16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

GRADIENT BASED APPROXIMATE JOINT DIAGONALIZATION BY ORTHOGONAL TRANSFORMS

Mikael Sørensen, Sylvie Icart, Pierre Comon and Luc Deneire

Laboratoire I3S, CNRS/UNSA
Les Algorithmes - Euclide-B, 06903 Sophia Antipolis, France
phone: +33-4 9294 2751
{sorensen, icart, pcomon, deneire}@i3s.unice.fr

ABSTRACT

Approximate Joint Diagonalization (AJD) of a set of symmetric matrices by an orthogonal transform is a popular problem in Blind Source Separation (BSS). In this paper we propose a gradient based algorithm which maximizes the sum of squares of diagonal entries of all the transformed symmetric matrices. Our main contribution is to transform the orthogonality constrained optimization problem into an unconstrained problem. This transform is performed in two steps: First by parameterizing the orthogonal transform matrix by the matrix exponential of a skew-symmetric matrix. Second, by introducing an isomorphism between the vector space of skew-symmetric matrices and the Euclidean vector space of appropriate dimension. This transform is then applied to a gradient based algorithm called GAEX to perform joint diagonalization of a set of symmetric matrices.

1. INTRODUCTION

Joint diagonalization of a set of symmetric matrices is a popular problem in BSS [2], [3], [5], [18].

A necessary and sufficient condition for the existence of an orthogonal matrix that will diagonalize all matrices in a finite set of symmetric matrices is that all the matrices contained in the set must mutually commute with each other [10], and hence only an approximate solution to the joint diagonalization problem can in general be obtained.

One of the first AJD algorithms to handle a set of symmetric matrices was proposed by De Leeuw in [6]. Since there in general doesn't exist an analytical solution to this problem a Jacobi procedure was proposed to numerically solve the problem. Another Jacobi procedure to this problem known as JADE, was later proposed by Cardoso in [3], [4].

Another way to try impose orthogonality of the transform matrix is by the penalty method and Joho proposed a gradient method in [7] to diagonalize a set of symmetric matrices by the aid of the penalty method. Furthermore Joho proposed a gradient and a Newton method in [8] to diagonalize a set of symmetric matrices by applying the projection based method proposed by Manton in [12] which projects every update to the manifold of orthogonal matrices.

Afsari [1] and Tanaka [16] proposed gradient based methods by exploiting the geometry of matrix Lie groups. Furthermore Yamada proposed in [17] to apply the Cayley transform as a global parameterization of the set of orthogonal matrices with positive determinant and eigenvalues different from minus one.

The approach taken in this paper is also to make use of a proper parameterization of a set orthogonal matrices which will transform the constrained optimization problem into an unconstrained problem by the use of appropriate transforms so that the transform matrix will remain orthogonal after each update. The difference between the approach in [17] and the one presented in this paper is that a local parameterization will also be considered here and secondly the matrix exponential will be applied as a parameterization of the orthogonal matrices. Before presenting a gradient based AJD algorithm some notation used throughout the paper and the problem formulation will be presented.

1.1 Notations

Let $\mathbb{R}$ and $\mathbb{R}_+$ denote the set of real and nonnegative real numbers respectively. Furthermore let $\mathrm{SO}_m(\mathbb{R})$, $\mathrm{S}_m(\mathbb{R})$ and $\mathrm{S}_m^{\perp}(\mathbb{R})$ denote the set of $m \times m$ orthogonal matrices with determinant equal to one, the symmetric matrices and skew-symmetric matrices respectively. Moreover let $(\cdot)^T$, $(\cdot)^H$, $\mathrm{Vec}(\cdot)$, $\mathrm{Unvec}(\cdot)$, $\otimes$ and $\|\cdot\|$ denote the transpose, conjugate-transpose, matrix vectorization operator, inverse matrix vectorization operator, the Kronecker product of two matrices and the Frobenius norm of a matrix respectively. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$, then let $\mathbf{A}_{ij}$ denote the ith row-jth column entry of $\mathbf{A}$ and let $\mathbf{e}_i \in \mathbb{R}^m$ denote the unit vector with one in the entry i and zero elsewhere. Let δ_{ij} be the Dirac delta function which is zero except when the two indices are equal, in which case $\delta_{ii} = 1$.

1.2 Problem Formulation

Given a finite set of symmetric matrices $\{\mathbf{T}_k\}_{k=1}^n \subset \mathrm{S}_m(\mathbb{R})$, then let the transform matrix be $\mathbf{U} \in \mathrm{SO}_m(\mathbb{R})$ and the transformed matrices be $\mathbf{K}_k = \mathbf{U}\mathbf{T}_k\mathbf{U}^T \forall k \in \{1,\ldots,n\}$, then the matrix $\mathbf{U}$ which would imply that $\mathbf{K}_k$ is as diagonal as possible for every $k \in \{1,\ldots,n\}$ is sought.

A measure on how diagonal the set of transformed matrices are is given by the functional $\Psi : \mathrm{SO}_m(\mathbb{R}) \to \mathbb{R}$ given by

$$\Psi(\mathbf{U}) \quad = \quad \frac{1}{2}\sum_{k=1}^{n} \|\mathbf{U}\mathbf{T}_k\mathbf{U}^T - \mathrm{diag}\left(\mathbf{U}\mathbf{T}_k\mathbf{U}^T\right)\|^2, \qquad (1)$$

THE WORK OF MIKAEL SØRENSEN IS SUPPORTED BY THE EU BY A MARIE-CURIE FELLOWSHIP (EST-SIGNAL PROGRAM : HTTP://EST-SIGNAL.I3S.UNICE.FR) UNDER CONTRACT NO MEST-CT-2005-021175.

where $\text{diag}(\mathbf{A}) \in \mathbb{R}^{m\times m}$ and $\text{diag}(\mathbf{A})_{ij} = \mathbf{A}_{ij}\delta_{ij}$.

Due to the invariance property of the Frobenius norm wrt. any orthogonal operator, minimizing (1) is equivalent to maximizing the functional $J : \text{SO}_m(\mathbb{R}) \to \mathbb{R}$ given as

$$J(\mathbf{U}) \quad = \quad \frac{1}{2}\sum_{k=1}^{n} \|\text{diag}\left(\mathbf{U}\mathbf{T}_k\mathbf{U}^T\right)\|^2.$$

A gradient based solution to the constrained maximization problem

$$\arg\max_{\mathbf{U}\in\text{SO}_m(\mathbb{R})} J(\mathbf{U}) \tag{2}$$

will be proposed in this paper.

2. GRADIENT BASED AJD

2.1 Parameterization and Update of an Orthogonal Matrix

It is well-known that the orthogonality property of the transform matrix can be preserved by parameterizing them by the matrix exponential $e^{\mathbf{X}} \triangleq \sum_{n=0}^{\infty} \frac{\mathbf{X}^n}{n!}$. Moreover since $\text{SO}_m(\mathbb{R})$ is a multiplicative group and the mapping $\text{e} : \text{S}_m^{\perp}(\mathbb{R}) \to \text{SO}_m(\mathbb{R})$ given by $\text{e}^{\mathbf{X}}$ is surjective [9], the following local parameterization of the transform matrix will ensure that the transform matrix will remain orthogonal after each update:

$$\mathbf{U}^{(n+1)} \quad = \quad \mathbf{U}^{(n)}\text{e}^{\mu\widetilde{\mathbf{U}}},$$

where $\widetilde{\mathbf{U}} \in \text{S}_m^{\perp}(\mathbb{R})$, $\mathbf{U}^{(n)} \in \text{SO}_m(\mathbb{R})$, $\mu \in \mathbb{R}_+$ and the upper index denotes the iteration number. Alternatively the following global parameterization could be used:

$$\mathbf{U}^{(n+1)} \quad = \quad \text{e}^{\widetilde{\mathbf{U}}^{(n)}+\mu\widetilde{\mathbf{U}}},$$

where $\widetilde{\mathbf{U}}^{(n)}, \widetilde{\mathbf{U}} \in \text{S}_m^{\perp}(\mathbb{R})$, $\mu \in \mathbb{R}_+$ and the upper index denotes the iteration number. By the use of the matrix exponential the problem has turned from finding an orthogonal matrix to finding a skew-symmetric matrix. In general the computation of the matrix exponential must be considered computationally demanding, but here only the computation of the matrix exponential wrt. a skew-symmetric matrix is required and this can for instance be done by an EigenValue Decomposition (EVD).

The following subsection will show that it is possible to turn the constrained problem (2) into an unconstrained problem by the use of a second transform introduced in the following subsection.

2.2 Isomorphism between $\mathbf{S}_n^{\perp}(\mathbb{R})$ and $\mathbb{R}^{\frac{n(n-1)}{2}}$

In [14] an explicit expression of an isomorphism between $\text{S}_n(\mathbb{R})$ and $\mathbb{R}^{\frac{(n+1)n}{2}}$ was derived. By a similar procedure, an explicit expression of an isomorphism between $\text{S}_n^{\perp}(\mathbb{R})$ and $\mathbb{R}^{\frac{(n-1)n}{2}}$ can be found.

Let $\mathbf{A} \in \text{S}_n^{\perp}(\mathbb{R})$ and let $w(\mathbf{A})$ denote the $\frac{(n-1)n}{2} \times 1$ vector that is obtained by eliminating all the diagonal and supradiagonal elements from $\mathbf{A}$ and thereafter stacking the infradiagonal elements into the vector

$$w(\mathbf{A}) = \left[\mathbf{A}_{12}, \cdots, \mathbf{A}_{1n}, \mathbf{A}_{23}, \cdots, \mathbf{A}_{2n}, \cdots, \mathbf{A}_{(n-1)n}\right]^T.$$

Definition 2.1. *(Skew-Symmetric Elimination Matrix)*

The skew-symmetric elimination matrix $\widetilde{\boldsymbol{L}} \in \mathbb{R}^{\frac{(n-1)n}{2}\times n^2}$ is defined to be the matrix representation of the linear mapping $\widetilde{L} : S_n^{\perp}(\mathbb{R}) \to \mathbb{R}^{\frac{(n-1)n}{2}}$ and it satisfies $\widetilde{\boldsymbol{L}}Vec(\boldsymbol{A}) = w(\boldsymbol{A})\forall \boldsymbol{A} \in S_n^{\perp}(\mathbb{R})$.

Definition 2.2. *(Inverse Skew-Symmetric Elimination Matrix)*

Let the inverse skew-symmetric elimination matrix $\widetilde{\boldsymbol{L}}^{-1} \in \mathbb{R}^{n^2\times\frac{(n-1)n}{2}}$ be defined to be the matrix representation of the linear mapping $\widetilde{L}^{-1} : \mathbb{R}^{\frac{(n-1)n}{2}} \to S_n^{\perp}(\mathbb{R})$ and it satisfies $Vec(\boldsymbol{A}) = \widetilde{\boldsymbol{L}}^{-1} w(\boldsymbol{A})\, \forall \boldsymbol{A} \in S_n^{\perp}(\mathbb{R})$.

Before giving an explicit expression of the skew-symmetric elimination matrix, let

$$\mathbf{I}_{\frac{(n-1)n}{2}} = \left[\tilde{\mathbf{u}}_{21}, \tilde{\mathbf{u}}_{31}, \ldots, \tilde{\mathbf{u}}_{n1}, \tilde{\mathbf{u}}_{32}, \ldots, \tilde{\mathbf{u}}_{n2}, \tilde{\mathbf{u}}_{43}, \ldots, \tilde{\mathbf{u}}_{n(n-1)}\right],$$

where $\mathbf{I}_m$ is the $m \times m$ identity matrix and and $\tilde{\mathbf{u}}_{ij} = \mathbf{e}_{(j-1)n+i-\frac{1}{2}j(j+1)} \in \mathbb{R}^{\frac{(n-1)n}{2}}$.

Proposition 2.3. *Let $\tilde{\boldsymbol{u}}_{ij} = \boldsymbol{e}_{(j-1)n+i-\frac{1}{2}j(j+1)} \in \mathbb{R}^{\frac{(n-1)n}{2}}$ and let $\mathbb{R}^{n\times n} \ni \boldsymbol{E}_{ij} = \boldsymbol{e}_i\boldsymbol{e}_j^T$ then $\widetilde{\boldsymbol{L}} = \sum_{i>j} \tilde{\boldsymbol{u}}_{ij} Vec\left(\boldsymbol{E}_{ij}\right)^T = \sum_{i>j} \tilde{\boldsymbol{u}}_{ij} \otimes \boldsymbol{e}_j^T \otimes \boldsymbol{e}_i^T$.*

Proof. By making use of the identities $\text{Vec}\left(\mathbf{y}\mathbf{x}^T\right) = \mathbf{y}\otimes\mathbf{x}$ and $\mathbf{x}\mathbf{y}^T = \mathbf{x}\otimes\mathbf{y}^T$ we get

$$\begin{aligned} w(\mathbf{A}) &= \sum_{i>j} \mathbf{A}_{ij}\tilde{\mathbf{u}}_{ij} \\ &= \sum_{i>j} \tilde{\mathbf{u}}_{ij}\mathbf{e}_i^T\mathbf{A}\mathbf{e}_j \\ &= \sum_{i>j} \tilde{\mathbf{u}}_{ij}\text{Vec}\left(\mathbf{E}_{ij}\right)^T \text{Vec}(\mathbf{A}) \\ &= \sum_{i>j} \left(\tilde{\mathbf{u}}_{ij}\otimes\mathbf{e}_j^T\otimes\mathbf{e}_i^T\right)\text{Vec}(\mathbf{A}). \end{aligned}$$

□

To obtain the matrix expression of the linear mapping $\widetilde{L}^{-1}$ the following lemma will be used.

Lemma 2.4. *Let $\widetilde{A}$ denote the lower triangular matrix consisting of the infradiagonal elements of $\boldsymbol{A}$ then $\widetilde{\boldsymbol{L}}^T w(\boldsymbol{A}) = Vec\left(\widetilde{\boldsymbol{A}}\right)$.*

Proof. Since $\tilde{\mathbf{u}}_{ij}^T \tilde{\mathbf{u}}_{hk} = \delta_{i-h,j-k}$ we get

$$\begin{aligned}
\widetilde{\mathbf{L}}^T \widetilde{\mathbf{L}} \mathrm{Vec}(\mathbf{A}) &= \sum_{i>j} \sum_{h>k} \left(\tilde{\mathbf{u}}_{ij}^T \otimes \mathbf{e}_j \otimes \mathbf{e}_i\right)\left(\tilde{\mathbf{u}}_{hk} \otimes \mathbf{e}_k^T \otimes \mathbf{e}_h^T\right) \mathrm{Vec}(\mathbf{A}) \\
&= \sum_{i>j} \sum_{h>k} \left(\tilde{\mathbf{u}}_{ij}^T \tilde{\mathbf{u}}_{hk} \otimes \mathbf{e}_j \mathbf{e}_k^T \otimes \mathbf{e}_i \mathbf{e}_h^T\right) \mathrm{Vec}(\mathbf{A}) \\
&= \sum_{i>j} \left(\mathbf{e}_j \mathbf{e}_j^T \otimes \mathbf{e}_i \mathbf{e}_i^T\right) \mathrm{Vec}(\mathbf{A}) \\
&= \sum_{i>j} \mathrm{Vec}\left(\mathbf{e}_i \mathbf{e}_i^T \mathbf{A} \mathbf{e}_j \mathbf{e}_j^T\right) \\
&= \mathrm{Vec}\left(\sum_{i>j} \mathbf{A}_{ij} \mathbf{E}_{ij}\right) \\
&= \mathrm{Vec}\left(\widetilde{\mathbf{A}}\right).
\end{aligned}$$

Hence

$$\begin{aligned}
\mathrm{Vec}\left(\widetilde{\mathbf{A}}\right) &= \widetilde{\mathbf{L}}^T \widetilde{\mathbf{L}} \mathrm{Vec}(\mathbf{A}) \\
&= \widetilde{\mathbf{L}}^T w(\mathbf{A}).
\end{aligned}$$

□

By using the above lemma it is possible to obtain an explicit expression of $\widetilde{L}^{-1}$, but before reaching this result the commutation matrix will be introduced.

Definition 2.5. *(Commutation Matrix [13], [15])*
Let $\mathbb{R}^{m\times n} \ni \boldsymbol{H}_{ij} \triangleq \boldsymbol{e}_{i,m} \boldsymbol{e}_{j,n}^T$, where $\boldsymbol{e}_{i,m}$ is the ith column of $\boldsymbol{I}_m$ and $\boldsymbol{e}_{j,n}$ is the jth column of $\boldsymbol{I}_n$. Then the commutation matrix $\boldsymbol{C}_{mn} \in \mathbb{R}^{mn\times mn}$ is defined to be

$$\boldsymbol{C}_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} \boldsymbol{H}_{ij} \otimes \boldsymbol{H}_{ij}^T.$$

The commutation matrix relates $\mathrm{Vec}(\mathbf{A})$ and $\mathrm{Vec}\left(\mathbf{A}^T\right)$ by the equation $\mathbf{C}_{mn}\mathrm{Vec}(\mathbf{A}) = \mathrm{Vec}\left(\mathbf{A}^T\right) \forall \mathbf{A} \in \mathbb{R}^{m\times n}$ [13].

Proposition 2.6. *The matrix representation of the linear map $\widetilde{L}^{-1}$ is given by the expression $\widetilde{\boldsymbol{L}}^{-1} = \widetilde{\boldsymbol{L}}^T - \boldsymbol{C}_{nn}\widetilde{\boldsymbol{L}}^T$.*

Proof. By lemma 2.4 and the commutation matrix we get

$$\begin{aligned}
\left(\widetilde{\mathbf{L}}^T - \mathbf{C}_{nn}\widetilde{\mathbf{L}}^T\right) w(\mathbf{A}) &= \widetilde{\mathbf{L}}^T w(\mathbf{A}) - \mathbf{C}_{nn}\widetilde{\mathbf{L}}^T w(\mathbf{A}) \\
&= \mathrm{Vec}\left(\widetilde{\mathbf{A}}\right) - \mathrm{Vec}\left(\widetilde{\mathbf{A}}^T\right) \\
&= \mathrm{Vec}(\mathbf{A}) \ \forall \mathbf{A} \in \mathrm{S}_n^{\perp}(\mathbb{R}).
\end{aligned}$$

□

By the use of the previously introduced isomorphism the search for a skew-symmetric matrix has now turned into a search for a vector in $\mathbf{R}^{\frac{(n-1)n}{2}}$. Hence any conventional iterative method operating in a vector space can now be applied to the problem. In particular a gradient ascent method will be described next.

2.3 Calculation of the Gradient

Now a matrix based formulation of the partial derivatives will be given. Let $\mathbf{X} \in \mathbb{R}^{m\times m}$, then we have the identities $\|\mathbf{X}\|^2 = \mathrm{Vec}(\mathbf{X})^T \mathrm{Vec}(\mathbf{X})$ and $\mathrm{Vec}(\mathrm{diag}(\mathbf{X})) = \mathrm{Diag}(\mathrm{Vec}(\mathbf{I}_m))\mathrm{Vec}(\mathbf{X})$, where $\mathrm{Diag}(\mathrm{Vec}(\mathbf{I}_m))$ is a diagonal matrix with $\mathrm{Diag}(\mathrm{Vec}(\mathbf{I}_m))_{ii} = \mathrm{Vec}(\mathbf{I}_m)_i$ and the matrix $\mathbf{P} \triangleq \mathrm{Diag}(\mathrm{Vec}(\mathbf{I}_m))$ is idempotent. By the use of the identities we will derive $\frac{\partial J(\widetilde{\mathbf{U}})}{\partial w(\widetilde{\mathbf{U}})^T}$ from the differential $\mathrm{d}(J(\mathbf{U}))$.

In [11] the differential of the matrix exponential wrt. a symmetric matrix was found. By a slight modification of their derivation the differential of the matrix exponential wrt. any square matrix can be found to be

$$\mathrm{d}\left(\mathrm{e}^{\mathbf{A}}\right) = \int_0^1 \mathrm{e}^{\mathbf{A}t} \mathrm{d}(\mathbf{A}) \mathrm{e}^{\mathbf{A}(1-t)} \mathrm{d}t. \tag{3}$$

Let $\mathbf{U}$ be fixed and let us recall the identity $\mathrm{Vec}(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right)\mathrm{Vec}(\mathbf{B})$, then by applying the introduced isomorphism and by equation (3) we get

$$\begin{aligned}
\mathrm{d}\left(\mathrm{Vec}\left(\mathbf{U}\mathrm{e}^{\widetilde{\mathbf{U}}}\right)\right) &= (\mathbf{I}_m \otimes \mathbf{U}) \mathrm{d}\left(\mathrm{Vec}\left(\mathrm{e}^{\widetilde{\mathbf{U}}}\right)\right) \\
&= (\mathbf{I}_m \otimes \mathbf{U}) \int_0^1 \mathrm{e}^{\widetilde{\mathbf{U}}(t-1)} \otimes \mathrm{e}^{\widetilde{\mathbf{U}}t} \mathrm{d}t \mathrm{Vec}\left(\mathrm{d}\left(\widetilde{\mathbf{U}}\right)\right) \\
&= (\mathbf{I}_m \otimes \mathbf{U}) \int_0^1 \mathrm{e}^{\widetilde{\mathbf{U}}(t-1)} \otimes \mathrm{e}^{\widetilde{\mathbf{U}}t} \mathrm{d}t \widetilde{\mathbf{L}}^{-1} \mathrm{d}\left(w(\widetilde{\mathbf{U}})\right),
\end{aligned} \tag{4}$$

where $\widetilde{\mathbf{L}}^{-1}$ denotes the inverse skew-symmetric elimination matrix.

Let $\mathbf{U}^{(n+1)} = \mathbf{U}^{(n)}\mathrm{e}^{\widetilde{\mathbf{U}}}$ and $\mathbf{K}_k = \mathbf{U}\mathbf{T}_k\mathbf{U}^T$, then the partial derivatives for the gradient ascent method can be found via the following differential and the use of (4) to be

$$\begin{aligned}
\mathrm{d}(J(\mathbf{U})) &= \frac{1}{2}\sum_{k=1}^{n} \mathrm{d}\left(\|\mathrm{diag}\left(\mathbf{U}\mathbf{T}_k\mathbf{U}^T\right)\|^2\right) \\
&= \sum_{k=1}^{n} \mathrm{Vec}(\mathbf{K}_k)^T \mathbf{P}\mathrm{d}\left(\mathrm{Vec}\left(\mathbf{U}\mathbf{T}_k\mathbf{U}^T\right)\right) \\
&= \sum_{k=1}^{n} \mathrm{Vec}(\mathbf{K}_k)^T \mathbf{P}\left((\mathbf{U}\mathbf{T}_k \otimes \mathbf{I}_m) + (\mathbf{I}_m \otimes \mathbf{U}\mathbf{T}_k)\mathbf{C}_{mm}\right) \\
&\times \ \mathrm{d}(\mathrm{Vec}(\mathbf{U})) \\
&= \sum_{k=1}^{n} \mathrm{Vec}(\mathbf{K}_k)^T \mathbf{P}\left((\mathbf{U}\mathbf{T}_k \otimes \mathbf{I}_m) + (\mathbf{I}_m \otimes \mathbf{U}\mathbf{T}_k)\mathbf{C}_{mm}\right) \\
&\times \ (\mathbf{I}_m \otimes \mathbf{U}) \int_0^1 \mathrm{e}^{\widetilde{\mathbf{U}}(t-1)} \otimes \mathrm{e}^{\widetilde{\mathbf{U}}t} \mathrm{d}t \widetilde{\mathbf{L}}^{-1} \mathrm{d}\left(w(\widetilde{\mathbf{U}})\right).
\end{aligned}$$

From the differential we can deduce that the gradient when evaluated at iteration n reduces to

$$\begin{aligned}
\frac{\partial J(\mathbf{U})}{\partial w(\widetilde{\mathbf{U}})^T} &= \sum_{k=1}^{n} \mathrm{Vec}(\mathbf{K}_k)^T \mathbf{P}\left((\mathbf{U}\mathbf{T}_k \otimes \mathbf{I}_m) + (\mathbf{I}_m \otimes \mathbf{U}\mathbf{T}_k)\mathbf{C}_{mm}\right) \\
&\times (\mathbf{I}_m \otimes \mathbf{U}) \widetilde{\mathbf{L}}^{-1},
\end{aligned}$$

since $\mathbf{U}^{(n)} = \mathbf{U}^{(n)}e^{\mathbf{0}}$ at iteration n. For the global parameterization $\mathbf{U}^{(n+1)} = \mathrm{e}^{\widetilde{\mathbf{U}}^{(n)}+\mu\widetilde{\mathbf{U}}}$, the gradient vector can be found by similar calculations to be

$$\frac{\partial J(\mathbf{U})}{\partial w(\widetilde{\mathbf{U}})^T} = \sum_{k=1}^{n} \mathrm{Vec}(\mathbf{K}_k)^T \mathbf{P}((\mathbf{U}\mathbf{T}_k \otimes \mathbf{I}_m) + (\mathbf{I}_m \otimes \mathbf{U}\mathbf{T}_k)\mathbf{C}_{mm}) \times \int_0^1 \mathrm{e}^{\widetilde{\mathbf{U}}(t-1)} \otimes \mathrm{e}^{\widetilde{\mathbf{U}}t} \mathrm{d}t \widetilde{\mathbf{L}}^{-1}.$$

Given the EVD $\widetilde{\mathbf{U}} = \mathbf{R}\Lambda\mathbf{R}^H$ the above integral can be expressed as

$$\int_0^1 \mathrm{e}^{\widetilde{\mathbf{U}}(t-1)} \otimes \mathrm{e}^{\widetilde{\mathbf{U}}t}\mathrm{d}t = (\mathbf{R}\otimes\mathbf{R}) \int_0^1 \mathrm{e}^{\Lambda(t-1)} \otimes \mathrm{e}^{\Lambda t}\mathrm{d}t\, (\mathbf{R}\otimes\mathbf{R})^H.$$

Furthermore the diagonal elements of $\int_0^1 \mathrm{e}^{\Lambda(t-1)} \otimes \mathrm{e}^{\Lambda t}\mathrm{d}t$ can be found to be

$$\begin{aligned}\int_0^1 \mathrm{e}^{\lambda_i(t-1)}\mathrm{e}^{\lambda_j t}\mathrm{d}t &= \int_0^1 \mathrm{e}^{(\lambda_i+\lambda_j)t-\lambda_i}\mathrm{d}t \\ &= \begin{cases} \frac{\mathrm{e}^{\lambda_j}-\mathrm{e}^{-\lambda_i}}{\lambda_i+\lambda_j} & ,\lambda_i \neq -\lambda_j \\ \mathrm{e}^{-\lambda_i} & ,\lambda_i = -\lambda_j \end{cases}\end{aligned}$$

where λ_i is the ith eigenvalue of $\widetilde{\mathbf{U}}$.
Applying the local parameterization, then the search direction is now set to

$$\widetilde{\mathbf{U}} = \mathrm{Unvec}\left(\widetilde{\mathbf{L}}^{-1} \frac{\partial J(\mathbf{U})}{\partial w(\widetilde{\mathbf{U}})^T} \Big/ \left\| \frac{\partial J(\mathbf{U})}{\partial w(\widetilde{\mathbf{U}})^T} \right\| \right).$$

Next a simple and inexact linesearch procedure will be applied to solve the maximization problem

$$\arg\max_{\mu\in\mathbb{R}_+} J\left(\mathbf{U}^{(n-1)}e^{\mu\widetilde{\mathbf{U}}}\right). \tag{5}$$

By selecting μ sufficiently small convergence to a critical point is guaranteed. More specifically let $\omega = \{0, \frac{1}{20}, \frac{2}{20}, \ldots, 1\}$, then

$$\mu^{(n)} = \arg\max_{\mu\in\omega} J\left(\mathbf{U}^{(n-1)}e^{\mu\widetilde{\mathbf{U}}}\right)$$

will be used as an approximation to (5) in the following simulation section. It should be pointed out that a guarantee of convergence to a global maximum is still an open problem.

3. SIMULATION

The performance of the proposed gradient based algorithm, which will be called GAEX, will be compared with the JADE algorithm proposed in [3].

To compare the different orthogonal simultaneous matrix diagonalization algorithms, the performance will be measured on a set of matrices $\{\mathbf{T}_k\}_{k=1}^n \subset \mathbb{R}^{m\times m}$, where $\mathbf{T}_k = \mathbf{U}\mathbf{D}_k\mathbf{U}^T + \beta\mathbf{E}$, where $\mathbf{U}$ is a randomly generated orthogonal matrix, $\mathbf{D}_k$ is a random diagonal matrix,

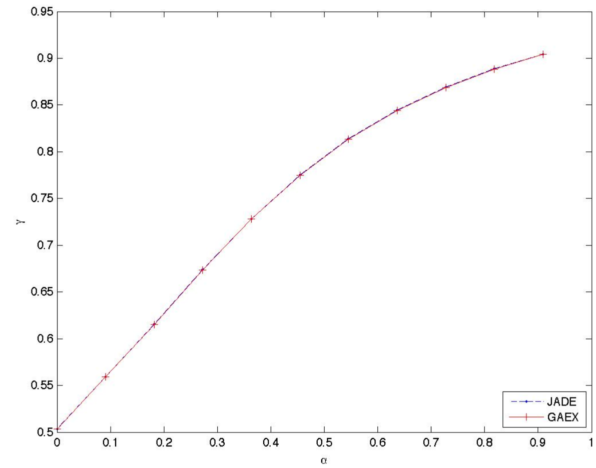

Figure 1: The diagonalization of a set random matrices where α is varying.

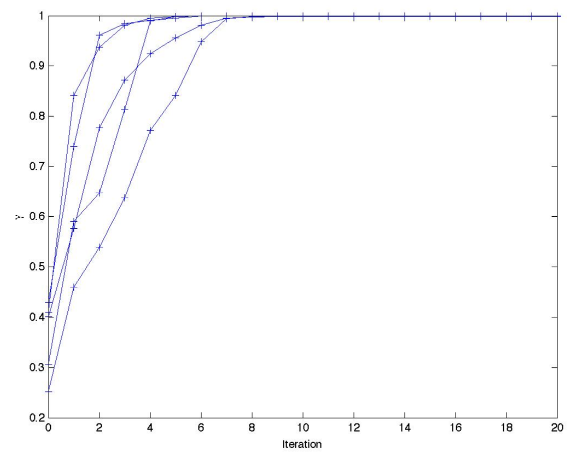

Figure 2: The convergence of the GAEX algorithm for different trials when $\beta = 0$.

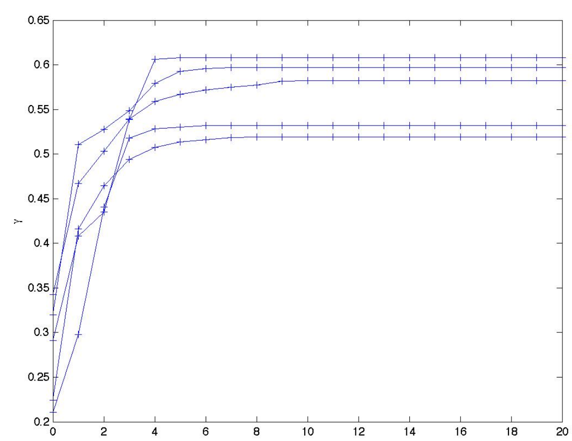

Figure 3: The convergence of the GAEX algorithm for different trials when $\alpha = 0.5$.

$\mathbf{E} \in \mathbb{R}^{m\times m}$ is a random matrix and $\beta \in \mathbb{R}$ is a gain factor. Let x be uniformly distributed

$$\mathrm{U}_{(a,b)}(x) \sim \begin{cases} \frac{1}{b-a} & ,a < x < b \\ 0 & ,\text{elsewhere} \end{cases}$$

Then let $\mathbf{D}_{k_{ii}}$ and $\mathbf{E}_{ij}$ be randomly drawn elements in $\mathrm{U}_{(-100,100)}(x)$ and $\mathbf{U}$ is an orthogonal basis of the subspace spanned by the columns of random matrices contained in $\mathbb{R}^{m\times m}$ and which entries are randomly drawn elements in $\mathrm{U}_{(-100,100)}(x)$. Moreover in all simulations $m = n = 5$ and the proposed gradient based method will be initialized with $\mathbf{U}^{(0)} = \mathbf{I}_m$.

A measure on how diagonal a set of matrices is, is the following

$$\gamma = \frac{\sum_{k=1}^{n} \|\mathrm{diag}\,(\mathbf{T}_k)\|^2}{\sum_{k=1}^{n} \|\mathbf{T}_k\|^2}.$$

This is indeed a measure on how diagonal the set of matrices is, since $0 \le \gamma \le 1$ and when orthogonal transforms are applied then γ decreases continuously as $\mathbf{T}_k$ deviates continuously from a diagonal form.

Let $\alpha = \frac{\sum_{k=1}^{n} \|\mathbf{D}_k\|^2}{\|\beta \mathbf{E}\|^2}$ and let α vary from 0 to 0.9 with a hop size of 0.1, then a comparison between the mentioned procedures as a function of α can be seen on figure 1.

Furthermore the convergence of the GAEX algorithm for different trials as a function of the iteration number when $\beta = 0$ and $\alpha = 0.5$ can be seen on figure 2 and figure 3 respectively.

4. SUMMARY

We have proposed a joint diagonalization algorithm which transforms the constrained optimization problem into an unconstrained problem. This means that any conventional iterative method operating in a vector space could be applied to the problem.

An explicit expression for an isomorphism between $\mathrm{S}_n^{\perp}(\mathbb{R})$ and $\mathbb{R}^{\frac{n(n-1)}{2}}$ was introduced. Furthermore expressions of the first order derivatives of the proposed method was also presented.

Simulations indicate that the gradient based method converges to the same point as the JADE algorithm does. Hence a potential application of introduced method is in the field of adaptive BSS and this topic is left open for future investigation.

REFERENCES

[1] B. Afsari and P.S.Krishnaprasad "Some Gradient Based Joint Diagonalization Methods for ICA," in *Proceedings of the 5th International Conference on Independent Component Analysis and Blind Source Separation*, Granada, Spain, September 22-24, 2004, pp.437–444.

[2] A. Belouchrani and K. Abed Meraim and J.F. Cardoso and E. Moulines, "A blind source separation technique based on second order statistics," *IEEE Trans. Signal Process.*, vol. 45, pp. 434–444. 1997.

[3] J. F. Cardoso and A. Souloumiac, "Blind Beamforming for non-Gaussian signals," *IEE Proceedings-F*, vol. 140, pp. 362–370. 1993.

[4] J. F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Mat. Anal. Appl.*, vol. 17, pp. 161–164. 1996.

[5] A. Cichocki and S-I. Amari, *Adaptive Blind Signal and Image Processing*. John Wiley & Sons, New York, 2002.

[6] J. De Leeuw, "A New Computational Method to Fit the Weighted Euclidian Distance Model," *Psychometrika*, vol. 43, pp. 479–489. 1978.

[7] M. Joho and K. Rahbar, "Joint diagonalization of correlation matrices by using a Gradient methods with application to blind signal separation," in *IEEE Proc. SAM*, Rosslyn, VA, August 4-6, 2002 pp. 273–277.

[8] M. Joho and H. Mathis, "Joint diagonalization of correlation matrices by using a Newton methods with application to blind signal separation," in *IEEE Proc. SAM*, Rosslyn, VA, August 4-6, 2002 pp. 403-407.

[9] J. Gallier, *Geometric Methods and Application: for computer science and engineering*. Springer, New York, 2000.

[10] R. A. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, New York, 1985.

[11] O. Linton and R. McCorie, "Differentation of an Exponential Matrix Function," *Econometric Theory*, vol. 11, pp.1182–1185. 1995.

[12] J. Manton, "Optimization Algorithms Exploiting Unitary Constraints," *IEEE Trans. Signal Process.*, vol. 50, pp. 635–650. 2002.

[13] J.R. Magnus and H. Neudecker, "The Commutation Matrix: Some Properties and Applications," *The Annals of Statistics*, vol. 7, pp. 381–394. 1979.

[14] J.R. Magnus and H. Neudecker, "The Elimination Matrix: Some Lemmas and Applications," *SIAM J. Alg. Disc. Meth.*, vol. 1, pp. 422–449. 1980.

[15] H. Neudecker and T. Wansbeck, "Some Results on Commutation Matrices, with Statistical Applications," *The Canadian Journal of Statistics*, vol.11, pp. 221–231. 1983.

[16] T. Tanaka and S. Fiori, "Least Squares Approximate Joint Diagonalization on the Orthogonal Group," in *ICASSP*, Honolulu, Hawaii, USA, April 15-20, 2007, pp. 649–652.

[17] I. Yamada and T. Ezaki "An Orthogonal Matrix Optimization by Dual Cayley Parameterization Technique, " in *Proceedings of the 4th International Conference on Independent Component Analysis and Blind Source Separation*, April 1 4, 2003, Nara, Japan, 2003, pp. 35–40.

[18] A. Ziehe and K. R. Müller, "TDSEP - an efficient algorithm for blind separation using time structure," in *Proc. Int. Conf. on Artificial Neural Networks*, September 2-4, 1998, Skövde, Sweden, pp. 273–277.