An Extension of the Canonical Correlation Analysis to the Case of Multiple Observations of Two Groups of Variables

Ronald Phlypo, Marco Congedo

An Extension of the Canonical Correlation Analysis to the Case of Multiple Observations of Two Groups of Variables

Ronald Phlypo and Marco Congedo

***Abstract*— In this contribution we present a method that extends the Canonical Correlation Analysis for two groups of variables to the case of multiple conditions. Contrary to the extensions in literature based on augmenting the number of variable groups, the addition of conditions allows for a more robust estimate of the canonical correlation structure inherently present in the data. Algorithms to solve the estimation problem are based on joint approximate diagonalization algorithms for matrix sets. Simulations show the performance of the proposed method under two different scenarios: the calculation of a latent canonical structure and the estimation of a bilinear mixture model.**

I. INTRODUCTION

In this work we present a method to extend the classical Canonical Correlation Analysis (CCA) of Hotelling for a single condition on two groups of variables [5] to the case of multiple conditions for two groups of variables. While a lot of work has focused on the extension of Hotelling's original proposal to multiple groups of variables, only little work has been carried out with respect to the multiplication of the conditions. However, many applications in engineering behave under this model, such as simultaneous recordings in two heterogeneous measurement spaces. Examples may be the simultaneous recordings of extracranial and intracranial electroencephalographic or electrocardiographic data or even the simultaneous recording of electroencephalographic data and the GAZE direction. In this contribution we aim at sketching the major contributions to CCA during the last decades and we propose a variant as to include the multiple conditions. Simulation studies will show the performance of the proposed algorithm under varying conditions.

II. METHODS

A. Canonical Correlation Analysis

Canonical Correlation Analysis (CCA) is one of the most useful methods for describing linear relationships between (the scores of) two groups of variables. Originally proposed in Hotelling's papers [5], it has ever since received much attention in psychometrics, chemometrics and other scientific domains related to the search for explanatory, latent variables underlying multiple observations. The basic CCA model for two groups of variables $\mathbf{X}_1 \in \mathbb{R}^{m_1 \times T}$ and $\mathbf{X}_2 \in \mathbb{R}^{m_2 \times T}$ is given as

$$\max_{\mathbf{t}_1, \mathbf{t}_2} \mathbf{t}_1^T \mathbf{P}_1^T \mathbf{P}_2 \mathbf{t}_2 \ , \tag{1}$$

where $\mathbf{P}_i$ is an orthogonal, unit L_2-norm basis for $\mathbf{X}_i$, which can be obtained as $\mathbf{P}_i = \mathbf{W}_i^T \mathbf{X}_i$ through, e.g., a singular value decomposition of $\mathbf{X}_i$. In other words, the CCA model strives to find those weighting vectors $\mathbf{t}_i$ that maximise the correlation $\langle \mathbf{P}_1 \mathbf{t}_1, \mathbf{P}_2 \mathbf{t}_2 \rangle$.

From the seminal work of Hotelling on, many efforts have been conducted to generalize model (1) to multiple observation sets. Horst [4] maximized the average correlation coefficient between linear combinations, i.e.,

$$\max_{\{\mathbf{t}_i\}} \sum_i \sum_{j \neq i} \mathbf{t}_i^T \mathbf{P}_i^T \mathbf{P}_j \mathbf{t}_j \tag{2}$$

and in the works of Carroll [2] we find an explicit maximization of the correlation between $\mathbf{P}_i \mathbf{t}_i$ and the auxiliary variable $\mathbf{z}$ by maximizing the function

$$f(\mathbf{z}) = \frac{\mathbf{z}^T \left(\sum_i \mathbf{X}_i^T (\mathbf{X}_i \mathbf{X}_i^T)^{-1} \mathbf{X}_i \right) \mathbf{z}}{\mathbf{z}^T \mathbf{z}}$$

over $\mathbf{z}$. In fact, the maximizer of the above is the eigenvector of $\sum_i \mathbf{X}_i^T (\mathbf{X}_i \mathbf{X}_i^T)^{-1} \mathbf{X}_i$ associated with its largest eigenvalue. An effort to regroup the above and similar algorithms using a limited number of representative functions and a single algorithmic solution can be found in [3]. But all of the above methods deal with the presence of multiple groups of data and only little attention has been given to the observation of two groups of variables under several conditions.

In addition, all of the previously introduced algorithms use a summing over the different correlation matrices. Unfortunately, by summing over the different correlation structures, information may go lost. Recent advances in (multi-)linear algebra and signal processing have shown that the joint diagonalization of the matrix set containing all correlation structures offers a more stable solution. This can be seen from the fact that the joint diagonalization of the matrix set is less biased than the diagonalization of the sum, since a perfect diagonalization of the sum $\mathbf{R}$ also diagonalizes the imperfections due to the integrated noise. A simple example may illustrate this phenomenon:

Example 1 (Matrix sum diagonalization): Take two (semi-positive definite) random symmetric matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{2 \times 2}$ and calculate a diagonalization of their sum.

This work has been supported through grant GAZE/EEG of the National Research Agency (ANR), France

R. Phlypo and M. Congedo are with the Vision and Brain Signal processing (ViBs) research group, GIPSA Lab, INPG/UMR 5216 CNRS, BP 46, 961, Rue de la Houille Blanche, 38402 Saint Martin d'Hères, France. ronald.phlypo@gipsa-lab.grenoble-inp.fr, marco.congedo@gmail.com

Now, by simple calculations we see that

$$\left\{\begin{array}{rcl}\mathbf{w}_1^T\mathbf{A}\mathbf{w}_2 &=& 0\\ \mathbf{w}_1^T\mathbf{B}\mathbf{w}_2 &=& 0\\ \mathbf{w}_2^T\mathbf{A}\mathbf{w}_1 &=& 0\\ \mathbf{w}_2^T\mathbf{B}\mathbf{w}_1 &=& 0\end{array}\right. \Rightarrow \left\{\begin{array}{rcl}\mathbf{w}_1^T(\mathbf{A}+\mathbf{B})\mathbf{w}_2 &=& 0\\ \mathbf{w}_2^T(\mathbf{A}+\mathbf{B})\mathbf{w}_1 &=& 0\end{array}\right. ,$$

where $\mathbf{w}_i$ is the i-th column of $\mathbf{W}$. But the inverse does not generally hold true! Actually, the conditions on the sum are necessary but not sufficient for the conditions on the individual terms to hold.

The diagonalization of a single matrix $\mathbf{R}$ has another important drawback, namely, the rotational ambiguity. It can be easily proved that if $\mathbf{W}$ diagonalizes $\mathbf{R}$ as $\mathbf{\Lambda} = \mathbf{W}^T\mathbf{R}\mathbf{W}$, then so do all $\mathbf{T} = \mathbf{W}\mathbf{Q}$, where $\mathbf{Q}$ is an arbitrary orthogonal matrix. For the above example, this means that there does exist a matrix $\mathbf{T} = \mathbf{W}\mathbf{Q}$ that diagonalizes the sum $(\mathbf{A}+\mathbf{B})$ and $\mathbf{A}$ and by consequence also $\mathbf{B}$. However, this does not extend to the case of more than two matrices, since it cannot be guaranteed that the eigenvalues remain real. This is one of the reasons why joint approximate diagonalization of a matrix set has received that much attention in the signal processing literature over the last decades. In what follows, we will show how the joint approximate diagonalization can also be employed for canonical correlation analysis when multiple observations are available of the same two groups of variables under various conditions.

B. Multiple Condition Canonical Correlation Analysis

Remark that maximizing the correlation of the observed sets with an auxiliary variable $\mathbf{z}$ does no longer make sense when we consider multiple conditions. Indeed, the response of a variable may change from condition to condition and it is no longer of primordial interest to maximize this correlation. Denote by $\mathbf{R}_{\mathbf{xx}}^{(k)}$ and $\mathbf{R}_{\mathbf{yy}}^{(k)}$ the covariance matrices under condition k of the variable group $\mathbf{x} \in \mathbb{R}^M$ and $\mathbf{y} \in \mathbb{R}^N$, respectively and denote by $\mathbf{R}_{\mathbf{xy}}^{(k)}$ their cross-covariances.

The joint approximate diagonalization of a matrix set can now directly (and independently) be carried out on $\mathbf{R}_{\mathbf{xx}}^{(k)}$ and $\mathbf{R}_{\mathbf{yy}}^{(k)}$ by requiring that the sets $\{\mathbf{W}_\mathbf{x}^T\mathbf{R}_{\mathbf{xx}}^{(k)}\mathbf{W}_\mathbf{x}\}_k$ and $\{\mathbf{W}_\mathbf{y}^T\mathbf{R}_{\mathbf{yy}}^{(k)}\mathbf{W}_\mathbf{y}\}_k$ are approximately diagonal. However, this does not guarantee the approximate joint diagonality of the set $\{\mathbf{W}_\mathbf{x}\mathbf{R}_{\mathbf{xy}}^{(k)}\mathbf{W}_\mathbf{y}\}_k$. As a consequence, we must impose the following ($\forall k$):

$$\left\{\begin{array}{l}\mathbf{W}_\mathbf{x}^T\mathbf{R}_{\mathbf{xx}}^{(k)}\mathbf{W}_\mathbf{x} \approx \mathbf{\Lambda}_{\mathbf{xx}}^{(k)}\\ \mathbf{W}_\mathbf{y}^T\mathbf{R}_{\mathbf{yy}}^{(k)}\mathbf{W}_\mathbf{y} \approx \mathbf{\Lambda}_{\mathbf{yy}}^{(k)}\\ \mathbf{W}_\mathbf{x}^T\mathbf{R}_{\mathbf{xy}}^{(k)}\mathbf{W}_\mathbf{y} \approx \mathbf{\Lambda}_{\mathbf{xy}}^{(k)}\end{array}\right. ,$$

with $\mathbf{\Lambda}_{\mathbf{xx}}^{(k)}, \mathbf{\Lambda}_{\mathbf{yy}}^{(k)}$ and $\mathbf{\Lambda}_{\mathbf{xy}}^{(k)}$ the appropriate diagonal matrices. In what follows, we will drop the superscript $\cdot^{(k)}$ to facilitate reading. All expressions should be understood $\forall k$.

Defining

$$\begin{array}{rcl}\mathbf{R}_{\mathbf{xy}}\mathbf{W}_\mathbf{y}\mathbf{W}_\mathbf{y}^T\mathbf{R}_{\mathbf{xy}}^T &=& \tilde{\mathbf{R}}_{\mathbf{xx}}\\ \mathbf{R}_{\mathbf{xy}}^T\mathbf{W}_\mathbf{x}\mathbf{W}_\mathbf{x}^T\mathbf{R}_{\mathbf{xy}} &=& \tilde{\mathbf{R}}_{\mathbf{yy}}\end{array}$$

it follows from the above that if $\mathbf{W}_\mathbf{x}^T\mathbf{R}_{\mathbf{xy}}\mathbf{W}_\mathbf{y}$ is approximately diagonal, then so are $\mathbf{W}_\mathbf{x}^T\tilde{\mathbf{R}}_{\mathbf{xx}}\mathbf{W}_\mathbf{x}$ and $\mathbf{W}_\mathbf{y}^T\tilde{\mathbf{R}}_{\mathbf{yy}}\mathbf{W}_\mathbf{y}$. Unfortunately, the inverse does not in general hold true. Although we have no proof, we conjecture that $\mathbf{W}_\mathbf{x}^T\mathbf{R}_{\mathbf{xy}}\mathbf{W}_\mathbf{y}$ at least approximately takes the form of a permutation matrix completed with zero columns ($M < N$) or rows ($M > N$). In other words, each row and each column of $\mathbf{W}_\mathbf{x}^T\mathbf{R}_{\mathbf{xy}}\mathbf{W}_\mathbf{y}$ has maximally one element that is approximately 1 in absolute value, with all other values in that row (respectively column) approximately zero. In the appendix, we give the outline of a possible proof.

The joint approximate diagonalization may now alternatingly be carried out on the two sets $\{\tilde{\mathbf{R}}_{\mathbf{xx}}^{(k)}, \mathbf{R}_{\mathbf{xx}}^{(k)}\}$ and $\{\tilde{\mathbf{R}}_{\mathbf{yy}}^{(k)}, \mathbf{R}_{\mathbf{yy}}^{(k)}\}$. To that purpose one can choose one of the available joint approximate diagonalization algorithms readily available in the literature such as [1], [8], [6]. We propose to alternatingly run one step of the chosen algorithm on each of the sets to avoid getting stuck in local minima.

C. Weighted Multiple Condition Canonical Correlation Analysis

Actually, the above trick is reminiscent to the trick used in [7] for spatio-temporal blind source separation. Let us, in analogy to [7], introduce the weighting value α, which balances between an independent joint approximate diagonalization in each of the measurement spaces and the diagonalization of the cross-covariance matrices. In other words, the sets of diagonalization matrices take the form $\{\alpha\tilde{\mathbf{R}}_{\mathbf{xx}}^{(k)}, (1-\alpha)\mathbf{R}_{\mathbf{xx}}^{(k)}\}$ and $\{\alpha\tilde{\mathbf{R}}_{\mathbf{yy}}^{(k)}, (1-\alpha)\mathbf{R}_{\mathbf{yy}}^{(k)}\}$. Remark that the above paragraph considered the form $\alpha = 0.5$.

D. A Welcome Byproduct

Suppose we consider the measurement spaces as the two dimensions in which we recorded e.g. an event-related potential (ERP). The different conditions then simply reduce to the different trials under which we have recorded the ERP. Our signals may be represented as $\mathbf{X}^{(k)} = \mathbf{R}_{\mathbf{xy}}$, where $\mathbf{x}$ accounts for the observations along the spatial dimension and $\mathbf{y}$ for the observations along the dimension of relative time after stimulation onset. The matrices $\mathbf{R}_{\mathbf{xx}}$ and $\mathbf{R}_{\mathbf{yy}}$ can then simply be obtained by marginalisation over the temporal, respectively the spatial dimension as $\mathbf{R}_{\mathbf{xx}} = E\{\mathbf{X}\mathbf{X}^T\}$ and $\mathbf{R}_{\mathbf{yy}} = E\{\mathbf{X}^T\mathbf{X}\}$. The joint diagonalization of $\mathbf{R}_{\mathbf{yy}}$, $\mathbf{R}_{\mathbf{xx}}$ and $\mathbf{R}_{\mathbf{xy}}$ is then closely related to an approximate singular value decomposition of the matrix set $\{\mathbf{X}^{(k)}\}$, possibly with non-orthogonal changes of bases (if we choose a non-orthogonal joint diagonalization algorithm, e.g. [8], [6]).

III. SIMULATIONS

Since we would like to display a wide variety of possible applications, we have chosen two scenarios as the bases for our simulations. All simulation suppose $M = 3, N = 5$ and $K = 25$. For each observation k we have a population of 10^3 samples. We used a non-orthogonal fast approximate joint diagonalization algorithm of Tichavský and Yeredor [8] alternating the iterations between the two sets $\{(1-\alpha)\mathbf{R}_{\mathbf{xx}}, \alpha\tilde{\mathbf{R}}_{\mathbf{xx}}\}$ and $\{(1-\alpha)\mathbf{R}_{\mathbf{yy}}, \alpha\tilde{\mathbf{R}}_{\mathbf{yy}}\}$ and repeated the experiment over 100 Monte Carlo realisations.

In both the scenarios we investigate the diagonality (up to permutation and scale) of the final matrices $\mathbf{W}_\mathbf{x}^T\mathbf{R}_{\mathbf{xx}}\mathbf{W}_\mathbf{x}$, $\mathbf{W}_\mathbf{y}^T\mathbf{R}_{\mathbf{yy}}\mathbf{W}_\mathbf{y}$ and $\mathbf{W}_\mathbf{x}^T\mathbf{R}_{\mathbf{xy}}\mathbf{W}_\mathbf{y}$ using the Moreau-Amari performance index for a matrix $\mathbf{G} \in \mathbb{R}^{D\times D}$ defined as

$$PI(\mathbf{G}) = \frac{1}{2D(D-1)}\Big(\sum_{j=1}^{D}\frac{\sum_{i=1}^{D}|g_{ij}| - \max_i(|g_{ij}|)}{\sum_{i=1}^{D}|g_{ij}|}\cdots + \sum_{i=1}^{D}\frac{\sum_{j=1}^{D}|g_{ij}| - \max_j(|g_{ij}|)}{\sum_{j=1}^{D}|g_{ij}|}\Big) \quad . \tag{3}$$

For the rectangular matrices $\mathbf{W}_\mathbf{x}^T\mathbf{R}_{\mathbf{xy}}\mathbf{W}_\mathbf{y}$ we calculate only the term corresponding to its smallest dimension, i.e., for $M < N$

$$PI(\mathbf{G}) = \frac{1}{D(D-1)}\left(\sum_{i=1}^{D}\frac{\sum_{j=1}^{D}|g_{ij}| - \max_j(|g_{ij}|)}{\sum_{j=1}^{D}|g_{ij}|}\right) \quad .$$

A. Scenario 1

The first scenario is the case where the observed variables behave just as in the model, i.e., variable (groups) $\mathbf{x}$ and $\mathbf{y}$ are created such that $\mathbf{R}_{\mathbf{xx}}$ and $\mathbf{R}_{\mathbf{yy}}$ are the identity matrices. The variables are then observed by the user through a linear mixture up to some additive noise as $\mathbf{x}^{(k)} = \beta\mathbf{U}_\mathbf{x}\mathbf{x} + (1-\beta)\mathbf{n}_\mathbf{x}^{(k)}$, where $\mathbf{n}_\mathbf{x}$ is a vector of the dimension of $\mathbf{x}$ containing unit variance Gaussian noise and $\mathbf{U}_\mathbf{x} \in \mathbb{R}^{M\times M}$ is a non-degenerate mixing matrix. The observations $\mathbf{y}^{(k)} = \beta\mathbf{U}_\mathbf{y}\mathbf{y}+(1-\beta)\mathbf{n}_\mathbf{y}^{(k)}$ are created analogously and we calculate for each k the corresponding covariance and cross-covariance matrices. The results are given in Figure 1.

It is clear from the measure $PI(\mathbf{W}_\mathbf{x}^T\mathbf{R}_{\mathbf{xy}}\mathbf{W}_y)$ that the influence of α cannot be neglected. Indeed, when $\alpha = 0$ – i.e. neglecting the cross-covariance structure –, the diagonalization of $\mathbf{W}_\mathbf{x}^T\mathbf{R}_{\mathbf{xy}}\mathbf{W}_y$ in the noiseless case is only approximate, whereas for $\alpha > 0$ it is quasi-exact.

B. Scenario 2

In this scenario we consider the observations $\mathbf{R}_{\mathbf{xy}}^{(k)} = \mathbf{X}^{(k)} = (1-\beta)\mathbf{U}_\mathbf{x}\mathbf{\Delta}^{(k)}\mathbf{U}_\mathbf{y}^T + \beta\mathbf{n}^{(k)}$ and calculate the marginalized covariances as described in paragraph II-D. The matrices $\mathbf{U}_\mathbf{x}$ and $\mathbf{U}_\mathbf{y}$ are also called the mixing matrices and its entries are drawn according to the normal distribution with zero mean and unit variance. The matrix $\mathbf{\Delta}^{(k)}$ is a diagonal matrix with entries drawn from the same normal distribution. Next to the Moreau-Amari performance index on the resulting approximately diagonalized matrices we also calculate the index for the matrices $\mathbf{W}_\mathbf{x}^T\mathbf{U}_\mathbf{x}$ and $\mathbf{W}_\mathbf{y}^T\mathbf{U}_\mathbf{y}$, which describes how well we can estimation our model. The results are given in Figure 2.

It is particularly interesting that in scenario 2, except for low noise, it seems that the influence of parameter α is inverted with respect to scenario 1. For higher noise levels, it seems that the exclusion of $\mathbf{R}_{\mathbf{xy}}$ – both in terms of diagonalization as in terms of model estimation – slightly augments the performance. This is contradictory to what is generally accepted, in that the singular value decomposition of the signal is judged more stable than the left or right eigenvalue decompositions of the covariance matrices.

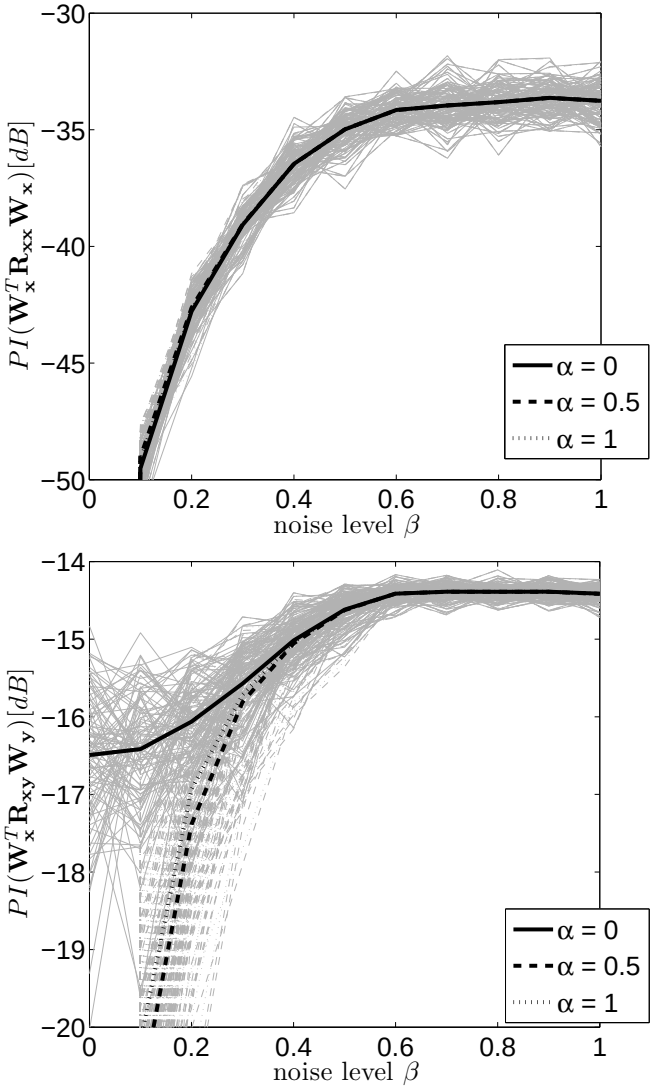

Fig. 1. The diagonalization criteria (3) for the first scenario. Gray lines are for the individual Monte Carlo realisations, black lines are the means over the realisations

IV. CONCLUSIONS

The extension of the canonical correlation analysis to multiple observations of two groups of variables seems a promising approach for signal processing applications. When estimating the latent canonical structure in the data, multiple observations may lower the impact of noise on the estimates. The proposed method allows for the estimation of the non-orthogonal matrices exposing the inherent correlation structure without a required pre-whitening of the observations. In addition, by introducing a weighting parameter, one may balance between the covariance and the cross-covariance diagonalization. At last, when the diagonalization of an observed matrix structure and its marginalized covariances is envisaged, the method is reminiscent to an approximately joint singular value decomposition.

Future works should focus on the application of the algorithm to joint recordings of GAZE and EEG signals,

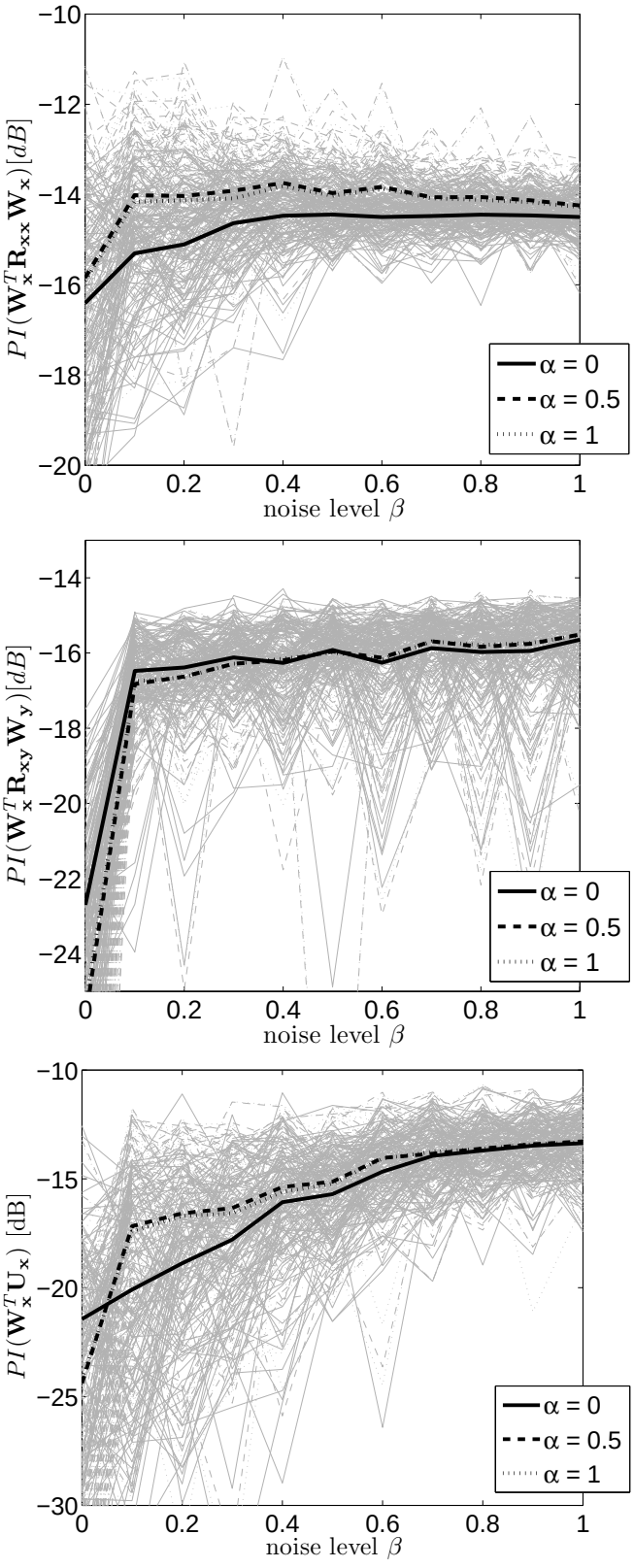

Fig. 2. The diagonalization and model estimation criteria (3) for the second scenario. Gray lines are for the individual Monte Carlo realisations, black lines are the means over the realisations

where we should principally investigate whether the linear dependence is sufficient to explore their relationships.

APPENDIX

A. Outline of a proving strategy

Assume, without loss of generality, that $M \geq N$ and that $\mathbf{R_{xx}}$ and $\mathbf{R_{yy}}$ are the identity matrices. Renaming $\mathbf{R} = \mathbf{W}_\mathbf{x}^T\mathbf{R_{xy}}\mathbf{W_y}$, we have

$$\mathbf{R}^T\mathbf{R} = \mathbf{\Lambda}$$
$$\mathbf{R}\mathbf{R}^T = \begin{pmatrix} \mathbf{\Lambda} & \mathbf{0}_{N\times M-N} \\ \mathbf{0}_{M-N\times N} & \mathbf{0}_{M-N\times M-N} \end{pmatrix}$$

from which we have

$$\mathbf{R} = \begin{pmatrix} \tilde{\mathbf{R}} \\ \mathbf{0}_{M-N\times N} \end{pmatrix}$$

with $\tilde{R} \in \mathbb{R}^{N\times N}$ an orthogonal matrix. The entries $\tilde{R}_{ij}$ contain the values $E\{x_i y_j\}$ and since $E\{x_i x_j\} = 0, E\{y_i y_j\} = 0 \forall i \neq j$. We have that each row and each column of $\mathbf{R}$ can contain at maximum one entry $|\tilde{R}_{ij}| > \cos(\pi/4) = 1/\sqrt{2}$. Suppose now that $\exists j_1 : |E\{x_m y_{j_1}\}| \approx |E\{x_{m'} y_{j_1}\}| \approx 1/\sqrt{2}$, then all other entries $E\{x_i y_{j_1}\} \approx 0, \forall i \neq m, m'$. Taking $\tilde{\mathbf{R}}\tilde{\mathbf{R}}^T = \mathbf{\Lambda}$, we see that $\Lambda_{mm} > 1/2$ and $\Lambda_{mm'} > 1/2$. Analogous reasoning brings us to $\exists i_1 : |E\{x_{i_1} y_n\}| \approx |E\{x_{i_1} y_{n'}\}| \approx 1/\sqrt{2} \Rightarrow E\{x_{i_1} y_j\} \approx 0, \forall j \neq n, n'$.Thus, $\mathbf{\Lambda}$ is not diagonal under the above assumptions and at least one of the assumed entries must be much smaller than $1/\sqrt{2}$.

REFERENCES

[1] Jean-François Cardoso and Antoine Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 17(1):161–164, 1996.

[2] J. Douglas Carroll. Generalization of canonical correlation analysis to three or more sets of variables. In *Proceedings of the 76th Annual Convention APA*, pages 227–228, 1968.

[3] Mohamed Hanafi and Henk A. L. Kiers. Analysis of k sets of data, with differential emphasis on agreement between and within sets. *Computational Statistics & Data Analysis*, 51:1491–1508, 2006.

[4] P. Horst. Relations among m sets of measures. *Psychometrika*, 26:129–149, 1961.

[5] H. Hotelling. Relation between two sets of variates. *Biometrika*, 28:321–377, 1936.

[6] A. Souloumiac. Nonorthogonal joint diagonalization by combining givens and hyperbolic rotations. 57(6):2222–2231, June 2009.

[7] Fabian J. Theis, Peter Gruber, Ingo R. Keck, Anke Meyer-Bäse, and Elmar W. Lang. Spatiotemporal blind source separation using double-sided approximate joint diagonalization. In *Proceedings EUSIPCO 2005*, Antalya, 2005.

[8] Petr Tichavský and Arie Yeredor. Fast approximate joint diagonalization incorporating weight matrices. 57(3):878–891, March 2009.

COPYRIGHT NOTICE