Orthonormal approximate joint block-diagonalization

Bloc-diagonalisation simultanée approchée avec contrainte orthonormale

Cédric Févotte
Fabian J. Theis

2007D007

2007

Département Traitement du Signal et des Images
Groupe Audio, Acoustique et Ondes

Technical report - Rapport de recherche
GET/Télécom Paris

Orthonormal approximate joint block-diagonalization

Cédric Févotte and Fabian J. Theis

Abstract: The aim of this work is to give a comprehensive overview of the problem of jointly block-diagonalizing a set of matrices. We discuss how to implement methods in the common case of only approximative block-diagonalizability using Jacobi algorithms. Standard Jacobi optimization techniques for diagonalization and joint diagonalization are reviewed first, before we study their generalizations to the block case and give some new theoretical insights on existence and uniqueness issues as well as on the interplay between block and standard diagonalization problems. Simulations on synthetic data show that in the block case convergence to the optimal solution is not always observed in practice and that the behavior of the Jacobi approach is very much dependent on the initialization of the orthonormal basis and also on the choice of the successive rotations.

Bloc-diagonalisation simultanée approchée avec contrainte orthonormale

Cédric Févotte and Fabian J. Theis

Résumé: Ce rapport présente de manière unifiée des techniques de diagonalisation, bloc-diagonalisation (BD), diagonalisation simultanée (DS) et bloc-diagonalisation simultanée (BDS) par méthode de Jacobi. Nos contributions principales concernent le problème de la bloc-diagonalisation simultanée. Les conditions d'existence et d'unicité des solutions sont étudiées. Il apparaît en pratique que la convergence des méthodes de Jacobi vers une solution optimale (minimisant le critère choisi), généralement observée dans le cas de la DS, n'est pas toujours observée pour la BDS, et qu'elle dépend largement de l'initialisation et du choix des rotations successives. A ce titre nous décrivons une nouvelle méthode de sélection des rotations qui maximise d'un point de vue empirique les chances de convergence vers une solution optimale (sans toutefois la garantir).

ORTHONORMAL APPROXIMATE JOINT BLOCK-DIAGONALIZATION

CÉDRIC FÉVOTTE * AND FABIAN J. THEIS †

Abstract. The aim of this work is to give a comprehensive overview of the problem of jointly block-diagonalizing a set of matrices. We discuss how to implement methods in the common case of only approximative block-diagonalizability using Jacobi algorithms. Standard Jacobi optimization techniques for diagonalization and joint diagonalization are reviewed first, before we study their generalizations to the block case and give some new theoretical insights on existence and uniqueness issues as well as on the interplay between block and standard diagonalization problems. Simulations on synthetic data show that in the block case convergence to the optimal solution is not always observed in practice and that the behavior of the Jacobi approach is very much dependent on the initialization of the orthonormal basis and also on the choice of the successive rotations.

Key words. simultaneous unitary diagonalization, Jacobi optimization, matrix factorization

1. Introduction.

Joint diagonalization techniques have received much attention in the last fifteen years within the field of signal processing, and more specifically within the fields of Independent Component Analysis (ICA) and Blind Source Separation (BSS). JADE, a standard ICA algorithm developed by Cardoso and Souloumiac [9], is based on joint diagonalization (JD) of a set of cumulant matrices. To this purpose the authors designed a Jacobi algorithm for approximate joint diagonalization of a set of matrices [10]. In a BSS parlance, JADE allows for separation of determined linear instantaneous mixtures of mutually independent sources, exploiting fourth-order statistics. Other standard BSS techniques involving joint diagonalization include the SOBI algorithm [2], TDSEP [18] and TFBSS [12], which all rely on second-order statistics of the sources, namely covariance matrices in the first and second case and spatial Wigner-Ville spectra in the third case.

Joint block-diagonalization (JBD) came into play in BSS when Abed-Meraim, Belouchrani and co-authors extended the SOBI algorithm to overdetermined convolutive mixtures [5]. Their idea was to turn the convolutive mixture into an overdetermined linear instantaneous mixture of block dependent sources, the second-order statistics matrices of the source vector thus becoming block-diagonal instead of diagonal. Hence, the joint-diagonalization step in SOBI needs to be replaced by a joint block diagonalization step. Another area of application can be found in the context of multidimensional or group ICA [8, 14]. Its goal is to linearly transform an observed multivariate random vector such that its image is decomposed into groups of stochastically independent vectors. It has been shown that by using fourth-order cumulants to measure the independence, JADE now translates into a JBD problem [17]; similarly also SOBI and other JD-based criteria can be extended to this group ICA setting [11, 16].

Abed-Meraim et al. have sketched several Jacobi strategies in [1, 3, 4]: the JBD problem is turned into a minimization problem, where the matrix parameter (the joint block-diagonalizer) is constrained to be unitary (because of spatial prewhitening). The minimizer is searched for iteratively, as a product of Givens rotations, each rotation minimizing a block-diagonality criterion around a fixed axis.

*GET/Télécom Paris (ENST), 37-39 rue Dareau, 75014 Paris, France (`fevotte@tsi.enst.fr`).

†Bernstein Center for Computational Neuroscience, MPI for Dynamics and Self-Organisation, Göttingen, Germany, (`fabian@theis.name`), partially supported by the DFG (grant GRK 638).

Convergence of the algorithm is easily shown (as seen in the following), but convergence to an optimal solution (which minimizes the chosen JBD criterion) is not guaranteed. In fact, we observed that results vary widely according to the choice of the successive rotations and the initialization of the algorithm, which is not discussed in previous works [1, 3, 4].

In this paper, we propose a review of Jacobi methods for approximate orthonormal diagonalization, joint diagonalization, block-diagonalization and joint block-diagonalization of matrices. Moreover, novel contributions are brought to the block case. In particular, we point out that the choice of rotations is a sensitive issue which greatly influences the convergence properties of the Jacobi algorithm, as illustrated on extensive simulations with synthetic data. We propose a new method for choosing the rotations, which, from empirical testing, offers better chances to converge to the optimal solution (while still not guaranteeing it), as compared to the standard cyclic Jacobi technique. We also point out the interest of initializing JBD with the output of JD, coroborating the idea that JD could in fact perform JBD up to permutations, as suggested by Cardoso in [8], more recently conjectured by Abed-Meraim and Belouchrani in [1], and partially proved in this paper. Existence and uniqueness conditions for the solutions of the JBD problem are also investigated.

This manuscript is organized as follows. Section 2 reviews 'historical' approaches for Jacobi diagonalization of a matrix. Section 3 reviews how the Jacobi approach to diagonalization was extended to the joint diagonalization problem by Cardoso and Souloumiac. Section 4 considers the extension of the Jacobi approach to the block-diagonalization of one matrix, and discusses related existence and uniqueness issues of the underlying block-diagonalization problem. Section 5 describes the generalization to the joint block-diagonalization problem, and gives some theoretical insight in the structure of its solutions. This section in particular reports and discusses various results of the proposed JBD algorithms, and theoretically connects them with the simpler JD problem. Section 6 gives brief conclusions.

2. Approximate diagonalization.

We first recall results on the diagonalization of complex and real matrices, and specifically review approximative solutions based on Jacobi algorithms.

2.1. Complex orthonormal basis.

Two standard Jacobi methods for the approximate diagonalization of a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ in a complex orthonormal basis are presented in the following. Further details can be found in [13] and in the references therein. Some notations and derivations are also inspired from [10]. If $\mathbf{A}$ is diagonalizable in an orthonormal basis, then the algorithms we describe essentially compute this basis, see next section. If $\mathbf{A}$ is not diagonalizable in an orthonormal basis, then the following algorithms yield an orthonormal basis in which $\mathbf{A}$ is the *most diagonal*, in a quadratic sense. In other words, we look for an orthonormal matrix $\mathbf{U} \in \mathbb{C}^{n\times n}$ such that

$$\mathbf{U}\mathbf{A}\mathbf{U}^H = \mathbf{D}$$

where $\mathbf{D} \in \mathbb{C}^{n\times n}$ minimizes the deviation from zero of the off-diagonal elements, i.e. the following diagonality criterion

$$\mathrm{off}\,(\mathbf{A}) := \sum_{1\leq i\neq j\leq n} |m_{ij}|^2 = \|\mathbf{A}\|_F - \sum_{1\leq i\leq n} |m_{ii}|^2, \tag{2.1}$$

where $\|\mathbf{A}\|_F$ denotes the Frobenius norm of the matrix $\mathbf{A}$. In other words, we look for $\mathbf{U}$ via minimization of the following criterion

$$C_{\mathrm{d}}(\mathbf{V};\mathbf{A}) := \mathrm{off}\left(\mathbf{V}\,\mathbf{A}\,\mathbf{V}^H\right) \tag{2.2}$$

with respect to the unitary matrix $\mathbf{V} \in U(n)$, i.e. $\mathbf{V} \in \mathbb{C}^{n\times n}$ with $\mathbf{V}\,\mathbf{V}^H = \mathbf{I}_n$.

2.1.1. Existence and uniqueness. The question whether a solution of the *approximate* diagonalization problem exists can be easily answered in the affirmative, however it is not clear how many minima of (2.1) exist. We answer this only in the limit case of perfect factorization, where off $(\mathbf{V}) = 0$: it is well known that a diagonalizer can be found if and only if $\mathbf{A}$ is normal i.e. $\mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A}$. Hence a sufficient condition for such a diagonalization to exist is that $\mathbf{A}$ is hermitian, cf. Sec. 2.1.6.

The number of solutions i.e. diagonalizing matrices $\mathbf{V}$ depends on the eigenvalue distribution. In the case of $\mathbf{A}$ having pairwise different eigenvalues, which we will denote by unispectral $\mathbf{A}$ in the following, all eigenspaces of $\mathbf{A}$ are of dimension one. But by construction, the rows of $\mathbf{V}$ consist of a basis of unit-length eigenvectors of $\mathbf{A}$, so $\mathbf{V}$ is unique except for permutations of the rows. Moreover the eigenvectors may be multiplied by a unit scalar, so $\mathbf{V}$ is unique except for left multiplication by a permutation matrix and a diagonal unitary matrix. In the case of larger-dimensional eigenspaces, additional indeterminacies arise, as non-trivial linear combinations from eigenvectors of a single eigenspace again produce a valid diagonalizer $\mathbf{M}'$. In many applications, $\mathbf{V}$ has to be determined from an observed $\mathbf{A}$, so this case may be avoided by adding additional matrices with possibly different eigenstructure into the diagonalization criterion in the form of joint diagonalization, see Sec. 3.

2.1.2. The Jacobi idea. Jacobi methods rely on the fact that any unitary matrix $\mathbf{V} \in U(n)$ can be written as the product of complex Givens (rotation) matrices $\mathbf{G}(p,q,c,s) \in U(n)$, defined for $p < q$ by

$$\mathbf{G}(p,q,c,s) = \begin{array}{cc} \begin{bmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \dots & c & \dots & \bar{s} & \dots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & -s & \dots & c & \dots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix} & \begin{array}{c} \\ \\ p \\ \\ q \\ \\ \\ \end{array} \\ \begin{array}{ccccccc} & & p & & q & & \end{array} & \end{array}$$

with $(c,s) \in \mathbb{R}\times\mathbb{C}$ such that $c^2 + |s|^2 = 1$. The Jacobi idea consists of successively applying Givens rotations to $\mathbf{A}$ in order to minimize criterion (2.2). For fixed p and q, one iteration of the method consists of the following two steps:

- select (c,s) such that $C_{\mathrm{d}}(\mathbf{G}(p,q,c,s);\mathbf{A})$ is minimal
- $\mathbf{A} \leftarrow \mathbf{G}(p,q,c,s)\,\mathbf{A}\,\mathbf{G}(p,q,c,s)^H$

An interesting aspect of this method is that the minimization step can be done algebraically.

2.1.3. Method. Let us check the effect of a Givens rotation on $\mathbf{A}$. For fixed p and q with $p < q$, we note that $\mathbf{B} = \mathbf{G}(p,q,c,s)\,\mathbf{A}\,\mathbf{G}(p,q,c,s)^H$. Simple calculations show that $\mathbf{B}$ equals $\mathbf{A}$ everywhere except on the p-th and q-th rows and columns, so

$$\mathbf{B} =$$

$$\begin{array}{cc} \left[\begin{array}{c|c|c|c|c} a_{ij} & c\,a_{ip}+s\,a_{iq} & a_{ij} & -\bar{s}\,a_{ip}+c\,a_{iq} & a_{ij} \\ \hline c\,a_{pj}+\bar{s}\,a_{qj} & \begin{array}{c} c^2\,a_{pp}+|s|^2\,a_{qq} \\ +c\,s\,a_{pq}+c\,\bar{s}\,a_{qp} \end{array} & c\,a_{pj}+\bar{s}\,a_{qj} & \begin{array}{c} c^2\,a_{pq}-\bar{s}^2\,a_{qp} \\ +c\,\bar{s}\,(a_{qq}-a_{pp}) \end{array} & c\,a_{pj}+\bar{s}\,a_{qj} \\ \hline a_{ij} & c\,a_{ip}+s\,a_{iq} & a_{ij} & -\bar{s}\,a_{ip}+c\,a_{iq} & a_{ij} \\ \hline -s\,a_{pj}+c\,a_{qj} & \begin{array}{c} c^2\,a_{qp}-s^2\,a_{pq} \\ +c\,s\,(a_{qq}-a_{pp}) \end{array} & -s\,a_{pj}+c\,a_{qj} & \begin{array}{c} c^2\,a_{qq}+|s|^2\,a_{pp} \\ -c\,s\,a_{pq}-c\,\bar{s}\,a_{qp} \end{array} & -s\,a_{pj}+c\,a_{qj} \\ \hline a_{ij} & c\,a_{ip}+s\,a_{iq} & a_{ij} & -\bar{s}\,a_{ip}+c\,a_{iq} & a_{ij} \end{array}\right] & \begin{array}{c} \\ p \\ \\ \\ q \\ \\ \\ \end{array} \\ \begin{array}{ccccc} \qquad & \qquad p \qquad & \qquad & \qquad q \qquad & \qquad \end{array} & \end{array} \tag{2.3}$$

Let us look for c and s that minimize $C_{\mathrm{d}}(\mathbf{G}(p,q,c,s);\mathbf{A})$. We have:

$$C_{\mathrm{d}}(\mathbf{G}(p,q,c,s);\mathbf{A}) = \mathrm{off}\,(\mathbf{B}) = \sum_{1\le i\ne j\le n} |b_{ij}|^2 = \|\mathbf{B}\|_F^2 - \sum_{i=1}^{n} |b_{ii}|^2$$

Invariance of the Frobenius norm under rotation guarantees $\|\mathbf{B}\|_F = \|\mathbf{A}\|_F$. Moreover, the diagonal terms of $\mathbf{B}$ are equal to the diagonal terms of $\mathbf{A}$ except for entries p and q. Hence, according to Eq. (2.1),

$$\begin{aligned} \mathrm{off}\,(\mathbf{B}) &= \|\mathbf{A}\|_F^2 - \left(\sum_{i=1}^{n} |a_{ii}|^2 - |a_{pp}|^2 - |a_{qq}|^2 + |b_{pp}|^2 + |b_{qq}|^2\right) \\ &= \mathrm{off}\,(\mathbf{A}) + |a_{pp}|^2 + |a_{qq}|^2 - |b_{pp}|^2 - |b_{qq}|^2 \end{aligned} \tag{2.4}$$

$\mathbf{A}$ does not depend on c and s, hence the minimization of $C_{\mathrm{d}}(\mathbf{G}(p,q,c,s);\mathbf{A})$ with respect to c and s amounts to the maximization of $|b_{pp}|^2 + |b_{qq}|^2$. Eq. (2.4) can be further expanded. Indeed, using the triangle equality, we get

$$\begin{aligned} |b_{pp}|^2 + |b_{qq}|^2 &= \frac{1}{2}\left(|b_{pp}+b_{qq}|^2 + |b_{pp}-b_{qq}|^2\right) \\ |a_{pp}|^2 + |a_{qq}|^2 &= \frac{1}{2}\left(|a_{pp}+a_{qq}|^2 + |a_{pp}-a_{qq}|^2\right) \end{aligned}$$

Moreover, the trace being invariant under rotation, we have $b_{pp} + b_{qq} = a_{pp} + a_{qq}$. Hence

$$\mathrm{off}\,(\mathbf{B}) = \mathrm{off}\,(\mathbf{A}) + \frac{1}{2}\left(|a_{pp}-a_{qq}|^2 - |b_{pp}-b_{qq}|^2\right) \tag{2.5}$$

and thus

$$\underset{c,s}{\mathrm{minimize}}\; C_{\mathrm{d}}(\mathbf{G}(p,q,c,s);\mathbf{A}) \iff \underset{c,s}{\mathrm{maximize}}\; C'_{\mathrm{d}}(c,s) := |b_{pp}-b_{qq}|^2$$

For the sake of clarity in the notations, we discard the dependence of $C'_{\mathrm{d}}(c,s)$ of p, q and $\mathbf{A}$. Let us now study the maximization of $C'_{\mathrm{d}}(c,s)$. From Eq. (2.3), we get

$$b_{pp} - b_{qq} = (c^2 - |s|^2)(a_{pp}-a_{qq}) + 2\,c\,s\,a_{pq} + 2\,c\,\bar{s}\,a_{qp} \tag{2.6}$$

Defining

$$\begin{aligned} \mathbf{v}(c,s) &:= \left(c^2 - |s|^2, c\,s + c\,\bar{s}, \mathrm{i}\,(c\,s - c\,\bar{s})\right)^T \\ \mathbf{h}(\mathbf{A}) &:= (a_{pp}-a_{qq}, a_{pq}+a_{qp}, \mathrm{i}\,(a_{qp}-a_{pq})) \end{aligned}$$

we have $b_{pp} - b_{qq} = \mathbf{h}(\mathbf{A})\,\mathbf{v}(c,s)$ and hence

$$C'_{\mathrm{d}}(c,s) = \mathbf{v}(c,s)^T\ \mathbf{h}(\mathbf{A})^H\mathbf{h}(\mathbf{A})\ \mathbf{v}(c,s)$$

Notice that $\mathbf{v}(c,s)$ is a *real* vector. Together with $\mathbf{h}(\mathbf{A})^H\mathbf{h}(\mathbf{A})$ being by construction a positive semidefinite matrix, we therefore get

$$C'_{\mathrm{d}}(c,s) = \mathbf{v}(c,s)^T\ \mathbf{Q}_{\mathrm{d}}\ \mathbf{v}(c,s)$$

with the positive semidefinite real matrix $\mathbf{Q}_{\mathrm{d}} = \mathrm{Re}\left\{\mathbf{h}(\mathbf{A})^H\mathbf{h}(\mathbf{A})\right\}$. Hence, the maximization of $C'_{\mathrm{d}}(c,s)$ boils down to the maximization of the non-negative quadratic form $q_{\mathrm{d}}(\mathbf{x}) = \mathbf{x}^T\,\mathbf{Q}_{\mathrm{d}}\,\mathbf{x}$ on the *real* domain $\{\mathbf{v}(c,s)\mid(c,s)\in\mathbb{R}\times\mathbb{C}, c^2+|s|^2=1\}$.

Let us now show that the maximization of $q_{\mathrm{d}}(\mathbf{x})$ on the latter domain is equivalent to the maximization of $q_{\mathrm{d}}(\mathbf{x})$ on the unit sphere, which will in turn amount to compute the eigenvector corresponding to the largest eigenvalue of $\mathbf{Q}_{\mathrm{d}}$.

LEMMA 2.1. *Let* $\mathcal{E} = \{\mathbf{v}(c,s)\mid(c,s)\in\mathbb{R}\times\mathbb{C}, c^2+|s|^2=1\}$ *and let* $\mathcal{S}^2 = \{(x,y,z)^T\in\mathbb{R}^3\mid x^2+y^2+z^2=1\}$ *be the 2-sphere. Then* $\mathcal{E}=\mathcal{S}^2$.

Proof. Let $(u,v,w)^T\in\mathcal{E}$. Let us show that $(u,v,w)^T\in\mathcal{S}^2$. By definition, there exists (c,s) such that $u=c^2-|s|^2, v=c\,s+c\,\bar{s}, w=\mathrm{i}\,(c\,s-c\,\bar{s})$. We have:

$$\begin{aligned}u^2+v^2+w^2 &= (c^2-|s|^2)^2+(c\,s+c\,\bar{s})^2-(c\,s-c\,\bar{s})^2\\ &= (c^2-|s|^2)^2+4\,c^2\,|s|^2=(c^2+|s|^2)^2=1\end{aligned}$$

Hence $(u,v,w)^T\in\mathcal{S}^2$, and thus $\mathcal{E}\subset\mathcal{S}^2$.

Now, let $(u,v,w)^T\in\mathcal{S}^2$. If $u\neq-1$ we define

$$c=\sqrt{\frac{u+1}{2}},\quad s=\frac{v-\mathrm{i}\,w}{\sqrt{2\,(u+1)}}, \tag{2.7}$$

and if $u=-1$ (and thus $v=w=0$) we set $c=0$ and $s=1$. Simple algebra shows that $c^2+|s|^2=1$ and $(u,v,w)^T=\left(c^2-|s|^2, c\,s+c\,\bar{s}, \mathrm{i}\,(c\,s-c\,\bar{s})\right)^T=\mathbf{v}(c,s)$. Hence $(u,v,w)^T\in\mathcal{E}$, and thus $\mathcal{S}^2\subset\mathcal{E}$. □

The maximization of $C'_{\mathrm{d}}(c,s)$ on $\mathcal{E}$ is thus equivalent to the maximization of $q_{\mathrm{d}}(\mathbf{x})$ on $\mathcal{S}^2$. Furthermore, the quadratic form $q_{\mathrm{d}}(\mathbf{x})$ is maximized on the unit sphere $\mathcal{S}^2$ by any unit-norm eigenvector corresponding to the largest eigenvalue of $\mathbf{Q}_{\mathrm{d}}$. Let $(u^\star_{\mathrm{d}}, v^\star_{\mathrm{d}}, w^\star_{\mathrm{d}})^T$ be such an eigenvector, chosen with $u^\star_{\mathrm{d}}\geq 0$. From Eq. (2.7), the values $(c^\star_{\mathrm{d}}, s^\star_{\mathrm{d}})$ which minimize criterion $C_{\mathrm{d}}(\mathbf{G}(p,q,c,s);\mathbf{A})$ with p and q fixed are thus

$$c^\star_{\mathrm{d}}=\sqrt{\frac{u^\star_{\mathrm{d}}+1}{2}},\quad s^\star_{\mathrm{d}}=\frac{v^\star_{\mathrm{d}}-\mathrm{i}\,w^\star_{\mathrm{d}}}{\sqrt{2\,(u^\star_{\mathrm{d}}+1)}}. \tag{2.8}$$

2.1.4. Choice of the rotations. Given p and q we have shown how to find the Givens rotation that guarantees maximum decrease of criterion C_{d}. In this paragraph, we briefly recall two strategies for the choice of p and q.

Classical Jacobi. This algorithm takes its full meaning for diagonalization of a hermitian matrix. It consists of choosing p and q at each iteration such that $|a_{pq}|^2$ is maximum, in order to ensure maximum decrease of C_{d} at each iteration, see Algorithm 1.

Cyclic Jacobi. The algorithm consists of methodically sweeping all the rows of $\mathbf{A}$, see Algorithm 2. The exploration of all nondiagonal positions (p,q), $p<q$, is called a *sweep*. For hermitian matrices, it can be shown that this algorithm is faster than Classical Jacobi, see [13].

Algorithm 1: Classical Jacobi

Input: matrix $\mathbf{A}$
Output: unitary diagonalizer $\mathbf{U}$

$\mathbf{U} \leftarrow \mathbf{I}_n$
$\epsilon \leftarrow \mathrm{tol} \cdot \|\mathbf{A}\|_F$
while off $(\mathbf{A}) > \epsilon$ **do**
  choose p and q such that $|a_{pq}|^2$ is maximum
  compute $c_{\mathrm{d}}^\star$ and $s_{\mathrm{d}}^\star$ according to Eq. (2.8)
  $\mathbf{A} \leftarrow \mathbf{G}(p, q, c_{\mathrm{d}}^\star, s_{\mathrm{d}}^\star)\, \mathbf{A}\, \mathbf{G}(p, q, c_{\mathrm{d}}^\star, s_{\mathrm{d}}^\star)^H$
  $\mathbf{U} \leftarrow \mathbf{U}\, \mathbf{G}(p, q, c_{\mathrm{d}}^\star, s_{\mathrm{d}}^\star)$

Algorithm 2: Cyclic Jacobi

Input: matrix $\mathbf{A}$
Output: unitary diagonalizer $\mathbf{U}$

$\mathbf{U} \leftarrow \mathbf{I}_n$
$\epsilon \leftarrow \mathrm{tol} \cdot \|\mathbf{A}\|_F$
while off $(\mathbf{A}) > \epsilon$ **do**
  for $p \leftarrow 1$ ***to*** $n-1$ **do**
    for $q \leftarrow p+1$ ***to*** n **do**
      compute $c_{\mathrm{d}}^\star$ and $s_{\mathrm{d}}^\star$ according to Eq. (2.8)
      $\mathbf{A} \leftarrow \mathbf{G}(p, q, c_{\mathrm{d}}^\star, s_{\mathrm{d}}^\star)\, \mathbf{A}\, \mathbf{G}(p, q, c_{\mathrm{d}}^\star, s_{\mathrm{d}}^\star)^H$
      $\mathbf{U} \leftarrow \mathbf{U}\, \mathbf{G}(p, q, c_{\mathrm{d}}^\star, s_{\mathrm{d}}^\star)$

2.1.5. Convergence of the algorithm. By construction, the algorithm ensures decrease of criterion C_{d} at each iteration. Indeed, by definition of $c_{\mathrm{d}}^\star$ and $s_{\mathrm{d}}^\star$, we get for all $(c, s) \in \mathbb{R} \times \mathbb{C}$ with $c^2 + |s|^2 = 1$ that

$$C_{\mathrm{d}}(\mathbf{G}(p, q, c_{\mathrm{d}}^\star, s_{\mathrm{d}}^\star)) \leq C_{\mathrm{d}}(\mathbf{G}(p, q, c, s)).$$

In particular, for $(c, s) = (1, 0)$ we find

$$\mathrm{off}\,(\mathbf{B}) \leq \mathrm{off}\,(\mathbf{A})\,.$$

At each iteration of the algorithm, the matrix $\mathbf{B}$ obtained after rotations is thus 'at least as diagonal as' matrix $\mathbf{A}$ at previous iteration. Since every bounded monotonic sequence in $\mathbb{R}$ converges, the convergence of our algorithm is guaranteed. However, this does not guarantee that the algorithm converges to the minimum of C_{d}, except when $\mathbf{A}$ is hermitian, as shown in the next paragraph.

2.1.6. Special case: hermitian matrices. If $\mathbf{A} \in \mathbb{C}^{n \times n}$ is hermitian, then it is diagonalizable in a unitary basis, and its eigenvalues are real. In that case it is possible to give an explicit form of off $(\mathbf{B})$ − off $(\mathbf{A})$, as a function of the coefficients of $\mathbf{A}$ only. Indeed, in this case, the vector $\mathbf{h}(\mathbf{A})$ is real and thus $\mathbf{Q}_{\mathrm{d}} = \mathbf{h}(\mathbf{A})^T \mathbf{h}(\mathbf{A})$. We saw previously that the maximum of $C'_{\mathrm{d}}(c, s)$ is equal to the largest eigenvalue of $\mathbf{Q}_{\mathrm{d}}$. Now, $\mathbf{Q}_{\mathrm{d}}$ is by construction a positive semidefinite matrix of rank 1, hence it has only one nonzero eigenvalue, equal to its trace. Thus

$$|b_{pp} - b_{qq}|^2 = \mathrm{trace}(\mathbf{Q}_{\mathrm{d}}) = |a_{pp} - a_{qq}|^2 + |a_{pq} + a_{qp}|^2 + |a_{pq} - a_{qp}|^2$$

Then, from Eq. (2.5):

$$\begin{aligned}\text{off}\,(\mathbf{B}) - \text{off}\,(\mathbf{A}) &= \frac{1}{2}\left(|a_{pp} - a_{qq}|^2 - |b_{pp} - b_{qq}|^2\right) = -\frac{1}{2}\left(|a_{pq} + a_{qp}|^2 + |a_{pq} - a_{qp}|^2\right)\\ &= -\left(|a_{pq}|^2 + |a_{qp}|^2\right) = -2\,|a_{pq}|^2.\end{aligned}$$

Hence, in the case of hermitian matrices, absolute decrease of the criterion is ensured at each iteration, as long as one of the nondiagonal entries of $\mathbf{A}$ is nonzero. This guarantees convergence of the algorithm to the minimum of C_{d}, which is 0 for hermitian matrices.

2.2. Real orthonormal basis. We now present a few simplifications of the previous method in the specific case where we want to diagonalize $\mathbf{A} \in \mathbb{C}^{n\times n}$ in an orthonormal basis, i.e $\mathbf{U}$ is from the orthogonal group $O(n)$ and $\mathbf{D} \in \mathbb{C}^{n\times n}$. Note that a perfect diagonalizer exists if and only if $\mathbf{A}$ is symmetric, and any solution is unique only up to left multiplication by a permutation and sign matrix.

In the real case, we look for an orthonormal basis in the form of a product of *real* Givens matrices $\mathbf{G}_r(p, q, c, s)$, defined for $p < q$ by

$$\mathbf{G}_r(p,q,c,s) = \begin{array}{c c} \left[\begin{array}{ccccccc} 1 & \dots & 0 & \dots & 0 & \dots & 0\\ \vdots & \ddots & \vdots & & \vdots & & \vdots\\ 0 & \dots & c & \dots & s & \dots & 0\\ \vdots & & \vdots & \ddots & \vdots & & \vdots\\ 0 & \dots & -s & \dots & c & \dots & 0\\ \vdots & & \vdots & & \vdots & \ddots & \vdots\\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{array}\right] & \begin{array}{c} \\ \\ p\\ \\ q\\ \\ \\ \end{array}\\ \begin{array}{ccccccc} & & p & & q & & \end{array} & \end{array} \tag{2.9}$$

with $(c, s) \in \mathbb{R}^2$ such that $c^2 + s^2 = 1$. As before, for fixed p and q, one iteration consists of the following steps:

- determine (c, s) with minimal $C_{\mathrm{d}}(\mathbf{G}_r(p, q, c, s); \mathbf{A})$
- set $\mathbf{A} \leftarrow \mathbf{G}_r(p, q, c, s)\,\mathbf{A}\,\mathbf{G}_r(p, q, c, s)^T$

Like in the complex case, minimizing $C_{\mathrm{d}}(\mathbf{G}_r(p, q, c, s); \mathbf{A})$ amounts to maximizing $C'_{\mathrm{d}}(c, s) = |b_{pp} - b_{qq}|^2$. However, from Eq. (2.6) in the real case we have:

$$b_{pp} - b_{qq} = (c^2 - s^2)(a_{pp} - a_{qq}) + 2\,c\,s\,(a_{pq} + a_{qp})$$

Defining

$$\begin{aligned}\mathbf{v}_r(c,s) &= \left(c^2 - s^2, 2\,c\,s\right)^T\\ \mathbf{h}_r(\mathbf{A}) &= (a_{pp} - a_{qq}, a_{pq} + a_{qp})\end{aligned}$$

we obtain $b_{pp} - b_{qq} = \mathbf{h}_r(\mathbf{A})\,\mathbf{v}_r(c, s)$ and hence

$$C'_{\mathrm{d}}(c,s) = \mathbf{v}_r(c,s)^T\ \mathbf{h}_r(\mathbf{A})^T\mathbf{h}_r(\mathbf{A})\ \mathbf{v}_r(c,s).$$

Like before, the maximization of $C'_{\mathrm{d}}(c, s)$ is equivalent to the maximization of the quadratic form $q_r(\mathbf{x}) = \mathbf{x}^T\,\mathbf{Q}_{dr}\,\mathbf{x}$ on the unit sphere, with $\mathbf{Q}_{dr} = \mathrm{Re}\left\{\mathbf{h}_r(\mathbf{A})^T\mathbf{h}_r(\mathbf{A})\right\}$. Let $(u^\star_{dr}, v^\star_{dr})^T$ be a unit-norm eigenvector of $\mathbf{Q}_{dr}$ corresponding to its largest eigenvalue, chosen such that $u^\star_{dr} \geq 0$. From Eq. (2.7), the expressions of $(c^\star_{dr}, s^\star_{dr})$ which minimize criterion $C_{\mathrm{d}}(\mathbf{G}_r(p, q, c, s); \mathbf{A})$ for given p and q are

$$c^\star_{dr} = \sqrt{\frac{u^\star_{dr} + 1}{2}}, \quad s^\star_{dr} = \frac{v^\star_{dr}}{\sqrt{2\,(u^\star_{dr} + 1)}}.$$

Note that it is possible to give algebraic expressions of $c^\star_{dr}$ and $s^\star_{dr}$ using closed form expressions of the eigenvectors of a 2×2 matrix.

3. Approximate joint diagonalization. Many applications [2, 5, 9, 12, 16, 17], especially in the field of blind signal processing, face the problem of recovering a diagonalizing orthonormal basis $\mathbf{U}$ of one or possibly multiple matrices $\mathbf{A}_k$. As discussed in Section 2.1.1, diagonalization of only a single matrix $\mathbf{A}_1$ might result in considerable indeterminacies due to a degenerate eigenvalue structure and of course due to estimation noise in $\mathbf{A}_1$ itself.

Consider for example

$$\mathbf{A}_1 = \mathbf{U}^H \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \mathbf{U}, \qquad \mathbf{A}_2 = \mathbf{U}^H \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \mathbf{U}.$$

Diagonalization of only $\mathbf{A}_1$ or $\mathbf{A}_2$ separately may produce solutions different from $\mathbf{U}$, because the basis in the two-dimensional eigenspace corresponding to the eigenvalue $k \in \{1, 2\}$ cannot be uniquely chosen. If however a common basis of these two matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ is to be found, $\mathbf{U}$ is the unique solution except for permutations and signs, see Section 3.1.1. Also, if the $\mathbf{A}_k$ are deteriorated by noise — coming for instance from estimation errors in practical settings — then again the search for an (approximate) common basis increases the statistical validity of the resulting diagonalizer as estimate of some underlying unknown basis. This is the fundamental reason for generalizing diagonalization problems to the joint diagonalization of multiple matrices.

3.1. Complex orthonormal basis. We consider the problem of approximate joint diagonalization, also encountered under the name of approximate simultaneous diagonalization. Let $\mathcal{A} = \{\mathbf{A}_1, \ldots, \mathbf{A}_K\}$ be a set of K complex $(n \times n)$-matrices. Our objective is to find a unitary matrix $\mathbf{U} \in U(n)$ such that for all $k = 1, \ldots, K$, the matrices

$$\mathbf{U}\,\mathbf{A}_k\,\mathbf{U}^H = \mathbf{D}_k$$

are as diagonal as possible, in the sense of criterion (2.1). In other words, we want to minimize

$$C_{\mathrm{jd}}(\mathbf{V}; \mathcal{A}) = \sum_{k=1}^{K} \operatorname{off}\left(\mathbf{V}\,\mathbf{A}_k\,\mathbf{V}^H\right) \tag{3.1}$$

with respect to $\mathbf{V} \in U(n)$.

3.1.1. Existence and uniqueness. Again, we will only discuss the limit case of perfect factorization, where $C_{\mathrm{jd}}(\mathbf{U}; \mathcal{A}) = 0$. According to the case of $K = 1$ from Section 2.1.1, a necessary condition for a joint diagonalizer $\mathbf{U}$ to exist is that each $\mathbf{A}_k$ is normal. In order to guarantee a common orthonormal basis, moreover it is sufficient for the $\mathbf{A}_k$ to commute, so a joint diagonalizer exists if and only if the normal $\mathbf{A}_k$ commute.

If one of the matrices $\mathbf{A}_k$ is unispectral, then the solution $\mathbf{U}$ is unique except for permutation and unit-length scalars according to Section 2.1.1. However, due to the fact that we now jointly diagonalize multiple matrices, this condition can be relaxed considerably. Indeed, $\mathbf{U}$ is unique except for the trivial indeterminacies from above if and only if for any two different eigenvectors (rows of $\mathbf{U}$) at least one $\mathbf{A}_k$ has distinct corresponding eigenvalues, see [2], theorem 3.

3.1.2. Method. As before the common orthonormal basis $\mathbf{U}$ is estimated iteratively. For fixed p and q one iteration of the method consists of

- finding (c, s) with minimal $C_{\text{jd}}(\mathbf{G}(p, q, c, s); \mathcal{A})$, and
- $\forall k : \mathbf{A}_k \leftarrow \mathbf{G}(p, q, c, s)\, \mathbf{A}_k\, \mathbf{G}(p, q, c, s)^H$.

If we set $\mathbf{B}_k = \mathbf{G}(p, q, c, s)\, \mathbf{A}_k\, \mathbf{G}(p, q, c, s)^H$ for all $k = 1, \dots, K$, then

$$C_{\text{jd}}(\mathbf{G}(p, q, c, s); \mathcal{A}) = \sum_{k=1}^{K} \text{off}\,(\mathbf{B}_k)\,.$$

Using the notation $\mathbf{A}_k = \{a_{kij}\}$, Eq. (2.4) shows

$$\begin{aligned} C_{\text{jd}}(\mathbf{G}(p, q, c, s); \mathcal{A}) &= \sum_{k=1}^{K} \text{off}\,(\mathbf{A}_k) + |a_{kpp}|^2 + |a_{kqq}|^2 - |b_{kpp}|^2 - |b_{kqq}|^2 \\ &= \sum_{k=1}^{K} \text{off}\,(\mathbf{A}_k) + \frac{1}{2}\left(|a_{kpp} - a_{kqq}|^2 - |b_{kpp} - b_{kqq}|^2\right). \end{aligned} \tag{3.2}$$

Like in Section 2.1.3, the minimization of $C_{\text{jd}}(\mathbf{G}(p, q, c, s); \mathcal{A})$ amounts to the maximization of $C'_{\text{jd}}(c, s) := \sum_{k=1}^{K} |b_{kpp} - b_{kqq}|^2$. With same notations as before we thus get

$$\begin{aligned} C'_{\text{jd}}(c, s) &= \sum_{k=1}^{K} \mathbf{v}(c, s)^T\ \mathbf{h}(\mathbf{A}_k)^H \mathbf{h}(\mathbf{A}_k)\ \mathbf{v}(c, s) \\ &= \mathbf{v}(c, s)^T\ \left(\sum_{k=1}^{K} \mathbf{h}(\mathbf{A}_k)^H \mathbf{h}(\mathbf{A}_k)\right)\ \mathbf{v}(c, s) \\ &= \mathbf{v}(c, s)^T\ \text{Re}\left\{\sum_{k=1}^{K} \mathbf{h}(\mathbf{A}_k)^H \mathbf{h}(\mathbf{A}_k)\right\}\ \mathbf{v}(c, s). \end{aligned}$$

The maximization of $C'_{\text{jd}}(c, s)$ on $\mathcal{E}$ as defined in lemma 2.1 is hence equivalent to the maximization of the quadratic form $q_{\text{jd}}(\mathbf{x}) = \mathbf{x}^T\, \mathbf{Q}_{\text{jd}} \mathbf{x}$ on the unit sphere $\mathcal{S}^2$, with

$$\mathbf{Q}_{\text{jd}} = \text{Re}\left\{\sum_{k=1}^{K} \mathbf{h}(\mathbf{A}_k)^H \mathbf{h}(\mathbf{A}_k)\right\}.$$

Let $(u^\star_{\text{jd}}, v^\star_{\text{jd}}, w^\star_{\text{jd}})^T$ be a unit-norm eigenvector of $\mathbf{Q}_{\text{jd}}$ corresponding to the largest eigenvalue, and chosen such that $u^\star_{\text{jd}} \geq 0$. From Eq.(2.7), the expressions of $(c^\star_{\text{jd}}, s^\star_{\text{jd}})$ which minimize criterion $C_{\text{jd}}(\mathbf{G}(p, q, c, s); A)$ for fixed p and q are

$$c^\star_{\text{jd}} = \sqrt{\frac{u^\star_{\text{jd}} + 1}{2}}, \quad s^\star_{\text{jd}} = \frac{v^\star_{\text{jd}} - \mathrm{i}\, w^\star_{\text{jd}}}{\sqrt{2\,(u^\star_{\text{jd}} + 1)}}.$$

3.1.3. Algorithm convergence. Like for the approximate diagonalization of a complex matrix (Section 2.1), the algorithm guarantees decrease of criterion C_{jd} at each iteration. Indeed, for all $(c, s) \in \mathbb{R} \times \mathbb{C}$ with $c^2 + |s|^2 = 1$, we have

$$C_{\text{jd}}(\mathbf{G}(p, q, c^\star_{\text{jd}}, s^\star_{\text{jd}})) \leq C_{\text{jd}}(\mathbf{G}(p, q, c, s)),$$

and in particular, for $(c, s) = (1, 0)$

$$\sum_{k=1}^{K} \text{off}\,(\mathbf{B}_k) \leq \sum_{k=1}^{K} \text{off}\,(\mathbf{A}_k)\,.$$

Cardoso and Souloumiac's algorithm [10] generalizes the Cyclic Jacobi algorithm of Section 2.1.4; all the off-diagonal entries (p, q) are swept row by row. A MATLAB implementation of their approach is available on Cardoso's webpage [7]. In this implementation the algorithm stops when all the values of $s^\star_{\text{jd}}$ within a sweep are lower than a given threshold (set by default to the square root of the machine precision). However, like for the diagonalization case, the convergence of the algorithm to the minimum of C_{jd} is not guaranteed; except in the case where $\mathcal{A}$ is a set of commuting hermitian matrices. A counter example is found in [6].

3.2. Real orthonormal basis. Like in Section 2.2, a few simplifications can be made when real coefficients are used, and $\mathbf{U} \in O(n)$. Then the expressions of $(c^\star_{\text{jdr}}, s^\star_{\text{jdr}})$ minimizing $C_{\text{jd}}(\mathbf{G}_r(p, q, c, s); \mathcal{A})$ for fixed p and q are

$$c^\star_{\text{jdr}} = \sqrt{\frac{u^\star_{\text{jdr}} + 1}{2}}, \quad s^\star_{\text{jdr}} = \frac{v^\star_{\text{jdr}}}{\sqrt{2\,(u^\star_{\text{jdr}} + 1)}}$$

where $(u^\star_{\text{jdr}}, v^\star_{\text{jdr}})^T$, $u^\star_{\text{jdr}} \geq 0$, is a unit-norm eigenvector of $\mathbf{Q}_{\text{jdr}} = \sum_{k=1}^{K} \mathbf{h}_r(\mathbf{A}_k)^T \mathbf{h}_r(\mathbf{A}_k)$ associated to the largest eigenvalue.

4. Approximate block-diagonalization. For the sake of clarity we now consider the extension of the latter techniques to approximate block diagonalization of a single matrix, before approximate *joint* block diagonalization is finally investigated in the next section. We will only deal with the case of fixed block-size, which is known in advance — a situation relevant to many practical applications [4, 5], although the problem of unknown block-structure is worth investigating by itself.

4.1. Complex orthonormal basis. Let $\mathbf{A} \in \mathbb{C}^{n \times n}$. Our objective is to find an orthonormal matrix $\mathbf{U} \in \mathbb{C}^{n \times n}$ such that

$$\mathbf{U}\,\mathbf{A}\,\mathbf{U}^H = \mathbf{D}$$

is as block-diagonal as possible. In the following we note L the (fixed) dimension of the diagonal blocks of $\mathbf{D}$ and $m = n/L$ the number of blocks. We decompose

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \dots & \mathbf{A}_{1m} \\ \vdots & & \vdots \\ \mathbf{A}_{m1} & \dots & \mathbf{A}_{mm} \end{bmatrix}$$

into matrices $\mathbf{A}_{ij}$ of dimension $L \times L$, where $i, j = 1, \dots, m$. Our block-diagonality criterion is chosen as

$$\text{boff}\,(\mathbf{A}) := \sum_{1 \leq i \neq j \leq m} \|\mathbf{A}_{ij}\|_F^2 \tag{4.1}$$

and we look for $\mathbf{U} \in U(n)$ via minimization of criterion

$$C_{\text{bd}}(\mathbf{V}; \mathbf{A}) = \text{boff}\left(\mathbf{V}\,\mathbf{A}\,\mathbf{V}^H\right) \tag{4.2}$$

with respect to $\mathbf{V} \in U(n)$.

4.1.1. Existence and uniqueness. Consider again the case of perfect factorization, where $C_{\rm bd}(\mathbf{U};\mathbf{A}) = 0$ i.e. where $\mathbf{U}\mathbf{A}\mathbf{U}^H$ is block-diagonal with blocks of size L. Obvious indeterminacies of $\mathbf{U}$ are left-multiplication by a unitary block-diagonal matrix as generalization of the unit-scaling indeterminacy in the case of block size 1 from Section 2.1.1. Moreover, a permutation of blocks of rows of $\mathbf{U}$ yields again a block-diagonalizer; in other words, an additional indeterminacy is given by the left-multiplication by a block-permutation matrix i.e. a matrix consisting of blocks that are either zero or $\mathbf{I}_L$ such that in each row and column exactly one block is non-zero.

Additional indeterminacies might come into play if the block size has not been chosen adequately: consider for example the situation

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 1 & 3 \end{bmatrix}, \mathbf{U}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \mathbf{U}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$

Then $\mathbf{A}$ is already block-diagonal with blocks of size 2, but of course $\mathbf{A}$ is real symmetric, so diagonalizable, and indeed $\mathbf{U}_1\mathbf{A}\mathbf{U}_1^T$ is diagonal with entries $4, 2, 4, 2$ on the diagonal. But so is $\mathbf{U}_2\mathbf{A}\mathbf{U}_2^T$ (with diagonal $(4, 4, 2, 2)$), and $\mathbf{U}_1$ and $\mathbf{U}_2$ do not differ only by a block-permutation. This additional indeterminacy comes from the fact that the block-decomposition of $\mathbf{A}$ into blocks of size 2 is not maximal — a finer decomposition, 1-diagonality in this case, may be chosen. Also note that this is no special case for symmetric matrices; indeed by replacing one block in $\mathbf{A}$ by some non-symmetric block, a finer $(1, 1, 2)$-decomposition may be found.

Hence in full generality, we have to treat different block sizes and therefore have to look for a maximal-length decomposition in that setting. It can be shown that such a decomposition is then unique except for scaling and permutation of blocks of the same size. Our interest in this manuscript however lies in fixed block sizes.

Regarding existence of a block-diagonalizer of block-size L we generalize the results from the case $L = 1$ as follows: consider a Schur decomposition $\mathbf{A} = \mathbf{U}^H\mathbf{S}\mathbf{U}$ of $\mathbf{A}$ [13]; here $\mathbf{U}$ is unitary and $\mathbf{S}$ upper triangular. The Schur decomposition may not be unique if there are zeros in the upper triangular part of $\mathbf{S}$. It may of course be interpreted as generalization of a diagonalizer, and we can show the following simple lemma:

LEMMA 4.1. *A matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is block-diagonalizable of block-size L if and only if there exists a Schur decomposition $\mathbf{A} = \mathbf{U}^H\mathbf{S}\mathbf{U}$ such that $\mathbf{S}$ is L-block-diagonal.*

Proof. If such a Schur decomposition exists, it already is an L-block diagonalization. Now assume the converse; let $\mathbf{B} := \mathbf{U}\mathbf{A}\mathbf{U}^H$ be block-diagonal with blocks $\mathbf{B}_{11}, \dots, \mathbf{B}_{mm}$ on the diagonal. Let $\mathbf{B}_{ii} = \mathbf{V}_i^H\mathbf{S}_i\mathbf{V}_i$ be Schur decompositions of the blocks $\mathbf{B}_{ii}$ on the diagonal, so $\mathbf{V}_i \in U(L)$ and $\mathbf{S}_i \in \mathbb{C}^{L\times L}$ upper triangular. Then putting together these matrices,

$$\mathbf{V} := \begin{bmatrix} \mathbf{V}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{V}_m \end{bmatrix}, \qquad \mathbf{S} := \begin{bmatrix} \mathbf{S}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{S}_m \end{bmatrix}$$

yields a unitary matrix $\mathbf{U} \in U(n)$ and an upper triangular, block-diagonal matrix $\mathbf{S} \in \mathbb{C}^{n\times n}$. Moreover, by construction $\mathbf{B} = \mathbf{V}^H\mathbf{S}\mathbf{V}$ and therefore

$$\mathbf{A} = \mathbf{U}^H\mathbf{B}\mathbf{U} = \mathbf{U}^H\mathbf{V}^H\mathbf{S}\mathbf{V}\mathbf{U} = (\mathbf{V}\mathbf{U})^H\mathbf{S}(\mathbf{V}\mathbf{U})$$

is a Schur decomposition of $\mathbf{A}$ with block-diagonal triangular part $\mathbf{S}$. □

This lemma characterizes block-diagonalizability of $\mathbf{A}$, however it cannot be used to test efficiently for it, simply because it states that $\mathbf{A}$ must have a block-diagonal Schur decomposition, but *not all* Schur decompositions share this property. For example, with

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 3 \end{bmatrix}, \mathbf{U} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{U}\mathbf{A}\mathbf{U}^H = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

we get an upper triangular, 2-block-diagonal matrix $\mathbf{A}$, which is a block-diagonal Schur decomposition of itself. But also $\mathbf{U}\mathbf{A}\mathbf{U}^H$ is upper-triangular and thus constitutes a Schur decomposition of $\mathbf{A}$ with non-block-diagonal upper triangular part (for $L = 2$). Of course, the desired block-diagonal decomposition may be reconstructed from $\mathbf{U}\mathbf{A}\mathbf{U}^H$ simply by permutations.

But how to choose the permutation? We propose using Algorithm 3; we denote its resulting permuted matrix by $\mathcal{P}(\mathbf{A})$ when applied to the input $\mathbf{A}$. By construction, $\mathcal{P}(\mathbf{A})$ is constructed from $\mathbf{A}$ by iteratively permuting columns and rows in order to guarantee that all non-zeros of $\mathbf{A}$ are clustered along the diagonal as closely as possible. Clearly, if applied to a block-diagonal matrix, it will stay block-diagonal and only involve permutations within blocks, which belong to the trivial indeterminacies of block-diagonalization as subsumed above.

Algorithm 3: Block-diagonality permutation finder

Input: $(n \times n)$-matrix $\mathbf{A}$
Output: block-diagonal matrix $\mathcal{P}(\mathbf{A}) := \mathbf{B}$ such that $\mathbf{B} = \mathbf{P}\mathbf{A}\mathbf{P}^T$ for some permutation matrix $\mathbf{P}$

```
B ← A
for i ← 1 to n do
    repeat
        if (j0 ← min{j | j ≥ i and a_ij = 0 and a_ji = 0}) exists then
            if (k0 ← min{k | k > j0 and (a_ik ≠ 0 or a_ki ≠ 0)}) exists then
                swap column j0 of B with column k0
                swap row j0 of B with row k0
    until no swap has occurred;
```

We conjecture the following theorem:

THEOREM 4.2. *A matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is maximal block-diagonalizable of block-size L if and only if for any Schur decomposition $\mathbf{A} = \mathbf{U}^H\mathbf{S}\mathbf{U}$, the matrix $\mathcal{P}(\mathbf{S})$ after application of Algorithm 3 is L-block-diagonal.*

One direction is obviously true: If we assume that we have found a Schur decomposition $\mathbf{A} = \mathbf{U}^H\mathbf{S}\mathbf{U}$ with block-diagonal $\mathcal{P}(\mathbf{S})$. Then $\mathcal{P}(\mathbf{S}) = \mathbf{P}\mathbf{S}\mathbf{P}^H$ for some permutation matrix $\mathbf{P}$, so $\mathbf{A} = (\mathbf{P}\mathbf{U})^H\mathcal{P}(\mathbf{S})\mathbf{P}\mathbf{U}$ is a block-diagonalization of $\mathbf{A}$.

However the converse cannot be shown as easily. We are currently working on proofing the theorem using the uniqueness result of block-diagonalizability mentioned above. Nevertheless, care has to be taken with respect to the maximality condition. In any case, extensive simulations confirm the validity of the above conjecture.

This theorem now allows for easy testing whether or not a given matrix is block-diagonalizable; simply take any Schur decomposition, apply Algorithm 3 and check if the resulting upper-triangular matrix is block-diagonal. Moreover, this also directly yields a simple algorithm for perfect block-diagonalization. However, it is not easily extended to approximate block-diagonalization, nor to joint block-diagonalization, so in the following we will introduce a different algorithm, based on the ideas presented in the previous Sections.

4.1.2. Notations. In order to simplify notation in the block case, we introduce the following notation: Let $I_1, \ldots, I_m$ define a partition of $[\![1,n]\!] := \{1,2,\ldots,n\}$ such that

$$\begin{aligned} I_1 &= [\![1, L]\!] \\ &\vdots \\ I_i &= [\![(i-1)\,L+1, i\,L]\!] \\ &\vdots \\ I_m &= [\![(m-1)\,L+1, m\,L]\!]\,, \end{aligned}$$

and let $i(k) := \lceil k/L \rceil$ for $k \in [\![1,n]\!]$, where $\lceil x \rceil$ is the smallest integer larger than or equal to x. So $i(k)$ gives the index i of the interval I_i to which k belongs.

4.1.3. Method. As before, we apply successive Givens rotations to $\mathbf{A}$, until the criterion (4.2) is minimal; for fixed p and q, one iteration of the method consists of

- minimizing $C_{\mathrm{bd}}(\mathbf{G}(p,q,c,s);\mathbf{A})$ with respect to (c,s), and
- updating $\mathbf{A} \leftarrow \mathbf{G}(p,q,c,s)\,\mathbf{A}\,\mathbf{G}(p,q,c,s)^H$.

Note that according to section 4.1.1, $\mathbf{U}$ can only be estimated up to a block-diagonal unitary matrix with m blocks of dimension L and possibly a permutation of these blocks. Let $(p,q) \in [\![1,n]\!]^2, p < q$ and $\mathbf{B} = \mathbf{G}(p,q,c,s)\,\mathbf{A}\,\mathbf{G}(p,q,c,s)^H$. From Eq. (4.1), Eq. (4.2) and with $\|\mathbf{B}\|_F^2 = \sum_{ij} \|\mathbf{B}_{ij}\|_F^2$, we have

$$C_{\mathrm{bd}}(\mathbf{G}(p,q,c,s);\mathbf{A}) = \mathrm{boff}\,(\mathbf{B}) = \|\mathbf{B}\|_F^2 - \sum_{i=1}^{m} \|\mathbf{B}_{ii}\|_F^2 = \|\mathbf{A}\|_F^2 - \sum_{i=1}^{m} \|\mathbf{B}_{ii}\|_F^2.$$

Assume that p and q belong to the same sub-interval I_i, i.e, $i(p) = i(q)$. From Eq. (2.3), $\mathbf{B}$ is everywhere equal to $\mathbf{A}$, except on the p^{th} and q^{th} rows and columns. Hence

$$\sum_{i=1}^{m} \|\mathbf{B}_{ii}\|_F^2 = \sum_{i=1}^{m} \|\mathbf{A}_{ii}\|_F^2 - \left\|\mathbf{A}_{i(p)i(p)}\right\|_F^2 + \left\|\mathbf{B}_{i(p)i(p)}\right\|_F^2.$$

Now, because p and q belong to the same sub-interval, $\left\|\mathbf{B}_{i(p)i(p)}\right\|_F = \left\|\mathbf{A}_{i(p)i(p)}\right\|_F$, which follows from the invariance of Frobenius norm under unitary transformation. Hence, if $i(p) = i(q)$ then $\sum_{i=1}^m \|\mathbf{B}_{ii}\|_F^2 = \sum_{i=1}^m \|\mathbf{A}_{ii}\|_F^2$, i.e. $\mathrm{boff}\,(\mathbf{B}) = \mathrm{boff}\,(\mathbf{A})$, so C_{bd} remains constant. In the following we will thus assume $i(p) \neq i(q)$, i.e, $q - p \geq L$. In that case we have

$$\begin{aligned} &C_{\mathrm{bd}}(\mathbf{G}(p,q,c,s);\mathbf{A}) = \\ &\quad \|\mathbf{A}\|_F^2 - \left\{ \sum_{k=1}^{m} \|\mathbf{A}_{kk}\|_F^2 - \left\|\mathbf{A}_{i(p)i(p)}\right\|_F^2 - \left\|\mathbf{A}_{i(q)i(q)}\right\|_F^2 + \left\|\mathbf{B}_{i(p)i(p)}\right\|_F^2 + \left\|\mathbf{B}_{i(q)i(q)}\right\|_F^2 \right\} \\ &= \mathrm{boff}\,(\mathbf{A}) + \left\|\mathbf{A}_{i(p)i(p)}\right\|_F^2 + \left\|\mathbf{A}_{i(q)i(q)}\right\|_F^2 - \left\|\mathbf{B}_{i(p)i(p)}\right\|_F^2 - \left\|\mathbf{B}_{i(q)i(q)}\right\|_F^2. \end{aligned} \tag{4.3}$$

Minimization of $C_{\text{bd}}(\mathbf{G}(p,q,c,s);\mathbf{A})$ is thus equivalent to maximization of $\left\|\mathbf{B}_{i(p)i(p)}\right\|_F^2 + \left\|\mathbf{B}_{i(q)i(q)}\right\|_F^2$. However, because only the p^{th} and q^{th} rows and columns of $\mathbf{B}$ depend on c and s, the minimization of C_{bd} finally amounts to the maximization of criterion

$$C'_{\text{bd}}(c,s) := |b_{pp}|^2 + |b_{qq}|^2 + \sum_{j\in I_{i(p)}, j\neq p} |b_{pj}|^2 + |b_{jp}|^2 + \sum_{j\in I_{i(q)}, j\neq q} |b_{qj}|^2 + |b_{jq}|^2,$$

where we recall from (2.3) the expressions of b_{pp}, b_{qq}, b_{pj}, b_{jp}, b_{qj}:

$$\begin{aligned}
b_{pp} &= c^2\, a_{pp} + |s|^2\, a_{qq} + c\, s\, a_{pq} + c\, \bar{s}\, a_{qp} \\
b_{qq} &= c^2\, a_{qq} + |s|^2\, a_{pp} - c\, s\, a_{pq} - c\, \bar{s}\, a_{qp} \\
b_{pj} &= c\, a_{pj} + \bar{s}\, a_{qj} \quad (j \in I_{i(p)}, j \neq p) \\
b_{jp} &= c\, a_{jp} + s\, a_{jq} \quad (j \in I_{i(p)}, j \neq p) \\
b_{qj} &= -s\, a_{pj} + c\, a_{qj} \quad (j \in I_{i(q)}, j \neq q) \\
b_{jq} &= -\bar{s}\, a_{jp} + c\, a_{jq} \quad (j \in I_{i(q)}, j \neq q)
\end{aligned}$$

Like before, for sake of clarity in the notations we discard the dependence of $C'_{\text{bd}}(c,s)$ on p, q and $\mathbf{A}$.

It may be shown [1,3] that the maximization of $C'_{\text{bd}}(c,s)$ boils down to the constrained maximization of a linear quadratic form. This optimization can be achieved using Lagrange multipliers. The computation of the latter requires solving a polynomial of degree 6 in the complex case (i.e, $\mathbf{U} \in \mathbb{C}^{n\times n}$), and of degree 4 in the real case (i.e, $\mathbf{U} \in \mathbb{R}^{n\times n}$). First order approximations of the criterion are also considered in [1,3] to simplify its maximization. A tensorial rank-1 approximation is also found in [15].

4.2. Real matrices. We now consider a somehow different approach for the particular case of the block-diagonalization of a real matrix in a real orthonormal basis (which is the problem usually encountered in BSS). Our approach, sketched in [15], also boils down to the calculation at each iteration of the roots of a polynomial of degree 4, however its derivation is straight-forward; for example no parametrization of $O(n)$ using Lagrangian methods was necessary to construct the polynomial. As will be seen in the simulations of Section 5, the important issue that will have to be addressed is not how to maximize $C'_{\text{bd}}(c,s)$ at each iteration, but rather how to choose the couples (p,q).

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a matrix that we want to block-diagonalize in a real orthonormal basis $\mathbf{U} \in \mathbb{R}^{n\times n}$. Let us shortly remark that the complex existence and uniqueness results from Section 4.1.1 also hold in the real case — slight care has to be taken due to the fact that in a real Schur decomposition, the middle matrix is only 'block-triangular' in the sense that 2×2-blocks might occur on the diagonal, which correspond to complex eigenvalues of $\mathbf{A}$. However in such a case, the containing block of a possible block diagonalization is necessarily also of size at least 2, and includes the corresponding two columns, because otherwise we would get a real eigenvalue. Hence, the uniqueness results from above can be translated to the real case in a similar fashion.

As before we now look for an orthogonal basis as a product of real Givens matrices $\mathbf{G}_r(p,q,c,s)$. $C'_{\text{bd}}(c,s)$ is then equal to $C'_{\text{bdr}}(c,s)$, defined by

$$C'_{\text{bdr}}(c,s) := (b_{pp})^2 + (b_{qq})^2 + \sum_{j\in I_{i(p)}, j\neq p} (b_{pj})^2 + (b_{jp})^2 + \sum_{j\in I_{i(q)}, j\neq q} (b_{qj})^2 + (b_{jq})^2 \quad (4.4)$$

with entries

$$\begin{aligned}
(b_{pp})^2 &= (c^2\,a_{pp} + s^2\,a_{qq} + c\,s\,(a_{pq} + a_{qp}))^2 \\
(b_{qq})^2 &= (c^2\,a_{qq} + s^2\,a_{pp} - c\,s\,(a_{pq} + a_{qp}))^2 \\
(b_{pj})^2 &= (c\,a_{pj} + s\,a_{qj})^2\,(c^2 + s^2) \quad (j \in I_{i(p)}, j \neq p) \\
(b_{jp})^2 &= (c\,a_{jp} + s\,a_{jq})^2\,(c^2 + s^2) \quad (j \in I_{i(p)}, j \neq p) \\
(b_{qj})^2 &= (-s\,a_{pj} + c\,a_{qj})^2\,(c^2 + s^2) \quad (j \in I_{i(q)}, j \neq q) \\
(b_{jq})^2 &= (-s\,a_{jp} + c\,a_{jq})^2\,(c^2 + s^2) \quad (j \in I_{i(q)}, j \neq q)
\end{aligned}$$

$C'_{\mathrm{bdr}}(c,s)$ is a polynomial in c and s of degree 4. The expressions of $(b_{pj})^2$,$(b_{jp})^2$,$(b_{qj})^2$ and $(b_{jq})^2$ have been multiplied by $(c^2+s^2)=1$ in order to obtain an expression of $C'_{\mathrm{bdr}}(c,s)$ homogeneous in c and s (i.e all the terms in $C'_{\mathrm{bdr}}(c,s)$ are on the form $\alpha_{ij} c^i\, s^j$ with $i+j=4$).

Using polar coordinates, for all $(c,s) \in \mathbb{R}^2$ with $c^2+s^2=1$, there exists a unique $\theta \in [0, 2\pi[$ such that $(c,s) = (\cos\theta, \sin\theta)$. $C'_{\mathrm{bdr}}(c,s)$ can thus be solely expressed as a function of θ. Expanding expression (4.4) gives the following polynomial, to be maximized in θ:

$$C'_{\mathrm{bdr}}(\theta) = q_{40}\,\cos^4\theta + q_{04}\,\sin^4\theta + q_{31}\,\cos^3\theta\sin\theta + q_{13}\,\cos\theta\sin^3\theta + q_{22}\,\cos^2\theta\sin^2\theta.$$

The expressions of the coefficients q_{ij} are given the Appendix.

The function $C'_{\mathrm{bdr}}(\theta)$ is periodical with period π. We may thus determine its maximum on the interval $]-\frac{\pi}{2},\frac{\pi}{2}]$, so $\cos\theta \geq 0$. The derivative of $C'_{\mathrm{bd}}(\theta)$ is given by

$$\begin{aligned}
\partial\, C'_{\mathrm{bd}}/\partial\theta(\theta) = {} & q_{31}\,\cos^4\theta - q_{13}\,\sin^4\theta - 2\,(2\,q_{40} - q_{22})\,\cos^3\theta\sin\theta + \\
& 2\,(2\,q_{04} - q_{22})\,\cos\theta\sin^3\theta + 3\,(q_{13} - q_{31})\,\cos^2\theta\sin^2\theta.
\end{aligned}$$

Dividing $C'_{\mathrm{bdr}}(\theta)$ by $\cos^4\theta$ for $\theta \neq \frac{\pi}{2}$, we get altogether:

$$\partial\, C'_{\mathrm{bdr}}/\partial\theta(\theta) = 0 \iff \begin{cases} \theta = \frac{\pi}{2} & \text{if } q_{13} = 0 \\ P(\tan\theta) = 0 & \text{if } \theta \in]-\frac{\pi}{2},\frac{\pi}{2}[\end{cases}$$

where $P(x)$ is the polynomial defined by

$$P(x) = q_{13}\;x^4 - 2\,(2\,q_{04} - q_{22})\;x^3 - 3\,(q_{13} - q_{31})\;x^2 + 2\,(2\,q_{40} - q_{22})\;x - q_{31}.$$

Thus, the maximization of $C'_{\mathrm{bdr}}(\theta)$ amounts to computing the roots the latter polynomial of degree 4. The roots of P can be calculated in closed form, however the expressions are unfeasibly long; in the following we will simply estimate them numerically and keep the real root $x^\star_{\mathrm{bdr}}$ whose inverse tangent $\theta^\star_{\mathrm{bdr}}$ maximizes $C'_{\mathrm{bdr}}(\theta)$.

5. Approximate joint block-diagonalization. We finally combine the ideas of simultaneous or joint diagonalization with the idea of block diagonalization. In this Section, this combination is discussed and an algorithm is proposed. Moreover, some results on reducing the joint block-diagonalization problem to joint diagonalization are presented.

5.1. Complex orthonormal basis. The approximate joint block-diagonalization of a set of K complex matrices $\mathcal{A} = \{\mathbf{A}_1, \ldots, \mathbf{A}_K\}$ is considered. The problem consist of finding an orthonormal matrix $\mathbf{U} \in \mathbb{C}^{n\times n}$ such that $\forall k \in [\![1, K]\!]$, the matrices

$$\mathbf{U}\,\mathbf{A}_k\,\mathbf{U}^H = \mathbf{B}_k$$

are as block-diagonal as possible, in the sense of criterion (4.1). Defining $\mathbf{A}_k = \{a_{kij}\}$ and, for $k \in [\![1, K]\!]$:

$$\mathbf{A}_k = \begin{bmatrix} \mathbf{A}_{k11} & \dots & \mathbf{A}_{k1m} \\ \vdots & & \vdots \\ \mathbf{A}_{km1} & \dots & \mathbf{A}_{kmm} \end{bmatrix}$$

where $\forall (i, j) \in [\![1, m]\!]^2$ and $\mathbf{A}_{kij}$ is of dimensions $L \times L$. We look for $\mathbf{U}$ by minimizing criterion

$$C_{\mathrm{jbd}}(\mathbf{V}; \mathcal{A}) := \sum_{i=1}^{K} \mathrm{boff}\left(\mathbf{V}\,\mathbf{A}_i\,\mathbf{V}^H\right)$$

with respect to the unitary matrix $\mathbf{V} \in U(n)$.

5.1.1. Existence and uniqueness. As before, we will only discuss the case of perfect factorization, where $C_{\mathrm{jbd}}(\mathbf{U}; \mathcal{A}) = 0$. First let us note that according to the results for $K = 1$ from Section 4.1.1, uniqueness can only hold up to block-permutation and unitary block-scaling, as these two transformations preserve block-diagonality. And indeed, in many situations these are already all indeterminacies.

For simplicity, we now only consider normal matrices $\mathbf{A}_k$ — a situation often encountered in practice. If we moreover assume that they are unispectral, then uniqueness follows already from Section 2.1.1, as all matrices $\mathbf{A}_k$ are diagonalizable. But how about existence? For this, at first in the case of $K = 2$, given an eigenvector $\mathbf{v}$ of $\mathbf{A}_1$, we define an index

$$\begin{aligned} \kappa(\mathbf{v}; \mathbf{A}_1, \mathbf{A}_2) := \min_k\ & \exists \text{ eigenvectors } \mathbf{v}_1, \dots, \mathbf{v}_{k-1} \neq \mathbf{v} \text{ of } \mathbf{A}_1 \\ & \exists \text{ eigenvectors } \mathbf{w}_1, \dots, \mathbf{w}_k \text{ of } \mathbf{A}_2 \text{ such that} \\ & \langle \mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}\rangle = \langle \mathbf{w}_1, \dots, \mathbf{w}_k\rangle. \end{aligned} \tag{5.1}$$

So $\kappa(\mathbf{v}; \mathbf{A}_1, \mathbf{A}_2)$ is the minimal dimension of a vector space generated by eigenvalues of $\mathbf{A}_1$ and containing $\mathbf{v}$ such that it may also be generated by eigenvalues of $\mathbf{A}_2$. In other words, it measures the minimal number of additional eigenvalues needed to group with $\mathbf{v}$ such that compatibility (or 'block-commutativity') with eigenvectors of $\mathbf{A}_2$ is achieved. Note that if we do not want to assume unispectral $\mathbf{A}$, the eigenvectors simply have to be replaced by maximal eigenspaces.

LEMMA 5.1. *The eigenvectors $\mathbf{v}_i$ and $\mathbf{w}_i$ from Eq.* (5.1) *are already uniquely determined by $\mathbf{v}$ except for permutation and unit scalars.*

Proof. Let $k = \kappa(\mathbf{v}; \mathbf{A}_1, \mathbf{A}_2)$, and consider two different minimal eigenvector representations

$$\begin{aligned} \langle \mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}\rangle &= \langle \mathbf{w}_1, \dots, \mathbf{w}_k\rangle \\ \langle \mathbf{v}'_1, \dots, \mathbf{v}'_{k-1}, \mathbf{v}\rangle &= \langle \mathbf{w}'_1, \dots, \mathbf{w}'_k\rangle. \end{aligned}$$

Then the intersection on each side must consist again of eigenvectors, as the corresponding eigenspaces $\langle \mathbf{v}_i\rangle$ are orthogonal, similarly for $\langle \mathbf{w}_i\rangle$. Moreover

$$\mathbf{v} \in \langle \mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}\rangle \cap \langle \mathbf{v}'_1, \dots, \mathbf{v}'_{k-1}, \mathbf{v}\rangle \neq \emptyset,$$

and

$$\langle \mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{v} \rangle \cap \langle \mathbf{v}'_1, \ldots, \mathbf{v}'_{k-1}, \mathbf{v} \rangle = \langle \mathbf{w}_1, \ldots, \mathbf{w}_k \rangle \cap \langle \mathbf{w}'_1, \ldots, \mathbf{w}'_k \rangle$$

is another equality according to Eq. (5.1). But this means that the dimension of the above intersection may not decrease (due to minimality), so

$$\langle \mathbf{v}_1, \ldots, \mathbf{v}_{k-1} \rangle = \langle \mathbf{v}'_1, \ldots, \mathbf{v}'_{k-1} \rangle$$
$$\langle \mathbf{w}_1, \ldots, \mathbf{w}_k \rangle = \langle \mathbf{w}'_1, \ldots, \mathbf{w}'_k \rangle.$$

and the claim follows. □

THEOREM 5.2. *Given a set of normal unispectral matrices $\mathcal{A} = \{\mathbf{A}_1, \ldots, \mathbf{A}_K\}$, a joint L-block-diagonalizer exists if for each eigenvector $\mathbf{v}$ of $\mathbf{A}_1$ and all $k = 2, \ldots, K$, we have $\kappa(\mathbf{v}; \mathbf{A}_1, \mathbf{A}_k) = L$.*

Proof. First assume $K = 2$. Then according to the uniqueness result of lemma 5.1, the space $\mathbb{R}^n$ may be decomposed into subspaces of dimension L:

$$\begin{aligned}\mathbb{R}^n =& \langle \mathbf{v}_1, \ldots, \mathbf{v}_L \rangle \oplus \ldots \oplus \langle \mathbf{v}_{m(L-1)+1}, \ldots, \mathbf{v}_{mL} \rangle \\ =& \langle \mathbf{w}_1, \ldots, \mathbf{w}_L \rangle \oplus \ldots \oplus \langle \mathbf{w}_{m(L-1)+1}, \ldots, \mathbf{w}_{mL} \rangle\end{aligned}$$

Here, the $\mathbf{v}_i$ and $\mathbf{w}_i$ are eigenvectors of $\mathbf{A}_1$ and $\mathbf{A}_2$ respectively. But the above decomposition implies that a unitary basis of $\mathbb{R}^n$ may be chosen that maps the above L-dimensional subspaces of the $\mathbf{v}_i$ onto those of the $\mathbf{w}_i$.

For $K > 2$, we may now choose an L-block-diagonalizer $\mathbf{U}_k$ according to the above for each tuple $(\mathbf{A}_1, \mathbf{A}_k)$. However due to the fact the $\mathbf{U}_k$ are essentially unique block-diagonalizers of $\mathbf{A}_1$, after possible block-scaling and permutation, we may assume that they are all equal, hence we get the desired joint block-diagonalizer. □

If $\mathcal{A}$ is joint L-block-diagonalizable, then necessarily $\kappa(\mathbf{v}; \mathbf{A}_1, \mathbf{A}_k) \leq L$, as the index $\kappa(\mathbf{v}; \mathbf{A}_1, \mathbf{A}_k)$ is a lower bound on the joint block size of the two matrices, where the block is given by the condition that it must contain $\mathbf{v}$. Only if the matrices are minimally L-block-diagonalizable, the equality of the two conditions holds.

Note that the case of varying block sizes may now easily be implemented by requiring different values of $\kappa(\mathbf{v}; \mathbf{A}_1, \mathbf{A}_k)$, which essentially measures the joint block size. The case of larger-dimensional eigenspaces follows easily by extending the definition of κ to include eigenspaces instead of eigenvalues only. An extension toward non-normal matrices may be realized as generalization of the case $K = 1$ from Section 4.1.1 by using the Schur decomposition, however it is not straightforward.

5.1.2. Method. For fixed p and q, one iteration of the algorithm again consists of the two steps

- minimizing $C_{\text{jbd}}(\mathbf{G}(p, q, c, s); \mathcal{A})$ with respect to (c, s), and
- updating $\mathbf{A}_k \leftarrow \mathbf{G}(p, q, c, s)\, \mathbf{A}_k\, \mathbf{G}(p, q, c, s)^H$ for all k.

If we define $\mathbf{B}_k = \mathbf{G}(p, q, c, s)\, \mathbf{A}_k\, \mathbf{G}(p, q, c, s)^H$, we get

$$C_{\text{jbd}}(\mathbf{G}(p, q, c, s); \mathcal{A}) = \sum_{k=1}^{K} \text{boff}\,(\mathbf{B}_k)\,.$$

From (4.3), for fixed p and q such that $i(p) \neq i(q)$, we therefore have

$$C_{\text{jbd}}(\mathbf{G}(p, q, c, s); \mathcal{A}) = \sum_{k=1}^{K} \text{boff}\,(\mathbf{A}_k) + \left\|\mathbf{A}_{ki(p)i(p)}\right\|_F^2 + \left\|\mathbf{A}_{ki(q)i(q)}\right\|_F^2 - \left\|\mathbf{B}_{ki(p)i(p)}\right\|_F^2 - \left\|\mathbf{B}_{ki(q)i(q)}\right\|_F^2. \quad (5.2)$$

The minimization of $C_{\text{jbd}}(\mathbf{G}(p,q,c,s);\mathcal{A})$ is thus equivalent to the maximization of $\sum_{k=1}^{K} \left\|\mathbf{B}_{ki(p)i(p)}\right\|_F^2 + \left\|\mathbf{B}_{ki(q)i(q)}\right\|_F^2$. However because only the p^{th} and q^{th} rows and columns of $\mathbf{B}_k$ depend on c and s for all k, the minimization of C_{jbd} amounts to maximizing the criterion

$$C'_{\text{jbd}}(c,s) := \sum_{k=1}^{K} \left\{ |b_{kpp}|^2 + |b_{kqq}|^2 + \sum_{j\in I_{i(p)}, j\neq p} |b_{kpj}|^2 + |b_{kjp}|^2 + \sum_{j\in I_{i(q)}, j\neq q} |b_{kqj}|^2 + |b_{kjq}|^2 \right\}$$

where the expressions of b_{kpp}, b_{kqq}, b_{kpj}, b_{kjp}, b_{kqj} and b_{kjq} are given by Eq. (2.3). As in Section 4.1, we now consider the particular case of real matrices to be joint-diagonalized in a real orthonormal basis.

5.2. Real matrices. If we assume that $\mathcal{A}$ is a set of real matrices and that we are looking for a common real basis $\mathbf{U} \in O(n)$, criterion $C_{\text{jbd}}(c,s)$ reduces to $C_{\text{jbdr}}(c,s)$ defined by

$$C'_{\text{jbdr}}(c,s) := \sum_{k=1}^{K} \left\{ (b_{kpp})^2 + (b_{kqq})^2 + \sum_{j\in I_{i(p)}, j\neq p} (b_{kpj})^2 + (b_{kjp})^2 + \sum_{j\in I_{i(q)}, j\neq q} (b_{kqj})^2 + (b_{kjq})^2 \right\}.$$

Using results of Section 4.2 and setting $(c,s) = (\cos\theta, \sin\theta)$, $C'_{\text{jbdr}}(c,s)$ can be written as a polynomial in $\cos\theta$ and $\sin\theta$:

$$C'_{\text{jbdr}}(\theta) = q_{40}\, \cos^4\theta + q_{04}\, \sin^4\theta + q_{31}\, \cos^3\theta \sin\theta + q_{13}\, \cos\theta\sin^3\theta + q_{22}\, \cos^2\theta\sin^2\theta$$

with coefficients given in the Appendix. The computation of the optimal value $\theta^\star_{\text{jbdr}}$ maximizing $C'_{\text{jbdr}}(\theta)$ is done as in Section 4.2.

5.3. Joint block-diagonalization by joint diagonalization. In this section we show that the form of the joint-diagonaliation minimization criterion $C_{\text{jd}}(\mathbf{U};\mathcal{A})$ may already imply joint block-diagonalization. For simplicity, we only treat the real case. Algorithmically this result implies that for JBD we may simply perform joint diagonalization and then permute the columns of $\mathbf{E}$ to achieve block-diagonality using Algorithm 3 — in experiments this turns out to be an efficient solution to JBD [1].

The result is based on a conjecture from [1] essentially claiming that a minimum of the JD cost function $C_{\text{jd}}(\mathbf{U};\mathcal{A})$ already is a JBD i.e. a minimum of the function $C_{\text{jbd}}(\mathbf{U};\mathcal{A})$ up to a permutation matrix. Indeed, in the conjecture it is required to use the Jacobi-update algorithm from [9], but indeed this is not necessary, and we can prove the conjecture partially:

We want to show that JD implies JBD up to permutation; i.e. if $\mathbf{U}$ is a minimum of $C_{\text{jd}}(\mathbf{U};\mathcal{A})$, then there exists a permutation $\mathbf{P}$ such that $C_{\text{jbd}}(\mathbf{PU};\mathcal{A}) = 0$ (given existence of a JBD). But of course $C_{\text{jd}}(\mathbf{PU};\mathcal{A}) = C_{\text{jd}}(\mathbf{U};\mathcal{A})$, so we will show why (certain) JBD solutions are minima of $C_{\text{jd}}(\mathbf{U};\mathcal{A})$. However, JD might have additional minima. First note that clearly not any JBD minimizes $C_{\text{jd}}(\mathbf{U};\mathcal{A})$, only those such that in each block, $C_{\text{jd}}(\mathbf{U};\mathcal{A})$ when restricted to the block is maximal over $\mathbf{U} \in O(L)$. We will call such a JBD block-optimal in the following.

THEOREM 5.3. *Any block-optimal JBD of $\mathcal{A}$ i.e. $\mathbf{U} \in O(L)$ with $C_{\text{jbd}}(\mathbf{U};\mathcal{A}) = 0$ is a local minimum of $C_{\text{jd}}(\mathbf{V};\mathcal{A})$.*

Proof. The local minimality can be shown using Lagrange-multipliers, but an even simpler (and somewhat sloppy but illustrative) method is to use the explicit parametrization of $O(n)$ by Givens matrices. Consider the infinitesimal, elementary real Givens rotation $\mathbf{G}_{pq}(\epsilon) := \mathbf{G}_r(p, q, \sqrt{1-\epsilon^2}, \epsilon)$ defined for $p < q$ and $0 \le \epsilon < 1$.

Let $\mathbf{U} \in O(n)$ be block-optimal with $C_{\text{jbd}}(\mathbf{U}; \mathcal{A}) = 0$. We have to show that $\mathbf{U}$ is a local minimum of $C_{\text{jd}}(\mathbf{V}; \mathcal{A})$ or equivalently a local maximum of $C_{\text{jd}}(\mathbf{V}; \mathcal{A})$, the sum of the transformed squared diagonals. After substituting each $\mathbf{A}_k$ by $\mathbf{U}\mathbf{A}_k\mathbf{U}^T$, we may already assume that $\mathbf{A}_k$ is m-block diagonal, so we have to show that $\mathbf{U} = \mathbf{I}$ is a maximum of C_{jd}.

The Givens rotations from above can now be used to construct local coordinates of the $d := n(n-1)/2$-dimensional manifold $O(n)$ at $\mathbf{I}$, simply by

$$\begin{aligned} \iota : (-1,1)^d &\longrightarrow O(n) \\ (\epsilon_{12}, \epsilon_{13}, \ldots, \epsilon_{n-1,n}) &\longmapsto \prod_{p<q} \mathbf{G}_{pq}(\epsilon_{pq}). \end{aligned}$$

This is an embedding, and $\iota(0) = \mathbf{I}$, so we only have to show that $h(\epsilon) := g(\iota(\epsilon))$ has a local maximum at $\epsilon = 0$. We do this by considering h partially in each coordinate. Let $p < q$. If p, q are in the same block ($i(p) = i(q)$), then h is locally maximal i.e. positive semi-definite at 0 in the direction ϵ_{pq} because of the assumption that $\mathbf{E} = \mathbf{I}$ is block-optimal

Now assume p and q are from different blocks. After possible permutation, we may assume that $p = q + 1$ so that each matrix $\mathbf{A}_k \in \mathcal{A}$ is of the form

$$\mathbf{A}_k = \begin{pmatrix} \ddots & \vdots & & 0 & \\ \cdots & a_k & 0 & & \\ & 0 & b_k & \cdots & \\ & & & & \\ 0 & & \vdots & \ddots & \end{pmatrix},$$

where a_k is located at index (p, p). Then $\mathbf{G}_{pq}(\epsilon)\mathbf{A}_k\mathbf{G}_{pq}(\epsilon)^T$ equals

$$\begin{pmatrix} \ddots & \vdots & & 0 \\ \cdots & a_k - (a_k - b_k)\epsilon^2 & (a_k - b_k)\epsilon\sqrt{1-\epsilon^2} & \cdots \\ \cdots & (a_k - b_k)\epsilon\sqrt{1-\epsilon^2} & b_k + (a_k - b_k)\epsilon^2 & \cdots \\ 0 & & \vdots & \ddots \end{pmatrix},$$

and entries on the diagonal other than at indices (p, p) and (q, q) are not changed, so

$$\begin{aligned} &\| \operatorname{diag}(\mathbf{G}_{pq}(\epsilon)\mathbf{A}_k\mathbf{G}_{pq}(\epsilon))^T\|^2 - \| \operatorname{diag}(\mathbf{A}_k)\|^2 = \\ &= -2a_k(a_k - b_k)\epsilon^2 + 2b_k(a_k - b_k)\epsilon^2 + 2(a_k - b_k)^2\epsilon^4 \\ &= -2(a_k^2 + b_k^2)\epsilon^2 + 2(a_k - b_k)^2\epsilon^4. \end{aligned}$$

Hence

$$h(0, \ldots, 0, \epsilon_{pq}, 0, \ldots, 0) - h(0) = -c\epsilon_{pq}^2 + d\epsilon_{pq}^4$$

with $c = 2\sum_{k=1}^K (a_k^2 + b_k^2)$ and $d = 2\sum_{k=1}^K (a_k - b_k)^2$. Now either $c = 0$, then also $d = 0$ and h is constant zero in the direction ϵ_{pq}. Or, more interestingly, $c \neq 0$, then $c > 0$ and therefore h is negative definite in the direction ϵ_{pq}.

Altogether we get a negative definite h at 0 except for 'trivial directions', and hence a local maximum at 0. $\square$

5.4. Simulations. The employed algorithms as well as some of the following examples are freely available for download at `http://www.biologie.uni-regensburg.de/Biophysik/Theis/researchjbd.html`. The programs have been realized in MATLAB, and sufficient documentation is given to reproduce the results and extend the algorithms.

As for the diagonalization cases, the convergence of the proposed (joint) block-diagonalization scheme is by construction guaranteed, whatever the chosen strategy for the selection of the couples (p, q). If convergence to the global minimum was in practice usually observed for the above diagonalization schemes, this is certainly not the case for block-diagonalization. To illustrate this point we have tested 3 strategies for the choice of rotations, applied to the JBD of real matrices in a real basis.

(M1) The first method is inspired from the Cyclic Jacobi approach of Algorithm 2, except for the fact that the couples (p, q) are chosen out of the diagonal blocks. The algorithm is initialized with the identity matrix, i.e $\mathbf{U} = \mathbf{I}_n$. The algorithm is stopped when all the values of $s^{\star}_{\mathrm{jbdr}} = \sin \theta^{\star}_{\mathrm{jbdr}}$ are lower than 10^{-4} within a sweep.

(M2) The second method is identical to (M1) except for the fact that the algorithm is initialized with the matrix $\mathbf{U}_{\mathrm{jdr}}$ provided by joint diagonalization of $\mathcal{A}$.

(M3) The third method is inspired from Classical Jacobi (Algorithm 1) and consists of choosing at each iteration the couple (p, q) ensuring a maximum decrease of criterion C_{jbd}. This requires computing all the differences $|\sum_{k=1}^{K} \mathrm{boff}\,(\mathbf{B}_k) - \mathrm{boff}\,(\mathbf{A}_k)\,|$ for all couples (p, q) and to pick up the couple which yields the largest difference value. The algorithm stops when 20 successive value of $s^{\star}_{\mathrm{jbdr}}$ are all lower than 10^{-4}.

The three methods are applied to 100 random draws of K real matrices exactly block-diagonalizable in a real common orthonormal basis. Various values of L (size of the blocks), m (number of blocks) and K (number of matrices) are considered. The number of failures over the 100 realizations (i.e, the number of times the methods do not converge to a solution such that $C_{\mathrm{jbdr}} = 0$) is reported in Table 5.1.

5.5. Discussion. The previous results emphasize the importance of the initialization and the choice of the rotations. Failure rates of (M1) are very high, in particular when m and L increase. (M2) and (M3), which are both initialized by joint-diagonalization, give much better results, with (M3) being in nearly every case more reliable than (M2). However, none of the two methods systematically converge to a global minimum of C_{jbdr} when $m \geq 3$, and, interestingly, the methods do not usually fail on the same sets of data. Also, Fig. 5.1 and Fig. 5.2 show that (M3) only need a few iterations after JD to minimize C_{jbdr}.

Indeed, and this indicates the validity of the claim from Section 5.3, JD minimizes the joint block-off-diagonality C_{jbdr}, however only up to a permutation. And in the above simulation, the permutation is then discovered by application of the JBD algorithm — this also explains why in Figures 5.1 and 5.2, the cost function after JD only decreases in discrete steps, corresponding to identified permutations of one block.

6. Conclusions. After reviewing existence and uniqueness results as well as algorithms for diagonalization and joint diagonalization, we proposed extensions to the setting of (joint) block-diagonalization, a problem of considerable importance for instance in the field of blind signal processing. The novel theoretical contributions are partial existence results as well as the perhaps astonishing relation from theorem 5.3

m	2														
L	2					4					6				
K	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
M1	1	4	4	1	2	32	33	25	10	11	55	33	21	24	16
M2	0	0	0	0	0	11	1	0	0	0	43	2	0	0	0
M3	0	0	0	0	0	5	0	0	0	0	14	0	0	0	0
m	3														
L	2					4					6				
K	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
M1	3	14	11	18	8	68	54	38	33	32	84	60	48	51	52
M2	0	0	0	0	0	29	5	1	2	0	53	10	8	7	8
M3	0	0	0	0	0	15	1	0	3	1	44	0	0	2	8
m	4														
L	2					4					6				
K	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
M1	5	30	21	19	16	87	75	68	60	59	99	83	77	77	75
M2	0	0	0	0	0	47	7	6	4	2	88	15	8	4	10
M3	0	0	0	0	0	21	5	4	2	3	65	8	2	0	5

TABLE 5.1

Number of failures of methods M1, M2 and M3 over 100 random realizations of K matrices exactly block-diagonalizable in a common orthonormal basis.

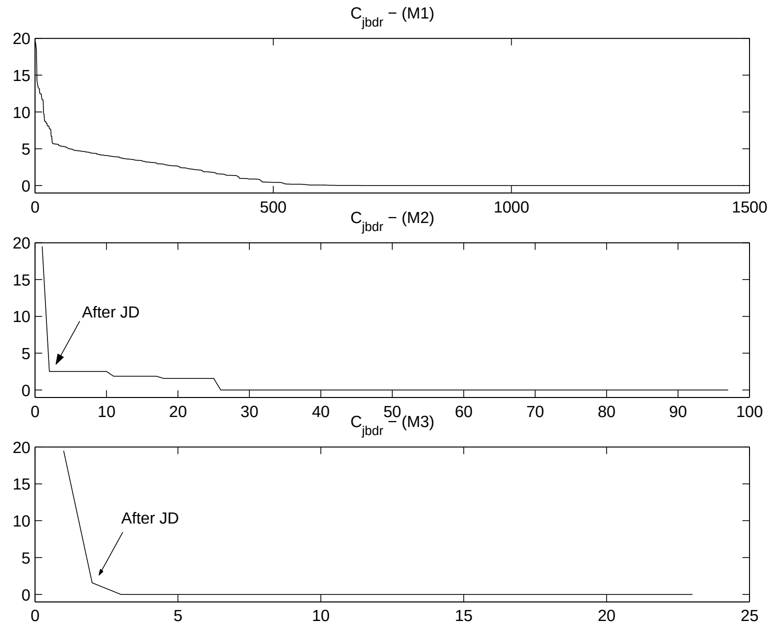

FIG. 5.1. *Evolution of criterion C_{jbdr} for a random set $\mathcal{A}$ such that $m = 3$, $L = 4$, $K = 3$. Using a 1.25 GHz Powerbook G4, the computation times for this particular dataset are: (M1 - 5.6s), (M2 - 1.7s), (M3 - 4.3s). The three methods succeed in minimizing the criterion.*

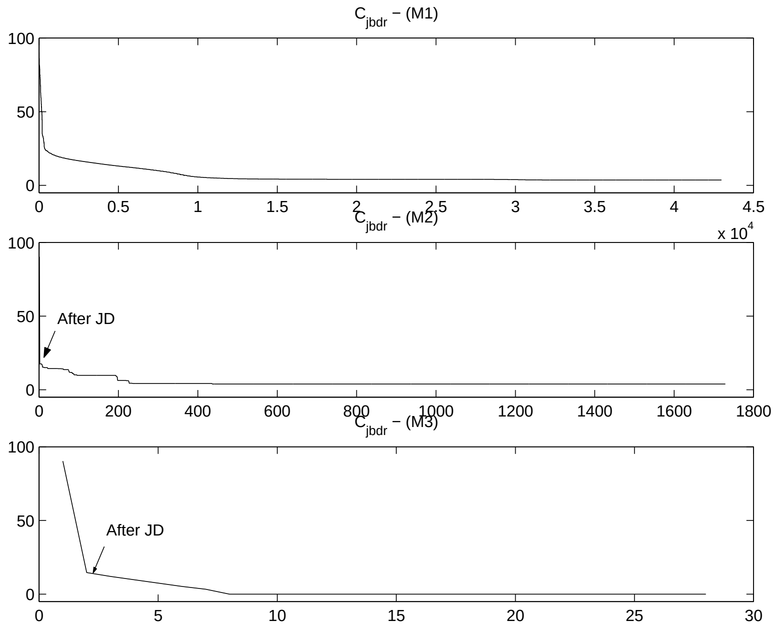

FIG. 5.2. *Evolution of criterion C_{jbdr} for a random set $\mathcal{A}$ such that $m = 4$, $L = 6$, $K = 3$. Using a 1.25 GHz Powerbook G4, the computation times for this particular dataset are: (M1 - 186s), (M2 - 10.2s), (M3 - 16.1s). Only (M3) succeeds in minimizing the criterion.*

between the well-known joint diagonalization problem and the similar task of joint block-diagonalization.

The main algorithmic conclusion of this report is: Jacobi algorithms for joint block-diagonalizaton bring up convergence problems that do not occur in joint diagonalization and that still need to be properly addressed. However we proposed a strategy (method (M3)) which consideraby reduces the failure rates of the straightforward approach (M1). The fact that lower failure rates are obtained with (M2) and (M3), which are initialized with joint-diagonalization, tend to corroborate the conjecture that JBD diagonalization could be achieved up to an arbitrary permutation of columns via JD [1, 8], but it still does not explain why this permutation cannot be solved by minimization of C_{jbdr}.

Acknowledgments. The authors thank Harold Gutch for the careful proofreading of the manuscript.

Appendix A. Expression of the polynomial coefficients. Below are the coefficients of the fourth-order polynomial that needs to be rooted at each iteration of the proposed method for joint block-diagonalization of a set of real matrices having a common real orthonormal basis. The particular block-diagonalization case of only one matrix simply corresponds to $K = 1$.

$$
\begin{aligned}
q_{40} =& \sum_{k=1}^{K} \left\{ a_{kpp}^2 + a_{kqq}^2 + \sum_{j \in I_{i(p)}, j \neq p} a_{kpj}^2 + a_{kjp}^2 + \sum_{j \in I_{i(q)}, j \neq q} a_{kqj}^2 + a_{kjq}^2 \right\} \\
q_{04} =& \sum_{k=1}^{K} \left\{ a_{kpp}^2 + a_{kqq}^2 + \sum_{j \in I_{i(p)}, j \neq p} a_{kqj}^2 + a_{kjq}^2 + \sum_{j \in I_{i(q)}, j \neq q} a_{kpj}^2 + a_{kjp}^2 \right\} \\
q_{31} =& 2 \sum_{k=1}^{K} \{ (a_{kpp} - a_{kqq}) \, (a_{kpq} + a_{kqp}) \\
& + \sum_{j \in I_{i(p)}, j \neq p} a_{kpj} \, a_{kqj} + a_{kjp} \, a_{kjq} - \sum_{j \in I_{i(q)}, j \neq q} a_{kpj} \, a_{kqj} + a_{kjp} \, a_{kjq} \} \\
q_{13} =& 2 \sum_{k=1}^{K} \{ (a_{kqq} - a_{kpp}) \, (a_{kpq} + a_{kqp}) \\
& + \sum_{j \in I_{i(p)}, j \neq p} a_{kpj} \, a_{kqj} + a_{kjp} \, a_{kjq} - \sum_{j \in I_{i(q)}, j \neq q} a_{kpj} \, a_{kqj} + a_{kjp} \, a_{kjq} \} \\
q_{22} =& \sum_{k=1}^{K} \{ 2 \, (a_{kpq} + a_{kqp})^2 + 4 \, a_{kpp} \, a_{kqq} \\
& + \sum_{j \in I_{i(p)}, j \neq p} (a_{kpj}^2 + a_{kqj}^2) + (a_{kjp}^2 + a_{kjq}^2) \\
& + \sum_{j \in I_{i(q)}, j \neq q} (a_{kpj}^2 + a_{kqj}^2) + (a_{kjp}^2 + a_{kjq}^2) \}
\end{aligned}
$$

REFERENCES

[1] K. Abed-Meraim and A. Belouchrani, *Algorithms for joint block diagonalization*, in Proc. EUSIPCO'04, Vienna, Austria, 2004, pp. 209–212.

[2] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and É. Moulines, *A blind source separation technique based on second order statistics*, IEEE Trans. Signal Processing, 45 (1997), pp. 434–444.

[3] A. Belouchrani, K. Abed-Meraim, and Y. Hua, *Jacobi-like algorithms for joint block diagonalization: Application to source localization*, in Proc. International Symposium on Intelligent Signal Processing and Communication Systems, Nov. 1998.

[4] A. Belouchrani, M. G. Amin, and K. Abed-Meraim, *Direction finding in correlated noise fields based on joint block-diagonalization of spatio-temporal correlation matrices*, IEEE Signal Processing Letters, 4 (1997), pp. 266–268.

[5] H. Bousbiah-Salah, A. Belouchrani, and K. Abed-Meraim, *Jacobi-like algorithm for blind signal separation of convolutive mixtures*, Electronics Letters, 37 (2001), pp. 1049–1050.

[6] A. Bunse-Gerstner, R. Byers, and V. Mehrmann, *Numerical methods for simultaneous diagonalization*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 927–949.

[7] J.-F. Cardoso, *Home page*. `http://www.tsi.enst.fr/~cardoso`.

[8] ———, *Multidimensional independent component analysis*, in Proc. ICASSP, 1998.

[9] J.-F. Cardoso and A. Souloumiac, *Blind beamforming for non Gaussian signals*, IEE Proceedings-F, 140 (1993), pp. 362–370.

[10] ———, *Jacobi angles for simultaneous diagonalization*, SIAM J. Mat. Anal. Appl., 17 (1996), pp. 161–164.

[11] C. Févotte and C. Doncarli, *A unified presentation of blind source separation methods for convolutive mixtures using block-diagonalization*, in Proc. 4$^{\text{th}}$ Symposium on Independent Component Analysis and Blind Source Separation (ICA'03), Nara, Japan, Apr. 2003.

[12] ———, *Two contributions to blind source separation using time-frequency distributions*, IEEE Signal Processing Letters, 11 (2004).

[13] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, third edition ed., 1996.

[14] L. De Lathauwer, D. Callaerts, B. De Moor, and J. Vandewalle, *Fetal electrocardiogram extraction by source subspace separation*, in Proc. IEEE Signal Processing / ATHOS Workshop on Higher-Order Statistics, Jun. 1995, pp. 134–138.

[15] L. De Lathauwer, C. Févotte, B. De Moor, and J. Vandewalle, *Jacobi algorithm for joint block diagonalization in blind identification*, in Proc. 23$^{\text{th}}$ Symposium on Information Theory in the Benelux, Louvain-la-Neuve, Belgium, Mai 2002, pp. 155–162.

[16] F.J. Theis, *Blind signal separation into groups of dependent signals using joint block diagonalization*, in Proc. ISCAS 2005, Kobe, Japan, 2005, pp. 5878–5881.

[17] ———, *Multidimensional independent component analysis using characteristic functions*, in Proc. EUSIPCO 2005, Antalya, Turkey, 2005.

[18] A. Ziehe and K.-R. Mueller, *TDSEP – an efficient algorithm for blind separation using time structure*, in Proc. of ICANN'98, L. Niklasson, M. Bodén, and T. Ziemke, eds., Skövde, Sweden, 1998, Springer Verlag, Berlin, pp. 675–680.

Dépôt légal : 2007 – 2ème^{r} trimestre
Imprimé à l'Ecole Nationale Supérieure des Télécommunications – Paris
ISSN 0751-1345 ENST D (Paris) (France 1983-9999)

Ecole Nationale Supérieure des Télécommunications
Groupe des Ecoles des Télécommunications - membre de ParisTech
46, rue Barrault - 75634 Paris Cedex 13 - Tél. + 33 (0)1 45 81 77 77 - www.enst.fr
Département TSI