Submitted to Bernoulli

Supplement to "Entrywise limit theorems for eigenvectors of signal-plus-noise matrix models with weak signals"

FANGZHENG XIE[1,a],

[1]*Department of Statistics, Indiana University, Bloomington, United States.* [a]*fxie@iu.edu*

The supplementary material contains the proofs for Section 3 and Section 4 in the manuscript.

Contents

S1. Technical preparations

The supplementary material begins with several auxiliary results that have already been established in the literature. We first present a theorem due to [1]. It is quite useful to obtain sharp concentration bounds for $\|\mathbf{U}_{\mathbf{A}_+}\|_{2\to\infty}$, $\|\mathbf{U}_{\mathbf{A}_-}\|_{2\to\infty}$, $\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty}$, $\|\mathbf{U}_{\mathbf{A}_-}^{(m)}\|_{2\to\infty}$, $\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathrm{sgn}(\mathbf{H}_+^{(m)}) - \mathbf{U}_{\mathbf{P}_+}\|_{2\to\infty}$, and

$\|\mathbf{U}_{\mathbf{A}_-}^{(m)}\mathrm{sgn}(\mathbf{H}_-^{(m)}) - \mathbf{U}_{\mathbf{P}_-}\|_{2\to\infty}$. Although it can also lead to sharp bounds for $\|\mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*\|_{2\to\infty}$ and $\|\mathbf{U}_{\mathbf{A}_-} - \mathbf{A}\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{W}_-^*\|_{2\to\infty}$ when $\log n \asymp n\rho_n$, it does not provide an enough control of the entrywise error $\|\mathbf{e}_m^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}\mathbf{W}^*)\|_2$ for each individual $m \in [n]$.

Theorem S1.1 (Theorem 2.1 in [1]). *Let $\mathbf{M}$ be an $n\times n$ symmetric random matrices with $\mathbb{E}\mathbf{M} = \mathbf{P}$. Suppose r, s are integers with $1 \le r \le n$, $0 \le s \le n-r$. Let $\mathbf{U}, \mathbf{U}_{\mathbf{P}} \in \mathbb{O}(n,r)$ be the eigenvector matrices of $\mathbf{M}$ and $\mathbf{P}$, respectively, such that $\mathbf{M}\mathbf{U} = \mathbf{U}\mathbf{S}$ and $\mathbf{P}\mathbf{U}_{\mathbf{P}} = \mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}$, where $\mathbf{S} = \mathrm{diag}\{\lambda_{s+1}(\mathbf{M}), \dots, \lambda_{s+r}(\mathbf{M})\}$, and $\mathbf{S}_{\mathbf{P}} = \mathrm{diag}\{\lambda_{s+1}(\mathbf{P}), \dots, \lambda_{s+r}(\mathbf{P})\}$. We adopt the convention that $\lambda_0(\mathbf{P}) = +\infty$ and $\lambda_n(\mathbf{P}) = -\infty$. Define the eigengap*

$$\Delta = \min\{\lambda_s(\mathbf{P}) - \lambda_{s+1}(\mathbf{P}), \lambda_{s+r}(\mathbf{P}) - \lambda_{s+r+1}(\mathbf{P})\} \wedge \min_{k\in[r]} |\lambda_{s+k}(\mathbf{P})|$$

and $\kappa = \max_{k\in[r]} \lambda_{s+k}(\mathbf{P})/\Delta$. Suppose there exists some $\bar{\gamma} > 0$ and a function $\omega(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$, such that the following conditions hold:

- (A1) *(Incoherence) $\|\mathbf{P}\|_{2\to\infty} \le \bar{\gamma}\Delta$.*
- (A2) *(Row and columnwise independence) For any $i \in [n]$, the entries in the ith row and column of $\mathbf{M}$ are independent of others.*
- (A3) *(Spectral norm concentration) $32\kappa \max\{\bar{\gamma}, \omega(\bar{\gamma})\} \le 1$ and $\mathbb{P}(\|\mathbf{M} - \mathbf{P}\|_2 > \bar{\gamma}\Delta) \le \delta_0$ for some $\delta_0 \in (0,1)$.*
- (A4) *(Row concentration) Suppose $\omega(x)$ is non-decreasing in $\mathbb{R}_+$ with $\omega(0) = 0$, $\omega(x)/x$ is non-increasing in $\mathbb{R}_+$. There exists some $\delta_1 \in (0,1)$, such that for all $i \in [n]$ and any $n\times r$ matrix $\mathbf{V}$,*

$$\mathbb{P}\left\{\|\mathbf{e}_i^{\mathrm{T}}(\mathbf{M} - \mathbf{P})\mathbf{V}\|_2 \le \Delta\|\mathbf{V}\|_{2\to\infty}\omega\left(\frac{\|\mathbf{V}\|_{\mathrm{F}}}{\sqrt{n}\|\mathbf{V}\|_{2\to\infty}}\right)\right\} \ge 1 - \frac{\delta_1}{n}.$$

Then with probability at least $1 - \delta_0 - 2\delta_1$, we have

$$\|\mathbf{U}\|_{2\to\infty} \lesssim \{\kappa + \omega(1)\}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty} + \frac{\bar{\gamma}\|\mathbf{P}\|_{2\to\infty}}{\Delta},$$

$$\|\mathbf{U}\mathrm{sgn}(\mathbf{U}^{\mathrm{T}}\mathbf{U}_{\mathbf{P}}) - \mathbf{U}_{\mathbf{P}}\|_{2\to\infty} \lesssim \kappa\{\kappa + \omega(1)\}\{\bar{\gamma} + \omega(\bar{\gamma})\}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty} + \frac{\bar{\gamma}\|\mathbf{P}\|_{2\to\infty}}{\Delta} + \omega(1)\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}.$$

We next state a vector version of the Bernstein's inequality due to [11]. The advantage of this concentration inequality is that it is dimension free.

Lemma S1.2 (Corollary 4.1 in [11]). *Let $\mathbf{y}_1, \dots, \mathbf{y}_n \in \mathbb{R}^d$ be a sequence of independent random vectors such that $\mathbb{E}(\mathbf{y}_i) = \mathbf{0}_d$ and $\|\mathbf{y}_i\|_2 \le U$ almost surely for all $1 \le i \le n$ and some $U > 0$. Denote $\tau^2 := \sum_{i=1}^n \mathbb{E}\|\mathbf{y}_i\|_2^2$. Then for all $t \ge (U + \sqrt{U^2 + 36\tau^2})/6$,*

$$\mathbb{P}\left(\left\|\sum_{i=1}^n \mathbf{y}_i\right\|_2 > t\right) \le 28\exp\left(-\frac{3t^2}{6\tau^2 + 2Ut}\right)$$

Lemma S1.3 below is a generic matrix Chernoff bound due to [17]. In the context of random dot product graphs, it allows us to construct the required function $\omega(\cdot)$ in condition A4 of Theorem S1.1. See Section S6.2 for more details.

Lemma S1.3 (Corollary 3.7 in [17]). *Let $\mathbf{Z}_1,\ldots,\mathbf{Z}_n$ be a sequence of symmetric independent random matrices in $\mathbb{R}^{d\times d}$. Assume that there is a function $g:(0,+\infty)\to[0,+\infty]$ and a sequence of deterministic symmetric matrices $\mathbf{M}_1,\ldots,\mathbf{M}_n\in\mathbb{R}^{d\times d}$ such that $\mathbb{E}e^{\theta\mathbf{Z}_i}\preceq e^{g(\theta)\mathbf{M}_i}$ for all $\theta>0$. Define the scale parameter $\rho=\lambda_{\max}\left(\sum_{i=1}^n\mathbf{M}_n\right)$. Then for all $t\in\mathbb{R}$,*

$$\mathbb{P}\left\{\lambda_{\max}\left(\sum_{i=1}^n\mathbf{Z}_i\right)\geq t\right\}\leq d\inf_{\theta>0}\exp\{-\theta t+g(\theta)\rho\}.$$

We conclude this section with the following Berry-Esseen bound for multivariate nonlinear statistics due to [15], which is useful for us to prove Theorem 3.1 and Theorem 4.6.

Theorem S1.4 (Corollary 2.2 in [15]). *Let $\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_n$ be independent random vectors in $\mathbb{R}^d$ such that $\mathbb{E}\boldsymbol{\xi}_j=\mathbf{0}_d$, $j\in[n]$ and $\sum_{j=1}^n\mathbb{E}(\boldsymbol{\xi}_j\boldsymbol{\xi}_j^{\mathrm{T}})=\mathbf{I}_d$. Let $\mathbf{T}=\sum_{j=1}^n\boldsymbol{\xi}_j+\mathbf{D}(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_n)$ be a nonlinear statistic, where $\mathbf{D}(\cdot)$ is a measurable function from $\mathbb{R}^{n\times d}\to\mathbb{R}^d$. Let O be an event and Δ be a random variable such that $\Delta\geq\|\mathbf{D}(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_n)\|_2\mathbb{1}(O)$, and suppose $\{\Delta^{(j)}\}_{j=1}^n$ are random variables such that $\Delta^{(j)}$ is independent of $\boldsymbol{\xi}_j$, $j\in[n]$. Denote $\gamma:=\sum_{j=1}^n\mathbb{E}(\|\boldsymbol{\xi}_j\|_2^3)$ and $\mathcal{A}$ the collection of all convex measurable sets in $\mathbb{R}^d$. Then*

$$\sup_{A\in\mathcal{A}}|\mathbb{P}(\mathbf{T}\in A)-\mathbb{P}(\mathbf{z}\in A)|\lesssim d^{1/2}\gamma+\mathbb{E}\left\{\left\|\sum_{j=1}^n\boldsymbol{\xi}_j\right\|_2\Delta\right\}+\sum_{j=1}^n\mathbb{E}\{\|\boldsymbol{\xi}_j\|_2|\Delta-\Delta^{(j)}|\}+\mathbb{P}(O^c).$$

S2. Auxiliary results

In this section, we introduce some technical tools that serve as the building blocks for our theory. We first present several useful results that are applied throughout the proofs.

Result S2.1 (Concentration of eigenvalues). Under Assumption 5, by Weyl's inequality, with probability at least $1-c_0n^{-\zeta}$, the p largest eigenvalues of $\mathbf{A}$ are bounded above by $(1/2)n\rho_n\lambda_d(\mathbf{\Delta}_n)$, the q smallest eigenvalues of $\mathbf{A}$ are bounded below by $-(1/2)n\rho_n\lambda_d(\mathbf{\Delta}_n)$, and the absolute values of the remaining eigenvalues of $\mathbf{A}$ are bounded by a constant multiple of $(n\rho_n)^{1/2}$. In other words, for sufficiently large n, with probability at least $1-c_0n^{-\zeta}$,

$$\begin{aligned}
&\lambda_1(\mathbf{A})\geq\ldots\geq\lambda_p(\mathbf{A})\geq\frac{1}{2}\lambda_p(\mathbf{P})=\frac{1}{2}n\rho_n\lambda_p(\mathbf{\Delta}_{n+})\geq\frac{1}{2}n\rho_n\lambda_d(\mathbf{\Delta}_n),\\
&\max_{p+1\leq i\leq n-q}|\lambda_i(\mathbf{A})|\leq K(n\rho_n)^{1/2},\\
&\lambda_n(\mathbf{A})\leq\ldots\leq\lambda_{n-q+1}(\mathbf{A})\leq\frac{1}{2}\lambda_{n-q+1}(\mathbf{P})=-\frac{1}{2}n\rho_n\lambda_q(\mathbf{\Delta}_{n-})\leq-\frac{1}{2}n\rho_n\lambda_d(\mathbf{\Delta}_n).
\end{aligned}\tag{1}$$

Result S2.2 (Concentration of $\mathbf{S}_{\mathbf{A}}$). Suppose Assumptions 1-5 hold. Then for sufficiently large n, $\|\mathbf{S}_{\mathbf{A}}\|_2\leq 2n\rho_n\lambda_1(\mathbf{\Delta}_n)$ and $\|\mathbf{S}_{\mathbf{A}}^{-1}\|_2\leq\{2n\rho_n\lambda_d(\mathbf{\Delta}_n)\}^{-1}$with probability at least $1-c_0n^{-\zeta}$, where $c_0>0,\zeta\geq 1$ are absolute constants. This can be implied by the concentration of eigenvalues in Result S2.1 and Assumption 5.

Result S2.3 (Eigenvector delocalization). $\mathbf{U_P}$ satisfies that $\|\mathbf{U_P}\|_{2\to\infty} \leq \|\mathbf{X}\|_{2\to\infty}/\{n\lambda_d(\mathbf{\Delta}_n)\}^{1/2}$. Consequently, $\lambda_d(\mathbf{\Delta}_n) \leq \|\mathbf{X}\|_{2\to\infty}^2/d$. To see why these results hold, we first observe that

$$\begin{aligned}\|\mathbf{U_P}\|_{2\to\infty} &= \|\rho_n^{1/2}[\mathbf{X}_+,\mathbf{X}_-]\mathrm{diag}(|\mathbf{S}_{\mathbf{P}_+}|,|\mathbf{S}_{\mathbf{P}_-}|)^{-1/2}\|_{2\to\infty}\\ &\leq \rho_n^{1/2}\|\mathbf{X}\|_{2\to\infty}\max(\|\mathbf{S}_{\mathbf{P}_+}^{-1}\|_2,\|\mathbf{S}_{\mathbf{P}_-}^{-1}\|_2)^{1/2} = \frac{\|\mathbf{X}\|_{2\to\infty}}{\sqrt{n}\lambda_d(\mathbf{\Delta}_n)^{1/2}}.\end{aligned}$$

Since $\|\mathbf{U_P}\|_{2\to\infty} \geq \sqrt{d/n}$ by the fact that $\mathbf{U_P} \in \mathbb{O}(n,d)$, we obtain $\lambda_d(\mathbf{\Delta}_n) \leq \|\mathbf{X}\|_{2\to\infty}^2/d$.

We next present a collection of auxiliary lemmas, the proofs of which are relegated to the Supplementary Material. Lemma S2.1 below essentially states the concentration property of $\mathbf{e}_i^{\mathrm{T}}\mathbf{E}\mathbf{V}$ for any deterministic matrix $\mathbf{V} \in \mathbb{R}^{n\times d}$ and can be proved using a matrix Bernstein's inequality [17].

Lemma S2.1. *Let $(y_i)_{i=1}^n$ be independent random variables, $|y_i| \leq 1$ with probability one, and $\max_{i\in[n]}\mathrm{var}(y_i) \leq \sigma^2\rho$ for some constant $\sigma^2 > 0$. Suppose $\mathbf{V} \in \mathbb{R}^{n\times d}$ is a deterministic matrix. Let $\mathbf{v}_i = \mathbf{V}^{\mathrm{T}}\mathbf{e}_i$, $i \in [n]$. Then there exist constants $C > 0$, such that for any $t \geq 1$,*

$$\begin{aligned}&\mathbb{P}\left\{\left\|\sum_{i=1}^n(y_i - \mathbb{E}y_i)\mathbf{v}_i\right\|_2 > 3t^2\|\mathbf{V}\|_{2\to\infty} + \sqrt{6}\sigma\rho^{1/2}t\|\mathbf{V}\|_{\mathrm{F}}\right\} \leq 28e^{-t^2},\\ &\mathbb{P}\left\{\left\|\sum_{i=1}^n(|y_i - \mathbb{E}y_i| - \mathbb{E}|y_i - \mathbb{E}y_i|)\mathbf{v}_i\right\|_2 > 3t^2\|\mathbf{V}\|_{2\to\infty} + \sqrt{6}\sigma\rho^{1/2}t\|\mathbf{V}\|_{\mathrm{F}}\right\} \leq 28e^{-t^2}.\end{aligned}$$

Lemma S2.2. *Let $(y_i)_{i=1}^n$ be independent random variables such that $\max_{i\in[n]}\|y_i - \mathbb{E}y_i\|_{\psi_2} \leq \sigma\rho^{1/2}$ for some constant $\sigma > 0$. Suppose $\mathbf{V} \in \mathbb{R}^{n\times d}$ is a deterministic matrix. Let $\mathbf{v}_i = \mathbf{V}^{\mathrm{T}}\mathbf{e}_i$, $i \in [n]$. Then there exist a constant $C_0 > 0$, such that for any $t \geq 1$,*

$$\begin{aligned}&\mathbb{P}\left\{\left\|\sum_{i=1}^n(y_i - p_i)\mathbf{v}_i\right\|_2 > C_0\sigma\rho^{1/2}t\|\mathbf{V}\|_{\mathrm{F}}\right\} \leq 2(d+1)e^{-t^2},\\ &\mathbb{P}\left\{\left\|\sum_{i=1}^n(|y_i - p_i| - \mathbb{E}|y_i - p_i|)\mathbf{v}_i\right\|_2 > 2C_0\sigma\rho^{1/2}t\|\mathbf{V}\|_{\mathrm{F}}\right\} \leq 2(d+1)e^{-t^2}.\end{aligned}$$

One of the difficulties in generalizing the perturbation bounds for a single eigenvector to an eigenvector matrix lies in the control of $\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+} - \mathbf{S}_{\mathbf{P}_+}\mathbf{W}_+^*$ and $\mathbf{W}_-^*\mathbf{S}_{\mathbf{A}_-} - \mathbf{S}_{\mathbf{P}_-}\mathbf{W}_-^*$ because the matrix multiplication is not commutative. The following Lemma S2.3 allows us to tackle this type of technical barrier.

Lemma S2.3. *Suppose Assumptions 1-5 hold. Then there exists a absolute constant $c_0 > 0$, such that for sufficiently large n, the following events hold with probability at least $1 - c_0n^{-\zeta} - c_0e^{-t}$ for all $t > 0$:*

$$\begin{aligned}\|\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+} - \mathbf{S}_{\mathbf{P}_+}\mathbf{W}_+^*\|_2 &\lesssim_\sigma \max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, d^{1/2}, t^{1/2}\right\},\\ \|\mathbf{W}_+^*|\mathbf{S}_{\mathbf{A}_+}|^{1/2} - |\mathbf{S}_{\mathbf{P}_+}|^{1/2}\mathbf{W}_+^*\|_{\mathrm{F}} &\lesssim_\sigma \frac{d^{1/2}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^{1/2}}\max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, d^{1/2}, t^{1/2}\right\},\end{aligned}$$

$$\leq \frac{\|\mathbf{X}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, d^{1/2}, t^{1/2}\right\},$$

$$\|\mathbf{W}_+^*|\mathbf{S}_{\mathbf{A}_+}|^{-1/2} - |\mathbf{S}_{\mathbf{P}_+}|^{-1/2}\mathbf{W}_+^*\|_{\mathrm{F}} \lesssim_\sigma \frac{\|\mathbf{X}\|_{2\to\infty}}{(n\rho_n)^{3/2}\lambda_d(\mathbf{\Delta}_n)^2}\max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, d^{1/2}, t^{1/2}\right\}.$$

The above concentration bounds also hold with $(\mathbf{W}_+^*, \mathbf{S}_{\mathbf{A}_+}, \mathbf{S}_{\mathbf{P}_+})$ *replaced by* $(\mathbf{W}_-^*, \mathbf{S}_{\mathbf{A}_-}, \mathbf{S}_{\mathbf{P}_-})$.

With the help of Lemma S2.4 below, we are able to provide a sharp control of several remainder terms. The analyses of these remainders are necessary, as will be seen in Section 3.2 (see lines (15), (16), and (17)).

Lemma S2.4. *Suppose Assumptions 1-5 hold. Then there exists an absolute constant* $c_0 > 0$*, such that for all* $t \geq 1$*,* $t \lesssim n\rho_n$*, the following events hold with probability at least* $1 - c_0 n^{-\zeta} - c_0 d e^{-t}$ *for sufficiently large* n*:*

$$\|\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+})\|_{2\to\infty} \lesssim_\sigma \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\max\left\{t^{1/2}, d^{1/2}, \frac{1}{\lambda_d(\mathbf{\Delta}_n)}\right\},$$

$$\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}(\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*)\|_2 \lesssim_\sigma \frac{t^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{3/2}\lambda_d(\mathbf{\Delta}_n)^2}\max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, d^{1/2}, t^{1/2}\right\},$$

$$\|(\mathbf{P} - \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}})\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1}\|_{2\to\infty} \lesssim_\sigma \frac{d^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{1}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}, d^{1/2}\right\}.$$

Also, for sufficiently large n*, with probability at least* $1 - c_0 n^{-\zeta}$*,*

$$\|\mathbf{U}_{\mathbf{P}_+}(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} - \mathbf{W}_+^*)\|_{2\to\infty} \lesssim \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^2}.$$

These concentration bounds hold with $(\mathbf{U}_{\mathbf{A}_+}, \mathbf{S}_{\mathbf{A}_+}, \mathbf{U}_{\mathbf{P}_+}, \mathbf{S}_{\mathbf{P}_+}, \mathbf{W}_+^*)$ *replaced by* $(\mathbf{U}_{\mathbf{A}_-}, \mathbf{S}_{\mathbf{A}_-}, \mathbf{U}_{\mathbf{P}_-}, \mathbf{S}_{\mathbf{P}_-}, \mathbf{W}_-^*)$.

We conclude this section with the following lemma, which asserts that the two-to-infinity norm of $\mathbf{U}_{\mathbf{A}}$ can be upper bounded by the two-to-infinity norm of $\mathbf{U}_{\mathbf{P}}$ with large probability.

Lemma S2.5. *Suppose Assumptions 1-5 hold. Then there exists an absolute constant* $c_0 > 0$*, such that for sufficiently large* n*, with probability at least* $1 - c_0 n^{-\zeta\wedge\xi}$*,*

$$\|\mathbf{U}_{\mathbf{A}}\|_{2\to\infty} \lesssim \{\kappa(\mathbf{\Delta}_n) + \varphi(1)\}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}.$$

S3. Proofs of the Lemmas in Section S2

In this section, we provide the proofs of Lemmas S2.1, S2.2, S2.3, and S2.4 in Section S2 above.

S3.1. Proof of Lemma S2.1

The proof is a straightforward application of the vector version of the Bernstein's inequality (Lemma S1.2). We set $\mathbf{y}_i = (y_i - \mathbb{E}y_i)\mathbf{v}_i$ and $\mathbf{y}_i = (|y_i - \mathbb{E}y_i| - \mathbb{E}|y_i - \mathbb{E}y_i|)\mathbf{v}_i$, respectively. Clearly,

$$\tau_1^2 = \sum_{i=1}^n \|\mathbf{v}_i\|_2^2 \mathrm{var}(y_i) \leq \sigma^2\rho\|\mathbf{V}\|_{\mathrm{F}}^2,$$

$$\tau_2^2 = \sum_{i=1}^n \|\mathbf{v}_i\|_2^2 \mathrm{var}(|y_i - \mathbb{E}y_i|) \leq \sum_{i=1}^n \|\mathbf{v}_i\|_2^2 \mathbb{E}\{(y_i - \mathbb{E}y_i)^2\} \leq \sigma^2\rho\|\mathbf{V}\|_{\mathrm{F}}^2$$

and we take $U = \|\mathbf{V}\|_{2\to\infty}$. Note that

$$\frac{1}{6}(U + \sqrt{U^2 + 36\tau_k^2}) \leq \frac{1}{3}U + \tau_k \leq \frac{1}{3}\|\mathbf{V}\|_{2\to\infty} + \sigma\rho^{1/2}\|\mathbf{V}\|_{\mathrm{F}} \leq 3t^2\|\mathbf{V}\|_{2\to\infty} + \sqrt{6}\sigma\rho^{1/2}t\|\mathbf{V}\|_{\mathrm{F}}$$

for any $t \geq 1$ and $k = 1, 2$. Then Lemma S1.2 implies

$$\begin{aligned}
&\mathbb{P}\left\{\left\|\sum_{i=1}^n (y_i - p_i)\mathbf{v}_i\right\|_2 > 3t^2\|\mathbf{V}\|_{2\to\infty} + \sqrt{6}\sigma\rho^{1/2}t\|\mathbf{V}\|_{\mathrm{F}}\right\}\\
&\leq 28\exp\left(-3\frac{9t^4\|\mathbf{V}\|_{2\to\infty}^2 + 6\sigma^2\rho t^2\|\mathbf{V}\|_{\mathrm{F}}^2 + 6\sqrt{6}\sigma\rho^{1/2}t^3\|\mathbf{V}\|_{\mathrm{F}}\|\mathbf{V}\|_{2\to\infty}}{6\sigma^2\rho\|\mathbf{V}\|_{\mathrm{F}}^2 + 6t^2\|\mathbf{V}\|_{2\to\infty}^2 + 2\sqrt{6}\sigma\rho^{1/2}t\|\mathbf{V}\|_{\mathrm{F}}\|\mathbf{V}\|_{2\to\infty}}\right)\\
&\leq 28\exp\left(-3\frac{9t^4\|\mathbf{V}\|_{2\to\infty}^2 + 6\sigma^2\rho t^2\|\mathbf{V}\|_{\mathrm{F}}^2 + 6\sqrt{6}\sigma\rho^{1/2}t^3\|\mathbf{V}\|_{\mathrm{F}}\|\mathbf{V}\|_{2\to\infty}}{9t^2\|\mathbf{V}\|_{2\to\infty}^2 + 6\sigma^2\rho\|\mathbf{V}\|_{\mathrm{F}}^2 + 6\sqrt{6}\sigma\rho^{1/2}t\|\mathbf{V}\|_{\mathrm{F}}\|\mathbf{V}\|_{2\to\infty}}\right) = 28e^{-3t^2},
\end{aligned}$$

and similarly,

$$\mathbb{P}\left\{\left\|\sum_{i=1}^n (|y_i - p_i| - \mathbb{E}|y_i - p_i|)\mathbf{v}_i\right\|_2 > 3t^2\|\mathbf{V}\|_{2\to\infty} + \sqrt{6}\sigma\rho^{1/2}t\|\mathbf{V}\|_{\mathrm{F}}\right\} \leq 28e^{-3t^2}.$$

S3.2. Proof of Lemma S2.2

We apply a "symmetric dilation" trick [1,12] and the matrix Chernoff bound (Lemma S1.3). Define

$$\mathbf{T}(\mathbf{v}_i) = \begin{bmatrix}\mathbf{0}_{d\times d} & \mathbf{v}_i\\ \mathbf{v}_i^{\mathrm{T}} & 0\end{bmatrix},\quad \mathbf{Z}_i = (y_i - p_i)\mathbf{T}(\mathbf{v}_i),\quad i = 1, 2, \ldots, n,$$

and let $\mathbf{S}_n = \sum_{i=1}^n \mathbf{Z}_i$. Clearly, $\|\mathbf{S}_n\|_2 = \max\{\lambda_{\max}(\mathbf{S}_n), \lambda_{\max}(-\mathbf{S}_n)\}$ and $-\mathbf{S}_n = \sum_{i=1}^n(-\mathbf{Z}_i)$. Observe that the spectral decomposition of $\mathbf{T}(\mathbf{v}_i)$ is given by

$$\mathbf{T}(\mathbf{v}_i) = \mathbf{Q}_i\begin{bmatrix}\|\mathbf{v}_i\|_2 & \\ & -\|\mathbf{v}_i\|_2\end{bmatrix}\mathbf{Q}_i^{\mathrm{T}} + 0\times\mathbf{Q}_{i\perp}\mathbf{Q}_{i\perp}^{\mathrm{T}},$$

where

$$\mathbf{Q}_i = \frac{1}{\sqrt{2}}\begin{bmatrix}\frac{\mathbf{v}_i}{\|\mathbf{v}_i\|_2} & \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|_2}\\ 1 & -1\end{bmatrix}$$

and $\mathbf{Q}_{i\perp} \in \mathbb{O}(d+1, d-1)$ is the orthogonal complement matrix of $\mathbf{Q}_i$. Then for any $\theta > 0$, we use the above spectral decomposition and Lemma 5.5 in [18] to compute the matrix moment generating function of $\mathbf{Z}_i$:

$$\begin{aligned}
\mathbb{E}e^{\theta\mathbf{Z}_i} &= \mathbb{E}\exp\{\theta(y_i - \mathbb{E}y_i)\mathbf{T}(\mathbf{v}_i)\}\\
&= \mathbb{E}\left\{\mathbf{Q}_i\begin{bmatrix}\exp\{\theta(y_i - \mathbb{E}y_i)\|\mathbf{v}_i\|_2\} & \\ & \exp\{-\theta(y_i - \mathbb{E}y_i)\|\mathbf{v}_i\|_2\}\end{bmatrix}\mathbf{Q}_i^{\mathrm{T}} + \mathbf{Q}_{i\perp}\mathbf{Q}_{i\perp}^{\mathrm{T}}\right\}
\end{aligned}$$

$$
\begin{aligned}
&\le \exp\{C_0\sigma^2\rho\theta^2\|\mathbf{v}_i\|_2^2\}\mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}} + \mathbf{Q}_{i\perp}\mathbf{Q}_{i\perp}^{\mathrm{T}}\\
&= \exp(C_0\sigma^2\rho\theta^2\|\mathbf{v}_i\|_2^2\mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}}) = \exp\{g(\theta)\mathbf{M}_i\},
\end{aligned}
$$

where $C_0 > 0$ is an absolute constant,

$$
g(\theta) = C_0\sigma^2\rho\theta^2, \quad \text{and} \quad \mathbf{M}_i = \|\mathbf{v}_i\|_2^2\mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}}.
$$

Similarly, $\mathbb{E}e^{\theta(-\mathbf{Z}_i)} \le \exp\{g(\theta)\mathbf{M}_i\}$. The corresponding scale parameter can be bounded by

$$
0 < \lambda_{\max}\left(\sum_{i=1}^{m}\mathbf{M}_i\right) \le \sum_{i=1}^{n}\|\mathbf{M}_i\|_2 \le \sum_{i=1}^{n}\|\mathbf{v}_i\|_2 = \|\mathbf{V}\|_{\mathrm{F}}^2.
$$

Therefore, by Lemma S1.3, for any $t > 0$,

$$
\begin{aligned}
\mathbb{P}(\|\mathbf{S}_n\|_2 > t) &\le \mathbb{P}\{\lambda_{\max}(\mathbf{S}_n) > t\} + \mathbb{P}\{\lambda_{\max}(-\mathbf{S}_n) > t\}\\
&\le 2(d+1)\exp\left\{\inf_{\theta>0}\left(-\theta t + C_0\theta^2\sigma^2\rho\|\mathbf{V}\|_{\mathrm{F}}^2\right)\right\}\\
&= 2(d+1)\exp\left(-\frac{t^2}{4C_0^2\sigma^2\rho\|\mathbf{V}\|_{\mathrm{F}}^2}\right).
\end{aligned}
$$

Now replacing t by $2C_0\sigma\rho^{1/2}\|\mathbf{V}\|_{\mathrm{F}}$ and adjust C_0 properly leads to the first assertion. The second assertion follows from the first assertion and the fact that

$$
\||y_i - \mathbb{E}y_i| - \mathbb{E}|y_i - \mathbb{E}y_i|\|_{\psi_2} \le 2\|y_i - \mathbb{E}y_i\|_{\psi_2} \le 2\sigma\rho^{1/2}.
$$

S3.3. Proof of Lemma S2.3

To prove Lemma S2.3, we first establish the following concentration bound for $\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}(\mathbf{A}-\mathbf{P})\mathbf{U}_{\mathbf{P}}\|_2$.

Lemma S3.1. *Suppose Assumptions 1-5 hold. Let $\mathbf{U}_1 \in \mathbb{O}(n,r)$, $\mathbf{U}_2 \in \mathbb{O}(n,s)$ be two matrices with $r,s \ge 1$, $r,s \le n$. Then there exists a constant $C > 0$ depending on σ, such that for all $t > 0$, with probability at least $1-(2+e)e^{-t}$,*

$$
\|\mathbf{U}_1^{\mathrm{T}}\mathbf{E}\mathbf{U}_2\|_2 \le C\max(r,s)^{1/2} + Ct^{1/2}.
$$

Proof of Lemma S3.1. By Assumption 3, we can write $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$, where $[\mathbf{E}_1]_{ij}$'s are independent bounded mean-zero random variables with $\mathrm{var}([\mathbf{E}_1]_{ij}) \le \sigma^2\rho_n$ for some constant $\sigma > 0$ and $[\mathbf{E}_2]_{ij}$'s are sub-Gaussian random variables whose sub-Gaussian norms are bounded by $\sigma\rho_n^{1/2}$. We apply a classical discretization trick to the spectral norm of $\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_k\mathbf{U}_2$, $k = 1,2$. By definition, $\|\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_k\mathbf{U}_2\|_2 \le \max_{\|\mathbf{v}_1\|_2,\|\mathbf{v}_2\|_2\le 1}|\mathbf{v}_1^{\mathrm{T}}\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_k\mathbf{U}_2\mathbf{v}_2|$. Now let $\mathcal{S}_\epsilon^{r-1}$ be an ϵ-net of the $(r-1)$-dimensional unit sphere $\mathcal{S}^{r-1} := \{\mathbf{v} : \|\mathbf{v}\|_2 = 1\}$, and similarly define $\mathcal{S}_\epsilon^{s-1}$. Clearly, for any $\mathbf{v}_1 \in \mathcal{S}^{r-1}$, $\mathbf{v}_2 \in \mathcal{S}^{s-1}$, there exists some $\mathbf{w}_1(\mathbf{v}_1) \in \mathcal{S}_\epsilon^{r-1}$ and $\mathbf{w}_2(\mathbf{v}_2) \in \mathcal{S}_\epsilon^{s-1}$, such that $\|\mathbf{v}_1 - \mathbf{w}_1(\mathbf{v}_1)\|_2 < \epsilon$, $\|\mathbf{v}_2 - \mathbf{w}_2(\mathbf{v}_2)\|_2 < \epsilon$, and

$$
\begin{aligned}
\|\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_k\mathbf{U}_2\|_2 &= \max_{\|\mathbf{v}_1\|_2,\|\mathbf{v}_2\|_2\le 1}|\mathbf{v}_1^{\mathrm{T}}\mathbf{U}_2^{\mathrm{T}}\mathbf{E}_k\mathbf{U}_2\mathbf{v}_2|\\
&= \max_{\|\mathbf{v}_1\|_2,\|\mathbf{v}_2\|_2\le 1}|\{\mathbf{v}_1 - \mathbf{w}_1(\mathbf{v}_1) + \mathbf{w}_1(\mathbf{v}_1)\}^{\mathrm{T}}\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_k\mathbf{U}_2\{\mathbf{v}_2 - \mathbf{w}_2(\mathbf{v}_2) + \mathbf{w}_2(\mathbf{v}_2)\}|_2
\end{aligned}
$$

$$\leq (\epsilon^2 + 2\epsilon)\|\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_k\mathbf{U}_2\|_2 + \max_{\mathbf{w}_1\in\mathcal{S}_\epsilon^{r-1},\mathbf{w}_2\in\mathcal{S}_\epsilon^{s-1}} |\mathbf{w}_1^{\mathrm{T}}\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_k\mathbf{U}_2\mathbf{w}_2|.$$

With $\epsilon = 1/3$, we have

$$\|\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_k\mathbf{U}_2\|_2 \leq \frac{9}{2}\max_{\mathbf{w}_1\in\mathcal{S}_{1/3}^{r-1},\mathbf{w}_2\in\mathcal{S}_{1/3}^{s-1}} |\mathbf{w}_1^{\mathrm{T}}\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_k\mathbf{U}_2\mathbf{w}_2|.$$

Furthermore, we know that $\mathcal{S}_{1/3}^{r-1}$ and $\mathcal{S}_{1/3}^{s-1}$ can be selected such that their cardinalities can be upper bounded by $|\mathcal{S}_{1/3}^{r-1}| \leq 18^r$ and $|\mathcal{S}_{1/3}^{s-1}| \leq 18^s$, respectively (see, for example, [13]). Now for fixed $\mathbf{w}_1 \in \mathcal{S}_{1/3}^{r-1}$ and $\mathbf{w}_2 \in \mathcal{S}_{1/3}^{s-1}$, let $\mathbf{z}_1 = \mathbf{U}_1\mathbf{w}_1 = [z_{11},\dots,z_{1n}]^{\mathrm{T}}$ and $\mathbf{z}_2 = \mathbf{U}_2\mathbf{w}_2 = [z_{21},\dots,z_{2n}]^{\mathrm{T}}$. Clearly, $\|\mathbf{z}_1\|_2, \|\mathbf{z}_2\|_2 \leq 1$, and

$$|\mathbf{w}_1^{\mathrm{T}}\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_k\mathbf{U}_2\mathbf{w}_2| = \left|\sum_{i=1}^n\sum_{j=1}^n [\mathbf{E}_k]_{ij}z_{1i}z_{2j}\right| \leq \left|\sum_{i<j}[\mathbf{E}_k]_{ij}(z_{1i}z_{2j} + z_{2i}z_{1j}) + \sum_{i=1}^n [\mathbf{E}_k]_{ii}z_{1i}z_{2i}\right|.$$

Denote $c_{ij} = z_{1i}z_{2j} + z_{1j}z_{2i}$ if $i \neq j$ and $c_{ii} = z_{1i}z_{2i}$. Note that

$$\sum_{i\leq j} c_{ij}^2 \leq \sum_{i<j}(2z_{1i}^2z_{2j}^2 + 2z_{2i}^2z_{1j}^2) + \sum_{i=1}^n z_{1i}^2z_{2i}^2 \leq 4\left(\sum_{i=1}^n z_{1i}^2\right)\left(\sum_{i=1}^n z_{2i}^2\right) + \sum_{i=1}^n z_{1i}^2 \leq 5.$$

For $\mathbf{E}_1$, by Hoeffding's inequality and a union bound over $\mathbf{w} \in \mathcal{S}_{1/3}^{d-1}$, we can pick an absolute constant $C > 0$, such that

$$\mathbb{P}\left\{\|\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_1\mathbf{U}_2\|_2 > C\max(r,s)^{1/2} + Ct^{1/2}\right\} \leq 2e^{-t}.$$

Appying Proposition 5.10 in [18] to $\|\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_2\mathbf{U}_2\|_2$ leads to a similar concentration inequality

$$\mathbb{P}\left\{\|\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_2\mathbf{U}_2\|_2 > C\max(r,s)^{1/2} + Ct^{1/2}\right\} \leq ee^{-t}.$$

with a potentially different multiplicative constant $C > 0$ depending on σ. The proof is completed by the inequality $\|\mathbf{U}_1^{\mathrm{T}}\mathbf{E}\mathbf{U}_2\|_2 \leq \|\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_1\mathbf{U}_2\|_2 + \|\mathbf{U}_1^{\mathrm{T}}\mathbf{E}_2\mathbf{U}_2\|_2$. □

Proof of Lemma S2.3. The proof is based on a modification of Lemma 49 in [2]. Following the decomposition there with the facts that $\mathbf{A}\mathbf{U}_{\mathbf{A}_+} = \mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}$, $\mathbf{A}\mathbf{U}_{\mathbf{A}_-} = \mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}$, $\mathbf{P}\mathbf{U}_{\mathbf{P}_+} = \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}$, and $\mathbf{P}\mathbf{U}_{\mathbf{P}_-} = \mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}$, we have

$$\begin{aligned}
\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+} - \mathbf{S}_{\mathbf{P}_+}\mathbf{W}_+^* &= (\mathbf{W}_+^* - \mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+})\mathbf{S}_{\mathbf{A}_+} + \mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}(\mathbf{A}-\mathbf{P})(\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+})\\
&\quad + \mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}(\mathbf{A}-\mathbf{P})\mathbf{U}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} + \mathbf{S}_{\mathbf{P}_+}(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} - \mathbf{W}_+^*),\\
\mathbf{W}_-^*\mathbf{S}_{\mathbf{A}_-} - \mathbf{S}_{\mathbf{P}_-}\mathbf{W}_-^* &= (\mathbf{W}_-^* - \mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-})\mathbf{S}_{\mathbf{A}_-} + \mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}(\mathbf{A}-\mathbf{P})(\mathbf{U}_{\mathbf{A}_-} - \mathbf{U}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-})\\
&\quad + \mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}(\mathbf{A}-\mathbf{P})\mathbf{U}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-} + \mathbf{S}_{\mathbf{P}_-}(\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-} - \mathbf{W}_-^*).
\end{aligned}$$

By Assumption 5, $\|\mathbf{A}-\mathbf{P}\|_2 \leq K(n\rho_n)^{1/2}$ with probability at least $1 - c_0n^{-\zeta}$ for all sufficiently large n for some absolute constants $K, c_0 > 0$, $\zeta \geq 1$. By Result S2.2, $\|\mathbf{S}_{\mathbf{A}}\|_2 \leq 2n\rho_n\lambda_1(\mathbf{\Delta}_n)$ with probability at least $1 - c_0 n^{-\zeta}$ for sufficiently large n, where $\zeta \geq 1$. By Lemma 6.7 in [4],

$$\|\mathbf{W}_+^* - \mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\|_2 \leq \|\sin\Theta(\mathbf{U}_{\mathbf{A}_+}, \mathbf{U}_{\mathbf{P}_+})\|_2^2, \quad \|\mathbf{W}_-^* - \mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}\|_2 \leq \|\sin\Theta(\mathbf{U}_{\mathbf{A}_-}, \mathbf{U}_{\mathbf{P}_-})\|_2^2.$$

Then by Davis-Kahan theorem in the form of [4], we have

$$\|\mathbf{W}_+^* - \mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\|_2 \leq \|\sin\Theta(\mathbf{U}_{\mathbf{A}_+}, \mathbf{U}_{\mathbf{P}_+})\|_2^2 \leq \frac{4\|\mathbf{A}-\mathbf{P}\|_2^2}{\{n\rho_n\lambda_d(\mathbf{\Delta}_n)\}^2} \leq \frac{4K^2}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^2},$$
$$\|\mathbf{W}_-^* - \mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}\|_2 \leq \|\sin\Theta(\mathbf{U}_{\mathbf{A}_-}, \mathbf{U}_{\mathbf{P}_-})\|_2^2 \leq \frac{4\|\mathbf{A}-\mathbf{P}\|_2^2}{\{n\rho_n\lambda_d(\mathbf{\Delta}_n)\}^2} \leq \frac{4K^2}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^2}$$

when $\|\mathbf{A}-\mathbf{P}\|_2 \leq K(n\rho_n)^{1/2}$, which occurs with probability at least $1 - c_0 n^{-\zeta}$, $\zeta \geq 1$. Also, observe that by Lemma 6.7 in [4] and Davis-Kahan theorem again, we have

$$\|\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\|_2 = \|\sin\Theta(\mathbf{U}_{\mathbf{A}_+}, \mathbf{U}_{\mathbf{P}_+})\|_2 \leq \frac{2\|\mathbf{A}-\mathbf{P}\|_2}{n\rho_n\lambda_d(\mathbf{\Delta}_n)} \leq \frac{2K}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)},$$
$$\|\mathbf{U}_{\mathbf{A}_-} - \mathbf{U}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}\|_2 = \|\sin\Theta(\mathbf{U}_{\mathbf{A}_-}, \mathbf{U}_{\mathbf{P}_-})\|_2 \leq \frac{2\|\mathbf{A}-\mathbf{P}\|_2}{n\rho_n\lambda_d(\mathbf{\Delta}_n)} \leq \frac{2K}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)},$$

provided that $\|\mathbf{A}-\mathbf{P}\|_2 \leq K(n\rho_n)^{1/2}$, which occurs with probability at least $1 - c_0 n^{-\zeta}$, $\zeta \geq 1$. Hence, for all sufficiently large n, we apply Lemma S3.1 to obtain

$$\begin{aligned}
\|\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+} - \mathbf{S}_{\mathbf{P}_+}\mathbf{W}_+^*\|_2 &\leq \|\mathbf{W}_+^* - \mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\|_2(\|\mathbf{S}_{\mathbf{A}_+}\|_2 + \|\mathbf{S}_{\mathbf{P}_+}\|_2) \\
&\quad + \|\mathbf{E}\|_2\|\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\|_2 + \|\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}\|_2 \\
&\lesssim_\sigma \frac{n\rho_n\lambda_1(\mathbf{\Delta}_n)}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^2} + \frac{(n\rho_n)^{1/2}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)} + \max(d^{1/2}, t^{1/2}) \\
&\lesssim \max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, d^{1/2}, t^{1/2}\right\}
\end{aligned}$$

with probability at least $1 - c_0 n^{-\zeta} - c_0 e^{-t}$, $\zeta \geq 1$. Following the same reasoning, we also see that the same concentration bound holds for $\|\mathbf{W}_-^*\mathbf{S}_{\mathbf{A}_-} - \mathbf{S}_{\mathbf{P}_-}\mathbf{W}_-^*\|_2$. This completes the proof of the first assertion.

We now turn to the second assertion. For any $k, l \in [p]$, write

$$\begin{aligned}
[\mathbf{W}_+^*|\mathbf{S}_{\mathbf{A}_+}|^{1/2} - |\mathbf{S}_{\mathbf{P}_+}|^{1/2}\mathbf{W}_+^*]_{kl} &= [\mathbf{W}_+^*]_{kl}\lambda_l(\mathbf{A})^{1/2} - \lambda_k^{1/2}(\mathbf{P})[\mathbf{W}_+^*]_{kl} \\
&= [\mathbf{W}_+^*]_{kl}\left\{\frac{\lambda_l(\mathbf{A}) - \lambda_k(\mathbf{P})}{\lambda_l(\mathbf{A})^{1/2} + \lambda_k(\mathbf{P})^{1/2}}\right\}.
\end{aligned}$$

Similarly, for any $k, l \in [q]$, we have

$$\begin{aligned}
[\mathbf{W}_-^*|\mathbf{S}_{\mathbf{A}_-}|^{1/2} - |\mathbf{S}_{\mathbf{P}_-}|^{1/2}\mathbf{W}_-^*]_{kl} &= [\mathbf{W}_-^*]_{kl}|\lambda_{n-q+l}(\mathbf{A})|^{1/2} - |\lambda_{n-q+k}(\mathbf{P})|^{1/2}[\mathbf{W}_-^*]_{kl} \\
&= [\mathbf{W}_-^*]_{kl}\left\{\frac{\lambda_{n-q+k}(\mathbf{P}) - \lambda_{n-q+l}(\mathbf{A})}{|\lambda_{n-q+l}(\mathbf{A})|^{1/2} + |\lambda_{n-q+k}(\mathbf{P})|^{1/2}}\right\}.
\end{aligned}$$

This immediately implies that

$$\|\mathbf{W}_+^*|\mathbf{S}_{\mathbf{A}_+}|^{1/2} - |\mathbf{S}_{\mathbf{P}_+}|^{1/2}\mathbf{W}_+^*\|_{\mathrm{F}} \le \frac{\sqrt{d}\|\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+} - \mathbf{S}_{\mathbf{P}_+}\mathbf{W}_+^*\|_2}{\{n\rho_n\lambda_d(\mathbf{\Delta}_n)\}^{1/2}} \le \frac{\|\mathbf{X}\|_{2\to\infty}\|\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+} - \mathbf{S}_{\mathbf{P}_+}\mathbf{W}_+^*\|_2}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)},$$

where we have used Result S2.3 that $d^{1/2} \le \|\mathbf{X}\|_{2\to\infty}/\lambda_d(\mathbf{\Delta}_n)^{1/2}$. Therefore, by the first assertion, for all sufficiently large n,

$$\|\mathbf{W}_+^*|\mathbf{S}_{\mathbf{A}_+}|^{1/2} - |\mathbf{S}_{\mathbf{P}_+}|^{1/2}\mathbf{W}_+^*\|_{\mathrm{F}} \lesssim \frac{\|\mathbf{X}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)} \max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, d^{1/2}, t^{1/2}\right\}$$

with probability at least $1 - c_0 n^{-\zeta} - c_0 e^{-t}$ for all $t > 0$, where $\zeta \ge 1$ is given by Assumption 5. We also see that the same concentration bound holds for $\|\mathbf{W}_-^*|\mathbf{S}_{\mathbf{A}_-}|^{1/2} - |\mathbf{S}_{\mathbf{P}_-}|^{1/2}\mathbf{W}_-^*\|_{\mathrm{F}}$ following the same reasoning.

The third assertion can be obtained in a similar fashion. By Result S2.2, for sufficiently large n, with probability at least $1 - c_0 n^{-\zeta}$, $\max(\|\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2, \|\mathbf{S}_{\mathbf{A}_-}^{-1}\|_2) \le \{2n\rho_n\lambda_d(\mathbf{\Delta}_n)\}^{-1}$. For any $k, l \in [p]$, we have

$$\begin{aligned}[\mathbf{W}_+^*|\mathbf{S}_{\mathbf{A}_+}|^{-1/2} - |\mathbf{S}_{\mathbf{P}_+}|^{-1/2}\mathbf{W}_+^*]_{kl} &= [\mathbf{W}_+^*]_{kl}\{\lambda_l(\mathbf{A})^{-1/2} - \lambda_k(\mathbf{P})^{-1/2}\} \\ &= \frac{[\mathbf{W}_+^*]_{kl}\{\lambda_k(\mathbf{P}) - \lambda_l(\mathbf{A})\}}{\lambda_l(\mathbf{A})^{1/2}\lambda_k(\mathbf{P})^{1/2}\{\lambda_l(\mathbf{A})^{1/2} + \lambda_k(\mathbf{P})^{1/2}\}}.\end{aligned}$$

For any $k, l \in [q]$, we have, similarly,

$$\begin{aligned}&[\mathbf{W}_-^*|\mathbf{S}_{\mathbf{A}_-}|^{-1/2} - |\mathbf{S}_{\mathbf{P}_-}|^{-1/2}\mathbf{W}_-^*]_{kl} \\ &\quad = [\mathbf{W}_-^*]_{kl}\{|\lambda_{n-q+l}(\mathbf{A})|^{-1/2} - |\lambda_{n-q+k}(\mathbf{P})|^{-1/2}\} \\ &\quad = \frac{[\mathbf{W}_-^*]_{kl}\{\lambda_{n-q+l}(\mathbf{A}) - \lambda_{n-q+k}(\mathbf{P})\}}{|\lambda_{n-q+l}(\mathbf{A})|^{1/2}|\lambda_{n-q+k}(\mathbf{P})|^{1/2}\{|\lambda_{n-q+l}(\mathbf{A})|^{1/2} + |\lambda_{n-q+k}(\mathbf{P})|^{1/2}\}}.\end{aligned}$$

Therefore, when $\|\mathbf{S}_{\mathbf{A}}^{-1}\|_2 \le 2\{(n\rho_n)\lambda_d(\mathbf{\Delta}_n)\}^{-1}$, by Result S2.3 that $d \le \|\mathbf{X}\|_{2\to\infty}^2/\lambda_d(\mathbf{\Delta}_n)$,

$$\begin{aligned}\|\mathbf{W}_+^*|\mathbf{S}_{\mathbf{A}_+}|^{-1/2} - |\mathbf{S}_{\mathbf{P}_+}|^{-1/2}\mathbf{W}_+^*\|_{\mathrm{F}}^2 &\le \|\mathbf{S}_{\mathbf{A}}^{-1}\|_2\|\mathbf{S}_{\mathbf{P}}^{-1}\|_2^2\|\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+} - \mathbf{S}_{\mathbf{P}_+}\mathbf{W}_+^*\|_{\mathrm{F}}^2 \\ &\le d\|\mathbf{S}_{\mathbf{A}}^{-1}\|_2\|\mathbf{S}_{\mathbf{P}}^{-1}\|_2^2\|\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+} - \mathbf{S}_{\mathbf{P}_+}\mathbf{W}_+^*\|_2^2 \\ &\le \frac{2\|\mathbf{X}\|_{2\to\infty}^2\|\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+} - \mathbf{S}_{\mathbf{P}_+}\mathbf{W}_+^*\|_2^2}{(n\rho_n)^3\lambda_d(\mathbf{\Delta}_n)^4}.\end{aligned}$$

The same concentration bound hold for $\|\mathbf{W}_-^*|\mathbf{S}_{\mathbf{A}_-}|^{-1/2} - |\mathbf{S}_{\mathbf{P}_-}|^{-1/2}\mathbf{W}_-^*\|_{\mathrm{F}}^2$ by the same reasoning. The proof of the third assertion is then completed by applying the first assertion. □

S3.4. Proof of Lemma S2.4

We first analyze the concentration bound for $\|\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}\|_2$, and the same reasoning leads to the same concentration bound for $\|\mathbf{S}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-} - \mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}\|_2$. Note that by definition of eigenvector matrices, $\mathbf{P}\mathbf{U}_{\mathbf{P}_+} = \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}$ and $\mathbf{A}\mathbf{U}_{\mathbf{A}_+} = \mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}$. Clearly, by Davis-Kahan theorem in the form of [4],

$$\begin{aligned}
\|\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}\|_2 &= \|\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{A}\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{P}\mathbf{U}_{\mathbf{A}_+}\|_2 \\
&= \|\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} + \mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{E}(\mathbf{I} - \mathbf{U}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}})\mathbf{U}_{\mathbf{A}_+}\|_2 \\
&\le \|\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}\|_2 + \|\mathbf{E}\|_2\|\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\|_2 \\
&\le \|\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}\|_2 + \|\mathbf{E}\|_2\|\sin\Theta(\mathbf{U}_{\mathbf{A}_+}, \mathbf{U}_{\mathbf{P}_+})\|_2 \\
&\le \|\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}\|_2 + \frac{2\|\mathbf{E}\|_2^2}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}.
\end{aligned}$$

For a realization of $\mathbf{A}$ with $\|\mathbf{E}\|_2 \lesssim (n\rho_n)^{1/2}$ and $\|\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}\|_2 \lesssim d^{1/2} + t^{1/2}$, which occurs with probability at least $1 - c_0 n^{-\zeta} - c_0 e^{-t}$ by Assumption 5 and Lemma S3.1, we have,

$$\|\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}\|_2 \lesssim \max\left\{t^{1/2}, d^{1/2}, \frac{1}{\lambda_d(\mathbf{\Delta}_n)}\right\}.$$

This event holds with probability at least $1 - c_0 n^{-\zeta} - c_0 e^{-t}$ for all $t > 0$. Then the first assertion is immediate by Result S2.2, Result S2.3, and the observation that

$$\begin{aligned}
&\|\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+})\|_{2\to\infty} \\
&\quad\le \|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2\|\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}\|_2.
\end{aligned}$$

For $\|\mathbf{U}_{\mathbf{P}_+}(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} - \mathbf{W}_+^*)\|_{2\to\infty}$, note that by Lemma 6.7 in [4] and Davis-Kahan theorem, we have

$$\begin{aligned}
\|\mathbf{U}_{\mathbf{P}_+}(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} - \mathbf{W}_+^*)\|_{2\to\infty} &\le \|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} - \mathbf{W}_+^*\|_2 \\
&\le \|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\sin\Theta(\mathbf{U}_{\mathbf{A}_+}, \mathbf{U}_{\mathbf{P}_+})\|_2^2 \\
&\le \frac{4\|\mathbf{E}\|_2^2}{(n\rho_n)^2\lambda_d(\mathbf{\Delta}_n)^2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}.
\end{aligned}$$

Then the fourth assertion follows from Assumption 5. The same concentration bound holds for $\|\mathbf{U}_{\mathbf{P}_-}(\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-} - \mathbf{W}_-^*)\|_{2\to\infty}$ by the same reasoning.

We now focus on $\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}(\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*)\|_2$. By Lemma S2.1, Lemma S2.2, for all $t \ge 1$ and $t \lesssim n\rho_n$, we have

$$\begin{aligned}
\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}\|_2 &= \left\|\sum_{j=1}^{n}(A_{mj} - \rho_n\mathbf{x}_m^{\mathrm{T}}\mathbf{x}_j)[\mathbf{U}_{\mathbf{P}_+}]_{j*}\right\|_2 \lesssim t\|\mathbf{U}_{\mathbf{P}_+}\|_{2\to\infty} + \sigma(\rho_n t)^{1/2}\|\mathbf{U}_{\mathbf{P}_+}\|_{\mathrm{F}} \\
&\lesssim_\sigma (n\rho_n t)^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}
\end{aligned}$$

with probability at least $1 - c_0 d e^{-t}$. Then by Lemma S2.3,

$$\begin{aligned}
\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}(\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*)\|_2 &\le \|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}\|_2\|\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*\|_2 \\
&\le \|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}\|_2\|\mathbf{S}_{\mathbf{P}_+}^{-1}\|_2\|\mathbf{S}_{\mathbf{P}_+}\mathbf{W}_+^* - \mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+}\|_2\|\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2
\end{aligned}$$

$$\lesssim_\sigma \frac{t^{1/2}\|\mathbf{U_P}\|_{2\to\infty}}{(n\rho_n)^{3/2}\lambda_d(\mathbf{\Delta}_n)^2}\max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, d^{1/2}, t^{1/2}\right\}$$

with probability at least $1-c_0n^{-\zeta}-c_0de^{-t}$ for sufficiently large n. The same concentration bound holds for $\|\mathbf{e}_m^{\mathrm{T}}\mathbf{EU}_{\mathbf{P}_-}(\mathbf{W}_-^*\mathbf{S}_{\mathbf{A}_-}^{-1}-\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{W}_-^*)\|_2$ by the same reasoning.
We finally deal with terms $\|(\mathbf{P}-\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}})\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1}\|_{2\to\infty}$ and $\|(\mathbf{P}-\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}})\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1}\|_{2\to\infty}$ by adopting the analysis in Appendix B.1 in [14]. By construction,

$$\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}=\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{A}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1}=\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{P}\mathbf{U}_{\mathbf{A}_+},$$
$$\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}=\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{A}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1}=\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{P}\mathbf{U}_{\mathbf{A}_-}$$

implying that

$$\begin{aligned}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}-\mathbf{S}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}&=\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\\&=\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{E}(\mathbf{I}_n-\mathbf{U}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}})\mathbf{U}_{\mathbf{A}_+}+\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+},\\\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}-\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}&=\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_-}\\&=\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{E}(\mathbf{I}_n-\mathbf{U}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}})\mathbf{U}_{\mathbf{A}_-}+\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}.\end{aligned}$$

Namely,

$$\mathbf{S}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}=\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}-\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_+},$$
$$\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}=\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}-\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_-}$$

and, by Lemma S3.1 and Davis-Kahan theorem, with probability at least $1-c_0n^{-\zeta}-c_0e^{-t}$,

$$\|\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\|_2\le\|\mathbf{E}\|_2\|\sin\Theta(\mathbf{U}_{\mathbf{A}_+},\mathbf{U}_{\mathbf{P}_+})\|_2+\|\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}\|_2\lesssim_\sigma\max\left\{\frac{1}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}, d^{1/2}\right\},$$
$$\|\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_-}\|_2\le\|\mathbf{E}\|_2\|\sin\Theta(\mathbf{U}_{\mathbf{A}_-},\mathbf{U}_{\mathbf{P}_-})\|_2+\|\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_-}\|_2\lesssim_\sigma\max\left\{\frac{1}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}, d^{1/2}\right\}$$

for sufficiently large n. For any $k\in[q]$ and $l\in[p]$, we have

$$[\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}-\mathbf{S}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}]_{kl}=[\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}]_{kl}\{\lambda_l(\mathbf{A})-\lambda_{n-q+k}(\mathbf{P})\},$$
$$[\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}-\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}]_{lk}=[\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}]_{lk}\{\lambda_{n-q+k}(\mathbf{A})-\lambda_l(\mathbf{P})\}$$

Note that by the concentration of eigenvalues Result S2.1

$$\min_{k\in[q],l\in[p]}\{\lambda_l(\mathbf{A})-\lambda_{n-q+k}(\mathbf{P})\}\ge n\rho_n\lambda_d(\mathbf{\Delta}_n),$$
$$\min_{k\in[q],l\in[p]}\{\lambda_l(\mathbf{P})-\lambda_{n-q+k}(\mathbf{A})\}\ge n\rho_n\lambda_d(\mathbf{\Delta}_n)$$

with probability at least $1-c_0n^{-\zeta}$ for large n. Therefore, by Lemma S3.1,

$$\|\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\|_{\mathrm{F}}\le\frac{1}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\|\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}-\mathbf{S}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\|_{\mathrm{F}}$$

$$
\begin{aligned}
&\leq \frac{\sqrt{d}\|\mathbf{E}\|_2\|\sin\Theta(\mathbf{U}_{\mathbf{A}_+},\mathbf{U}_{\mathbf{P}_+})\|_2 + \sqrt{d}\|\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}\|_2}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\\
&\lesssim_\sigma \frac{d^{1/2}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{1}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}, d^{1/2}\right\}
\end{aligned}
$$

with probability at least $1 - c_0 n^{-\zeta} - c_0 d e^{-t}$ for sufficiently large n. The same concentration bound also holds for $\|\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}\|_{\mathrm{F}}$ by the same reasoning. Therefore,

$$
\begin{aligned}
\|(\mathbf{P} - \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}})\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1}\|_{2\to\infty} &= \|\mathbf{U}_{\mathbf{P}_-}(\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1})\|_{2\to\infty}\\
&\leq \|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\|_2 + \|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\|_2\|\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2\\
&\lesssim_\sigma \frac{d^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{1}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}, d^{1/2}\right\}
\end{aligned}
$$

with probability at least $1 - c_0 n^{-\zeta} - c_0 d e^{-t}$ for sufficiently large n. The same concentration bound holds for $\|(\mathbf{P} - \mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}})\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1}\|_{2\to\infty}$ as well by the same reasoning. The proof is thus completed.

S4. Proofs for Section 3

S4.1. Proof of Lemma 3.3

The proof of Lemma 3.3 is slightly involved and is more difficult than Lemma 1 and Lemma 3 in [1]. The underlying reason is that we replace the mth row and mth column in $\mathbf{A}$ by their expected values in $\mathbf{A}^{(m)}$, but the construction of $\mathbf{A}^{(m)}$ in [1] is to zero out the mth row and column of $\mathbf{A}$. As $\mathbb{E}\mathbf{A}^{(m)}$ is the same as $\mathbb{E}\mathbf{A} = \mathbf{P}$, we can borrow the entrywise eigenvector analysis there to $\mathbf{U}_{\mathbf{A}}^{(m)}$. By the construction of $\mathbf{A}^{(m)}$,

$$
[\mathbf{A} - \mathbf{A}^{(m)}]_{ij} = \begin{cases} 0, & \text{if } i \neq m \text{ and } j \neq m,\\ A_{ij} - \mathbb{E}A_{ij}, & \text{if } i = m \text{ or } j = m. \end{cases}
$$

It follows that

$$
\|\mathbf{A} - \mathbf{A}^{(m)}\|_2 \leq \|\mathbf{A} - \mathbf{A}^{(m)}\|_{\mathrm{F}} \leq \left\{2\sum_{j=1}^n (A_{mj} - \mathbb{E}A_{mj})^2\right\}^{1/2} \leq \sqrt{2}\|\mathbf{E}\|_{2\to\infty} \lesssim (n\rho_n)^{1/2}
$$

with probability at least $1 - c_0 n^{-\zeta}$ by Assumption 5. Denote $\mathbf{I}_+ = \mathbf{I}_p$ and $\mathbf{I}_- = \mathbf{I}_q$ for convenience. Now viewing $\mathbf{A}^{(m)}$ as a perturbed version of $\mathbf{P}$, we apply Davis-Kahan theorem to obtain

$$
\begin{aligned}
\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)} - \mathbf{U}_{\mathbf{P}_+}\|_2 &= \|\{\mathbf{U}_{\mathbf{A}_+}^{(m)}(\mathbf{U}_{\mathbf{A}_+}^{(m)})^{\mathrm{T}} - \mathbf{I}_+\}\mathbf{U}_{\mathbf{P}_+}\|_2 = \|\{\mathbf{U}_{\mathbf{A}_+}^{(m)}(\mathbf{U}_{\mathbf{A}_+}^{(m)})^{\mathrm{T}} - \mathbf{I}_+\}\mathbf{U}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\|_2\\
&= \|\sin\Theta(\mathbf{U}_{\mathbf{A}_+}^{(m)}, \mathbf{U}_{\mathbf{P}_+})\|_2 \leq \frac{2\|\mathbf{A}^{(m)} - \mathbf{P}\|_2}{n\rho_n\lambda_d(\mathbf{\Delta}_n)} \leq \frac{2\|\mathbf{A} - \mathbf{P}\|_2 + 2\|\mathbf{A}^{(m)} - \mathbf{A}\|_2}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\\
&\lesssim \frac{1}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}
\end{aligned}
$$

with probability at least $1 - c_0 n^{-\zeta}$. The same concentration bound also holds for $\|\mathbf{U}_{\mathbf{A}_-}^{(m)}\mathbf{H}_-^{(m)} - \mathbf{U}_{\mathbf{P}_-}\|_2$ by the same reasoning. This completes the proof of the first assertion.

We now turn the focus to $\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty}$, $\|\mathbf{U}_{\mathbf{A}_-}^{(m)}\|_{2\to\infty}$ as well as $\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathrm{sgn}(\mathbf{H}_+^{(m)}) - \mathbf{U}_{\mathbf{P}_+}\|_{2\to\infty}$ and $\|\mathbf{U}_{\mathbf{A}_-}^{(m)}\mathrm{sgn}(\mathbf{H}_-^{(m)}) - \mathbf{U}_{\mathbf{P}_-}\|_{2\to\infty}$. This is the place where we apply Theorem S1.1 with $\mathbf{M} = \mathbf{A}^{(m)}$. We set

$$\bar{\gamma}_+ = \frac{\max\{3K, \|\mathbf{X}\|_{2\to\infty}^2\}}{(n\rho_n)^{1/2}\lambda_{\min}(\mathbf{\Delta}_{n+})}, \quad \bar{\gamma}_- = \frac{\max\{3K, \|\mathbf{X}\|_{2\to\infty}^2\}}{(n\rho_n)^{1/2}\lambda_{\min}(\mathbf{\Delta}_{n-})},$$

where $K > 0$ is the constant selected such that $\|\mathbf{E}\|_2 \le K(n\rho_n)^{1/2}$ with probability at least $1 - c_0 n^{-\zeta}$ according to Assumption 5 and $c_0 > 0, \zeta \ge 1$ are absolute constants. Denote $\Delta_+ = \lambda_p(\mathbf{P}) = n\rho_n\lambda_p(\mathbf{\Delta}_{n+})$, $\Delta_- = -\lambda_{n-q+1}(\mathbf{P}) = n\rho_n\lambda_q(\mathbf{\Delta}_{n-})$, $\kappa_+ = \lambda_1(\mathbf{P})/\Delta_+ = \kappa(\mathbf{\Delta}_{n+})$, and $\kappa_- = |\lambda_n(\mathbf{P})|/\Delta_- = \kappa(\mathbf{\Delta}_{n-})$. Note that $\kappa_\pm \le \kappa(\mathbf{\Delta}_n)$. We take the function $\omega(\cdot)$ in Theorem S1.1 to be the same the $\varphi(\cdot)$ given in Assumption 4. For condition A1, we see that

$$\begin{aligned}\|\mathbf{P}\|_{2\to\infty} &\le \rho_n\|\mathbf{X}\|_{2\to\infty}\|\mathbf{X}\|_2 \le \sqrt{n}\rho_n\|\mathbf{X}\|_{2\to\infty}^2\\ &\le \frac{\max\{3K, \|\mathbf{X}\|_{2\to\infty}^2\}n\rho_n}{(n\rho_n)^{1/2}} = \bar{\gamma}_+\Delta_+ = \bar{\gamma}_-\Delta_-.\end{aligned}$$

Condition A2 automatically holds because for each fixed i, the ith row and column of $\mathbf{A}^{(m)}$ are either the ith row and column of $\mathbf{A}$, or their expected values. By the construction of $\mathbf{A}$, we see that the ith row and column of $\mathbf{A}^{(m)}$ are independent of the rest of the random variables in $\mathbf{A}^{(m)}$. For condition A3, since $\max(\gamma_+, \gamma_-) \le \gamma$ and $\varphi(\cdot)$ is non-decreasing, we see that

$$\max[32\kappa_+\max\{\bar{\gamma}_+, \omega(\bar{\gamma}_+)\}, 32\kappa_-\max\{\bar{\gamma}_-, \omega(\bar{\gamma}_-)\}] \le 32\kappa(\mathbf{\Delta}_n)\max\{\gamma, \varphi(\gamma)\} \le 1$$

by Assumption 5. Again, by Assumption 5,

$$\|\mathbf{A}^{(m)} - \mathbf{P}\|_2 \le \|\mathbf{A} - \mathbf{A}^{(m)}\|_2 + \|\mathbf{E}\|_2 \le 3K(n\rho_n)^{1/2} \le \bar{\gamma}_+\Delta_+ = \bar{\gamma}_-\Delta_-$$

with probability at least $1 - \delta_0$, where $\delta_0 = c_0 n^{-\zeta}$ for constants $c_0 > 0$ and $\zeta \ge 1$. For condition A4, it automatically holds by Assumption 4 with probability $1 - \delta_1 n^{-1}$ because $\mathbf{A}^{(m)} - \mathbf{P} = \mathbf{E}^{(m)}$, where $\delta_1 = c_0 n^{-\xi}$ and $\xi \ge 1$. Note that $\varphi(\cdot)$ is non-decreasing, implying that

$$\max\{\bar{\gamma}_+ + \varphi(\bar{\gamma}_+), \bar{\gamma}_- + \varphi(\bar{\gamma}_-)\} \le \gamma\ + \varphi(\gamma).$$

Also, note that

$$\max\left\{\frac{\bar{\gamma}_+\|\mathbf{P}\|_{2\to\infty}}{\Delta_+}, \frac{\bar{\gamma}_-\|\mathbf{P}\|_{2\to\infty}}{\Delta_-}\right\} \lesssim \frac{\|\mathbf{X}\|_{2\to\infty}^2 n\rho_n\lambda_1(\mathbf{\Delta}_n)}{(n\rho_n)^{3/2}\lambda_d(\mathbf{\Delta}_n)^2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty} \le \kappa(\mathbf{\Delta}_n)\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}.$$

Hence, we obtain from Theorem S1.1 that, with probability at least $1 - c_0 n^{-\zeta\wedge\xi}$,

$$\begin{aligned}&\max(\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty}, \|\mathbf{U}_{\mathbf{A}_-}^{(m)}\|_{2\to\infty}) \lesssim \{\kappa(\mathbf{\Delta}_n) + \varphi(1)\}\,\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty},\\ &\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathrm{sgn}(\mathbf{H}_+^{(m)}) - \mathbf{U}_{\mathbf{P}_+}\|_{2\to\infty} \lesssim [\kappa(\mathbf{\Delta}_n)\{\kappa(\mathbf{\Delta}_n) + \varphi(1)\}\{\gamma + \varphi(\gamma)\} + \kappa(\mathbf{\Delta}_n) + \varphi(1)]\,\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty},\\ &\|\mathbf{U}_{\mathbf{A}_-}^{(m)}\mathrm{sgn}(\mathbf{H}_-^{(m)}) - \mathbf{U}_{\mathbf{P}_-}\|_{2\to\infty} \lesssim [\kappa(\mathbf{\Delta}_n)\{\kappa(\mathbf{\Delta}_n) + \varphi(1)\}\{\gamma + \varphi(\gamma)\} + \kappa(\mathbf{\Delta}_n) + \varphi(1)]\,\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty},\end{aligned}$$

which are the second and the third assertion.

We then focus on the last assertion regarding $\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)} - \mathbf{U}_{\mathbf{A}_+}\mathbf{H}_+\|_2$ and $\|\mathbf{U}_{\mathbf{A}_-}^{(m)}\mathbf{H}_-^{(m)} - \mathbf{U}_{\mathbf{A}_-}\mathbf{H}_-\|_2$. By the concentration of eigenvalues in Result S2.1, we know that $\lambda_p(\mathbf{A}) \geq (1/2)n\rho_n\lambda_d(\mathbf{\Delta}_n)$ and $\lambda_{n-q+1}(\mathbf{A}) \leq -(1/2)n\rho_n\lambda_d(\mathbf{\Delta}_n)$ with probability at least $1 - c_0 n^{-\zeta}$ for sufficiently large n. By Weyl's inequality, for sufficiently large n,

$$\lambda_p(\mathbf{A}^{(m)}) \geq \lambda_p(\mathbf{A}) - \|\mathbf{A} - \mathbf{A}^{(m)}\|_2 \geq \frac{1}{4}n\rho_n\lambda_d(\mathbf{\Delta}_n),$$
$$\lambda_{n-q+1}(\mathbf{A}^{(m)}) \leq \lambda_{n-q+1}(\mathbf{A}) + \|\mathbf{A} - \mathbf{A}^{(m)}\|_2 \leq -\frac{1}{4}n\rho_n\lambda_d(\mathbf{\Delta}_n)$$

with probability at least $1 - c_0 n^{-\zeta}$, where we have used the assumption that $n\rho_n\lambda_d(\mathbf{\Delta}_n)^2 \to \infty$ and the fact that $\|\mathbf{A} - \mathbf{A}^{(m)}\|_2 \leq \sqrt{2}K(n\rho_n)^{1/2} \leq (1/4)n\rho_n\lambda_d(\mathbf{\Delta}_n)$ with probability at least $1 - c_0 n^{-\zeta}$ for large n by Assumption 5. On the other hand, applying Weyl's inequality to $\lambda_{p+1}(\mathbf{A})$ and $\lambda_{n-q}(\mathbf{A})$ yields

$$\lambda_{p+1}(\mathbf{A}) \leq \lambda_{p+1}(\mathbf{P}) + \|\mathbf{E}\|_2 \leq K(n\rho_n)^{1/2} \leq \frac{1}{8}n\rho_n\lambda_d(\mathbf{\Delta}_n),$$
$$\lambda_{n-q}(\mathbf{A}) \geq \lambda_{n-q}(\mathbf{P}) - \|\mathbf{E}\|_2 \geq -K(n\rho_n)^{1/2} \geq -\frac{1}{8}n\rho_n\lambda_d(\mathbf{\Delta}_n)$$

with probability at least $1 - c_0 n^{-\zeta}$ for sufficiently large n. We thus obtain that

$$\lambda_p(\mathbf{A}^{(m)}) - \lambda_{p+1}(\mathbf{A}) \geq \frac{1}{8}(n\rho_n)\lambda_d(\mathbf{\Delta}_n) \quad \text{and} \quad \lambda_{n-q}(\mathbf{A}) - \lambda_{n-q+1}(\mathbf{A}^{(m)}) \geq \frac{1}{8}(n\rho_n)\lambda_d(\mathbf{\Delta}_n)$$

with probability at least $1 - c_0 n^{-\zeta}$ for large n. Hence, by a version of the Davis-Kahan theorem (See Theorem VII.3.4 in [3]),

$$\begin{aligned}
\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)} - \mathbf{U}_{\mathbf{A}_+}\mathbf{H}_+\|_2 = \|\sin\Theta(\mathbf{U}_{\mathbf{A}_+}^{(m)}, \mathbf{U}_{\mathbf{A}_+})\|_2 &\leq \frac{\|(\mathbf{A} - \mathbf{A}^{(m)})\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_2}{\lambda_p(\mathbf{A}^{(m)}) - \lambda_{p+1}(\mathbf{A})}\\
&\leq \frac{8\|(\mathbf{A} - \mathbf{A}^{(m)})\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{\mathrm{F}}}{(n\rho_n)\lambda_d(\mathbf{\Delta}_n)},\\
\|\mathbf{U}_{\mathbf{A}_-}^{(m)}\mathbf{H}_-^{(m)} - \mathbf{U}_{\mathbf{A}_-}\mathbf{H}_-\|_2 = \|\sin\Theta(\mathbf{U}_{\mathbf{A}_-}^{(m)}, \mathbf{U}_{\mathbf{A}_-})\|_2 &\leq \frac{\|(\mathbf{A} - \mathbf{A}^{(m)})\mathbf{U}_{\mathbf{A}_-}^{(m)}\|_2}{\lambda_{n-q}(\mathbf{A}) - \lambda_{n-q+1}(\mathbf{A}^{(m)})}\\
&\leq \frac{8\|(\mathbf{A} - \mathbf{A}^{(m)})\mathbf{U}_{\mathbf{A}_-}^{(m)}\|_{\mathrm{F}}}{(n\rho_n)\lambda_d(\mathbf{\Delta}_n)}
\end{aligned}$$

with probability at least $1 - c_0 n^{-\zeta}$ for large n. We now focus on $\|(\mathbf{A} - \mathbf{A}^{(m)})\mathbf{U}_{\mathbf{A}}^{(m)}\|_{\mathrm{F}}$. The key idea is that the non-zero entries of $\mathbf{A} - \mathbf{A}^{(m)}$ are the centered version of the mth row and mth column of $\mathbf{A}$, namely, $\{A_{ij} - \mathbb{E}A_{ij} : i = m \text{ or } j = m\}$. This is a collection of random variables that are independent of $\mathbf{A}^{(m)}$. Since $\mathbf{U}_{\mathbf{A}}^{(m)}$ is the eigenvector matrix of $\mathbf{A}^{(m)}$ corresponding to the eigenvalues in $\mathbf{S}_{\mathbf{A}}$, it follows that $(\mathbf{A} - \mathbf{A}^{(m)})$ and $(\mathbf{U}_{\mathbf{A}_+}^{(m)}, \mathbf{U}_{\mathbf{A}_-}^{(m)})$ are independent. Below, we will work with $\mathbf{U}_{\mathbf{A}_+}^{(m)}$, noticing that the same reasoning will lead to the same concentration bound for $\|\mathbf{U}_{\mathbf{A}_-}^{(m)}$ Write

$$\|(\mathbf{A}-\mathbf{A}^{(m)})\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{\mathrm{F}}^2 = \sum_{i\neq m}(A_{im}-\mathbb{E}A_{im})^2\|[\mathbf{U}_{\mathbf{A}_+}^{(m)}]_{m*}\|_2^2 + \left\|\sum_{j=1}^n (A_{mj}-\mathbb{E}A_{mj})[\mathbf{U}_{\mathbf{A}_+}^{(m)}]_{j*}\right\|_2^2$$
$$\leq \|\mathbf{E}\|_{2\to\infty}^2\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty}^2 + \left\|\sum_{j=1}^n (A_{mj}-\mathbb{E}A_{mj})[\mathbf{U}_{\mathbf{A}_+}^{(m)}]_{j*}\right\|_2^2.$$

By Assumption 5 and the second assertion, for large n, we know that with probability at least $1-c_0n^{-\zeta\wedge\xi}$,

$$\|\mathbf{E}\|_{2\to\infty}\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty} \leq \|\mathbf{E}\|_2\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty} \lesssim \{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}(n\rho_n)^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}.$$

For the second part, for any $t\geq 1$ and $t\lesssim n\rho_n$, we consider the following two events:

$$\mathcal{E}_1 = \left\{\mathbf{A}:\left\|\sum_{j=1}^n (A_{mj}-\mathbb{E}A_{mj})[\mathbf{U}_{\mathbf{A}_+}^{(m)}]_{j*}\right\|_2 \leq C_0t\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty} + C_0\sigma(\rho_nt)^{1/2}\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{\mathrm{F}}\right\},$$
$$\mathcal{E}_2 = \left\{\mathbf{A}:\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty} \leq C_0\{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\right\}.$$

Here, $C_0>0$ is a constant that will be determined later. By the independence between $\mathbf{A}-\mathbf{A}^{(m)}$ and $\mathbf{U}_{\mathbf{A}_+}^{(m)}$, Lemma S2.1, and Lemma S2.2, we can select C_0 depending on σ such that

$$\mathbb{P}(\mathcal{E}_1) = \sum_{\mathbf{A}^{(m)}}\mathbb{P}(\mathcal{E}_1\mid\mathbf{A}^{(m)})p(\mathbf{A}^{(m)}) \geq \sum_{\mathbf{A}^{(m)}}\{1-c_0de^{-t}\}p(\mathbf{A}^{(m)}) = 1-c_0de^{-t}.$$

Also, by the second assertion, $\mathbb{P}(\mathcal{E}_2)\geq 1-c_0n^{-\zeta\wedge\xi}$ for sufficiently large n. Now we consider a realization $\mathbf{A}\in\mathcal{E}_1\cap\mathcal{E}_2$. Then

$$\begin{aligned}\left\|\sum_{j=1}^n (A_{mj}-\mathbb{E}A_{mj})[\mathbf{U}_{\mathbf{A}_+}^{(m)}]_{j*}\right\|_2 &\leq C_0t\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty} + C_0\sigma(\rho_nt)^{1/2}\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{\mathrm{F}}\\
&\leq C_0t\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty} + C_0\sigma(n\rho_nt)^{1/2}\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty}\\
&\lesssim_\sigma (n\rho_nt)^{1/2}\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty}\\
&\lesssim_\sigma (n\rho_nt)^{1/2}\{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}.\end{aligned}$$

Such a realization occurs with probability at least $1-c_0n^{-\zeta\wedge\xi}-c_0de^{-t}$ for large n. Hence, we conclude that

$$\begin{aligned}\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)}-\mathbf{U}_{\mathbf{A}_+}\mathbf{H}_+\|_2 &= \|\sin\Theta(\mathbf{U}_{\mathbf{A}_+}^{(m)},\mathbf{U}_{\mathbf{A}_+})\|_2\\
&\leq \frac{8\|\mathbf{E}\|_{2\to\infty}\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)} + \frac{8\left\|\sum_{j\neq m}(A_{mj}-\mathbb{E}A_{mj})[\mathbf{U}_{\mathbf{A}_+}^{(m)}]_{j*}\right\|_2}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\end{aligned}$$

$$
\begin{aligned}
&\lesssim_\sigma \frac{\{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}\{(n\rho_n)^{1/2}+(n\rho_n t)^{1/2}\}}{(n\rho_n)\lambda_d(\mathbf{\Delta}_n)}\|\mathbf{U}_\mathbf{P}\|_{2\to\infty}\\
&\lesssim \frac{\{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}t^{1/2}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\|\mathbf{U}_\mathbf{P}\|_{2\to\infty}
\end{aligned}
$$

with probability at least $1-c_0n^{-\zeta\wedge\xi}-c_0de^{-t}$ for sufficiently large n. The proof is thus completed.

S4.2. Proof of Theorem 3.2

As discussed in Section 3.2, a crucial step in controlling the row-wise perturbation bound of the term $\mathbf{U}_\mathbf{A}-\mathbf{A}\mathbf{U}_\mathbf{P}\mathbf{S}_\mathbf{P}^{-1}\mathbf{W}^*$ lies in a sharp control of $\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_\mathbf{A}-\mathbf{U}_\mathbf{P}\mathbf{W}^*)$. This result is established in Lemma S4.1 below with the help of the decoupling technique in Section 3.2 and Lemma 3.3.

Lemma S4.1. *Suppose Assumptions 1-5 hold. Let $m\in[n]$ be any fixed row index. Then there exists an absolute constant $c_0>0$, such that for all $t\geq 1$, $t\lesssim n\rho_n$,*

$$
\begin{aligned}
\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_+}-\mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^*)\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2 &\lesssim_\sigma \frac{\|\mathbf{U}_\mathbf{P}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\\
&\quad\times\max\left\{\frac{\{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)},\frac{\kappa(\mathbf{\Delta}_n)+\varphi(1)}{\lambda_d(\mathbf{\Delta}_n)^2},\chi t\right\}
\end{aligned}
$$

holds with probability at least $1-c_0n^{-\zeta\wedge\xi}-c_0de^{-t}$ for sufficiently large n, where

$$
\chi:=\varphi(1)+\frac{\|\mathbf{X}\|_{2\to\infty}^2\vee 1}{\lambda_d(\mathbf{\Delta}_n)}.
$$

The above concentration bound also holds for $\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_-}-\mathbf{U}_{\mathbf{P}_-}\mathbf{W}_-^*)\mathbf{S}_{\mathbf{A}_-}^{-1}\|_2$

Proof of Lemma S4.1. Below, we will only work with $\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_+}-\mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^*)\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2$, noting that the same reasoning leads to the same concentration bound for $\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_-}-\mathbf{U}_{\mathbf{P}_-}\mathbf{W}_-^*)\mathbf{S}_{\mathbf{A}_-}^{-1}\|_2$. Let $\mathbf{H}_+$, $\mathbf{H}_-$, $\mathbf{A}^{(m)}$, $\mathbf{U}_{\mathbf{A}_+}^{(m)}$, $\mathbf{U}_{\mathbf{A}_-}^{(m)}$, $\mathbf{H}_+^{(m)}$, and $\mathbf{H}_-^{(m)}$ be defined as in Section 3.2. By inequality (20) in Section 3.2, we immediately obtain

$$
\begin{aligned}
\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_+}-\mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^*)\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2 &\leq \|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\{\mathrm{sgn}(\mathbf{H}_+)-\mathbf{H}_+\}\|_2\|\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2\\
&\quad+\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_+}\mathbf{H}_+-\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)})\|_2\|\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2\\
&\quad+\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)}-\mathbf{U}_{\mathbf{P}_+})\|_2\|\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2.
\end{aligned}
$$

We first focus on $\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\|_2$. By Assumption 5 and Lemma 2 in [1], we know that $\|\mathbf{H}_+^{-1}\|_2\leq 2$ with probability at least $1-c_0n^{-\zeta}$ for sufficiently large n. Then by Lemma 3.3, Lemma S2.1, Lemma S2.2, and the fact that $\mathbf{e}_m^{\mathrm{T}}\mathbf{E}$ and $\mathbf{U}_{\mathbf{A}_+}^{(m)}$ are independent, we have, for sufficiently large n,

$$
\begin{aligned}
\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\|_2 &\leq \|\mathbf{H}_+^{-1}\|_2\{\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_+}\mathbf{H}_+-\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)})\|_2+\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_2\}\\
&\lesssim_\sigma \|\mathbf{E}\|_2\|\mathbf{U}_{\mathbf{A}_+}\mathbf{H}_+-\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)}\|_2+(n\rho_n t)^{1/2}\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty}
\end{aligned}
$$

$$\lesssim_\sigma \frac{\{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)}\|\mathbf{U}_\mathbf{P}\|_{2\to\infty} + \{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}(n\rho_n t)^{1/2}\|\mathbf{U}_\mathbf{P}\|_{2\to\infty}$$

$$\lesssim_\sigma \{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}(n\rho_n t)^{1/2}\|\mathbf{U}_\mathbf{P}\|_{2\to\infty}$$

with probability at least $1 - c_0 n^{-\zeta\wedge\xi} - c_0 d e^{-t}$, where the last inequality is due to the fact that $\lambda_d(\mathbf{\Delta}_n)^{-1} \lesssim (n\rho_n)^{1/2}$. Letting $t = \{1 + (\xi\wedge\zeta)\}\log n$, we see that

$$\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\|_2 \lesssim_\sigma \{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}(n\rho_n\log n)^{1/2}\|\mathbf{U}_\mathbf{P}\|_{2\to\infty}$$

with probability at least $1 - c_0 n^{-\zeta\wedge\xi}$ for large n. We now work on the first term, By Assumption 5, Result S2.2, Lemma 6.7 in [4], and Davis-Kahan theorem, for large n,

$$\begin{aligned}
\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\{\mathrm{sgn}(\mathbf{H}_+)-\mathbf{H}_+\}\|_2\|\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2 &\le \|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\|_2\|\sin\Theta(\mathbf{U}_{\mathbf{A}_+},\mathbf{U}_{\mathbf{P}_+})\|_2^2\|\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2\\
&\le \|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\|_2\frac{4\|\mathbf{E}\|_2^2}{(n\rho_n)^2\lambda_d(\mathbf{\Delta}_n)^2}\|\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2\\
&\lesssim_\sigma (n\rho_n)\{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}\|\mathbf{U}_\mathbf{P}\|_{2\to\infty}\\
&\quad\times\frac{n\rho_n}{(n\rho_n)^2\lambda_d(\mathbf{\Delta}_n)^2}\times\frac{1}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\\
&= \frac{\{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^3}\|\mathbf{U}_\mathbf{P}\|_{2\to\infty}
\end{aligned}$$

with probability at least $1 - c_0 n^{-\zeta\wedge\xi}$. For the second term, for all $t \ge 1$ and $t \lesssim n\rho_n$, by Assumption 5, Result S2.2, and Lemma 3.3, for large n,

$$\begin{aligned}
\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_+}\mathbf{H}_+ - \mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)})\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2 &\le \|\mathbf{E}\|_{2\to\infty}\|\mathbf{U}_{\mathbf{A}_+}\mathbf{H}_+ - \mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)}\|_2\|\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2\\
&\le \|\mathbf{E}\|_2\|\mathbf{U}_{\mathbf{A}_+}\mathbf{H}_+ - \mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)}\|_2\|\mathbf{S}_{\mathbf{A}}^{-1}\|_2\\
&\lesssim (n\rho_n)^{1/2}\times\frac{\{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}t^{1/2}\|\mathbf{U}_\mathbf{P}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\times\frac{1}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\\
&= \frac{\{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}t^{1/2}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^2}\|\mathbf{U}_\mathbf{P}\|_{2\to\infty}
\end{aligned}$$

with probability at least $1 - c_0 n^{-\zeta\wedge\xi} - c_0 d e^{-t}$.

We now focus on the third term. Denote $\mathbf{V}_+^{(m)} = \mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)} - \mathbf{U}_{\mathbf{P}_+}$. Let $t \ge 1$ and $t \lesssim n\rho_n$. Consider the following events:

$$\mathcal{E}_1 = \left\{\mathbf{A} : \|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{V}_+^{(m)}\|_2 \le C_0 t\|\mathbf{V}_+^{(m)}\|_{2\to\infty} + C_0\sigma(\rho_n t)^{1/2}\|\mathbf{V}_+^{(m)}\|_2\right\},$$

$$\mathcal{E}_2 = \left\{\mathbf{A} : \|\mathbf{V}_+^{(m)}\|_2 \le \frac{C_0}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}, \|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty} \le C_0\{\kappa(\mathbf{\Delta}_n)+\varphi(1)\}\|\mathbf{U}_\mathbf{P}\|_{2\to\infty}\right\},$$

$$\mathcal{E}_3 = \left\{\mathbf{A} : \|\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathrm{sgn}(\mathbf{H}_+^{(m)}) - \mathbf{U}_{\mathbf{P}_+}\|_{2\to\infty} \le C_0\chi\|\mathbf{U}_\mathbf{P}\|_{2\to\infty}\right\},$$

where

$$\chi := \varphi(1) + \frac{\|\mathbf{X}\|_{2\to\infty}^2 \vee 1}{\lambda_d(\mathbf{\Delta}_n)}.$$

and $C_0 > 0$ is a constant that will be determined later. Note that

$$\begin{aligned}\|\mathbf{V}_+^{(m)}\|_{2\to\infty} &\leq \|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty}\|\mathrm{sgn}(\mathbf{H}_+^{(m)}) - \mathbf{H}_+^{(m)}\|_2 + \|\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathrm{sgn}(\mathbf{H}_+^{(m)}) - \mathbf{U}_{\mathbf{P}_+}\|_{2\to\infty}\\ &\leq 2\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty} + \|\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathrm{sgn}(\mathbf{H}_+^{(m)}) - \mathbf{U}_{\mathbf{P}_+}\|_{2\to\infty}.\end{aligned}$$

By Lemma 3.3, we can select the constant $C_0 > 0$, such that $\mathbb{P}(\mathcal{E}_2) \geq 1 - c_0 n^{-\zeta\wedge\xi}$, $\mathbb{P}(\mathcal{E}_3) \geq 1 - c_0 n^{-\zeta\wedge\xi}$ for sufficiently large n. For event $\mathcal{E}_1$, we use the conditional distribution and the fact that $\mathbf{e}_m^{\mathrm{T}}\mathbf{E}$ is independent of $\mathbf{V}_+^{(m)}$, together with Lemma S2.1 and Lemma S2.2, to obtain

$$\mathbb{P}(\mathcal{E}_1) = \sum_{\mathbf{A}^{(m)}} \mathbb{P}(\mathcal{E}_1 \mid \mathbf{A}^{(m)})p(\mathbf{A}^{(m)}) \geq \sum_{\mathbf{A}^{(m)}} (1 - c_0 de^{-t})p(\mathbf{A}^{(m)}) = 1 - c_0 de^{-t}.$$

Then over the event $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$, we apply the fact that $\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty} \geq \sqrt{d/n}$ to obtain

$$\begin{aligned}\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{V}_+^{(m)}\|_2 &\leq C_0 t\|\mathbf{V}_+^{(m)}\|_{2\to\infty} + C_0\sigma(\rho_n t)^{1/2}\|\mathbf{V}_+^{(m)}\|_{\mathrm{F}}\\ &\leq C_0 t\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\|_{2\to\infty} + C_0 t\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathrm{sgn}(\mathbf{H}_+^{(m)}) - \mathbf{U}_{\mathbf{P}_+}\|_{2\to\infty} + C_0\sigma(d\rho_n t)^{1/2}\|\mathbf{V}_+^{(m)}\|_2\\ &\lesssim_\sigma \chi t\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty} + \frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)}\sqrt{\frac{d}{n}}\\ &\lesssim \chi t\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}.\end{aligned}$$

It follows from Result S2.2 that

$$\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)} - \mathbf{U}_{\mathbf{P}_+})\|_2\|\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2 \lesssim_c \frac{\chi t\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}$$

with probability at least $1 - c_0 n^{-\zeta\wedge\xi} - c_0 de^{-t}$ for sufficiently large n. The proof is completed by combining the above concentration bounds. □

We are now in a position to prove Theorems 3.2.

Proof of Theorem 3.2. The proof follows from Lemmas S4.1 and S2.4. Below, we will only work with $\mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*$, noting that the same reasoning leads to the same concentration bound for $\mathbf{U}_{\mathbf{A}_-} - \mathbf{A}\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{W}_-^*$, and combining the two concentration bounds finishes the proof. Following the decomposition of $\mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*$ in Section 3.2, we have

$$\begin{aligned}\|\mathbf{e}_m^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*)\| &\leq \|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^*)\mathbf{S}_{\mathbf{A}_+}^{-1}\|_{2\to\infty}\\ &\quad + \|\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+})\|_{2\to\infty}\\ &\quad + \|\mathbf{U}_{\mathbf{P}_+}(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} - \mathbf{W}_+^*)\|_{2\to\infty}\\ &\quad + \|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}(\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*)\|_2\end{aligned}$$

$$+ \|(\mathbf{P} - \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}})\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1}\|_{2\to\infty}.$$

By Lemma S4.1, the first term on the right-hand side above satisfies

$$\begin{aligned}\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^*)\mathbf{S}_{\mathbf{A}_+}^{-1}\|_2 &\lesssim_\sigma \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\\ &\quad\times \max\left\{\frac{\{\kappa(\mathbf{\Delta}_n) + \varphi(1)\}t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)}, \frac{\{\kappa(\mathbf{\Delta}_n) + \varphi(1)\}}{\lambda_d(\mathbf{\Delta}_n)^2}, \chi t\right\}\end{aligned}$$

with probability at least $1 - c_0 n^{-\zeta\wedge\xi} - c_0 d e^{-t}$ for sufficiently large n. We also know from Lemma S2.4 that the following events hold with probability at least $1 - c_0 n^{-\zeta\wedge\xi} - c_0 d e^{-t}$ for $t \geq 1$, $t \lesssim n\rho_n$:

$$\begin{aligned}&\|\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+})\|_{2\to\infty} \lesssim \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\max\left\{t^{1/2}, d^{1/2}, \frac{1}{\lambda_d(\mathbf{\Delta}_n)}\right\},\\ &\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}_+}(\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*)\|_2 \lesssim_\sigma \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d^2(\mathbf{\Delta}_n)}\max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}, d^{1/2}\right\},\\ &\|\mathbf{U}_{\mathbf{P}_+}(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} - \mathbf{W}_+^*)\|_{2\to\infty} \lesssim \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^2},\\ &\|(\mathbf{P} - \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}})\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1}\|_{2\to\infty} \lesssim \frac{d^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{1}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}, d^{1/2}\right\}\end{aligned}$$

where we have used the fact that $t^{1/2}/(n\rho_n)^{1/2} \lesssim 1$ and $t/(n\rho_n)^{1/2} \lesssim t^{1/2}$. Then Lemmas S4.1 and S2.4 immediately imply that

$$\begin{aligned}&\|\mathbf{e}_m^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*)\|_2\\ &\quad\lesssim_\sigma \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{\{\kappa(\mathbf{\Delta}_n) + \varphi(1)\}t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)}, \frac{\{\kappa(\mathbf{\Delta}_n) + \varphi(1)\}}{\lambda_d(\mathbf{\Delta}_n)^2}, \chi t\right\}\\ &\qquad+ \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\max\left\{t^{1/2}, d^{1/2}, \frac{1}{\lambda_d(\mathbf{\Delta}_n)}\right\} + \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^2}\\ &\qquad+ \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^2}\max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}, d^{1/2}\right\}\\ &\qquad+ \frac{d^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{1}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}, d^{1/2}\right\}\\ &\quad\leq \frac{\chi\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)}, \frac{1}{\lambda_d(\mathbf{\Delta}_n)^2}, t\right\}\end{aligned}$$

with probability at least $1 - c_0 n^{-\zeta\wedge\xi} - c_0 d e^{-t}$ for sufficiently large n. Here, we have used the fact that $d^{1/2} \leq d \leq \|\mathbf{X}\|_{2\to\infty}^2/\lambda_d(\mathbf{\Delta}_n)$ from Result S2.3 and

$$\kappa(\mathbf{\Delta}_n) = \frac{\lambda_1(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)} \leq \frac{\|\mathbf{X}\|_{\mathrm{F}}^2}{n\lambda_d(\mathbf{\Delta}_n)} \leq \frac{\|\mathbf{X}\|_{2\to\infty}^2}{\lambda_d(\mathbf{\Delta}_n)}.$$

This completes the first assertion. For the entrywise perturbation bound for the scaled eigenvectors, we recall the decomposition (6)

$$\widetilde{\mathbf{X}}_+\mathbf{W}_+ - \frac{\mathbf{A}\mathbf{X}_+(\mathbf{X}_+^{\mathrm{T}}\mathbf{X}_+)^{-1}}{\rho_n^{1/2}} = \mathbf{U}_{\mathbf{A}_+}(\mathbf{W}_+^*|\mathbf{S}_{\mathbf{A}_+}|^{1/2} - |\mathbf{S}_{\mathbf{P}_+}|^{1/2}\mathbf{W}_+^*)^{\mathrm{T}}\mathbf{W}_{\mathbf{X}_+}$$
$$+ (\mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*)(\mathbf{W}_+^*)^{\mathrm{T}}|\mathbf{S}_{\mathbf{P}_+}|^{1/2}\mathbf{W}_{\mathbf{X}_+}.$$

Then for each fixed row index $m \in [n]$, for all $t \geq 1$ and $t \lesssim n\rho_n$, we apply the first assertion above, Lemma S2.5, and Lemma S2.3 to conclude that

$$\begin{aligned}
&\|\mathbf{e}_m^{\mathrm{T}}(\widetilde{\mathbf{X}}_+\mathbf{W}_+ - \rho_n^{-1/2}\mathbf{A}\mathbf{X}_+(\mathbf{X}_+^{\mathrm{T}}\mathbf{X}_+)^{-1})\|_2\\
&\quad\leq \|\mathbf{U}_{\mathbf{A}_+}\|_{2\to\infty}\|\mathbf{W}_+^*|\mathbf{S}_{\mathbf{A}_+}|^{1/2} - |\mathbf{S}_{\mathbf{P}_+}|^{1/2}\mathbf{W}_+^*\|_2 + \|\mathbf{e}_m^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*)\|_2\|\mathbf{S}_{\mathbf{P}_+}\|_2^{1/2}\\
&\quad\lesssim \frac{\{\kappa(\mathbf{\Delta}_n) + \varphi(1)\}\|\mathbf{X}\|_{2\to\infty}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}, d^{1/2}\right\}\\
&\quad\quad + \frac{\chi\|\mathbf{X}\|_{2\to\infty}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)}, \frac{1}{\lambda_d(\mathbf{\Delta}_n)^2}, t\right\}\\
&\quad\lesssim \frac{\chi\|\mathbf{X}\|_{2\to\infty}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{(\|\mathbf{X}\|_{2\to\infty}^2 \vee 1)t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, t\right\}
\end{aligned}$$

with probability at least $1 - c_0 n^{-\zeta\wedge\xi} - c_0 d e^{-t}$ for sufficiently large n. This completes the proof of the second assertion. The third and fourth assertions regarding the concentrations of

$$\|\mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*\|_{2\to\infty} \quad\text{and}\quad \|\widetilde{\mathbf{X}}_+\mathbf{W}_+ - \rho_n^{-1/2}\mathbf{A}\mathbf{X}_+(\mathbf{X}_+^{\mathrm{T}}\mathbf{X}_+)^{-1}\|_{2\to\infty}$$

are immediate from the first two assertions and a union bound over $m \in [n]$ because $\zeta \wedge \xi$ is strictly greater than 1. □

S4.3. Proof of Theorem 3.1

By Theorem 3.2 and decompositions (4), (5) in the manuscript, for each fixed $i \in [n]$, we have

$$\sqrt{n}(\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i)_+ - \rho_n^{1/2}(\mathbf{x}_i)_+) = \frac{1}{\sqrt{n\rho_n}}\sum_{j=1}^n (A_{ij} - \mathbb{E}A_{ij})\mathbf{\Delta}_{n+}^{-1}(\mathbf{x}_j)_+ + \sqrt{n}\mathbf{R}_{\mathbf{X}_+}^{\mathrm{T}}\mathbf{e}_i,$$

$$\sqrt{n}(\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i)_- - \rho_n^{1/2}(\mathbf{x}_i)_-) = \frac{-1}{\sqrt{n\rho_n}}\sum_{j=1}^n (A_{ij} - \mathbb{E}A_{ij})\mathbf{\Delta}_{n-}^{-1}(\mathbf{x}_j)_- + \sqrt{n}\mathbf{R}_{\mathbf{X}_-}^{\mathrm{T}}\mathbf{e}_i,$$

$$n\rho_n^{1/2}\mathbf{W}_{\mathbf{X}_+}^{\mathrm{T}}(\mathbf{W}_+^*[\mathbf{U}_{\mathbf{A}_+}]_{i*} - [\mathbf{U}_{\mathbf{P}_+}]_{i*}) = \frac{1}{\sqrt{n\rho_n}}\sum_{j=1}^n (A_{ij} - \mathbb{E}A_{ij})\mathbf{\Delta}_{n+}^{-3/2}(\mathbf{x}_j)_+ + n\rho_n^{1/2}\{\mathbf{R}_{\mathbf{U}_+}(\mathbf{W}_+^*)^{\mathrm{T}}\mathbf{W}_{\mathbf{X}_+}\}^{\mathrm{T}}\mathbf{e}_i,$$

$$n\rho_n^{1/2}\mathbf{W}_{\mathbf{X}_-}^{\mathrm{T}}(\mathbf{W}_-^*[\mathbf{U}_{\mathbf{A}_-}]_{i*} - [\mathbf{U}_{\mathbf{P}_-}]_{i*}) = \frac{-1}{\sqrt{n\rho_n}}\sum_{j=1}^n (A_{ij} - \mathbb{E}A_{ij})\mathbf{\Delta}_{n-}^{-3/2}(\mathbf{x}_j)_- + n\rho_n^{1/2}\{\mathbf{R}_{\mathbf{U}_-}(\mathbf{W}_-^*)^{\mathrm{T}}\mathbf{W}_{\mathbf{X}_+}\}^{\mathrm{T}}\mathbf{e}_i,$$

where $\mathbf{R}_{\mathbf{X}_+} = \widetilde{\mathbf{X}}_+\mathbf{W}_+ - \rho_n^{-1/2}\mathbf{A}\mathbf{X}_+(\mathbf{X}_+^{\mathrm{T}}\mathbf{X}_+)^{-1}$, $\mathbf{R}_{\mathbf{X}_-} = \widetilde{\mathbf{X}}_-\mathbf{W}_- + \rho_n^{-1/2}\mathbf{A}\mathbf{X}_-(\mathbf{X}_-^{\mathrm{T}}\mathbf{X}_-)^{-1}$, $\mathbf{R}_{\mathbf{U}_+} = \mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*$, and $\mathbf{R}_{\mathbf{U}_-} = \mathbf{U}_{\mathbf{A}_-} - \mathbf{A}\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{W}_-^*$. Equivalently, we have

$$\sqrt{n}\mathbf{\Sigma}_{ni}^{-1/2}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i) = \frac{1}{\sqrt{n\rho_n}}\sum_{j=1}^n (A_{ij} - \mathbb{E}A_{ij})\mathbf{\Sigma}_{ni}^{-1/2}\mathbf{I}_{p,q}\mathbf{\Delta}_n^{-1}\mathbf{x}_j + \sqrt{n}\mathbf{\Sigma}_{ni}^{-1/2}\mathbf{R}_{\mathbf{X}}^{\mathrm{T}}\mathbf{e}_i,$$

$$\begin{aligned} n\rho_n^{1/2}\mathbf{\Gamma}_{ni}^{-1/2}\mathbf{W}_{\mathbf{X}}^{\mathrm{T}}(\mathbf{W}^*[\mathbf{U}_{\mathbf{A}}]_{i*} - [\mathbf{U}_{\mathbf{P}}]_{i*}) &= \frac{1}{\sqrt{n\rho_n}}\sum_{j=1}^n (A_{ij} - \mathbb{E}A_{ij})\mathbf{\Gamma}_{ni}^{-1/2}\mathbf{I}_{p,q}\mathbf{\Delta}_n^{-3/2}\mathbf{x}_j \\ &\quad + n\rho_n^{1/2}\mathbf{\Gamma}_{ni}^{-1/2}\{\mathbf{R}_{\mathbf{U}}(\mathbf{W}^*)^{\mathrm{T}}\mathbf{W}_{\mathbf{X}}\}^{\mathrm{T}}\mathbf{e}_i, \end{aligned}$$

where $\mathbf{R}_{\mathbf{X}} = [\mathbf{R}_{\mathbf{X}_+}, \mathbf{R}_{\mathbf{X}_-}]$ and $\mathbf{R}_{\mathbf{U}} = [\mathbf{R}_{\mathbf{U}_+}, \mathbf{R}_{\mathbf{U}_-}]$. To apply Theorem S1.4, we take

$$\begin{aligned} \xi_j &= \frac{1}{\sqrt{n\rho_n}}(A_{ij} - \mathbb{E}A_{ij})\mathbf{\Sigma}_{ni}^{-1/2}\mathbf{I}_{p,q}\mathbf{\Delta}_n^{-1}\mathbf{x}_j, \quad \mathbf{D} = \sqrt{n}\mathbf{\Sigma}_{ni}^{-1/2}\mathbf{R}_{\mathbf{X}}^{\mathrm{T}}\mathbf{e}_i, \\ \Delta^{(j)} &= \Delta = C\frac{\chi\|\mathbf{\Sigma}_{ni}^{-1/2}\|_2\|\mathbf{X}\|_{2\to\infty}^2}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^{3/2}}\max\left\{\frac{(\|\mathbf{X}\|_{2\to\infty}^2 \vee 1)(\log n\rho_n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, \log n\rho_n\right\}, \\ O &= \{\mathbf{A} : \Delta > \|\mathbf{D}\|_2\}, \end{aligned}$$

and

$$\begin{aligned} \xi_j' &= \frac{1}{\sqrt{n\rho_n}}(A_{ij} - \mathbb{E}A_{ij})\mathbf{\Gamma}_{ni}^{-1/2}\mathbf{I}_{p,q}\mathbf{\Delta}_n^{-3/2}\mathbf{x}_j, \quad \mathbf{D}' = \sqrt{n}\mathbf{\Gamma}_{ni}^{-1/2}\{\mathbf{R}_{\mathbf{U}}(\mathbf{W}^*)^{\mathrm{T}}\mathbf{W}_{\mathbf{X}}\}^{\mathrm{T}}\mathbf{e}_i, \\ \Delta^{(j)\prime} &= \Delta' = C\frac{\chi\|\mathbf{\Gamma}_{ni}^{-1/2}\|_2\|\mathbf{X}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^{3/2}}\max\left\{\frac{(\log n\rho_n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)}, \frac{1}{\lambda_d(\mathbf{\Delta}_n)^2}, \log n\rho_n\right\}, \\ O' &= \{\mathbf{A} : \Delta' > \|\mathbf{D}'\|_2\}. \end{aligned}$$

Here $C > 0$ is an absolute constant. In particular, we can select $C > 0$, which may depend on σ, such that $\mathbb{P}(O^c) \lesssim d/(n\rho_n)$ and $\mathbb{P}\{(O')^c\} \lesssim d/(n\rho_n)$ for sufficiently large n according to Theorem 3.2. Note that Assumption 3 implies that $\mathbb{E}[\mathbf{E}_{ij}]^2 \lesssim \sigma^2\rho_n$ for all $i, j \in [n]$, so that

$$\mathbf{\Sigma}_{ni} = \mathbf{I}_{p,q}\mathbf{\Delta}_n^{-1}\left\{\frac{1}{n\rho_n}\sum_{j=1}^n \mathbb{E}[\mathbf{E}_{ij}]^2\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\right\}\mathbf{\Delta}_n^{-1}\mathbf{I}_{p,q} \preceq \sigma^2\mathbf{I}_{p,q}\mathbf{\Delta}_n^{-1}\mathbf{I}_{p,q} \Longrightarrow \|\mathbf{\Sigma}_{ni}^{-1/2}\|_2 \geq \frac{1}{\sigma}\lambda_1(\mathbf{\Delta}_n)^{1/2},$$

$$\mathbf{\Gamma}_{ni} = \mathbf{I}_{p,q}\mathbf{\Delta}_n^{-3/2}\left\{\frac{1}{n\rho_n}\sum_{j=1}^n \mathbb{E}[\mathbf{E}_{ij}]^2\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\right\}\mathbf{\Delta}_n^{-3/2}\mathbf{I}_{p,q} \preceq \sigma^2\mathbf{I}_{p,q}\mathbf{\Delta}_n^{-2}\mathbf{I}_{p,q} \Longrightarrow \|\mathbf{\Gamma}_{ni}^{-1/2}\|_2 \geq \frac{1}{\sigma}\lambda_1(\mathbf{\Delta}_n).$$

Therefore, by Result S2.3, we have

$$\mathbb{P}(O) \lesssim \frac{d}{n\rho_n} \leq \frac{\|\mathbf{X}\|_{2\to\infty}^2}{n\rho_n\lambda_d(\mathbf{\Delta}_n)} \leq \frac{\|\mathbf{X}\|_{2\to\infty}^2\lambda_1(\mathbf{\Delta}_n)^{1/2}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^{3/2}} \lesssim_\sigma \frac{\chi d^{1/2}\|\mathbf{\Sigma}_{ni}^{-1/2}\|_2\|\mathbf{X}\|_{2\to\infty}^2}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^{3/2}} \lesssim_\sigma d^{1/2}\Delta,$$

$$\mathbb{P}(O') \lesssim \frac{d}{n\rho_n} \leq \frac{d^{1/2}\|\mathbf{X}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^{1/2}} \leq \frac{d^{1/2}\|\mathbf{X}\|_{2\to\infty}\lambda_1(\mathbf{\Delta}_n)}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^{3/2}} \lesssim_\sigma \frac{\chi d^{1/2}\|\mathbf{\Gamma}_{ni}^{-1/2}\|_2\|\mathbf{X}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^{3/2}} \lesssim_\sigma d^{1/2}\Delta'.$$

Note that $|\Delta - \Delta^{(j)}| = 0$, $|\Delta' - \Delta^{(j)'}| = 0$, and $\Delta^{(j)}$'s, $\Delta^{(j)'}$'s are constant random variables so that $\Delta^{(j)}$ and $\boldsymbol{\xi}_j$ are independent, and $\Delta^{(j)'}$ and $\boldsymbol{\xi}'_j$ are independent as well. Furthermore, $\mathbb{E}(\boldsymbol{\xi}_j) = \mathbb{E}(\boldsymbol{\xi}'_j) = 0$ and by the definition of $\boldsymbol{\Sigma}_{ni}$, $\boldsymbol{\Gamma}_{ni}$,

$$\sum_{j=1}^n \mathbb{E}(\boldsymbol{\xi}_j\boldsymbol{\xi}_j^{\mathrm{T}}) = \boldsymbol{\Sigma}_{ni}^{-1/2}\mathbf{I}_{p,q}\boldsymbol{\Delta}_n^{-1}\left\{\frac{1}{n\rho_n}\sum_{j=1}^n \mathbb{E}[\mathbf{E}_{ij}]^2\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\right\}\boldsymbol{\Delta}_n^{-1}\mathbf{I}_{p,q}\boldsymbol{\Sigma}_{ni}^{-1/2} = \mathbf{I}_d,$$

$$\sum_{j=1}^n \mathbb{E}\{(\boldsymbol{\xi}'_j)(\boldsymbol{\xi}'_j)^{\mathrm{T}}\} = \boldsymbol{\Gamma}_{ni}^{-1/2}\mathbf{I}_{p,q}\boldsymbol{\Delta}_n^{-3/2}\left\{\frac{1}{n\rho_n}\sum_{j=1}^n \mathbb{E}[\mathbf{E}]_{ij}^2\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\right\}\boldsymbol{\Delta}_n^{-3/2}\mathbf{I}_{p,q}\boldsymbol{\Gamma}_{ni}^{-1/2} = \mathbf{I}_d.$$

We now proceed to $\sum_{j=1}^n \mathbb{E}(\|\boldsymbol{\xi}_j\|_2^3)$, $\sum_{j=1}^n \mathbb{E}(\|\boldsymbol{\xi}'_j\|_2^3)$, and $\mathbb{E}(\|\sum_{j=1}^n \boldsymbol{\xi}_j\|_2)$, $\mathbb{E}(\|\sum_{j=1}^n \boldsymbol{\xi}'_j\|_2)$. For the first two terms, under Assumption 3 (i), we have

$$\begin{aligned}\sum_{j=1}^n \mathbb{E}(\|\boldsymbol{\xi}_j\|_2^3) &= \frac{1}{(n\rho_n)^{3/2}}\sum_{j=1}^n \mathbb{E}|[\mathbf{E}]_{ij}|^3\|\boldsymbol{\Sigma}_{ni}^{-1/2}\mathbf{I}_{p,q}\boldsymbol{\Delta}_n^{-1}\mathbf{x}_j\|_2\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\Delta}_n^{-1}\mathbf{I}_{p,q}\boldsymbol{\Sigma}_{ni}^{-1}\mathbf{I}_{p,q}\boldsymbol{\Delta}_n^{-1}\mathbf{x}_j\\
&\le \frac{\|\boldsymbol{\Sigma}_{ni}^{-1/2}\|_2\|\mathbf{X}\|_{2\to\infty}}{(n\rho_n)^{3/2}\lambda_d(\boldsymbol{\Delta}_n)}\sum_{j=1}^n \mathbb{E}|[\mathbf{E}]_{ij}|^3\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\Delta}_n^{-1}\mathbf{I}_{p,q}\boldsymbol{\Sigma}_{ni}^{-1}\mathbf{I}_{p,q}\boldsymbol{\Delta}_n^{-1}\mathbf{x}_j,\\
\sum_{j=1}^n \mathbb{E}(\|\boldsymbol{\xi}'_j\|_2^3) &= \frac{1}{(n\rho_n)^{3/2}}\sum_{j=1}^n \mathbb{E}|[\mathbf{E}]_{ij}|^3\|\boldsymbol{\Gamma}_{ni}^{-1/2}\mathbf{I}_{p,q}\boldsymbol{\Delta}_n^{-3/2}\mathbf{x}_j\|_2\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\Delta}_n^{-3/2}\mathbf{I}_{p,q}\boldsymbol{\Gamma}_{ni}^{-1}\mathbf{I}_{p,q}\boldsymbol{\Delta}_n^{-3/2}\mathbf{x}_j.\\
&\le \frac{\|\boldsymbol{\Gamma}_{ni}^{-1/2}\|_2\|\mathbf{X}\|_{2\to\infty}}{(n\rho_n)^{3/2}\lambda_d(\boldsymbol{\Delta}_n)^{3/2}}\sum_{j=1}^n \mathbb{E}|[\mathbf{E}]_{ij}|^3\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\Delta}_n^{-3/2}\mathbf{I}_{p,q}\boldsymbol{\Gamma}_{ni}^{-1}\mathbf{I}_{p,q}\boldsymbol{\Delta}_n^{-3/2}\mathbf{x}_j.\end{aligned}$$

For $\mathbb{E}(\|\sum_{j=1}^n \boldsymbol{\xi}_j\|_2)$, we use Jensen's inequality to write

$$\mathbb{E}\left(\left\|\sum_{j=1}^n \boldsymbol{\xi}_j\right\|_2\right) \le \left\{\mathbb{E}\left(\left\|\sum_{j=1}^n \boldsymbol{\xi}_j\right\|_2^2\right)\right\}^{1/2} = \left(\sum_{j=1}^n \mathbb{E}\|\boldsymbol{\xi}_j\|_2^2\right)^{1/2} = \left[\mathrm{tr}\left\{\sum_{j=1}^n \mathbb{E}(\boldsymbol{\xi}_j\boldsymbol{\xi}_j^{\mathrm{T}})\right\}\right]^{1/2} = d^{1/2}.$$

Similarly, we also have $\mathbb{E}(\|\sum_{j=1}^n \boldsymbol{\xi}'_j\|_2) \le d^{1/2}$. This immediately implies that

$$\mathbb{E}\left(\left\|\sum_{j=1}^n \boldsymbol{\xi}_j\right\|_2 \Delta\right) \le d^{1/2}\Delta \quad\text{and}\quad \mathbb{E}\left(\left\|\sum_{j=1}^n \boldsymbol{\xi}'_j\right\|_2 \Delta'\right) \le d^{1/2}\Delta'.$$

We now apply Theorem S1.4 and the aforementioned results to conclude that

$$\begin{aligned}&\sup_{A\in\mathcal{A}}\left|\mathbb{P}\left\{\sqrt{n}\boldsymbol{\Sigma}_{ni}^{-1/2}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i)\in A\right\} - \mathbb{P}(\mathbf{z}\in A)\right|\\
&\quad\lesssim_\sigma \frac{\chi d^{1/2}\|\boldsymbol{\Sigma}_{ni}^{-1/2}\|_2\|\mathbf{X}\|_{2\to\infty}^2}{(n\rho_n)^{1/2}\lambda_d(\boldsymbol{\Delta}_n)^{3/2}}\max\left\{\frac{(\|\mathbf{X}\|_{2\to\infty}^2\vee 1)(\log n\rho_n)^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \frac{1}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \log n\rho_n\right\}\end{aligned}$$

$$+\frac{d^{1/2}\|\boldsymbol{\Sigma}_{ni}^{-1/2}\|_2\|\mathbf{X}\|_{2\to\infty}}{(n\rho_n)^{3/2}\lambda_d(\boldsymbol{\Delta}_n)}\sum_{j=1}^n \mathbb{E}|[\mathbf{E}]_{ij}|^3\mathrm{tr}\left(\boldsymbol{\Sigma}_{ni}^{-1/2}\mathbf{I}_{p,q}\boldsymbol{\Delta}_n^{-1}\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\Delta}_n^{-1}\mathbf{I}_{p,q}\boldsymbol{\Sigma}_{ni}^{-1/2}\right)$$

and

$$\begin{aligned}
&\sup_{A\in\mathcal{A}}\left|\mathbb{P}\left\{n\rho_n^{1/2}\boldsymbol{\Gamma}_{ni}^{-1/2}\mathbf{W}_{\mathbf{X}}^{\mathrm{T}}(\mathbf{W}^*[\mathbf{U}_{\mathbf{A}}]_{i*}-[\mathbf{U}_{\mathbf{P}}]_{i*})\in A\right\}-\mathbb{P}(\mathbf{z}\in A)\right|\\
&\lesssim_\sigma \frac{d^{1/2}\chi\|\boldsymbol{\Gamma}_{ni}^{-1/2}\|_2\|\mathbf{X}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\boldsymbol{\Delta}_n)^{3/2}}\max\left\{\frac{(\log n\rho_n)^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)},\frac{1}{\lambda_d(\boldsymbol{\Delta}_n)^2},\log n\rho_n\right\}\\
&\quad+\frac{d^{1/2}\|\boldsymbol{\Gamma}_{ni}^{-1/2}\|_2\|\mathbf{X}\|_{2\to\infty}}{(n\rho_n)^{3/2}\lambda_d(\boldsymbol{\Delta}_n)^{3/2}}\sum_{j=1}^n\mathbb{E}|[\mathbf{E}]_{ij}|^3\mathrm{tr}\left(\boldsymbol{\Gamma}_{ni}^{-1/2}\mathbf{I}_{p,q}\boldsymbol{\Delta}_n^{-3/2}\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\Delta}_n^{-3/2}\mathbf{I}_{p,q}\boldsymbol{\Gamma}_{ni}^{-1/2}\right)
\end{aligned}$$

for sufficiently large n. This completes the proof.

S5. Proofs for Section 4.1

To prove Theorem 4.2, we need to verify Assumptions 1-5. The technical tools we applied here are based on Section 3.3 of the Supplementary Material of [1]. By the conditions of Theorem 4.2, Assumptions 1 and 2 hold automatically. For Assumption 3, we let $[\mathbf{E}_1]_{ij}=\mathbf{x}_i^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{x}_j(I_{ij}-\rho_n)$ and $[\mathbf{E}_2]_{ij}=\epsilon_{ij}I_{ij}/\rho_n$. Clearly, $[\mathbf{E}]_{ij}=[\mathbf{E}_1]_{ij}+[\mathbf{E}_2]_{ij}$ and $[\mathbf{E}_1]_{ij}$ satisfies Assumption 3 (i). Since

$$\|[\mathbf{E}_2]_{ij}\|_{\psi_2}\leq\frac{1}{\rho_n}\sup_{p\geq1}\frac{1}{\sqrt{p}}\left(\mathbb{E}\left|\epsilon_{ij}I_{ij}\right|^p\right)^{1/p}\leq\frac{1}{\rho_n}\sup_{p\geq1}\frac{1}{\sqrt{p}}(\mathbb{E}|\epsilon_{ij}|^p)^{1/p}\times\sup_{p\geq1}(\mathbb{E}|I_{ij}|^p)^{1/p}\lesssim\tau\rho_n,$$

we see that $[\mathbf{E}_2]_{ij}$ satisfies Assumption 3 (ii). We now work with Assumptions 4 and 5. Define

$$\bar{\varphi}(x)=\begin{cases}\dfrac{4\|\mathbf{X}\|_{2\to\infty}^2}{\lambda_d(\boldsymbol{\Delta}_n)}\sqrt{\dfrac{\log n}{n\rho_n}}\max\left(x,\sqrt{\dfrac{\log n}{n\rho_n}}\right), & \text{if } x>0,\\ 0, & \text{if } x=0,\end{cases}$$

$$\widetilde{\varphi}(x)=\begin{cases}\dfrac{4\|\mathbf{X}\|_{2\to\infty}^2}{\lambda_d(\boldsymbol{\Delta}_n)}\sqrt{\dfrac{\log n}{n\rho_n}}\dfrac{\tau\rho_n}{\|\mathbf{X}\|_{2\to\infty}^2}, & \text{if } x>0,\\ 0, & \text{if } x=0,\end{cases}$$

Let $\varphi(x)=\bar{\varphi}(x)+\widetilde{\varphi}(x)$. Clearly, $\varphi(0)=0$ and $\varphi(x)/x$ is non-increasing in $(0,+\infty)$. Without loss of generality, we may assume that $\mathbf{V}\neq\mathbf{0}_{n\times d}$. By Lemma 16 in [1],

$$\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}\mathbf{V}\|_2\leq n\rho_n\lambda_d(\boldsymbol{\Delta}_n)\varphi\left(\frac{\|\mathbf{V}\|_{\mathrm{F}}}{\sqrt{n}\|\mathbf{V}\|_{2\to\infty}}\right)$$

with probability at least $1-c_0n^{-(1+\xi)}$, where $\xi=1$ and $c_0=5$. To show that the same concentration bound holds for $\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^{(m)}\mathbf{V}\|_2$, we consider $[\mathbf{E}_1^{(m)}]_{ij}$ and $[\mathbf{E}_2^{(m)}]_{ij}$ separately. We may assume that $i\neq m$ without loss of generality. Exploiting the proof of the first assertion of Lemma 16 in [1], we see that

$$\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}_1^{(m)}\mathbf{V}\|_2\leq n\rho_n\lambda_d(\boldsymbol{\Delta}_n)\bar{\varphi}\left(\frac{\|\mathbf{V}\|_{\mathrm{F}}}{\sqrt{n}\|\mathbf{V}\|_{2\to\infty}}\right)$$

with probability at least $1 - 2n^{-(1+\xi)}$, where $\xi = 1$. By the proof of the second assertion of Lemma 16 in [1], we have

$$\begin{aligned}\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}_2^{(m)}\mathbf{V}\|_2 &\leq \tau\rho_n^2\|\mathbf{V}\|_{2\to\infty}\sqrt{\frac{12(n-1)\log(n-1)}{\rho_n}} \leq \tau\rho_n^2\|\mathbf{V}\|_{2\to\infty}\sqrt{\frac{12n\log n}{\rho_n}}\\ &\leq n\rho_n\lambda_d(\mathbf{\Delta}_n)\left\{\frac{4\|\mathbf{X}\|_{2\to\infty}^2}{\lambda_d(\mathbf{\Delta}_n)}\sqrt{\frac{\log n}{n\rho_n}}\frac{\tau\rho_n}{\|\mathbf{X}\|_{2\to\infty}^2}\right\} = n\rho_n\lambda_d(\mathbf{\Delta}_n)\widetilde{\varphi}\left(\frac{\|\mathbf{V}\|_{\mathrm{F}}}{\sqrt{n}\|\mathbf{V}\|_{2\to\infty}}\right)\end{aligned}$$

with probability at least $1 - 4n^{-2}$ for sufficiently large n. Therefore,

$$\begin{aligned}\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^{(m)}\mathbf{V}\|_2 &\leq \|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}_1^{(m)}\mathbf{V}\|_2 + \|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}_2^{(m)}\mathbf{V}\|_2\\ &\leq n\rho_n\lambda_d(\mathbf{\Delta}_n)\left\{\bar{\varphi}\left(\frac{\|\mathbf{V}\|_{\mathrm{F}}}{\sqrt{n}\|\mathbf{V}\|_{2\to\infty}}\right) + \widetilde{\varphi}\left(\frac{\|\mathbf{V}\|_{\mathrm{F}}}{\sqrt{n}\|\mathbf{V}\|_{2\to\infty}}\right)\right\}\\ &= n\rho_n\lambda_d(\mathbf{\Delta}_n)\varphi\left(\frac{\|\mathbf{V}\|_{\mathrm{F}}}{\sqrt{n}\|\mathbf{V}\|_{2\to\infty}}\right)\end{aligned}$$

with probability at least $1 - 6n^{-(\xi+1)}$ with $\xi = 1$. Hence, Assumption 4 holds. For Assumption 5 , we let

$$\gamma = \frac{c_1(\tau\rho_n + \|\mathbf{X}\|_{2\to\infty}^2)}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)},$$

where $c_1 > 0$ is a constant to be determined later. By Lemma 14 in [1], we have

$$\|\mathbf{E}\|_2 \leq c_2(\|\mathbf{X}\|_{2\to\infty}^2 + \tau\rho_n)(n\rho_n)^{1/2}$$

with probability at least $1 - 4n^{-\zeta}$ with $\zeta = 1$, where $c_2 > 1$ is a constant. Then $\|\mathbf{A} - \mathbb{E}\mathbf{A}\|_2 \leq K(n\rho_n)^{1/2}$ with with probability at least $1 - 4n^{-\zeta}$ ($\zeta = 1$) if we select $K = c_2(\|\mathbf{X}\|_{2\to\infty}^2 + \tau)$. Now set $c_1 = \max\{3K, \|\mathbf{X}\|_{2\to\infty}^2\}/(\tau\rho_n + \|\mathbf{X}\|_{2\to\infty}^2)$. By the conditions of Theorem 4.2, we have

$$\gamma = \frac{\max(3K, \|\mathbf{X}\|_{2\to\infty}^2)}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)} \lesssim \sqrt{\frac{\log n}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^2}} \to 0.$$

Note that

$$c_1 \geq \frac{3K}{\tau\rho_n + \|\mathbf{X}\|_{2\to\infty}^2} = \frac{3c_2(\|\mathbf{X}\|_{2\to\infty}^2 + \tau)}{\tau\rho_n + \|\mathbf{X}\|_{2\to\infty}^2} \geq 3c_2 \geq 1.$$

Then by Lemma 12 in [1], we know that

$$\begin{aligned}\varphi(\gamma) &\leq 4\gamma\sqrt{\log n}(1 + \gamma\sqrt{\log n})\\ &\leq \frac{4c_1(\tau\rho_n + \|\mathbf{X}\|_{2\to\infty}^2)}{\lambda_d(\mathbf{\Delta}_n)}\sqrt{\frac{\log n}{n\rho_n}}\left\{1 + \frac{c_1(\tau\rho_n + \|\mathbf{X}\|_{2\to\infty}^2)}{\lambda_d(\mathbf{\Delta}_n)}\sqrt{\frac{\log n}{n\rho_n}}\right\}\\ &\leq \frac{4c_1(\tau\rho_n + \|\mathbf{X}\|_{2\to\infty}^2)}{\lambda_d(\mathbf{\Delta}_n)}\sqrt{\frac{\log n}{n\rho_n}}\left\{1 + \frac{c_1(\tau\rho_n + \|\mathbf{X}\|_{2\to\infty}^2)\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}\sqrt{\frac{\log n}{n\rho_n}}\right\}\end{aligned}$$

$$\leq \frac{8c_1(\tau\rho_n + \|\mathbf{X}\|_{2\to\infty}^2)}{\lambda_d(\mathbf{\Delta}_n)}\sqrt{\frac{\log n}{n\rho_n}}$$

for sufficiently large n by the condition of Theorem 4.2. Note that $\gamma \leq \varphi(\gamma)$. It follows that

$$32\kappa(\mathbf{\Delta}_n)\max\{\gamma,\varphi(\gamma)\} \leq \frac{256c_1(\tau\rho_n + \|\mathbf{X}\|_{2\to\infty}^2)\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}\sqrt{\frac{\log n}{n\rho_n}} \to 0.$$

Thus, Assumptions 1-5 hold, allowing us to apply Theorem 3.1. Again, by Lemma 12 in [1], we have,

$$\varphi(1) \lesssim \gamma\sqrt{\log n} \lesssim \frac{\tau\rho_n + \|\mathbf{X}\|_{2\to\infty}^2}{\lambda_d(\mathbf{\Delta}_n)}\sqrt{\frac{\log n}{n\rho_n}} \lesssim_\tau \frac{\|\mathbf{X}\|_{2\to\infty}^2}{\lambda_d(\mathbf{\Delta}_n)}.$$

Observe that $\kappa(\mathbf{\Delta}_n) = \lambda_1(\mathbf{\Delta}_n)/\lambda_d(\mathbf{\Delta}_n) \leq \|\mathbf{\Delta}_n\|_{\mathrm{F}}/\lambda_d(\mathbf{\Delta}_n) \leq \|\mathbf{X}\|_{2\to\infty}^2/\lambda_d(\mathbf{\Delta}_n)$. Therefore,

$$\chi = \varphi(1) + \frac{\|\mathbf{X}\|_{2\to\infty}^2 \vee 1}{\lambda_d(\mathbf{\Delta}_n)} \lesssim \frac{\|\mathbf{X}\|_{2\to\infty}^2 \vee 1}{\lambda_d(\mathbf{\Delta}_n)} \leq \frac{\|\mathbf{X}\|_{2\to\infty}^4 \vee 1}{\lambda_d(\mathbf{\Delta}_n)}.$$

We now work with the normal approximation and error estimate for $\widetilde{p}_{ij} - p_{ij}$, where $p_{ij} = \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{x}_j$. By construction and the block diagonal structure of $\mathbf{W}$, we have $\mathbf{W}\mathbf{I}_{p,q}\mathbf{W}^{\mathrm{T}} = \mathbf{W}^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{W} = \mathbf{I}_{p,q}$, and hence,

$$\begin{aligned}\widetilde{p}_{ij} - p_{ij} &= \rho_n^{1/2}\mathbf{x}_i^{\mathrm{T}}\mathbf{I}_{p,q}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j) + \rho_n^{1/2}\mathbf{x}_j^{\mathrm{T}}\mathbf{I}_{p,q}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i)\\ &\quad + \theta\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i\|_2\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j\|_2,\end{aligned}$$

where $|\theta| \in [0,1]$ by Cauchy-Schwarz inequality. Let $\mathbf{r}_i = \{\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{-1/2}\mathbf{A}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{I}_{p,q}\}^{\mathrm{T}}\mathbf{e}_i$. By Theorem 3.2 and Result S2.3, we have

$$\begin{aligned}\max(\|\mathbf{r}_i\|_2, \|\mathbf{r}_j\|_2) &\lesssim_\tau \frac{\chi\|\mathbf{X}\|_{2\to\infty}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\sqrt{n\rho_n}\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{(\|\mathbf{X}\|_{2\to\infty}^2 \vee 1)\sqrt{\log n\rho_n}}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, \log n\rho_n\right\}\\ &\lesssim \frac{(\|\mathbf{X}\|_{2\to\infty}^6 \vee 1)}{n\rho_n^{1/2}\lambda_d(\mathbf{\Delta}_n)^{5/2}}\max\left\{\frac{(\|\mathbf{X}\|_{2\to\infty}^2 \vee 1)\sqrt{\log n\rho_n}}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, \log n\rho_n\right\}\end{aligned}$$

with probability at least $1 - c_0 d/(n\rho_n)$ for some absolute constant $c_0 > 0$. For simplicity, let $\beta_n(\mathbf{X})/(n\rho_n^{1/2})$ denote the upper bound above for $\|\mathbf{r}_i\|_2$ and $\|\mathbf{r}_j\|_2$. We can thus further write

$$\sqrt{\frac{n}{\rho_n}}\frac{(\widetilde{p}_{ij} - p_{ij})}{\sigma_{nij}} = \sum_{l=1}^n \frac{[\mathbf{E}]_{il}\left\{\mathbf{x}_j^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_l + \mathbb{1}(l=j)\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_i\right\}}{\sigma_{nij}\sqrt{n\rho_n}} + \sum_{l\neq i}^n \frac{[\mathbf{E}]_{jl}\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_l}{\sigma_{nij}\sqrt{n\rho_n}} + \sqrt{\frac{n}{\rho_n}}\frac{r_{p_{ij}}}{\sigma_{nij}}$$

for $i \neq j$ and

$$\sqrt{\frac{n}{\rho_n}}\frac{(\widetilde{p}_{ii} - p_{ii})}{\sigma_{nii}} = \frac{2}{\sigma_{nii}\sqrt{n\rho_n}}\sum_{l=1}^n [\mathbf{E}]_{il}\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_l + \sqrt{\frac{n}{\rho_n}}\frac{r_{p_{ii}}}{\sigma_{nii}},$$

where

$$r_{p_{ij}} = \begin{cases} \rho_n^{1/2}\mathbf{x}_i^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{r}_j + \rho_n^{1/2}\mathbf{x}_j^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{r}_i + \theta\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i\|_2\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j\|_2, & \text{if } i \neq j, \\ 2\rho_n^{1/2}\mathbf{x}_i^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{r}_i + \theta\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i\|_2^2, & \text{if } i = j. \end{cases}$$

Recall the decomposition

$$\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i = \frac{1}{n\rho_n^{1/2}}\sum_{l=1}^n [\mathbf{E}_1]_{il}\mathbf{I}_{p,q}\mathbf{\Delta}_n^{-1}\mathbf{x}_l + \frac{1}{n\rho_n^{1/2}}\sum_{l=1}^n [\mathbf{E}_2]_{il}\mathbf{I}_{p,q}\mathbf{\Delta}_n^{-1}\mathbf{x}_l + \mathbf{r}_i,$$

where the first term satisfies

$$\begin{aligned} \left\|\frac{1}{n\rho_n^{1/2}}\sum_{l=1}^n [\mathbf{E}_1]_{il}\mathbf{I}_{p,q}\mathbf{\Delta}_n^{-1}\mathbf{x}_l\right\|_2 &\lesssim \frac{\log n\rho_n}{n\rho_n^{1/2}}\|\mathbf{\Delta}_n^{-1}\|_2\|\mathbf{X}\|_{2\to\infty} + \frac{\|\mathbf{X}\|_{2\to\infty}^2\rho_n^{1/2}\sqrt{\log n\rho_n}}{n\rho_n^{1/2}}\sqrt{n}\|\mathbf{\Delta}_n^{-1}\|_2\|\mathbf{X}\|_{2\to\infty} \\ &= \frac{\log n\rho_n}{n\rho_n^{1/2}}\|\mathbf{\Delta}_n^{-1}\|_2\|\mathbf{X}\|_{2\to\infty} + \|\mathbf{X}\|_{2\to\infty}^3\|\mathbf{\Delta}_n^{-1}\|_2\sqrt{\frac{\log n\rho_n}{n}} \\ &\lesssim \frac{\|\mathbf{X}\|_{2\to\infty}^3 \vee 1}{\lambda_d(\mathbf{\Delta}_n)}\sqrt{\frac{\log n\rho_n}{n}} \end{aligned}$$

with probability at least $1 - c_0/(n\rho_n)$ for some absolute constant $c_0 > 0$ by Lemma S2.1, and the second term satisfies

$$\left\|\frac{1}{n\rho_n^{1/2}}\sum_{l=1}^n [\mathbf{E}_2]_{il}\mathbf{I}_{p,q}\mathbf{\Delta}_n^{-1}\mathbf{x}_l\right\|_2 \lesssim_\tau \frac{\rho_n\sqrt{\log n\rho_n}}{n\rho_n^{1/2}}\sqrt{n}\|\mathbf{\Delta}_n^{-1}\|_2\|\mathbf{X}\|_{2\to\infty} \lesssim \frac{\|\mathbf{X}\|_{2\to\infty}^3 \vee 1}{\lambda_d(\mathbf{\Delta}_n)}\sqrt{\frac{\log n\rho_n}{n}}$$

with probability at least $1 - c_0/(n\rho_n)$ for some absolute constant $c_0 > 0$ by Lemma S2.2. This implies that for any fixed $(i, j) \in [n] \times [n]$,

$$\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i\|_2\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j\|_2 \lesssim_\tau \frac{(\|\mathbf{X}\|_{2\to\infty}^6 \vee 1)}{\lambda_d(\mathbf{\Delta}_n)^2}\frac{\log n\rho_n}{n} + \frac{\beta_n^2(\mathbf{X})}{n^2\rho_n}$$

with probability at least $1 - c_0/(n\rho_n)$ for some absolute constant $c_0 > 0$. Therefore, for any fixed $(i, j) \in [n] \times [n]$, we obtain

$$\begin{aligned} \left|\sqrt{\frac{n}{\rho_n}}\frac{r_{p_{ij}}}{\sigma_{nij}}\right| &\lesssim \sqrt{\frac{n}{\rho_n}}\frac{\rho_n^{1/2}\|\mathbf{X}\|_{2\to\infty}\max(\|\mathbf{r}_i\|_2, \|\mathbf{r}_j\|_2)}{\sigma_{nij}} + \sqrt{\frac{n}{\rho_n}}\frac{\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i\|_2\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j\|_2}{\sigma_{nij}} \\ &\lesssim_\tau \frac{\beta_n(\mathbf{X})\|\mathbf{X}\|_{2\to\infty}}{\sigma_{nij}\sqrt{n\rho_n}} + \frac{(\|\mathbf{X}\|_{2\to\infty}^6 \vee 1)\log n\rho_n}{\sigma_{nij}\lambda_d(\mathbf{\Delta}_n)^2\sqrt{n\rho_n}} + \frac{\beta_n(\mathbf{X})^2}{\sigma_{nij}(n\rho_n)^{3/2}}. \end{aligned}$$

In addition, we have

$$\begin{aligned} \mathbb{E}|[\mathbf{E}]_{ij}|^3 &\lesssim \mathbb{E}|[\mathbf{E}_1]_{ij}|^3 + \mathbb{E}|[\mathbf{E}_2]_{ij}|^3 \lesssim \|\mathbf{X}\|_{2\to\infty}^2(\mathbf{x}_i^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{x}_j)^2\rho_n(1-\rho_n) + \tau^3\rho_n^4 \\ &\lesssim_\tau (\|\mathbf{X}\|_{2\to\infty}^2 \vee 1)\{\rho_n(1-\rho_n)(\mathbf{x}_i^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{x}_j)^2 + \tau^2\rho_n^3\}, \end{aligned}$$

which implies that

$$
\begin{aligned}
&\frac{1}{\sigma_{nij}^3(n\rho_n)^{3/2}}\sum_{l=1}^n\mathbb{E}|[\mathbf{E}]_{il}|^3\left|\mathbf{x}_j^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_l+\mathbb{1}(l=j)\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_i\right|^3+\frac{1}{\sigma_{nij}^3(n\rho_n)^{3/2}}\sum_{l\neq i}^n\mathbb{E}|[\mathbf{E}]_{jl}|^3\left|\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_l\right|^3\\
&\quad\lesssim\frac{(\|\mathbf{X}\|_{2\to\infty}^2\vee 1)}{\sigma_{nij}^3(n\rho_n)^{3/2}}\sum_{l=1}^n\{\rho_n(1-\rho_n)(\mathbf{x}_i^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{x}_l)^2+\tau^2\rho_n^3\}\left|\mathbf{x}_j^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_l+\mathbb{1}(l=j)\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_i\right|^3\\
&\qquad+\frac{(\|\mathbf{X}\|_{2\to\infty}^2\vee 1)}{\sigma_{nij}^3(n\rho_n)^{3/2}}\sum_{l\neq i}^n\{\rho_n(1-\rho_n)(\mathbf{x}_j^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{x}_l)^2+\tau^2\rho_n^3\}\left|\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_l\right|^3\\
&\quad\lesssim\frac{(\|\mathbf{X}\|_{2\to\infty}^4\vee 1)}{\sigma_{nij}^3(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\frac{1}{n}\sum_{l=1}^n\{(1-\rho_n)(\mathbf{x}_i^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{x}_l)^2+\tau^2\rho_n^2\}\left\{\mathbf{x}_j^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_l+\mathbb{1}(l=j)\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_i\right\}^2\\
&\qquad+\frac{(\|\mathbf{X}\|_{2\to\infty}^4\vee 1)}{\sigma_{nij}^3(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\frac{1}{n}\sum_{l\neq i}^n\{(1-\rho_n)(\mathbf{x}_j^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{x}_l)^2+\tau^2\rho_n^2\}\left(\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_l\right)^2\\
&\quad=\frac{(\|\mathbf{X}\|_{2\to\infty}^4\vee 1)}{\sigma_{nij}\lambda_d(\mathbf{\Delta}_n)(n\rho_n)^{1/2}},
\end{aligned}
$$

and similarly,

$$
\frac{1}{\sigma_{nii}^3(n\rho_n)^{3/2}}\sum_{l=1}^n\mathbb{E}|[\mathbf{E}]_{il}|^3\left|2\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_l\right|^3\lesssim_\tau\frac{(\|\mathbf{X}\|_{2\to\infty}^4\vee 1)}{\sigma_{nii}\lambda_d(\mathbf{\Delta}_n)(n\rho_n)^{1/2}}.
$$

We finally apply Theorem S1.4 with

$$
(\xi_{il})_{l=1}^n=\left(\frac{[\mathbf{E}]_{il}\left\{\mathbf{x}_j^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_l+\mathbb{1}(l=j)\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_i\right\}}{\sigma_{nij}\sqrt{n\rho_n}}\right)_{l=1}^n,\quad(\xi_{jl})_{l\neq i}^n=\left(\frac{[\mathbf{E}]_{jl}\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_l}{\sigma_{nij}\sqrt{n\rho_n}}\right)_{l\neq i}^n
$$

if $i\neq j$,

$$
(\xi_{il})_{l=1}^n=\left(\frac{2[\mathbf{E}]_{il}\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_l}{\sigma_{nii}\sqrt{n\rho_n}}\right)_{l=1}^n
$$

if $i=j$, and

$$
\begin{aligned}
&\Delta=\Delta^{(il)}=\Delta^{(jl)}=\Delta^{(l)}=C_\tau\left\{\frac{\beta_n(\mathbf{X})\|\mathbf{X}\|_{2\to\infty}}{\sigma_{nij}\sqrt{n\rho_n}}+\frac{(\|\mathbf{X}\|_{2\to\infty}^6\vee 1)\log n\rho_n}{\sigma_{nij}\lambda_d(\mathbf{\Delta}_n)^2\sqrt{n\rho_n}}+\frac{\beta_n(\mathbf{X})^2}{\sigma_{nij}(n\rho_n)^{3/2}}\right\},\\
&D=\sqrt{\frac{n}{\rho_n}}\frac{r_{p_{ij}}}{\sigma_{nij}},\quad\mathcal{O}=\{\mathbf{A}:|D|\leq\Delta\},
\end{aligned}
$$

where $C_\tau>0$ is a constant depending on τ, to conclude that

$$
\sup_{A\in\mathcal{A}}\left|\mathbb{P}\left\{\sqrt{\frac{n}{\rho_n}}\frac{(\widetilde{p}_{ij}-p_{ij})}{\sigma_{nij}}\in A\right\}-\mathbb{P}(z\in A)\right|
$$

$$
\begin{aligned}
&\lesssim_\tau \frac{(\|\mathbf{X}\|_{2\to\infty}^4 \vee 1)}{\sigma_{nij}\lambda_d(\mathbf{\Delta}_n)(n\rho_n)^{1/2}} + \frac{\beta_n(\mathbf{X})\|\mathbf{X}\|_{2\to\infty}}{\sigma_{nij}\sqrt{n\rho_n}} + \frac{(\|\mathbf{X}\|_{2\to\infty}^6 \vee 1)\log n\rho_n}{\sigma_{nij}\lambda_d(\mathbf{\Delta}_n)^2\sqrt{n\rho_n}} + \frac{\beta_n(\mathbf{X})^2}{\sigma_{nij}(n\rho_n)^{3/2}}\\
&\lesssim \frac{\beta_n(\mathbf{X})\|\mathbf{X}\|_{2\to\infty}}{\sigma_{nij}\sqrt{n\rho_n}} + \frac{(\|\mathbf{X}\|_{2\to\infty}^6 \vee 1)\log n\rho_n}{\sigma_{nij}\lambda_d(\mathbf{\Delta}_n)^2\sqrt{n\rho_n}} + \frac{\beta_n(\mathbf{X})^2}{\sigma_{nij}(n\rho_n)^{3/2}}.
\end{aligned}
$$

The proof is thus completed.

S6. Proofs for Section 4.2

S6.1. A sharp concentration inequality

The key technical challenge for the application of Theorem S1.1 lies in finding the function $\varphi(\cdot)$ satisfying condition A4. In the context of a two-block stochastic block model, the authors of [1] showed in Lemma 7 there that $\varphi(x) \propto [\max\{1, \log(1/x)\}]^{-1}$. Lemma S6.1 below is a generalization of Lemma 7 in [1] to general dimension d. Note that it does not follow from the vector Bernstein's inequality (Lemma S1.2) but provides a sharper control of the sum of vector-scaled independent centered Bernoulli random variables.

Lemma S6.1. *Let $y_i \sim \mathrm{Bernoulli}(p_i)$ independently for all $i = 1, \ldots, n$, and suppose $\mathbf{V}$ is a deterministic matrix. Let $\mathbf{v}_i = \mathbf{V}^{\mathrm{T}}\mathbf{e}_i$, $i \in [n]$ and $\rho = \max_{i\in[n]} p_i$. Then for any $\alpha > 0$,*

$$
\mathbb{P}\left\{\left\|\sum_{i=1}^n (y_i - p_i)\mathbf{v}_i\right\|_2 > \frac{(2+\alpha)n\rho\|\mathbf{V}\|_{2\to\infty}}{\mathrm{Log}(\sqrt{n}\|\mathbf{V}\|_{2\to\infty}/\|\mathbf{V}\|_{\mathrm{F}})}\right\} \le 2(d+1)e^{-\alpha n\rho},
$$

where $\mathrm{Log}(x) := \max\{1, \log x\}$.

Proof of Lemma S6.1. The proof is a non-trivial generalization of Lemma 7 in [1]. We follow the "symmetric dilation" trick [1,12] applied in the proof of Lemma S2.2 together with a sharp control of the moment generating function of y_i, which is motivated by [1,8]. Without loss of generality, we may assume that $\|\mathbf{V}\|_{2\to\infty} = 1$, since the event of interest is invariant to rescaling of $\mathbf{V}$. Let

$$
\mathbf{T}(\mathbf{v}_i) = \begin{bmatrix} \mathbf{0}_{d\times d} & \mathbf{v}_i \\ \mathbf{v}_i^{\mathrm{T}} & 0 \end{bmatrix}, \quad \mathbf{Z}_i = (y_i - p_i)\mathbf{T}(\mathbf{v}_i), \quad i = 1, 2, \ldots, n,
$$

and let $\mathbf{S}_n = \sum_{i=1}^n \mathbf{Z}_i$. Clearly, $\|\mathbf{S}_n\|_2 = \max\{\lambda_{\max}(\mathbf{S}_n), \lambda_{\max}(-\mathbf{S}_n)\}$ and $-\mathbf{S}_n = \sum_{i=1}^n(-\mathbf{Z}_i)$. Observe that the spectral decomposition of $\mathbf{T}(\mathbf{v}_i)$ is given by

$$
\mathbf{T}(\mathbf{v}_i) = \mathbf{Q}_i \begin{bmatrix} \|\mathbf{v}_i\|_2 & \\ & -\|\mathbf{v}_i\|_2 \end{bmatrix} \mathbf{Q}_i^{\mathrm{T}} + 0 \times \mathbf{Q}_{i\perp}\mathbf{Q}_{i\perp}^{\mathrm{T}},
$$

where

$$
\mathbf{Q}_i = \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|_2} & \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|_2} \\ 1 & -1 \end{bmatrix}
$$

and $\mathbf{Q}_{i\perp} \in \mathbb{O}(d+1, d-1)$ is the orthogonal complement matrix of $\mathbf{Q}_i$. Then we use the above spectral decomposition to compute the matrix exponentials

$$
\mathbb{E}e^{\theta\mathbf{Z}_i} = p_i\exp\{(1-p_i)\theta\mathbf{T}(\mathbf{v}_i)\} + (1-p_i)\exp\{-p_i\theta\mathbf{T}(\mathbf{v}_i)\}
$$

$$= \mathbf{Q}_i \begin{bmatrix} p_i e^{(1-p_i)\theta\|\mathbf{v}_i\|_2} + (1-p_i)e^{-p_i\theta\|\mathbf{v}_i\|_2} & \\ & p_i e^{-(1-p_i)\theta\|\mathbf{v}_i\|_2} + (1-p_i)e^{p_i\theta\|\mathbf{v}_i\|_2} \end{bmatrix} \mathbf{Q}_i^{\mathrm{T}} + \mathbf{Q}_{i\perp}\mathbf{Q}_{i\perp}^{\mathrm{T}}$$

and

$$\begin{aligned} \mathbb{E}e^{\theta(-\mathbf{Z}_i)} &= p_i \exp\{-(1-p_i)\theta\mathbf{T}(\mathbf{v}_i)\} + (1-p_i)\exp\{p_i\theta\mathbf{T}(\mathbf{v}_i)\} \\ &= \mathbf{Q}_i \begin{bmatrix} p_i e^{-(1-p_i)\theta\|\mathbf{v}_i\|_2} + (1-p_i)e^{p_i\theta\|\mathbf{v}_i\|_2} & \\ & p_i e^{(1-p_i)\theta\|\mathbf{v}_i\|_2} + (1-p_i)e^{-p_i\theta\|\mathbf{v}_i\|_2} \end{bmatrix} \mathbf{Q}_i^{\mathrm{T}} \\ &\quad + \mathbf{Q}_{i\perp}\mathbf{Q}_{i\perp}^{\mathrm{T}}. \end{aligned}$$

Observe the following two basic inequalities: $1 + x \le e^x$ for $x > -1$ and $e^x \le 1 + x + e^r x^2/2$ for $|x| \le r$. We then obtain

$$\begin{aligned} & p_i e^{(1-p_i)\theta\|\mathbf{v}_i\|_2} + (1-p_i)e^{-p_i\theta\|\mathbf{v}_i\|_2} \\ &\quad = \{(1-p_i) + p_i e^{\theta\|\mathbf{v}_i\|_2}\}e^{-p_i\theta\|\mathbf{v}_i\|_2} \le \exp\{p_i(e^{\theta\|\mathbf{v}_i\|_2} - 1) - p_i\theta\|\mathbf{v}_i\|_2\} \\ &\quad \le \exp\left\{p_i\theta\|\mathbf{v}_i\|_2 + p_i\frac{e^{\theta\|\mathbf{V}\|_{2\to\infty}}}{2}\theta^2\|\mathbf{v}_i\|_2^2 - p_i\theta\|\mathbf{v}_i\|_2\right\} = \exp\left\{\frac{e^{\theta\|\mathbf{V}\|_{2\to\infty}}}{2}\theta^2\|\mathbf{v}_i\|_2^2 p_i\right\}, \\ & p_i e^{-(1-p_i)\theta\|\mathbf{v}_i\|_2} + (1-p_i)e^{p_i\theta\|\mathbf{v}_i\|_2} \\ &\quad = \{(1-p_i) + p_i e^{-\theta\|\mathbf{v}_i\|_2}\}e^{p_i\theta\|\mathbf{v}_i\|_2} \le \exp\{p_i(e^{-\theta\|\mathbf{v}_i\|_2} - 1) + p_i\theta\|\mathbf{v}_i\|_2\} \\ &\quad \le \exp\left\{-p_i\theta\|\mathbf{v}_i\|_2 + p_i\frac{e^{\theta\|\mathbf{V}\|_{2\to\infty}}}{2}\theta^2\|\mathbf{v}_i\|_2^2 + p_i\theta\|\mathbf{v}_i\|_2\right\} = \exp\left\{\frac{e^{\theta\|\mathbf{V}\|_{2\to\infty}}}{2}\theta^2\|\mathbf{v}_i\|_2^2 p_i\right\} \end{aligned}$$

for any $\theta > 0$. Namely,

$$\begin{aligned} \mathbb{E}e^{\theta\mathbf{Z}_i} &\preceq \exp\left\{\frac{e^{\theta\|\mathbf{V}\|_{2\to\infty}}}{2}\theta^2\|\mathbf{v}_i\|_2^2 p_i\right\}\mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}} + \mathbf{Q}_{i\perp}\mathbf{Q}_{i\perp}^{\mathrm{T}} \\ &= \exp\left(\frac{1}{2}\theta^2 e^{\theta}\|\mathbf{v}_i\|_2^2 p_i\right)\mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}} + \mathbf{Q}_{i\perp}\mathbf{Q}_{i\perp}^{\mathrm{T}} \\ &= \exp\{g(\theta)\mathbf{M}_i\}, \end{aligned}$$

where

$$g(\theta) = \frac{1}{2}\theta^2 e^{\theta}, \quad \mathbf{M}_i = (p_i\|\mathbf{v}_i\|_2^2)\mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}}.$$

Similarly, we also have $\mathbb{E}e^{\theta(-\mathbf{Z}_i)} \preceq \exp\{g(\theta)\mathbf{M}_i\}$. Now we compute the scale parameter

$$\rho = \lambda_{\max}\left(\sum_{i=1}^n \mathbf{M}_i\right) \le \sum_{i=1}^n \|\mathbf{M}_i\|_2 \le \rho\|\mathbf{V}\|_{\mathrm{F}}^2.$$

Since $\mathbf{M}_i \succeq \mathbf{0}_{d\times d}$, we also see that $\rho > 0$. Now applying Lemma S1.3 yields

$$\mathbb{P}(\|\mathbf{S}_n\|_2 > t) \le \mathbb{P}\{\lambda_{\max}(\mathbf{S}_n) > t\} + \mathbb{P}\{\lambda_{\max}(-\mathbf{S}_n) > t\}$$

$$\leq 2(d+1)\exp\left(-\theta t + \frac{1}{2}\rho\theta^2 e^\theta \|\mathbf{V}\|_{\mathrm{F}}^2\right)$$

for any $\theta > 0$ and $t \in \mathbb{R}$. Set $\theta = \mathrm{Log}(\sqrt{n}/\|\mathbf{V}\|_{\mathrm{F}})$. Since $\|\mathbf{V}\|_{\mathrm{F}} \leq \sqrt{n}\|\mathbf{V}\|_{2\to\infty} = \sqrt{n}$, we see that $\log(\sqrt{n}/\|\mathbf{V}\|_{\mathrm{F}}) > 0$, and hence, $\theta \leq 1 + \log(\sqrt{n}/\|\mathbf{V}\|_{\mathrm{F}})$. It follows that

$$\frac{\rho\theta^2}{2}e^\theta\|\mathbf{V}\|_{\mathrm{F}}^2 \leq \frac{\rho\theta^2}{2}e\sqrt{n}\|\mathbf{V}\|_{\mathrm{F}} = \frac{e\rho n}{2}\frac{\|\mathbf{V}\|_{\mathrm{F}}}{\sqrt{n}}\left\{\mathrm{Log}\left(\frac{\sqrt{n}}{\|\mathbf{V}\|_{\mathrm{F}}}\right)\right\}^2 \leq \frac{en\rho}{2},$$

where we have applied the basic inequality $\mathrm{Log}\, x \leq \sqrt{x}$ for $x \geq 1$. With $t = \{\mathrm{Log}(\sqrt{n}/\|\mathbf{V}\|_{\mathrm{F}})\}^{-1}(2+\alpha)n\rho$, we then obtain

$$\mathbb{P}\left\{\|\mathbf{S}_n\|_2 > \frac{(2+\alpha)n\rho}{\mathrm{Log}(\sqrt{n}/\|\mathbf{V}\|_{\mathrm{F}})}\right\} \leq d\exp\left\{-(2+\alpha)n\rho + \frac{en\rho}{2}\right\} \leq 2(d+1)e^{-\alpha n\rho}.$$

The proof is thus completed. □

S6.2. Proof of Theorem 4.4

We first present two useful results for random graph models.

Result S6.1 (Spectral norm concentration for random graphs). If $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ and the conditions of Theorem 4.4 hold, then for any $c > 0$, there exists some constant $K_c > 0$ only depending on c, such that $\|\mathbf{A}-\mathbf{P}\|_2 \leq K_c(n\rho_n)^{1/2}$ with probability at least $1 - n^{-c}$. This follows exactly from Theorem 5.2 in [7].

Result S6.2 (Concentration bound for $\|\mathbf{E}\|_\infty$ for random graphs). Suppose $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ and the conditions of Theorem 4.4 hold. For any $c > 0$, there exists some constant $K_c > 0$, such that with probability at least $1 - 2n^{-c}$, $\|\mathbf{E}\|_\infty \leq K_c n\rho_n$. This is a consequence of Bernstein's inequality. To see this, we first observe that $||[\mathbf{E}]_{ij}| - \mathbb{E}|[\mathbf{E}]_{ij}|| \leq 1$, $\mathbb{E}|[\mathbf{E}]_{ij}| = 2\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j) \leq 2\rho_n$, and

$$\sum_{j=1}^n \mathbb{E}(|[\mathbf{E}]_{ij}| - \mathbb{E}|[\mathbf{E}]_{ij}|)^2 \leq \sum_{j=1}^n \mathbb{E}(A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)^2 \leq n\rho_n.$$

Then for any $C > 0$, an application of Bernstein's inequality yields

$$\mathbb{P}\left\{\sum_{j=1}^n |[\mathbf{E}]_{ij}| > (C+2)n\rho_n\right\} \leq \mathbb{P}\left\{\sum_{j=1}^n (|[\mathbf{E}]_{ij}| - \mathbb{E}|[\mathbf{E}]_{ij}|) > Cn\rho_n\right\} \leq 2\exp\left(-\frac{3C^2 n\rho_n}{6+4C}\right).$$

The constant C can be selected such that $3C^2n\rho_n/(6+4C) \geq (c+1)\log n$. Now taking $K_c = C + 2$ and applying a union bound over $i \in [n]$ yields that $\|\mathbf{E}\|_\infty \leq K_c n\rho_n$ with probability at least $1 - 2n^{-c}$.

To prove Theorem 4.4, we apply Theorem 3.1 by first verifying Assumptions 1-5. By the definition of random dot product graphs, Assumption 1 automatically holds because $\|\mathbf{X}\|_{2\to\infty} \leq 1$. Assumption 2 also holds automatically by the conditions of Theorem 4.4. Assumption 3 also holds because one can set $[\mathbf{E}_2]_{ij} = 0$ and $[\mathbf{E}_1]_{ij} = A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j$. It remains to verify Assumptions 4 and 5. Let $c \geq 1$ be any fixed constant. By Result S6.1, there exists a constant $K_c \geq 1$ that depends on $c > 0$, such that

$\mathbb{P}\{\|\mathbf{E}\|_2 \le K_c(n\rho_n)^{1/2}\} \ge 1 - n^{-c}$. Set $\varphi(x) = (2+\beta_c)\{\mathrm{Log}(1/x)\}^{-1}\lambda_d(\mathbf{\Delta}_n)^{-1}$ for a constant $\beta_c > 0$ such that $\beta_c n\rho_n \ge (c+2)\log n$. Then with

$$\gamma = \frac{\max\{3K_c, \|\mathbf{X}\|_{2\to\infty}^2\}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)} = \frac{3K_c}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)},$$

we immediately see that

$$32\kappa(\mathbf{\Delta}_n)\{\gamma, \varphi(\gamma)\} \lesssim_c \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)} \max\left\{\frac{1}{(n\rho_n)^{1/2}}, \frac{1}{\log(n\rho_n\lambda_d(\mathbf{\Delta}_n)^2)}\right\} \to 0$$

by the condition of Theorem 4.4. This shows that Assumption 5 holds with $\zeta = c \ge 1$ and $c_0 = 1$. It remains to show that Assumption 4 holds with the previously selected $\varphi(\cdot)$ function. By Lemma S6.1, for any deterministic $\mathbf{V} \in \mathbb{R}^{n\times d}$, we have

$$\begin{aligned}
&\mathbb{P}\left\{\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}\mathbf{V}\|_2 \le n\rho_n\lambda_d(\mathbf{\Delta}_n)\|\mathbf{V}\|_{2\to\infty}\varphi\left(\frac{\|\mathbf{V}\|_{\mathrm{F}}}{\sqrt{n}\|\mathbf{V}\|_{2\to\infty}}\right)\right\}\\
&\quad = \mathbb{P}\left\{\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}\mathbf{V}\|_2 \le \frac{n\rho_n\lambda_d(\mathbf{\Delta}_n)(2+\beta_c)\|\mathbf{V}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)\,\mathrm{Log}(\sqrt{n}\|\mathbf{V}\|_{2\to\infty}/\|\mathbf{V}\|_{\mathrm{F}})}\right\}\\
&\quad \ge 1 - 2d\exp(-\beta_c n\rho_n) \ge 1 - c_0 n^{-(1+\xi)},
\end{aligned}$$

where $\xi = c$ and $c_0 = 2$. To show that the same concentration bound also holds for $\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^{(m)}\mathbf{V}\|_2$, we simply observe that $[\mathbf{E}^{(m)}]_{im}$ can be viewed as a centered Bernoulli random variable whose success probability is zero. Then applying Lemma S6.1 leads to that

$$\mathbb{P}\left\{\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^{(m)}\mathbf{V}\|_2 \le n\rho_n\lambda_d(\mathbf{\Delta}_n)\|\mathbf{V}\|_{2\to\infty}\varphi\left(\frac{\|\mathbf{V}\|_{\mathrm{F}}}{\sqrt{n}\|\mathbf{V}\|_{2\to\infty}}\right)\right\} \ge 1 - c_0 n^{-(1+\xi)},$$

where $c_0 = 2$ and $\xi = c$. To finish the proof, we observe that $\mathbb{E}|[\mathbf{E}]_{ij}|^3 \le \mathbb{E}[\mathbf{E}]_{ij}^2 = \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)$, implying that

$$\begin{aligned}
&\frac{1}{n\rho_n}\sum_{j=1}^n \mathbb{E}|[\mathbf{E}]_{ij}|^3\mathbf{x}_j^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{\Sigma}_{ni}^{-1}\mathbf{\Delta}_n^{-1}\mathbf{x}_j \le \mathrm{tr}(\mathbf{I}_d) = d \le \frac{\|\mathbf{X}\|_{2\to\infty}^2}{\lambda_d(\mathbf{\Delta}_n)} \le \frac{1}{\lambda_d(\mathbf{\Delta}_n)},\\
&\frac{1}{n\rho_n}\sum_{j=1}^n \mathbb{E}|[\mathbf{E}]_{ij}|^3\mathbf{x}_j^{\mathrm{T}}\mathbf{\Delta}_n^{-3/2}\mathbf{\Gamma}_{ni}^{-1}\mathbf{\Delta}_n^{-3/2}\mathbf{x}_j \le \mathrm{tr}(\mathbf{I}_d) = d \le \frac{\|\mathbf{X}\|_{2\to\infty}^2}{\lambda_d(\mathbf{\Delta}_n)} \le \frac{1}{\lambda_d(\mathbf{\Delta}_n)}.
\end{aligned}$$

Also, observe that $\varphi(1) \lesssim \lambda_d(\mathbf{\Delta}_n)^{-1}$ and $\chi = \varphi(1) + (\|\mathbf{X}\|_{2\to\infty}^2 \vee 1)/\lambda_d(\mathbf{\Delta}_n) \lesssim \lambda_d(\mathbf{\Delta}_n)^{-1}$. Then by Theorem 3.1, we have, for each fixed index $i \in [n]$ and for any sufficiently large n,

$$\begin{aligned}
&\sup_{A\in\mathcal{A}}\left|\mathbb{P}\left\{\sqrt{n}\mathbf{\Sigma}_n(\mathbf{x}_i)^{-1/2}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i) \in A\right\} - \mathbb{P}(\mathbf{z}\in A)\right|\\
&\quad \lesssim \frac{d^{1/2}\|\mathbf{\Sigma}_n(\mathbf{x}_i)^{-1/2}\|_2}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^{5/2}}\max\left\{\frac{(\log n\rho_n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, \log n\rho_n\right\},
\end{aligned}$$

$$\sup_{A\in\mathcal{A}}\left|\mathbb{P}\left\{n\rho_n^{1/2}\mathbf{\Gamma}_n(\mathbf{x}_i)^{-1/2}\mathbf{W}_{\mathbf{X}}^{\mathrm{T}}(\mathbf{W}^*[\mathbf{U}_{\mathbf{A}}]_{i*}-[\mathbf{U}_{\mathbf{P}}]_{i*})\in A\right\}-\mathbb{P}(\mathbf{z}\in A)\right|$$
$$\lesssim\frac{d^{1/2}\|\mathbf{\Gamma}_n(\mathbf{x}_i)^{-1/2}\|_2}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^{5/2}}\max\left\{\frac{(\log n\rho_n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)},\frac{1}{\lambda_d(\mathbf{\Delta}_n)^2},\log n\rho_n\right\},$$

where $\mathcal{A}$ is the collection of all convex measurable sets in $\mathbb{R}^d$, and $\mathbf{z}\sim\mathrm{N}_d(\mathbf{0}_d,\mathbf{I}_d)$. The proof is thus completed.

S6.3. Proof of Corollary 4.1

From the proof in Section S6.2, we see that Assumptions 1-5 hold with $\xi=\zeta=c$ and $\varphi(x)\propto\{\mathrm{Log}(1/x)\}^{-1}\lambda_d(\mathbf{\Delta}_n)^{-1}$. By Theorem 3.2 with $t=(c+1)\log n$ and a union bound over $m\in[n]$ for sufficiently large n, we have

$$\|\mathbf{U}_{\mathbf{A}}-\mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}\mathbf{W}^*\|_{2\to\infty}\lesssim_c\frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^2}\max\left\{\frac{(\log n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)},\frac{1}{\lambda_d(\mathbf{\Delta}_n)^2},\log n\right\},$$
$$\left\|\widetilde{\mathbf{X}}\mathbf{W}-\frac{\mathbf{A}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}}{\rho_n^{1/2}}\right\|_{2\to\infty}\lesssim_c\frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^2}\max\left\{\frac{(\log n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)},\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2},\log n\right\}$$

with probability at least $1-c_0n^{-c}$. Also, we observe that

$$\begin{aligned}\|\mathbf{U}_{\mathbf{A}}-\mathbf{U}_{\mathbf{P}}\mathbf{W}^*\|_{2\to\infty}&\leq\|\mathbf{U}_{\mathbf{A}}-\mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}\mathbf{W}^*\|_{2\to\infty}+\|(\mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}-\mathbf{U}_{\mathbf{P}})\mathbf{W}^*\|_{2\to\infty}\\&=\|\mathbf{U}_{\mathbf{A}}-\mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}\mathbf{W}^*\|_{2\to\infty}+\|\mathbf{E}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}\|_{2\to\infty}\\&\leq\|\mathbf{U}_{\mathbf{A}}-\mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}\mathbf{W}^*\|_{2\to\infty}+\frac{\|\mathbf{E}\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}.\end{aligned}$$

By Lemma S2.1 and a union bound, for any $a>0$, we have

$$\begin{aligned}&\mathbb{P}\left\{\|\mathbf{E}\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}>3a\log n\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}+(6a\rho_n\log n)^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{\mathrm{F}}\right\}\\&\leq\sum_{m=1}^n\mathbb{P}\left\{\left\|\sum_{j=1}^n(A_{mj}-\rho_n\mathbf{x}_m^{\mathrm{T}}\mathbf{x}_j)(\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{e}_j)\right\|_2>3a\log n\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}+(6a\rho_n\log n)^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{\mathrm{F}}\right\}\\&\leq 28ne^{-3a\log n}=28n^{-(3a-1)}.\end{aligned}$$

Now we can set $a=(c+1)/3$ to obtain that

$$\begin{aligned}\frac{\|\mathbf{E}\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}&\lesssim_c\frac{(\log n)\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}+\frac{(6\rho_n\log n)^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{\mathrm{F}}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\\&\lesssim\frac{(n\rho_n\log n)^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}+\frac{(n\rho_n\log n)^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\\&=\frac{2(\log n)^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\end{aligned}$$

with probability at least $1 - c_0 n^{-3c}$. Then by the concentration bound for $\|\mathbf{U_A} - \mathbf{A}\mathbf{U_P}\mathbf{S_P}^{-1}\mathbf{W}^*\|_{2\to\infty}$, we have

$$\|\mathbf{U_A} - \mathbf{U_P}\mathbf{W}^*\|_{2\to\infty} \le \|\mathbf{U_A} - \mathbf{A}\mathbf{U_P}\mathbf{S_P}^{-1}\mathbf{W}^*\|_{2\to\infty} + \frac{\|\mathbf{E}\mathbf{U_P}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}$$
$$\lesssim_c \|\mathbf{U_A} - \mathbf{A}\mathbf{U_P}\mathbf{S_P}^{-1}\mathbf{W}^*\|_{2\to\infty} + \frac{\|\mathbf{E}\mathbf{U_P}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)} + \frac{(\log n)^{1/2}\|\mathbf{U_P}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}$$

with probability at least $1 - c_0 n^{-c}$ for sufficiently large n. This completes the proof of the concentration bound for the unscaled eigenvectors $\|\mathbf{U_A} - \mathbf{U_P}\mathbf{W}^*\|_{2\to\infty}$.

For the fourth assertion, we recall the decomposition (4)

$$\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X} = \rho_n^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} + \{\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{-1/2}\mathbf{A}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\}.$$

For the first term, we apply Lemma S2.1 with $t = (c\log n)^{1/2}$ and a union bound over $m \in [n]$ to obtain

$$\|\rho_n^{-1/2}(\mathbf{A} - \mathbf{P})\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\|_{2\to\infty} = \|\mathbf{E}\mathbf{U_P}\mathbf{S_P}^{-1/2}\|_{2\to\infty} \le \max_{m\in[n]} \|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U_P}\|_2\|\mathbf{S_P}^{-1/2}\|_2$$
$$\lesssim_c \frac{(n\rho_n\log n)^{1/2}\|\mathbf{U_P}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^{1/2}} = \frac{(\log n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^{1/2}}\|\mathbf{U_P}\|_{2\to\infty}$$

with probability at least $1 - c_0 n^{-2c}$. The proof is thus completed.

S7. Proofs for Section 4.3

S7.1. Outline of the proof of Theorem 4.7

We first present the outline the proof of Theorem 4.7, which is a non-trivial extension of [19] to sparse graphs. Recall that the ith row of the one-step refinement $\widehat{\mathbf{x}}_i = \widetilde{\mathbf{x}}_i + \rho_n^{1/2}\mathcal{I}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}})^{-1}\nabla_{\mathbf{x}_i}\ell_{\mathbf{A}}(\rho_n^{-1/2}\widetilde{\mathbf{X}})$. Then a simple computation leads to the following decomposition of $\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i$:

$$\begin{aligned}
\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i &= \mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i + \rho_n^{1/2}\mathcal{I}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})^{-1}\nabla_{\mathbf{x}_i}\ell_{\mathbf{A}}(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W}) \\
&= \mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i + \rho_n^{1/2}\mathcal{I}_i(\mathbf{X})^{-1}\left\{\nabla_{\mathbf{x}_i}\ell_{\mathbf{A}}(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W}) - \nabla_{\mathbf{x}_i}\ell_{\mathbf{A}}(\mathbf{X})\right\} &(2)\\
&\quad + \rho_n^{1/2}\left\{\mathcal{I}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})^{-1} - \mathcal{I}_i(\mathbf{X})^{-1}\right\}\nabla_{\mathbf{x}_i}\ell_{\mathbf{A}}(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W}) &(3)\\
&\quad + \rho_n^{1/2}\mathcal{I}_i(\mathbf{X})^{-1}\nabla_{\mathbf{x}_i}\ell_{\mathbf{A}}(\mathbf{X}). &(4)
\end{aligned}$$

Since $\|\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W} - \mathbf{X}\|_{2\to\infty} = o_{\mathbb{P}}(1)$ by Corollary 4.1, it is expected that $\mathcal{I}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})^{-1} \approx \mathcal{I}_i(\mathbf{X})^{-1}$ by the continuous mapping theorem, and hence, term (3) should be comparatively small. Term (4) corresponds to the first term on the right-hand side of (26) and is a sum of independent mean-zero random variables. The non-trivial part is the analysis of the term in line (2). The intuition is that

$$\nabla_{\mathbf{x}_i}\ell_{\mathbf{A}}(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W}) - \nabla_{\mathbf{x}_i}\ell_{\mathbf{A}}(\mathbf{X}) \approx -\mathcal{I}_i(\mathbf{X})(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \mathbf{x}_i)$$

by a first-order Taylor approximation of $\nabla_{\mathbf{x}_i}\ell_{\mathbf{A}}$. However, making the above approximation precise is technically involved because $\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j \neq \mathbf{x}_j$ for all $j \neq i$. In [19], the authors assumed $n\rho_n^5 = \omega((\log n)^2)$ and their proof technique is no longer applicable when $n\rho_n = \Omega(\log n)$. In the present work, we overcome this difficulty by taking advantage of the decoupling strategy developed in Section 3, together with a delicate second-order Taylor approximation analysis.

Now for any $\epsilon > 0$, denote

$$\mathcal{X}_n(\epsilon) := \left\{\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]^{\mathrm{T}} \in \mathbb{R}^{n\times d} : \mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j \in [\epsilon, 1-\epsilon] \quad \text{for all } i, j \in [n]\right\}$$

and for each fixed index $i \in [n]$, define the matrix-valued function $\mathbf{H}_i : \mathcal{X}_n(\delta/2) \to \mathbb{R}^{d\times d}$ by

$$\mathbf{H}_i(\mathbf{V}) = \frac{1}{n}\sum_{j=1}^n \frac{\mathbf{v}_j\mathbf{v}_j^{\mathrm{T}}}{\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j(1-\rho_n\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j)}. \tag{5}$$

We continue the decomposition of $\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_{0i}$ mentioned earlier:

$$\begin{aligned}
\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i &= \mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i + \rho_n^{1/2}\mathcal{I}_i(\mathbf{X})^{-1}\left\{\nabla_{\mathbf{x}_i}\ell_{\mathbf{A}}(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W}) - \nabla_{\mathbf{x}_i}\ell_{\mathbf{A}}(\mathbf{X})\right\}\\
&\quad + \rho_n^{1/2}\left\{\mathcal{I}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})^{-1} - \mathcal{I}_i(\mathbf{X})^{-1}\right\}\nabla_{\mathbf{x}_i}\ell_{\mathbf{A}}(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})\\
&\quad + \rho_n^{1/2}\mathcal{I}_i(\mathbf{X})^{-1}\nabla_{\mathbf{x}_i}\ell_{\mathbf{A}}(\mathbf{X})\\
&= \frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n \frac{[\mathbf{E}]_{ij}\mathbf{G}_n(\mathbf{x}_i)^{-1}\mathbf{x}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\\
&\quad + \mathbf{G}_n(\mathbf{x}_i)^{-1}\{\mathbf{G}_n(\mathbf{x}_i)(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i) + \mathbf{r}_{i1}\} + \mathbf{R}_{i2}\mathbf{r}_{i1} + \mathbf{r}_{i3},
\end{aligned}$$

where

$$\mathbf{r}_{i1} = \frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\left\{\frac{(A_{ij} - \widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j)(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j)}{\rho_n^{-1}\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j(1-\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j)} - \frac{(A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\}, \tag{6}$$

$$\mathbf{R}_{i2} = \mathbf{H}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})^{-1} - \mathbf{G}_n(\mathbf{x}_i)^{-1}, \tag{7}$$

$$\mathbf{r}_{i3} = \frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n \frac{(A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\mathbf{R}_{i2}\mathbf{x}_j.$$

The most challenging part is a sharp concentration bound for $\mathbf{r}_{i1}$. We now sketch the argument for bounding $\mathbf{r}_{i1}$. For any constant $\epsilon \in (0, 1/2)$, define $\mathcal{X}_2(\epsilon) = \{(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^d \times \mathbb{R}^d : \|\mathbf{u}\|_2, \|\mathbf{v}\|_2 \leq 1, \epsilon \leq \mathbf{u}^{\mathrm{T}}\mathbf{v} \leq 1-\epsilon\}$. For each $(\mathbf{u}, \mathbf{v}) \in \mathcal{X}_2(\delta/2)$, define the following functions:

$$\begin{aligned}
\mathbf{g}(\mathbf{u}, \mathbf{v}) &= \frac{\mathbf{v}}{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})},\\
\mathbf{h}_{ij}(\mathbf{u}, \mathbf{v}) &= \frac{(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j - \mathbf{u}^{\mathrm{T}}\mathbf{v})\mathbf{v}}{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})},\\
\boldsymbol{\phi}_{ij}(\mathbf{u}, \mathbf{v}) &= (A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{g}(\mathbf{u}, \mathbf{v}) + \rho_n\mathbf{h}_{ij}(\mathbf{u}, \mathbf{v}).
\end{aligned}$$

Applying a first-order Taylor expansion to $\mathbf{g}$ and $\mathbf{h}$ yields

$$
\begin{aligned}
\mathbf{g}(\mathbf{u},\mathbf{v}) - \mathbf{g}(\mathbf{x}_i,\mathbf{x}_j) &= -\frac{(1-2\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}}{\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\}^2}(\mathbf{u}-\mathbf{x}_i)\\
&\quad + \left\{\frac{\mathbf{I}_d}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)} - \frac{(1-2\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}}{\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\}^2}\right\}(\mathbf{v}-\mathbf{x}_j)\\
&\quad + \mathbf{r}_{\mathbf{g}}(\mathbf{u},\mathbf{v},\mathbf{x}_i,\mathbf{x}_j),\\
\mathbf{h}_{ij}(\mathbf{u},\mathbf{v}) - \mathbf{h}_{ij}(\mathbf{x}_i,\mathbf{x}_j) &= -\frac{\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}(\mathbf{u}-\mathbf{x}_i)}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)} - \frac{\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}(\mathbf{v}-\mathbf{x}_j)}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)} + \mathbf{r}_{\mathbf{h}_{ij}}(\mathbf{u},\mathbf{v},\mathbf{x}_i,\mathbf{x}_j),
\end{aligned}
$$

where $\mathbf{r}_{\mathbf{g}},\mathbf{r}_{\mathbf{h}_{ij}}$ are higher-order remainders of $\mathbf{g}$ and $\mathbf{h}_{ij}$. Then we can write $\mathbf{r}_{i1}$ as

$$
\begin{aligned}
\mathbf{r}_{i1} &= \frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n\left\{\boldsymbol{\phi}_{ij}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i,\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j) - \boldsymbol{\phi}_{ij}(\mathbf{x}_i,\mathbf{x}_j)\right\}\\
&= -\frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n\frac{\rho_n\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\mathbf{x}_i) \qquad (8)\\
&\quad -\frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n\frac{\rho_n\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\mathbf{x}_j) \qquad (9)\\
&\quad -\frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n\left[\frac{[\mathbf{E}]_{ij}(1-2\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}}{\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\}^2}\right](\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\mathbf{x}_i) \qquad (10)\\
&\quad +\frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n\frac{[\mathbf{E}]_{ij}\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{I}_d-(1-2\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}\}}{\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\}^2}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\mathbf{x}_j) \qquad (11)\\
&\quad +\frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n[\mathbf{E}]_{ij}\mathbf{r}_{\mathbf{g}} + \frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n\rho_n\mathbf{r}_{\mathbf{h}_{ij}}, \qquad (12)
\end{aligned}
$$

where we have compressed the notation

$$
\mathbf{r}_{\mathbf{g}} = \mathbf{r}_{\mathbf{g}}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i,\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j,\mathbf{x}_i,\mathbf{x}_j),\quad \mathbf{r}_{\mathbf{h}_{ij}} = \mathbf{r}_{\mathbf{h}_{ij}}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i,\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j,\mathbf{x}_i,\mathbf{x}_j).
$$

Term (8) is the same as $\mathbf{G}_n(\mathbf{x}_i)(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_i)$. In what follows, we are going to work on terms (9), (10), (11), and (12), respectively, and provide sharp concentration bounds for them.

S7.2. Some technical preparations

In this section, we make some technical preparations for the proof of Theorem 4.7. The following lemma provides a concentration bound for $\|\widetilde{\mathbf{X}}\mathbf{W}-\rho_n^{-1/2}\mathbf{A}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\|_{\mathrm{F}}$.

Lemma S7.1. *Let* $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ *and assume the conditions of Theorem 4.4 hold. Then there exists an absolute constant* $c_0 > 0$*, such that given any fixed* $c > 0$*, for all sufficiently large* n *and for all* $t > 0$*, the following event holds with probability at least* $1 - c_0 n^{-c} - c_0 e^{-t}$*:*

$$\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{A}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\|_{\mathrm{F}} \lesssim_c \frac{1}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)} \max\left\{t^{1/2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}\right\},$$
$$\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}\|_{\mathrm{F}} \lesssim_c \frac{1}{\lambda_d(\mathbf{\Delta}_n)}.$$

Proof of Lemma S7.1. We first remark that this lemma does not follow from Theorem 3.2. Instead, we rely on the following matrix decomposition due to [2] and [16]:

$$\begin{aligned}
\widetilde{\mathbf{X}} - \mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}^* &= \mathbf{E}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}\mathbf{W}^* - \mathbf{U}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}}\mathbf{W}^*\mathbf{S}_{\mathbf{A}}^{-1/2} \\
&\quad + (\mathbf{I} - \mathbf{U}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}})\mathbf{E}(\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*)\mathbf{S}_{\mathbf{A}}^{-1/2} + \mathbf{U}_{\mathbf{P}}(\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}} - \mathbf{W}^*)\mathbf{S}_{\mathbf{A}}^{-1/2} \\
&\quad + \mathbf{U}_{\mathbf{P}}(\mathbf{W}^*\mathbf{S}_{\mathbf{A}}^{1/2} - \mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}^*) - \mathbf{E}\mathbf{U}_{\mathbf{P}}(\mathbf{S}_{\mathbf{P}}^{-1/2}\mathbf{W}^* - \mathbf{W}^*\mathbf{S}_{\mathbf{A}}^{-1/2}).
\end{aligned}$$

Denote $\mathbf{E} = \mathbf{A} - \mathbf{P}$. Since $\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}^* = \rho_n^{1/2}\mathbf{X}\mathbf{W}^{\mathrm{T}}$ and $\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}\mathbf{W}^* = \rho_n^{-1/2}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{W}^{\mathrm{T}}$, it follows that

$$\begin{aligned}
\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X} &= \rho_n^{-1/2}\mathbf{E}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} - \mathbf{U}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}}\mathbf{W}^*\mathbf{S}_{\mathbf{A}}^{-1/2}\mathbf{W} \\
&\quad + (\mathbf{I} - \mathbf{U}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}})\mathbf{E}(\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*)\mathbf{S}_{\mathbf{A}}^{-1/2}\mathbf{W} + \mathbf{U}_{\mathbf{P}}(\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}} - \mathbf{W}^*)\mathbf{S}_{\mathbf{A}}^{-1/2}\mathbf{W} \\
&\quad + \mathbf{U}_{\mathbf{P}}(\mathbf{W}^*\mathbf{S}_{\mathbf{A}}^{1/2} - \mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}^*)\mathbf{W} - \mathbf{E}\mathbf{U}_{\mathbf{P}}(\mathbf{S}_{\mathbf{P}}^{-1/2}\mathbf{W}^* - \mathbf{W}^*\mathbf{S}_{\mathbf{A}}^{-1/2})\mathbf{W}.
\end{aligned}$$

Denote $\mathbf{R}_{\mathbf{X}} = \widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{-1/2}\mathbf{A}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$. Using Davis-Kahan theorem and the fact that $\mathbf{R}_{\mathbf{X}} = \widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X} - \rho_n^{-1/2}\mathbf{E}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$, we obtain

$$\begin{aligned}
\|\mathbf{R}_{\mathbf{X}}\|_{\mathrm{F}} &\le \sqrt{d}\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}}\|_2\|\mathbf{S}_{\mathbf{A}}^{-1/2}\|_2 + \|\mathbf{E}\|_2\|\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*\|_{\mathrm{F}}\|\mathbf{S}_{\mathbf{A}}^{-1/2}\|_2 \\
&\quad + \sqrt{d}\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}} - \mathbf{W}^*\|_2\|\mathbf{S}_{\mathbf{A}}^{-1/2}\|_2 + \|\mathbf{W}^*\mathbf{S}_{\mathbf{A}}^{1/2} - \mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}^*\|_{\mathrm{F}} \\
&\quad + \|\mathbf{E}\|_2\|\mathbf{W}^*\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{S}_{\mathbf{P}}^{-1/2}\mathbf{W}^*\|_{\mathrm{F}} \\
&\lesssim \sqrt{d}\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}}\|_2\|\mathbf{S}_{\mathbf{A}}^{-1/2}\|_2 + \frac{\sqrt{d}\|\mathbf{E}\|_2^2}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\|\mathbf{S}_{\mathbf{A}}^{-1/2}\|_2 + \frac{\sqrt{d}\|\mathbf{E}\|_2^2}{(n\rho_n)^2\lambda_d(\mathbf{\Delta}_n)^2}\|\mathbf{S}_{\mathbf{A}}^{-1/2}\|_2 \\
&\quad + \|\mathbf{W}^*\mathbf{S}_{\mathbf{A}}^{1/2} - \mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}^*\|_{\mathrm{F}} + \|\mathbf{E}\|_2\|\mathbf{W}^*\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{S}_{\mathbf{P}}^{-1/2}\mathbf{W}^*\|_{\mathrm{F}}.
\end{aligned}$$

By Lemma S3.1, Lemma S2.3, Result S6.1, and Result S2.2, for sufficiently large n,

$$\begin{aligned}
&\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}}\|_2 \lesssim_c d^{1/2} + t^{1/2}, \quad \|\mathbf{E}\|_2 \lesssim_c (n\rho_n)^{1/2}, \quad \|\mathbf{S}_{\mathbf{A}}^{-1/2}\|_2 \lesssim_c \frac{1}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^{1/2}} \\
&\|\mathbf{W}^*\mathbf{S}_{\mathbf{A}}^{1/2} - \mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}^*\|_{\mathrm{F}} \lesssim_c \frac{1}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)} \max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}\right\}, \\
&\|\mathbf{W}^*\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{S}_{\mathbf{P}}^{-1/2}\mathbf{W}^*\|_{\mathrm{F}} \lesssim_c \frac{1}{(n\rho_n)^{3/2}\lambda_d(\mathbf{\Delta}_n)^2} \max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}\right\}
\end{aligned}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$. Here, we have used the fact that

$$d^{1/2} \le d \le \lambda_d(\mathbf{\Delta}_n)^{-1}\|\mathbf{X}\|_{2\to\infty}^2 \le \kappa(\mathbf{\Delta}_n)/\lambda_d(\mathbf{\Delta}_n)$$

from Result S2.3. Hence, we conclude that

$$\begin{aligned}
\|\mathbf{R}_{\mathbf{X}}\|_{\mathrm{F}} &\lesssim_c \frac{d + (dt)^{1/2}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^{1/2}} + \frac{d^{1/2}(n\rho_n)}{(n\rho_n)^{3/2}\lambda_d(\mathbf{\Delta}_n)^{3/2}} \\
&\quad + \frac{d^{1/2}(n\rho_n)}{(n\rho_n)^{5/2}\lambda_d(\mathbf{\Delta}_n)^{5/2}} + \frac{1}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)} \max\left\{t^{1/2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}\right\} \\
&\quad + \frac{1}{(n\rho_n)\lambda_d(\mathbf{\Delta}_n)^2} \max\left\{t^{1/2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}\right\} \\
&\lesssim \frac{1}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)} \max\left\{t^{1/2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}\right\}
\end{aligned}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$ for sufficiently large n. This completes the proof of the first assertion. The second assertion follows from the fact that

$$\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}\|_{\mathrm{F}} \le \frac{1}{n\rho_n^{1/2}}\|\mathbf{E}\|_2\|\mathbf{X}\|_{\mathrm{F}}\|\mathbf{\Delta}_n^{-1}\|_2 + \|\mathbf{R}_{\mathbf{X}}\|_{\mathrm{F}},$$

Result S6.1, and the assumption that $\kappa(\mathbf{\Delta}_n)/\{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)\} \to 0$. □

Lemma S7.2. *Suppose* $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ *and let the conditions of Theorem 4.4 hold. Then there exists an absolute constant* $c_0 > 0$*, such that given any fixed* $c > 0$*, for each fixed* $i \in [n]$*, for sufficiently large* n *and for all* $t \ge 1$*,* $t \lesssim n\rho_n$*,*

$$\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i\|_2 \lesssim_c \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)} \max\left\{\frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, t\right\}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$.

Proof of Lemma S7.2. By Theorem 3.2 and Lemma S2.1, for sufficiently large n and for all $t \ge 1$, $t \lesssim n\rho_n$,

$$\begin{aligned}
\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i\|_2 &\le \frac{1}{n\rho_n^{1/2}}\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}\mathbf{X}\mathbf{\Delta}_n^{-1}\|_2 + \|\mathbf{e}_i^{\mathrm{T}}(\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{-1/2}\mathbf{A}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1})\|_2 \\
&\lesssim_c \frac{(n\rho_n t)^{1/2}\|\mathbf{X}\|_{2\to\infty}\|\mathbf{\Delta}_n^{-1}\|_2}{n\rho_n^{1/2}} \\
&\quad + \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^2} \max\left\{\frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, t\right\} \\
&\le \frac{t^{1/2}}{\sqrt{n}\lambda_d(\mathbf{\Delta}_n)} + \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)} \max\left\{\frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, t\right\} \\
&\lesssim \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)} \max\left\{\frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, t\right\}
\end{aligned}$$

with probability at least $1-c_0n^{-c}-c_0e^{-t}$. Note that the tail probability c_0de^{-t} can be replaced by c_0e^{-t} when A_{ij}'s are Bernoulli random variables because the vector Bernstein's inequality (Lemma S2.1) is dimension free. The proof is thus completed. □

We finally present the following lemma that characterizes the Taylor expansion behavior of the functions $\mathbf{g}$ and $\mathbf{h}_{ij}$ defined in Section S7.1.

Lemma S7.3. *Let $\mathcal{X}_2(\epsilon)$, $\mathbf{g}(\mathbf{u},\mathbf{v})$, $\mathbf{h}_{ij}(\mathbf{u},\mathbf{v})$, $\mathbf{r}_{\mathbf{g}}(\mathbf{u},\mathbf{v},\mathbf{x}_i,\mathbf{x}_j)$, and $\mathbf{r}_{\mathbf{h}_{ij}}(\mathbf{u},\mathbf{v},\mathbf{x}_i,\mathbf{x}_j)$ be defined as in Section S7.1. Suppose $(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{X}(\delta)$. Then:*

(a) *For all $(\mathbf{u},\mathbf{v})\in\mathcal{X}_2(\delta/2)$,*

$$\left\|\mathbf{g}(\mathbf{u},\mathbf{v})-\mathbf{g}(\mathbf{x}_i,\mathbf{x}_j)\right\|_2\lesssim\frac{1}{\delta^4}(\|\mathbf{u}-\mathbf{x}_i\|_2+\|\mathbf{v}-\mathbf{x}_j\|_2),$$
$$\left\|\mathbf{h}_{ij}(\mathbf{u},\mathbf{v})-\mathbf{h}_{ij}(\mathbf{x}_i,\mathbf{x}_j)\right\|_2\lesssim\frac{1}{\delta^4}(\|\mathbf{u}-\mathbf{x}_i\|_2+\|\mathbf{v}-\mathbf{x}_j\|_2).$$

(b) *For all $(\mathbf{u},\mathbf{v})\in\mathcal{X}_2(\delta/2)$,*

$$\|\mathbf{r}_{\mathbf{g}}(\mathbf{u},\mathbf{v},\mathbf{x}_i,\mathbf{x}_j)\|_2\lesssim\frac{d^{1/2}}{\delta^6}(\|\mathbf{u}-\mathbf{x}_i\|_2^2+\|\mathbf{v}-\mathbf{x}_j\|_2^2),$$
$$\|\mathbf{r}_{\mathbf{h}_{ij}}(\mathbf{u},\mathbf{v},\mathbf{x}_i,\mathbf{x}_j)\|_2\lesssim\frac{d^{1/2}}{\delta^6}(\|\mathbf{u}-\mathbf{x}_i\|_2^2+\|\mathbf{v}-\mathbf{x}_j\|_2^2).$$

Proof of Lemma S7.3. For each $k\in[d]$, denote

$$g_k(\mathbf{u},\mathbf{v})=\frac{\mathbf{e}_k^{\mathrm{T}}\mathbf{v}}{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})},\quad h_{ijk}(\mathbf{u},\mathbf{v})=\frac{(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j-\mathbf{u}^{\mathrm{T}}\mathbf{v})\mathbf{e}_k^{\mathrm{T}}\mathbf{v}}{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})}.$$

A simple algebra shows that the gradients of g_k and h_{ijk} are

$$\begin{aligned}
&\frac{\partial g_k}{\partial\mathbf{u}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})=\frac{(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\mathbf{e}_k^{\mathrm{T}}\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^2},\quad\frac{\partial g_k}{\partial\mathbf{v}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})=\frac{\mathbf{I}_d}{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})}+\frac{(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\mathbf{e}_k^{\mathrm{T}}\mathbf{v}\mathbf{u}^{\mathrm{T}}}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^2},\\
&\frac{\partial h_{ijk}}{\partial\mathbf{u}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})=-\frac{\mathbf{e}_k^{\mathrm{T}}\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})}-\frac{(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j-\mathbf{u}^{\mathrm{T}}\mathbf{v})(\mathbf{e}_k^{\mathrm{T}}\mathbf{v})\mathbf{v}^{\mathrm{T}}}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^2},\\
&\frac{\partial h_{ijk}}{\partial\mathbf{v}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})=\frac{(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j-\mathbf{u}^{\mathrm{T}}\mathbf{v})\mathbf{e}_k^{\mathrm{T}}}{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})}-\frac{\mathbf{e}_k^{\mathrm{T}}\mathbf{v}\mathbf{u}^{\mathrm{T}}}{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})}-\frac{(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j-\mathbf{u}^{\mathrm{T}}\mathbf{v})(\mathbf{e}_k^{\mathrm{T}}\mathbf{v})\mathbf{u}^{\mathrm{T}}}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^2}.
\end{aligned}$$

Clearly,

$$\sup_{(\mathbf{u},\mathbf{v})\in\mathcal{X}_2(\delta/2)}\max\left\{\left\|\frac{\partial\mathbf{g}}{\partial\mathbf{u}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})\right\|_2+\left\|\frac{\partial\mathbf{g}}{\partial\mathbf{v}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})\right\|_2,\left\|\frac{\partial\mathbf{h}_{ij}}{\partial\mathbf{u}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})\right\|_2+\left\|\frac{\partial\mathbf{h}_{ij}}{\partial\mathbf{v}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})\right\|_2\right\}\lesssim\frac{1}{\delta^4}.$$

Then assertion (a) then follows directly from the mean-value inequality for vector-valued functions. To prove assertion (b), we need to first compute the Hessian of g_k:

$$\frac{\partial^2 g_k}{\partial\mathbf{u}\partial\mathbf{u}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})=\left[\frac{2\rho_n(\mathbf{e}_k^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^2}+\frac{2(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})^2(\mathbf{e}_k^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^3}\right]\mathbf{v}\mathbf{v}^{\mathrm{T}},$$

$$\frac{\partial^2 g_k}{\partial \mathbf{v}\partial \mathbf{u}^{\mathrm{T}}}(\mathbf{u},\mathbf{v}) = \left[\frac{2\rho_n(\mathbf{e}_k^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^2} + \frac{2(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})^2(\mathbf{e}_k^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^3}\right]\mathbf{u}\mathbf{v}^{\mathrm{T}}$$
$$- \frac{(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v}\}^2}(\mathbf{e}_k\mathbf{v}^{\mathrm{T}} + \mathbf{e}_k^{\mathrm{T}}\mathbf{v}\mathbf{I}_d),$$
$$\frac{\partial^2 g_k}{\partial \mathbf{v}\partial \mathbf{v}^{\mathrm{T}}}(\mathbf{u},\mathbf{v}) = \left[\frac{2\rho_n(\mathbf{e}_k^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^2} + \frac{2(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})^2(\mathbf{e}_k^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^3}\right]\mathbf{u}\mathbf{u}^{\mathrm{T}}$$
$$- \frac{(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^2}(\mathbf{e}_k\mathbf{u}^{\mathrm{T}} + \mathbf{u}\mathbf{e}_k^{\mathrm{T}}).$$

Since $(\mathbf{u},\mathbf{v}) \in \mathcal{X}_2(\delta/2)$ and $\rho_n \in (0,1]$, we see that

$$\sup_{(\mathbf{u},\mathbf{v})\in\mathcal{X}_2(\delta/2)} \max\left\{\left\|\frac{\partial^2 g_k}{\partial \mathbf{u}\partial \mathbf{u}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})\right\|_2, \left\|\frac{\partial^2 g_k}{\partial \mathbf{v}\partial \mathbf{u}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})\right\|_2, \left\|\frac{\partial^2 g_k}{\partial \mathbf{v}\partial \mathbf{v}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})\right\|_2\right\} \lesssim \frac{1}{\delta^6}.$$

By the mean-value inequality, for any $(\mathbf{u}_1,\mathbf{v}_1),(\mathbf{u}_2,\mathbf{v}_2) \in \mathcal{X}_2(\delta/2)$,

$$\left\|\begin{bmatrix}\frac{\partial g_k}{\partial \mathbf{u}}(\mathbf{u}_1,\mathbf{v}_1)\\ \frac{\partial g_k}{\partial \mathbf{v}}(\mathbf{u}_1,\mathbf{v}_1)\end{bmatrix} - \begin{bmatrix}\frac{\partial g_k}{\partial \mathbf{u}}(\mathbf{u}_2,\mathbf{v}_2)\\ \frac{\partial g_k}{\partial \mathbf{v}}(\mathbf{u}_2,\mathbf{v}_2)\end{bmatrix}\right\|_2 \lesssim \frac{1}{\delta^6}\left\|\begin{bmatrix}\mathbf{u}_1-\mathbf{u}_2\\ \mathbf{v}_1-\mathbf{v}_2\end{bmatrix}\right\|_2.$$

Namely, the gradient of g_k is Lipschitz continuous over $\mathcal{X}_2(\delta/2)$ with a Lipschitz constant upper bounded by an absolute constant factor of $1/\delta^6$. By Taylor's theorem, for any $(\mathbf{u},\mathbf{v}) \in \mathcal{X}_2(\delta/2)$,

$$|\mathbf{e}_k^{\mathrm{T}}\mathbf{r}_{\mathbf{g}}(\mathbf{u},\mathbf{v},\mathbf{x}_i,\mathbf{x}_j)| \lesssim \frac{\|\mathbf{u}-\mathbf{x}_i\|_2^2 + \|\mathbf{v}-\mathbf{x}_j\|_2^2}{\delta^6},$$

and hence,

$$\|\mathbf{r}_{\mathbf{g}}(\mathbf{u},\mathbf{v},\mathbf{x}_i,\mathbf{x}_j)\|_2 = \left(\sum_{k=1}^d |\mathbf{e}_k^{\mathrm{T}}\mathbf{r}_{\mathbf{g}}(\mathbf{u},\mathbf{v},\mathbf{x}_i,\mathbf{x}_j)|^2\right)^{1/2} \lesssim \frac{\sqrt{d}}{\delta^6}(\|\mathbf{u}-\mathbf{x}_i\|_2^2 + \|\mathbf{v}-\mathbf{x}_j\|_2^2).$$

The Hessian of h_{ijk} can be computed similarly:

$$\frac{\partial^2 h_{ijk}}{\partial \mathbf{u}\partial \mathbf{u}^{\mathrm{T}}}(\mathbf{u},\mathbf{v}) = \left[\frac{2(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v}) + 2\rho_n(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j - \mathbf{u}^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^2} + \frac{2(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})^2(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j - \mathbf{u}^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^3}\right]$$
$$\times (\mathbf{e}_k^{\mathrm{T}}\mathbf{v})\mathbf{v}\mathbf{v}^{\mathrm{T}},$$
$$\frac{\partial^2 h_{ijk}}{\partial \mathbf{v}\partial \mathbf{u}^{\mathrm{T}}}(\mathbf{u},\mathbf{v}) = \left[\frac{2(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v}) + 2\rho_n(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j - \mathbf{u}^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^2} + \frac{2(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})^2(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j - \mathbf{u}^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^3}\right]$$
$$\times (\mathbf{e}_k^{\mathrm{T}}\mathbf{v})\mathbf{u}\mathbf{v}^{\mathrm{T}} - \left[\frac{1}{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})} + \frac{(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j - \mathbf{u}^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v}\}^2}\right]$$
$$\times (\mathbf{e}_k\mathbf{v}^{\mathrm{T}} + \mathbf{e}_k^{\mathrm{T}}\mathbf{v}\mathbf{I}_d),$$
$$\frac{\partial^2 h_{ijk}}{\partial \mathbf{v}\partial \mathbf{v}^{\mathrm{T}}}(\mathbf{u},\mathbf{v}) = -\left[\frac{1}{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})} + \frac{(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j - \mathbf{u}^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v}\}^2}\right](\mathbf{u}\mathbf{e}_k^{\mathrm{T}} + \mathbf{e}_k\mathbf{u}^{\mathrm{T}})$$

$$+\left[\frac{2(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})+2\rho_n(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j-\mathbf{u}^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^2}+\frac{2(1-2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})^2(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j-\mathbf{u}^{\mathrm{T}}\mathbf{v})}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^3}\right]$$
$$\times(\mathbf{e}_k^{\mathrm{T}}\mathbf{v})\mathbf{u}\mathbf{u}^{\mathrm{T}}.$$

This implies that

$$\sup_{(\mathbf{u},\mathbf{v})\in\mathcal{X}_2(\delta/2)}\max\left\{\left\|\frac{\partial^2 h_{ijk}}{\partial\mathbf{u}\partial\mathbf{u}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})\right\|_2,\left\|\frac{\partial^2 h_{ijk}}{\partial\mathbf{v}\partial\mathbf{u}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})\right\|_2,\left\|\frac{\partial^2 h_{ijk}}{\partial\mathbf{v}\partial\mathbf{v}^{\mathrm{T}}}(\mathbf{u},\mathbf{v})\right\|_2\right\}\lesssim\frac{1}{\delta^6}.$$

An identical argument shows that

$$\|\mathbf{r}_{\mathbf{h}_{ij}}(\mathbf{u},\mathbf{v},\mathbf{x}_i,\mathbf{x}_j)\|_2\lesssim\frac{\sqrt{d}}{\delta^6}(\|\mathbf{u}-\mathbf{x}_i\|_2^2+\|\mathbf{v}-\mathbf{x}_j\|_2^2).$$

□

S7.3. Concentration bound for (9)

Lemma S7.4. *Let* $\mathbf{A}\sim\mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ *with and assume the conditions of Theorem 4.7 hold. Then there exists an absolute constant* $c_0>0$*, such that given any fixed* $c>0$*, for each fixed row index* $i\in[n]$*, for all* $t\geq 1$*,* $t\lesssim n\rho_n$*, and for sufficiently large* n*, with probability at least* $1-c_0n^{-c}-c_0e^{-t}$*,*

$$\left\|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\frac{\rho_n\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\mathbf{x}_j)\right\|\lesssim_c\frac{1}{n\rho_n^{1/2}\delta^2\lambda_d(\boldsymbol{\Delta}_n)}\max\left\{\frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)},t^{1/2}\right\}.$$

Proof of Lemma S7.4. Denote $\mathbf{R}_{\mathbf{X}}=\widetilde{\mathbf{X}}\mathbf{W}-\rho_n^{-1/2}\mathbf{A}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$. By the decomposition (4), for any $j\in[n]$, we have

$$\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\rho_n^{1/2}\mathbf{x}_j=\frac{1}{n\sqrt{\rho_n}}\sum_{a=1}^n(A_{ja}-\rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_a)\boldsymbol{\Delta}_n^{-1}\mathbf{x}_a+\mathbf{R}_{\mathbf{X}}^{\mathrm{T}}\mathbf{e}_j.$$

It follows that

$$\begin{aligned}
&\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\frac{\rho_n\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\mathbf{x}_j)\\
&=\frac{1}{n^2\sqrt{\rho_n}}\sum_{j=1}^n\sum_{a=1}^n\frac{\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\Delta}_n^{-1}\mathbf{x}_a}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}(A_{ja}-\rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_a)+\frac{1}{n}\sum_{j=1}^n\frac{\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}\mathbf{R}_{\mathbf{X}}^{\mathrm{T}}\mathbf{e}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\\
&=\frac{1}{n\sqrt{\rho_n}}\left\{\sum_{j\leq a}\mathbf{z}_{ija}+\sum_{j>a}\mathbf{z}_{ija}\right\}+\frac{1}{n}\sum_{j=1}^n\frac{\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}\mathbf{R}_{\mathbf{X}}^{\mathrm{T}}\mathbf{e}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)},
\end{aligned}$$

where

$$\mathbf{z}_{ija}=\frac{(A_{ja}-\rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_a)\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\Delta}_n^{-1}\mathbf{x}_a}{n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}.$$

By Lemma S2.1, with $t \geq 1$ and $t \lesssim n\rho_n$, we see that

$$\frac{1}{n\sqrt{\rho_n}}\left\|\sum_{j\leq a}\mathbf{z}_{ija}\right\|_2 + \frac{1}{n\sqrt{\rho_n}}\left\|\sum_{j>a}\mathbf{z}_{ija}\right\|_2 \lesssim_c \left\{\frac{t + (n^2\rho_n t)^{1/2}}{n\sqrt{\rho_n}}\right\}\max_{j,a\in[n]}\left\|\frac{\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_a}{n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\|_2$$

$$\lesssim \frac{t^{1/2}}{n\delta^2\lambda_d(\mathbf{\Delta}_n)}$$

with probability at least $1 - c_0e^{-t}$. In addition, by Lemma S7.1, for sufficiently large n,

$$\|\mathbf{R}_{\mathbf{X}}\|_{\mathrm{F}} \lesssim \frac{1}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}\right\}$$

with probability at least $1 - c_0n^{-c} - c_0e^{-t}$ for all $t > 0$. By Cauchy-Schwarz inequality, we have

$$\left\|\frac{1}{n}\sum_{j=1}^n\frac{\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}\mathbf{R}_{\mathbf{X}}^{\mathrm{T}}\mathbf{e}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\|_2 \leq \frac{1}{n}\sum_{j=1}^n\left\|\frac{\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\|_2\|\mathbf{R}_{\mathbf{X}}^{\mathrm{T}}\mathbf{e}_j\|_2$$

$$\leq \frac{1}{n}\left[\sum_{j=1}^n\left\|\frac{\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\|_2^2\right]^{1/2}\|\mathbf{R}_{\mathbf{X}}\|_{\mathrm{F}} \leq \frac{\|\mathbf{R}_{\mathbf{X}}\|_{\mathrm{F}}}{\sqrt{n}\delta^2}$$

$$\lesssim_c \frac{1}{n\rho_n^{1/2}\delta^2\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}\right\},$$

with probability at least $1 - c_0n^{-c} - c_0e^{-t}$ for all $t > 0$ whenever n is sufficiently large. Therefore, we conclude that

$$\left\|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\frac{\rho_n\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \mathbf{x}_j)\right\|$$

$$\lesssim_c \frac{t^{1/2}}{n\delta^2\lambda_d(\mathbf{\Delta}_n)} + \frac{1}{n\rho_n^{1/2}\delta^2\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}\right\}$$

$$\lesssim \frac{1}{n\rho_n^{1/2}\delta^2\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}\right\}$$

with probability at least $1 - c_0n^{-c} - c_0e^{-t}$ for all $t \geq 1$, $t \lesssim n\rho_n$, provided that n is sufficiently large. The proof is thus completed. $\square$

S7.4. Concentration bound for (10)

Lemma S7.5. *Let* $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ *with and assume the conditions of Theorem 4.7 hold. Suppose* $\{\mathbf{B}_{nij} : i, j \in [n]\}$ *is a collection of deterministic* $d \times d$ *matrices with* $\sup_{i,j\in[n]}\|\mathbf{B}_{nij}\|_{\mathrm{F}} \leq \delta^{-4}$. *Then given any fixed* $c > 0$, *for each fixed row index* $i \in [n]$, *for all* $t > 0$, $t \lesssim n\rho_n$, *and sufficiently large* n,

$$\left\|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n[\mathbf{E}]_{ij}\mathbf{B}_{nij}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \mathbf{x}_i)\right\|_2$$

$$\lesssim_c \frac{\|\mathbf{U_P}\|_{2\to\infty}}{(n\rho_n)^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)} \max\left\{\frac{t}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, t^{3/2}\right\}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$, *where* $c_0 > 0$ *is an absolute constant.*

Proof. First observe that by definition of the matrix norm,

$$\begin{aligned}
&\left\|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n [\mathbf{E}]_{ij}\mathbf{B}_{nij}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \mathbf{x}_i)\right\|_2 \\
&\quad\le \frac{1}{\sqrt{\rho_n}}\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i\|_2 \left\|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n \mathbf{B}_{nij}(A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\right\|_2 \\
&\quad\le \frac{1}{\sqrt{\rho_n}}\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i\|_2 \left\|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n \mathrm{vec}(\mathbf{B}_{nij})(A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\right\|_2 .
\end{aligned}$$

By Lemma S2.1, with for all $t \lesssim n\rho_n$ and $t \ge 1$, we have

$$\left\|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n \mathrm{vec}(\mathbf{B}_{nij})(A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\right\|_2 \lesssim \frac{(n\rho_n t)^{1/2}}{n\rho_n^{1/2}} \max_{j\in[n]} \|\mathbf{B}_{nij}\|_{\mathrm{F}} \le \frac{t^{1/2}}{\sqrt{n}\delta^4}$$

with probability at least $1 - c_0 e^{-t}$. Hence, by Lemma S7.2 for sufficiently large n,

$$\begin{aligned}
&\left\|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n [\mathbf{E}]_{ij}\mathbf{B}_{nij}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \mathbf{x}_i)\right\| \\
&\quad\le \frac{\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i\|_2}{\sqrt{\rho_n}} \left\|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n \mathrm{vec}(\mathbf{B}_{nij})(A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\right\|_2 \\
&\quad\lesssim_c \frac{\|\mathbf{U_P}\|_{2\to\infty}}{(n\rho_n)^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)} \max\left\{\frac{t}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, t^{3/2}\right\}
\end{aligned}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$. The proof is thus completed. □

S7.5. Concentration bound for (11)

Lemma S7.6. *Let* $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ *with and assume the conditions of Theorem 4.7 hold. Suppose* $\{\mathbf{B}_{nij} : i, j \in [n]\}$ *is a collection of deterministic* $d \times d$ *matrices such that* $\sup_{i,j\in[n]} \|\mathbf{B}_{nij}\|_2 \le \delta^{-4}$. *Then given any fixed* $c > 0$, *for each fixed index* $i \in [n]$, *for all* $t \ge 1$, $t \lesssim n\rho_n$ *and sufficiently large* n,

$$\begin{aligned}
&\left\|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n [\mathbf{E}]_{ij}\mathbf{B}_{nij}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \mathbf{x}_j)\right\|_2 \\
&\quad\lesssim_c \frac{\|\mathbf{U_P}\|_{2\to\infty}}{(n\rho_n)^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)} \max\left\{\frac{\kappa(\mathbf{\Delta}_n)^{1/2}t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, t\right\}
\end{aligned}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$ *for sufficiently large* n*, where* $c_0 > 0$ *is an absolute constant*

Proof of Lemma S7.6. Let $\mathbf{A}^{(m)}$, $\mathbf{U}_{\mathbf{A}}^{(m)}$, and $\mathbf{H}^{(m)}$, $m = 1,\ldots,n$ be the auxiliary matrices defined in Section 3 of the manuscript. Now we fix the row index $i \in [n]$. Observe that

$$\begin{aligned}\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X} &= \mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{1/2}(\mathbf{W}^*)^{\mathrm{T}}\mathbf{W}_{\mathbf{X}} - \mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}} = (\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{1/2}\mathrm{sgn}(\mathbf{H}) - \mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{1/2})\mathbf{W}_{\mathbf{X}}\\
&= \mathbf{U}_{\mathbf{A}}\{\mathbf{S}_{\mathbf{A}}^{1/2}\mathrm{sgn}(\mathbf{H}) - \mathrm{sgn}(\mathbf{H})\mathbf{S}_{\mathbf{P}}^{1/2}\}\mathbf{W}_{\mathbf{X}} + \mathbf{U}_{\mathbf{A}}\{\mathrm{sgn}(\mathbf{H}) - \mathbf{H}\}\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\\
&\quad + (\mathbf{U}_{\mathbf{A}}\mathbf{H} - \mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)})\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}} + (\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}})\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}.\end{aligned}$$

This immediately leads to the following decomposition of the quantity of interest:

$$\begin{aligned}&\left\{\frac{1}{n\rho_n}\sum_{j=1}^n[\mathbf{E}]_{ij}\mathbf{B}_{nij}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j)\right\}^{\mathrm{T}}\\
&\quad= \frac{1}{n\rho_n}\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{e}_j^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\{\mathbf{S}_{\mathbf{A}}^{1/2}\mathrm{sgn}(\mathbf{H}) - \mathrm{sgn}(\mathbf{H})\mathbf{S}_{\mathbf{P}}^{1/2}\}\mathbf{W}_{\mathbf{X}}\mathbf{B}_{nij}^{\mathrm{T}} &(13)\\
&\qquad+ \frac{1}{n\rho_n}\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{e}_j^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\{\mathrm{sgn}(\mathbf{H}) - \mathbf{H}\}\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\mathbf{B}_{nij}^{\mathrm{T}} &(14)\\
&\qquad+ \frac{1}{n\rho_n}\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{e}_j^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}}\mathbf{H} - \mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)})\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\mathbf{B}_{nij}^{\mathrm{T}} &(15)\\
&\qquad+ \frac{1}{n\rho_n}\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{e}_j^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}})\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\mathbf{B}_{nij}^{\mathrm{T}}. &(16)\end{aligned}$$

For term (13), for all $t > 0$, we apply Result S6.2, Lemma S2.3, and Lemma S2.5 to obtain that for sufficiently large n,

$$\begin{aligned}&\left\|\frac{1}{n\rho_n}\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{e}_j^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\{\mathbf{S}_{\mathbf{A}}^{1/2}\mathrm{sgn}(\mathbf{H}) - \mathrm{sgn}(\mathbf{H})\mathbf{S}_{\mathbf{P}}^{1/2}\}\mathbf{W}_{\mathbf{X}}\mathbf{B}_{nij}^{\mathrm{T}}\right\|_2\\
&\quad\leq \frac{1}{n\rho_n}\sum_{j=1}^n|[\mathbf{E}]_{ij}|\|\mathbf{U}_{\mathbf{A}}\|_{2\to\infty}\|\mathbf{W}^*\mathbf{S}_{\mathbf{A}}^{1/2} - \mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}^*\|_2\|\max_{j\in[n]}\|\mathbf{B}_{nij}\|_2\\
&\quad= \frac{1}{n\rho_n}\|\mathbf{E}\|_\infty\|\mathbf{U}_{\mathbf{A}}\|_{2\to\infty}\|\mathbf{W}^*\mathbf{S}_{\mathbf{A}}^{1/2} - \mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}^*\|_2\|\max_{j\in[n]}\|\mathbf{B}_{nij}\|_2\\
&\quad\lesssim_c \frac{1}{n\rho_n\delta^4}(n\rho_n)\frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)}\frac{1}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}\right\}\\
&\quad= \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)^2}\max\left\{\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2}\right\}\end{aligned}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$. For term (14), we invoke Lemma S2.3, Lemma 6.7 in [4], and the Davis-Kahan theorem to obtain that for sufficiently large n,

$$
\begin{aligned}
&\left\|\frac{1}{n\rho_n}\sum_{j=1}^n (A_{ij} - \rho_n \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{e}_j^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\{\mathrm{sgn}(\mathbf{H}) - \mathbf{H}\}\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\mathbf{B}_{nij}^{\mathrm{T}}\right\|_2\\
&\quad\le \frac{1}{n\rho_n}\sum_{j=1}^n |[\mathbf{E}]_{ij}|\|\mathbf{U}_{\mathbf{A}}\|_{2\to\infty}\|\mathbf{W}^* - \mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\|_2\|\mathbf{S}_{\mathbf{P}}\|_2^{1/2}\max_{j\in[n]}\|\mathbf{B}_{nij}\|_2\\
&\quad\le \frac{1}{n\rho_n}\|\mathbf{E}\|_\infty\|\mathbf{U}_{\mathbf{A}}\|_{2\to\infty}\frac{4\|\mathbf{E}\|_2^2}{\lambda_d(\mathbf{P})^2}\|\mathbf{S}_{\mathbf{P}}\|_2^{1/2}\max_{j\in[n]}\|\mathbf{B}_{nij}\|_2\\
&\quad\lesssim_c \frac{1}{n\rho_n}(n\rho_n)\frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)}\frac{(n\rho_n)}{(n\rho_n)^2\lambda_d(\mathbf{\Delta}_n)^2}\{(n\rho_n)\lambda_1(\mathbf{\Delta}_n)\}^{1/2}\frac{1}{\delta^4}\\
&\quad\lesssim \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)^3}
\end{aligned}
$$

with probability at least $1 - 4n^{-c}$. We now turn the focus to the more complicated terms (15) and (16). Denote $\mathbf{\Omega}^{(i)} = [\boldsymbol{\omega}_1^{(i)}, \ldots, \boldsymbol{\omega}_n^{(i)}]^{\mathrm{T}}$, where $[\boldsymbol{\omega}_j^{(i)}]^{\mathrm{T}} = \mathbf{e}_j^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{A}}\mathbf{H})\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\mathbf{B}_{nij}^{\mathrm{T}}$, $j = 1, \ldots, n$. Then for the term (15), we invoke Result S6.1 and Lemma 3.3 to obtain that for sufficiently large n, for all $t \ge 1$, $t \lesssim n\rho_n$,

$$
\begin{aligned}
&\left\|\frac{1}{n\rho_n}\sum_{j=1}^n (A_{ij} - \rho_n \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{e}_j^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}}\mathbf{H} - \mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)})\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\mathbf{B}_{nij}^{\mathrm{T}}\right\|_2\\
&\quad= \frac{1}{n\rho_n}\left\|\sum_{j=1}^n [\mathbf{E}]_{ij}\boldsymbol{\omega}_j^{(i)}\right\|_2 = \frac{1}{n\rho_n}\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}\mathbf{\Omega}^{(i)}\|_2\\
&\quad\le \frac{1}{n\rho_n}\|\mathbf{E}\|_{2\to\infty}\|\mathbf{\Omega}^{(i)}\|_{\mathrm{F}} = \frac{1}{n\rho_n}\|\mathbf{E}\|_{2\to\infty}\left(\sum_{j=1}^n \|\boldsymbol{\omega}_j^{(n)}\|_2^2\right)^{1/2}\\
&\quad\le \frac{1}{n\rho_n}\|\mathbf{E}\|_{2\to\infty}\left(\sum_{j=1}^n \|\mathbf{e}_j(\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{A}}\mathbf{H})\|_2^2\|\mathbf{S}_{\mathbf{P}}\|_2\max_{j\in[n]}\|\mathbf{B}_{nij}\|_2^2\right)^{1/2}\\
&\quad\le \frac{1}{n\rho_n}\|\mathbf{E}\|_2\left(d\|\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{A}}\mathbf{H}\|_2^2\|\mathbf{S}_{\mathbf{P}}\|_2\max_{j\in[n]}\|\mathbf{B}_{nij}\|_2^2\right)^{1/2}\\
&\quad\lesssim_c \frac{d^{1/2}}{(n\rho_n)^{1/2}}\{n\rho_n\lambda_1(\mathbf{\Delta}_n)\}^{1/2}\frac{1}{\delta^4}\frac{t^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^2}\\
&\quad\le \frac{\kappa(\mathbf{\Delta}_n)^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}t^{1/2}}{(n\rho_n)^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)^2}
\end{aligned}
$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$. We finally turn our attention to the term (16). Denote $\mathbf{\Theta}^{(i)} = [\boldsymbol{\theta}_1^{(i)}, \ldots, \boldsymbol{\theta}_n^{(i)}]^{\mathrm{T}}$, where $(\boldsymbol{\theta}_j^{(i)})^{\mathrm{T}} = \mathbf{e}_j^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}})\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\mathbf{B}_{nj}^{\mathrm{T}}$, $j = 1, \ldots, n$. Let $t \ge 1$ and $t \lesssim n\rho_n$. We take advantage of the fact that $\mathbf{\Theta}^{(i)}$ and $(A_{ij} - \rho_n \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j)_{j=1}^n$ are independent and consider the following events:

$$
\begin{aligned}
\mathcal{E}_1 &= \left\{ \mathbf{A} : \left\| \sum_{j=1}^n (A_{ij} - \rho_n \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j) \boldsymbol{\theta}_j^{(i)} \right\|_2 \le t \|\mathbf{\Theta}^{(i)}\|_{2\to\infty} + (2t\rho_n)^{1/2} \|\mathbf{\Theta}^{(i)}\|_{\mathrm{F}} \right\}, \\
\mathcal{E}_2 &= \left\{ \mathbf{A} : \|\mathbf{U}_{\mathbf{A}}^{(i)} \mathrm{sgn}(\mathbf{H}^{(i)}) - \mathbf{U}_{\mathbf{P}}\|_{2\to\infty} \le \frac{C_c \|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)}, \|\mathbf{U}_{\mathbf{A}}^{(i)}\|_{2\to\infty} \le \frac{C_c \|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)} \right\}, \\
\mathcal{E}_3 &= \left\{ \mathbf{A} : \|\mathbf{U}_{\mathbf{A}}^{(i)} \mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}}\|_2 \le \frac{C_c}{(n\rho_n)^{1/2} \lambda_d(\mathbf{\Delta}_n)} \right\}.
\end{aligned}
$$

Here, $C_c > 0$ is a constant only depending on c that will be determined later. By Lemma S2.1,

$$
\mathbb{P}(\mathcal{E}_1) = \sum_{\mathbf{A}^{(m)}} \mathbb{P}(\mathcal{E}_1 \mid \mathbf{A}^{(m)}) p(\mathbf{A}^{(m)}) \ge (1 - 28e^{-t}) \sum_{\mathbf{A}^{(m)}} p(\mathbf{A}^{(m)}) = 1 - 28e^{-t}.
$$

By Lemma 3.3, for sufficiently large n, $\mathbb{P}(\mathcal{E}_2) \ge 1 - 6n^{-c}$ and $\mathbb{P}(\mathcal{E}_3) \ge 1 - 3n^{-c}$. Hence, over the event $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$, which has probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$, we have

$$
\begin{aligned}
&\left\| \frac{1}{n\rho_n} \sum_{j=1}^n (A_{ij} - \rho_n \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j) \mathbf{e}_j^{\mathrm{T}} (\mathbf{U}_{\mathbf{A}}^{(i)} \mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}}) \mathbf{S}_{\mathbf{P}}^{1/2} \mathbf{W}_{\mathbf{X}} \mathbf{B}_{nij}^{\mathrm{T}} \right\|_2 \\
&\quad \le \frac{3t}{n\rho_n} \|\mathbf{\Theta}^{(i)}\|_{2\to\infty} + \frac{(6\rho_n t)^{1/2}}{n\rho_n} \left(\sum_{j=1}^n \|\boldsymbol{\theta}_j^{(i)}\|_2^2 \right)^{1/2} \\
&\quad = \frac{3t}{n\rho_n} \max_{j\in[n]} \|\mathbf{e}_j^{\mathrm{T}} (\mathbf{U}_{\mathbf{A}}^{(i)} \mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}}) \mathbf{S}_{\mathbf{P}}^{1/2} \mathbf{W}_{\mathbf{X}} \mathbf{B}_{nij}^{\mathrm{T}}\|_2 \\
&\qquad + \frac{(6\rho_n t)^{1/2}}{n\rho_n} \left(\sum_{j=1}^n \|\mathbf{e}_j^{\mathrm{T}} (\mathbf{U}_{\mathbf{A}}^{(i)} \mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}}) \mathbf{S}_{\mathbf{P}}^{1/2} \mathbf{W}_{\mathbf{X}} \mathbf{B}_{nij}^{\mathrm{T}}\|_2^2 \right)^{1/2} \\
&\quad \le \frac{3t}{n\rho_n} (2\|\mathbf{U}_{\mathbf{A}}^{(i)}\|_{2\to\infty} + \|\mathbf{U}_{\mathbf{A}}^{(i)} \mathrm{sgn}(\mathbf{H}^{(i)}) - \mathbf{U}_{\mathbf{P}}\|_{2\to\infty}) \|\mathbf{S}_{\mathbf{P}}\|_2^{1/2} \max_{j\in[n]} \|\mathbf{B}_{nij}\|_2 \\
&\qquad + \frac{(6\rho_n t)^{1/2}}{n\rho_n} \left(\|\mathbf{U}_{\mathbf{A}}^{(i)} \mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}}\|_{\mathrm{F}}^2 \|\mathbf{S}_{\mathbf{P}}\|_2^{1/2} \max_{j\in[n]} \|\mathbf{B}_{nij}^{\mathrm{T}}\|_2^2 \right)^{1/2} \\
&\quad \lesssim_c \frac{t}{n\rho_n} \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)} \{n\rho_n \lambda_1(\mathbf{\Delta}_n)\}^{1/2} \frac{1}{\delta^4} + \frac{(d\rho_n t)^{1/2}}{n\rho_n} \frac{\{n\rho_n \lambda_1(\mathbf{\Delta}_n)\}^{1/2}}{(n\rho_n)^{1/2} \lambda_d(\mathbf{\Delta}_n)} \frac{1}{\delta^4} \\
&\quad = \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty} t}{(n\rho_n)^{1/2} \delta^4 \lambda_d(\mathbf{\Delta}_n)} + \frac{t^{1/2}}{(n\rho_n)^{1/2} \delta^4 \lambda_d(\mathbf{\Delta}_n)} \sqrt{\frac{d}{n}} \lesssim \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty} t}{(n\rho_n)^{1/2} \delta^4 \lambda_d(\mathbf{\Delta}_n)}.
\end{aligned}
$$

Combining the aforementioned concentration bounds for (13), (14), (15), and (16) completes the proof. □

S7.6. Concentration bound for (12)

We now focus on the concentration bound for term (12) by taking advantage of the auxiliary matrices $\mathbf{U}_{\mathbf{A}}^{(i)}, \mathbf{H}^{(i)}$ defined in Section 3. Observe that term (12) consists of two terms:

$$\frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n [\mathbf{E}]_{ij}\mathbf{r}_{\mathbf{g}} \quad \text{and} \quad \frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n \rho_n \mathbf{r}_{\mathbf{h}_{ij}}.$$

The second term is relatively easy to analyze, whereas the first term is more involved. Recall that we assume

$$\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)(n\rho_n)^{1/2}} \to 0, \quad \frac{1}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^{3/2}}\max\left\{\frac{(\log n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^3}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^3}, \frac{\log n}{\lambda_d(\mathbf{\Delta}_n)}\right\} \to 0.$$

By Corollary 4.1, given any fixed $c > 0$, for sufficiently large n,

$$\begin{aligned}
\|\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W} - \mathbf{X}\|_{2\to\infty} &\lesssim_c \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n^{1/2}\rho_n\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{(\log n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^3}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^3}, \frac{\log n}{\lambda_d(\mathbf{\Delta}_n)}\right\}\\
&\quad + \frac{(\log n)^{1/2}}{\rho_n^{1/2}\lambda_d(\mathbf{\Delta}_n)^{1/2}}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\\
&\leq \frac{1}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^{3/2}}\max\left\{\frac{(\log n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^3}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^3}, \frac{\log n}{\lambda_d(\mathbf{\Delta}_n)}\right\}\\
&\quad + \frac{(\log n)^{1/2}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}
\end{aligned}$$

with probability at least $1 - c_0 n^{-c}$. By assumption,

$$\frac{1}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^{3/2}}\max\left\{\frac{(\log n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^3}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^3}, \frac{\log n}{\lambda_d(\mathbf{\Delta}_n)}\right\} \quad \Longrightarrow \quad \frac{(\log n)^{1/2}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)} \to 0.$$

Therefore, for sufficiently large n,

$$(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i, \rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j) \in \mathcal{X}_2(\delta/2) \quad \text{for all } i, j \in [n]$$

with probability at least $1 - c_0 n^{-c}$. Then we can apply Lemma S7.3 to further obtain

$$\begin{aligned}
&\left\|\frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n [\mathbf{E}]_{ij}\mathbf{r}_{\mathbf{g}}\right\|_2 + \left\|\frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n \rho_n\mathbf{r}_{\mathbf{h}_{ij}}\right\|_2\\
&\quad \lesssim \frac{d^{1/2}}{n\rho_n^{1/2}\delta^6}\|\mathbf{E}\|_\infty\|\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \mathbf{x}_i\|_2^2 + \frac{d^{1/2}}{n\rho_n^{3/2}\delta^6}\sum_{j=1}^n |[\mathbf{E}]_{ij}|\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j\|_2^2\\
&\quad + \frac{d^{1/2}}{n\rho_n^{1/2}\delta^6}\sum_{j=1}^n \|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i\|_2^2 + \frac{d^{1/2}}{n\rho_n^{1/2}\delta^6}\sum_{j=1}^n \|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j\|_2^2
\end{aligned}$$

$$= \frac{d^{1/2}(\|\mathbf{E}\|_\infty + n\rho_n)}{n\rho_n^{3/2}\delta^6}\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i\|_2^2 + \frac{d^{1/2}}{n\rho_n^{3/2}\delta^6}\sum_{j=1}^n |[\mathbf{E}]_{ij}|\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j\|_2^2$$
$$+ \frac{d^{1/2}}{n\rho_n^{1/2}\delta^6}\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}\|_{\mathrm{F}}^2$$

with probability at least $1 - c_0 n^{-c}$ for large n. By Result S6.2, Lemma S7.1, and Lemma S7.2, for sufficiently large n,

$$\|\mathbf{E}\|_\infty \lesssim_c n\rho_n \quad \text{with probability at least } 1 - c_0 n^{-c},$$
$$\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}\|_{\mathrm{F}} \lesssim_c \frac{1}{\lambda_d(\mathbf{\Delta}_n)} \quad \text{with probability at least } 1 - c_0 n^{-c},$$
$$\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i\|_2 \lesssim_c \frac{\|\mathbf{U_P}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, t\right\}$$
$$\text{with probability at least } 1 - c_0 n^{-c} - c_0 e^{-t}.$$

It suffices to provide a concentration bound for

$$\frac{1}{n\rho_n^{3/2}}\sum_{j=1}^n |[\mathbf{E}]_{ij}|\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j\|_2^2.$$

Lemma S7.7. *Let $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ and assume the conditions of Theorem 4.7 hold. Then given any fixed $c > 0$, for each fixed index $i \in [n]$, for all $t \geq 1$, $t \lesssim n\rho_n$, and sufficiently large n,*

$$\frac{1}{n\rho_n^{3/2}}\sum_{j=1}^n |[\mathbf{E}]_{ij}|\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j\|_2^2 \lesssim_c \frac{\|\mathbf{U_P}\|_{2\to\infty}^2}{\rho_n^{1/2}\lambda_d(\mathbf{\Delta}_n)^2}\max\left\{t, \frac{1}{\lambda_d(\mathbf{\Delta}_n)^2}\right\},$$
$$\frac{1}{n\rho_n}\sum_{j=1}^n |[\mathbf{E}]_{ij}|\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j\|_2$$
$$\lesssim_c \frac{1}{\sqrt{n}\lambda_d(\mathbf{\Delta}_n)} + \frac{\|\mathbf{U_P}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\max\left\{t, \frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}\right\}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$ for sufficiently large n, where $c_0 > 0$ is an absolute constant.

Proof of Lemma S7.7. The proof is quite similar to that of Lemma S7.6 modulus some slight modifications. Following the decomposition

$$\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X} = \mathbf{U_A}\{\mathbf{S_A}^{1/2}\mathrm{sgn}(\mathbf{H}) - \mathrm{sgn}(\mathbf{H})\mathbf{S_P}^{1/2}\}\mathbf{W_X} + \mathbf{U_A}\{\mathrm{sgn}(\mathbf{H}) - \mathbf{H}\}\mathbf{S_P}^{1/2}\mathbf{W_X}$$
$$+ (\mathbf{U_A}\mathbf{H} - \mathbf{U_A}^{(i)}\mathbf{H}^{(i)})\mathbf{S_P}^{1/2}\mathbf{W_X} + (\mathbf{U_A}^{(i)}\mathbf{H}^{(i)} - \mathbf{U_P})\mathbf{S_P}^{1/2}\mathbf{W_X},$$

we obtain from the Cauchy-Schwarz inequality that $(a + b + c + d)^2 \leq 4a^2 + 4b^2 + 4c^2 + 4d^2$ and the triangle inequality that

$$\frac{1}{n\rho_n^{3/2}}\sum_{j=1}^n |[\mathbf{E}]_{ij}|\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j\|_2^2$$

$$
\begin{aligned}
&\le \frac{1}{n\rho_n^{3/2}}\sum_{j=1}^{n}(|[\mathbf{E}]_{ij}|-\mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\rho_n^{1/2}\mathbf{x}_j\|_2^2+\frac{2}{n\rho_n^{1/2}}\|\widetilde{\mathbf{X}}\mathbf{W}-\rho_n^{1/2}\mathbf{X}\|_{\mathrm{F}}^2\\
&\le \frac{4}{n\rho_n^{3/2}}\sum_{j=1}^{n}(|[\mathbf{E}]_{ij}|-\mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{e}_j^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\{\mathbf{S}_{\mathbf{A}}^{1/2}\mathrm{sgn}(\mathbf{H})-\mathrm{sgn}(\mathbf{H})\mathbf{S}_{\mathbf{P}}^{1/2}\}\mathbf{W}_{\mathbf{X}}\|_2^2 &(17)\\
&\quad+\frac{4}{n\rho_n^{3/2}}\sum_{j=1}^{n}(|[\mathbf{E}]_{ij}|-\mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{e}_j^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\{\mathrm{sgn}(\mathbf{H})-\mathbf{H}\}\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\|_2^2 &(18)\\
&\quad+\frac{4}{n\rho_n^{3/2}}\sum_{j=1}^{n}(|[\mathbf{E}]_{ij}|-\mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}}\mathbf{H}-\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)})\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\|_2^2 &(19)\\
&\quad+\frac{4}{n\rho_n^{3/2}}\sum_{j=1}^{n}(|[\mathbf{E}]_{ij}|-\mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)}-\mathbf{U}_{\mathbf{P}})\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\|_2^2 &(20)\\
&\quad+\frac{2}{n\rho_n^{1/2}}\|\widetilde{\mathbf{X}}\mathbf{W}-\rho_n^{1/2}\mathbf{X}\|_{\mathrm{F}}^2 &(21)
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{n\rho_n}\sum_{j=1}^{n}|[\mathbf{E}]_{ij}|\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\rho_n^{1/2}\mathbf{x}_j\|_2\\
&\le \frac{1}{n\rho_n}\sum_{j=1}^{n}(|[\mathbf{E}]_{ij}|-\mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\rho_n^{1/2}\mathbf{x}_j\|_2+\frac{2}{n}\sum_{j=1}^{n}\|\mathbf{e}_j^{\mathrm{T}}(\widetilde{\mathbf{X}}\mathbf{W}-\rho_n^{1/2}\mathbf{X})\|_2\\
&\le \frac{1}{n\rho_n}\sum_{j=1}^{n}(|[\mathbf{E}]_{ij}|-\mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{e}_j^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\{\mathbf{S}_{\mathbf{A}}^{1/2}\mathrm{sgn}(\mathbf{H})-\mathrm{sgn}(\mathbf{H})\mathbf{S}_{\mathbf{P}}^{1/2}\}\mathbf{W}_{\mathbf{X}}\|_2 &(22)\\
&\quad+\frac{1}{n\rho_n}\sum_{j=1}^{n}(|[\mathbf{E}]_{ij}|-\mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{e}_j^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\{\mathrm{sgn}(\mathbf{H})-\mathbf{H}\}\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\|_2 &(23)\\
&\quad+\frac{1}{n\rho_n}\sum_{j=1}^{n}(|[\mathbf{E}]_{ij}|-\mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}}\mathbf{H}-\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)})\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\|_2 &(24)\\
&\quad+\frac{1}{n\rho_n}\sum_{j=1}^{n}(|[\mathbf{E}]_{ij}|-\mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)}-\mathbf{U}_{\mathbf{P}})\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\|_2 &(25)\\
&\quad+\frac{2}{\sqrt{n}}\|\widetilde{\mathbf{X}}\mathbf{W}-\rho_n^{1/2}\mathbf{X}\|_{\mathrm{F}}. &(26)
\end{aligned}
$$

For terms (21) and (26), we know from Lemma S7.1 that

$$
\frac{1}{n\rho_n^{1/2}}\|\widetilde{\mathbf{X}}\mathbf{W}-\rho_n^{1/2}\mathbf{X}\|_{\mathrm{F}}^2\lesssim_c\frac{1}{n\rho_n^{1/2}\lambda_d(\mathbf{\Delta}_n)^2},\quad \frac{1}{\sqrt{n}}\|\widetilde{\mathbf{X}}\mathbf{W}-\rho_n^{1/2}\mathbf{X}\|_{\mathrm{F}}\lesssim_c\frac{1}{\sqrt{n}\lambda_d(\mathbf{\Delta}_n)}
$$

with probability at least $1 - c_0 n^{-c}$ for sufficiently large n. For terms (17) and (22), by Result S6.2, Lemma S2.3, and Lemma S2.5, for all $t \geq 1$, $t \lesssim n\rho_n$,

$$
\begin{aligned}
&\frac{4}{n\rho_n^{3/2}} \sum_{j=1}^{n} (|[\mathbf{E}]_{ij}| - \mathbb{E}|[\mathbf{E}]_{ij}|) \| \mathbf{e}_j^{\mathrm{T}} \mathbf{U}_{\mathbf{A}} \{ \mathbf{S}_{\mathbf{A}}^{1/2} \mathrm{sgn}(\mathbf{H}) - \mathrm{sgn}(\mathbf{H}) \mathbf{S}_{\mathbf{P}}^{1/2} \} \mathbf{W}_{\mathbf{X}} \|_2^2 \\
&\quad \leq \frac{4}{n\rho_n^{3/2}} \sum_{j=1}^{n} |[\mathbf{E}]_{ij}| \|\mathbf{U}_{\mathbf{A}}\|_{2\to\infty}^2 \|\mathbf{W}^* \mathbf{S}_{\mathbf{A}}^{1/2} - \mathbf{S}_{\mathbf{P}}^{1/2} \mathbf{W}^*\|_2^2 \\
&\quad = \frac{4}{n\rho_n^{3/2}} \|\mathbf{E}\|_\infty \|\mathbf{U}_{\mathbf{A}}\|_{2\to\infty}^2 \|\mathbf{W}^* \mathbf{S}_{\mathbf{A}}^{1/2} - \mathbf{S}_{\mathbf{P}}^{1/2} \mathbf{W}^*\|_2^2 \\
&\quad \lesssim_c \frac{1}{n\rho_n^{3/2}} n\rho_n \left\{ \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\boldsymbol{\Delta}_n)} \right\}^2 \left[\frac{1}{(n\rho_n)^{1/2} \lambda_d(\boldsymbol{\Delta}_n)} \max\left\{ \frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)}, t^{1/2} \right\} \right]^2 \\
&\quad = \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}^2}{n\rho_n^{3/2} \lambda_d(\boldsymbol{\Delta}_n)^4} \max\left\{ \frac{\kappa(\boldsymbol{\Delta}_n)^2}{\lambda_d(\boldsymbol{\Delta}_n)^2}, t \right\} = \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}^2}{\rho_n^{1/2} \lambda_d(\boldsymbol{\Delta}_n)^2} \max\left\{ \frac{\kappa(\boldsymbol{\Delta}_n)^2}{n\rho_n \lambda_d(\boldsymbol{\Delta}_n)^4}, \frac{t}{n\rho_n \lambda_d(\boldsymbol{\Delta}_n)^2} \right\} \\
&\quad \lesssim \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}^2}{\rho_n^{1/2} \lambda_d(\boldsymbol{\Delta}_n)^2} \max\left\{ \frac{1}{\lambda_d(\boldsymbol{\Delta}_n)^2}, t \right\}
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{n\rho_n} \sum_{j=1}^{n} (|[\mathbf{E}]_{ij}| - \mathbb{E}|[\mathbf{E}]_{ij}|) \| \mathbf{e}_j^{\mathrm{T}} \mathbf{U}_{\mathbf{A}} \{ \mathbf{S}_{\mathbf{A}}^{1/2} \mathrm{sgn}(\mathbf{H}) - \mathrm{sgn}(\mathbf{H}) \mathbf{S}_{\mathbf{P}}^{1/2} \} \mathbf{W}_{\mathbf{X}} \|_2 \\
&\quad \leq \frac{1}{n\rho_n} \sum_{j=1}^{n} |[\mathbf{E}]_{ij}| \|\mathbf{U}_{\mathbf{A}}\|_{2\to\infty} \|\mathbf{W}^* \mathbf{S}_{\mathbf{A}}^{1/2} - \mathbf{S}_{\mathbf{P}}^{1/2} \mathbf{W}^*\|_2 \\
&\quad = \frac{1}{n\rho_n} \|\mathbf{E}\|_\infty \|\mathbf{U}_{\mathbf{A}}\|_{2\to\infty} \|\mathbf{W}^* \mathbf{S}_{\mathbf{A}}^{1/2} - \mathbf{S}_{\mathbf{P}}^{1/2} \mathbf{W}^*\|_2 \\
&\quad \lesssim_c \frac{1}{n\rho_n} n\rho_n \left\{ \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\boldsymbol{\Delta}_n)} \right\} \frac{1}{(n\rho_n)^{1/2} \lambda_d(\boldsymbol{\Delta}_n)} \max\left\{ \frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)}, t^{1/2} \right\} \\
&\quad = \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2} \lambda_d(\boldsymbol{\Delta}_n)} \max\left\{ \frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \frac{t^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)} \right\}
\end{aligned}
$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$. For terms (18) and (23), by Lemma S2.5, Lemma 6.7 in [4], and the Davis-Kahan theorem,

$$
\begin{aligned}
&\frac{4}{n\rho_n^{3/2}} \sum_{j=1}^{n} (|[\mathbf{E}]_{ij}| - \mathbb{E}|[\mathbf{E}]_{ij}|) \| \mathbf{e}_j^{\mathrm{T}} \mathbf{U}_{\mathbf{A}} \{ \mathrm{sgn}(\mathbf{H}) - \mathbf{H} \} \mathbf{S}_{\mathbf{P}}^{1/2} \mathbf{W}_{\mathbf{X}} \|_2^2 \\
&\quad \leq \frac{4}{n\rho_n^{3/2}} \sum_{j=1}^{n} |[\mathbf{E}]_{ij}| \|\mathbf{U}_{\mathbf{A}}\|_{2\to\infty}^2 \|\mathbf{W}^* - \mathbf{U}_{\mathbf{P}}^{\mathrm{T}} \mathbf{U}_{\mathbf{A}}\|_2^2 \|\mathbf{S}_{\mathbf{P}}\|_2 \\
&\quad \leq \frac{4}{n\rho_n^{3/2}} \|\mathbf{E}\|_\infty \|\mathbf{U}_{\mathbf{A}}\|_{2\to\infty}^2 \|\sin\Theta(\mathbf{U}_{\mathbf{A}}, \mathbf{U}_{\mathbf{P}})\|_2^4 \|\mathbf{S}_{\mathbf{P}}\|_2
\end{aligned}
$$

$$
\begin{aligned}
&\le \frac{4}{n\rho_n^{3/2}}\|\mathbf{E}\|_\infty\|\mathbf{U_A}\|_{2\to\infty}^2\frac{\|\mathbf{E}\|_2^4}{\lambda_d(\mathbf{P})^4}\|\mathbf{S_P}\|_2\\
&\lesssim_c \frac{n\rho_n}{n\rho_n^{3/2}}\left\{\frac{\|\mathbf{U_P}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)}\right\}^2\frac{(n\rho_n)^2}{(n\rho_n)^4\lambda_d(\mathbf{\Delta}_n)^4}(n\rho_n)\lambda_1(\mathbf{\Delta}_n) = \frac{\kappa(\mathbf{\Delta}_n)\|\mathbf{U_P}\|_{2\to\infty}^2}{n\rho_n^{3/2}\lambda_d(\mathbf{\Delta}_n)^5}\\
&= \frac{\|\mathbf{U_P}\|_{2\to\infty}^2}{\rho_n^{1/2}\lambda_d(\mathbf{\Delta}_n)^2}\left\{\frac{\kappa(\mathbf{\Delta}_n)}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^3}\right\} \le \frac{\|\mathbf{U_P}\|_{2\to\infty}^2}{\rho_n^{1/2}\lambda_d(\mathbf{\Delta}_n)^2}\left\{\frac{1}{\lambda_d(\mathbf{\Delta}_n)}\right\}.
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{n\rho_n}\sum_{j=1}^n(|[\mathbf{E}]_{ij}| - \mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{e}_j^\mathrm{T}\mathbf{U_A}\{\mathrm{sgn}(\mathbf{H}) - \mathbf{H}\}\mathbf{S_P}^{1/2}\mathbf{W_X}\|_2\\
&\quad\le \frac{1}{n\rho_n}\sum_{j=1}^n|[\mathbf{E}]_{ij}|\|\mathbf{U_A}\|_{2\to\infty}\|\mathbf{W}^* - \mathbf{U_P^\mathrm{T}U_A}\|_2\|\mathbf{S_P}\|_2^{1/2}\\
&\quad\le \frac{1}{n\rho_n}\|\mathbf{E}\|_\infty\|\mathbf{U_A}\|_{2\to\infty}\|\sin\Theta(\mathbf{U_A},\mathbf{U_P})\|_2^2\|\mathbf{S_P}\|_2^{1/2}\\
&\quad\le \frac{1}{n\rho_n}\|\mathbf{E}\|_\infty\|\mathbf{U_A}\|_{2\to\infty}\frac{\|\mathbf{E}\|_2^2}{\lambda_d(\mathbf{P})^2}\|\mathbf{S_P}\|_2^{1/2}\\
&\quad\lesssim_c \frac{n\rho_n}{n\rho_n}\left\{\frac{\|\mathbf{U_P}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)}\right\}\frac{(n\rho_n)}{(n\rho_n)^2\lambda_d(\mathbf{\Delta}_n)^2}(n\rho_n)^{1/2}\lambda_1(\mathbf{\Delta}_n)^{1/2}\\
&\quad= \frac{\|\mathbf{U_P}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\left\{\frac{1}{\lambda_d(\mathbf{\Delta}_n)^2}\right\}.
\end{aligned}
$$

with probability at least $1 - c_0n^{-c}$ for sufficiently large n. For terms (19) and (24), we invoke Lemma 3.3 to obtain that

$$
\begin{aligned}
&\frac{4}{n\rho_n^{3/2}}\sum_{j=1}^n(|[\mathbf{E}]_{ij}| - \mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{e}_j^\mathrm{T}(\mathbf{U_A H} - \mathbf{U_A^{(\mathit{i})}H^{(\mathit{i})}})\mathbf{S_P}^{1/2}\mathbf{W_X}\|_2^2\\
&\lesssim_c \frac{1}{n\rho_n^{3/2}}\sum_{j=1}^n\|\mathbf{e}_j^\mathrm{T}(\mathbf{U_A H} - \mathbf{U_A^{(\mathit{i})}H^{(\mathit{i})}})\|_2^2\|\mathbf{S_P}\|_2 = \frac{d}{n\rho_n^{3/2}}\|\mathbf{U_A H} - \mathbf{U_A^{(\mathit{i})}H^{(\mathit{i})}}\|_2^2\|\mathbf{S_P}\|_2\\
&\lesssim \frac{d}{n\rho_n^{3/2}}\frac{t\|\mathbf{U_P}\|_{2\to\infty}^2}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^4}n\rho_n\lambda_1(\mathbf{\Delta}_n) \le \frac{\kappa(\mathbf{\Delta}_n)\|\mathbf{U_P}\|_{2\to\infty}^2 t}{n\rho_n^{3/2}\lambda_d(\mathbf{\Delta}_n)^4}\\
&= \frac{\|\mathbf{U_P}\|_{2\to\infty}^2}{\rho_n^{1/2}\lambda_d(\mathbf{\Delta}_n)^2}\left\{\frac{\kappa(\mathbf{\Delta}_n)t}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^2}\right\} \lesssim \frac{\|\mathbf{U_P}\|_{2\to\infty}^2 t}{\rho_n^{1/2}\lambda_d(\mathbf{\Delta}_n)^2}
\end{aligned}
$$

and

$$
\frac{1}{n\rho_n}\sum_{j=1}^n(|[\mathbf{E}]_{ij}| - \mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{e}_j^\mathrm{T}(\mathbf{U_A H} - \mathbf{U_A^{(\mathit{i})}H^{(\mathit{i})}})\mathbf{S_P}^{1/2}\mathbf{W_X}\|_2
$$

$$
\begin{aligned}
&\lesssim_c \frac{1}{n\rho_n}\sum_{j=1}^n \|\mathbf{e}_j^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}}\mathbf{H} - \mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)})\|_2\|\mathbf{S}_{\mathbf{P}}\|_2^{1/2} = \frac{d^{1/2}}{n\rho_n}\|\mathbf{U}_{\mathbf{A}}\mathbf{H} - \mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)}\|_2\|\mathbf{S}_{\mathbf{P}}\|_2^{1/2}\\
&\lesssim \frac{d^{1/2}}{n\rho_n}\frac{t^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\boldsymbol{\Delta}_n)^2}(n\rho_n)^{1/2}\lambda_1(\boldsymbol{\Delta}_n)^{1/2} \leq \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\kappa(\boldsymbol{\Delta}_n)^{1/2}t^{1/2}}{n\rho_n\lambda_d(\boldsymbol{\Delta}_n)^2}\\
&= \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\boldsymbol{\Delta}_n)^2}\left\{\frac{\kappa(\boldsymbol{\Delta}_n)^{1/2}t^{1/2}}{(n\rho_n)^{1/2}}\right\} \lesssim \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\boldsymbol{\Delta}_n)^2}\kappa(\boldsymbol{\Delta}_n)^{1/2}
\end{aligned}
$$

with probability at least $1 - c_0n^{-c} - c_0e^{-t}$ for sufficiently large n. Finally, for terms (20) and (25), we denote $\boldsymbol{\Theta}^{(i)} = [\boldsymbol{\theta}_1^{(i)}, \ldots, \boldsymbol{\theta}_n^{(i)}]^{\mathrm{T}}$, where $(\boldsymbol{\theta}_j^{(i)})^{\mathrm{T}} = \mathbf{e}_j^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}})\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}$, $j = 1, \ldots, n$ and consider the following events that are similar to those in the proof of Lemma S7.6:

$$
\begin{aligned}
\mathcal{E}_1 &= \left\{\mathbf{A} : \sum_{j=1}^n(|[\mathbf{E}]_{ij}| - \mathbb{E}|[\mathbf{E}]_{ij}|)\|\boldsymbol{\theta}_j^{(i)}\|_2^2 \leq t\max_{j\in[n]}\|\boldsymbol{\theta}_j^{(i)}\|_2^2 + (2\rho_n t)^{1/2}\left(\sum_{j=1}^n\|\boldsymbol{\theta}_j^{(i)}\|_2^4\right)^{1/2}\right\},\\
\mathcal{E}_1' &= \left\{\mathbf{A} : \sum_{j=1}^n(|[\mathbf{E}]_{ij}| - \mathbb{E}|[\mathbf{E}]_{ij}|)\|\boldsymbol{\theta}_j^{(i)}\|_2 \leq t\max_{j\in[n]}\|\boldsymbol{\theta}_j^{(i)}\|_2 + (2\rho_n t)^{1/2}\|\boldsymbol{\Theta}_j^{(i)}\|_{\mathrm{F}}\right\},\\
\mathcal{E}_2 &= \left\{\mathbf{A} : \|\mathbf{U}_{\mathbf{A}}^{(i)}\mathrm{sgn}(\mathbf{H}^{(i)}) - \mathbf{U}_{\mathbf{P}}\|_{2\to\infty} \leq \frac{C_c\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\boldsymbol{\Delta}_n)}, \|\mathbf{U}_{\mathbf{A}}^{(i)}\|_{2\to\infty} \leq \frac{C_c\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\boldsymbol{\Delta}_n)}\right\},\\
\mathcal{E}_3 &= \left\{\mathbf{A} : \|\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}}\|_2 \leq \frac{C_c}{(n\rho_n)^{1/2}\lambda_d(\boldsymbol{\Delta}_n)}\right\},
\end{aligned}
$$

where C_c is a constant only depending on c and will be selcted later. By Lemma S2.1 and the independence between $\boldsymbol{\Theta}^{(i)}$ and $([\mathbf{E}]_{ij})_{j=1}^n$, we have $\mathbb{P}(\mathcal{E}_1) \geq 1 - 28e^{-t}$ and $\mathbb{P}(\mathcal{E}_1') \geq 1 - 28e^{-t}$. By Lemma 3.3, for sufficiently large n, $\mathbb{P}(\mathcal{E}_2) \geq 1 - 6n^{-c}$ and $\mathbb{P}(\mathcal{E}_3) \geq 1 - 3n^{-c}$. Hence, over the event $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$, we have

$$
\begin{aligned}
&\frac{4}{n\rho_n^{3/2}}\sum_{j=1}^n(|[\mathbf{E}]_{ij}| - \mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}})\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}\|_2^2\\
&\quad\leq \frac{4t}{n\rho_n^{3/2}}\max_{j\in[n]}\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}})\|_2^2\|\mathbf{S}_{\mathbf{P}}\|_2\\
&\quad\quad + \frac{4(2\rho_n t)^{1/2}}{n\rho_n^{3/2}}\left\{\sum_{j=1}^n\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}})\|_2^4\|\mathbf{S}_{\mathbf{P}}\|_2^2\right\}^{1/2}\\
&\quad\leq \frac{4t}{n\rho_n^{3/2}}\|\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}}\|_{2\to\infty}^2\|\mathbf{S}_{\mathbf{P}}\|_2\\
&\quad\quad + \frac{4(2\rho_n t)^{1/2}}{n\rho_n^{3/2}}\|\mathbf{S}_{\mathbf{P}}\|_2\|\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}}\|_{\mathrm{F}}\\
&\quad\leq \frac{4t}{n\rho_n^{3/2}}\|\mathbf{U}_{\mathbf{A}}^{(i)}\mathbf{H}^{(i)} - \mathbf{U}_{\mathbf{P}}\|_{2\to\infty}^2\|\mathbf{S}_{\mathbf{P}}\|_2
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \frac{4(2d\rho_n t)^{1/2}}{n\rho_n^{3/2}}\|\mathbf{S_P}\|_2\|\mathbf{U_A}^{(i)}\mathbf{H}^{(i)} - \mathbf{U_P}\|_{2\to\infty}\|\mathbf{U_A}^{(i)}\mathbf{H}^{(i)} - \mathbf{U_P}\|_2\\
&\le \frac{4t}{n\rho_n^{3/2}}(2\|\mathbf{U_A}^{(i)}\|_{2\to\infty} + \|\mathbf{U_A}^{(i)}\mathrm{sgn}(\mathbf{H}^{(i)}) - \mathbf{U_P}\|_{2\to\infty})^2\|\mathbf{S_P}\|_2\\
&\quad + \frac{4(2d\rho_n t)^{1/2}}{n\rho_n^{3/2}}\|\mathbf{S_P}\|_2\|\mathbf{U_A}^{(i)}\mathbf{H}^{(i)} - \mathbf{U_P}\|_{2\to\infty}\|\mathbf{U_A}^{(i)}\mathbf{H}^{(i)} - \mathbf{U_P}\|_2\\
&\lesssim_c \frac{t}{n\rho_n^{3/2}}\frac{\|\mathbf{U_P}\|_{2\to\infty}^2}{\lambda_d(\mathbf{\Delta}_n)^2}n\rho_n\lambda_1(\mathbf{\Delta}_n) + \frac{(dt)^{1/2}n\rho_n\lambda_1(\mathbf{\Delta}_n)}{n\rho_n}\frac{\|\mathbf{U_P}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)}\frac{1}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\\
&= \frac{t\|\mathbf{U_P}\|_{2\to\infty}^2}{\rho_n^{1/2}\lambda_d(\mathbf{\Delta}_n)^2} + \frac{t^{1/2}\|\mathbf{U_P}\|_{2\to\infty}}{\rho_n^{1/2}\lambda_d(\mathbf{\Delta}_n)^2}\sqrt{\frac{d}{n}} \lesssim \frac{t\|\mathbf{U_P}\|_{2\to\infty}^2}{\rho_n^{1/2}\lambda_d(\mathbf{\Delta}_n)^2}.
\end{aligned}
$$

Similarly, over the event $\mathcal{E}_1' \cap \mathcal{E}_2 \cap \mathcal{E}_3$, we have

$$
\begin{aligned}
&\frac{1}{n\rho_n}\sum_{j=1}^n(|[\mathbf{E}]_{ij}| - \mathbb{E}|[\mathbf{E}]_{ij}|)\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{U_A}^{(i)}\mathbf{H}^{(i)} - \mathbf{U_P})\mathbf{S_P}^{1/2}\mathbf{W_X}\|_2\\
&\le \frac{t}{n\rho_n}\max_{j\in[n]}\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{U_A}^{(i)}\mathbf{H}^{(i)} - \mathbf{U_P})\|_2\|\mathbf{S_P}\|_2^{1/2}\\
&\quad + \frac{(2\rho_n t)^{1/2}}{n\rho_n}\left\{\sum_{j=1}^n\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{U_A}^{(i)}\mathbf{H}^{(i)} - \mathbf{U_P})\|_2^2\|\mathbf{S_P}\|_2\right\}^{1/2}\\
&\le \frac{t}{n\rho_n}\|\mathbf{U_A}^{(i)}\mathbf{H}^{(i)} - \mathbf{U_P}\|_{2\to\infty}\|\mathbf{S_P}\|_2^{1/2} + \frac{(2\rho_n t)^{1/2}}{n\rho_n}\|\mathbf{S_P}\|_2^{1/2}\|\mathbf{U_A}^{(i)}\mathbf{H}^{(i)} - \mathbf{U_P}\|_{\mathrm{F}}\\
&\le \frac{t}{n\rho_n}\|\mathbf{U_A}^{(i)}\mathbf{H}^{(i)} - \mathbf{U_P}\|_{2\to\infty}\|\mathbf{S_P}\|_2^{1/2} + \frac{(2d\rho_n t)^{1/2}}{n\rho_n}\|\mathbf{S_P}\|_2^{1/2}\|\mathbf{U_A}^{(i)}\mathbf{H}^{(i)} - \mathbf{U_P}\|_2\\
&\le \frac{t}{n\rho_n}(2\|\mathbf{U_A}^{(i)}\|_{2\to\infty} + \|\mathbf{U_A}^{(i)}\mathrm{sgn}(\mathbf{H}^{(i)}) - \mathbf{U_P}\|_{2\to\infty})\|\mathbf{S_P}\|_2^{1/2}\\
&\quad + \frac{(2d\rho_n t)^{1/2}}{n\rho_n}\|\mathbf{S_P}\|_2^{1/2}\|\mathbf{U_A}^{(i)}\mathbf{H}^{(i)} - \mathbf{U_P}\|_2\\
&\lesssim_c \frac{t}{n\rho_n}\frac{\|\mathbf{U_P}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)}\{n\rho_n\lambda_1(\mathbf{\Delta}_n)\}^{1/2} + \frac{(d\rho_n t)^{1/2}(n\rho_n)^{1/2}\lambda_1(\mathbf{\Delta}_n)^{1/2}}{n\rho_n}\frac{1}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\\
&= \frac{t\|\mathbf{U_P}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)} + \frac{t^{1/2}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\sqrt{\frac{d}{n}} \lesssim \frac{t\|\mathbf{U_P}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}.
\end{aligned}
$$

The events $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ and $\mathcal{E}_1' \cap \mathcal{E}_2 \cap \mathcal{E}_3$ both occur with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$. The proof is completed by combining the concentration bounds above. □

We now combine the aforementioned analysis to obtain the concentration bound for (12).

Lemma S7.8. *Let* $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ *and assume the conditions of Theorem 4.7 hold. Then given any fixed* $c > 0$*, for each fixed index* $i \in [n]$*, for all* $t \geq 1$*,* $t \lesssim n\rho_n$*, and sufficiently large n,*

$$\left\|\frac{1}{n\rho_n^{1/2}}\sum_{j=1}^{n}[\mathbf{E}]_{ij}\mathbf{r}_{\mathbf{g}} + \frac{1}{n\rho_n^{1/2}}\rho_n\mathbf{r}_{\mathbf{h}_{ij}}\right\|_2 \lesssim_c \frac{d^{1/2}\|\mathbf{U_P}\|_{2\to\infty}^2}{\rho_n^{1/2}\delta^6\lambda_d(\boldsymbol{\Delta}_n)^2}\max\left\{\frac{t}{\lambda_d(\boldsymbol{\Delta}_n)^4}, \frac{\kappa(\boldsymbol{\Delta}_n)^2}{\lambda_d(\boldsymbol{\Delta}_n)^4}, t^2\right\}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$ *for sufficiently large n, where* $c_0 > 0$ *is an absolute constant.*

S7.7. Concentration bound for (6)

We are now in a position to obtain a concentration bound for term (6) by collecting the results in Sections S7.3, S7.4, S7.5, and S7.6.

Lemma S7.9. *Let* $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ *and assume the conditions of Theorem 4.7 hold. Let* $\boldsymbol{\phi}_{ij}(\mathbf{u},\mathbf{v})$*,* $\mathbf{g}$*,* $\mathbf{h}_{ij}$ *be defined as in Section S7.1 and* $\mathbf{r}_{i1}$ *be defined as in (6). Then given any fixed* $c > 0$*, for each fixed row index* $i \in [n]$*, for all* $t \geq 1$*,* $t \lesssim \log n$*, and sufficiently large n,*

$$\|\mathbf{r}_{i1}\|_2 \lesssim_c \frac{\|\mathbf{U_P}\|_{2\to\infty}}{\delta^4\lambda_d(\boldsymbol{\Delta}_n)}\max\left\{\frac{t^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)^2}, t\right\},$$

$$\begin{aligned}
&\left\|\mathbf{r}_{i1} + \frac{1}{n\rho_n^{1/2}}\sum_{j=1}^{n}\frac{\rho_n\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \mathbf{x}_i)}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\|_2\\
&\quad\lesssim_c \frac{d^{1/2}\|\mathbf{U_P}\|_{2\to\infty}}{(n\rho_n)^{1/2}\delta^6\lambda_d(\boldsymbol{\Delta}_n)^{5/2}}\max\left\{\frac{t}{\lambda_d(\boldsymbol{\Delta}_n)^4}, \frac{\kappa(\boldsymbol{\Delta}_n)^2}{\lambda_d(\boldsymbol{\Delta}_n)^4}, t^2\right\}
\end{aligned}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$ *for sufficiently large n, where* $c_0 > 0$ *is an absolute constants.*

Proof of Lemma S7.9. For convenience, we denote $\widetilde{\mathbf{y}}_i = \rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i$, $i \in [n]$. We first show that for sufficiently large n, $(\widetilde{\mathbf{y}}_i, \widetilde{\mathbf{y}}_j) \in \mathcal{X}_2(\delta/2)$ for all $(i,j) \in [n]\times[n]$ with large probability. By Corollary 4.1, for all $c > 0$,

$$\begin{aligned}
\max_{i\in[n]}\|\widetilde{\mathbf{y}}_i - \mathbf{x}_i\|_2 &= \rho_n^{-1/2}\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}\|_{2\to\infty}\\
&\lesssim_c \frac{1}{n\rho_n\lambda_d(\boldsymbol{\Delta}_n)^{3/2}}\max\left\{\frac{(\log n)^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)^3}, \frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)^3}, \frac{\log n}{\lambda_d(\boldsymbol{\Delta}_n)}\right\} + \frac{(\log n)^{1/2}}{(n\rho_n)^{1/2}\lambda_d(\boldsymbol{\Delta}_n)}.
\end{aligned}$$

with probability at least $1 - c_0 n^{-c}$ for sufficiently large n. The upper bound on the preceeding display converges to 0 as $n \to \infty$ by our assumption. Therefore, for sufficiently large n, with probability at least $1 - c_0 n^{-c}$,

$$\max_{i\in[n]}\|\widetilde{\mathbf{y}}_i - \mathbf{x}_i\|_2 \leq \frac{\delta}{6}.$$

Then for any $i, j \in [n]$, with probability at least $1 - c_0 n^{-c}$,

$$\begin{aligned}
\widetilde{\mathbf{y}}_i^{\mathrm{T}}\widetilde{\mathbf{y}}_j &= (\widetilde{\mathbf{y}}_i - \mathbf{x}_i + \mathbf{x}_i)^{\mathrm{T}}(\widetilde{\mathbf{y}}_j - \mathbf{x}_j + \mathbf{x}_j) \leq \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j + \|\widetilde{\mathbf{y}}_i - \mathbf{x}_i\|_2^2 + 2\|\widetilde{\mathbf{y}}_i - \mathbf{x}_i\|_2\\
&\leq \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j + 3\|\widetilde{\mathbf{y}}_i - \mathbf{x}_i\|_2 \leq 1 - \delta + 3(\delta/6) = 1 - \delta/2,
\end{aligned}$$

$$\widetilde{\mathbf{y}}_i^{\mathrm{T}}\widetilde{\mathbf{y}}_j \geq \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j - 3\|\widetilde{\mathbf{y}}_i - \mathbf{x}_i\|_2 \geq \delta - 3(\delta/6) = \delta/2.$$

This further implies that $\|\widetilde{\mathbf{y}}_i\|_2 \leq 1$ for all $i \in [n]$, and hence, $(\widetilde{\mathbf{y}}_i, \widetilde{\mathbf{y}}_j) \in \mathcal{X}_2(\delta/2)$ for all $(i,j) \in [n] \times [n]$ with probability at least $1 - c_0 n^{-c}$ for sufficiently large n.

We now proceed to the first assertion. Invoking assertion (a) of Lemma S7.3, we have

$$\left\|\boldsymbol{\phi}_{ij}(\widetilde{\mathbf{y}}_i, \widetilde{\mathbf{y}}_j) - \boldsymbol{\phi}_{ij}(\mathbf{x}_i, \mathbf{x}_j)\right\|_2 \lesssim (|[\mathbf{E}]_{ij}| + \rho_n)\|\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W} - \mathbf{X}\|_{2\to\infty}$$

with probability at least $1 - c_0 n^{-c}$ for sufficiently large n. It follows from Result S6.2, Lemma S7.2, and Lemma S7.7 that, for any $t \geq 1$ and $t \lesssim \log n$,

$$\begin{aligned}
&\frac{1}{n\rho_n^{1/2}}\left\|\sum_{j=1}^n \left\{\boldsymbol{\phi}_{ij}(\widetilde{\mathbf{y}}_i, \widetilde{\mathbf{y}}_j) - \boldsymbol{\phi}_{ij}(\mathbf{x}_i, \mathbf{x}_j)\right\}\right\|_2\\
&\quad\lesssim \frac{1}{\delta^4}(\|\mathbf{E}\|_\infty + n\rho_n)\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i\|_2 + \frac{1}{n\rho_n\delta^4}\sum_{j=1}^n |[\mathbf{E}]_{ij}|\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j\|_2\\
&\quad\quad + \frac{1}{n\delta^4}\sum_{j=1}^n \|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_j\|_2\\
&\quad\lesssim_c \frac{n\rho_n}{n\rho_n\delta^4}\frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\boldsymbol{\Delta}_n)}\max\left\{\frac{t^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)^2}, t\right\} + \frac{1}{\sqrt{n}\delta^4\lambda_d(\boldsymbol{\Delta}_n)}\\
&\quad\quad + \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\boldsymbol{\Delta}_n)}\max\left\{t, \frac{t^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)}, \frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)^2}\right\} + \frac{1}{\sqrt{n}\delta^4\lambda_d(\boldsymbol{\Delta}_n)}\\
&\quad\lesssim_c \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\delta^4\lambda_d(\boldsymbol{\Delta}_n)}\max\left\{\frac{t^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)^2}, t\right\} + \frac{1}{\sqrt{n}\delta^4\lambda_d(\boldsymbol{\Delta}_n)}\\
&\quad\quad + \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\boldsymbol{\Delta}_n)}\max\left\{\frac{t}{(n\rho_n)^{1/2}}, \frac{t^{1/2}}{(n\rho_n)^{1/2}\lambda_d(\boldsymbol{\Delta}_n)}, \frac{\kappa(\boldsymbol{\Delta}_n)}{(n\rho_n)^{1/2}\lambda_d(\boldsymbol{\Delta}_n)^2}\right\}\\
&\quad\lesssim \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\delta^4\lambda_d(\boldsymbol{\Delta}_n)}\max\left\{\frac{t^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)^2}, t\right\}
\end{aligned}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$ for sufficiently large n. This completes the proof of the first assertion. For the second assertion, we recall that

$$\left\|\mathbf{r}_{i1} + \frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n \frac{\rho_n\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1 - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \mathbf{x}_i)\right\|_2$$

can be decomposed into the four terms (9), (10), (11), and (12). More specifically,

$$\begin{aligned}
&\mathbf{r}_{i1} + \frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n \frac{\rho_n\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1 - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \mathbf{x}_i)\\
&\quad = -\frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n \frac{\rho_n\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1 - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \mathbf{x}_j)
\end{aligned}$$

$$- \frac{1}{n\rho_n^{1/2}} \sum_{j=1}^n \left[\frac{[\mathbf{E}]_{ij}(1-2\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}}{\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\}^2} \right] (\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \mathbf{x}_i)$$
$$+ \frac{1}{n\rho_n^{1/2}} \sum_{j=1}^n \frac{[\mathbf{E}]_{ij}\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{I}_d - (1-2\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}\}}{\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\}^2} (\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \mathbf{x}_j)$$
$$+ \frac{1}{n\rho_n^{1/2}} \sum_{j=1}^n [\mathbf{E}]_{ij}\mathbf{r}_{\mathbf{g}} + \frac{1}{n\rho_n^{1/2}} \sum_{j=1}^n \rho_n \mathbf{r}_{\mathbf{h}_{ij}}.$$

By Lemma S7.4, for all $t \geq 1$, $t \lesssim n\rho_n$, for sufficiently large n, with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$,

$$\left\| \frac{1}{n\rho_n^{1/2}} \sum_{j=1}^n \frac{\rho_n \mathbf{x}_j \mathbf{x}_i^{\mathrm{T}} (\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \mathbf{x}_j)}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)} \right\|_2 \lesssim_c \frac{1}{n\rho_n^{1/2}\delta^2\lambda_d(\mathbf{\Delta}_n)} \max\left\{ \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)}, t^{1/2} \right\}.$$

We next apply Lemma S7.5 with

$$\mathbf{B}_{nij} = \frac{(1-2\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}}{\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\}^2}$$

to obtain that for sufficiently large n, for all $t \geq 1$, $t \lesssim n\rho_n$, with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$,

$$\left\| \frac{1}{n\rho_n^{1/2}} \sum_{j=1}^n [\mathbf{E}]_{ij} \frac{(1-2\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \mathbf{x}_j)}{\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\}^2} \right\|_2$$
$$\lesssim_c \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)} \max\left\{ \frac{t}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, t^{3/2} \right\}.$$

In addition, by Lemma S7.6, with

$$\mathbf{B}_{nij} = \frac{\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{I}_d - (1-2\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}\}}{\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\}^2},$$

for all $t \geq 1$, $t \lesssim n\rho_n$,

$$\left\| \frac{1}{n\rho_n^{1/2}} \sum_{j=1}^n \frac{[\mathbf{E}]_{ij}\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{I}_d - (1-2\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j\mathbf{x}_i^{\mathrm{T}}\}}{\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\}^2} (\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \mathbf{x}_j) \right\|_2$$
$$\lesssim_c \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)} \max\left\{ \frac{\kappa(\mathbf{\Delta}_n)^{1/2}t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, t \right\}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$ for sufficiently large n. Finally, by Lemma S7.8, for all $t \geq 1$, $t \lesssim n\rho_n$,

$$\left\| \frac{1}{n\rho_n^{1/2}} \sum_{j=1}^n [\mathbf{E}]_{ij}\mathbf{r}_{\mathbf{g}} + \frac{1}{n\rho_n^{1/2}} \rho_n \mathbf{r}_{\mathbf{h}_{ij}} \right\|_2 \lesssim_c \frac{d^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\delta^6\lambda_d(\mathbf{\Delta}_n)^{5/2}} \max\left\{ \frac{t}{\lambda_d(\mathbf{\Delta}_n)^4}, \frac{\kappa(\mathbf{\Delta}_n)^2}{\lambda_d(\mathbf{\Delta}_n)^4}, t^2 \right\}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$ for sufficiently large n. The proof is then completed by combining the aforementioned concentration bounds. □

S7.8. Concentration bound for (7)

In this section, we work on a concentration bound for term (7). Since $\widetilde{\mathbf{X}}\mathbf{W}$ is close to $\rho_n^{1/2}\mathbf{X}$ in the stringent two-to-infinity norm distance by Theorem 3.2, it is expected that term (7) is asymptotically negligible by the continuous mapping theorem. A formal description of this result requires some work. To begin with, we first oberve the following fact that guarantees that $\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W}$ is close to $\mathbf{X}$ in the two-to-infinity norm distance.

Result S7.1. By Corollary 4.1, given any fixed $c > 0$,

$$\|\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W} - \mathbf{X}\|_{2\to\infty} \lesssim_c \frac{1}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^{3/2}} \max\left\{\frac{(\log n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^3}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^3}, \frac{\log n}{\lambda_d(\mathbf{\Delta}_n)}\right\} + \frac{(\log n)^{1/2}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}$$

with probability at least $1 - c_0 n^{-c}$ for sufficiently large n. By assumption,

$$\frac{1}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^{3/2}} \max\left\{\frac{(\log n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^3}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^3}, \frac{\log n}{\lambda_d(\mathbf{\Delta}_n)}\right\} \quad\Longrightarrow\quad \frac{(\log n)^{1/2}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)} \to 0.$$

Therefore, for sufficiently large n,

$$(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i, \rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j) \in \mathcal{X}_2(\delta/2) \quad \text{for all } i, j \in [n]$$

with probability at least $1 - c_0 n^{-c}$.

Lemma S7.10. *Let* $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ *and assume the conditions of Theorem 4.7 hold. Then given any fixed* $c > 0$*, for each fixed row index* $i \in [n]$*, for all* $t \geq 1$*,* $t \lesssim n\rho_n$*, and sufficiently large* n*,*

$$\left\|\mathbf{H}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W}) - \mathbf{H}_i(\mathbf{X})\right\|_2 \lesssim_c \frac{1}{(n\rho_n)^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)^{3/2}} \max\left\{\frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, t\right\}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$ *for sufficiently large* n*.*

Proof of Lemma S7.10. For any $\mathbf{u}, \mathbf{v}$ with $\|\mathbf{u}\|_2, \|\mathbf{v}\|_2 \leq 1$ and $\delta/2 \leq \mathbf{u}^{\mathrm{T}}\mathbf{v} \leq 1 - \delta/2$, define

$$\mathbf{H}(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1 - \rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})}.$$

By the matrix differential calculus (see, e.g., [9]), we can compute

$$\frac{\partial}{\partial\mathbf{u}^{\mathrm{T}}}\mathrm{vec}\{\mathbf{H}(\mathbf{u}, \mathbf{v})\} = \frac{(1 - 2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})(\mathbf{v}\otimes\mathbf{v})\mathbf{u}^{\mathrm{T}}}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1 - \rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^2},$$

$$\frac{\partial}{\partial\mathbf{v}^{\mathrm{T}}}\mathrm{vec}\{\mathbf{H}(\mathbf{u}, \mathbf{v})\} = \frac{(1 - 2\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})(\mathbf{v}\otimes\mathbf{v})\mathbf{v}^{\mathrm{T}}}{\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1 - \rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^2} + \frac{\mathbf{I}_d\otimes\mathbf{v} + \mathbf{v}\otimes\mathbf{I}_d}{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1 - \rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})}.$$

Clearly,

$$\sup_{(\mathbf{u},\mathbf{v})\in\mathcal{X}_2(\delta/2)}\left\{\left\|\frac{\partial}{\partial\mathbf{u}^{\mathrm{T}}}\mathrm{vec}\{\mathbf{H}(\mathbf{u},\mathbf{v})\}\right\|_2+\left\|\frac{\partial}{\partial\mathbf{u}^{\mathrm{T}}}\mathrm{vec}\{\mathbf{H}(\mathbf{u},\mathbf{v})\}\right\|_2\right\}\leq\frac{40}{\delta^4}.$$

Then by the mean-value inequality, for any $\mathbf{u},\mathbf{v}$ with $\|\mathbf{u}\|_2,\|\mathbf{v}\|_2\leq 1$, $\delta/2\leq\mathbf{u}^{\mathrm{T}}\mathbf{v}\leq 1-\delta/2$,

$$\|\mathbf{H}(\mathbf{u},\mathbf{v})-\mathbf{H}(\mathbf{x}_i,\mathbf{x}_j)\|_2\leq\frac{40}{\delta^4}(\|\mathbf{u}-\mathbf{x}_i\|_2+\|\mathbf{v}-\mathbf{x}_j\|_2)$$

Denote $\widetilde{\mathbf{y}}_i=\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i$, $i\in[n]$. We then apply Result S7.1, Lemma S7.1, and Lemma S7.2 to obtain that for all $t\geq 1$, $t\lesssim n\rho_n$,

$$\begin{aligned}
&\|\mathbf{H}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})-\mathbf{H}_i(\mathbf{X})\|_2\\
&\quad\leq\frac{1}{n}\sum_{j=1}^n\|\mathbf{H}(\widetilde{\mathbf{y}}_i,\widetilde{\mathbf{y}}_j)-\mathbf{H}(\mathbf{x}_i,\mathbf{x}_j)\|_2\lesssim\frac{1}{n\delta^4}\sum_{j=1}^n(\|\widetilde{\mathbf{y}}_i-\mathbf{x}_i\|_2+\|\widetilde{\mathbf{y}}_j-\mathbf{x}_j\|_2)\\
&\quad\leq\frac{1}{\rho_n^{1/2}\delta^4}\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_i\|_2+\frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\rho_n^{1/2}\mathbf{x}_j\|_2\\
&\quad\leq\frac{1}{\rho_n^{1/2}\delta^4}\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_i\|_2+\frac{1}{(n\rho_n)^{1/2}\delta^4}\|\widetilde{\mathbf{X}}\mathbf{W}-\rho_n^{1/2}\mathbf{X}\|_{\mathrm{F}}\\
&\quad\lesssim_c\frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\rho_n^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2},\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2},t\right\}+\frac{1}{(n\rho_n)^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)}\\
&\quad\lesssim\frac{1}{(n\rho_n)^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)^{3/2}}\max\left\{\frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2},\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2},t\right\}
\end{aligned}$$

with probability at least $1-c_0n^{-c}-c_0e^{-t}$ for sufficiently large n. The proof is thus completed. □

S7.9. Proofs of Theorems 4.7 and 4.6

Proof of Theorem 4.7. We first recall the following decomposition of $\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_i$ in Section S7.1:

$$\mathbf{G}_n(\mathbf{x}_i)^{1/2}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_i)=\frac{1}{n\rho_n^{1/2}}\sum_{j=1}^n\frac{[\mathbf{E}]_{ij}\mathbf{G}_n(\mathbf{x}_i)^{-1/2}\mathbf{x}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}+\widehat{\mathbf{r}}_i,$$

where

$$\begin{aligned}
\widehat{\mathbf{r}}_i&=\mathbf{G}_n(\mathbf{x}_i)^{-1/2}\{\mathbf{G}_n(\mathbf{x}_i)(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_i)+\mathbf{r}_{i1}\}+\mathbf{G}_n(\mathbf{x}_i)^{1/2}\mathbf{R}_{i2}\mathbf{r}_{i1}+\mathbf{G}_n(\mathbf{x}_i)^{1/2}\mathbf{r}_{i3},\\
\mathbf{r}_{i1}&=\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\left\{\frac{(A_{ij}-\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j)(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j)}{\rho_n^{-1}\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j(1-\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j)}-\frac{(A_{ij}-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\},\\
\mathbf{R}_{i2}&=\mathbf{H}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})^{-1}-\mathbf{G}_n(\mathbf{x}_i)^{-1},
\end{aligned}$$

$\mathbf{H}_i(\cdot)$ is the function defined in (5), and

$$\mathbf{r}_{i3} = \frac{1}{n\sqrt{\rho_n}} \sum_{j=1}^{n} \frac{(A_{ij} - \rho_n \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j)}{\mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j (1 - \rho_n \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j)} \mathbf{R}_{i2} \mathbf{x}_j.$$

By Lemma S7.9, for all $t \geq 1$, $t \lesssim \log n$,

$$\begin{aligned}
\|\mathbf{r}_{i1}\|_2 &\lesssim_c \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\delta^4 \lambda_d(\boldsymbol{\Delta}_n)} \max\left\{\frac{t^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)^2}, t\right\} \\
&\leq \frac{1}{\sqrt{n}\delta^4 \lambda_d(\boldsymbol{\Delta}_n)^{3/2}} \max\left\{\frac{t^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)^2}, t\right\}
\end{aligned}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$ for sufficiently large n, and

$$\begin{aligned}
\left\|\mathbf{r}_{i1} + \mathbf{G}_n(\mathbf{x}_i)(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i)\right\|_2 &\lesssim_c \frac{d^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\delta^6 \lambda_d(\boldsymbol{\Delta}_n)^{5/2}} \max\left\{\frac{t}{\lambda_d(\boldsymbol{\Delta}_n)^4}, \frac{\kappa(\boldsymbol{\Delta}_n)^2}{\lambda_d(\boldsymbol{\Delta}_n)^4}, t^2\right\} \\
&\leq \frac{d^{1/2}}{n\rho_n^{1/2}\delta^6 \lambda_d(\boldsymbol{\Delta}_n)^3} \max\left\{\frac{t}{\lambda_d(\boldsymbol{\Delta}_n)^4}, \frac{\kappa(\boldsymbol{\Delta}_n)^2}{\lambda_d(\boldsymbol{\Delta}_n)^4}, t^2\right\}
\end{aligned}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$. We next focus on $\mathbf{R}_{i2}$. Observe that

$$\mathbf{G}_n(\mathbf{x}_i) = \frac{1}{n}\sum_{j=1}^{n} \frac{\mathbf{x}_j \mathbf{x}_j^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1 - \rho_n \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)} = \mathbf{H}_i(\mathbf{X}) \succeq \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j^{\mathrm{T}} = \boldsymbol{\Delta}_n.$$

By definition of $\mathbf{H}_i(\cdot)$ and Result S7.1,

$$\begin{aligned}
\mathbf{H}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W}) &= \mathbf{W}^{\mathrm{T}}\left\{\frac{1}{n}\sum_{j=1}^{n}\frac{\widetilde{\mathbf{x}}_j\widetilde{\mathbf{x}}_j^{\mathrm{T}}}{\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j(1 - \widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j)}\right\}\mathbf{W} \\
&= \mathbf{W}^{\mathrm{T}}\left[\frac{1}{n}\sum_{j=1}^{n}\frac{\widetilde{\mathbf{x}}_j\widetilde{\mathbf{x}}_j^{\mathrm{T}}}{\rho_n(\rho_n^{-1/2}\widetilde{\mathbf{x}}_i^{\mathrm{T}}\rho_n^{-1/2}\widetilde{\mathbf{x}}_j)(1 - \rho_n\rho_n^{-1/2}\widetilde{\mathbf{x}}_i^{\mathrm{T}}\rho_n^{-1/2}\widetilde{\mathbf{x}}_j)}\right]\mathbf{W} \\
&\gtrsim \frac{1}{\rho_n}\mathbf{W}^{\mathrm{T}}\left(\frac{1}{n}\sum_{j=1}^{n}\widetilde{\mathbf{x}}_j\widetilde{\mathbf{x}}_j^{\mathrm{T}}\right)\mathbf{W} = \frac{1}{n\rho_n}\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{T}})\mathbf{W}
\end{aligned}$$

with probability at least $1 - c_0 n^{-c}$ for sufficiently large n. Namely,

$$\lambda_d\left\{\mathbf{H}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})\right\} \gtrsim \frac{1}{n\rho_n}\lambda_d(\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{X}}) = \frac{1}{n\rho_n}\lambda_d(\mathbf{A}),$$

and hence, by Result S2.2,

$$\|\mathbf{H}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})^{-1}\|_2 \lesssim n\rho_n\|\mathbf{S}_{\mathbf{A}}^{-1}\|_2 \lesssim \frac{1}{\lambda_d(\boldsymbol{\Delta}_n)}$$

with probability at least $1-c_0n^{-c}$ for sufficiently large n. Also, by Lemma S7.10, for all $t\geq 1$, $t\lesssim n\rho_n$,

$$\|\mathbf{H}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})-\mathbf{H}_i(\mathbf{X})\|_2\lesssim_c \frac{1}{(n\rho_n)^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)^{3/2}}\max\left\{\frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2},\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2},t\right\}$$

with probability at least $1-c_0n^{-c}-c_0e^{-t}$ for sufficiently large n. It follows that

$$\begin{aligned}
\|\mathbf{G}_n(\mathbf{x}_i)^{1/2}\mathbf{R}_{i2}\|_2&=\|\mathbf{G}_n(\mathbf{x}_i)^{-1/2}\{\mathbf{H}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})-\mathbf{G}_n(\mathbf{x}_i)\}\mathbf{H}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})^{-1}\|_2\\
&\leq\|\mathbf{G}_n(\mathbf{x}_i)^{-1/2}\|_2\|\mathbf{H}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})-\mathbf{G}_n(\mathbf{x}_i)\|_2\|\mathbf{H}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})^{-1}\|_2\\
&\lesssim\frac{1}{\lambda_d(\mathbf{\Delta}_n)^{3/2}}\|\mathbf{H}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})-\mathbf{G}_n(\mathbf{x}_i)\|_2\\
&\lesssim_c\frac{1}{(n\rho_n)^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)^{3}}\max\left\{\frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2},\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2},t\right\}
\end{aligned}$$

with probability at least $1-c_0n^{-c}-c_0e^{-t}$ for sufficiently large n. We then move forward to the analysis of $\mathbf{r}_{i3}$. Write

$$\|\mathbf{G}_n(\mathbf{x}_i)^{1/2}\mathbf{r}_{i3}\|_2\leq\|\mathbf{G}_n(\mathbf{x}_i)^{1/2}\mathbf{R}_{i2}\|_2\frac{1}{n\sqrt{\rho_n}}\left\|\sum_{j=1}^n\frac{(A_{ij}-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\|_2.$$

By Lemma S2.1, for all $t\geq 1$ and $t\lesssim n\rho_n$, we have

$$\begin{aligned}
\left\|\sum_{j=1}^n\frac{(A_{ij}-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\|_2&\leq 3t\max_{j\in[n]}\left\|\frac{\mathbf{x}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\|_2\\
&\quad+(6\rho_nt)^{1/2}\left\{\sum_{j=1}^n\left\|\frac{\mathbf{x}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\|_2^2\right\}^{1/2}\\
&\lesssim(n\rho_nt)^{1/2}\max_{j\in[n]}\left\|\frac{\mathbf{x}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\|_2\lesssim\frac{(n\rho_nt)^{1/2}}{\delta^2}
\end{aligned}$$

with probability at least $1-c_0e^{-t}$. It follows immediately that

$$\begin{aligned}
\|\mathbf{G}_n(\mathbf{x}_i)^{1/2}\mathbf{r}_{i3}\|_2&\lesssim_c\frac{1}{(n\rho_n)^{1/2}\delta^4\lambda_d(\mathbf{\Delta}_n)^{3}}\max\left\{\frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2},\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2},t\right\}\frac{(n\rho_nt)^{1/2}}{n\rho_n^{1/2}\delta^2}\\
&=\frac{1}{n\rho_n^{1/2}\delta^6\lambda_d(\mathbf{\Delta}_n)^{3}}\max\left\{\frac{t}{\lambda_d(\mathbf{\Delta}_n)^2},\frac{\kappa(\mathbf{\Delta}_n)t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2},t^{3/2}\right\}
\end{aligned}$$

with probability at least $1-c_0n^{-c}-c_0e^{-t}$ for all sufficiently large n.

Summarizing the above large probability bounds, for all $t\geq 1$, $t\lesssim\log n$,

$$\|\widehat{\mathbf{r}}_i\|_2\leq\|\mathbf{G}_n(\mathbf{x}_i)^{-1/2}\|_2\|\mathbf{G}_n(\mathbf{x}_i)(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_i)+\mathbf{r}_{i1}\|_2$$

$$
\begin{aligned}
&+ \|\mathbf{G}_n(\mathbf{x}_i)^{1/2}\mathbf{R}_{i2}\|_2\|\mathbf{r}_{i1}\|_2 + \|\mathbf{G}_n(\mathbf{x}_i)^{1/2}\mathbf{r}_{i3}\|_2\\
&\lesssim_c \frac{d^{1/2}}{n\rho_n^{1/2}\delta^6\lambda_d(\boldsymbol{\Delta}_n)^{7/2}}\max\left\{\frac{t}{\lambda_d(\boldsymbol{\Delta}_n)^4}, \frac{\kappa(\boldsymbol{\Delta}_n)^2}{\lambda_d(\boldsymbol{\Delta}_n)^4}, t^2\right\}\\
&\quad + \frac{1}{(n\rho_n)^{1/2}\delta^4\lambda_d(\boldsymbol{\Delta}_n)^3}\max\left\{\frac{t^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)^2}, t\right\}\\
&\quad \times \frac{1}{\sqrt{n}\delta^4\lambda_d(\boldsymbol{\Delta}_n)^{3/2}}\max\left\{\frac{t^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)^2}, t\right\}\\
&\quad + \frac{1}{n\rho_n^{1/2}\delta^6\lambda_d(\boldsymbol{\Delta}_n)^3}\max\left\{\frac{t}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \frac{\kappa(\boldsymbol{\Delta}_n)t^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)^2}, t^{3/2}\right\}\\
&\le \frac{d^{1/2}}{n\rho_n^{1/2}\delta^6\lambda_d(\boldsymbol{\Delta}_n)^{7/2}}\max\left\{\frac{t}{\lambda_d(\boldsymbol{\Delta}_n)^4}, \frac{\kappa(\boldsymbol{\Delta}_n)^2}{\lambda_d(\boldsymbol{\Delta}_n)^4}, t^2\right\}\\
&\quad + \frac{1}{n\rho_n^{1/2}\delta^8\lambda_d(\boldsymbol{\Delta}_n)^{9/2}}\max\left\{\frac{t}{\lambda_d(\boldsymbol{\Delta}_n)^4}, \frac{\kappa(\boldsymbol{\Delta}_n)^2}{\lambda_d(\boldsymbol{\Delta}_n)^4}, t^2\right\}\\
&\quad + \frac{1}{n\rho_n^{1/2}\delta^6\lambda_d(\boldsymbol{\Delta}_n)^3}\max\left\{\frac{t}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \frac{\kappa(\boldsymbol{\Delta}_n)t^{1/2}}{\lambda_d(\boldsymbol{\Delta}_n)^2}, t^{3/2}\right\}\\
&\lesssim \frac{1}{n\rho_n^{1/2}\delta^8\lambda_d(\boldsymbol{\Delta}_n)^{9/2}}\max\left\{\frac{t}{\lambda_d(\boldsymbol{\Delta}_n)^4}, \frac{\kappa(\boldsymbol{\Delta}_n)^2}{\lambda_d(\boldsymbol{\Delta}_n)^4}, t^2\right\}
\end{aligned}
$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$ for sufficiently large n. The proof is thus completed. □

Proof of Theorem 4.6. We apply Theorem S1.4 to obtain the desired Berry-Esseen bound. Let

$$
\begin{aligned}
&\boldsymbol{\xi}_j = \frac{(A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{G}_n(\mathbf{x}_i)^{-1/2}\mathbf{x}_j}{\sqrt{n\rho_n}\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}, \quad \mathbf{D} = \sqrt{n}\widehat{\mathbf{r}}_i,\\
&\Delta^{(j)} = \Delta = \frac{C}{n\rho_n^{1/2}\delta^8\lambda_d(\boldsymbol{\Delta}_n)^{9/2}}\max\left\{\frac{\log n\rho_n}{\lambda_d(\boldsymbol{\Delta}_n)^4}, \frac{\kappa(\boldsymbol{\Delta}_n)^2}{\lambda_d(\boldsymbol{\Delta}_n)^4}, (\log n\rho_n)^2\right\},\\
&\mathcal{O} = \{\mathbf{A} : \Delta > \|\mathbf{D}\|_2\},
\end{aligned}
$$

where $C > 0$ is some absolute constant. By definition of $\boldsymbol{\xi}_j$ and $\mathbf{G}_n(\mathbf{x}_i)$, $\mathbb{E}(\boldsymbol{\xi}_j) = 0$ and $\boldsymbol{\Sigma}_n(\mathbf{x}_i)$,

$$
\sum_{j=1}^n \mathbb{E}_0(\boldsymbol{\xi}_j\boldsymbol{\xi}_j^{\mathrm{T}}) = \frac{1}{n}\sum_{j=1}^n \frac{\mathbf{G}_n(\mathbf{x}_i)^{-1/2}\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\mathbf{G}_n(\mathbf{x}_i)^{-1/2}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)} = \mathbf{I}_d.
$$

We now proceed to $\sum_{j=1}^n \mathbb{E}(\|\boldsymbol{\xi}_j\|_2^3)$ and $\mathbb{E}(\|\sum_{j=1}^n \boldsymbol{\xi}_j\|_2)$. For the first term, we have

$$
\sum_{j=1}^n \mathbb{E}(\|\boldsymbol{\xi}_j\|_2^3) = \frac{1}{(n\rho_n)^{3/2}}\sum_{j=1}^n \mathbb{E}|A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j|^3\left\|\frac{\mathbf{G}_n(\mathbf{x}_i)^{-1/2}\mathbf{x}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\|_2^3
$$

$$
\begin{aligned}
&\leq \frac{\|\mathbf{G}_n(\mathbf{x}_i)^{-1/2}\|_2}{(n\rho_n)^{3/2}\delta^2}\sum_{j=1}^n \mathbb{E}\{(A_{ij}-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)^2\}\left\|\frac{\mathbf{G}_n(\mathbf{x}_i)^{-1/2}\mathbf{x}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\|_2^2\\
&\leq \frac{1}{(n\rho_n)^{3/2}\delta^2\lambda_d(\boldsymbol{\Delta}_n)^{1/2}}\sum_{j=1}^n \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathrm{tr}\left[\frac{\mathbf{G}_n(\mathbf{x}_i)^{-1/2}\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\mathbf{G}_n(\mathbf{x}_i)^{-1/2}}{\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\}^2}\right]\\
&= \frac{1}{(n\rho_n)^{1/2}\delta^2\lambda_d(\boldsymbol{\Delta}_n)^{1/2}}\mathrm{tr}\left[\mathbf{G}_n(\mathbf{x}_i)^{-1/2}\left\{\frac{1}{n}\sum_{j=1}^n\frac{\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\right\}\mathbf{G}_n(\mathbf{x}_i)^{-1/2}\right]\\
&= \frac{1}{(n\rho_n)^{1/2}\delta^2\lambda_d(\boldsymbol{\Delta}_n)^{1/2}}\mathrm{tr}(\mathbf{I}_d) = \frac{d}{(n\rho_n)^{1/2}\delta^2\lambda_d(\boldsymbol{\Delta}_n)^{1/2}}.
\end{aligned}
$$

For $\mathbb{E}(\|\sum_{j=1}^n \boldsymbol{\xi}_j\|_2)$, we use Jensen's inequality to write

$$
\mathbb{E}\left(\left\|\sum_{j=1}^n\boldsymbol{\xi}_j\right\|_2\right)\leq\left\{\mathbb{E}\left(\left\|\sum_{j=1}^n\boldsymbol{\xi}_j\right\|_2^2\right)\right\}^{1/2}=\left(\sum_{j=1}^n\mathbb{E}\|\boldsymbol{\xi}_j\|_2^2\right)^{1/2}=\left\{\mathrm{tr}\left(\sum_{j=1}^n\mathbb{E}\boldsymbol{\xi}_j\boldsymbol{\xi}_j^{\mathrm{T}}\right)\right\}^{1/2}=d^{1/2}.
$$

This immediately implies that

$$
\mathbb{E}\left(\left\|\sum_{j=1}^n\boldsymbol{\xi}_j\right\|_2\Delta\right)\lesssim\frac{d^{1/2}}{n\rho_n^{1/2}\delta^8\lambda_d(\boldsymbol{\Delta}_n)^{9/2}}\max\left\{\frac{\log n\rho_n}{\lambda_d(\boldsymbol{\Delta}_n)^4},\frac{\kappa(\boldsymbol{\Delta}_n)^2}{\lambda_d(\boldsymbol{\Delta}_n)^4},(\log n\rho_n)^2\right\}
$$

Finally, for $\mathbb{P}(O^c)$, the concentration bound in Theorem 4.7 implies that $\mathbb{P}(O^c)\lesssim(n\rho_n)^{-1/2}$ for sufficiently large n. We hence conclude from Theorem S1.4 that

$$
\begin{aligned}
&\sup_{A\in\mathcal{A}}\left|\mathbb{P}\left\{\sqrt{n}\mathbf{G}_n(\mathbf{x}_i)^{-1/2}(\mathbf{W}_n^{\mathrm{T}}\widehat{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_i)\in A\right\}-\mathbb{P}(\mathbf{z}\in A)\right|\\
&\lesssim d^{1/2}\gamma+\mathbb{E}\left(\left\|\sum_{j=1}^n\boldsymbol{\xi}_j\right\|_2\Delta\right)+\mathbb{P}(O^c)\\
&\lesssim\frac{d^{1/2}}{n\rho_n^{1/2}\delta^8\lambda_d(\boldsymbol{\Delta}_n)^{9/2}}\max\left\{\frac{\log n\rho_n}{\lambda_d(\boldsymbol{\Delta}_n)^4},\frac{\kappa(\boldsymbol{\Delta}_n)^2}{\lambda_d(\boldsymbol{\Delta}_n)^4},(\log n\rho_n)^2\right\}.
\end{aligned}
$$

The proof is thus completed. □

S8. Proofs for Section 4.4

S8.1. Proof of Theorem 4.9

We follow the notations and definitions in Sections 2 and 3.

■ **Row-wise concentration bound for the membership profile matrix estimate.** First note that $\sigma_1(\boldsymbol{\Theta})\leq\sqrt{n}$ because

$$
\sigma_1(\boldsymbol{\Theta})\leq\|\boldsymbol{\Theta}\|_{\mathrm{F}}\leq\sqrt{n}\|\boldsymbol{\Theta}\|_{2\to\infty}\leq\sqrt{n}\|\boldsymbol{\Theta}\|_\infty=\sqrt{n}.
$$

Also, we have $\mathbf{\Theta}^{\mathrm{T}}\mathbf{\Theta} \le c_1^2 n\mathbf{I}_d$ by the condition of Theorem 4.9. Therefore,

$$\lambda_d(\mathbf{\Delta}_n) = \frac{1}{n}\lambda_d\{(\mathbf{X}^*)^{\mathrm{T}}\mathbf{\Theta}^{\mathrm{T}}\mathbf{\Theta}(\mathbf{X}^*)\} \ge c_1^2\lambda_d\{(\mathbf{X}^*)^{\mathrm{T}}(\mathbf{X}^*)\} \ge c_1^4.$$

Namely, $\lambda_d(\mathbf{\Delta}_n)$ is bounded away from 0. By Corollary 4.1, there exists constants $K_1, c_1 > 0$, such that

$$\|\mathbf{U_A} - \mathbf{U_P}\mathbf{W}^*\|_{2\to\infty} \le \frac{K_1(\log n)^{1/2}}{n\rho_n^{1/2}} \quad \text{with probability at least } 1 - c_1 n^{-2}.$$

By Lemma S2.5, Result S6.1, and Davis-Kahan theorem,

$$\begin{aligned}
\|\mathbf{U_A}\mathbf{U_A}^{\mathrm{T}} - \mathbf{U_P}\mathbf{U_P}^{\mathrm{T}}\|_{2\to\infty} &\le \|\mathbf{U_A} - \mathbf{U_P}\mathbf{W}^*\|_{2\to\infty}\|\mathbf{U_A}^{\mathrm{T}}\|_2 + \|\mathbf{U_P}\|_{2\to\infty}\|\mathbf{U_A} - \mathbf{U_P}\mathbf{W}^*\|_2 \\
&\le \|\mathbf{U_A} - \mathbf{U_P}\mathbf{W}^*\|_{2\to\infty}\|\mathbf{U_A}^{\mathrm{T}}\|_2 + \|\mathbf{U_P}\|_{2\to\infty}\frac{4\|\mathbf{E}\|_2}{n\rho_n\lambda_d(\mathbf{\Delta}_n)} \\
&\le \frac{K_2(\log n)^{1/2}}{n\rho_n^{1/2}} \quad \text{with probability at least } 1 - c_0 n^{-2}
\end{aligned}$$

for constants $K_2, c_0 > 0$ for sufficiently large n. By Lemma 2.1 in [10], $\mathbf{U_P} = \mathbf{\Theta}\mathbf{V_P}$, where $\mathbf{V_P} \in \mathbb{R}^{d\times d}$ is the submatrix of $\mathbf{U_P}$ corresponding to the pure node indices $\{i_1, \ldots, i_d\}$. By Lemma II.3 in [10], we have $\sigma_1(\mathbf{V_P}) \le (c_1^2 n)^{-1/2}$ and $\sigma_d(\mathbf{V_P}) \ge n^{-1/2}$. Since

$$\begin{aligned}
&\mathbb{P}\left\{\|\mathbf{U_A}\mathbf{U_A}^{\mathrm{T}} - \mathbf{U_P}\mathbf{U_P}^{\mathrm{T}}\|_{2\to\infty} \le \frac{K_2(\log n)^{1/2}}{n\rho_n^{1/2}}\right\} \ge 1 - c_0 n^{-2}, \\
&\sigma_d(\mathbf{U_P}\mathbf{V_P}^{\mathrm{T}}) = \sigma_d(\mathbf{V_P}) \ge n^{-1/2}, \quad \|(\mathbf{U_P}\mathbf{V_P})^{\mathrm{T}}\|_{2\to\infty} = \|\mathbf{V_P}^{\mathrm{T}}\|_{2\to\infty} \le \sigma_1(\mathbf{V_P}) \le \frac{1}{\sqrt{n}}, \\
&\frac{K_2(\log n)^{1/2}}{n\rho_n^{1/2}} = o(n^{-1/2}) \le \frac{1}{\sqrt{n}}\min\left(\frac{1}{2\sqrt{d-1}}, \frac{1}{4}\right)\left(1 + 80\frac{n^{-1}}{n^{-1}}\right)^{-1} \\
&\qquad\qquad \le \sigma_d(\mathbf{U_P}\mathbf{V_P}^{\mathrm{T}})\min\left(\frac{1}{2\sqrt{d-1}}, \frac{1}{4}\right)\left(1 + 80\frac{\|(\mathbf{U_P}\mathbf{V_P})^{\mathrm{T}}\|_{2\to\infty}^2}{\sigma_d(\mathbf{U_P}\mathbf{V_P}^{\mathrm{T}})^2}\right)^{-1},
\end{aligned}$$

then by Theorem 3 in [6], there exists a permutation matrix $\mathbf{\Pi}_n \in \{0,1\}^{d\times d}$ such that

$$\max_{k\in[d]}\|(\mathbf{U_A}\mathbf{V_A}^{\mathrm{T}} - \mathbf{U_P}\mathbf{V_P}^{\mathrm{T}}\mathbf{\Pi}_n)\mathbf{e}_k\|_2 \le \frac{K_3(\log n)^{1/2}}{n\rho_n^{1/2}} \quad \text{with probability at least } 1 - c_0 n^{-2}$$

for constants $K_3, c_0 > 0$ for sufficiently large n. It follows that

$$\begin{aligned}
\|\mathbf{V_A} - \mathbf{\Pi}_n^{\mathrm{T}}\mathbf{V_P}\mathbf{U_P}^{\mathrm{T}}\mathbf{U_A}\|_{2\to\infty} &= \max_{k\in[d]}\|\mathbf{e}_k^{\mathrm{T}}(\mathbf{V_A}\mathbf{U_A}^{\mathrm{T}} - \mathbf{\Pi}_n^{\mathrm{T}}\mathbf{V_P}\mathbf{U_P}^{\mathrm{T}})\mathbf{U_A}\|_2 \\
&\le \max_{k\in[d]}\|(\mathbf{U_A}\mathbf{V_A}^{\mathrm{T}} - \mathbf{U_P}\mathbf{V_P}^{\mathrm{T}}\mathbf{\Pi}_n)\mathbf{e}_k\|_2 \\
&\le \frac{K_3(\log n)^{1/2}}{n\rho_n^{1/2}} \quad \text{with probability at least } 1 - c_0 n^{-2}
\end{aligned}$$

for sufficiently large n. By Lemma 6.7 in [4], Result S2.1, and Davis-Kahan theorem, we further have

$$
\begin{aligned}
\|\mathbf{V}_\mathbf{A} - \mathbf{\Pi}_n^\mathrm{T}\mathbf{V}_\mathbf{P}\mathbf{W}^*\|_2 &\leq \|\mathbf{V}_\mathbf{A} - \mathbf{\Pi}_n^\mathrm{T}\mathbf{V}_\mathbf{P}\mathbf{U}_\mathbf{P}^\mathrm{T}\mathbf{U}_\mathbf{A}\|_2 + \|\mathbf{\Pi}_n^\mathrm{T}\mathbf{V}_\mathbf{P}(\mathbf{U}_\mathbf{P}^\mathrm{T}\mathbf{U}_\mathbf{A} - \mathbf{W}^*)\|_2\\
&\leq \|\mathbf{V}_\mathbf{A} - \mathbf{\Pi}_n^\mathrm{T}\mathbf{V}_\mathbf{P}\mathbf{U}_\mathbf{P}^\mathrm{T}\mathbf{U}_\mathbf{A}\|_2 + \|\mathbf{V}_\mathbf{P}\|_2\|\sin\Theta(\mathbf{U}_\mathbf{A}, \mathbf{U}_\mathbf{P})\|_2^2\\
&\leq \sqrt{d}\|\mathbf{V}_\mathbf{A} - \mathbf{\Pi}_n^\mathrm{T}\mathbf{V}_\mathbf{P}\mathbf{U}_\mathbf{P}^\mathrm{T}\mathbf{U}_\mathbf{A}\|_{2\to\infty} + \frac{4\|\mathbf{E}\|_2^2}{c_1\sqrt{n}(n\rho_n)^2\lambda_d(\mathbf{\Delta}_n)^2}\\
&\leq \frac{K_4(\log n)^{1/2}}{n\rho_n^{1/2}} \quad \text{with probability at least } 1 - c_0 n^{-2}
\end{aligned}
$$

for sufficiently large n, where $K_4, c_0 > 0$ are constants. By Weyl's inequality, we have $\sigma_d(\mathbf{V}_\mathbf{A}) \geq (1/2)\sigma_d(\mathbf{V}_\mathbf{P}) \geq (1/2)n^{-1/2}$ with probability at least $1 - c_0 n^{-2}$ for sufficiently large n since $\log n = o(n\rho_n)$. Namely,

$$
\|\mathbf{V}_\mathbf{A}^{-1}\|_2 \leq 2\sqrt{n} \quad \text{with probability at least } 1 - c_0 n^{-2}
$$

for sufficiently large n. Hence, we have

$$
\begin{aligned}
\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\mathbf{\Pi}_n\|_{2\to\infty} &= \max_{i\in[n]}\|\mathbf{e}_i^\mathrm{T}(\mathbf{U}_\mathbf{A}\mathbf{V}_\mathbf{A}^{-1} - \mathbf{U}_\mathbf{P}\mathbf{V}_\mathbf{P}^{-1}\mathbf{\Pi}_n)\|_2\\
&\leq \max_{i\in[n]}\|\mathbf{e}_i^\mathrm{T}(\mathbf{U}_\mathbf{A} - \mathbf{U}_\mathbf{P}\mathbf{W}^*)\|_2\|\mathbf{V}_\mathbf{A}^{-1}\|_2 + \|\mathbf{U}_\mathbf{P}\|_{2\to\infty}\|\mathbf{W}^* - \mathbf{U}_\mathbf{P}^\mathrm{T}\mathbf{U}_\mathbf{A}\|_2\|\mathbf{V}_\mathbf{A}^{-1}\|_2\\
&\quad + \max_{i\in[n]}\|\mathbf{e}_i^\mathrm{T}\mathbf{U}_\mathbf{P}\mathbf{U}_\mathbf{P}^\mathrm{T}\mathbf{U}_\mathbf{A}(\mathbf{\Pi}_n^\mathrm{T}\mathbf{V}_\mathbf{P}\mathbf{U}_\mathbf{P}^\mathrm{T}\mathbf{U}_\mathbf{A})^{-1}(\mathbf{\Pi}_n^\mathrm{T}\mathbf{V}_\mathbf{P}\mathbf{U}_\mathbf{P}^\mathrm{T}\mathbf{U}_\mathbf{A} - \mathbf{V}_\mathbf{A})\mathbf{V}_\mathbf{A}^{-1}\|_2\\
&\leq \|\mathbf{U}_\mathbf{A} - \mathbf{U}_\mathbf{P}\mathbf{W}^*\|_{2\to\infty}\|\mathbf{V}_\mathbf{A}^{-1}\|_2 + \|\mathbf{U}_\mathbf{P}\|_{2\to\infty}\|\sin\Theta(\mathbf{U}_\mathbf{A}, \mathbf{U}_\mathbf{P})\|_2^2\|\mathbf{V}_\mathbf{A}^{-1}\|_2\\
&\quad + \max_{i\in[n]}\|\mathbf{e}_i^\mathrm{T}\mathbf{U}_\mathbf{P}\mathbf{V}_\mathbf{P}^{-1}\mathbf{\Pi}_n(\mathbf{\Pi}_n^\mathrm{T}\mathbf{V}_\mathbf{P}\mathbf{U}_\mathbf{P}^\mathrm{T}\mathbf{U}_\mathbf{A} - \mathbf{V}_\mathbf{A})\mathbf{V}_\mathbf{A}^{-1}\|_2\\
&\leq \|\mathbf{U}_\mathbf{A} - \mathbf{U}_\mathbf{P}\mathbf{W}^*\|_{2\to\infty}\|\mathbf{V}_\mathbf{A}^{-1}\|_2 + \|\mathbf{U}_\mathbf{P}\|_{2\to\infty}\frac{4\|\mathbf{E}\|_2^2}{(n\rho_n)^2\lambda_d(\mathbf{\Delta}_n)^2}\|\mathbf{V}_\mathbf{A}^{-1}\|_2\\
&\quad + \|\mathbf{\Theta}\mathbf{\Pi}_n\|_\infty\|\mathbf{\Pi}_n^\mathrm{T}\mathbf{V}_\mathbf{P}\mathbf{U}_\mathbf{P}^\mathrm{T}\mathbf{U}_\mathbf{A} - \mathbf{V}_\mathbf{A}\|_{2\to\infty}\|\mathbf{V}_\mathbf{A}^{-1}\|_2\\
&\leq \frac{K_1(\log n)^{1/2}}{n\rho_n^{1/2}} \times 2\sqrt{n} + \frac{1}{\sqrt{n}\lambda_d(\mathbf{\Delta}_n)^{1/2}} \times \frac{K_5}{(n\rho_n)\lambda_d(\mathbf{\Delta}_n)^2} \times 2\sqrt{n}\\
&\quad + \frac{K_4(\log n)^{1/2}}{n\rho_n^{1/2}} \times 2\sqrt{n}\\
&\leq K\sqrt{\frac{\log n}{n\rho_n}} \quad \text{with probability at least } 1 - c_0 n^{-2}
\end{aligned}
$$

for sufficiently large n for constants $K, K_5, c_0 > 0$. This completes the proof for the two-to-infinity norm error bounds on the membership profile matrix estimation.

■ **Asympototic normality of the estimators for the pure nodes.** Let $j_k \in [n]$ be the row index such that $\mathbf{\Pi}_n^\mathrm{T}\boldsymbol{\theta}_{j_k} = \mathbf{e}_k$. For each $k \in [d]$, define

$$
\mathcal{I}_k = \{i \in [n] : \mathbf{e}_i^\mathrm{T}\mathbf{\Theta} = \mathbf{e}_k\}, \quad \mathcal{J}_k = \left\{i \in [n] : \|\mathbf{e}_i^\mathrm{T}\widehat{\mathbf{\Theta}} - \mathbf{e}_k^\mathrm{T}\|_2 \leq \eta\right\}.
$$

Let $\pi_n \in \mathcal{S}_d$ be the permutation such that $\mathbf{\Pi}_n \mathbf{e}_k = \mathbf{e}_{\pi_n(k)}$. We claim that $\mathcal{I}_{\pi_n(k)} = \mathcal{J}_k$ with probability at least $1 - c_0 n^{-2}$ for sufficiently large n. For any $i \in \mathcal{I}_{\pi_n(k)}$, we know that $\mathbf{e}_i^{\mathrm{T}} \mathbf{\Theta} = \mathbf{e}_{\pi_n(k)} = \mathbf{\Pi}_n \mathbf{e}_k$. Therefore,

$$\begin{aligned}
\|\mathbf{e}_i^{\mathrm{T}} \widehat{\mathbf{\Theta}} - \mathbf{e}_k^{\mathrm{T}}\|_2 &= \|\mathbf{e}_i^{\mathrm{T}} \widehat{\mathbf{\Theta}} \mathbf{\Pi}_n^{\mathrm{T}} - \mathbf{e}_k^{\mathrm{T}} \mathbf{\Pi}_n^{\mathrm{T}}\|_2 = \|\mathbf{e}_i^{\mathrm{T}} \widehat{\mathbf{\Theta}} \mathbf{\Pi}_n^{\mathrm{T}} - \mathbf{e}_i^{\mathrm{T}} \mathbf{\Theta}\|_2 \\
&= \|\mathbf{e}_i^{\mathrm{T}} (\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\mathbf{\Pi}_n)\|_2 \leq \|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\mathbf{\Pi}_n\|_{2\to\infty} \\
&\leq K\sqrt{\frac{\log n}{n\rho_n}} \leq \eta \quad \text{with probability at least } 1 - c_0 n^{-2}
\end{aligned}$$

for sufficiently large n. This shows that $\mathcal{I}_{\pi_n(k)} \subset \mathcal{I}_j$ with probability at least $1 - c_0 n^{-2}$ for sufficiently large n. Conversely, for any $j \in \mathcal{J}_k$, we have

$$\begin{aligned}
\|\boldsymbol{\theta}_j - \mathbf{e}_{\pi_n(k)}\|_2 &= \|\mathbf{e}_j^{\mathrm{T}} \mathbf{\Theta} - \mathbf{e}_k^{\mathrm{T}} \mathbf{\Pi}_n^{\mathrm{T}}\|_2 = \|\mathbf{e}_j^{\mathrm{T}} \mathbf{\Theta}\mathbf{\Pi}_n - \mathbf{e}_k\|_2 \leq \|\mathbf{e}_j^{\mathrm{T}} (\mathbf{\Theta}\mathbf{\Pi}_n - \widehat{\mathbf{\Theta}})\|_2 + \|\mathbf{e}_j^{\mathrm{T}} \widehat{\mathbf{\Theta}} - \mathbf{e}_k^{\mathrm{T}}\|_2 \\
&\leq \|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\mathbf{\Pi}_n\|_{2\to\infty} + \eta \leq K\sqrt{\frac{\log n}{n\rho_n}} + \frac{c_2}{2} \quad \text{with probability at least } 1 - c_0 n^{-2}
\end{aligned}$$

for sufficiently large n. Then for sufficiently large n, we have

$$\|\boldsymbol{\theta}_j - \mathbf{e}_{\pi_n(k)}\|_2 \leq \frac{3c_2}{4} < \min_{i \in [n], \boldsymbol{\theta}_i \neq \mathbf{e}_{\pi_n(k)}} \|\boldsymbol{\theta}_i - \mathbf{e}_{\pi_n(k)}\|_2,$$

with probability at least $1 - c_0 n^{-2}$, implying that for sufficiently large n, $\boldsymbol{\theta}_j = \mathbf{e}_{\pi_n(k)}$ by the condition of the theorem. Therefore, $\mathcal{J}_k \subset \mathcal{I}_{\pi_n(k)}$ with probability at least $1 - c_0 n^{-2}$ for sufficiently large n, implying that

$$\iota_k = \min_{j \in \mathcal{J}_k} j = \min_{i \in \mathcal{I}_{\pi_n(k)}} i = i_{\pi_n(k)} \quad \text{with probability at least } 1 - c_0 n^{-2}$$

for sufficiently large n. Now for any $k \in [d]$, define

$$\widetilde{\mathbf{t}}_{nk} = \sqrt{n}\mathbf{\Sigma}_n(\mathbf{x}_k^*)^{-1/2}(\mathbf{W}^{\mathrm{T}} \widetilde{\mathbf{x}}_{i_k} - \rho_n^{1/2} \mathbf{x}_k^*) \quad \text{and} \quad \widehat{\mathbf{t}}_{nk} = \sqrt{n}\mathbf{G}_n(\mathbf{x}_k^*)^{1/2}(\mathbf{W}^{\mathrm{T}} \widehat{\mathbf{x}}_{i_k} - \rho_n^{1/2} \mathbf{x}_k^*).$$

By Theorem 4.4, we know that for any convex measurable $A \subset \mathbb{R}^d$,

$$\mathbb{P}\left(\widetilde{\mathbf{t}}_{nk} \in A\right) \to \mathbb{P}(\mathbf{z} \in A) \quad \text{and} \quad \mathbb{P}\left(\widehat{\mathbf{t}}_{nk} \in A\right) \to \mathbb{P}(\mathbf{z} \in A),$$

where $\mathbf{z} \sim \mathrm{N}_d(\mathbf{0}_d, \mathbf{I}_d)$. This implies that

$$\begin{aligned}
&\max_{\pi \in \mathcal{S}_d} \mathbb{P}\left(\widetilde{\mathbf{t}}_{n\pi(k)} \in A\right) \to \mathbb{P}(\mathbf{z} \in A), \quad \min_{\pi \in \mathcal{S}_d} \mathbb{P}\left(\widetilde{\mathbf{t}}_{n\pi(k)} \in A\right) \to \mathbb{P}(\mathbf{z} \in A), \\
&\max_{\pi \in \mathcal{S}_d} \mathbb{P}\left(\widehat{\mathbf{t}}_{n\pi(k)} \in A\right) \to \mathbb{P}(\mathbf{z} \in A), \quad \min_{\pi \in \mathcal{S}_d} \mathbb{P}\left(\widehat{\mathbf{t}}_{n\pi(k)} \in A\right) \to \mathbb{P}(\mathbf{z} \in A).
\end{aligned}$$

Hence, we have

$$\begin{aligned}
&\mathbb{P}\left\{\sqrt{n}\mathbf{\Sigma}_n(\mathbf{x}_{\pi_n(k)}^*)^{-1/2}(\mathbf{W}^{\mathrm{T}} \widetilde{\mathbf{x}}_{\iota_k} - \rho_n^{1/2} \mathbf{x}_{\pi_n(k)}^*) \in A\right\} \\
&\quad \leq \mathbb{P}\left\{\sqrt{n}\mathbf{\Sigma}_n(\mathbf{x}_{\pi_n(k)}^*)^{-1/2}(\mathbf{W}^{\mathrm{T}} \widetilde{\mathbf{x}}_{\iota_k} - \rho_n^{1/2} \mathbf{x}_{\pi_n(k)}^*) \in A, \iota_k = i_{\pi_n(k)}\right\} + \mathbb{P}(\iota_k \neq i_{\pi_n(k)})
\end{aligned}$$

$$
\begin{aligned}
&\le \mathbb{P}\left\{\sqrt{n}\boldsymbol{\Sigma}_n(\mathbf{x}^*_{\pi_n(k)})^{-1/2}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_{i_{\pi_n(k)}} - \rho_n^{1/2}\mathbf{x}^*_{\pi_n(k)}) \in A\right\} + c_0 n^{-2}\\
&\le \max_{\pi\in\mathcal{S}_d} \mathbb{P}(\widetilde{\mathbf{t}}_{n\pi(k)} \in A) + c_0 n^{-2} \to \mathbb{P}(\mathbf{z}\in A),
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}\left\{\sqrt{n}\boldsymbol{\Sigma}_n(\mathbf{x}^*_{\pi_n(k)})^{-1/2}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_{\iota_k} - \rho_n^{1/2}\mathbf{x}^*_{\pi_n(k)}) \in A\right\}\\
&\quad\ge \mathbb{P}\left\{\sqrt{n}\boldsymbol{\Sigma}_n(\mathbf{x}^*_{\pi_n(k)})^{-1/2}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_{\iota_k} - \rho_n^{1/2}\mathbf{x}^*_{\pi_n(k)}) \in A, \iota_k = i_{\pi_n(k)}\right\}\\
&\quad= \mathbb{P}\left(\widetilde{\mathbf{t}}_{n\pi_n(k)} \in A\right) + \mathbb{P}\left(\iota_k = i_{\pi_n(k)}\right) - \mathbb{P}\left[\left\{\widetilde{\mathbf{t}}_{n\pi_n(k)} \in A\right\} \cup \left\{\iota_k = i_{\pi_n(k)}\right\}\right]\\
&\quad\ge \mathbb{P}\left(\widetilde{\mathbf{t}}_{n\pi_n(k)} \in A\right) + 1 - c_0 n^{-2} - 1\\
&\quad\ge \min_{\pi\in\mathcal{S}_d} \mathbb{P}(\widetilde{\mathbf{t}}_{n\pi(k)} \in A) - c_0 n^{-2} \to \mathbb{P}(\mathbf{z}\in A).
\end{aligned}
$$

This implies that $\widetilde{\mathbf{t}}_{n\pi_n(k)} \to \mathrm{N}_d(\mathbf{0}_d, \mathbf{I}_d)$. The same reasoning also implies that $\widehat{\mathbf{t}}_{n\pi_n(k)} \to \mathrm{N}_d(\mathbf{0}_d, \mathbf{I}_d)$, and the proof is thus completed.

S8.2. Proof of Theorem 4.10

By the proof of Theorem 4.4, Theorem 3.2, and Theorem 4.7, under the condition that d is fixed and $\lambda_d(\boldsymbol{\Delta}_n)$ is bounded away from 0, we have

$$
\begin{aligned}
(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_{0i}) &= \frac{1}{n\rho_n^{1/2}}\boldsymbol{\Delta}_n^{-1}\sum_{a=1}^n (A_{ia} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)\mathbf{x}_a + o_{\mathbb{P}}(n^{-1/2}),\\
(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_{0i}) &= \frac{1}{n\rho_n^{1/2}}\sum_{a=1}^n \frac{(A_{ia} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)\mathbf{x}_a}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)} + o_{\mathbb{P}}(n^{-1/2}).
\end{aligned}
$$

Since $|A_{ij} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j| \le 1$ and $n\rho_n^{1/2} = \omega(n^{1/2})$, we have,

$$
\begin{aligned}
\sqrt{n}\left\{\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j) - \rho_n^{1/2}(\mathbf{x}_i - \mathbf{x}_j)\right\} &= \frac{1}{(n\rho_n)^{1/2}}\boldsymbol{\Delta}_n^{-1}\sum_{a=1}^n (A_{ia} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)\mathbf{x}_a\\
&\quad - \frac{1}{(n\rho_n)^{1/2}}\boldsymbol{\Delta}_n^{-1}\sum_{b\ne i}^n (A_{jb} - \rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b)\mathbf{x}_b + o_{\mathbb{P}}(1),\\
\sqrt{n}\left\{\mathbf{W}^{\mathrm{T}}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j) - \rho_n^{1/2}(\mathbf{x}_i - \mathbf{x}_j)\right\} &= \frac{1}{(n\rho_n)^{1/2}}\sum_{a=1}^n \frac{(A_{ia} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)\mathbf{x}_a}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)}\\
&\quad - \frac{1}{(n\rho_n)^{1/2}}\sum_{b\ne i}^n \frac{(A_{jb} - \rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b)\mathbf{x}_b}{\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b(1-\rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b)} + o_{\mathbb{P}}(1).
\end{aligned}
$$

Note that

$$
\frac{1}{(n\rho_n)^{1/2}}\boldsymbol{\Delta}_n^{-1}\sum_{a=1}^n (A_{ia} - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)\mathbf{x}_a - \frac{1}{(n\rho_n)^{1/2}}\boldsymbol{\Delta}_n^{-1}\sum_{b\ne i}^n (A_{jb} - \rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b)\mathbf{x}_b \quad \text{and}
$$

$$\frac{1}{(n\rho_n)^{1/2}}\sum_{a=1}^{n}\frac{(A_{ia}-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)\mathbf{x}_a}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)}-\frac{1}{(n\rho_n)^{1/2}}\sum_{b\neq i}^{n}\frac{(A_{jb}-\rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b)\mathbf{x}_b}{\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b(1-\rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b)}$$

are sums of mean-zero independent random vectors. In addition, observe that $\mathbf{\Sigma}_n(\mathbf{x}_i)$, $\mathbf{\Sigma}_n(\mathbf{x}_i)^{-1}$, $\mathbf{G}_n(\mathbf{x}_i)$, and $\mathbf{G}_n(\mathbf{x}_i)^{-1}$ are all $O(1)$ and $\Omega(1)$, and that

$$\begin{aligned}
&\operatorname{var}\left\{\frac{1}{(n\rho_n)^{1/2}}\mathbf{\Delta}_n^{-1}\sum_{a=1}^{n}(A_{ia}-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)\mathbf{x}_a-\frac{1}{(n\rho_n)^{1/2}}\mathbf{\Delta}_n^{-1}\sum_{b\neq i}^{n}(A_{jb}-\rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b)\mathbf{x}_b\right\}\\
&\quad=\mathbf{\Sigma}_n(\mathbf{x}_i)+\mathbf{\Sigma}_n(\mathbf{x}_j)+o(1),\\
&\operatorname{var}\left\{\frac{1}{(n\rho_n)^{1/2}}\sum_{a=1}^{n}\frac{(A_{ia}-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)\mathbf{x}_a}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)}-\frac{1}{(n\rho_n)^{1/2}}\sum_{b\neq i}^{n}\frac{(A_{jb}-\rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b)\mathbf{x}_b}{\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b(1-\rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b)}\right\}\\
&\quad=\mathbf{G}_n(\mathbf{x}_i)^{-1}+\mathbf{G}_n(\mathbf{x}_j)^{-1}+o(1),\\
&\sum_{a=1}^{n}\mathbb{E}\left\|\mathbf{\Delta}_n^{-1}\frac{(A_{ia}-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)\mathbf{x}_a}{(n\rho_n)^{1/2}}\right\|_2^3+\sum_{b\neq i}^{n}\mathbb{E}\left\|\mathbf{\Delta}_n^{-1}\frac{(A_{jb}-\rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b)\mathbf{x}_b}{(n\rho_n)^{1/2}}\right\|_2^3\\
&\quad\leq\frac{\|\mathbf{\Delta}_n^{-1}\|_2^3}{(n\rho_n)^{1/2}}\left\{\frac{1}{n}\sum_{a=1}^{n}\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)\|\mathbf{x}_a\|_2^3+\frac{1}{n}\sum_{b\neq i}^{n}\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b(1-\rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b)\|\mathbf{x}_b\|_2^3\right\}\\
&\quad\leq\frac{\|\mathbf{\Delta}_n^{-1}\|_2^3}{(n\rho_n)^{1/2}}\to0,\\
&\sum_{a=1}^{n}\mathbb{E}\left\|\frac{(A_{ia}-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)\mathbf{x}_a}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)(n\rho_n)^{1/2}}\right\|_2^3+\sum_{b\neq i}^{n}\mathbb{E}\left\|\frac{(A_{jb}-\rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b)\mathbf{x}_b}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)(n\rho_n)^{1/2}}\right\|_2^3\\
&\quad\leq\frac{1}{(n\rho_n)^{1/2}}\left\{\frac{1}{n}\sum_{a=1}^{n}\frac{\|\mathbf{x}_a\|_2^3}{\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_a)\}^2}+\frac{1}{n}\sum_{b\neq i}^{n}\frac{\|\mathbf{x}_b\|_2^3}{\{\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b(1-\rho_n\mathbf{x}_j^{\mathrm{T}}\mathbf{x}_b)\}^2}\right\}\\
&\quad\leq\frac{2}{(n\rho_n)^{1/2}\delta^4}\to0.
\end{aligned}$$

Therefore, by Lyapunov's central limit theorem (see, for example, Theorem 7.1.2 in [5]),

$$\sqrt{n}\{\mathbf{\Sigma}_n(\mathbf{x}_i)+\mathbf{\Sigma}_n(\mathbf{x}_j)\}^{-1/2}\left\{\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i-\widetilde{\mathbf{x}}_j)-\rho_n^{1/2}(\mathbf{x}_i-\mathbf{x}_j)\right\}\xrightarrow{\mathcal{L}}\mathrm{N}(\mathbf{0}_d,\mathbf{I}_d),$$

$$\sqrt{n}\{\mathbf{G}_n(\mathbf{x}_i)^{-1}+\mathbf{G}_n(\mathbf{x}_j)^{-1}\}^{-1/2}\left\{\mathbf{W}^{\mathrm{T}}(\widehat{\mathbf{x}}_i-\widehat{\mathbf{x}}_j)-\rho_n^{1/2}(\mathbf{x}_i-\mathbf{x}_j)\right\}\xrightarrow{\mathcal{L}}\mathrm{N}(\mathbf{0}_d,\mathbf{I}_d).$$

We next show that

$$\begin{aligned}
\rho_n^{-1}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{\Delta}}_n\mathbf{W}&=\mathbf{\Delta}_n+o_{\mathbb{P}}(1),\\
\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{\Sigma}}_n(\widetilde{\mathbf{x}}_i)\mathbf{W}&=\mathbf{\Sigma}_n(\mathbf{x}_i)+o_{\mathbb{P}}(1),\\
\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{G}}_n(\widetilde{\mathbf{x}}_i)\mathbf{W}&=\mathbf{G}_n(\mathbf{x}_i)+o_{\mathbb{P}}(1).
\end{aligned}$$

For the first equation, by Lemma S7.1, we have

$$\|\rho_n^{-1}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{\Delta}}_n\mathbf{W} - \mathbf{\Delta}_n\|_2 \le \frac{1}{n\rho_n}\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}\|_{\mathrm{F}}\|\widetilde{\mathbf{X}}\mathbf{W}\|_2 + \frac{1}{n\rho_n}\|\rho_n^{1/2}\mathbf{X}\|_2\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}\|_2$$
$$= \frac{1}{n\rho_n}O_{\mathbb{P}}(1)\times O_{\mathbb{P}}((n\rho_n)^{1/2}) = O_{\mathbb{P}}\left(\frac{1}{(n\rho_n)^{1/2}}\right).$$

For the second equation, we denote

$$\widetilde{\mathbf{D}}_i = \frac{1}{\rho_n}\mathrm{diag}\left\{\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_1(1-\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_1),\ldots,\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_n(1-\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_n)\right\},$$
$$\mathbf{D}_i = \mathrm{diag}\left\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_1(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_1),\ldots,\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_n(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_n)\right\}.$$

By Result S7.1 and Corollary 4.1, we have

$$\max_{i,j\in[n]}|\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j| \le \left(\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}\|_{2\to\infty} + 2\max_{i,j\in[n]}\|\rho_n^{1/2}\mathbf{X}\|_{2\to\infty}\right)\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}\|_{2\to\infty}$$
$$= O_{\mathbb{P}}\left(\rho_n\sqrt{\frac{\log n}{n\rho_n}}\right)$$

and

$$\max_{i,j\in[n]}|(\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j)^2 - (\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)^2| \le \max_{i,j\in[n]}|\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j|\left(|\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j| + 2\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j\right)$$
$$= O_{\mathbb{P}}\left(\rho_n^2\sqrt{\frac{\log n}{n\rho_n}}\right).$$

It follows that

$$\|\widetilde{\mathbf{D}}_i - \mathbf{D}_i\|_2 \le \rho_n^{-1}\max_{i,j\in[n]}|\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j| + \rho_n^{-1}\max_{i,j\in[n]}|(\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j)^2 - (\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)^2| = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n\rho_n}}\right).$$

Therefore, by Lemma S7.1,

$$\left\|\mathbf{W}^{\mathrm{T}}\frac{1}{n\rho_n^2}\sum_{j=1}^n\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j(1-\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j)\widetilde{\mathbf{x}}_j\widetilde{\mathbf{x}}_j^{\mathrm{T}}\mathbf{W} - \frac{1}{n}\sum_{j=1}^n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\right\|_2$$
$$= \frac{1}{n}\left\|(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W})^{\mathrm{T}}\widetilde{\mathbf{D}}_i(\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W}) - \mathbf{X}^{\mathrm{T}}\mathbf{D}_i\mathbf{X}\right\|_2$$
$$\le \frac{1}{n}\left\|\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W} - \mathbf{X}\right\|_2\|\widetilde{\mathbf{D}}_i\|_2\|\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W}\|_2 + \frac{1}{n}\|\mathbf{X}\|_2\|\widetilde{\mathbf{D}}_i - \mathbf{D}_i\|_2\|\rho_n^{-1/2}\mathbf{X}\mathbf{W}\|_2$$
$$+ \frac{1}{n}\|\mathbf{X}\|_2\|\mathbf{D}_i\|_2\|\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W} - \mathbf{X}\|_2$$
$$= \frac{1}{n}\times O_{\mathbb{P}}(\rho_n^{-1/2})\times O_{\mathbb{P}}(1)\times O_{\mathbb{P}}(n^{1/2}) + \frac{1}{n}\times O(n^{1/2})\times O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n\rho_n}}\right)\times O_{\mathbb{P}}(n^{1/2})$$

$$+\frac{1}{n}\times O(n^{1/2})\times O(1)\times O_{\mathbb{P}}(\rho_n^{-1/2}) = o_{\mathbb{P}}(1).$$

Hence, for the second equation, we have

$$\begin{aligned}
\mathbf{W}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}_n(\widetilde{\mathbf{x}}_i)\mathbf{W} &= (\rho_n^{-1}\mathbf{W}^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}}_n\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\left\{\frac{1}{n\rho_n^2}\sum_{j=1}^n \widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j(1-\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j)\widetilde{\mathbf{x}}_j\widetilde{\mathbf{x}}_j^{\mathrm{T}}\mathbf{W}\right\}(\rho_n^{-1}\mathbf{W}^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}}_n\mathbf{W})^{-1}\\
&= \{\boldsymbol{\Delta}_n + o_{\mathbb{P}}(1)\}^{-1}\left\{\frac{1}{n}\sum_{j=1}^n \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}} + o_{\mathbb{P}}(1)\right\}\{\boldsymbol{\Delta}_n + o_{\mathbb{P}}(1)\}^{-1}\\
&= \boldsymbol{\Delta}_n^{-1}\left\{\frac{1}{n}\sum_{j=1}^n \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\right\}\boldsymbol{\Delta}_n^{-1} + o_{\mathbb{P}}(1)\\
&= \boldsymbol{\Sigma}_n(\mathbf{x}_i) + o_{\mathbb{P}}(1).
\end{aligned}$$

For the third equation, it follows directly from Lemma S7.10 with $t = \log n$. Hence, we conclude that

$$\begin{aligned}
\mathbf{W}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}_{ij}\mathbf{W} &= \mathbf{W}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}_n(\widetilde{\mathbf{x}}_i)\mathbf{W} + \mathbf{W}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}_n(\widetilde{\mathbf{x}}_j)\mathbf{W} = \boldsymbol{\Sigma}_n(\mathbf{x}_i) + \boldsymbol{\Sigma}_n(\mathbf{x}_j) + o_{\mathbb{P}}(1),\\
\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{G}}_{ij}\mathbf{W} &= \mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{G}}_n(\widetilde{\mathbf{x}}_i)^{-1}\mathbf{W} + \mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{G}}_n(\widetilde{\mathbf{x}}_j)^{-1}\mathbf{W} = \mathbf{G}_n(\mathbf{x}_i)^{-1} + \mathbf{G}_n(\mathbf{x}_j)^{-1} + o_{\mathbb{P}}(1).
\end{aligned}$$

By Slutsky's lemma, under the null distribution $H_0 : \mathbf{x}_i = \mathbf{x}_j$, we have

$$\begin{aligned}
T_{ij}^{(\mathrm{ASE})} &= n(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}_{ij}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)\\
&= n\{\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)\}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}_{ij}\mathbf{W}\{\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)\}\\
&= n\{\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)\}^{\mathrm{T}}[\{\boldsymbol{\Sigma}_n(\mathbf{x}_i) + \boldsymbol{\Sigma}_n(\mathbf{x}_j)\}^{-1} + o_{\mathbb{P}}(1)]\{\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)\}\\
&\xrightarrow{\mathcal{L}} \chi_d^2,\\
T_{ij}^{(\mathrm{OSE})} &= n(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)^{\mathrm{T}}\widetilde{\mathbf{G}}_{ij}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)\\
&= n\{\mathbf{W}^{\mathrm{T}}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)\}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{G}}_{ij}\mathbf{W}\{\mathbf{W}^{\mathrm{T}}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)\}\\
&= n\{\mathbf{W}^{\mathrm{T}}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)\}^{\mathrm{T}}[\{\mathbf{G}_n(\mathbf{x}_i)^{-1} + \mathbf{G}_n(\mathbf{x}_j)^{-1}\}^{-1} + o_{\mathbb{P}}(1)]\{\mathbf{W}^{\mathrm{T}}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)\}\\
&\xrightarrow{\mathcal{L}} \chi_d^2.
\end{aligned}$$

We now consider the distributions of $T_{ij}^{(\mathrm{ASE})}$ and $T_{ij}^{(\mathrm{OSE})}$ under the alternative $H_A : \mathbf{x}_i \neq \mathbf{x}_j$ but $(n\rho_n)^{1/2}(\mathbf{x}_i - \mathbf{x}_j) \to \boldsymbol{\mu} \neq \mathbf{0}_d$. Under the condition that $\boldsymbol{\Sigma}_n(\mathbf{x}_i) \to \boldsymbol{\Sigma}_i$ and $\mathbf{G}_n(\mathbf{x}_i) \to \mathbf{G}_i$, we have,

$$\begin{aligned}
\sqrt{n}\{\boldsymbol{\Sigma}_n(\mathbf{x}_i) + \boldsymbol{\Sigma}_n(\mathbf{x}_j)\}^{-1/2}\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j) &\xrightarrow{\mathcal{L}} \mathrm{N}\left(\boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j)^{-1}\boldsymbol{\mu}, \mathbf{I}_d\right),\\
\sqrt{n}\{\mathbf{G}_n(\mathbf{x}_i)^{-1} + \mathbf{G}_n(\mathbf{x}_j)^{-1}\}^{-1/2}\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j) &\xrightarrow{\mathcal{L}} \mathrm{N}\left(\boldsymbol{\mu}^{\mathrm{T}}(\mathbf{G}_i + \mathbf{G}_j)^{-1}\boldsymbol{\mu}, \mathbf{I}_d\right).
\end{aligned}$$

Since $\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j) = O_{\mathbb{P}}(n^{-1/2})$ and $\mathbf{W}^{\mathrm{T}}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j) = O_{\mathbb{P}}(n^{-1/2})$, it follows that under $H_A : \mathbf{x}_i \neq \mathbf{x}_j$ but $(n\rho_n)^{1/2}(\mathbf{x}_i - \mathbf{x}_j) \to \boldsymbol{\mu} \neq \mathbf{0}_d$,

$$\begin{aligned}
T_{ij}^{(\mathrm{ASE})} &= n(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}_{ij}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)\\
&= n\{\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)\}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}_{ij}\mathbf{W}\{\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)\}\\
&= n\{\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)\}^{\mathrm{T}}[\{\boldsymbol{\Sigma}_n(\mathbf{x}_i) + \boldsymbol{\Sigma}_n(\mathbf{x}_j)\}^{-1} + o_{\mathbb{P}}(1)]\{\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)\}\\
&= \left\|\sqrt{n}\{\boldsymbol{\Sigma}_n(\mathbf{x}_i) + \boldsymbol{\Sigma}_n(\mathbf{x}_j)\}^{-1/2}\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)\right\|_2^2 + o_{\mathbb{P}}(n\|\mathbf{W}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)\|_2^2)\\
&\overset{\mathcal{L}}{\to} \chi_d^2(\boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j)^{-1}\boldsymbol{\mu}),\\
T_{ij}^{(\mathrm{OSE})} &= n(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)^{\mathrm{T}}\widetilde{\mathbf{G}}_{ij}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)\\
&= n\{\mathbf{W}^{\mathrm{T}}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)\}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{G}}_{ij}\mathbf{W}\{\mathbf{W}^{\mathrm{T}}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)\}\\
&= n\{\mathbf{W}^{\mathrm{T}}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)\}^{\mathrm{T}}[\{\mathbf{G}_n(\mathbf{x}_i)^{-1} + \mathbf{G}_n(\mathbf{x}_j)^{-1}\}^{-1} + o_{\mathbb{P}}(1)]\{\mathbf{W}^{\mathrm{T}}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)\}\\
&= \left\|\sqrt{n}\{\mathbf{G}_n(\mathbf{x}_i)^{-1} + \mathbf{G}_n(\mathbf{x}_j)^{-1}\}^{-1/2}\mathbf{W}^{\mathrm{T}}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)\right\|_2^2 + o_{\mathbb{P}}(n\|\mathbf{W}^{\mathrm{T}}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)\|_2^2)\\
&\overset{\mathcal{L}}{\to} \chi_d^2(\boldsymbol{\mu}^{\mathrm{T}}(\mathbf{G}_i^{-1} + \mathbf{G}_j^{-1})^{-1}\boldsymbol{\mu}).
\end{aligned}$$

The proof is thus completed.

References

[1] ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* **48** 1452–1474. https://doi.org/10.1214/19-AOS1854

[2] ATHREYA, A., FISHKIND, D. E., TANG, M., PRIEBE, C. E., PARK, Y., VOGELSTEIN, J. T., LEVIN, K., LYZINSKI, V., QIN, Y. and SUSSMAN, D. L. (2018). Statistical Inference on Random Dot Product Graphs: a Survey. *J. Mach. Learn. Res.* **18** 1–92.

[3] BHATIA, R. (1997). *Matrix analysis* **169**. Springer Science & Business Media. https://doi.org/10.1007/978-1-4612-0653-8

[4] CAPE, J., TANG, M. and PRIEBE, C. E. (2019). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Ann. Statist.* **47** 2405–2439. https://doi.org/10.1214/18-AOS1752

[5] CHUNG, K. L. (2001). *A course in probability theory*. Academic press.

[6] GILLIS, N. and VAVASIS, S. A. (2014). Fast and Robust Recursive Algorithms for Separable Nonnegative Matrix Factorization. *IEEE Trans. Pattern Anal. Mach. Intell.* **36** 698–714. https://doi.org/10.1109/TPAMI.2013.226

[7] LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. https://doi.org/10.1214/14-AOS1274

[8] LEI, L. (2019). Unified $\ell_{2\to\infty}$ Eigenspace Perturbation Theory for Symmetric Random Matrices. *arXiv preprint:1909.04798*. https://doi.org/10.48550/arXiv.1909.04798

[9] MAGNUS, J. R. and NEUDECKER, H. (1985). Matrix differential calculus with applications to simple, hadamard, and kronecker products. *J. Math. Psychol.* **29** 474–492. https://doi.org/10.1016/0022-2496(85)90006-9

[10] MAO, X., SARKAR, P. and CHAKRABARTI, D. (2020). Estimating Mixed Memberships With Sharp Eigenvector Deviations. *J. Amer. Statist. Assoc.* **0** 1–13. https://doi.org/10.1080/01621459.2020.1751645

[11] MINSKER, S. (2017). On some extensions of Bernstein's inequality for self-adjoint operators. *Stat. Probab. Lett.* **127** 111–119. https://doi.org/10.1016/j.spl.2017.03.020

[12] PAULSEN, V. (2002). *Completely bounded maps and operator algebras* **78**. Cambridge University Press. https://doi.org/10.1017/CBO9780511546631

[13] POLLARD, D. (1990). Empirical processes: theory and applications. In *NSF-CBMS Regional Conf. Ser. Probab. Statist.* i–86. https://doi.org/10.1214/cbms/1462061091

[14] RUBIN-DELANCHY, P., CAPE, J., TANG, M. and PRIEBE, C. E. (2022). A statistical interpretation of spectral embedding: The generalised random dot product graph. *J. R. Stat. Soc., B: Stat. Methodol.* **84** 1446–1473. https://doi.org/10.1111/rssb.12509

[15] SHAO, Q.-M. and ZHANG, Z.-S. (2022). Berry–Esseen bounds for multivariate nonlinear statistics with applications to M-estimators and stochastic gradient descent algorithms. *Bernoulli* **28** 1548–1576. https://doi.org/10.3150/21-BEJ1336

[16] TANG, M., ATHREYA, A., SUSSMAN, D. L., LYZINSKI, V., PARK, Y. and PRIEBE, C. E. (2017). A Semiparametric Two-Sample Hypothesis Testing Problem for Random Graphs. *J. Comput. Graph. Statist.* **26** 344–354. https://doi.org/10.1080/10618600.2016.1193505

[17] TROPP, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12** 389–434. https://doi.org/10.1007/s10208-011-9099-z

[18] VERSHYNIN, R. (2012). *Introduction to the non-asymptotic analysis of random matrices* In *Compressed Sensing: Theory and Applications* 210–268. Cambridge University Press. https://doi.org/10.1017/CBO9780511794308.006

[19] XIE, F. and XU, Y. (2021). Efficient Estimation for Random Dot Product Graphs via a One-Step Procedure. *Journal of the American Statistical Association* **0** 1–14. https://doi.org/10.1080/01621459.2021.1948419