

Stability of Certainty-Equivalent Adaptive LQR for Linear Systems with Unknown Time-Varying Parameters

Marcell Bartos^{1,2}

Johannes Köhler^{1,3}

Florian Dörfler²

Melanie N. Zeilinger¹

MBARTOS@ETHZ.CH

J.KOHLER@IMPERIAL.AC.UK

DOERFLER@CONTROL.EE.ETHZ.CH

MZEILINGER@ETHZ.CH

¹: Institute for Dynamic Systems and Control, ETH Zürich, Zürich, Switzerland

²: Automatic Control Laboratory, ETH Zürich, Zürich, Switzerland

³: Department of Mechanical Engineering, Imperial College London, London, UK

Editors: G. Sukhatme, L. Lindemann, S. Tu, A. Wierman, N. Atanasov

Abstract

Standard model-based control design deteriorates when the system dynamics change during operation. To overcome this challenge, online and adaptive methods have been proposed in the literature. In this work, we consider the class of discrete-time linear systems with unknown time-varying parameters. We propose a simple, modular, and computationally tractable approach by combining two classical and well-known building blocks from estimation and control: the least mean square filter and the certainty-equivalent linear quadratic regulator. Despite both building blocks being simple and off-the-shelf, our analysis shows that they can be seamlessly combined to a powerful pipeline with stability guarantees. Namely, finite-gain ℓ^2 -stability of the closed-loop interconnection of the unknown system, the parameter estimator, and the controller is proven, despite the presence of unknown disturbances and time-varying parametric uncertainties. Real-world applicability of the proposed algorithm is showcased by simulations carried out on a nonlinear planar quadrotor.

Keywords: adaptive control, parametric uncertainty, online learning for control, unknown systems, linear time-varying systems, linear quadratic regulator

1. Introduction

Many successful control designs, such as the linear quadratic regulator (LQR) (Anderson and Moore, 2007), or model predictive control (Rawlings et al., 2020), require a known model of the controlled system. However, in real world scenarios, it is often cumbersome and expensive to derive accurate models from first principles.

Data-driven methods (Hou and Wang, 2013) have been proposed to address this issue, leveraging an offline-collected dataset to either directly synthesize the controller (De Persis and Tesi, 2019; van Waarde et al., 2025), or indirectly, via a system identification step (Ljung, 1998; van Overschee and De Moor, 2012). To account for the remaining model uncertainty, usually a robust control approach is considered (Berberich et al., 2020; van Waarde et al., 2020), which is only feasible if the error is small. In addition, offline designs fail if the system behavior changes online, or if no stable open-loop experiments can be conducted. To overcome these challenges, online and adaptive algorithms are needed that update the model and the policy to improve performance over time.

Reinforcement learning (Sutton and Barto, 1998) is another popular approach to address the lack of an accurate model. These methods aim to learn an optimal policy via repeated interaction with the environment, usually in an episodic and simulated manner, leveraging the availability of a vast

collection of data and increasing computational power. However, the vast majority of reinforcement learning methods work offline (Levine et al., 2020), rendering them incapable of adapting to online changes in the system dynamics. Furthermore, they generally only achieve the desired guarantees *after training*, whereas failure (e.g. in terms of instability or arbitrarily large cost) is admissible *during training*. Transferring these techniques to the online setting where learning happens through real-world experiments is challenging, as in this case transient behavior is important, and resetting the experiments at the start of every episode is not always possible.

In this work, we address the optimal control of discrete-time linear systems, when the model parameters are unknown and time-varying.

1.1. Related Work

Classical adaptive control: The well-established field of adaptive control (Aström and Wittenmark, 1994; Krstic et al., 1995; Narendra and Annaswamy, 2012) aims to stabilize systems with parametric uncertainties by continuously adapting the controller. Limitations include the fact that the parameter adaptation rule and the control policy have to be simultaneously designed and tailored to specific, often restrictive problem classes (e.g., assuming that the system is minimum phase). Additionally, optimality of the controller is rarely an objective of classical adaptive control, and most results are limited to continuous-time systems. For example, discrete-time gradient descent requires a good choice of the step size, unlike gradient flows in continuous time.

Online learning the LQR: In the past decade there has been an increased interest in applying online learning techniques (Shalev-Shwartz et al., 2012) to learn the LQR for linear time-invariant (LTI) systems. These include methods from the optimism in the face of uncertainty framework (Abbasi-Yadkori and Szepesvári, 2011; Cohen et al., 2019; Bartos et al., 2026); policy gradient methods (Fazel et al., 2018; Zhao et al., 2025); certainty-equivalent control (Wang and Janson, 2021); and policy iteration (Song and Iannelli, 2024). The methods of Abbasi-Yadkori and Szepesvári (2011); Wang and Janson (2021); Bartos et al. (2026) focus on deriving regret bounds, while Cohen et al. (2019); Song and Iannelli (2024); Zhao et al. (2025) also certify stability of the closed loop. However, with the exception of Zhao et al. (2025); Wang and Janson (2021), they carry out updates of the policy in an episodic manner, i.e., only after some (usually increasing) number of time steps. The methods in Zhao et al. (2025); Wang and Janson (2021) require access to a known stabilizing initial controller and work by injecting exploratory noise into the control input to ensure that the data is persistently exciting (Narendra and Annaswamy, 2012; Willems et al., 2005). Finally, a common limitation of all these methods is that they consider the time-invariant case, i.e., they cannot handle online-changing parameters.

Online learning in the time-varying case: A few recent works have addressed the control problem of time-varying systems from the online learning perspective. The methods proposed in Qu et al. (2021); Lin et al. (2021) come with performance guarantees, but require that the true parameter value is revealed to them every time step before selecting the control input. The methods of Minasyan et al. (2021); Gradu et al. (2023) assume that the system is open-loop stable, and Noori and Babazadeh (2025) only consider the finite-horizon problem. Under the assumption of infrequently changing or slowly drifting dynamics, the method of Yu et al. (2023) guarantees stability of the closed-loop system, however, it requires solving a second-order cone program every time step, where the number of constraints grows linearly with time, limiting its real-world applicability.

1.2. Contributions

We propose a simple, modular, and computationally tractable approach by combining two classical and well-known building blocks from estimation and control: the least mean square filter (LMS) (Aström and Wittenmark, 1994) and the certainty-equivalent LQR. Although the LMS and the LQR are widely used, stability guarantees for corresponding adaptive controllers for time-varying parameters are not available in the literature.

In contrast with the aforementioned online-learning-based methods, the proposed method is applicable to linear systems with continuously changing parameters and operates in a non-episodic manner, i.e., by adapting the policy every time step. Moreover, unlike classical adaptive control, our approach is modular, general, uses off-the-shelf building blocks, and addresses the discrete-time setting.

Despite the simplicity of the proposed approach, robust stability of the closed-loop interconnection is proven, even without assuming that the data is persistently exciting. In the case of persistently exciting data, simulation results suggest that the proposed algorithm is also able to learn the model and achieve optimality in the asymptotic sense.

1.3. Outline and Notation

Outline: Section 2 defines the considered problem setting, and Section 3 presents the proposed method as well as its stability analysis. Section 4 contains simulation results on a planar nonlinear quadrotor model. Section 5 concludes the paper and provides an outlook on future directions.

Notation: Let \mathbb{R} and \mathbb{N} denote the set of real and nonnegative integer numbers, respectively. I_n represents the $n \times n$ identity matrix, positive (semi-)definiteness of a matrix Q is indicated by $Q \succ (\succeq) 0$, and the Kronecker product of two matrices is denoted by \otimes . The 2-norm of vector x is denoted by $\|x\|$, while $\|x\|_P := \sqrt{x^\top P x}$ for $P = P^\top \succ 0$. For matrix $B \in \mathbb{R}^{n \times m}$, $\|B\|$ denotes its spectral norm, and its vectorization is denoted by $\text{vec}(B) \in \mathbb{R}^{nm}$, which is obtained by stacking the columns of B on top of one another. For vectors $a \in \mathbb{R}^n, b \in \mathbb{R}^m$, we define $(a, b) := [a^\top b^\top]^\top \in \mathbb{R}^{n+m}$. For a sequence $\{a_0, a_1, \dots\} = \{a_k\}_{k=0}^\infty = \{a\}$, we use $\{a\} \in \ell^1$ or $\{a\} \in \ell^2$ to denote that it is summable or square-summable, respectively.

2. Problem Setup

Consider the discrete-time linear time-varying (LTV) system

$$x_{k+1} = A(\theta_k)x_k + B(\theta_k)u_k + w_k, \quad (1)$$

with state $x_k \in \mathbb{R}^n$, input $u_k \in \mathbb{R}^m$, and unknown disturbance $w_k \in \mathbb{R}^n$ at time step $k \in \mathbb{N}$, and initial condition $x_0 \in \mathbb{R}^n$. We assume that the state can be measured directly and that the matrices $A : \mathbb{R}^p \rightarrow \mathbb{R}^{n \times n}$ and $B : \mathbb{R}^p \rightarrow \mathbb{R}^{n \times m}$ are known functions of the unknown time-varying parameter $\theta_k \in \mathbb{R}^p$. Let $\Delta\theta_k := \theta_{k+1} - \theta_k$.

Assumption 1 (Bounded disturbances) *The disturbances are uniformly bounded, i.e., $\exists \bar{W} \geq 0$ such that $\|w_k\| \leq \bar{W}, \forall k \in \mathbb{N}$.*

Assumption 2 (Parameters) *The parameter vector is contained in a known compact convex set: $\theta_k \in \Theta, \forall k \in \mathbb{N}$. The set Θ is pointwise stabilizable, i.e., $\forall \theta \in \Theta$, the pair $(A(\theta), B(\theta))$ is stabilizable. The matrices $A(\theta), B(\theta)$ depend affinely on θ .*

Note that the pointwise stabilizability condition in Assumption 2 is significantly weaker than assuming the existence of a single controller that simultaneously stabilizes every element of Θ .

Our primary goal is to control system (1) while ensuring ℓ^2 -stability of the closed-loop system with respect to the disturbances w_k , despite the presence of unknown time-varying parameters θ_k . Definition 1 provides the exact definition of ℓ^2 -stability.

Definition 1 (Finite-gain ℓ^2 -stability, adapted from Khalil (2002, Definition 5.1)) Closed-loop system (1) under a controller $u_k = \pi_k(x_k)$ is finite-gain ℓ^2 -stable with respect to the disturbance signal $\{w\}$, if $\exists c_1, c_2 \geq 0$ such that

$$\sum_{k=0}^T \|x_k\|^2 \leq c_1 \sum_{k=0}^T \|w_k\|^2 + c_2, \quad \forall T \in \mathbb{N}, \forall \{w\} \in \ell^2.$$

If the closed-loop system is finite-gain ℓ^2 -stable, then $\lim_{k \rightarrow \infty} x_k = 0$, as long as $\{w\} \in \ell^2$. In our analysis, the constant c_2 in Definition 1 will depend on the parameter variation signal $\{\Delta\theta\}$. Additionally, the policy is to be adapted with the aim to minimize the asymptotic average cost $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^T (x_k^\top Q x_k + u_k^\top R u_k)$, where $Q, R \succ 0$.

3. Proposed Method

The proposed method consists of two parts: the model learner and the policy update. The learner continuously updates an estimate $\hat{\theta}_k$ of the unknown parameter θ_k based on the most recent state transition measurement. This parameter estimate is then fed into the policy update rule, which in turn continuously updates the control policy every time step (see Figure 1).

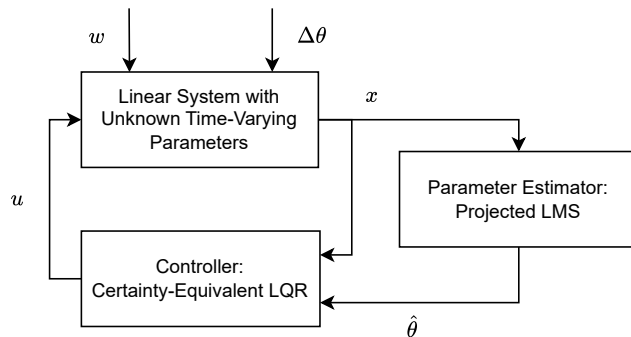


Figure 1: The interconnection of the unknown system, the parameter estimator, and the controller.

3.1. Model Learner

Before presenting the proposed model learner, we equivalently reformulate (1) in a form akin to affine linear regression:

$$x_{k+1} = \delta(x_k, u_k) + D(x_k, u_k)\theta_k + w_k, \tag{2}$$

where $\delta(x_k, u_k) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ and $D(x_k, u_k) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{n \times p}$ are linear and known functions. In the following, system representations (1) and (2) are used interchangeably, and the shorthand $D_k := D(x_k, u_k)$ will be used when convenient.

Define the 1-step nominal prediction of the state at time k as $\hat{x}_{1|k} = A(\hat{\theta}_k)x_k + B(\hat{\theta}_k)u_k$. The proposed model learner is the projected least mean square (LMS) estimator:

$$\hat{\theta}_k = \Pi_{\Theta}[\hat{\theta}_{k-1} + \mu D(x_{k-1}, u_{k-1})^\top (x_k - \hat{x}_{1|k-1})], \quad (3)$$

where $\Pi_{\Theta}[\theta] = \operatorname{argmin}_{\tilde{\theta} \in \Theta} \|\tilde{\theta} - \theta\|$ represents the projection onto the convex set Θ , and $\mu > 0$ is the step size. For the analysis in the following sections, we introduce the following definitions. Let φ_k denote the parameter estimation error at time k , defined as $\varphi_k := \hat{\theta}_k - \theta_k$. Given φ_k , we can define the 1-step nominal prediction error as

$$e_{1|k} := x_{k+1} - \hat{x}_{1|k} - w_k = D_k \theta_k - D_k \hat{\theta}_k = -D_k \varphi_k.$$

An interesting insight is that in the noiseless case ($w_k \equiv 0$), the LMS update is equivalent to a gradient descent step on the squared 1-step prediction error $\|e_{1|k}\|^2$.

The (unprojected) LMS filter is a fundamental tool in recursive parameter estimation (Aström and Wittenmark, 1994). The projected version has recently been applied for parameter estimation in adaptive model predictive control (Lorenzen et al., 2019; Degner et al., 2026). While for general convex sets Θ the projected LMS involves solving a convex optimization problem at every time step, in the common case where Θ is a hyperrectangle, projection reduces to clipping the input vector in an elementwise manner. An advantage of the LMS compared to the more popular recursive least squares (RLS) method is that RLS is unsuitable for the estimation of continuously changing parameters due to its inability to forget. A popular approach to overcome this limitation of RLS is to use exponential forgetting. However, ensuring desirable behavior of the resulting adaptive controller becomes challenging, as one needs to consider how fast the parameters change and tune the forgetting factor accordingly.

3.2. Policy Update

Given the current estimate $\hat{\theta}_k$, the policy is updated to the corresponding certainty-equivalent LQR. In our case, certainty equivalence means that the algorithm pretends that the true parameter does not change over time, and it also disregards the presence of disturbances. The policy update is given by

$$u_k = K_{\text{LQR}}(\hat{\theta}_k)x_k, \quad (4)$$

where $K_{\text{LQR}}(\theta)$ represents the LQR gain corresponding to the LTI system $A(\theta)$, $B(\theta)$ and weight matrices $Q, R \succ 0$, defined as $K_{\text{LQR}}(\theta) = -(R + B(\theta)^\top P(\theta)B(\theta))^{-1}B(\theta)^\top P(\theta)A(\theta)$. Here $P(\theta)$ denotes the unique positive definite solution to the discrete-time algebraic Riccati equation

$$P(\theta) = A(\theta)^\top P(\theta)A(\theta) - A(\theta)^\top P(\theta)B(\theta)(R + B(\theta)^\top P(\theta)B(\theta))^{-1}B(\theta)^\top P(\theta)A(\theta) + Q, \quad (5)$$

and it serves as the Lyapunov function associated with the LQR:

$$V(x, \theta) = x^\top P(\theta)x, \quad (6)$$

In summary, at each time step, the algorithm pretends to know the model parameter perfectly and that this parameter does not change over time, and selects the corresponding optimal infinite horizon controller. As shown in the following sections, this results in a stable closed-loop behavior. Since both the LMS and LQR are computationally cheap, the policy can be updated at every sampling time, even at high rates. The proposed algorithm is summarized in Algorithm 1.

Algorithm 1: Certainty-Equivalent Adaptive LQR

Input: Parameter set Θ , initial state x_0 , initial parameter estimate $\hat{\theta}_0 \in \Theta$, step size $\mu > 0$;

- 1 **for** $k \in \mathbb{N}$ **do**
- 2 Update policy to the certainty-equivalent LQR: $K_k = K_{\text{LQR}}(\hat{\theta}_k)$;
- 3 Apply control input $u_k = K_k x_k$;
- 4 Measure new state x_{k+1} ;
- 5 Update the parameter estimate $\hat{\theta}_{k+1}$ according to the projected LMS estimator (3);
- 6 **end**

3.3. Closed-loop Analysis

Although system (1) is linear, its closed-loop interconnection with the projected LMS estimator (3) and the certainty-equivalent LQR (4) is nonlinear and time-varying (due to the fact that the control policy is a nonlinear function of the past and the system is time-varying). Consequently, we resort to a Lyapunov-type stability analysis. First, we define the diameter of the set Θ as $d = \max_{\theta_1, \theta_2 \in \Theta} \|\theta_1 - \theta_2\|$, and restate relevant results on the LMS from the literature.

Proposition 2 (LMS bounds (Degner et al., 2026, Proposition 1)) *Let Assumptions 1 and 2 hold. Suppose that the state and the input are uniformly bounded: $\|x_k\| \leq X, \|u_k\| \leq U, \forall k \in \mathbb{N}$. If the step size satisfies*

$$\frac{1}{\mu} \geq \sup_{\|x\| \leq X, \|u\| \leq U} \|D(x, u)\|^2, \quad (7)$$

then

a) the parameter estimation error $\varphi_k = \hat{\theta}_k - \theta_k$ satisfies

$$\|\varphi_{k+1}\|^2 - \|\varphi_k\|^2 \leq -\mu \|e_{1|k}\|^2 + \mu \|w_k\|^2 + 2d \|\Delta\theta_k\|, \quad \forall k \in \mathbb{N}, \quad (8)$$

b) the distance between consecutive estimates is bounded by

$$\|\hat{\theta}_{k+1} - \hat{\theta}_k\| \leq \sqrt{\mu} \|e_{1|k} + w_k\|, \quad \forall k \in \mathbb{N}, \quad (9)$$

c) and $\forall T \in \mathbb{N}$ it holds that

$$\sum_{k=0}^T \|e_{1|k}\|^2 \leq \frac{1}{\mu} \|\varphi_0\|^2 + \sum_{k=0}^T \left(\|w_k\|^2 + \frac{2d}{\mu} \|\Delta\theta_k\| \right). \quad (10)$$

Note that $\|\varphi_k\|^2$ can be thought of as a Lyapunov function for the estimator dynamics (3), and (8) as the corresponding Lyapunov inequality. The right-hand side consists of increase terms as functions of the disturbances w_k and $\Delta\theta_k$, and a semidefinite decrease term: $\|e_{1|k}\|^2 = \|D(x_k, u_k)\varphi_k\|^2 = \varphi_k^\top D_k^\top D_k \varphi_k$ with $D_k^\top D_k \succeq 0$.

Inequality (10) provides a finite-horizon upper bound on the prediction error. Taking the limit $T \rightarrow \infty$, this result implies that the prediction error sequence is square-summable ($\{e\} \in \ell^2$), as long as $\{w\} \in \ell^2$ and $\{\Delta\theta\} \in \ell^1$. Inequality (9) provides a bound on the variation of the parameter

estimates $\hat{\theta}_k$, which will be important in the following closed-loop analyses. Finally, note that Proposition 2 does not require persistency of excitation.

Having derived bounds and a Lyapunov inequality for the LMS estimator, we proceed to derive similar results for the policy update. We will make use of the following technical lemma:

Lemma 3 (Lipschitz continuity of the LQR over Θ) *Let Assumption 2 hold. Then $\exists L_K > 0$ such that*

$$\|K_{\text{LQR}}(\theta') - K_{\text{LQR}}(\theta)\| \leq L_K \|\theta' - \theta\|, \quad \forall \theta, \theta' \in \Theta.$$

Proof Lipschitz continuity can be proven by showing that the Jacobians $\partial \text{vec}(P(\theta)) / \partial \text{vec}(A(\theta))$ and $\partial \text{vec}(P(\theta)) / \partial \text{vec}(B(\theta))$ derived in East et al. (2020, Proposition 2) always exist and are uniformly upper bounded. The detailed proof can be found in Appendix A.2. ■

With the help of Lemma 3, the following result can be proven:

Proposition 4 (LQR Lyapunov decrease) *Let Assumptions 1 and 2 hold. Suppose that the state is uniformly bounded: $\|x_k\| \leq X, \forall k \in \mathbb{N}$. Consider any sequence of parameter estimates $\hat{\theta}_k$ that satisfy the bound (9), and consider the instantaneous and certainty-equivalent LQR update (4). Then there exist constants $\alpha, \beta, \gamma > 0$ such that the following Lyapunov decrease condition is satisfied:*

$$V(x_{k+1}, \hat{\theta}_{k+1}) - V(x_k, \hat{\theta}_k) \leq -\alpha \|x_k\|^2 + \beta \|e_{1|k}\|^2 + \gamma \|w_k\|^2, \quad \forall k \in \mathbb{N}, \quad (11)$$

where $V(x, \theta)$ is defined in (6).

Proof The proof can be found in Appendix A.3. ■

The obvious limitation of directly applying Propositions 2 and 4 is that we do not have an a priori guarantee that the state and the input are bounded at all times. This is in contrast to prior works that utilize the LMS estimator for adaptive model predictive control (Lorenzen et al., 2019; Degner et al., 2026), and overcoming this challenge is one of the key technical contributions of our work. To this end, we define the Lyapunov function candidate for the joint state (x_k, φ_k) as

$$\tilde{V}(x_k, \varphi_k, k) := V(x_k, \varphi_k + \theta_k) + \frac{\beta}{\mu} \|\varphi_k\|^2 = x_k^\top P(\varphi_k + \theta_k) x_k + \frac{\beta}{\mu} \|\varphi_k\|^2, \quad (12)$$

i.e., it is a linear combination of the Lyapunov functions from Propositions 2 and 4. Combining Propositions 2 and 4 results in the main theoretical result of the paper.

Theorem 5 (Stability of the closed loop) *Let Assumptions 1 and 2 hold, and assume that the diameter d of the parameter set Θ is small enough. Consider the closed-loop interconnection of system (1), projected LMS estimator (3), and certainty-equivalent LQR update (4). There exists a small enough step size $\mu > 0$ such that*

- a) *the state and the input are uniformly bounded in time: $\|x_k\| \leq X, \|u_k\| \leq U, \forall k \in \mathbb{N}$ for some $X, U > 0$,*
- b) *there exist constants $\tilde{\alpha}, \tilde{\gamma}, \tilde{\delta}, \tilde{c}, \bar{c} > 0$ satisfying*

$$\tilde{c} \|(x_k, \varphi_k)\|^2 \leq \tilde{V}(x_k, \varphi_k, k) \leq \bar{c} \|(x_k, \varphi_k)\|^2, \quad \forall k \in \mathbb{N}, \quad (13)$$

$$\tilde{V}(x_{k+1}, \varphi_{k+1}, k+1) - \tilde{V}(x_k, \varphi_k, k) \leq -\tilde{\alpha} \|x_k\|^2 + \tilde{\gamma} \|w_k\|^2 + \tilde{\delta} \|\Delta \theta_k\|, \quad \forall k \in \mathbb{N}, \quad (14)$$

c) and there exist constants $c_{1,2,3,4} > 0$ such that $\forall T \in \mathbb{N}$ it holds that

$$\sum_{k=0}^T \|x_k\|^2 \leq c_1 \|x_0\|^2 + c_2 \|\varphi_0\|^2 + c_3 \sum_{k=0}^T \|w_k\|^2 + c_4 \sum_{k=0}^T \|\Delta\theta_k\|, \quad (15)$$

i.e., the state is finite-gain ℓ^2 -stable with respect to the disturbance $\{w\}$ as long as $\{\Delta\theta\} \in \ell^1$.

Proof The proof can be found in Appendix A.4. ■

Inequality (15) provides a finite-horizon upper bound on the state. Taking the limit $T \rightarrow \infty$, this result implies that the state trajectory is square-summable ($\{x\} \in \ell^2$), as long as $\{w\} \in \ell^2$ and $\{\Delta\theta\} \in \ell^1$, similarly to the corresponding result for the prediction error in Proposition 2.

3.4. Discussion

Compared to other non-episodic methods that learn the LQR online (Wang and Janson, 2021; Zhao et al., 2025), the main advantage of the proposed method is its ability to handle continuously changing unknown parameters. Moreover, it does not need access to a stabilizing controller for the unknown system, instead, we require that the parameters lie inside the small enough known set Θ . Additionally, other methods commonly use (recursive) least squares to estimate the parameter, and assume that the disturbance follows an independent and identically distributed zero-mean Gaussian distribution. In contrast, the LMS used by the proposed method is agnostic to distributional assumptions and instead requires boundedness of the disturbances (cf. Assumption 1).

Furthermore, existing methods (e.g., Wang and Janson (2021); Zhao et al. (2025)) need to inject noise to ensure persistently exciting data. In contrast, the proposed method lifts this requirement, allowing it to purely apply the LQR feedback without injecting additional noise. Note that we require a small enough known parameter set Θ . Explicitly characterizing the exact condition on its size is difficult in practice. However, as also showcased in Section 4, this requirement is often satisfied by reasonable uncertainty sets in practical scenarios. A resulting limitation of the proposed method is that it does not guarantee convergence of the parameter estimates. Consequently, the policy is also not guaranteed to converge to the optimal one. On the other hand, this also highlights a further advantage: stability of the closed-loop is guaranteed even without correctly learning the unknown system, which is beneficial whenever the unknown system is difficult-to-learn (Tsiamis and Pappas, 2021). Finally, simulation results (see Section 4) suggest that, by ensuring persistent excitation, the proposed method is also able to learn the true parameter. This is not surprising, as the reason that the proposed method does not achieve perfect identification is that the LMS estimator only learns as long as there is a prediction error $e_{1|k}$. We expect that under persistent excitation $e_{1|k} \equiv 0$ implies the absence of parameter error, i.e., $\phi_k \equiv 0$.

4. Numerical Example

In order to demonstrate the real-world applicability of the proposed algorithm, we consider the stabilization of the planar quadrotor model used in Singh et al. (2023), which is an unstable underactuated nonlinear system.¹ The discrete-time nonlinear dynamics obtained via forward Euler discretization can be written as

1. The code is available online: <https://gitlab.ethz.ch/ics/adaptive-lqr-for-unknown-ltv-systems>, doi:10.3929/ethz-c-000798810

$$x_{k+1} = \begin{bmatrix} p_{k+1}^x \\ p_{k+1}^z \\ \psi_{k+1} \\ v_{k+1}^x \\ v_{k+1}^z \\ \omega_{k+1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & T_s & 0 & 0 \\ 0 & 1 & 0 & 0 & T_s & 0 \\ 0 & 0 & 1 & 0 & 0 & T_s \\ 0 & 0 & -gT_s & 1 & 0 & 0 \\ 0 & 0 & -[\theta_k]_1 T_s & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{A(\theta_k)} x_k + \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ T_s/m & T_s/m \\ T_s l [\theta_k]_2 & -T_s l [\theta_k]_2 \end{bmatrix}}_{B(\theta_k)} u_k + w_k^{\text{nl}}, \quad (16)$$

where the nonlinearities are captured by

$$w_k^{\text{nl}} = T_s \begin{bmatrix} v_k^x (\cos(\psi_k) - 1) - v_k^z \sin(\psi_k) \\ v_k^x \sin(\psi_k) + v_k^z (\cos(\psi_k) - 1) \\ 0 \\ v_k^z \omega_k - g \sin(\psi_k) + [\theta_k]_1 \cos(\psi_k) + g \psi_k \\ -v_k^x \omega_k - g \cos(\psi_k) - [\theta_k]_1 \sin(\psi_k) + w_k^z + [\theta_k]_1 \psi_k + [1/m \ 1/m] u_{\text{eq}} \\ w_k^\psi \end{bmatrix}. \quad (17)$$

The state $x \in \mathbb{R}^6$ consists of horizontal and vertical positions p^x, p^z , velocities in the body frame v^x, v^z , and pitch angle and angular velocity ψ, ω , and the input $u \in \mathbb{R}^2$ denotes deviation from the hovering thrust $u_{\text{eq}} = \frac{mg}{2} [1 \ 1]^\top$. Disturbances $(w^z, w^\psi) \in [-1, 1]^2$ represent inaccuracies in the actuation. The unknown parameter vector $\theta = ([\theta]_1, [\theta]_2) \in \Theta = [-10, 10] \times [50, 500]$ consists of a time-varying horizontal wind force ($[\theta]_1$), and the inverse of the moment of inertia of the out-of-plane axis ($[\theta]_2$). Note that the set Θ is of considerable size. Parameters $g = 9.81$, $m = 0.5$, $l = 0.25$ corresponding to the gravitational acceleration, mass, and length of one arm, respectively, are assumed to be known and constant, and the discretization time is $T_s = 0.1$ (all in SI units). It can be shown that $\forall \theta \in \Theta$ the pair $(A(\theta), B(\theta))$ is stabilizable, i.e., Assumption 2 is satisfied.

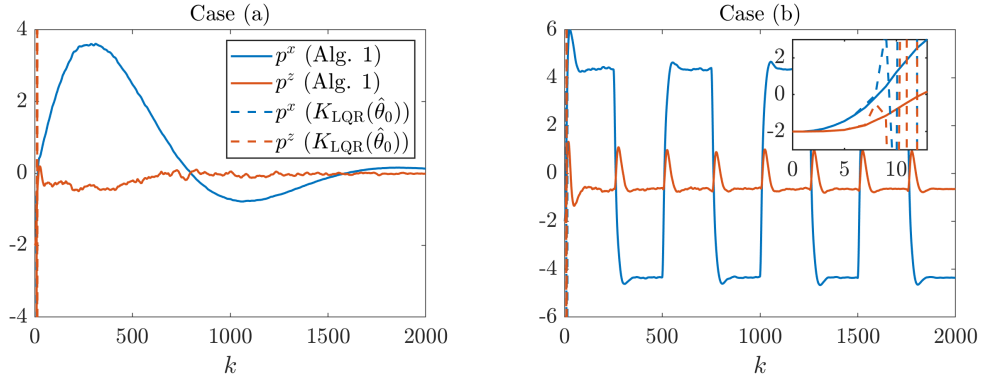


Figure 2: Closed-loop position trajectories of the proposed adaptive controller (Alg. 1) (solid) and a non-adaptive baseline (diverging dashed, see zoomed inset on the right) applied to the nonlinear planar quadrotor (16) for Cases (a) and (b).

Regarding the unknown parameters, we consider two cases. The wind $[\theta_k]_1$ follows a decaying pattern in Case (a), and a persistent square waveform with constant amplitude in Case (b) (see Figure 3 for the wind profiles), and the true value for the inverse of the inertia is $[\theta_k]_2 \equiv 250$ in both

cases. First, we apply Algorithm 1 to control the true nonlinear system (16), with $\mu = 50$, $Q = I_6$, $R = 10I_2$. The initial conditions are $p_0^x = p_0^y = -2$ and $\hat{\theta}_0 = [0 \ 100]^\top$, and disturbances (w_k^z, w_k^ψ) are sampled independently from a uniform distribution over $[-1, 1] \cdot \exp(-0.001k)$. As Figure 2 shows, the proposed method achieves asymptotic stability of the closed-loop in Case (a), and keeps the system from diverging in Case (b). Figure 2 also highlights the need for online adaptation: the initial policy $K_{\text{LQR}}(\hat{\theta}_0)$ is unable to stabilize the system.

In order to investigate the convergence of the parameter estimates, we test our algorithm on the linearized dynamics $x_{k+1} = A(\theta_k)x_k + B(\theta_k)u_k + w_k^{\text{lin}}$ with $w_k^{\text{lin}} = T_s[0 \ 0 \ 0 \ [\theta_k]_1 \ w_k^z \ w_k^\psi]^\top$. For this example, we apply the input $u_k = K_k x_k + \epsilon_k$, where the exploratory noise $\epsilon_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \text{diag}(0.5^2, 0.1^2))$ is injected to the system to ensure persistency of excitation. As illustrated by Figure 3, the proposed method is able to learn the true value of the moment of inertia and track the unknown time-varying wind parameter. This highlights a trade-off that one can make: if only stability is important, then there is no need to inject noise, but learning the parameter can also be achieved by injecting additional noise.

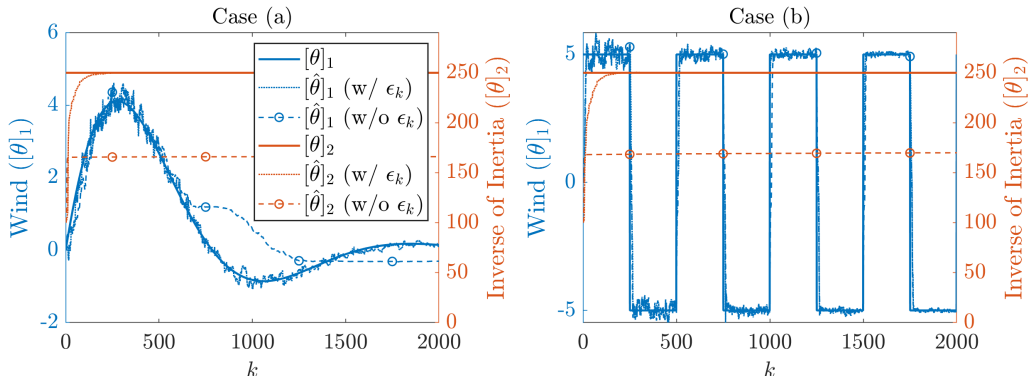


Figure 3: Comparison of the proposed adaptive controller (Alg. 1) with (dotted) and without (dashed) additional exploratory noise ϵ_k in terms of convergence of the estimates to the true values (solid) on the linearized dynamics for Cases (a) and (b).

5. Conclusion and Outlook

In this work, we proposed an adaptive control scheme for discrete-time linear systems with unknown time-varying parameters. The proposed method is a combination of two off-the-shelf building blocks from the literature: the certainty-equivalent linear quadratic regulator and the projected least mean square estimator. Despite its simplicity, ℓ^2 -stability of the closed-loop interconnection was proven, even without assuming that the data is persistently exciting. The real-world applicability of the proposed method was demonstrated on a nonlinear planar quadrotor model. Furthermore, simulation results also suggest that the proposed method is able to learn the unknown parameter if the data is persistently exciting, resulting in asymptotically optimal behavior. The formal investigation of this hypothesis is the subject of future work.

Appendix A. Technical Proofs

A.1. Notation and Preliminaries

We will often use the shorthand notations $A_k := A(\theta_k)$, $B_k := B(\theta_k)$, $\hat{A}_k := A(\hat{\theta}_k)$, $\hat{B}_k := B(\hat{\theta}_k)$, $P_k := P(\hat{\theta}_k)$, $K_k := K_{\text{LQR}}(\hat{\theta}_k)$, $\hat{A}_{\text{cl},k} := \hat{A}_k + \hat{B}_k K_k$, and $\hat{Q} := \text{diag}(Q, R)$. Since Θ is bounded and pointwise stabilizable (cf. Assumption 2), we can define the following uniform bounds:

$$\begin{aligned} \|K_{\text{LQR}}(\theta)\| &\leq \bar{K}, \quad P(\theta) \preceq \bar{p}I, \quad R \preceq rI, \quad q \preceq Q \preceq qI, \quad \forall \theta \in \Theta, \\ \|A(\theta) + B(\theta)K_{\text{LQR}}(\theta')\| &\leq \bar{a}, \quad \forall \theta, \theta' \in \Theta. \end{aligned} \quad (18)$$

Throughout the proofs we will make use of the following lemma (in addition to Lemma 3):

Lemma 6 (Young's inequality) For any $0 < M \in \mathbb{R}^{n \times n}$, $\forall \varepsilon > 0$ and $\forall x, y \in \mathbb{R}^n$, it holds that

$$\|x + y\|_M^2 \leq (1 + \varepsilon)\|x\|_M^2 + \left(1 + \frac{1}{\varepsilon}\right)\|y\|_M^2.$$

We will also use the following convergent series:

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}, \quad \sum_{k=0}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2}, \quad \sum_{k=0}^{\infty} k^2 q^{k-1} = \frac{1+q}{(1-q)^3}, \quad \forall |q| < 1. \quad (19)$$

Finally, we remark that, since $A(\theta)$ and $B(\theta)$ are affine, Lemma 3 implies that $\exists L_A > 0$ such that

$$\|\hat{A}_{\text{cl},j} - \hat{A}_{\text{cl},k}\| \leq L_A \|\hat{\theta}_j - \hat{\theta}_k\|, \quad \forall \hat{\theta}_j, \hat{\theta}_k \in \Theta. \quad (20)$$

A.2. Proof of Lemma 3

Proof As shown in East et al. (2020, Proposition 2), the Jacobians $\partial \text{vec}(P(\theta)) / \partial \text{vec}(A(\theta))$ and $\partial \text{vec}(P(\theta)) / \partial \text{vec}(B(\theta))$ exist and are given by

$$\frac{\partial \text{vec}(P(\theta))}{\partial \text{vec}(A(\theta))} = Z_1^{-1} Z_2, \quad \frac{\partial \text{vec}(P(\theta))}{\partial \text{vec}(B(\theta))} = Z_1^{-1} Z_3,$$

as long as Z_1 is invertible, where

$$\begin{aligned} Z_1 &= I_{n^2} - (A^\top \otimes A^\top)(I_{n^2} - (PBM_2B^\top \otimes I_n) - (I_n \otimes PBM_2B^\top) \\ &\quad + (PB \otimes PB)(M_2 \otimes M_2)(B^\top \otimes B^\top)), \\ Z_2 &= (V_{n,n} + I_{n^2})(I_{n^2} \otimes A^\top M_1), \\ Z_3 &= (A^\top \otimes A^\top)((PB \otimes PB)(M_2 \otimes M_2)(I_{m^2} - V_{m,m}(I_m \otimes B^\top P)) \\ &\quad - (I_{n^2} + V_{n,n}(PBM_2 \otimes P)), \\ M_1 &= P - PBM_2B^\top P, \quad M_2 = M_3^{-1}, \quad M_3 = R + B^\top PB, \end{aligned}$$

where the permutation matrix $V_{n,m} \in \mathbb{R}^{nm \times nm}$ is implicitly defined by $V_{n,m} \text{vec}(A) = \text{vec}(A^\top)$, and we omit the argument θ for ease of notation. Since

$$Z_1 = I_{n^2} - (A^\top - A^\top PBM_2B^\top) \otimes (A^\top - A^\top PBM_2B^\top) = I_{n^2} - (A + BK)^\top \otimes (A + BK)^\top,$$

and $A + BK$ is Schur stable, Z_1 is invertible over Θ , moreover, there exist $C_Z > 0$ and $0 \leq \rho_Z < 1$ such that

$$\|((A + BK)^\top \otimes (A + BK)^\top)^k\| \leq C_Z \rho_Z^k, \forall k \in \mathbb{N}, \forall \theta \in \Theta.$$

Therefore,

$$\begin{aligned} \|Z_1^{-1}\| &= \|(I_{n^2} - (A + BK)^\top \otimes (A + BK)^\top)^{-1}\| \\ &\stackrel{\text{Neumann series}}{\leq} \sum_{k=0}^{\infty} \|((A + BK)^\top \otimes (A + BK)^\top)^k\| \leq \sum_{k=0}^{\infty} C_Z \rho_Z^k = \frac{C_Z}{1 - \rho_Z}, \end{aligned}$$

which, combined with compactness of Θ , proves that the Jacobians are uniformly upper bounded over Θ .

Based on this, the Jacobians $\partial \text{vec}(K_{\text{LQR}}(\theta)) / \partial \text{vec}(A(\theta))$ and $\partial \text{vec}(K_{\text{LQR}}(\theta)) / \partial \text{vec}(B(\theta))$ can be derived using the same techniques as used in [East et al. \(2020\)](#), and the corresponding uniform upper bounds also exist. Since $A(\theta)$ and $B(\theta)$ are affine in θ , this implies that the Jacobian $\partial \text{vec}(K_{\text{LQR}}(\theta)) / \partial \theta$ is uniformly upper bounded over Θ , proving Lipschitz continuity over Θ . ■

A.3. Proof of Proposition 4

Proof In order to bound the difference $V(x_{k+1}, \hat{\theta}_{k+1}) - V(x_k, \hat{\theta}_k)$, we split it into two parts:

$$V(x_{k+1}, \hat{\theta}_{k+1}) - V(x_k, \hat{\theta}_k) = \underbrace{V(x_{k+1}, \hat{\theta}_{k+1}) - V(x_{k+1}, \hat{\theta}_k)}_{(*)} + \underbrace{V(x_{k+1}, \hat{\theta}_k) - V(x_k, \hat{\theta}_k)}_{(**)}.$$

The difference $(**)$ can be upper bounded as follows:

$$\begin{aligned} V(x_{k+1}, \hat{\theta}_k) - V(x_k, \hat{\theta}_k) &= \|x_{k+1}\|_{P_k}^2 - \|x_k\|_{P_k}^2 \\ &\leq \left(1 + \frac{1}{\varepsilon_1}\right) \|e_{1|k} + w_k\|_{P_k}^2 + (1 + \varepsilon_1) \|\hat{x}_{1|k}\|_{P_k}^2 - \|x_k\|_{P_k}^2, \end{aligned}$$

for arbitrary $\varepsilon_1 > 0$, due to $x_{k+1} = \hat{x}_{1|k} + e_{1|k} + w_k$ and Lemma 6. As P_k solves (5),

$$\|\hat{x}_{1|k}\|_{P_k}^2 = \|(\hat{A}_k + \hat{B}_k K_k)x_k\|_{P_k}^2 = \|x_k\|_{P_k}^2 - \|x_k\|_{Q + K_k^\top R K_k}^2,$$

which results in the following upper bound for $(**)$:

$$\begin{aligned} V(x_{k+1}, \hat{\theta}_k) - V(x_k, \hat{\theta}_k) &\leq \left(1 + \frac{1}{\varepsilon_1}\right) \|e_{1|k} + w_k\|_{P_k}^2 + \varepsilon_1 \|x_k\|_{P_k}^2 - (1 + \varepsilon_1) \|x_k\|_{Q + K_k^\top R K_k}^2 \\ &\stackrel{(18)}{\leq} -((1 + \varepsilon_1)\underline{q} - \varepsilon_1 \bar{p}) \|x_k\|^2 + \left(1 + \frac{1}{\varepsilon_1}\right) \bar{p} \|e_{1|k} + w_k\|^2. \end{aligned} \quad (21)$$

In order to bound $(*)$, note that $V(x, \theta)$ is the infinite-horizon cost-to-go corresponding to system $A(\theta), B(\theta)$ under controller $K_{\text{LQR}}(\theta)$:

$$V(x, \theta) = x^\top P(\theta)x = \sum_{j=0}^{\infty} \|(x_j, K_{\text{LQR}}(\theta)x_j)\|_Q^2, \quad (22)$$

where $x_{j+1} = (A(\theta) + B(\theta)K_{\text{LQR}}(\theta))x_j$, $x_0 = x$. To express $V(x_{k+1}, \hat{\theta}_{k+1})$ and $V(x_{k+1}, \hat{\theta}_k)$ via (22), define the auxiliary systems $z_{i+1} = \hat{A}_{\text{cl},k+1}z_i$ and $y_{i+1} = \hat{A}_{\text{cl},k}y_i$ initialized at $z_0 = y_0 = x_{k+1}$, and let

$$\xi_i := z_i - y_i, \quad \eta_i := K_{k+1}z_i - K_k y_i. \quad (23)$$

Then,

$$\begin{aligned} \eta_i &= K_{k+1}\xi_i + (K_{k+1} - K_k)y_i \\ \xi_{i+1} &= \hat{A}_{\text{cl},k+1}\xi_i + (\hat{A}_{\text{cl},k+1} - \hat{A}_{\text{cl},k})y_i, \end{aligned}$$

that is,

$$\begin{aligned} \xi_j &= \sum_{i=0}^{j-1} \hat{A}_{\text{cl},k+1}^i (\hat{A}_{\text{cl},k+1} - \hat{A}_{\text{cl},k}) \hat{A}_{\text{cl},k}^{j-1-i} x_{k+1}, \\ \eta_j &= (K_{k+1} - K_k) \hat{A}_{\text{cl},k}^j x_{k+1} + K_{k+1} \sum_{i=0}^{j-1} \hat{A}_{\text{cl},k+1}^i (\hat{A}_{\text{cl},k+1} - \hat{A}_{\text{cl},k}) \hat{A}_{\text{cl},k}^{j-1-i} x_{k+1}, \end{aligned}$$

implying

$$\|\xi_j\| \leq \|\hat{A}_{\text{cl},k+1} - \hat{A}_{\text{cl},k}\| \sum_{i=0}^{j-1} \|\hat{A}_{\text{cl},k+1}^i\| \|\hat{A}_{\text{cl},k}^{j-1-i}\| \|x_{k+1}\|, \quad (24)$$

$$\begin{aligned} \|\eta_j\| &\leq \|K_{k+1} - K_k\| \|\hat{A}_{\text{cl},k}^j\| \|x_{k+1}\| \\ &\quad + \|K_{k+1}\| \|\hat{A}_{\text{cl},k+1} - \hat{A}_{\text{cl},k}\| \sum_{i=0}^{j-1} \|\hat{A}_{\text{cl},k+1}^i\| \|\hat{A}_{\text{cl},k}^{j-1-i}\| \|x_{k+1}\|, \end{aligned} \quad (25)$$

Since $\hat{A}_{\text{cl},k}$ and $\hat{A}_{\text{cl},k+1}$ are Schur stable, there exist $C > 0$ and $0 \leq \rho < 1$ such that

$$\|\hat{A}_{\text{cl},k}^i\| \leq C\rho^i, \quad \|\hat{A}_{\text{cl},k+1}^i\| \leq C\rho^i, \quad \forall i \geq 0, \quad \forall \hat{\theta}_k \in \Theta,$$

resulting in

$$\sum_{i=0}^{j-1} \|\hat{A}_{\text{cl},k+1}^i\| \|\hat{A}_{\text{cl},k}^{j-1-i}\| \leq \sum_{i=0}^{j-1} C^2 \rho^{j-1} = C^2 j \rho^{j-1}.$$

Consequently,

$$\|\xi_j\| \stackrel{(24)}{\leq} C^2 \|\hat{A}_{\text{cl},k+1} - \hat{A}_{\text{cl},k}\| \|x_{k+1}\| j \rho^{j-1}, \quad (26)$$

$$\|\eta_j\| \stackrel{(25)}{\leq} C^2 \|K_{k+1} - K_k\| \|x_{k+1}\| \rho^j + C^2 \|K_{k+1}\| \|\hat{A}_{\text{cl},k+1} - \hat{A}_{\text{cl},k}\| \|x_{k+1}\| j \rho^{j-1}. \quad (27)$$

Note that

$$\|x_{k+1}\|^2 = \|(A_k + B_k K_k)x_k + w_k\|^2 \stackrel{\text{Lemma 6}}{\leq} 2\|(A_k + B_k K_k)x_k\|^2 + 2\|w_k\|^2 \leq 2\bar{a}^2 X^2 + 2\bar{W}^2, \quad (28)$$

where the last inequality is due to (18) and the assumption that $\|x_k\| \leq X$. Taking the squares of (26)-(27), and using $\rho^j \leq \rho^{j-1}$, Lemma 3, and (18), (20), and (28) results in

$$\begin{aligned}\|\xi_j\|^2 &\leq 2C^2(\bar{a}^2X^2 + \bar{W}^2)L_A^2j^2(\rho^2)^{j-1}\|\hat{\theta}_{k+1} - \hat{\theta}_k\|^2, \\ \|\eta_j\|^2 &\leq 2C^2(\bar{a}^2X^2 + \bar{W}^2)(L_K^2(\rho^2)^j + 2\bar{K}L_KL_AL_Aj(\rho^2)^{j-1} + \bar{K}^2L_A^2j^2(\rho^2)^{j-1})\|\hat{\theta}_{k+1} - \hat{\theta}_k\|^2.\end{aligned}\quad (29)$$

From (22), it can be seen that $\forall \varepsilon_2 > 0$

$$\begin{aligned}V(x_{k+1}, \hat{\theta}_{k+1}) - V(x_{k+1}, \hat{\theta}_k) &= \sum_{j=0}^{\infty} \left(\|(z_j, K_{k+1}z_j)\|_{\bar{Q}}^2 - \|(y_j, K_k y_j)\|_{\bar{Q}}^2 \right) \\ &\stackrel{(23)}{=} \sum_{j=0}^{\infty} \left(\|(y_j + \xi_j, K_k y_j + \eta_j)\|_{\bar{Q}}^2 - \|(y_j, K_k y_j)\|_{\bar{Q}}^2 \right) \\ &\stackrel{\text{Lemma 6}}{\leq} \sum_{j=0}^{\infty} \left(\varepsilon_2 \|(y_j, K_k y_j)\|_{\bar{Q}}^2 + \left(1 + \frac{1}{\varepsilon_2}\right) \|(\xi_j, \eta_j)\|_{\bar{Q}}^2 \right) \\ &= \varepsilon_2 \|x_{k+1}\|_{P_k}^2 + \left(1 + \frac{1}{\varepsilon_2}\right) \sum_{j=0}^{\infty} (\|\xi_j\|_{\bar{Q}}^2 + \|\eta_j\|_{\bar{R}}^2) \\ &\stackrel{(18)}{\leq} \varepsilon_2 \bar{p} \|x_{k+1}\|^2 + \left(1 + \frac{1}{\varepsilon_2}\right) \sum_{j=0}^{\infty} (q\|\xi_j\|^2 + r\|\eta_j\|^2).\end{aligned}\quad (30)$$

Combining (29) with the infinite sums (19) yields:

$$\begin{aligned}\sum_{j=0}^{\infty} \|\xi_j\|^2 &\leq 2C^2(\bar{a}^2X^2 + \bar{W}^2)\|\hat{\theta}_{k+1} - \hat{\theta}_k\|^2 L_A^2 \frac{1 + \rho^2}{(1 - \rho^2)^3}, \\ \sum_{j=0}^{\infty} \|\eta_j\|^2 &\leq 2C^2(\bar{a}^2X^2 + \bar{W}^2)\|\hat{\theta}_{k+1} - \hat{\theta}_k\|^2 \left(\frac{L_K^2}{1 - \rho^2} + \frac{2\bar{K}L_KL_AL_A}{(1 - \rho^2)^2} + \bar{K}^2L_A^2 \frac{1 + \rho^2}{(1 - \rho^2)^3} \right).\end{aligned}\quad (31)$$

Applying Lemma 6 with $\varepsilon_3 > 0$ on $\|x_{k+1}\|^2 = \|\hat{x}_{1|k} + e_{1|k} + w_k\|^2$ gives

$$\|x_{k+1}\|^2 \leq \left(1 + \frac{1}{\varepsilon_3}\right) \|e_{1|k} + w_k\|^2 + (1 + \varepsilon_3)\bar{a}^2\|x_k\|^2.\quad (32)$$

Finally, combining (30), (31), (32), and (9) results in

$$\begin{aligned}V(x_{k+1}, \hat{\theta}_{k+1}) - V(x_{k+1}, \hat{\theta}_k) & \\ &\leq \left(\left(1 + \frac{1}{\varepsilon_2}\right) \mu(C_1X^2 + C_2\bar{W}^2) + \varepsilon_2 \left(1 + \frac{1}{\varepsilon_3}\right) \bar{p} \right) \|e_{1|k} + w_k\|^2 + \varepsilon_2(1 + \varepsilon_3)\bar{a}^2\bar{p}\|x_k\|^2,\end{aligned}\quad (33)$$

for some $C_1, C_2 > 0$.

Summing (33) and (21) leads to

$$\begin{aligned}V(x_{k+1}, \hat{\theta}_{k+1}) - V(x_k, \hat{\theta}_k) &\leq -((1 + \varepsilon_1)q - \varepsilon_2(1 + \varepsilon_3)\bar{a}^2\bar{p} - \varepsilon_1\bar{p})\|x_k\|^2 \\ &\quad + \left(\left(1 + \frac{1}{\varepsilon_2}\right) \mu(C_1X^2 + C_2\bar{W}^2) + \left(\varepsilon_2 \left(1 + \frac{1}{\varepsilon_3}\right) + 1 + \frac{1}{\varepsilon_1} \right) \bar{p} \right) \|e_{1|k} + w_k\|^2,\end{aligned}\quad (34)$$

$\forall \varepsilon_{1,2,3} > 0$. By choosing $\varepsilon_{1,2,3} > 0$ small enough and once again using Lemma 6 on $\|e_{1|k} + w_k\|^2$, we obtain $\alpha, \beta, \gamma > 0$ satisfying the Lyapunov decrease condition (11), where β is of the form

$$\beta = \nu_2 \mu X^2 + \nu_3 \mu + \nu_4 \quad (35)$$

with $\nu_{2,3,4} > 0$ constants independent of X and μ . ■

A.4. Proof of Theorem 5

Proof First, we address item b) of Theorem 5 by showing that the candidate Lyapunov function (12) satisfies (13)-(14) for a given k , assuming $\|x_k\| \leq X$ for some $X \geq 1$. Then, we show that sublevel sets of the Lyapunov function are invariant for the joint state (x_k, φ_k) . Lastly, we prove that if $d > 0$ (the diameter of Θ) is small enough, there is a sufficiently large $X > 0$ such that containment in the invariant set implies $\|x_{k+1}\| \leq X$, proving items a)-b) of Theorem 5. Finally, we prove item c).

If $\|x_k\| \leq X$, then $\|u_k\| = \|K(\hat{\theta}_k)x_k\|$ is also bounded, since $\hat{\theta}_k \in \Theta$. Thus, by selecting $\mu = \nu_1/X^2$, where $\nu_1 > 0$ is some constant depending on the known functions $A(\theta)$ and $B(\theta)$, the step size $\mu > 0$ satisfies (7), ensuring that the conditions in Propositions 2 and 4 hold for k , resulting in inequalities (8) and (11) for k . Multiplying (8) by $\beta/\mu > 0$ and adding it to (11) yields

$$\tilde{V}(x_{k+1}, \varphi_{k+1}, k+1) - \tilde{V}(x_k, \varphi_k, k) \leq -\alpha \|x_k\|^2 + (\beta + \gamma) \|w_k\|^2 + \frac{2d\beta}{\mu} \|\Delta\theta_k\|, \quad (36)$$

proving (14) with $\tilde{\alpha} = \alpha$, $\tilde{\gamma} = \gamma + \beta$, $\tilde{\delta} = 2d\beta/\mu$. It can also be seen that $\tilde{V}(x_k, \varphi_k, k)$ satisfies the uniform lower and upper bounds (13):

$$\underline{q} \|x_k\|^2 + \frac{\beta}{\mu} \|\varphi_k\|^2 \leq \tilde{V}(x_k, \varphi_k, k) \leq \bar{p} \|x_k\|^2 + \frac{\beta}{\mu} \|\varphi_k\|^2. \quad (37)$$

Define the time-varying sublevel set of $\tilde{V}(x_k, \varphi_k, k)$ as

$$\mathcal{S}_k := \{ (x_k, \varphi_k) \mid \tilde{V}(x_k, \varphi_k, k) \leq b \},$$

where

$$b := \max \left\{ \bar{p} \|x_0\|^2; \left(\frac{\bar{p}}{\alpha} + 1 \right) \left((\gamma + \beta) \bar{W}^2 + 2 \frac{\beta}{\mu} d^2 \right) \right\} + \frac{\beta}{\mu} d^2,$$

ensuring that $(x_0, \varphi_0) \in \mathcal{S}_0$. We use case distinction to show that $(x_k, \varphi_k) \in \mathcal{S}_k \implies (x_{k+1}, \varphi_{k+1}) \in \mathcal{S}_{k+1}$:

Case 1: if $\|x_k\|^2 \geq (\gamma + \beta) \frac{\bar{W}^2}{\alpha} + \frac{2\beta d^2}{\mu\alpha}$, then, noting that $\|\Delta\theta_k\| = \|\theta_{k+1} - \theta_k\| \leq d$, it holds that

$$\tilde{V}(x_{k+1}, \varphi_{k+1}, k+1) - \tilde{V}(x_k, \varphi_k, k) \stackrel{(36)}{\leq} -\alpha \|x_k\|^2 + (\beta + \gamma) \|w_k\|^2 + \frac{2d\beta}{\mu} \|\Delta\theta_k\| \leq 0,$$

implying $\tilde{V}(x_{k+1}, \varphi_{k+1}, k+1) \leq \tilde{V}(x_k, \varphi_k, k) \leq b$, i.e., $(x_{k+1}, \varphi_{k+1}) \in \mathcal{S}_{k+1}$.

Case 2: if $\|x_k\|^2 < (\gamma + \beta)\frac{\bar{W}^2}{\alpha} + \frac{2\beta d^2}{\mu\alpha}$, then

$$\begin{aligned}
 \tilde{V}(x_{k+1}, \varphi_{k+1}, k+1) &\stackrel{(36)}{\leq} \tilde{V}(x_k, \varphi_k, k) - \alpha\|x_k\|^2 + (\beta + \gamma)\|w_k\|^2 + \frac{2d\beta}{\mu}\|\Delta\theta_k\| \\
 &\stackrel{(37)}{\leq} \bar{p}\|x_k\|^2 + \frac{\beta}{\mu}d^2 + (\beta + \gamma)\bar{W}^2 + \frac{2\beta}{\mu}d^2 \\
 &< \bar{p}\left((\gamma + \beta)\frac{\bar{W}^2}{\alpha} + \frac{2\beta}{\mu\alpha}d^2\right) + \frac{\beta}{\mu}d^2 + (\beta + \gamma)\bar{W}^2 + \frac{2\beta}{\mu}d^2 \\
 &= \left(\frac{\bar{p}}{\alpha} + 1\right)\left((\gamma + \beta)\bar{W}^2 + 2\frac{\beta}{\mu}d^2\right) + \frac{\beta}{\mu}d^2 \leq b,
 \end{aligned}$$

i.e., $(x_{k+1}, \varphi_{k+1}) \in \mathcal{S}_{k+1}$.

That is $(x_k, \varphi_k) \in \mathcal{S}_k \implies (x_{k+1}, \varphi_{k+1}) \in \mathcal{S}_{k+1}$, further implying $\|x_{k+1}\|^2 \leq b/q$ due to (37). Note that in order for Proposition 4 to hold, β needs to satisfy $\beta \geq \nu_2\mu X^2 + \nu_3\mu + \nu_4$ with $\nu_{2,3,4} > 0$ (see (35)). Since $\mu = \nu_1/X^2$ and $X \geq 1$, this means that $\beta = \nu_1(\nu_2 + \nu_3) + \nu_4 =: \nu_5$ is a valid choice, in which case $\beta/\mu = \nu_5 X^2/\nu_1$. As α is independent of X , μ , and d , and we are free to select $\gamma = \beta$, b is of the form

$$b = \max\{\bar{p}\|x_0\|^2; \nu_6 + \nu_7 d^2 X^2\} + \nu_8 d^2 X^2,$$

with constants $\nu_{6,7,8} > 0$ independent of d and X . Consequently, to ensure that $(x_k, \varphi_k) \in \mathcal{S}_k$ implies $\|x_k\| \leq X$, it must hold that

$$qX^2 \geq b = \max\{\bar{p}\|x_0\|^2; \nu_6 + \nu_7 d^2 X^2\} + \nu_8 d^2 X^2,$$

which has a solution $X \geq 1$ as long as d is small enough.

In summary, for sufficiently small d , there exists a large enough choice for X such that $(x_k, \varphi_k) \in \mathcal{S}_k$ implies both $\|x_k\| \leq X$ and $(x_{k+1}, \varphi_{k+1}) \in \mathcal{S}_{k+1}$. As $(x_0, \varphi_0) \in \mathcal{S}_0$ due to (37), we have proven item **b)** of Theorem 5 by induction. Additionally, $\|x_k\| \leq X$, $\forall k \in \mathbb{N}$ implies $\|u_k\| \leq U$, $\forall k \in \mathbb{N}$ for some $U > 0$, proving item **a)**. Finally, regarding item **c)**, bound (15) can be proven by summing up (14) from $k = 0$ to $k = T - 1$, simplifying the resulting telescopic sum, and using bounds (13) on $\tilde{V}(x_T, \varphi_T, T)$ and $\tilde{V}(x_0, \varphi_0, 0)$. ■

Acknowledgments

This work was supported as a part of NCCR Automation, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant number 51NF40_225155).

References

- Yasin Abbasi-Yadkori and Csaba Szepesvári. Regret bounds for the adaptive control of linear quadratic systems. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 1–26. JMLR Workshop and Conference Proceedings, 2011.
- Brian D.O. Anderson and John B. Moore. *Optimal control: linear quadratic methods*. Courier Corporation, 2007.

- Karl J. Aström and Björn Wittenmark. *Adaptive control*. Addison-Wesley Longman Publishing Co., Inc., 1994.
- Marcell Bartos, Bruce D. Lee, Lenart Treven, Andreas Krause, Florian Dörfler, and Melanie N. Zeilinger. Optimistic online LQR via intrinsic rewards. *arXiv preprint arXiv:2603.28938*, 2026.
- Julian Berberich, Johannes Köhler, Matthias A. Müller, and Frank Allgöwer. Data-driven model predictive control with stability and robustness guarantees. *IEEE transactions on automatic control*, 66(4):1702–1717, 2020.
- Alon Cohen, Tomer Koren, and Yishay Mansour. Learning linear-quadratic regulators efficiently with only \sqrt{T} regret. In *International Conference on Machine Learning*, pages 1300–1309. PMLR, 2019.
- Claudio De Persis and Pietro Tesi. Formulas for data-driven control: Stabilization, optimality, and robustness. *IEEE Transactions on Automatic Control*, 65(3):909–924, 2019.
- Maximilian Degner, Raffaele Soloperto, Melanie N. Zeilinger, John Lygeros, and Johannes Köhler. Adaptive economic model predictive control: Performance guarantees for nonlinear systems. *IEEE Transactions on Automatic Control*, 2026.
- Sebastian East, Marco Gallieri, Jonathan Masci, Jan Koutnik, and Mark Cannon. Infinite-horizon differentiable model predictive control. *Proceedings of ICLR 2020*, 2020.
- Maryam Fazel, Rong Ge, Sham Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *International conference on machine learning*, pages 1467–1476. PMLR, 2018.
- Paula Gradu, Elad Hazan, and Edgar Minasyan. Adaptive regret for control of time-varying dynamics. In *Learning for Dynamics and Control Conference*, pages 560–572. PMLR, 2023.
- Zhong-Sheng Hou and Zhuo Wang. From model-based control to data-driven control: Survey, classification and perspective. *Information Sciences*, 235:3–35, 2013.
- Hassan K. Khalil. *Nonlinear systems*, volume 3. Upper Saddle River, NJ: Prentice hall, 2002.
- Miroslav Krstic, Petar V. Kokotovic, and Ioannis Kanellakopoulos. *Nonlinear and adaptive control design*. John Wiley & Sons, Inc., 1995.
- Sergey Levine, Aviral Kumar, George Tucker, and Justin Fu. Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*, 2020.
- Yiheng Lin, Yang Hu, Guanya Shi, Haoyuan Sun, Guannan Qu, and Adam Wierman. Perturbation-based regret analysis of predictive control in linear time varying systems. *Advances in Neural Information Processing Systems*, 34:5174–5185, 2021.
- Lennart Ljung. System identification. In *Signal analysis and prediction*, pages 163–173. Springer, 1998.
- Matthias Lorenzen, Mark Cannon, and Frank Allgöwer. Robust MPC with recursive model update. *Automatica*, 103:461–471, 2019.

- Edgar Minasyan, Paula Gradu, Max Simchowitz, and Elad Hazan. Online control of unknown time-varying dynamical systems. *Advances in Neural Information Processing Systems*, 34:15934–15945, 2021.
- Kumpati S. Narendra and Anuradha M. Annaswamy. *Stable adaptive systems*. Courier Corporation, 2012.
- Sahel Vahedi Noori and Maryam Babazadeh. A data-ensemble-based approach for sample-efficient LQ control of linear time-varying systems. *Journal of the Franklin Institute*, page 108118, 2025.
- Guannan Qu, Yuanyuan Shi, Sahin Lale, Anima Anandkumar, and Adam Wierman. Stable online control of linear time-varying systems. In *Learning for Dynamics and Control*, pages 742–753. PMLR, 2021.
- James Blake Rawlings, David Q. Mayne, Moritz Diehl, et al. *Model predictive control: theory, computation, and design*, volume 2. Nob Hill Publishing Madison, WI, 2020.
- Shai Shalev-Shwartz et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.
- Sumeet Singh, Benoit Landry, Anirudha Majumdar, Jean-Jacques Slotine, and Marco Pavone. Robust feedback motion planning via contraction theory. *The International Journal of Robotics Research*, 42(9):655–688, 2023.
- Bowen Song and Andrea Iannelli. The role of identification in data-driven policy iteration: A system theoretic study. *International Journal of Robust and Nonlinear Control*, 2024.
- Richard S. Sutton and Andrew G. Barto. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.
- Anastasios Tsiamis and George J. Pappas. Linear systems can be hard to learn. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 2903–2910. IEEE, 2021.
- Peter van Overschee and Bart L. De Moor. *Subspace identification for linear systems: Theory—Implementation—Applications*. Springer Science & Business Media, 2012.
- Henk J. van Waarde, M. Kanat Camlibel, and Mehran Mesbahi. From noisy data to feedback controllers: Nonconservative design via a matrix s-lemma. *IEEE Transactions on Automatic Control*, 67(1):162–175, 2020.
- Henk J. van Waarde, M. Kanat Camlibel, and Harry L. Trentelman. *Data-Based Linear Systems and Control Theory*. Kindle Direct Publishing, 2025.
- Feicheng Wang and Lucas Janson. Exact asymptotics for linear quadratic adaptive control. *Journal of Machine Learning Research*, 22(265):1–112, 2021.
- Jan C. Willems, Paolo Rapisarda, Ivan Markovskiy, and Bart L. De Moor. A note on persistency of excitation. *Systems & Control Letters*, 54(4):325–329, 2005.

Jing Yu, Varun Gupta, and Adam Wierman. Online adversarial stabilization of unknown linear time-varying systems. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 8320–8327. IEEE, 2023.

Feiran Zhao, Alessandro Chiuso, and Florian Dörfler. Policy gradient adaptive control for the LQR: Indirect and direct approaches. *arXiv preprint arXiv:2505.03706*, 2025.