Supplementary Materials

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 (Stewart and Sun, 1990, 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 $s$

The following lemma gives some fundamental results for $s$

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

$$\| \sin \Theta(U, V) \| = \|U_c^T V\| = \|U^T V_c\|.$$ 

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 $U_{i=1}^3 \{z \in \mathbb{C} \mid |z - t_i| \leq r\}$, 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 $A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ -q & -p & 0 \end{bmatrix}$, $\tilde{t}_i, \tilde{t}_j, \tilde{t}_3$ 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 Kahan et al. (1982, Theorem 8), we know that for each $\tilde{t}_i$, there exists a $t_j$ such that

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

(11)

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

6 Proofs

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 $R(A) = N(A^T)^\perp = R(C^\top)$.

Proof. Using (1), for any $v \in N(A^T)$, we have $C_i x = A \Sigma_i A^T x = 0$, similarly, $C_i^T x = 0$. Therefore, $N(A^T) \subset N(C)$. Next, we show $\sigma_p(C) > 0$ by contradiction. If $\sigma_p(C) = 0$, there exists a nonzero vector $v \notin N(A^T)$ such that $C_i v = 0$. Let $w = A^T v$, 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 $C_i v = 0$ that

$$0 = C_i v = A \Sigma_i A^T v = A \Sigma_i w = A \begin{bmatrix} \Sigma_i^{(1)} w_1 \\ \vdots \\ \Sigma_i^{(\ell)} w_\ell \end{bmatrix}.$$ 

(12)
Therefore, we have $\Sigma_i^{(i)} w_i = 0$ for all $i$. Similarly, $w_i^T \Sigma_i^{(i)} = 0$ for all $i$. Let $w_i^\perp \in \mathbb{R}^{p_i \times (p_i - 1)}$ be such that $[w_1, w_i^\perp]$ be nonsingular, then

$$[w_1, w_i^\perp]^T \Sigma_i^{(i)} [w_1, w_i^\perp] = \begin{bmatrix} 0 & 0 \\ 0 & \ast \end{bmatrix}, \quad \text{for } i = 1, \ldots, m,$$

i.e., $C_1 = \{\Sigma_i^{(i)}\}_{i=1}^m$ 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 $\dim(\mathcal{N}(A^T)) = d - p$ and $\mathcal{N}(A^T) \subset \mathcal{N}(C)$, we have $\mathcal{N}(A^T) = \mathcal{N}(C)$. Then it follows that

$$\beta(A) = \mathcal{N}(A^T)^\perp = \mathcal{N}(C)^\perp = \mathcal{A}(C^T)$$

This completes the proof. $\square$

### 6.2 Proof of Theorem 2.2

**Theorem 2.2.** Let $(\tau_p, A)$ be a solution to BJBDP for $C$. Let the columns of $V_2$ be an orthonormal basis for $\mathcal{N}(A^T)$, $\phi_1 \geq \cdots \geq \phi_d$ 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 $\tilde{U}_1 = [\tilde{u}_1, \ldots, \tilde{u}_p]$, $\tilde{V}_1 = [\tilde{v}_1, \ldots, \tilde{v}_p]$, 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

$$\|\sin(\beta(A), \beta(\tilde{V}_1))\| \leq \frac{\|\tilde{U}_1^T E V_2\|}{\phi_p}.$$  

**Proof.** First, by Theorem 2.1, we know that $\phi_{p+1} = \cdots = \phi_d = 0$. 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 $\|E\| < \frac{\phi_p}{2}$, we have $\tilde{\phi}_p \geq \phi_p - \|E\| > \frac{\phi_p}{2} > \|E\| \geq \tilde{\phi}_{p+1}$. Thus, $\beta(\tilde{V}_1)$ is well defined. By calculations, we have

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

where (a) uses $\text{diag}(\tilde{\phi}_1, \ldots, \tilde{\phi}_p) \tilde{V}_1^T = \tilde{U}_1^T \tilde{C}$, (b) uses $CV_2 = 0$. Then using Lemma 5.2, we get

$$\|\sin(\beta(A), \beta(\tilde{V}_1))\| = \|\tilde{V}_1^T V_2\| = \|\text{diag}(\tilde{\phi}_1, \ldots, \tilde{\phi}_p)^{-1} \tilde{U}_1^T E V_2\| \leq \frac{\|\tilde{U}_1^T E V_2\|}{\tilde{\phi}_p}.$$  

The proof is completed. $\square$

### 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 $V_1^T V_1 = I_p$, $\beta(V_1) = \beta(C^T)$. Denote $B_i = V_1^T C_i V_1$, $B = \{B_i\}_{i=1}^m$. Then $C_i$’s can be factorized as in (1) with $\beta(A) = \beta(C^T)$ 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 $j \neq k$.  

Yunfeng Cai, Ping Li
Proof. (⇒) (Sufficiency) Let $W = A^T V_1$. Since $\mathcal{H}(C_i^T) = \mathcal{H}(A) = \mathcal{H}(V_1)$, and $V_1$, $A$ both have full column rank, we know that $W$ is nonsingular. Let

$$X = W^{-1} \Gamma W = W^{-1} \text{diag}(\gamma_1 I_{p_1}, \ldots, \gamma_{\ell} I_{p_{\ell}}) W,$$

(15)

where $\gamma_1, \ldots, \gamma_{\ell}$ be $\ell$ distinct real numbers. For all $1 \leq i \leq m$, we have

$$B_i X \overset{(a)}{=} W^T \Sigma_i WW^{-1} \Gamma W = W^T \Sigma_i \Gamma W = W^T \Gamma W^{-T} W^T \Sigma_i W \overset{(b)}{=} X^T B_i,$$

where both (a) and (b) use $W = A^T V_1$, (1) and (15). Therefore, $X \in \mathcal{N}(B)$, and it is of form (3).

(⇐) (Necessity) Substituting (3) into $B_i X = X^T B_i$, we get

$$B_i Y \text{diag}(X_{11}, \ldots, X_{1\ell}) Y^{-1} = Y^{-T} \text{diag}(X_{11}^T, \ldots, X_{1\ell}^T) Y^T B_i,$$

(16)

Partition $Y^T B_i Y = [\Sigma_i^{(j,k)}]$ with $\Sigma_i^{(j,k)} \in \mathbb{R}^{p_j \times p_k}$, then it follows from (16) that

$$\Sigma_i^{(j,k)} X_{kk} = X_{jj}^T \Sigma_i^{(j,k)}, \quad \text{for} \quad j, k = 1, 2, \ldots, \ell.$$

(17)

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

$$V_i^T C_i V_1 = B_i = Y^{-T} \Sigma_i Y^{-1},$$

(18)

where $\Sigma_i = \text{diag}(\Sigma_i^{(1,1)}, \ldots, \Sigma_i^{(\ell,\ell)})$. Using $\mathcal{H}(C_i) = \mathcal{H}(V_1)$, we know that $\mathcal{R}(C_i) \subset \mathcal{H}(V_1)$ and $\mathcal{R}(C_i^T) \subset \mathcal{H}(V_1)$. Then it follows from (18) that

$$C_i = V_i Y^{-T} \Sigma_i Y^{-1} V_1^T.$$

Set $A = V_1 Y^{-T}$, the conclusion follows immediately. \hfill \Box

6.4 Proof of Theorem 2.4

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

Proof. (⇒) (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

$$\Gamma_{jj} = W_j \text{diag}(\Gamma_{jj}^{(1)}, \Gamma_{jj}^{(2)}) W_j^{-1},$$

(19)

where $\Gamma_{jj}^{(1)}$ and $\Gamma_{jj}^{(2)}$ are two real matrices and $\lambda(\Gamma_{jj}^{(1)}) \cap \lambda(\Gamma_{jj}^{(2)}) = \emptyset$. Using $\text{vec}(\Gamma_{jj}) \in \mathcal{N}(G_{jj})$, we have

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

(20)

Substituting (19) into (20), we get

$$\Sigma_i^{(j,j)} \text{diag}(\Gamma_{jj}^{(1)}, \Gamma_{jj}^{(2)}) - \text{diag}(\Gamma_{jj}^{(1)}, \Gamma_{jj}^{(2)})^T \Sigma_i^{(j,j)} = 0, \quad \text{for} \quad 1 \leq i \leq m.$$  

(21)

where $\Sigma_i^{(j,j)} = W_j^T \Sigma_i^{(j,j)} W_j$. Similar to the proof of necessity for Theorem 2.3, using $\lambda(\Gamma_{jj}^{(1)}) \cap \lambda(\Gamma_{jj}^{(2)}) = \emptyset$, 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.,

$$\begin{bmatrix} \Sigma_i^{(j,j)} & 0 \\ 0 & \Sigma_i^{(k,k)} \end{bmatrix} \begin{bmatrix} 0 & \Gamma_{jk} \\ \Gamma_{kj} & 0 \end{bmatrix} - \begin{bmatrix} 0 & \Gamma_{jk}^T \\ \Gamma_{kj} & 0 \end{bmatrix} \begin{bmatrix} \Sigma_i^{(j,j)} & 0 \\ 0 & \Sigma_i^{(k,k)} \end{bmatrix} = 0.$$  

(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

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

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 \( W_{jk}^T \begin{bmatrix} \Sigma_i^{(jj)} & 0 \\ 0 & \Sigma_i^{(kk)} \end{bmatrix} W_{jk} \) for all \( 1 \leq i \leq m \) are all block diagonal. For the ease of notation, let \( j = 1, k = 2 \). Denote \( \hat{A} = A \text{diag}(W_{12}^T, I_{p_1}, \ldots, I_{p_k}) \). We know that \( A, \hat{A} \) are not equivalent since \( W_{12} \) is not \((p_1, p_2)\)-block diagonal. This contradicts with the assumption that BJBDP for \( C \) is uniquely \( \tau_p \)-block-diagonalizable, completing the proof of sufficiency.

(\( \Leftarrow \)) (Necessity) Let \((\tau_p, A)\) and \((\hat{\tau}_p, \hat{A})\) be two solutions to the BJBDP for \( C \), i.e., it holds that

\[
C_i = A \Sigma_i A^T = \hat{A} \hat{\Sigma}_i \hat{A}^T,
\]

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

Let \( \tau_p = (p_1, \ldots, p_\ell) \), \( \hat{\tau}_p = (\hat{p}_1, \ldots, \hat{p}_\ell) \). As \((\tau_p, A)\) and \((\hat{\tau}_p, \hat{W})\) are both solutions, it holds that \( \ell = \hat{\ell} \). By Theorem 2.1, we know that \( \hat{\mathcal{R}}(C^T) = \mathcal{R}(A) = \hat{\mathcal{R}}(A) \). Since \( A \) and \( \hat{A} \) are both of full column rank, we know that \( p = \hat{p} \) and there exists nonsingular matrix \( Z \) such that \( \hat{A} = AY^{-T} \). Then it follows from (24) that

\[
\hat{\Sigma}_i = Y^T \Sigma_i Y \quad \text{for} \quad 1 \leq i \leq m.
\]

Let \( \Gamma = Y \text{diag}(\gamma_1 I_{\hat{p}_1}, \ldots, \gamma_\ell I_{\hat{p}_\ell})Y^{-1} \), where \( \gamma_1, \ldots, \gamma_\ell \) are distinct real numbers. Using (25), we have

\[
\Sigma_i \Gamma = Y^{-T} (Y^T \Sigma_i Y) \text{diag}(\gamma_j I_{\hat{p}_j})Y^{-1} = Y^{-T} \text{diag}(\gamma_j I_{\hat{p}_j})(Y^T \Sigma_i Y)Y^{-1} = \Gamma^T \Sigma_i,
\]

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

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 \( \tau_p \)-block diagonal; using (P1), \( \Gamma = Y \text{diag}(\gamma_j I_{\hat{p}_j})Y^{-1} \) and \( \cup_{j=1}^\ell \lambda(\Gamma_{jj}) = \lambda(\Gamma) \), we know that \( \lambda(\Gamma_{k_j k_j}) = \lambda(\gamma_j I_{\hat{p}_j}) \) 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 \( \hat{\tau}_p = \tau_p \Pi_\ell \). Let \( \Pi \in \mathbb{R}^{p \times p} \) be the permutation matrix associated with \( \Pi_\ell \). Then

\[
\text{diag}(\gamma_1 I_{p_{k_1}}, \ldots, \gamma_\ell I_{p_{k_\ell}}) = \Pi^T \text{diag}(\gamma_1 I_{p_1}, \ldots, \gamma_\ell I_{p_\ell}) \Pi.
\]

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

\[
\text{diag}(\Gamma_1, \ldots, \Gamma_\ell) = Y \Pi^T \text{diag}(\gamma_1 I_{p_1}, \ldots, \gamma_\ell I_{p_\ell})(Y \Pi^T)^{-1}.
\]

Noticing that the columns of \( Y \Pi^T \) are eigenvectors of \( \Gamma \), we know that \( Y \Pi^T \) is \( \tau_p \)-block-diagonal. Therefore, we can rewrite \( \hat{A} = AY^{-T} \) as \( \hat{A} = A(Y \Pi^T)^{-1} \Pi \), in which \( (Y \Pi^T)^{-1} \) is \( \tau_p \)-block-diagonal, \( \Pi \) is the permutation matrix associated with \( \Pi_\ell \). So, \((\tau_p, A)\) and \((\hat{\tau}_p, \hat{A})\) are equivalent. The proof is completed. \( \square \)
6.5 Proof of Theorem 2.5

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

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

(II) If \( D \) has a nontrivial diagonalizer, then \( \text{OPT}(D) \) 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 of (I) via its the contrapositive. If the feasible set of \( \text{OPT}(D) \) is not empty, then it has a solution \( X_* \). Using \( \text{tr}(X_*) = 0 \), \( \text{tr}(X_*^T) = q > 0 \), 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, \( D \) 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 \( X_* \in \mathcal{M}(D) \), 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

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

Since \( \mu > 0 \) has full column rank, we know that \( \sum_{i=1}^m |z^H D_i z|^2 = 0 \). Therefore, \( \gamma \) is real. And 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

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

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 \( \sum_j \gamma_j = 0 \) and \( \sum_j \gamma_j^2 = q \). 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 \( t_1 = t_2 \) 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

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

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

\[
\text{tr}(X_*^4) = \|u\|^2 \|v\|^2 / q \geq (u^T v)^2 / q = (q_1 t_1^2 + q_2 t_2^2 + q_3 t_3^2) / q = q,
\]
and the equality holds if and only if $u$ and $v$ are co-linear. Using the first two equalities of (31), $q_1$, $q_2$, $q_3$ cannot 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 $X_*$ 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 $\mathcal{N}(D)$. If $D$ has a nontrivial diagonalizer, then there exists a matrix $Z$ such that $D_i = Z\Phi_iZ^\top$, where $\Phi_i$’s are $\tau_q = (q_1, q_2)$-block diagonal. Since $D$ has full column rank, $Z$ is nonsingular. Let $X = Z^{-T}\text{diag}(\sqrt{q_1}/q_1^2, \sqrt{q_2}/q_2^2)Z^\top$. It is easy to see that $\text{tr}(X) = 0$, $\text{tr}(X^2) = 1$ and $X \in \mathcal{N}(D)$. 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 $\gamma_1 = \sqrt{\frac{q_2}{q_1}}$, $\gamma_2 = -\sqrt{\frac{q_2}{q_1}}$. 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{Proof of Theorem 2.6}

\textbf{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.

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

\begin{enumerate}[(1)]
    \item 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, $\text{card}(\hat{\tau}) = 2$.
    \item Denote $\hat{Z}^{-1}Di\hat{Z}^{-\top} = \text{diag}(D_i^{(1)}, D_i^{(2)})$, $D^{(1)} = \{D_i^{(1)}\}$ and $D^{(2)} = \{D_i^{(2)}\}$. Then $D^{(1)}$ and $D^{(2)}$ both have full column rank.
\end{enumerate}

Fact (1) is because when $\text{card}(\hat{\tau}) > 1$, $D$ can be block diagonalized. Fact (2) is due to the fact $\hat{Z}$ is nonsingular and $\hat{Z}^{-1}Di\hat{Z}^{-\top} = \text{diag}(D_i^{(1)}, D_i^{(2)})$.

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

$$\hat{\tau}_p = (\hat{p}_1, \ldots, \hat{p}_\ell), \quad C_i = \hat{A}\hat{\Sigma}_i\hat{A}^\top = \hat{A}\text{diag}(\hat{\Sigma}_i^{(1)}, \ldots, \hat{\Sigma}_i^{(\ell)})\hat{A}^\top, \quad i = 1, \ldots, m,$$

(32)

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

Using (1) and (32), we have

$$B_i = V_1^\top\hat{A}\hat{\Sigma}_i\hat{A}^\top V_1 = \hat{Z}\hat{\Sigma}_i\hat{Z}^\top = V_1^\top\hat{A}\Sigma_i\hat{A}^\top V_1 = Z\Sigma_iZ^\top.$$

(33)

where $\hat{Z} = V_1^\top\hat{A}$, $Z = V_1^\top A$. By Theorem 2.1, we know that $A(V_1) = \mathcal{R}(C_i) = \mathcal{R}(A)$. By the construction of $\hat{A}$, we know $A(V_1) = \mathcal{R}(\hat{A})$. 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

$$\hat{\Sigma}_i = Y^\top\Sigma_iY, \quad \text{for } 1 \leq i \leq m.$$

(34)
where \( Y = Z^\top \hat{Z}^{-\top} \). Let \( \Gamma = Y \operatorname{diag}(\gamma_1 I_{\tilde{p}_1}, \ldots, \gamma_\ell I_{\tilde{p}_\ell}) Y^{-1} \), where \( \gamma_1, \ldots, \gamma_\ell \) are distinct real numbers. Using (34), we have

\[
\Sigma_i \Gamma = Y^{-\top} (Y^\top \Sigma_i Y) \operatorname{diag}(\gamma_j I_{\tilde{p}_j}) Y^{-1} = Y^{-\top} \operatorname{diag}(\gamma_j I_{\tilde{p}_j})(Y^\top \Sigma_i Y) Y^{-1} = \Gamma^\top \Sigma_i, 
\]

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

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 \( \tau_p \)-block diagonal; using (P1), \( \Gamma = Y \operatorname{diag}(\gamma_j I_{\tilde{p}_j}) Y^{-1} \) 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 \( \gamma_k \). Without loss of generality, let \( \Gamma_{11}, \Gamma_{22} \) correspond to \( \gamma_1 \), the remaining blocks correspond to other \( \gamma_k \)'s. Then using \( \Gamma = Y \operatorname{diag}(\gamma_1 I_{\tilde{p}_1}, \ldots, \gamma_\ell I_{\tilde{p}_\ell}) Y^{-1} \), we know that \( Y = \operatorname{diag}(Y_{11}, Y_{22}) \), where \( Y_{11} \in \mathbb{R}^{\tilde{p}_1 \times \tilde{p}_1} \) and \( \tilde{p}_1 = p_1 + p_2 \). Using \( Y = Z^\top \hat{Z}^{-\top} \) and (35), we get

\[
\hat{\Sigma}_i = Y^\top \Sigma_i Y = \operatorname{diag}(Y_{11}, Y_{22})^\top \Sigma_i \operatorname{diag}(Y_{11}, Y_{22}), \quad \text{for} \ 1 \leq i \leq m.
\]

Therefore, we have

\[
\hat{\Sigma}_i^{(11)} = Y_{11}^\top \operatorname{diag}(\Sigma_1^{(11)}, \Sigma_1^{(22)}) Y_{11}, \quad \text{for} \ 1 \leq i \leq m,
\]

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

### 6.7 Proof of Theorem 2.7

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

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

(II) If \( \widetilde{D} \) has a nontrivial \( \delta \)-diagonalizer, then \( \operatorname{OPT}(\widetilde{D}, \delta) \) has a solution \( X_\star \). In addition, assume

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

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

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

**Proof.** First, we show of (I) via its the contrapositive. If the feasible set of \( \operatorname{OPT}(\widetilde{D}, \delta) \) is not empty, then \( \operatorname{OPT}(\widetilde{D}, \delta) \) has a solution \( X_\star \), which can be factorized into \( X_\star = Y \operatorname{diag}(\Gamma_1, \Gamma_2) Y^{-1} \) (since \( \text{tr}(X_\star) = 0 \) and \( \text{tr}(X_\star^2) = q \)), 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 \( Z = Y^{-\top}, \Phi_1 = \operatorname{diag}(Y_1^\top \widetilde{D}_1 Y_1, Y_2^\top \widetilde{D}_2 Y_2), \)

\[
g = \min \frac{\|\Gamma_1^\top X_\star - X_\star \Gamma_2^\top\|^F}{\|X_\star\|^F} \quad \text{and} \quad \kappa = \kappa_2(Y) = \frac{\sigma_{\max}(Y)}{\sigma_{\min}(Y)}. \]

By calculations, we have

\[
\|X_\star\|^2 = \text{tr}(Y^{-\top} \operatorname{diag}(\Gamma_1^\top, \Gamma_2^\top) Y Y \operatorname{diag}(\Gamma_1, \Gamma_2) Y^{-1}) \\
\leq \|Y\|^2 \text{tr}(Y^{-\top} \operatorname{diag}(\Gamma_1^\top, \Gamma_2^\top) \operatorname{diag}(\Gamma_1, \Gamma_2) Y^{-1}) \\
= \|Y\|^2 \text{tr}(\operatorname{diag}(\Gamma_1, \Gamma_2) Y^{-1} Y^{-\top} \operatorname{diag}(\Gamma_1^\top, \Gamma_2^\top)) \\
\leq \kappa^2 \text{tr}(\operatorname{diag}(\Gamma_1, \Gamma_2) \operatorname{diag}(\Gamma_1^\top, \Gamma_2^\top)) = \kappa^2 \text{tr}(X_\star^2) = \kappa^2 q,
\]

\[ (36) \]
where (a) uses $X_* \in \mathcal{M}_d(\bar{D})$, (b) uses the definition of $g$. Then it follows from (36) and (37) that

$$
\sum_{i=1}^{m} \|\bar{D}_i - Z\Phi_i Z^\top\|_F^2 \leq \frac{\kappa^4 \|X_*\|_F^2}{g^2} \delta^2 \leq \frac{\kappa^6}{g^4} \delta^2.
$$

This completes the proof of (I).

Next, we show (II). If $\bar{D}$ has a nontrivial $\delta$-diagonalizer, then there exists a matrix $Z$ such that $\sum_{i=1}^{m} \|\bar{D}_i - Z\Phi_i Z^\top\|_F^2 \leq \frac{1}{4} \delta^2$ (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 $X = Z^{-\top} \Gamma Z^\top$, where $\Gamma = \text{diag}(\sqrt{\frac{F}{q_1}}, -\sqrt{\frac{F}{q_2}})$.

By calculations, we have

$$
\|\textbf{L}(\bar{D})\vec{X}\|^2 = \sum_{i=1}^{m} \|\bar{D}_i X - X^\top \bar{D}_i\|_F^2 \leq 2 \sum_{i=1}^{m} \|\bar{D}_i - Z\Phi_i Z^\top\|_F \|X - X^\top (\bar{D}_i - Z\Phi_i Z^\top)\|_F^2 \\
\leq 4\|X\|_F^2 \sum_{i=1}^{m} \|\bar{D}_i - Z\Phi_i Z^\top\|_F^2 \leq \|X\|_F^2 \delta^2,
$$

where (a) uses $Z\Phi_i Z^\top X - X^\top Z\Phi_i Z^\top = 0$. Therefore, $\|\textbf{L}(\bar{D})\vec{X}\|_F \leq \frac{\|X\|_F \delta}{\|X\|_F} \leq \delta$. Also note that $\text{tr}(X) = 0$ and $\text{tr}(X^2) = q$, then the feasible set of $\text{opt}(\bar{D}, \delta)$ is nonempty. Consequently, $\text{opt}(\bar{D}, \delta)$ has a solution $X_*$. Let $\gamma$ be an arbitrary eigenvalue of $X_*$, and $z$ be the corresponding unit-length eigenvector. By calculations, we have

$$
\kappa^2 q^2 \delta^2 \geq \delta^2 \|X_*\|_F^2 \geq \|\textbf{L}(\bar{D})\vec{X}\|^2 \geq \sum_{i=1}^{m} \|\bar{D}_i X_* - X_*^\top \bar{D}_i\|_F^2 \\
\geq \sum_{i=1}^{m} \|z^H \bar{D}_i x_* - z^H X_*^\top \bar{D}_i z\|_F^2 = |\gamma - \bar{\gamma}|^2 \sum_{i=1}^{m} |z^H \bar{D}_i z|^2 \geq \mu^2 |\gamma - \bar{\gamma}|^2,
$$

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

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

$$
\text{tr}(X_*) = \sum_{j=1}^{q} \gamma_j = 0, \quad \text{tr}(X_*^2) = \sum_{j=1}^{q} (\gamma_j^2 - \eta_j^2) = q, \quad \text{tr}(X_*^4) = \sum_{j=1}^{q} (\gamma_j^4 + \eta_j^4 - 6\gamma_j^2 \eta_j^2).
$$
Using the method of Lagrange multipliers, we consider
\[ L(\gamma_1, \eta_1, \ldots, \gamma_q, \eta_q; \mu_1, \mu_2) = \sum_{j=1}^{q} (\gamma_j^2 + \eta_j^2 - 6\gamma_j^2\eta_j^2) + \mu_1 \sum_{j=1}^{q} \gamma_j + \mu_2 \left( \sum_{j=1}^{q} (\gamma_j^2 - \eta_j^2) - q \right), \]
where \( \mu_1, \mu_2 \) are Lagrange multipliers. By calculations, we have
\[ \frac{\partial L}{\partial \gamma_j} = 4\gamma_j^3 + 2(\mu_2 - 6\eta_j^2)\gamma_j + \mu_1 = 0. \] (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 \( \gamma_j \subset \bigcup_{i=1}^{3} \{ z \mid |z - t_i| \leq O(\delta) \} \), 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 \( \lambda(X_*) \subset \bigcup_{i=1,2} \text{Rect}_i \).

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

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

Let \( u = [\sqrt{q_1} t_1, \sqrt{q_2} t_2, \sqrt{q_3} t_3]^\top \), \( v = [\sqrt{q_1}, \sqrt{q_2}, \sqrt{q_3}]^\top \). Then we have \( \|u\|^2 + O(\delta) = \text{tr}(X_*^2), \|v\| = \sqrt{q} \). Using Cauchy’s inequality, we get
\[ \text{tr}(X_*^4) + O(\delta) = \|u\|^2 = \|u\|^2 \|v\|^2 / q \geq (u^\top v)^2 / q = (q_1^2 t_1^2 + q_2^2 t_2^2 + q_3^2 t_3^2)^2 / q = q + O(\delta), \]
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 \( \bigcup_{i=1,2,3,q_i \neq 0} \{ z \mid |z - t_i| \leq O(\delta) \} \). Otherwise, \( q_1, q_2 \) and \( q_3 \) are all positive integers. Therefore, \( t_2^2 = t_2^2 = t_3^2 \), 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 \( \bigcup_{i=1,2} \text{Rect}_i \).

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

If \( \tilde{D} \) has a nontrivial \( \delta \)-diagonalizer, then there exists a matrix \( Z \) such that \( \sum_{i=1}^{m} \| \tilde{D}_i - Z \Phi_i Z^\top \|^2_F \leq \frac{1}{4} \delta^2 \), where \( \Phi_i \)'s are all \( \tau_q = (q_1, q_2) \) block diagonal matrices. Let \( X = Z^{-1} \Gamma Z^\top \), where \( \Gamma = \text{diag}(\sqrt{q_1} I_{q_1}, -\sqrt{q_2} I_{q_2}) \). 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 \( \rho_1 = \sqrt{\frac{q_1}{q_1}} + O(\delta), \rho_2 = -\sqrt{\frac{q_2}{q_2}} + O(\delta) \). Then it follows that
\[ \rho_1 - \rho_2 = \sqrt{\frac{q_1}{q_1}} + \sqrt{\frac{q_2}{q_2}} + O(\delta) \geq 2 + O(\delta), \]
completing the proof. \( \square \)
6.8 Proof of Theorem 2.8

**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 $\tilde{C} = \{\tilde{C}_i\}_{i=1}^m = \{C_i + E_i\}_{i=1}^m$ be a perturbed matrix set of $\mathcal{C}$. Denote

$$
\tau_p = (p_1, \ldots, p_t), \quad \tilde{\tau}_p = (\tilde{p}_1, \ldots, \tilde{p}_\ell), \quad A = [A_1, \ldots, A_t], \quad \tilde{A} = [\tilde{A}_1, \ldots, \tilde{A}_t],
$$

where $(\tilde{\tau}_p, \tilde{A})$ is the output of Algorithm 4. Assume $\mathcal{N}(G_{jj}) = \mathcal{R}(\text{vec}(I_{p_j}))$ 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,

$$
\epsilon = \frac{\|E\|}{\phi_p}, \quad r = \frac{\sqrt{2(d + 2C)}}{\sigma^2_{\min}(A)(1 - \epsilon^2)}, \quad g_j = \frac{\sqrt{2j}}{(\ell - 1)\kappa \sqrt{p}} - \max\left\{\kappa \frac{1}{\omega_{\text{sur}}}, \frac{1}{\omega_{\text{ir}}}\right\} r, \quad j = 1, 2,
$$

where $C$ and $\kappa$ are two constants.

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

(II) Further assume $g_2 > \frac{\epsilon}{\omega_{\text{ir}}}$, then there exists a $\tau_p$-block diagonal matrix $D$ such that

$$
\| [\tilde{A}_{1'}, \ldots, \tilde{A}_{\ell'}] - AD \|_F \leq \frac{\epsilon r}{g_2 - \frac{\epsilon}{\omega_{\text{ir}}}} \| A \|_F + (\frac{\epsilon^2}{\sqrt{1 - \epsilon^2}} + \epsilon) \| \tilde{A} \|_F = O(\epsilon),
$$

where $c$ is a constant.

**Proof.** Using $\|E\| < \epsilon \phi_p$ and Theorem 2.2, we have

$$
\delta = \phi_{p+1} \leq \|E\| \leq \epsilon \phi_p, \quad \|\sin \Theta(\mathcal{R}(A), \mathcal{R}(\tilde{V}_1))\| \leq \frac{\|T_1^\top E V_2\|}{\phi_p} \leq \|E\| \leq \epsilon. \tag{42}
$$

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 $[T_c \ T_s]$ is orthonormal, $\|T_s\| \leq \epsilon$, $\sigma_{\min}(T_c) = \sqrt{1 - \|\sin \Theta(V_1, \tilde{V}_1)\|^2} \geq \sqrt{1 - \epsilon^2}$. Therefore, $T_c$ is nonsingular. Let $B_t = V_1^\top C_t V_1, ~ \tilde{B}_t = \tilde{V}_1^\top C_t \tilde{V}_1$. And by calculations, we have

$$
\| \tilde{B}_t - T_c^\top B_t T_c \|_F = \| \tilde{V}_1^\top (C_i + E_i) \tilde{V}_1 - T_c^\top V_1^\top C_i V_1 T_c \|_F
\leq \| \tilde{V}_1^\top C_i \tilde{V}_1 - T_c^\top V_1^\top C_i V_1 T_c + \tilde{V}_1^\top E_i \tilde{V}_1 \|_F
\leq (a) T_c^\top V_1^\top C_i V_2 T_s + T_s^\top V_2^\top C_i V_1 T_c + T_c^\top V_2^\top C_i V_2 T_s + T_s^\top V_1^\top E_i \tilde{V}_1 \|_F
\leq (b) \| E_i \|_F, \tag{43}
$$

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

$$
T_c^\top B_t T_c = T_c^\top V_1^\top A \Sigma_i A^\top V_1 T_c = Z \Sigma_i Z^\top. \tag{44}
$$

On the other hand, on output of Algorithm 4, it holds that

$$
\sum_{i=1}^m \| \tilde{B}_i - \tilde{Z} \Sigma_i \tilde{Z}^\top \|_F^2 \leq C \delta^2 \leq C \phi_p^2 \| E \|_F = C \phi_p^2 \epsilon^2, \tag{45}
$$

where $\Sigma_i = \text{diag}(\Sigma_{i1}, \ldots, \Sigma_{i\ell'})$’s are all $\tilde{\tau}_p = (\tilde{p}_1, \ldots, \tilde{p}_\ell)$-block diagonal, and for each $1 \leq j \leq \tilde{\ell}$, $\{\Sigma_{ij}\}_{i=1}^m$ does not have $\delta$-block diagonalizer.
Using (43), (44) and (45), we have
\[
\sum_{i=1}^{m} \|Z_{i}Z_{i}^{\top} - \hat{Z} \hat{Z}^{\top}\|_{F}^{2} \leq 2 \sum_{i=1}^{m} \left( \|Z_{i}Z_{i}^{\top} - \hat{B}_{i}\|_{F}^{2} + \|\hat{B}_{i} - \hat{Z} \hat{Z}^{\top}\|_{F}^{2} \right)
\leq 2 \left( \sum_{i=1}^{m} \|E_{i}\|_{F}^{2} + C \phi_{p}^{2} \epsilon^{2} \right) = \|E\|_{F}^{2} + 2C \phi_{p}^{2} \epsilon^{2}
\leq (d + 2C) \phi_{p}^{2} \epsilon^{2}. \tag{46}
\]

As \(T_{c}\) is nonsingular, \(A\) has full column rank, \(\mathcal{R}(V_{1}) = \mathcal{R}(A)\), we know that \(Z\) is nonsingular. \(\hat{Z}\) is also nonsingular since it is the product of a sequence of nonsingular matrices. Then we may let \(Y = Z^{\top} \hat{Z}^{\top}\), \(\Gamma = Y \hat{Y}^{-1} = \frac{1}{\hat{\epsilon}} \text{diag}(\gamma_{1} I_{p_{1}}, \ldots, \gamma_{\ell} I_{p_{\ell}})Y^{-1}\), where \(\gamma_{j} = -1 + \frac{2(j-1)}{\ell-1}\) for \(j = 1, \ldots, \ell\), \(\hat{\epsilon} = \|Y \text{diag}(\gamma_{1} I_{p_{1}}, \ldots, \gamma_{\ell} I_{p_{\ell}})Y^{-1}\|_{F}\). It follows
\[
\hat{\epsilon} = \sqrt{\sum_{j=1}^{\ell} \hat{\gamma}_{j}^{2}} \leq \sqrt{\kappa(Y)} \sqrt{\hat{\epsilon}}. \tag{47}
\]

Denote \(F_{i} = Z_{i}Z_{i}^{\top} - \hat{Z} \hat{Z}^{\top}\) for all \(i\). Direct calculations give rise to
\[
\sum_{i=1}^{m} \|\Sigma_{i} \Gamma^{\top} \Sigma_{i} \|_{F}^{2} = \sum_{i=1}^{m} \left( \|Z_{i}Z_{i}^{\top} - \hat{Z} \hat{Z}^{\top}\|^2_{F} - \|Z_{i}Z_{i}^{\top} - \hat{Z} \hat{Z}^{\top}\|_{F}^{2} \right)
= \sum_{i=1}^{m} \|Z_{i}Z_{i}^{\top} - \hat{Z} \hat{Z}^{\top}\|_{F}^{2} - \left( \|Z_{i}Z_{i}^{\top} - \hat{Z} \hat{Z}^{\top}\|_{F}^{2} \right)
\leq 2 \|\Gamma\|_{F}^{2} \sum_{i=1}^{m} \|Z_{i}Z_{i}^{\top} - \hat{Z} \hat{Z}^{\top}\|_{F}^{2}
\leq 2 \|\Gamma\|_{F}^{2} \sum_{i=1}^{m} \|Z_{i}Z_{i}^{\top} - \hat{Z} \hat{Z}^{\top}\|_{F}^{2} \leq \frac{2(d + 2C) \phi_{p}^{2} \epsilon^{2}}{\sigma_{\min}(Z)} \leq r^{2}, \tag{48}
\]

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

Partition \(\Gamma = [\Gamma_{jk}]\) with \(\Gamma_{jk} \in \mathbb{R}_{\ell \times \ell}\), and recall (4) and (5). Using (48), we get
\[
\sum_{j=1}^{\ell} \|G_{jj} \text{vec}(\Gamma_{jj})\|^{2} + \sum_{1 \leq j < k \leq \ell} \left\| G_{jk} \begin{bmatrix} \text{vec}(\Gamma_{jk}) \\ -\text{vec}(\Gamma_{kj}) \end{bmatrix} \right\|^{2} = \sum_{j=1}^{m} \|\Sigma_{j} \Gamma^{\top} \Sigma_{j} \|_{F}^{2} \leq r^{2}. \tag{49}
\]

Let \(r_{jj} = G_{jj} \text{vec}(\Gamma_{jj})\), the eigenvalues of \(\Gamma_{jj}\) be \(\gamma_{j1}, \ldots, \gamma_{jp_{j}}\), for \(j = 1, \ldots, \ell\). Then we have
\[
\Gamma_{jj} = \hat{\Gamma}_{jj} + \hat{\gamma}_{jj} I_{p_{j}},
\]
where \(\hat{\Gamma}_{jj} = \text{reshape}(G_{jj}^{\top} r_{jj}, p_{j}, p_{j})\). And it follows that
\[
\sum_{k=1}^{p_{j}} |\gamma_{jk} - \hat{\gamma}_{jk}|^{2} \leq \|\hat{\Gamma}_{jj}\|_{F}^{2} \leq \frac{\|r_{jj}\|^{2}}{\omega_{ir}^{2}}. \tag{50}
\]

Let \(r_{jk} = G_{jk} \begin{bmatrix} \text{vec}(\Gamma_{jk}) \\ -\text{vec}(\Gamma_{kj}) \end{bmatrix}\), for \(1 \leq j < k \leq \ell\). Then we have
\[
\|\Gamma_{jk}\|_{F}^{2} + \|\Gamma_{kj}\|_{F}^{2} \leq \|G_{jk}^{\top} r_{jk}\|^{2} \leq \frac{\|r_{jk}\|^{2}}{\omega_{neq}^{2}}. \tag{51}
\]
Let \( \mu_{jk} = \arg\min_{\gamma \in \{\gamma_1, \ldots, \gamma_p\}} |\frac{2}{\theta} - \gamma_{jk}| \). By Sun (1996, Remark 3.3, (2)), it holds that

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

(52)

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

\[
\sum_{j=1}^{\ell} \sum_{k=1}^{p_j} \left| \frac{\mu_{jk}}{\theta} - \gamma_{jk} \right|^2 \leq \sum_{j=1}^{\ell} \sum_{k=1}^{p_j} \left| \frac{\mu_{jk}}{\theta} - \gamma_{jk} \right|^2 + \sum_{j=1}^{\ell} \sum_{k=1}^{p_j} \left| \gamma_{jk} - \gamma_j \right|^2 
\]

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

(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 correspond to \( \hat{\gamma}_j \), then we have

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

(54)

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

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

\[
R = [R_{jk}] = \text{OffBlkdiag}_{\tau_{\rho}}(\text{OffBlkdiag}_{\tau_{\rho}}(\Gamma^\top Y^{-\top}) + \text{diag}(\Gamma_{11} - \gamma_1 I, \ldots, \Gamma_{\ell \ell} - \gamma_\ell I)) \text{OffBlkdiag}_{\tau_{\rho}}(Y^{-\top}),
\]

where \( Y_{jk}, R_{jk} \in \mathbb{R}^{p_j \times p_k} \). Using \( \Gamma = Y(Y)Y^{-1} = \frac{1}{\theta} Y \text{diag}(\gamma_1 I_{p_1}, \ldots, \gamma_\ell I_{p_\ell})Y^{-1} \), we have \( \Gamma^\top Y^{-\top} = Y^{-\top} \Gamma \), whose off-block diagonal part reads

\[
\text{diag}(\gamma_1 I, \ldots, \gamma_\ell I) \text{OffBlkdiag}_{\tau_{\rho}}(Y^{-\top}) = \text{OffBlkdiag}_{\tau_{\rho}}(Y^{-\top}) \frac{1}{\theta} \text{diag}(\gamma_1 I, \ldots, \gamma_\ell I) = -R.
\]

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

\[
\|Y_{jk}\|_F = \frac{\|R_{jk}\|_F}{|\gamma_j - \frac{2}{\theta}|} \leq \frac{\|R_{jk}\|_F}{|\gamma_j - \frac{2}{\theta}|} \leq \frac{\|R_{jk}\|_F}{\theta (\ell - 1)^2} \leq \frac{1}{\theta} \|\Gamma^\top Y^{-\top} + \frac{1}{\omega_{\text{ir}}} \sqrt{\sum_j \|r_{jj}\|^2} \|\text{OffBlkdiag}_{\tau_{\rho}}(Y^{-\top})\|_F
\]

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

\[
\|\text{OffBlkdiag}_{\tau_{\rho}}(Y^{-\top})\|_F \leq \frac{\|R\|_F}{\theta}
\]

\[
\leq \frac{1}{\theta} \left( \|\text{OffBlkdiag}_{\tau_{\rho}}(\Gamma^\top)\|_F \|Y^{-\top}\| + \frac{1}{\omega_{\text{ir}}} \sqrt{\sum_j \|r_{jj}\|^2} \|\text{OffBlkdiag}_{\tau_{\rho}}(Y^{-\top})\|_F \right),
\]

and hence

\[
\|\text{OffBlkdiag}_{\tau_{\rho}}(Y^{-\top})\|_F \leq \frac{\|\text{OffBlkdiag}_{\tau_{\rho}}(\Gamma^\top)\|_F \|Y^{-\top}\| + \frac{1}{\omega_{\text{ir}}} \sqrt{\sum_j \|r_{jj}\|^2} \|\text{OffBlkdiag}_{\tau_{\rho}}(Y^{-\top})\|_F}{\theta - \frac{r_{\text{neq}}}{\omega_{\text{ir}}}} \leq \frac{r_{\text{neq}}}{\theta (\ell - 1)^2} \leq \frac{r_{\text{neq}}}{\theta (\ell - 1)^2 \kappa(Y)},
\]

(55)
where the last inequality uses (50) and (51).

Finally, by calculations, we have

\[ \hat{A} = \tilde{V}_1 \tilde{Z} = (V_1 T_c + V_2 T_s) \tilde{Z} = (V_1 T_c^{-\top} (I - T_s^\top T_s) + V_2 T_s) \tilde{Z} \]
\[ = V_1 T_c^{-\top} Y Y^- \top + (-V_1 T_c^{-\top} (T_s^\top T_s) + V_2 T_s) \tilde{Z} \]
\[ = A Y Y^- \top + (-V_1 T_c^{-\top} (T_s^\top T_s) + V_2 T_s) \tilde{Z} \]
\[ = A \text{diag}(Y_{11}, \ldots, Y_{\ell\ell}) + A \text{OffBlkdiag}_\tau (Y^- \top) + (-V_1 T_c^{-\top} (T_s^\top T_s) + V_2 T_s) \tilde{Z}, \]

and it follows that

\[ \| \hat{A} - A \text{diag}(Y_{11}, \ldots, Y_{\ell\ell}) \|_F \leq \| A \| \| \text{OffBlkdiag}_\tau (Y^- \top) \|_F + (\| T_c^{-\top} (T_s^\top T_s) + \| T_s \| ) \| \tilde{Z} \|_F \]
\[ \leq \| A \| \frac{\epsilon}{\omega_{\text{min}}} \| Y^- \top \| + (\epsilon^2 \sqrt{1 - \epsilon^2} + \epsilon) \| A \|_F. \]

The proof is completed.