A ROBUST RLS ALGORITHM FOR ADAPTIVE CANONICAL CORRELATION ANALYSIS

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 data sets. In this paper we show that CCA can be reformulated as a pair of coupled least squares (LS) problems. By exploiting this idea, we first present an iterative batch procedure to extract all the canonical vectors through a regression procedure. Then, we derive a Recursive Least Squares (RLS) algorithm for on-line CCA. This algorithm can be further improved to increase its robustness against outliers and impulsive noise. The proposed algorithm is applied to blind identification of multichannel FIR systems, and its performance is illustrated through simulations.

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 [1] as a way of measuring the linear relationship between two multidimensional sets of variables. 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. Although several batch algorithms for the extraction of canonical variates and canonical correlations have been recently presented [2–4], the generalization of these algorithms to real-time signal processing applications has been only suggested [4]. In this paper we exploit the reformulation of CCA as a pair of coupled least squares regression problems to develop a new iterative batch CCA algorithm. Moreover, this batch technique can be easily extended to adaptive CCA by applying a recursive least squares (RLS) algorithm. The proposed on-line algorithm can be viewed as a particular case of a fixed-point method for GEV problems presented in [5]. However, by looking at CCA as a couple of regression problems some advantages appear: first, we are able to derive a true RLS algorithm, secondly, the adaptive algorithm can be used to extract simultaneously all the canonical vectors and, finally, the availability of a reference signal allows us to derive a robust version of the algorithm, which is similar to the Recursive Least M-Estimate algorithm proposed in [6].

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

The proposed adaptive CCA algorithm is applied to the blind identification of single-input multiple-output (SIMO) time-varying channels, which is a common problem encountered in communications, sonar and seismic signal processing.

2. CANONICAL CORRELATION ANALYSIS

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

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

where $\mathbf{R}_{ij} = \mathbf{X}_i^T\mathbf{X}_j$ is an estimate of the correlation matrix. Problem (1) can be formulated as the following constrained optimization problem

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

The solution of (2) is given by the eigenvector corresponding to the largest eigenvalue of the following generalized eigenvalue problem (GEV) [3]

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

where ρ is the maximum correlation between the two sets of variables and $\mathbf{v} = [\mathbf{w}_1^T, \mathbf{w}_2^T]^T$ is the eigenvector. Alternatively, the solution can be obtained as the eigenvector corresponding to the largest eigenvalue of the matrix $\mathbf{C}$

$$\mathbf{C} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{0} \end{bmatrix}.$$

The remaining eigenvectors, $\mathbf{v}_i = [\mathbf{w}_{1,i}^T, \mathbf{w}_{2,i}^T]^T$, and eigenvalues, ρ_i, of $\mathbf{C}$ are the subsequent canonical vectors and correlations respectively. The corresponding canonical variates $\mathbf{y}_{1,i} = \mathbf{X}_1\mathbf{w}_{1,i}$ and $\mathbf{y}_{2,i} = \mathbf{X}_2\mathbf{w}_{2,i}$ are maximally correlated and orthogonal for different pairs of canonical vectors, that is, for $i \neq j$

$$\begin{aligned} \mathbf{y}_{1,i}^T\mathbf{y}_{1,j} &= \mathbf{w}_{1,i}^T\mathbf{R}_{11}\mathbf{w}_{1,j} = 0, \\ \mathbf{y}_{2,i}^T\mathbf{y}_{2,j} &= \mathbf{w}_{2,i}^T\mathbf{R}_{22}\mathbf{w}_{2,j} = 0, \\ \mathbf{y}_{1,i}^T\mathbf{y}_{2,j} &= \mathbf{w}_{1,i}^T\mathbf{R}_{12}\mathbf{w}_{2,j} = 0. \end{aligned} \tag{4}$$

Many linear algebra techniques exist in the literature to solve this problem, however, besides their high computationally cost, they are not well suited for adaptive processing.

3. CCA THROUGH ITERATIVE REGRESSION

3.1. Extraction of the main eigenvector

In this paper we propose a new iterative technique for solving the CCA problem. To develop the method, let us start by noting that $\mathbf{R}_{11}^{-1}\mathbf{R}_{12} = \mathbf{X}_1^{+}\mathbf{X}_2$ and $\mathbf{R}_{22}^{-1}\mathbf{R}_{21} = \mathbf{X}_2^{+}\mathbf{X}_1$, where $\mathbf{X}_j^{+} = (\mathbf{X}_j^T\mathbf{X}_j)^{-1}\mathbf{X}_j^T$ is the pseudoinverse of $\mathbf{X}_j$. Therefore, the GEV problem (3) can be viewed as two coupled LS regression problems

$$\rho\mathbf{w}_1 = \mathbf{X}_1^{+}\mathbf{X}_2\mathbf{w}_2 = \mathbf{X}_1^{+}\mathbf{y}_2, \tag{5}$$

$$\rho\mathbf{w}_2 = \mathbf{X}_2^{+}\mathbf{X}_1\mathbf{w}_1 = \mathbf{X}_2^{+}\mathbf{y}_1. \tag{6}$$

The basic idea of the batch algorithm is to solve both regression problems iteratively: at each iteration we form a LS regression problem using as desired output a linear combination of the canonical variates obtained in the previous iteration. Specifically, at the kth iteration we construct the reference signals $\tilde{\mathbf{y}}_1(k) = (1-\alpha)\mathbf{y}_1(k-1) + \alpha\mathbf{y}_2(k-1)$, and $\tilde{\mathbf{y}}_2(k) = (1-\alpha)\mathbf{y}_2(k-1) + \alpha\mathbf{y}_1(k-1)$ where $0 \le \alpha \le 1$. For instance, for $\alpha = 0.5$ the desired output for the regression problem is the mean of the previous output estimates. Now, the canonical vectors are obtained by solving the following pair of LS problems

$$\underset{\beta(k),\mathbf{w}_1(k)}{\operatorname{argmin}} \quad J_1(k) = \|\tilde{\mathbf{y}}_2(k) - \beta(k)\mathbf{X}_1\mathbf{w}_1(k)\|_2^2, \tag{7}$$

$$\underset{\beta(k),\mathbf{w}_2(k)}{\operatorname{argmin}} \quad J_2(k) = \|\tilde{\mathbf{y}}_1(k) - \beta(k)\mathbf{X}_2\mathbf{w}_2(k)\|_2^2, \tag{8}$$

where $\beta(k) = \alpha + (1-\alpha)\rho(k)$, and whose solutions are given by

$$\beta(k)\mathbf{w}_1(k) = \mathbf{X}_1^{+}\tilde{\mathbf{y}}_2(k), \tag{9}$$

$$\beta(k)\mathbf{w}_2(k) = \mathbf{X}_2^{+}\tilde{\mathbf{y}}_1(k). \tag{10}$$

By grouping the canonical vectors into the eigenvector $\mathbf{v}(k) = [\mathbf{w}_1^T(k), \mathbf{w}_2^T(k)]^T$, and substituting the reference signals $\tilde{\mathbf{y}}_1(k)$ and $\tilde{\mathbf{y}}_2(k)$ into (9) and (10), we arrive to the following equation, which describes the convergence of the algorithm

$$\beta(k)\mathbf{v}(k) = [\alpha\mathbf{I} + (1-\alpha)\mathbf{C}]\,\mathbf{v}(k-1). \tag{11}$$

In this way, after each iteration we obtain a scaled version of the main eigenvector $\beta(k)\mathbf{v}(k)$. Taking into account that $\beta(k) = \alpha + (1-\alpha)\rho(k)$, the correlation coefficient at iteration k can be readily estimated.

Although derived in different way, this technique is equivalent to the well-known power method to extract the main eigenvector and eigenvalue of matrix $\mathbf{C}$. However, as we will show later, this LS regression framework can be modified in a straightforward manner to derive an adaptive CCA algorithm, which can be of interest in applications where the statistics change over time.

3.2. Extraction of the remaining eigenvectors

To extract the remaining canonical vectors we will resort to a deflation technique [7]. Specifically, the regression problems (5) and (6) can be generalized as

$$\rho_i\mathbf{w}_{1,i} = \mathbf{X}_1^{+}\mathbf{X}_2\mathbf{w}_{2,i} = \mathbf{X}_1^{+}\mathbf{y}_{2,i}, \qquad i = 1,2,\ldots,p$$

$$\rho_i\mathbf{w}_{2,i} = \mathbf{X}_2^{+}\mathbf{X}_1\mathbf{w}_{1,i} = \mathbf{X}_2^{+}\mathbf{y}_{1,i}, \qquad i = 1,2,\ldots,p$$

where $p = \min(m_1, m_2)$. This leads to the following iterative equations

$$\beta_i(k)\mathbf{w}_{1,i}(k) = \mathbf{X}_1^{+}\tilde{\mathbf{y}}_{2,i}(k),$$

$$\beta_i(k)\mathbf{w}_{2,i}(k) = \mathbf{X}_2^{+}\tilde{\mathbf{y}}_{1,i}(k).$$

where $\beta_i(k) = \alpha + (1-\alpha)\rho_i(k)$ and $\tilde{\mathbf{y}}_{1,i}(k)$ and $\tilde{\mathbf{y}}_{2,i}(k)$ are the new reference signals, which must be constrained to fulfill the orthogonality conditions (4). Specifically, $\tilde{\mathbf{y}}_{1,i}(k)$ and $\tilde{\mathbf{y}}_{2,i}(k)$ are now constructed as

$$\tilde{\mathbf{y}}_{1,i}(k) = (1-\alpha)\overline{\mathbf{P}}_{1,i}(k)\mathbf{y}_{1,i}(k-1) + \alpha\mathbf{y}_{2,i}(k-1),$$

$$\tilde{\mathbf{y}}_{2,i}(k) = (1-\alpha)\overline{\mathbf{P}}_{2,i}(k)\mathbf{y}_{2,i}(k-1) + \alpha\mathbf{y}_{1,i}(k-1),$$

where $\overline{\mathbf{P}}_{j,i}(k)$ $(j = 1,2)$ denotes the projection matrix onto the subspace orthogonal to the previously extracted canonical variates $\mathbf{Y}_{j,i}(k) = [\mathbf{y}_{j,1}(k), \ldots \mathbf{y}_{j,i-1}(k)]$

$$\overline{\mathbf{P}}_{j,i}(k) = \mathbf{I} - \mathbf{Y}_{j,i}(k)\left(\mathbf{Y}_{j,i}^T(k)\mathbf{Y}_{j,i}(k)\right)^{-1}\mathbf{Y}_{j,i}^T(k).$$

4. ON-LINE RLS ALGORITHM

4.1. Extraction of the main eigenvector

The iterative regression framework introduced in the previous section is applied here to derive an adaptive CCA technique. To obtain an on-line algorithm, the LS regression problems (7) and (8) are now rewritten as the following cost functions

$$\underset{\beta(n),\mathbf{w}_1(n)}{\operatorname{argmin}} \quad J_1(n) = \sum_{l=1}^{n}\lambda^{n-l}\left(\tilde{y}_2(l) - \beta(n)\mathbf{x}_1^T(l)\mathbf{w}_1(n)\right)^2,$$

$$\underset{\beta(n),\mathbf{w}_2(n)}{\operatorname{argmin}} \quad J_2(n) = \sum_{l=1}^{n}\lambda^{n-l}\left(\tilde{y}_1(l) - \beta(n)\mathbf{x}_2^T(l)\mathbf{w}_2(n)\right)^2,$$

where $\tilde{y}_j(n)$ are the reference signals, and $0 < \lambda \le 1$ is the forgetting factor. For notational convenience, and without loss of generality, we will use $\alpha = 0$, i.e. $\tilde{y}_j(n) = \mathbf{x}_j^T(n)\mathbf{w}_j(n-1)$ and $\beta(n) = \rho(n)$, then, a direct application of the RLS algorithm yields, for $j = 1,2$

$$\rho(n)\mathbf{w}_j(n) = \rho(n-1)\mathbf{w}_j(n-1) + \mathbf{k}_{\mathbf{x}_j}(n)\tilde{e}_j(n), \tag{12}$$

where

$$\tilde{e}_1(n) = \tilde{y}_2(n) - \rho(n-1)\mathbf{x}_1^T(n)\mathbf{w}_1(n-1), \tag{13}$$

$$\tilde{e}_2(n) = \tilde{y}_1(n) - \rho(n-1)\mathbf{x}_2^T(n)\mathbf{w}_2(n-1), \tag{14}$$

are the *a priori* errors, and the Kalman gain vector $\mathbf{k}_{\mathbf{x}_j}(n)$ of the process $\mathbf{x}_j$ is updated with the well-known equations

$$\mathbf{k}_{\mathbf{x}_j}(n) = \frac{\mathbf{P}_{\mathbf{x}_j}(n-1)\mathbf{x}_j(n)}{\lambda + \mathbf{x}_j^T(n)\mathbf{P}_{\mathbf{x}_j}(n-1)\mathbf{x}_j(n)},$$

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

Initialize $\mathbf{P}_{\mathbf{x}_1}, \mathbf{P}_{\mathbf{x}_2}, \mathbf{P}_{\mathbf{y}_{1,i}}, \mathbf{P}_{\mathbf{y}_{2,i}}, \rho_i$ and $\mathbf{u}_i$ for $i = 1, 2, \ldots, p$.
for $n = 1, 2, \ldots$ **do**
  Obtain $\mathbf{x}_j(n)$ and update $\mathbf{k}_{\mathbf{x}_j}$ and $\mathbf{P}_{\mathbf{x}_j}$ for $j = 1, 2$.
  for $i = 1, 2, \ldots, p$ **do**
    Update $\mathbf{k}_{\mathbf{y}_{j,i}}$ and $\mathbf{P}_{\mathbf{y}_{j,i}}$ with $\mathbf{y}_{j,i}(n)$ for $j = 1, 2$.
    Apply (15) to update ρ_i and $\mathbf{u}_i$ with $\|\mathbf{u}_i\| = 1$.
    Update $y_{1,i}(n)$ and $y_{2,i}(n)$.
  end for
end for

Algorithm 1: Summary of the proposed adaptive CCA Algorithm.

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

Although derived in a different way, this procedure is equivalent to the fixed-point algorithm for generalized eigendecomposition (GED) proposed in [5]. However, it is interesting to point out that our method is a true RLS algorithm, which uses a reference signal specifically constructed for CCA. This reference signal can be used, for instance, to develop a robust version of the algorithm as we will describe later.

4.2. Extraction of the remaining eigenvectors

The generalization of Eq. (12) for multiple pairs of canonical vectors is, for $j = 1, 2$

$$\rho_i(n)\mathbf{w}_{j,i}(n) = \rho_i(n-1)\mathbf{w}_{j,i}(n-1) + \mathbf{k}_{\mathbf{x}_j}(n)\tilde{e}_{j,i}(n),$$

where the *a priori* errors are defined again as (13) and (14), but now the reference signals are obtained by means of a deflation technique, which resembles the APEX algorithm [7]

$$\tilde{y}_{j,i}(n) = \mathbf{x}_j^T(n)\mathbf{w}_{j,i}(n-1) - \mathbf{y}_{j,i}^T(n)\rho_i(n-1)\mathbf{c}_{j,i}(n-1),$$

where $\mathbf{y}_{j,i}(n) = [y_{j,1}(n), \ldots, y_{j,i-1}(n)]^T$ contains the canonical variates, i.e., $y_{j,i}(n) = \mathbf{x}_j^T(n)\mathbf{w}_{j,i}(n)$, and $\rho_i(n-1)\mathbf{c}_{j,i}(n-1)$ imposes the orthogonality conditions (4).

Similarly to principal component analysis (PCA) [7], there are many techniques to update the coefficients $\rho_i(n-1)\mathbf{c}_{j,i}(n)$, in this work we propose the following RLS-based update procedure

$$\rho_i(n)\mathbf{c}_{j,i}(n) = \rho_i(n-1)\mathbf{c}_{j,i}(n-1) + \mathbf{k}_{\mathbf{y}_{j,i}}(n)\tilde{y}_{j,i}(n),$$

where $\tilde{y}_{j,i}(n)$ can be seen as the *a priori* error of the new RLS problem. Finally, we can write the overall algorithm (see Algorithm 1) in matrix form as

$$\begin{aligned}\rho_i(n)\mathbf{u}_i(n) = \rho_i(n-1)\mathbf{u}_i(n-1) + \\ + \mathbf{K}_i(n)\mathbf{\Gamma}_i(n-1)\mathbf{Z}_i^T(n)\mathbf{u}_i(n-1),\end{aligned} \qquad (15)$$

where $\mathbf{u}_i(n) = [\mathbf{w}_{1,i}^T(n), \mathbf{w}_{2,i}^T(n), \mathbf{c}_{1,i}^T(n), \mathbf{c}_{2,i}^T(n)]^T$, and

$$\mathbf{K}_i(n) = \begin{bmatrix} \mathbf{k}_{\mathbf{x}_1}(n) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{k}_{\mathbf{x}_2}(n) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{k}_{\mathbf{y}_{1,i}}(n) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{k}_{\mathbf{y}_{2,i}}(n) \end{bmatrix},$$

$$\mathbf{\Gamma}_i(n) = \begin{bmatrix} -\rho_i(n) & 1 & 0 & -\rho_i(n) \\ 1 & -\rho_i(n) & -\rho_i(n) & 0 \\ 1 & 0 & -\rho_i(n) & 0 \\ 0 & 1 & 0 & -\rho_i(n) \end{bmatrix},$$

$$\mathbf{Z}_i(n) = \begin{bmatrix} \mathbf{x}_1(n) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_2(n) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{y}_{1,i}(n) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{y}_{2,i}(n) \end{bmatrix}.$$

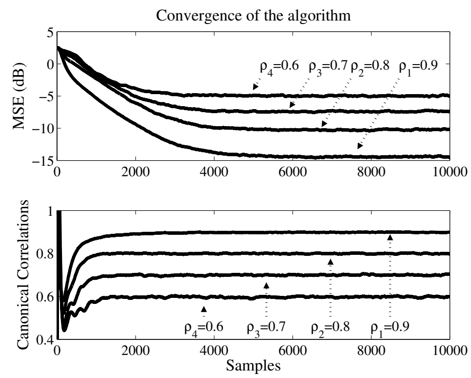

Fig. 1. Convergence of the eigenvectors and eigenvalues.

4.3. On-line algorithm robust to outliers

An advantage of the proposed LS framework is that it allows the development of CCA algorithms robust to outliers. In particular, we propose to apply the RLS algorithm only if the error between the estimated canonical variates

$$e_i(n) = \mathbf{x}_1^T(n)\mathbf{w}_{1,i}(n-1) - \mathbf{x}_2^T(n)\mathbf{w}_{2,i}(n-1),$$

is under an adaptive threshold $\xi_i(n)$. In this way, this procedure is equivalent to the recently proposed Recursive Least M-Estimate algorithm [6], where the classical LS cost function is replaced by the modified Huber M-Estimate function. The adaptive selection of the threshold is based on the median filter proposed in [6]

$$\xi_i(n) = k_{\xi_i}\sigma_i(n),$$

where k_{ξ_i} is a parameter that controls the outlier suppression level and $\sigma_i(n)$ is estimated as

$$\sigma_i^2(n) = \lambda_{\sigma_i}\sigma_i^2(n-1) + (1-\lambda_{\sigma_i})c_1\text{med}(A_{e_i}(n)),$$

where $A_{e_i}(n) = [e_i^2(n), \ldots, e_i^2(n-N_w+1)]$ denotes a window of errors of length N_w, and $c_1 = 1.483(1 + 5/(N_w - 1))$.

5. SIMULATION RESULTS

Three examples are shown in this section to illustrate the performance of the proposed method. In all the simulations we show the averaged results of 300 independent simulations. Finally, the initialization parameters for the three examples are

- $\mathbf{P}_{\mathbf{x}_1}, \mathbf{P}_{\mathbf{x}_2}, \mathbf{P}_{\mathbf{y}_{1,i}}$ and $\mathbf{P}_{\mathbf{y}_{2,i}}$, for $i = 1, \ldots, p$, are initialized as $10^5\mathbf{I}$, where $\mathbf{I}$ is the identity matrix.
- $\mathbf{w}_{1,i}$ and $\mathbf{w}_{2,i}$, for $i = 1, \ldots, p$, are initialized as random vectors.
- $\mathbf{c}_{1,i}, \mathbf{c}_{2,i}, \rho_i$ and σ_i, for $i = 1, \ldots, p$, are initialized to zero.

5.1. Algorithm convergence

In the first example we consider a stationary environment without outliers. We simulate two data sets of dimensions $m_1 = 30$ and $m_2 = 20$ for which the first four canonical correlations are $\rho_1 = 0.9$, $\rho_2 = 0.8$, $\rho_3 = 0.7$ and $\rho_4 = 0.6$. The application of the RLS-based algorithm with forgetting factor $\lambda = 0.99$ gives the results shown in Fig. 1, where we can see that both the estimated canonical vectors and the estimated canonical correlations, converge very fast to the theoretical values.

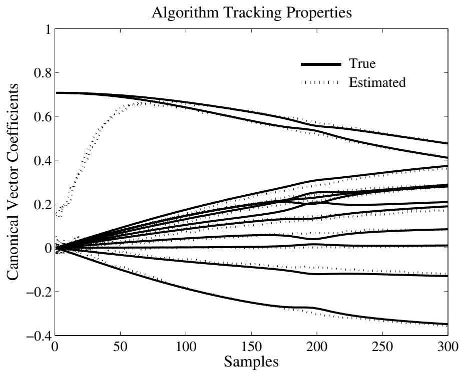

Fig. 2. Tracking properties of the RLS-CCA algorithm.

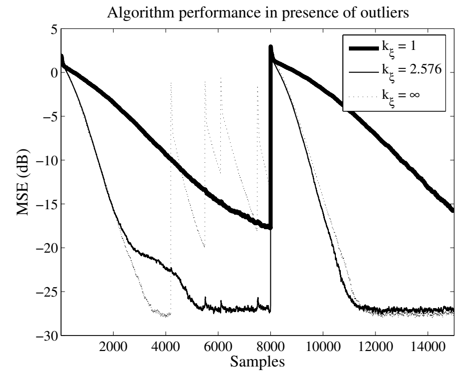

Fig. 3. Performance in presence of outliers.

5.2. Tracking properties

The second example considers a blind SIMO channel estimation problem [8], but in a time-varying environment. The channels change linearly their coefficients in order to evaluate the tracking behavior of the algorithm. The initial channels are $\mathbf{h}_1(0) = [1, 0, 0, 0, 0, 0]$ and $\mathbf{h}_2(0) = [0, 0, 0, 0, 0, 1]$, whereas the channels after 300 samples are $\mathbf{h}_1(300) = [1, 0.6, -0.3, 0, 0.4, 0.8]$ and $\mathbf{h}_2(300) = [0.7, -0.9, 0.7, 0.5, 0.2, 1]$. Both channels are driven by a common white Gaussian signal, and their outputs are corrupted by white Gaussian noise to get a final $SNR = 20dB$. Our goal is to obtain the main canonical vectors between the outputs of both channels: this is an alternative estimation technique to that proposed in [8]. Figure 2 shows the evolution of the true and estimated canonical vectors weights using a forgetting factor of $\lambda = 0.9$.

5.3. Performance in the presence of outliers

In the final example we consider again a SIMO channel estimation problem but corrupted now by impulsive noise. The channels are

$$\mathbf{h}_1 = [0.8, -0.4, 0.6, -0.8, 1, -0.8, 0.6, -0.4, 0.8],$$
$$\mathbf{h}_2 = [0.7, 0.2, -0.3, 0.5, -1, 0.5, -0.3, 0.2, 0.7],$$

and the signal to noise ratio is $SNR = 30dB$. In addition to the additive white Gaussian noise, some outliers are added at some specific samples. In particular, outliers modeled as Gaussian noise with standard deviation $\sigma = 10$ are added at samples 4200 and 7500 for channel 1 and at samples 5500 and 6100 for channel 2. Additionally, the channel response $\mathbf{h}_2$ abruptly changes to $-\mathbf{h}_2$ at sample 8000. The forgetting factor is $\lambda = 0.95$, the median filter window length is $N_w = 20$, and three different values of k_{ξ_i} have been used. The results are shown in Fig. 3, where we can see the trade-off between the convergence rate and the outlier suppression level. From Fig. 3 we also conclude that the selection of an appropriate value of k_{ξ_i} plays a determinant role in the performance of the algorithm in the presence of outliers. For instance, the algorithm with $k_{\xi_i} = 2.576$ outperforms the algorithms with $k_{\xi_i} = \infty$ (no outlier suppression) and $k_{\xi_i} = 1$ (very low threshold thus meaning a higher false alarm probability and correspondingly a slower convergence).

6. CONCLUSIONS

In this paper we have developed a new RLS algorithm for adaptive CCA, which is based on the reformulation of CCA as a pair of coupled LS regression problems. By changing the LS cost function by a Huber M-Estimate function, the proposed algorithm offers an increased robustness against outliers and impulsive noise. The performance of the proposed algorithm has been demonstrated through simulations in time-varying blind SIMO channel identification problems. Further investigation lines include the extension of the proposed method to kernel CCA (KCCA), and the generalization to multiple data sets.

7. REFERENCES

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

[2] D. R. Hardoon, S. Szedmak and J. Shawe-Taylor, "Canonical correlation analysis; An overview with application to learning methods". *Neural Computation*, vol. 16, no. 12, pp.2639-2664, December 2004.

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

[4] M. A. Hasan, "Information Criteria for Reduced Rank Canonical Correlation Analysis", *Proc. of the Int. Joint Conf. on Neural Networks, 2004*, Budapest, Hungary, July 2004.

[5] 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

[6] S.C. Chan and Y. Zou, "A Recursive Least M-Estimate Algorithm for Robust Adaptive Filtering in Impulsive Noise: Fast Algorithm and Convergence Performance Analysis", *IEEE Trans. Signal Processing*, vol. 52, no. 4, pp. 975-991, April 2004.

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

[8] G. Xu, H. Liu, L. Tong, T. Kailath, "A least-squares approach to blind channel equalization". *IEEE Trans. Signal Processing*, vol. 43, pp. 2982-2993, Dec. 1995.