5-2001

Finite-length MIMO adaptive equalization using canonical correlations

Aleksandar Dogandžić
Iowa State University, ald@iastate.edu

Arye Nehorai
University of Illinois at Chicago

FINITE-LENGTH MIMO ADAPTIVE EQUALIZATION USING CANONICAL CORRELATIONS

Aleksandar Dogandžić and *Arye Nehorai*

EECS Department
University of Illinois at Chicago, Chicago, Illinois 60607
email: {adogandz, nehorai}@eecs.uic.edu

ABSTRACT

We propose finite-length multi-input multi-output adaptive equalization methods for "smart" antenna arrays using the statistical theory of canonical correlations. We show that the proposed methods are related to maximum likelihood reduced-rank channel and noise estimation algorithms in unknown spatially correlated noise, and to several recently proposed adaptive equalization schemes.

1. INTRODUCTION

Multi-input multi-output (MIMO) channel equalization has recently attracted much attention due to recent popularity of antenna arrays applied at the receiver [1] and transmitter [2]. Adaptive and non-adaptive decision-feedback (DFE) MIMO equalizers have been recently proposed in [3] and [4], respectively (see also references therein). In this paper, we present methods for finite-length MIMO adaptive spatial and temporal equalization based on canonical correlation analysis [5], [6]. These methods are multivariate extensions of the adaptive equalization algorithms in [7], [8], [9], classical finite-length adaptive equalization in [10], and blind adaptive beamforming methods which use finite alphabet [11] and constant modulus [12] properties of the received signal. We show a relationship between the proposed methods and reduced-rank multivariate linear regression problem solved in [13].

First, in Section 2, we briefly review the maximum likelihood (ML) channel and noise estimation in [13]. Then, we describe the proposed adaptive equalization algorithm in Section 3, and discuss its application when training data is available (see Section 4) or not available (i.e. blind scenario, see Section 5).

2. REDUCED-RANK ML ESTIMATION

We review the ML estimation in [13] for a reduced-rank channel. As is [13], we model the received signal as a linear combination of basis functions, which includes various wireless channel models as special cases, see [1], [14]. However, unlike [13], where the measurements are real and basis functions are known, here we consider the measurement model with complex data and parametric basis functions. The proposed parametric basis function model is useful in blind equalization and symbol detection, i.e. when training data is not available, see Section 5.

Denote by $\boldsymbol{y}(t)$ an $m \times 1$ data vector received by an array of m antennas at time t and assume that we have collected N temporal data vectors. Then, we consider the following measurement model:

$$\boldsymbol{y}(t) = H\boldsymbol{\phi}(t,\boldsymbol{\eta}) + \boldsymbol{e}(t), \qquad t = 1, \ldots, N, \tag{2.1}$$

where H is an $m \times d$ channel response matrix of rank $r \leq \min(m,d)$, $\boldsymbol{\phi}(t,\boldsymbol{\eta})$ is a $d \times 1$ vector of basis functions, and $\boldsymbol{e}(t)$ is zero-mean Gaussian, temporally white and spatially correlated noise with unknown positive definite covariance Σ. The basis functions $\boldsymbol{\phi}(t,\boldsymbol{\eta})$ are chosen to describe the signal of interest, and $\boldsymbol{\eta}$ is a vector of unknown basis-function parameters, which may be the unknown symbols or phases of the received signal in constant-modulus scenario (see Section 5.1.1).

To present the ML estimates of H and Σ, it is useful to define $Y = [\boldsymbol{y}(1) \cdots \boldsymbol{y}(N)]$, $\Phi(\boldsymbol{\eta}) = [\boldsymbol{\phi}(1,\boldsymbol{\eta}) \cdots \boldsymbol{\phi}(N,\boldsymbol{\eta})]$, $\widehat{R}_{yy} = (1/N) \cdot YY^*$, $\widehat{R}_{\phi\phi} = (1/N) \cdot \Phi(\boldsymbol{\eta})\Phi(\boldsymbol{\eta})^*$, $\widehat{R}_{y\phi} = \widehat{R}_{\phi y}^* = (1/N) \cdot Y\Phi(\boldsymbol{\eta})^*$, and

$$\widehat{C}_{y\phi} = \widehat{R}_{yy}^{-1/2}\widehat{R}_{y\phi}\widehat{R}_{\phi\phi}^{-1/2}, \tag{2.2}$$

which is the estimated cross-correlation between the vectors $\widehat{R}_{yy}^{-1/2}\boldsymbol{y}(t)$ and $\widehat{R}_{\phi\phi}^{-1/2}\boldsymbol{\phi}(t)$ or the estimated *coherence matrix* between $\boldsymbol{y}(t)$ and $\boldsymbol{\phi}(t,\boldsymbol{\eta})$, see [5, Section VII]. Also, $A^{1/2}$ denotes a Hermitian square root of a Hermitian matrix A, and $A^{-1/2} = (A^{1/2})^{-1}$; this notation will be used throughout the paper. Note that $\widehat{R}_{y\phi}$ and $\widehat{R}_{\phi\phi}$ are functions of $\boldsymbol{\eta}$. To simplify the notation, we omit these dependencies throughout this paper. Consider now the singular value

This work was supported by the Air Force Office of Scientific Research under Grants F49620-97-1-0481 and F49620-99-1-0067, the National Science Foundation under Grant MIP-9615590, and the Office of Naval Research under Grant N00014-98-1-0542.

decomposition (SVD) of $\widehat{C}_{y\phi}$:

$$
\begin{aligned}
\widehat{C}_{y\phi} &= \widehat{U}\widehat{\Lambda}\widehat{V}^*, && (2.3a)\\
\widehat{U}^*\widehat{U} &= \widehat{U}\widehat{U}^* = I_m, \quad \widehat{V}^*\widehat{V} = \widehat{V}\widehat{V}^* = I_d, && (2.3b)\\
\widehat{\Lambda} &= \begin{cases} [\widehat{\Lambda}(m), 0], & m < d \\ [\widehat{\Lambda}(d), 0]^T, & m > d \end{cases}, && (2.3c)\\
\widehat{\Lambda}(m) &= \mathrm{diag}\{\widehat{\lambda}(1), \widehat{\lambda}(2), \dots, \widehat{\lambda}(m)\}, && (2.3d)
\end{aligned}
$$

where "$*$" denotes a conjugate transpose and $1 \geq \widehat{\lambda}(1) \geq \widehat{\lambda}(2) \geq \dots \geq \widehat{\lambda}(\min(m,d)) \geq 0$. Again, for notational simplicity we omit the dependence of the above quantities on $\boldsymbol{\eta}$.

We now present the ML estimate of the reduced-rank channel matrix H. First, we adopt the following notation: $\widehat{U}(r)$ and $\widehat{V}(r)$ are the matrices containing the first r columns of $\widehat{U}$ and $\widehat{V}$, respectively. For the model in (2.1) with known $\boldsymbol{\eta}$, the ML estimates of H and Σ are

$$
\begin{aligned}
\widehat{H}(\boldsymbol{\eta}) &= \widehat{R}_{yy}^{1/2}\widehat{U}(r)\widehat{\Lambda}(r)\widehat{V}(r)^*\widehat{R}_{\phi\phi}^{-1/2}, && (2.4a)\\
\widehat{\Sigma}(\boldsymbol{\eta}) &= \widehat{R}_{yy} - \widehat{R}_{yy}^{1/2}\widehat{U}(r)\widehat{\Lambda}^2(r)\widehat{U}(r)^*\widehat{R}_{yy}^{1/2}, && (2.4b)
\end{aligned}
$$

see [13], [14]. If $\boldsymbol{\eta}$ is unknown, its ML estimate $\widehat{\boldsymbol{\eta}}$ is obtained by maximizing the concentrated likelihood

$$
l(\boldsymbol{\eta}) = \prod_{i=1}^{r} \frac{1}{1 - \widehat{\lambda}^2(i)}, \qquad (2.5)
$$

see [14, eq. (4.1)] and [13, eq. (35)]. To find the ML estimates of H and Σ, replace $\boldsymbol{\eta}$ in (2.4) by $\widehat{\boldsymbol{\eta}}$.

In the following, we propose an alternative criterion, which is maximized for the same estimate of $\boldsymbol{\eta}$ as the concentrated likelihood function (2.5). This criterion is motivated by the MIMO equalization scheme in Figure 1.

3. MIMO ADAPTIVE EQUALIZATION

We analyze the adaptive MIMO equalization scheme depicted in Figure 1. We wish to find an $r \times m$ beamforming matrix B and an $r \times d$ basis-function filtering matrix W that minimize the error between the beamformed data and filtered basis functions $\boldsymbol{\epsilon}(t,\boldsymbol{\eta}) = B\boldsymbol{y}(t) - W\boldsymbol{\phi}(t,\boldsymbol{\eta})$ in the mean-square sense. Define the error matrix as $\mathcal{E}(\boldsymbol{\eta}) = [\boldsymbol{\epsilon}(1,\boldsymbol{\eta}) \cdots \boldsymbol{\epsilon}(N,\boldsymbol{\eta})]$. In the following, we show that this problem is related to canonical correlation analysis.

We propose to estimate B, W, and $\boldsymbol{\eta}$ by maximizing the inverse of estimated geometric mean-squared error of $\boldsymbol{\epsilon}(t,\boldsymbol{\eta})$:

$$
l(\boldsymbol{\eta}, B, W) = \frac{1}{|(1/N)\cdot\mathcal{E}(\boldsymbol{\eta})\cdot\mathcal{E}(\boldsymbol{\eta})^*|} \qquad (3.1)
$$

subject to the normalizing constraint

$$
B\widehat{R}_{yy}B^* = I_r. \qquad (3.2)
$$

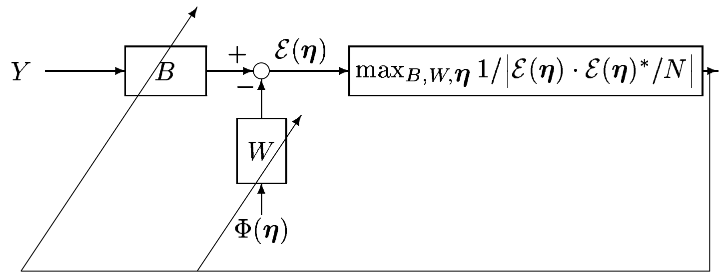

Fig. 1. Proposed MIMO adaptive equalization scheme.

Here, $|\cdot|$ denotes the determinant. The normalizing constraint prevents trivial solution (in which B and W equal zero), and imposes the estimated beamformed signals BY to be uncorrelated.

It can be shown that, under $B\widehat{R}_{yy}B^* = I_r$, all the eigenvalues of $\mathcal{E}(\boldsymbol{\eta})\cdot\mathcal{E}(\boldsymbol{\eta})^*/N$ are simultaneously minimized for

$$
\begin{aligned}
\widehat{B}(\boldsymbol{\eta}) &= [\widehat{\boldsymbol{b}}_1(\boldsymbol{\eta}), \widehat{\boldsymbol{b}}_2(\boldsymbol{\eta}) \cdots \widehat{\boldsymbol{b}}_r(\boldsymbol{\eta})]^*\\
&= \widehat{U}(r)^*\widehat{R}_{yy}^{-1/2}, && (3.3a)\\
\widehat{W}(\boldsymbol{\eta}) &= [\widehat{\boldsymbol{w}}_1(\boldsymbol{\eta}), \widehat{\boldsymbol{w}}_2(\boldsymbol{\eta}) \cdots \widehat{\boldsymbol{w}}_r(\boldsymbol{\eta})]^*\\
&= \widehat{B}(\boldsymbol{\eta})\widehat{R}_{y\phi}\widehat{R}_{\phi\phi}^{-1} = \widehat{\Lambda}(r)\widehat{V}(r)^*\widehat{R}_{\phi\phi}^{-1/2}; && (3.3b)
\end{aligned}
$$

where $\widehat{U}(r)$ and $\widehat{V}(r)$ are the matrices containing the first r columns of $\widehat{U}$ and $\widehat{V}$, respectively (see [15]). Therefore, $B = \widehat{B}(\boldsymbol{\eta})$ and $W = \widehat{W}(\boldsymbol{\eta})$ maximize (3.1), yielding $l(\boldsymbol{\eta}, \widehat{B}(\boldsymbol{\eta}), \widehat{W}(\boldsymbol{\eta})) = l(\boldsymbol{\eta})$, which is the concentrated likelihood function in (2.5). Note that $\widehat{B}(\boldsymbol{\eta}) \cdot \widehat{H}(\boldsymbol{\eta}) = \widehat{W}(\boldsymbol{\eta})$, where $\widehat{H}(\boldsymbol{\eta})$ is the ML estimate of the channel in (2.4a). Also, $\widehat{\boldsymbol{y}}_{\mathrm{c}}(t,\boldsymbol{\eta}) = [\widehat{y}_{\mathrm{c},1}(t,\boldsymbol{\eta}), \widehat{y}_{\mathrm{c},2}(t,\boldsymbol{\eta}), \dots, \widehat{y}_{\mathrm{c},r}(t,\boldsymbol{\eta})]^T = \widehat{B}(\boldsymbol{\eta})\boldsymbol{y}(t)$ and $\widehat{\boldsymbol{\phi}}_{\mathrm{c}}(t,\boldsymbol{\eta}) = [\widehat{\phi}_{\mathrm{c},1}(t,\boldsymbol{\eta}), \widehat{\phi}_{\mathrm{c},2}(t,\boldsymbol{\eta}), \dots, \widehat{\phi}_{\mathrm{c},r}(t,\boldsymbol{\eta})]^T = \widehat{W}(\boldsymbol{\eta})\boldsymbol{\phi}(t,\boldsymbol{\eta})$ can be viewed as estimated *canonical coordinates* of the data and basis functions, respectively, whereas $\widehat{\lambda}(i)$ are the estimated *canonical correlations*, see [5]. This allows for an elegant interpretation of the proposed equalization scheme in the context of canonical correlation analysis, see e.g. [6, ch. 12]. The first estimated canonical coordinates $\widehat{y}_{\mathrm{c},1}(t,\boldsymbol{\eta}) = \widehat{\boldsymbol{b}}_1(\boldsymbol{\eta})^*\boldsymbol{y}(t)$ and $\widehat{\phi}_{\mathrm{c},1}(t,\boldsymbol{\eta}) = \widehat{\boldsymbol{w}}_1(\boldsymbol{\eta})^*\boldsymbol{\phi}(t,\boldsymbol{\eta})$ have the largest estimated correlation $\widehat{\lambda}(1)$ among all possible linear combinations of $\boldsymbol{y}(t)$ and $\boldsymbol{\phi}(t,\boldsymbol{\eta})$. Further, $\widehat{y}_{\mathrm{c},2}(t,\boldsymbol{\eta}) = \widehat{\boldsymbol{b}}_2(\boldsymbol{\eta})^*\boldsymbol{y}(t)$ and $\widehat{\phi}_{\mathrm{c},2}(t,\boldsymbol{\eta}) = \widehat{\boldsymbol{w}}_2(\boldsymbol{\eta})^*\boldsymbol{\phi}(t,\boldsymbol{\eta})$ have the largest estimated correlation $\widehat{\lambda}(2)$ among all possible linear combinations of $\boldsymbol{y}(t)$ and $\boldsymbol{\phi}(t,\boldsymbol{\eta})$ that are uncorrelated with $\widehat{y}_{\mathrm{c},1}(t,\boldsymbol{\eta})$ and $\widehat{\phi}_{\mathrm{c},1}(t,\boldsymbol{\eta})$, and so on.

For a single sensor with $\boldsymbol{y}(t) = [y(t), y(t-1), \dots, y(t-m+1)]^T$, rank $r = 1$, and basis functions chosen to model the multipath effect by uniformly discretizing the time-delay spread (i.e. $\boldsymbol{\phi}(t,\boldsymbol{\eta}) = [s(t), s(t-1), \dots, s(t-d+1)]^T$), the equalizers in (3.3) become row vectors, i.e. $\widehat{B}(\boldsymbol{\eta}) = \widehat{\boldsymbol{b}}(\boldsymbol{\eta})^*$ and $\widehat{W}(\boldsymbol{\eta}) = \widehat{\boldsymbol{w}}(\boldsymbol{\eta})^*$, where $\widehat{\boldsymbol{b}}(\boldsymbol{\eta})$ can be interpreted as a feedforward filter, which shapes the channel to the desired impulse response $\widehat{\boldsymbol{w}}(\boldsymbol{\eta})$; this is a classical adaptive equalization scheme in [10].

For unknown $\boldsymbol{\eta}$, the maximization of (2.5) can be performed by iteration, as described in Section 5.

4. MIMO EQUALIZATION AND SYMBOL DETECTION USING TRAINING DATA

If training data is available, we can separate the equalization and detection tasks as follows: use training data to estimate B and W [see (3.3)] and then detect the unknown sequence by applying *metric combining* (MC) [16, sec. IV.A] to the equalized data and basis functions. We show that this procedure is equivalent to estimating Σ and H from the training data [using (2.4)] and detecting the unknown sequence using *interference rejection combining* (IRC) [16, eq. (8)] (see also [17]).

Let $Y_{\rm K} = [\boldsymbol{y}_{\rm K}(1) \cdots \boldsymbol{y}_{\rm K}(N_{\rm K})]$ be the data set containing the known (or training) sequence, described by known basis functions $\Phi_{\rm K} = [\boldsymbol{\phi}_{\rm K}(1) \cdots \boldsymbol{\phi}_{\rm K}(N_{\rm K})]$. Further, by analogy with (2.2) and (2.3), define $\widehat{C}_{{\rm K}y\phi} = \widehat{R}_{{\rm K}yy}^{-1/2} \widehat{R}_{{\rm K}y\phi} \widehat{R}_{{\rm K}\phi\phi}^{-1/2} = \widehat{U}_{\rm K} \widehat{\Lambda}_{\rm K} \widehat{V}_{\rm K}^*$, where $\widehat{R}_{{\rm K}yy} = Y_{\rm K} Y_{\rm K}{}^*/N_{\rm K}$, $\widehat{R}_{{\rm K}\phi\phi} = \Phi_{\rm K} \Phi_{\rm K}{}^*/N_{\rm K}$, and $\widehat{R}_{{\rm K}y\phi} = Y_{\rm K} \Phi_{\rm K}{}^*/N_{\rm K}$. Define also the singular values of $\widehat{C}_{{\rm K}y\phi}$: $1 \geq \widehat{\lambda}_{\rm K}(1) \geq \widehat{\lambda}_{\rm K}(2) \geq \ldots \geq \widehat{\lambda}_{\rm K}(\min(m,d)) \geq 0$ and $\widehat{\Lambda}_{\rm K}(r) = \mathrm{diag}\{\widehat{\lambda}_{\rm K}(1), \widehat{\lambda}_{\rm K}(2), \ldots, \widehat{\lambda}_{\rm K}(r)\}$. Then, from (3.3), the estimates of B and W based on the training data are $\widehat{B}_{\rm K} = \widehat{U}_{\rm K}(r)^* \widehat{R}_{{\rm K}yy}^{-1/2}$ and $\widehat{W}_{\rm K} = \widehat{\Lambda}_{\rm K}(r) \widehat{V}_{\rm K}(r)^* \widehat{R}_{{\rm K}\phi\phi}^{-1/2}$, and the channel and noise estimates follow from (2.4) as $\widehat{H}_{\rm K} = \widehat{R}_{{\rm K}yy}^{1/2} \widehat{U}_{\rm K}(r) \widehat{\Lambda}_{\rm K}(r) \widehat{V}_{\rm K}(r)^* \widehat{R}_{{\rm K}\phi\phi}^{-1/2}$ and $\widehat{\Sigma}_{\rm K} = \widehat{R}_{{\rm K}yy} - \widehat{R}_{{\rm K}yy}^{1/2} \widehat{U}_{\rm K}(r) \widehat{\Lambda}_{\rm K}^2(r) \widehat{U}_{\rm K}(r)^* \widehat{R}_{{\rm K}yy}^{1/2}$. Note also that $(1/N_{\rm K})[\widehat{B}_{\rm K} Y_{\rm K} - \widehat{W}_{\rm K} \Phi_{\rm K}] \cdot [\widehat{B}_{\rm K} Y_{\rm K} - \widehat{W}_{\rm K} \Phi_{\rm K}]^* = I_r - \widehat{\Lambda}_{\rm K}^2(r)$.

Now, apply the beamformer $\widehat{B}_{\rm K}$ to the data Y containing the unknown sequence and and the basis-function filter $\widehat{W}_{\rm K}$ to the hypothesized basis functions $\boldsymbol{\phi}(t,\boldsymbol{\eta})$, yielding $\widehat{\boldsymbol{y}}_{\rm Kc}(t) = [\widehat{y}_{{\rm Kc},1}(t), \widehat{y}_{{\rm Kc},2}(t), \ldots, \widehat{y}_{{\rm Kc},r}(t)]^T = \widehat{B}_{\rm K} \boldsymbol{y}(t)$ and $\widehat{\boldsymbol{\phi}}_{\rm Kc}(t,\boldsymbol{\eta}) = [\widehat{\phi}_{{\rm Kc},1}(t,\boldsymbol{\eta}), \widehat{\phi}_{{\rm Kc},2}(t,\boldsymbol{\eta}), \ldots, \widehat{\phi}_{{\rm Kc},r}(t,\boldsymbol{\eta})]^T = \widehat{W}_{\rm K} \boldsymbol{\phi}(t,\boldsymbol{\eta})$. To find the unknown sequence $\boldsymbol{\eta}$, we use the metric combining of the estimated canonical coordinates $\widehat{\boldsymbol{y}}_{\rm Kc}(t)$ and $\widehat{\boldsymbol{\phi}}_{\rm Kc}(t,\boldsymbol{\eta})$, i.e. minimize the following cost function:

$$C_{\rm Kc}^{\rm MC}(\boldsymbol{\eta}) = \sum_{i=1}^{r} \; [1 - \widehat{\lambda}_{\rm K}^2(i)]^{-1} \; \sum_{t=1}^{N} \left| \widehat{y}_{{\rm Kc},i}(t) - \widehat{\phi}_{{\rm Kc},i}(t,\boldsymbol{\eta}) \right|^2,$$

see [15]. Interestingly, the above cost function is equal to the following IRC cost function:

$$C_{\rm K}^{\rm IRC}(\boldsymbol{\eta}) = \sum_{t=1}^{N} [\boldsymbol{y}(t) - \widehat{H}_{\rm K} \boldsymbol{\phi}(t,\boldsymbol{\eta})]^* \; \widehat{\Sigma}_{\rm K}^{-1} [\boldsymbol{y}(t) - \widehat{H}_{\rm K} \boldsymbol{\phi}(t,\boldsymbol{\eta})],$$

i.e. $C_{\rm Kc}^{\rm MC}(\boldsymbol{\eta}) = C_{\rm K}^{\rm IRC}(\boldsymbol{\eta})$, see [15]. If $\boldsymbol{\eta}$ is the unknown sequence to be detected, maximum likelihood sequence estimation (MLSE) can be used to minimize the above cost functions with respect to $\boldsymbol{\eta}$, along the lines of [16].

For rank-1 channels (i.e. $r = 1$) and basis functions chosen to model the multipath effect by uniformly discretizing the time-delay spread (i.e. $\boldsymbol{\phi}_{\rm K}(t) = [s_{\rm K}(t), s_{\rm K}(t-1), \ldots, s_{\rm K}(t-d+1)]^T$ and similarly for $\boldsymbol{\phi}(t,\boldsymbol{\eta})$), the above equalization and detection algorithms become very similar those in [7], [8], [9] (where the differences arise because the normalizing constraints in [7], [8], [9] differ from (3.2)).

5. BLIND MIMO EQUALIZATION

Two iterative procedures for blind MIMO equalization and symbol detection follow from the results of Sections 2 and 3.

The first iteration is based on the ML results in Section 2: first fix $\boldsymbol{\eta}$ and compute $H = \widehat{H}(\boldsymbol{\eta})$ and $\Sigma = \widehat{\Sigma}(\boldsymbol{\eta})$ using (2.4), then fix H and Σ and minimize the interference rejection combining cost function $\sum_{t=1}^{N} [\boldsymbol{y}(t) - H\boldsymbol{\phi}(t,\boldsymbol{\eta})]^* \cdot \Sigma^{-1} \cdot [\boldsymbol{y}(t) - H\boldsymbol{\phi}(t,\boldsymbol{\eta})]$ with respect to $\boldsymbol{\eta}$. Iterate between the above two steps as long as there is a significant increase in (2.5).

An alternative iterative method is based on (3.1): first fix $\boldsymbol{\eta}$ and compute $B = \widehat{B}(\boldsymbol{\eta})$ and $W = \widehat{W}(\boldsymbol{\eta})$ using (3.3), then fix B and W and maximize (3.1) with respect to $\boldsymbol{\eta}$; iterate as long as there is a significant increase in (3.1). In the following section, we consider the full-rank channel with $r = d$ co-channel signals, which allows for further simplifications of this iteration.

5.1. Full-rank Channel with $r = d$ Co-channel Signals

Consider now an important special case of a full-rank channel with $r = d$ narrowband co-channel signals $s_1(t,\boldsymbol{\eta})$, $s_2(t,\boldsymbol{\eta}), \ldots, s_r(t,\boldsymbol{\eta})$ impinging on the array. Then, the basis function vector becomes $\boldsymbol{\phi}(t,\boldsymbol{\eta}) = [s_1(t,\boldsymbol{\eta}), s_2(t,\boldsymbol{\eta}), \ldots, s_r(t,\boldsymbol{\eta})]^T$. Since W in (3.1) reduces to a square (and generally non-singular) matrix, we can recover the signals $\Phi(\boldsymbol{\eta})$ by computing $\Omega = W^{-1}BY$, once B and W are estimated. Note that (2.3c) simplifies to $\widehat{\Lambda} = [\widehat{\Lambda}(r), 0]^T$, and $\widehat{V}(r) = \widehat{V}$, implying that $\widehat{C}_{y\phi} = \widehat{U}(r)\widehat{\Lambda}(r)\widehat{V}^*$, $\widehat{W}(\boldsymbol{\eta}) = \widehat{\Lambda}(r)\widehat{V}^*\widehat{R}_{\phi\phi}^{-1/2}$, $\widehat{H}(\boldsymbol{\eta}) = \widehat{R}_{y\phi}\widehat{R}_{\phi\phi}^{-1}$, and $\widehat{\Sigma}(\boldsymbol{\eta}) = \widehat{R}_{yy} - \widehat{R}_{y\phi}\widehat{R}_{\phi\phi}^{-1}\widehat{R}_{y\phi}^*$. Also, (2.5) reduces to $l(\boldsymbol{\eta}) = |\widehat{R}_{\phi\phi}|/|\widehat{R}_{\phi\phi} - \widehat{R}_{y\phi}^*\widehat{R}_{yy}^{-1}\widehat{R}_{y\phi}| = |\widehat{R}_{yy}|/|\widehat{R}_{yy} - \widehat{R}_{y\phi}\widehat{R}_{\phi\phi}^{-1}\widehat{R}_{y\phi}^*|$, which can be viewed as estimated geometric signal-to-noise ratio, see [18]. Define

$$\widehat{\Omega}(\boldsymbol{\eta}) = \widehat{B}_{\rm WLS}(\boldsymbol{\eta})Y, \tag{5.1}$$

where $\widehat{B}_{\rm WLS}(\boldsymbol{\eta}) = \widehat{W}(\boldsymbol{\eta})^{-1}\widehat{B}(\boldsymbol{\eta}) = [\widehat{H}(\boldsymbol{\eta})^*\widehat{\Sigma}(\boldsymbol{\eta})^{-1}\widehat{H}(\boldsymbol{\eta})]^{-1} \cdot \widehat{H}(\boldsymbol{\eta})^*\widehat{\Sigma}(\boldsymbol{\eta})^{-1} = [\widehat{H}(\boldsymbol{\eta})^*\widehat{R}_{yy}^{-1}\widehat{H}(\boldsymbol{\eta})]^{-1}\widehat{H}(\boldsymbol{\eta})^*\widehat{R}_{yy}^{-1} = \widehat{R}_{\phi\phi}^{1/2}\widehat{V}\widehat{\Lambda}(r)^{-1}\widehat{U}(r)^*\widehat{R}_{yy}^{-1/2}$ is the (estimated) weighted least squares (WLS) beamformer. Thus, $\widehat{\Omega}(\boldsymbol{\eta})$ can be viewed as an WLS estimate of the basis function matrix $\Phi(\boldsymbol{\eta})$. Since $\widehat{B}_{\text{WLS}}(\boldsymbol{\eta})\widehat{H}(\boldsymbol{\eta}) = I_r$, $\widehat{B}_{\text{WLS}}(\boldsymbol{\eta})$ is a left inverse of $\widehat{H}(\boldsymbol{\eta})$.

The second iterative procedure described in the previous section simplifies as follows: first fix $\boldsymbol{\eta}$ and compute $\Omega = \widehat{\Omega}(\boldsymbol{\eta})$ using (5.1). Then, fix Ω and find $\boldsymbol{\eta}$ that maximizes $|\Xi(\boldsymbol{\eta})|^{-1}$, where $\Xi(\boldsymbol{\eta}) = (1/N) \cdot [\Omega - \Phi(\boldsymbol{\eta})] \cdot [\Omega - \Phi(\boldsymbol{\eta})]^*$. Iterate as long as there is a significant increase in (3.1) between consecutive steps.

A sub-optimal second step may be to simply project Ω onto finite alphabet to demodulate the unknown symbols $\boldsymbol{\eta}$; this would effectively minimize the diagonal entries of $\Xi(\boldsymbol{\eta})$ and therefore its trace, but not necessarily the determinant (for $r = d = 1$ this is optimal, see the following section).

5.1.1. Single Source

In the case of a single source, we have $r = d = 1$ and the basis function matrix degenerates to a row vector $\Phi(\boldsymbol{\eta}) = [s(1,\boldsymbol{\eta}), s(2,\boldsymbol{\eta}), \ldots, s(N,\boldsymbol{\eta})]$. Then, $\widehat{R}_{y\phi} = \widehat{\boldsymbol{r}}_{y\phi} = (1/N) \cdot \sum_{t=1}^{N} \boldsymbol{y}(t)s(t,\boldsymbol{\eta})^*$ and $\widehat{R}_{\phi\phi} = \widehat{r}_{\phi\phi} = (1/N) \cdot \sum_{t=1}^{N} |s(t,\boldsymbol{\eta})|^2$, and the concentrated likelihood function in (2.5) becomes $l(\boldsymbol{\eta}) = \widehat{r}_{\phi\phi}/(\widehat{r}_{\phi\phi} - \widehat{\boldsymbol{r}}_{y\phi}^* \widehat{R}_{yy}^{-1} \widehat{\boldsymbol{r}}_{y\phi})$. After the monotonic transformation $1 - 1/l(\boldsymbol{\eta})$, the concentrated likelihood function further simplifies to $l_1(\boldsymbol{\eta}) = \widehat{\boldsymbol{r}}_{y\phi}^* \widehat{R}_{yy}^{-1} \widehat{\boldsymbol{r}}_{y\phi}/\widehat{r}_{\phi\phi}$. This concentrated likelihood function can be maximized using the iterative procedure from the previous section. For fixed $\boldsymbol{\eta}$, the first step consists of computing [see (5.1)]

$$\omega(t) = \widehat{\omega}(t,\boldsymbol{\eta}) = (\widehat{r}_{\phi\phi}/\widehat{\boldsymbol{r}}_{y\phi}^* \widehat{R}_{yy}^{-1} \widehat{\boldsymbol{r}}_{y\phi}) \cdot \widehat{\boldsymbol{r}}_{y\phi}^* \widehat{R}_{yy}^{-1} \boldsymbol{y}(t), \quad (5.2)$$

for $t = 1, 2, \ldots, N$. Then, in the second step, fix $\omega(t)$ and minimize $\Xi(\boldsymbol{\eta}) = (1/N) \cdot \sum_{t=1}^{N} |\omega(t) - s(t,\boldsymbol{\eta})|^2$ with respect to $\boldsymbol{\eta}$. If $\boldsymbol{\eta}$ contains unknown symbols and each time snapshot corresponds to a different symbol, each term in the above summation can be minimized separately; we can view the second step as projection onto finite alphabet. In this case, the above iteration is identical to the recently proposed decoupled weighted iterative least squares with projection (DW-ILSP) [11].

In the case when the signal $s(t,\boldsymbol{\eta})$ is modeled only by using a constant modulus property, we can choose $\Phi(\boldsymbol{\eta}) = [\exp[j\vartheta(1)], \exp[j\vartheta(2)], \ldots, \exp[j\vartheta(N)]]$ and thus $\boldsymbol{\eta} = [\vartheta(1), \vartheta(2), \ldots, \vartheta(N)]^T$. Note that here $\widehat{r}_{\phi\phi} = 1$, and the first step of the iteration consists of computing $\omega(t) = \widehat{\omega}(t,\boldsymbol{\eta}) = \widehat{\boldsymbol{r}}_{y\phi}^* \widehat{R}_{yy}^{-1} \boldsymbol{y}(t)/\widehat{\boldsymbol{r}}_{y\phi}^* \widehat{R}_{yy}^{-1} \widehat{\boldsymbol{r}}_{y\phi}$ for $t = 1, 2, \ldots, N$. Then, in the second step, fix $\omega(t)$, $t = 1, 2, \ldots, N$ and compute $\widehat{\boldsymbol{\eta}} = [\angle\omega(1), \angle\omega(2), \ldots, \angle\omega(N)]^T$, which minimizes $\Xi(\boldsymbol{\eta}) = (1/N) \cdot \sum_{t=1}^{N} |\omega(t) - \exp[\vartheta(t)]|^2$, yielding $\Xi(\widehat{\boldsymbol{\eta}}) = (1/N) \cdot \sum_{t=1}^{N} |\omega(t) - \exp[\widehat{\vartheta}(t)]|^2 = (1/N) \cdot \sum_{t=1}^{N} |\omega(t) - \omega(t)/|\omega(t)||^2$, which is an estimated *mean-squared amplitude fluctuation* of the beamformer's output $\omega(t)$. The obtained algorithm is identical to the *least-squares constant modulus algorithm* (LSCMA) in [12].

6. REFERENCES

[1] A.J. Paulraj and C.B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Mag.*, vol. 14, no. 6, pp. 49–83, Nov. 1997.

[2] G.J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[3] A. Maleki-Tehrani, B. Hassibi, and J.M. Cioffi, "Adaptive equalization of multiple-input multiple-output (MIMO) channels," in *Proc. ICC*, New Orleans, LA, June 2000, pp. 1670–1674.

[4] N. Al-Dhahir and A.H. Sayed, "The finite-length multi-input multi-output MMSE-DFE," *IEEE Trans. Signal Processing*, vol. 48, pp. 2921–2936, Oct. 2000.

[5] L.L. Scharf and J.K. Thomas, "Wiener filters in canonical coordinates for transform coding, filtering, and quantizing," *IEEE Trans. Signal Processing*, vol. 46, pp. 647–654, Mar. 1998.

[6] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, 2nd edition, 1984.

[7] F. Pipon, P. Chevalier, P. Vila, and J.J. Monot, "Joint spatial and temporal equalization for channels with ISI and CCI," in *Proc. 1st IEEE SP Workshop SPAWC*, Paris, France, Apr. 1997, pp. 309–312.

[8] M.A. Lagunas, A.I. Perez-Neira, and J. Vidal, "Optimal array combiner for sequence detectors," in *Proc. ICASSP*, Seattle, WA, May 1998, pp. 3341–3344.

[9] D. Giancola *et al.*, "Space-time processing for time varying co-channel interference rejection and channel estimation in GSM/DCS systems," in *Proc. 49th Veh. Technol. Conf.*, Houston, TX, May 1999, pp. 1317–1323.

[10] D.D. Falconer and F.R. Magee Jr., "Adaptive channel memory truncation for maximum likelihood sequence estimation," *Bell Syst. Tech. J.*, vol. 52, pp. 1541–1562, Nov. 1973.

[11] A. Ranheim, "A decoupled approach to adaptive signal separation using an antenna array," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 676–682, May 1999.

[12] B.G. Agee, "The least-squares CMA: a new technique for rapid correction of constant modulus signals," in *Proc. ICASSP*, Tokyo, Japan, Apr. 1986, pp. 953–956.

[13] P. Stoica and M. Viberg, "Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 3069–3078, Dec. 1996.

[14] A. Dogandžić and A. Nehorai, "Space-time fading channel estimation in unknown spatially correlated noise," in *Proc. 37th Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Sept. 1999, pp. 948–957.

[15] A. Dogandžić and A. Nehorai, "Finite-length multi-input multi-output adaptive equalization using canonical correlation analysis," submitted, 2000.

[16] G.E. Bottomley and K. Jamal, "Adaptive arrays and MLSE equalization," in *Proc. 45th IEEE Veh. Technol. Conf.*, Chicago, IL, July 1995, pp. 50–54.

[17] M. Stojanovic, J. Catipovic, and J.G. Proakis, "Adaptive multichannel combining and equalization for underwater acoustic communications," *J. Acoust. Soc. Am.*, vol. 94, pp. 1621–1631, Sept. 1993.

[18] J.M. Cioffi, P.H. Algoet, and P.S. Chow, "Combined equalization and coding with finite-length decision feedback equalization," in *Proc. Globecom Conf.*, San Diego, CA, Dec. 1990, pp. 1664–1668.