Blind Sequence Estimation

Lang Tong, *Member, IEEE*

***Abstract*—Estimating the data sequence from the received signal without knowing the transmission channel is referred to as blind sequence estimation. A new blind sequence estimation scheme is proposed by exploiting the second order statistical properties of the source and the algebraic structures of the data sequence. An optimal source (deterministic) correlation estimator and the Viterbi algorithm are used to achieve blind sequence estimation. The proposed approach also has a special structure particularly attractive for time varying channels.**

***Index Terms*—Channel equalization, multipath channels, viterbi algorithm, sequence estimation.**

I. Introduction

BLIND identification and equalization have recently attracted increasing research attention. One of the advantages of blind equalization methods is that the training period of the transmission over an intersymbol interference (ISI) channel may be shortened or eliminated. Therefore, there is a potential increase in the transmission efficiency, particularly for those rapidly varying channels whose training signals must be transmitted periodically for conventional equalization methods. In mobile communication, for example, the GSM standard currently adopted in Europe requires a 28 bits training sequence for every 116 information bits [7]. Certain military applications have even higher overhead associated with the transmission of channel probes.

Most current blind equalization schemes are not adequate for mobile communication channels that require channel identification and equalization within about 100 symbols. Since Sato [16] presented the idea of self-recovering (blind) equalization, various algorithms (e.g., [5], [6], [13], [16], [18], [23]), have been proposed based on optimizing certain criteria involving higher order statistics (HOS). Blind channel identification algorithms using HOS were also explored by Hatzinakos and Nikias [8], and by Porat and Friedlander [14]. Simulations have shown that the convergence time of these HOS algorithms (in thousands of symbols) is much too long for mobile communication channels. Several blind channel identification methods based on second-order cyclostationary properties have recently been proposed [11], [15], [19], [21], [22]. The closed-form identification algorithms proposed in [19], [20], [21], [22], for example, have a convergence rate of about 100 symbols in simulation studies. It has been shown, however, that the channel is not identifiable when the channel has certain special zeros [19], although perhaps such cases are rare. Nonetheless, this limitation is theoretically fundamental.

B. Proposed New Approach and Related Work

Existing approaches to blind equalization require the identification of the channel or its inverse. In this paper, we propose a new approach focusing on blind sequence estimation rather than channel identification. In fact, the proposed approach does not require a channel identification. The key idea is to estimate the source (deterministic) correlation from the observation (without knowing the channel) and apply the Viterbi algorithm to reconstruct the input symbols. As we shall show in this paper, the estimation of signal correlation function requires only the orthogonalization of the channel, which is much simpler than the estimation of the channel. The proposed approach also applies to both single and multiple receiver structures. This is valuable because spatial diversities offered by multiple receivers are important for fading channels. Balaban and Salz presented a rather extensive theoretical and simulation study of an optimal diversity combining an equalization scheme with conventional adaptive equalizations (with training sequence or under decision directed mode) [2]. Our approach naturally fits into the platform of multiple receivers.

Blind sequence estimation without channel identification distinguishes the proposed approach from several existing sequence estimation methods. The Viterbi algorithm was applied to combat ISI first by Forney [3], [4], where the channel was assumed to be known. The blind techniques involving the Viterbi algorithm were proposed by Magee and Proakis [9] and Ungerboeck [24] where the channel is estimated in a decision-directed mode. These approaches, however, may suffer error propagation. More recently, Seshadri [17] and Zervas *et al.* [26] proposed the techniques of joint data and channel estimation by quantizing the channel. When converged properly, these techniques offer the optimal solution at the expense of computation complexity.

II. Multipath Fading Channels and Assumptions

In this section, a vector representation of a multipath fading channel is presented. A time varying channel between the transmitter and the ith receiver, as illustrated in Fig. 1, is a

Paper approved by A. Duel-Hallen, the Editor for Communications of the IEEE Communications Society. Manuscript received June 7, 1993; revised March 14, 1994, June 21, 1994, and October 10, 1994. This work was partially supported by the National Science Foundation under Grant NCR-9321813, the Advanced Research Projects Agency monitored by the Federal Bureau of Investigation under Contract J-FBI-94-221 and the Naval Research Laboratory under Contract N00014-94-1-G003. This paper was presented in part at the 26th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, Nov. 1992.

L. Tong is with the Department of Electrical and Systems Engineering, University of Connecticut, U-157, Storrs, CT 06269-3157 USA (email: ltong@eng2.uconn.edu).

IEEE Log Number 9414787 USA.

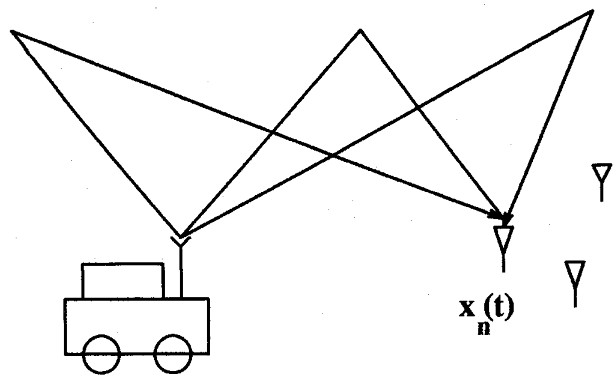

Fig. 1. A multipath environment.

multipath Rayleigh fading channel with an envelope impulse response

$$c_i(t) = \sum_{k=1}^{K} \alpha_{ik}(t)\delta(t - \tau_{ik}) \quad (1)$$

where K is the number of multipaths ranging from 2 to 70 or more, $\alpha_{ik}(t)$ are zero-mean Gaussian processes, and τ_{ik} are the corresponding path delays. For slowly varying channels, $\alpha_{ik}(t) \approx \alpha_{ik}$ within a duration of about 100 symbols, see [2]. With the QAM signaling, the received baseband signal at the ith receiver has the following form

$$x_i(t) = \sum_n s_n p(t - nT_s) * c_i(t) + n_i(t) \quad (2)$$

$$= \sum_n s_n h_i(t - nT_s) + n_i(t) \quad (3)$$

where s_n is the symbol sequence, T_s is the symbol interval, $p(t)$ is the impulse response of the pulse shaping filter, and $h_i(t)$ is the composite channel that includes the channel, the shaping filter, and perhaps the receiver filter, and finally, $n_i(t)$ is the additive noise. When $x_i(t)$ is sampled at $t = l\Delta$ with some unknown timing t_0, we have a discrete-time model

$$x_i(l) = \sum_n s_n h_i(l - nT) + n_i(l) \quad (4)$$

where, under a mild abuse of notation, $x_i(l) = x_i(l\Delta + t_0), h_i(l) = h_i(l\Delta + t_0)$, and $n_i(l) = n_i(l\Delta + t_0)$. $T = \frac{T_s}{\Delta}$ is assumed to be an integer.

With the assumption that $h_i(l)$ lasts d symbol intervals, i.e., $h_i(l) = 0$, for $l < 0$, and $l > dT$, a vector representation of the received signal at the ith receiver is obtained by letting

$$\mathbf{x}_i(t) = [x_i(tT), \cdots, x_i((t+1)T - 1)]^t \quad (5)$$

$$\mathbf{n}_i(t) = [n_i(tT), \cdots, n_i((t+1)T - 1)]^t \quad (6)$$

$$\mathbf{h}_i(t) = [h_i(tT), \cdots, h_i((t+1)T - 1)]^t \quad (7)$$

$$\mathbf{H}_i = [\mathbf{h}_i(0), \cdots, \mathbf{h}_i(d-1)] \quad (8)$$

$$\mathbf{s}(t) = [s_t, s_{t-1}, \cdots, s_{t-d+1}]^t. \quad (9)$$

We then have

$$\mathbf{x}_i(t) = \mathbf{H}_i\mathbf{s}(t) + \mathbf{n}(t). \quad (10)$$

Putting all the data from M receivers in a single matrix form, let

$$\mathbf{x}(t) = \begin{pmatrix} \mathbf{x}_1(t) \\ \vdots \\ \mathbf{x}_M(t) \end{pmatrix} \in \mathcal{C}^{N\times 1}, \mathbf{n}(t) = \begin{pmatrix} \mathbf{n}_1(t) \\ \vdots \\ \mathbf{n}_M(t) \end{pmatrix} \in \mathcal{C}^{N\times 1} \quad (11)$$

$$\mathbf{H}(t) = \begin{pmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_M \end{pmatrix} \in \mathcal{C}^{N\times d}. \quad (12)$$

The vector representation is given by

$$\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t) + \mathbf{n}(t). \quad (13)$$

The problem of blind sequence estimation is to estimate the information symbols s_n *without* the knowledge of the channel parameter matrix $\mathbf{H}$.

Remark:

1) The above model captures a wide range of fading channels, particularly those involving a large number of fading paths.
2) At each receiver, one can form $\mathbf{x}_i(t)$ using more data than that from a single symbol interval.

For example, data from K symbol intervals can be used to form $\mathbf{x}_i(t)$ by

$$\mathbf{x}_i(t) = [x_i(tT), \cdots, x_i((t+1)T - 1)| \cdots |x_i((t-K)T), \cdots, x_i((t-K-1)T - 1)]^t. \quad (14)$$

The corresponding vector representation has the same form, and the approach developed in this paper applies also to the case of single receiver.

A. Assumptions

We assume the following assumptions in the sequel:

1) The information symbol sequence is zero mean, and $E(s_i s_j^*) = \delta(i - j)$, where $\delta(\cdot)$ is the discrete-time unit pulse.
2) The noise process $n_j(\cdot)$ is zero mean for all j, and $E(n_i(t_1)n_j^*(t_2)) = \sigma^2\delta(i-j)\delta(t_1 - t_2)$.
3) The noise process $n_j(\cdot)$ is uncorrelated with the source symbol sequence $\{s_k\}$.
4) The channel parameter matrix $\mathbf{H}$ is an $N \times d$ matrix with full column rank.

The assumptions A.1–A.3 are fairly common. Assumption A.4 deserves closer examination. First, A.4 implies that all channels have finite impulse responses, which is reasonable for most wireless communication cases. It has been shown that the full rank condition is necessary and sufficient for the channel to be identifiable using only the second order statistics of the observation [19], [21]. In fact, A.4 is almost always satisfied except perhaps in some pathological cases[1]. In practice, however, the channel matrix $\mathbf{H}$ is most likely to be close to singular due to small tails of the channel impulse

[1] The necessary and sufficient condition for $\mathbf{H}$ be rank deficient is that all M channels share the same uniformly $\frac{2\pi}{T}$-spaced zeros [19], [21].

responses. Although our derivation and analyses are based on the full rank condition, the algorithm shows a surprising robustness when $\mathbf{H}$ is close to singular. This is due perhaps to the application of the Viterbi algorithm where the algebraic structure of the source is exploited.

III. Blind Sequence Estimation

In this section, we present a new blind sequence detection scheme that does not require a complete channel identification. We first motivate our approach by showing that the source (deterministic) correlation can be obtained from the (deterministic) correlation of the received signal. It is then pointed out that the Viterbi algorithm can be applied to obtain the source symbols. Next, we develop an optimum estimator for the source correlation.

A. Mahalanobis Orthogonalization

The so-called Mahalanobis orthogonalization plays a key role in our approach. The essential idea is to orthogonalize the channels using only the observation data. Consider the vector representation (13). Denoting

$$\mathbf{R}_x(k) = E(\mathbf{x}(t)\mathbf{x}^*(t-k)) \tag{15}$$

we have, from (13) and assumptions (A.1–A.4)

$$\mathbf{R}_x(0) = \mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I}. \tag{16}$$

It is well known that the singular value decomposition (SVD) of $\mathbf{R}_x(0)$ must then take the following form

$$\mathbf{R}_x(0) = \mathbf{U}\mathrm{diag}(\pi_1^2, \cdots, \pi_N^2)\mathbf{U}^* \tag{17}$$

$$= \mathbf{U_s}\mathrm{diag}(\lambda_1^2, \cdots, \lambda_d^2)\mathbf{U_s}^* + \sigma^2\mathbf{I} \tag{18}$$

where $\mathbf{U}_s$ is the submatrix consisting of the first d columns of $\mathbf{U}$ and $\pi_i^2 = \lambda_i^2 + \sigma^2$. The Mahalanobis orthogonalization transform is defined by

$$\mathbf{y}(t) = \mathbf{T}_m\mathbf{x}(t) \tag{19}$$

where $\mathbf{T}_m$ is obtained from the SVD of $\mathbf{R}_x(0)$ by

$$\mathbf{T}_m = \mathbf{\Lambda}_s^{-1}\mathbf{U}_s^* \tag{20}$$

$$\mathbf{\Lambda_s} = \mathrm{diag}(\lambda_1, \cdots, \lambda_d). \tag{21}$$

Note that, given $\mathbf{R}_x(0)$, the noise variance σ^2 and the signal dimension d can both be determined from the singular values of the $N \times N$ matrix $\mathbf{R}_x(0)$ when $d < N$ because the last $N - d$ singular values will all be σ^2.

The Mahalanobis transform matrix $\mathbf{T}_m$ has the following simple but important property.

Property 1: The Mahalanobis transform matrix $\mathbf{T}_m$ orthogonalizes the channel parameter matrix $\mathbf{H}$, i.e., there is an orthogonal matrix $\mathbf{V}$ such that

$$\mathbf{T}_m\mathbf{H} = \mathbf{V}. \tag{22}$$

Proof: Comparing (16) and (18), we have

$$\mathbf{H}\mathbf{H}^* = \mathbf{U}_s\mathbf{\Lambda}_s^2\mathbf{U}_s^*. \tag{23}$$

Therefore

$$\mathbf{T}_m\mathbf{H}(\mathbf{T}_m\mathbf{H})^* = \mathbf{I} \tag{24}$$

which leads to (22). □□□

The important consequence of this property is when there is no noise, the (deterministic) correlation of the orthogonalized observation

$$\mathbf{y}(t) = \mathbf{T}_m\mathbf{x}(t) = \mathbf{V}\mathbf{s}(t) \tag{25}$$

preserves the (deterministic) correlation of the source, i.e.,

$$\mathbf{y}^*(t)\mathbf{y}(t-k) = \mathbf{s}^*(t)\mathbf{s}(t-k) \qquad \forall k. \tag{26}$$

This implies that the correlation function of $\mathbf{s}(t)$ can be recovered from the observation process without knowing the channel.

B. Application of Viterbi Algorithm

It is the inner product preserving property (26) that enables a direct application of the Viterbi algorithm to the transformed observation $\mathbf{y}(t)$. Denote

$$r_y^{(k)}(t) = \mathbf{y}^*(t)\mathbf{y}(t-k) \tag{27}$$

$$r_s^{(k)}(t) = \mathbf{s}^*(t)\mathbf{s}(t-k) \tag{28}$$

$$= \sum_{l=0}^{d-1} s_{t-l}^* s_{t-l-k}. \tag{29}$$

We have

$$r_y^{(k)}(t) = r_s^{(k)}(t) + w^{(k)}(t) \tag{30}$$

where $w^{(k)}(t)$ accounts for the contribution of noise

$$w^{(k)}(t) = \mathbf{n}^*(t)\mathbf{T}_m^*\mathbf{T}_m\mathbf{H}\mathbf{s}(t-k) + \mathbf{s}^*(t)\mathbf{T}_m^*\mathbf{H}^*\mathbf{T}_m\mathbf{n}(t) + \mathbf{n}^*(t)\mathbf{T}_m^*\mathbf{T}_m\mathbf{n}(t-k). \tag{31}$$

An optimal sequence detection can be defined in the following sense

$$\min_{\{s_n\}} \sum_t |r_y^{(k)}(t) - r_s^{(k)}(t)|^2. \tag{32}$$

This optimization can be achieved by applying the Viterbi algorithm to a K^{d+k-1}-state trellis for a K-QAM signaling.

Remark:

1) There are inherent ambiguities in all blind estimation schemes. In blind sequence estimation, $\{s_k\}$ and $\{s_k e^{j\theta}\}$ have identical autocorrelations. Such an ambiguity cannot be eliminated without further information of the source. A common practice is to employ differential encoding of the information sequence. Of course, knowing a single symbol eliminates the ambiguity completely. This can be achieved by perhaps using a few flag bits at the beginning of the transmission.

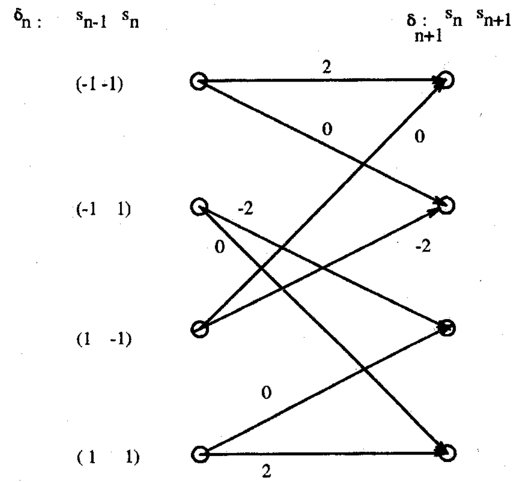

Fig. 2. The state transition diagram.

2) The delay parameter k in (32) should be chosen as small as possible so that the number of states is minimum. Obviously, for $k = 0$, $\{s_k\}$ cannot be uniquely determined from $r_y^{(0)}(t)$ since each symbol can have an arbitrary phase without affecting $r_y^{(0)}(t)$. Setting $k = 1$ is usually sufficient for the unique determination of the information sequence. In this case, the ambiguity is reduced to a common phase ambiguity for all symbols. The optimal sequence detection scheme can be extended with multiple selections of k's by using

$$\min_{\{s_n\}} \sum_{t,k} |r_y^{(k)}(t) - r_s^{(k)}(t)|^2 \tag{33}$$

as the objective function. Such an approach may improve the performance of the algorithm with added complexities.

Example: Blind Sequence Estimation of A BPSK Source: Consider an example when $d = 2$ with the BPSK constellation. We have

$$r_y^{(1)}(t) = s_t s_{t-1} + s_{t-1} s_{t-2} + w^{(1)}(t). \tag{34}$$

Define the state δ_n by

$$\delta_n = (s_{n-1}, s_n). \tag{35}$$

The state transition diagram is shown in Fig. 2. Each transition path from δ_n to δ_{n+1} is labeled by $r_s^{(1)}(n+1)$. Fig. 3 shows a typical sequence $\{s_n\} = \{-1, -1, 1, 1, 1\}$, the correlation sequence $r_y^{(1)}(t)$ is the estimate of $r_s^{(1)}(t)$, and the corresponding optimal path found by the Viterbi algorithm.

C. An Optimal Estimator of the Source Correlation

There is a potential problem from the above approach when the channel parameter matrix $\mathbf{H}$ is ill-conditioned. In such a case, $\mathbf{\Lambda}_s$ in (21) is close to singular, and the Mahalanobis transform $\mathbf{T}_m$ defined in (20) tends to greatly enhance the noise. In addition, the Mahalanobis transform enables us to find the source correlation from the correlation of the observation $\mathbf{y}(t)$ only when there is no noise (26). Such an approach does not result in an optimal estimation of the source correlation.

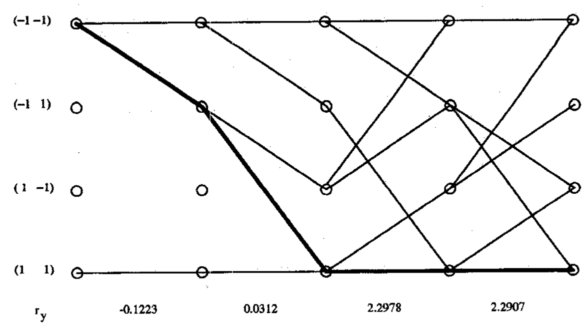

Fig. 3. The trellis diagram and the optimal path of a typical example.

The key idea of the proposed approach is to obtain an optimal estimation of the source (deterministic) correlation $r_s(t)$

$$r_s(t) = \mathbf{s}^*(t)\mathbf{s}(t-1). \tag{36}$$

The Viterbi algorithm can then be applied to the estimated $r_s(t)$ to recover the source sequence. Motivated by the Mahalanobis transformation, we propose to estimate $r_s(t)$ from the (deterministic) correlation function of a transformed observation. Denote

$$\mathbf{y}(t) = \mathbf{T}\mathbf{x}(t) \tag{37}$$

$$r_y(t) = \mathbf{y}^*(t)\mathbf{y}(t-1). \tag{38}$$

The objective is to find the optimal transformation matrix $\mathbf{T}_o$ that minimizes

$$J(\mathbf{T}) = E(|r_y(t) - r_s(t)|^2). \tag{39}$$

Fortunately, under some mild conditions, the above optimization can be obtained in closed form.

Theorem 1: Assume, in addition to A.1–A.4, the information sequence $\{s_k\}$ is i.i.d., and the noise $n_i(\cdot)$ is independent of $n_j(\cdot)$ and the source $\{s_k\}$, a matrix $\mathbf{T}_o$ that minimizes $J(\mathbf{T})$, is given by

$$\mathbf{T}_o = (\sigma^2\mathbf{I} + \mathbf{\Lambda}_s^2)^{-1}\mathbf{\Lambda}_s\mathbf{U}_s^* \tag{40}$$

where σ^2, $\mathbf{\Lambda}_s$, and $\mathbf{U}_s$ are obtained from the SVD of $\mathbf{R}_x(0)$ in (17)–(18). The minimum estimation error is given by

$$J(\mathbf{T}_o) = \sum_{i=1}^{d}(1 - \frac{\lambda_i^4}{\pi_i^4}) \tag{41}$$

where λ_i^2 and π_i^2, also defined in (17)–(18), are the singular values of $\mathbf{R}_x(0)$ for the noise-free and noisy observation, respectively (17).

Remark: When there is no noise, the transform (40) reduces to the Mahalanobis transformation and the source correlation can be estimated with no error. When the source is uncorrelated but not i.i.d., $\mathbf{T}_o$ has a different interpretation. See Section III.D.

Proof: In the Appendix, we show that

$$J(\mathbf{T}) = tr(\mathbf{PQPQ}) + \sigma^4 tr(\mathbf{P}^2) + 2\sigma^2 tr(\mathbf{P}^2\mathbf{Q}) - 2tr(\mathbf{PQ}) + d \quad (42)$$

where d is the signal subspace dimension (it is also the column rank of $\mathbf{H}$), and $\mathbf{P}$ and $\mathbf{Q}$ are both Hermitian matrices

$$\mathbf{P} = \mathbf{T}^*\mathbf{T} \quad (43)$$

$$\mathbf{Q} = \mathbf{HH}^*. \quad (44)$$

To find $\mathbf{T}_o$, we find first the optimum $\mathbf{P}_o$ that minimizes $J(\mathbf{T})$ in (42) by setting

$$\frac{\partial J(\mathbf{T})}{\partial \mathbf{P}} = \mathbf{0}. \quad (45)$$

We have

$$\mathbf{QPQ} + \sigma^4\mathbf{P} + 2\sigma^2\mathbf{PQ} - \mathbf{Q} = \mathbf{0}. \quad (46)$$

From (16)–(18), and (44)

$$\mathbf{Q} = \mathbf{HH}^* \quad (47)$$

$$= \mathbf{R}_x(0) - \sigma^2\mathbf{I} \quad (48)$$

$$= \mathbf{U}_s\Lambda_s^2\mathbf{U}_s^*. \quad (49)$$

It can then be easily verified that a solution of (46) is given by

$$\mathbf{P}_o = \mathbf{U}_s\mathbf{\Sigma}_p^2\mathbf{U}_s^* \quad (50)$$

$$\mathbf{\Sigma}_p^2 = (\sigma^2\mathbf{I} + \mathbf{\Lambda}_s^2)^{-2}\mathbf{\Lambda}_s^2. \quad (51)$$

Therefore, from (43), a transformation matrix $\mathbf{T}_o$ (among infinite many) that minimizes $J(\mathbf{T})$ is given by

$$\mathbf{T}_o = (\sigma^2\mathbf{I} + \mathbf{\Lambda}_s^2)^{-1}\mathbf{\Lambda}_s\mathbf{U}_s^*. \quad (52)$$

Substituting $\mathbf{P}_o$ and $\mathbf{Q}$ into (42), we have the estimation error

$$J(\mathbf{T}_o) = tr(\mathbf{\Sigma}_p^4\mathbf{\Lambda}_s^4 + \sigma^4\mathbf{\Sigma}_p^4 + 2\sigma^2\mathbf{\Sigma}_p^4\mathbf{\Lambda}_s^2 - 2\mathbf{\Sigma}_p^2\mathbf{\Lambda}_s^2) + d \quad (53)$$

$$= d + tr(\mathbf{\Sigma}_P^4(\sigma^2\mathbf{I} + \mathbf{\Lambda}_s^2)^2 - 2\mathbf{\Sigma}_p^2\mathbf{\Lambda}_s^2). \quad (54)$$

Substituting Σ_p^2 in (51) into the above equation, we have

$$J(\mathbf{T}_o) = d - tr((\sigma^2\mathbf{I} + \mathbf{\Lambda}_s^2)^{-2}\mathbf{\Lambda}_s^4) \quad (55)$$

$$= d - \sum_{i=1}^{d} \frac{\lambda_i^4}{(\sigma^2 + \lambda_i^2)^2} \quad (56)$$

$$= \sum_{i=1}^{d}(1 - \frac{\lambda_i^4}{\pi_i^4}). \quad (57)$$

□□□

D. A Simple Interpretation of the Optimal Estimation

The optimal estimation of source correlation has a simple interpretation that provides some insight. From Property 1, we have

$$\mathbf{H} = \mathbf{U_s\Lambda_sV} \quad (58)$$

where $\mathbf{U_s}$ and $\mathbf{\Lambda_s}$ can be computed from $\mathbf{R_x}(0)$ and $\mathbf{V}$ is some unknown orthogonal matrix. Substituting (58) into (13), we obtain

$$\mathbf{x}(t) = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{z}(t) + \mathbf{n}(t), \quad (59)$$

$$\mathbf{z}(t) = \mathbf{Vs}(t). \quad (60)$$

Since $\mathbf{V}$ is orthogonal, the correlation function of $\mathbf{z}(t)$ is the same as the correlation function of $\mathbf{s}(t)$

$$\mathbf{z}^*(t)\mathbf{z}(t-k) = \mathbf{s}^*(t)\mathbf{s}(t-k). \quad (61)$$

Note that $\mathbf{U_s}$ and $\mathbf{\Lambda_s}$ are known. Therefore, the Manalanobis transform

$$\mathbf{y}(t) = \mathbf{T}_m\mathbf{x}(t) \quad (62)$$

$$= \mathbf{z}(t) + \mathbf{\Lambda}_s^{-1}\mathbf{U}_s^*\mathbf{n}(t) \quad (63)$$

can be viewed as the least-square estimate of $\mathbf{z}(t)$. When $\mathbf{\Lambda_s}$ is close to singular, noise is greatly enhanced.

A better approach is using the minimum variance estimate $\hat{\mathbf{z}}(t)$ of $\mathbf{z}(t)$, and estimate $r_s(t) = \mathbf{s}^*(t)\mathbf{s}(t-1)$ by $r_{\hat{z}}(t) = \hat{\mathbf{z}}^*(t)\hat{\mathbf{z}}(t-1)$. Interestingly, the linear transform $\mathbf{T}_{mv}$ that gives the minimum variance estimate of $\mathbf{z}(t)$ turns out to be the same matrix that provides the minimum variance estimate of the source correlation r_x given by (40).

From (59), the minimum variance linear estimate of $\mathbf{z}(t)$ is given by [12]

$$\hat{\mathbf{z}}(t) = \mathbf{T}_{mv}\mathbf{x}(t), \quad (64)$$

$$\mathbf{T}_{mv} = E(\mathbf{z}(t)\mathbf{x}^*(t))\mathbf{R}_x^{-1}(0). \quad (65)$$

Substituting $\mathbf{x}(t)$ from (59) and $\mathbf{R}_x(0)$ from (16) into (65), we have

$$\mathbf{T}_{mv} = \mathbf{\Lambda}_s\mathbf{U}_s^*\mathbf{U}\text{diag}^{-1}(\pi_1^2, \cdots, \pi_N^2)\mathbf{U}^* \quad (66)$$

$$= (\sigma^2\mathbf{I} + \mathbf{\Lambda}_s^2)^{-1}\mathbf{\Lambda}_s\mathbf{U}_s^* = \mathbf{T}_o. \quad (67)$$

E. Summary of the Algorithm and Implementation

A schematic of the approach proposed in this section is shown in Fig. 4. The transformation matrix $\mathbf{T}_o$ is first estimated, perhaps continuously, from the windowed data by computing the singular value decomposition of the sampled covariance $\hat{\mathbf{R}}_x(0)$. The signal dimension d and the noise variance σ^2 are also estimated from this decomposition. The correlation function of the transformed data is computed and fed into the Viterbi algorithm. One of the most important features of this approach is that the Viterbi algorithm is time invariant even for time varying channels. This, of course, means significant simplifications of implementation where the Viterbi algorithm can be incorporated as a stand alone part even when the channel varies with time.

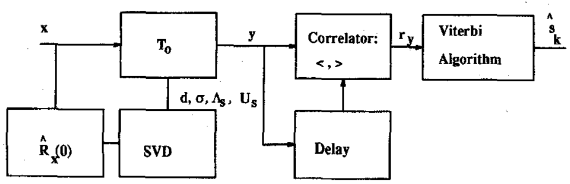

Fig. 4. A schematic of the proposed approach.

1) Blind Sequence Estimation Algorithm—An Outline:

a) Form the data vector $\mathbf{x}_i(t)$ by sampling the received signal at each receiver at a rate higher, perhaps, than the symbol rate

$$\mathbf{x}_i(t) = [x_i(tT), \cdots, x_i((t+1)T - 1)]^t. \tag{68}$$

b) Collecting data from each receiver and form the total data vector $\mathbf{x}(t)$ by

$$\mathbf{x}(t) = [\mathbf{x}_1^t(t), \cdots \mathbf{x}_M^t(t)]^t. \tag{69}$$

c) Estimate $\mathbf{R}_x(0)$ by

$$\hat{\mathbf{R}}_x(0) = \frac{1}{L}\sum_{i=0}^{L-1} \mathbf{x}(t-i)\mathbf{x}^*(t-i). \tag{70}$$

Appropriate data windows can be used to estimate $\mathbf{R}_x(0)$ continuously.

d) Compute the SVD of $\hat{\mathbf{R}}_x(0)$

$$\hat{\mathbf{R}}_x(0) = \mathbf{U}\mathrm{diag}(\pi_1^2, \cdots \pi_N^2)\mathbf{U}^*. \tag{71}$$

e) Estimate the signal dimension d from the singular values $\{\pi_i^2\}$. There are many detection schemes [25], [10]. See the discussion followed by this outline and the simulation examples.

f) Estimate the noise variance σ^2 by

$$\hat{\sigma}^2 = \frac{1}{N-d}\sum_{i=d+1}^{N} \pi_i^2. \tag{72}$$

g) Extract the first d singular vectors $\mathbf{u}_i$ and let

$$\mathbf{U}_s = [\mathbf{u}_1, \cdots, \mathbf{u}_d] \tag{73}$$

$$\mathbf{\Lambda}_s = \mathrm{diag}(\sqrt{\pi_1^2 - \hat{\sigma}^2}, \cdots \sqrt{\pi_d^2 - \hat{\sigma}^2}). \tag{74}$$

h) Form the optimal transform matrix $\mathbf{T}_o$ by

$$\mathbf{T}_o = (\mathbf{\Lambda}_s^2 + \hat{\sigma}^2\mathbf{I})^{-1}\mathbf{\Lambda}_s\mathbf{U}_s^*. \tag{75}$$

i) Apply the linear transform to the data vector

$$\mathbf{y}(t) = \mathbf{T}_o\mathbf{x}(t). \tag{76}$$

j) Compute the (deterministic) correlation

$$r_y(t) = \mathbf{y}^*(t)\mathbf{y}(t-1). \tag{77}$$

k) Estimate the source symbol sequence via the Viterbi algorithm by minimizing

$$\min_{\{s_k\}} \sum_t |r_y(t) - r_s(t)|^2 \tag{78}$$

where $r_s(t) = \sum_{i=0}^{d} s_{t-i}^* s_{t-i-1}$.

2) Implementation Issues: Indeterminacy: It is well known the blind equalization can only be achieved up to an unknown constant phase and delay [1]. The use of the Viterbi algorithm reduces the ambiguity somewhat because the source symbols are constrained. For example, there is a sign ambiguity associated with the estimated symbols when the BPSK signal is used.

Signal Subspace Dimension: The determination of the dimension d of the signal subspace appears to be quite important from our experience in simulation. Since channel impulse responses often have small tails, exact determination of d is hardly possible, nor does it seem to be necessary. In practice, using lower dimension approximation is often satisfactory. More importantly, smaller d results in smaller trellis in the Viterbi algorithm. In our simulation, we start with small d and increase d if necessary. Of course, there is a trade-off between better performance and computation cost.

IV. Simulation Examples

We present in this section, a simulation example to demonstrate the performance of the proposed approach. We considered 3-ray multipath channels. We assumed to have three receivers and the sampling rate at each receiver was four times the symbol rate ($T = 4$). The composite channels were given by

$$\begin{aligned} h_i(t) =& \alpha_{i1}p(t) + \alpha_{i2}p(t-\tau_1) + \alpha_{i3}p(t-\tau_2), \\ i =& 1, 2, 3 \end{aligned} \tag{79}$$

where $p(\cdot)$ was a raised-cosine pulse with 90% roll-off. The delays were $\tau_1 = 0.7T$ and $\tau_2 = 1.2T$. The gain α_{ij} at the ith receiver with respect to the jth multipath is seen in (80), generated from a zero mean unit variance Gaussian distribution (see equation (80) at the bottom of the page). The real and imaginary parts of the channel are shown in Fig. 5. The corresponding channel matrix $\mathbf{H}$, obtained by truncating the channel impulse responses up to six symbol intervals was given by (81) (please see equation (81) at the bottom of the next page). The truncation errors for these channels, as one can see from the first and last columns of $\mathbf{H}$, are negligible. We note that all the baud rate sampled channels are non-minimum phase. It is clear from the row vectors of $\mathbf{H}$ that the intersymbol interference is severe for this multipath channel. Noise was added at the receiver with signal-to-noise ratio (SNR)

$$[\alpha_{ij}] = \begin{pmatrix} 1.2094 + 0.8260i & 0.6251 - 0.1450i & -0.9660 + 1.6894i \\ -0.5123 - 1.3070i & 0.1163 + 0.3929i & 0.2007 - 0.0755i \\ 0.8922 + 0.9049i & -0.0075 - 0.0219i & 0.1018 + 0.1685i \end{pmatrix} \tag{80}$$

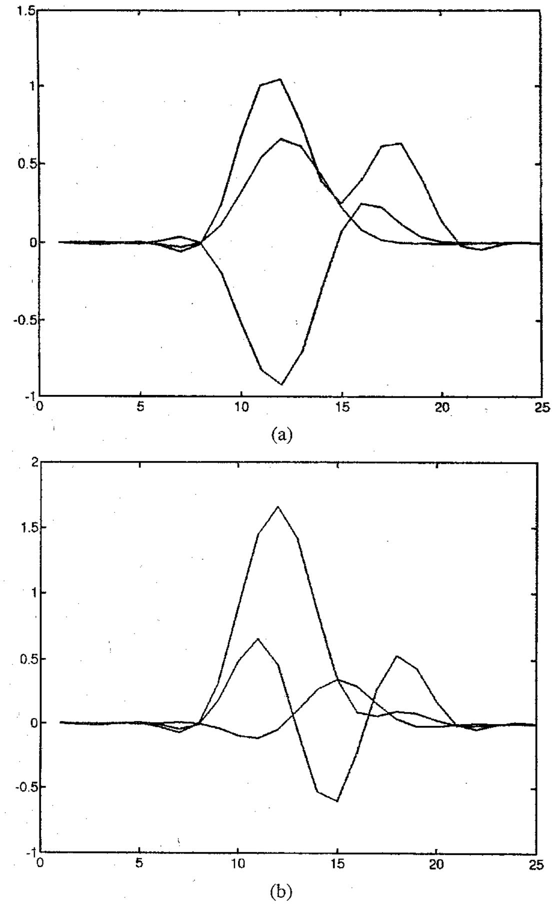

Fig. 5. Channel impulse responses. (a) Real parts of the channel impulse responses for the three receivers. (b) Imaginary parts of the channel impulse responses for the three receivers.

defined by

$$SNR = 10\log_{10}\frac{E(||\mathbf{Hs}(t)||^2)}{E(||\mathbf{n}(t)||^2)}. \tag{82}$$

The input source was a BPSK signal.

The algorithm was executed in a batch manner. For every 100 bits of data, we estimated the covariance matrix $\mathbf{R}_x(0)$. We did not assume that the dimension of the signal space was known. Starting with signal subspace dimension $d = 1$, the transformation matrix $\mathbf{T}_o$ was formed from the SVD of the estimated covariance matrix $\hat{\mathbf{R}}_x(0)$. The correlation function of the transformed data was computed and fed into the Viterbi algorithm. We did not assume the initial state of trellis was known, although such an assumption may be reasonable and will improve the performance. If there were errors in the symbol estimation, the signal dimension d was increased to obtain the sequence estimate with minimum error. Such a scheme may be implemented in practice if the source were coded with error detection capability. We used this scheme for the purpose of finding the best performance that can be offered by this algorithm. Fig. 6 shows the bit-error-rate vs. SNR. Total 10^5 bits were tested for SNR below 6 dB, 10^6 bits for $SNR = 8$ dB and 10^7 bits for $SNR = 9$ dB. For $SNR = 10$ dB, there was no error found in 10^7 testing bits. It appears that using 100 symbols to estimate the necessary statistics was adequate.

V. Conclusion

We presented a blind sequence estimation method that does not require complete channel identification. It exploits both statistical and algebraic properties of the source. It can be applied to both single and multiple receiver structures. From numerical simulations, the algorithm seems to perform well using a relatively small amount of data which shows the potential application of this approach to wireless communications. One of the important future works is the performance analysis of this algorithm.

Appendix

Here we will derive (42).

$$\begin{aligned} J(\mathbf{T}) =& tr(\mathbf{PQPQ}) + \sigma^4 tr(\mathbf{P}^2) \\ &+ 2\sigma^2 tr(\mathbf{P}^2\mathbf{Q}) - 2tr(\mathbf{PQ}) + d. \end{aligned} \tag{83}$$

$$\mathbf{H} = \begin{pmatrix} 0.0009+0.0028i & 0.0072+0.0103i & 0.2430+0.1870i & 0.7540-0.0495i & 0.6062+0.2760i & -0.0191-0.0091i \\ -0.0055-0.0024i & -0.0164-0.0059i & 0.6565+0.4884i & 0.3909-0.5281i & 0.6253+0.5310i & -0.0417-0.0436i \\ -0.0100-0.0078i & -0.0553-0.0410i & 1.0065+0.6577i & 0.2543-0.6012i & 0.4159+0.4348i & -0.0081-0.0068i \\ -0.0010-0.0003i & -0.0060-0.0075i & 1.0453+0.4638i & 0.3991-0.2262i & 0.1434+0.1755i & 0.0071+0.0125i \\ -0.0004-0.0009i & 0.0015-0.0027i & 0.1167-0.0361i & 0.6059+0.1097i & 0.0194+0.1615i & -0.0050-0.0025i \\ -0.0031+0.0000i & -0.0102-0.0000i & 0.3228-0.0922i & 0.4313+0.2751i & 0.0006+0.0392i & -0.0022+0.0031i \\ -0.0040+0.0011i & -0.0254+0.0067i & 0.5396-0.1144i & 0.2303+0.3498i & -0.0034-0.0175i & -0.0003-0.0004i \\ -0.0001-0.0003i & 0.0002+0.0008i & 0.6543-0.0464i & 0.0864+0.2950i & -0.0055-0.0180i & -0.0008-0.0027i \\ -0.0002-0.0003i & -0.0043+0.0062i & -0.1883+0.3240i & -0.7068+1.4093i & 0.2313+0.0643i & 0.0025-0.0133i \\ 0.0041-0.0079i & 0.0135-0.0256i & -0.5144+0.8889i & -0.2989+0.8589i & 0.1242+0.0987i & 0.0015-0.0146i \\ 0.0061-0.0113i & 0.0393-0.0702i & -0.8265+1.4448i & 0.0762+0.3477i & 0.0390+0.0831i & -0.0011-0.0020i \\ -0.0004-0.0002i & -0.0002-0.0011i & -0.9203+1.6637i & 0.2532+0.0924i & 0.0065+0.0263i & -0.0009+0.0012i \end{pmatrix} \tag{81}$$

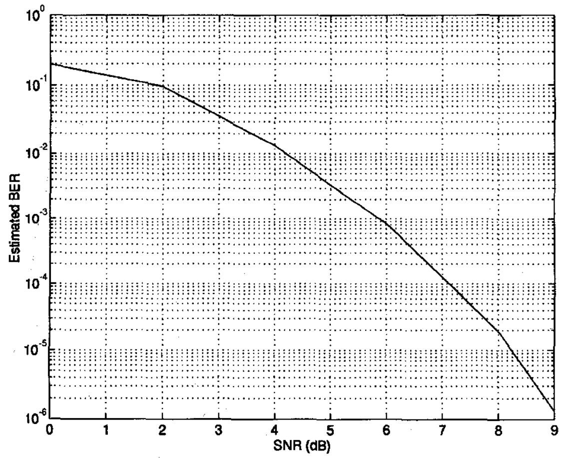

Fig. 6. BER versus SNR for a fixed channel. Signal statistics were estimated every 100 symbols. 10^5 bits were tested for $SNR \leq 6$ dB, 10^6 bits for $SNR = 8$ dB and 10^7 bits for $SNR = 9$ dB.

Recall that

$$\mathbf{x}(t) = \mathbf{Hs}(t) + \mathbf{n}(t) \quad (84)$$

$$\mathbf{y}(t) = \mathbf{Tx}(t). \quad (85)$$

Denote

$$r_s = \mathbf{s}^*(t)\mathbf{s}(t-1) \quad (86)$$

$$r_y = \mathbf{y}(t)^*\mathbf{y}(t-1) \quad (87)$$

$$\mathbf{P} = \mathbf{T}^*\mathbf{T} \quad (88)$$

$$\mathbf{Q} = \mathbf{HH}^* \quad (89)$$

$$J = E(|r_y - r_{\tilde{s}}|^2). \quad (90)$$

We have

$$J = E(|r_y - r_s|^2) \quad (91)$$

$$= E(|r_y|^2) - E(r_y r_s^*) - E(r_y^* r_s) + E(|r_s|^2). \quad (92)$$

We shall now proceed to evaluate each term in (91).

Computation of $E(|r_y|^2)$: From (84), (85), (88), we have

$$r_y = \mathbf{x}(t)^*\mathbf{Px}(t-1), \quad (93)$$

$$= \underbrace{\mathbf{s}^*(t)\mathbf{H}^*\mathbf{PHs}(t-1)}_{r_y^{(i)}} + \underbrace{\mathbf{n}^*(t)\mathbf{Pn}(t-1)}_{r_y^{(ii)}} + \underbrace{\mathbf{s}^*(t)\mathbf{H}^*\mathbf{Pn}(t-1)}_{r_y^{(iii)}} + \underbrace{\mathbf{n}^*(t)\mathbf{P}^*\mathbf{Hs}(t-1)}_{r_y^{(iv)}} \quad (94)$$

Based on the assumptions given in Theorem 1, it is easily shown that

$$E(|r_y|^2) = E(|r_y^{(i)}|^2) + E(|r_y^{(ii)}|^2) + E(|r_y^{(iii)}|^2) + E(|r_y^{(iv)}|^2). \quad (95)$$

We then have

$$E(|r_y^{(i)}|^2) = E(\mathbf{s}^*(t)\mathbf{H}^*\mathbf{PHs}(t-1)\mathbf{s}^* \cdot (t-1)\mathbf{H}^*\mathbf{PHs}(t)) \quad (96)$$

$$= tr(\mathbf{H}^*\mathbf{PH}E(\mathbf{s}(t-1)\mathbf{s}^* \cdot (t-1)\mathbf{H}^*\mathbf{PHs}(t)\mathbf{s}^*(t))). \quad (97)$$

Let $\mathbf{\Pi} = \mathbf{H}^*\mathbf{PH} = [\pi_{ij}]$, where π_{ij} is the $(i,j)^{th}$ component of $\mathbf{\Pi}$. We have

$$E(|r_y^{(i)}|^2) = tr(\mathbf{\Pi}^* E(\mathbf{s}(t-1)\mathbf{s}^*(t)\mathbf{s}^*(t-1)\mathbf{\Pi s}(t))). \quad (98)$$

Recalling from (9), that $\mathbf{s}(t) = [s_t, \cdots, s_{t-d+1}]^t$, and that $\{s_t\}$ is an i.i.d. sequence, the (i,j)th component of $E(\mathbf{s}(t-1)\mathbf{s}^*(t)\mathbf{s}^*(t-1)\mathbf{\Pi s}(t))$ is given by

$$E(s_{t-i}s_{t-j+1}^* \sum_{k,l} \pi_{kl} s_{t-k}^* s_{t-l+1}) = \pi_{ij}. \quad (99)$$

Hence

$$E(|r_y^{(i)}|^2) = tr(\mathbf{\Pi}^*\mathbf{\Pi}) \quad (100)$$

$$= tr(\mathbf{H}^*\mathbf{P}^*\mathbf{HH}^*\mathbf{PH}) \quad (101)$$

$$= tr(\mathbf{P}^*\mathbf{QPQ}^*). \quad (102)$$

Next, we evaluate the second term in (95)

$$E(|r_y^{(ii)}|^2) = E(\mathbf{n}^*(t)\mathbf{Pn}(t-1)\mathbf{n}^* \cdot (t-1)\mathbf{P}^*\mathbf{n}(t)) \quad (103)$$

$$= tr(\mathbf{P}E(n(t-1)n^* \cdot (t-1)\mathbf{P}^*n(t)n^*(t))). \quad (104)$$

From the assumption that the noise processes are uncorrelated and white, we have

$$E(|r_y^{(ii)}|^2) = \sigma^2 tr(\mathbf{PP}^*). \quad (105)$$

The third and the fourth terms are evaluated similarly

$$E(|r_y^{(iii)}|^2) = E(\mathbf{s}^*(t)\mathbf{H}^*\mathbf{Pn}(t-1)\mathbf{n}^*(t-1) \mathbf{P}^*\mathbf{Hs}(t)) \quad (106)$$

$$= \sigma^2 tr(\mathbf{H}^*\mathbf{PP}^*\mathbf{H}) \quad (107)$$

$$= \sigma^2 tr(\mathbf{PP}^*\mathbf{Q}) \quad (108)$$

$$E(|r_y^{(iv)}|^2) = E(\mathbf{n}^*(t)\mathbf{P}^*\mathbf{Hs}(t-1)\mathbf{s}^*(t-1) \mathbf{H}^*\mathbf{Pn}(t)) \quad (109)$$

$$= tr(\mathbf{P}^*\mathbf{H}E(\mathbf{s}(t-1)\mathbf{s}^*(t-1)) \mathbf{H}^*\mathbf{P}E(\mathbf{n}(t)\mathbf{n}^*(t))) \quad (110)$$

$$= \sigma^2 tr(\mathbf{QPP}^*) \quad (111)$$

$$= E(|r_y^{(iii)}|^2). \quad (112)$$

Hence, with both $\mathbf{P}$ and $\mathbf{Q}$ being Hermitian, we have

$$E(|r_y|^2) = tr(\mathbf{PQPQ}) + \sigma^4(\mathbf{P}^2) + 2\sigma^2 tr(\mathbf{P}^2\mathbf{Q}). \quad (113)$$

Computation of $E(r_s^* r_y)$ and $E(r_y^* r_s)$: From (84), (85), (88), (86), (87), we have

$$r_s^* r_y = \mathbf{s}^*(t)\mathbf{H}^*\mathbf{PHs}(t-1)\mathbf{s}^*(t-1)\mathbf{s}(t) + \mathbf{n}^*(t)\mathbf{Pn}(t-1)\mathbf{s}^*(t-1)\mathbf{s}(t) \quad (114)$$

$$+ \mathbf{s}^*(t)\mathbf{H}^*\mathbf{Pn}(t-1)\mathbf{s}^*(t-1)\mathbf{s}(t) + \mathbf{n}^*(t)\mathbf{PHs}(t-1)\mathbf{s}^*(t-1)\mathbf{s}(t). \quad (115)$$

Again, using assumptions **A.1**, **A.2**, and **A.3**, we have

$$E(r_s^* r_y) = tr(\mathbf{H}^*\mathbf{PH}) \quad (116)$$

$$= tr(\mathbf{PQ}) \quad (117)$$

$$E(r_y^* r_s) = tr(\mathbf{PQ}). \quad (118)$$

The Computation of $E(|r_s|^2)$: Under the assumption that the source symbols are independent, we have

$$E(|r_s|^2) = \sum_{i,j=1}^{d} E(s^*_{t+1-i}s_{t-i}s^*_{t-j}s_{t+1-j}) \quad (119)$$

$$= \sum_{i=1}^{d} E(|s_{t+1-i}|^2|s_{t-i}|^2) = d. \quad (120)$$

We now have, from (113), (117), (118)

$$J = tr(\mathbf{PQPQ}) + \sigma^4 tr(\mathbf{P}^2) + 2\sigma^2 tr(\mathbf{P}^2\mathbf{Q}) - 2tr(\mathbf{PQ}) + d. \quad (121)$$

□□□

ACKNOWLEDGMENT

The author wishes to thank the referee for the constructive comments and the suggestions of using multiple delays in the proposed approach.

REFERENCES

[1] M. Goursat, A. Benveniste, and G. Ruget, "Robust identification of a nonminimum phase system: blind adjustment of a linear equalizer in data communication," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 385–399, June 1980.

[2] P. Balaban and J. Salz, "Optimum diversity combining and equalization in digital data transmission with applications to cellular mobile radio - part I: Theoretical considerations," *IEEE Trans. Commun.*, vol. 40, pp. 885–894, May 1992.

[3] G. D. Forney, "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Inform. Theory*, vol. IT-18, pp. 363–378, May 1972.

[4] ——, "The Viterbi algorithm," *IEEE Proc.*, vol. 61, pp. 268–278, Mar. 1972.

[5] G. J. Foschini, "Equalizing without altering or detecting data," *Bell Syst. Tech. J.*, vol. 64, pp. 1885–1911, Oct. 1985.

[6] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867–1875, Nov. 1980.

[7] D. J. Goodman, "Trends in cellular and cordless communications," *IEEE Commun. Mag.*, pp. 31–40, June 1991.

[8] D. Hatzinakos and C.Nikias, "Estimation of multipath channel response in frequency selective channels," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 12–19, Jan. 1989.

[9] F. R. Magee Jr. and J.G. Proakis, "Adaptive maximum-likelihood sequence estimation for signaling in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 120–124, Jan. 1973.

[10] K. Konstantinides and K. Yao, "Statistical analysis of effective singular values in matrix rank determination," *IEEE Trans. ASSP*, vol. 36, pp. 757–763, May 1988.

[11] Y. Li and Z. Ding, "Blind channel identification based on second order cyclostationary statistics," in *Proc. ICASSP Conf.*, 1993.

[12] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: John Wiley, 1969.

[13] G. Picci and G.Prati, "Blind equalization and carrier recovery using a stop-and-go decision directed algorithm," *IEEE Trans. Commun.*, vol. COM-35, pp. 877–887, Sept. 1987.

[14] B. Porat and B. Friedlander, "Blind equalization of digital communication channels using high-order moments," *IEEE Trans. Signal Processing*, vol. 39, pp. 522–526, 1991.

[15] S. Prakriya and D. Hatzinakos, "Efficient nonlinear channel identification using cyclostationary signal analysis," in *Proc. ICASSP Conf.*, vol. 4, 1993, pp. 264–267.

[16] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation," *IEEE Trans. Commun.*, vol. COM-23, pp. 679–682, June 1975.

[17] N. Seshadri, "Joint data and channel estimation using fast blind trellis search techniques," in *Proc. Globecom*, 1991, pp. 1659–1663.

[18] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Inform. Theory*, vol. 36, pp. 312–320, Mar. 1990.

[19] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind identification and equalization of multipath channels: a frequency domain approach," submitted to *IEEE Trans. Inform. Theory*.

[20] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *25th Asilomar Conf.*, 1991.

[21] ——, and T. Kailath, "Fast blind equalization of multipath channels via antenna arrays," in *Proc. ICASSP*, 1993.

[22] ——, and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, Mar. 1994.

[23] J.R. Treichler and B.G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. ASSP.*, vol. ASSP-31, pp. 459–472, Apr. 1983.

[24] G. Ungerboeck, "Adaptive maximum likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. Commun.*, vol. COM-22, pp. 624–636, May 1974.

[25] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. ASSP*, vol. ASSP-33, pp. 387–392, Apr. 1985.

[26] E. Zervas, J. Proakis, and V. Eyuboğlu, "A "quantized" channel approach to blind equalization," in *Proc. ICC*, vol. 3, 1992, pp. 1539–1643.

Lang Tong (S'88–M'90) received the B.E. degree in control engineering from Tsinghua University, Beijing, China, in 1985, and M.S. and Ph.D. degrees in electrical engineering, from the University of Notre Dame, South Bend, Indiana, in 1987 and 1990, respectively.

After being a Postdoctoral Research Affiliate at the Information Systems Laboratory, Stanford University, he joined the Department of Electrical and Computer Engineering, West Virginia University. Since 1993, he has been with the Department of Electrical and Systems Engineering at the University of Connecticut, Storrs, CT. He also held a Visiting Assistant Professor position at Stanford University in the summer of 1992. His research interests are in the areas of digital communication, signal processing, and system theory.

Dr. Tong received the *Best Graduating Student Award* from the University of Notre Dame, and the *Outstanding Young Author Award* from the IEEE Circuits and Systems Society.