2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Practical Stability of Approximating Discrete-Time Filters with Respect to Model Mismatch Using Relative Entropy Concepts

Onvaree Techakesari, Jason J. Ford, and Dragan Nešić

***Abstract*— This paper establishes practical stability results for an important range of approximate discrete-time filtering problems involving mismatch between the true system and the approximating filter model. Using local consistency assumption, the practical stability established is in the sense of an asymptotic bound on the amount of bias introduced by the model approximation. Significantly, these practical stability results do not require the approximating model to be of the same model type as the true system. Our analysis applies to a wide of range of estimation problems and justifies the common practice of approximating intractable infinite dimensional nonlinear filters by simpler computationally tractable filters.**

I. INTRODUCTION

Many filtering problems involve estimation of system quantities from noisy measurements in situations where the exact (or true) model of the system is either unknown or is more complicated than can be handled using standard techniques. In these types of problems, approximate filters are often designed on the basis of an approximating system that reasonably represents the true dynamics. For example, using this informal idea, nonlinear dynamics are sometimes approximated by hidden Markov model (HMM) or linear dynamics. These two type of approximating models lead to the application of HMM filters or Kalman filters in a wide range of signal and image processing applications [1]–[3].

Despite the successful application of approximate filters in a large number of applications, conditions that ensure reasonable filter behaviour have not yet been completely established in many situations. Specifically, only a small number of stability results in situations involving model errors have been presented. These include stability with respect to model mismatch for Kalman filters [4], [5] and particle filters [6]–[8]. Further, some convergence and stochastic stability results for extended Kalman filters are presented in [9], [10]. Moreover, practical stability of general nonlinear observer in a deterministic setting is presented in [17]

In recent work, relative entropy concepts have been exploited to determine the similarity of different model descriptions [3], [14], [15]. Specifically, in [14], it has been shown that the relative entropy rate (RER) between the joint state and measurement processes of two HMMs is related to the probabilistic distance (or relative likelihood) between the HMMs. This relationship suggests a connection between RER and the filter performance, and importantly, allows the use of RER in the design of HMMs to approximate uncertain nonlinear dynamics [3]. These results motivate consideration of relative entropy concepts in general filtering problems.

In this paper, we extend the deterministic practical stability result in [17] to a stochastic setting. We establish practical stability of general approximating filters with respect to modelling errors under some mild assumptions, including one-step (or local) consistency and forgetting properties. Moreover, we show how relative entropy concepts can be exploited to establish the required local consistency conditions without reference to specific property of the filtering equations. The results of this paper are established using the local consistency techniques that have previously been used to establish semi-global practical stability results for discretisation of nonlinear controllers [16].

This paper is structured as follows: In Section II, we introduce our nominal dynamics, our information state concepts, our modelling approximations, and the concepts of relative entropy rate. In Section III, we establish some important consistency results and the main practical stability results of this paper are established in Section IV. In Section V, we illustrate our results in the case of HMM approximation. Some conclusions are presented in Section VI.

II. PROMBLEM FORMULATION

A. Dynamics

For the time step $k > 0$, we will consider the following state process $x_k \in \mathbb{R}^n$ and measurement process $y_k \in \mathbb{R}^m$,

$$\begin{aligned} x_k &= f(x_{k-1}) + v_k \\ y_k &= c(x_k) + w_k \end{aligned} \tag{1}$$

where x_0 has *a priori* distribution σ_0, $f(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$, and $c(\cdot) : \mathbb{R}^n \to \mathbb{R}^m$. Here, $v_k \in \mathbb{R}^n$ and $w_k \in \mathbb{R}^m$ are sequences of independent and identically distributed *i.i.d.* random variables with densities $\phi_v(\cdot)$ and $\phi_w(\cdot)$, respectively. The random variables v_k, w_k, and x_0 are assumed to be mutually independent for all k. We will use the shorthand $x_{[\ell,m]}$ to denote the state sequences $\{x_\ell, \ldots, x_m\}$. We likewise define $y_{[\ell,m]}$.

Throughout the rest of this paper, we will consider all processes to be defined on a probability space $(\Omega, \mathcal{F}, P)$ where Ω consists of all infinite sequences $\{x_0, \ldots, x_k, \ldots; y_1, \ldots, y_k, \ldots\}$ (with elements $\omega \in \Omega$), $\mathcal{F}$ is a σ-algebra generated by these sequences, and P is a probability measure given by Kolmogorov extension theorem applied to these sequences [18]. Finally, $\mathcal{Y}_{[1,k]}$ denotes the

O. Techakesari and J.J. Ford are with the School of Engineering, Queensland University of Technology, Brisbane QLD 4001, Australia. o.techakesari@qut.edu.au, j2.ford@qut.edu.au
D. Nešić is with the Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville VIC 3010, Australia, dnesic@unimelb.edu.au

complete filtration generated by the sequence $y_{[1,k]}$, see [11, p. 18].

In filtering, we are often interested in the conditional mean estimate of x_k given the measurements $y_{[1,k]}$ and the *a priori* distribution σ_0, which can be defined, when it exists, as:

$$\hat{x}^e_{k|[1,k],\sigma_0} \triangleq E\left[x_k \middle| \mathcal{Y}_{[1,k]}, \sigma_0\right] \tag{2}$$

for all $k > 0$, where $E[\cdot]$ denotes the expectation operation corresponding to P. Similarly, $\hat{x}^e_{k|[\ell,m],\sigma_{\ell-1}}$ will denote the conditional mean estimate at time k, given the measurement sequences $y_{[\ell,m]}$ and the distribution $\sigma_{\ell-1}$ of $x_{\ell-1}$ at time $\ell - 1$.

Unfortunately, in many situations, it may not be possible to implement a filter that produces $\hat{x}^e_{k|[1,k],\sigma_0}$ (for example, such a filter may be computationally intractable). In this paper, we are interested in the performance of sub-optimal or approximate filters that provide approximate estimates for our system state, x_k.

B. Normalised Information State

We now introduce some information state concepts that describe our estimation operations. Consider the space $L^\infty(\mathbb{R}^n)$ which includes $L^1(\mathbb{R}^n)$; see [19] for an introduction into vector space concepts. We will introduce the $\langle\cdot,\cdot\rangle$ notation to denote the operation of $\xi(\cdot) \in L^1(\mathbb{R}^n)$ and $\gamma(\cdot) \in L^\infty(\mathbb{R}^n)$ as $\langle \xi, \gamma\rangle \triangleq \int_{\mathbb{R}^n} \xi(x)\gamma(x)dx$. We will also introduce the L_1 norm on information state [19]:

$$||\xi(\cdot)||_1 \triangleq \int_{x\in\mathbb{R}^n} |\xi(x)|dx. \tag{3}$$

Let $\bar{L}^1(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$ denote functions in $L^1(\mathbb{R}^n)$ that have L_1 norm equal to 1 in that $\bar{L}^1(\mathbb{R}^n) \triangleq \left\{\xi(\cdot) : \xi(\cdot) \in L^1(\mathbb{R}^n) \text{ and } ||\xi(\cdot)||_1 = 1\right\}$. We can now define a normalised information state process $\sigma^e_k(\cdot) \in \bar{L}^1(\mathbb{R}^n) : \mathbb{R}^n \to \mathbb{R}$, based on the true model, by

$$\langle \sigma^e_k, \gamma\rangle = E\left[\gamma(x_k) \middle| \mathcal{Y}_{[1,k]}, \sigma_0\right] \tag{4}$$

for all $k > 0$, and all test functions $\gamma(\cdot) \in L^\infty(\mathbb{R}^n)$, where $\sigma_0 \in \bar{L}^1(\mathbb{R}^n)$ is the *a priori* distribution of x_0. This definition highlights that the normalised information state $\sigma^e_k(\cdot)$ can be interpreted as a conditional probability density function of x_k given measurement sequences $y_{[1,k]}$ and *a priori* distribution σ_0. In particular, when it exists, we can write our conditional mean estimate as

$$\hat{x}^e_{k|[1,k],\sigma_0} = \int_{x\in\mathbb{R}^n} \sigma^e_k(x)x dx. \tag{5}$$

We also consider an unnormalised information state $\sigma^{e|u}_k(\cdot) \in L^1(\mathbb{R}^n)$ which provides a method of calculating $\sigma^e_k(\cdot)$. For all $k > 0$, the unnormalised information state is given by [11, Ch.5]

$$\sigma^{e|u}_k(x) = \frac{\phi_w(y_k - c(x))}{\phi_w(y_k)} \int_{\mathbb{R}^n} \phi_v(x - f(z))\sigma^e_{k-1}(z)dz \tag{6}$$

for $x \in \mathbb{R}^n$, where $\sigma^e_0 \in L^1(\mathbb{R}^n) = \sigma_0$. The normalised information state $\sigma^e_k(\cdot) \in \bar{L}^1(\mathbb{R}^n)$ can then be written as

$$\sigma^e_k(\cdot) = N^{-1}_k \sigma^{e|u}_k(\cdot) \tag{7}$$

where $N_k = \left|\left|\sigma^{e|u}(\cdot)\right|\right|_1$ is a normalisation factor. We highlight that (6) and (7) together evolve $\sigma^e_{k-1}(\cdot) \in \bar{L}^1(\mathbb{R}^n)$ to produce $\sigma^e_k(\cdot) \in \bar{L}^1(\mathbb{R}^n)$. Here, when required to highlight the initial condition, we will write $\sigma^e_{k|[1,k],\sigma_0}(\cdot)$ to denote the normalised information state $\sigma^e_k(\cdot)$ after evolution by measurements $y_{[1,k]}$ from initial distribution σ_0 at time $k = 0$ (and sometimes further shortened to $\sigma^e_k(\sigma_0)$, especially when used in sub-scripts of other quantities). Similarly, $\sigma^e_{k|[\ell+1,k],\sigma^e_\ell}(\cdot)$ will denote $\sigma^e_k(\cdot)$ after evolution by measurements $y_{[\ell+1,k]}$ from distribution $\sigma^e_\ell(\cdot)$ at time $k = \ell$. Importantly, the distributive nature of the information state recursions means that $\sigma^e_{k|[\ell+1,k],\sigma^e_\ell(\sigma_0)}(\cdot) = \sigma^e_{k|[1,k],\sigma_0}(\cdot)$. We also define $\sigma^e_{0|[1,0],\sigma_0} \triangleq \sigma_0$ for all $\sigma_0 \in \bar{L}^1(\mathbb{R}^n)$. We highlight that, although not explicitly shown in our notation, all these information state quantities are also $\mathcal{Y}_{[1,k]}$-measurable random variables.

C. Parameterised Class of Approximating Models

Let $h > 0$ parameterises a class of approximating models (for example, h might be a spatial discretisation size). For each h, let us consider the following approximating model of x_k and y_k (for time step $k > 0$):

$$\begin{aligned} x_k &= f^h(x_{k-1}) + v^h_k \\ y_k &= c^h(x_k) + w^h_k \end{aligned} \tag{8}$$

where x_0 has *a priori* distribution σ^h_0, $f^h(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$, and $c^h(\cdot) : \mathbb{R}^n \to \mathbb{R}^m$. Here, $v^h_k \in \mathbb{R}^n$ and $w^h_k \in \mathbb{R}^m$ are *i.i.d.* random variables with densities $\phi^h_v(\cdot)$ and $\phi^h_w(\cdot)$, respectively, and v^h_k, w^h_k, and x_0 are assumed to be mutually independent. Corresponding to each approximating model, we introduce a new probability measure P^h which allows us to relate the true model and these approximating models on the common measure space $(\Omega, \mathcal{F})$; such a measure can be defined through the Kolmogorov extension theorem [18].

For a given $h > 0$, we can also define the conditional mean estimate associated with the approximate model as:

$$\hat{x}^h_{k|[1,k],\sigma^h_0} \triangleq E^h\left[x_k \left| \mathcal{Y}_{[1,k]}, \sigma^h_0\right.\right], \tag{9}$$

where $E^h[\cdot]$ denotes the expectation operation defined by measure P^h.

Similar to the true model, we also define a normalised information state process $\sigma^h_k(\cdot) \in \bar{L}^1(\mathbb{R}^n) : \mathbb{R}^n \to \mathbb{R}$, for our h-class of models, as

$$\langle \sigma^h_k, \gamma\rangle = E^h\left[\gamma(x_k) \left| \mathcal{Y}_{[1,k]}, \sigma^h_0\right.\right] \tag{10}$$

for all $k > 0$, all $h > 0$ and all test functions $\gamma(\cdot) \in L^\infty(\mathbb{R}^n)$ where $\sigma^h_0 \in \bar{L}^1(\mathbb{R}^n)$ is the *a priori* distribution of x_0. Furthermore, we can define a recursion for the unnormalised information state process $\sigma^{h|u}_k(\cdot) \in L^1(\mathbb{R}^n)$ as $\sigma^{h|u}_k(x) = \frac{\phi^h_w(y_k - c^h(x))}{\phi^h_w(y_k)} \int_{\mathbb{R}^n} \phi^h_v(x - f^h(z))\sigma^h_{k-1}(z)dz$ so that a normalised information state $\sigma^h_k(\cdot) \in \bar{L}^1(\mathbb{R}^n)$ can be written as $\sigma^h_k(\cdot) = \bar{N}^{-1}_k \sigma^{h|u}_k(\cdot)$ where $\bar{N}^{-1}_k = \left|\left|\sigma^{h|u}_k(\cdot)\right|\right|_1$. Again, we highlight that $\sigma^h_{k-1}(\cdot) \in \bar{L}^1(\mathbb{R}^n)$ evolves to $\sigma^h_k(\cdot) \in \bar{L}^1(\mathbb{R}^n)$. As above, we also write $\sigma^h_{k|[\ell+1,k],\sigma^h_\ell(\sigma^h_0)}(\cdot) = \sigma^h_{k|[1,k],\sigma^h_0}(\cdot)$, and we define $\sigma^h_{0|[1,0],\sigma_0} \triangleq \sigma_0$ for all $\sigma_0 \in \bar{L}^1(\mathbb{R}^n)$.

We are interested in the situations where the quality of the approximation improves as $h \to 0$ (the meaning of this asymptotic behaviour will be discussed in more detail later). Before this discussion, we introduce some relative entropy concepts.

D. The Relative Entropy between Models

Consider two probability measures μ and ν on the measurable space $(\Omega, \mathcal{F})$. The relative entropy $\mathcal{D}(\mu \| \nu)$ of μ with respect to ν is defined as [20]

$$\mathcal{D}(\mu \| \nu) \triangleq \begin{cases} \int_\Omega \left(\log \frac{d\mu}{d\nu}\right) d\mu, & \text{if } \mu \ll \nu \text{ and} \\ & \left|\log\left(\frac{d\mu}{d\nu}\right)\right| \text{ is integrable} \\ +\infty, & \text{otherwise,} \end{cases} \tag{11}$$

where $(d\mu/d\nu)$ is the Radon-Nikodym derivative of μ with respect to ν. Here, we use $\mu \ll \nu$ to indicate that μ is absolutely continuous with respect to ν, in the sense that $\mu = 0$ whenever $\nu = 0$. The relative entropy $\mathcal{D}(\mu \| \nu)$ provides a pseudo-distance measure between μ and ν (not a true distance because it is non-symmetric and does not satisfy the triangle inequality).

In the following, we will consider the relative entropy between densities or information states (and this will be understood to mean the relative entropy between the measures corresponding to the densities, when such measures exist).

We now introduce some assumptions about our approximating models.

III. APPROXIMATING FILTER ASSUMPTIONS

A function $\psi(\cdot)$ is of class-$\mathscr{K}$ if it is continuous, strictly increasing and $\psi(0) = 0$. Moreover, function $\beta(\cdot,\cdot)$ is of class-$\mathscr{KL}$ if $\beta(\cdot,t)$ is of class-$\mathscr{K}$ for each $t \geq 0$ and $\beta(s,\cdot)$ is decreasing to zero for each $s > 0$, (see [23, Ch. 4] for descriptions of system stablity involving such functions).

We now introduce some assumptions that will be used to establish the important multi-step consistency of filters and our main practical stability result.

A1) Let $N \subset \mathbb{R}^+$ be a bounded set containing the origin. The class of approximating filters $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ is asymptotically stable with respect to initial conditions in N if, there exists a $H > 0$ and a $\beta(\cdot,\cdot) \in \mathscr{KL}$ such that, for all $h \in (0, H]$, all $||\sigma_0 - \bar{\sigma}_0|| \in N$ (with $\sigma_0, \bar{\sigma}_0 \in \bar{L}^1(\mathbb{R}^n)$), and all $k \geq 0$, we have that

$$\begin{aligned} &\left|\left|\sigma^h_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\bar{\sigma}_0}(\cdot)\right|\right|_1 \\ &\quad \leq \beta\left(||\sigma_0 - \bar{\sigma}_0||_1, k\right) \qquad P\text{-a.s..} \end{aligned} \tag{12}$$

A2) The class of approximating filters $\sigma^h_{k|[k],\sigma_{k-1}}(\cdot)$ is Lipschitz continuous with respect to prior information if there exists a $H > 0$ such that, for all $h \in (0, H]$, all prior information $\sigma_{k-1}, \bar{\sigma}_{k-1} \in \bar{L}^1(\mathbb{R}^n)$, and all $k > 0$, we have that

$$\begin{aligned} &\left|\left|\sigma^h_{k|[k],\sigma_{k-1}}(\cdot) - \sigma^h_{k|[k],\bar{\sigma}_{k-1}}(\cdot)\right|\right|_1 \\ &\quad \leq K\,||\sigma_{k-1}(\cdot) - \bar{\sigma}_{k-1}(\cdot)||_1 \qquad P\text{-a.s.} \end{aligned}$$

where $K > 0$ is a finite constant.

Remark 1: Assumption A1 is an abstract version of the asymptotic stability property with respect to initial conditions (or exponential forgeting of initial conditions) that is often encountered in discussion of filter behaviour (for example, see [4], [5], [22]). As an example, if $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ corresponds to a class of Kalman filters, then under controllability, observability and other mild conditions, exponential forgetting of covariance matrix and conditional mean estimate, with respect to initial conditions, can be shown [4], [5]. Hence, using the definition of the L_1 norm, and various algebraic manipulations, it can be shown that

$$\left|\left|\sigma^h_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\bar{\sigma}_0}(\cdot)\right|\right|_1 \leq \beta\left(||\sigma_0 - \bar{\sigma}_0||_1, k\right) P^h\text{-a.s.}$$

where $\beta(||\sigma_0 - \bar{\sigma}_0||_1, k) = \alpha_1 ||\sigma_0 - \bar{\sigma}_0||_1 e^{-\alpha_2 k}$ for some $\alpha_1 > 0$ and $\alpha_2 > 0$.

Remark 2: We highlight that asymptotic stability properties with respect to initial conditions (such as those that could be shown when using Kalman filter approximations) would generally be established P^h-a.s. (not P-a.s. as expressed in Assumption A1). However, the mild additional condition that $P^h \gg P$ can be used on results that hold P^h-a.s. to imply that they also hold P-a.s.. Note that $P^h \gg P$ seems to be a natural prerequisite for approximation.

Remark 3: In many situations, the bounded set N appeared in A1 includes all $\sigma_0, \bar{\sigma}_0 \in \bar{L}^1(\mathbb{R}^n)$ and hence, A1 will imply A2.

We will now introduce some important definitions.

Definition 3.1: The class of approximating filters $\sigma^h_{k|[k],\sigma_{k-1}}(\cdot)$ is said to be one-step or locally consistent with respect to the true filter $\sigma^e_{k|[k],\sigma_{k-1}}(\cdot)$ if, for each finite $\rho > 0$, there exists a $H > 0$ such that, for all $h \in (0, H]$, all initial conditions $\sigma_{k-1} \in \bar{L}^1(\mathbb{R}^n)$, and all $k > 0$, we have that

$$\left|\left|\sigma^e_{k|[k],\sigma_{k-1}}(\cdot) - \sigma^h_{k|[k],\sigma_{k-1}}(\cdot)\right|\right|_1 \leq \rho \qquad P\text{-a.s..} \tag{13}$$

Lemma 3.1: Consider a state process $x_{[0,k]}$ and a measurement process $y_{[1,k]}$ generated by the true system (1). The L_1 norm of the error between the true filter $\sigma^e_{k|[k],\sigma_{k-1}}(\cdot)$ and the class of approximating filters $\sigma^h_{k|[k],\sigma_{k-1}}(\cdot)$ is bounded by the relative entropy of conditional probability density functions in the sense that

$$\begin{aligned} &\left|\left|\sigma^e_{k|[k],\sigma_{k-1}}(\cdot) - \sigma^h_{k|[k],\sigma_{k-1}}(\cdot)\right|\right|_1 \\ &\leq B\sqrt{\mathcal{D}\left(\sigma^e_{k|[k],\sigma_{k-1}}(\cdot)\middle\|\sigma^h_{k|[k],\sigma_{k-1}}(\cdot)\right)} \quad P\text{-a.s.} \end{aligned} \tag{14}$$

for all initial conditions $\sigma_{k-1} \in \bar{L}^1(\mathbb{R}^n)$ and all $k > 0$, where $B > 0$ is a finite constant.

Proof: From [20, Lemma 11.6.1], we obtain

$$\begin{aligned} &\left(\left|\left|\sigma^e_{k|[k],\sigma_{k-1}}(\cdot) - \sigma^h_{k|[k],\sigma_{k-1}}(\cdot)\right|\right|_1\right)^2 \\ &\leq \bar{B}\mathcal{D}\left(\sigma^e_{k|[k],\sigma_{k-1}}(\cdot)\middle\|\sigma^h_{k|[k],\sigma_{k-1}}(\cdot)\right) \quad P\text{-a.s.} \end{aligned} \tag{15}$$

for some positive finite constant $\bar{B}$ (which is independent of h). The lemma statement then follows under the square root operation. ∎

Lemma 3.1 establishes that relative entropy provides a method for establishing one-step consistency for a class of estimators (as long as $\mathcal{D}\left(\sigma^e_{k|[k],\sigma_{k-1}}(\cdot)\|\sigma^h_{k|[k],\sigma_{k-1}}(\cdot)\right) \to 0$ as $h \to 0$). We now introduce the concept of multi-step consistency.

Definition 3.2: The class of approximating filters $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ is said to be multi-step consistent with the respect to the true filter $\sigma^e_{k|[1,k],\sigma_0}(\cdot)$ if, for each finite $L \geq 2$ and each finite $\eta(L) > 0$, there exists a $H > 0$ such that, for all $h \in (0, H]$, all initial conditions $\sigma_0 \in \bar{L}^1(\mathbb{R}^n)$ and all $k \in [1, L]$ we have that

$$\left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma_0}(\cdot)\right|\right|_1 \leq \eta(L) \qquad P\text{-a.s.}. \quad (16)$$

We now establish conditions under which this multi-step consistency condition holds.

Lemma 3.2: Consider a state process $x_{[0,k]}$ and a measurement process $y_{[1,k]}$ generated by the true system (1). Consider a class of approximating filters $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$. Assume that A2 and the one-step consistency condition hold, then the class of approximating filters is multi-step consistent with respect to the true filter $\sigma^e_{k|[1,k],\sigma_0}(\cdot)$.

Proof: From our definition of one-step consistency, we have that for each $\rho > 0$, there is a $H > 0$ such that, at time $k = 1$,

$$\left|\left|\sigma^e_{1|[1],\sigma_0}(\cdot) - \sigma^h_{1|[1],\sigma_0}(\cdot)\right|\right|_1 \leq \rho \qquad P\text{-a.s.} \quad (17)$$

for all $h \in (0, H]$ and all initial conditions $\sigma_0 \in \bar{L}^1(\mathbb{R}^n)$.

At time $k = 2$, we have that

$$\begin{aligned}
&\left|\left|\sigma^e_{2|[1,2],\sigma_0}(\cdot) - \sigma^h_{2|[1,2],\sigma_0}(\cdot)\right|\right|_1 \\
&= \left|\left|\sigma^e_{2|[2],\sigma^e_1(\sigma_0)}(\cdot) - \sigma^h_{2|[2],\sigma^h_1(\sigma_0)}(\cdot)\right|\right|_1 \\
&\leq \left|\left|\sigma^e_{2|[2],\sigma^e_1(\sigma_0)}(\cdot) - \sigma^h_{2|[2],\sigma^e_1(\sigma_0)}(\cdot)\right|\right|_1 \\
&\quad + \left|\left|\sigma^h_{2|[2],\sigma^e_1(\sigma_0)}(\cdot) - \sigma^h_{2|[2],\sigma^h_1(\sigma_0)}(\cdot)\right|\right|_1 \\
&\leq \rho + K\left|\left|\sigma^e_1(\sigma_0) - \sigma^h_1(\sigma_0)\right|\right|_1 \qquad P\text{-a.s.} \\
&= \rho + K\left|\left|\sigma^e_{1|[1],\sigma_0}(\cdot) - \sigma^h_{1|[1],\sigma_0}(\cdot)\right|\right|_1 \qquad P\text{-a.s.} \\
&\leq (1 + K)\rho \qquad P\text{-a.s.}. \qquad (18)
\end{aligned}$$

The 2nd step comes from Minkowski's inequality [18, p. 242]. In the 3rd step, we have applied one-step consistency assumption that $\left|\left|\sigma^e_{2|[2],\sigma^e_1(\sigma_0)}(\cdot) - \sigma^h_{2|[2],\sigma^e_1(\sigma_0)}(\cdot)\right|\right|_1 \leq \rho$ and using Assumption A2, we have that $\left|\left|\sigma^h_{2|[2],\sigma^e_1(\sigma_0)}(\cdot) - \sigma^h_{2|[2],\sigma^h_1(\sigma_0)}(\cdot)\right|\right|_1 \leq K\left|\left|\sigma^e_1(\sigma_0) - \sigma^h_1(\sigma_0)\right|\right|_1$. In the 4th step, we have used that $\sigma^e_1(\sigma_0)$ is shorthand for $\sigma^e_{1|[1],\sigma_0}(\cdot)$, etc.

Now at time $k = 3$, we have that

$$\begin{aligned}
&\left|\left|\sigma^e_{3|[1,3],\sigma_0}(\cdot) - \sigma^h_{3|[1,3],\sigma_0}(\cdot)\right|\right|_1 \\
&\leq \left|\left|\sigma^e_{3|[3],\sigma^e_2(\sigma_0)}(\cdot) - \sigma^h_{3|[3],\sigma^e_2(\sigma_0)}(\cdot)\right|\right|_1 \\
&\quad + \left|\left|\sigma^h_{3|[3],\sigma^e_2(\sigma_0)}(\cdot) - \sigma^h_{3|[3],\sigma^h_2(\sigma_0)}(\cdot)\right|\right|_1 \\
&\leq \rho + K\left|\left|\sigma^e_2(\sigma_0)(\cdot) - \sigma^h_2(\sigma_0)(\cdot)\right|\right|_1 \qquad P\text{-a.s.} \\
&= \rho + K\left|\left|\sigma^e_{2|[1,2],\sigma_0}(\cdot) - \sigma^h_{2|[1,2],\sigma_0}(\cdot)\right|\right|_1 \qquad P\text{-a.s.} \\
&\leq \rho + K(1 + K)\rho \qquad P\text{-a.s.} \\
&= (1 + K + K^2)\rho \qquad P\text{-a.s.}. \qquad (19)
\end{aligned}$$

By induction, for each $L \geq 2$ and each $\eta(L) > 0$, from our one-step consistency, we can select a $\rho > 0$, such that, for all $h \in (0, H]$ and all initial conditions $\sigma_0 \in \bar{L}^1(\mathbb{R}^n)$, we have that

$$\begin{aligned}
\left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma_0}(\cdot)\right|\right|_1 &\leq \rho\sum_{i=0}^{L-1} K^i \qquad P\text{-a.s.} \\
&\leq \eta(L) \qquad P\text{-a.s.}. \quad (20)
\end{aligned}$$

This establishes the lemma statement. ■

IV. PRACTICAL STABILITY OF APPROXIMATING FILTERS

We now establish the main result of this paper.

Theorem 4.1: (Practical asymptotic stability with respect to modelling errors) Consider a state process $x_{[0,k]}$ and a measurement process $y_{[1,k]}$ generated by the true system (1). Assume that A1 holds and that the class of approximating filters $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ is multi-step consistent with respect to the true filter $\sigma^e_{k|[1,k],\sigma_0}(\cdot)$. Then, for any selected $R > 0$, there exists a $H > 0$ such that, for all $h \in (0, H]$, all initial conditions $\left|\left|\sigma_0 - \sigma^h_0\right|\right|_1 \in N$ (with $\sigma_0, \sigma^h_0 \in \bar{L}^1(\mathbb{R}^n)$), and all $k \geq 0$, the class of approximating filters is practically stable in the presence of modelling errors in the sense that

$$\begin{aligned}
&\left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma^h_0}(\cdot)\right|\right|_1 \\
&\leq \beta\left(\left|\left|\sigma_0 - \sigma^h_0\right|\right|_1, k\right) + R \qquad P\text{-a.s.}. \qquad (21)
\end{aligned}$$

Proof: For the selected R and the bounded set N appearing in Assumption A1, let $L > 0$ and $\eta(L)$ be such that

1) $2\eta(L) \in N$,
2) $R > \frac{3}{2}\beta(2\eta(L), 0)$, and
3) $\beta(M, L) \leq \eta(L)$ where $M = \sup_{\xi\in N}|\xi|$.

Then note that, from Assumption A1 with $k = 0$, $\beta(s, 0) \geq s$. Hence, $R > \frac{3}{2}\beta(2\eta(L), 0)$ ensures that $\eta(L) \leq \frac{1}{3}R$.

Now for the selected L and $\eta(L)$, there is a finite $H > 0$ such that for all $h \in (0, H]$, all $\left|\left|\sigma_0 - \sigma^h_0\right|\right|_1 \in N$ and all $k \in [1, L]$, we have that

$$\left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma_0}(\cdot)\right|\right|_1 \leq \eta(L) \qquad P\text{-a.s.}. \quad (22)$$

Thus, from Assumption A1 and (22), we have for such $\left(h, k, \left|\left|\sigma_0 - \sigma_0^h\right|\right|_1\right)$ that

$$\begin{aligned}
&\left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma_0^h}(\cdot)\right|\right|_1 \\
&\leq \left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma_0}(\cdot)\right|\right|_1 \\
&\quad + \left|\left|\sigma^h_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma_0^h}(\cdot)\right|\right|_1 \\
&\leq \eta(L) + \beta\left(\left|\left|\sigma_0 - \sigma_0^h\right|\right|_1, k\right) \qquad P\text{-a.s..} \qquad (23)
\end{aligned}$$

Since $\eta(L) < \frac{1}{3}R$, we have that (21) holds for all $h \in (0, H]$, all $\left|\left|\sigma_0 - \sigma_0^h\right|\right|_1 \in N$, and all $k \in [1, L]$. It remains to establish that this holds for larger k.

We highlight that from our choice of L and $\eta(L)$, (23), and noting that $N \subset M$, we have at time $k = L$ that

$$\left|\left|\sigma^e_{L|[1,L],\sigma_0}(\cdot) - \sigma^h_{L|[1,L],\sigma_0}(\cdot)\right|\right|_1 \leq 2\eta(L) \qquad P\text{-a.s..} \qquad (24)$$

Now consider the time interval $k \in [L+1, 2L]$ and let $\bar{k} = k - L$. From Assumption A1, time-invariance of the true system (1) and the approximating model (8), the bounds previously established in (22) and (24), and that $2\eta(L) \in N$, we have for all $k \in [L+1, 2L]$ that

$$\begin{aligned}
&\left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma_0^h}(\cdot)\right|\right|_1 \\
&= \left|\left|\sigma^e_{\bar{k}|[L+1,k],\sigma^e_L(\sigma_0)}(\cdot) - \sigma^h_{\bar{k}|[L+1,k],\sigma^h_L(\sigma_0^h)}(\cdot)\right|\right|_1 \\
&\leq \left|\left|\sigma^e_{\bar{k}|[L+1,k],\sigma^e_L(\sigma_0)}(\cdot) - \sigma^h_{\bar{k}|[L+1,k],\sigma^e_L(\sigma_0)}(\cdot)\right|\right|_1 \\
&\quad + \left|\left|\sigma^h_{\bar{k}|[L+1,k],\sigma^e_L(\sigma_0)}(\cdot) - \sigma^h_{\bar{k}|[L+1,k],\sigma^h_L(\sigma_0^h)}(\cdot)\right|\right|_1 \\
&\leq \eta(L) + \beta\left(\left|\left|\sigma^e_L(\sigma^e_0) - \sigma^h_L(\sigma^h_0)\right|\right|_1, \bar{k}\right) \quad P\text{-a.s.} \\
&\leq \eta(L) + \beta\left(2\eta(L), \bar{k}\right) \quad P\text{-a.s..} \qquad (25)
\end{aligned}$$

Here, Assumption A1 is used in the 2nd step. In the 3rd step, we have used that $\sigma^e_L(\sigma^e_0)$ is shorthand for $\sigma^e_{L|[1,L],\sigma_0}(\cdot)$, etc. The 4th step follows from (24).

Now we highlight that, for all $k \in [L+1, 2L]$,

$$\begin{aligned}
\left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma_0^h}(\cdot)\right|\right|_1 &\leq \eta(L) + \beta\left(2\eta(L), 0\right) P\text{-a.s.} \\
&\leq \frac{R}{3} + \frac{2R}{3} \qquad P\text{-a.s.} \\
&= R \qquad P\text{-a.s.} \qquad (26)
\end{aligned}$$

We also note that at the end of the interval $[L+1, 2L]$, we have that

$$\begin{aligned}
&\left|\left|\sigma^e_{2L|[1,2L],\sigma_0}(\cdot) - \sigma^h_{2L|[1,2L],\sigma_0^h}(\cdot)\right|\right|_1 \\
&\leq \eta(L) + \beta\left(2\eta(L), L\right) \qquad P\text{-a.s.} \\
&\leq 2\eta(L) \qquad P\text{-a.s..} \qquad (27)
\end{aligned}$$

The result then follows by induction. ∎

The importance of Theorem 4.1 is that if the approximating filter is asymptotically stable with respect to initial conditions and locally consistent with respect to the true filter, then the error between the true and approximating filters is asymptotically small. This means that approximations can be designed that, asymptotically, have any desired level of relative performance. We stress that we only require the error between the approximating filter and differently initialised versions of the approximating filter itself is bounded by $\beta(\cdot, \cdot)$. This assumption, combined with multi-step consistency property, is used to establish the practical stability result.

Remark 4: The role of relative entropy in the presented practical asymptotic stability result is limited to establishing the useful one-step consistency property and we acknowledge that it may be possible to establish the required one-step consistency using other techniques. However, we highlight that relative entropy concepts allow the important one-step consistency property to be established without appealing to the specific nature of the filtering recursions (in comparison, we note that previous stability results of this type have only been established by appealing to specific features of the recursions involved, for example the stability results for particle filters established in [6]–[8]).

A. Models with Sufficiently Informative Observations

Let us now introduce an assumption under which our true and approximation models will be said to have sufficiently informative observations.

C1) The difference between distributions $\Delta\sigma^h_k(\cdot) = \sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma_0^h}(\cdot)$ has light tails if there exists a $H > 0$ such that, for all $h \in (0, H]$, all initial conditions $\sigma_0, \sigma_0^h \in \bar{L}^1(\mathbb{R}^n)$, and all $k \geq 0$, we have that

$$\left|\left|\Delta\sigma^h_k(x)\right|\right|_1 \geq B\left|\left|\Delta\sigma^h_k(x)x\right|\right|_1 \quad P\text{-a.s.}$$

for some finite constant B (which is independent of h).

Remark 5: Assumption C1 implies that observations should be sufficiently informative so that the information states corresponding to the true and approximating models sufficiently match outside some compact set. As simple example, C1 automatically holds if w_k has compact support. Moreover, this condition also holds when w_k has Gaussian density (which tends to "localise" the state values enough for C1 to hold). Although we admit this condition seems difficult to establish without examination of the specific filters involved, we highlight that Assumption C1 seems no more restrictive than the usual type of observation model assumptions that appears in establishment of the asymptotic stability with respect to initial conditions, see [12], [13].

We can now establish a result related to the conditional mean estimates produced by the approximate filters.

Theorem 4.2: Consider a state process $x_{[0,k]}$ and a measurement process $y_{[1,k]}$ generated by the true system (1). Assume A1, C1, and that the class of approximating filters $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ is multi-step consistent with respect to the true filter $\sigma^e_{k|[1,k],\sigma_0}(\cdot)$. Then, for any selected $R > 0$, there is a $H > 0$ such that, for all $h \in (0, H]$, all initial conditions $\left|\left|\sigma_0 - \sigma_0^h\right|\right|_1 \in N$ (with $\sigma_0, \sigma_0^h \in \bar{L}^1(\mathbb{R}^n)$), and all $k \geq 0$, the class of approximating filters is practically stable in the presence of modelling errors in the sense that

$$\left|\hat{x}^e_{k|[1,k],\sigma_0} - \hat{x}^h_{k|[1,k],\sigma_0^h}\right| \leq \beta\left(\left|\left|\sigma_0 - \sigma_0^h\right|\right|_1, k\right) + R \quad P\text{-a.s..} \tag{28}$$

Proof: We first note that

$$\begin{aligned}\hat{x}^e_{k|[1,k],\sigma_0} &- \hat{x}^h_{k|[1,k],\sigma_0^h} \\ &= \int_{x\in\mathbb{R}^n} \left(\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma_0^h}(\cdot)\right) x dx. \end{aligned} \tag{29}$$

By taking the magnitude, we obtain

$$\begin{aligned}&\left|\hat{x}^e_{k|[1,k],\sigma_0} - \hat{x}^h_{k|[1,k],\sigma_0^h}\right| \\ &= \left|\int_{x\in\mathbb{R}^n} \left(\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma_0^h}(\cdot)\right) x dx\right| \\ &\leq \left|\left|\left(\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma_0^h}(\cdot)\right) x\right|\right|_1 \\ &\leq \bar{B}\left|\left|\left(\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma_0^h}(\cdot)\right)\right|\right|_1 \end{aligned} \tag{30}$$

for some finite positive constant $\bar{B}$ (which is independent of h).

In the 2nd step, we have used the definition of $||\cdot||_1$ operation and the integral property that $||\xi(\cdot)x||_1 \geq \left|\int_{x\in\mathbb{R}^n}\xi(\cdot)xdx\right|$. The 3rd step follows from Assumption C1. The theorem statement then follows from the result of Theorem 4.1. ■

V. CASE STUDY

This paper's stability results provide the first theoretical justification to the widespread application of HMMs in various approximation problems [1]–[3]. We now illustrate these results in the approximation of a scalar continuous-valued nonlinear system by a hidden Markov model (HMM). For this purpose, we will create a special interpretation of the HMM's underlying discrete-state process via a spatial "blurred" version of the state process.

For presentation purposes, we limit our example to a scalar example (but this approach can be generalised). Consider a scalar true model with dynamics $x_k \in \mathbb{R}$ described by (1), where $f(x) = \text{mod}(ax+b, 2b) - b$ for some $a \in \bar{N}$ and $b > 0$, $c(x) = x$, $\phi_w(\cdot)$ is a zero-mean unit-variance Gaussian density, and $\phi_v(\cdot)$ is some density function with support only in the interval$[-1, 1]$ (here $\text{mod}(\cdot,\cdot)$ is the modulus operation). Under these assumptions $x_k \in S_x = [-(b+1), b+1)$ for all k.

The restriction of dynamics to a finite region of state-space is somewhat limiting but is also understandable considering the nature of HMMs (also, admittedly, the restriction to bounded region immediately implies that filtering errors are finite, but our results establish the tighter error bound).

We will now introduce a HMM process which approximates the true system described above (repeating the construction of [3]). Let $e_i = [0,\ldots,0,1,0,\ldots,0]' \in \mathbb{R}^{\bar{N}}$ denote an indicator vector with 1 in the ith position and zero elsewhere, and let $\bar{N}$ denotes the number of HMM states (to simplify later construction, we will assume $\bar{N}$ is even). At time k, we will let $X_k \in \{e_1, e_2, \ldots, e_{\bar{N}}\}$ denote the state of the HMM process. This HMM state process is described by a transition probabilities matrix A with ijth element, $A_{ij} = p(X_{k+1} = e_i|X_k = e_j)$, where $p(\cdot)$ is the probability law describing our HMM state process. The HMM state process is also assumed to have an initial probabilities vector π with ith element, $\pi_i = p(X_0 = e_i)$. The measurement process y_k associated with the HMM state process is described by an output probability matrix $B(y_k)$ given by $B(y_k) = \text{diag}([p(y_k|X_k = e_1), \ldots, p(y_k|X_k = e_{\bar{N}})])$, where $\text{diag}(x)$ is the diagonal matrix with x on its diagonal.

The HMM state process is assumed to exist on the following spatial discretisation of S_x. Let G be the spatial grid (with $\bar{N}$ grid points) that approximates S_x such that $G = \{x : x = \pm mh\}$ where $m = 1, \ldots, (\bar{N}/2)$, and $h = 2(b+1)/\bar{N}$ is spacing parameter. This allows us to relate each grid point with a HMM state value. We will use $G(e_i)$ to denote the specific location on G corresponding to state value e_i.

We will now introduce a blurred approximating process associated with this HMM state process. Let $C(e_i)$ denote a h-sized cell containing grid location $G(e_i)$. The cell $C(e_i)$ is used to describe the region of S_x represented by the state value e_i. The cells are assumed designed to completely cover S_x in the sense that for all $x \in S_x$, $x \in C(e_i)$ for some e_i. Conversely, let $e(x)$ be the indicator vector denoting the cell containing the value x (ie. the inverse association), that is $x \in C(e(x))$ for all possible values of $x \in S_x$. We will also assume that the boundaries between adjacent cells are not shared. For approximation purposes, we define x_k^a to be a blurred version of X_k, with the properties that, for all $k \geq 0$, $x_k^a \in C(X_k)$, $X_k = e(x_k^a)$ and x_k^a has uniform distribution over the cell $C(X_k)$. Further, we assume that $p(y_k|x_k^a) = p(y_k|X_k = e(x_k^a))$ for all possible values of y_k and x_k^a. We also assume that the grid points $G(e_i)$ are centred in their corresponding cells $C(e_i)$ so that $E[x_k^a|X_k = e(x_k^a)] = G^h(e(x_k^a))$ for all possible values of x_k^a. At time k, the information state associated with this blurred process, given the measurements $y_{[1,k]}$ and an blurred initial condition corresponding π, can be written as:

$$\sigma^{a,h}_{k|[1,k],\sigma_0}(x) = \frac{1}{h}e(x)'\hat{X}_{k|[1,k],\pi}$$

where $\hat{X}_{k|[1,k],\pi}$ denotes the HMM filter estimate at time k given the measurements $y_{[1,k]}$ and the initial condition π (see [11]). Note that the inner product $e(x)'\hat{X}_{k|[1,k],\pi}$ simply extracts the element of $\hat{X}_{k|[1,k],\pi}$ corresponding to the filtered probability of being in cell $C(x)$.

We now consider the application of Theorem 4.1 and Theorem 4.2 to the approximate filter $\sigma^{a,h}_{k|[1,k],\sigma_0}(x)$ by introducing some assumptions about the approximation model. Let $x^+_{k+1} = E[x_{k+1} - x_k|x_k]$. Let us choose an irreducible and aperiodic A such that

$$\begin{aligned}A_{ij} = &\frac{1}{2}\max\left[0, \min\left(G(e_i) + \frac{h}{2}, f\left(G(e_j)\right) + 1\right)\right. \\ &\left. - \max\left(G(e_i) - \frac{h}{2}, f\left(G(e_j)\right) - 1\right)\right],\end{aligned}$$

so that the state dynamics are matched in means and variances (that is, locally consistent as suggested in [21]) in the sense that

1) $E[G(X_{k+1}) - G(X_k)|X_k = e(x_k)] = x^+_{k+1} + \alpha_1 h$ for some $\alpha_1 > 0$ and for all x_k,
2) $E[(G(X_{k+1}) - G(X_k) - x^+_{k+1})^2|X_k = e(x_k)] = E[(x_{k+1} - x_k - x^+_{k+1})^2|x_k] + \alpha_2 h$ for some $\alpha_2 > 0$ and for all x_k.

Assume the observation model $p(y_k|X_k) = \phi_w(y_k - G(X_k))$. Under these assumptions on HMM parameters (specifically, irreducible and aperiodic A and positive observation density $p(y_k|X_k)$), Theorem 2.2 of [22] shows that corresponding HMM filter is exponential forgetting in the sense that

$$\left|\left|\hat{X}_{k|[1,k],\pi} - \hat{X}_{k|[1,k],\pi_0}\right|\right|_1 \leq \beta^{HMM}\left(||\pi - \pi_0||_1, k\right) P\text{-a.s.},$$

where π, π_0 are two different initial conditions. Here, $\beta^{HMM}(s,k) = \alpha\bar{\varepsilon}^k||s||_1$ where $\bar{\varepsilon} < 1$ and α is a finite constant. Under our definition of x^a_k, the same exponential stability with respect to initial conditions also holds for the blurred process x^a_k, P-a.s.. Also, Assumption A2 holds because HMM conditional mean estimates are linear in previous estimate, see [11].

Now we note that the bounded nature of S_x implies that the information states $\sigma^e_k(\cdot)$ and $\sigma^h_k(\cdot)$ have compact support (which is independent of h) and hence, Assumption C1 holds because

$$\begin{aligned}\|\Delta\sigma^h_k(x_k)x_k\|_1 &= \int \left|\Delta\sigma^h_k x_k\right| dx_k \\ &\leq |x_{max}| \int \left|\Delta\sigma^h_k\right| dx_k \\ &= B\|\Delta\sigma^h_k\|_1\end{aligned}$$

where x_{max} is the largest absolute value of x in the support of $\Delta\sigma_h$. Hence, if we select our HMM design so that $\mathcal{D}\left(\sigma^e_{k|[k],\sigma_{k-1}(\cdot)}\middle\|\sigma^h_{k|[k],\sigma_{k-1}(\cdot)}\right) \to 0$ as $h \to 0$, then Assumptions A2 can be used to give that the approximating HMM model is multi-step consistent with the true model. Consequently, Assumptions A1, A2, C1 and the multi-step consistency property allow us to apply Theorem 4.1 (and Theorem 4.2) to establish practical stability of the approximating filters with respect to modelling errors.

VI. CONCLUSION

In this paper, we present practical stability with respect to modelling errors results for a range of approximate filtering problems. The results are established using some important filter local consistency and relative entropy concepts. We illustrated the application of our practical stability results in the case of hidden Markov model based approximations.

VII. ACKNOWLEDGMENTS

This research was supported under Australian Research Council's Linkage Projects funding scheme (project number LP100100302) and the Smart Skies Project, which is funded, in part, by the Queensland State Government Smart State Funding Scheme. The work of the third author was supported by the ARC under the Future Fellow and Discovery Grants programs.

REFERENCES

[1] R. L. Streit and R. F. Barrett, "Frequency line tracking using hidden Markov models," *IEEE Trans. Acoustic, Speech, and Signal Process.*, vol. 38(4), pp. 2659-2676, Apr. 1999.
[2] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77(2), pp. 257-286, Feb. 1989.
[3] J. Lai and J. J. Ford, "Relative entropy rate based multiple hidden Markov model approximation," *IEEE Trans. Signal Process.*, vol. 58(1), pp. 165-174, Jan. 2010.
[4] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, New York: Academic Press, 1970.
[5] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Englewood Cliffs: Prentice-Hall, 1979.
[6] D. Crisan and K. Heine, "Stability of the discrete time filter in terms of the tails of noise distributions," *J. London Math. Soc.*, vol. 78(2), pp. 441-458, Jul. 2008.
[7] B. Azimi-Sadjadi and P. S. Krishnaprasad, "Approximate nonlinear filtering and its application in navigation," *Automatica*, vol. 41(6), pp. 945-956, Jun. 2005.
[8] K. Heine and D. Crisan, "Uniform approximations of discrete-time filters," *Adv. Appl. Prob.*, vol. 40(4), pp. 979-1001, 2008.
[9] K. Reif, S. Gunther, and E. Yaz, "Stochastic stability of the discrete-time extended Kalman filter," *IEEE Trans. Autom. Control*, vol. 44(4), pp. 714-728, Apr. 1999.
[10] B. F. La Scala, R. R. Bitmead, and M. R. James, "Conditions for stability of the extended Kalman filter and their application to the frequency tracking problem," *Mathematics of Control, Signals, and Systems*, vol. 8(1), pp. 1-26, 1995.
[11] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*, Berlin: Spring-Verlag, 1995.
[12] R. van Handel, "Discrete time nonlinear filters with informative observations are stable," *Elect. Comm. in Probab.*, vol. 13, pp. 562-575, Oct. 2008.
[13] A. Budhiraja and D. Ocone, "Exponential stability in discrete-time filtering for non-ergodic signals," *Stochastic Processes and their Applications*, vol. 82(2), pp. 245-257, Aug. 1999.
[14] L. Xie, V. A. Ugrinovskii, and I. R. Petersen, "Probabilistic distances between finite-state finite-alphabet hidden Markov models," *IEEE Trans. Autom. Control*, vol. 50(4), pp. 505-511, Apr. 2005.
[15] M. Vidyasagar, "Bounds on the Kullback-Leibler divergence rate between hidden Markov models," In *Proc. 46th IEEE Conf. on Decision and Control*, pp. 6160-6165, 2007.
[16] D. Nešić, A. R. Teel, and P. V. Kokotovic, "Sufficient conditions for stabilization of sampled-data nonlinear systems via discrete-time approximations," *Syst. Control Lett.*, vol. 38(4-5), pp. 259-270, Dec. 1999.
[17] M. Arcak and D. Nešić, "A framework for nonlinear sampled-data observer design via approximate discrete-time models and emulation," *Automatica*, vol. 40, pp. 1931-1938, 2004.
[18] P. Billingsley, *Probability and Measure*, 3rd ed., New York: Wiley, 1995.
[19] D. G. Luenberger, *Optimization by Vector Space Methods*, New York: Wiley, 1969.
[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., New York: Wiley, 2005.
[21] H. J. Kushner and P. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, 2nd ed., New York: Springer-Verlag, 2001.
[22] F. Le Gland and L. Mevel, "Exponential forgetting and geometric ergodicity in hidden Markov models," *Mathematics of Control, Signals, and Systems*, vol. 13(1), pp. 63-93, 2000.
[23] H. K. Khalil, *Nonlinear Systems*, 3rd ed., Upper Saddle River, N.J.: Prentice-Hall, 2002.