Blind Equalization of IIR Channels Using Hidden Markov Models and Extended Least Squares

Vikram Krishnamurthy, *Member, IEEE*, Subhrakanti Dey, and James P. LeBlanc

***Abstract*—In this paper, we present a blind equalization algorithm for noisy IIR channels when the channel input is a finite state Markov chain. The algorithm yields estimates of the IIR channel coefficients, channel noise variance, transition probabilities, and state of the Markov chain. Unlike the optimal maximum likelihood estimator which is computationally infeasible since the computing cost increases exponentially with data length, our algorithm is computationally inexpensive. Our algorithm is based on combining a recursive hidden Markov model (HMM) estimator with a relaxed SPR (strictly positive real) extended least squares (ELS) scheme. In simulation studies we show that the algorithm yields satisfactory estimates even in low SNR. We also compare the performance of our scheme with a truncated FIR scheme and the constant modulus algorithm (CMA) which is currently a popular algorithm in blind equalization.**

I. Introduction

The hidden Markov model (HMM) consists of a Markov chain corrupted by additive white Gaussian noise (WGN). For this simple case, the expectation maximization (EM) algorithm (Baum–Welch re-estimation formulae) [8] is a commonly used numerical scheme to obtain ML estimates of the HMM parameters. However, in many applications, e.g., communication systems, the Markov chain is first filtered by an unknown channel and then corrupted by additive WGN. The problem of estimating the coefficients of such an unknown channel with unknown (stochastic) channel inputs is termed "blind equalization."

If the channel can be modeled as a finite impulse response (FIR) filter and the channel input as a finite-state Markov chain, then ML estimates of the coefficients of the channel can straightforwardly be obtained using the EM algorithm (see [2]). Also recursive versions of the EM algorithm can be used to obtain on-line equalization schemes for the FIR channel [3]. Maximum *a posteriori* state estimates can be obtained and a Viterbi algorithm can be used to obtain ML state sequence estimates.

Manuscript received December 7, 1993; revised May 29, 1995. This work was partially funded by ATERB and an Australian Research Council grant. The activities of the Cooperative Research Center for Robust and Adaptive Systems were supported by the Australian Commonwealth Government under the Cooperative Research Centers Program. The associate editor coordinating the review of this paper and approving it for publication was Dr. R. D. Preuss.

V. Krishnamurthy is with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria, Australia.

S. Dey is with the Department of Systems Engineering, Research School of Information Sciences and Engineering, Australian National University, Canberra, Australia.

J. P. LeBlanc is with the Department of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA.

IEEE Log Number 9415292.

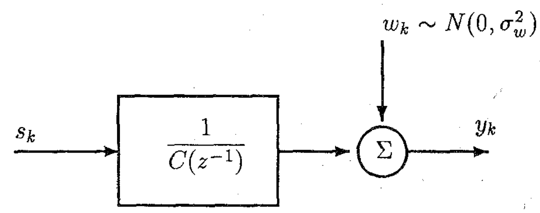


Fig. 1. Signal model.

In this paper, we consider the case when the channel is *infinite impulse response* (IIR). The EM algorithm can no longer be used because the E-step requires computation of the smoothed (posterior) probability density function. In the case of a Markov chain corrupted by white noise, an inductive calculation is possible [12] using the forward-backward scheme. However, in the case of a noisy IIR filtered Markov chain, the lack of Markovianity of the channel output does not permit an inductive calculation of the smoothed or even the filtered probability density function. Direct computation of these densities are computationally prohibitive since they require computing a weighted sum of the joint probability density functions of the observations over all N^T realizations of a N-state T-point Markov chain. Similarly, the (optimal) Viterbi algorithm cannot be applied for state sequence estimation.

Therefore, the only feasible algorithms for the estimation of the state and parameters of noisy IIR filtered Markov chains are suboptimal algorithms. In this paper, we propose a suboptimal algorithm which couples a recursive HMM estimator with an extended least squares (ELS) estimator. We call our algorithm the *HMM-ELS* algorithm.

Let us first describe our signal model and estimation objective. Then we discuss applications, related works and show why some alternative estimation approaches are infeasible for our problem.

Signal Model

The model we shall consider is schematically shown in Fig. 1 and can be described as follows:

The observations $y_k, k = 1, 2, \cdots, T$ are obtained as

$$y_k = \frac{s_k}{C(z^{-1})} + w_k, \qquad w_k \sim N(0, \sigma_w^2) \tag{1.1}$$

where w_k is zero mean white Gaussian noise (WGN) with variance σ_w^2.

$C(z^{-1}) = 1 - \Sigma_{i=1}^{p} c_i z^{-i}$ (where z^{-1} is the delay operator) denotes the *unknown* autoregressive IIR channel.

1053-587X/95$04.00 © 1995 IEEE

We assume that $C(z^{-1})$ is stable, i.e., it has all its zeros outside the unit circle and that p is known. s_k denotes a N-state discrete–time homogeneous first-order Markov chain. Consequently, the state s_k at time k is one of N *known* state levels $q = (q_1\ q_2 \cdots q_N)'$. The transition probability matrix is $A = (a_{ij})$ where $a_{ij} = P(s_{t+1} = q_j | s_t = q_i)$. Of course $a_{ij} \geq 0, \Sigma_{j=1}^{N}\ a_{ij} = 1$, for each i. We assume that s_k is ergodic (see Section III for details). Let π denote the initial state probability vector: $\pi = (\pi_i), \pi_i = P(s_1 = q_i)$.

Let us denote the T length noisy filtered observation sequence as $Y_T = (y_1, \cdots, y_T)'$.

Remark: We assume an autoregressive (AR) channel only for notational convenience. As described later in this section, dealing with ARMA channels (minimum or nonminimum phase) is a straightforward extension.

Aim

Given the observations Y_T, the aim of our blind equalization algorithm is two-fold:

1) *State Estimates:* Obtain filtered estimates $\hat{s}_k$ of the state of the Markov chain at time k.
2) *Parameter Estimates:* Estimate the unknown IIR filter coefficients $C = (c_1 \cdots c_p)'$, channel noise variance σ_w^2 and transition probability matrix A. Let $\phi = (C, \sigma_w^2, A)$ denote the unknown parameter vector.

Highlights of Our HMM-ELS Algorithm

The following are some of the highlights of our HMM-ELS blind equalization algorithm:

1) Methodology: The crux of our equalization problem lies in the fact that due to the IIR channel, the Markov chain is imbedded in colored noise (or equivalently the channel output is non-Markovian) which can be seen by rewriting (1.1) as $C(z^{-1})y_k = s_k + w_k - c_1 w_{k-1} - \cdots - c_p w_{k-p}$. Standard HMM signal processing assumes the noise to be white and cannot be used. The presence of colored noise suggests using the ELS algorithm in conjunction with a hidden Markov model estimator. Before spelling out the details of our approach let us first briefly recall ELS and recursive HMM estimation.

- The ELS algorithm is widely used in adaptive control for linear ARMAX system identification (see Ljung [9], p. 317 for details). It yields consistent estimates of the system parameters. The ELS algorithm has a certain strictly positive real (SPR) condition on the polynomial $C(z^{-1})$ for almost sure convergence in the parameter estimates. In many cases this SPR condition is not satisfied. For this reason modifications of the ELS algorithm have been proposed to relax the SPR condition [1], [5]. We shall use the relaxed SPR ELS scheme suggested in [1].
- The HMM estimator yields optimal filtered estimates of the Markov state and parameters of a Markov chain in white noise. The state estimator is based on the forward filter [8]. On-line parameter estimates including transition probabilities A and channel noise variance σ_w^2 are obtained via the recursive EM algorithm [4], [16].

Our HMM-ELS algorithm is based on cross-coupling the relaxed SPR ELS and HMM estimators resulting in a computationally efficient recursive (on-line) scheme. The algorithm is schematically shown in Fig. 2 and can be briefly described as follows: At each time instant, ELS estimates of the noise and channel coefficients C are passed to the HMM estimator which yields Markov state estimates and also parameter estimates of A, σ_w^2. The Markov state estimates are in turn passed to the ELS estimator at the next time instant, and so on. Heuristically one would expect that the HMM-ELS scheme yields satisfactory estimates when the initial estimates are sufficiently close to the true values. The same philosophy is used in adaptive control and is called the "certainty equivalence" [20, p. 180]. We show in extensive simulation studies, the HMM-ELS algorithm is extremely robust to initial conditions and yields excellent estimates even in low signal to noise ratio (SNR).

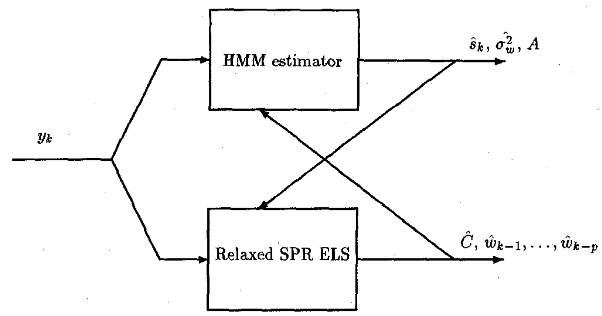


Fig. 2. HMM-ELS blind equalization algorithm.

We have been unable to prove convergence of the HMM-ELS algorithm. However, in Section III we present convergence results for certain special cases of the algorithm.

2) Computational Requirements: Our HMM-ELS scheme requires $O(N^2T) + O(p^2T)$ computations for a T-point data sequence. Recall that the ML estimator requires $O(N^T)$ computations. Also, if the IIR channel was approximated by a FIR channel of length $\tilde{p}$ then the ML estimator would require $O(N^{\tilde{p}}T)$ computations which is still significant, and hence, impractical for large $\tilde{p}$.

3) Simulated Performance: In extensive simulations studies we show that our HMM-ELS algorithm yields excellent estimates even in low SNR. Our simulations are conducted on oversampled BPSK signals (two-state Markov chains) and also Markov chains with three and five states. Various IIR channels are considered including time-varying channels with jump changing coefficients. We show that the HMM-ELS algorithm performs significantly better than standard HMM schemes that assume white noise. Also the HMM-ELS scheme is compared with a truncated FIR approximation algorithm and the constant modulus algorithm (CMA).

Applications

1) *Blind equalization of IIR channels:* Often in communication systems, due to coding and oversampling, the input to the channel is a finite-state Markov chain rather than IID (independent and identically distributed). Examples include phase-shift-keyed (PSK) and frequency-shift-keyed (FSK) signals. Of course IID inputs can also be

handled using the techniques of this paper, since an IID chain is merely a special case when all elements of the transition probability matrix A are $1/N$ (see Section 5.2 for simulation results and comparison with CMA).

2) Most often the input alphabets i.e., the levels q of the Markov chain, are known *a priori.* Hence, we do not consider the estimation of q in this paper. The transition probability matrix A depends on the type of coding and oversampling used. In the case when the transition probabilities are unknown, we estimate them via the recursive EM algorithm [4].
3) Because of the widespread use of the constant modulus algorithm (CMA) in blind-equalization [14], [15], we compare the performance of the HMM-ELS algorithm with CMA in computer simulations.
4) 2. *ARMAX HMM's:* Our signal model (1.1) is a special case of the nonlinear ARMAX system (nonlinear because of the Markov chain input s_k)

$$A(z^{-1})y_k = B(z^{-1})s_k + C(z^{-1})w_k \qquad (1.2)$$

with $A(z^{-1}) = C(z^{-1})$ and $B(z^{-1}) = 1$. In time-series jargon our model is an *Output-Error* model [9], p. 75. The HMM-ELS algorithm can be viewed as a suboptimal scheme for estimating Markov chains in FIR filtered colored noise $C(z^{-1})w_k$. Notice if the noise is IIR filtered, ML estimates can be obtained via the EM algorithm [6].
5) By a straightforward extension, the HMM-ELS algorithm can be used with arbitrary (minimum or non-minimum phase) polynomials $B(z^{-1})$ in (1.2) since $B(z^{-1})s_k$ itself is a Markov chain with N^B states (where B is the degree of the polynomial $B(z^{-1})$). This model is very similar to the "Nonstationary Time Series" model recently proposed for studying business cycles in the econometrics literature [19].
6) *Other applications:* Virtually the same problem as blind-equalization arises in geophysics in the modeling of seismic impedances [12]. Also in biophysics, channel currents in cell membranes are often modeled as finite-state Markov chains [7]. Due to thermal noise and filtering effect of the measurement probes, the measurements can be modeled similarly to (1.1).

Other Approaches

Our signal model (1.1) is a special case of what is termed in the statistics literature as a stochastic dynamic linear model (DLM) [17]. Broadly speaking, there are four classes of suboptimal schemes in the literature that are used to estimate DLM's [16]: decision directed schemes, probabilistic editor, probabilistic teacher, and quasi-Bayes techniques. Of these the most commonly used schemes in the communications literature are *Decision Directed Schemes* which involve cut-off rules that assign a state estimate at each time instant. Our HMM-ELS scheme falls in this category. Other examples include decision feedback equalizers (DFE), suboptimal Viterbi algorithms including reduced-sequence estimation schemes [21], [23], and the CMA. The reduced-sequence estimation schemes have computational cost of $O(N^{\tilde{p}})$ for an IIR approximation of length $\tilde{p}$.

In [13], a truncated maximum likelihood scheme is proposed for estimating DLM's. Again, this scheme is a finite-memory approximation to the IIR filter and so is computationally expensive; the computational cost is $O(N^{\tilde{p}})$ for a truncation of length $\tilde{p}$. In [12], an off-line estimation scheme is presented which maximizes the joint probability density function of the Markov chain and observations. The scheme involves a continuous-discrete optimization problem which is quite complicated.

We have not found any works in the literature that use our approach of combining a linear estimator (ELS in our case) with an optimal nonlinear estimator (HMM estimator, in our case).

This paper is organized as follows: In Section II, we present details of our HMM-ELS blind equalization algorithm. In Section III, we present some preliminary convergence results. Section IV presents the results of computer simulation studies, and Section V compares the HMM-ELS algorithm with a truncated FIR algorithm and the constant modulus algorithm (CMA). Section VI lists some conclusions.

II. HMM-ELS Blind Equalization Algorithm

The HMM-ELS algorithm proposed in this paper combines a relaxed SPR ELS scheme and recursive HMM estimator resulting in a suboptimal computationally efficient recursive (on-line) scheme.

Before presenting details, let us first briefly give a rationalization of the HMM-ELS algorithm.

Notice that our signal model (1.1) can be rewritten as

$$C(z^{-1})y_k = s_k + w_k - c_1 w_{k-1} - \cdots - c_p w_{k-p}. \qquad (2.1)$$

If at each time instant k, the Markov chain state s_k was exactly known in (2.1) then the estimation problem reduces to a standard ARMAX estimation problem (more specifically an "output error" model, [5]). Then ELS yields asymptotically consistent estimates of C and also estimates of the previous noise values.

On the other hand, if the previous noise values $w_{k-1}, \cdots, w_{k-p}$ and also C were exactly known in (2.1), then these values could be subtracted from the observations resulting in s_k corrupted by *white* noise w_k, which is a standard HMM problem. The HMM estimator then yields optimal filtered state estimates. Also, via the recursive EM algorithm online estimates of the parameters A, σ_w^2 can be obtained.

As shown in Fig. 2, the HMM-ELS algorithm combines these two steps as follows:

1) At time k, the recursive HMM estimator yields estimate of the state of s_k, noise variance σ_w^2 and transition probabilities A.
2) The relaxed SPR ELS estimator gives on-line estimates of the channel parameters c_i and $w_{k-i}, i \in \{1, 2, \cdots, p\}$, denoted by $\hat{c}_i^{(k)}$ and $\hat{w}_{k-i}$, respectively.

The two steps are described below in the following two subsections.

Let $\hat{\phi}^{(k)} = (\hat{C}^{(k)}, \hat{\sigma}_w^{2(k)}, \hat{A}^{(k)})$ denote the model estimates at time k.

A. Recursive HMM Estimator

At time k, we have $\hat{w}_{k-1}, \cdots, \hat{w}_{k-p}$ and $\hat{C}^{(k-1)}(z^{-1})$ available from the ELS scheme described in Section II-B. Therefore, the HMM to be estimated is

HMM Signal Model:

$$\hat{C}^{(k-1)}(z^{-1})y_k + \sum_{i=1}^{p} \hat{c}_i^{(k-1)} \hat{w}_{k-i} = s_k + w_k \qquad (2.2)$$

We shall use a recursive HMM estimator to estimate the parameters σ_w^2, A and the state s_k.

State Estimation: Let $\hat{W}_{k-p}^{k-1} = (\hat{w}_{k-1}, \cdots, \hat{w}_{k-p})'$. Define the symbol probability density function

$$\begin{aligned} b_n(y_k; \hat{W}_{k-p}^{k-1}, \hat{\phi}^{(k)}) &\triangleq f(y_k | Y_{k-p}^{k-1}, s_k = q_n, \hat{W}_{k-p}^{k-1}, \hat{\phi}^{(k)}), \\ &\qquad n \in \{1, 2, \cdots, N\} \\ &= \frac{1}{\sqrt{2\pi\hat{\sigma}_w^{2(k)}}} \cdot \exp\left(- \frac{\left(y_k - \sum_{i=1}^{p} \hat{c}_i^{(k)} y_{k-i} - q_n + \sum_{i=1}^{p} \hat{c}_i^{(k)} \hat{w}_{k-i} \right)^2}{2\hat{\sigma}_w^{2(k)}} \right) \end{aligned} \qquad (2.3)$$

where $f(\cdot)$ denotes the density function and the second equation follows since $w_k \sim N[0, \sigma_w^2]$. For convenience we shall denote $b_n(y_k; \hat{W}_{k-p}^{k-1}, \hat{\phi}^{(k)})$ as $b_n(y_k)$.

Define the unnormalized filtered density $\alpha_k(m)$ and the filtered state estimate $\hat{s}_k$ as

$$\alpha_k(m) = f(s_k = q_m, Y_k | \hat{W}_{k-1}, \hat{\phi}^{(k)}) \qquad (2.4)$$

$$\hat{s}_k = \boldsymbol{E}\{s_k = q_m | Y_k, \hat{W}_{k-1}, \hat{\phi}^{(k)}\}. \qquad (2.5)$$

Lemma 1: The unnormalized filtered density $\alpha_k(m)$, the normalized filtered density $\gamma_{k|k}(m), m \in \{1, 2, \cdots, N\}$, and the filtered state estimate $\hat{s}_k$ can be computed recursively as follows:

$$\alpha_k(n) = \sum_{m=1}^{N} \alpha_{k-1}(m) a_{mn} b_n(y_k), \quad \alpha_1(m) = \pi_m b_m(y_1) \qquad (2.6)$$

$$\gamma_{k|k}(m) = \frac{\alpha_k(m)}{\sum_{m=1}^{N} \alpha_k(m)} \qquad (2.7)$$

$$\begin{aligned} \hat{s}_k &= \boldsymbol{E}\{s_k = q_m | Y_k, \hat{W}_{k-1}, \hat{\phi}^{(k)}\} \\ &= \frac{\sum_{m=1}^{N} \alpha_k(m) q_m}{\sum_{m=1}^{N} \alpha_k(m)}, \qquad m \in \{1, 2, \cdots, N\}. \end{aligned} \qquad (2.8)$$

Proof: The proof is almost identical to that in [8]. □

Remark: The state estimate $\hat{s}_k$ computed in (2.8) is called the conditional-mean (CM) state estimate. Extensive simulations have confirmed that using CM state estimates always results in better performance than maximum aposteriori MAP state estimates. A heuristic reasoning is that unlike CM estimates, MAP estimates are discrete valued. Therefore, errors in the MAP estimate introduce a bursty noise signal which degrades the performance of the subsequent ELS step.

Parameter Estimation

We use the recursive EM algorithm [4] to obtain on-line estimates of σ_w^2 and A. For brevity, we omit details of the algorithm (see [4], and the references therein).

Noise variance: Using the recursive EM algorithm update for the variance (e.g., (3.35) in [4]) we have (2.9), which appears at the bottom of the page, where $\hat{c}_i^{(k-1)}$ and $\hat{w}_{k-i}$ are the estimates of the channel parameters and the past noise values, $i \in \{1, 2, \cdots, p\}$ and $\hat{C}^{(k-1)}(z^{-1}) = 1 - \Sigma_i \, \hat{c}_i^{(k-1)} z^{-i}$.

Transition probabilities: The update equation for the $\hat{a}_{mn}^{(k)}$ is somewhat complicated by the constraints $\Sigma_n \, \hat{a}_{mn}^{(k)} = 1$ and $\hat{a}_{mn}^{(k)} \geq 0$. As described in [6], these constraints can be taken into account by dealing with square roots:

$$s_{mn}(k) = \sqrt{\hat{a}_{mn}^{(k)}}, \qquad m, n \in \{1, 2, \cdots, N\}. \qquad (2.10)$$

The advantage of this new parametrization is that the constraint manifold is differentiable at all points, and we now only have the equality constraint $\Sigma_{n=1}^{N} \, s_{mn}^2(k) = 1$. The recursive EM update equations for the transition probabilities are given as (see [6])

$$\begin{aligned} s_{mn}^{U}(k) &= s_{mn}(k-1) + \frac{g_{mn}(k)}{\mu_{mn}(k)} \\ \hat{a}_{ij}^{(k)} &= s_{ij}^2(k) = \frac{(s_{ij}^{U}(k))^2}{\sum_{j=1}^{N} (s_{ij}^{U}(k))^2} \end{aligned} \qquad (2.11)$$

$$\hat{\sigma}_w^{2(k)} = \hat{\sigma}_w^{2(k-1)} + \frac{\sum_{m=1}^{N} \gamma_{k|k}(m) (\hat{C}^{(k-1)}(z^{-1}) y_k - q_m + \sum_{i=1}^{p} \hat{c}_i^{(k-1)} \hat{w}_{k-i})^2 - \hat{\sigma}_w^{2(k-1)}}{k} \qquad (2.9)$$

which ensures that $\hat{a}_{ij}^{(k)}$ satisfies the above constraints. In (2.11)

$$g_{mn}(k) = \frac{2\zeta_{k|k}(m,n)}{s_{mn}(k-1)} - 2\gamma_{k|k}(m)s_{mn}(k-1),$$

$$\mu_{mn}(k) = \lambda_A \mu_{mn}(k-1) + \frac{2\zeta_{k|k}(m,n)}{s_{mn}^2(k-1)} + 2\gamma_{k|k}(m)$$

$$\zeta_{k|k}(m,n) = \frac{\gamma_{k|k}(m)\hat{a}_{mn}^{(k-1)} b_n(y_{k+1})}{\sum_{m=1}^{N}\sum_{n=1}^{N}\gamma_{k|k}(m)\hat{a}_{mn}^{(k-1)} b_n(y_{k+1})}. \quad (2.12)$$

B. Relaxed SPR ELS Algorithm

The HMM estimator described above yields filtered estimates $\hat{s}_k$ of the Markov chain s_k.

Let e_k denote the error in the Markov state estimate, i.e., $e_k \triangleq s_k - \hat{s}_k$. We can rewrite (1.1) as

$$C(z^{-1})y_k = \hat{s}_k + C(z^{-1})w_k + e_k \quad (2.13)$$

Assumption: Assume that the noise terms $C(z^{-1})w_k + e_k$ can be represented as $D(z^{-1})w_k$ where

$$D(z^{-1}) = 1 - \sum_{i=1}^{r} d_i z^{-i} \quad (2.14)$$

for some $r \geq p$ (we give a heuristic justification for this at the end of the section).

Then the ARMAX model to be estimated is

ARMAX Model:

$$C(z^{-1})y_k = \hat{s}_k + D(z^{-1})w_k. \quad (2.15)$$

We shall use a relaxed SPR ELS algorithm to estimate the parameters of (2.15). The standard ELS algorithm is too restrictive because to ensure almost sure convergence (we discuss convergence in Section III) in parameter estimates it requires that $\{1/D(z^{-1}) - 1/2\}$ be SPR, i.e.,

$$\mathrm{Re}\left[\frac{1}{D(e^{j\omega})} - \frac{1}{2}\right] > 0 \quad \forall\omega \quad -\pi < \omega \leq \pi. \quad (2.16)$$

In many cases, this SPR condition will not be satisfied. Hence, we shall use a relaxed SPR algorithm to estimate the parameters as follows:

Let us first transform (2.15) to the equivalent model

Transformed Model:

$$F_d(z^{-1})C(z^{-1})y_k = F_d(z^{-1})\hat{s}_k - G_d(z^{-1})w_{k-M} + w_k \quad (2.17)$$

where $F_d(z^{-1}) = \Sigma_{i=0}^{M-1} f_i q^{-i}$ with $f_0 = 1$ is the unique $(M-1)$th degree truncation of $D^{-1}(z^{-1})$ and $G_d(z^{-1})$ is the unique remainder term given by $G_d(z^{-1}) = \Sigma_{i=0}^{r-1} g_i q^{-i}$, i.e.,

$$1 = D(z^{-1})F_d(z^{-1}) + z^{-N}G_d(z^{-1}). \quad (2.18)$$

The advantage of the transformed model (2.17) is that now the SPR condition becomes relaxed

$$\frac{1}{1 - G_d(z^{-1})} \text{ is SPR}, \quad (2.19)$$

which is less restrictive than (2.16).

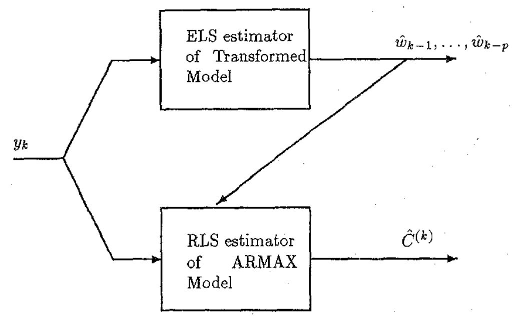


Fig. 3. Relaxed SPR ELS algorithm.

By choosing M suitably in (2.18), the SPR condition (2.19) will be satisfied. We give a design rule for selecting M, in Section III.

The relaxed SPR ELS algorithm operates with the following two steps carried out in parallel at each time instant k (see Fig. 3):

Step 1: ELS Estimation of Transformed Model (2.17)

Assume that M has been chosen sufficiently large so that (2.19) is satisfied.

Let θ denote the vector of parameters associated with the the coefficients of (2.17)

$$\text{i.e. } \theta = \text{coefficients of } F_d(z^{-1})C(z^{-1}), F_d(z^{-1}), G_d(z^{-1}). \quad (2.20)$$

Then ELS parameter estimation is carried out on the transformed model (2.17) yielding estimates of the coefficients of the polynomials $F_d(z^{-1})C(z^{-1}), F_d(z^{-1})$ and $G_d(z^{-1})$ and the past noise estimates $w_{k-i}, i \in \{1, 2, \cdots, r\}$ as follows:

$$\begin{aligned}
\psi_k &= (y_{k-1} \cdots y_{k-p-M+1}, \hat{s}_{k-1} \cdots \hat{s}_{k-M+1}, \\
&\quad -\hat{w}_{k-M} \cdots -\hat{w}_{k-r-M+1})' \\
\hat{w}_k &= y_k - \psi_k'\hat{\theta}_{k-1} - \hat{s}_k \\
P_k &= \frac{1}{\lambda}\left\{P_{k-1} - \frac{P_{k-1}\psi_k\psi_k' P_{k-1}}{\lambda + \psi_k' P_{k-1}\psi_k}\right\}, \qquad P_0 > 0 \\
\hat{\theta}_k &= \hat{\theta}_{k-1} + P_k\psi_k\hat{w}_k
\end{aligned} \quad (2.21)$$

where $\hat{\theta}_k$ denotes the estimate of θ at time k, ψ_k is the regression vector, λ is the forgetting factor with $0 < \lambda \leq 1$. P_0 in (2.21) is initialized to a positive definite symmetric matrix.

Step 2: Recursive Least Squares (RLS) Parameter Recovery

The above ELS step gives us consistent estimates of the coefficients of the transformed model (2.17) providing the relaxed SPR condition (2.19) is satisfied. The parameters of the original system (2.15) denoted as

$$\overline{\theta} = (c_1 \cdots c_p, d_1 \cdots d_r)' \quad (2.22)$$

are obtained using the following RLS algorithm operating in parallel to the above ELS algorithm.

$$\overline{\psi}_k = (y_{k-1} \cdots y_{k-p}, -\hat{w}_{k-1} \cdots -\hat{w}_{k-r})'$$

$$\overline{P}_k = \frac{1}{\overline{\lambda}}\left\{\overline{P}_{k-1} - \frac{\overline{P}_{k-1}\overline{\psi}_k\overline{\psi}_k'\overline{P}_{k-1}'}{\overline{\lambda}+\overline{\psi}_k'\overline{P}_{k-1}\overline{\psi}_k}\right\}, \qquad P_0 > 0$$

$$\hat{\overline{\theta}}_k = \hat{\overline{\theta}}_{k-1} + \overline{P}_k\overline{\psi}_k(y_k - \hat{s}_k - \overline{\psi}_k'\hat{\overline{\theta}}_{k-1}) \tag{2.23}$$

where $\hat{\overline{\theta}}_k$ denotes the estimate of $\overline{\theta}$ at time k and $\overline{\lambda}$ is the forgetting factor such that $0 < \overline{\lambda} \leq 1$.

Note that the noise terms $\hat{w}_{k-1}, \cdots, \hat{w}_{k-r}$ in the regression vector $\overline{\psi}_k$ are regarded as measurable. Thus, the algorithm (2.23) has an almost standard least squares form. The only nonstandard feature is that the regression vector $\overline{\psi}_k$ differs from the true one where w_{k-i} would be present instead of $\hat{w}_{k-i}$. In the next section, we shall present a theorem (which we proved in [1]), that as long as the ELS algorithm converges, this difference is asymptotically negligible, i.e., it does not affect either the consistency or the asymptotic rate of convergence of the RLS scheme.

Justification of Assumption (2.14)

Assumption (2.14) allows for estimation errors e_k in the HMM estimator. It models e_k as some stationary finite moving average process driven by w_k. We have been unable to give a rigorous proof. Certainly, simulations show that using a parametric model such as (2.15) yields significantly better estimates than assuming $C(z^{-1}) = D(z^{-1})(e_k = 0)$.

Notice that our assumption is weaker than assuming e_k is white. However, it assumes a finite moving average parametrization.

For small e_k and $\phi^{(k)} = \phi$ (the true model) the assumption holds as heuristically explained below: By the innovations theorem [11] e_k is white. Then $C(z^{-1})w_k + e_k$ can be represented as $D(z^{-1})u_k$ for some stable polynomial $D(z^{-1})$ where u_k is white (see [10], Theorem 2.1, pp. 214–215). For small $e_k, u_k \rightarrow w_k$.

Interpretation of Relaxed SPR Condition (2.19)

The effect of increasing M in (2.18) is to relax the SPR condition (2.19) by increasing the SPR region. Fig. 4 shows the SPR conditions for a IIR(2) channel, i.e., $N = 2$ for $M = 1$ (standard ELS), $M = 2, M = 4$, and $M = 8$. Also shown is the stability triangle, i.e., the region where $C(z^{-1})$ is stable. Fig. 4 shows the benefit of working with $M > 1$ as far as the SPR condition is concerned. Notice that for large M, e.g., $M = 8$, there can only be "marginal" failure of the SPR condition.

C. Computational Complexity

The cost for a T-length data sequence is:
Recursive HMM estimator: $O(N^2T)$
Relaxed SPR ELS: $O((p+r+M)^2T)$.
Total cost: $O(N^2T) + O((p+r+M)^2T)$.
Note, if the standard ELS is used, $M = 1$.

III. Convergence Results

We have been unable to prove convergence of the ELS-HMM algorithm. Indeed, proving convergence is extremely difficult because any proof requires showing first that the filtered estimate α_k is exponentially stable. This itself has not been proved in the literature yet.

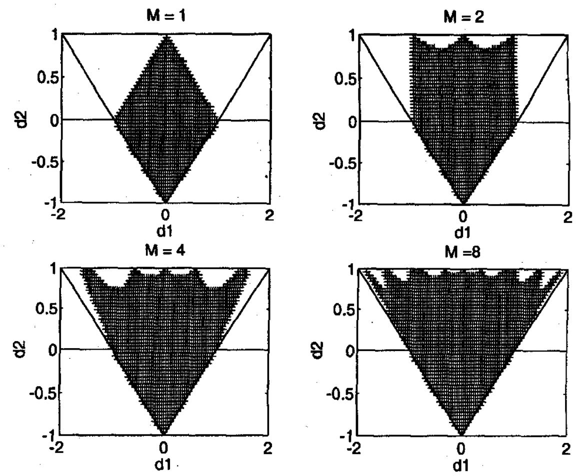


Fig. 4. Relaxed SPR regions for IIR(2) channel ($N = 2$).

However, we list here convergence results under the following special cases:

A. If $s_k = s_k$ for All k

If the Markov chain was estimated without any errors, i.e., $e_k = 0$ in (2.13) then the ARMAX model (2.15) becomes an output-error model with $C(z^{-1}) = D(z^{-1})$. We then have the following convergence results for the relaxed SPR ELS algorithm; see [5] for proofs.

The following two theorems show that the relaxed SPR ELS algorithm gives consistent estimates providing the SPR condition is satisfied and the input Markov chain is persistently exciting.

Theorem 2: Consider the ELS algorithm (2.21) associated with the transformed signal model (2.17). If M is chosen sufficiently large such that the SPR condition (2.19) is satisfied and if the input $\hat{s}_k$ is persistently exciting, then

$$||\theta - \hat{\theta}_k||^2 = O(k^{-1}\log k) \text{ a.s.}$$

$$\sum_{i=1}^{k} |w_i - \hat{w}_i|^2 = O(\log \lambda_{\max} P_k^{-1}). \tag{3.1}$$

Proof: The proof is presented in [5]. The only condition to be checked in our case is that the input Markov chain s_k is persistently exciting. From [20], p. 73, Lemma 3.4.5, an input is (weakly) persistently exciting of order p, if its two-sided spectrum is nonzero at p points or more.

Our assumption in Section I that s_k is a homogeneous ergodic (more precisely "mixing" [22, pp. 32–33]) Markov chain ensures that s_k is persistently exciting. This is because a homogeneous Markov chain is ergodic if it contains a single recurrent class of states that is aperiodic [18, Theorem 3.2.5, p. 191]. This, for example, precludes s_k being a constant valued process that is not persistently exciting.

Theorem 3: Consider the RLS algorithm (2.23) with signal model (2.15) under the relaxed SPR condition (2.19), where

$\hat{w}_k$ is generated from the ELS algorithm (2.21). Then

$$\|\bar{\theta} - \bar{\theta}_k\|^2 \leq O(k^{-1} \log k) \text{ a.s.} \tag{3.2}$$

Proof: See [5]. □

Remark: It should be pointed out, in general, Theorem 2 also requires that for persistence of excitation of the regression vector ψ_k, the noise polynomial and input polynomial are coprime. This condition is automatically satisfied here, since input polynomial (that multiplying s_k) is unity.

We now give a design rule for selecting M in (2.18). The following theorem is proved in ([5]).

Theorem 4: Consider the polynomial $D(z^{-1})$ $(= C(z^{-1})$ since $\hat{s}_k = s_k, \forall k)$

$$D(z^{-1}) = \sum_{i=0}^{r} d_i z^{-i} = \prod_{i=1}^{r} (1 - z_i z^{-1}) \text{ with } d_0 = 1 \tag{3.3}$$

such that $|z_i| \leq R < 1$ for all i. Consider also for any M a polynomial pair $\{F_d(z^{-1}), G_d(z^{-1})\}$ with degrees $M - 1$ and $r - 1$, respectively, defined uniquely by the long division (2.18). Then, there exists an integer $M_0(R)$ such that for all $M \geq M_0(R)$, the relaxed SPR condition (2.19) is satisfied. Moreover, $M_0(R)$ can be defined as the smallest values of M such that

$$R^M r(M + 2r - 1)^r 2^{r-1} < 1, \quad R\frac{(M + 2r)^r}{(M + 2r - 1)^r} < 1. \tag{3.4}$$

The above theorem says that if the zeros of the colored noise polynomial $D(z^{-1})$ lie inside a circle centered at the origin of radius R, then selecting $M = M_0(R)$ will always result in the SPR condition being satisfied. So $M_0(R)$ can be used as a design rule (albeit a very conservative rule [5]) for selecting M if *a priori* knowledge is available that the roots of D lie in a circle of radius R.

B. If Assumption (2.13) Holds

Again, Theorems 2–4 hold. For persistence of excitation of the regressor, it is required that $C(z^{-1})$ and $D(z^{-1})$ are coprime.

C. If $w_k = \hat{w}_k$ for All k

If $(\hat{w}_{k-1}, \cdots, \hat{w}_{k-p}) = (w_{k-1}, \cdots, w_{k-p})$, then it can be proved that the recursive EM algorithm is a Gauss–Newton scheme for maximizing the Kullback information measure [4].

In the following theorem, let ϕ denote the true model.

Theorem 5: The recursions (2.9) and (2.11) are derived by using a Gauss–Newton algorithm to maximize the Kullback–Leibler information measure $J(\phi^{(k)}) = E\{\log f(Y_k|\phi^{(k)})|\phi\}$. Moreover, under sufficient regularity $\phi^{(k)} \to \phi$ a.s. and in mean square.

Proof: See [4] and [16, pp. 205–207] for the regularity conditions required. □

IV. Simulation Studies of HMM-ELS Algorithm

In this section, we present detailed computer simulation studies to evaluate the performance of our HMM-ELS algorithm. This section is organized as follows: We present simulation results for various IIR channels, including time-varying channels with over-sampled binary phase-shift-keyed (BPSK) signal inputs, which can be modeled as a two-state Markov chain. We then show that the HMM-ELS equalizer also yields excellent estimates for higher order Markov inputs (three and five states). The necessity of using a relaxed SPR ELS criterion in the HMM-ELS algorithm is illustrated. Finally, a comparison between the error probability in state estimates as obtained via the standard HMM algorithm (which assumes the noise is white) and HMM-ELS algorithm is presented.

For each of the channel models considered in this section, by replicating each simulation experiment 50 times, the mean estimate and root mean square (rms) error were computed as

$$\text{mean estimate} = \frac{1}{50} \sum_{i=1}^{50} \hat{c}_i$$
$$\text{rms error} = \frac{1}{50} \sum_{i=1}^{50} (c_i - \hat{c}_i)^2. \tag{4.1}$$

In all cases initial channel estimates were chosen as $c_i^{(0)} = 0$.

A. Blind Equalization of IIR Channels for BPSK Signaling Scheme

Consider a binary phase-shift keyed (BPSK) signal of the form (pp. 394–403 [24])

$$s_t = \sum_{k=1}^{T} \tilde{a}_k \Pi_{T_s}[t - (k-1)T_s] \cos(w_s t) \tag{4.2}$$

where T_s is the bit duration, $w_s = 2\pi/T_s, \Pi_{T_s}$ is the "boxcar" function

$$\Pi_{T_s}[t] = \begin{cases} 1 & 0 \leq t \leq T_s \\ -1 & \text{elsewhere} \end{cases} \tag{4.3}$$

and $\tilde{a}_i$ is a two-state Markov chain with levels $\{-1, +1\}$ and transition probability matrix $\tilde{v}$. To a good approximation [2], s_k can be regarded as a two-state Markov chain with $q = (-1, 1)'$ and

$$A = \begin{pmatrix} a_{11} & 1 - a_{11} \\ 1 - a_{22} & a_{22} \end{pmatrix} \quad \text{where} \quad a_{ii} = 1 - \frac{1 - \tilde{v}_{ii}}{T_s}. \tag{4.4}$$

Thus, oversampling (i.e., increasing the bit duration T_s) results in a diagonally dominant A.

Accordingly, the results in this section are obtained by simulating a two-state Markov chain input

$$A = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}, \quad q = \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \tag{4.5}$$

In addition, we chose $M = 2$ and $r = 4$.

TABLE I
HMM-ELS PARAMETER ESTIMATION ERROR FOR IIR(2) CHANNEL, $N = 2, \sigma_w = 0.6$

	true	mean	rms error
c_1	0.4	0.3526	0.0476
c_2	-0.6	-0.5629	0.0375

	true	mean	rms error
c_1	0.8	0.7787	0.0220
c_2	-0.8	-0.7836	0.0176

	true	mean	rms error
c_1	0.4	0.4783	0.0789
c_2	0.5	0.4223	0.0785

	true	mean	rms error
c_1	0.2	0.3046	0.1048
c_2	0.7	0.5997	0.1045

	true	mean	rms error
c_1	0.6	0.6582	0.0593
c_2	0.3	0.2422	0.0589

	true	mean	rms error
c_1	0.5	0.4730	0.0286
c_2	-0.4	-0.3765	0.0280

	true	mean	rms error
c_1	1.5	1.4988	0.0041
c_2	-0.9	-0.8951	0.0062

	true	mean	rms error
c_1	1.0	1.0275	0.0285
c_2	-0.3	-0.3270	0.0283

TABLE II
HMM-ELS PARAMETER ESTIMATION ERROR FOR IIR(4) CHANNEL, $N = 2$

$\sigma_w = 0.2$

	true	HMM-ELS		std HMM
		mean	rms error	estimate
c_1	0.6	0.5981	0.0093	0.4931
c_2	-0.5	-0.4973	0.0122	-0.3094
c_3	0.3	0.2965	0.0194	0.1444
c_4	-0.16	-0.1544	0.0366	-0.0907

$\sigma_w = 1.0$

	true	HMM-ELS		std HMM
		mean	rms error	estimate
c_1	0.6	0.6037	0.0186	0.1895
c_2	-0.5	-0.4697	0.0426	0.0204
c_3	0.3	0.2577	0.0513	0.0164
c_4	-0.16	-0.1540	0.0182	-0.0122

$\sigma_w = 0.4$

	true	HMM-ELS		std HMM
		mean	rms error	estimate
c_1	-0.5	-0.5119	0.0195	-0.3619
c_2	-0.4	-0.3951	0.0135	-0.3303
c_3	0.3	0.3013	0.0096	0.2151
c_4	0.2	0.2149	0.0386	0.0920

$\sigma_w = 1.0$

	true	HMM-ELS		std HMM
		mean	rms error	estimate
c_1	-0.5	-0.5082	0.0222	-0.1898
c_2	-0.4	-0.2996	0.1017	-0.1864
c_3	0.3	0.2361	0.0661	0.0723
c_4	0.2	0.1663	0.0424	-0.0068

TABLE III
HMM-ELS PARAMETER ESTIMATION FOR IIR(4) CHANNEL, $N = 2$

$\sigma_w = 1.4$

	true	mean	rms error
c_1	0.6	0.6121	0.0514
c_2	-0.5	-0.4026	0.1258
c_3	0.3	0.1703	0.1507
c_4	-0.16	-0.1357	0.0459

$\sigma_w = 1.4$

	true	mean	rms error
c_1	-0.5	0.4672	0.0511
c_2	-0.4	-0.2685	0.1334
c_3	0.3	0.1999	0.1030
c_4	0.2	0.1314	0.0817

TABLE IV
HMM-ELS PARAMETER ESTIMATION ERROR FOR IIR(8) CHANNEL, $N = 2$

$\sigma_w = 0.2$

	true	mean	rms error
c_1	0.1	0.0982	0.0027
c_2	-0.6	-0.5967	0.0043
c_3	0.59	0.5862	0.0053
c_4	-0.19	-0.1834	0.0090
c_5	0.07	0.0624	0.0093
c_6	-0.054	-0.0485	0.0066
c_7	-0.012	-0.0171	0.0065
c_8	0.032	0.0346	0.0039

$\sigma_w = 0.4$

	true	mean	rms error
c_1	0.1	0.0803	0.0196
c_2	-0.6	-0.5683	0.0314
c_3	0.59	0.5555	0.0323
c_4	-0.19	-0.1441	0.0460
c_5	0.07	0.0233	0.0451
c_6	-0.054	-0.0242	0.0287
c_7	-0.012	-0.0347	0.0233
c_8	0.032	0.0449	0.0137

$\sigma_w = 0.2$

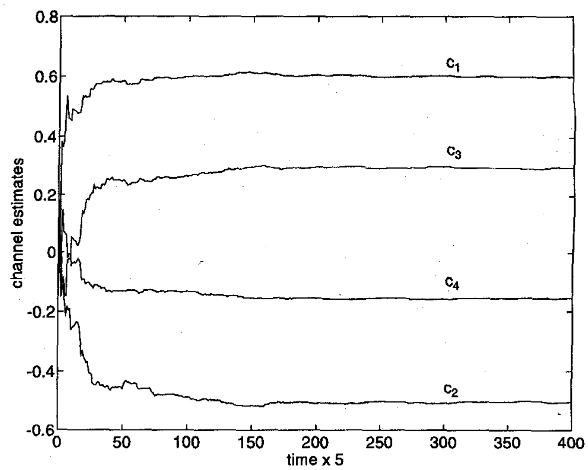
	true	mean	rms error
c_1	0.5	0.4950	0.0052
c_2	-0.8	-0.7914	0.0102
c_3	0.95	0.9405	0.0123
c_4	-0.47	-0.4593	0.0148
c_5	0.17	0.1621	0.0138
c_6	-0.174	-0.1698	0.0101
c_7	-0.032	-0.0319	0.0074
c_8	0.072	0.0706	0.0004

$\sigma_w = 0.4$

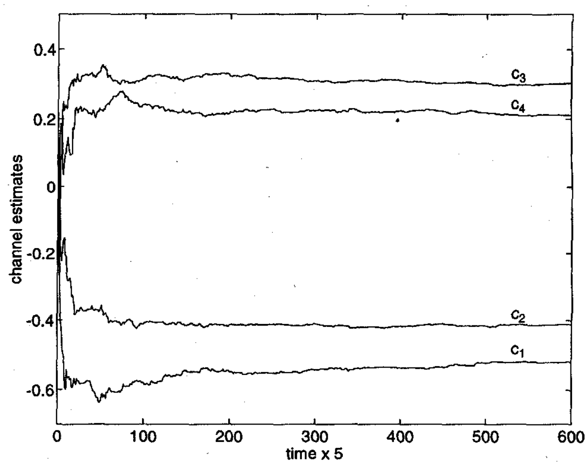
	true	mean	rms error
c_1	0.5	0.4822	0.0236
c_2	-0.8	-0.7696	0.0421
c_3	0.95	0.9320	0.0407
c_4	-0.47	-0.4675	0.0492
c_5	0.17	0.1698	0.0510
c_6	-0.174	-0.1902	0.0388
c_7	-0.032	-0.0169	0.0349
c_8	0.072	0.0620	0.0165

Mean Estimate and rms Error: Tables I–IV show the mean estimates and rms errors of the channel coefficient estimates for IIR(2), IIR(4), and IIR(8) channels, respectively. For comparison, Table II also lists the corresponding estimates obtained via the standard HMM algorithm, which are seen to be unsatisfactory and get even worse with increasing noise.

Notice that despite the fact that the channels $C = (0.8, -0.8)', (1.5, -0.9)', (0.5, -0.4)'$ in Table I are non-SPR, the HMM-ELS algorithm (with relaxed ELS algorithms) yields excellent estimates. Similarly, all the channels in Tables II and III are non-SPR; still the estimates are excellent.



$N = 2$, $C = (0.6\ -0.5\ 0.3\ -0.16)'$, $\sigma_w = 0.2$



$N = 2$, $C = (-0.5\ -0.4\ 0.3\ 0.2)'$, $\sigma_w = 0.4$

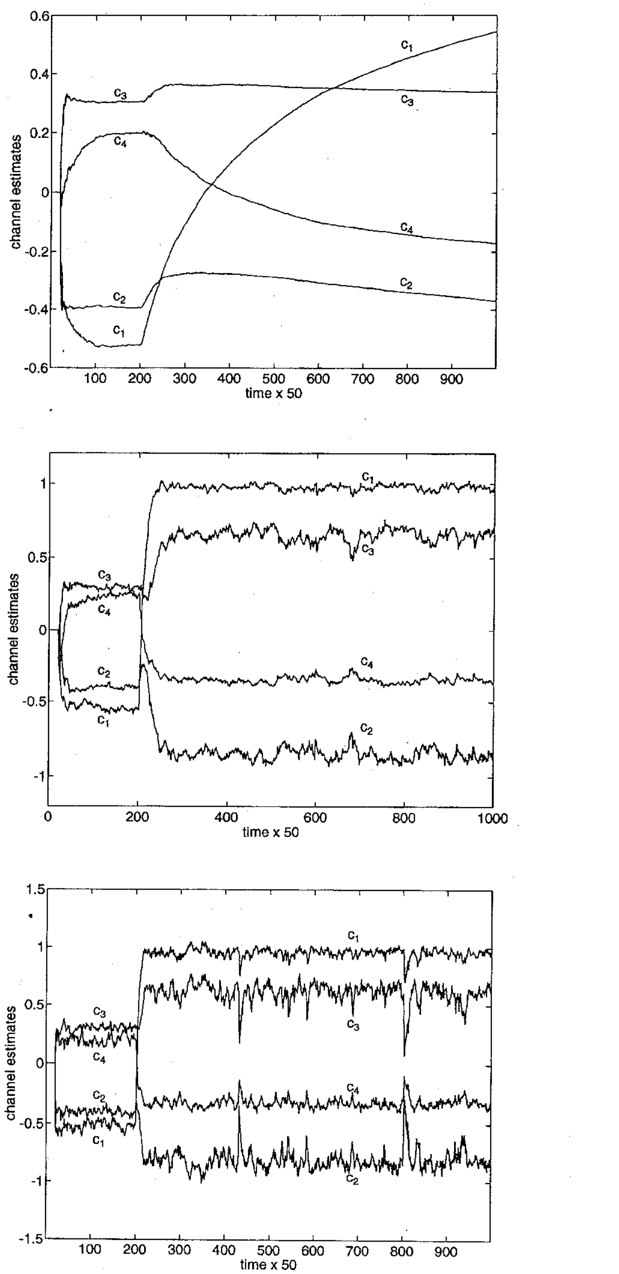
Fig. 5. Time evolution of IIR(4) channel parameters estimates.

Fig. 5 show how the estimates of the channel parameters converge with time for two different time-invariant IIR(4) channel models. The noise variance and the channel parameters are specified below each figure.

Equalization of time-varying channel: Consider a jump time varying IIR(4) channel with coefficients

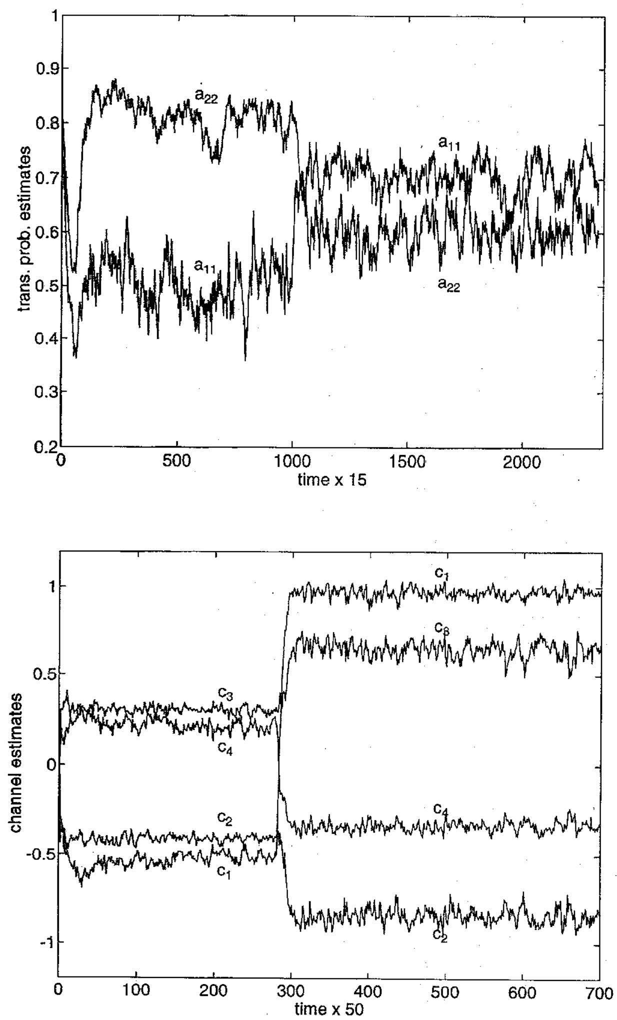
$$C = \begin{cases} (-0.5\ -0.4\ 0.3\ 0.2)' & 1 \le k \le 10000 \\ (1.0\ -0.9\ 0.7\ -0.36)' & 10000 < k \le 50000 \end{cases}$$
$$\text{and} \quad \sigma_w = 0.6. \tag{4.6}$$

Fig. 6 shows how the HMM-ELS algorithm tracks the channel coefficients for different forgetting factors. With no forgetting ($\lambda = 1$), the convergence is slow and the estimates do not converge to the true values even after 50 000 samples. Fig. 6 also shows that by using a forgetting factor $\lambda = 0.998$ and $\lambda = 0.995$, the convergence is significantly faster. As expected with increasing forgetting (smaller λ), there is increasing local drift in the estimates around the true parameter values.



$\lambda = 1.0$, $\lambda = 0.998$, $\lambda = 0.995$ respectively

Fig. 6. Equalization of "jump" time-varying channel.



$\lambda_A = 0.995$, $\lambda = 0.99$

Fig. 7. Equalization of jump time-varying channel with jump time-varying transition probabilities.

Equalization of Time-Varying Channel with Time-Varying Input Statistics: We consider the case where the channel coefficients, as well as the Markov chain transition probabilities jump change as follows:

$$C = \begin{cases} (-0.5\ -0.4\ 0.3\ 0.2)' & 1 \le k \le 15000 \\ (1.0\ -0.9\ 0.7\ -0.36)' & 15000 < k \le 50000 \end{cases} \tag{4.7}$$

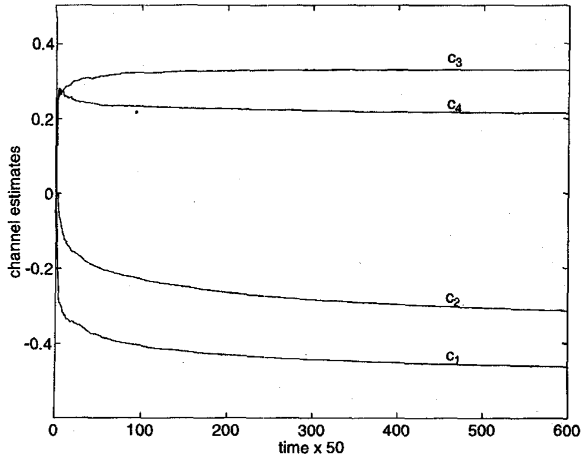
$$\begin{aligned} a_{11} &= 0.5, \quad a_{22} = 0.8 \qquad 1 \le k \le 15000 \\ a_{11} &= 0.7, \quad a_{22} = 0.6 \qquad 15000 < k \le 50000. \end{aligned} \tag{4.8}$$

In addition, $\sigma_w = 0.6$. Fig. 7 shows the time evolution of the channel and transition probability estimates with different forgetting factors.

Very Low SNR Performance: With increasing noise variance, the bias in the estimates increases, as can be seen from the tables, particularly from Table III as compared to Table II. However, since the SNR in typical communication systems is much higher than this example, the authors are confident of the performance of the HMM-ELS in such systems.

B. HMM with Higher Number of States

In this subsection, estimates obtained for Markov chains with higher number of states are presented. Recall that the computational cost is $O(N^2)$ which is much less than using a FIR channel approximation to the IIR channel, where the complexity is $O(N^{\tilde{p}})$ for a $\tilde{p}$ length approximation.



$N = 5, C = (-0.5\ -0.4\ 0.3\ 0.2)', \sigma_w = 0.5$

Fig. 8. IIR(4) channel estimates for five-state Markov input.

Three State Markov Chains ($N = 3$): Table V is for three different IIR(4) channels with $M = 2, r = 6$.

$$\sigma_w = 0.5, \quad A = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}, \quad q = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \quad (4.9)$$

Five State Markov Chains ($N = 5$): Fig. 8 shows the channel estimates of an IIR(4) channel with $M = 2, r = 6$,

$$\sigma_w = 0.5, \quad A = \begin{pmatrix} 0.5 & 0.05 & 0.15 & 0.15 & 0.15 \\ 0.15 & 0.5 & 0.05 & 0.15 & 0.15 \\ 0.15 & 0.15 & 0.5 & 0.05 & 0.15 \\ 0.15 & 0.15 & 0.15 & 0.5 & 0.05 \\ 0.05 & 0.15 & 0.15 & 0.15 & 0.5 \end{pmatrix},$$

$$q = \begin{pmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{pmatrix} \quad (4.10)$$

C. Comparative Study of Unrelaxed and Relaxed SPR Algorithms

The HMM-ELS algorithm was was applied to a non-SPR IIR(2) channel with $C = (-1\ -0.9)'$

$$A = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}, \quad q = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad \text{and} \quad \sigma_w = 1.0 \quad (4.11)$$

Fig. 9 plots the channel coefficient estimates using the HMM-ELS algorithm with i) standard (unrelaxed) ELS with $r = 2$, ii) standard overparametrized ELS with $r = 20$, and iii) relaxed SPR ELS algorithms $r = 4$, $M = 2$, respectively. Fig. 9 shows that the bias in the estimates is quite large for the standard ELS scheme and reduces somewhat when overparametrization is applied. In comparison, the relaxed SPR algorithm performs extremely well to give an estimate of $\hat{C} = (-0.9493\ -0.84)'$. Numerous simulations show that overparametrization does not always help, and hence, relaxed SPR ELS algorithm emerges as the obvious choice.

TABLE V
HMM-ELS Parameter Estimation Error for IIR(4) Channel, $N = 3$

$\sigma_w = 0.5$

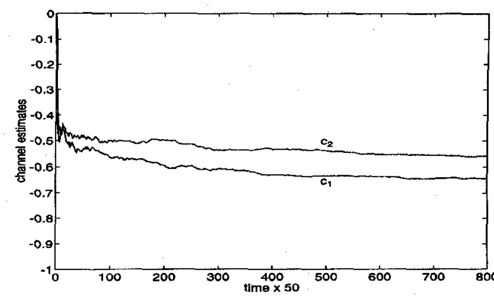
	true	mean	rms error
c_1	1	0.9820	0.0212
c_2	-0.9	-0.9086	0.0202
c_3	0.7	0.7208	0.0276
c_4	-0.36	-0.3961	0.0377

$\sigma_w = 0.5$

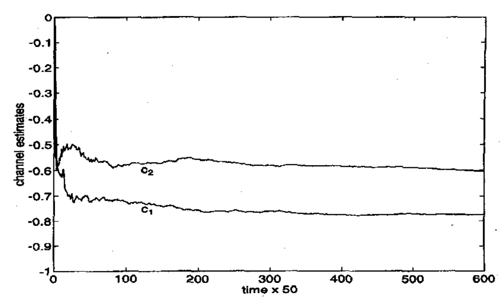
	true	mean	rms error
c_1	0.6	0.5836	0.0194
c_2	-0.5	-0.5097	0.0220
c_3	0.3	0.3104	0.0259
c_4	-0.16	-0.1753	0.0211

$\sigma_w = 0.5$

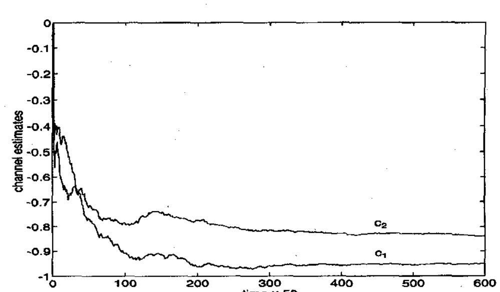
	true	mean	rms error
c_1	-0.5	-0.4944	0.0117
c_2	-0.4	-0.3901	0.0180
c_3	0.3	0.2806	0.0222
c_4	0.2	0.1692	0.00323



HMM-ELS algorithm with standard ELS, $r = 2$



HMM-ELS algorithm with overparametrized ELS, $r = 20$

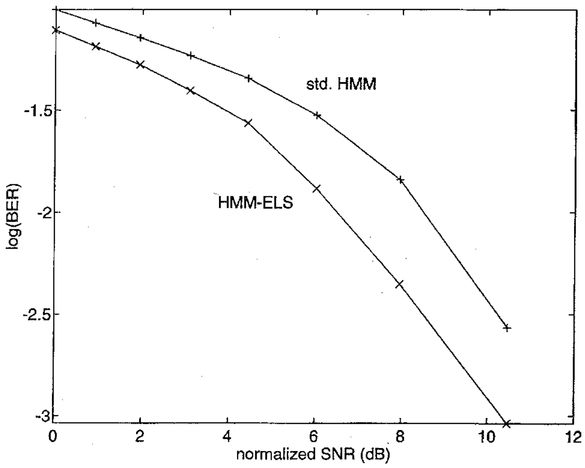


HMM-ELS algorithm with relaxed SPR ELS
$N = 2, C = (-1\ -0.9)', \sigma_w = 1.0$

Fig. 9. Necessity of using relaxed SPR ELS.

D. Error Probability Comparison with Standard HMM Algorithm

Since one of the objectives of the HMM-ELS algorithm is to obtain the filtered state estimates, it is of interest to see how



$N = 2,\ q = (-1\ 1),\ C = (0.6\ -0.5\ 0.3\ -0.16)'$

Fig. 10. Error probability in state estimates versus SNR.

they are compared to state estimates as obtained via standard HMM algorithm assuming the actually colored noise to be white. HMM-ELS algorithm performs substantially better than the standard HMM algorithm as is seen from Fig. 10. The bit-error rate (BER) is obtained as the fraction of number of errors out of 50 000 sample points averaged over 50 simulation runs. The signal to noise ratio (SNR) is computed as

$$SNR(dB) = -10 \log \sigma_w^2. \tag{4.12}$$

This is the normalized SNR (normalized wrt $\sigma_w = 1$) since the channel $C = (0.6\ -0.5\ 0.3\ -0.16)'$ is fixed, and the Markov chain has a fixed transition probability matrix (as specified in Fig. 10).

We do not study state estimation and error probabilities in detail because once the channel has been equalized by the HMM-ELS algorithm, various "standard" methods (such as Adaptive Viterbi-type schemes) may be more suitable for data recovery.

V. Comparison of HMM-ELS Algorithm with Other Algorithms

We now compare the HMM-ELS algorithm with a truncated FIR approximation algorithm and the CMA. While comparing with CMA, IID, as well as Markov chain inputs, are considered.

A. Comparison of HMM-ELS Algorithm with Truncated FIR Approximation

As mentioned in Section I, it is possible to approach our equalization problem by approximating the IIR channel by a FIR channel and then estimating the FIR channel coefficients using the recursive EM algorithm as in [2]. We now compare this approach with the HMM-ELS algorithm.

Consider the IIR(4) channel $C = (-0.5\ -0.4\ 0.3\ 0.2)'$ of Table V. For the same data as in Table V, assuming a FIR(4) channel, we ran the recursive EM algorithm in [2] to estimate the FIR coefficients. The estimate of the FIR channel after

TABLE VI
CMA Parameter Estimation Error for IIR(2) Channel, $N = 2, \sigma_w = 0.6$

	true	mean	rms error
c_1	0.4	-0.3771	0.7789
c_2	-0.6	-0.6513	0.0677

	true	mean	rms error
c_1	0.8	0.1841	0.6180
c_2	-0.8	-0.7670	0.0614

	true	mean	rms error
c_1	0.4	-0.0240	0.4242
c_2	0.5	0.8139	0.3142

	true	mean	rms error
c_1	0.2	0.0174	0.1830
c_2	0.7	0.7927	0.0942

	true	mean	rms error
c_1	0.6	-0.0690	0.6691
c_2	0.3	0.8414	0.5416

	true	mean	rms error
c_1	0.5	-0.9368	1.4385
c_2	-0.4	-0.7330	0.3360

	true	mean	rms error
c_1	1.5	1.1153	0.3881
c_2	-0.9	-0.8074	0.0973

	true	mean	rms error
c_1	1.0	-3.0463	4.0545
c_2	-0.3	-3.1359	2.8477

50 000 points is $1 - 0.5559z^{-1} - 0.0446z^{-2} + 0.4148z^{-3} - 0.3350z^{-4}$. The equivalent IIR(4) channel obtained by long division is $C = (-0.5559\ -0.3536\ 0.1934\ -0.0127)'$. This estimate is much worse than the HMM-ELS estimate (see Table V).

To obtain better estimates, longer FIR channel approximations are required. This involves exponentially increasing computational cost; for a $\tilde{p}$ length FIR approximation the cost is $O(N^{\tilde{p}}T)$. This demonstrates the attractiveness of our HMM-ELS algorithm.

B. Comparison of HMM-ELS Algorithm with Constant Modulus Algorithm

We now compare the HMM-ELS and CMA algorithms. We compare the mean and rms error of the channel estimates once the algorithms have converged.

We do not compare convergence rates of the two algorithms because i) the convergence of the CMA is highly dependent on initial conditions, and ii) the convergence rate of the CMA can drastically be changed by choosing different step sizes. Similarly, by using different forgetting factors, the convergence rate of the HMM-ELS algorithm can also be changed.

The CMA simulations were run for data lengths of 100 000 points. Each simulation experiment was replicated fifty times. In addition, the CMA was initialized at the true channel inverse, i.e., $\kappa(1 - c_1 \cdots -c_p)'$ where κ is the scale factor so chosen that the variance of the output of the equalizer matches the known unit variance of the source. Notice that κ will change as a function of channel and correlation.

Markov Chain Input: Tables VI and VII show the mean estimates and rms errors of the channel coefficient estimates using CMA for a Markov chain input. They are to be compared with Tables I and II, which show the corresponding results using the HMM-ELS algorithm.

IID Input: Since CMA was originally designed for IID data, it is worthwhile comparing the performance of HMM-ELS and CMA for IID input (i.e., when $a(i,j) = 1/N$ for all i, j).

TABLE VII
CMA PARAMETER ESTIMATION ERROR FOR IIR(4) CHANNEL, $N = 2$

$\sigma_w = 0.2$

	true	mean	rms error
c_1	0.6	0.1704	0.4300
c_2	-0.5	-0.2059	0.2943
c_3	0.3	0.0736	0.2265
c_4	-0.16	-0.0891	0.0712

$\sigma_w = 1.0$

	true	mean	rms error
c_1	0.6	-1.9175	2.5263
c_2	-0.5	-2.3782	1.9051
c_3	0.3	-1.7365	2.0514
c_4	-0.16	-1.1848	1.0304

$\sigma_w = 0.16$

	true	mean	rms error
c_1	-0.5	-1.6764	1.1982
c_2	-0.4	-5.6491	5.3053
c_3	0.3	3.9455	4.2730
c_4	0.2	2.8070	3.0265

$\sigma_w = 1.0$

	true	mean	rms error
c_1	-0.5	-1.2577	0.7994
c_2	-0.4	-1.7417	0.3890
c_3	0.3	-1.6927	2.0125
c_4	0.2	1.3354	1.5452

TABLE VIII
HMM-ELS PERFORMANCE FOR i.i.d. IPUT AND IIR(2) MODEL, $N = 2, \sigma_w = 0.6$

	true	mean estimate	rms error
c_1	0.6	0.5864	0.0152
c_2	0.3	0.3134	0.0151

	true	mean estimate	rms error
c_1	1.0	0.9485	0.0524
c_2	-0.3	-0.2485	0.0524

TABLE IX
CMA PERFORMANCE FOR i.i.d. INPUT WITH IIR(2) CHANNEL, $N = 2, \sigma_w = 0.6$

	true	mean estimate	rms error
c_1	0.6	0.5277	0.0762
c_2	0.3	0.3499	0.0534

	true	mean estimate	rms error
c_1	1.0	0.7224	0.2804
c_2	-0.3	-0.0712	0.2317

Table VIII shows the HMM-ELS parameter estimation error for IID data for an IIR(2) channel. Table IX shows the parameter estimation errors obtained by CMA using the same input.

Discussion: By comparing the tables we conclude the following:

1) Even at high SNR, the HMM-ELS algorithm yields significantly better channel estimates.
2) At moderate to low SNR, the rms errors from the HMM-ELS are orders of magnitude lower than CMA. (Compare Table II with Table VII when $\sigma_w = 1$).
3) When the source is IID instead of Markov, the CMA performance improves (compare Table IX with Table VI). However, HMM-ELS still performs significantly better than CMA (compare Table VIII with Table IX).
4) Of course, CMA involves $O(p)$ computations at each time instant, which is computationally much cheaper than HMM-ELS.

VI. CONCLUSION

We have presented a suboptimal computationally efficient recursive blind equalization algorithm for IIR channels with finite-state Markov inputs. The algorithm combines a hidden Markov model estimator with a relaxed SPR extended least squares estimator, and is termed the HMM-ELS algorithm. Although, we have not been able to prove convergence of the algorithm, simulations show that it performs extremely well, even in low SNR.

We believe that similar combinations of cross-coupled linear estimators (e.g., Kalman filters) and nonlinear estimators (e.g., HMM estimator) can be used for a variety of other problems like speech coding and pulse train de-interleaving.

As a future research topic, it is worthwhile treating the above equalization problem as a linear estimation problem with correlated noise (filtered Markov chain). It may be possible to use instrumental variable techniques [9] to effectively whiten this noise and then estimate the channel coefficients.

ACKNOWLEDGMENT

The authors would like to thank Prof. C. R. Johnson of the Electrical Engineering Department of Cornell University and Prof. J. B. Moore of the Department of Systems Engineering at Australian National University for useful discussions.

REFERENCES

[1] V. Krishnamurthy, B. Wahlberg, and J. B. Moore, "Factorizations that relax the positive real condition in continuous-time and fast-sampled ELS schemes," *Int. J. Adaptive Contr. Signal Processing,* vol. 4, no. 5, pp. 389–414, Sept. 1990.

[2] V. Krishnamurthy and L. B. White, "Blind equalization of FIR channels with Markov inputs," presented at *IFAC Int. Conf. Adaptive Syst. Contr. Signal Processing,* Grenoble, France, July 1992.

[3] L. B. White and V. Krishnamurthy, "Adaptive blind equalization of FIR channels using hidden Markov models," presented at *IEEE Conf. Commun.,* Geneva, Switzerland, May 1993.

[4] 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, Aug. 1993.

[5] J. B. Moore, M. Niedzwiecki, and L. Xia, "Identication/prediction algorithms for ARMAX models with relaxed positive real conditions," *Int. J. Adaptive Contr. Signal Processing,* vol. 4, no. 1, pp. 49–67, 1990.

[6] 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. 42, no. 12, pp. 3535–3539, Dec. 1994.

[7] S. H. Chung, V. Krishnamurthy, and J. B. Moore, "Adaptive processing techniques based on hidden Markov models for characterizing very small channel currents buried in noise and deterministic interferences," *Philo. Trans. Royal Soc.,* London B, 1991, pp. 357–384.

[8] 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.

[9] L. Ljung, *System Identification Theory for the User.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[10] P. E. Caines, *Linear Stochastic Systems.* New York: Wiley, 1988.

[11] R. J. Elliott, *Stochastic Calculus and Applications,.* New York: Springer-Verlag, 1982.

[12] J. Goutsias and J. M. Mendel, "Optimal simultaneous detection and estimation of filtered discrete semi-Markov chains," *IEEE Trans. Inform. Theory,* vol. 34, no. 3, pp. 551–568, May 1988.

[13] J. Tugnait, "Adaptive estimation and identification for discrete systems with Markov jump parameters," *IEEE Trans Automat. Contr.,* vol. AC-27, no. 5, pp. 1054–1065, Oct. 1982.

[14] A. Benveniste and M. Goursat, "Blind equalizers," *IEEE Trans. Commun.,* vol. COM-32, no. 8, pp. 871–883, Aug. 1984.

[15] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.,* vol. COM-28, no. 11, pp. 1867–1875, Nov. 1980.

[16] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions.* New York: Wiley, 1985.

[17] P. J. Harrison and C. F. Stevens, "Bayesian forecasting," *J. Royal Stat. Soc. B,* vol. 38, pp. 205–247, 1976.

[18] H. J. Larson and B. O. Shubert, *Probabilistic Models in Engineering Sciences.* New York: Wiley, 1979, vol. 2.

[19] J. D. Hamilton, "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica,* vol. 57, no. 2, pp. 357–384, 1989.

[20] G. C. Goodwin and K. W. Sin, *Adaptive Filtering Prediction and Control.* Englewood Cliffs, NJ: Prentice-Hall, 1984.

[21] M. V. Eyuboglu and S. V. H. Quereshi, "Reduced state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, no. 36, pp. 12–20, 1988.

[22] P. Billingsley, *Ergodic Theory and Information*. New York: Wiley, 1965.

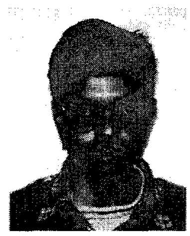
[23] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Trans. Commun.*, no. 37, pp. 428–436, 1989.

[24] K. Shanmugam, *Digital and Analog Conmmunication Systems*. New York: Wiley, 1985.

Vikram Krishnamurthy (S'90–M'91) was born in India in 1966. He received the bachelor's degree in electrical engineering (first class honors) from the University of Auckland, New Zealand, in 1987 and the doctoral degree from the Department of Systems Engineering, Australian National University, Canberra, in 1991.

From 1992 to 1993, he was a senior research engineer at the Center for Robust and Adaptive Systems at Australian National University. He is currently a senior lecturer with the Department of Electrical Engineering, University of Melbourne, Melbourne, Australia. His research interests are in estimation theory, time-series analysis, statistical signal processing, and systems theory.

Dr. Krishnamurthy was awarded the prestigious Queen Elizabeth II Fellowship in 1993.



Subhrakanti Dey was born in Calcutta, India, in 1968. He received the Bachelor of Technology and Master of Technology degrees from the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, India, in 1991 and 1993, respectively. Since July 1993, he has been pursuing the Ph.D. degree from the Department of Systems Engineering, Australian National University, Canberra.

His current research interests are statistical and adaptive signal processing, adaptive control, nonlinear filtering, and control and robust filtering.

James P. LeBlanc received the B.S. degree in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1983. In 1986, he was awarded Raytheon's Miccioli Scholarship and attended Cornell University, Ithaca, NY, USA, receiving the Master of Engineering degree in 1987.

After working a number of years in the industry with Raytheon Company's Submarine Signal Division and later with Motorola's Digital Signal Processing Operation, he returned to graduate school in 1991. Since then, he has been with Cornell University's Electrical Engineering Department as a Ph.D. student and research assistant. His research interests include adaptive signal processing (particularly blind equalization), statistical signal processing, communications systems theory, parameter estimation, and application of algebraic geometry techniques to systems theory.