4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), April 2003, Nara, Japan

CRITICAL POINT ANALYSIS OF JOINT DIAGONALIZATION CRITERIA

Gen Hori † *and Jonathan H. Manton* ‡

†Brain Science Institute, RIKEN, Saitama 351-0198, Japan
hori@bsp.brain.riken.go.jp

‡Department of Electrical and Electronic Engineering,
University of Melbourne, Victoria, 3010, Australia
jon@ee.mu.oz.au

ABSTRACT

The stability and sensitivity of joint diagonalization criteria are analyzed using the Hessian of the criteria at their critical points. The sensitivity of some known joint diagonalization criteria is shown to be weak when they are applied to the matrices with closely placed eigenvalues. In such a situation, a large deviation in the diagonalizer cause a slight change in the criterion which makes the estimation of the diagonalizer difficult. To overcome the problem, a new joint diagonalization criterion is introduced. The criterion is linear with respect to the target matrices and has ability to cancel the effect of closely placed eigenvalues.

1. INTRODUCTION

Joint diagonalization is a problem of finding a matrix which similarly transform a set of given matrices as diagonal as possible simultaneously. It has applications in independent component analysis where typical targets are the fourth order cumulant matrices[3] and the delayed correlation matrices[7] of prewhitened observed signals as well as in other research areas such as structured eigenvalue problem[2]. A joint diagonalization algorithm is a combination of a joint diagonality criterion and an optimization method used for minimizing the criterion. A joint diagonality criterion is a function defined on the space of the diagonalizer which takes its minimum when the given matrices are jointly diagonalized. The present paper surveys three joint diagonality criteria in Section 2. Typical optimization methods applied to joint diagonality criteria include the Jacobi method, gradient method and the Newton method.

The purpose of the present paper is to analyze the stability of the critical points of joint diagonalization criteria, point out that known criteria have some problem when it is applied to the joint diagonalization of matrices with closely placed eigenvalues, and introduce a new criterion (defined in Section 2.3) to circumvent the problem. The stability analysis of the critical points using the Hessian of the criterion reveals that, when the given matrices have closely placed eigenvalues, the sensitivity of the joint diagonalization criterion seriously goes down, that is, a large deviation in the diagonalizer causes an extremely small amount of change in the criterion, which is a problem in the context of independent component analysis where the main purpose is not the calculation of the eigenvalues of given matrices but the estimation of the diagonalizer. The effect of the closeness of the eigenvalues to the sensitivity of the widely used "off" criterion is quadratic. To circumvent the problem, we propose to use a new joint diagonality criterion (defined in Section 2.3). The criterion is parameterized by constant diagonal matrices of the same size with the target matrices and its sensitivity can be controlled by the diagonal elements of the constant diagonal matrices. On the other hand, unless the diagonal elements are appropriately adjusted, the critical points of the criterion become unstable, that is, the iterative minimization of the criterion does not converge to the desired matrix. Since the appropriate adjustment of the diagonal elements can not be done without knowing the orders of the eigenvalues of the target matrices, we have to start the optimization with some other criterion like "off" and switch to the minimization of the new criterion after the target matrices come close to diagonal matrices so that they expose the order of their eigenvalues.

The rest of the present paper is organized as follows. Section 2.1 and 2.2 survey two known joint diagonalization criteria and Section 2.3 introduces the new criterion. Section 3 analyzes the critical points of the two know criteria and the new criterion to examine the stability and the sensitivity of them. Section 4 demonstrates the advantage of the new criterion using simulation. Section 5 contains concluding remarks.

2. JOINT DIAGONALIZATION CRITERIA

Consider a set of K matrices

$$\{A_1, A_2, \ldots, A_K\}, \quad A_k \in M(n, C).$$

Joint diagonalization of the set of matrices is a problem of finding a matrix U which makes

$$U^* A_1 U, U^* A_2 U, \ldots, U^* A_K U$$

as diagonal as possible simultaneously where U is typically a unitary matrix. If all the matrices of the set commute each other then they can be diagonalized simultaneously. If they do not then they can not be diagonalized simultaneously but instead we can try to find a unitary matrix which minimizes a joint diagonality criterion defined to solve specific problems. We call the former and the latter "exact joint diagonalization" and "approximate joint diagonalization" respectively.

A diagonality measure of a single matrix is a function defined on the set of matrices $\mathfrak{A}$ which meets the following condition,

$$\psi(A) = \min_{A \in \mathfrak{A}} \psi(A) \Rightarrow A \text{ is diagonal.}$$

A joint diagonality criterion of a set of matrices is then defined using the diagonality measure of each matrix,

$$\phi(U) = \sum_{k=1}^{K} \psi(U^* A_k U).$$

Once a joint diagonality criterion is defined, a joint diagonalization problem can be approached using iterative optimization methods such as the Jacobi method, the gradient method and the Newton method. The rest of the section briefly surveys two known joint diagonality criteria and introduces a new criterion.

2.1. Sum of squares of off-diagonals

The sum of squares of the off-diagonal elements,

$$\psi_1(A) = \mathbf{off}(A) = \sum_{i \neq j} |a_{ij}|^2,$$

usually called "off", is a widely used measure of diagonality of a matrix. Based on "off", a joint diagonality criterion is defined as

$$\phi_1(U) = \sum_{k=1}^{K} \mathbf{off}(U^* A_k U). \tag{1}$$

The criterion is easy to understand as well as theoretically reasonable when it is applied to the joint diagonalization of the fourth order cumulant matrices[3]. Several optimization methods have been applied to the criterion including the extended Jacobi method[1][3] and the Newton method[10].

2.2. Log likelihood criterion

Pham[5] introduced a diagonality measure of a positive definite Hermitian matrix A defined as

$$\psi_2(A) = \log\det\operatorname{diag}(A) - \log\det(A),$$

motivated by the maximum likelihood method used in the common principal component estimation[6]. The maximum likelihood estimation in the problem amounts to minimization of the joint diagonality criterion

$$\phi_2(U) = \sum_{k=1}^{K} n_k (\log\det\operatorname{diag}(U^* A_k U) - \log\det(U^* A_k U)) \tag{2}$$

where n_k is a positive number representing the size of observations from the distribution with a covariance matrix A_k. The criterion has an invariance with respect to a scale change which is not shared with "off". Pham[5] considered the case without orthogonality constraint and developed a Jacobi-like method for the optimization of the criterion. In the following section, we consider the case with orthogonality constraint where the second term of the criterion becomes constant.

2.3. Linear criterion

To overcome the shortcomings of the previous two criteria which will be discussed in the following sections, we introduce a new joint diagonality criterion based on the diagonality measure of a matrix defined as

$$\psi_3(A) = -\operatorname{tr} CA = -\sum_{i=1}^{n} c_i a_{ii},$$

where $C = \operatorname{diag}(c_1, \ldots, c_n)$ is a constant diagonal matrix. This diagonality measure is related the work by Brockett[8] in which it was shown that the gradient flow on the adjoint orbit minimizing the measure globally converges to a diagonal matrix whose diagonal elements are ordered similarly to the diagonal elements of the constant matrix C. Based on the diagonality measure, we introduce a linear joint diagonality criterion defined as

$$\phi_3(U) = -\sum_{k=1}^{K} \operatorname{tr} C_k (U^* A_k U), \tag{3}$$

where $C_k = \operatorname{diag}(c_1^{(k)}, \ldots, c_n^{(k)}), k = 1, 2, \ldots, K$ are constant diagonal matrices. As we will see in the analysis in the following section, a critical point of the joint diagonality criterion is stable only if the diagonal elements of $U^* A_k U$ and the diagonal elements of C_k are similarly ordered for all k. Because it is generally impossible to know the order of the diagonal elements of

$U^* A_k U$ at critical points in advance, we need to start the optimization with some other criterion like "off" and switch to the linear criterion when the target matrices become close to diagonal so that their diagonal elements expose their order at critical points. Because any choice of the diagonal matrices C_k whose diagonal elements are appropriately ordered makes the critical points stable, we have room to control the sensitivity of the criterion by the diagonal elements of C_k.

3. CRITICAL POINT ANALYSIS

We analyze the stability and sensitivity at the critical points of the three joint diagonality criteria defined in the previous section, restricting our attention to the exact problems of real matrices. To calculate the Hessian of the criteria, we use the following coordinate system defined on the orthogonal group. Let E_{ij} denote an $n \times n$ matrix whose (i, j)-th element is 1 and all the other elements are 0's. Suppose that X_{ij} is a skew symmetric matrix defined as

$$X_{ij} = E_{ji} - E_{ij} \quad (i < j),$$

then a plane rotation is expressed in an exponential of the skew symmetric matrix X_{ij},

$$e^{\vartheta X_{ij}} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & \cos\vartheta & & -\sin\vartheta & & \\ & & & \ddots & & & \\ & & \sin\vartheta & & \cos\vartheta & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}.$$

We introduce a local coordinate system,

$$\theta = (\theta_{12}, \theta_{13}, \ldots, \theta_{1n}, \theta_{23}, \ldots, \theta_{2n}, \ldots, \theta_{n-1,n}),$$

on the orthogonal group around the identity defined as

$$\begin{aligned} U(\theta) \;=\; & e^{\theta_{12}X_{12}} \cdot e^{\theta_{13}X_{13}} \cdots e^{\theta_{1n}X_{1n}} \\ & \cdot e^{\theta_{23}X_{23}} \cdots e^{\theta_{2n}X_{2n}} \\ & \cdots \\ & \cdot e^{\theta_{n-1,n}X_{n-1,n}}. \end{aligned}$$

In the coordinate system, the identity matrix I is expressed as $\theta = 0$. Let us put

$$A(\theta) = U(\theta)^* A_0 U(\theta),$$

then the first and second order derivatives of $A(\theta)$ at A_0 are calculated as

$$\partial_{ij} A(\theta)|_{\theta=0} = [A_0, X_{ij}],$$

$$\partial_{ij}\partial_{kl} A(\theta)|_{\theta=0} = [[A_0, X_{ij}], X_{kl}],$$

where ∂_{ij} denotes the differentiation with respect to θ_{ij},

$$\partial_{ij} = \frac{\partial}{\partial \theta_{ij}}.$$

Note that the order of X_{ij} and X_{kl} in the right hand side of the formula of the second order derivative is the same as the order appeared in the definition of $U(\theta)$ and does not depend on the order of differentiation.

We consider an obvious exact joint diagonalization problem of a set of matrices,

$$\{\Lambda_1, \Lambda_2, \ldots, \Lambda_K\},$$

where

$$\Lambda_k = \operatorname{diag}(\lambda_1^{(k)}, \lambda_2^{(k)}, \ldots, \lambda_n^{(k)}), \quad k = 1, 2, \ldots, K,$$

and examine the stability and the sensitivity of joint diagonality criteria around the critical point $U = I$. From a viewpoint of the stability analysis, we do not loose generality by this setting. To calculate the Hessian of a diagonality criterion at diagonal points, we put

$$A_k(\theta) = U(\theta)^* \Lambda_k U(\theta), \quad k = 1, \ldots, K.$$

Then the first and second order derivatives of $A_k(\theta)$ at Λ_k are calculated as

$$\begin{aligned} \partial_{ij} A_k(\theta)|_{\theta=0} &= [\Lambda_k, X_{ij}] \\ &= -(\lambda_i^{(k)} - \lambda_j^{(k)})(E_{ij} + E_{ji}), \end{aligned} \tag{4}$$

$$\begin{aligned} \partial_{ij}\partial_{kl} A_k(\theta)|_{\theta=0} &= [[\Lambda_k, X_{ij}], X_{kl}] \\ &= \begin{cases} -2(\lambda_i^{(k)} - \lambda_j^{(k)})(E_{ii} - E_{jj}) & (i = k, j = l) \\ (\lambda_i^{(k)} - \lambda_j^{(k)})(E_{jl} + E_{lj}) & (i = k, j \neq l) \\ -(\lambda_i^{(k)} - \lambda_j^{(k)})(E_{ik} + E_{ki}) & (i \neq k, j = l) \\ (\lambda_i^{(k)} - \lambda_j^{(k)})(E_{il} + E_{li}) & (k = j, i < l) \\ 0 & (\text{otherwise}) \end{cases}. \end{aligned} \tag{5}$$

The Hessian of the "off" joint diagonality criterion (1) at $U = I$ is calculated as follows. Minimizing "off" is equivalent to maximizing the sum of square of the diagonal elements under an orthogonal transform. Therefore the negative of the sum of square of diagonal elements is equivalent to "off" from the viewpoint of the calculation of the Hessian. Then we have

$$\partial_{ij}\partial_{kl}\phi_1(U(\theta))|_{\theta=0} = -\sum_{r=1}^{K}\sum_{p=1}^{n} \partial_{ij}\partial_{kl}(a_{pp}^{(r)})^2$$

$$= \; -\sum_{r=1}^{K}\sum_{p=1}^{n} 2(a_{pp}^{(r)}\partial_{ij}\partial_{kl}a_{pp}^{(r)} + \partial_{ij}a_{pp}^{(r)}\partial_{kl}a_{pp}^{(r)})$$

where $a_{ij}^{(k)}$ denotes the (i,j)-th element of the k-th matrix $A_k(\theta)$. Using (4) and (5), we obtain

$$\partial_{ij}\partial_{kl}\phi_1(U(\theta))|_{\theta=0} = \begin{cases} 4\sum_{r=1}^{K}(\lambda_i^{(r)}-\lambda_j^{(r)})^2 & (i=k, j=l) \\ 0 & (\text{otherwise}) \end{cases}.$$

The obtained Hessian is already diagonalized and its $n(n-1)/2$ diagonal elements are given as

$$4\sum_{k=1}^{K}(\lambda_i^{(k)}-\lambda_j^{(k)})^2 \quad (i<j).$$

From this we see that all the diagonal elements of the Hessian are positive and the critical point $U=I$ is stable when each Λ_k has distinct diagonal elements while the sensitivity of the criterion becomes weak when $\lambda_i^{(r)}$ and $\lambda_j^{(r)}$ are close for the most of r for some pair (i,j).

For the log likelihood criterion (2), the Hessian is calculated as follows. Since the second term of the diagonality criterion is invariant under an orthogonal transform, the criterion is equivalent to

$$\tilde{\psi}_2(A) = \log\det\mathrm{diag}(A) = \sum_{p=1}^{n}\log a_{pp}.$$

Then the Hessian is obtained as

$$\partial_{ij}\partial_{kl}\phi_2(U(\theta))|_{\theta=0} = \sum_{r=1}^{K}\sum_{p=1}^{n} n_r\partial_{ij}\partial_{kl}\log(a_{pp}^{(r)})$$
$$= \sum_{r=1}^{K}\sum_{p=1}^{n} n_r \frac{a_{pp}^{(r)}\partial_{ij}\partial_{kl}a_{pp}^{(r)} - \partial_{ij}a_{pp}^{(r)}\partial_{kl}a_{pp}^{(r)}}{(a_{pp}^{(r)})^2}$$
$$= \begin{cases} 2\sum_{r=1}^{K} n_r \dfrac{(\lambda_i^{(r)}-\lambda_j^{(r)})^2}{\lambda_i^{(r)}\lambda_j^{(r)}} & (i=k, j=l) \\ 0 & (\text{otherwise}) \end{cases}.$$

The obtained Hessian is also diagonalized in this case. Note that the log likelihood criterion is for the positive definite Hermitian matrices and all the diagonal elements of Λ_k are assumed to be positive. We then see that the critical points of the log likelihood criterion have the stability and the sensitivity similar to those of the "off" criterion.

For the linear joint diagonality criterion (3), the Hessian is calculated as

$$\partial_{ij}\partial_{kl}\phi_3(U(\theta))|_{\theta=0} = -\sum_{r=1}^{K}\sum_{p=1}^{n} c_p^{(r)}\,\partial_{ij}\partial_{kl}a_{pp}^{(r)}$$
$$= \begin{cases} 2\sum_{r=1}^{K}(c_i^{(r)}-c_j^{(r)})(\lambda_i^{(r)}-\lambda_j^{(r)}) & (i=k, j=l) \\ 0 & (\text{otherwise}) \end{cases}.$$

The obtained Hessian is also diagonalized in this case. We see that the condition for the critical point $U=I$ to be stable is that C_k and Λ_k have distinct diagonal elements and they are similarly ordered for all k. While the diagonal elements of C_k need to meet the above condition to make the criterion stable, they can be used to control the sensitivity of the criterion. For example, if $\lambda_i^{(r)}$ and $\lambda_j^{(r)}$ are close for the most of r for some pair (i,j) and this spoils the sensitivity of the criterion, then we can set $c_i^{(r)}$ and $c_j^{(r)}$ relatively apart to cancel the effect of the closeness and recover the sensitivity.

4. SIMULATION

We compare the performance of the three joint diagonality criteria $\phi_1(U)$, $\phi_2(U)$ and $\phi_3(U)$ using the following two joint diagonalization problems,

Problem (A) - without closely placed eigenvalues

$$\{\mathrm{diag}(1,2,3,4,5), \mathrm{diag}(5,4,3,2,1)\},$$

Problem (B) - with closely placed eigenvalues

$$\{\mathrm{diag}(1,2,3,4,4.1), \mathrm{diag}(5,4,3,2,1.9)\}.$$

Obviously, the optimal solutions of those joint diagonalization problems are permutation matrices, including the identity matrix. We simulate gradient equations of the three joint diagonality criteria for the two joint diagonalization problem starting from an orthogonal matrix very close to the identity matrix (which must converge to the identity matrix) using the Euler method. For the linear criteria $\phi_3(U)$, we compare the following two options of constant matrices,

$$C_1 = \mathrm{diag}(1,2,3,4,5), C_2 = \mathrm{diag}(5,4,3,2,1),$$

and

$$C_1 = \mathrm{diag}(1,2,3,4,14), C_2 = \mathrm{diag}(14,13,12,11,1).$$

From the consideration in the previous section, it is expected that the former suffers the sensitivity reduction for Problem (B) while the latter cancels the effect of the closely placed eigenvalues.

Figs.1-4 depict the time-evolution of the "off" criterion $\phi_1(U)$ and the deviation from the identity matrix $d(U,I) = |\log(U)|$ for the gradient equations of the joint diagonality criteria. To measure the convergence of the target matrices, we use $\phi_1(U)$ in common for the sake of comparison. The convergence of the diagonalizer is measured by $d(U,I)$. Where $\phi_1(U)$ is small enough but $d(U,I)$ is not, the target matrices are converged to diagonal matrices but the diagonalizer is not converged to the optimal solution.

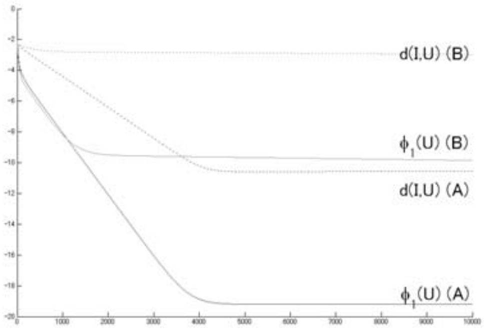

Figure 1. Gradient equation of $\phi_1(U)$

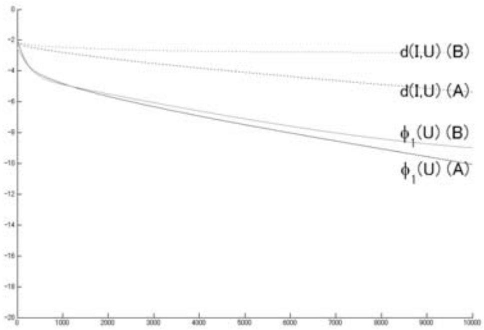

Figure 2. Gradient equation of $\phi_2(U)$

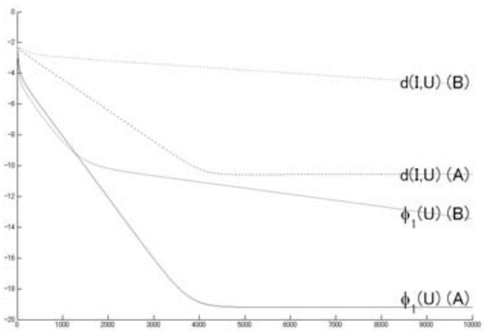

Figure 3. Gradient equation of $\phi_3(U)$
($C_1 = \text{diag}(1, 2, 3, 4, 5), C_2 = \text{diag}(5, 4, 3, 2, 1)$)

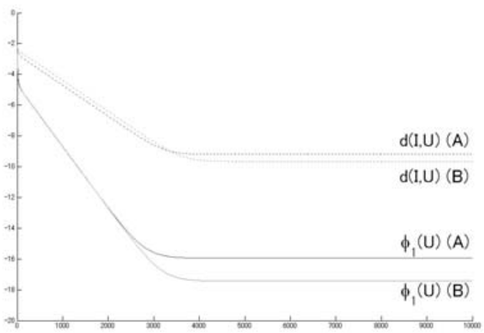

Figure 4. Gradient equation of $\phi_3(U)$
($C_1 = \text{diag}(1, 2, 3, 4, 14), C_2 = \text{diag}(14, 13, 12, 11, 1)$)

Figs.1-3 share the tendency that the convergence of both $\phi_1(U)$ and $d(U, I)$ become dull for Problem (B), comparing to Problem (A). Especially, the convergence of the diagonalizer is totally spoiled by the the effect of the closely placed eigenvalues. On the other hand, in Fig.4, the convergence of both $\phi_1(U)$ and $d(U, I)$, and hence of the diagonalizer, are not affected by the closely placed eigenvalues.

5. CONCLUDING REMARKS

The stability and sensitivity of two known joint diagonality criteria were analyzed using the Hessian and it was pointed out that their sensitivity is weakened when the target matrices have closely placed eigenvalues. To circumvent the problem, we introduced the linear criterion and considered its stability condition and sensitivity control. The consideration was validated by simulation of the gradient equations of the joint diagonality criteria.

To add to that the linear criterion has ability to cancel the effect of closely placed eigenvalues, its linearity makes the joint diagonalization problem more tractable in many aspects of optimization methods. For example, when the linear criterion is optimized using the Jacobi iteration, the optimal rotation angle is given by a much simpler formula (almost the same formula of the rotation angle for a single matrix) comparing to the formula for the "off" criterion obtained by Cardoso and Souloumiac[4]. On the other hand, the linear criterion is only for the final phase of iteration and not for the global optimization. We have to start with an optimization of some other criterion and switch to the linear criterion when the target matrices come close to diagonal matrices.

Finally, we note that there is a unified way of looking at the three joint diagonality criteria referred to in the present paper. Let us extend the constant matrices C_k in the definition of $\phi_3(U)$ to time varying matrices and consider the following two cases,

$$\begin{aligned} C_k &= \text{diag}(a_{11}^{(k)}, a_{22}^{(k)}, \ldots, a_{33}^{(k)}), \\ C_k &= -\text{diag}(1/a_{11}^{(k)}, 1/a_{22}^{(k)}, \ldots, 1/a_{33}^{(k)}), \end{aligned}$$

then the former and the latter correspond to the "off" criterion $\phi_1(U)$ and the log likelihood criterion $\phi_2(U)$ respectively. From this viewpoint, switching joint diagonalization criteria can be regarded as switching the matrices C_k and extended to more flexible scheduling of C_k.

6. REFERENCES

[1] A. Bunse-Gerstner, R. Byers, and V. Mehrmann. Numerical methods for simultaneous diagonalization. *SIAM J. Mat. Anal. Appl.*, 14:927–949, 1993.

[2] A. Bunse-Gerstner, R. Byers, and V. Mehrmann. A chart of numerical methods for structured eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 13:419–453, 1992.

[3] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-gaussian signals. *IEE Proc. F*, 140:362–370, 1993.

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

[5] D.T. Pham. Joint approximate diagonalization of positive definite Hermitian matrices. *SIAM J. Matrix Anal. Appl.*, 22:1136–1152, 2001.

[6] B.N. Flury. Common principal components in k groups. *J. Amer. Statist. Assoc.*, 79:892–897, 1984.

[7] L. Molgedey and H.G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.*, 72:3634–3537, 1994.

[8] R.W. Brockett. Dynamical systems that sort lists, diagonalize matrices and solve linear programming problems. *Linear Algebra Appl.*, 146:79–91, 1991.

[9] G. Hori. A new approach to joint diagonalization. In *Proc. 2nd Intl. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, pages 151–155.

[10] M. Nikpour, J.H. Manton, and G. Hori. Algorithms on stiefel manifold for joint diagonalisation. In *Proc. of ICASSP2002*, pages 1481–1484, 2002.

[11] J.H. Manton. Optimisation algorithms exploiting unitary constraints. *IEEE Trans. Signal Processing*, 50:635–650, 2002.

[12] A. Edelman, T.A. Arias, and S.T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20:303–353, 1998.