Supplementary Material for “Efficient Estimation of Random Dot Product Graphs via a One-Step Procedure”

This document contains the following supplementary materials:

- A comprehensive list of notations used throughout document (Section A).
- Proofs of the main theorems, namely, Theorems 1 through 9 (Sections B through F).
- One-step estimator in the context of positive definite stochastic block models (Section G).
- Further discussion of the applicability of the proposed theory to sparse graphs (Section H).
- Additional simulated examples (Section I).

Contents

A List of Notations

In Table 1 we prepare a comprehensive lists of notations that are repeatedly used throughout the paper as well as the supplementary material.

Table 1: Table of Notations

$\mathbf{I}_d$	$d \times d$ identity matrix
$\mathbf{0}$	The Eulidean vector of all zeros in its coordinates
$\mathbf{1}$	The Eulidean vector of all ones in its coordinates
$[n]$	$[n] = \{1, 2, \ldots, n\}$; In this work $[n]$ may denote the set of all vertices.
$a \lesssim b$	$a \lesssim b$ if $a \leq Cb$ for some constant $C > 0$
$a \gtrsim b$	$a \gtrsim b$ if $a \geq Cb$ for some constant $C > 0$
$a \vee b$	$a \vee b = \max(a, b)$
$a \wedge b$	$a \wedge b = \min(a, b)$
$a \asymp b$	$a \asymp b$ if $a \lesssim b$ and $a \gtrsim b$
$\mathbb{O}(n, d)$	$\mathbb{O}(n, d) = \{\mathbf{U} \in \mathbb{R}^{n \times d} : \mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_d\}$, where $n \geq d$
$\mathbb{O}(d)$	$\mathbb{O}(d) = \mathbb{O}(d, d)$
$\mathbf{x} \leq \mathbf{y}$	For vectors $\mathbf{x} = [x_1, \ldots, x_d]^{\mathrm{T}}$ and $\mathbf{y} = [y_1, \ldots, y_d]^{\mathrm{T}}$ in $\mathbb{R}^d$, the inequality $\mathbf{x} \leq \mathbf{y}$ means that $x_k \leq y_k$ for all $k = 1, 2, \ldots, d$
$\|\mathbf{x}\|$	For a vector $\mathbf{x} = [x_1, \ldots, x_d]^{\mathrm{T}} \in \mathbb{R}^d$, $\|\mathbf{x}\|_2 = (\sum_k x_k^2)^{1/2}$
$\mathbf{\Sigma}_1 \preceq \mathbf{\Sigma}_2$	$\mathbf{\Sigma}_2 - \mathbf{\Sigma}_1$ is positive semidefinite
$\mathbf{\Sigma}_1 \succeq \mathbf{\Sigma}_2$	$\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ is positive semidefinite
$\mathbf{\Sigma}_1 \prec \mathbf{\Sigma}_2$	$\mathbf{\Sigma}_2 - \mathbf{\Sigma}_1$ is positive definite
$\mathbf{\Sigma}_1 \succ \mathbf{\Sigma}_2$	$\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ is positive definite
$\sigma_k(\mathbf{Z})$	kth largest singular value of a matrix $\mathbf{Z}$
$\|\mathbf{Z}\|_2$	Spectral norm of a matrix $\mathbf{Z}$: $\|\mathbf{Z}\|_2 = \sigma_1(\mathbf{Z})$

Table 1: Table of Notations (continued)

$\\|\mathbf{Z}\\|_{\mathrm{F}}$	Frobenius norm of a matrix $\mathbf{Z} = [Z_{ik}]_{n\times d}$: $\\|\mathbf{Z}\\|_{\mathrm{F}} = (\sum_{i,k} Z_{ik}^2)^{1/2}$
$\\|\mathbf{Z}\\|_{2\to\infty}$	Two-to-infinity norm of a matrix $\mathbf{Z} = [Z_{ik}]_{n\times d}$: $\\|\mathbf{Z}\\|_{2\to\infty} = \max_i(\sum_k z_{ik}^2)^{1/2}$
$\mathrm{diag}(\mathbf{z})$	For a vector $\mathbf{z} = [x_1, \ldots, x_n]^{\mathrm{T}} \in \mathbb{R}^n$, the matrix $\mathrm{diag}(\mathbf{x})$ is a diagonal matrix with the ith diagonal entries being z_i, $i = 1, \ldots, n$
$\mathcal{X}$	Latent space $\mathcal{X} = \{\mathbf{x} = [x_1, \ldots, x_d] \in \mathbb{R}^d : x_1, \ldots, x_d > 0, \\|\mathbf{x}\\| < 1\}$
$\mathcal{X}(\delta)$	A subset of $\mathcal{X}$ such that $\mathbf{x}, \mathbf{u} \in \mathcal{X}(\delta)$ implies $\mathbf{x}^{\mathrm{T}}\mathbf{u} \in [\delta, 1-\delta]$ for some $\delta > 0$
$\mathcal{X}^n$	n-fold cartesion product of $\mathcal{X}$: $\mathcal{X}^n = \{\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^{\mathrm{T}} \in \mathbb{R}^{n\times d} : \mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{X}\}$
n	Number of vertices
d	Dimension of underlying latent positions
$\mathbf{X}$	Latent position matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^{\mathrm{T}} \in \mathcal{X}^n$
$\mathbf{x}_i$	The ith row of $\mathbf{X}$ as a column vector in $\mathbb{R}^d$
$\mathbf{A}$	Adjacency matrix drawn from a random dot product graph
ρ_n	Sparsity factor $\rho_n \in (0, 1]$ such that $A_{ij} \sim \mathrm{Bernoulli}(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)$
$\mathbf{X}_0$	True latent position matrix $\mathbf{X}_0 = [\mathbf{x}_{01}, \ldots, \mathbf{x}_{0n}]^{\mathrm{T}} \in \mathcal{X}^n$
$\mathbf{x}_{0i}$	The ith row of $\mathbf{X}_0$ as a column vector in $\mathbb{R}^d$
F_n	Empirical distribution function of $(\mathbf{x}_{0i})_{i=1}^n$: $F_n(\mathbf{x}) = (1/n)\sum_{i=1}^n \mathbf{x}_{\mathrm{i}} \leq \mathbf{x}$
F	The cumulative distribution function such that (2.1) holds
$\mathbf{W}$	A generic orthogonal matrix in $\mathbb{O}(d)$
$\boldsymbol{\Delta}$	Second moment matrix of F: $\boldsymbol{\Delta} = \int_{\mathcal{X}} \mathbf{x}\mathbf{x}^{\mathrm{T}} F(\mathrm{d}\mathbf{x})$
τ	Cluster assignment function $\tau : [n] \to [K]$ in a K-block stochastic block model
$\mathbf{B}$	A symmetric $K \times K$ block probability matrix for a K-block stochastic block model
$\widehat{\mathbf{X}}^{(\mathrm{ASE})}$	The adjacency spectral embedding of $\mathbf{A}$ into $\mathbb{R}^d$
$\boldsymbol{\nu}_k$	A unique latent position in a positive semidefinite stochastic block model
$\widehat{\mathbf{x}}_i^{(\mathrm{ASE})}$	The ith row of $\widehat{\mathbf{X}}^{(\mathrm{ASE})}$ as a column vector in $\mathbb{R}^d$
ρ	$\rho = \lim_{n\to\infty} \rho_n$; In this work we assume $\rho \in \{0, 1\}$.

Table 1: Table of Notations (continued)

$\mathbf{\Sigma}(\mathbf{x})$	$\mathbf{\Delta}^{-1}\int_{\mathcal{X}}\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_1(1-\rho\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)\}\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}F(\mathrm{d}\mathbf{x}_1)\mathbf{\Delta}^{-1}$, where $\mathbf{x}\in\mathcal{X}(\delta)$
$\mathbf{G}(\mathbf{x})$	$\int_{\mathcal{X}}\left\{\frac{\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}_1(1-\rho\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)}\right\}F(\mathrm{d}\mathbf{x}_1)$, where $\mathbf{x}\in\mathcal{X}(\delta)$
$\mathbf{\Delta}_n$	$\int_{\mathcal{X}}\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}F_n(\mathrm{d}\mathbf{x})$, and can be alternatively expressed as $(1/n)\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0$.
$\mathbf{\Sigma}_n(\mathbf{x})$	$\int_{\mathcal{X}}\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_1(1-\rho\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)\}\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}F_n(\mathrm{d}\mathbf{x})$, where $\mathbf{x}\in\mathcal{X}$
$\mathbf{G}_n(\mathbf{x})$	$\int_{\mathcal{X}}\left\{\frac{\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}_1(1-\rho\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)}\right\}F_n(\mathrm{d}\mathbf{x})$, where $\mathbf{x}\in\mathcal{X}(\delta)$.
$\mathbf{\Psi}_n(\mathbf{x})$	Score function $\mathbf{\Psi}_n(\mathbf{x})=\frac{1}{n}\sum_{j\neq i}^{n}\frac{(A_{ij}-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})}$
$\widetilde{\mathbf{x}}_i$	Initial estimator of $\mathbf{x}_i$ for the one-step procedures (3.1) and (3.2)
$\widetilde{\mathbf{X}}$	Initial estimator of $\mathbf{X}$ for the one-step procedure in (3.2)
$\widehat{\mathbf{X}}$	The one-step estimator for $\mathbf{X}$ defined by (3.2)
$\widehat{\mathbf{x}}_i$	The ith row of $\widehat{\mathbf{X}}$ as a column vector in $\mathbb{R}^d$
$\mathcal{L}(\mathbf{M})$	The normalized Laplacian of a square matrix $\mathbf{M}$ defined by $\mathcal{L}(\mathbf{M})=(\mathrm{diag}(\mathbf{M}\mathbf{1}))^{-1/2}\mathbf{M}(\mathrm{diag}(\mathbf{M}\mathbf{1}))^{-1/2}$
$\mathbf{Y}$	The population Laplacian spectral embedding: $\mathbf{Y}=\mathbf{Y}(\mathbf{X})=(\mathrm{diag}(\mathbf{X}\mathbf{1}))^{-1/2}\mathbf{X}$
$\mathbf{y}_i$	The ith row of $\mathbf{Y}$ as a column vector in $\mathbb{R}^d$
$\mathbf{Y}_0$	The true value of $\mathbf{Y}$ given by $\mathbf{Y}_0=(\mathrm{diag}(\mathbf{X}_0\mathbf{X}_0\mathbf{1}))^{-1/2}\mathbf{X}_0$
$\mathbf{y}_{0i}$	The ith row of $\mathbf{Y}_0$ as a column vector in $\mathbb{R}^d$
$\breve{\mathbf{X}}$	The (sample) Laplacian spectral embedding of $\mathbf{A}$ into $\mathbb{R}^d$
$\boldsymbol{\mu}$	$\int_{\mathcal{X}}\mathbf{x}F(\mathrm{d}\mathbf{x})$
$\widetilde{\mathbf{\Delta}}$	$\int_{\mathcal{X}}\frac{\mathbf{x}\mathbf{x}^{\mathrm{T}}}{\mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}}F(\mathrm{d}\mathbf{x})$
$\widetilde{\mathbf{\Sigma}}(\mathbf{x})$	$\left(\widetilde{\mathbf{\Delta}}^{-1}-\frac{\mathbf{x}\boldsymbol{\mu}^{\mathrm{T}}}{2\mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}}\right)\int_{\mathcal{X}}\frac{\mathbf{x}^{\mathrm{T}}\mathbf{x}_1(1-\rho\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}F(\mathrm{d}\mathbf{x}_1)}{\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}_1)^2}\left(\widetilde{\mathbf{\Delta}}^{-1}-\frac{\mathbf{x}\boldsymbol{\mu}^{\mathrm{T}}}{2\mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}}\right)^{\mathrm{T}}$, $\mathbf{x}\in\mathcal{X}(\delta)$
$\widehat{\mathbf{Y}}$	The one-step estimator for the population LSE: $\widehat{\mathbf{Y}}=\{\mathrm{diag}(\widehat{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{X}}\mathbf{1})\}^{-1/2}\widehat{\mathbf{X}}$, where $\widehat{\mathbf{X}}$ is the one-step estimator for $\mathbf{X}$, and $\widetilde{\mathbf{X}}$ is an initial estimator for $\mathbf{X}$
$\widehat{\mathbf{y}}_i$	The ith row of $\widehat{\mathbf{Y}}$ as a column vector in $\mathbb{R}^d$

Table 1: Table of Notations (continued)

$\boldsymbol{\mu}_n$	$\int_{\mathcal{X}} \mathbf{x} F_n(\mathrm{d}\mathbf{x})$, can be equivalently expressed as $(1/n)\sum_{i=1}^n \mathbf{x}_{0i}$
$\widetilde{\mathbf{G}}(\mathbf{x})$	$\dfrac{1}{\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}}\left(\mathbf{I}_d - \dfrac{\mathbf{x}\boldsymbol{\mu}^{\mathrm{T}}}{2\mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}}\right)\mathbf{G}^{-1}(\mathbf{x})\left(\mathbf{I}_d - \dfrac{\mathbf{x}\boldsymbol{\mu}^{\mathrm{T}}}{2\mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}}\right)^{\mathrm{T}}$, $\mathbf{x}\in\mathcal{X}(\delta)$
$\widetilde{\boldsymbol{\Delta}}_n$	$\int_{\mathcal{X}}\left(\dfrac{\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}}{\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_1}\right)F_n(\mathrm{d}\mathbf{x}_1)$, can be equivalently expressed as $\dfrac{1}{n}\sum_{j=1}^n \dfrac{\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0j}}$
$\widetilde{\boldsymbol{\Sigma}}_n(\mathbf{x})$	$\left(\widetilde{\boldsymbol{\Delta}}_n^{-1} - \dfrac{\mathbf{x}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}}\right)\dfrac{1}{n}\sum_{j=1}^n \dfrac{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x})(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}_{0j})^2}\left(\widetilde{\boldsymbol{\Delta}}_n^{-1} - \dfrac{\mathbf{x}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}}\right)^{\mathrm{T}}$, $\mathbf{x}\in\mathcal{X}(\delta)$
$C(F_k, F_l)$	Chernoff information between distributions F_k and F_l, defined by $\sup_{t\in(0,1)}\left\{-\log\int f_k(\mathbf{x})^t f_l(\mathbf{x})^{1-t}\mathrm{d}\mathbf{x}\right\}$, where $F_k(\mathbf{x}) = f_k(\mathbf{x})\mathrm{d}\mathbf{x}$, $F_l(\mathbf{x}) = f_l(\mathbf{x})\mathrm{d}\mathbf{x}$
$\lambda_k(\mathbf{C})$	The kth largest eigenvalue of a (square) positive semidefinite matrix $\mathbf{C}$
$[\mathbf{z}]_k$	The kth coordinate of a Euclidean vector $\mathbf{z}$
$C, C_1, C_2, C_c, c, c_1, c_2, ...$	Generic constants that may change from line to line but is independent of the quantities of interest
$\mathbf{V}_n(\mathbf{x})$	$(1/n)\sum_{j\neq i}\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}$
$\mathbf{V}(\mathbf{x})$	$\int_{\mathcal{X}}\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_1(1-\rho\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)\}\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}F(\mathrm{d}\mathbf{x}_1)$
$\mathbf{P}_0$	$\rho_n\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}$
$\mathbf{E}$	$\mathbf{A} - \rho_n\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}} = \mathbf{A} - \mathbf{P}_0$
$\mathbf{H}$	$(n\rho_n)^{-1/2}(\mathbf{A} - \rho_n\mathbf{X}_0\mathbf{X}_0)^{\mathrm{T}} = (n\rho_n)^{-1/2}\mathbf{E}$
H_{ij}	The (i,j)th element of $\mathbf{H}$
$\mathbf{e}_i$	The standard basis vector in $\mathbb{R}^n$ with all 0 in coordinates except the ith coordinate being 1
$\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}$	The spectral decomposition of $\mathbf{P}_0$, where $\mathbf{S}_{\mathbf{P}} = \mathrm{diag}\{\lambda_1(\mathbf{P}_0),\ldots,\lambda_d(\mathbf{P}_0)\}$, and $\mathbf{U}_{\mathbf{P}}\in\mathbb{O}(n,d)$
$\mathbf{u}_{0k}$	The kth column of $\mathbf{U}_{\mathbf{P}}$, where $k\in[d]$
$\sum_{i=1}^n\widehat{\lambda}_i\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i^{\mathrm{T}}$	The spectral decomposition of $\mathbf{A}$, where $\lvert\widehat{\lambda}_1\rvert\geq\ldots\geq\lvert\widehat{\lambda}_n\rvert\geq 0$, and $\widehat{\mathbf{u}}_i^{\mathrm{T}}\widehat{\mathbf{u}}_j = \mathbb{1}(i=j)$
$\mathbf{U}_{\mathbf{A}}$	$[\widehat{\mathbf{u}}_1,\ldots,\widehat{\mathbf{u}}_d]$
$\mathbf{S}_{\mathbf{A}}$	$\mathrm{diag}\{\lvert\widehat{\lambda}_1\rvert,\ldots,\lvert\widehat{\lambda}_d\rvert\}$

Table 1: Table of Notations (continued)

ν, ν_1, ν_2	A generic positive constant that may change from line to line but does not depend on n
$\mathbf{W}^*$	The product of the left and right singular factor of $\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}$: Let $\mathbf{W}_1\mathbf{S}_{\mathbf{PA}}\mathbf{W}_2^{\mathrm{T}}$ be the singular value decomposition of $\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}$. Then $\mathbf{W}^* = \mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}$
$\mathbf{W}_{\mathbf{X}}$	The orthogonal matrix $\mathbf{W}_{\mathbf{X}} \in \mathbb{O}(d)$ such that $\mathbf{X}_0 = \mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}}$
$M_n(\mathbf{x})$	$\frac{1}{n}\sum_{j\neq i}^{n}\{A_{ij}\log(\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}) + (1-A_{ij})\log(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}$, $\mathbf{x}\in\mathcal{X}(\delta)$
$\widetilde{\mathbf{V}}_n(\mathbf{x})$	$\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x})\mathbf{x}_{0i}\mathbf{x}_{0i}^{\mathrm{T}}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^2}$
$\widetilde{\mathbf{V}}(\mathbf{x})$	$\int_{\mathcal{X}}\frac{\mathbf{x}_1^{\mathrm{T}}\mathbf{x}(1-\rho_n\mathbf{x}_1^{\mathrm{T}}\mathbf{x})\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}}{(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\mu}_n)^2}F(\mathrm{d}\mathbf{x}_1)$
$\mathbf{\Pi}$	A generic permutation matrix
$\boldsymbol{\nu}_k$	A unique latent position in a positive semidefinite stochastic block model
$M(\mathbf{x})$	$\mathbb{E}_0\{M_n(\mathbf{x})\}$
$\mathbf{\Psi}_n(\mathbf{x})$	$\frac{\partial M_n}{\partial\mathbf{x}}(\mathbf{x})$
$\mathbf{\Psi}(\mathbf{x})$	$\mathbb{E}_0\{\mathbf{\Psi}_n(\mathbf{x})\}$
$\bar{M}(\mathbf{x})$	$\lim_{n\to\infty}M(\mathbf{x})$ (note that $M(\mathbf{x})$ depends on n implicitly)
$\dot{\mathbf{\Psi}}_0$	$\frac{1}{n}\sum_{j\neq i}^{n}\frac{\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}$
$\dot{\mathbf{\Psi}}_{n,0}$	$\frac{1}{n}\sum_{j\neq i}^{n}\frac{\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{\widetilde{\mathbf{x}}_{i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\widetilde{\mathbf{x}}_{i}^{\mathrm{T}}\mathbf{x}_{0j})}$
$\mathrm{vec}(\mathbf{\Sigma})$	the vectorization of the matrix $\mathbf{\Sigma}$ defined to be the vector formed by stacking the columns of $\mathbf{\Sigma}$ consecutively
$H_i(\mathbf{X})$	$\frac{1}{n}\sum_{j=1}^{n}\frac{\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}$
$\mathbf{T}$	$\mathrm{diag}(\rho_n\mathbf{P}_0\mathbf{1}) = \mathrm{diag}(\rho_n\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}\mathbf{1})$
$\mathbf{D}$	$\mathrm{diag}(\mathbf{A}\mathbf{1})$
$\widetilde{\mathbf{E}}$	$n\rho_n\{\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}_0)\}$, where $\mathcal{L}(\mathbf{M}) = (\mathrm{diag}(\mathbf{M}\mathbf{1}))^{-1/2}\mathbf{M}(\mathrm{diag}(\mathbf{M}\mathbf{1}))^{-1/2}$
$\sum_{i=1}^{n}\widetilde{\lambda}_i(\widetilde{\mathbf{u}}_{\mathbf{A}})_i(\widetilde{\mathbf{u}}_{\mathbf{A}})_i^{\mathrm{T}}$	The spectral decomposition of $\mathcal{L}(\mathbf{A})$, where $\lvert\widetilde{\lambda}_1\rvert \geq \ldots \geq \lvert\widetilde{\lambda}_n\rvert$, and $(\widetilde{\mathbf{u}}_{\mathbf{A}})_i^{\mathrm{T}}(\widetilde{\mathbf{u}}_{\mathbf{A}})_j = \mathbb{1}(i=j)$
$\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}\widetilde{\mathbf{U}}_{\mathbf{P}}$	The spectral decomposition of $\mathcal{L}(\mathbf{P}_0)$, where $\widetilde{\mathbf{U}}_{\mathbf{P}}\in\mathbb{O}(n,d)$, and

Table 1: Table of Notations (continued)

	$\widetilde{\mathbf{S}}_{\mathbf{P}} = \mathrm{diag}[\lambda_1\{\mathcal{L}(\mathbf{P}_0)\}, \ldots, \lambda_d\{\mathcal{L}(\mathbf{P}_0)\}]$
$\widetilde{\mathbf{u}}_{0k}$	The kth column of $\widetilde{\mathbf{U}}_{\mathbf{P}}$ as a vector in $\mathbb{R}^n$
$\widetilde{u}_{0jk}$	The jth element of the kth column of $\widetilde{\mathbf{U}}_{\mathbf{P}}$
$\widetilde{\mathbf{U}}_{\mathbf{A}}$	$[(\widetilde{\mathbf{u}}_{\mathbf{A}})_1, \ldots, (\widetilde{\mathbf{u}}_{\mathbf{A}})_2]$
$\widetilde{\mathbf{S}}_{\mathbf{A}}$	$\mathrm{diag}(\lvert\widetilde{\lambda}_1\rvert, \ldots, \lvert\widetilde{\lambda}_d\rvert)$
$\lVert\mathbf{Z}\rVert_\infty$	Infinity norm of a matrix $\mathbf{Z} = [Z_{ik}]_{n\times d}$: $\lVert\mathbf{Z}\rVert_\infty = \max_i \sum_k \lvert z_{ik}\rvert$
$\lVert\mathbf{z}\rVert_\infty$	Infinity norm of a vector $\mathbf{z} = [z_1, \ldots, z_n]^{\mathrm{T}}$: $\lVert\mathbf{z}\rVert_\infty = \max_i \lvert z_i\rvert$
$[\mathbf{Z}]_{*k}$	The kth column of a matrix $\mathbf{Z}$ as a column vector
$[\mathbf{Z}]_{i*}$	The ith row of a matrix $\mathbf{Z}$ as a column vector

B Proof of Theorem 1 (Limit Theorem for the ASE)

Theorem 1 is slightly different than those presented in Tang and Priebe (2018) and Athreya et al. (2016), as the latent positions $\mathbf{x}_{01}, \ldots, \mathbf{x}_{0n}$ are deterministic in the current setup. For comparison, we first state the limit theorem for the ASE when the latent positions are independent and identically distributed random variables in Theorem B.1 below, which is originally due to Athreya et al. (2016):

Theorem B.1 (Theorems 2.1 and 2.2 of Tang and Priebe, 2018) *Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{\text{i.i.d.}}{\sim} F$ for some distribution F supported on $\mathcal{X}$, and suppose $\mathbf{A} \mid \mathbf{X} \sim \mathrm{RDPG}(\mathbf{X})$ with a sparsity factor ρ_n. Suppose either $\rho_n \equiv 1$ for all n or $\rho_n \to 0$ but $(\log n)^4/(n\rho_n) \to 0$ as $n \to \infty$, and denote $\rho = \lim_{n\to\infty} \rho_n$. Let $\widehat{\mathbf{X}}^{(\mathrm{ASE})} = [\widehat{\mathbf{x}}_1^{(\mathrm{ASE})}, \ldots, \widehat{\mathbf{x}}_n^{(\mathrm{ASE})}]^{\mathrm{T}}$ be the ASE defined by (2.2). Let $\boldsymbol{\Delta}$ and $\boldsymbol{\Sigma}(\mathbf{x})$ be given as in Theorem 1 and assume that $\boldsymbol{\Delta}$ and $\boldsymbol{\Sigma}(\mathbf{x})$ are strictly positive definite for all $\mathbf{x} \in \mathcal{X}$. Then there exists a sequence of orthogonal matrices $(\mathbf{W}_n)_{n=1}^\infty \in \mathbb{R}^{d\times d}$ such that*

$$\lVert\widehat{\mathbf{X}}^{(\mathrm{ASE})}\mathbf{W}_n - \rho_n^{1/2}\mathbf{X}_0\rVert_{\mathrm{F}}^2 \overset{a.s.}{\to} \int_{\mathcal{X}} \mathrm{tr}\{\boldsymbol{\Sigma}(\mathbf{x})\}F(\mathrm{d}\mathbf{x}),$$

and for any fixed index $i \in [n]$, $\sqrt{n}(\mathbf{W}_n^{\mathrm{T}}\widehat{\mathbf{x}}_i^{(\mathrm{ASE})} - \rho_n^{1/2}\mathbf{x}_{0i})$ converges to a normal mixture distribution with density

$$\int_{\mathcal{X}} \frac{1}{\sqrt{\det\{2\pi\boldsymbol{\Sigma}(\mathbf{x}_1)\}}} \exp\left\{-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathbf{x}_1)^{-1}\mathbf{x}\right\} F(\mathrm{d}\mathbf{x}_1).$$

We breakdown the proof into the proof of the limit(2.3) and the proof of the asymptotic normality of the rows of the ASE (2.4).

B.1 Proof of the Limit (2.3)

Proof of the limit (2.3)**.** The proof of (2.3) is very similar to the proof given in Appendix A of Tang and Priebe (2018) and here we only present the sketch. In the case where $\mathbf{A} \sim \mathrm{RDPG}(\mathbf{X}_0)$ where $\mathbf{X}_0 = [\mathbf{x}_{01}, \ldots, \mathbf{x}_{0n}]^{\mathrm{T}}$, Theorem A.5 in Tang et al. (2017a) yields

$$\|\widehat{\mathbf{X}}^{(\mathrm{ASE})}\mathbf{W} - \rho_n^{1/2}\mathbf{X}_0\|_{\mathrm{F}} = \rho_n^{-1/2}\|(\mathbf{A} - \rho_n\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}})\mathbf{X}_0(\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0)^{-1}\|_{\mathrm{F}} + O_{\mathbb{P}_0}((n\rho_n)^{-1/2}).$$

Denote $\zeta = \rho_n^{-1}\|(\mathbf{A} - \rho_n\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}})\mathbf{X}_0(\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0)^{-1}\|_{\mathrm{F}}^2$. Then Lemma A.5 in Tang et al. (2017a) further shows that $\zeta - \mathbb{E}_0(\zeta)$ is asymptotically negligible. Apppendix A of Tang and Priebe (2018) also shows that

$$\mathbb{E}_0(\zeta) = \mathrm{tr}\left[n(\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0)^{-1}\{n^{-2}\rho_n^{-1}\mathbf{X}_0^{\mathrm{T}}\mathbb{E}_0\{(\mathbf{A} - \rho_n\mathbf{X}_0\mathbf{X}_0)^2\}\mathbf{X}_0\}n(\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0)^{-1}\right],$$

where

$$\mathbb{E}_0\{(\mathbf{A} - \rho_n\mathbf{X}_0\mathbf{X}_0)^2\} = \mathrm{diag}\left\{\sum_{j\neq i}\rho_n\mathbf{x}_{01}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{01}^{\mathrm{T}}\mathbf{x}_{0j}), \ldots, \sum_{j\neq i}\rho_n\mathbf{x}_{0n}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0n}^{\mathrm{T}}\mathbf{x}_{0j})\right\}.$$

In what follows we make use of the condition (2.1) to derive $\mathbb{E}_0(\zeta)$. Condition (2.1) immediately implies that $n(\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0)^{-1} \to \boldsymbol{\Delta}^{-1}$. Furthermore,

$$n^{-2}\rho_n^{-1}\mathbf{X}_0^{\mathrm{T}}\mathbb{E}_0\{(\mathbf{A} - \rho_n\mathbf{X}_0\mathbf{X}_0)^2\}\mathbf{X}_0 = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_{0i}\left\{\frac{1}{n}\sum_{j\neq i}^n \mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\right\}\mathbf{x}_{0i}^{\mathrm{T}}.$$

Define a matrix-valued fucntion $\mathbf{V}_n(\mathbf{x}) = (1/n)\sum_{j\neq i}^n\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}$. Using the argument for proving Lemma D.7, one can also show that $\mathbf{V}_n(\mathbf{x}) \to \mathbf{V}(\mathbf{x})$ uniformly for all $\mathbf{x} \in \bar{\mathcal{X}}$ as well, where $\bar{\mathcal{X}}$ is the closure of $\mathcal{X}$ (which is compact), and

$$\mathbf{V}(\mathbf{x}) = \int_{\mathcal{X}}\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_1(1 - \rho\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)\}\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}F(\mathrm{d}\mathbf{x}_1).$$

Since $F_n = (1/n)\sum_{j=1}^n \delta_{\mathbf{x}_{0i}}$ converges strongly to F, it follows that (see, for example, Exercise 3 in Section 4.4 of Chung, 2001) that

$$\frac{1}{n}\sum_{i=1}^n \mathbf{x}_{0i}\left\{\frac{1}{n}\sum_{j\neq i}^n \mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\right\}\mathbf{x}_{0i}^{\mathrm{T}} = \frac{1}{n}\sum_{i=1}^n \mathbf{V}_n(\mathbf{x}_i) = \int_{\mathcal{X}}\mathbf{V}_n(\mathbf{x}_1)F_n(\mathrm{d}\mathbf{x}_1) \to \int_{\mathcal{X}}\mathbf{V}(\mathbf{x}_1)F(\mathrm{d}\mathbf{x}).$$

Hence we conclude that

$$
\begin{aligned}
\mathbb{E}_0(\zeta) &= \mathrm{tr}\left[n(\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0)^{-1}\{n^{-2}\rho_n^{-1}\mathbf{X}_0^{\mathrm{T}}\mathbb{E}_0\{(\mathbf{A}-\rho_n\mathbf{X}_0\mathbf{X}_0)^2\}\mathbf{X}_0\}n(\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0)^{-1}\right] \\
&\to \mathrm{tr}\left\{\mathbf{\Delta}^{-1}\int_{\mathcal{X}}\mathbf{V}(\mathbf{x}_1)F(\mathrm{d}\mathbf{x}_1)\mathbf{\Delta}^{-1}\right\} \\
&= \int_{\mathcal{X}}\mathrm{tr}\left\{\mathbf{\Delta}^{-1}\mathbf{V}(\mathbf{x})\mathbf{\Delta}^{-1}\right\}F(\mathrm{d}\mathbf{x}) \\
&= \int_{\mathcal{X}}\mathrm{tr}\left[\mathbf{\Delta}^{-1}\int_{\mathcal{X}}\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_1(1-\rho\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)\}\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}F(\mathrm{d}\mathbf{x}_1)\mathbf{\Delta}^{-1}\right]F(\mathrm{d}\mathbf{x}) \\
&= \int_{\mathcal{X}}\mathrm{tr}\{\mathbf{\Sigma}(\mathbf{x})\}F(\mathrm{d}\mathbf{x}).
\end{aligned}
$$

□

B.2 Proof of the Asymptotic Normality (2.4)

The proof of the asymptotic normality (2.4) is enormously different than that presented in Appendix A in Tang and Priebe (2018) and Athreya et al. (2016). Here we follow the proof strategy adopted in Cape et al. (2019), which addresses the asymptotic normality of the rows of eigenvectors of random matrices. To this end, we need the following technical lemma.

Lemma B.1 *Let* $\mathbf{P}_0 = \rho_n\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}$ *and* $\mathbf{E} = \mathbf{A} - \mathbf{P}_0$. *Then for deterministic vector* $\mathbf{v} \in \mathbb{R}$, *any* $k = 1, \ldots, \lceil \log n \rceil$, *and* $p \leq \lceil (\log n)^2 \rceil$, *there exists a constant* $C_{\mathbf{E}} > 0$ *such that*

$$
\mathbb{E}_0\left(|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^k\mathbf{v}|^p\right) \leq (n\rho_n)^{kp/2}(2kp)^{kp}\|\mathbf{v}\|_\infty^p.
$$

Proof. The proof is very similar to that of Lemma 5.4 in Mao et al. (2020), which originates from the proof of Lemma 7.10 in Erdös et al. (2013). Denote $\mathbf{H} = (n\rho_n)^{-1/2}\mathbf{E} = (n\rho_n)^{-1/2}(\mathbf{A} - \mathbf{P}_0)$, and let H_{ij} be the (i,j)th entry of $\mathbf{H}$. The proof is based on the following two observations:

- $\mathbb{E}_0(|H_{ij}|^m) \leq 1/n$ for all $m \geq 2$. This is because $n\rho_n \to \infty$, implying that

  $$
  |H_{ij}| = \frac{|A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}|}{\sqrt{n\rho_n}} \leq \frac{1}{\sqrt{n\rho_n}} \leq 1.
  $$

  Therefore, for any $m \geq 2$

  $$
  \mathbb{E}_0(|H_{ij}|^m) \leq \mathbb{E}_0(H_{ij}^2) = \frac{1}{n\rho_n}\mathbb{E}_0\{(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})^2\} \leq \frac{1}{n}.
  $$

- To compute an upper bound for the pth moment of $|\mathbf{e}_i^{\mathrm{T}}\mathbf{H}^k\mathbf{v}|$, the authors of Erdös et al. (2013) use a multigraph construction technique by partition the summation indices into equivalent classes according to whether the indexed variables in the same equivalent class are the same or not. Then the number of summand is further upper bounded by the use of a spanning tree of the associated multigraph. Since the upper bound there uses the moment of the absolute values of H_{ij}'s and the vector $\mathbf{v}$ is deterministic, we directly apply the proof there to the upper bound yields

$$E(|\mathbf{e}_i^{\mathrm{T}}\mathbf{H}^k\mathbf{v}|^p) \leq (2kp)^{kp}\|\mathbf{v}\|_\infty^p.$$

The factor 2 comes from the fact that $\mathbf{H}$ is diagonal and only the upper triangular part of $\mathbf{H}$ are independent random variables.

Since $\mathbf{H} = (n\rho_n)^{-1/2}\mathbf{E}$, it follows that

$$\mathbb{E}_0\left(|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^k\mathbf{v}|^p\right) \leq (n\rho_n)^{kp/2}(2kp)^{kp}\|\mathbf{v}\|_\infty^p.$$

The proof is thus completed. □

Proof of the asymptotic normality (2.4). Denote $\mathbf{E} = \mathbf{A} - \mathbf{P}_0 = \mathbf{A} - \rho_n\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}$. Let $\mathbf{A} = \sum_{i=1}^n \widehat{\lambda}_i\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i^{\mathrm{T}}$ be the spectral decomposition of $\mathbf{A}$ with $|\widehat{\lambda}_1| \geq \ldots \geq |\widehat{\lambda}_n|$. Denote $\mathbf{S}_\mathbf{A} = \mathrm{diag}(|\widehat{\lambda}_1|,\ldots,|\widehat{\lambda}_d|)$ and $\mathbf{U}_\mathbf{A} = [\widehat{\mathbf{u}}_1,\ldots,\widehat{\mathbf{u}}_d]$. Similarly, let $\mathbf{P}_0 = \mathbf{U}_\mathbf{P}\mathbf{S}_\mathbf{P}\mathbf{U}_\mathbf{P}^{\mathrm{T}}$ be the spectral decomposition of $\mathbf{P}_0$, where $\mathbf{U}_\mathbf{P} \in \mathbb{O}(n,d)$ and $\mathbf{S}_\mathbf{P} = \mathrm{diag}\{\lambda_1(\mathbf{P}_0),\ldots,\lambda_d(\mathbf{P}_0)\}$. Let $\mathbf{U}_\mathbf{P} = [\mathbf{u}_{01},\ldots,\mathbf{u}_{0d}]$. In order to prove (2.4), we first list several useful facts:

(a) For any $c > 0$, there exists $C_1 > 0$ such that $\|\mathbf{A} - \mathbf{P}\|_2 \leq C_1(n\rho_n)^{1/2}$ with probability at least $1 - n^{-c}$, and there exists constant $C_2 > 0$ such that $\|\mathbf{A} - \mathbf{P}\|_2 \leq C_2(n\rho_n)^{1/2}\log n$ with probability at least $1 - \exp\{-c(\log n)^2\}$. The first result is obtained from Lei and Rinaldo (2015), and the second result is a consequence of the matrix Bernstein's inequality.

(b) For any $c > 0$, there exists $C > 0$ such that $\|\mathbf{S}_\mathbf{A}\|_2 \leq Cn\rho_n$ and $\|\mathbf{S}_\mathbf{A}^{-1}\|_2 \leq C(n\rho_n)^{-1}$ with probability at least $1 - \exp\{-c(\log n)^2\}$. This fact can be implied by fact (a) and Weyl's inequality. Furthermore, deterministically, $\|\mathbf{S}_\mathbf{P}\|_2 \lesssim n\rho_n$ and $\|\mathbf{S}_\mathbf{P}^{-1}\|_2 \lesssim (n\rho_n)^{-1}$. This result follows from the fact that the smallest eigenvalue of $\mathbf{S}_\mathbf{P}$ is the same as the smallest eigenvalue of $\rho_n(\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0)$, and $(1/n)(\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0) \to \boldsymbol{\Delta}$ for some deterministic positive definite matrix $\boldsymbol{\Delta}$.

(c) For any $c > 0$, there exists $C > 0$ such that $\|\mathbf{U}_\mathbf{P}^{\mathrm{T}}\mathbf{E}\mathbf{U}_\mathbf{P}\|_{\mathrm{F}} \leq C\log n$ with probability at least $1 - n^{-c}$, and $\|\mathbf{U}_\mathbf{P}^{\mathrm{T}}\mathbf{E}\mathbf{U}_\mathbf{P}\|_{\mathrm{F}} = O_{\mathbb{P}_0}(1)$. This follows from the union bound and Hoeffding's inequality (see equation (50) in Athreya et al., 2018).

(d) There exists some constant $c > 0$, such that for any fixed $i \in [n]$,

$$\mathbb{P}\left(\bigcup_{k=1}^d \left\{|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^2\mathbf{u}_{0k}| > (n\rho_n)(\log n)^2\|\mathbf{u}_{0k}\|_\infty\right\}\right) \leq \exp\left(-c\log n\right).$$

This result can be derived as follows: By Lemma B.1 and Markov's inequality, with $p = \lfloor(\log n)/8\rfloor$, for every $k = 1,\ldots,d$, we have

$$\begin{aligned}\mathbb{P}_0\left(|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^2\mathbf{u}_{0k}| > (n\rho_n)(\log n)^2\|\mathbf{u}_{0k}\|_\infty\right) &\leq \frac{(4p)^{2p}(n\rho_n)^p\|\mathbf{u}_{0k}\|_\infty^p}{(n\rho_n)^p(\log n)^{2p}\|\mathbf{u}_{0k}\|_\infty^p}\\ &= \left(\frac{4p}{\log n}\right)^{2p} = \exp\left(-c\log n\right)\end{aligned}$$

for some constant $c > 0$. The result then follows from the union bound applied to $k \in \{1, \dots, d\}$.

(e) There exists a constant $\nu > 0$, such that for all $\xi \in (1, 2]$,

$$\mathbb{P}_0\left(\bigcup_{t=1}^{\lceil \log n\rceil}\bigcup_{i=1}^{n}\bigcup_{k=1}^{d}\left\{|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^t\mathbf{u}_{0k}| > (n\rho_n)^{t/2}(\log n)^{\xi t}\|\mathbf{u}_{0k}\|_\infty\right\}\right) \leq \exp\{-\nu(\log n)^\xi\}.$$

The derivation is similar to (d), but for each fixed $i \in [n]$, $k \in [d]$, and $t = 1, \dots, \lceil \log n\rceil$, we replace the choice of $p = \lfloor(\log n)/8\rfloor$ by $\lfloor(\log n)^\xi/(4t)\rfloor$, and invoke Lemma B.1 to derive

$$\begin{aligned}\mathbb{P}_0\left(|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^t\mathbf{u}_{0k}| > (n\rho_n)^{t/2}(\log n)^{\xi t}\|\mathbf{u}_{0k}\|_\infty\right) &\leq \frac{(n\rho_n)^{pt/2}(2pt)^{pt}\|\mathbf{u}_{0k}\|_\infty^p}{(n\rho_n)^{pt/2}(\log n)^{\xi tp}\|\mathbf{u}_{0k}\|_\infty^p}\\ &= \left\{\frac{2pt}{(\log n)^\xi}\right\}^{pt} = \exp\left\{-c(\log n)^\xi\right\}.\end{aligned}$$

Since $\xi > 1$ and $(\log n)^\xi \gg \log n$, taking the union bound over $i \in [n]$, $t = 1, \dots, \lceil \log n\rceil$, and $k \in [d]$ leads to the desired result.

(f) $\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty} \lesssim n^{-1/2}$. This is because $\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty} = \|\rho_n^{-1/2}\mathbf{X}_0\mathbf{S}_{\mathbf{P}}^{-1/2}\|_{2\to\infty} \leq \rho_n^{-1/2}\|\mathbf{X}_0\|_{2\to\infty}\|\mathbf{S}_{\mathbf{P}}^{-1/2}\| \lesssim n^{-1/2}$.

Let $\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}$ have singular value decomposition $\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}} = \mathbf{W}_1\mathbf{S}_{\mathbf{PA}}\mathbf{W}_2^{\mathrm{T}}$, and let $\mathbf{W}^* = \mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}$. We begin the proof by observing the following decomposition of $\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*$ originating from Cape et al. (2019):

$$\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^* = \mathbf{E}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}\mathbf{W}^* + \mathbf{R},$$

where $\mathbf{R} = \mathbf{R}^{(1)} + \mathbf{R}^{(2)} + \mathbf{R}_2^{(1)} + \mathbf{R}_2^{(2)} + \mathbf{R}_2^{(\infty)}$,

$$\begin{aligned}&\mathbf{R}^{(1)} = \mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}\mathbf{R}^{(3)}, \quad \mathbf{R}^{(3)} = \mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{-1} - \mathbf{S}_{\mathbf{P}}^{-1}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}},\\ &\mathbf{R}^{(2)} = \mathbf{U}_{\mathbf{P}}(\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}} - \mathbf{W}^*), \quad \mathbf{R}_2^{(1)} = \mathbf{E}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}(\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{-2} - \mathbf{S}_{\mathbf{P}}^{-2}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}),\\ &\mathbf{R}_2^{(2)} = \mathbf{E}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}(\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}} - \mathbf{W}^*), \quad \mathbf{R}_2^{(\infty)} = \sum_{t=2}^{\infty}\mathbf{E}^t\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{-(t+1)}.\end{aligned}$$

The above derivation is directly obtained from the proof of Theorem 2 in Cape et al. (2019), and the key idea is that the spectra of $\mathbf{S}_{\mathbf{A}}$ and $\mathbf{E}$ are disjoint with high probability, such that one can apply the von Neumann trick to write

$$\mathbf{U}_{\mathbf{A}} = \sum_{t=0}^{\infty}\mathbf{E}^t\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{-(t+1)}$$

Also see (S3) in the Supplementary Material of Cape et al. (2019).

Let $\xi \in (1, 2]$ be a constant to be specified later. We now fix the row index $i \in [n]$ and show that

$$\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*)\|_2 = O_{\mathbb{P}_0}\left\{\frac{(\log n)^2}{\sqrt{n}}\right\}. \tag{B.1}$$

This can be done by the establishing the following results:

(1) $\|\mathbf{E}\mathbf{R}^{(1)}\|_{2\to\infty} = o_{\mathbb{P}_0}(n^{-1/2})$. In fact, following Cape et al. (2019), we see that

$$\|\mathbf{R}^{(3)}\|_2 \leq d\|\mathbf{S}_{\mathbf{P}}^{-1}\|_2\|\mathbf{S}_{\mathbf{A}}^{-1}\|_2\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}} - \mathbf{S}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\|_2 = O_{\mathbb{P}_0}\{(n\rho_n)^{-2}(\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}}\|_2 + 1)\}.$$

By result (e) with $t = 1$, we have $\|\mathbf{E}\mathbf{U}_{\mathbf{P}}\|_{2\to\infty} = O_{\mathbb{P}_0}((n\rho)^{1/2}(\log n)^{\xi}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty})$. Therefore, by results (b), (c), and (f), we obtain

$$\begin{aligned}\|\mathbf{E}\mathbf{R}^{(1)}\|_{2\to\infty} &\leq \|\mathbf{E}\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{S}_{\mathbf{P}}\|_2\|\mathbf{R}^{(3)}\|_2 = O_{\mathbb{P}_0}\left\{\frac{(\log n)^{\xi}(\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{P}}\|_2 + 1)\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}}\right\}\\ &= \frac{1}{\sqrt{n}}O_{\mathbb{P}_0}\left\{\frac{(\log n)^{\xi}}{(n\rho_n)^{1/2}}\right\} = o_{\mathbb{P}_0}(n^{-1/2})\end{aligned}$$

(2) $\|\mathbf{E}\mathbf{R}^{(2)}\|_{2\to\infty} = o_{\mathbb{P}_0}(n^{-1/2})$. Since $\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}} - \mathbf{W}^*\|_2 = O_{\mathbb{P}_0}((n\rho_n)^{-1})$, which is a simple consequence of the Davis-Kahan theorem (Yu et al., 2015), it follows from result (e) with $t = 1$ and result (f) that

$$\|\mathbf{E}\mathbf{R}^{(2)}\|_{2\to\infty} \leq \|\mathbf{E}\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}} - \mathbf{W}^*\|_2 = O_{\mathbb{P}_0}\left((n\rho_n)^{1/2}(\log n)^{\xi}n^{-1/2}(n\rho_n)^{-1}\right) = o_{\mathbb{P}_0}(n^{-1/2}).$$

(3) $\mathbf{e}_i^{\mathrm{T}}\mathbf{E}\mathbf{R}_2^{(1)} = o_{\mathbb{P}_0}(n^{-1/2})$. By the results (d) and (f), we have

$$\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^2\mathbf{U}_{\mathbf{P}}\|_2 = O_{\mathbb{P}_0}\{n^{-1/2}(n\rho_n)(\log n)^2\}.$$

From Cape et al. (2019), we have that

$$\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{-2} - \mathbf{S}_{\mathbf{P}}^{-2}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\|_2 = O_{\mathbb{P}_0}((n\rho_n)^{-5/2}).$$

Therefore,

$$\begin{aligned}\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}\mathbf{R}_2^{(1)}\|_2 &\leq \|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^2\mathbf{U}_{\mathbf{P}}\|_2\|\mathbf{S}_{\mathbf{P}}\|_2\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{-2} - \mathbf{S}_{\mathbf{P}}^{-2}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\|_2\\ &= O_{\mathbb{P}_0}\{n^{-1/2}(n\rho_n)(\log n)^2\}O_{\mathbb{P}_0}((n\rho_n)^{-3/2}) = O_{\mathbb{P}_0}\left\{\frac{(\log n)^2}{\sqrt{n(n\rho_n)}}\right\} = o_{\mathbb{P}_0}(n^{-1/2}).\end{aligned}$$

(4) $\|\mathbf{E}\mathbf{R}_2^{(2)}\|_{2\to\infty} = o_{\mathbb{P}_0}(1)$. By result (e) with $t = 2$, the fact that $\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}} - \mathbf{W}^*\|_2 = O_{\mathbb{P}_0}((n\rho_n)^{-1})$, the results (b) and (f), we obtain

$$\begin{aligned}\|\mathbf{E}\mathbf{R}_2^{(2)}\|_{2\to\infty} &\leq \|\mathbf{E}^2\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{S}_{\mathbf{P}}^{-1}\|_2\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}} - \mathbf{W}^*\|_2\\ &= O_{\mathbb{P}_0}\left\{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}(\log n)^{2\xi}(n\rho_n)(n\rho_n)^{-1}(n\rho_n)^{-1}\right\}\\ &= O_{\mathbb{P}_0}\left\{n^{-1/2}\frac{(\log n)^{2\xi}}{n\rho_n}\right\} = o_{\mathbb{P}_0}(n^{-1/2}).\end{aligned}$$

(5) $\|\mathbf{E}\mathbf{R}_2^{(\infty)}\|_{2\to\infty} = O_{\mathbb{P}_0}\{n^{-1/2}(\log n)^2\}$. Denote $t(n) = 2\lceil(\log n)/(\log n\rho_n)\rceil$. Clearly, $t(n) \ll \log n$ since

$n\rho_n \to \infty$. Write

$$\begin{aligned}\|\mathbf{E}\mathbf{R}_2^{(\infty)}\|_{2\to\infty} &= \left\|\sum_{t=2}^{\infty}\mathbf{E}^{t+1}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{-(t+1)}\right\|_{2\to\infty}\\
&\le \sum_{t=2}^{t(n)+1}\|\mathbf{E}^{t+1}\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{S}_{\mathbf{P}}\|_2\|\mathbf{S}_{\mathbf{A}}^{-1}\|_2^{t+1} + \sum_{t=t(n)+2}^{\infty}\|\mathbf{E}\|_2^{t+1}\|\mathbf{S}_{\mathbf{P}}\|_2\|\mathbf{S}_{\mathbf{A}}^{-1}\|_2^{t+1}.\end{aligned}$$

By the results (b) and (e), for any $c > 0$, there exist constant $C_{\mathbf{E}} > 0$, such that with probability at least $1 - n^{-c}$, we have

$$\begin{aligned}\sum_{t=2}^{t(n)+1}\|\mathbf{E}^{t+1}\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{S}_{\mathbf{P}}\|_2\|\mathbf{S}_{\mathbf{A}}^{-1}\|_2^{t+1} &\le \sum_{t=3}^{t(n)+2}\|\mathbf{E}^{t}\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{S}_{\mathbf{P}}\|_2\|\mathbf{S}_{\mathbf{A}}^{-1}\|_2^{t}\\
&\le \sum_{t=3}^{\infty}C_{\mathbf{E}}^t(n\rho_n)^{t/2}(\log n)^{t\xi}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{S}_{\mathbf{P}}\|_2(n\rho_n)^{-t}\\
&\lesssim \frac{(n\rho_n)}{\sqrt{n}}\sum_{t=3}^{\infty}\{C_{\mathbf{E}}(n\rho_n)^{-1/2}(\log n)^{\xi}\}^t\\
&\lesssim \frac{n\rho_n}{\sqrt{n}}\frac{(\log n)^{3\xi}C_{\mathbf{E}}^3}{(n\rho_n)^{3/2}} \asymp \frac{(\log n)^{3\xi}}{\sqrt{n}(n\rho_n)^{1/2}}.\end{aligned}$$

For the second infinite sum, by the results (a) and (b), with probability at least $1 - n^{-c}$, we directly compute

$$\begin{aligned}\sum_{t=t(n)+2}^{\infty}\|\mathbf{E}\|_2^{t+1}\|\mathbf{S}_{\mathbf{P}}\|_2\|\mathbf{S}_{\mathbf{A}}^{-1}\|_2^{t+1} &\lesssim \sum_{t=t(n)+2}^{\infty} n\rho_n\{C(n\rho_n)^{1/2}(n\rho_n)^{-1}\}^{t+1}\\
&= n\rho_n\sum_{t=t(n)+2}^{\infty}\{C(n\rho_n)^{-1/2}\}^{t+1}\\
&= \exp\left[\{t(n)+3\}\log C - \frac{t(n)+3}{2}\log(n\rho_n) + \log(n\rho_n)\right]\\
&\le \exp\left[\left\{2\left\lceil\frac{\log n}{\log(n\rho_n)}\right\rceil + 3\right\}\log C - \frac{1}{2}\log(n\rho_n) - \log n\right]\\
&\lesssim \exp\left\{-\frac{1}{2}\log(n\rho_n) - \frac{1}{2}\log n\right\} = o(n^{-1/2}).\end{aligned}$$

Therefore, with $\xi = 4/3$, we conclude that

$$\|\mathbf{E}\mathbf{R}_2^{(\infty)}\|_{2\to\infty} = O_{\mathbb{P}_0}\left\{\frac{1}{\sqrt{n}}\frac{(\log n)^4}{(n\rho_n)^{1/2}}\right\} = O_{\mathbb{P}_0}\left\{\frac{(\log n)^2}{\sqrt{n}}\frac{(\log n)^2}{(n\rho_n)^{1/2}}\right\} = O_{\mathbb{P}_0}\left\{\frac{(\log n)^2}{\sqrt{n}}\right\}.$$

(6) $\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^2\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}\mathbf{W}^*\|_2 = O_{\mathbb{P}_0}\{n^{-1/2}(\log n)^2\}$. Since

$$\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^2\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}\mathbf{W}^*\|_2 \le \|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^2\mathbf{U}_{\mathbf{P}}\|_2\|\mathbf{S}_{\mathbf{P}}^{-1}\|_2 \lesssim (n\rho_n)^{-1}\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^2\mathbf{U}_{\mathbf{P}}\|_2,$$

then by result (d), with probability going to 1, and result (f),

$$\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^2\mathbf{U_P}\mathbf{S_P}\mathbf{W}^*\|_2 \lesssim (n\rho_n)^{-1}(n\rho_n)(\log n)^2\|\mathbf{U_P}\|_{2\to\infty} = O_{\mathbb{P}_0}\left\{\frac{(\log n)^2}{\sqrt{n}}\right\}.$$

We are finally in a position to prove the asymptotic normality (2.4) of the ith row of the ASE. Let $\mathbf{W_X}$ be the orthogonal matrix such that $\mathbf{X}_0 = \mathbf{U_P}\mathbf{S_P}^{1/2}\mathbf{W_X}$, and let $\mathbf{W} = (\mathbf{W}^*)^{\mathrm{T}}\mathbf{W_X}$. Following the derivation in Appendix A.1 and A.2 in Athreya et al. (2018), we obtain the following decomposition:

$$\begin{aligned}\sqrt{n}\mathbf{e}_i^{\mathrm{T}}\{\widehat{\mathbf{X}}^{(\mathrm{ASE})}\mathbf{W} - \rho_n^{1/2}\mathbf{X}_0\} &= \sqrt{n}\mathbf{e}_i^{\mathrm{T}}(\mathbf{A}-\mathbf{P})\mathbf{U_P}\mathbf{S_P}^{-1/2}\mathbf{W_X} + \sqrt{n}\mathbf{e}_i^{\mathrm{T}}\mathbf{E}\mathbf{U_P}(\mathbf{W}^*\mathbf{S_A}^{-1/2} - \mathbf{S_P}^{-1/2}\mathbf{W}^*)\mathbf{W}\\ &\quad - \sqrt{n}\mathbf{e}_i^{\mathrm{T}}\mathbf{U_P}(\mathbf{U_P}^{\mathrm{T}}\mathbf{E}\mathbf{U_P})\mathbf{W}^*\mathbf{S_A}^{-1/2}\mathbf{W} + \sqrt{n}\mathbf{e}_i^{\mathrm{T}}\mathbf{E}(\mathbf{U_A} - \mathbf{U_P}\mathbf{W}^*)\mathbf{S_A}^{-1/2}\mathbf{W}\\ &\quad - \sqrt{n}\mathbf{e}_i^{\mathrm{T}}\mathbf{U_P}\mathbf{U_P}^{\mathrm{T}}\mathbf{E}(\mathbf{U_A} - \mathbf{U_P}\mathbf{W}^*)\mathbf{S_A}^{-1/2}\mathbf{W} + \sqrt{n}\mathbf{e}_i^{\mathrm{T}}(\mathbf{R}_1\mathbf{S_A}^{1/2} + \mathbf{U_P}\mathbf{R}_2),\end{aligned}$$

where

$$\mathbf{R}_1 = \mathbf{U_P}(\mathbf{U_P}^{\mathrm{T}}\mathbf{U_A} - \mathbf{W}^*),\quad \mathbf{R}_2 = \mathbf{W}^*\mathbf{S_A}^{1/2} - \mathbf{S_P}^{1/2}\mathbf{W}^*.$$

By Appendix A in Tang and Priebe (2018),

$$\{\sqrt{n}\mathbf{e}_i^{\mathrm{T}}(\mathbf{A}-\mathbf{P})\mathbf{U_P}\mathbf{S_P}^{-1/2}\mathbf{W_X}\}^{\mathrm{T}} \xrightarrow{\mathcal{L}} \mathrm{N}(\mathbf{0}, \mathbf{\Sigma}(\mathbf{x}_{0i})).$$

It suffices to argue that the remainders are asymptotically negligible.

(7) $\sqrt{n}\|\mathbf{E}\mathbf{U_P}(\mathbf{W}^*\mathbf{S_A}^{-1/2} - \mathbf{S_P}^{-1/2}\mathbf{W}^*)\mathbf{W}\|_{2\to\infty} = o_{\mathbb{P}_0}(1)$. By a similar argument for proving Lemma 49 of Athreya et al. (2018), we have

$$\|\mathbf{W}^*\mathbf{S_A}^{-1/2} - \mathbf{S_P}^{-1/2}\mathbf{W}^*\| = O_{\mathbb{P}_0}\left\{(n\rho_n)^{-3/2}\log n\right\}.$$

Furthermore, by result (e) with $t = 1$ and result (f),

$$\|\mathbf{E}\mathbf{U_P}\|_{2\to\infty} = O_{\mathbb{P}_0}\{n^{-1/2}(n\rho_n)^{1/2}(\log n)^{4/3}\}.$$

Therefore,

$$\begin{aligned}\sqrt{n}\|\mathbf{E}\mathbf{U_P}(\mathbf{W}^*\mathbf{S_A}^{-1/2} - \mathbf{S_P}^{-1/2}\mathbf{W}^*)\mathbf{W}\|_{2\to\infty} &\le O_{\mathbb{P}_0}\{(n\rho_n)^{1/2}(\log n)^{4/3}\}O_{\mathbb{P}_0}\left\{(n\rho_n)^{-3/2}\log n\right\}\\ &= O_{\mathbb{P}_0}\{(n\rho_n)^{-1}(\log n)^{7/3}\} = o_{\mathbb{P}_0}(1).\end{aligned}$$

(8) $\sqrt{n}\|\mathbf{U_P}(\mathbf{U_P}^{\mathrm{T}}\mathbf{E}\mathbf{U_P})\mathbf{W}^*\mathbf{S_A}^{-1/2}\mathbf{W}\|_{2\to\infty} = o_{\mathbb{P}_0}(1)$. By the results (b), (c), and (f), we have

$$\begin{aligned}\sqrt{n}\|\mathbf{U_P}(\mathbf{U_P}^{\mathrm{T}}\mathbf{E}\mathbf{U_P})\mathbf{W}^*\mathbf{S_A}^{-1/2}\mathbf{W}\|_{2\to\infty} &\le \sqrt{n}\|\mathbf{U_P}\|_{2\to\infty}\|\mathbf{U_P}^{\mathrm{T}}\mathbf{E}\mathbf{U_P}\|_2\|\mathbf{S_A}^{-1/2}\|_2\\ &\lesssim \|\mathbf{U_P}^{\mathrm{T}}\mathbf{E}\mathbf{U_P}\|_2\|\mathbf{S_A}^{-1/2}\|_2 = O_{\mathbb{P}_0}(\log n)O_{\mathbb{P}_0}((n\rho_n)^{-1/2}) = o_{\mathbb{P}_0}(1).\end{aligned}$$

(9) $\sqrt{n}\mathbf{e}_i^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*)\mathbf{S}_{\mathbf{A}}^{-1/2}\mathbf{W} = o_{\mathbb{P}_0}(1)$. By the previously claimed result (B.1), we have

$$\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*)\|_2 = O_{\mathbb{P}_0}\left\{\frac{(\log n)^2}{\sqrt{n}}\right\}.$$

Together with result (b), we obtain

$$\begin{aligned}\|\sqrt{n}\mathbf{e}_i^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*)\mathbf{S}_{\mathbf{A}}^{-1/2}\mathbf{W}\|_2 &\leq \sqrt{n}\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*)\|_2\|\mathbf{S}_{\mathbf{A}}^{-1/2}\|_2\\ &= O_{\mathbb{P}_0}\left\{\frac{(\log n)^2}{\sqrt{n\rho_n}}\right\} = o_{\mathbb{P}_0}(1).\end{aligned}$$

(10) $\|\sqrt{n}\mathbf{U}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*)\mathbf{S}_{\mathbf{A}}^{-1/2}\mathbf{W}\|_{2\to\infty} = o_{\mathbb{P}_0}(1)$. By the Davis-Kahan theorem and results (a) and (b), $\|\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*\|_2 = O_{\mathbb{P}_0}((n\rho_n)^{-1/2})$. Hence,

$$\begin{aligned}\|\sqrt{n}\mathbf{U}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*)\mathbf{S}_{\mathbf{A}}^{-1/2}\mathbf{W}\|_{2\to\infty} &\leq \sqrt{n}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{E}\|_2\|\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*\|_2\|\mathbf{S}_{\mathbf{A}}^{-1/2}\|_2\\ &= O_{\mathbb{P}_0}\{(n\rho_n)^{1/2}(n\rho_n)^{-1/2}(n\rho_n)^{-1/2}\} = o_{\mathbb{P}_0}(1).\end{aligned}$$

(11) $\|\sqrt{n}(\mathbf{R}_1\mathbf{S}_{\mathbf{A}}^{1/2} + \mathbf{U}_{\mathbf{P}}\mathbf{R}_2)\|_{2\to\infty} = o_{\mathbb{P}_0}(1)$. By Lemma 49 of Athreya et al. (2018) (which also holds for deterministic $\mathbf{x}_{01}, \ldots, \mathbf{x}_{0n}$), $\|\mathbf{R}_2\|_{\mathrm{F}} = O_{\mathbb{P}_0}((n\rho)^{-1/2}\log n)$. Hence,

$$\sqrt{n}\|\mathbf{U}_{\mathbf{P}}\mathbf{R}_2\|_{2\to\infty} \leq \sqrt{n}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{R}_2\|_2 = O_{\mathbb{P}_0}((n\rho_n)^{-1/2}\log n) = o_{\mathbb{P}_0}(1).$$

In addition, we obtain

$$\sqrt{n}\|\mathbf{R}_1\mathbf{S}_{\mathbf{A}}^{1/2}\|_{2\to\infty} \leq \sqrt{n}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}\|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}} - \mathbf{W}^*\|_2\|\mathbf{S}_{\mathbf{A}}^{1/2}\|_2 \lesssim \|\mathbf{U}_{\mathbf{P}}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}} - \mathbf{W}^*\|_2\|\mathbf{S}_{\mathbf{A}}^{1/2}\|_2 = O_{\mathbb{P}_0}((n\rho_n)^{-1})$$

by result (b).

The proof is completed by summarizing the above results. □

C Proof of Theorems 2 and 3 (Estimating One Latent Position)

C.1 Proofs of Theorem 2

Before presenting the proof of Theorem 2, we need to establish the following real analysis result. We note that this result may not be completely original and can be proved using a standard real analysis argument.

Lemma C.1 *Let $(f_n)_{n=1}^{\infty} \subset C^2(D)$ be a sequence of functions in $C^2(D)$, where D is a convex compact subset of $\mathbb{R}^d$ with non-empty interior, and $C^2(D)$ is the class of twice continuously differentiable functions on D. Let $f \in C^2(D)$ as well. Assume the following conditions hold:*

(i) f_n converges uniformly to f within D, i.e., $\sup_{\mathbf{x}\in D}|f_n(\mathbf{x}) - f(\mathbf{x})| \to 0$ *as $n \to \infty$;*

(ii) f_n *and* f *both have continuous Hessians*

$$\frac{\partial^2 f_n}{\partial \mathbf{x} \partial \mathbf{x}^{\mathrm{T}}}(\mathbf{x}), \quad \frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}^{\mathrm{T}}}(\mathbf{x})$$

and the negative Hessians are positive definite for all $\mathbf{x} \in D$*.*

Let $\widehat{\mathbf{x}}_n$ *be the unique maximizer of* f_n*, and* $\widehat{\mathbf{x}}_0$ *be the unique maximizer of* f*. Then:*

(a) The sequence $(\widehat{\mathbf{x}}_n)_{n=1}^{\infty}$ *converges to the unique maximizer* $\widehat{\mathbf{x}}_0$ *of the limit function* f*;*

(b) For all $\epsilon > 0$*, there exists a positive* $\eta = \eta(\epsilon)$ *that does not depend on* n*, and a positive integer* $N = N(\epsilon)$ *that only depends on* ϵ*, such that for all* $n \geq N$*, and all* $\mathbf{x}$ *satisfying* $\|\mathbf{x} - \widehat{\mathbf{x}}_n\|_2 > \epsilon$*,*

$$f_n(\widehat{\mathbf{x}}_n) - f_n(\mathbf{x}) \geq \eta(\epsilon)$$

holds.

Proof. For conclusion (a) we take $(\widehat{\mathbf{x}}_{n_k})_{k=1}^{\infty}$ to be any subsequence of $(\widehat{\mathbf{x}}_n)_{n=1}^{\infty}$ that converges. Suppose $\widehat{\mathbf{x}}_{n_k} \to \widehat{\mathbf{x}}_0^*$ as $n \to \infty$. It follows that

$$\begin{aligned} |f_{n_k}(\widehat{\mathbf{x}}_{n_k}) - f(\widehat{\mathbf{x}}_0^*)| &\leq |f_{n_k}(\widehat{\mathbf{x}}_{n_k}) - f(\widehat{\mathbf{x}}_{n_k})| + |f(\widehat{\mathbf{x}}_{n_k}) - f(\widehat{\mathbf{x}}_0^*)| \\ &\leq \sup_{\mathbf{x} \in D} |f_{n_k}(\mathbf{x}) - f(\mathbf{x})| + |f(\widehat{\mathbf{x}}_{n_k}) - f(\widehat{\mathbf{x}}_0^*)|. \end{aligned}$$

The sequence $(\sup_{\mathbf{x} \in D} |f_n(\mathbf{x}) - f(\mathbf{x})|)_{n=1}^{\infty}$ converges to 0 as $n \to \infty$, and the function f is continuous. Therefore, the two terms on the right-hand side of the previous display converges to 0 as $k \to \infty$ because $(\sup_{\mathbf{x} \in D} |f_{n_k}(\mathbf{x}) - f(\mathbf{x})|)_{k=1}^{\infty}$ is a subsequence of $(\sup_{\mathbf{x} \in D} |f_n(\mathbf{x}) - f(\mathbf{x})|)_{n=1}^{\infty}$, and hence, $f_{n_k}(\widehat{\mathbf{x}}_{n_k}) \to f(\widehat{\mathbf{x}}_0^*)$ as $n \to \infty$. By definition, $\widehat{\mathbf{x}}_{n_k}$ is the maximizer of f_{n_k}, implying that $f_{n_k}(\widehat{\mathbf{x}}_{n_k}) \geq f_{n_k}(\mathbf{x})$ for all $\mathbf{x} \in D$. Therefore,

$$f(\widehat{\mathbf{x}}_0^*) = \lim_{n \to \infty} f_{n_k}(\widehat{\mathbf{x}}_{n_k}) \geq \lim_{n \to \infty} f_{n_k}(\mathbf{x}) = f(\mathbf{x})$$

for any $\mathbf{x} \in D$, where we have used the fact that f_{n_k} converges point-wise to f as $k \to \infty$. This further shows that $\widehat{\mathbf{x}}_0^*$ is the maximizer of f. Since the maximizer of f is unique, we conclude that $\widehat{\mathbf{x}}_{n_k} \to \widehat{\mathbf{x}}_0^* = \widehat{\mathbf{x}}_0$ as $k \to \infty$. Note that any converging subsequence $(\widehat{\mathbf{x}}_{n_k})_{k=1}^{\infty}$ of $(\widehat{\mathbf{x}}_n)_{n=1}^{\infty}$ converges to the same limit point $\widehat{\mathbf{x}}_0$. Therefore we conclude that the entire sequence $\widehat{\mathbf{x}}_n$ converges to $\widehat{\mathbf{x}}_0$.

For part (b), we first observe that $\widehat{\mathbf{x}}_0$ is a strict maximizer of f because the negative Hessian of f is positive definite, and $-f$ is convex because the Hessian of $-f$ is always positive definite. Therefore, there exists some $\delta > 0$, such that $f(\widehat{\mathbf{x}}_0) > f(\mathbf{x})$ for all $\mathbf{x} \in \{\mathbf{x} \in D : 0 < \|\mathbf{x} - \widehat{\mathbf{x}}_0\| < \delta\}$. Now we claim that there exists some constant $\xi > 0$ that only depends on $\delta > 0$, such that

$$f(\mathbf{x}) + \xi \leq f(\widehat{\mathbf{x}}_0) \quad \text{for any } \mathbf{x} \in \{\mathbf{x} \in D : \|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 \geq \delta\}. \tag{C.1}$$

In fact, if one assumes otherwise, then for all $\xi > 0$, there exists some $\mathbf{x}_\xi$ with $\|\mathbf{x}_\xi - \widehat{\mathbf{x}}_0\|_2 \geq \delta$, such that

$f(\mathbf{x}_\xi) + \xi > f(\widehat{\mathbf{x}}_0)$. Taking a sequence $(\xi_j)_{j=1}^\infty = (1/j)_{j=1}^\infty$ yields a sequence $(\mathbf{x}_j)_{j=1}^\infty$ such that

$$f(\mathbf{x}_j) + \frac{1}{j} > f(\widehat{\mathbf{x}}_0).$$

Let $(\mathbf{x}_{j_k})_{k=1}^\infty$ be a converging subsequence of $(\mathbf{x}_j)_{j=1}^\infty$ that converges to some point $\mathbf{y} \in D$. Then the continuity of f leads to

$$f(\mathbf{y}) = \lim_{k\to\infty} \left\{ f(\mathbf{x}_{j_k}) + \frac{1}{j_k} \right\} \geq f(\widehat{\mathbf{x}}_0).$$

Since $(\mathbf{x}_{j_k})_{k=1}^\infty \subset \{\mathbf{x} \in D : \|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 \geq \delta\}$, and the latter superset is compact, it follows that $\|\mathbf{y} - \widehat{\mathbf{x}}_0\|_2 \geq \delta$, and hence, it must be the case that $f(\mathbf{y}) < f(\widehat{\mathbf{x}}_0)$ due to the uniqueness of the maximizer of f. Therefore we conclude that there exists some constant $\xi > 0$ that only depends on $\delta > 0$, such that

$$f(\mathbf{x}) + \xi \leq f(\widehat{\mathbf{x}}_0) \quad \text{for all } \mathbf{x} \in \{\mathbf{x} \in D : \|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 \geq \delta\}.$$

Now we further claim that for any $\epsilon > 0$, there exists some $\eta(\epsilon) > 0$ such that

$$f(\widehat{\mathbf{x}}_0) \geq f(\mathbf{x}) + 4\eta(\epsilon) \quad \text{for all } \mathbf{x} \in \{\mathbf{x} \in D : \|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 > \epsilon\}. \tag{C.2}$$

Now let $\epsilon > 0$ be arbitrarily given. We consider two cases: If $\epsilon \geq \delta$, then we can take $\eta(\epsilon) = \xi/4$ directly. Then the previous claim applies; If $\epsilon < \delta$, then we see that

$$\{\mathbf{x} \in D : \|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 > \epsilon\} = \{\mathbf{x} \in D : \epsilon < \|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 < \delta\} \cup \{\mathbf{x} \in D : \|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 \geq \delta\}.$$

For any $\mathbf{x} \in \{\mathbf{x} \in D : \epsilon < \|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 < \delta\}$, by the fact that $\widehat{\mathbf{x}}_0$ is a strict maximizer of f, it must be the case that

$$\sup_{\mathbf{x}:\epsilon < \|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 < \delta} f(\mathbf{x}) < f(\widehat{\mathbf{x}}_0).$$

Now take

$$\eta(\epsilon) = \min\left[\frac{1}{4} \left\{ f(\widehat{\mathbf{x}}_0) - \sup_{\mathbf{x}:\epsilon < \|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 < \delta} f(\mathbf{x}) \right\}, \frac{\xi}{4} \right].$$

It follows from (C.1) that

$$\begin{aligned}
&\sup_{\mathbf{x}\in D:\|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 > \epsilon} f(\mathbf{x}) + 4\eta(\epsilon) \\
&\quad = \max\left[\sup_{\mathbf{x}\in D:\epsilon < \|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 < \delta} f(\mathbf{x}) + 4\eta(\epsilon), \sup_{\mathbf{x}\in D:\|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 \geq \delta} f(\mathbf{x}) + 4\eta(\epsilon) \right] \\
&\quad \leq \max\left[\sup_{\mathbf{x}\in D:\epsilon < \|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 < \delta} f(\mathbf{x}) + f(\widehat{\mathbf{x}}_0) - \sup_{\mathbf{x}\in D:\epsilon < \|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 < \delta} f(\mathbf{x}), \sup_{\mathbf{x}\in D:\|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 \geq \delta} f(\mathbf{x}) + \xi \right] \\
&\quad = f(\widehat{\mathbf{x}}_0),
\end{aligned}$$

and this completes the proof of (C.2).

We finally prove the desired result. Since $f_n \to f$ uniformly as $n \to \infty$, then there exists some large positive constant $N_1 = N_1(\epsilon)$, such that

$$\sup_{\mathbf{x}\in D} |f_n(\mathbf{x}) - f(\mathbf{x})| < \eta(\epsilon/2) \quad \text{for all } n > N_1.$$

Furthermore, by part (a) and the continuity of f, there exists some large positive constant $N_2 = N_2(\epsilon)$, such that

$$\|\widehat{\mathbf{x}}_n - \widehat{\mathbf{x}}_0\| < \frac{\epsilon}{2} \quad \text{and} \quad |f(\widehat{\mathbf{x}}_n) - f(\widehat{\mathbf{x}}_0)| < \eta(\epsilon/2) \quad \text{for all } n > N_2.$$

Note that

$$\{\mathbf{x} \in D : \|\mathbf{x} - \widehat{\mathbf{x}}_n\|_2 > \epsilon\} \subset \{\mathbf{x} \in D : \|\mathbf{x} - \widehat{\mathbf{x}}_0\|_2 > \epsilon/2\}$$

for $n \geq N_2(\epsilon)$. Hence, for all $n \geq N_2(\epsilon)$, we obtain

$$\begin{aligned}
\sup_{\mathbf{x}\in D:\|\mathbf{x}-\widehat{\mathbf{x}}_n\|_2>\epsilon} f_n(\mathbf{x}) &\leq \sup_{\mathbf{x}\in D:\|\mathbf{x}-\widehat{\mathbf{x}}_n\|_2>\epsilon} |f_n(\mathbf{x}) - f(\mathbf{x})| + \sup_{\mathbf{x}\in D:\|\mathbf{x}-\widehat{\mathbf{x}}_n\|_2>\epsilon} f(\mathbf{x}) \\
&\leq \sup_{\mathbf{x}\in D} |f_n(\mathbf{x}) - f(\mathbf{x})| + \sup_{\mathbf{x}\in D:\|\mathbf{x}-\widehat{\mathbf{x}}_n\|_2>\epsilon} f(\mathbf{x}) \\
&\leq \sup_{\mathbf{x}\in D} |f_n(\mathbf{x}) - f(\mathbf{x})| + \sup_{\mathbf{x}\in D:\|\mathbf{x}-\widehat{\mathbf{x}}_0\|_2>\epsilon/2} f(\mathbf{x}) \\
&\leq \sup_{\mathbf{x}\in D} |f_n(\mathbf{x}) - f(\mathbf{x})| + f(\widehat{\mathbf{x}}_0) - 4\eta(\epsilon/2),
\end{aligned}$$

where we have used (C.2) at the end of the previous display. We further write

$$\begin{aligned}
\sup_{\mathbf{x}\in D:\|\mathbf{x}-\widehat{\mathbf{x}}_n\|_2>\epsilon} f_n(\mathbf{x}) &\leq \sup_{\mathbf{x}\in D} |f_n(\mathbf{x}) - f(\mathbf{x})| + f(\widehat{\mathbf{x}}_0) - 4\eta(\epsilon/2) \\
&= \sup_{\mathbf{x}\in D} |f_n(\mathbf{x}) - f(\mathbf{x})| + f(\widehat{\mathbf{x}}_0) - f(\widehat{\mathbf{x}}_n) + f(\widehat{\mathbf{x}}_n) - f_n(\widehat{\mathbf{x}}_n) + f_n(\widehat{\mathbf{x}}_n) - 4\eta(\epsilon/2) \\
&\leq \sup_{\mathbf{x}\in D} |f_n(\mathbf{x}) - f(\mathbf{x})| + |f(\widehat{\mathbf{x}}_0) - f(\widehat{\mathbf{x}}_n)| + \sup_{\mathbf{x}\in D} |f(\mathbf{x}) - f_n(\mathbf{x})| + f_n(\widehat{\mathbf{x}}_n) - 4\eta(\epsilon/2)
\end{aligned}$$

for all $n \geq N_2(\epsilon)$. Hence for any $n \geq N = N(\epsilon) = \max\{N_1(\epsilon), N_2(\epsilon)\}$,

$$\begin{aligned}
\sup_{\mathbf{x}\in D:\|\mathbf{x}-\widehat{\mathbf{x}}_n\|_2>\epsilon} f_n(\mathbf{x}) &\leq 2\sup_{\mathbf{x}\in D} |f_n(\mathbf{x}) - f(\mathbf{x})| + |f(\widehat{\mathbf{x}}_0) - f(\widehat{\mathbf{x}}_n)| + f_n(\widehat{\mathbf{x}}_n) - 4\eta(\epsilon/2) \\
&\leq f_n(\widehat{\mathbf{x}}_n) + 2\eta(\epsilon/2) + \eta(\epsilon/2) - 4\eta(\epsilon/2) = f_n(\widehat{\mathbf{x}}_n) - \eta(\epsilon/2).
\end{aligned}$$

The proof is completed by observing that both N and $\eta(\epsilon/2)$ only depend on ϵ so that we can replace $\eta(\epsilon/2)$ by $\eta(\epsilon)$ without changing the proof, and these two quantities does not depend on n. $\square$

Proof of Theorem 2. We begin the proof with writing down the likelihood function for $\mathbf{x}_i$:

$$\ell_{\mathbf{A}}(\mathbf{x}_i) = \sum_{j\neq i}^{n} \{A_{ij} \log(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_{0j}) + (1 - A_{ij}) \log(1 - \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_{0j})\}.$$

For convenience we denote the following functions:

$$
\begin{aligned}
M_n(\mathbf{x}) &= \frac{1}{n}\ell_{\mathbf{A}}(\mathbf{x}) = \frac{1}{n}\sum_{j\neq i}^n \{A_{ij}\log(\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}) + (1-A_{ij})\log(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\},\\
M(\mathbf{x}) &= \mathbb{E}_0\{M_n(\mathbf{x})\} = \frac{1}{n}\sum_{j\neq i}^n \{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}\log(\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}) + (1-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\log(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\},\\
\boldsymbol{\Psi}_n(\mathbf{x}) &= \frac{\partial M_n}{\partial \mathbf{x}}(\mathbf{x}) = \frac{1}{n}\sum_{j\neq i}^n \left\{\frac{(A_{ij}-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})}\right\}\mathbf{x}_{0j},\\
\boldsymbol{\Psi}(\mathbf{x}) &= \mathbb{E}_0\{\boldsymbol{\Psi}_n(\mathbf{x})\} = \mathbb{E}_0\left\{\frac{\partial M_n}{\partial \mathbf{x}}(\mathbf{x})\right\} = \frac{1}{n}\sum_{j\neq i}^n \left\{\frac{(\mathbf{x}_{0i}-\mathbf{x})^{\mathrm{T}}\mathbf{x}_{0j}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})}\right\}\mathbf{x}_{0j}.
\end{aligned}
$$

Note that the function M itself also depends on n implicitly. Denote Ψ_{nk} the kth component of $\boldsymbol{\Psi}_n$, *i.e.*,

$$
\Psi_{nk}(\mathbf{x}) = \frac{1}{n}\sum_{j\neq i}^n \left\{\frac{(A_{ij}-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})}\right\}x_{0jk},\quad k=1,2,\ldots,d,
$$

where $\mathbf{x}_{0j} = [x_{0j1},\ldots,x_{0jd}]^{\mathrm{T}} \in \mathbb{R}^d$. Simple algebra shows that

$$
\begin{aligned}
\frac{\partial^2 M_n}{\partial\mathbf{x}\partial\mathbf{x}^{\mathrm{T}}}(\mathbf{x}) &= -\frac{1}{n}\sum_{j\neq i}^n \frac{\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})} - \frac{1}{n}\sum_{j\neq i}^n \frac{(A_{ij}-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})(1-2\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}^2},\\
\frac{\partial^2 M}{\partial\mathbf{x}\partial\mathbf{x}^{\mathrm{T}}}(\mathbf{x}) &= -\frac{1}{n}\sum_{j\neq i}^n \left\{\frac{(\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})^2 - 2\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0i}\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j} + \mathbf{x}^{\mathrm{T}}\mathbf{x}_{0i}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})}\right\}\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}},\\
\frac{\partial^2 \Psi_{nk}}{\partial\mathbf{x}\partial\mathbf{x}^{\mathrm{T}}}(\mathbf{x}) &= \frac{1}{n}\sum_{j\neq i}^n \frac{x_{0jk}(1-2\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})}{\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}^2}\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}} + \frac{1}{n}\sum_{j\neq i}^n \frac{x_{0jk}\{(1-2\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}) + 2(A_{ij}-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}}{\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}^2}\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}\\
&\quad + \frac{1}{n}\sum_{j\neq i}^n \frac{x_{0jk}\{2(A_{ij}-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})(1-2\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})^2\}}{\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}^3}\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}.
\end{aligned}
$$

Clearly, $\mathcal{X}(\delta)$ is compact and $M_n(\mathbf{x})$ is continuous. Therefore $\widehat{\mathbf{x}}_i^{(\mathrm{MLE})} = \arg\max_{\mathbf{x}\in\mathcal{X}(\delta)} M_n(\mathbf{x})$ exists with probability one. Furthermore, by Shannon's lemma (see, for example, Lemma 2.2.1 in Bickel and Doksum, 2015), we know that $M(\mathbf{x})$ is maximized at $\mathbf{x}=\mathbf{x}_{0i}$. Since $\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j} \in [\delta, 1-\delta]$ for all $j\in[n]$, implying that

$$
\begin{aligned}
-\frac{\partial^2 M}{\partial\mathbf{x}\partial\mathbf{x}^{\mathrm{T}}}(\mathbf{x}) &= \frac{1}{n}\sum_{j\neq i}^n \left\{\frac{(\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})^2 - 2\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0i}\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j} + \mathbf{x}^{\mathrm{T}}\mathbf{x}_{0i}}{(\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})^2(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})^2}\right\}\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}\\
&\succ \frac{1}{n}\sum_{j\neq i}^n \left\{\frac{(\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})^2 - 2\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0i}\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j} + (\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0i})^2}{(\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})^2(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})^2}\right\}\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}\\
&= \frac{1}{n}\sum_{j\neq i}^n \left\{\frac{(\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j} - \mathbf{x}^{\mathrm{T}}\mathbf{x}_{0i})^2}{(\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})^2(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})^2}\right\}\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}} \succeq \mathbf{O},
\end{aligned}
$$

it follows that $\mathbf{x}_{0i}$ is the unique maximizer of M because $-M$ is strictly convex, and $\mathcal{X}(\delta)$ is convex and compact. Observe that the function M (implicitly) depends on n. Since the gradient of M is

$$\sup_{n\in\mathbb{N}_+}\sup_{\mathbf{x}\in\mathcal{X}}\left\|\frac{\partial M}{\partial \mathbf{x}}(\mathbf{x})\right\|_2 \leq \sup_{n\in\mathbb{N}_+}\frac{1}{n}\sum_{j\neq i}\sup_{\mathbf{x}\in\mathcal{X}}\left\|\frac{(\mathbf{x}_{0i}-\mathbf{x})^{\mathrm{T}}\mathbf{x}_{0j}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})}\mathbf{x}_{0j}\right\|_2 \leq \frac{2}{\delta^2} < \infty.$$

Therefore the function class $\{M(\mathbf{x})\}_n$ (each function M depends on n implicitly) is equicontinuous. We also observe that

$$\begin{aligned}\lim_{n\to\infty} M(\mathbf{x}) &= \lim_{n\to\infty}\frac{1}{n}\sum_{j\neq i}^{n}\{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}\log(\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}) + (1-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\log(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}\\ &= \bar{M}(\mathbf{x}) := \int_{\mathcal{X}}\left\{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_1\log(\mathbf{x}^{\mathrm{T}}\mathbf{x}_1) + (1-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_1)\log(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)\right\}F(\mathrm{d}\mathbf{x}_1)\end{aligned}$$

converges for all $\mathbf{x}\in\mathcal{X}(\delta)$, and $\mathcal{X}(\delta)$ is a compact subset of $\mathbb{R}^d$. Hence, by the Arzela-Ascoli theorem, the convergence $M\to\bar{M}$ as $n\to\infty$ is also uniform. Note that the Hessian of $-M$ is strictly positive definite. Also note that

$$\begin{aligned}&\sup_{\mathbf{x}_1,\mathbf{x}\in\mathcal{X}}\left\|\frac{\partial}{\partial\mathbf{x}}\left\{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_1\log(\mathbf{x}^{\mathrm{T}}\mathbf{x}_1) + (1-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_1)\log(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)\right\}\right\|_2\\ &\quad= \sup_{\mathbf{x}_1,\mathbf{x}\in\mathcal{X}}\left\|\frac{(\mathbf{x}_{0i}-\mathbf{x})^{\mathrm{T}}\mathbf{x}_1}{\mathbf{x}^{\mathrm{T}}\mathbf{x}_1(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)}\mathbf{x}_1\right\|_2 \leq \frac{2}{\delta^2} < \infty,\\ &\sup_{\mathbf{x}_1,\mathbf{x}\in\mathcal{X}}\left\|\frac{\partial^2}{\partial\mathbf{x}\partial\mathbf{x}}\left\{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_1\log(\mathbf{x}^{\mathrm{T}}\mathbf{x}_1) + (1-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_1)\log(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)\right\}\right\|_2\\ &\quad= \sup_{\mathbf{x}_1,\mathbf{x}\in\mathcal{X}}\left\|\frac{(\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)^2 - 2\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0i}\mathbf{x}^{\mathrm{T}}\mathbf{x}_1 + \mathbf{x}^{\mathrm{T}}\mathbf{x}_{0i}}{\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_1(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)\}^2}\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}\right\|_2 \leq \frac{4}{\delta^4} < \infty,\end{aligned}$$

and hence, by the Lebesgue dominating convergence theorem, the negative Hessian of $\bar{M}$ equals

$$-\frac{\partial^2\bar{M}(\mathbf{x})}{\partial\mathbf{x}\partial\mathbf{x}^{\mathrm{T}}}(\mathbf{x}) = \int_{\mathcal{X}}\frac{(\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)^2 - 2\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0i}\mathbf{x}^{\mathrm{T}}\mathbf{x}_1 + \mathbf{x}^{\mathrm{T}}\mathbf{x}_{0i}}{\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_1(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)\}^2}\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}F(\mathrm{d}\mathbf{x}_1),$$

which is also strictly positive definite. Hence $-\bar{M}$ is also strictly convex, and the maximizer for $\bar{M}$ is unique in $\mathcal{X}(\delta)$ because $\mathcal{X}(\delta)$ is convex and compact. Therefore, we apply Lemma C.1 to obtain the following conclusion: For any $\epsilon > 0$, there exists an $\eta(\epsilon) > 0$ that depends on $\epsilon > 0$ but does not depend on n, and a positive integer $N = N(\epsilon)$ that depends on $\epsilon > 0$, such that

$$\sup_{\|\mathbf{x}-\mathbf{x}_{0i}\|>\epsilon} M(\mathbf{x}) + \eta(\epsilon) \leq M(\mathbf{x}_{0i}) \tag{C.3}$$

for all $n > N(\epsilon)$.

We first claim that

$$\sup_{\mathbf{x}\in\mathcal{X}(\delta)}|M_n(\mathbf{x}) - M(\mathbf{x})| \overset{\mathbb{P}_0}{\to} 0. \tag{C.4}$$

Define a stochastic process $\{J(\mathbf{x}) = M_n(\mathbf{x}) - M(\mathbf{x}) : \mathbf{x}\in\mathcal{X}(\delta)\}$. Since for any $\mathbf{x}_1,\mathbf{x}_2\in\mathcal{X}(\delta)$, there exists a constant K_δ only depending on $\delta > 0$, such that

$$\begin{aligned}\left|\log\left(\frac{\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_{0j}}{1-\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_{0j}}\right) - \log\left(\frac{\mathbf{x}_2^{\mathrm{T}}\mathbf{x}_{0j}}{1-\mathbf{x}_2^{\mathrm{T}}\mathbf{x}_{0j}}\right)\right| &\leq \sup_{\mathbf{x}\in\mathcal{X}(\delta), j\in[n]}\left\|\frac{\partial}{\partial\mathbf{x}}\log\left(\frac{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}}{1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}}\right)\right\|\|\mathbf{x}_1-\mathbf{x}_2\| \\ &\leq K_\delta\|\mathbf{x}_1-\mathbf{x}_2\|,\end{aligned}$$

it follows from Hoeffding's inequality that

$$\mathbb{P}_0\left(|J(\mathbf{x}_1) - J(\mathbf{x}_2)| > t\right) \leq 2\exp\left(-\frac{2nt^2}{K_\delta^2\|\mathbf{x}_1-\mathbf{x}_2\|^2}\right),$$

implying that $J(\cdot)$ is a sub-Gaussian process with respect to $K_\delta n^{-1/2}\|\cdot\|$. Hence the packing entropy can also be bounded: There exists some large constant $C > 0$, such that

$$\log\mathcal{D}\left(\epsilon, \mathcal{X}(\delta), \frac{K_\delta}{\sqrt{n}}\|\cdot\|\right) \leq d\log\left(\frac{C}{\epsilon\sqrt{n}}\right).$$

Hence, by the fact that $\sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathcal{X}(\delta)} K_\delta n^{-1/2}\|\mathbf{x}_1-\mathbf{x}_2\| \leq cn^{-1/2}$ for some constant $c\in(0, C)$, a maximum inequality for sub-Gaussian process (see, for example, Corollary 8.5 in Kosorok, 2008), and the change of variable $u = \log\{C/(\epsilon\sqrt{n})\}$, we have

$$\begin{aligned}\mathbb{E}_0\left(\sup_{\mathbf{x}\in\mathcal{X}(\delta)}|J(\mathbf{x})|\right) &\lesssim \mathbb{E}_0(|J(\mathbf{x}_{0i})|) + \int_0^{cn^{-1/2}}\sqrt{\log\mathcal{D}\left(\epsilon,\mathcal{X}(\delta),\frac{M}{\sqrt{n}}\|\cdot\|\right)}\mathrm{d}\epsilon \\ &\lesssim \sqrt{\mathrm{var}_0(J(\mathbf{x}_{0i}))} + \int_0^{cn^{-1/2}}\sqrt{\log\frac{C}{\epsilon\sqrt{n}}}\mathrm{d}\epsilon \\ &= \left\{\frac{1}{n^2}\sum_{j\neq i}^n \mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\log\frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{1-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}\right\}^{1/2} + \frac{C}{\sqrt{n}}\int_{\log\frac{C}{c}}^{\infty}\sqrt{u}\mathrm{e}^{-u}\mathrm{d}u \to 0\end{aligned}$$

as $n\to\infty$. Therefore we conclude that $\sup_{\mathbf{x}\in\mathcal{X}(\delta)}|J(\mathbf{x})| = o_{\mathbb{P}_0}(1)$.
In the proof below we shall drop the superscript (MLE) from $\widehat{\mathbf{x}}_i^{(\mathrm{MLE})}$ and write $\widehat{\mathbf{x}}_i = \widehat{\mathbf{x}}_i^{(\mathrm{MLE})}$ for short. We next use the claim (C.4) to show that $\widehat{\mathbf{x}}_i$ is consistent for $\mathbf{x}_{0i}$. The proof here is quite similar to that of Theorem 5.7 in Van der Vaart (2000) and presented here for completeness. In fact, this implies that $M_n(\mathbf{x}_{0i}) - M(\mathbf{x}_{0i}) \overset{\mathbb{P}_0}{\to} 0$. Furthermore, $\widehat{\mathbf{x}}_i$ is the maximizer of M_n, implying that

$$\begin{aligned}M(\mathbf{x}_{0i}) - M(\widehat{\mathbf{x}}_i) &= M_n(\mathbf{x}_{0i}) + o_{\mathbb{P}_0}(1) - M(\widehat{\mathbf{x}}_i) \leq M_n(\widehat{\mathbf{x}}_i) - M(\widehat{\mathbf{x}}_i) + o_{\mathbb{P}_0}(1) \\ &\leq \sup_{\mathbf{x}\in\mathcal{X}(\delta)}|J(\mathbf{x})| + o_{\mathbb{P}_0}(1) = o_{\mathbb{P}_0}(1).\end{aligned}$$

This shows that $\mathbb{P}_0(M(\mathbf{x}_{0i}) - M(\widehat{\mathbf{x}}_i) \geq \eta) \to 0$ for all $\eta > 0$. Recall that by (C.3) for all $\epsilon > 0$, there exists some $\eta(\epsilon) > 0$ not depending on n, such that $\|\widehat{\mathbf{x}} - \mathbf{x}_{0i}\| > \epsilon$ implies $M(\widehat{\mathbf{x}}_i) \leq M(\mathbf{x}_{0i}) - \eta(\epsilon)$, although the function $M(\mathbf{x}_{0i})$ implicitly depends on n. Namely, for all $\epsilon > 0$, there exists some $\eta = \eta(\epsilon) > 0$ such that

$$\mathbb{P}_0\left(\|\widehat{\mathbf{x}}_i - \mathbf{x}_{0i}\| > \epsilon\right) \leq \mathbb{P}_0\left\{M(\mathbf{x}_{0i}) - M(\widehat{\mathbf{x}}_i) \geq \eta(\epsilon)\right\} \to 0$$

as $n \to \infty$. This completes the proof of consistency of $\widehat{\mathbf{x}}_i$ for $\mathbf{x}_{0i}$.

We finally show the asymptotic normality of $\widehat{\mathbf{x}}_i$. Since $\widehat{\mathbf{x}}_i$ is consistent for $\mathbf{x}_{0i}$, it follows that with probability tending to one, $\widehat{\mathbf{x}}_i$ is in the interior of $\mathcal{X}(\delta)$ since $\mathbf{x}_{0i}$ is. Assume this event occurs. By Taylor's expansion, we have, for $k = 1, 2, \ldots, d$, that

$$0 = \Psi_{nk}(\widehat{\mathbf{x}}_i) = \Psi_{nk}(\mathbf{x}_{0i}) + \frac{\partial \Psi_{nk}}{\partial \mathbf{x}^{\mathrm{T}}}(\mathbf{x}_{0i})(\widehat{\mathbf{x}}_i - \mathbf{x}_{0i}) + \frac{1}{2}(\widehat{\mathbf{x}}_i - \mathbf{x}_{0i})^{\mathrm{T}} \left\{ \frac{\partial^2 \Psi_{nk}}{\partial \mathbf{x} \partial \mathbf{x}^{\mathrm{T}}}(\widetilde{\mathbf{x}}_k) \right\} (\widehat{\mathbf{x}}_i - \mathbf{x}_{0i}),$$

where $\widetilde{\mathbf{x}}_k$ lies on the line segment linking $\mathbf{x}_{0i}$ and $\widehat{\mathbf{x}}_{0i}$. Since for any $\mathbf{x} \in \mathcal{X}(\delta)$,

$$\begin{aligned}
\left\| \frac{\partial^2 \Psi_{nk}}{\partial \mathbf{x} \partial \mathbf{x}^{\mathrm{T}}}(\mathbf{x}) \right\| &\leq \frac{1}{n} \sum_{j \neq i}^{n} \frac{\{1 + 2(1-\delta)\} \|\mathbf{x}_{0j}\|^2}{\delta^2 (1-\delta)^2} + \frac{1}{n} \sum_{j \neq i}^{n} \frac{\{(3 - 2\delta) + 2(2-\delta)\} \|\mathbf{x}_{0j}\|^2}{\delta^2 (1-\delta)^2} \\
&\quad + \frac{1}{n} \sum_{j \neq i}^{n} \frac{2(2-\delta)(3-2\delta)^2 \|\mathbf{x}_{0j}\|^2}{\delta^3 (1-\delta)^3} \lesssim \frac{1}{n} \|\mathbf{X}_0\|_{\mathrm{F}}^2 \leq 1,
\end{aligned}$$

it follows that the Hessian of $\Psi_{nk}(\widetilde{\mathbf{x}})$ is bounded in probability. Observe that,

$$\mathbb{E}_0 \left\{ \frac{\partial \mathbf{\Psi}_n(\mathbf{x}_{0i})}{\partial \mathbf{x}^{\mathrm{T}}} \right\} = -\frac{1}{n} \sum_{j \neq i}^{n} \frac{\mathbf{x}_{0j} \mathbf{x}_{0j}^{\mathrm{T}}}{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})},$$

and for any $s, t \in [d]$,

$$\mathrm{var}_0 \left\{ \frac{\partial \Psi_{ns}(\mathbf{x}_{0i})}{\partial x_t} \right\} = \frac{1}{n^2} \sum_{j \neq i}^{n} \frac{(1 - 2\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})^2 (x_{0js} x_{0jt})^2}{\{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})\}^3} \to 0$$

as $n \to \infty$, we obtain from the law of large numbers that

$$\frac{\partial \mathbf{\Psi}_n(\mathbf{x}_{0i})}{\partial \mathbf{x}^{\mathrm{T}}} = -\mathbf{G}_n(\mathbf{x}_{0i}) + o_{\mathbb{P}_0}(1).$$

Therefore, we conclude from the Taylor's expansion and $\widehat{\mathbf{x}}_i - \mathbf{x}_{0i} = o_{\mathbb{P}_0}(1)$ that

$$\begin{aligned}
-\mathbf{\Psi}_n(\mathbf{x}_{0i}) &= \left\{ -\mathbf{G}_n(\mathbf{x}_{0i}) + o_{\mathbb{P}_0}(1) + \frac{1}{2}(\widehat{\mathbf{x}}_i - \mathbf{x}_{0i})^{\mathrm{T}} O_{\mathbb{P}_0}(1) \right\} (\widehat{\mathbf{x}}_i - \mathbf{x}_{0i}) \\
&= \{ -\mathbf{G}_n(\mathbf{x}_{0i}) + o_{\mathbb{P}_0}(1) \} (\widehat{\mathbf{x}}_i - \mathbf{x}_{0i}).
\end{aligned}$$

Namely,

$$\sqrt{n}(\widehat{\mathbf{x}}_i - \mathbf{x}_{0i}) = \{ \mathbf{G}_n(\mathbf{x}_{0i}) + o_{\mathbb{P}_0}(1) \}^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{j \neq i}^{n} \frac{(A_{ij} - \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j}) \mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})} \right\}.$$

Observe that

$$\sum_{j\neq i}^{n}\mathbb{E}_0\left\{\left\|\frac{1}{\sqrt{n}}\frac{(A_{ij}-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\right\|^3\right\}\leq\frac{1}{n^{3/2}}\sum_{j\neq i}^{n}\frac{(2-\delta)^3\|\mathbf{x}_{0j}\|^3}{\{\delta(1-\delta)\}^3}\to 0,$$

$$\mathrm{var}_0\left(\frac{1}{\sqrt{n}}\sum_{j\neq i}^{n}\frac{(A_{ij}-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\right)=\mathbf{G}_n(\mathbf{x}_{0i})\to\mathbf{G}(\mathbf{x}_{0i})\quad\text{as } n\to\infty,$$

it follows from Lyapunov's central limit theorem that $\sqrt{n}(\widehat{\mathbf{x}}_i-\mathbf{x}_{0i})\xrightarrow{\mathcal{L}}\mathrm{N}(\mathbf{0},\mathbf{G}(\mathbf{x}_{0i})^{-1})$.

We finally show that $\boldsymbol{\Sigma}(\mathbf{x})-\mathbf{G}(\mathbf{x})^{-1}$ is positive semidefinite for all $\mathbf{x}\in\mathcal{X}(\delta)$. Recall that for the ith row $\widehat{\mathbf{x}}_i^{(\mathrm{ASE})}$ of the ASE, $\sqrt{n}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i^{(\mathrm{ASE})}-\mathbf{x}_{0i})\xrightarrow{\mathcal{L}}\mathrm{N}(\mathbf{0},\boldsymbol{\Sigma}(\mathbf{x}_{0i}))$ for a sequence of orthogonal matrices $(\mathbf{W})_{n=1}^{\infty}=(\mathbf{W}_n)_{n=1}^{\infty}\subset\mathbb{O}(d)$ by Theorem 1. Since $F_n(\cdot)=(1/n)\sum_{i=1}^{n}\mathbb{1}\{\mathbf{x}_i\leq\cdot\}$ converges to F strongly according to condition (2.1), it follows that for any $\mathbf{x}\in\mathcal{X}(\delta)$,

$$\begin{aligned}
\boldsymbol{\Delta}_n&:=\int_{\mathcal{X}}\mathbf{x}\mathbf{x}^{\mathrm{T}}F_n(\mathrm{d}\mathbf{x})=\frac{1}{n}\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0\to\boldsymbol{\Delta},\\
\boldsymbol{\Sigma}_n(\mathbf{x})&:=\boldsymbol{\Delta}_n^{-1}\left[\int_{\mathcal{X}}\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_1(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)\}\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}F_n(\mathrm{d}\mathbf{x}_1)\right]\boldsymbol{\Delta}_n^{-1}\\
&=\left(\frac{1}{n}\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0\right)^{-1}\left(\frac{1}{n}\mathbf{X}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})\mathbf{X}_0\right)\left(\frac{1}{n}\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0\right)^{-1}\to\boldsymbol{\Sigma}(\mathbf{x}_{0i}),\\
\mathbf{G}_n(\mathbf{x})&:=\int_{\mathcal{X}}\left\{\frac{\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}_1(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_1)}\right\}F_n(\mathrm{d}\mathbf{x}_1)=\frac{1}{n}\mathbf{X}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})^{-1}\mathbf{X}_0\to\mathbf{G}(\mathbf{x}),
\end{aligned}$$

where $\mathbf{D}_n(\mathbf{x})=\mathrm{diag}\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{01}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{01}),\ldots,\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0n}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0n})\}$. Now let $\mathbf{X}_0$ yield singular value decomposition $\mathbf{X}_0=\mathbf{U}_0\mathbf{S}_0^{1/2}\mathbf{V}_0^{\mathrm{T}}$ with $\mathbf{U}_0\in\mathbb{O}(n,d)$, $\mathbf{S}_0^{1/2}$ being diagonal, and $\mathbf{V}_0\in\mathbb{O}(d)$. We see immediately that

$$\begin{aligned}
\boldsymbol{\Sigma}_n(\mathbf{x})&=\left(\frac{1}{n}\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0\right)^{-1}\left(\frac{1}{n}\mathbf{X}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})\mathbf{X}_0\right)\left(\frac{1}{n}\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0\right)^{-1}\\
&=n(\mathbf{V}_0\mathbf{S}_0^{-1}\mathbf{V}_0^{\mathrm{T}})(\mathbf{V}_0\mathbf{S}_0^{1/2}\mathbf{U}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})\mathbf{U}_0\mathbf{S}_0^{1/2}\mathbf{V}_0^{\mathrm{T}})(\mathbf{V}_0\mathbf{S}_0^{-1}\mathbf{V}_0^{\mathrm{T}})\\
&=n\mathbf{V}_0\mathbf{S}_0^{-1/2}(\mathbf{U}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})\mathbf{U}_0)\mathbf{S}_0^{-1/2}\mathbf{V}_0^{\mathrm{T}},\\
\mathbf{G}_n(\mathbf{x})^{-1}&=n(\mathbf{X}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})^{-1}\mathbf{X}_0)^{-1}=n(\mathbf{V}_0\mathbf{S}_0^{1/2}\mathbf{U}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})^{-1}\mathbf{U}_0\mathbf{S}_0^{1/2}\mathbf{V}_0^{\mathrm{T}})^{-1}\\
&=n\mathbf{V}_0\mathbf{S}_0^{-1/2}(\mathbf{U}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})^{-1}\mathbf{U}_0)^{-1}\mathbf{S}_0^{-1/2}\mathbf{V}_0^{\mathrm{T}}.
\end{aligned}$$

Since $\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_0=\mathbf{I}_d$, it follows that $\mathbf{U}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})\mathbf{U}_0-(\mathbf{U}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})^{-1}\mathbf{U}_0)^{-1}$ is positive semidefinite (Marshall and Olkin, 1990), and hence, $\boldsymbol{\Sigma}(\mathbf{x})-\mathbf{G}(\mathbf{x})^{-1}=\lim_{n\to\infty}\{\boldsymbol{\Sigma}_n(\mathbf{x}_{0i})-\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\}$ is positive semidefinite for any $\mathbf{x}\in\mathcal{X}(\delta)$. The proof is thus completed. $\square$

C.2 Proof of Theorem 3

Proof of Theorem 3. The idea of the proof is very similar to that of Theorem 5.45 in Van der Vaart (2000). Using the notation there (which coincides with the notation in Section C.1), we denote

$$\mathbf{\Psi}_n(\mathbf{x}) = \frac{1}{n}\sum_{j\neq i}^{n}\frac{(A_{ij} - \mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})},\quad \dot{\mathbf{\Psi}}_0 = \frac{1}{n}\sum_{j\neq i}^{n}\frac{\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})},\quad \dot{\mathbf{\Psi}}_{n,0} = \frac{1}{n}\sum_{j\neq i}^{n}\frac{\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{\widetilde{\mathbf{x}}_i^{\mathrm{T}}\mathbf{x}_{0j}(1 - \widetilde{\mathbf{x}}_i^{\mathrm{T}}\mathbf{x}_{0j})}.$$

The two main ingredients to prove the asymptotic normality of the one-step estimator are:

(i) $\dot{\mathbf{\Psi}}_{n,0} - \dot{\mathbf{\Psi}}_0 \overset{\mathbb{P}_0}{\to} 0$. To prove this convergence result, we denote

$$\mathbf{\Gamma}_j(\mathbf{x}) = \frac{\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})}.$$

Denote the (k,l)-th element of $\mathbf{\Gamma}_j$ by Γ_{jkl}. Clearly,

$$\frac{\partial \Gamma_{jkl}(\mathbf{x})}{\partial \mathbf{x}} = \frac{(1 - 2\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}x_{0jk}x_{jl}}{\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}^2},$$

where $\mathbf{x}_{0j} = [x_{0j1},\ldots,x_{0jd}]^{\mathrm{T}}$. This implies that there exists some $\epsilon > 0$, such that for all $\mathbf{x}$ in a neighborhood of $\mathbf{x}_{0i}$ within radius ϵ, denoted by $B(\mathbf{x}_{0i},\epsilon) = \{\mathbf{x}\in\mathbb{R}^d : \|\mathbf{x} - \mathbf{x}_{0i}\| < \epsilon\}$,

$$\max_{j\in[n]}\sup_{\mathbf{x}\in B(\mathbf{x}_{0i},\epsilon)}\sum_{k,l\in[d]}\left\|\frac{\partial \Gamma_{jkl}(\mathbf{x})}{\partial \mathbf{x}}\right\|_{\mathrm{F}} \leq C$$

for some constant C not depending on n. Therefore, we have, over the event $\{\widetilde{\mathbf{x}}_i \in B(\mathbf{x}_{0i},\epsilon)\}$, that

$$\begin{aligned}
\|\dot{\mathbf{\Psi}}_{n,0} - \dot{\mathbf{\Psi}}_0\|_{\mathrm{F}} &\leq \frac{1}{n}\sum_{j\neq i}\|\mathbf{\Gamma}_j(\widetilde{\mathbf{x}}_i) - \mathbf{\Gamma}_j(\mathbf{x}_{0i})\|_{\mathrm{F}}\\
&\leq \frac{1}{n}\sum_{j\neq i}\left\{\max_{j\in[n]}\sup_{\mathbf{x}\in B(\mathbf{x}_{0i},\epsilon)}\sum_{k,l\in[d]}\left\|\frac{\partial \Gamma_{jkl}(\mathbf{x})}{\partial \mathbf{x}}\right\|_{\mathrm{F}}\right\}\|\widetilde{\mathbf{x}}_i - \mathbf{x}_{0i}\|_2\\
&\leq C\|\widetilde{\mathbf{x}}_i - \mathbf{x}_{0i}\|_2.
\end{aligned}$$

Then, for any $t > 0$, we have

$$\begin{aligned}
\mathbb{P}_0\left(\|\dot{\mathbf{\Psi}}_{n,0} - \dot{\mathbf{\Psi}}_0\|_{\mathrm{F}} > t\right) &= \mathbb{P}_0\left(\|\dot{\mathbf{\Psi}}_{n,0} - \dot{\mathbf{\Psi}}_0\|_{\mathrm{F}} > t, \widetilde{\mathbf{x}}_i \in B(\mathbf{x}_{0i},\epsilon)\right)\\
&\quad + \mathbb{P}_0\left(\|\dot{\mathbf{\Psi}}_{n,0} - \dot{\mathbf{\Psi}}_0\|_{\mathrm{F}} > t, \widetilde{\mathbf{x}}_i \notin B(\mathbf{x}_{0i},\epsilon)\right)\\
&\leq \mathbb{P}_0\left(C\|\widetilde{\mathbf{x}}_i - \mathbf{x}_{0i}\|_2 > t, \widetilde{\mathbf{x}}_i \in B(\mathbf{x}_{0i},\epsilon)\right) + \mathbb{P}_0(\|\widetilde{\mathbf{x}}_i - \mathbf{x}_{0i}\|_2 \geq \epsilon)\\
&\leq \mathbb{P}_0\left(\|\widetilde{\mathbf{x}}_i - \mathbf{x}_{0i}\|_2 \geq \frac{t}{C}\right) + \mathbb{P}_0(\|\widetilde{\mathbf{x}}_i - \mathbf{x}_{0i}\|_2 \geq \epsilon) \to 0,
\end{aligned}$$

where we have used the fact that $\widetilde{\mathbf{x}}_i - \mathbf{x}_{0i} = O_{\mathbb{P}_0}(n^{-1/2})$ in the last line of the previous display. This completes the proof of (i).

(ii) $\sqrt{n}\|\mathbf{\Psi}_n(\widetilde{\mathbf{x}}_i) - \mathbf{\Psi}_n(\mathbf{x}_{0i}) - \dot{\mathbf{\Psi}}_0(\widetilde{\mathbf{x}}_i - \mathbf{x}_{0i})\| \overset{\mathbb{P}_0}{\to} 0$. To prove this convergence result, we need to argue that $\partial\mathbf{\Psi}_n(\mathbf{x}_{0i})/\partial\mathbf{x}^{\mathrm{T}}$ is sufficiently close to $\dot{\mathbf{\Psi}}_0$, and the Taylor expansion of $\mathbf{\Psi}_n$ at $\mathbf{x}_{0i}$ is valid. We first argue that $\partial\mathbf{\Psi}_n(\mathbf{x}_{0i})/\partial\mathbf{x}^{\mathrm{T}}$ is sufficiently close to $\dot{\mathbf{\Psi}}_0$. By the proof of Theorem 2,

$$\frac{\partial\mathbf{\Psi}_n^{\mathrm{T}}}{\partial\mathbf{x}}(\mathbf{x}) = -\frac{1}{n}\sum_{j\neq i}^{n}\frac{\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})} - \frac{1}{n}\sum_{j\neq i}^{n}\frac{(A_{ij}-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})(1-2\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}^2}.$$

Hence

$$\frac{\partial\mathbf{\Psi}_n^{\mathrm{T}}}{\partial\mathbf{x}}(\mathbf{x}_{0i}) = \dot{\mathbf{\Psi}}_0 - \frac{1}{n}\sum_{j\neq i}^{n}\frac{(A_{ij}-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})(1-2\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{\{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\}^2}.$$

Note that

$$\begin{aligned}\mathrm{var}_0\left[\mathrm{vec}\left\{\frac{\partial\mathbf{\Psi}_n^{\mathrm{T}}}{\partial\mathbf{x}}(\mathbf{x}_{0i}) - \dot{\mathbf{\Psi}}_0\right\}\right] &= \frac{1}{n^2}\sum_{j\neq i}^{n}\frac{(1-2\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})^2}{\{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\}^3}\mathrm{vec}(\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}})\mathrm{vec}(\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}})^{\mathrm{T}}\\ &\asymp \frac{1}{n},\end{aligned}$$

where $\mathrm{vec}(\mathbf{\Sigma})$ is the vectorization of the matrix $\mathbf{\Sigma}$ defined to be the vector formed by stacking the columns of $\mathbf{\Sigma}$ consecutively. Furthermore,

$$\sum_{j\neq i}^{n}\mathbb{E}_0\left[\left\|\frac{1}{\sqrt{n}}\frac{(A_{ij}-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})(1-2\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{\{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\}^2}\right\|_{\mathrm{F}}^3\right] \asymp \frac{1}{n^{1/2}} \to 0.$$

Therefore, by Lyapunov's central limit theorem,

$$\frac{\partial\mathbf{\Psi}_n^{\mathrm{T}}}{\partial\mathbf{x}}(\mathbf{x}_{0i}) - \dot{\mathbf{\Psi}}_0 = O_{\mathbb{P}_0}\left(\frac{1}{\sqrt{n}}\right).$$

We now establish the Taylor expansion of $\mathbf{\Psi}_n$ at $\mathbf{x}_{0i}$. Let Ψ_{nk} be the kth coordinate function of $\mathbf{\Psi}_n$, $k = 1,\ldots,d$. By the proof of Theorem 2,

$$\begin{aligned}\frac{\partial^2\Psi_{nk}}{\partial\mathbf{x}\partial\mathbf{x}^{\mathrm{T}}}(\mathbf{x}) &= \frac{1}{n}\sum_{j\neq i}^{n}\frac{x_{0jk}(1-2\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})}{\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}^2}\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}\\ &\quad + \frac{1}{n}\sum_{j\neq i}^{n}\frac{x_{0jk}\{(1-2\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}) + 2(A_{ij}-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}}{\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}^2}\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}\\ &\quad + \frac{1}{n}\sum_{j\neq i}^{n}\frac{x_{0jk}\{2(A_{ij}-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})(1-2\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})^2\}}{\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}^3}\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}.\end{aligned}$$

Therefore, there exists some $\epsilon > 0$, such that for all $\mathbf{x} \in B(\mathbf{x}_{0i}, \epsilon)$, we have

$$\Psi_{nk}(\mathbf{x}) - \Psi_{nk}(\mathbf{x}_{0i}) = \frac{\partial\Psi_{nk}}{\partial\mathbf{x}^{\mathrm{T}}}(\mathbf{x}_{0i})(\mathbf{x}-\mathbf{x}_{0i}) + \frac{1}{2}(\mathbf{x}-\mathbf{x}_{0i})^{\mathrm{T}}\left\{\frac{\partial^2\Psi_{nk}}{\partial\mathbf{x}\partial\mathbf{x}^{\mathrm{T}}}(\widetilde{\boldsymbol{\theta}}_k)\right\}(\mathbf{x}-\mathbf{x}_{0i}),$$

where $\widetilde{\boldsymbol{\theta}}_k$ lies on the line segment linking $\mathbf{x}_{0i}$ and $\mathbf{x}$. Again, by the proof of Theorem 2, we see that

$$\max_{k\in[d]}\sup_{\mathbf{x}\in B(\mathbf{x}_{0i},\epsilon)}\left\|\frac{\partial^2\Psi_{nk}}{\partial\mathbf{x}\partial\mathbf{x}^{\mathrm{T}}}(\mathbf{x})\right\|\leq C$$

for some constant $C>0$ not depending on n. Therefore,

$$\left\|\boldsymbol{\Psi}_n(\mathbf{x})-\boldsymbol{\Psi}_n(\mathbf{x}_{0i})-\frac{\partial\boldsymbol{\Psi}_n}{\partial\mathbf{x}^{\mathrm{T}}}(\mathbf{x}_{0i})(\mathbf{x}-\mathbf{x}_{0i})\right\|_{\mathrm{F}}\leq Cd\|\mathbf{x}-\mathbf{x}_{0i}\|_2^2 \tag{C.5}$$

whenever $\mathbf{x}\in B(\mathbf{x}_{0i},\epsilon)$. Now we are able to prove (ii). Write

$$\begin{aligned}
&\sqrt{n}\|\boldsymbol{\Psi}_n(\widetilde{\mathbf{x}}_i)-\boldsymbol{\Psi}_n(\mathbf{x}_{0i})-\dot{\boldsymbol{\Psi}}_0(\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i})\|_2\\
&\quad\leq\sqrt{n}\left\|\boldsymbol{\Psi}_n(\widetilde{\mathbf{x}}_i)-\boldsymbol{\Psi}_n(\mathbf{x}_{0i})-\frac{\partial\boldsymbol{\Psi}_n}{\partial\mathbf{x}^{\mathrm{T}}}(\mathbf{x}_{0i})(\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i})\right\|_2+\sqrt{n}\left\|\frac{\partial\boldsymbol{\Psi}_n^{\mathrm{T}}}{\partial\mathbf{x}}(\mathbf{x}_{0i})-\dot{\boldsymbol{\Psi}}_0\right\|_2\|\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i}\|_2\\
&\quad=\sqrt{n}\left\|\boldsymbol{\Psi}_n(\widetilde{\mathbf{x}}_i)-\boldsymbol{\Psi}_n(\mathbf{x}_{0i})-\frac{\partial\boldsymbol{\Psi}_n}{\partial\mathbf{x}^{\mathrm{T}}}(\mathbf{x}_{0i})(\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i})\right\|_2+O_{\mathbb{P}_0}(n^{-1/2}).
\end{aligned}$$

Now over the event $\{\widetilde{\mathbf{x}}_i\in B(\mathbf{x}_{0i},\epsilon)\}$, we can apply the Taylor expansion (C.5) and write

$$\left\|\boldsymbol{\Psi}_n(\widetilde{\mathbf{x}}_i)-\boldsymbol{\Psi}_n(\mathbf{x}_{0i})-\frac{\partial\boldsymbol{\Psi}_n}{\partial\mathbf{x}^{\mathrm{T}}}(\mathbf{x}_{0i})(\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i})\right\|_2\leq Cd\|\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i}\|_2^2.$$

Therefore, for any $t>0$,

$$\begin{aligned}
&\mathbb{P}_0\left(\sqrt{n}\left\|\boldsymbol{\Psi}_n(\widetilde{\mathbf{x}}_i)-\boldsymbol{\Psi}_n(\mathbf{x}_{0i})-\frac{\partial\boldsymbol{\Psi}_n}{\partial\mathbf{x}^{\mathrm{T}}}(\mathbf{x}_{0i})(\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i})\right\|_2>t\right)\\
&\quad\leq\mathbb{P}_0\left(\sqrt{n}\left\|\boldsymbol{\Psi}_n(\widetilde{\mathbf{x}}_i)-\boldsymbol{\Psi}_n(\mathbf{x}_{0i})-\frac{\partial\boldsymbol{\Psi}_n}{\partial\mathbf{x}^{\mathrm{T}}}(\mathbf{x}_{0i})(\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i})\right\|_2>t,\widetilde{\mathbf{x}}_i\in B(\mathbf{x}_{0i},\epsilon)\right)\\
&\qquad+\mathbb{P}_0\left(\|\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i}\|_2>\epsilon\right)\\
&\quad\leq\mathbb{P}_0\left(Cd\|\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i}\|_2^2>\frac{t}{\sqrt{n}}\right)+\mathbb{P}_0\left(\|\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i}\|_2>\epsilon\right)\to 0,
\end{aligned}$$

where we have applied the fact that $\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i}=O_{\mathbb{P}_0}(n^{-1/2})$ to the last line of the previous display. The claim (ii) is thus established.

We are now in a position to prove Theorem 3. By definition and (ii), we have

$$\begin{aligned}
\boldsymbol{\Psi}_{n,0}\sqrt{n}(\widehat{\mathbf{x}}_i^{(\mathrm{OS})}-\mathbf{x}_{0i})&=\boldsymbol{\Psi}_{n,0}\sqrt{n}(\widetilde{\mathbf{x}}_i-\boldsymbol{\Psi}_{n,0}^{-1}\boldsymbol{\Psi}_n(\widetilde{\mathbf{x}}_i)-\mathbf{x}_{0i})\\
&=\boldsymbol{\Psi}_{n,0}\sqrt{n}(\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i})-\sqrt{n}\{\boldsymbol{\Psi}_n(\widetilde{\mathbf{x}}_i)-\boldsymbol{\Psi}_n(\mathbf{x}_{0i})\}+\sqrt{n}\boldsymbol{\Psi}_n(\mathbf{x}_{0i})\\
&=(\boldsymbol{\Psi}_{n,0}-\dot{\boldsymbol{\Psi}}_0)\sqrt{n}(\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i})+\sqrt{n}\boldsymbol{\Psi}_n(\mathbf{x}_{0i})+o_{\mathbb{P}_0}(1).
\end{aligned}$$

By (i) and the fact that $\sqrt{n}(\widetilde{\mathbf{x}}_i - \mathbf{x}_{0i}) = O_{\mathbb{P}_0}(1)$, we see that the first term on the right-hand side of the previous display is also $o_{\mathbb{P}_0}(1)$. Therefore,

$$\mathbf{\Psi}_{n,0}\sqrt{n}(\widehat{\mathbf{x}}_i^{(\mathrm{OS})} - \mathbf{x}_{0i}) = \sqrt{n}\mathbf{\Psi}_n(\mathbf{x}_{0i}) + o_{\mathbb{P}_0}(1) = \frac{1}{\sqrt{n}}\sum_{j\neq i}^n \frac{(A_{ij} - \mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})} + o_{\mathbb{P}_0}(1).$$

Using the proof of Theorem 2, we see that the first term in the right-hand side of the above display converges to $\mathrm{N}(\mathbf{0}, \mathbf{G}(\mathbf{x}_{0i}))$ in distribution. Therefore, by (i), the fact that $\dot{\mathbf{\Psi}}_0 \to \mathbf{G}(\mathbf{x}_{0i})$ as $n\to\infty$, and Slutsky's theorem, we conclude that

$$\sqrt{n}(\widehat{\mathbf{x}}_i^{(\mathrm{OS})} - \mathbf{x}_{0i}) = \mathbf{\Psi}_{n,0}^{-1}\frac{1}{\sqrt{n}}\sum_{j\neq i}^n \frac{(A_{ij} - \mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})} + o_{\mathbb{P}_0}(1) \xrightarrow{\mathcal{L}} \mathrm{N}(\mathbf{0}, \mathbf{G}(\mathbf{x}_{0i})^{-1}).$$

□

D Proofs of Theorems 4 and 5

D.1 Some Technical Lemmas for the One-Step Procedure

Before proceeding to the proof of Theorem 4, we first establish a collection of technical lemmas for bounding the remainder $\widehat{\mathbf{R}}_i$ in (3.5).

Lemma D.1 *Let $\mathbf{A} \sim \mathrm{RDPG}(\mathbf{X}_0)$ and assume the conditions in Theorem 4 holds. Let an estimator $\widetilde{\mathbf{X}} \in \mathbb{R}^{n\times d}$ satisfy the approximate linearization property (3.3) with an orthogonal matrix $\mathbf{W} \in \mathbb{O}(d)$ (possibly depending on n). Then*

$$\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}_0\|_{2\to\infty} = O_{\mathbb{P}_0}\left(\frac{(\log n)^{(1\vee\omega)/2}}{\sqrt{n\rho_n}}\right).$$

Proof. The proof of this lemma is similar to that of Lemma 2 in Tang et al. (2017b), except that we consider the case where a sparsity factor ρ_n is taken into account, and the proof is presented here for the sake of completeness. Recall from (3.3) that

$$\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}_0\|_{2\to\infty} \leq \rho_n^{-1/2}\sqrt{d}\max_{i\in[n],k\in[d]}\left|\sum_{j=1}^n (A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\zeta_{ijk}\right| + \|\widetilde{\mathbf{R}}\|_{\mathrm{F}}.$$

where $\boldsymbol{\zeta}_{ij} = [\zeta_{ij1},\ldots,\zeta_{ijd}]^{\mathrm{T}} \in \mathbb{R}^d$. By Hoeffding's inequality, the union bound, and the condition that $\sup_{i,j\in[n]}\|\boldsymbol{\zeta}_{ij}\| \lesssim 1/n$, for any $t>0$, we have,

$$\mathbb{P}_0\left(\max_{i\in[n],k\in[k]}\left|\sum_{j=1}^n (A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\zeta_{ijk}\right| > t\right) \leq 2nd\exp\left(-\frac{2t^2}{\sum_{j=1}^n\zeta_{ijk}^2}\right) = 2nd\exp\left\{-Knt^2\right\}$$

for some constant $K>0$. Therefore, for any $c>0$, there exists some constant $C_c>0$ and $n_c\in\mathbb{N}_+$, such that for all $n \geq n_c$.

$$\mathbb{P}_0\left(\rho_n^{-1/2}\sqrt{d}\max_{i\in[n],k\in[d]}\left|\sum_{j=1}^n (A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\zeta_{ijk}\right| > C_c\sqrt{\frac{\log n}{n\rho_n}}\right) \leq \frac{1}{n^c}.$$

This shows that

$$\rho_n^{-1/2}\sqrt{d}\max_{i\in[n],k\in[d]}\left|\sum_{j=1}^n (A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\zeta_{ijk}\right| = O_{\mathbb{P}_0}\left(\sqrt{\frac{\log n}{n\rho_n}}\right).$$

The proof is completed by applying the condition that $\|\widetilde{\mathbf{R}}\|_{\mathrm{F}} = O_{\mathbb{P}_0}\left((n\rho_n)^{-1/2}(\log n)^{\omega/2}\right)$. $\square$

Lemma D.2 *Let* $\mathbf{A} \sim \mathrm{RDPG}(\mathbf{X}_0)$ *with sparsity factor* ρ_n*, and assume the conditions of Theorem 4 hold. Let an estimator* $\widetilde{\mathbf{X}} \in \mathbb{R}^{n\times d}$ *satisfy the approximate linearization property (3.3) with an orthogonal matrix* $\mathbf{W} \in \mathbb{O}(d)$ *(possibly depending on* n*). Then*

$$\max_{i\in[n]}\left\|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n \frac{\rho_n\mathbf{x}_{0j}\mathbf{x}_{0i}^{\mathrm{T}}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \mathbf{x}_{0j})\right\| = O_{\mathbb{P}_0}\left(\frac{(\log n)^{(1\vee\omega)/2}}{n\rho_n^{1/2}}\right).$$

Proof. Recall by condition (3.3) that for any $j \in [n]$,

$$[\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_{0j}]_k = \rho_n^{-1/2}\sum_{a=1}^n (A_{ja}-\rho_n\mathbf{x}_{0j}^{\mathrm{T}}\mathbf{x}_{0a})\zeta_{iak} + \widetilde{R}_{jk}, \quad k = 1,2,\ldots,d,$$

where $\boldsymbol{\zeta}_{ij} = [\zeta_{ij1},\ldots,\zeta_{ijd}]^{\mathrm{T}}$. It follows that for $k = 1,2,\ldots,d$,

$$\begin{aligned}
&\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\sum_{s=1}^d \frac{\rho_n x_{0jk}x_{0is}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}[\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \mathbf{x}_{0j}]_s\\
&\quad= \frac{1}{n\sqrt{\rho_n}}\sum_{s=1}^d\sum_{j=1}^n\sum_{a=1}^n \frac{x_{0jk}x_{0is}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}(A_{ja}-\rho_n\mathbf{x}_{0j}^{\mathrm{T}}\mathbf{x}_{0a})\zeta_{ias}\\
&\quad\quad+ \frac{1}{n}\sum_{j=1}^n\sum_{s=1}^d \frac{x_{0jk}x_{0is}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\widetilde{R}_{js}\\
&\quad= \frac{1}{n\sqrt{\rho_n}}\sum_{s=1}^d\left\{\sum_{j<a} Z_{iksja} + \sum_{j>a} Z_{iksja} + \sum_{j=1}^n Z_{iksjj}\right\} + \frac{1}{n}\sum_{j=1}^n\sum_{s=1}^d \frac{x_{0jk}x_{0is}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\widetilde{R}_{js},
\end{aligned}$$

where

$$Z_{iksja} = \frac{x_{0jk}x_{0is}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}(A_{ja}-\rho_n\mathbf{x}_{0j}^{\mathrm{T}}\mathbf{x}_{0a})\zeta_{ias}.$$

Observe that by Hoeffding's inequality and the union bound,

$$\begin{aligned}\mathbb{P}_0\left(\frac{1}{n\sqrt{\rho_n}}\max_{i\in[n]}\left|\sum_{j<a}Z_{iksja}\right|>t\right)&\leq 2n\exp\left[-2n^2\rho_nt^2\left\{\sum_{j<a}\left(\frac{\zeta_{ias}x_{0jk}x_{0is}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\right)^2\right\}^{-1}\right]\\&\leq 2n\exp\left(-Kn^2\rho_nt^2\right).\end{aligned}$$

This shows that

$$\frac{1}{n\sqrt{\rho_n}}\max_{i\in[n]}\left|\sum_{j<a}Z_{iksja}\right|=O_{\mathbb{P}_0}\left(\sqrt{\frac{\log n}{n^2\rho_n}}\right),$$

and hence, a similar argument yields that

$$\frac{1}{n\sqrt{\rho_n}}\max_{i\in[n]}\left|\sum_{j=1}^n\sum_{a=1}^nZ_{iksja}\right|=O_{\mathbb{P}_0}\left(\sqrt{\frac{\log n}{n^2\rho_n}}\right).$$

In addition, by the fact that $\|\widetilde{\mathbf{R}}\|_{\mathrm{F}}^2=O_{\mathbb{P}_0}((n\rho_n)^{-1}(\log n)^{\omega})$ we have

$$\begin{aligned}&\max_{i\in[n]}\left|\frac{1}{n}\sum_{j=1}^n\sum_{s=1}^d\frac{x_{0jk}x_{0is}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\widetilde{R}_{js}\right|\\&\quad\leq\frac{1}{n}\left[\sum_{j=1}^n\sum_{s=1}^d\max_{i\in[n]}\left\{\frac{x_{0jk}x_{0is}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\right\}^2\right]^{1/2}\|\widetilde{\mathbf{R}}\|_{\mathrm{F}}\lesssim\frac{1}{\sqrt{n}}\|\widetilde{\mathbf{R}}\|_{\mathrm{F}}=O_{\mathbb{P}_0}\left(\frac{(\log n)^{\omega/2}}{n\rho_n^{1/2}}\right).\end{aligned}$$

Therefore, we conclude that

$$\begin{aligned}&\max_{i\in[n]}\left\|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\frac{\rho_n\mathbf{x}_{0j}\mathbf{x}_{0i}^{\mathrm{T}}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\mathbf{x}_{0j})\right\|\\&\quad\lesssim\sum_{k=1}^d\sum_{s=1}^d\max_{i\in[n]}\left|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\sum_{a=1}^nZ_{iksja}\right|+\sum_{k=1}^d\max_{i\in[n]}\left|\frac{1}{n}\sum_{j=1}^n\sum_{s=1}^d\frac{x_{0jk}x_{0is}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\widetilde{R}_{js}\right|\\&\quad=O_{\mathbb{P}_0}\left(\frac{(\log n)^{(1\vee\omega)/2}}{n\rho_n^{1/2}}\right),\end{aligned}$$

and the proof is thus completed. □

Lemma D.3 *Let $\mathbf{A}\sim\mathrm{RDPG}(\mathbf{X}_0)$ with sparsity factor ρ_n and assume the conditions of Theorem 4 hold.Let an estimator $\widetilde{\mathbf{X}}\in\mathbb{R}^{n\times d}$ satisfy the approximate linearization property (3.3) with an orthogonal matrix $\mathbf{W}\in\mathbb{O}(d)$ (possibly depending on n). Suppose $\{\boldsymbol{\alpha}_{ijk}:i,j\in[n],k\in[d]\}$ is a collection of deterministic* *vectors in* $\mathbb{R}^d$ *with* $\sup_{i,j\in[n],k\in[d]}\|\boldsymbol{\alpha}_{ijk}\|<\infty$*. Then*

$$\max_{i\in[n],k\in[d]}\frac{1}{n\sqrt{\rho_n}}\left|\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\alpha}_{ijk}^{\mathrm{T}}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i})\right|=O_{\mathbb{P}_0}\left(\frac{(\log n)^{1/2+(1\vee\omega)/2}}{n\rho_n}\right).$$

Proof. First observe that by Cauchy-Schwarz inequality,

$$\begin{aligned}&\left|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\alpha}_{ijk}^{\mathrm{T}}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i})\right|\\&\quad\le\frac{1}{\sqrt{\rho_n}}\|\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_{0i}\|\left\|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\boldsymbol{\alpha}_{ijk}(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\right\|.\end{aligned}$$

By Hoeffding's inequality and the union bound, for all $r=1,2,\ldots,d$,

$$\mathbb{P}_0\left(\max_{i\in[n],k\in[d]}\left|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n[\boldsymbol{\alpha}_{ijk}]_r(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\right|>t\right)\le 2nd\exp\left(-Kn\rho_nt^2\right)$$

for some constant $K>0$. This shows that

$$\max_{i\in[n],k\in[d]}\left|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n[\boldsymbol{\alpha}_{ijk}]_r(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\right|=O_{\mathbb{P}_0}\left(\sqrt{\frac{\log n}{n\rho_n}}\right)$$

Hence, we conclude from Lemma D.1 that

$$\begin{aligned}&\max_{i\in[n]}\left|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\alpha}_{ijk}^{\mathrm{T}}(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\mathbf{x}_{0i})\right|\\&\quad\lesssim\frac{\|\widetilde{\mathbf{X}}\mathbf{W}-\rho_n^{1/2}\mathbf{X}_0\|_{2\to\infty}}{\sqrt{\rho_n}}\sum_{r=1}^d\left|\max_{i\in[n],k\in[d]}\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n[\boldsymbol{\alpha}_{ijk}]_r(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\right|=O_{\mathbb{P}_0}\left(\frac{(\log n)^{(1/2)+(1\vee\omega)/2}}{n\rho_n}\right).\end{aligned}$$

The proof is thus completed. ∎

Lemma D.4 *Let* $\mathbf{A}\sim\mathrm{RDPG}(\mathbf{X}_0)$ *with sparsity factor* ρ_n*, and assume the conditions of Theorem 4 hold. Let an estimator* $\widetilde{\mathbf{X}}\in\mathbb{R}^{n\times d}$ *satisfy the approximate linearization property (3.3) with an orthogonal matrix* $\mathbf{W}\in\mathbb{O}(d)$ *(possibly depending on* n*). Suppose* $\{\boldsymbol{\beta}_{ijk}:i,j\in[n],k\in[d]\}$ *is a collection of deterministic vectors in* $\mathbb{R}^d$ *such that* $\sup_{i,j\in[n],k\in[d]}\|\boldsymbol{\beta}_{ijk}\|<\infty$*. Then for each individual* $i\in[n]$*,*

$$\left|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\mathbf{x}_{0j})^{\mathrm{T}}\boldsymbol{\beta}_{ijk}\right|=O_{\mathbb{P}_0}\left(\frac{(\log n)^{\omega/2}}{n\rho_n^{3/2}}\right)$$

and

$$\sum_{i=1}^{n}\left\{\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^{n}(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\mathbf{x}_{0j})^{\mathrm{T}}\boldsymbol{\beta}_{ijk}\right\}^2=O_{\mathbb{P}_0}\left(\frac{(\log n)^{\omega}}{n\rho_n^3}\right).$$

Proof. Denote $\boldsymbol{\beta}_{ijk}=[\beta_{ijk1},\ldots,\beta_{ijkd}]^{\mathrm{T}}$. Recall the approximate linearization property (3.3) that

$$[\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\rho_n^{1/2}\mathbf{x}_{0j}]_s=\rho_n^{-1/2}\sum_{a=1}^{n}(A_{ja}-\rho_n\mathbf{x}_{0j}^{\mathrm{T}}\mathbf{x}_{0a})\zeta_{ias}+\widetilde{R}_{js},\quad s=1,2,\ldots,d,$$

where $\boldsymbol{\zeta}_{ia}=[\zeta_{ia1},\ldots,\zeta_{iad}]^{\mathrm{T}}$. It follows that

$$\begin{aligned}Q_{ik}&:=\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^{n}(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\mathbf{x}_{0j})^{\mathrm{T}}\boldsymbol{\beta}_{ijk}\\&=\frac{1}{n\rho_n^{3/2}}\sum_{s=1}^{d}\sum_{j=1}^{n}\sum_{a=1}^{n}z_{iksja}+\frac{1}{n\rho_n}\sum_{s=1}^{d}\sum_{j=1}^{n}(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\beta_{ijks}\widetilde{R}_{js},\end{aligned}$$

where $z_{iksja}=\zeta_{ias}\beta_{ijks}(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})(A_{ja}-\rho_n\mathbf{x}_{0j}^{\mathrm{T}}\mathbf{x}_{0a})$. Clearly,

$$\frac{1}{n^2\rho_n^3}\max_{i\in[n]}\mathbb{E}_0\left\{\left(\sum_{s=1}^{d}\sum_{j=1}^{n}\sum_{a=1}^{n}z_{iksja}\right)^2\right\}\lesssim\frac{1}{n^2\rho_n^3}\sum_{s=1}^{d}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{h=1}^{n}\sum_{b=1}^{n}\max_{i\in[n]}\mathbb{E}_0(z_{iksja}z_{ikshb}).$$

We now argue that the summation $\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{h=1}^{n}\sum_{b=1}^{n}\max_{i\in[n]}\max_{i\in[n]}\mathbb{E}_0(z_{iksja}z_{ikshb})$ is upper bounded by $\sup_{i,j}\|\boldsymbol{\zeta}_{ij}\|^2n^2\rho_n$ up to a multiplicative constant. Note that as the indices j,a,h,b ranging over $[n]$, $\mathbb{E}_0(z_{iksja}z_{ikshb})$ is nonzero only if the cardinality of the collection of random variables $\{A_{ij},A_{ih},A_{aj},A_{bh}\}$ is 2 or 1. These cases occur only if either one of the following cases happens:

1. A_{ij} and A_{ih} are the same random variable, and A_{aj},A_{bh} are the same random variable. This happens only if one the following cases occur:

   (a) $(i,j)=(i,h),(a,j)=(b,h)\Rightarrow j=h,\ a=b$, and the number of terms is $O(n^2)$;

   (b) $(i,j)=(h,i),(a,j)=(b,h)\Rightarrow i=j=h,\ a=b$, and the number of terms is $O(n)$;

   (c) $(i,j)=(i,h),(a,j)=(h,b)\Rightarrow j=h=a=b$, and the number of terms is $O(n)$;

   (d) $(i,j)=(h,i),(a,j)=(h,b)\Rightarrow i=j=h=a=b$, and the number of terms is 1;

2. A_{ij} and A_{aj} are the same random variable, and A_{ih},A_{bh} are the same random variable. This happens only if one the following cases occur:

   (a) $(i,j)=(a,j),(i,h)=(b,h)\Rightarrow i=a=b$, and the number of terms is $O(n^2)$;

   (b) $(i,j)=(j,a),(i,h)=(b,h)\Rightarrow i=j=a=b$, and the number of terms is $O(n)$;

   (c) $(i,j)=(a,j),(i,h)=(h,b)\Rightarrow i=h=a=b$, and the number of terms is $O(n)$;

   (d) $(i,j)=(j,a),(i,h)=(h,b)\Rightarrow i=j=h=a=b$, and the number of terms is 1;

3. A_{ij} and A_{bh} are the same random variable, and A_{ih}, A_{aj} are the same random variable. This happens only if one the following cases occur:

   (a) $(i,j) = (b,h), (i,h) = (a,j) \Rightarrow i = b = a$, $h = j$, and the number of terms is $O(n)$;

   (b) $(i,j) = (h,b), (i,h) = (a,j) \Rightarrow i = j = h = a = b$, and the number of terms is 1;

   (c) $(i,j) = (b,h), (i,h) = (j,a) \Rightarrow j = j = h = a = b$, and the number of terms is 1;

   (d) $(i,j) = (h,b), (i,h) = (j,a) \Rightarrow i = j = h = a = b$, and the number of terms is 1.

Therefore, the number of nonzero terms in the summation

$$\sum_{j=1}^n\sum_{a=1}^n\sum_{h=1}^n\sum_{b=1}^n \max_{i\in[n]}\max_{i\in[n]}\mathbb{E}_0(z_{iksja}z_{ikshb})$$

is $O(n^2)$. Furthermore, the centered second and fourth moments of Bernoulli$(\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})$ is upper bounded by ρ_n. Therefore, we obtain that

$$\frac{1}{n^2\rho_n^3}\max_{i\in[n]}\mathbb{E}_0\left\{\left(\sum_{s=1}^d\sum_{j=1}^n\sum_{a=1}^n z_{iksja}\right)^2\right\} \lesssim \frac{1}{n^2\rho_n^3}\sup_{i,j\in[n]}\|\boldsymbol{\zeta}_{ij}\|^2 n^2\rho_n \lesssim \frac{1}{(n\rho_n)^2}.$$

In addition,

$$\begin{aligned}\max_{i\in[n]}\left|\frac{1}{n\rho_n}\sum_{s=1}^d\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\beta_{ijks}R_{js}\right| &\leq \frac{1}{n\rho_n}\max_{i\in[n]}\left\{\sum_{s=1}^d\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})^2\beta_{jks}^2\right\}^{1/2}\|\widetilde{\mathbf{R}}\|_{\mathrm{F}}\\ &= O_{\mathbb{P}_0}\left(\frac{(\log n)^{\omega/2}}{n\rho_n^{3/2}}\right).\end{aligned}$$

Namely, this implies that for each individual $i \in [n]$,

$$\left|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \mathbf{x}_{0j})^{\mathrm{T}}\boldsymbol{\beta}_{ijk}\right| = O_{\mathbb{P}_0}\left(\frac{(\log n)^{\omega/2}}{n\rho_n^{3/2}}\right)$$

Furthermore,

$$\sum_{i=1}^n\mathbb{E}_0\left\{\left(\frac{1}{n\rho_n^{3/2}}\sum_{s=1}^d\sum_{j=1}^n\sum_{a=1}^n z_{iksja}\right)^2\right\} \leq \frac{1}{n\rho_n^3}\max_{i\in[n]}\mathbb{E}_0\left\{\left(\sum_{s=1}^d\sum_{j=1}^n\sum_{a=1}^n z_{iksja}\right)^2\right\} \lesssim \frac{1}{n\rho_n^2},$$

implying that

$$\sum_{i=1}^n\left\{\frac{1}{n\rho_n^{3/2}}\sum_{s=1}^d\sum_{j=1}^n\sum_{a=1}^n z_{iksja}\right\}^2 = O_{\mathbb{P}_0}\left(\frac{1}{n\rho_n^2}\right)$$

by Markov's inequality. Therefore, we conclude that

$$\sum_{i=1}^n\left\{\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})(\rho_n^{-1/2}\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\mathbf{x}_{0j})^{\mathrm{T}}\boldsymbol{\beta}_{ijk}\right\}^2$$

$$\leq 2\sum_{i=1}^n\left\{\frac{1}{n\rho_n^{3/2}}\sum_{s=1}^d\sum_{j=1}^n\sum_{a=1}^n z_{iksja}\right\}^2+2n\max_{i\in[n]}\left|\frac{1}{n\rho_n}\sum_{s=1}^d\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\beta_{ijks}R_{js}\right|^2=O_{\mathbb{P}_0}\left(\frac{(\log n)^\omega}{n\rho_n^3}\right).$$

The proof is thus completed. □

Lemma D.5 *Let* $\mathbf{A}\sim\mathrm{RDPG}(\mathbf{X}_0)$ *and assume the conditions in Theorem 5 holds. Denote* $Z=Z(\mathbf{A})=\sum_{i=1}^n\|\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij}\|^2$*, where* $\{\boldsymbol{\gamma}_{ij}:i,j\in[n]\}$ *is a collection of deterministic vectors in* $\mathbb{R}^d$ *such that* $\sup_{i,j\in[n]}\|\boldsymbol{\gamma}_{ij}\|\lesssim(n\sqrt{\rho_n})^{-1}$*. Then* $Z=\mathbb{E}_0(Z)+o_{\mathbb{P}_0}(1)$*.*

Proof. The proof of Lemma D.5 relies on the following logarithmic Sobolev concentration inequality:

Lemma D.6 (Theorem 6.7 in Boucheron et al., 2013) *Let* $\mathbf{A},\mathbf{A}'\in\{0,1\}^{n\times n}$ *be two symmetric hollow random adjacency matrices and* $Z=Z(\mathbf{A})$ *be a measurable function of of* $\mathbf{A}$*. Denote by* $\mathbf{A}^{(kl)}$ *the adjacency matrix obtained by replacing the* (k,l) *and* (l,k) *entries of* $\mathbf{A}$ *by those of* $\mathbf{A}'$*, and* $Z_{kl}=Z(\mathbf{A}^{(kl)})$*. If there exists a constant* $v>0$ *such that*

$$\mathbb{P}\left(\sum_{k<l}(Z-Z_{kl})^2>v\right)\leq\eta,$$

then for all $\epsilon>0$*,* $\mathbb{P}(|Z-\mathbb{E}(Z)|>t)\leq 2\exp\{-t^2/(2v)\}+\eta$*.*

Let $\mathbf{A}'$ be another symmetric hollow random adjacency matrix. Denote by $\mathbf{A}^{(kl)}$ the adjacency matrix obtained by replacing the (k,l) and (l,k) entries of $\mathbf{A}$ by those of $\mathbf{A}'$, and $Z_{kl}=Z(\mathbf{A}^{(kl)})$ Since that $\mathbf{A}$ and $\mathbf{A}^{(kl)}$ only differs by the (k,l) and (l,k) entries, and that the entries of $\mathbf{A}$ and $\mathbf{A}'$ are binary, we see that when $Z-Z_{kl}\neq 0$, $(A_{kl}-A'_{kl})(Z-Z_{kl})=C_{1kl}+C_{2kl}+c_{kl}$, where

$$C_{1kl}=2\sum_{a=1}^n(A_{ka}-\rho_n\mathbf{x}_{0k}^{\mathrm{T}}\mathbf{x}_{0a})\boldsymbol{\gamma}_{kl}^{\mathrm{T}}\boldsymbol{\gamma}_{ka},\quad C_{2kl}=2\sum_{a=1}^n(A_{la}-\rho_n\mathbf{x}_{0l}^{\mathrm{T}}\mathbf{x}_{0a})\boldsymbol{\gamma}_{lk}^{\mathrm{T}}\boldsymbol{\gamma}_{la},$$

and $c_{kl}=(1-2\rho_n\mathbf{x}_{0k}^{\mathrm{T}}\mathbf{x}_{0l})(\|\boldsymbol{\gamma}_{kl}\|^2+\|\boldsymbol{\gamma}_{lk}\|^2)-2(A_{kl}-\rho_n\mathbf{x}_{0k}^{\mathrm{T}}\mathbf{x}_{0l})\|\boldsymbol{\gamma}_{kl}\|^2-2(A_{lk}-\rho_n\mathbf{x}_{0l}^{\mathrm{T}}\mathbf{x}_{0k})\|\boldsymbol{\gamma}_{lk}\|^2$. Since

$$\begin{aligned}\sum_{k<l}\mathbb{E}_0(C_{1kl}^2)&=4\sum_{k<l}\sum_{a=1}^n\sum_{b=1}^n\mathbb{E}_0\{(A_{ka}-\rho_n\mathbf{x}_{0k}^{\mathrm{T}}\mathbf{x}_{0a})(A_{kb}-\rho_n\mathbf{x}_{0k}^{\mathrm{T}}\mathbf{x}_{0b})\}(\boldsymbol{\gamma}_{kl}^{\mathrm{T}}\boldsymbol{\gamma}_{ka})(\boldsymbol{\gamma}_{kl}^{\mathrm{T}}\boldsymbol{\gamma}_{kb})\\&=4\sum_{k<l}\sum_{a=1}^n\mathbb{E}_0\{(A_{ka}-\rho_n\mathbf{x}_{0k}^{\mathrm{T}}\mathbf{x}_{0a})^2\}(\boldsymbol{\gamma}_{kl}^{\mathrm{T}}\boldsymbol{\gamma}_{ka})^2\\&\leq 4\sum_{k<l}\sum_{a=1}^n\rho_n\mathbf{x}_{0k}^{\mathrm{T}}\mathbf{x}_{0a}(1-\rho_n\mathbf{x}_{0k}^{\mathrm{T}}\mathbf{x}_{0a})\|\boldsymbol{\gamma}_{kl}\|^2\|\boldsymbol{\gamma}_{ka}\|^2\lesssim\frac{1}{n\rho_n},\end{aligned}$$

$$
\begin{aligned}
\sum_{k<l}\mathbb{E}_0(C_{2kl}^2) &= 4\sum_{k<l}\sum_{a=1}^n\sum_{b=1}^n\mathbb{E}_0\{(A_{la}-\rho_n\mathbf{x}_{0l}^{\mathrm{T}}\mathbf{x}_{0a})(A_{lb}-\rho_n\mathbf{x}_{0l}^{\mathrm{T}}\mathbf{x}_{0b})\}(\boldsymbol{\gamma}_{lk}^{\mathrm{T}}\boldsymbol{\gamma}_{la})(\boldsymbol{\gamma}_{lk}^{\mathrm{T}}\boldsymbol{\gamma}_{lb})\\
&= 4\sum_{k<l}\sum_{a=1}^n\mathbb{E}_0\{(A_{la}-\rho_n\mathbf{x}_{0l}^{\mathrm{T}}\mathbf{x}_{0a})^2\}(\boldsymbol{\gamma}_{lk}^{\mathrm{T}}\boldsymbol{\gamma}_{la})^2 \lesssim \frac{1}{n\rho_n},\\
\sum_{k<l}\mathbb{E}_0(c_{kl}^2) &\leq 6\sum_{k<l}(1-2\rho_n\mathbf{x}_{0k}^{\mathrm{T}}\mathbf{x}_{0l})(\|\boldsymbol{\gamma}_{kl}\|^4+\|\boldsymbol{\gamma}_{lk}\|^4) + 6\sum_{k<l}\mathbb{E}_0\{(A_{kl}-\rho_n\mathbf{x}_{0k}^{\mathrm{T}}\mathbf{x}_{0l})^2\}(\|\boldsymbol{\gamma}_{kl}\|^4+\|\boldsymbol{\gamma}_{lk}\|^4)\\
&\lesssim \frac{1}{n^2\rho_n^2}+\frac{\rho_n}{n^2\rho_n^2}\lesssim\frac{1}{n^2\rho_n},
\end{aligned}
$$

we conclude that $\mathbb{E}_0\{\sum_{k<l}(Z-Z_{kl})^2\}\leq C/(n\rho_n)$ for some constant $C>0$. Therefore, by Markov's inequality,

$$
\mathbb{P}\left(\sum_{k<l}(Z-Z_{kl})^2>\frac{1}{\log n}\right)\leq\frac{C\log n}{n\rho_n}\leq\frac{C(\log n)^2}{n\rho_n^5}\to 0.
$$

Invoking Lemma D.6, we obtain that

$$
\mathbb{P}_0\left(|Z-\mathbb{E}_0(Z)|>\epsilon\right)=\mathbb{P}_0\left[\left|Z-\frac{1}{n}\sum_{i=1}^n\mathrm{tr}\{\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\}\right|>\epsilon\right]\leq 2\exp\left(-\frac{1}{2}\epsilon^2\log n\right)+\frac{C\log n}{n\rho_n}\to 0
$$

for all $\epsilon>0$. The proof is thus completed. □

Lemma D.7 *Let $\mathbf{G}_n(\mathbf{x})$ be defined as in Theorem 4, $\mathbf{G}(\mathbf{x})$ defined as in Theorem 2, $\widetilde{\mathbf{G}}(\mathbf{x})$ be defined as in Theorem 9. Denote*

$$
\widetilde{\mathbf{G}}_n(\mathbf{x})=\frac{1}{\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}}\left(\mathbf{I}_d-\frac{\mathbf{x}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}_n}\right)\mathbf{G}_n(\mathbf{x})^{-1}\left(\mathbf{I}_d-\frac{\mathbf{x}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}_n}\right),
$$

where $\boldsymbol{\mu}_n=(1/n)\sum_{i=1}^n\mathbf{x}_{0i}$. Let $\mathcal{X}(\delta)$ be the set of all $\mathbf{x}\in\mathcal{X}$ such that any $\mathbf{x},\mathbf{u}\in\mathcal{X}(\delta)$ satisfy $\delta\leq\mathbf{x}^{\mathrm{T}}\mathbf{u}\leq 1-\delta$, where $\delta>0$ is some small constant independent of n. Then

$$
\sup_{\mathbf{x}\in\mathcal{X}(\delta)}\|\mathbf{G}_n(\mathbf{x})^{-1}-\mathbf{G}(\mathbf{x})^{-1}\|_{\mathrm{F}}\to 0,\quad \textit{and}\quad \sup_{\mathbf{x}\in\mathcal{X}(\delta)}\|\widetilde{\mathbf{G}}_n(\mathbf{x})-\widetilde{\mathbf{G}}(\mathbf{x})\|_{\mathrm{F}}\to 0
$$

as $n\to\infty$.

Proof. We first show that $\mathbf{G}_n(\mathbf{x})\to\mathbf{G}(\mathbf{x})$ as $n\to\infty$ uniformly for all $\mathbf{x}\in\mathcal{X}(\delta)$. It suffices to show that for all $s,t\in[d]$,

$$
\sup_{\mathbf{x}\in\mathcal{X}(\delta)}|\mathbf{e}_s^{\mathrm{T}}\mathbf{G}_n(\mathbf{x})\mathbf{e}_t-\mathbf{e}_s^{\mathrm{T}}\mathbf{G}(\mathbf{x})\mathbf{e}_s|\to 0
$$

as $n\to\infty$. Observe that $\mathbf{e}_s^{\mathrm{T}}\mathbf{G}_n(\mathbf{x})\mathbf{e}_t-\mathbf{e}_s^{\mathrm{T}}\mathbf{G}(\mathbf{x})\mathbf{e}_s\to 0$ as $n\to\infty$ for each $\mathbf{x}\in\mathcal{X}(\delta)$. Furthermore,

$$
\begin{aligned}
\sup_{n\geq 1,s,t,\in[d]}\left\|\frac{\partial}{\partial\mathbf{x}}\mathbf{e}_s^{\mathrm{T}}\mathbf{G}_n(\mathbf{x})\mathbf{e}_t\right\|_2 &\leq \sup_{n\geq 1,s,t,\in[d]}\frac{1}{n}\sum_{j=1}^n\left\|\frac{[\mathbf{x}_{0j}]_s[\mathbf{x}_{0j}]_t(1-2\rho_n\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}}{\{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})\}^2}\right\|_2\\
&\leq \sup_{n\geq 1,s,t,\in[d]}\frac{1}{n}\sum_{j=1}^n\frac{3}{\delta^4}=\frac{3}{\delta^4}<\infty.
\end{aligned}
$$

This means that the function class $(\mathbf{e}_s^{\mathrm{T}}\mathbf{G}_n(\mathbf{x})\mathbf{e}_t)_{n=1}^{\infty}$ is a uniformly Lipschitz function class and hence is equicontinuous. Since $\mathcal{X}(\delta)$ is compact, it follows from the Arzela-Ascoli theorem that the convergence $\mathbf{e}_s^{\mathrm{T}}\mathbf{G}_n(\mathbf{x})\mathbf{e}_t - \mathbf{e}_s^{\mathrm{T}}\mathbf{G}(\mathbf{x})\mathbf{e}_s \to 0$ is also uniform.

We next show that $\sup_{\mathbf{x}\in\mathcal{X}(\delta)}\|\mathbf{G}_n(\mathbf{x})^{-1} - \mathbf{G}(\mathbf{x})^{-1}\|_{\mathrm{F}} \to 0$. This immediately follows from the inequality

$$\sup_{\mathbf{x}\in\mathcal{X}(\delta)}\|\mathbf{G}_n(\mathbf{x}) - \mathbf{G}(\mathbf{x})\|_{\mathrm{F}} \leq \sup_{\mathbf{x}\in\mathcal{X}(\delta)}\|\mathbf{G}_n(\mathbf{x})^{-1}\|_{\mathrm{F}}\|\mathbf{G}_n(\mathbf{x}) - \mathbf{G}(\mathbf{x})\|_{\mathrm{F}}\|\mathbf{G}(\mathbf{x})^{-1}\|_{\mathrm{F}},$$

the fact that $\mathbf{G}_n(\mathbf{x}) \succeq (1/2)\boldsymbol{\Delta}$ and $\mathbf{G}(\mathbf{x}) \succeq \boldsymbol{\Delta}$ for sufficiently large n, and the uniform convergence of $\mathbf{G}_n(\mathbf{x}) \to \mathbf{G}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}(\delta)$.

We finally show that $\widetilde{\mathbf{G}}_n(\mathbf{x}) - \widetilde{\mathbf{G}}(\mathbf{x})$ uniformly for all $\mathbf{x} \in \mathcal{X}(\delta)$. This result also follows from the fact that

$$\sup_{\mathbf{x}\in\mathcal{X}(\delta)}\left|\frac{1}{\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}} - \frac{1}{\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}}\right| \leq \frac{1}{\delta^2}\|\boldsymbol{\mu}_n - \boldsymbol{\mu}\| \to 0, \quad \sup_{\mathbf{x}\in\mathcal{X}(\delta)}\left\|\frac{\mathbf{x}\boldsymbol{\mu}_n^{\mathrm{T}}}{(\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x})^2} - \frac{\mathbf{x}\boldsymbol{\mu}^{\mathrm{T}}}{(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x})^2}\right\|_{\mathrm{F}} \leq \frac{3}{\delta^4}\|\boldsymbol{\mu}_n - \boldsymbol{\mu}\| \to 0,$$

and the uniform convergence result $\sup_{\mathbf{x}\in\mathcal{X}(\delta)}\|\mathbf{G}_n(\mathbf{x})^{-1} - \mathbf{G}(\mathbf{x})^{-1}\|_{\mathrm{F}} \to 0$. The proof is thus completed. □

D.2 Proof of Theorem 4

Proof of Theorem 4. Let $\mathbf{W}$ be the matrix satisfying (3.3). For any $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^{\mathrm{T}} \in \mathbb{R}^{n\times d}$, denote $\mathbf{H}_i(\mathbf{X}) = (1/n)\sum_{j=1}^n \mathbf{x}_j\{(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)(1 - \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\}^{-1}\mathbf{x}_j^{\mathrm{T}}$. By definition,

$$\begin{aligned}
\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_{0i} &= (\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_{0i}) \\
&\quad + \mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})^{-1}\mathbf{W}\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}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\right\} \\
&\quad + \frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\frac{(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}(\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})^{-1}\mathbf{W})\mathbf{x}_{0j} \\
&= (\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_{0i}) \\
&\quad + \mathbf{G}_n(\mathbf{x}_{0i})^{-1}\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}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\right\} \\
&\quad + \{\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})^{-1}\mathbf{W} - \mathbf{G}_n(\mathbf{x}_{0i})^{-1}\} \\
&\quad \times \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}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\right\} \\
&\quad + \frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\frac{(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\{\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})^{-1}\mathbf{W} - \mathbf{G}_n(\mathbf{x}_{0i})^{-1}\}\mathbf{x}_{0j} \\
&\quad + \frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\frac{(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{x}_{0j} \\
&= (\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_{0i}) + \mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{R}_{i1} + \mathbf{R}_{i2}\mathbf{R}_{i1} + \mathbf{R}_{i3} \\
&\quad + \frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\frac{(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{x}_{0j},
\end{aligned}$$

where

$$\mathbf{R}_{i1} = \frac{1}{n\sqrt{\rho_n}} \sum_{j=1}^n \left\{ \frac{(A_{ij} - \widetilde{\mathbf{x}}_i^{\mathrm{T}} \widetilde{\mathbf{x}}_j)(\rho_n^{-1/2} \mathbf{W}^{\mathrm{T}} \widetilde{\mathbf{x}}_j)}{\rho_n^{-1} \widetilde{\mathbf{x}}_i^{\mathrm{T}} \widetilde{\mathbf{x}}_j (1 - \widetilde{\mathbf{x}}_i^{\mathrm{T}} \widetilde{\mathbf{x}}_j)} - \frac{(A_{ij} - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j}) \mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})} \right\},$$

$$\mathbf{R}_{i2} = \mathbf{W}^{\mathrm{T}} \mathbf{H}_i (\widetilde{\mathbf{X}})^{-1} \mathbf{W} - \mathbf{G}_n (\mathbf{x}_{0i})^{-1},$$

$$\mathbf{R}_{i3} = \frac{1}{n\sqrt{\rho_n}} \sum_{j=1}^n \frac{(A_{ij} - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})}{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})} \mathbf{R}_{i2} \mathbf{x}_{0j}.$$

We first analyze $\mathbf{R}_{i1}$. Denote the function $\boldsymbol{\phi}_{ij} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ by

$$\boldsymbol{\phi}_{ij}(\mathbf{u}, \mathbf{v}) = \frac{(A_{ij} - \rho_n \mathbf{u}^{\mathrm{T}} \mathbf{v}) \mathbf{v}}{\mathbf{u}^{\mathrm{T}} \mathbf{v} (1 - \rho_n \mathbf{u}^{\mathrm{T}} \mathbf{v})}, \quad i, j \in [n],$$

and let $\boldsymbol{\phi}_{ij} = [\phi_{ij1}, \ldots, \phi_{ijd}]^{\mathrm{T}}$. By Taylor's expansion, we have, if $\|\mathbf{u} - \mathbf{x}_{0i}\| < \epsilon$ and $\|\mathbf{v} - \mathbf{x}_{0j}\| < \epsilon$ for sufficiently small $\epsilon > 0$, and $\delta \leq \min_{i,j \in [n]} \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} \leq \max_{i,j \in [n]} \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} \leq 1 - \delta$ for some constant $\delta > 0$, then

$$\begin{aligned}
&\phi_{ijk}(\mathbf{u}, \mathbf{v}) - \phi_{ijk}(\mathbf{x}_{0i}, \mathbf{x}_{0j}) \\
&\quad = -\left\{ \frac{\rho_n x_{0jk} \mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})} \right\}^{\mathrm{T}} (\mathbf{u} - \mathbf{x}_{0i}) - \left\{ \frac{\rho_n \mathbf{x}_{0i} x_{0jk}}{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})} \right\}^{\mathrm{T}} (\mathbf{v} - \mathbf{x}_{0j}) \\
&\qquad - \left[\frac{x_{0jk} (A_{ij} - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})(1 - 2\rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j}) \mathbf{x}_{0j}}{\{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})\}^2} \right]^{\mathrm{T}} (\mathbf{u} - \mathbf{x}_{0i}) \\
&\qquad + \left[\frac{(A_{ij} - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j}) \{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j}) \mathbf{e}_k - x_{0jk} (1 - 2\rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j}) \mathbf{x}_{0i}\}}{\{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})\}^2} \right]^{\mathrm{T}} (\mathbf{v} - \mathbf{x}_{0j}) + R_{ijk},
\end{aligned}$$

where $\max_{i,j \in [n], k \in [d]} |R_{ijk}| \leq C_\delta \max(\|\mathbf{u} - \mathbf{u}_0\|^2, \|\mathbf{v} - \mathbf{v}_0\|^2)$ for some constant C_δ only depending on δ. Applying the above fact to $\mathbf{R}_{i1}$, we derive

$$\begin{aligned}
\mathbf{R}_{i1} &= -\frac{1}{n\sqrt{\rho_n}} \sum_{j=1}^n \frac{\rho_n \mathbf{x}_{0j} \mathbf{x}_{0j}^{\mathrm{T}}}{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})} (\rho_n^{-1/2} \mathbf{W}^{\mathrm{T}} \widetilde{\mathbf{x}}_i - \mathbf{x}_{0i}) \\
&\quad - \frac{1}{n\sqrt{\rho_n}} \sum_{j=1}^n \frac{\rho_n \mathbf{x}_{0j} \mathbf{x}_{0i}^{\mathrm{T}}}{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})} (\rho_n^{-1/2} \mathbf{W}^{\mathrm{T}} \widetilde{\mathbf{x}}_j - \mathbf{x}_{0j}) \\
&\quad - \frac{1}{n\sqrt{\rho_n}} \sum_{j=1}^n \left[\frac{(A_{ij} - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})(1 - 2\rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j}) \mathbf{x}_{0j} \mathbf{x}_{0j}^{\mathrm{T}}}{\{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})\}^2} \right] (\rho_n^{-1/2} \mathbf{W}^{\mathrm{T}} \widetilde{\mathbf{x}}_i - \mathbf{x}_{0i}) \\
&\quad + \frac{1}{n\sqrt{\rho_n}} \sum_{j=1}^n \frac{(A_{ij} - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{ij}) \{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j}) \mathbf{I}_d - (1 - 2\rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j}) \mathbf{x}_{0j} \mathbf{x}_{0i}^{\mathrm{T}}\}}{\{\mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j} (1 - \rho_n \mathbf{x}_{0i}^{\mathrm{T}} \mathbf{x}_{0j})\}^2} (\rho_n^{-1/2} \mathbf{W}^{\mathrm{T}} \widetilde{\mathbf{x}}_j - \mathbf{x}_{0j}) \\
&\quad + \frac{1}{n\sqrt{\rho_n}} \sum_{j=1}^n \mathbf{R}_{ij} \\
&= -\mathbf{R}_{i11} - \mathbf{R}_{i12} - \mathbf{R}_{i13} + \mathbf{R}_{i14} + \frac{1}{n\sqrt{\rho_n}} \sum_{j=1}^n \mathbf{R}_{ij},
\end{aligned}$$

where $\mathbf{R}_{ij}$'s are such that $\max_{i,j\in[n]}\|\mathbf{R}_{ij}\| \lesssim \|\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W} - \mathbf{X}_0\|_{2\to\infty}^2$ when $\|\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W} - \mathbf{X}_0\|_{2\to\infty}$ is sufficiently small. Clearly,

$$\mathbf{R}_{i11} = \frac{1}{n}\sum_{j=1}^n \frac{\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_{0i}) = \mathbf{G}_n(\mathbf{x}_{0i})(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_{0i}).$$

Furthermore, Lemma D.2 shows that $\max_{i\in[n]}\|\mathbf{R}_{i12}\| = O_{\mathbb{P}_0}((n\sqrt{\rho_n})^{-1}(\log n)^{(1\vee\omega)/2})$. In addition, we have $\max_{i\in[n]}\|\mathbf{R}_{i13}\| = O_{\mathbb{P}_0}((n\rho_n)^{-1}(\log n)^{1/2+(1\vee\omega)/2})$ by Lemma D.3, $\|\mathbf{R}_{i14}\| = O_{\mathbb{P}_0}((n\rho_n^{3/2})^{-1}(\log n)^{\omega/2})$ by Lemma D.4, and

$$\max_{i\in[n]}\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\max_{j\in[n]}\|\mathbf{R}_{ij}\| = O_{\mathbb{P}_0}\left(\frac{(\log n)^{1\vee\omega}}{n\rho_n^{5/2}}\right)$$

by Lemma D.1. This shows that

$$\begin{aligned}\|\mathbf{R}_{i1} + \mathbf{R}_{i11}\| &= \left\|\mathbf{R}_{i1} + \mathbf{G}_n(\mathbf{x}_{0i})(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_{0i})\right\|\\ &\le \|\mathbf{R}_{i12}\| + \|\mathbf{R}_{i13}\| + \|\mathbf{R}_{i14}\| + \frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\max_{i,j\in[n]}\|\mathbf{R}_{ij}\| = O_{\mathbb{P}_0}\left(\frac{(\log n)^{1\vee\omega}}{n\rho_n^{5/2}}\right),\end{aligned}$$

and hence,

$$\|\mathbf{R}_{i1}\| \le \|\mathbf{G}_n(\mathbf{x}_{0i})\|\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}_0\|_{2\to\infty} + O_{\mathbb{P}_0}\left(\frac{(\log n)^{1\vee\omega}}{n\rho_n^{5/2}}\right) = O_{\mathbb{P}_0}\left(\sqrt{\frac{(\log n)^{1\vee\omega}}{n\rho_n}}\right)$$

by Lemma D.1.

Next we focus on $\mathbf{R}_{i2}$. Since the function $(\mathbf{u},\mathbf{v})\mapsto\{\mathbf{u}^{\mathrm{T}}\mathbf{v}(1-\rho_n\mathbf{u}^{\mathrm{T}}\mathbf{v})\}^{-1}\mathbf{v}\mathbf{v}^{\mathrm{T}}$ is Lipschitz continuous in a neighborhood of $(\mathbf{x}_{0i},\mathbf{x}_{0j})$, it follows immediately that

$$\max_{i\in[n]}\|\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})\mathbf{W} - \mathbf{G}_n(\mathbf{x}_{0i})\|_{\mathrm{F}} \lesssim \|\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W} - \mathbf{X}_0\|_{2\to\infty}$$

when $\|\rho_n^{-1/2}\widetilde{\mathbf{X}}\mathbf{W} - \mathbf{X}_0\|_{2\to\infty} \le C_c\rho_n^{-1}\sqrt{n^{-1}}(\log n)^{(1\vee\omega)/2}$. Namely,

$$\max_{i\in[n]}\|\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})\mathbf{W} - \mathbf{G}_n(\mathbf{x}_{0i})\|_{\mathrm{F}} = O_{\mathbb{P}_0}\left(\frac{(\log n)^{(1\vee\omega)/2}}{\rho_n\sqrt{n}}\right)$$

by Lemma D.1. Furthermore, since $\mathbf{G}_n(\mathbf{x}_{0i}) \to \mathbf{G}(\mathbf{x}_{0i})$ as $n\to\infty$, $\mathbf{G}_n(\mathbf{x}_{0i}) - \boldsymbol{\Delta}$ is positive definite for sufficiently large n, and

$$|\lambda_d(\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})\mathbf{W}) - \lambda_d(\mathbf{G}_n(\mathbf{x}_{0i}))| \le \|\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})\mathbf{W} - \mathbf{G}_n(\mathbf{x}_{0i})\|_{\mathrm{F}}^2,$$

(see, for example, Hoffman and Wielandt, 2003), we conclude that

$$\min_{i\in[n]}\lambda_d(\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})\mathbf{W}) \ge \min_{i\in[n]}\lambda_d(\mathbf{G}_n(\mathbf{x}_{0i})) - \max_{i\in[n]}\|\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})\mathbf{W} - \mathbf{G}_n(\mathbf{x}_{0i})\|_{\mathrm{F}}^2 \ge \lambda_d(\boldsymbol{\Delta}) - o_{\mathbb{P}_0}(1),$$

namely, $\max_{i\in[n]}\lambda_d^{-1}(\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})\mathbf{W}) = O_{\mathbb{P}_0}(1)$. Therefore,

$$
\begin{aligned}
\max_{i\in[n]}\|\mathbf{R}_{i2}\|_{\mathrm{F}} &= \max_{i\in[n]}\|\{\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})\mathbf{W}\}^{-1}\{\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})\mathbf{W} - \mathbf{G}_n(\mathbf{x}_{0i})\}\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\|_{\mathrm{F}}\\
&\lesssim \max_{i\in[n]}\lambda_d^{-1}(\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})\mathbf{W})\max_{i\in[n]}\|\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\|_{\mathrm{F}}\max_{i\in[n]}\|\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})\mathbf{W} - \mathbf{G}_n(\mathbf{x}_{0i})\|_{\mathrm{F}}\\
&\leq O_{\mathbb{P}_0}(1)\|\mathbf{\Delta}^{-1}\|_{\mathrm{F}}\max_{i\in[n]}\|\mathbf{W}^{\mathrm{T}}\mathbf{H}_i(\widetilde{\mathbf{X}})\mathbf{W} - \mathbf{G}_n(\mathbf{x}_{0i})\|_{\mathrm{F}} = O_{\mathbb{P}_0}\left(\frac{(\log n)^{(1\vee\omega)/2}}{\rho_n\sqrt{n}}\right)
\end{aligned}
$$

We finally move forward to analyze $\mathbf{R}_{i3}$. Since

$$
\max_{i\in[n]}\|\mathbf{R}_{i3}\|_{\mathrm{F}} \leq \max_{i\in[n]}\|\mathbf{R}_{i2}\|_{\mathrm{F}}\sum_{k=1}^d\frac{1}{n\sqrt{\rho_n}}\max_{i\in[n]}\left|\sum_{j=1}^n\frac{(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})x_{0jk}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\right|,
$$

and by Hoeffding's inequality and the union bound,

$$
\begin{aligned}
\mathbb{P}_0\left(\max_{i\in[d]}\left|\sum_{j=1}^n\frac{(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})x_{0jk}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\right| > t\sqrt{n\log n}\right) &= 2n\exp\left[-\frac{2t^2n\log n}{\sum_{j=1}^n x_{0jk}^2\{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\}^{-2}}\right]\\
&\leq 2\exp\left\{-(Mt^2 - 1)\log n\right\}
\end{aligned}
$$

for some constant $M > 0$. Hence,

$$
\sum_{k=1}^d\frac{1}{n\sqrt{\rho_n}}\max_{i\in[n]}\left|\sum_{j=1}^n\frac{(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})x_{0jk}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\right| = O_{\mathbb{P}_0}\left(\sqrt{\frac{\log n}{n\rho_n}}\right),
$$

and hence, $\max_{i\in[n]}\|\mathbf{R}_{i3}\|_{\mathrm{F}} = O_{\mathbb{P}_0}(\rho_n^{-3/2}n^{-1}(\log n)^{1/2 + (1\vee\omega)/2})$. Therefore, we conclude that

$$
\begin{aligned}
\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_{0i} &= (\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_{0i}) + \mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{R}_1 + \mathbf{R}_2\mathbf{R}_1 + \mathbf{R}_3\\
&\quad + \frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\frac{(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{x}_{0j}\\
&= \frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\frac{(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{x}_{0j} + \widehat{\mathbf{R}}_i,
\end{aligned}
$$

where

$$
\begin{aligned}
\|\widehat{\mathbf{R}}_i\| &= \left\|\mathbf{G}_n(\mathbf{x}_{0i})^{-1}(\mathbf{R}_{i1} + \mathbf{R}_{i11}) + \mathbf{R}_{i2}\mathbf{R}_{i1} + \mathbf{R}_{i3}\right\| \leq \|\mathbf{\Delta}_n^{-1}\|_2\|\|\mathbf{R}_{i1} + \mathbf{R}_{i11}\| + (\|\mathbf{R}_{i2}\|_2\|\mathbf{R}_{i1}\| + \|\mathbf{R}_{i3}\|)\\
&= O_{\mathbb{P}_0}\left(\frac{(\log n)^{1\vee\omega}}{n\rho_n^{5/2}}\right) + O_{\mathbb{P}_0}\left(\frac{(\log n)^{1\vee\omega}}{n\rho_n^{3/2}}\right) + O_{\mathbb{P}_0}\left(\frac{(\log n)^{1/2 + (1\vee\omega)/2}}{n\rho_n^{3/2}}\right) = O_{\mathbb{P}_0}\left(\frac{(\log n)^{1\vee\omega}}{n\rho_n^{5/2}}\right).
\end{aligned}
$$

We now proceed to prove that $\sum_{i=1}^n\|\widehat{\mathbf{R}}_i\|^2 = O_{\mathbb{P}_0}((n\rho_n^5)^{-1}(\log n)^{2(1\vee\omega)})$. Observe that by Lemma D.4 we have

$$\begin{aligned}\sum_{i=1}^n \left\|\mathbf{G}_n(\mathbf{x}_{0i})^{-1}(-\mathbf{R}_{i12}-\mathbf{R}_{i13}+\mathbf{R}_{i14})\right\|^2 &\leq 3n\left\|\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\right\|_{\mathrm{F}}^2\left\{\max_{i\in[n]}\|\mathbf{R}_{i12}\|^2+\max_{i\in[n]}\|\mathbf{R}_{i13}\|^2\right\}+3\sum_{i=1}^n\|\mathbf{R}_{i14}\|^2\\
&\leq n\left\|\mathbf{\Delta}^{-1}\right\|_{\mathrm{F}}^2\left\{\max_{i\in[n]}\|\mathbf{R}_{i12}\|^2+\max_{i\in[n]}\|\mathbf{R}_{i13}\|^2\right\}+\sum_{i=1}^n\|\mathbf{R}_{i14}\|^2\\
&= O_{\mathbb{P}_0}\left(\frac{(\log n)^{2(1\vee\omega)}}{n\rho_n^3}\right),\end{aligned}$$

and that

$$\sum_{i=1}^n\left\|\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\mathbf{R}_{ij}\right\|^2\leq n\|\mathbf{\Delta}^{-1}\|_{\mathrm{F}}^2\left(\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\max_{i\in[n]}\|\mathbf{R}_{ij}\|\right)^2=O_{\mathbb{P}_0}\left(\frac{(\log n)^{2(1\vee\omega)}}{n\rho_n^5}\right).$$

Besides, by the above derivation we have

$$\begin{aligned}\sum_{i=1}^n\|\mathbf{R}_{i2}\mathbf{R}_{i1}\|^2&\leq\max_{i\in[n]}\|\mathbf{R}_{i2}\|_{\mathrm{F}}^2\sum_{i=1}^n\|\mathbf{R}_{i1}\|^2\\
&\leq O_{\mathbb{P}_0}\left(\frac{(\log n)^{1\vee\omega}}{n^2\rho_n}\right)\left\{n\max_{i\in[n]}\|\mathbf{G}_n(\mathbf{x}_{0i})\|_{\mathrm{F}}^2\|\widetilde{\mathbf{X}}\mathbf{W}-\rho_n^{1/2}\mathbf{X}_0\|_{\mathrm{F}}^2\right\}\\
&\quad+O_{\mathbb{P}_0}\left(\frac{(\log n)^{1\vee\omega}}{n^2\rho_n}\right)\left\{\sum_{i=1}^n\left\|\mathbf{R}_{i12}+\mathbf{R}_{i13}-\mathbf{R}_{i14}-\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\mathbf{R}_{ij}\right\|^2\right\}\\
&=O_{\mathbb{P}_0}\left(\frac{(\log n)^{2(1\vee\omega)}}{n^2\rho_n^2}\right),\end{aligned}$$

and

$$\sum_{i=1}^n\|\mathbf{R}_{i3}\|_{\mathrm{F}}^2\leq n\max_{i\in[n]}\|\mathbf{R}_{i3}\|_{\mathrm{F}}^2=O_{\mathbb{P}_0}\left(\frac{(\log n)^{2(1\vee\omega)}}{n\rho_n^3}\right).$$

Therefore, we conclude that

$$\begin{aligned}\sum_{i=1}^n\|\widehat{\mathbf{R}}_i\|_{\mathrm{F}}^2\leq 4\sum_{i=1}^n\left\|\mathbf{G}_n(\mathbf{x}_{0i})^{-1}(-\mathbf{R}_{i12}-\mathbf{R}_{i13}+\mathbf{R}_{i14})\right\|^2+4\sum_{i=1}^n\left\|\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n\mathbf{R}_{ij}\right\|^2\\
+4\sum_{i=1}^n\|\mathbf{R}_{i2}\mathbf{R}_{i1}\|^2+4\sum_{i=1}^n\|\mathbf{R}_{i3}\|_{\mathrm{F}}^2=O_{\mathbb{P}_0}\left(\frac{(\log n)^{2(1\vee\omega)}}{n\rho_n^5}\right).\end{aligned}$$

The proof is thus completed. □

D.3 Proof of Theorem 5

Proof of Theorem 5. Let $(\mathbf{W})_{n=1}^{\infty} = (\mathbf{W}_n)_{n=1}^{\infty} \subset \mathbb{O}(d)$ be a sequence of orthogonal matrices satisfying (3.3). Denote

$$\boldsymbol{\gamma}_{ij} = \frac{1}{n\sqrt{\rho_n}} \frac{\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}.$$

First note that $\mathbf{G}_n(\mathbf{x}_{0i})^{-1} \succeq \boldsymbol{\Delta}$ for sufficiently large n, and hence,

$$\sup_{i,j\in[n]} \|\boldsymbol{\gamma}_{ij}\| \leq \frac{1}{n\sqrt{\rho_n}} \sup_{i,j\in[n]} \frac{\|\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\|\|\mathbf{x}_{0j}\|}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})} \leq \frac{1}{n\sqrt{\rho_n}} \frac{\|\boldsymbol{\Delta}\|_2}{\delta(1-\delta)} \lesssim \frac{1}{n\sqrt{\rho_n}}. \tag{D.1}$$

Also observe that

$$\begin{aligned}
\mathbb{E}_0\left(\sum_{i=1}^n \left\|\sum_{j=1}^n (A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij}\right\|^2\right) &= \sum_{i=1}^n\sum_{a=1}^n\sum_{b=1}^n \mathbb{E}_0\left\{(A_{ia} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0a})(A_{ib} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0b})\boldsymbol{\gamma}_{ia}^{\mathrm{T}}\boldsymbol{\gamma}_{ib}\right\}\\
&= \frac{1}{n^2\rho_n}\sum_{i=1}^n\sum_{a=1}^n \frac{\mathbb{E}_0\{(A_{ia} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0a})^2\}}{\{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0a}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0a})\}^2}\mathbf{x}_{0a}^{\mathrm{T}}\mathbf{G}_n^{-2}(\mathbf{x}_{0i})\mathbf{x}_{0a}\\
&= \frac{1}{n}\sum_{i=1}^n\frac{1}{n}\sum_{a=1}^n \frac{\mathrm{tr}\{\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{x}_{0a}\mathbf{x}_{0a}^{\mathrm{T}}\mathbf{G}_n^{-1}(\mathbf{x}_{0i})\}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0a}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0a})}\\
&= \frac{1}{n}\sum_{i=1}^n \mathrm{tr}\{\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\}.
\end{aligned}$$

By Theorem 4 and Lemma D.5, we can write

$$\begin{aligned}
\left\|\widehat{\mathbf{X}}\mathbf{W} - \mathbf{X}_0\right\|_{\mathrm{F}}^2 &= \sum_{i=1}^n\left\|\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij}\right\|^2 + 2\sum_{i=1}^n\widehat{\mathbf{R}}_i^{\mathrm{T}}\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij} + \sum_{i=1}^n\|\widehat{\mathbf{R}}_i\|^2\\
&= \frac{1}{n}\sum_{i=1}^n\mathrm{tr}\{\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\} + 2\sum_{i=1}^n\widehat{\mathbf{R}}_i^{\mathrm{T}}\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij} + o_{\mathbb{P}_0}(1) + O_{\mathbb{P}_0}\left(\frac{(\log n)^{2(1\vee\omega)}}{n\rho_n^5}\right)\\
&= \frac{1}{n}\sum_{i=1}^n\mathrm{tr}\{\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\} + 2\sum_{i=1}^n\widehat{\mathbf{R}}_i^{\mathrm{T}}\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij} + o_{\mathbb{P}_0}(1).
\end{aligned}$$

By Cauchy-Schwarz inequality,

$$\begin{aligned}
\left|\sum_{i=1}^n\widehat{\mathbf{R}}_i^{\mathrm{T}}\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij}\right| &\leq \sum_{i=1}^n\|\widehat{\mathbf{R}}_i\|\left\|\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\gamma_{ijk}\right\|\\
&\leq \left(\sum_{i=1}^n\|\widehat{\mathbf{R}}_i\|^2\right)^{1/2}\left\{\sum_{i=1}^n\left\|\sum_{j=1}^n(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij}\right\|^2\right\}^{1/2}\\
&= O_{\mathbb{P}_0}\left(\frac{(\log n)^{1\vee\omega}}{\sqrt{n\rho_n^5}}\right)\left\{\frac{1}{n}\sum_{i=1}^n\mathrm{tr}\{\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\} + o_{\mathbb{P}_0}(1)\right\}^{1/2} = o_{\mathbb{P}_0}(1).
\end{aligned}$$

Hence, by condition (2.1) and the uniform convergence of $\mathbf{G}_n(\mathbf{x})^{-1} \to \mathbf{G}(\mathbf{x})^{-1}$ for all $\mathbf{x}$ (Lemma D.7), we obtain (see, for example, Exercise 3 in Section 4.4 of Chung, 2001)

$$\frac{1}{n}\sum_{i=1}^n \mathrm{tr}\{\mathbf{G}_n(\mathbf{x}_{0i})\} = \int \mathrm{tr}\{\mathbf{G}_n(\mathbf{x})^{-1}\}F_n(\mathrm{d}\mathbf{x}) \to \int_{\mathcal{X}} \mathrm{tr}\{\mathbf{G}(\mathbf{x})^{-1}\}F(\mathrm{d}\mathbf{x}).$$

This completes the first part of the theorem. For the second part, we observe that

$$\begin{aligned}
&\sum_{j=1}^n \mathbb{E}_0\left\{\left\|\frac{1}{\sqrt{n\rho_n}}\frac{(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{x}_{0j}\right\|^3\right\}\\
&\quad\le \frac{1}{(n\rho_n)^{3/2}}\sum_{j=1}^n \frac{\mathbb{E}_0\{|A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}|^3\}}{\{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\}^3}\|\mathbf{G}_n^{-1}(\mathbf{x}_{0i})\|_2^3\|\mathbf{x}_{0j}\|^3 \lesssim \frac{1}{\sqrt{n\rho_n}} \lesssim \frac{(\log n)^{1\vee\omega}}{\sqrt{n\rho_n^5}} \to 0,
\end{aligned}$$

and

$$\begin{aligned}
\sum_{j=1}^n \mathrm{var}_0\left\{\frac{1}{\sqrt{n\rho_n}}\frac{(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\right\} &= \sum_{j\neq i}^n \mathrm{var}_0\left\{\frac{1}{\sqrt{n\rho_n}}\frac{(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{x}_{0j}\right\}\\
&= \frac{1}{n\rho_n}\sum_{j\neq i}^n \frac{\rho_n\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}\mathbf{G}_n(\mathbf{x}_{0i})^{-1}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\\
&= \mathbf{G}_n(\mathbf{x}_{0i})^{-1} \to \mathbf{G}(\mathbf{x}_{0i})^{-1}.
\end{aligned}$$

It follows from the Lyapunov's central limit theorem and Theorem 4 that

$$\sqrt{n}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i - \mathbf{x}_{0i}) = \sum_{j=1}^n \frac{1}{\sqrt{n\rho_n}}\frac{(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{x}_{0j} + o_{\mathbb{P}_0}(1) \xrightarrow{\mathcal{L}} \mathrm{N}(\mathbf{0}, \mathbf{G}(\mathbf{x}_{0i})^{-1}).$$

The proof is thus completed. □

E Proof of Theorems 6 and 7 (Limit Theorems for the LSE)

In this section of the proofs, we introduce following notations: Let $\mathbf{1} = [1,\ldots,1]^{\mathrm{T}} \in \mathbb{R}^n$ be the n-dimensional vector with coordinates being ones, $\mathbf{P}_0 = \mathbb{E}_0(\mathbf{A}) = \rho_n\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}$, $\mathbf{T} = \mathrm{diag}(\rho_n\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}\mathbf{1})$, $\mathbf{D} = \mathrm{diag}(\mathbf{A1})$, and $\widetilde{\mathbf{E}} = n\rho_n\{\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}_0)\}$. We further denote $\mathcal{L}(\mathbf{A}) = \sum_{i=1}^n \widetilde{\lambda}_i(\widetilde{\mathbf{u}}_{\mathbf{A}})_i(\widetilde{\mathbf{u}}_{\mathbf{A}})_i^{\mathrm{T}}$ the spectral decomposition of $\mathcal{L}(\mathbf{A})$ with $|\widetilde{\lambda}_1| \ge \ldots \ge |\widetilde{\lambda}_n|$, $\mathcal{L}(\mathbf{P}_0) = \widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}\widetilde{\mathbf{U}}_{\mathbf{P}}^{\mathrm{T}}$ the (compact) spectral decomposition of $\mathcal{L}(\mathbf{P}_0)$, where $\widetilde{\mathbf{U}}_{\mathbf{P}} \in \mathbb{O}(n,d)$, and $\widetilde{\mathbf{S}}_{\mathbf{P}} = \mathrm{diag}[\lambda_1\{\mathcal{L}(\mathbf{P}_0)\},\ldots,\lambda_d\{\mathcal{L}(\mathbf{P}_0)\}]$. We observe that the LSE can be written as $\breve{\mathbf{X}} = \widetilde{\mathbf{U}}_{\mathbf{A}}\widetilde{\mathbf{S}}_{\mathbf{A}}^{1/2}$, where $\widetilde{\mathbf{U}}_{\mathbf{A}} = [(\widetilde{\mathbf{u}}_{\mathbf{A}})_1,\ldots,(\widetilde{\mathbf{u}}_{\mathbf{A}})_d]$, and $\widetilde{\mathbf{S}}_{\mathbf{A}} = \mathrm{diag}(|\widetilde{\lambda}_1|,\ldots,|\widetilde{\lambda}_d|)$.

E.1 Proof of the Limit (4.3)

Proof of the limit (4.3). The proof is almost the same as the proof in Tang and Priebe (2018) and we only sketch the argument and present the difference. Using the proof there, we obtain the following local expansion of $\breve{\mathbf{X}}\mathbf{W} - \mathbf{Y}_0$ for some orthogonal alignment matrix $\mathbf{W}$ that may depend on n:

$$\breve{\mathbf{X}}\mathbf{W} - \mathbf{Y}_0 = \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\mathbf{Y}_0(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1} + \frac{1}{2}\mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\mathbf{Y}_0 + \breve{\mathbf{R}}, \tag{E.1}$$

where $\|\breve{\mathbf{R}}\|_{\mathrm{F}} = O((n\rho_n)^{-1})$ with high probability. Using argument developed in Appendix B.4 in Tang and Priebe (2018) through the application of the concentration inequalities developed in Boucheron et al. (2003) (Theorem 5 and Theorem 6 there), we obtain that $n\rho_n\|\breve{\mathbf{X}}\mathbf{W} - \mathbf{Y}_0\|_{\mathrm{F}}^2$ is concentrated around

$$\begin{aligned}
&n\rho_n\mathbb{E}_0\left\{\left\|\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\mathbf{Y}_0(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1} + \frac{1}{2}\mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\mathbf{Y}_0\right\|_{\mathrm{F}}^2\right\}\\
&\quad= n\rho_n\mathbb{E}_0\left\{\|\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\mathbf{Y}_0(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1}\|_{\mathrm{F}}^2\right\} + \frac{n\rho_n}{4}\mathbb{E}_0\left\{\|\mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\mathbf{Y}_0\|_{\mathrm{F}}^2\right\}\\
&\quad\quad+ n\rho_n\mathrm{tr}\left[\mathbb{E}_0\left\{\mathbf{Y}_0^{\mathrm{T}}\mathbf{T}^{-1}(\mathbf{T} - \mathbf{D})\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\mathbf{Y}_0(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1}\right\}\right]
\end{aligned}$$

almost surely. The only difference is that the expected values in Tang and Priebe (2018) is taken with regard to both $\mathbf{A}$ and $\mathbf{P}$, whereas the expected values here is taken with respect to $\mathbf{A}$ only as $\mathbf{P}$ is deterministic in the current setup. Furthermore, using the argument developed in page 2407 of Tang and Priebe (2018) where the expected value is taken with regard to $\mathbf{A}$ conditioning on $\mathbf{P}_0$, we obtain

$$\mathbb{E}_0\left\{\|\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\mathbf{Y}_0(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1}\|_{\mathrm{F}}^2\right\} = \mathrm{tr}\left\{(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1}\mathbf{Y}_0^{\mathrm{T}}\widetilde{\mathbf{M}}\mathbf{Y}_0(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1}\right\},$$

where

$$\widetilde{\mathbf{M}} = \mathrm{diag}\left\{\frac{1}{n^2\rho_n}\sum_{j=1}^n\frac{\mathbf{x}_{01}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{01}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}_{01}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}, \ldots, \frac{1}{n^2\rho_n}\sum_{j=1}^n\frac{\mathbf{x}_{0n}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0n}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}_{0n}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}\right\}.$$

The key difference is that instead of using the strong law of large numbers applied to the random latent positions $\mathbf{x}_{01}, \ldots, \mathbf{x}_{0n}$, we make use of the assumption (2.1) to show the desired convergence results. By construction of $\mathbf{Y}_0$, it is easy to see that $\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0 \to \widetilde{\boldsymbol{\Delta}}$ as $n \to \infty$. Furthermore,

$$\begin{aligned}
\mathbf{Y}_0^{\mathrm{T}}\widetilde{\mathbf{M}}\mathbf{Y}_0 &= \frac{1}{n^2\rho_n}\sum_{i=1}^n\mathbf{y}_{0i}\mathbf{y}_{0i}^{\mathrm{T}}\sum_{j=1}^n\frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}\\
&= \frac{1}{n^3\rho_n}\sum_{i=1}^n\frac{\mathbf{x}_{0i}\mathbf{x}_{0i}^{\mathrm{T}}}{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n}\sum_{j=1}^n\frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}\\
&= \frac{1}{n^3\rho_n}\sum_{j=1}^n\sum_{i=1}^n\frac{\mathbf{x}_{0i}\mathbf{x}_{0i}^{\mathrm{T}}}{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n}\frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}\\
&= \frac{1}{n^2\rho_n}\sum_{j=1}^n\frac{1}{\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}\frac{1}{n}\sum_{i=1}^n\frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0i}\mathbf{x}_{0i}^{\mathrm{T}}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^2}\\
&= \frac{1}{n\rho_n}\left[\frac{1}{n}\sum_{j=1}^n\widetilde{\mathbf{V}}_n(\mathbf{x}_{0j})\right],
\end{aligned}$$

where

$$\widetilde{\mathbf{V}}_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n \frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x})\mathbf{x}_{0i}\mathbf{x}_{0i}^{\mathrm{T}}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^2} \to \widetilde{\mathbf{V}}(\mathbf{x}) = \int_{\mathcal{X}} \frac{\mathbf{x}_1^{\mathrm{T}}\mathbf{x}(1-\rho\mathbf{x}_1^{\mathrm{T}}\mathbf{x})\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}}{(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\mu})^2}F(\mathrm{d}\mathbf{x}_1)$$

uniformly over $\mathbf{x}\in\mathcal{X}$ (the argument for proving the uniform convergence is the same as that in the proof of Lemma D.7). Hence we conclude that

$$n\rho_n\mathrm{tr}\{(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1}\mathbf{Y}_0^{\mathrm{T}}\widetilde{\mathbf{M}}\mathbf{Y}_0(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1}\} \to \mathrm{tr}\left\{\widetilde{\boldsymbol{\Delta}}^{-1}\int_{\mathcal{X}}\mathbf{V}(\mathbf{x})F(\mathrm{d}\mathbf{x})\widetilde{\boldsymbol{\Delta}}^{-1}\right\}.$$

A similar argument can be applied to derive that

$$\frac{n\rho_n}{4}\mathbb{E}_0\left\{\|\mathbf{T}^{-1}(\mathbf{T}-\mathbf{D})\mathbf{Y}_0\|_{\mathrm{F}}^2\right\} \to \frac{1}{4}\mathrm{tr}\left\{\int_{\mathcal{X}}\frac{\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}}{(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\mu})^2}\left(1-\frac{\rho\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\Delta}\mathbf{x}_1}{\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\mu}}\right)F(\mathrm{d}\mathbf{x}_1)\right\}$$

and that

$$\begin{aligned}
&n\rho_n\mathrm{tr}\left[\mathbb{E}_0\left\{\mathbf{Y}_0^{\mathrm{T}}\mathbf{T}^{-1}(\mathbf{T}-\mathbf{D})\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}\mathbf{Y}_0(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1}\right\}\right]\\
&\quad\to \rho\mathrm{tr}\left[\iint_{\mathcal{X}^2}\frac{(\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2\mathbf{x}_2^{\mathrm{T}}\mathbf{x}_1)}{(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\mu})^2(\mathbf{x}_2^{\mathrm{T}}\boldsymbol{\mu})}\mathbf{x}_1\mathbf{x}_2^{\mathrm{T}}F(\mathrm{d}\mathbf{x}_1)F(\mathrm{d}\mathbf{x}_2)\widetilde{\boldsymbol{\Delta}}^{-1}\right] - \mathrm{tr}\left\{\int_{\mathcal{X}}\frac{\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}}{(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\mu})^2}F(\mathrm{d}\mathbf{x}_1)\right\}
\end{aligned}$$

The proof is completed by combining the above derivations.

□

E.2 Proof of the Asymptotic Normality (4.4) Under the Sparse regime (ii)

Proof of the asymptotic normality (4.4) under the sparse regime (ii). We first prove the asymptotic normality of the rows of the LSE under the condition (ii), and this is a simple modification of Appendix B.1 in Tang and Priebe (2018).

Let $\tau:[n]\to[K]$ be the cluster assignment function such that $\tau(i)=k$ when $\mathbf{x}_{0i}=\boldsymbol{\nu}_k$, and let $n_k=\sum_{i=1}^n\mathbb{1}(\mathbf{x}_{0i}=\boldsymbol{\nu}_k)$. Clearly, $n_k/n\to\pi_k$ for each $k\in[K]$. Since $F=\sum_{k=1}^K\pi_k\delta_{\boldsymbol{\nu}_k}$, then there exists a permutation matrix $\boldsymbol{\Pi}$ such that $[\boldsymbol{\nu}_1,\ldots,\boldsymbol{\nu}_1,\boldsymbol{\nu}_2,\ldots,\boldsymbol{\nu}_2,\ldots,\boldsymbol{\nu}_K,\ldots,\boldsymbol{\nu}_K]^{\mathrm{T}}=\boldsymbol{\Pi}^{\mathrm{T}}\mathbf{X}_0=\boldsymbol{\Pi}^{\mathrm{T}}[\mathbf{x}_{01},\ldots,\mathbf{x}_{0n}]^{\mathrm{T}}$.

Now suppose for a fixed $i\in[n]$, $\mathbf{x}_{0i}=\boldsymbol{\nu}_k$ for a $k\in[K]$ that depends on i. Let $\boldsymbol{\Pi}_k$ be the permutation matrix within the kth cluster. In the case where $1<k<K$, $\boldsymbol{\Pi}_k$ has the form

$$\boldsymbol{\Pi}_k=\boldsymbol{\Pi}\begin{bmatrix}\mathbf{I}_{n_1} & & & & & & \\ & \ddots & & & & & \\ & & \mathbf{I}_{n_{k-1}} & & & & \\ & & & \mathbf{Q}_k & & & \\ & & & & \mathbf{I}_{n_{k+1}} & & \\ & & & & & \ddots & \\ & & & & & & \mathbf{I}_{n_K}\end{bmatrix}\boldsymbol{\Pi}^{\mathrm{T}}$$

for some $n_k\times n_k$ permutation matrix $\mathbf{Q}_k$. A similar block diagonal $\boldsymbol{\Pi}_k$ can be explicitly written down when

$k = 1$ or $k = K$. Without loss of generality, we may assume that $k \neq 1$ and $k \neq K$. Furthermore, the adjacency matrix $\mathbf{A}$ can be written as the following block matrix form

$$\mathbf{\Pi}^{\mathrm{T}}\mathbf{A}\mathbf{\Pi} = \begin{bmatrix} \mathbf{A}_{\star 11} & \mathbf{A}_{\star 1k} & \mathbf{A}_{\star 12} \\ \mathbf{A}_{\star 1k}^{\mathrm{T}} & \mathbf{A}_{kk} & \mathbf{A}_{\star 2k}^{\mathrm{T}} \\ \mathbf{A}_{\star 21} & \mathbf{A}_{\star 2k} & \mathbf{A}_{\star 22} \end{bmatrix},$$

where $\mathbf{A}_{kk} = [A_{mj} : \tau(m) = \tau(j) = k]$, $\mathbf{A}_{\star 11} = [A_{mj} : \tau(m), \tau(j) < k]$, $\mathbf{A}_{\star 22} = [A_{mj} : \tau(m), \tau(j) > k]$, $\mathbf{A}_{\star 12} = [A_{mj} : \tau(m) < k, \tau(j) > k]$, and

$$\begin{bmatrix} \mathbf{A}_{\star 1k} \\ \mathbf{A}_{\star 2k} \end{bmatrix} = [\mathbf{A}_{mj} : \tau(m) \neq k, \tau(j) = k].$$

Clearly, the entries of $\mathbf{A}_{kk}$ are independent Bernoulli$(\rho_n\langle \boldsymbol{\nu}_k, \boldsymbol{\nu}_k\rangle)$ random variables, where $\langle\cdot,\cdot\rangle$ denotes the Euclidean inner product, and the entries of $\mathbf{A}_{\star 1k}, \mathbf{A}_{\star 2k}$ are independent Bernoulli$(\rho_n\langle \boldsymbol{\nu}_l, \boldsymbol{\nu}_k\rangle)$ random variables. Therefore,

$$\mathbf{A}_{\star kk} \overset{\mathcal{L}}{=} \mathbf{Q}_k\mathbf{A}_{kk}\mathbf{Q}_k^{\mathrm{T}}, \quad \mathbf{A}_{\star 1k} \overset{\mathcal{L}}{=} \mathbf{A}_{\star 1k}\mathbf{Q}_k^{\mathrm{T}}, \quad \mathbf{A}_{\star 2k} \overset{\mathcal{L}}{=} \mathbf{A}_{\star 2k}\mathbf{Q}_k^{\mathrm{T}}.$$

Hence,

$$\mathbf{\Pi}_k\mathbf{A}\mathbf{\Pi}_k^{\mathrm{T}} = \mathbf{\Pi}\begin{bmatrix} \mathbf{A}_{\star 11} & \mathbf{A}_{\star 1k}\mathbf{Q}_k^{\mathrm{T}} & \mathbf{A}_{\star 12} \\ \mathbf{Q}_k\mathbf{A}_{\star 1k}^{\mathrm{T}} & \mathbf{Q}_k\mathbf{A}_{kk}\mathbf{Q}_k^{\mathrm{T}} & \mathbf{Q}_k\mathbf{A}_{\star 2k}^{\mathrm{T}} \\ \mathbf{A}_{\star 21} & \mathbf{A}_{\star 2k}\mathbf{Q}_k^{\mathrm{T}} & \mathbf{A}_{\star 22} \end{bmatrix}\mathbf{\Pi}^{\mathrm{T}} \overset{\mathcal{L}}{=} \mathbf{\Pi}\begin{bmatrix} \mathbf{A}_{\star 11} & \mathbf{A}_{\star 1k} & \mathbf{A}_{\star 12} \\ \mathbf{A}_{\star 1k}^{\mathrm{T}} & \mathbf{A}_{kk} & \mathbf{A}_{\star 2k}^{\mathrm{T}} \\ \mathbf{A}_{\star 21} & \mathbf{A}_{\star 2k} & \mathbf{A}_{\star 22} \end{bmatrix}\mathbf{\Pi}^{\mathrm{T}} = \mathbf{A}.$$

Therefore, $\mathbf{\Pi}_k\breve{\mathbf{X}} \overset{\mathcal{L}}{=} \breve{\mathbf{X}}$, $\mathbf{\Pi}_k\mathbf{X}_0 = \mathbf{X}_0$, and $\mathbf{\Pi}_k\mathbf{Y}_0 = \mathbf{Y}_0$. Observe that the orthogonal alignment matrix $\mathbf{W}$ does not change with the permutation matrix $\mathbf{\Pi}_k$. Hence, this further implies the exchangeability of the rows within the kth cluster for the remainder matrix $\mathbf{R}$:

$$\mathbf{\Pi}_k\breve{\mathbf{R}} = \mathbf{\Pi}_k(\breve{\mathbf{X}}\mathbf{W} - \mathbf{Y}_0) = \mathbf{\Pi}_k\breve{\mathbf{X}}\mathbf{W} - \mathbf{Y}_0 \overset{\mathcal{L}}{=} \breve{\mathbf{X}}\mathbf{W} - \mathbf{Y}_0 = \breve{\mathbf{R}}.$$

Let $\breve{\mathbf{r}}_j$ be the jth row of $\breve{\mathbf{R}}$, $j \in [n]$. Clearly, by the exchangeability within clusters, for any $i, j \in [n]$ such that $\tau(j) = \tau(i) = k$, we see that $\breve{\mathbf{r}}_j = \breve{\mathbf{r}}_i$. Using the fact that $\mathbf{E}_0(\|\breve{\mathbf{R}}\|_{\mathrm{F}}^2) = O((n\rho_n)^{-2})$ (which can be easily derived using the fact that with probability at least $1 - n^{-3}$, $\|\breve{\mathbf{R}}\|_{\mathrm{F}} \lesssim (n\rho_n)^{-1}$; also see Appendix B.1 in Tang and Priebe, 2018), we see that for any $i \in [n]$ with $\tau(i) = k$,

$$n^2\rho_n\mathbb{E}_0(\|\breve{\mathbf{r}}_i\|_2^2) = n^2\rho_n\frac{1}{n_k}\sum_{j:\tau(j)=\tau(i)=k}\mathbb{E}_0(\|\breve{\mathbf{r}}_j\|_2^2) \leq \frac{n^2\rho_n}{n_k}\mathbb{E}(\|\breve{\mathbf{R}}\|_{\mathrm{F}}^2) \lesssim \frac{n\rho_n}{\pi_k + o(1)}(n\rho_n)^{-2} \to 0.$$

This shows that $\breve{\mathbf{r}}_i = o_{\mathbb{P}_0}(n\rho_n^{1/2})$.

The rest of the proof is devoted to prove the asymtotic normality of the ith row of

$$\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}\mathbf{Y}_0(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1}+\frac{1}{2}\mathbf{T}^{-1}(\mathbf{T}-\mathbf{D})\mathbf{Y}_0,$$

and is almost the same as that of Appendix B.1 in Tang and Priebe (2018). Let $\breve{\mathbf{x}}_i$ be the ith row of $\breve{\mathbf{X}}$. In essence, using the expansion (E.1), we have

$$n\rho_n^{1/2}(\mathbf{W}^{\mathrm{T}}\breve{\mathbf{x}}_i-\mathbf{y}_{0i})=\sum_{j\neq i}\left\{\widetilde{\mathbf{\Delta}}_n^{-1}\frac{\mathbf{x}_{0j}}{\sqrt{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}}-\frac{\mathbf{x}_{0i}}{2(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}}\right\}\frac{A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{\sqrt{n\rho_n}}+o_{\mathbb{P}_0}(1),$$

where $\widetilde{\mathbf{\Delta}}_n=\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0$. The sum of the third moment of the sum of independent random vectors in the first term of the right-hand side of the previous display is bounded above by

$$\frac{1}{(n\rho_n)^{3/2}}\sum_{j\neq i}^{n}\left\|\widetilde{\mathbf{\Delta}}_n^{-1}\frac{1}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}-\frac{\mathbf{x}_{0i}}{2(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}}\right\|_2^3\mathbb{E}_0(|A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}|^3)\lesssim\frac{n\rho_n}{(n\rho_n)^{3/2}}\to 0.$$

The variance of the sum of these independent random vectors is

$$\begin{aligned}
&\frac{1}{n}\sum_{j\neq i}\left\{\widetilde{\mathbf{\Delta}}_n^{-1}\frac{\mathbf{x}_{0j}}{\sqrt{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}}-\frac{\mathbf{x}_{0i}}{2(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}}\right\}\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\left\{\widetilde{\mathbf{\Delta}}_n^{-1}\frac{\mathbf{x}_{0j}}{\sqrt{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}}-\frac{\mathbf{x}_{0i}}{2(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}}\right\}^{\mathrm{T}}\\
&=\frac{1}{n}\sum_{j\neq i}\left\{\frac{\widetilde{\mathbf{\Delta}}_n^{-1}(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)\mathbf{x}_{0j}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}}-\frac{\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0j}}{2(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}}\right\}\frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)^2}\left\{\frac{\widetilde{\mathbf{\Delta}}_n^{-1}(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)\mathbf{x}_{0j}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}}-\frac{\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0j}}{2(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}}\right\}^{\mathrm{T}}\\
&=\frac{1}{n(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)}\sum_{j\neq i}\left(\widetilde{\mathbf{\Delta}}_n^{-1}-\frac{\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n}\right)\frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)^2}\left(\widetilde{\mathbf{\Delta}}_n^{-1}-\frac{\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n}\right)^{\mathrm{T}}\\
&\to\widetilde{\mathbf{\Sigma}}(\mathbf{x}_{0i}).
\end{aligned}$$

The proof of the asymptotic normality of $\mathbf{W}^{\mathrm{T}}\breve{\mathbf{x}}_i-\mathbf{y}_{0i}$ is thus completed by the Lyapunov's central limit theorem. $\square$

E.3 Some Technical Lemmas for the LSE

We next focus on the proof of the asymptotic normality (4.4) of the rows of the LSE under the condition (i). The proof strategy is enormously different than that presented in Tang and Priebe (2018), where the authors make use of the exchangeability property of $\mathbf{A}$ and $\mathbf{x}_{01},\ldots,\mathbf{x}_{0n}$ and such a exchangeability no longer holds in our current setup. Here we shall take a different approach and follow the framework of Cape et al. (2019) to establish the asymptotic normality (4.4). In preparation for doing so, we need to establish a collection of preliminary results first.

Lemma E.1 *Assume the conditions of Theorem 6 hold under the dense regime (i). Denote*

$$\widetilde{\mathbf{E}}_1=n\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}+\frac{n}{2}\mathbf{T}^{-1/2}\mathbf{P}_0\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})+\frac{n}{2}\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})\mathbf{P}_0\mathbf{T}^{-1/2}.$$

Let $\widetilde{\mathbf{u}}_{01}, \ldots, \widetilde{\mathbf{u}}_{0d}$ be the column vectors of $\widetilde{\mathbf{U}}_{\mathbf{P}}$. Then there exist constants $C_{\widetilde{\mathbf{E}}} > 0$, $\nu > 0$, such that

$$\mathbb{P}_0\left[\bigcup_{t=1}^{2}\bigcup_{i=1}^{n}\bigcup_{k=1}^{d}\left\{|\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}_1^t\widetilde{\mathbf{u}}_{0k}| > C_{\widetilde{\mathbf{E}}}^t(\log n)^{2t}n^{t/2}\|\widetilde{\mathbf{u}}_{0k}\|_\infty\right\}\right] \leq \exp\{-\nu(\log n)^2\}.$$

Proof. Let $H_{ij} = n^{-1/2}(A_{ij} - \mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})$ and $\mathbf{H} = [H_{ij}]_{n\times n}$. We first consider the case of $t = 1$. Let $\widetilde{\mathbf{u}}_{0k} = [\widetilde{u}_{01k}, \ldots, \widetilde{u}_{0nk}]^{\mathrm{T}}$. Then for any $i \in [n]$ and $k \in [d]$, we have

$$\begin{aligned}
|\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}_1\widetilde{\mathbf{u}}_{0k}| &\leq |\mathbf{e}_i^{\mathrm{T}}n\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\widetilde{\mathbf{u}}_{0k}| + \frac{n}{2}|\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}\mathbf{P}_0\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{u}}_{0k}| \\
&\quad + \frac{n}{2}|\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}_0\mathbf{T}^{-1/2}\widetilde{\mathbf{u}}_{0k}| \\
&\leq |\mathbf{e}_i^{\mathrm{T}}n\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\widetilde{\mathbf{u}}_{0k}| + n\|\mathbf{T}^{-1/2}\|_\infty\|\mathbf{P}_0\|_\infty\|\mathbf{T}^{-3/2}\|_\infty\|\mathbf{T} - \mathbf{D}\|_\infty\|\widetilde{\mathbf{u}}_{0k}\|_\infty.
\end{aligned}$$

Since $\|\mathbf{T}^{-1}\|_\infty = \|\mathbf{T}^{-1/2}\|_2 = O(n^{-1/2})$, $\|\mathbf{P}_0\|_\infty = O(n)$, and $\|\widetilde{\mathbf{S}}_{\mathbf{P}}\|_\infty = \|\widetilde{\mathbf{S}}_{\mathbf{P}}\|_2 = O(1)$, the second term of the previous display is upper bounded by a constant multiple of $\|\widetilde{\mathbf{u}}_{0k}\|_\infty\|\mathbf{T} - \mathbf{D}\|_\infty$. By the Chernoff bound, we see that $\|\mathbf{T} - \mathbf{D}\|_\infty$ is upper bounded by a constant multiple of $\sqrt{n}\log n$ with probability at least $1 - \exp\{-c(\log n)^2\}$ for a constant $c > 0$. Hence we conclude

$$n\|\mathbf{T}^{-1/2}\|_\infty\|\mathbf{P}_0\|_\infty\|\mathbf{T}^{-3/2}\|_\infty\|\mathbf{T} - \mathbf{D}\|_\infty\|\widetilde{\mathbf{u}}_{0k}\|_\infty \lesssim (\log n)n^{1/2}\|\widetilde{\mathbf{u}}_{0k}\|_\infty$$

with probability at least $1 - \exp\{-c(\log n)^2\}$. For the first term, it has the form

$$\left|\sum_{j=1}^{n}\frac{\{A_{ij} - \mathbb{E}_0(A_{ij})\}\widetilde{u}_{0jk}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}\right|,$$

which can be upper bounded by a constant of multiple of $\log n \asymp (\log n)n^{1/2}\|\widetilde{\mathbf{u}}_{0k}\|_\infty$ with probability at least $1 - \exp\{-c(\log n)^2\}$ for a constant $c > 0$ by Hoeffding's inequality. Hence we conclude from the union bound over $i \in [n]$ and $k \in [d]$ that

$$\mathbb{P}_0\left[\bigcup_{i=1}^{n}\bigcup_{k=1}^{d}\left\{|\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}_1\widetilde{\mathbf{u}}_{0k}| > C_{\widetilde{\mathbf{E}}_1}(\log n)^2n^{1/2}\|\widetilde{\mathbf{u}}_{0k}\|_\infty\right\}\right] \leq \exp\{-\nu_1(\log n)^2\}$$

for some constants $C_{\widetilde{\mathbf{E}}_1} > 0$ and $\nu_1 > 0$ because $ne^{-c(\log n)^2} \leq e^{-c/2(\log n)^2}$ for sufficiently large n.
We next consider the case where $t = 2$, which is slightly more involved. Write

$$\begin{aligned}
\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}_1^2\widetilde{\mathbf{u}}_{0k} &= n^2\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\widetilde{\mathbf{u}}_{0k} + \frac{n^2}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1}\mathbf{P}_0\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{u}}_{0k} \\
&\quad + \frac{n^2}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-2}(\mathbf{T} - \mathbf{D})\mathbf{P}_0\mathbf{T}^{-1/2}\widetilde{\mathbf{u}}_{0k} \\
&\quad + \frac{n^2}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}\mathbf{P}_0\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\widetilde{\mathbf{u}}_{0k} \\
&\quad + \frac{n^2}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}_0\mathbf{T}^{-1}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\widetilde{\mathbf{u}}_{0k} + \mathbf{e}_i^{\mathrm{T}}\mathbf{R}_{\widetilde{\mathbf{E}}_1}^{(2)}\widetilde{\mathbf{u}}_{0k},
\end{aligned}$$

where

$$\|\mathbf{R}^{(2)}_{\widetilde{\mathbf{E}}_1}\|_\infty \lesssim n^2\|\mathbf{T}^{-1/2}\|_\infty^2\|\mathbf{P}_0\|_\infty^2\|\mathbf{T}^{-3/2}\|_\infty^2\|\mathbf{T}-\mathbf{D}\|_\infty^2 \lesssim \|\mathbf{T}-\mathbf{D}\|_\infty^2.$$

Since Chernoff bound implies $\|\mathbf{T}-\mathbf{D}\|_\infty^2 \lesssim n(\log n)^2$ with probability at least $1-2\exp\{-c(\log n)^2\}$, it follows that

$$|\mathbf{e}_i^{\mathrm{T}}\mathbf{R}^{(2)}_{\widetilde{\mathbf{E}}_1}\widetilde{\mathbf{u}}_{0k}| \leq \|\mathbf{R}^{(2)}_{\widetilde{\mathbf{E}}_1}\|_\infty\|\widetilde{\mathbf{u}}_{0k}\|_\infty \lesssim n(\log n)^2\|\widetilde{\mathbf{u}}_{0k}\|_\infty$$

with probability at least $1-2\exp\{-c(\log n)^2\}$. We then analyze the first five terms on the right-hand side of $\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}_1^2\widetilde{\mathbf{u}}_{0k}$ separately. Let $i\in[n]$ and $k\in[d]$ be fixed.

(1) For the first term, we observe that

$$\begin{aligned}
&n^2\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}\widetilde{\mathbf{u}}_{0k}\\
&\quad=\sum_{i_1=1}^n\sum_{i_2=1}^n\frac{\{A_{ii_1}-\mathbb{E}_0(A_{ii_1})\}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0i_1}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}\frac{\{A_{i_1i_2}-\mathbb{E}_0(A_{i_1i_2})\}}{(\mathbf{x}_{0i_1}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0i_2}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}\widetilde{u}_{0i_2k}\\
&\quad=n\sum_{i_1=1}^n\sum_{i_2=1}^n H_{ii_1}H_{i_1i_2}\frac{\widetilde{u}_{0i_2k}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0i_1}^{\mathrm{T}}\boldsymbol{\mu}_n)(\mathbf{x}_{0i_2}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}.
\end{aligned}$$

By exploiting the proof of Lemma B.1, we immediately conclude that

$$\mathbb{E}_0\{|n^2\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}\widetilde{\mathbf{u}}_{0k}|^p\} \leq n^p(2Cp)^{2p}\|\widetilde{\mathbf{u}}_{0k}\|_\infty^p$$

for any p for a constant $C\geq 1$, and hence, with $p=\lfloor(\log n)^2/(4C)\rfloor$,

$$\begin{aligned}
&\mathbb{P}_0\left\{|n^2\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}\widetilde{\mathbf{u}}_{0k}| > C^2(\log n)^4 n\|\widetilde{\mathbf{u}}_{0k}\|_\infty\right\}\\
&\quad\leq \frac{n^p(2Cp)^{2p}\|\widetilde{\mathbf{u}}_{0k}\|_\infty^p}{C^{2p}(\log n)^{4p}n^p\|\widetilde{\mathbf{u}}_{0k}\|_\infty^p} = \left\{\frac{2p}{(\log n)^2}\right\}^{2p} = \exp\{-c(\log n)^2\}
\end{aligned}$$

by Markov's inequality for a constant $c>0$.

(2) For the second term, write

$$\begin{aligned}
&n^2\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1}\mathbf{P}_0\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})\widetilde{\mathbf{u}}_{0k}\\
&\quad=-\sum_{j=1}^n\sum_{l=1}^n\sum_{m=1}^n\frac{\{A_{ij}-\mathbb{E}_0(A_{ij})\}\{A_{lm}-\mathbb{E}_0(A_{lm})\}\mathbf{y}_{0j}^{\mathrm{T}}\mathbf{y}_{0l}\widetilde{u}_{0lk}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n}\\
&\quad=-\left[\sum_{j=1}^n\frac{\{A_{ij}-\mathbb{E}_0(A_{ij})\}\mathbf{y}_{0j}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}\right]^{\mathrm{T}}\left[\sum_{l=1}^n\sum_{m=1}^n\frac{\{A_{lm}-\mathbb{E}_0(A_{lm})\}\mathbf{y}_{0l}\widetilde{u}_{0lk}}{\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n}\right]\\
&\quad=-\left[\frac{1}{\sqrt{n}}\sum_{j=1}^n\frac{\{A_{ij}-\mathbb{E}_0(A_{ij})\}\mathbf{x}_{0j}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}\right]^{\mathrm{T}}\left[\sum_{l=1}^n\sum_{m=1}^n\frac{\{A_{lm}-\mathbb{E}_0(A_{lm})\}\mathbf{x}_{0l}\widetilde{u}_{0lk}}{\sqrt{n}(\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}}\right]
\end{aligned}$$

For the first factor on the right-hand side of the last display, we obtain directly from Hoeffding's inequality that

$$\left\| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\{A_{ij} - \mathbb{E}_0(A_{ij})\}\mathbf{x}_{0j}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n} \right\|_2 \leq C \log n$$

with probability at least $1 - 2\exp\{-c(\log n)^2\}$ for constants $c, C > 0$. The second factor can be written as

$$\begin{aligned}
&\sum_{l=1}^{n}\sum_{m=1}^{n} \frac{\{A_{lm} - \mathbb{E}_0(A_{lm})\}\mathbf{x}_{0l}\widetilde{u}_{0lk}}{\sqrt{n}(\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}} \\
&\quad = \sum_{l<m} \left\{ \frac{\mathbf{x}_{0l}\widetilde{u}_{0lk}}{\sqrt{n}(\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}} + \frac{\mathbf{x}_{0m}\widetilde{u}_{0mk}}{\sqrt{n}(\mathbf{x}_{0m}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}} \right\} \{A_{lm} - \mathbb{E}_0(A_{lm})\} - \sum_{l=1}^{n} \frac{(\mathbf{x}_{0l}^{\mathrm{T}}\mathbf{x}_{0l})\mathbf{x}_{0l}\widetilde{u}_{0lk}}{\sqrt{n}(\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}} \\
&\quad = \sum_{l<m} \left\{ \frac{\mathbf{x}_{0l}\widetilde{u}_{0lk}}{\sqrt{n}(\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}} + \frac{\mathbf{x}_{0m}\widetilde{u}_{0mk}}{\sqrt{n}(\mathbf{x}_{0m}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}} \right\} \{A_{lm} - \mathbb{E}_0(A_{lm})\} + O(\sqrt{n}\|\widetilde{\mathbf{u}}_{0k}\|_\infty).
\end{aligned}$$

Similarly, the first sum on the right-hand side of the previous display is bounded by a constant multiple of $\sqrt{n}\|\widetilde{\mathbf{u}}_{0k}\|_\infty \log n$ with probability at least $1 - 2\exp\{-c(\log n)^2\}$ for a constant $c > 0$ by Hoeffding's inequality. Hence we conclude that

$$\begin{aligned}
&\mathbb{P}_0 \left\{ |\mathbf{e}_i^{\mathrm{T}} n^2 \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1}\mathbf{P}_0\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{u}}_{0k}| > Cn(\log n)^4 \|\widetilde{\mathbf{u}}_{0k}\|_\infty \right\} \\
&\quad \leq \mathbb{P}_0 \left\{ |\mathbf{e}_i^{\mathrm{T}} n^2 \mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1}\mathbf{P}_0\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\widetilde{\mathbf{u}}_{0k}| > C\sqrt{n}(\log n)^2 \|\widetilde{\mathbf{u}}_{0k}\|_\infty \right\} \\
&\quad \leq 2\exp\{-c(\log n)^2\}
\end{aligned}$$

for some constants $C, c > 0$.

(3) The third term can be written as

$$\begin{aligned}
&\frac{n^2}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{T}^{-1}\mathbf{D})(\mathbf{Y}_0\mathbf{Y}_0^{\mathrm{T}}\widetilde{\mathbf{u}}_{0k}) \\
&\quad = \frac{n}{2}\sum_{j=1}^{n} \left\{ \frac{A_{ij} - \mathbb{E}_0(A_{ij})}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}} \right\} \left\{ 1 - \frac{\sum_{l=1}^{n} A_{jl}}{\sum_{l=1}^{n}\mathbb{E}_0(A_{jl})} \right\} \mathbf{e}_j^{\mathrm{T}}\mathbf{Y}_0\mathbf{Y}_0^{\mathrm{T}}\widetilde{\mathbf{u}}_{0k} \\
&\quad = -\frac{1}{2}\sum_{j=1}^{n}\sum_{l=1}^{n} \left\{ \frac{A_{ij} - \mathbb{E}_0(A_{ij})}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}} \right\} \left\{ \frac{A_{jl} - \mathbb{E}_0(A_{jl})}{\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n} \mathbf{e}_j^{\mathrm{T}}\mathbf{Y}_0\mathbf{Y}_0^{\mathrm{T}}\widetilde{\mathbf{u}}_{0k} \right\} \\
&\quad = -\frac{n}{2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} H_{ii_1}H_{i_1i_2} \frac{\mathbf{e}_{i_1}^{\mathrm{T}}\mathbf{Y}_0\mathbf{Y}_0^{\mathrm{T}}\widetilde{\mathbf{u}}_{0k}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0i_1}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}}.
\end{aligned}$$

Since

$$\max_{i,j\in[n]} \left| \frac{\mathbf{e}_j^{\mathrm{T}}\mathbf{Y}_0\mathbf{Y}_0^{\mathrm{T}}\widetilde{\mathbf{u}}_{0k}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}} \right| \lesssim \max_{j\in[n]} \left| \mathbf{y}_{0j}^{\mathrm{T}} \sum_{m=1}^{n} \mathbf{y}_{0m}(\mathbf{e}_m^{\mathrm{T}}\mathbf{u}_{0k}) \right| \lesssim \|\widetilde{\mathbf{u}}_{0k}\|_\infty,$$

we then apply the proof of Lemma B.1 and obtain that the pth moment on the right-hand side can be bounded as follows

$$\mathbb{E}_0\left\{\left|\frac{n^2}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{T}^{-1}\mathbf{D})(\mathbf{Y}_0\mathbf{Y}_0^{\mathrm{T}}\widetilde{\mathbf{u}}_{0k})\right|^p\right\}\leq n^p(2Cp)^{2p}\|\widetilde{\mathbf{u}}_{0k}\|_\infty^p.$$

Hence we conclude by Markov's inequality that

$$\mathbb{P}_0\left\{\left|\frac{n^2}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{T}^{-1}\mathbf{D})(\mathbf{Y}_0\mathbf{Y}_0^{\mathrm{T}}\widetilde{\mathbf{u}}_{0k})\right|>Cn(\log n)^4\|\widetilde{\mathbf{u}}_{0k}\|_\infty\right\}\leq\exp\{-c(\log n)^2\}$$

for some constants $C,c>0$. The argument for the detailed derivation is exactly the same as (1).

(4) For the fourth term, write

$$\begin{aligned}
&n^2\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}\mathbf{P}_0\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}\widetilde{\mathbf{u}}_{0k}\\
&=\mathbf{y}_{0i}^{\mathrm{T}}[\mathbf{y}_{01},\ldots,\mathbf{y}_{0n}]\mathrm{diag}\left\{\left(\frac{\sum_{l=1}^n\mathbb{E}_0(A_{jl})-\sum_{l=1}^nA_{jl}}{\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}\right)_{j=1}^n\right\}\left[\frac{A_{jm}-\mathbb{E}_0(A_{jm})}{(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0m}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}\right]_{n\times n}\widetilde{\mathbf{u}}_{0k}\\
&=-\frac{1}{n}\sum_{j=1}^n\sum_{l=1}^n\sum_{m=1}^n\frac{\{(A_{jl}-\mathbb{E}_0(A_{jl})\}\{(A_{jm}-\mathbb{E}_0(A_{jm})\}\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0m}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}\widetilde{u}_{0mk}\\
&=-\frac{1}{n}\sum_{l=1}^n\sum_{j=1}^n\sum_{m=1}^n\frac{\{(A_{lj}-\mathbb{E}_0(A_{lj})\}\{(A_{jm}-\mathbb{E}_0(A_{jm})\}\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0m}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}\widetilde{u}_{0mk}=-\frac{1}{n}\sum_{l=1}^nz_{ikl},
\end{aligned}$$

where

$$\begin{aligned}
z_{ikl}&=\sum_{j=1}^n\sum_{m=1}^n\frac{\{(A_{lj}-\mathbb{E}_0(A_{lj})\}\{(A_{jm}-\mathbb{E}_0(A_{jm})\}\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0m}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}\widetilde{u}_{0mk}\\
&=n\sum_{i_1=1}^n\sum_{i_2=1}^nH_{li_1}H_{i_1i_2}\frac{\widetilde{u}_{0i_2k}}{(\mathbf{x}_{0i_1}^{\mathrm{T}}\boldsymbol{\mu}_n)(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0i_1}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0i_1}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0i_2}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}.
\end{aligned}$$

An argument similar to the proof of Lemma B.1 yields

$$\mathbb{P}_0(|z_{ikl}|>Cn(\log n)^4\|\widetilde{\mathbf{u}}_{0k}\|_\infty)\leq\exp\{-c(\log n)^2\}$$

for some constants $C,c>0$ that does not depend on i,k,l. Since $\log n\ll(\log n)^2$, we apply the union bound to obtain

$$\mathbb{P}_0\left(\left|\frac{1}{n}\sum_{l=1}^nz_{ikl}\right|>Cn(\log n)^2\|\widetilde{\mathbf{u}}_{0k}\|_\infty\right)\leq\sum_{l=1}^n\mathbb{P}_0(|z_{ikl}|>Cn(\log n)^4\|\widetilde{\mathbf{u}}_{0k}\|_\infty)\leq\exp\{-c(\log n)^2\},$$

for some constant $c>0$.

(5) We write the fifth term as

$$\begin{aligned}
&\frac{n^2}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})\mathbf{P}_0\mathbf{T}^{-1}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}\widetilde{\mathbf{u}}_{0k}\\
&\quad=\frac{1}{2}\left\{\sum_{j=1}^n\frac{\mathbb{E}_0(A_{ij})-A_{ij}}{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n}\right\}\mathbf{e}_i^{\mathrm{T}}\mathbf{Y}_0\mathbf{Y}_0^{\mathrm{T}}\left[\frac{A_{lm}-\mathbb{E}_0(A_{lm})}{(\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0m}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}\right]_{n\times n}\widetilde{\mathbf{u}}_{0k}\\
&\quad=\frac{1}{2}\left\{\sum_{j=1}^n\frac{\mathbb{E}_0(A_{ij})-A_{ij}}{\sqrt{n}\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n}\right\}\frac{1}{\sqrt{n}}\sum_{l=1}^n\sum_{m=1}^n\frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0l}\{A_{lm}-\mathbb{E}_0(A_{lm})\}\widetilde{u}_{0mk}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0m}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}.
\end{aligned}$$

By Hoeffding's inequality, the first factor can be bounded:

$$\mathbb{P}_0\left(\left|\sum_{j=1}^n\frac{\mathbb{E}_0(A_{ij})-A_{ij}}{\sqrt{n}\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n}\right|>C\log n\right)\leq 2\exp\{-c(\log n)^2\}.$$

We further write the second factor as

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{l=1}^n\sum_{m=1}^n\frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0l}\{A_{lm}-\mathbb{E}_0(A_{lm})\}\widetilde{u}_{0mk}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0m}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}\\
&\quad=\sum_{l<m}\frac{\zeta_{iklm}}{\sqrt{n}}\{A_{lm}-\mathbb{E}_0(A_{lm})\}-\sum_{l=1}^m\frac{\mathbf{x}_{0l}^{\mathrm{T}}\mathbf{x}_{0l}\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0l}\widetilde{u}_{0lk}}{\sqrt{n}(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n}\\
&\quad=\sum_{l<m}\frac{\zeta_{iklm}}{\sqrt{n}}\{A_{lm}-\mathbb{E}_0(A_{lm})\}+O(\sqrt{n}\|\widetilde{\mathbf{u}}_{0k}\|_\infty),
\end{aligned}$$

where

$$\zeta_{iklm}=\left\{\frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0l}\widetilde{u}_{0mk}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0m}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}+\frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0m}\widetilde{u}_{0lk}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0m}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0m}^{\mathrm{T}}\boldsymbol{\mu}_n\mathbf{x}_{0l}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}}\right\}.$$

Clearly, $\max_{i,k,l,m}|\zeta_{iklm}|\lesssim\|\widetilde{\mathbf{u}}_{0k}\|_\infty$. It follows from Hoeffding's inequality that

$$\mathbb{P}_0\left\{\left|\sum_{l<m}\frac{\zeta_{iklm}}{\sqrt{n}}\{A_{lm}-\mathbb{E}_0(A_{lm})\}\right|>C\sqrt{n}\|\widetilde{\mathbf{u}}_{0k}\|_\infty(\log n)\right\}\leq\exp\{-c(\log n)^2\}$$

for some constants $C,c>0$. We thus conclude that

$$\mathbb{P}_0\left\{\left|\frac{n^2}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})\mathbf{P}_0\mathbf{T}^{-1}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}\widetilde{\mathbf{u}}_{0k}\right|>C\sqrt{n}\|\widetilde{\mathbf{u}}_{0k}\|_\infty(\log n)^2\right\}\leq 2\exp\{-c(\log n)^2\}$$

for some constants $C,c>0$.

Now combining the analyses (1)-(5) together with the bound on $|\mathbf{e}_i^{\mathrm{T}}\mathbf{R}^{(2)}_{\widetilde{\mathbf{E}}_1}\widetilde{\mathbf{u}}_{0k}|$ yields

$$\mathbb{P}_0\left\{|\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}_1^2\widetilde{\mathbf{u}}_{0k}|>Cn(\log n)^4\|\widetilde{\mathbf{u}}_{0k}\|_\infty\right\}\leq\exp\{-c(\log n)^2\}$$

for some constants $C, c > 0$. Hence, we can take the union bound over all $i \in [n]$ and $k \in [d]$ to conclude that

$$\mathbb{P}_0\left[\bigcup_{i\in[n]}\bigcup_{k\in[d]}\left\{|\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}_1^2\widetilde{\mathbf{u}}_{0k}| > C_{\widetilde{\mathbf{E}}_1}^2 n(\log n)^4\|\widetilde{\mathbf{u}}_{0k}\|_\infty\right\}\right] \leq \exp\{-\nu_2(\log n)^2\}$$

The proof is completed by taking the union bound over $t = 1$ and $t = 2$. $\square$

Lemma E.2 *Assume the conditions of Theorem 6 hold and assume that $\rho_n^{-1} \lesssim (\log n)^\epsilon$ for some $\epsilon > 0$. Then for any fixed $i \in [n]$,*

$$\left[\frac{1}{\sqrt{n\rho_n}}\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}(\sqrt{n}\mathbf{Y}_0)\right]^{\mathrm{T}} = \frac{1}{\sqrt{n\rho_n}}\sum_{j=1}^n \widetilde{E}_{ij}\sqrt{n}\mathbf{y}_{0j} \overset{\mathcal{L}}{\to} \mathrm{N}(\mathbf{0}, \widetilde{\boldsymbol{\Delta}}\widetilde{\boldsymbol{\Sigma}}(\mathbf{x}_{0i})\widetilde{\boldsymbol{\Delta}})$$

Proof. We begin the proof by first remarking that Lemma B.4 of Tang and Priebe (2018) also holds when $\mathbf{A} \sim \mathrm{RDPG}(\mathbf{X}_0)$ with a deterministic latent position matrix $\mathbf{X}_0$ and recalling the decomposition of the random vector of interest in Appendix B of Tang and Priebe (2018):

$$\begin{aligned}
\frac{1}{\sqrt{n\rho_n}}\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}(\sqrt{n}\mathbf{Y}_0) &= \sqrt{n\rho_n}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}(\sqrt{n}\mathbf{Y}_0) + \frac{\sqrt{n\rho_n}}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})\mathbf{P}_0\mathbf{T}^{-1/2}(\sqrt{n}\mathbf{Y}_0)\\
&\quad + \frac{\sqrt{n\rho_n}}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}\mathbf{P}_0\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})(\sqrt{n}\mathbf{Y}_0)\\
&\quad + \frac{\sqrt{n\rho_n}}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)(\mathbf{D}^{-1/2}-\mathbf{T}^{-1/2})(\sqrt{n}\mathbf{Y}_0)\\
&\quad + \frac{\sqrt{n\rho_n}}{2}\mathbf{e}_i^{\mathrm{T}}(\mathbf{D}^{-1/2}-\mathbf{T}^{-1/2})(\mathbf{A}-\mathbf{P}_0)\mathbf{D}^{-1/2}(\sqrt{n}\mathbf{Y}_0)\\
&\quad + \frac{\sqrt{n\rho_n}}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})\mathbf{P}_0(\mathbf{D}^{-1/2}-\mathbf{T}^{-1/2})(\sqrt{n}\mathbf{Y}_0)\\
&\quad + \frac{\sqrt{n\rho_n}}{2}\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{R}}^{(-1/2)}(\sqrt{n}\mathbf{Y}_0),
\end{aligned}\tag{E.2}$$

where $\|\widetilde{\mathbf{R}}^{(-1/2)}\|_\infty = O_{\mathbb{P}_0}((n\rho_n)^{-1}\log n)$ by exploiting the proof of (B.11) and (B.13) in Appendix B.2 of Tang and Priebe (2018), together with the fact that $\|\mathbf{P}_0\|_\infty = O(n\rho_n)$. We prove the desired asymptotic normality result by showing that the decomposition on the right-hand side of (E.2) is dominated by the first two terms, whereas the remainders are asymptotically negligible.

We next show that the remainders are $o_{\mathbb{P}_0}(1)$ by establishing the following results.

(1) $\sqrt{n\rho_n}\|\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}\mathbf{P}_0\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})(\sqrt{n}\mathbf{Y}_0)\|_2 = o_{\mathbb{P}_0}(1)$. Observing that $\|\widetilde{\mathbf{S}}_{\mathbf{P}}\|_2 = O(1)$ and $\|\widetilde{\mathbf{U}}_{\mathbf{P}}\|_{2\to\infty} = \|\mathbf{Y}_0\mathbf{S}_{\mathbf{P}}^{-1/2}\|_{2\to\infty} \lesssim n^{-1/2}$, we then write

$$\begin{aligned}
\sqrt{n\rho_n}\|\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}\mathbf{P}_0\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})(\sqrt{n}\mathbf{Y}_0)\|_2 &\leq n\sqrt{\rho_n}\|\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}\widetilde{\mathbf{U}}_{\mathbf{P}}^{\mathrm{T}}(\mathbf{T}-\mathbf{D})\mathbf{T}^{-1}\widetilde{\mathbf{U}}_{\mathbf{P}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\|_{2\to\infty}\\
&\lesssim n\sqrt{\rho_n}\|\widetilde{\mathbf{U}}_{\mathbf{P}}\|_{2\to\infty}\|\widetilde{\mathbf{U}}_{\mathbf{P}}^{\mathrm{T}}(\mathbf{T}-\mathbf{D})\mathbf{T}^{-1}\widetilde{\mathbf{U}}_{\mathbf{P}}\|_2\\
&\lesssim \|\sqrt{n\rho_n}\widetilde{\mathbf{U}}_{\mathbf{P}}^{\mathrm{T}}(\mathbf{T}-\mathbf{D})\mathbf{T}^{-1}\widetilde{\mathbf{U}}_{\mathbf{P}}\|_2.
\end{aligned}$$

Let $\widetilde{u}_{0jk}$ be the (j,k) entry of $\widetilde{\mathbf{U}}_{\mathbf{P}}$. Clearly, $\max_{j,k}|\widetilde{u}_{0jk}|\leq\|\widetilde{\mathbf{U}}_{\mathbf{P}}\|_{2\to\infty}\lesssim n^{-1/2}$. We consider the (k,l) entry of $\sqrt{n\rho_n}\widetilde{\mathbf{U}}_{\mathbf{P}}^{\mathrm{T}}(\mathbf{T}-\mathbf{D})\mathbf{T}^{-1}\widetilde{\mathbf{U}}_{\mathbf{P}}$, which can be written as

$$
\begin{aligned}
\sum_{i=1}^n\sum_{j=1}^n\frac{\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}-A_{ij}}{\sqrt{n\rho_n(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)^2}}\widetilde{u}_{0ik}\widetilde{u}_{0il}&=-2\sum_{i<j}\frac{A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{\sqrt{n\rho_n}(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)}\widetilde{u}_{0ik}\widetilde{u}_{0il}+\sum_{i=1}^n\frac{\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{\sqrt{n\rho_n}(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)}\widetilde{u}_{0ik}\widetilde{u}_{0il}\\
&=-2\sum_{i<j}\frac{A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{\sqrt{n\rho_n}(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)}\widetilde{u}_{0ik}\widetilde{u}_{0il}+o\left(1\right).
\end{aligned}
$$

By Hoeffding's inequality, for any $t>0$,

$$
\mathbb{P}_0\left(\left|\sum_{i<j}\frac{A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{\sqrt{n\rho_n}(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)}\widetilde{u}_{0ik}\widetilde{u}_{0il}\right|>t\right)\leq 2\exp\left\{-t^2\left(\sum_{i<j}\frac{\widetilde{u}_{0ik}^2\widetilde{u}_{0il}^2}{n\rho_n}\right)^{-1}\right\}=2\exp(-n\rho_nt^2)\to 0.
$$

Hence we conclude that

$$
\sqrt{n\rho_n}\|\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}\mathbf{P}_0\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})(\sqrt{n}\mathbf{Y}_0)\|_2\lesssim\sum_{k,l\in[d]}\left|\sum_{i<j}\frac{A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{\sqrt{n\rho_n}(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)}\widetilde{u}_{0ik}\widetilde{u}_{0il}\right|+o\left(1\right)=o_{\mathbb{P}_0}(1).
$$

(2) $\sqrt{n\rho_n}\|\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)(\mathbf{D}^{-1/2}-\mathbf{T}^{-1/2})(\sqrt{n}\mathbf{Y}_0)\|_2=o_{\mathbb{P}_0}(1)$. By Lemma B.4 of Tang and Priebe (2018),

$$
\mathbf{D}^{-1/2}-\mathbf{T}^{-1/2}=\frac{1}{2}\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})+O_{\mathbb{P}_0}((n\rho_n)^{-3/2}\log n).
$$

Therefore,

$$
\begin{aligned}
&\sqrt{n\rho_n}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)(\mathbf{D}^{-1/2}-\mathbf{T}^{-1/2})\sqrt{n}\mathbf{Y}_0\\
&\quad=\frac{1}{2}\sqrt{n\rho_n}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)(\mathbf{T}-\mathbf{D})\mathbf{T}^{-3/2}\sqrt{n}\mathbf{Y}_0+\sqrt{n}O_{\mathbb{P}_0}\{\|\mathbf{A}-\mathbf{P}_0\|_2(n\rho_n)^{-3/2}\log n\}\\
&\quad=\frac{1}{2}\sqrt{n\rho_n}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)(\mathbf{T}-\mathbf{D})\mathbf{T}^{-3/2}\sqrt{n}\mathbf{Y}_0+O_{\mathbb{P}_0}(\sqrt{n}(n\rho_n)^{-1}\log n)\\
&\quad=\frac{1}{2}\sqrt{n\rho_n}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)(\mathbf{T}-\mathbf{D})\mathbf{T}^{-3/2}\sqrt{n}\mathbf{Y}_0+o_{\mathbb{P}_0}(1).
\end{aligned}
$$

Let $[\mathbf{Y}_0]_{*k}$ be the kth column of $\mathbf{Y}_0$, and $\widetilde{\boldsymbol{\beta}}_k=\sqrt{n}\mathbf{T}^{-3/2}[\mathbf{Y}_0]_{*k}$. Then we consider

$$
Z_{ik}=-\frac{1}{2}\sqrt{n\rho_n}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)(\mathbf{T}-\mathbf{D})\mathbf{T}^{-3/2}\sqrt{n}[\mathbf{Y}_0]_{*k}.
$$

Observe that

$$
\mathbb{E}_0(Z_{ik})=\frac{1}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{P}_0-\mathbf{P}_0\circ\mathbf{P}_0)\mathbf{T}^{-3/2}\sqrt{n}[\mathbf{Y}_0]_{*k},
$$

by the computation of $\mathbb{E}_0\{(\mathbf{A}-\mathbf{P}_0)(\mathbf{T}-\mathbf{D})\}$, where $\circ$ is the Hadamard (entry-wise) matrix product operator, and that $\|\mathbf{Y}_0\|_\infty\lesssim n^{-1/2}$, we obtain

$$
|\mathbb{E}_0(Z_{ik})|\leq\frac{1}{2}\|\mathbf{T}^{-1/2}\|_\infty\|\mathbf{P}_0\|_\infty\|\mathbf{T}^{-3/2}\|_\infty\sqrt{n}\|\mathbf{Y}_0\|_\infty=O\{(n\rho_n)^{-1/2}(n\rho_n)(n\rho_n)^{-3/2}\}=o(1).
$$

Hence $Z_{ik} = o_{\mathbb{P}_0}(1)$ by Markov's inequality, and hence,

$$\sqrt{n\rho_n}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})\sqrt{n}\mathbf{Y}_0 = o_{\mathbb{P}_0}(1).$$

(3) $\sqrt{n\rho_n}\|\mathbf{e}_i^{\mathrm{T}}(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})(\mathbf{A} - \mathbf{P}_0)\mathbf{D}^{-1/2}(\sqrt{n}\mathbf{Y}_0)\|_2 = o_{\mathbb{P}_0}(1)$. Again, by Lemma B.4 of Tang and Priebe (2018), we have $\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2} = O_{\mathbb{P}_0}((n\rho_n)^{-1}\log n)$. Therefore,

$$\begin{aligned}
&\sqrt{n\rho_n}\|\mathbf{e}_i^{\mathrm{T}}(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})(\mathbf{A} - \mathbf{P}_0)\mathbf{D}^{-1/2}(\sqrt{n}\mathbf{Y}_0)\|_2\\
&\quad\leq \sqrt{n\rho_n}\|\mathbf{e}_i^{\mathrm{T}}(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}(\sqrt{n}\mathbf{Y}_0)\|_2\\
&\qquad+ \sqrt{n\rho_n}\|\mathbf{e}_i^{\mathrm{T}}(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})(\mathbf{A} - \mathbf{P}_0)(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})(\sqrt{n}\mathbf{Y}_0)\|_2\\
&\quad\leq \sqrt{n\rho_n}\|\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2}\|_2\left\|\sum_{j=1}^n \frac{A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{\sqrt{n\rho_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}}\frac{\mathbf{x}_{0j}}{\sqrt{\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}}\right\|_2\\
&\qquad+ \sqrt{n\rho_n}\|\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2}\|_2^2\|\mathbf{A} - \mathbf{P}_0\|_2\|\sqrt{n}\mathbf{Y}_0\|_2\\
&\quad= O_{\mathbb{P}_0}\left((n\rho_n)^{-1/2}\log n\right)\left\|\sum_{j=1}^n \frac{A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{\sqrt{n\rho_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}}\frac{\mathbf{x}_{0j}}{\sqrt{\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}}\right\|_2\\
&\qquad+ O_{\mathbb{P}_0}\left\{(n\rho_n)^{1/2}(n\rho_n)^{-2}(\log n)^2(n\rho_n)^{1/2}\sqrt{n}\right\}\\
&\quad= O_{\mathbb{P}_0}\left((n\rho_n)^{-1/2}\log n\right)\left\|\sum_{j=1}^n \frac{A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{\sqrt{n\rho_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}}\frac{\mathbf{x}_{0j}}{\sqrt{\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}}\right\|_2 + o_{\mathbb{P}_0}(1).
\end{aligned}$$

By the Lyapunov's central limit theorem,

$$\sum_{j=1}^n \frac{A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{\sqrt{n\rho_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}}\frac{\mathbf{x}_{0j}}{\sqrt{\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}} = O_{\mathbb{P}_0}(1).$$

Hence we conclude that $\sqrt{n\rho_n}\|\mathbf{e}_i^{\mathrm{T}}(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})(\mathbf{A} - \mathbf{P}_0)\mathbf{D}^{-1/2}(\sqrt{n}\mathbf{Y}_0)\|_2 = o_{\mathbb{P}_0}(1)$.

(4) $\sqrt{n\rho_n}\|\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}_0(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})(\sqrt{n}\mathbf{Y}_0)\|_2 = o_{\mathbb{P}_0}(1)$. Since $\mathbf{D}$ and $\mathbf{T}$ are diagonal matrices, we obtain

$$\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2} = O_{\mathbb{P}_0}((n\rho_n)^{-1}\log n) \quad\Longrightarrow\quad \|\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2}\|_\infty = O_{\mathbb{P}_0}((n\rho_n)^{-1}\log n).$$

Similarly, $\|\mathbf{T} - \mathbf{D}\|_\infty = O_{\mathbb{P}_0}((n\rho_n)^{1/2}\log n)$. It follows that

$$\begin{aligned}
&\sqrt{n\rho_n}\|\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}_0(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})(\sqrt{n}\mathbf{Y}_0)\|_2\\
&\quad\leq \sqrt{n\rho_n}\|\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}_0(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})(\sqrt{n}\mathbf{Y}_0)\|_\infty\\
&\quad\leq \sqrt{n\rho_n}\|\mathbf{T}^{-3/2}\|_\infty\|\mathbf{T} - \mathbf{D}\|_\infty\|\mathbf{P}_0\|_\infty\|\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2}\|_\infty\sqrt{n}\|\mathbf{Y}_0\|_\infty\\
&\quad\lesssim (n\rho_n)^{-1}O_{\mathbb{P}_0}((n\rho_n)^{1/2}\log n)(n\rho_n)O_{\mathbb{P}_0}((n\rho_n)^{-1}\log n)\\
&\quad= O_{\mathbb{P}_0}\{(n\rho_n)^{-1/2}(\log n)^2\} = o_{\mathbb{P}_0}(1).
\end{aligned}$$

(5) $\sqrt{n\rho_n}\|\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{R}}^{(-1/2)}(\sqrt{n}\mathbf{Y}_0)\|_2 = o_{\mathbb{P}_0}(1)$. This is a simple consequence of the result $\|\widetilde{\mathbf{R}}^{(-1/2)}\|_\infty = O_{\mathbb{P}_0}((n\rho_n)^{-1}\log n)$ and $\|\mathbf{Y}_0\|_\infty \lesssim n^{-1/2}$.

We next show the asymptotic normality of the first two terms. Denote $\widetilde{\boldsymbol{\Delta}}_n = \mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0$. According to the aforementioned analysis, we write

$$
\begin{aligned}
\frac{1}{\sqrt{n\rho_n}}\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}(\sqrt{n}\mathbf{Y}_0) &= \sqrt{n\rho_n}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-1/2}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}(\sqrt{n}\mathbf{Y}_0)\\
&\quad + \frac{\sqrt{n\rho_n}}{2}\mathbf{e}_i^{\mathrm{T}}\mathbf{T}^{-3/2}(\mathbf{T}-\mathbf{D})\mathbf{P}\mathbf{T}^{-1/2}(\sqrt{n}\mathbf{Y}_0) + o_{\mathbb{P}_0}(1)\\
&= \frac{1}{\sqrt{n\rho_n}}\sum_{j=1}^n \frac{A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)}\mathbf{x}_{0j}^{\mathrm{T}}\\
&\quad - \frac{1}{2\sqrt{n\rho_n}(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}}\sum_{j=1}^n (A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})(\widetilde{\boldsymbol{\Delta}}_n\mathbf{x}_{0i})^{\mathrm{T}} + o_{\mathbb{P}_0}(1).
\end{aligned}
$$

Equivalently, we have

$$
\left[\frac{1}{\sqrt{n\rho_n}}\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}(\sqrt{n}\mathbf{Y}_0)\right]^{\mathrm{T}} = \frac{1}{\sqrt{n\rho_n}}\sum_{j=1}^n\left\{\frac{\mathbf{x}_{0j}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{1/2}(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)} - \frac{1}{2}\frac{\widetilde{\boldsymbol{\Delta}}_n\mathbf{x}_{0i}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)^{3/2}}\right\}(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}) + o_{\mathbb{P}_0}(1).
$$

The first term on the right-hand side of the previous display is a sum of independent mean-zero random vectors, the variance of which is

$$
\begin{aligned}
&\frac{1}{n\rho_n}\sum_{j=1}^n\left\{\frac{\mathbf{x}_{0j}}{(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)} - \frac{1}{2}\frac{\widetilde{\boldsymbol{\Delta}}_n\mathbf{x}_{0i}}{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n}\right\}\frac{\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)}\left\{\frac{\mathbf{x}_{0j}}{(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)} - \frac{1}{2}\frac{\widetilde{\boldsymbol{\Delta}}_n\mathbf{x}_{0i}}{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n}\right\}^{\mathrm{T}}\\
&= \frac{1}{n}\sum_{j=1}^n\left\{\mathbf{x}_{0j} - \frac{1}{2}\frac{\widetilde{\boldsymbol{\Delta}}_n\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n}\right\}\frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)^2}\left\{\mathbf{x}_{0j} - \frac{1}{2}\frac{\widetilde{\boldsymbol{\Delta}}_n\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n}\right\}^{\mathrm{T}}\\
&= \widetilde{\boldsymbol{\Delta}}_n\frac{1}{n}\sum_{j=1}^n\left\{\widetilde{\boldsymbol{\Delta}}_n^{-1} - \frac{1}{2}\frac{\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}}{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n}\right\}\frac{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n)(\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n)^2}\left\{\widetilde{\boldsymbol{\Delta}}_n^{-1} - \frac{1}{2}\frac{\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}}{\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n}\right\}^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}}_n\\
&\to \widetilde{\boldsymbol{\Delta}}\widetilde{\boldsymbol{\Sigma}}(\mathbf{x}_{0i})\widetilde{\boldsymbol{\Delta}}.
\end{aligned}
$$

The proof is then completed by the Lyapunov's central limit theorem. □

E.4 Proof of the Asymptotic Normality (4.4) Under the Dense Regime (i)

Proof of the asymptotic normality (4.4) under the dense regime (i). Let $\widetilde{E}_{ij}$ be the (i,j)th entry. In particular, we need to verify that the following conditions hold:

A1. $\rho_n \to 0$ with $n\rho_n \gtrsim (\log n)^{c_2}$ for some $c_2 > 0$. This condition automatically holds since $\rho \equiv 1$ under the dense regime (i).

A2. $\lambda_d\{n\rho_n\mathcal{L}(\mathbf{P}_0)\} \gtrsim n\rho_n$ and $\lambda_d\{n\rho_n\mathcal{L}(\mathbf{P}_0)\}^{-1}\lambda_1\{n\rho_n\mathcal{L}(\mathbf{P}_0)\} = O(1)$. We see that

$$\lambda_k\{n\mathcal{L}(\mathbf{P}_0)\} = n\lambda_k(\mathbf{Y}_0\mathbf{Y}_0^{\mathrm{T}}) = n\lambda_k(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0) \asymp n\lambda_k(\widetilde{\boldsymbol{\Delta}}) \asymp n$$

because $\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0 \to \widetilde{\boldsymbol{\Delta}}$ as $n \to \infty$ and $\widetilde{\boldsymbol{\Delta}}$ is strictly positive definite.

A3. There exists constants $C, c > 0$ such that $\|\widetilde{\mathbf{E}}\|_2 \le C(n\rho_n)^{1/2}$ with probability at least $1 - n^{-c}$ for all $n \ge n_0(C, c)$. This is the result of Oliveira (2009) (also see Lemma B.1 of Tang and Priebe, 2018).

A4. There exists constants $C_{\widetilde{\mathbf{E}}}, \nu > 0, \xi > 1$ such that for all $1 \le t \le 2$, for each standard basis vector $\mathbf{e}_i$, and for each column vector $\widetilde{\mathbf{u}}_{0k}$ of $\widetilde{\mathbf{U}}_0$,

$$|\langle \mathbf{e}_i, \widetilde{\mathbf{E}}^t\widetilde{\mathbf{u}}_{0k}\rangle| \le (C_{\mathbf{E}}n\rho_n)^{t/2}(\log n)^{t\xi}\|\widetilde{\mathbf{u}}_{0k}\|_\infty$$

with probability at least $1 - \exp\{-\nu(\log n)^\xi\}$, provided that $n \ge n_0(C_{\widetilde{\mathbf{E}}}, \nu, \xi)$. We show that this condition holds with $\xi = 2$ by applying Lemma E.1. Following the proof of Lemma E.2 (in particular, the derivation of equation (E.2)), we obtain

$$\widetilde{\mathbf{E}} = \widetilde{\mathbf{E}}_1 + \widetilde{\mathbf{E}}_2,$$

where

$$\begin{aligned}
\widetilde{\mathbf{E}}_1 &= n\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2} + \frac{n}{2}\mathbf{T}^{-1/2}\mathbf{P}_0\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D}) + \frac{n}{2}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}_0\mathbf{T}^{-1/2},\\
\widetilde{\mathbf{E}}_2 &= \frac{n}{2}\mathbf{T}^{-1/2}(\mathbf{A} - \mathbf{P}_0)(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2}) + \frac{n}{2}(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})(\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\\
&\quad + \frac{n}{2}(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})(\mathbf{A} - \mathbf{P}_0)(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2}) + \frac{n}{2}\mathbf{T}^{-3/2}(\mathbf{T} - \mathbf{D})\mathbf{P}_0(\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2})\\
&\quad + \frac{n}{2}\widetilde{\mathbf{R}}^{(-1/2)},
\end{aligned}$$

and the remainder $\widetilde{\mathbf{R}}^{(-1/2)}$ satisfies $\max\{\|\widetilde{\mathbf{R}}^{(-1/2)}\|_2, \|\widetilde{\mathbf{R}}^{(-1/2)}\|_\infty\} \le C(\log n)/n$ with probability at least $1 - 2\exp\{-c(\log n)^2\}$ for some constants $C, c > 0$. Since Chernoff bound together with the matrix Bernstein's inequality imply that with probability at least $1 - \exp\{-c(\log n)^2\}$,

$$\begin{aligned}
&\|\mathbf{D} - \mathbf{T}\|_\infty = \|\mathbf{D} - \mathbf{T}\|_2 \le C\sqrt{n}\log n\\
&\|\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2}\|_\infty = \|\mathbf{D}^{-1/2} - \mathbf{T}^{-1/2}\|_2 \le Cn^{-1}\log n,\\
&\|\mathbf{A} - \mathbf{P}_0\|_2 \le C\sqrt{n}\log n
\end{aligned}$$

for some constants $C, c > 0$, then we immediately obtain that

$$\|\widetilde{\mathbf{E}}_1\|_2 \le 3C\sqrt{n}\log n$$

with probability at least $1 - \exp\{-c(\log n)^2\}$. Furthermore, one exploits the analysis of Appendix B.2 in Tang and Priebe (2018) and conclude that there exists constant $C_{\widetilde{\mathbf{E}}_2}, \nu > 0$, such that with probability at least $1 - \exp\{-\nu(\log n)^2\}$,

$$\|\widetilde{\mathbf{E}}_2\|_2 \le C_{\widetilde{\mathbf{E}}_2}(\log n)^2$$

for sufficiently large n.

Therefore, by Lemma E.1 with $t = 2$, with probability at least $1 - \exp\{-\nu(\log n)^2\}$ for some constant $\nu > 0$, we have

$$\begin{aligned}
|\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}^2\widetilde{\mathbf{u}}_{0k}| &\lesssim |\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}_1^2\widetilde{\mathbf{u}}_{0k}| + \|\widetilde{\mathbf{E}}_1\|_2\|\widetilde{\mathbf{E}}_2\|_2 + \|\widetilde{\mathbf{E}}_2\|_2^2\\
&\lesssim (\log n)^4 n\|\widetilde{\mathbf{u}}_{0k}\|_\infty + (\log n)^3\sqrt{n} + (\log n)^4\\
&\lesssim (\log n)^4 n\|\widetilde{\mathbf{u}}_{0k}\|_\infty
\end{aligned}$$

for all $i \in [n]$ and $k \in [d]$, where we have used the fact that $\|\widetilde{\mathbf{u}}_{0k}\|_\infty \ge n^{-1/2}$.

We now consider the case with $t = 1$. Then by Lemma E.1 again with $t = 1$, we obtain with probability at least $1 - \exp\{-\nu(\log n)^2\}$, for all $i \in [n]$ and $k \in [d]$,

$$\begin{aligned}
|\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}\widetilde{\mathbf{u}}_{0k}| &\le |\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}_1\widetilde{\mathbf{u}}_{0k}| + |\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}_2\widetilde{\mathbf{u}}_{0k}| \le |\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}_1\widetilde{\mathbf{u}}_{0k}| + \|\widetilde{\mathbf{E}}_2\|_2\\
&\le C_{\widetilde{\mathbf{E}}_1} n^{1/2}(\log n)^2\sqrt{n}\|\widetilde{\mathbf{u}}_{0k}\|_\infty + C_{\widetilde{\mathbf{E}}_2}(\log n)^2\\
&\le (C_{\widetilde{\mathbf{E}}_1} + C_{\widetilde{\mathbf{E}}_2})(\log n)^2\sqrt{n}\|\widetilde{\mathbf{u}}_{0k}\|_\infty.
\end{aligned}$$

Then we apply the union bound over $t = 1$ and $t = 2$ to conclude that there exists constants $\widetilde{C}_{\widetilde{\mathbf{E}}}, \nu > 0$, such that for sufficiently large n, with probability at least $1 - \exp\{-\nu(\log n)^2\}$,

$$|\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}^t\widetilde{\mathbf{u}}_{0k}| \le (C_{\widetilde{\mathbf{E}}})^t(\log n)^{2t}n^{t/2}\|\widetilde{\mathbf{u}}_{0k}\|_\infty$$

for all $i \in [n], k \in [d]$, and $t = 1, 2$.

A5. $\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0 \to \widetilde{\mathbf{\Delta}}$, and for each fixed $i \in [n]$,

$$\left[\frac{1}{\sqrt{n\rho_n}}\mathbf{e}_i^{\mathrm{T}}\widetilde{\mathbf{E}}(\sqrt{n}\mathbf{Y}_0)\right]^{\mathrm{T}} = \frac{1}{\sqrt{n\rho_n}}\sum_{j=1}^n \widetilde{E}_{ij}\sqrt{n}\mathbf{y}_{0j}$$

converges in distribution to a centered multivariate normal distribution. This is exactly the result of Lemma E.2. In particular, the covariance matrix is given by $\widetilde{\mathbf{\Delta}}\widetilde{\mathbf{\Sigma}}(\mathbf{x}_{0i})\widetilde{\mathbf{\Delta}}$.

A6. $(\log n)^{2\xi}/(n\rho_n) \to 0$ and

$$\rho_n^{-1/2}\max\{(\log n)^{2\xi}, \|\widetilde{\mathbf{U}}_0^{\mathrm{T}}\widetilde{\mathbf{E}}\widetilde{\mathbf{U}}_0\|_2 + 1\}\|\widetilde{\mathbf{U}}_0\|_{2\to\infty} = o_{\mathbb{P}_0}(1).$$

By (B.17) in Lemma B.4 of Tang and Priebe (2018), we see that $\|\widetilde{\mathbf{U}}_0^{\mathrm{T}}\widetilde{\mathbf{E}}\widetilde{\mathbf{U}}_0\|_2 = O_{\mathbb{P}_0}(n^{-1})$. With $\xi = 2$ and $\rho_n \equiv 1$, the above statement is equivalent to

$$\max\{(\log n)^{2\xi}, \|\widetilde{\mathbf{U}}_0^{\mathrm{T}}\widetilde{\mathbf{E}}\widetilde{\mathbf{U}}_0\|_2 + 1\}n^{-1/2} = o_{\mathbb{P}_0}(1),$$

which automatically holds.

Hence, by Theorem 3 in Cape et al. (2019), there exists orthogonal matrices $\mathbf{W}^*$, $\mathbf{W}_\mathbf{X}$ (possibly depending on n), such that for any fixed $i \in [n]$,

$$n\mathbf{W}_\mathbf{X}^\mathrm{T}\{\mathbf{W}^*(\widetilde{\mathbf{U}}_\mathbf{A}^\mathrm{T}\mathbf{e}_i) - (\widetilde{\mathbf{U}}_\mathbf{P}^\mathrm{T}\mathbf{e}_i)\} \overset{\mathcal{L}}{\to} \mathrm{N}(\mathbf{0}, \widetilde{\mathbf{\Delta}}^{-1/2}\widetilde{\mathbf{\Sigma}}(\mathbf{x}_{0i})\widetilde{\mathbf{\Delta}}^{-1/2}).$$

In particular, the matrix $\mathbf{W}_\mathbf{X}$ can be taken such that $\mathbf{Y}_0 = \widetilde{\mathbf{U}}_\mathbf{P}\widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\mathbf{W}_\mathbf{X}$, and $\mathbf{W}^*$ can be taken as the product of the left and right singular vector matrices of $\widetilde{\mathbf{U}}_\mathbf{P}^\mathrm{T}\widetilde{\mathbf{U}}_\mathbf{A}$. Then we proceed to compute

$$\begin{aligned}
n\{\mathbf{W}_\mathbf{X}^\mathrm{T}\mathbf{W}^*(\check{\mathbf{X}}^\mathrm{T}\mathbf{e}_i) - \mathbf{y}_{0i}\} &= n\{\mathbf{W}_\mathbf{X}^\mathrm{T}\mathbf{W}^*(\check{\mathbf{X}}^\mathrm{T}\mathbf{e}_i) - (\mathbf{Y}_0^\mathrm{T}\mathbf{e}_i)\} \\
&= n\{\mathbf{W}_\mathbf{X}^\mathrm{T}\mathbf{W}^*(\widetilde{\mathbf{S}}_\mathbf{A}^{1/2}\widetilde{\mathbf{U}}_\mathbf{A}^\mathrm{T}\mathbf{e}_i) - (\mathbf{W}_\mathbf{X}^\mathrm{T}\widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\widetilde{\mathbf{U}}_\mathbf{P}^\mathrm{T}\mathbf{e}_i)\} \\
&= n\mathbf{W}_\mathbf{X}^\mathrm{T}\{\mathbf{W}^*(\widetilde{\mathbf{S}}_\mathbf{A}^{1/2}\widetilde{\mathbf{U}}_\mathbf{A}^\mathrm{T}\mathbf{e}_i) - \widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\widetilde{\mathbf{U}}_\mathbf{P}^\mathrm{T}\mathbf{e}_i\} \\
&= n\mathbf{W}_\mathbf{X}^\mathrm{T}\{(\mathbf{W}^*\widetilde{\mathbf{S}}_\mathbf{A}^{1/2} - \widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\mathbf{W}^*)(\widetilde{\mathbf{U}}_\mathbf{A}^\mathrm{T}\mathbf{e}_i) + \widetilde{\mathbf{S}}_\mathbf{P}^{1/2}(\mathbf{W}^*\widetilde{\mathbf{U}}_\mathbf{A}^\mathrm{T}\mathbf{e}_i - \widetilde{\mathbf{U}}_\mathbf{P}^\mathrm{T}\mathbf{e}_i)\} \\
&= n\mathbf{W}_\mathbf{X}^\mathrm{T}\{(\mathbf{W}^*\widetilde{\mathbf{S}}_\mathbf{A}^{1/2} - \widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\mathbf{W}^*)(\widetilde{\mathbf{U}}_\mathbf{A}^\mathrm{T}\mathbf{e}_i)\} + n\mathbf{W}_\mathbf{X}^\mathrm{T}\widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\{\mathbf{W}^*(\mathbf{U}_\mathbf{A}^\mathrm{T}\mathbf{e}_i) - (\widetilde{\mathbf{U}}_\mathbf{P}^\mathrm{T}\mathbf{e}_i)\}.
\end{aligned}$$

By Proposition B.2 and Lemma B.3 Tang and Priebe (2018), we have

$$\begin{aligned}
\|\mathbf{W}^*\widetilde{\mathbf{S}}_\mathbf{A}^{1/2} - \widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\mathbf{W}^*\|_2 &\le \|(\mathbf{W}^* - \widetilde{\mathbf{U}}_\mathbf{P}^\mathrm{T}\widetilde{\mathbf{U}}_\mathbf{A})\widetilde{\mathbf{S}}_\mathbf{A}^{1/2}\|_2 + \|\widetilde{\mathbf{U}}_\mathbf{P}^\mathrm{T}\widetilde{\mathbf{U}}_\mathbf{A}\mathbf{S}_\mathbf{A}^{1/2} - \widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\widetilde{\mathbf{U}}_\mathbf{P}^\mathrm{T}\widetilde{\mathbf{U}}_\mathbf{A}\|_2 + \|\mathbf{S}_\mathbf{P}^{1/2}(\widetilde{\mathbf{U}}_\mathbf{P}^\mathrm{T}\widetilde{\mathbf{U}}_\mathbf{A} - \mathbf{W}^*)\|_2 \\
&\le \|\mathbf{W}^* - \widetilde{\mathbf{U}}_\mathbf{P}^\mathrm{T}\widetilde{\mathbf{U}}_\mathbf{A}\|_2(\|\widetilde{\mathbf{S}}_\mathbf{A}^{1/2}\|_2 + \|\widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\|_2) + O_{\mathbb{P}_0}(n^{-1}) \\
&= O_{\mathbb{P}_0}(n^{-1}).
\end{aligned}$$

By Theorem 1 of Cape et al. (2019), we have

$$\begin{aligned}
\|\widetilde{\mathbf{U}}_\mathbf{A}\|_{2\to\infty} &\le \|\widetilde{\mathbf{U}}_\mathbf{P}\|_{2\to\infty} + \|\widetilde{\mathbf{U}}_\mathbf{A} - \widetilde{\mathbf{U}}_\mathbf{P}\mathbf{W}^*\|_{2\to\infty} \\
&\le \|\widetilde{\mathbf{U}}_\mathbf{P}\|_{2\to\infty} + O_{\mathbb{P}_0}\left(\frac{1}{\sqrt{n}}\min\{\sqrt{d}(\log n)^2\|\mathbf{U}_0\|_{2\to\infty}, 1\}\right) \\
&\lesssim \frac{1}{\sqrt{n}} + O_{\mathbb{P}_0}\left(\frac{(\log n)^2}{n}\right) = O_{\mathbb{P}_0}\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
\|n\mathbf{W}_\mathbf{X}^\mathrm{T}\{(\mathbf{W}^*\widetilde{\mathbf{S}}_\mathbf{A}^{1/2} - \widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\mathbf{W}^*)(\widetilde{\mathbf{U}}_\mathbf{A}^\mathrm{T}\mathbf{e}_i)\}\|_2 &\le n\|\mathbf{W}^*\widetilde{\mathbf{S}}_\mathbf{A}^{1/2} - \widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\mathbf{W}^*\|_2\|\widetilde{\mathbf{U}}_\mathbf{A}^\mathrm{T}\mathbf{e}_i\|_2 \\
&\le n\|\mathbf{W}^*\widetilde{\mathbf{S}}_\mathbf{A}^{1/2} - \widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\mathbf{W}^*\|_2\|\widetilde{\mathbf{U}}_\mathbf{A}\|_{2\to\infty} \\
&= o_{\mathbb{P}_0}(1).
\end{aligned}$$

This further implies that

$$\begin{aligned}
n\{\mathbf{W}_\mathbf{X}^\mathrm{T}\mathbf{W}^*(\check{\mathbf{X}}^\mathrm{T}\mathbf{e}_i) - \mathbf{y}_{0i}\} &= n\mathbf{W}_\mathbf{X}^\mathrm{T}\widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\{\mathbf{W}^*(\mathbf{U}_\mathbf{A}^\mathrm{T}\mathbf{e}_i) - (\widetilde{\mathbf{U}}_\mathbf{P}^\mathrm{T}\mathbf{e}_i)\} + o_{\mathbb{P}_0}(1) \\
&= n\mathbf{W}_\mathbf{X}^\mathrm{T}\widetilde{\mathbf{S}}_\mathbf{P}^{1/2}\mathbf{W}_\mathbf{X}[\mathbf{W}_\mathbf{X}^\mathrm{T}\{\mathbf{W}^*(\mathbf{U}_\mathbf{A}^\mathrm{T}\mathbf{e}_i) - (\widetilde{\mathbf{U}}_\mathbf{P}^\mathrm{T}\mathbf{e}_i)\}] + o_{\mathbb{P}_0}(1).
\end{aligned}$$

Using the asymptotic normality

$$n\mathbf{W}_{\mathbf{X}}^{\mathrm{T}}\{\mathbf{W}^*(\mathbf{U}_{\mathbf{A}}^{\mathrm{T}}\mathbf{e}_i) - (\widetilde{\mathbf{U}}_{\mathbf{P}}^{\mathrm{T}}\mathbf{e}_i)\} \to \mathrm{N}(\mathbf{0}, \widetilde{\boldsymbol{\Delta}}^{-1/2}\widetilde{\boldsymbol{\Sigma}}(\mathbf{x}_{0i})\widetilde{\boldsymbol{\Delta}}^{-1/2})$$

and the fact that $\mathbf{W}_{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\mathbf{W}_{\mathbf{X}} \to \widetilde{\boldsymbol{\Delta}}^{1/2}$, we apply Slutsky's theorem to conclude that

$$n\{\mathbf{W}^{\mathrm{T}}(\breve{\mathbf{X}}^{\mathrm{T}}\mathbf{e}_i) - \mathbf{y}_{0i}\} \overset{\mathcal{L}}{\to} \mathrm{N}(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}(\mathbf{x}_{0i}))$$

with the choice of $\mathbf{W} = (\mathbf{W}_{\mathbf{X}}^{\mathrm{T}}\mathbf{W}^*)^{\mathrm{T}}$. □

E.5 Proof of Theorem 7

Proof of Theorem 7. We adopt the notations used in Section E. Following the decomposition E.1, we immediately see that

$$\begin{aligned}
\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}_0 &= \mathbf{D}^{1/2}\breve{\mathbf{X}}\mathbf{W} - \mathbf{T}^{1/2}\mathbf{Y}_0\\
&= (\mathbf{D}^{1/2} - \mathbf{T}^{1/2})\breve{\mathbf{X}}\mathbf{W} + \mathbf{T}^{1/2}(\breve{\mathbf{X}}\mathbf{W} - \mathbf{Y}_0)\\
&= (\mathbf{D}^{1/2} - \mathbf{T}^{1/2})\breve{\mathbf{X}}\mathbf{W} + (\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\mathbf{Y}_0(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1} + \frac{1}{2}\mathbf{T}^{-1/2}(\mathbf{T} - \mathbf{D})\mathbf{Y}_0 + \mathbf{T}^{1/2}\breve{\mathbf{R}},
\end{aligned}$$

where the remainder $\breve{\mathbf{R}}$ satisfies the following concentration property: For any constant $c > 0$, there exists a constant $C > 0$ such that $\|\breve{\mathbf{R}}\|_{\mathrm{F}} \le C(n\rho_n)^{-1}$ with probability at least $1 - n^{-c}$. Let t_i denote the ith diagonal element of $\mathbf{T}$ and d_i the ith diagonal element of $\mathbf{D}$. Clearly,

$$\left|\sqrt{d_i} - \sqrt{t_i} - \frac{d_i - t_i}{\sqrt{2}t_i}\right| = \frac{(d_i - t_i)^2}{2\sqrt{t_i}(\sqrt{d_i} + \sqrt{t_i})^2} \le \frac{(d_i - t_i)^2}{2t_i^{3/2}}.$$

Since $\min_{i\in[n]} t_i^{3/2} = O((n\rho_n)^{3/2})$, it follows from Hoeffding's inequality and the union bound that

$$\left\|\mathbf{D}^{1/2} - \mathbf{T}^{1/2} - \frac{1}{2}\mathbf{T}^{-1/2}(\mathbf{D} - \mathbf{T})\right\|_2 = O_{\mathbb{P}_0}\left\{\frac{(\log n)^2}{(n\rho_n)^{1/2}}\right\}.$$

Observe that $\widetilde{\boldsymbol{\Delta}}_n = \mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0$, that $\mathbf{Y}_0$ and $\breve{\mathbf{X}}$ have spectra bounded away from 0 and ∞, and that $\|\mathbf{T}\|_2 = O(n\rho_n)$. Therefore,

$$\begin{aligned}
\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}_0 &= (\mathbf{D}^{1/2} - \mathbf{T}^{1/2})\breve{\mathbf{X}}\mathbf{W} + (\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\mathbf{Y}_0(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1} + \frac{1}{2}\mathbf{T}^{-1/2}(\mathbf{T} - \mathbf{D})\mathbf{Y}_0 + \mathbf{T}^{1/2}\breve{\mathbf{R}}\\
&= \frac{1}{2}\mathbf{T}^{-1/2}(\mathbf{D} - \mathbf{T})\breve{\mathbf{X}}\mathbf{W} + O_{\mathbb{P}_0}\left\{\frac{(\log n)^2}{(n\rho_n)^{1/2}}\right\} + (\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\mathbf{Y}_0\boldsymbol{\Delta}_n^{-1}\\
&\quad + \frac{1}{2}\mathbf{T}^{-1/2}(\mathbf{T} - \mathbf{D})\mathbf{Y}_0 + O_{\mathbb{P}_0}\{(n\rho_n)^{-1/2}\}\\
&= (\mathbf{A} - \mathbf{P}_0)\mathbf{T}^{-1/2}\mathbf{Y}_0\boldsymbol{\Delta}_n^{-1} + \frac{1}{2}\mathbf{T}^{-1/2}(\mathbf{D} - \mathbf{T})(\breve{\mathbf{X}}\mathbf{W} - \mathbf{Y}_0) + O_{\mathbb{P}_0}\left\{\frac{(\log n)^2}{(n\rho_n)^{1/2}}\right\},
\end{aligned}$$

where the remainders $O_{\mathbb{P}_0}\{(n\rho_n)^{-1/2}(\log n)^2\}$ are with respect to $\|\cdot\|_{\mathrm{F}}$. Since we obtain from Appendix B.2 and Appendix B.3 in Tang and Priebe (2018) that

$$\|\mathbf{A}-\mathbf{P}_0\|_2 = O_{\mathbb{P}_0}\{(n\rho_n)^{1/2}\},\quad \|\mathbf{T}-\mathbf{D}\|_2 = O_{\mathbb{P}_0}\{(n\rho_n)^{1/2}\log n\},\quad \|\mathbf{Y}_0\|_{\mathrm{F}} = O(1),\quad \|\widetilde{\boldsymbol{\Delta}}_n^{-1}\|_{\mathrm{F}} = O(1),$$

we further write

$$\begin{aligned}
\|\mathbf{T}^{-1/2}(\mathbf{D}-\mathbf{T})(\breve{\mathbf{X}}\mathbf{W}-\mathbf{Y}_0)\|_{\mathrm{F}} &\leq \|\mathbf{T}^{-1/2}\|_2\|\mathbf{D}-\mathbf{T}\|_2\|\breve{\mathbf{X}}\mathbf{W}-\mathbf{Y}_0\|_{\mathrm{F}}\\
&\leq \|\mathbf{T}^{-1/2}\|_2\|\mathbf{D}-\mathbf{T}\|_2\|\mathbf{T}^{-1/2}\|_2\|\mathbf{A}-\mathbf{P}_0\|_2\|\mathbf{T}^{-1/2}\|_2\|\mathbf{Y}_0\|_{\mathrm{F}}\|\widetilde{\boldsymbol{\Delta}}_n^{-1}\|_2\\
&\quad + \frac{1}{2}\|\mathbf{T}^{-1/2}\|_2\|\mathbf{D}-\mathbf{T}\|_2\|\mathbf{T}^{-1}\|_2\|\mathbf{T}-\mathbf{D}\|_2\|\mathbf{Y}_0\|_{\mathrm{F}}\\
&\quad + \|\mathbf{T}^{-1/2}\|_2\|\mathbf{D}-\mathbf{T}\|_2\|\breve{\mathbf{R}}\|_{\mathrm{F}}\\
&= O_{\mathbb{P}_0}\left\{(n\rho_n)^{-1/2}(\log n)^2\right\}.
\end{aligned}$$

Therefore, we can write

$$\|\widetilde{\mathbf{X}}\mathbf{W}-\rho_n^{1/2}\mathbf{X}_0-(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}\mathbf{Y}_0(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1}\|_{\mathrm{F}} = O_{\mathbb{P}_0}\left\{(n\rho_n)^{-1/2}(\log n)^2\right\}.$$

Since for each fixed i,

$$\begin{aligned}
\mathbf{e}_i^{\mathrm{T}}(\mathbf{A}-\mathbf{P}_0)\mathbf{T}^{-1/2}\mathbf{Y}_0(\mathbf{Y}_0^{\mathrm{T}}\mathbf{Y}_0)^{-1} &= \sum_{j=1}^n\left(\frac{A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{\sqrt{n\rho_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}}\right)\mathbf{y}_{0j}^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}}_n^{-1} = \sum_{j=1}^n\left(\frac{A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}}{\sqrt{n\rho_n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}}\right)\frac{\mathbf{x}_{0j}^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}}_n^{-1}}{\sqrt{n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}}\\
&= \rho_n^{-1/2}\sum_{j=1}^n\left(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}\right)\frac{\mathbf{x}_{0j}^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}}_n^{-1}}{n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n},
\end{aligned}$$

the proof is completed by observing that

$$\sup_{i,j\in[n]}\left\|\frac{\widetilde{\boldsymbol{\Delta}}_n^{-1}\mathbf{x}_{0j}}{n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n}\right\|_2 \leq \sup_{i,j\in[n]}\frac{\|\widetilde{\boldsymbol{\Delta}}_n^{-1}\|_2\|\mathbf{x}_{0j}\|_2}{n\mathbf{x}_{0j}^{\mathrm{T}}\boldsymbol{\mu}_n} \lesssim \frac{1}{n}.$$

□

F Proofs of Theorems 8 and 9

F.1 Proof of Theorem 8

Proof of Theorem 8. Let $(\mathbf{W})_{n=1}^\infty = (\mathbf{W}_n)_{n=1}^\infty \subset \mathbb{O}(d)$ be the sequence of orthogonal matrices satisfying (3.3). Define a function $\mathbf{h}:\mathbb{R}^n\times\mathbb{R}^{n\times d}\to\mathbb{R}^d$ by

$$\mathbf{h}(\mathbf{x},\mathbf{Z}) = [h_1(\mathbf{x},\mathbf{Z}),\ldots,h_d(\mathbf{x},\mathbf{Z})]^{\mathrm{T}} = \frac{\mathbf{x}}{\sqrt{(1/n)\sum_{j=1}^n\mathbf{x}^{\mathrm{T}}\mathbf{z}_j}},\quad \text{where } \mathbf{Z} = [\mathbf{z}_1,\ldots,\mathbf{z}_n]^{\mathrm{T}}\in\mathbb{R}^{n\times d}.$$

Simple algebra shows that for $k = 1, \ldots, d$

$$
\begin{aligned}
\frac{\partial h_k}{\partial \mathbf{x}^{\mathrm{T}}}(\rho_n^{1/2}\mathbf{x}_{0i}, \rho_n^{1/2}\mathbf{X}_0) &= \rho_n^{-1/2}\left(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\right)^{-3/2}\left(\frac{1}{n}\sum_{j=1}^n \mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}\mathbf{e}_k^{\mathrm{T}} - \frac{1}{2n}\sum_{j=1}^n \mathbf{e}_k^{\mathrm{T}}\mathbf{x}_{0i}\mathbf{x}_{0j}^{\mathrm{T}}\right),\\
\frac{\partial h_k}{\partial \mathbf{z}_j^{\mathrm{T}}}(\rho_n^{1/2}\mathbf{x}_{0i}, \rho_n^{1/2}\mathbf{X}_0) &= -\frac{1}{2n\sqrt{\rho_n}}\left(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\right)^{-3/2}\mathbf{e}_k^{\mathrm{T}}\mathbf{x}_{0i}\mathbf{x}_{0i}^{\mathrm{T}},\\
\frac{\partial^2 h_k}{\partial \mathbf{x}\partial \mathbf{x}^{\mathrm{T}}}(\rho_n^{1/2}\mathbf{x}_{0i}, \rho_n^{1/2}\mathbf{X}_0) &= \rho_n^{-1}(\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i})^{-3/2}\left\{-\frac{1}{2}\mathbf{e}_k\boldsymbol{\mu}_n^{\mathrm{T}} - \frac{1}{2}\boldsymbol{\mu}_n\mathbf{e}_k^{\mathrm{T}} + \frac{3}{4}(\mathbf{e}_k^{\mathrm{T}}\mathbf{x}_{0i})(\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i})^{-1}\boldsymbol{\mu}_n\boldsymbol{\mu}_n^{\mathrm{T}}\right\},\\
\frac{\partial^2 h_k}{\partial \mathbf{x}\partial \mathbf{z}_j^{\mathrm{T}}}(\rho_n^{1/2}\mathbf{x}_{0i}, \rho_n^{1/2}\mathbf{X}_0) &= -\frac{1}{2n\rho_n}(\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i})^{-3/2}\left\{\mathbf{e}_k\mathbf{x}_{0i}^{\mathrm{T}} + (\mathbf{e}_k^{\mathrm{T}}\mathbf{x}_{0i})\mathbf{I} - \frac{3}{2n}(\mathbf{e}_k^{\mathrm{T}}\mathbf{x}_{0i})(\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i})^{-1}\mathbf{x}_{0i}\mathbf{x}_{0i}^{\mathrm{T}}\right\},\\
\frac{\partial^2 h_k}{\partial \mathbf{z}_l\partial \mathbf{z}_j^{\mathrm{T}}}(\rho_n^{1/2}\mathbf{x}_{0i}, \rho_n^{1/2}\mathbf{X}_0) &= \frac{3}{4n^2\rho_n}(\mathbf{e}_k^{\mathrm{T}}\mathbf{x}_{0i})(\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i})^{-5/2}\mathbf{x}_{0i}\mathbf{x}_{0i}^{\mathrm{T}}.
\end{aligned}
$$

Note that

$$
\sup_{j\in[n]}\left\|\frac{\partial^2 h_k}{\partial \mathbf{x}\partial \mathbf{z}_j^{\mathrm{T}}}(\rho_n^{1/2}\mathbf{x}_{0i}, \rho_n^{1/2}\mathbf{X}_0)\right\|_{\mathrm{F}} = O\left(\frac{1}{n\rho_n}\right),\quad \sup_{j,l\in[n]}\left\|\frac{\partial^2 h_k}{\partial \mathbf{z}_l\partial \mathbf{z}_j^{\mathrm{T}}}(\rho_n^{1/2}\mathbf{x}_{0i}, \rho_n^{1/2}\mathbf{X}_0)\right\|_{\mathrm{F}} = O\left(\frac{1}{n^2\rho_n}\right).
$$

It follows from Taylor's expansion that

$$
\begin{aligned}
\mathbf{h}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i, \widetilde{\mathbf{X}}\mathbf{W}) &= \mathbf{h}(\rho_n^{1/2}\mathbf{x}_{0i}, \rho_n^{1/2}\mathbf{X}_0) + \frac{\partial \mathbf{h}}{\partial \mathbf{x}^{\mathrm{T}}}(\rho_n^{1/2}\mathbf{x}_{0i}, \rho_n^{1/2}\mathbf{X}_0)(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_{0i})\\
&\quad + \sum_{j=1}^n \frac{\partial \mathbf{h}}{\partial \mathbf{z}_j^{\mathrm{T}}}(\rho_n^{1/2}\mathbf{x}_{0i}, \rho_n^{1/2}\mathbf{X}_0)(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j - \rho_n^{1/2}\mathbf{x}_{0j}) + \mathbf{R}_{\mathbf{x}_i} + \sum_{j=1}^n \mathbf{R}_{\mathbf{x}_i\mathbf{z}_j} + \sum_{j=1}^n\sum_{l=1}^n \mathbf{R}_{\mathbf{z}_j\mathbf{z}_l},
\end{aligned}
$$

where

$$
\max_{i\in[n]}\|\mathbf{R}_{\mathbf{x}_i}\| \lesssim \frac{(\log n)^{1\vee\omega}}{\rho_n^2 n},\quad \sup_{i,j\in[n]}\|\mathbf{R}_{\mathbf{x}_i\mathbf{z}_j}\| \lesssim \frac{(\log n)^{1\vee\omega}}{n^2\rho_n^2},\quad \sup_{i,j,l\in[n]}\|\mathbf{R}_{\mathbf{z}_l\mathbf{z}_j}\| \lesssim \frac{(\log n)^{1\vee\omega}}{n^3\rho_n^2}
$$

provided that

$$
\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}_0\|_{2\to\infty} \leq C_c\frac{(\log n)^{(1\vee\omega)/2}}{\sqrt{n\rho_n}},\quad \|\widehat{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}_0\|_{2\to\infty} \leq C_c\frac{(\log n)^{(1\vee\omega)/2}}{\sqrt{n\rho_n}}
$$

for some constant $C_c > 0$. Note that by Theorem 4, we have

$$
\|\widehat{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}_0\|_{2\to\infty} \leq \sum_{k=1}^d \max_{i\in[n]}\left|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n \frac{(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})[\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{x}_{0j}]_k}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\right| + O_{\mathbb{P}_0}\left(\frac{(\log n)^{1\vee\omega}}{n\rho_n^{5/2}}\right).
$$

By Hoeffding's inequality and the union bound, we see that

$$
\max_{i\in[n]}\left|\frac{1}{n\sqrt{\rho_n}}\sum_{j=1}^n \frac{(A_{ij} - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})[\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{x}_{0j}]_k}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1 - \rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}\right| = O_{\mathbb{P}_0}\left(\sqrt{\frac{\log n}{n\rho_n}}\right).
$$

Thus, we conclude that $\|\widehat{\mathbf{X}}\mathbf{W}-\rho_n^{1/2}\mathbf{X}_0\|_{2\to\infty}=O_{\mathbb{P}_0}((n\rho_n)^{-1/2}(\log n)^{(1\vee\omega)/2})$. Invoking this fact and Lemma D.1, we see that

$$
\begin{aligned}
\sqrt{n}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{y}}_i-\mathbf{y}_{0i}) &= \mathbf{h}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i,\widetilde{\mathbf{X}}\mathbf{W})-\mathbf{h}(\rho_n^{1/2}\mathbf{x}_{0i},\rho_n^{1/2}\mathbf{X}_0)\\
&= \rho_n^{-1/2}\left(\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i}\right)^{-3/2}\left\{\frac{1}{n}\sum_{j=1}^n\left(\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}\mathbf{I}_d-\frac{1}{2}\mathbf{x}_{0i}\mathbf{x}_{0j}^{\mathrm{T}}\right)\right\}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_{0i})\\
&\quad+\mathbf{R}_{i1}^{(\mathrm{L})}+\mathbf{R}_{i2}^{(\mathrm{L})},
\end{aligned}
$$

where

$$
\mathbf{R}_{i1}^{(\mathrm{L})}=\sum_{j=1}^n\boldsymbol{\Xi}_{ij}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\rho_n^{1/2}\mathbf{x}_{0j}),\quad \boldsymbol{\Xi}_{ij}=[\boldsymbol{\xi}_{ij1},\ldots,\boldsymbol{\xi}_{ijd}]^{\mathrm{T}}=-\frac{1}{2n\sqrt{\rho_n}}\left(\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n\right)^{-3/2}\mathbf{x}_{0i}\mathbf{x}_{0i}^{\mathrm{T}},
$$

and

$$
\max_{i\in[n]}\|\mathbf{R}_{i2}^{(\mathrm{L})}\|\leq\max_{i\in[n]}\|\mathbf{R}_{\mathbf{x}_i}\|+n\max_{j\in[n]}\|\mathbf{R}_{\mathbf{x}_i\mathbf{z}_j}\|+n^2\max_{j,l}\|\mathbf{R}_{\mathbf{z}_j\mathbf{z}_l}\|=O_{\mathbb{P}_0}\left(\frac{(\log n)^{1\vee\omega}}{n\rho_n^2}\right).
$$

By an argument that is similar to the proof of Lemma D.2, we see that

$$
\max_{i\in[n]}\|\mathbf{R}_{i1}^{(\mathrm{L})}\|\lesssim\sum_{k=1}^d\max_{i\in[n]}\left|\sum_{j=1}^n\boldsymbol{\xi}_{ijk}^{\mathrm{T}}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_j-\rho_n^{1/2}\mathbf{x}_{0j})\right|=O_{\mathbb{P}_0}\left(\frac{(\log n)^{(1\vee\omega)/2}}{n\rho_n}\right).
$$

Hence we conclude that

$$
\begin{aligned}
\sqrt{n}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{y}}_i-\mathbf{y}_{0i}) &= \rho_n^{-1/2}\left(\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i}\right)^{-3/2}\left\{\frac{1}{n}\sum_{j=1}^n\left(\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}\mathbf{I}_d-\frac{1}{2}\mathbf{x}_{0i}\mathbf{x}_{0j}^{\mathrm{T}}\right)\right\}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_{0i})+\mathbf{R}_i^{(\mathrm{L})}\\
&= \rho_n^{-1/2}\frac{1}{\sqrt{\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i}}}\left(\mathbf{I}_d-\frac{\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i}}\right)(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_{0i})+\mathbf{R}_i^{(\mathrm{L})},
\end{aligned}
$$

where $\max_{i\in[n]}\|\mathbf{R}_i^{(\mathrm{L})}\|=O_{\mathbb{P}_0}((n\rho_n^2)^{-1}(\log n)^{1\vee\omega})$. This further implies that

$$
\sum_{i=1}^n\|\mathbf{R}_i^{(\mathrm{L})}\|^2=O_{\mathbb{P}_0}\left((n\rho_n^4)^{-1}(\log n)^{2(1\vee\omega)}\right).
$$

The proof is thus completed. □

F.2 Proof of Theorem 9

Proof of Theorem 9. Let $(\mathbf{W})_{n=1}^\infty=(\mathbf{W}_n)_{n=1}^\infty\subset\mathbb{O}(d)$ be the sequence of orthogonal matrices satisfying (3.3). Denote

$$
\boldsymbol{\gamma}_{ij}=\frac{1}{n\sqrt{\rho_n}}(\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i})^{-1/2}\left(\mathbf{I}_d-\frac{\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}}{\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i}}\right)\frac{\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\mathbf{x}_{0j}}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})}.
$$

First note that $\|\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\|_2 \leq \|\mathbf{\Delta}^{-1}\|_2$ for sufficiently large n, and hence,

$$\sup_{i,j\in[n]}\|\boldsymbol{\gamma}_{ij}\| \leq \frac{1}{n\sqrt{\rho_n}}\delta^{-1/2}\left(1+\frac{1}{\delta}\right)\sup_{i,j\in[n]}\frac{\|\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\|\|\mathbf{x}_{0j}\|}{\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})} \lesssim \frac{1}{n\sqrt{\rho_n}}. \tag{F.1}$$

Also observe that

$$\begin{aligned}
\mathbb{E}_0\left(\sum_{i=1}^n\left\|\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij}\right\|^2\right) &= \sum_{i=1}^n\sum_{a=1}^n\sum_{b=1}^n\mathbb{E}_0\left\{(A_{ia}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0a})(A_{ib}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0b})\boldsymbol{\gamma}_{ia}^{\mathrm{T}}\boldsymbol{\gamma}_{ib}\right\}\\
&= \frac{1}{n}\sum_{i=1}^n\mathrm{tr}\left[\frac{1}{(\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i})}\left(\mathbf{I}_d-\frac{\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}}{\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i}}\right)\mathbf{G}_n(\mathbf{x}_{0i})^{-1}\left(\mathbf{I}_d-\frac{\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}}{\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i}}\right)^{\mathrm{T}}\right]\\
&= \frac{1}{n}\sum_{i=1}^n\mathrm{tr}\{\widetilde{\mathbf{G}}_n(\mathbf{x}_{0i})\}.
\end{aligned}$$

Denote

$$\widehat{\mathbf{R}}_i^{(\mathrm{L})} = (\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i})^{-1/2}\left(\mathbf{I}_d-\frac{\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\mathbf{x}_{0i}^{\mathrm{T}}\boldsymbol{\mu}_n}\right)\widehat{\mathbf{R}}_i+\rho_n^{1/2}\mathbf{R}_i^{(\mathrm{L})}.$$

Clearly, $\sum_{i=1}^n\|\widehat{\mathbf{R}}_i^{(\mathrm{L})}\|^2 = O_{\mathbb{P}_0}((n\rho_n^5)^{-1}(\log n)^2)$. By Theorem 8 and Lemma D.5, we can write

$$\begin{aligned}
&n\rho_n\left\|\widehat{\mathbf{Y}}\mathbf{W}-\mathbf{Y}_0\right\|_{\mathrm{F}}^2\\
&\quad= \sum_{i=1}^n\left\|\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij}\right\|^2+2\sum_{i=1}^n(\widehat{\mathbf{R}}_i^{(\mathrm{L})})^{\mathrm{T}}\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij}+\sum_{i=1}^n\|\widehat{\mathbf{R}}_i^{(\mathrm{L})}\|^2\\
&\quad= \frac{1}{n}\sum_{i=1}^n\mathrm{tr}\{\widetilde{\mathbf{G}}_n(\mathbf{x}_{0i})\}+2\sum_{i=1}^n(\widehat{\mathbf{R}}_i^{(\mathrm{L})})^{\mathrm{T}}\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij}+o_{\mathbb{P}_0}(1)+O_{\mathbb{P}_0}\left(\frac{(\log n)^{2(1\vee\omega)}}{n\rho_n^5}\right)\\
&\quad= \frac{1}{n}\sum_{i=1}^n\mathrm{tr}\{\widetilde{\mathbf{G}}_n(\mathbf{x}_{0i})\}+2\sum_{i=1}^n(\widehat{\mathbf{R}}_i^{(\mathrm{L})})^{\mathrm{T}}\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij}+o_{\mathbb{P}_0}(1).
\end{aligned}$$

By Cauchy-Schwarz inequality and Lemma D.5,

$$\begin{aligned}
\left|\sum_{i=1}^n(\widehat{\mathbf{R}}_i^{(\mathrm{L})})^{\mathrm{T}}\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij}\right| &\leq \sum_{i=1}^n\|\widehat{\mathbf{R}}_i^{(\mathrm{L})}\|\left\|\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\gamma_{ijk}\right\|\\
&\leq \left(\sum_{i=1}^n\|\widehat{\mathbf{R}}_i^{(\mathrm{L})}\|^2\right)^{1/2}\left\{\sum_{i=1}^n\left\|\sum_{j=1}^n(A_{ij}-\rho_n\mathbf{x}_{0i}^{\mathrm{T}}\mathbf{x}_{0j})\boldsymbol{\gamma}_{ij}\right\|^2\right\}^{1/2}\\
&= o_{\mathbb{P}_0}(1).
\end{aligned}$$

Furthermore, by condition (2.1) and Lemma D.7, we see that

$$\frac{1}{n}\sum_{i=1}^n\mathrm{tr}\{\widetilde{\mathbf{G}}_n(\mathbf{x}_{0i})\} = \int\mathrm{tr}\{\widetilde{\mathbf{G}}_n(\mathbf{x})\}F_n(\mathrm{d}\mathbf{x}) \to \int\mathrm{tr}\{\widetilde{\mathbf{G}}(\mathbf{x})\}F(\mathrm{d}\mathbf{x}).$$

This completes the proof of the first part of the theorem. For the second part, we see that

$$\frac{1}{\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i}}\left(\mathbf{I}_d-\frac{\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i}}\right)\mathbf{G}(\mathbf{x}_{0i})^{-1}\left(\mathbf{I}_d-\frac{\mathbf{x}_{0i}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0i}}\right)=\widetilde{\mathbf{G}}_n(\mathbf{x}_{0i})\to\widetilde{\mathbf{G}}(\mathbf{x}_{0i}).$$

The result directly follows from the asymptotic normality of $\sqrt{n}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_{0i})$.
We are now left with the assertion that $\widetilde{\boldsymbol{\Sigma}}(\mathbf{x})-\widetilde{\mathbf{G}}(\mathbf{x})$ is positive semidefinite. Denote

$$\boldsymbol{\Lambda}=\mathrm{diag}\{(\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{01})^{-1},\ldots,(\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0n})^{-1}\},\quad \widetilde{\boldsymbol{\Delta}}_n=\int_{\mathcal{X}}\left(\frac{\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}}{\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_1}\right)F_n(\mathrm{d}\mathbf{x}_1)=\frac{1}{n}\sum_{j=1}^n\frac{\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}}{\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}_{0j}}=\frac{1}{n}\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X}_0.$$

Clearly, $\widetilde{\boldsymbol{\Delta}}_n\to\widetilde{\boldsymbol{\Delta}}$ by (2.1). Suppose $\mathbf{X}_0$ yields the singular value decomposition $\mathbf{X}_0=\mathbf{U}_0\mathbf{S}_0^{1/2}\mathbf{V}_0^{\mathrm{T}}$ with $\mathbf{U}_0\in\mathbb{O}(n,d)$, $\mathbf{S}_0^{1/2}$ being diagonal, and $\mathbf{V}_0\in\mathbb{O}(d)$. By Corollary 2.1 in Pecaric et al. (1996), we have

$$(\mathbf{U}_0^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{U}_0)(\mathbf{U}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})\mathbf{U}_0)(\mathbf{U}_0^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{U}_0)\succeq\mathbf{U}_0^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{D}_n(\mathbf{x})\boldsymbol{\Lambda}\mathbf{U}_0,$$

implying that

$$\begin{aligned}
\widetilde{\boldsymbol{\Delta}}_n^{-1}\left(\frac{1}{n}\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{D}_n(\mathbf{x})\boldsymbol{\Lambda}\mathbf{X}_0\right)\widetilde{\boldsymbol{\Delta}}_n^{-1}&=n(\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X}_0)^{-1}\left(\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{D}_n(\mathbf{x})\boldsymbol{\Lambda}\mathbf{X}_0\right)(\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X}_0)^{-1}\\
&=n\mathbf{V}_0\mathbf{S}_0^{-1/2}(\mathbf{U}_0^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{U}_0)^{-1}(\mathbf{U}_0^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{D}_n(\mathbf{x})\boldsymbol{\Lambda}\mathbf{U}_0)(\mathbf{U}_0^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{U}_0)^{-1}\mathbf{S}_0^{-1/2}\mathbf{V}_0^{\mathrm{T}}\\
&\succeq n\mathbf{V}_0\mathbf{S}_0^{-1/2}(\mathbf{U}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})^{-1}\mathbf{U}_0)^{-1}\mathbf{S}_0^{-1/2}\mathbf{V}_0^{\mathrm{T}}\\
&=n(\mathbf{V}_0\mathbf{S}_0^{1/2}\mathbf{U}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})^{-1}\mathbf{U}_0\mathbf{S}_0^{1/2}\mathbf{V}_0^{\mathrm{T}})^{-1}\\
&=\left(\frac{1}{n}\mathbf{X}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})^{-1}\mathbf{X}_0\right)^{-1}.
\end{aligned}$$

Since $\widetilde{\boldsymbol{\Delta}}_n\boldsymbol{\mu}_n=\boldsymbol{\mu}_n$, it follows that

$$\begin{aligned}
\widetilde{\boldsymbol{\Sigma}}_n(\mathbf{x})&:=\frac{1}{\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}}\left(\widetilde{\boldsymbol{\Delta}}_n^{-1}-\frac{\mathbf{x}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}}\right)\left\{\frac{1}{n}\sum_{j=1}^n\frac{\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j}(1-\rho_n\mathbf{x}^{\mathrm{T}}\mathbf{x}_{0j})}{(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}_{0j})^2}\mathbf{x}_{0j}\mathbf{x}_{0j}^{\mathrm{T}}\right\}\left(\widetilde{\boldsymbol{\Delta}}_n^{-1}-\frac{\mathbf{x}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}}\right)^{\mathrm{T}}\\
&=\frac{1}{\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}}\left(\widetilde{\boldsymbol{\Delta}}_n^{-1}-\frac{\mathbf{x}\boldsymbol{\mu}_n^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}}_n^{-1}}{2\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}}\right)\left(\frac{1}{n}\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{D}_n(\mathbf{x})\boldsymbol{\Lambda}\mathbf{X}_0\right)\left(\widetilde{\boldsymbol{\Delta}}_n^{-1}-\frac{\mathbf{x}\boldsymbol{\mu}_n^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}}_n^{-1}}{2\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}}\right)^{\mathrm{T}}\\
&=\frac{1}{\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}}\left(\mathbf{I}_d-\frac{\mathbf{x}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}}\right)\widetilde{\boldsymbol{\Delta}}_n^{-1}\left(\frac{1}{n}\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{D}_n(\mathbf{x})\boldsymbol{\Lambda}\mathbf{X}_0\right)\widetilde{\boldsymbol{\Delta}}_n^{-1}\left(\mathbf{I}_d-\frac{\mathbf{x}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}}\right)^{\mathrm{T}}\\
&\succeq\frac{1}{\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}}\left(\mathbf{I}_d-\frac{\mathbf{x}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}}\right)\left(\frac{1}{n}\mathbf{X}_0^{\mathrm{T}}\mathbf{D}_n(\mathbf{x})^{-1}\mathbf{X}_0\right)^{-1}\left(\mathbf{I}_d-\frac{\mathbf{x}\boldsymbol{\mu}_n^{\mathrm{T}}}{2\boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{x}}\right)^{\mathrm{T}}\to\widetilde{\mathbf{G}}(\mathbf{x})
\end{aligned}$$

as $n\to\infty$. Therefore, $\widetilde{\boldsymbol{\Sigma}}(\mathbf{x})=\lim_{n\to\infty}\widetilde{\boldsymbol{\Sigma}}_n(\mathbf{x})\succeq\widetilde{\mathbf{G}}(\mathbf{x})$. The proof is thus completed. □

G Positive Definite Stochastic Block Models

In this section, we show that the asymptotic covariance matrix of the one-step estimator (3.2) and the ASE, under the conditions of Theorem 4, are identical, when the underlying random dot product graph coincides with a stochastic block model with a positive definite block probability matrix. This is established in Theorem G.1 below.

Theorem G.1 *Let* $\mathbf{A} \sim \mathrm{RDPG}(\mathbf{X}_0)$ *with a sparsity factor* ρ_n *for some* $\mathbf{X}_0 = [\mathbf{x}_{01}, \ldots, \mathbf{x}_{0n}]^{\mathrm{T}} \in \mathcal{X}^n$. *Assume that the conditions of Theorem 4 hold, and denote* $\rho = \lim_{n\to\infty} \rho_n$. *Further assume that there exist* d *linearly independent vectors* $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_d \in \mathcal{X}$ *and a probability vector* $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_d]$, *such that for each* $i \in [n]$, $\mathbf{x}_{0i} = \boldsymbol{\nu}_k$ *for some* $k \in [d]$ *with*

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(\mathbf{x}_{0i} = \boldsymbol{\nu}_k) = \pi_k, \quad k \in [d].$$

Let τ *be the cluster assignment function* $\tau : [n] \to [d]$ *such that* $\tau(i) = k$ *if and only if* $\mathbf{x}_{0i} = \boldsymbol{\nu}_k$. *Denote* $\widehat{\mathbf{X}} = [\widehat{\mathbf{x}}_1, \ldots, \widehat{\mathbf{x}}_n]^{\mathrm{T}}$ *the one-step estimator (3.2) based on an initial estimator* $\widetilde{\mathbf{X}}$ *that satisfies the approximate linearization property. Then there exists a sequence of orthogonal matrices* $(\mathbf{W})_{n=1}^{\infty} = (\mathbf{W}_n)_{n=1}^{\infty} \subset \mathbb{O}(d)$, *such that for each fixed* $i \in [n]$,

$$\sqrt{n}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i - \boldsymbol{\nu}_k) \xrightarrow{\mathcal{L}} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\nu}_k))$$

if $\tau(i) = k$, *where* $\boldsymbol{\Sigma}(\boldsymbol{\nu}_k)$ *is the same as the asymptotic covariance matrix of the* i*th row of the ASE, and can be computed using the formula in Theorem 1:*

$$\boldsymbol{\Sigma}(\boldsymbol{\nu}_k) = \boldsymbol{\Delta}^{-1}\left[\sum_{l=1}^{d} \pi_l\{\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_l(1-\rho\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_l)\}\boldsymbol{\nu}_l\boldsymbol{\nu}_l^{\mathrm{T}}\right]\boldsymbol{\Delta}^{-1}, \quad \textit{and} \quad \boldsymbol{\Delta} = \sum_{l=1}^{d}\pi_l\boldsymbol{\nu}_l\boldsymbol{\nu}_l^{\mathrm{T}}.$$

In other words, the asymptotic covariance matrix of the one-step estimator is the same as that of the ASE.

Proof. According to Theorem 5 in the manuscript, we have

$$\sqrt{n}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i - \boldsymbol{\nu}_k) \xrightarrow{\mathcal{L}} \mathrm{N}(\mathbf{0}, \mathbf{G}(\boldsymbol{\nu}_k)^{-1})$$

if $\mathbf{x}_{0i}$ matches with $\boldsymbol{\nu}_k$, where

$$\mathbf{G}(\boldsymbol{\nu}_k) = \sum_{l=1}^{d}\frac{\pi_l\boldsymbol{\nu}_l\boldsymbol{\nu}_l^{\mathrm{T}}}{\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_l(1-\rho\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_l)}.$$

It is therefore sufficient to show that $\mathbf{G}(\boldsymbol{\nu}_k)^{-1} = \boldsymbol{\Sigma}(\boldsymbol{\nu}_k)$ for all $k \in [K]$. Since $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_d$ are linearly independent, it follows that $\mathbf{N}_0 = [\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_d]^{\mathrm{T}}$ is invertible. Thus, we can write

$$\boldsymbol{\Delta} = \sum_{l=1}^{d}\pi_l\boldsymbol{\nu}_l\boldsymbol{\nu}_l^{\mathrm{T}} = \mathbf{N}_0^{\mathrm{T}}\mathrm{diag}(\boldsymbol{\pi})\mathbf{N}_0, \quad \boldsymbol{\Delta}^{-1} = \mathbf{N}_0^{-1}\mathrm{diag}(\boldsymbol{\pi})^{-1}\mathbf{N}_0^{-\mathrm{T}},$$

where $\mathbf{N}_0^{-\mathrm{T}}$ is the shorthand notation for $(\mathbf{N}_0^{-1})^{\mathrm{T}} = (\mathbf{N}_0^{\mathrm{T}})^{-1}$. Therefore,

$$\begin{aligned}
\mathbf{G}(\boldsymbol{\nu}_k)^{-1} &= \left[\sum_{l=1}^{d} \frac{\pi_l \boldsymbol{\nu}_l \boldsymbol{\nu}_l^{\mathrm{T}}}{\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_l(1-\rho\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_l)}\right]^{-1} = \left[\mathbf{N}_0^{\mathrm{T}}\mathrm{diag}\left\{\frac{\pi_1}{\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_1(1-\rho\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_1)},\ldots,\frac{\pi_d}{\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_d(1-\rho\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_d)}\right\}\mathbf{N}_0\right]^{-1}\\
&= \mathbf{N}_0^{-1}\mathrm{diag}\left\{\frac{\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_1(1-\rho\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_1)}{\pi_1},\ldots,\frac{\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_d(1-\rho\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_d)}{\pi_d}\right\}\mathbf{N}_0^{-\mathrm{T}}\\
&= \boldsymbol{\Delta}^{-1}\mathbf{N}_0^{\mathrm{T}}\mathrm{diag}(\boldsymbol{\pi})\mathrm{diag}\left\{\frac{\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_1(1-\rho\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_1)}{\pi_1},\ldots,\frac{\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_d(1-\rho\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_d)}{\pi_d}\right\}\mathrm{diag}(\boldsymbol{\pi})\mathbf{N}_0\boldsymbol{\Delta}^{-1}\\
&= \boldsymbol{\Delta}^{-1}\mathbf{N}_0^{\mathrm{T}}\mathrm{diag}\left\{\pi_1\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_1(1-\rho\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_1),\ldots,\pi_d\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_d(1-\rho\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_d)\right\}\mathbf{N}_0\boldsymbol{\Delta}^{-1}\\
&= \boldsymbol{\Delta}^{-1}\left[\sum_{l=1}^{d}\pi_l\{\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_l(1-\rho\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_l)\}\boldsymbol{\nu}_l\boldsymbol{\nu}_l^{\mathrm{T}}\right]\boldsymbol{\Delta}^{-1} = \boldsymbol{\Sigma}(\boldsymbol{\nu}_k).
\end{aligned}$$

The proof is thus completed. □

H Further Discussion of Sparse Graphs

This section of the Supplementary Material provides further discussion on the decaying rate of the sparsity factor ρ_n. The discussion is motivated by the comparison of different conditions of the average degrees in the random graph model for different estimators. In Theorem 4 and Theorem 8, which establish the asymptotic characterizations of the one-step estimators, we have imposed the assumption that the sparsity factor ρ_n for the random dot product graph is either constantly 1, or converges to 0 with the requirement that $n\rho_n$ grows at a polynomial rate of n. Here we use the polynomial rate of n to describe the case where $n\rho_n \gtrsim n^\eta$ for a strictly positive constant $\eta > 0$. In contrast, the limit theorem of the ASE (*i.e.*, Theorem 1) only requires that $n\rho_n$ grows at a polynomial rate of $\log n$. From the graph theory perspective, $n\rho_n$ controls the average expected degree of the random graph model, *i.e.*, the sparsity level of the graph. Therefore, the limit theorem of the ASE (Theorem 1) allows sparser graphs because the average expected degree grows at a polynomial rate of $\log n$, whereas the limit theorems of the one-step estimators (Theorems 4 and 8) require the graphs to be relatively denser as the average expected degree grows at a polynomial rate of n. The graphs in the latter scenario are only considered as moderately sparse.

This relatively stronger sparsity condition is a consequence of the proof technique employed here. The condition we propose may be weakened. Nevertheless, the proof strategy is a standard approach for establishing the asymptotic normality of the one-step estimators (see Section 5.7 of Van der Vaart, 2000), and the standard proof strategy inevitably leads to the requirement that $n\rho_n$ must be lower bounded by a polynomial of n. To illustrate this result, we consider the following simplified problem: Assume that the underlying random dot product graph is of dimension 1, *i.e.*, the latent position matrix $\mathbf{X}_0$ is an $n \times 1$ column vector, and suppose we focus on estimating a single latent position x_{0i} with the knowledge of the rest of the latent positions $(x_{0j})_{j\neq i}$. Formally, we take the initial estimator $\widetilde{\mathbf{X}}$ for the entire latent position matrix to be of the form

$$e_j^{\mathrm{T}}\widetilde{\mathbf{X}} = \begin{cases}\rho_n^{1/2}x_{0i}, & \text{if } j \neq i,\\ \mathbf{e}_i^{\mathrm{T}}\mathbf{X}^{(\mathrm{ASE})}, & \text{if } j = i.\end{cases}.$$

Here we choose $\widetilde{x}_i$ for x_{0i} as the ith row of the ASE, *i.e.*, $\widetilde{x}_i = \mathbf{e}_i^{\mathrm{T}}\mathbf{X}^{\mathrm{ASE}}$, but it is also possible to consider more general estimators. Denote $\widehat{x}_i^{(\mathrm{ASE})} = \mathbf{e}_i^{\mathrm{T}}\widehat{\mathbf{X}}^{(\mathrm{ASE})}$. Then the ith row of the one-step estimator (3.2) can be written as

$$\begin{aligned}\widehat{x}_i = \widehat{x}_i^{(\mathrm{ASE})} &+ \left\{\frac{1}{n}\sum_{j\neq i}\frac{\rho_n^{1/2}x_{0j}}{\widehat{x}_i^{(\mathrm{ASE})}(1-\rho_n^{1/2}\widehat{x}_i^{(\mathrm{ASE})}x_{0j})} + \frac{1}{n(1-(\widehat{x}_i^{(\mathrm{ASE})})^2)}\right\}^{-1}\\ &\times\left\{\frac{1}{n}\sum_{j\neq i}\frac{A_{ij}-\rho_n^{1/2}\widehat{x}_i^{(\mathrm{ASE})}x_{0j}}{\widehat{x}_i^{(\mathrm{ASE})}(1-\rho_n^{1/2}\widehat{x}_i^{(\mathrm{ASE})}x_{0j})} - \frac{\widehat{x}_i^{(\mathrm{ASE})}}{n(1-(\widehat{x}_i^{(\mathrm{ASE})})^2)}\right\}.\end{aligned}$$

Since $x_i^{(\mathrm{ASE})} = \rho_n^{1/2}x_{0i} + O_{\mathbb{P}_0}(n^{-1/2}) = O_{\mathbb{P}_0}(\rho_n^{1/2})$ by the asymptotic normality and the fact that $n\rho_n \geq 1$, it follows that $1-(x_i^{(\mathrm{ASE})})^2$ stays bounded away from 0 and 1 with probability going to 1, and hence,

$$\frac{1}{n(1-(\widehat{x}_i^{(\mathrm{ASE})})^2)} = O_{\mathbb{P}_0}(n^{-1}),\quad\text{and}\quad \frac{\widehat{x}_i^{(\mathrm{ASE})}}{n(1-(\widehat{x}_i^{(\mathrm{ASE})})^2)} = O_{\mathbb{P}_0}(n^{-1}).$$

It follows that

$$\begin{aligned}\widehat{x}_i &= \widehat{x}_i^{(\mathrm{ASE})} + \left[\left\{\frac{1}{n}\sum_{j\neq i}\frac{\rho_n^{1/2}x_{0j}}{\widehat{x}_i^{(\mathrm{ASE})}(1-\rho_n^{1/2}\widehat{x}_i^{(\mathrm{ASE})}x_{0j})}\right\}^{-1} + O_{\mathbb{P}_0}(n^{-1})\right]\\ &\quad\times\left\{\frac{1}{n}\sum_{j\neq i}\frac{(A_{ij}-\rho_n^{1/2}\widehat{x}_i^{(\mathrm{ASE})}x_{0j})}{\widehat{x}_i^{(\mathrm{ASE})}(1-\rho_n^{1/2}\widehat{x}_i^{(\mathrm{ASE})}x_{0j})} + O_{\mathbb{P}_0}(n^{-1})\right\}\\ &= \widehat{x}_i^{(\mathrm{ASE})} + \left\{\frac{1}{n}\sum_{j\neq i}\frac{\rho_n^{1/2}x_{0j}}{\widehat{x}_i^{(\mathrm{ASE})}(1-\rho_n^{1/2}\widehat{x}_i^{(\mathrm{ASE})}x_{0j})}\right\}^{-1}\left\{\frac{1}{n}\sum_{j\neq i}\frac{A_{ij}-\rho_n^{1/2}\widehat{x}_i^{(\mathrm{ASE})}x_{0j}}{\widehat{x}_i^{(\mathrm{ASE})}(1-\rho_n^{1/2}\widehat{x}_i^{(\mathrm{ASE})}x_{0j})}\right\} + O_{\mathbb{P}_0}(n^{-1}).\end{aligned}$$

Now we use the notation of Section 5.7 of Van der Vaart (2000). Denote

$$\Psi_n(x) = \frac{1}{n}\sum_{j\neq i}\frac{A_{ij}-\rho_n^{1/2}xx_{0j}}{x(1-\rho_n^{1/2}xx_{0j})}, \dot{\Psi}_{n,0} = -\frac{1}{n}\sum_{j\neq i}\frac{\rho_n^{1/2}x_{0j}}{\widehat{x}_i^{(\mathrm{ASE})}(1-\rho_n^{1/2}\widehat{x}_i^{(\mathrm{ASE})}x_{0j})}, \dot{\Psi}_0 = -\int_{\mathcal{X}}\frac{x_1}{1-\rho x_{0i}x_1}F(\mathrm{d}x_1).$$

Then the above equation can be written as

$$\widehat{x}_i = \widehat{x}_i^{(\mathrm{ASE})} - \dot{\Psi}_{n,0}^{-1}\Psi_n(\widehat{x}_i^{(\mathrm{ASE})}) + O_{\mathbb{P}_0}(n^{-1}),$$

and following the method in Van der Vaart (2000),

$$\begin{aligned}
\dot\Psi_{n,0}\sqrt{n}(\widehat{x}_i-\rho_n^{1/2}x_{0i}) &= \dot\Psi_{n,0}\sqrt{n}(\widehat{x}_i^{(\mathrm{ASE})}-\rho_n^{1/2}x_{0i}) - \sqrt{n}\Psi_n(\widehat{x}_i^{(\mathrm{ASE})}) + O_{\mathbb{P}_0}(n^{-1/2})\\
&= \dot\Psi_{n,0}\sqrt{n}(\widehat{x}_i^{(\mathrm{ASE})}-\rho_n^{1/2}x_{0i}) - \sqrt{n}\left\{\Psi_n(\widehat{x}_i^{(\mathrm{ASE})}) - \Psi_n(\rho_n^{1/2}x_{0i})\right\}\\
&\quad - \sqrt{n}\Psi_n(\rho_n^{1/2}x_{0i}) + O_{\mathbb{P}_0}(n^{-1/2})\\
&= (\dot\Psi_{n,0}-\dot\Psi_0)\sqrt{n}(\widehat{x}_i^{(\mathrm{ASE})}-\rho_n^{1/2}x_{0i})\\
&\quad + \dot\Psi_0\sqrt{n}(\widehat{x}_i^{(\mathrm{ASE})}-\rho_n^{1/2}x_{0i}) - \sqrt{n}\left\{\Psi_n(\widehat{x}_i^{(\mathrm{ASE})}) - \Psi_n(\rho_n^{1/2}x_{0i})\right\}\\
&\quad - \sqrt{n}\Psi_n(\rho_n^{1/2}x_{0i}) + O_{\mathbb{P}_0}(n^{-1/2})\\
&= (\dot\Psi_{n,0}-\dot\Psi_0)O_{\mathbb{P}_0}(1)\\
&\quad + \dot\Psi_0\sqrt{n}(\widehat{x}_i^{(\mathrm{ASE})}-\rho_n^{1/2}x_{0i}) - \sqrt{n}\left\{\Psi_n(\widehat{x}_i^{(\mathrm{ASE})}) - \Psi_n(\rho_n^{1/2}x_{0i})\right\}\\
&\quad - \sqrt{n}\Psi_n(\rho_n^{1/2}x_{0i}) + O_{\mathbb{P}_0}(n^{-1/2}).
\end{aligned}$$

To establish the asymptotic normality of $\sqrt{n}(\widehat{x}_i-\rho_n^{1/2}x_{0i})$, the two main standard ingredients employed in Section 5.7 of Van der Vaart (2000) are:

(a) $\dot\Psi_{n,0}\overset{\mathbb{P}_0}{\to}\dot\Psi_0$. This condition holds without requiring that $n\rho_n$ grows polynomially in n. Instead, we only need $n\rho_n\to\infty$. Similar to the proof of Theorem 3 in Section C.2, if we denote

$$\Gamma_j(x) = \frac{\rho_n^{1/2}x_{0j}}{x(1-\rho_n^{1/2}xx_{0j})},$$

then the derivative is

$$\Gamma_j'(x) = -\frac{\rho_n^{1/2}x_{0j}(1-2\rho_n^{1/2}xx_{0j})}{x^2(1-\rho_n^{1/2}xx_{0j})^2},$$

and this implies that there exists a sufficiently small $\epsilon>0$, such that for all $x\in\mathbb{R}$ with $\sqrt{n}|x-\rho_n^{1/2}x_{0i}|<\epsilon(n\rho_n)^{1/2}$,

$$\begin{aligned}
1-2\rho_n^{1/2}xx_{0j} &\le 1+2\rho_n^{1/2}|x-\rho_n^{1/2}x_{0i}|x_{0j}+2\rho_nx_{0i}x_{0j}\le 1+4\rho_n = O(1),\\
1-\rho_n^{1/2}xx_{0j} &\ge 1-\rho_n^{1/2}|x-\rho_n^{1/2}x_{0i}|x_{0j}-\rho_nx_{0i}x_{0j}\ge 1-\rho_n(\epsilon+x_{0i}x_{0j}) = O(1),\\
x^2 &\ge (x-\rho_n^{1/2}x_{0i}+\rho_n^{1/2}x_{0i})^2\ge(\rho_n^{1/2}x_{0i}-|x-\rho_n^{1/2}x_{0i}|)^2\\
&\ge(\rho_n^{1/2}x_{0i}-\rho_n^{1/2}\epsilon)^2 = \rho_n(x_{0i}-\epsilon)^2\gtrsim\rho_n,
\end{aligned}$$

and hence,

$$\max_{j\in[n]}\sup_{x:\sqrt{n}|x-\rho_n^{1/2}x_{0i}|<\epsilon(n\rho_n)^{1/2}}|\Gamma_j'(x)|\lesssim\rho_n^{-1/2}.$$

Therefore, by the mean-value theorem, over the event $\{\sqrt{n}|\widehat{x}_i^{(\mathrm{ASE})} - \rho_n^{1/2}x_{0i}| < \epsilon(n\rho_n)^{1/2}\}$,

$$
\begin{aligned}
|\dot{\Psi}_{n,0} - \dot{\Psi}_0| &\le \frac{1}{n}\sum_{j\ne i}|\Gamma_j(\widehat{x}_i^{(\mathrm{ASE})}) - \Gamma_j(\rho_n^{1/2}x_{0i})| + \left|\frac{1}{n}\sum_{j\ne i}\Gamma_j(\rho_n^{1/2}x_{0i}) - \dot{\Psi}_0\right| \\
&\le \max_{j\in[n]} \sup_{x:\sqrt{n}|x-\rho_n^{1/2}x_{0i}|<\epsilon(n\rho_n)^{1/2}} |\Gamma_j'(x)||\widehat{x}_i^{(\mathrm{ASE})} - \rho_n^{1/2}x_{0i}| + \left|\frac{1}{n}\sum_{j\ne i}\Gamma_j(\rho_n^{1/2}x_{0i}) - \dot{\Psi}_0\right| \\
&\le C\rho_n^{-1/2}|\widehat{x}_i^{(\mathrm{ASE})} - \rho_n^{1/2}x_{0i}| + o(1)
\end{aligned}
$$

for some constant $C > 0$. This is because the second term goes to 0 as $n \to \infty$ due condition (2.1) in the main text. Therefore, for any $t > 0$ and sufficiently large n,

$$
\begin{aligned}
\mathbb{P}_0\left(|\dot{\Psi}_{n,0} - \dot{\Psi}_0| > t\right) &= \mathbb{P}_0\left(|\dot{\Psi}_{n,0} - \dot{\Psi}_0| > t, \sqrt{n}|\widehat{x}_i^{(\mathrm{ASE})} - \rho_n^{1/2}x_{0i}| < \epsilon(n\rho_n)^{1/2}\right) \\
&\quad + \mathbb{P}_0\left(|\dot{\Psi}_{n,0} - \dot{\Psi}_0| > t, \sqrt{n}|\widehat{x}_i^{(\mathrm{ASE})} - \rho_n^{1/2}x_{0i}| \ge \epsilon(n\rho_n)^{1/2}\right) \\
&\le \mathbb{P}_0\left(C\rho_n^{-1/2}|\widehat{x}_i^{(\mathrm{ASE})} - \rho_n^{1/2}x_{0i}| > t/2, \sqrt{n}|\widehat{x}_i^{(\mathrm{ASE})} - \rho_n^{1/2}x_{0i}| < \epsilon(n\rho_n)^{1/2}\right) \\
&\quad + \mathbb{P}_0(\sqrt{n}|\widehat{x}_i^{(\mathrm{ASE})} - \rho_n^{1/2}x_{0i}| \ge \epsilon(n\rho_n)^{1/2}) \\
&\le \mathbb{P}_0\left(\sqrt{n}|\widehat{x}_i^{(\mathrm{ASE})} - \rho_n^{1/2}x_{0i}| > \frac{(n\rho_n)^{1/2}t}{2C}\right) \\
&\quad + \mathbb{P}_0(\sqrt{n}|\widehat{x}_i^{(\mathrm{ASE})} - \rho_n^{1/2}x_{0i}| \ge \epsilon(n\rho_n)^{1/2}) \to 0
\end{aligned}
$$

where we have used the fact $\sqrt{n}(\widehat{x}_i^{(\mathrm{ASE})} - \rho_n^{1/2}x_{0i}) = O_{\mathbb{P}_0}(1)$ and $n\rho_n \to \infty$. This completes the verification of $\dot{\Psi}_{n,0} \overset{\mathbb{P}_0}{\to} \dot{\Psi}_0$.

(b) $\sqrt{n}\{\Psi_n(\widehat{x}_i^{(\mathrm{ASE})}) - \Psi_n(\rho_n^{1/2}x_{0i})\} - \dot{\Psi}_0\sqrt{n}(\widehat{x}_i^{(\mathrm{ASE})} - \rho_n^{1/2}x_{0i}) = o_{\mathbb{P}_0}(1)$. We remark that this condition holds provided that $n\rho_n^3 \to 0$ as $n \to \infty$. Observe that

$$
\begin{aligned}
\Psi_n'(x) &= \frac{1}{n}\sum_{j\ne i}\frac{-\rho_n^{1/2}xx_{0j}(1-\rho_n^{1/2}xx_{0j}) - (A_{ij} - \rho_n^{1/2}xx_{0j})(1-2\rho_n^{1/2}xx_{0j})}{x^2(1-\rho_n^{1/2}xx_{0j})^2} \\
&= -\frac{1}{n}\sum_{j\ne i}\Gamma_j(x) - \frac{1}{n}\sum_{j\ne i}\frac{(A_{ij} - \rho_n^{1/2}xx_{0j})(1-2\rho_n^{1/2}xx_{0j})}{x^2(1-\rho_n^{1/2}xx_{0j})^2}
\end{aligned}
$$

and that

$$
\begin{aligned}
\dot{\Psi}_0 &= \left\{\dot{\Psi}_0 + \frac{1}{n}\sum_{j\ne i}\Gamma_j(\rho_n^{1/2}x_{0j})\right\} - \left\{\frac{1}{n}\sum_{j\ne i}\Gamma_j(\rho_n^{1/2}x_{0j}) + \Psi_n'(\rho_n^{1/2}x_{0i})\right\} + \Psi'(\rho_n^{1/2}x_{0i}) \\
&= o(1) - \frac{1}{n}\sum_{j\ne i}\frac{(A_{ij} - \rho_n x_{0i}x_{0j})(1-2\rho_n x_{0i}x_{0j})}{\rho_n x_{0i}^2(1-\rho_n x_{0i}x_{0j})^2} + \Psi_n'(\rho_n^{1/2}x_{0i}) \\
&= o_{\mathbb{P}_0}(1) + \Psi_n'(\rho_n^{1/2}x_{0i})
\end{aligned}
$$

by condition (2.1) and the fact that

$$\operatorname{var}\left\{\frac{1}{n}\sum_{j\neq i}\frac{(A_{ij}-\rho_n x_{0i}x_{0j})(1-2\rho_n x_{0i}x_{0j})}{\rho_n x_{0i}^2(1-\rho_n x_{0i}x_{0j})^2}\right\} = \frac{1}{n^2}\sum_{j\neq i}\frac{\rho_n x_{0i}x_{0j}(1-\rho_n x_{0i}x_{0j})(1-2\rho_n x_{0i}x_{0j})^2}{\rho_n^2 x_{0i}^4(1-\rho_n x_{0i}x_{0j})^4}$$
$$\lesssim \frac{1}{n\rho_n}\to 0.$$

Then we apply the fact that $\sqrt{n}(\widehat{x}_i^{(\mathrm{ASE})}-\rho_n^{1/2}x_{0i}) = O_{\mathbb{P}_0}(1)$ and the mean-value theorem to derive

$$\begin{aligned}
&\sqrt{n}\{\Psi_n(\widehat{x}_i^{(\mathrm{ASE})})-\Psi_n(\rho_n^{1/2}x_{0i})\} - \dot{\Psi}_0\sqrt{n}(\widehat{x}_i^{(\mathrm{ASE})}-\rho_n^{1/2}x_{0i})\\
&\quad= \sqrt{n}\{\Psi_n(\widehat{x}_i^{(\mathrm{ASE})})-\Psi_n(\rho_n^{1/2}x_{0i})\} - \Psi_n'(\rho_n^{1/2}x_{0i})\sqrt{n}(\widehat{x}_i^{(\mathrm{ASE})}-\rho_n^{1/2}x_{0i})\\
&\qquad - o_{\mathbb{P}_0}(1)\sqrt{n}(\widehat{x}_i^{(\mathrm{ASE})}-\rho_n^{1/2}x_{0i})\\
&\quad= \sqrt{n}\{\Psi_n(\widehat{x}_i^{(\mathrm{ASE})})-\Psi_n(\rho_n^{1/2}x_{0i})\} - \Psi_n'(\rho_n^{1/2}x_{0i})\sqrt{n}(\widehat{x}_i^{(\mathrm{ASE})}-\rho_n^{1/2}x_{0i}) + o_{\mathbb{P}_0}(1)\\
&\quad= \Psi_n''(\widetilde{x}_i)\sqrt{n}(\widehat{x}_i^{(\mathrm{ASE})}-\rho_n^{1/2}x_{0i})^2 + o_{\mathbb{P}_0}(1)\\
&\quad= \Psi_n''(\widetilde{x}_i)O_{\mathbb{P}_0}(n^{-1/2}) + o_{\mathbb{P}_0}(1),
\end{aligned}$$

where $\widetilde{x}_i$ lies between $\rho_n^{1/2}x_{0i}$ and $\widehat{x}_i^{(\mathrm{ASE})}$. Since the above derivation is equivalent, we see that ingredient (b) holds if and only if $\Psi''(\widetilde{x}_i) = o_{\mathbb{P}_0}(n^{1/2})$. Thus we compute

$$\begin{aligned}
\Psi_n''(\widetilde{x}_i) &= -\frac{1}{n}\sum_{j\neq i}\Gamma_j'(\widetilde{x}_i) - \frac{\rho_n^{1/2}}{\widetilde{x}_i^2}\frac{1}{n}\sum_{j\neq i}\frac{x_{0j}(1-2\rho_n^{1/2}\widetilde{x}_i x_{0j}) + 2(A_{ij}-\rho_n^{1/2}\widetilde{x}_i x_{0j})x_{0j}}{(1-\rho_n^{1/2}\widetilde{x}_i x_{0j})^2}\\
&\quad - \frac{1}{\widetilde{x}_i^3}\frac{1}{n}\sum_{j\neq i}\frac{2(A_{ij}-\rho_n^{1/2}\widetilde{x}_i x_{0j})(1-2\rho_n^{1/2}\widetilde{x}_i x_{0j})^2(1-\rho_n^{1/2}\widetilde{x}_i x_{0j})}{(1-\rho_n^{1/2}\widetilde{x}_i x_{0j})^4}\\
&= -\frac{1}{n}\sum_{j\neq i}\Gamma_j'(\widetilde{x}_i) - \frac{\rho_n^{1/2}}{n\widetilde{x}_i^2}\sum_{j\neq i}\frac{x_{0j}(1-2\rho_n^{1/2}\widetilde{x}_i x_{0j}) + 2(A_{ij}-\rho_n^{1/2}\widetilde{x}_i x_{0j})x_{0j}}{(1-\rho_n^{1/2}\widetilde{x}_i x_{0j})^2}\\
&\quad - \frac{1}{n\widetilde{x}_i^3}\sum_{j\neq i}\frac{2\rho_n^{1/2}(\rho_n^{1/2}x_{0i}-\widetilde{x}_i)x_{0j}(1-2\rho_n^{1/2}\widetilde{x}_i x_{0j})^2(1-\rho_n^{1/2}\widetilde{x}_i x_{0j})}{(1-\rho_n^{1/2}\widetilde{x}_i x_{0j})^4}\\
&\quad - \frac{1}{n\widetilde{x}_i^3}\sum_{j\neq i}\frac{2(A_{ij}-\rho_n x_{0i}x_{0j})(1-2\rho_n^{1/2}\widetilde{x}_i x_{0j})^2(1-\rho_n^{1/2}\widetilde{x}_i x_{0j})}{(1-\rho_n^{1/2}\widetilde{x}_i x_{0j})^4}.
\end{aligned}$$

Similarly, there exists a sufficiently small $\epsilon > 0$, such that over the event $\{\sqrt{n}|\widehat{x}_i^{(\mathrm{ASE})}-\rho_n^{1/2}x_{0i}| < \epsilon(n\rho_n)^{1/2}\}$, we have

$$\begin{aligned}
|\rho_n^{1/2}x_{0i}-\widetilde{x}_i| &\leq |\rho_n^{1/2}x_{0i}-\widehat{x}_i^{(\mathrm{ASE})}| < \epsilon\rho_n^{1/2},\\
|1-2\rho_n^{1/2}\widetilde{x}_i x_{0j}| &\leq 1 + 2\rho_n^{1/2}|\widetilde{x}_i-\rho_n^{1/2}x_{0i}|x_{0j} + 2\rho_n x_{0i}x_{0j} \leq 1+4\rho_n = O(1),\\
1-\rho_n^{1/2}\widetilde{x}_i x_{0j} &\geq 1-\rho_n^{1/2}|\widetilde{x}_i-\rho_n^{1/2}x_{0i}|x_{0j} - \rho_n x_{0i}x_{0j} \geq 1-\rho_n(\epsilon + x_{0i}x_{0j}) = O(1),\\
|\widetilde{x}_i| &\geq -|\widetilde{x}_i-\rho_n^{1/2}x_{0i}| + \rho_n^{1/2}x_{0i} \geq \rho_n^{1/2}x_{0i} - \rho_n^{1/2}\epsilon \gtrsim \rho_n^{1/2},
\end{aligned}$$

and

$$\max_{j\in[n]}|\Gamma_j'(\widetilde{x}_i)|\lesssim\rho_n^{-1/2}.$$

Since the event $\{\sqrt{n}|\widehat{x}_i^{(\mathrm{ASE})}-\rho_n^{1/2}x_{0i}|<\epsilon(n\rho_n)^{1/2}\}$ has probability going to 1, then with probability going to 1,

$$\begin{aligned}
\Psi_n''(\widetilde{x}_i) &= -\frac{1}{n}\sum_{j\neq i}\Gamma_j'(\widetilde{x}_i)-\frac{\rho_n^{1/2}}{\widetilde{x}_i^2}\frac{1}{n}\sum_{j\neq i}\frac{x_{0j}(1-2\rho_n^{1/2}\widetilde{x}_ix_{0j})+2(A_{ij}-\rho_n^{1/2}\widetilde{x}_ix_{0j})x_{0j}}{(1-\rho_n^{1/2}\widetilde{x}_ix_{0j})^2}\\
&\quad-\frac{\rho_n^{1/2}}{\widetilde{x}_i^3}\frac{1}{n}\sum_{j\neq i}\frac{2(\rho_n^{1/2}x_{0i}-\widetilde{x}_i)x_{0j}(1-2\rho_n^{1/2}\widetilde{x}_ix_{0j})^2(1-\rho_n^{1/2}\widetilde{x}_ix_{0j})}{(1-\rho_n^{1/2}\widetilde{x}_ix_{0j})^4}\\
&\quad-\frac{1}{\widetilde{x}_i^3}\frac{1}{n}\sum_{j\neq i}\frac{2(A_{ij}-\rho_nx_{0i}x_{0j})(1-2\rho_n^{1/2}\widetilde{x}_ix_{0j})^2(1-\rho_n^{1/2}\widetilde{x}_ix_{0j})}{(1-\rho_n^{1/2}\widetilde{x}_ix_{0j})^4}\\
&= O(\rho_n^{-1/2})+O(\rho_n^{-1/2})+O(\rho_n^{-1})+O(\rho_n^{-3/2})=O(\rho_n^{-3/2}).
\end{aligned}$$

Hence, we see that $\Psi_n''(\widetilde{x}_i)=O_{\mathbb{P}_0}(\rho_n^{-3/2})$. Since we need to require $\Psi_n''(\widetilde{x}_i)=o_{\mathbb{P}_0}(\sqrt{n})$ in order to establish the asymptotic normality of the one-step estimator, it is hence essential to require that $\Psi_n''(\widetilde{x}_i)=O_{\mathbb{P}_0}(\rho_n^{-3/2})=o_{\mathbb{P}_0}(\sqrt{n})$, which in turn requires that $\rho_n^3n\to\infty$. This is equivalent to require that $n\rho_n=\{n^2(n\rho_n^3)\}^{1/3}=n^{2/3}(n\rho_n^3)^{1/3}\gg n^{2/3}$.

Through the above derivation, we observe that in order the standard technique introduced in Section 5.7 of Van der Vaart (2000) works for establishing the asymptotic normality of the one-step estimator, it is necessary to require that $n\rho_n$ is lower bounded by a polynomial of n, *i.e.*, the graph is moderately sparse.

We also remark that the condition $n\rho_n$ being lower bounded by a polynomial of n is not a sufficient condition for the asymptotic normality of the one-step estimator. The following theorem illustrates this result by providing an example showing that the asymptotic normality of the one-step estimator also occurs under specific setups when $(\log n)^4/(n\rho_n)\to 0$. The proof technique employed below requires a non-standard treatment of the score function and the Fisher information matrix in contrast to the classical technique in Van der Vaart (2000). Note that generalizing the technique below to random dot product graphs with latent dimension d greater than 1 is non-trivial.

Theorem H.1 *Let $\mathbf{A}\sim\mathrm{RDPG}(\mathbf{X}_0)$ with a sparsity factor ρ_n, where the latent position matrix is $\mathbf{X}_0=[p,\ldots,p]^{\mathrm{T}}$ for some $p\in(0,1)$. Fix $i\in[n]$ and let $\widehat{\mathbf{X}}^{(\mathrm{ASE})}$ be the ASE of $\mathbf{A}$ into $\mathbb{R}^1$. Consider an initial estimator $\widetilde{\mathbf{X}}$ of the form*

$$e_j^{\mathrm{T}}\widetilde{\mathbf{X}}=\begin{cases}\rho_n^{1/2}p, & \text{if } j\neq i,\\ e_i^{\mathrm{T}}\mathbf{X}^{(\mathrm{ASE})}, & \text{if } j=i.\end{cases}$$

Let $\widehat{\mathbf{X}}=[\widehat{x}_1,\ldots,\widehat{x}_n]^{\mathrm{T}}$ be the one-step estimator defined by (3.2) with the initial estimator $\widetilde{\mathbf{X}}=[\widetilde{x}_1,\ldots,\widehat{x}_n]^{\mathrm{T}}$ defined as above. Under the condition that $\rho_n\to 0$ but $(\log n)^4/(n\rho_n)\to 0$, we have

$$\sqrt{n}(\widehat{x}_j-\rho_n^{1/2}p)\overset{\mathcal{L}}{\to}\mathrm{N}(0,1),\quad j=1,\ldots,n.$$

Proof of Theorem H.1. We first consider $\widehat{x}_i$. Write

$$
\begin{aligned}
\widehat{x}_i &= \widetilde{x}_i + \left\{\frac{1}{n}\sum_{j=1}^n \frac{\widetilde{x}_j^2}{\widetilde{x}_i\widetilde{x}_j(1-\widetilde{x}_i\widetilde{x}_j)}\right\}^{-1}\left\{\frac{1}{n}\sum_{j=1}^n \frac{(A_{ij}-\widetilde{x}_i\widetilde{x}_j)\widetilde{x}_j}{\widetilde{x}_i\widetilde{x}_j(1-\widetilde{x}_i\widetilde{x}_j)}\right\}\\
&= \widetilde{x}_i + \left\{\frac{1}{n}\sum_{j=1}^n \frac{\widetilde{x}_j}{(1-\widetilde{x}_i\widetilde{x}_j)}\right\}^{-1}\left\{\frac{1}{n}\sum_{j=1}^n \frac{(A_{ij}-\widetilde{x}_i\widetilde{x}_j)}{(1-\widetilde{x}_i\widetilde{x}_j)}\right\}\\
&= \widetilde{x}_i + \left\{\frac{1}{n}\sum_{j\neq i} \frac{\rho_n^{1/2}p}{1-\rho_n^{1/2}\widetilde{x}_i p} + \frac{\widetilde{x}_i}{n(1-\widetilde{x}_i^2)}\right\}^{-1}\left\{\frac{1}{n}\sum_{j\neq i} \frac{(A_{ij}-\rho_n^{1/2}\widetilde{x}_i p)}{(1-\rho_n^{1/2}\widetilde{x}_i p)} - \frac{\widetilde{x}_i^2}{n(1-\widetilde{x}_i^2)}\right\}\\
&= \widetilde{x}_i + \left\{\frac{p}{1-\rho_n^{1/2}\widetilde{x}_i p} - \frac{p}{n(1-\rho_n^{1/2}\widetilde{x}_i p)} + \frac{\widetilde{x}_i}{n\rho_n^{1/2}(1-\widetilde{x}_i^2)}\right\}^{-1}\\
&\quad\times\left\{\frac{1}{n\rho_n^{1/2}}\sum_{j\neq i} \frac{(A_{ij}-\rho_n^{1/2}\widetilde{x}_i p)}{(1-\rho_n^{1/2}\widetilde{x}_i p)} - \frac{\widetilde{x}_i^2}{n\rho_n^{1/2}(1-\widetilde{x}_i^2)}\right\}.
\end{aligned}
$$

Observe that $\widetilde{x}_i = \rho_n^{1/2}p + O_{\mathbb{P}_0}(n^{-1/2})$, $1-\rho_n^{1/2}\widetilde{x}_i p = 1-\rho_n p^2 + o_{\mathbb{P}_0}(1)$, $1-\widetilde{x}_i^2 = 1-\rho_n p^2 + o_{\mathbb{P}_0}(1)$ by the continuous mapping theorem, we see that

$$
\begin{aligned}
-\frac{p}{n(1-\rho_n^{1/2}\widetilde{x}_i p)} + \frac{\widetilde{x}_i}{n\rho_n^{1/2}(1-\widetilde{x}_i^2)} &= \frac{-p}{n\{1-\rho_n p^2 + o_{\mathbb{P}_0}(1)\}} + \frac{\rho_n^{1/2}p + O_{\mathbb{P}_0}(n^{-1/2})}{n\rho_n^{1/2}\{1-\rho_n p^2 + o_{\mathbb{P}_0}(1)\}}\\
&= \frac{-p}{n\{1-\rho_n p^2 + o_{\mathbb{P}_0}(1)\}} + \frac{\rho_n^{1/2}[p + O_{\mathbb{P}_0}\{(n\rho_n)^{-1/2}\}]}{n\rho_n^{1/2}\{1-\rho_n p^2 + o_{\mathbb{P}_0}(1)\}} = O_{\mathbb{P}_0}(n^{-1}),\\
\frac{\widetilde{x}_i^2}{n\rho_n^{1/2}(1-\widetilde{x}_i^2)} &= \frac{\rho_n[p + O_{\mathbb{P}_0}\{(n\rho_n)^{-1/2}]^2\}}{n\rho_n^{1/2}\{1-\rho_n p^2 + o_{\mathbb{P}_0}(1)\}} = O_{\mathbb{P}_0}\left(\frac{\rho_n^{1/2}}{n}\right).
\end{aligned}
$$

Therefore, we proceed to compute

$$
\begin{aligned}
\widehat{x}_i &= \widetilde{x}_i + \left\{\frac{p}{1-\rho_n^{1/2}\widetilde{x}_i p} - \frac{p}{n(1-\rho_n^{1/2}\widetilde{x}_i p)} + \frac{\widetilde{x}_i}{n\rho_n^{1/2}(1-\widetilde{x}_i^2)}\right\}^{-1}\\
&\quad\times\left\{\frac{1}{n\rho_n^{1/2}}\sum_{j\neq i} \frac{(A_{ij}-\rho_n^{1/2}\widetilde{x}_i p)}{(1-\rho_n^{1/2}\widetilde{x}_i p)} - \frac{\widetilde{x}_i^2}{n\rho_n^{1/2}(1-\widetilde{x}_i^2)}\right\}\\
&= \widetilde{x}_i + \left\{\frac{p}{1-\rho_n^{1/2}\widetilde{x}_i p} + O_{\mathbb{P}_0}(n^{-1})\right\}^{-1}\left\{\frac{1}{n\rho_n^{1/2}}\sum_{j\neq i} \frac{(A_{ij}-\rho_n^{1/2}\widetilde{x}_i p)}{(1-\rho_n^{1/2}\widetilde{x}_i p)} + O_{\mathbb{P}_0}(n^{-1})\right\}\\
&= \widetilde{x}_i + \left\{\frac{1-\rho_n^{1/2}p\widetilde{x}_i}{p} + o_{\mathbb{P}_0}(1)\right\}\left\{\frac{1}{n\rho_n^{1/2}}\sum_{j\neq i} \frac{(A_{ij}-\rho_n^{1/2}\widetilde{x}_i p)}{(1-\rho_n^{1/2}\widetilde{x}_i p)} + O_{\mathbb{P}_0}(n^{-1})\right\},
\end{aligned}
$$

where the last inequality is due to the continuous mapping theorem. Since

$$
\begin{aligned}
\frac{1}{n\rho_n^{1/2}}\sum_{j\neq i}\frac{(A_{ij}-\rho_n^{1/2}\widetilde{x}_i p)}{(1-\rho_n^{1/2}\widetilde{x}_i p)} &= \frac{1}{n\rho_n^{1/2}}\sum_{j\neq i}\frac{(A_{ij}-\rho_n p^2)+(\rho_n p^2-\rho_n^{1/2}\widetilde{x}_i p)}{(1-\rho_n^{1/2}\widetilde{x}_i p)}\\
&= \frac{1}{(1-\rho_n^{1/2}\widetilde{x}_i p)}\sum_{j\neq i}\frac{(A_{ij}-\rho_n p^2)}{n\rho_n^{1/2}} + \frac{n-1}{n\rho_n^{1/2}}\frac{(\rho_n p^2-\rho_n^{1/2}\widetilde{x}_i p)}{(1-\rho_n^{1/2}\widetilde{x}_i p)}\\
&= \frac{O_{\mathbb{P}_0}(n^{-1/2})}{1-\rho_n p+o_{\mathbb{P}_0}(1)} + \frac{n-1}{n}\frac{p(\rho_n^{1/2}p-\widetilde{x}_i)}{\{1-\rho_n p^2+o_{\mathbb{P}_0}(1)\}} = O_{\mathbb{P}_0}(n^{-1/2}),
\end{aligned}
$$

it follows that

$$
\begin{aligned}
\widehat{x}_i &= \widetilde{x}_i + \frac{1}{n\rho_n^{1/2}}\sum_{j\neq i}\frac{(A_{ij}-\rho_n^{1/2}\widetilde{x}_i p)}{p} + \frac{1-\rho_n^{1/2}p\widetilde{x}_i}{p}O_{\mathbb{P}_0}(n^{-1}) + o_{\mathbb{P}_0}(n^{-1/2})\\
&= \widetilde{x}_i + \frac{1}{n\rho_n^{1/2}}\sum_{j\neq i}\frac{(A_{ij}-\rho_n^{1/2}\widetilde{x}_i p)}{p} + o_{\mathbb{P}_0}(n^{-1/2}) = \frac{1}{n}\widetilde{x}_i + \frac{1}{n\rho_n^{1/2}}\sum_{j\neq i}\frac{A_{ij}}{p} + o_{\mathbb{P}_0}(n^{-1/2}).
\end{aligned}
$$

Hence we conclude that

$$
\sqrt{n}(\widehat{x}_i-\rho_n^{1/2}p) = \frac{1}{\sqrt{(n-1)\rho_n}}\sum_{j\neq i}\frac{A_{ij}-\rho_n p^2}{p} + o_{\mathbb{P}_0}(1) \xrightarrow{\mathcal{L}} \mathrm{N}(0,1-\rho p^2) = \mathrm{N}(0,1).
$$

by the central limit theorem. This establishes the asymptotic normality for $\widehat{x}_j$ with $j=i$. Next we consider $\widehat{x}_k$ with $k\neq i$. Write

$$
\begin{aligned}
\widehat{x}_k &= \widetilde{x}_k + \left\{\frac{1}{n}\sum_{j=1}^n\frac{\widetilde{x}_j}{\widetilde{x}_k(1-\widetilde{x}_k\widetilde{x}_j)}\right\}^{-1}\left\{\frac{1}{n}\sum_{j=1}^n\frac{A_{kj}-\widetilde{x}_k\widetilde{x}_j}{\widetilde{x}_k(1-\widetilde{x}_k\widetilde{x}_j)}\right\}\\
&= \rho_n^{1/2}p + \left\{\frac{1}{n(1-\rho_n p^2)} + \frac{\widetilde{x}_i}{n\rho_n^{1/2}p(1-\rho_n^{1/2}p\widetilde{x}_i)} + \frac{1}{n}\sum_{j\notin\{i,k\}}\frac{\rho_n^{1/2}p}{\rho_n^{1/2}p(1-\rho_n p^2)}\right\}^{-1}\\
&\qquad\times\left\{\frac{1}{n}\sum_{j\notin\{i,k\}}\frac{A_{kj}-\rho_n p^2}{\rho_n^{1/2}p(1-\rho_n p^2)} + \frac{1}{n}\frac{A_{ki}-\rho_n^{1/2}p\widetilde{x}_i}{\rho_n^{1/2}p(1-\rho_n^{1/2}p\widetilde{x}_i)} - \frac{\rho_n^{1/2}p}{n(1-\rho_n p^2)}\right\}\\
&= \rho_n^{1/2}p + \left\{\frac{1}{n(1-\rho_n p^2)} + \frac{\rho_n^{1/2}[p+O_{\mathbb{P}_0}\{(n\rho_n)^{-1/2}\}]}{n\rho_n^{1/2}p(1-\rho_n p^2+O_{\mathbb{P}_0}\{(n\rho_n)^{-1/2}\})} + \frac{n-2}{n}\frac{1}{1-\rho_n p^2}\right\}^{-1}\\
&\qquad\times\left\{\frac{1}{n}\sum_{j\notin\{i,k\}}\frac{A_{kj}-\rho_n p^2}{\rho_n^{1/2}p(1-\rho_n p^2)} + \frac{1}{n}\frac{A_{ki}-\rho_n^{1/2}p\widetilde{x}_i}{\rho_n^{1/2}p(1-\rho_n^{1/2}p\widetilde{x}_i)} - \frac{\rho_n^{1/2}p}{n(1-\rho_n p^2)}\right\}\\
&= \rho_n^{1/2}p + \left\{o_{\mathbb{P}_0}(1)+(1-\rho_n p^2)\right\}\left\{\frac{1}{n\rho_n^{1/2}}\sum_{j\notin\{i,k\}}\frac{A_{kj}-\rho_n p^2}{p(1-\rho_n p^2)} + O_{\mathbb{P}_0}\left(\frac{1}{\sqrt{n}\sqrt{n\rho_n}}\right)\right\}\\
&= \rho_n^{1/2}p + o_{\mathbb{P}_0}(n^{-1/2}) + \frac{1}{n\rho_n^{1/2}}\sum_{j\neq\{i,k\}}\frac{A_{kj}-\rho_n p^2}{p}.
\end{aligned}
$$

Therefore,

$$\sqrt{n}(\widehat{x}_k - \rho_n^{1/2}p) = \frac{1}{\sqrt{n\rho_n}}\sum_{j\neq k}\frac{A_{kj}-\rho_n p^2}{p} + o_{\mathbb{P}_0}(1) \overset{\mathcal{L}}{\to} \mathrm{N}(0,1)$$

because $\mathrm{var}_0(A_{kj}/\rho_n^{1/2}) = p^2(1-\rho_n p^2) \to p^2$ as $\rho_n \to 0$. The proof is thus completed. $\square$

I Additional Numerical Examples

I.1 Clustering performance in stochastic block models via Chernoff information

In this section, we consider the stochastic block models with positive semidefinite block probability matrices in the context of random dot product graphs and focus on vertex clustering as a subsequent inference task of interest after obtaining estimates of the latent positions or the population LSE. In particular, the following four estimates are considered: the ASE (2.2), the one-step estimator (3.2) initialized at the ASE, abbreviated as OSE-A, the LSE (4.1), and the one-step estimator (4.5) for the population LSE, abbreviated as OSE-L. These estimates are then used as input features for vertex clustering.

Our goal is to compare the vertex clustering using these four estimates rather than the performance of specific clustering algorithms. Hence, we need a criterion that is independent of the choice of the clustering algorithm, but focuses on distributions of the input features. To this end, we introduce the concept of *minimum pairwise Chernoff distance*. Generically, let $\mathbf{x}_1,\ldots,\mathbf{x}_n$ be i.i.d. random variables following a distribution $F \in \{F_1,\ldots,F_K\}$, where $F_k(\mathrm{d}\mathbf{x}) = f_k(\mathbf{x})\mathrm{d}\mathbf{x}$, $k \in [K]$, and suppose the task is to determine whether $F = F_k$ for $k \in [K]$. Assume that $F = F_k$ with prior probability π_k, $k \in [K]$. Then for any decision rule u, the risk of u is $r(u) = \sum_{k=1}^K \pi_k \sum_{l\neq k} p_{kl}(u)$, where $p_{kl}(u)$ is the probability that the decision rule u assigns $F = F_l$ when the underlying true distribution is $F = F_k$. In the context of vertex clustering, the decision rule u plays the role of a clustering algorithm, and $\mathbf{x}_i$'s are treated as the rows of one of the aforementioned four estimates. Since we are interested in a criterion that does not depend on u, it is natural to investigate the behavior of the risk when the optimal decision rule (clustering algorithm) is applied. The following result characterized the optimal error rate (Leang and Johnson, 1997):

$$\inf_u \lim_{n\to\infty} \frac{1}{n} r(u) = -\min_{k\neq l} C(F_k, F_l),$$

where $C(F_k, F_l)$ is the *Chernoff information* between F_k and F_l defined by (Chernoff, 1952, 1956)

$$C(F_k, F_l) = \sup_{t\in(0,1)} \left\{-\log\int f_k(\mathbf{x})^t f_l(\mathbf{x})^{1-t}\mathrm{d}\mathbf{x}\right\}, \tag{I.1}$$

and $\min_{k\neq l} C(F_k, F_l)$ is the minimum pairwise Chernoff distance. This quantity describes the asymptotic decaying rate of the error for the optimal decision rule, with larger values indicating smaller optimal error rate. In our context, since the asymptotic distributions of the rows of the four estimators are multivariate normal, it is useful to derive the Chernoff information for two multivariate normal distributions:

$$C(F_k, F_l) = \sup_{t\in(0,1)} \left\{ \frac{t(1-t)}{2}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_l)^{\mathrm{T}} \mathbf{V}_t^{-1}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_l) + \frac{1}{2}\log\frac{|\mathbf{V}_t|}{|\mathbf{V}_k|^t|\mathbf{V}_l|^{1-t}} \right\},$$

where $F_k = \mathrm{N}(\boldsymbol{\mu}_k, \mathbf{V}_k)$ and $F_l = \mathrm{N}(\boldsymbol{\mu}_l, \mathbf{V}_l)$, and $\mathbf{V}_t = t\mathbf{V}_k + (1-t)\mathbf{V}_l$.

For a K-block stochastic block model with a positive semidefinite block probability matrix $\mathbf{B} = (\mathbf{X}_0^*)(\mathbf{X}_0^*)^{\mathrm{T}}$, where $\mathbf{X}_0^* = [\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_K]^{\mathrm{T}} \in \mathbb{R}^{K\times d}$, $d \le K$, and a cluster assignment function $\tau : [n] \to [K]$ satisfying $(1/n)\sum_{i=1}^n \mathbb{1}\{\tau(i) = k\} \to \pi_k$ for $k \in [K]$ and $\sum_{k=1}^K \pi_k = 1$, we define the following quantities for the ASE, the LSE, the OSE-A, and the OSE-L, respectively:

$$\begin{aligned}
\rho^*_{\mathrm{ASE}} &= \min_{k\neq l}\sup_{t\in(0,1)} \frac{nt(1-t)}{2}(\boldsymbol{\nu}_k - \boldsymbol{\nu}_{0l})^{\mathrm{T}}\boldsymbol{\Sigma}_{kl}^{-1}(t)(\boldsymbol{\nu}_k - \boldsymbol{\nu}_l),\\
\rho^*_{\mathrm{LSE}} &= \min_{k\neq l}\sup_{t\in(0,1)} \frac{n^2t(1-t)}{2}(\mathbf{y}^*_{0k} - \mathbf{y}^*_{0l})^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}_{kl}^{-1}(t)(\mathbf{y}^*_{0k} - \mathbf{y}^*_{0l}),\\
\rho^*_{\mathrm{OSE-A}} &= \min_{k\neq l}\sup_{t\in(0,1)} \frac{nt(1-t)}{2}(\boldsymbol{\nu}_k - \boldsymbol{\nu}_{0l})^{\mathrm{T}}\mathbf{G}_{kl}^{-1}(t)(\boldsymbol{\nu}_k - \boldsymbol{\nu}_{0l}),\\
\rho^*_{\mathrm{OSE-L}} &= \min_{k\neq l}\sup_{t\in(0,1)} \frac{n^2t(1-t)}{2}(\mathbf{y}^*_{0k} - \mathbf{y}^*_{0l})^{\mathrm{T}}\widetilde{\mathbf{G}}_{kl}^{-1}(t)(\mathbf{y}^*_{0k} - \mathbf{y}^*_{0l}),
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\Sigma}_{kl}(t) &= t\boldsymbol{\Sigma}(\boldsymbol{\nu}_k) + (1-t)\boldsymbol{\Sigma}(\boldsymbol{\nu}_l), & \widetilde{\boldsymbol{\Sigma}}_{kl}(t) &= t\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\nu}_k) + (1-t)\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\nu}_l),\\
\mathbf{G}_{kl}(t) &= t\mathbf{G}(\boldsymbol{\nu}_k)^{-1} + (1-t)\mathbf{G}(\boldsymbol{\nu}_l)^{-1}, & \widetilde{\mathbf{G}}_{kl}(t) &= t\widetilde{\mathbf{G}}_k(\boldsymbol{\nu}_k) + (1-t)\widetilde{\mathbf{G}}(\boldsymbol{\nu}_l),
\end{aligned}$$

and $\mathbf{y}^*_{0k} = \boldsymbol{\nu}_k(\sum_{l=1}^K n\pi_l\boldsymbol{\nu}_k^{\mathrm{T}}\boldsymbol{\nu}_l)^{-1/2}$. These quantities are motivated by the use of the minimum pairwise Chernoff distance for measuring clustering performance. Note that for all $t \in (0,1)$, we have seen in Section 3 and Section 4 that $\boldsymbol{\Sigma}_{kl}(t) \succeq \mathbf{G}_{kl}(t)$ and $\widetilde{\boldsymbol{\Sigma}}_{kl}(t) \succeq \widetilde{\mathbf{G}}_{kl}(t)$. It follows that $\rho^*_{\mathrm{ASE}} \le \rho^*_{\mathrm{OSE-A}}$ and $\rho^*_{\mathrm{LSE}} \le \rho^*_{\mathrm{OSE-L}}$ regardless of the choice of the underlying true latent positions. Namely, the decaying rate of the optimal decision error using the OSE-A is always smaller than that using the ASE, and the same conclusion holds for the comparison between the OSE-L and the LSE. We also note that the above criteria are independent of the choice of the clustering algorithm and only depend on the distribution of the input features.

Example 1 *We revisit Example 1 in Section 3. Consider the following rank-one stochastic block model example with two communities on n vertices. The block probability matrix is*

$$\mathbf{B} = \rho_n \begin{bmatrix} p^2 & pq \\ pq & q^2 \end{bmatrix},$$

where $\rho_n \in (0,1]$ is the sparsity factor, $p, q \in (0,1)$, and the cluster assignment function $\tau : [n] \to [2]$ satisfies $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{\tau(i) = 1\} = \pi_1$, $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{\tau(i) = 2\} = \pi_2$, where $[\pi_1, \pi_2]^{\mathrm{T}}$ is a probability vector with $\pi_1 + \pi_2 = 1$. The distribution F satisfying condition (2.1) can be explicitly computed: $F(\mathrm{d}x) = \pi_1\delta_p(\mathrm{d}x) + \pi_2\delta_q(\mathrm{d}x)$ with $\pi_1 + \pi_2 = 1$, $p, q \in (0,1)$.

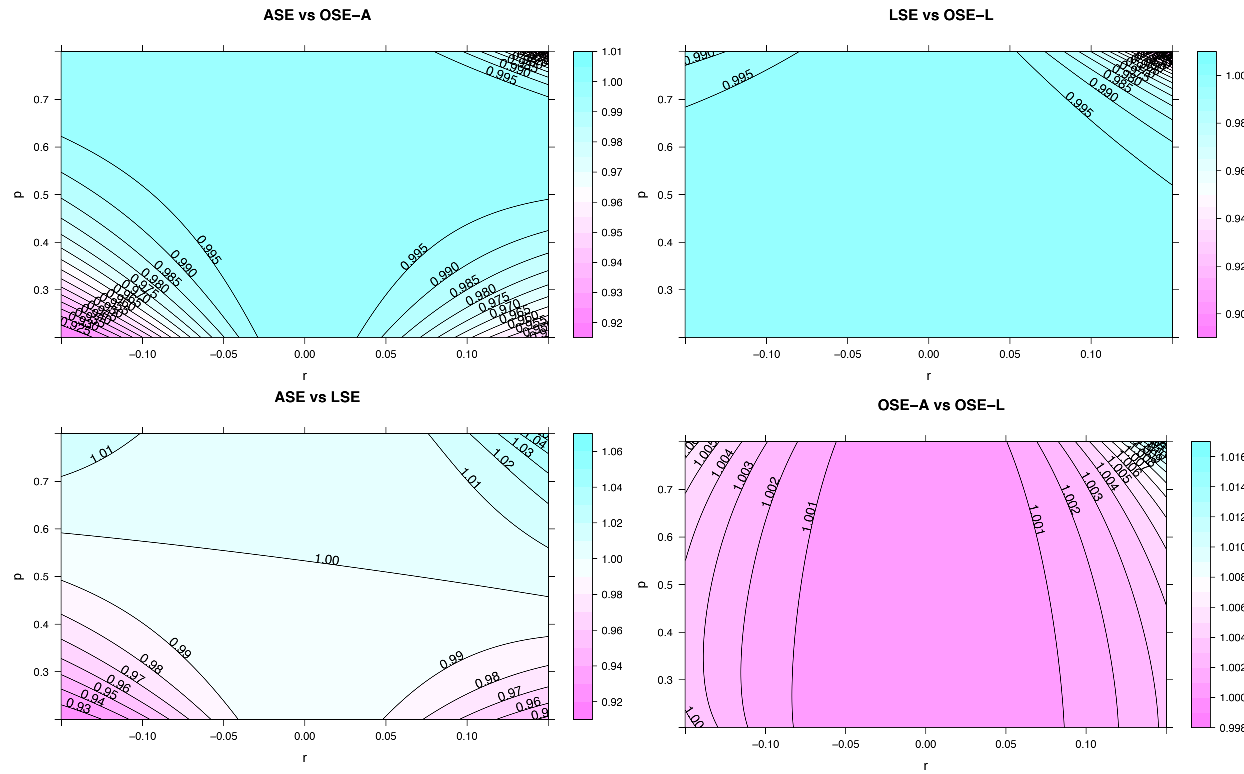


Figure 1: Heatmap and level curves of the ratios $\rho^*_{\mathrm{ASE}}/\rho^*_{\mathrm{OSE-A}}$, $\rho^*_{\mathrm{LSE}}/\rho^*_{\mathrm{OSE-L}}$, $\rho^*_{\mathrm{ASE}}/\rho^*_{\mathrm{LSE}}$, and $\rho^*_{\mathrm{OSE-A}}/\rho^*_{\mathrm{OSE-L}}$ for $p \in [0.2, 0.8]$ and $r \in [-0.15, 0.15]\backslash\{0\}$ for Example 1.

We first consider the dense setting where $\rho_n \equiv 1$ for all n. Then simple algebra yields

$$\rho^*_{\mathrm{OSE-A}} = \frac{n(p-q)^2}{2}\{G(p)^{-1/2} + G(q)^{-1/2}\}^{-2},$$
$$\rho^*_{\mathrm{OSE-L}} = \frac{n(p-q)^2}{2}\left\{\frac{\sqrt{p}+\sqrt{q}}{2\sqrt{p}}G(p)^{-1/2} + \frac{\sqrt{p}+\sqrt{q}}{2\sqrt{q}}G(q)^{-1/2}\right\}^{-2}.$$

where $G(p) = \frac{\pi_1 p^2}{p^2(1-p^2)} + \frac{\pi_2 q^2}{pq(1-pq)}$ and $G(q) = \frac{\pi_1 p^2}{pq(1-pq)} + \frac{\pi_2 q^2}{q^2(1-q^2)}$. We have already shown that $\rho^*_{\mathrm{ASE}} \leq \rho^*_{\mathrm{OSE-A}}$ and $\rho^*_{\mathrm{LSE}} \leq \rho^*_{\mathrm{OSE-L}}$ always hold for stochastic block models. Furthermore, in this specific example, one can show that $\rho^*_{\mathrm{OSE-A}} \geq \rho_{\mathrm{OSE-L}}$ always holds regardless of the choice of p, q, π_1, and π_2. This means that the OSE-A dominates the OSE-L in terms of the optimal error rate in this specific rank-one stochastic block model example. Note that OSE-A does not necessarily dominate the OSE-L in general and another example is provided in the Supplementary Material. To visualize these findings, we fix $\pi_1 = 0.6, \pi_2 = 0.4$, let p range over $[0.2, 0.8]$, $r = q - p$ range over in $[-0.15, 0.15]\backslash\{0\}$, compute the ratios $\rho^*_{\mathrm{ASE}}/\rho^*_{\mathrm{OSE-A}}$, $\rho^*_{\mathrm{LSE}}/\rho^*_{\mathrm{OSE-L}}$, $\rho^*_{\mathrm{ASE}}/\rho^*_{\mathrm{LSE}}$, $\rho^*_{\mathrm{OSE-A}}/\rho^*_{\mathrm{OSE-L}}$, and plot the numerical results in Figure 1.

Besides the aforementioned large sample conclusion, we perform two finite-sample experiments as well. We first compute the four estimates with $n = 200$, $p = 0.6, q = 0.4$ and $n = 200$, $p = 0.45$, $q = 0.6$, respectively, in the dense setting where $\rho_n \equiv 1$. In each of the scenarios, we choose the Gaussian-mixture-

Table 2: Rand indices of the GMM-based clustering algorithm using different estimates for Example 1 (dense). For each setup $p = 0.6, q = 0.4$ and $p = 0.45$, $q = 0.6$, the Rand indices are averaged over 1000 Monte Carlo replicates of adjacency matrices. The standard errors are included in the parentheses.

Estimates	ASE	OSE-A
$p = 0.6$, $q = 0.4$	0.8985 (1.13×10^{-3})	**0.9022** (1.12×10^{-3})
$p = 0.45$, $q = 0.6$	0.7635 (2.06×10^{-3})	**0.7899** (1.70×10^{-3})
Estimates	LSE	OSE-L
$p = 0.6$, $q = 0.4$	0.8966 (1.26×10^{-3})	0.8972 (1.25×10^{-3})
$p = 0.45$, $q = 0.6$	0.7742 (1.64×10^{-3})	0.7863 (1.37×10^{-3})

model-based (GMM-based) clustering algorithm (Fraley et al., 2012; Fraley and Raftery, 2002), which is recommended in Tang and Priebe (2018), for the subsequent vertex clustering task. To evaluate the finite-sample clustering results, we adopt the Rand index (Rand, 1971) as a measurement of the clustering accuracy. Formally, given two partitions $\mathcal{C}_1 = \{c_{11}, \dots, c_{1r}\}$ *and* $\mathcal{C}_2 = \{c_{21}, \dots, c_{2s}\}$ *of* $[n]$*, let a be the number of pairs in* $[n]$ *that are both in the same block in partition* $\mathcal{C}_1$ *and in the same block in partition* $\mathcal{C}_2$*, and b the number of pairs in* $[n]$ *that are neither in the same block in* $\mathcal{C}_1$ *nor in the same block in* $\mathcal{C}_2$*. Then the Rand index (between* $\mathcal{C}_1$ *and* $\mathcal{C}_2$*) is defined by* $\mathrm{RI}(\mathcal{C}_1, \mathcal{C}_2) = 2(a + b)/\{n(n-1)\}$*. The Rand index ranges between* 0 *and* 1*, with higher value suggesting a better agreement between* $\mathcal{C}_1$ *and* $\mathcal{C}_2$*. Table 2 reports the average Rand indices of the four cluster assignment estimates in comparison with the underlying true cluster assignment based on* 1000 *Monte Carlo replicates, together with the corresponding standard errors. The differences in the Rand indices are statistically significant at level* $\alpha = 0.01$*, and the results are in accordance with the aforementioned large sample conclusion.*

In addition, we also consider finite-sample experiments in a sparse setting where the sparsity factor ρ_n *decays to* 0 *at the rate* $\rho_n = (\log n)^{-1}$*. The number of vertices n is set to* 1000*, and we consider two setups* $p = 0.6$*,* $q = 0.4$ *and* $p = 0.45$*,* $q = 0.6$*, respectively. For each setup, we also apply the GMM-based clustering algorithm to the ASE, the OSE-A, the LSE, and the OSE-L, and compute the Rand indices to evaluate the clustering accuracy. The experiments are repeated independently for* 1000 *Monte Carlo replicates. We tabulate the resulting average Rand indices of the four estimates in comparison with the true cluster assignment in Table 3, together with the corresponding standard errors, based on* 1000 *Monte Carlo replicates. The differences are also statistically significant at level* $\alpha = 0.01$*. Note that the Rand indices of the LSE and the OSE-L are similar in the sparse regime. This is because the asymptotic variance of the LSE and that of the OSE-L coincide when* $\rho_n \to \infty$*, namely,* $\widetilde{\Sigma}(p) = \widetilde{G}(p)$ *and* $\widetilde{\Sigma}(q) = \widetilde{G}(q)$ *for all* $p, q \in (0, 1)$*.*

Table 3: Rand indices of the GMM-based clustering algorithm using different estimates for Example 1 (sparse). For each setup $p = 0.6, q = 0.4$ and $p = 0.45$, $q = 0.6$, the Rand indices are averaged over 1000 Monte Carlo replicates of adjacency matrices. The standard errors are included in the parentheses.

Estimates	ASE	OSE-A
$p = 0.6$, $q = 0.4$	0.5453 (1.2×10^{-3})	0.5688 (1.3×10^{-3})
$p = 0.45$, $q = 0.6$	0.4012 (1.0×10^{-3})	**0.4105** (1.1×10^{-3})
Estimates	LSE	OSE-L
$p = 0.6$, $q = 0.4$	**0.6126** (1.2×10^{-3})	**0.6140** (1.1×10^{-3})
$p = 0.45$, $q = 0.6$	0.3788 (1.4×10^{-3})	0.3813 (1.2×10^{-3})

Example 2 *As another example, we consider the following rank-two stochastic block model with three communities. The block probability matrix* $\mathbf{B}$ *is given by* $\mathbf{B} = (\mathbf{X}_0^*)(\mathbf{X}_0^*)^{\mathrm{T}}$*, where*

$$\mathbf{X}_0^* = \begin{bmatrix} q & q & p \\ q & p & q \end{bmatrix}.$$

Let $\tau : [n] \to \{1, 2, 3\}$ *be the corresponding cluster assignment function satisfying*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\tau(i) = 1\} = 0.8, \quad \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\tau(i) = 2\} = 0.1, \quad \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\tau(i) = 2\} = 0.1.$$

We let p *range over* $[0.3, 0.6]$ *and* $q = p - r$ *with* $r \in [-0.2, -0.01]$*. We then explore the minimum pairwise Chernoff distance via the computation of the ratios* $\rho^*_{\mathrm{ASE}}/\rho^*_{\mathrm{OSE-A}}$*,* $\rho^*_{\mathrm{LSE}}/\rho^*_{\mathrm{OSE-L}}$*,* $\rho^*_{\mathrm{ASE}}/\rho^*_{\mathrm{LSE}}$*,* $\rho^*_{\mathrm{OSE-A}}/\rho^*_{\mathrm{OSE-L}}$*, and plot the ratios in Figure 2. Panels (a) and (b) in Figure 2 show that the OSE-A outperforms the ASE and the OSE-L outperforms the LSE, respectively, in terms of the optimal clustering error rates. Panel (c) of Figure 2 indicate that the LSE outperforms the ASE for sparser stochastic block models corresponding to the lower-left region of the heatmap, and a similar conclusion of the comparison between the OSE-A and OSE-L can be drawn from panel (d) of Figure 2 as well.*

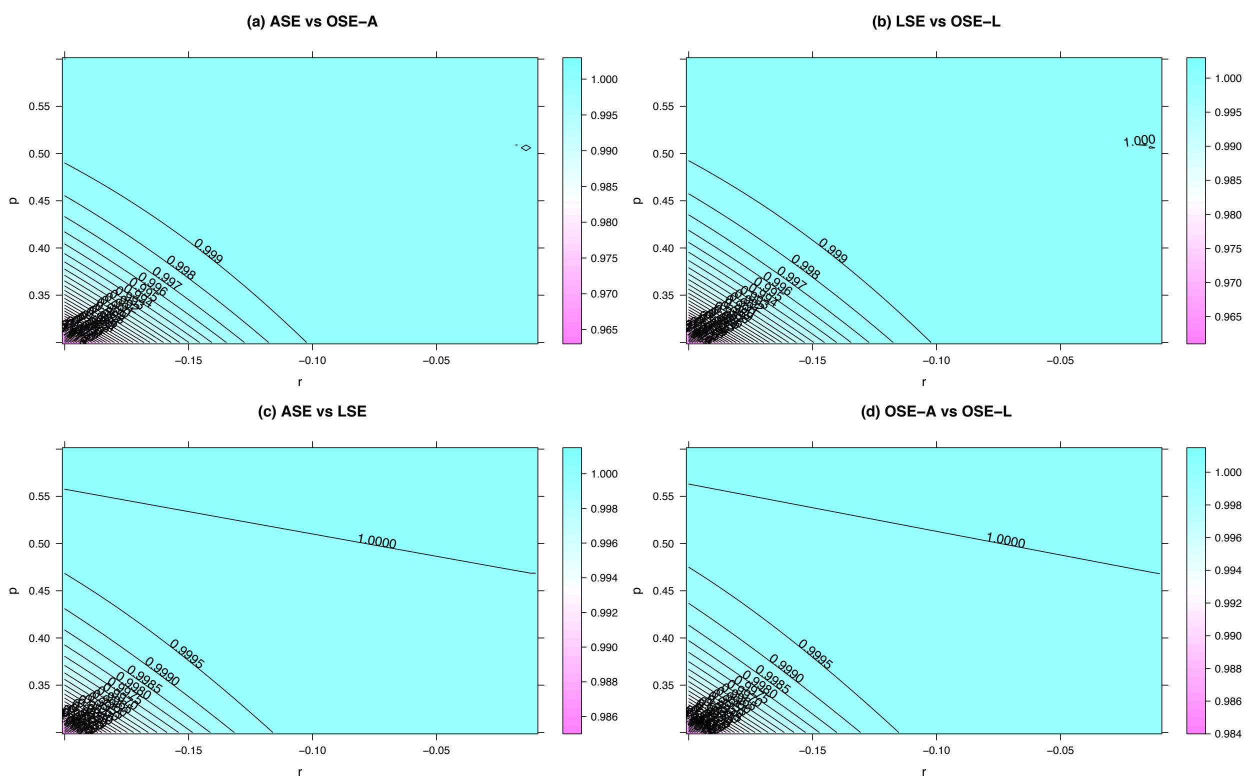


Figure 2: Heatmap and level curves of the ratios $\rho^*_{\mathrm{ASE}}/\rho^*_{\mathrm{OSE-A}}$, $\rho^*_{\mathrm{LSE}}/\rho^*_{\mathrm{OSE-L}}$, $\rho^*_{\mathrm{ASE}}/\rho^*_{\mathrm{LSE}}$, and $\rho^*_{\mathrm{OSE-A}}/\rho^*_{\mathrm{OSE-L}}$ for $p \in [0.2, 0.8]$ and $r \in [-0.15, 0.15]\backslash\{0\}$ for Example 2.

Remark 1 *Unlike the minimum pairwise Chernoff distance, which is an asymptotic criterion for comparing the performance of different estimators in terms of the subsequent optimal clustering rates and does not depend on the clustering algorithm, the Rand index can only reflect the behavior of the clustering result in a finite-sample experiment and may depend on the clustering method we choose.*

I.2 A Dense Three-Block Stochastic Block Model Example

We next consider the following three-block stochastic block model on n vertices with the block probability matrix $\mathbf{B} = (\mathbf{X}_0^*)(\mathbf{X}_0^*)^{\mathrm{T}}$, where

$$(\mathbf{X}_0^*)^{\mathrm{T}} = \begin{bmatrix} 0.3 & 0.3 & 0.6 \\ 0.3 & 0.6 & 0.3 \end{bmatrix},$$

and a cluster assignment function $\tau : [n] \to [3]$, such that as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\tau(i) = 1\} \to 0.3, \quad \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\tau(i) = 2\} \to 0.3, \quad \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\tau(i) = 3\} \to 0.4.$$

The corresponding distribution F satisfying condition (2.1) is $F(\mathrm{d}\mathbf{x}) = \sum_{k=1}^{3} \pi_k \delta_{\boldsymbol{\nu}_k}(\mathrm{d}\mathbf{x})$, where $\pi_1 = \pi_2 = 0.3$, $\pi_3 = 0.4$, $\boldsymbol{\nu}_1 = [0.3, 0.3]^{\mathrm{T}}$, $\boldsymbol{\nu}_2 = [0.3, 0.6]^{\mathrm{T}}$, and $\boldsymbol{\nu}_3 = [0.6, 0.3]^{\mathrm{T}}$. For each $n \in \{500, 600, \ldots, 1200\}$, we generate 10000 replicates of the simulated adjacency matrices from the above sampling model, and then compute the following four estimates: the ASE (2.2), the one-step estimate (3.2) initialized at the ASE (OSE-A), the LSE (4.1), and the one-step estimate (4.5) for the population LSE (OSE-L). The goal is to compare the performance of vertex clustering by applying the GMM-based clustering algorithm to these estimates.

Figure 3 and Table 4 present the Rand indices of clustering results obtained by applying the GMM-based clustering algorithm to the four estimates against the underlying true cluster assignment, and these Rand indices are averaged based on 10000 Monte Carlo replicates. The standard errors corresponding to the Monte Carlo replicates are tabulated in the parentheses of Table 4. When the number of vertices $n \in \{500, 600, 700, 800\}$, the clustering results based on the ASE outperform the rest of the competitors. However, as n increases with $n \geq 900$, the best result is given by either the OSE-A or the OSE-L, and the differences in the Rand indices are statistically significant at level $\alpha = 0.01$. In particular, when $n \in \{1100, 1200\}$, the OSE-A and the OSE-L yield better results than the ASE and the LSE, respectively. These numerical results are in accordance with the fact that asymptotically, the ASE and the LSE are dominated by the OSE-A and OSE-L, respectively.

For each $n \in \{600, 900, 1200\}$, we also compute the OSE-A $\widehat{\mathbf{X}}$ and the OSE-L $\widehat{\mathbf{Y}}$ for each block, as well as the corresponding cluster-specific sample covariance matrices after applying the appropriate orthogonal transformation towards the underlying true $\mathbf{X}_0$ and $\mathbf{Y}_0$, for one randomly selected instance among the 10000 replicated adjacency matrices. The results are tabulated in Table 5 and Table 6, respectively, in comparison with the limit covariance matrices given by Theorem 1 (for the ASE), Theorem 5 (for the OSE-A), Theorem 6 (for the LSE), and Theorem 9 (for the OSE-L). It can be seen that as n increases, the sample covariance matrices converge to their corresponding cluster-specific limit covariance matrices. The scatter points of $\widehat{\mathbf{X}}$

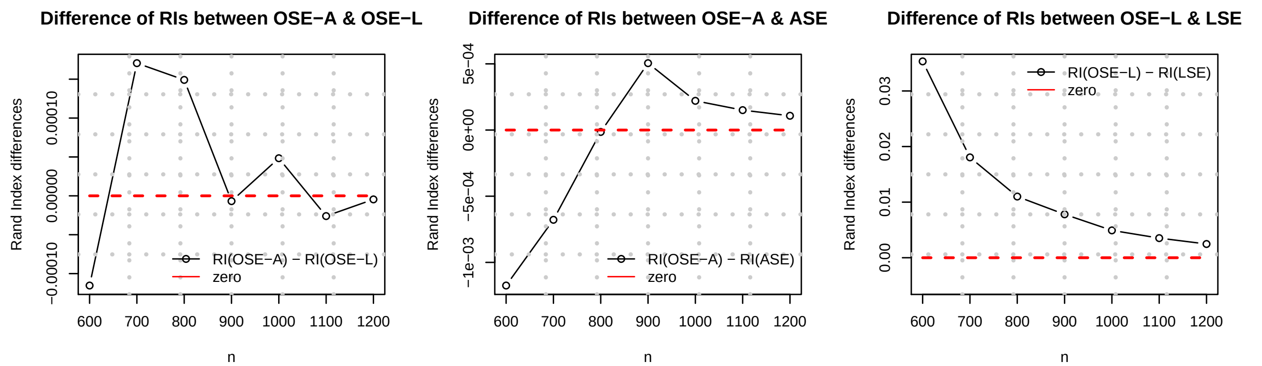


Figure 3: Differences of the Rand indices for Subsection I.2: The Rand indices are obtained by applying the GMM-model-based clustering method to different estimates (the ASE, the OSE-A, the LSE, and the OSE-L) when the number of vertices n ranges in $\{500, 600, \ldots, 1200\}$. The three panels are the plots of the differences of Rand indices between the OSE-A and the OSE-L, those between the OSE-A and the ASE, and those between the OSE-L and the LSE, respectively. The results are averaged based on 10000 Monte Carlo replicates.

Table 4: Rand indices for Subsection I.2: The Rand indices are obtained by applying the GMM-model-based clustering method to different estimates (the ASE, the OSE-A, the LSE, and the OSE-L) when the number of vertices n ranges in $\{500, 600, \ldots, 1200\}$, and for each n, the Rand indices are averaged over 10000 Monte Carlo replicates of adjacency matrices, with the standard errors included in parentheses.

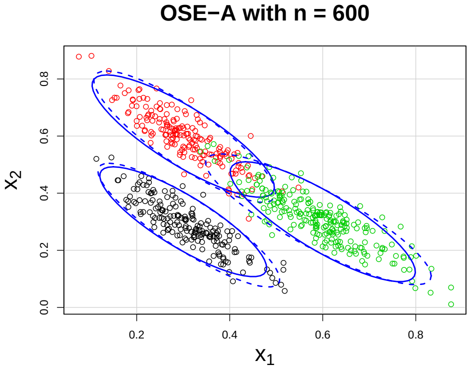
Estimates	ASE	OSE-A	LSE	OSE-L
$n = 500$	**0.89783** (3.3×10^{-4})	0.89199 (4.0×10^{-4})	0.82708 (5.9×10^{-4})	0.89761 (3.4×10^{-4})
$n = 600$	**0.93329** (1.9×10^{-4})	0.93212 (2.2×10^{-4})	0.89691 (4.2×10^{-4})	0.93224 (2.0×10^{-4})
$n = 700$	**0.95445** (1.3×10^{-4})	0.95378 (1.4×10^{-4})	0.93555 (2.1×10^{-4})	0.95361 (1.4×10^{-4})
$n = 800$	**0.96857** (8.6×10^{-5})	0.96856 (8.6×10^{-5})	0.95740 (1.2×10^{-4})	0.96841 (8.8×10^{-5})
$n = 900$	0.97811 (6.4×10^{-5})	0.97861 (6.1×10^{-5})	0.97081 (8.2×10^{-5})	**0.97862** (6.0×10^{-5})
$n = 1000$	0.98443 (5.0×10^{-5})	**0.98465** (4.9×10^{-5})	0.97968 (6.1×10^{-5})	0.98460 (4.8×10^{-5})
$n = 1100$	0.98894 (4.0×10^{-5})	0.98909 (3.9×10^{-5})	0.98559 (4.6×10^{-5})	**0.98911** (3.9×10^{-5})
$n = 1200$	0.99213 (3.1×10^{-5})	0.99224 (3.0×10^{-5})	0.98978 (3.6×10^{-5})	**0.99225** (3.0×10^{-5})

and $\widehat{\mathbf{Y}}$ after applying the orthogonal alignment matrix $\mathbf{W}$ towards $\mathbf{X}_0$ and $\mathbf{Y}_0$ are visualized in Figure 4, along with the cluster-specific 95% empirical and asymptotic confidence ellipses in dashed lines and solid lines, respectively. These figures also validate the aforementioned limit results empirically.

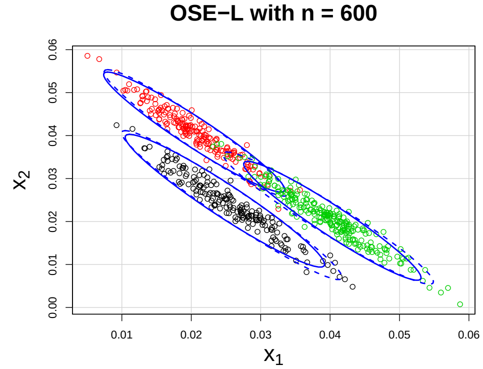
I.3 A Sparser Three-Block Stochastic Block Model

We continue with the stochastic block model considered in I.2 and make modification as follows. The underlying sampling model is still a random dot product graph, but the sparsity level of the graph changes with the number of vertices. The block probability matrix is given by a constant multiple of $\mathbf{B}$, namely,

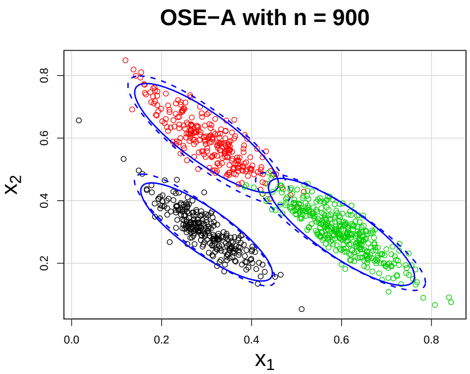
$$\mathbf{B} = (\mathbf{X}_0^*)(\mathbf{X}_0^*)^{\mathrm{T}}, \quad \text{where} \quad (\mathbf{X}_0^*)^{\mathrm{T}} = \alpha_n \begin{bmatrix} 0.3 & 0.3 & 0.6 \\ 0.3 & 0.6 & 0.3 \end{bmatrix},$$



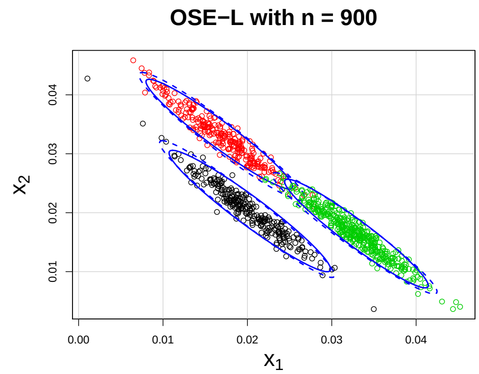
OSE-A with $n = 600$



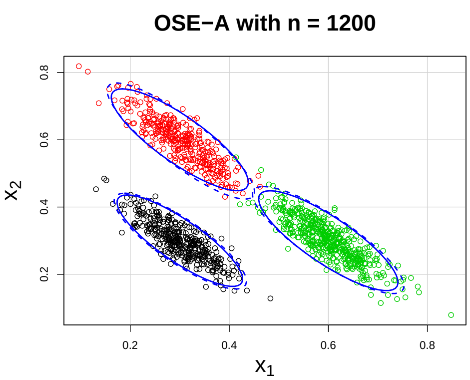
OSE-L with $n = 600$



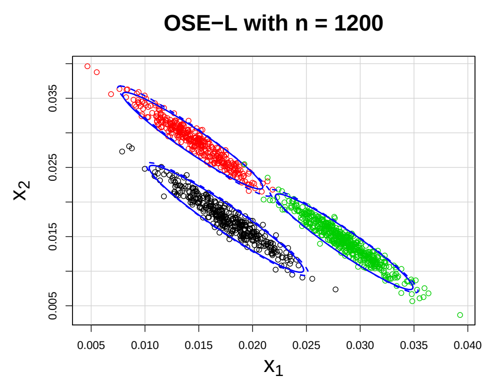
OSE-A with $n = 900$



OSE-L with $n = 900$



OSE-A with $n = 1200$



OSE-L with $n = 1200$

Figure 4: Scatter plots of the OSE-A and OSE-L in the three-block stochastic block model example with n vertices in Section I.2, with $n \in \{600, 900, 1200\}$. The scatter points are colored according to the cluster assignment of the corresponding vertices. For each specific cluster, the 95% empirical confidence ellipses are displayed by the dashed lines, along with the 95% asymptotic confidence ellipses drawn using the solid lines, as provided by Theorem 5 and Theorem 9.

Table 5: Three-block stochastic block model example in Section I.2: the cluster-specific sample covariance matrices for the OSE-A with the number of vertices $n \in \{600, 900, 1200\}$, in comparison with the limit covariance matrix of the OSE-A and the ASE.

	$n = 600$	$n = 900$	$n = 1200$
$\mathbf{G}_n(\boldsymbol{\nu}_1)^{-1}$	$\begin{bmatrix} 3.762811 & -3.651978 \\ -3.651978 & 4.621907 \end{bmatrix}$	$\begin{bmatrix} 3.647588 & -3.669011 \\ -3.669011 & 4.756269 \end{bmatrix}$	$\begin{bmatrix} 3.550210 & -3.169485 \\ -3.169485 & 4.032291 \end{bmatrix}$
	$\mathbf{G}(\boldsymbol{\nu}_1)^{-1}$	$\boldsymbol{\Sigma}(\boldsymbol{\nu}_1)$	
Limit Covariances	$\begin{bmatrix} 3.220559 & -2.898602 \\ -2.898602 & 3.703496 \end{bmatrix}$	$\begin{bmatrix} 3.221615 & -2.895962 \\ -2.895962 & 3.710096 \end{bmatrix}$	
	$n = 600$	$n = 900$	$n = 1200$
$\mathbf{G}_n(\boldsymbol{\nu}_2)^{-1}$	$\begin{bmatrix} 3.765898 & -3.803725 \\ -3.803725 & 5.213684 \end{bmatrix}$	$\begin{bmatrix} 4.846434 & -4.701554 \\ -4.701554 & 6.290449 \end{bmatrix}$	$\begin{bmatrix} 4.403550 & -4.438183 \\ -4.438183 & 5.918178 \end{bmatrix}$
	$\mathbf{G}(\boldsymbol{\nu}_2)^{-1}$	$\boldsymbol{\Sigma}(\boldsymbol{\nu}_2)$	
Limit Covariances	$\begin{bmatrix} 3.844914 & -3.518540 \\ -3.518540 & 4.590484 \end{bmatrix}$	$\begin{bmatrix} 3.844943 & -3.519037 \\ -3.519037 & 4.598917 \end{bmatrix}$	
	$n = 600$	$n = 900$	$n = 1200$
$\mathbf{G}_n(\boldsymbol{\nu}_3)^{-1}$	$\begin{bmatrix} 5.812994 & -4.747638 \\ -4.747638 & 5.186311 \end{bmatrix}$	$\begin{bmatrix} 5.322016 & -4.583552 \\ -4.583552 & 5.301519 \end{bmatrix}$	$\begin{bmatrix} 4.664418 & -4.140929 \\ -4.140929 & 5.045003 \end{bmatrix}$
	$\mathbf{G}(\boldsymbol{\nu}_3)^{-1}$	$\boldsymbol{\Sigma}(\boldsymbol{\nu}_3)$	
Limit Covariances	$\begin{bmatrix} 3.969281 & -3.495403 \\ -3.495403 & 4.414981 \end{bmatrix}$	$\begin{bmatrix} 3.966424 & -3.496907 \\ -3.496907 & 4.414189 \end{bmatrix}$	

and α_n is a scaling factor that varies with the number of vertices n specified later. The cluster assignment function $\tau : [n] \to [3]$ is such that

$$\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{\tau(i) = 1\} \to 0.3, \quad \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{\tau(i) = 2\} \to 0.3, \quad \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{\tau(i) = 3\} \to 0.4,$$

as $n \to \infty$. The specification of τ remains the same as in Section I.2. Let $\boldsymbol{\pi} = [0.3, 0.3, 0.4]^{\mathrm{T}}$. The scaling factor α_n is defined as follows:

$$\alpha_n = \sqrt{\frac{\mathcal{D}}{n\boldsymbol{\pi}^{\mathrm{T}}\widetilde{\mathbf{B}}\boldsymbol{\pi}}}, \quad \text{where} \quad \widetilde{\mathbf{B}} = \begin{bmatrix} 0.3 & 0.3 \\ 0.3 & 0.6 \\ 0.6 & 0.3 \end{bmatrix}\begin{bmatrix} 0.3 & 0.3 & 0.6 \\ 0.3 & 0.6 & 0.3 \end{bmatrix}.$$

It is designed such that the average expected degree of vertices in the resulting graph is fixed at a constant $\mathcal{D}$ when the number of vertices n varies. Here we set $\mathcal{D} = 300$. The remaining settings are the same as those in Section I.2, and we compute the Rand indices of the GMM-based clustering algorithm applied to the four estimates (the ASE, the OSE-A, the LSE, and the OSE-L). The results are tabulated in Table 7 and visualized in Figure 5. Note that in contrast to Table 4 and Figure 3 in the manuscript, the Rand

Table 6: Three-block stochastic block model example in Section I.2: the cluster-specific sample covariance matrices for the OSE-L with the number of vertices $n \in \{600, 900, 1200\}$, in comparison with the limit covariance matrix of the OSE-L and the LSE.

k	$n = 600$	$n = 900$	$n = 1200$
$\widetilde{\mathbf{G}}_n(\boldsymbol{\nu}_1)$	$\begin{bmatrix} 14.90763 & -15.75143 \\ -15.75143 & 17.78524 \end{bmatrix}$	$\begin{bmatrix} 14.41713 & -15.60931 \\ -15.60931 & 18.02098 \end{bmatrix}$	$\begin{bmatrix} 13.55155 & -14.03450 \\ -14.03450 & 15.8845 \end{bmatrix}$
	$\widetilde{\mathbf{G}}(\boldsymbol{\nu}_1)$	$\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\nu}_1)$	
Limit Covariances	$\begin{bmatrix} 12.40965 & -12.78420 \\ -12.78420 & 14.37281 \end{bmatrix}$	$\begin{bmatrix} 12.41030 & -12.78353 \\ -12.78353 & 14.37349 \end{bmatrix}$	
	$n = 600$	$n = 900$	$n = 1200$
$\widetilde{\mathbf{G}}_n(\boldsymbol{\nu}_2)$	$\begin{bmatrix} 10.29810 & -11.05184 \\ -11.05184 & 12.82275 \end{bmatrix}$	$\begin{bmatrix} 12.95708 & -13.73205 \\ -13.73205 & 15.79435 \end{bmatrix}$	$\begin{bmatrix} 12.12355 & -13.08517 \\ -13.08517 & 15.15464 \end{bmatrix}$
	$\widetilde{\mathbf{G}}(\boldsymbol{\nu}_2)$	$\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\nu}_2)$	
Limit Covariances	$\begin{bmatrix} 10.22658 & -10.48123 \\ -10.48123 & 11.73625 \end{bmatrix}$	$\begin{bmatrix} 10.23471 & -10.48190 \\ -10.48190 & 11.73631 \end{bmatrix}$	
	$n = 600$	$n = 900$	$n = 1200$
$\widetilde{\mathbf{G}}_n(\boldsymbol{\nu}_3)$	$\begin{bmatrix} 13.64657 & -13.33241 \\ -13.33241 & 14.05080 \end{bmatrix}$	$\begin{bmatrix} 12.79163 & -12.98354 \\ -12.98354 & 14.20550 \end{bmatrix}$	$\begin{bmatrix} 11.36609 & -11.81191 \\ -11.81191 & 13.30482 \end{bmatrix}$
	$\widetilde{\mathbf{G}}(\boldsymbol{\nu}_3)$	$\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\nu}_3)$	
Limit Covariances	$\begin{bmatrix} 9.821792 & -10.16463 \\ -10.16463 & 11.50649 \end{bmatrix}$	$\begin{bmatrix} 9.823044 & -10.16911 \\ -10.16911 & 11.52254 \end{bmatrix}$	

indices here decrease as the number of vertices n increases. This is because the overall sparsity of the graphs increase as n increases, and the clustering accuracy fundamentally depends on the overall sparsity of the stochastic block model (see, for example, Zhang et al., 2016). We see that the one-step estimators (both for the latent positions and for the population LSE) outperform the spectral estimators (the ASE and the LSE) when n increases. The differences in the Rand indices are statistically significant at level $\alpha = 0.01$ for $n \geq 900$. In particular, when $n \geq 900$, the OSE-A and the OSE-L yield better results than the ASE and the LSE, respectively. These numerical results are in accordance with the fact that asymptotically, the ASE and the LSE are dominated by the OSE-A and OSE-L, respectively, even when the stochastic block model exhibits increasing sparsity level as the number of vertices increases.

I.4 Additional simulated examples

This subsection provides an additional example in addition to Subsection 5.1. We still consider a random dot product graph with one-dimensional latent positions and number of vertices being $n = 1000$. The latent positions $\mathbf{X}_0 = [x_{01}, \ldots, x_{0n}]^{\mathrm{T}}$ are generated as follows. We first generate latent variables $z_i = 0.5133\mathbb{1}\{i \leq 348\} + 0.95\mathbb{1}\{i > 348\} + w_i$, $i = 1, \ldots, n$, where $w_1, \ldots, w_n$ are independent $\mathrm{Unif}(-0.0125, 0.0125)$ random variables. Then a Matér Gaussian process regression model with the roughness parameter $5/2$ is applied to

Table 7: Rand indices for Subsection I.3: The Rand indices are obtained by applying the GMM-model-based clustering method to different estimates (the ASE, the OSE-A, the LSE, and the OSE-L) when the number of vertices n ranges in $\{500, 600, \ldots, 1200\}$. The Rand indices are averaged over 10000 Monte Carlo replicates of adjacency matrices, with the standard errors included in parentheses.

Estimates	ASE	OSE-A	LSE	OSE-L
$n = 500$	**0.99942** (1.3×10^{-5})	0.99709 (2.8×10^{-5})	0.99879 (1.9×10^{-5})	0.99708 (2.8×10^{-5})
$n = 600$	**0.99557** (3.3×10^{-5})	0.99551 (3.3×10^{-5})	0.99295 (4.5×10^{-5})	0.99548 (3.4×10^{-5})
$n = 700$	0.99011 (4.7×10^{-5})	**0.99015** (4.7×10^{-5})	0.98585 (6.0×10^{-5})	0.99008 (4.8×10^{-5})
$n = 800$	0.98467 (5.4×10^{-5})	**0.98470** (5.5×10^{-5})	0.97885 (7.0×10^{-5})	0.98464 (5.5×10^{-5})
$n = 900$	0.97939 (6.1×10^{-5})	0.97982 (5.9×10^{-5})	0.97258 (7.6×10^{-5})	**0.97990** (5.8×10^{-5})
$n = 1000$	0.97516 (6.7×10^{-5})	0.97565 (6.4×10^{-5})	0.96761 (8.7×10^{-5})	**0.97576** (6.4×10^{-5})
$n = 1100$	0.97138 (7.0×10^{-5})	0.97179 (6.8×10^{-5})	0.96249 (9.2×10^{-5})	**0.97192** (6.7×10^{-5})
$n = 1200$	0.96816 (6.7×10^{-5})	0.96848 (6.6×10^{-5})	0.95886 (8.7×10^{-5})	**0.96866** (6.3×10^{-5})

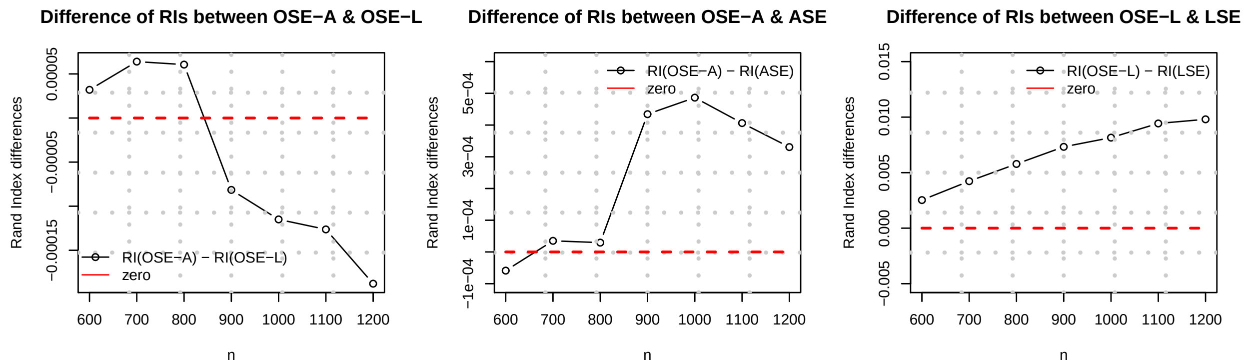


Figure 5: Differences of the Rand indices for Subsection I.3: The Rand indices are obtained by applying the GMM-model-based clustering method to different estimates (the ASE, the OSE-A, the LSE, and the OSE-L) when the number of vertices n ranges in $\{500, 600, \ldots, 1200\}$. The three panels are the plots of the differences of Rand indices between the OSE-A and the OSE-L, those between the OSE-A and the ASE, and those between the OSE-L and the LSE, respectively. The results are averaged based on 10000 Monte Carlo replicates.

the data points $(i/n, z_i)_{i=1}^n$ and we set x_{0i} to be the estimated regression curve evaluated at i/n, $i = 1, \ldots, n$. The specific values adopted above are carefully selected such that the difference between SSE_{ASE} and $SSE_{\text{OSE}-\text{A}}$ and the difference between SSE_{LSE} and $SSE_{\text{OSE}-\text{L}}$ are asymptotically maximized. The four estimates invovled are the ASE $\widehat{\mathbf{X}}^{(\text{ASE})}$, the OSE-A $\widehat{\mathbf{X}}$, the LSE $\breve{\mathbf{X}}$, and the OSE-L $\widehat{\mathbf{Y}}$. We generate 1000 independent adjacency matrices from the aforementioned random dot product graph as Monte Carlo replicates. For each realization of the adjacency matrix, we compute the sum-of-squares errors (SSEs) of the four estimates, namely, $SSE_{\text{ASE}} = \inf_{\mathbf{W} \in \{\pm 1\}} \|\widehat{\mathbf{X}}^{(\text{ASE})}\mathbf{W} - \mathbf{X}_0\|_2^2$, $SSE_{\text{OSE}-\text{A}} = \inf_{\mathbf{W} \in \{\pm 1\}} \|\widehat{\mathbf{X}}\mathbf{W} - \mathbf{X}_0\|_2^2$, $SSE_{\text{LSE}} = \inf_{\mathbf{W} \in \{\pm 1\}} \|\breve{\mathbf{X}}\mathbf{W} - \mathbf{Y}_0\|_2^2$, and $SSE_{\text{OSE}-\text{L}} = \inf_{\mathbf{W} \in \{\pm 1\}} \|\widehat{\mathbf{Y}}\mathbf{W} - \mathbf{Y}_0\|_2^2$. Figure 6 below visualizes the SSEs of the four estimates across 1000 Monte Carlo replicates. Specifically, in Figure 6, panels (a) and (b) visualize the boxplots of the SSEs, panels (c) and (d) illustrate the scatter points of these SSEs along with their respective limit values, and panels (e) and (f) show the scatter points of the differences of the SSEs,

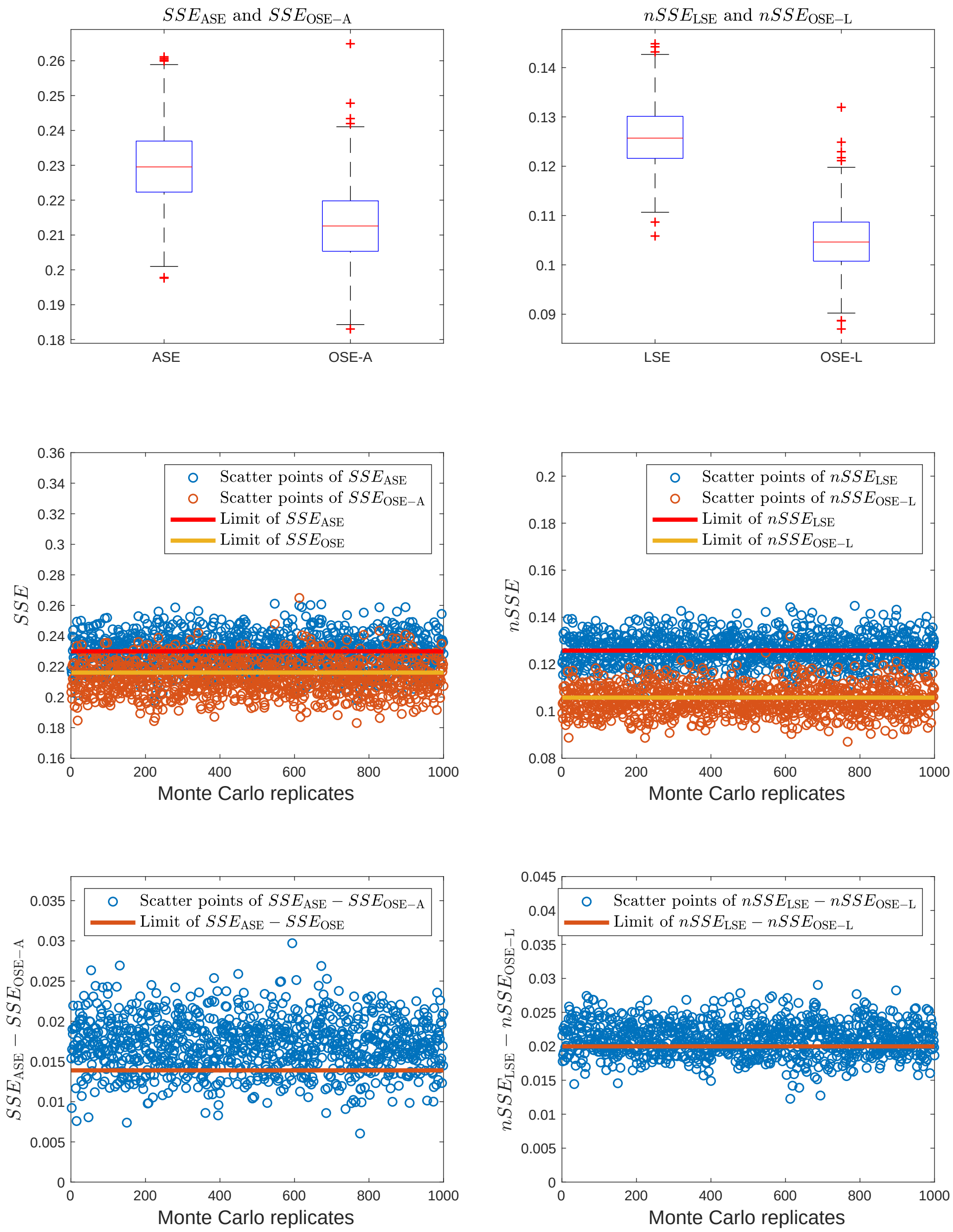


Figure 6: Visualization of the SSEs for Subsection I.4: Panels (a) and (b) are the boxplots of SSE_{ASE} versus $SSE_{\mathrm{OSE-A}}$ and $nSSE_{\mathrm{LSE}}$ versus $nSSE_{\mathrm{OSE-L}}$ across 1000 Monte Carlo replicates, respectively; Panels (c) and (d) are the scatter points of SSE_{ASE} versus $SSE_{\mathrm{OSE-A}}$ and $nSSE_{\mathrm{LSE}}$ versus $nSSE_{\mathrm{OSE-L}}$ across 1000 Monte Carlo replicates, where the solid lines represent the asymptotic values of these SSEs; Panels (e) and (f) are the scatter points of $SSE_{\mathrm{ASE}} - SSE_{\mathrm{OSE-A}}$ and $nSSE_{\mathrm{LSE}} - nSSE_{\mathrm{OSE-L}}$ across 1000 Monte Carlo replicates, respectively.

namely, $SSE_{\mathrm{ASE}} - SSE_{\mathrm{OSE-A}}$ and $nSSE_{\mathrm{LSE}} - nSSE_{\mathrm{OSE-L}}$, respectively. From Figure 6, it can be clearly seen that OSE-A and OSE-L provide substantial improvement over the ASE and the LSE, respectively, in terms of the SSEs.

References

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. *Journal of Machine Learning Research*, 18(226):1–92.

Athreya, A., Priebe, C. E., Tang, M., Lyzinski, V., Marchette, D. J., and Sussman, D. L. (2016). A limit theorem for scaled eigenvectors of random dot product graphs. *Sankhya A*, 78(1):1–18.

Bickel, P. J. and Doksum, K. A. (2015). *Mathematical statistics: basic ideas and selected topics*, volume 2. CRC Press.

Boucheron, S., Lugosi, G., and Massart, P. (2003). Concentration inequalities using the entropy method. *Ann. Probab.*, 31(3):1583–1614.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.

Cape, J., Tang, M., and Priebe, C. E. (2019). Signal-plus-noise matrix models: eigenvector deviations and fluctuations. *Biometrika*, 106(1):243–250.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, 23(4):493–507.

Chernoff, H. (1956). Large-sample theory: Parametric case. *The Annals of Mathematical Statistics*, 27(1):1–22.

Chung, K. L. (2001). *A course in probability theory*. Academic press.

Erdös, L., Knowles, A., Yau, H.-T., and Yin, J. (2013). Spectral statistics of erdős–rényi graphs i: Local semicircle law. *Ann. Probab.*, 41(3B):2279–2375.

Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458):611–631.

Fraley, C., Raftery, A. E., Murphy, T. B., and Scrucca, L. (2012). mclust version 4 for r: normal mixture modeling for model-based clustering, classification, and density estimation. Technical report, Technical report.

Hoffman, A. J. and Wielandt, H. W. (2003). The variation of the spectrum of a normal matrix. In *Selected Papers Of Alan J Hoffman: With Commentary*, pages 118–120. World Scientific.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer.

Leang, C. C. and Johnson, D. H. (1997). On the asymptotics of m-hypothesis bayesian detection. *IEEE Transactions on Information Theory*, 43(1):280–282.

Lei, J. and Rinaldo, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.*, 43(1):215–237.

Mao, X., Sarkar, P., and Chakrabarti, D. (2020). Estimating mixed memberships with sharp eigenvector deviations. *Journal of the American Statistical Association*, 0(0):1–13.

Marshall, A. W. and Olkin, I. (1990). Matrix versions of the cauchy and kantorovich inequalities. *aequationes mathematicae*, 40(1):89–93.

Oliveira, R. I. (2009). Concentration of the adjacency matrix and of the laplacian in random graphs with independent edges. *arXiv preprint arXiv:0911.0600*.

Pecaric, J. E., Puntanen, S., and Styan, G. P. (1996). Some further matrix extensions of the cauchy-schwarz and kantorovich inequalities, with some statistical applications. *Linear Algebra and its Applications*, 237-238:455 – 476. Linear Algebra and Statistics: In Celebration of C. R. Rao's 75th Birthday (September 10, 1995).

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850.

Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V., Park, Y., and Priebe, C. E. (2017a). A semiparametric two-sample hypothesis testing problem for random graphs. *Journal of Computational and Graphical Statistics*, 26(2):344–354.

Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V., and Priebe, C. E. (2017b). A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli*, 23(3):1599–1630.

Tang, M. and Priebe, C. E. (2018). Limit theorems for eigenvectors of the normalized Laplacian for random graphs. *Ann. Statist.*, 46(5):2360–2415.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Yu, Y., Wang, T., and Samworth, R. J. (2015). A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323.

Zhang, A. Y., Zhou, H. H., et al. (2016). Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, 44(5):2252–2280.