BLOCK JACOBI-TYPE METHODS FOR NON-ORTHOGONAL JOINT DIAGONALISATION

Hao Shen[1] *and Knut Hüper*[2]

[1]Department of Information Engineering, The Australian National University, ACT 0200, Australia.
Email: hao.shen@rsise.anu.edu.au

[2] Max Planck Institute for Biological Cybernetics, Spemannstraße 38, 72076 Tübingen, Germany.
Email: knut.hueper@tuebingen.mpg.de

ABSTRACT

In this paper, we study the problem of non-orthogonal joint diagonalisation of a set of real symmetric matrices. A family of block Jacobi-type methods are proposed to optimise two popular cost functions for the non-orthogonal joint diagonalisation, namely, the off-norm function and the log-likelihood function. By exploiting the appropriate underlying manifold, namely the so-called oblique manifold, rigorous analysis shows that, under the exact non-orthogonal joint diagonalisation setting, the proposed methods converge locally quadratically fast to a joint diagonaliser. Finally, performance of our methods is investigated by numerical experiments for both exact and approximate non-orthogonal joint diagonalisation.

Index Terms— Independent component analysis (ICA), non-orthogonal joint diagonalisation (NoJD), oblique manifold, block Jacobi-type method, local quadratic convergence.

1. INTRODUCTION

In recent years, joint diagonalisation of a set of matrices has attracted considerable attention in the areas of statistical signal processing and multivariate statistics. It has become a prominent tool for the problem of linear independent component analysis (ICA). In general, after a whitening process of the source signals, usually via principal component analysis (PCA), a linear ICA problem is expected to be solved by a joint diagonalisation of a set of matrices, which are derived from certain statistics of the whitened sources, with an orthogonality constraint on the diagonalising matrix [6]. Unfortunately, it has been shown in [4] that linear ICA performed by an orthogonal joint diagonalisation (OJD) might have a serious limit of degraded performance, especially when additive noise is present. Furthermore, some diagonality criterion, namely the weighted least square (WLS) based criterion, might result in poor performance of joint diagonalisation, since this criterion is practically distorted by the whitening process [12].

To avoid the aforementioned limits of OJD, a natural relaxation of OJD, namely the non-orthogonal joint diagonalisation (NoJD), has been recently proposed and studied with increasing attention. Generally speaking, criterion or measure of diagonality of a set of matrices can be constructed mainly in three different forms, specifically, off-norm formulation [7], log-likelihood formulation [8], and subspace fitting formulation [10]. Note that log-likelihood based criteria only apply to a set of positive definite matrices. Nowadays, various numerical algorithms for optimising these diagonality criteria have been developed in the community. To list a few, they include Newton-type methods [7], Gauss-Newton methods [10], Jacobi-type methods [8, 11], and so on. In this work, we focus on developing Jacobi-type NoJD methods.

Other than the algebra-like approaches taken in [8, 11], we treat Jacobi-type method as an optimisation approach on smooth manifold. It is known that, by the construction of the so-called Exact NoJD and certain diagonality criteria, the oblique manifold is an appropriate setting for these scenarios [1, 3]. Recent work in [5] proposes a block Jacobi-type method on a smooth manifold, which has been applied successfully to the problem of independent subspace analysis (ISA) in [9]. The task of this paper is to develop block Jacobi-type NoJD methods from a geometric optimisation perspective, which optimise two particular NoJD cost functions, specifically the off-norm function and the log-likelihood function.

The paper is organised as follows. Section 2 provides a brief introduction to the NoJD problem and some basic concepts of the oblique manifold required in our derivation. In Section 3, we analyse the critical point set of two NoJD cost functions under the Exact NoJD setting. By exploiting the block diagonal structure of the Hessians at the joint diagonaliser, a family of block Jacobi-type NoJD methods is proposed. Meanwhile, local convergence properties of these methods are presented as well. Finally, in Section 4, performance of the proposed algorithms is investigated by numerical experiments for both the Exact NoJD and Approximate NoJD settings.

2. MATHEMATICAL PRELIMINARIES

Given a set of $m\times m$ real symmetric matrices $\{C_i\}_{i=1}^n$, constructed by

$$C_i = A\Lambda_i A^\top, \qquad i = 1,\ldots,n, \tag{1}$$

where $\Lambda_i = \operatorname{diag}\left(\lambda_{i1},\ldots,\lambda_{im}\right) \in \mathbb{R}^{m\times m}$ with $\lambda_{ij} \neq 0$ for $j = 1,\ldots,m$ and $A \in \mathbb{R}^{m\times m}$ non-singular. The problem of estimating the matrix A given only the set $\{C_i\}_{i=1}^n$ is equivalent to finding a non-singular matrix $X \in \mathbb{R}^{m\times m}$ such that the set of matrices $\{Y_i\}_{i=1}^n$, computed by

$$Y_i = X^\top C_i X, \qquad i = 1,\ldots,n, \tag{2}$$

are simultaneously diagonalised. We refer to this problem as the Exact NoJD problem. It is clear that, a joint diagonaliser X can only be determined up to column-wise permutation and scaling, i.e., if X is a diagonaliser, so is any XDP where D is an $m\times m$ invertible diagonal matrix and P an $m\times m$ permutation matrix. For any $j = 1,\ldots,m$, let us denote $\overline{\lambda}_j = (\lambda_{ij})_{i=1}^n \in \mathbb{R}^n$. According to theorem 2.3 in [3], the Exact NoJD problem (2) has a unique joint diagonaliser, which refers to that column-wise permutation and scaling are the only indeterminacies in identifying the diagonaliser, if and only if any pair $(\overline{\lambda}_j, \overline{\lambda}_k)$ for $j \neq k$ are linearly independent.

To deal with the scaling ambiguity, the oblique manifold is shown to be an appropriate setting for Exact NoJD [1, 3]. Let denote the $m\times m$ oblique manifold by

$$\mathcal{OB}(m) := \left\{X \in \mathbb{R}^{m\times m} \,|\, \operatorname{ddiag}(X^\top X) = I_m, \operatorname{rk} X = m\right\}, \tag{3}$$

where $\operatorname{ddiag}(Z)$ forms a diagonal matrix, whose diagonal entries are just those of Z, and I_m is the $m \times m$ identity matrix. Thus, we define the off-norm function [7], a NoJD cost function, as follows

$$f_1\colon \mathcal{OB}(m) \to \mathbb{R}, \quad X \mapsto \frac{1}{4}\sum_{i=1}^{n} \left\| \operatorname{off}(X^\top C_i X) \right\|_{\mathrm{F}}^2, \qquad (4)$$

where $\operatorname{off}(Z) = Z - \operatorname{ddiag}(Z)$ is a matrix by setting the diagonal entries of Z to zero, and $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm of a matrix.

In the specific application of linear ICA, however, due to the presence of additive noise or estimation errors, exact joint diagonalisation is hardly possible to achieve. Thus in general, the task is alternatively to find a matrix $X \in \mathbb{R}^{m\times m}$ such that the matrices $\{Y_i\}_{i=1}^n$ under the construction (2) are jointly as diagonal as possible. We refer to this problem as the Approximate NoJD problem. It is important to notice that the off-norm function (4) is not *column-wise scale invariant* with respect to the matrix X, except that X is an exact joint diagonaliser. So to overcome this limit, a log-likelihood based cost function is proposed, see [8],

$$f_2\colon \mathcal{OB}(m) \to \mathbb{R}, \quad X \mapsto \frac{1}{2}\sum_{i=1}^{n} \log \frac{\det \operatorname{ddiag}\left(X^\top C_i X\right)}{\det\left(X^\top C_i X\right)}. \qquad (5)$$

Although this function is nicely *column-wise scale invariant*, it unfortunately applies only to the scenario where the set of matrices to be jointly diagonalised are positive definite.

More recently, to avoid these drawbacks of the functions (4) and (5), a new subspace fitting based cost function is developed in [2]. Following our analysis in the next section, however, we can show that this cost function does not share the features, which we will derive for both the off-norm function and the log-likelihood function. In other words, the block Jacobi-type NoJD method developed in this work does not apply to it. Therefore, discussion and analysis on the subspace fitting function in [2] are omitted.

Before we continue, let us review some basic concepts of $\mathcal{OB}(m)$. It is known that $\mathcal{OB}(m)$ is an $m(m-1)$ dimensional smooth manifold. The tangent space of $\mathcal{OB}(m)$ at $X \in \mathcal{OB}(m)$ is defined by

$$T_X\mathcal{OB}(m) = \left\{H \in \mathbb{R}^{m\times m} \mid \operatorname{ddiag}(X^\top H) = 0\right\}. \qquad (6)$$

Let denote the set of all $m \times m$ matrices with all diagonal entries equal to zero by

$$\mathfrak{off}(m) = \left\{Z \in \mathbb{R}^{m\times m} \mid z_{ii} = 0, \text{ for } i = 1, \ldots, m\right\}. \qquad (7)$$

Lemma 1 *For every point $X \in \mathcal{OB}(m)$, the following map*

$$\begin{aligned}&\mu_X\colon \mathfrak{off}(m) \to \mathcal{OB}(m),\\ &Z \mapsto X(I_m + Z)\operatorname{diag}\left\{\tfrac{1}{\|X(e_1+z_1)\|}, \ldots, \tfrac{1}{\|X(e_m+z_m)\|}\right\},\end{aligned} \qquad (8)$$

where $Z = [z_1, \ldots, z_m] \in \mathfrak{off}(m)$ and e_i is the i-th standard basis vector of $\mathbb{R}^m$, is a local, smooth paramterisation around X with the following properties. For a given $X = [x_1, \ldots, x_m] \in \mathcal{OB}(m)$ and $\Theta = [\theta_1, \ldots, \theta_m] \in \mathfrak{off}(m)$,

(i) $\mu_X(0) = X$, *and*

(ii) $H := \mathrm{D}\,\mu_X(0)(\Theta) \in T_X\mathcal{OB}(m)$, *where*

$$H = \left[\Pi(x_1)X\theta_1, \ldots, \Pi(x_m)X\theta_m\right]. \qquad (9)$$

Here $\Pi(x_i) := I_m - x_i x_i^\top$ is the orthogonal projection operator onto the complement of $\operatorname{span}(x_i)$.

3. DERIVATION OF BLOCK JACOBI-TYPE METHODS FOR EXACT NOJD

All analysis and derivation in this section are undertaken under the Exact NoJD setting. In Section 3.1, we first provide critical point analysis of the functions (4) and (5). The Hessians of both functions at a joint diagonaliser are shown to be positive definite and share a nice structure of being block diagonal, which leads to a family of block Jacobi-type NoJD methods in Section 3.2.

3.1. Analysis of Cost Functions for Exact NoJD

3.1.1. The Off-norm Function

Taking the first derivative of f_1, we get

$$\mathrm{D}\, f_1(X)H = \sum_{i=1}^{n} \operatorname{tr}\left(\operatorname{off}\left(X^\top C_i X\right) X^\top C_i H\right), \qquad (10)$$

where $H \in T_X\mathcal{OB}(m)$. Let $X^* \in \mathcal{OB}(m)$ be a joint diagonaliser. It is trival to see that

$$\mathrm{D}\, f_1(X^*)H = 0, \qquad (11)$$

i.e., any diagonaliser of the Exact NoJD problem (2) is a critical point of f_1. We now calculate the Hessian of f_1 at such an X^*, i.e. the symmetric bilinear form $\mathsf{H}_{f_1}(X^*)\colon T_{X^*}\mathcal{OB}(m) \times T_{X^*}\mathcal{OB}(m) \to \mathbb{R}$. Using Lemma 1, we get

$$\begin{aligned}\mathsf{H}_{f_1}(X^*)(H,H) &= \tfrac{\mathrm{d}}{\mathrm{d}\,t}(f_1 \circ \mu_{X^*})(t\Theta)\big|_{t=0}\\ &= \sum_{i=1}^{n} \operatorname{tr}\left(\left(\operatorname{off}\left(X^{*\top} C_i H\right) + \operatorname{off}\left(H^\top C_i X^*\right)\right) X^{*\top} C_i H\right).\end{aligned} \qquad (12)$$

By the construction of Exact NoJD (1), we know

$$X^* = A^{-\top}\Delta \in \mathcal{OB}(m), \qquad (13)$$

where $\Delta = \operatorname{diag}\left\{\delta_1, \ldots, \delta_m\right\} \in \mathbb{R}^{m\times m}$ normalises each column of $A^{-\top}$, is a joint diagonaliser. It is then equivalent to $A = (X^*)^{-\top}\Delta$. Thus, we can compute

$$\begin{aligned}X^{*\top} C_i H &= X^{*\top}(X^*)^{-\top}\Delta\Lambda_i\Delta(X^*)^{-1}H\\ &= \Delta\Lambda_i\Delta(X^*)^{-1}H.\end{aligned} \qquad (14)$$

Let us denote $D_i = \operatorname{diag}\left\{d_{i1}, \ldots, d_{im}\right\} = \Delta\Lambda_i\Delta$, i.e. $d_{ij} = \delta_j^2\lambda_{ij}$. By recalling Eq (9), a direct computation gives

$$\begin{aligned}\mathsf{H}_{f_1}(X^*)(H,H) &= \sum_{i=1}^{n}\sum_{1\le j<k\le m}\left(d_{ij}\theta_{jk} + d_{ik}\theta_{kj}\right)^2\\ &= \sum_{1\le j<k\le m}\begin{bmatrix}\theta_{jk}\\ \theta_{kj}\end{bmatrix}^\top \underbrace{\begin{bmatrix}\sum_{i=1}^{n} d_{ij}^2 & \sum_{i=1}^{n} d_{ij}d_{ik}\\ \sum_{i=1}^{n} d_{ij}d_{ik} & \sum_{i=1}^{n} d_{ik}^2\end{bmatrix}}_{=:B_1^{jk}}\begin{bmatrix}\theta_{jk}\\ \theta_{kj}\end{bmatrix}.\end{aligned} \qquad (15)$$

Clearly, the Hessian of f_1 at a joint diagonaliser 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 $\mathbb{R}^{m\times(m-1)}$. Then, the definiteness of the Hessian depends on the determinant of B_1^{jk}'s, which is computed by

$$\begin{aligned}\det(B_1^{jk}) &= \Big(\sum_{i=1}^{n} d_{ij}^2\Big)\Big(\sum_{i=1}^{n} d_{ik}^2\Big) - \Big(\sum_{i=1}^{n} d_{ij}d_{ik}\Big)^2\\ &= \delta_j^4\delta_k^4\left(\Big(\sum_{i=1}^{n}\lambda_{ij}^2\Big)\Big(\sum_{i=1}^{n}\lambda_{ik}^2\Big) - \Big(\sum_{i=1}^{n}\lambda_{ij}\lambda_{ik}\Big)^2\right).\end{aligned} \qquad (16)$$

By the Cauchy-Schwarz inequality, it can be shown that $\det(B_1^{jk})$ is certainly non-negative, and equal to zero if and only if $\overline{\lambda}_j = (\lambda_{ij})_{i=1}^n \in \mathbb{R}^n$ and $\overline{\lambda}_k$ are linearly dependent. Thus, recall the uniqueness condition of solutions of Exact NoJD, i.e. theorem 2.3 in [3], we conclude

Lemma 2 *Let the Exact NoJD problem* (2) *has a unique joint diagonaliser. Then the Hessian of the off-norm function* (4) *at the joint diagonaliser is positive definite.*

3.1.2. The Log-likelihood Function

Now we apply the same analysis to the log-likelihood function f_2 (5). Due to the similarities with Section 3.1.1, the results in the follwoing are presented more briefly. Note that, we now assume that the set of matrices $\{C_i\}_{i=1}^n$ are all positive definite, i.e. $\lambda_{ij} > 0$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

The first derivative of f_2 is computed by

$$\mathrm{D} f_2(X)H = \sum_{i=1}^{n} 2\,\mathrm{tr}\Big(\big(\mathrm{ddiag}(X^\top C_i X)\big)^{-1}\big(X^\top C_i H\big) - X^{-1}H\Big). \quad (17)$$

It can be shown that a joint diagonaliser $X^* \in \mathcal{OB}(m)$ of the Exact NoJD (2) is a critical point of f_2, i.e., $\mathrm{D}\, f_2(X^*)H = 0$. Similarly, the Hessian of f_2 at X^*, i.e. the symmetric bilinear form $\mathsf{H}_{f_2}(X^*)\colon T_{X^*}\mathcal{OB}(m) \times T_{X^*}\mathcal{OB}(m) \to \mathbb{R}$, can be computed by

$$\begin{aligned} \mathsf{H}_{f_2}(X^*)(H,H) &= \tfrac{\mathrm{d}}{\mathrm{d}\,t}(f_2 \circ \mu_{X^*})(t\Theta)\big|_{t=0} \\ &= \sum_{i=1}^{n} \sum_{1\le j<k\le m} \left(\frac{d_{ij}}{d_{ik}}\theta_{jk}^2 + 2\theta_{jk}\theta_{kj} + \frac{d_{ik}}{d_{ij}}\theta_{kj}^2\right) \\ &= \sum_{1\le j<k\le m} \begin{bmatrix}\theta_{jk}\\ \theta_{kj}\end{bmatrix}^\top \underbrace{\begin{bmatrix} \sum_{i=1}^n \frac{d_{ij}}{d_{ik}} & n \\ n & \sum_{i=1}^n \frac{d_{ik}}{d_{ij}} \end{bmatrix}}_{=:B_2^{jk}} \begin{bmatrix}\theta_{jk}\\ \theta_{kj}\end{bmatrix}. \end{aligned} \quad (18)$$

Following the fact that $d_{ij} = \delta_j^2 \lambda_{ij}$, we compute

$$\begin{aligned} \det(B_2^{jk}) &= \left(\sum_{i=1}^{n} \frac{d_{ij}}{d_{ik}}\right)\left(\sum_{i=1}^{n} \frac{d_{ik}}{d_{ij}}\right) - n^2 \\ &= \left(\sum_{i=1}^{n} \frac{\lambda_{ij}}{\lambda_{ik}}\right)\left(\sum_{i=1}^{n} \frac{\lambda_{ik}}{\lambda_{ij}}\right) - n^2. \end{aligned} \quad (19)$$

By using the Chebyshev's sum inequality, it can be shown similarly that $\det(B_2^{jk})$ is non-negative, and equal to zero if and only if $\overline{\lambda}_j$ and $\overline{\lambda}_k$ are linearly dependent. Thus we conclude

Lemma 3 *Let the set of matrices* $\{C_i\}_{i=1}^n$ *constructed by* (1) *be all positive definite, and the Exact NoJD problem* (2) *has a unique joint diagonaliser. Then the Hessian of the log-likelihood function* (5) *at the joint diagonaliser is positive definite.*

It is important to notice that, the Hessian of the log-likelihood function f_2 (5) at a joint diagonaliser shares the same block diagonal structure as the Hessian of the off-norm function f_1 (4).

3.2. Block Jacobi-type Exact NoJD Algorithm

In this subsection, we develop a family of block Jacobi-type methods, which optimise the functions f_1 and f_2, by using the block diagonal structure of the Hessians at a joint diagonaliser shared by both functions.

A block Jacobi-type method consists of an iterative application of so-called *grouped variable sweep* operations. Let $\Theta_{jk} = (\theta_{jk})_{i,j=1}^m \in \mathfrak{off}(m)$ be a zero matrix except for the jk- and kj-th entries, and denote $\overline{\theta} = [\theta_{jk}\ \theta_{kj}]^\top \in \mathbb{R}^2$. For any point $X \in \mathcal{OB}(m)$, we construct a family of maps $\{\nu_{jk}^{(X)}\}_{1\le j<k\le m}^m$ by

$$\nu_{jk}^{(X)}\colon \mathbb{R}^2 \to \mathcal{OB}(m), \qquad \overline{\theta} \mapsto \mu_X(\Theta_{jk}). \quad (20)$$

Then, a block Jacobi-type method for optimising the functions f_1 and f_2 is described as

Algorithm 1 *Block Jacobi-type NoJD Method*

Step 1: Given an initial guess $X_0 \in \mathcal{OB}(m)$ and set $s = 0$.
Step 2: Set $s = s + 1$ and let $X_s = X_{s-1}$.
For $1 \le j < k \le m$, update

$$X_s \leftarrow \nu_{jk}^{(X_s)}(\overline{\theta}^*),$$

where

$$\overline{\theta}^* = \begin{cases} \underset{\overline{\theta}\in\mathbb{R}^2}{\operatorname{argmin}}\ f_t \circ \nu_{jk}^{(X_s)}(\overline{\theta}), & \text{if } f_t \circ \nu_{jk}^{(X_s)}(\overline{\theta}) \not\equiv f_t(X_s); \\ 0, & \text{otherwise.} \end{cases}$$

Step 3: If $\|X_s - X_{s-1}\|$ is small enough, stop.
Otherwise, go to Step 2.

Here, $t \in \{1, 2\}$ and $\|\cdot\|$ is any matrix norm.

Let denote the image of the derivative $\mathrm{D}\,\nu_{jk}^{(X)}(0)\colon \mathbb{R}^2 \to T_X\mathcal{OB}(m)$ by $V_{jk}^{(X)} := \mathrm{im}\,\mathrm{D}\,\nu_{jk}^{(X)}(0)$. It is clear that the set $\{V_{jk}^{(X)}\}_{1\le j<k\le m}^m$ forms a direct sum decomposition of the tangent space $T_X\mathcal{OB}(m)$. By recalling the computations (15) and (18), it can be shown that the vector subspaces $V_{jk}^{(X)}$ are *mutually orthonormal* with respect to the Hessians of both functions f_1 and f_2 at a joint diagonaliser, which is a nondegenerate critical point of both functions by Lemma 2 and 3. Thus, following theorem 2.1 in [5], we summarise the local convergence property without proof in the following.

Theorem 1 *Let the Exact NoJD problem* (2) *has a unique joint diagonaliser. Then the block Jacobi-type method, i.e. Algorithm 1, converges locally quadratically fast to the joint diagonaliser.*

Let us look at Algorithm 1 a bit closer. Within a sweep (Step 2) with $X_s = [x_1^{(s)}, \ldots, x_m^{(s)}] \in \mathcal{OB}(m)$, the (j,k)-th iteration solves essentially a NoJD problem of the set of 2×2 symmetric matrices $\{B_{jk}^{(i)}\}_{i=1}^n$, constructed by

$$B_{jk}^{(i)} = \begin{bmatrix} b_{jj}^{(i)} & b_{jk}^{(i)} \\ b_{jk}^{(i)} & b_{kk}^{(i)} \end{bmatrix} \in \mathbb{R}^2, \quad \text{where } b_{jk}^{(i)} = x_j^{(s)\top} C_i x_k^{(s)}. \quad (21)$$

It has been shown in [11] that, under the Exact NoJD setting, a joint diagonaliser of such a subproblem (21) can be obtained in closed-form. In other words, the so-called DNJD algorithm in [11] is a special case of Algorithm 1. Therefore, the local convergence property of the DNJD follows directly from Theorem 1. Nevertheless, in general, the (j,k)-th iteration minimises the following functions, either

$$\begin{aligned} & f_1 \circ \nu_{jk}^{(X)}\colon \mathbb{R}^2 \to \mathbb{R}, \\ & \qquad \overline{\theta} \mapsto \sum_{i=1}^{n} \frac{(b_{jk}^{(i)} + b_{jj}^{(i)}\theta_{jk} + b_{kk}^{(i)}\theta_{kj} + b_{jk}^{(i)}\theta_{jk}\theta_{kj})^2}{\|x_j + \theta_{kj}x_k\|^2\,\|x_k + \theta_{jk}x_j\|^2}, \end{aligned} \quad (22)$$

or

$$\begin{aligned} f_2 \circ \nu_{jk}^{(X)}\colon \mathbb{R}^2 \to \mathbb{R}, \quad \overline{\theta} \mapsto & \sum_{i=1}^{n} \log(b_{jj}^{(i)} + 2b_{jk}^{(i)}\theta_{kj} + b_{kk}^{(i)}\theta_{kj}^2) \\ & + \log(b_{jj}^{(i)}\theta_{jk}^2 + 2b_{jk}^{(i)}\theta_{jk} + b_{kk}^{(i)}) - \log(1 - \theta_{jk}\theta_{kj})^2. \end{aligned} \quad (23)$$

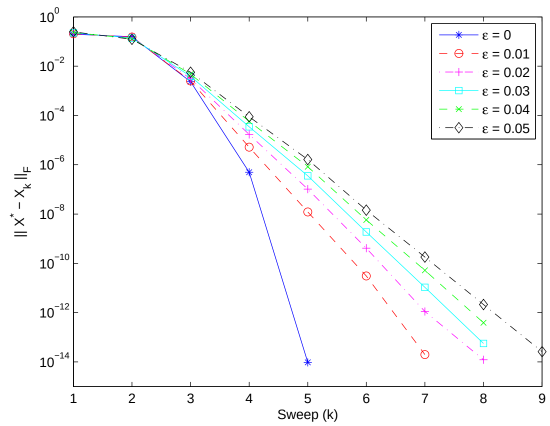

Fig. 1. Convergence properties of Gauss-Newton-Jacobi NoJD.

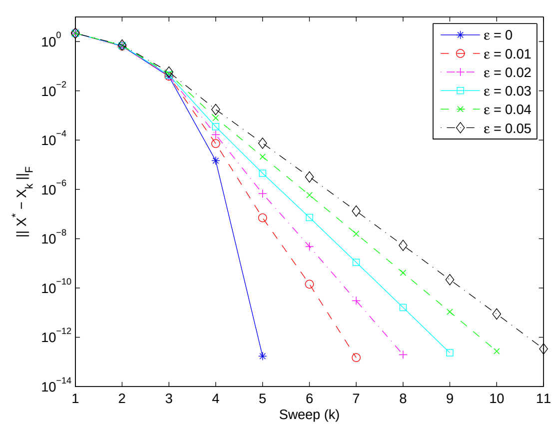

Fig. 2. Convergence properties of Pham's algorithm.

4. NUMERICAL EXPERIMENTS

It is well known that performance of block Jacobi-type methods significantly depends on the techniques to solve the subproblems, e.g. in our case, minimising either (22) or (23). In this section, we will demonstrate local convergence properties of two concrete Jacobi-type NoJD methods under both the Exact and Approximate NoJD settings.

For the off-norm function based methods, i.e. minimising the function (22), although the DNJD algorithm in [11], which occupies a closed-form update derived under the Exact NoJD setting, has demonstrated its effectiveness even for Approximate NoJD by numerical evidence. There is however no theory to guarantee its convergence. In our experiment, we utilise a Gauss-Newton method to minimise (22). In the following, we refer to the resulting Jacobi-type method as Gauss-Newton-Jacobi NoJD method.

For the log-likelihood function based methods, it seems almost impossible to derive a closed-form solution for minimising (23), even under the Exact NoJD setting. Nevertheless, an efficient algorithm, developed in a Jacobi-type style by Pham [8], minimises the upper bound of (23) instead, whose minimum point can be found in closed-form, and appears to coincide with the minimum point of (23) under the Exact NoJD setting. In our experiment, we adapt Pham's algorithm on the oblique manifold.

The task of our experiment is to jointly diagonalise a set of symmetric matrices $\{\widetilde{C}_i\}_{i=1}^n$, constructed by

$$\widetilde{C}_i = A\Lambda_i A^\top + \varepsilon E_i, \qquad i = 1, \ldots, n, \tag{24}$$

where $A \in \mathbb{R}^{m\times m}$ is a randomly picked matrix in $\mathcal{OB}(m)$, diagonal entries of Λ_i are drawn from a uniform distribution on the interval $(9, 11)$, $E_i \in \mathbb{R}^{m\times m}$ is the symmetric part of an $m \times m$ matrix, whose entries are generated from a uniform distribution on the unit interval $(-0.5, 0.5)$, representing the additive noise, and $\varepsilon \in \mathbb{R}$ is the noise level. We set $m = 5$, $n = 20$, and run six tests, for both the Gauss-Newton-Jacobi algorithm and Pham's algorithm, in accordance with increasing noise, by using $\varepsilon = t \times 10^{-2}$ where $t = 0, \ldots, 5$.

The convergence of algorithms is measured by the distance of the accumulation point $X^* \in \mathcal{OB}(m)$ to the current iterate $X_k \in \mathcal{OB}(m)$, i.e., by $\|X_k - X^*\|_{\mathrm{F}}$. According to Fig. 1 and 2, it is clear that both algorithms converge locally quadratically fast to a joint diagonaliser under the Exact NoJD setting, i.e., $\varepsilon = 0$. Although Pham's algorithm was claimed in [8] to converge locally quadratically fast to a joint diagonaliser under the Approximate NoJD setting, such a property indeed holds no longer for both algorithms with the presence of noise. They appear to converge only linearly fast.

5. REFERENCES

[1] P.-A. Absil and K. A. Gallivan, "Joint diagonalization on the oblique manifold for independent component analysis," in *Proceedings of ICASSP 2006*, pp. V945–V948.

[2] B. Afsari, "Simple LU and QR based non-orthogonal matrix joint diagonalization," in *Lecture Notes in Computer Science, Proceedings of ICA 2006*, vol. 3889, pp. 1–7.

[3] B. Afsari, "Sensitivity analysis for the problem of matrix joint diagonalization," to appear at *SIAM journal on matrix analysis and applications* (Special Issue on Tensorial Methods), 2008. http://www.math.umd.edu/~bijan/

[4] J.-F. Cardoso, "On the performance of orthogonal source separation algorithms," in *Proceedings of EUSIPCO 1994*, pp. 776–779.

[5] K. Hüper, "A dynamical system approach to matrix eigenvalue algorithms," in *the IMA Volumes in Mathematics and its Applications, Mathematical Systems Theory in Biology, Communications, Computation*, vol. 134, pp. 257–274, 2003.

[6] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley, New York, 2001.

[7] M. Joho and K. Rahbar, "Joint diagonalization of correlation matrices by using Newton methods with application to blind signal separation," in *Proceedings of SAM 2002*, pp. 403–407.

[8] D. T. Pham, "Joint approximate diagonalization of positive definite Hermitian matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 22, pp. 1136–1152, 2001.

[9] H. Shen, M. Kleinsteuber, and K. Hüper, "Block Jacobi-type methods for log-likelihood based linear independent subspace analysis," in *Proceedings of MLSP 2007*, pp. 133–138.

[10] A.-J. van der Veen, "Joint diagonalization via subspace fitting techniques," in *Proceedings of ICASSP 2001*, pp. 2773–2776.

[11] F. Wang, Z. Liu, and J. Zhang, "Nonorthogonal joint diagonalization algorithm based on trigonometric parameterization," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5299–5308, 2007.

[12] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1545–1553, 2002.