Linear Algebra and its Applications 581 (2019) 163–197

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Perturbation analysis for matrix joint block diagonalization

Yunfeng Cai [a,1], Ren-Cang Li [b,*,2]

[a] *Cognitive Computing Lab, Baidu Research, Beijing 100193, PR China*
[b] *Department of Mathematics, University of Texas at Arlington, P.O. Box 19408, Arlington, TX 76019-0408, USA*

ARTICLE INFO

Article history:
Received 23 January 2019
Accepted 8 July 2019
Available online 17 July 2019
Submitted by V. Mehrmann

MSC:
65F99
49Q12
15A23
15A69

Keywords:
Matrix joint block-diagonalization
Perturbation analysis
Modulus of uniqueness
Modulus of non-divisibility
MICA

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 matrix joint diagonalization problem (JDP) as a special case for which all $W^{\mathrm{T}}A_iW$ are required diagonal. Generically, such a matrix W may not exist, but there are practical applications such as multidimensional independent component analysis (MICA) for which it does exist under the ideal situation, i.e., 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. The main goal of this paper is 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 had been attempted for JBDP because, in large part, there is no quantitative way to describe solution uniqueness of JBDP until 2017 when Cai and Liu (2017) [9] successfully obtained a necessary and sufficient uniqueness condition. Based on the condition, in this article, we will establish a perturbation theory for JBDP. Our main contributions include an error bound for the approximate block-diagonalizer $\widetilde{W}$

* Corresponding author.
E-mail addresses: yfcai1116@gmail.com (Y. Cai), rcli@uta.edu (R.-C. Li).
[1] Supported in part by NNSFC grants 11671023, 11301013, and 11421101.
[2] Supported in part by NSF grants CCF-1527104 and DMS-1719620.

https://doi.org/10.1016/j.laa.2019.07.012

and a backward error analysis for JBDP. Numerical tests are presented to validate the theoretical results.

1. Introduction

The matrix 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., [1–4]) and semidefinite programming (see, e.g., [5–8]). Tremendous efforts have been devoted to solving JBDP and, as a result, 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 [9–12] and references therein. The MATLAB toolbox for tensor computation – TENSORLAB [13] 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 study. Previously, there are only a handful papers in the literature that studied the perturbation analysis of the matrix joint 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:

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

means that $n_1, n_2, \ldots, n_t$ 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 τ_n, denoted by $t = \mathrm{card}(\tau_n)$.

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

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

and define its τ_n-*block-diagonal part* and τ_n-*off-block-diagonal part* as

$$\mathrm{Bdiag}_{\tau_n}(X) = \mathrm{diag}(X_{11}, \ldots, X_{tt}), \quad \mathrm{OffBdiag}_{\tau_n}(X) = X - \mathrm{Bdiag}_{\tau_n}(X).$$

The matrix X is referred to as *a τ_n-block-diagonal matrix* if $\mathrm{OffBdiag}_{\tau_n}(X) = 0$. The set of all τ_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 τ_n is to find a nonsingular matrix $W \in \mathbb{R}^{n\times n}$ such that all $W^{\mathrm{T}}A_iW$ are τ_n-block-diagonal, i.e.,

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

where $A_i^{(jj)} \in \mathbb{R}^{n_j\times n_j}$. When (1.3) holds, we say that $\mathcal{A}$ is *τ_n-block-diagonalizable* and W is a *τ_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 $\tau_n = (1, 1, \ldots, 1)$, the word "*$\tau_n$-block*" is dropped from all relevant terms. For example, "τ_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_j 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 *τ_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 τ_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 Π can be expressed as $\Pi = \mathrm{diag}(\Pi_1, \ldots, \Pi_t)$, where Π_j is an $n_j \times n_j$ permutation matrix. But not all $\Pi \in \mathbb{P}_{\tau_n}$ also belong to $\mathbb{D}_{\tau_n}$. 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 $\Pi \in \mathbb{R}^{n\times n}$ is in $\mathbb{P}_{\tau_n}$ when $\tau_n = (1, 1, \ldots, 1)$. It can be proved that for given $\Pi \in \mathbb{P}_{\tau_n}$, there is a permutation π of $\{1, 2, \ldots, t\}$ such that

$$\Pi^{\mathrm{T}}D\Pi \in \mathbb{D}_{\tau_n} = \mathrm{diag}(\Pi_1^{\mathrm{T}}D_{\pi(1)}\Pi_1, \Pi_2^{\mathrm{T}}D_{\pi(2)}\Pi_2, \ldots, \Pi_t^{\mathrm{T}}D_{\pi(t)}\Pi_t)$$

for any $D = \mathrm{diag}(D_1, D_2, \ldots, D_t) \in \mathbb{D}_{\tau_n}$. Specifically, the subblocks of Π, if partitioned as in (1.2), are all 0 blocks, except those at the block positions $(\pi(j), j)$, which are $n_j \times n_j$ permutation matrices Π_j. As a consequence, $n_j = n_{\pi(j)}$ for all $1 \leq j \leq t$.

It is not hard to verify that if W is a τ_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, τ_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 τ_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 τ_n-block-diagonalizable* if it has a τ_n-block-diagonalizer and if any two of its τ_n-block-diagonalizers are equivalent.

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

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

This doesn't loss any generality because $W[\mathrm{Bdiag}_{\tau_n}(W^{\mathrm{T}}W)]^{-1/2} \in \mathbb{W}_{\tau_n}$ 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 ΔA_i is a perturbation to A_i. Assume $\mathcal{A} = \{A_i\}_{i=1}^m$ is τ_n-block-diagonalizable and $W \in \mathbb{W}_{\tau_n}$ is a τ_n-block-diagonalizer and (1.3) holds. Let $\widetilde{W} \in \mathbb{W}_{\tau_n}$ be an approximate τ_n-block-diagonalizer of $\widetilde{\mathcal{A}}$ in the sense that all $\widetilde{W}^{\mathrm{T}}\widetilde{A}_i\widetilde{W}$ are approximately τ_n-block-diagonal. How much does $\widetilde{W}$ differ from the block-diagonalizer W of $\mathcal{A}$?

There are two important aspects that need clarification regarding this perturbation problem. First, $\widetilde{\mathcal{A}}$ may or may not be τ_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 to allow this. 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 for $\widetilde{\mathcal{A}}$, that always exist. It turns out this workaround is exactly what some practical applications call for because most practical JBDP come from block-diagonalizable JBDP but contaminated with noises to become approximately block-diagonalizable and an approximate diagonalizer for the noisy JBDP gets computed numerically. In such a scenario, 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 block-diagonalizers, given that they are not unique. In view of Definition 1.2 and the discussion in the paragraph immediately proceeding it, we propose to use

$$\operatorname{dist}(W, \widetilde{W}) := \min_{\substack{D\in\mathbb{D}_{\tau_n}, D^{\mathrm{T}}D=I \\ \Pi\in\mathbb{P}_{\tau_n}}} \|W - \widetilde{W}D\Pi\| \tag{1.5}$$

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 theoretical analysis below, we use both $\|\cdot\|_2$, the matrix spectral norm, and $\|\cdot\|_{\mathrm{F}}$, the matrix Frobenius norm [14], but use only $\|\cdot\|_{\mathrm{F}}$ in our numerical tests because then (1.5) is computable. Additionally, in using (1.5), we usually restrict W and $\widetilde{W}$ to $\mathbb{W}_{\tau_n}$. In fact, we can show that $\operatorname{dist}(\cdot,\cdot)$ is a metric over $\mathbb{W}_{\tau_n}$ for any unitary invariant norm $\|\cdot\|_{\mathrm{ui}}$ as follows: first, the non-negativity $\operatorname{dist}(W, \widetilde{W}) \geq 0$ is obvious; second, $\operatorname{dist}(W, \widetilde{W}) = 0$ if and only if W and $\widetilde{W}$ are equivalent; third, $\operatorname{dist}(W, \widetilde{W}) = \operatorname{dist}(\widetilde{W}, W)$ holds since $\|\cdot\|_{\mathrm{ui}}$ is unitary invariant and D, Π are unitary; fourth, for any $W_1, W_2, W_3 \in \mathbb{W}_{\tau_n}$, let

$$(D_{12}, \Pi_{12}) = \operatorname*{argmin}_{D,\Pi} \operatorname{dist}(W_1, W_2),\ (D_{23}, \Pi_{23}) = \operatorname*{argmin}_{D,\Pi} \operatorname{dist}(W_2, W_3).$$

It can be seen that $\widetilde{D} = D_{23}\Pi_{23}D_{12}\Pi_{23}^{\mathrm{T}} \in \mathbb{D}_{\tau_n}$ and is also orthogonal, and $\widetilde{\Pi} = \Pi_{23}\Pi_{12} \in \mathbb{P}_{\tau_n}$, and finally

$$\begin{aligned}
\operatorname{dist}(W_1, W_3) &\leq \|W_1 - W_3\widetilde{D}\widetilde{\Pi}\|_{\mathrm{ui}} \\
&\leq \|W_1 - W_2D_{12}\Pi_{12}\|_{\mathrm{ui}} + \|W_2D_{12}\Pi_{12} - W_3\widetilde{D}\widetilde{\Pi}\|_{\mathrm{ui}} \\
&= \operatorname{dist}(W_1, W_2) + \operatorname{dist}(W_2, W_3).
\end{aligned}$$

1.3. A case study: MICA

MICA [1,15,4] aims at separating linearly mixed unknown sources into statistically independent groups of signals. A basic MICA model can be stated as

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

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 $s = \left[s_1^{\mathrm{T}}, \ldots, s_t^{\mathrm{T}}\right]^{\mathrm{T}}$ with $s_j \in \mathbb{R}^{n_j}$ for $j = 1, 2, \ldots, t$, and $v = [\nu_1, \ldots, \nu_n]^{\mathrm{T}}$. 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 $\nu_1, \ldots, \nu_n$ are real stationary white random signals, mutually uncorrelated with the same variance σ^2, and independent of the sources. 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 Π is a block-wise permutation matrix. In this sense, there are certain degree

of freedoms in the determination of M. Such indeterminacy of the solution is natural, and does not matter in applications. We have the following statements.

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

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

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 $R_{ss} \in \mathbb{D}_{\tau_n}$. Often σ can be very well estimated: $\sigma \approx \hat{\sigma}$. Then we have

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

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

(b) The kurtosis[3] $\mathcal{C}_x^4$ 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

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

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

$$\widetilde{\mathcal{A}} = \{R_{xx} - \hat{\sigma} I\} \cup \{\mathcal{C}_x^4(i_1, i_2, :, :)\}_{i_1, i_2 = 1}^{n},$$

for which $W := M^{-\mathrm{T}}$ is an exact τ_n-block-diagonalizer when no noise is presented. When we attempt to numerically block-diagonalize $\widetilde{\mathcal{A}}$, what we do is to calculate an approximation $\widetilde{W}$ of $M^{-\mathrm{T}}D\Pi$ 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 there is no noise).

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 block-diagonalizable and, in the end, an approximate block-diagonalizer $\widetilde{W}$ of $\widetilde{\mathcal{A}}$ gets computed. In other words, 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 block-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.

[3] Other cumulants can also be considered.

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 written on the perturbations of JDP only. For O-JDP, Cardoso [1] presented a first order perturbation bound for a set of commuting matrices, and the result was later generalized by Russo [16]. For general JDP, using gradient flows, Afsari [17] studied sensitivity via cost functions and obtained first order perturbation bounds for the diagonalizer. Shi and Cai [18] 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 [19–22]. The uniqueness conditions of tensor decompositions, which is strongly connected to the sensitivity of tensor decompositions, received much attention recently (see, e.g., [20,23–30]). However, as to the perturbation theory for tensor decompositions, despite its importance, few results exist. Recently in [31] and [32], the condition numbers for the so-called *canonical polyadic decomposition* (CPD) and *join decomposition problem* (including the *Waring decomposition*, and some specific types of BTD, etc.) are investigated. Nonetheless, more studies are in need in the perturbation theory for various types of tensor decompositions.

1.5. Our contribution and the organization of this paper

A biggest 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, unlike many other matrix perturbation problems surveyed in [33]. Quite recently, in the sense of Definition 1.2, Cai and Liu [9] 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 widely 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, $\mathrm{eig}(A)$ 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 them. $\kappa_2(A)=\frac{\|A\|_2}{\sigma_{\min}(A)}$ denotes the matrix spectral condition number.

2. Uniquely block-diagonalizable jbdp

We begin by fixing two universal notations, throughout the rest of this paper, for the sets of matrices of interest. Let $\mathcal{A}=\{A_i\}_{i=1}^m$ be the set of m matrices, where each $A_i\in\mathbb{R}^{n\times n}$. When $\mathcal{A}$ is τ_n-block diagonalizable, i.e., (1.3) holds, we define matrix sets

$$\mathcal{A}_j=\{A_i^{(jj)}\}_{i=1}^m \quad \text{for } j=1,2,\ldots,t=\mathrm{card}(\tau_n). \tag{2.1}$$

In [9], 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 τ_n is not given but instead it asks for finding a partition τ_n with the largest cardinality such that $\mathcal{A}$ is τ_n-block-diagonalizable and at the same time a τ_n-block-diagonalizer. Via an algebraic approach, necessary and sufficient conditions [9, 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 *([9]). Given partition τ_n of n, suppose that the* JBDP *of $\mathcal{A}$ is τ_n-block diagonalizable and W is its τ_n-block-diagonalizer satisfying* (1.3)*, and assume that every $\mathcal{A}_j$ cannot be further block diagonalized,*[4] *i.e., for any partition τ_{n_j} of n_j with* $\mathrm{card}(\tau_{n_j})\geq 2$*, $\mathcal{A}_j$ is not τ_{n_j}-block-diagonalizable. Then the* JBDP *of $\mathcal{A}$ is uniquely τ_n-block-diagonalizable if and only if the matrix*

$$M_{jk}=\sum_{i=1}^m\begin{bmatrix} I_{n_k}\otimes[(A_i^{(jj)})^{\mathrm{T}}A_i^{(jj)}+A_i^{(jj)}(A_i^{(jj)})^{\mathrm{T}}] & A_i^{(kk)}\otimes A_i^{(jj)}+(A_i^{(kk)})^{\mathrm{T}}\otimes(A_i^{(jj)})^{\mathrm{T}} \\ A_i^{(kk)}\otimes A_i^{(jj)}+(A_i^{(kk)})^{\mathrm{T}}\otimes(A_i^{(jj)})^{\mathrm{T}} & [(A_i^{(kk)})^{\mathrm{T}}A_i^{(kk)}+A_i^{(kk)}(A_i^{(kk)})^{\mathrm{T}}]\otimes I_{n_j}\end{bmatrix} \tag{2.2}$$

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

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

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

$$\mathscr{N}(\mathcal{A}) := \left\{ Z \in \mathbb{R}^{n\times n} \ : \ A_i Z - Z^{\mathrm{T}} A_i = 0 \text{ for } 1 \le i \le m \right\} \tag{2.3}$$

has played an important role in the proof of [9, 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

$$A_i = \operatorname{diag}(A_i^{(11)}, \ldots, A_i^{(tt)}) \quad \text{for } 1 \le i \le m \tag{2.4}$$

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

$$\operatorname{diag}(A_i^{(11)}, \ldots, A_i^{(tt)}) Z - Z^{\mathrm{T}} \operatorname{diag}(A_i^{(11)}, \ldots, A_i^{(tt)}) = 0 \quad \text{for } 1 \le i \le m. \tag{2.5}$$

Partition Z conformally as $Z = [Z_{jk}]$, where $Z_{jk} \in \mathbb{R}^{n_j \times n_k}$. Blockwise, (2.5) can be rewritten as

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

These equations can be decoupled into two sets of matrix equations. The first set is, for $1 \le j \le t$,

$$A_i^{(jj)} Z_{jj} - Z_{jj}^{\mathrm{T}} A_i^{(jj)} = 0 \quad \text{for } 1 \le i \le m, \tag{2.7a}$$

and the second set is, for $1 \le j < k \le t$,

$$A_i^{(jj)} Z_{jk} - Z_{kj}^{\mathrm{T}} A_i^{(kk)} = 0, \quad A_i^{(kk)} Z_{kj} - Z_{jk}^{\mathrm{T}} A_i^{(jj)} = 0 \quad \text{for } 1 \le i \le m. \tag{2.7b}$$

Consider first (2.7b) for $1 \le j < k \le t$. With the help of the Kronecker product (see, e.g., [34]), they are equivalent to

$$G_{jk} \begin{bmatrix} \operatorname{vec}(Z_{jk}) \\ -\operatorname{vec}(Z_{kj}^{\mathrm{T}}) \end{bmatrix} = 0, \tag{2.8a}$$

where

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

Notice that M_{jk} defined in (2.2) simply equals to $G_{jk}^{\mathrm{T}}G_{jk}$. Thus, according to Theorem 2.1, $\mathcal{A}$ is uniquely τ_n-block-diagonalizable if and only if the smallest singular value $\sigma_{\min}(G_{jk}) > 0$, provided all $\mathcal{A}_j$ cannot be further block diagonalized.

Next, we note that (2.7a) is equivalent to

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

where

$$G_{jj} = \begin{bmatrix} I_{n_j} \otimes A_1^{(jj)} - \big[(A_1^{(jj)})^{\mathrm{T}} \otimes I_{n_j}\big]\Pi_j \\ \vdots \\ I_{n_j} \otimes A_m^{(jj)} - \big[(A_m^{(jj)})^{\mathrm{T}} \otimes I_{n_j}\big]\Pi_j \end{bmatrix}, \tag{2.9b}$$

and $\Pi_j \in \mathbb{R}^{n_j^2}$ is the perfect shuffle permutation matrix [35, Subsection 1.2.11] that enables $\Pi_j\,\mathrm{vec}(Z_{jj}^{\mathrm{T}}) = \mathrm{vec}(Z_{jj})$.

Theorem 2.2. *Suppose $\mathcal{A}$ is already in the* JBD *form with respect to $\tau_n = (n_1, \ldots, n_t)$, i.e., A_i are given by (2.4). The following statements hold.*

(a) $G_{jj}\,\mathrm{vec}(I_{n_j}) = 0$, *i.e., G_{jj} is rank-deficient;*
(b) *$\mathcal{A}_j$ cannot be further block-diagonalized if and only if for any $Z_{jj} \in \mathcal{N}(\mathcal{A}_j)$, its eigenvalues are either a single real number or a single pair of two complex conjugate numbers.*
(c) *If $\dim \mathcal{N}(\mathcal{A}_j) = 1$ 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.*

Proof. Item (a) holds because $Z = I_{n_j}$ clearly satisfies (2.7a).

For item (b), we will prove both sufficiency and necessity by contradiction.

($\Rightarrow$) Suppose there exists a $Z_{jj} \in \mathcal{N}(\mathcal{A}_j)$ such that its eigenvalues are neither a single real number nor a single pair of two complex conjugate numbers. Then Z_{jj} can be decomposed into $Z_{jj} = W_j\,\mathrm{diag}(D_1^{(j)}, D_2^{(j)})W_j^{-1}$, where W_j, $D_1^{(j)}$, $D_2^{(j)}$ are all real matrices and $\mathrm{eig}(D_1^{(j)}) \cap \mathrm{eig}(D_2^{(j)}) = \emptyset$. Then substituting the decomposition into (2.7a), we can conclude that $W_j^{\mathrm{T}}A_i^{(jj)}W_j$ for $i = 1, 2, \ldots, m$ are all block-diagonal matrices, contradicting that $\mathcal{A}_j$ cannot be further block diagonalized.

($\Leftarrow$) Assume, to the contrary, that $\mathcal{A}_j$ can be further block-diagonalized, i.e., there exists a nonsingular W_j such that $W_j^{\mathrm{T}}A_i^{(jj)}W_j = \mathrm{diag}(B_i^{(j1)}, B_i^{(j2)})$, where B_i^{j1}, $B_i^{(j2)}$ are of order n_{j1} and n_{j2}, respectively. Then

$$Z_{jj} = W_j\,\mathrm{diag}(\gamma_1 I_{n_{j1}}, \gamma_2 I_{n_{j2}})W_j^{-1} \in \mathcal{N}(\mathcal{A}_j),$$

where γ_1, γ_2 are arbitrary real numbers. That is that some $Z_{jj} \in \mathcal{N}(\mathcal{A}_j)$ can have distinct real eigenvalues, a contradiction.

Lastly for item (c), assume, to the contrary, that $\mathcal{A}_j$ can be further block-diagonalized. Without loss of generality, we may assume that there exists a nonsingular matrix $W_j \in \mathbb{R}^{n_j \times n_j}$ such that $W_j^{\mathrm{T}} A_i^{(jj)} W_j = \mathrm{diag}(A_i^{(jj1)}, A_i^{(jj2)})$ for $i = 1, 2, \ldots, m$, where $A_i^{(jj1)}$ and $A_i^{(jj2)}$ are respectively of order n_{j1} and n_{j2}. Then (2.7a) has at least two linearly independent solutions $W_j \,\mathrm{diag}(I_{n_{j1}}, 0) W_j^{-1}$, $W_j \,\mathrm{diag}(0, I_{n_{j2}}) W_j^{-1}$. Therefore, (2.9a) has two linearly 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 τ_n-block-diagonalizable $\mathcal{A}$.

Definition 2.3. Suppose that $\mathcal{A}$ is τ_n-block-diagonalizable and let $W \in \mathbb{W}_{\tau_n}$ be a τ_n-block-diagonalizer of $\mathcal{A}$ such that (1.3) holds.

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

$$\omega_{\mathrm{uq}} \equiv \omega_{\mathrm{uq}}(\mathcal{A}; W) = \min_{1 \le j < k \le t} \sigma_{\min}(G_{jk}), \tag{2.10}$$

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

(b) Suppose that none of $\mathcal{A}_j$ can be further block-diagonalized. 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 τ_n-block-diagonalizer W is defined by $\omega_{\mathrm{nd}} = \infty$ if $\tau_n = (1, 1, \ldots, 1)$ and

$$\omega_{\mathrm{nd}} = \min_{n_j > 1} \{\text{the smallest nonzero singular value of } G_{jj}\} \tag{2.11}$$

otherwise, where G_{jj} is given by (2.9b).

Note the notion of the modulus of non-divisibility is defined under the condition that none of $\mathcal{A}_j$ can be further block-diagonalized. It is needed because in order for (2.11) to be well-defined, we need to make sure that G_{jj} has at least one nonzero singular value in the case when $n_j > 1$. Indeed, $G_{jj} \neq 0$ whenever $n_j > 1$, if none of $\mathcal{A}_j$ can be further block-diagonalized. To see this, we note $G_{jj} = 0$ implies that any matrix Z_{jj} of order n_j is a solution to (2.7a) 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 the assumption that none of $\mathcal{A}_j$ can be further block-diagonalized.

The proposition below partially justifies Definition 2.3.

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

(a) *$\mathcal{A}$ is uniquely τ_n-block-diagonalizable if* $\omega_{\rm uq}(\mathcal{A};W)>0$.
(b) *None of $\mathcal{A}_j$ can be further block-diagonalized and*

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

Remark 2.5. A few comments are in order.

(a) The definition of $\omega_{\rm uq}$ is a natural generation of the modulus of uniqueness in [18] for JDP (i.e., when $\tau_n=(1,1,\ldots,1)$).
(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 (2.11), 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 $\dim\mathcal{N}(\mathcal{A}_j)=2$, i.e., the second smallest singular value of G_{jj} is still 0. As an example, we consider $A_i=\left[\begin{smallmatrix}\alpha_i & \beta_i\\ \beta_i & -\alpha_i\end{smallmatrix}\right]\in\mathbb{R}^{2\times2}$ for $i=1,2,\ldots,m$, where $\beta_i\neq0$ for all i and α_i/β_i's are not a constant. Then $\mathcal{A}$ cannot be simultaneously diagonalized but $\mathcal{N}(\mathcal{A})=\operatorname{span}\{I_2,\left[\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right]\}$, i.e., $\dim\mathcal{N}(\mathcal{A})=2$.

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

Theorem 2.6. *If $\mathcal{A}$ is uniquely τ_n-block-diagonalizable, then $\omega_{\rm uq}$ and $\omega_{\rm nd}$ are both independent of the choice of diagonalizers $W\in\mathbb{W}_{\tau_n}$.*

Proof. Let $W\in\mathbb{W}_{\tau_n}$ be a τ_n-block-diagonalizer of $\mathcal{A}$. Then all possible τ_n-block-diagonalizers 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 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 Π_j is a permutation matrix of order n_j. 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 (2.8b), but in terms of $\widetilde{A}_i^{(jj)}$ and $\widetilde{A}_i^{(kk)}$, $\widetilde{G}_{jj}$, accordingly as G_{jj} in (2.9b), but in terms of $\widetilde{A}_i^{(jj)}$. Then by calculations, we have

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

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. □

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.

Recall $\mathcal{A} = \{A_i\}_{i=1}^m$ which is the unperturbed 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 $t = \mathrm{card}(\tau_n) \geq 2$. We assume that

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

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).

Suppose that $\mathcal{A}$ 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 τ_n-block-diagonalizable for $m \geq 3$. This means that $\widetilde{\mathcal{A}}$ may not be τ_n-block-diagonalizable regardless how tiny $\delta_{\mathcal{A}}$ may be. For this reason, we will not assume that $\widetilde{\mathcal{A}}$ is τ_n-block-diagonalizable, but, instead, it has an approximate τ_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 has two advantages. Firstly, it serves all practical purposes well, because in any likely practical situations we usually end up with $\widetilde{\mathcal{A}}$ which is close to some τ_n-block-diagonalizable $\mathcal{A}$, which is not actually available due to unavoidable noises such as in MICA, and, at the same time, an approximate τ_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 actually τ_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 $\max_j |\gamma_j| = 1$, and define the τ_n-block-diagonalzablility 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 Γ is. The rationale behind defining these residuals is in the following proposition.

Proposition 3.1. *$\widetilde{W}^{\mathrm{T}}\widetilde{A}_i\widetilde{W}$ is τ_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 Γ with distinct diagonal entries suffices. But later, we will see that our upper bound depends on Γ, which makes us wonder what the best Γ is for the best possible bound. Unfortunately, this is not a trivial task and would be an interesting subject for future studies. Later in our numerical tests, we use a few random Γ along with the following one

$$\Gamma = \mathrm{diag}(-I_{n_1}, (-1 + \frac{2}{t-1})I_{n_2}, (-1 + \frac{4}{t-1})I_{n_3}, \ldots, I_{n_t}). \tag{3.5}$$

Nonetheless, we still keep Γ as a parameter to choose in our main result in hope that it may come to help in certain circumstance. We restrict γ_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. Set

$$g = \min_{j\neq k} |\gamma_j - \gamma_k|, \quad \tilde{r} = \left(\sum_{i=1}^{m} \|\widetilde{R}_i\|_{\mathrm{F}}^2\right)^{1/2}. \tag{3.6}$$

The quantity $\tilde{r}$ will be used to measure how good $\widetilde{W}$ is in approximately diagonalizing $\widetilde{\mathcal{A}}$. In fact, it can be verified that

$$g\left(\sum_{i=1}^{m} \|\,\mathrm{OffBdiag}_{\tau_n}(\widetilde{W}^{\mathrm{T}}\widetilde{A}_i\widetilde{W})\|_F^2\right)^{\frac{1}{2}} \le \tilde{r} \le 2\left(\sum_{i=1}^{m} \|\,\mathrm{OffBdiag}_{\tau_n}(\widetilde{W}^{\mathrm{T}}\widetilde{A}_i\widetilde{W})\|_F^2\right)^{\frac{1}{2}}, \tag{3.7}$$

which implies that $\tilde{r}$ is *comparable* to $\left(\sum_{i=1}^{m} \|\,\mathrm{OffBdiag}_{\tau_n}(\widetilde{W}^{\mathrm{T}}\widetilde{A}_i\widetilde{W})\|_F^2\right)^{\frac{1}{2}}$, provided g is not too tiny. For the choice (3.5), $g = 2/(t-1)$.

The next proposition is due to an anonymous referee and it improves our earlier version.

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

$$\frac{\|\mathcal{E}\|_{\mathrm{F}}}{\|\widetilde{\mathcal{A}}\|_{\mathrm{F}}} \le \frac{\|\widetilde{W}^{-1}\|_2^2}{\|\widetilde{\mathcal{A}}\|_{\mathrm{F}}} \cdot \left(\sum_{i=1}^{m} \| \operatorname{OffBdiag}_{\tau_n}(\widetilde{W}^{\mathrm{T}} \widetilde{A}_i \widetilde{W}) \|_F^2 \right)^{\frac{1}{2}} =: \varepsilon_{\mathrm{bker}}(\widetilde{\mathcal{A}}; \widetilde{W}), \tag{3.8}$$

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. Let $\widetilde{E}_i = \operatorname{OffBdiag}_{\tau_n}(\widetilde{W}^{\mathrm{T}} \widetilde{A}_i \widetilde{W})$ and $E_i = -\widetilde{W}^{-\mathrm{T}} \widetilde{E}_i \widetilde{W}^{-1}$ for $1 \le i \le m$. Then

$$\begin{aligned} \operatorname{OffBdiag}_{\tau_n}(\widetilde{W}^{\mathrm{T}}[\widetilde{A}_i + E_i]\widetilde{W}) &= \operatorname{OffBdiag}_{\tau_n}(\widetilde{W}^{\mathrm{T}} \widetilde{A}_i \widetilde{W} - \widetilde{E}_i) \\ &= \operatorname{OffBdiag}_{\tau_n}(\widetilde{W}^{\mathrm{T}} \widetilde{A}_i \widetilde{W}) - \widetilde{E}_i = 0. \end{aligned}$$

Thus $\widetilde{W}$ is an exact τ_n-block-diagonalizer of the matrix set $\{\widetilde{A}_i + E_i\}_{i=1}^m$. It can be verified that $\mathcal{E} = \{E_i\}_{i=1}^m$ satisfies (3.8). □

3.2. Main result

With the setup, we are ready to state our main result.

Theorem 3.3. *Assume* (3.1) *and that* $\mathcal{A}$ *is perturbed to* $\widetilde{\mathcal{A}}$, *for which* $\widetilde{W}$ *is an approximate* τ_n*-block-diagonalizer. Let* $Q = W^{-1}\widetilde{W}$, *and let* ω_{uq} *and* ω_{nd} *be defined in* Definition 2.3, *and*[5]

$$\tau = \frac{\sqrt{2}-1}{\sqrt{t-1}}, \quad \alpha = \frac{2\tau}{(\sqrt{2}+\tau)^2}, \tag{3.9}$$

$$\delta = \|Q^{-1}\|_2^2\, \tilde{r} + 2\|Q^{-1}\|_2 \|W\|_2 \|\widetilde{W}\|_2\, \delta_{\mathcal{A}}, \quad \epsilon_* = \frac{\tau \kappa_2(Q)\delta}{\alpha g\, \omega_{\mathrm{uq}}}, \tag{3.10}$$

where g *and* $\tilde{r}$ *are given by* (3.6). *If*

$$\delta < \min\left\{ \frac{\alpha g\, \omega_{\mathrm{uq}}}{\kappa_2(Q)}, \ \frac{(1-2\alpha) g\, \omega_{\mathrm{nd}}}{\sqrt{2}} \right\}, \tag{3.11}$$

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

$$\min_{\substack{D \in \mathbb{D}_{\tau_n}, D^{\mathrm{T}} D = I \\ \Pi \in \mathbb{P}_{\tau_n}}} \frac{\|W - \widetilde{W} D \Pi\|_{\mathrm{p}}}{\|\widetilde{W}\|_{\mathrm{p}}} \le \frac{1 + \sqrt{t}\,\epsilon_*}{\sqrt{1 - 2\sqrt{t-1}\epsilon_* - (t-1)\epsilon_*^2}} - 1 \tag{3.12}$$

$$= \frac{\tau}{\alpha} \cdot \frac{(\sqrt{t} + \sqrt{t-1})\kappa_2(Q)\delta}{g\, \omega_{\mathrm{uq}}} + O(\delta^2) := \varepsilon_{\mathrm{ub}}. \tag{3.13}$$

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

As we commented before, there is a hidden matrix parameter Γ to choose in applying this theorem. It is there to make the theorem more versatile for a better bound sometimes. A weaker version of this theorem is to specialize Γ to the one in (3.5). Specifically, Theorem 3.3 remains valid upon replacing g by $2/(t-1)$ and $\tilde{r}$ by $2\left(\sum_{i=1}^{m} \|\operatorname{OffBdiag}_{\tau_n}(\widetilde{W}^{\mathrm{T}}\widetilde{A}_i\widetilde{W})\|_F^2\right)^{\frac{1}{2}}$.

Next we look at two illustrating examples and discuss the implications of Theorem 3.3.

Example 3.1. Let $A_1 = I_2$, $A_2 = \operatorname{diag}(1, 1+\varsigma)$, where $\varsigma > 0$ is a parameter. It is clear that $W = I_2$ is a diagonalizer of $\mathcal{A} = \{A_1, A_2\}$ with respect to $\tau_2 = (1,1)$. By calculations according to (2.10), we get

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

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 see that if θ and ϵ are sufficiently small, $\widetilde{W}$ is a good block-diagonalizer. Now choose $\Gamma = \operatorname{diag}(-1, 1)$. We have

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

Thus, if $\theta = \epsilon$ and $\epsilon \ll 1$, then (3.11) is satisfied. Thus, by (3.12), for $\mathrm{p} \in \{2, \mathrm{F}\}$

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

Therefore, as long as ς is not too small, ω_{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 ς is small, say $\varsigma = \epsilon^{\phi}$ with $0 < \phi < 1$, then $\widetilde{W}$ is always a good block-diagonalizer, independent of θ, 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 good upper bound for the relative error in $\widetilde{W}$.

Example 3.2. Let $A_1 = \mathrm{diag}(I_2, \begin{bmatrix} 1 & 1+\varsigma \\ 1 & 1 \end{bmatrix})$, $A_2 = \mathrm{diag}(\begin{bmatrix} 1 & 1+\varsigma \\ 1 & 1 \end{bmatrix}, I_2)$, where $\varsigma > 0$ is a parameter. Then $W = I_4$ is a τ_4-block-diagonalizer of $\mathcal{A} = \{A_1, A_2\}$, where $\tau_4 = (2, 2)$. According to (2.10), we calculate ω_{uq} to get

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

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

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

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

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

Therefore, if θ and ϵ are sufficiently small, then $\widetilde{W}$ is a good block-diagonalizer. Now let $\Gamma = \mathrm{diag}(-I_2, I_2)$. 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 ς is not too small, then (3.11) is satisfied. Thus, by (3.12), for $\mathrm{p} \in \{2, \mathrm{F}\}$

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

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 ς is small, say $\varsigma = \epsilon^{\phi}$ with $\phi > 0$, then the condition (3.11) of Theorem 3.3 is likely violated, and consequently, Theorem 3.3 is no longer applicable.

From these two examples, we can see that the bound $\varepsilon_{\mathrm{ub}}$ in (3.12) is good in the sense that it can be in the same order of magnitude as the relative error. But when ω_{uq} and/or ω_{nd} is small, Theorem 3.3 may not provide a good bound or even fails to give a bound. This observation is more or less expected. In fact, when ω_{uq} and/or ω_{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 change 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*

$$\min_{\substack{D\in\mathbb{D}_{\tau_n},D^{\mathrm{T}}D=I\\ \Pi\in\mathbb{P}_{\tau_n}}}\frac{\|W-\widetilde{W}D\Pi\|_{\mathrm{p}}}{\|\widetilde{W}\|_{\mathrm{p}}}\leq\frac{\tau}{\alpha}\cdot\frac{(\sqrt{t}+\sqrt{t-1})\delta}{g\,\omega_{\mathrm{uq}}}+O(\delta^2). \tag{3.14}$$

Some of the quantities in the right-hand side of (3.12) 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 ω_{uq} and ω_{nd} which are intrinsic to the JBDP for $\mathcal{A}$ may well be estimated by those of $\widehat{\mathcal{A}}=\left\{\mathrm{Bdiag}_{\tau_n}(\widetilde{W}^{\mathrm{T}}\widetilde{A}\widetilde{W})\right\}_{i=1}^{m}$. Finally, $\|W\|_2\geq 1$ for any $W\in\mathbb{W}_{\tau_n}$, because $\|W\|_2$ is equal to the square root of the largest eigenvalue of $W^{\mathrm{T}}W$ and the latter is no smaller than the largest diagonal entry of $W^{\mathrm{T}}W$, which is 1. On the other hand, write $W=[W_1,W_2,\ldots,W_t]$ with $W_j\in\mathbb{R}^{n\times n_j}$ and $W_j^{\mathrm{T}}W_j=I_{n_j}$. For any $x=\left[x_1^{\mathrm{T}},x_2^{\mathrm{T}},\ldots,x_t^{\mathrm{T}}\right]^{\mathrm{T}}$ with $x_j\in\mathbb{R}^{n_j}$, we have

$$\begin{aligned}\|Wx\|_2&=\|W_1x_1+\cdots+W_tx_t\|_2\leq\|W_1x_1\|_2+\cdots+\|W_tx_t\|_2\\&=\|x_1\|_2+\cdots+\|x_t\|_2\leq\sqrt{t}\sqrt{\|x_1\|_2^2+\cdots+\|x_t\|_2^2}=\sqrt{t}\,\|x\|_2,\end{aligned}$$

which implies $\|W\|_2\leq\sqrt{t}$. Therefore, we get

$$1\leq\|W\|_2\leq\sqrt{t}. \tag{3.15}$$

The same holds for $\widetilde{W}$, too.

Remark 3.5. Several comments are in order.

(a) The quantity δ in (3.10) consists of two parts: the first part indicates how good $\widetilde{W}$ is in approximately block-diagonalizing $\widetilde{\mathcal{A}}$, and the second part indicates how large the perturbation is. Therefore, the condition (3.11) means that the block-diagonalizer $\widetilde{W}$ has to be sufficiently good and the perturbation has to be sufficiently small so that δ does not exceed the right-hand side of (3.11), which is proportional to the moduli ω_{uq} and ω_{nd}. Although the modulus of non-divisibility ω_{nd} does not appear explicitly in the upper bound, it limits the size of δ.

(b) In (3.12), $\varepsilon_{\mathrm{ub}}$ is a monotonically increasing function in δ and $\kappa_2(Q)$. If W (or $\widetilde{W}$) is ill-conditioned, then both δ and $\kappa_2(Q)$ may be large, as a result, $\varepsilon_{\mathrm{ub}}$ can be large.

(c) If $\delta\ll 1$, by (3.12), we have

$$\min_{D,\Pi}\frac{\|W-\widetilde{W}D\Pi\|_{\mathrm{p}}}{\|\widetilde{W}\|_{\mathrm{p}}}\leq\frac{\tau}{\alpha}\cdot\frac{(\sqrt{t}+\sqrt{t-1})\kappa_2(Q)}{\omega_{\mathrm{uq}}}\cdot\frac{\delta}{g}+O(\delta^2). \tag{3.16}$$

(d) As far as sensitivity is concerned, the factor τ/α in (3.13) is insignificant because it can be bounded as

$$1 < \frac{\tau}{\alpha} = \frac{1}{2}\left(\sqrt{2} + \frac{\sqrt{2}-1}{\sqrt{t-1}}\right)^2 \le \frac{1}{2}(2\sqrt{2}-1)^2$$

for $t \ge 2$. Thus $\varepsilon_{\rm ub}$ is proportional to $\sqrt{t}$, other things being equal.

(e) Recalling how $\widetilde{R}_i$ and g are defined, we find that Γ can affect the quality of the upper bound provided by Theorem 3.3 significantly. We need all γ_j well-separated from each other, lest g will be tiny. Ideally, we would like to minimize the upper bound over Γ, which does not seem to be an easy thing to do. In both Examples 3.1 and 3.2, we picked the particular Γ in (3.5).

(f) 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 block-diagonalizers, i.e., both JBDP 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 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 (3.16)

$$\min_{D,\Pi} \frac{\|W - \widetilde{W}D\Pi\|_{\rm p}}{\|\widetilde{W}\|_{\rm p}} \le \frac{\tau}{\alpha} \cdot \frac{(\sqrt{t}+\sqrt{t-1})\kappa_2(Q)\|Q^{-1}\|_2^2}{\omega_{\rm uq}} \cdot \frac{\tilde{r}}{g} + O(\tilde{r}^2). \tag{3.17}$$

This error bound is $O(\frac{\tilde{r}}{\omega_{\rm uq}})$, which is in agreement with the error bound when applied to JDP in [18, Corollary 3.2].

4. Proof of Theorem 3.3

Recall the assumptions: $\mathcal{A}$ is τ_n-block-diagonalizable and $W \in \mathbb{W}_{\tau_n}$ is a τ_n-block-diagonalizer such that (1.3) holds. The modulus of uniqueness $\omega_{\rm uq}$ and the modulus of non-divisibility $\omega_{\rm nd}$ 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 τ_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*

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

for $1 \le i \le m$. *Partition* $Z = [Z_{jk}]$ *with* $Z_{jk} \in \mathbb{R}^{n_j \times n_k}$ *and let* $\mathrm{eig}(Z_{jj}) = \{\mu_{jk}\}_{k=1}^{n_j}$. *The following statements hold.*

(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 $R_i = \left[R_i^{(jk)}\right]$ conformally with respect to τ_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 (2.8b). Put them all together to get

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

where

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

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

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

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

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

where G_{jj} is defined by (2.9b). Since $\dim \mathscr{N}(\mathcal{A}_j) = 1$ by assumption, we know by Theorem 2.2(a) that the null space of G_{jj} is spanned by $\mathrm{vec}(I_{n_j})$, and thus there exists a real number $\hat{\mu}_j$ such that

$$\mathrm{vec}(Z_{jj}) = G_{jj}^{\dagger} r_{jj} + \hat{\mu}_j\,\mathrm{vec}(I_{n_j}),$$

where $G_{jj}^{\dagger}$ is the Moore-Penrose inverse [36, p.102] of G_{jj}. It follows immediately that

$$Z_{jj} = \widehat{Z}_{jj} + \hat{\mu}_j I_{n_j},$$

where $\widehat{Z}_{jj} = \mathrm{reshape}(G_{jj}^{\dagger} r_{jj}, n_j, n_j)$. In particular, $\mathrm{eig}(\widehat{Z}_{jj}) = \{\mu_{jk} - \hat{\mu}_j\}_{k=1}^{n_j}$ and hence

$$\sum_{k=1}^{n_j} |\mu_{jk} - \hat{\mu}_j|^2 \le \|\widehat{Z}_{jj}\|_{\mathrm{F}}^2 \le \frac{\|r_{jj}\|_2^2}{\omega_{\mathrm{nd}}^2} \le \frac{\sum_{i=1}^m \|R_i^{(jj)}\|_{\mathrm{F}}^2}{\omega_{\mathrm{nd}}^2} \le \frac{\sum_{i=1}^m \|\,\mathrm{Bdiag}_{\tau_n}(R_i)\|_{\mathrm{F}}^2}{\omega_{\mathrm{nd}}^2}.$$

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 $B = \mathrm{Bdiag}_{\tau_n}(Z)$ and $E = -\,\mathrm{OffBdiag}_{\tau_n}(Z)$. Let τ and α be as in (3.9) and g as in (3.6). If*

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

then there exists a τ_n-block-diagonal matrix $\widetilde{B} = \mathrm{diag}(\widetilde{B}_{11}, \ldots, \widetilde{B}_{tt})$ and a nonsingular matrix $P = [P_{jk}]$ 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 $j = 1, 2, \ldots, t$

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

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

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

$$\widehat{P}_j = \left[P_{1j}^{\mathrm{T}}, \quad \ldots, \quad P_{j-1,j}^{\mathrm{T}}, \quad 0_{n_j \times n_j}, \quad P_{j+1,j}^{\mathrm{T}}, \quad \ldots, \quad P_{tj}^{\mathrm{T}} \right]^{\mathrm{T}}. \tag{4.6c}$$

Proof. It suffices to show that 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

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

where $\mathrm{sep}_{\mathrm{F}}(\cdots)$ is the separation of two matrices [36, p.247]. Let $\tilde{g} = g - \|E_{11}\|_{\mathrm{F}} - \|E_{22}\|_{\mathrm{F}}$. By [36, Theorem 2.8 on p.238], we conclude that if

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

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

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

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),

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

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

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

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

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 $\mathrm{eig}(\widetilde{B}_{11})=\{\tilde{\mu}_{1k}\}_{k=1}^{n_1}$, 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 $x_j=-\sum_{\substack{k=1\\k\neq j}}^{t}P_{jk}x_k$ and thus

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

Therefore

$$\begin{aligned}\|x\|_2^2=\sum_{j=1}^{t}\|x_j\|_2^2 &\le (t-1)\sum_{j=1}^{t}\sum_{\substack{k=1\\k\neq j}}^{t}\|P_{jk}\|_2^2\|x_k\|_2^2\\ &= (t-1)\sum_{k=1}^{t}\sum_{\substack{j=1\\j\neq k}}^{t}\|P_{jk}\|_2^2\|x_k\|_2^2\end{aligned}$$

$$\leq (t-1)\sum_{k=1}^{t}\|\widehat{P}_k\|_{\mathrm{F}}^2\|x_k\|_2^2$$
$$< (t-1)\tau^2\|x\|_2^2 < \|x\|_2^2,$$

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 γ_j.

Lemma 4.4. *Let $P = [P_{jk}]$ 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 τ is defined by (3.9). 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 σ 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

$$P = [P_1, \quad \ldots, \quad P_t] \text{ with } P_j \in \mathbb{R}^{n\times n_j}.$$

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 jth 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 $1 \le j' \le t$ as a result of the permutation Π. Partition $\widetilde{W}$ as $\widetilde{W} = \big[\widetilde{W}_1, \ldots, \widetilde{W}_t\big]$ 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 $\big\|\widetilde{W}_j^{\mathrm{T}}\widetilde{W}_\ell\big\|_2 \le 1$. To see the last inequality, we note

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

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 $\|\widehat{P}_{j'}\|_{\mathrm{F}} \le \epsilon$, we have

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

Combining it with (4.16), we get

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

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

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

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

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 $\mathrm{eig}(Z_{jj}) = \{\mu_{jk}\}_{k=1}^{n_j}$. The proof will be completed in the following four steps:

Step 1. We will show that Z is approximately τ_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 π of $\{1, 2, \ldots, t\}$ such that

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

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

Step 3. We will show that there exist a permutation $\Pi \in \mathbb{P}_{\tau_n}$ and a nonsingular $P \equiv [P_{jk}] \in \mathbb{R}^{n\times n}$ 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 (3.12).

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

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

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

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

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

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

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

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

By a result of Kahan [37] (see also [38, Remark 3.3]), we have

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

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

$$\begin{aligned}4\alpha^2g^2 &> 4\kappa_2^2(Q)\frac{\delta^2}{\omega_{\rm uq}^2} && \text{(by (3.11))}\\ &\ge 4\|Q^{-1}\operatorname{OffBdiag}_{\tau_n}(Z)Q\|_{\rm F}^2 && \text{(by (4.18))} \qquad (4.23a)\\ &\ge 2\sum_{j=1}^{t}\sum_{k=1}^{n_j}|\mu_{jk}-\gamma_{\ell_{jk}}|^2 && \text{(by (4.22))}\\ &\ge 2(|\mu_{j1}-\gamma_{\ell_{j1}}|^2+|\mu_{j2}-\gamma_{\ell_{j2}}|^2)\\ &\ge (|\mu_{j1}-\gamma_{\ell_{j1}}|+|\mu_{j2}-\gamma_{\ell_{j2}}|)^2\\ &\ge (|\gamma_{\ell_{j1}}-\gamma_{\ell_{j2}}|-|\mu_{j1}-\mu_{j2}|)^2\\ &\ge \left(g-\sqrt{2}\,\frac{\delta}{\omega_{\rm nd}}\right)^2 && \text{(by (4.21))}\\ &> [1-(1-2\alpha)]^2g^2 && \text{(by (3.11))}\\ &= 4\alpha^2g^2, && (4.23b)\end{aligned}$$

a contradiction. Now using (4.22), (4.18) and (3.11), we get

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

Thus, we know that each $j \in \{1, 2, \ldots, t\}$ corresponds to 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). □

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

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

Denote by $\mathrm{eig}(\widetilde{B}_{jj}) = \{\tilde{\mu}_{jk}\}_{k=1}^{n_j}$. By (4.6b), (4.18) and (3.11), we know

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

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

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

Let Π be the permutation matrix such that

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

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

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

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

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

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

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 σ of $\widetilde{D}$, and (4.14) holds with

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

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 $U = \mathrm{diag}(U_1,\ldots,U_t)$, $V = \mathrm{diag}(V_1,\ldots,V_t)$ and $D = \Pi V U^{\mathrm{T}}\Pi^{\mathrm{T}}$. It can be verified that D is orthogonal and τ_n-block-diagonal. It follows from $W\widetilde{D} = \widetilde{W}P\Pi$ that

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

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

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

Combine it with (4.28) to conclude the proof of (3.12). □

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, 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 the number m of matrices, we generate the matrix sets $\mathcal{A}$ 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 τ_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 τ_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 ξ is a parameter for controlling noise level. $\widetilde{\mathcal{A}}$ is likely not τ_n-block-diagonalizable but it is approximately. An approximate block-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 [22] 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

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

over orthogonal $D_{\pi(j)}$, permutations π 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 notation a little bit, we let π be the one that achieve the optimal in (5.1), 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 computed as

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

with D as above and $\Pi \in \mathbb{P}_{\tau_n}$ as determined by the optimal π. There 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 Γ to use. Ideally, we should use the one that minimize the right-hand side of (3.12), but we don't have a simple way of doing that. For the tests below, we use 50 different Γ and pick the best bound. Specifically, we use a particular one in (3.5)

$$\Gamma = \mathrm{diag}(-I_{n_1}, (-1 + \frac{2}{t-1})I_{n_2}, (-1 + \frac{4}{t-1})I_{n_3}, \ldots, I_{n_t}) \tag{3.5}$$

as well as 49 random ones with their diagonal entries $\gamma_1, \ldots, \gamma_t$ randomly drawn from the interval $[-1, 1]$ with the uniform distribution. Our experience suggests that the particular

Table 5.1
Bound vs. m, the number of matrices in $\mathcal{A}$ for $\tau_9 = (3, 3, 3)$.

m	ω_{uq}	ω_{nd}	δ	Ratio	$\varepsilon_{\text{bker}}$	ε_{ub}	Error
4	2.4e+00	2.3e+00	6.3e−10	1.3e−09	1.0e−09	1.2e−09	2.3e−11
8	4.0e+00	4.1e+00	8.6e−10	1.1e−09	8.1e−10	9.9e−10	2.4e−11
16	6.7e+00	6.2e+00	1.3e−09	9.6e−10	6.8e−10	8.9e−10	2.3e−11
32	1.1e+01	1.1e+01	2.0e−09	8.9e−10	6.3e−10	8.2e−10	2.5e−11
64	1.7e+01	1.7e+01	1.4e−09	1.8e−09	9.1e−10	1.7e−09	4.0e−11
128	2.5e+01	2.5e+01	3.0e−09	1.3e−09	7.9e−10	1.2e−09	3.4e−11
256	3.6e+01	3.6e+01	3.5e−09	4.9e−10	2.9e−10	4.5e−10	1.4e−11

Table 5.2
Bound vs. m, the number of matrices in $\mathcal{A}$ for $\tau_6 = (1, 2, 3)$.

m	ω_{uq}	ω_{nd}	δ	Ratio	$\varepsilon_{\text{bker}}$	ε_{ub}	Error
4	2.1e+00	3.4e+00	6.0e−11	4.2e−10	2.2e−11	3.9e−10	3.5e−12
8	3.0e+00	5.2e+00	8.9e−11	2.9e−10	1.9e−11	2.7e−10	3.6e−12
16	6.2e+00	7.7e+00	1.5e−10	1.7e−10	2.5e−11	1.6e−10	3.8e−12
32	8.9e+00	1.1e+01	1.9e−10	2.2e−10	2.5e−11	2.1e−10	3.8e−12
64	1.3e+01	1.5e+01	2.4e−10	1.4e−10	1.3e−11	1.3e−10	1.7e−12
128	1.7e+01	2.2e+01	4.2e−10	1.2e−10	1.3e−11	1.1e−10	1.7e−12
256	2.4e+01	3.3e+01	4.7e−10	1.3e−10	9.2e−12	1.2e−10	1.1e−12

Γ in (3.5) usually leads to bounds having the same order as the best one produced by the 49 random Γ. However, it can happen that the best one is much better than and up to one tenth of that by the particular Γ, although such extremes do not happen very often.

We will report 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), and varying noise levels (test 4). We will examine these quantities: the modulus of uniqueness ω_{uq}, the modulus of non-divisibility ω_{nd}, the quantity δ as defined in (3.10), the *ratio* as the quotient of δ over the right hand side of (3.11) (to make sure that (3.11) is satisfied), the upper bound $\varepsilon_{\text{bker}} \equiv \varepsilon_{\text{bker}}(\widetilde{\mathcal{A}}; \widetilde{W})$ on the relative backward error as in (3.8), the perturbation bound ε_{ub} as in (3.12), and finally the estimated true *error* in $\widetilde{W}$ as in (5.2).

Test 1: number of matrices. In this test, we fix $\xi = 10^{-12}$ and vary the number m of matrices in the matrix set $\mathcal{A}$. The numerical results are displayed in Tables 5.1 and 5.2 for the two different partitions $\tau_9 = (3, 3, 3)$ and $\tau_6 = (1, 2, 3)$, respectively. We summarize our observations from Tables 5.1 and 5.2 as follows.

1. For all m, the *ratio*s are far less than 1. In other words, (3.11) is satisfied for all, and hence the bound by (3.12) can be used.
2. For all m, ε_{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.

Table 5.3
Bound vs. matrix size $n = 9p$ for $\tau_n = p \times (3, 3, 3)$.

n	ω_{uq}	ω_{nd}	δ	Ratio	$\varepsilon_{\text{bker}}$	ε_{ub}	Error
9	7.1e+00	7.3e+00	3.9e−10	4.0e−10	9.5e−11	3.7e−10	8.3e−12
18	1.1e+01	1.0e+01	1.4e−09	6.6e−10	2.6e−10	6.1e−10	1.4e−11
27	1.2e+01	1.2e+01	3.9e−09	1.6e−09	7.7e−10	1.4e−09	2.5e−11
36	1.5e+01	1.5e+01	1.2e−07	5.2e−08	7.9e−08	4.8e−08	9.6e−10
45	1.6e+01	1.6e+01	8.9e−09	3.4e−09	3.8e−09	3.2e−09	4.8e−11
54	1.8e+01	1.8e+01	3.8e−07	1.1e−07	2.0e−07	9.8e−08	1.6e−09
63	1.9e+01	1.9e+01	9.9e−09	4.2e−09	3.0e−09	3.9e−09	4.4e−11

Table 5.4
Bound vs. matrix size $n = 6p$ for $\tau_n = p \times (1, 2, 3)$.

n	ω_{uq}	ω_{nd}	δ	Ratio	$\varepsilon_{\text{bker}}$	ε_{ub}	Error
6	4.4e+00	7.2e+00	1.1e−10	6.6e−10	2.5e−11	6.1e−10	3.4e−12
12	5.8e+00	6.2e+00	1.3e−09	1.6e−09	5.3e−10	1.5e−09	2.4e−11
18	8.5e+00	8.0e+00	1.3e−09	1.9e−09	7.7e−10	1.8e−09	2.5e−11
24	9.7e+00	9.1e+00	6.8e−09	5.6e−09	4.5e−09	5.2e−09	6.3e−11
30	9.8e+00	9.1e+00	2.5e−09	2.1e−09	1.3e−09	1.9e−09	2.0e−11
36	1.1e+01	9.4e+00	4.3e−09	2.5e−09	1.3e−09	2.3e−09	2.9e−11
42	1.2e+01	1.1e+01	3.7e−09	1.5e−09	9.3e−10	1.4e−09	1.6e−11

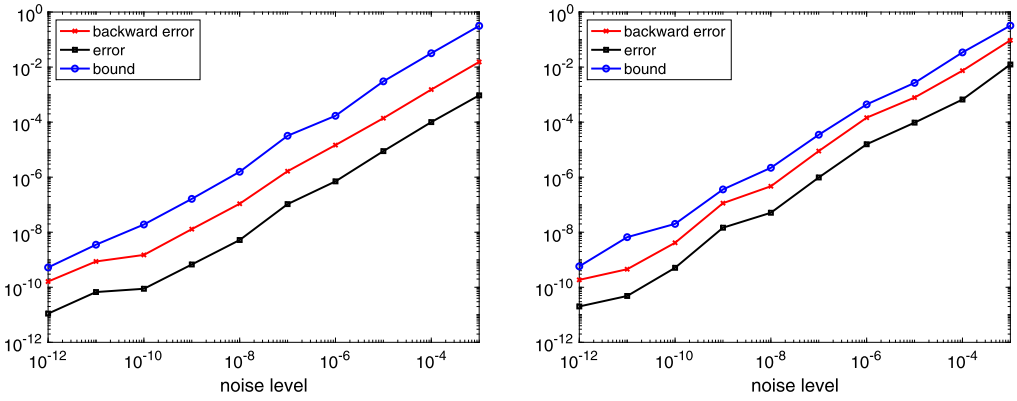
Table 5.5
Bound vs. number of diagonal blocks.

t	ω_{uq}	ω_{nd}	δ	Ratio	$\varepsilon_{\text{bker}}$	ε_{ub}	Error
3	5.0e+00	9.9e+00	1.8e−10	3.0e−10	3.4e−11	2.7e−10	1.1e−12
4	5.6e+00	5.2e+00	4.4e−10	6.7e−10	4.8e−11	6.0e−10	4.8e−12
5	4.0e+00	8.5e+00	6.0e−10	8.2e−09	1.6e−10	7.2e−09	1.4e−11
6	3.2e+00	9.5e+00	2.1e−09	1.2e−08	8.7e−10	1.0e−08	3.2e−11
7	3.3e+00	8.3e+00	2.5e−09	1.7e−08	1.4e−09	1.4e−08	1.1e−10
8	6.5e+00	5.8e+00	8.6e−09	3.7e−08	8.1e−09	3.2e−08	3.2e−10
9	5.9e+00	6.1e+00	3.5e−09	2.0e−08	2.0e−09	1.7e−08	5.0e−11

Test 2: matrix sizes. In this test, we fix $\xi = 10^{-12}$, $m = 16$, and use two partitions $\tau_n = p \times (3, 3, 3)$ or $\tau_n = p \times (1, 2, 3)$, where $p = 1, 2, \ldots, 7$. Then the matrix size $n = 9p$ or $6p$ will increase as p increases. We display the numerical results in Tables 5.3 and 5.4. We can see from Tables 5.3 and 5.4 that ε_{ub} provides a very good upper bound on the *error* for different sizes of matrices.

Test 3: number of diagonal blocks. In this test, we fix $\xi = 10^{-12}$, $m = 16$, and generate the partition τ_n randomly using MATLAB command `randi(5,t,1)`. In other words, the block-diagonal matrices D_j have t diagonal blocks and the order of the ith block is $\tau_n(i)$, randomly drawn from $\{1, 2, \ldots, 5\}$ with the uniform distribution. For $t = 3, 4, \ldots, 9$, we display the numerical results in Table 5.5. We can see from Table 5.5 that ε_{ub} provides a very good upper bound on the *error* as the numbers of diagonal blocks varies.

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 Fig. 5.1, we plot $\varepsilon_{\text{bker}}$ (*backward error*), *error* and ε_{ub} (*bound*) versus different noise levels. We can see from Fig. 5.1 that as ξ



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

Fig. 5.1. Backward error $\varepsilon_{\text{bker}}$, *error*, and bound ε_{ub} vs. noise level.

increases, $\varepsilon_{\text{bker}}$, *error* and ε_{ub} all increase almost linearly. For all noise levels, ε_{ub} indeed provides a good upper bound on the *error*.

6. Concluding remarks

In this paper, we developed a perturbation theory for JBDP. An upper bound is obtained for the relative distance (1.5) between a block-diagonalizer W for the original JBDP of $\mathcal{A}$ that is block-diagonalizable and an approximate diagonalizer $\widetilde{W}$ for its perturbed JBDP of $\widetilde{\mathcal{A}}$. The backward error is also derived for JBDP. Numerical tests that validate the theoretical results are presented.

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 not much simplification in results and arguments can be gained but all the results remain valid after minor changes to the definitions of G_{jk} in (2.8b): remove the second, fourth, ..., block rows as now all $A_i^{(jj)}$ are symmetric.

We have been limiting all matrices to real ones, but this is not a limitation. 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.

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 τ_n-block-diagonal. A similar development that are very much parallel to those in [9] and in this paper can be worked out. A major change will be to redefine the subspace $\mathscr{N}(\mathcal{A})$ in (2.3) as

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

We omit the detail.

Declaration of Competing Interest

No competing interest.

Acknowledgements

We wish to thank the anonymous referees for their careful reading and suggestions that improve our presentation significantly. The simple proof of the second inequality in (3.15) is taken from one of the referees' report to replace our earlier proof that read more complicated. Proposition 3.2 is also due to the referee and it improves our earlier version.

References

[1] J.-F. Cardoso, Multidimensional independent component analysis, in: Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, 1998, vol. 4, IEEE, Washington, DC, 1998, pp. 1941–1944.
[2] L. De Lathauwer, B. De Moor, J. Vandewalle, Fetal electrocardiogram extraction by blind source subspace separation, IEEE Trans. Biomed. Eng. 47 (5) (2000) 567–572.
[3] F.J. Theis, Blind signal separation into groups of dependent signals using joint block diagonalization, in: 2005 IEEE International Symposium on Circuits and Systems, ISCAS 2005, IEEE, 2005, pp. 5878–5881.
[4] F.J. Theis, Towards a general independent subspace analysis, in: Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 2006, pp. 1361–1368.
[5] Y. Bai, E. de Klerk, D. Pasechnik, R. Sotirov, Exploiting group symmetry in truss topology optimization, Optim. Eng. 10 (3) (2009) 331–349.
[6] E. De Klerk, D.V. Pasechnik, A. Schrijver, Reduction of symmetric semidefinite programs using the regular $*$-representation, Math. Program. 109 (2–3) (2007) 613–624.
[7] E. De Klerk, R. Sotirov, Exploiting group symmetry in semidefinite programming relaxations of the quadratic assignment problem, Math. Program. 122 (2) (2010) 225–246.
[8] K. Gatermann, P.A. Parrilo, Symmetry groups, semidefinite programs, and sums of squares, J. Pure Appl. Algebra 192 (1) (2004) 95–128.
[9] Y. Cai, C. Liu, An algebraic approach to nonorthogonal general joint block diagonalization, SIAM J. Matrix Anal. Appl. 38 (1) (2017) 50–71.
[10] G. Chabriel, M. Kleinsteuber, E. Moreau, H. Shen, P. Tichavsky, A. Yeredor, Joint matrices decompositions and blind source separation: a survey of methods, identification, and applications, IEEE Signal Process. Mag. 31 (3) (2014) 34–43.
[11] L. De Lathauwer, A survey of tensor methods, in: 2009 IEEE International Symposium on Circuits and Systems, IEEE, 2009, pp. 2773–2776.
[12] P. Tichavský, A.-H. Phan, A. Cichocki, Non-orthogonal tensor diagonalization, Signal Process. 138 (2017) 313–320.
[13] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, L. De Lathauwer, Tensorlab 3.0, available online at http://www.tensorlab.net, Mar. 2016.
[14] J.W. Demmel, Applied Numerical Linear Algebra, SIAM, Philadelphia, PA, 1997.
[15] B. Póczos, A. Lőrincz, Independent subspace analysis using k-nearest neighborhood distances, in: Artificial Neural Networks: Formal Models and Their Applications-ICANN 2005, Springer, 2005, pp. 163–168.
[16] F.G. Russo, A note on joint diagonalizers of shear matrices, Sci. Asia 38 (2012) 401–407.
[17] B. Afsari, Sensitivity analysis for the problem of matrix joint diagonalization, SIAM J. Matrix Anal. Appl. 30 (3) (2008) 1148–1171.
[18] D. Shi, Y. Cai, S. Xu, Some perturbation results for a normalized non-orthogonal joint diagonalization problem, Linear Algebra Appl. 484 (2015) 457–476.
[19] 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) (2008) 1022–1032.

[20] L. De Lathauwer, Decompositions of a higher-order tensor in block terms-part II: definitions and uniqueness, SIAM J. Matrix Anal. Appl. 30 (3) (2008) 1033–1066.
[21] L. De Lathauwer, D. Nion, Decompositions of a higher-order tensor in block terms-part III: alternating least squares algorithms, SIAM J. Matrix Anal. Appl. 30 (3) (2008) 1067–1083.
[22] D. Nion, A tensor framework for nonunitary joint block diagonalization, IEEE Trans. Signal Process. 59 (10) (2011) 4585–4594.
[23] I. Domanov, 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) (2013) 855–875.
[24] I. Domanov, 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) (2013) 876–903.
[25] J.B. Kruskal, Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics, Linear Algebra Appl. 18 (2) (1977) 95–138.
[26] M. Sørensen, L. De Lathauwer, New uniqueness conditions for the canonical polyadic decomposition of third-order tensors, SIAM J. Matrix Anal. Appl. 36 (4) (2015) 1381–1403.
[27] M. Sørensen, L. De Lathauwer, Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-$(L_{r,n}, L_{r,n}, 1)$ terms–part I: uniqueness, SIAM J. Matrix Anal. Appl. 36 (2) (2015) 496–522.
[28] A. Stegeman, On uniqueness of the canonical tensor decomposition with some form of symmetry, SIAM J. Matrix Anal. Appl. 32 (2) (2011) 561–583.
[29] M. Yang, On partial and generic uniqueness of block term tensor decompositions, Ann. Univ. Ferrara 60 (2) (2014) 465–493.
[30] M. Yang, W. Li, M. Xiao, On identifiability of higher order block term tensor decompositions of rank $l_r\otimes$ rank-1, Linear Multilinear Algebra (2018) 1–23.
[31] P. Breiding, N. Vannieuwenhoven, The condition number of join decompositions, SIAM J. Matrix Anal. Appl. 39 (1) (2018) 287–309.
[32] N. Vannieuwenhoven, Condition numbers for the tensor rank decomposition, Linear Algebra Appl. 535 (2017) 35–86.
[33] R.-C. Li, Matrix perturbation theory, in: L. Hogben, R. Brualdi, G.W. Stewart (Eds.), Handbook of Linear Algebra, 2nd edition, CRC Press, Boca Raton, FL, 2014, Ch. 21.
[34] C.F. Van Loan, The ubiquitous Kronecker product, J. Comput. Appl. Math. 123 (1–2) (2000) 85–100.
[35] C.F. Van Loan, G.H. Golub, Matrix Computations, 4th edition, Johns Hopkins University Press, Baltimore, MD, 2012.
[36] G.W. Stewart, J.-G. Sun, Matrix Perturbation Theory, Academic Press, Boston, 1990.
[37] W. Kahan, Spectra of nearly Hermitian matrices, Proc. Amer. Math. Soc. 48 (1) (1975) 11–17.
[38] J.-G. Sun, On the variation of the spectrum of a normal matrix, Linear Algebra Appl. 246 (1996) 215–223.