CANONICAL CORRELATION ANALYSIS (CCA) ALGORITHMS FOR MULTIPLE DATA SETS: APPLICATION TO BLIND SIMO EQUALIZATION

Javier Vía, Ignacio Santamaría and Jesús Pérez

Dept. of Communications Engineering. University of Cantabria
39005 Santander, Cantabria, Spain
E-mail: {jvia,nacho,jperez}@gtas.dicom.unican.es

ABSTRACT

Canonical Correlation Analysis (CCA) is a classical tool in statistical analysis that measures the linear relationship between two or several data sets. In [1] it was shown that CCA of $M = 2$ data sets can be reformulated as a pair of coupled least squares (LS) problems. Here, we generalize this idea to $M > 2$ data sets. First, we present a batch algorithm to extract all the canonical vectors through an iterative regression procedure, which at each iteration uses as desired output the mean of the outputs obtained in the previous iteration. Furthermore, this alternative formulation of CCA as M coupled regression problems allows us to derive in a straightforward manner a recursive least squares (RLS) algorithm for on-line CCA. The proposed batch and on-line algorithms are applied to blind equalization of single-input multiple-output (SIMO) channels. Some simulation results show that the CCA-based algorithms outperform other techniques based on second-order statistics for this particular application.

1. INTRODUCTION

Canonical Correlation Analysis (CCA) is a well-known technique in multivariate statistical analysis, which has been widely used in economics, meteorology, and in many modern information processing fields, such as communication theory, statistical signal processing, and Blind Source Separation (BSS).

CCA was developed by H. Hotelling [2] as a way of measuring the linear relationship between two multidimensional sets of variables and was later extended to several data sets [3]. Typically, CCA is formulated as a generalized eigenvalue (GEV) problem; however, a direct application of eigendecomposition techniques is often unsuitable for high dimensional data sets as well as for adaptive environments due to their high computational cost.

Recently, several adaptive algorithms have been developed for the case of $M = 2$ data sets [1, 4, 5]. In particular, in [1] an interpretation of CCA as a pair of LS regression problems was exploited to derive batch and on-line algorithms. Here we extend this approach to $M > 2$ data sets. Specifically, we propose an alternative formulation of the Maximum Variance (MAXVAR) generalization of CCA of several data sets proposed in the classic work by Kettenring [3]. Similarly to the $M = 2$ case, this reformulation considers CCA as a set of M coupled least squares problems and can be exploited to derive in a straightforward manner batch and adaptive algorithms.

The proposed CCA algorithms turn out to be particularly suitable for blind equalization of single-input multiple-output (SIMO) channels, which is a common problem encountered in communications, sonar and seismic signal processing. SIMO channels appear either when the signal is oversampled at the receiver or from the use of an array of antennas. It is well known that, if the input signal is informative enough and the FIR channels are co-prime, second order statistics (SOS) are sufficient for blind equalization. In this paper we show that maximizing the correlation among the outputs of the equalizers (i.e. CCA) is a reasonable equalization criterion, which outperforms other well-known blind equalization techniques such as the Modified Second Order Statistic Algorithm (MSOSA) described in [6].

This work was supported by MCYT (Ministerio de Ciencia y Tecnología) under grants TIC2001-0751-C04-03 and TEC2004-06451-C05-02.

2. OVERVIEW OF CCA OF $M = 2$ DATA SETS

Let $\mathbf{X}_1 \in \mathbb{R}^{N\times m_1}$ and $\mathbf{X}_2 \in \mathbb{R}^{N\times m_2}$ be two known full-rank data matrices. Canonical Correlation Analysis (CCA) can be defined as the problem of finding two canonical vectors: $\mathbf{h}_1$ of size $m_1 \times 1$ and $\mathbf{h}_2$ of size $m_2 \times 1$, such that the canonical variates $\mathbf{z}_1 = \mathbf{X}_1\mathbf{h}_1$ and $\mathbf{z}_2 = \mathbf{X}_2\mathbf{h}_2$ are maximally correlated, i.e.,

$$\underset{\mathbf{h}_1,\mathbf{h}_2}{\text{argmax}}\ \rho = \frac{\mathbf{z}_1^T\mathbf{z}_2}{\|\mathbf{z}_1\|\|\mathbf{z}_2\|} = \frac{\mathbf{h}_1^T\mathbf{R}_{12}\mathbf{h}_2}{\sqrt{\mathbf{h}_1^T\mathbf{R}_{11}\mathbf{h}_1\mathbf{h}_2^T\mathbf{R}_{22}\mathbf{h}_2}}, \tag{1}$$

where $\mathbf{R}_{kl} = \mathbf{X}_k^T\mathbf{X}_l$ is an estimate of the crosscorrelation matrix. Problem (1) is equivalent to the following constrained optimization problem

$$\begin{aligned} &\underset{\mathbf{h}_1,\mathbf{h}_2}{\text{argmax}}\ \rho = \mathbf{h}_1^T\mathbf{R}_{12}\mathbf{h}_2 \\ &\text{subject to} \qquad \mathbf{h}_1^T\mathbf{R}_{11}\mathbf{h}_1 = \mathbf{h}_2^T\mathbf{R}_{22}\mathbf{h}_2 = 1. \end{aligned} \tag{2}$$

The solution of this problem is given by the eigenvector corresponding to the largest eigenvalue of the following generalized eigenvalue problem (GEV) [7]

$$\begin{bmatrix} \mathbf{0} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{0} \end{bmatrix}\mathbf{h} = \rho \begin{bmatrix} \mathbf{R}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{22} \end{bmatrix}\mathbf{h}, \tag{3}$$

where ρ is the canonical correlation and $\mathbf{h} = [\mathbf{h}_1^T, \mathbf{h}_2^T]^T$ is the eigenvector. The remaining eigenvectors and eigenvalues of (3) are the subsequent canonical vectors and correlations, respectively. The corresponding canonical variates are maximally correlated and orthogonal for different pairs of canonical vectors.

For the two data sets case, it can be easily proved that constraint (2) is equivalent to

$$\frac{\mathbf{h}_1^T\mathbf{R}_{11}\mathbf{h}_1 + \mathbf{h}_2^T\mathbf{R}_{22}\mathbf{h}_2}{2} = 1.$$

This alternative constraint will be used in the next section to generalize CCA to several complex data sets.

3. CCA OF $M > 2$ DATA SETS

Let $\mathbf{X}_k \in \mathbb{C}^{N\times m_k}$ for $k = 1,\ldots,M$ be full-rank matrices. If we denote the canonical vectors and variables as $\mathbf{h}_k$ and $\mathbf{z}_k = \mathbf{X}_k\mathbf{h}_k$, respectively; and the estimated crosscorrelation matrices as $\mathbf{R}_{kl} = \mathbf{X}_k^H\mathbf{X}_l$, then, the generalization of the CCA problem to $M > 2$ data sets can be formulated as

$$\begin{aligned} &\underset{\mathbf{h}_1,\ldots,\mathbf{h}_M}{\text{argmax}}\ \rho = \frac{1}{M(M-1)}\sum_{\substack{k,l=1\\k\neq l}}^{M}\mathbf{z}_k^H\mathbf{z}_l = \frac{1}{M(M-1)}\sum_{\substack{k,l=1\\k\neq l}}^{M}\mathbf{h}_k^H\mathbf{R}_{kl}\mathbf{h}_l \\ &\text{subject to} \qquad \frac{1}{M}\sum_{k=1}^{M}\mathbf{h}_k^H\mathbf{R}_{kk}\mathbf{h}_k = 1, \end{aligned} \tag{4}$$

and solving by the method of Lagrange multipliers we obtain the following GEV problem

$$\frac{1}{M-1}(\mathbf{R}-\mathbf{D})\mathbf{h} = \rho\mathbf{D}\mathbf{h}, \tag{5}$$

where

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \cdots & \mathbf{R}_{1M} \\ \vdots & \ddots & \vdots \\ \mathbf{R}_{M1} & \cdots & \mathbf{R}_{MM} \end{bmatrix} \quad , \quad \mathbf{D} = \begin{bmatrix} \mathbf{R}_{11} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{R}_{MM} \end{bmatrix}, \tag{6}$$

$\mathbf{h} = [\mathbf{h}_1^T, \ldots, \mathbf{h}_M^T]^T$, and ρ is the generalized canonical correlation. Then, the main CCA solution is obtained as the eigenvector associated to the largest eigenvalue of (5), and the remaining eigenvectors and eigenvalues are the subsequent solutions of the CCA problem.

Although formulated in a different way, in the appendix we prove that problem (4) is equivalent to the Maximum Variance (MAXVAR) generalization of CCA proposed by Kettenring in [3].

Many linear algebra techniques exist in the literature to solve this GEV problem, however, besides their high computational cost, they are not well suited for adaptive processing. In the following we describe a LS framework which avoids the need of eigendecomposition techniques.

4. CCA THROUGH ITERATIVE REGRESSION

4.1 Batch LS Algorithm

Let us start by rewriting (5) as

$$\frac{1}{M}\mathbf{R}\mathbf{h} = \beta\mathbf{D}\mathbf{h}, \tag{7}$$

where

$$\beta = \frac{1+(M-1)\rho}{M}.$$

Now, by noting that $\mathbf{R}_{kk}^{-1}\mathbf{R}_{kl} = \mathbf{X}_k^+\mathbf{X}_l$, where $\mathbf{X}_k^+ = (\mathbf{X}_k^H\mathbf{X}_k)^{-1}\mathbf{X}_k^H$ is the pseudoinverse of $\mathbf{X}_k$, the GEV problem (7) can be viewed as M coupled LS regression problems

$$\beta\mathbf{h}_k = \mathbf{X}_k^+\mathbf{z}, \qquad k = 1, \ldots, M,$$

where $\mathbf{z} = \frac{1}{M}\sum_{k=1}^M \mathbf{X}_k\mathbf{h}_k$.

The key idea of the batch algorithm is to solve these regression problems iteratively: at each iteration t we form M LS regression problems using as desired output

$$\mathbf{z}(t) = \frac{1}{M}\sum_{k=1}^M \mathbf{X}_k\mathbf{h}_k(t-1),$$

and a new solution is thus found by solving

$$\beta(t)\mathbf{h}_k(t) = \mathbf{X}_k^+\mathbf{z}(t) \qquad k = 1, \ldots, M.$$

Finally, $\beta(t)$ and $\mathbf{h}_k(t)$ can be obtained through a straightforward normalization step, which forces $\mathbf{h}(t)$ to satisfy either (4) or $\|\mathbf{h}(t)\| = 1$. The subsequent CCA eigenvectors can be obtained by means of a deflation technique [8], similarly to the technique used in [1], which forces the orthogonality condition among the subsequent solutions $\mathbf{z}(t)$.

It is easy to realize that this technique is equivalent to the well-known power method to extract the main eigenvector and eigenvalue of (7). However, an advantage of this alternative formulation is that it allows us to derive an adaptive CCA algorithm in a straightforward manner.

Initialize $\mathbf{P}_k(0) = \delta^{-1}\mathbf{I}$, with $\delta \ll 1$ for $k = 1, \ldots, M$.
Initialize $\mathbf{h}_i(0)$, $\mathbf{c}_i(0) = \mathbf{0}$ and $\rho_i(0) = 0$ for $i = 1, \ldots, p$.
for $n = 1, 2, \ldots$ **do**
  Update $\mathbf{k}_k(n)$ and $\mathbf{P}_k(n)$ with $\mathbf{x}_k(n)$ for $k = 1, \ldots, M$.
  for $i = 1, \ldots, p$ **do**
    Obtain $z_i(n)$, $\tilde{z}_i(n)$ and $\mathbf{e}_i(n)$.
    Obtain $\beta_i(n)\mathbf{h}_i(n)$ with (8) and update $\mathbf{c}_i(n)$.
    Estimate $\beta_i(n)$ and $\mathbf{h}_i(n)$ considering $\|\mathbf{h}_i(n)\| = 1$.
  end for
end for

Algorithm 1: Summary of the proposed adaptive CCA algorithm.

4.2 On-Line RLS Algorithm

To obtain an on-line algorithm, the LS regression problems are now rewritten as the following cost functions

$$\underset{\beta(n),\mathbf{h}_k(n)}{\text{argmin}} \quad J_k(n) = \sum_{l=1}^n \lambda^{n-l}\left\|z(l) - \beta(n)\mathbf{x}_k^H(l)\mathbf{h}_k(n)\right\|^2,$$

where $z(n) = \frac{1}{M}\sum_{k=1}^M \mathbf{x}_k^H(n)\mathbf{h}_k(n-1)$ is the reference signal, and $0 < \lambda \leq 1$ is the forgetting factor.

A direct application of the RLS algorithm yields, for $k = 1, \ldots, M$

$$\beta(n)\mathbf{h}_k(n) = \beta(n-1)\mathbf{h}_k(n-1) + \mathbf{k}_k(n)e_k(n),$$

where

$$e_k(n) = z(n) - \beta(n-1)\mathbf{x}_k^H(n)\mathbf{h}_k(n-1),$$

is the *a priori* error for the kth data set, and the Kalman gain vector $\mathbf{k}_k(n)$ of the process $\mathbf{x}_k$ is updated with the well-known equations

$$\mathbf{k}_k(n) = \frac{\mathbf{P}_k(n-1)\mathbf{x}_k(n)}{\lambda + \mathbf{x}_k^H(n)\mathbf{P}_k(n-1)\mathbf{x}_k(n)},$$

$$\mathbf{P}_k(n) = \lambda^{-1}\left(\mathbf{I} - \mathbf{k}_k(n)\mathbf{x}_k^H(n)\right)\mathbf{P}_k(n-1),$$

where $\mathbf{P}_k(n) = \mathbf{\Phi}_k^{-1}(n)$ is the inverse of the autocorrelation matrix $\mathbf{\Phi}_k(n) = \sum_{l=1}^n \lambda^{n-l}\mathbf{x}_k(l)\mathbf{x}_k^H(l)$.

To extract the subsequent CCA solutions we resort to a deflation technique, which resembles the APEX algorithm [8] and extends to $M > 2$ data sets the algorithm presented in [1]. Specifically, denoting the estimated ith eigenvector of (7) as $\mathbf{h}_i(n) = [\mathbf{h}_{1,i}^T(n), \ldots, \mathbf{h}_{M,i}^T(n)]^T$, the reference signal is obtained as

$$\tilde{z}_i(n) = z_i(n) - \mathbf{z}_i^H(n)\mathbf{c}_i(n-1),$$

where $\mathbf{z}_i(n) = [z_1(n), \ldots, z_{i-1}(n)]^H$ is a vector containing the extracted signal, $z_i(n)$ is given by

$$z_i(n) = \frac{1}{M}\sum_{k=1}^M \mathbf{x}_k^H(n)\mathbf{h}_{k,i}(n-1),$$

and $\mathbf{c}_i(n-1)$, which imposes the orthogonality conditions among the solutions $z_i(n)$, is updated through the RLS algorithm similarly to [1]. Finally, grouping the *a priori* errors into the vector $\mathbf{e}_i(n) = [e_{1,i}(n), \ldots, e_{M,i}(n)]^T$, we can write the overall algorithm (see Algorithm 1) in matrix form as

$$\beta_i(n)\mathbf{h}_i(n) = \beta_i(n-1)\mathbf{h}_i(n-1) + \mathbf{K}(n)\mathbf{e}_i(n), \tag{8}$$

where

$$\mathbf{K}(n) = \begin{bmatrix} \mathbf{k}_1(n) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{k}_M(n) \end{bmatrix}.$$

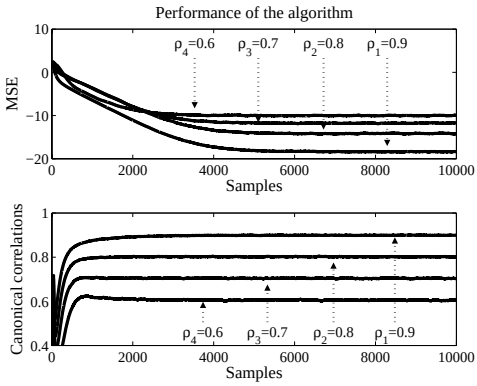

Figure 1: Convergence of the eigenvectors and eigenvalues for the adaptive CCA-RLS algorithm. $\lambda = 0.99$.

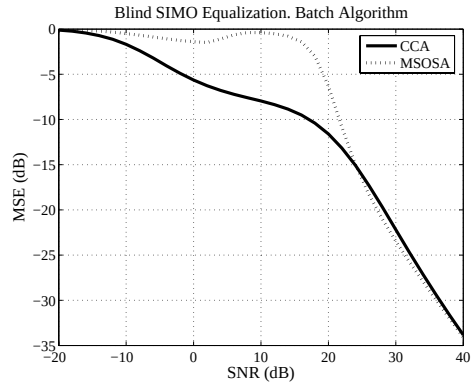

Figure 2: Blind SIMO Equalization. Performance of the batch algorithms with colored signals.

Unlike other recently proposed adaptive algorithms for GEV problems [9], our method is a true RLS algorithm, which uses a reference signal specifically constructed for CCA and derived from the regression framework. This reference signal opens the possibility of new improvements of CCA algorithms: for instance, it can be used to develop robust versions of the algorithm [1], or to construct a soft decision signal useful in blind equalization problems [10].

5. APPLICATION OF CCA TO BLIND EQUALIZATION OF SIMO CHANNELS

An interesting application of CCA is the blind equalization of single-input multiple-output (SIMO) channels. Let us suppose a system where an unknown source signal $s[n]$ is sent through K different and unknown finite impulse response (FIR) channels, $h_k[n]$ for $k = 1, \cdots, K$, with maximum length L. Denoting the observations as $\mathbf{x}_r[n] = [x_1[n], \ldots, x_K[n]]^T$, where $x_k[n] = s[n] * h_k[n]$ is the k-th received signal, the system equations can be written as $\mathbf{x}[n] = \mathbf{H}\mathbf{s}[n]$, where we have used the following definitions

$$\mathbf{h}[n] = [h_1[n], \ldots, h_K[n]]^T,$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}[0] & \cdots & \mathbf{h}[L-1] & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{h}[0] & \cdots & \mathbf{h}[L-1] \end{bmatrix},$$

$$\mathbf{s}[n] = \left[s[n], \cdots, s[n - L_{eq} - L + 2]\right]^T,$$

$$\mathbf{x}[n] = \left[\mathbf{x}_r^T[n], \cdots, \mathbf{x}_r^T[n - L_{eq} + 1]\right]^T,$$

and where L_{eq} is a parameter determining the dimensions of the vectors and matrices. It has been proved in [6] that, under mild assumptions, there exists a matrix $\mathbf{W} = \left[\mathbf{w}_1, \cdots, \mathbf{w}_{L_{eq}+L-1}\right]$, of dimensions $KL_{eq} \times (L_{eq} + L - 1)$, such that

$$\mathbf{w}_k^H \mathbf{x}[n+k] = \mathbf{w}_l^H \mathbf{x}[n+l], \qquad k, l = 1, \ldots, L_{eq} + L - 1. \tag{9}$$

In addition, Eq. (9) holds iff

$$\mathbf{s}[n] = \mathbf{W}^H \mathbf{x}[n].$$

Then, CCA can be applied to the $M = L_{eq} + L - 1$ data sets $\mathbf{x}[n+k]$ to maximize the correlation among the outputs of the M equalizers $\mathbf{w}_k$, and finally, the equalized output $\hat{s}[n]$ will be obtained as the mean of the M outputs. The advantages of CCA over other SOS equalization techniques such as the Modified Second Order Statistics (MSOSA) [6] can be explained from the fact that CCA provides the best one-dimensional PCA representation of unit-norm canonical variates (see the appendix). This produces a better performance for short data registers, a mitigation of the noise enhancement problem for ill-conditioned channels and a faster convergence in time-varying environments.

6. SIMULATION RESULTS

Three examples are shown in this section to illustrate the performance of the CCA algorithms. In all the simulations the results of 300 independent realizations are averaged. The parameters of the algorithm are initialized as follows:

- $\mathbf{P}_k(0) = 10^5\mathbf{I}$, for $k = 1, \cdots, M$.
- $\mathbf{h}_i$ is initialized with random values, for $i = 1, \ldots, p$, where p is the number of canonical vectors of interest.
- $\mathbf{c}_i$ and ρ_i, for $i = 1, \ldots, p$, are initialized to zero.

In the first example, four complex data sets of dimensions $m_1 = 40$, $m_2 = 30$, $m_3 = 20$ and $m_4 = 10$ have been generated. The first four generalized canonical correlations are $\rho_1 = 0.9$, $\rho_2 = 0.8$, $\rho_3 = 0.7$ and $\rho_4 = 0.6$. Fig. 1 shows the results obtained by the RLS-based algorithm with forgetting factor $\lambda = 0.99$. We can see that both the estimated canonical vectors and the estimated canonical correlations converge very fast to the theoretical values.

The next two examples consider the blind SIMO equalization problem: the first one illustrates the performance of the CCA and MSOSA batch algorithms for colored source signals. Specifically, we consider a SIMO system with 2 channels: $\mathbf{h}_1 = [1, 0.5, 0.2]$ and $\mathbf{h}_2 = [1, -0.2, 0.5]$ excited by a speech signal sampled at 7418Hz. Fig. 2 shows that for low and moderate SNRs the batch CCA algorithm clearly outperforms the MSOSA.

In the third example the performance of the adaptive algorithms is analyzed. Now a SIMO system with three channels whose impulse responses are shown in Table 1 and a 16-QAM source signal have been considered. The signal to noise ratio is SNR=30dB and the forgetting factor is $\lambda = 0.95$. The results of Fig. 3 compare the convergence speed of the CCA-RLS algorithm and the MSOSA with different step sizes (the value $\mu = 2 \cdot 10^{-4}$ is the largest one ensuring convergence for all the trials). The improvement in speed over the MSOSA method is remarkable.

7. CONCLUSIONS

In this paper, the problem of CCA of multiple data sets has been reformulated as a set of coupled LS regression problems. It has been proved that the proposed formulation is, in fact, equivalent to the CCA-MAXVAR problem described by Kettenring. However, the LS regression point of view allows us to derive batch an online (RLS-based) adaptive algorithms for CCA in a straightforward manner. The performance of the algorithm has been demonstrated through simulations in blind SIMO channel equalization problems, where the proposed CCA algorithms outperform other blind equalization techniques based on second-order statistics. Further investigation lines include the application of these ideas to blind equalization of multiple-input multiple-output (MIMO) channels, and the extension to nonlinear processing through kernel CCA (KCCA).

n	$\mathbf{h}_1[n]$	$\mathbf{h}_2[n]$	$\mathbf{h}_3[n]$
0	$1.786-j1.989$	$0.245+j0.974$	$0.873-j1.234$
1	$-2.113+j3.153$	$-2.223+j1.595$	$-0.939+j0.914$
2	$0.256+j0.484$	$0.428-j0.485$	$0.302+j0.090$
3	$2.230+j0.109$	$3.061-j0.564$	$1.077-j1.980$
4	$-1.359-j1.326$	$-0.186+j0.199$	$-1.155-j0.238$
5	$-0.665-j2.047$	$-1.089+j0.419$	$0.592+j0.908$
6	$1.198-j2.016$	$0.865+j1.288$	$0.157-j0.497$

Table 1: Impulse response of the SIMO channel used in the second blind SIMO equalization example.

APPENDIX

In this appendix we show that the proposed generalization of the CCA problem to $M > 2$ data sets given by (4) (or equivalently by (5)) is equivalent to the maximum variance (MAXVAR) generalization proposed by Kettenring in [3]. The MAXVAR generalization of CCA is formulated in [3] as the problem of finding a set of vectors $\mathbf{f}_k$ and the corresponding projections $\mathbf{y}_k = \mathbf{X}_k\mathbf{f}_k$, which admit the best possible one-dimensional PCA representation and subject to the constraint $\|\mathbf{y}_k\| = 1$; i.e. the cost function to be minimized is

$$J_{PCA}(\mathbf{f}) = \min_{\mathbf{z},\mathbf{a}} \frac{1}{M}\sum_{k=1}^{M} \|\mathbf{z} - a_k\mathbf{y}_k\|^2 \quad \text{subject to} \quad \|\mathbf{a}\|^2 = M, \quad (10)$$

where $\mathbf{a} = [a_1, \ldots, a_M]^T$ is the PCA vector providing the weights for the best combination of the outputs and $\mathbf{f} = [\mathbf{f}_1^T, \ldots, \mathbf{f}_M^T]^T$. Notice first that the canonical vectors defined in the paper are related to $\mathbf{f}_k$ through $\mathbf{h}_k = a_k\mathbf{f}_k$, whereas $\mathbf{z}_k = a_k\mathbf{y}_k$.

Taking the derivative of (10) with respect to $\mathbf{z}$ and equating to zero we get

$$\mathbf{z} = \frac{1}{M}\mathbf{Y}\mathbf{a}, \quad (11)$$

where matrix $\mathbf{Y}$ has been defined as $\mathbf{Y} = [\mathbf{y}_1|\cdots|\mathbf{y}_M]$.

Now, substituting (11) into (10), the cost function becomes

$$J_{PCA}(\mathbf{f}) = 1 - \frac{\mathbf{a}^H\mathbf{Y}^H\mathbf{Y}\mathbf{a}}{M^2} = 1 - \beta,$$

which is minimized when β is the largest eigenvalue of $\mathbf{Y}^H\mathbf{Y}/M$ and $\mathbf{a}$ is its corresponding eigenvector scaled to satisfy $\|\mathbf{a}\|^2 = M$.

Using the singular value decomposition (SVD) of $\mathbf{X}_k = \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^H$, we can write

$$\mathbf{y}_k = \mathbf{X}_k\mathbf{f}_k = \mathbf{U}_k\mathbf{g}_k$$

where $\mathbf{g}_k = \mathbf{\Sigma}_k\mathbf{V}_k^H\mathbf{f}_k$ is a unit norm vector.

Defining $\mathbf{X} = [\mathbf{X}_1 \cdots \mathbf{X}_M]$ and $\mathbf{U} = [\mathbf{U}_1 \cdots \mathbf{U}_M]$, β can be rewritten as

$$\beta = \frac{1}{M^2}\mathbf{b}^H\mathbf{U}^H\mathbf{U}\mathbf{b},$$

where $\mathbf{b} = [\mathbf{b}_1^T, \ldots, \mathbf{b}_M^T]^T$, with $\mathbf{b}_k = a_k\mathbf{g}_k$, consequently the squared norm of $\mathbf{b}$ is $\|\mathbf{b}\|^2 = M$.

After the SVD, the solution $\mathbf{b}$ that maximizes β is the eigenvector of $\mathbf{U}^H\mathbf{U}/M$ associated to its largest eigenvalue. In order to obtain the CCA solution directly from $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ we write

$$\frac{1}{M}\mathbf{U}^H\mathbf{U}\mathbf{b} = \frac{1}{M}\mathbf{\Sigma}^{-1}\mathbf{V}^H\mathbf{X}^H\mathbf{X}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{b} = \beta\mathbf{b}, \quad (12)$$

where $\mathbf{\Sigma}$ and $\mathbf{V}$ are block-diagonal matrices with elements $\mathbf{\Sigma}_i$ and $\mathbf{V}_i$, respectively.

Left-multiplying (12) by $\mathbf{V}\mathbf{\Sigma}^{-1}$ we have

$$\frac{1}{M}\mathbf{V}\mathbf{\Sigma}^{-2}\mathbf{V}^H\mathbf{X}^H\mathbf{X}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{b} = \beta\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{b}.$$

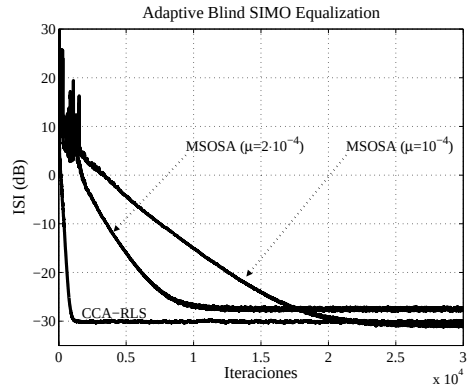

Figure 3: Blind SIMO Equalization. Performance of the adaptive algorithms. $\lambda = 0.95$, SNR=30dB.

Finally, defining $\mathbf{h} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{b}$ and taking into account that $\mathbf{R} = \mathbf{X}^H\mathbf{X}$ and $\mathbf{D} = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^H$ are the matrices defined in (6), the GEV problem (7) is obtained. Obviously, Eq. (7) has the same solutions (eigenvectors) as the proposed generalized CCA problem in (5): this concludes the proof. From this discussion we also find that the best desired output for the proposed regression procedure is given by

$$\mathbf{z} = \frac{1}{M}\mathbf{Y}\mathbf{a} = \frac{1}{M}\mathbf{X}\mathbf{h} = \frac{1}{M}\sum_{k=1}^{M}\mathbf{z}_k.$$

REFERENCES

[1] J. Vía, I. Santamaría and J. Pérez, "A robust RLS algorithm for adaptive Canonical Correlation Analysis". *Proc. of 2005 IEEE International Conference on Acoustics, Speech and Signal Processing*, Philadelphia, USA, March 2005.

[2] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, pp. 321-377, 1936.

[3] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, pp. 433-451, 1971.

[4] A. Pezeshki, M. R. Azimi-Sadjadi and L. L. Scharf, "A network for recursive extraction of canonical coordinates," *Neural Networks*, vol.16, no. 5-6, pp. 801-808, June 2003.

[5] A. Pezeshki, L. L. Scharf, M. R. Azimi-Sadjadi, Y. Hua, "Two-channel constrained least squares problems: solutions using power methods and connections with canonical coordinates," *IEEE Trans. Signal Processing*, vol. 53, pp. 121-135, Jan. 2005.

[6] Y. Li and K. J. R. Liu, "Blind adaptive spatial-temporal equalization algorithms for wireless communications using antenna arrays," *IEEE Communication Letters*, vol. 1, pp. 25-27, Jan. 1997.

[7] M. Borga, *Learning Multidimensional Signal Processing*, PhD thesis, Linköping University, Linköping, Sweden, 1998.

[8] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks, Theory and Applications*, New York: Wiley, 1996.

[9] Y. N. Rao, J. C. Principe and T. F. Wong, "Fast RLS-like algorithm for generalized eigendecomposition and its applications", *J. VLSI Signal Process. Syst. Signal Image Video Technol.*, vol. 37, no. 2, pp. 333-344, 2004

[10] J. Vía, I. Santamaría and J. Pérez, "Adaptive Blind Equalization of SIMO Systems Based On Canonical Correlation Analysis", Submitted to *IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2005)*.