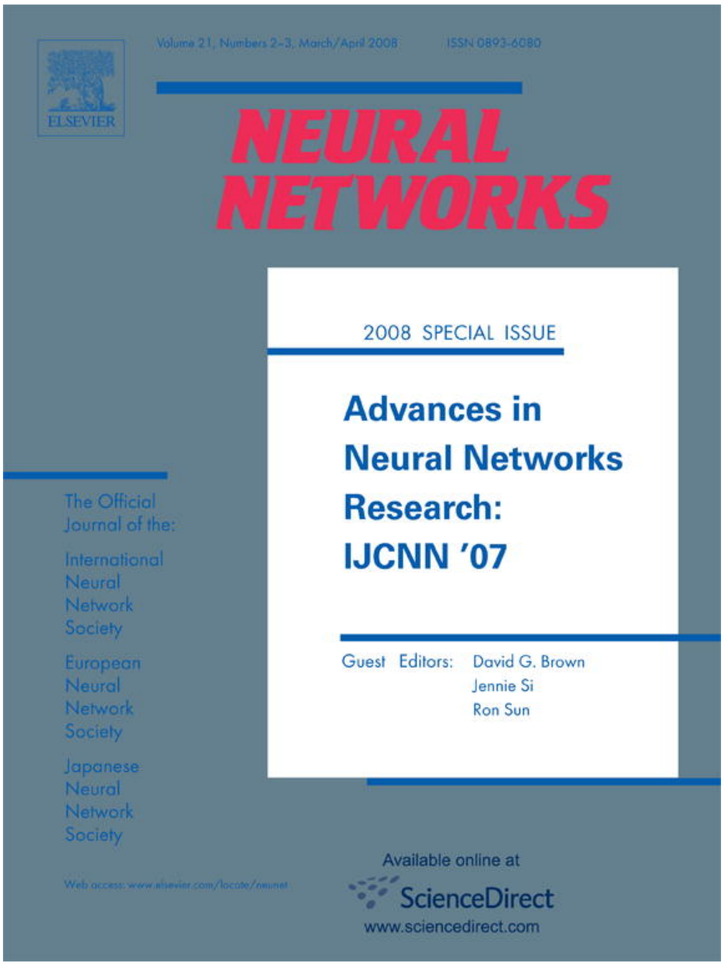

Volume 21, Numbers 2–3, March/April 2008
ISSN 0893-6080
ELSEVIER
NEURAL NETWORKS
2008 SPECIAL ISSUE
Advances in Neural Networks Research: IJCNN '07
Guest Editors: David G. Brown
Jennie Si
Ron Sun
The Official Journal of the:
International Neural Network Society
European Neural Network Society
Japanese Neural Network Society
Web access: www.elsevier.com/locate/neunet
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com
ScienceDirect
Neural Networks 21 (2008) 493–501
Neural Networks
www.elsevier.com/locate/neunet

2008 Special Issue

Iterative multi-channel coherence analysis with applications☆

Bryan D. Thompson*, Mahmood R. Azimi-Sadjadi

Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523, USA

Received 8 August 2007; received in revised form 1 December 2007; accepted 13 December 2007

Abstract

An iterative learning algorithm for performing Multi-Channel Coherence Analysis (MCCA) is developed in this paper. MCCA is an extension of the well-known Canonical Correlation Analysis (CCA) that allows for more than two data channels to be analyzed. This paper discusses a standard method for performing MCCA and compares it to a newly developed data-driven and iterative approach. The proposed algorithm is then tested on two examples and its performance is evaluated in terms of estimation errors with respect to the values obtained using the standard MCCA algorithm. The first example uses a synthesized data set while the second example uses a real data set based on multi-spectral satellite imagery of the Earth's surface.

Keywords: Iterative learning; Multi-channel coherence analysis; Canonical correlation analysis; Least mean square learning; Multi-spectral satellite imagery

1. Introduction

Canonical correlation analysis (CCA) (Hotelling, 1936; Rencher, 2002) provides an elegant framework for analyzing and solving many problems in signal processing (Pezeshki, Scharf, Azimi-Sadjadi, & Hua, 2005), communications (Scharf & Mullis, 2000), radar or sonar (Pezeshki, Azimi-Sadjadi, & Scharf, 2007), and satellite imagery (Falcone, Azimi-Sadjadi, & Kankiewicz, 2007). Using CCA, the analysis of linear dependence and coherence between two data channels may easily be carried out based upon their canonical correlations. Consequently, an extension from the two-channel case to include multiple channels has been a topic of interest since the early 1950s and was first explored in Vinograde (1950).

Multi-Channel Coherence Analysis (MCCA) (Kettingring, 1971) may be viewed as an extension of CCA to include multiple data channels. Developing this extension is not a trivial task for many reasons. CCA requires maximizing one cross-correlation term, subject to a particular constraint, to obtain each canonical coordinate pair and the corresponding correlation. However, for the multi-channel extension, there are multiple cross-correlation terms to be maximized simultaneously. Even if these terms are maximized correctly, the resulting multi-channel coordinates and correlations are not "canonical" as in the two-channel case. CCA produces 'valid' canonical coordinates and correlations because the correlations measure cosines of principal angles between the two data channels (Pezeshki et al., 2005). Moreover, the principal angles are the angles between the associated canonical coordinates (Scharf & Mullis, 2000). Consequently, the canonical correlations are invariant to affine transformations applied to the data channels. Thus, any scaling and/or rotation transformation on the data channels does not change the canonical correlations.

The difficulty with making the multi-channel extension of CCA is that the resultant problem is not as well defined (Kettingring, 1971) and therefore has more than one possible formation. Many variations of the MCCA problem have been investigated in the past (Horst, 1961; Kettingring, 1971; Nielsen, 2002; Steel, 1951; Vinograde, 1950). More specifically, these include the exploration of optimizing one of the five main objective functions subject to some meaningful constraints. In Horst (1961), optimizing the variance of the mapped composite data channel was considered. It was shown that this variance simplifies to maximizing the sum of the correlations between all pairs of mapped data channels. In

☆ An abbreviated version of some portions of this article appeared in Thompson and Azimi-Sadjadi (2007) as part of the IJCNN 2007 Conference Proceedings, published under IEE copyright.

* Corresponding author.
E-mail addresses: BryanDavidThompson@gmail.com (B.D. Thompson), azimi@engr.colostate.edu (M.R. Azimi-Sadjadi).

0893-6080/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2007.12.025

addition to this analysis, optimizing the sum of the squared correlations was introduced in Kettingring (1971). The effect this has on the eigenvalues of the correlation matrix of the composite coordinate vector was also shown. The remaining three objective functions focused on optimizing eigenvalue-based measures of the correlation matrix and are not explored in this paper. In Nielsen (2002), various combinations of these five objective functions and constraints were analyzed leading to many optimization problems.

This paper defines and analyzes the MCCA problem using the objective function and constraint introduced in Horst (1961), and later extended in Kettingring (1971). This formulation has also been explored in Nielsen (2002). However, the traditional method for solving this formulation of MCCA requires the knowledge of correlation matrices between all pairs of data channels and having to solve eigenvalue problems. Clearly, the computation of all correlation sums is inefficient, especially when only the principal ones are of interest. This paper develops an algorithm to iteratively perform MCCA. This algorithm provides the ability to extract multi-channel coordinates and correlation sums directly from the data without the knowledge of correlation matrices between all pairs of data channels. Furthermore, in contrast to the standard method for MCCA, this algorithm allows for the computation of a subset of the dominant multi-channel correlation sums and the associated coordinates hence providing a possible savings in computational burden.

The organization of this paper is as follows. Section 2 analyzes the aforementioned objective function and constraint for extracting the multi-channel coordinates and correlations. This leads to a constrained optimization problem, uniquely defined and further examined. Section 3 discusses the relationship between the mutual information carried by the data channels and the multi-channel correlation sums. The development and introduction of the proposed data-driven iterative approach for recursive extraction of multi-channel coordinates and correlation sums is then discussed in Section 4. Advantages and disadvantages of this method over the standard approach for MCCA are also reviewed. Section 5 provides an example that demonstrates the effectiveness of this iterative approach on a synthesized data set. In addition, a second example is provided that uses a real data set based on multi-spectral satellite imagery obtained via Meteosat 8-SEVIRI (Schmetz et al., 2002): an Eumetsat satellite used to capture images from portions of the Earth's surface. This example exploits coherence information across spectral bands to extract the coordinates and corresponding correlations. Convergence of the developed algorithm is demonstrated experimentally on both examples. Finally, conclusions are presented in Section 6.

2. Multi-channel coherence analysis reviewed

Consider n zero mean random vectors, $\mathbf{x}_1, \mathbf{x}_2, \ldots,$ and $\mathbf{x}_n$, representing multiple data channels comprising the composite data channel $\mathbf{z} = \left[\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_n^T\right]^T \in \mathbb{R}^{d\times 1}$. Let each channel $\mathbf{x}_j \in \mathbb{R}^{d_j\times 1}$ be of dimension d_j, where it is assumed that $\mathbf{x}_1$ is of the smallest dimension. The correlation matrix of the composite data channel $\mathbf{z}$ is given by

$$R_{\mathbf{zz}} = E\left[\mathbf{z}\mathbf{z}^T\right] = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ R_{21} & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ R_{n1} & R_{n2} & \cdots & R_{nn} \end{bmatrix}, \quad (1)$$

where $R_{jk} = E\left[\mathbf{x}_j\mathbf{x}_k^T\right]$ is the cross-correlation matrix of data channels $\mathbf{x}_j$ and $\mathbf{x}_k$. Clearly, we have $R_{jk} = R_{kj}^T$.

Similar to CCA, the ith multi-channel coordinate of the jth channel is found by searching for the ith coordinate mapping vector, $\mathbf{a}_{i,j}$, of data channel $\mathbf{x}_j$. This produces the ith multi-channel coordinate for the jth channel,

$$v_{ij} = \mathbf{a}_{i,j}^T\mathbf{x}_j, \quad (2)$$

which must satisfy certain conditions discussed later. If the ith coordinate mapping vectors are found for all n channels, then we obtain the following composite coordinate mapping vector

$$\mathbf{a}_i = [\mathbf{a}_{i,1}^T, \mathbf{a}_{i,2}^T, \ldots, \mathbf{a}_{i,n}^T]^T, \quad (3)$$

and composite coordinate vector

$$\mathbf{v}_i = \begin{bmatrix} \mathbf{a}_{i,1}^T\mathbf{x}_1 \\ \mathbf{a}_{i,2}^T\mathbf{x}_2 \\ \vdots \\ \mathbf{a}_{i,n}^T\mathbf{x}_n \end{bmatrix} = \begin{bmatrix} v_{i,1} \\ v_{i,2} \\ \vdots \\ v_{i,n} \end{bmatrix}. \quad (4)$$

The composite coordinate vector (4) consists of the ith multi-channel coordinate of every channel. The associated correlation matrix of (4) is given by

$$R_{\mathbf{v}_i\mathbf{v}_i} = E\left[\mathbf{v}_i\mathbf{v}_i^T\right] = \begin{bmatrix} \mathbf{a}_{i,1}^T R_{11}\mathbf{a}_{i,1} & \mathbf{a}_{i,1}^T R_{12}\mathbf{a}_{i,2} & \cdots & \mathbf{a}_{i,1}^T R_{1n}\mathbf{a}_{i,n} \\ \mathbf{a}_{i,2}^T R_{21}\mathbf{a}_{i,1} & \mathbf{a}_{i,2}^T R_{22}\mathbf{a}_{i,2} & \cdots & \mathbf{a}_{i,2}^T R_{2n}\mathbf{a}_{i,n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_{i,n}^T R_{n1}\mathbf{a}_{i,1} & \mathbf{a}_{i,n}^T R_{n2}\mathbf{a}_{i,2} & \cdots & \mathbf{a}_{i,n}^T R_{nn}\mathbf{a}_{i,n} \end{bmatrix}. \quad (5)$$

The elements of (5) are correlations between all possible pairs of the coordinates defined in (2).

In CCA the first canonical coordinate pair and correlation are found by maximizing only one correlation (Hotelling, 1936; Rencher, 2002), namely, $\mathrm{corr}\{v_{1,1}, v_{1,2}\}$, while imposing the unit variance constraints $\mathrm{var}\{v_{1,1}\} = \mathrm{var}\{v_{1,2}\} = 1$. In the multi-channel extension, there are more correlations to maximize. Thus, to extend the two-channel case to the multi-channel case, all correlations must be maximized simultaneously (Horst, 1961). Moreover, we desire that this extension reduces to the standard case of CCA when only two channels are considered, i.e., when $n = 2$. Another desired feature (Kettingring, 1971) is that the method selects the multi-channel coordinates, one from each data channel, so as to optimize some function of their correlation matrix, $R_{\mathbf{v}_i\mathbf{v}_i}$. To do this, many sensible objective functions of $R_{\mathbf{v}_i\mathbf{v}_i}$ have been proposed and analyzed in the past (Horst, 1961; Kettingring, 1971; Steel, 1951; Vinograde, 1950). This paper analyzes the SUMCOR objective function (Horst, 1961) that maximizes the variance of the mapped composite data channel, $\mathbf{a}_i^{\mathrm{T}}\mathbf{z}$. This can be written as

$$\begin{aligned}\mathrm{var}\{\mathbf{a}_i^{\mathrm{T}}\mathbf{z}\} &= E[(\mathbf{a}_i^{\mathrm{T}}\mathbf{z} - E[\mathbf{a}_i^{\mathrm{T}}\mathbf{z}])^2] = \mathbf{a}_i^{\mathrm{T}} R_{\mathbf{zz}}\mathbf{a}_i \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} \mathbf{a}_{i,j}^{\mathrm{T}} R_{jk}\mathbf{a}_{i,k} = \sum_{j=1}^{n}\sum_{k=1}^{n} [R_{\mathbf{v}_i\mathbf{v}_i}]_{j,k} \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} corr\{\mathbf{a}_{i,j}^{\mathrm{T}}\mathbf{x}_j, \mathbf{a}_{i,k}^{\mathrm{T}}\mathbf{x}_k\} \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} corr\{v_{i,j}, v_{i,k}\}. \end{aligned} \tag{6}$$

Hence, maximizing the variance of the mapped composite data channel simplifies to maximizing the sum of the correlations between all pairs of mapped channels given by (2). This method is referred to as the "maximum correlation method" in Horst (1961). By using the same constraint as in CCA, i.e., unit variance, this objective function was examined in Horst (1961) and Kettingring (1971) and resulted in an eigenvalue problem lacking a unique solution. Other constraints have since been introduced and used (Nielsen, 1994) in the optimization process. This study examines the unit trace constraint that constrains the sum of the variances of all mapped data channels to be unity. However, this sum is equivalent to the trace of matrix $R_{\mathbf{v}_i\mathbf{v}_i}$. Thus, the constraint yields

$$\sum_{j=1}^{n} \mathrm{var}\{v_{i,j}\} = \mathrm{tr}\{R_{\mathbf{v}_i\mathbf{v}_i}\} = 1, \tag{7}$$

where $\mathrm{tr}\{Q\}$ denotes the trace of matrix Q.

It can easily be shown that when $n = 2$, finding the first coordinate pair and correlation by optimizing (6) with the unit trace constraint (7) yields the same problem as that obtained with CCA (Hotelling, 1936; Rencher, 2002) (with the exception of a scaling factor). This helps to validate the extension of CCA to MCCA with this objective function/constraint pair. However, the problem of finding higher-order canonical coordinate pairs and corresponding correlations requires a deflation procedure (Pezeshki, Azimi-Sadjadi, & Scharf, 2003) to remove the contributions of the previously found coordinates from the data channels. An equivalent procedure that can be applied to MCCA does not exist (Kettingring, 1971). However a similar procedure suggested in Kettingring (1971) and used in this work requires that the ith composite coordinate mapping vector, $\mathbf{a}_i$, is orthogonal to $\mathbf{a}_p$, for $p \neq i$. Consequently, the problem of finding the first multi-channel coordinate set and correlation sum can be extended to that of finding the ith coordinate set and correlation sum by incorporating this requirement of orthogonality.

Thus, the problem of optimizing the SUMCOR objective function with the unit trace constraint for finding the first composite coordinate mapping vector $\mathbf{a}_1 = [\mathbf{a}_{1,1}^{\mathrm{T}}, \mathbf{a}_{1,2}^{\mathrm{T}}, \ldots, \mathbf{a}_{1,n}^{\mathrm{T}}]^{\mathrm{T}}$ is

$$\mathbf{a}_1 = \arg\max_{\tilde{\mathbf{a}}_1} \sum_{j=1}^{n}\sum_{k=1}^{n} \tilde{\mathbf{a}}_{1,j}^{\mathrm{T}} R_{jk}\tilde{\mathbf{a}}_{1,k}, \tag{8}$$

subject to

$$\mathrm{tr}\{R_{\mathbf{v}_1\mathbf{v}_1}\} = \sum_{j=1}^{n} \mathbf{a}_{1,j}^{\mathrm{T}} R_{jj}\mathbf{a}_{1,j} = 1.$$

The "˜" notation denotes some intermediate (non-optimal) value. Using a Lagrange multiplier method, the constrained optimization problem in (8) can be converted to maximizing

$$c_1(\tilde{\mathbf{a}}_1) = \sum_{j=1}^{n}\sum_{k=1}^{n} \tilde{\mathbf{a}}_{1,j}^{\mathrm{T}} R_{jk}\tilde{\mathbf{a}}_{1,k} - \lambda_1\left(\sum_{j=1}^{n} \tilde{\mathbf{a}}_{1,j}^{\mathrm{T}} R_{jj}\tilde{\mathbf{a}}_{1,j} - 1\right), \tag{9}$$

where λ_1 is the 1st Lagrange multiplier. At the solution, the gradient of $c_1(\tilde{\mathbf{a}}_1)$ with respect to $\tilde{\mathbf{a}}_1$ will equal zero, i.e.,

$$\begin{aligned} \nabla c_1(\mathbf{a}_1) &= 2\sum_{k=1}^{n} R_{jk}\mathbf{a}_{1,k} - 2\lambda_1 R_{jj}\mathbf{a}_{1,j} = \mathbf{0}, \\ &\quad \text{for } j = 1, 2, \ldots, n, \\ &\sum_{k=1}^{n} R_{jk}\mathbf{a}_{1,k} = \lambda_1 R_{jj}\mathbf{a}_{1,j}, \quad \text{or} \\ &A\mathbf{a}_1 = \lambda_1 B\mathbf{a}_1 \end{aligned} \tag{10}$$

where

$$A = R_{\mathbf{zz}} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ R_{21} & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ R_{n1} & R_{n2} & \cdots & R_{nn} \end{bmatrix}, \quad \text{and}$$

$$B = \begin{bmatrix} R_{11} & 0 & \cdots & 0 \\ 0 & R_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{nn} \end{bmatrix}. \tag{11}$$

The result in (10) is an eigenvalue problem and standard methods for solving such a problem are well-known (Golub & Loan, 1996). To find higher-order multi-channel coordinate mapping vectors, e.g., $\mathbf{a}_i$ for $i = 1, 2, \ldots, u$, we must satisfy the orthogonality requirement (Kettingring, 1971) of $\langle \mathbf{a}_p, \mathbf{a}_i \rangle = \mathbf{a}_i^{\mathrm{T}}\mathbf{a}_p = 0$, for $p \neq i$. To do this, we find the u largest eigenvalues of $A\mathbf{a}_i = \lambda_i B\mathbf{a}_i$ that are distinct, where we define $u \leq d_1$ as the number of distinct eigenvalues. Since the eigenvalues are distinct, the corresponding eigenvectors are therefore orthogonal to one another. Thus, the ith multi-channel composite coordinate mapping vector, $\mathbf{a}_i$, is simply the ith eigenvector with corresponding distinct eigenvalue λ_i, where $\lambda_1 > \lambda_2 > \cdots > \lambda_u$.

If we left-multiply $A\mathbf{a}_i = \lambda_i B\mathbf{a}_i$ by the transpose of the ith multi-channel composite coordinate mapping vector, we obtain a relationship between the ith multi-channel correlation sum, $c_i(\mathbf{a}_i)$, and the ith largest distinct eigenvalue, λ_i. That is,

$$c_i(\mathbf{a}_i) = \sum_{j=1}^{n}\sum_{k=1}^{n} \mathbf{a}_{i,j}^{\mathrm{T}} R_{jk}\mathbf{a}_{i,k} = \mathbf{a}_i^{\mathrm{T}} A\mathbf{a}_i = \lambda_i \mathbf{a}_i^{\mathrm{T}} B\mathbf{a}_i$$
$$= \lambda_i \sum_{j=1}^{n} \mathbf{a}_{i,j}^{\mathrm{T}} R_{jj}\mathbf{a}_{i,j} = \lambda_i. \quad (12)$$

Thus, at solution the ith multi-channel correlation sum is equal to the ith largest distinct eigenvalue. Furthermore, the multi-channel correlation sums are the largest u distinct eigenvalues of the eigenvalue problem $A\mathbf{a}_i = \lambda_i B\mathbf{a}_i$, and the corresponding eigenvectors contain the coordinate mapping vectors for each data channel. These mapping vectors may be used to obtain the multi-channel coordinates via (2).

3. Relationship to multi-channel mutual information

Here we develop a relationship between the mutual information carried by all n data channels and the multi-channel correlation sums. According to Kay (2000), the mutual information for the composite data vector $\mathbf{z} = \left[\mathbf{x}_1^{\mathrm{T}}, \mathbf{x}_2^{\mathrm{T}}, \ldots, \mathbf{x}_n^{\mathrm{T}}\right]^{\mathrm{T}}$ is defined as

$$I = \sum_{i=1}^{n} H_{\mathbf{x}_i} - H_{\mathbf{z}}, \quad (13)$$

where $H_{\mathbf{x}_i}$ and $H_{\mathbf{z}}$ denote the Shannon (Shannon, 1948) entropies of $\mathbf{x}_i$ and $\mathbf{z}$, respectively. If we assume $\mathbf{z}$ to be Gaussian distributed, i.e. $\mathbf{z} : N(\mathbf{0}, R_{\mathbf{zz}})$, with density function $f(\mathbf{z})$, then the Shannon entropy is

$$H_{\mathbf{z}} = E[\log f(\mathbf{z})] = \frac{d}{2}\log(2\pi e) + \frac{1}{2}\log\det[R_{\mathbf{zz}}], \quad (14)$$

where $d = \sum_{i=1}^{n} d_i$. Using $H_{\mathbf{z}}$ in (14) and similar expressions for $H_{\mathbf{x}_i}$'s, we may obtain the group mutual information I as

$$I = \frac{1}{2}\sum_{i=1}^{n}\log\det[R_{ii}] - \frac{1}{2}\log\det[R_{\mathbf{zz}}]$$
$$= -\frac{1}{2}\log\left(\frac{\det[A]}{\det[B]}\right), \quad (15)$$

where A and B are defined by (11). Now, from the eigenvalue problem $A\mathbf{a}_i = \lambda_i B\mathbf{a}_i$, we can easily show that (15) reduces to

$$I = -\frac{1}{2}\sum_{i=1}^{d}\log\lambda_i. \quad (16)$$

Note that in the two-channel case $\det[R_{\mathbf{zz}}] = \det[R_{11}]\det[R_{22}]\det[I - KK^{\mathrm{T}}]$ where K is the canonical correlation matrix of canonical correlations, k_i. This result suggests that $\lambda_i = 1 - k_i^2$ in (16) for the CCA.

4. An iterative learning algorithm for multi-channel coherence analysis

This section introduces an iterative method for the recursive extraction of multi-channel coordinates and correlations directly from the data. This method avoids the computation of correlation matrices for all pairs of data channels and having to solve eigenvalue problems. In addition, this algorithm allows for a subset containing the dominant correlation sums and associated coordinates to be computed, which may in turn yield computational savings when compared to the standard approach. However, this potential computational savings is clearly dependent on the data channel dimensions, as well as the number of multi-channel coordinates and correlation sums to be computed. Similar to the definition used in Section 2, this approach also solves the MCCA problem by optimizing the SUMCOR objective function with the unit trace constraint.

We start by iteratively searching for the first multi-channel coordinate of each channel and the correlation sum. This leads to updating equations that, when iterated, will converge to the desired coordinates and correlation sum. The steepest descent algorithm (Golub & Loan, 1996) is used in the development of these updating equations. Next, we extend the problem of finding the first set of coordinates and corresponding correlation sum to finding the ith set, for $i = 2, 3, \ldots, u$, of coordinates and corresponding correlation sum, given that the previous $(i-1)$ have been successfully computed. Then, the updating equations are modified so that the calculation of sample correlation matrices between all data channel pairs is not required. This reduces the computation time and overall algorithm complexity. The result is the updating equations that are obtained by using the Least-Mean-Square (LMS) algorithm (Golub & Loan, 1996).

To develop an iterative method for finding the first multi-channel coordinates and corresponding correlation sum, we start from the optimization problem in (9). For finding the first set of coordinate mapping vectors, we consider the gradient of $c_1(\tilde{\mathbf{a}}_{1,j})$ with respect to $\tilde{\mathbf{a}}_{1,j}$, for $j = 1, 2, \ldots, n$, given by

$$\nabla c_1(\tilde{\mathbf{a}}_{1,j}) = 2\sum_{k=1}^{n} R_{jk}\tilde{\mathbf{a}}_{1,k} - 2\lambda_1 R_{jj}\tilde{\mathbf{a}}_{1,j}. \quad (17)$$

Using this gradient, the updating equation for the jth coordinate mapping vector approximation, $\tilde{\mathbf{a}}_{1,j}$, using the steepest descent algorithm is

$$\tilde{\mathbf{a}}_{1,j}(l+1) = \tilde{\mathbf{a}}_{1,j}(l) + \frac{\beta}{2}\left(\nabla c_1(\tilde{\mathbf{a}}_{1,j}(l))\right)$$
$$= \tilde{\mathbf{a}}_{1,j}(l) + \beta\left(\sum_{k=1}^{n} R_{jk}\tilde{\mathbf{a}}_{1,k}(l) - \tilde{\lambda}_1(l+1)R_{jj}\tilde{\mathbf{a}}_{1,j}(l)\right), \quad (18)$$

where l is the index of iteration and β is the step size. In a more compact form ($\forall j = 1, 2, \ldots, n$) we can rewrite (18) as

$$\tilde{\mathbf{a}}_1(l+1) = \tilde{\mathbf{a}}_1(l) + \beta\left(A\tilde{\mathbf{a}}_1(l) - \tilde{\lambda}_1(l+1)B\tilde{\mathbf{a}}_1(l)\right), \quad (19)$$

where the composite coordinate mapping vector estimate,

$$\tilde{\mathbf{a}}_1(l) = [\tilde{\mathbf{a}}_{1,1}^{\mathrm{T}}(l), \tilde{\mathbf{a}}_{1,2}^{\mathrm{T}}(l), \ldots, \tilde{\mathbf{a}}_{1,n}^{\mathrm{T}}(l)]^{\mathrm{T}}, \quad (20)$$

updates all coordinate mapping vector estimates simultaneously. To create an updating equation for the first Lagrange multiplier approximation, $\tilde{\lambda}_1$, we must first find the value of this multiplier once the solution is reached. In (12), we showed that at the solution,

$$\lambda_1 = \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{a}_{1,j}^{\mathrm{T}} R_{jk} \mathbf{a}_{1,k} = \mathbf{a}_1^{\mathrm{T}} A \mathbf{a}_1. \tag{21}$$

Using (20) and (21), the updating equation for $\tilde{\lambda}_1$ can be formulated as

$$\tilde{\lambda}_1(l+1) = \sum_{j=1}^{n} \sum_{k=1}^{n} \tilde{\mathbf{a}}_{1,j}^{\mathrm{T}}(l) R_{jk} \tilde{\mathbf{a}}_{1,k}(l) = \tilde{\mathbf{a}}_1^{\mathrm{T}}(l) A \tilde{\mathbf{a}}_1(l). \tag{22}$$

Together, updating Eqs. (19) and (22) comprise an iterative process for extracting the first multi-channel coordinate mapping vectors and correlation sum from n data channels. We use the " $\dot{}$ " notation to denote the converged values of these variables to within some acceptable error tolerance. That is, for sufficiently large l, $\tilde{\mathbf{a}}_i(l) = \dot{\mathbf{a}}_i$ and $\tilde{\lambda}_i(l) = \dot{\lambda}_i$, where $\epsilon_1 \geq \|\mathbf{a}_i - \dot{\mathbf{a}}_i\|_2^2 / \|\mathbf{a}_i\|_2^2$ and $\epsilon_2 \geq (\lambda_i - \dot{\lambda}_i)^2/\lambda_i^2$ are acceptable normalized errors. After convergence, $\dot{\mathbf{a}}_{1,j}$ closely approximates the true mapping vector $\mathbf{a}_{1,j}$, for $j = 1, 2, \ldots, n$, which allows for finding the first multi-channel coordinates. Moreover, $\dot{\lambda}_1$ closely approximates the first multi-channel correlation sum, λ_1.

To develop an iterative method for finding the ith set, for $i = 2, 3, \ldots, u$, of multi-channel coordinate mapping vectors and correlation sum given that the previous $(i-1)$ have been successfully computed, it is desired to find $\mathbf{a}_i = [\mathbf{a}_{i,1}^{\mathrm{T}}, \mathbf{a}_{i,2}^{\mathrm{T}}, \ldots, \mathbf{a}_{i,n}^{\mathrm{T}}]^{\mathrm{T}}$ such that

$$\mathbf{a}_i = \arg\max_{\tilde{\mathbf{a}}_i} \sum_{j=1}^{n} \sum_{k=1}^{n} \tilde{\mathbf{a}}_{i,j}^{\mathrm{T}} R_{jk} \tilde{\mathbf{a}}_{i,k}, \tag{23}$$

subject to the following two constraints

$$\mathrm{tr}\{R_{\mathbf{v}_i\mathbf{v}_i}\} = \sum_{j=1}^{n} \mathbf{a}_{i,j}^{\mathrm{T}} R_{jj} \mathbf{a}_{i,j} = 1, \quad \text{and} \quad \langle \mathbf{a}_p, \mathbf{a}_i \rangle = 0,$$
$$\text{for } p = 1, 2, \ldots, i-1.$$

The latter constraint ensures that the requirement of orthogonality among composite coordinate mapping vectors is satisfied, as previously discussed in Section 2. Using a penalty parameter and corresponding penalty function (Chong & Zak, 2001), the constrained optimization problem in (23) can be converted to maximizing

$$\begin{aligned} c_i(\tilde{\mathbf{a}}_i) &= \sum_{j=1}^{n} \sum_{k=1}^{n} \tilde{\mathbf{a}}_{i,j}^{\mathrm{T}} R_{jk} \tilde{\mathbf{a}}_{i,k} - \gamma \left(\left(\sum_{j=1}^{n} \tilde{\mathbf{a}}_{i,j}^{\mathrm{T}} R_{jj} \tilde{\mathbf{a}}_{i,j} - 1 \right)^2 \right. \\ &\quad \left. + \left(\sum_{p=1}^{i-1} \langle \mathbf{a}_p, \tilde{\mathbf{a}}_i \rangle^2 \right) \right) \\ &= \tilde{\mathbf{a}}_i^{\mathrm{T}} A \tilde{\mathbf{a}}_i - \gamma g(\tilde{\mathbf{a}}_i) \end{aligned} \tag{24}$$

where $\gamma > 0$ is a penalty parameter with corresponding penalty function

$$\begin{aligned} g(\tilde{\mathbf{a}}_i) &= \left(\sum_{j=1}^{n} \tilde{\mathbf{a}}_{i,j}^{\mathrm{T}} R_{jj} \tilde{\mathbf{a}}_{i,j} - 1 \right)^2 + \left(\sum_{p=1}^{i-1} \langle \mathbf{a}_p, \tilde{\mathbf{a}}_i \rangle^2 \right) \\ &= \left(\tilde{\mathbf{a}}_i^{\mathrm{T}} B \tilde{\mathbf{a}}_i - 1 \right)^2 + \sum_{p=1}^{i-1} (\tilde{\mathbf{a}}_i^{\mathrm{T}} \mathbf{a}_p)^2. \end{aligned} \tag{25}$$

The larger the value of the penalty parameter γ, the closer the approximation $\tilde{\mathbf{a}}_i$ will be to its true value $\mathbf{a}_i$ (Chong & Zak, 2001). Note that $g(\mathbf{a}_i) = 0$ and $\nabla g(\mathbf{a}_i) = \mathbf{0}$. Next, we consider the gradient of $c_i(\tilde{\mathbf{a}}_i)$ with respect to $\tilde{\mathbf{a}}_i$, given by

$$\nabla c_i(\tilde{\mathbf{a}}_i) = 2A\tilde{\mathbf{a}}_i - \gamma \left(4B\tilde{\mathbf{a}}_i(\tilde{\mathbf{a}}_i^{\mathrm{T}} B \tilde{\mathbf{a}}_i - 1) + 2\sum_{p=1}^{i-1} (\mathbf{a}_p^{\mathrm{T}} \tilde{\mathbf{a}}_i) \mathbf{a}_p \right). \tag{26}$$

Therefore, an updating equation for the ith composite coordinate mapping vector approximation can be formulated as follows,

$$\begin{aligned} \tilde{\mathbf{a}}_i(l+1) &= \tilde{\mathbf{a}}_i(l) + \frac{\beta}{2} \left(\nabla c_i(\tilde{\mathbf{a}}_i(l)) \right) \\ &= \tilde{\mathbf{a}}_i(l) + \beta A \tilde{\mathbf{a}}_i(l) - 2\gamma\beta B \tilde{\mathbf{a}}_i(l) \left(\tilde{\mathbf{a}}_i^{\mathrm{T}}(l) B \tilde{\mathbf{a}}_i(l) - 1 \right) \\ &\quad - \gamma\beta \sum_{p=1}^{i-1} (\dot{\mathbf{a}}_p^{\mathrm{T}} \tilde{\mathbf{a}}_i(l)) \dot{\mathbf{a}}_p \end{aligned} \tag{27}$$

where $\dot{\mathbf{a}}_p$ for $p = 1, 2, \ldots, i-1$ are the previously obtained composite coordinate mapping vectors. Clearly, the updating equation for the ith multi-channel correlation sum is given by

$$\tilde{\lambda}_i(l+1) = \tilde{\mathbf{a}}_i^{\mathrm{T}}(l) A \tilde{\mathbf{a}}_i(l). \tag{28}$$

The steepest descent-based updating equations in (27) and (28) require knowledge of two exact correlation matrices, namely, A and B, which must be computed beforehand. This may be computationally demanding for some applications. However, by using time updating equations for these correlation matrices we can arrive at the LMS-based learning for computing all the multi-channel coordinates and correlation sums directly from the multi-channel data samples. To accomplish this we substitute the matrices A and B in (27) with their instantaneous rank-one approximations $\hat{A}(l)$ and $\hat{B}(l)$ given by

$$\begin{aligned} \hat{A}(l) &= \mathbf{z}(l)\mathbf{z}(l)^{\mathrm{T}}, \quad \text{and} \\ \hat{B}(l) &= \mathrm{Diag}[\mathbf{x}_1(l)\mathbf{x}_1(l)^{\mathrm{T}}, \ldots, \mathbf{x}_n(l)\mathbf{x}_n(l)^{\mathrm{T}}], \end{aligned} \tag{29}$$

where $\mathbf{z}(l) = [\mathbf{x}_1(l)^{\mathrm{T}}, \mathbf{x}_2(l)^{\mathrm{T}}, \ldots, \mathbf{x}_n(l)^{\mathrm{T}}]^{\mathrm{T}}$ is the lth sample of the composite data vector. This leads to an LMS-based updating for $\tilde{\mathbf{a}}_i$, as

$$\begin{aligned} \tilde{\mathbf{a}}_i(l+1) &= \tilde{\mathbf{a}}_i(l) + \beta \hat{A}(l) \tilde{\mathbf{a}}_i(l) \\ &\quad - 2\gamma\beta \hat{B}(l) \tilde{\mathbf{a}}_i(l) \left(\tilde{\mathbf{a}}_i^{\mathrm{T}}(l) \hat{B}(l) \tilde{\mathbf{a}}_i(l) - 1 \right) \\ &\quad - \gamma\beta \sum_{p=1}^{i-1} (\dot{\mathbf{a}}_p^{\mathrm{T}} \tilde{\mathbf{a}}_i(l)) \dot{\mathbf{a}}_p. \end{aligned} \tag{30}$$

To find an updating equation for $\tilde{\lambda}_i$ that does not depend on the exact correlation matrix A, we simply substitute the correlation matrix A with its corresponding sample equivalent (see Gubner (2006)), i.e.

$$A \approx \tilde{A}(l) = \frac{1}{l} \sum_{h=1}^{l} \mathbf{z}(h)\mathbf{z}^{\mathrm{T}}(h) \tag{31}$$

where l is the index of the most recently obtained data sample. Hence, the updating equation for $\tilde{\lambda}_i$ becomes

$$\begin{aligned}\tilde{\lambda}_i(l+1) &= \tilde{\mathbf{a}}_i^{\mathrm{T}}(l)\tilde{A}(l+1)\tilde{\mathbf{a}}_i(l)\\ &= \tilde{\mathbf{a}}_i^{\mathrm{T}}(l)\left(\frac{1}{l+1}\sum_{h=1}^{l+1}\mathbf{z}(h)\mathbf{z}^{\mathrm{T}}(h)\right)\tilde{\mathbf{a}}_i(l)\\ &= \frac{1}{l+1}\tilde{\mathbf{a}}_i^{\mathrm{T}}(l)\left(\sum_{h=1}^{l}\mathbf{z}(h)\mathbf{z}^{\mathrm{T}}(h)+\mathbf{z}(l+1)\mathbf{z}^{\mathrm{T}}(l+1)\right)\tilde{\mathbf{a}}_i(l)\\ &= \frac{1}{l+1}\tilde{\mathbf{a}}_i^{\mathrm{T}}(l)\left(l\tilde{A}(l)+\mathbf{z}(l+1)\mathbf{z}^{\mathrm{T}}(l+1)\right)\tilde{\mathbf{a}}_i(l)\\ &= \frac{l}{l+1}\tilde{\lambda}_i(l)+\frac{1}{l+1}(\tilde{\mathbf{a}}_i^{\mathrm{T}}(l)\mathbf{z}(l+1))^2. \end{aligned}\tag{32}$$

Clearly, the LMS-based updating Eqs. (30) and (32) comprise an iterative learning process for the recursive extraction of the ith set of multi-channel coordinates and corresponding correlation sum, for $i = 2, 3, \ldots, u$. After convergence, $\tilde{\mathbf{a}}_i = \dot{\mathbf{a}}_i$ will be the ith composite coordinate mapping vector that yields the correlation sum $\tilde{\lambda}_i = \dot{\lambda}_i$, each to within some acceptable error as previously discussed. The convergence property of the updating equations in (30) and (32) follows that of the standard LMS learning (Widrow & Stearns, 1985). Note that the LMS versions of (19) and (22) can easily be developed using the same approach. To start this iterative learning we select appropriate values for the step size β and penalty parameter γ, set $l = 0$, $\tilde{\lambda}_i(0) = 0$, and randomly initialize $\tilde{\mathbf{a}}_i(0)$ such that it has unit length, for every new i.

5. Implementation and results

In this section, we show the usefulness of the proposed iterative MCCA algorithm for recursive extraction of multi-channel coordinate mapping vectors and correlation sums. The performance of the algorithm is demonstrated by comparing the sample coordinate mapping vectors and sample correlation sums that are obtained by directly solving the eigenvalue problem $A\mathbf{a}_i = \lambda_i B\mathbf{a}_i$ with those obtained via the proposed iterative algorithm. This comparison is made by plotting the normalized norm squared error of the sample coordinate mapping vector estimates and the normalized squared error of the sample correlation sum estimates. The normalized norm squared error of the ith sample multi-channel coordinate mapping vector estimation, for $i = 1, 2, \ldots, u$, at iteration l is

$$e_{\mathbf{a}_i}^2(l) = \left(\frac{\|\mathbf{a}_i - \tilde{\mathbf{a}}_i(l)\|_2}{\|\mathbf{a}_i\|_2}\right)^2. \tag{33}$$

Also, the normalized squared error of the ith sample multi-channel correlation sum estimation is

$$e_{\lambda_i}^2(l) = \left(\frac{\lambda_i - \tilde{\lambda}_i(l)}{\lambda_i}\right)^2, \tag{34}$$

where $\mathbf{a}_i$ and λ_i are obtained via the standard method of MCCA discussed in Section 2, and $\tilde{\mathbf{a}}_i(l)$ and $\tilde{\lambda}_i(l)$ are the estimated ith sample multi-channel coordinate mapping vector and correlation sum, respectively, at iteration l generated using the iterative MCCA. In what follows we present the results on two examples that demonstrate the effectiveness of the proposed iterative MCCA algorithm.

5.1. Synthesized data

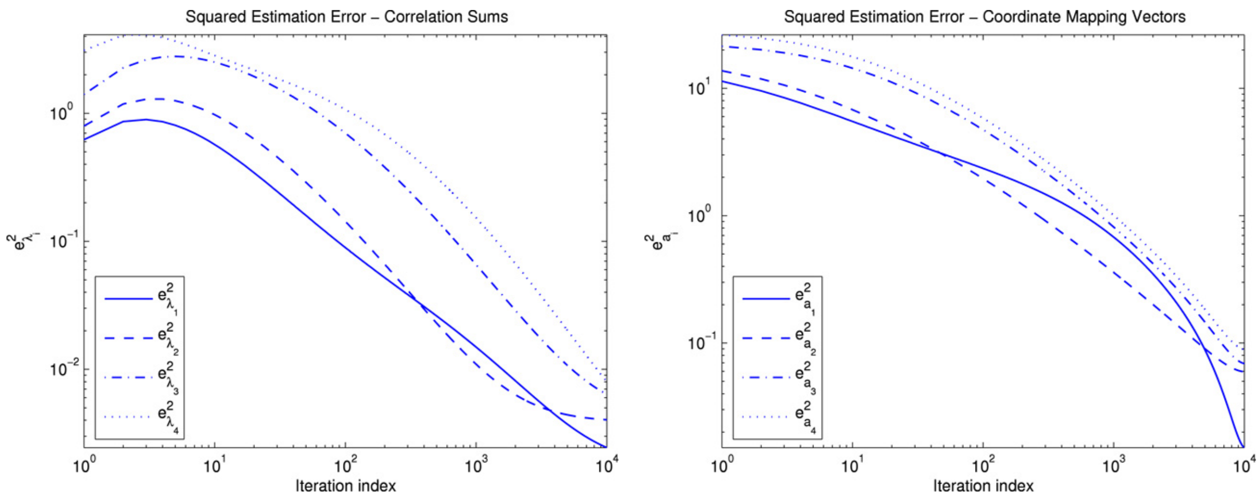
A synthesized three-channel data set consisting of 10^3 samples from each channel, $\mathbf{x}_1 \in \mathbb{R}^{4\times 1}$, $\mathbf{x}_2 \in \mathbb{R}^{5\times 1}$, and $\mathbf{x}_3 \in \mathbb{R}^{6\times 1}$, is used. These data channels are generated by the linear models

$$\begin{aligned}\mathbf{x}_1 &= H_1\eta_{\mathbf{x}_1}\\ \mathbf{x}_2 &= H_2\eta_{\mathbf{x}_2} + H_{21}\mathbf{x}_1\\ \mathbf{x}_3 &= H_3\eta_{\mathbf{x}_3} + H_{31}\mathbf{x}_1 + H_{32}\mathbf{x}_2,\end{aligned}\tag{35}$$

so that linear dependence exists between the data channels. Matrices $H_1 \in \mathbb{R}^{4\times 4}$, $H_2 \in \mathbb{R}^{5\times 5}$, $H_3 \in \mathbb{R}^{6\times 6}$, $H_{21} \in \mathbb{R}^{5\times 4}$, $H_{31} \in \mathbb{R}^{6\times 4}$, and $H_{32} \in \mathbb{R}^{6\times 5}$ are used to synthesize the three data channels from $\eta_{\mathbf{x}_1} : N(\boldsymbol{\mu}_1, C_{11})$, $\eta_{\mathbf{x}_2} : N(\boldsymbol{\mu}_2, C_{22})$, and $\eta_{\mathbf{x}_3} : N(\boldsymbol{\mu}_3, C_{33})$, which are three independent Gaussian random vectors in $\mathbb{R}^{4\times 1}$, $\mathbb{R}^{5\times 1}$, and $\mathbb{R}^{6\times 1}$, respectively. A total of 10^4 updating iterations were executed for this simulation. The step size β was linearly varied from 10^{-3} to 10^{-5}, decrementing at each iteration.

The squared estimation errors of all 4 multi-channel correlations and coordinate mapping vectors are plotted versus iteration index (log–log scale) in Fig. 1(a) and (b), respectively. From these plots, it can be seen that the estimated values approach their true values as the number of updating iterations increases, hence validating the usefulness of the proposed algorithm. It is noteworthy to mention that the squared estimation errors of the first and second coordinate mapping vectors and the correlation sums approach zero faster than the third and fourth. This is due to the recursive nature of this algorithm and accumulation of the round-off or computational errors for iterative extraction of each additional mapping vector and correlation sum.

To investigate the importance of correlation sums in measuring multi-channel coherence, similar to the role of canonical correlations in CCA, two special extreme cases of linear dependence between the data channels were implemented. For the first case, linearly independent channels were synthesized by setting H_{21}, H_{31}, and H_{32} to zero matrices. Taking into account the symmetry of $R_{\mathbf{v}_i\mathbf{v}_i}$ in (5) and the unit trace constraint, $(\lambda_i - 1)/2$ captures the sum of all cross-correlations between the ith multi-channel coordinates. For this case, these were 0.0751, 0.0598, 0.0458, and 0.0317. For the second case, linearly dependent channels were synthesized by setting H_2 and H_3 to nearly zero matrices to avoid singularity problems of A and B matrices. The sum of all cross-correlations between the ith multi-channel coordinates were 1.0000, 0.9995, 0.9992, and 0.9985, which indicate much larger and consistent multi-channel correlation sums when compared to those of the linearly independent case. This implies that, similar to canonical correlations in CCA, the correlation sums in MCCA are measures of coherence between the multi-channel data hence demonstrating the usefulness of MCCA for many applications that require coherence-based feature



(a) Correlation estimation errors.

(b) Coordinate mapping vector estimation errors.

Fig. 1. Estimation errors versus index of updating epoch.

Table 1
Characteristics of Meteosat 8-SEVIRI spectral bands

#	Type	Central wavelength (μm)	Primary use
1	Visible	0.635	Land/Cloud/Aerosol boundaries
2	Visible	0.81	Land/Cloud/Aerosol boundaries
3	Visible	1.64	Land/Cloud/Aerosol boundaries
4	Infrared	3.90	Surface/Cloud temperature
5	Water vapor	6.25	Water vapor
6	Water vapor	7.35	Water vapor
7	Infrared	8.70	Cloud properties
8	Infrared	9.66	Ozone
9	Infrared	10.80	Surface/Cloud temperature
10	Infrared	12.00	Surface/Cloud temperature
11	Infrared	13.40	Carbon dioxide

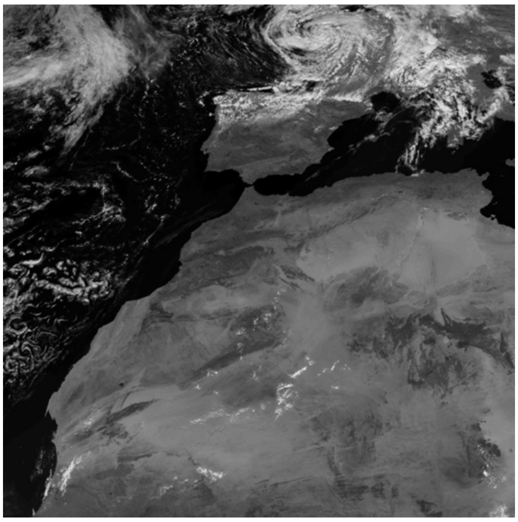
extraction (Thompson, Cartmill, Azimi-Sadjadi, & Schock, 2006), e.g. the second example in this paper.

5.2. Multi-spectral satellite imagery data

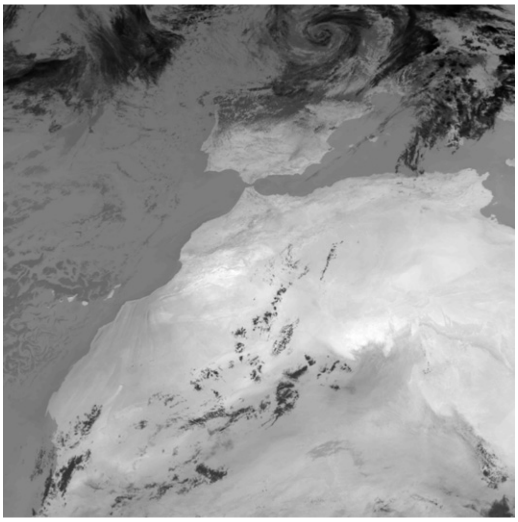
This example is intended to further demonstrate the real usefulness of the proposed MCCA algorithm on the geostationary Meteosat 8-SEVIRI satellite imagery data (Schmetz et al., 2002) obtained on July 8th, 2004. This Meteosat Second Generation (MSG) satellite produces images of the planet, a region centered on 0 degree longitude, at high temporal frequency and resolution. For every fifteen minute time segment between the hours of 5:00 and 19:00 Universal Time Coordinated (UTC), Meteosat 8-SEVIRI collects multi-spectral high resolution images at eleven different spectral bands over a particular region. Table 1 lists the Meteosat 8-SEVIRI spectral bands and their corresponding characteristics.

The images acquired (Falcone et al., 2007) at 12:00 UTC, July 8th, 2004, for all eleven spectral bands are used in this study to form three channels of data, $\mathbf{x}_1 \in \mathbb{R}^{3\times1}$, $\mathbf{x}_2 \in \mathbb{R}^{4\times1}$, and $\mathbf{x}_3 \in \mathbb{R}^{4\times1}$, respectively. The partitioning of the eleven multi-spectral bands into these three channels is decided based upon the physical and meteorological characteristics of the data at each band. More specifically, the first data channel contains the three spectral bands primarily used for land/cloud/aerosol boundary analysis (visible bands 1–3 in Table 1 while the second data channel contains the four spectral bands primarily used for surface/cloud temperature and cloud property analysis (IR bands 4, 7, 9 and 10). Finally, the third data channel contains the remaining four spectral bands (bands 5, 6, 8 and 11) primarily used for water vapor, ozone, and carbon dioxide analysis. Fig. 2(a)–(c) show an image from each of these three channels obtained from Meteosat 8-SEVIRI for spectral bands 2, 9, and 6, respectively. The area covered in these figures extend from Spain at the top in the middle, north-eastern Africa in central and bottom right, Italy and the Mediterranean Sea in the top right, and part of the Atlantic ocean on the left side of the image.

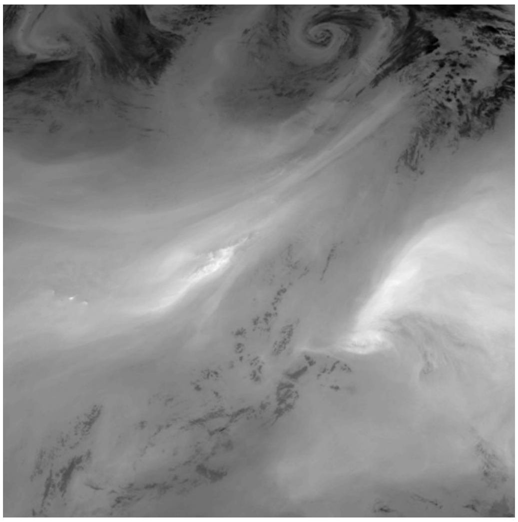
The main goal of performing MCCA on this data set in the manner described above is to find coherence across grouped spectral channels, i.e. the extent of dependence (or mutual information) across the three groupings of the Meteosat 8-SEVIRI images. A potential practical application is to quantify the importance of not only spectral channels but also specific spectral features for cloud/land data analysis. In this case, the particular MCCA implementation in this study would allow one to investigate the importance of, for example, channel 3



(a) Spectral band 2 (Land/Cloud/Aerosol boundaries).



(b) Spectral band 9 (Surface/Cloud temperature).



(c) Spectral band 6 (Water vapor).

Fig. 2. Meteosat 8-SEVIRI images.

features on this analysis. Note that MCCA could also be used to determine incoherence (or independence) across the spectral channels by examining the subdominant correlation sums.

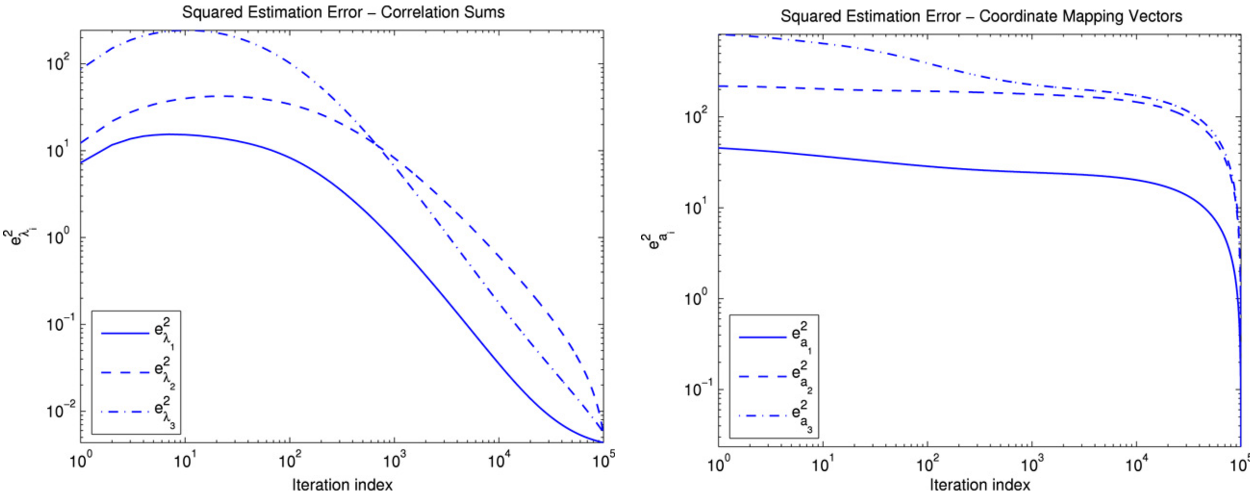
A subset of the image pixels in all the spectral bands is randomly selected to perform iterative MCCA. This subset constitutes only 50% of all available samples per each image. A total of 10^5 updating iterations were performed for this example. The step size β was linearly varied from 10^{-3} to 10^{-6} during the course of the learning. The penalty parameter γ was empirically chosen to be 10.

The squared estimation errors of all 3 multi-channel correlations and coordinate mapping vectors are plotted versus iteration index (log–log scale) in Fig. 3(a) and (b), respectively. Similar results observed from the first example are seen from these plots. As can be seen from these figures, the estimated values approach their true values as the number of updating iterations increases, hence further validating the usefulness of the proposed algorithm on real-life problems. The final values for the norm squared estimation errors of the three sample multi-channel coordinate mapping vectors are $e^2_{\mathbf{a}_1} = 0.0236$, $e^2_{\mathbf{a}_2} = 0.0249$, and $e^2_{\mathbf{a}_3} = 0.0632$, respectively. These values also validate the same conclusion made in the first example which indicated that the final values for the estimation errors of the sample multi-channel coordinate mapping vectors and the corresponding correlation sums increase in the order they are obtained.

6. Conclusion

The MCCA method (Horst, 1961; Kettingring, 1971; Nielsen, 2002; Steel, 1951; Vinograde, 1950), that allows for coherence analysis to be performed on more than two data channels in a manner similar to two-channel CCA, is considered in this paper. A common formulation of the MCCA problem (Horst, 1961; Kettingring, 1971; Nielsen, 2002; Steel, 1951) with respect to the choice of objective function and constraint was briefly reviewed. This standard formulation of the MCCA problem requires the knowledge of correlation matrices between all pairs of data channels and solving eigenvalue problems. Moreover, all correlation sums and the associated coordinates are computed even though in most practical applications only the dominant ones are of prime interest. In this paper an iterative algorithm that implements a data-driven version of MCCA is developed, analyzed, and tested. This LMS-based learning enables the recursive extraction of multi-channel coordinates and correlation sums directly from the data while avoiding the computation of the correlation matrices for all pairs of data channels as well as solving eigenvalue problems. Moreover, the ability to obtain only the principal coordinates and correlation sums with the highest contribution to coherence is provided by the proposed iterative approach.

The effectiveness of the proposed iterative method was demonstrated via two examples that used synthesized data and multi-spectral satellite imagery data. The convergence of the algorithm was empirically verified via plots of norm squared estimation errors of the MCCA coordinate mapping vectors and



(a) Correlation sum estimation errors.

(b) Coordinate mapping vector estimation errors.

Fig. 3. Estimation errors versus iteration index.

the squared error of the correlation sums versus iteration index. The theoretical convergence of the proposed iterative algorithm follows that of the LMS algorithm (Widrow & Stearns, 1985). The results showed that, the estimates of the MCCA coordinate mapping vectors and correlation sums did approach their true values as the number of iterations increased. Nonetheless, compared to the standard MCCA method in Section 2, the calculated coordinates and correlation sums obtained using this iterative algorithm exhibit some errors due to the accumulation of numerical/rounding errors of prior coordinate mapping vectors and correlation sums. The numerical error accumulation effect was demonstrated in the results of both examples, where the final values for the norm squared estimation error of the ith sample multi-channel coordinate mapping vector, and the squared estimation error of the corresponding sample correlation sum were shown to increase as i increased.

Acknowledgments

The authors would like to acknowledge the Office of Naval Research (ONR-321) for their support under contract #N00014-05-1-0014. The authors would also like to thank the Cooperative Institute for Research in the Atmosphere (CIRA) for providing the Meteosat 8-SEVIRI data set for this study.

References

Chong, E., & Zak, S. (2001). *An introduction to optimization* (2nd ed.). New York, NY, USA: Wiley-Interscience.

Falcone, A., Azimi-Sadjadi, M., & Kankiewicz, A. (2007). Dual-satellite cloud product generation using temporally updated canonical coordinate features. *IEEE Transactions on Geoscience and Remote Sensing*, *45*, 1046–1060.

Golub, G., & Loan, C. V. (1996). *Matrix computations* (3rd ed.). John Hopkins University Press.

Gubner, J. (2006). *Probability and random processes for electrical and computer engineers*. Cambridge University Press.

Horst, P. (1961). Relations among *m* sets of measures. *Psychometrika*, *26*, 129–149.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, *28*, 321–377.

Kay, J. (2000). *Neural networks for unsupervised learning based on information theory*. New York, NY, USA: Oxford University Press, Inc.

Kettingring, J. (1971). Canonical analysis of several sets of variables. *Biometrika*, *58*, 433–451.

Nielsen, A. (1994) Analysis of regularly and irregularly sampled spatial, multivariate, and multi-temporal data. *Ph.D. thesis*, Informatics and Mathematical Modeling, Technical University of Denmark, Richard Petersens Plads, Building 321, DK-2800 Kongens Lyngby, Denmark.

Nielsen, A. (2002). Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data. *IEEE Transactions on Image Processing*, *11*, 293–305.

Pezeshki, A., Azimi-Sadjadi, M., & Scharf, L. (2003). A network for recursive extraction of canonical coordinates. *Neural Networks*, *16*, 801–808.

Pezeshki, A., Azimi-Sadjadi, M., & Scharf, L. (2007). Undersea target classification using canonical correlation analysis. *IEEE Journal of Oceanic Engineering*, *32*(4), 948–955.

Pezeshki, A., Scharf, L., Azimi-Sadjadi, M., & Hua, Y. (2005). Two-channel constrained least squares problems: Solutions using power methods and connections with canonical coordinates. *IEEE Transactions on Signal Processing*, *53*(1), 121–135.

Rencher, A. C. (2002). *Methods of multivariate analysis* (2nd ed.). New York, NY, USA: Wiley-Interscience.

Scharf, L., & Mullis, C. (2000). Canonical coordinates and the geometry of inference, rate, and capacity. *IEEE Transactions on Signal Processing*, *48*(3), 824–831.

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., et al. (2002). An introduction to meteosat second generation (msg). *Bulletin of the American Meteorological Society*, *83*, 977–992.

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423. 623–656.

Steel, R. (1951). Minimum generalized variance for a set of linear functions. *Annals of Mathematical Statistics*, *22*, 456–460.

Thompson, B., Cartmill, J., Azimi-Sadjadi, M., & Schock, S. (2006). A multichannel canonical correlation analysis feature extraction with application to buried underwater target classification. In *Proc. of the intern. joint conf. on neural networks* (pp. 4413–4420).

Thompson, B., & Azimi-Sadjadi, M. (2007). Aniterative learning algorithm for multi-channel coherence analysis. In *Proceedings of IJCNN '07*.

Vinograde, B. (1950). Canonical positive definite matrices under internal linear transformations. *Proceedings of the American Mathematical Society*, *11*, 159–161.

Widrow, B., & Stearns, S. (1985). *Adaptive signal processing*. Prentice Hall.