MATHEMATICS OF COMPUTATION
Volume 00, Number 0, Pages 000–000
S 0025-5718(XX)0000-0

# PERTURBATION ANALYSIS FOR MATRIX JOINT BLOCK DIAGONALIZATION

YUNFENG CAI AND REN-CANG LI

Abstract. The matrix joint block diagonalization problem (JBDP) of a given matrix set $\mathcal{A} = \{A_i\}_{i=1}^m$ is about finding a nonsingular matrix $W$ such that all $W^{\mathrm{T}}A_iW$ are block diagonal. It includes the joint diagonalization problem (JDP) as a special case for which all $W^{\mathrm{T}}A_iW$ are diagonal. Generically, such a matrix $W$ may not exist, but there are practically applications such as multidimensional independent component analysis (MICA) for which it does under the ideal situation, ie., no noise is presented. However, in practice noises do get in and, as a consequence, the matrix set is only approximately block diagonalizable, i.e., one can only make all $\widetilde{W}^{\mathrm{T}}A_i\widetilde{W}$ nearly block diagonal at best, where $\widetilde{W}$ is an approximation to $W$, obtained usually by computation. This motivates us to develop a perturbation theory for JBDP to address, among others, the question: how accurate this $\widetilde{W}$ is. Previously such a theory for JDP has been discussed, but no effort has been attempted for JBDP yet. In this paper, with the help of a necessary and sufficient condition for the uniqueness of JBDP developed in [3], we are able to establish the error bound, perform backward error analysis, and propose a condition number for JBDP. Numerical tests validate the theoretical results.

## 1. Introduction

The joint block diagonalization problem (JBDP) is about jointly block diagonalizing a set of matrices. In recent years, it has found many applications in independent subspace analysis, also known as *multidimensional independent component analysis* (MICA) (see, e.g., [4, 11, 28, 29]) and semidefinite programming (see, e.g., [2, 6, 7, 16]). Tremendous efforts have been devoted to solving JBDP and several numerical methods have been proposed. The purpose of this paper, however, is to develop a perturbation theory for JBDP. For this reason, we will not delve into numerical methods, but refer the interested reader to [3, 5, 10, 30] and references therein. Also, the MATLAB toolbox for tensor computation – TENSORLAB [33] can also be used for the purpose.

In the rest of this section, we will formally introduce JBDP and formulate its associated perturbation problem, along with some notations and definitions. Through a case study on the basic MICA model, we rationalize our formulations and provide our motivations for current studies in this paper. Previously, there are only a handful papers in the literature that studied the perturbation analysis of the joint

---

Received by the editor February 27, 2017.
2010 *Mathematics Subject Classification.* 15A23, 15A69, 49Q12.
Y. Cai was supported in part by NSFC grants 11301013, 11671023 and 11421101. R.-C. Li was supported in part by NSF grants DMS-1317330 and CCF-1527104, and NSFC grant 11428104.

diagonalization problem (JDP). Briefly, we will review these existing works and their limitations. Finally, we explain our contribution and the organization of this paper.

1.1. **Joint Block Diagonalization (JBD).** A *partition* of positive integer $n$:

$$\tau_n = (n_1, \dots, n_t) \tag{1.1}$$

means that $n_1, n_2, \dots, n_t$ are all positive integers and their sum is $n$, i.e., $\sum_{i=1}^t n_i = n$. The integer $t$ is called the *cardinality* of the partition $\tau_n$, denoted by $\mathrm{card}(\tau_n)$.

Given a partition $\tau_n$ as in (1.1) and a matrix $A \in \mathbb{R}^{n\times n}$ (the set of $n \times n$ real matrices), we partition $A$ by

$$A = \begin{array}{c c} & \begin{array}{cccc} n_1 & n_2 & \cdots & n_t \end{array} \\ \begin{array}{c} n_1 \\ n_2 \\ \vdots \\ n_t \end{array} & \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1t} \\ A_{21} & A_{22} & \cdots & A_{2t} \\ \vdots & \vdots & & \vdots \\ A_{t1} & A_{t2} & \cdots & A_{tt} \end{bmatrix} \end{array} \tag{1.2}$$

and define its $\tau_n$*-block diagonal part* and $\tau_n$*-off-block diagonal part* as

$$\mathrm{Bdiag}_{\tau_n}(A) = \mathrm{diag}(A_{11}, \dots, A_{tt}), \quad \mathrm{OffBdiag}_{\tau_n}(A) = A - \mathrm{Bdiag}_{\tau_n}(A).$$

The matrix $A$ is referred to as *a* $\tau_n$*-block diagonal matrix* if $\mathrm{OffBdiag}_{\tau_n}(A) = 0$. The set of all $\tau_n$-block diagonal matrices is denoted by $\mathbb{D}_{\tau_n}$.

**The Joint Block Diagonalization Problem** (JBDP). Let $\mathcal{A} = \{A_i\}_{i=1}^m$ be the set of $m$ matrices, where each $A_i \in \mathbb{R}^{n\times n}$. The JBDP for $\mathcal{A}$ with respect to $\tau_n$ is to find a nonsingular matrix $W \in \mathbb{R}^{n\times n}$ such that all $W^{\mathrm{T}} A_i W$ are $\tau_n$-block diagonal, i.e.,

$$W^{\mathrm{T}} A_i W = \mathrm{diag}(A_i^{(11)}, \dots, A_i^{(tt)}) \quad \text{for} \quad i = 1, 2, \dots, m, \tag{1.3}$$

where $A_i^{(jj)} \in \mathbb{R}^{n_j \times n_j}$ for $j = 1, 2, \dots, t$. When (1.3) holds, we say that $\mathcal{A}$ is $\tau_n$*-block diagonalizable* and $W$ is a $\tau_n$*-block diagonalizer* of $\mathcal{A}$. If $W$ is also required to be orthogonal, this JBDP is referred to as an *orthogonal* JBDP (O-JBDP).

By convention, if $\tau_n = (1, 1, \dots, 1)$, the word "$\tau_n$*-block*" is dropped from all relevant terms. For example, "$\tau_n$-block diagonal" is reduced to just "diagonal". Correspondingly, the letter "B" is dropped from all abbreviations. For example, "JBDP" becomes "JDP". This convention is adopted throughout this article.

Generically, JBDP often has no solution for $m \geq 3$ and $n_i$ not so unevenly distributed, simply by counting the number of equations implied by (1.3) and the number of unknowns. For example, when $m = 3$ and $n_1 = n_2 = n_3 = n/3$, there are $m(n^2 - \sum_{i=1}^t n_i^2) = 2n^2$ equations but only $n^2$ unknowns in $W$. However, in certain practical applications such as MICA without noises, solvable JBDP do arise.

**Definition 1.1.** A permutation matrix $\Pi \in \mathbb{R}^{n\times n}$ is called $\tau_n$*-block diagonal preserving* if $\Pi^{\mathrm{T}} D \Pi \in \mathbb{D}_{\tau_n}$ for any $D \in \mathbb{D}_{\tau_n}$. The set of all $\tau_n$-block diagonal preserving permutation matrices is denoted by $\mathbb{P}_{\tau_n}$.

Evidentally, any permutation matrix $\Pi \in \mathbb{D}_{\tau_n}$ is in $\mathbb{P}_{\tau_n}$. This is because such a $\Pi$ can be expressed as $\Pi = \mathrm{diag}(\Pi_1, \dots, \Pi_t)$, where $\Pi_j$ is an $n_j \times n_j$ permutation matrix. But not all $\Pi \in \mathbb{P}_{\tau_n}$ are also in $\mathbb{D}_{\tau_n}$. For example, for $n = 4$ and $\tau_4 = (2, 2)$, $\Pi = \begin{bmatrix} 0 & I_2 \\ I_2 & 0 \end{bmatrix} \in \mathbb{P}_{\tau_4}$ but $\Pi \notin \mathbb{D}_{\tau_4}$. In particular, any permutation matrix $\Pi \in \mathbb{P}_{\tau_n}$

when $\tau = (1, 1, \ldots, 1)$. It can be proved that for given $\Pi \in \mathbb{P}_{\tau_n}$, there is a permutation $\pi$ if $\{1, 2, \ldots, t\}$ such that $\Pi^{\mathrm{T}} D \Pi \in \mathbb{D}_{\tau_n} = \operatorname{diag}(D_{\pi(1)}, D_{\pi(2)}, \ldots, D_{\pi(t)})$ for any $D = \operatorname{diag}(D_1, D_2, \ldots, D_t) \in \mathbb{D}_{\tau_n}$. Thus the subblocks of $\Pi$, if partitioned as in (1.2), are all 0 blocks, except those at the positions $(\pi(j), j)$, which are $n_j \times n_j$ permutation matrices. As a consequence, $n_j = n_{\pi(j)}$ for all $1 \le j \le t$.

It is not hard to verify that if $W$ is a $\tau_n$-block diagonalizer of $\mathcal{A}$, then so is $WD\Pi$ for any given $D \in \mathbb{D}_{\tau_n}$ and $\Pi \in \mathbb{P}_{\tau_n}$. In view of this, $\tau_n$-block diagonalizers, if exist, are not unique because any diagonalizer brings out a class of *equivalent* diagonalizers in the form of $WD\Pi$. For this reason, we introduce the following definition for uniquely block diagonalizable JBDP.

**Definition 1.2.** Two $\tau_n$-block diagonalizers $W$ and $\widetilde{W}$ of $\mathcal{A}$ are *equivalent* if there exist a nonsingular matrix $D \in \mathbb{D}_{\tau_n}$ and $\Pi \in \mathbb{P}_{\tau_n}$ such that $\widetilde{W} = WD\Pi$. The JBDP for $\mathcal{A}$ is said *uniquely $\tau_n$-block diagonalizable* if it has a $\tau_n$-block diagonalizer and if any two of its $\tau_n$-block diagonalizers are equivalent.

To further reduce freedoms for the sake of comparing two diagonalizers, we restrict our considerations of diagonalizers from the matrix set:

$$\mathbb{W}_{\tau_n} := \{W \in \mathbb{R}^{n\times n} \ : \ W \text{ is nonsingular and } \operatorname{Bdiag}_{\tau_n}(W^{\mathrm{T}}W) = I_n\}. \tag{1.4}$$

This doesn't loss any generality because $W[\operatorname{Bdiag}_{\tau_n}(W^{\mathrm{T}}W)]^{-1/2} \in \mathbb{W}_{\tau_n}$ for any nonsingular $W \in \mathbb{R}^{n\times n}$.

1.2. **Perturbation Problem for JBDP.** Let $\widetilde{\mathcal{A}} = \{\widetilde{A}_i\}_{i=1}^m = \{A_i + \Delta A_i\}_{i=1}^m$, where $\Delta A_i$ is a perturbation to $A_i$. Assume $\mathcal{A} = \{A_i\}_{i=1}^m$ is $\tau_n$-block diagonalizable and $W \in \mathbb{W}_{\tau_n}$ is a $\tau_n$-block diagonalizer and (1.3) holds. Let $\widetilde{W} \in \mathbb{W}_{\tau_n}$ be an approximate $\tau_n$-block diagonalizer of $\widetilde{\mathcal{A}}$ in the sense that all $\widetilde{W}^{\mathrm{T}}\widetilde{A}_i\widetilde{W}$'s are approximately $\tau_n$-block diagonal. How much does $\widetilde{W}$ differ from the block diagonalizer $W$ of $\mathcal{A}$?

There are two important aspects that needs clarification regarding this perturbation problem. First, $\widetilde{\mathcal{A}}$ may or may not be $\tau_n$-block diagonalizable. Although allowing this counters the common sense that one can only gauge the difference between diagonalizers that exist, it is for a good reason and important practically. As we argued above, a generic JBDP is usually not block diagonalizable, and thus even if the JBDP for $\mathcal{A}$ has a diagonalizer, its arbitrarily perturbed problem is potentially not block diagonalizable no matter how tiny the perturbation may be. This leads to an impossible task: to compare the block diagonalizer $W$ of the unperturbed $\mathcal{A}$, that does exist, to a diagonalizer $\widetilde{W}$ of the perturbed matrix set $\widetilde{\mathcal{A}}$, that may not exist. We get around this dilemma by talking about an approximate diagonalizer of $\widetilde{\mathcal{A}}$, that always exist. It turns out this workaround is exactly what some practical applications calls for because most practical JBDP come from block diagonalizable JBDP but tainted with noises to become approximately block diagonalizable and an approximate diagonalizer for the noisy JBDP gets computed numerically. It is important to get a sense as how far the computed diagonalizer is from the exact diagonalizer of the clean albeit unknown JBDP, had the noises not presented.

The second aspect is about what metric to use in order to measure the difference between two diagonalizers, given that diagonalizers are not unique. In view of Definition 1.2 and the discussion in the paragraph immediately proceeding it, we propose to use

$$\min_{D\in\mathbb{D}_{\tau_n},\Pi\in\mathbb{P}_{\tau_n}} \frac{\|W-\widetilde{W}D\Pi\|}{\|\widetilde{W}\|} \tag{1.5}$$

for the purpose, where $\|\cdot\|$ is some matrix norm. Usually which norm to use is determined by the convenience of any particular analysis, but for all practical purpose, any norm is just as good as another. In our analysis, below, we use either $\|\cdot\|_2$, the matrix spectral norm, or $\|\cdot\|_{\rm F}$, the matrix Frobenius norm [13]. Additionally, in using (1.5), we usually restrict $W$ and $\widetilde{W}$ to $\mathbb{W}_{\tau_n}$.

1.3. **A Case Study: MICA.** MICA [4, 20, 29] aims at separating linearly mixed unknown sources into statistically independent groups of signals. A basic MICA model can be given by

$$x = Ms + v, \tag{1.6}$$

where $x\in\mathbb{R}^n$ is the observed mixture, $M\in\mathbb{R}^{n\times n}$ is a nonsingular matrix (often called *the mixing matrix*), $s\in\mathbb{R}^n$ is the source signal, and $v\in\mathbb{R}^n$ is the noise vector. We would like to recover the source $s$ from the observed mixture $x$. Let $s=\left[s_1^{\rm T},\dots,s_t^{\rm T}\right]^{\rm T}$ with $s_j\in\mathbb{R}^{n_j}$ for $j=1,2,\dots,t$, and $v=[\nu_1,\dots,\nu_n]^{\rm T}$. To recover the source signal $s$, it suffices to find $M$ or its inverse from the observed mixture $x$. Notice that if $M$ is a solution, then so is $MD\Pi$, where $D$ is a block diagonal scaling matrix and $\Pi$ is a block-wise permutation matrix. In this sense, there is certain degree of freedom in the determination of $M$. Such indeterminacy of the solution is natural, and does not matter in applications.

Assume that all $s_j$ are independent of each other, and each $s_j$ has mean 0 and contains no lower-dimensional independent component, and among all $s_j$, there exists at most one Gaussian component. Assume further that the noises $\nu_1,\dots,\nu_n$ are real stationary white random signals, mutually uncorrelated with the same variance $\sigma^2$, and independent of the sources. We have the following statements.

(a) The covariance matrix $R_{xx}$ of $x$ satisfies

$$R_{xx}=\mathbb{E}(xx^{\rm T})=M\mathbb{E}(ss^{\rm T})M^{\rm T}+\mathbb{E}(vv^{\rm T})=MR_{ss}M^{\rm T}+\sigma^2 I, \tag{1.7}$$

where $\mathbb{E}(\,\cdot\,)$ stands for the mathematical expectation, and $R_{ss}$ is the covariance matrix of $s$. By the above assumptions, we know that $R_{ss}\in\mathbb{D}_{\tau_n}$. Assume that $\sigma$ is accurately estimated as $\hat\sigma$. Then we have

$$R_{xx}-\hat\sigma^2 I\approx MR_{ss}M^{\rm T}. \tag{1.8}$$

In particular, in the absence of noises, i.e., $\sigma=0$, (1.8) becomes an equality.

(b) The kurtosis[1] $\mathcal{C}_x^4$ of $x$ is a tensor of dimension $n\times n\times n\times n$. Fixing two indices, say the first two, and varying the last two, we have

$$\mathcal{C}_x^4(i_1,i_2,:,:)=M\mathcal{C}_s^4(i_1,i_2,:,:)M^{\rm T}, \tag{1.9}$$

where $\mathcal{C}_s^4$ is the kurtosis of $s$ and it can be shown that $\mathcal{C}_s^4(i_1,i_2,:,:)\in\mathbb{D}_{\tau_n}$.

Together, they result in a JBDP for $\widetilde{\mathcal{A}}=\{R_{xx}-\hat\sigma I\}\cup\{\mathcal{C}_x^4(i_1,i_2,:,:)\}_{i_1,i_2=1}^n$, and $\widetilde{W}:=M^{-\rm T}$ is its approximate $\tau_n$-block diagonalizer. When we attempt to compute it, all we can do is to calculate an approximation of $M^{-\rm T}D\Pi$ for some $D\in\mathbb{D}_{\tau_n}$ and $\Pi\in\mathbb{P}_{\tau_n}$, which corresponds to the indeterminacy of MICA (even in the case when $\sigma=0$, i.e., there is no noise).

---

[1] Other cumulants can also be considered.

The point we try to make from this case study is that, in practical applications, due to measurement errors, we only get to work with $\widetilde{\mathcal{A}} = \{\widetilde{A}_i\}$ that are, in general, only approximately jointly block diagonalizable and, in the end, an approximate block diagonalizer $\widetilde{W}$ of $\widetilde{\mathcal{A}}$ gets computed. In the other word, we usually don't have $\mathcal{A}$ which is known block diagonalizable in theory but what we do have is $\widetilde{\mathcal{A}}$ which may or may not be block diagonalizable and for which we have an approximate diagonalizer $\widetilde{W}$. Then how far this $\widetilde{W}$ is from the exact diagonalizer $W$ of $\mathcal{A}$ becomes a central question, in order to gauge the quality of $\widetilde{W}$. This is what we set out to do in this paper. Our result is an upper bound on the measure in (1.5). Such an upper bound will also help us understand what are the inherent factors that affect the sensitivity of JBDP.

1.4. **Related works.** Though tremendous efforts have gone to solve JDP/JBDP, their perturbation problems had received little or no attention in the past. In fact, today there are only a handful articles available on the perturbations of JDP only. For O-JDP, Cardoso [4] presented a first order perturbation bound for a set of commuting matrices, and the result was later generalized by Russo [21]. For general JDP, using gradient flows, Afsari [1] studied sensitivity via cost functions and obtained first order perturbation bounds for the diagonalizer. Shi and Cai [22] investigated a normalized JDP through a constrained optimization problem, and obtained an upper bound on certain distance between an approximate diagonalizer of a perturbed optimization problem and an exact diagonalizer of the unperturbed optimization problem.

JBDP can also be regarded as a particular case of the *block term decomposition* (BTD) of third order tensors [8, 9, 12, 19]. The uniqueness conditions of tensor decompositions, which is strongly connected to the sensitivity of tensor decompositions, received much attention recently (see, e.g., [9, 14, 15, 18, 24, 23, 25]). However, perturbation theories for tensor decompositions, often referred to as *identifiability of tensors*, up to now, is only discussed for the so-called *canonical polyadic decomposition* (CPD) (see [32] and references therein). Perturbation theories for other models of tensor decompositions, e.g., the tucker decomposition and BTD, have not been touched yet. More work is obviously needed in this area.

1.5. **Our contribution and the organization of this paper.** A major reason as to why no available perturbation analysis for JBDP is, perhaps, due to lacking perfect ways to uniquely describe block diagonalizers, not to mention no available uniqueness condition to nail them down. Quite recently, in the sense of Definition 1.2, Cai and Liu [3] established necessary and sufficient conditions for a JBDP to be uniquely block diagonalizable. These conditions are the cornerstone for our current investigation in this paper. Unlike the results in existing literatures, the result in this paper does not involve any cost function, which makes it applicable to any approximate diagonalizer computed from min/maximizing a cost function. The result also reveals the inherent factors that affect the sensitivity of JBDP.

The rest of this paper is organized as follows. In section 2, we discuss properties of a uniquely block diagonalizable JBDP and introduce the concepts of the moduli of uniqueness and non-divisibility that play key roles in our later development. Our main result is presented in section 3, along with detailed discussions on its numerous implications. The proof of the main result is rather long and technical and thus is deferred to section 4. We validate our theoretical contributions by numerical tests reported in section 5. Finally, concluding remarks are given in section 6.

**Notation**. $\mathbb{R}^{m\times n}$ is the set of all $m\times n$ real matrices and $\mathbb{R}^m=\mathbb{R}^{m\times 1}$. $I_n$ is the $n\times n$ identity matrix, and $0_{m\times n}$ is the $m$-by-$n$ zero matrix. When their sizes are clear from the context, we may simply write $I$ and $0$. The symbol $\otimes$ denotes the Kronecker product. The operation $\mathrm{vec}(X)$ turns a matrix $X$ into a column vector formed by the first column of $X$ followed by its second column and then its third column and so on. Inversely, $\mathrm{reshape}(x,m,n)$ turns the $mn$-by-1 vector $x$ into an $m$-by-$n$ matrix in such a way that $\mathrm{reshape}(\mathrm{vec}(X),m,n)=X$ for any $X\in\mathbb{R}^{m\times n}$. The spectral norm and Frobenius norm of a matrix are denoted by $\|\cdot\|_2$ and $\|\cdot\|_{\mathrm{F}}$, respectively. For a square matrix $A$, $\lambda(A)$ is the set of all eigenvalues of $A$, counting algebraic multiplicities. For convenience, we will agree that any matrix $A\in\mathbb{R}^{m\times n}$ has $n$ singular values and $\sigma_{\min}(A)$ is the smallest one among all.

## 2. Uniquely block diagonalizable JBDP

In [3], a classification of JBDP is proposed. Among all and besides the one in subsection 1.1, there is the so-called *general* JBDP (GJBDP) for $\mathcal{A}$ for which a partition $\tau_n$ is not given but instead it asks for finding a partition $\tau_n$ with the largest cardinality such that $\mathcal{A}$ is $\tau_n$-block diagonalizable and at the same time a $\tau_n$-block diagonalizer. Via an algebraic approach, necessary and sufficient conditions [3, Theorem 2.5] are obtained for the uniqueness of (equivalent) block diagonalizers of the GJBDP for $\mathcal{A}$. As a corollary, we have the following result.

**Theorem 2.1** ([3])**.** *Given partition $\tau_n$ of $n$, suppose that the* JBDP *of $\mathcal{A}=\{A_i\}_{i=1}^m$ is $\tau_n$-block diagonalizable and $W$ is its $\tau_n$-block diagonalizer satisfying* (1.3)*. Let $\mathcal{A}_j=\{A_i^{(jj)}\}_{i=1}^m$ for $j=1,2,\ldots,t$ and assume that every $\mathcal{A}_j$ cannot be further block diagonalized*[2]*, i.e., for any partition $\tau_{n_j}$ of $n_j$ with $\mathrm{card}(\tau_{n_j})\geq 2$, $\mathcal{A}_j$ is not $\tau_{n_j}$-block diagonalizable. Then the* JBDP *of $\mathcal{A}=\{A_i\}_{i=1}^m$ is uniquely $\tau_n$-block diagonalizable if and only if the matrix*

(2.1)
$$M_{jk}=\sum_{i=1}^m\begin{bmatrix} I_{n_k}\otimes\Big[(A_i^{(jj)})^{\mathrm{T}}A_i^{(jj)}+A_i^{(jj)}(A_i^{(jj)})^{\mathrm{T}}\Big] & A_i^{(kk)}\otimes A_i^{(jj)}+(A_i^{(kk)})^{\mathrm{T}}\otimes(A_i^{(jj)})^{\mathrm{T}} \\ A_i^{(kk)}\otimes A_i^{(jj)}+(A_i^{(kk)})^{\mathrm{T}}\otimes(A_i^{(jj)})^{\mathrm{T}} & \Big[(A_i^{(kk)})^{\mathrm{T}}A_i^{(kk)}+A_i^{(kk)}(A_i^{(kk)})^{\mathrm{T}}\Big]\otimes I_{n_j}\end{bmatrix}$$

*is nonsingular for all $1\leq j<k\leq t$.*

The following subspace of $\mathbb{R}^{n\times n}$

$$\mathscr{N}(\mathcal{A}):=\big\{Z\in\mathbb{R}^{n\times n}\ :\ A_iZ-Z^{\mathrm{T}}A_i=0 \text{ for } 1\leq i\leq m\big\} \tag{2.2}$$

has played an important role in the proof of [3, Theorem 2.5], and it will also contribute to our perturbation analysis later in a big way.

Next, let us examine some fundamental properties of $Z\in\mathscr{N}(\mathcal{A})$ with

$$A_i=\mathrm{diag}(A_i^{(11)},\ldots,A_i^{(tt)})\quad\text{for } 1\leq i\leq m \tag{2.3}$$

already. Any $Z\in\mathscr{N}(\mathcal{A})$ satisfies

$$\mathrm{diag}(A_i^{(11)},\ldots,A_i^{(tt)})Z-Z^{\mathrm{T}}\,\mathrm{diag}(A_i^{(11)},\ldots,A_i^{(tt)})=0\quad\text{for } 1\leq i\leq m. \tag{2.4}$$

[2]For the MICA model, this assumption is equivalent to say that each component $s_j$ has no lower dimensional component.

Partition $Z$ conformally as $Z = [Z_{jk}]$, where $Z_{jk} \in \mathbb{R}^{n_j \times n_k}$. Blockwise, (2.4) can be rewritten as

$$A_i^{(jj)} Z_{jk} - Z_{kj}^{\mathrm{T}} A_i^{(kk)} = 0 \text{ for } 1 \le i \le m,\ 1 \le j, k \le t. \tag{2.5}$$

These equations can be decoupled into

$$A_i^{(jj)} Z_{jj} - Z_{jj}^{\mathrm{T}} A_i^{(jj)} = 0 \quad \text{for } 1 \le i \le m \tag{2.6a}$$

and $1 \le j \le t$, and

$$A_i^{(jj)} Z_{jk} - Z_{kj}^{\mathrm{T}} A_i^{(kk)} = 0, \quad A_i^{(kk)} Z_{kj} - Z_{jk}^{\mathrm{T}} A_i^{(jj)} = 0 \quad \text{for } 1 \le i \le m \tag{2.6b}$$

and $1 \le j < k \le t$.

Consider first (2.6b). Together they are equivalent to

$$G_{jk} \begin{bmatrix} \operatorname{vec}(Z_{jk}) \\ -\operatorname{vec}(Z_{kj}^{\mathrm{T}}) \end{bmatrix} = 0, \tag{2.7}$$

where

$$G_{jk} = \begin{bmatrix} I_{n_k} \otimes A_1^{(jj)} & (A_1^{(kk)})^{\mathrm{T}} \otimes I_{n_j} \\ I_{n_k} \otimes (A_1^{(jj)})^{\mathrm{T}} & A_1^{(kk)} \otimes I_{n_j} \\ \vdots & \vdots \\ I_{n_k} \otimes A_m^{(jj)} & (A_m^{(kk)})^{\mathrm{T}} \otimes I_{n_j} \\ I_{n_k} \otimes (A_m^{(jj)})^{\mathrm{T}} & A_m^{(kk)} \otimes I_{n_j} \end{bmatrix}. \tag{2.8}$$

Notice that $M_{jk}$ defined in (2.1) simply equals to $G_{jk}^{\mathrm{T}} G_{jk}$. Thus, according to Theorem 2.1, $\mathcal{A}$ is uniquely $\tau_n$-block diagonalizable if and only if the smallest singular value $\sigma_{\min}(G_{jk}) > 0$.

Next, we note that (2.6a) is equivalent to

$$G_{jj} \operatorname{vec}(Z_{jj}) = 0, \tag{2.9}$$

where

$$G_{jj} = \begin{bmatrix} I_{n_j} \otimes A_1^{(jj)} - \big[(A_1^{(jj)})^{\mathrm{T}} \otimes I_{n_j}\big]\Pi_j \\ \vdots \\ I_{n_j} \otimes A_m^{(jj)} - \big[(A_m^{(jj)})^{\mathrm{T}} \otimes I_{n_j}\big]\Pi_j \end{bmatrix}, \tag{2.10}$$

and $\Pi_j \in \mathbb{R}^{n_j^2}$ is the perfect shuffle permutation matrix [31, Subsection 1.2.11] that enables $\Pi_j \operatorname{vec}(Z_{jj}^{\mathrm{T}}) = \operatorname{vec}(Z_{jj})$.

**Theorem 2.2.** *Suppose $\mathcal{A} = \{A_i\}_{i=1}^m$ is already in the* JBD *form with respect to $\tau_n = (n_1, \dots, n_t)$, i.e., $A_i$ are given by (2.3). The following statements hold.*

(a) *$G_{jj} \operatorname{vec}(I_{n_j}) = 0$, i.e., $G_{jj}$ is rank deficient;*
(b) *If $\mathcal{A}_j$ cannot be further block diagonalized, then the eigenvalues of $Z_{jj} \in \mathscr{N}(\mathcal{A}_j)$ are either a single real number or a single pair of two complex conjugate numbers.*
(c) *If $\dim \mathscr{N}(\mathcal{A}_j) = 1$ which means either $n_j = 1$ or the second smallest singular value of $G_{jj}$ is positive, then $\mathcal{A}_j$ cannot be further block diagonalized.*

item (b) is a necessary and sufficient condition when $Z_{jj}$ is arbitrary. The proof is simple. Not simple to me. Let us strengthen item (b) and add the proof in.

*Proof.* Item (a) holds because $Z = I_{n_j}$ clearly satisfies (2.6a). For item (b), if the conclusion were not true, then $Z_{jj}$ can be decomposed into $Z_{jj} = W_j \operatorname{diag}(D_1^{(j)}, D_2^{(j)}) W_j^{-1}$, where $W_j$, $D_1^{(j)}$, $D_2^{(j)}$ are all real matrices and $\lambda(D_1^{(j)}) \cap \lambda(D_2^{(j)}) = \emptyset$. Then substituting the decomposition into (2.6a), we can conclude that $W_j^{\mathrm{T}} A_i^{(jj)} W_j$ for $i = 1, 2, \dots, m$ are all block diagonal matrices, a contradiction. Lastly for item (c), assume, to the contrary, that $\mathcal{A}_j$ can be further block diagonalized. Without loss of generosity, we may assume that there exists a nonsingular matrix $W_j \in \mathbb{R}^{n_j \times n_j}$ such that $W_j^{\mathrm{T}} A_i^{(jj)} W_j = \operatorname{diag}(A_i^{(jj1)}, A_i^{(jj2)})$ for $i = 1, 2, \dots, m$, where $A_i^{(jj1)}$ and $A_i^{(jj2)}$ are respectively of order $n_{j1}$ and $n_{j2}$. Then (2.6a) has at least two linear independent solutions $W_j \operatorname{diag}(I_{n_{j1}}, 0) W_j^{-1}$, $W_j \operatorname{diag}(0, I_{n_{j2}}) W_j^{-1}$. Therefore, (2.9) has two linear independent solutions, which implies that the second smallest singular value of the coefficient matrix $G_{jj}$ must be 0, a contradiction. □

In view of Theorems 2.1 and 2.2, we introduce the moduli of uniqueness and non-divisibility for $\mathcal{A}$ that is $\tau_n$-block diagonalizable.

**Definition 2.3.** Let $W \in \mathbb{W}_{\tau_n}$ be a $\tau_n$-block diagonalizer of $\mathcal{A} = \{A_i\}_{i=1}^m$ such that (1.3) holds.

(a) The *modulus of uniqueness* of the JBDP for $\mathcal{A}$ with respective to the $\tau_n$-block diagonalizer $W$ is defined by

$$\omega_{\mathrm{uq}} \equiv \omega_{\mathrm{uq}}(\mathcal{A}; W) = \min_{1 \le j < k \le t} \sigma_{\min}(G_{jk}), \tag{2.11}$$

where $G_{jk}$ is given by (2.8).

(b) Suppose that $\mathcal{A}$ is uniquely $\tau_n$-block diagonalizable. The *modulus of non-divisibility* $\omega_{\mathrm{nd}} \equiv \omega_{\mathrm{nd}}(\mathcal{A}; W)$ of the JBDP for $\mathcal{A}$ with respective to the $\tau_n$-block diagonalizer $W$ is defined by $\omega_{\mathrm{nd}} = \infty$ if $\tau_n = (1, 1, \dots, 1)$ and

$$\omega_{\mathrm{nd}} = \min_{n_j > 1} \{\text{the smallest nonzero singular value of } G_{jj}\}, \tag{2.12}$$

otherwise, where $G_{jj}$ is given by (2.10).

Note the notion of the modulus of non-divisibility is defined under the condition of unique $\tau_n$-block diagonalizability of $\mathcal{A}$. It is needed because in order for (2.12) to be well-defined, we need to make sure that $G_{jj}$ has at least one nonzero singular value in the case when $n_j > 1$. In deed, $G_{jj} \ne 0$ whenever $n_j > 1$, if $\mathcal{A}$ is uniquely $\tau_n$-block diagonalizable. To see this, we note $G_{jj} = 0$ implies that any matrix $Z_{jj}$ of order $n_j$ is a solution to (2.6a) and thus $A_i^{(jj)}$ for $1 \le i \le m$ are diagonal, which means that $\mathcal{A}_j$ can be further (block) diagonalized. This contradicts to the assumption that $\mathcal{A}$ is uniquely $\tau_n$-block diagonalizable.

The corollary below partially justifies Definition 2.3.

**Corollary 2.4.** *Let $W \in \mathbb{W}_{\tau_n}$ be a $\tau_n$-block diagonalizer of $\mathcal{A} = \{A_i\}_{i=1}^m$ such that (1.3) holds, and let $\mathcal{A}_j = \{A_i^{(jj)}\}_{i=1}^m$. Suppose $\dim \mathscr{N}(\mathcal{A}_j) = 1$ for all $1 \le j \le t$, and let $\sigma_{-2}^{(j)}$ be the second smallest singular value of $G_{jj}$ for $j = 1, 2, \dots, t$ whenever $n_j > 1$. Then the following statement holds.*

How about now?

(a) *$\mathcal{A}$ is uniquely $\tau_n$-block diagonalizable if $\omega_{\mathrm{uq}}(\mathcal{A}; W) > 0$.*

(b) *None of $\mathcal{A}_j$ can be block diagonalized and*

$$\omega_{\mathrm{nd}} \equiv \omega_{\mathrm{nd}}(\mathcal{A}; W) = \min_{n_j > 1} \sigma_{-2}^{(j)} > 0.$$

*Remark* 2.5. A few comments are in order.

(a) The definition of $\omega_{\mathrm{uq}}$ is a natural generation of the modulus of uniqueness in [22] for JDP (i.e., when $\tau_n = (1, 1, \dots, 1)$).

(b) By Theorem 2.2(a), we know the smallest singular value of $G_{jj}$ is always 0. Thus it seems natural that in defining $\omega_{\rm nd}$ in (2.12), one would expect using the *second* smallest singular value of $G_{jj}$. It turns out that there are examples for which $\mathcal{A}_j$ cannot be further block diagonalized and yet $\dim \mathscr{N}(\mathcal{A}_j) = 2$, i.e., the second smallest singular value of $G_{jj}$ is still 0.

Consider $A_i = \left[\begin{smallmatrix} \alpha_i & \beta_i \\ \beta_i & -\alpha_i \end{smallmatrix}\right]$ for $i = 1, 2, \ldots, m$, where all $\alpha_i, \beta_i \neq 0 \in \mathbb{R}$ and $\frac{\alpha_i}{\beta_i}$ are not a constant. Then $\mathcal{A} = \{A_i\}_{i=1}^m$ cannot be simultaneously diaognalized and $\mathscr{N}(\mathcal{A}) = \operatorname{span}\{I_2, \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]\}$, i.e., $\dim \mathscr{N}(\mathcal{A}) = 2$.

The moduli $\omega_{\rm uq}$ and $\omega_{\rm nd}$, as defined in Definition 2.3, depend on the choice of the diaognalizer $W$. But, as the following theorem shows, in the case when $\mathcal{A} = \{A_i\}_{i=1}^m$ is uniquely $\tau_n$-block diagonalizable, their dependency on diagonalizer $W \in \mathbb{W}_{\tau_n}$ can be removed.

**Theorem 2.6.** *If $\mathcal{A} = \{A_i\}_{i=1}^m$ is uniquely $\tau_n$-block diagonalizable, then $\omega_{\rm uq}$ and $\omega_{\rm nd}$ are both independent of the choice of diagonalizer $W \in \mathbb{W}_{\tau_n}$.*

*Proof.* Let $W \in \mathbb{W}_{\tau_n}$ be a $\tau_n$-block diagonalizer of $\mathcal{A}$. Then all possible $\tau_n$-block diagonalizer of $\mathcal{A}$ from $\mathbb{W}_{\tau_n}$ take the form $\widetilde{W} = WD\Pi$ for some $D \in \mathbb{D}_{\tau_n}$ and $\Pi \in \mathbb{P}_{\tau_n}$. We will show that $\omega_{\rm uq}(\mathcal{A}; \widetilde{W}) = \omega_{\rm uq}(\mathcal{A}; W)$ and $\omega_{\rm nd}(\mathcal{A}; \widetilde{W}) = \omega_{\rm nd}(\mathcal{A}; W)$.

We can write $D = \operatorname{diag}(D_1, \ldots, D_t)$, where $D_j \in \mathbb{R}^{n_j \times n_j}$. All $D_j$ are all orthogonal since $W$, $\widetilde{W} \in \mathbb{W}_{\tau_n}$. We have

$$\begin{aligned}
\widetilde{W}^{\rm T} A_i \widetilde{W} &= \Pi^{\rm T} \operatorname{diag}(D_1^{\rm T} A_i^{(11)} D_1, \ldots, D_t^{\rm T} A_i^{(tt)} D_t)\Pi \\
&= \operatorname{diag}(\Pi_1^{\rm T} D_{\ell_1}^{\rm T} A_i^{(\ell_1 \ell_1)} D_{\ell_1} \Pi_1, \ldots, \Pi_t^{\rm T} D_{\ell_t}^{\rm T} A_i^{(\ell_t \ell_t)} D_{\ell_t} \Pi_t),
\end{aligned}$$

where $\{\ell_1, \ell_2, \ldots, \ell_t\}$ is a permutation of $\{1, 2, \ldots, t\}$, and $\Pi_j$ is a permutation matrix of order $n_j$ for $j = 1, \ldots, t$. Denote by $\widetilde{A}_i^{(jj)} = \Pi_j^{\rm T} D_{\ell_j}^{\rm T} A_i^{(\ell_j \ell_j)} D_{\ell_j} \Pi_j$, and define $\widetilde{G}_{jk}$, accordingly as $G_{jk}$ in (2.8), but in terms of $\widetilde{A}_i^{(jj)}$ and $\widetilde{A}_i^{(kk)}$, $\widetilde{G}_{jj}$, accordingly as $G_{jj}$ in (2.10), but in terms of $\widetilde{A}_i^{(jj)}$. Then by calculations, we have

$$\begin{aligned}
\widetilde{G}_{jk} &= \big[I_{2m} \otimes (\Pi_k D_{\ell_k})^{\rm T} \otimes (\Pi_j D_{\ell_j})^{\rm T})\big] G_{jk} \big[I_2 \otimes (\Pi_k D_{\ell_k}) \otimes (\Pi_j D_{\ell_j}))\big], \\
\widetilde{G}_{jj} &= \big[I_m \otimes (\Pi_j D_{\ell_j})^{\rm T} \otimes (\Pi_j D_{\ell_j})^{\rm T})\big] G_{jj} \big[(\Pi_k D_{\ell_k}) \otimes (\Pi_j D_{\ell_j})\big],
\end{aligned}$$

which imply that the singular values of $\widetilde{G}_{jk}$ and $\widetilde{G}_{jj}$ are the same as those of $G_{jk}$ and $G_{jj}$, respectively. The conclusion follows. $\square$

## 3. Main Perturbation Results

In this section, we present our main theorem, along with some illustrating examples and discussions on its implications. We defer its lengthy proof to section 4.

3.1. **Set up the Stage.** In what follows, we will set up the groundwork for our perturbation analysis and explain some of our assumptions.

As before, $\mathcal{A} = \{A_i\}_{i=1}^n$ is the upperturbed matrix set, where all $A_i \in \mathbb{R}^{n \times n}$, and $\tau_n = (n_1, \ldots, n_t)$ is a partition of $n$ with $t \geq 2$. We assume that

$$\boxed{\begin{array}{l}\mathcal{A} \text{ is } \tau_n\text{-block diagonalizable, } W \in \mathbb{W}_{\tau_n} \text{ is its } \tau_n\text{-block diagonalizer} \\ \text{such that (1.3) holds, and, moreover, } \dim \mathscr{N}(\mathcal{A}_j) = 1 \text{ for all } j, \\ \text{where } \mathcal{A}_j = \{A_i^{(jj)}\}_{i=1}^m,\ 1 \leq j \leq t.\end{array}} \tag{3.1}$$

The assumption that $\dim \mathscr{N}(\mathcal{A}_j) = 1$ implies that $\mathcal{A}_j$ cannot be further block diagonalized by Theorem 2.2(c).

$\mathcal{A} = \{A_i\}_{i=1}^n$ is perturbed to $\widetilde{\mathcal{A}} = \{\widetilde{A}_i\}_{i=1}^m \equiv \{A_i + \Delta A_i\}_{i=1}^m$, and let

$$\|\mathcal{A}\|_{\mathrm{F}} := \left(\sum_{i=1}^m \|A_i\|_{\mathrm{F}}^2\right)^{1/2}, \quad \delta_{\mathcal{A}} := \left(\sum_{i=1}^m \|\Delta A_i\|_{\mathrm{F}}^2\right)^{1/2}. \tag{3.2}$$

Previously, we commented on that, more often than not, a generic JBDP may not be diagonalizable for $m \geq 3$. This means that $\widetilde{\mathcal{A}}$ may not be $\tau_n$-block diagonalizable regardless how tiny $\delta_{\mathcal{A}}$ may be. For this reason, we will not assume that $\widetilde{\mathcal{A}}$ is $\tau_n$-block diagonalizable, but, instead, it has an approximate $\tau_n$-block diagonalizer $\widetilde{W} \in \mathbb{W}_{\tau_n}$ in the sense that

$$\text{all } \widetilde{W}^{\mathrm{T}}\widetilde{A}_i\widetilde{W} \text{ are nearly } \tau_n\text{-block diagonal.} \tag{3.3}$$

Doing so makes sense and actually has two advantages. Firstly, this can serve all practical purposes well, because in any likely practical situations we usually end up knowing $\widetilde{\mathcal{A}}$ which is close to some $\tau_n$-block diagonalizable $\mathcal{A}$ that is not available due to unavoidable noises such as in MICA, and, at the same time, an approximate $\tau_n$-block diagonalizer can be made available by computation. Secondly, it is general enough to cover the case when the JBDP for $\widetilde{\mathcal{A}}$ is $\tau_n$-block diagonalizable.

We have to quantify the statement (3.3) in order to proceed. To this end, we pick a diagonal matrix $\Gamma = \mathrm{diag}(\gamma_1 I_{n_1}, \ldots, \gamma_t I_{n_t})$, where $\gamma_1, \ldots, \gamma_t$ are distinct real numbers with all $|\gamma_j| \leq 1$, and define the $\tau_n$-block diagonalizablility residuals

$$\widetilde{R}_i = \widetilde{W}^{\mathrm{T}}\widetilde{A}_i\widetilde{W}\Gamma - \Gamma\widetilde{W}^{\mathrm{T}}\widetilde{A}_i\widetilde{W} \quad \text{for } i = 1, 2, \ldots, m. \tag{3.4}$$

Notice $\mathrm{Bdiag}_{\tau_n}(\widetilde{R}_i) = 0$ always no matter what $\Gamma$ is. The rationale behind defining these residuals is the following proposition.

**Proposition 3.1.** *$\widetilde{W}^{\mathrm{T}}\widetilde{A}_i\widetilde{W}$ is $\tau_n$-block diagonal, i.e., $\mathrm{OffBdiag}_{\tau_n}(\widetilde{W}^{\mathrm{T}}\widetilde{A}_i\widetilde{W}) = 0$ if and only if $\widetilde{R}_i = 0$.*

As far as this proposition is concerned, any diagonal $\Gamma$ with distinct diagonal entries suffices. But later, we will see that our upper bound depends on $\Gamma$, which makes us wonder what the best $\Gamma$ is for the best possible bound. Unfortunately, this is not a trivial task and would be an interesting subject for future studies. We will return to this later in our numerical example section. We restrict $\gamma_i$ to real numbers for consistency consideration since $\mathcal{A}$ and $\widetilde{\mathcal{A}}$ are assumed real. All developments below work equally well even if they are complex. For later use, we set

$$g = \min_{j \neq k} |\gamma_j - \gamma_k|, \quad \tilde{r} = \left(\sum_{i=1}^m \|\widetilde{R}_i\|_{\mathrm{F}}^2\right)^{1/2}. \tag{3.5}$$

In addition to Proposition 3.1, another benefit of defining the residuals $\widetilde{R}_i$ can be viewed through backward error analysis. In fact, all $\widetilde{R}_i$ being nearly zeros, i.e., tiny $\tilde{r}$, implies that $\widetilde{\mathcal{A}}$ is nearby an exact $\tau_n$-block diagonalizable matrix set.

**Proposition 3.2.** *$\widetilde{W}$ is an exact $\tau_n$-block diagonalizer of the matrix set $\{\widetilde{A}_i + E_i\}_{i=1}^m$ with relative backward error*

$$\frac{\|\mathcal{E}\|_{\rm F}}{\|\widetilde{\mathcal{A}}\|_{\rm F}} \le \frac{\|\widetilde{W}^{-1}\|_2^2}{\|\widetilde{\mathcal{A}}\|_{\rm F}} \cdot \frac{\tilde{r}}{g} =: \varepsilon_{\rm bker}(\widetilde{\mathcal{A}};\widetilde{W}), \tag{3.6}$$

*where $\mathcal{E} = \{E_i\}_{i=1}^m$ which will be referred to as the* backward perturbation *to $\widetilde{\mathcal{A}}$ with respect to the approximate diagonalizer $\widetilde{W}$.*

*Proof.* Partition $\widetilde{R}_i$ as $\widetilde{R}_i = \big[\widetilde{R}_i^{(jk)}\big]$ with $\widetilde{R}_i^{(jk)} \in \mathbb{R}^{n_j\times n_k}$. Then (3.4) can be rewritten as

$$\widetilde{W}^{\rm T}(\widetilde{A}_i + E_i)\widetilde{W}\Gamma - \Gamma\widetilde{W}^{\rm T}(\widetilde{A}_i + E_i)\widetilde{W} = 0, \tag{3.7}$$

where $E_i = \widetilde{W}^{-\rm T}\big[E_i^{(jk)}\big]\widetilde{W}^{-1}$ with $E_i^{(jj)} = 0$ and $E_i^{(jk)} = \frac{\widetilde{R}_i^{(jk)}}{\gamma_k-\gamma_j}$ for $j\ne k$. Let $\mathcal{E} = \{E_i\}_{i=1}^m$ which satisfies (3.6). □

3.2. **Main Result.** With the set up, we are ready to state our main result.

**Theorem 3.3.** *Adopt the setup in* subsection 3.1 *up to* (3.4)*. Let $Q = W^{-1}\widetilde{W}$, and let $\omega_{\rm uq}$ and $\omega_{\rm nd}$ be defined in* Definition 2.3*, and*[3]

$$\tau = \frac{\sqrt{2}-1}{\sqrt{t-1}}, \quad \alpha = \frac{2\tau}{(\sqrt{2}+\tau)^2}, \tag{3.8}$$

$$\delta = \|Q^{-1}\|_2^2\,\tilde{r} + 2\|Q^{-1}\|_2\|W\|_2\|\widetilde{W}\|_2\,\delta_{\mathcal{A}}, \quad \epsilon_* = \frac{\tau\kappa_2(Q)\delta}{\alpha g\,\omega_{\rm uq}}. \tag{3.9}$$

*If*

$$\delta < \min\left\{\frac{\alpha g\,\omega_{\rm uq}}{\kappa_2(Q)}, \frac{(1-2\alpha)g\,\omega_{\rm nd}}{\sqrt{2}}\right\}, \tag{3.10}$$

*then for* $\mathrm{p}\in\{2,\mathrm{F}\}$

$$\begin{aligned}\min_{\substack{D\in\mathbb{D}_{\tau_n}, D^{\rm T}D=I\\ \Pi\in\mathbb{P}_{\tau_n}}} \frac{\|W-\widetilde{W}D\Pi\|_{\rm p}}{\|\widetilde{W}\|_{\rm p}} &\le \frac{1+\sqrt{t}\,\epsilon_*}{\sqrt{1-2\sqrt{t-1}\epsilon_* - (t-1)\epsilon_*^2}} - 1 \\ &= \frac{\tau}{\alpha}\cdot\frac{(\sqrt{t}+\sqrt{t-1})\kappa_2(Q)\delta}{g\,\omega_{\rm uq}} + O(\delta^2) := \varepsilon_{\rm ub}.\end{aligned} \tag{3.11}$$

In what follows, we first look at two illustrating examples, then discuss the implications of Theorem 3.3.

**Example 3.1.** Let $A_1 = I_2$, $A_2 = \mathrm{diag}(1, 1+\varsigma)$, where $\varsigma > 0$ is a parameter. It is obvious that $W = I_2$ is a diagonalizer of $\mathcal{A} = \{A_1, A_2\}$ with respect to $\tau_2 = (1,1)$. By calculations, we get

$$\omega_{\rm uq} = \sqrt{\varsigma^2 + 2\varsigma + 4 - (\varsigma+2)\sqrt{\varsigma^2+4}} = \frac{\varsigma}{\sqrt{2}} + O(\varsigma^{3/2}), \quad \omega_{\rm nd} = \infty.$$

[3]Recall that $t \ge 2$. The quantity $\tau$ decreases as $t$ increases and thus $\tau \le \sqrt{2}-1$. Since $\alpha$ increases as $\tau$ does, $\alpha$ decreases as $t$ increases and thus $\alpha \le 2(\sqrt{2}-1)/(2\sqrt{2}-1)^2 < 1/4$.

Perturb $\mathcal{A}$ to $\widetilde{\mathcal{A}} = \{\widetilde{A}_1, \widetilde{A}_2\}$, where $\widetilde{A}_1 = A_1 + \epsilon E$ and $\widetilde{A}_2 = A_2 - \epsilon E$, with $E = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$, and $\epsilon \geq 0$ is a parameter for controlling the level of perturbation. Consider

$$c = \cos\theta, \quad s = \sin\theta, \quad \widetilde{W} = \begin{bmatrix} c & s \\ -s & c \end{bmatrix},$$

where $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ is a parameter that controls the quality of approximate diagonalizer $\widetilde{W}$ of $\widetilde{\mathcal{A}}$. Simple calculations give

$$\widetilde{W}^{\mathrm{T}} \widetilde{A}_1 \widetilde{W} = \begin{bmatrix} 1+\epsilon & \epsilon \\ -\epsilon & 1+\epsilon \end{bmatrix}, \quad \widetilde{W}^{\mathrm{T}} \widetilde{A}_2 \widetilde{W} = \begin{bmatrix} 1+\varsigma s^2 - \epsilon & -\epsilon - \varsigma cs \\ \epsilon - \varsigma cs & 1 + \varsigma c^2 - \epsilon \end{bmatrix}$$

from which we can see that if $\theta$ and $\epsilon$ are sufficiently small, $\widetilde{W}$ is a good block diagonalizer. Now let $\Gamma = \operatorname{diag}(-1, 1)$. Simple calculations give

$$g = 2, \quad \kappa_2(Q) = 1, \quad \tilde{r} = \sqrt{16\epsilon^2 + 8\varsigma^2 c^2 s^2}, \quad \delta_{\mathcal{A}} = 2\sqrt{2}\epsilon, \quad \delta = \tilde{r} + 2\delta_{\mathcal{A}}.$$

Thus, if $\theta = \epsilon$ and $\epsilon \ll 1$, then (3.10) is satisfied. By Theorem 3.3, we have (3.11) which gives

$$\min_{D,\Pi} \frac{\|W - \widetilde{W} D\Pi\|_{\mathrm{p}}}{\|\widetilde{W}\|_{\mathrm{p}}} = 2\sin\frac{\theta}{2} \approx \epsilon, \quad \varepsilon_{\mathrm{ub}} \approx \frac{(1+5\sqrt{2})(\sqrt{16+8\varsigma^2} + 4\sqrt{2})\epsilon}{4\omega_{\mathrm{uq}}}.$$

Therefore, as long as $\varsigma$ is not too small, $\omega_{\mathrm{uq}}$ is not small, and then $\varepsilon_{\mathrm{ub}} = O(\epsilon)$, i.e., the relative error in $\widetilde{W}$ and the upper bound $\varepsilon_{\mathrm{ub}}$ have the same order of magnitude. However, if $\epsilon \ll 1$ and $\varsigma$ is small, say $\varsigma = \epsilon^{\phi}$ with $0 < \phi < 1$, then $\widetilde{W}$ is always a good block diagonalizer, independent of $\theta$, in the sense that $\tilde{r}$ is always small. But now we have $\varepsilon_{\mathrm{ub}} = O(\epsilon^{1-\phi})$, which does not provide a sharp upper bound for the relative error in $\widetilde{W}$.

**Example 3.2.** Let $A_1 = \operatorname{diag}(I_2, \begin{bmatrix} 1 & 1+\varsigma \\ 1 & 1 \end{bmatrix})$, $A_2 = \operatorname{diag}(\begin{bmatrix} 1 & 1+\varsigma \\ 1 & 1 \end{bmatrix}, I_2)$, where $\varsigma > 0$ is a parameter. Then $W = I_4$ is a $\tau_4$-block diagonalizer of $\mathcal{A} = \{A_1, A_2\}$, where $\tau_4 = (2, 2)$. By calculations, we have

$$\omega_{\mathrm{uq}} \approx 0.5858 + O(\varsigma), \quad \omega_{\mathrm{nd}} = \varsigma.$$

Perturb $\mathcal{A}$ to $\widetilde{\mathcal{A}} = \{\widetilde{A}_1, \widetilde{A}_2\}$, where $\widetilde{A}_1 = A_1 + \epsilon E$, $\widetilde{A}_2 = A_2 - \epsilon E$, where $E$ is a 4-by-4 matrix of all ones and $\epsilon \geq 0$. Consider

$$U = \operatorname{diag}\Big(\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\Big), \quad \widetilde{W} = U \operatorname{diag}\Big(1, \begin{bmatrix} c & s \\ -s & c \end{bmatrix}, 1\Big),$$

where $c = \cos\theta$, $s = \sin\theta$, and $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Then

$$\sum_{i=1}^{2} \big\| \operatorname{OffBdiag}_{\tau_n}(\widetilde{W}^{\mathrm{T}} \widetilde{A}_i \widetilde{W}) \big\|_{\mathrm{F}}^2 = 4s^2c^2(2+\varsigma)^2 + 4\varsigma^2 s^2 + 16(1+s^2)c^2\epsilon^2.$$

Therefore, if $\theta$ and $\epsilon$ are sufficiently small, then $\widetilde{W}$ is a good block diagonalizer. Now let $\Gamma = \operatorname{diag}(-I_2, I_2)$. By simple calculations, we get

$$g = 2, \quad \kappa_2(Q) = 1, \quad \delta_{\mathcal{A}} = 4\sqrt{2}\epsilon, \quad \delta = \tilde{r} + 2\delta_{\mathcal{A}},$$
$$\tilde{r} = 2\sqrt{4s^2c^2(2+\varsigma)^2 + 4\varsigma^2 s^2 + 16(1+s^2)c^2\epsilon^2}.$$

If $\theta = \epsilon \ll 1$ and $\varsigma$ is not too small, then (3.10) is satisfied and so does (3.11) by Theorem 3.3. We have

$$\min_{D,\Pi} \frac{\|W - \widetilde{W} D\Pi\|_{\mathrm{p}}}{\|\widetilde{W}\|_{\mathrm{p}}} = 2\sin\frac{\theta}{2} \approx \epsilon, \quad \varepsilon_{\mathrm{ub}} \approx \frac{(1+5\sqrt{2})\delta}{4\omega_{\mathrm{uq}}} = O(\epsilon),$$

i.e., the relative error in $\widetilde{W}$ and the upper bound $\varepsilon_{\mathrm{ub}}$ have the same order of magnitude. However, if $\theta = \frac{\pi}{2} - \epsilon$ with $\epsilon \ll 1$ and $\varsigma$ is small, say $\varsigma = \epsilon^{\phi}$ with $\phi > 0$, then the condition (3.10) of Theorem 3.3 is likely violated, and consequently, Theorem 3.3 is no longer applicable.

From the above two examples, we can see that the bound $\varepsilon_{\mathrm{ub}}$ in (3.11) is *sharp* in the sense that it can be in the same order of magnitude as the relative error. But when $\omega_{\mathrm{uq}}$ and/or $\omega_{\mathrm{nd}}$ is small, Theorem 3.3 may not provide a sharp bound or even fails to give a bound. This observation is more or less expected. In fact, when $\omega_{\mathrm{uq}}$ and/or $\omega_{\mathrm{nd}}$ is small, the JBDP for $\mathcal{A}$ can be thought of as an ill-conditioned problem in the sense that any small perturbation can result in huge error in the solution.

When solving an O-JBDP, diagonalizers $W$, $\widetilde{W}$ are orthogonal, and thus $\delta = \tilde{r} + 2\delta_{\mathcal{A}}$. Theorem 3.3 yields

**Corollary 3.4.** *In* Theorem 3.3*, if $W$ and $\widetilde{W}$ are assumed orthogonal, then*

$$\min_{\substack{D\in\mathbb{D}_{\tau_n}, D^{\mathrm{T}}D=I \\ \Pi\in\mathbb{P}_{\tau_n}}} \frac{\|W - \widetilde{W} D\Pi\|_{\mathrm{p}}}{\|\widetilde{W}\|_{\mathrm{p}}} \le \frac{\tau}{\alpha}\cdot\frac{(\sqrt{t}+\sqrt{t-1})\delta}{g\,\omega_{\mathrm{uq}}} + O(\delta^2). \tag{3.12}$$

Some of the quantities in the right-hand side of (3.11) are not computable, unless $W$ is known. But it can still be useful in assessing roughly how good the approximate bock diagonalizer $\widetilde{W}$ may be. Suppose that $\tilde{r}$ is sufficiently tiny. Then it is plausible to assume $\|Q^{-1}\|_2 = O(1)$. The moduli $\omega_{\mathrm{uq}}$ and $\omega_{\mathrm{nd}}$ which are intrinsic to the JBDP for $\mathcal{A}$ may well be estimated by those of $\widehat{\mathcal{A}} = \big\{\operatorname{Bdiag}_{\tau_n}(\widetilde{W}^{\mathrm{T}}\widetilde{A}\widetilde{W})\big\}_{i=1}^{m}$. Finally, for $W \in \mathbb{W}_{\tau_n}$

$$1 \le \|W\|_2 \le \sqrt{t}. \tag{3.13}$$

The same holds for $\widetilde{W}$, too. We will justify (3.13) after Lemma 4.4 in section 4 in order to use some of the techniques arising in its proof.

*Remark* 3.5. Several comments are in order.

(a) The quantity $\delta$ in (3.9) consists of two parts: the first part indicates how good the block diagonalizer $\widetilde{W}$ is, and the second part indicates how large the perturbation is. Therefore, the condition (3.10) means that the block diagonalizer $\widetilde{W}$ has to be sufficiently good and the perturbation has to be sufficiently small so that $\delta$ does not exceed the right-hand side of (3.10), which is proportional to the modulus of uniqueness $\omega_{\mathrm{uq}}$. Although the modulus of non-divisibility $\omega_{\mathrm{nd}}$ does not appear explicitly in the upper bound, it limits the size of $\delta$.

(b) In (3.11), $\varepsilon_{\mathrm{ub}}$ is a monotonic increasing function in $\delta$ and $\kappa_2(Q)$. If $W$ (or $\widetilde{W}$) is ill-conditioned, then both $\delta$ and $\kappa_2(Q)$ can be large, as a result, $\varepsilon_{\mathrm{ub}}$ can be large.

(c) If $\delta \ll 1$, by (3.11), we have

$$\min_{D,\Pi} \frac{\|W - \widetilde{W} D\Pi\|_{\mathrm{p}}}{\|\widetilde{W}\|_{\mathrm{p}}} \le \frac{\tau}{\alpha} \cdot \frac{(\sqrt{t} + \sqrt{t-1})\kappa_2(Q)}{\omega_{\mathrm{uq}}} \cdot \frac{\delta}{g} + O(\delta^2). \tag{3.14}$$

(d) A natural assumption when performing a perturbation analysis for JBDP is to assume that both the original matrix set $\mathcal{A}$ and its perturbed one $\widetilde{\mathcal{A}}$ admit exact diagonalizers, i.e., both JBDPs are solvable. Theorem 3.3 covers such a scenario as a special case with $\tilde{r} = 0$.

Theorem 3.3, as a perturbation theorem for JBDP, can be used to yield an error bound for an approximate block diagonalizer of a block diagonalizable $\mathcal{A}$ by simply letting all $\widetilde{A}_i = A_i$, i.e., $\delta_{\mathcal{A}} = 0$. In fact, when $\delta_{\mathcal{A}} = 0$, $\delta = \|Q^{-1}\|_2^2\, \tilde{r}$. If also $\tilde{r} \ll 1$, then $\delta \ll 1$ and thus by (3.14)

$$\min_{D,\Pi} \frac{\|W - \widetilde{W} D\Pi\|_{\mathrm{p}}}{\|\widetilde{W}\|_{\mathrm{p}}} \le \frac{\tau}{\alpha} \cdot \frac{(\sqrt{t} + \sqrt{t-1})\kappa_2(Q)\|Q^{-1}\|_2^2}{\omega_{\mathrm{uq}}} \cdot \frac{\tilde{r}}{g} + O(\tilde{r}^2). \tag{3.15}$$

This error bound is $O(\frac{\tilde{r}}{\omega_{\mathrm{uq}}})$, which in agreement with the error bound when applied to JDP in [22, Corollary 3.2].

3.3. **Condition Number.** A widely accepted way to define condition number is through some kind of first order expansion. To explain the idea, we use the explanation in [13, p.4] for a real-valued differentiable function $f(x)$ of real variable $x$. Now if $x$ is perturbed to $x + \delta x$, we have, to the first order,

$$\frac{|f(x+\delta x) - f(x)|}{|f(x)|} \approx \frac{|f'(x)| \cdot |x|}{|f(x)|} \cdot \frac{|\delta x|}{|x|}.$$

In words, this says that the relative change to the function value $f(x)$ is about the relative change to the input $x$ magnified by the factor $|f'(x)| \cdot |x|/|f(x)|$ which defines the (relative) condition number of $f(x)$ at $x$. A prerequisite for this line of definition is that $f$ is well-defined in some neighborhood of $x$.

In generalizing this framework to more broad content. The above scalar-valued function $f$ is translated into some mapping that maps inputs which are usually much more general than a single scalar $x$ to some output. In the context of JBDP, naturally the input is the set $\mathcal{A}$ of matrices and the output is the block diagonalizer $W$. But then the framework does not work because any generic and arbitrarily small perturbation to $\mathcal{A}$ will render one that is not $\tau_n$-block diagonalizable, i.e., the mapping that takes in $\mathcal{A}$ is not well-defined in any neighborhood of $\mathcal{A}$.

We have to seek some other way. Recall the rule of thumb:

$$\text{forward error} \lesssim \text{condition number} \times \text{backward error}.$$

We will use this as a guideline. Consider $\mathcal{A}$ and $\widetilde{\mathcal{A}}$ which is some tiny perturbation away from $\mathcal{A}$ and suppose both are $\tau_n$-block diagonalizable with $\tau_n$-block diagonalizer $W$ and $\widetilde{W}$ from $\mathbb{W}_{\tau_n}$, respectively. Apply Theorem 3.3 with $\tilde{r} = 0$ and sufficiently tiny $\delta_{\mathcal{A}}$ to get

$$\begin{aligned}&\min_{\substack{D\in\mathbb{D}_{\tau_n},D^{\mathrm{T}}D=I\\ \Pi\in\mathbb{P}_{\tau_n}}}\frac{\|W-\widetilde{W}D\Pi\|_{\mathrm{p}}}{\|\widetilde{W}\|_{\mathrm{p}}}\\ &\qquad\lesssim\frac{\tau}{\alpha}\cdot\frac{(\sqrt{t}+\sqrt{t-1})\kappa_2(Q)\|Q^{-1}\|_2\|W\|_2\|\widetilde{W}\|_2\|\mathcal{A}\|_{\mathrm{F}}}{g\,\omega_{\mathrm{uq}}}\cdot\frac{\delta_{\mathcal{A}}}{\|\mathcal{A}\|_{\mathrm{F}}}.\end{aligned}$$

Thinking about as $\widetilde{\mathcal{A}}$ goes to $\mathcal{A}$, we may let $\widetilde{W}$ go to $W$ and the right-hand side approaches to

$$\frac{\tau}{\alpha}\cdot\frac{(\sqrt{t}+\sqrt{t-1})\|W\|_2^2\|\mathcal{A}\|_{\mathrm{F}}}{g\,\omega_{\mathrm{uq}}}\cdot\frac{\delta_{\mathcal{A}}}{\|\mathcal{A}\|_{\mathrm{F}}}.$$

which suggests that we may define the $\tau_n$-*condition number* of JBDP for $\mathcal{A}$ as

$$\operatorname{cond}(\mathcal{A})=\frac{\tau}{\alpha}\cdot\frac{(\sqrt{t}+\sqrt{t-1})\|W\|_2^2\|\mathcal{A}\|_{\mathrm{F}}}{\omega_{\mathrm{uq}}}. \tag{3.16}$$

A few remarks are in order for this condition number $\operatorname{cond}(\mathcal{A})$.

(a) As it appears, the right-hand side of (3.16) depends on the $\tau_n$-block diagonalizer $W\in\mathbb{W}_{\tau_n}$. But it isn't. This is because $\omega_{\mathrm{uq}}$ is independent of the choice of the block diagonalizer $W\in\mathbb{W}_{\tau_n}$ (Theorem 2.6) and so is $\|W\|_2$ (Lemma 3.6 below).

(b) Given $\beta\neq 0$, let $\beta\mathcal{A}=\{\beta A_i\}_{i=1}^m$. It can be seen that $\operatorname{cond}(\mathcal{A})=\operatorname{cond}(\beta\mathcal{A})$, i.e., the condition number $\operatorname{cond}(\mathcal{A})$ is scale invariant.

(c) Suppose $\|A_i\|_{\mathrm{F}}=1$ for $i=1,2,\ldots,m$ and consider the condition number $\operatorname{cond}(\widehat{\mathcal{A}})$ of the JBDP for $\widehat{\mathcal{A}}=\{\beta_iA_i\}_{i=1}^m$, where $\beta_j$'s are positive real numbers. Recall the definition of $G_{jk}$ in (2.8) and the definition of $\omega_{\mathrm{uq}}$. $W$, as a $\tau_n$-block diagonalizer of $\mathcal{A}$, is also one of $\widehat{\mathcal{A}}$. Now define $\widehat{G}_{jk}$ for $\widehat{\mathcal{A}}$, similarly to $G_{jk}$ for $\mathcal{A}$. We have

$$\widehat{G}_{jk}=\big[\operatorname{diag}(\beta_1,\ldots,\beta_m)\otimes I_{2n_jn_k}\big]G_{jk}. \tag{3.17}$$

Let $\beta_{\max}=\max_{1\le j\le t}\beta_j$ and $\beta_{\min}=\min_{1\le j\le t}\beta_j$. We have $\sigma_{\min}(\widehat{G}_{jk})\ge\beta_{\min}\sigma_{\min}(G_{jk})$. Thus, $\hat{\omega}_{\mathrm{uq}}:=\omega_{\mathrm{uq}}(\widehat{\mathcal{A}})\ge\beta_{\min}\omega_{\mathrm{uq}}$. Therefore

$$\operatorname{cond}(\widehat{\mathcal{A}})=\frac{\tau}{\alpha}\cdot\frac{(\sqrt{t}+\sqrt{t-1})\|W\|_2^2\left(\sum_{i=1}^m\|\beta_iA_i\|_{\mathrm{F}}^2\right)^{1/2}}{\hat{\omega}_{\mathrm{uq}}}\le\frac{\beta_{\max}}{\beta_{\min}}\operatorname{cond}(\mathcal{A}). \tag{3.18}$$

As an upper bound of $\operatorname{cond}(\widehat{\mathcal{A}})$, the right hand side of (3.18) is minimized if all $\beta_j$'s are equal. This tells us that when solving JBDP, it would be a good idea to first normalize each $A_i$ to have $\|A_i\|_{\mathrm{F}}=1$ for all $i$.

(d) It is easy to see that the modulus of uniqueness $\omega_{\mathrm{uq}}$ is an monotonic increasing function of the number of matrices in $\mathcal{A}$. How it affects the condition number $\operatorname{cond}(\mathcal{A})$ is in general unclear. In our numerical tests in section 5, as we put more matrices into the matrix set $\mathcal{A}$, the condition number $\operatorname{cond}(\mathcal{A})$ first decreases then remains almost unchanged.

(e) Compared with the condition number $\operatorname{cond}_\lambda$ introduced in [22] for JDP only, our condition number here is about the square root of $\operatorname{cond}_\lambda$ there, and thus more realistic.

**Lemma 3.6.** *For any two $W, \widetilde{W} \in \mathbb{W}_{\tau_n}$, if $\widetilde{W} = WD\Pi$ for some $D \in \mathbb{D}_{\tau_n}$ and $\Pi \in \mathbb{P}_{\tau_n}$, then $D$ is orthogonal and, as a result, $\|\widetilde{W}\|_2 = \|W\|_2$.*

*Proof.* Since $D \in \mathbb{D}_{\tau_n}$, $D = \operatorname{diag}(D_1, D_2, \ldots, D_t)$ with $D_j \in \mathbb{R}^{n_j \times n_j}$. It suffices to show each $D_j$ is orthogonal. Write $W = [W_1, W_2, \ldots, W_t]$ and $\widetilde{W} = [\widetilde{W}_1, \widetilde{W}_2, \ldots, \widetilde{W}_t]$, where $W_j$, $\widetilde{W}_j \in \mathbb{R}^{n \times n_j}$. Because $W$, $\widetilde{W} \in \mathbb{W}_{\tau_n}$ by assumption, we have

$$W_j^{\mathrm{T}} W_j = I_{n_j}, \quad \operatorname{Bdiag}_{\tau_n}(\widetilde{W}^{\mathrm{T}} \widetilde{W}) = I_n.$$

Notice $\operatorname{Bdiag}_{\tau_n}(\Pi^{\mathrm{T}} D^{\mathrm{T}} W^{\mathrm{T}} W D \Pi) = \operatorname{Bdiag}_{\tau_n}(D^{\mathrm{T}} W^{\mathrm{T}} W D)$ to get

$$I_{n_j} = D_j^{\mathrm{T}} W_j^{\mathrm{T}} W_j D_j = D_j^{\mathrm{T}} D_j,$$

i.e., $D_j$ is orthogonal for all $j$, as expected. □

Thus, if a JBDP is uniquely $\tau_n$-block diagonalizable, then all $\tau_n$-block diagonalizers in $\mathbb{W}_{\tau_n}$ can be written in the form $WD\Pi$, where $W \in \mathbb{W}_{\tau_n}$ is a particular $\tau_n$-block diagonalizer, $D \in \mathbb{D}_{\tau_n}$ is orthogonal and $\Pi \in \mathbb{P}_{\tau_n}$.

## 4. Proof of Theorem 3.3

Recall the assumptions: $\mathcal{A} = \{A_i\}_{i=1}^m$ is $\tau_n$-block diagonalizable and $W \in \mathbb{W}_{\tau_n}$ is a $\tau_n$-block diagonalizer such that (1.3) holds. The modulus of uniqueness $\omega_{\mathrm{uq}}$ and the modulus of non-divisibility $\omega_{\mathrm{nd}}$ for the block diagonalization of $\mathcal{A}$ by $W$ are defined by Definition 2.3. The perturbed matrix set is $\widetilde{\mathcal{A}} = \{\widetilde{A}_i\}_{i=1}^m$ and $\widetilde{W}$ is an approximate $\tau_n$-block diagonalizer of $\widetilde{\mathcal{A}}$. $\Gamma = \operatorname{diag}(\gamma_1 I_{n_1}, \ldots, \gamma_t I_{n_t})$, where $\gamma_1, \ldots, \gamma_t$ are distinct real numbers with all $|\gamma_j| \le 1$, and $\widetilde{R}_i$ are defined by (3.4).

### 4.1. Three Lemmas.

The three lemmas in this subsection may have interest of their own, although their roles here are to assist the proof of Theorem 3.3.

**Lemma 4.1.** *For given $Z \in \mathbb{R}^{n \times n}$, denote by*

$$R_i = \operatorname{diag}(A_i^{(11)}, \ldots, A_i^{(tt)}) Z - Z^{\mathrm{T}} \operatorname{diag}(A_i^{(11)}, \ldots, A_i^{(tt)}) \tag{4.1}$$

*for $1 \le i \le m$. Partition $Z = \begin{bmatrix} Z_{jk} \end{bmatrix}$ with $Z_{jk} \in \mathbb{R}^{n_j \times n_k}$ and let $\lambda(Z_{jj}) = \{\mu_{jk}\}_{k=1}^{n_j}$.*

(a) *If $\omega_{\mathrm{uq}} > 0$, then*

$$\|\operatorname{OffBdiag}_{\tau_n}(Z)\|_{\mathrm{F}}^2 \le \frac{\sum_{i=1}^m \|\operatorname{OffBdiag}_{\tau_n}(R_i)\|_{\mathrm{F}}^2}{\omega_{\mathrm{uq}}^2}. \tag{4.2}$$

(b) *If $\dim \mathscr{N}(\mathcal{A}_j) = 1$, then there exists a real number $\hat{\mu}_j$ such that*

$$\sum_{k=1}^{n_j} |\mu_{jk} - \hat{\mu}_j|^2 \le \frac{\sum_{i=1}^m \|\operatorname{Bdiag}_{\tau_n}(R_i)\|_{\mathrm{F}}^2}{\omega_{\mathrm{nd}}^2}. \tag{4.3}$$

*Proof.* Partition $R_i = \begin{bmatrix} R_i^{(jk)} \end{bmatrix}$ conformally with respect to $\tau_n$. First, we show (4.2). For any pair $(j, k)$ with $j < k$, it follows from (4.1) that

$$G_{jk} \begin{bmatrix} \operatorname{vec}(Z_{jk}) \\ -\operatorname{vec}(Z_{kj}^{\mathrm{T}}) \end{bmatrix} = \begin{bmatrix} \operatorname{vec}(R_1^{(jk)}) \\ -\operatorname{vec}((R_1^{(kj)})^{\mathrm{T}}) \\ \vdots \\ \operatorname{vec}(R_m^{(jk)}) \\ -\operatorname{vec}((R_m^{(kj)})^{\mathrm{T}}) \end{bmatrix} =: r_{jk},$$

where $G_{jk}$ is defined by (2.8). Put them all together to get

$$M_{\rm uq} z_{\rm uq} = r_{\rm uq},$$

where

$$\begin{aligned}
M_{\rm uq} &= \operatorname{diag}\big(G_{12},\dots,G_{1t},G_{23},\dots,G_{2t},\dots,G_{t-1,t}\big),\\
z_{\rm uq} &= \big[\operatorname{vec}(Z_{12})^{\rm T},\,-\operatorname{vec}(Z_{21}^{\rm T})^{\rm T},\dots,\operatorname{vec}(Z_{1t})^{\rm T},-\operatorname{vec}(Z_{t1}^{\rm T})^{\rm T},\\
&\qquad \operatorname{vec}(Z_{23})^{\rm T},-\operatorname{vec}(Z_{32}^{\rm T})^{\rm T},\dots,\operatorname{vec}(Z_{2t})^{\rm T},-\operatorname{vec}(Z_{t2}^{\rm T})^{\rm T},\dots,\\
&\qquad \operatorname{vec}(Z_{t-1,t})^{\rm T},\operatorname{vec}(Z_{t,t-1}^{\rm T})^{\rm T}\big]^{\rm T},\\
r_{\rm uq} &= \big[r_{12}^{\rm T},\dots,r_{1t}^{\rm T},r_{23}^{\rm T},\dots,r_{2t}^{\rm T},\dots,r_{t-1,t}^{\rm T}\big]^{\rm T}.
\end{aligned}$$

We have $\sigma_{\min}(M_{\rm uq}) = \min_{j<k}\sigma_{\min}(G_{jk}) = \omega_{\rm uq} > 0$, and thus

$$\|\operatorname{OffBdiag}_{\tau_n}(Z)\|_{\rm F}^2 = \|z_{\rm uq}\|_2^2 \le \frac{\|r_{\rm uq}\|_2^2}{\omega_{\rm uq}^2} = \frac{\sum_{i=1}^m \|\operatorname{OffBdiag}_{\tau_n}(R_i)\|_{\rm F}^2}{\omega_{\rm uq}^2},$$

as expected. Next, we show (4.3). For $j = k$, using (4.1), we have

$$G_{jj}\operatorname{vec}(Z_{jj}) = \begin{bmatrix}\operatorname{vec}(R_1^{(jj)})\\ \vdots \\ \operatorname{vec}(R_m^{(jj)})\end{bmatrix} =: r_{jj},$$

where $G_{jj}$ is defined by (2.10). Since $\dim\mathscr{N}(\mathcal{A}_j) = 1$ by assumption, we know that the null space of $G_{jj}$ is spanned by $\operatorname{vec}(I_{n_j})$, and thus there exists a real number $\hat\mu_j$ such that

$$\operatorname{vec}(Z_{jj}) = G_{jj}^{\dagger} r_{jj} + \hat\mu_j \operatorname{vec}(I_{n_j}),$$

where $G_{jj}^{\dagger}$ is the Moore-Penrose inverse [26, p.102] of $G_{jj}$. It follows immediately that

$$Z_{jj} = \widehat Z_{jj} + \hat\mu_j I_{n_j},$$

where $\widehat Z_{jj} = \operatorname{reshape}(G_{jj}^{\dagger} r_{jj}, n_j, n_j)$. In particular, $\lambda(\widehat Z_{jj}) = \{\mu_{jk} - \hat\mu_j\}_{k=1}^{n_j}$ and hence

$$\sum_{k=1}^{n_j} |\mu_{jk} - \hat\mu_j|^2 \le \|\widehat Z_{jj}\|_{\rm F}^2 \le \frac{\|r_{jj}\|_2^2}{\omega_{\rm nd}^2} \le \frac{\sum_{i=1}^m \|R_i^{(jj)}\|_{\rm F}^2}{\omega_{\rm nd}^2} \le \frac{\sum_{i=1}^m \|\operatorname{Bdiag}_{\tau_n}(R_i)\|_{\rm F}^2}{\omega_{\rm nd}^2}.$$

This completes the proof. □

Previously in Theorem 3.3, $Q$ is set to $W^{-1}\widetilde W$, but the one in the next lemma can be any given nonsingular matrix.

**Lemma 4.2.** *For any given nonsingular $Q \in \mathbb{R}^{n\times n}$, let $Z = Q\Gamma Q^{-1}$ and write $Z = B - E$ with $B = \operatorname{Bdiag}_{\tau_n}(Z)$ and $E = -\operatorname{OffBdiag}_{\tau_n}(Z)$. Let $\tau$ and $\alpha$ be as in (3.8). If*

$$g > \|Q^{-1}EQ\|_{\rm F}/\alpha, \tag{4.4}$$

*then there exists a $\tau_n$-block diagonal matrix $\widetilde B = \operatorname{diag}(\widetilde B_{11},\dots,\widetilde B_{tt})$ and a nonsingular matrix $P = \big[P_{jk}\big]$ with $P_{jk} \in \mathbb{R}^{n_j\times n_k}$ and $P_{jj} = I_{n_j}$ such that*

$$B(QP) = (QP)\widetilde B, \tag{4.5}$$

*and for* $j = 1, 2, \ldots, t$

$$\|\widehat{P}_j\|_{\rm F} \le \frac{\tau}{\alpha} \cdot \frac{\|Q^{-1}EQ\|_{\rm F}}{g}, \tag{4.6a}$$

$$\sum_{k=1}^{n_j} |\tilde{\mu}_{jk} - \gamma_j|^2 < (1+\tau^2) \cdot \|Q^{-1}EQ\|_{\rm F}^2, \tag{4.6b}$$

*where* $\tilde{\mu}_{j1}, \ldots, \tilde{\mu}_{jn_j}$ *are the eigenvalues of* $\widetilde{B}_{jj}$*, and*

$$\widehat{P}_j = \begin{bmatrix} P_{1j}^{\rm T}, & \ldots, & P_{j-1,j}^{\rm T}, & 0_{n_j \times n_j}, & P_{j+1,j}^{\rm T}, & \ldots, & P_{tj}^{\rm T} \end{bmatrix}^{\rm T}. \tag{4.6c}$$

*Proof.* It suffices to show there exist $\widehat{P}_1 \in \mathbb{R}^{n \times n_1}$ and $\widetilde{B}_{11} \in \mathbb{R}^{n_1 \times n_1}$ such that

$$Q^{-1}BQ \begin{bmatrix} I_{n_1} \\ \widehat{P}_1 \end{bmatrix} \equiv (\Gamma + Q^{-1}EQ) \begin{bmatrix} I_{n_1} \\ \widehat{P}_1 \end{bmatrix} = \begin{bmatrix} I_{n_1} \\ \widehat{P}_1 \end{bmatrix} \widetilde{B}_{11}, \tag{4.7}$$

(4.6) for $j = 1$ holds, and $P$ is nonsingular.

Partition $Q^{-1}EQ = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix}$ with $E_{11} \in \mathbb{R}^{n_1 \times n_1}$, $E_{22} \in \mathbb{R}^{(n-n_1)\times(n-n_1)}$. A direct calculation gives

$$\operatorname{sep}_{\rm F}(\gamma_1 I_{n_1}, \operatorname{diag}(\gamma_2 I_{n_2}, \ldots, \gamma_t I_{n_t})) = \min_{2 \le j \le t} |\gamma_j - \gamma_1| \ge g,$$

where $\operatorname{sep}_{\rm F}(\cdots)$ is the separation of two matrices [26, p.247]. Let $\tilde{g} = g - \|E_{11}\|_{\rm F} - \|E_{22}\|_{\rm F}$. By [26, Theorem 2.8 on p.238], we conclude that if

$$\tilde{g} > 0, \quad \frac{\|E_{21}\|_{\rm F}\|E_{12}\|_{\rm F}}{\tilde{g}^2} < \frac{1}{4}, \tag{4.8}$$

then there is a unique $\widehat{P}_1 \in \mathbb{R}^{(n-n_1)\times n_1}$ such that

$$\|\widehat{P}_1\|_{\rm F} \le \frac{2\|E_{21}\|_{\rm F}}{\tilde{g} + \sqrt{\tilde{g}^2 - 4\|E_{21}\|_{\rm F}\|E_{12}\|_{\rm F}}} \tag{4.9}$$

and (4.7) holds. We have to show that the assumption (4.4) ensures (4.8) and that (4.9) implies (4.6a) for $j = 1$. In fact, under (4.4),

$$\begin{aligned} \tilde{g} &\ge g - \sqrt{2(\|E_{11}\|_{\rm F}^2 + \|E_{22}\|_{\rm F}^2)} \\ &\ge g - \sqrt{2}\|Q^{-1}EQ\|_{\rm F} \\ &> (1 - \sqrt{2}\,\alpha) g \\ &> 0, \end{aligned} \tag{4.10}$$

$$\begin{aligned} \frac{\|E_{21}\|_{\rm F}\|E_{12}\|_{\rm F}}{\tilde{g}^2} &\le \frac{\|E_{21}\|_{\rm F}^2 + \|E_{12}\|_{\rm F}^2}{2\tilde{g}^2} \\ &< \frac{\|E_{21}\|_{\rm F}^2 + \|E_{12}\|_{\rm F}^2}{2(1-\sqrt{2}\,\alpha)^2 g^2} \\ &\le \frac{\|Q^{-1}EQ\|_{\rm F}^2}{2(1-\sqrt{2}\,\alpha)^2 g^2} \\ &\le \frac{\alpha^2}{2(1-\sqrt{2}\,\alpha)^2} \\ &< \frac{1}{4}. \end{aligned} \tag{4.11}$$

They give (4.8). It follows from (4.9), (4.10), and (4.11) that

$$\begin{aligned}\|\widehat{P}_1\|_{\mathrm{F}} &\le \frac{2}{(1-\sqrt{2}\,\alpha)+\sqrt{(1-\sqrt{2}\,\alpha)^2-2\alpha^2}}\cdot\frac{\|Q^{-1}EQ\|_{\mathrm{F}}}{g}\\ &= \frac{\tau}{\alpha}\cdot\frac{\|Q^{-1}EQ\|_{\mathrm{F}}}{g} \qquad (4.12)\\ &< \tau.\end{aligned}$$

The inequality (4.6a) for $j=1$ is a result of (4.12).

Next we show (4.6b) for $j=1$. Pre-multiply (4.7) by $[I_{n_1},0]$ to get, after rearrangement,

$$\widetilde{B}_{11}-\gamma_1 I_{n_1} = [I_{n_1},0]Q^{-1}EQ\begin{bmatrix} I_{n_1}\\ P_1\end{bmatrix}.$$

Since $\lambda(\widetilde{B}_{11}) = \{\tilde{\mu}_{1k}\}_{k=1}^{n_1}$, we have

$$\begin{aligned}\sum_{k=1}^{n_1}|\tilde{\mu}_{1k}-\gamma_1|^2 &\le \left\|[I_{n_1}\ 0]Q^{-1}EQ\begin{bmatrix} I_{n_1}\\ \widehat{P}_1\end{bmatrix}\right\|_{\mathrm{F}}^2\\ &\le \left\|\begin{bmatrix} I_{n_1}\\ \widehat{P}_1\end{bmatrix}\right\|_2^2\|Q^{-1}EQ\|_{\mathrm{F}}^2\\ &\le (1+\|\widehat{P}_1^{\mathrm{T}}\widehat{P}_1\|_2)\|Q^{-1}EQ\|_{\mathrm{F}}^2\\ &\le (1+\tau^2)\cdot\|Q^{-1}EQ\|_{\mathrm{F}}^2,\end{aligned}$$

as was to be shown.

Finally, we show that $P$ is nonsingular by contradiction. If $P$ were singular, let $x=[x_1^{\mathrm{T}}\ \dots\ x_t^{\mathrm{T}}]^{\mathrm{T}}$ be a nonzero vector with $x_j\in\mathbb{R}^{n_j}$ such that $Px=0$. We then have $x_j = -\sum_{\substack{k=1\\k\ne j}}^{t} P_{jk}x_k$ and thus

$$\|x_j\|_2^2 = \Big(\Big\|\sum_{\substack{k=1\\k\ne j}}^{t} P_{jk}x_k\Big\|_2\Big)^2 \le \Big(\sum_{\substack{k=1\\k\ne j}}^{t}\|P_{jk}\|_2\|x_k\|_2\Big)^2 \le (t-1)\sum_{\substack{k=1\\k\ne j}}^{t}\|P_{jk}\|_2^2\|x_k\|_2^2.$$

Therefore

$$\begin{aligned}\|x\|_2^2 = \sum_{j=1}^{t}\|x_j\|_2^2 &\le (t-1)\sum_{j=1}^{t}\sum_{\substack{k=1\\k\ne j}}^{t}\|P_{jk}\|_2^2\|x_k\|_2^2\\ &= (t-1)\sum_{k=1}^{t}\sum_{\substack{j=1\\j\ne k}}^{t}\|P_{jk}\|_2^2\|x_k\|_2^2\\ &\le (t-1)\sum_{k=1}^{t}\|\widehat{P}_k\|_{\mathrm{F}}^2\|x_k\|_2^2\\ &< (t-1)\tau^2\|x\|_2^2 < \|x\|_2^2,\end{aligned}$$

a contradiction. This completes the proof. □

*Remark* 4.3. Lemma 4.2 implies that when the off-block diagonal part of $Z$ is sufficiently small, $QP$ is the eigenvector matrix of $B=\mathrm{Bdiag}_{\tau_n}(Z)$ with $P\approx I$, and for each $j$ there are $n_j$ eigenvalues of $B$ that cluster around $\gamma_j$.

**Lemma 4.4.** *Let* $P = \begin{bmatrix} P_{jk} \end{bmatrix}$ *with* $P_{jk} \in \mathbb{R}^{n_j \times n_k}$, $P_{jj} = I_{n_j}$, *and* $\|\widehat{P}_j\|_{\mathrm{F}} \leq \epsilon$, *where* $\widehat{P}_j$ *is defined as in* (4.6c), $0 \leq \epsilon < \tau$, *and* $\tau$ *is defined by* (3.8). *Then*

$$\|P - I\|_{\mathrm{F}} \leq \sqrt{t}\,\epsilon. \tag{4.13}$$

*Furthermore, let* $W, \widetilde{W} \in \mathbb{W}_{\tau_n}$, $\widetilde{D} = \mathrm{diag}(\widetilde{D}_{11}, \ldots, \widetilde{D}_{tt}) \in \mathbb{D}_{\tau_n}$, *and* $\Pi \in \mathbb{P}_{\tau_n}$. *If* $W\widetilde{D} = \widetilde{W}P\Pi$, *then* $\widetilde{D}$ *is nonsingular and*

$$\sqrt{1 - 2\sqrt{t-1}\,\epsilon - (t-1)\epsilon^2} \leq \sigma \leq \sqrt{1 + 2\sqrt{t-1}\,\epsilon + (t-1)\epsilon^2}. \tag{4.14}$$

*for each singular value* $\sigma$ *of* $\widetilde{D}$.

*Proof.* Since $P - I = \left[\widehat{P}_1, \ldots, \widehat{P}_t\right]$, we have

$$\|P - I\|_{\mathrm{F}} = \left( \sum_{j=1}^{t} \left\|\widehat{P}_j\right\|_{\mathrm{F}}^2 \right)^{1/2} \leq \sqrt{t}\epsilon,$$

which is (4.13).

Next we show that $\widetilde{D}$ is nonsingular and (4.14) holds. Write $P = \begin{bmatrix} P_1, & \ldots, & P_t \end{bmatrix}$ with $P_j \in \mathbb{R}^{n \times n_j}$. Using $W\widetilde{D} = \widetilde{W}P\Pi$, we get

$$\widetilde{D}^{\mathrm{T}} W^{\mathrm{T}} W \widetilde{D} = \Pi^{\mathrm{T}} P^{\mathrm{T}} \widetilde{W}^{\mathrm{T}} \widetilde{W} P \Pi. \tag{4.15}$$

Since $W \in \mathbb{W}_{\tau_n}$, the $j$th diagonal blocks at both sides of (4.15) read

$$\widetilde{D}_{jj}^{\mathrm{T}} \widetilde{D}_{jj} = P_{j'}^{\mathrm{T}} \widetilde{W}^{\mathrm{T}} \widetilde{W} P_{j'}, \tag{4.16}$$

where $1 \leq j' \leq t$ as a result of the permutation $\Pi$. Partition $\widetilde{W}$ as $\widetilde{W} = [\widetilde{W}_1, \ldots, \widetilde{W}_t]$ with $\widetilde{W}_j \in \mathbb{R}^{n \times n_j}$. We infer from $\widetilde{W} \in \mathbb{W}_{\tau_n}$ that $\widetilde{W}_j^{\mathrm{T}} \widetilde{W}_j = I_{n_j}$ and $\|\widetilde{W}_j^{\mathrm{T}} \widetilde{W}_\ell\|_2 \leq 1$. To see the last inequality, we note

$$|x_j^{\mathrm{T}} \widetilde{W}_j^{\mathrm{T}} \widetilde{W}_\ell x_\ell| \leq \|\widetilde{W}_j x_j\|_2 \|\widetilde{W}_\ell x_\ell\|_2 = \|x_j\|_2 \|x_\ell\|_2 = 1 \tag{4.17}$$

for any unit vectors $x_j \in \mathbb{R}^{n_j}$ and $x_\ell \in \mathbb{R}^{n_\ell}$. Now using $P_{j'j'} = I_{n_{j'}}$ and $\|\widehat{P}_{j'}\|_{\mathrm{F}} \leq \epsilon$, we have

$$\begin{aligned}
\|P_{j'}^{\mathrm{T}} \widetilde{W}^{\mathrm{T}} \widetilde{W} P_{j'} - I_{n_{j'}}\|_{\mathrm{F}} &= \|\widetilde{W}_{j'}^{\mathrm{T}} \widetilde{W} \widehat{P}_{j'} + \widehat{P}_{j'}^{\mathrm{T}} \widetilde{W}^{\mathrm{T}} \widetilde{W}_{j'} + \widehat{P}_{j'}^{\mathrm{T}} \widetilde{W}^{\mathrm{T}} \widetilde{W} \widehat{P}_{j'}\|_{\mathrm{F}} \\
&\leq 2 \Big\| \sum_{\ell \neq j'} \widetilde{W}_{j'}^{\mathrm{T}} \widetilde{W}_\ell P_{\ell j'} \Big\|_{\mathrm{F}} + \Big\| \sum_{k \neq j'} \sum_{\ell \neq j'} P_{kj'}^{\mathrm{T}} \widetilde{W}_k^{\mathrm{T}} \widetilde{W}_\ell P_{\ell j'} \Big\|_{\mathrm{F}} \\
&\leq 2 \sum_{\ell \neq j'} \|P_{\ell j'}\|_{\mathrm{F}} + \sum_{k \neq j'} \sum_{\ell \neq j'} \|P_{kj'}\|_{\mathrm{F}} \|P_{\ell j'}\|_{\mathrm{F}} \\
&= 2 \sum_{\ell \neq j'} \|P_{\ell j'}\|_{\mathrm{F}} + \Big( \sum_{k \neq j'} \|P_{kj'}\|_{\mathrm{F}} \Big)^2 \\
&\leq 2 \Big[ (t-1) \sum_{k \neq j'} \|P_{kj'}\|_{\mathrm{F}}^2 \Big]^{1/2} + (t-1) \sum_{k \neq j'} \|P_{kj'}\|_{\mathrm{F}}^2 \\
&\leq 2\sqrt{t-1}\,\epsilon + (t-1)\epsilon^2.
\end{aligned}$$

Combining it with (4.16), we get

$$\|\widetilde{D}_{jj}^{\mathrm{T}} \widetilde{D}_{jj} - I_{n_j}\|_{\mathrm{F}} \leq 2\sqrt{t-1}\,\epsilon + (t-1)\epsilon^2 < 2\sqrt{t-1}\tau + (t-1)\tau^2 = 1,$$

which implies that $\widetilde{D}_{jj}$ is nonsingular, and for any singular value $\sigma$ of $\widetilde{D}_{jj}$, it holds that

$$-2\sqrt{t-1}\,\epsilon - (t-1)\epsilon^2 \le \sigma^2 - 1 \le 2\sqrt{t-1}\,\epsilon + (t-1)\epsilon^2.$$

The conclusion follows immediately since $\widetilde{D} \in \mathbb{D}_{\tau_n}$. $\square$

We now present a proof of (3.13). Since $\|\widetilde{W}\|_2$ is equal to the square root of the largest eigenvalue of $\widetilde{W}^{\mathrm{T}}\widetilde{W}$ and the latter is no smaller than the largest diagonal entry of $\widetilde{W}^{\mathrm{T}}\widetilde{W}$, we have $\|\widetilde{W}\|_2 \ge 1$. Let $x = \big[x_1^{\mathrm{T}}, x_2^{\mathrm{T}}, \dots, x_t^{\mathrm{T}}\big]^{\mathrm{T}}$ with $x_j \in \mathbb{R}^{n_j}$. Similarly to (4.17), we find

$$x^{\mathrm{T}}\widetilde{W}^{\mathrm{T}}\widetilde{W}x = \sum_{j,\,\ell} x_j^{\mathrm{T}}\widetilde{W}_j^{\mathrm{T}}\widetilde{W}_\ell x_\ell \le \sum_{j,\,\ell} \|x_j\|_2\|x_\ell\|_2 \le \frac{1}{2}\sum_{j,\,\ell}\big(\|x_j\|_2^2 + \|x_\ell\|_2^2\big) = t\|x\|_2^2,$$

and thus $\|\widetilde{W}\|_2 \le \sqrt{t}$.

4.2. **Proof of Theorem 3.3.** Recall $Q = W^{-1}\widetilde{W}$ and let $Z = Q\Gamma Q^{-1}$. Partition $Z = \big[Z_{jk}\big]$ with $Z_{jk} \in \mathbb{R}^{n_j\times n_k}$, and let $\lambda(Z_{jj}) = \{\mu_{jk}\}_{k=1}^{n_j}$. The proof will be completed in the following four steps:

Step 1. We will show that $Z$ is approximately $\tau_n$-block diagonal. Specifically, we show

$$\|\operatorname{OffBdiag}_{\tau_n}(Z)\|_{\mathrm{F}} \le \frac{\big(\sum_{i=1}^m \|\operatorname{OffBdiag}_{\tau_n}(R_i)\|_{\mathrm{F}}^2\big)^{1/2}}{\omega_{\mathrm{uq}}} \le \frac{\delta}{\omega_{\mathrm{uq}}}, \tag{4.18}$$

where $R_i$ is given by (4.1).

Step 2. We will show that the eigenvalues of $Z_{jj}$ cluster around a unique $\gamma_{j'}$ by showing that there exists a permutation $\pi$ of $\{1, 2, \dots, t\}$ such that

$$|\mu_{jk} - \gamma_{\pi(j)}| < \frac{g}{2}, \quad |\mu_{jk} - \gamma_i| > \frac{g}{2}, \quad \text{for any } i \ne \pi(j). \tag{4.19}$$

In the other word, each of the $t$ disjoint intervals $(\gamma_i - g/2, \gamma_i + g/2)$ contains one and only one $\lambda(Z_{jj})$.

Step 3. We will show that there exist a permutation $\Pi \in \mathbb{P}_{\tau_n}$ and a nonsingular $P \equiv \big[P_{jk}\big] \in \mathbb{R}^{n\times n}$ with $P_{jk} \in \mathbb{R}^{n_j\times n_k}$ and $P_{jj} = I_{n_j}$, satisfying (4.6a), such that $\widetilde{D} = QP\Pi \in \mathbb{D}_{\tau_n}$.

Step 4. We will prove (3.11).

*Proof* of Step 1. Recall $\widetilde{R}_i = \widetilde{W}^{\mathrm{T}}\widetilde{A}_i\widetilde{W}\Gamma - \Gamma\widetilde{W}^{\mathrm{T}}\widetilde{A}_i\widetilde{W}$ of (3.4). We have

$$\begin{aligned}\widetilde{R}_i &= \widetilde{W}^{\mathrm{T}}A_i\widetilde{W}\Gamma - \Gamma\widetilde{W}^{\mathrm{T}}A_i\widetilde{W} + \widetilde{W}^{\mathrm{T}}\Delta A_i\widetilde{W}\Gamma - \Gamma\widetilde{W}^{\mathrm{T}}\Delta A_i\widetilde{W}\\ &= Q^{\mathrm{T}}W^{\mathrm{T}}A_iWQ\Gamma - \Gamma Q^{\mathrm{T}}W^{\mathrm{T}}A_iWQ + \widetilde{W}^{\mathrm{T}}\Delta A_i\widetilde{W}\Gamma - \Gamma\widetilde{W}^{\mathrm{T}}\Delta A_i\widetilde{W},\end{aligned}$$

from which it follows that

$$\begin{aligned}R_i &= W^{\mathrm{T}}A_iWZ - Z^{\mathrm{T}}W^{\mathrm{T}}A_iW\\ &= Q^{-\mathrm{T}}\widetilde{R}_iQ^{-1} - W^{\mathrm{T}}\Delta A_i\widetilde{W}\Gamma Q^{-1} + Q^{-\mathrm{T}}\Gamma\widetilde{W}^{\mathrm{T}}\Delta A_iW.\end{aligned}$$

Putting all of them for $1 \le i \le m$ together, we get

$$\begin{bmatrix} R_1 \\ \vdots \\ R_m \end{bmatrix} = (I_m \otimes Q^{-\mathrm{T}}) \begin{bmatrix} \widetilde{R}_1 \\ \vdots \\ \widetilde{R}_m \end{bmatrix} Q^{-1} - (I_m \otimes W^{\mathrm{T}}) \begin{bmatrix} \Delta A_1 \\ \vdots \\ \Delta A_m \end{bmatrix} \widetilde{W}^{\mathrm{T}} \Gamma Q^{-1} + \left[ I_m \otimes (Q^{-\mathrm{T}} \Gamma \widetilde{W}^{\mathrm{T}}) \right] \begin{bmatrix} \Delta A_1 \\ \vdots \\ \Delta A_m \end{bmatrix} W.$$

Consequently,

$$\left( \sum_{i=1}^{m} \|R_i\|_{\mathrm{F}}^2 \right)^{1/2} \le \|Q^{-1}\|_2^2 \, \tilde{r} + 2\|Q^{-1}\|_2 \|W\|_2 \|\widetilde{W}\|_2 \delta_{\mathcal{A}} = \delta.$$

Combine it with (4.2) in Lemma 4.1 to conclude (4.18). □

*Proof* of Step 2. Using Lemma 4.1, we know that there exists $\hat{\mu}_j$ such that

$$\sum_{k=1}^{n_j} |\mu_{jk} - \hat{\mu}_j|^2 \le \frac{\sum_{i=1}^{m} \| \operatorname{Bdiag}_{\tau_n}(R_i)\|_{\mathrm{F}}^2}{\omega_{\mathrm{nd}}^2} \le \left( \frac{\delta}{\omega_{\mathrm{nd}}} \right)^2. \tag{4.20}$$

Then for any $\mu_{j\,k_1}$, $\mu_{j\,k_2}$, we have

$$\begin{aligned} |\mu_{j\,k_1} - \mu_{j\,k_2}|^2 &\le (|\mu_{j\,k_1} - \hat{\mu}_j| + |\mu_{j\,k_2} - \hat{\mu}_j|)^2 \\ &\le 2(|\mu_{j\,k_1} - \hat{\mu}_j|^2 + |\mu_{j\,k_2} - \hat{\mu}_j|^2) \\ &\le 2 \sum_{k=1}^{n_j} |\mu_{jk} - \hat{\mu}_j|^2 \\ &\le 2 \left( \frac{\delta}{\omega_{\mathrm{nd}}} \right)^2. \end{aligned} \tag{4.21}$$

Let $\operatorname{argmin}_\ell |\mu_{jk} - \gamma_\ell| = \ell_{jk}$. Noticing that

$$\Gamma = Q^{-1} Z Q = Q^{-1} \operatorname{Bdiag}_{\tau_n}(Z) Q + Q^{-1} \operatorname{OffBdiag}_{\tau_n}(Z) Q.$$

By a result of Kahan [17] (see also [27, Remark 3.3]), we have

$$\sum_{j=1}^{t} \sum_{k=1}^{n_j} |\mu_{jk} - \gamma_{\ell_{jk}}|^2 \le 2\|Q^{-1} \operatorname{OffBdiag}_{\tau_n}(Z) Q\|_{\mathrm{F}}^2. \tag{4.22}$$

Now we declare $\ell_{j1} = \cdots = \ell_{jn_j} = j'$ for all $j = 1, 2, \ldots, t$. Because otherwise, say $\ell_{j1} \neq \ell_{j2}$, we have

$$
\begin{aligned}
4\alpha^2 g^2 &> 4\kappa_2^2(Q)\frac{\delta^2}{\omega_{\mathrm{uq}}^2} && \text{(by (3.10))} \\
&\geq 4\|Q^{-1}\operatorname{OffBdiag}_{\tau_n}(Z)Q\|_{\mathrm{F}}^2 && \text{(by (4.18))} \qquad (4.23a)\\
&\geq 2\sum_{j=1}^{t}\sum_{k=1}^{n_j}|\mu_{jk} - \gamma_{\ell_{jk}}|^2 && \text{(by (4.22))} \\
&\geq 2(|\mu_{j1} - \gamma_{\ell_{j1}}|^2 + |\mu_{j2} - \gamma_{\ell_{j2}}|^2) \\
&\geq (|\mu_{j1} - \gamma_{\ell_{j1}}| + |\mu_{j2} - \gamma_{\ell_{j2}}|)^2 \\
&\geq (|\gamma_{\ell_{j1}} - \gamma_{\ell_{j2}}| - |\mu_{j1} - \mu_{j2}|)^2 \\
&\geq \left(g - \sqrt{2}\,\frac{\delta}{\omega_{\mathrm{nd}}}\right)^2 && \text{(by (4.21))} \\
&> [1 - (1 - 2\alpha)]^2 g^2 && \text{(by (3.10))} \\
&= 4\alpha^2 g^2, && \qquad (4.23b)
\end{aligned}
$$

a contradiction. Now using (4.22), (4.18) and (3.10), we get

$$
\begin{aligned}
\max_k |\mu_{jk} - \gamma_{j'}| &\leq \left(\sum_{k=1}^{n_j}|\mu_{jk} - \gamma_{j'}|^2\right)^{1/2} \leq \sqrt{2}\|Q^{-1}\operatorname{OffBdiag}_{\tau_n}(Z)Q\|_{\mathrm{F}} \\
&\leq \sqrt{2}\kappa_2(Q)\|\operatorname{OffBdiag}_{\tau_n}(Z)\|_{\mathrm{F}} \leq \frac{\sqrt{2}\kappa_2(Q)\delta}{\omega_{\mathrm{uq}}} < \sqrt{2}\alpha g < \frac{1}{2}g.
\end{aligned}
$$

Thus, we know that each $j \in \{1, 2, \ldots, t\}$ corresponds a unique $j'$ satisfying that $|\mu_{jk} - \gamma_{j'}| < g/2$ and $|\mu_{jk} - \gamma_i| > g/2$ for any $i \neq j'$. This is (4.19). $\square$

*Proof* of Step 3. Notice that (4.23a) implies that $\|Q^{-1}\operatorname{OffBdiag}_{\tau_n}(Z)Q\|_{\mathrm{F}} \leq \alpha g$, i.e., (4.4) holds. By Lemma 4.2, there exists a $\tau_n$-block diagonal matrix $\widetilde{B} = \operatorname{diag}(\widetilde{B}_{11}, \ldots, \widetilde{B}_{tt})$ and a nonsingular matrix $P \equiv \begin{bmatrix} P_{jk} \end{bmatrix}$ with $P_{jk} \in \mathbb{R}^{n_j \times n_k}$ and $P_{jj} = I_{n_j}$, satisfying (4.6), such that

$$
\operatorname{Bdiag}_{\tau_n}(Z)(QP) = (QP)\widetilde{B}. \tag{4.24}
$$

Denote by $\lambda(\widetilde{B}_{jj}) = \{\tilde{\mu}_{jk}\}_{k=1}^{n_j}$. By (4.6b), (4.18) and (3.10), we know

$$
\begin{aligned}
\max_k |\tilde{\mu}_{jk} - \gamma_j| &\leq \sqrt{\sum_k |\tilde{\mu}_{jk} - \gamma_j|^2} \\
&\leq (1 + \tau^2)\kappa_2(Q)\|\operatorname{OffBdiag}_{\tau_n}(Z)\|_{\mathrm{F}} \\
&< (1 + \tau^2)\kappa_2(Q)\frac{\delta}{\omega_{\mathrm{uq}}} < (1 + \tau^2)\alpha g < \frac{g}{2}.
\end{aligned}
$$

What this means is that each of the $t$ disjoint intervals $(\gamma_i - g/2, \gamma_i + g/2)$ contains one and only one $\lambda(\widetilde{B}_{jj})$. Previously in Step 2, we proved that each of the $t$ disjoint intervals $(\gamma_i - g/2, \gamma_i + g/2)$ contains one and only one $\lambda(Z_{jj})$ as well. On the other hand, we also have $\lambda(\operatorname{Bdiag}_{\tau_n}(Z)) = \lambda(\widetilde{B})$ by (4.24). Therefore, there is permutation $\pi$ of $\{1, 2, \ldots, t\}$ such that

$$\lambda(\widetilde{B}_{\pi(j)\pi(j)}) = \lambda(Z_{jj}) \quad \text{for } 1 \le j \le t. \tag{4.25}$$

Let $\Pi$ be the permutation matrix such that

$$\Pi^{\mathrm{T}} \widetilde{B} \Pi = \operatorname{diag}(\widetilde{B}_{\pi(1)\pi(1)}, \ldots, \widetilde{B}_{\pi(t)\pi(t)}). \tag{4.26}$$

It can be seen that $\Pi \in \mathbb{P}_{\tau_n}$, i.e., it is $\tau_n$-block structure preserving. Finally by (4.25) and (4.26),

$$\begin{aligned} \operatorname{diag}(Z_{11}, \ldots, Z_{tt})(QP\Pi) &= QP\widetilde{B}\Pi \\ &= (QP\Pi)\Pi^{\mathrm{T}}\widetilde{B}\Pi \\ &= (QP\Pi) \operatorname{diag}(\widetilde{B}_{\pi(1)\pi(1)}, \ldots, \widetilde{B}_{\pi(t)\pi(t)}). \end{aligned} \tag{4.27}$$

Let $\widetilde{D} = QP\Pi \equiv \big[\widetilde{D}_{jk}\big]$ with $\widetilde{D}_{jk} \in \mathbb{R}^{n_j \times n_k}$. The equation (4.27) becomes

$$\operatorname{diag}(Z_{11}, \ldots, Z_{tt})\widetilde{D} = \widetilde{D} \operatorname{diag}(\widetilde{B}_{\pi(1)\pi(1)}, \ldots, \widetilde{B}_{\pi(t)\pi(t)})$$

which yields $Z_{jj}\widetilde{D}_{jk} = \widetilde{D}_{jk}\widetilde{B}_{\pi(k)\pi(k)}$. Recalling (4.25) and $\lambda(Z_{jj}) \cap \lambda(Z_{kk}) = \emptyset$ for $j \ne k$ by (4.19), we conclude that $\widetilde{D}_{jk} = 0$ for $j \ne k$, i.e., $\widetilde{D}$ is $\tau_n$-block diagonal. $\square$

*Proof* of Step 4. Noticing that $Q = W^{-1}\widetilde{W}$ and $\widetilde{D} = QP\Pi$ in Step 3, we have $W\widetilde{D} = \widetilde{W}P\Pi$. Then using Lemma 4.4, we know that $\widetilde{D}$ is nonsingular and for any singular value $\sigma$ of $\widetilde{D}$, (4.14) holds with

$$\epsilon = \frac{\tau}{\alpha} \cdot \frac{\|Q^{-1} \operatorname{OffBdiag}_{\tau_n}(Z) Q\|_{\mathrm{F}}}{g}.$$

By (4.18), we have

$$\epsilon \le \frac{\tau}{\alpha} \cdot \frac{\kappa_2(Q)\delta}{g\,\omega_{\mathrm{uq}}} = \epsilon_*. \tag{4.28}$$

Now let $\widetilde{D}_{jj} = U_j \Sigma_j V_j^{\mathrm{T}}$ be the SVD of $\widetilde{D}_{jj}$. Denote by $U = \operatorname{diag}(U_1, \ldots, U_t)$, $V = \operatorname{diag}(V_1, \ldots, V_t)$ and $D = \Pi V U^{\mathrm{T}} \Pi^{\mathrm{T}}$. It can be verified that $D$ is orthogonal and $\tau_n$-block diagonal. It follows from $W\widetilde{D} = \widetilde{W}P\Pi$ that

$$\begin{aligned} W = \widetilde{W}P\Pi\widetilde{D}^{-1} &= \widetilde{W}(\Pi\widetilde{D}^{-1}\Pi^{\mathrm{T}})\Pi + \widetilde{W} \operatorname{OffBdiag}_{\tau_n}(P)\Pi\widetilde{D}^{-1} \\ &= \widetilde{W}D\Pi + \widetilde{W}(\Pi\widetilde{D}^{-1}\Pi^{\mathrm{T}} - D)\Pi + \widetilde{W} \operatorname{OffBdiag}_{\tau_n}(P)\Pi\widetilde{D}^{-1} \\ &= \widetilde{W}D\Pi + \widetilde{W}\Pi V(\Sigma^{-1} - I)U + \widetilde{W} \operatorname{OffBdiag}_{\tau_n}(P)\Pi\widetilde{D}^{-1}. \end{aligned}$$

Using Lemma 4.4, we have for $\mathrm{p} \in \{2, \mathrm{F}\}$

$$\begin{aligned} \big\|W - \widetilde{W}D\Pi\big\|_{\mathrm{p}} &= \big\|\widetilde{W}\Pi V(\Sigma^{-1} - I)U + \widetilde{W} \operatorname{OffBdiag}_{\tau_n}(P)\Pi\widetilde{D}^{-1}\big\|_{\mathrm{p}} \\ &\le \big\|\widetilde{W}\big\|_{\mathrm{p}} \left( \frac{1 + \sqrt{t}\,\epsilon_*}{\sqrt{1 - 2\sqrt{t-1}\epsilon_* - (t-1)\epsilon_*^2}} - 1 \right) \\ &= \big\|\widetilde{W}\big\|_{\mathrm{p}} \big[(\sqrt{t} + \sqrt{t-1})\epsilon + O(\epsilon^2)\big]. \end{aligned}$$

Combine it with (4.28) to conclude the proof of (3.11). $\square$

## 5. Numerical examples

In this section, we present some random numerical tests to validate our theoretical results. All numerical examples were carried out using MATLAB R2014b, with machine unit roundoff $2^{-53} \approx 1.1 \times 10^{-16}$.

Let us start by explain how the testing examples are constructed. Given a partition $\tau_n = (n_1, \ldots, n_t)$ of $n$ and $m$, the number of matrices, we generate the matrix sets $\mathcal{A} = \{A_i\}_{i=1}^m$ and $\widetilde{\mathcal{A}} = \{\widetilde{A}_i\}_{i=1}^m$ as follows.

(1) Randomly generate $W \equiv [W_1, \ldots, W_t] \in \mathbb{W}_{\tau_n}$. This is done by first generating an $n \times n$ random matrix from the standard normal distribution and then orthonormalizing its first $n_1$ columns, the next $n_2$ columns, ..., and the last $n_t$ columns, respectively. Set $V = W^{-\mathrm{T}}$;
(2) Generate $m$ $\tau_n$-block diagonal matrices $D_j$ randomly from the standard normal distribution and set $A_j = VD_jV^{\mathrm{T}}$ for $1 \le j \le m$. This makes sure that $\mathcal{A}$ is $\tau_n$-block diagonalizable.
(3) Generate $m$ noise matrices $N_j$ also randomly from the standard normal distribution and set $\widetilde{A}_j = A_j + \xi N_j$, where $\xi$ is a parameter for controlling noise level. $\widetilde{\mathcal{A}}$ is likely not $\tau_n$-block diagonalizable but it is approximately. An approximate diagonalizer $\widetilde{W} \equiv [\widetilde{W}_1, \ldots, \widetilde{W}_t] \in \mathbb{W}_{\tau_n}$ of $\widetilde{\mathcal{A}}$ is computed by JBD-NCG [19] followed by orthonormalization as in item (1) above.

For comparison purpose, we estimate the relative error between $\widetilde{W}$ and $W$ as measured by (1.5) for $\mathrm{p} = \mathrm{F}$ as follows. We have to minimize

$$\|W - \widetilde{W}D\Pi\|_{\mathrm{F}}^2 = \|W\|_{\mathrm{F}}^2 - 2\,\mathrm{trace}(W^{\mathrm{T}}\widetilde{W}D\Pi) + \|\widetilde{W}\|_{\mathrm{F}}^2$$

over orthogonal $D \in \mathbb{D}_{\tau_n}$ and $\Pi \in \mathbb{P}_{\tau_n}$, which is equivalent to maximizing

$$\sum_{j=1}^{t} \mathrm{trace}(W_j^{\mathrm{T}}\widetilde{W}_{\pi(j)}D_{\pi(j)})$$

over orthogonal $D_{\pi(j)}$ and permutations $\pi$ of $\{1, 2, \ldots, t\}$, subject to $n_j = n_{\pi(j)}$, which again is equivalent to

$$\max_{\pi} \sum_{j=1}^{t} (\text{the sum of the singular values of } W_j^{\mathrm{T}}\widetilde{W}_{\pi(j)}) \tag{5.1}$$

subject to $n_j = n_{\pi(j)}$. Abusing the notation a little bit, we let $\pi$ be the one that achieve the optimal in (5.1) and perform the singular value decomposition $\widetilde{W}_{\pi(j)}^{\mathrm{T}}W_j = U_j\Sigma_jV_j^{\mathrm{T}}$, and set $D = \mathrm{diag}(U_{\pi(1)}V_{\pi(1)}^{\mathrm{T}}, \ldots, U_{\pi(t)}V_{\pi(t)}^{\mathrm{T}})$. Finally, the error (1.5) for $\mathrm{p} = \mathrm{F}$ is given by

$$\frac{\|W - \widetilde{W}D\Pi\|_{\mathrm{F}}}{\|\widetilde{W}\|_{\mathrm{F}}} \tag{5.2}$$

with $D$ as above and $\Pi \in \mathbb{P}_{\tau_n}$ as determined by the optimal $\pi$. It doesn't seem to be a simple way to compute (1.5) for $\mathrm{p} = 2$.

To generate error bounds by Theorem 3.3, we have to decide what $\Gamma$ to use. Ideally, we should use the one that minimize the right-hand side of (3.11), but we don't have an easy way to do that. For the tests below, we use 50 different $\Gamma$ and pick the best bound. Specifically, we use a particular one

$$\Gamma = \operatorname{diag}(-1, -1 + \frac{2}{t-1}, -1 + \frac{4}{t-1}, \dots, 1) \tag{5.3}$$

and 49 random ones with their diagonal entries $\gamma_1, \dots, \gamma_t$ randomly drawn from the interval $(-1, 1)$ with the uniform distribution. Our experience suggests that the particular $\Gamma$ in (5.3) usually leads to bounds having the same order as the best one produced by the 49 random $\Gamma$. However, it can happen that the best one is much better, although not very often.

Now we are ready to perform our numerical tests according to five different testing scenarios: varying numbers of matrices (test 1), varying matrix sizes (test 2), varying numbers of diagonal blocks (test 3), varying noise levels (test 4), and varying condition numbers $\operatorname{cond}(\mathcal{A})$ (test 5). We will examine the following quantities: the modulus of uniqueness $\omega_{\rm uq}$, the modulus of non-divisibility $\omega_{\rm nd}$, $\delta$ as defined in (3.9), the *ratio* as the quotient of $\delta$ over the right hand side of (3.10) (if it is less than 1, then (3.10) is satisfied), $\varepsilon_{\rm bker} \equiv \varepsilon_{\rm bker}(\widetilde{\mathcal{A}}; \widetilde{W})$ the upper bound as in (3.6) for the backward error, $\operatorname{cond}(\mathcal{A})$ the condition number as defined in (3.16), $\varepsilon_{\rm ub}$ as in (3.11), and finally the *error* in $\widetilde{W}$ as in (5.2).

| $m$ | $\omega_{\rm uq}$ | $\omega_{\rm nd}$ | $\delta$ | *ratio* | $\varepsilon_{\rm bker}$ | cond($\mathcal{A}$) | $\varepsilon_{\rm ub}$ | *error* |
|---|---|---|---|---|---|---|---|---|
| 4 | 1.7e+00 | 1.9e+00 | 4.8e-10 | 1.4e-09 | 3.4e-10 | 2.4e+03 | 1.3e-09 | 1.9e-11 |
| 8 | 3.8e+00 | 3.9e+00 | 2.2e-10 | 1.5e-09 | 3.2e-10 | 1.6e+03 | 1.4e-09 | 1.9e-11 |
| 16 | 6.6e+00 | 6.4e+00 | 9.8e-10 | 7.3e-10 | 3.3e-10 | 1.3e+03 | 6.8e-10 | 1.9e-11 |
| 32 | 1.0e+01 | 1.0e+01 | 8.5e-10 | 6.5e-10 | 2.7e-10 | 1.2e+03 | 6.0e-10 | 1.8e-11 |
| 64 | 1.6e+01 | 1.6e+01 | 1.3e-09 | 4.2e-10 | 1.8e-10 | 1.2e+03 | 3.8e-10 | 1.2e-11 |
| 128 | 2.5e+01 | 2.5e+01 | 2.2e-09 | 4.4e-10 | 2.1e-10 | 1.2e+03 | 4.0e-10 | 1.4e-11 |
| 256 | 3.6e+01 | 3.6e+01 | 1.8e-09 | 4.2e-10 | 1.7e-10 | 1.2e+03 | 3.9e-10 | 1.1e-11 |

TABLE 1. Bound vs. $m$, the number of matrices in $\mathcal{A}$ for case $\tau_9 = (3, 3, 3)$

**Test 1: number of matrices.** In this test, we fix $\xi = 10^{-12}$ and vary the number $m$ of matrices in the matrix set $\mathcal{A}$, and use two different partitions $\tau_9 = (3, 3, 3)$ and $\tau_6 = (1, 2, 3)$. the numerical results are displayed in Tables 1 and 2 for the two different partitions, respectively. We summarize our observations from Tables 1

| $m$ | $\omega_{\rm uq}$ | $\omega_{\rm nd}$ | $\delta$ | *ratio* | $\varepsilon_{\rm bker}$ | cond($\mathcal{A}$) | $\varepsilon_{\rm ub}$ | *error* |
|---|---|---|---|---|---|---|---|---|
| 4 | 8.1e-01 | 2.7e+00 | 1.4e-10 | 8.8e-10 | 3.6e-11 | 9.7e+04 | 8.1e-10 | 7.3e-12 |
| 8 | 3.0e+00 | 4.7e+00 | 1.7e-10 | 5.6e-10 | 7.3e-11 | 2.8e+04 | 5.2e-10 | 7.5e-12 |
| 16 | 5.9e+00 | 7.4e+00 | 2.0e-10 | 4.5e-10 | 7.7e-11 | 1.8e+04 | 4.1e-10 | 5.8e-12 |
| 32 | 8.0e+00 | 1.1e+01 | 3.3e-10 | 4.0e-10 | 7.9e-11 | 1.8e+04 | 3.7e-10 | 6.4e-12 |
| 64 | 9.7e+00 | 1.6e+01 | 4.1e-10 | 3.4e-10 | 5.3e-11 | 1.9e+04 | 3.1e-10 | 5.9e-12 |
| 128 | 1.6e+01 | 2.3e+01 | 4.7e-10 | 3.2e-10 | 3.9e-11 | 1.7e+04 | 2.9e-10 | 4.3e-12 |
| 256 | 2.2e+01 | 3.2e+01 | 5.7e-10 | 4.3e-10 | 3.3e-11 | 1.7e+04 | 3.9e-10 | 3.4e-12 |

TABLE 2. Bound vs. $m$, the number of matrices in $\mathcal{A}$ for case $\tau_6 = (1, 2, 3)$

and 2 as follows.

(1) For all $m$, the *ratio*s are far less than 1. In the other word, (3.10) is satisfied for all, and hence the bound (3.11) holds.

(2) For all $m$, $\varepsilon_{\rm ub}$ provides a very good upper bound on the *error*.
(3) As $m$ increases, i.e., as we expand the matrix set $\mathcal{A}$, the modulus of uniqueness and modulus of non-divisibility increase as well, and the condition number cond($\mathcal{A}$) decreases at first, then remains almost the same.

**Test 2: matrix sizes.** In this test, we fix $\xi = 10^{-12}$, $m = 16$, and test for two partitions $\tau_n = p \times (3,3,3)$ or $\tau_n = p \times (1,2,3)$, where $p = 1,2,\dots,7$. Then the matrix size $n = 9p$ or $6p$ will increase as $p$ increases. We display the numerical results in Tables 3 and 4. We can see from Tables 3 and 4 that $\varepsilon_{\rm ub}$ provides a very good upper bound on the *error* for different sizes of matrices.

| $n$ | $\omega_{\rm uq}$ | $\omega_{\rm nd}$ | $\delta$ | *ratio* | $\varepsilon_{\rm bker}$ | cond($\mathcal{A}$) | $\varepsilon_{\rm ub}$ | *error* |
|---|---|---|---|---|---|---|---|---|
| 9 | 6.8e+00 | 6.8e+00 | 2.1e-10 | 7.7e-10 | 4.5e-11 | 2.4e+02 | 7.1e-10 | 3.9e-12 |
| 18 | 1.1e+01 | 1.1e+01 | 2.5e-09 | 2.1e-09 | 1.3e-09 | 6.3e+03 | 2.0e-09 | 5.6e-11 |
| 27 | 1.2e+01 | 1.2e+01 | 1.1e-08 | 5.1e-09 | 4.3e-09 | 1.7e+04 | 4.7e-09 | 1.2e-10 |
| 36 | 1.4e+01 | 1.4e+01 | 6.7e-09 | 2.3e-09 | 1.2e-09 | 5.6e+03 | 2.1e-09 | 3.2e-11 |
| 45 | 1.6e+01 | 1.6e+01 | 3.1e-09 | 2.0e-09 | 1.2e-09 | 4.4e+03 | 1.8e-09 | 1.8e-11 |
| 54 | 1.8e+01 | 1.8e+01 | 1.7e-08 | 4.7e-09 | 6.1e-09 | 2.6e+04 | 4.4e-09 | 5.7e-11 |
| 63 | 1.9e+01 | 1.9e+01 | 2.1e-07 | 5.4e-08 | 7.2e-08 | 9.4e+03 | 5.0e-08 | 7.7e-10 |

TABLE 3. Bound vs. matrix size $n$ for case $\tau_n = p \times (3,3,3)$

| $n$ | $\omega_{\rm uq}$ | $\omega_{\rm nd}$ | $\delta$ | *ratio* | $\varepsilon_{\rm bker}$ | cond($\mathcal{A}$) | $\varepsilon_{\rm ub}$ | *error* |
|---|---|---|---|---|---|---|---|---|
| 6 | 4.2e+00 | 5.7e+00 | 1.8e-10 | 3.7e-10 | 2.6e-11 | 1.0e+02 | 3.4e-10 | 4.6e-12 |
| 12 | 6.8e+00 | 6.7e+00 | 3.5e-10 | 7.9e-10 | 7.6e-11 | 4.8e+02 | 7.3e-10 | 6.0e-12 |
| 18 | 8.8e+00 | 9.4e+00 | 5.7e-10 | 1.6e-09 | 3.5e-10 | 5.5e+03 | 1.4e-09 | 1.2e-11 |
| 24 | 9.0e+00 | 8.5e+00 | 4.7e-09 | 3.1e-09 | 1.5e-09 | 4.4e+03 | 2.8e-09 | 5.0e-11 |
| 30 | 9.5e+00 | 9.0e+00 | 9.2e-09 | 4.8e-09 | 3.6e-09 | 7.2e+03 | 4.4e-09 | 5.5e-11 |
| 36 | 1.2e+01 | 1.0e+01 | 3.8e-09 | 4.4e-09 | 2.3e-09 | 1.9e+03 | 4.1e-09 | 4.4e-11 |
| 42 | 1.3e+01 | 1.2e+01 | 6.9e-09 | 4.7e-09 | 6.5e-09 | 1.2e+05 | 4.4e-09 | 4.5e-11 |

TABLE 4. Bound vs. matrix size $n$ for case $\tau_n = p \times (1,2,3)$

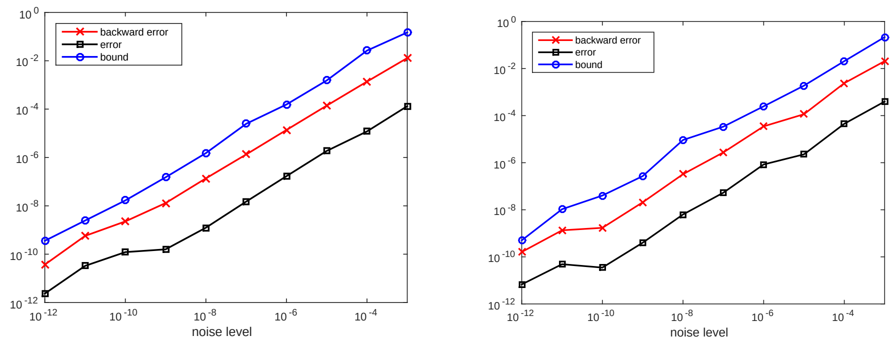
**Test 3: number of diagonal blocks.** In this test, we fix $\xi = 10^{-12}$, $m = 16$, and generate the partition $\tau_n$ randomly using MATLAB command `randi(5,t,1)`, where $t$ is the number of diagonal blocks. In the other word, the block diagonal matrices $D_j$ have $t$ diagonal blocks and the order of the $i$th block is $\tau_n(i)$, randomly drawn from $\{1,2,\dots,5\}$ with the uniform distribution. For $t = 3,4,\dots,9$, we display the numerical results in Table 5. We can see from Table 5 that $\varepsilon_{\rm ub}$ provides a very good upper bound on the *error* for the different numbers of diagonal blocks.

**Test 4: noise level.** In this test, we fix the number of matrices $m = 16$. For different partitions $\tau_n = (3,3,3)$ and $\tau_n = (1,2,3)$, in Figure 1, we plot $\varepsilon_{\rm bker}$, *error* and $\varepsilon_{\rm ub}$ versus different noise levels. We can see from Figure 1 that as $\xi$ increases, $\varepsilon_{\rm bker}$, *error* and $\varepsilon_{\rm ub}$ all increase almost linearly. For all noise levels, $\varepsilon_{\rm ub}$ indeed provides a good upper bound on the *error*.

**Test 5: condition number.** In this test, we fix $m = 16$, $\xi = 10^{-12}$. For two different partitions $\tau_n = (3,3,3)$ and $\tau_n = (1,2,3)$, we ran the tests 100 times for each partition. In Figure 2, we plot the quotient $\varepsilon_{\rm ub}/$*error* versus the condition

| $t$ | $\omega_{\mathrm{uq}}$ | $\omega_{\mathrm{nd}}$ | $\delta$ | *ratio* | $\varepsilon_{\mathrm{bker}}$ | $\mathrm{cond}(\mathcal{A})$ | $\varepsilon_{\mathrm{ub}}$ | *error* |
|---|---|---|---|---|---|---|---|---|
| 3 | 5.7e+00 | 7.6e+00 | 6.7e-10 | 5.9e-10 | 1.9e-10 | 1.8e+04 | 5.4e-10 | 1.1e-11 |
| 4 | 3.5e+00 | 7.1e+00 | 5.7e-10 | 4.1e-09 | 6.2e-10 | 4.2e+03 | 3.7e-09 | 5.2e-11 |
| 5 | 3.8e+00 | 5.8e+00 | 8.3e-10 | 3.8e-09 | 8.1e-10 | 4.4e+03 | 3.3e-09 | 1.8e-11 |
| 6 | 4.0e+00 | 6.0e+00 | 8.0e-10 | 3.5e-09 | 6.7e-10 | 2.2e+04 | 3.0e-09 | 1.2e-11 |
| 7 | 5.8e+00 | 6.5e+00 | 1.9e-09 | 7.1e-09 | 2.7e-09 | 1.2e+04 | 6.1e-09 | 3.7e-11 |
| 8 | 4.4e+00 | 8.1e+00 | 2.4e-09 | 1.5e-08 | 3.0e-09 | 3.5e+04 | 1.3e-08 | 3.6e-11 |
| 9 | 3.9e+00 | 8.4e+00 | 1.1e-09 | 9.5e-09 | 8.7e-10 | 1.3e+04 | 8.1e-09 | 1.3e-11 |

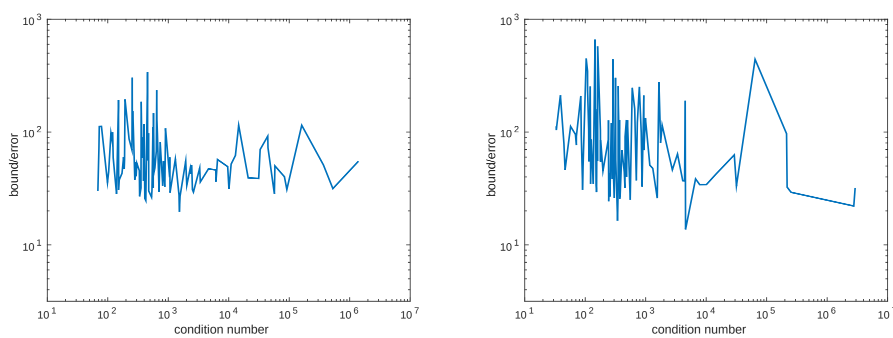
TABLE 5. Bound vs. number of diagonal blocks




$\tau_n = (3,3,3)$ $\tau_n = (1,2,3)$

FIGURE 1. Backward error $\varepsilon_{\mathrm{bker}}$, *error* and $\varepsilon_{\mathrm{ub}}$ vs. noise level

number $\mathrm{cond}(\mathcal{A})$. The smaller the quotient is, the sharper $\varepsilon_{\mathrm{ub}}$ estimates the *error*. We can see from Figure 2 that $\varepsilon_{\mathrm{ub}}$ provides a good upper bound on the *error*, even as the condition number becomes large.




$\tau_n = (3,3,3)$ $\tau_n = (1,2,3)$

FIGURE 2. $\varepsilon_{\mathrm{ub}}/$*error* vs. condition number

## 6. Concluding Remarks

In this paper, we developed a perturbation theory for JBDP. An upper bound is given for the distance between a block diagonalizer for the original JBDP that is block diagonalizable and an approximate diagonalizer for its perturbed JBDP. The backward error and condition number are also derived and discussed for JBDP. Numerical tests validate the theoretical results.

The JBDP of interest in this paper is for block diagonalization via congruence transformations which are known to preserve symmetry. Yet our development so far does not assume that all $A_i$ are symmetric. What will happen to all the results if they are symmetric. It turns out that all the results and development remain valid after minor changes to the definitions of $G_{jk}$ in (2.8): remove the second, fourth, ... block rows as now all $A_i^{(jj)}$ are symmetric.

We have been limiting all the matrices to real ones. This is not a limitation, either. In fact, if all matrices are complex, the change that needs to be made is simply to replace all transposes T by complex conjugate transposes H, but for simplicity we still would like to keep all $\gamma_i$, the diagonal entries of $\Gamma$ real so that we don't have change the definition of the gap $g$ in (3.5).

Conceivably, we might use similarity transformation for block diagonalization, i.e., instead of (1.3), we may seek a nonsingular matrix $W \in \mathbb{R}^{n\times n}$ such that all $W^{-1}A_iW$ are $\tau_n$-block diagonal. A similar development that are very much parallel to those in [3] and in this paper can be worked out. A major change will be to redefine the subspace $\mathscr{N}(\mathcal{A})$ in (2.2) as

$$\mathscr{N}(\mathcal{A}) := \big\{Z \in \mathbb{R}^{n\times n} \ : \ A_iZ - ZA_i = 0 \text{ for } 1 \le i \le m\big\}.$$

We omit the detail.

## References


[1] B. Afsari. Sensitivity analysis for the problem of matrix joint diagonalization. *SIAM J. Matrix Anal. Appl.*, 30(3):1148–1171, 2008.

[2] Y. Bai, E. de Klerk, D. Pasechnik, and R. Sotirov. Exploiting group symmetry in truss topology optimization. *Optim. Engrg.*, 10(3):331–349, 2009.

[3] Y. Cai and C. Liu. An algebraic approach to nonorthogonal general joint block diagonalization. *SIAM J. Matrix Anal. Appl.*, 38(1):50–71, 2017.

[4] J.-F. Cardoso. Multidimensional independent component analysis. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 4, pages 1941–1944. IEEE, Washinton, DC, 1998.

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

[6] E. De Klerk, D. V. Pasechnik, and A. Schrijver. Reduction of symmetric semidefinite programs using the regular $*$-representation. *Math. Program.*, 109(2-3):613–624, 2007.

[7] E. De Klerk and R. Sotirov. Exploiting group symmetry in semidefinite programming relaxations of the quadratic assignment problem. *Math. Program.*, 122(2):225–246, 2010.

[8] L. De Lathauwer. Decompositions of a higher-order tensor in block terms–part I: Lemmas for partitioned matrices. *SIAM J. Matrix Anal. Appl.*, 30(3):1022–1032, 2008.

[9] L. De Lathauwer. Decompositions of a higher-order tensor in block terms–part II: Definitions and uniqueness. *SIAM J. Matrix Anal. Appl.*, 30(3):1033–1066, 2008.

[10] L. De Lathauwer. A survey of tensor methods. In *2009 IEEE International Symposium on Circuits and Systems*, pages 2773–2776. IEEE, 2009.

[11] L. De Lathauwer, B. De Moor, and J. Vandewalle. Fetal electrocardiogram extraction by blind source subspace separation. *IEEE Trans. Biomedical Engrg.*, 47(5):567–572, 2000.

[12] L. De Lathauwer and D. Nion. Decompositions of a higher-order tensor in block terms–part III: Alternating least squares algorithms. *SIAM J. Matrix Anal. Appl.*, 30(3):1067–1083, 2008.
[13] J. W. Demmel. *Applied Numerical Linear Algebra*. SIAM, Philadelphia, PA, 1997.
[14] I. Domanov and L. De Lathauwer. On the uniqueness of the canonical polyadic decomposition of third-order tensors—part I: Basic results and uniqueness of one factor matrix. *SIAM J. Matrix Anal. Appl.*, 34(3):855–875, 2013.
[15] I. Domanov and L. De Lathauwer. On the uniqueness of the canonical polyadic decomposition of third-order tensors—part II: Uniqueness of the overall decomposition. *SIAM J. Matrix Anal. Appl.*, 34(3):876–903, 2013.
[16] K. Gatermann and P. A. Parrilo. Symmetry groups, semidefinite programs, and sums of squares. *J. Pure Appl. Algebra*, 192(1):95–128, 2004.
[17] W. Kahan. Spectra of nearly hermitian matrices. *Proc. Amer. Math. Soc.*, 48(1):11–17, 1975.
[18] J. B. Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.*, 18(2):95–138, 1977.
[19] D. Nion. A tensor framework for nonunitary joint block diagonalization. *IEEE Trans. Signal Process.*, 59(10):4585–4594, 2011.
[20] B. Póczos and A. Lőrincz. Independent subspace analysis using k-nearest neighborhood distances. In *Artificial Neural Networks: Formal Models and Their Applications–ICANN 2005*, pages 163–168. Springer, 2005.
[21] F. G. Russo. On an argument of j.–f. cardoso dealing with perturbations of joint diagonalizers. *preprint, http://arXiv:1103.3670*, 2011.
[22] D. C. Shi, Y. F. Cai, and S. F. Xu. Some perturbation results for a normalized non-orthogonal joint diagonalization problem. *Linear Algebra Appl.*, 484:457–476, 2015.
[23] M. Sørensen and L. De Lathauwer. Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-(Lr,n,Lr,n,1) terms—part I: Uniqueness. *SIAM J. Matrix Anal. Appl.*, 36(2):496–522, 2015.
[24] M. Sørensen and L. De Lathauwer. New uniqueness conditions for the canonical polyadic decomposition of third-order tensors. *SIAM J. Matrix Anal. Appl.*, 36(4):1381–1403, 2015.
[25] A. Stegeman. On uniqueness of the canonical tensor decomposition with some form of symmetry. *SIAM J. Matrix Anal. Appl.*, 32(2):561–583, 2011.
[26] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Academic Press, Boston, 1990.
[27] J. G. Sun. On the variation of the spectrum of a normal matrix. *Linear Algebra Appl.*, 246:215 – 223, 1996.
[28] F. J. Theis. Blind signal separation into groups of dependent signals using joint block diagonalization. In *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*, pages 5878–5881. IEEE, 2005.
[29] F. J. Theis. Towards a general independent subspace analysis. In *Advances in Neural Information Processing Systems*, pages 1361–1368, MIT Press, Cambridge, MA, 2006.
[30] P. Tichavsky, A. H. Phan, and A. Cichocki. Non-orthogonal tensor diagonalization. *preprint, http://arXiv:1402.1673v3*, 2014.
[31] C. F. Van Loan and G. H. Golub. *Matrix computations*. Johns Hopkins University Press, Baltimore, MD, 4th edition, 2012.
[32] N. Vannieuwenhoven. A condition number for the tensor rank decomposition. *preprint, http://arXiv:1604.00052*, 2016.
[33] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer. Tensorlab 3.0, Mar. 2016. Available online at http://www.tensorlab.net.

LMAM & School of Mathematical Sciences, Peking Univ., Beijing, P.R. China, 100871.
*E-mail address*: yfcai@math.pku.edu.cn

Department of Mathematics, University of Texas at Arlington, P.O.Box 19408, Arlington, TX 76019, United States.
*E-mail address*: rcli@uta.edu