engineering [4], and finance [5].

Model identification adaptive control (MIAC) extends system identification by simultaneously learning system parameters and controlling the system to complete a desired objective [6]. Though its online parameter estimation allows it to adapt to external disturbances and time-varying dynamics, typical MIAC methods assume full observability [7]–[9], which is problematic in settings with limited state information. Moreover, their extension to continuous state and action spaces can introduce intense computational effort.

We address planning in partially observable environments with continuous state and action spaces. We treat system parameters as hidden state variables, which we estimate with informative input design, which selects controls that maximize information gain [2]. We reformulate system identification under informative input design and MIAC as partially observable Markov decision processes with belief-state dependent rewards (ρ-POMDPs [10]), where belief-states include both the mean and uncertainty of true states. Since belief-state dynamics are nonlinear, stochastic, and underactuated, we efficiently solve the ρ-POMDP using an adapted belief-state iterative Linear Quadratic Regulator (BiLQR) that plans over Gaussian beliefs and penalizes uncertainty [11]. This approach quickly identifies system parameters by prioritizing information-gathering actions while maintaining effective control through robust state estimation.

We present our approach on both fully and partially observable problems for a cart-pole balancing example and a steady aircraft flight example. In these problems, BiLQR outperforms common approaches that combine filtering or regression for informative input design and model predictive control approaches for MIAC. BiLQR not only achieves control objectives while identifying system parameters more quickly and accurately, but also adapts to time-varying dynamics. To our knowledge, this is the first work to formulate MIAC as a ρ-POMDP jointly over system states and parameters, providing efficient solutions to MIAC problems while reasoning over uncertainty about system parameters.

In summary, our contributions include:

- formulating system identification and MIAC as a ρ-POMDP over system states and parameters,
- solving the ρ-POMDP using BiLQR more accurately than the baselines, and
- demonstrating our approach in fully and partially observable settings, and with time-varying dynamics.

In this paper, we first provide an overview of system identification, MIAC, and POMDPs. We then discuss our approach for formulating MIAC problems as ρ-POMDPs and solving them with BiLQR. Finally, we demonstrate our approach against various baselines in fully and partially observable environments with unknown system parameters.

II. Background

A. System Identification and Adaptive Control

System identification uses observed data to learn parameters θ that dictate a dynamics model $f(s,a)$ [1]. Methods for fully observable systems include regression [12], parametric modeling [13], and informative input design, which selects controls to maximize information about model parameters [2]. Some filtering methods, like Kalman Filtering, can address partial observability with these methods [14]. However, decoupling parameter and state estimation can lead to suboptimal information flow, limiting the accuracy of both parameter and state estimates.

Several methods also combine model identification and control. Many decouple system identification from control for quicker estimation, such as by using expectation maximization [15], planning for worst-case parameter estimation or distributions [7], optimizing data generation for better estimation [8], or sampling-based methods [9]. Model reference

[1] Stanford University, Stanford, CA 94305 USA {mtho, arec, mykel}@stanford.edu

[2] TerraAI, Redwood City, CA 94063 USA

*Toyota Research Institute provided funds to support this work.

adaptive control (MRAC) selects controls so the system follows a reference model that represents the desired behavior [16]. Model identification adaptive control simultaneously learns system parameters and controls the system [6]. Compared to MRAC, MIAC's online estimation allows flexible responses to disturbances and time-varying dynamics [6]. Still, existing MIAC methods do not explicitly account for partial observability or consider actions that balance control and parameter uncertainty reduction simultaneously.

B. POMDPs

A partially observable Markov decision process (POMDP) is a formulation for systems with observations that provide incomplete information about the state. A finite horizon POMDP is defined by the tuple $(\mathscr{S},\mathscr{A},\mathscr{O},T,O,R,\tau)$ consisting of state, action, and observation spaces $\mathscr{S},\mathscr{A},\mathscr{O}$, a dynamics model T mapping states and actions to a distribution over resulting states, an observation model O mapping an underlying transition to a distribution over emitted observations, a reward function R that maps the underlying transition to an incremental reward, and finite time horizon τ. A policy π generates actions from an initial state distribution b_0, a history of actions $a_{0:t}$, and observations $o_{1:t}$, which together can be represented concisely as an instantaneous belief distribution over states b_t where $b_t(s) = p(s_t = s \mid b_0, a_{0:t}, o_{1:t})$. An optimal policy maximizes the expected cumulative reward over the time horizon:

$$\max_{\pi} V^{\pi}(b_0) = \mathbb{E}_{T,O}\left[\sum_{t=0}^{\tau} R(b_t,a_t) \mid b_0\right], \tag{1}$$

where the belief-based reward function returns the expected reward from transitions from states in the belief [17].

Traditional POMDPs model state-based rewards and cannot explicitly model an agent's goal to reduce state uncertainty. To address this, ρ-POMDPs modify the standard POMDP framework by introducing belief-dependent rewards that favor actions that reduce the uncertainty (i.e., entropy) in the belief state [10]. They have been applied to active sensing problems to balance between strategically gathering information and completing tasks [18]. This work can be seen as a dual control problem with belief-dependent rewards.

Exact planning in large POMDPs is generally intractable. While small, discrete POMDPs can be solved offline with approximate methods [19], larger ones require online, receding-horizon planning [20]. Sampling-based tree search extends to continuous state and action spaces [9] but struggles in high-dimensional control problems. Trial-and-error techniques like reinforcement learning can learn effective policies in large problems, but can be very data inefficient [21]. Under certain assumptions, these problems can be addressed using optimal control methods like model predictive control (MPC) or Linear Quadratic Regulator (LQR) from a mean state estimate [22]. These methods can be extended to plan over nonlinear dynamics with iterative linearization (e.g. iLQR), or to plan over full belief-state dynamics rather than planning from a mean (BiLQR) [11].

III. Methodology

A. Problem Formulation

As stated previously, the goal of system identification is to learn the system parameters. The goal of *informative input design* for system identification is to choose actions that decrease uncertainty over system parameters. Consider a prior distribution over possible system parameters $p(\theta)$ with shorthand b_θ. Let $b_{\theta,\tau}$ be the shorthand for the posterior distribution over system parameters consistent with observations through the time horizon τ. With an uncertainty measure H of the belief (e.g., Shannon entropy), we can formulate the objective for active system identification as:

$$\min_{\pi} \mathbb{E}[H(b_{\theta,\tau})]. \tag{2}$$

That is, the optimal policy π results in minimal expected uncertainty in the parameters at the final time step τ.

We may also consider identifying system parameters quickly, possibly in the presence of large, instantaneous disturbances to system parameters. We can accomplish this with an objective that minimizes stagewise uncertainty:

$$\min_{\pi} \mathbb{E}\left[\sum_{t=1}^{\tau} H(b_{\theta,\tau})\right]. \tag{3}$$

The goal in MIAC is to reduce uncertainty in the parameters and simultaneously control the system. Consider the system state x and stagewise costs $C_t(x,a)$. The objective is:

$$\min_{\pi} \mathbb{E}\left[\sum_{t=1}^{\tau} C_t(x_t,a_t) + \lambda H(b_{\theta,\tau})\right], \tag{4}$$

with trade-off hyperparameter λ. That is, we wish to determine the policy π that optimizes the control objective over the time horizon while reducing the uncertainty in the system parameters at the final time step.

B. ρ-POMDP Formulation

We propose solving the MIAC objective in Eq. (4) by formulating the problem as a ρ-POMDP that combines system states and parameters. Given a joint state $s = [x, \theta]$, we can define belief-based rewards as the stagewise cost of the MIAC objective in Eq. (4):

$$R(b,a) = -\mathbb{E}_{x\sim b}[C(x,a)] - \lambda H(b_\theta), \tag{5}$$

where b_θ marginalizes the joint belief over θ. As is common in informative input design [18], we assume Gaussian beliefs.

$$b(s) = \mathscr{N}\left(s \mid \mu_s = \begin{pmatrix}\mu_x \\ \mu_\theta\end{pmatrix}, \Sigma_s = \begin{pmatrix}\Sigma_{xx} & \Sigma_{x\theta} \\ \Sigma_{\theta x} & \Sigma_{\theta\theta}\end{pmatrix}\right), \tag{6}$$

with shorthand $b = [\mu_s\ \Sigma_s^*]^\top$, where Σ_s^* is the covariance matrix Σ_s flattened. Assuming nonlinear Gaussian dynamics and measurements, we write the joint transition function as:

$$T(s' \mid s,a) = \mathscr{N}\left(\begin{pmatrix}x' \\ \theta'\end{pmatrix} \mid \mu = \begin{pmatrix}f(x,a,\theta) \\ \theta\end{pmatrix},\right. \tag{7}$$

$$\left.\Sigma = \begin{pmatrix}W_x & 0 \\ 0 & W_\theta\end{pmatrix}\right) \tag{8}$$

where $f(x,a,\theta)$ is the deterministic part of the dynamics function, and W_x and W_θ are process noise matrices for x and θ respectively. We can plan for disturbances to system parameters by assuming a nonzero W_θ. We will use $\bar{f}(s,a)$ as a shorthand for the deterministic part of the joint transition.

We can also define a Gaussian observation function:

$$O(o \mid s,a) = O(o \mid x,a,\theta) \tag{9}$$
$$= \mathcal{N}(o \mid \mu = g(x,a), \Sigma = V), \tag{10}$$

where $g(x,a)$ is the deterministic part of the observation function and V is the observation noise. For fully observable systems, we approximate the true observation function $O(o \mid x,\theta,a) = \delta(o = x)$ with a Gaussian centered at x' with an arbitrarily small, user-defined covariance.

C. Belief-state iLQR

We propose solving this reformulated MIAC problem by adapting Platt Jr. *et al.*'s BiLQR approach to plan over system state and parameter beliefs with a receding horizon.

Algorithm 1 Belief-state iLQR planning applied to MIAC

```
Require: Nominal control sequence, $(\bar{a}_0, \ldots, \bar{a}_{\tau-1})$, Current Belief State, $b_0 = b_0(x,\theta) = [\mu_{x,0}, \mu_{\theta,0}, \Sigma^*_{x,0}]$
Set $\bar{b}_0 = b_0$ and $\delta a_t = 0$ for all $t \in \{0, \ldots, \tau-1\}$
while not converged
    Forward pass: $\forall t \in \{0, \ldots, \tau-1\}$
    Compute nominal trajectory $\bar{b}_{t+1} = F(\bar{b}_t, \bar{a}_t + \delta a_t)$
    Set $\bar{a}_t \leftarrow \bar{a}_t + \delta a_t$
    Compute matrices $\tilde{A}_t$ and $\tilde{B}_t$ from Eq. (13)
    Compute reward, $R_t(b,a)$ around $\bar{b}_t, \bar{a}_t$
    Backward pass: $\forall t \in \{\tau-1, \ldots, 0\}$
    Update value approximation as in Eq. 15 of [11]
    Update feedback law $\delta a_t$ as in Eq. 16 of [11]
return $\bar{a}_0$
```

Algorithm 1 depicts the adaptation of the BiLQR planner to MIAC. It initializes a nominal trajectory over a receding time horizon by forward-propagating the initial belief with zero control input. Then, it executes forward and backward passes, starting with the initial belief state to update the actions until the optimal sequence is found. In the initialization and forward pass, the belief-state dynamics $F(b,a)$ propagate the belief state over the time horizon. The dynamics are

$$b_{t+1} = F(b_t, a_t) = \begin{pmatrix} \bar{f}(\mu_{s,t}, a_t) \\ \mu_{\theta,t} \\ \Sigma^*_{t+1} \end{pmatrix}, \tag{11}$$

$$\Sigma_{t+1} = (I - (A_t \Sigma_t A_t^\top + W_x) C_t^\top (C_t (A_t \Sigma_t A_t^\top + W_x) C_t^\top + V)^{-1} C_t)(A_t \Sigma_t A_t^\top + W_x), \tag{12}$$

where A_t is the joint dynamics Jacobian matrix and C_t is the observation Jacobian matrix. The belief-state dynamics are linearized around the nominal trajectory. The linearized belief-state space matrices $\tilde{A}$ and $\tilde{B}$ are

$$\begin{aligned} \tilde{A}_t &= \frac{\partial F}{\partial b} \Big|_{(\bar{b}_t, \bar{a}_t)} = \begin{pmatrix} \frac{\partial f}{\partial x_t} & 0 & 0 \\ 0 & I & 0 \\ \frac{\partial \Sigma^*_t}{\partial x_t} & \frac{\partial \Sigma^*_t}{\partial \theta_t} & \frac{\partial \Sigma^*_t}{\partial \Sigma_t} \end{pmatrix} \\ \tilde{B}_t &= \frac{\partial F}{\partial a} \Big|_{(\bar{b}_t, \bar{a}_t)} = \begin{pmatrix} \frac{\partial f}{\partial a_t} \\ 0 \\ 0 \end{pmatrix}, \end{aligned} \tag{13}$$

where Σ^*_t is the stacked column representation of the covariance matrix Σ. The forward pass propagates the mean of the belief forward with the linearized state dynamics and the covariance through an extended Kalman filter assuming maximum-likelihood observations $O = g(x,a)$.

The backward pass refines the policy using the quadratic reward that updates the optimal actions from the initial guess. In maximizing the reward, the resulting plan better aligns with the system dynamics and mitigates uncertainty.

If only system identification is considered, the goal is to reduce the uncertainty in the system parameters. Thus, the per-time-step reward function is

$$R_t(b,a) = \mathbf{1}(t = \tau) {\Sigma^*}^\top_{\theta\theta,t} \Lambda \Sigma^*_{\theta\theta,t}, \tag{14}$$

where Λ is a negative semi-definite matrix that penalizes uncertainty in system parameters at time τ, measured by $\Sigma_{\theta\theta,\tau}$, the sub-vector of Σ^*_t for just the covariance between the unknown system parameters.

For MIAC, the stagewise reward function is

$$\begin{aligned} R_t(b,a) = {} & (\mu_{x,t} - \mu_{goal})^\top \hat{Q}_t (\mu_{x,t} - \mu_{goal}) \\ & + a_t^\top \hat{R}_t a_t + \mathbf{1}(t = \tau) {\Sigma^*}^\top_{\theta\theta,t} \Lambda \Sigma^*_{\theta\theta,t}, \end{aligned} \tag{15}$$

which incorporates negative semi-definite $\hat{Q}_t$ and $\hat{R}_t$ state and action cost matrices from traditional LQR. This formulation is equivalent to Eq. (4), with $H(b_\theta)$ as a quadratic cost function on system parameter uncertainty. The forward and backward passes are performed until the optimal control sequence converges. Only the first action from the plan is executed. We update the belief and replan after receiving new state information. Platt Jr. *et al.* show that once the belief covariance is sufficiently reduced, the system enters a locally linear regime where LQR guarantees exponential convergence of the belief mean to the target. BiLQR, like iLQR, is not globally optimal but provides locally optimal solutions around a nominal belief trajectory.

The ρ-POMDP formulation of MIAC and the BiLQR planner allows efficient solutions in both fully and partially observable domains with continuous state spaces.

IV. Experiments

Our experiments consider both fully and partially observable systems to empirically demonstrate our adapted BiLQR's performance. As expected, performance degrades under partial observability compared to full observability, but BiLQR is more successful, and this case provides a strong lower-bound baseline for comparison. We compare informative input design performance against filtering and regression baselines and MIAC performance against model predictive control (MPC) with regression and extended Kalman filter (EKF) baselines. Moreover, we demonstrate robustness to time-varying disturbances in system parameters.

We use the `POMDPs.jl` framework for our experiments [23]. For full experimental details, including cost matrices and hyperparameters, see `https://github.com/sisl/MIAC_BiLQR/`.

A. Problem Domains

We discuss the problem domains used in our experiments, the POMDP definition, their fully and partially observable variations, and the system parameters to be identified.

1) Cart-pole: In this one-dimensional traditional control problem, a cart moves along a track, and a pole, attached by a pivot, must be balanced upright. The state of the system x_t is described by the cart's position p_t, the pole's angle ψ_t, and their respective velocities $\dot{p}_t$ and $\dot{\psi}_t$. Control is achieved by applying a one-dimensional force a to the cart [24]. We assume the mass of the pole is unknown and seek to identify the log mass, $\theta = \log m_p$.

The dynamics follow $x_{t+1} = x_t + \dot{x}_t \delta t$ where

$$\dot{x}_t = \begin{bmatrix} \dot{p}_t, \\ \dot{\psi}_t, \\ \frac{m_p \sin\psi_t \left(L\dot{\psi}_t^2 + g\cos\psi_t\right) + a}{h} \\ -\frac{(m_c+m_p) g \sin\psi_t + m_p L \dot{\psi}_t^2 \sin\psi_t \cos\psi_t + a\cos\psi_t}{hL} \end{bmatrix}. \tag{16}$$

Reward is given for minimizing control effort and keeping the pole within 12° of the vertical [24]. The partially observable setting includes noisy observations of cart and pole position, i.e., $g(x,a) = [p, \psi]^\top$ with Gaussian noise V.

2) Aircraft Steady Flight: We adapt the linear aircraft informative input design problem from Ott *et al.* to a longitudinal motion model with three degrees of freedom. The goal is to identify the linear state-space dynamics defined by the transition matrix Φ_1 and input matrix Φ_2. The state x_t consists of horizontal and vertical velocity perturbations u_t and w_t, angle of attack α_t, and pitch rate $\dot{\alpha}_t$. The action a_t includes elevator deflection δ_e and throttle δ_{th}. To test the method on multiple unknown parameters while balancing computational efficiency, we treat the first columns of Φ_1 and Φ_2 as the unknown parameters.

The state evolution is governed by the equation:

$$x_{t+1} = \Phi_1 x_t + \Phi_2 a_t.$$

Reward is given for keeping α_t within 0° to 30°, maintaining non-negative vertical velocity, and minimizing control effort. In the partially observable setting, all states except pitch rate are observed, i.e., $g(x,a) = [u, w, \alpha]^\top$ with Gaussian noise V.

B. Experiments and Discussion

1) Informative Input Design Comparison: We evaluate BiLQR on informative input design in fully and partially observable environments using (1) the trace of the final system parameter covariance, $\Sigma_{\theta\theta,T}$ (zero indicates complete uncertainty reduction), and (2) the log likelihood of the true parameter, $\log p_{\hat{\theta},\tau}$ (higher implies more accurate estimates). We benchmark BiLQR against approximate least squares (regression) and an EKF applied to the joint state under a random policy sampled from $\mathscr{U}(a_{\min}, a_{\max})$.

Table I summarizes the performance of the algorithms on the domains, averaged across 150 simulations per experiment. BiLQR significantly outperforms both baselines at reducing uncertainty in system parameters, as shown by the lower trace of the system parameter covariance matrix. Additionally, the log likelihood at the final time step for the system parameters is the highest for BiLQR, indicating a higher probability that the system parameters learned by BiLQR are accurate. In the partially observable environments, the uncertainty is higher, and the log likelihood is lower than in the fully observable counterparts, indicating that it was more difficult to identify the system parameters. Still, BiLQR outperformed the EKF baseline in this setting.

2) Model Identification Adaptive Control Comparison: Our evaluation metrics for MIAC are the trace of $\Sigma_{\theta\theta,T}$, $\log p_{\hat{\theta},\tau}$, and the expected reward over the time horizon $\mathbb{E}[R(b,a)]$, indicating success in achieving the control objective. For the cart-pole domain, the objective is to balance the pole at $\psi = \pi/2$. For the aircraft domain, the objective is to reach a forward speed of 100 m/s and reduce vertical velocity, angle of attack, and pitch angle rate to zero. We benchmark BiLQR against both a least squares approximation and using an EKF when following model predictive control (by solving a quadratic program) and also using an EKF when following a random policy.

Table II summarizes the performance of the algorithms on the different domains, averaged across 150 simulations per domain and observability variation. In both problem domains with both full and partial observability, we see that BiLQR reduces uncertainty, as shown by the low trace of the system parameter covariance matrix. It performed comparably to MPC with EKF updates, indicating it learned the system parameters just as effectively. Also, as indicated by the higher average reward per time step, BiLQR drives the system to achieve the objective better than all of the baselines.

Figure 1 shows performance over time for the cart-pole fully observable environment. BiLQR does better than EKF and both MPC baselines in reducing the uncertainty about the pole mass, as shown by the higher final log likelihood. This measure indicates that the BiLQR estimate achieved for the mass at the final time step is most accurate. In this plot, BiLQR's log likelihood remains higher than the other baselines for most of the time horizon, indicating that it finds more accurate estimates earlier than the baselines.

Figure 2 shows performance over time for the partially observable aircraft environment. BiLQR performs better than EKF and MPC with EKF in reducing the uncertainty about unknown elements of the linear state space Φ_1 and Φ_2 matrices, as shown by the higher log likelihood across most of the time horizon and at the final time step. This log likelihood was calculated from the multivariate Gaussian distribution defined by the vector of means for the unknown elements and the associated covariance matrix. BiLQR's likelihood also

	Cart-Pole Full Obs		Cart-Pole Partial Obs		Aircraft Full Obs		Aircraft Partial Obs	
Solver	$\mathrm{tr}(\Sigma_{\theta\theta,\tau})$	$\log p_{\hat{\theta},\tau}$	$\mathrm{tr}(\Sigma_{\theta\theta,\tau})$	$\log p_{\hat{\theta},\tau}$	$\mathrm{tr}(\Sigma_{\theta\theta,\tau})$	$\log p_{\hat{\theta},\tau}$	$\mathrm{tr}(\Sigma_{\theta\theta,\tau})$	$\log p_{\hat{\theta},\tau}$
BiLQR	**0.012**±0.001	**0.514**±0.147	**0.108**±0.007	**−2.396**±1.824	**7.994**±0.470	**−4.976**±0.425	**22.604**±0.816	**−7.792**±1.202
Random + EKF	0.022±0.002	−39.945±29.861	0.132±0.011	−4.507±2.932	32.329±0.148	−31.219±2.613	45.970±0.079	−17.837±1.937
Regression	0.172±0.081	−35.235±3.966	–	–	49.480±0.234	−28.865±1.362	–	–

TABLE I: Informative input design with ρ-POMDP planning, reported mean and standard error of $\Sigma_{\theta\theta,T}$ and $\log p_{\hat{\theta},\tau}$ over the predicted distribution across 150 simulations using BiLQR with a covariance minimizing objective, random policy with EKF, and linear regression.

	Cart-Pole Full Obs			Cart-Pole Partial Obs		
Solver	$\mathrm{tr}(\Sigma_{\theta\theta,\tau})$	$\log p_{\hat{\theta},\tau}$	$R(b,a)$	$\mathrm{tr}(\Sigma_{\theta\theta,\tau})$	$\log p_{\hat{\theta},\tau}$	$R(b,a)$
BiLQR	**0.010**±0.001	**0.022**±0.328	**4.251**±2.050	**0.140**±0.010	**−0.699**±0.169	**4.035**±2.046
MPC + Regression	0.096±0.012	−46.671±6.270	3.687±1.635	-	-	-
MPC + EKF	0.011±0.001	−6.497±5.102	3.748±2.185	0.145±0.010	−13.114±9.782	3.863±2.052
Random + EKF	0.022±0.002	−39.945±29.861	3.584±1.804	0.132±0.011	−4.507±2.932	3.606±2.141

	Aircraft Full Obs			Aircraft Partial Obs		
Solver	$\mathrm{tr}(\Sigma_{\theta\theta,\tau})$	$\log p_{\hat{\theta},\tau}$	$R(b,a)$	$\mathrm{tr}(\Sigma_{\theta\theta,\tau})$	$\log p_{\hat{\theta},\tau}$	$R(b,a)$
BiLQR	**2.057**±0.120	**−3.377**±3.353	**77.953**±12.905	**22.609**±0.334	**−7.919**±0.900	**78.429**±12.471
MPC + Regression	44.974±0.289	−180.014±10.128	3.707±0.902	–	–	–
MPC + EKF	9.440±0.212	−7.578±3.010	56.040±18.239	20.741±0.146	−8.035±6.288	58.033±19.242
Random + EKF	32.329±0.148	−31.219±2.613	51.707±17.305	45.970±0.079	−17.837±1.937	51.967±16.678

TABLE II: MIAC with ρ-POMDP planning, reported mean and standard error of $\Sigma_{\theta\theta,T}$ and $\log p_{\hat{\theta},\tau}$ over 150 simulations using BiLQR with a covariance-minimizing objective, MPC with linear regression and EKF, and a random policy with EKF.

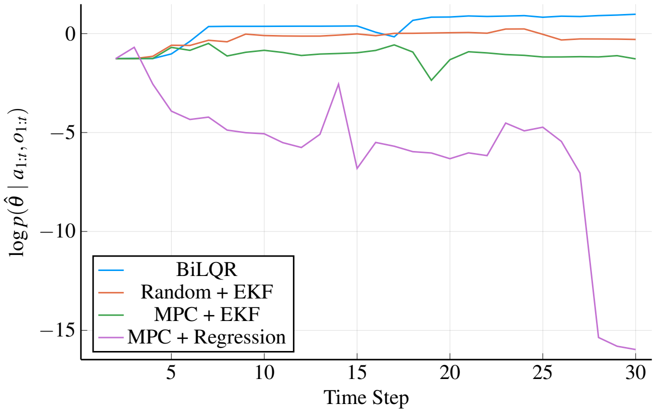

Fig. 1: MIAC $\log p(\hat{\theta} \mid a_{1:t}, o_{1:t})$ for BiLQR, Random + EKF, MPC + EKF, and MPC + Regression in one example simulation of the fully observable cart-pole environment (showing the mean trend with error would obscure the BiLQR trend, since the baseline uncertainties in $\log p_{\hat{\theta},\tau}$ are orders of magnitude larger than BiLQR).

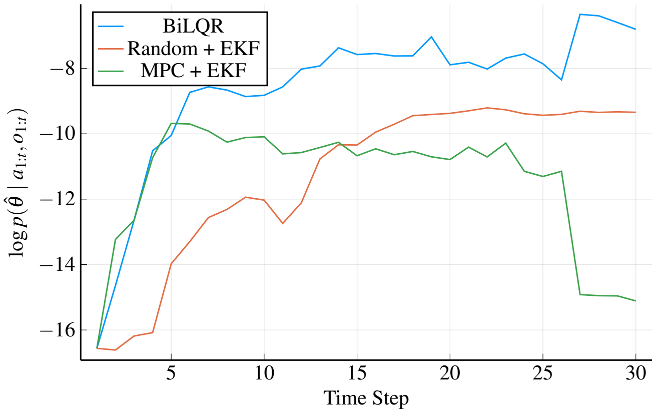

Fig. 2: MIAC $\log p(\hat{\theta} \mid a_{1:t}, o_{1:t})$ for BiLQR, Random + EKF, and MPC + EKF in one example simulation of the partially observable aircraft environment.

rises faster than the baselines, so more accurate estimates were found earlier in this example as well.

3) System Identification during Disturbances: In the ρ-POMDP formulation, a nonzero W_θ enables robust planning against parameter disturbances. For example, although the cart-pole mass remains constant in practice, a large disturbance may alter its modeled value to better model the dynamics. We simulate a sudden change in mass to test our approach with non-stationary parameters. As shown in Figure 3, BiLQR quickly tracks the change and reduces uncertainty about the pole's mass, with a slight delay to first detect and then adjust to the abrupt shift using observed data.

V. Conclusion

We framed informative input design and model identification adaptive control in fully and partially observable systems as a ρ-POMDP, with system parameters as hidden variables and a belief-based objective to balance control with system identification. Our experimental results on both fully observable and partially observable domains demonstrate that

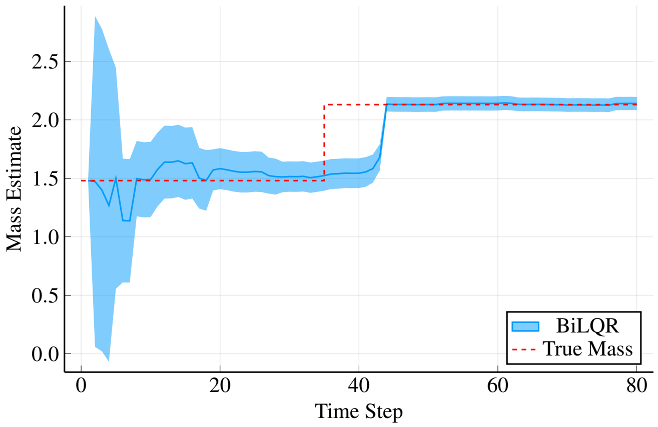

Fig. 3: Tracking change in pole mass that changes suddenly at time $t = t_c$ for one example simulation with BiLQR.

our adapted BiLQR significantly outperforms baselines, particularly in scenarios with high measurement noise. We also demonstrate robustness to time-varying system parameters.

Limitations of our approach include its local optimality and the computational challenges of scaling to high-dimensional state or parameter spaces. Real-time feasibility may be improved through warm-started linearizations or low-rank approximations. Framing MIAC as a ρ-POMDP also invites opportunities for incorporating constraints and exploring other belief-space planning approaches.

References

[1] L. Ljung, *System Identification: Theory for the User*. Prentice Hall, 1999.

[2] J. Ott, M. J. Kochenderfer, and S. Boyd, "Informative input design for dynamic mode decomposition," in *Conference on Learning for Dynamics and Control (L4DC)*, 2025.

[3] R. Johansson, A. Robertsson, K. Nilsson, and M. Verhaegen, "State-space system identification of robot manipulator dynamics," *Mechatronics*, vol. 10, no. 3, pp. 403–418, 2000.

[4] A. D. Wiens and O. T. Inan, "A novel system identification technique for improved wearable hemodynamics assessment," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 5, pp. 1345–1354, 2015.

[5] C. A. Los, "System identification in noisy data environments: An application to six Asian stock markets," *Journal of Banking & Finance*, vol. 30, no. 7, pp. 1997–2024, 2006.

[6] Z. Öreg, H.-S. Shin, and A. Tsourdos, "Model identification adaptive control: Implementation case studies for a high manoeuvrability aircraft," in *Mediterranean Conference on Control and Automation*, 2019, pp. 559–564.

[8] B. Wahlberg, H. Hjalmarsson, and M. Annergren, "On optimal input design in system identification for control," in *IEEE Conference on Decision and Control (CDC)*, 2010, pp. 5548–5553.

[9] P. Slade, P. Culbertson, Z. Sunberg, and M. Kochenderfer, "Simultaneous active parameter estimation and control using sampling-based Bayesian reinforcement learning," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 804–810.

[7] H. Nishimura, N. Mehr, A. Gaidon, and M. Schwager, "RAT iLQR: A risk auto-tuning controller to optimally account for stochastic model mismatch," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 763–770, 2021.

[10] M. Araya, O. Buffet, V. Thomas, and F. Charpillet, "A POMDP extension with belief-dependent rewards," in *Advances in Neural Information Processing Systems (NIPS)*, 2010.

[11] R. Platt Jr., R. Tedrake, L. Kaelbling, and T. Lozano-Perez, "Belief space planning assuming maximum likelihood observations," in *Robotics: Science and Systems*, 2010.

[12] R. E. Kopp and R. J. Orford, "Linear regression applied to system identification for adaptive control systems," *AIAA Journal*, vol. 1, no. 10, pp. 2300–2306, 1963.

[13] D. Gedon, N. Wahlström, T. B. Schön, and L. Ljung, "Deep state space models for nonlinear system identification," in *International Federation of Automated Control Symposium on System Identification*, 2021, pp. 481–486.

[14] J. Valasek and W. Chen, "Observer/Kalman filter identification for online system identification of aircraft," *AIAA Journal of Guidance, Control, and Dynamics*, vol. 26, pp. 347–353, 2003.

[15] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains," *The Annals of Mathematical Statistics*, vol. 41, no. 1, pp. 164 –171, 1970.

[16] M. Schreier, "Modeling and adaptive control of a quadrotor," in *IEEE International Conference on Mechatronics and Automation*, 2012, pp. 383–390.

[17] K. J. Åström, "Optimal Control of Markov Processes with Incomplete State Information I," *Journal of Mathematical Analysis and Applications*, vol. 10, pp. 174–205, 1965.

[18] J. Ott, E. Balaban, and M. J. Kochenderfer, "Sequential bayesian optimization for adaptive informative path planning with multimodal sensing," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[19] H. Kurniawati, D. Hsu, and W. S. Lee, "SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces," in *Robotics: Science and Systems*, 2009.

[20] D. Silver and J. Veness, "Monte-Carlo planning in large POMDPs," in *Advances in Neural Information Processing Systems (NIPS)*, 2010.

[21] M. P. Deisenroth and C. E. Rasmussen, "PILCO: A model-based and data-efficient approach to policy search," in *International Conference on Machine Learning (ICML)*, 2011, 465–472.

[22] O. R. González and A. G. Kelkar, "Robust multivariable control," in *The Electrical Engineering Handbook*, Academic Press, 2005, pp. 1037–1047.

[23] M. Egorov, Z. N. Sunberg, E. Balaban, T. A. Wheeler, J. K. Gupta, and M. J. Kochenderfer, "POMDPs.jl: A framework for sequential decision making under uncertainty," *Journal of Machine Learning Research*, vol. 18, no. 26, pp. 1–5, 2017.

[24] A. G. Barto, R. S. Sutton, and C. W. Anderson, "Neuronlike adaptive elements that can solve difficult learning control problems," *IEEE Transactions on Systems, Man, and Cybernetics*, pp. 834–846, 1983.