A Block-Jacobi Algorithm for Non-Symmetric Joint Diagonalization of Matrices

Hao Shen and Martin Kleinsteuber

Department of Electrical and Computer Engineering
Technische Universität München, Germany
{hao.shen,kleinsteuber}@tum.de

Abstract. This paper studies the problem of Non-symmetric Joint Diagonalization (NsJD) of matrices, namely, jointly diagonalizing a set of complex matrices by one matrix multiplication from the left and one multiplication with possibly another matrix from the right. In order to avoid ambiguities in the solutions, these two matrices are restricted to lie in the complex oblique manifold. We derive a necessary and sufficient condition for the uniqueness of solutions, and characterize the Hessian of global minimizers of the off-norm cost function, which is typically used to measure joint diagonality. By exploiting existing results on Jacobi algorithms, we develop a block-Jacobi method that guarantees local convergence to a pair of joint diagonalizers at a quadratic rate. The performance of the proposed algorithm is investigated by numerical experiments.

Key words: Non-symmetric joint diagonalization of matrices, complex oblique manifold, uniqueness conditions, block Jacobi algorithm, local quadratic convergence.

1 Introduction

Joint Diagonalization (JD) of a set of matrices has attracted considerable attentions in the areas of statistical signal processing and multivariate statistics. Its applications include linear blind source separation (BSS), beamforming, and direction of arrival (DoA) estimation, cf. [1]. Classic literature focuses on the problem of symmetric joint diagonalization (SJD) of matrices. Namely, a set of matrices are to be diagonalized via matrix congruence transforms, i.e. multiplication from the left and the right with a matrix and its (Hermitian) transposed, respectively.

In this work, we consider the problem of *Non-symmetric Joint Diagonalization (NsJD)* of matrices, where the two matrices multiplied from the left and the right are different. Such a general form has been studied in the scheme of multiple-access multiple-input multiple-output (MIMO) wireless transmission, cf. [2]. In the work of [3], NsJD approaches have demonstrated its application in solving the problem of independent vector analysis. Moreover, the problem of NsJD is closely related to the problem of the canonical polyadic decomposition of tensors. We refer to [4, 5] for further discussions on the latter subject.

One contribution of this work is the development of a block Jacobi method for solving the problem of NsJD at a super-linear rate of convergence. Jacobi-type methods have a long history in solving problems of matrix joint diagonalization. Early work in [6] employs a Jacobi algorithm for solving the unitary joint diagonalization problem based on the common off-norm cost function. In the non-unitary setting, Jacobi-type methods have been developed for both the log-likelihood formulation, cf. [7], and the common off-norm case, cf. [8, 9]. The current work provides an extension of [10], which only considers the problem in the real symmetric setting.

2   Uniqueness of non-symmetric JD

In this work, we denote by $(\cdot)^{\mathsf{T}}$ the matrix transpose, $(\cdot)^{\mathsf{H}}$ the Hermitian transpose, $\overline{(\cdot)}$ the complex conjugate of entries of a matrix, and by $Gl(m)$ the set of all $m \times m$ invertible complex matrices. Let $\{C_i\}_{i=1}^n$ be a set of $m \times m$ complex matrices, constructed by

$$C_i = A_L \Omega_i A_R^{\mathsf{H}}, \quad i = 1, \ldots n, \tag{1}$$

where $A_L, A_R \in Gl(m)$ and $\Omega_i = \mathrm{diag}\left(\omega_{i1}, \ldots, \omega_{im}\right) \in \mathbb{C}^{m \times m}$ with $\Omega_i \neq 0$. It is worth noticing that there is no relation between the matrices A_L and A_R. The task of NsJD is to find a pair of matrices $X_L, X_R \in Gl(m)$ such that the set of matrices

$$\left\{ X_L^{\mathsf{H}} C_i X_R \big| i = 1, \ldots, n \right\} \tag{2}$$

are jointly diagonalized. In order to investigate properties of any algorithm that aims at finding such a joint diagonalizer, it is fundamental to understand under what conditions there exists a unique solution. In the remainder of this section, we therefore elaborate the uniqueness properties of the non-symmetric joint diagonalization problem (2).

As it is known from the SJD case, there is also an inherent permutation and scaling ambiguity here. Let $D_L, D_R \in Gl(m)$ be diagonal and $P \in Gl(m)$ be a permutation matrix. If $X_L^* \in Gl(m)$ and $X_R^* \in Gl(m)$ are the joint diagonalizers of problem (2), then so is the pair of $(X_L^* D_L P, X_R^* D_R P)$. In other words, the joint diagonalizers can only be identified up to individual scaling and a joint permutation. We define the set of two jointly column-wise permuted diagonal $(m \times m)$-matrices by

$$\mathcal{G}(m) := \big\{ (D_1 P, D_2 P) \big| D_1, D_2 \in Gl(m) \text{ are diagonal and} \\ P \in Gl(m) \text{ is a permutation matrix} \big\}. \tag{3}$$

As the set $\mathcal{G}(m)$ admits a group structure, we can define an equivalence class on $Gl(m) \times Gl(m)$ as follows.

Definition 1 (Essential Equivalence). *Let $(X_L, X_R) \in Gl(m) \times Gl(m)$, then (X_L, X_R) is said to be* essentially equivalent to *$(Y_L, Y_R) \in Gl(m) \times Gl(m)$, and vice versa, if there exists $(E_L, E_R) \in \mathcal{G}(m)$ such that*

$$X_L = Y_L E_L \qquad and \qquad X_R = Y_R E_R. \tag{4}$$

Moreover, we say that the solution of problem (2) *is* essentially unique, *if it admits a unique solution on the set of equivalence classes.*

Due to the fact that

$$X_L^{\mathsf{H}} C_i X_R = (X_L^{\mathsf{H}} A_L) \Omega_i (A_R^{\mathsf{H}} X_R), \tag{5}$$

we assume without loss of generality that the $C_i = \Omega_i$, $i = 1, \ldots, n$, are already diagonal. In other words, we investigate the question of under what conditions the set $\mathcal{G}(m)$ admits the only solutions to the joint diagonalization problem (2), when the C_i's are already diagonal.

In order to characterize the uniqueness conditions, we need to define a measure of collinearity for diagonal matrices. Recall $\Omega_i = \mathrm{diag}(\omega_{i1}, \ldots, \omega_{im}) \in \mathbb{C}^{m \times m}$ for $i = 1, \ldots, n$. For a fixed diagonal position k, we denote by $z_k := [\omega_{1k}, \ldots, \omega_{nk}]^{\mathsf{T}} \in \mathbb{C}^n$ the vector consisting of the k-th diagonal element of each matrix, respectively. Then, the collinearity measure for the set of Ω_i's is defined by

$$\rho(\Omega_1, \ldots, \Omega_n) := \max_{1 \leq k < l \leq n} |c(z_k, z_l)|, \tag{6}$$

where $c(z_k, z_l)$ is the cosine of the complex angle between two vectors $v, w \in \mathbb{C}^n$, computed as

$$c(v, w) := \begin{cases} \frac{v^{\mathsf{H}} w}{\|v\| \|w\|} & \text{if } v \neq 0 \wedge w \neq 0, \\ 1 & \text{otherwise.} \end{cases} \tag{7}$$

Here, $\|v\|$ denotes the Euclidean norm of a vector v. Note, that $0 \leq \rho \leq 1$ and that $\rho = 1$ if and only if there exists a complex scalar $\omega \in \mathbb{C}$ and a pair z_k, z_l, $k \neq l$ so that $z_k = \omega z_l$. In other words, $\rho = 1$ if and only if there exist two positions (k, k) and (l, l) such that the corresponding entries in the matrices Ω_i only differ by multiplication with a complex scalar ω. We adopt the methods for uniqueness analysis of symmetric joint diagonalization cases, developed in [11], to the NsJD setting.

Lemma 1. *Let $\Omega_i \in \mathbb{C}^{m \times m}$, for $i = 1, \ldots, n$, be diagonal, and let $X_L, X_R \in Gl(m)$ so that $X_L^{\mathsf{H}} \Omega_i X_R$ is diagonal as well. Then the pair (X_L, X_R) is essentially unique if and only if $\rho(\Omega_1, \ldots, \Omega_n) < 1$.*

Proof. First, consider the case $m = 2$ and let

$$X_L = \begin{bmatrix} l_1 & l_2 \\ l_3 & l_4 \end{bmatrix} \in Gl(2), \qquad \text{and} \qquad X_R = \begin{bmatrix} r_1 & r_2 \\ r_3 & r_4 \end{bmatrix} \in Gl(2). \tag{8}$$

Then $X_L^{\mathsf{H}} \Omega_i X_R$ is diagonal for $i = 1, \ldots, n$, if and only if

$$\begin{cases} \omega_{i1} \overline{l_1} r_2 + \omega_{i2} \overline{l_3} r_4 = 0 \\ \omega_{i1} \overline{l_2} r_1 + \omega_{i2} \overline{l_4} r_3 = 0 \end{cases} \tag{9}$$

for $i = 1, \ldots, n$. The corresponding system of linear equations reads as

$$\begin{bmatrix} \omega_{11} & \omega_{21} & \ldots & \omega_{n1} \\ \omega_{12} & \omega_{22} & \ldots & \omega_{n2} \end{bmatrix}^{\mathsf{H}} \begin{bmatrix} l_1 \overline{r_2} & l_2 \overline{r_1} \\ l_3 \overline{r_4} & l_4 \overline{r_3} \end{bmatrix} = 0, \tag{10}$$

which only has a unique trivial solution if and only if the coefficient matrix in ω's has rank 2, which in turn is equivalent to $\rho(\Omega_1, \ldots, \Omega_n) < 1$. Specifically, the trivial solution, i.e.

$$l_1 \overline{r_2} = l_2 \overline{r_1} = l_3 \overline{r_4} = l_4 \overline{r_3} = 0, \tag{11}$$

together with the invertibility of X_L and X_R yields that either

$$l_1 = r_1 = l_4 = r_4 = 0, \quad \text{or} \quad l_2 = r_2 = l_3 = r_3 = 0. \tag{12}$$

Therefore, one can conclude that $(X_L, X_R) \in \mathcal{G}(2)$. For the case $m > 2$, if $\rho = 1$ then there exists a pair (k, l) such that $|c(z_k, z_l)| = 1$ and the same argument as above shows that $\rho = 1$ implies the non-uniqueness of the joint diagonalizer.

For the reverse direction of the statement, we assume that the joint diagonalizer is not in $\mathcal{G}(m)$. We further assume that one of the Ω_i's, say Ω_1, is invertible. Then

$$X_L^{\mathsf{H}} \Omega_i X_R (X_L^{\mathsf{H}} \Omega_1 X_R)^{-1} = X_L^{\mathsf{H}} \Omega_i \Omega_1^{-1} (X_L^{\mathsf{H}})^{-1}, \tag{13}$$

for $i = 1, \ldots, n$, gives the simultaneous eigendecomposition of the diagonal matrices $\Omega_i \Omega_1^{-1}$. Since we assume $(X_L, X_R) \notin \mathcal{G}(m)$, this eigendecomposition is not unique and thus each matrix $\Omega_i \Omega_1^{-1}$ must have at least two identical eigenvalues and these eigenvalues must be located at the same positions (k, k) and (l, l) for all the matrices $\Omega_i \Omega_1^{-1}$. In other words, there exists a pair (k, l) with $k \neq l$ such that

$$\frac{\omega_{ik}}{\omega_{1k}} = \frac{\omega_{il}}{\omega_{1l}}, \tag{14}$$

which is equivalent to $|c(z_k, z_l)| = 1$ and hence $\rho(\Omega_1, \ldots, \Omega_n) = 1$. If all the Ω_i's are singular, we distinguish between two cases. Firstly, assume that there is a position on the diagonals, say k, where all $\omega_{ik} = 0$. Then $|c(z_k, z_l)| = 1$ holds true for any $k \neq l$ and thus $\rho = 1$. Secondly, if there is no common position where all the Ω_i's have a zero entry, there exists an invertible linear combination, say Ω_0, which can also be diagonalized via the same transformations. Then by considering a new set $\{\Omega_i\}_{i=0}^n$, the same argument as from (13) to (14) for the invertible case applies by replacing Ω_1 with Ω_0. $\blacksquare$

3   Analysis of the joint diagonality measure

To deal with the scaling ambiguity, we restrict the search space for the diagonalizing matrices to the quotient space

$$Op(m) = Gl(m)/\{D \in Gl(m) \mid D \text{ is diagonal.}\}, \tag{15}$$

cf. [12] for further details. As representatives for one equivalent class, we choose elements from the set of complex oblique matrices, i.e.

$$\mathcal{OB}(m) := \{X \in \mathbb{C}^{m \times m} \mid \operatorname{ddiag}(X^{\mathsf{H}} X) = I_m, \operatorname{rk} X = m\}, \tag{16}$$

where $\mathrm{ddiag}(Z)$ forms a diagonal matrix, whose diagonal entries are just those of Z, and I_m is the $m \times m$ identity matrix. As a measure for the joint diagonality, we choose

$$f : Op(m) \times Op(m) \to \mathbb{R}, \qquad (X_L, X_R) \mapsto \frac{1}{4} \sum_{i=1}^{n} \left\| \mathrm{off}(X_L^{\mathsf{H}} C_i X_R) \right\|_F^2, \quad (17)$$

where $\mathrm{off}(Z) = Z - \mathrm{ddiag}(Z)$ and $\| \cdot \|_{\mathrm{F}}$ is the Frobenius norm of matrices.

The convergence rate of the Jacobi method depends on a non-degenerated Hessian form of the cost function at the optimal solution. In what follows, we therefore characterize the critical points of (17) and specify their Hessian form. Let us denote the set of all $m \times m$ matrices with zero diagonal entries by

$$\mathfrak{off}(m) = \left\{ Z \in \mathbb{C}^{m \times m} | z_{ii} = 0, \text{ for } i = 1, \ldots, m \right\}. \tag{18}$$

Then, for any $X \in Op(m)$, the following map

$$\mu_X : \mathfrak{off}(m) \to Op(m), \ Z \mapsto X(I_m + Z) \, \mathrm{diag} \left\{ \tfrac{1}{\|X(e_1 + z_1)\|}, \ldots, \tfrac{1}{\|X(e_m + z_m)\|} \right\}, \ (19)$$

where $Z = [z_1, \ldots, z_m] \in \mathfrak{off}(m)$ and e_i is the i-th standard basis vector, is a local and smooth parameterization around X. Let $\mathcal{X} := (X_L, X_R) \in Op^2(m) := Op(m) \times Op(m)$ and denote by $\mathcal{H} := (H_L, H_R) \in T_{\mathcal{X}} Op^2(m)$ the tangent vector at $\mathcal{X}$. The first derivative of f evaluated at $\mathcal{X}$ in tangent direction $\mathcal{H}$ yields

$$\mathrm{D}f(\mathcal{X})(\mathcal{H}) = \sum_{i=1}^{n} \mathrm{tr} \Big(\mathrm{off} \left(X_L^{\mathsf{H}} C_i X_R \right) X_R^{\mathsf{H}} C_i^{\mathsf{H}} H_L + \mathrm{off} \left(H_L^{\mathsf{H}} C_i X_R \right) X_R^{\mathsf{H}} C_i^{\mathsf{H}} X_L + $$
$$\mathrm{off} \left(X_L^{\mathsf{H}} C_i X_R \right) H_R^{\mathsf{H}} C_i^{\mathsf{H}} X_L + \mathrm{off} \left(X_L^{\mathsf{H}} C_i H_R \right) X_R^{\mathsf{H}} C_i^{\mathsf{H}} X_L \Big). \tag{20}$$

Since any joint diagonalizer $\mathcal{X}^*$ is a global minimum, it satisfies $\mathrm{D}f(\mathcal{X}^*)\mathcal{H} = 0$. We use $\mu_{\mathcal{X}} := (\mu_{X_L}, \mu_{X_R})$ as the local parameterization of $Op^2(m)$ and compute the Hessian form $\mathsf{H}_f(\mathcal{X}^*) : T_{\mathcal{X}^*} Op^2(m) \times T_{\mathcal{X}^*} Op^2(m) \to \mathbb{R}$. Let $\mathcal{X}^* = (X_L^*, X_R^*)$ be a pair of joint diagonalizers. Without loss of generality assume that $X_L^{*\mathsf{H}} A_L$ is diagonal and denoted by $\Lambda_L := \mathrm{diag}(\lambda_{L1}, \ldots, \lambda_{Lm})$. Similarly, we denote $\Lambda_R := \mathrm{diag}(\lambda_{R1}, \ldots, \lambda_{Rm}) = A_R^{\mathsf{H}} X_R^*$. A tedious but direct calculation leads to

$$\mathsf{H}_f(\mathcal{X}^*)(\mathcal{H}, \mathcal{H}) = \frac{\mathrm{d}^2}{\mathrm{d}t^2} (f \circ \mu_{\mathcal{X}^*})(t\Theta, t\Xi) \Big|_{t=0} = \sum_{j \neq k}^{m} \begin{bmatrix} \theta_{jk} \\ \xi_{kj} \end{bmatrix}^{\mathsf{T}} \underbrace{\begin{bmatrix} \sum_{i=1}^{n} |\delta_{ij}|^2 & \sum_{i=1}^{n} \delta_{ij}\overline{\delta_{ik}} \\ \sum_{i=1}^{n} \overline{\delta_{ij}}\delta_{ik} & \sum_{i=1}^{n} |\delta_{ik}|^2 \end{bmatrix}}_{=: B_{jk}} \begin{bmatrix} \theta_{jk} \\ \xi_{kj} \end{bmatrix}, \ (21)$$

where δ_{ij} is the j-th diagonal entry of the diagonal matrix $\Delta_i := X_L^* C_i X_R^* = \Lambda_L \Omega_i \Lambda_R$.

Clearly, the Hessian of f at $\mathcal{X}^*$ is at least positive semi-definite, and diagonal in terms of 2×2 blocks, with respect to the standard basis of the parameter

space $\mathfrak{off}(m) \times \mathfrak{off}(m)$. Then, the definiteness of the Hessian simply depends on the determinant of B_{jk}'s, which is computed by

$$
\begin{aligned}
\det(B_{jk}) &= \left(\sum_{i=1}^{n} |\delta_{ij}|^2 \right) \left(\sum_{i=1}^{n} |\delta_{ik}|^2 \right) - \left| \sum_{i=1}^{n} \delta_{ij} \overline{\delta_{ik}} \right|^2 \\
&= |\lambda_{Lj}|^2 \, |\lambda_{Rj}|^2 \, |\lambda_{Lk}|^2 \, |\lambda_{Rk}|^2 \left(\left(\sum_{i=1}^{n} |\omega_{ij}|^2 \right) \left(\sum_{i=1}^{n} |\omega_{ik}|^2 \right) - \left| \sum_{i=1}^{n} \omega_{ij} \overline{\omega_{ik}} \right|^2 \right).
\end{aligned} \tag{22}
$$

By the Cauchy-Schwarz inequality, $\det(B_{jk})$ is non-negative, and is equal to zero if and only if there is a pair of positions (j,k), so that $z_j \in \mathbb{R}^n$ and z_k are linearly dependent, i.e. $\rho(\Omega_1, \ldots, \Omega_n) = 1$. Thus, we conclude

Lemma 2. *Let the NsJD problem* (2) *have a unique joint diagonalizer. Then the Hessian of the off-norm function* (17) *at the joint diagonalizer is positive definite.*

4 Block-Jacobi for non-symmetric joint diagonalization

In this section, we develop a block Jacobi algorithm to minimize the cost function (17). Firstly, let us define the complex one dimensional subspace

$$
V_{ij} := \big\{ Z = (z_{kl}) \in \mathbb{C}^{m \times m} | z_{kl} = 0 \text{ for } (k,l) \neq (i,j) \big\}. \tag{23}
$$

It is clear that $\oplus_{i \neq j}(V_{ij} \times V_{ji}) = \mathfrak{off}(m) \times \mathfrak{off}(m)$. We then define

$$
\mathcal{V}_{ij}(\mathcal{X}) := \{ \tfrac{\mathrm{d}}{\mathrm{d}t} \, \mu_{\mathcal{X}}(t \cdot Z)|_{t=0} \, | Z \in (V_{ij} \times V_{ji}) \}, \tag{24}
$$

yielding a vector space decomposition of the tangent space $T_{\mathcal{X}} Op^2(m)$.

The block Jacobi-type method iteratively employs the search along the subspaces $\mathcal{V}_{ij}(X)$ in a cyclic manner. More precisely, let $(\Theta_{jk}, \Theta_{kj}) \in V_{ij} \times V_{ji}$ and denote $\theta = [\theta_{jk} \ \theta_{kj}]^{\mathsf{T}} \in \mathbb{C}^2$. For any point $\mathcal{X} \in Op^2(m)$, we construct a family of maps $\{\nu_{jk}^{(\mathcal{X})}\}_{j \neq k}^{m}$ by

$$
\nu_{jk}^{(\mathcal{X})} : \mathbb{C}^2 \to Op^2(m), \qquad \theta \mapsto \mu_{\mathcal{X}}(\Theta_{jk}, \Theta_{kj}). \tag{25}
$$

The algorithm is presented in Algorithm 1. It is readily seen with (21) that the $\mathcal{V}_{ij}$ are orthogonal with respect to $\mathsf{H}_f(\mathcal{X}^*)$. Therefore, the following result guarantees the super linear convergence rate of the block Jacobi method to an exact joint diagonalizer in case that this diagonalizer is essentially unique.

Theorem 1 ([13]). *Let M be an n-dimensional manifold and let x^* be a local minimum of the smooth cost function $f : M \to \mathbb{R}$ with nondegenerate Hessian $\mathsf{H}_f(x^*)$. Let μ_x be a family of local parameterizations of M and let $\oplus_i V_i$ be a decomposition of $\mathbb{R}^n$. If the subspaces $\mathcal{V}_i := T_0 \mu_{x^*}(V_i) \subset T_{x^*} M$ are orthogonal with respect to $\mathsf{H}_f(x^*)$, then the Block-Jacobi method is locally quadratic convergent to x^*.*

Algorithm 1 *Jacobi Algorithm for Non-Symmetric Joint Diagonalization*

Step 1: Given an initial guess $\mathcal{X}^{(0)} = (X_L^{(0)}, X_R^{(0)}) \in Op^2(m)$ and set $s = 0$.

Step 2: Set $s = s + 1$ and let $\mathcal{X}_s = \mathcal{X}_{s-1}$.

      For $1 \leq j < k \leq m$, update

$$\mathcal{X}_s \leftarrow \nu_{jk}^{(\mathcal{X}_s)}(\theta^*),$$

    with $\theta^* = -\mathsf{H}_\varphi(0)^{-1}\nabla_\varphi(0)$. Here, H and ∇ denote the usual Hessian and the gradient of the function

$$\varphi \colon \mathbb{C}^2 \to \mathbb{R}, \quad \theta \mapsto f \circ \nu_{jk}^{(\mathcal{X}_s)}(\theta)$$

Step 3: If $\|\mathcal{X}_s - \mathcal{X}_{s-1}\|$ is small enough, stop.

      Otherwise, go to Step 2.

For an experimantal evaluation of this result, we consider the task of jointly diagonalizing a set of non-symmetric matrices $\{\widetilde{C}_i\}_{i=1}^n$, constructed by

$$\widetilde{C}_i = A_L \Lambda_i A_R^{\mathsf{H}} + \varepsilon E_i, \qquad i = 1, \ldots, n, \tag{26}$$

where $A \in \mathbb{C}^{m \times m}$ is a randomly picked matrix in $\mathcal{OB}(m)$, the modulus of diagonal entries of Λ_i are drawn from a uniform distribution on the interval $(9, 11)$, $E_i \in \mathbb{C}^{m \times m}$ represents the additive noise, whose entries are generated from a uniform distribution on the unit interval $(-0.5, 0.5)$, and $\varepsilon \in \mathbb{R}$ is the noise level. We set $m = 5$, $n = 20$, and run six tests in accordance with increasing noise, by using $\varepsilon = d \times 10^{-2}$ where $d = 0, \ldots, 5$. Each experiment was initialized with the same point, which is randomly drawn within an appropriate neighbourhood of the true joint diagonalizer.

The convergence of algorithms is measured by the distance of the accumulation point $\mathcal{X}^* := (X_L^*, X_R^*)$ to the current iterate $\mathcal{X}^{(k)} := (X_L^{(k)}, X_R^{(k)})$, i.e., by $\|X_L^{(k)} - X_L^*\|_F + \|X_R^{(k)} - X_R^*\|_F$. According to Fig. 1, our proposed algorithm converges locally quadratically fast to a pair of joint diagonalizers under the NsJD setting, i.e., when $\varepsilon = 0$, whereas with an increasing level of noise, the convergence rate slows down accordingly with a tendency of more gradual slopes.

References

1. Chabriel, G., Kleinsteuber, M., Moreau, E., Shen, H., Tichavsky, P., Yeredor, A.: Joint matrices decompositions and blind source separation: A survey of methods, identification, and applications. IEEE Signal Processing Magazine **31**(3) (2014) 34–43

2. Chabriel, G., Barrere, J.: Non-symmetrical joint zero-diagonalization and mimo zero-division multiple access. IEEE Transactions on Signal Processing **59**(5) (2011) 2296–2307

3. Shen, H., Kleinsteuber, M.: A matrix joint diagonalization approach for complex independent vector analysis. In Proceedings of the 10th International Conference
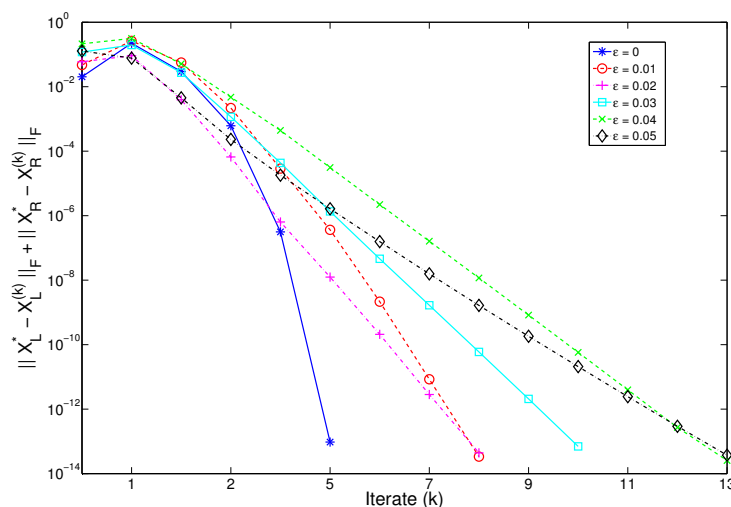
Fig. 1. Convergence properties of the proposed block Jacobi algorithm.

on Latent Variable Analysis and Signal Separation (LVA/ICA). Volume 7191 of Lecture Notes in Computer Science., (2012) 66–73

4. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Review **51**(3) (2009) 455–500

5. Cichocki, A., Mandic, D.P., Phan, A.H., Caiafa, C.F., Zhou, G., Zhao, Q., de Lath-auwer, L.: Tensor decompositions for signal processing applications: From two-way to multiway component analysis. IEEE Signal Processing Magazine **32**(2) (2015) 145–163

6. Cardoso, J.F., Souloumiac, A.: Jacobi angles for simultaneous diagonalisation. SIAM Journal of Matrix Analysis and Application **17**(1) (1996) 161–164

7. Pham, D.T.: Joint approximate diagonalization of positive definite Hermitian ma-trices. SIAM Journal on Matrix Analysis and Applications **22**(4) (2001) 1136–1152

8. Wang, F., Liu, Z., Zhang, J.: Nonorthogonal joint diagonalization algorithm based on trigonometric parameterization. IEEE Transactions on Signal Processing **55**(11) (2007) 5299–5308

9. Souloumiac, A.: Nonorthogonal joint diagonalization by combining Givens and hyperbolic rotations. IEEE Transactions on Signal Processing **57**(6) (2009) 2222–2231

10. Shen, H., Hüper, K.: Block Jacobi-type methods for non-orthogonal joint diago-nalisation. In: Proceedings of the 34th IEEE ICASSP. (2009) 3285–3288

11. Kleinsteuber, M., Shen, H.: Uniqueness analysis of non-unitary matrix joint diag-onalization. IEEE Transactions on Signal Processing **61**(7) (2013) 1786–1796

12. Kleinsteuber, M., Shen, H.: A geometric framework for non-unitary joint diagonal-ization of complex symmetric matrices. In: Proceedings of the 1st conference on Geometric Science of Information (GSI). Volume 8085 of Lecture Notes in Com-puter Science. (2013) 353–360

13. Kleinsteuber, M., Shen, H.: Block-jacobi methods with newton-steps andnon-unitary joint matrix diagonalization. In: Accepted at the 2nd conference on Geo-metric Science of Information (GSI). (2015)