Online Learning of the Kalman Filter With Logarithmic Regret

Anastasios Tsiamis, *Student Member, IEEE*, and George J. Pappas, *Fellow, IEEE*

***Abstract*—In this article, we consider the problem of predicting observations generated online by an unknown, partially observable linear system, which is driven by Gaussian noise. In the linear Gaussian setting, the optimal predictor in the mean square error sense is the celebrated Kalman filter, which can be explicitly computed when the system model is known. When the system model is unknown, we have to learn how to predict observations online based on finite data, suffering possibly a nonzero regret with respect to the Kalman filter's prediction. We show that it is possible to achieve a regret of the order of poly $\log(N)$ with high probability, where N is the number of observations collected. This is achieved using an online least-squares algorithm, which exploits the approximately linear relation between future observations and past observations. The regret analysis is based on the stability properties of the Kalman filter, recent statistical tools for finite sample analysis of system identification, and classical results for the analysis of least-squares algorithms for time series. Our regret analysis can also be applied to other predictors, e.g., multiple step-ahead prediction, or prediction under exogenous inputs including closed-loop prediction. A fundamental technical contribution is that our bounds hold even for the class of nonexplosive systems (including marginally stable systems), which was not addressed before in the case of online prediction.**

***Index Terms*—Adaptive estimation, Kalman filter, online learning, statistical learning.**

I. Introduction

THE celebrated Kalman filter has been a fundamental approach for estimation and prediction of time-series data, with diverse applications ranging from control systems and robotics [1], [2] to computer vision [3] and economics [4]. Given a known system model with known noise statistics, the Kalman filter predicts future observations of a *partially observable* dynamical process by filtering past observations. When the underlying process is linear and the noise is Gaussian, the Kalman filter is optimal in the sense that it minimizes the mean square prediction error. Since Kalman's seminal paper [5], the stability and statistical properties of the Kalman filter have been well studied when the system model is *known*.

Learning to predict *unknown* partially observable systems is a significantly more challenging problem. Even in the case of linear systems, learning directly the model parameters of the system results in nonlinear, nonconvex problems [6]. Adaptive filtering algorithms address the problem of making observation predictions when the system model or the noise statistics are unknown or changing [7], [8], [9], [10]. These adaptive filtering approaches are usually based on variations of extended least squares. The statistical analysis of adaptive algorithms has relied on asymptotic tools, which assume that the number of collected data N is infinite. However, our asymptotic tools, e.g., the central limit theorem or law of large numbers, do not always capture all aspects of finite sample performance [11, Ch. 2]. Moreover, the dependence of prediction performance on various system-theoretic parameters has been hidden under the big-O notation.

In this article, we consider the problem of predicting observations generated by an *unknown*, partially observable linear dynamical system under *finite samples*. The system dynamics and observation map are corrupted by Gaussian noise. For the theoretical analysis, we adopt the notion of regret [12], which captures the finite sample performance of online prediction. It measures how far our online predictions are from the optimal Kalman Filter predictions that has access to the full system model. Our goal is to find an online prediction algorithm that has provably small regret. Our technical contributions are as follows.

1) *System-theoretic regret:* We define a notion of regret that has a natural, system-theoretic interpretation. The prediction error of an online prediction algorithm is compared against the prediction error of the Kalman filter that has access to the exact model, which is allowed to be arbitrary. Previous regret definitions [13] required the model to lie in a finite set.
2) *Logarithmic regret for the Kalman filter:* We present the first online prediction algorithm with provable logarithmic regret upper bounds for the classical Kalman filter. We prove that with high probability, the regret of our algorithm is of the order of $\tilde{O}(1)$, where $\tilde{O}$ hides poly $\log N$ terms and N is the number of observations collected. Our algorithm is inspired by subspace identification system parameterizations [14]. Instead of optimizing over the state-space parameters, which is a nonconvex problem, we convexify the problem by establishing an approximate regression between the next observation and past observations. Our analysis is based on the stability properties

Manuscript received 11 May 2022; accepted 1 September 2022. Date of publication 19 September 2022; date of current version 26 April 2023. This work was supported in part by NSF CPS grant: Robust Learning for Perception-Based Autonomous Systems. Recommended by Senior Editor Tetsuya Iwasaki and Guest Editors Anuradha M. Annaswamy, Manfred Morari, Claire J. Tomlin, Rene Vidal, and Melanie N. Zeilinger. *(Corresponding author: Anastasios Tsiamis.)*

Anastasios Tsiamis is with the Automatic Control Laboratory, ETH Zurich, Switzerland (e-mail: atsiamis@control.ee.ethz.ch).

George J. Pappas is with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: pappasg@seas.upenn.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2022.3207670.

Digital Object Identifier 10.1109/TAC.2022.3207670

0018-9286 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

of the Kalman filter, tools for self-normalized martingales and matrices, and high-dimensional statistics [11].

3) *Logarithmic regret for nonexplosive systems:* Our regret guarantees do not require persistency of excitation and hold for the class of nonexplosive systems, which includes marginally stable linear systems. This proves that online prediction performance does not depend on the system stability gap $(1/(1-\rho)$, where ρ is the spectral radius of the system). Recently, it has been shown that the stability gap does not affect system identification [15]. However, the question, whether the stability gap affects online prediction under stochastic noise, had not been addressed before.
4) *Refined regret bounds for stable systems:* When the system is stable, we can provide tighter logarithmic regret upper bounds by exploiting stationarity and new results for persistency of excitation developed in this article.
5) *Regret analysis for other predictors:* Our approach directly carries over to various interesting online predictors. For example, our analysis can be directly extended to the case of f-step ahead prediction of observations, or to the case when we have exogenous inputs. Both predictors enjoy similar logarithmic regret bounds.
6) *Learning gap between LQR and Kalman filter:* One of the implications of our bounds is that learning to predict observations like the Kalman filter is provably easier than solving the online linear quadratic regulator (LQR) problem, which in general requires $\Omega(\sqrt{N})$ regret [16], [17]. This might not be surprising due to the fact that, in the case of exogenous inputs, we need to inject exploratory signals into the system.

A. Related Work

Recently, there have been important results addressing the regret of the adaptive LQR problem [18], [19], [20], [21], [22], [23], [24]. The best regret for LQR is sublinear and of the order of $\tilde{O}(\sqrt{N})$, where N is the numbers of state samples collected; an in-depth survey can be found in [25]. When the system model is *known*, then the Kalman filter is the dual of the LQR, suggesting that this duality can be exploited in deriving the regret of the Kalman filter. However, when the system model is *unknown*, the LQR and the Kalman filter are not dual problems [26]. As the state is fully observed in LQR, the system identification in adaptive LQR reduces to a simple least-squares problem. In the adaptive Kalman filter, the state is *partially* observed resulting in nonconvex system identification problems requiring us to consider a different approach.

A related but different problem focuses on online prediction algorithms for systems without internal states (such as autoregressive moving average) [27]. Although such results suggest that logarithmic regret is possible, they do not directly address state-space systems and require the observations to be bounded. Prediction of observations generated by state-space models in the case of exogenous inputs and adversarial noise but with a bounded budget was studied in [28]. Recently, Kozdoba et al. [13] introduced regret bounds for the Kalman Filter in the restricted context of scalar and bounded observations. The regret is shown to be of the order of $\sqrt{N}$ along with a small linear term. Here, we improve the state of the art to logarithmic bounds for general observations. Concurrently and independently [29] also proved logarithmic regret bounds for the Kalman filter, under an algorithm similar to ours. Contrary to [29], our algorithm progresses in epochs, and is agnostic to the total number N of samples collected. The order of the logarithm in our regret upper bound is independent of the size of the system. Our analysis also establishes persistency of excitation guarantees, providing refined regret bounds in the case of stable systems. After this article was written, regret bounds were extended to the case where the Kalman filter closed-loop matrix is close to instability [30]. Finally, Lale et al. [31] proved logarithmic regret bounds for the linear quadratic Gaussian (LQG) control problem, under the assumption that the optimal policy is persistently exciting. The LQG control problem is different from the prediction problem studied here.

Our online algorithm is inspired by subspace identification techniques [32]. The technical approach is based on classical results for the analysis of the least-squares estimator for time series [33], as well as modern results for finite sample analysis of system identification in both the fully observed [34], [35], [36] and the partially observed cases [15], [31], [37], [38], [39], [40].

Article organization: In Section II, we provide some background on the classical Kalman filter and formulate the regret problem considered in this article. In Section III, we introduce the online learning algorithm while the regret guarantees are presented in Section IV. In Section V, we illustrate the performance of our algorithm via numerical simulations. In Section VI, we present discussion and extensions. Finally, Sections VII concludes this article. The proofs can be found in the Appendix.

Notation: The term universal constant is used for numbers that are independent of the system's (algorithm's) parameters. The operator $\succeq$ denotes comparison in the positive semidefinite cone. With $\|\|_2$, we denote the Euclidean norm for vectors and the spectral norm for matrices. The spectral radius of a matrix A is denoted by $\rho(A)$. The smallest singular value of a matrix A is denoted by $\sigma_{\min}(A)$. By A^*, we denote the transpose of A. The big-O notation $O(f(N))$ represents a function that grows at most as fast as $f(N)$. The $\tilde{O}(\cdot)$ notation hides polylogarithmic terms. The poly(x) notation means a polynomial function of x.

II. Problem Formulation

The Kalman filter considers the problem of predicting observations generated by the following *state-space system*:

$$\begin{aligned} x_{k+1} &= Ax_k + w_k, \quad w_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, Q) \\ y_k &= Cx_k + v_k, \quad v_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, R) \end{aligned} \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the state, $y_k \in \mathbb{R}^m$ and $m \leq n$ are the observations (outputs), $A \in \mathbb{R}^{n\times n}$ is the system matrix, and $C \in \mathbb{R}^{m\times n}$ is the observation matrix. The time series w_k and v_k represent the process and measurement noise, respectively, and are modeled as zero-mean independent identically distributed (i.i.d.) Gaussian variables, independent of each other, with covariances Q and R, respectively. The initial state is zero-mean Gaussian with covariance Σ_0 and independent of the noises. The following assumption holds throughout this article.

Assumption 1: System (1) is nonexplosive, namely the spectral radius is $\rho(A) \leq 1$. The state matrices are bounded $\|A\|_2, \|C\|_2, \|Q\|_2, \|R\|_2 \leq M$, for some $M \geq 0$. Letting $A = SJS^{-1}$ be the Jordan form decomposition of A, then the similarity transformation is bounded $\|S\|_2, \|S^{-1}\|_2 \leq M$.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

Let $\mathcal{F}_k \triangleq \sigma(y_0, \dots, y_k)$ be the filtration generated by the observations $y_0, \dots, y_k$. Given the observations up to time k, the optimal prediction $\hat{y}_{k+1}$ at time $k+1$ in the minimum mean square error sense is defined as

$$\hat{y}_{k+1} \triangleq \arg\min_{z \in \mathcal{F}_k} \mathbb{E}\left[\|y_{k+1} - z\|_2^2 \,|\mathcal{F}_k\right]. \tag{2}$$

In the case of system (1), the optimal predictor admits a recursive expression, known as the *Kalman filter*:

$$\begin{aligned} \hat{x}_{k+1} &= A\hat{x}_k + Ke_k,\ \hat{x}_0 = 0 \\ \hat{y}_k &= C\hat{x}_k,\ y_k = \hat{y}_k + e_k \end{aligned} \tag{3}$$

where $e_k \triangleq y_k - C\hat{x}_k$ is the innovation noise process. Matrix $K \in \mathbb{R}^{n\times m}$ is called the Kalman filter gain, and can be computed based on A, C, Q and R [see (6)].

The Kalman filter requires the system matrices A and C and noise covariances Q and R to be known. In this article, we seek online learning algorithms that can predict observations based only on past observation data, without any knowledge of system matrices or noise covariances. To quantify the online prediction performance, we define the regret of our online learning algorithm with respect to the Kalman filter (3) that has full knowledge of system model (1). Our goal is to achieve sublinear regret, as defined in the following problem statement.

Problem 1: Assume that A, C, Q, R, Σ_0 and n in system model (1) are unknown. Consider a sequence $y_0, y_1, \dots,$ of observations generated by system (1). Let $\tilde{y}_k \in \mathcal{F}_{k-1}$ be the prediction of an online learning algorithm based on the history $y_{k-1}, \dots, y_0$ and $\hat{y}_k$ be the Kalman filter prediction (3) that has full knowledge of model (1). Define the regret

$$\mathcal{R}_N \triangleq \sum_{k=1}^{N} \|y_k - \tilde{y}_k\|^2 - \sum_{k=1}^{N} \|y_k - \hat{y}_k\|^2. \tag{4}$$

Fix a failure probability $\delta > 0$. Our goal is to find a learning algorithm such that with probability at least $1-\delta$:

$$\mathcal{R}_N \le \text{poly}(\log 1/\delta)o(N).$$

The regret captures the average suboptimality of the online predictor. If the regret is sublinear, this implies average convergence since $\mathcal{R}_N/N = o(1)$.

In the following section, we provide some background on the Kalman filter and specify some standard assumptions, which guarantee that the Kalman filter is well defined.

A. Kalman Filter Background

The Kalman filter enjoys two critical properties, namely closed-loop stability and innovation orthogonality, that are now reviewed. The following standard assumption holds throughout this article and guarantees that the Kalman filter is stable.

Assumption 2: R is strictly positive definite. The system matrix pair (A, C) is observable, i.e., the observability matrix

$$\mathcal{O}_k \triangleq \begin{bmatrix} C^* & A^*C^* & \dots & (A^*)^{k-1}C^* \end{bmatrix}^* \tag{5}$$

has rank n for all $k \ge n$. The pair $(A, Q^{1/2})$ is controllable, i.e., the pair $(A^*, Q^{1/2})$ is observable.

The following result shows that under Assumption 2, the closed loop matrix $A - KC$ of the Kalman filter is stable.

Proposition 1 ([41]): Consider system (1) under Assumption 2. The Kalman filter gain in (3) is computed by

$$K = APC^* \left(CPC^* + R\right)^{-1}$$

where P is the positive definite solution to

$$P = (A - KC)P(A - KC)^* + Q + KRK^*. \tag{6}$$

Moreover, the closed-loop matrix $A - KC$ is stable, i.e., it has spectral radius $\rho(A - KC) < 1$.

Proposition 1 implies that the Kalman filter reaches the steady state exponentially fast. This allows us to make the following standard assumption, which guarantees that the Kalman filter has stabilized to its steady state and (3) is well defined.

Assumption 3: We assume that the initial state covariance is $\Sigma_0 = P$, where P is defined in (6).

The following assumption is for notational simplicity. It assumes that the largest eigenvalue of $A - KC$ is simple. It also assumes that the responses $C(A - KC)^t K$ have bounded norm. We could remove it at the expense of more complicated bounds in the regret analysis.

Assumption 4: For all $t \ge 0$, the closed-loop matrix satisfies $\|(A - KC)^t\|_2 \le M\rho(A - KC)^t$. The Kalman filter gain K and the innovation covariance $\bar{R} \triangleq CPC^* + R$ are bounded $\|K\|_2, \|\bar{R}\|_2 \le M$. The responses are upper bounded $\sum_{t\ge 0} \|C(A - KC)^t K\| \le M(1 - \rho(A - KC))^{-1}$.

Our final assumption makes sure that system (3) is minimal.

Assumption 5: The pair (A, K) is controllable.

The above assumption is only needed to provide refined regret guarantees for stable systems (see Theorem 2 and Lemma I.4).

In addition to the previous stability properties, another property of the Kalman filter is that the innovation sequence $e_k = y_k - \hat{y}_k$ is orthogonal and, by Gaussianity, also i.i.d. By the law of large numbers, this implies that the ℓ_2 accumulative error $\sum_{k=0}^{N} \|y_k - \hat{y}_k\|_2^2$ will be of the order of $\Omega(N)$ almost surely. Predicting the true observations exactly is impossible in the stochastic noise setting, even if we know the system.

Systems (1) and (3) are statistically equivalent, i.e., they generate observations with the same distribution. This implies that the noise parameterization is not unique [42]. Another source of ill-posedness is that the state-space parameterization is nonunique. Any similar system $S^{-1}AS, CS, S^{-1}QS^{-*}$ generates the same observations. These problems will be addressed later by considering an alternative system representation.

III. Online Prediction Algorithm

Our online prediction algorithm is based on a system response representation (Markov parameterization) that has been used in subspace system identification [32]. Let p be an integer that represents how far we look into the past. We define the vector of past observations at time k:

$$Z_{k,p} \triangleq \begin{bmatrix} y_{k-p}^* & \dots & y_{k-1}^* \end{bmatrix}^*,\ k \ge p. \tag{7}$$

Define also the matrix of closed-loop responses:

$$G_p \triangleq \begin{bmatrix} C(A - KC)^{p-1}K & \dots & CK \end{bmatrix} \tag{8}$$

which are essentially the Markov parameters of the stochastic system. By expanding the Kalman filter (3) p-steps into the past, the observation at time k can be rewritten as

$$y_k = G_p Z_{k,p} + \underbrace{C(A - KC)^p \hat{x}_{k-p}}_{\text{bias}} + e_k. \tag{9}$$

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: Online Prediction Algorithm.

Input: β, λ, T_{init}
for $k = 0, \dots, T_{\text{init}} - 1$ **do**
  Observe y_k
end for
for $i = 1, 2, \dots$ **do**
  $T = 2^{i-1} T_{\text{init}}$
  $p = \beta \log T$
  $\bar{V}_{T-1} = \lambda I + \sum_{t=p}^{T-1} Z_{t,p} Z_{t,p}^*$
  $\tilde{G}_{T-1} = (\sum_{t=p}^{T-1} y_t Z_{t,p}^*) \bar{V}_{T-1}^{-1}$
  for $k = T, \dots, 2T-1$ **do**
    Predict $\tilde{y}_k = \tilde{G}_{k-1} Z_{k,p}$
    Observe y_k
    Update $\bar{V}_k = \bar{V}_{k-1} + Z_{k,p} Z_{k,p}^*$
    $\tilde{G}_k = \tilde{G}_{k-1} + (y_k - \tilde{y}_k) Z_{k,p}^* \bar{V}_k^{-1}$
  end for
end for

Instead of optimizing over system parameters A, C, and K, which results in a nonconvex optimization problem, we optimize over (the higher dimensional) G_p, which makes the problem convex. From an online learning perspective, this technique is also known as improper learning. Using this lifting, we can learn a least-squares estimate $\tilde{G}_{k,p}$ by regressing outputs y_t to past outputs $Z_{t,p}$ for $t \leq k$

$$\tilde{G}_{k,p} = \sum_{t=p}^{k} y_t Z_{t,p}^* \left(\lambda I + \sum_{t=p}^{k} Z_{t,p} Z_{t,p}^* \right)^{-1} \tag{10}$$

where $\lambda > 0$ is a regularization parameter. A similar procedure was employed in [31], [40], and [43] for system identification. To predict the next observation, we compute

$$\tilde{y}_{k+1} = \tilde{G}_{k,p} Z_{k+1,p}. \tag{11}$$

Due to the stability properties of the Kalman filter (see Section II-A), the bias term in (9) will decay fast with p, since it is premultiplied by $\rho(A - KC)^p$. Notice that for nonexplosive systems, the state $\hat{x}_{k-p}$ can grow polynomially fast in the worst case. Even if $\hat{x}_{k-p}$ remains bounded, keeping the past p constant would lead to a nonvanishing bias error (linear regret). Thus, to make sure that the prediction error decreases, we need to gradually increase the past horizon p. For this reason, inspired by the "doubling trick" [12], we divide the learning in epochs, where every epoch is twice longer than the previous one. During every epoch we keep the past horizon constant. Since $\rho(A - KC)^p$ is exponentially decreasing, it is sufficient to slowly increase the past as $p = O(\log T)$, where T is the epoch duration.

The pseudocode of our online prediction approach can be found in Algorithm 1. Each epoch lasts from time $T_i, \dots, 2T_i - 1$, where $i = 1, \dots,$ is the epoch, $T_i = 2^{i-1} T_{\text{init}}$, and T_{init} is a design parameter (the length of the first epoch). During every epoch, we keep the past $p_i = \beta \log(T_i)$ constant, where β is a design parameter. Initially, from time 0 to $T_{\text{init}} - 1$, we have a warm-up phase where we gather enough observations to start predicting. To make sure that $p_i < T_i$, we tune T_{init} accordingly. Based on Lemma III.1 it is sufficient to select $T_{\text{init}} = 2\beta \log 2\beta$. Within an epoch, the least-squares-based predictor (11) can be implemented in a recursive way:

$$\bar{V}_{k,p_i} = \bar{V}_{k-1,p_i} + Z_{k,p_i} Z_{k,p_i}^*$$
$$\tilde{G}_{k,p_i} = \tilde{G}_{k-1,p_i} + (y_k - \tilde{y}_k) Z_{k,p_i}^* \bar{V}_k^{-1}$$

which requires polynomial complexity and at most $O(\log T_i)$ memory. In the beginning of an epoch, when p_i is updated, we reinitialize the predictor based on the whole past, which requires polynomial complexity and $O(T_i)$ memory. In total, after N collected samples, the runtime is $\Theta(N \log N)$ and the memory requirement is linear $O(N)$.

No knowledge about the dynamics or even the state dimension n is required. We only need to know upper bounds on n and $(1 - \rho(A - KC))^{-1}$ in order to tune the past horizon p—see Theorem 1. There is a tradeoff between the bias error and statistical efficiency. Increasing the past horizon by selecting larger β leads to smaller bias error, but increases the variance of the estimate of G_p since we have more unknowns.

IV. Regret Analysis

In this section, we prove that with high probability, the prediction regret is not only sublinear, but also of the order of poly $\log N$ (or $\tilde{O}(1)$), where N is the number of observations collected so far. The challenge in the nonexplosive regime is that the observations grow unbounded polynomially fast ($\Omega(\sqrt{N})$). Meanwhile, recent work in finite sample analysis of system identification [15], [38], [39], [40] shows that the model parameters can be learned at a slower rate ($O(1/\sqrt{N})$). Therefore, these system identification results cannot be directly applied to obtain regret bounds for our problem. Nonetheless, we show that our online Algorithm 1 mitigates the effect of unbounded observations. As a result, the logarithmic regret bound of $\tilde{O}(1)$ remains valid even as we approach instability.

We provide two results, one for nonexplosive systems ($\rho(A) \leq 1$) and one for stable systems ($\rho(A) < 1$). Before we present the regret results, let us introduce some notion. Let $a(s) = s^d - a_{d-1} s^{d-1} \cdots - a_0$ be the minimal polynomial of matrix A, i.e., the minimum degree polynomial such that $a(A) = 0$. Denote its degree by d. We define the ℓ_1 norm of its coefficients as $\|a\|_1 \triangleq 1 + \sum_{i=0}^{d-1} |a_i|$; the ℓ_2 norm $\|a\|_2$ is defined in a similar way. Let κ be the dimension of the largest Jordan block of A. In general, $\kappa \leq d \leq n$. For brevity, we omit the first T_{init} terms from the regret. More precise bounds can be inferred by the proofs, at the cost of more complicated and less intuitive expressions.

Theorem 1 (Regret for nonexplosive systems): Consider system (3) and let Assumptions 1–4 be in effect. Fix a horizon N and a failure probability $\delta > 0$. Assume Algorithm 1 runs for N time steps with $\lambda = 1$ and let

$$\beta = \Omega\left(\frac{\kappa}{\log(1/\rho(A - KC))} \right) \tag{12}$$

and $T_{\text{init}} = 2\beta \log 2\beta$. With probability at least $1 - \delta$

$$\mathcal{R}_N \leq \text{poly}(\mathcal{C}_{\text{diff}}) O(\log^6 N) \tag{13}$$

where $\mathcal{C}_{\text{diff}} \triangleq d^\kappa n \beta \|a\|_2 \frac{M}{1 - \rho(A-KC)} \log 1/\delta$.

Theorem 1 provides the first logarithmic regret upper bounds for the problem of Kalman filter prediction. Concurrently and independently, logarithmic regret upper bounds were also proved in [29], where the upper bound is of the order of poly$(\log^\kappa T)$,

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

i.e., it scales exponentially with the size of the Jordan blocks. Here, we provide an improved logarithmic rate of $\log^6 T$. Our bounds do not depend on the stability gap $1/(1-\rho(A))$ of the open loop matrix A and they do not degrade as we approach instability $\rho(A)=1$. However, they degrade with the stability gap $1/(1-\rho(A-KC))$ of the Kalman filter closed-loop matrix. As shown in subsequent work [30], it is possible to also remove this dependence in the special case when $A-KC$ has real eigenvalues, by considering filtered versions of the past observations. Note that our regret bounds do not require persistency of excitation. Hence, they do not degrade as R approaches singularity.

The upper bound also depends on the quantities d^{κ} and $\|a\|$, both of which can be exponential in the dimension of the system state n in the worst case. This can happen, for example, if $\kappa=n$, e.g., if the system is an nth order integrator. Dependence of learning performance on the coefficients of the characteristic or minimal polynomial has been found in related work [37]. This dependence can be improved—see, for example, [28], [15], [29]. However, it is an open question whether it is possible to avoid the exponential dependence on κ. It might be possible that systems with long chain structure, e.g., integrators, are harder to learn; such systems are difficult to observe even in the known model case. In open-loop system identification [15], such dependence also appears. It might be an inherent limitation of the problem, since fundamental quantities of the system, for example, matrix A^i or the observability matrix $\mathcal{O}_i$ have norms that scale with i^{κ}—see Lemma I.1.

We can provide a tighter logarithmic rate and avoid the above issues in the case of stable systems ($\rho(A)<1$), by exploiting stationarity and persistency of excitation. Let $\Gamma_k \triangleq \mathbb{E}\hat{x}_k\hat{x}_k^*$ be the covariance of the state with $\Gamma_\infty=\lim_{k\to\infty}\Gamma_k$ the steady-state covariance. Let the minimum singular value of R be denoted by σ_R. Let the *mixing time* τ_{mix} be the minimum time such that the system is close to the steady state:

$$\|\Gamma_\infty^{-1/2}\Gamma_k\Gamma_\infty^{-1/2}\|_2 \geq 1/2, \text{ for all } k \geq \tau_{\text{mix}}. \tag{14}$$

Based on Lemma I.4, the mixing time scales with

$$\tau_{\text{mix}} \leq \text{poly}((\log 1/\rho(A))^{-1}, \kappa, \log \tfrac{\sigma_{\max}(\Gamma_\infty)}{\sigma_{\min}(\Gamma_\infty)}).$$

We obtain the following guarantees.

Theorem 2 (Regret for stable systems): Consider system (3) and let Assumptions 1–5 be in effect with $\rho(A)<1$. Fix a horizon N and a failure probability $\delta>0$. Assume that Algorithm 1 runs for N time steps with $\lambda=1$, β as in (12), and $T_{\text{init}}=2\beta\log 2\beta$. There exists

$$N_0^{\text{s}} = \text{poly}\,(n,\beta,\log 1/\delta,\tau_{\text{mix}},1/\sigma_R)$$

such that with probability at least $1-\delta$, if $N>N_0^{\text{s}}$, then

$$\mathcal{R}_N \leq \text{poly}(\mathcal{C}_{\text{diff}}^{\text{s}})O(\log^4 N) \tag{15}$$

where $\mathcal{C}_{\text{diff}}^{\text{s}} \triangleq n\beta\frac{M}{1-\rho(A-KC)}\log 1/\delta$.

Notice that for stable systems the dominant logarithmic term in the upper bound is smaller. Also, we no longer have quantities that depend exponentially on the dimension n. The main bound (15) does not depend on the stability gap $1/(1-\rho(A))$. However, via τ_{mix} in N_0^{s}, the guarantees depend logarithmically on the inverse radius $\log 1/\rho(A)$. This quantity reflects the mixing time needed for a stable system to approach stationarity. The burn-in time N_0^{s} is also related to persistency of excitation conditions, i.e., we need enough samples to guarantee that the Gram matrix $\bar{V}_k$ is well conditioned.

The proofs of Theorems 1 and 2 can be found in the Appendix. In the next section, we provide an overview of the regret analysis and explain why the quantities d^{κ} and $\|a\|_2$ appear in the bound in Theorem 1. We also explain how we obtain tighter bounds by exploiting stationarity and persistency of excitation in the case of stable systems.

A. Regret Analysis for General Nonexplosive Systems

Recall the definition of the innovation error $e_k=y_k-\hat{y}_k$. For brevity, we also define the error $\tilde{e}_k \triangleq \tilde{y}_k-\hat{y}_k$ between the online prediction of Algorithm 1 and the Kalman Filter prediction. Adding and subtracting $\hat{y}_k$ in the first term

$$\begin{aligned}\mathcal{R}_N &= \sum_{k=T_{\text{init}}}^{N} \|e_k+\hat{y}_k-\tilde{y}_k\|_2^2 - \|e_k\|_2^2 \\ &= \underbrace{\sum_{k=T_{\text{init}}}^{N}\|\hat{y}_k-\tilde{y}_k\|_2^2}_{\mathcal{L}_N} + 2\underbrace{\sum_{k=T_{\text{init}}}^{N} e_k^*\,(\hat{y}_k-\tilde{y}_k)}_{\text{martingale term}}.\end{aligned}$$

It is sufficient to prove that the cumulative square loss

$$\mathcal{L}_N \triangleq \sum_{k=T_{\text{init}}}^{N}\|\hat{y}_k-\tilde{y}_k\|_2^2 \tag{16}$$

is logarithmic in N. Because the innovations are i.i.d., we have a martingale structure for the second term since $e_k\in\mathcal{F}_k$, while $\tilde{e}_k\in\mathcal{F}_{k-1}$. The martingale term is small and can be bounded in terms of the square loss $\mathcal{L}_N$. In particular, the quantity

$$(\mathcal{L}_N+1)^{-1/2}\sum_{k=T_{\text{init}}}^{N} e_k^*\,(\hat{y}_k-\tilde{y}_k)$$

is a self-normalized martingale and can be analyzed based on the following theorem, which can be found in [40]. It is an extension of [44], Th. 1] and [36], Prop. 8.2].

Theorem 3 (Self-normalized martingales [40]): Let $\{\mathcal{F}_t\}_{t=0}^{\infty}$ be a filtration. Let $\eta_t\in\mathbb{R}^m$, $t\geq 0$ be $\mathcal{F}_t$-measurable, independent of $\mathcal{F}_{t-1}$. Suppose also that $\eta_t\sim\mathcal{N}(0,I)$ is isotropic Gaussian. Let $X_t\in\mathbb{R}^d$, $t\geq 0$ be $\mathcal{F}_{t-1}-$measurable. Assume that W is a $d\times d$ positive definite matrix. For any $t\geq 0$, define

$$\bar{W}_t = W+\sum_{s=1}^{t}X_sX_s^*, \qquad D_t=\sum_{s=1}^{t}X_sH_s^*$$
$$H_s^* = [\eta_s^* \quad \dots \quad \eta_{s+r-1}^*]\in\mathbb{R}^{rm}$$

for some integer r. Then, for any $\delta>0$, with probability at least $1-\delta$, for all $t\geq 0$

$$\left\|\bar{W}_t^{-1/2}D_t\right\|_2^2 \leq 8r\left(\log\frac{r5^m}{\delta}+\log\det\bar{W}_tW^{-1}\right).$$

An application of the above theorem with $\bar{W}_t=\mathcal{L}_t$, $H_t=e_t$ (see the proof of Theorem 1 in the Appendix) implies that

$$\sum_{k=T_{\text{init}}}^{N} e_k^*\,(\hat{y}_k-\tilde{y}_k) = \tilde{O}(\sqrt{\mathcal{L}_N})$$

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

with high probability. Hence, we focus on bounding $\mathcal{L}_N$.

To complete the proof, it is sufficient to show that within a fixed epoch, the cumulative square loss is logarithmic—see Theorem I.2. Then, the same will hold for the total square loss $\mathcal{L}_N$, since the number of epochs is logarithmic with N. Hence, for the remaining section, we will assume that we are within one epoch i, i.e., the past horizon $p = p_i$ and $T = 2^{i-1}T_{\text{init}}$ are constant. For brevity, we omit the subscript p from all variables and write $G, \tilde{G}_k,$ and Z_k instead of $G_p, \tilde{G}_{k,p},$ and $Z_{k,p}$.

Define $S_{k-1} \triangleq \sum_{i=p}^{k-1} e_i Z_i^*$ and $\bar{V}_{k-1} \triangleq \lambda I + \sum_{i=p}^{k-1} Z_i Z_i^*$. Then, the error between our online prediction and the Kalman filter prediction can be written as

$$\begin{aligned}\tilde{e}_k &= (\tilde{G}_{k-1} - G)Z_k - C(A-KC)^p \hat{x}_{k-p} \\ &= \underbrace{S_{k-1}\bar{V}_{k-1}^{-1} Z_k}_{\text{regression}} + \underbrace{\lambda G \bar{V}_{k-1}^{-1} Z_k}_{\text{regularization}} \\ &\quad + \underbrace{C(A-KC)^p \left(\sum_{i=T}^{k-1} \hat{x}_{i-p} Z_i^* \bar{V}_{k-1}^{-1} Z_k - \hat{x}_{k-p} \right)}_{\text{truncation bias}}. \quad (17)\end{aligned}$$

The regression term is due to the noise e_k perturbing the system. The truncation bias is due to using only p past observations and not all of them. The key ingredients to analyze the cumulative error $\mathcal{L}_N$ are 1) the stability properties of the closed-loop matrix $A - KC$; 2) self-normalization properties of predictor (11).

Regression term: We can rewrite the regression term as a product of three separate terms:

$$S_{k-1}\bar{V}_{k-1}^{-1} Z_k = (S_{k-1}\bar{V}_{k-1}^{-1/2})(\bar{V}_{k-1}^{-1/2}\bar{V}_k^{1/2})(\bar{V}_k^{-1/2} Z_k).$$

Hence, the regression term is upper bounded by

$$\sqrt{Z_k^* \bar{V}_k^{-1} Z_k} \sup_{T \le t \le 2T-1} \left(\|S_{t-1}\bar{V}_{t-1}^{-1/2}\| \|\bar{V}_{t-1}^{-1/2}\bar{V}_t^{1/2}\|_2 \right).$$

The first term $Z_k^* \bar{V}_k^{-1} Z_k$ is self-normalized since $\bar{V}_k$ is the Gram matrix of $Z_k, \dots, Z_p$. We control its cumulative growth using the following lemma that is inspired by [33].

Lemma 1: Recall that $\bar{V}_k = \lambda I + \sum_{i=p}^{k} Z_i Z_i^*$. We have

$$\sum_{k=T}^{2T-1} Z_k \bar{V}_k^{-1} Z_k \le \log\det(\bar{V}_{2T-1}\bar{V}_{T-1}^{-1}).$$

Intuitively, $Z_k Z_k^*$ appears in $\bar{V}_k$ and, hence, it cancels out the effect of Z_k. Hence, to prove that the cumulative square loss is logarithmic, it is sufficient to upper-bound the supremum of the last terms. The second term, $S_{k-1}\bar{V}_{k-1}^{-1/2}$ is a self-normalized martingale and can be bounded based on Theorem 3, where $\bar{V}_t, S_t, Z_t,$ and e_t assume the role of $\bar{W}_t, D_t, X_t,$ and $H_t,$ respectively—see Theorem I.1 in the Appendix for details. Thus, the term $\|S_{k-1}\bar{V}_{k-1}^{-1/2}\|_2$ grows at most logarithmically with k. The martingale property comes from the fact that the innovation process e_k of the Kalman filter is i.i.d.—see Section II-A. The third term can be rewritten as

$$\|\bar{V}_{k-1}^{-1/2}\bar{V}_k^{1/2}\|_2^2 = 1 + \|\bar{V}_{k-1}^{-1/2} Z_k\|_2^2$$

where the equality follows from

$$\bar{V}_{k-1}^{-1/2}\bar{V}_k\bar{V}_{k-1}^{-1/2} = I + \bar{V}_{k-1}^{-1/2} Z_k Z_k^* \bar{V}_{k-1}^{-1/2}.$$

Now, we can exploit the fact that Z_k does not change too fast compared to the most recent past $Z_{k-1}, \dots, Z_{k-n}$.

Lemma 2 (ARMA-like representation): Let $y_0, y_1, \dots$ be observations generated by system (1). Fix a past horizon p and let a be the minimal polynomial of A with degree d. Then, the past observations satisfy the following recursion:

$$Z_k = a_{d-1} Z_{k-1} + \cdots + a_0 Z_{k-d} + \delta_k \quad (18)$$

where $\delta_k \in \mathcal{F}_{k-1}$ with

$$\|\delta_k\|_2 \le \Delta \sup_{i \le k-1} \|e_i\|_2 \quad (19)$$

and $\Delta = O(d^{\kappa-1}\|a\|_1\sqrt{p})$.

Intuitively, the unbounded components of Z_k are captured by the recent history $Z_{k-1}, \dots, Z_{k-d}$ and the residual δ_k is bounded. Replacing Z_k with (18), we obtain by two Cauchy–Schwarz inequality applications

$$\begin{aligned}\|\bar{V}_{k-1}^{-1/2} Z_k\|_2^2 &\le 2\|a\|_2^2 \sum_{i=0}^{d-1} \|\bar{V}_{k-1}^{-1/2} Z_{k-d+i}\|_2^2 + 2\|\bar{V}_{k-1}^{-1/2}\delta_k\|_2^2 \\ &\le 2\|a\|_2^2 d + 2\lambda\Delta^2 \sup_{i \le k-1} \|e_i\|_2^2\end{aligned}$$

where we used the fact that the terms $\bar{V}_{k-1}^{-1/2} Z_{k-d+i}$ in the sum are normalized and their norm is upper-bounded by 1.

Regularization and truncation terms: For the regularization term, we follow the same steps as with the regression one. Since matrix $A - KC$ is stable, the truncation term decreases exponentially with p. System quantity κ governs how fast the observations grow polynomially. Parameter β should be large enough to cancel out this polynomial rate. This explains why κ affects the choice of β in (12)—see Theorem I.2 for details.

B. Refined Analysis for Stable Systems

If $\rho(A) < 1$, it is sufficient to write the regression term as a product of two terms

$$S_{k-1}\bar{V}_{k-1}^{-1} Z_k = (S_{k-1}\bar{V}_{k-1}^{-1/2})(\bar{V}_{k-1}^{-1/2} Z_k).$$

We exploit the fact that Z_k converges exponentially fast to a stationary distribution. Hence, the term $\bar{V}_{k-1}^{-1/2} Z_k$ will effectively be self-normalized, without the need to express Z_k as a function of the past observations. The key ingredient toward providing stronger regret guarantees is proving that persistency of excitation for the past observations holds with high probability. In particular, we prove a novel persistency of excitation result, which strengthens prior results—see for example [40]. Similar to the definition of the state covariance, we define the past outputs' covariance $\Gamma_{Z,k} \triangleq \mathbb{E} Z_k Z_k^*$. We have the following.

Lemma 3 (Uniform persistency of excitation: Stable case): Consider the conditions of Theorem 2. Select a failure probability $\delta > 0$. Let $T = 2^{i-1}T_{\text{init}}$ for some fixed epoch i with $p = \beta \log T$ the corresponding past horizon. There exists a $N_0^{\text{s}} = \text{poly}(n, \beta, \tau_{\text{mix}}, 1/\sigma_R, \log 1/\delta)$ such that if $T \ge N_0^{\text{s}}$, with probability at least $1 - \delta$

$$\sum_{j=p}^{k-1} Z_j Z_j^* \succeq \frac{k}{8}\Gamma_{Z,\infty} \quad (20)$$

uniformly for all $T \le k \le 2T - 1$.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

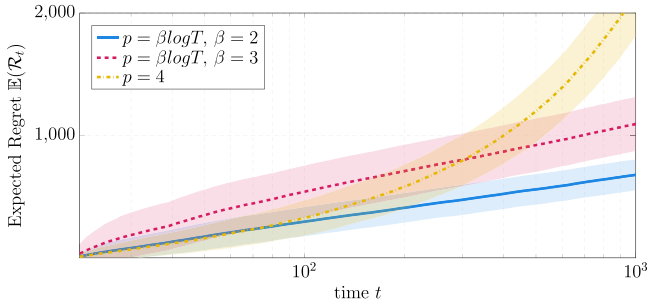


Fig. 1. (Empirical) expected regret for various choices of the past horizon. The shaded regions capture one (empirical) standard deviation. We indeed obtain logarithmic regret (the x-axis is logarithmic) if the past horizon increases logarithmically as $\beta \log t$. On the other hand, if we keep the past horizon constant ($p = 4$), we suffer from linear regret.

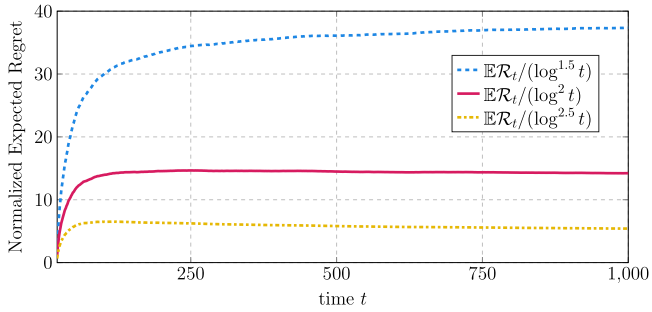


Fig. 2. Expected regret, normalized with different logarithmic rates. It seems that the logarithmic rate of our algorithm is $\log^2 T$.

Hence, the term $\bar{V}_{k-1}^{-1/2} Z_k$ can be bounded by

$$\left\| \bar{V}_{k-1}^{-1/2} Z_k \right\|_2 \leq O\left(\frac{1}{\sqrt{k}}\right) \left\| \Gamma_{Z,k}^{-1/2} Z_k \right\|_2$$

where now the normalized term $\Gamma_{Z,k}^{-1/2} Z_k$ behaves like a standard isotropic Gaussian variable—see Theorem II.1.

V. Simulations

In this section, we perform simulations to verify that our algorithm exhibits logarithmic regret. We also study how the past horizon p affects performance. Let

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0.9 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, Q = I, R = 1.$$

The system is nonexplosive with $\kappa = 2$. We select $T_{\text{init}} = 20$, $\lambda = 1$, and $N = 1000$. We ignore the errors for the initial epoch until T_{init}. We perform 1000 Monte Carlo simulations. In Fig. 1, we show the (empirical) expected regret for various choices of the past horizon, where the shaded regions capture one (empirical) standard deviation. We indeed obtain logarithmic regret (note that the x-axis is logarithmic), if we let the past horizon p increase slowly with time as $\beta \log t$. Note that this slow growth of p is essential for obtaining logarithmic regret under our algorithm. As shown in Fig. 1, under a constant past horizon, e.g., $p = 4$, we suffer from linear regret. Nonetheless, if the past horizon increases too fast, even if the regret is logarithmic, it might exhibit large variance. For instance, if we choose $\beta = 2$, we obtain smaller regret with smaller variance compared to $\beta = 3$. In Fig. 2, with $\beta = 2$, we investigate the logarithmic order of the regret by normalizing with different logarithmic powers. It seems that the actual rate of our algorithm is of the order of $\log^2 T$, suggesting that our regret upper bounds could be further improved. We observed similar results for both stable and nonexplosive systems.

VI. Discussion and Extensions

In this section, we discuss implications and generalizations of Algorithm 1 and the regret analysis.

Comparison with online LQR: Our bounds show that the problem of learning the Kalman filter online is provably easier than the online LQR, in the case of unknown systems. The latter requires in general regret of the order of $\Omega(\sqrt{N})$ [16], [17]. This is another reason why the problems are not dual in the unknown model case [26]. This gap might be expected since in the case of control, there is a need for exogenous exploratory signals, which increase the LQR control cost.

f-steps ahead predictor: An immediate generalization of Algorithm 1 is to consider the f-steps ahead predictor, where f is some future horizon. Instead of predicting only the next observation, we predict the sequence $y_k, \ldots, y_{k+f}$. Denote the future observations and noises by

$$Y_k = \begin{bmatrix} y_k^* & \cdots & y_{k+f}^* \end{bmatrix}^*, \quad E_k^+ = \begin{bmatrix} e_k^* & \cdots & e_{k+f}^* \end{bmatrix}^*.$$

Similar to (7), we can establish a regression:

$$Y_k = \mathcal{O}_f \mathcal{K}_p Z_k + \mathcal{O}_f (A - KC)^p \hat{x}_{k-p} + \mathcal{T}_f E_k^+$$

where $\mathcal{K}_p \triangleq [(A - KC)^{p-1} K \ \cdots \ K]$, and $\mathcal{T}_f$ is a lower triangular block Toeplitz matrix generated by $I, CK, \ldots, CA^{f-2}K$. The optimal Kalman filter predictor is

$$\hat{Y}_k = \mathcal{O}_f \mathcal{K}_p Z_k + \mathcal{O}_f (A - KC)^p \hat{x}_{k-p}.$$

Hence, the regret can be defined as in (4), with the lowercase y replaced with uppercase Y. The online predictor (11) can be adapted here:

$$\tilde{Y}_k = \tilde{G}_{k,f,p} Z_k$$

where $\tilde{G}_{k,f,p}$ is obtained similar to (10) by regressing future observations Y_t to past observations Z_t from time p up to $k - f$. The logarithmic regret guarantees of $\tilde{O}(1)$ also hold with the final bound depending polynomially on f and $\|\mathcal{T}_f\|_2$.

Prediction with inputs: We can apply the same algorithm in the case of exogenous inputs u_k:

$$x_{k+1} = Ax_k + Bu_k + w_k$$
$$y_k = Cx_k + v_k$$

where B is some *unknown* input matrix. Define now the vector

$$Z_{k,p}^{\text{c}} \triangleq \begin{bmatrix} y_{k-p}^* & \cdots & y_{k-1}^* & u_{k-p}^* & \cdots & u_{t-1}^* \end{bmatrix}^*$$

of both past outputs and past inputs. Following identical arguments as in the derivation of (9), we arrive at

$$y_k = \begin{bmatrix} G_p & G_p^{\text{in}} \end{bmatrix} Z_{k,p}^{\text{c}} + e_k + C(A - KC)^p \hat{x}_{k-p}$$

where G_p is defined in (8) and

$$G_p^{\text{in}} \triangleq \begin{bmatrix} C(A - KC)^{p-1} B & \cdots & CB \end{bmatrix}.$$

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

TABLE 1
NOTATIONS FOR FIXED PAST HORIZON p

Z_t	$\triangleq \begin{bmatrix} y_{t-p}^* & \cdots & y_{t-1}^* \end{bmatrix}^*$	Past outputs at time t
E_t	$\triangleq \begin{bmatrix} e_{t-p}^* & \cdots & e_{t-1}^* \end{bmatrix}^*$	Past noises at time t
$\bar{Z}_k$	$\triangleq \begin{bmatrix} Z_p & \cdots & Z_k \end{bmatrix}$	Batch past outputs up to time k
$\bar{E}_k$	$\triangleq \begin{bmatrix} E_p & \cdots & E_k \end{bmatrix}$	Batch past noises up to time k
$\bar{X}_k$	$\triangleq \begin{bmatrix} \hat{x}_0 & \cdots & \hat{x}_{k-p} \end{bmatrix}$	Batch past states up to time k
S_k	$\triangleq \sum_{t=p}^k e_t Z_t^*$	Correlation of current noise with past outputs
V_k	$\triangleq \bar{Z}_k \bar{Z}_k^* = \sum_{t=p}^k Z_t Z_t^*$	Gram matrix of past outputs
$\bar{V}_k$	$\triangleq \lambda I + V_k$	Regularized Gram matrix of past outputs
$\bar{R}$	$\triangleq \mathbb{E} e_k e_k^*$	Covariance of innovations
$\mathcal{T}_k$	$\triangleq \mathrm{Toep}\left(I, CK, \ldots, CA^{k-2}K\right)$	See (I.4), Toeplitz matrix of I_m and system responses CA^tK
Σ_E	$\triangleq \mathbb{E}\mathcal{T}_p E_t E_t^* T_p^*$	Covariance of weighted past noises
σ_R	$\triangleq \sigma_{\min}(R)$	Smallest singular value of R
Γ_t	$\triangleq \mathbb{E}\hat{x}_t \hat{x}_t^*$	Covariance of Kalman filter state prediction
$\Gamma_{Z,t}$	$\triangleq \mathbb{E} Z_t Z_t^*$	Covariance of past outputs
G	$\triangleq \begin{bmatrix} C(A-KC)^{p-1}K & \ldots & CK \end{bmatrix}$	Kalman filter responses

To predict, we follow exactly Algorithm 1, where $Z_{k,p}$ is replaced with $Z_{k,p}^{\mathrm{c}}$, i.e., instead of using only the past outputs, we also use the past inputs. We can obtain logarithmic regret guarantees and the proof (nonexplosive case) still holds, as long as we can establish an upper bound on the input energy [similar to (35)]. Some sufficient conditions include 1) bounded inputs or 2) closed-loop inputs (see [43]) that render the closed-loop system stable. Due to space limitations, we leave the detailed analysis for future work.

VII. CONCLUSION

In this article, we provided the first logarithmic regret upper bounds for learning the classical Kalman filter of an unknown system with unknown stochastic noise. Our regret analysis holds for nonexplosive systems and our bounds do not degrade with the system stability gap. Going forward, our article opens up several research directions. Our current bounds are mainly qualitative and data independent, focusing on how various system-theoretic properties affect learning. It is an open problem to develop sharper data-dependent bounds that are more suitable for control applications, possibly at the cost of losing some interpretability. Another interesting direction is to study how the learning performance is affected by system-theoretic properties, such as the exponential quantity d^κ in the case of systems with long chain structure, e.g., κ-order integrators. Analyzing the regret of other online algorithms, e.g., extended least squares, is also an open problem. Another challenging problem for both prediction and system identification is the case of explosive systems. Although in the fully observed case, this problem has been studied [34], [36], it remains open in the case of partially observable systems.

APPENDIX I
REGRET ANALYSIS FOR NONEXPLOSIVE SYSTEMS

A. Notation and Organization of the Appendix

Structure: In Sections I-B and I-C, we review fundamental results from system theory and statistics. These provide the main tools for proving Theorems 1 and 2. In Section I-D, we provide finite sample complexity bounds and persistency of excitation (PE) results for a fixed time k and fixed past horizon p. In Section I-E, we generalize those results from pointwise to uniform over all times k in one epoch. In Section I-F, we prove Lemma 1. By combining the uniform bounds and Lemma 1, we prove in Section I-G that the square loss suffered within one epoch is logarithmic with respect to the length of the epoch. Hence, we can now prove Theorem 1—see Section I-H. In Section II, we analyze the case of stable systems and prove Theorem 2. Section III includes some technical results about logarithmic inequalities, which are used to show that the burn-in time N_0 depends polynomially on the various system parameters.

Notation: A summary of the notation can be found in Table 1. We will analyze the performance of Algorithm 1 based mainly on a fixed epoch i. Since the past horizon p is kept constant during an epoch, we drop the index p from $Z_{k,p}$, G_p, $\tilde{G}_{k,p}$, and $\bar{V}_{k,p}$ and write Z_k, G, $\tilde{G}_k$, and $\bar{V}_k$ instead. Similar to the past outputs Z_k, we also define the past noises:

$$E_k \triangleq \begin{bmatrix} e_{t-p}^* & \cdots & e_{t-1}^* \end{bmatrix}^*. \tag{21}$$

The batch past outputs, batch past noises, and batch past Kalman filter states are defined as

$$\bar{Z}_k \triangleq \begin{bmatrix} Z_p & \cdots & Z_k \end{bmatrix}, \bar{E}_k \triangleq \begin{bmatrix} E_p & \cdots & E_k \end{bmatrix}$$
$$\bar{X}_k \triangleq \begin{bmatrix} \hat{x}_0 & \cdots & \hat{x}_{k-p} \end{bmatrix}. \tag{22}$$

Recall the definition of the correlations between the current innovation and the past outputs $S_k \triangleq \sum_{t=p}^k e_t Z_t^*$ and the regularized autocorrelations of past outputs $\bar{V}_k \triangleq \lambda I + \bar{Z}_k \bar{Z}_k^*$. The innovation sequence e_k is i.i.d. zero mean Gaussian. Its covariance has a closed-form expression and is defined as

$$\bar{R} \triangleq \mathbb{E} e_k e_k^* = CPC^* + R \tag{23}$$

where P is the solution to the Riccati equation (6). Next, we define the Toeplitz matrix $\mathcal{T}_k$, for some $k \geq 1$

$$\mathcal{T}_k \triangleq \begin{bmatrix} I_m & 0 & & 0 \\ CK & I_m & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ CA^{k-2}K & CA^{k-3}K & \cdots & I_m \end{bmatrix}. \tag{24}$$

The past outputs can be written as

$$Z_t = \mathcal{O}_p \hat{x}_{t-p} + \mathcal{T}_p E_t \tag{25}$$

The covariance of $\mathcal{T}_p E_t$ is denoted by

$$\Sigma_E \triangleq \mathbb{E}\mathcal{T}_p E_t E_t^* \mathcal{T}_p^* = \mathcal{T}_p \,\mathrm{diag}(\bar{R}, \ldots, \bar{R}) \mathcal{T}_p^*. \tag{26}$$

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

We define the covariance of the state predictions

$$\Gamma_k \triangleq \mathbb{E}\hat{x}_k\hat{x}_k^* \tag{27}$$

and the covariance of the past outputs

$$\Gamma_{Z,k} \triangleq \mathbb{E}Z_kZ_k^* = \mathcal{O}_p\Gamma_{k-p}\mathcal{O}_p^* + \Sigma_E. \tag{28}$$

B. System-Theoretic Bounds

1) Bounds for System and Covariance Matrices:

Lemma I.1: Consider matrix A with Jordan form SJS^{-1}. Let κ be the largest Jordan block of A. Under Assumption 1

$$\left\|A^i\right\|_2 \le M^2\rho(A)^{i-\kappa+1}\left(\frac{ei}{\kappa-1}\right)^{\kappa-1}. \tag{29}$$

As a result the following bounds hold:

$$\left\|A^t\right\|_2 \le M^2O\left(t^{\kappa-1}\right), \qquad \|\mathcal{O}_t\|_2 \le M^3O\left(t^{\kappa}\right)$$
$$\|\mathcal{T}_t\|_2 \le M^3O\left(t^{\kappa}\right), \qquad \|\Gamma_t\|_2 \le M^7O\left(t^{2\kappa-1}\right).$$

Proof: Recall the Jordan form of $A = SJS^{-1}$. For simplicity, assume that J is equal to an $n \times n$ Jordan block with eigenvalue $\lambda = \rho(A)$. Then, we have that

$$J^i = \begin{bmatrix} \lambda^i & \binom{i}{1}\lambda^{i-1} & \dots & \binom{i}{n-1}\lambda^{i-n+1} \\ 0 & \lambda^i & \dots & \binom{i}{n-2}\lambda^{i-n+2} \\ & & \ddots & \\ 0 & 0 & \dots & \lambda^i \end{bmatrix}.$$

By Lemma I.2 and Assumption 1, we obtain

$$\left\|A^i\right\|_2 \le M^2\lambda^{i-n+1}\sum_{k=0}^{n-1}\binom{i}{k} \le M^2\lambda^{i-n+1}\left(\frac{ei}{n-1}\right)^{n-1}$$

where the second inequality is standard [11, Exercise 0.0.5]. This completes the proof of (29). The proof for the other cases is similar since the Jordan matrix is block diagonal.

The bounds on $\mathcal{O}_t, \mathcal{T}_t, \Gamma_t$ follow from (29). □

Lemma I.2 (Toeplitz norm [40]): Let $L \in \mathbb{R}^{m_1n\times m_2n}$ for some integers n, m_1, and m_2 be a block triangular Toeplitz matrix:

$$L = \text{Toep}(L_1, L_2, \dots, L_n)$$

where $L_i \in \mathbb{R}^{m_1\times m_2}$. Then, $\|L\|_2 \le \sum_{i=1}^n \|L_i\|_2$.

Lemma I.3 (Monotonicity [40]): Consider system (3), with $\Gamma_k \triangleq \mathbb{E}\hat{x}_k\hat{x}_k^*$. The sequence Γ_k is increasing in the positive semidefinite cone.

Lemma I.4 (Mixing time): Assume that system (3) is stable $\rho(A) < 1$. Recall the definition of $\bar{R}$ in (23). Under Assumption 5, the sequence Γ_k converges to $\Gamma_\infty \succ 0$, the unique positive definite solution of the Lyapunov equation:

$$\Gamma_\infty = A\Gamma_\infty A^* + K\bar{R}K^*.$$

Let the mixing time τ_{mix} be defined as in (14). Then,

$$\tau_{\text{mix}} \le \frac{1}{\log 1/\rho(A)}\tilde{O}(\max\{\log \text{cond}(\Gamma_\infty), \kappa\})$$

with $\text{cond}(\Gamma_\infty) = \frac{\sigma_{\max}(\Gamma_\infty)}{\sigma_{\min}(\Gamma_\infty)}$.

Proof: Since A is stable, $\Gamma_\infty = \sum_{k=0}^\infty A^kK\bar{R}K^*(A^*)^k$ is well defined and solves the Lyapunov equation. Since (A, K) is controllable (by Assumption 5), Γ_∞ is strictly positive definite: $\Gamma_\infty \succeq \mathcal{K}_n\mathcal{K}_n^* \succ 0$, where the following matrix has full rank:

$$\mathcal{K}_n \triangleq \begin{bmatrix} K\bar{R}^{1/2} & AK\bar{R}^{1/2} & \dots A^{n-1}K\bar{R}^{1/2} \end{bmatrix}.$$

It is unique since the operator $\mathcal{L}(M) = M - AMA^*$ is invertible; its eigenvalues are bounded below by $1 - \rho^2(A)$.

Notice that $\Gamma_0 = 0 \preceq \Gamma_\infty$ and by induction, we can show that $\Gamma_k \preceq \Gamma_\infty$. Since Γ_k is also monotone, it converges to the unique Γ_∞. Now after some algebra, we obtain

$$\Gamma_k - \Gamma_\infty = -A^k\Gamma_\infty(A^*)^k.$$

It is sufficient to find a τ_{mix} such that

$$\|A^{\tau_{\text{mix}}}\|^2\,\sigma_{\max}(\Gamma_\infty) \le \frac{\sigma_{\min}(\Gamma_\infty)}{2}.$$

Since the norm of matrix A grows as fast as $\|A^{\tau_{\text{mix}}}\|_2 = O(\rho(A)^{\tau_{\text{mix}}-\kappa+1}{\tau_{\text{mix}}}^\kappa)$, it is sufficient to pick

$$\tau_{\text{mix}} \ge \frac{\kappa\log\tau_{\text{mix}}}{\log\frac{1}{\rho(A)}} + \frac{\log\text{cond}(\Gamma_\infty)/2}{\log\frac{1}{\rho(A)}} + \kappa - 1.$$

By Lemma III.1, the order of τ_{mix} is

$$\tau_{\text{mix}} = \frac{1}{\log 1/\rho(A)}\tilde{O}(\max\{\log\text{cond}(\Gamma_\infty), \kappa\}).$$

2) Proof of Lemma 2:

For the plain observations y_k

$$y_{k-t} = CA^{d-t}\hat{x}_{k-d} + \sum_{s=t+1}^{d} CA^{s-t-1}Ke_{k-s} + e_{k-t}$$

for $t = 0, \dots, d$. By the definition of the minimal polynomial

$$CA^d\hat{x}_k = a_{d-1}CA^{d-1}\hat{x}_{k-1} + \dots + a_0C\hat{x}_{k-d}$$

which leads to

$$y_k = a_{d-1}y_{k-1} + \dots + a_0y_{k-d} + \sum_{s=0}^{d} L_se_{k-s}$$

with $L_0 = I$ and

$$L_s = -a_{d-s}I_m - \sum_{t=1}^{s-1} a_{d-s+t}CA^{t-1}K + CA^{s-1}K.$$

The norm of the above matrices is upper-bounded by

$$\|L_s\|_2 \le \|a\|_1\,\|C\|_2\,\|K\|_2 \max_{0\le i\le d}\left\|A^{i-1}\right\|_2 \tag{30}$$

where $\|a\|_1$ denotes the ℓ_1 norm of the polynomial coefficients. The same will now hold for the past outputs:

$$Z_k = a_{d-1}Z_{k-1} + \dots + a_0Z_{k-p} + \sum_{s=0}^{d}\text{diag}(L_s, \dots, L_s)E_{k-s}$$

where E_k is the vector of past noises. We can bound the residual $\delta_k \triangleq \sum_{s=0}^d \text{diag}(L_s, \dots, L_s)E_{k-s}$ by

$$\|\delta_k\|_2 \le \Delta\,\|e_s\|_2, \text{ where}$$
$$\Delta \triangleq (d+1)\max_{0\le s\le d}\|L_s\|_2\sqrt{p}. \tag{31}$$

From (29) and (30), it follows that $\Delta = O(d^{\kappa-1}\|a\|_1\sqrt{p})$. □

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

C. Statistical Toolbox

1) Least Singular Value of Toeplitz Matrix: Let $u_t \in \mathbb{R}^m$, $t = 0, \dots,$ be an i.i.d. sequence, where $u_k \sim \mathcal{N}(0, I)$ are isotropic Gaussians. The following result, which is adapted from [45], shows that the Toeplitz matrix of u_t is well conditioned with high probability. Similar results appeared in [38] and [39].

Lemma I.5 (Toeplitz Isometry [45]): Let $u_t \in \mathbb{R}^m, t = 0, \dots,$ be an i.i.d. sequence of Gaussian variables with unit covariance matrix. Consider the Toeplitz matrix

$$U = \begin{bmatrix} u_{k-p} & u_{k-p-1} & \dots & u_0 \\ u_{k-p+1} & u_{k-p} & \dots & u_1 \\ \vdots & & & \\ u_{k-1} & u_{k-2} & \dots & u_{p-1} \end{bmatrix}.$$

Fix a failure probability $0 < \delta < 1/2$. There exists a universal constant C such that if

$$k \geq Cpm \log(pm/\delta)$$

then with probability at least $1 - \delta$

$$\tfrac{1}{2} kI \preceq UU^* \preceq \tfrac{3}{2} kI. \tag{32}$$

Proof: Define $\bar{U} = [U \quad U_0]$, where

$$U_0 = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ u_0 & 0 & \dots & 0 & 0 \\ \vdots & & & & \\ u_{p-2} & u_{p-3} & \dots & u_0 & 0 \end{bmatrix}.$$

Notice that we have

$$\|UU^* - kI\|_2 \leq \|\bar{U}\bar{U}^* - kI\|_2 + \|U_0 U_0^*\|_2$$

By [45, Th. A.2], it follows that with probability at least $1 - \delta/2$

$$\left\|\bar{U}\bar{U}^* - kI\right\|_2 \leq C'(pm \log(k/\delta) + \sqrt{kpm \log(k/\delta)})$$

for some universal constant C'. To bound $U_0 U_0^*$, we follow the same steps as in the proof of [45, Th. A.2] (see bounds on $U_2 U_2^*$ there). With probability at least $1 - \delta/2$

$$\|U_0 U_0^*\|_2 \leq C'' pm \log(pm/\delta)$$

for some other universal constant C''. By a union bound and if we select $k \geq Cpm \log(pm/\delta)$, with C large enough, we get that with probability at least $1 - \delta$

$$\|UU^* - kI\|_2 \leq k/2.$$

The condition on k makes use of Lemma III.1. □

2) Gaussian Suprema:

Lemma I.6: Consider $v_t \in \mathbb{R}^d \sim \mathcal{N}(0, I)$ i.i.d., for $t = 1, \dots, k$. Let $X_k \in \mathbb{R}^q$ be a linear combination:

$$X_k \triangleq \sum_{t=1}^{k} M_{k,t} v_t, \text{ for } k = 1, \dots, T$$

where $M_{t,k} \in \mathbb{R}^{q \times d}$. For some $\mu > 0$, define

$$\Sigma_k \triangleq \mu I + \mathbb{E} X_k X_k^*.$$

Fix a failure probability $\delta > 0$. With probability at least $1 - \delta$

$$\sup_{k=1,\dots,T} \|\Sigma_k^{-1/2} X_k\|_2 \leq \sqrt{q} + \sqrt{2 \log \frac{T}{\delta}}. \tag{33}$$

If $\mathbb{E} X_k X_k^*$ and $k \leq T$ are invertible, the result holds for $\mu = 0$.

Proof: Fix a k. An application of Jensen's inequality gives

$$\mathbb{E}\|\Sigma_k^{-1/2} X_k\|_2 \leq \sqrt{\mathbb{E} X_k^* \Sigma_k^{-1} X_k} = \sqrt{\operatorname{tr} \Sigma_k^{-1} \mathbb{E} X_k X_k^*} \leq \sqrt{q}.$$

Observe that we have $\mathbb{E} X_k X_k^* = \sum_{t=1}^{k} M_t M_t^*$. As a result,

$$\left\|\Sigma_k^{-1/2} \begin{bmatrix} M_1 & \dots & M_k \end{bmatrix}\right\|_2 \leq 1$$

since by definition $\Sigma_k \succeq \mathbb{E} X_k X_k^*$. Hence, the function $\|\Sigma_k^{-1/2} X_k\|_2$ is $1-$Lipschitz with respect to $v_{t,i}$, for $t = 1, \dots, k$, $i = 1, \dots, d$. By concentration of Lipschitz functions of independent Gaussian variables [46, Th. 5.6]

$$P(\|\Sigma_k^{-1/2} X_k\|_2 > \sqrt{q} + t) \leq e^{-t^2/2}.$$

Now select $t = \sqrt{2 \log \frac{T}{\delta}}$ and take a union bound over k. □

D. Finite Sample Bounds for Fixed-Time, Fixed-Past

In this section, we include results for persistence of excitation and for identification of the system parameters for a fixed time instance k and a fixed past horizon p.

Theorem I.1 (Finite-sample identification): Fix a time k. Consider system (3) with observations $y_0, \dots, y_k$. Fix a past horizon p, a failure probability $\delta > 0$ and recall the notation of Table 1 and the universal constant C in Lemma I.5. Let

$$k_1(p, \delta) \triangleq Cpm \log(pm/\delta)$$

$$k_2(k, p, \delta) \triangleq \frac{512pn}{\min\{4, \sigma_R\}} \log\left(\frac{5p}{\delta} \frac{n \|\mathcal{O}_p\|_2^2 \|\Gamma_{k-p}\|_2 + \delta}{\delta}\right). \tag{34}$$

1) With probability at least $1 - 2\delta$, the following events $\mathcal{E}_{\bar{Z}} \cap \mathcal{E}_{\text{cross}}$ hold simultaneously:
   i) output upper bounds

$$\mathcal{E}_{\bar{Z}} \triangleq \left\{\bar{Z}_k \bar{Z}_k^* \preceq k \frac{mp}{\delta} \Gamma_{Z,k}\right\} \tag{35}$$

   ii) correlation upper bounds

$$\mathcal{E}_{\text{cross}} \triangleq \left\{\|S_k \bar{V}_k^{-1/2}\|_2^2 \leq \varepsilon(k, p, \delta)\right\}, \text{ where} \tag{36}$$

$$\varepsilon(k, p, \delta) \triangleq 16 \|\bar{R}\| mp \log \frac{5mpk\left(\|\Gamma_{Z,k}\|_2 \lambda^{-1} + 1\right)}{\delta}. \tag{37}$$

2) If k is large enough

$$k \geq k_3(k, p, \delta) \triangleq \max\{k_1(p, \delta), k_2(k, p, \delta)\} \tag{38}$$

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

then with probability at least $1-3\delta$, the following PE condition is true:

$$\mathcal{E}_{\mathrm{PE}} \triangleq \left\{ \begin{array}{l} \mathcal{T}_p \bar{E}_k \bar{E}_k^* \mathcal{T}_p^* \succeq \frac{k}{2}\Sigma_E \succeq \frac{k}{2}\sigma_R I, \\ \bar{Z}_k \bar{Z}_k^* \succeq \frac{1}{2}\mathcal{O}_p \bar{X}_k \bar{X}_k^* \mathcal{O}_p^* + \frac{1}{2}\mathcal{T}_p \bar{E}_k \bar{E}_k^* \mathcal{T}_p^* \end{array} \right\}. \quad (39)$$

Proof: Let $\bar{W}_t, \bar{V}_t$ be the Gram matrices:

$$\bar{W}_k \triangleq \bar{X}_k \bar{X}_k^* + W, \qquad W \triangleq \frac{k}{\|\mathcal{O}_p\|_2^2} I \quad (40)$$

$$\bar{V}_k \triangleq \bar{Z}_k \bar{Z}_k^* + V, \qquad V \triangleq \lambda I. \quad (41)$$

Define the base events:

$$\mathcal{E}_{\bar{X}} \triangleq \left\{ \bar{X}_k \bar{X}_k^* \preceq k\frac{n}{\delta}\Gamma_{k-p} \right\}, \ \mathcal{E}_E \triangleq \left\{ \mathcal{T}_p \bar{E}_k \bar{E}_k^* \mathcal{T}_p^* \succeq \frac{k}{2}\Sigma_E \right\}$$

$$\mathcal{E}_{XE} \triangleq \left\{ \left\| \bar{W}_k^{-1/2} \bar{X}_k \bar{E}_k^* \mathcal{T}_p^* \Sigma_E^{-1/2} \right\|_2^2 \right.$$
$$\left. \leq \ 8p \log \frac{p5^m \det(\bar{W}_k W^{-1})}{\delta} \right\}.$$

$$\mathcal{E}_{EZ} \triangleq \left\{ \left\| \bar{R}^{-1/2} S_k \bar{V}_k^{-1/2} \right\|_2^2 \leq 8 \log \frac{5^m \det(\bar{V}_k V^{-1})}{\delta} \right\}$$

We will show that $\mathcal{E}_{\bar{Z}}$ and all of the base events occur with probability at least $1-\delta$ each. Moreover

$$\mathcal{E}_{\mathrm{PE}} \supseteq \mathcal{E}_{\bar{X}} \cap \mathcal{E}_E \cap \mathcal{E}_{XE}, \ \mathcal{E}_{\mathrm{cross}} \supseteq \mathcal{E}_{\bar{Z}} \cap \mathcal{E}_{ZE}.$$

Hence, by two union bounds

$$\mathbb{P}(\mathcal{E}_{\mathrm{PE}}) \geq 1-3\delta, \ \mathbb{P}(\mathcal{E}_{\bar{Z}} \cap \mathcal{E}_{\mathrm{cross}}) \geq 1-2\delta.$$

1) *Base events:* The fact that $\mathbb{P}(\mathcal{E}_{\bar{X}}) \geq 1-\delta$, $\mathbb{P}(\mathcal{E}_{\bar{Z}}) \geq 1-\delta$ follows by a Markov inequality argument—see [35, Sec. 3]. The fact that $\mathbb{P}(\mathcal{E}_E) \geq 1-\delta$ follows from Lemma I.7. For each of the remaining events, we apply Theorem 3; note that $\bar{R}^{-1/2} e_k$ and $\Sigma_E^{-1/2}\mathcal{T}_p E_k$ are isotropic.
2) *Event $\mathcal{E}_{\mathrm{PE}}$:* From Lemma I.8, we have that $\mathcal{E}_{\mathrm{PE}} \supseteq \mathcal{E}_{\bar{X}} \cap \mathcal{E}_E \cap \mathcal{E}_{XE}$ if k satisfies (38).
3) *Event $\mathcal{E}_{cross}$:* We show that $\mathcal{E}_{\mathrm{cross}} \supseteq \mathcal{E}_{\bar{\mathrm{Z}}} \cap \mathcal{E}_{\mathrm{ZE}}$.

Conditioned on $\mathcal{E}_{\bar{\mathrm{Z}}} \cap \mathcal{E}_{\mathrm{ZE}}$, we have

$$\|S_k V_k^{-1/2}\|_2^2 \leq 8 \left\|\bar{R}\right\|_2 \left(\log \frac{5^m}{\delta} + \log\det \bar{V}_k V^{-1} \right)$$
$$\leq 8 \left\|\bar{R}\right\|_2 \left(\log \frac{5^m}{\delta} + mp \log\left(k\frac{mp}{\delta} \|\Gamma_{Z,k}\|_2 \lambda^{-1} + 1 \right) \right)$$

where the second inequality follows from $|\det L| \leq \|L\|_2^{mp}$ for any matrix $L \in \mathbb{R}^{mp \times mp}$. The final bound is simplified using $\frac{\delta}{mpk} < 1$ and $\log \frac{5^m}{\delta} \leq mp \log \frac{5mp}{\delta}$. □

Next, we prove the PE results that are required in the proof of the above theorem.

Lemma I.7 (Noise PE): Consider the conditions of Theorem I.1 and the definition of $k_1(p,\delta)$. If

$$k \geq k_1(p,\delta) \quad (42)$$

then with probability at least $1-\delta$

$$\frac{k}{2}\sigma_R I \preceq \frac{k}{2}\Sigma_E \preceq \mathcal{T}_p \bar{E}_k \bar{E}_k^* \mathcal{T}_p^* \preceq \frac{3k}{2}\Sigma_E.$$

Proof: Matrices $U_k \triangleq \Sigma_E^{-1/2}\mathcal{T}_p E_k$ satisfy the conditions of Lemma I.5. Hence, under (42), with probability at least $1-\delta$

$$\frac{k}{2} I \preceq \sum_{t=p}^{k} U_t U_t^* \preceq \frac{3k}{2} I.$$

Multiplying both sides with $\Sigma_E^{1/2}$ gives

$$\frac{k}{2}\Sigma_E \preceq \mathcal{T}_p \bar{E}_k \bar{E}_k^* \mathcal{T}_p^* \preceq \frac{3k}{2}\Sigma_E.$$

Finally, from [40, Lemma 2], we have $\Sigma_E \succeq \sigma_R I$. □

Next, we prove PE for the past outputs.

Lemma I.8 (Output PE): Consider the conditions of Theorem I.1 and the definition of $k_2(k,p,\delta)$, $\mathcal{E}_E$, $\mathcal{E}_{\bar{X}}$, and $\mathcal{E}_{XE}$. If

$$k \geq k_2(k,p,\delta) \quad (43)$$

then the following output PE condition holds:

$$\left\{ \bar{Z}_k \bar{Z}_k^* \succeq \frac{1}{2}\mathcal{O}_p \bar{X}_k \bar{X}_k^* \mathcal{O}_p + \frac{1}{2}\mathcal{T}_p \bar{E}_k \bar{E}_k \mathcal{T}_p \right\} \supseteq \mathcal{E}_E \cap \mathcal{E}_{\bar{X}} \cap \mathcal{E}_{XE}.$$

Proof: The batch past outputs can be written as

$$\bar{Z}_k = \mathcal{O}_p \bar{X}_k + \mathcal{T}_p \bar{E}_k.$$

As a result, the sample-covariance matrix $\bar{Z}_k \bar{Z}_k^*$ will be

$$O_p \bar{X}_k \bar{X}_k^* \mathcal{O}_p^* + \mathcal{T}_p \bar{E}_k \bar{E}_k^* \mathcal{T}_p^* + \mathcal{O}_p \bar{X}_k \bar{E}_k^* \mathcal{T}_p^* + \mathcal{T}_p \bar{E}_k \bar{X}_k^* \mathcal{O}_p^*.$$

The proof proceeds in two steps. First, we bound the cross-terms based on the events $\mathcal{E}_{\bar{X}}$ and $\mathcal{E}_{XE}$. Second, we show that if k is large enough, then the cross-terms are dominated by the autocorrelation terms:

$$\mathcal{O}_p \bar{X}_k \bar{E}_k^* \mathcal{T}_p^* + \mathcal{T}_p \bar{E}_k \bar{X}_k^* \mathcal{O}_p^* \preceq \frac{1}{2}\left(O_p \bar{X}_k \bar{X}_k^* \mathcal{O}_p^* + \mathcal{T}_p \bar{E}_k \bar{E}_k^* \mathcal{T}_p^* \right).$$

Cross-term bounds: For simplicity, we rewrite $\Sigma_E^{-1/2}\mathcal{T}_p \bar{E}_k = \bar{U}_k$, where $\bar{U}_k$ is defined similarly to $\bar{E}_k$ but has unit variance components. Conditioned on $\mathcal{E}_{\bar{X}}$

$$\log\det \bar{W}_k W^{-1} \leq n \log\left(\frac{n\|\mathcal{O}_p\|_2^2 \|\Gamma_{k-p}\|_2}{\delta} + 1 \right)$$

where we used the property $|\det L| \leq \|L\|_2^n$, for any matrix $L \in \mathbb{R}^n$. Conditioned also on $\mathcal{E}_{XE}$

$$\left\| \bar{W}_k^{-1/2} \bar{X}_k \bar{U}_k^* \right\|_2^2 \leq \mathcal{C}_{XE}^2$$

where we define

$$\mathcal{C}_{XE} \triangleq \sqrt{8p\left(\log \frac{p5^m}{\delta} + n \log\left(\frac{n\|\mathcal{O}_p\|_2^2 \|\Gamma_{k-p}\|_2}{\delta} + 1 \right) \right)}.$$

Let now $u \in \mathbb{R}^{mp}$, $\|u\|_2 = 1$ be an arbitrary unit vector. Then, consider the quadratic form

$$q(u) \triangleq \frac{1}{k}\left(u^* \mathcal{O}_p \bar{X}_k \bar{U}_k^* \Sigma_E^{1/2} u + u^* \Sigma_E^{1/2} \bar{U}_k \bar{X}_k^* \mathcal{O}_p^* u \right).$$

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

Conditioned on $\{\|\bar{W}_k^{-1/2}\bar{X}_k\bar{U}_k^*\|_2^2 \le \mathcal{C}_{XE}\} \cap \mathcal{E}_E \cap \mathcal{E}_X$ and using $I = \bar{W}_k\bar{W}_k^{-1}$, we can bound the quadratic form by

$$|q(u)| \le \frac{2}{k}\left\|u^*\mathcal{O}_p\bar{W}_k^{1/2}\right\|_2\left\|\bar{W}_k^{-1/2}\bar{X}_k\bar{U}_k^*\right\|_2\left\|\Sigma_E^{1/2}u\right\|_2$$
$$\le 2\sqrt{\frac{1}{k}u^*\mathcal{O}_p\bar{X}_k\bar{X}_k^*\mathcal{O}_p^*u + 1}\frac{\mathcal{C}_{XE}}{\sqrt{k}}\left\|\Sigma_E^{1/2}u\right\|_2.$$

Cross-terms are dominated: Define variables

$$a \triangleq \frac{1}{k}u^*\mathcal{O}_p\bar{X}_k\bar{X}_k^*\mathcal{O}_p^*u, \quad b \triangleq u^*\Sigma_E u.$$

To complete the proof, it is sufficient to show that for any unit vector u if (43) holds, then

$$2\sqrt{a+1}\frac{\mathcal{C}_{XE}}{\sqrt{k}}\sqrt{b} \le \frac{1}{2}\left(a + \frac{1}{k}u^*\Sigma_E^{1/2}\bar{U}_k\bar{U}_k^*\Sigma_E^{1/2}u\right).$$

But on $\mathcal{E}_E$, we have $\Sigma_E^{1/2}\bar{U}_k\bar{U}_k^*\Sigma_E^{1/2} \succeq k/2\Sigma_E$. Thus, it is sufficient to show

$$2\sqrt{a+1}\frac{\mathcal{C}_{XE}}{\sqrt{k}}\sqrt{b} \le \frac{a}{2} + \frac{b}{4}.$$

To guarantee the inequality, we apply the following Lemma I.9, where we exploit the fact that $b \ge \sigma_{\min}(\Sigma_E) \ge \sigma_R$. It follows that it is sufficient to have

$$\mathcal{C}_{XE/}\sqrt{k} \le \min\{2, \sqrt{\sigma_R}\}/8.$$

To obtain the final expression for $k_2(k,p,\delta)$, we use the simplification $8p\log(p5^m/\delta) \le 8pn\log(5p/\delta)$, since $m \le n$.

Lemma I.9: Let $a \ge 0$ and $b \ge \sigma_R > 0$. Then, if

$$\gamma \le \frac{\min\{2, \sqrt{\sigma_R}\}}{8}$$

then $f(a,b) \triangleq \frac{a}{2} + \frac{b}{4} - 2\sqrt{a+1}\sqrt{b}\gamma \ge 0$.

Proof: By minimizing over a, we obtain

$$\min_{0\le a} f(a,b) = \left\{\begin{array}{ll} b/4 - 2\sqrt{b}\gamma, & \text{if } 2\gamma\sqrt{b} \le 1 \\ b\left(1/4 - 2\gamma^2\right) - 1/2, & \text{if } 2\gamma\sqrt{b} > 1 \end{array}\right\}.$$

If $2\gamma\sqrt{b} \le 1$, then we have

$$\min_{0\le a,\sigma_R\le b} f(a,b) = \frac{b}{4} - 2\sqrt{b}\gamma \ge \frac{b - \sqrt{b\sigma_R}}{4} \ge 0.$$

Since $\gamma \le 1/4$, the case $2\gamma\sqrt{b} > 1$ can occur only if $b > 4$. But then, for $b > 4$

$$b\left(\frac{1}{4} - 2\gamma^2\right) - \frac{1}{2} \ge b\left(\frac{1}{4} - \frac{1}{8}\right) - \frac{1}{2} = \frac{b-4}{8} \ge 0.$$

□

E. Uniform Bounds

The result of Theorem I.1 applies only for a fixed time step k within an epoch. Here, we directly extend the results to hold uniformly for all times steps k within an epoch. Note that we do not require PE.

Lemma I.10 (Uniform bounds): Consider the conditions of Theorem 1. Fix a failure probability $\delta > 0$. Let $T = 2^{i-1}T_{\text{init}}$ for some fixed epoch i with $p = \beta\log T$ the corresponding past horizon. Consider the definition of $\varepsilon(k,p,\delta)$ in (37). With probability at least $1 - 2\sum_{k=T}^{2T-1}\frac{1}{k^2}\delta$, the following events hold:

$$\mathcal{E}_{\text{unif}} \triangleq \bigcup_{k=T}^{2T-1}\left\{\begin{array}{ll} \bar{Z}_k\bar{Z}_k^* & \preceq \frac{k^3mp}{\delta}\Gamma_{Z,k} \\ \left\|S_k\bar{V}_k^{-1/2}\right\|_2 & \le \varepsilon(k,p,\delta/k^2) \end{array}\right\}. \quad (44)$$

Proof: Fix a k such that $T \le k \le 2T-1$ apply Theorem I.1 for δ replaced with δ/k^2. The results follows from (35) and (36), by taking the union bound over all $T \le k \le 2T-1$. □

F. Proof of Lemma 1

Recall that $\bar{V}_{k-1} = \bar{V}_k - Z_kZ_k^*$. Using the identity $\det(I + FB) = \det(I + BF)$, we obtain

$$\det\bar{V}_{k-1} = \det\bar{V}_k\det\left(I - \bar{V}_k^{-1/2}Z_kZ_k^*\bar{V}_k^{-1/2}\right)$$
$$= \det\bar{V}_k\left(1 - Z_k^*\bar{V}_k^{-1}Z_k\right).$$

Rearranging the terms gives

$$Z_k^*\bar{V}_k^{-1}Z_k = 1 - \frac{\det\bar{V}_{k-1}}{\det\bar{V}_k} \le \log\det\bar{V}_k - \log\det\bar{V}_{k-1}$$

where the inequality follows from the fact that the sequence $\bar{V}_k \succeq \bar{V}_{k-1}$ is increasing and the elementary inequality:

$$1 - x \le \log 1/x, \text{ for } x \le 1.$$

Since the upper bound telescopes, we finally get

$$\sum_{k=T}^{2T-1} Z_k^*\bar{V}_k^{-1}Z_k \le \log\det\bar{V}_{2T-1} - \log\det\bar{V}_{T-1}.$$

□

G. Analysis Within One Epoch

We will analyze the square loss for the duration of one epoch, from time T up to time $2T-1$ with fixed past horizon $p = \beta\log T$. Note that PE is not required.

Consider the cumulative square loss within the epoch:

$$\mathcal{L}_T^{2T-1} \triangleq \sum_{k=T}^{2T-1}\|\hat{y}_k - \tilde{y}_k\|_2^2. \quad (45)$$

Based on the notation of Table 1, the error between the Kalman filter prediction and our online algorithm is

$$\tilde{y}_k - \hat{y}_k = S_{k-1}\bar{V}_{k-1}^{-1}Z_k + \lambda G\bar{V}_{k-1}^{-1}Z_k$$
$$+ C(A-KC)^p\left(\bar{X}_{k-1}\bar{Z}_{k-1}\bar{V}_{k-1}^{-1}Z_k - \hat{x}_{k-p}\right).$$

By Cauchy–Schwarz, the submultiplicative property of norm, and the fact that $\|\bar{Z}_{k-1}\bar{V}_{k-1}^{-1/2}\|_2 \le 1$ is normalized

$$\|\tilde{y}_k - \hat{y}_k\|_2^2 \le 4\Big(\|S_{k-1}\bar{V}_{k-1}^{-1/2}\|_2^2 + \|\lambda G\bar{V}_k^{-1/2}\|_2^2$$
$$+ \|C(A-KC)^p\|_2^2\|\bar{X}_{k-1}\|_2^2\Big)\|\bar{V}_{k-1}^{-1/2}\bar{V}_k^{1/2}\|_2^2\|\bar{V}_k^{-1/2}Z_k\|_2^2$$
$$+ 4\|C(A-KC)^p\|_2^2\|\hat{x}_{k-p}\|_2^2. \quad (46)$$

To obtain a bound on the square loss, it is sufficient to bound the following three terms:

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

1) the supremum over $T \le t \le 2T-1$ of:

$$\|S_{t-1}\bar{V}_{t-1}^{-1/2}\|_2^2 + \|\lambda G \bar{V}_t^{-1/2}\|_2^2 + \|C(A-KC)^p\|_2^2 \left\|\bar{X}_{t-1}\right\|_2^2$$

2) the sum $\|C(A-KC)^p\|_2^2 \sum_{k=T}^{2T-1} \|\hat{x}_{k-p}\|_2^2$;
3) the supremum over $T \le t \le 2T-1$ of

$$\|\bar{V}_{t-1}^{-1/2}\bar{V}_t^{1/2}\|_2^2 = 1 + \|\bar{V}_{t-1}^{-1/2} Z_t\|_2^2$$

where the equality follows from

$$\bar{V}_{t-1}^{-1/2}\bar{V}_t\bar{V}_{t-1}^{-1/2} = I + \bar{V}_{t-1}^{-1/2} Z_t Z_t^* \bar{V}_{t-1}^{-1/2};$$

4) the sum $\sum_{k=T}^{2T-1} \|\bar{V}_k^{-1/2} Z_k\|_2^2$.

Theorem I.2 (Square loss within epoch): Consider the conditions of Theorem 1. Let a be the minimal polynomial of A with degree d, Δ defined as in (31). Fix a failure probability $\delta > 0$. Let $T = 2^{i-1}T_{\text{init}}$ for some fixed epoch $i \ge 1$ with $p = \beta \log T$ the corresponding past horizon. Recall the definition of $\mathcal{C}_{\text{diff}}$ in Theorem 1. With probability at least $1 - 4\sum_{k=T}^{2T-1} \frac{1}{k^2}\delta$

$$\mathcal{L}_T^{2T-1} \le \text{poly}(\mathcal{C}_{\text{diff}}) \left(O(\log^5 T) + \tilde{O}(\rho(A-KC)^p T^{2\kappa}) \right).$$

Proof: Consider the uniform event $\mathcal{E}_{\text{unif}}$ defined in (44) and define the events:

$$\mathcal{E}_x \triangleq \left\{ \sup_{k \le 2T-1} \|\Gamma_k^{-1/2}\hat{x}_k\|_2 \le \sqrt{n} + \sqrt{2\log\frac{4T}{\delta_1}} \right\}$$

$$\mathcal{E}_e = \left\{ \sup_{k \le 2T-1} \|\bar{R}^{-1/2} e_k\|_2 \le \sqrt{m} + \sqrt{2\log\frac{2T}{\delta_1}} \right\}$$

where $\delta_1 = \sum_{k=T}^{2T-1} \delta/k^2$. Based on Lemma I.10, Lemma I.6, and a union bound, the events $\mathcal{E}_{\text{unif}} \cap \mathcal{E}_x \cap \mathcal{E}_e$ occur with probability at least $1 - 4\sum_{k=T}^{2T-1} \frac{1}{k^2}\delta$. We will bound all terms of the square loss based on the above events.

1) a) For the term $S_{k-1}\bar{V}_{k-1}^{-1/2}$, based on event $\mathcal{E}_{\text{unif}}$

$$\begin{aligned}\|S_{k-1}\bar{V}_{k-1}^{-1/2}\|_2^2 &\le \varepsilon(2T, p, \delta/(2T)^2) \\ &= \text{poly}(\mathcal{C}_{\text{diff}}) O(\log^2 T)\end{aligned}$$

where the exponent of 2 is due to the fact that the dominant term is $p \log \|\Gamma_Z, k\|_2$, with $\log \|\Gamma_Z, k\|_2 = O(\log T)$ in the general case of nonexplosive systems.

b) Regularization term: $\|\lambda G \bar{V}_t^{-1/2}\|_2^2 \le \lambda \|G\|_2^2$.

c) To bound the term $\|C(A-KC)^p\|_2^2 \|\bar{X}_{t-1}\|_2^2$, we use the inequality $\|\bar{X}_{t-1}\|_2^2 \le 2T \sup_{k \le 2T-1} \|\hat{x}_k\|_2^2$. Hence, it is sufficient to upper-bound the norm of the states. Since the covariances Γ_k are increasing

$$\sup_{k \le 2T-1} \|\hat{x}_k\|_2^2 \le \|\Gamma_{2T-1}\|_2 \sup_{k \le 2T-1} \left\|\Gamma_k^{-1/2}\hat{x}_k\right\|_2^2.$$

Hence, conditioned on $\mathcal{E}_x$ and since $\|\Gamma_{2T-1}\|_2 = O(T^{2\kappa-1})$

$$\begin{aligned}&\|C(A-KC)^p\|_2^2 \left\|\bar{X}_{t-1}\right\|_2^2 \\ &\quad \le \text{poly}(\mathcal{C}_{\text{diff}}) \tilde{O}(\rho(A-KC)^p T^{2\kappa}).\end{aligned}$$

2) To bound the sum $\|C(A-KC)^p\|_2^2 \sum_{k=T}^{2T-1} \|\hat{x}_{k-p}\|_2^2$, we use the exact same steps as above since

$$\sum_{k=T}^{2T-1} \|\hat{x}_{k-p}\|_2^2 \le T \sup_{k \le 2T-1} \|\hat{x}_k\|_2^2.$$

3) To bound the norm of $\|\bar{V}_{k-1}^{-1/2} Z_k\|_2^2$, we exploit the ARMA-like representation in Lemma 2. Replacing $Z_k = a_{d-1}Z_{k-1} + \cdots + a_0 Z_{k-d} + \delta_k$ by two applications of Cauchy–Schwarz

$$\begin{aligned}\|\bar{V}_{k-1}^{-1/2} Z_k\|_2^2 &\le 2\|a\|_2^2 \sum_{i=1}^{d} Z_{k-i}^* \bar{V}_{k-1}^{-1} Z_{k-i} \\ &\quad + 2\|\bar{V}_{k-1}^{-1/2}\delta_k\|_2^2 \\ &\le 2\|a\|_2^2 d + 2\lambda^{-1}\Delta^2 \sup_{i \le k-1} \|e_i\|_2^2 \\ &= \text{poly}(\mathcal{C}_{\text{diff}}) O(\log T)\end{aligned}$$

where the bound on $\|e_i\|_2^2$ follows from $\mathcal{E}_e$.

4) Finally, by Lemma 1, we have

$$\sum_{k=T}^{2T-1} \left\|\bar{V}_k^{-1/2} Z_k\right\|_2^2 \le \log\det(\bar{V}_{2T}\lambda^{-1}).$$

Based on $\mathcal{E}_{\text{unif}}$ we obtain

$$\log\det(\bar{V}_{2T}\lambda^{-1}) = \text{poly}(\mathcal{C}_{\text{diff}}) O(\log^2 T).$$

H. Proof of Theorem 1

Recall that the regret can be decomposed into two terms

$$\mathcal{R}_N = \mathcal{L}_N + 2\sum_{k=T_{\text{init}}}^{N} e_k^* (\hat{y}_k - \tilde{y}_k)$$

where $\mathcal{L}_N$ is the square loss and the other term is a martingale.

Square loss bound: Without loss of generality assume that $N = 2T_i - 1 = T_{\text{init}} 2^i$ is the end of an epoch, where i is the total number of epochs. The cumulative square loss can be rewritten as $\mathcal{L}_N = \sum_{j=1}^{i-1} \mathcal{L}_{T_j}^{2T_j-1}$. Next, we select

$$\beta \ge 4\frac{\kappa}{\log 1/\rho(A-KC)} \tag{47}$$

which under Assumption 4 makes the bias term in Theorem I.2 logarithmic. By Theorem I.2 and a union bound over all epochs, with probability at least $1 - 7\frac{\pi^2}{6}\delta$

$$\mathcal{L}_N = \text{poly}(\mathcal{C}_{\text{diff}}) O(\log^6 N).$$

Martingale term bound: Denote $u_k \triangleq \bar{R}^{-1/2} e_k$ and $z_k \triangleq R^{1/2}(\hat{y}_k - \tilde{y}_k)$. Then, $\sum_{t=1}^{N} e_t^*(\hat{y}_t - \tilde{y}_t) = \sum_{t=1}^{N} u_t^* z_t = \sum_{t=1}^{N}\sum_{i=1}^{m} u_{t,i} z_{t,i}$. To apply Theorem 3, we need to slightly modify the definition of the filtration. Let $\mathcal{F}_{t,i} \triangleq \sigma(\mathcal{F}_t \cup \{u_{t+1,1}, \ldots, u_{t+1,i}\})$, with $\mathcal{F}_{t+1} \equiv \mathcal{F}_{t,m}$ and define

$$\tilde{\mathcal{F}}_0 = \mathcal{F}_0 \tag{48}$$

$$\tilde{\mathcal{F}}_s = \mathcal{F}_{t,s-tm}, \text{ if } tm+1 \le s \le (t+1)m. \tag{49}$$

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

By applying Theorem 3 with $\tilde{\mathcal{F}}_s$, we can bound the sum in terms of the square loss $\mathcal{L}_N$. With probability at least $1-\delta$

$$\left(\sum_{t=1}^{N} z_t^* z_t + 1\right)^{-1/2} \sum_{t=1}^{N} u_t^* z_t \leq 8 \log\left(\frac{5}{\delta}\left(\sum_{t=1}^{N} z_t^* z_t + 1\right)\right)$$

or using the fact that $z_k^* z_k \leq \|\bar{R}\|_2 \|\hat{y}_k - \tilde{y}_k\|_2^2$

$$\sum_{t=1}^{N} u_t^* z_t \leq \left(\left\|\bar{R}\right\|_2 \mathcal{L}_N + 1\right)^{1/2} 8 \log\left(\frac{5}{\delta}\left(\left\|\bar{R}\right\|_2 \mathcal{L}_N + 1\right)\right).$$

By a union bound, with probability at least $1 - (7\frac{\pi^2}{6} + 1)\delta$

$$\mathcal{R}_N = \text{poly}(\mathcal{C}_{\text{diff}}) O(\log^6 N).$$

To complete the proof, we rescale δ and we rewrite the bounds with respect $\tilde{\delta} = (7\frac{\pi^2}{6} + 1)^{-1}\delta$. □

APPENDIX II
STABLE CASE

In the case of stable systems, stationarity allows us to prove stronger PE results.

Lemma II.1 (Stable: Output PE): Consider system (3) with observations $y_0, \dots, y_k$. Let $\tau = \tau_{\text{mix}} + p$, where τ_{mix} is the mixing time defined in (14). Recall the universal constant C in Lemma I.5 and the notation of Table 1. Define

$$k_4(p,\delta) \triangleq C\tau m \log(\tau m/\delta)$$

$$k_5(k,p,\delta) \triangleq \frac{512pn}{\min\{4,\sigma_R\}} \log\left(\frac{5p}{\delta}\frac{n\left\|\mathcal{O}_\tau\right\|_2^2 \left\|\Gamma_{k-\tau}\right\|_2}{\delta} + 1\right). \tag{50}$$

With probability at least $1-3\delta$, if

$$k \geq k_6(k,p,\delta) \triangleq \max\{k_4(p,\delta), k_5(k,p,\delta)\} \tag{51}$$

then the following output PE condition holds:

$$\bar{Z}_k \bar{Z}_k^* \succeq \frac{k}{4}\Gamma_{Z,\tau} \succeq \frac{k}{8}\Gamma_{Z,\infty}. \tag{52}$$

Proof: Define the controllability matrix

$$\mathcal{C}_t \triangleq \begin{bmatrix} A^{t-1}K & \dots & AK & K \end{bmatrix}, t \geq 1.$$

The state covariance at any time can be conveniently expressed in terms of the controllability matrix

$$\Gamma_t = \mathcal{C}_t \,\text{diag}(\bar{R}, \dots, \bar{R}) \mathcal{C}_t^*.$$

Combining the above equality with (28), we obtain that

$$\Gamma_{Z,\tau} = \begin{bmatrix} O_p \mathcal{C}_{\tau_{\text{mix}}} & \mathcal{T}_p \end{bmatrix} \text{diag}(\bar{R}, \dots, \bar{R}) \begin{bmatrix} O_p \mathcal{C}_{\tau_{\text{mix}}} & \mathcal{T}_p \end{bmatrix}^*.$$

Hence, by the definition of mixing time

$$\Gamma_{Z,\tau} \succeq \Gamma_{Z,\infty}/2. \tag{53}$$

What remains is to show the first inequality in (52).

The proof is similar to the one of Lemma I.8. We only need an additional step, to further unroll $\hat{x}_{x-p}$ for τ_{mix} time steps into the past in (25). Extend the definition of the past noises

$$E_t^\tau \triangleq \begin{bmatrix} e_{t-\tau}^* & \cdots & e_{t-1}^* \end{bmatrix}^*.$$

Rolling out the state equations in (25), we obtain

$$Z_t = \mathcal{O}_p A^{\tau_{\text{mix}}} \hat{x}_{t-\tau} + \begin{bmatrix} O_p \mathcal{C}_{\tau_{\text{mix}}} & \mathcal{T}_p \end{bmatrix} E_t^\tau. \tag{54}$$

Define the isotropic variables

$$U_t \triangleq \Gamma_{Z,\tau}^{-1/2} \begin{bmatrix} O_p \mathcal{C}_{\tau_{\text{mix}}} & \mathcal{T}_p \end{bmatrix} E_t^\tau$$

which are well defined since $\Gamma_{Z,\tau} \succeq \Sigma_E \succeq \sigma_R I$. This enables us to rewrite (54) as

$$Z_t = \mathcal{O}_p A^{\tau_{\text{mix}}} \hat{x}_{t-\tau} + \Gamma_{Z,\tau}^{1/2} U_t. \tag{55}$$

Expanding the correlations gives

$$Z_t Z_t^* = \mathcal{O}_p A^{\tau_{\text{mix}}} \hat{x}_{t-\tau} \hat{x}_{t-\tau}^* (\mathcal{O}_p A^{\tau_{\text{mix}}})^* + \Gamma_{Z,\tau}^{1/2} U_t U_t^* \Gamma_{Z,\tau}^{1/2} + \mathcal{O}_p A^{\tau_{\text{mix}}} \hat{x}_{t-\tau} U_t^* \Gamma_{Z,\tau}^{1/2} + \Gamma_{Z,\tau}^{1/2} U_t \hat{x}_{t-\tau}^* (\mathcal{O}_p A^{\tau_{\text{mix}}})^*.$$

The remaining proof is now identical to Lemma I.8 and is, thus, omitted. To simplify the notation for the expression of k_5, we used $\|\mathcal{O}_p A^{\tau_{\text{mix}}}\|_2 \leq \|\mathcal{O}_\tau\|_2$ ($\mathcal{O}_p A^{\tau_{\text{mix}}}$ is a submatrix of $\mathcal{O}_\tau$). We also used $\sigma_{\min}(\Gamma_{Z,\infty}) \geq \sigma_{\min}(\Sigma_E) \geq \sigma_R$ in the step where we apply the technical Lemma I.9.

A. Proof of Lemma 3

We prove a more general version.

Lemma II.2 (Stable case: Uniform PAC bounds): Consider the conditions of Theorem 1 with $\rho(A) < 1$. Select a failure probability $\delta > 0$. Let $T = 2^{i-1} T_{\text{init}}$ for some fixed epoch i with $p = \beta \log T$ the corresponding past horizon. Consider also the definition of $\varepsilon(k,p,\delta)$ in (37). There exists an $N_0^{\text{s}} = \text{poly}(n, \beta, \log 1/\delta, \tau_{\text{mix}}, 1/\sigma_R)$ such that with probability at least $1 - 5\sum_{k=T}^{2T-1} \frac{1}{k^2}\delta$ the following events hold:

$$\mathcal{E}_{\text{unif}} \triangleq \bigcap_{k=T}^{2T-1} \left\{ \begin{array}{cc} \bar{Z}_k \bar{Z}_k^* & \preceq \frac{k^3 mp}{\delta}\Gamma_{Z,k} \\ \left\|S_k \bar{V}_k^{-1/2}\right\|_2 & \leq \varepsilon(k,p,\delta/k^2) \end{array} \right\} \tag{56}$$

$$\mathcal{E}_{\text{unif}}^{\text{PE}} \triangleq \bigcap_{k=\max\{N_0^{\text{s}}, T\}}^{2T-1} \left\{ \bar{Z}_k \bar{Z}_k^* \succeq \frac{k}{8}\Gamma_{Z,\infty} \right\}. \tag{57}$$

Proof: Recall $k_6(k,p,\delta)$ defined in (51) and let

$$N_0^{\text{s}} \triangleq \min\left\{t : \; k \geq k_6(k,p,\delta/k^2), \text{ for all } k \geq t\right\}. \tag{58}$$

Now, fix a k such that $T \leq k \leq 2T-1$ apply Theorem I.1 and Lemma II.1 for δ replaced with δ/k^2. Taking the union bound over all $T \leq k \leq 2T-1$, from (52), (35), and (36), we obtain that

$$\mathbb{P}(\mathcal{E}_{\text{unif}} \cap \mathcal{E}_{\text{unif}}^{\text{PE}}) \geq 1 - 5\sum_{k=T}^{2T-1} \frac{1}{k^2}\delta.$$

What remains to show is that N_0^{s} depends polynomially on $\beta, n, \log 1/\delta, 1/\sigma_R, \tau_{\text{mix}}$. By Lemma I.1, the covariance matrix Γ_k increases at most as fast as $M^4 k^{2\kappa-1}$ (a similar result holds for the observability matrix). Hence, the dominant term in k_5 is of the order of

$$k_5(k, \beta \log k, \delta/k^2) \leq \beta n \kappa \frac{M}{\sigma_R} \log(1/\delta)\, O(\log^2 k).$$

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

The dominant term in k_4 is

$$k_4(\beta \log k, \delta/k^2) \leq (\tau_{\text{mix}} + \beta \log k) m \log \frac{k}{\delta}.$$

To simplify the final expression, define

$$N^{\text{s}}_{\text{diff}} = \tau_{\text{mix}} \beta n \kappa \tfrac{M}{\sigma_R} \log(1/\delta).$$

By Lemmas III.1 and III.2, it follows that N^{s}_0 is of the order of

$$N^{\text{s}}_0 \leq 4C \log^2 4\, C N^{\text{s}}_{\text{diff}} \log^2 \; N^{\text{s}}_{\text{diff}}$$

where C is the universal constant defined in Lemma I.5. □

B. Proof of Theorem 2

Similar to the nonexplosive case, we analyze the square loss for a single epoch. We additionally exploit the sharper PE condition.

Theorem II.1 (Square loss within epoch): Consider the conditions of Theorem 2 and the definition of $\mathcal{C}^{\text{s}}_{\text{diff}}$ therein. Fix a failure probabilities $\delta > 0$ and consider N^{s}_0 defined as in (58). Let $T = 2^{i-1} T_{\text{init}}$ for some fixed epoch $i \geq 1$ with $p = \beta \log T$ the corresponding past horizon. Let β satisfy (12). Then, with probability at least $1 - 7\sum_{k=T}^{2T-1} \frac{1}{k^2}\delta$

$$\mathcal{L}_T^{2T-1} \leq \text{poly}(\mathcal{C}_{\text{diff}}) O(\log^5 \, T).$$

If moreover $T \geq N^{\text{s}}_0$, then

$$\mathcal{L}_T^{2T-1} \leq \text{poly}(\mathcal{C}^{\text{s}}_{\text{diff}}) O(\log^3 \, T).$$

Proof: For $T < N^{\text{s}}_0$, the result follows from Theorem I.2. Let $T \geq N^{\text{s}}$ and consider the uniform events $\mathcal{E}_{\text{unif}}$ and $\mathcal{E}^{\text{PE}}_{\text{unif}}$ defined in (56) and (II.8). Define also

$$\mathcal{E}_x = \left\{ \sup_{k \leq 2T-1} \|\Gamma_k^{-1/2} \hat{x}_k\|_2 \leq \sqrt{n} + \sqrt{2 \log \frac{4T}{\delta_1}} \right\}$$

$$\mathcal{E}_z = \left\{ \sup_{k \leq 2T-1} \|\Gamma_{Z,k}^{-1/2} Z_k\|_2 \leq \sqrt{pm} + \sqrt{2 \log \frac{2T}{\delta_1}} \right\}$$

where $\delta_1 = \sum_{k=T}^{2T-1} \delta/k^2$. Based on Lemmas II.2 and I.6, and a union bound, all events $\mathcal{E}_{\text{unif}} \cap \mathcal{E}^{\text{PE,s}}_{\text{unif}} \cap \mathcal{E}_x \cap \mathcal{E}_z$ occur with probability at least $1 - 7\sum_{k=T}^{2T-1} \frac{1}{k^2}\delta$. Now, we proceed as in the proof of Theorem I.2. Instead of separately bounding the quantities in steps 3) and 4) in the proof of Theorem I.2, here we provide a combined bound on the sum of $\|\bar{V}_{k-1}^{-1/2} Z_k\|_2^2$.

The sum $\|\bar{V}_{k-1}^{-1/2} Z_k\|_2^2$ is upper bounded by

$$\left(\sum_{k=T}^{2T-1} \left\| \bar{V}_{k-1}^{-1/2} \Gamma_{Z,k}^{1/2} \right\|_2^2 \right) \sup_{k \leq 2T-1} \left\| \Gamma_{Z,k}^{-1/2} Z_k \right\|_2^2.$$

However, since $T \geq N^{\text{s}}_0$, based on the event $\mathcal{E}^{\text{PE}}_{\text{unif}}$

$$\sum_{k=T}^{2T-1} \left\| \bar{V}_{k-1}^{-1/2} \Gamma_{Z,k}^{1/2} \right\|_2^2 \leq 8 \frac{2T}{T-1}.$$

Finally, based on $\mathcal{E}_z$, we have $\|\Gamma_{Z,k}^{-1/2} Z_k\|_2^2 = O(\log T)$.

The remaining steps are the same as in the proof of Theorem 1.

APPENDIX III
TECHNICAL LEMMAS

Lemma III.1: Let $c > 0$. The inequality

$$k \geq c \log k$$

is true if $k \geq \max\{2c \log 2c, 1\}$.

Proof: If $c \leq e$, then the inequality is satisfied for all $k > 0$. To see why this holds, consider $f(k) = k - e \log k$. The minimum is attained at $f(e) = e - e \log e = 0$. Hence, $k \geq e \log k \geq c \log k$. If $c > e$, then the function $k - c \log k$ is increasing for $k \geq c$. Moreover, $2c \log 2c \geq c$. As a result, if $k \geq 2c \log 2c$, then also

$$k - c \log k \geq 2c \log 2c - c \log(2c \log 2c)$$
$$\geq \left(c - \frac{c}{e} \right) \log 2c \geq 0$$

where we used Lemma III.3.

Lemma III.2: Let $c \geq 0$. The inequality

$$k \geq c \log^2 k$$

is true if $k \geq \max\{4c \log^2 4c, 4c \log 4c, 1\}$.

Proof: If $c \leq 1$, then the inequality is satisfied for $k \geq 1$. To see why this holds, define $f(k) = k - \log^2 k$. Its derivative $f'(k) = 1 - 2\frac{\log k}{k}$ is always positive for $k \geq 1$ since from the proof of Lemma III.1 $k \geq 2 \log k$. Hence, $f(k) \geq f(1) = 1$.

Consider now the case $c > 1$ and define $g(k) = k - c \log^2 k$. Its derivative is $g'(k) = 1 - 2c\frac{\log k}{k}$. From Lemma III.1, $g'(k) \geq 0$, for $k \geq \max\{4c \log 4c, 1\}$. Now, pick $k_1 = 4c \log^2 4c$ and observe that $k_1 \geq 4c \log 4c$ since $4c > e$ and $\log 4c > 1$. Since g is increasing for $k \geq k_1$, it is sufficient to prove that $g(k_1) > 0$. Invoking Lemma III.3, we obtain the inequality

$$c \log^2(k_1) \leq c \left(\log 4c + \frac{1}{e} \log 4c \right)^2 \leq 4c \log^2 4c = k_1$$

where (i) follows from Lemma III.3.

Lemma III.3: Let $c \geq e$, then the following inequality holds:

$$\log \log c \leq \frac{1}{e} \log c.$$

Proof: Consider function $f(c) = \frac{1}{e} \log c - \log \log c$ and compute the derivative

$$f'(c) = \frac{1}{ec} - \frac{1}{c \log c}.$$

The minimum is attained at e^e. Hence,

$$f(c) \geq f(e^e) = 0, \;\; \text{for all } c \geq e.$$

ACKNOWLEDGMENT

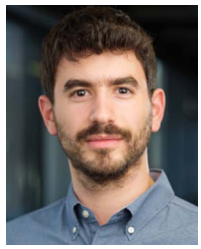
The authors would like to thank Nikolai Matni for useful discussions.

REFERENCES

[1] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 4th ed. vol. 1. Belmont, MA, USA: Athena Scientific, 2017.
[2] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robot. Automat. Mag.*, vol. 13, no. 2, pp. 99–110, Jun. 2006.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.

[3] H. Coskun, F. Achilles, R. DiPietro, N. Navab, and F. Tombari, "Long short-term memory Kalman filters: Recurrent neural estimators for pose regularization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 5524–5532.
[4] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge, U.K.: Cambridge Univ. Press, 1990.
[5] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME–J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960.
[6] C. Yu, L. Ljung, and M. Verhaegen, "Identification of structured state-space models," *Automatica*, vol. 90, pp. 54–61, 2018.
[7] L. Ljung, "Convergence of an adaptive filter algorithm," *Int. J. Control*, vol. 27, no. 5, pp. 673–693, 1978.
[8] J. Moore and G. Ledwich, "Multivariable adaptive parameter and state estimators with convergence analysis," *ANZIAM J.*, vol. 21, no. 2, pp. 176–197, 1979.
[9] T. L. Lai and Z. Ying, "Recursive identification and adaptive prediction in linear stochastic systems," *SIAM J. Control Optim.*, vol. 29, no. 5, pp. 1061–1090, 1991.
[10] F. Ding, Y. Shi, and T. Chen, "Performance analysis of estimation algorithms of nonstationary ARMA processes," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1041–1053, Mar. 2006.
[11] R. Vershynin, *High-Dimensional Probability: An Introduction With Applications in Data Science*. vol. 47. Cambridge, U.K.: Cambridge Univ. Press, 2018.
[12] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. Cambridge, U.K.: Cambridge Univ. Press, 2006.
[13] M. Kozdoba, J. Marecek, T. Tchrakian, and S. Mannor, "On-line learning of linear dynamical systems: Exponential forgetting in Kalman filters," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 4098–4105.
[14] S. J. Qin, "An overview of subspace identification," *Comput. Chem. Eng.*, vol. 30, no. 10–12, pp. 1502–1513, 2006.
[15] M. Simchowitz, R. Boczar, and B. Recht, "Learning linear dynamical systems with semi-parametric least squares," in *Proc. Conf. Learn. Theory*, 2019.
[16] M. Simchowitz and D. Foster, "Naive exploration is optimal for online LQR," in *Proc. Int. Conf. Mach. Learn.*, 2020.
[17] I. Ziemann and H. Sandberg, "On a phase transition of regret in linear quadratic control: The memoryless case," *IEEE Contr. Syst. Lett.*, vol. 5, no. 2, pp. 695–700, Apr. 2021.
[18] Y. Abbasi-Yadkori and C. Szepesvári, "Regret bounds for the adaptive control of linear quadratic systems," in *Proc. 24th Annu. Conf. Learn. Theory*, 2011, pp. 1–26.
[19] M. K. S. Faradonbeh, A. Tewari, and G. Michailidis, "Optimism-based adaptive regulation of linear-quadratic systems," *IEEE Trans. Autom. Control*, vol. 66, no. 4, pp. 1802–1808, 2020.
[20] Y. Ouyang, M. Gagrani, and R. Jain, "Learning-based control of unknown linear systems with Thompson sampling," 2017, *arXiv:1709.04047*.
[21] M. Abeille and A. Lazaric, "Improved regret bounds for Thompson sampling in linear quadratic control problems," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1–9.
[22] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "Regret bounds for robust adaptive control of the linear quadratic regulator," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 4188–4197.
[23] H. Mania, S. Tu, and B. Recht, "Certainty equivalent control of LQR is efficient," *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.
[24] A. Cohen, T. Koren, and Y. Mansour, "Learning linear-quadratic regulators efficiently with only $\sqrt{T}$ regret," in *Proc. Int. Conf. Mach. Learn.*, 2019.
[25] N. Matni, A. Proutiere, A. Rantzer, and S. Tu, "From self-tuning regulators to reinforcement learning and back again," in *Proc. IEEE 58th Conf. Decis. Control*, 2019, pp. 3724–3740.
[26] A. Tsiamis, N. Matni, and G. J. Pappas, "Sample complexity of Kalman filtering for unknown systems," *Learn. Dyn. Control*, pp. 435–444, 2020.
[27] O. Anava, E. Hazan, S. Mannor, and O. Shamir, "Online learning for time series prediction," in *Proc. Conf. Learn. Theory*, 2013, pp. 172–184.
[28] E. Hazan, H. Lee, K. Singh, C. Zhang, and Y. Zhang, "Spectral filtering for general linear dynamical systems," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 4634–4643.
[29] U. Ghai, H. Lee, K. Singh, C. Zhang, and Y. Zhang, "No-regret prediction in marginally stable systems," in *Proc. Conf. Learn. Theory*., 2020, pp. 1714–1757.
[30] P. Rashidinejad, J. Jiao, and S. Russell, "Slip: Learning to predict in unknown dynamical systems with long-term memory," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 5716–5728, 2020.
[31] S. Lale, K. Azizzadenesheli, B. Hassibi, and A. Anandkumar, "Logarithmic regret bound in partially observable linear dynamical systems," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 20876–20888, 2020.
[32] D. Bauer, M. Deistler, and W. Scherrer, "Consistency and asymptotic normality of some subspace algorithms for systems without observed inputs," *Automatica*, vol. 35, no. 7, pp. 1243–1254, 1999.
[33] T. L. Lai and C. Z. Wei, "Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems," *Ann. Statist.*, vol. 10, no. 1, pp. 154–166, 1982.
[34] M. K. S. Faradonbeh, A. Tewari, and G. Michailidis, "Finite time identification in unstable linear systems," *Automatica*, vol. 96, pp. 342–353, 2018.
[35] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht, "Learning without mixing: Towards a sharp analysis of linear system identification," in *Proc. Conf. Learn. Theory*, 2018, pp. 439–473.
[36] T. Sarkar and A. Rakhlin, "Near optimal finite time identification of arbitrary linear dynamical systems," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 5610–5618.
[37] M. Hardt, T. Ma, and B. Recht, "Gradient descent learns linear dynamical systems," *J. Mach. Learn. Res.*, vol. 19, no. 29, pp. 1–44, 2018.
[38] S. Oymak and N. Ozay, "Non-asymptotic identification of LTI systems from a single trajectory," in *Proc. IEEE Amer. Control Conf.*, 2019, pp. 5655–5661.
[39] T. Sarkar, A. Rakhlin, and M. A. Dahleh, "Finite time LTI system identification," *J. Mach. Learn. Res.*, vol. 22, no. 26, pp. 1–61, 2021.
[40] A. Tsiamis and G. J. Pappas, "Finite sample analysis of stochastic system identification," in *Proc. IEEE 58th Conf. Decis. Control*, 2019, pp. 3648–3654.
[41] B. Anderson and J. Moore, *Optimal Filtering*. Mineola, NY, USA: Dover, 2005.
[42] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems: Theory–Implementation–Applications.* Berlin, Germany: Springer Sci. Business Media, 2012.
[43] B. Lee and A. Lamperski, "Non-asymptotic closed-loop system identification using autoregressive processes and Hankel model reduction," in *Proc. 59th IEEE Conf. Decis. Control*, 2020, pp. 3419–3424.
[44] Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári, "Improved algorithms for linear stochastic bandits," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 2312–2320.
[45] H. Lee, "Improved rates for identification of partially observed linear dynamical systems," in *Proc. Int. Conf. Algorithmic Learn. Theory*, 2022, pp. 668–698.
[46] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford, U.K.: Oxford Univ. Press, 2013.



Anastasios Tsiamis (Student Member, IEEE) received the diploma degree in electrical and computer engineering from the National Technical University of Athens, Athens, Greece, in 2014. He is currently working toward the Ph.D. degree with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA.

His research interests include control theory and machine learning.

Mr. Tsiamis was a finalist for the IFAC Young Author Prize in IFAC 2017 World Congress and a finalist for the Best Student Paper Award in ACC 2019.



George J. Pappas (Fellow, IEEE) received the Ph.D. degree in electrical engineering and computer science from the University of California, Berkeley, Berkeley, CA, USA, in 1998.

He is currently the UPS Foundation Professor and the Chair of the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA. He is a member of the GRASP Lab and the PRECISE Center. He was the Deputy Dean for Research with the School of Engineering and Applied Science. His research interests include control systems, robotics, and machine learning.

Dr. Pappas was the recipient of numerous awards, such as the IEEE CSS Antonio Ruberti Young Researcher Prize, the George S. Axelby Outstanding Paper Award, the Hugo Schuck Best Paper Award, the George H. Heilmeier Award, the NSF PECASE award, and Best Papers Awards at ACC, CDC, ICRA, ICCPS, ICASSP. Most of his former students and postdocs are now with leading academic institutions around the world.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:30:44 UTC from IEEE Xplore. Restrictions apply.