



HAL
open science

Approximate Joint Singular Value Decomposition of an Asymmetric Rectangular Matrix Set

Marco Congedo, Ronald Phlypo, Dinh-Tuan Pham

► **To cite this version:**

Marco Congedo, Ronald Phlypo, Dinh-Tuan Pham. Approximate Joint Singular Value Decomposition of an Asymmetric Rectangular Matrix Set. *IEEE Transactions on Signal Processing*, 2011, 59 (1), pp.415-424. 10.1109/TSP.2010.2087018 . hal-00559251

HAL Id: hal-00559251

<https://hal.science/hal-00559251>

Submitted on 25 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Approximate Joint Singular Value Decomposition of an Asymmetric Rectangular Matrix Set

Marco Congedo, Ronald Phlypo, *Member, IEEE*, and Dinh-Tuan Pham, *Member, IEEE*,

Abstract—The singular value decomposition $C = U\Lambda V^T$ is among the most useful and widespread tools in linear algebra. Often in engineering a multitude of matrices with common latent structure are available. Suppose we have a set of K matrices $\{C_1, \dots, C_K\}$ for which we wish to find two orthogonal matrices U and V such that all products $U^T C_k V$ are as close as possible to rectangular diagonal form. We show that the problem can be solved efficiently by iterating either power iterations followed by an orthogonalization process or Givens rotations. The two proposed algorithms can be seen as a generalization of approximate joint diagonalization (AJD) algorithms to the bilinear orthogonal forms. Indeed, if the input matrices are symmetric and $U = V$, the optimization problem reduces to that of orthogonal AJD. The effectiveness of the algorithms is shown with numerical simulations and the analysis of a large database of 84 electroencephalographic recordings. The proposed algorithms open the road to new applications of the blind source separation framework, of which we give some example for electroencephalographic data.

Index Terms—bilinear orthogonal approximate joint diagonalization, singular value decomposition, Lödwin orthogonalization, power iteration, Jacobi iterations, Givens rotations

I. INTRODUCTION

Whereas the blind source separation (BSS) problem can be solved by the joint diagonalization of two symmetric matrices [1], [2], it has been soon discovered that more robust and efficient blind source separation algorithms can be obtained by joint diagonalization of a symmetric matrix set [3], [4], [5] [6, For a review see]. The joint diagonalization of a matrix set perturbed by noise (e.g., due to estimates of data statistics) does not have a closed form solution and can be obtained only approximately. From the point of view of linear algebra, the approximate joint diagonalization (AJD) extends the eigenvalue-eiegnvector decomposition (single matrix diagonalization) and its generalized version (two matrix diagonalization) to a matrix set including three or more matrices. As such, it is a very general and flexible tool. Iterative algorithms solving the AJD problem are studied intensively in the signal processing community [5], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19]. These advances have allowed for Blind Source Separation (BSS) applications

M. Congedo and R. Phlypo are with the ViBS (Vision and Brain Signal Processing) Research Group, GIPSA lab, CNRS, INPG, 961, rue de la Houille Blanche, BP 46, 38402 Grenoble, France. E-mail: marco.congedo@gmail.com, ronald.phlypo@gipsa-lab.grenoble-inp.fr

D.-T. Pham is with IMAG, CNRS, 51 rue des Mathématiques, BP 53, 38041 Grenoble CEDEX 9, France. E-mail: Dinh-Tuan.Pham@imag.fr

Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org

in several fields such as speech enhancement, image processing, geophysical data analysis, wireless communication and biological signal analysis [20].

The AJD algorithms proposed so far differ according to the restrictions imposed on the matrices that can be diagonalized (Hermitian/symmetric, positive semi-definite, normal), the restrictions imposed on the joint diagonalizer sought (unitary/orthogonal or full-column rank), their convergence rate and computational complexity per iteration. More importantly, they differ in terms of the cost function to be optimized. For the sake of efficiency some algorithms rely on heuristics (e.g., [19], [9], [14]), whereas others have focused on the more natural formulation based on Frobenius norm off-diagonal minimization [15], [16], [8], [17], [10], [21]. Restricting the joint diagonalizer to the unitary or orthogonal case, a straightforward least-squares objective function exists [4], [21]. In this case, BSS is achieved with the classical two-step procedure often encountered in independent component analysis (ICA); data is whitened first, leaving the solution up to a rotation, which is then found by unitary/orthogonal AJD algorithms [4], [22], [23]. The work described in this paper operates within this latter framework. Historically, AJD algorithms have addressed the joint diagonalization of Hermitian or symmetric matrices. Therefore a joint diagonalizer is sought so as to span both the column space and the row space of the matrix set to be diagonalized¹. As a consequence, the decomposition of the input signal is performed along one dimension, typically taken as the space, for example in electroencephalography (EEG), or the time, for example in functional magnetic resonance imaging. In this paper, we pose a different AJD problem, which is actually a bi-linear AJD. Indeed, we no longer require a single joint diagonalization matrix spanning both the column and row space, instead, we require a pair of joint diagonalizers with mutually orthogonal columns. One of the matrices spans the column space while the other the row space of the matrix set. This problem allows the AJD of non-symmetric, possibly rectangular matrices. From the point of view of linear algebra the resulting decomposition is an extension of the singular value decomposition to a set comprising more than two matrices, which we name Approximate Joint Singular Value Decomposition (AJSVD). As such, it promises to become an even more general and flexible tool, which reduces to the previous AJD case for symmetric matrix sets. To solve

¹There has been also some work on the approximate joint diagonalization of normal matrices (possibly asymmetric), for which the column space is spanned by the columns of the joint diagonalizer and the row space by the rows of its inverse [18]. However, such decomposition does not find application in the BSS context.

the AJSVD problem we propose two iterative algorithms addressing the least-squares objective function, opening the road to new BSS approaches. For instance, we may construct matrix sets living in two dimensions, for example, space and time, yielding as solution both bases simultaneously. We are also enabled to study double sensor array recordings, wherein the matrix set can be formed by cross-statistics between two sensor arrays in order to concentrate on their similarity. So far the bilinear AJD problem has been addressed only in the computer vision community. In [24], Jacobi iterations were proposed for the square case without allowing for dimension reduction. In [25], a gradient approach to the bilinear AJD problem has been suggested, based on the natural gradient approach. This algorithm has shown to exhibit slow convergence, since only small step sizes are allowed to keep the updates in the manifold of orthonormal matrices.

This paper is organized as it follows: in section II we report the notation used throughout. In section III we motivate our study with the practical example of BSS in electroencephalography (EEG) and we illustrate possible applications of our algorithms in the context of EEG. In section IV we formalize the bi-linear AJD in the real domain and we pose the optimization problem within a least-squares framework. We then show that it is possible to decouple the optimization problem in a number of simple quadratic form optimizations, alternating between the estimation of the two joint diagonalizers to be found. Then, we present two simple and efficient iterative algorithms; the first optimizes the least-squares objective function by power iterations and imposes unitarity/orthogonality at the end of each iteration, extending upon previous work in [15], [16], [26]. The second operates by Givens rotations, extending upon the seminal work of [4]. We also provide a suggestion for initialization, which applies in general to all unitary/orthogonal AJD algorithms. In section V we present results of numerical simulations and the diagonalization of cospectral matrices of a large (N=84) EEG normative database, both witnessing in favor of the good performance of the proposed algorithms. Finally, section VI includes our discussion, where we indicate the extension to the complex case and we address the limitations of our work.

II. NOTATIONS

In the following we will indicate matrices by upper case bold face characters (\mathbf{A}), matrix sets by calligraphic letters (\mathcal{A}), vectors by lower case bold face (\mathbf{a}), indices and scalars by lower case italic (a) characters. The i -th column of \mathbf{A} will be denoted by \mathbf{a}_i , unless ambiguity between rows or columns may exist, in which case we will use \mathbf{A}_i and $\mathbf{A}_{\cdot i}$ to denote the i -th row, respectively column of \mathbf{A} . Constants will be noted in upper case italics (A). Notations $tr(\cdot)$, $(\cdot)^T$ and $\|\cdot\|$ indicate the trace of a matrix, the transpose operator and the Frobenius norm, respectively. The operators $\text{Diag}(\cdot)$ and $\text{Off}(\cdot)$ return a matrix nullifying the off-diagonal or the diagonal elements of its argument, respectively. $\text{Diag}_A(\cdot)$ nullifies not only the off-diagonal elements of its argument but also the elements on the diagonal with column index greater than A . All these operators are defined for both square and rectangular matrices.

The identity matrix in $\mathbb{R}^{A \times A}$ is denoted by \mathbf{I}_A , the matrix of zeros in $\mathbb{R}^{A \times A}$ by $\mathbf{0}_A$ and $|a|$ is the absolute value of a . $\lambda_n(\mathbf{A})$ will denote the n -th eigenvalue of the matrix \mathbf{A} , where n is an arbitrary indexation. We will use the terms orthonormal and orthogonal for matrices $\mathbf{A} \in \mathbb{R}^{A \times A}$ for which $\mathbf{A}^T \mathbf{A} = \mathbf{A} \mathbf{A}^T = \mathbf{I}_A$ and matrices $\mathbf{A} \in \mathbb{R}^{A_1 \times A_2}$, $A_1 \geq A_2$ for which $\mathbf{A}^T \mathbf{A} = \mathbf{I}_{A_2}$, respectively.

III. MOTIVATION

BSS by AJD of second-order sensor statistics offers a flexible approach to separate EEG sources based on their characteristic non-stationarity, spectral profile and/or experimental condition membership. The sources of interest are equivalent electric dipoles formed by assemblies of locally synchronized pyramidal cells in the cortex. The currents induced by these dipoles diffuses through the head and reaches the scalp sensors (electrodes). This process can be seen as a linear mixing process in good approximation. In the EEG literature the mixing phenomenon is known as the "volume conduction effect", see [26] for a review.

As a consequence, we may write the observed potentials $\mathbf{x}(t) \in \mathbb{R}^P$ as

$$\mathbf{x}(t) = \mathbf{A} \mathbf{s}(t) , \quad (1)$$

where \mathbf{A} is a time-invariant, full column rank mixing matrix and $\mathbf{s}(t) \in \mathbb{R}^N$, $N \leq P$ holds the unknown time courses of the sources components. The AJD scheme for BSS consists in estimating a number of second-order statistical matrices directly from the multivariate observed measurements $\mathbf{x}(t)$, e.g., Fourier cospectral matrices, covariance matrices estimated under different experimental conditions or covariance matrices estimated in different temporal windows [26]. Following Eq. (1), such matrices have theoretical form $\mathbf{C}_k = \mathbf{A} \mathbf{\Lambda}_k \mathbf{A}^T$, where the $\mathbf{\Lambda}_k$ represent the diagonal matrices of the unknown, true source statistics, thus we work under the assumption that the sources are spatially uncorrelated.

Under rather general conditions, it can then be shown that the AJD of the matrix set results in the *identification* of the demixing matrix $\mathbf{B}^T = \widehat{\mathbf{A}}^{-1}$, yielding source estimates

$$\hat{\mathbf{s}}(t) = \mathbf{B}^T \mathbf{x}(t) = \mathbf{B}^T \mathbf{A} \mathbf{s}(t) ,$$

up to the usual order and scaling indeterminacies [20], [23], i.e., $\mathbf{B}^T \mathbf{A} = \mathbf{P} \mathbf{\Lambda}$, where \mathbf{P} and $\mathbf{\Lambda}$ are respectively a permutation and a diagonal scaling matrix. This identification scheme is named blind because it does not rely on the knowledge of the mixing process. Furthermore, it does not depend on the mixing process itself [27], but only on the uniqueness of the source statistics profiles $\lambda_n(\mathbf{C}_k)$, $k = 1, 2 \dots K$ along the diagonalization set. Precisely, following [28] (see also [29], [30]) let us define

$$\rho = \max_{m,n} \left| \frac{\sum_k \lambda_n(\mathbf{C}_k) \lambda_m(\mathbf{C}_k)}{\sqrt{\sum_k [\lambda_n(\mathbf{C}_k) \lambda_n(\mathbf{C}_k)] \sum_k [\lambda_m(\mathbf{C}_k) \lambda_m(\mathbf{C}_k)]}} \right| ,$$

as the modulus of uniqueness of a diagonalization set; then $\rho < 1$ (no two such profiles are colinear) is the necessary and sufficient condition for source identification $\hat{\mathbf{s}} = \mathbf{P} \mathbf{\Lambda} \mathbf{s}$.

Upon congruence transformation of $\mathcal{C} = \{\mathbf{C}_k, k = 1, 2 \dots K\}$, the set $\{\mathbf{\Lambda}_k = \mathbf{B}^T \mathbf{C}_k \mathbf{B}\}$ contains zero off-diagonal elements

and thus outputs uncorrelated sources, effectively removing the nuisance effects of volume conduction.

So far, identification has relied only on squared matrices of source auto-statistics Λ_k . In this work we allow the Λ_k to be rectangular, so as to allow diagonalization sets of the form $\mathbf{C}_k = \mathbf{U}^T \Lambda_k \mathbf{V}$. We then require the identification of both \mathbf{U} and \mathbf{V} under orthogonality constraints, yielding the approximate joint diagonalization of an asymmetric and possibly rectangular matrix set.

In this framework, Λ_k no longer need to be restricted to source auto-statistics, but may hold cross-statistics between two observation spaces. Thanks to this generalization, new approaches to blind identification are allowed. For instance, increasing attention is currently being paid to the integration and fusion of data acquired synchronously or asynchronously from different modalities [31], such as EEG and magnetoencephalography, EEG and functional magnetic resonance imaging. Data fusion refers to the idea to use both modalities in a combined (joint) analysis, trying to find commonalities between them. A special case of data fusion is hyperscanning, wherein the same modality is recorded synchronously between two individuals. Then, after whitening the data of each observation, one may jointly diagonalize a set of their cross-statistics, allowing for a robust variant of the canonical correlation analysis. Again, one may use Fourier cospectral matrices at different frequencies or covariance matrices in different time windows. Note that the above extension to CCA (which is restricted to the case $K = 1$ or to summation over the matrix set [32]), generally lowers the modulus of uniqueness by adding matrices to the set $\{\mathbf{C}_k\}$.

Another example of this framework is a space-time analysis. Consider evoked response potential studies, where $\mathbf{C}_k, k = 1 \dots K$ are the time-locked observations whose columns are the observed potentials as a function of the time instants after stimulus onset and whose rows are with respect to the different electrode locations. After suitable transformation, one may jointly diagonalize directly the matrices $\{\mathbf{C}_k\}$ (without resorting to second or higher order statistical estimations), yielding jointly the spatial filters and the time domain patterns.

IV. METHODS

A. Model

Consider the following generative model

$$\mathbf{C}_k = \mathbf{U} \Lambda_k \mathbf{V}^T + \mathbf{N}_k ,$$

where $\mathbf{U} \in \mathbb{R}^{P \times P}$ and $\mathbf{V} \in \mathbb{R}^{Q \times Q}$ are orthogonal matrices, the matrices Λ_k are diagonal matrices in $\mathbb{R}^{P \times Q}$ with $N \leq \min(P, Q)$ non-zero elements and \mathbf{N}_k are matrices containing measurement noise and estimation errors.

Typically, the set

$$\mathcal{C} = \{\mathbf{C}_k; k = 1 \dots K, \mathbf{C}_k \in \mathbb{R}^{P \times Q}\}$$

is composed of statistical estimations of the signal part $\mathbf{U} \Lambda_k \mathbf{V}^T$ sharing a common latent structure and a randomly varying noise contribution \mathbf{N}_k . We wish to identify the two orthonormal matrices \mathbf{U} and \mathbf{V} determining the latent structure, such that

$$\mathbf{U}^T \mathbf{C}_k \mathbf{V} \approx \Lambda_k, \forall k , \quad (2)$$

where the approximation is mainly due to the non-diagonalizable noise part.

When $K = 1$, the above decomposition is the singular value decomposition (SVD) of \mathbf{C}_1 where only the N major singular values are retained on the diagonal of Λ_1 and the corresponding left and right eigenvectors are the columns of \mathbf{U} and \mathbf{V} , respectively. $\mathbf{U} \Lambda_1 \mathbf{V}^T$ is then known as the best rank- N approximation to \mathbf{C}_1 [33]. Only when $K = 1$ and $N = \text{rank}(\mathbf{C}_1)$, the equality in Eq. (2) holds, whereas for the general case where $K > 1$ it may hold only approximately.

Whenever $P = Q = N$ and $\mathbf{U} = \mathbf{V}$, the above problem reduces to the joint approximate orthogonal diagonalization of approximate Hermitian/symmetric matrices [4], [34]. In this paper, we tackle the more general problem when $P \neq Q, N \leq \min(P, Q)$ and $\mathbf{C}_k \not\approx \mathbf{C}_k^T$. We will treat the real case only, addressing the extension to the complex case in Section VI.

B. Objective Function

As is the case for the SVD of a single matrix, we shall work under the norm ℓ_2 . We consider the least-squares objective function

$$\mathcal{J}^{\text{LS}}(\mathbf{U}, \mathbf{V}, \{\Lambda_k\} | \mathcal{C}) = \sum_{k=1}^K \|\mathbf{C}_k - \mathbf{U} \Lambda_k \mathbf{V}^T\|^2 \quad (3)$$

with respect to \mathbf{U}, \mathbf{V} and the set $\{\Lambda_k\}$ of diagonal matrices with N non-zero entries.

Since $\|\mathbf{A}\|^2 = \text{tr}(\mathbf{A}^T \mathbf{A})$ and because of the invariability of the trace under orthogonal transformations, the above criterion can be rewritten as

$$\sum_{k=1}^K \left(\|\mathbf{C}_k\|^2 + \|\Lambda_k\|^2 - 2 \text{tr}(\Lambda_k \mathbf{U}^T \mathbf{C}_k \mathbf{V}) \right) .$$

For fixed \mathbf{U} and \mathbf{V} it is easily seen that the above criterion is minimized whenever

$$\Lambda_k = \text{Diag}_N(\mathbf{U}^T \mathbf{C}_k \mathbf{V}), \forall k .$$

The attained minimum is

$$\mathcal{J}^{\text{LS}}(\{\Lambda_k\} | \mathbf{U}, \mathbf{V}, \mathcal{C}) = \sum_{k=1}^K \left(\|\mathbf{C}_k\|^2 - \text{tr} \left(\left[\text{Diag}_N(\mathbf{U}^T \mathbf{C}_k \mathbf{V}) \right]^2 \right) \right)$$

One is thus led to the problem of maximizing

$$\mathcal{J}^{\text{OFF}}(\mathbf{U}, \mathbf{V} | \mathcal{C}) = \sum_{k=1}^K \text{tr} \left(\left[\text{Diag}_N(\mathbf{U}^T \mathbf{C}_k \mathbf{V}) \right]^2 \right) = \sum_{k=1}^K \sum_{i=1}^N \left[\mathbf{u}_i^T \mathbf{C}_k \mathbf{v}_i \right]^2 , \quad (4)$$

where \mathbf{u}_i and \mathbf{v}_i are the i -th columns of \mathbf{U} and \mathbf{V} , respectively, satisfying the orthonormality constraints $\mathbf{u}_i^T \mathbf{u}_j = \mathbf{v}_i^T \mathbf{v}_j = \delta_{ij}$, with δ_{ij} the Kronecker delta, taking 1 if and only if $i = j$ and 0 otherwise. For similar developments in the case of higher-order cumulants, see [21].

The expansion of Eq. (4) gives

$$\sum_{n=1}^N \sum_{k=1}^K \left(\mathbf{u}_n^T \mathbf{C}_k \mathbf{v}_n \right)^2 = \sum_{n=1}^N \mathbf{u}_n^T \mathcal{M}_n(\mathbf{V}) \mathbf{u}_n \quad (5a)$$

$$= \sum_{n=1}^N \mathbf{v}_n^T \mathcal{M}_n(\mathbf{U}) \mathbf{v}_n , \quad (5b)$$

where

$$\mathcal{M}_q(\mathbf{V}) = \sum_k \mathbf{C}_k \mathbf{v}_q \mathbf{v}_q^T \mathbf{C}_k^T, \forall q = 1 \dots Q, \quad (6a)$$

$$\mathcal{M}_p(\mathbf{U}) = \sum_k \mathbf{C}_k^T \mathbf{u}_p \mathbf{u}_p^T \mathbf{C}_k, \forall p = 1 \dots P. \quad (6b)$$

Following Eq. (5), the search over the matrices \mathbf{U} and \mathbf{V} can be expressed as the maximization of

$$\begin{aligned} \mathcal{J}^{\text{Diag}}(\mathbf{U}|\mathbf{V}, C) &= \sum_n \mathbf{u}_n^T \mathcal{M}_n(\mathbf{V}) \mathbf{u}_n, \\ \mathcal{J}^{\text{Diag}}(\mathbf{V}|\mathbf{U}, C) &= \sum_n \mathbf{v}_n^T \mathcal{M}_n(\mathbf{U}) \mathbf{v}_n. \end{aligned}$$

C. Optimization by power iterations

Eq. (5a) and (5b) are maximized by taking \mathbf{u}_n as the principal eigenvector of $\mathcal{M}_n(\mathbf{V})$ ($\forall n$) and \mathbf{v}_n as the principal eigenvector of $\mathcal{M}_n(\mathbf{U})$ ($\forall n$).

We have the following theorem:

Theorem 1: Let $\mathbf{C}_k = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T + \mathbf{N}_k$ where the entries of \mathbf{N}_k are uncorrelated with mean 0 and variance σ^2 . The matrices $\mathcal{M}_n(\mathbf{V})$ and $\mathcal{M}_n(\mathbf{U})$ have eigenvalues $E\{\lambda_n^2\} + \sigma^2$ with algebraic multiplicity 1 and σ^2 with algebraic multiplicity $P - 1$ and $Q - 1$, respectively.

Proof: We only prove the case $K \rightarrow +\infty$ and $Q \gg P$ (or $P \gg Q$). Instead of the sum over k , we use the expected value $E\{\cdot\}$.

The matrices $\mathcal{M}_n(\mathbf{V})$ can be written as $E\left\{(\mathbf{U}\mathbf{\Lambda}_k\mathbf{V}^T + \mathbf{N}_k)\mathbf{v}_n\mathbf{v}_n^T(\mathbf{U}\mathbf{\Lambda}_k\mathbf{V}^T + \mathbf{N}_k)^T\right\} = E\{\lambda_n^2\mathbf{u}_n\mathbf{u}_n^T + \sigma^2\mathbf{I}_P$. The eigenvalues are thus $E\{\lambda_n^2\} + \sigma^2$ and σ^2 with algebraic multiplicity 1 and $P - 1$, respectively, whereas the eigenvectors remain unchanged with respect to the noiseless case.

The proof goes analogously for the matrices $\mathcal{M}_n(\mathbf{U})$. ■

If $E\{\lambda_n^2\} > 0$, the principal eigenvectors of the matrices $\mathcal{M}_n(\mathbf{V})$ and $\mathcal{M}_n(\mathbf{U})$ can efficiently be obtained through power iterations. Denoting by i the iteration index, we have

$$\begin{aligned} \mathbf{u}_n^{[r+1]} &= \mathcal{M}_n(\mathbf{V})\mathbf{u}_n^{[r]} \\ \mathbf{v}_n^{[r+1]} &= \mathcal{M}_n(\mathbf{U})\mathbf{v}_n^{[r]} \end{aligned}$$

Based on theorem 1, the convergence rate of the power method for finding the principal eigenvector \mathbf{v}_n of $\mathcal{M}_n(\mathbf{U})$ is given by

$$\begin{aligned} \left| E\left\{\lambda_n^{[r]2}\right\} - E\{\lambda_n^2\} \right| &\leq \tan(\theta_0) E\{\lambda_n^2\} \left(\frac{\lambda_2(\mathcal{M}_n(\mathbf{U}))^2}{\lambda_1(\mathcal{M}_n(\mathbf{U}))^2} \right)^{2i} \\ &\propto E\{\lambda_n^2\} \left(\frac{\sigma^2}{E\{\lambda_n^2\} + \sigma^2} \right)^{2i} \\ &\approx \frac{1}{K} \sum_{k'} \lambda_n^2(\mathbf{C}_{k'}) \frac{\sigma^2}{1/K \sum_k \lambda_n^2(\mathbf{C}_k) + \sigma^2}, \end{aligned}$$

where θ_0 is the angle between the initialization $\mathbf{v}_n^{[0]}$ and the eigenvector \mathbf{v}_n .

However, \mathbf{U} and \mathbf{V} are not known and current estimates must be used, yielding

$$\begin{aligned} \mathbf{u}_n^{[r+1]} &= \mathcal{M}_n(\mathbf{V}^{[r]})\mathbf{u}_n^{[r]} \\ \mathbf{v}_n^{[r+1]} &= \mathcal{M}_n(\mathbf{U}^{[r+1]})\mathbf{v}_n^{[r]} \end{aligned}$$

Theorem 1 shows that the power iteration will converge in the neighbourhood of the solution.

Now, if the series $\mathbf{V}^{[r]}$ converges to \mathbf{V} , then $\mathbf{u}_n^{[r]}$ converges almost surely to the principal eigenvector of $\mathcal{M}_n(\mathbf{V})$ under the condition that the initial estimate contains a component in its direction. The power method has already proven useful in higher dimensional problems [35], although only for best (in the ℓ_2 -norm sense) rank-1 approximations to the data and has also been found effective in symmetric AJD algorithms [15], [16]. The repeated updating of $\mathbf{u}_n^{[r]}$ and $\mathbf{v}_n^{[r]}$ does not guarantee that $\mathbf{U}^{[r]}$ and $\mathbf{V}^{[r]}$ are orthogonal. To correct divergence from orthonormality, the matrices $\mathbf{U}^{[r+1]}$ and $\mathbf{V}^{[r+1]}$ need to be taken back into the space of orthonormal matrices after each sweep.

Orthogonalization: For transforming the matrices \mathbf{U} and \mathbf{V} to the space of orthonormal matrices after each power iteration, we have considered the Gram-Schmidt orthogonalization scheme, which is widely used in the signal processing community, and the Lödwin orthogonalization transformation [36]. While the Gram-Schmidt orthogonalization proceeds sequentially, thus propagating errors in the estimation of the vectors throughout the orthogonalization process, Lödwin's method has proved more robust. Let $\hat{\mathbf{U}}^{[r]}$ be the matrix formed by the column vectors $\mathbf{u}_n^{[r]}$ obtained through the N power iterations on $\mathbf{u}_n^{[r-1]}$ and let $\hat{\mathbf{U}}^{[r]} = \phi_{\hat{\mathbf{U}}^{[r]}} \Sigma_{\hat{\mathbf{U}}^{[r]}} \psi_{\hat{\mathbf{U}}^{[r]}}^T$ be its SVD. The projection of $\hat{\mathbf{U}}^{[r]}$ on the space of orthonormal matrices under the ℓ_2 -norm is then given by $\mathbf{U}^{[r]} = \phi_{\hat{\mathbf{U}}^{[r]}} \psi_{\hat{\mathbf{U}}^{[r]}}^T$. Similarly, we obtain $\hat{\mathbf{V}}^{[r]} = \phi_{\hat{\mathbf{V}}^{[r]}} \Sigma_{\hat{\mathbf{V}}^{[r]}} \psi_{\hat{\mathbf{V}}^{[r]}}^T$ as the SVD of $\hat{\mathbf{V}}^{[r]}$ and $\mathbf{V}^{[r]} = \phi_{\hat{\mathbf{V}}^{[r]}} \psi_{\hat{\mathbf{V}}^{[r]}}^T$. As compared with the Gram-Schmidt orthogonalization, we do not fix any vector, nor an update order.

Stopping Criterion: From the eigenvalue equations $\mathcal{M}_n(\mathbf{V})\mathbf{u}_n = \sum_k \lambda_n^2(\mathbf{C}_k)\mathbf{u}_n$ and $\mathcal{M}_n(\mathbf{U})\mathbf{v}_n = \sum_k \lambda_n^2(\mathbf{C}_k)\mathbf{v}_n$, it follows that $\|\mathcal{M}_n(\mathbf{V})\mathbf{u}_n\| = \|\mathcal{M}_n(\mathbf{U})\mathbf{v}_n\| = \sum_k \lambda_n^2(\mathbf{C}_k) = \mathbf{u}_n^T \mathcal{M}_n(\mathbf{V})\mathbf{u}_n = \mathbf{v}_n^T \mathcal{M}_n(\mathbf{U})\mathbf{v}_n$. The sum

$$\gamma^{[r+1]} = \sum_n \|\mathcal{M}_n(\mathbf{V}^{[r]})\mathbf{u}_n^{[r]}\| + \|\mathcal{M}_n(\mathbf{U}^{[r]})\mathbf{v}_n^{[r]}\| \quad (10)$$

can thus be used to determine convergence without explicit computation of the gradient of the cost function. Particularly, the algorithm can be stopped when $(1 - \gamma^{[r+1]}/\gamma^{[r]})$ approaches 0. Notice that this stopping criterion is an average criterion over the $P + Q$ decoupled optimizations.

The algorithm has been summarized in Algorithm 1.

D. Optimization by Givens rotations

To avoid the projection step after each update, one can consider updates that optimize over \mathbf{U} and \mathbf{V} in the space of orthonormal matrices. A well known approach consists of subsequent multiplicative updates within the group of rotational matrices (a subset of the orthonormal matrices consisting of those orthonormal matrices whose determinant equals 1), also known as the Jacobi iterations or subsequent Givens rotations [37].

Since left and right multiplication of \mathbf{C}_k as $\mathbf{U}^T \mathbf{C}_k \mathbf{V}$ yields a bilinear orthonormal transform of \mathbf{C}_k , it can be carried out by successive planar rotations. Let $\tilde{\mathbf{C}}_k =$

Algorithm 1 AJSVD by power iterations

Input: set of K matrices $C = \{C_k \in \mathbb{R}^{P \times Q}; k = 1 \dots K\}$
 Initialize $\mathbf{U}^{[0]} \in \mathbb{R}^{P \times P}$ and $\mathbf{V}^{[0]} \in \mathbb{R}^{Q \times Q}$ (if no information available let them be the identity matrices in $\mathbb{R}^{P \times P}$ and $\mathbb{R}^{Q \times Q}$)
 $N = \min(P, Q), r \leftarrow 0$
while no convergence **do**
 $r \leftarrow r + 1$
 for $n = 1$ to N **do**
 Calculate $\mathcal{M}_n(\mathbf{V}^{[r-1]})$ from Eq. (6a)
 $\mathbf{u}_n^{[r]} = \mathcal{M}_n(\mathbf{V}^{[r-1]})\mathbf{u}_n^{[r-1]}$
 end for
 Orthogonalize $\mathbf{U}^{[r]}$
 for $n = 1$ to N **do**
 Calculate $\mathcal{M}_n(\mathbf{U}^{[r]})$ from Eq. (6b)
 $\mathbf{v}_n^{[r]} = \mathcal{M}_n(\mathbf{U}^{[r]})\mathbf{v}_n^{[r-1]}$
 end for
 Orthogonalize $\mathbf{V}^{[r]}$
end while

$(\mathbf{U}^{[r-1]})^T C_k (\mathbf{V}^{[r-1]})$ be the transformed matrices at iteration m of the algorithm. The idea is to pick an index pair $(i, j) \in \{1 \dots n \dots N\}^2 \setminus \{(1, 1) \dots (n, n) \dots (N, N)\}$ and rotate the columns \mathbf{u}_i and \mathbf{u}_j as

$$[\mathbf{u}_i \ \mathbf{u}_j] \leftarrow [\mathbf{u}_i \cos \theta - \mathbf{u}_j \sin \theta \quad \mathbf{u}_i \sin \theta + \mathbf{u}_j \cos \theta]$$

while keeping all other columns of \mathbf{U} as well as all columns of \mathbf{V} fixed. Accordingly, the i -th row $\tilde{C}_{k;i}$ and the j -th row $\tilde{C}_{k;j}$ of \tilde{C}_k are transformed as

$$\begin{bmatrix} \tilde{C}_{k;i} \\ \tilde{C}_{k;j} \end{bmatrix} \leftarrow \begin{bmatrix} \tilde{C}_{k;i} \cos \theta - \tilde{C}_{k;j} \sin \theta \\ \tilde{C}_{k;i} \sin \theta + \tilde{C}_{k;j} \cos \theta \end{bmatrix}.$$

The optimal rotation angle should be chosen as to maximize the criterion in Eq. (4). If we choose $i < j < N$, the criterion depends on the rows $\tilde{C}_{k;i}$ and $\tilde{C}_{k;j}$ only through the terms $\sum_k \tilde{C}_{k;ii}^2$ and $\sum_k \tilde{C}_{k;jj}^2$. The angle that maximizes Eq. (4) is given by

$$\arg \max_{\theta} \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}^T \left\{ \sum_{k=1}^K \begin{bmatrix} \tilde{C}_{k;ii}^2 + \tilde{C}_{k;jj}^2 & \tilde{C}_{k;jj}\tilde{C}_{k;ij} - \tilde{C}_{k;ii}\tilde{C}_{k;ji} \\ \tilde{C}_{k;jj}\tilde{C}_{k;ij} - \tilde{C}_{k;ii}\tilde{C}_{k;ji} & \tilde{C}_{k;ji}^2 + \tilde{C}_{k;ij}^2 \end{bmatrix} \right\} \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \left\{ \begin{array}{l} \mathbf{U}^T \sum_k (C_k C_k^T) \mathbf{U} \\ \mathbf{V}^T \sum_k (C_k^T C_k) \mathbf{V} \end{array} \right\} = \sum_k \Lambda_k^2 \quad (17)$$

The criterion $\mathcal{J}^{\text{Diag}}(\mathbf{u}_i, \mathbf{u}_j | \{\mathbf{u}_k\}, \mathbf{V}, C)$ is thus maximized by choosing $[\cos \theta \ \sin \theta]^T$ as the normalized major eigenvector of the central matrix in the above expression. Similarly, when a pair $i \leq N < j \leq P$ is chosen, only the term $\sum_k \tilde{C}_{k;ii}^2$ is affected. Hence, by the above result, the criterion is maximized by choosing $[\cos \theta \ \sin \theta]^T$ as the normalized major eigenvector of

$$\sum_{k=1}^K \begin{bmatrix} \tilde{C}_{k;ii}^2 & -\tilde{C}_{k;ii}\tilde{C}_{k;ji} \\ -\tilde{C}_{k;ii}\tilde{C}_{k;ji} & \tilde{C}_{k;ji}^2 \end{bmatrix}. \quad (12)$$

Analogous reasoning yield the update rules for the columns of \tilde{V} , acting on the columns $\tilde{C}_{k;i}$ and $\tilde{C}_{k;j}$ of \tilde{C}_k . The optimal angle is found by choosing the $[\cos \theta \ \sin \theta]^T$ as the normalized

major eigenvector of

$$\sum_{k=1}^K \begin{bmatrix} \tilde{C}_{k;ii}^2 + \tilde{C}_{k;jj}^2 & \tilde{C}_{k;jj}\tilde{C}_{k;ji} - \tilde{C}_{k;ii}\tilde{C}_{k;ij} \\ \tilde{C}_{k;jj}\tilde{C}_{k;ji} - \tilde{C}_{k;ii}\tilde{C}_{k;ij} & \tilde{C}_{k;ji}^2 + \tilde{C}_{k;ij}^2 \end{bmatrix}. \quad (13)$$

whenever $i < j \leq N$ or as the normalized major eigenvector of

$$\sum_{k=1}^K \begin{bmatrix} \tilde{C}_{k;ij}^2 & -\tilde{C}_{k;ii}\tilde{C}_{k;ij} \\ -\tilde{C}_{k;ii}\tilde{C}_{k;ij} & \tilde{C}_{k;ij}^2 \end{bmatrix}. \quad (14)$$

for $i \leq N < j < Q$.

Note that the explicit calculation of θ is not needed, since one can form the matrix

$$\mathbf{G} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$$

directly from the normalized major eigenvectors and apply the following update rules

$$[\mathbf{u}_i \ \mathbf{u}_j] \leftarrow [\mathbf{u}_i \ \mathbf{u}_j] \mathbf{G} \quad (15a)$$

$$\begin{bmatrix} \tilde{C}_{k;i} \\ \tilde{C}_{k;j} \end{bmatrix} \leftarrow \mathbf{G}^T \begin{bmatrix} \tilde{C}_{k;i} \\ \tilde{C}_{k;j} \end{bmatrix} \quad (15b)$$

and

$$[\mathbf{v}_i \ \mathbf{v}_j] \leftarrow [\mathbf{v}_i \ \mathbf{v}_j] \mathbf{G} \quad (16a)$$

$$\begin{bmatrix} \tilde{C}_{k;i} \\ \tilde{C}_{k;j} \end{bmatrix} \leftarrow \begin{bmatrix} \tilde{C}_{k;i} \\ \tilde{C}_{k;j} \end{bmatrix} \mathbf{G}. \quad (16b)$$

Moreover, the eigenvalues and corresponding eigenvectors can be obtained algebraically at a low computational cost.

Stopping Criterion: Convergence occurs when only negligible changes to the matrices \mathbf{U} and \mathbf{V} take place. This is equivalent to have small rotational angles only, or $|\sin \theta| < \varepsilon$ for all index pairs of \mathbf{U} and \mathbf{V} , with ε some arbitrary small value. The algorithm has been summarized in Algorithm 2.

E. Initialization

From Eq. (4), we have

Diagonality of the products in Eq. (17) is a necessary, albeit not sufficient condition. Thus, for both algorithms we can initialize \mathbf{U} and \mathbf{V} as the matrices whose columns are the eigenvectors of $\sum_k C_k C_k^T$ and $\sum_k C_k^T C_k$, respectively. A numerically preferable, albeit equivalent procedure is to initialize \mathbf{U} as the left singular vectors of $[C_1 \dots C_k \dots C_K]$ and \mathbf{V} as the left singular vectors of $[C_1^T \dots C_k^T \dots C_K^T]$. Note that, if $\sigma^2 = 0$, this actually solves the AJSVD problem. However, as σ^2 increases, this initialization will progressively be more distant from the solution, thus it is worthy only for data with relatively high signal to noise ratios. In general, AJD algorithms converge regardless the initialization and the two proposed algorithms are no exception. The initialization proposed here is thus only useful to speed up convergence.

Algorithm 2 AJSVD with Givens rotations

Initialize \mathbf{U} and \mathbf{V} (if no information available let them be the identity matrices in $\mathbb{R}^{P \times P}$ and $\mathbb{R}^{Q \times Q}$)
while no convergence **do**
 for i from 1 to N ; j from $i + 1$ to P **do**
 if $j \leq N$ **then**
 perform update according to (15) with \mathbf{G} obtained from Eq. (11)
 else
 perform update according to (15) with \mathbf{G} obtained from Eq. (12)
 end if
 end for
 for i from 1 to N ; j from $i + 1$ to Q **do**
 if $j \leq N$ **then**
 perform update according to (16) with \mathbf{G} obtained from Eq. (13)
 else
 perform update according to (16) with \mathbf{G} obtained from Eq. (14)
 end if
 end for
end while

V. RESULTS

To test AJSVD we run the algorithm on simulated data and on a large database comprising 84 electroencephalographic recordings.

Since C is based on the common structures \mathbf{U} and \mathbf{V} , we can evaluate the algorithm's performance by taking the performance criterion

$$\Phi_{\mathbf{U}}(\mathbf{U}^{[r]}) = PI\left[\left(\tilde{\mathbf{U}}^{[r]}\right)^T \tilde{\mathbf{U}}\right] \quad (18a)$$

$$\Phi_{\mathbf{V}}(\mathbf{V}^{[r]}) = PI\left[\left(\tilde{\mathbf{V}}^{[r]}\right)^T \tilde{\mathbf{V}}\right] \quad (18b)$$

where $\tilde{\mathbf{U}}$, $\tilde{\mathbf{V}}$, $\tilde{\mathbf{U}}^{[r]}$ and $\tilde{\mathbf{V}}^{[r]}$ contain only N columns of \mathbf{U} , \mathbf{V} , $\mathbf{U}^{[r]}$ and $\mathbf{V}^{[r]}$ respectively and $PI(\mathbf{H})$ is Moreau's performance index [38] defined as

$$PI(\mathbf{H}) = \frac{1}{2(N-1)} \left(\sum_{\alpha} \left[\frac{\sum_{\beta} |h_{\alpha\beta}|}{\max_{\delta} |h_{\alpha\delta}|} - 1 \right] + \sum_{\gamma} \left[\frac{\sum_{\eta} |h_{\eta\gamma}|}{\max_{\kappa} |h_{\kappa\gamma}|} - 1 \right] \right) \quad (19)$$

When the ground truth is not available, as is the case when we run the algorithm on real data with unknown latent structure, we use the following performance criterion:

$$\Phi_{\text{Diag}}(\mathbf{U}^{[r]}, \mathbf{V}^{[r]}) = \frac{1}{K(M-1)} \sum_k \frac{\left\| \text{Off} \left[\left(\tilde{\mathbf{U}}^{[r]} \right)^T \mathbf{C}_k \tilde{\mathbf{V}}^{[r]} \right] \right\|^2}{\left\| \text{Diag} \left[\left(\tilde{\mathbf{U}}^{[r]} \right)^T \mathbf{C}_k \tilde{\mathbf{V}}^{[r]} \right] \right\|^2} \quad (20)$$

where $M = \max(P, Q)$. For a given noise variance and singular value matrices $\{\mathbf{\Lambda}_k\}$, the expected performance ($K \rightarrow +\infty$) is

given by

$$\Phi_{\text{Diag}}(\mathbf{U}, \mathbf{V}) = \frac{\sigma^2}{\frac{1}{N} \sum_n E\{\lambda_n^2(\mathbf{C}_k)\} + \sigma^2} \quad (21)$$

A. Simulated Data

For a first evaluation of the two proposed algorithms for the approximate diagonalization of a set of asymmetric rectangular matrices, we generate a set C of K input matrices in $\mathbb{R}^{12 \times 16}$. Each input matrix is generated as $\mathbf{C}_k = \mathbf{U}\mathbf{\Lambda}_k\mathbf{V}^T + \mathbf{N}_k$, where the diagonal elements of $\mathbf{\Lambda}_k$ are i.i.d. following a normal distribution with mean zero and standard deviation one and the entries of \mathbf{N}_k are i.i.d. following a normal distribution with mean zero and standard deviation $\sigma = 0.1$. Furthermore, \mathbf{U} and \mathbf{V} are randomly generated orthonormal matrices in $\mathbb{R}^{P \times P}$ and $\mathbb{R}^{Q \times Q}$, respectively, obtained through a QR-decomposition of random matrices with i.i.d. elements drawn from the normal distribution $\mathcal{N}(0, 1)$. The main objective of this first experiment will be to show the accuracy of the theoretical performance estimation using Eq. (21) when the number of matrices $K \rightarrow +\infty$. For comparison we choose $K = 50$ and $K = 250$.

The theoretically expected ($K \rightarrow +\infty$) performance $\Phi_{\text{Diag}}(\mathbf{U}, \mathbf{V})$ in the above case with $E\{\lambda_n^2\} = 1$ is simply given by

$$\Phi_{\text{Diag}}(\mathbf{U}, \mathbf{V}) = \frac{\sigma^2}{1 + \sigma^2} \quad (22)$$

Fig. 1(a)&(b) show the results for both the power method followed by Lödwin's orthogonal correction method and the Jacobi method using Givens rotations. Both the 100 individual realisations and their average are shown.

The results in Fig. 1 confirm the accuracy of the theoretical expectation of the performance measure in Eq. (21). It can also be seen that the Jacobi iterations have a better performance than the power iterations with respect to the off-diagonal criterion for this settings.

A second approach consists in the evaluation of the performance indices in Eq. (18). To this goal we realize 100 runs for i.i.d. $\lambda_n(\mathbf{C}_k)$ taken from a normal distribution having zero mean and unit variance, under varying levels of i.i.d. zero mean, normally distributed noise with varying variance and for a varying number of matrices K ($P = 12, Q = 16$). Since we do not want to evaluate the influence of the initialization stage, both \mathbf{U} and \mathbf{V} were initialized as the identity matrices in their respective dimensions. The results can be found in Fig. 2 and Table I. A stopping criterion of 200 iterations has been used here, such that the algorithms could be compared by their asymptotic convergence.

From both Fig. 1 and 2 we observe that, despite the difference of the two algorithms with respect to the off-diagonal criterion (see Fig. 1), the identification of the principal N columns of \mathbf{U} and \mathbf{V} performs almost equally well. We have been able to show the global convergence of the power iteration in the neighbourhood of the solution, while this is not the case for the Jacobi iterations. The obtained results witness in favour of the good performance of Jacobi iterations, even in the rectangular case.

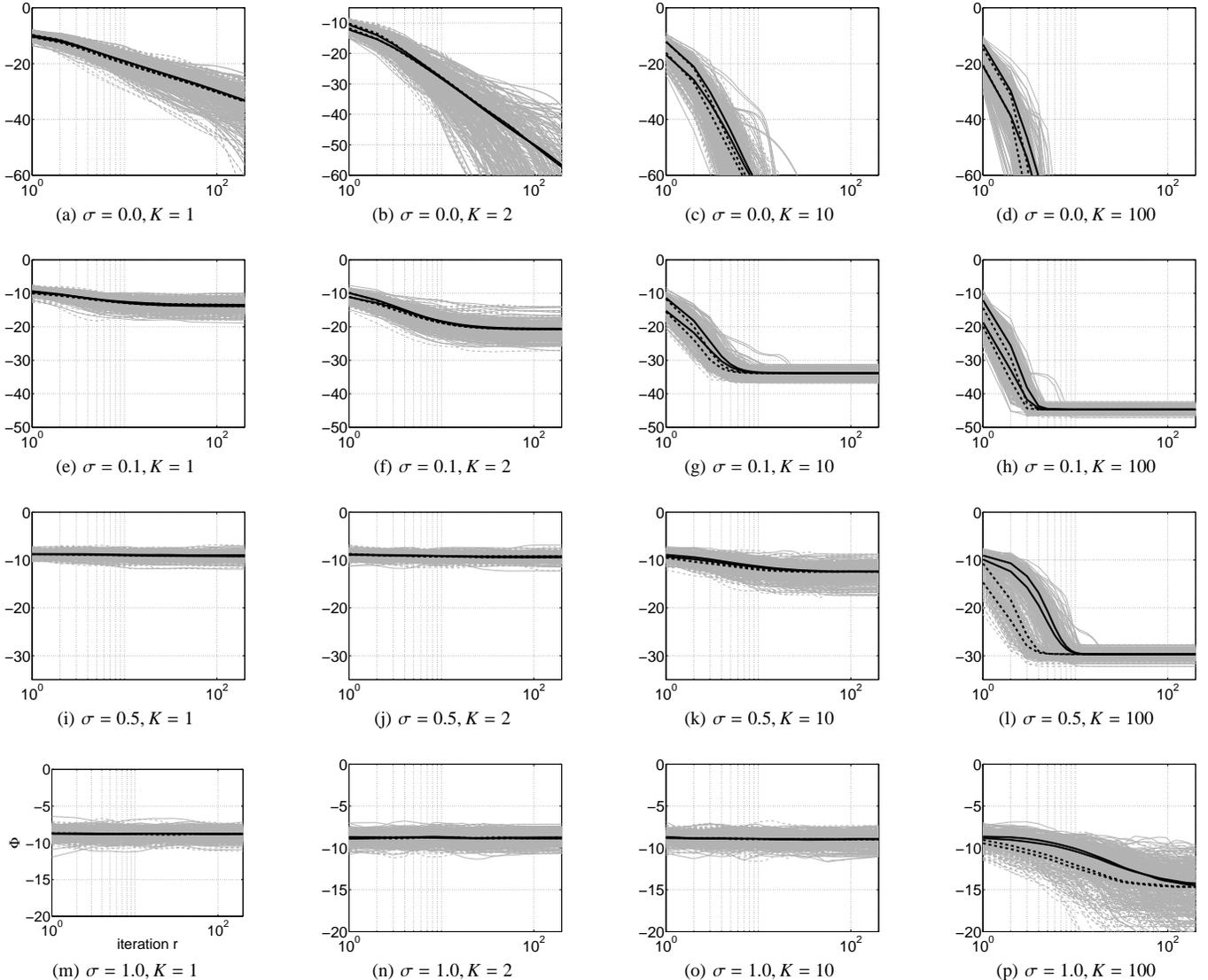


Fig. 2: Moreau-Amari’s performance index [Eq. (18)] for the estimated linear transformations. From left to right we have $K = 1, 2, 10, 100$ matrices and top-down we have a noise level with standard deviation $\sigma = 0, 0.1, 0.5$ and 1 . The continuous lines correspond to the power method and the discontinuous lines to the Given’s rotations. X-axes range from iteration 0 to iteration 200. Equal scaling of the Y-axis (in dB) in each row has been used for ease of comparison.

B. EEG Database

We use the NTE (Nova Tech EEG, Inc.) normative database comprising 84 EEG recordings obtained on healthy individuals aged in between 17 and 35 years old at rest with the eyes closed. The data was acquired from 19 electrodes placed according to the standard international 10/20 system and sampled at 128 samples per second. After artifact rejection, about 90 seconds of EEG was available for each recording. We apply the blind source separation procedure described in [26]. Fourier cospectral matrices at 0.5 Hz resolution were estimated by FFT on 2-sec windows with 50% overlap using the Welch tapering [39]. The cospectra in between 1 and 28 Hz were summed and the sum whitened. The whitening transform was then applied to each of the cospectra in the 1-28 Hz range and the set was subsequently subjected to the two proposed

algorithms. The average (with standard deviation bars) non-diagonality [Eq. 20] and number of iterations applying five different stopping criteria are displayed in Fig. 3. The non-diagonality function reaches its minimum using as stop criterion 10^{-6} for Lödwin and 10^{-3} for Jacobi. For those values the average number of iterations is respectively 198 and 260 and their standard deviations are comparable. For this real dataset, it appears that Algorithm 1 converges faster (in terms of number of iterations).

VI. DISCUSSION

The algorithms we propose are simple to implement, allowing quick testing of new BSS approaches in the community. Their numerical complexity per iteration is $O(42KN^3)$ for the power iterations and $O(6KMN^2)$ for the Jacobi iterations,

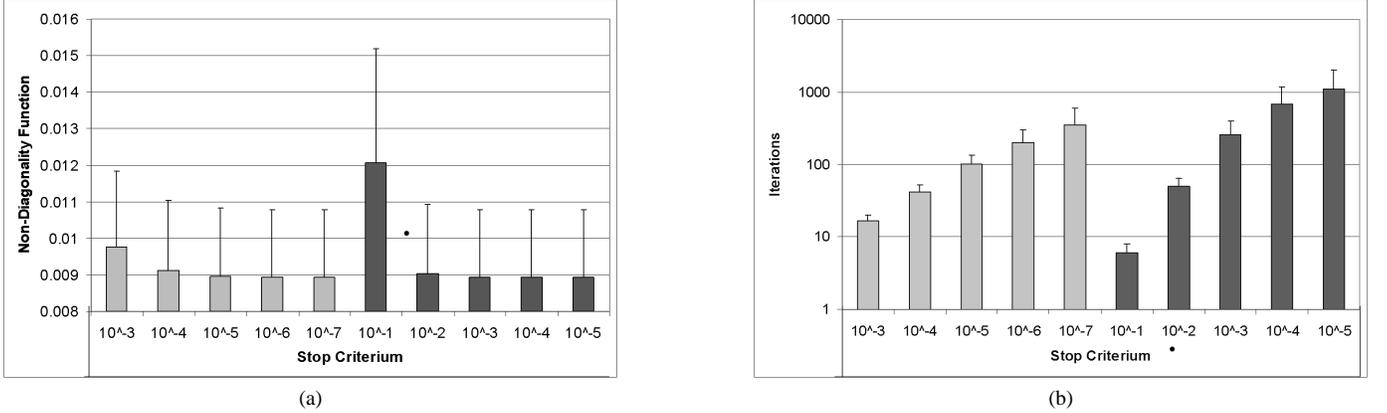


Fig. 3: (a) The non-diagonality measured through Eq. (22) for solving the AJSVD by power iterations with Lödwin orthogonalization (light grey bars) and by Givens rotations (dark grey bars) as obtained on the Nova Tech EEG normative database. Each bar represents the average (on a total of 84) and standard deviation obtained with the stopping criterion set to 10^{-3} , 10^{-4} , 10^{-5} , 10^{-6} and 10^{-7} for the power iterations and 10^{-1} , 10^{-2} , 10^{-3} , 10^{-4} and 10^{-5} for the Givens Rotations. (b) The corresponding average and standard deviation of the number of iterations needed for convergence.

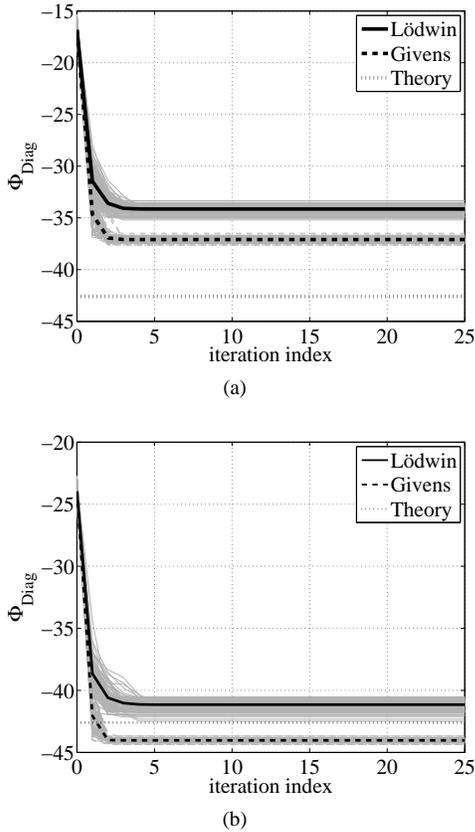


Fig. 1: The performance index of Eq. (20) as a function of the iteration index for the data as described in section V-A. The theoretical expected value of Eq. (22) is indicated with the dotted line. In gray, the 100 individual realisations are plotted and in black the mean over these realisations. The continuous lines are used for the power iteration method (AJSVD Algorithm 1) and the discontinuous lines for the Jacobi method (AJSVD Algorithm 2).

TABLE I: The average of the performance index [Eq. (18a)] (in dB) over 100 Monte Carlo runs as a function of the noise level and the number of matrices using (a) Lödwin symmetric orthogonalization and (b) Jacobi iterations.

(a) Lödwin						
$\sigma \backslash K$	1	2	5	10	50	100
0.00	-11.52	-18.99	-32.52	-55.27	-157.27	-157.38
0.10	-6.78	-10.22	-14.85	-16.98	-20.90	-22.38
0.20	-5.23	-6.53	-11.02	-13.71	-17.78	-19.38
0.30	-4.80	-5.35	-7.81	-11.16	-15.93	-17.38
0.40	-4.52	-4.89	-5.66	-8.14	-14.37	-15.99
0.50	-4.54	-4.62	-5.08	-6.14	-13.14	-14.87
0.60	-4.41	-4.53	-4.74	-5.19	-11.44	-13.78
0.70	-4.42	-4.45	-4.63	-4.82	-8.55	-12.41
0.80	-4.35	-4.40	-4.54	-4.66	-6.24	-10.07
0.90	-4.35	-4.46	-4.46	-4.60	-5.39	-7.33
1.00	-4.34	-4.35	-4.38	-4.49	-4.96	-5.90

(b) Jacobi						
$\sigma \backslash K$	1	2	5	10	50	100
0.00	-11.78	-18.75	-35.76	-59.59	-158.29	-158.46
0.10	-6.69	-10.22	-14.89	-16.98	-20.90	-22.38
0.20	-5.30	-6.45	-11.07	-13.70	-17.78	-19.38
0.30	-4.82	-5.38	-7.88	-11.20	-15.93	-17.38
0.40	-4.52	-4.87	-5.73	-8.29	-14.37	-15.99
0.50	-4.48	-4.63	-5.04	-6.17	-13.14	-14.87
0.60	-4.39	-4.54	-4.74	-5.20	-11.78	-13.79
0.70	-4.35	-4.45	-4.63	-4.83	-9.75	-12.69
0.80	-4.33	-4.43	-4.51	-4.62	-6.88	-11.54
0.90	-4.33	-4.44	-4.44	-4.57	-5.64	-8.97
1.00	-4.37	-4.32	-4.46	-4.45	-5.01	-6.47

which is in line with existing AJD algorithms. However, the convergence rate of our algorithm is rather low, thus extensions to the bi-linear case of optimization schemes may be investigated in the future, as well as non-orthogonal solutions to the AJSVD problem, avoiding the whitening step [9], [14], [19].

The appropriate stopping criterion for AJD algorithms with linear convergence depends on the data at hand. The diag-

onalization of the large EEG normative database we have performed (Fig. 3) provides an indication, but this is a mere empirical approach. Future research could focus on an adaptive procedure to set the optimal stopping criterion, balancing between computational resources and the required accuracy of the offered solution.

The AJSVD algorithm 1 based on power iterations can easily be extended to the complex domain; it suffices to define the Frobenius norm of a matrix $\mathbf{C} \in \mathbb{C}^{P \times Q}$ as $\|\mathbf{C}\|^2 = \sum_p \sum_q |\mathbf{C}_{pq}|^2 = \text{tr} \mathbf{C}^H \mathbf{C}$, where H denotes the conjugate transpose. It is then easily seen that the equations in section IV transform accordingly and that we are left with the maximization of

$$\mathcal{J}^{OFF}(\mathbf{U}, \mathbf{V} | \mathbf{C}) = \sum_{k=1}^K \text{tr} |\text{Diag}_N(\mathbf{U}^H \mathbf{C}_k \mathbf{V})|^2 = \sum_{k=1}^K \sum_{i=1}^N |\mathbf{u}_i^H \mathbf{C}_k \mathbf{v}_i|^2$$

which is equivalent to the maximization by parts of

$$\begin{aligned} \mathcal{J}^{\text{Diag}}(\mathbf{U} | \mathbf{V}, \mathbf{C}) &= \sum_n \mathbf{u}_n^H \mathcal{M}_n(\mathbf{V}) \mathbf{u}_n \\ \mathcal{J}^{\text{Diag}}(\mathbf{V} | \mathbf{U}, \mathbf{C}) &= \sum_n \mathbf{v}_n^H \mathcal{M}_n(\mathbf{U}) \mathbf{v}_n . \end{aligned}$$

Now, since the above equations contain the Hermitian positive semidefinite matrices $\mathcal{M}_n(\mathbf{V})$ and $\mathcal{M}_n(\mathbf{U})$, their eigenvalues are real positive and the power iteration methods can be used as defined for the real case. In contrast, the Jacobi iterations need to be adapted appropriately to the complex case. To this end, we propose the use of the algorithm for complex Givens rotations as has been described in [34].

VII. CONCLUSION

The algorithms we propose are simple and numerically efficient. Numerical simulations show good convergence properties using either the power method combined with Lödwin symmetric orthonormalization or Jacobi iterations. In addition, we have shown heuristically that both the power method and the Jacobi iterations have global convergent behaviour in the bilinear case, with an asymptotic convergence close to the theoretical value. The analysis of a large EEG database also shows good performance. The new framework opens the road to new blind source separation applications.

ACKNOWLEDGEMENTS

This research has been partially supported by ANR (Agence National de Recherche) projects GAZE&EEG, OpenVibe2 & RoBiK. The electroencephalographic data for analysis was kindly provided by Nova Tech EEG, Inc., MESA, AZ.

REFERENCES

- [1] L. Parra and P. Sadja, "Blind source separation via generalized eigenvalue decomposition," *Journal of Machine Learning Research*, vol. 4, pp. 1261–1269, 2003.
- [2] L. Tong, R.-W. Liu, V. C. Soon, and Y.-F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans on Circuits and Systems*, vol. 38, no. 5, pp. 499–509, 1991.
- [3] A. Belouchrani, K. Abed-Meraim, J. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans on Sign Proc*, vol. 45, no. 2, pp. 434–444, 1997.

- [4] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc.-F*, vol. 140, no. 6, pp. 362–370, 1993.
- [5] E. Moreau, "A generalization of joint-diagonalization criteria for source separation," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 530–541, March 2001.
- [6] F. J. Theis and Y. Inouye, "On the use of joint diagonalization in blind signal processing," in *Proc. IEEE International Symposium on Circuits and Systems ISCAS 2006*, Kos, Greece, 2006, pp. 3589–3592.
- [7] D. T. Pham, "Joint approximate diagonalization of positive definite matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 22, no. 4, pp. 1136–1152, 2000.
- [8] R. Vollgraf and K. Obermayer, "Quadratic optimization for simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3270–3278, Sept. 2006.
- [9] P. Tichavský and A. Yeredor, "Fast approximate joint diagonalization incorporating weight matrices," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 878–891, March 2009.
- [10] E. M. Fadaili, N. T. Moreau, and E. Moreau, "Nonorthogonal joint diagonalization/zero diagonalization for source separation based on time-frequency distributions," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1673–1687, May 2007.
- [11] X.-L. Li and X.-D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1803–1814, May 2007.
- [12] F. Wang, Z. Liu, and J. Zhang, "Nonorthogonal joint diagonalization algorithm based on trigonometric parameterization," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5299–5308, Nov. 2007.
- [13] G. Zhou, Z. Yang, Z. Wu, and J. Zhang, "Non-orthogonal joint diagonalization with diagonal constraints," *Progress in Natural Science*, vol. 18, pp. 735–739, 2008.
- [14] A. Souloumiac, "Nonorthogonal joint diagonalization by combining givens and hyperbolic rotations," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2222–2231, June 2009.
- [15] M. Congedo and D.-T. Pham, "Least-squares joint diagonalization of a matrix set by a congruence transformation," in *Singaporean-French IPAL Symposium (SinFra'09)*, 2009.
- [16] D.-T. Pham, , and M. Congedo, "Least square joint diagonalization of matrices under an intrinsic scale constraint," in *8th International Conference on Independent Component Analysis and Signal Separation (ICA 2008)*, 2008, pp. 298–305.
- [17] S. Degerine and E. Kane, "A comparative study of approximate joint diagonalization algorithms for blind source separation in presence of additive noise," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3022–3031, June 2007.
- [18] R. Iferroudjene, K. Abed-Meraim, and A. Belouchrani, "A new jabobi-like method for joint diagonalization of arbitrary non-defective matrices," *Applied Mathematics and Computation*, vol. 211, pp. 363–373, 2009.
- [19] A. Ziehe, P. Laskov, G. Nolte, and K.-R. Müller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *Journal of Machine Learning Research*, vol. 5, pp. 777–800, 2004.
- [20] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation. Independent Component Analysis and Applications*. Oxford: Academic Press, 2010.
- [21] M. Wax and J. Sheinvald, "A least-squares approach to joint diagonalization," *IEEE Signal Process. Lett.*, vol. 4, no. 2, pp. 52–53, Feb. 1997.
- [22] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, 2005th ed. Wiley, 2002.
- [23] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley Interscience, 2001.
- [24] B. Pesquet-Popescu, J.-C. Pesquet, and A. P. Petropulu, "Joint singular value decomposition - a new tool for separable representation of images," in *Proc. Int Image Processing Conf*, vol. 2, 2001, pp. 569–572.
- [25] G. Hori, "Comparison of two main approaches to joint svd," in *Independent Component Analysis and Signal Separation*, ser. Lecture Notes in Computer Science, T. Adali, C. Jutten, J. a. Romano, and A. Barros, Eds. Springer Berlin / Heidelberg, 2009, vol. 5441, pp. 42–49.
- [26] M. Congedo, C. Gouy-Pailler, and C. Jutten, "On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics," *Clinical Neurophysiology*, vol. 119, no. 12, pp. 2677–2686, 2008.
- [27] J.-F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, 1996.

- [28] B. Afsari, "Sensitivity analysis for the problem of matrix joint diagonalization," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1148–1171, 2008.
- [29] A. Aïssa-El-Bey, K. Abed-Meraim, Y. Grenier, and Y. Hua, "A general framework for second-order blind separation of stationary colored sources," *Signal Processing*, vol. 88, pp. 2123–2137, 2008.
- [30] F. Theis, "Uniqueness of complex and multidimensional independent component analysis," *Signal Processing*, vol. 84, pp. 951–956, 2004.
- [31] N. Correa, T. Adali, Y. Li, and V. Calhoun, "Canonical correlation analysis for data fusion and group interferences," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 39–50, 2010.
- [32] J. D. Carroll, "Generalization of canonical correlation analysis to three or more sets of variables," in *Proceedings of the 76th Annual Convention APA*, 1968, pp. 227–228.
- [33] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, pp. 211–218, 1936.
- [34] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM Journal on Matrix Analysis and Applications*, vol. 17, no. 1, pp. 161–164, 1996.
- [35] L. De Lathauwer and P. Comon, "Higher-order power method," in *Nonlinear Theory and its Applications, NOLTA'95*, vol. 1, Las Vegas, 1995, pp. 91–96.
- [36] B. C. Carlson and J. M. Keller, "Orthogonalization procedures and the localization of Wannier functions," *Physical Review*, vol. 105, no. 1, pp. 102–103, 1957.
- [37] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, 1996.
- [38] O. Macchi and E. Moreau, "Self-adaptive source separation by direct and recursive networks," in *Proc. International Conference on Digital Signal Processing (DSP93)*, 1993, pp. 1154–1159.
- [39] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–74, 1967.



Marco Congedo was born in Bari, Italy, in 1972. In 2003 he obtained the Ph.D. degree in Biological Psychology with a minor in Statistics from the University of Tennessee, Knoxville. From 2003 to 2006 he has been a post-doc fellow at the French National Institute for Research in Informatics and Control (INRIA) and at France Telecom R&D, in France. Since 2007 Dr. Congedo is a Research Scientist at the Centre National de la Recherche Scientifique (CNRS) in the GIPSA Laboratory, Grenoble, France.

Dr. Congedo has been the recipient of several awards, scholarships and research grants. He is interested in basic human electroencephalography (EEG) and magnetoencephalography (MEG), real-time neuroimaging (neurofeedback and brain computer interface), EEG hyperscanning and multivariate statistical tools useful for EEG and MEG such as inverse solutions and blind source separation.

Dr. Congedo is a Fellow of the International Society for Neurofeedback and Research and a Consulting Editor for the *Journal of Neurotherapy*.



Ronald Phlypo (S06M10) was born in Ostend, Belgium, in 1981. He graduated in industrial engineering from KHBO, Ostend, in 2003 and received the masters degree in artificial intelligence from KULeuven, Leuven, Belgium, in 2004. He obtained the Ph.D. degree in Engineering Sciences from the University of Ghent, Belgium, in 2009.

During 2006–2009, he has been a visitor with the Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis, France.

Currently, he is a Postdoctoral Researcher at the GIPSA laboratories, Grenoble, France. His research interests include (blind) statistical signal and array processing and its application to biomedical signal processing problems.



Dinh-Tuan Pham (M88) was born in Hanoi, Vietnam, on February 10, 1945. He graduated from the Engineering School of Applied Mathematics and Computer Science (ENSIMAG), Polytechnic Institute of Grenoble, Grenoble, France, in 1968. He received the Ph.D. degree in statistics from the University of Grenoble, Grenoble, France, in 1975.

He was a Postdoctoral Fellow at the Department of Statistics, University of California, Berkeley, in 1977/1978 and a Visiting Professor at the Department of Mathematics, Indiana University, Bloomington, in 1979/1980. Currently, he is Director of Research at the French Centre National de Recherche Scientifique (CNRS), Laboratoire Jean Kuntzmann. His research interests include time-series analysis, signal modeling, blind source separation, nonlinear (particle) filtering, and biomedical signal processing.