Supplemental materials to "Robust and Computationally Feasible Community Detection in the Presence of Arbitrary Outliers"

0.1. *Proof of Lemma 6.6.* The proof of inequalities (6.2), (6.3), (6.4) and (6.5) are given as follows step by step.

proof of (6.2)*:.*
The off-diagonal entries of each row of $\boldsymbol{K}_{ii}$ are $l_i - 1$ IID random variables $x_1, \dots, x_{l_i-1}$ obeying

$$\mathbb{P}(x_k = 1) = B_{ii}, \quad \mathbb{P}(x_k = 0) = 1 - B_{ii}.$$

By Chernoff's inequality (Lemma 6.3), we have

$$\mathbb{P}\left(\sum_{j=1}^{l_i-1} x_j \le (l_i - 1)B_{ii} - t\right) \le e^{-\frac{t^2}{2(l_i-1)B_{ii}}}.$$

By letting $t = 2\sqrt{(l_i - 1)B_{ii}\log n}$, with probability at least $1 - \frac{1}{n^2}$,

$$\sum_{j=1}^{l_i-1} x_j \ge (l_i - 1)B_{ii} - 2\sqrt{(l_i - 1)B_{ii}\log n}.$$

Then, with probability at least $1 - \frac{1}{n}$, for all $i = 1, \dots, r$, there holds

$$\boldsymbol{K}_{ii}\mathbf{1}_{l_i} \ge \left((l_i - 1)B_{ii} - 2\sqrt{(l_i - 1)B_{ii}\log n}\right)\mathbf{1}_{l_i}.$$

Then the inequality (6.2) is proven.

proof of (6.3) *and* (6.4)*:.*
The elements of each row of $\boldsymbol{K}_{jk}$ have the same distribution as IID random variables $x_1, \dots, x_{l_k}$ obeying

$$\mathbb{P}(x_i = 1) = B_{jk}, \quad \mathbb{P}(x_i = 0) = 1 - B_{jk}.$$

Chernoff's inequalities (Lemma 6.3) yields

$$\mathbb{P}\left(\sum_{i=1}^{l_k} x_i \ge l_k B_{jk} + t\right) \ge e^{-\frac{t^2}{2\left(l_k B_{jk} + t/3\right)}}.$$

By letting $t = 2\log n + \sqrt{6B_{jk}l_k\log n}$, with probability at least $1 - \frac{1}{n^3}$, we have

$$\sum_{i=1}^{l_k} x_i \le l_k B_{jk} + 2\log n + \sqrt{6B_{jk}l_k\log n}.$$

By taking the uniform bound for all $\boldsymbol{K}_{jk}, 1 \leq j < k \leq r$, with probability at least $1 - \frac{r}{n^2}$, for all $1 \leq j < k \leq r$,

$$\boldsymbol{K}_{jk}\mathbf{1}_{l_k} \leq \left(l_k B_{jk} + 2\log n + \sqrt{6 l_k B_{jk} \log n}\right)\mathbf{1}_{l_j}.$$

The assumption (6.1) implies $\delta > C\frac{2\log n + \sqrt{6 l_k B_{jk}\log n}}{l_k}$, and then the inequality (6.3) is proven.

Similarly, with probability at least $1 - \frac{r}{n^2}$, for all $1 \leq j < k \leq r$, the inequality (6.4) holds.

proof of (6.5)*:.*
The elements of $\boldsymbol{K}_{jk}$ have the same distribution as a collection of IID random variables $x_1, \ldots, x_{l_k l_j}$ obeying

$$\mathbb{P}(x = 1) = B_{jk}, \quad \mathbb{P}(x = 0) = 1 - B_{jk}.$$

Chernoff's inequalities (Lemma 6.3) implies

$$\mathbb{P}\left(\mathbf{1}_{l_j}^T \boldsymbol{K}_{jk}\mathbf{1}_{l_k} \leq l_k l_j B_{jk} - t\right) \leq e^{-\frac{t^2}{2 l_k l_j B_{jk}}}.$$

Then, with probability at least $1 - \frac{1}{n}$, we have

$$\mathbf{1}_{l_j}^T \boldsymbol{K}_{jk}\mathbf{1}_{l_k} \geq l_k l_j B_{jk} - \sqrt{2 l_k l_j B_{jk}\log n}.$$

By the assumption (6.1), there holds $\delta > C\sqrt{\frac{2B_{jk}\log n}{l_k l_j}}$, which implies the inequality (6.3).

0.2. *Proof of Lemma 6.7.* First we prove a fact about zero-mean Bernoulli random variable. Suppose u is a zero-mean Bernoulli random variable which satisfies $\mathbb{P}(u = -\rho) = 1 - \rho$ and $\mathbb{P}(u = 1 - \rho) = \rho$. Then it is straightforward to calculate that

$$\mathrm{Var}(u) = \rho^2(1-\rho) + (1-\rho)^2\rho = \rho(1-\rho) \leq \rho.$$

Now let us prove the lemma. By the calculation of the variances of zero-mean Bernoulli random variables, $B_{ii}\left(\boldsymbol{J}_{l_i} - \boldsymbol{I}_{l_i}\right) - \boldsymbol{K}_{ii}$ satisfies the condition in Corollary 6.5 with $\sigma = \sqrt{B_{ii}}$. Therefore, with probability at least $1 - \sum_{i=1}^r \frac{c}{l_i^4}$, we have

$$\|B_{ii}\left(\boldsymbol{J}_{l_i} - \boldsymbol{I}_{l_i}\right) - \boldsymbol{K}_{ii}\| \leq C_0\left(\sqrt{l_i B_{ii}\log l_i} + \log l_i\right), \quad 1 \leq i \leq r,$$

The condition $p^- \geq C\left(\frac{\log n}{n_{\min}}\right)$ implies the inequalities (6.6).

Moreover, $\boldsymbol{U}$ satisfies the condition in Corollary 6.5 with $\sigma = \sqrt{q^+}$. Therefore, with probability at least $1 - \frac{c}{n^4}$, the inequality (6.7) holds.

0.3. *Proof of Lemma 6.8.* Define $\boldsymbol{A}_1 = \begin{bmatrix} \boldsymbol{K} & \boldsymbol{Z} \\ \boldsymbol{Z}^T & \boldsymbol{W} \end{bmatrix}$ and $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{A}_1\boldsymbol{P}^T$. Moreover, define

$$\boldsymbol{E}_1 = \alpha\boldsymbol{I}_N - (1-\lambda)\boldsymbol{A}_1 + \lambda(\boldsymbol{J}_N - \boldsymbol{I}_N - \boldsymbol{A}_1),$$

and $\boldsymbol{E} = \boldsymbol{P}\boldsymbol{E}_1\boldsymbol{P}^T$. Since $\boldsymbol{P}\boldsymbol{I}_N\boldsymbol{P}^T = \boldsymbol{I}_N$ and $\boldsymbol{P}\boldsymbol{J}_N\boldsymbol{P}^T = \boldsymbol{J}_N$, we have

$$\boldsymbol{E} = \alpha\boldsymbol{I}_N - (1-\lambda)\boldsymbol{A} + \lambda(\boldsymbol{J}_N - \boldsymbol{I}_N - \boldsymbol{A})$$

which is in accordance with the definition (2.4). For any $N \times N$ Hermitian matrix $\widetilde{\boldsymbol{X}}$, it is feasible to (2.3) if and only if $\boldsymbol{P}^T\widetilde{\boldsymbol{X}}\boldsymbol{P}$ is feasible to (2.3). Moreover, we have

$$\langle \widetilde{\boldsymbol{X}}, \boldsymbol{E}\rangle = \langle \boldsymbol{P}^T\widetilde{\boldsymbol{X}}\boldsymbol{P}, \boldsymbol{P}^T\boldsymbol{E}\boldsymbol{P}\rangle = \langle \boldsymbol{P}^T\widetilde{\boldsymbol{X}}\boldsymbol{P}, \boldsymbol{E}_1\rangle.$$

This implies that $\widehat{\boldsymbol{X}}$ is a solution to (2.3) if and only if $\boldsymbol{P}^T\widehat{\boldsymbol{X}}\boldsymbol{P}$ is a solution to (2.3) by replacing $\boldsymbol{E}$ with $\boldsymbol{E}_1$, or equivalently, replacing $\boldsymbol{A}$ with $\boldsymbol{A}_1$. Suppose Theorem 3.1 is true for $\boldsymbol{P} = \boldsymbol{I}_n$, which means $\boldsymbol{P}^T\widehat{\boldsymbol{X}}\boldsymbol{P}$ must be of the form

$$\boldsymbol{P}^T\widehat{\boldsymbol{X}}\boldsymbol{P} = \begin{bmatrix} \boldsymbol{J}_{l_1} & & & \widehat{\boldsymbol{Z}}_1 \\ & \ddots & & \vdots \\ & & \boldsymbol{J}_{l_r} & \widehat{\boldsymbol{Z}}_r \\ \widehat{\boldsymbol{Z}}_1^T & \cdots & \widehat{\boldsymbol{Z}}_r^T & \widehat{\boldsymbol{W}} \end{bmatrix}$$

which implies

$$\widehat{\boldsymbol{X}} = \boldsymbol{P} \begin{bmatrix} \boldsymbol{J}_{l_1} & & & \widehat{\boldsymbol{Z}}_1 \\ & \ddots & & \vdots \\ & & \boldsymbol{J}_{l_r} & \widehat{\boldsymbol{Z}}_r \\ \widehat{\boldsymbol{Z}}_1^T & \cdots & \widehat{\boldsymbol{Z}}_r^T & \widehat{\boldsymbol{W}} \end{bmatrix} \boldsymbol{P}^T.$$

Our proof is therefore done.

0.4. *Proof of Lemma 6.9 .* Since $0 < \lambda < 1$, we have

$$\|-(1-\lambda)\boldsymbol{W}+\lambda(\boldsymbol{J}_m-\boldsymbol{I}_m-\boldsymbol{W})\|_{op} \le \|-(1-\lambda)\boldsymbol{W}+\lambda(\boldsymbol{J}_m-\boldsymbol{I}_m-\boldsymbol{W})\|_F \le m,$$

and by the assumption $\alpha \ge 2m$,

$$\widetilde{\boldsymbol{W}} = \alpha\boldsymbol{I}_m - (1-\lambda)\boldsymbol{W} + \lambda(\boldsymbol{J}_m - \boldsymbol{I}_m - \boldsymbol{W}) \succeq (\alpha - m)\boldsymbol{I}_m \succ \boldsymbol{0}. \tag{0.19}$$

This implies that the objective function of (6.10) is strongly convex. The constraint of (6.10) is evidently convex and compact, so the solution exists uniquely.

Obviously, there are feasible points to (6.10) with all inequalities holding strictly. Therefore, by the constraints qualification under the Slater's condition, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ satisfy the KKT condition, which are (6.11), (6.12) and (6.13). By equality (6.11) $\widetilde{\boldsymbol{W}}\boldsymbol{x}_i + \widetilde{\boldsymbol{Z}}_i^T \mathbf{1}_{l_i} = \boldsymbol{\beta}_i - \boldsymbol{\Xi}\boldsymbol{x}_i$ and the inequality (6.13) $\langle \boldsymbol{x}_i, \boldsymbol{\beta}_i \rangle = 0$, we have

$$\boldsymbol{x}_i^T(\widetilde{\boldsymbol{W}} + \boldsymbol{\Xi})\boldsymbol{x}_i = -\boldsymbol{x}_i^T \widetilde{\boldsymbol{Z}}_i^T \mathbf{1}_{l_i} \leq m l_i.$$

Since the matrix $\boldsymbol{\Xi}$ is a diagonal matrix whose diagonal entries are all non-negative, $\widetilde{\boldsymbol{W}} + \boldsymbol{\Xi}$ is positive definite. By Cauchy-Schwarz inequality, for all $1 \leq j, k \leq r$, we have

$$\boldsymbol{x}_j^T(\widetilde{\boldsymbol{W}} + \boldsymbol{\Xi})\boldsymbol{x}_k \leq \left(\boldsymbol{x}_k^T(\widetilde{\boldsymbol{W}} + \boldsymbol{\Xi})\boldsymbol{x}_k\right)^{\frac{1}{2}} \left(\boldsymbol{x}_j^T(\widetilde{\boldsymbol{W}} + \boldsymbol{\Xi})\boldsymbol{x}_j\right)^{\frac{1}{2}} \leq m\sqrt{l_j l_k}.$$

Notice that the equation (6.11) is equivalent to

$$(\alpha \boldsymbol{I}_m + \lambda(\boldsymbol{J}_m - \boldsymbol{I}_m) - \boldsymbol{W} + \boldsymbol{\Xi})\boldsymbol{x}_i = \boldsymbol{\beta}_i - (\lambda l_i)\mathbf{1}_m + \boldsymbol{Z}_i^T \mathbf{1}_{l_i}.$$

Taking its jth row yields

$$(\alpha - \lambda)x_{i_j} + \lambda \sum_{k=1}^m x_{i_k} + \xi_j x_{i_j} + \lambda l_i = \sum_{i=1}^m W_{jk} x_{i_k} + \beta_{i_j} + \boldsymbol{e}^T \boldsymbol{Z}_i^T \mathbf{1}_{l_i}.$$

The non-negativity of $\boldsymbol{W}$ implies (6.15). Finally, since $x_{i_j}\beta_{i_j} = 0$, from the above equality, we know once $\beta_{i_j} > 0$, there holds $\beta_{i_j} \leq (m - 1 + l_i)\lambda$, which implies (6.16).

0.5. *Proof of Lemma 6.10.* Here we provide some intuition why $\boldsymbol{X}$ is a solution to (2.3). There are two objects to notice. One is the objective function $f(\widetilde{\boldsymbol{X}}) = \langle \boldsymbol{X}, \boldsymbol{E} \rangle$, and the other one is the constraint set $\mathcal{M} := \left\{\widetilde{\boldsymbol{X}} : \mathbf{0} \leq \widetilde{\boldsymbol{X}} \leq \boldsymbol{J}, \widetilde{\boldsymbol{X}} \succeq \mathbf{0}\right\}$. To guarantee that $\boldsymbol{X}$ is the solution of (2.3), we need to show that at the point $\boldsymbol{X}$, the level set of $f(\widetilde{\boldsymbol{X}})$ is tangent to the boundary of $\mathcal{M}$. In other words, the normal vector of $f(\widetilde{\boldsymbol{X}})$, i.e., $-\boldsymbol{E}$, lies in the normal cone of the boundary of $\mathcal{M}$ at point $\boldsymbol{X}$.

Now let us investigate the normal vectors of $\mathcal{M}$ at point $\boldsymbol{X}$. Write $\mathcal{M} = \mathcal{M}_1 \cap \mathcal{M}_2$, where $\mathcal{M}_1 := \left\{\widetilde{\boldsymbol{X}} : \mathbf{0} \leq \widetilde{\boldsymbol{X}} \leq \boldsymbol{J}\right\}$ and $\mathcal{M}_2 := \left\{\widetilde{\boldsymbol{X}} : \widetilde{\boldsymbol{X}} \succeq \mathbf{0}\right\}$. Suppose $\boldsymbol{\Lambda}_1$ is a normal vector of $\mathcal{M}_1$ at $\boldsymbol{X}$, then $\boldsymbol{\Lambda}_1$ must have the following property: $\Lambda_{1_{ij}} \leq 0$ if $X_{ij} = 0$, $\Lambda_{1_{ij}} = 0$ if $0 < X_{ij} < 1$ and $\Lambda_{1_{ij}} \geq 0$ if $X_{ij} = 1$. As to $\mathcal{M}_2$, suppose $\boldsymbol{\Lambda}_2$ is a normal vector of $\mathcal{M}_2$ at $\boldsymbol{X}$. Then $\boldsymbol{\Lambda}_2 \preceq \mathbf{0}$ and $\boldsymbol{\Lambda}_2\boldsymbol{X} = \mathbf{0}$. The normal vectors of $\mathcal{M}$ at point $\boldsymbol{X}$ is of the form $\boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2$.

Then we have the equation $-\boldsymbol{E} = \boldsymbol{\Lambda}_1+\boldsymbol{\Lambda}_2$, or equivalently $-\boldsymbol{\Lambda}_2 = \boldsymbol{E}+\boldsymbol{\Lambda}_1$. It is obvious that if (6.17) holds, $\boldsymbol{\Lambda}_2 = -\boldsymbol{\Lambda}$ satisfies the required equation. The only thing to check is that $\beta_{ij} = 0$ when $x_{ij} > 0$ and $\xi_i = 0$ when the ith diagonal entry of $\sum_{i=1}^r \boldsymbol{x}_i\boldsymbol{x}_i^T$ is less than one. If so, we know that $\boldsymbol{\Lambda} - \boldsymbol{E}$ lies in the normal cone of $\mathcal{M}_1$ at the point $\boldsymbol{X}$. As desired, these requirements are assured by (6.12) and (6.13).

Proof. Suppose $\widehat{\boldsymbol{X}}$ is solution to (2.3). We define an $N \times N$ matrix $\boldsymbol{H}$ as follows:

$$\widehat{\boldsymbol{X}} = \boldsymbol{X}+\boldsymbol{H} = \begin{bmatrix} \boldsymbol{J}_{l_1} + \boldsymbol{H}_{11} & \dots & \boldsymbol{H}_{1r} & \mathbf{1}_{l_1}\boldsymbol{x}_1^T + \boldsymbol{H}_1 \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{H}_{1r}^T & \dots & \boldsymbol{J}_{l_r} + \boldsymbol{H}_{rr} & \mathbf{1}_{l_r}\boldsymbol{x}_r^T + \boldsymbol{H}_r \\ \boldsymbol{x}_1\mathbf{1}_{l_1}^T + \boldsymbol{H}_1^T & \cdots & \boldsymbol{x}_r\mathbf{1}_{l_r}^T + \boldsymbol{H}_r^T & \boldsymbol{x}_1\boldsymbol{x}_1^T + \cdots + \boldsymbol{x}_r\boldsymbol{x}_r^T + \boldsymbol{H}_0 \end{bmatrix}.$$

As discussed in Section 6.4.1, $\boldsymbol{X}$ is feasible to (2.3). By definition we know $\boldsymbol{X} + \boldsymbol{H}$ is also feasible to (2.3). This implies $\boldsymbol{H}_{ii} \leq \mathbf{0}$ for $i = 1, \dots, r$ and $\boldsymbol{H}_{jk} \geq \mathbf{0}$ for $1 \leq j < k \leq r$.

By the feasibility of both $\boldsymbol{X}$ and $\boldsymbol{X} + \boldsymbol{H}$, and the optimality of $\boldsymbol{X} + \boldsymbol{H}$ to (2.3), we have $\langle \boldsymbol{X} + \boldsymbol{H}, \boldsymbol{E}\rangle \leq \langle \boldsymbol{X}, \boldsymbol{E}\rangle$, which implies

$$\langle \boldsymbol{H}, \boldsymbol{E}\rangle \leq 0. \tag{0.20}$$

Define

$$\boldsymbol{\Upsilon} := \begin{bmatrix} -\boldsymbol{\Psi}_{11} & \dots & \boldsymbol{\Phi}_{1r} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\Phi}_{1r}^T & \dots & -\boldsymbol{\Psi}_{rr} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \boldsymbol{\Gamma} := \begin{bmatrix} \mathbf{0} & \dots & \mathbf{0} & \frac{1}{l_1}\mathbf{1}_{l_1}\boldsymbol{\beta}_1^T \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \frac{1}{l_r}\mathbf{1}_{l_r}\boldsymbol{\beta}_r^T \\ \frac{1}{l_1}\boldsymbol{\beta}_1\mathbf{1}_{l_1}^T & \dots & \frac{1}{l_r}\boldsymbol{\beta}_r\mathbf{1}_{l_r}^T & -\boldsymbol{\Xi} \end{bmatrix}.$$

Then we have

$$\boldsymbol{E} = \boldsymbol{\Lambda} + \boldsymbol{\Upsilon} + \boldsymbol{\Gamma}.$$

The inequality (0.20) is equivalent to

$$\langle \boldsymbol{H}, \boldsymbol{\Lambda} + \boldsymbol{\Upsilon} + \boldsymbol{\Gamma}\rangle \leq 0.$$

In the sequel we intend to prove that $\langle \mathbf{H}, \boldsymbol{\Gamma}\rangle \geq \mathbf{0}$, $\langle \mathbf{H}, \boldsymbol{\Lambda}\rangle \geq \mathbf{0}$ and $\langle \mathbf{H}, \boldsymbol{\Upsilon}\rangle \geq \mathbf{0}$:

Proof of $\langle \mathbf{H}, \mathbf{\Gamma} \rangle \geq \mathbf{0}$:.
By the feasibility of $\boldsymbol{X}+\boldsymbol{H}$ and the non-negativity of $\boldsymbol{\beta}_i$, $1 \leq i \leq r$, we have

$$\left\langle \mathbf{1}_{l_i}\boldsymbol{x}_i^T + \boldsymbol{H}_i, \frac{1}{l_i}\mathbf{1}_{l_i}\boldsymbol{\beta}_i^T \right\rangle \geq 0.$$

By (6.13), i.e., $\langle \boldsymbol{x}_i, \boldsymbol{\beta}_i \rangle = 0$, we have

$$\left\langle \boldsymbol{H}_i, \frac{1}{l_i}\mathbf{1}_{l_i}\boldsymbol{\beta}_i^T \right\rangle \geq 0.$$

On the other hand, by the feasibility of $\boldsymbol{X}+\boldsymbol{H}$ and the non-negativity of $\mathbf{\Xi}$, we have

$$\left\langle \boldsymbol{J}_m - \left(\boldsymbol{x}_1\boldsymbol{x}_1^T + \cdots + \boldsymbol{x}_r\boldsymbol{x}_r^T + \boldsymbol{H}_0\right), \mathbf{\Xi} \right\rangle \geq 0.$$

By (6.12), i.e.

$$\left\langle \boldsymbol{J}_m - \left(\boldsymbol{x}_1\boldsymbol{x}_1^T + \cdots + \boldsymbol{x}_r\boldsymbol{x}_r^T\right), \mathbf{\Xi} \right\rangle = 0,$$

we have

$$\langle \boldsymbol{H}_0, -\mathbf{\Xi} \rangle \geq 0.$$

In summary, we have

$$\langle \boldsymbol{H}, \mathbf{\Gamma} \rangle \geq 0.$$

Proof of $\langle \mathbf{H}, \mathbf{\Lambda} \rangle \geq \mathbf{0}$:.
By the feasibility condition $\boldsymbol{X}+\boldsymbol{H} \succeq \mathbf{0}$ and $\mathbf{\Lambda}\boldsymbol{V} = \mathbf{0}$, we have

$$0 \leq \langle \boldsymbol{X}+\boldsymbol{H}, \mathbf{\Lambda} \rangle \leq \langle \boldsymbol{V}\boldsymbol{V}^T + \boldsymbol{H}, \mathbf{\Lambda} \rangle = \langle \boldsymbol{H}, \mathbf{\Lambda} \rangle.$$

Proof of $\langle \mathbf{H}, \mathbf{\Upsilon} \rangle \geq \mathbf{0}$:.
By the facts $\boldsymbol{H}_{ii} \leq \mathbf{0}$ and $\mathbf{\Psi}_{ii} > \mathbf{0}$ for $i = 1, \ldots, r$, we have

$$\langle \boldsymbol{H}_{ii}, -\mathbf{\Psi}_{ii} \rangle \geq 0.$$

Moreover, by the facts $\boldsymbol{H}_{jk} \geq \mathbf{0}$ and $\mathbf{\Phi}_{jk} > \mathbf{0}$ for $i = 1 \leq j < k \leq r$, we have

$$\langle \boldsymbol{H}_{jk}, \mathbf{\Phi}_{jk} \rangle \geq 0.$$

Consequently, we have

$$\langle \boldsymbol{H}, \mathbf{\Upsilon} \rangle \geq 0.$$

In conclusion, we have proven $\langle \mathbf{H}, \mathbf{\Gamma} \rangle \geq \mathbf{0}$, $\langle \mathbf{H}, \mathbf{\Lambda} \rangle \geq \mathbf{0}$ and $\langle \mathbf{H}, \mathbf{\Upsilon} \rangle \geq \mathbf{0}$. Since we also have proven $\langle \boldsymbol{H}, \mathbf{\Lambda} + \mathbf{\Upsilon} + \mathbf{\Gamma} \rangle \leq 0$, we know equalities hold in all these inequalities. In particular, we have $\langle \boldsymbol{H}_{ii}, -\mathbf{\Psi}_{ii} \rangle = 0$ and $\langle \boldsymbol{H}_{jk}, \mathbf{\Phi}_{jk} \rangle = 0$. The nonpositivity of $\boldsymbol{H}_{ii}$ and the strict positivity of $\mathbf{\Psi}_{ii}$ imply that $\boldsymbol{H}_{ii} = \mathbf{0}$. Similarly, the nonnegativity of $\boldsymbol{H}_{jk}$, $j < k$ and the strict positivity of $\boldsymbol{\Phi}_{jk}$ imply that $\boldsymbol{H}_{jk} = \mathbf{0}$. Therefore, $\widehat{\boldsymbol{X}}$ is of the form (3.3).

There is a byproduct: $\langle \boldsymbol{H}, \boldsymbol{\Lambda} + \boldsymbol{\Upsilon} + \boldsymbol{\Gamma} \rangle = \langle \boldsymbol{H}, \boldsymbol{E} \rangle = 0$ implies $\langle \widehat{\boldsymbol{X}}, \boldsymbol{E} \rangle = \langle \boldsymbol{X}, \boldsymbol{E} \rangle$. By the optimality of $\widehat{\boldsymbol{X}}$ and feasibility of $\boldsymbol{X}$ in (2.3), $\boldsymbol{X}$ is also a solution to this optimization problem. □

0.6. *Proof of Lemma 6.11.* We first give candidates of $\boldsymbol{\Psi}_{ii}$ for $1 \le i \le r$ and $\boldsymbol{\Phi}_{jk}$ for $1 \le j < k \le r$ and hence $\boldsymbol{\Lambda}$, such that $\boldsymbol{\Lambda}$ is a particular solution to $\boldsymbol{\Lambda V} = \mathbf{0}$. After that, we prove our constructed $\boldsymbol{\Lambda}$ satisfies other inequalities required in Lemma 6.10.

The equality $\boldsymbol{\Lambda V} = \mathbf{0}$ amounts to

$$\boldsymbol{\Lambda v}_i = \mathbf{0}, \quad i = 1, \dots, r;$$

that is, for all $i = 1, \dots, r$ and $1 \le j < k \le r$,

$$\begin{cases} \left(\widetilde{\boldsymbol{Z}}_i^T - \frac{1}{l_i}\boldsymbol{\beta}_i \mathbf{1}_{l_i}^T\right)\mathbf{1}_{l_i} + \left(\widetilde{\boldsymbol{W}} + \boldsymbol{\Xi}\right)\boldsymbol{x}_i = \mathbf{0}, & (0.21) \\ \left((\alpha - \lambda)\boldsymbol{I}_{l_i} + \lambda \boldsymbol{J}_{l_i} - \boldsymbol{K}_{ii} + \boldsymbol{\Psi}_{ii}\right)\mathbf{1}_{l_i} + \left(\widetilde{\boldsymbol{Z}}_i - \frac{1}{l_i}\mathbf{1}_{l_i}\boldsymbol{\beta}_i^T\right)\boldsymbol{x}_i = \mathbf{0}, & (0.22) \\ \left(\lambda \boldsymbol{J}_{(l_j,l_k)} - \boldsymbol{K}_{jk} - \boldsymbol{\Phi}_{jk}\right)\mathbf{1}_{l_k} + \left(\widetilde{\boldsymbol{Z}}_j - \frac{1}{l_j}\mathbf{1}_{l_j}\boldsymbol{\beta}_j^T\right)\boldsymbol{x}_k = 0, & (0.23) \\ \left(\lambda \boldsymbol{J}_{(l_k,l_j)} - \boldsymbol{K}_{jk}^T - \boldsymbol{\Phi}_{jk}^T\right)\mathbf{1}_{l_j} + \left(\widetilde{\boldsymbol{Z}}_k - \frac{1}{l_k}\mathbf{1}_{l_k}\boldsymbol{\beta}_k^T\right)\boldsymbol{x}_j = 0. & (0.24) \end{cases}$$

Obviously, (0.21) is equivalent to the equation (6.11). In the following, we will construct $\boldsymbol{\Psi}_{ii}$ satisfying (0.22) and $\boldsymbol{\Phi}_{jk}$ satisfying both (0.23) and (0.24).

First, let us give $\boldsymbol{\Psi}_{ii}$ explicitly for $i = 1, \dots, r$. The equality (0.22) is equivalent to

$$\begin{aligned} \boldsymbol{\Psi}_{ii}\mathbf{1}_{l_i} &= -\left((\alpha - \lambda)\boldsymbol{I}_{l_i} + \lambda \boldsymbol{J}_{l_i} - \boldsymbol{K}_{ii}\right)\mathbf{1}_{l_i} - \left(\widetilde{\boldsymbol{Z}}_i - \frac{1}{l_i}\mathbf{1}_{l_i}\boldsymbol{\beta}_i^T\right)\boldsymbol{x}_i \\ &= (-(\alpha - \lambda) - \lambda l_i)\mathbf{1}_{l_i} + \boldsymbol{K}_{ii}\mathbf{1}_{l_i} - \widetilde{\boldsymbol{Z}}_i \boldsymbol{x}_i \\ &= (-(\alpha - \lambda) - \lambda l_i)\mathbf{1}_{l_i} + \boldsymbol{K}_{ii}\mathbf{1}_{l_i} - (\lambda \boldsymbol{J}_{(l_i,m)} - \boldsymbol{Z}_i)\boldsymbol{x}_i, \end{aligned}$$

where the second equality is due to (6.13), i.e., $\boldsymbol{x}_i^T \boldsymbol{\beta}_i = 0$. Since we need to construct an $\boldsymbol{\Psi}_{ii} > \mathbf{0}$, we propose a candidate of the form $\boldsymbol{\Psi}_{ii} = \tau \boldsymbol{J}_{ii} + \boldsymbol{D}_{ii}$, where $\boldsymbol{D}_{ii}$ is a diagonal matrix. It is easy to verify that

(0.25)

$$\boldsymbol{\Psi}_{ii} := Diag(\boldsymbol{K}_{ii}\mathbf{1}_{l_i} + \boldsymbol{Z}_i \boldsymbol{x}_i) + \frac{\delta}{16 l_i}\boldsymbol{J}_{l_i} - \left(\lambda(\boldsymbol{x}_i^T \mathbf{1}_m) + (l_i - 1)\lambda + \alpha + \frac{\delta}{16}\right)\boldsymbol{I}_{l_i}$$

satisfies the above equality constraint.

Next, let us construct $\boldsymbol{\Phi}_{jk}$ satisfying both (0.23) and (0.24). The equality (0.23) is equivalent to

$$\begin{aligned}\boldsymbol{\Phi}_{jk}\mathbf{1}_{l_k} &= \left(\lambda l_k - \frac{\boldsymbol{\beta}_j^T\boldsymbol{x}_k}{l_j}\right)\mathbf{1}_{l_j} - \boldsymbol{K}_{jk}\mathbf{1}_{l_k} + \widetilde{\boldsymbol{Z}}_j\boldsymbol{x}_k \\ &= \left(\lambda l_k - \frac{(\mathbf{1}_{l_j}^T\widetilde{\boldsymbol{Z}}_j + \boldsymbol{x}_j^T(\boldsymbol{\Xi}+\widetilde{\boldsymbol{W}}))\boldsymbol{x}_k}{l_j}\right)\mathbf{1}_{l_j} - \boldsymbol{K}_{jk}\mathbf{1}_{l_k} + \widetilde{\boldsymbol{Z}}_j\boldsymbol{x}_k := \boldsymbol{a}\end{aligned}$$

where the second equality is due to (6.11). Similarly, the equality (0.24) is equivalent to

$$\begin{aligned}\boldsymbol{\Phi}_{jk}^T\mathbf{1}_{l_j} &= \left(\lambda l_j - \frac{\boldsymbol{\beta}_k^T\boldsymbol{x}_j}{l_k}\right)\mathbf{1}_{l_k} - \boldsymbol{K}_{jk}^T\mathbf{1}_{l_j} + \widetilde{\boldsymbol{Z}}_k\boldsymbol{x}_j \\ &= \left(\lambda l_j - \frac{(\mathbf{1}_{l_k}^T\widetilde{\boldsymbol{Z}}_k + \boldsymbol{x}_k^T(\boldsymbol{\Xi}+\widetilde{\boldsymbol{W}}))\boldsymbol{x}_j}{l_k}\right)\mathbf{1}_{l_k} - \boldsymbol{K}_{jk}^T\mathbf{1}_{l_j} + \widetilde{\boldsymbol{Z}}_k\boldsymbol{x}_j := \boldsymbol{b}\end{aligned}$$

A necessary condition of the existence of such matrix $\boldsymbol{\Phi}_{jk}$ is that

$$\mathbf{1}_{l_j}^T\boldsymbol{a} = \mathbf{1}_{l_k}^T\boldsymbol{b}.$$

This is easy to check. In fact, by the above formulas of $\boldsymbol{a}$ and $\boldsymbol{b}$, we have

$$\mathbf{1}_{l_j}^T\boldsymbol{a} = \lambda l_k l_j - \boldsymbol{x}_k^T(\boldsymbol{\Xi}+\widetilde{\boldsymbol{W}}))\boldsymbol{x}_j - \mathbf{1}_{l_j}^T\boldsymbol{K}_{jk}\mathbf{1}_{l_k} = \mathbf{1}_{l_k}^T\boldsymbol{b}.$$

We denote $s = \lambda l_k l_j - \boldsymbol{x}_k^T(\boldsymbol{\Xi}+\widetilde{\boldsymbol{W}}))\boldsymbol{x}_j - \mathbf{1}_{l_j}^T\boldsymbol{K}_{jk}\mathbf{1}_{l_k} = \mathbf{1}_{l_j}^T\boldsymbol{\Phi}_{jk}\mathbf{1}_{l_k}$. It is easy to check that one particular solution to the linear system $\boldsymbol{\Phi}_{jk}\mathbf{1}_{l_k} = \boldsymbol{a}$ and $\boldsymbol{\Phi}_{jk}^T\mathbf{1}_{l_j} = \boldsymbol{b}$ is

$$\boldsymbol{\Phi}_{jk} = \frac{1}{l_k}\boldsymbol{a}\mathbf{1}_{l_k}^T + \frac{1}{l_j}\mathbf{1}_{l_j}\boldsymbol{b}^T - \frac{s}{l_j l_k}\boldsymbol{J}_{(l_j,l_k)}.$$

After simplification, we have

$$\begin{aligned}\boldsymbol{\Phi}_{jk} &:= \left(\frac{1}{l_j}\mathbf{1}_{l_j}\boldsymbol{x}_j^T\widetilde{\boldsymbol{Z}}_k^T + \frac{1}{l_k}\widetilde{\boldsymbol{Z}}_j\boldsymbol{x}_k\mathbf{1}_{l_k}^T\right) - \left(\frac{1}{l_j}\boldsymbol{J}_{l_j}\boldsymbol{K}_{jk} + \frac{1}{l_k}\boldsymbol{K}_{jk}\boldsymbol{J}_{l_k}\right) \\ &\quad + \frac{1}{l_k l_j}(l_k l_j\lambda + \mathbf{1}_{l_j}^T\boldsymbol{K}_{jk}\mathbf{1}_{l_k} - \mathbf{1}_{l_k}^T\widetilde{\boldsymbol{Z}}_k\boldsymbol{x}_j - \mathbf{1}_{l_j}^T\widetilde{\boldsymbol{Z}}_j\boldsymbol{x}_k - \boldsymbol{x}_j^T(\boldsymbol{\Xi}+\widetilde{\boldsymbol{W}})\boldsymbol{x}_k)\boldsymbol{J}_{(l_j,l_k)}. \qquad (0.26)\end{aligned}$$

It suffices to prove that $\boldsymbol{\Psi}_{ii} > \mathbf{0}$, $\boldsymbol{\Phi}_{jk} > \mathbf{0}$, and $\boldsymbol{\Lambda} \succeq \mathbf{0}$. We will prove these constraints one by one. By the assumption $p^- \geq C\left(\frac{\log n}{n_{\min}}\right)$ and $p^- > q^+$, we have the

$$\delta > C\sqrt{\frac{p^- \log n}{n_{\min}}} \geq C\left(\sqrt{\frac{q^+ \log n}{n_{\min}}} + \frac{\log n}{n_{\min}}\right).$$

Therefore, with probability at least $1 - \frac{1}{n} - \frac{2r}{n^2} - \frac{cr}{n_{\min}^4}$, the inequalities (6.2), (6.3), (6.4) and (6.5) in Lemma 6.6, as well as the inequality (6.7) in Lemma 6.7 hold. Next, we prove the inequalities $\boldsymbol{\Psi}_{ii} > \mathbf{0}$, $\boldsymbol{\Phi}_{jk} > \mathbf{0}$, and $\boldsymbol{\Lambda} \succeq \mathbf{0}$ in the following three steps.

Step 1: $\boldsymbol{\Psi}_{ii} > \mathbf{0}$.
By the inequality (6.2) $\boldsymbol{K}_{ii}\mathbf{1}_{l_i} \geq \left((l_i-1)B_{ii} - 2\sqrt{(l_i-1)B_{ii}\log n}\right)\mathbf{1}_{l_i}$, $\boldsymbol{Z}_i \geq \mathbf{0}$, $\boldsymbol{x}_i \geq \mathbf{0}$ and $\alpha > m > \lambda m$, we have

$$\begin{aligned}
&\boldsymbol{\Psi}_{ii} - \frac{\delta}{16 l_i}\boldsymbol{J}_{l_i} \\
&= \mathrm{Diag}(\boldsymbol{K}_{ii}\mathbf{1}_{l_i} + \boldsymbol{Z}_i\boldsymbol{x}_i) - \left(\lambda(\boldsymbol{x}_i^T\mathbf{1}_{l_m}) + (l_i-1)\lambda + \alpha + \frac{\delta}{16}\right)\boldsymbol{I}_{l_i} \\
&\geq \text{and} \succeq \left(\left((l_i-1)B_{ii} - 2\sqrt{(l_i-1)B_{ii}\log n}\right) - \left((l_i-1)\lambda + 2\alpha + \frac{\delta}{16}\right)\right)\boldsymbol{I}_{l_i} \\
&= \left(\left((l_i-1)(B_{ii}-\lambda) - 2\sqrt{(l_i-1)B_{ii}\log n}\right) - \left(2\alpha + \frac{\delta}{16}\right)\right)\boldsymbol{I}_{l_i} \\
&:= f(\sqrt{B_{ii}})\boldsymbol{I}_{l_i},
\end{aligned}$$

where f is a quadratic function. By the basic properties of quadratic functions, the fact $B_{ii} \geq p^- \geq C\left(\frac{\log n}{n_{\min}}\right)$ implies $f(\sqrt{B_{ii}}) \geq f(\sqrt{p^-})$. Then

$$\begin{aligned}
&\boldsymbol{\Psi}_{ii} - \frac{\delta}{16 l_i}\boldsymbol{J}_{l_i} \\
&\geq \text{and} \succeq \left(\left((l_i-1)(p^- - \lambda) - 2\sqrt{(l_i-1)p^-\log n}\right) - \left(2\alpha + \frac{\delta}{16}\right)\right)\boldsymbol{I}_{l_i} \\
&\geq \text{and} \succeq \left(\left((l_i-1)\frac{\delta}{4} - 2\sqrt{(l_i-1)p^-\log n}\right) - \left(2\alpha + \frac{\delta}{16}\right)\right)\boldsymbol{I}_{l_i}.
\end{aligned}$$

The assumption $\delta > C\left(\sqrt{\frac{p^-\log n}{n_{\min}}} + \frac{\alpha}{n_{\min}}\right)$ implies $\boldsymbol{\Psi}_{ii} - \frac{\delta}{16 l_i}\boldsymbol{J}_{l_i} \geq \text{and} \succeq \mathbf{0}$, and hence we have $\boldsymbol{\Psi}_{ii} > \mathbf{0}$. As a byproduct, we have

$$\boldsymbol{\Psi}_{ii} \succeq \left(\left((l_i-1)(B_{ii}-\lambda) - 2\sqrt{(l_i-1)B_{ii}\log n}\right) - \left(2\alpha + \frac{\delta}{16}\right)\right)\boldsymbol{I}_{l_i}.$$

Step 2:. $\boldsymbol{\Phi}_{jk} > \mathbf{0}$

Recall the definition $\widetilde{\boldsymbol{Z}}_i = \lambda \boldsymbol{J}_{(l_i,m)} - \boldsymbol{Z}_i$. Since $\boldsymbol{Z}_i$ comes from the adjacency matrix, and $0 < \lambda < 1$, we have $\|\widetilde{\boldsymbol{Z}}_i\|_\infty \leq 1$. Notice that $\boldsymbol{\Phi}_{jk}$ can be represented as a sum of four terms as follows:

$$\begin{aligned}\boldsymbol{\Phi}_{jk} := &\left(\lambda + \frac{1}{l_k l_j}\mathbf{1}_{l_j}^T \boldsymbol{K}_{jk}\mathbf{1}_{l_k}\right)\boldsymbol{J}_{(l_j,l_k)} + \left(\frac{1}{l_j}\mathbf{1}_{l_j}\boldsymbol{x}_j^T\widetilde{\boldsymbol{Z}}_k^T + \frac{1}{l_k}\widetilde{\boldsymbol{Z}}_j\boldsymbol{x}_k\mathbf{1}_{l_k}^T\right)\\ &- \left(\frac{1}{l_j}\boldsymbol{J}_{l_j}\boldsymbol{K}_{jk} + \frac{1}{l_k}\boldsymbol{K}_{jk}\boldsymbol{J}_{l_k}\right)\\ &- \frac{1}{l_k l_j}(\mathbf{1}_{l_k}^T\widetilde{\boldsymbol{Z}}_k\boldsymbol{x}_j + \mathbf{1}_{l_j}^T\widetilde{\boldsymbol{Z}}_j\boldsymbol{x}_k + \boldsymbol{x}_j^T(\boldsymbol{\Xi} + \widetilde{\boldsymbol{W}})\boldsymbol{x}_k)\boldsymbol{J}_{(l_j,l_k)}.\end{aligned}$$

We will give the lower bound of the first term and give upper bounds to the infinity norms of the later three terms. By (6.5), i.e. $\mathbf{1}_{l_j}^T\boldsymbol{K}_{jk}\mathbf{1}_{l_k} \geq \left(B_{jk} - \frac{\delta}{16}\right) l_k l_j$, and the assumption $\lambda \geq B_{jk} + \frac{\delta}{4}$, we have

$$\left(\lambda + \frac{1}{l_k l_j}\mathbf{1}_{l_j}^T\boldsymbol{K}_{jk}\mathbf{1}_{l_k}\right)\boldsymbol{J}_{(l_j,l_k)} \geq \left(2B_{jk} + \frac{3\delta}{16}\right)\boldsymbol{J}_{(l_j,l_k)}.$$

Since $\|\widetilde{\boldsymbol{Z}}_i\|_\infty \leq 1$ and $\|\boldsymbol{x}_i\|_\infty \leq 1$, we have

$$\left\|\frac{1}{l_j}\mathbf{1}_{l_j}\boldsymbol{x}_j^T\widetilde{\boldsymbol{Z}}_k^T + \frac{1}{l_k}\widetilde{\boldsymbol{Z}}_j\boldsymbol{x}_k\mathbf{1}_{l_k}^T\right\|_\infty \leq \frac{2m}{n_{\min}}.$$

By inequality (6.3) $\boldsymbol{K}_{jk}\mathbf{1}_{l_k} \leq \left(B_{jk} + \frac{\delta}{16}\right) l_k\mathbf{1}_{l_j}$ and inequality (6.4) $\boldsymbol{K}_{jk}^T\mathbf{1}_{l_j} \leq \left(B_{jk} + \frac{\delta}{16}\right) l_j\mathbf{1}_{l_k}$, we have

$$\frac{1}{l_j}\boldsymbol{J}_{l_j}\boldsymbol{K}_{jk} + \frac{1}{l_k}\boldsymbol{K}_{jk}\boldsymbol{J}_{l_k} \leq \left(2B_{jk} + \frac{\delta}{8}\right)\boldsymbol{J}_{(l_j,l_k)}.$$

By inequality (6.14), i.e., $\boldsymbol{x}_j^T(\widetilde{\boldsymbol{W}} + \boldsymbol{\Xi})\boldsymbol{x}_k \leq m\sqrt{l_j l_k}$, we have

$$\begin{aligned}&\left\|\frac{1}{l_k l_j}(\mathbf{1}_{l_k}^T\widetilde{\boldsymbol{Z}}_k\boldsymbol{x}_j + \mathbf{1}_{l_j}^T\widetilde{\boldsymbol{Z}}_j\boldsymbol{x}_k + \boldsymbol{x}_j^T(\boldsymbol{\Xi} + \widetilde{\boldsymbol{W}})\boldsymbol{x}_k)\boldsymbol{J}_{(l_j,l_k)}\right\|_\infty\\ &\leq \frac{ml_k + ml_j + m\sqrt{l_k l_j}}{l_k l_j} \leq \frac{3m}{n_{\min}}.\end{aligned}$$

By adding these four terms together, we have

$$\boldsymbol{\Phi}_{jk} \geq \left(\frac{\delta}{16} - \frac{5m}{n_{\min}}\right)\boldsymbol{J}_{(l_j,l_k)}.$$

By the assumption $\delta \geq C\frac{\alpha}{n_{\min}} > C\frac{m}{n_{\min}}$, we have $\boldsymbol{\Phi}_{jk} > \mathbf{0}$.

Step 3:. $\boldsymbol{\Lambda} \succeq \mathbf{0}$

Suppose the eigenvalues of $\boldsymbol{\Lambda}$ are $\lambda_1(\boldsymbol{\Lambda}) \geq \ldots \geq \lambda_N(\boldsymbol{\Lambda})$. The condition $\boldsymbol{\Lambda V} = \mathbf{0}$ implies $\text{rank}(\boldsymbol{\Lambda}) \leq n - r$. Assuming $\lambda_{N-r}(\boldsymbol{\Lambda}) > 0$, we must have $\lambda_{N-r+1} = \ldots = \lambda_N = 0$, and hence $\boldsymbol{\Lambda} \succeq \mathbf{0}$. Therefore, it suffices to prove $\lambda_{N-r}(\boldsymbol{\Lambda}) > 0$. We first define

$$\widehat{\boldsymbol{V}} = \begin{bmatrix} \frac{1}{\sqrt{l_1}}\mathbf{1}_{l_1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sqrt{l_2}}\mathbf{1}_{l_2} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \frac{1}{\sqrt{l_r}}\mathbf{1}_{l_r} \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{N\times r}.$$

Obviously, $\widehat{\boldsymbol{V}}$ is a basis matrix, i.e., the columns of $\widehat{\boldsymbol{V}}$ are a orthonormal basis of the column space of $\widehat{\boldsymbol{V}}$. Define $\widehat{\boldsymbol{V}}_\perp \in \mathbb{R}^{N\times(N-r)}$, such that $\boldsymbol{U} = [\widehat{\boldsymbol{V}}_\perp, \widehat{\boldsymbol{V}}]$ is an orthogonal matrix. Define

$$\widetilde{\boldsymbol{\Lambda}} := \begin{bmatrix} \widetilde{\boldsymbol{\Lambda}}_1 & \widetilde{\boldsymbol{\Lambda}}_2 \\ \widetilde{\boldsymbol{\Lambda}}_2^T & \widetilde{\boldsymbol{W}} + \boldsymbol{\Xi} \end{bmatrix}$$

$$:= \begin{bmatrix} (\alpha-\lambda)\boldsymbol{I}_{l_1} + B_{11}\boldsymbol{J}_{l_1} - \boldsymbol{K}_{11} + \boldsymbol{\Psi}_{11} & \ldots & B_{1r}\boldsymbol{J}_{(l_1,l_r)} - \boldsymbol{K}_{1r} & \widetilde{\boldsymbol{Z}}_1 - \frac{1}{l_1}\mathbf{1}_{l_1}\boldsymbol{\beta}_1^T \\ \vdots & \ddots & \vdots & \vdots \\ B_{1r}\boldsymbol{J}_{(l_r,l_1)} - \boldsymbol{K}_{1r}^T & \ldots & (\alpha-\lambda)\boldsymbol{I}_{l_r} + B_{rr}\boldsymbol{J}_{l_r} - \boldsymbol{K}_{rr} + \boldsymbol{\Psi}_{rr} & \widetilde{\boldsymbol{Z}}_r - \frac{1}{l_r}\mathbf{1}_{l_r}\boldsymbol{\beta}_r^T \\ \widetilde{\boldsymbol{Z}}_1^T - \frac{1}{l_1}\boldsymbol{\beta}_1\mathbf{1}_{l_1}^T & \ldots & \widetilde{\boldsymbol{Z}}_r^T - \frac{1}{l_r}\boldsymbol{\beta}_r\mathbf{1}_{l_r}^T & \widetilde{\boldsymbol{W}} + \boldsymbol{\Xi} \end{bmatrix}.$$

The matrix $\widetilde{\boldsymbol{\Lambda}}$ is closely tied up with $\boldsymbol{\Lambda}$ in the sense that

$$\widetilde{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda} = \begin{bmatrix} (B_{11}-\lambda)\boldsymbol{J}_{l_1} & \ldots & (B_{1r}-\lambda)\boldsymbol{J}_{(l_1,l_r)} + \boldsymbol{\Phi}_{1r} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ (B_{1r}-\lambda)\boldsymbol{J}_{(l_r,l_1)} + \boldsymbol{\Phi}_{1r}^T & \ldots & (B_{rr}-\lambda)\boldsymbol{J}_{l_r} & \mathbf{0} \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

By the construction of $\boldsymbol{\Phi}_{jk}$, it can be written as $\boldsymbol{\Phi}_{jk} = \mathbf{1}_{l_j}\boldsymbol{a}^T + \boldsymbol{b}\mathbf{1}_{l_k}^T$. Therefore, straightforward calculation yields $\widehat{\boldsymbol{V}}_\perp^T(\boldsymbol{\Lambda} - \widetilde{\boldsymbol{\Lambda}})\widehat{\boldsymbol{V}}_\perp = \mathbf{0}$, which implies that

$$\widehat{\boldsymbol{V}}_\perp^T\boldsymbol{\Lambda}\widehat{\boldsymbol{V}}_\perp = \widehat{\boldsymbol{V}}_\perp^T\widetilde{\boldsymbol{\Lambda}}\widehat{\boldsymbol{V}}_\perp.$$

Since $\boldsymbol{U}^T\boldsymbol{\Lambda}\boldsymbol{U} = \begin{bmatrix} \widehat{\boldsymbol{V}}_\perp^T\boldsymbol{\Lambda}\widehat{\boldsymbol{V}}_\perp & \widehat{\boldsymbol{V}}_\perp^T\boldsymbol{\Lambda}\widehat{\boldsymbol{V}} \\ \widehat{\boldsymbol{V}}^T\boldsymbol{\Lambda}\widehat{\boldsymbol{V}}_\perp & \widehat{\boldsymbol{V}}^T\boldsymbol{\Lambda}\widehat{\boldsymbol{V}} \end{bmatrix}$ has the same spectrum as $\boldsymbol{\Lambda}$ does, by Lemma 6.2, there holds

$$\lambda_{N-r}(\boldsymbol{\Lambda}) = \lambda_{N-r}(\boldsymbol{U}^T\boldsymbol{\Lambda}\boldsymbol{U}) \geq \lambda_{N-r}(\widehat{\boldsymbol{V}}_\perp^T\boldsymbol{\Lambda}\widehat{\boldsymbol{V}}_\perp) = \lambda_{N-r}(\widehat{\boldsymbol{V}}_\perp^T\widetilde{\boldsymbol{\Lambda}}\widehat{\boldsymbol{V}}_\perp). \tag{0.27}$$

Since $\boldsymbol{U}^T\widetilde{\boldsymbol{\Lambda}}\boldsymbol{U} = \begin{bmatrix} \widehat{\boldsymbol{V}}_\perp^T\widetilde{\boldsymbol{\Lambda}}\widehat{\boldsymbol{V}}_\perp & \widehat{\boldsymbol{V}}_\perp^T\widetilde{\boldsymbol{\Lambda}}\widehat{\boldsymbol{V}} \\ \widehat{\boldsymbol{V}}^T\widetilde{\boldsymbol{\Lambda}}\widehat{\boldsymbol{V}}_\perp & \widehat{\boldsymbol{V}}^T\widetilde{\boldsymbol{\Lambda}}\widehat{\boldsymbol{V}} \end{bmatrix}$, by Lemma 6.2 again, we have

$$\lambda_{N-r}(\widehat{\boldsymbol{V}}_\perp^T\widetilde{\boldsymbol{\Lambda}}\widehat{\boldsymbol{V}}_\perp) \geq \lambda_N(\boldsymbol{U}^T\widetilde{\boldsymbol{\Lambda}}\boldsymbol{U}) = \lambda_N(\widetilde{\boldsymbol{\Lambda}}). \tag{0.28}$$

By considering the above inequalities (0.27) and (0.28), in order to prove $\boldsymbol{\Lambda} \succeq \boldsymbol{0}$, it suffices to prove $\lambda_N(\widetilde{\boldsymbol{\Lambda}}) > 0$, i.e., $\widetilde{\boldsymbol{\Lambda}} \succ \boldsymbol{0}$.

Define

$$\begin{aligned}
\widetilde{\boldsymbol{\Lambda}}_1 &:= \boldsymbol{F}_1 + \boldsymbol{F}_2 \\
&= \begin{bmatrix} (\alpha-\lambda)\boldsymbol{I}_{l_1} + B_{11}\boldsymbol{J}_{l_1} - \boldsymbol{K}_{11} + \boldsymbol{\Psi}_{11} & \dots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \dots & (\alpha-\lambda)\boldsymbol{I}_{l_r} + B_{11}\boldsymbol{J}_{l_r} - \boldsymbol{K}_{rr} + \boldsymbol{\Psi}_{rr} \end{bmatrix} \\
&+ \begin{bmatrix} \boldsymbol{0} & \dots & B_{1r}\boldsymbol{J}_{(l_1,l_r)} - \boldsymbol{K}_{1r} \\ \vdots & \ddots & \vdots \\ B_{1r}\boldsymbol{J}_{(l_1,l_r)} - \boldsymbol{K}_{1r}^T & \dots & \boldsymbol{0} \end{bmatrix}.
\end{aligned}$$

In the first step, we proved

$$\boldsymbol{\Psi}_{ii} \succeq \left(\left((l_i-1)(B_{ii}-\lambda) - 2\sqrt{(l_i-1)B_{ii}\log n}\right) - \left(2\alpha + \frac{\delta}{16}\right)\right)\boldsymbol{I}_{l_i}.$$

By Lemma 6.7, we have

$$\|B_{ii}\left(\boldsymbol{J}_{l_i} - \boldsymbol{I}_{l_i}\right) - \boldsymbol{K}_{ii}\| \leq C_0\sqrt{l_iB_{ii}\log l_i}.$$

This implies that

$$\begin{aligned}
&(\alpha-\lambda)\boldsymbol{I}_{l_i} + B_{ii}\boldsymbol{J}_{l_i} - \boldsymbol{K}_{ii} + \boldsymbol{\Psi}_{ii} \\
&\succeq \left(\left(l_i(B_{ii}-\lambda) - (C_0+2)\sqrt{l_iB_{ii}\log n}\right) - \left(\alpha + \frac{\delta}{16}\right)\right)\boldsymbol{I}_{l_i} \\
&:= h(\sqrt{B_{ii}})\boldsymbol{I}_{l_i}
\end{aligned}$$

where h is a quadratic function. By basic properties of quadratic functions, the condition $B_{ii} \geq p^- \geq (C_0+2)\left(\frac{\log n}{n_{\min}}\right)$ implies $h(\sqrt{B_{ii}}) \geq h(\sqrt{p^-})$. Therefore,

$$\begin{aligned}
&(\alpha-\lambda)\boldsymbol{I}_{l_i} + B_{ii}\boldsymbol{J}_{l_i} - \boldsymbol{K}_{ii} + \boldsymbol{\Psi}_{ii} \\
&\succeq \left(\left(l_i(p^- - \lambda) - (C_0+2)\sqrt{l_ip^-\log n}\right) - \left(\alpha + \frac{\delta}{16}\right)\right)\boldsymbol{I}_{l_i} \\
&\succeq \left(\left(l_i\frac{\delta}{4} - (C_0+2)\sqrt{l_ip^-\log n}\right) - \left(\alpha + \frac{\delta}{16}\right)\right)\boldsymbol{I}_{l_i} := g(\sqrt{l_i})\boldsymbol{I}_{l_i}
\end{aligned}$$

where g is a quadratic function. By basic properties of quadratic functions, $\sqrt{l_i} \geq \sqrt{n_{\min}} > C\frac{\sqrt{p^- \log n}}{\delta}$ implies $g(\sqrt{l_i}) \geq g(\sqrt{n_{\min}})$. Therefore,

$$\boldsymbol{F}_1 \succeq \left(\left(n_{\min}\frac{\delta}{4} - (C_0+2)\sqrt{n_{\min}p^- \log n} \right) - \left(\alpha + \frac{\delta}{16} \right) \right) \boldsymbol{I}_n.$$

On the other hand, by Lemma 6.7, we have $\|\boldsymbol{F}_2\|_{op} \leq C_0 \left(\sqrt{nq^+ \log n} + \log n \right)$, which implies

$$\begin{aligned}
\widetilde{\boldsymbol{\Lambda}}_1 &= \boldsymbol{F}_1 + \boldsymbol{F}_2 \\
&\succeq \left(\delta\left(\frac{n_{\min}}{4} - \frac{1}{16} \right) - (C_0+2)\sqrt{n_{\min}p^- \log n} - \alpha - C_0\left(\sqrt{nq^+ \log n} + \log n \right) \right) \boldsymbol{I}_n.
\end{aligned}$$

By the assumption $p^- \geq C\frac{\log n}{n_{\min}}$ and

$$\delta > C\left(\sqrt{\frac{p^- \log n}{n_{\min}}} + \frac{\alpha}{n_{\min}} + \frac{\sqrt{nq^+ \log n}}{n_{\min}} \right),$$

when C is large enough, we have

$$\widetilde{\boldsymbol{\Lambda}}_1 \succeq \frac{n_{\min}\delta}{8}\boldsymbol{I}_n.$$

Therefore, in order to guarantee $\widetilde{\boldsymbol{\Lambda}} = \begin{bmatrix} \widetilde{\boldsymbol{\Lambda}}_1 & \widetilde{\boldsymbol{\Lambda}}_2 \\ \widetilde{\boldsymbol{\Lambda}}_2^T & \widetilde{\boldsymbol{W}} + \boldsymbol{\Xi} \end{bmatrix} \succ 0$, it suffices to prove

$$\begin{bmatrix} \frac{n_{\min}\delta}{8}\boldsymbol{I}_n & \widetilde{\boldsymbol{\Lambda}}_2 \\ \widetilde{\boldsymbol{\Lambda}}_2^T & \widetilde{\boldsymbol{W}} + \boldsymbol{\Xi} \end{bmatrix} \succ \boldsymbol{0}.$$

By multiplying $\begin{bmatrix} w\boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_m \end{bmatrix}$ with some $w > 0$ on both sides, it is suffices to prove

$$\widetilde{\widetilde{\boldsymbol{\Lambda}}} := \begin{bmatrix} \frac{n_{\min}\delta w^2}{8}\boldsymbol{I}_n & w\widetilde{\boldsymbol{\Lambda}}_2 \\ w\widetilde{\boldsymbol{\Lambda}}_2^T & \widetilde{\boldsymbol{W}} + \boldsymbol{\Xi} \end{bmatrix} \succ \boldsymbol{0}.$$

Here we choose $w = \frac{33m}{n_{\min}\delta}$. We would like to prove the positive definiteness of $\widetilde{\widetilde{\boldsymbol{\Lambda}}}$ by the well known Gershgorin Theorem, that is, for each row of $\widetilde{\widetilde{\boldsymbol{\Lambda}}}$, the sum of absolute values of the off-diagonal entries is less than the corresponding diagonal entry. Let us first investigate the first n rows of $\widetilde{\widetilde{\boldsymbol{\Lambda}}}$. Recall that

$$w\widetilde{\boldsymbol{\Lambda}}_2 = w\begin{bmatrix} \lambda \boldsymbol{J}_{(l_1,m)} - \boldsymbol{Z}_1 - \frac{1}{l}{}_1 \boldsymbol{1}_{l_1}\boldsymbol{\beta}_1^T \\ \vdots \\ \lambda \boldsymbol{J}_{(l_r,m)} - \boldsymbol{Z}_r - \frac{1}{l}{}_r \boldsymbol{1}_{l_r}\boldsymbol{\beta}_r^T \end{bmatrix}.$$

Since $\mathbf{0} \le \boldsymbol{Z}_i \le \boldsymbol{J}_{(l_i,m)}$ and (6.16), i.e., $\mathbf{0} \le \boldsymbol{\beta}_i \le (m + l_i - 1)\lambda \mathbf{1}_m$, the sum of absolute values of each row of $w\widetilde{\boldsymbol{\Lambda}}_2$ is no larger than

$$w\left(\lambda m + m + \frac{m(m + l_i - 1)\lambda}{l_i}\right) \le 4wm.$$

The above inequality is due to $1 \ge \delta \ge C\frac{\sqrt{r}m}{n_{\min}}$ with sufficiently large C. Since all the diagonal entries are $w^2\frac{n_{\min}\delta}{8}$. Then the Gershgorin condition holds by $w^2\frac{n_{\min}\delta}{8} > 4wm$, which is guaranteed by the definition of w.

Now let us study the bottom m rows of $\widetilde{\widetilde{\boldsymbol{\Lambda}}}$. Notice that

$$w\widetilde{\boldsymbol{\Lambda}}_2 = w\left[\lambda \boldsymbol{J}_{(m,l_1)} - \boldsymbol{Z}_1^T - \tfrac{1}{l_1}\boldsymbol{\beta}_1\mathbf{1}_{l_1}^T, \ldots, \lambda \boldsymbol{J}_{(m,l_r)} - \boldsymbol{Z}_r^T - \tfrac{1}{l_r}\boldsymbol{\beta}_r\mathbf{1}_{l_r}^T\right].$$

By (6.15), i.e.,

$$\beta_{i_j} + \boldsymbol{e}_j^T\boldsymbol{Z}_i^T\mathbf{1}_{l_i} \le (\alpha - \lambda + \xi_j)x_{i_j} + \lambda l_i + \lambda\sum_{k=1}^m x_{i_k},$$

the sum of absolute values of the jth row is no larger than

$$\begin{aligned}
& w\left(n\lambda + \sum_{i=1}^r\left(\boldsymbol{e}_j^T\boldsymbol{Z}_i^T\mathbf{1}_{l_i} + \beta_{i_j}\right)\right) \\
& \le w\left(n\lambda + (\alpha - \lambda + \xi_j)\sum_{i=1}^r x_{i_j} + \lambda n + \lambda\sum_{k=1}^m\sum_{i=1}^r x_{i_k}\right) \\
& \le w\left(n\lambda + (\alpha - \lambda + \xi_j)\sqrt{r}\sqrt{\sum_{i=1}^r x_{i_j}^2} + \lambda n + \lambda\sum_{k=1}^m\sqrt{r}\sqrt{\sum_{i=1}^r x_{i_k}^2}\right) \\
& \le \frac{33m}{n_{\min}\delta}\left((2n + (m-1)\sqrt{r})\lambda + (\alpha + \xi_j)\sqrt{r}\right),
\end{aligned}$$

where the final inequality is due to $\sum_{i=1}^r x_{i_k}^2 \le 1$ for all $k = 1, \ldots, m$. This is the constraint in the optimization (6.10). On the other hand, since

$$\widetilde{\boldsymbol{X}} + \boldsymbol{\Xi} = \alpha\boldsymbol{I}_m + \lambda\left(\boldsymbol{J}_m - \boldsymbol{I}_m\right) - \boldsymbol{W} + \boldsymbol{\Xi},$$

its diagonal entry in the jth row is $\alpha + \xi_j$ while the sum of absolute values of the off-diagonal entries in the jth row is no larger than $m - 1$. Back to the $(n + j)$th row of $\widetilde{\widetilde{\boldsymbol{\Lambda}}}$, the Gershgorin condition holds if

$$\alpha + \xi_j > m - 1 + \frac{33m}{n_{\min}\delta}\left((2n + (m-1)\sqrt{r})\lambda + (\alpha + \xi_j)\sqrt{r}\right)$$

i.e.

$$\left(1 - \frac{33m\sqrt{r}}{n_{\min}\delta}\right)(\alpha + \xi_j) > m - 1 + \frac{33m}{n_{\min\delta}}(2n + (m-1)\sqrt{r})\lambda.$$

By the condition $\delta > C\frac{m\sqrt{r}}{n_{\min}}$ with sufficiently large C, and the fact $\xi_j \geq 0$, the above inequality can be guaranteed by

$$\alpha > 2m + \frac{66m}{n_{\min\delta}}(2n + (m-1)\sqrt{r})\lambda.$$

This can be guaranteed by $\delta > C\frac{nmp^-}{(\alpha-2m)n_{\min}} > C\frac{nm\lambda}{(\alpha-2m)n_{\min}}$, since $m\sqrt{r} < \frac{n_{\min}\delta}{C} < n$ when $C \geq 1$.

In a word, when δ satisfies (6.18), $\widetilde{\widetilde{\mathbf{\Lambda}}} \succ \mathbf{0}$, and then $\widetilde{\mathbf{\Lambda}} \succ \mathbf{0}$, and hence $\mathbf{\Lambda} \succeq \mathbf{0}$.

400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340,
E-mail: tcai@wharton.upenn.edu; xiaodli@wharton.upenn.edu