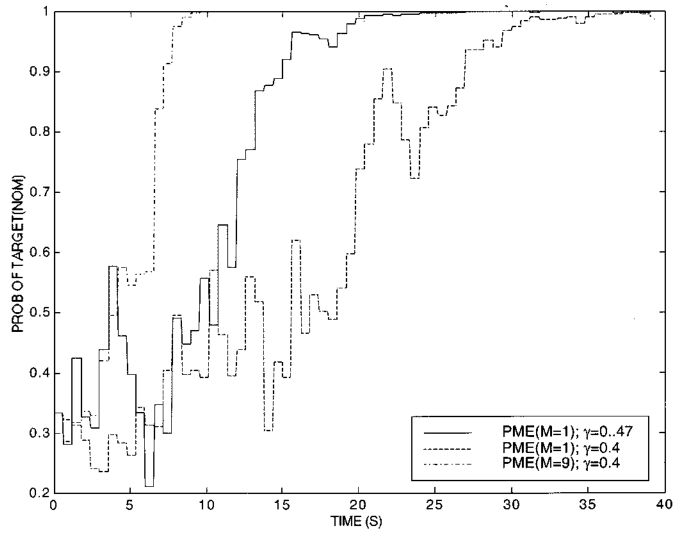

Fig. 2. Target recognition using shape analysis in the PME.

quality. As γ changes from 0.33 to 0.4, the time required for 90% identification is reduced in the PME$(M = 9)$ from 9 to 6 s. As the image processing becomes better, the advantage of the symmetric PME becomes less since the contribution of $\{y_k\}$ to the modal update becomes smaller.

IV. Conclusion

For problems in which the modal state is hidden in the modal measurement, a symmetric modification of the PME improves the speed of modal identification. The symmetric PME achieves its advantage by incorporating the correlation between target type and tracking error into the modal estimate. Base-state estimation is less sensitive to modal identification than is often supposed. Only when the application requires high-quality modal estimation is the symmetric PME required.

References

[1] J. S. Evans and R. J. Evans, "Image-enhanced multiple model tracking," *Automatica*, vol. 35, pp. 1769–1786, 1999.

[2] M. I. Miller, A. Srivastava, and U. Grenander, "Conditional-mean estimation via jump-diffusion processes in multiple target tracking/recognition," *IEEE Trans. Signal Processing*, vol. 43, pp. 2678–2690, Nov. 1995.

[3] D. D. Sworder and J. E. Boyd, "Jump-diffusion processes in tracking/recognition," *IEEE Trans. Signal Processing*, vol. 46, pp. 235–239, Jan. 1998.

[4] ——, *Estimation Problems in Hybrid Systems*. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[5] H. Wang, T. Kirubarajan, and Y. Bar-Shalom, "Large scale air traffic surveillance using IMM estimators with assignment," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, pp. 255–266, Jan. 1999.

Finite-Length MIMO Equalization Using Canonical Correlation Analysis

Aleksandar Dogandžić and Arye Nehorai

***Abstract*—We propose finite-length multi-input multi-output (MIMO) equalization methods for "smart" antenna arrays using the statistical theory of canonical correlations. We show that the proposed methods are related to maximum likelihood (ML) reduced-rank channel and noise estimation algorithms in unknown spatially correlated noise as well as to several recently developed equalization schemes.**

***Index Terms*—MIMO equalization, space-time processing.**

I. Introduction

Multi-input multi-output (MIMO) channel equalization has recently been of interest due to the popularity of antenna arrays (also known as smart antennas) applied at the receiver [1], [2] and transmitter [3]. Adaptive and nonadaptive MIMO decision-feedback equalizers (DFEs) have been recently proposed in [4]–[6] (see also references therein). In addition, reduced-rank channel estimation [7]–[11] and equalization [12] have recently attracted considerable attention, as low-rank channels may appear in practical situations [13], [14]. In this correspondence (see also [15]), we present a framework for finite-length MIMO spatial and temporal equalization based on the canonical correlation analysis [16]–[18]. This framework allows for multivariate extensions of the array combining algorithms in [19]–[24], classical finite-length equalization in [25], [26], joint data and channel estimation algorithm in [27], and blind adaptive beamforming methods that use finite alphabet [28] and constant modulus [29] properties of the received signal. We show a relationship between the proposed methods and the maximum likelihood (ML) reduced-rank channel and noise estimation for unknown spatially correlated noise in [7].

First, in Section II, we briefly review the reduced-rank ML channel and noise estimation in [7]. In Section III, we describe the proposed equalization criterion and relate it to the ML estimation results from Section II. Then, we discuss its application when training data is available (see Section IV) or not available (i.e., blind scenario; see Section V).

II. Reduced-Rank ML Estimation

We review the ML estimation in [7] for a reduced-rank channel and spatially correlated noise with unknown covariance. Similarly to [7], we model the received signal as a linear combination of basis functions, which includes various wireless channel models as special cases; see [1] and [8]. (Similar reduced-rank models for channel estimation using known basis functions were also used in [9, Sec. 2.6] [10, Sec. III], and [11]. Unlike [7], where the measurements are real the and the basis functions are

Manuscript received December 7, 2000; revised January 2, 2001. This work was supported by the Air Force Office of Scientific Research under Grants F49620-99-1-0067 and F49620-00-1-0083, the National Science Foundation under Grant CCR-0105334, the Office of Naval Research under Grant N00014-01-1-0681, and the UIC Graduate Student Fellowship. The associate editor coordinating the review of this paper and approving it for publication was Dr. Kristine L. Bell.

A. Dogandžić was with the Department of Electrical Engineering and Computer Science, University of Illinois at Chicago, Chicago, IL, 60607 USA. He is now with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: ald@iastate.edu).

A. Nehorai is with the Department of Electrical and Computer Engineering, , University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: nehorai@ece.uic.edu).

Publisher Item Identifier S 1053-587X(02)02395-4.

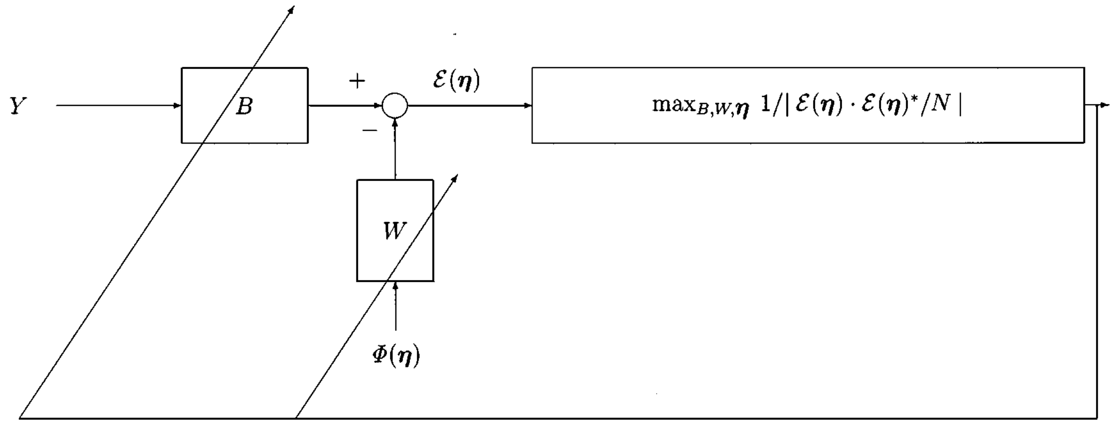

Fig. 1. Proposed multi-input multi-output equalization scheme.

known, here, we consider the measurement model with complex data and parametric basis functions (see also [8], [31]). The proposed parametric basis-function model is useful for blind equalization and symbol detection, i.e., when training data is not available; see Section V

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:

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

where

- H unknown $m \times d$ channel response matrix of rank $r \leq \min(m,d)$;
- $\boldsymbol{\phi}(t,\boldsymbol{\eta})$ $d \times 1$ vector of basis functions;
- $\boldsymbol{e}(t)$ zero-mean complex 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 V-A1). Observe that the channel matrix H in (2.1) is assumed to be constant in the observation interval $t = 1, 2, \ldots, N$, which may be restrictive in fast fading scenarios. This problem can be overcome by a proper choice of basis functions $\boldsymbol{\phi}(t,\boldsymbol{\eta})$, which would incorporate Doppler effect; see [30] and [31].

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

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

which is the estimated cross-correlation between the vectors $\hat{R}_{yy}^{-1/2}\boldsymbol{y}(t)$ and $\hat{R}_{\phi\phi}^{-1/2}\boldsymbol{\phi}(t)$ or the estimated *coherence matrix* between $\boldsymbol{y}(t)$ and $\boldsymbol{\phi}(t,\boldsymbol{\eta})$; see ([16, Sec. VII]). Here, "*" denotes a conjugate transpose. In addition, $A^{1/2}$ 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 $\hat{R}_{y\phi}$ and $\hat{R}_{\phi\phi}$ are functions of $\boldsymbol{\eta}$. To simplify the notation, we omit these dependencies. Consider the singular value decomposition (SVD) of $\hat{C}_{y\phi}$:

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

where $1 \geq \hat{\lambda}(1) \geq \hat{\lambda}(2) \geq \ldots \geq \hat{\lambda}(\min(m,d)) \geq 0$. 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: $\hat{U}(r)$ and $\hat{V}(r)$ are the matrices containing the first r columns of $\hat{U}$ and $\hat{V}$, respectively. For the model in (2.1) with known $\boldsymbol{\eta}$, the ML estimates of H and Σ are

$$\hat{H}(\boldsymbol{\eta}) = \hat{R}_{yy}^{1/2}\hat{U}(r)\hat{\Lambda}(r)\hat{V}(r)^*\hat{R}_{\phi\phi}^{-1/2} \tag{2.4a}$$
$$\hat{\Sigma}(\boldsymbol{\eta}) = \hat{R}_{yy} - \hat{R}_{yy}^{1/2}\hat{U}(r)\hat{\Lambda}^2(r)\hat{U}(r)^*\hat{R}_{yy}^{1/2} \tag{2.4b}$$

see [7] and [8]. Methods for efficiently computing the above expressions when $r \ll \min(m,d)$ are derived in [32]. If $\boldsymbol{\eta}$ is unknown, its ML estimate $\hat{\boldsymbol{\eta}}$ is obtained by maximizing the concentrated likelihood

$$l_{\text{CLF}}(\boldsymbol{\eta}) = \prod_{i=1}^{r}\frac{1}{1-\hat{\lambda}^2(i)} \tag{2.5}$$

(see [8, Eq. (4.1)] and [7, Eq. (35)]). To find the ML estimates of H and Σ, replace $\boldsymbol{\eta}$ in (2.4) by $\hat{\boldsymbol{\eta}}$. Based on [33, Sec. 9.2] (see also [16, Sec. IV]), we can view $\log[l_{\text{CLF}}(\boldsymbol{\eta})]$ as a measure of (estimated) mutual information between $\boldsymbol{y}(t)$ and $\boldsymbol{\phi}(t,\boldsymbol{\eta})$.

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 shown in Fig. 1.

III. MIMO EQUALIZATION

We analyze the MIMO equalization scheme depicted in Fig. 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 $\epsilon(t,\boldsymbol{\eta}) = B\boldsymbol{y}(t) - W\boldsymbol{\phi}(t,\boldsymbol{\eta})$ in the mean-square sense. Define the error matrix as $\mathcal{E}(\boldsymbol{\eta}) = [\epsilon(1,\boldsymbol{\eta})\cdots\epsilon(N,\boldsymbol{\eta})] = BY - W\Phi(\boldsymbol{\eta})$. 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 the estimated geometric mean-squared error of $\epsilon(t,\boldsymbol{\eta})$

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

subject to the normalizing constraint

$$B\hat{R}_{yy}B^* = I_r. \tag{3.2}$$

Here, $|\cdot|$ denotes the determinant. The normalizing constraint prevents the trivial solution (in which B and W equal zero) and imposes the estimated beamformed signals BY to be uncorrelated. It also constrains the dynamic range of the beamformed data, which is highly desirable for implementation. Note that the classical equalization schemes in [25] and [26] and array combining algorithm in [19] use normalizing constraints applied to the basis-function filtering matrix W.

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

$$\begin{aligned}\hat{B}(\boldsymbol{\eta}) &= [\hat{\boldsymbol{b}}_1(\boldsymbol{\eta}), \hat{\boldsymbol{b}}_2(\boldsymbol{\eta}) \cdots \hat{\boldsymbol{b}}_r(\boldsymbol{\eta})]^* \\ &= \hat{U}(r)^* \hat{R}_{yy}^{-1/2} \end{aligned} \tag{3.3a}$$

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

(see the Appendix). Therefore, $B = \hat{B}(\boldsymbol{\eta})$ and $W = \hat{W}(\boldsymbol{\eta})$ maximize (3.1), yielding

$$\begin{aligned}\frac{1}{N}\hat{\mathcal{E}}(\boldsymbol{\eta}) \cdot \hat{\mathcal{E}}(\boldsymbol{\eta})^* &= \frac{1}{N} \cdot [\hat{B}(\boldsymbol{\eta})Y - \hat{W}(\boldsymbol{\eta})\Phi(\boldsymbol{\eta})] \\ &\quad \cdot [\hat{B}(\boldsymbol{\eta})Y - \hat{W}(\boldsymbol{\eta})\Phi(\boldsymbol{\eta})]^* \\ &= \frac{1}{N}\sum_{t=1}^{N}[\hat{\boldsymbol{y}}_c(t,\boldsymbol{\eta}) - \hat{\boldsymbol{\phi}}_c(t,\boldsymbol{\eta})] \\ &\quad \cdot [\hat{\boldsymbol{y}}_c(t,\boldsymbol{\eta}) - \hat{\boldsymbol{\phi}}_c(t,\boldsymbol{\eta})]^* \\ &= I_r - \hat{\Lambda}^2(r) \end{aligned} \tag{3.4}$$

and $l(\boldsymbol{\eta}, \hat{B}(\boldsymbol{\eta}), \hat{W}(\boldsymbol{\eta})) = l_{\mathrm{CLF}}(\boldsymbol{\eta})$, which is the concentrated likelihood function in (2.5). Note that $\hat{B}(\boldsymbol{\eta}) \cdot \hat{H}(\boldsymbol{\eta}) = \hat{W}(\boldsymbol{\eta})$, where $\hat{H}(\boldsymbol{\eta})$ is the ML estimate of the channel in (2.4a); thus, $\hat{W}(\boldsymbol{\eta})$ is an estimate of the channel matrix *after* applying the beamforming matrix $\hat{B}(\boldsymbol{\eta})$ to the received data. In addition, $\hat{\boldsymbol{y}}_c(t,\boldsymbol{\eta}) = [\hat{y}_{c,1}(t,\boldsymbol{\eta}), \hat{y}_{c,2}(t,\boldsymbol{\eta}), \ldots, \hat{y}_{c,r}(t,\boldsymbol{\eta})]^T = \hat{B}(\boldsymbol{\eta})\boldsymbol{y}(t)$ and $\hat{\boldsymbol{\phi}}_c(t,\boldsymbol{\eta}) = [\hat{\phi}_{c,1}(t,\boldsymbol{\eta}), \hat{\phi}_{c,2}(t,\boldsymbol{\eta}), \ldots, \hat{\phi}_{c,r}(t,\boldsymbol{\eta})]^T = \hat{W}(\boldsymbol{\eta})\boldsymbol{\phi}(t,\boldsymbol{\eta})$ can be viewed as the estimated *canonical coordinate* vectors of the data and basis functions, respectively, whereas $\hat{\lambda}(i)$ are the estimated *canonical correlations*; see [16]. This allows for an elegant interpretation of the proposed equalization scheme in the context of canonical correlation analysis, as follows (see also [18, ch. 12]). The first estimated canonical coordinates $\hat{y}_{c,1}(t,\boldsymbol{\eta}) = \hat{\boldsymbol{b}}_1(\boldsymbol{\eta})^*\boldsymbol{y}(t)$ and $\hat{\phi}_{c,1}(t,\boldsymbol{\eta}) = \hat{\boldsymbol{w}}_1(\boldsymbol{\eta})^*\boldsymbol{\phi}(t,\boldsymbol{\eta})$ have the largest estimated correlation $\hat{\lambda}(1)$ among all possible linear combinations of $\boldsymbol{y}(t)$ and $\boldsymbol{\phi}(t,\boldsymbol{\eta})$. Here, we define the estimated correlation between $\boldsymbol{b}^*\boldsymbol{y}(t)$ and $\boldsymbol{w}^*\boldsymbol{\phi}(t,\boldsymbol{\eta})$ as

$$\widehat{\mathrm{corr}}\{\boldsymbol{b}^*\boldsymbol{y}(t), \boldsymbol{w}^*\boldsymbol{\phi}(t,\boldsymbol{\eta})\} = \frac{|\boldsymbol{b}^*\hat{R}_{y\phi}\boldsymbol{w}|}{\sqrt{\boldsymbol{b}^*\hat{R}_{yy}\boldsymbol{b}} \cdot \sqrt{\boldsymbol{w}^*\hat{R}_{\phi\phi}\boldsymbol{w}}}. \tag{3.5}$$

Further, $\hat{y}_{c,2}(t,\boldsymbol{\eta}) = \hat{\boldsymbol{b}}_2(\boldsymbol{\eta})^*\boldsymbol{y}(t)$ and $\hat{\phi}_{c,2}(t,\boldsymbol{\eta}) = \hat{\boldsymbol{w}}_2(\boldsymbol{\eta})^*\boldsymbol{\phi}(t,\boldsymbol{\eta})$ have the largest estimated correlation $\hat{\lambda}(2)$ among all possible linear combinations of $\boldsymbol{y}(t)$ and $\boldsymbol{\phi}(t,\boldsymbol{\eta})$ that are uncorrelated with $\hat{y}_{c,1}(t,\boldsymbol{\eta})$ and $\hat{\phi}_{c,1}(t,\boldsymbol{\eta})$, and so on;[1] see Fig. 2. In addition, from (3.4), it follows that the estimated mean-square error matrix $\hat{\mathcal{E}}(\boldsymbol{\eta}) \cdot \hat{\mathcal{E}}(\boldsymbol{\eta})^*/N$ is diagonal with diagonal terms sorted in nonincreasing order.

Interestingly, $\hat{B}(\boldsymbol{\eta})$ and $\hat{W}(\boldsymbol{\eta})$ in (3.3) also maximize the following criterion:

$$\begin{aligned}\mathrm{SINR}(\boldsymbol{\eta}, B, W) &= \frac{|B\hat{R}_{yy}B^*|}{|(1/N) \cdot \mathcal{E}(\boldsymbol{\eta}) \cdot \mathcal{E}(\boldsymbol{\eta})^*|} \\ &= \frac{|(1/N) \cdot BYY^*B^*|}{|(1/N) \cdot [BY - W\Phi(\boldsymbol{\eta}] \cdot [BY - W\Phi(\boldsymbol{\eta})]^*|} \end{aligned} \tag{3.6}$$

which is a multivariate extension of the signal-to-interference-plus-noise ratio (SINR) definition in [24, Eq. (3)]); see the Appendix. Note that $\mathrm{SINR}(\boldsymbol{\eta}, \hat{B}(\boldsymbol{\eta}), \hat{W}(\boldsymbol{\eta})) = l_{\mathrm{CLF}}(\boldsymbol{\eta})$. In addition, $\hat{B}(\boldsymbol{\eta})$ and $\hat{W}(\boldsymbol{\eta})$

[1]Note that the estimated variances of $\hat{y}_{c,i}(t,\boldsymbol{\eta})$ are normalized to one [due to the constraint (3.2)], whereas the estimated variances of $\hat{\phi}_{c,i}(t,\boldsymbol{\eta})$ are $\hat{\lambda}^2(i)$ for $i = 1, 2, \ldots, r$. For simplicity, in the literature, the variances of $\hat{\phi}_{c,i}(t,\boldsymbol{\eta})$ are often normalized to one as well, in which case, the correlation and covariance between the canonical coordinates are equivalent.

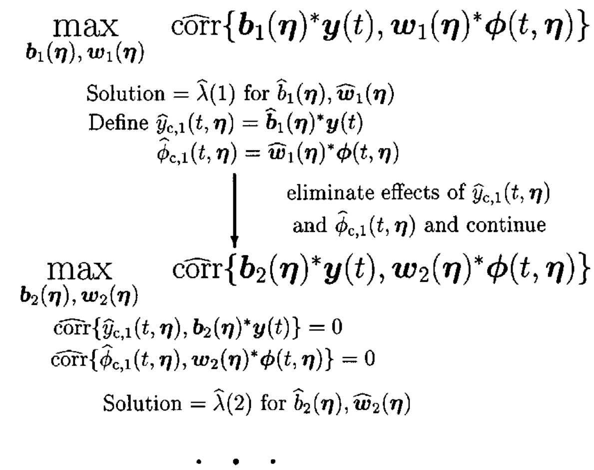

Fig. 2. Estimated canonical correlations.

in (3.3) are only one possible choice of B and W that maximize (3.6); other solutions can be obtained by premultiplying both B and W by an arbitrary nonsingular square matrix.

For a single antenna with $\boldsymbol{y}(t) = [y(t), y(t-1), \ldots, y(t-m+1)]^T$, rank $r = 1$, and basis functions chosen to model multipath effect by uniformly discretizing the time-delay spread, i.e., $\boldsymbol{\phi}(t,\boldsymbol{\eta}) = [s(t), s(t-1), \ldots, s(t-d+1)]^T$, the equalizers in (3.3) become row vectors, i.e., $\hat{B}(\boldsymbol{\eta}) = \hat{\boldsymbol{b}}(\boldsymbol{\eta})^*$ and $\hat{W}(\boldsymbol{\eta}) = \hat{\boldsymbol{w}}(\boldsymbol{\eta})^*$, where $\hat{\boldsymbol{b}}(\boldsymbol{\eta})$ can be interpreted as a feedforward filter, which shapes the channel to the desired impulse response $\hat{\boldsymbol{w}}(\boldsymbol{\eta})$; this is a classical equalization scheme in [25] and [26].

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

IV. MIMO Equalization and Symbol Detection Using Training Data

If training data is available, we can separate the estimation and detection tasks as follows: Use the training data to estimate Σ and H [see (2.4)], and then detect the unknown sequence using *interference rejection combining* (IRC) [34, Eq. (8)] (see also [35] and [36]). We show that this procedure is equivalent to estimating B and W from the training data [see (3.3)] and detecting the unknown sequence by applying *metric combining* (MC) [34, Sec. IV.A] to the equalized data and basis functions.

For known H and Σ, maximizing the likelihood function for the measurement model (2.1) is equivalent to minimizing the following interference rejection combining cost function:

$$\begin{aligned}C^{\mathrm{IRC}}(\boldsymbol{\eta}; Y; H, \Sigma) = \sum_{t=1}^{N}[\boldsymbol{y}(t) - H\boldsymbol{\phi}(t,\boldsymbol{\eta})]^* \\ \cdot \Sigma^{-1} \cdot [\boldsymbol{y}(t) - H\boldsymbol{\phi}(t,\boldsymbol{\eta})]. \end{aligned} \tag{4.1}$$

Note that the measurement model in Section II [and, consequently, the above IRC cost function] assumes stationary noise (i.e., the spatially noise covariance Σ is constant in time, as in [35] and [36]), whereas [34] allows for the noise covariance to change in time. Thus, (4.1) is a stationary version of the IRC cost function in [34, Eq. (8)]. Similarly, a stationary metric combining cost function follows from (4.1) if Σ is a diagonal matrix.

Let $Y_K = [\boldsymbol{y}_K(1) \cdots \boldsymbol{y}_K(N_K)]$ be the data set containing the known (or training) sequence described by known basis functions $\Phi_K = [\boldsymbol{\phi}_K(1) \cdots \boldsymbol{\phi}_K(N_K)]$. Further, by analogy with (2.2) and (2.3), define $\hat{C}_{Ky\phi} = \hat{R}_{Kyy}^{-1/2}\hat{R}_{Ky\phi}\hat{R}_{K\phi\phi}^{-1/2} = \hat{U}_K\hat{\Lambda}_K\hat{V}_K^*$, where $\hat{R}_{Kyy} = Y_K Y_K{}^*/N_K$, $\hat{R}_{K\phi\phi} = \Phi_K\Phi_K{}^*/N_K$, and $\hat{R}_{Ky\phi} =$

$Y_K \Phi_K{}^*/N_K$. In addition, define the singular values of $\hat{C}_{Ky\phi} : 1 \geq \hat{\lambda}_K(1) \geq \hat{\lambda}_K(2) \geq \ldots \geq \hat{\lambda}_K(\min(m,d)) \geq 0$ and $\hat{\Lambda}_K(r) = \mathrm{diag}\{\hat{\lambda}_K(1), \hat{\lambda}_K(2), \ldots, \hat{\lambda}_K(r)\}$. Then, from (3.3), the estimates of B and W based on the training data are $\hat{B}_K = \hat{U}_K(r)^* \hat{R}_{Kyy}^{-1/2}$ and $\hat{W}_K = \hat{\Lambda}_K(r)\hat{V}_K(r)^* \hat{R}_{K\phi\phi}^{-1/2}$, and the channel and noise estimates follow from (2.4) as $\hat{H}_K = \hat{R}_{Kyy}^{1/2}\hat{U}_K(r)\hat{\Lambda}_K(r)\hat{V}_K(r)^*\hat{R}_{K\phi\phi}^{-1/2}$ and $\hat{\Sigma}_K = \hat{R}_{Kyy} - \hat{R}_{Kyy}^{1/2}\hat{U}_K(r)\hat{\Lambda}_K^2(r)\hat{U}_K(r)^*\hat{R}_{Kyy}^{1/2}$. Note that $(1/N_K) \cdot [\hat{B}_K Y_K - \hat{W}_K \Phi_K] \cdot [\hat{B}_K Y_K - \hat{W}_K \Phi_K]^* = I_r - \hat{\Lambda}_K^2(r)$, i.e., the estimated error covariance matrix is diagonal; see also (3.4).

To find the unknown sequence $\boldsymbol{\eta}$, we can use the IRC cost function (4.1) applied to the data Y containing the unknown sequence and the channel and noise covariance estimates $\hat{H}_K$ and $\hat{\Sigma}_K$:

$$C^{\mathrm{IRC}}(\boldsymbol{\eta}; Y; \hat{H}_K, \hat{\Sigma}_K) = \sum_{t=1}^{N} [\boldsymbol{y}(t) - \hat{H}_K \boldsymbol{\phi}(t,\boldsymbol{\eta})]^* \times \hat{\Sigma}_K^{-1}[\boldsymbol{y}(t) - \hat{H}_K \boldsymbol{\phi}(t,\boldsymbol{\eta})] \quad (4.2)$$

or we can use metric combining applied to the beamformed data $\hat{B}_K Y$ with the channel and noise covariance matrices chosen as $\hat{W}_K$ and $I_r - \hat{\Lambda}_K^2(r)$:

$$\begin{aligned} C^{\mathrm{IRC}}&(\boldsymbol{\eta}; \hat{B}_K Y; \hat{W}_K, I_r - \hat{\Lambda}_K^2(r)) \\ &= \sum_{t=1}^{N} [\boldsymbol{y}_{Kc}(t) - \boldsymbol{\phi}_{Kc}(t,\boldsymbol{\eta})]^* \\ &\quad \cdot [I_r - \hat{\Lambda}_K^2(r)]^{-1} \cdot [\boldsymbol{y}_{Kc}(t) - \boldsymbol{\phi}_{Kc}(t,\boldsymbol{\eta})] \\ &= \sum_{i=1}^{r} [1 - \hat{\lambda}_K^2(i)]^{-1} \sum_{t=1}^{N} \left| \hat{y}_{Kc,i}(t) - \hat{\phi}_{Kc,i}(t,\boldsymbol{\eta}) \right|^2 \end{aligned} \quad (4.3)$$

where

- $\hat{\boldsymbol{y}}_{Kc}(t) = [\hat{y}_{Kc,1}(t), \hat{y}_{Kc,2}(t), \ldots, \hat{y}_{Kc,r}(t)]^T = \hat{B}_K \boldsymbol{y}(t)$;
- $\hat{\boldsymbol{\phi}}_{Kc}(t,\boldsymbol{\eta}) = [\hat{\phi}_{Kc,1}(t,\boldsymbol{\eta}), \hat{\phi}_{Kc,2}(t,\boldsymbol{\eta}), \ldots, \hat{\phi}_{Kc,r}(t,\boldsymbol{\eta})]^T = \hat{W}_K \boldsymbol{\phi}(t,\boldsymbol{\eta})$.

Interestingly, the above cost functions are equal, i.e.,

$$C^{\mathrm{IRC}}(\boldsymbol{\eta}; Y; \hat{H}_K, \hat{\Sigma}_K) = C^{\mathrm{IRC}}(\boldsymbol{\eta}; \hat{B}_K Y; \hat{W}_K, I_r - \hat{\Lambda}_K^2(r)). \quad (4.4)$$

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 [34].

For rank-1 channels (i.e., $r = 1$) and basis functions chosen to model the multipath effect of a single user by uniformly discretizing the time-delay spread (i.e., $\boldsymbol{\phi}_K(t) = [s_K(t), s_K(t-1), \ldots, s_K(t-d+1)]^T$ and similarly for $\boldsymbol{\phi}(t,\boldsymbol{\eta})$), the above equalization and detection algorithms become very similar those in [19]–[24] [where the differences arise because the normalizing constraints in [19]–[24] differ from (3.2)]. A simulation study in [23] shows that the above rank-1 equalization algorithms (followed by an MLSE detector) achieve performance (in terms of probability of error) comparable with that of a full-rank IRC. The simulations in [23] were performed following the global standard for mobile communications (GSM) scenario. Note that in rank-1 schemes, a relatively small number of parameters needs to be estimated [since $\hat{B}(\boldsymbol{\eta})$ and $\hat{W}(\boldsymbol{\eta})$ become row vectors], which may be important if the amount of training data is limited (as in GSM).

V. Blind MIMO Equalization and Symbol Detection

In this section, we discuss the case when training data is unavailable; then, joint equalization and symbol detection can be applied to the received data by maximizing the concentrated likelihood in (2.5). Two iterative procedures for blind MIMO equalization and symbol detection follow from the results of Sections II and III. The first procedure provides a framework for extending the joint data and channel estimation algorithm in [27] to account for multiple receiver antennas and spatially correlated noise. The second procedure outlines a way for extending the decoupled weighted iterative least squares with projection (DW-ILSP) algorithm [28] and least-squares constant modulus algorithm (LSCMA) [29] to multiuser scenario.

The first procedure is based on the ML results in Section II: First, fix $\boldsymbol{\eta}$, and compute $H = \hat{H}(\boldsymbol{\eta})$ and $\Sigma = \hat{\Sigma}(\boldsymbol{\eta})$ 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). For one receiver antenna and basis functions of the form $\boldsymbol{\phi}(t,\boldsymbol{\eta}) = [s(t), s(t-1), \ldots, s(t-d+1)]^T$ [where $s(t)$ are unknown symbols to be detected], the above procedure is equivalent to the one proposed in [27].

An alternative iterative method is based on (3.1): First, fix $\boldsymbol{\eta}$, and compute $B = \hat{B}(\boldsymbol{\eta})$ and $W = \hat{W}(\boldsymbol{\eta})$ 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 $r = d$, which allows for further simplifications of this iteration.

A. Full-Rank Channel

Consider now an important special case where $r = d$, i.e., the channel matrix has full rank $r = d$. Since W in (3.1) reduces to a square (and generally nonsingular) matrix, we can recover the basis-function matrix $\Phi(\boldsymbol{\eta})$ by computing $\Omega = W^{-1}BY$, once B and W are estimated. Note that (2.3c) simplifies to $\hat{\Lambda} = [\hat{\Lambda}(r), 0]^T$, and $\hat{V}(r) = \hat{V}$, implying that $\hat{C}_{y\phi} = \hat{U}(r)\hat{\Lambda}(r)\hat{V}^*$, $\hat{W}(\boldsymbol{\eta}) = \hat{\Lambda}(r)\hat{V}^*\hat{R}_{\phi\phi}^{-1/2}$, $\hat{H}(\boldsymbol{\eta}) = \hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1}$, and $\hat{\Sigma}(\boldsymbol{\eta}) = \hat{R}_{yy} - \hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1}\hat{R}_{y\phi}^*$. In addition, (2.5) reduces to

$$l_{\mathrm{CLF}}(\boldsymbol{\eta}) = \frac{|\hat{R}_{\phi\phi}|}{|\hat{R}_{\phi\phi} - \hat{R}_{y\phi}^*\hat{R}_{yy}^{-1}\hat{R}_{y\phi}|} = \frac{|\hat{R}_{yy}|}{|\hat{R}_{yy} - \hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1}\hat{R}_{y\phi}^*|} \quad (5.1)$$

which can be viewed as estimated geometric signal-to-noise ratio (SNR); see [37]. The above expression is a multivariate extension (accounting for multiple receiver antennas and spatially correlated noise) of the concentrated-likelihood multiuser detector in [38]. For one basis function (i.e., $d = 1$) and one receiver antenna (i.e., $m = 1$), it further reduces to the standard noncoherent detector (see, e.g., [39, Sect. 5.4]). Define

$$\hat{\Omega}(\boldsymbol{\eta}) = \hat{B}_{\mathrm{WLS}}(\boldsymbol{\eta})Y \quad (5.2)$$

where

$$\begin{aligned} \hat{B}_{\mathrm{WLS}}(\boldsymbol{\eta}) &= \hat{W}(\boldsymbol{\eta})^{-1}\hat{B}(\boldsymbol{\eta}) \\ &= [\hat{H}(\boldsymbol{\eta})^*\hat{\Sigma}(\boldsymbol{\eta})^{-1}\hat{H}(\boldsymbol{\eta})]^{-1} \cdot \hat{H}(\boldsymbol{\eta})^*\hat{\Sigma}(\boldsymbol{\eta})^{-1} \\ &= [\hat{H}(\boldsymbol{\eta})^*\hat{R}_{yy}^{-1}\hat{H}(\boldsymbol{\eta})]^{-1} \cdot \hat{H}(\boldsymbol{\eta})^*\hat{R}_{yy}^{-1} \\ &= \hat{R}_{\phi\phi}^{1/2}\hat{V}\hat{\Lambda}(r)^{-1}\hat{U}(r)^*\hat{R}_{yy}^{-1/2} \end{aligned} \quad (5.3)$$

is the (estimated) weighted least squares (WLS) beamformer. Thus, $\hat{\Omega}(\boldsymbol{\eta})$ can be viewed as an WLS estimate of the basis-function matrix $\Phi(\boldsymbol{\eta})$. Since $\hat{B}_{\mathrm{WLS}}(\boldsymbol{\eta})\hat{H}(\boldsymbol{\eta}) = I_r$, $\hat{B}_{\mathrm{WLS}}(\boldsymbol{\eta})$ is a left inverse of $\hat{H}(\boldsymbol{\eta})$.

The second iterative procedure described in the previous section simplifies as follows: First, fix $\boldsymbol{\eta}$, and compute $\Omega = \hat{\Omega}(\boldsymbol{\eta})$ using (5.2). 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). Simulation analysis of the above algorithms will be published elsewhere.

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, $\hat{R}_{y\phi} = \hat{\boldsymbol{r}}_{y\phi} = (1/N)\cdot\sum_{t=1}^{N}\boldsymbol{y}(t)s(t,\boldsymbol{\eta})^*$, and $\hat{R}_{\phi\phi} = \hat{r}_{\phi\phi} = (1/N)\cdot\sum_{t=1}^{N}|s(t,\boldsymbol{\eta})|^2$, and the concentrated likelihood function in (2.5) becomes $l_{\mathrm{CLF}}(\boldsymbol{\eta}) = \hat{r}_{\phi\phi}/(\hat{r}_{\phi\phi} - \hat{\boldsymbol{r}}_{y\phi}^*\hat{R}_{yy}^{-1}\hat{\boldsymbol{r}}_{y\phi})$. After the monotonic transformation $1 - 1/l_{\mathrm{CLF}}(\boldsymbol{\eta})$, the concentrated likelihood function further simplifies to $l_{\mathrm{CLF1}}(\boldsymbol{\eta}) = \hat{\boldsymbol{r}}_{y\phi}^*\hat{R}_{yy}^{-1}\hat{\boldsymbol{r}}_{y\phi}/\hat{r}_{\phi\phi}$. 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 (5.2)]

$$\omega(t) = \hat{\omega}(t,\boldsymbol{\eta}) = (\hat{r}_{\phi\phi}/\hat{\boldsymbol{r}}_{y\phi}^*\hat{R}_{yy}^{-1}\hat{\boldsymbol{r}}_{y\phi})\cdot\hat{\boldsymbol{r}}_{y\phi}^*\hat{R}_{yy}^{-1}\boldsymbol{y}(t) \tag{5.4}$$

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, then 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) [28].

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, $\hat{r}_{\phi\phi} = 1$, and the first step of the iteration consists of computing $\omega(t) = \hat{\omega}(t,\boldsymbol{\eta}) = \hat{\boldsymbol{r}}_{y\phi}^*\hat{R}_{yy}^{-1}\boldsymbol{y}(t)/\hat{\boldsymbol{r}}_{y\phi}^*\hat{R}_{yy}^{-1}\hat{\boldsymbol{r}}_{y\phi}$ for $t = 1, 2, \ldots, N$. Then, in the second step, fix $\omega(t), t = 1, 2, \ldots, N$, and compute $\hat{\boldsymbol{\eta}} = [\angle\omega(1), \angle\omega(2), \ldots, \angle\omega(N)]^T$, which minimizes $\Xi(\boldsymbol{\eta}) = (1/N)\cdot\sum_{t=1}^{N}|\omega(t) - \exp[\vartheta(t)]|^2$, yielding $\Xi(\hat{\boldsymbol{\eta}}) = (1/N)\cdot\sum_{t=1}^{N}|\omega(t) - \exp[\hat{\vartheta}(t)]|^2 = (1/N)\cdot\sum_{t=1}^{N}|\omega(t) - \omega(t)/|\omega(t)||^2$, 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 [29].

VI. Concluding Remarks

We presented finite-length MIMO equalization methods for "smart" antenna arrays using the statistical theory of canonical correlations. We showed that these methods are closely related to maximum likelihood reduced-rank channel and noise estimation for unknown spatially correlated noise and to several (blind and nonblind) equalization schemes.

Further research will include: simulation and convergence analysis of the proposed algorithms, their adaptive implementation, and rank estimation. It is of interest to consider extensions of the above results to account for temporally correlated interference as well; computationally efficient methods can be derived for this case if we adopt a space-time separable noise model in, e.g., [40].

Appendix A
Derivation of Expressions for Beamforming and Filtering Matrices

We show that under $B\hat{R}_{yy}B^* = I_r$, all the eigenvalues of $(1/N)\cdot[BY - W\Phi(\boldsymbol{\eta})]\cdot[BY - W\Phi(\boldsymbol{\eta})]^*$ are simultaneously minimized for $B = \hat{B}(\boldsymbol{\eta})$ and $W = \hat{W}(\boldsymbol{\eta})$ in (3.3).

First, we obtain the optimal estimate of W for given B as follows:

$$\frac{1}{N}\cdot[BY - W\Phi(\boldsymbol{\eta})]\cdot[BY - W\Phi(\boldsymbol{\eta})]^*$$
$$= B\hat{R}_{yy}B^* - W\hat{R}_{y\phi}^*B^* - B\hat{R}_{y\phi}W^* + W\hat{R}_{\phi\phi}W^* \tag{A.1a}$$
$$= B\hat{R}_{yy}B^* - B\hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1}\hat{R}_{y\phi}^*B^* \tag{A.1b}$$
$$+ (B\hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1} - W)\hat{R}_{\phi\phi}(B\hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1} - W)^*$$
$$\geq B\hat{R}_{yy}B^* - B\hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1}\hat{R}_{y\phi}^*B^* \tag{A.1c}$$

with equality in (A.1c) for $W = B\hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1}$. Using $B\hat{R}_{yy}B^* = I_r$, the expression in (A.1c) further simplifies to $I_m - B\hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1}\hat{R}_{y\phi}^*B^*$. Denote by $\rho_i(P)$ the ith largest eigenvalue of a Hermitian matrix P, i.e., $\rho_1(P) \geq \rho_2(P) \geq \cdots$. Note also that for two Hermitian matrices P and Q such that $P - Q$ is positive semidefinite (i.e., $P \geq Q$), we have $\rho_i(P) \geq \rho_i(Q)$ for all i; see [41, result 1.19, p. 36]. Then

$$\rho_{r-i+1}\left\{\frac{1}{N}\cdot[BY - W\Phi(\boldsymbol{\eta})]\cdot[BY - W\Phi(\boldsymbol{\eta})]^*\right\}$$
$$\geq \rho_{r-i+1}\left\{I_m - B\hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1}\hat{R}_{y\phi}^*B^*\right\} \tag{A.2a}$$
$$= 1 - \rho_i\left\{B\hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1}\hat{R}_{y\phi}^*B^*\right\} \tag{A.2b}$$
$$\geq 1 - \rho_i\left\{\hat{R}_{yy}^{-1/2}\hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1}\hat{R}_{y\phi}^*\hat{R}_{yy}^{-1/2}\right\} \tag{A.2c}$$
$$= 1 - \hat{\lambda}^2(i) \tag{A.2d}$$

where the last inequality follows from the Poincaré separation theorem (see [42, pp. 64 and 65] and [43]). In (A.2c), the equality holds for $B\hat{R}_{yy}^{1/2} = \hat{U}(r)^*$. Then, the results (3.3) easily follow.

We now show that $\hat{B}(\boldsymbol{\eta})$ and $\hat{W}(\boldsymbol{\eta})$ in (3.3) maximize the SINR expression in (3.6). Again, we first find the optimal estimate of W for given B: $W = B\hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1}$ [which follows from (A.1c)], and concentrate the SINR with respect to it:

$$\mathrm{SINR}(\boldsymbol{\eta}, B, B\hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1}) = \frac{|B\hat{R}_{yy}B^*|}{|B\hat{R}_{yy}B^* - B\hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1}\hat{R}_{y\phi}^*B^*|}. \tag{A.3}$$

Without loss of generality, we can choose $B = Z\hat{U}^*\hat{R}_{yy}^{-1/2}$, where Z is an $r \times m$ matrix of full rank r. Then, (A.3) becomes

$$\frac{|ZZ^*|}{|Z\hat{U}^*(I_m - \hat{C}_{y\phi}\hat{C}_{y\phi}^*)\hat{U}Z^*|} = \frac{|ZZ^*|}{|Z(I_m - \hat{\Lambda}\hat{\Lambda}^T)Z^*|} \tag{A.4}$$

which is maximized for $Z = Z_0 = [I_r, 0]$, giving exactly the estimates in (3.3).

Acknowledgment

The authors are grateful to the anonymous reviewers for their helpful comments and to G. Ginis from Stanford University for bringing [6] to our attention.

References

[1] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Mag.*, vol. 14, pp. 49–83, Nov. 1997.

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

[3] G. 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.

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

[5] 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.

[6] J. M. Cioffi and G. D. Forney, "Generalized DFE for packet transmission with ISI and gaussian noise," in *Communications, Computation, Control and Signal Processing*, A. Paulraj, *et al.*, Ed. Norwell, MA: Kluwer, 1997, pp. 79–127.

[7] P. Stoica and M. Viberg, "Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions," *IEEE Trans. Signal Processing*, vol. 44, pp. 3069–3078, Dec. 1996.

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

[9] E. Lindskog, “Space-time processing and equalization for wireless communications,” Ph.D. dissertation, Uppsala Univ., Uppsala, Sweden, 1999.
[10] E. Lindskog and C. Tidestav, “Reduced rank channel estimation,” in *Proc. 49th Veh. Technol. Conf.*, Houston, TX, May 1999, pp. 1126–1130.
[11] D. Giancola, A. Sanguanini, and U. Spagnolini, “Variable rank receiver structures for low-rank space-time channels,” in *Proc. 49th Veh. Technol. Conf.*, Houston, TX, May 1999, pp. 65–69.
[12] E. Lindskog and C. Tidestav, “Reduced rank space-time equalization,” in *Proc. IEEE Int. Symp. PIMRC*, Boston, MA, Sept. 1998, pp. 1081–1085.
[13] D. Chizhik, G. J. Foschini, and R. A. Valenzuela, “Capacities of multi-element transmit and receive antennas: Correlations and keyholes,” *Electron. Lett.*, vol. 36, pp. 1099–1100, June 2000.
[14] D. Gesbert, H. Bolcskei, D. A. Gore, and A. J. Paulraj, “MIMO wireless channels: Capacity and performance prediction,” in *Proc. Globecom Conf.*, San Francisco, CA, Nov. 2000, pp. 1083–1088.
[15] A. Dogandžić and A. Nehorai, “Finite-length MIMO adaptive equalization using canonical correlations,” in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Salt Lake, UT, May 2001, pp. 2149–2152.
[16] L. L. Scharf and J. K. Thomas, “Canonical coordinates and the geometry of inference, rate, and capacity,” *IEEE Trans. Signal Processing*, vol. 46, pp. 647–654, Mar. 1998.
[17] L. L. Scharf and C. T. Mullis, “Wiener filters in canonical coordinates for transform coding, filtering, and quantizing,” *IEEE Trans. Signal Processing*, vol. 48, pp. 824–831, Mar. 2000.
[18] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd ed. New York: Wiley, 1984.
[19] F. Pipon, P. Chevalier, P. Vila, and J. J. Monot, “Joint spatial and temporal equalization for channels with ISI and CCI-theoretical and experimental results for a base station reception,” in *Proc. 1st IEEE Signal Process. Workshop SPAWC*, Paris, France, Apr. 1997, pp. 309–312.
[20] J.-W. Liang, J.-T. Chen, and A. J. Paulraj, “A two-stage hybrid approach for CCI/ISI reduction with space-time processing,” *IEEE Commun. Lett.*, vol. 1, pp. 163–165, Nov. 1997.
[21] M. A. Lagunas, A. I. Perez Neira, and J. Vidal, “Joint beamforming and Viterbi equalizer in wireless communications,” in *Proc. 31st Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 1997, pp. 915–919.
[22] ——, “Optimal array combiner for sequence detectors,” in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Seattle, WA, May 1998, pp. 3341–3344.
[23] M. A. Lagunas, J. Vidal, and A. Pérez-Neira, “Joint array combining and MLSE for single-user receivers in multipath Gaussian multiuser channels,” *IEEE J. Select. Areas Commun.*, vol. 18, pp. 2252–2259, Nov. 2000.
[24] D. Giancola, U. Girola, S. Parolari, A. Picciriello, U. Spagnolini, and D. Vincenzoni, “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.
[25] D. D. Falconer and F. R. Magee, “Adaptive channel memory truncation for maximum likelihood sequence estimation,” *Bell Syst. Tech. J.*, vol. 52, pp. 1541–1562, Nov. 1973.
[26] S. U. H. Qureshi, “Adaptive equalization,” *Proc. IEEE*, vol. 73, pp. 1349–1387, Sept. 1985.
[27] N. Seshadri, “Joint data and channel estimation using blind trellis search techniques,” *IEEE Trans. Commun.*, vol. 42, pp. 1000–1011, Feb.–Apr. 1994.
[28] A. Ranheim, “A decoupled approach to adaptive signal separation using an antenna array,” *IEEE Trans. Veh. Technol.*, vol. 48, pp. 676–682, May 1999.
[29] B. G. Agee, “The least-squares CMA: A new technique for rapid correction of constant modulus signals,” in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Tokyo, Japan, Apr. 1986, pp. 953–956.
[30] A. M. Sayeed and B. Aazhang, “Joint multipath-Doppler diversity in mobile wireless communications,” *IEEE Trans. Commun.*, vol. 47, pp. 123–132, Jan. 1999.
[31] A. Dogandžić and A. Nehorai, “Space-time fading channel estimation and symbol detection in unknown spatially correlated noise,” *IEEE Trans. Signal Processing*, vol. 50, pp. 457–474, Mar. 2002.
[32] Y. B. Hua, M. Nikpour, and P. Stoica, “Optimal reduced-rank estimation and filtering,” *IEEE Trans. Signal Processing*, vol. 49, pp. 457–469, Mar. 2001.
[33] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco, CA: Holden-Day, 1964.
[34] G. E. Bottomley and K. Jamal, “Adaptive arrays and MLSE equalization,” in *Proc. 45th Veh. Technol. Conf.*, Chicago, IL, July 1995, pp. 50–54.
[35] M. Stojanovic, J. Catipovic, and J. G. Proakis, “Adaptive multichannel combining and equalization for underwater acoustic communication,” *J. Acoust. Soc. Amer.*, vol. 94, pp. 1621–1631, Sept. 1993.
[36] C. Tidestav and E. Lindskog, “Bootstrap equalization,” in *Proc. IEEE Int. Conf. Universal Pers. Commun.*, Florence, Italy, Oct. 1998, pp. 1221–1225.
[37] 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.
[38] E. Visotsky and U. Madhow, “Noncoherent multiuser detection for CDMA systems with nonlinear modulation: A non-Bayesian approach,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 1352–1367, May 2001.
[39] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.
[40] J. Vidal, M. Cabrera, and A. Agustin, “Full exploitation of diversity in space-time MMSE receivers,” in *Proc. 52nd Veh. Technol. Conf.*, Boston, MA, Sept. 2000, pp. 2497–2502.
[41] M. S. Srivastava and C. G. Khatri, *An Introduction to Multivariate Statistics* New York, North Holland, 1979.
[42] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley, 1973.
[43] ——, “Separation theorems for singular values of matrices and their applications in multivariate analysis,” *J. Multivariate Anal.*, vol. 9, pp. 362–377, 1979.