A Channel Subspace Post-Filtering Approach to Adaptive Least-Squares Estimation

Raj Nadakuditi, *Student Member, IEEE,* and James C. Preisig, *Member, IEEE*

***Abstract*—A major challenge while communicating in dynamic channels, such as the underwater acoustic channel, is the large amount of time-varying inter-symbol interference (ISI) due to multipath. In many realistic channels, the fluctuations between different taps of the sampled channel impulse response are correlated. Traditional least-squares algorithms used for adapting channel equalizers do not exploit this correlation structure. A channel subspace post-filtering algorithm is presented that treats the least-squares channel estimate as a noisy time series and exploits the channel correlation structure to reduce the channel estimation error. The improvement in performance of the post-filtered channel estimator is predicted theoretically and demonstrated using both simulation and experimental data. Experimental data is also used to demonstrate the improvement in performance of a channel estimate-based decision feedback equalizer that uses this post-filtered channel estimate to determine the equalizer coefficients.**

***Index Terms*—Adaptive equalization, channel estimate-based decision feedback equalization, least-squares algorithms, multipath channel, reduced subspace methods, system identification, underwater acoustic communication.**

I. Introduction

TIME-VARYING intersymbol interference (ISI) due to multipath is one of the major problems encountered while communicating in dynamic channels [1]. Since it is generally not possible to use universal precoding techniques that eliminate the effect of this ISI, active research has emphasized the design of adaptive receiver algorithms that are able to track and compensate for this time-varying ISI. Least-squares-based algorithms [2], [3] are often used to track such time-varying systems.

There is an extensive body of literature dealing with the formulation and the analysis of commonly used least-squares-based methods such as the exponentially windowed recursive least-squares (EW-RLS) and the sliding window recursive least-squares (SW-RLS) algorithms. For such techniques, selecting an appropriate rate of adaptation involves a tradeoff between the tracking error variance and the observation noise induced error variance [3]. In channels with relatively short multipath responses and high received signal-to-noise ratios (SNRs), in the absence of computational constraints, optimal algorithm design using one of these methods primarily involves selecting the appropriate rate of adaptation based on the assumed channel-output model. On the other hand, for channels with long multipath responses and relatively low SNRs, merely varying the rate of adaptation leads to unsatisfactory algorithm performance.

There have, however, been attempts at improving algorithm performance by using techniques that exploit the characteristics of the transmitted data or that of the channel itself. These techniques can be broadly categorized into either subspace methods [4]–[6], which exploit the low-rank nature of the transmitted data covariance matrix or sparsing methods [7]–[9], which exploit the sparse nature of the energetic taps of the sampled channel impulse response. The method presented here exploits a more general form of channel sparseness, that is, the low-rank of the channel impulse response correlation matrix, to improve algorithm performance.

The main result presented in this paper utilizes a different paradigm introduced earlier in [10]–[12] while formulating and implementing what will be referred to as the channel subspace post-filtering (CSF) approach to adaptive least-squares estimation. This approach exploits the correlations between the different taps of the sampled channel impulse response and any low-rank nature of the channel subspace, that is, a consequence of the correlated tap fluctuations in long multipath channels, to improve the tracking performance of the channel estimator. The resultant improved channel estimate is then used to determine the coefficients of a standard channel estimate-based decision feedback equalizer (DFE). As a result, the performance of this DFE is improved as well. The CSF formulation is primarily motivated by the need to improve the tracking performance of classes of channels with low-rank multidimensional subspace structures. However, it will also prove to be relevant when analyzing and improving the tracking performance of single tap least-squares algorithms.

The paper is organized as follows. Section II introduces a simplified channel-output model that is used in the subsequent analysis, whereas Section III presents the metrics for assessing the performance of least-squares tracking algorithms. Section IV introduces and analyzes the tracking performance of the EW-RLS algorithm using an extended state space representation of an equivalent dynamic system. The EW-RLS channel estimate is treated as a noisy time series and is post-filtered using the channel subspace post-filtering (CSF) approach that is presented in Section V. An analogously constructed but

Manuscript received January 7, 2003; revised July 9, 2003. This work was supported by an ONR Ocean Acoustics Graduate Traineeship Award under Contract N00014-00-10049 and an ONR Ocean Acoustics Young Faculty Award under Contract N00014-00-10048. This paper is WHOI contribution number 10846. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Naofal M. W. Al-Dhahir.

R. Nadakuditi is with the Massachusetts Institute of Technology and the Woods Hole Oceanographic Institution, Woods Hole, MA 02543 USA (e-mail: raj@mit.edu).

J. Preisig is with the Woods Hole Oceanographic Institution, Woods Hole, MA 02543 USA (e-mail: jpreisig@whoi.edu).

Digital Object Identifier 10.1109/TSP.2004.828926

decidedly suboptimal post-filter, referred to as the abrupt rank reduction (ARR) filter, is presented in Section VI. Section VII analyzes the problem of channel order mismatch, where the assumed number of taps when formulating the EW-RLS algorithm is different from the actual number of taps in the true channel model. It will be shown that the CSF approach remains effective in such situations. Section VIII casts the CSF problem into a familiar deterministic least-squares minimization problem that leads to the development of an adaptive channel subspace filtering (ACSF) algorithm. Section IX demonstrates the improvement in channel estimator performance due to the CSF approach on a single tap channel, a two-tap channel with channel order mismatch, and on a low-rank multitap channel, using simulated data. Experimental data is used in Section X to demonstrate the improvement in channel estimator performance due to the ACSF approach in a practical setting. Section XI uses experimental data to show how these improved ACSF channel estimates can also improve the performance of a standard channel estimate-based DFE. Section XII summarizes the main contributions of this paper.

II. System and Channel Model

A simplified time-varying first-order Gauss–Markov model[1] for system identification [2] is used as the channel model. The unknown dynamic system is modeled as a transversal filter whose tap-weight vector $\underline{h}(n)$ (i.e., discrete time impulse response) evolves according to a first-order Markov process written in vector form as

$$\underline{h}(n+1) = \alpha \underline{h}(n) + \underline{v}(n+1) \tag{1}$$

where α is a possibly complex scalar, the underscore denotes vectors, and all vectors are $N \times 1$ column vectors. The time-varying channel impulse response at time n is represented by the vector $\underline{h}(n)$ with the process noise vector $\underline{v}(n)$ having a correlation matrix $\mathbf{R}_{vv} = E[\underline{v}(n)\underline{v}^H(n)]$. The output of the system is given by

$$y(n) = \underline{h}^H(n)\underline{x}(n) + w(n) \tag{2}$$

where the superscript H represents the conjugate transpose (Hermitian), the received data is $y(n)$, $\underline{x}(n)$ denotes the white transmitted data with correlation matrix $\mathbf{R}_{xx} = E[\underline{x}(n)\underline{x}^H(n)] = \sigma_x^2\mathbf{I}$, and $w(n)$ denotes additive white, Gaussian observation noise with a variance of σ^2. If $|\alpha| < 1$, (1) and (2) collectively describe a time-varying system with stationary statistics.

If $\hat{\underline{h}}(n \mid n-1)$ denotes an estimate of the true channel impulse response $\underline{h}(n)$ obtained using data up to time $n-1$, then the predicted data at time n, $\hat{y}(n \mid n-1)$ is given by $\hat{y}(n \mid n-1) = \hat{\underline{h}}^H(n \mid n-1)\underline{x}(n)$. The corresponding prediction error $\xi(n \mid n-1)$ is given by

$$\begin{aligned}\xi(n \mid n-1) &= y(n) - \hat{y}(n \mid n-1) \\ &= y(n) - \hat{\underline{h}}^H(n \mid n-1)\underline{x}(n).\end{aligned} \tag{3}$$

[1]This model has been chosen since it greatly simplifies the theoretical analysis that is to be presented in Sections IV and V. This resulting simplification facilitates the derivation of closed-form analytical expressions for the performance metrics that are to be introduced in Section III.

The prediction error is used to the adapt the weights of the channel estimate $\hat{\underline{h}}(n \mid n-1)$. Throughout the remainder of this paper, for notational simplicity, $\hat{\underline{h}}(n-1)$ is used to denote $\hat{\underline{h}}(n \mid n-1)$. Additionally, possibly time-varying quantities are explicitly described using a functional dependence on the time index n (e.g., $\mathbf{R}_{zz}(n)$), whereas the omission of the time dependence is used to implicitly denote time-invariant quantities (e.g., $\mathbf{R}_{zz}$). It is assumed that the process noise vector $\underline{v}(l)$, the input (data) vector $\underline{x}(m)$, the observation noise $w(n)$, and the initial channel impulse response $\underline{h}(k)|_{k=0}$ are independent of one another for all l, m, and n. Without loss of generality, an assumption is made that the number of taps in the unknown system $\underline{h}(n)$ is the same as the number of taps N in the adaptive filter $\hat{\underline{h}}(n)$ used to model and track this system. In Section VII, it will be shown that the proposed approach remains effective even when there is a channel order mismatch between the assumed and true channel model. These assumptions will be used in Sections III–XI when analyzing the tracking performance of the EW-RLS algorithm.

III. Criteria for Tracking Performance Assessment

The channel estimation error vector, which is also referred to as the tracking error vector, may be defined as

$$\underline{\epsilon}(n \mid n-1) = \underline{h}(n) - \hat{\underline{h}}(n-1). \tag{4}$$

The relationship between the channel estimation error vector and the prediction error, assuming the linear channel output model given in (2), can be written as

$$\begin{aligned}\xi(n \mid n-1) &= \underline{h}^H(n)\underline{x}(n) + w(n) - \hat{\underline{h}}^H(n-1)\underline{x}(n) \\ &= \left(\underline{h}(n) - \hat{\underline{h}}(n-1)\right)^H \underline{x}(n) + w(n).\end{aligned} \tag{5}$$

Substituting the expression for the channel estimation error vector given in (4) results in

$$\xi(n \mid n-1) = \underline{\epsilon}^H(n \mid n-1)\underline{x}(n) + w(n). \tag{6}$$

The mean-square channel estimation error is defined as

$$\begin{aligned}\mathcal{D}(n) &= E\left[\|\underline{h}(n) - \hat{\underline{h}}(n-1)\|^2\right] \\ &= E\left[\|\underline{\epsilon}(n \mid n-1)\|^2\right].\end{aligned} \tag{7}$$

In the subsequent analysis, it is assumed that n is large enough for any initialization transient of the least-squares algorithm to have passed. Equation (7) may be alternately written as

$$\mathcal{D}(n) = tr[\mathbf{R}_{\epsilon\epsilon}(n)] \tag{8}$$

where $\mathbf{R}_{\epsilon\epsilon}(n)$ is the correlation matrix of the error vector $\underline{\epsilon}(n \mid n-1)$ defined as

$$\mathbf{R}_{\epsilon\epsilon}(n) = E\left[\underline{\epsilon}(n \mid n-1)\underline{\epsilon}^H(n \mid n-1)\right]. \tag{9}$$

An additional metric $\mathbf{R}_{\epsilon\hat{h}}(n) = \mathbf{R}_{\hat{h}\epsilon}^H(n)$, which is the cross correlation matrix of the error vector $\underline{\epsilon}(n \mid n-1)$ and the channel estimate vector $\hat{\underline{h}}(n-1)$, is defined as

$$\mathbf{R}_{\epsilon\hat{h}}(n) = E\left[\underline{\epsilon}(n \mid n-1)\hat{\underline{h}}^H(n-1)\right]. \tag{10}$$

The orthogonality principle states that if $\mathbf{R}_{\epsilon\hat{h}}(n) \neq \mathbf{0}$, the channel estimate is not a minimum mean-square error estimate of the true channel impulse response.

The mean-square signal prediction error can also be expressed in terms of the mean-square channel estimation error $\mathcal{D}(n)$ using (6) such that

$$\begin{aligned} &E[\|\xi(n \mid n-1)\|^2] \\ &\quad = tr\left(E[\underline{\epsilon}(n)\underline{\epsilon}^H(n)\underline{x}(n)\underline{x}^H(n)]\right) + E[w^2(n)] \\ &\quad = tr\left(\underbrace{E[\underline{\epsilon}(n)\underline{\epsilon}^H(n)]}_{\mathbf{R}_{\epsilon\epsilon}(n)}\underbrace{E[\underline{x}(n)\underline{x}^H(n)]}_{\sigma_x^2\mathbf{I}}\right) \\ &\qquad + E[w^2(n)] \\ &\quad = \sigma_x^2\mathcal{D}(n) + \sigma^2. \end{aligned} \tag{11}$$

The transmitted data vector $\underline{x}(n)$ is assumed to be independent of the channel impulse response vector $\underline{h}(n)$, previous samples of the transmitted data vector $\underline{x}(l)$, and the additive observation noise $w(l)$ for $l = 1,\ldots,n-1$. It is thus also independent of previous samples of the received data $y(l)$, given by (2), for $l = 1,\ldots,n-1$. Hence, by the definition of $\hat{\underline{h}}(n-1)$ and the expression for the channel estimation error vector in (4), the transmitted data vector $\underline{x}(n)$ is independent of $\underline{\epsilon}(n \mid n-1)$. The second equality in (11) follows from this fact. Equation (11), which expresses the relationship between the channel estimation error variance and the mean-square prediction error, states that a reduction in channel estimation error leads to a corresponding scaled decrease in the prediction error. In a realistic scenario where the true channel impulse response is unknown, it is difficult to ascertain the true channel estimation error. The mean-square prediction error can therefore be used as a surrogate for the mean-square channel estimation error in gauging the performance of an adaptive algorithm. This equivalence will be used later on to replace an unobservable cost function expressed in terms of the channel estimation error with an observable cost function expressed in terms of the prediction error.

The mean-square channel estimation error can be expressed as a sum of two components

$$\mathcal{D}(n) = \mathcal{D}_1(n) + \mathcal{D}_2(n) \tag{12}$$

where the first term $\mathcal{D}_1(n) = tr(\mathbf{D}_1(n))$ is referred to as the tracking error variance, and $\mathcal{D}_2(n) = tr(\mathbf{D}_2(n))$ is referred to as the observation noise induced error variance. The terms in $\mathcal{D}_1(n)$ depend on the dynamics of the time-varying channel, whereas the terms in $\mathcal{D}_2(n)$ depend on the observation noise variance. Separating these error terms, as above, provides insight into the tradeoff involved when designing a least-squares algorithm to track a time-varying system. This decomposition is similar to that in [2].

For a given adaptive algorithm, best performance is achieved by selecting a rate of adaptation that balances the improvement due to any reduction in the tracking error variance $\mathcal{D}_1(n)$ with any resultant deterioration due to an increase in the observation noise induced error variance $\mathcal{D}_2(n)$. Section IV formally introduces a commonly used least-squares tracking algorithm, uses an extended state space framework to theoretically analyze its tracking performance in terms of the channel estimation error, and provides a basis for evaluating the improvement in its performance due to channel subspace post-filtering.

IV. EW-RLS Algorithm and Its Tracking Performance

The exponentially windowed least-squares estimate $\hat{\underline{h}}(n)$ of the channel impulse response $\underline{h}(n)$ is given by

$$\hat{\underline{h}}(n) = \underset{\underline{h}}{\operatorname{argmin}} \sum_{k=1}^{n} \lambda^{n-k}\|y(k) - \underline{h}^H\underline{x}(k)\|^2 \tag{13}$$

where λ is a positive constant close to but less than 1. The recursive update equation for such a least-squares algorithm is given by

$$\hat{\underline{h}}(n) = \hat{\underline{h}}(n-1) + \underline{\mathbf{K}}(n)\xi^*(n \mid n-1) \tag{14}$$

where the superscript $*$ denotes complex conjugation, $\underline{\mathbf{K}}(n)$ is an adaptation gain vector, and the prediction error $\xi(n \mid n-1)$ is defined as in (3). The adaptation gain vector may be computed as

$$\underline{\mathbf{K}}(n) = \left(\sum_{i=1}^{n} \lambda^{n-i}\underline{x}(i)\underline{x}^H(i)\right)^{-1}\underline{x}(n), \text{ for } \lambda \in [0,1]. \tag{15}$$

Subtracting (1) from (14) and using the definition of the channel estimation error vector in (4) results in

$$\begin{aligned} \underline{\epsilon}(n+1 \mid n) = \alpha\underline{h}(n) + \underline{v}(n+1) \\ - \left(\hat{\underline{h}}(n-1) + \underline{\mathbf{K}}(n)\xi^*(n \mid n-1)\right). \end{aligned} \tag{16}$$

Substituting the expressions for $\xi(n \mid n-1)$ and $\underline{\epsilon}(n \mid n-1)$ to rearrange the terms results in

$$\begin{aligned} \underline{\epsilon}(n+1 \mid n) = (\alpha-1)\underline{h}(n) - \underline{\mathbf{K}}(n)w^*(n) + \underline{v}(n+1) \\ + \left(\mathbf{I} - \underline{\mathbf{K}}(n)\underline{x}^H(n)\right)\underline{\epsilon}(n \mid n-1). \end{aligned} \tag{17}$$

Equations (1) and (17) together form a coupled state space system that models the dynamics of the unknown system as well as that of the channel estimation error "system." Since their dynamics are coupled, their behavior in steady state may be analyzed by studying the dynamics of the equivalent extended state space system in steady state. Assuming that $\alpha = 1$ reduces (17) to the form given in [2] and [3]. Consider the augmented vector

$$\underline{g}(n) = \begin{bmatrix} \underline{h}(n) \\ \underline{\epsilon}(n \mid n-1) \end{bmatrix}. \tag{18}$$

Combining (1) and (17), the extended state space system governing the evolution of $\underline{g}(n)$ may be described as

$$\begin{aligned} \begin{bmatrix} \underline{h}(n+1) \\ \underline{\epsilon}(n+1 \mid n) \end{bmatrix} = \begin{pmatrix} \alpha\mathbf{I} & \mathbf{0} \\ (\alpha-1)\mathbf{I} & \mathbf{I} - \underline{\mathbf{K}}(n)\underline{x}^H(n) \end{pmatrix} \\ \times \begin{bmatrix} \underline{h}(n) \\ \underline{\epsilon}(n \mid n-1) \end{bmatrix} \\ + \begin{pmatrix} \underline{v}(n+1) \\ \underline{v}(n+1) - \underline{\mathbf{K}}(n)w^*(n) \end{pmatrix}. \end{aligned} \tag{19}$$

The system may be rewritten as

$$\underline{g}(n+1) = \mathbf{A}(n)\underline{g}(n) + \underline{B}(n) \tag{20}$$

where $\mathbf{A}(n)$ and $\underline{B}(n)$ are given by

$$\mathbf{A}(n) = \begin{pmatrix} \alpha\mathbf{I} & \mathbf{0} \\ (\alpha-1)\mathbf{I} & \mathbf{I} - \underline{\mathbf{K}}(n)\underline{x}^H(n) \end{pmatrix} \tag{21}$$

$$\underline{B}(n) = \begin{pmatrix} \underline{v}(n+1) \\ \underline{v}(n+1) - \underline{\mathbf{K}}(n)w^*(n) \end{pmatrix}. \tag{22}$$

Equation (20) is a stochastic difference equation in the augmented vector $\underline{g}(n)$, where the system matrix $\mathbf{A}(n)$ is a random quantity. From [3], $E\left[\underline{\mathbf{K}}(n)\underline{x}^H(n)\right] = (1-\lambda)\mathbf{I}$ so that $E\left[\mathbf{I} - \underline{\mathbf{K}}(n)\underline{x}^H(n)\right] = \lambda\mathbf{I}$. Thus

$$E[\mathbf{A}(n)] = \begin{pmatrix} \alpha\mathbf{I} & \mathbf{0} \\ (\alpha-1)\mathbf{I} & \lambda\mathbf{I} \end{pmatrix}. \tag{23}$$

The direct averaging method [13] may be used to study the convergence behavior of such a stochastic difference equation in an average sense. Based on this method, the solution of the stochastic difference (20), operating under the assumption of α and λ close to 1, is related to the solution of another stochastic difference equation whose system matrix is equal to the ensemble average given in (23).

More specifically, the stochastic difference equation in (20) may be replaced with another stochastic difference equation described as

$$\underline{g}(n+1) = E[\mathbf{A}(n)]\underline{g}(n) + \underline{B}(n). \tag{24}$$

Generally, the notation in (24) should be different from that in the original difference equation in (20). However, for the sake of notational convenience, this has not been done here. The correlation matrix $\mathbf{R}_{gg}$ is given by

$$\mathbf{R}_{gg} = E[\underline{g}(n+1)\underline{g}^H(n+1)] \tag{25}$$

where the expectation is taken with respect to the transmitted data vector $\underline{x}(n)$, the process noise vector $\underline{v}(n)$, and the observation noise $w(n)$. The correlation matrix $\mathbf{R}_{gg} = E[\underline{g}(n+1)\underline{g}^H(n+1)]$ in steady state is given by

$$\begin{aligned} \mathbf{R}_{gg} &= E[\mathbf{A}(n)]\mathbf{R}_{gg}E[\mathbf{A}^H(n)] + E[\underline{B}(n)\underline{B}^H(n)] \\ &= \begin{pmatrix} \alpha\mathbf{I} & \mathbf{0} \\ (\alpha-1)\mathbf{I} & \lambda\mathbf{I} \end{pmatrix}\mathbf{R}_{gg}\begin{pmatrix} \alpha\mathbf{I} & \mathbf{0} \\ (\alpha-1)\mathbf{I} & \lambda\mathbf{I} \end{pmatrix}^H \\ &\quad + E[\underline{B}(n)\underline{B}^H(n)] \end{aligned} \tag{26}$$

where $E[\mathbf{A}(n)]$ is substituted from (23). The correlation matrix $\mathbf{R}_{gg}$ of the augmented vector $\underline{g}(n)$ under steady state can also be written in terms of its submatrices as

$$\mathbf{R}_{gg} = \begin{pmatrix} \mathbf{R}_{hh} & \mathbf{R}_{h\epsilon} \\ \mathbf{R}_{\epsilon h} & \mathbf{R}_{\epsilon\epsilon} \end{pmatrix} \tag{27}$$

where $\mathbf{R}_{hh} = E[\underline{h}(n)\underline{h}^H(n)]$, $\mathbf{R}_{h\epsilon} = E[\underline{h}(n)\underline{\epsilon}^H(n)]$, $\mathbf{R}_{\epsilon h} = E[\underline{\epsilon}(n)\underline{h}^H(n)]$, and $\mathbf{R}_{\epsilon\epsilon} = E[\underline{\epsilon}(n)\underline{\epsilon}^H(n)]$ are the corresponding correlation submatrices. For notational convenience, $\underline{\epsilon}(n)$ will be used in place of $\underline{\epsilon}(n \mid n-1)$ to represent the channel estimation error vector given in (4). The correlation matrix $\mathbf{R}_{gg}$ of the augmented vector $\underline{g}(n)$ under steady state can be written as

$$\begin{aligned} \begin{pmatrix} \mathbf{R}_{hh} & \mathbf{R}_{h\epsilon} \\ \mathbf{R}_{\epsilon h} & \mathbf{R}_{\epsilon\epsilon} \end{pmatrix} &= E[\underline{B}(n)\underline{B}^H(n)] \\ &\quad + \begin{pmatrix} \alpha\mathbf{I} & \mathbf{0} \\ (\alpha-1)\mathbf{I} & \lambda\mathbf{I} \end{pmatrix}\begin{pmatrix} \mathbf{R}_{hh} & \mathbf{R}_{h\epsilon} \\ \mathbf{R}_{\epsilon h} & \mathbf{R}_{\epsilon\epsilon} \end{pmatrix} \\ &\quad \times \begin{pmatrix} \alpha\mathbf{I} & \mathbf{0} \\ (\alpha-1)\mathbf{I} & \lambda\mathbf{I} \end{pmatrix}^H. \end{aligned} \tag{28}$$

It is assumed that the additive observation noise $w(n)$ is zero mean and uncorrelated with both the transmitted data vector $\underline{x}(n)$ and the process noise vector $\underline{v}(n+1)$. Furthermore, the independence between the Gaussian additive noise $w(n)$ and the transmitted data vector $\underline{x}(n)$ is exploited in deriving the expression [3]

$$\begin{aligned} E[\underline{\mathbf{K}}(n)w^*(n)\underline{\mathbf{K}}^H(n)w(n)] &= E[\underline{\mathbf{K}}(n)\underline{\mathbf{K}}^H(n)]E[w^2(n)] \\ &= \sigma^2(1-\lambda)^2\mathbf{R}_{xx}^{-1} \end{aligned} \tag{29}$$

so that $E[\underline{B}(n)\underline{B}^H(n)]$ in (28) can be written as

$$E[\underline{B}(n)\underline{B}^H(n)] = \begin{pmatrix} \mathbf{R}_{vv} & \mathbf{R}_{vv} \\ \mathbf{R}_{vv} & \mathbf{R}_{vv} + \sigma^2(1-\lambda)^2\mathbf{R}_{xx}^{-1} \end{pmatrix}. \tag{30}$$

By expanding terms on the right-hand side of (26) and relating them to the corresponding terms of the resultant submatrices on the left-hand side of the expression, the following relationships are obtained:

$$\mathbf{R}_{hh} = \frac{1}{(1-|\alpha|^2)}\mathbf{R}_{vv} \tag{31}$$

$$\mathbf{R}_{h\epsilon} = \alpha(\alpha-1)^*\mathbf{R}_{hh} + \alpha\lambda\mathbf{R}_{h\epsilon} + \mathbf{R}_{vv} \tag{32}$$

$$\mathbf{R}_{\epsilon h} = \alpha^*(\alpha-1)\mathbf{R}_{hh} + \alpha^*\lambda\mathbf{R}_{\epsilon h} + \mathbf{R}_{vv} \tag{33}$$

$$\begin{aligned} \mathbf{R}_{\epsilon\epsilon} &= |\alpha-1|^2\mathbf{R}_{hh} + (\alpha-1)\lambda\mathbf{R}_{h\epsilon} + (\alpha-1)^*\lambda\mathbf{R}_{\epsilon h} \\ &\quad + \lambda^2\mathbf{R}_{\epsilon\epsilon} + \sigma^2(1-\lambda)^2\mathbf{R}_{xx}^{-1} + \mathbf{R}_{vv}. \end{aligned} \tag{34}$$

Combining (31) through (34) and eliminating common terms results in

$$\mathbf{R}_{h\epsilon} = \mathbf{R}_{\epsilon h}^H = \frac{(1-\alpha)}{(1-|\alpha|^2)(1-\alpha\lambda)}\mathbf{R}_{vv} \tag{35}$$

$$\begin{aligned} \mathbf{R}_{\epsilon\epsilon} &= \underbrace{\frac{(1-\alpha\lambda)(1-\alpha^*) + (1-\alpha^*\lambda)(1-\alpha)}{(1-|\alpha|^2)(1+\lambda)(1-\alpha\lambda)(1-\alpha^*\lambda)}\mathbf{R}_{vv}}_{\mathbf{D}_1} \\ &\quad + \underbrace{\frac{(1-\lambda)}{(1+\lambda)}\sigma^2\mathbf{R}_{xx}^{-1}}_{\mathbf{D}_2} \end{aligned} \tag{36}$$

where the underbraced terms in (36) represent the aforementioned tracking error variance $\mathcal{D}_1 = tr(\mathbf{D}_1)$ and the observation noise-induced error variance $\mathcal{D}_2 = tr(\mathbf{D}_2)$ terms, respectively. Recalling that $\underline{\epsilon}(n) = \underline{h}(n) - \underline{\hat{h}}(n-1)$, it is apparent by inspection that for nontrivial cases, $\mathbf{R}_{\epsilon\hat{h}} = \mathbf{R}_{\epsilon h} - \mathbf{R}_{\epsilon\epsilon} \neq \mathbf{0}$. By the orthogonality principle, this indicates that the channel estimate $\underline{\hat{h}}(n)$ is not optimal in a mean-square sense. An optimal post-filter may thus be constructed, as detailed in Section V, for

which expressions for $\mathbf{R}_{h\hat{h}}$ and $\mathbf{R}_{\hat{h}\hat{h}}$ will prove to be useful. These may be written in terms of $\mathbf{R}_{h\epsilon}$ and $\mathbf{R}_{hh}$ as

$$\mathbf{R}_{h\hat{h}} = \mathbf{R}_{\hat{h}h}^H = \mathbf{R}_{hh} - \mathbf{R}_{h\epsilon} \tag{37}$$

$$\mathbf{R}_{\hat{h}\hat{h}} = \mathbf{R}_{hh} - \mathbf{R}_{h\epsilon} - \mathbf{R}_{\epsilon h} + \mathbf{R}_{\epsilon\epsilon} \tag{38}$$

where the fact that $\mathbf{R}_{h\hat{h}} = \mathbf{R}_{\hat{h}h}^H$ follows from $\mathbf{R}_{h\epsilon} = \mathbf{R}_{\epsilon h}^H$. After substituting (31), (35), and (36) into (37) and (38), the following simplified expressions are obtained:

$$\mathbf{R}_{h\hat{h}} = \frac{\alpha(1-\lambda)}{(1-|\alpha|^2)(1-\alpha\lambda)}\mathbf{R}_{vv} \tag{39}$$

$$\mathbf{R}_{\hat{h}\hat{h}} = \frac{(1-|\alpha|^2\lambda^2)(1-\lambda)}{(1-|\alpha|^2)(1+\lambda)(1-\alpha\lambda)(1-\alpha^*\lambda)}\mathbf{R}_{vv} + \frac{(1-\lambda)}{(1+\lambda)}\sigma^2\mathbf{R}_{xx}^{-1}. \tag{40}$$

The steady state mean-square channel estimation error $tr(\mathbf{R}_{\epsilon\epsilon})$ expressed in (36) captures the inherent tradeoff between tracking error variance $\mathcal{D}_1$ and the observation noise-induced error variance $\mathcal{D}_2$ when a rate of adaptation λ is chosen. Furthermore, the channel estimate $\underline{\hat{h}}(n-1)$ and the channel impulse response vector $\underline{h}(n)$ are correlated, and their cross correlation matrix $\mathbf{R}_{h\hat{h}} = \mathbf{R}_{\hat{h}h}^H$ is given by (39).

V. Channel Subspace Post-Filtering

The EW-RLS channel estimate (13) may be treated as a noisy time series so that from (4)

$$\underline{\hat{h}}(n-1) = \underline{h}(n) - \underline{\epsilon}(n). \tag{41}$$

This noisy channel estimate can be post-filtered so that

$$\underline{\hat{h}}_p(n-1) = \mathbf{F}\underline{\hat{h}}(n-1) \tag{42}$$

where $\underline{\hat{h}}_p(n)$ is the post-filtered channel estimate, and $\mathbf{F}$ is the post-filter. The optimal filter $\mathbf{F}$ is given by

$$\mathbf{F} = \underset{\mathbf{F}}{\operatorname{argmin}} E\left[\|\underline{h}(n) - \mathbf{F}\underline{\hat{h}}(n-1)\|^2\right] = \mathbf{R}_{h\hat{h}}\mathbf{R}_{\hat{h}\hat{h}}^{-1}. \tag{43}$$

Let the eigenvalue decomposition of $\mathbf{R}_{vv}$ be given by $\mathbf{R}_{vv} = \mathbf{U}\Sigma_{vv}\mathbf{U}^H$, where the columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{R}_{vv}$, and Σ_{vv} is a diagonal matrix. Since from (31), $\mathbf{R}_{hh} = \mathbf{R}_{vv}/(1-|\alpha|^2)$, the eigenvectors of the process noise correlation matrix $\mathbf{R}_{vv}$ are also the eigenvectors of the channel correlation matrix $\mathbf{R}_{hh}$. In adition, given the assumption $\mathbf{R}_{xx} = \sigma_x^2\mathbf{I} = \sigma_x^2\mathbf{U}\mathbf{U}^H$, examination of (39) and (40) makes it apparent that the columns of $\mathbf{U}$ are also the eigenvectors of the matrices $\mathbf{R}_{h\hat{h}}$ and $\mathbf{R}_{\hat{h}\hat{h}}$. Hence, the post-filter $\mathbf{F}$ can be expressed in terms of these eigenvectors $\mathbf{U}$ as $\mathbf{F} = \mathbf{U}\Sigma_f\mathbf{U}^H$, where $\Sigma_f = \text{diag}(\sigma_{f1}, \sigma_{f2}, \ldots, \sigma_{fN})$. The post-filter $\mathbf{F}$ is referred to as a channel subspace filter (CSF) because its eigenvectors $\mathbf{U}$ are the same as those of the channel correlation matrix. The CSF coefficients, which are the diagonal elements of Σ_f, are given by

$$\sigma_{\text{fi}} = \frac{\alpha(1+\lambda)(1-\alpha^*\lambda){\sigma_{hi}}^2\sigma_x^2}{(1-|\alpha|^2\lambda^2)\sigma_{hi}^2\sigma_x^2 + (1-\alpha\lambda)(1-\alpha^*\lambda)\sigma^2} \tag{44}$$

where $i = 1, 2, 3, \ldots, N$, $\sigma_{hi}^2 = \sigma_{vi}^2/(1-|\alpha|^2)$ is the channel impulse response energy corresponding to the ith eigenvector, and $\sigma_{h1}^2 \geq \sigma_{h2}^2 \geq \cdots \geq \sigma_{hN}^2$. From (44), it may be noted that for real α, σ_{fi} is also real. For complex α, however, σ_{fi} is generally complex. The post-filter $\mathbf{F} = \mathbf{U}\Sigma_f\mathbf{U}^H$ may thus have complex eigenvalues and is, hence, non-Hermitian for complex α and Hermitian for real α. The CSF coefficients depend on the parameters α and λ, the observation noise energy σ^2, and the energy distribution in channel subspaces (subspace profile). This Wiener channel subspace filter weights the subspaces with higher energy more favorably than subspaces with lower energy and eliminates the observation noise error associated with any null subspaces. In general, the CSF coefficients shape the amplitude and phase of the subspaces, thereby compensating for the EW-RLS algorithm-induced amplitude and phase distortion.

The correlation matrix of the channel estimation error after post-filtering is given by

$$\mathbf{R}_{\epsilon f} = \mathbf{R}_{hh} - \mathbf{R}_{h\hat{h}}\mathbf{R}_{\hat{h}\hat{h}}^{-1}\mathbf{R}_{h\hat{h}}^H. \tag{45}$$

$\mathbf{R}_{\epsilon f}$ is diagonalized by the eigenvectors $\mathbf{U}$ so that $\mathbf{R}_{\epsilon f} = \mathbf{U}\Sigma_{\epsilon f}\mathbf{U}^H$, where the diagonal elements of the matrix $\Sigma_{\epsilon f}$ are

$$\sigma_{\epsilon fi}^2 = \sigma_{hi}^2 \times \left(1 - \frac{|\alpha|^2(1-\lambda^2)\sigma_{hi}^2\sigma_x^2}{(1-|\alpha|^2\lambda^2)\sigma_{hi}^2\sigma_x^2 + (1-\alpha\lambda)(1-\alpha^*\lambda)\sigma^2}\right) \tag{46}$$

where $i = 1, 2, \ldots, N$. The channel estimation error of the CSF algorithm is $E[\|\underline{\epsilon}_f(n)\|^2] = \sum_{i=1}^N \sigma_{\epsilon fi}^2$. From (36), the eigenvectors $\mathbf{U}$ also diagonalize the EW-RLS channel estimation error correlation matrix $\mathbf{R}_{\epsilon\epsilon}$. The equivalent diagonal matrix for the unfiltered EW-RLS estimate Σ_ϵ has diagonal elements

$$\sigma_{\epsilon i}^2 = \frac{(1-\alpha\lambda)(1-\alpha^*) + (1-\alpha^*\lambda)(1-\alpha)}{(1+\lambda)(1-\alpha\lambda)(1-\alpha^*\lambda)}\sigma_{hi}^2 + \frac{(1-\lambda)}{(1+\lambda)}\frac{\sigma^2}{\sigma_x^2} \tag{47}$$

where $i = 1, 2, \ldots, N$. The channel estimation error of the EW-RLS algorithm is $E[\|\underline{\epsilon}(n)\|^2] = \sum_{i=1}^N \sigma_{\epsilon i}^2$. Equations (42), (46), and (47) will be used in Section IX to compute the optimal theoretical channel estimation error and compare them with the simulation results.

VI. Suboptimal Channel Subspace Post-Filtering

The channel subspace post-filter is the solution to the MMSE problem formulated in (43). This post-filter $\mathbf{F}$ can be expressed in terms of the eigenvectors $\mathbf{U}$ and the channel subspace filtering coefficients σ_{fi} given in (44). The optimal post-filter is thus characterized by the channel subspace filtering coefficients, which, as discussed earlier, weight the subspaces with higher energy more heavily than subspaces with lower energy and compensate for amplitude and phase distortion. For a low-rank channel, a simplified but suboptimal filter may be considered.

Consider such a low-rank channel where $\mathbf{R}_{hh}$ has only r significant eigenvalues and $r \leq N$, i.e., significant energy is limited to an r dimensional subspace. A suboptimal post-filter,

referred to as the abrupt rank reduction (ARR) filter, may be constructed in such a manner that only the r channel subspaces with significant eigenvalues are retained so that

$$\mathbf{F}_a = \mathbf{U}\Sigma_{aa}\mathbf{U}^H \tag{48}$$

where $\Sigma_{aa} = \text{diag}(\underbrace{1,1,\ldots,1}_{r},\underbrace{0,0,\ldots,0}_{N-r})$. The ARR channel estimation error vector $\underline{\epsilon}_a(n) = \underline{h}(n) - \mathbf{F}_a\hat{\underline{h}}(n-1)$ has a correlation matrix $\mathbf{R}_{\epsilon a}$. This correlation matrix is diagonalized by the eigenvectors $\mathbf{U}$ so that $\mathbf{R}_{\epsilon a} = \mathbf{U}\Sigma_{\epsilon a}\mathbf{U}^H$, where the elements of the diagonal matrix $\Sigma_{\epsilon a}$ are

$$\sigma^2_{\epsilon ai} = \frac{(1-\alpha\lambda)(1-\alpha^*) + (1-\alpha^*\lambda)(1-\alpha)}{(1+\lambda)(1-\alpha\lambda)(1-\alpha^*\lambda)}\sigma^2_{hi} + \frac{(1-\lambda)}{(1+\lambda)}\frac{\sigma^2}{\sigma^2_x}, \quad \text{for } i = 1,\ldots,r \tag{49}$$

$$= \frac{1}{1-|\alpha|^2}{\sigma_{vi}}^2, \quad \text{for } i = r+1,\ldots,N. \tag{50}$$

The ARR channel estimation error is $E[\|\underline{\epsilon}_a(n)\|^2] = \sum_{i=1}^{N}\sigma^2_{\epsilon ai}$. Any improvement in channel estimation error using this filter is the result of the elimination of any observation noise-induced error variance associated with the null subspace of the channel impulse response correlation matrix. The ARR filter is simply a projection onto a reduced-dimensional subspace.

VII. Channel Order Mismatch

The criteria for assessing tracking performance were introduced in Section III, with the assumption that the number of taps in the adaptive filter used to model the unknown channel was the same as the number of taps in the true channel. When this assumption does not hold, there is a channel order mismatch. If the number of taps in the adaptive filter is greater than the number of taps in the true channel model, there in an overestimation of the channel order. Conversely, when the number of taps in the adaptive filter is less than the number of taps in the true channel model, there is an underestimation of the channel order. This section analyzes both these situations and shows that the CSF approach is applicable even when there is overestimation or underestimation of the channel order.

A. Overestimation of Channel Order

When there is an overestimation of the channel order, the channel model in (1) may be written as

$$\underline{h}_o(n+1) = \alpha\underline{h}_o(n) + \underline{v}_o(n+1) \tag{51}$$

where, if the number of taps in the assumed channel model is $N_o > N$, the augmented $N_o \times 1$ vectors $\underline{h}_o(n)$ and $\underline{v}_o(n)$ may be viewed, without loss of generality, as the vectors $\underline{h}(n)$ and $\underline{v}(n)$ appended by an $(N_o - N)\times 1$ column vector of zeros $\underline{0}_{N_o-N}$, respectively. In vector notation, $\underline{h}_o^T(n) = [\underline{h}^T(n)\,\underline{0}_{N_o-N}^T]$, and $\underline{v}_o^T(n) = [\underline{v}^T(n)\,\underline{0}_{N_o-N}^T]$. The channel output in (2) be alternately expressed as

$$y(n) = \underline{h}_o^H(n)\underline{x}_o(n) + w(n) \tag{52}$$

where, by inspection, $\underline{x}_o^T(n) = \left[\underline{x}^T(n)\,\underline{0}_{N_o-N}^T\right]$. The correlation matrix $\mathbf{R}_{v_ov_o}$ can be expressed as $\mathbf{R}_{v_ov_o} = \mathbf{U}_o\Sigma_{v_ov_o}\mathbf{U}_o^H$. The eigenvectors $\mathbf{U}_o$ of the augmented process correlation matrix $\mathbf{R}_{v_ov_o}$ may be expressed in terms of the eigenvectors $\mathbf{U}$ of the true process correlation matrix $\mathbf{R}_{vv}$ as

$$\mathbf{U}_o = \begin{pmatrix}\mathbf{U} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{N_o-N}\end{pmatrix}. \tag{53}$$

Similarly, it is easy to verify that the first N eigenvalues of $\mathbf{R}_{v_ov_o}$ are identical to those of $\mathbf{R}_{vv}$, whereas the remaining $N_o - N$ eigenvalues are identically equal to zero, i.e., $\sigma^2_{voi} = \sigma^2_{vi}$ for $i = 1,\ldots,N$ and $\sigma^2_{voi} = 0$ for $i = N+1,\ldots,N_o$. Applying the formulas from Section V yields

$$\sigma_{foi} = \frac{\alpha(1+\lambda)(1-\alpha^*\lambda)\sigma^2_{hoi}\sigma^2_x}{(1-|\alpha|^2\lambda^2)\sigma^2_{hoi}\sigma^2_x + (1-\alpha\lambda)(1-\alpha^*\lambda)\sigma^2} \tag{54}$$

$$\sigma^2_{\epsilon oi} = \frac{(1-\alpha\lambda)(1-\alpha^*) + (1-\alpha^*\lambda)(1-\alpha)}{(1+\lambda)(1-\alpha\lambda)(1-\alpha^*\lambda)}\sigma^2_{hoi} + \frac{(1-\lambda)}{(1+\lambda)}\frac{\sigma^2}{\sigma^2_x} \tag{55}$$

and

$$\sigma^2_{\epsilon foi} = \sigma^2_{hi} \times\left(1 - \frac{|\alpha|^2(1-\lambda^2)\sigma^2_{hoi}\sigma^2_x}{(1-|\alpha|^2\lambda^2)\sigma^2_{hoi}\sigma^2_x + (1-\alpha\lambda)(1-\alpha^*\lambda)\sigma^2}\right) \tag{56}$$

for $i = 1,2,\ldots,N_o$. The channel estimation error of the CSF algorithm is $E[\|\underline{\epsilon}_{fo}(n)\|^2] = \sum_{i=1}^{N_o}\sigma^2_{\epsilon foi}$. Note that $\sigma_{foi} = 0$ for $i = N+1,\ldots,N_o$. Furthermore, from (55) and (56), $\sigma^2_{\epsilon oi} = (1-\lambda)\sigma^2/(1+\lambda)\sigma^2_x$, whereas $\sigma^2_{\epsilon foi} = 0$ for $i = N+1,\ldots,N_o$. By comparing (47) and (56), it is apparent that overestimation of the channel order results in an increased EW-RLS channel estimation error because of the observation noise-induced error variance associated with the $N_o - N$ extra taps. Equation (46) and (56) imply, however, that the CSF post-filter, to within the limits of the direct averaging method, improves the EW-RLS channel estimator performance by eliminating the observation noise-induced error variance due to these extra taps and by compensating for the EW-RLS algorithm-introduced phase and magnitude distortion.

B. Underestimation of Channel Order

When the channel order has been underestimated, the channel model in (1) may be equivalently represented as

$$\begin{bmatrix}\underline{h}_u(n+1)\\ \underline{h}_r(n+1)\end{bmatrix} = \alpha\begin{bmatrix}\underline{h}_u(n)\\ \underline{h}_r(n)\end{bmatrix} + \begin{bmatrix}\underline{v}_u(n+1)\\ \underline{v}_r(n+1)\end{bmatrix} \tag{57}$$

where if the number of taps in the assumed channel model is $N_u < N$, then the modified $N_u \times 1$ underestimated vectors $\underline{h}_u(n)$ and $\underline{v}_u(n)$ in (57) may be viewed, without loss of generality, as the first N_u elements of the vectors $\underline{h}(n)$ and $\underline{v}(n)$, respectively. In vector notation, $\underline{h}^T(n) = \left[\underline{h}_u^T(n)\,\underline{h}_r^T(n)\right]$ and $\underline{v}^T(n) = \left[\underline{v}_u^T(n)\,\underline{v}_r^T(n)\right]$, where $\underline{h}_r(n)$ and $\underline{v}_r(n)$ are the residual $N - N_u$ taps of the true channel impulse response vector $\underline{h}(n)$ and the process noise vector $\underline{v}(n)$, respectively.

The process correlation matrix $\mathbf{R}_{vv}$ can be expressed in terms of its submatrices as

$$\mathbf{R}_{vv} = \begin{bmatrix} \mathbf{R}_{v_u v_u} & \mathbf{R}_{v_u v_r} \\ \mathbf{R}_{v_r v_u} & \mathbf{R}_{v_r v_r} \end{bmatrix} \tag{58}$$

where $\mathbf{R}_{v_u v_u} = E\left[\underline{v}_u(n)\underline{v}_u^H(n)\right]$, $\mathbf{R}_{v_u v_r} = \mathbf{R}_{v_r v_u}^H = E\left[\underline{v}_u(n)\underline{v}_r^H(n)\right]$, and $\mathbf{R}_{v_r v_r} = E\left[\underline{v}_r(n)\underline{v}_r^H(n)\right]$ are the corresponding correlation submatrices. The channel correlation matrix $\mathbf{R}_{hh}$ can also be expressed as

$$\mathbf{R}_{hh} = \begin{bmatrix} \mathbf{R}_{h_u h_u} & \mathbf{R}_{h_u h_r} \\ \mathbf{R}_{h_r h_u} & \mathbf{R}_{h_r h_r} \end{bmatrix} \tag{59}$$

where the correlation submatrices are analogously defined as in (58). Using the notational decomposition of the channel impulse response vector $\underline{h}(n)$ into the underestimated and residual vector introduced above, the channel output in (2) may be written as

$$\begin{aligned} y(n) &= \underline{h}_u^H(n)\underline{x}_u(n) + \underline{h}_r^H(n)\underline{x}_r(n) + w(n) \\ &= \underline{h}_u^H(n)\underline{x}_u(n) + w_u(n) \end{aligned} \tag{60}$$

where $w_u(n) = \underline{h}_r^H(n)\underline{x}_r(n) + w(n)$, and $\underline{x}_u(n)$ and $\underline{x}_r(n)$ are, respectively, the samples of the data vector $\underline{x}^T(n) = \left[\underline{x}_u^T(n)\ \underline{x}_r^T(n)\right]$ associated with the taps of the underestimated channel impulse response vector $\underline{h}_u(n)$ and its residual $\underline{h}_r(n)$, respectively.

Recalling that the additive white Gaussian observation noise $w(n)$ is uncorrelated with $\underline{x}(n)$ and $\underline{h}(n)$, it is clear that $w(n)$ is also uncorrelated with and independent of $\underline{x}_u(n)$, $\underline{x}_r(n)$, $\underline{h}_u(n)$, and $\underline{h}_r(n)$ as well. Furthermore, since the elements of the data vector $\underline{x}(n)$ are assumed to be independent and uncorrelated, $\left(\mathbf{R}_{xx} = \sigma_x^2\mathbf{I}\right)$, and it can readily be verified that $E\left[\underline{h}_u^H(n)\underline{x}_u(n)\underline{x}_r^H(n)\underline{h}_r(n)\right] = 0$. Hence, $w_u(n)$, as defined, is independent and uncorrelated with the term $\underline{h}_u^H(n)\underline{x}_u(n)$ in (60). It is therefore the *equivalent* additive white Gaussian observation noise for the underestimated channel response vector $\underline{h}_u(n)$, whose channel model is embedded in (57). Its variance $E\left[w_u^2(n)\right] = \sigma_u^2$ can be shown to be given by

$$\sigma_u^2 = \sigma^2 + \sigma_x^2 tr(\mathbf{R}_{h_r h_r}) \geq \sigma^2. \tag{61}$$

The channel estimation error vector $\underline{\epsilon}(n \mid n-1)$ is given by

$$\underline{\epsilon}(n \mid n-1) = \underbrace{\begin{bmatrix} \underline{h}_u(n) \\ \underline{h}_r(n) \end{bmatrix}}_{\underline{h}(n)} - \begin{bmatrix} \hat{\underline{h}}(n-1) \\ \underline{0}_{N-N_u} \end{bmatrix} = \begin{bmatrix} \underline{\epsilon}_u(n) \\ \underline{h}_r(n) \end{bmatrix} \tag{62}$$

where $\underline{\epsilon}_u(n) = \underline{h}_u(n) - \hat{\underline{h}}(n-1)$, and $\underline{h}_r(n)$ is the residual true channel impulse response vector defined earlier. The formula of Section V can be applied to the reduced order system described in (57) and (60), resulting in a filter that reduces the estimation error $\underline{\epsilon}_u(n)$ in (62). Channel order underestimation leads to an increase in the equivalent observation noise, reflected in (61), which adversely impacts the performance of the EW-RLS algorithm by increasing the observation noise induced error variance. While the CSF algorithm improves the performance of the EW-RLS algorithm by weighting the subspaces with higher energy more favorably than the subspaces with lower energy, this improvement is diminished by the constant offset term $tr(\mathbf{R}_{h_r h_r})$, due to the untracked taps of the true channel impulse response. The energy in these untracked taps, relative to the energy in the taps that are being tracked, ultimately determines the extent to which the CSF algorithm can improve the performance of the EW-RLS algorithm.

VIII. Adaptive Channel Subspace Filtering Algorithm

The CSF approach described above relies on explicit knowledge of the system parameters for the assumed channel model in (1). In a realistic scenario, even if the simplistic channel model in (1) were assumed to be valid, the parameters α, $\mathbf{U}$, Σ_{vv}, and σ^2 are unknown or, as in some instances, not directly observable. Merely the received data $y(n)$ and the transmitted data $\underline{x}(n)$ are known. From (11), it is clear that minimization of the unobservable mean-square channel estimation error also minimizes the observable mean-square prediction error. The CSF formulation in (43) expressed as a minimization of an unobservable cost function can hence be rewritten as a minimization of an observable cost function

$$\mathbf{F} = \underset{\mathbf{F}}{\operatorname{argmin}} E\left[\left\|y(n) - \hat{\underline{h}}^H(n)\mathbf{F}^H\underline{x}(n)\right\|^2\right]. \tag{63}$$

In a realistic scenario of unknown channel system parameters, the MMSE formulation in (63) is the basis for a deterministic least-squares problem expressed as

$$\hat{\mathbf{F}}(n) = \underset{\mathbf{F}}{\operatorname{argmin}} \sum_{k=1}^{n} \beta(n,k)\left\|y(k) - \hat{\underline{h}}^H(k)\mathbf{F}^H\underline{x}(k)\right\|^2 \tag{64}$$

where $\beta(n,k)$ is the weighting function, and $\hat{\mathbf{F}}(n)$ is the estimated post-filter. If $\beta(n,k) = \gamma^{n-k}$, then (64) takes on an EW-RLS form. For the remainder of this section and the discussions that follow, it will be assumed that $\beta(n,k) = 1$ only for $k = n - M + 1, \ldots, n$ and 0 elsewhere. Equation (64) thus takes on an SW-RLS form, which is written as

$$\hat{\mathbf{F}}(n) = \underset{\mathbf{F}}{\operatorname{argmin}} \sum_{k=n-M+1}^{n} \left\|y(k) - \hat{\underline{h}}^H(k)\mathbf{F}^H\underline{x}(k)\right\|^2 \tag{65}$$

where M is the length of the averaging window. This SW-RLS form was chosen for convenience.

Since $\mathbf{F}$ is an $N \times N$ matrix, even for moderate to large N, the formulation in (65) would require an unreasonably large number of samples of $y(n)$ for the solution $\hat{\mathbf{F}}(n)$ to converge. The number of parameters to be estimated can be reduced by exploiting the fact that for the channel model in (1), the channel estimate correlation matrix in (40) has the same eigenvectors as the CSF filter computed as $\mathbf{F} = \mathbf{U}\Sigma_f\mathbf{U}^H$. If the channel estimate time series were used to estimate

eigenvectors $\hat{\mathbf{U}}(n)$,[2] then the post-filter, which is written as $\hat{\mathbf{F}} = \hat{\mathbf{U}}(n)\hat{\Sigma}_f(n)\hat{\mathbf{U}}^H(n)$, could be computed using an estimate of the N CSF coefficients that form the diagonal matrix $\hat{\Sigma}_f(n)$. The modal decomposition of $\hat{\mathbf{F}}$ can be used to express the least-squares problem in (65) in terms of the required CSF coefficients such that we have (66), shown at the bottom of the page, where $\underline{\hat{u}}_i(n)$ is the ith column of the estimated eigenvector matrix $\hat{\mathbf{U}}(n)$ and the $\hat{\mathbf{F}} = \sum_{i=1}^{\hat{r}} \hat{\sigma}_{\text{fi}} \underline{\hat{u}}_i(n)\underline{\hat{u}}_i^H(n)$. Equation (66) is a least-squares problem that can be solved for a given M. In several communications applications, it may be necessary to minimize the n_0-step prediction error. A more general way of writing (66) is thus (67), shown at the bottom of the page. In the communications system application of interest here, if there are N_{fs} samples per transmitted data symbol, then $n_0 = N_{fs}$. Regardless of the value of n_0 used to calculate the optimal post-filter $\hat{\mathbf{F}}(n)$, the notation $\underline{\hat{h}}_p(n) = \hat{\mathbf{F}}(n)\underline{\hat{h}}(n)$ will be used.

The deterministic least-squares problem posed in (67) constitutes the basis for an adaptive CSF algorithm. However, the intended use of such an ACSF algorithm in low-rank channels and the sensitivity of subspace-based methods in such channels motivates a desire to introduce the channel rank as an additional parameter for the proposed ACSF algorithm. In low-rank channels, from (44), a number of the CSF coefficients will have an expected weight of zero. Thus, if the presumed rank of the channel is $\hat{r}$, then only the corresponding $\hat{r}$ CSF coefficients will have to be estimated, whereas the remaining $N - \hat{r}$ coefficients will be assigned zero weight. To fulfill this criterion, (67) can be rewritten as (68), shown at the bottom of the page. This equation can be solved for a given M and $\hat{r}$ and forms the basis for the proposed ACSF algorithm. Tables I and II summarize a proposed causal and noncausal ACSF algorithm.[3] Regardless of the causality of the solution, introducing the following variables allows (68) to be posed as traditional least-squares problem

$$\underline{\sigma}_f(n;\hat{r}) = [\sigma_{f1}(n;\hat{r}), \sigma_{f2}(n;\hat{r}) \ldots \sigma_{f\hat{r}}(n;\hat{r})]^T \quad (69)$$

$$\underline{z}(k+n_0;\hat{r}) = \begin{bmatrix} \underline{\hat{h}}^H(k)\underline{\hat{u}}_1(n) \\ \cdots \\ \underline{\hat{h}}^H(k)\underline{\hat{u}}_{\hat{r}}(n) \end{bmatrix} \cdot * \begin{bmatrix} \underline{\hat{u}}_1^H(n)\underline{x}(k+n_0) \\ \cdots \\ \underline{\hat{u}}_{\hat{r}}^H(n)\underline{x}(k+n_0) \end{bmatrix} \quad (70)$$

where it may be recalled that $\hat{r}$ is the conjectured rank of the channel and that the $(\cdot *)$ operation denotes an element-by-element multiplication of the adjacent vectors. Using these variables, the deterministic least-squares problem can be posed as (71), shown at the bottom of the page, whose conjectured rank $\hat{r}$ dependent solution is given by

$$\underline{\hat{\sigma}}_f(n;\hat{r}) = \mathbf{R}_{zz}^{-1}(n;\hat{r})\mathbf{R}_{zy}(n;\hat{r}) \quad (72)$$

where $\mathbf{R}_{zz}(n;\hat{r})$ and $\mathbf{R}_{zy}(n;\hat{r})$ are computed as

$$\mathbf{R}_{zz}(n;\hat{r}) = \sum_{k=n-M+1}^{n} \underline{z}(k+n_0;\hat{r})\underline{z}^H(k+n_0;\hat{r}) \quad (73)$$

$$\mathbf{R}_{zy}(n;\hat{r}) = \sum_{k=n-M+1}^{n} \underline{z}(k+n_0;\hat{r})y^*(k+n_0). \quad (74)$$

[2]This could be done by first computing the (unnormalized) channel estimate sample correlation matrix $\mathbf{R}_{\hat{h}\hat{h}}(n \mid n-1) = \sum_{k=1}^{n-1} \underline{\hat{h}}(k)\underline{\hat{h}}^H(k)$, then computing the eigenvalue (or singular value) decomposition, yielding eigenvectors $\hat{\mathbf{U}}(n)$.

[3]Following the notation introduced earlier, use of the time index n denotes functional time dependence, whereas its omission denotes functional time invariance. To maintain generality, the time index n will be retained for the remainder of this section.

$$\text{diag}(\hat{\Sigma}_f)(n) = \underset{\text{diag}(\Sigma_f)}{\arg\min} \sum_{k=n-M+1}^{n} \left\| y(k) - \sum_{i=1}^{N} \sigma_{\text{fi}}^* \underline{\hat{h}}^H(k)\underline{\hat{u}}_i(n)\underline{\hat{u}}_i^H(n)\underline{x}(k) \right\|^2 \quad (66)$$

$$\text{diag}(\hat{\Sigma}_f)(n) = \underset{\text{diag}(\Sigma_f)}{\arg\min} \sum_{k=n-M+1}^{n} \left\| y(k+n_0) - \sum_{i=1}^{N} \sigma_{\text{fi}}^* \underline{\hat{h}}^H(k)\underline{\hat{u}}_i(n)\underline{\hat{u}}_i^H(n)\underline{x}(k+n_0) \right\|^2. \quad (67)$$

$$\text{diag}(\hat{\Sigma}_f)(n;\hat{r}) = \underset{\text{diag}(\Sigma_f)}{\arg\min} \sum_{k=n-M+1}^{n} \left\| y(k+n_0) - \sum_{i=1}^{\hat{r}} \sigma_{\text{fi}}^* \underline{\hat{h}}^H(k)\underline{\hat{u}}_i(n)\underline{\hat{u}}_i^H(n)\underline{x}(k+n_0) \right\|^2. \quad (68)$$

$$\underline{\hat{\sigma}}_f(n;\hat{r}) = \underset{\underline{\sigma}_f(n;\hat{r})}{\arg\min} \sum_{k=n-M+1}^{n} \left\| y(k+n_0) - \underline{\sigma}_f^H(n;\hat{r})\underline{z}(k+n_0;\hat{r}) \right\|^2 \quad (71)$$

TABLE I
SUMMARY OF CAUSAL ACSF ALGORITHM

Initialization: Pick M, $n_0 = N_{fs}$, $\hat{r}$. Set $\hat{\mathbf{R}}_{\hat{h}\hat{h}} = \mathbf{0}$.

Operation: For every $n = 1, \ldots, L$,

1) Compute $\hat{\underline{h}}(n-1)$, the RLS channel estimate.

2) Update $\hat{\mathbf{R}}_{\hat{h}\hat{h}}(n) = \hat{\mathbf{R}}_{\hat{h}\hat{h}}(n) + \hat{\underline{h}}(n-1)\hat{\underline{h}}^H(n-1)$.

3) Compute/update eigenvectors $\hat{\mathbf{U}}(n)$ of $\hat{\mathbf{R}}_{\hat{h}\hat{h}}(n)$.

4) Compute $\mathbf{R}_{zz}(n;\hat{r})$ and $\mathbf{R}_{zy}(n;\hat{r})$ using $\hat{\mathbf{U}}(n)$ as in (70), (73) and (74).

5) Calculate ACSF coefficient vector $\hat{\underline{\sigma}}_f(n;\hat{r})$ using $\mathbf{U}(n)$, $\mathbf{R}_{zz}(n;\hat{r})$ and $\mathbf{R}_{zy}(n;\hat{r})$ as given in (72).

6) Compute post-filter $\hat{\mathbf{F}}(n;\hat{r}) = \sum_{i=1}^{\hat{r}} \sigma_{fi}(n;\hat{r})\hat{\underline{u}}_i(n)\hat{\underline{u}}_i^H(n)$.

7) Post-filter the RLS estimate $\hat{\underline{h}}_p(n;\hat{r}) = \hat{\mathbf{F}}(n;\hat{r})\hat{\underline{h}}(n)$.

TABLE II
SUMMARY OF NONCAUSAL ACSF ALGORITHM

Initialization: Pick $n_0 = N_{fs}$, $\hat{r}$.

Operation: Given RLS estimates $\hat{\underline{h}}(n-1)$ for $n = 1, \ldots, L$, set $M = L$ and,

1) Compute $\hat{\mathbf{R}}_{\hat{h}\hat{h}} = \sum_{n=1}^{L} \hat{\underline{h}}(n-1)\hat{\underline{h}}^H(n-1)$.

2) Compute eigenvectors $\hat{\mathbf{U}}$ of $\hat{\mathbf{R}}_{\hat{h}\hat{h}}$.

3) Compute $\mathbf{R}_{zz}(\hat{r})$ and $\mathbf{R}_{zy}(\hat{r})$ using $\hat{\mathbf{U}}$ as in (70), (73) and (74) setting $n = L$.

4) Calculate ACSF coefficient vector $\hat{\underline{\sigma}}_f(\hat{r})$ using $\mathbf{U}$, $\mathbf{R}_{zz}(\hat{r})$ and $\mathbf{R}_{zy}(\hat{r})$ as given in (72), (73) and (74).

5) Compute post-filter $\hat{\mathbf{F}}(\hat{r}) = \sum_{i=1}^{\hat{r}} \sigma_{fi}(\hat{r})\hat{\underline{u}}_i\hat{\underline{u}}_i^H$.

6) For $n = 1, \ldots, L$, post-filter the RLS estimate $\hat{\underline{h}}_p(n;\hat{r}) = \hat{\mathbf{F}}(\hat{r})\hat{\underline{h}}(n)$.

The ACSF coefficient vector $\hat{\underline{\sigma}}_f(n;\hat{r})$, as given in (72), can be calculated using an RLS algorithm using $\underline{z}(n)$ and $y(n)$ as its inputs. The estimated ACSF is computed as $\hat{\mathbf{F}}(n;\hat{r}) = \sum_{i=1}^{\hat{r}} \hat{\sigma}_{\text{fi}}(n;\hat{r})\hat{\underline{u}}_i(n)\hat{\underline{u}}_i^H(n)$. In experimental scenarios, the performance of the ACSF algorithm versus $\hat{r}$ can be evauated using the rank-$\hat{r}$-dependent ACSF prediction error metric $(1/N)\sum_{n=1}^{N} |\xi_f^{n_0}(n;\hat{r})|^2$. Here, $\xi_f^{n_0}(n;\hat{r}) = y(n+n_0) - \hat{\underline{h}}_p^H(n;\hat{r})\underline{x}(n+n_0)$, and $\hat{\underline{h}}_p(n;\hat{r})$ depends on the rank $\hat{r}$ used to compute the ACSF post-filter $\hat{\mathbf{F}}(n;\hat{r})$. For the channel model in (1), it is reasonable to expect this rank $\hat{r}^*$ to correspond to the true rank of the channel. The noncausal ACSF algorithm introduced is a benchmark for evaluating the performance of the practically implementable causal ACSF algorithm. The causal and noncausal ARR algorithms are similar to their ACSF variants, with the exception that the rank-$\hat{r}$-dependent filtering coefficients are, by definition, identically equal to 1, with the remaining filtering coefficients equal to 0.

IX. SIMULATION RESULTS

The models for the time-varying system described in (1) and (2) were used to generate data used in the simulations. Equiprobable ± 1 BPSK symbols were used to generate samples of the transmitted data vector $\underline{x}(n)$. For multitap simulations, the eigenvectors $\mathbf{U}$ of the correlation matrix $\mathbf{R}_{vv}$ were chosen randomly. The performance metrics for the simulations were calculated on a single channel realization using a large number of generated samples of received data $y(n)$. Only the samples of the channel estimate vector corresponding to the steady-state regime of the EW-RLS algorithm, i.e., after the transient phase of the algorithm was assumed to have been completed, were used for these performance evaluations. The ACSF algorithm, when used for comparison, refers to the noncausal variant described in Table II.

A. Single Tap Channel Simulation

A single tap channel was simulated with the following parameters: $\alpha = 0.9995\, e^{j2\pi/300}$, SNR $= 5$ dB, and $\sigma_v^2 = 1$. From (31), this corresponds to σ_h^2 of about 30 dB[4] and a noise variance σ^2 of about 25 dB. Channel estimates were generated using the EW-RLS algorithm and post-filtered using a scalar CSF filter that was computed using (44). Fig. 1 shows the agreement between the theoretical and simulated results for the optimal performance of the EW-RLS, CSF, and ACSF algorithms. The CSF algorithm improves performance by about 4 dB. The ACSF algorithm matches the performance of the CSF algorithm to within 0.1 dB. Fig. 2 shows the magnitude and phase of the optimal CSF coefficients for this example.

For the parameter α in this example, the channel dynamics as described using (1) primarily consist of phase rotations. The coherent averaging of the channel impulse response over the equivalent averaging window [14] of the EW-RLS algorithm introduces a phase and magnitude distortion in the channel estimate. For smaller values of λ, of up to about 0.975, the magnitude of the CSF coefficient is fairly close to 1, whereas

[4]Although dB is a logarithmic measure that is conventionally computed from dimensionless quantities, it has been used dually in this paper to express the magnitude of dimensionless quantities such as SNR and quantities such as variance and theoretical channel estimation error that are clearly not dimensionless. Its use in this manner will be obvious from the context; in such situations, the logarithmic dB measure computed as $10 * \log 10(\cdot)$ is relative to a unit 1 measure of the relevant quantity.

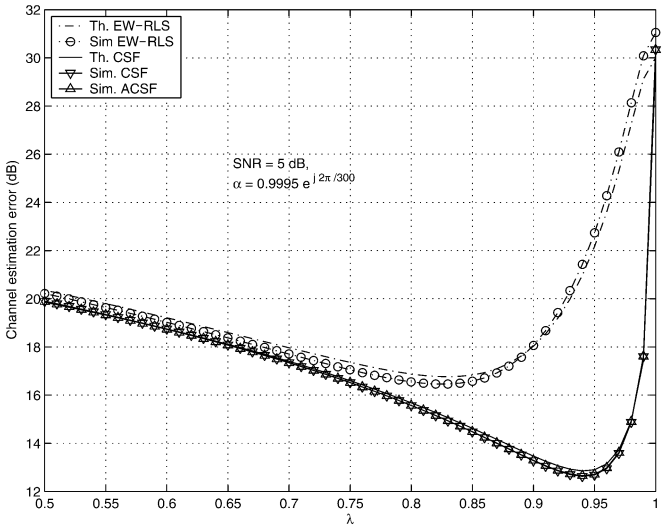


Fig. 1. Tracking performance of the EW-RLS, CSF, and ACSF algorithms on data simulating a one tap channel.

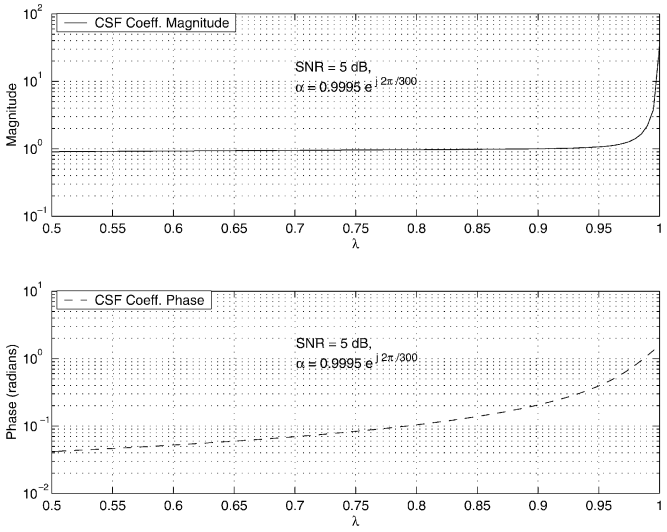


Fig. 2. Channel subspace filtering coefficients for simulated one-tap channel data.

its phase increases with λ. In this regime, the CSF algorithm primarily compensates for the phase distortion due to the coherent averaging of the EW-RLS algorithm. At higher values of λ, the CSF algorithm needs to account for both the phase distortion and the magnitude reduction in the channel estimate caused by the averaging of the rotating phasor over a longer window.

The minimum point in the performance curve for the CSF algorithm occurs at a greater value of λ than that of the EW-RLS algorithm, as seen in Fig. 1. At this higher value of λ, for the EW-RLS algorithm, the observation noise-induced error variance decreases but cannot offset the increased tracking error variance due to the greater phase and magnitude distortion introduced by the larger coherent averaging window [14]. The CSF algorithm is, however, able to compensate for this phase and magnitude distortion and, hence, improve the overall performance of the EW-RLS channel estimator. This constitutes the basis for the improved performance of the CSF algorithm in tracking channel dynamics composed of phase rotations. Section IX-B analyzes the performance of the CSF algorithm when there is a channel order mismatch.

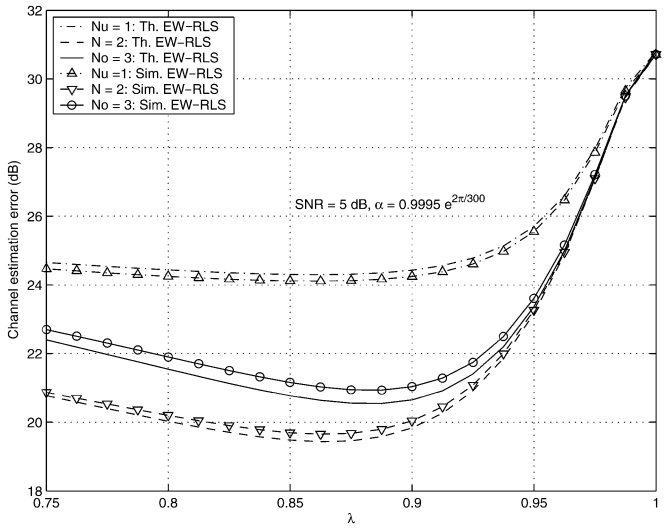
B. Channel Order Mismatch Simulation

With the previous example in mind, a two tap channel was simulated with the following parameters: $\alpha = 0.9995\ e^{j2\pi/300}$, SNR $= 5$ dB, and $\mathbf{R}_{vv} = [1\ 0.25; 0.25\ 0.2]$. The eigenvalues of the process correlation matrix are $\sigma_{v1}^2 = 1.0717$ and $\sigma_{v2}^2 = 0.1283$. As before, channel estimates were generated using the EW-RLS algorithm and post-filtered using a CSF filter. However, unlike before, three sets of channel estimates were generated corresponding to an underestimation of the channel order (with $N_u = 1$), perfect estimation of the channel order (with $N = 2$), and overestimation of the channel order (with $N_o = 3$).

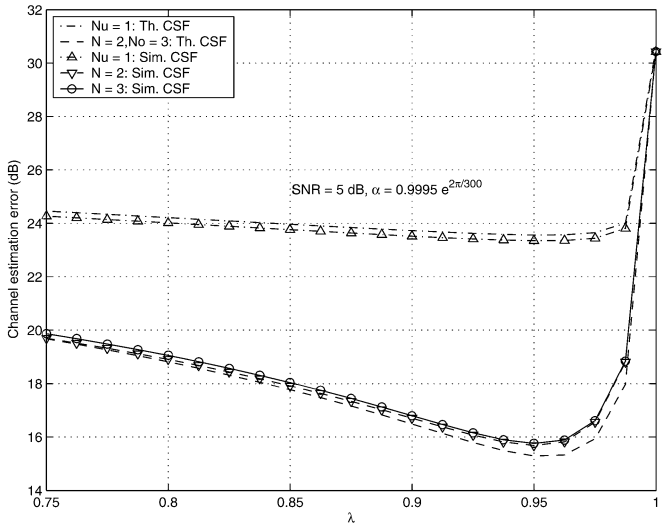
Fig. 3(a) shows the deterioration in EW-RLS tracking performance due to overestimation and underestimation of the channel order. Underestimation of the channel order results in a tracking performance that is 5 dB worse relative to the tracking performance of the $N = 2$ tap EW-RLS algorithm. Overestimation of the channel order, on the other hand, results in a 1.5-dB degradation in performance that is primarily due to the increased noise in the subspace associated with the extra tap. At higher values of λ, where the observation noise-induced error variance of the EW-RLS algorithm decreases, the performance of the $N_o = 3$ tap EW-RLS algorithm gets increasingly closer to the performance of the $N = 2$ tap EW-RLS algorithm. At values of λ very close to 1, where the observation noise-induced error variance term vanishes the $N_u = 1$ tap EW-RLS algorithm, the $N = 2$ tap EW-RLS algorithm and the $N_o = 3$ tap EW-RLS algorithm have the same channel estimation error of about -30.97 dB. This is equal to the energy $\mathrm{tr}(R_{hh})$ of the true channel impulse response $\underline{h}(n)$, indicating that all of the estimators are unable to track the fluctuations in the channel impulse response.

Fig. 3(b) shows the improvement in performance due to the CSF algorithm. There is also an excellent agreement between the CSF and ACSF performance that is omitted from Fig. 3(b) for clarity. The CSF algorithm applied to the $N_u = 1$ tap EW-RLS algorithm improves performance by about 0.6 dB. It can also be seen that CSF algorithm applied to the $n_0 = 3$ tap EW-RLS algorithm results in tracking performance very close to the CSF post-filtered $N = 2$ tap EW-RLS algorithm. This improvement in performance is about 4 dB relative to the $N = 2$ tap EW-RLS algorithm and about 5 dB relative to the $N_o = 3$ tap EW-RLS algorithm. This improved performance occurs at a higher value of λ, where the CSF algorithm is able to compensate for the increased tracking error variance due to the phase distortion induced by the EW-RLS algorithm.

It is insightful to compare the performance of the CSF algorithm on the underestimated channel to the one tap example presented earlier. As in the previous example, the minimum point in the performance curve for the CSF algorithm applied to the underestimated channel occurs at a larger value of λ than that of the EW-RLS algorithm. As discussed earlier, at this higher value of λ, for the EW-RLS algorithm, the observation noise-induced variance decreases but is offset by the increased tracking error variance due to the greater phase and magnitude distortion introduced by the larger coherent averaging window.



(a) Channel estimator performance vs λ for EW-RLS algorithm.



(b) Channel estimator performance vs λ for CSF algorithm.

Fig. 3. Comparison of impact of channel order mismatch on performance of EW-RLS and CSF algorithms.

The CSF algorithm compensates for this phase and magnitude distortion and improves the overall performance of the $N_u = 1$ tap EW-RLS algorithm. While this improvement in estimator performance relative to the tap in question is approximately 4 dB, from the previous example, the net improvement is more than offset by the large offset term due to the untracked tap. As a result, while the CSF algorithm still improves the performance of the EW-RLS algorithm, the magnitude of the improvement is diminished.

These simulations confirm that even when the channel order is underestimated or overestimated, the CSF algorithm is able to improve the performance of the EW-RLS algorithm. Section IX-C analyzes the performance of the CSF algorithm in tracking a multitap low-rank channel.

C. Multitap Channel Simulation

The time-varying channel impulse response vector $\underline{h}(n)$ was modeled using a 35 tap transversal vector $(N = 35)$. The eigenvalues of the driving process, i.e., the diagonal elements of Σ_{vv}, determine the energy associated with each of the channel subspace eigenvectors. The correlation matrix $\mathbf{R}_{vv}$ was modeled as having a rank of 5. The eigenvalues of the correlation matrix were normalized so that the energy in the driving process was unity, i.e., $\mathrm{tr}(\Sigma_{vv}) = 1$. The energy of the nonnull subspaces, i.e., the nonzero eigenvalues of Σ_{vv}, were assigned to be $[0.6056, 0.1831, 0.1217, 0.0608, 0.0303]$ such that, as described earlier, $tr(\Sigma_{vv}) = 1$. The time-varying system was evolved as (1) with the parameter $\alpha = 0.9995$. This resulted in the time-varying channel impulse response $\underline{h}(n)$ having an energy given by (31) as $\mathrm{tr}(\mathbf{R}_{hh}) = \mathrm{tr}(\mathbf{R}_{vv})/(1 - |\alpha|^2)$, which, for $\alpha = 0.9995$ corresponds to an an energy of about 30 dB. The output of the system $y(n)$ was generated using (2) with additive Gaussian observation noise $w(n)$ with a variance of σ^2, yielding an SNR of 5 dB. Fig. 4 compares the performance of these algorithms. The figure shows the theoretical values of channel estimation error computed with (47), (49), and (46) labeled as Th. EW-RLS, Th. ARR, and Th. CSF, respectively.

The rank 5 ARR filter, as constructed, exploits only the reduced dimensionality of the channel. The rank 5 CSF filter, on the other hand, implicitly exploits both the reduced dimensionality of the channel as well as the correlation between the channel estimate and the true channel impulse response. This is manifested in the improved performance of the CSF filter for all values of the tracking parameter λ, as evidenced in Fig. 4. It may be noted that best performance is achieved at a smaller value of λ. For this example, a smaller value of λ decreases the tracking error variance and increases the observation noise-induced error variance for the EW-RLS algorithm. When α is real, it can be shown from (44) that $\sigma_{\mathrm{fi}} \leq 1$ when only a magnitude correction is needed. The rank 5 CSF algorithm improves the channel estimation error by eliminating a major part of this increased observation noise-induced error variance associated with the null subspaces and by weighting the subspaces with higher energy more favorably than those with lower energy. The rank 5 ARR filter is also able to achieve comparable performance by eliminating this increased observation noise-induced error variance associated with the null subspaces. In this case, the performance of the rank 5 ARR filter is comparable to that of the CSF filter, suggesting that in some situations, the use of an ARR filter might provide the expected gain in computational efficiency without sacrificing too much algorithmic performance. Both the ARR and the CSF algorithms improve the channel estimator performance by about 3 dB.

Fig. 4 shows the agreement between predicted (Th. EW-RLS, Th. ARR, and Th. CSF) and realized channel estimation error (Sim. EW-RLS, Sim. ARR, and Sim. CSF) as well as the equivalence between the ACSF (Sim. ACSF) and the CSF (Sim. CSF) solutions. This figure also shows the mismatch between the predicted performance and the simulation performance for smaller values of λ, where the assumptions of the direct averaging method begin to break down.

Fig. 5 demonstrates that the performance of the ACSF algorithm is sensitive to eigenvector estimation errors. Here, for a given 50 000-point data window, the eigenvectors used to determine the ACSF solution are computed using an estimate of the channel estimate correlation matrix composed of 50 000, 500, 100, and 50 equally spaced channel estimates, respectively. It

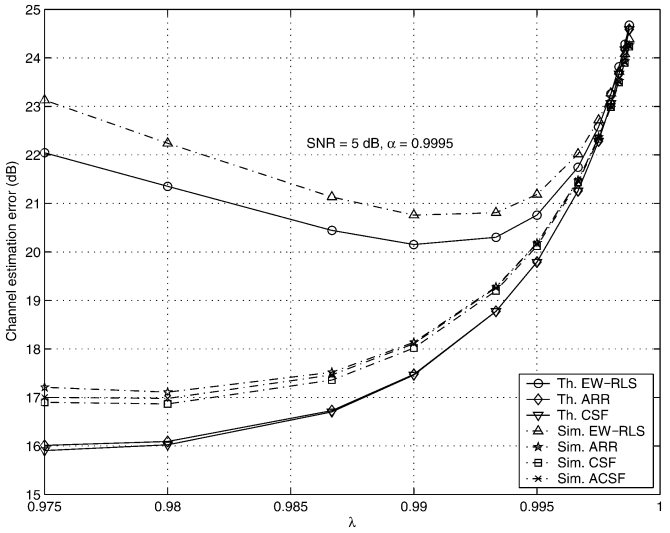


Fig. 4. Tracking performance of the EW-RLS, CSF, and ACSF algorithms on multitap channel simulated data. The curves from top to bottom at $\lambda = 0.975$ are Sim. EW-RLS, Th. EW-RLS, Sim. ARR, Sim. ACSF, Sim CSF, Th. ARR, and Th. CSF, respectively.

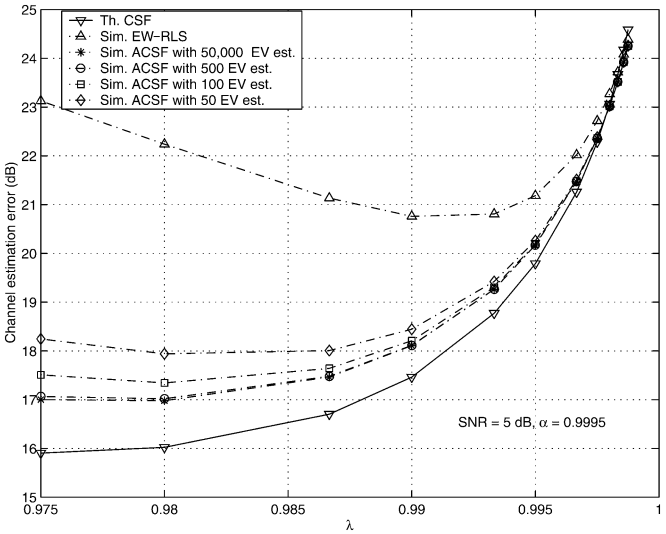


Fig. 5. Effect of eigenvector estimation errors on performance of ACSF algorithm using multitap channel simulated data.

can be seen that there is a deterioration in algorithm performance of approximately 1.2 dB. While a rigorous analysis of the performance of the ACSF algorithm is beyond the scope of this paper, simulation results such as Fig. 5 suggest that eigenvector estimation errors do indeed affect the performance of the ACSF algorithm.

X. Results on Experimental Data

The ACSF algorithm presented earlier was used to process experimental data collected in acoustic communication experiments [15]. This section presents the results of processing this data over a range of operating channel conditions. This will serve to demonstrate that even though a simplified theoretical channel model was initially used to formulate the algorithm, an ACSF algorithm is able to exploit the assumed channel subspace structure to improve the performance of the overall channel estimator.

The received signal with a carrier frequency of 2.25 kHz was sampled, brought to baseband, lowpass filtered, and downsampled to a rate of 2.5 kHz corresponding to two baseband samples per transmitted data symbol interval. The baseband samples were processed to compensate for any Doppler shift [16]. A 88 tap model was used to represent the time-varying baseband channel impulse response. Fig. 6 shows the magnitude of a sample baseband channel impulse response estimate generated using the EW-RLS algorithm with $\lambda = 0.955$. The ACSF coefficients were computed using $n_0 = 2$, whereas its performance was evaluated using $n_0 = 3$. Fig. 7(a) shows the improvement in performance obtained by using an ACSF approach. The rank 3 ACSF solution is within 0.1 dB of the best performing rank 75 ACSF solution. Fig. 7(b) shows the performance versus λ curve and demonstrates the 9-dB improvement in performance due to the rank 7 ACSF algorithm. It can also be seen that both the causal and noncausal processing demonstrate this improved performance. Fig. 7(a) shows that the tracking performance of the rank 3 ACSF filter is comparable with the rank 88 ACSF filter. The rank 7 ARR filter achieves a performance gain of

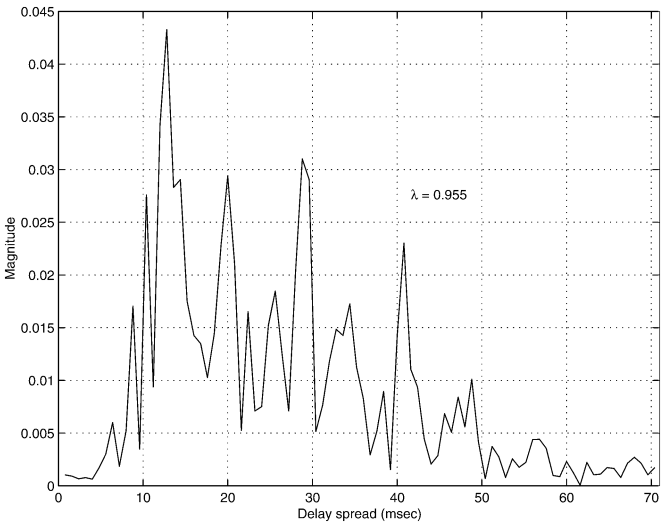
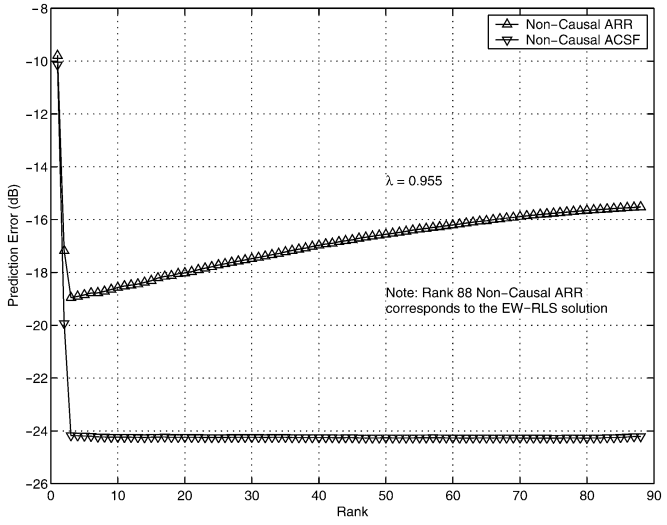


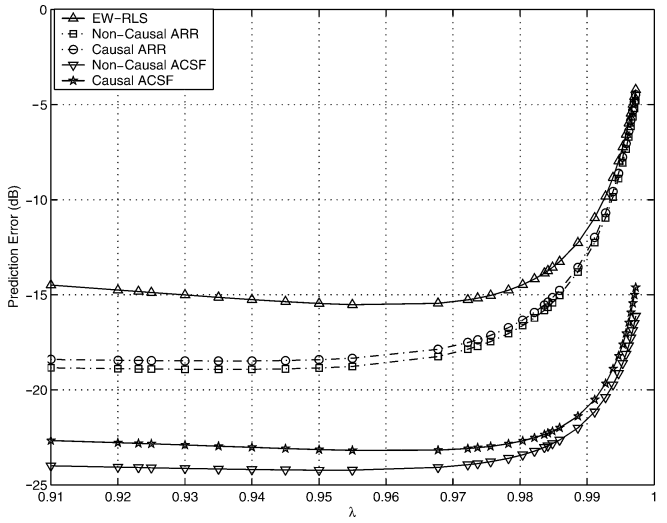
Fig. 6. Magnitude of a sample baseband channel impulse response estimate.

about 1.5 dB. This demonstrates that the ACSF coefficients used to compute the post-filter achieves the additional 7.5 dB of performance gain by compensating for the magnitude and phase distortion due to coherent averaging. Fig. 8 compares the magnitude and phase of the ACSF coefficients computed using an *a priori* assumed rank of 3 and 88, respectively. Since the rank 3 ACSF coefficients in Fig. 8 have a magnitude close to 1 and a phase of about -0.0873 rad (or about -5°), the performance gain is primarily due to a phase compensation, rather than a magnitude correction. This is similar to the single tap channel example presented in Section IX.

The improvement due to the ACSF approach is a result of exploiting both the reduced dimensionality of the channel subspace and the correlation between the channel estimate and the true channel impulse response. Hence, despite the relative computational efficiency of the ARR algorithm, its performance degradation relative to the ACSF algorithm is severe. The performance of the causal and the noncausal versions of the ACSF algorithm are within 1 dB of each other.



(a) Channel estimator performance vs Rank ($\hat{r}$).



(a) Channel estimator performance vs λ ($\hat{r} = 7$).

Fig. 7. Channel estimator performance on experimental data.

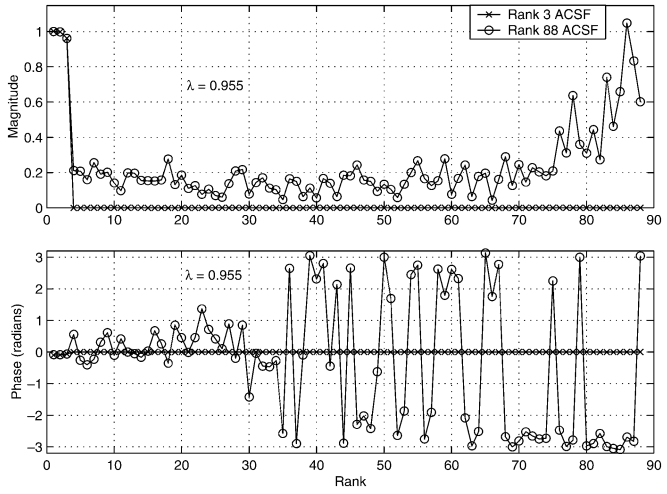


Fig. 8. Channel subspace filtering coefficients.

Improvement in estimator performance on additional channels is demonstrated in [12].

XI. Application to a Channel Estimate-Based DFE

Prior work [17] has shown that the performance of channel estimate-based DFEs is adversely affected by channel estimation errors. This suggests that improved channel estimation due to the CSF approach would lead to improved performance of the channel estimate-based DFE. The details of the channel estimate-based DFE are omitted here for the sake of brevity. The performance of the standard channel estimate based DFE [18] in Section XI-A is compared on experimental data, where the DFE feedforward and feedback taps are computed using either ACSF post-filtered or (unfiltered) EW-RLS channel estimates.

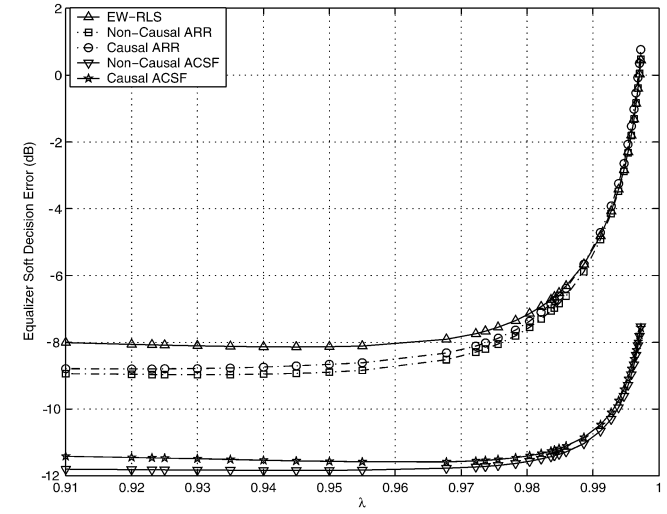
A. DFE Performance on Experimental Data

The channel estimate-based DFE algorithm was used to process the same experimental data used in Section X. Fig. 6 shows the magnitude of one of the sampled complex baseband channel impulse response estimates that was generated using this data. For the experimental results presented, the channel estimates were obtained using the EW-RLS algorithm. The channel estimate time series is postprocessed using the ACSF algorithm as described earlier. The postprocessed channel estimate is used to compute the coefficients of the feedback and the feedforward filter, as in [18].

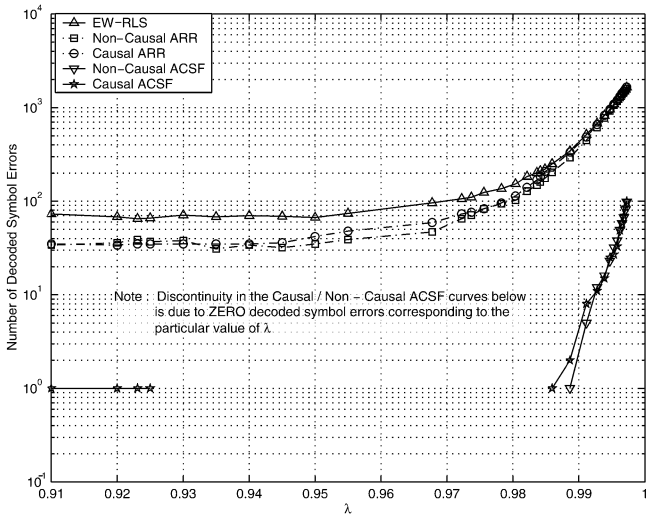
Fig. 9(a) compares the equalizer performance, i.e., the soft decision error in dB, of a channel estimate based DFE using post-filtered channel estimates and conventional EW-RLS channel estimates. Fig. 9(b) compares these channel estimate-based DFEs in terms of the number of incorrect decoded symbols. For this data set, 3200 quadrature phase-shift keying (QPSK) data symbols were transmitted at a rate of 1250 symbols/s and subsequently decoded. The performance of the ACSF post-filtered channel estimate-based DFE, configured using 40 feedback and 24 feedforward taps (ten anticausal and 14 causal taps, respectively), is demonstrably superior to that of the conventional EW-RLS channel estimate-based DFE.

XII. Conclusion

A channel subspace post-filtering algorithm has been presented that treats the least-squares channel estimate as a noisy time series and exploits the correlation structure of the channel subspace to improve the tracking performance of the EW-RLS tracking algorithm. This improvement in performance has been demonstrated both analytically and using simulation data. An adaptive CSF algorithm has been proposed that has been observed to closely match the performance of the CSF algorithm in the simulations presented. The simulation results suggest that the CSF approach is useful in channels with moderate to low SNRs, where there is either significant phase rotation, a channel order mismatch, or an inherent reduced dimensionality of the channel subspace. Although the channel model in (1) was fairly simplistic, experimental data demonstrates the applicability of the CSF approach on practical channels. A channel estimate-based DFE that uses these post-filtered channel estimates to determine the equalizer coefficients was shown to have improved performance on the same experimental data. Additional results in [10] and [11] show that the ACSF algorithm improves the performance of the SW-RLS channel estimation algorithm on the same experimental data presented here.



(a) Equalizer performance vs λ



(b) Number of incorrectly decoded symbols vs λ.

Fig. 9. Performance on experimental data.

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for their comments that helped improve this paper.

REFERENCES

[1] H. V. Poor and G. W. Wornell, Eds., *Wireless Communications: Signal Processing Perspectives*. Upper Saddle River, NJ: Prentice-Hall, 1998.

[2] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[3] E. Eleftheriou and D. D. Falconer, "Tracking properties and steady state performance of RLS adaptive filter algorithms," in *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, 1986, pp. 1097–1110.

[4] E. Moulines and P. Duhamel, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.

[5] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Trans. Inform. Theory*, vol. 43, pp. 499–511, Mar. 1997.

[6] L. L. Scharf and D. W. Tufts, "Rank reduction for modeling stationary signals," in *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, Mar. 1987, pp. 350–355.

[7] D. Kocic, D. Brady, and M. Stojanovic, "Sparse equalization for real-time digital underwater acoustic communications," in *Proc. OCEANS*, 1995, pp. 133–139.

[8] I. J. Fevrier, S. B. Gelfand, and M. P. Fitz, "Reduced complexity decision feedback equalization for multipath channels with large delay spreads," *IEEE Trans. Commun.*, vol. 47, pp. 927–937, June 1999.

[9] Y. F. Cheng and D. M. Etter, "Analysis of an adaptive technique for modeling sparse systems," *IEEE Trans. Signal Processing*, vol. 37, pp. 254–264, Feb. 1989.

[10] R. Nadakuditi and J. C. Preisig, "A channel subspace filtering approach to adaptive equalization of realistic acoustic channels," *J. Acoust. Soc. Amer.*, pt. 2, vol. 109, no. 5, pp. 2476–2476, May 2001.

[11] ——, "A channel subspace filtering approach to adaptive equalization of highly dynamic realistic channels," in *Proc. 35th Asilomar Conf. Signals Syst.*, vol. 2, Nov. 2001, pp. 1616–1623.

[12] R. Nadakuditi, "A channel subspace post-filtering approach to adaptive equalization," Master's thesis, Mass. Inst. Technol., Cambridge, MA, 2002.

[13] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes With Applications to Stochastic System Theory*. Cambridge, MA: MIT Press, 1984.

[14] J. G. Proakis, H. Lev-Ari, J. Lin, and F. Ling, "Optimal tracking of time-varying channels: A frequency domain approach for known and new algorithms," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 141–154, Jan. 1995.

[15] L. Freitag, M. Johnson, M. Stojanovic, D. Nagle, and J. Catipovic, "Survey and analysis of underwater acoustic channels for coherent communication in the medium-frequency band," in *Proc. OCEANS*, vol. 1, Providence, RI, Sept. 2000, pp. 123–128.

[16] M. Johnson, L. Freitag, and M. Stojanovic, "Improved doppler tracking and correction for underwater acoustic communication," in *Proc. ICASSP*, vol. 1, Munich, Germany, Apr. 1997, pp. 575–578.

[17] M. Stojanovic, "Analysis of the impact of channel estimation errors on the performance of a decision-feedback equalizer in fading multipath channels," *IEEE Trans. Commun.*, vol. 37, pp. 877–886, Feb. 1995.

[18] M. Stojanovic, L. Freitag, and M. Johnson, "Channel-estimate based adaptive equalization of underwater acoustic signals," in *Proc. Oceans*, Seattle, WA, 1999.

Raj Nadakuditi (S'01) received the B.S. degree in electrical engineering from Lafayette College, Easton, PA, in 1999 and the S.M. degree in electrical engineering and Computer Science from the Massachusetts Institute of Technology (M.I.T.), Cambridge, in 2001. He is currently working toward the Ph.D. degree at the Department Electrical Engineering and Computer Science at M.I.T. and in the Joint Program of Applied Ocean Science and Engineering at the Woods Hole Oceanographic Institution, Woods Hole, MA. His research interests are in the general area of adaptive signal processing, information theory, and random matrix theory with an emphasis on wireless communications applications.

Mr. Nadakuditi is a member of Eta Kappa Nu, Tau Beta Pi, Phi Beta Kappa, Pi Mu Epsilon, and Omicron Delta Epsilon. He was a recipient of the ONR Graduate Traineeship Award in 1999.

James C. Presig (M'91) received the B.S. degree in electrical engineering from the United States Coast Guard Academy, New London, CT, in 1980, the S.M. and E.E. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1988, and the Ph.D. degree in electrical and ocean engineering from the Massachusetts Institute of Technology/Woods Hole Oceanographic Institution (WHOI) Joint Program in Oceanography and Oceanographic Engineering, Woods Hole, MA, in 1992.

He was a Postdoctoral Investigator at WHOI from 1992 to 1994 and a Visiting Assistant Professor at Northeastern University, Boston, MA, from 1994 to 1997. Since July 1997, he has been on the scientific staff of the Department of Applied Ocean Physics and Engineering at WHOI. His research interests are in the areas of adaptive signal processing, underwater acoustic propagation modeling, underwater acoustic communications, and numerical optimization.

Dr. Presig is the recipient of the 1999 ONR Ocean Acoustics Young Faculty Award and is a member of the IEEE Sensor Array and Multichannel Signal Processing Technical Committee and the Acoustical Society of America Underwater Acoustics and Signal Processing Technical Committees. He is also an associate editor of the IEEE JOURNAL OF OCEANIC ENGINEERING.