Identification of Matrix Joint Block Diagonalization

Yunfeng Cai, Ping Li

Cognitive Computing Lab
Baidu Research
No.10 Xibeiwang East Road, Beijing 100193, China
10900 NE 8th St. Bellevue, Washington 98004, USA
{caiyunfeng, liping11}@baidu.com

Abstract

Given a set $\mathcal{C} = \{C_i\}_{i=1}^m$ of square matrices, the matrix blind joint block diagonalization problem (BJBDP) is to find a full column rank matrix $A$ such that $C_i = A\Sigma_i A^\top$ for all $i$, where $\Sigma_i$’s are all block diagonal matrices with as many diagonal blocks as possible. The BJBDP plays an important role in independent subspace analysis (ISA). This paper considers the identification problem for BJBDP, that is, under what conditions and by what means, we can identify the diagonalizer $A$ and the block diagonal structure of $\Sigma_i$, especially when there is noise in $C_i$’s. In this paper, we propose a “bi-block diagonalization” method to solve BJBDP, and establish sufficient conditions under which the method is able to accomplish the task. Numerical simulations validate our theoretical results. To the best of the authors’ knowledge, existing numerical methods for BJBDP have no theoretical guarantees for the identification of the exact solution, whereas our method does.

1 Introduction

The matrix joint block diagonalization problem (JBDP) is a particular block term decomposition of a third order tensor [18, 23]. Over the past two decades, it has become a fundamental tool in independent subspace analysis (ISA) (e.g., [7, 29]). ISA has found many applications in machine learning tasks, e.g., subspace clustering [33, 27, 32], face recognition/verification [21, 20, 19, 4], learning of disentangled representations [2, 26], etc. In this paper, we consider the identification problem for a blind JBDP (BJBDP), then show how the problem arises in ISA.

1.1 Problem Statement

To introduce the identification problem of BJBDP, we need the following definitions.

**Definition 1.** We call $\tau_p = (p_1, \ldots, p_\ell)$ a partition of positive integer $p$ if $p_1, \ldots, p_\ell$ are all positive integers and $\sum_{i=1}^\ell p_i = p$. The integer $\ell$ is called the cardinality of the partition $\tau_p$, denoted by $\ell = \text{card}(\tau_p)$. Two partitions $\tau_p = (p_1, \ldots, p_\ell)$, $\tilde{\tau}_p = (\tilde{p}_1, \ldots, \tilde{p}_\ell)$ are said to be equivalent, denoted by $\tau_p \sim \tilde{\tau}_p$, if $\ell = \tilde{\ell}$ and there exists a permutation $\Pi_\ell$ such that $\tau_p = \tilde{\tau}_p \Pi_\ell$.

For example, $\tau_p = \{3, 1, 5, 2\}$, $\ell = \text{card}(\tau_p) = 4$, $p = 11$, and $\tilde{\tau}_p = \{1, 5, 2, 3\}$ is equivalent to $\tau_p$.

**Definition 2.** Given a partition $\tau_p = (p_1, \ldots, p_\ell)$ and a matrix $X \in \mathbb{R}^{p \times p}$, partition $X$ as $X = [X_{ij}]$ with $X_{ij} \in \mathbb{R}^{p_i \times p_j}$. Define the $\tau_p$-block diagonal part and $\tau_p$-off-block diagonal part of $X$, respectively, as

\[
\text{BlkDiag}_{\tau_p}(X) \triangleq \text{diag}(X_{11}, X_{22}, \ldots, X_{\ell\ell}),
\]

\[
\text{OffBlkdiag}_{\tau_p}(X) \triangleq X - \text{BlkDiag}_{\tau_p}(X).
\]

The matrix $X$ is referred to as a $\tau_p$-block diagonal matrix if $\text{OffBlkdiag}_{\tau_p}(X) = 0$. 

The Joint Block Diagonalization Problem (JBDP) Given a matrix set \( C = \{C_i\}_{i=1}^m \) with \( C_i \in \mathbb{R}^{d \times d} \) for \( 1 \leq i \leq m \). The JBDP for \( C \) with respect to a partition \( \tau_p \) is to find a full column rank matrix \( A = A(\tau_p) \in \mathbb{R}^{d \times p} \) such that all \( C_i \)'s can be factorized as

\[
C_i = A\Sigma_iA^\top = A\text{diag}(\Sigma_i^{(1)}, \ldots, \Sigma_i^{(\ell)})A^\top, \quad \forall i,
\]

where \( \Sigma_i \)'s are all \( \tau_p \)-block diagonal. When (1) holds, we say that \( C \) is \( \tau_p \)-block diagonalizable and \( A \) is a \( \tau_p \)-block diagonalizer of \( C \).

The Blind JBDP (BJBDP) Given a matrix set \( C = \{C_i\}_{i=1}^m \) with \( C_i \in \mathbb{R}^{d \times d} \) for \( 1 \leq i \leq m \). The BJBDP for \( C \) is to find a partition \( \tau_p \) and a full column rank matrix \( A = A(\tau_p) \) such that \( C \) is \( \tau_p \)-block diagonalizable and \( \text{card}(\tau_p) \) is maximized. A solution to the BJBDP is denoted by \( (\tau_p, A) \).

Uniqueness of BJBDP If \( (\tau_p, A) \) with \( \tau_p = (p_1, \ldots, p_\ell) \) and \( A \in \mathbb{R}^{d \times p} \) is a solution to BJBDP, then \( (\hat{\tau}_p, \hat{A}) = (\tau_p\Pi_\ell, AD\Pi) \) is also a solution, where \( \Pi_\ell \in \mathbb{R}^{\ell \times \ell} \) is a permutation matrix, \( D \) is any nonsingular \( \tau_p \)-block diagonal matrix, \( \Pi \in \mathbb{R}^{p \times p} \) is a permutation matrix associated with \( \Pi_\ell \), which permutes the column blocks of \( A \) as \( \Pi_\ell \) permutes \( \tau_p \). In fact, \( \Pi \) can be obtained by replacing the 1 and 0 elements in the \( j \)-th column of \( \Pi_\ell \) by \( I_{p_j} \) and zero matrices of right sizes, respectively. If \( (\hat{\tau}_p, \hat{A}) = (\tau_p\Pi_\ell, AD\Pi) \), we say that \( (\tau_p, A) \) and \( (\hat{\tau}_p, \hat{A}) \) are equivalent, denoted by \( (\tau_p, A) \sim (\hat{\tau}_p, \hat{A}) \). If any two solutions to BJBDP are equivalent, we say that the solution to the BJBDP is unique, the BJBDP for \( C \) is uniquely \( \tau_p \)-block-diagonalizable.

Identifiability of BJBDP Let \( (\tau_p, A) \) be a solution to the BJBDP for \( C \). Let \( \tilde{C} = \{\tilde{C}_i\}_{i=1}^m = \{C_i + E_i\}_{i=1}^m \), where \( E_i \in \mathbb{R}^{d \times d} \) is a perturbation to \( C_i \) for \( 1 \leq i \leq m \). Under what conditions, and by what means, we can find a \( (\tilde{\tau}_p, \tilde{A}) \) such that

\[
\tilde{C}_i \approx \tilde{A}\Sigma_i\tilde{A}^\top = \tilde{A}\text{diag}(\tilde{\Sigma}_i^{(1)}, \ldots, \tilde{\Sigma}_i^{(\ell)})\tilde{A}^\top, \quad \forall i,
\]

where \( \tilde{\Sigma}_i \)'s are all \( \tilde{\tau}_p \)-block diagonal matrices with \( \tilde{\tau}_p \sim \tau_p \), and \( \tilde{A} \) is close to \( A \) (up to block permutation and block diagonal scaling).

1.2 ISA: A Case Study

Independent Subspace Analysis (ISA) aims at separating linearly mixed unknown sources into statistically independent groups of signals. A basic model can be stated as

\[
x = As,
\]

where \( x \in \mathbb{R}^d \) is the observed mixture, \( A \in \mathbb{R}^{d \times p} \) is the unknown mixing matrix and has full column rank, \( s \in \mathbb{R}^p \) is the source signal vector. Let \( s = [s_1^\top, \ldots, s_\ell^\top]^\top \) with \( s_j \in \mathbb{R}^{p_j} \) for \( j = 1, \ldots, \ell \). Assume that each \( s_j \) has mean 0 and contains no lower-dimensional independent component, all \( s_j \) are independent of each other. ISA attempts to recover \( s \) from \( x \). Obviously, it holds that

\[
C_{xx} = E(xx^\top) = A\text{E}(ss^\top)A^\top = AC_{ss}A^\top,
\]

where \( E(\cdot) \) stands for expectation, and \( C_{xx}, C_{ss} \) are the covariance matrices of \( x \) and \( s \), respectively. By assumption, \( C_{ss} = \text{diag}(C_{s_1s_1}, \ldots, C_{s_\ell s_\ell}) \) is \( \tau_p \)-block diagonal, where \( C_{sjsj} \) is the covariance matrix of \( s_j \).
Now let \( x(a), \ldots, x(T) \) be \( T \) samples. In a piecewise stationary model \([16, 17]\), the samples are partitioned into \( m \) non-overlapping domains \( \{ T_i \}_{i=1}^m \), where \( T_i \) contains \( t_i \) samples, and \( \sum_i t_i = T \). Let \( \tilde{C}_i \triangleq \frac{1}{t_i} \sum_{t \in T_i} x(t)x(t)^\top \) and \( \tilde{C} = \{ \tilde{C}_i \}_{i=1}^m \). Ideally, \( A \) is a \( \tau_p \)-block diagonalizer of \( \tilde{C} \). The question is that whether we can find \((\tilde{\tau}_p, \tilde{A})\) by solving the BJD for \( \tilde{C} \) such that \( \tilde{\tau}_p \sim \tau_p \), and \( \tilde{A} \) is “close” to \( A \)? Under what conditions? And how?

1.3 A Short Review and Our Contribution

The identification problem is closely related to the uniqueness of the problem. In the context of ISA, it is shown that the decomposition of a random vector with existing covariance into independent, irreducible components is unique up to order and invertible transformations within the components (referred to as “trivial indeterminacy” hereafter) and an invertible transformation in possibly higher dimensional Gaussian component \([12, 13]\). In the context of JBDP, when the matrices have additional structure, a local indeterminacy may occur \([11, 13]\). As JBDP is a particular block term decomposition of a third-order tensor, solution to JBDP is unique up to trivial determinacy almost surely \([18]\).

Algorithmically, JBDP is usually formulated as an optimization problem, then solved via optimization-based numerical methods (e.g., \([23, 9]\)). However, without the information of the block diagonal structure, it is difficult to formulate the cost function. As a result, for BJD, a two-stage procedure is proposed – first apply a joint diagonalization method (e.g., \([8, 34]\)), then reveal the block diagonal structure by certain clustering method (e.g., \([30]\)). However, such a procedure is based on a conjecture \([1]\) that the JD and JBD problems share the same minima. But this conjecture is only partially proved \([29]\). Three algebraic methods are proposed to solve BJD: When the diagonalizer is orthogonal, using matrix \( * \)-algebra, an error controlled method is proposed in \([22]\). Then the results are non-trivially generalized to the non-orthogonal diagonalizer case in \([6]\); Using the matrix polynomial, a three-stage method is proposed in \([5]\).

To the best of the authors’ knowledge, current numerical methods for BJD have no theoretical guarantees for a good identification of the exact solution \((\tau_p, A)\), i.e., for the computed solution \((\tilde{\tau}_p, \tilde{A})\), \( \tilde{\tau}_p \) is equivalent to \( \tau_p \) and \( \tilde{A} \) is “close” to \( A \). In this paper, we will answer this fundamental question. For both noiseless and noisy cases, we first find the range space the diagonalizer via a (truncated) singular value decomposition (SVD), then reveal the block diagonal structure by a bi-diagonalization procedure. Under proper assumptions, we show that the proposed method is able to identify \((\tau_p, A)\). Numerical simulations validate our theoretical results.

The rest of this paper is organized as follows. In Section 2, we establish the identification condition of the range space of \( A \) and the block diagonal structure for both noiseless and noisy cases. Numerical experiments are presented in Section 3. Concluding remarks are given in Section 4.

Notation. \( I_n \) 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 vec\((X)\) transforms 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. The spectral norm and Frobenius norm of a matrix are denoted by \( \| \cdot \|_2 \) and \( \| \cdot \|_F \), respectively. For a matrix \( X \), \( \mathcal{R}(X) \) and \( \mathcal{N}(X) \) stand for the range space and null space of \( X \), respectively. For any square matrix set \( D = \{ D_i \}_{i=1}^m \), we denote \( \tilde{D} = [D_1^\top, D_1, \ldots, D_m^\top, D_m]^\top \). For a subspace \( \mathcal{V} \) of \( \mathbb{R}^n \), its orthogonal complement is defined as \( \mathcal{V}^\perp = \{ w \in \mathbb{R}^n \mid w^\top v = 0, \forall v \in \mathcal{V} \} \).

2 Main Results

In this section, we establish the identification conditions for BJD. First, we identify \( \mathcal{R}(A) \) in Section 2.1, then the block diagonal structure in Section 2.2.

2.1 Identification of \( \mathcal{R}(A) \)

The following theorem identifies \( \mathcal{R}(A) \) for the noiseless case, i.e., \( E_i = 0 \) for all \( 1 \leq i \leq m \).

**Theorem 2.1.** Let \((\tau_p, A)\) be a solution to BJD for \( C \). Then \( \mathcal{R}(A) = \mathcal{N}(C) = \mathcal{R}(C^\top) \).
By Theorem 2.1, it is natural for us to approximate of $R(A)$ by the subspace spanned by the first $p$ right singular vectors of $\tilde{C}$. The so called canonical angle is needed to state the result.

**Canonical Angles between Two Subspaces** Let $\mathcal{X}, \mathcal{Y}$ be $k$ and $\ell$ dimensional subspaces of $\mathbb{R}^n$, respectively, and $k \geq \ell$. Let $X \in \mathbb{R}^{n \times k}, Y \in \mathbb{R}^{n \times \ell}$ be the orthonormal basis matrices of $\mathcal{X}$ and $\mathcal{Y}$, respectively. Denote the singular values of $Y^\top X$ by $\omega_1, \ldots, \omega_{\ell}$, and they are in a non-decreasing order, i.e., $\omega_1 \leq \cdots \leq \omega_{\ell}$. The canonical angles $\theta_j(\mathcal{X}, \mathcal{Y})$ between $\mathcal{X}$ and $\mathcal{Y}$ are defined by

$$0 \leq \theta_j(\mathcal{X}, \mathcal{Y}) \triangleq \arccos \omega_j \leq \frac{\pi}{2}, \quad \text{for } 1 \leq j \leq \ell.$$  

They are in a non-increasing order, i.e., $\theta_1(\mathcal{X}, \mathcal{Y}) \geq \cdots \geq \theta_{\ell}(\mathcal{X}, \mathcal{Y})$. Set

$$\Theta(\mathcal{X}, \mathcal{Y}) \triangleq \text{diag}(\theta_1(\mathcal{X}, \mathcal{Y}), \ldots, \theta_{\ell}(\mathcal{X}, \mathcal{Y})).$$

It is worth mentioning here that the canonical angles defined above are independent of the choices of the orthonormal basis matrices $X$ and $Y$.

**Theorem 2.2.** Let $(\tau_p, A)$ be a solution to BJBBDP for $C$. Let the columns of $V_2$ be an orthonormal basis for $\mathcal{N}(A^\top)$, $\phi_1 \geq \cdots \geq \phi_d$ and $\tilde{\phi}_1 \geq \cdots \geq \tilde{\phi}_d$ be the singular values of $C$ and $\tilde{C}$, respectively. Then

$$\tilde{\phi}_p \geq \phi_p - \|E\|, \quad \tilde{\phi}_{p+1} \leq \|E\|. \quad (2)$$

In addition, let $\tilde{U}_1 = [\tilde{u}_1, \ldots, \tilde{u}_p], \tilde{V}_1 = [\tilde{v}_1, \ldots, \tilde{v}_p]$, where $\tilde{u}_j, \tilde{v}_j$ are the left and right singular vector of $\tilde{C}$ corresponding to $\tilde{\phi}_j$, respectively, and $\tilde{U}_1, \tilde{V}_1$ are both orthonormal. If $\|E\| < \frac{\phi_p}{2}$, then

$$\| \sin \Theta(\mathcal{R}(A), \mathcal{R}(\tilde{V}_1)) \| \leq \frac{\| \tilde{U}_1^\top E V_2 \|}{\tilde{\phi}_p}.$$  

By Theorem 2.2, when $\|E\|$ is sufficiently small compared with $\phi_p$, we are able to find the correct $p$, and $\mathcal{R}(\tilde{V}_1)$ is a good approximation for $\mathcal{R}(A)$.

**2.2 Identification of the Block Diagonal Structure**

In this section, we first discuss the identification of the block diagonal structure for the noiseless case, then the noisy case.

**2.2.1 The Noiseless Case**

This section is organized as follows:

(a) Firstly, we present a necessary and sufficient condition for when $C_i$’s can be factorized in the form (1);

(b) Secondly, we present a way to determine whether the solution to the BJBBDP is unique;

(c) Finally, we show how to find a solution to the BJBBDP, and establish the theoretical guarantee.

**Remark 1.** The results for (a) and (b) are given below by Theorems 2.3 and 2.4, respectively. We need to emphasize here that Theorem 2.3 is rewritten from [6, Lemma 2.3], and Theorem 2.4 is partially rewritten from [6, Theorem 2.5]. The difference between Theorem 2.3 and Lemma 2.3 is that the diagonalizer here is rectangular rather than square. The main difference between Theorem 2.4 and Theorem 2.5 is the proof. The proof here is simpler, more importantly, the proof is constructive and explainable. Borrowing those two results from [6] should not undermine the contribution of this paper, since they are the start point for our main contribution – the algorithms (Algorithms 2 and 4) to identify the solution of BJBBDP with theoretical guarantees (Theorems 2.6 and 2.8).
The following linear space will play an important role in the analysis.

**Definition 3.** Given a matrix set \( \mathcal{D} = \{ D_i \}_{i=1}^m \) with \( D_i \in \mathbb{R}^{q \times q} \), define
\[
\mathcal{N}(\mathcal{D}) \triangleq \{ X \in \mathbb{R}^{q \times q} \mid D_i X - X^T D_i = 0, \ 1 \leq i \leq m \}.
\]

Now we present a necessary and sufficient condition for when \( C_i \)'s can be factorized in the form (1).

**Theorem 2.3.** Given \( C = \{ C_i \}_{i=1}^m \) with \( C_i \in \mathbb{R}^{d \times d} \). Let \( V_1 \in \mathbb{R}^{d \times p} \) be such that \( V_1^T V_1 = I_p \), \( \mathcal{R}(V_1) = \mathcal{R}(C^\top) \). Denote \( \mathcal{B} = \{ B_i \}_{i=1}^m \). Then \( C_i \)'s can be factorized as in (1) with \( \mathcal{R}(A) = \mathcal{R}(C^\top) \) if and only if there exists a matrix \( X \in \mathcal{N}(\mathcal{B}) \), which can be factorized into
\[
X = Y \text{diag}(X_{11}, \ldots, X_{\ell\ell}) Y^{-1},
\]
where \( Y \in \mathbb{R}^{p \times p} \) is nonsingular, \( X_{jj} \in \mathbb{R}^{p_j \times p_j} \) for \( 1 \leq j \leq \ell \) and \( \lambda(X_{jj}) \cap \lambda(X_{kk}) = \emptyset \) for \( j \neq k \).

According to Theorem 2.3, once we find an \( X \in \mathcal{N}(\mathcal{B}) \) which has a factorization in form (3), we can find a \((\tau_p, A)\) satisfying (1). Next, we examine some fundamental properties of \( \Gamma \in \mathcal{N}(\{ \Sigma_i \}) \) with \( \Sigma_i = \text{diag}(\Sigma_i^{(1)}, \ldots, \Sigma_i^{(1)}) \), based on which we can determine whether \((\tau_p, A)\) is a solution to the BJBDP for \( \mathcal{C} \).

Partition \( \Gamma \) as \( \Gamma = [\Gamma_{jk}] \), where \( \Gamma_{jk} \in \mathbb{R}^{p_j \times p_k} \). Using \( \Sigma_i \Gamma - \Gamma \Sigma_i = 0 \), we have two sets of matrix equations. The first set is for \( 1 \leq j = k \leq \ell \):
\[
\Sigma_i^{(j)} \Gamma_{jj} - \Gamma_{jj}^\top \Sigma_i^{(j)} = 0, \quad \text{for } 1 \leq i \leq m; \quad (4a)
\]

The second set is for \( 1 \leq j < k \leq \ell \):
\[
\begin{align*}
\Sigma_i^{(j)} \Gamma_{jk} - \Gamma_{kj}^\top \Sigma_i^{(k)} &= 0, \\
\Sigma_i^{(k)} \Gamma_{kj} - \Gamma_{jk}^\top \Sigma_i^{(j)} &= 0,
\end{align*} \quad \text{for } 1 \leq i \leq m. \quad (4b)
\]

With the help of the Kronecker product, the first set of equations are equivalent to

\[
G_{jj} \text{vec}(\Gamma_{jj}) = 0, \quad (5a)
\]

where
\[
G_{jj} = \begin{bmatrix}
I_{p_j} \otimes \Sigma_i^{(j)} - [\Sigma_i^{(j)} \otimes I_{p_j}] \Sigma_i^{(j)}
& \cdots \\
& I_{p_j} \otimes \Sigma_i^{(j)} - [\Sigma_i^{(j)} \otimes I_{p_j}] \Sigma_i^{(j)}
\end{bmatrix},
\]

\( \Pi_j \in \mathbb{R}^{p_j \times p_j} \) is the perfect shuffle permutation matrix [31, Subsection 1.2.11] that enables \( \Pi_j \text{vec}(Z_{jj}^\top) = \text{vec}(Z_{jj}) \). The second set of equations are equivalent to

\[
G_{jk} \begin{bmatrix}
\text{vec}(\Gamma_{jk}) \\
-\text{vec}(\Gamma_{kj})
\end{bmatrix} = 0, \quad (5b)
\]

where
\[
G_{jk} = \begin{bmatrix}
I_{p_k} \otimes \Sigma_i^{(j)} & (\Sigma_i^{(k)})^\top \otimes I_{p_j} \\
I_{p_k} \otimes (\Sigma_i^{(j)})^\top & \Sigma_i^{(k)} \otimes I_{p_j}
& \cdots \\
& I_{p_k} \otimes (\Sigma_i^{(j)})^\top & \Sigma_i^{(k)} \otimes I_{p_j}
\end{bmatrix}.
\]

For \( G_{jj} \) and \( G_{jk} \), we introduce the following two properties:
(P1) For $1 \leq j \leq \ell$, for any $\text{vec}(\Gamma_{jj}) \in \mathcal{N}(G_{jj})$, the eigenvalues of $\Gamma_{jj}$ are the same real number or the same complex conjugate pair.

(P2) For $1 \leq j < k \leq \ell$, $G_{jk}$ has full column rank.

The uniqueness of the solution to the BJBDP is closely related to (P1) and (P2). In fact, we have the following theorem.

**Theorem 2.4.** Let $A$ be a $\tau_p$-block diagonalizer of $C$ i.e., (1) holds. Then $(\tau_p, A)$ is the unique solution to the BJBDP for $C$ if and only if both (P1) and (P2) hold.

Several important remarks follow in order.

**Remark 2.** Based on Theorem 2.4, once we get a $\tau_p$-block diagonalizer $A$ that factorizes $C_i$ as (1), we can determine whether $(\tau_p, A)$ is the unique solution to the BJBDP by checking (P1) and (P2).

**Remark 3.** By the proof of Theorem 2.4, we have the following facts to help the understanding of (P1) and (P2).

1) If (P1) does not hold for some $j$, then $\{\Gamma_{jj}^\alpha\}_{j=1}^m$ can be further block diagonalized. This is because if (P1) does not hold for some $j$, there exists $\Gamma_{jj} \in \mathbb{R}^{p_j \times p_j}$ such that $\text{vec}(\Gamma_{jj}) \in \mathcal{N}(G_{jj})$ and a nonsingular $W_j \in \mathbb{R}_j^{p_j \times p_j}$ such that

$$
\Gamma_{jj} = W_j \text{diag}(\Gamma_{jj}^\alpha, \Gamma_{jj}^\beta) W_j^{-1},
$$

where $\Gamma_{jj}^\alpha$ and $\Gamma_{jj}^\beta$ are two real matrices and $\lambda(\Gamma_{jj}^\alpha) \cap \lambda(\Gamma_{jj}^\beta) = \emptyset$. Using $\text{vec}(\Gamma_{jj}) \in \mathcal{N}(G_{jj})$, we have

$$
\Sigma_i^{(jj)} \Gamma_{jj} - \Gamma_{jj} \Sigma_i^{(jj)} = 0, \quad \text{for } 1 \leq i \leq m.
$$

Substituting (6) into the above equality, we get

$$
\Sigma_i^{(jj)} \text{diag}(\Gamma_{jj}^\alpha, \Gamma_{jj}^\beta) - \Sigma_i^{(jj)} \text{diag}(\Gamma_{jj}^\alpha, \Gamma_{jj}^\beta) \Sigma_i^{(jj)} = 0, \quad \text{for } 1 \leq i \leq m,
$$

where $\Sigma_i^{(jj)} = W_j^\top \Sigma_i^{(jj)} W_j$ for $i = 1, \ldots, m$. Partition $\Sigma_i^{(jj)}$ as $\Sigma_i^{(jj)} = \begin{bmatrix} \Sigma_i^{(j1)} & \Sigma_i^{(j2)} \\ \Sigma_i^{(j2)} \Sigma_i^{(j2)} \end{bmatrix}$. Then it follows that

$$
\begin{bmatrix} \Sigma_i^{(j1)} \\ \Sigma_i^{(j2)} \Sigma_i^{(j2)} \end{bmatrix} \Gamma_{jj}^\beta - \begin{bmatrix} \Gamma_{jj}^\alpha \Sigma_i^{(j1)} \\ \Sigma_i^{(j2)} \Sigma_i^{(j2)} \end{bmatrix} = 0, \quad \text{for } 1 \leq i \leq m.
$$

Using $\lambda(\Gamma_{jj}^\alpha) \cap \lambda(\Gamma_{jj}^\beta) = \emptyset$, we have $\Sigma_i^{(j12)} = 0$ and $\Sigma_i^{(j21)} = 0$. In other words, $\Sigma_i^{(jj)}$ for $1 \leq i \leq m$ can be further block diagonalized.

2) If (P2) does not hold for some $j \neq k$, then $\{\text{diag}(\Sigma_i^{(jj)}, \Sigma_i^{(kk)})\}_{i=1}^m$ has a diagonalizer that is not $(I_{p_j}, I_{p_k})$-block diagonal. For example, let $a_i$’s, $b_i$’s and $c_i$’s be arbitrary real numbers, it holds that

$$
\text{diag} \left( \begin{bmatrix} 0 & a_i \\ a_i & b_i \end{bmatrix}, \begin{bmatrix} 0 & a_i \\ a_i & c_i \end{bmatrix} \right) \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{diag} \left( \begin{bmatrix} 0 & a_i \\ a_i & b_i \end{bmatrix}, \begin{bmatrix} 0 & a_i \\ a_i & c_i \end{bmatrix} \right) \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^\top,
$$

in which the diagonalizer is not equivalent to $I_4$.

**Remark 4.** In the context of ISA, (P1) essentially requires the irreducibility [12, 13] of the independent component; (P2) generalizes the concept of “local indeterminacy/simple component” [11, 13], and is much more mathematically strict.

Next, we consider how to solve the BJBDP.
Given a set $\mathcal{D} = \{D_i\}_{i=1}^{m}$ of $q$-by-$q$ matrices with $\mathcal{D}$ having full column rank. When $\mathcal{D}$ has a $\tau_\mathcal{D} = (q_1, q_2)$-block diagonalizer $Z$, i.e., $D_i$'s can be factorized as $D_i = Z\Phi_i Z^\top$, where $\Phi_i$'s are $\tau_\mathcal{D}$-block diagonal, then set $X_\star = Z^{-\top} \text{diag}(I_{q_1}, -I_{q_2})Z^\top$ ($Z$ is nonsingular since $\mathcal{D}$ has full column rank), it holds that

$$X_\star \in \mathcal{N}(\mathcal{D}), (X_\star - I)(X_\star + I) = 0, X_\star \neq \pm I.$$  

Conversely, once we find such an $X_\star$, factorize $X_\star$ into $X_\star = Y \text{diag}(I_{q_1}, -I_{q_2})Y^{-1}$, then $Y^{-\top}$ is a $\tau_\mathcal{D}$-block diagonalizer. In what follows, we formulate the problem of finding such an $X_\star$ as a constrained optimization problem.

Note that

$$\begin{align*}
(X - I)(X + I) &= 0 \\
\Leftrightarrow \min_X \text{tr}((X - I)^2(X + I)^2) &\Leftrightarrow \min_X \text{tr}(X^4) - 2 \text{tr}(X^2) + q,
\end{align*}$$

and $\text{tr}(X) = 0$ together with $\text{tr}(X^2) = q$ ensure $X \neq 0$ and the eigenvalues of $X$ lie in both left and right complex plane, as a result, $X$ is not a scalar matrix. So, we propose to find $X_\star$ by solving the following optimization problem:

$$\begin{align*}
\text{opt}(\mathcal{D}) : &\quad \min_X \text{tr}(X^4), \\
\text{subject to} &\quad X \in \mathcal{N}(\mathcal{D}), \text{tr}(X) = 0, \text{tr}(X^2) = q.
\end{align*}$$

(7)

For $\text{opt}(\mathcal{D})$, we have the following result.

**Theorem 2.5.** Given a set $\mathcal{D} = \{D_i\}_{i=1}^{m}$ of $q$-by-$q$ matrices with $\mathcal{D}$ having full column rank.

(I) If $\mathcal{D}$ does not have a nontrivial diagonalizer, then the feasible set of $\text{opt}(\mathcal{D})$ is empty.

(II) If $\mathcal{D}$ has a nontrivial diagonalizer, then $\text{opt}(\mathcal{D})$ has a solution $X_\star$. In addition, assume

$$\mu = \min_{\|z\| = 1} \sqrt{\sum_{i=1}^{m} |z^H D_i z|^2} > 0,$$

then $X_\star$ has two distinct real eigenvalues, and the gap between them are no less than two.

**Remark 5.** If $\mathcal{D}$ has full column rank, then $\mu > 0$ almost surely. Therefore, (II) holds almost surely without the assumption $\mu > 0$.

Based on Theorem 2.5, we present Algorithm 1, which will find a $\tau_\mathcal{D}$-diagonalizer $Z$ for a matrix set $\mathcal{D} = \{D_i\}$ with $\text{card}(\tau_\mathcal{D}) = 2$ whenever $\mathcal{D}$ can be block-diagonalized.

**Algorithm 1** Bi-Block Diagonalization (Bi-BD)

1. **Input:** A matrix set $\mathcal{D} = \{D_i\}_{i=1}^{m}$ of $q$-by-$q$ matrices.

2. **Output:** $(\tau_\mathcal{D}, Z)$ such that $Z$ is a $\tau_\mathcal{D}$-block diagonalizer of $\mathcal{D}$ with $\tau_\mathcal{D} = (q_1, q_2)$ or $\tau_\mathcal{D} = (q)$.

3. if feasible set of $\text{opt}(\mathcal{D})$ is empty then set $\tau_\mathcal{D} = (q)$, $Z = I_q$;

4. else Solve $\text{opt}(\mathcal{D})$, denote the solution by $X_\star$;

5. Compute $X_\star = Y \text{diag}(\Gamma_1, \Gamma_2)Y^{-1}$, where $\Gamma_1 \in \mathbb{R}^{q_1 \times q_1}$, $\Gamma_2 \in \mathbb{R}^{q_2 \times q_2}$, both $\lambda(\Gamma_1)$ and $\lambda(\Gamma_2)$ contain only one real number, and the two real numbers are different.

6. Set $\tau_\mathcal{D} = (q_1, q_2)$, $Z = Y^{-\top}$.
Line 5 in Algorithm 1 can be computed via Algorithm 7.6.3 in [31]. The central task is to solve \( \text{OPT}(\mathcal{D}) \).
Using the Kronecker product, \( X \in \mathcal{M}(\mathcal{D}) \) if and only if \( \mathbf{L}(\mathcal{D})\text{vec}(X) = 0 \), where
\[
\mathbf{L}(\mathcal{D}) \triangleq \begin{bmatrix}
    I_q \otimes D_1 - D_1^\top \otimes I_q \Pi_q \\
    \vdots \\
    I_q \otimes D_m - D_m^\top \otimes I_q \Pi_q
\end{bmatrix} \in \mathbb{R}^{m_q \times q^2}.
\]
(8)

Here \( \Pi_q \in \mathbb{R}^{q\times q} \) is the perfect shuffle permutation. The restarted Lanczos bi-diagonalization method [3] (MATLAB script \texttt{svds}), which is usually used to compute a few smallest/largest singular values and the corresponding singular vectors of a large scale matrix, is well suited here, since only the right singular vectors corresponding with the smallest singular value zero are needed. From the right singular vectors corresponding to zero, we can construct an orthonormal basis \( \{X_1, \ldots, X_s\} \) for \( \mathcal{M}(\mathcal{D}) \), where \( s = \dim \mathcal{M}(\mathcal{D}) \).

Now let \( \mathcal{M} = [M_{ijkl}] \in \mathbb{R}^{s \times s \times s \times s} \), \( K = [K_{ij}] \in \mathbb{R}^{s \times s} \) with \( M_{ijkl} = \text{tr}(X_i X_j X_k X_l), K_{ij} = \text{tr}(X_i X_j) \), the optimization problem \( \text{OPT}(\mathcal{D}) \) is reduced into
\[
\min_{\alpha \in \mathbb{R}^s} M\alpha^4, \quad \text{subject to} \quad \alpha^\top K\alpha = 1,
\]
(9)

where \( M\alpha^4 \triangleq \sum_{i,j,k,l} M_{ijkl}\alpha_i\alpha_j\alpha_k\alpha_l \). Let \( K = G^\top G \) be the Cholesky factorization of \( K \) (by definition, \( K \) is symmetric positive definite), and denote \( \beta = G\alpha, \mathcal{N} = \mathcal{M} \times_1 G^{-\top} \times_2 G^{-\top} \times_3 G^{-\top} \times_4 G^{-\top} \), where \( \times_i \) denotes the modal product [31]. Then (9) can be rewritten as
\[
\min_{\beta \in \mathbb{R}^s} \mathcal{N}\beta^3, \quad \text{subject to} \quad \beta^\top \beta = 1,
\]
(10)

whose KKT condition is \( \mathcal{N}\beta^3 = \lambda \beta \), which is a Z-eigenvalue problem [24] of an order-4 tensor. Using the shifted power method [15, 10], the eigenvector \( \beta_* \) corresponding with the smallest eigenvalue can be computed.

Then \( X_* \) can obtained \( X_* = \sum_{j=1}^s \alpha_j X_j \), where \( \alpha = G^{-1}\beta_* \).

With the help of Algorithm 1, we may find a solution to BJBBDP recursively. We summarize the method in Algorithm 2.

**Algorithm 2 BJBBDP via BI-BD**

1: **Input:** A matrix set \( \mathcal{C} = \{C_i\}_{i=1}^m \) of \( d \)-by-\( d \) matrices.

2: **Output:** \( (\tilde{\tau}, \vec{A}) \), a solution to the BJBBDP of \( \mathcal{C} \).

3: Compute \( V_1 \), whose columns form an orthonormal basis for \( \mathcal{C}^\top \);

4: Compute \( \mathcal{B} = \{B_i\}_{i=1}^m = \{V_i^\top C V_i\}_{i=1}^m \);

5: Initialize \( \tilde{\tau} = (p), \vec{A} = V_1, \text{list} = [0] \);

6: **while** \( \exists 0 \) in list \**do**

7: Find \( t = \text{argmax}\{\tilde{\tau}(i) \mid \text{list}(i) = 0\} \);

8: Set \( k_1 = \sum_{i=1}^{t-1} \tilde{\tau}(i) + 1, k_2 = \sum_{i=1}^t \tilde{\tau}(i), D_i = B_i(k_1 : k_2, k_1 : k_2) \) and \( \mathcal{D} = \{D_i\} \);

9: Call Algorithm 1 with input \( \mathcal{D} \), denote the output by \( (\hat{\tau}, \hat{Z}) \);

10: **if** \( \text{card}(\hat{\tau}) = 1 \) **then** Update list(\( t \)) = 1;

11: **else** Update list and \( \tilde{\tau} \) by replacing their \( t \)-th entry by \([0, 0]\) and \( \hat{\tau} \), respectively;

12: Update \( B_i(k_1 : k_2, k_1 : k_2) = \hat{Z}^{-1} D_i \hat{Z}^{-\top}, \vec{A}(:,k_1 : k_2) = \hat{A}(,:,k_1 : k_2) \hat{Z} \).

13: **end if**

14: **end while**

Under proper assumptions, we can show that Algorithm 2 is able to identify the solution to BJBBDP.

**Theorem 2.6.** Assume that the BJBBDP for \( \mathcal{C} \) is uniquely \( \tau_p \)-block-diagonalizable, and let \( (\tau_p, A) \) be a solution satisfying (1). Then \( (\tau_p, A) \) can be identified via Algorithm 2, almost surely.
2.2.2 The Noisy Case

In this section, we discuss the identification of the block diagonal structure with the presence of noise. According to Theorem 2.2, a good approximation for $\mathcal{A}(A)$ can be obtained when the perturbation is small. Given a perturbed matrix set $\bar{D} = \{D_i + E_i\}_{i=1}^m$, where $D = \{D_i\}_{i=1}^m$ can be block diagonalized, $E_i$ is a perturbation to $D_i$. Inspired by the noiseless case, we consider an approximation of $\mathcal{A}(\bar{D})$ to approximately block-diagonalize $\bar{D}$. The subspace $\mathcal{A}(\bar{D})$ seems to be a natural choice, however, due to the presence of the noise, $\mathcal{A}(\bar{D})$ in general only has a trivial element – the scalar matrix, which is useless for matrix joint block diagonalization. Recall (8), let $\tilde{v}_1, \ldots, \tilde{v}_{\mu_2}$ be the right singular vectors of $L(\bar{D})$ corresponding to the singular values $\tilde{\sigma}_1, \ldots, \tilde{\sigma}_{\mu_2}$, respectively, and the singular values be in a non-decreasing order. We define

$$\mathcal{A}_\delta(\bar{D}) \triangleq \{\text{reshape}(v, q, q) \mid v \in \mathcal{A}([\tilde{v}_1, \ldots, \tilde{v}_k]), \tilde{\sigma}_k \leq \delta < \tilde{\sigma}_{k+1}\}.$$  

Note that if $\delta = 0$ and $E_i = 0$ for all $i$, then $\mathcal{A}_\delta(\bar{D}) = \mathcal{A}(\bar{D})$. Therefore, we may say that $\mathcal{A}_\delta(\bar{D})$ is a generalization of $\mathcal{A}(\bar{D})$. In what follows, we will let $\mathcal{A}_\delta(\bar{D})$ play the role of $\mathcal{A}(\bar{D})$. We also generalize the definition of diagonalizer as follows.

**Definition 4.** Given a set $\bar{D} = \{\bar{D}_i\}_{i=1}^m$ of $q$-by-$q$ matrices. We call $Z$ a $(\tau_q, \delta)$-diagonalizer (also referred to as $\delta$-diagonalizer when $\tau_q$ is clear from the context) of $\bar{D}$ if

$$\sum_{i=1}^m \|\bar{D}_i - Z\Phi_i Z^T\|_F^2 \leq C \delta^2,$$

where $\Phi_i$’s are all $\tau_q$-block diagonal matrices, and $C$ is a constant.

Rewrite the optimization problem $\text{OPT}(\bar{D})$ as

$$\text{OPT}(\bar{D}, \delta) : \min_X \text{tr}(X^4),$$

subject to $X \in \mathcal{A}_\delta(\bar{D}), \text{tr}(X) = 0, \text{tr}(X^2) = q$.

Then similar to Theorem 2.5, we have the next Theorem.

**Theorem 2.7.** Given a set $\bar{D} = \{\bar{D}_i\}_{i=1}^m$ of $q$-by-$q$ matrices with $\bar{D}$ having full column rank. Let $\delta = o(1)$ be a small real number.

(I) If $\bar{D}$ does not have a nontrivial $\delta$-diagonalizer, then the feasible set of $\text{OPT}(\bar{D}, \delta)$ is empty.

(II) If $\bar{D}$ has a nontrivial $\delta$-diagonalizer, then $\text{OPT}(\bar{D}, \delta)$ has a solution $X_*$. In addition, assume

$$\mu = \min_{\|z\|=1} \sqrt{\sum_{i=1}^m |z^H \bar{D}_i z|^2} = O(1),$$

and for $i = 1, 2$, let

$$\text{Rect}_i \triangleq \{z \in \mathbb{C} \mid \text{Re}(z) - \rho_i \leq a, |\text{Im}(z)| \leq b\},$$

where $a = O(\delta)$, $b = O(\delta)$. Then

$$\lambda(X_*) \subset \bigcup_{i=1}^2 \text{Rect}_i, \quad \rho_1 - \rho_2 \geq 2 + O(\delta).$$

Based on Theorem 2.7, we have Algorithms 3 and 4. Specifically, Algorithm 3 finds a $(\tau_q, \delta)$-diagonalizer $Z$ for a matrix set $\bar{D} = \{\bar{D}_i\}$ with $\text{card}(\tau_q) = 2$ whenever $\bar{D}$ can be approximately block-diagonalized; Algorithm 4 finds an approximate solution with the presence of noise.
Algorithm 3 Approximate Bi-Block Diagonalization (A-BI-BD)

1: **Input:** A matrix set $\mathcal{D} = \{\hat{D}_i\}^m_{i=1}$ of $q$-by-$q$ matrices, and a parameter $\delta$.
2: **Output:** $(\tau_q, Z)$ such that $Z$ is a $(\tau_q, \delta)$-block diagonalizer of $\hat{D}$ with $\tau_q = (q_1, q_2)$ or $\tau_q = (q)$.
3: if feasible set of $\text{OPT}(\hat{D}, \delta)$ is empty then set $\tau_q = (q)$, $Z = I_q$;
4: else Solve $X_*$;
5: Compute $X_* = Y \text{diag}(\Gamma_1, \Gamma_2)Y^{-1}$, where $\Gamma_1 \in \mathbb{R}^{q_1 \times q_1}$, $\Gamma_2 \in \mathbb{R}^{q_2 \times q_2}$, and the distance between $\lambda(\Gamma_1)$ and $\lambda(\Gamma_2)$ is approximately two.
6: Set $\tau_q = (q_1, q_2)$, $Z = Y^{-\top}$.
7: end if

Algorithm 4 BBJDP via A-BI-BD

1: **Input:** A matrix set $\mathcal{C} = \{\hat{C}_i\}^m_{i=1}$ of $d$-by-$d$ matrices.
2: **Output:** $(\hat{\tau}_p, \hat{A})$ such that $\hat{A}$ is a $(\hat{\tau}_p, \delta)$-block diagonalizer, where $\delta$ is a parameter.
3: Compute singular values $\tilde{\phi}_1 \geq \cdots \geq \tilde{\phi}_n$ and the corresponding right singular vectors $\tilde{v}_1, \ldots, \tilde{v}_n$ of $\hat{C}$, set $\tilde{V}_1 = [\tilde{v}_1, \ldots, \tilde{v}_p]$ with $\tilde{\phi}_{p+1} < \xi \tilde{\phi}_p$, where $\xi < 1$ is a real parameter, say $\xi = 0.1$;
4: Compute $\tilde{B} = \{\tilde{B}_i\}^m_{i=1} = [\tilde{V}_1^\top \tilde{C}_i \tilde{V}_1]_{i=1}^m$;
5: Initialize $\tilde{\tau}_p = (p)$, $\hat{A} = \tilde{V}_1$, list = $[0]$;
6: while $\exists 0$ in list do
7: Find $t = \arg\max\{\tilde{\tau}_p(i) | \text{list}(i) = 0\}$;
8: Set $k_1 = \sum_{i=1}^{t-1} \tilde{\tau}_p(i) + 1$, $k_2 = \sum_{i=1}^t \tilde{\tau}_p(i)$, $\tilde{D}_i = \tilde{B}_i(k_2 : k_1, k_1 : k_2)$ and $\tilde{D} = \{\tilde{D}_i\}$;
9: Call Algorithm 3 with input $\mathcal{D}$ and $\delta$, denote the output by $(\hat{\tau}, \hat{Z})$;
10: if $\text{card}(\hat{\tau}) = 1$ then Update list$t = 1$;
11: else Update list and $\hat{\tau}_p$ by replacing their $t$th entry by $[0, 0]$ and $\hat{\tau}$, respectively;
12: Update $B_i(k_1 : k_2, k_1 : k_2) = \hat{Z}^{-\top} D_i \hat{Z}^{-\top}$, $\hat{A}(; k_1 : k_2) = \hat{A}(; k_1 : k_2) \hat{Z}$.
13: end if
14: end while

Finally, we establish the identifiability for BBJDP with the presence of noise. The modulus of irreducibility and nonequivalence defined below are needed.

**Definition 5.** Let $(\tau_p, A)$ be a solution to BBJDP for $\mathcal{C}$ with $\text{BlkDiag}_{\tau_p}(A^\top A) = I_p$. Let $G_{jj}$, $G_{jk}$ be the same as in (5). The **modulus of irreducibility and nonequivalence** for $\mathcal{C}$ with respective to the diagonalizer $A$ are respectively defined as

$$
\omega_{ir} = \begin{cases} 
\infty, & \tau_p = (1, \ldots, 1), \\
\min\{|\sigma| : \sigma \in \sigma(G_{jj}), \sigma \neq 0\}, & \text{otherwise},
\end{cases}
$$

$$
\omega_{neq} = \min_{1 \leq j \leq k \leq l} \sigma_{\min}(G_{jk}).
$$

**Remark 6.** The moduli $\omega_{ir}$ and $\omega_{neq}$ depend on the choice of the diagonalizer $A$. When the solution to BBJDP for $\mathcal{C}$ is unique, we can show that their dependency on diagonalizer $A$ can be removed.

**Remark 7.** The modulus of irreducibility measures how far away the small blocks can be further block diagonalized; the modulus of nonequivalence measures how far away the BBJDP may have nonequivalent solutions.

The following theorem tells that when the noise is sufficiently small, $(\tau_p, A)$ can be identified.
Theorem 2.8. Assume that the BJBDP for $\mathcal{C} = \{C_i\}_{i=1}^m$ is uniquely $\tau_p$-block-diagonalizable, and let $(\tau_p, A)$ be a solution satisfying (1). Let $\tilde{\mathcal{C}} = \{\tilde{C}_i\}_{i=1}^m = \{C_i + E_i\}_{i=1}^m$ be a perturbed matrix set of $\mathcal{C}$. Denote
\[
\tau_p = (p_1, \ldots, p_\ell), \quad \hat{\tau}_p = (\hat{p}_1, \ldots, \hat{p}_\ell), \quad A = [A_1, \ldots, A_\ell], \quad \tilde{A} = [\tilde{A}_1, \ldots, \tilde{A}_\ell],
\]
where $(\hat{\tau}_p, \tilde{A})$ is the output of Algorithm 4. Assume $\mathcal{N}(G_{jj}) = R(\vec{I}_{p_j})$ for all $j$, where $G_{jj}$ is defined in (5a). Also assume that $p$ is correctly identified in Line 3 of Algorithm 4. Let the singular values of $\tilde{C}$ be the same as in Theorem 2.2,
\[
\epsilon = \frac{\|E\|}{\phi_p}, \quad r = \sqrt{2}(d + 2C) \frac{\hat{\phi}_p \epsilon}{\sigma_{\min}(A)(1 - \epsilon^2)}, \quad g_j = \frac{\sqrt{2j}}{(\ell - 1)\kappa \sqrt{p}} - \max\left\{\frac{\kappa}{\omega_{\inf}}, \frac{1}{\omega_{ir}}\right\} r, \quad \text{for } j = 1, 2,
\]
where $C$ and $\kappa$ are two constants.

(I) If $g_1 > 0$, then $\hat{\ell} = \ell$, and there exists a permutation $\{1', 2', \ldots, \ell'\}$ of $\{1, 2, \ldots, \ell\}$ such that $p_j = \hat{p}_{j'}$. In order words, $\hat{\tau}_p \sim \tau_p$.

(II) Further assume $g_2 > \frac{1}{\omega_{ir}}$, then there exists a $\tau_p$-block diagonal matrix $D$ such that
\[
\left\|\left[\tilde{A}_{1'}, \ldots, \tilde{A}_{\ell'}\right] - AD\right\|_F \leq \frac{cr}{g_2 - \frac{1}{\omega_{ir}}} \|A\|_F + \left(\frac{2}{\sqrt{1 - \epsilon^2}} + \epsilon\right)\|\tilde{A}\|_F = O(\epsilon),
\]
where $c$ is a constant.

### 3 Numerical Experiment

In this section, we present several numerical examples. All numerical tests are carried out using MATLAB. Our method (BI-BD) is compared with two BJBDP methods, namely, JBD-LM [9] and JBD-NCG [23], which are optimization based and need to know $\tau_p$ in advance.

**Example 1.** Given $\tau_p = (p_1, \ldots, p_\ell)$, we generate the matrix set $\tilde{\mathcal{C}} = \{\tilde{C}_i\}_{i=1}^m$ as follows:
\[
\tilde{C}_i = AD_iA^\top + N_i, \quad i = 1, \ldots, m,
\]
where $A \in \mathbb{R}^{n \times p}$, $D_i \in \mathbb{R}^{p \times p}$ is $\tau_p$-block diagonal and $N_i \in \mathbb{R}^{n \times n}$. The entries of $A$ and $D_i$ (block diagonal part) are drawn from $\mathcal{N}(0, 1)$, and the entries of $N_i$ from $\mathcal{N}(0, \sigma^2)$. The signal-to-noise ratio (SNR) is defined as $\text{SNR} = 10 \log_{10} 1/\sigma^2$. We carried out the tests with $m = 10$, $n = 15$, $p = 10$, $\tau_p = (2, 3, 3, 4)$, SNR = 40, 60, 80, 100. All tests are repeated 20 times, and the average results are reported in Figures 1 to 4.

![Figure 1: Singular values under different SNRs](image)

Figure 1 plots the smallest six singular values of $\tilde{C}$ for different SNRs, showing a big gap between the third and fourth singular values. The larger SNR is, the larger the gap is. Therefore, we can find the correct $p$. 

\[
\log_{10} \text{Singular value index}
\]

\[
\begin{array}{cccc}
\text{SNR=40} & \text{SNR=60} & \text{SNR=80} & \text{SNR=100} \\
\hline
\text{Singular value index} & 1 & 2 & 3 & 4 & 5 & 6
\end{array}
\]
Figure 2 plots the principle angle between $\mathcal{R}(A)$ and the range space spanned by the right singular vectors of $\hat{C}$ corresponding to the largest six singular vectors. We can see that $\mathcal{R}(A)$ is well estimated in all cases; the larger SNR is, the better the estimation is.

Figure 2: Principle angles under different SNRs

Figure 3 plots $\log(|\hat{A}^\top A|)$, where $\hat{A}$ is a diagonalizer obtained by BI-BD for SNR=40. We can see that the resulting matrix is approximately $\tau_p$ block diagonal up to permutation.

Figure 3: Structure of $|\hat{A}^\top A|$
Example 2. Consider three pieces of 3D independent sources. 6000 sample points were generated from noise free 3D wire-frames (as shown in the first row of Figure 5), then whitened. A random 9-by-9 matrix was used to mix the sources, and the mixed sources are shown in the second row of Figure 5. Our BI-BD method was applied to the mixed sources, and the recovered signals are shown in the last row of Figure 5. We can see that our method is able to recover the sources successfully.

![Figure 5: The original source signals, the mixed source signals, and the recovered signals](image)

4 Conclusion

In this paper, we studied the identification problem for matrix joint block diagonalization. We propose a numerical method called BI-BD to solve the problem, in which the block diagonal structure is revealed step by step via solving an optimization problem. Under the assumption that the solution is unique, we show that BI-BD is able to identify the true solution when the noise is sufficiently small. Two parameters, namely, the modulus of irreducibility (which measures how far away the small blocks can be further block diagonalized) and the modulus of nonequivalence (which measures how far away the BJBDP may have nonequivalent solutions), are introduced. According to Theorem 2.8, those two parameters determine the noise level that our BI-BD algorithm is able to identify the solution successfully. To the best of the authors' knowledge, our algorithm is the first method that has theoretical guarantees to find a good solution. Numerical simulations validate our theoretical results.
References

[1] Karim Abed-Meraim and Adel Belouchrani. Algorithms for joint block diagonalization. In Proceedings of the 2004 12th European Signal Processing Conference (EUSIPCO), pages 209–212, Vienna, Austria, 2004.

[2] Maren Awiszus, Hanno Ackermann, and Bodo Rosenhahn. Learning disentangled representations via independent subspaces. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshops, pages 560–568, Seoul, Korea (South), 2019.

[3] James Baglama and Lothar Reichel. Augmented implicitly restarted lanczos bidiagonalization methods. SIAM J. Sci. Comput., 27(1):19–42, 2005.

[4] Xin Yuan Cai, Chunheng Wang, Buihua Xiao, Xue Chen, and Ji Zhou. Deep nonlinear metric learning with independent subspace analysis for face verification. In Proceedings of the 20th ACM Multimedia Conference (MM), pages 749–752, Nara, Japan, 2012.

[5] Yunfeng Cai, Guanghui Cheng, and Decai Shi. Solving the general joint block diagonalization problem via linearly independent eigenvectors of a matrix polynomial. Numerical Linear Algebra with Applications, 26(4):e2238, 2019.

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

[7] Jean-François Cardoso. Multidimensional independent component analysis. In Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1941–1944, Seattle, WA, 1998.

[8] Jean-François Cardoso and Antoine Souloumiac. Blind beamforming for non-gaussian signals. In IEE proceedings F (radar and signal processing), volume 140, pages 362–370. IET, 1993.

[9] Omar Cherrak, Hicham Ghennioni, El Hossein Abarkan, and Nadège Thirion-Moreau. Non-unitary joint block diagonalization of matrices using a levenberg-marquardt algorithm. In Proceedings of the 21st European Signal Processing Conference (EUSIPCO), pages 1–5, Marrakech, Morocco, 2013.

[10] Stefano Cipolla, Michela Redivo-Zaglia, and Francesco Tudisco. Shifted and extrapolated power methods for tensor \( \ell^p \)-eigenpairs. arXiv preprint arXiv:1909.11964, 2019.

[11] Harold W. Gutch, Takanori Maehara, and Fabian J. Theis. Second order subspace analysis and simple decompositions. In Proceedings of the 9th International Conference on Latent Variable Analysis and Signal Separation (LVA / ICA), pages 370–377, St. Malo, France, 2010.

[12] Harold W. Gutch and Fabian J. Theis. Independent subspace analysis is unique, given irreducibility. In Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation (ICA), pages 49–56, London, UK, 2007.

[13] Harold W. Gutch and Fabian J. Theis. Uniqueness of linear factorizations into independent subspaces. J. Multivar. Anal., 112:48–62, 2012.

[14] W Kahan, BN Parlett, and Erxiong Jiang. Residual bounds on approximate eigensystems of nonnormal matrices. SIAM Journal on Numerical Analysis, 19(3):470–484, 1982.

[15] Tamara G. Kolda and Jackson R. Mayo. Shifted power method for computing tensor eigenpairs. SIAM J. Matrix Anal. Appl., 32(4):1095–1124, 2011.

[16] Dana Lahat, Jean-François Cardoso, and Hagit Messer. Second-order multidimensional ICA: performance analysis. IEEE Trans. Signal Process., 60(9):4598–4610, 2012.

[17] Dana Lahat, Jean-François Cardoso, and Hagit Messer. Blind separation of multi-dimensional components via subspace decomposition: Performance analysis. IEEE Trans. Signal Process., 62(11):2894–2905, 2014.
[18] Lieven De Lathauwer. Decompositions of a higher-order tensor in block terms - part II: definitions and uniqueness. *SIAM J. Matrix Anal. Appl.*, 30(3):1033–1066, 2008.

[19] Quoc V. Le, Will Y. Zou, Serena Y. Yeung, and Andrew Y. Ng. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In *Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3361–3368, Colorado Springs, CO, 2011.

[20] Stan Z. Li, Xiaoguang Lu, Xinwen Hou, Xianhua Peng, and Qiansheng Cheng. Learning multiview face subspaces and facial pose estimation using independent component analysis. *IEEE Trans. Image Process.*, 14(6):705–712, 2005.

[21] Stan Z. Li, Xiaoguang Lv, and Hongjiang Zhang. View-based clustering of object appearances based on independent subspace analysis. In *Proceedings of the Eighth International Conference On Computer Vision (ICCV)*, pages 295–300, Vancouver, Canada, 2001.

[22] Takanori Maehara and Kazuo Murota. Algorithm for error-controlled simultaneous block-diagonalization of matrices. *SIAM J. Matrix Anal. Appl.*, 32(2):605–620, 2011.

[23] Dimitri Nion. A tensor framework for nonunitary joint block diagonalization. *IEEE Trans. Signal Process.*, 59(10):4585–4594, 2011.

[24] Liqun Qi. Eigenvalues of a real supersymmetric tensor. *J. Symb. Comput.*, 40(6):1302–1324, 2005.

[25] Gilbert W. Stewart and Ji-Guang Sun. *Matrix Perturbation Theory*. Academic Press, Boston, 1990.

[26] Jan Stuehmer, Richard E. Turner, and Sebastian Nowozin. Independent subspace analysis for unsupervised learning of disentangled representations. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1200–1210, Online [Palermo, Sicily, Italy], 2020.

[27] Chun Chen Su, Zongze Wu, Ming Yin, Kaixin Li, and Weijun Sun. Subspace clustering via independent subspace analysis network. In *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP)*, pages 4217–4221, Beijing, China, 2017.

[28] Ji-guang Sun. On the variation of the spectrum of a normal matrix. *Linear algebra and its applications*, 246:215–223, 1996.

[29] Fabian J. Theis. Towards a general independent subspace analysis. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1361–1368, Vancouver, Canada, 2006.

[30] Petr Tichavský, Anh Huy Phan, and Andrzej Cichocki. Non-orthogonal tensor diagonalization. *Signal Process.*, 138:313–320, 2017.

[31] Charles F. Van Loan and Gene H. Golub. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 4th edition, 2012.

[32] Xing Wang, Jun Wang, Carlotta Domeniconi, Guoxian Yu, Guoqiang Xiao, and Maozu Guo. Multiple independent subspace clusterings. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI)*, pages 5353–5360, Honolulu, HI, 2019.

[33] Wei Ye, Samuel Maurus, Nina Hubig, and Claudia Plant. Generalized independent subspace clustering. In *Proceedings of the IEEE 16th International Conference on Data Mining (ICDM)*, pages 569–578, Barcelona, Spain, 2016.

[34] Andreas Ziehe, Pavel Laskov, Guido Nolte, and Klaus-Robert Müller. A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. *J. Mach. Learn. Res.*, 5:777–800, 2004.
Appendix

5 Preliminary

In this section, we present some preliminary results that will be used in subsequent proofs.

The following lemma is the well-known Weyl theorem (e.g., [25, p.203]).

**Lemma 5.1.** For two Hermitian matrices $A, \tilde{A} \in \mathbb{C}^{n \times n}$, let $\lambda_1 \leq \cdots \leq \lambda_n, \tilde{\lambda}_1 \leq \cdots \leq \tilde{\lambda}_n$ be eigenvalues of $A, \tilde{A}$, respectively. Then

$$|\lambda_j - \tilde{\lambda}_j| \leq \|A - \tilde{A}\|, \quad \text{for } 1 \leq j \leq n.$$ 

The following lemma gives some fundamental results for $\sin \Theta(U, V)$, which can be easily verified via definition.

**Lemma 5.2.** Let $[U, U_c]$ and $[V, V_c]$ be two orthogonal matrices with $U \in \mathbb{R}^{n \times k}, V \in \mathbb{R}^{n \times \ell}$. Then

$$\|\sin \Theta(U, V)\| = \|U_c^TV\| = \|U^TV_c\|.$$

The following lemma discusses the perturbation bound for the roots of a third order equation.

**Lemma 5.3.** Given a perturbed third order equation $t^3 + (p + \epsilon)t + q = 0$, where $p, q \in \mathbb{R}$ and $\epsilon \in \mathbb{R}$ is a small perturbation. Denote the roots of $t^3 + pt + q = 0$ by $t_1, t_2, t_3$, and assume that the multiplicity of each root is no more than two. Then the roots of $t^3 + (p + \epsilon)t + q = 0$ lie in $\cup_{i=1}^3 \{ z \in \mathbb{C} \mid |z - t_i| \leq r \}$, where $r = O(\sqrt{\epsilon})$.

**Proof.** Let the roots of $t^3 + (p + \epsilon)t + q = 0$ be $\tilde{t}_1, \tilde{t}_2, \tilde{t}_3$. Notice that $t_1, t_2$ and $t_3$ are the eigenvalues of $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -q & -p & 0 \end{bmatrix}$, $\tilde{t}_1, \tilde{t}_2, \tilde{t}_3$ are the eigenvalues of $\tilde{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -q & -p - \epsilon & 0 \end{bmatrix}$. Since the multiplicity of $t_i$ is no more than two, the size of each diagonal block of the Jordan canonical form of $A$ is no more than two. Using [14, Theorem 8], we know that for each $\tilde{t}_i$, there exists a $t_j$ such that

$$\frac{1}{1 + |\tilde{t}_i - t_j|^{s-1}} \leq o(1) \left\| \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \epsilon & 0 \end{bmatrix} \right\| = O(\epsilon), \quad (11)$$

where $s = 1$ or $2$. Therefore, $|\tilde{t}_i - t_j| \leq O(\sqrt{\epsilon})$. The conclusion follows. \qed

6 Proof

In this section, we present the proofs of the theoretical results in the paper.

6.1 Proof of Theorem 2.1

**Theorem 2.1.** Let $\tau_p, A$ be a solution to BJBDP for $C$. Then $\mathcal{R}(A) = \mathcal{N}(C) = \mathcal{N}(C^T)$.

**Proof.** Using (1), for any $v \in \mathcal{N}(A^T)$, we have $C_i^Tv = \Sigma_iA^Tv = 0$, similarly, $C_i^Tv = 0$. Therefore, $\mathcal{N}(C^T) \subset \mathcal{N}(C)$.

Next, we show $\sigma_p(C) > 0$ by contradiction. If $\sigma_p(C) = 0$, there exists a nonzero vector $v \notin \mathcal{N}(A^T)$ such that $Cv = 0$. Let $w = A^Tv$, we know that $w \neq 0$. Partition $w$ as $w = [w_1^T, \ldots, w_\ell^T]^T$, where $w_j \in \mathbb{R}^{p_j}$ for $j = 1, \ldots, \ell$. Then there at least exists one $w_j \neq 0$. Without loss of generality, assume $w_1 \neq 0$. It follows from $Cv = 0$ that

$$0 = C_i v = A \Sigma_i A^T v = A \Sigma_i w = A \begin{bmatrix} \Sigma_i^{(1)} w_1 \\ \vdots \\ \Sigma_i^{(\ell)} w_\ell \end{bmatrix}, \quad (12)$$
Therefore, we have $\Sigma^{(11)}_i w_1 = 0$ for all $i$. Similarly, $w_i^T \Sigma^{(11)}_i = 0$ for all $i$. Let $w_i^\top \in \mathbb{R}^{p_1 \times (p_1 - 1)}$ be such that $[w_1, w_i^\top]$ be nonsingular, then

$$[w_1, w_i^\top] \Sigma^{(11)}_i [w_1, w_i^\top]^T = \begin{bmatrix} 0 & 0 \\ 0 & * \end{bmatrix} = \begin{bmatrix} 0 \\ * \end{bmatrix}, \quad \text{for } i = 1, \ldots, m,$$

i.e., $C_i = \{\Sigma^{(11)}_i\}_{i=1}^m$ can be further block diagonalized, which contradicts with the assumption that $(\tau_p, A)$ is a solution to the BJBDP.

Now we have $\dim(\mathcal{N}(C)) \leq d - p$. Combining it with $\dim(\mathcal{N}(A^\top)) = d - p$ and $\mathcal{N}(A^\top) \subset \mathcal{N}(C)$, we have $\mathcal{N}(A^\top) = \mathcal{N}(C)$. Then it follows that

$$\mathcal{R}(A) = \mathcal{N}(A^\top)^\perp = \mathcal{N}(C)^\perp = \mathcal{R}(C^\top).$$

This completes the proof. \qed

### 6.2 Proof of Theorem 2.2

**Theorem 2.2.** Let $(\tau_p, A)$ be a solution to the BJBDP for $C$. Let the columns of $V_2$ be an orthonormal basis for $\mathcal{N}(A^\top)$, $\phi_1 \geq \cdots \geq \phi_d$ and $\tilde{\phi}_1 \geq \cdots \geq \tilde{\phi}_d$ be the singular values of $C$ and $\tilde{C}$, respectively. Then

$$\tilde{\phi}_p \geq \phi_p - \|E\|, \quad \tilde{\phi}_{p+1} \leq \|E\|. \quad (13)$$

In addition, let $\tilde{V}_1 = [\tilde{u}_1, \ldots, \tilde{u}_p]$, $\tilde{V}_1 = [\tilde{v}_1, \ldots, \tilde{v}_p]$, where $\tilde{u}_j, \tilde{v}_j$ are the left and right singular vector of $\tilde{C}$ corresponding to $\tilde{\phi}_j$, respectively, and $\tilde{U}_1, \tilde{V}_1$ are both orthonormal. If $\|E\| < \frac{\tilde{\phi}_p}{2}$, then

$$\|\sin \Theta(\mathcal{R}(A), \mathcal{R}(\tilde{V}_1))\| \leq \frac{\|\tilde{U}_1^\top EV_2\|}{\tilde{\phi}_p}.$$

**Proof.** First, by Theorem 2.1, we know that $\tilde{\phi}_{p+1} = \cdots = \tilde{\phi}_d = 0$. On the other hand, by Lemma 5.1, we have

$$|\tilde{\phi}_j - \phi_j| \leq \|\tilde{C} - C\| = \|E\|, \quad \text{for } j = 1, \ldots, d.$$ 

Then (2) follows.

Second, using (2) and $\|E\| < \frac{\tilde{\phi}_p}{2}$, we have $\tilde{\phi}_p \geq \phi_p - \|E\| > \frac{\phi_p}{2} > \|E\| \geq \tilde{\phi}_{p+1}$. Thus, $\mathcal{R}(\tilde{V}_1)$ is well defined. By calculations, we have

$$\text{diag}(\tilde{\phi}_1, \ldots, \tilde{\phi}_p) \tilde{V}_1^\top V_2 \stackrel{(a)}{=} \tilde{U}_1^\top \tilde{C} V_2 = \tilde{U}_1^\top (C + E) V_2 \stackrel{(b)}{=} \tilde{U}_1^\top V_2,$$

where (a) uses $\text{diag}(\tilde{\phi}_1, \ldots, \tilde{\phi}_p) \tilde{V}_1^\top V_2 = \tilde{U}_1^\top \tilde{C}$, (b) uses $\tilde{C} V_2 = 0$. Then using Lemma 5.2, we get

$$\|\sin \Theta(\mathcal{R}(A), \mathcal{R}(\tilde{V}_1))\| = \|V_1^\top V_2\| = \|\text{diag}(\tilde{\phi}_1, \ldots, \tilde{\phi}_p)^{-1} \tilde{U}_1^\top V_2\| \leq \frac{\|\tilde{U}_1^\top EV_2\|}{\tilde{\phi}_p}.$$

The proof is completed. \qed

### 6.3 Proof of Theorem 2.3

**Theorem 2.3.** Given $C = \{C_i\}_{i=1}^m$ with $C_i \in \mathbb{R}^{d \times d}$. Let $V_1 \in \mathbb{R}^{d \times p}$ be such that $V_1^\top V_1 = I_p$, $\mathcal{R}(V_1) = \mathcal{R}(C^\top)$. Denote $B_i = V_1^\top C_i V_1$, $B = \{B_i\}_{i=1}^m$. Then $C_i$'s can be factorized as in (1) with $\mathcal{R}(A) = \mathcal{R}(C^\top)$ if and only if there exists a matrix $X \in \mathcal{N}(B)$, which can be factorized into

$$X = Y \text{ diag}(X_{11}, \ldots, X_{\ell \ell}) Y^{-1}, \quad (14)$$

where $Y \in \mathbb{R}^{p \times p}$ is nonsingular, $X_{jj} \in \mathbb{R}^{p_j \times p_j}$ for $1 \leq j \leq \ell$ and $\lambda(X_{jj}) \cap \lambda(X_{kk}) = \emptyset$ for $j \neq k$. 

17
Proof. \((\Rightarrow)\) (Sufficiency) Let \(W = A^\top V_1\). Since \(\mathcal{R}(C^\top) = \mathcal{R}(A) = \mathcal{R}(V_1)\), and \(V_1, A\) both have full column rank, we know that \(W\) is nonsingular. Let
\[
X = W^{-1} \Gamma W = W^{-1} \text{diag}(\gamma_1 I_{p_1}, \ldots, \gamma_\ell I_{p_\ell}) W,
\]
where \(\gamma_1, \ldots, \gamma_\ell\) be \(\ell\) distinct real numbers. For all \(1 \leq i \leq m\), we have
\[
B_i X \stackrel{(a)}{=} W^\top \Sigma_i W W^{-1} \Gamma W = W^\top \Sigma_i \Gamma W = W^\top \Gamma \Sigma_i W = W^\top \Gamma W^\top W^\top \Sigma_i W = X^\top B_i,
\]
where both (a) and (b) use \(W = A^\top V_1\), (1) and (15). Therefore, \(X \in \mathcal{N}(B)\), and it is of form (3).

\((\Leftarrow)\) (Necessity) Substituting (3) into (20), we get
\[
B_i X = X^\top B_i,
\]
for all \(1 \leq i \leq m\). Let \(B_i Y \stackrel{(3)}{=} [\Sigma_i^{(j,k)}] \in \mathbb{R}^{p_i \times p_k}\), then it follows from (16) that
\[
\Sigma_i^{(j,k)} X_{kk} = X_{jj}^\top \Sigma_i^{(j,k)}, \quad \text{for} \quad j, k = 1, 2, \ldots, \ell.
\]
Consequently, for \(j \neq k\), we know that \(\Sigma_i^{(j,k)} = 0\) since \(\lambda(X_{jj}) \cap \lambda(X_{kk}) = \emptyset\). Then we know that
\[
V_1^\top C_i V_1 = B_i = Y^{-\top} \Sigma_i Y^{-1},
\]
where \(\Sigma_i = \text{diag}(\Sigma_i^{(1,1)}, \ldots, \Sigma_i^{(\ell,\ell)})\). Using \(\mathcal{R}(C_i^\top) = \mathcal{R}(V_1)\), we know that \(\mathcal{R}(C_i) \subset \mathcal{R}(V_1)\) and \(\mathcal{R}(C_i^\top) \subset \mathcal{R}(V_1)\). Then it follows from (18) that
\[
C_i = V_1 Y^{-\top} \Sigma_i Y^{-1} V_1^\top.
\]
Set \(A = V_1 Y^{-\top}\), the conclusion follows immediately. \(\square\)

### 6.4 Proof of Theorem 2.4

**Theorem 2.4.** Let \(A\) be a \(\tau_p\)-block diagonalizer of \(C\) i.e., (1) holds. Then \((\tau_p, A)\) is the unique solution to the BJBDP for \(C\) if and only if both (P1) and (P2) hold.

**Proof.** \((\Rightarrow)\) (Sufficiency) First, we show (P1) by contradiction. If (P1) doesn’t hold, there exists \(\Gamma_{jj} \in \mathbb{R}^{p_j \times p_j}\) such that \(\text{vec}(\Gamma_{jj}) \in \mathcal{N}(G_{jj})\) and a nonsingular \(W_j \in \mathbb{R}^{p_j \times p_j}\) such that
\[
\Gamma_{jj} = W_j \text{diag}(\Gamma_{jj}^{(a)}, \Gamma_{jj}^{(b)}) W_j^{-1},
\]
where \(\Gamma_{jj}^{(a)}\) and \(\Gamma_{jj}^{(b)}\) are two real matrices and \(\lambda(\Gamma_{jj}^{(a)}) \cap \lambda(\Gamma_{jj}^{(b)}) = \emptyset\). Using \(\text{vec}(\Gamma_{jj}) \in \mathcal{N}(G_{jj})\), we have
\[
\Sigma_i^{(j,j)} \Gamma_{jj} - \Gamma_{jj}^\top \Sigma_i^{(j,j)} = 0, \quad \text{for} \quad 1 \leq i \leq m.
\]
Substituting (19) into (20), we get
\[
\Sigma_i^{(j,j)} \text{diag}(\Gamma_{jj}^{(a)}, \Gamma_{jj}^{(b)}) - \text{diag}(\Gamma_{jj}^{(a)}, \Gamma_{jj}^{(b)})^\top \Sigma_i^{(j,j)} = 0, \quad \text{for} \quad 1 \leq i \leq m.
\]
where \(\Sigma_i^{(j,j)} = W_j^\top \Sigma_i^{(j,j)} W_j\). Similar to the proof of necessity for Theorem 2.3, using \(\lambda(\Gamma_{jj}^{(a)}) \cap \lambda(\Gamma_{jj}^{(b)}) = \emptyset\), we have \(\Sigma_i^{(j,j)}\) for \(1 \leq i \leq m\) are all block diagonal matrices. In other words, \(C_i\)'s can be simultaneously block diagonalizable with more than \(\ell\) blocks. This contradicts with the fact \((\tau_p, A)\) is the solution to the BJBDP.
Next, we show \((P2)\), also by contradiction. Since \(G_{jk}\) is rank deficient, then there exist two matrices \(\Gamma_{jk}, \Gamma_{kj}\), which are not zero at the same time, such that (4b) holds, i.e.,

\[
\begin{bmatrix}
\Sigma_i^{(j)} & 0 \\
0 & \Sigma_i^{(kk)}
\end{bmatrix}
\begin{bmatrix}
0 & \Gamma_{jk} \\
\Gamma_{kj} & 0
\end{bmatrix}
\begin{bmatrix}
\Sigma_i^{(j)} & 0 \\
0 & \Sigma_i^{(kk)}
\end{bmatrix} = 0.
\]

(22)

Since \(\begin{bmatrix} 0 & \Gamma_{jk} \\ \Gamma_{kj} & 0 \end{bmatrix} \neq 0\), it has at least a nonzero eigenvalue. Now let \(\lambda\) be a nonzero eigenvalue of \(\begin{bmatrix} 0 & \Gamma_{jk} \\ \Gamma_{kj} & 0 \end{bmatrix}\), and \(\begin{bmatrix} x \\ y \end{bmatrix}\) be the corresponding eigenvector. Then it is easy to see that \(-\lambda\) is also an eigenvalue, and the corresponding eigenvector is \(\begin{bmatrix} -x \\ y \end{bmatrix}\). In addition, \(x \neq 0\) and \(y \neq 0\). Therefore, there exists a nonsingular matrix \(W_{jk}\), which is not \((p_j, p_k)\)-block diagonal, such that

\[
\begin{bmatrix}
0 & \Gamma_{jk} \\
\Gamma_{kj} & 0
\end{bmatrix} = W_{jk}
\begin{bmatrix}
\Upsilon & 0 \\
0 & -\Upsilon
\end{bmatrix} W_{jk}^{-1},
\]

(23)

where \(\Upsilon\) is nonsingular, \(\lambda(\Upsilon) \cap \lambda(-\Upsilon) = \emptyset\) and \(W_{jk}\) is not \((p_j, p_k)\)-block diagonal. Plugging (23) into (22), similar to the proof of necessity for Theorem 2.3, we can show that \(W_{jk}^\top \Sigma_i^{(j)j} \Sigma_i^{(k)k} W_{jk}\) for all \(1 \leq i \leq m\) are all block diagonal. For the ease of notation, let \(j = 1, k = 2\). Denote \(\tilde{A} = A \text{diag}(W_{12}^{-\top}, I_{p_3}, \ldots, I_{p_\ell})\). We know that \(A, \tilde{A}\) are not equivalent since \(W_{12}\) is not \((p_1, p_2)\)-block diagonal. This contradicts with the assumption that \(\text{BJBDP}\) for \(\mathcal{C}\) is uniquely \(\tau_p\)-block-diagonalizable, completing the proof of sufficiency.

\((\Leftarrow)\) (Necessity) Let \((\tilde{\tau}_p, \tilde{A})\) be a solution to the \(\text{BJBDP}\) for \(\mathcal{C}\). Then it holds that

\[
C_i = A\Sigma_i A^\top = \tilde{A}\Sigma_i \tilde{A}^\top,
\]

(24)

where \(\Sigma_i\)'s are all \(\tau_p\)-block diagonal, \(\tilde{\Sigma}_i\)'s are all \(\tilde{\tau}_p\)-block-diagonal. It suffices if we can show that \((\tau_p, A)\) and \((\tilde{\tau}_p, \tilde{A})\) are equivalent.

Let \(\tau_p = (p_1, \ldots, p_\ell), \tilde{\tau}_p = (\hat{p}_1, \ldots, \hat{p}_\ell)\). As \((\tilde{\tau}_p, \tilde{A})\) is a solution to \(\text{BJBDP}\), it holds that \(\ell \leq \hat{\ell}\). By virtue of Theorem 2.1, we know that \(\mathcal{R}(C_i^\top) = \mathcal{R}(A) = \mathcal{R}(\tilde{A})\). Since \(A\) and \(\tilde{A}\) are both of full column rank, we know that \(p = \hat{p}\) and there exists a nonsingular matrix \(Z\) such that \(\tilde{A} = AY^{-\top}\). Then it follows from (24) that

\[
\tilde{\Sigma}_i = Y^\top \Sigma_i Y, \quad \text{for } 1 \leq i \leq m.
\]

(25)

Let \(\Gamma = Y \text{diag}(\gamma_1 I_{p_1}, \ldots, \gamma_\ell I_{p_\ell})Y^{-1}\), where \(\gamma_1, \ldots, \gamma_\ell\) are distinct real numbers. Using (25), we have

\[
\Sigma_i \Gamma = Y^{-\top} (Y^\top \Sigma_i Y) \text{diag}(\gamma_j I_{p_j})Y^{-1} = Y^{-\top} \text{diag}(\gamma_j I_{p_j})(Y^\top \Sigma_j Y)Y^{-1} = \Gamma^\top \Sigma_i,
\]

(26)

i.e., \(\Gamma \in \mathcal{M}^\ell(\{\Sigma_i\})\).

Partitioning \(\Gamma = [\Gamma_{jk}] \in \mathbb{R}^{p_j \times p_k}\). Recall (4) and (5), by (P2), we have \(\Gamma_{jk} = 0\) for \(j \neq k\), i.e., \(\Gamma\) is \(\tau_p\)-block diagonal; using (P1), \(\Gamma = Y \text{diag}(\gamma_j I_{p_j})Y^{-1}\) and \(\cup_{j=1}^\ell \lambda(\Gamma_{jj}) = \lambda(\Gamma)\), we know that \(\ell = \hat{\ell}\), \(\lambda(\Gamma_{kk}) = \lambda(\gamma_j I_{p_j})\) for \(1 \leq j \leq \ell\), where \(\{k_1, k_2, \ldots, k_\ell\}\) is a permutation of \(\{1, 2, \ldots, \ell\}\). Thus, \(\hat{p}_j = p_{k_j}\) for \(1 \leq j \leq \ell\). In other words, there exists a permutation \(\Pi_{\ell} \in \mathbb{R}^{\ell \times \ell}\) such that \(\tilde{\tau}_p = \tau_p \Pi_{\ell}\). Let \(\Pi \in \mathbb{R}^{p \times p}\) be the permutation matrix associated with \(\Pi_{\ell}\). Then

\[
\text{diag}(\gamma_1 I_{p_1}, \ldots, \gamma_\ell I_{p_\ell}) = \Pi^\top \text{diag}(\gamma_1 I_{p_1}, \ldots, \gamma_\ell I_{p_\ell})\Pi.
\]

(27)

where \(\gamma_j'\) is the eigenvalue of \(\Gamma_{jj}\). Then it follows that

\[
\text{diag}(\Gamma_{11}, \ldots, \Gamma_{\ell\ell}) = YY^\top \text{diag}(\gamma_1 I_{p_1}, \ldots, \gamma_\ell I_{p_\ell})(YY^\top)^{-1}.
\]

(28)

Noticing that the columns of \(YY^\top\) are eigenvectors of \(\Gamma\), we know that \(YY^\top\) is \(\tau_p\)-block-diagonal. Therefore, we can rewrite \(\tilde{A} = AY^{-\top}\) as \(\tilde{A} = A(YY^\top)^{-\top}\Pi\), in which \((YY^\top)^{-\top}\Pi\) is \(\tau_p\)-block-diagonal, \(\Pi\) is the permutation matrix associated with \(\Pi_{\ell}\). Thus, \((\tau_p, A)\) and \((\tilde{\tau}_p, \tilde{A})\) are equivalent. The proof is completed. □
6.5 Proof of Theorem 2.5

**Theorem 2.5.** Given a set \( \mathcal{D} = \{D_i\}_{i=1}^m \) of \( q \)-by-\( q \) matrices with \( \mathcal{D} \) having full column rank.

(I) If \( \mathcal{D} \) does not have a nontrivial diagonalizer, then the feasible set of \( \text{OPT}(\mathcal{D}) \) is empty.

(II) If \( \mathcal{D} \) has a nontrivial diagonalizer, then \( \text{OPT}(\mathcal{D}) \) has a solution \( X_* \). In addition, assume

\[
\mu = \min_{\|z\|=1} \sqrt{\sum_{i=1}^m |z^H D_i z|^2} > 0,
\]

then \( X_* \) has two distinct real eigenvalues, and the gap between them are no less than two.

**Proof.** First, we show (I) via its the contrapositive. If the feasible set of \( \text{OPT}(\mathcal{D}) \) is not empty, then it has a solution \( X_* \). Using \( \text{tr}(X_*) = 0, \text{tr}(X_*^2) = q > 0 \), we know that \( X_* \) can be factorized into \( X_* = Y \text{diag}(\Gamma_1, \Gamma_2)Y^{-1} \), where \( \Gamma_1, \Gamma_2 \) are real matrices and \( \lambda(\Gamma_1), \lambda(\Gamma_2) \) lie in the open left and closed right complex planes, respectively. Therefore, \( \lambda(\Gamma_1) \cap \lambda(\Gamma_2) = \emptyset \). By Theorem 2.3, \( \mathcal{D} \) has a nontrivial diagonalizer, completing the proof of (I).

Next, we show (II). Let \( \gamma \) be an arbitrary eigenvalue of \( X_* \), and \( z \) be the corresponding eigenvector. Using \( X_* \in \mathcal{M}(\mathcal{D}) \), we have

\[
0 = z^H D_i X_* z - z^H X_*^T D_i z = (\gamma - \bar{\gamma}) z^H D_i z, \quad \text{for } 1 \leq i \leq m.
\]

Then it follows that

\[
(\gamma - \bar{\gamma}) \sum_{i=1}^\ell |z^H D_i z|^2 = 0.
\]

Since \( \mu > 0 \) has full column rank, we know that \( \sum_{i=1}^\ell |z^H D_i z|^2 = 0 \). Therefore, \( \gamma \) is real. It follows \( \lambda(X_*) \subset \mathbb{R} \).

Now we show that \( X_* \) has two distinct eigenvalues. Denote the eigenvalues of \( X_* \) by \( \gamma_1 \leq \cdots \leq \gamma_q \). Then

\[
\text{tr}(X_*) = \sum_{j=1}^q \gamma_j = 0, \quad \text{tr}(X_*^2) = \sum_{j=1}^q \gamma_j^2 = q, \quad \text{tr}(X_*^4) = \sum_{j=1}^q \gamma_j^4.
\]

Using the method of Lagrange multipliers, we consider

\[
L(\gamma_1, \ldots, \gamma_q; \mu_1, \mu_2) = \sum_{j=1}^q \gamma_j^4 + \mu_1 \sum_{j=1}^q \gamma_j + \mu_2 \left( \sum_{j=1}^q \gamma_j^2 - q \right),
\]

where \( \mu_1, \mu_2 \) are Lagrange multipliers. By calculations, we have

\[
\frac{\partial L}{\partial \gamma_j} = 4\gamma_j^3 + \mu_1 + 2\mu_2 \gamma_j = 0.
\]

Noticing that \( \gamma_j \)'s are the real roots of the third order equation \( 4\gamma^3 + 2\mu_2 \gamma + \mu_1 = 0 \), which has one real root or three real roots, we know that either \( \gamma_j \)'s are identical to the unique real root or \( \gamma_j \) is one of the three real roots for all \( j \). The former case is impossible since \( \sum_{j=1}^q \gamma_j = 0 \) and \( \sum_{j=1}^q \gamma_j^2 = q \). For the latter case, set \( \gamma_1 = \cdots = \gamma_{q_1} = t_1, \gamma_{q_1+1} = \cdots = \gamma_{q_1+q_2} = t_2 \) and \( \gamma_{q_1+q_2+1} = \cdots = \gamma_q = t_3 \), where \( t_1 \leq t_2 \leq t_3 \) are the three real roots, \( q_1, q_2 \) and \( q_3 \) are respectively the multiplicities of \( t_1, t_2 \) and \( t_3 \) as eigenvalues of \( X_* \). If \( t_1 = t_2 < t_3 \) or \( t_2 = t_3 \), \( X_* \) has two distinct eigenvalues. In what follows we assume \( t_1 < t_2 < t_3 \).

Using (29), we get

\[
q_1 t_1^3 + q_2 t_2^3 + q_3 t_3^3 = 0, \quad q_1 t_1^2 + q_2 t_2^2 + q_3 t_3^2 = q, \quad \text{tr}(X_*^4) = q_1 t_1^4 + q_2 t_2^4 + q_3 t_3^4.
\]

Introduce two vectors \( u = [\sqrt{q_1} t_1^2, \sqrt{q_2} t_2^2, \sqrt{q_3} t_3^2]^T, v = [\sqrt{q_1}, \sqrt{q_2}, \sqrt{q_3}]^T \). Then we have \( \|u\| = \sqrt{\text{tr}(X_*^2)}, \|v\| = \sqrt{q} \). Using Cauchy's inequality, we get

\[
\text{tr}(X_*^4) = \|u\|^2\|v\|^2/q \geq (u^\top v)^2/q = (q_1 t_1^2 + q_2 t_2^2 + q_3 t_3^2)^2/q = q.
\]
and the equality holds if and only if $u$ and $v$ are co-linear. Using the first two equalities of (31), $q_1, q_2, q_3$ can not have more than one zeros. If one of $q_1, q_2, q_3$ is zero, $X_*$ has two distinct eigenvalues. Otherwise, $q_1, q_2$ and $q_3$ are all positive integers. Therefore, $t_1^2 = t_2^2 = t_3^2$, which implies that $A_*$ has two distinct eigenvalues.

The above proof essentially show that the optimal value is achieved at $X = X_*$. The following statements show that such an $X$ is feasible in $N(D)$. If $D$ has a nontrivial diagonalizer, then there exists a matrix $Z$ such that $D = ZF_1Z^\top$, where $F_i$'s are $q_i (q_1, q_2)$-block diagonal. Since $D$ has full column rank, $Z$ is nonsingular. Let $X = Z^{-\top} \text{diag}(\sqrt{q_1}I_{q_1}, -\sqrt{q_2}I_{q_2})Z^\top$. It is easy to see that $\text{tr}(X) = 0$, $\text{tr}(X^2) = 1$ and $X \in N(D)$. In other words, there exists a feasible $X$ which has two distinct real eigenvalues. Therefore, we may declare that $\text{OPT}(D)$ is minimized at $X = X_*$, with $X_*$ having two distinct real eigenvalues.

Lastly, let $\gamma_1 > \gamma_2$ be the distinct real eigenvalues of $X_*$, with multiplicities $q_1$ and $q_2$, respectively, we show $\gamma_1 - \gamma_2 \geq 2$. Rewrite the first equalities of (29) as

$$q_1 \gamma_1 + q_2 \gamma_2 = 0, \quad q_1 \gamma_1^2 + q_2 \gamma_2^2 = q.$$ 

By calculations, we get $\gamma_1 = \sqrt{\frac{q_2}{q_1}}$, $\gamma_2 = \frac{\sqrt{q_1} + \sqrt{q_2}}{q_2}$. Then it follows that

$$\gamma_1 - \gamma_2 = \sqrt{\frac{q_2}{q_1}} + \sqrt{\frac{q_1}{q_2}} \geq 2,$$

completing the proof.

\section{6.6 Proof of Theorem 2.6}

Theorem 2.6. Assume that the BJBDP for $C$ is uniquely $\tau_p$-block-diagonalizable, and let $(\tau_p, A)$ be a solution satisfying (1). Then $(\tau_p, A)$ can be identified via Algorithm 2, almost surely.

Proof. If we can show $\text{card}(\tau_p) = \text{card}(\tau_p)$, then $(\tau_p, A)$ is also a solution to the BJBDP for $C$. Since the BJBDP is uniquely $\tau_p$-block-diagonalizable, we know that $(\tau_p, A)$ is equivalent to $(\tau_p, A)$, i.e., $(\tau_p, A)$ is identified. Next, we show $\text{card}(\tau_p) = \text{card}(\tau_p)$. The following two facts are needed.

(I) Given a matrix set $D$ with $D$ having full column rank. If $D$ does not have any $\tau_q$-block diagonalizer with $\text{card}(\tau_q) \geq 2$, then $\hat{\tau}$ on Line 9 of Algorithm 2 satisfies $\text{card}(\hat{\tau}) = 1$; Otherwise, $\text{card}(\hat{\tau}) = 2$.

(II) Denote $\hat{Z}^{-1}D_1\hat{Z}^{-\top} = \text{diag}(D_1^{(a)}, D_1^{(b)}), \ D_1^{(a)} = \{D_1^{(a)}\}$ and $D_1^{(b)} = \{D_1^{(b)}\}$. Then $D_1^{(a)}$ and $D_1^{(b)}$ both have full column rank.

Fact (I) is because when $\text{card}(\hat{\tau}) > 1$, $D$ can be block diagonalized. Fact (II) is due to the fact $\hat{Z}$ is nonsingular and $\hat{Z}^{-1}D_1\hat{Z}^{-\top} = \text{diag}(D_1^{(a)}, D_1^{(b)})$.

Now assume that the solution $(\hat{\tau}_p, A)$ returned by Algorithm 2 satisfies

$$\hat{\tau}_p = (\hat{\tau}_p^1, \ldots, \hat{\tau}_p^\ell), \quad C_i = \hat{A}\Sigma_i\hat{A}^\top = \hat{A}\Sigma_i^{(11)} \ldots \Sigma_i^{(\ell\ell)}\hat{A}^\top, \quad i = 1, \ldots, m,$$

where $\hat{\Sigma}_i$'s are all $\hat{\tau}_p^\ell$-block diagonal. Then $\ell \leq \ell$ and $\Sigma_i^{(jj)}$ can be further block diagonalized for all $j = 1, \ldots, \ell$. Next, we show $\text{card}(\hat{\tau}_p) = \ell = \text{card}(\tau_p)$ by contradiction.

Using (1) and (32), we have

$$B_i = V_1^\top \hat{A}\Sigma_i\hat{A}^\top V_i = \hat{Z}\Sigma_i\hat{Z}^\top = V_1^\top A\Sigma_i A^\top V_i = \hat{Z}\Sigma_i\hat{Z}^\top.$$  

where $\hat{Z} = V_1^\top \hat{A}, \ Z = V_1^\top A$. By Theorem 2.1, we know that $\mathcal{R}(V_1) = \mathcal{R}(C^\top) = \mathcal{R}(A)$. By the construction of $\hat{A}$, we know $\mathcal{R}(V_1) = \mathcal{R}(\hat{A})$. Since $V_1, A, \hat{A}$ all have full column rank, we know that $\hat{Z}$ and $Z$ are both nonsingular. Then it follows from (33) that

$$\Sigma_i = Y^\top \Sigma_i Y, \quad \text{for } 1 \leq i \leq m.$$
where $Y = Z^T \tilde{Z}^{-\top}$. Let $\Gamma = Y \text{diag}(\gamma_1 I_{\tilde{p}_1}, \ldots, \gamma_\ell I_{\tilde{p}_\ell})Y^{-1}$, where $\gamma_1, \ldots, \gamma_\ell$ are distinct real numbers. Using (34), we have

$$\Sigma_i \Gamma = Y^{-\top} (Y^\top \Sigma_i Y) \text{diag}(\gamma_j I_{\tilde{p}_j}) Y^{-1} = Y^{-\top} \text{diag}(\gamma_j I_{\tilde{p}_j})(Y^\top \Sigma_i Y) Y^{-1} = \Gamma^\top \Sigma_i, \quad (35)$$

i.e., $\Gamma \in \mathcal{N}(\{\Sigma_i\})$.

Partition $\Gamma = [\Gamma_{jk}]$ with $\Gamma_{jk} \in \mathbb{R}^{p_j \times p_k}$. Recall (4) and (5), by (P2), we have $\Gamma_{jk} = 0$ for $j \neq k$, i.e., $\Gamma$ is $\tau_j$-block diagonal; using (P1), $\Gamma = Y \text{diag}(\gamma_j I_{\tilde{p}_j})Y^{-1}$ and $\cup_{j=1}^\ell \lambda(\Gamma_{jj}) = \lambda(\Gamma)$, we know that for each $\Gamma_{jj}$ ($j = 1, \ldots, \ell$), its eigenvalues are all $\gamma_k$ ($1 \leq k \leq \ell$). If $\ell < \ell$, there exist at least two blocks of $\Gamma_{jj}$’s corresponding to the same $\gamma_k$’s. Without loss of generality, let $\Gamma_{11}, \Gamma_{22}$ correspond to $\gamma_1$, the remaining blocks correspond to other $\gamma_k$’s. Then using $\Gamma = Y \text{diag}(\gamma_1 I_{\tilde{p}_1}, \ldots, \gamma_\ell I_{\tilde{p}_\ell})Y^{-1}$, we know that $Y = \text{diag}(\gamma_{11} Y_{11}, \gamma_{22} Y_{22})$, where $Y_{11} \in \mathbb{R}^{\tilde{p}_1 \times \tilde{p}_1}$ and $\tilde{p}_1 = p_1 + p_2$. Using $Y = Z^T \tilde{Z}^{-\top}$ and (35), we get

$$\tilde{\Sigma}_i = Y^\top \Sigma_i Y = \text{diag}(Y_{11}^\top, Y_{22}^\top) \Sigma_i \text{diag}(Y_{11}, Y_{22}), \quad \text{for } 1 \leq i \leq m.$$ Therefore, we have

$$\tilde{\Sigma}_i^{(11)} = Y_{11}^\top \text{diag}(\Sigma_i^{(11)}, \Sigma_i^{(22)}) Y_{11}, \quad \text{for } 1 \leq i \leq m,$$

which contradicts with the fact that $\{\tilde{\Sigma}_i^{(11)}\}_{i=1}^m$ can not be further block diagonalized. The proof is completed.

\[\square\]

### 6.7 Proof of Theorem 2.7

**Theorem 2.7.** Given a set $\tilde{D} = \{\tilde{D}_i\}_{i=1}^m$ of $q$-by-$q$ matrices with $\tilde{D}$ having full column rank. Let $\delta = o(1)$ be a small real number.

\(\textbf{(I)}\) If $\tilde{D}$ does not have a nontrivial $\delta$-diagonalizer, then the feasible set of $\text{opt}(\tilde{D}, \delta)$ is empty.

\(\textbf{(II)}\) If $\tilde{D}$ has a nontrivial $\delta$-diagonalizer, then $\text{opt}(\tilde{D}, \delta)$ has a solution $X_\ast$. In addition, assume

$$\mu = \min_{\|z\| = 1} \sqrt{\sum_{i=1}^m |z^H \tilde{D}_i z|^2} = O(1),$$

and for $i = 1, 2$, let

$$\text{Rect}_i \triangleq \{z \in \mathbb{C} | |\text{Re}(z) - \rho_i| \leq a, |\text{Im}(z)| \leq b\},$$

where $a = O(\delta)$, $b = O(\delta)$. Then

$$\lambda(X_\ast) \subset \cup_{i=1}^2 \text{Rect}_i, \quad \rho_1 - \rho_2 \geq 2 + O(\delta).$$

**Proof.** First, we show of (I) via its the contrapositive. If the feasible set of $\text{opt}(\tilde{D}, \delta)$ is not empty, then $\text{opt}(\tilde{D}, \delta)$ has a solution $X_\ast$, which can be factorized into $X_\ast = Y \text{diag}(\Gamma_1, \Gamma_2)Y^{-1}$ (since $\text{tr}(X_\ast) = 0$ and $\text{tr}(X^2_\ast) = q$), where $Y$ is nonsingular, $\Gamma_1 \in \mathbb{R}^{q_1 \times q_1}$, $\Gamma_2 \in \mathbb{R}^{q_2 \times q_2}$ and $\lambda(\Gamma_1) \cap \lambda(\Gamma_2) = \emptyset$. Set $Z = Y^{-\top}$, $\Phi_i = \text{diag}(Y_1^\top \tilde{D}_1 Y_1, Y_2^\top \tilde{D}_1 Y_2)$, $g = \min \frac{||Y^2_1 X - X Y_2||_F}{||X||_F}$ and $\kappa = \kappa_2(Y) = \frac{\sigma_{\max}(Y)}{\sigma_{\min}(Y)}$. By calculations, we have

$$||X_\ast||_F^2 = \text{tr}(Y^{-\top} \text{diag}(\Gamma_1^\top, \Gamma_2^\top) Y \text{diag}(\Gamma_1, \Gamma_2) Y^{-1}) \leq ||Y||^2 \text{tr}(Y^{-\top} \text{diag}(\Gamma_1^\top, \Gamma_2^\top) \text{diag}(\Gamma_1, \Gamma_2) Y^{-1}) = ||Y||^2 \text{tr}(\text{diag}(\Gamma_1, \Gamma_2) Y^{-\top} \text{diag}(\Gamma_1^\top, \Gamma_2^\top)) \leq \kappa^2 \text{tr}(\text{diag}(\Gamma_1, \Gamma_2) \text{diag}(\Gamma_1^\top, \Gamma_2^\top)) = \kappa^2 \text{tr}(X_\ast^2) = \kappa^2 q,$$  
\[\text{(36)}\]
and
\[ \delta^2 \| \text{vec}(X_*) \|^2 \geq \| \mathbf{L}(\widetilde{D}) \text{vec}(X_*) \|^2 = \sum_{i=1}^{m} \| \widetilde{D}_i X_* - X_*^T \widetilde{D}_i \|^2_F, \]
\[ \geq \sum_{i=1}^{m} \| Z(Y^T \widetilde{D}_i Y) \|_{F} \geq \frac{g^2}{\kappa^6} \sum_{i=1}^{m} \| \widetilde{D}_i - Z \Phi_i Z^T \|^2_F, \]
where (a) uses \( X_* \in \mathcal{A}(\widetilde{D}) \), (b) uses the definition of \( g \). Then it follows from (36) and (37) that
\[ \sum_{i=1}^{m} \| \widetilde{D}_i - Z \Phi_i Z^T \|^2_F \leq \frac{\kappa^4 \| X_* \|^2_F}{g^2} \delta^2 \leq \frac{\kappa^6}{g^2 q^2} \delta^2. \]
This completes the proof of (I).

Next, we show (II). If \( \widetilde{D} \) has a nontrivial \( \delta \)-diagonalizer, then there exists a matrix \( Z \) such that
\[ \sum_{i=1}^{m} \| \widetilde{D}_i - Z \Phi_i Z^T \|^2_F \leq \frac{1}{4} \delta^2 \] (by setting \( \delta = \frac{1}{2 \sqrt{\kappa}} \delta \), the constant becomes \( \frac{1}{4} \), and by definition, \( Z \) is still a \( \delta \)-diagonalizer), where \( \Phi_i \)'s are all \( \tau_q = (q_1, q_2) \) block diagonal matrices. Let \( X = Z^{-1} \Gamma Z^T \), where \( \Gamma = \text{diag}(\sqrt{\frac{\tau_q}{q_1}} I_{q_1}, -\sqrt{\frac{\tau_q}{q_2}} I_{q_2}) \). By calculations, we have
\[ \| \mathbf{L}(\widetilde{D}) \text{vec}(X) \|^2 \leq \sum_{i=1}^{m} \| \widetilde{D}_i X - Z \Phi_i Z^T X \|^2_F \leq 4 \| X \|^2 \sum_{i=1}^{m} \| \widetilde{D}_i - Z \Phi_i Z^T \|^2_F \leq \| X \|^2 \delta^2, \]
where (a) uses \( Z \Phi_i Z^T X - X^T Z \Phi_i Z^T = 0 \). Therefore, \( \| \mathbf{L}(\widetilde{D}) \text{vec}(X) \| \leq \| X \| \delta \leq \delta \). Also note that \( \text{tr}(X) = 0 \) and \( \text{tr}(X^2) = q \), then the feasible set of \( \text{OPT}(\widetilde{D}, \delta) \) is nonempty. Consequently, \( \text{OPT}(\widetilde{D}, \delta) \) has a solution \( X_* \).

Let \( \gamma \) be an arbitrary eigenvalue of \( X_* \), and \( z \) be the corresponding unit-length eigenvector. By calculations, we have
\[ \kappa^2 q \delta^2 \geq \delta^2 \| X_* \|^2_F = \| \mathbf{L}(\widetilde{D}) \text{vec}(X) \|^2 \geq \sum_{i=1}^{m} \| \widetilde{D}_i X_* - X_*^T \widetilde{D}_i \|^2_F \]
\[ \geq \sum_{i=1}^{m} \| z^H \widetilde{D}_i X_* z - z^H X_*^T \widetilde{D}_i z \|^2_F = |\gamma - \bar{\gamma}|^2 \sum_{i=1}^{m} \| z^H \widetilde{D}_i z \|^2 \geq \mu^2 |\gamma - \bar{\gamma}|^2, \]
where \( \mu \) is the smallest singular value of \( \widetilde{D} \).

Then we know that the imaginary part of \( \mu \) is no more than \( \sqrt{\frac{\tau_q}{2q}} = O(\delta) \).

Now let the eigenvalues of \( X_* \) be \( \mu_j + \eta_j \sqrt{-1} \) for \( j = 1, \ldots, q \), where \( \mu_j, \eta_j \in \mathbb{R} \). Then
\[ \text{tr}(X_*) = \sum_{j=1}^{q} \gamma_j = 0, \quad \text{tr}(X_*) = \sum_{j=1}^{q} (\gamma_j^2 - \eta_j^2) = q, \quad \text{tr}(X_*) = \sum_{j=1}^{q} (\gamma_j^4 + \eta_j^4 - \sqrt{\frac{\tau_q}{2q}} \gamma_j^2 \eta_j^2). \]
Using the method of Lagrange multipliers, we consider

\[ L(\gamma_1, \eta_1, \ldots, \gamma_q, \eta_q; \mu_1, \mu_2) = \sum_{j=1}^{q} (\gamma_j^2 + \eta_j^2 - 6\gamma_j \eta_j^2) + \mu_1 \sum_{j=1}^{q} \gamma_j + \mu_2 \left( \sum_{j=1}^{q} \gamma_j^2 - \eta_j^2 \right) - q, \]

where \( \mu_1, \mu_2 \) are Lagrange multipliers. By calculations, we have

\[ \frac{\partial L}{\partial \gamma_j} = 4\gamma_j^3 + 2(\mu_2 - 6\eta_j^2)\gamma_j + \mu_1 = 0. \]  

(40)

Take (40) as perturbed third order equations of \( 4t^3 + 2\mu_2 t + \mu_1 = 0 \). Using Lemma 5.3 and \( |\eta_j| \leq O(\delta) \), we know that \( \gamma_j \in \cup_{t=1}^{3} \{ z \mid |z - t_i| \leq O(\delta) \} \), where \( t_1, t_2 \) and \( t_3 \) are the roots of \( 4t^3 + 2\mu_2 t + \mu_1 = 0 \).

Next, we consider the following cases:

Case (1) \( t_1 = t_2 \notin \mathbb{R}, t_3 \in \mathbb{R} \).
In this case, set \( \rho_1 = \text{Re}(t_1), \rho_2 = t_3 \), then \( \lambda(X_*) \subset \cup_{i=1,2} \text{Rect}_i \).

Case (2) \( t_1, t_2, t_3 \in \mathbb{R}, t_1 = \xi + O(\delta) \) for \( i = 1, 2, 3 \).
In this case, using \( t_1 + t_2 + t_3 = 0 \) (by Vieta’s formulas), we get \( \xi = O(\delta) \). Then it follows that \( |\gamma_j| = O(\delta) \) for all \( j \). Using (39) and \( \eta_j = O(\delta) \), we get \( q \times O(\delta^2) = q \), which contradicts with \( \delta = o(1) \).

Case (3) \( t_1, t_2, t_3 \in \mathbb{R}, t_1 = \xi + O(\delta) \) for \( i = 1, 2 \).
In this case, set \( \rho_1 = \xi, \rho_2 = t_3 \), then \( \lambda(X_*) \subset \cup_{i=1,2} \text{Rect}_i \).

Case (4) \( t_1, t_2, t_3 \in \mathbb{R}, |t_i - t_j| > O(\delta) \) for \( i \neq j \).
In this case, without loss of generality, assume \( t_1 < t_2 < t_3 \), and there are \( p_i \) eigenvalues of \( X_* \) lie in \( \{ z \mid |z - t_i| \leq O(\delta) \} \), for \( i = 1, 2, 3 \). Using \( \eta_j = O(\delta) \) and (39), we get

\[ \text{tr}(X_*) = q_1 t_1 + q_2 t_2 + q_3 t_3 + O(\delta) = 0, \]

\[ \text{tr}(X^2) = q_1 t_1^2 + q_2 t_2^2 + q_3 t_3^2 + O(\delta) = q, \]

\[ \text{tr}(X^4) = q_1 t_1^4 + q_2 t_2^4 + q_3 t_3^4 + O(\delta). \]

Let \( u = [\sqrt{q_1} t_1^2, \sqrt{q_2} t_2^2, \sqrt{q_3} t_3^2]^\top, v = [\sqrt{q_1}, \sqrt{q_2}, \sqrt{q_3}]^\top. \) Then we have \( ||u||^2 + O(\delta) = \text{tr}(X^4) \), \( ||v|| = \sqrt{q} \).
Using Cauchy’s inequality, we get

\[ \text{tr}(X^4) + O(\delta) = ||u||^2 = ||u||^2 ||v||^2 / q \geq (u^\top v)^2 / q = (q_1 t_1^2 + q_2 t_2^2 + q_3 t_3^2)^2 / q = q + O(\delta), \]

and the equality holds if and only if \( u \) and \( v \) are co-linear. Using the first two equalities of (41), \( q_1, q_2, q_3 \) can not have more than one zeros. If one of \( q_1, q_2, q_3 \) is zero, say \( q_3 = 0 \), then the eigenvalues of \( X_* \) lie in two disks \( \cup_{i=1,2,3,q_i \neq 0} \{ z \mid |z - t_i| \leq O(\delta) \} \). Otherwise, \( q_1, q_2 \) and \( q_3 \) are all positive integers. Therefore, \( t_1^2 = t_2^2 = t_3^2 \), which implies that \( t_2 = t_1 \) or \( t_2 = t_3 \). This contradicts with \( t_1 < t_2 < t_3 \). To summarize, the eigenvalues of \( X_* \) lie in \( \cup_{i=1,2} \text{Rect}_i \).

The above proof essentially show that the optimal value is achieved at \( X = X_* \), with its eigenvalues lie in \( \cup_{i=1,2} \text{Rect}_i \). The following statements show that such an \( X \) is feasible in \( \mathcal{M}_2(D) \).

If \( \tilde{D} \) has a nontrivial \( \delta \)-diagonalizer, then there exists a matrix \( Z \) such that \( \sum_{i=1}^{p} \| \tilde{D}_i - Z \Phi_i Z^\top \|^2_F \leq \frac{1}{4} \delta^2 \),

where \( \Phi_i \’s \) are all \( r_2 = (q_1, q_2) \) block diagonal matrices. Let \( X = Z^\top \Gamma Z \), where \( \Gamma = \text{diag}(\sqrt{\frac{q_1}{q_2}} I_{q_1}, -\sqrt{\frac{q_2}{q_3}} I_{q_2}) \).
We know that \( X \) is also feasible. Therefore, we may declare that \( \text{OPT}(\tilde{D}, \delta) \) is minimized at \( X = X_* \), with the eigenvalues of \( X_* \) lying in two disks.

Lastly, let \( (\rho_1, 0), (\rho_2, 0) \) be the centers of the two disks, and there are \( q_1, q_2 \) eigenvalues of \( X_* \) lie \( \text{Disk}_1, \text{Disk}_2 \), respectively. We show \( \rho_1 - \rho_2 \geq 2 + O(\delta) \). Rewrite the first two equalities of (41) as

\[ q_1 \rho_1 + q_2 \rho_2 = O(\delta), \quad q_1 \rho_1^2 + q_2 \rho_2^2 = q + O(\delta). \]
By calculations, we get \( \rho_1 = \sqrt{\frac{2q_1}{q_1}} + O(\delta) \), \( \rho_2 = -\sqrt{\frac{2q_2}{q_2}} + O(\delta) \). Then it follows that
\[
\rho_1 - \rho_2 = \sqrt{\frac{q_2}{q_1}} + \sqrt{\frac{q_1}{q_2}} + O(\delta) \geq 2 + O(\delta),
\]
completing the proof.

6.8 Proof of Theorem 2.8

**Theorem 2.8.** Assume that the BBDP for \( \mathcal{C} = \{C_i\}_{i=1}^m \) is uniquely \( \tau_p \)-block-diagonalizable, and let \( (\tau_p, A) \) be a solution satisfying (1). Let \( \tilde{\mathcal{C}} = \{\tilde{C}_i\}_{i=1}^m = \{C_i + E_i\}_{i=1}^m \) be a perturbed matrix set of \( \mathcal{C} \). Denote
\[
\tau_p = (p_1, \ldots, p_\ell), \quad \tilde{\tau}_p = (\tilde{p}_1, \ldots, \tilde{p}_\ell), \quad A = [A_1, \ldots, A_\ell], \quad \tilde{A} = [\tilde{A}_1, \ldots, \tilde{A}_\ell],
\]
where \( (\tilde{\tau}_p, \tilde{A}) \) is the output of Algorithm 4. Assume \( \mathcal{N}(G_{jj}) = \mathcal{R}(\text{vec}(I_p)) \) for all \( j \), where \( G_{jj} \) is defined in (5a). Also assume that \( p \) is correctly identified in Line 3 of Algorithm 4. Let the singular values of \( \tilde{C} \) be the same as in Theorem 2.2,
\[
\epsilon = \frac{\|E\|}{\phi_p}, \quad r = \sqrt{\frac{2(d + 2C)}{\sigma^2_{\min}(A)(1 - \epsilon^2)}}, \quad g_j = \frac{\sqrt{2j}}{(\ell - 1) \kappa \sqrt{p}} - \max\{\frac{\kappa}{\omega_{\text{neg}}}, \frac{1}{\omega_{\text{ir}}}\} r, \quad \text{for } j = 1, 2,
\]
where \( C \) and \( \kappa \) are two constants.

**Proof.** Using \( \|E\| < \epsilon \phi_p \) and Theorem 2.2, we have
\[
\delta = \phi_{p+1} \leq \|E\| \leq \epsilon \phi_p, \quad \|\sin \Theta(\mathcal{R}(A), \mathcal{R}(\tilde{V}_1))\| \leq \frac{\|\tilde{V}_1^\top EE_2\|}{\phi_p} \leq \frac{\|E\|}{\phi_p} \leq \epsilon.
\]
Let \( [V_1, V_2] \) be an orthogonal matrix such that \( \mathcal{R}(V_1) = \mathcal{R}(A), \mathcal{R}(V_2) = \mathcal{N}(A^\top) \). Then we can write \( \tilde{V}_1 = V_1 T_c + V_2 T_s \), where \( \begin{bmatrix} T_c \\ T_s \end{bmatrix} \) is orthonormal, \( \|T_s\| = \|\sin \Theta(V_1, \tilde{V}_1)\| \leq \epsilon, \sigma_{\text{min}}(T_c) = \sqrt{1 - \|\sin \Theta(V_1, \tilde{V}_1)\|^2} \geq \sqrt{1 - \epsilon^2} \). Therefore, \( T_c \) is nonsingular. Let \( B_i = V_1^\top C_i V_1, \tilde{B}_i = \tilde{V}_1^\top \tilde{C}_i \tilde{V}_1 \). By calculations, we have
\[
\|\tilde{B}_i - T_c^\top B_c T_c\|_F = \|\tilde{V}_1^\top (C_i + E_i) \tilde{V}_1 - T_c^\top V_1^\top C_i V_1 T_c\|_F \\
\leq \|\tilde{V}_1^\top C_i \tilde{V}_1 - T_c^\top V_1^\top C_i V_1 T_c + \tilde{V}_1^\top E_i \tilde{V}_1\|_F \\
\leq \|T_c^\top V_1^\top C_i V_1 T_c + T_s^\top V_2^\top C_i V_1 T_c + T_s^\top V_2^\top C_i V_2 T_s + \tilde{V}_1^\top E_i \tilde{V}_1\|_F \\
= \|E_i\|_F,
\]
where (a) uses \( \tilde{V}_1 = V_1 T_c + V_2 T_s \), (b) uses \( A^\top V_2 = 0 \) (by Theorem 2.1).

On one hand, let \( Z = T_c^\top V_1^\top A \), using (1), we have
\[
T_c^\top B_c T_c = T_c^\top V_1^\top A \Sigma_i A^\top V_1 T_c = Z \Sigma_i Z^\top.
\]
On the other hand, on output of Algorithm 4, it holds that

$$\sum_{i=1}^{m} \| \tilde{B}_i - \tilde{Z} \tilde{\Sigma}_i \tilde{Z}^\top \|^2_F \leq C \delta^2 = C \hat{\phi}_p^2 \leq C \hat{\phi}_p^2 \epsilon^2,$$

(45)

where \( \tilde{\Sigma}_i = \text{diag}(\Sigma_{i1}, \ldots, \Sigma_{i\ell}) \)'s are all \( \hat{\tau}_p = (\hat{\rho}_1, \ldots, \hat{\rho}_\ell) \)-block diagonal, and for each \( 1 \leq j \leq \ell \), \( \{ \Sigma_{ij} \}_{i=1}^{m} \) does not have \( \delta \)-block diagonalizer.

Using (43), (44) and (45), we have

$$\sum_{i=1}^{m} \| Z \Sigma_i Z^\top - \tilde{Z} \tilde{\Sigma}_i \tilde{Z}^\top \|^2_F \leq 2 \sum_{i=1}^{m} (\| Z \Sigma_i Z^\top - \tilde{B}_i \|^2_F + \| \tilde{B}_i - \tilde{Z} \tilde{\Sigma}_i \tilde{Z}^\top \|^2_F)$$

$$\leq 2 \sum_{i=1}^{m} (\| E_i \|^2_F + C \hat{\phi}_p^2 \epsilon^2) = \| E \|^2_F + 2C \hat{\phi}_p^2 \epsilon^2 \leq d \| E \|^2 + 2C \hat{\phi}_p^2 \epsilon^2$$

$$\leq (d + 2C) \hat{\phi}_p^2 \epsilon^2.$$

(46)

As \( T_e \) is nonsingular, \( A \) has full column rank, \( \mathcal{R}(V_1) = \mathcal{R}(A) \), we know that \( Z \) is nonsingular. \( \tilde{Z} \) is also nonsingular since it is the product of a sequence of nonsingular matrices. Then we may let \( Y = Z^\top \tilde{Z}^{-\top}, \Gamma = Y \tilde{Y}^{-1} = \frac{1}{\epsilon} \text{diag}(\gamma_1 I_{\hat{p}_1}, \ldots, \gamma_\ell I_{\hat{p}_\ell}) Y^{-1} \), where \( \gamma_j = -1 + \frac{2(j-1)}{\ell-1} \) for \( j = 1, \ldots, \ell \), \( \varrho = \| \text{diag}(\gamma_1 I_{\hat{p}_1}, \ldots, \gamma_\ell I_{\hat{p}_\ell}) Y^{-1} \|_F \). It follows that

$$\varrho = \varrho \| \Gamma \|_F = \| Y \text{diag}(\gamma_1 I_{\hat{p}_1}, \ldots, \gamma_\ell I_{\hat{p}_\ell}) Y^{-1} \|_F \leq \kappa(Y) \sqrt{\sum_{j=1}^{\ell} \hat{\rho}_j \gamma_j^2} \leq \kappa(Y) \sqrt{\bar{p}}.$$

(47)

Denote \( F_i = Z \Sigma_i Z^\top - \tilde{Z} \tilde{\Sigma}_i \tilde{Z}^\top \) for all \( i \). Direct calculations give rise to

$$\sum_{i=1}^{m} \| \Sigma_i \Gamma - \Gamma^\top \Sigma_i \|^2_F = \sum_{i=1}^{m} \| Z^{-1}(Z \Sigma_i Z^\top \tilde{Z}^{-\top} \Gamma \tilde{Z}^\top - \tilde{Z} \Gamma \tilde{Z}^{-\top} Z \Sigma_i Z^\top) Z^{-\top} \|^2_F$$

$$= \sum_{i=1}^{m} \| Z^{-1}(\tilde{Z} \Sigma_i \tilde{Z}^\top + F_i) \tilde{Z}^{-\top} \Gamma \tilde{Z}^\top - \tilde{Z} \Gamma \tilde{Z}^{-\top} (\tilde{Z} \Sigma_i \tilde{Z}^\top + F_i) Z^{-\top} \|^2_F$$

$$= \sum_{i=1}^{m} \| Z^{-1} F_i Z^{-\top} \Gamma - \Gamma^\top Z^{-1} F_i Z^{-\top} \|^2_F$$

$$\leq 2 \| \Gamma \|^2_F \sum_{i=1}^{m} \| Z^{-1} F_i Z^{-\top} \|^2 \leq \frac{2(d + 2C) \hat{\phi}_p^2 \epsilon^2}{\sigma_{\min}^2(Z)} \leq r^2,$$

(48)

where (a) uses (46), \( \| \Gamma \|_F = 1 \) and (b) uses the definition of \( r \) and \( \sigma_{\min}(T_e) \geq \sqrt{1 - \epsilon^2} \).

Partition \( \Gamma = [\Gamma_{jk}] \) with \( \Gamma_{jk} \in \mathbb{R}^{p_j \times k_j} \), and recall (4) and (5). Using (48), we get

$$\sum_{j=1}^{\ell} \| G_{jj} \text{vec}(\Gamma_{jj}) \|^2 + \sum_{1 \leq j < k \leq \ell} \| G_{jk} \text{vec}(\Gamma_{kk}^{\top}) \|^2 = \sum_{i=1}^{m} \| \Sigma_i \Gamma - \Gamma^\top \Sigma_i \|^2_F \leq r^2.$$

(49)

Let \( r_{jj} = G_{jj} \text{vec}(\Gamma_{jj}) \), the eigenvalues of \( \Gamma_{jj} \) be \( \gamma_j, \ldots, \gamma_{j p_j} \), for \( j = 1, \ldots, \ell \). Then we have

$$\Gamma_{jj} = \hat{\Gamma}_{jj} + \hat{\gamma}_j I_{p_j},$$

where \( \hat{\Gamma}_{jj} = \text{reshape}(G_{jj} r_{jj}, p_j, p_j) \). It follows that

$$\sum_{k=1}^{p_j} | \gamma_{jk} - \hat{\gamma}_j |^2 \leq \| \hat{\Gamma}_{jj} \|^2_F \leq \frac{r_{jj}^2}{\omega_{ir}^2}.$$

(50)
Let \( r_{jk} = G_{jk} \left[ \begin{smallmatrix} \vec(\Gamma_{jk}) \\ -\vec(\Gamma_{kj}) \end{smallmatrix} \right] \) for \( 1 \leq j < k \leq \ell \). Then we have

\[
\|\Gamma_{jk}\|_F^2 + \|\Gamma_{kj}\|_F^2 \leq \|G_{jk}r_{jk}\|_F^2 \leq \frac{\|r_{jk}\|_F^2}{\omega_{neq}^2}.
\] (51)

Let \( \mu_{jk} = \text{argmin}_{\gamma \in (\gamma_1, \ldots, \gamma_{\ell})} \| r_{jk} - \gamma_{jk} \|_F^2 \). By [28, Remark 3.3, (b)], it holds that

\[
\sum_{j=1}^\ell \sum_{k=1}^{p_j} \left( \frac{\mu_{jk}}{\varrho} - \gamma_{jk} \right)^2 \leq \kappa^2(Y) \sum_{j<k} (\|\Gamma_{jk}\|_F^2 + \|\Gamma_{kj}\|_F^2)
\] (52)

Using (50), (51) and (52), we have

\[
\sum_{j=1}^\ell \sum_{k=1}^{p_j} \left( \frac{\mu_{jk}}{\varrho} - \gamma_{jk} \right)^2 \leq \sum_{j=1}^\ell \sum_{k=1}^{p_j} \left( \frac{\mu_{jk}}{\varrho} - \gamma_{jk} \right)^2 + \sum_{j=1}^\ell \sum_{k=1}^{p_j} |\gamma_{jk} - \hat{\gamma}_{jk}|^2
\]

\[
\leq \kappa^2(Y) \sum_{j<k} \|r_{jk}\|_F^2 + \frac{1}{\omega_{ir}} \sum_j \|r_{jj}\|_F^2 \leq \max \left\{ \frac{\kappa^2(Y)}{\omega_{neq}}, \frac{1}{\omega_{ir}} \right\} r^2.
\] (53)

Now we declare that for any \( j \), it holds that \( \mu_{j1} = \mu_{j2} = \cdots = \mu_{jp_j} \). Because otherwise, without loss of generality, say \( \mu_{j1} = \gamma_1, \mu_{j2} = \gamma_2 \), and they corresponds to \( \hat{\gamma}_j \), then we have

\[
\sum_{j=1}^\ell \sum_{k=1}^{p_j} \left( \frac{\mu_{jk}}{\varrho} - \gamma_{jk} \right)^2 \geq \left| \frac{\gamma_1}{\varrho} - \hat{\gamma}_j \right|^2 + \left| \frac{\gamma_2}{\varrho} - \hat{\gamma}_j \right|^2 \geq \frac{\left| \gamma_1 - \gamma_2 \right|^2}{2\varrho^2} \geq \frac{2}{(\ell - 1)^2 \kappa^2(Y) p},
\] (54)

where the last inequality uses the definition of \( \gamma_j \) and also (47). Combining (53) and (54), we get \( \max \left\{ \frac{\kappa(Y)}{\omega_{neq}}, \frac{1}{\omega_{ir}} \right\} r^2 \geq \frac{1}{(\ell - 1)^2 \kappa(Y)} \sqrt{\frac{2}{\varrho^2}} \), which contradicts to the assumption that \( g_1 > 0 \). Therefore, \( \hat{\ell} = \ell \), and there exists a permutation \( \{1', 2', \ldots, \ell'\} \) of \( \{1, 2, \ldots, \ell\} \) such that \( p_j = p_j \).

Without loss of generality, let \( j' = j \) for all \( j = 1, \ldots, \ell \). Let \( Y^{-\top} = [Y_{jk}] \),

\[
R = [R_{jk}] = \text{OffBlkdiag}_{p_j}(\text{OffBlkdiag}_{p_j}(\Gamma^{\top}Y^{-\top}) + \text{diag}(\Gamma_{11} - \hat{\gamma}_1 I, \ldots, \Gamma_{\hat{\ell} \ell} - \hat{\gamma}_{\hat{\ell}} I) \text{OffBlkdiag}_{p_j}(Y^{-\top}) - \hat{\gamma}_j I) \text{OffBlkdiag}_{p_j}(Y^{-\top}) - \hat{\gamma}_j I) \text{OffBlkdiag}_{p_j}(Y^{-\top})\frac{1}{\varrho} \text{diag}(\gamma_1 I, \ldots, \gamma_\ell I) = -R.
\]

Then it follows that \( (\hat{\gamma}_j - \frac{2\varrho}{\sqrt{\varrho}}) Y_{jk} = R_{jk} \) for \( j \neq k \). By calculations, we have

\[
\|Y_{jk}\|_F = \frac{\|R_{jk}\|_F}{|\gamma_j - \gamma_k/\varrho|} \leq \frac{\|R_{jk}\|_F}{|\gamma_j - \gamma_k/\varrho|} \leq \frac{\|R_{jk}\|_F}{\frac{2(\ell - 1)}{\varrho}} = \frac{\|R_{jk}\|_F}{g_2},
\]

\[
\|R\|_F \leq \|\text{OffBlkdiag}_{p_j}(\Gamma^{\top})\| \|Y^{-\top}\| + \max_j \|\Gamma_{jj} - \hat{\gamma}_j I\| \|\text{OffBlkdiag}_{p_j}(Y^{-\top})\|_F
\]\n
\[
\leq \|\text{OffBlkdiag}_{p_j}(\Gamma^{\top})\| \|Y^{-\top}\| + \sqrt{\sum_j \|r_{jj}\|_F^2} \|\text{OffBlkdiag}_{p_j}(Y^{-\top})\|_F,
\]

where (a) uses the definition of \( \gamma_j \), (b) uses (47) and (53), (c) uses (50). Therefore,

\[
\|\text{OffBlkdiag}_{p_j}(Y^{-\top})\|_F \leq \frac{\|R\|_F}{g_2}
\]\n
\[
\leq \frac{1}{g_2} \left( \|\text{OffBlkdiag}_{p_j}(\Gamma^{\top})\| \|Y^{-\top}\| + \sqrt{\sum_j \|r_{jj}\|_F^2} \|\text{OffBlkdiag}_{p_j}(Y^{-\top})\|_F \right),
\]

27
and hence
\[ \| \text{OffBlkdiag}_{\tau_p}(Y^{-\top}) \|_F \leq \frac{\| \text{OffBlkdiag}_{\tau_p}(I^\top) \|_F \| Y^{-\top} \|}{g_2 - \frac{\sqrt{\sum_{j=1}^{\ell} \| r_{ij} \|_2^2}}{\omega_{ir}}} \leq \frac{r}{\omega_{ireg}} \| Y^{-1} \|, \] (55)

where the last inequality uses (50) and (51).

Finally, by calculations, we have
\[
\tilde{A} = \tilde{V}_1 \tilde{Z} = (V_1 T_c + V_2 T_s) \tilde{Z} = (V_1 T_c^{-\top} (I - T_s^\top T_s) + V_2 T_s) \tilde{Z} \\
= V_1 T_c^{-\top} Z Y^{-\top} + (-V_1 T_c^{-\top} (T_s^\top T_s) + V_2 T_s) \tilde{Z} \\
= AY^{-\top} + (-V_1 T_c^{-\top} (T_s^\top T_s) + V_2 T_s) \tilde{Z} \\
= A \text{diag}(Y_{11}, \ldots, Y_{\ell\ell}) + A \text{OffBlkdiag}_{\tau_p}(Y^{-\top}) + (-V_1 T_c^{-\top} (T_s^\top T_s) + V_2 T_s) \tilde{Z},
\]

and it follows that
\[
\| \tilde{A} - A \text{diag}(Y_{11}, \ldots, Y_{\ell\ell}) \|_F \leq \| A \| \| \text{OffBlkdiag}_{\tau_p}(Y^{-\top}) \|_F + (\| T_c^{-\top} T_s^\top T_s \| + \| T_s \|) \| \tilde{Z} \|_F \\
\leq \frac{r}{\omega_{ireg}} \| Y^{-\top} \| + \frac{r \epsilon^2}{\sqrt{1 - \epsilon^2}} + \epsilon \| A \|_F.
\]

The proof is completed. \qed