Proceedings of Machine Learning Research vol 120:1–10, 2020 2nd Annual Conference on Learning for Dynamics and Control

Sample Complexity of Kalman Filtering for Unknown Systems

Anastasios Tsiamis ATSIAMIS@SEAS.UPENN.EDU
Nikolai Matni NMATNI@SEAS.UPENN.EDU
George J. Pappas PAPPASG@SEAS.UPENN.EDU

Editors: A. Bayen, A. Jadbabaie, G. J. Pappas, P. Parrilo, B. Recht, C. Tomlin, M.Zeilinger

Abstract

In this paper, we consider the task of designing a Kalman Filter (KF) for an unknown and partially observed autonomous linear time invariant system driven by process and sensor noise. To do so, we propose studying the following two step process: first, using system identification tools rooted in subspace methods, we obtain coarse finite-data estimates of the state-space parameters, and Kalman gain describing the autonomous system; and second, we use these approximate parameters to design a filter which produces estimates of the system state. We show that when the system identification step produces sufficiently accurate estimates, or when the underlying true KF is sufficiently robust, that a Certainty Equivalent (CE) KF, i.e., one designed using the estimated parameters directly, enjoys provable sub-optimality guarantees. We further show that when these conditions fail, and in particular, when the CE KF is marginally stable (i.e., has eigenvalues very close to the unit circle), that imposing additional robustness constraints on the filter leads to similar sub-optimality guarantees. We further show that with high probability, both the CE and robust filters have mean prediction error bounded by $\tilde{O}(1/\sqrt{N})$, where N is the number of data points collected in the system identification step. To the best of our knowledge, these are the first end-to-end sample complexity bounds for the Kalman Filtering of an unknown system.

Keywords: Kalman Filter, System Identification, Sample Complexity, Certainty Equivalence

1. Introduction

Time series prediction is a fundamental problem across control theory (Kailath et al., 2000), economics (Bauer and Wagner, 2002) and machine learning. In the case of autonomous linear time invariant (LTI) systems driven by Gaussian process and sensor noise:

$$z_{k+1} = Az_k + w_k,\ y_k = Cz_k + v_k \tag{1}$$

the celebrated Kalman Filter (KF) has been the standard method for prediction (Anderson and Moore, 2005) . When model (1) is known, the KF minimizes the mean square prediction error. However, in many practical cases of interest (e.g., tracking moving objects, stock price forecasting), the state-space parameters are not known and must be learned from time-series data. This system identification step, based on a finite amount of data, inevitably introduces parametric errors in model (1), which leads to a KF with suboptimal prediction performance (El Ghaoui and Calafiore, 2001).

In this paper, we study this scenario, and provide finite-data estimation guarantees for the Kalman Filtering of an unknown autonomous LTI system (1). We consider a simple two step procedure. In

This work is supported by the AFOSR under grant FA9550-19-1-0265 (Assured Autonomy in Contested Environments), and the NSF under grant CPS 1837253.

© 2020 A. Tsiamis, N. Matni & G.J. Pappas.

the first step, using system identification tools rooted in subspace methods, we obtain finite-data estimates of the state-space parameters, and Kalman gain describing system (1). Then, in the second step, we use these approximate parameters to design a filter which predicts the system state. We provide an end-to-end analysis of this two-step procedure, and characterize the sub-optimality of the resulting filter in terms of the number of samples used during the system identification step, where the sub-optimality is measured in terms of the mean square prediction error of the filter. A key insight that emerges from our analysis is that using a Certainty Equivalent (CE) Kalman Filter, i.e., using a KF computed directly from estimated parameters, can yield poor estimation performance if the resulting CE KF has eigenvalues close to the unit circle. To address this issue, we propose a Robust Kalman Filter that mitigates these effects and that still enjoys provable sub-optimality guarantees.

Our main contributions are that: i) we show that if the system identification step produces sufficiently accurate estimates, or if the underlying true KF is sufficiently robust, then the CE KF has near optimal mean square prediction error, ii) we show when the CE KF is marginally stable, i.e., when it has eigenvalues close to the unit circle, that a Robust KF synthesized by explicitly imposing bounds on the magnitude of certain closed loop maps of the system enjoys similar mean square prediction error bounds as the CE KF, while demonstrating improved stability properties, and iii) we integrate the above results with the finite-data system identification guarantees of Tsiamis and Pappas (2019), to provide, to the best of our knowledge, the first end-to-end sample complexity bounds for the Kalman Filtering of an unknown system. In particular, we show that the mean square estimation error of both the Certainty Equivalent and Robust Kalman filter produced by the two step procedure described above is, with high probability, bounded by $\tilde{O}(1/\sqrt{N})$, where N is the number of samples collected in the system identification step.

Related work. A similar two step process was studied for the Linear Quadratic (LQ) control of an unknown system in Dean et al. (2017); Mania et al. (2019). While LQ optimal control and Kalman Filtering are known to be dual problems, this duality breaks down when the state-space parameters describing the system dynamics are not known. In particular, the LQ optimal control problem assumes full state information, making the system identification step much simpler – in particular, it reduces to a simple least-squares problem. In contrast, in the KF setting, as only partial observations are available, the additional challenge of finding an appropriate system order and state-space realization must be addressed. On the other hand, in the KF problem one can directly estimate the KF gain from data, which makes analyzing performance of the CE KF simpler than the performance of the CE LQ optimal controller (Mania et al., 2019).

System identification of autonomous LTI systems (1) is referred to as stochastic system identification (Van Overschee and De Moor, 2012). Classical results consider the asymptotic consistency of stochastic subspace system identification, as in Deistler et al. (1995); Bauer et al. (1999), whereas contemporary results seek to provide finite data guarantees (Tsiamis and Pappas, 2019; Lee and Lamperski, 2019). Finite data guarantees for system identification of partially observed systems can also be found in Oymak and Ozay (2018); Simchowitz et al. (2019); Sarkar et al. (2019), but these results focus on learning the non-stochastic part of the system, assuming that a user specified input is used to persistently excite the dynamics.

Classical approaches to robust Kalman Filtering can be found in El Ghaoui and Calafiore (2001); Sayed et al. (2001); Levy and Nikoukhah (2012), where parametric uncertainty is explicitly taken into account during the filter synthesis procedure. Although similar in spirit to our robust KF procedure, these approaches assume fixed parametric uncertainty, and do not characterize the effects of parametric uncertainty on estimation performance, with this latter step being key in providing

end-to-end sample complexity bounds. We also note that although not directly comparable to our work, the filtering problem for an unknown LTI system was also recently studied in the adversarial noise setting in Hazan et al. (2018), where a spectral filtering technique is used to directly predict the output bypassing the system identification step. In the stochastic noise case, online-learning of the Kalman Filter was studied in Kozdoba et al. (2019), where the goal is to predict a scalar output

Paper structure. In Sec. 2, we formulate the problem, and in Sec. 3 and 4, we derive performance guarantees for the proposed CE and Robust Kalman filters. In Sec. 5, we provide end-to-end sample complexity bounds for our two step procedure, and demonstrate the effectiveness of our pipeline with a numerical example in Sec. 6. We end with a discussion of future work in Sec. 7. An extended version of the paper, containing all proofs and results, is available on arXiv under the same title.

Notation. We let bold symbols denote the frequency representation of signals. For example, $\mathbf{\Phi} = \sum_{t=0}^{\infty} \Phi_t z^{-t}$. If M is stable with spectral radius $\rho(M) < 1$, then we denote its resolvent by $\mathfrak{R}_M \triangleq (zI - M)^{-1}$. The $\mathcal{H}_2$ system norm is defined by $\|\mathbf{\Phi}\|_{\mathcal{H}_2}^2 \triangleq \sum_{t=0}^{\infty} \|\Phi_t\|_F^2$, where $\|\cdot\|_F$ is the Frobenius norm. The $\mathcal{H}_\infty$ system norm is defined by $\|\mathbf{\Phi}\|_{\mathcal{H}_\infty} \triangleq \sup_{\|z\|=1} \|\mathbf{\Phi}(z)\|_2$, where $\|\cdot\|_2$ is the spectral norm. Let $\frac{1}{z}\mathcal{RH}_\infty$ be the set of real rational stable strictly proper transfer matrices.

2. Problem Formulation

For the remainder of the paper, we consider the Kalman Filter form of system (1):

$$x_{k+1} = Ax_k + Ke_k,\ y_k = Cx_k + e_k, \tag{2}$$

where $x_k \in \mathbb{R}^n$ is the prediction (state), $y_k \in \mathbb{R}^m$ is the output, and $e_k \in \mathbb{R}^n$ is the innovation process. The innovations e_k are assumed to be i.i.d. zero mean Gaussians, with positive definite covariance matrix R, and the initial state is assumed to be $x_0 = 0$. In general, the system (1) driven by i.i.d. zero mean Gaussian process and sensor noise is equivalent to system (2) for a suitable gain matrix K, as both noise models produce outputs with identical statistical properties (Van Overschee and De Moor, 2012, Chapter 3). We make the following assumption throughout the rest of the paper.

Assumption 1 *Matrices A, C, K, R are unknown, and the pair (A, C) is observable. Both the matrices A and $A - KC$ have spectral radius less than 1, i.e., $\rho(A) < 1$ and $\rho(A - KC) < 1$.*

The observability assumption is standard, and the stability of $A - KC$ follows from the properties of the Kalman filter (Anderson and Moore, 2005). We note that the filter synthesis procedures we propose can be applied even if $\rho(A) \geq 1$ – however, in this case, we are unable to guarantee bounded estimation error for the resulting CE and robust KFs (see Theorem 1 and Lemma 2).

Our goal is to provide end-to-end sample complexity bounds for the two step pipeline illustrated in Fig. 1. First, we collect a trajectory $\{y_t\}_{t=0}^N$ of length N from system (2), and use system identification tools with finite data guarantees to learn the parameters $\hat{A}, \hat{C}, \hat{K}, \hat{R}$ and bound the corresponding parameter uncertainties by $(\epsilon_A, \epsilon_C, \epsilon_K, \epsilon_R)$. Second, we use these approximate parameters to synthesize a filter from the following class:

$$\tilde{x}_k = \hat{A}\tilde{x}_{k-1} + \sum_{t=1}^{k} L_t(y_{k-t} - \hat{C}\tilde{x}_{k-t}), \quad \tilde{J} \triangleq \sqrt{\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T} \|\tilde{x}_k - x_k\|_2^2} \tag{3}$$

where $\{L_t\}_{t=1}^{\infty}$ are to be designed and $\tilde{J}$ is the filter's mean square prediction error as defined with respect to the optimal KF. Note that the predictor class above includes the CE KF – see Section 3

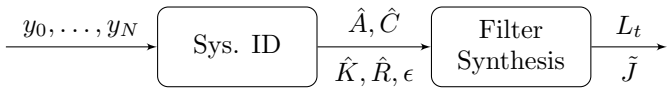


Figure 1: The proposed identification and filter synthesis pipeline. Using a single trajectory of N samples $\{y_t\}_{t=0}^N$ generated by system (2), a system identification algorithm computes estimates of $(\hat{A}, \hat{C}, \hat{K}, \hat{R})$ with corresponding identification error bounds $\epsilon := \max(\epsilon_A, \epsilon_C, \epsilon_K, \epsilon_R)$. Then, using these estimates, we synthesize a filter defined by dynamic gains $\{L_t\}_{t=1}^\infty$, which has mean square prediction error $\tilde{J}$, defined in (3).

– and that if the the true system parameters are known, i.e., if $\hat{A} = A$, $\hat{C} = C$, $\hat{K} = K$, then the optimal mean squared prediction error $\tilde{J} = 0$ is achieved.

Problem 1 (End-to-end Sample Complexity) *Fix a failure probability $\delta > 0$. Given a single trajectory $y_0, \ldots, y_N$ of system (2), compute system parameter estimates $\hat{A}, \hat{C}, \hat{K}, \hat{R}$, and design a Kalman filter in class (3), defined by gains $\{L_t\}_{t=1}^\infty$, such that with probability at least $1 - \delta$, we have that $\tilde{J} \leq \epsilon_J$, so long as $N \geq \mathrm{poly}(1/\epsilon_J, \log(1/\delta))$.*

To address Problem 1, we will: i) leverage recent results regarding the the sample complexity of stochastic system identification, ii) provide estimation guarantees for certainty equivalent as well as robust Kalman filter designed using the identified system parameters (see Problem 2 below), and (iii) provide end-to-end performance guarantees by integrating steps (i) and (ii) (see Problem 1 above).

Recently Tsiamis and Pappas (2019) provided a finite sample analysis for stochastic system identification which provides bounds on the identification error $\epsilon := \max(\epsilon_A, \epsilon_C, \epsilon_K, \epsilon_R)$. Leveraging these results, we focus next on solving the Filter Synthesis task described below using both a certainty equivalent Kalman filter as well as a robust Kalman filter.

Problem 2 (Near Optimal Kalman Filtering of an Uncertain System) *Consider system (2). Let $\hat{A}, \hat{C}, \hat{K}, \hat{R}$ be estimates satisfying*[1] $\|A - \hat{A}\|_2 \leq \epsilon_A$, $\|C - \hat{C}\|_2 \leq \epsilon_C$, $\|K - \hat{K}\|_2 \leq \epsilon_K$, $\|R - \hat{R}\|_2 \leq \epsilon_R$. *Design a Kalman filter in class (3), defined by gains $\{L_t\}_{t=0}^\infty$, with mean square prediction error decaying with the size of the parameter uncertainty, i.e., such that $\tilde{J} \leq O(\epsilon_A, \epsilon_C, \epsilon_K, \epsilon_R)$.*

3. Estimation Guarantees for Certainty Equivalent Kalman Filtering

For the certainty equivalent Kalman filter, we directly use the estimated state-space parameters from the system identification step. Based on the estimated $\hat{K}, \hat{R}$ we compute the covariance:

$$\begin{bmatrix} \hat{Q} & \hat{S} \\ \hat{S}^* & \hat{R} \end{bmatrix} \triangleq \mathbb{E} \begin{bmatrix} \hat{K} e_k \\ e_k \end{bmatrix} \begin{bmatrix} e_k^* \hat{K}^* & e_k^* \end{bmatrix} = \begin{bmatrix} \hat{K} \hat{R}^{1/2} \\ \hat{R}^{1/2} \end{bmatrix} \begin{bmatrix} \hat{R}^{1/2} \hat{K}^* & \hat{R}^{1/2} \end{bmatrix}.$$

Then, based on standard Kalman filter theory, we compute the stabilizing solution[2] of the following Riccati equation with correlation terms (Kailath et al., 2000):

$$P = \hat{A} P \hat{A}^* + \hat{Q} - (\hat{A} P \hat{C}^* + \hat{S})(\hat{C} P \hat{C}^* + \hat{R})^{-1} (\hat{C} P \hat{A}^* + \hat{S}^*).$$

1. In practice, estimating the parameters of a partially observed system (2) is ill-posed, in that any similarity transformation S can be applied to generate parameters $(S^{-1}AS, CS, S^{-1}K, R)$ describing the same system, and the bounds described hold for *some* similarity transformation S. All results in this paper apply nearly as is to the general case of $S \neq I$ under suitable assumptions – more details can be found in the extended version.
2. A stabilizing solution P to the Riccati equation defines a Kalman gain L_{CE} such that $\rho(\hat{A} - L_{CE}\hat{C}) < 1$.

Then, the CE Kalman filter gain is static and takes the form

$$L_1 = L_{CE} \triangleq (\hat{A}P\hat{C}^* + \hat{S})(\hat{C}P\hat{C}^* + \hat{R})^{-1},\ L_t = 0,\ \text{for } t = 2, \ldots. \tag{4}$$

Trivially, if $\rho(\hat{A} - \hat{K}\hat{C}) < 1$, then the stabilizing solution of the Riccati equation is $P = 0$ with $L_{CE} = \hat{K}$; the solution does not depend on $\hat{R}$. The next result shows that if the underlying true Kalman filter is sufficiently robust, as measured by a spectral decay rate, and that estimation parameter errors are sufficiently small, then the CE Kalman filter achieves near optimal performance.

Theorem 1 (Near Optimal Certainty Equivalent Kalman Filtering) *Consider Problem 2 and the CE KF (4). For any $\rho(A - KC) \leq \rho < 1$, define $\tau(A - KC, \rho) \triangleq \sup_{t\geq 0} \left\|(A - KC)^t\right\|_2 \rho^{-t}$. If the robustness condition $2\tau(A - KC, \rho) \cdot (\epsilon_A + \epsilon_C(\|K\|_2 + \epsilon_K) + \epsilon_K \|C\|_2) \leq 1 - \rho$ is satisfied, then $L_{CE} = \hat{K}$ and:*

$$\tilde{J} \leq \sqrt{3}\bar{\mathcal{C}}\epsilon \left\| \begin{bmatrix} \mathfrak{R}_A K \\ I \end{bmatrix} R^{1/2} \right\|_{\mathcal{H}_2}$$

where $\epsilon = \max\{\epsilon_A, \epsilon_C, \epsilon_K\}$, $\bar{\mathcal{C}} = 2\frac{\tau(A-KC,\rho)}{1-\rho}(1 + \|K\|_2 + \epsilon_K)$ and $\mathfrak{R}_A = (zI - A)^{-1}$.

When CE Kalman filtering fails. The transient behavior of the CE Kalman filter is governed by the closed loop eigenvalues of $\hat{A} - \hat{K}\hat{C}$, with performance degrading as eigenvalues approach the unit circle. This may occur if the estimation errors $(\epsilon_A, \epsilon_C, \epsilon_K)$ are large enough to cause $\rho(\hat{A} - \hat{K}\hat{C}) \approx 1$ even if the true system has spectral radius $\rho(A - KC) < 1$. We show in the next section that this undesirable scenario can be avoided by explicitly constraining the transient response of the resulting Kalman filter to satisfy certain robustness constraints.

4. Estimation Guarantees for Robust Kalman Filtering

To address the possible poor performance of the CE Kalman filter when model uncertainty is large, we propose to search over dynamic filters (3) subject to additional robustness constraints on their transient response. Using the System Level Synthesis (SLS) framework (Wang et al., 2019; Anderson et al., 2019) for Kalman Filtering (Wang et al., 2015), we parameterize the class of dynamic filters (3) subject to additional robustness constraints in a way that leads to convex optimization problems.

For a given dynamic predictor $\boldsymbol{L}(z) = \sum_{t=0}^{\infty} z^{-t} L_{t+1}$, we define the closed loop *system responses*:

$$\boldsymbol{\Phi}_w(z) \triangleq (zI - \hat{A} + \boldsymbol{L}\hat{C})^{-1},\ \boldsymbol{\Phi}_v(z) \triangleq -(zI - \hat{A} + \boldsymbol{L}\hat{C})^{-1}\boldsymbol{L}. \tag{5}$$

In (Wang et al., 2015), it is shown that these responses are in fact the closed loop maps from process and sensor noise $(\boldsymbol{w}, \boldsymbol{v})$ to state estimation error, and that the filter gain achieving the desired behavior can be recovered via $\boldsymbol{L} = -\boldsymbol{\Phi}_w^{-1}\boldsymbol{\Phi}_v$ so long as the responses $(\boldsymbol{\Phi}_w, \boldsymbol{\Phi}_v)$ are constrained to lie in an affine space defined by the system dynamics. By expressing the mean squared prediction error of the filters (3) in terms of their system responses, we are able to clearly delineate the effects of parametric uncertainty from the cost of deviating from the CE Kalman filter.

Lemma 2 (Error analysis) *Consider system (2). Let $\Delta_A \triangleq A - \hat{A}$, $\Delta_C \triangleq C - \hat{C}$, $\Delta_K \triangleq K - \hat{K}$. Any filter (3) with parameterization (5) has mean squared prediction error given by*

$$\tilde{J} = \left\| \begin{bmatrix} \boldsymbol{\Phi}_w & \boldsymbol{\Phi}_v \end{bmatrix} \left\{ \begin{bmatrix} \Delta_A & \Delta_K \\ \Delta_C & 0 \end{bmatrix} \begin{bmatrix} \mathfrak{R}_A K \\ I \end{bmatrix} + \begin{bmatrix} \hat{K} \\ I \end{bmatrix} \right\} R^{1/2} \right\|_{\mathcal{H}_2}$$

Based on the previous lemma, we can upper bound the mean squared prediction error of filters (3) by

$$\tilde{J} \leq \underbrace{\sqrt{3}\epsilon \left\| \begin{bmatrix} \mathbf{\Phi}_w & \mathbf{\Phi}_v \end{bmatrix} \right\|_{\mathcal{H}_2} \left\| \begin{bmatrix} \mathfrak{R}_A K \\ I \end{bmatrix} R^{1/2} \right\|_{\mathcal{H}_\infty}}_{\text{parameter uncertainty term}} + \underbrace{\left\| \mathbf{\Phi}_w \hat{K} + \mathbf{\Phi}_v \right\|_{\mathcal{H}_2} \left\| R^{1/2} \right\|_2}_{\text{suboptimality term}},$$

where $\epsilon = \max\{\epsilon_A, \epsilon_C, \epsilon_K\}$. This upper bound clearly separates the effects of parameter uncertainty, as captured by the first term, and the performance cost incurred by the filter $\boldsymbol{L}$ due to its deviation from the CE Kalman gain $\hat{K}$, as captured by the second. In order to optimally tradeoff between these two terms, we propose the following robust SLS optimization problem:

$$\begin{aligned} &\min_{\mathbf{\Phi}_w, \mathbf{\Phi}_v} \left\| \mathbf{\Phi}_w \hat{K} + \mathbf{\Phi}_v \right\|_{\mathcal{H}_2} \\ &\text{s.t. } \left\| \begin{bmatrix} \mathbf{\Phi}_w & \mathbf{\Phi}_v \end{bmatrix} \right\|_{\mathcal{H}_2} \leq \mathcal{C}, \mathbf{\Phi}_w(zI - \hat{A}) - \mathbf{\Phi}_v \hat{C} = I, \ \mathbf{\Phi}_w, \ \mathbf{\Phi}_v \in \frac{1}{z}\mathcal{RH}_\infty \end{aligned} \tag{6}$$

where the constant $\mathcal{C}$ is a regularization parameter, and the affine constraint $\mathbf{\Phi}_w(zI - \hat{A}) - \mathbf{\Phi}_v \hat{C} = I$, $\mathbf{\Phi}_w$, $\mathbf{\Phi}_v \in \frac{1}{z}\mathcal{RH}_\infty$ parameterizes all filters of the form (3) that have bounded mean squared prediction error (see Wang et al. (2015) for more details). As we formalize in the following theorem, for appropriately selected regularization parameter $\mathcal{C}$ and sufficiently accurate estimation errors (ϵ_A, ϵ_C), the robust KF has near optimal mean square estimation error.

Theorem 3 (Robust Kalman Filter) *Consider Problem 2 with Kalman filters from class (3) synthesized using the robust SLS optimization problem (6). If the regularization parameter is chosen such that $\mathcal{C} \geq 2(1 + \|K\|_2) \|\mathfrak{R}_{A-KC}\|_{\mathcal{H}_2}$, and further, the estimation errors (ϵ_A, ϵ_C) are such that*

$$(\epsilon_A + \epsilon_C \|K\|_2) \|\mathfrak{R}_{A-KC}\|_{\mathcal{H}_\infty} \leq 1/2 \tag{7}$$

then the robust SLS optimization problem is feasible, and the synthesized robust Kalman filter has mean squared prediction error upper-bounded by

$$\tilde{J} \leq \sqrt{3}\mathcal{C}\epsilon \left\| \begin{bmatrix} \mathfrak{R}_A K \\ I \end{bmatrix} \right\|_{\mathcal{H}_\infty} \|R^{1/2}\|_2 + 2\epsilon \|\mathfrak{R}_{A-KC}\|_{\mathcal{H}_2} \|R^{1/2}\|_2, \tag{8}$$

where $\epsilon = \max\{\epsilon_A, \epsilon_C, \epsilon_K\}$.

We further note that whenever the system responses induced by the CE Kalman filter $\tilde{\mathbf{\Phi}}_w \triangleq (zI - \hat{A} + \hat{K}\hat{C})^{-1}$, $\tilde{\mathbf{\Phi}}_v \triangleq -(zI - \hat{A} + \hat{K}\hat{C})^{-1}\hat{K}$ are a feasible solution to optimization problem (6), they are also optimal, resulting in a filter $\boldsymbol{L} = \hat{K}$ with performance identical to the CE setting.

5. End-to-End Sample Complexity for the Kalman Filter

Theorems 1 and 3 provide two different solutions to Problem 2. Combining these theorems with the finite data system identification guarantees of Tsiamis and Pappas (2019), we now derive, to the best of our knowledge, the first end-to-end sample complexity bounds for the Kalman filtering of an unknown system. For both the CE and robust Kalman filter, we show that the mean squared estimation error defined in (3) decreases with rate $O(1/\sqrt{N})$ up to logarithmic terms, where N is the number of samples collected during the system identification step. The formal statement of the following theorem which addresses Problem 1 can be found in the arXiv version of the paper.

Theorem 4 (End-to-end guarantees, informal) *Fix a failure probability $\delta \in (0,1)$, and assume that we are given a sample trajectory $\{y_t\}_{t=0}^N$ generated by system (2). Then as long as $N \geq \mathrm{poly}(\log(1/\delta))$, we have with probability at least $1-\delta$ that the identification and filter synthesis pipeline of Fig. 1, with system identification performed as in Tsiamis and Pappas (2019) and filter synthesis performed as in Sections 3, 4, achieves mean squared prediction error satisfying*

$$\tilde{J} \leq \mathcal{C}_{ID}\mathcal{C}_{KF}\tilde{O}(\sqrt{\frac{\log 1/\delta}{N}}), \text{ where } \mathcal{C}_{KF} = \frac{1}{1-\rho(A-KC)}\frac{1}{1-\rho(A)}$$

and $\mathcal{C}_{ID}$ captures the difficulty of identifying system (2)(see Tsiamis and Pappas (2019) or extended version). Here, $\tilde{O}$ hides constants, other system parameters, and logarithmic terms.

The bound derived in Theorem 4 highlights an interesting tension between how easy it is to identify the unknown system, and the robustness of the underlying optimal Kalman filter. The constant $\mathcal{C}_{KF}$ captures how robust the underlying open loop system A and closed loop Kalman filter $A-KC$ are, as measured by their spectral gaps $1-\rho(A)$ and $1-\rho(A-KC)$. In particular, we expect $\mathcal{C}_{KF}$ to be small for systems that admit optimal KFs with favorable robustness and transient performance. In contrast, the constant $\mathcal{C}_{ID}$ captures how easy it is to identify a system: recent results for the fully observed setting (Simchowitz et al., 2018; Sarkar and Rakhlin, 2018) suggest that systems with *larger* spectral radius are in fact easier to identify, as they provide more "signal" to the identification algorithm. In this way, our upper bound suggests that systems which properly balance between these two properties, robust transient performance and ease of identification, enjoy favorable sample complexity.

We also note that the degradation of our bound with the inverse of the spectral gap $1-\rho(A)$ appears to be a limitation of the proposed offline two step architecture – indeed, Lemma 2 suggests that any estimation error in the state-space parameters (A, C) causes an increase in mean squared prediction error as $\|\mathfrak{R}_A\| \propto (1-\rho(A))^{-1}$ increases. It remains open as to whether other prediction architectures would suffer from the same limitation.

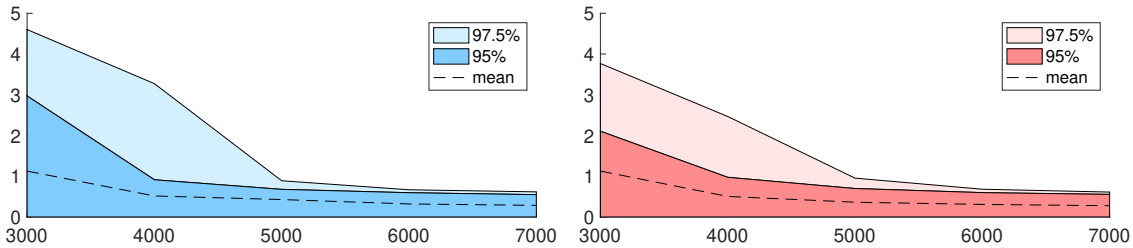
6. Simulations

We perform Monte Carlo simulations of the proposed pipeline for the system

$$A = \begin{bmatrix} 0.8 & 1 & 0 \\ 0 & 0.9 & 1 \\ 0 & 0 & 0.9 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, K = \begin{bmatrix} 1.5320 & 0.9401 & 0.1923 \end{bmatrix}^*, R = 10.6414.$$

for varying sample lengths N. We simulate both the CE and robust Kalman filters, and set the regularization parameter to $\mathcal{C} = 10$ in the robust SLS optimization problem (6). For each iteration, we first simulate system (2) to obtain N output samples. Then, we perform system identification to obtain the system parameters, after which we synthesize both CE and robust Kalman filters. Finally, we compute the mean prediction error of the designed filters. We refer the reader to the extended version of the paper for a detailed description of our experimental implementation.

Figure 2 (a) and (b) show the empirically computed mean squared prediction errors of the CE and Robust Kalman filters, with the mean, 95th, and 97.5th percentiles being shown. Notice that both errors decrease with a rate of $1/\sqrt{N}$, and that while the average behavior of both filters is quite similar, there is a noticeable gap in their tail behaviors. We observe that the most significant gap



(a) CE Kalman Filter (b) Robust Kalman Filter

Figure 2: The 95% and 97.5% empirical percentiles for the mean squared prediction error $\tilde{J}$ of the CE and Robust Kalman filters. We run 1000 Monte Carlo simulations for different sample lengths N (x-axis, number of samples).

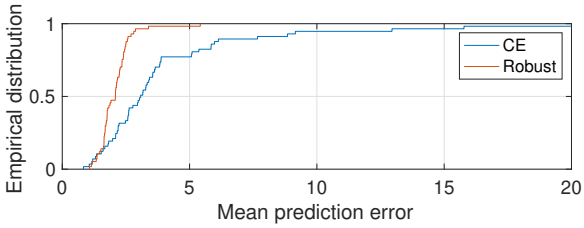


Figure 3: Performance improvement for the robust KF conditioned on the event that $\rho(\hat{A} - L_{CE}\hat{C}) > 0.97$.

between the CE and Robust Kalman filters occurs when the eigenvalues of the CE matrix $\hat{A} - L_{CE}\hat{C}$ are close to the unit circle. Fig. 3 shows the empirical distribution of mean squared prediction errors conditioned on the event that $\rho(\hat{A} - L_{CE}\hat{C}) > 0.97$. In this case, the CE filter can exhibit *extremely* poor mean squared prediction error, with the worst observed error (not shown in Fig. 3 in the interst of space) approximately equal to 70 – in contrast, the worst error exhibited by the robust Kalman filter was approximately equal to 5. Thus, we were able to achieve a 14x reduction in worst-case mean squared error. For some simulations the robust KF can exhibit worse performance compared to the CE Kalman Filter. However, over all simulations, the robust KF was at most 1.64x suboptimal compared to the CE Kalman filter.

7. Conclusions & Future work

In this paper, we proposed and analyzed a system identification and filter synthesis pipeline. Leveraging contemporary finite data guarantees from system identification (Tsiamis and Pappas, 2019), as well as novel parameterizations of robust Kalman filters (Wang et al., 2015), we provided, to the best of our knowledge, the first end-to-end sample complexity bounds for the Kalman filtering of an unknown LTI autonomous system. Our analysis revealed that, depending on the spectral properties of the CE Kalman filter, that a robust Kalman filter approach may lead to improved performance. In future work, we would like to explore how to improve robustness and performance by further exploiting information about system uncertainty, as well as how to integrate our results into an optimal control framework, such as Linear Quadratic Gaussian control. We would also like to explore more options for the regularization parameter $\mathcal{C}$, for example, let it depend on the data or the number of samples.

References

B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Dover Publications, 2005.

James Anderson, John C Doyle, Steven H Low, and Nikolai Matni. System level synthesis. *Annual Reviews in Control*, 2019.

Dietmar Bauer and Martin Wagner. Estimating cointegrated systems using subspace algorithms. *Journal of Econometrics*, 111(1):47–84, 2002.

Dietmar Bauer, Manfred Deistler, and Wolfgang Scherrer. Consistency and asymptotic normality of some subspace algorithms for systems without observed inputs. *Automatica*, 35(7):1243–1254, 1999.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *arXiv preprint arXiv:1710.01688*, 2017.

Manfred Deistler, K Peternell, and Wolfgang Scherrer. Consistency and relative efficiency of subspace methods. *Automatica*, 31(12):1865–1875, 1995.

Laurent El Ghaoui and Giuseppe Calafiore. Robust filtering for discrete-time systems with bounded noise and parametric uncertainty. *IEEE Transactions on Automatic Control*, 46(7):1084–1089, 2001.

Elad Hazan, Holden Lee, Karan Singh, Cyril Zhang, and Yi Zhang. Spectral filtering for general linear dynamical systems. In *Advances in Neural Information Processing Systems*, pages 4634–4643, 2018.

Thomas Kailath, Ali H Sayed, and Babak Hassibi. *Linear estimation*. Prentice Hall, 2000.

Mark Kozdoba, Jakub Marecek, Tigran Tchrakian, and Shie Mannor. On-line learning of linear dynamical systems: Exponential forgetting in Kalman filters. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4098–4105, 2019.

Bruce Lee and Andrew Lamperski. Non-asymptotic Closed-Loop System Identification using Autoregressive Processes and Hankel Model Reduction. *arXiv preprint arXiv:1909.02192*, 2019.

Bernard C Levy and Ramine Nikoukhah. Robust state space filtering under incremental model perturbations subject to a relative entropy tolerance. *IEEE Transactions on Automatic Control*, 58 (3):682–695, 2012.

Horia Mania, Stephen Tu, and Benjamin Recht. Certainty equivalent control of LQR is efficient. *arXiv preprint arXiv:1902.07826*, 2019.

Samet Oymak and Necmiye Ozay. Non-asymptotic Identification of LTI Systems from a Single Trajectory. *arXiv preprint arXiv:1806.05722*, 2018.

Tuhin Sarkar and Alexander Rakhlin. Near optimal finite time identification of arbitrary linear dynamical systems. *arXiv preprint arXiv:1812.01251*, 2018.

Tuhin Sarkar, Alexander Rakhlin, and Munther A Dahleh. Finite-Time System Identification for Partially Observed LTI Systems of Unknown Order. *arXiv preprint arXiv:1902.01848*, 2019.

Ali H Sayed et al. A framework for state-space estimation with uncertain models. *IEEE Transactions on Automatic Control*, 46(7):998–1013, 2001.

Max Simchowitz, Horia Mania, Stephen Tu, Michael I Jordan, and Benjamin Recht. Learning Without Mixing: Towards A Sharp Analysis of Linear System Identification. *arXiv preprint arXiv:1802.08334*, 2018.

Max Simchowitz, Ross Boczar, and Benjamin Recht. Learning Linear Dynamical Systems with Semi-Parametric Least Squares. *arXiv preprint arXiv:1902.00768*, 2019.

Anastasios Tsiamis and George J Pappas. Finite Sample Analysis of Stochastic System Identification. In *IEEE 58th Conference on Decision and Control (CDC)*, 2019.

Peter Van Overschee and Bart De Moor. *Subspace identification for linear systems: Theory–Implementation–Applications*. Springer Science & Business Media, 2012.

Yuh-Shyang Wang, Seungil You, and Nikolai Matni. Localized distributed Kalman filters for large-scale systems. *IFAC-PapersOnLine*, 48(22):52–57, 2015.

Yuh-Shyang Wang, Nikolai Matni, and John C Doyle. A system level approach to controller synthesis. *IEEE Transactions on Automatic Control*, 2019.