Quantification of Mismatch Error in Randomly Switching Linear State-Space Models

Parisa Karimi, Zhizhen Zhao, Mark D. Butala, and Farzad Kamalabadi

***Abstract*—Switching Kalman Filters (SKF) are well known for solving switching linear dynamic system (SLDS), i.e., piece-wise linear estimation problems. Practical SKFs are heuristic, approximate filters and require more computational resources than a single-mode Kalman filter (KF). On the other hand, applying a single-mode mismatched KF to an SLDS results in erroneous estimation. This letter quantifies the average error an SKF can eliminate compared to a mismatched, single-mode KF before collecting measurements. Derivations of the first and second moments of the estimators' errors are provided and compared. One can use these derivations to quantify the average performance of filters beforehand and decide which filter to run in operation to have the best performance in terms of estimation error and computation complexity. We further provide simulation results that verify our mathematical derivations.**

***Index Terms*—Switching Kalman filter, recursive estimation, detection, switching linear dynamic systems, model mismatch.**

I. Introduction

ESTIMATING the multidimensional state of a dynamic system from indirect, noisy measurements is a fundamental problem encountered in all branches of engineering and physical sciences, with specific examples such as time-dependent imaging and tomography [1], [2], geophysical data assimilation [3], gene expression modeling [4], and economic forecasting [5]. Given the initial state distribution and state-space model parameters, state estimates may be optimally recovered using Bayesian inference algorithms [6]. The Kalman filter [7] provides the optimal solution for linear state-space models with additive Gaussian noise [8].

Dynamic systems are often driven by external factors that cause shifts in behavior, e.g., a shock resulting in greater activity before settling back to a more quiescent baseline mode. Such system dynamics are often naturally modeled as a switching linear dynamic system (SLDS) [9] which augments a linear state-space model with a hidden, discrete random variable that indicates the system dynamical mode. The inference problem then involves the time-dependent, joint estimation of the discrete system mode and continuous state. Finding the exact posterior and optimal filtering in this scenario is computationally intractable [10] since the belief state grows exponentially with time, and a practical switching Kalman filter (SKF) formulation (see, e.g., [10]–[12], and the references within) involves certain approximations, e.g., time horizon truncation. Characterization of estimation error vs. computational cost is, therefore, of paramount importance.

In this work, we quantify the estimation error reduction when an SKF is used instead of a mismatched KF, i.e., when an SLDS is approximated as a single-mode system. The result has several immediate practical uses. For one, it reveals those situations when the SKF mean square error (MSE) reduction is significant in comparison to the additional computational burden. As another, it provides a means to reduce SKF computational cost by coalescing SLDS modes with the least impact on the MSE (an approach we explore in [13]). We show that SKF MSE performance is a function of 1) the detection rate at each time step, and 2) the mismatch bias whenever the algorithm detects a mode incorrectly. Both are functions of the switching distributions and transition probabilities. Due to space restrictions, we focus the scope of this work on the scenarios where an estimate or an upper bound of the detection rate is assumed to be known, and calculate the MSE accordingly. Detection rate approximation as a function of problem specification using approximate metrics [14]–[17] will be studied in a future publication.

The performance of mismatched KFs for a single-mode, linear dynamic system has been studied previously [18]–[22]. In the case of switching dynamics, [23] explores the conditions under which the instantaneous mode detection is successful based on the statistics of the predicted and collected measurement residuals. Also, in [24]–[27] the convergence of a mode-based KF is studied and the conditions under which the steady state bias term would converge to zero are investigated. However, to the best of our knowledge, the work presented in this paper is the first to quantify the transient error in SLDS state estimation based on model parameters alone and without measurements. In particular, the novel and significant contribution of this paper is the analysis of the MSE of a multi-modal Maximum A Posteriori (MAP) estimator, formulated as an SKF, in comparison to a mismatched, single-mode MAP estimator, formulated as KF, before collecting measurements.

Throughout the paper, we use the following convention for the notation. A random vector $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{C})$ follows a Gaussian distribution with mean $\boldsymbol{m}$ and covariance $\boldsymbol{C}$. The expectation, covariance, and probability are denoted by $\mathbb{E}, \mathbb{C}, p$ respectively, and $\boldsymbol{I}$ is the identity matrix.

The remainder of the paper is organized as follows. The SLDS signal model and KF/SKF formulations are reviewed in Sections II and III, respectively. Section IV derives estimation error for mismatched KFs and SKFs. Section V provides remarks on implementation. Simulations verify the derivations in Section VI, and conclusions are presented in Section VII.

Manuscript received July 23, 2021; revised September 16, 2021; accepted September 17, 2021. Date of publication September 29, 2021; date of current version October 20, 2021. This work was supported by the Grainger College of Engineering UIUC-ZJU Institute. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Kai Liu. *(Corresponding author: Parisa Karimi.)*

Parisa Karimi, Zhizhen Zhao, and Farzad Kamalabadi are with the University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA (e-mail: parisa2@illinois.edu; zhizhenz@illinois.edu; farzadk@illinois.edu).

Mark D. Butala is with the Zhejiang University, Hangzhou, Zhejiang 310027, China (e-mail: markbutala@intl.zju.edu.cn).

Digital Object Identifier 10.1109/LSP.2021.3116504

II. Signal Model

The state-space model for an SLDS is

$$\boldsymbol{x}_n = \boldsymbol{A}_n \boldsymbol{x}_{n-1} + \boldsymbol{\nu}_n, \quad (1)$$

$$\boldsymbol{y}_n = \boldsymbol{H}_n \boldsymbol{x}_n + \boldsymbol{\omega}_n, \quad (2)$$

where $\boldsymbol{\nu}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_n)$ and $\boldsymbol{\omega}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_n)$ are additive Gaussian noises such that $\boldsymbol{\nu}_n \perp \boldsymbol{\nu}_{n'} \, \forall n \neq n'$, $\boldsymbol{\omega}_n \perp \boldsymbol{\omega}_{n'} \, \forall n \neq n'$, $\boldsymbol{\nu}_n \perp \boldsymbol{\omega}_{n'} \, \forall n, n'$, and $\boldsymbol{Q}_n = \mathbb{E}[\boldsymbol{\nu}_n \boldsymbol{\nu}_n^T]$, $\boldsymbol{R}_n = \mathbb{E}[\boldsymbol{\omega}_n \boldsymbol{\omega}_n^T]$. The evolution matrices $\boldsymbol{A}_n = \boldsymbol{A}(s_n)$ are randomly drawn from the set of $d \times d$ evolution matrices $\mathcal{A} = \{\boldsymbol{A}(b) : b = 1, \ldots, r\}$, the covariance matrices $\boldsymbol{Q}_n = \boldsymbol{Q}(s_n)$ are randomly drawn from the set $\mathcal{Q} = \{\boldsymbol{Q}(b) : b = 1, \ldots, r\}$, and the random variable $s_n \in \{1, \ldots, r\}$ refers to the mode of the dynamic system at time n, which is drawn from a discrete Markov process assuming known initial prior and transition matrix. $\boldsymbol{x}_n \in \mathbb{R}^d$, the hidden d-dimensional state variable at time step n, is then estimated given the m-dimensional measurement vector $\boldsymbol{y}_n$, the $m \times d$ measurement matrix $\boldsymbol{H}_n$, $\mathcal{Q}, \mathcal{A}$, and $\boldsymbol{R}_n$. (m is the number of measurements, d is the state dimension, and r is the number of modes the system may switch between).

III. Kalman Filter/Switching Kalman Filter

The KF is the optimal estimator for linear dynamic systems. Let $\boldsymbol{y}_1^n \equiv [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n]$, $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{x}_{0|0}, \boldsymbol{P}_{0|0})$, $\boldsymbol{x}_{n|n} = \mathbb{E}[\boldsymbol{x}_n | \boldsymbol{y}_1^n]$, and $\boldsymbol{P}_{n|n} = \mathbb{C}(\boldsymbol{x}_n | \boldsymbol{y}_1^n)$. The "$Filter$" operator is defined as

$$(\boldsymbol{x}_{n|n}, \boldsymbol{P}_{n|n}) = Filter(\boldsymbol{A}_n, \boldsymbol{H}_n, \boldsymbol{x}_{n-1|n-1}, \boldsymbol{P}_{n-1|n-1}, \boldsymbol{Q}_n, \boldsymbol{R}_n, \boldsymbol{y}_1^n), \quad (3)$$

which involves the repeated application of a time update step

$$\boldsymbol{x}_{n|n-1} = \boldsymbol{A}_n \boldsymbol{x}_{n-1|n-1}, \quad (4)$$

$$\boldsymbol{P}_{n|n-1} = \boldsymbol{A}_n \boldsymbol{P}_{n-1|n-1} \boldsymbol{A}_n^T + \boldsymbol{Q}_n; \quad (5)$$

and a measurement update step such that $\boldsymbol{\epsilon}_n = \boldsymbol{y}_n - \boldsymbol{H}_n \boldsymbol{x}_{n|n-1}$, and $\boldsymbol{K}_n = \boldsymbol{P}_{n|n-1} \boldsymbol{H}_n^T (\boldsymbol{H}_n \boldsymbol{P}_{n|n-1} \boldsymbol{H}_n^T + \boldsymbol{R}_n)^{-1}$ is the Kalman gain,

$$\boldsymbol{x}_{n|n} = \boldsymbol{x}_{n|n-1} + \boldsymbol{K}_n \boldsymbol{\epsilon}_n, \quad (6)$$

$$\boldsymbol{P}_{n|n} = (\boldsymbol{I} - \boldsymbol{K}_n \boldsymbol{H}_n) \boldsymbol{P}_{n|n-1}. \quad (7)$$

The application of a KF to an SLDS (1)–(2) results in erroneous estimates. The well known SKF formulation detects the switching mode s_n and its corresponding model parameters $(\boldsymbol{A}_n, \boldsymbol{Q}_n)$ at each time step, and estimates the state variables accordingly. Upon perfect detection of the sequence of modes $\{s_n\}_{n=1}^T$ in T time steps, one obtains an optimal estimate of the state variable $\boldsymbol{x}_n$ in terms of both MAP and MSE metrics. Due to the combinatoric explosion in system mode trajectories involved in optimal SKFs [10], several approximate SKF algorithms have been proposed (e.g. [10], [12]).

In this paper, we assume the probability of correct mode detection is known and the SKF algorithm considered throughout the paper refers to the MAP estimator, which uses the KF corresponding to the mode with the largest probability at each time step. Calculating the measures studied in this paper for the MMSE-based SKF estimator is a more challenging task due to the resulting complex, non-Gaussian distributions. Since $\mathrm{MSE}(\boldsymbol{x}_{n|n}^{\mathrm{KF}}) \geq \mathrm{MSE}(\boldsymbol{x}_{n|n}^{\mathrm{MAP}}) \geq \mathrm{MSE}(\boldsymbol{x}_{n|n}^{\mathrm{MMSE}})$, it is clear that the MSE quantified in this paper provides an upper bound for the additional error caused by merging modes in an MMSE estimator. Determining the tightness of this inequality is a challenging undertaking and would be the subject of future work.

IV. Derivation of the Mean Squared Error

To quantify the performance of each filter, we first study the estimation error imposed by applying a mismatched model to a single-mode, linear dynamic model in Section IV-A. Then, in Sections IV-B and IV-C, the effect of applying a single-mode mismatched KF and SKF to an SLDS is quantified using prior and transition probabilities of the SLDS.

A. Mismatched Kalman Filter Error

Instead of the state-space equations (1)–(2) with the correct dynamic evolution model $(\boldsymbol{A}_n, \boldsymbol{Q}_n) = (\boldsymbol{A}, \boldsymbol{Q})$ at time n, consider a mismatched model using $(\widetilde{\boldsymbol{A}}_n, \widetilde{\boldsymbol{Q}}_n) = (\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{Q}})$; ($\widetilde{\boldsymbol{X}}$ refers to the parameter $\boldsymbol{X}$ in the mismatched model). It is well known that KF estimates are unbiased and optimal, but this is true when the correct model is used. In order to determine how far estimates deviate from the correct model estimates, we study the error term $\boldsymbol{e}_n = \boldsymbol{x}_n - \widetilde{\boldsymbol{x}}_{n|n}$, where $\boldsymbol{x}_n$ is the ground truth state variable, $\widetilde{\boldsymbol{x}}_{n|n}$ is the estimate from the mismatched model, and $\widetilde{\boldsymbol{K}}_n = \widetilde{\boldsymbol{K}}$ is the Kalman gain at time n obtained based on equations (4)–(7) using the mismatched dynamic model $(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{Q}})$, as follows

$$\boldsymbol{x}_n = \boldsymbol{A}\boldsymbol{x}_{n-1} + \boldsymbol{\nu}_n, \quad (8)$$

$$\begin{aligned} \widetilde{\boldsymbol{x}}_{n|n} &= \widetilde{\boldsymbol{A}}\boldsymbol{x}_{n-1|n-1} + \widetilde{\boldsymbol{K}}(\boldsymbol{y}_n - \boldsymbol{H}_n \widetilde{\boldsymbol{x}}_{n|n-1}) \\ &= \widetilde{\boldsymbol{A}}\boldsymbol{x}_{n-1|n-1} + \widetilde{\boldsymbol{K}}\big[\boldsymbol{H}_n(\boldsymbol{A}\boldsymbol{x}_{n-1} + \boldsymbol{\nu}_n) + \boldsymbol{\omega}_n \\ &\quad - \boldsymbol{H}_n \widetilde{\boldsymbol{A}}\boldsymbol{x}_{n-1|n-1}\big], \end{aligned} \quad (9)$$

$$\begin{aligned} \boldsymbol{e}_n &= \boldsymbol{A}\boldsymbol{x}_{n-1} + \boldsymbol{\nu}_n - \widetilde{\boldsymbol{A}}\boldsymbol{x}_{n-1|n-1} - \widetilde{\boldsymbol{K}}\big[\boldsymbol{H}_n(\boldsymbol{A}\boldsymbol{x}_{n-1} + \boldsymbol{\nu}_n) \\ &\quad + \boldsymbol{\omega}_n - \boldsymbol{H}_n \widetilde{\boldsymbol{A}}\boldsymbol{x}_{n-1|n-1}\big] \\ &= \big[\widetilde{\boldsymbol{B}}_n(\boldsymbol{A} - \widetilde{\boldsymbol{A}})\big]\boldsymbol{x}_{n-1} + \widetilde{\boldsymbol{B}}_n(\widetilde{\boldsymbol{A}}\boldsymbol{e}_{n-1} + \boldsymbol{\nu}_n) - \widetilde{\boldsymbol{K}}\boldsymbol{\omega}_n. \end{aligned} \quad (10)$$

where $\widetilde{\boldsymbol{B}}_n = \boldsymbol{I} - \widetilde{\boldsymbol{K}}\boldsymbol{H}_n$. This defines a new state space model where the noise terms are white Gaussian (but note that the measurement noise $\boldsymbol{\omega}_n$ and state evolution noise $\widetilde{\boldsymbol{B}}_n \boldsymbol{\nu}_n - \widetilde{\boldsymbol{K}}\boldsymbol{\omega}_n$ are dependent) and the input is $\big[\widetilde{\boldsymbol{B}}_n(\boldsymbol{A} - \widetilde{\boldsymbol{A}})\big]\boldsymbol{x}_{n-1}$.

This error term may be studied in terms of its mean and covariance. The mean is given by

$$\mathbb{E}[\boldsymbol{e}_n] = \widetilde{\boldsymbol{B}}_n(\boldsymbol{A} - \widetilde{\boldsymbol{A}})\,\mathbb{E}[\boldsymbol{x}_{n-1}] + \widetilde{\boldsymbol{B}}_n \widetilde{\boldsymbol{A}}\,\mathbb{E}[\boldsymbol{e}_{n-1}]. \quad (11)$$

To calculate $\mathbb{C}(\boldsymbol{e}_n)$, we need to first calculate the following covariance terms:

$$\mathbb{C}(\boldsymbol{e}_0) = \mathbb{C}(\boldsymbol{x}_0 - \boldsymbol{x}_{0|0}) = \boldsymbol{P}_0, \quad \mathbb{C}(\boldsymbol{x}_0) = \boldsymbol{P}_0,$$

$$\mathbb{C}(\boldsymbol{x}_n) = \mathbb{C}(\boldsymbol{A}\boldsymbol{x}_{n-1} + \boldsymbol{\nu}_n) = \boldsymbol{A}\mathbb{C}(\boldsymbol{x}_{n-1})(\boldsymbol{A})^T + \boldsymbol{Q}_n,$$

$$\mathbb{C}(\boldsymbol{e}_n, \boldsymbol{x}_n) = \mathbb{C}(\boldsymbol{x}_n - \widetilde{\boldsymbol{x}}_{n|n}, \boldsymbol{x}_n) = \mathbb{C}(\boldsymbol{x}_n) - \mathbb{C}(\widetilde{\boldsymbol{x}}_{n|n}, \boldsymbol{x}_n).$$

We denote $\mathbb{C}(\widetilde{\boldsymbol{x}}_{n|n}, \boldsymbol{x}_n)$ by $\boldsymbol{u}_n$ and let $\widetilde{\boldsymbol{x}}_{n|n-1} = \widetilde{\boldsymbol{A}}\boldsymbol{x}_{n-1|n-1}$, and then obtain the following recursion,

$$\begin{aligned} \boldsymbol{u}_n &= \mathbb{C}(\widetilde{\boldsymbol{A}}\boldsymbol{x}_{n-1|n-1} + \widetilde{\boldsymbol{K}}(\boldsymbol{y}_n - \boldsymbol{H}_n \widetilde{\boldsymbol{x}}_{n|n-1}), \boldsymbol{A}\boldsymbol{x}_{n-1} + \boldsymbol{\nu}_n) \\ &= \widetilde{\boldsymbol{B}}_n \widetilde{\boldsymbol{A}}\boldsymbol{u}_{n-1}\boldsymbol{A}^T + \widetilde{\boldsymbol{K}}\boldsymbol{H}_n \boldsymbol{A}\mathbb{C}(\boldsymbol{x}_{n-1})\boldsymbol{A}^T + \widetilde{\boldsymbol{K}}\boldsymbol{H}_n \boldsymbol{Q}_n, \end{aligned}$$

$$\begin{aligned} \widetilde{\boldsymbol{K}} &= \big[\widetilde{\boldsymbol{A}}\boldsymbol{P}_{n-1|n-1}(\widetilde{\boldsymbol{A}})^T + \widetilde{\boldsymbol{Q}}\big] \\ &\quad \times \boldsymbol{H}_n^T \big(\boldsymbol{H}_n \big[\widetilde{\boldsymbol{A}}\boldsymbol{P}_{n-1|n-1}(\widetilde{\boldsymbol{A}})^T + \widetilde{\boldsymbol{Q}}\big]\boldsymbol{H}_n^T + \boldsymbol{R}_n\big)^{-1}, \end{aligned}$$

with $\boldsymbol{u}_0 = \mathbb{C}(\boldsymbol{x}_{0|0},\, \boldsymbol{x}_{0|0} + \boldsymbol{\nu}_0) = \mathbf{0}$. Therefore, using (10) and calculating the covariance, we have

$$\mathbb{C}(\boldsymbol{e}_n) = \widetilde{\boldsymbol{K}}\boldsymbol{R}_n(\widetilde{\boldsymbol{K}})^T + \boldsymbol{J}_n\mathbb{C}(\boldsymbol{x}_{n-1})\boldsymbol{J}_n^T + \widetilde{\boldsymbol{B}}_n\boldsymbol{Q}_n\widetilde{\boldsymbol{B}}_n^T + \widetilde{\boldsymbol{B}}_n\widetilde{\boldsymbol{A}}\mathbb{C}(\boldsymbol{e}_{n-1})(\widetilde{\boldsymbol{A}})^T\widetilde{\boldsymbol{B}}_n^T + \boldsymbol{J}_n\mathbb{C}(\boldsymbol{x}_{n-1},\, e_{n-1})\widetilde{\boldsymbol{B}}_n^T + \widetilde{\boldsymbol{B}}_n\mathbb{C}(\boldsymbol{x}_{n-1},\, e_{n-1})\boldsymbol{J}_n^T. \tag{12}$$

where $\boldsymbol{J}_n = (\boldsymbol{I} - \widetilde{\boldsymbol{K}}\boldsymbol{H}_n)(\boldsymbol{A} - \widetilde{\boldsymbol{A}})$. All the above variables can be calculated recursively and exactly, given the prior statistics $\boldsymbol{x}_{n-1|n-1}$ and $\boldsymbol{P}_{n-1|n-1}$.

B. Single-Mode Kalman Filter Error in an SLDS

In this section, an arbitrary single-mode KF with parameters $\boldsymbol{A}, \boldsymbol{Q}$ is applied to an SLDS, and the MSE is calculated. Let l_n refer to a trajectory from the set of all possible r^n trajectories of discrete modes that may occur, where r is the number of possible modes to occur at each time step and n is the time step. Also, let $\boldsymbol{e}_n^{l_n}$ be the conditional error of the KF with trajectory l_n, and note that its mean and covariance can be calculated based on Section IV-A recursively given the trajectory. Assuming $l_n = [l_{n-1}, i]$ s.t. $i \in \{1, \ldots, r\}$ and $n > 1$, the error at each time is given by:

$$\boldsymbol{e}_n^{l_n} = \big[(\boldsymbol{I} - \boldsymbol{K}_n\boldsymbol{H}_n)(\boldsymbol{A}(i) - \boldsymbol{A})\boldsymbol{x}_{n-1}^{l_{n-1}} - \boldsymbol{K}_n\boldsymbol{\omega}_n + (\boldsymbol{I} - \boldsymbol{K}_n\boldsymbol{H}_n)(\boldsymbol{A}\boldsymbol{e}_{n-1}^{l_{n-1}} + \boldsymbol{\nu}_n)\big], \tag{13}$$

$$\boldsymbol{e}_n = \sum_{l_n} \delta_{l_n}\boldsymbol{e}_n^{l_n}, \qquad \mathbb{E}[\boldsymbol{e}_n] = \sum_{l_n} \pi_{l_n}\,\mathbb{E}[\boldsymbol{e}_n^{l_n}], \tag{14}$$

where $\boldsymbol{x}_{n-1}^{l_{n-1}}$ and $\boldsymbol{e}_{n-1}^{l_{n-1}}$ refer to the ground truth state variable and error for trajectory l_{n-1}, $\boldsymbol{K}_n$ is the Kalman gain for the single-mode KF at time n, δ_{l_n} equals one when trajectory l_n occurs and is zero otherwise, and π_{l_n} is the probability of trajectory l_n and can be calculated based on the given SLDS transition probabilities and priors. Similarly, we compute the covariance $\mathbb{C}(\boldsymbol{e}_n) = \sum_{l_n}\pi_{l_n}\,\mathbb{E}[\boldsymbol{e}_n^{l_n}(\boldsymbol{e}_n^{l_n})^T] - \mathbb{E}[\boldsymbol{e}_n]\,\mathbb{E}[\boldsymbol{e}_n]^T$. This formulation can be used to calculate the performance of an arbitrary single-mode KF in an SLDS.

C. Switching Kalman Filter Error in SLDS

We now calculate the MSE when a SKF algorithm is applied to an SLDS. Let l_n and q_n refer to the trajectory that occurs (the true trajectory) and is detected using the SKF, respectively, each taking values in the set of all possible r^n trajectories of length n such that $l_n = [l_{n-1}, i]$, $q_n = [q_{n-1}, j]$, where $i, j \in \{1, \ldots, r\}$. Also, let $\boldsymbol{e}_n^{(l_n;q_n)}$ be the conditional error based on these trajectories, which can be calculated recursively based on results from Section IV-A given the trajectories l_n and q_n:

$$\boldsymbol{e}_n^{(l_n;q_n)} = (\boldsymbol{I} - \boldsymbol{K}_n^{[q_{n-1},j]}\boldsymbol{H}_n)\big[(\boldsymbol{A}(i) - \boldsymbol{A}(j))\boldsymbol{x}_{n-1}^{l_{n-1}} + \boldsymbol{\nu}_n + \boldsymbol{A}(j)\boldsymbol{e}_{n-1}^{(l_{n-1};q_{n-1})}\big] - \boldsymbol{K}_n^{[q_{n-1},j]}\boldsymbol{\omega}_n, \tag{15}$$

where $\boldsymbol{x}_{n-1}^{l_{n-1}}$ and $\boldsymbol{e}_{n-1}^{(l_{n-1};q_{n-1})}$ refer to the ground truth state and error when trajectory l_{n-1} occurs and trajectory q_{n-1} is detected, $\boldsymbol{K}_n^{[q_{n-1},j]}$ is the KF gain assuming detected trajectory of $[q_{n-1}, j]$ at time n. The error then may be written as

$$\boldsymbol{e}_n = \sum_{l_n}\sum_{q_n} \delta_{l_n;q_n}\boldsymbol{e}_n^{(l_n;q_n)}, \tag{16}$$

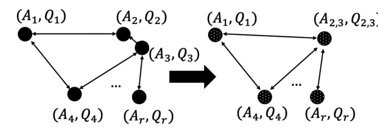


Fig. 1. Using the quantified error in this paper, SLDS model with r modes with evolution matrix and covariance matrix pair $(A_i, Q_i), \forall i \in \{1, \ldots, r\}$ may be reduced to SLDS model with $(r-1)$ modes, where $(A_{2,3}, Q_{2,3})$ refers to the evolution matrix and covariance matrix pair of the mode that is obtained from merging modes 2,3.

where $\delta_{l_n;q_n}$ equals one when l_n occurs and q_n is detected, and 0 otherwise. The mean of this random process is calculated as $\mathbb{E}[\boldsymbol{e}_n] = \sum_{l_n}\sum_{q_n}\pi_{l_n,q_n}\,\mathbb{E}[\boldsymbol{e}_n^{(l_n;q_n)}]$ where π_{l_n,q_n} is the probability that trajectory l_n occurs and trajectory q_n is detected, which may also be calculated recursively. Similarly, covariance of the error is given by: $\mathbb{C}(\boldsymbol{e}_n) = \sum_{l_n}\sum_{q_n}\pi_{l_n,q_n}\,\mathbb{E}[\boldsymbol{e}_n^{(l_n;q_n)}(\boldsymbol{e}_n^{(l_n;q_n)})^T] - \mathbb{E}[\boldsymbol{e}_n]\,\mathbb{E}[\boldsymbol{e}_n]^T$.

V. Discussion

Some challenges in calculating the derived statistics in practice are discussed below.

1) Calculating the error statistics in section IV-C considering the random variable $\boldsymbol{K}_n^{[q_{n-1},j]}$ requires computing and adding an exponentially increasing number of terms (equivalent to running a Monte Carlo simulation). As an approximation, we assume it to be a deterministic variable approximated by $\boldsymbol{K}_n^{[q_{n-1},j]} \sim \boldsymbol{K}_n^{[j]_{1\times n}}$ (where $[j]_{1\times n}$ is a $1 \times n$ vector and all its elements equal j), or calculating $\boldsymbol{K}_n$ given $\mathbb{E}_{[l_{n-1};q_{n-1}]}[P_{n-1|n-1}]$. Simulation results show that the statistics calculated using both approximations can closely mimic the variations of the estimation error.

2) Applying the formulation to multi-modal SLDS enables making the optimal decision on which modes to keep in an SKF framework based on the quantified errors, but at a huge computational cost due to having r^n trajectories at time n. A sub-optimal, feasible solution to the problem calculates the marginal transition probability between each pair of modes and applies the formulation to each pair. In this framework, the collection of switching dynamic systems is represented using a graph network where each node is a dynamic system mode, as shown in Fig. 1. For each pair of nodes: if using an SKF for the pair does not provide a significant improvement over a single-mode KF using the proposed formulation, they are merged, and if not, both nodes are kept. As a result, a multi-modal problem is divided into multiple bi-modal problems.

3) The recursive calculation of (13) and (15) for all possible trajectories is practically infeasible as n becomes large. We consider solutions for the following two scenarios:

(3.i) If the transition probabilities between modes are equal, the detection rate at each time step for each mode depends only on the mode that occurs at that time step, and the Kalman gains are calculated using the considerations discussed in (1), then the bias at time n based on (16) for a multi-modal system can be derived recursively using the same notation as in Section IV-C:

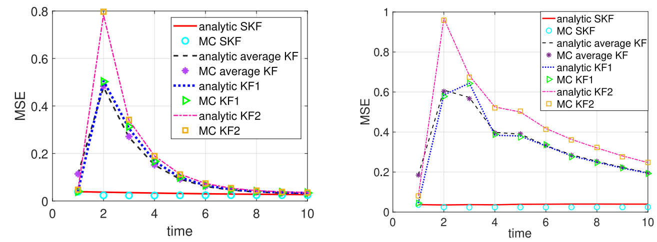
$$\boldsymbol{e}_n = \sum_{[l_{n-1},i]}\sum_{[q_{n-1},j]} \delta_{[l_{n-1},i];[q_{n-1},j]}\boldsymbol{e}_n^{([l_{n-1},i];[q_{n-1},j])}$$
$$= \sum_{i,j\in\{1,\ldots,r\}}\sum_{l_{n-1}}\sum_{q_{n-1}} \delta_{i|l_{n-1}}\delta_{[l_{n-1};q_{n-1}]}\delta_{j|i,[l_{n-1};q_{n-1}]}$$

$$
\begin{aligned}
&\times \boldsymbol{e}_n^{([l_{n-1},i];[q_{n-1},j])}\\
&\overset{(a)}{=} \sum_{i,j\in\{1,\dots,r\}} \sum_{l_{n-1}} \sum_{q_{n-1}} \delta_{i|l_{n-1}} \delta_{[l_{n-1};q_{n-1}]} \delta_{j|i,[l_{n-1};q_{n-1}]}\\
&\quad \times (\boldsymbol{I} - \boldsymbol{K}_n^{[q_{n-1},j]} \boldsymbol{H}_n)\big[(\boldsymbol{A}(i) - \boldsymbol{A}(j))\boldsymbol{x}_{n-1}^{l_{n-1}} + \boldsymbol{\nu}_n\\
&\quad + \boldsymbol{A}(j)\boldsymbol{e}_{n-1}^{(l_{n-1};q_{n-1})}\big] - \boldsymbol{K}_n^{[q_{n-1},j]}\boldsymbol{\omega}_n,\\
&\overset{(b)}{=} \sum_{i,j\in\{1,\dots,r\}} \delta_i^* \delta_{j|i}\\
&\quad \times (\boldsymbol{I} - \boldsymbol{K}_n^{[j]_{1\times n}} \boldsymbol{H}_n)\Big[(\boldsymbol{A}(i) - \boldsymbol{A}(j)) \sum_{l_{n-1}} \delta_{l_{n-1}} \boldsymbol{x}_{n-1}^{l_{n-1}} + \boldsymbol{\nu}_n\\
&\quad + \boldsymbol{A}(j) \sum_{l_{n-1}} \sum_{q_{n-1}} \delta_{[l_{n-1};l_{n-1}]} \boldsymbol{e}_{n-1}^{(l_{n-1};q_{n-1})}\Big] - \boldsymbol{K}_n^{[j]_{1\times n}}\boldsymbol{\omega}_n.
\end{aligned}
\tag{17}
$$

where $\delta_{i|\kappa}, \delta_{j|\kappa}$ equal one if mode i occurs and if mode j is detected given the condition κ occurs, respectively, and they equal zero otherwise. The equality (a) is concluded based on the defined state-space model in Section II. The equality (b) holds if we assume $\delta_{i|l_{n-1}} = \delta_i^* \forall i \in \{1,\dots,r\}$ such that $p(\delta_i^* = 1) = \frac{1}{r} \forall i \in \{1,\dots,r\}$, which means that the occurrence of each mode at each time step is deterministic and equiprobable due to equal transition probabilities. Also, we assume $\delta_{j|i,[l_{n-1},q_{n-1}]} = \delta_{j|i}$, which means that detection rate of each mode at time n only depends on the mode that occurs at time n. Moreover, $\boldsymbol{K}_n^{[q_{n-1},j]} \approx \boldsymbol{K}_n^{[j]_{1\times n}}$, which means that the Kalman gain in the derivations is calculated considering the approximations discussed in (1). Once Eq. (17) is verified, it is obvious that knowledge of the mean and (cross-)covariance of $\boldsymbol{x}_{n-1} = \sum_{l_{n-1}} \delta_{l_{n-1}} \boldsymbol{x}_{n-1}^{l_{n-1}}$ and $\boldsymbol{e}_{n-1} = \sum_{l_{n-1}} \sum_{q_{n-1}} \delta_{[l_{n-1};l_{n-1}]} \boldsymbol{e}_{n-1}^{(l_{n-1};q_{n-1})}$, as well as the noise statistics, is sufficient to recursively calculate the MSE using the first and second order statistics derived in IV-C. The same reasoning also applies to the formulation of a single-mode KF in an SLDS, as a special case of this general scenario.

(3.ii) When the transition probabilities are not equal, the probability of each trajectory is a deterministic function of the transition matrix and the initial probabilities. To minimize the number of exponentially increasing trajectories that need to be considered for the exact calculation of MSE as in IV-C, one may approximate the additional error by keeping N_t trajectories with the largest probabilities and ignoring the rest. The calculated MSE using the N_t trajectories will be a biased estimate of the desired value depending on the number of considered trajectories, the initial prior of each mode, and the transition matrix between the modes. The number of trajectories required to keep for an approximation with bounded error increases as time increases and as the transition probabilities tend to equality.

Mode detection in an SKF has its largest error when all modes have the same occurrence probability. To verify this, let l_n be the random variable referring to different r^n trajectories that may have occurred by time n. If the probability of all possible trajectories is equal to $\frac{1}{r^n}$, the uncertainty of the random variable l_n is maximized, and therefore, its detection error is maximized [28]. Therefore, studying the performance of an SKF for an SLDS that satisfies the assumptions discussed above is a computationally feasible and robust metric for the purpose of comparison between the performance of an SKF and a single-mode Kalman filter.



(a) diagonal evolution model (b) random evolution model

Fig. 2. MSEs for different filtering schemes in Section VI for (a) a deterministic state evolution model and (b) a random evolution model.

VI. Numerical Results

A bi-modal, 4-dimensional state-space model is generated via Monte Carlo (MC) simulations, and the error statistics calculated using analytic derivations are compared with MC simulations to verify the proposed formulation. An approximation of the detection rate is assumed to be given in these simulations and the SKF gain at time n when mode j is detected is approximated by the gain of the KF corresponding to detected mode trajectory $[j]_{1\times n}$ at time n (the ground truth trajectory can be $[q_{n-1}, j]$ where $q_{n-1} \in \{1,\dots,r\}^{n-1}$). Let the state-space equations be as presented in (1)–(2), where $\boldsymbol{A}^{(1)} = 0.9\boldsymbol{I}_{4\times4}$, $\boldsymbol{A}^{(2)} = -0.46\boldsymbol{I}_{4\times4}$ in Fig. 2(a), whereas in Fig. 2(b) $\boldsymbol{A}^{(1)}$ and $\boldsymbol{A}^{(2)}$ are generated randomly. For both cases, $\boldsymbol{Q}^{(1)} = \boldsymbol{Q}^{(2)} = 0.01\boldsymbol{I}_{4\times4}$, $\boldsymbol{R} = 0.01\boldsymbol{I}_{4\times4}$, $\boldsymbol{x}_0 = [1]_{1\times4}^T$, $\boldsymbol{P}_0 = \boldsymbol{I}_{4\times4}$, and $\boldsymbol{H} = \boldsymbol{I}_{4\times4}$ with the transition matrix $[0.5, 0.5; 0.5, 0.5]$ and the prior $[0.5, 0.5]$.

The single-mode KFs using models (1) and (2), as well as the "average KF" $\boldsymbol{A}_n = \pi_1\boldsymbol{A}^{(1)} + \pi_2\boldsymbol{A}^{(2)}$, where π_i is the probability of mode i at each time n (calculated based on the prior and transition probabilities of the discrete mode), are used for estimation. Intuitively, if the switching distributions are close to each other (e.g., in the KL divergence sense), the average filter's estimates are close to the optimal solution. Alternatively, if the distributions are far from each other, the average KF's estimates are poor.

Using 20 k MC simulations, $\mathbb{E}[\boldsymbol{e}_n]$ and $\mathbb{C}[\boldsymbol{e}_n]$ are computed, and the MSE is calculated as $\mathbb{E}[\boldsymbol{e}_n]^T\,\mathbb{E}[\boldsymbol{e}_n] + \mathrm{Tr}[\mathbb{C}(\boldsymbol{e}_n)]$ (Tr is the matrix trace) for the SKF, average KF, and mismatched single-mode KFs (KF1 and KF2). Fig. 2(a)–2(b) show the consistency between the analytic and MC-calculated MSEs for both presented case studies. Superior performance of the SKF over the single-mode KFs is clearly observed in both cases. Simulations for state dimensions $d > 4$ and different system dynamics also produced consistent results verifying the MSE derivations.

VII. Conclusion and Future Work

The MSE performance of the SKF and an arbitrary single-mode mismatched KF were derived analytically and compared using a recursive formulation. This formulation may be used to decide which filter to run operationally for a specific SLDS. This work is a step towards automating the filter decision process for a specific SLDS scenario by evaluating the accuracy versus computation trade-off.

REFERENCES

[1] R. A. Frazin, M. D. Butala, A. Kemball, and F. Kamalabadi, "Time-dependent reconstruction of nonstationary objects with tomographic or interferometric measurements," *Astrophysical J.*, vol. 635, no. 2, pp. L197–L200, Dec. 2005.

[2] M. D. Butala, F. Kamalabadi, R. A. Frazin, and Y. Chen, "Dynamic tomographic imaging of the solar corona," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 5, pp. 755–766, Oct. 2008.

[3] G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *J. Geophysical Res.: Oceans*, vol. 99, no. C5, pp. 10143–10162, 1994.

[4] K. Murphy, "Modelling gene expression data using dynamic Bayesian networks," Uni. California Berkeley, Berkeley, CA, USA., Tech. Rep., 2006.

[5] Y. Zeng S. Wu, *State-Space Models: Applications in Economics and Finance*. Berlin, Germany: Springer, 2013.

[6] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. 2nd ed., Ser. Springer Ser. Statist. New York, NY, USA: Springer, 1985. [Online]. Available: https://cds.cern.ch/record/1327974

[7] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960. [Online]. Available: https://doi.org/10.1115/1.3662552

[8] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[9] K. P. Murphy and S. Russell, "Dynamic Bayesian networks: Representation, inference and learning," Ph.D. dissertation, Uni. of California, Berkeley CA, USA, 2002.

[10] K. P. Murphy, "Switching Kalman filters," Uni. California, Berkeley, CA, USA Tech. Rep., 1998.

[11] R. J. Elliott, F. Dufour, and W. P. Malcolm, "State and mode estimation for discrete-time jump Markov systems," *SIAM J. Control Optim.*, vol. 44, no. 3, pp. 1081–1104, 2005. [Online]. Available: https://doi.org/10.1137/S0363012904442628

[12] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: A survey," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 1, pp. 103–123, Jan. 1998.

[13] P. Karimi, M. Butala, and F. Kamalabadi, "Efficient model selection in switching linear dynamic systems," 2021, *arXiv:2012.04543*.

[14] M. M. H. El Ayadi, M. S. Kamel, and F. Karray, "Toward a tight upper bound for the error probability of the binary Gaussian classification problem," *Pattern Recognit.*, vol. 41, no. 6, pp. 2120–2132, Jun. 2008. [Online]. Available: https://doi.org/10.1016/j.patcog.2007.10.028

[15] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, vol. 15, no. 1, pp. 52–60, Feb. 1967.

[16] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, vol. COM-15, no. 1, pp. 52–60, Feb. 1967.

[17] F. Nielsen, "Generalized Bhattacharyya and Chernoff upper bounds on Bayes error using quasi-arithmetic means," *Pattern Recognit. Lett.*, vol. 42, pp. 25–34, Jun 2014.

[18] B. Anderson and J. Moore, *Optimal Filtering*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.

[19] Q. Ge, T. Shao, Z. Duan, and C. Wen, "Performance analysis of the Kalman filter with mismatched noise covariances," *IEEE Trans. Autom. Control*, vol. 61, no. 12, pp. 4014–4019, Dec. 2016.

[20] S. Sangsuk-Iam and T. E. Bullock, "Analysis of discrete-time Kalman filtering under incorrect noise covariances," *IEEE Trans. Autom. Control*, vol. 35, no. 12, pp. 1304–1309, Dec. 1990.

[21] T. Nishimura, "On the a priori information in sequential estimation problems," *IEEE Trans. Autom. Control*, vol. AC-11, no. 2, pp. 197–204, Apr. 1966.

[22] P. D. Hanlon and P. S. Maybeck, "Characterization of Kalman filter residuals in the presence of mismodeling," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 1, pp. 114–131, Jan. 2000.

[23] I. Hwang, H. Balakrishnan, and C. Tomlin, "Performance analysis of hybrid estimation algorithms," in *Proc. 42nd IEEE Int. Conf. Decis. Control*, 2003, pp. 5353–5359.

[24] W. Zhang and B. Natarajan, "On the statistical convergence of bias in mode-based Kalman filter for switched systems," *EURASIP J. Adv. Signal Process.*, vol. 2018, no. 1, Nov. 2018, Art. no. 71.

[25] W. Zhang and B. Natarajan "Quantifying the bias dynamics in a mode-based Kalman filter for stochastic hybrid systems," in *Proc. Annu. Amer. Control Conf.*, 2018, pp. 5849–5856.

[26] W. Zhang and B. Natarajan, "Impact of time correlated mode mismatch on Markov jump linear system state estimation," *IEEE Control Syst. Lett.*, vol. 2, no. 3, pp. 489–494, Jul. 2018.

[27] W. Zhang and B. Natarajan, "Bias analysis in Kalman filter with correlated mode mismatch errors," *Signal Process.*, vol. 154, pp. 232–237, 2018.

[28] *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2005, ch. 7, pp. 183–241.