# Reduced Complexity On-line Estimation of Hidden Markov Model Parameters.

J.B. Moore[1] and J.J. Ford[1]

[1] Department of Systems Engineering, Research School of Information Sciences and Engineering, Australian National University, Canberra, ACT 0200, Australia.

**Abstract**

In this paper we propose and study low complexity algorithms for on-line estimation of hidden Markov model (HMM) parameters. The estimates approach the true model parameters as the measurement noise approaches zero, but otherwise give improved estimates, albeit with bias. On a finite data set in the high noise case, the bias may not be significantly more severe than for a higher complexity asymptotically optimal scheme. Our algorithms require $O(N^3)$ calculations per time instant, where $N$ is the number of states. Previous algorithms based on earlier hidden Markov model signal processing methods, including the expectation-maximumisation (EM) algorithm require $O(N^4)$ calculations per time instant.

## 1 Introduction

A discrete time homogeneous Markov chain, with a finite state space having $N$ elements, is supposed observed in noise. Such a situation is termed a hidden Markov model or HMM. A recent treatment can be found in Elliott[3].

In addition to estimating the state of the chain given the observations, it is often of importance to estimate the state values, transition probability matrix and the noise characteristics, see Collings[2], Elliott[3], Ford[4, 5] and Moore[12, 13]. To estimate the transition matrix of the Markov chain using the well established classical Baum-Welch algorithm and related expectation-maximisation (EM) algorithms based on data up until time $T$ and initial model parameter estimates, it is usual to estimate the number of jumps $\mathcal{J}_T^{ij}$ of the chain from state $i$ to state $j$, for $1 \leq i, j \leq N$, up to the time $T$ and the occupation times $\mathcal{O}_T^i$ is state $i$ for $1 \leq i \leq N$. Here $N$ is the number of states in the model. Then $\mathcal{J}_T^{ij}(\mathcal{O}_T^i)^{-1}$ gives an improved estimate in a likelihood sense for the probability of the state switching from $i$ to $j$, ie. $a_{ij}$. To estimate other model parameters involved in the measurement equation, related transitions $\mathcal{T}_T^i$ from states $i$ to outputs are required. Using the improved model parameters the process is repeated until convergence to a (local) maximum of the likelihood function. The EM algorithm re-estimates model parameters using forward and backward passes through the data, and so is not really an on-line scheme. Also, the theory gives only local convergence to a local maximum of the likelihood function.

In Moore[12], almost-sure consistent convergent parameter estimators are proposed for estimating hidden Markov model parameters online. The almost-sure asymptotic convergence results are obtained via standard martingale convergence results, refer to Meyer[11] and Neveu[14], and the Kronecker lemma, refer to Loeve[10] and Neveu[14], stability properties for HMM filters are used, see Dey[1], Shue[17], and ordinary differential equation theory for stochastic approximation, see Ljung[9], Kushner[7], Gerencsér[6] and LeGland[8]. The optimal parameter estimates converges almost surely to the true parameter values Of course, there must be persistence of excitation in the models, (and estimators), to achieve this property. The results presented in Moore[12] contrast the Baum-Welch re-estimation results which are only guaranteed to converge to a local maximum of the likelihood function.

Here, we propose reduced complexity estimation schemes based on the consistent schemes in Moore[12] with the view to reducing (additional) computational effort at each step from $O(N^4)$ to $O(N^3)$ The schemes become consistent as the noise level approaches zero, but otherwise giving improved estimates, albeit with bias.

In Section 2, the discrete-time HMM signal model is defined, on-line estimators are introduced, and associated convergence properties reviewed, including the almost sure convergence result from Moore[12]. In Section 3, reduced complexity algorithms are proposed. Simulations are given in Section 4, and conclusions in Section 5.

## 2 Dynamics, Measure Change, and Martingale Properties

### *2.1 Dynamics*

Our time parameter set is the non-negative integers $Z^+ = \{0, 1, 2, \ldots\}$. On a probability space $(\Omega, \mathcal{F}, P)$ we suppose we have a finite state, time homogeneous Markov chain $\{X_\ell\}, \ell \in Z^+$. As pointed out in Elliott[3], without loss of generality we can take the state space of the Markov chain to be the set $S = \{e_1, e_2, \ldots, e_n\}$ of unit vectors in $\mathbb{R}^N$. Here $e_i = (0, 0, \ldots, 1, \ldots, 0)' \in \mathbb{R}^N$.

Consequently, at each $\ell, X_\ell \in S$. Consider the state space model

$$X_k = AX_{k-1} + V_k \tag{1}$$
$$y_k = (C'X_{k-1}) + (D'X_{k-1})w_k. \tag{2}$$

Here $A \in \mathbb{R}^{N\times N}$ is the stochastic matrix with $\underline{1}'A = \underline{1} := [1, 1, \ldots, 1]'$ elements $a_{ij} > 0$, $1 \le i, j \le N$, and satisfying $E[X_k] = AE[x_{k-1}]$, $a_{ij} = P(X_\ell | X_{\ell-1} = e_j)$. Also $V_k$ is a martingale increment satisfying

$$E\left[V_{k+1}|X_k\right] = 0 \in \mathbb{R}^N.$$

Here, $w_k$ is a sequence of *i.i.d.* $N(0, 1)$ random variables defined on $(\Omega, \mathcal{F}, P)$. That the measurements $y_k \in \mathbb{R}$ are linear in $X_{k-1}$ is not really a restriction since nonlinear operations $f(X)$ are linear in $X$ as with f=$(f(e_1), f(e_2), \ldots, f(e_N))$ , $f(X) = \langle f, X\rangle$. Denote, $C = (c_1, \ldots, c_N)'$ , $D = (d_1, \ldots, d_N)'$.

Write $\mathcal{G}_k :=$ the sigma field generated by $\{X_0, X_1, \ldots, X_k, y_1, \ldots, y_k\}$ and $\mathcal{Y}_k :=$ the sigma field generated by $\{y_1, \ldots, y_k\}$. We shall write $M$ for the model determined by these parameters $(a_{ji}, c_i, d_i)$ , $1 \le i, j, \le N$. We assume throughout that the model order $N$ is known.

### *2.2 Measure Change*

From Elliott[3] recall that a probability $\bar{P}$ is introduced such that under $\bar{P}$, $X$ is still a Markov chain with transition matrix $A$, but the random variables $y_k$ are themselves independent and normally distributed as $N(0, 1)$.

$$\lambda_\ell(X_{\ell-1}) = \frac{\phi\left((y_\ell - C'X_{\ell-1})/(D'X_{\ell-1})\right)}{(D'X_{\ell-1})\phi(y_\ell)}$$

where $\phi(x)$ is, for example $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.

With $\Lambda_0 = 1$ and $\Lambda_k = \Pi_{\ell=1}^k \lambda_\ell(X_{\ell-1})$, a probability measure $\bar{P}$ can be defined on $(\Omega, \mathcal{G}_\infty)$ by putting $\frac{dP}{d\bar{P}} \mid_{\mathcal{G}_k} = \bar{\Lambda}_k$.

One can then show (see Elliott[3]), that under $P$, $\{w_k\}$ is a sequence of independent $N(0, 1)$ random variables, where $w_k := (y_k - (C'X_{k-1}))/(D'X_{k-1})$.

Furthermore, under $P$ the state $X$ remains a Markov chain with transition matrix $A$. Let us denote the model (1) (2) as

$$M = M(A, C, D, \pi_0)$$

where $\pi_0 = E[X_0]$.

### 2.3 *Transitions and Occupation Times*

For $1 \leq r, s \leq N$ write $\mathcal{J}_k^{rs} = \sum_{\ell=1}^{k} \langle X_{\ell-1}, e_r \rangle \langle X_\ell, e_s \rangle$, then $\mathcal{J}_k^{rs}$ is the number of jumps of $\{X_k\}$ from state $r$ to state $s$ up to time $k$. The Markov chain $X_k$ is not observed directly, but only through the observations $j$. The occupation times for being in state $r$ up until time $k$ are $\mathcal{O}_k^r = \sum_{s=1}^{n} \mathcal{J}_k^{rs}$.

Let us denote

$$\begin{aligned} \mathcal{J}_k &:= (\mathcal{J}_k^{rs}), \quad 1 \leq r, s \leq N \in \mathbb{R}^{N \times N} && (3) \\ \mathcal{O}_k &:= \mathcal{J}_k \underline{1} \qquad\qquad\qquad \in \mathbb{R}^N && (4) \end{aligned}$$

Thus in obvious notation

$$\begin{aligned} \mathcal{J}_k &= \sum_{\ell=1}^{k} X_{\ell-1} X'_\ell, && (5) \\ (\mathcal{O}_k)_{\text{diag}} &= \sum_{\ell=1}^{k} X_{\ell-1} X'_{\ell-1} = \sum_{\ell=1}^{k} (X_{\ell-1})_{\text{diag}} && (6) \end{aligned}$$

Post multiplication of (1) by $X'_{k-1}$ and summing yields

$$k^{-1} \mathcal{J}'_k = A(k^{-1} \mathcal{O}_k)_{\text{diag}} + k^{-1} \sum_{\ell=1}^{k} V_\ell X'_{\ell-1} \tag{7}$$

Given knowledge of $\{X_k\}$, it makes sense to estimate $A$ as

$$\bar{A}_k = \mathcal{J}'_k (\mathcal{O}_k)^{-1}_{\text{diag}} \tag{8}$$

at least when the inverse exists (as when all states are excited). We say that the states are *persistently exciting* when there is an finite integer $M > 0$, such that $\mathcal{O}_k^i$ is $k > M$ for $i = 1, \ldots, N$ and a constant $B < \infty$ such that for all $K > M$ and $1 \leq i \leq N$ then $k(\mathcal{O}_k^i)^{-1} < B$. This is equivalent to the condition that

$$\lim_{k \to \infty} \sup_{i,j,k} (\frac{1}{k} \mathcal{O}_{k-1})^{-1}_{\text{diag}} < \infty \tag{9}$$

In a parallel manner to the above, one can define transitions $\mathcal{T}_k^r(f) = \sum_{\ell=1}^{k} \langle X_{\ell-1}, e_r \rangle f(y_\ell)$ where $f(y)$ is either $y$, or $y^2$, depending on application. So define the row vector with elements $\mathcal{T}_k^r$ as

$$\mathcal{T}'_k(f) = \sum_{\ell=1}^{k} f(y_\ell) X'_{\ell-1}. \tag{10}$$

Now (2) leads to

$$k^{-1} \mathcal{T}'_k(y) = C' k^{-1} (\mathcal{O}_k)_{\text{diag}} + D' k^{-1} \sum_{\ell=1}^{k} (X'_{\ell-1})_{\text{diag}} w_\ell \tag{11}$$

and thus estimates can be defined as

$$\bar{C}_k = \mathcal{T}'_k(y) (\mathcal{O}_k)^{-1}_{\text{diag}}. \tag{12}$$

Likewise, squaring (2) and post-multiplying by $X'_{k-1}$ we have:

$$\begin{aligned} &\frac{1}{k} \mathcal{T}'_k(y^2) = [c_1^2, c_2^2, \ldots] \frac{1}{k} (\mathcal{O}_k)_{\text{diag}} + (d_1^2, d_2^2, \ldots) \frac{1}{k} \sum_{\ell=1}^{k} (X_{\ell-1})_{\text{diag}} w_\ell^2 \\ &+ 2(c_1 d_1, c_2 d_2, \ldots) \frac{1}{k} \sum_{\ell=1}^{k} (X_{\ell-1})_{\text{diag}} w_\ell. \end{aligned} \tag{13}$$

Estimates $\bar{D}$ can be constructed from

$$(\bar{d}_1^2, \bar{d}_2^2, \ldots, \bar{d}_N^2) = \mathcal{T}_k'(y^2)(\mathcal{O}_k)_{\text{diag}} - (\bar{c}_1^2, \bar{c}_1^2, \ldots, \bar{c}_N^2). \tag{14}$$

### *2.4 Parameter Estimation*

Consider model "estimates", possibly time varying, denoted $\widehat{\mathcal{M}}_k = \widehat{M}_1, \widehat{M}_2, \ldots, \widehat{M}_k$. Let us denote associated conditional mean estimates based on the correct model and "incorrect" model as

$$\begin{aligned}
\widehat{X}_{k|k} &= E[X_k \mid \mathcal{Y}_k, M]; \quad \widehat{X}_{k|k,\widehat{\mathcal{M}}_{k-1}} = E\left[X_k \mid \mathcal{Y}_k, \widehat{\mathcal{M}}_{k-1}\right] \\
\widehat{\mathcal{J}}_{k|k} &= E[\mathcal{J}_k \mid \mathcal{Y}_k, M]; \quad \widehat{\mathcal{J}}_{k|k,\widehat{\mathcal{M}}_{k-1}} = E\left[\mathcal{J}_k \mid \mathcal{Y}_k, \widehat{\mathcal{M}}_{k-1}\right] \\
\widehat{\mathcal{O}}_{k|k} &= E[\mathcal{O}_k \mid \mathcal{Y}_k, M]; \quad \widehat{\mathcal{O}}_{k|k,\widehat{\mathcal{M}}_{k-1}} = E\left[\mathcal{O}_k \mid \mathcal{Y}_k, \widehat{\mathcal{M}}_{k-1}\right] \\
\widehat{\mathcal{T}}_{k|k}(f) &= E[\mathcal{T}_k(f) \mid \mathcal{Y}_k, M]; \quad \widehat{\mathcal{T}}_{k|k,\widehat{\mathcal{M}}_{k-1}}(f) = E\left[\mathcal{T}_k(f) \mid \mathcal{Y}_k, \widehat{\mathcal{M}}_{k-1}\right]
\end{aligned} \tag{15}$$

Similar notation will denote one step ahead predictions $\widehat{X}_{k|k-1}$ etc., or smoothed estimates $\widehat{X}_{k|k+d}$ etc.

Now, in seeking on-line estimates of $\mathcal{J}_k$, such as conditional mean filtered, it turns out, perhaps surprisingly, to be essential to work first with on-line estimates of $\mathcal{J}_k$ given $X_k = e_i$ for each $i$ and then derive the desired estimates from these. Thus define

$$\mathcal{J}_k^X := X_k \text{row vec } \mathcal{J}_k \in \mathbb{R}^{N \times N^2} \tag{16}$$

$$\text{row vec } \mathcal{J}_k = \underline{1}' \mathcal{J}_k^X \in \mathbb{R}^{1 \times N^2}. \tag{17}$$

Let us define conditional mean estimates and associated unnormalized forms under $\bar{P}$ for $\mathcal{J}_k$ as follows

$$\widehat{\mathcal{J}}_{k|k}^X := E[\mathcal{J}_k \mid \mathcal{Y}_k, M], \quad \sigma(\mathcal{J}_{k|k}^X) := \bar{E}\left[\Lambda_k \mathcal{J}_k^X \mid \mathcal{Y}_k, M\right]. \tag{18}$$

With $\widehat{\mathcal{T}}_{k|k}^X$, $\widehat{\mathcal{O}}_{k|k}^X$, $\sigma(\mathcal{T}_{k|k}^X)$ and $\sigma(\mathcal{O}_{k|k}^X)$ similarly defined. The following lemma is a convenient re-packaging of the following results of Chapter 3 of Elliot[3] to a matrix form.

**Lemma 1** *Consider the HMM signal model (1) (2) denoted by $M$. The conditional mean filtered estimates for $\mathcal{J}_k^X$, $\mathcal{J}_k$, $\mathcal{T}_k^X$, $\mathcal{T}_k$, under $\bar{P}$, are obtained from:*

$$\begin{aligned}
&\sigma(\mathcal{J}_{k|k}^X) = AB(y_k)\sigma(\mathcal{J}_{k-1|k-1}^X) + ((Ae_1)_{\text{diag}}, (Ae_2)_{\text{diag}}, \ldots, (Ae_N)_{\text{diag}})\,(B(y_k)(\alpha_{k-1})_{\text{diag}} \otimes I) \\
&\quad \text{row vec } \sigma(\mathcal{J}_{k|k}) = \underline{1}'\sigma(\mathcal{J}_{k|k}^X)
\end{aligned} \tag{19}$$

$$\begin{aligned}
&\sigma(\mathcal{T}_{k|k}^X) = AB(y_k)\sigma(\mathcal{T}_{k-1|k-1}^X) + AB(y_k)(\alpha_{k-1})_{\text{diag}} f(y) \\
&\quad \text{row vec } \sigma(\mathcal{T}_{k|k}) = \underline{1}'\sigma(\mathcal{T}_{k|k}^X).
\end{aligned} \tag{20}$$

**Proof:** See Moore[12]
**Remarks:**

1. Exponential stablilty or initial-condition forgetting of the filters (19) and (20) follows by appealing to the generalised Perron-Frobenius result, see Seneta[16], in the same way as in Dey[1], LeGland[8], Shue[17]

2. The computational difficulty is that the $\sigma(\mathcal{J}_{k|k}^X)$ calculation requires $N^4$ multiplications for each up-date. See Elliott[3] for alternative form of the filters which can be easier to implement in practice.

3. Note $\mathcal{O}_k = \mathcal{J}_k \underline{1}$, so that $\sigma(\mathcal{O}_{k|k}) = \sigma(\mathcal{J}_{k|k})\underline{1}$. (Note also that $\alpha_k$ can be derived from $\sigma(\mathcal{J}^X_{k|k})$ by summing operations).

Now consider parameter estimates

$$\begin{aligned} \bar{A}_{k|k,\widehat{\mathcal{M}}_{k-1}} &= \widehat{\mathcal{J}}'_{k|k,\widehat{\mathcal{M}}_{k-1}} \left(\widehat{\mathcal{O}}_{k|k,\widehat{\mathcal{M}}_{k-1}}\right)^{-1}_{\text{diag}} \\ \bar{C}_{k|k,\widehat{\mathcal{M}}_{k-1}} &= \widehat{\mathcal{T}}'_{k|k,\widehat{\mathcal{M}}_{k-1}}(f) \left(\widehat{\mathcal{O}}_{k|k,\widehat{\mathcal{M}}_{k-1}}\right)^{-1}_{\text{diag}} \end{aligned} \tag{21}$$

and likewise for $\bar{D}_{k|k,\widehat{\mathcal{M}}_{k-1}}$ via (14) and then introduce the persistently excitation condition associated with the model $M$ and its "estimate" $\widehat{\mathcal{M}}_k$ as

$$\lim_{k\to\infty} \sup \left(\frac{1}{k}\widehat{\mathcal{O}}_{k|k,\widehat{\mathcal{M}}_{k-1}}\right)^{-1}_{\text{diag}} < \infty \tag{22}$$

The case studied in Moore[12] is where $\widehat{M}_{k-1}$ is given adaptively from estimates $\bar{A}_{k|k,\widehat{\mathcal{M}}_{k-1}}$, $\bar{C}_{k|k,\widehat{\mathcal{M}}_{k-1}}$ and $\bar{D}_{k|k,\widehat{\mathcal{M}}_{k-1}}$.

### *2.5 Convergence Results*

Before proceeding to propose reduced-complexity algorithms for estimating HMM parameters we repeat here the almost sure convergence results for estimating the transition probability matrix $A$, stated in Moore[12].

**Theorem 1** *Consider the HMM of (1) (2) and a particular observation sequence and state sequence outcome, $\{y_k\}$ and $\{X_k\}$, of the HMM with all states in $\{X_k\}$ persistently exciting in that (9) holds. Consider the (somewhat artificial) case where the conditional mean estimates based on the true model $M$ are available. Then*

$$\lim_{k\to\infty} \bar{A}_{k|k,M}, \bar{C}_{k|k,M}, \bar{D}_{k|k,M} = A, C, D, \quad \text{a.s.} \tag{23}$$

*The almost sure convergence rate guaranteed is $\bar{\rho}(k)^{\frac{1}{2}}$ which is like $\frac{1}{\sqrt{k}}(lnk(lnlnk)^{\alpha})^{\frac{1}{2}}$, for $k > 4$ and for any $\alpha > 1$, and the mean square rate is $(k_{-})^{-\frac{1}{2}}$ where $(k_{-})^{-\frac{1}{2}}$ is arbitrarily slower than $k^{-\frac{1}{2}}$.*

**Proof:** See Moore [12]. The proof is based on martingale convergence results and the Kronecker Lemma.

**Theorem 2** *Consider the HMM of (1) (2) and a particular observation sequence and state sequence outcome, $\{y_k\}$ and $\{X_k\}$, of the HMM with all states in $\{X_k\}$ persistently exciting in that (9) holds. Consider also an assumed model set $\widehat{\mathcal{M}}_k$ where $\widehat{M}_k$ is adaptively updated using $\bar{A}_{k|k,\widehat{\mathcal{M}}_{k-1}}$ which we suppose is persistently exciting, along with $M$, in that (22) holds and $C$ and $D$ are known. Then*

$$\lim_{k\to\infty} \bar{A}_{k|k,\widehat{\mathcal{M}}_{k-1}} = A \ \text{ a.s,} \tag{24}$$

**Proof:** See Moore[12]. The proof is based on the ordinary differential equation approach.

**Remark:** See Moore[12] for the case when $\bar{C}_{k|k,\widehat{\mathcal{M}}_{k-1}}$ and $\bar{D}_{k|k,\widehat{\mathcal{M}}_{k-1}}$ are also to be estimated.

## 3 Reduced Complexity Algorithms

One way to reduce the computational requirements of the overall estimation algorithm (21) is by reducing the computational effort used to produce estimates of the conditional mean. We note that there is a subset of models $\{\bar{M}\}$ for which calculation of the conditional mean estimates is computational easier than for other models. Hence, the key idea of this paper is to calculate the conditional mean estimates of $\mathcal{J}_k$ and $\mathcal{O}_k$ corresponding to one of these special models $\bar{M}$ (or a sequence of these models) rather than generic model "estimates" such as the adaptive $M_k$ above. This will reduce the computational effort required to implement the overall estimation algorithm (21) .

An example of one such $\bar{M}$ is the *i.i.d.* sequence model which is a subset of the valid HMMs. For the purpose of generating the conditional mean estimates, $\widehat{\mathcal{J}}_{k|k,\widehat{\mathcal{M}}_{k-1}}$ and $\widehat{\mathcal{O}}_{k|k,\widehat{\mathcal{M}}_{k-1}}$, the state sequence can be modelled as an *i.i.d.* sequence, leading to a model estimate with $\widehat{M}_k = M_{iid}$ for all $k$ as

$$M_{iid} = \{A_{iid}, C, D, \pi_0\}, \quad A_{iid} = \frac{1}{N}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \cdots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}. \tag{25}$$

The filters now denoted $\widehat{\mathcal{J}}_{k|k,M_{iid}}$ and $\widehat{\mathcal{O}}_{k|k,M_{iid}}$ require $O(N^3)$ calculations per time instant. This is a reduction from the $O(N^4)$ calculations per time instant required to implement the the general filters $\widehat{\mathcal{J}}_{k|k,\widehat{\mathcal{M}}_{k-1}}$ and $\widehat{\mathcal{O}}_{k|k,\widehat{\mathcal{M}}_{k-1}}$ using estimates $\bar{A}_{k|k,\widehat{\mathcal{M}}_{k-1}}$.

The filters for the model set $M_{\alpha\beta} = \{\alpha A_{iid} + \beta I_{N\times N}, C, D, \pi_0\}$ for scalar $\alpha, \beta$ where $\{\alpha, \beta : \underline{1}'A = \underline{1}\}$ also requires only $O(N^3)$ calculations per time instant. This set can approximate a larger class of models and results in reduced estimation error.

**Remark**

1. Even though $A = I$ may appear a likely candidate model of $\{\bar{M}\}$ for reducing complexity the persistently excitation condition (22) is not satisfied for this model and the convergence no longer holds.

2. Correct estimation occurs in low noise because $\widehat{X}_{k|k,\widehat{A}} \approx X_k$. and invariant of $\widehat{A}$.

## 4 Simulations Studies

### *4.1 Reduced Complexity Estimation*

A 3000 point, 2-state HMM was generated
with parameters: $a_{ii} = 0.8,\ i = 1, 2$ and $a_{ij} = 0.2\ \ \forall i, j\ i \neq j, C = [2, 4]'$ and $D = 0.1\,\underline{1}$. Estimation was performed in two ways: using the simplified model approximation (25), and using $\widehat{M}_k$ corresponding to the best available parameter estimates. At this noise level little bias is introduced by the approximation.

### *4.2 Threshold*

To examine the bias introduced into the parameter estimation a 10000 point, 2-state HMM generated with parameters: $a_{ii} = 0.8,\ i = 1, 2$ and $a_{ij} = 0.2\ \ \forall i, j\ i \neq j,\ C = [2, 4]'$ was simulated at various SNRs. Estimation of $A$ is performed assuming $M = M_{iid}$. Figure 1 shows estimation error verses SNR. From this figure it appears that for SNR $> 12$ or $D < 0.2$ the bias introduced is insignificant.

### *4.3 More Complicated Approximations*

A 10000 point, 2-state HMM was generated with parameters: $a_{ii} = 0.8$, $i = 1,2$ and $a_{ij} = 0.2$ $\forall i, j$ $i \neq j$, $C = [2,4]'$ and $D = 0.7$ or SNR=6.3. Estimation of $A$ was performed using three methods: using the model set $M_{\alpha\beta}$, using the model set $M_{iid}$, and using the adaptive scheme of Moore [12]. Figure 2 shows the convergence of parameters estimates over time At this noise level it appears using models $M_{\alpha\beta}$ reduces the estimation error.

## 5 Conclusions

In this paper we present reduced complexity algorithms for on-line estimation of hidden Markov model parameters. The presented algorithm requires only $O(N^3)$ calculations per time instant compared to the $O(N^4)$ calculations required to implement the estimation algorithm presented in Elliott[3]. The various estimates are not guaranteed to be consistent, but simulation studies indicate their usefulness.

## References


[1] 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*, New Orleans, Dec, 1995.

[2] I.B. Collings, V. Krishnamurthy and J.B. Moore, "Online Identification of Hidden Markov Models via Recursive Prediction Error Techniques", *IEEE Transactions on Signal Processing*, Dec, 1994.

[3] R.J. Elliott, L. Aggoun and J.B. Moore, *Hidden Markov Models, Estimation and Control*, Springer-Verlag, 1995.

[4] J.J. Ford and J.B. Moore, "On Adaptive HMM state Estimation", To appear in *IEEE Transactions on Signal Processing.*

[5] J.J. Ford and J.B. Moore, "Adaptive Estimation of HMM Transition Probabilities", To appear in *IEEE Transactions on Signal Processing.*

[6] L. Gerencsér, "Rate of Convergence of recursive estimators", *SIAM J. Control and Optimization*, Vol. 30, No. 5, pp. 120-1227, Sept. 1992.

[7] H.J. Kushner and D.S. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Optimization*, Springer-Verlag, Berlin, New York, 1978.

[8] F. LeGland and L. Mevel, "Geonmetric ergodicity in hidden Markov models", *Internal Publication 1024*, July 1996. ftp://ftp.irisa,fr/techreports/1996/PI-1028.ps.gz.

[9] L. Ljung, "Analysis of recursive stochastic algorithms", *IEEE Trans. Automatic Control*, vol. AC-22, pp. 551-575, Aug 1977.

[10] M. Loeve, *Probability Theory*, 2nd Ed., Van Nostrand, Princeton, NJ, 1960.

[11] P. Meyer, *Martingales and Stochastic Integrals - I*, Lecture Notes in Mathematics Series, No. 284, Springer-Vertag, New York, 1972.

[12] J.B. Moore, J.J. Ford & R.J. Elliott, "On-line consistent estimation of hidden Markov model parameters", submitted to *IEEE Transactions on Signal Processing.*

[13] J.B. Moore, V. Krishnamurthy, "On-line Estimation of Hidden Markov Model Parameters Based on the Kullback-Leibler Information Measure", *IEEE Transactions on Signal Processing*, Vol. 41, No. 3, pp 2557-2573, Aug 1993.

[14] J. Neveu, *Discrete Parameter Martingales*, Amsterdam, Holland: North-Holland, 1975.

[15] L.R. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition", *Proc. IEEE*, Vol.77, No.2, pp 257-285, 1989.

[16] E. Seneta, *Non-negative matrices and Markov Chains*, 2nd ed., Chapters 3-4, Springer-Verlag, 1981.

[17] L. Shue, B.D.O. Anderson and S. Dey, "Exponential stability of filters and smoothers for hidden Markov models" , *Proceedings of the 4th European Control Conference*, 1997.

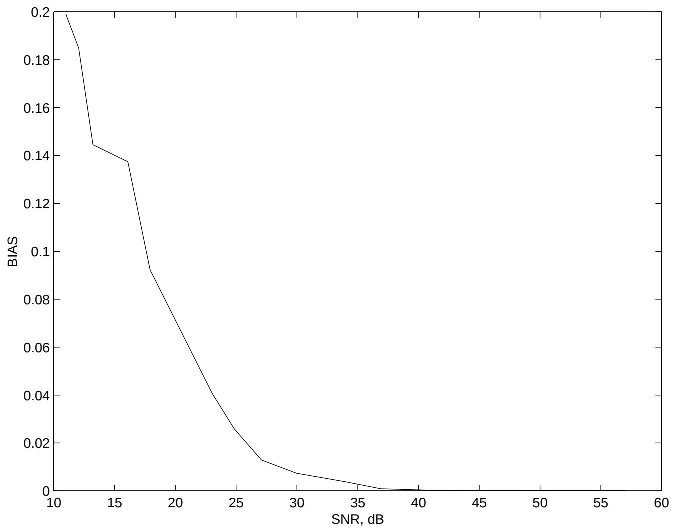


Figure 1: The empirical calculated of bias after 10000 points.

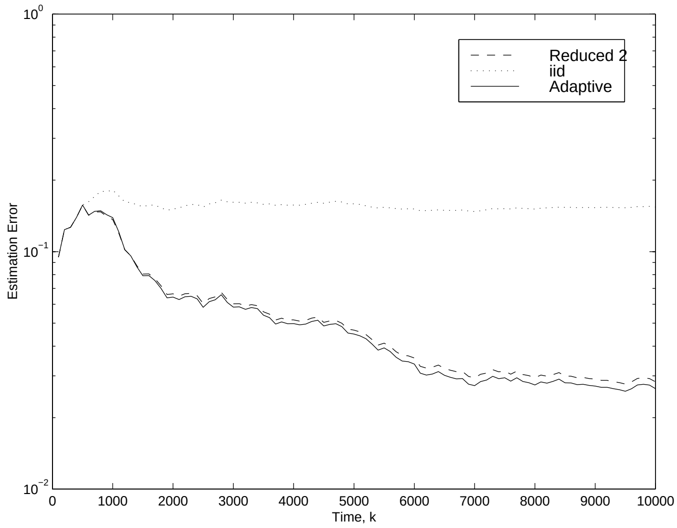


Figure 2: An empircal comparsion