Practical Stability of Approximating Discrete-Time Filters with Respect to Model Mismatch ☆

Onvaree Techakesari[a,*], Jason J. Ford[a], Dragan Nešić[b]

[a]*School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane QLD 4001, Australia*
[b]*Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville VIC 3010, Australia*

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. Practical stability is established in the sense of an asymptotic bound on the amount of bias introduced by the model approximation. Our analysis applies to a wide range of estimation problems and justifies the common practice of approximating intractable infinite dimensional nonlinear filters by simpler computationally tractable filters.

Keywords: Asymptotic Stability; Discrete-time Systems; Filter Stability; Model Approximation; Modelling Errors; Nonlinear Filters; Practical Stability

1. 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 filtering problems, tractable filters are often proposed on the *ad hoc* basis of an approximating system that reasonably represents the true dynamics. For example, using this informal idea, hidden Markov model (HMM) filters and Kalman filters have been exploited in a wide range of signal and image processing applications, see [1, 2, 3]. Despite the successful application of approximate filters in a large number of applications, conditions that ensure reasonable filter behaviour have not been completely established in many situations.

When considering filtering behaviour, there are two basic types of stability properties of interest: asymptotic stability with respect to initialisation errors and stability properties in the presence of modelling errors. The first type of stability properties is important because initial conditions are rarely known perfectly, whilst the second type of property is important because system models are usually not completely known (or are too complex). Fortunately, asymptotic stability of filters with respect to erroneous initial conditions has been established in many situations including Kalman filter [4, 5, 6], risk-sensitive filters [7], as well as some general asymptotic stability results provided in [8, 9, 10]. In comparison, only a small number of stability type results for situations involving model mismatch have been established. These include stability with respect to model mismatch for Kalman filters [4, 5], HMM filters [11], and particle filters [12, 13, 14]. Further, some convergence and stochastic stability type results for extended Kalman filters are presented in [15]. Beyond these results, the expected performance properties of approximating filters can be indirectly characterised through performance limits provided by optimal filters (for example, lower error bounds of optimal filters are established in [16, 17, 18, 19] for the purpose of characterising the possible performance of approximating filters). However, this type of analysis only provides lower error bound for approximating filters and, of course, any specific filter might perform considerably worse than the lower bound.

In this paper, we investigate stability of general approximating filters in the presence of modelling errors. In this situation, the recursive nature of the filtering process might suggest that the inclusion of modelling error, at each filtering step, could lead to an unbounded growth in estimation error. However, if the approximating filter exhibits some initial condition forgetting properties, and the error introduced by the model approximation is bounded on a finite time interval, then we can show that the filtering error is bounded forever. Moreover, under some additional multi-step consistency assumptions, practical stability of the approximating filters can be established. The stability proofs used here are similar in nature to the proofs used in the important nonlinear control stability results estab-

☆An earlier version of this paper was presented at the IEEE CDC 2011 Conference. 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 last author is supported by the Australian Research Council under the Future Fellow and Discover Project schemes.

*Tel.: +61 7 3138 8496; fax: +61 7 3138 4528.

Email addresses: o.techakesari@qut.edu.au (Onvaree Techakesari), j2.ford@qut.edu.au (Jason J. Ford), dnesic@unimelb.edu.au (Dragan Nešić)

Preprint submitted to Automatica
Received June 20, 2012 23:15:57 PST

lished in [20].

This paper is structured as follows: In Section 2, we introduce our filter approximation problem. In Section 3, we establish a preliminary bound result for approximating filters, before our main practical stability result is established in Section 4. In Section 5, an example is presented and some conclusions are then provided in Section 6.

2. Problem Formulation

2.1. Dynamics

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

$$\begin{aligned} x_{k+1} &= f(x_k) + w_{k+1} \\ y_{k+1} &= c(x_k) + v_{k+1} \end{aligned} \tag{1}$$

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, $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are sequences of independent and identically distributed *i.i.d.* random variables with strictly positive densities $\phi_w(\cdot)$ and $\phi_v(\cdot)$, respectively. The random variables w_k, v_k, and x_0 are assumed to be mutually independent for all k. We will use the shorthand $y_{[\ell,m]}$ to denote the measurement sequences $\{y_\ell, \ldots, y_m\}$. We likewise define $x_{[\ell,m]}$.

Throughout the rest of this paper, we will consider processes defined on a probability space $(\Omega, \mathcal{F}, P)$ where Ω is defined to consist 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 endowed on Ω, and on the basis of distributions described by (1), P will be the probability measure defined on $(\Omega, \mathcal{F})$ given by the Kolmogorov existence theorem [21]. Finally, we let $\mathcal{Y}_{[1,k]}$ denote the complete filtration generated by the sequence $y_{[1,k]}$, see [22, 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:

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

for all $k > 0$, where $E[\cdot]$ denotes the expectation operation corresponding to P.

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 approximate filters that provide approximate estimates for our system state, x_k.

2.2. Normalised Information State

We now introduce some information state concepts that describe our estimation operations. Consider the spaces $L^\infty(\mathbb{R}^n)$ and $L^1(\mathbb{R}^n)$; see [23] 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 [23]:

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

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 the sense that $\bar{L}^1(\mathbb{R}^n) \triangleq \{\xi(\cdot) : \xi(\cdot) \in L^1(\mathbb{R}^n) \text{ and } ||\xi(\cdot)||_1 = 1\}$. 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

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

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. Note that information states are also $(\Omega, \mathcal{F}, P)$ random variables, but we will suppress the dependency on ω. 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 [22, Thm 4.4 of Ch. 5]

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

for $x \in \mathbb{R}^n$, where $\sigma^{e|u}_0 = \sigma_0 \in L^1(\mathbb{R}^n)$. We first note that $\sigma^{e|u}_k(\cdot) \in L^1(\mathbb{R}^n)$ for all k because, using that $\phi_v(\cdot)$ is strictly positive density, (6) defines a bounded linear operator as argued in [24]. In this paper, we will say that $y_{[1,T]}$ is a T-feasible measurement sequence for the true model if $||\sigma^{e|u}_k(\cdot)||_1 > 0$ for all $k \in [1, T]$. Note that this property automatically holds because we have assumed strictly positive densities. We also note that if $y_{[1,T]}$ is a T-feasible measurement sequence for the true model, then a normalised information state $\sigma^e_k(\cdot) \in \bar{L}^1(\mathbb{R}^n)$ can then be written as

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

for all $k \in [1, T]$, where $N_k = ||\sigma^{e|u}_k(\cdot)||_1$ is a normalisation factor. We highlight that this ensures $\sigma^e_k(\cdot) \in \bar{L}^1(\mathbb{R}^n)$. 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 subscripts 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 $\sigma^e_{k|[\ell+1,k],\sigma^e_\ell(\sigma_0)}(\cdot) = \sigma^e_{k|[1,k],\sigma_0}(\cdot)$. Finally, we define $\sigma^e_{0|[1,0],\sigma_0} \triangleq \sigma_0$ for all $\sigma_0 \in \bar{L}^1(\mathbb{R}^n)$.

Preprint submitted to Automatica
Received June 20, 2012 23:15:57 PST

2.3. Parameterised Class of Approximating Models

Let $h > 0$ parameterise a class of approximating models (for example, if considering a class of hidden Markov models, then 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 $k \geq 0$):

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

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, $w^h_k \in \mathbb{R}^n$ and $v^h_k \in \mathbb{R}^m$ are *i.i.d.* random variables with strictly positive densities $\phi^h_w(\cdot)$ and $\phi^h_v(\cdot)$, respectively, and w^h_k, v^h_k, and x_0 are assumed to be mutually independent. Corresponding to each approximating model (8), based on the distributions described by (8), P^h will be the probability measure defined on $(\Omega, \mathcal{F})$ given by the Kolmogorov existence theorem [21].

For a given $h > 0$, we can also define the conditional mean estimate associated with the approximating model given the measurements $y_{[1,k]}$ and the *a priori* distribution σ^h_0, when it exists, as:

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

for all $k > 0$, 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 an approximating model, as

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

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

$$\sigma^{h|u}_k(x) = \int_{\mathbb{R}^n} \frac{\phi^h_v(y_k - c^h(z))}{\phi^h_v(y_k)} \phi^h_w(x - f^h(z))\sigma^{h|u}_{k-1}(z)dz \tag{11}$$

for all $x \in \mathbb{R}^n$, when $\sigma^{h|u}_0 = \sigma^h_0 \in L^1(\mathbb{R}^n)$. Similar to the argument above, $\sigma^{h|u}_k(\cdot) \in L^1(\mathbb{R}^n)$ for all k. We will also say that $y_{[1,T]}$ is a T-feasible measurement sequence for the approximating model if $||\sigma^{h|u}_k(\cdot)||_1 > 0$ for all $k \in [1, T]$ (note that this holds automatically because we have assumed strictly positive densities). Here, if $y_{[1,T]}$ is a T-feasible measurement sequences for the approximating model, then for all $k \in [1, T]$, 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 $\bar{N}^{-1}_k = ||\sigma^{h|u}_k(\cdot)||_1$. Again, we highlight that $\sigma^h_k(\cdot) \in \bar{L}^1(\mathbb{R}^n)$. As above, we can 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 define $\sigma^h_{0|[1,0],\sigma_0} \triangleq \sigma_0$ for all $\sigma_0 \in \bar{L}^1(\mathbb{R}^n)$.

2.4. Measurements and Information State Support

Let us consider a subset $\Gamma \in \mathcal{F}$ that has desirable properties. In the following, we will establish results on Γ. This type of restriction is not unreasonable, and for example we would only expect to establish performance results on feasible measurement sequences. Moreover, sometimes it is useful to establish error bounds in specific measurement situations.

A set of $\bar{L}^1(\mathbb{R}^n)$ functions denoted by $G(\mathbb{R}^n)$ will be called a support of an information state filter if, for all initial conditions $\sigma_0 \in G(\mathbb{R}^n)$ and all $\omega \in \Gamma$, the filter stays within this set in the sense that $\sigma_k(\sigma_0) \in G(\mathbb{R}^n)$ for all k. In this paper, we consider two support sets $G^e(\mathbb{R}^n)$ and $G^h(\mathbb{R}^n)$ for the information states $\sigma^e_k(\cdot)$ and $\sigma^h_k(\cdot)$, respectively. Because the information states remain in these support sets, we will only require filtering properties to hold on $G^e(\mathbb{R}^n)$ and $G^h(\mathbb{R}^n)$ rather than the entire $\bar{L}^1(\mathbb{R}^n)$.

We now define a projection operator between support sets as $\pi^h(\sigma) = \min_{\bar{\sigma} \in G^h(\mathbb{R}^n)} ||\sigma - \bar{\sigma}||_1$ for any $\sigma \in G^e(\mathbb{R}^n)$ (when a minimum exists). Otherwise, when a unique minimum does not exist, for the purpose of this paper we can always define a non-unique projection operator as $\pi^h(\sigma) \in \{\bar{\sigma} \in G^h(\mathbb{R}^n) : \inf_{\bar{\sigma} \in G^h(\mathbb{R}^n)} ||\sigma - \bar{\sigma}||_1 \leq \delta\}$ for any $\sigma \in G^e(\mathbb{R}^n)$ and for some small $\delta > 0$.

In the next two sections, we will introduce some important definitions and establish our main results.

3. Bounded Error of an Approximating Filter

A function ψ is said to be of class-$\mathscr{K}$ if it is continuous, strictly increasing, and $\psi(0) = 0$. Moreover, function β 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 [25, Ch. 4] for descriptions of system stability involving such functions).

Consider a given set Γ and $\pi^h(\cdot)$ operator. We will now introduce some important definitions.

Definition 3.1. *(Asymptotic stability of an approximating filter with respect to initial conditions) Consider an approximating filter $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ (some fixed $h > 0$). For a given set Γ, the approximating filter $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ is said to be asymptotically stable with respect to initial conditions if there exists a $\beta(\cdot, \cdot) \in \mathscr{KL}$ such that, for all $\sigma_0, \bar{\sigma}_0 \in G^h(\mathbb{R}^n)$, and all $k \geq 0$, we have that*

$$\left|\left|\sigma^h_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\bar{\sigma}_0}(\cdot)\right|\right|_1 \leq \beta\left(||\sigma_0 - \bar{\sigma}_0||_1, k\right) \tag{12}$$

for all $\omega \in \Gamma$.

Remark 1. *Definition 3.1 is an abstract version of the asymptotic stability property with respect to initial conditions that is often encountered in discussion of filter behaviour (for example, see [4, 5, 11]). As an example, if $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ corresponds to a Kalman filter, then under observability and other mild conditions, asymptotic stability of covariance matrix and conditional mean estimate with*

Preprint submitted to Automatica
Received June 20, 2012 23:15:57 PST

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 for all $\omega \in \Gamma$ and all $\sigma_0, \bar{\sigma}_0 \in G(\mathbb{R}^n)$ (the set of Gaussian densities)

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

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

Definition 3.2. *(Finite filter error on finite interval) Consider an approximating filter $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ (some fixed $h > 0$). For a given set Γ and projection operator $\pi^h(\cdot)$, the approximating filter $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ is said to have finite error on a finite interval $[1, L]$ with respect to the true filter $\sigma^e_{k|[1,k],\sigma_0}(\cdot)$ if, for all initial conditions $\sigma_0 \in G^e(\mathbb{R}^n)$, all $\omega \in \Gamma$, and all $k \in [1, L]$, we have that*

$$\left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\pi^h(\sigma_0)}(\cdot)\right|\right|_1 \leq \eta(L) \tag{13}$$

where $\eta(L) > 0$ is finite.

Remark 2. *The quantity $\eta(L)$ may depend on the model parameter h, but this dependency has been suppressed in our notation to aid presentation (and such dependency does not have a role in the following theorem).*

Definition 3.3. *(Error growth matched filters) Consider an approximating filter $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ (some fixed $h > 0$). For a given set Γ and projection operator $\pi^h(\cdot)$, we will say that the approximating filter $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ and the true filter $\sigma^e_{k|[1,k],\sigma_0}(\cdot)$ are L^*-error growth matched filters in Γ, if the approximating filter is asymptotically stable with respect to initial conditions and has finite error on the finite interval $[1, L^*]$ with respect to the true filter $\sigma^e_{k|[1,k],\sigma_0}(\cdot)$, and*

$$\beta(M, L^*) \leq \eta(L^*) \tag{14}$$

where $M = \sup_{\sigma,\bar{\sigma} \in G^h(\mathbb{R}^n)} ||\sigma - \bar{\sigma}||_1$.

We now establish a key result.

Theorem 3.1. *(Asymptotically bounded error in the presence of modelling errors) Consider a state process $x_{[0,k]}$ and a measurement process $y_{[1,k]}$ generated by the true system (1), and an approximating filter $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ (some fixed $h > 0$). For a given set Γ and projection operator $\pi^h(\cdot)$, assume that the approximating filter $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ and the true filter $\sigma^e_{k|[1,k],\sigma_0}(\cdot)$ are L^*-error growth matched filters. Then, the asymptotic error introduced by the approximating filter $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ is bounded by $R = 2\eta(L^*)$, in the sense that, for all $k \geq 0$,*

$$\begin{aligned}&\left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma^h_0}(\cdot)\right|\right|_1 \\ &\quad \leq \beta\left(\left|\left|\pi^h(\sigma_0) - \sigma^h_0\right|\right|_1, k\right) + R\end{aligned} \tag{15}$$

for all $\omega \in \Gamma$ and all initial conditions $\sigma_0 \in G^e(\mathbb{R}^n), \sigma^h_0 \in G^h(\mathbb{R}^n)$.

Proof. First we have for all $k \in [0, L^*]$, all $\omega \in \Gamma$, and all initial conditions $\sigma_0 \in G^e(\mathbb{R}^n)$, $\sigma^h_0 \in G^h(\mathbb{R}^n)$ that

$$\begin{aligned}&\left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma^h_0}(\cdot)\right|\right|_1 \\ &\leq \left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\pi^h(\sigma_0)}(\cdot)\right|\right|_1 \\ &\quad + \left|\left|\sigma^h_{k|[1,k],\pi^h(\sigma_0)}(\cdot) - \sigma^h_{k|[1,k],\sigma^h_0}(\cdot)\right|\right|_1 \\ &\leq \eta(L^*) + \beta\left(\left|\left|\pi^h(\sigma_0) - \sigma^h_0\right|\right|_1, k\right).\end{aligned} \tag{16}$$

In the 1st step, we have used Minkowski's inequality. In the last step, we have used Definitions 3.1 and 3.2. Since $\eta(L^*) = \frac{1}{2}R$, we have that (15) holds for all $k \in [0, L^*]$. It remains to establish that this holds for larger k.

Now consider the time interval $k \in [1, L^* + 1]$ and let $\bar{k} = k - 1$. From Definition 3.1, time-invariance of the true system (1) and the approximating model (8), we can write

$$\begin{aligned}&\left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma^h_0}(\cdot)\right|\right|_1 \\ &= \left|\left|\sigma^e_{\bar{k}|[1,\bar{k}],\sigma^e_1(\sigma_0)}(\cdot) - \sigma^h_{\bar{k}|[1,\bar{k}],\sigma^h_1(\sigma^h_0)}(\cdot)\right|\right|_1 \\ &\leq \left|\left|\sigma^e_{\bar{k}|[1,\bar{k}],\sigma^e_1(\sigma_0)}(\cdot) - \sigma^h_{\bar{k}|[1,\bar{k}],\pi^h(\sigma^e_1(\sigma_0))}(\cdot)\right|\right|_1 \\ &\quad + \left|\left|\sigma^h_{\bar{k}|[1,\bar{k}],\pi^h(\sigma^e_1(\sigma_0))}(\cdot) - \sigma^h_{\bar{k}|[1,\bar{k}],\sigma^h_1(\sigma^h_0)}(\cdot)\right|\right|_1 \\ &\leq \eta(L^*) + \beta\left(\left|\left|\pi^h(\sigma^e_1(\sigma_0)) - \sigma^h_1(\sigma^h_0)\right|\right|_1, \bar{k}\right)\end{aligned} \tag{17}$$

where we have again used Minskowski inequality and Definitions 3.1 and 3.2. Hence, at time $\bar{k} = L^*$, that is $k = L^* + 1$, we have that

$$\begin{aligned}&\left|\left|\sigma^e_{L^*+1|[1,L^*+1],\sigma_0}(\cdot) - \sigma^h_{L^*+1|[1,L^*+1],\sigma^h_0}(\cdot)\right|\right|_1 \\ &\leq \eta(L^*) + \beta\left(\left|\left|\pi^h(\sigma^e_1(\sigma_0)) - \sigma^h_1(\sigma^h_0)\right|\right|_1, L^*\right) \\ &\leq 2\eta(L^*) \\ &\leq R + \beta\left(\left|\left|\pi^h(\sigma_0) - \sigma^h_0\right|\right|_1, L^* + 1\right)\end{aligned} \tag{18}$$

where we have used Definition 3.3 and that $R = 2\eta(L^*)$.

Hence, (15) holds for $k \in [0, L^* + 1]$. Repeating application of steps (17) and (18) on intervals $k \in [n, L^* + n]$ for $n = 2, 3, \ldots$ then establishes that (15) holds for all $k \geq 0$. □

The importance of Theorem 3.1 is that it establishes that the additional estimation error introduced through filter approximation is asymptotically bounded by R.

Remark 3. *Theorem 3.1 does not bound the error in conditional mean estimates (extra compactness of support assumptions are required on information states).*

4. Practical Stability of Approximating Filters

Let us first introduce some assumptions on the class of approximating filters.

Preprint submitted to Automatica
Received June 20, 2012 23:15:57 PST

Definition 4.1. *(Asymptotic stability of a class of approximating filters with respect to initial conditions) For a given set* Γ*, the class of approximating filters* $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ *is said to be asymptotically stable with respect to initial conditions if there exists a* $H > 0$ *and a* $\beta(\cdot,\cdot) \in \mathscr{K}\mathscr{L}$ *such that, for all* $h \in (0, H]$*, all* $\sigma_0, \bar{\sigma}_0 \in G^h(\mathbb{R}^n)$*, all* $\omega \in \Gamma$*, and all* $k \geq 0$*, we have that*

$$\left|\left|\sigma^h_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\bar{\sigma}_0}(\cdot)\right|\right|_1 \leq \beta\left(||\sigma_0 - \bar{\sigma}_0||_1, k\right). \quad (19)$$

Definition 4.2. *(Multi-step consistency) For a given set* Γ *and projection operator* $\pi^h(\cdot)$*, the class of approximating filters* $\sigma^h_{k|[1,k],\sigma_0}(\cdot)$ *is said to be multi-step consistent with respect to the true filter* $\sigma^e_{k|[1,k],\sigma_0}(\cdot)$ *if, for each finite* $L \geq 2$ *and each* $\eta(L) > 0$*, there exists a* $H > 0$ *such that for all* $h \in (0, H]$*, all initial conditions* $\sigma_0 \in G^e(\mathbb{R}^n)$*, all* $\omega \in \Gamma$*, and all* $k \in [1, L]$*, we have that*

$$\left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\pi^h(\sigma_0)}(\cdot)\right|\right|_1 \leq \eta(L). \quad (20)$$

We will now establish a key practical stability result.

Theorem 4.1. *Consider a state process* $x_{[0,k]}$ *and a measurement process* $y_{[1,k]}$ *generated by the true system (1). For a given set* Γ *and projection operator* $\pi^h(\cdot)$*, assume that the class of approximating filters* $\sigma^h_{k|[1,k],\sigma^h_0}(\cdot)$ *is asymptotically stable with respect to initial conditions and that the class of approximating filters* $\sigma^h_{k|[1,k],\sigma^h_0}(\cdot)$ *is multi-step consistent with the true filter* $\sigma^e_{k|[1,k],\sigma_0}(\cdot)$*. Then, the class of approximating filters* $\sigma^h_{k|[1,k],\sigma^h_0}(\cdot)$ *is practically stable in the presence of modelling errors, in the sense that for any selected* $R_p > 0$*, there exists a* $H > 0$ *such that for all* $h \in (0, H]$*,*

$$\begin{aligned}&\left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\sigma^h_0}(\cdot)\right|\right|_1\\ &\quad\leq \beta\left(\left|\left|\pi^h(\sigma_0) - \sigma^h_0\right|\right|_1, k\right) + R_p\end{aligned} \quad (21)$$

for all $k \geq 0$*, all* $\omega \in \Gamma$*, and all initial conditions* $\sigma_0 \in G^e(\mathbb{R}^n), \sigma^h_0 \in G^h(\mathbb{R}^n)$*.*

Proof. For the selected R_p, let η_0 be such that $2\eta_0 \leq R_p$ (a small enough η_0 always exists). Let $L^* \geq 2$ be such that $\beta(M, L^*) \leq \eta_0$ where $M = \sup_{\sigma,\bar{\sigma}\in G^h(\mathbb{R}^n)} ||\sigma - \bar{\sigma}||_1$ (a large enough L^* always exists). Then, from the multi-step consistency assumption, there exists a $H > 0$ such that for all $h \in (0, H]$, for all $k \in [1, L^*]$, and all $\sigma_0 \in G^e(\mathbb{R}^n)$, we have $\left|\left|\sigma^e_{k|[1,k],\sigma_0}(\cdot) - \sigma^h_{k|[1,k],\pi^h(\sigma_0)}(\cdot)\right|\right|_1 \leq \eta_0$ for all $\omega \in \Gamma$. Hence, there exists a $H > 0$, such that for each $h \in (0, H]$, the approximating and true filters are L^*-error growth matched filters with $\eta(L^*) = \eta_0$. Theorem 3.1 can then be applied, for each $h \in (0, H]$, to give the bound (15), with $R = 2\eta_0$. The theorem result then follows because η_0 was chosen so that $R \leq R_p$. □

The importance of Theorem 4.1 is that if the class of approximating filters is asymptotically stable with respect to initial conditions, and multi-step consistent with the true filter (a condition that often holds), then an approximate filter can be selected that, asymptotically, has arbitrarily small error compared to the true filter.

5. Example

For $k = 0, \ldots, T-1$, consider a true and an approximating systems of the form

$$\begin{aligned}x_{k+1} &= ax_k + v_{k+1}\\ y_{k+1} &= cx_k + w_{k+1}\end{aligned} \quad (22)$$

where w_k and v_k are zero-mean Gaussian noises, and $T = 100$. Consider a true system with $a^e = 0.95, c^e = 1$, and process and measurement noises with covariances $Q^e = 1$ and $R^e = 1$, respectively. Also consider an approximating system with $a^h = 0.99, c^h = 1$, and process and measurement noises with covariances $Q^h = 1.2$ and $R^h = 1.2$, respectively.

Let us consider the set $\Gamma = \{\omega : |y_k| \leq B_m$ for all $k \in [1, T]\}$, where $B_m = 10$. Note that for all elements of Γ, $y_{[1,T]}$ are T-feasible measurement sequences of both the true and approximating models. For $\omega \in \Gamma$, $G^e(\mathbb{R}^n)$ and $G^h(\mathbb{R}^n)$ are supports for the filter recursions (6) and (11), respectively, and these supports consist of Gaussian densities with the means between -10 and 10. Here, we can select projection operator $\pi^h(\sigma) = \sigma$ for all $\sigma \in G^e(\mathbb{R}^n)$.

For the given set Γ, projection operator $\pi^h(\cdot)$, and set of functions $G^h(\mathbb{R}^n)$, it can be shown that Definition 3.1 holds with $\beta(s,t) = \min\{2, 5\exp(-0.3t)\}$ for all $s \geq 0$. The bound $\beta(\cdot,\cdot)$ was determined from examination of the worst initial conditions in $G^e(\mathbb{R}^n)$ and $G^h(\mathbb{R}^n)$ of $\hat{x}^{e-}_0 = 10, \hat{x}^{h-}_0 = -10, P^e_0 = 0, P^h_0 = 0$, where $\hat{x}^{e-}_0, \hat{x}^{h-}_0, P^{e-}_0$, and P^{h-}_0 denote initial values of the prior estimates and prior covariances for the true and approximating Kalman filters, respectively (see [26, p. 40] for Kalman filter details). We note that the Kalman filter recursion for covariance can be determined independent of the data, and solved immediately [27, Eqn 5.2.3-5]. Now let $\hat{x}^{e-}_k$ and $\hat{x}^{h-}_k$ denote the true and approximating prior mean estimates at time k, respectively, and let K^e_k and K^h_k denote the Kalman filter gains of the true and approximating filters, respectively. For the chosen set Γ, algebraic manipulation of the standard Kalman filter recursions can be used to show that the difference between the true and approximating prior mean estimates $\Delta\hat{x}^-_k = \hat{x}^{e-}_k - \hat{x}^{h-}_k$ can be bounded by

$$\begin{aligned}\left|\Delta\hat{x}^-_{k+1}\right| \leq{} & |a^e(1 - K^e_k c^e)|\left|\Delta\hat{x}^-_k\right|\\ &+ \left|a^e - a^h - a^e K^e_k c^e + a^h K^h_k c^h\right| B^h_k\\ &+ \left|a^e K^e_k - a^h K^h_k\right| B_m\end{aligned} \quad (23)$$

where $B^h_k = |a^h(1 - K^h_{k-1}c^h)|B^h_{k-1} + |a^h K^h_{k-1}|B_m$ with $B^h_0 = |\hat{x}^{h-}_0|$. Here, B^h_k is a bound on the worst prior mean estimates of the approximating filter and B_m is the worst measurements. Note that the models used in this paper

Preprint submitted to Automatica
Received June 20, 2012 23:15:57 PST

(1) and (8) have a one-step measurement delay, and hence the $\hat{x}_k^{e-}$ and $\hat{x}_k^{h-}$ quantities from the usual Kalman filter are the same as the quantities given in (2) and (9), respectively. The prior covariance information and the difference between prior means (23) can be used to determine the difference between information states initialised with the same initial condition and hence determine the $\eta(L)$ required in Definition 3.2. From our chosen $\beta(\cdot,\cdot)$ and $\eta(L)$, the smallest value of L^* such that Definition 3.3 holds is $L^* = 10$, giving $\eta(L^*) = 0.39$. Hence, for all initial conditions $\sigma_0 \in G^e(\mathbb{R}^n)$ and $\sigma_0^h \in G^h(\mathbb{R}^n)$, the conditions of Theorem 3.1 hold with $R = 0.78$, and hence the L_1-norm difference between the true and approximating systems is bounded as described by (15). In this Kalman filter mismatch problem, because covariance information does not depend on measurement sequence, it can also be shown that the error in prior mean estimates with respect to model mismatch is asymptotically bounded by 1.34. We highlight that this bound is much tighter than the bound directly implied by our set Γ or indirectly implied by the separately bounded responses of the two filters (because our bound exploits the similarity of the two filters' responses to specific measurements as described by Definition 3.2).

It is important to highlight that if the measurements are allowed to be unbounded, then the error in estimates is unbounded (this is a fundamental property of filter, as suggested by (23), and is not a limitation of our result). In practice, the presented analysis can be used to understand error properties of approximate Kalman filters during periods in which $|y_k| \leq B_m$. Due to the nature of the measurement process, periods containing sequences of bounded measurements can only be separated by a finite number of $|y_k| > B_m$ events (if B_m is large enough).

6. Conclusion

This paper establishes an asymptotic error bound on filtering performance in the situations involving model mismatch. We also present results on practical stability of a filter with respect to modelling errors. The results are established using forgetting and consistency properties, and are illustrated for a Kalman filter mismatch example.

References

[1] Rabiner, L. R., (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of IEEE, 77*(2), 257-286.

[2] Lai, J., & Ford, J. J. (2010). Relative entropy rate based multiple hidden Markov model approximation. *IEEE Transactions on Signal Processing, 58*(1), 165-174.

[3] Barniv, Y., & Kella, O. (1987). Dynamic programming solution for detecting dim moving targets. Part II: Analysis. *IEEE Transactions on Aerospace and Electronic Systems, AES-23*(6), 776-788.

[4] Jazwinski, A. H. (1970). *Stochastic processes and filtering theory.* New York: Academic Press.

[5] Anderson, B. D. O., & Moore, J. B. (1979). *Optimal filtering.* Englewood Cliffs: Prentice-Hall.

[6] Ocone, D., & Pardoux, E. (1996). Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM Journal on Control and Optimization, 34*(1), 226-243.

[7] Dey, S., & Charalambous, C. D. (2000). On asymptotic stability of continuous-time risk-sensitive filters with respect to initial conditions. *Systems & Control Letters, 41*(1), 9-18.

[8] Budhiraja, A. (2001). Ergodic properties of the nonlinear filter. *Stochastic Processes and their Applications, 95*(1), 1-24.

[9] Budhiraja, A., & Ocone, D. (1999). Exponential stability in discrete-time filtering for non-ergodic signals. *Stochastic Processes and their Applications, 82*(2), 245-257.

[10] Clark, J. M. C., Ocone, D. L., & Coumarbatch, C. (1999). Relative entropy and error bounds for filtering of Markov processes. *Mathematics of Control, Signals, and Systems, 12*(4), 346-360.

[11] Le Gland, F., & Mevel, L. (2000). Exponential forgetting and geometric ergodicity in hidden Markov models. *Mathematics of Control, Signals, and Systems, 13*(1), 63-93.

[12] Crisan, D., & Heine, K. (2008). Stability of the discrete time filter in terms of the tails of noise distributions. *Journal of the London Mathematical Society, 78*(2), 441-458.

[13] Azimi-Sadjadi, B., & Krishnaprasad, P. S. (2005). Approximate nonlinear filtering and its application in navigation. *Automatica, 41*(6), 945-956.

[14] Heine, K., & Crisan, D. (2008). Uniform approximations of discrete-time filters. *Advances in Applied Probability, 40*(4), 979-1001.

[15] Reif, K., Gunther, S., & Yaz, E. (1999). Stochastic stability of the discrete-time extended Kalman filter. *IEEE Transactions on Automatic Control, 44*(4), 714-728.

[16] Zakai, M., & Ziv, J. (1972). Lower and upper bounds on the optimal filtering error of certain diffusion processes. *IEEE Transactions on Information Theory, 18*(3), 325-331.

[17] Bobrovsky, B., & Zakai, M. (1976). A lower bound on the estimation error for certain diffusion processes. *IEEE Transactions on Information Theory, 22*(1), 45-52.

[18] Barak, H. (1986). A lower bound on the integral of the nonlinear filtering error. *Systems and Control Letters, 7*(2), 109-116.

[19] Zeitouni, O. (1984). On the tightness of some error bounds for the nonlinear filtering problems. *IEEE Transactions on Automatic Control, 29*(9), 854-857.

[20] Nešić, D., Teel, A. R., & Kokotović, P. V. (1999). Sufficient conditions for stabilization of sampled-data nonlinear systems via discrete-time approximations. *Systems & Control Letters, 38*(4-5), 259-270.

[21] Billingsley, P. (1995). *Probability and Measure* (3rd ed.). New York: Wiley.

[22] Elliott, R. J., Aggoun, L., & Moore, J. B. (1995). *Hidden Markov models: Estimation and control.* (1st ed.). Berlin: Spring-Verlag.

[23] Luenberger, D. G. (1969). *Optimization by vector space methods.* New York: Wiley.

[24] James, M. R., Baras, J. S., & Elliott, R. J. (1994). Risk-sensive control and dynamics games for partially observed discrete-time nonlinear systems. *IEEE Transactions on Automatic Control, 39*(4), 780-792.

[25] Khalil, H. K. (2002), *Nonlinear Systems* (3rd ed.). Upper Saddle River, N.J.: Prentice-Hall.

[26] Brown, R. G., & Hwang, P. Y. C. (1997). *Introduction to random signals and applied Kalman filtering with Matlab exercises and solutions* (3rd ed.). New York: Wiley.

[27] Bar-Shalom, Y., Li, X. R., & Kirubarajan, T. (2001). *Estimation with applications to tracking and navigation.* New York: Wiley.

Preprint submitted to Automatica
Received June 20, 2012 23:15:57 PST