

EXPONENTIAL STABILITY OF FILTERS AND SMOOTHERS FOR HIDDEN MARKOV MODELS

L. Shue[†], B. D. O. Anderson and S. Dey

*Department of Systems Engineering
and Cooperative Research Centre for Robust and Adaptive Systems,
Research School of Information Sciences and Engineering,
Australian National University,
Canberra ACT 0200, Australia.*

[†]*Fax: +61-6-279-8688 and e-mail: louis.shue@syseng.anu.edu.au*

Keywords : Stochastic, Stability, Estimation

Abstract

In this paper, we address the problem of exponential stability of filters and fixed-lag smoothers for discrete-time and discrete-state Hidden Markov Models (HMMs). By appealing to a generalised Perron-Frobenius result for nonnegative matrices, we demonstrate exponential forgetting for both the recursive filters and smoothers, and obtain overbounds on the rate of forgetting. Simulation studies are carried out to substantiate the results.

1 Introduction

The purpose of this paper is to explain how some important results in Kalman filtering and smoothing can be carried over to contemporary problems involving Hidden Markov Models (HMMs). Presently, our attention is restricted to discrete-time models, with finite and discrete state and observation sets; tackling the problem for continuous-time models involve new tools which we are seeking to develop.

The problems we consider are of two kinds, but related. Most Kalman filters have an exponential stability property [4], even in nonstationary situations. This ensures that, if the filter is initialised at some finite time, the initial conditions (or equivalently, old measurements) are forgotten exponentially fast, thus the remote past cannot significantly influence the present. The absence of this property in a given filter is likely to lead to unreliable estimates, as round-off errors may overpower the measurements.

The first results of the paper extend the exponential forgetting result to HMMs. Similar results along these lines can be found in [1], and, of a more preliminary nature, in [6]. Our results are probably more comprehensive, appealing to more recent developments in properties of products

of positive and non-negative matrices [8], and allowing the calculation of convergence rates.

The second main thrust of the paper is to deal with fixed-lag smoothing, where measurements up to the present time are used to infer information about the state Δ time intervals in the past (Δ being a fixed quantity). In smoothing problems which can be tackled using Kalman filter ideas, it has been found that as Δ increases, there is less and less additional benefit gained in terms of the quality of the estimates. The fall-off in the rate of increase of benefit is exponential, and with Δ equal to four or five times the dominant time constant of the Kalman filter, practically all the benefit derivable from smoothing is achieved [2].

In this paper, we establish equations for the HMM fixed-lag smoother. Again, we are able to show that the rate of increase of benefit from smoothing falls off exponentially as Δ increases, with the same rate as the forgetting rate of the HMM filter, which is, furthermore, at least as fast as that of the original hidden Markov signal model. The results were predicted, and the subject of simulation study, more than twenty years ago [7]. They constitute important guidelines for the use of fixed-lag smoothing, as opposed to filtering, of HMMs.

In Section 2, we define the signal model, and demonstrate exponential forgetting of initial conditions for recursive filters. In Section 3, the filtering ideas are extended to smoothed estimation scheme. Section 4 contains some simulation results and discussions, in which the performance of filtered and smoothed estimates for a two-state HMM are compared. In Section 5 we present some concluding remarks.

2 Filtering

In this section we will consider the problem of filtering. Formally, a filtered estimate is a conditional estimate of the state of the HMM, given a series of measurements.

2.1 Signal Model

Consider a first order discrete-time and discrete-state Markov process $X_k = \{1, 2, 3, \dots, N\}$, k denoting time. At each time instant k , a corresponding signal $Y_k = \{1, 2, \dots, N\}$ is observed, again having discrete values. We will adopt the convention that a lower-case x_k denotes the actual state value, and likewise for y_k . The probability vectors for X_k and Y_k are updated by the system matrices A and C , where $A = \{a_{ij}\} = \{\Pr(X_{k+1} = i | X_k = j)\}$ and $C = \{c_{mn}\} = \{\Pr(Y_k = m | X_k = n)\}$, and $\sum_{i=1}^N a_{ij} = 1$, $\sum_{m=1}^M c_{mn} = 1$. Further, unless otherwise stated, $a_{ij} > 0$ and $c_{mn} > 0$, $\forall i, j, n \in \{1, 2, \dots, N\}$, $\forall m \in \{1, 2, \dots, M\}$.

2.2 Evolution of Filtered Distributions

Let $\Pi_{k|k}$ and $\Pi_{k+1|k}$ be the filtered and one-step prediction probability vector, with the i -th entry being $\Pr(X_k = i | Y_0, Y_1, \dots, Y_k)$ and $\Pr(X_{k+1} = i | Y_0, Y_1, \dots, Y_k)$ respectively. The time evolution relations for the filtered probability vector are

$$\Pi_{k+1|k} = A \Pi_{k|k} \quad (1)$$

$$\Pi_{k+1|k+1} = \frac{1}{1'_N C_{y_{k+1}} \Pi_{k+1|k}} C_{y_{k+1}} \Pi_{k+1|k} \quad (2)$$

where $1'_N C_{y_{k+1}} \Pi_{k+1|k} = [1 \dots 1] C_{y_{k+1}} \Pi_{k+1|k}$ is a scalar normalising constant to ensure the entries of $\Pi_{k+1|k+1}$ sum to 1, and $C_{y_{k+1}} = \text{diag}(c_{11} \ c_{12} \ c_{13} \ \dots \ c_{1N})$, when $y_{k+1} = l$.

2.3 Exponential Forgetting

By iterating (1) and (2), the filtered probability vector at time k can be expressed in terms of an arbitrarily chosen initial distribution $\Pi_{0|0}$:

$$\begin{aligned} \Pi_{k|k} &= \frac{C_{y_k} A \Pi_{k-1|k-1}}{1'_N C_{y_k} A \Pi_{k-1|k-1}} \\ &= \frac{(C_{y_k} A C_{y_{k-1}} A \dots C_{y_1} A) \Pi_{0|0}}{1'_N C_{y_k} A C_{y_{k-1}} A \dots C_{y_1} A \Pi_{0|0}} \\ &= \frac{T_{1,k} \Pi_{0|0}}{1'_N T_{1,k} \Pi_{0|0}} \end{aligned} \quad (3)$$

where

$$T_{1,k} = C_{y_k} A C_{y_{k-1}} A \dots C_{y_1} A \quad (4)$$

We now proceed to derive initial-condition forgetting of the filter by appealing to the generalised Perron-Frobenius result [8] for an inhomogeneous product of matrices. Broadly, this theorem states that, under certain conditions, a product of positive matrices, of the form $T_{1,r} = H_r H_{r-1} \dots H_1$, may become dyadic¹ as $r \rightarrow \infty$.

As stated previously, $A > 0$ and $C > 0$, therefore (4), a product of successive $C_y A$, is strictly positive, and the

¹The term dyadic means that a matrix is of rank 1.

requirements of the aforementioned theorem are automatically satisfied. This means that, as $k \rightarrow \infty$,

$$T_{1,k} \rightarrow U(k) V' \quad (5)$$

for some positive column vector $U(k)$, and positive row vector V' . Without loss of generality, let $v_1 = 1$. Hence the normalised filter probability vector becomes

$$\begin{aligned} \Pi_{k|k} &= \frac{1}{1'_N T_{1,k} \Pi_{0|0}} T_{1,k} \Pi_{0|0} \\ &\rightarrow \frac{1}{1'_N T_{1,k} \Pi_{0|0}} \begin{bmatrix} u_1(k) \\ u_2(k) \\ \vdots \\ u_n(k) \end{bmatrix} (1 \ v_2 \ v_3 \ \dots \ v_n) \Pi_{0|0} \\ &\rightarrow \frac{1}{N} U(k) \\ &\quad \sum_{i=1}^N u_i(k) \end{aligned}$$

This is independent of the initial distribution $\Pi_{0|0}$. The rate of convergence of (5) is exponential and rates are computable [8].

Similar convergence properties also exist when the conditions $A > 0$ and $C > 0$ are relaxed. In particular, the current results can be applied to the following cases:

1. $C > 0$, $A \geq 0$ and $A > 0$ failing but primitive², and
2. $A > 0$, $C \geq 0$ and $C > 0$ failing.

3 Smoothing

Smoothing is an extension of filtering, in the sense that the conditional probability vector for X_j uses more measurements, not just up to time j , but till some later time $k > j$. Since more measurements are used, better estimates should result, although there is an inherent delay before the smoothed estimates are available.

In this section, we will implement two fixed-lag smoothing schemes for an HMM by formulating the problem in terms of fixed-point smoothing, and subsequently varying the fixed index j to obtain a fixed-lag smoothing scheme. The first approach involves the construction of a fictitious augmented state model, as outlined in [4, 7]; embedded within the filtered probability vector for the augmented HMM is the smoothed probability vector for the original HMM. In the second method, the smoothed estimates consist of merging the filtered probabilities from forward- and backward-models, using a procedure outlined in [5].

3.1 Smoother From Augmented Signal Model

Definition 3.1 For each $k \geq j$, let $\mathcal{Z}_k = Z_{j,k} = (X_j \ X_k)'$ be an augmented state vector, consisting of the states of

²Primitivity means that $A^k > 0$, for $k > 0$

the original Markov process as defined in Section 2.1, at a fixed time j , and a variable time k .

From Definition 3.1, it can be seen that \mathcal{Z}_k can only assume the values $(1, 1)$, $(1, 2)$, \dots , $(1, N)$, $(2, 1)$, \dots , (N, N) . By denoting the probability vector that \mathcal{Z}_k is in each of the N^2 possible states at time k as \mathcal{P}_k , it is seen that these hold:

$$\mathcal{P}_{k+1} = \mathcal{A}\mathcal{P}_k$$

where

$$\mathcal{A} = \begin{bmatrix} A & 0 & \dots & 0 \\ 0 & A & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & A \end{bmatrix} = I_N \otimes A$$

Further

$$\mathcal{P}_j = \begin{bmatrix} e_1 \Pr(X_j = 1) \\ e_2 \Pr(X_j = 2) \\ \vdots \\ e_N \Pr(X_j = N) \end{bmatrix}$$

where e_i is a unit vector, and recalling that \mathcal{P}_j is an $(N \times 1)$ column vector.

By arguing that the output process of the original HMM with state X_k can also be regarded as the output process of an HMM with state \mathcal{Z}_k , for $k \geq j$, the observation matrix \mathcal{C} for the augmented state is

$$\mathcal{C} = [C \ C \ \dots \ C] = 1'_N \otimes C$$

and, similarly

$$\mathcal{C}_{y_{k+1}} = \begin{bmatrix} C_{y_{k+1}} & 0 & \dots & 0 \\ 0 & C_{y_{k+1}} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & C_{y_{k+1}} \end{bmatrix} = I_N \otimes C_{y_{k+1}}$$

The filtered probability vector for \mathcal{Z}_k , denoted as $\hat{\Pi}_{j,k|k}$, evolve according to the following recursions:

$$\hat{\Pi}_{j,k+1|k} = \mathcal{A}\hat{\Pi}_{j,k|k} \quad (6)$$

$$\hat{\Pi}_{j,k+1|k+1} = \frac{1}{1'_{N^2} \mathcal{C}_{y_{k+1}} \hat{\Pi}_{j,k+1|k}} \mathcal{C}_{y_{k+1}} \hat{\Pi}_{j,k+1|k} \quad (7)$$

where $1'_{N^2} \mathcal{C}_{y_{k+1}} \hat{\Pi}_{j,k+1|k}$ is a scalar normalising constant.

Since, by definition, the i -th entry of the smoothed probability vector $\Pi_{j|j+\Delta}$ at time j with lag Δ for the unaugmented HMM is just

$$\begin{aligned} & \Pr(X_j = i | Y_0, \dots, Y_{j+\Delta}) \\ &= \sum_{l=1}^N \Pr(X_j = i, X_{j+\Delta} = l | Y_0, \dots, Y_{j+\Delta}), \end{aligned}$$

the smoothed probability vector $\Pi_{j|j+\Delta}$ for the unaugmented HMM can be evaluated by summing appropriate

terms in the filtered probability vector for the augmented model:

$$\begin{aligned} \Pi_{j|j+\Delta} &= \begin{bmatrix} 1 \dots 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \dots 1 \end{bmatrix} \hat{\Pi}_{j,j+\Delta|j+\Delta} \\ &= (I_N \otimes 1'_N) \hat{\Pi}_{j,j+\Delta|j+\Delta} \\ &= \frac{[(I_N \otimes 1'_N)(\mathcal{C}_{y_{j+\Delta}} \mathcal{A} \mathcal{C}_{y_{j+\Delta-1}} \mathcal{A} \dots \mathcal{C}_{y_{j+1}} \mathcal{A})] \hat{\Pi}_{j,j|j}}{\Theta_{j,j+\Delta}} \\ &= \frac{[I_N \otimes (1'_N T_{j+1,j+\Delta})] \hat{\Pi}_{j,j|j}}{\Theta_{j,j+\Delta}} \end{aligned} \quad (8)$$

where $\Theta_{j,j+\Delta} = 1'_{N^2} [I_N \otimes (1'_N T_{j+1,j+\Delta})] \hat{\Pi}_{j,j|j}$, and

$$T_{j+1,j+\Delta} = C_{y_{j+\Delta}} \mathcal{A} C_{y_{j+\Delta-1}} \mathcal{A} \dots C_{y_{j+1}} \mathcal{A} \quad (9)$$

Note that the property $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ has been used, and

$$\begin{aligned} \hat{\Pi}_{j,j|j} &= \begin{bmatrix} \Pr(X_j = 1, X_j = 1 | Y_0, \dots, Y_j) \\ \Pr(X_j = 1, X_j = 2 | Y_0, \dots, Y_j) \\ \vdots \\ \Pr(X_j = 2, X_j = 1 | Y_0, \dots, Y_j) \\ \vdots \\ \Pr(X_j = N, X_j = N-1 | Y_0, \dots, Y_j) \\ \Pr(X_j = N, X_j = N | Y_0, \dots, Y_j) \end{bmatrix} \\ &= [e_1 \Pi_{j|j}(1) \ \dots \ e_N \Pi_{j|j}(N)]' \end{aligned} \quad (10)$$

with $\Pi_{j|j}(i)$ denoting the i -th entry in $\Pi_{j|j}$.

With j fixed, (9) has the same structure as (4), hence as $\Delta \rightarrow \infty$,

$$T_{j+1,j+\Delta} \rightarrow U(\Delta) V' \quad (11)$$

Furthermore, the same rate of convergence as for the filter determines how fast $T_{j+1,j+\Delta}$ becomes dyadic.

Remark 3.1 Disregarding the j -dependence, $U(\Delta) = \{u_i(\Delta)\}$ is a Δ -dependent term, whereas $V = \{v_i\}$ is a vector of constants, $\forall i \in \{1, 2, \dots, N\}$.

The unnormalised smoothed probability vector becomes

$$\begin{aligned} \tilde{\Pi}_{j|j+\Delta} &= \left[I_N \otimes \left(\sum_{i=1}^N u_i(\Delta) \right) V' \right] \hat{\Pi}_{j,j|j} \\ &= \left[\sum_{i=1}^N u_i(\Delta) \right] \begin{bmatrix} v_1 \Pi_{j|j}(1) \\ v_2 \Pi_{j|j}(2) \\ \vdots \\ v_N \Pi_{j|j}(N) \end{bmatrix} \end{aligned}$$

Once normalised, it can be seen that the Δ -dependent terms are cancelled, thus establishing the Δ -independence of the smoothing equation for large lags, or Δ 's. This means that all the improvement that can be gained by smoothing is attained after some finite Δ , and no significant gains can be made as the lag is extended further.

3.2 Smoother From Forward-Backward Filters

3.2.1 Backwards Model

Following ideas in [3], an analogous backward Markov state process and HMM, which evolves backwards in time, can be constructed from a forward Markov state process.

Definition 3.2 For a given forward Markov state process, characterised by a state transition probability matrix A , the associated backward process has a system matrix A^b with elements a_{ij}^b , such that

$$\begin{aligned} a_{ij}^b &= \Pr(X_k = i | X_{k+1} = j) \\ &= \Pr(X_k = i) \Pr(X_{k+1} = j | X_k = i) \Pr(X_{k+1} = j)^{-1} \end{aligned}$$

In other words, and further assuming stationarity so that $\Pr(X_k = i) = \Pr(X_{k+1} = i) = \Pr(X = i)$,

$$A^b = \Lambda A' \Lambda^{-1}$$

where $\Lambda = \text{diag}(\Pi)$, and Π is the steady-state distribution, such that $A\Pi = \Pi$.

3.2.2 Smoother from Forward-Backward Model

Definition 3.3 Denote a reverse filter probability and the associated analog of the one-step prediction probability vector by $\Pi_{k|k}^+$ and $\Pi_{k|k+1}^+$ respectively, with the i -th entry being $\Pr(X_k = i | Y_k, Y_{k+1}, \dots, Y_{k+\Delta})$ and $\Pr(X_k = i | Y_{k+1}, Y_{k+1}, \dots, Y_{k+\Delta})$, where $k + \Delta$ is the maximum time at which measurements are available.

In accordance with Definition 3.3, $\Pi_{k+\Delta|k+\Delta}^+(i) = \Pr(X_{k+\Delta} = i | Y_{k+\Delta})$ and $\Pi_{k+\Delta|k+\Delta+1}^+(i) = \Pr(X_{k+\Delta} = i)$, since $k + \Delta$ is the maximum time at which measurements are available. By further assuming stationarity, it can be seen that

$$\Pi_{k+\Delta|k+\Delta+1}^+(i) = \Pr(X = i)$$

The reverse-time filter obeys the following updating equations:

$$\Pi_{k|k+1}^+ = A^b \Pi_{k+1|k+1}^+ \quad (12)$$

$$\Pi_{k|k}^+ = \frac{1}{1'_N C_{y_k} \Pi_{k|k+1}^+} C_{y_k} \Pi_{k|k+1}^+ \quad (13)$$

Using the fact that the sequences (Y_0, Y_1, \dots, Y_j) and $(Y_{j+1}, Y_{j+2}, \dots, Y_{j+\Delta})$ are conditionally independent given X_j [5], the smoothed probability vector can be expressed as a product of two terms, the forward and reverse filtered probability vectors:

$$\begin{aligned} \Pr(X_j = i | Y_0, \dots, Y_{j+\Delta}) &= \frac{\Pr(X_j = i | Y_0, \dots, Y_j)}{\Theta \Pr(X_j = i)} \times \\ &\Pr(X_j = i | Y_{j+1}, \dots, Y_{j+\Delta}) \quad (14) \end{aligned}$$

where Θ is a normalising constant.

From (12), (13) and (14), $\Pi_{j|j+\Delta}(i)$ can be rewritten as

$$\begin{aligned} &\Theta \Pr(X_j = i) \Pi_{j|j+\Delta}(i) \\ &= \Pi_{j|j}(i) \Pi_{j|j+1}^+(i) \\ &= \Pi_{j|j}(i) \left[A^b C_{y_{j+1}} A^b \dots A^b C_{y_{j+\Delta}} \Pi_{j+\Delta|j+\Delta+1}^+ \right](i) \\ &= \Pi_{j|j}(i) \left[T_{j+1, j+\Delta}^b \Pi_{j+\Delta|j+\Delta+1}^+ \right](i) \quad (15) \end{aligned}$$

with $\Pi_{j|j+\Delta}(i)$ denoting the i -th entry of $\Pi_{j|j+\Delta}$, etc.

The Δ -dependence lies entirely in $T_{j+1, j+\Delta}^b$; the apparent Δ -dependence in $\Pi_{j+\Delta|j+\Delta+1}^+$ will be lost due to the initialising process of the reverse filter. Now, since A^b and C both are strictly positive, as $\Delta \rightarrow \infty$

$$\begin{aligned} T_{j+1, j+\Delta}^b &= A^b C_{y_{j+1}} A^b C_{y_{j+2}} \dots A^b C_{y_{j+\Delta}} \\ &= (\Lambda A' \Lambda^{-1}) C_{y_{j+1}} (\Lambda A' \Lambda^{-1}) C_{y_{j+2}} \dots C_{y_{j+\Delta}} \\ &= \Lambda (A' C_{y_{j+1}} A' C_{y_{j+2}} \dots A' C_{y_{j+\Delta}}) \Lambda^{-1} \\ &\rightarrow \Lambda (U(\Delta) V')' \Lambda^{-1} \\ &\equiv U^b V^{b'}(\Delta) \end{aligned}$$

The third equality follows from the commutativity of diagonal matrices. By the same arguments in Section 3.1, the Δ -dependency in the smoothed estimate (15) can again be removed after normalisation.

4 Simulations of a 2-state HMM

The basic system investigated is a two-state HMM, with a symmetric A matrix (also known as random telegraph wave). The simulations aim to illustrate the benefits of smoothing over filtering in environments with different signal and noise contents. Results were obtained for two cases: fixing A and varying C , and then fixing C and varying A . The respective fixed matrices are: $A = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$

$$, C = \begin{bmatrix} 0.1 & 0.8 \\ 0.9 & 0.2 \end{bmatrix}.$$

The lag required for the smoothing error variance to reach steady-state can be estimated by using the so-called Birkhoff coefficient (see [8]), denoted as $\tau_B(\cdot)$, which is a measure of how close a given matrix is to having rank 1. This means that

$$\tau_B(T_{j+1, j+\Delta}) \leq [\tau_B(T_{j+1, j+\Delta})_{\text{ob}}]^{\Delta-1}$$

where $\tau_B(T_{j+1, j+\Delta})_{\text{ob}}$ provides an overbound on the $\tau_B(\cdot)$ of the individual terms which make up the product $T_{j+1, j+\Delta}$. Using the fact that $\tau_B(\cdot)$ of a positive diagonal matrix is 1, an approximate bound is

$$\tau_B(T_{j+1, j+\Delta})_{\text{ob}} = \tau_B(A),$$

whereas a tighter bound can be found by grouping three terms at a time in $T_{j+1, j+\Delta}$, and is

$$\ln \tau_B(T_{j+1, j+\Delta})_{\text{ob}} = \frac{1}{2} E[\ln \tau_B(AC_{y_k} A)]$$

The criteria for convergence, or $T_{j+1,j+\Delta}$ having rank 1, is to determine $\Delta = \Delta_{crit}$, such that $\tau_B(\cdot)_{ob}^{\Delta_{crit}/4} = 1/e$. The associated overbounds $\tau_B(\cdot)_{ob}$ and the expected lag before convergence have been listed in Table 1.

Lastly, in analogy with the Kalman filtering case, we define the error variance to be

$$\frac{1}{2}E[(\Pi_k - \Pi_{k|k})'(\Pi_k - \Pi_{k|k})|\mathcal{Y}_k] = \frac{1}{2} \left[1 - \Pi_{k|k}'\Pi_{k|k} \right]$$

where $\Pi_k(i) = \Pr(X_k = i)$ and $\mathcal{Y}_k = \{Y_0, Y_1, \dots, Y_k\}$.

4.1 A constant, C variable

In this series of simulations, the C matrices used are:

$$C_1 = \begin{bmatrix} 0.1 & 0.8 \\ 0.9 & 0.2 \end{bmatrix}, \quad C_2 = \begin{bmatrix} 0.45 & 0.4 \\ 0.55 & 0.6 \end{bmatrix}$$

In Figs. 1-2, there is rapid initial improvement in the smoothing error which then reaches some steady-state value. The rate of convergence with Δ of the smoothing error³ is relatively insensitive to variations in C , in accordance with the fact that $\tau_B(\cdot)$ of a diagonal matrix is 1. From Table 1, it can be seen that Δ_{crit} provides an adequate, albeit slightly pessimistic, bound to the maximum Δ before convergence.

As C becomes ‘whiter’ (ie. given a certain x_k , each possible y_k is equally likely, which means columns in C are identical), the magnitude of the errors for both the filtered and smoothed estimates increases and approaches the maximum value of 0.25. This means that in environments with high measurement noise, it is difficult to obtain better estimates, by either filtering or smoothing, than estimating via steady-state distributions alone.

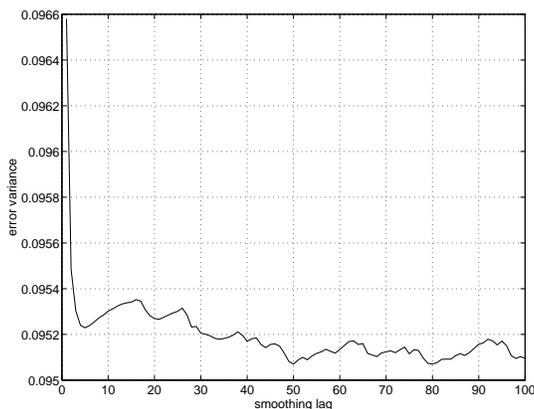


Figure 1: Variation of smoothing error with Δ : A fixed, C_1 ; filter error: 0.1094

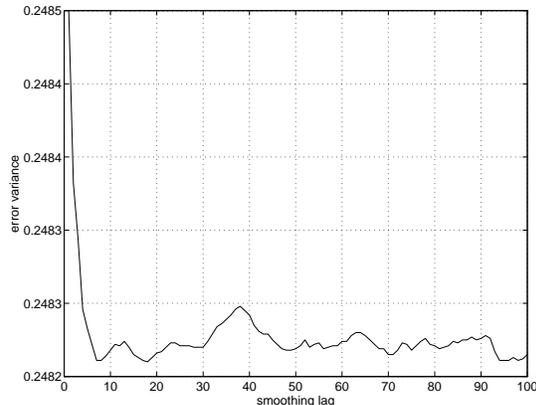


Figure 2: Variation of smoothing error with Δ : A fixed, C_3 ; filter error: 0.2487

4.2 A variable, C fixed

This series of simulations illustrate the dependence of smoothing/filtering on the correlation between X_k 's from one time instant to the next. The following A matrices have been used:

$$A_1 = \begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix}$$

As the X process varies from an almost deterministic system (A has a dominant entry in every row) to close to becoming a white process (A having identical columns), the minimum Δ required for steady-state decreases (Fig. 3-4), until any gain/loss by smoothing is almost instantaneous. This is indicated by the increased fluctuations in Fig. 4, and also confirmed by the decrease in Δ_{crit} (see Table 1), which is a measure of how readily previous states are forgotten.

| | $\tau_B(T_{j+1,j+\Delta})_{ob}$ | | Δ_{crit} | |
|--------------------|---------------------------------|----------------|-----------------|----------------|
| | approx. bound | 3-term average | approx. bound | 3-term average |
| (A fixed) C_1 | 0.8 | 0.729 | 17.9 | 12.7 |
| C_2 | 0.8 | 0.800 | 17.9 | 17.9 |
| (C fixed) A_1 | 0.9 | 0.861 | 38.0 | 26.6 |
| A_2 | 0.2 | 0.144 | 2.49 | 2.06 |

Table 1: Overbounds on $\tau_B(T_{j+1,j+\Delta})$ and the expected smoothing lag before convergence.

5 Conclusions

Natural extensions of the present work includes a more complete treatment of systems with both A and C both being nonnegative, investigations of continuous-time models, and also continuous-state/output HMMs, which

³The same arguments apply equally to the rate of forgetting of initial-conditions for filters.

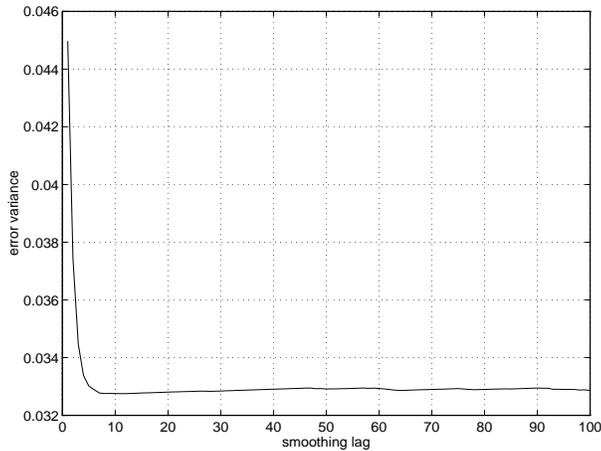


Figure 3: Variation of smoothing error with Δ : C fixed, A_1 ; filter error: 0.0644

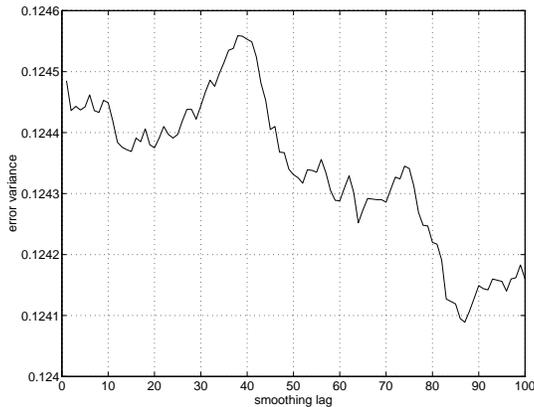


Figure 4: Variation of smoothing error with Δ : C fixed, A_4 ; filter error: 0.1258

makes possible direct comparisons between the performance of the filters/smoothers with SNR of the system. However, for continuous-time systems, an analog of the limiting theorem for a product of nonnegative matrices is required. Other indirect applications include the Viterbi algorithm, where the concept of a Δ_{crit} may help to guide the choice of truncation point, and hybrid systems which involve HMMs such as arising from, for example, multi-target tracking.

Acknowledgements:

The authors wish to acknowledge the funding of the activities of the Cooperative Research Centre for Robust and Adaptive Systems by the Australian Commonwealth Government under the Cooperative Research Centres Program.

References

- [1] A. Arapostathis and S. I. Marcus, "Analysis of an identification algorithm arising in the adaptive estimation of Markov chains", *Mathematics of Control, Signals and Systems*, vol. 3, pp. 1-29, (Jan 1990).
- [2] B. D. O. Anderson, "Properties of optimal linear smoothing", *IEEE Trans. Automatic Control*, vol. AC-14, no. 11, pp. 114-115, (Feb 1969).
- [3] B. D. O. Anderson and T. Kailath, "Forwards and backwards models for finite-state Markov processes", *Advanced Applied Probability*, vol. 11, pp. 118-133, (1979).
- [4] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice Hall, Inc. , (1979).
- [5] B. D. O. Anderson and I. Rhodes, "Smoothing algorithms for nonlinear finite-dimensional systems", *Stochastics*, vol. 9, pp. 139-165, (1983).
- [6] R. K. Boel, J. B. Moore and S. Dey, "Geometric convergence of filters for Hidden Markov Models", *Proceedings of the 34th Conference on Decision and Control*, pp. 69-74, (1995).
- [7] D. Clements and B. D. O. Anderson, "A nonlinear fixed-lag smoother for finite-state Markov processes", *IEEE Trans. Information Theory*, vol. 2, pp. 446-452, (1975).
- [8] E. Seneta, *Non-negative matrices and Markov Chains, 2nd Edition*, Chapters 3-4, Springer-Verlag, (1981).