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, $\widetilde{A} \in \mathbb{C}^{n\times n}$, let $\lambda_1 \le \cdots \le \lambda_n$, $\tilde{\lambda}_1 \le \cdots \le \tilde{\lambda}_n$ be eigenvalues of A, $\widetilde{A}$, respectively. Then*

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

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

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

$$\|\sin\Theta(U, V)\| = \|U_{\mathrm{c}}^\top V\| = \|U^\top V_{\mathrm{c}}\|.$$

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 $\cup_{i=1}^3 \{z \in \mathbb{C} \mid |z - t_i| \le r\}$, 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 $\widetilde{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -q & -p-\epsilon & 0 \end{bmatrix}$. 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

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

where $s = 1$ or 2. Therefore, $|\tilde{t}_i - t_j| \le O(\sqrt{\epsilon})$. 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 (τ_p, A) be a solution to BJBDP for $\mathcal{C}$. Then $\mathscr{R}(A) = \mathscr{N}(\underline{C})^\perp = \mathscr{R}(\underline{C}^\top)$.

Proof. Using (1), for any $v \in \mathscr{N}(A^\top)$, we have $C_i x = A\Sigma_i A^\top x = 0$, similarly, $C_i^\top x = 0$. Therefore, $\mathscr{N}(A^\top) \subset \mathscr{N}(\underline{C})$.

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

$$0 = C_i v = A\Sigma_i A^\top v = A\Sigma_i w = A \begin{bmatrix} \Sigma_i^{(11)} w_1 \\ \vdots \\ \Sigma_i^{(\ell\ell)} w_t \end{bmatrix}. \tag{12}$$

Therefore, we have $\Sigma_i^{(11)} w_1 = 0$ for all i. Similarly, $w_1^\top \Sigma_i^{(11)} = 0$ for all i. Let $w_1^c \in \mathbb{R}^{p_1 \times (p_1 - 1)}$ be such that $[w_1, w_1^c]$ be nonsingular, then

$$[w_1, w_1^c]^\top \Sigma_i^{(11)} [w_1, w_1^c] = \begin{bmatrix} 0 & 0 \\ 0 & * \end{bmatrix}, \quad \text{for } i = 1, \dots, m,$$

i.e., $\mathcal{C}_1 = \{\Sigma_i^{(11)}\}_{i=1}^m$ can be further block diagonalized, which contradicts with the assumption that (τ_p, A) is a solution to the BJBDP.

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

$$\mathscr{R}(A) = \mathscr{N}(A^\top)^\perp = \mathscr{N}(\underline{C})^\perp = \mathscr{R}(\underline{C}^\top)$$

This completes the proof. □

6.2 Proof of Theorem 2.2

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

$$\tilde{\phi}_p \geq \phi_p - \|\underline{E}\|, \qquad \tilde{\phi}_{p+1} \leq \|\underline{E}\|. \tag{13}$$

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

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

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

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

Then (2) follows.

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

$$\operatorname{diag}(\tilde{\phi}_1, \dots, \tilde{\phi}_p) \widetilde{V}_1^\top V_2 \overset{(a)}{=} \widetilde{U}_1^\top \underline{\widetilde{C}} V_2 = \widetilde{U}_1^\top (\underline{C} + \underline{E}) V_2 \overset{(b)}{=} \widetilde{U}_1^\top \underline{E} V_2,$$

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

$$\|\sin\Theta(\mathscr{R}(A), \mathscr{R}(\tilde{V}_1))\| = \|\widetilde{V}_1^\top V_2\| = \|\operatorname{diag}(\tilde{\phi}_1, \dots, \tilde{\phi}_p)^{-1} \widetilde{U}_1^\top \underline{E} V_2\| \leq \frac{\|\widetilde{U}_1^\top \underline{E} V_2\|}{\tilde{\phi}_p}.$$

The proof is completed. □

6.3 Proof of Theorem 2.3

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

$$X = Y \operatorname{diag}(X_{11}, \dots, X_{\ell\ell}) Y^{-1}, \tag{14}$$

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 $j \neq k$.

Proof. (⇒) (Sufficiency) Let $W = A^\top V_1$. Since $\mathscr{R}(\underline{C}^\top) = \mathscr{R}(A) = \mathscr{R}(V_1)$, and V_1, A both have full column rank, we know that W is nonsingular. Let

$$X = W^{-1}\Gamma W = W^{-1}\operatorname{diag}(\gamma_1 I_{p_1}, \dots, \gamma_\ell I_{p_\ell})W, \tag{15}$$

where $\gamma_1, \dots, \gamma_\ell$ be ℓ distinct real numbers. For all $1 \le i \le m$, we have

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

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

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

$$B_i Y \operatorname{diag}(X_{11}, \dots, X_{\ell\ell}) Y^{-1} = Y^{-\top}\operatorname{diag}(X_{11}^T, \dots, X_{\ell\ell}^T) Y^\top B_i. \tag{16}$$

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

$$\Sigma_i^{(jk)} X_{kk} = X_{jj}^\top \Sigma_i^{(jk)}, \quad \text{for} \quad j, k = 1, 2, \dots, \ell. \tag{17}$$

Consequently, for $j \ne k$, we know that $\Sigma_i^{(jk)} = 0$ since $\lambda(X_{jj}) \cap \lambda(X_{kk}) = \emptyset$. Then we know that

$$V_1^\top C_i V_1 = B_i = Y^{-\top}\Sigma_i Y^{-1}, \tag{18}$$

where $\Sigma_i = \operatorname{diag}(\Sigma_i^{(11)}, \dots, \Sigma_i^{(\ell\ell)})$. Using $\mathscr{R}(\underline{C}^\top) = \mathscr{R}(V_1)$, we know that $\mathcal{R}(C_i) \subset \mathscr{R}(V_1)$ and $\mathcal{R}(C_i^\top) \subset \mathscr{R}(V_1)$. 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. □

6.4 Proof of Theorem 2.4

Theorem 2.4. Let (τ_p, A) be a solution to the BJBDP for $\mathcal{C}$, i.e., (1) holds. Then the BJBDP for $\mathcal{C}$ is uniquely τ_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 $\operatorname{vec}(\Gamma_{jj}) \in \mathscr{N}(G_{jj})$ and a nonsingular $W_j \in \mathbb{R}^{p_j \times p_j}$ such that

$$\Gamma_{jj} = W_j \operatorname{diag}(\Gamma_{jj}^{(1)}, \Gamma_{jj}^{(2)}) W_j^{-1}, \tag{19}$$

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

$$\Sigma_i^{(jj)}\Gamma_{jj} - \Gamma_{jj}^\top \Sigma_i^{(jj)} = 0, \quad \text{for } 1 \le i \le m. \tag{20}$$

Substituting (19) into (20), we get

$$\widetilde{\Sigma}_i^{(jj)} \operatorname{diag}(\Gamma_{jj}^{(1)}, \Gamma_{jj}^{(2)}) - \operatorname{diag}(\Gamma_{jj}^{(1)}, \Gamma_{jj}^{(2)})^\top \widetilde{\Sigma}_i^{(jj)} = 0, \quad \text{for } 1 \le i \le m. \tag{21}$$

where $\widetilde{\Sigma}_i^{(jj)} = W_j^\top \Sigma_i^{(jj)} W_j$. Similar to the proof of necessity for Theorem 2.3, using $\lambda(\Gamma_{jj}^{(1)}) \cap \lambda(\Gamma_{jj}^{(2)}) = \emptyset$, we have $\widetilde{\Sigma}_i^{(jj)}$ for $1 \le i \le m$ are all block diagonal matrices. In other words, C_i's can be simultaneously block diagonalizable with more than ℓ blocks. This contradicts with the fact (τ_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 Γ_{jk}, Γ_{kj}, which are not zero at the same time, such that (4b) holds, i.e.,

$$\begin{bmatrix} \Sigma_i^{(jj)} & 0 \\ 0 & \Sigma_i^{(kk)} \end{bmatrix} \begin{bmatrix} 0 & \Gamma_{jk} \\ \Gamma_{kj} & 0 \end{bmatrix} - \begin{bmatrix} 0 & \Gamma_{kj}^\top \\ \Gamma_{jk}^\top & 0 \end{bmatrix} \begin{bmatrix} \Sigma_i^{(jj)} & 0 \\ 0 & \Sigma_i^{(kk)} \end{bmatrix} = 0. \tag{22}$$

Since $\begin{bmatrix} 0 & \Gamma_{jk} \\ \Gamma_{kj} & 0 \end{bmatrix} \neq 0$, it has at least a nonzero eigenvalue. Now let λ 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

$$\begin{bmatrix} 0 & \Gamma_{jk} \\ \Gamma_{kj} & 0 \end{bmatrix} = W_{jk} \begin{bmatrix} \Upsilon & 0 & 0 \\ 0 & -\Upsilon & 0 \\ 0 & 0 & 0 \end{bmatrix} W_{jk}^{-1}, \tag{23}$$

where Υ 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 how that $W_{jk}^{\top} \begin{bmatrix} \Sigma_i^{(jj)} & 0 \\ 0 & \Sigma_i^{(kk)} \end{bmatrix} W_{jk}$ for all $1 \leq i \leq m$ are all block diagonal. For the ease of notation, let $j = 1$, $k = 2$. Denote $\widehat{A} = A \operatorname{diag}(W_{12}^{-\top}, I_{p_3}, \ldots, I_{p_\ell})$. We know that A, $\widehat{A}$ are not equivalent since W_{12} is not (p_1, p_2)-block diagonal. This contradicts with the assumption that BJBDP for $\mathcal{C}$ is uniquely τ_p-block-diagonalizable, completing the proof of sufficiency.

($\Leftarrow$) (Necessity) Let (τ_p, A) and $(\hat{\tau}_{\hat{p}}, \widehat{A})$ be two solutions to the BJBDP for $\mathcal{C}$, i.e., it holds that

$$C_i = A \Sigma_i A^{\top} = \widehat{A} \widehat{\Sigma}_i \widehat{A}^{\top}, \tag{24}$$

where Σ_i's are all τ_p-block diagonal, $\widehat{\Sigma}_i$'s are all $\hat{\tau}_{\hat{p}}$-block-diagonal. It suffices if we can show that (τ_p, A) and $(\hat{\tau}_{\hat{p}}, \widehat{W})$ are equivalent.

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

$$\widehat{\Sigma}_i = Y^{\top} \Sigma_i Y, \quad \text{for } 1 \leq i \leq m. \tag{25}$$

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

$$\Sigma_i \Gamma = Y^{-\top}(Y^{\top} \Sigma_i Y) \operatorname{diag}(\gamma_j I_{\hat{p}_j}) Y^{-1} = Y^{-\top} \operatorname{diag}(\gamma_j I_{\hat{p}_j})(Y^{\top} \Sigma_i Y) Y^{-1} = \Gamma^{\top} \Sigma_i, \tag{26}$$

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

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., Γ is τ_p-block diagonal; using **(P1)**, $\Gamma = Y \operatorname{diag}(\gamma_j I_{\hat{p}_j}) Y^{-1}$ 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 Π_ℓ. Then

$$\operatorname{diag}(\gamma_1 I_{p_{k_1}}, \ldots, \gamma_\ell I_{p_{k_\ell}}) = \Pi^{\top} \operatorname{diag}(\gamma_1' I_{p_1}, \ldots, \gamma_\ell' I_{p_\ell}) \Pi. \tag{27}$$

where γ_j' is the eigenvalue of Γ_{jj}. Then it follows that

$$\operatorname{diag}(\Gamma_{11}, \ldots, \Gamma_{\ell\ell}) = Y \Pi^{\top} \operatorname{diag}(\gamma_1' I_{p_1}, \ldots, \gamma_\ell' I_{p_\ell})(Y \Pi^{\top})^{-1}. \tag{28}$$

Noticing that the columns of $Y\Pi^{\top}$ are eigenvectors of Γ, we know that $Y\Pi^{\top}$ is τ_p-block-diagonal. Therefore, we can rewrite $\widehat{A} = AY^{-\top}$ as $\widehat{A} = A(Y\Pi^{\top})^{-\top}\Pi$, in which $(Y\Pi^{\top})^{-\top}$ is τ_p-block-diagonal, Π is the permutation matrix associated with Π_ℓ. So, (τ_p, A) and $(\hat{\tau}_p, \widehat{A})$ are equivalent. The proof is completed. $\square$

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 $\underline{D}$ 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

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

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 $\text{OPT}(\mathcal{D})$ is not empty, then it has a solution X_*. Using $\mathrm{tr}(X_*) = 0$, $\mathrm{tr}(X_*^2) = q > 0$, we know that X_* can be factorized into $X_* = Y \operatorname{diag}(\Gamma_1, \Gamma_2) Y^{-1}$, where Γ_1, Γ_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 γ be an arbitrary eigenvalue of X_*, and z be the corresponding eigenvector. Using $X_* \in \mathscr{N}(\mathcal{D})$, we have

$$0 = z^{\mathrm{H}} D_i X_* z - z^{\mathrm{H}} X_*^{\top} D_i z = (\gamma - \bar{\gamma}) z^{\mathrm{H}} D_i z, \quad \text{for } 1 \le i \le m.$$

Then it follows that

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

Since $\mu > 0$ has full column rank, we know that $\sum_{i=1}^{\ell} |z^{\mathrm{H}} D_i z|^2 = 0$. Therefore, γ 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 $\gamma_1 \le \cdots \le \gamma_q$. Then

$$\mathrm{tr}(X_*) = \sum_{j=1}^q \gamma_j = 0, \quad \mathrm{tr}(X_*^2) = \sum_{j=1}^q \gamma_j^2 = q, \quad \mathrm{tr}(X_*^4) = \sum_{j=1}^q \gamma_j^4. \tag{29}$$

Using the method of Lagrange multipliers, we consider

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

where μ_1, μ_2 are Lagrange multipliers. By calculations, we have

$$\frac{\partial L}{\partial \gamma_j} = 4\gamma_j^3 + \mu_1 + 2\mu_2 \gamma_j = 0. \tag{30}$$

Noticing that γ_j's are the real roots of the third order equation $4t^3 + 2\mu_2 t + \mu_1 = 0$, which has one real root or three real roots, we know that either γ_j's are identical to the unique real root or γ_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 $t_1 \le t_2 \le t_3$ 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

$$q_1 t_1 + q_2 t_2 + q_3 t_3 = 0, \quad q_1 t_1^2 + q_2 t_2^2 + q_3 t_3^2 = q, \quad \mathrm{tr}(X_*^4) = q_1 t_1^4 + q_2 t_2^4 + q_3 t_3^4. \tag{31}$$

Introduce two vectors $u = [\sqrt{q_1} t_1^2, \sqrt{q_2} t_2^2, \sqrt{q_3} t_3^2]^{\top}$, $v = [\sqrt{q_1}, \sqrt{q_2}, \sqrt{q_3}]^{\top}$. Then we have $\|u\| = \sqrt{\mathrm{tr}(X_*^4)}$, $\|v\| = \sqrt{q}$. Using Cauchy's inequality, we get

$$\mathrm{tr}(X_*^4) = \|u\|^2 \|v\|^2 / q \ge (u^{\top} v)^2 / q = (q_1 t_1^2 + q_2 t_2^2 + q_3 t_3^2)^2 / q = q,$$

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 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 $\mathscr{N}(\mathcal{D})$. If $\mathcal{D}$ has a nontrivial diagonalizer, then there exists a matrix Z such that $D_i = Z\Phi_i Z^\top$, where Φ_i's are $\tau_q = (q_1, q_2)$-block diagonal. Since $\underline{D}$ has full column rank, Z is nonsingular. Let $X = Z^{-T}\operatorname{diag}(\sqrt{\frac{q_2}{q_1}} I_{q_1}, -\sqrt{\frac{q_1}{q_2}} I_{q_2}) Z^\top$. It is easy to see that $\operatorname{tr}(X) = 0$, $\operatorname{tr}(X^2) = 1$ and $X \in \mathscr{N}(\mathcal{D})$. In other words, there exists a feasible X which has two distinct real eigenvalues. Therefore, we may declare that $\textsc{opt}(\mathcal{D})$ 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 $\gamma_1 = \sqrt{\frac{q_2}{q_1}}$, $\gamma_2 = -\sqrt{\frac{q_1}{q_2}}$. 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. □

6.6 Proof of Theorem 2.6

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

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

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

(2) Denote $\widehat{Z}^{-1} D_i \widehat{Z}^{-\top} = \operatorname{diag}(D_i^{(1)}, D_i^{(2)})$, $\mathcal{D}^{(1)} = \{D_i^{(1)}\}$ and $\mathcal{D}^{(2)} = \{D_i^{(2)}\}$. Then $\underline{D}^{(1)}$ and $\underline{D}^{(2)}$ both have full column rank.

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

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

$$\hat{\tau}_p = (\hat{p}_1, \ldots, \hat{p}_{\hat{\ell}}), \quad C_i = \widehat{A}\widehat{\Sigma}_i\widehat{A}^\top = \widehat{A}\operatorname{diag}(\widehat{\Sigma}_i^{(11)}, \ldots, \widehat{\Sigma}_i^{(\hat{\ell}\hat{\ell})})\widehat{A}^\top, \quad i = 1, \ldots, m, \tag{32}$$

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

Using (1) and (32), we have

$$B_i = V_1^\top \widehat{A}\widehat{\Sigma}_i\widehat{A}^\top V_1 = \widehat{Z}\widehat{\Sigma}_i\widehat{Z}^\top = V_1^\top A\Sigma_i A^\top V_1 = Z\Sigma_i Z^\top. \tag{33}$$

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

$$\widehat{\Sigma}_i = Y^\top \Sigma_i Y, \quad \text{for } 1 \leq i \leq m. \tag{34}$$

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

$$\Sigma_i \Gamma = Y^{-\top}(Y^\top \Sigma_i Y) \operatorname{diag}(\gamma_j I_{\hat{p}_j}) Y^{-1} = Y^{-\top} \operatorname{diag}(\gamma_j I_{\hat{p}_j})(Y^\top \Sigma_i Y) Y^{-1} = \Gamma^\top \Sigma_i, \tag{35}$$

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

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., Γ is τ_p-block diagonal; using **(P1)**, $\Gamma = Y \operatorname{diag}(\gamma_j I_{\hat{p}_j}) Y^{-1}$ and $\cup_{j=1}^{\ell} \lambda(\Gamma_{jj}) = \lambda(\Gamma)$, we know that for each Γ_{jj} $(j = 1, \dots, \ell)$, its eigenvalues are all γ_k $(1 \leq k \leq \hat{\ell})$. If $\hat{\ell} < \ell$, there exist at least two blocks of Γ_{jj}'s corresponding to the same γ_k. Without loss of generality, let Γ_{11}, Γ_{22} correspond to γ_1, the remaining blocks correspond to other γ_k's. Then using $\Gamma = Y \operatorname{diag}(\gamma_1 I_{\hat{p}_1}, \dots, \gamma_\ell I_{\hat{p}_{\hat{\ell}}}) Y^{-1}$, we know that $Y = \operatorname{diag}(Y_{11}, Y_{22})$, where $Y_{11} \in \mathbb{R}^{\hat{p}_1 \times \hat{p}_1}$ and $\hat{p}_1 = p_1 + p_2$. Using $Y = Z^\top \widehat{Z}^{-\top}$ and (35), we get

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

Therefore, we have

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

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

6.7 Proof of Theorem 2.7

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

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

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

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

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

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

Proof. First, we show of (I) via its the contrapositive. If the feasible set of $\text{OPT}(\widetilde{\mathcal{D}}, \delta)$ is not empty, then $\text{OPT}(\widetilde{\mathcal{D}}, \delta)$ has a solution X_*, which can be factorized into $X_* = Y \operatorname{diag}(\Gamma_1, \Gamma_2) Y^{-1}$ (since $\operatorname{tr}(X_*) = 0$ and $\operatorname{tr}(X_*^2) = q$), 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 $Z = Y^{-\top}$, $\Phi_i = \operatorname{diag}(Y_1^\top \widetilde{D}_i Y_1, Y_2^\top \widetilde{D}_i Y_2)$, $g = \min \frac{\|\Gamma_1^\top X - X\Gamma_2\|_F}{\|X\|_F}$ and $\kappa = \kappa_2(Y) = \frac{\sigma_{\max}(Y)}{\sigma_{\min}(Y)}$. By calculations, we have

$$\begin{aligned}
\|X_*\|_F^2 &= \operatorname{tr}(Y^{-\top} \operatorname{diag}(\Gamma_1^\top, \Gamma_2^\top) Y^\top Y \operatorname{diag}(\Gamma_1, \Gamma_2) Y^{-1}) \\
&\leq \|Y\|^2 \operatorname{tr}(Y^{-\top} \operatorname{diag}(\Gamma_1^\top, \Gamma_2^\top) \operatorname{diag}(\Gamma_1, \Gamma_2) Y^{-1} \\
&= \|Y\|^2 \operatorname{tr}(\operatorname{diag}(\Gamma_1, \Gamma_2) Y^{-1} Y^{-\top} \operatorname{diag}(\Gamma_1^\top, \Gamma_2^\top)) \\
&\leq \kappa^2 \operatorname{tr}(\operatorname{diag}(\Gamma_1, \Gamma_2) \operatorname{diag}(\Gamma_1^\top, \Gamma_2^\top)) = \kappa^2 \operatorname{tr}(X_*^2) = \kappa^2 q,
\end{aligned} \tag{36}$$

and

$$\begin{aligned}
\delta^2\|\mathrm{vec}(X_*)\|^2 &\overset{(a)}{\geq} \|\mathbf{L}(\widetilde{\mathcal{D}})\mathrm{vec}(X_*)\|^2 = \sum_{i=1}^m \|\widetilde{D}_i X_* - X_*^\top \widetilde{D}_i\|_F^2 \\
&= \sum_{i=1}^m \|Z(Y^\top \widetilde{D}_i Y \operatorname{diag}(\Gamma_1,\Gamma_2) - \operatorname{diag}(\Gamma_1^\top,\Gamma_2^\top) Y^\top \widetilde{D}_i Y) Z^\top\|_F^2 \\
&\geq \frac{1}{\|Y\|^4}\sum_{i=1}^m \|Y^\top \widetilde{D}_i Y \operatorname{diag}(\Gamma_1,\Gamma_2) - \operatorname{diag}(\Gamma_1^\top,\Gamma_2^\top) Y^\top \widetilde{D}_i Y\|_F^2 \\
&\geq \frac{1}{\|Y\|^4}\sum_{i=1}^m \Big(\|Y_1^\top \widetilde{D}_i Y_2 \Gamma_2 - \Gamma_1^\top Y_1^\top \widetilde{D}_i Y_2\|_F^2 + \|Y_2^\top \widetilde{D}_i Y_1 \Gamma_1 - \Gamma_2^\top Y_2^\top \widetilde{D}_i Y_1\|_F^2\Big) \\
&\overset{(b)}{\geq} \frac{g^2}{\|Y\|^4}\sum_{i=1}^m \Big(\|Y_1^\top \widetilde{D}_i Y_2\|_F^2 + \|Y_2^\top \widetilde{D}_i Y_1\|_F^2\Big) \geq \frac{g^2}{\kappa^4}\sum_{i=1}^m \|Z(Y^\top \widetilde{D}_i Y - \Phi_i)Z^\top\|_F^2 \\
&= \frac{g^2}{\kappa^4}\sum_{i=1}^m \|\widetilde{D}_i - Z\Phi_i Z^\top\|_F^2, \qquad (37)
\end{aligned}$$

where (a) uses $X_* \in \mathscr{N}_\delta(\widetilde{\mathcal{D}})$, (b) uses the definition of g. Then it follows from (36) and (37) that

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

This completes the proof of (I).

Next, we show (II). If $\widetilde{\mathcal{D}}$ has a nontrivial δ-diagonalizer, then there exists a matrix Z such that $\sum_{i=1}^m \|\widetilde{D}_i - Z\Phi_i Z^\top\|_F^2 \leq \frac{1}{4}\delta^2$ (by setting $\delta = \frac{1}{2\sqrt{C}}\delta$, the constant becomes $\frac{1}{4}$, and by definition, Z is still a δ-diagonalizer), where Φ_i's are all $\tau_q = (q_1, q_2)$ block diagonal matrices. Let $X = Z^{-\top}\Gamma Z^\top$, where $\Gamma = \operatorname{diag}(\sqrt{\frac{q_2}{q_1}} I_{q_1}, -\sqrt{\frac{q_1}{q_2}} I_{q_2})$. By calculations, we have

$$\begin{aligned}
\|\mathbf{L}(\widetilde{\mathcal{D}})\mathrm{vec}(X)\|^2 &= \sum_{i=1}^m \|\widetilde{D}_i X - X^\top \widetilde{D}_i\|_F^2 \overset{(a)}{\leq} 2\sum_{i=1}^m \|(\widetilde{D}_i - Z\Phi_i Z^\top)X - X^\top(\widetilde{D}_i - Z\Phi_i Z^\top)\|_F^2 \\
&\leq 4\|X\|^2 \sum_{i=1}^m \|\widetilde{D}_i - Z\Phi_i Z^\top\|_F^2 \leq \|X\|^2\delta^2,
\end{aligned}$$

where (a) uses $Z\Phi_i Z^\top X - X^\top Z\Phi_i Z^\top = 0$. Therefore, $\frac{\|\mathbf{L}(\widetilde{\mathcal{D}})\mathrm{vec}(X)\|}{\|\mathrm{vec}(X)\|} \leq \frac{\|X\|\delta}{\|X\|_F} \leq \delta$. Also note that $\mathrm{tr}(X) = 0$ and $\mathrm{tr}(X^2) = q$, then the feasible set of $\textsc{opt}(\widetilde{\mathcal{D}}, \delta)$ is nonempty. Consequently, $\textsc{opt}(\widetilde{\mathcal{D}}, \delta)$ has a solution X_*.

Let γ be an arbitrary eigenvalue of X_*, and z be the corresponding unit-length eigenvector. By calculations, we have

$$\begin{aligned}
\kappa^2 q\delta^2 \geq \delta^2\|X_*\|_F^2 &= \|\mathbf{L}(\widetilde{\mathcal{D}})\mathrm{vec}(X)\|^2 \geq \sum_{i=1}^m \|\widetilde{D}_i X_* - X_*^\top \widetilde{D}_i\|_F^2 \\
&\geq \sum_{i=1}^m \|z^{\mathrm{H}}\widetilde{D}_i X_* z - z^{\mathrm{H}} X_*^\top \widetilde{D}_i z\|_F^2 = |\gamma - \bar{\gamma}|^2 \sum_{i=1}^m |z^{\mathrm{H}}\widetilde{D}_i z|^2 \geq \mu^2|\gamma - \bar{\gamma}|^2, \qquad (38)
\end{aligned}$$

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

Now let the eigenvalues of X_* be $\mu_j + \eta_j\sqrt{-1}$ for $j = 1, \ldots, q$, where μ_j, $\eta_j \in \mathbb{R}$. Then

$$\mathrm{tr}(X_*) = \sum_{j=1}^q \gamma_j = 0, \quad \mathrm{tr}(X_*^2) = \sum_{j=1}^q (\gamma_j^2 - \eta_j^2) = q, \quad \mathrm{tr}(X_*^4) = \sum_{j=1}^q (\gamma_j^4 + \eta_j^4 - 6\gamma_j^2\eta_j^2). \qquad (39)$$

Using the method of Lagrange multipliers, we consider

$$L(\gamma_1,\eta_1,\dots,\gamma_q,\eta_q;\mu_1,\mu_2)=\sum_{j=1}^q(\gamma_j^4+\eta_j^4-6\gamma_j^2\eta_j^2)+\mu_1\sum_{j=1}^q\gamma_j+\mu_2\Big(\sum_{j=1}^q(\gamma_j^2-\eta_j^2)-q\Big),$$

where μ_1, μ_2 are Lagrange multipliers. By calculations, we have

$$\frac{\partial L}{\partial\gamma_j}=4\gamma_j^3+2(\mu_2-6\eta_j^2)\gamma_j+\mu_1=0. \tag{40}$$

Take (40) as perturbed third order equations of $4t^3+2\mu_2t+\mu_1=0$. Using Lemma 5.3 and $|\eta_j|\le O(\delta)$, we know that $\gamma_j\subset\cup_{i=1}^3\{z\mid |z-t_i|\le O(\delta)\}$, where t_1, t_2 and t_3 are the roots of $4t^3+2\mu_2t+\mu_1=0$. Next, we consider the following cases:

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

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 $\lambda(X_*)\subset\cup_{i=1,2}\mathrm{Rect}_i$.

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

$$\mathrm{tr}(X_*)=q_1t_1+q_2t_2+q_3t_3+O(\delta)=0, \tag{41a}$$
$$\mathrm{tr}(X_*^2)=q_1t_1^2+q_2t_2^2+q_3t_3^2+O(\delta)=q, \tag{41b}$$
$$\mathrm{tr}(X_*^4)=q_1t_1^4+q_2t_2^4+q_3t_3^4+O(\delta). \tag{41c}$$

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 $\|u\|^2+O(\delta)=\mathrm{tr}(X_*^4)$, $\|v\|=\sqrt{q}$. Using Cauchy's inequality, we get

$$\mathrm{tr}(X_*^4)+O(\delta)=\|u\|^2=\|u\|^2\|v\|^2/q\ge(u^\top v)^2/q=(q_1t_1^2+q_2t_2^2+q_3t_3^2)^2/q=q+O(\delta),$$

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 $\cup_{i=1,2,3,q_i\ne0}\{z\mid |z-t_i|\le O(\delta)\}$. 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 $\cup_{i=1,2}\mathrm{Rect}_i$.

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

If $\widetilde{\mathcal{D}}$ has a nontrivial δ-diagonalizer, then there exists a matrix Z such that $\sum_{i=1}^m\|\widetilde{D}_i-Z\Phi_iZ^\top\|_F^2\le\frac14\delta^2$, where Φ_i's are all $\tau_q=(q_1,q_2)$ block diagonal matrices. Let $X=Z^{-\top}\Gamma Z^\top$, where $\Gamma=\mathrm{diag}(\sqrt{\frac{q_2}{q_1}}I_{q_1},-\sqrt{\frac{q_1}{q_2}}I_{q_2})$. We know that X is also feasible. Therefore, we may declare that $\textsc{opt}(\widetilde{\mathcal{D}},\delta)$ 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 Disk_1, Disk_2, respectively. We show $\rho_1-\rho_2\ge 2+O(\delta)$. 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 $\rho_1=\sqrt{\frac{q_2}{q_1}}+O(\delta)$, $\rho_2=-\sqrt{\frac{q_1}{q_2}}+O(\delta)$. Then it follows that

$$\rho_1-\rho_2=\sqrt{\frac{q_2}{q_1}}+\sqrt{\frac{q_1}{q_2}}+O(\delta)\ge 2+O(\delta),$$

completing the proof. □

6.8 Proof of Theorem 2.8

Theorem 2.8. Assume that the BJBDP for $\mathcal{C} = \{C_i\}_{i=1}^m$ is uniquely τ_p-block-diagonalizable, and let (τ_p, A) be a solution satisfying (1). Let $\widetilde{\mathcal{C}} = \{\widetilde{C}_i\}_{i=1}^m = \{C_i + E_i\}_{i=1}^m$ be a perturbed matrix set of $\mathcal{C}$. Denote

$$\tau_p = (p_1, \dots, p_\ell), \quad \hat{\tau}_p = (\hat{p}_1, \dots, \hat{p}_{\hat{\ell}}), \quad A = [A_1, \dots, A_\ell], \quad \widehat{A} = [\widehat{A}_1, \dots, \widehat{A}_{\hat{\ell}}],$$

where $(\hat{\tau}_p, \widehat{A})$ is the output of Algorithm 4. Assume $\mathscr{N}(G_{jj}) = \mathscr{R}(\mathrm{vec}(I_{p_j}))$ 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 $\underline{\widetilde{C}}$ be the same as in Theorem 2.2,

$$\epsilon = \frac{\|\underline{E}\|}{\tilde{\phi}_p}, \quad r = \frac{\sqrt{2(d+2C)}\,\tilde{\phi}_p\,\epsilon}{\sigma_{\min}^2(A)(1-\epsilon^2)}, \quad g_j = \frac{\sqrt{2j}}{(\hat{\ell}-1)\kappa\sqrt{p}} - \max\{\frac{\kappa}{\omega_{\mathrm{neq}}}, \frac{1}{\omega_{\mathrm{ir}}}\}r, \quad \text{for } j = 1, 2,$$

where C and κ are two constants.

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

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

$$\|[\widehat{A}_{1'}, \dots, \widehat{A}_{\ell'}] - AD\|_F \le \frac{\frac{c\,r}{\omega_{\mathrm{neq}}}}{g_2 - \frac{r}{\omega_{\mathrm{ir}}}}\|A\|_F + (\frac{\epsilon^2}{\sqrt{1-\epsilon^2}} + \epsilon)\|\widehat{A}\|_F = O(\epsilon),$$

where c is a constant.

Proof. Using $\|\underline{E}\| < \epsilon\tilde{\phi}_p$ and Theorem 2.2, we have

$$\delta = \tilde{\phi}_{p+1} \le \|\underline{E}\| \le \epsilon\tilde{\phi}_p, \qquad \|\sin\Theta(\mathscr{R}(A), \mathscr{R}(\widetilde{V}_1))\| \le \frac{\|\widetilde{U}_1^\top \underline{E} V_2\|}{\tilde{\phi}_p} \le \frac{\|\underline{E}\|}{\tilde{\phi}_p} \le \epsilon. \tag{42}$$

Let $[V_1, V_2]$ be an orthogonal matrix such that $\mathscr{R}(V_1) = \mathscr{R}(A)$, $\mathscr{R}(V_2) = \mathscr{N}(A^\top)$. Then we can write $\widetilde{V}_1 = V_1 T_c + V_2 T_s$, where $\begin{bmatrix} T_c \\ T_s \end{bmatrix}$ is orthonormal, $\|T_s\| = \|\sin\Theta(V_1, \widetilde{V}_1)\| \le \epsilon$, $\sigma_{\min}(T_c) = \sqrt{1 - \|\sin\Theta(V_1, \widetilde{V}_1)\|^2} \ge \sqrt{1-\epsilon^2}$. Therefore, T_c is nonsingular. Let $B_i = V_1^\top C_i V_1$, $\widetilde{B}_i = \widetilde{V}_1^\top \widetilde{C}_i \widetilde{V}_1$. And by calculations, we have

$$\begin{aligned}
\|\widetilde{B}_i - T_c^\top B_i T_c\|_F &= \|\widetilde{V}_1^\top (C_i + E_i)\widetilde{V}_1 - T_c^\top V_1^\top C_i V_1 T_c\|_F \\
&\le \|\widetilde{V}_1^\top C_i \widetilde{V}_1 - T_c^\top V_1^\top C_i V_1 T_c + \widetilde{V}_1^\top E_i \widetilde{V}_1\|_F \\
&\overset{(a)}{\le} \|T_c^\top V_1^\top C_i V_2 T_s + T_s^\top V_2^\top C_i V_1 T_c + T_s^\top V_2^\top C_i V_2 T_s + \widetilde{V}_1^\top E_i \widetilde{V}_1\|_F \\
&\overset{(b)}{=} \|E_i\|_F,
\end{aligned} \tag{43}$$

where (a) uses $\widetilde{V}_1 = V_1 T_c + V_2 T_s$, (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

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

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

$$\sum_{i=1}^m \|\widetilde{B}_i - \widehat{Z}\widehat{\Sigma}_i\widehat{Z}^\top\|_F^2 \le C\delta^2 = C\tilde{\phi}_{p+1}^2 \le C\tilde{\phi}_p^2\epsilon^2, \tag{45}$$

where $\widehat{\Sigma}_i = \mathrm{diag}(\Sigma_{i1}, \dots, \widehat{\Sigma}_{i\hat{\ell}})$'s are all $\hat{\tau}_p = (\hat{p}_1, \dots, \hat{p}_{\hat{\ell}})$-block diagonal, and for each $1 \le j \le \hat{\ell}$, $\{\Sigma_{ij}\}_{i=1}^m$ does not have δ-block diagonalizer.

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

$$\begin{aligned}\sum_{i=1}^{m}\|Z\Sigma_i Z^\top - \widehat{Z}\widehat{\Sigma}_i\widehat{Z}^\top\|_F^2 &\le 2\sum_{i=1}^{m}(\|Z\Sigma_i Z^\top - \widetilde{B}_i\|_F^2 + \|\widetilde{B}_i - \widehat{Z}\widehat{\Sigma}_i\widehat{Z}^\top\|_F^2)\\ &\le 2(\sum_{i=1}^{m}\|E_i\|_F^2 + C\tilde{\phi}_p^2\epsilon^2) = \|\underline{E}\|_F^2 + 2C\tilde{\phi}_p^2\epsilon^2 \le d\|\underline{E}\|^2 + 2C\tilde{\phi}_p^2\epsilon^2\\ &\le (d+2C)\tilde{\phi}_p^2\epsilon^2. \end{aligned} \tag{46}$$

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

$$\varrho = \varrho\|\Gamma\|_F = \|Y\operatorname{diag}(\gamma_1 I_{\hat{p}_1},\ldots,\gamma_\ell I_{\hat{p}_{\hat{\ell}}})Y^{-1}\|_F \le \kappa(Y)\sqrt{\sum_{j=1}^{\hat{\ell}}\hat{p}_j\gamma_j^2} \le \kappa(Y)\sqrt{p}. \tag{47}$$

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

$$\begin{aligned}\sum_{i=1}^{m}\|\Sigma_i\Gamma - \Gamma^\top\Sigma_i\|_F^2 &= \sum_{i=1}^{m}\|Z^{-1}(Z\Sigma_i Z^\top\widehat{Z}^{-\top}\widehat{\Gamma}\widehat{Z}^\top - \widehat{Z}\widehat{\Gamma}^\top\widehat{Z}^{-1}Z\Sigma_i Z^\top)Z^{-\top}\|_F^2\\ &= \sum_{i=1}^{m}\|Z^{-1}((\widehat{Z}\widehat{\Sigma}_i\widehat{Z}^\top + F_i)\widehat{Z}^{-\top}\widehat{\Gamma}\widehat{Z}^\top - \widehat{Z}\widehat{\Gamma}^\top\widehat{Z}^{-1}(\widehat{Z}\widehat{\Sigma}_i\widehat{Z}^\top + F_i))Z^{-\top}\|_F^2\\ &= \sum_{i=1}^{m}\|Z^{-1}F_i Z^{-\top}\Gamma - \Gamma^\top Z^{-1}F_i Z^{-\top}\|_F^2\\ &\le 2\|\Gamma\|_F^2\sum_{i=1}^{m}\|Z^{-1}F_i Z^{-\top}\|^2 \overset{(a)}{\le} \frac{2(d+2C)\tilde{\phi}_p^2\epsilon^2}{\sigma_{\min}^4(Z)} \overset{(b)}{\le} r^2, \end{aligned} \tag{48}$$

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

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

$$\sum_{j=1}^{\ell}\|G_{jj}\mathrm{vec}(\Gamma_{jj})\|^2 + \sum_{1<j<k\le\ell}\left\|G_{jk}\begin{bmatrix}\mathrm{vec}(\Gamma_{jk})\\ -\mathrm{vec}(\Gamma_{kj}^\top)\end{bmatrix}\right\|^2 = \sum_{i=1}^{m}\|\Sigma_i\Gamma - \Gamma^\top\Sigma_i\|_F^2 \le r^2. \tag{49}$$

Let $r_{jj} = G_{jj}\mathrm{vec}(\Gamma_{jj})$, the eigenvalues of Γ_{jj} be $\gamma_{j1},\ldots,\gamma_{jp_j}$, for $j = 1,\ldots,\ell$. Then we have

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

where $\widehat{\Gamma}_{jj} = \mathrm{reshape}(G_{jj}^\dagger r_{jj}, p_j, p_j)$. And it follows that

$$\sum_{k=1}^{p_j}|\gamma_{jk} - \hat{\gamma}_j|^2 \le \|\widehat{\Gamma}_{jj}\|_F^2 \le \frac{\|r_{jj}\|^2}{\omega_{\mathrm{ir}}^2}. \tag{50}$$

Let $r_{jk} = G_{jk}\begin{bmatrix}\mathrm{vec}(\Gamma_{jk})\\ -\mathrm{vec}(\Gamma_{kj}^\top)\end{bmatrix}$, for $1 \le j < k < \ell$. Then we have

$$\|\Gamma_{jk}\|_F^2 + \|\Gamma_{kj}\|_F^2 \le \|G_{jk}^\dagger r_{jk}\|^2 \le \frac{\|r_{jk}\|^2}{\omega_{\mathrm{neq}}^2}. \tag{51}$$

Let $\mu_{jk} = \operatorname{argmin}_{\gamma\in\{\gamma_1,\ldots,\gamma_{\hat\ell}\}}|\frac{\gamma}{\varrho} - \gamma_{jk}|$. By Sun (1996, Remark 3.3, (2)), it holds that

$$\sum_{j=1}^{\ell}\sum_{k=1}^{p_j}|\frac{\mu_{jk}}{\varrho} - \gamma_{jk}|^2 \le \kappa^2(Y)\sum_{j<k}(\|\Gamma_{jk}\|_F^2 + \|\Gamma_{kj}\|_F^2) \tag{52}$$

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

$$\begin{aligned}\sum_{j=1}^{\ell}\sum_{k=1}^{p_j}|\frac{\mu_{jk}}{\varrho} - \hat\gamma_j|^2 &\le \sum_{j=1}^{\ell}\sum_{k=1}^{p_j}|\frac{\mu_{jk}}{\varrho} - \gamma_{jk}|^2 + \sum_{j=1}^{\ell}\sum_{k=1}^{p_j}|\gamma_{jk} - \hat\gamma_j|^2\\ &\le \frac{\kappa^2(Y)}{\omega_{\text{neq}}^2}\sum_{j<k}\|r_{jk}\|^2 + \frac{1}{\omega_{\text{ir}}^2}\sum_j\|r_{jj}\|^2 \le \max\{\frac{\kappa^2(Y)}{\omega_{\text{neq}}^2}, \frac{1}{\omega_{\text{ir}}^2}\}r^2.\end{aligned} \tag{53}$$

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

$$\sum_{j=1}^{\ell}\sum_{k=1}^{p_j}|\frac{\mu_{jk}}{\varrho} - \gamma_{jk}|^2 \ge |\frac{\gamma_1}{\varrho} - \hat\gamma_j|^2 + |\frac{\gamma_2}{\varrho} - \hat\gamma_j|^2 \ge \frac{|\gamma_1-\gamma_2|^2}{2\varrho^2} \ge \frac{2}{(\hat\ell-1)^2\kappa^2(Y)p}, \tag{54}$$

where the last inequality uses the definition of γ_j and also (47). Combining (53) and (54), we get $\max\{\frac{\kappa(Y)}{\omega_{\text{neq}}}, \frac{1}{\omega_{\text{ir}}}\}r \ge \frac{1}{(\hat\ell-1)\kappa(Y)}\sqrt{\frac{2}{p}}$, 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 $p_j = \hat p_{j'}$, completing the proof of (I).

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

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

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

$$\operatorname{diag}(\hat\gamma_1 I, \ldots, \hat\gamma_\ell I)\operatorname{OffBlkdiag}_{\tau_p}(Y^{-\top}) - \operatorname{OffBlkdiag}_{\tau_p}(Y^{-\top})\frac{1}{\varrho}\operatorname{diag}(\gamma_1 I, \ldots, \gamma_\ell I) = -R.$$

Then it follows that $(\hat\gamma_j - \frac{\gamma_k}{\varrho})Y_{jk} = R_{jk}$ for $j \ne k$. By calculations, we have

$$\begin{aligned}\|Y_{jk}\|_F &= \frac{\|R_{jk}\|_F}{|\hat\gamma_j - \gamma_k/\varrho|} \le \frac{\|R_{jk}\|_F}{|\gamma_j/\varrho - \gamma_k/\varrho| - |\hat\gamma_j - \gamma_j/\varrho|} \overset{(a)}{\le} \frac{\|R_{jk}\|_F}{\frac{2|j-k|}{\varrho(\ell-1)} - |\hat\gamma_j - \gamma_j/\varrho|} \overset{(b)}{\le} \frac{\|R_{jk}\|_F}{g_2},\\ \|R\|_F &\le \|\operatorname{OffBlkdiag}_{\tau_p}(\Gamma^\top)\|\|Y^{-\top}\| + \max_j\|\Gamma_{jj} - \hat\gamma_j I\|\|\operatorname{OffBlkdiag}_{\tau_p}(Y^{-\top})\|_F\\ &\overset{(c)}{\le} \|\operatorname{OffBlkdiag}_{\tau_p}(\Gamma^\top)\|\|Y^{-\top}\| + \frac{\sqrt{\sum_j\|r_{jj}\|^2}}{\omega_{\text{ir}}}\|\operatorname{OffBlkdiag}_{\tau_p}(Y^{-\top})\|_F,\end{aligned}$$

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

$$\begin{aligned}&\|\operatorname{OffBlkdiag}_{\tau_p}(Y^{-\top})\|_F \le \frac{\|R\|_F}{g_2}\\ \le&\frac{1}{g_2}\Big(\|\operatorname{OffBlkdiag}_{\tau_p}(\Gamma^\top)\|_F\|Y^{-\top}\| + \frac{\sqrt{\sum_j\|r_{jj}\|^2}}{\omega_{\text{ir}}}\|\operatorname{OffBlkdiag}_{\tau_p}(Y^{-\top})\|_F\Big),\end{aligned}$$

and hence

$$\|\operatorname{OffBlkdiag}_{\tau_p}(Y^{-\top})\|_F \le \frac{\|\operatorname{OffBlkdiag}_{\tau_p}(\Gamma^\top)\|_F\|Y^{-\top}\|}{g_2 - \frac{\sqrt{\sum_j\|r_{jj}\|^2}}{\omega_{\text{ir}}}} \le \frac{\frac{r}{\omega_{\text{neq}}}\|Y^{-1}\|}{g_2 - \frac{r}{\omega_{\text{ir}}}}, \tag{55}$$

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

Finally, by calculations, we have

$$\begin{aligned}
\widehat{A} &= \widetilde{V}_1\widehat{Z} = (V_1T_c + V_2T_s)\widehat{Z} = (V_1T_c^{-\top}(I - T_s^{\top}T_s) + V_2T_s)\widehat{Z} \\
&= V_1T_c^{-\top}ZY^{-\top} + (-V_1T_c^{-\top}(T_s^{\top}T_s) + V_2T_s)\widehat{Z} \\
&= AY^{-\top} + (-V_1T_c^{-\top}(T_s^{\top}T_s) + V_2T_s)\widehat{Z} \\
&= A\,\mathrm{diag}(Y_{11},\ldots,Y_{\ell\ell}) + A\,\mathrm{OffBlkdiag}_{\tau_p}(Y^{-\top}) + (-V_1T_c^{-\top}(T_s^{\top}T_s) + V_2T_s)\widehat{Z},
\end{aligned}$$

and it follows that

$$\begin{aligned}
\|\widehat{A} - A\,\mathrm{diag}(Y_{11},\ldots,Y_{\ell\ell})\|_F &\leq \|A\|\|\,\mathrm{OffBlkdiag}_{\tau_p}(Y^{-\top})\|_F + (\|T_c^{-\top}T_s^{\top}T_s\| + \|T_s\|)\|\widehat{Z}\|_F \\
&\leq \|A\|\frac{\frac{r}{\omega_{\mathrm{neq}}}\|Y^{-\top}\|}{g_2 - \frac{r}{\omega_{\mathrm{ir}}}} + (\frac{\epsilon^2}{\sqrt{1-\epsilon^2}} + \epsilon)\|\widehat{A}\|_F.
\end{aligned}$$

The proof is completed. □