International Journal of Wireless Information Networks, Vol. 14, No. 1, March 2007 (2006)
DOI: 10.1007/s10776-006-0052-z

Channel and Data Estimation for Ad Hoc Networks and Cognitive Radio

Larry N. Singh,[1] and Galigekere R. Dattatreya[2,3]

Received: 26 April 2005; Accepted: 23 October 2006

Estimation of channel and data characteristics by the receiver is important in adaptive wireless transmission protocols and in cognitive radio. This paper formulates the estimation problem with the help of an illustrative example from the IEEE 802.11a OFDM standard. The problem reduces to the estimation of the common component variance and mixing probabilities in a “nite Gaussian mixture, with known values for component means. Using the known component means, $l_1, \ldots, l_M$, a set of non-linear transformations, $\sinh l_i x$ and $\cosh l_i x$ of the data (mixture random variable X) are used to develop convergent and computationally e -cient estimators for both the noise variance and the vector of symbol probabilities. The estimation equations can be implemented recursively or with a batch processing algorithm. Asymptotic variances of the estimates and the Cramer…Rao minimum variance bounds are derived. The estimates converge to true unknowns even when the sequences of noise and data symbols are dependent sequences. The OFDM example is simulated with parameters corresponding to the highest acceptable error rate. For a time-varying channel model chosen from the literature, it is shown that our estimator receives considerably more than adequate amount of data during an average time interval of unchanging channel characteristics. Analytical results, numerical results and related issues are discussed.

KEY WORDS: Ad hoc networks; cognitive radio; channel and data estimation; adaptive wireless transmission; asymptotic variance

1. INTRODUCTION

Packet radio networks are being deployed in such a wide array of diverse applications, that concerns of a crisis in spectrum-availability are being raised. Cognitive radio [1], i.e. radio systems with adaptive intelligence, is receiving increasing attention by researchers and developers in an attempt to overcome such spectrum congestion bottlenecks. Cognitive radio takes advantage of the rapidly increasing sophistication of radio equipments• capabilities to sense the spectral environment and adaptively change the operating frequency, power, modulation, and other parameters, in order to dynamically reuse available spectrum without a ect-ing other (possibly licensed) users. Currently, Software De“ned Radio[2] is a technology developed to permit radio equipment to change the operating frequencies during their operation, should the need arise. This technology is ideally suited for cognitive radio. Rules for the operation of cognitive radio are being developed by regulatory bodies such as the Federal Communications Commission (FCC).

As a consequence of the proposed design of cognitive radio, multiple users (possibly belonging to different groups or organizations) may operate their radios with time-varying but overlapping space, frequency and modulation schemes. Cognitive radio approaches for secondary (unlicensed), opportunistic

[1] School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[2] Department of Computer Science, The University of Texas at Dallas, MS EC 3.1, 2601 North Floyd Road, P.O. Box 830688 Richardson, TX75083-0688, USA
[3] E-mail: datta@utdallas.edu

users, which take advantage of the available spectrum without interfering with primary (licensed) users• transmissions are being developed [1]. Multiple groups of secondary users may need to compete to use the available spectrum intermittently. Secondary users may also alter their spectrum or modulation techniques, or both, in order to continue using radio frequencies in conjunction with primary users or other secondary users. Of course, in the unlicensed Industrial, Scienti“c and Medical (ISM) band of the radio spectrum, there are no primary users and a free-for-all situation arises, which is subject to common regulations.

In situations wherein users need to dynamically change and adapt operating parameters as described above, channel and user assessment and monitoring by the radio receiver is imperative. Several different channel and data parameter estimation problems become important for the successful operation of such radio networks. Brodersen et al. [1] specify the following channel quality estimation procedures as part of the system functions of the radio equipment. The transmitter parameters are based upon the results of the channel sounding during link setup. The physical layer should also continuously estimate the quality of the sub-channels by analyzing the data packets received during an ongoing communication.

Adaptive wireless transmission protocols have also been proposed for current wireless communication technologies in Pursley and Wilkins [3], and Gass et al. [4]. In [3], Pursley and Wilkins propose and evaluate variable power, code rate and symbol rate adaptation to match the multipath pro“le and propagation loss of the wireless channel. The approach requires the receiver to estimate some measure of postdetection signal quality (PDSQ). Pursley and Wilkins show how to obtain the signal-to-noise-density ratio (S/N_0) from the PDSQ for di erent combinations of ranges of S/N_0 and numbers of multipath components. The approach in [3] is to map the PDSQ to S/N_0 starting from any method for computing PDSQ. The value of the PDSQ depends on the parameters of the data and the noise at a point past the demodulation scheme. For a complete monitoring of the channel, all the parameters of noise and data are useful. The objective of this paper is to comprehensively develop a scheme for monitoring the channel and to analyze the performance of such a scheme.

The technique developed in this paper is suitable for implementation at the receiver. The development and analysis are very general and completely mathematical in nature. The performance of the estimators developed in this paper are thoroughly analyzed.

1.1. Literature Survey

The demodulated data from the received signal turns out to be a mixture random variable. Research interest in mixture densities and applications have steadily grown over the last few decades. The EM algorithm [5], kernel density estimation [6], the method of moments [7] and of moments of transformations [8] and the minimization of some information measures [9] are the common approaches. Tugnait, Tong and Ding [10] provide an overview of the techniques applicable to blind estimation in wireless data communication. Wang and Chen [11] also discuss applications of blind turbo equalization in noise. Monin and Salut [12] deals with minimum variance estimation of parameters when the parameter to be estimated is constrained by bounds.

The problem of parameter estimation continues to be of practical importance in digital wireless communication and signal processing. Proposals for cognitive radio are expected to increase the motivation to develop simple estimation approaches on the one hand and sophisticated techniques on the other hand. Simpler approaches would be necessary for cursory examination of operating conditions to monitor various subchannels for possible changeover to a different subchannel, should the environment demand such a preemption of a transmitter from a currently used subchannel. More accurate estimation of parameters is required for the receiver to process the received signal and detect the sequence of symbols with the optimal decision scheme. Multiple input and multiple output (MIMO) data communication system have similar requirements. Pham and Zoubir [13] present a sequential M-estimation algorithm to estimate an unknown parameter vector. The observation sequence is a known linear transform of an unknown vector plus a noise sequence. The noise probability density function is a mixture of a Gaussian density and a minor contamination by an unknown symmetric function. They cite the relevance of this problem in indoor and urban radio communication. Farquharson et al. [14] introduce an algorithm for estimating parameters of polynomial phase signals in white Gaussian noise. The cited applications are in radar, sonar, telemetry, communications, and power systems. Deng et al. [15]

study a decision-directed channel estimation scheme in a MIMO system. The received signal at the m-th antenna is due to a block of L symbols corrupted by Gaussian noise, possibly colored. The optimal parameter estimation requires inversion of a square matrix of L rows. Furthermore, naturally, the matrix is dependent on transmitted data symbols. Instead of inverting the matrix, they find approximate iterative solutions. Different decompositions of the matrix result in different well known techniques such as Jacobi, Gauss-Seidel, and successive overrelaxation iteration. To overcome the lack of knowledge of the transmitted symbols, they use pilot symbols to obtain the initial estimate of the channel and soft decision feedback to improve the channel estimate after each iteration. Our unsupervised estimation approach (without requiring a knowledge of transmitted symbols) developed in the present paper find applications in the estimation of such MIMO channels. Other recent publications on the use of estimation theory in communication and signal processing are by Jeong and Gray [16] and Vorobyov et al. [17]. Jeong and Gray [16] compare the EM algorithm and a recent alternative, the GMVQ (Gauss Mixture Vector Quantization) approach for image retrieval. They show that although both algorithms have about the same performance, GMVQ technique has half the computational complexity of the EM algorithm. Vorobyov et al. [17] develop an estimator for a random vector when it is corrupted by additive white noise. Both the unknown vector and the noise are assumed to be zero mean Gaussian with known covariance matrices. The result is a linear estimator that minimizes the mean squared error with a certain selected probability.

Dattatreya and Fang [8] propose and study Gaussian transformations of the data and develop a method of moments of transformations to estimate the additive noise variance and symbol probabilities when the signal means for different symbols are known. The choice of a single control parameter is crucial there and there is no guaranteed approach to verify the existence of a viable parameter or to evaluate this parameter if one exists. Simpler data transformations wherein control parameters (if any) can be easily selected, and corresponding estimators with guaranteed convergence are desirable. In this paper, we propose and study hyperbolic transforms of the data. The approach needs no external control parameters and is guaranteed to converge. Results are summarized in the following section.

1.2. Summary and Organization of the Paper

Section 2 develops the estimation problem starting from a practical example of a communication system organized with an IEEE standard protocol. Section 3 defines the hyperbolic data transformations and derives the estimator equations for the common noise variance and the symbol probabilities. The equations connect the expectations of the data transformations and the M unknowns (the $M-1$ symbol probabilities and the noise variance). These equations are manipulated to eliminate the variance; the resulting $M-1$ equations are linear in the $M-1$ symbol probabilities. Section 4 is devoted to the analysis of the estimator. The linear equations for symbol probabilities are first shown to be linearly independent, thus verifying the sufficiency of the proposed data transformations, and giving us a computationally efficient closed form solution, provided the exact values of the expectations of data transformations are known. Substituting sample averages of the data transformations for the exact expectations gives us corresponding estimator equations. Even though the estimator equations for the symbol probabilities are linear, their closed form vector estimator is a non-linear function of the multiple data transformations. The asymptotic covariance matrix of the estimator for symbol probabilities and noise variance is derived. Simulation experiments are conducted and the Cramer–Rao minimum variance bounds are evaluated for the estimation problem, for comparison, in section 5. An alternative estimator in an earlier publication is also simulated for comparison; the results and properties of the two estimators are compared and contrasted. Section 6 concludes the paper.

2. PROBLEM FORMULATION

As a practical example, the IEEE 802.11a standard Orthogonal Frequency Division Multiplexing (OFDM) scheme, with a data rate of 36 Mega-bits per second is used as a reference, where appropriate. In this standard, data are represented by complex numbers. Each of the real and imaginary parts of a complex number can be one of four possible values, giving 16 possible complex numbers for 16 different symbols. Therefore, each instance of a symbol represents 4 bits of the original binary sequence of the data. The real and imaginary components of a symbol are represented as sine and cosine waves of a specified duration

(3.2 μs). The sine wave is constructed with one of two different peak values and one of two different phases, 0 and π radians. This approach is referred to as Pulse Amplitude Modulation (PAM) with four pulse levels. The cosine wave has similar options as the sine wave. The combination of the sine and cosine waves is called Quadrature Amplitude Modulation (QAM). The sum of such a sine and cosine wave is sometimes called a complex wave and can take on one of the 16 possible forms for the 16 different values that a symbol can attain. A total of 48 successive symbols are transformed into 48 such complex waves, each of a different frequency. The bandwidth is actually divided into 52 sub-carriers. The other 4 sub-carriers are used for pilot signals. The sum of these 48 complex waves for the specified duration of 3.2 μs constitutes an OFDM symbol. A complete data packet is constructed from a sequence of such OFDM symbols, preceded by control signals. Each data packet begins with a preamble sequence of several known symbols. The purpose of repeating the same sequence of symbols in every packet is to allow the receiver to perform synchronization, frequency acquisition and Automatic Gain Control (AGC). The AGC process is formulated later, in Eq. (2). The maximum number of bytes (octets) allowed in a packet is 4095 in the OFDM standard. Consider 1024 bytes (octets) of a payload data in a packet, which corresponds to 512 complex numbers or 2048 real numbers. Each real number takes on one of four values. A 1024 byte frame takes about 0.25 ms of time to transmit.

Before the packet arrives at the receiver, the transmitted packet undergoes attenuation, phase shift, and is also corrupted by noise. The known preamble sequence of symbols allows the receiver to compute fairly accurate values for the attenuation and phase shift. The sine and cosine portions of each frequency in each segment are separated and their amplitudes and phases obtained through orthogonal transformation processes. A sine wave multiplied by a sine wave and integrated over the segment gives the required amplitude and sign. Unfortunately, the signal is corrupted by additive noise and this offsets the value obtained for the sine wave by a small but unknown value. Similarly, all other component values of all the complex numbers are obtained at the receiver and these are slightly corrupted versions of the original components of the transmitted complex numbers. The corruption mechanism is usually modeled as addition of zero mean Gaussian random variables for the real and imaginary parts of the complex number, with justification (see [18]).

A symbol value for each received complex number is computed and assigned through a detection algorithm such as a maximum a posteriori probability detection process. This detection process requires a priori symbol probabilities. The statistical structure of the data sequence can also be time-varying. For example, over TCP connections (Tanenbaum [19]) between the same pair of transmitter and receiver, different types of data, such as computer files, voice, or image signals may be transmitted. If the receiver can estimate the statistics of the data and of the additive noise as it functions and operates on incoming data signals, its performance will be superior. This will favorably affect the overall performance of the data networks in the form of reduced retransmissions of data frames at the data link or of the data packets at the transport layer. Such an adaptive estimation with no other aid from the transmitter is called blind or unsupervised parameter estimation. Such a channel and data parameter estimation procedure would (i) satisfy one of the system function requirements specified in cognitive radio [1], and (ii) provide the needed data for the postdetection signal quality (PDSQ) calculation in [3].

Consider a PAM scheme with M symbols $S_1, \ldots, S_M$ and corresponding probabilities of occurrence $P_1, \ldots, P_M$. For the above OFDM example, $M = 4$. Each symbol may represent two bits in a bit stream of data. For the symbol S_i, the transmitter constructs a sinusoid of a carrier frequency f_c and an amplitude a_i for the duration T of the symbol. The quantity Tf_c is a large integer. In the earlier OFDM example, Tf_c is approximately 150,000. The a_i values typically have equal increments and are spread symmetrically around the origin. The sinusoids for a_i and $-a_i$ are out of phase with each other. A sequence of symbols produces a succession of such segments of sinusoids resulting in the data signal. A data frame [19] consists of a sequence of possibly several hundreds of symbols. A sequence of prearranged symbols is included at the beginning of the data segment. This is the preamble in the OFDM example. The receiver has full knowledge of such a preamble symbol sequence. Let $s(t)$ be the time function modulating the carrier sinusoid of the entire data frame. The time function $s(t)\sin(2\pi f_c t)$ is the pulse amplitude modulated data signal. The receiver receives a time function

$$a s(t)\sin(2\pi f_c t + \phi) + g(t) \qquad (1)$$

where a is the unknown attenuation due to the transmission distance and any multi-path fading, / is the phase shift, and g(t) is the zero mean additive noise signal. At the receiver, the parameter a is estimated fairly accurately with the help of the known pilot carrier symbol sequence. Similarly, the phase shift / is obtained from the sinusoids of the known preamble sequence. The received signal is multiplied by the quantity $\frac{g}{a}$ where g is a receiver parameter. This is automatic gain control. The resulting signal is

$$r(t) = gs(t)\sin(2pf_c t + /\ ) + \frac{g}{a} g(t). \qquad (2)$$

The demodulator multiplies the above processed signal r(t) by $h\sin(2pf_c t + /\ )$ and integrates over the symbol time period T to result in

$$\begin{aligned} y &= \int_0^T r(t) h \sin(2pf_c t + /\ ) dt \\ &= \frac{T}{2} gha_i + \frac{gh}{a}\int_0^T \sin(2pf_c t + /\ ) g(t) dt \qquad (3) \end{aligned}$$

where a_i corresponds to the symbol S_i transmitted.

$$t = \frac{gh}{a}\int_0^T \sin(2pf_c t + /\ ) g(t) dt \qquad (4)$$

corresponds to the outcome of the equivalent noise random variable Υ that a ects the detection scheme. The “nal composite random variable input to the detector is Y which is one of $\frac{Tgh}{2} a_i$ with the symbol probability P_i plus the zero mean noise random variable Υ . The noise Υ is usually modeled as a Gaussian random variable with good justi“cation. Let the variance of this noise be q^2. The set $\{\frac{Tgh}{2} a_1, \ldots, \frac{Tgh}{2} a_M\}$ is known at the receiver. Thus, in the M symbol system, the received random variable Y is an M component Gaussian mixture with known component means $b_i = \frac{Tgh}{2} a_i$ and unknown component variance q^2. If Y is scaled and translated as X = u + wY with known constants u and w, X is a Gaussian mixture with component means l_i = u + $w\beta_i$ and the variance of each component is $r^2 = w^2 q^2$. Therefore, without loss of generality, we can and do assume that $l_M = 0$ and $l_i > 0$, i = 1,..., M −1. Thus, the parameter estimation problem reduces to the estimation of the unknown noise variance, r^2, and the M unknown data (or symbol) probabilities P_1, ..., P_M, from a sequence of identically distributed mixture samples $\{X_1, \ldots, X_n\}$, each sample having the density of the random variable X given by

$$p(x) = \sum_{j=1}^{M} P_j \frac{1}{\sqrt{2pr^2}} \exp\left(-\frac{(x - l_j)^2}{2r^2}\right), \qquad (5)$$

with $l_i > 0$, i = 1,..., M −1 and $l_M = 0$. To ensure that the component densities are distinct, the component means are assumed to be distinct. In order for the mixture data to be in”uenced by all the components, $P_i > 0$, for all i, is required.

3. DEVELOPMENT OF THE ESTIMATOR

3.1. Data Transformations

We propose the following non-linear functions (transformations), using the hyperbolic sine (sinh) and hyperbolic cosine (cosh) functions

$$g_i(x) = \sinh(l_i x),\ i = 1, \ldots, M-1, \text{ and} \qquad (6)$$

$$h_i(x) = \cosh(l_i x),\ i = 1, \ldots, M-1. \qquad (7)$$

The expectations of $g_i(x)$ and $h_i(x)$ are obtained from the characteristic function of the Gaussian random variable [20]. De“ne

$$a_i = E[\sinh l_i X] = e^{\frac{r^2 l_i^2}{2}} \sum_{k=1}^{M-1} P_k \sinh l_i l_k \qquad (8)$$

$$\begin{aligned} b_i &= E[\cosh l_i X] \\ &= e^{\frac{r^2 l_i^2}{2}} \left(1 + \sum_{k=1}^{M-1} P_k (\cosh l_i l_k - 1)\right), \qquad (9) \\ & i = 1, \ldots, M-1, \end{aligned}$$

where $\sinh l_i l_k$ is interpreted as $\sinh(l_i l_k)$. Similarly, $\cosh l_i l_k$ is used to denote $\cosh(l_i l_k)$.

3.2. Linear Equations for P

Dividing Eq. (8) by Eq. (9) and rearranging, for each i,

$$\sum_{j=1}^{M-1} P_j \left(b_i \sinh l_i l_j + a_i (1 - \cosh l_i l_j)\right) = a_i,$$
$$i = 1, \ldots, M-1. \qquad (10)$$

Let Γ be the (M −1) × (M −1) matrix with elements,

$$c_{i,j} = b_i \sinh l_i l_j + a_i (1 - \cosh l_i l_j), \qquad (11)$$

P be the column vector of mixing proportions $\{P_1, ..., P_{M-1}\}$, and A be the column vector $\{a_1, ..., a_{M-1}\}^T$. Therefore,

$$\Gamma P = A. \tag{12}$$

Substitute the values for a_i and b_i in Eqs. (8), (9) and (10) to give

$$\left(\sum_{j=1}^{M-1} P_j \sinh l_i l_j\right) e^{\frac{l_i^2 r^2}{2}} + e^{\frac{l_i^2 r^2}{2}} \sum_{j=1}^{M-1}\sum_{k=1}^{M-1} P_j P_k \big((\cosh l_i l_k - 1)\sinh l_i l_j + (1 - \cosh l_i l_j)\sinh l_i l_k\big) = a_i \tag{13}$$

Consider the double summation in the previous equation:

$$\sum_{j=1}^{M-1}\sum_{k=1}^{M-1} P_j P_k \cosh l_i l_k \sinh l_i l_j - \sum_{j=1}^{M-1}\sum_{k-1}^{M-1} P_j P_k \cosh l_i l_j \sinh l_i l_k - \sum_{j=1}^{M-1}\sum_{k-1}^{M-1} P_j P_k \sinh l_i l_j + \sum_{j=1}^{M-1}\sum_{k-1}^{M-1} P_j P_k \sinh l_i l_k = 0, \tag{14}$$

by interchanging the roles of j and k in one pair. Therefore, Eq. (13) reduces to

$$\sum_{j=1}^{M-1} P_j \sinh l_i l_j e^{\frac{l_i^2 r^2}{2}} = a_i, \tag{15}$$

for every i which is identical to the corresponding Eq. (8).

In summary, Eq. (15) is obtained from Eq. (13), which is obtained from Eq. (10), separately for each i. Since (15) is identical to (8) for each i, uniqueness of a solution for P in (8) implies uniqueness of solution for P in (10). The following subsection shows the uniqueness of solution for (8). De“ne $\mathcal{S}$ as the $(M-1) \times (M-1)$ matrix with elements $\sinh l_i l_j$ for $1 \le i, j \le M-1$.

3.3. Proof that $\mathcal{S}$ is Non-singular

The Maclaurin series expansion for the hyperbolic sine function, $\sinh l_i l_j$ is

$$l_i l_j + \frac{l_i^3 l_j^3}{3!} + \frac{l_i^5 l_j^5}{5!} + \cdots \tag{16}$$

De“ne the following in“nite sequence

$$\left(\mathcal{U}_i = \frac{l_i^{2k+1}}{\sqrt{(2k+1)!}},\ k = 0, 1, 2, \ldots\right) \tag{17}$$

where l_i is a real number. For the problem at hand, there are $M-1$ sequences, $\mathcal{U}_1, \ldots, \mathcal{U}_{M-1}$.

Consider the vector space $\mathcal{V}$ consisting of the complex “eld of scalars, the $M-1$ in“nite sequences $\{\mathcal{U}_1, \ldots, \mathcal{U}_{M-1}\} \subset \mathcal{V}$, and the in“nite sequence of zeros being the zero vector 0. If $u = \{u_0, u_1, ...\}$ and $v = \{v_0, v_1, ...\}$ are vectors in $\mathcal{V}$ and c is a scalar, the scalar multiplication operation is de“ned as

$$c \cdot u = \{cu(0), cu(1), \ldots\}, \tag{18}$$

and the vector addition operation is de“ned as

$$u + v = \{u(0) + v(0), u(1) + v(1), \cdots\}. \tag{19}$$

In other words, $\mathcal{V}$ is simply the linear span generated by the vectors $\{\mathcal{U}_1, \ldots, \mathcal{U}_{M-1}\}$, and the “eld of complex numbers.

Lemma 1: The set of $M-1$ sequences $\{\mathcal{U}_1, \ldots, \mathcal{U}_{M-1}\}$ is a basis for $\mathcal{V}$.

Proof: Suppose the linear combination,

$$a_1\mathcal{U}_1 + a_2\mathcal{U}_2 + \ldots + a_{M-1}\mathcal{U}_{M-1} = 0, \tag{20}$$

where $a_1, a_2, ..., a_{M-1}$ are arbitrary, complex coe - cients. This implies,

$$a_1\{l_1, \frac{l_1^3}{\sqrt{3!}}, \ldots\} + \cdots + a_{M-1}\{l_{M-1}, \frac{l_{M-1}^3}{\sqrt{3!}}, \ldots\} = 0. \tag{21}$$

Therefore, the “rst $M-1$ equations in Eq. (21), can be written in a matrix form as

$$\begin{pmatrix} l_1 & l_2 & \ldots & l_M \\ \vdots & \vdots & \vdots & \vdots \\ \frac{l_1^{2i-1}}{\sqrt{(2i-1)!}} & \frac{l_2^{2i-1}}{\sqrt{(2i-1)!}} & \ldots & \frac{l_{M-1}^{2i-1}}{\sqrt{(2i-1)!}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{l_1^{2M-1}}{\sqrt{(2M-1)!}} & \frac{l_2^{2M-1}}{\sqrt{(2M-1)!}} & \ldots & \frac{l_{M-1}^{2M-1}}{\sqrt{(2M-1)!}} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_{M-1} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{22}$$

Post-multiply the coef“cient matrix in Eq. (22) by

$$\mathrm{Diag}[l_i]_{i=1}^{M-1}. \tag{23}$$

Let $h_i = l_i^2$, so the resulting product matrix is

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \ddots & \frac{1}{\sqrt{(2i-1)!}} & \vdots \\ 0 & \dots & 0 & \frac{1}{\sqrt{(2M-1)!}} \end{pmatrix} \begin{pmatrix} h_1 & h_2 & \dots & h_{M-1} \\ \vdots & \vdots & \vdots & \vdots \\ h_1^M & h_2^M & \dots & h_{M-1}^{M-1} \end{pmatrix}. \tag{24}$$

The constant matrix on the left is clearly non-singular since it is diagonal and all the diagonal entries are non-zero. The $(M-1) \times (M-1)$ matrix in (3) is a Vandermonde matrix [21] and hence, is invertible since each l_i, i = 1,..., $M-1$ is non-zero and distinct. Therefore, the only solution to Eq. (22) and hence to Eq. (20), is $a_1 = \dots = a_{M-1} = 0$. This implies that the set of $M-1$ vectors, $\{\mathcal{U}_1, \dots, \mathcal{U}_{M-1}\}$ is linearly independent, thus completing the proof. □

Equip $\mathcal{V}$ with the following inner product

$$\langle \mathcal{U}_i, \mathcal{U}_j \rangle = \sum_{k=0}^{\infty} \mathcal{U}_i(k) \overline{\mathcal{U}}_j(k) \tag{25}$$

$$= \sum_{k=0}^{\infty} \left(\frac{l_i^{2k+1}}{\sqrt{(2k+1)!}} \right) \left(\frac{l_j^{2k+1}}{\sqrt{(2k+1)!}} \right), \tag{26}$$

where $\bar{z}$ represents the complex conjugate of z. Notice that by this de“nition

$$\langle \mathcal{U}_i, \mathcal{U}_j \rangle = \sinh l_i l_j. \tag{27}$$

It is not dif“cult to verify that $\langle \cdot, \cdot \rangle$ is a valid inner product in $\mathcal{V}$, and hence $\mathcal{V}$ is a unitary space [21].

Lemma 2: The matrix $\mathcal{S}$ is non-singular for all distinct, positive component mean values.

Proof: The matrix with element at row i and column j equal to $\sinh l_i l_j$ may be written as the Hermitian matrix

$$\mathcal{S} = \begin{pmatrix} \langle \mathcal{U}_1, \mathcal{U}_1 \rangle & \langle \mathcal{U}_1, \mathcal{U}_2 \rangle & \dots & \langle \mathcal{U}_1, \mathcal{U}_{M-1} \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \mathcal{U}_{M-1}, \mathcal{U}_1 \rangle & \langle \mathcal{U}_{M-1}, \mathcal{U}_2 \rangle & \dots & \langle \mathcal{U}_{M-1}, \mathcal{U}_{M-1} \rangle \end{pmatrix}. \tag{28}$$

This matrix is also known as a Gram matrix. By Lemma 1 the vectors $\{\mathcal{U}_1, \dots, \mathcal{U}_{M-1}\}$ are linearly independent. Therefore, by theorem 2, page 110 of Lancaster and Tismenetsky [21] the Gram determinant of $\mathcal{S}$ is positive and thus, $\mathcal{S}$ is non-singular. □

Therefore, Eq. (10) uniquely determines P.

3.4. Estimation of P and r^2 from Data

If there is a sequence of n samples of data, $\{x_1, x_2, \dots, x_n\}$, an estimator for a_i, de“ned in Eq. (8) is

$$\hat{a}_i = \frac{1}{n} \sum_{k=1}^{n} \sinh l_i x_k. \tag{29}$$

Likewise, an estimator for b_i, de“ned in Eq. (9) is given by

$$\hat{b}_i = \frac{1}{n} \sum_{k=1}^{n} \cosh l_i x_k. \tag{30}$$

Using $\hat{a}_i(n)$ and $\hat{b}_i(n)$ as the estimates for a_i and b_i, respectively, given n samples of data, the $\hat{\Gamma}^{-1}$ matrix can be evaluated. $\hat{\Gamma}^{-1}(n)$ is de“ned as the estimate of Γ^{-1} given n samples of data. De“ne $\hat{P}(n)$ as the estimate of the vector P, given n samples of data. From Eq. (12) this gives,

$$\hat{P}(n) = \hat{\Gamma}^{-1}(n) \hat{A}(n), \tag{31}$$

our estimator to obtain the mixing proportions.

For the component variance,

$$r^2 = E[X^2] - [l_1^2, \dots, l_{M-1}^2] P, \tag{32}$$

so that

$$\hat{r}^2(n) = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - [l_1^2, \dots, l_{M-1}^2] \hat{P}(n). \tag{33}$$

4. ASYMPTOTIC ANALYSIS

4.1. Properties of the Non-linear Data Transformations

Consider the case of independent sequences of noise and symbols. Let $c = E[X^2]$ and

$$\hat{c}(n) = \frac{1}{n} \sum_{k=1}^{n} x_k^2. \tag{34}$$

Also, let

$$m_i = \sum_{k=1}^{M-1} P_k l_k^i. \tag{35}$$

Hence, the mean of X^2 is

$$E[X^2] = c \tag{36}$$

$$= r^2 + m_2, \tag{37}$$

and the variance of X^2 is

$$\mathrm{Var}[X^2] = E[X^4] - E^2[X^2] \tag{38}$$

$$= 3r^4 + 6r^2 m_2 + m_4 - (r^2 + m_2)^2 \tag{39}$$

$$= 2r^4 + 4r^2 m_2 + m_4 - m_2^2. \tag{40}$$

From the central limit theorem [22], $\hat{c}(n)$ is asymptotically normal with mean $r^2 + \nu_2$ and variance $\frac{1}{n}(2r^4 + 4r^2 m_2 + m_4 - m_2^2)$, written as $AN(r^2 + m_2, \frac{1}{n}(2r^4 + 4r^2 m_2 + m_4 - m_2^2))$.

Let

$$Y_n = (Y_1(n), \ldots, Y_{2M-1}) \tag{41}$$

$$= (\hat{a}_1(n), \ldots, \hat{a}_{M-1}(n), \hat{b}_1(n), \ldots, \hat{b}_{M-1}(n), \hat{c}(n)) \tag{42}$$

be the random vector consisting of the estimates of a_i, b_i and c, and de“ne the covariance matrix Σ as having elements $\Sigma_{i,j} = n\mathrm{Cov}(Y_i(n), Y_j(n))$. Let $y = \{y_1, \ldots, y_{2M-1}\}$ be the outcome of the random vector Y_n. The asymptotic variances of the estimators $\hat{P}$ and $\hat{r}^2$ are functions of Σ and their derivatives with respect to y [22]. The following develops the asymptotic covariance matrix T.

$$\mathrm{Cov}(\sinh l_i X, \sinh l_j X) = E[\sinh l_i X \sinh l_j X] - E[\sinh l_i X]E[\sinh l_j X] \tag{43}$$

$$= \frac{1}{2}E[\cosh(l_i - l_j)X] - \frac{1}{2}E[\cosh(l_i + l_j)X] - a_i a_j \tag{44}$$

$$= \frac{-1}{2}e^{\frac{r^2(l_i-l_j)^2}{2}}\left[1 + \sum_{k=1}^{M-1} P_k\left(\cosh(l_k(l_i - l_j)) - 1\right)\right] + \frac{1}{2}e^{\frac{r^2(l_i+l_j)^2}{2}}\left[1 + \sum_{k=1}^{M-1} P_k\left(\cosh(l_k(l_i + l_j)) - 1\right)\right] - a_i a_j. \tag{45}$$

Similarly,

$$\mathrm{Cov}(\sinh l_i X, \cosh l_j X) = \frac{1}{2}e^{\frac{r^2(l_i-l_j)^2}{2}}\sum_{k=1}^{M-1} P_k \sinh l_k(l_i - l_j) + \frac{1}{2}e^{\frac{r^2(l_i+l_j)^2}{2}}\sum_{k=1}^{M-1} P_k \sinh l_k(l_i + l_j) - a_i b_j \tag{46}$$

$$\mathrm{Cov}(\cosh l_i X, \cosh l_j X) = \frac{1}{2}e^{\frac{r^2(l_i-l_j)^2}{2}}\left[1 + \sum_{k=1}^{M-1} P_k\left(\cosh(l_k(l_i - l_j)) - 1\right)\right] + \frac{1}{2}e^{\frac{r^2(l_i+l_j)^2}{2}}\left[1 + \sum_{k=1}^{M-1} P_k\left(\cosh(l_k(l_i + l_j)) - 1\right)\right] - b_i b_j \tag{46}$$

for all $1 \le i, j \le M-1$. The above derivations use hyperbolic identities and the characteristic function of the Gaussian random variable. Cumbersome details have been reduced for conciseness.

In order to compute $\mathrm{Cov}(\sinh l_i X, X^2)$ and $\mathrm{Cov}(\cosh l_i X, X^2)$, expressions are required for $E[X^2 e^{l_k X}]$ and $E[X^2 e^{-l_k X}]$ where $1 \le k \le M-1$. Firstly, an expression for $E[Z^2 e^{aZ}]$ is derived where Z is a Gaussian random variable with mean l and variance q, and a is an arbitrary real constant. From probability theory,

$$E[Z^2 e^{aZ}] = \frac{1}{\sqrt{2pq^2}}\int_{-\infty}^{\infty} z^2 e^{az} e^{-\frac{(z-l)^2}{2q^2}} dz \tag{47}$$

$$= \frac{1}{\sqrt{2pq^2}}\int_{-\infty}^{\infty} z^2 e^{\frac{-[z^2 - 2(l + aq^2) + l^2]}{2q^2}} dz \tag{48}$$

$$= \frac{1}{\sqrt{2pq^2}}\int_{\infty}^{\infty} z^2 e^{\frac{-[z-(l+aq^2)]^2 + (a^2q^4 + 2alq^2)}{2q^2}} dz \tag{49}$$

$$= e^{al + \frac{a^2q^2}{2}} \frac{1}{\sqrt{2pq^2}}\int_{\infty}^{\infty} z^2 e^{\frac{-[z-(l+aq^2)]^2}{2q^2}} dz \tag{50}$$

$$= e^{al + \frac{a^2q^2}{2}}\left[q^2 + (l + aq^2)^2\right] \tag{51}$$

$$= e^{al + \frac{a^2q^2}{2}}\left[l^2 + (1 + 2al)q^2 + a^2q^4\right]. \tag{52}$$

Thus,

$$E[Z^2 e^{-aZ}] = e^{-al + \frac{a^2q^2}{2}}\left[l^2 + (1 - 2al)q^2 + a^2q^4\right] \tag{53}$$

and hence,

$$E[Z^2 \sinh aZ] = \frac{1}{2}(E[Z^2 e^{aZ}] - E[Z^2 e^{-aZ}]) \tag{54}$$

$$= \frac{1}{2} e^{\frac{a^2 q^2}{2}} \big(-e^{-al} \, (a^2 q^4 + l^2 + q^2(1 - 2al)) + e^{al} \, (a^2 q^4 + l^2 + q^2(1 + 2al)) \big), \quad (55)$$

and

$$E[Z^2 \cosh aZ] = \frac{1}{2} (E[Z^2 e^{aZ}] + E[Z^2 e^{-aZ}]) \quad (56)$$

$$= \frac{1}{2} e^{\frac{a^2 q^2}{2}} \big(e^{-al} \, (a^2 q^4 + l^2 + q^2(1 - 2al)) + e^{al} \, (a^2 q^4 + l^2 + q^2(1 + 2al)) \big), \quad (57)$$

Therefore, for the Gaussian mixture,

$$E[X^2 \sinh l_i X] = \frac{1}{2} e^{\frac{l_i^2 r^2}{2}} \sum_{k=1}^{M-1} P_k \big(-e^{-l_i l_k} (l_i^2 r^4 + l_k^2 + r^2(1 - 2l_i l_k)). + e^{l_i l_k} (l_i^2 r^4 + l_k^2 + r^2(1 + 2l_i l_k)) \big) \quad (58)$$

and,

$$E[X^2 \cosh l_i X] = \frac{1}{2} e^{\frac{l_i^2 r^2}{2}} \sum_{k=1}^{M-1} P_k \big(e^{-l_i l_k} (l_i^2 r^4 f + l_k^2 + r^2 (1 - 2l_i l_k)). + e^{l_i l_k} (l_i^2 r^4 + l_k^2 + r^2(1 + 2l_i l_k)) \big) \quad (59)$$

for all $1 \le i \le M-1$. The required covariances are,

$$\mathrm{Cov}(X^2, \sinh l_i X) = E[X^2 \sinh l_i X] - E[X^2] E[\sinh l_i X] \quad (60)$$

and

$$\mathrm{Cov}(X^2, \cosh l_i X) = E[X^2 \cosh l_i X] - E[X^2] E[\cosh l_i X], \quad (61)$$

for all $1 \le i \le M-1$. Let $\boldsymbol{\eta} = \{a_1, \ldots, a_{M-1}, b_1, \ldots, b_{M-1}, c\}$, and thus, Y_n is $AN(\boldsymbol{\eta}, \frac{1}{n}\Sigma)$, by the central limit theorem.

4.2. Convergence Properties of the Estimators

The function in Eq. (31) is continuous everywhere except for values of Y_n for which $\gamma(n)$ is singular. The function $\det(\gamma(n))$ is a “nite-order polynomial function of the elements of matrix $\gamma(n)$. Observe that $\gamma(n)$ is comprised of elements that are linear combinations of the outcomes of continuous random variables. As such, $\det(\gamma(n))$ is a continuous, “nite-order polynomial function of Y_n. The equation $\det(\gamma(n))$ has a “nite number of roots. Each root of $\det(\gamma(n))$ is an outcome of the vector random variable Y_n. Observe that Y_n is a continuous random vector with no non-zero probability value at any outcome. Thus, the set of singularities of $\gamma(n)$ is a “nite, nowhere dense set [23]. The probability measure of this set of singularities is clearly zero, since Y_n is a continuous random vector and does not take on non-zero probability value at a single point, and the set of singularities is formed by the “nite union of these points. Hence, $\gamma(n)$ is invertible with probability one (wp1), and Eq. (31) is continuous wp1. Y_n converges to $\boldsymbol{\eta}$ wp1, by the Theorem on page 24 of Ser”ing [22]. This gives rise to the following theorem.

Theorem 1: $\hat{P}(n)$ converges to P wp1 and $\hat{r}^2(n)$ converges to r^2 wp1.

Proof: The “rst part of the theorem follows from the above explanation. The second part is due to the fact that $\frac{1}{n}\sum_{i=1}^{n} x_i^2$ converges to c wp1 in Eq. (33), and utilizing Application D on page 26 of Ser”ing [22]. □

The theorems from Ser”ing [22] used above to show the convergence of $\hat{P}$ and $\hat{r}^2$ do not require the sequences of random variables to be independent. Therefore, $\hat{P}$ and $\hat{r}^2$ developed here converge even if the noise sequence is dependent and/or the sequence of data symbols form a regular Markov chain.

4.3. Asymptotic Variance of $\hat{P}$ for iid Sequences

Theorem 2: $\hat{P}(n)$ is $AN(P, \frac{1}{n} D_P \Sigma D_P^T)$ for independent sequences of data symbols and noise.

Proof: Di erentiating Eq. (31) with respect to y with matrix di erentiation de“ned as in Judge et al. [24] yields,

$$D_P = \left[\frac{\partial \hat{P}}{\partial y_j(n)} \Big|_{y=\boldsymbol{\eta}} \right]_{(M-1)\times(2M-1)} \quad (62)$$

$$= \left[\Gamma^{-1} \left(\frac{\partial A}{\partial y_j(n)} - \frac{\partial \Gamma}{\partial y_j(n)} \hat{P} \right) \Big|_{y=\boldsymbol{\eta}} \right]_{(M-1)\times(2M-1)}. \quad (63)$$

For $y = \boldsymbol{\eta}$, $\hat{P}$ on the right-hand side of Eq. (63), evaluates to P, which is a constant vector independent of y. Therefore,

$$D_P = \left[\Gamma^{-1}\left(\frac{\partial A}{\partial y_j(n)} - \frac{\partial \Gamma \hat{P}}{\partial y_j(n)}\right)\right]_{(M-1)\times(2M-1)}. \quad (64)$$

Observe that

$$\left[\frac{\partial A}{\partial y_j(n)}\right]_{1\times(2M-1)} = \left[\frac{\partial A}{\partial y_1(n)}, \ldots, \frac{\partial A}{\partial y_{2M-1}(n)}\right] \quad (65)$$

$$= \begin{pmatrix} 1 & 0 & & 0 & 0 & & 0 \\ 0 & 1 & & 0 & 0 & & 0 \\ & & \ddots & & & \ddots & \\ 0 & 0 & & 1 & 0 & & 0 \end{pmatrix}_{(M-1)\times(2M-1)}. \quad (66)$$

In the above matrix, the entries are all 1•s on the main diagonal and 0•s elsewhere. Also,

So,

$$\Gamma P = \begin{pmatrix} \sum_{k=1}^{M-1} P_k[a_1(1-\cosh l_1 l_k) + b_1 \sinh l_1 l_k] \\ \vdots \\ \sum_{k=1}^{M-1} P_k[a_{M-1}(1-\cosh l_{M-1} l_k) + b_{M-1}\sinh l_{M-1} l_k] \end{pmatrix}. \quad (67)$$

$$\frac{\partial(\Gamma P)}{\partial y} = \begin{pmatrix} \ddots & & & \ddots & & \vdots \\ & \sum_{k=1}^{M-1} P_k(1-\cosh l_i l_k) & & & \sum_{k=1}^{M-1} P_k \sinh l_i l_k & 0 \\ & & \ddots & & \ddots & \vdots \end{pmatrix} \quad (68)$$

is a $(M-1)\times(2M-1)$ matrix. The elements of matrix $\frac{\partial(\Gamma P)}{\partial y}$ in row i and column j

$$= \begin{cases} \sum_{k=1}^{M-1} P_k(1-\cosh l_i l_k), & i = j \\ \sum_{k=1}^{M-1} P_k \sinh l_i l_k, & i = j - M + 1 \\ 0, & \text{otherwise.} \end{cases} \quad (69)$$

Substituting Eqs. (66) and (68) into Eq. (64) and

$$D_P = \Gamma^{-1}\begin{pmatrix} & \ddots & & & \ddots & & \vdots \\ 1 - \sum_{k=1}^{M-1} P_k(1-\cosh l_i l_k) & & & -\sum_{k=1}^{M-1} P_k \sinh l_i l_k & & & 0 \\ & & \ddots & & & \ddots & \vdots \end{pmatrix}. \quad (70)$$

simplifying yields

In summary, D_P is an $(M-1)\times(2M-1)$ matrix that is the product of Γ^{-1} and an $(M-1)\times(2M-1)$ matrix with elements in row i and column j

$$= \begin{cases} 1 - \sum_{k=1}^{M-1} P_k(1-\cosh l_i l_k), & i = j \\ -\sum_{k=1}^{M-1} P_k \sinh l_i l_k, & i = j - M + 1 \\ 0, & \text{otherwise.} \end{cases} \quad (71)$$

Each row of D_P is non-zero for $y = \boldsymbol{\eta}$, therefore, by Theorem A in Section 3.3 of Ser"ing [22], $\hat{P}$ is $AN(P, \frac{1}{n} D_P \Sigma D_P^T)$. D_P is an $(M-1)\times(2M-1)$ matrix and Σ is a $(2M-1)\times(2M-1)$ matrix, thus, the product $D\Sigma D^T$ is a $(M-1)\times(M-1)$ matrix, which is dimensionally consistent. □

4.4. Asymptotic Variance of $\hat{\boldsymbol{\sigma}}^2$ for iid Sequences

Theorem 3: $\hat{r}^2(n)$ is $AN(r^2, \frac{1}{n} D_{r^2} \Sigma D_{r^2}^T)$ for independent sequences of data symbols and noise.

Proof: The matrix of the derivatives of r^2 with respect to y is given as

$$D_{r^2} = \left[\frac{\partial E[X^2]}{\partial y} - [l_1^2, \ldots, l_{M-1}^2] D_P \big|_{y=\boldsymbol{\eta}}\right]_{1\times(2M-1)} \quad (72)$$

$$= [0, \ldots, 0, 1]_{1\times(2M-1)} - [l_1^2, \ldots, l_{M-1}^2] D_P. \quad (73)$$

As in the previous subsection, D_{r^2} is non-zero at $y = \boldsymbol{\eta}$, and by Theorem A in Section 3.3 of Ser"ing [22], $\hat{r}^2$ is $AN(r^2, \frac{1}{n} D_{r^2} \Sigma D_{r^2}^T)$. □

As in any (non-degenerate) estimation problem, the asymptotic variances of the estimates of unknown parameters are functions of the unknown parameters themselves. In practice, estimates of asymptotic variances are obtained by substituting the estimates of unknown parameters in places of their unknown values.

5. SIMULATION EXPERIMENTS AND DISCUSSION

A simulation experiment is designed to mimic the OFDM example mentioned in Section 1. The real or imaginary part of the complex wave is a PAM signal with four equally-spaced levels. Since the estimator is invariant to level translations, the four component mean values of the mixture are chosen to be 1, 2, 3 and 0. The OFDM noise tolerance standard is speci"ed as a maximum packet error rate of 10% for a 1000 byte packet. This corresponds to a symbol error probability of 5×10^{-5} in the four-level PAM scheme with iid symbols of equal probabilities. Following [18], for a PAM system with four equi-probable symbols whose successive mean values are separated by 1 and for a probability of symbol error of 0.5×10^{-4}, the variance of the noise random variable evaluates to approximately 0.0156. This value is used for the common component variance in our experiment. The symbol probabilities are chosen to be all di erent to illustrate the estimation process. The sequence of samples is iid.

In order to gauge the effectiveness of the estimators, the time-varying channel model in adaptive wireless transmission protocol of Gass et al. [4] is used. The channel model in [4] is a two state discrete parameter Markov chain (between successive transmissions) with a probability of 0.05 for a change in the states of the channel. This translates into an average of twenty successive data frames. In our OFDM transmission scheme, 4096 bytes is the maximum frame size. If a frame size of 1024 bytes is used, 20 frames correspond to $2048 \times 20 = 40{,}960$ real number samples. The conclusion is that the channel is stationary for an average of a few tens of thousands of such real number samples. Estimation experiments are conducted for a maximum of 5000 samples in this paper. Figures 1…4 show plots of estimates of P_1, P_2, P_3 and P_4, respectively, versus number of data samples for n up to 1000. Figure 5 is a plot of an estimate of $\hat{r}(n)^2$ for n up to 1000. For the particular example experiment shown, the values of the symbol probabilities are chosen as $\hat{P} = \{0.32, 0.27, 0.16, 0.25\}$. The values for the estimates were obtained by running the simulation for 1000 times and averaging the values produced. The standard deviation envelopes over the 1000 trials are plotted on the upper and lower sides of the plot of the averages. By comparison, the EM algorithm converges after approximately 20 iterations, given the same problem and 2500 samples of data. The EM algorithm is not

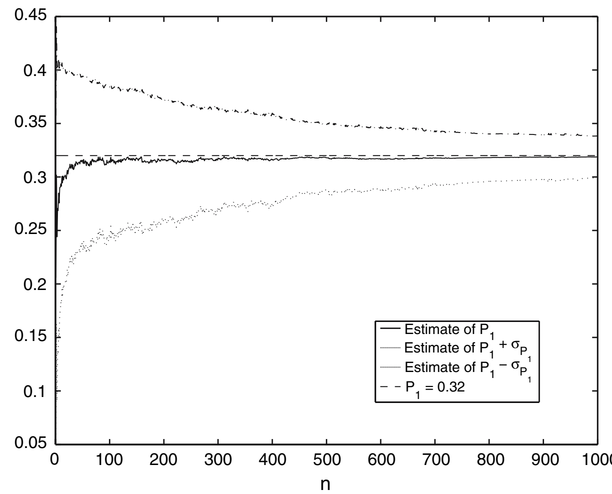

Fig. 1. Plot of $\hat{P}_1(n)$ versus n.

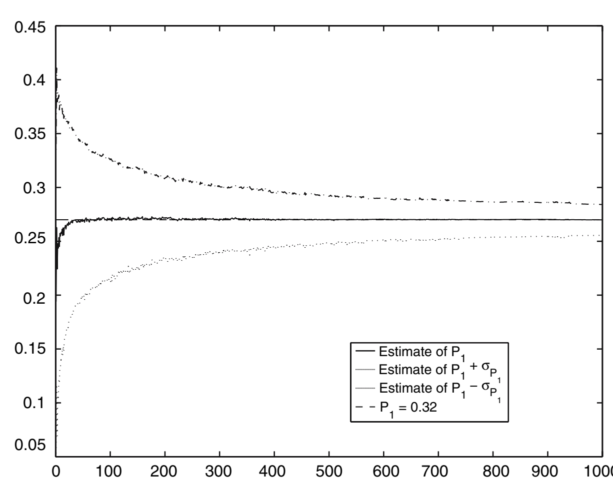

Fig. 2. Plot of $\hat{P}_2(n)$ versus n.

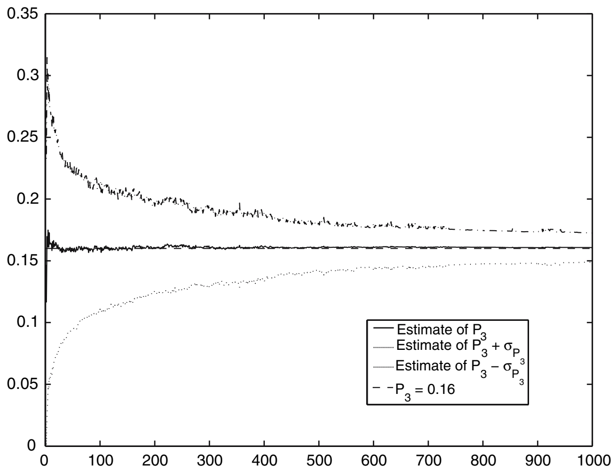

Fig. 3. Plot of $\hat{P}_3(n)$ versus n.

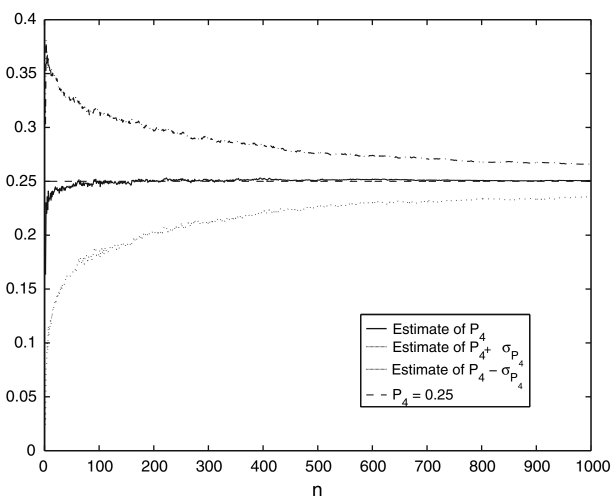

Fig. 4. Plot of $\hat{P}_4(n)$ versus n.

practical for large data sets and for data sets with varying sizes.

Our estimators are recursive so that cumulative results of intermediate steps in the estimation algorithm can be stored in accumulators and updated estimates evaluated as and when required. Over larger time intervals during which the parameters may be varying, a common approach used in practice is to use exponential forgetting, as follows. Let y_i be estimates of a parameter over successive non-overlapping time windows, $i = 1, \ldots$.

$$z_i = wy_i + (1 - w)z_{i-1} \tag{74}$$

where $0 < w < 1$ is the weight assigned to the current window. The estimate z_i uses all the windows

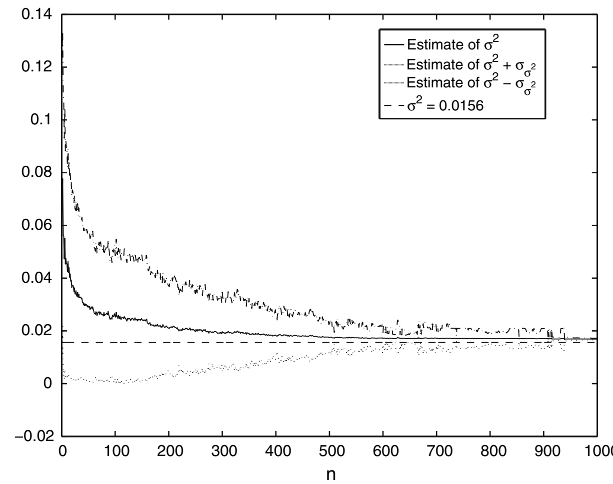

Fig. 5. Plot of $\hat{r}(n)$ and versus n.

up to the present but with exponentially decreasing weights for earlier windows.

The asymptotic variances of the estimates of the “rst three mixing proportions for this mixture example analytically evaluate to $\frac{1}{n}[0.5048, 0.2330, 0.1392]$. Similarly, the asymptotic variance of the estimate of the common component variance evaluates to $\frac{1}{n}0.0142$. As mentioned earlier, simulation to obtain estimates of such variances is conducted with 1000 experiments, each experiment using 5000 mixture samples. For 1000 experiments, the estimates of the variances of the “rst three mixing proportions are found to be $\frac{1}{5000}[0.3777, 0.1968, 0.1483]$. For r^2, the corresponding estimate of its variance is found to be $\frac{1}{5000}0.3791$. These “gures did not change signi“cantly when the same experiment, each with 5000 data samples, was repeated 2000 times instead of 1000 times. For 5000 data samples and 2000 experiments, the corresponding values are $\frac{1}{5000}[0.3648, 0.1982, 0.1413]$ and $\frac{1}{5000}0.3319$.

Analytical expressions for the elements of the Hessian of the log-likelihood function (van Trees [25]) for the estimation problem (with $M = 4$) are obtained with the help of Mathematica software. Their expectations are numerically evaluated for the above simulation problem and this gives the Fisher Information matrix for the numerical problem at hand. The Cramer…Rao minimum variance bounds (MVBs) are evaluated by inverting the Fisher Information matrix and picking its diagonal elements. They are $\frac{1}{n}0.3200$, $\frac{1}{n}0.5401$, $\frac{1}{n}0.1100$, and $\frac{1}{n}0.0007$, for P_1, P_2, P_3, and r^2, respectively.

Ef“cient estimators that achieve MVBs are not realizable for many practical estimation problems, for example, for the estimation of the prior probabilities of a “nite mixture only, given complete information about the component densities, as shown in Dattatreya and Kanal [26]. In view of this, the variances of the estimators developed here, given in the above paragraph, are assessed to be good.

5.1. Comparisons with other Estimators

The estimator in [8] is implemented for the above problem of estimating the noise variance, in a companion simulation experiment. As in the earlier case, 1000 experiments are conducted. The average of the 1000 experiments and its one standard deviation envelope above and below are drawn as a function of the number of data samples upto 1000, in Figure 6. The estimates for several initial number of samples ”uctuate very heavily. They are not plotted upto 10 samples, to ensure that the scale of the plots allow them to be illustrative. The deviation of these estimates are higher than the corresponding ones in Figure 5, for larger numbers of samples (around 1000).

It should be emphasized that the distinct superiority of the approach in the present paper is the convergence without any control parameter. The estimator in [8] uses a control parameter and convergence is not guaranteed for arbitrary choices of the control parameter; neither is there a good approach to “nd one, if it exists.

A comparison of our present estimator with a simple decision directed scheme is also illustrative. In a simple decision directed parameter estimation algorithm, decisions on received symbols are made using current estimates of parameters in the maximum

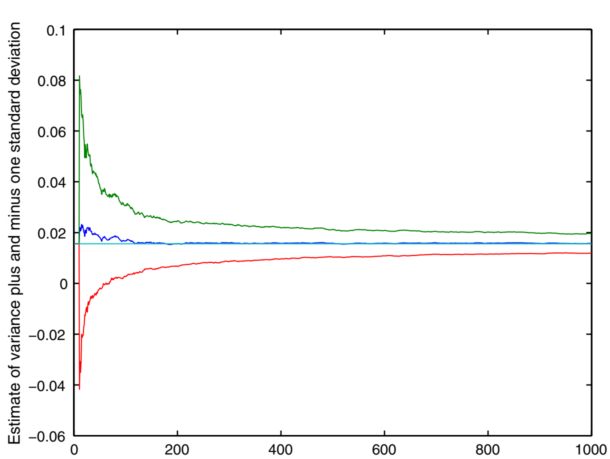

Fig. 6. Plot of $\hat{r}(n)$ and versus n using the estimator in [8].

a posteriori probability scheme. These decisions are then considered to be true symbols and the parameter estimates are updated. Let the decision making algorithm use an optimal scheme. Such a scheme is “ctitious, since the parameters of the optimal decision making scheme are unknown. But this helps us to evaluate the best possible properties of the (“ctitious) decision directed estimator. The expected value of the estimate of noise variance by such a scheme is given by

$$E[\hat{r}^2_{dd}] = \sum_{i=1}^{M}\left(\int_{x\in D(x_i)} \sum_{j=1}^{M} P_j p(x|x_j)(x - l_j)^2 dx\right) \tag{75}$$

where ω_j denotes the jth of M symbols and $D(\omega_j)$ denotes the region of the observation space for which the optimal decision is ω_j. The probability density function of the received random variable, condition on the transmitted symbol ω_j is $p(x|\ \omega_j)$. The true value of the noise variance is given by

$$E[\hat{r}^2] = \sum_{i=1}^{M} P_i \int_{x=-\infty}^{\infty} \sum_{j=1}^{M} p(x|x_j)(x - l_i)^2 dx. \tag{76}$$

Clearly, the decision directed estimator of the noise variance does not converge to its true value. The original analysis of a simple two symbol decision directed estimation scheme in Davisson and Schwartz [27] shows that there is also a small but non-zero probability of a runaway in which one of the symbol probabilities converges to 1.

6. CONCLUDING REMARKS

The main contribution in this paper is the development of convergent and computationally-ef“cient estimators for the mixing proportions and the common component variance of a “nite Gaussian mixture with known component means. This case of known component means occurs in QAM digital wireless communication with automatic gain control. Motivation for the study is comprehensively developed with an IEEE 802.11a OFDM standard example. The coef“cients for the unknowns and the values on the right-hand side of the linear equations for P_j are expectations of non-linear data transformations $\sinh l_i x$ and $\cosh l_i x$ for $i = 1, \ldots, M$, and X^2. These are well-behaved functions and their expectations are estimated by the corresponding sample averages. As a consequence, the vector estimator possesses a su cient statistic vector. Therefore, the estimator can be implemented recursively or in a batch processing algorithm. The sample averages of non-linear data transformations are convergent and asymptotically normal for the case of iid data sequences. Their “nite sample variances are also easily evaluated. However, obviously, the solution to the linear equations to evaluate the vector estimate involves multiplication of the inverse of a matrix and a column vector, both composed of the sample averages of data transformations. Therefore, evaluation of the “nite sample variances of the estimates is very di cult. Their asymptotic variances are evaluated with considerable e ort. A practical OFDM numerical example is simulated. The simulation plots the averages of the estimates based on several independent trials. The standard deviations above and below the average plots are also included. The Cramer…Rao minimum variance bounds are evaluated for the simulation problems for comparison. The estimator developed here is a superior alternative to the one developed by one of the present authors with another co-author [8] in the following ways. In [8], the estimator requires a control parameter that should be determined by examining a plot of a contour over a rectangular area. Such plotting is easily accomplished with the help of a computer. The authors have not been able to construct an example problem for which no control parameter exists in [8]. However, there is no proof there that every problem case possesses a control parameter value for convergent estimation of the unknowns.

The present estimator is superior to the EM algorithm in the following ways. The EM algorithm is a batch processing algorithm only and does not possess suf“cient statistics. All the data samples are required to be retained for every iteration of the EM algorithm. The probability density values of every data sample is evaluated for the current estimates of the unknown parameters, in every iteration of the EM algorithm. Its implementation requires supplying initial values for all the unknown parameters and only a local convergence is guaranteed, in general. As a consequence, variances or asymptotic variances of the estimates cannot be obtained for the EM algorithm estimator. Of course, the general EM algorithm does not have the advantage of knowledge of component mean values that the estimators in the present paper enjoy.

REFERENCES

1. R. W. Brodersen, A. Wolisz, D. Cabric, S. M. Mishra, and D. Willkomm, Corvus: A cognitive radio approach for usage of virtual unlicensed spectrum. http://www.bwrc.eecs.berkeley.edu/Research/MCMA/CR_White_paper_“nal1.pdf, White Paper, July 2004.
2. J. Mitola III, Software Radio … Cognitive Radio. http://www.ourworld.compuserve.com/homepages/jmitola.
3. M. B. Pursley, and C. S. Wilkins, Adaptive transmission for direct-sequence spread-spectrum communications over multipath channels, International Journal of Wireless Information Networks, Vol. 7, No. 2, pp. 69…77, 2000.
4. J. H. Gass Jr., M. B. Pursley, H. B. Russell, and J. S. Wysocarski, An adaptive-transmission protocol for frequency-hop wireless communication Networks, Wireless Networks, Vol. 7, pp. 487…495, 2001.
5. G. McLachlan, and T. Krishnan, The EM Algorithm and Extensions, Wiley Interscience, NY, USA, 1997.
6. B. W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall, FL, USA, 1986.
7. B. S. Everitt, and D. J. Hand, Finite Mixture Distributions, Chapman & Hall, FL, USA, 1981.
8. G. R. Dattatreya, and X. Fang, Parameter estimation: known vector signals in unknown Gaussian noise, Pattern Recognition, Vol. 36, pp. 2317…2332, 2003.
9. V. Krishnamurthy, and B. Moore, On-line estimation of hidden Markov model parameters based on Kullback…Leibler information measure, IEEE Transactions on Signal Processing, Vol. 41, pp. 557…573, 1993.
10. J. K. Tugnait, L. Tong, and Z. Ding, Single-user channel estimation and equalization, IEEE Signal Processing Magazine, Vol. 17, pp. 17…28, 2000.
11. X. Wang, and R. Chen, Blind turbo equalization in Gaussian and impulsive noise, IEEE Transactions on Vehicular Technology, Vol. 50, pp. 1092…1105, 2001.
12. A. Monin, and G. Salut, Minimum variance estimation of parameters constrained by bounds, IEEE Transactions on Signal Processing, Vol. 49, pp. 246…248, 2001.
13. D. S. Pham, and A. M. Zoubir, A sequential algorithm for robust parameter estimation, IEEE Signal Processing Letters, Vol. 12, pp. 21…24, 2005.
14. M. Farquharson, P. O•Shea, and G. Ledwich, A computationally e cient technique for estimating the parameters of polynomial-phase signals from noisy observation, IEEE Transactions on Signal Processing, Vol. 53, pp. 3337…3342, 2005.
15. X. Deng, A. M. Haimovich, and J. Garcia-Frias, Decision Directed iterative channel estimation for MIMO systems. IEEE International Conference on Communications (ICC •03), pp. 2326…2329, 2003.
16. S. Jeong and R. M. Gray. A comparison of EM and GMVQ in estimating Gauss mixtures: application to probabilistic image retrieval. Proceedings of ICASSP 2005, Vol. 5, v/361…v/364, 2005.
17. S. A. Vorobyov, Y. C. Eldar, A. Nemirovski, and A. B. Gershman, Probability-constrained approach to estimation of random Gaussian parameters 1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, pp. 101…104, 2005.
18. J. G. Proakis, Digital Communications, McGraw Hill, NY, USA, 1995.
19. A. S. Tanenbaum, Computer Networks, Prentice-Hall, NJ, USA, 2003.

20. A. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw Hill, NY, USA, 1984.
21. P. Lancaster, and M. Tismenetsky, The Theory of Matrices, 2th ed. Academic Press Inc., CA, USA, 1985.
22. R. J. Ser"ing, Approximation Theorems of Mathematical Statistics, John Wiley & Sons, NY, USA, 1980.
23. G. B. Folland, Real Analysis: Modern Techniques and Their Applications, 2th ed. John Wiley & Sons, NY, USA, 1999.
24. G. G. Judge, R. C. Hill, W. E. Gri ths, H. Lu ¨tkepohl, and T. Lee, Introduction to the Theory and Practice of Econometrics, John Wiley & Sons, NY, USA, 1982.
25. H. L. Trees, Detection, Estimation and Modulation Theory, Part 1, John Wiley & Sons, NY, USA, 1968.
26. G. R. Dattatreya, and L. N. Kanal, Asymptotically e cient estimation of prior probabilities in multiclass "nite mixtures, IEEE Transactions on Information Theory, Vol. 37, No. 3, pp. 482…489, 1990.
27. L. D. Davisson, and S. C. Schwartz, Analysis of a decision-directed receiver with unknown priors, IEEE Transactions on Information Theory, Vol. 16, pp. 270…276, 1970.

G. R. Dattatreya received the B. Tech. degree in Electrical Engineering from the Indian Institute of Technology, Madras, M.E. in Electrical Communication Engineering, and Ph.D. from the Department of Computer Science and Automation, Indian Institute of Science, Bangalore, India. He is currently an Associate Professor in the Department of Computer Science, University of Texas at Dallas. He has served as a visiting faculty member in (1) University of Maryland, College Park, (2) Malaysia Polytechnic Project, Batu Pahat, Johor, Malaysia, and (3) ITESM, Monterry, Mexico. Before joining academia, he was a Senior Scientist at the Scienti"c Analysis Group, Delhi, India. His current research interests are Stochastic Modeling, Parameter Estimation, and Adaptive Optimization in Communication, Signal Processing, and Computer Network Systems.

Larry N. Singh graduated from the University of Toronto with an Hons. B.Sc. degree in Computer Science in 1997. He completed his M.S. and Ph. D. degrees in Computer Science at The University of Texas at Dallas (UTD) in 1999 and 2004. His professional experience includes being a consultant for such companies as Europay International in Belgium, American Express International and the Royal Bank of Canada. He has been recognized for his extensive community service work; he is the recipient of the 2004 UTD Alumni Ring Award, which is the highest award given to UTD students for academic achievement and school and community involvement. His research interests include parameter estimation, stochastic modeling, queuing theory, numerical optimization and simulation, and bioinformatics.