Community recovery: Recap

Definition 1. *Let n be a positive integer (the number of vertices), k be a positive integer (the number of communities), $p = (p_1, \ldots, p_k)$ be a probability vector on $[k] := \{1, \ldots, k\}$ (the prior on the k communities) and W be a $k \times k$ symmetric matrix with entries in $[0, 1]$ (the connectivity probabilities). The pair (X, G) is drawn under $\mathrm{SBM}(n, p, W)$ if X is an n-dimensional random vector with i.i.d. components distributed under p, and G is an n-vertex simple graph where vertices i and j are connected with probability W_{X_i, X_j}, independently of other pairs of vertices. We also define the community sets by $\Omega_i = \Omega_i(X) := \{v \in [n] : X_v = i\}, i \in [k]$.*

Definition 2. *(X, G) is drawn under $\mathrm{SSBM}(n, k, q_{\mathrm{in}}, q_{\mathrm{out}})$ if W takes value q_{in} on the diagonal and q_{out} off the diagonal, and if the community prior is $p = \{1/k\}^k$ in the Bernoulli model, and X is drawn uniformly at random with the constraints $|\{v \in [n] : X_v = i\}| = n/k$, n a multiple of k, in the uniform or strictly balanced model.*

Definition 3 (Agreement and normalized agreement)**.** *The agreement between two community vectors $x, y \in [k]^n$ is obtained by maximizing the common components between x and any relabelling of y, i.e.,*

$$A(x, y) = \max_{\pi \in S_k} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(x_i = \pi(y_i)), \tag{9}$$

Definition 4. *Let $(X, G) \sim \mathrm{SBM}(n, p, W)$. The following recovery requirements are solved if there exists an algorithm that takes G as an input and outputs $\hat{X} = \hat{X}(G)$ such that*

- **Exact recovery:** $\mathbb{P}\{A(X, \hat{X}) = 1\} = 1 - o(1)$,
- **Almost exact recovery:** $\mathbb{P}\{A(X, \hat{X}) = 1 - o(1)\} = 1 - o(1)$,

Theorem 3. *[ABH14, MNS14a] Exact recovery in $\mathrm{SSBM}(n, 2, a\log(n)/n, b\log(n)/n)$ is solvable and efficiently so if $|\sqrt{a} - \sqrt{b}| > \sqrt{2}$ and unsolvable if $|\sqrt{a} - \sqrt{b}| < \sqrt{2}$.*

Theorem 3. *[ABH14, MNS14a] Exact recovery in* $\mathrm{SSBM}(n, 2, a\log(n)/n, b\log(n)/n)$ *is solvable and efficiently so if* $|\sqrt{a} - \sqrt{b}| > \sqrt{2}$ *and unsolvable if* $|\sqrt{a} - \sqrt{b}| < \sqrt{2}$.

In the two-community case, denoting by N_{in} and N_{out} the number of edges inside and across the clusters respectively,

$$\mathbb{P}\{G = g | \Omega = \omega\} \propto \left(\frac{q_{\text{out}}(1 - q_{\text{in}})}{q_{\text{in}}(1 - q_{\text{out}})} \right)^{N_{out}} . \tag{19}$$

Assuming $q_{\text{in}} \geq q_{\text{out}}$, we have $\frac{q_{\text{out}}(1-q_{\text{in}})}{q_{\text{in}}(1-q_{\text{out}})} \leq 1$ and thus

MAP is equivalent to finding a min-bisection of G,

i.e., a balanced partition with the least number of crossing edges.

Spectral relaxations. Consider again the symmetric SBM with strictly balanced communities. Recall that MAP maximizes

$$\max_{\substack{x \in \{+1,-1\}^n \\ x^t 1^n = 0}} x^t A x, \tag{20}$$

since this counts the number of edges inside the clusters minus the number of edges across the clusters, which is equivalent to the min-bisection problem (the total number of edges being fixed by the graph). The general idea behind spectral methods is to

relax the integral constraint to an Euclidean constraint on real valued vectors. This leads to looking for a maximizer of

$$\max_{\substack{x \in \mathbb{R}^n : \|x\|_2^2 = n \\ x^t 1^n = 0}} x^t A x. \tag{21}$$

Without the constraint $x^t 1^n = 0$, the above maximization gives precisely the eigenvector corresponding to the largest eigenvalue of A. Note that $A1^n$ is the vector containing the degrees of each node in g, and when g is an instance of the symmetric SBM, this concentrates to the same value for each vertex, and 1^n is close to an eigenvector of A. Since A is real symmetric, this suggests that the constraint $x^t 1^n = 0$ leads the maximization (21) to focus on the eigenspace orthogonal to the first eigenvector, and thus to the eigenvector corresponding to the second largest eigenvalue. Thus it is reasonable to take the second largest eigenvector $\phi_2(A)$ of A and round it to obtain an efficient relaxation of MAP:

$$\hat{X}_{\text{spec}} = \begin{cases} 1 & \text{if } \phi_2(A) \geq 0, \\ 2 & \text{if } \phi_2(A) < 0. \end{cases} \tag{22}$$

Now we describe a spectral method. To simplify presentation, it is assumed without loss of generality that: $x_i^\star = 1$ for any $1 \le i \le n/2$, and $x_i^\star = -1$ for any $i > n/2$.

A starting point for the algorithm design is to examine the mean of the adjacency matrix, given as follows

$$\mathbb{E}[\boldsymbol{A}] = \begin{bmatrix} p\,\mathbf{1}_{n/2}\mathbf{1}_{n/2}^\top & q\,\mathbf{1}_{n/2}\mathbf{1}_{n/2}^\top \\ q\,\mathbf{1}_{n/2}\mathbf{1}_{n/2}^\top & p\,\mathbf{1}_{n/2}\mathbf{1}_{n/2}^\top \end{bmatrix} - p\boldsymbol{I}.$$

As revealed by the above calculation, the matrix constructed below

$$\boldsymbol{M} = \boldsymbol{A} - \frac{p+q}{2}\mathbf{1}_n\mathbf{1}_n^\top + p\boldsymbol{I} \tag{3.30}$$

exhibits an approximate rank-1 structure, in the sense that its mean

$$\boldsymbol{M}^\star := \mathbb{E}[\boldsymbol{M}] = \frac{p-q}{2}\begin{bmatrix} \mathbf{1}_{n/2} \\ -\mathbf{1}_{n/2} \end{bmatrix}\begin{bmatrix} \mathbf{1}_{n/2}^\top & -\mathbf{1}_{n/2}^\top \end{bmatrix} \tag{3.31}$$

is a rank-1 matrix. The leading eigenvalue of $\boldsymbol{M}^\star$ and its associated eigenvector are given respectively by

$$\lambda^\star := \frac{(p-q)n}{2}, \quad \text{and} \quad \boldsymbol{u}^\star := \frac{1}{\sqrt{n}} \begin{bmatrix} \mathbf{1}_{n/2} \\ -\mathbf{1}_{n/2} \end{bmatrix}. \tag{3.32}$$

Crucially, the eigenvector $\boldsymbol{u}^\star$ encapsulates the precise community structure we seek to recover: all positive entries of $\boldsymbol{u}^\star$ correspond to vertices from one community, while the remaining ones form another community.

Inspired by the above calculation, a candidate spectral clustering algorithm consists of eigendecomposition followed by entrywise rounding:

1. Compute the leading eigenvector $\boldsymbol{u}$ of $\boldsymbol{M}$ (constructed in (3.30));

2. Compute the estimate $\boldsymbol{x} = [x_i]_{1 \le i \le n}$ such that for any $1 \le i \le n$,

$$x_i = \mathsf{sgn}(u_i) = \begin{cases} 1, & \text{if } u_i > 0, \\ -1, & \text{if } u_i \le 0. \end{cases} \tag{3.33}$$

Theorem 3.8. Consider the setting in Section 3.4.1, and suppose that

$$p \gtrsim \frac{\log n}{n}, \qquad \text{and} \qquad \sqrt{\frac{p}{n}} = o(p-q). \tag{3.35}$$

With probability exceeding $1 - O(n^{-8})$, the spectral method achieves

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{x_i = x_i^\star\} = 1 - o(1), \quad \text{or} \quad \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{x_i = -x_i^\star\} = 1 - o(1).$$

Special case 1: $p - q \gg \sqrt{\log n}\,/\,n, \qquad \text{if } p \asymp (\log n)\,/\,n.$

Special case 2: $p - q \gg 1/\sqrt{n}, \qquad \text{if } p \asymp 1,$

Theorem 3.4. Consider a *symmetric* random matrix $\boldsymbol{X} = [X_{i,j}]_{1 \le i,j \le n}$ in $\mathbb{R}^{n \times n}$, whose entries are independently generated and obey

$$\mathbb{E}[X_{i,j}] = 0, \quad \text{and} \quad |X_{i,j}| \le B, \qquad 1 \le i, j \le n. \tag{3.6}$$

Define

$$\nu := \max_i \sum\nolimits_j \mathbb{E}[X_{i,j}^2]. \tag{3.7}$$

Then there exists some universal constant $c > 0$ such that for any $t \ge 0$,

$$\mathbb{P}\Big\{\|\boldsymbol{X}\| \ge 4\sqrt{\nu} + t\Big\} \le n \exp\Big(-\frac{t^2}{cB^2}\Big). \tag{3.8}$$

This result, which appeared in Bandeira and Van Handel (2016, Remark 3.13), can be established

- As a useful corollary, if we know *a priori* that $\mathbb{E}[X_{i,j}^2] \le \sigma^2$ for all $1 \le i, j \le n$, then Theorem 3.4 implies that

$$\|\boldsymbol{X}\| \le 4\sigma\sqrt{n} + \widetilde{c}B\sqrt{\log n} \tag{3.9}$$

with probability at least $1 - n^{-8}$ for some constant $\widetilde{c} > 0$. To see this, it suffices to set $\widetilde{c} = \sqrt{9c}$ and take $t = B\sqrt{9c \log n}$ in (3.8).

2.3.1 Setup and notation

Let $\boldsymbol{M}^\star$ and $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{E}$ be two $n \times n$ real symmetric matrices. We express the eigendecomposition of $\boldsymbol{M}^\star$ and $\boldsymbol{M}$ as follows

$$\boldsymbol{M}^\star = \sum_{i=1}^{n} \lambda_i^\star \boldsymbol{u}_i^\star \boldsymbol{u}_i^{\star\top} = \begin{bmatrix} \boldsymbol{U}^\star & \boldsymbol{U}_\perp^\star \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}^\star & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_\perp^\star \end{bmatrix} \begin{bmatrix} \boldsymbol{U}^{\star\top} \\ \boldsymbol{U}_\perp^{\star\top} \end{bmatrix}; \tag{2.11}$$

$$\boldsymbol{M} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{U}_\perp \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_\perp \end{bmatrix} \begin{bmatrix} \boldsymbol{U}^\top \\ \boldsymbol{U}_\perp^\top \end{bmatrix}. \tag{2.12}$$

Here, $\{\lambda_i\}$ (resp. $\{\lambda_i^\star\}$) denote the eigenvalues of $\boldsymbol{M}$ (resp. $\boldsymbol{M}^\star$), and $\boldsymbol{u}_i$ (resp. $\boldsymbol{u}_i^\star$) stands for the eigenvector associated with the eigenvalue λ_i (resp. $\lambda_i^\star$). Additionally, we take

$$\boldsymbol{U} \coloneqq [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r] \in \mathbb{R}^{n \times r}, \qquad \boldsymbol{U}_\perp \coloneqq [\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_n] \in \mathbb{R}^{n \times (n-r)},$$
$$\boldsymbol{\Lambda} \coloneqq \mathsf{diag}([\lambda_1, \cdots, \lambda_r]), \qquad \boldsymbol{\Lambda}_\perp \coloneqq \mathsf{diag}([\lambda_{r+1}, \cdots, \lambda_n]).$$

The matrices $\boldsymbol{U}^\star$, $\boldsymbol{U}_\perp^\star$, $\boldsymbol{\Lambda}^\star$, and $\boldsymbol{\Lambda}_\perp^\star$ are defined analogously.

$$\mathsf{dist}(\boldsymbol{U}, \boldsymbol{U}^\star) \coloneqq \min_{\boldsymbol{R} \in \mathcal{O}^{r \times r}} \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|; \tag{2.10a}$$

$$\mathsf{dist}_{\mathrm{F}}(\boldsymbol{U}, \boldsymbol{U}^\star) \coloneqq \min_{\boldsymbol{R} \in \mathcal{O}^{r \times r}} \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|_{\mathrm{F}}. \tag{2.10b}$$

Corollary 2.8. Consider the settings in Section 2.3.1. Suppose that $|\lambda_1^\star| \geq |\lambda_2^\star| \geq \cdots \geq |\lambda_r^\star| > |\lambda_{r+1}^\star| \geq \cdots \geq |\lambda_n^\star|$ and $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$ (i.e., the eigenvalues are sorted by their magnitudes). If $\|\boldsymbol{E}\| < (1 - 1/\sqrt{2})(|\lambda_r^\star| - |\lambda_{r+1}^\star|)$, then

$$\mathsf{dist}(\boldsymbol{U}, \boldsymbol{U}^\star) \leq \sqrt{2}\|\sin\boldsymbol{\Theta}\| \leq \frac{2\|\boldsymbol{E}\boldsymbol{U}^\star\|}{|\lambda_r^\star| - |\lambda_{r+1}^\star|} \leq \frac{2\|\boldsymbol{E}\|}{|\lambda_r^\star| - |\lambda_{r+1}^\star|}; \tag{2.18a}$$

$$\mathsf{dist}_{\mathrm{F}}(\boldsymbol{U}, \boldsymbol{U}^\star) \leq \sqrt{2}\|\sin\boldsymbol{\Theta}\|_{\mathrm{F}} \leq \frac{2\|\boldsymbol{E}\boldsymbol{U}^\star\|_{\mathrm{F}}}{|\lambda_r^\star| - |\lambda_{r+1}^\star|} \leq \frac{2\sqrt{r}\|\boldsymbol{E}\|}{|\lambda_r^\star| - |\lambda_{r+1}^\star|}. \tag{2.18b}$$

Theorem 4.6. Fix any constant $\varepsilon > 0$, and consider the setting of Section 3.4.1. Suppose $p = \frac{\alpha \log n}{n}$ and $q = \frac{\beta \log n}{n}$ for some sufficiently large constants $\alpha > \beta > 0$.[1] In addition, assume that

$$\left(\sqrt{p} - \sqrt{q}\right)^2 \geq 2\left(1 + \varepsilon\right) \frac{\log n}{n}. \tag{4.45}$$

With probability $1 - o(1)$, the spectral method in Section 3.4.2 yields

$$x_i = x_i^\star \quad \text{for all } 1 \leq i \leq n, \qquad \text{or} \qquad x_i = -x_i^\star \quad \text{for all } 1 \leq i \leq n.$$

Next, we turn to vector and matrix norms. For any vector $\boldsymbol{v}$, we denote by $\|\boldsymbol{v}\|_2$, $\|\boldsymbol{v}\|_1$ and $\|\boldsymbol{v}\|_\infty$ its ℓ_2 norm, ℓ_1 norm and ℓ_∞ norm, respectively. For any matrix $\boldsymbol{A} = [A_{i,j}]_{1\le i\le m, 1\le j\le n}$, we let $\|\boldsymbol{A}\|$, $\|\boldsymbol{A}\|_*$, $\|\boldsymbol{A}\|_{\mathrm{F}}$ and $\|\boldsymbol{A}\|_\infty$ represent respectively its spectral norm (i.e., the largest singular value of $\boldsymbol{A}$), its nuclear norm (i.e., the sum of singular values of $\boldsymbol{A}$), its Frobenius norm (i.e., $\|\boldsymbol{A}\|_{\mathrm{F}} \coloneqq \sqrt{\sum_{i,j} A_{i,j}^2}$), and its entrywise ℓ_∞ norm (i.e., $\|\boldsymbol{A}\|_\infty \coloneqq \max_{i,j} |A_{i,j}|$). We also refer to $\|\boldsymbol{A}\|_{2,\infty}$ as the $\ell_{2,\infty}$ norm of $\boldsymbol{A}$, defined as $\|\boldsymbol{A}\|_{2,\infty} \coloneqq \max_i \|\boldsymbol{A}_{i,\cdot}\|_2$. Similarly, we define the $\ell_{\infty,2}$ norm of $\boldsymbol{A}$ as $\|\boldsymbol{A}\|_{\infty,2} \coloneqq \|\boldsymbol{A}^\top\|_{2,\infty}$. In addition, for any matrices $\boldsymbol{A} = [A_{i,j}]_{1\le i\le m, 1\le j\le n}$ and $\boldsymbol{B} = [B_{i,j}]_{1\le i\le m, 1\le j\le n}$, the inner product of $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined as and denoted by $\langle \boldsymbol{A}, \boldsymbol{B}\rangle = \sum_{1\le i\le m, 1\le j\le n} A_{i,j}B_{i,j} = \mathsf{Tr}(\boldsymbol{A}^\top \boldsymbol{B})$.

Ground truth. Consider a rank-r symmetric matrix $\boldsymbol{M}^\star \in \mathbb{R}^{n\times n}$ with eigenvectors $\boldsymbol{u}_1^\star, \cdots, \boldsymbol{u}_n^\star$ and associated eigenvalues $\lambda_1^\star, \cdots, \lambda_n^\star$ obeying

$$|\lambda_1^\star| \geq |\lambda_2^\star| \geq \cdots \geq |\lambda_r^\star| > 0 \quad \text{and} \quad \lambda_{r+1}^\star = \cdots = \lambda_n^\star = 0. \tag{4.22}$$

We shall write the eigendecomposition $\boldsymbol{M}^\star = \boldsymbol{U}^\star \boldsymbol{\Lambda}^\star \boldsymbol{U}^{\star\top}$ as usual, where $\boldsymbol{\Lambda}^\star := \mathsf{diag}([\lambda_1^\star, \cdots, \lambda_r^\star])$ and $\boldsymbol{U}^\star := [\boldsymbol{u}_1^\star, \cdots, \boldsymbol{u}_r^\star] \in \mathbb{R}^{n\times r}$. Denote the condition number of $\boldsymbol{M}^\star$ as

$$\kappa := |\lambda_1^\star| \,/\, |\lambda_r^\star|. \tag{4.23}$$

Akin to Definition 3.1, the incoherence parameter of $\boldsymbol{M}^\star$ is defined as

$$\mu := \frac{n\|\boldsymbol{U}^\star\|_{2,\infty}^2}{r}, \tag{4.24}$$

a parameter that captures how well the energy of $\boldsymbol{U}^\star$ is spread out across all rows and that obeys (see Remark 3.12)

$$1 \leq \mu \leq n/r. \tag{4.25}$$

Observed data. What we observe is a corrupted version

$$\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{E} \in \mathbb{R}^{n\times n}, \tag{4.26}$$

where $\boldsymbol{E}$ is a symmetric noise matrix. We denote by $\{\lambda_i\}_{1\le i\le n}$ the set of eigenvalues of $\boldsymbol{M}$ obeying

$$|\lambda_1| \ge |\lambda_2| \ge \cdots \ge |\lambda_n|, \tag{4.27}$$

and let $\boldsymbol{u}_i$ be the eigenvector of $\boldsymbol{M}$ associated with λ_i. We shall introduce the diagonal matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{r\times r}$ as $\boldsymbol{\Lambda} \coloneqq \mathsf{diag}([\lambda_1, \cdots, \lambda_r])$.

Noise assumptions. This section aims to cover a fairly broad class of scenarios of independent noise. In particular, the noise matrix considered herein is assumed to satisfy the mild conditions listed below.

Assumption 4.1. The entries in the lower triangular part of $\boldsymbol{E} = [E_{i,j}]_{1\le i,j\le n}$ are independently generated obeying

$$\mathbb{E}[E_{i,j}] = 0, \quad \mathbb{E}[E_{i,j}^2] =: \sigma_{i,j}^2 \le \sigma^2, \quad |E_{i,j}| \le B, \quad \text{for all } i \ge j. \tag{4.28}$$

In particular, σ^2 is taken to be the smallest choice satisfying (4.28). Further, it is assumed that

$$c_{\mathsf{b}} \coloneqq \frac{B}{\sigma\sqrt{n/(\mu\log n)}} = O(1). \tag{4.29}$$

Goal and algorithm. We seek to estimate $\boldsymbol{U}^\star$ based on $\boldsymbol{M}$. Towards this, a simple spectral method computes the matrix $\boldsymbol{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r] \in \mathbb{R}^{n\times r}$ that comprises the top-r leading eigenvectors of $\boldsymbol{M}$.

4.2.2 $\ell_{2,\infty}$ and ℓ_∞ theoretical guarantees

The leave-one-out argument introduced before, when properly strengthened, enables powerful $\ell_{2,\infty}$ performance guarantees for the spectral estimate $\boldsymbol{U}$, which concern row-wise perturbation of the eigenspace. Before continuing, we remind the readers of the global rotation ambiguity, namely, in general we cannot expect $\boldsymbol{U}$ to be close to $\boldsymbol{U}^\star$ unless suitable global rotation is taken into account. In light of this, we introduce the following notation that helps identify a proper rotation matrix.

Definition 4.1. For any matrix $\boldsymbol{Z}$ with SVD $\boldsymbol{Z} = \boldsymbol{U}_Z\boldsymbol{\Sigma}_Z\boldsymbol{V}_Z^\top$ (where $\boldsymbol{U}_Z$ and $\boldsymbol{V}_Z$ represent respectively the left and right singular matrices of $\boldsymbol{Z}$, and $\boldsymbol{\Sigma}_Z$ is a diagonal matrix composed of the singular values), define

$$\mathsf{sgn}(\boldsymbol{Z}) \coloneqq \boldsymbol{U}_Z\boldsymbol{V}_Z^\top \tag{4.30}$$

to be the matrix sign function of $\boldsymbol{Z}$.

Remark 4.1. The matrix sign function is commonly encountered when aligning two matrices—classically known as the orthogonal Procrustes problem (Schönemann, 1966). Consider any two matrices $\widehat{\boldsymbol{B}}, \boldsymbol{B} \in \mathbb{R}^{n\times r}$ with $r \le n$. Among all rotation matrices, the one that best aligns $\widehat{\boldsymbol{B}}$ with $\boldsymbol{B}$ is precisely $\mathsf{sgn}(\widehat{\boldsymbol{B}}^\top\boldsymbol{B})$ (see, e.g., (Ma *et al.*, 2020, Appendix D.2.1)), namely,

$$\mathsf{sgn}(\widehat{\boldsymbol{B}}^\top\boldsymbol{B}) = \arg\min_{\boldsymbol{O}\in\mathcal{O}^{r\times r}} \|\widehat{\boldsymbol{B}}\boldsymbol{O} - \boldsymbol{B}\|_{\mathrm{F}}^2.$$

Theorem 4.2. Consider the settings and assumptions in Section 4.2.1. Define $\boldsymbol{H} \coloneqq \boldsymbol{U}^\top \boldsymbol{U}^\star$. With probability exceeding $1 - O(n^{-5})$, one has

$$\left\|\boldsymbol{U}\mathsf{sgn}(\boldsymbol{H}) - \boldsymbol{U}^\star\right\|_{2,\infty} \lesssim \frac{\sigma\kappa\sqrt{\mu r} + \sigma\sqrt{r\log n}}{|\lambda_r^\star|}, \tag{4.31a}$$

$$\begin{aligned}&\left\|\boldsymbol{U}\mathsf{sgn}(\boldsymbol{H}) - \boldsymbol{M}\boldsymbol{U}^\star\left(\boldsymbol{\Lambda}^\star\right)^{-1}\right\|_{2,\infty} \\ &\qquad \lesssim \frac{\sigma\kappa\sqrt{\mu r}}{|\lambda_r^\star|} + \frac{\sigma^2\sqrt{rn\log n} + \sigma B\sqrt{\mu r\log^3 n}}{\left(\lambda_r^\star\right)^2},\end{aligned} \tag{4.31b}$$

provided that $\sigma\sqrt{n\log n} \le c_\sigma|\lambda_r^\star|$ for some sufficiently small constant $c_\sigma > 0$.