

RISK-SENSITIVE FILTERS FOR IDENTIFICATION OF HIDDEN MARKOV MODELS

Jeremy Thorne and John B. Moore

Department of Systems Engineering, RSISE, The ANU
Canberra, ACT 0200, Australia
Email: Jeremy.Thorne@syseng.anu.edu.au

ABSTRACT

In this paper we derive risk-sensitive filters which can be used for both on-line and off-line identification of hidden Markov models (HMMs). The identification is achieved by taking risk-sensitive conditional mean estimates of the number of state transitions (jumps) and occupation times and then using these values to estimate the parameters of the system.

Furthermore we demonstrate that the risk-sensitive filters approach the existing asymptotically optimal (risk-neutral) filters in the limit of the risk-sensitive parameter.

1. INTRODUCTION

The basic theory of hidden Markov models (HMMs) was introduced by Baum and colleagues in a series of classic papers stretching from 1966 to 1972 [3, 4, 5, 6, 7]. The states of an HMM are assumed to belong to a discrete set and state transition probabilities are organised to form the so-called transition matrix. The measurements of the state are noisy and or possibly quantized.

Although HMMs were soon recognised to have applications to speech recognition as introduced by Baker in 1975 [8], it has taken time for HMM techniques to be applied to other signal processing problems. Now HMMs have become an integral part of the signal processing fields armoury, with a growing number of applications including speech processing and recognition [8, 9, 10, 11], biological signal processing [12] and digital communication systems [13, 14].

When making use of HMMs in the applications cited previously it is essential (for optimum performance) to have an accurate estimate of the model parameters. As a result there have been a number of off-line and on-line identification schemes proposed. On-line identification schemes are an essential requirement to some signal processing applications due to the fluctuation of state transition probabilities with time as considered by Krishnamurthy and Moore in [15]. On-line identification schemes can also be ap-

plied to off-line identification problems with minor changes to the algorithms.

There have been a number of parameter estimation schemes proposed for HMMs. These include the off-line Baum-Welch algorithm [3], recursive prediction error methods [16, 17, 18], and a conditional mean estimate approach [19].

In [19] certain summations of non-linear functions of the state are proposed which when the system is fully observed can be used as a step to achieve sub-optimal estimates of the HMM parameters. The summations represent such quantities as the number of jumps between certain states, or the state occupation times. For partially observed HMMs recursive filters give conditional mean estimates of the summations and then use these estimates as a step to achieve estimates of the parameters of the HMM. It is shown in [19] that the estimates of the summations are asymptotically optimal and that they provide on-line consistent parameter estimation which is asymptotically optimal.

The existing on-line estimation schemes have a number of limitations associated with them, these include:

1. They are only asymptotically optimal when the true model is contained in the model set, e.g. if the observation noise has some unmodelled colouration.
2. Despite being asymptotically optimal when the true model is in the model set, estimates from limited data based on poor model parameter estimates can be high in error and it may take many samples for the estimates to converge sufficiently enough to be useful. We refer to this as poor transient performance. We can think of the error between the true model parameters and the estimated model parameters at initialisation of the estimation schemes, as a high noise situation where the additional 'noise' is coloured.
3. When only finite data sets are available there

may not be sufficient information to provide useful estimates.

It has been shown for state estimation of HMMs [20] in cases where the parameters of the model used for state estimation are not the true parameters of the system that there is a significant improvement in performance when some attempt at risk-sensitive state estimation is used. The robustness property of the risk-sensitive state estimation scheme is desirable in parameter estimation schemes, since the identification will be robust to model uncertainty. This is the motivation behind the estimation scheme proposed in this paper and in an earlier risk-sensitive parameter estimation scheme proposed in [21].

In [21] the authors proposed pseudo risk-sensitive filters which were a modified version of the filters proposed in [19]. The modifications were ad hoc but in the spirit of those rigorously established for risk-sensitive state estimation [20]. It is important to note that [21] did not rigorously justify the proposed filters. It was shown in [21] through simulation that the risk-sensitive filters proposed did significantly improve the transient performance of both on-line and off-line parameter estimation of HMMs, but there was no attempt made to investigate the robustness of the filters to colouration of noise etc.

The risk-sensitive identification scheme proposed in this paper is derived rigorously in that we develop risk-sensitive conditional mean estimates of the non-linear functions of the state proposed in [19].

This paper is organised as follows: In Section 2 we introduce the HMM signal model and the change of measure. In Section 3 we define the summations of the non-linear functions of the state (such as the number of transitions between certain states) used for the estimation of the HMM parameters and derive the filters used for the risk-sensitive conditional mean estimates of the non-linear functions. In Section 4 we investigate the limiting results of the risk-sensitive identification scheme, in the limit of the risk-sensitive parameter. Our conclusions are presented in Section 5.

2. HIDDEN MARKOV MODEL

2.1. State Space Model

Let X_k be a discrete-time homogeneous, first-order Markov process belonging to a finite-discrete set. Define $S = \{e_1, \dots, e_N\}$ where $e_i = (0, \dots, 1, \dots, 0)' \in \mathbb{R}^N$ with 1 in the i -th position. Without loss of generality, we can assume $X_k \in S$. We consider this process to be defined on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with $\mathcal{F}_k^0 = \sigma\{X_0, \dots, X_k\}$ and the complete filtration $\{\mathcal{F}_k\}$. The state space model is then defined by

$$X_k = AX_{k-1} + W_k \quad (1)$$

$$y_k = CX_{k-1} + v_k \quad (2)$$

where $W_k, k \in \mathbb{Z}^+$ is a sequence of \mathcal{F}_k -martingale increments and hence $E[W_k|\mathcal{F}_{k-1}] = 0$. Also y_k is continuous valued belonging to \mathbb{R}^P and $v_k \in \mathbb{R}^P, k \in \mathbb{Z}^+$ is i.i.d with a strictly positive density function ϕ_k .

Due to the Markov nature of X_k , we can write

$$E[X_k|\mathcal{F}_{k-1}] = E[X_k|X_{k-1}] = AX_{k-1} \quad (3)$$

where the entries $\{a_{ij}\}$ are defined as $P(X_k = e_i | X_{k-1} = e_j)$. Obviously $a_{ij} > 0, \forall i, j$ and $\sum_{j=1}^N a_{ij} = 1, \forall i$. We also assume that X_0 or its distribution is known.

2.2. Change of Measure

We define a new measure \bar{P} where $\{y_k\}, k \in \mathbb{Z}^+$ is a sequence of i.i.d. random variables having a density function ϕ_k . Define

$$\bar{\lambda}_k = \frac{\phi_k(y_k - CX_k)}{\phi_k(y_k)} \quad (4)$$

$$\bar{\Lambda}_k = \prod_{l=0}^k \bar{\lambda}_l \quad (5)$$

If we set the Radon-Nikodym derivative $\frac{d\bar{P}}{dP}|_{\mathcal{G}_k} = \bar{\Lambda}_k$ then under \bar{P} , the random variables $v_k, k \in \mathbb{Z}^+$ are i.i.d. with density functions ϕ_k . Here $\{\mathcal{G}_k\}$ is the complete filtration generated by $X_k^* = (X_0, \dots, X_k)$ and $\mathcal{Y}_{k-1} = (y_0, \dots, y_{k-1})$ (see [1]).

3. PARAMETER ESTIMATION

In this section of the paper we introduce the risk-sensitive cost function and risk-sensitive conditional mean estimates of the summations of non-linear functions of the states used to estimate the HMM parameters.

3.1. Parameters Estimation for Fully Observed Case

Here we first recall the approach of [19].

Multiplying (1) by X_{l-1}' and summing over k we find

$$\sum_{l=0}^k X_l X_{l-1}' = A \sum_{l=0}^k X_{l-1} X_{l-1}' + \sum_{l=0}^k W_l X_{l-1}' \quad (6)$$

or

$$J_k = A O_k + \sum_{l=0}^k W_l X'_{l-1} \quad (7)$$

where

$$J_k = \sum_{l=0}^k X_l X'_{l-1} \text{ and } O_k = \sum_{l=0}^k X_{l-1} X'_{l-1} \quad (8)$$

Hence, under a nonsingular condition on O_k we obtain a reasonable estimate of A from

$$\hat{A}_k = J_k O_k^{-1} \quad (9)$$

Clearly O_k is diagonal and $O_k = \sum_{l=0}^k (X_{l-1})_{diag}$. Furthermore [19] shows that

$$(J'_k \mathbf{1})_{diag} = O_k \quad (10)$$

The martingale movement property for W_l ensures that if $O_k^{-1} \rightarrow 0$ as $k \rightarrow \infty$, then $\hat{A}_k \rightarrow A$ as $k \rightarrow \infty$ almost surely.

Post-multiplication of (2) by X'_k and summing over k we obtain

$$\sum_{l=0}^k y_l X'_{l-1} = C \sum_{l=0}^k X_{l-1} X'_{l-1} + \sum_{l=0}^k v_l X'_{l-1} \quad (11)$$

or

$$T_k = C O_k + \sum_{l=0}^k v_l X'_{l-1} \quad (12)$$

where

$$T_k = \sum_{l=0}^k y_l X'_{l-1} \quad (13)$$

Hence, we may obtain a reasonable estimate of C from

$$\hat{C}_k = T_k O_k^{-1} \quad (14)$$

Again, a martingale movement property of the measured noise v_l ensures that $\hat{C}_k \rightarrow C$ as $k \rightarrow \infty$ (assuming $O_k^{-1} \rightarrow 0$).

Remark 3.1 It must be noted that (9) and (14) only hold for the case of persistent excitation. That is all states of the system are consistently visited.

3.2. Risk-sensitive Conditional Mean Estimates

In this section we find risk-sensitive conditional mean estimates of the summations defined in (8) and (13). The technique for taking risk-sensitive conditional mean estimates is demonstrated in [20], an overview of this method is contained in Section 3.2.1.

3.2.1. Risk Sensitive Cost Function and State Estimates

The risk-sensitive state estimation problem is to determine the state \hat{X}_k so that the risk-sensitive cost function is minimised. As shown in (15)

$$\hat{X}_k = \arg \min_{\zeta} J_k(\sigma) \quad \forall k = 0, 1, \dots \quad (15)$$

where

$$J_k(\zeta) = E[\exp(\theta \Psi_{0,k}(\zeta)) | \mathcal{Y}_k] \quad \forall k = 0, 1, \dots \quad (16)$$

is the risk sensitive cost function. Here,

$$\Psi_{0,k}(\zeta) = \hat{\Psi}_{0,k-1} + \frac{1}{2} (X_k - \zeta)' Q_k (X_k - \zeta) \quad (17)$$

where

$$Q_k \geq 0 \quad \forall k = 0, 1, 2, \dots$$

$$\hat{\Psi}_{m,n} = \frac{1}{2} \sum_{i=m}^n (X_i - \hat{X}_i)' Q_k (X_i - \hat{X}_i) \quad (18)$$

and $\theta \geq 0$ is the risk-sensitive parameter.

The risk-sensitive state estimation problem has already been solved in [20]. The technique used was to take the conditional mean estimate of the state as shown in (19) and then to use the recursion found (20) in the minimisation of the conditional mean estimate (15) and (16).

$$\alpha_{k+1}^{RS}(e_i) = \bar{E}[\bar{\Lambda}_k \exp(\theta \Psi_{0,k}) < X_{k+1}, e_j > | \mathcal{Y}_k] \quad (19)$$

$$\alpha_{k+1}^{RS} = A D_k B_k \alpha_k \quad (20)$$

where $\alpha_k^{RS} = [\alpha_k^{RS}(e_1), \dots, \alpha_k^{RS}(e_N)]'$ and

$$D_k = \text{diag}\{\exp\left[\frac{\theta}{2}(e_1 - \hat{X}_k)' Q_k (e_1 - \hat{X}_k)\right], \dots, \exp\left[\frac{\theta}{2}(e_N - \hat{X}_k)' Q_k (e_N - \hat{X}_k)\right]\}$$

$$\hat{X}_k = e_m \quad (21)$$

$$m^* = \arg \min_m \sum_{j=1}^N \frac{\phi_k(y_k - \hat{C}_{k-1}|k-1 e_j)}{\phi_k(y_k)} \exp\left[\frac{\theta}{2}[(e_j - e_m)' Q_k (e_j - e_m)]\right] \alpha_k^{RS}(e_j) \quad (22)$$

where $m \in \{1, 2, \dots, N\}$

3.2.2. Recursive Filters for Risk-sensitive Conditional Mean Estimates of \mathcal{J}_k

It was shown in [19] that to find the conditional mean estimate of \mathcal{J}_k , that we must first find the conditional mean estimate of the augmented matrix $X_k \text{row vec } \mathcal{J}_k$ or \mathcal{J}_k^X where $\text{row vec } \mathcal{J}_k = \mathbf{1}' \mathcal{J}_k^X$. To find the risk-sensitive conditional mean estimate of \mathcal{J}_k^X directly is a difficult problem, so the approach we take is to find the risk-sensitive estimates of \mathcal{J}_k^X 's columns i.e. $(\mathcal{J}_k^X)^r$ $r = 1, \dots, N^2$ or $X_k \mathcal{J}_k^{mn}$ $m, n = 1, \dots, N$.

Let us define (23) as the risk-sensitive conditional mean estimates of $(\mathcal{J}_k^X)^r$.

$$(\hat{\mathcal{J}}_k^X)^r = E[\exp(\theta \Psi_{0,k-1}) \mathcal{J}_k^{mn} < X_k, e_j > | \mathcal{Y}_k] \quad (23)$$

where \mathcal{J}_k^{mn} are the elements of \mathcal{J}_k with $m, n = 1, \dots, N$ and the inner product operation $< X_k, e_j >$ is the equivalent of the augmentation performed in [19].

Using a version of Baye's Theorem and the measure change we find

$$\begin{aligned} E[\exp(\theta \Psi_{0,k-1}) \mathcal{J}_k^{mn} | \mathcal{Y}_k] \\ = \frac{\bar{E}[\bar{\Lambda}_k \exp(\theta \Psi_{0,k-1}) \mathcal{J}_k^{mn} | \mathcal{Y}_k]}{\bar{E}[\bar{\Lambda}_k < X_k, e_j > | \mathcal{Y}_k]} \end{aligned} \quad (24)$$

Definition 3.1 Define the measures $\kappa_k^{mn}(e_j)$ to be the unnormalised information states such that

$$\kappa_{k-1|k-1}^{mn}(e_j) = \bar{E}[\bar{\Lambda}_{k-1} \exp(\theta \Psi_{0,k-2}) \mathcal{J}_{k-1}^{mn} < X_k, e_j > | \mathcal{Y}_{k-1}] \quad (25)$$

Lemma 3.1 The information states $\kappa_{k|k} = [\kappa_{k|k}^{11}, \dots, \kappa_{k|k}^{NN}]'$ obeys the following recursion

$$\begin{aligned} \kappa_{k|k} = G_k \kappa_{k-1|k-1} + [(e_1)_{diag} G_k (\alpha_{k-1}^{RS})_{diag}, \dots \\ \dots, (e_N)_{diag} G_k (\alpha_{k-1}^{RS})_{diag}] \end{aligned} \quad (26)$$

where

$$\begin{aligned} \kappa_{k|k}^{mn} &= [\kappa_{k|k}^{mn}(e_1), \dots, \kappa_{k|k}^{mn}(e_N)]' \\ G_k &= \hat{A}_{k-1|k-1} B_k D_{k-1} \\ B_k &= \text{diag} \left\{ \frac{\phi_k(y_k - \hat{C}_{k-1|k-1} e_1)}{\phi_k(y_k)}, \dots, \frac{\phi_k(y_k - \hat{C}_{k-1|k-1} e_N)}{\phi_k(y_k)} \right\} \end{aligned}$$

Proof

$$\begin{aligned} \kappa_{k|k}^{mn}(e_i) &= \bar{E}[\bar{\Lambda}_k \exp(\theta \Psi_{0,k-1}) \mathcal{J}_k^{mn} < X_k, e_i > | \mathcal{Y}_k] \\ &= \bar{E}[\bar{E}[\bar{\Lambda}_k X'_{k-1} \hat{A}'_{k-1|k-1} e_i \\ &\quad \exp\{\frac{\theta}{2}(X_{k-1} - \hat{X}_{k-1})' Q_{k-1} (X_{k-1} - \hat{X}_{k-1})\} \\ &\quad | \mathcal{Y}_k] \bar{\Lambda}_{k-1} \exp(\theta \Psi_{0,k-2}) \mathcal{J}_{k-1}^{mn} | \mathcal{Y}_{k-1}] \\ &\quad + \bar{E}[\bar{E}[\bar{\Lambda}_k (X_k X'_{k-1})^{mn} X'_{k-1} \hat{A}'_{k-1|k-1} e_i \\ &\quad \exp\{\frac{\theta}{2}(X_{k-1} - \hat{X}_{k-1})' Q_{k-1} (X_{k-1} - \hat{X}_{k-1})\} \\ &\quad | \mathcal{Y}_k] \bar{\Lambda}_{k-1} \exp(\theta \Psi_{0,k-2}) | \mathcal{Y}_{k-1}] \end{aligned}$$

(Exploited y_k being i.i.d., thus expectation within expectation. Substitutions made using (5), (8), (17) and (3), noting that in (3) we use the best available estimate of A , which is at time $k-1$.)

$$\begin{aligned} &= \bar{E}[\sum_{j=1}^N \{b_{jj} d_{jj} \hat{a}_{ij}\} \bar{\Lambda}_{k-1} \exp(\theta \Psi_{0,k-1}) \mathcal{J}_{k-1}^{mn} | \mathcal{Y}_{k-1}] \\ &\quad + \bar{E}[\sum_{j=1}^N \{b_{jj} d_{jj} (e_i e'_j)^{mn} \hat{a}_{ij}\} \\ &\quad \bar{\Lambda}_{k-1} \exp(\theta \Psi_{0,k-1}) | \mathcal{Y}_{k-1}] \forall i \end{aligned}$$

(Applied the formula for a discrete expectation, where b_{jj} , d_{jj} and \hat{a}_{ij} are the elements of B_k , D_{k-1} and $A_{k-1|k-1}$ respectively.)

$$\begin{aligned} &= \sum_{j=1}^N [\hat{a}_{ij} b_{jj} d_{jj} \kappa_{k-1|k-1}^{mn}(e_j) \\ &\quad + \hat{a}_{ij} b_{jj} d_{jj} (e_i e'_j)^{mn} \alpha_{k-1}^{RS}(e_j)] \forall i \end{aligned} \quad (27)$$

Writing in matrix notation completes the proof. \square

Definition 3.2 The normalised estimate of row vec $\hat{\mathcal{J}}_{k|k}$ is given by

$$\text{row vec } \hat{\mathcal{J}}_{k|k} = \frac{\mathbf{1}' \kappa_{k|k}}{\mathbf{1}' \alpha_k} \quad (28)$$

where $\alpha_k = \hat{A}_{k-1|k-1} B_k \alpha_{k-1}$ is the MAP state estimate

Now that we have the filter for the risk-sensitive conditional mean estimate of \mathcal{J}_k it is trivial to use (10) to find a risk-sensitive conditional mean estimate

of \mathcal{O}_k given by (29) and then substitute these results into (9) to find the risk-sensitive estimate of A as given by (30)

$$\hat{\mathcal{O}}_{k|k} = (\hat{\mathcal{J}}'_{k|k} \mathbf{1})_{diag} \quad (29)$$

$$\hat{A}_{k|k} = \hat{\mathcal{J}}_{k|k} \hat{\mathcal{O}}_{k|k}^{-1} \quad (30)$$

3.2.3. Recursive Filters for Risk-sensitive Conditional Mean Estimates of \mathcal{T}_k

In this section we present the filter for the estimation of the non-linear function \mathcal{T}_k . As is expected the technique for finding the risk-sensitive filter for estimation of \mathcal{T}_k is the same as used when finding the filter for \mathcal{J}_k . Because of this we omit the proof for the results presented here in Lemma 3.2

Lemma 3.2 *The information state $\gamma_{k|k} = [\gamma_{k|k}^{11}, \dots, \gamma_{k|k}^{PN}]'$ obeys the following recursion*

$$\gamma_{k|k} = G_k \gamma_{k-1|k-1} + y'_k \otimes [G_k(\alpha_{k-1}^{RS})_{diag}] \quad (31)$$

where $A \otimes B$ denotes the Kronecker product, with $(A \otimes B)_{ij} = a_{ij} b$

Definition 3.3 *The normalised estimate of row vec $\hat{\mathcal{T}}_{k|k}$ is given by*

$$\text{row vec } \hat{\mathcal{T}}_{k|k} = \frac{\mathbf{1}' \gamma_{k|k}}{\mathbf{1}' \alpha_k^{RS}} \quad (32)$$

Now that we have the filter for the risk-sensitive conditional mean estimate of \mathcal{T}_k and have \mathcal{O}_k given by (29) it is trivial to substitute these results into (14) to find the risk-sensitive estimate of C as given by (33)

$$\hat{C}_{k|k} = \hat{\mathcal{T}}_{k|k} \hat{\mathcal{O}}_{k|k}^{-1} \quad (33)$$

4. LIMITING RESULTS

In this section of the paper we discuss the limiting results of our risk-sensitive filter in the limit of the risk-sensitive variable θ . Let us now consider (34).

$$\begin{aligned} \lim_{\theta \rightarrow 0} G_k &= \hat{A}_{k-1|k-1} B_k D_{k-1} \\ &\rightarrow \hat{A}_{k-1|k-1} B_k \end{aligned} \quad (34)$$

since

$$\lim_{\theta \rightarrow 0} D_{k-1} \rightarrow I_N$$

where I_N is an $N \times N$ identity matrix.

This result means that the recursive filters for the risk-sensitive conditional mean estimates of \mathcal{J}_k^X and \mathcal{T}_k^X , approach the recursive filters for the risk-neutral conditional mean estimates (presented in [19]) as the risk-sensitive parameter θ approaches zero.

Furthermore when $\theta = 0$ the risk-sensitive filters are identical to the risk-neutral filters given in [19].

5. CONCLUSIONS

In this paper we have introduced the state space representation of hidden Markov models, which allowed the application of the measure change technique to give us a sequence of i.i.d. observations $\{y_k\}$. From this model basis we were able to rigorously derive a risk-sensitive identification scheme for HMMs, by drawing upon the parameter estimation techniques used in [1, 19] and by applying the idea of risk-sensitive conditional mean estimates which had previously been used only for state estimation in HMMs [20].

6. REFERENCES

- [1] R.J.Elliot, L.Aggoun and J.B.Moore, *Hidden Markov Models, Estimation and Control*, Springer-Verlag, 1995.
- [2] M.Loeve, *Probability Theory*, D. Van Nostrand Company, Princeton, NJ, 1963.
- [3] L.E.Baum and T.Petrie, "Statistical inference for probabilistic functions of finite state Markov chains" *Ann. Math. Stat.*, Vol. 37, pp 1554-1563, 1966.
- [4] L.E.Baum and J.A.Egon, "An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model ecology", *Bull. Amer. Meterol. Soc.*, vol. 73, pp 360-363, 1967.
- [5] L.E.Baum and G.R.Sell, "Growth functions for transformations on manifolds", *Pac. J. Math.*, vol. 27, No. 2, pp 211-227, 1968.
- [6] L.E.Baum, T. Petrie, G.Soules and N.Weiss, "A maximisation technique occurring in the statistical analysis of probabilistic functions of Markov Chains", *Ann. Math. Stat.*, vol. 41, No. 1, pp 164-171, 1970.

- [7] L.E.Baum, "An inequality and associated maximisation technique in statistical estimation for probabilistic functions of Markov processes", *Inequalities*, vol. 3, pp 1-8, 1972.
- [8] J.K.Baker, "The dragon system - an overview", *Acoust. Speech Signal Processing*, vol. ASSP-23, no. 1, pp 24-29, 1975.
- [9] L.R.Bahl and F.Jelinik, "Decoding of channels with insertions, deletions and substitutions with applications to speech recognition", *IEEE Trans. Informat. Theory*, vol. IT-21, pp 404-411, 1975.
- [10] F.Jelinik, "Continuous speech recognition by statistical methods", *Proc. IEEE*, vol. 64, pp 532-536, Apr. 1976.
- [11] L.R.Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition", *Proc. of IEEE* Vol. 77, No. 2, pp 257-285, 1989.
- [12] S.H.Chung, V.Krishnamurthy and J.B.Moore, "Adaptive processing techniques based on Hidden Markov Models for characterising very small channel currents buried in noise and deterministic interferences", *Philosophical Transactions of the Royal Society of London, Series B*, Vol. 334, pp 357-384, 1991.
- [13] I.B. Collings and J.B.Moore, "An Adaptive Hidden Markov Model Approach to FM and M-ary DPSK Demodulation in Noisy Fading Channels", *Signal Processing*, Vol. 47, No. 1, pp 71-84, Nov. 1995.
- [14] V.Krishnamurthy, S.Dey and J.P. LeBlanc, "A Hidden Markov Extended Least Squares Blind Equalization Algorithm for IIR channels", *Proc. 1994 IEEE International Symposium of Information Theory*, Trondheim, Norway, pp 354, July 1994.
- [15] V.Krishnamurthy, and J.B.Moore, "On-Line Estimation of Hidden Markov Model Parameters Based on the Kullback-Leibler Information Measure", *IEEE Trans. Signal Processing*, Vol. 41, No. 8, pp 2557-2573, 1993.
- [16] I.B. Collings, V. Krishnamurthy and J.B.Moore, "On-line Identification of Hidden Markov Models via Recursive Prediction Error Techniques", *IEEE Trans. Signal Processing*, Vol. SP-42, No. 12, pp 3535-3539, 1994.
- [17] J.J.Ford and J.B.Moore, "On Adaptive HMM state estimation", *IEEE Trans. Signal Processing*, Vol. 46, No. 2, pp 475-486, Feb. 1998.
- [18] J.J. Ford and J.B.Moore, "Adaptive Estimation of HMM Transition Probabilities", *IEEE Trans. Signal Processing*, Vol. 46, No. 5, pp 1374-1385, May 1998.
- [19] J.B. Moore, J.J.Ford and R.J.Elliot, "On-line Consistent Estimation of Hidden Markov Models", submitted to *IEEE Trans. on Signal Processing*
- [20] S.Dey and J.B.Moore, "Risk-sensitive filtering and smoothing for hidden Markov models", *Systems Control Lett.* Vol. 25, pp 361-366, 1995.
- [21] S.Athuraliya J.Ford and J.B.Moore, "Faster parameter estimation using risk-sensitive filters" to appear in CDC '98.