Uncertainty quantification in the Bradley–Terry–Luce model

Chao Gao[†]

Department of Statistics, University of Chicago, Chicago, IL 60637, USA
[†]Corresponding author. Email: chaogao@uchicago.edu

Yandi Shen

Department of Statistics, University of Chicago, Chicago, IL 60637, USA

and

Anderson Y. Zhang

Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104, USA

The Bradley–Terry–Luce (BTL) model is a benchmark model for pairwise comparisons between individuals. Despite recent progress on the first-order asymptotics of several popular procedures, the understanding of uncertainty quantification in the BTL model remains largely incomplete, especially when the underlying comparison graph is sparse. In this paper, we fill this gap by focusing on two estimators that have received much recent attention: the maximum likelihood estimator (MLE) and the spectral estimator. Using a unified proof strategy, we derive sharp and uniform non-asymptotic expansions for both estimators in the sparsest possible regime (up to some poly-logarithmic factors) of the underlying comparison graph. These expansions allow us to obtain: (i) finite-dimensional central limit theorems for both estimators; (ii) construction of confidence intervals for individual ranks; (iii) optimal constant of ℓ_2 estimation, which is achieved by the MLE but not by the spectral estimator. Our proof is based on a self-consistent equation of the second-order remainder vector and a novel leave-two-out analysis.

Keywords: Bradley–Terry–Luce model; uncertainty quantification; central limit theorem; maximum likelihood estimator; spectral estimator.

1. Introduction

1.1 *Overview*

In this paper, we study the problem of uncertainty quantification in the Bradley–Terry–Luce (BTL) model. Given n individuals with unknown merit scores $\theta^* = (\theta_1^*, \ldots, \theta_n^*)^\top$ and an Erdoős–Rényi comparison graph $A = \{A_{ij}\}_{i<j}$ with success probability p, individuals i and j are compared L times if $A_{ij} = 1$, and each binary result $y_{ij\ell}$ is independently in favour of the former with probability

$$\mathbb{P}(y_{ij\ell} = 1) \equiv \frac{e^{\theta_i^*}}{e^{\theta_i^*} + e^{\theta_j^*}}.$$

Equipped with the realization of the comparison graph A and all the pairwise comparison results, the goal is to conduct inference on the unknown merit vector θ^*. We give a detailed review of the BTL model

in Section 2 ahead. Along with its generalizations [20, 21, 30] and close cousins in assortive networks [3, 15, 29], the BTL model has found extensive applications in web search [9, 10], competitive sports [23, 34] and social network analysis [25, 32].

Among many statistical procedures developed for the BTL model, the following two approaches have received particular attention in recent years due to their wide applicability in practice.

- (Maximum likelihood estimator) In the current formulation, the maximum likelihood estimator (MLE) leads to a convex program and can be solved by efficient algorithms developed in [12, 16, 18, 44].

- (Spectral estimator) The spectral method, also named *rank centrality* in its original discovery [24], associates the pairwise comparisons with a Markov chain on the underlying comparison graph. The estimator is then taken to be the stationary measure of the corresponding sample transition probability matrix, and hence can be solved efficiently via power iterations [24].

In the past few years, a series of works [4–6, 24] have studied in depth the theoretical properties of these two estimators, with focus on their ℓ_2/ℓ_∞ estimation accuracy and performance in partial/full ranking; see Section 2 ahead for a detailed review of related results.

Compared to the above thorough study of the first-order asymptotics of the estimators, the understanding of the associated limiting distribution theory is much less complete. A critical step in establishing optimal error bounds for partial recovery of ranking in [5] is sharp tail probability estimates for both the MLE and the spectral method. These results suggest, though do not directly imply, asymptotic normality. In the seminal work [36], asymptotic normality was first rigourously proved for the MLE when the comparison graph is fully connected (i.e. the underlying Erdoős–Rényi comparison graph satisfies $p = 1$). Using the same proof strategy, [14] later extended the above result to the regime $p \asymp n^{-1/10}$ (up to some poly-logarithmic factors). Since the Erdoős–Rényi graph is connected with high probability as soon as $np \geq (1 + \varepsilon) \log n$ for any $\varepsilon > 0$, this leaves open the most challenging but practically relevant regime $p \asymp n^{-1}$ (up to some poly-logarithmic factors), as many real world networks only have near constant degrees. The goal of this paper is to fill this gap by providing a sharp characterization of both estimators in this regime.

1.2 *Main contribution*

Using a unified proof strategy, our main theoretical results provide a non-asymptotic expansion of both estimators in the regime $p \asymp n^{-1}$ (up to some poly-logarithmic factors). Let us start with the MLE, which we will denote by $\widehat{\theta}$. Our first main result states that (see Theorem 2 ahead for the formal statement)

$$\widehat{\theta}_i - \theta_i^* \approx \frac{b_i}{d_i}, \tag{1.1}$$

where

$$b_i = \sum_{j:j\neq i} A_{ij}\big(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\big), \quad d_i = \sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*).$$

Here $\bar{y}_{ij} \equiv L^{-1}\sum_{\ell=1}^{L} y_{ij\ell}$ is the averaged comparison result between individuals i and j when $A_{ij} = 1$, $\psi(x) \equiv e^x/(1 + e^x)$ is the logistic function, and the approximation $\approx$ in (1.1) is uniform over the

coordinates $i \in [n]$. To the best of our knowledge, (1.1) as formally stated in Theorem 2 is the first result in the literature with an explicit non-asymptotic expansion in the sparse regime $p \asymp n^{-1}$ (up to some poly-logarithmic factors).

Thanks to the tractable form of the main terms $\{b_i/d_i\}$ (we will defer their interpretation to Section 2 ahead), we are able to characterize the asymptotic behavior of the MLE $\widehat{\theta}$. In particular, we will present three concrete applications of the expansion in (1.1): (i) a finite-dimensional central limit theorem (CLT) of $\widehat{\theta}$; (ii) construction of confidence intervals for individual ranks; (iii) optimal constant of ℓ_2 estimation in the BTL model; see Section 4 ahead for details.

Using the same proof strategy underlying (1.1), we are able to derive a similar expansion for the spectral estimator which we denote by $\widetilde{\theta}$ (see Theorem 3 ahead for the formal statement):

$$\widetilde{\theta}_i - \theta_i^* \approx \frac{\widetilde{b}_i}{\widetilde{d}_i}, \tag{1.2}$$

where

$$\widetilde{b}_i = \sum_{j:j \neq i} A_{ij}(\pi_i^* + \pi_j^*)\big(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\big), \quad \widetilde{d}_i = \pi_i^* \cdot \sum_{j:j \neq i} A_{ij}\psi(\theta_j^* - \theta_i^*).$$

Here $\pi^* = (\pi_1^*, \ldots, \pi_n^*)^\top$ is the stationary measure of the Markov chain associated with the spectral estimator (see Section 2.1 ahead for details), and the approximation $\approx$ in (1.2) is again uniform over the coordinates $i \in [n]$. To the best of our knowledge, (1.2) is the first result in the literature that gives a sharp non-asymptotic expansion of the spectral estimator in the sparse regime $p \asymp n^{-1}$ (up to some poly-logarithmic factors). Analogous to the MLE $\widehat{\theta}$, the expansion (1.2) also leads to a finite-dimensional CLT for the spectral estimator and its exact constant of ℓ_2 error. Interestingly, along with a lower bound in the local minimax framework (cf. Theorem 9 ahead), we are able to conclude that *the MLE achieves the optimal constant of ℓ_2 estimation in the BTL model but the spectral method does not*. An analogous observation in terms of ranking performance has previously been made in [5].

Let us now discuss briefly the proof strategy underlying expansions (1.1) and (1.2). The previous proof strategy adopted in [14, 36] for the MLE $\widehat{\theta}$ proceeds by building and then inverting a self-consistent equation of $\widehat{\theta}$, where the key technical step is to approximate the inverse of the Hessian of the likelihood equation in an entrywise manner. In the fully connected case $p = 1$, such a technical result is available from [35], but its approximation accuracy deteriorates quickly as p gets smaller, resulting in the final condition $n^{1/10}p \to \infty$ in [14]. In this paper we adopt a higher-order analogue of the above approach by first building a self-consistent equation over the remainder vector

$$\delta_i \equiv \widehat{\theta}_i - \frac{b_i}{d_i}$$

in the case of the MLE $\widehat{\theta}$. This approach allows us to bypass the technical difficulty of approximating the inverse of the Hessian, and reduces the problem to a sharp enough ℓ_∞ control of δ, which we then tackle with a leave-two-out analysis. As will be detailed in Section 3 ahead, leaving two out instead of one as in previous works [5, 6] is essential for our analysis. We refer to Section 3 for a detailed discussion of proof strategies.

In a broader context, this paper can be categorized under the general theme of uncertainty quantification/statistical inference for models with growing/infinite dimensions, see [7, 11, 17, 22, 26–28, 31, 37, 40, 41, 45] for an incomplete list of recent works in other benchmark statistical models. This paper, together with the prior works [14, 36], resolves uncertainty quantification in the BTL model in both sparse and dense regimes.

1.3 *Organization*

The rest of the paper is organized as follows. Section 2 starts with a review of the BTL model and then presents our main results. Section 3 discusses in detail our proof strategy. Section 4 develops three concrete applications of our main theoretical expansions. Proofs for the MLE and spectral method are given in Sections 5 and 6 respectively, followed by proofs for the applications in Sections 7–9. Some additional auxiliary results are presented in Appendix A.

1.4 *Notation*

For any positive integer n, let $[n]$ denote the set $\{1, \dots, n\}$. For $a, b \in \mathbb{R}$, $a \vee b \equiv \max\{a, b\}$ and $a \wedge b \equiv \min\{a, b\}$. For $a \in \mathbb{R}$, let $a_+ \equiv a \vee 0$ and $a_- \equiv (-a) \vee 0$. For $x \in \mathbb{R}^n$, let $\|x\|_q = \|x\|_{\ell_q(\mathbb{R}^n)}$ denote its q-norm ($0 < q \leq \infty$) with $\|x\|_2$ abbreviated as $\|x\|$. By $\mathbf{1}_n$ we denote the vector of all ones in $\mathbb{R}^n$. For a matrix $M \in \mathbb{R}^{n \times n}$, let $\|M\|_{\mathrm{op}}$ and $\|M\|_F$ denote the spectral and Frobenius norms of M respectively. Let $M^\dagger$ denote its pseudo-inverse. We use $\{e_j\}$ to denote the canonical basis, whose dimension should be clear from the context. We use $\mathbb{I}$ to denote the indicator function.

For two nonnegative sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$ or $a_n = \mathcal{O}(b_n)$ if $a_n \leq Cb_n$ for some absolute constant C. We write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. We write $a_n \ll b_n$ or $a_n = \mathfrak{o}(b_n)$ (respectively $a_n \gg b_n$) if $\lim_{n \to \infty}(a_n/b_n) = 0$ (respectively $\lim_{n \to \infty}(a_n/b_n) = \infty$). We follow the convention that $0/0 = 0$.

Let φ, Φ be the density and the cumulative distribution function of a standard normal random variable. For any $\alpha \in (0, 1)$, let z_α be the normal quantile defined by $\mathbb{P}(\mathcal{N}(0, 1) > z_\alpha) = \alpha$.

2. Main results

2.1 *BTL model: a review*

We start by laying down the details of our problem setting. Consider n individuals where each one is associated with some latent merit parameter $\theta_i^* \in \mathbb{R}$ for $i \in [n]$. Comparisons among these n individuals are then made on a random Erdős–Rényi graph $A = \{A_{ij}\}_{i<j}$ with edge probability p, i.e. $\{A_{ij}\}_{i<j}$ are i.i.d. Bernoulli variables with success probability p. For each connected edge $A_{ij} = 1$, L independent comparisons denoted by $\{y_{ij\ell}\}_{\ell=1}^L$ are made between individuals i and j, where $y_{ij\ell}$ are i.i.d. Bernoulli variables with success probability $\psi(\theta_i^* - \theta_j^*)$ with ψ being the logistic function $\psi(t) \equiv e^t/(1+e^t)$. We observe the comparison graph A and the comparison results $\{y_{ij\ell}\}$, and the goal is to conduct statistical inference of the merit parameters $\theta^* = (\theta_1^*, \dots, \theta_n^*)^\top$.

Throughout the paper, we will consider the following parameter space for θ^*.

$$\Theta(\kappa) \equiv \left\{ \theta^* \in \mathbb{R}^n : \max_{i \in [n]} \theta_i^* - \min_{i \in [n]} \theta_i^* \leq \kappa, \mathbf{1}_n^\top \theta^* = 0 \right\}. \tag{2.1}$$

Here κ is known as the *dynamic range*, and since the BTL model is only identifiable up to a global shift in the merit parameter θ^*, we make the centering condition $\mathbf{1}_n^\top \theta^* = 0$. In our theory below, n is taken to be the main asymptotic parameter, and all other problem parameters p, θ^*, κ, L are allowed to change with n. We will mostly be interested in the following regime:

$$\kappa = \mathcal{O}(1), \quad np \gg (\log n)^\alpha \text{ for some } \alpha \geq 1. \tag{2.2}$$

We make a few remarks on the above conditions.

1. Here $\kappa = \mathcal{O}(1)$ is assumed to facilitate theoretical analysis. Direct conveniences brought by this condition include that the comparison success probability $\psi(\theta_i^* - \theta_j^*)$ is bounded away from 0 and 1, and the Hessian of the log-likelihood function $\ell_n(\theta)$ (see (3.2) below) will be well-conditioned. This condition is commonly assumed in the existing literature [5, 6, 8].

2. It is well-known that the underlying Erdős–Rényi graph is connected with high probability when $np \geq (1 + \varepsilon) \log n$ and disconnected if $np \leq (1 - \varepsilon) \log n$ for any $\varepsilon > 0$. Therefore, the second condition allows the sparsest possible regime (without losing identifiability) of the graph up to some poly-logarithmic factors.

More discussions of these conditions will follow after the statement of the main results; see Remark 2 ahead.

In the BTL model, arguably the most natural procedure is the MLE, which can be traced back to [12, 44]. After some elementary algebra, the (normalized) negative log-likelihood function is given by

$$\ell_n(\theta) = \sum_{i<j} A_{ij} \left[\bar{y}_{ij} \log \frac{1}{\psi(\theta_i - \theta_j)} + \bar{y}_{ji} \log \frac{1}{\psi(\theta_j - \theta_i)} \right]. \tag{2.3}$$

Here for each $i < j$, we let $\bar{y}_{ij} \equiv \left(\sum_{\ell=1}^L y_{ij\ell} \right)/L$ and we take $\bar{y}_{ji} = 1 - \bar{y}_{ij}$ by convention. In accordance with the identifiability condition $\mathbf{1}_n^\top \theta^* = 0$, we define the MLE to be

$$\widehat{\theta} \in \operatorname{argmin}_{\theta: \mathbf{1}_n^\top \theta = 0} \ell_n(\theta). \tag{2.4}$$

Under conditions in (2.2) with any $\alpha \geq 1$, the MLE exists and is unique with probability $1 - \mathcal{O}(n^{-10})$ since with the prescribed probability, the above optimization can be constrained to the compact set $\{\theta : \|\theta - \theta^*\|_\infty \leq 5\}$ (cf. [5, Proposition 8.1]) and the likelihood is strongly convex on this set (cf. Lemma 15 ahead). In what follows, we work on the event where $\widehat{\theta}$ is well-defined.

Another popular approach, named *rank centrality* in its original discovery [24], associates the pairwise comparisons with a random walk on the underlying Erdős–Rényi graph. More precisely, consider a Markov chain with n states (corresponding to the n individuals), and let the sample transition matrix P be defined by

$$P_{ij} \equiv \begin{cases} \frac{1}{d} A_{ij} \bar{y}_{ji}, & i \neq j, \\ 1 - \frac{1}{d} \sum_{k: k \neq i} A_{ik} \bar{y}_{ki}, & i = j. \end{cases}$$

Here we take $d = 2np$ throughout so that with probability $1 - \mathcal{O}(n^{-10})$, the matrix P has non-negative entries. Its population version conditioning on A is

$$
P_{ij}^* \equiv \mathbb{E}(P_{ij}|A) = \begin{cases} \frac{1}{d} A_{ij} \psi(\theta_j^* - \theta_i^*), & i \neq j, \\ 1 - \frac{1}{d} \sum_{k:k\neq i} A_{ik} \psi(\theta_k^* - \theta_i^*), & i = j; \end{cases}
$$

hence, the imaginary random walker on the graph has a higher tendency of moving to an adjacent node (individual) with a larger merit parameter. By direct verification, this population transition matrix admits the stationary measure

$$
\pi^* \equiv \left(\frac{e^{\theta_1^*}}{\sum_{k=1}^n e^{\theta_k^*}}, \ldots, \frac{e^{\theta_n^*}}{\sum_{k=1}^n e^{\theta_k^*}} \right), \tag{2.5}
$$

and the approach of rank centrality estimates it using the stationary measure $\widehat{\pi}$ of P:

$$
\widehat{\pi}^\top P = \widehat{\pi}^\top. \tag{2.6}
$$

In accordance with the identifiability condition $\mathbf{1}_n^\top \theta^* = 0$, the associated estimator of θ^* is defined as

$$
\widetilde{\theta}_i \equiv \log \widehat{\pi}_i - \frac{1}{n} \sum_{k=1}^n \log \widehat{\pi}_k. \tag{2.7}
$$

Since the above rank centrality algorithm builds on a long list of spectral ranking methods (see [43] for a comprehensive survey), we will henceforth call it the spectral method. For the rest of the paper, we reserve the notation $\widehat{\theta}$ for the MLE and $\widetilde{\theta}$ for the spectral method.

In the last few years, a series of works have made significant contribution towards understanding the performances of both estimators, in terms of both ℓ_2 and ℓ_∞ accuracy.

PROPOSITION 1. [5, 6, 24] Suppose the conditions in (2.2) hold with $\alpha = 1$. Then, there exists some $C = C(\kappa) > 0$ such that the following hold with probability at least $1 - \mathcal{O}(n^{-10})$ uniformly over $\theta^* \in \Theta(\kappa)$.

$$
\text{(MLE)} \quad \|\widehat{\theta} - \theta^*\|^2 \leq \frac{C}{pL} \quad \text{and} \quad \|\widehat{\theta} - \theta^*\|_\infty^2 \leq \frac{C \log n}{npL},
$$

$$
\text{(spectral)} \quad \|\widetilde{\theta} - \theta^*\|^2 \leq \frac{C}{pL} \quad \text{and} \quad \|\widetilde{\theta} - \theta^*\|_\infty^2 \leq \frac{C \log n}{npL}.
$$

In fact, both error rates are known to be minimax rate optimal [6, 24] over $\Theta(\kappa)$; see also [5, Lemmas 11.1 and 11.3] for some more refined ℓ_∞ tail estimates. Other aspects of the estimators, notably their ranking performances, have also been studied in depth in [5, 6, 24].

2.2 *Non-asymptotic expansions*

Compared to the above progress in the first-order asymptotics, the understanding of limiting distribution theory in the BTL model is much less complete. In the seminal work [36], the authors derived the ℓ_∞

consistency and asymptotic normality of the MLE in the full comparison case $p = 1$. Adopting a similar proof technique, the recent work [14] extended [36] to the sparsity regime $n^{1/10}p \to \infty$ (up to some poly-logarithmic factors). This leaves open the most challenging but practically relevant sparsity regime (2.2) where the average number of comparisons is only slowly growing.

The following first main result of this paper provides a sharp non-asymptotic expansion of the MLE $\widehat{\theta}$ in the regime (2.2), from which asymptotic normality and a few other interesting corollaries can be readily deduced. We present its proof in Section 5.

THEOREM 2. Suppose that the conditions in (2.2) hold with $\alpha = 3/2$. Then for any $i \in [n]$, we have

$$\widehat{\theta}_i - \theta_i^* = (1 + \varepsilon_{1,i})\frac{b_i}{d_i} + \varepsilon_{2,i}. \tag{2.8}$$

Here $\varepsilon_1, \varepsilon_2$ are two vectors in $\mathbb{R}^n$ such that $\|\varepsilon_1\|_\infty = \mathfrak{o}(1)$ and $\|\varepsilon_2\|_\infty = \mathfrak{o}(1/\sqrt{npL})$ with probability $1 - \mathcal{O}(n^{-10})$, and $b, d \in \mathbb{R}^n$ are given by

$$b_i = \sum_{j:j\neq i} A_{ij}\big(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\big), \quad d_i = \sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*).$$

REMARK 1. Using standard concentration arguments, we have $b_i \asymp \sqrt{np/L}$ and $d_i \asymp np$ for each $i \in [n]$, so the main term satisfies $b_i/d_i \asymp 1/\sqrt{npL}$. Hence, upon proper normalization, ε_2 is a small term in an entrywise manner.

REMARK 2. We comment on two possible generalizations of the above theorem.

1. (*Heterogeneity*) A direct generalization of the current BTL model is to allow heterogeneity in the pairwise comparisons, that is, replace the Erdős–Rényi probability p by p_{ij} and the number of comparisons L by L_{ij} for each pair (i, j) of comparison. With these changes, the above theorem continues to hold upon changing the conditions in (2.2) to $\min_{i<j} p_{ij} \gg (\log n)^\alpha/n$ and $\max_{i<j} p_{ij}/\min_{i<j} p_{ij} \leq C$ for some universal $C > 0$, and the main term b_i/d_i to

$$\frac{\sum_{j:j\neq i} A_{ij}(y_{ij} - L_{ij}\psi(\theta_i^* - \theta_j^*))}{\sum_{j:j\neq i} A_{ij}L_{ij}\psi'(\theta_i^* - \theta_j^*)},$$

where y_{ij} is now $\sum_{\ell=1}^{L_{ij}} y_{ij\ell}$; see [1, Section 3.1] for some related probabilistic properties of the inhomogeneous Erdős-Rényi graph. The applications to be introduced in Section 4 ahead will also hold upon proper modifications.

2. (Sparsity condition) The slightly worse exponent in $np \gg (\log n)^{3/2}$ (instead of 1) results from the bound $\|\delta\|_\infty \lesssim \sqrt{(\log n)^3/(np)^2}\ \mathcal{O}(1/\sqrt{npL})$ in Lemma 6, which is derived from an earlier rougher estimate $\|\delta\|_\infty \lesssim \sqrt{(\log n)^1/(np)^0}\mathcal{O}(1/\sqrt{npL})$ in Lemma 5. Here δ, defined in (5.3) for the MLE and (6.3) for the spectral method, is the key remainder vector we are trying to bound; see the proof idea in Section 3 ahead for more details. At the cost of a much lengthier proof, we may be able to iterate the above arguments and improve the condition to $np \gg (\log n)^{(3+2k)/(2+2k)}$ after the kth iteration, reaching exponent $1+\eta$ after $\lceil 1/(2\eta) \rceil$ iterations. An interesting theoretical question is whether the expansion (2.8) still holds under the condition $np \gg \log n$, or even better,

under $np \geq C \log n$ with some large universal $C > 0$ as established for some first-order results in [5]. We leave the complete resolution of this problem to a future work.

To the best of our knowledge, Theorem 2 is the first result in the literature with an explicit non-asymptotic expansion in the regime (2.2). To help with its understanding, we now explain the form $\{b_i/d_i\}$ of the main term. For each fixed $i \in [n]$, let $\ell_n^{(i)}(\theta_i|\theta_{-i})$ be the local negative likelihood of the ith individual given the rest of the merit parameters θ_{-i}:

$$\ell_n^{(i)}(\theta_i|\theta_{-i}) = \sum_{j:j \neq i} A_{ij}\Big[\bar{y}_{ij} \log \frac{1}{\psi(\theta_i - \theta_j)} + \bar{y}_{ji} \log \frac{1}{\psi(\theta_j - \theta_i)}\Big].$$

It can be readily verified that: (i) $\widehat{\theta}_i$ is a minimizer of $\ell_n^{(i)}(\theta_i|\widehat{\theta}_{-i})$ and (ii) $b_i = -\partial_{\theta_i}\ell_n^{(i)}(\theta_i^*|\theta_{-i}^*)$ and $d_i = \partial_{\theta_i}^2 \ell_n^{(i)}(\theta_i^*|\theta_{-i}^*)$, so a local quadratic expansion of $\ell_n^{(i)}(\theta_i|\widehat{\theta}_{-i})$ around θ_i^* yields that

$$\widehat{\theta}_i \approx \mathrm{argmin}_{\theta_i}\Big[\ell_n^{(i)}(\theta_i^*|\widehat{\theta}_{-i}) + \partial_{\theta_i}\ell_n^{(i)}(\theta_i^*|\widehat{\theta}_{-i}) \cdot (\theta_i - \theta_i^*) + \frac{1}{2}\partial_{\theta_i}^2 \ell_n^{(i)}(\theta_i^*|\widehat{\theta}_{-i}) \cdot (\theta_i - \theta_i^*)^2\Big]$$

$$\approx \mathrm{argmin}_{\theta_i}\Big[\ell_n^{(i)}(\theta_i^*|\widehat{\theta}_{-i}) - b_i(\theta_i - \theta_i^*) + \frac{1}{2}d_i(\theta_i - \theta_i^*)^2\Big] = \theta_i^* + \frac{b_i}{d_i}. \tag{2.9}$$

The key task of Theorem 2 is to give a tight control of the above approximation error; see Section 5 ahead for details.

As we will elaborate in Section 3 ahead, the previous proof idea adopted in [14, 36] is no longer applicable in the sparse regime (2.2), and our Theorem 2 is based on a completely different proof technique. Interestingly, this new proof technique also allows us to derive the following expansion of the spectral estimator $\widetilde{\theta}$; see Section 6 for its proof.

THEOREM 3. Suppose that the conditions in (2.2) hold with $\alpha = 3/2$. Then for any $i \in [n]$, we have

$$\widetilde{\theta}_i - \theta_i^* = (1 + \widetilde{\varepsilon}_{1,i})\frac{\widetilde{b}_i}{\widetilde{d}_i} + \widetilde{\varepsilon}_{2,i}. \tag{2.10}$$

Here $\widetilde{\varepsilon}_1, \widetilde{\varepsilon}_2$ are two vectors in $\mathbb{R}^n$ such that $\|\widetilde{\varepsilon}_1\|_\infty = \mathfrak{o}(1)$ and $\|\widetilde{\varepsilon}_2\|_\infty = \mathfrak{o}(1/\sqrt{npL})$ with probability $1 - \mathcal{O}(n^{-10})$, and $\widetilde{b}, \widetilde{d} \in \mathbb{R}^n$ are given by

$$\widetilde{b}_i = \sum_{j:j \neq i} A_{ij}(\pi_i^* + \pi_j^*)\big(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\big), \quad \widetilde{d}_i = \pi_i^* \cdot \sum_{j:j \neq i} A_{ij}\psi(\theta_j^* - \theta_i^*).$$

Here π^* is the population stationary measure defined in (2.5).

We now explain the form $\{\widetilde{b}_i/\widetilde{d}_i\}$ of the main term. Using a first-order Taylor expansion, we have

$$\widetilde{\theta}_i - \theta_i^* = \log\Big(\frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} + 1\Big) - \frac{1}{n}\sum_{k=1}^n \log\Big(\frac{\widehat{\pi}_k - \pi_k^*}{\pi_k^*} + 1\Big) \approx \frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*}. \tag{2.11}$$

Then it follows from the definition of $\widehat{\pi}$ that

$$\widehat{\pi}_i = \frac{\sum_{j:j\neq i} A_{ij}\bar{y}_{ij}\widehat{\pi}_j}{\sum_{j:j\neq i} A_{ij}\bar{y}_{ji}} \approx \frac{\sum_{j:j\neq i} A_{ij}\bar{y}_{ij}\pi_j^*}{\sum_{j:j\neq i} A_{ij}\bar{y}_{ji}}. \tag{2.12}$$

Combining the above two displays, we have

$$\widetilde{\theta}_i - \theta_i^* \approx \frac{\sum_{j:j\neq i} A_{ij}(\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^*)}{\pi_i^* \cdot \sum_{j:j\neq i} A_{ij}\bar{y}_{ji}} \approx \frac{\sum_{j:j\neq i} A_{ij}(\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^*)}{\pi_i^* \cdot \sum_{j:j\neq i} A_{ij}\psi(\theta_j^* - \theta_i^*)} = \frac{\widetilde{b}_i}{\widetilde{d}_i},$$

upon noting that $\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^* = (\pi_i^* + \pi_j^*)(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))$. The key task of Theorem 3 is to give a tight control of the above approximation error; see Section 6 ahead for details.

Theorems 2 and 3 provide us with valuable information on the asymptotic behavior of the two estimators and allow us to compare the two procedures on a much more refined basis than existing literature [5, 6, 24]. In particular, we will show that the MLE $\widehat{\theta}$ achieves the optimal constant of ℓ_2 estimation in the BTL model but the spectral estimator $\widetilde{\theta}$ does not; see Section 4.3 for more details.

3. Proof Strategy

This section discusses the unified high level proof idea underlying our main theorems, focusing on its difference from the previous one adopted in [14, 36]; detailed proofs are given in Sections 5 and 6 ahead.

3.1 *Summary of previous proof strategy*

We start by briefly summarizing the proof outline of [36] when $p = 1$ and explaining the additional technical difficulties in the sparse regime (2.2). Differentiating the likelihood in (2.3) leads to the score equations:

$$\frac{a_i}{L} = \sum_{j:j\neq i} \frac{e^{\widehat{\theta}_i}}{e^{\widehat{\theta}_i} + e^{\widehat{\theta}_j}}, \quad i \in [n],$$

with its population version given by $\mathbb{E}a_i = L \cdot \sum_{j:j\neq i} e^{\theta_i^*}/(e^{\theta_i^*} + e^{\theta_j^*})$. Here $a_i = \sum_{j:j\neq i} y_{ij}$ is the number of victories of the ith individual. Let $\Delta u_i \equiv (e^{\widehat{\theta}_i} - e^{\theta_i^*})/e^{\theta_i^*}$, which can be seen as a proxy for $\widehat{\theta}_i - \theta_i^*$ using a first-order Taylor expansion. Then, some algebra yields that

$$\frac{e^{\widehat{\theta}_i}}{e^{\widehat{\theta}_i} + e^{\widehat{\theta}_j}} - \frac{e^{\theta_i^*}}{e^{\theta_i^*} + e^{\theta_j^*}} = \frac{e^{\theta_i^*} e^{\theta_j^*}(\Delta u_i - \Delta u_j)}{(e^{\theta_i^*} + e^{\theta_j^*})^2} \times \frac{e^{\theta_i^*} + e^{\theta_j^*}}{e^{\widehat{\theta}_i} + e^{\widehat{\theta}_j}}$$

$$\approx \psi'(\theta_i^* - \theta_j^*)(\Delta u_i - \Delta u_j).$$

Combining the above two displays yields, in matrix form, that

$$a - \mathbb{E}a \approx H(\theta^*)\Delta u \approx H(\theta^*)(\widehat{\theta} - \theta^*), \tag{3.1}$$

where $a = (a_1, \ldots, a_n)^\top$, $\Delta u = (\Delta u_1, \ldots, \Delta u_n)^\top$, and $H(\theta^*) \in \mathbb{R}^{n \times n}$ is the Hessian given by

$$H(\theta^*) \equiv \frac{\partial^2 \ell_n(\theta)}{\partial \theta^2}\Big|_{\theta=\theta^*} = \Big(\sum_{k:k\neq i} A_{ik} \psi'(\theta_i^* - \theta_k^*) \mathbb{I}_{i=j} - A_{ij} \psi'(\theta_i^* - \theta_j^*) \mathbb{I}_{i\neq j} \Big)_{i,j}, \tag{3.2}$$

Given that the sample-mean type quantities $\{a_i - \mathbb{E}a_i\}$ are easy to analyze, it is natural to invert the Hessian $H(\theta^*)$ and find a close approximation S of ideally simple form such that (note that $H(\theta^*)$ is singular along $\mathbf{1}_n$ so we take the pseudo-inverse)

$$\widehat{\theta} - \theta^* \approx (H(\theta^*))^\dagger (a - \mathbb{E}a) \approx S(a - \mathbb{E}a),$$

an expansion similar to our (2.8). The construction of such an S is the main technical ingredient of [36], whose entrywise closeness to $(H(\theta^*))^\dagger$ at the order of n^{-2} is highly non-trivial to verify even in the case $p = 1$ [35]. For general p, this approximation error deteriorates to the order $(n^2 p^3)^{-1}$ as reported by [14, Lemma 7], which eventually leads to the sub-optimal condition $n^{1/10} p \to \infty$.

   We also mention a recent result by [19] that achieves the sparsity level $p \asymp \frac{\log n}{n}$ via a debiasing technique. However, their method requires L to be of greater order than $(n \log n)^2$, while our result and the results of [14, 35] are all valid even for $L = 1$.

3.2   *New proof strategy: remainder expansion and leave-two-out technique*

To bypass the above difficulty of approximating $H^\dagger(\theta^*)$, we take the following three-step approach. To avoid redundancy, we elaborate on the MLE $\widehat{\theta}$ and briefly discuss the spectral estimator $\widetilde{\theta}$.

  1. With the approximation remainder vector defined by

$$\delta_i \equiv \widehat{\theta}_i - \theta_i^* - \frac{b_i}{d_i},$$

     instead of directly expanding the target $\widehat{\theta} - \theta^*$ as in (3.1), we seek to construct a self-consistent equation for the remainder vector δ.

  2. From the above self-consistent equation of δ, derive a sharp bound for $\|\delta\|$.

  3. Starting from the above bound for $\|\delta\|$ and a rough bound on $\|\delta\|_\infty$, we derive a sharp bound for $\|\delta\|_\infty$ in the regime (2.2). This step leverages a leave-two-out technique, which is inspired by, and improves substantially upon, the leave-one-out technique in [5, 6].

   We now discuss these three steps in more detail.

3.2.1   *Step 1: self-consistent equation of δ*   By the definitions of b_i and d_i (in Theorem 2), we have

$$\delta_i = \widehat{\theta}_i - \theta_i^* - \frac{\sum_{j:j\neq i} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))}{\sum_{j:j\neq i} A_{ij} \psi'(\theta_i^* - \theta_j^*)}. \tag{3.3}$$

On the other hand, a local quadratic expansion of the likelihood (see (2.9) and recall $\ell_n^{(i)}(\theta_i|\theta_{-i})$ defined therein) yields that

$$\widehat{\theta}_i - \theta_i^* \approx -\frac{\partial_{\theta_i}\ell_n^{(i)}(\theta_i^*|\widehat{\theta}_{-i})}{\partial_{\theta_i}^2\ell_n^{(i)}(\theta_i^*|\theta_{-i}^*)} \approx \frac{\sum_{j:j\neq i}A_{ij}\big(\bar{y}_{ij} - \psi(\theta_i^* - \widehat{\theta}_j)\big)}{\sum_{j:j\neq i}A_{ij}\psi'(\theta_i^* - \theta_j^*)}.$$

Combining the above two displays and Taylor expanding once more in the numerator yields that

$$\delta_i \approx \frac{\sum_{j:j\neq i}A_{ij}\big(\psi(\theta_i^* - \theta_j^*) - \psi(\theta_i^* - \widehat{\theta}_j)\big)}{\sum_{j:j\neq i}A_{ij}\psi'(\theta_i^* - \theta_j^*)}$$

$$\approx \frac{\sum_{j:j\neq i}A_{ij}\psi'(\theta_i^* - \theta_j^*)(\widehat{\theta}_j - \theta_j^*)}{\sum_{j:j\neq i}A_{ij}\psi'(\theta_i^* - \theta_j^*)}.$$

To make δ also appear on the right side, we plug in $\widehat{\theta}_j - \theta_j^* = \delta_j + (b_j/d_j)$ for each $j \neq i$ in the above display, which yields the following self-consistent equation of δ:

$$d_i \cdot \delta_i \approx \sum_{j:j\neq i}A_{ij}\psi'(\theta_i^* - \theta_j^*)\delta_j + \sum_{j:j\neq i}A_{ij}\psi'(\theta_i^* - \theta_j^*)\frac{b_j}{d_j}. \tag{3.4}$$

In the next two steps, we will use this equation to obtain sharp bounds of $\|\delta\|$ and $\|\delta\|_\infty$.

The derivation for the spectral estimator $\widetilde{\theta}$ is analogous. The approximation remainder δ is now defined as

$$\delta_i \equiv \frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} - \frac{1}{n}\sum_{k=1}^{n}\frac{\widehat{\pi}_k - \pi_k^*}{\pi_k^*} - \frac{\widetilde{b}_i}{\widetilde{d}_i'},$$

where $\widetilde{d}_i' \equiv \pi_i^* \cdot \sum_{j:j\neq i}A_{ij}\bar{y}_{ij}$ is the sample version of $\widetilde{d}_i$. Due to the preliminary expansion in (2.11), we have $\delta_i \approx \widetilde{\theta}_i - \theta_i^* - (\widetilde{b}_i/\widetilde{d}_i)$. On the other hand, the definition of $\widehat{\pi}$ (see (2.12)) yields that

$$\frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} = \frac{\sum_{j:j\neq i}A_{ij}\big(\bar{y}_{ij}\widehat{\pi}_j - \bar{y}_{ji}\pi_i^*\big)}{\pi_i^* \cdot \sum_{j:j\neq i}A_{ij}\bar{y}_{ji}} = \frac{\sum_{j:j\neq i}A_{ij}\bar{y}_{ij}(\widehat{\pi}_j - \pi_j^*)}{\pi_i^* \cdot \sum_{j:j\neq i}A_{ij}\bar{y}_{ji}} + \frac{\widetilde{b}_i}{\widetilde{d}_i'}.$$

Combining the above two displays (and ignoring the small term $n^{-1}\sum_{k=1}^{n}(\widehat{\pi}_k - \pi_k^*)/\pi_k^*$) yields that

$$\delta_i \approx \frac{\sum_{j:j\neq i}A_{ij}\bar{y}_{ij}(\widehat{\pi}_j - \pi_j^*)}{\pi_i^* \cdot \sum_{j:j\neq i}A_{ij}\bar{y}_{ji}}.$$

By plugging in $\widehat{\pi}_j - \pi_j^* \approx \pi_j^*\big(\delta_j + (\widetilde{b}_j/\widetilde{d}_j)\big)$, we arrive at the following self-consistent equation for the spectral estimator:

$$\widetilde{d}_i \cdot \delta_i \approx \sum_{j:j\neq i} A_{ij}\bar{y}_{ij}\pi_j^* \cdot \delta_j + \sum_{j:j\neq i} A_{ij}\bar{y}_{ij}\pi_j^* \frac{\widetilde{b}_j}{\widetilde{d}_j}. \tag{3.5}$$

3.2.2 *Step 2: ℓ_2 control of δ by leave-one-out analysis* This is an essential intermediate step towards the bound of $\|\delta\|_\infty$ in the next step. By rearranging the self-consistent equation in (3.4), we have in matrix form

$$H(\theta^*)\delta \approx R, \tag{3.6}$$

where $H(\theta^*)$ is the Hessian in (3.2) and $R \in \mathbb{R}^n$ is some cumulated approximation error. This equation, which expands δ instead of $\widehat{\theta}$ directly, can be seen as a higher-order analogue of (3.1). The bound $\|\delta\| = \mathfrak{o}(1/\sqrt{pL})$ (cf. Proposition 11) is now readily obtained by (i) the bound of $\|R\|$ by preliminary estimates in Proposition 1 obtained via a leave-one-out analysis and (ii) the eigenvalue bound of $H(\theta^*)$ in Lemma 15, i.e. with probability $1 - \mathscr{O}(n^{-10})$,

$$np \lesssim \lambda_{\min,\perp}\big(H(\theta^*)\big) \lesssim \lambda_{\max}\big(H(\theta^*)\big) \lesssim np;$$

see Section 5.2 for details. Here $\lambda_{\min,\perp}\big(H(\theta^*)\big) = \min_{x\in\mathbb{R}^n:\|x\|=1,x^\top\mathbf{1}_n=0} x^\top H(\theta^*)x$ is the smallest eigenvalue of $H(\theta^*)$ orthogonal to the direction $\mathbf{1}_n$.

The derivation for the spectral estimator $\widetilde{\theta}$ is analogous: by rearranging (3.5), we have

$$L\delta \approx \widetilde{R},$$

where L is defined by $L_{ij} \equiv -A_{ij}\bar{y}_{ij}\pi_j^*$ for $i \neq j$ and $L_{ii} \equiv \sum_{j:j\neq i} A_{ij}\bar{y}_{ji}\pi_i^*$, and $\widetilde{R} \in \mathbb{R}^n$ is some cumulated approximation error. Note that even though L is not symmetric, its population version $\mathbb{E}(L|A)$ is a symmetric Laplacian matrix similar to $H(\theta^*)$; hence, the bound for $\|\delta\|$ can be similarly obtained as above.

3.2.3 *Step 3: ℓ_∞ control of δ by leave-two-out analysis* For the more refined ℓ_∞ bound of δ, we need to bound directly the right side of (3.4). For the terms therein, the term $\sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)(b_j/d_j)$ (and other approximation error terms) can be readily bounded using concentration arguments, and the most difficult term is

$$\sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)\delta_j.$$

If $\{A_{ij}\}_{j:j\neq i}$ was independent of $\{\delta_j\}_{j:j\neq i}$, then the above term could be bounded by standard concentration arguments along with the bound on $\|\delta\|$ from the previous step. To decorrelate these two terms, we will define a proxy $\delta_j^{(i)}$ of each δ_j by leaving out the ith observation, so that $\delta_j^{(i)}$ is independent of $\{A_{ij}\}_{j:j\neq i}$ while being close to δ_j; see Section 5.3.1 for details. The definition and analysis of $\delta_j^{(i)}$ further requires the analysis of a leave-two-out version of $\widehat{\theta}$, which is essential for our purpose and

fundamentally different from the bound given by the original leave-one-out analysis used in [5, 6]. Indeed, as commented in Remark 2, a leave-one-out analysis as in [5, 6] will only yield the rough bound

$$\|\delta\|_\infty \lesssim \sqrt{\frac{(\log n)^1}{(np)^0}} \mathcal{O}(\frac{1}{\sqrt{npL}}), \tag{3.7}$$

which is not useful (i.e. of the order $\mathfrak{o}(1/\sqrt{npL})$) even for the case $p = 1$. Starting from the second stage of this hierarchy (i.e. leave-two-out analysis), this bound is improved to $\sqrt{(\log n)^3/(np)^2}\mathcal{O}(1/\sqrt{npL})$ (cf. Proposition 10), which is accurate enough under the sparsity condition $np \gg (\log n)^{1.5}$. We believe a higher-order analysis (i.e. leave-k-out for a large but fixed k) will yield a better (but fixed) exponent closer to 1, but we will not pursue this direction here to avoid digression.

To summarize, our new proof technique adds two novel technical components to the existing theoretical analysis in [5, 6, 24, 35]: (i) a self-consistent equation for the remainder vector δ of first-order approximation; (ii) a leave-two-out technique that plays an essential role in the bound of $\|\delta\|_\infty$. The first component allows us to bypass the technical difficulty of approximating the inverse Hessian as in [35] in the case of vanishing p, which leads to a sub-optimal condition as shown by an earlier analysis. The second component allows us to improve over the rough bound (3.7), which is not needed for rate-optimal first-order bounds [5, 6, 24]. The rest of the proof combines concentration inequalities and the rate-optimal bounds in Proposition 1; we refer to Sections 5 and 6 for proof details of the MLE and spectral estimator, respectively.

4. Applications of the main expansions

In this section, we present three applications of the expansions in Theorems 2 and 3. Section 4.1 is devoted to a finite-dimensional CLT for the MLE $\widehat{\theta}$ and spectral estimator $\widetilde{\theta}$, followed by a construction of confidence intervals for the rank vector in Section 4.2. Lastly, we discuss the implication for the optimal constant in ℓ_2-estimation in Section 4.3. Some simulation results are presented in Section 4.4.

4.1 *Application I: Finite-dimensional CLT*

A direct consequence of the expansions in Theorems 2 and 3 is the following CLTs for any finite k-dimensional components of the estimators. We start with the MLE $\widehat{\theta}$. Without loss of generality, we state it for the first k components.

PROPOSITION 4. Let $\widehat{\theta}$ be the MLE in (2.4) and suppose the conditions of Theorem 2 hold. Then for any $\theta^* \in \Theta(\kappa)$ and fixed $k \in [n]$, we have

$$\Big(\rho_1(\bar{\theta})(\widehat{\theta}_1 - \theta_1^*), \ldots, \rho_k(\bar{\theta})(\widehat{\theta}_k - \theta_k^*)\Big) \rightsquigarrow \mathcal{N}_k(0, I_k) \tag{4.1}$$

for both $\bar{\theta} \in \{\widehat{\theta}, \theta^*\}$. Here the sequence $\{\rho_i(\theta)\}_{i=1}^n$ can be either

$$\rho_i(\theta) = \sqrt{L \cdot \sum_{j:j\neq i} A_{ij}\psi'(\theta_i - \theta_j)} \quad \text{or} \quad \sqrt{pL \cdot \sum_{j:j\neq i} \psi'(\theta_i - \theta_j)}.$$

The proof of the above result is presented in Section 7. Using the completely data-driven version of $\rho_i(\widehat{\theta})$, the above result can be used to produce multiple and simultaneous confidence intervals for finite-dimensional contrasts of the truth vector θ^*.

We note that previous works [14, 36] that consider the asymptotic normality of the MLE are based on a different identifiability condition. Namely, they assume that there are $n + 1$ individuals with merit parameters $(\theta_0^*, \ldots, \theta_n^*)$, θ_0^* and $\widehat{\theta}_0$ are set to be 0 by convention, and asymptotics is studied on the rest of the coordinates $(\widehat{\theta}_1, \ldots, \widehat{\theta}_n)$. As we will now show, our CLT in Proposition 4 can be naturally transformed to be applicable in their setting as well. Therefore, our result is a strict improvement over previous results in the regime (2.2).

Let $(\widehat{\theta}_0, \ldots, \widehat{\theta}_n) \in \operatorname{argmin}_{\theta \in \mathbb{R}^{n+1} : \mathbf{1}_{n+1}^\top \theta = 0} \ell_{n+1}(\theta)$ be the MLE under the truth $\mathbf{1}_n^\top (\theta_0^*, \ldots, \theta_n^*) = 0$, where $\ell_{n+1}(\theta)$ is the variant of (2.3) for $n + 1$ individuals. By Proposition 4, we have, heuristically speaking, $(\widehat{\theta}_0 - \theta_0^*, \ldots, \widehat{\theta}_k - \theta_k^*) \rightsquigarrow \mathscr{N}_{k+1}(0, V_{k+1}^{-1}(\theta^*))$, where

$$V_{k+1}(\theta^*) = L \cdot \operatorname{diag}\Big(\sum_{\substack{\ell : 0 \le \ell \le n \\ \ell \neq 0}} A_{0\ell} \psi'(\theta_0^* - \theta_\ell^*), \ldots, \sum_{\substack{\ell : 0 \le \ell \le n \\ \ell \neq k}} A_{k\ell} \psi'(\theta_k^* - \theta_\ell^*) \Big);$$

note that this representation is not completely rigourous since $V_{k+1}(\theta^*)$ depends on n. Now under the identifiability condition ($\theta_0^* = 0$) in [36], the MLE for $(\theta_1^*, \ldots, \theta_k^*)$ becomes

$$\widehat{\theta}^{\mathrm{SY}} \equiv (\widehat{\theta}_1 - \widehat{\theta}_0, \ldots, \widehat{\theta}_k - \widehat{\theta}_0),$$

where 'SY' stands for Simons–Yao. Then we have $(\widehat{\theta}_1^{\mathrm{SY}} - \theta_1^*, \ldots, \widehat{\theta}_k^{\mathrm{SY}} - \theta_k^*) \rightsquigarrow \mathscr{N}_k(0, (V^{\mathrm{SY}}(\theta^*))^{-1})$, with $(V^{\mathrm{SY}}(\theta^*))^{-1} \in \mathbb{R}^{k \times k}$ given by

$$(V^{\mathrm{SY}}(\theta^*))_{ii}^{-1} = \frac{1}{L \cdot \sum_{\ell : 0 \le \ell \le n, \ell \neq i} A_{i\ell} \psi'(\theta_i^* - \theta_\ell^*)} + \frac{1}{L \cdot \sum_{1 \le \ell \le n} A_{0\ell} \psi'(\theta_\ell^*)},$$

$$(V^{\mathrm{SY}}(\theta^*))_{ij}^{-1} = \frac{1}{L \cdot \sum_{1 \le \ell \le n} A_{0\ell} \psi'(\theta_\ell^*)} \qquad \text{if } i \neq j.$$

In the special case $p = 1$, this recovers [36, Theorem 2].

Similar to Proposition 4, we have the following finite-dimensional CLT for the spectral estimator $\widetilde{\theta}$.

PROPOSITION 5. Let $\widetilde{\theta}$ be the spectral estimator $\widetilde{\theta}$ defined in (2.7) and suppose the conditions of Theorem 3 hold. Then for any $\theta^* \in \Theta(\kappa)$ and fixed $k \in [n]$, we have

$$\left(\bar{\rho}_1(\bar{\theta})(\widetilde{\theta}_1 - \theta_1^*), \ldots, \bar{\rho}_k(\bar{\theta})(\widetilde{\theta}_k - \theta_k^*) \right) \rightsquigarrow \mathscr{N}_k(0, I_k) \tag{4.2}$$

for both $\bar{\theta} \in \{\widetilde{\theta}, \theta^*\}$. Here the sequence $\{\bar{\rho}_i(\theta)\}_{i=1}^n$ can be either

$$\bar{\rho}_i(\theta) = \sqrt{L \cdot \frac{\left(\sum_{j : j \neq i} A_{ij}(e^{\theta_i} + e^{\theta_j}) \psi'(\theta_i - \theta_j) \right)^2}{\sum_{j : j \neq i} A_{ij}(e^{\theta_i} + e^{\theta_j})^2 \psi'(\theta_i - \theta_j)}}}$$

or its deterministic version with all $\{A_{ij}\}_{i \neq j}$ replaced by p.

A direct application of Cauchy–Schwarz yields that

$$\bar{\rho}_i(\theta) \leq \rho_i(\theta),$$

and hence, the spectral estimator has a larger asymptotic variance than the MLE. We will rigorously justify the optimality of the MLE using the local minimax framework in Section 4.3 ahead.

4.2 *Application II: Confidence regions in ranking*

Consider the BTL model in the context of sports analytics, where after observing the outcome of a tournament, one wishes to construct a confidence interval for the rank of her team of interest. More concretely, let $r^{-1}(1), \ldots, r^{-1}(n)$ be the indices of the teams with the highest merit parameter, the second highest merit parameter and so on, so that

$$\theta^*_{r^{-1}(1)} \geq \theta^*_{r^{-1}(2)} \geq \cdots \geq \theta^*_{r^{-1}(n)},$$

where ties are broken arbitrarily. Correspondingly, $\{r(1), \ldots, r(n)\}$ is the rank vector of teams 1 to n, and we are interested in building a (discrete) confidence interval for the rank of a pre-fixed team. Without loss of generality, we choose the first team with merit parameter θ^*_1 (not necessarily the largest one).

Since rank is a global object, we need to construct confidence intervals for all n merit parameters, instead of just for θ^*_1. To this end, a straightforward idea is to use the ℓ_∞ bound in Proposition 1. This approach has two disadvantages: (i) the constant C therein is implicit and could potentially be large after complicated steps of analysis; (ii) the confidence intervals for all teams will have the same length, which is not ideal since certain teams will have more comparisons than others so their confidence intervals should heuristically be shorter.

We now describe a different procedure based on the non-asymptotic expansions in Section 2. The key is to exploit the explicit main terms $\{b_i/d_i\}$ therein to obtain short confidence intervals with data-driven lengths. With a fixed confidence level $1 - \alpha$, we will construct confidence intervals such that the target θ^*_1 belongs to its interval with asymptotical probability $1 - \alpha$, and all other teams belong to their respective (slightly more conservative) intervals with overwhelming probability. For θ^*_1, let

$$\mathscr{C}_1 \equiv \left[\widehat{\theta}_1 - \rho_1(\widehat{\theta})z_{1-\alpha/2}, \widehat{\theta}_1 + \rho_1(\widehat{\theta})z_{1-\alpha/2}\right],$$

where $z_{1-\alpha/2}$ is the normal quantile such that $\mathbb{P}(\mathcal{N}(0, 1) \geq z_{1-\alpha/2}) = \alpha/2$. It follows directly from the CLT in Proposition 4 that $\theta^*_1 \in \mathscr{C}_1$ with asymptotic probability $1 - \alpha$. For all other $i \neq 1$, let

$$\tau_i \equiv (1 + c_0)\sqrt{2 \log n \cdot \rho_i^{-2}(\widehat{\theta})},$$

where $\rho_i(\theta)$ denotes the sample version (with $\{A_{ij}\}$) defined in Proposition 4 and c_0 is an arbitrarily small but fixed constant. We then define the confidence intervals for $\{\widehat{\theta}_i\}_{i=2}^n$ to be

$$\mathscr{C}_i \equiv \left[\widehat{\theta}_i - \tau_i, \widehat{\theta}_i + \tau_i\right].$$

The constant in τ_i comes simply from the application of a (conditional) Bernstein's inequality to the main term b_i/d_i in the expansion of Theorem 2, so that by construction $\theta_i^* \in \mathscr{C}_i$ simultaneously for $i \in \{2, \ldots, n\}$ with overwhelming probability. Since ρ_i^{-2} is smaller for those teams i with more comparisons, the above confidence intervals have the desirable property of data-driven length.

The above construction of confidence intervals leads naturally to a confidence interval for the rank $r(1)$. We say two intervals satisfy $\mathscr{C}_i \leq \mathscr{C}_j$ if the upper end of $\mathscr{C}_i$ is smaller than the lower end of $\mathscr{C}_j$. Let n_1 be the number of $\mathscr{C}_i$'s such that $\mathscr{C}_1 \leq \mathscr{C}_i$, and n_2 be the number such that $\mathscr{C}_i \leq \mathscr{C}_1$. By definition, n_1, n_2 take integer values between 0 and $n - 1$. The confidence interval for the rank $r(1)$ is then taken to be all integers inside

$$[n_1 + 1, n - n_2].$$

The following result guarantees the confidence level of the above interval. Details of its proof are given in Section 8.

PROPOSITION 6. Suppose the conditions of Theorem 2 hold. Then with asymptotic probability at least $1 - \alpha$, we have $r(1) \in [n_1 + 1, n - n_2]$.

4.3 *Application III: Optimal constant of ℓ_2 estimation*

With the main expansion in Theorem 2, we are able to pin down the exact constants of ℓ_2 errors of the MLE $\widehat{\theta}$ and the spectral estimator $\widetilde{\theta}$. The proofs of the following two results are given in Sections 9.1 and 9.2, respectively.

PROPOSITION 7. Suppose the conditions in (2.2) hold with $\alpha = 1$. Then it holds with probability at least $1 - \mathcal{O}(n^{-10})$ that

$$\|\widehat{\theta} - \theta^*\|^2 = \frac{1 + \mathfrak{o}(1)}{pL} \cdot \sum_{i=1}^{n} \Big(\sum_{k:k \neq i} \psi'(\theta_i^* - \theta_k^*) \Big)^{-1},$$

where the $\mathfrak{o}(1)$ term is uniform over $\theta^* \in \Theta(\kappa)$ in (2.1).

PROPOSITION 8. Suppose the conditions in (2.2) hold with $\alpha = 1$. Then it holds with probability at least $1 - \mathcal{O}(n^{-10})$ that

$$\|\widetilde{\theta} - \theta^*\|^2 = \frac{1 + \mathfrak{o}(1)}{pL} \cdot \sum_{i=1}^{n} \frac{\sum_{j:j \neq i}(e^{\theta_i^*} + e^{\theta_j^*})^2 \psi'(\theta_i^* - \theta_j^*)}{\left(\sum_{j:j \neq i}(e^{\theta_i^*} + e^{\theta_j^*}) \psi'(\theta_i^* - \theta_j^*) \right)^2},$$

where the $\mathfrak{o}(1)$ term is uniform over $\theta^* \in \Theta(\kappa)$ in (2.1).

REMARK 3. Note that as opposed to the main expansions in Section 2, the above results only require the (almost) minimal sparsity condition $np \gg \log n$ for the underlying Erdős–Rényi graph to be connected. This is because the above results only require the ℓ_2 control of the remainder vector δ in (5.3) instead of the ℓ_∞ control; see Sections 5.2 and 9.1 for details in the case of MLE.

The above results are much more refined than existing ℓ_2 guarantees in the literature [5, 6, 24], whose optimalities are all based on a (global) minimax framework. In particular, they allow us to compare the

two estimators $\widehat{\theta}$ and $\widetilde{\theta}$ on a more fine-grained level. For example, while both $\widehat{\theta}$ and $\widetilde{\theta}$ achieve the (global) minimax rate $1/\sqrt{pL}$ in terms of ℓ_2 accuracy [24], a direct application of Cauchy–Schwarz yields that the error of the spectral estimator in Proposition 8 is always larger than that of the MLE in Proposition 7, and equality only holds in the case where all θ_i^*'s are equal.

In fact, the following result states that the ℓ_2 accuracy of the MLE cannot be improved asymptotically in terms of the local minimax framework. For any $\theta^* \in \mathbb{R}^n$ and $\varepsilon > 0$, we use $B(\theta^*, \varepsilon)$ to denote the hyperrectangle $\times_{i=1}^n [\theta_i^* - \varepsilon, \theta_i^* + \varepsilon]$.

THEOREM 9. Let $\kappa = \mathcal{O}(1)$ and ε_n be any sequence such that $\varepsilon_n \gg (npL)^{-1/2}$. Then for any θ^* such that $\max_i \theta_i^* - \min_i \theta_i^* \leq \kappa/2$, it holds that

$$\inf_{\widehat{\theta}} \sup_{\theta \in B(\theta^*, \varepsilon_n) \cap \Theta(\kappa)} \mathbb{E}_\theta \|\widehat{\theta} - \theta\|^2 \geq \big(1 + \mathfrak{o}(1)\big) \cdot \frac{1}{pL} \sum_{i=1}^n \Big(\sum_{j: j \neq i} \psi'(\theta_i^* - \theta_j^*) \Big)^{-1},$$

where the infimum of $\widehat{\theta}$ is taken over all estimators of θ.

REMARK 4. It is clear that the localization radius $(npL)^{-1/2}$ cannot be improved in terms of its order. Indeed, if localized in a neighbourhood of θ^* of size $\delta_n (npL)^{-1/2}$ for some $\delta_n = \mathfrak{o}(1)$, then the trivial estimator $\widehat{\theta}_{\text{fix}} \equiv \theta^*$ satisfies $\mathbb{E}_\theta \|\widehat{\theta}_{\text{fix}} - \theta\|^2 = \mathfrak{o}\big((pL)^{-1}\big)$ uniformly over this neighbourhood.

The proof of Theorem 9 is given in Section 9.3 and is based on van Trees' inequality [13]. Together with Proposition 7, they imply that the MLE is asymptotically and locally minimax optimal. This agrees with the conventional wisdom in classical parametric statistics that MLE is the most efficient estimator [42].

4.4 *Simulation*

We close this section with some simulation studies to verify the theoretical results above. Throughout the simulation, we set the number of comparison to be $L = 1$ and the dynamic range to be $\kappa = 2$. We first verify the finite-dimensional CLTs established in Section 4.1. For a given number n of individuals, we sample $\theta^* = (\theta_1^*, \ldots, \theta_n^*)^\top$ i.i.d. from Unif($[0, 2]$), and examine the distribution of $\widehat{\theta}_1$ and $\widetilde{\theta}_1$ prescribed by Propositions 4 and 5. For numerical reasons, we use the slightly larger success probability $p = (\log n)^3/n$ for the Erdős–Rényi comparison graph, and the corresponding QQ-plots against the nominal normal quantiles are presented in Figure 1. We see that starting from $n = 500$, both QQ-plots align very well with the diagonal line, validating both CLTs established in Propositions 4 and 5.

Next we validate the ℓ_2 risk predictions prescribed by Propositions 7 and 8. For numerical reasons, we take $p = (\log n)^{3.5}/n$ to generate the Erdős–Rényi comparison graph. From the risk plot in Figure 2, we observe that the theoretical predictions in Propositions 7 and 8 are very accurate, and also that the MLE indeed achieves a smaller risk than the spectral estimator.

5. **Proof of the main expansion for MLE**

The goal of this section is to prove Theorem 2. We introduce some preliminaries in Section 5.1, followed by two main steps of the proof in Sections 5.2 and 5.3, respectively. We refer to Section 3 for the general proof idea.
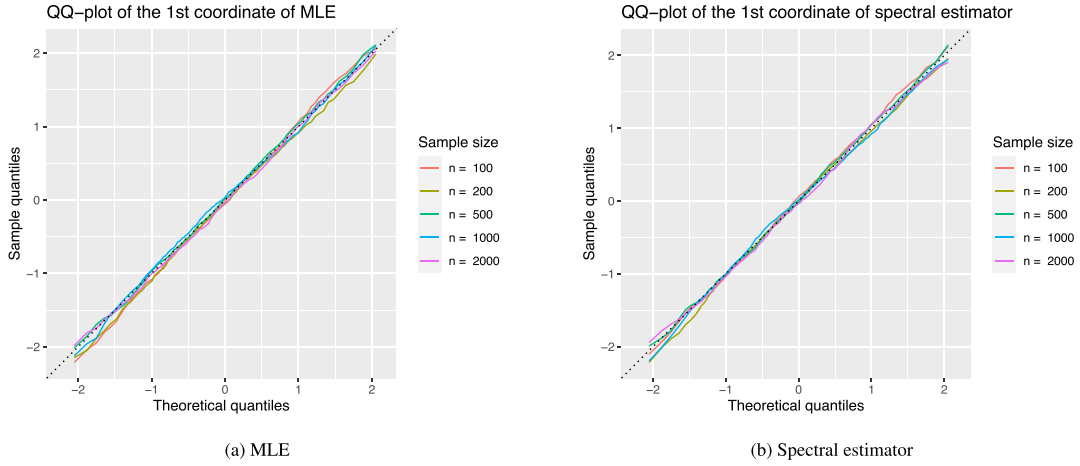
FIG. 1.   QQ-plots comparing the theoretical and sample quantiles of $\widehat{\theta}_1$ and $\widetilde{\theta}_1$. Simulation parameters: $L = 1$, $n \in \{100, 200, 500, 1000, 2000\}$ and $p = (\log n)^3/n$. For each n, $\theta^* = (\theta_1^*, \dots, \theta_n^*)^\top$ are i.i.d. from Unif($[0, 2]$). The empirical quantile curves are averaged over 1000 replications.

5.1   *Preliminary*

We will analyze $\widehat{\theta}$ in a coordinate-wise manner. Fix any $i \in [n]$, and decompose the total likelihood in (2.3) as $\ell_n(\theta) = \ell_n^{(-i)}(\theta_{-i}) + \ell_n^{(i)}(\theta_i | \theta_{-i})$, where

$$\ell_n^{(-i)}(\theta_{-i}) \equiv \sum_{j<k:j,k\neq i} A_{jk}\left[\bar{y}_{jk}\log\frac{1}{\psi(\theta_j - \theta_k)} + \bar{y}_{kj}\log\frac{1}{\psi(\theta_k - \theta_j)}\right],$$

$$\ell_n^{(i)}(\theta_i|\theta_{-i}) \equiv \sum_{j:j\neq i} A_{ij}\left[\bar{y}_{ij}\log\frac{1}{\psi(\theta_i - \theta_j)} + \bar{y}_{ji}\log\frac{1}{\psi(\theta_j - \theta_i)}\right].$$

Next define

$$f^{(i)}(\theta_i|\theta_{-i}) \equiv \frac{\partial}{\partial\theta_i}\ell_n^{(i)}(\theta_i|\theta_{-i}) = -\sum_{j:j\neq i} A_{ij}\big(\bar{y}_{ij} - \psi(\theta_i - \theta_j)\big),$$

$$g^{(i)}(\theta_i|\theta_{-i}) \equiv \frac{\partial^2}{\partial\theta_i^2}\ell_n^{(i)}(\theta_i|\theta_{-i}) = \sum_{j:j\neq i} A_{ij}\psi'(\theta_i - \theta_j). \tag{5.1}$$

Some preliminary estimates regarding $\widehat{\theta}$, $f^{(i)}(\theta_i^*|\widehat{\theta}_{-i})$ and $g^{(i)}(\theta_i^*|\widehat{\theta}_{-i})$ are summarized in Lemmas 16 and 18 in the Appendix. Let

$$\bar{\theta}_i \equiv \theta_i^* - \frac{f^{(i)}(\theta_i^*|\widehat{\theta}_{-i})}{g^{(i)}(\theta_i^*|\widehat{\theta}_{-i})}, \tag{5.2}$$
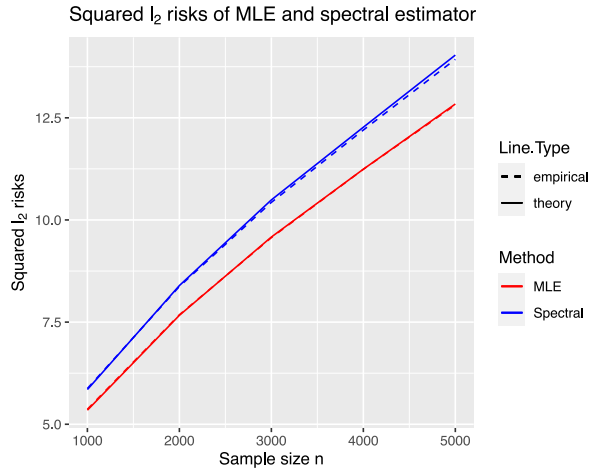
FIG. 2. Squared ℓ_2 risks of the MLE $\widehat{\theta}$ (red) and the spectral estimator $\widetilde{\theta}$ (blue), with solid lines denoting the theoretical predictions by Proposition 7 and 8, and dashed lines denoting the empirical values. Simulation parameters: $L = 1$, $n \in \{1000, 2000, 3000, 4000, 5000\}$ and $p = (\log n)^{3.5}/n$. For each n, $\theta^* = (\theta_1^*, \ldots, \theta_n^*)^\top$ are i.i.d. from Unif($[0, 2]$). The empirical risk curves are averaged over 500 replications.

whose definition is motivated by the local quadratic expansion in (2.9). Lastly define the remainder vector $\delta \in \mathbb{R}^n$ via

$$\widehat{\theta}_i - \theta_i^* = -\frac{f^{(i)}(\theta_i^*|\theta_{-i}^*)}{g^{(i)}(\theta_i^*|\theta_{-i}^*)} + \delta_i. \tag{5.3}$$

We have the following ℓ_∞ bound for δ.

PROPOSITION 10. Suppose that $\kappa = \mathscr{O}(1)$ and $np \gg (\log n)^{3/2}$. The following holds with probability $1 - \mathscr{O}(n^{-10})$.

$$\max_i \left| \delta_i - \varepsilon_i \frac{f^{(i)}(\theta_i^*|\theta_{-i}^*)}{g^{(i)}(\theta_i^*|\theta_{-i}^*)} \right| = \mathfrak{o}\left(\sqrt{\frac{1}{npL}}\right).$$

Here $\varepsilon \in \mathbb{R}^n$ satisfies $\|\varepsilon\|_\infty = \mathfrak{o}(1)$.

Combining the above result with (5.3) immediately completes the proof. The rest of this section is devoted to proving Proposition 10. As an essential intermediate step, we need to first give a sharp ℓ_2 bound for (a leave-one-out version of) δ. We present the ℓ_2 bound in the next subsection and the main proof of Proposition 10 in Section 5.3 ahead.

5.2 Control of $\|\delta\|$

The goal of this subsection is to prove the following ℓ_2 bound for δ defined in (5.3).

PROPOSITION 11. Suppose that $\kappa = \mathscr{O}(1)$ and $np \gg \log n$. Then it holds with probability $1 - \mathscr{O}(n^{-10})$ that

$$\|\delta\| \lesssim \frac{1}{\sqrt{pL}} \cdot \left[\sqrt{\frac{\log n}{np}} + \left(\frac{\log n}{npL}\right)^{1/4}\right] = \mathfrak{o}\left(\frac{1}{\sqrt{pL}}\right).$$

Proof of Proposition 11 Fix $m \in [n]$. By definition, δ_m can be decomposed as follows:

$$
\begin{aligned}
\delta_m &= \frac{f^{(m)}(\theta_m^* | \theta_{-m}^*)}{g^{(m)}(\theta_m^* | \theta_{-m}^*)} - \frac{f^{(m)}(\theta_m^* | \widehat{\theta}_{-m})}{g^{(m)}(\theta_m^* | \widehat{\theta}_{-m})} + \widehat{\theta}_m - \bar{\theta}_m \\
&= \frac{f^{(m)}(\theta_m^* | \theta_{-m}^*) - f^{(m)}(\theta_m^* | \widehat{\theta}_{-m})}{g^{(m)}(\theta_m^* | \widehat{\theta}_{-m})} \\
&\quad + f^{(m)}(\theta_m^* | \theta_{-m}^*)\left(\frac{1}{g^{(m)}(\theta_m^* | \theta_{-m}^*)} - \frac{1}{g^{(m)}(\theta_m^* | \widehat{\theta}_{-m})}\right) + \widehat{\theta}_m - \bar{\theta}_m.
\end{aligned}
\tag{5.4}
$$

For the first term in the above display, Taylor expansion yields that

$$
\begin{aligned}
f^{(m)}(\theta_m^* | \theta_{-m}^*) - f^{(m)}(\theta_m^* | \widehat{\theta}_{-m}) &= \sum_{i:i\neq m} A_{mi}\big(\psi(\theta_m^* - \theta_i^*) - \psi(\theta_m^* - \widehat{\theta}_i)\big) \\
&= \sum_{i:i\neq m} A_{mi}\psi'(\theta_m^* - \theta_i^*)(\widehat{\theta}_i - \theta_i^*) - \frac{1}{2}\sum_{i:i\neq m} A_{mi}\psi''(\theta_m^* - \xi_{m,i})(\widehat{\theta}_i - \theta_i^*)^2,
\end{aligned}
$$

where for each $i \neq m$, $\xi_{m,i}$ is some number between $\widehat{\theta}_i$ and θ_i^*. Now using again the decomposition of $\widehat{\theta}_m$ in (5.3), we have

$$
\begin{aligned}
&f^{(m)}(\theta_m^* | \theta_{-m}^*) - f^{(m)}(\theta_m^* | \widehat{\theta}_{-m}) \\
&= -\sum_{i:i\neq m} A_{mi}\psi'(\theta_m^* - \theta_i^*)\frac{f^{(i)}(\theta_i^* | \theta_{-i}^*)}{g^{(i)}(\theta_i^* | \theta_{-i}^*)} + \sum_{i:i\neq m} A_{mi}\psi'(\theta_m^* - \theta_i^*)\delta_i \\
&\quad - \frac{1}{2}\sum_{i:i\neq m} A_{mi}\psi''(\theta_m^* - \xi_{m,i})(\widehat{\theta}_i - \theta_i^*)^2.
\end{aligned}
$$

Go back to the definition of δ_m, and the above calculation yields that

$$
\begin{aligned}
g^{(m)}(\theta_m^*|\widehat{\theta}_{-m})\delta_m &= \big(f^{(m)}(\theta_m^*|\theta_{-m}^*) - f^{(m)}(\theta_m^*|\widehat{\theta}_{-m})\big) \\
&\quad + f^{(m)}(\theta_m^*|\theta_{-m}^*)\Big(\frac{g^{(m)}(\theta_m^*|\widehat{\theta}_{-m})}{g^{(m)}(\theta_m^*|\theta_{-m}^*)} - 1\Big) + g^{(m)}(\theta_m^*|\widehat{\theta}_{-m})\big(\widehat{\theta}_m - \bar{\theta}_m\big) \\
&= -\sum_{i:i\neq m} A_{mi}\psi'(\theta_m^* - \theta_i^*)\frac{f^{(i)}(\theta_i^*|\theta_{-i}^*)}{g^{(i)}(\theta_i^*|\theta_{-i}^*)} + \sum_{i:i\neq m} A_{mi}\psi'(\theta_m^* - \theta_i^*)\delta_i \\
&\quad - \frac{1}{2}\sum_{i:i\neq m} A_{mi}\psi''(\theta_m^* - \xi_{m,i})(\widehat{\theta}_i - \theta_i^*)^2 \\
&\quad + f^{(m)}(\theta_m^*|\theta_{-m}^*)\Big(\frac{g^{(m)}(\theta_m^*|\widehat{\theta}_{-m})}{g^{(m)}(\theta_m^*|\theta_{-m}^*)} - 1\Big) + g^{(m)}(\theta_m^*|\widehat{\theta}_{-m})\big(\widehat{\theta}_m - \bar{\theta}_m\big). \qquad (5.5)
\end{aligned}
$$

Rearranging the terms yields that

$$
\begin{aligned}
&g^{(m)}(\theta_m^*|\theta_{-m}^*)\delta_m - \sum_{i:i\neq m} A_{mi}\psi'(\theta_m^* - \theta_i^*)\delta_i + \big(g^{(m)}(\theta_m^*|\widehat{\theta}_{-m}) - g^{(m)}(\theta_m^*|\theta_{-m}^*)\big)\delta_m \\
&= -\sum_{i:i\neq m} A_{mi}\psi'(\theta_m^* - \theta_i^*)\frac{f^{(i)}(\theta_i^*|\theta_{-i}^*)}{g^{(i)}(\theta_i^*|\theta_{-i}^*)} - \frac{1}{2}\sum_{i:i\neq m} A_{mi}\psi''(\theta_m^* - \xi_{m,i})(\widehat{\theta}_i - \theta_i^*)^2 \\
&\quad + f^{(m)}(\theta_m^*|\theta_{-m}^*)\Big(\frac{g^{(m)}(\theta_m^*|\widehat{\theta}_{-m})}{g^{(m)}(\theta_m^*|\theta_{-m}^*)} - 1\Big) + g^{(m)}(\theta_m^*|\widehat{\theta}_{-m})\big(\widehat{\theta}_m - \bar{\theta}_m\big) \\
&\equiv R_{1,m} + R_{2,m} + R_{3,m} + R_{4,m}.
\end{aligned}
$$

Here R_1-R_4 are vectors in $\mathbb{R}^n$. Let H be the Laplacian matrix defined by $H_{ii} \equiv \sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)$ and $H_{ij} \equiv -A_{ij}\psi'(\theta_i^* - \theta_j^*)$ for $i \neq j$, and $D = \mathrm{diag}(D_1,\dots,D_n)$ be a diagonal matrix with $D_m \equiv g^{(m)}(\theta_m^*|\widehat{\theta}_{-m}) - g^{(m)}(\theta_m^*|\theta_{-m}^*)$. Then in matrix form,

$$
(H + D)\delta = R_1 + R_2 + R_3 + R_4. \qquad (5.6)
$$

Note that: (i) $\lambda_{\min,\perp}(H) \gtrsim np$ with the prescribed probability by Lemma 15; (ii) By Lemma 18, the operator norm of D satisfies

$$
\|D\|_{\mathrm{op}} = \max_{m\in[n]} |g^{(m)}(\theta_m^*|\widehat{\theta}_{-m}) - g^{(m)}(\theta_m^*|\theta_{-m}^*)| \lesssim \sqrt{\frac{np}{L}} + \sqrt{\frac{(\log n)^3}{npL}} = \mathfrak{o}(np).
$$

Hence, for large enough n, with

$$\text{ave}(u) \equiv \frac{1}{n} \sum_{i=1}^{n} u_i, \quad \forall u \in \mathbb{R}^n, \tag{5.7}$$

the left side of (5.6) can be lower bounded by

$$\|H\delta\| - \|D\|_{\text{op}} \|\delta\| = \|H(\delta - \text{ave}(\delta)\mathbf{1}_n)\| - \mathfrak{o}(np)\|\delta\|$$

$$\gtrsim np\|\delta - \text{ave}(\delta)\mathbf{1}_n\| - \mathfrak{o}(np)\|\delta\| \geq \frac{np}{2}\|\delta\| - np\sqrt{n}|\text{ave}(\delta)|. \tag{5.8}$$

Hence by (5.6), Lemma 1 for the bound for $|\text{ave}(\delta)|$, and Lemma 2 for the bounds for $\|R_1\|$-$\|R_4\|$, we have

$$\|\delta\| \lesssim \sqrt{n}|\text{ave}(\delta)| + (np)^{-1} \cdot \left(\|R_1\| + \|R_2\| + \|R_3\| + \|R_4\|\right)$$

$$\lesssim \frac{1}{\sqrt{pL}} \cdot \sqrt{\frac{\log n}{n}} + (np)^{-1} \cdot \left[\sqrt{\frac{n\log n}{L}} + \left(\frac{n}{L}\sqrt{\frac{np\log n}{L}}\right)^{1/2}\right]$$

$$\asymp \frac{1}{\sqrt{pL}} \cdot \left[\sqrt{\frac{\log n}{np}} + \left(\frac{\log n}{npL}\right)^{1/4}\right],$$

as desired. □

LEMMA 1. Suppose that $\kappa = \mathscr{O}(1)$ and $np \gg \log n$. Then the following holds with probability at least $1 - \mathscr{O}(n^{-10})$.

$$|\text{ave}(\delta)| \lesssim \frac{1}{\sqrt{npL}} \cdot \sqrt{\frac{\log n}{n}} = \mathfrak{o}\left(\frac{1}{\sqrt{npL}}\right).$$

Proof of Lemma 1 Since both $\widehat{\theta}$ and θ^* are centered by definition, the definition of δ in (5.3) yields that

$$\text{ave}(\delta) = \frac{1}{n} \sum_{m=1}^{n} \frac{f^{(m)}(\theta_m^* | \theta_{-m}^*)}{g^{(m)}(\theta_m^* | \theta_{-m}^*)}.$$

Let $w_{mi} \equiv A_{mi}/\big(\sum_{j:j\neq m} A_{mj}\psi'(\theta_m^* - \theta_j^*)\big)$. Then

$$
\begin{aligned}
\mathrm{ave}(\delta) &= \frac{1}{n}\sum_{m=1}^{n} \frac{\sum_{i:i\neq m} A_{mi}\big(\bar{y}_{mi} - \psi(\theta_m^* - \theta_i^*)\big)}{\sum_{i:i\neq m} A_{mi}\psi'(\theta_m^* - \theta_i^*)} \\
&= \frac{1}{n}\sum_{i\neq m} w_{mi}\big(\bar{y}_{mi} - \psi(\theta_m^* - \theta_i^*)\big) \\
&= \frac{1}{n}\sum_{i<m} w_{mi}\big(\bar{y}_{mi} - \psi(\theta_m^* - \theta_i^*)\big) + w_{im}\big(\bar{y}_{im} - \psi(\theta_i^* - \theta_m^*)\big) \\
&\overset{(*)}{=} \frac{1}{n}\sum_{i<m} (w_{mi} - w_{im})\big(\bar{y}_{mi} - \psi(\theta_m^* - \theta_i^*)\big) \\
&= \frac{1}{nL}\sum_{i<m}\sum_{\ell=1}^{L} (w_{mi} - w_{im})\big(y_{mi\ell} - \psi(\theta_m^* - \theta_i^*)\big),
\end{aligned}
$$

where $(*)$ follows as $\bar{y}_{im} - \psi(\theta_i^* - \theta_m^*) = (1 - \bar{y}_{mi}) - (1 - \psi(\theta_m^* - \theta_i^*)) = -(\bar{y}_{mi} - \psi(\theta_m^* - \theta_i^*))$. Now conditioning on the graph A, $\{w_{mi}\}$ are deterministic and the summands in the above display are independent across $i < m, \ell$. Hence, Hoeffding's inequality yields that for any $t > 0$,

$$
\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{m=1}^{n} \frac{f^{(m)}(\theta_m^*|\theta_{-m}^*)}{g^{(m)}(\theta_m^*|\theta_{-m}^*)}\Big| \geq t \Big| A\Big) \leq 2\exp\Big(-\frac{Cn^2Lt^2}{\sum_{i<m}(w_{mi}-w_{im})^2}\Big).
$$

By Lemma 13, it holds with the prescribed probability that

$$
\sum_{i<m} w_{mi}^2 = \sum_{i<m} \Big(\frac{A_{mi}}{\sum_{j:j\neq m} A_{mj}\psi'(\theta_m^* - \theta_j^*)}\Big)^2 \lesssim (np)^{-2}\sum_{i<m} A_{mi} \lesssim p^{-1}.
$$

Since a similar bound holds for $\sum_{i<m} w_{im}^2$, we have $\sum_{i<m}(w_{mi}-w_{im})^2 \lesssim p^{-1}$ with the prescribed probability. By choosing $t \asymp \sqrt{\log n/(n^2 pL)}$, the above two displays yield that with the prescribed probability $|\mathrm{ave}(\delta)| \lesssim n^{-1}\sqrt{\log n/(Lp)} = \sqrt{\log n/n}\cdot(1/\sqrt{npL})$. The proof is complete. $\square$

LEMMA 2. Recall R_1-R_4 defined in the proof of Proposition 11. Suppose $\kappa = \mathcal{O}(1)$ and $np \gg \log n$. Then the following hold with probability $1 - \mathcal{O}(n^{-10})$.

$$
\|R_1\|^2 \vee \|R_2\|^2 \vee \|R_3\|^2 \lesssim \frac{n\log n}{L}, \quad \|R_4\|^2 \lesssim \frac{n}{L}\sqrt{\frac{np\log n}{L}}.
$$

Proof of Lemma 2 (**Bound of R_1**) For each $m \in [n]$, let

$$
w_{ij;m} \equiv \frac{\psi'(\theta_m^* - \theta_i^*)A_{mi}A_{ij}}{\sum_{k:k\neq i} A_{ik}\psi'(\theta_i^* - \theta_k^*)}.
$$

Then, we have

$$
\begin{aligned}
R_{1,m} &= \sum_{i:i\neq m} A_{mi}\psi'(\theta_m^* - \theta_i^*)\frac{\sum_{j:j\neq i} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))}{\sum_{k:k\neq i} A_{ik}\psi'(\theta_i^* - \theta_k^*)} \\
&= \sum_{i,j:j\neq i, i\neq m} \left(\frac{\psi'(\theta_m^* - \theta_i^*)A_{mi}A_{ij}}{\sum_{k\neq i} A_{ik}\psi'(\theta_i^* - \theta_k^*)}\right)(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)) \\
&= \sum_{i,j:j\neq i, i,j\neq m} w_{ij;m}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)) + \sum_{i:i\neq m} w_{im;m}(\bar{y}_{im} - \psi(\theta_i^* - \theta_m^*)) \\
&= \sum_{i,j:i<j, i,j\neq m} w_{ij;m}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)) + w_{ji;m}(\bar{y}_{ji} - \psi(\theta_j^* - \theta_i^*)) \\
&\qquad + \sum_{i:i\neq m} w_{im;m}(\bar{y}_{im} - \psi(\theta_i^* - \theta_m^*)) \\
&\overset{(*)}{=} \frac{1}{L}\sum_{i,j:i<j, i,j\neq m}\sum_{\ell=1}^{L}(w_{ij;m} - w_{ji;m})(y_{ij\ell} - \psi(\theta_i^* - \theta_j^*)) \\
&\qquad + \frac{1}{L}\sum_{i:i\neq m}\sum_{\ell=1}^{L} w_{im;m}(y_{im\ell} - \psi(\theta_i^* - \theta_m^*)).
\end{aligned}
$$

Here $(*)$ follows as $\bar{y}_{ji} - \psi(\theta_j^* - \theta_i^*) = (1 - \bar{y}_{ij}) - (1 - \psi(\theta_i^* - \theta_j^*)) = -(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))$. Now conditioning on the graph A, $\{w_{ij;m}\}$ are deterministic and the above two sums are independent across their summands. Hence Hoeffding's inequality yields that

$$
\left|\frac{1}{L}\sum_{i,j:i<j, i,j\neq m}\sum_{\ell=1}^{L}(w_{ij;m} - w_{ji;m})(y_{ij\ell} - \psi(\theta_i^* - \theta_j^*))\right|
$$
$$
\lesssim \sqrt{\frac{\log n}{L}\cdot\sum_{i,j:i<j, i,j\neq m}(w_{ij;m} - w_{ji;m})^2},
$$
$$
\left|\frac{1}{L}\sum_{i:i\neq m}\sum_{\ell=1}^{L} w_{im;m}(y_{im\ell} - \psi(\theta_i^* - \theta_m^*))\right| \lesssim \sqrt{\frac{\log n}{L}\cdot\sum_{i:i\neq m} w_{im;m}^2}
$$

with the prescribed probability. Now by Lemma 13, we have

$$
\sum_{i,j:i<j,i,j\neq m} (w_{ij;m} - w_{ji;m})^2 \lesssim \sum_{i,j:i<j,i,j\neq m} \frac{A_{ij}A_{mi}}{\left(\sum_{k:k\neq i} A_{ik}\psi'(\theta_i^* - \theta_k^*)\right)^2}
$$

$$
\lesssim (np)^{-2} \sum_{i,j:i<j,i,j\neq m} A_{ij}A_{mi} \lesssim 1,
$$

$$
\sum_{i:i\neq m} w_{im;m}^2 \lesssim \sum_{i:i\neq m} \frac{A_{im}}{\left(\sum_{k:k\neq i} A_{ik}\psi'(\theta_i^* - \theta_k^*)\right)^2} \lesssim \frac{1}{np}.
$$

Combining the above two displays yields that $\max_{m\in[n]} |R_{1,m}| \lesssim \sqrt{\log n/L}$, and hence $\|R_1\| \lesssim \sqrt{n\log n/L}$ with the prescribed probability.

(**Bound of R_2**) Using $\sup_{u\in\mathbb{R}} |\psi''(u)| \lesssim 1$, we have $|R_{2,m}| \lesssim (\sum_{i:i\neq m} A_{im}|\widehat{\theta}_i - \theta_i^*|) \cdot \|\widehat{\theta} - \theta^*\|_\infty$. Hence, by Lemma 15 and Proposition 1, we have

$$
\|R_2\|^2 \lesssim \sum_{m=1}^n \left(\sum_{i:i\neq m} A_{mi}|\widehat{\theta}_i - \theta_i^*|\right)^2 \cdot \|\widehat{\theta} - \theta^*\|_\infty^2
$$

$$
\leq \|A\|_{op}^2 \cdot \|\widehat{\theta} - \theta^*\|^2 \cdot \|\widehat{\theta} - \theta^*\|_\infty^2
$$

$$
\lesssim (np)^2 \cdot \frac{1}{pL} \cdot \frac{\log n}{npL} = \frac{n\log n}{L^2}.
$$

(**Bound of R_3**) By Lemma 18, it holds with the prescribed probability that

$$
|R_{3,m}| \leq \left|f^{(m)}(\theta_m^*|\theta_{-m}^*)\right| \cdot \frac{\max_{m\in[n]}\left|g^{(m)}(\theta_m^*|\theta_{-m}^*) - g^{(m)}(\theta_m^*|\widehat{\theta}_{-m})\right|}{g^{(m)}(\theta_m^*|\theta_{-m}^*)}
$$

$$
\lesssim \left|f^{(m)}(\theta_m^*|\theta_{-m}^*)\right| \cdot \frac{1}{np}\left(\sqrt{\frac{np}{L}} + \sqrt{\frac{(\log n)^3}{npL}}\right).
$$

Hence, again by Lemma 18,

$$
\|R_3\|^2 \lesssim \sum_{m=1}^n \left(f^{(m)}(\theta_m^*|\theta_{-m}^*)\right)^2 \cdot \frac{1}{(np)^2}\left(\frac{np}{L} + \frac{(\log n)^3}{npL}\right)
$$

$$
\lesssim \frac{n^2 p}{L} \cdot \frac{1}{(np)^2}\left(\frac{np}{L} + \frac{(\log n)^3}{npL}\right) = \frac{n}{L^2} + \frac{n\log n}{L} \cdot \left(\frac{\log n}{np}\right)^2.
$$

(**Bound of R_4**) Using the bound $g^{(m)}(\theta_m^*|\theta_{-m}^*) = \sum_{i:i\neq m} A_{mi}\psi'(\theta_m^* - \theta_i^*) \lesssim np$ by Lemma 13, it follows from Lemma 19 that

$$\left[g^{(m)}(\theta_m^*|\theta_{-m}^*)(\widehat{\theta}_m - \bar{\theta}_m)\right]^2 \lesssim (np)^2 \left(\frac{|f^{(m)}(\theta_m^*|\widehat{\theta}_{-m})|}{np}\right)^3 = \frac{\left|f^{(m)}(\theta_m^*|\widehat{\theta}_{-m})\right|^3}{np}.$$

Moreover, it follows from Lemma 18-(4) that $g^{(m)}(\theta_m^*|\widehat{\theta}_{-m}) = (1 + \mathfrak{o}(1))g^{(m)}(\theta_m^*|\widehat{\theta}_{-m})$. Hence, by Lemma 18-(1), we have

$$
\begin{aligned}
\|R_4\|^2 &= \sum_{m=1}^n \left[g^{(m)}(\theta_m^*|\widehat{\theta}_{-m})(\widehat{\theta}_m - \bar{\theta}_m)\right]^2 \\
&\lesssim \sum_{m=1}^n \left[g^{(m)}(\theta_m^*|\theta_{-m}^*)(\widehat{\theta}_m - \bar{\theta}_m)\right]^2 \lesssim \frac{1}{np}\sum_{m=1}^n \left|f^{(m)}(\theta_m^*|\widehat{\theta}_{-m})\right|^3 \\
&\le \frac{1}{np}\cdot \max_{m\in[n]}\left|f^{(m)}(\theta_m^*|\widehat{\theta}_{-m})\right| \cdot \sum_{m=1}^n \left|f^{(m)}(\theta_m^*|\widehat{\theta}_{-m})\right|^2 \\
&\lesssim \frac{1}{np}\cdot \sqrt{\frac{np\log n}{L}}\cdot \frac{n^2 p}{L} = \frac{n}{L}\sqrt{\frac{np\log n}{L}}.
\end{aligned}
$$

The proof is complete. $\qquad\square$

5.3 *Control of $\|\delta\|_\infty$*

We will now prove Proposition 10. We start by introducing the leave-two-out technique and some additional notation in Section 5.3.1, followed by some preliminary estimates of the corresponding quantities in Section 5.3.2. The main proof of Proposition 10 is given in Section 5.3.3.

5.3.1 *Leave-two-out preliminary*
Recall that δ satisfies the following self-consistent equation in (5.5): for each $i \in [n]$,

$$g^{(i)}(\theta_i^*|\widehat{\theta}_{-i})\delta_i = \sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)\delta_j + \bar{R}_i$$

for some error term $\bar{R} \in \mathbb{R}^n$. In the previous section, we derived the bound for $\|\delta\|$ by treating the above system as a whole and solving for δ. Here for $\|\delta\|_\infty$, we need to bound the right side directly, and the key difficulty lies in controlling the term $\sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)\delta_j$. By standard concentration arguments, this term could be properly controlled if the sequence $\{A_{ij}\}_{j:j\neq i}$ was independent of $\{\delta_j\}_{j:j\neq i}$. For this reason, we introduce in the following a leave-one-out version $\delta^{(i)}$ of δ that is independent of the ith individual (and therefore $\{A_{ij}\}_{j:j\neq i}$).

Fix any index $i \in [n]$ to be left out. Recall the leave-one-out likelihood

$$\ell_n^{(-i)}(\theta_{-i}) = \sum_{j<k:j,k\neq i} A_{jk}\Big[\bar{y}_{jk}\log\frac{1}{\psi(\theta_j-\theta_k)} + \bar{y}_{kj}\log\frac{1}{\psi(\theta_k-\theta_j)}\Big].$$

The leave-one-out MLE $\widehat{\theta}^{(i)} \in \mathbb{R}^{n-1}$ is defined by

$$\widehat{\theta}^{(i)} \equiv \mathrm{argmin}_{\theta_{-i}\in\mathbb{R}^{n-1}:\mathrm{ave}(\theta_{-i})=\mathrm{ave}(\theta_{-i}^*)}\ell_n^{(-i)}(\theta_{-i}). \tag{5.9}$$

We start by finding an approximation $\bar{\theta}_j^{(i)}$ of $\widehat{\theta}_j^{(i)}$ for each $j \neq i$, analogous to the way $\bar{\theta}_j$ approximates $\widehat{\theta}_j$. By isolating the jth individual, we have the following leave-two-out decomposition $\ell_n^{(-i)}(\theta_{-i}) = \ell_n^{(-i,-j)}(\theta_{-i,-j}) + \ell_n^{(-i,j)}(\theta_j|\theta_{-i,-j})$, where

$$\ell_n^{(-i,-j)}(\theta_{-i,-j}) = \sum_{k<\ell:k,\ell\notin\{i,j\}} A_{k\ell}\Big[\bar{y}_{k\ell}\log\frac{1}{\psi(\theta_k-\theta_\ell)} + \bar{y}_{\ell k}\log\frac{1}{\psi(\theta_\ell-\theta_k)}\Big],$$

$$\ell_n^{(-i,j)}(\theta_j|\theta_{-i,-j}) = \sum_{k\notin\{i,j\}} A_{jk}\Big[\bar{y}_{jk}\log\frac{1}{\psi(\theta_j-\theta_k)} + \bar{y}_{kj}\log\frac{1}{\psi(\theta_k-\theta_j)}\Big].$$

Analogous to $f^{(i)}(\theta_i|\theta_{-i})$ and $g^{(i)}(\theta_i|\theta_{-i})$, define for each $j \neq i$

$$f^{(-i,j)}(\theta_j|\theta_{-i,-j}) \equiv \frac{\partial}{\partial\theta_j}\ell_n^{(-i,j)}(\theta_j|\theta_{-i,-j}) = -\sum_{k\notin\{i,j\}} A_{jk}\big(\bar{y}_{jk}-\psi(\theta_j-\theta_k)\big),$$

$$g^{(-i,j)}(\theta_j|\theta_{-i,-j}) \equiv \frac{\partial^2}{\partial\theta_j^2}\ell_n^{(-i,j)}(\theta_j|\theta_{-i,-j}) = \sum_{k\notin\{i,j\}} A_{jk}\psi'(\theta_j-\theta_k). \tag{5.10}$$

Using a similar reasoning for $\bar{\theta}$, define for each $j \neq i$

$$\bar{\theta}_j^{(i)} \equiv \theta_j^* - \frac{f^{(-i,j)}(\theta_j^*|\widehat{\theta}_{-j}^{(i)})}{g^{(-i,j)}(\theta_j^*|\widehat{\theta}_{-j}^{(i)})}.$$

This leads to the definition of a leave-one-out version of δ: for each $j \neq i$, let $\delta_j^{(i)}$ be defined by

$$\widehat{\theta}_j^{(i)} - \theta_j^* \equiv -\frac{f^{(-i,j)}(\theta_j^*|\theta_{-i,-j}^*)}{g^{(-i,j)}(\theta_j^*|\theta_{-i,-j}^*)} + \delta_j^{(i)}. \tag{5.11}$$

The following lemma is an analogue of Lemma 18 for the quantities $f^{(-i,j)}(\theta_j|\theta_{-i,-j})$ and $g^{(-i,j)}(\theta_j|\theta_{-i,-j})$. Its proof relies on the analysis of the leave-two-out MLE defined as

$$\widehat{\theta}^{(i,j)} \equiv \operatorname{argmax}_{\theta_{-i,-j}\in\mathbb{R}^{n-2}:\operatorname{ave}(\theta_{-i,-j})=\operatorname{ave}(\theta^*_{-i,-j})}\ell_n^{(-i,-j)}(\theta_{-i,-j}). \tag{5.12}$$

LEMMA 3. Suppose $\kappa = \mathscr{O}(1)$ and $np \gg \log n$. Then the following hold with probability at least $1 - \mathscr{O}(n^{-10})$.

1. There exists some $C = C(\kappa) > 0$ such that

$$\max_{i\in[n]} \max_{j:j\neq i} \left|f^{(-i,j)}(\theta_j^*|\widehat{\theta}_{-j}^{(i)})\right| \leq C\sqrt{\frac{np\log n}{L}}$$

and

$$\max_{i\in[n]} \sum_{j:j\neq i} \left(f^{(-i,j)}(\theta_j^*|\widehat{\theta}_{-j}^{(i)})\right)^2 \leq C\frac{n^2 p}{L}.$$

2. There exist some positive $c = c(\kappa)$ and $C = C(\kappa)$ such that

$$cnp \leq \min_{i\in[n]} \min_{j:j\neq i} g^{(-i,j)}(\theta_i^*|\theta_{-i,-j}^*) \leq \max_{i\in[n]} \max_{j:j\neq i} g^{(-i,j)}(\theta_i^*|\theta_{-i,-j}^*) \leq Cnp.$$

3. There exists some $C = C(\kappa) > 0$ such that

$$\max_{i\in[n]} \max_{j:j\neq i} \left|g^{(-i,j)}(\theta_j^*|\widehat{\theta}_{-j}^{(i)}) - g^{(-i,j)}(\theta_j^*|\theta_{-i,-j}^*)\right| \leq C\left(\sqrt{\frac{np}{L}} + \sqrt{\frac{(\log n)^3}{npL}}\right).$$

Proof of Lemma 3

(1) For the first inequality, fix any $i \in [n]$ and then j such that $j \neq i$. Then

$$-f^{(-i,j)}(\theta_j^*|\widehat{\theta}_{-j}^{(i)}) = \sum_{k\notin\{i,j\}} A_{jk}\left(\bar{y}_{jk} - \psi(\theta_j^* - \widehat{\theta}_k^{(i)})\right)$$

$$= \sum_{k\notin\{i,j\}} A_{jk}\left(\bar{y}_{jk} - \psi(\theta_j^* - \theta_k^*)\right) + \sum_{k\notin\{i,j\}} A_{jk}\left(\psi(\theta_j^* - \theta_k^*) - \psi(\theta_j^* - \widehat{\theta}_k^{(i)})\right)$$

$$\equiv (I_{ij}) + (II_{ij}).$$

By Lemma 14, we have $\max_{i\in[n]} \max_{j:j\neq i} |(I_{ij})| \lesssim \sqrt{np \log n/L}$ with the prescribed probability. On the other hand, as $\psi(\cdot)$ is Lipschitz with a universal constant, Lemmas 13 and 16 yield that

$$\max_{i\in[n]} \max_{j:j\neq i} |(II_{ij})| = \max_{i\in[n]} \max_{j:j\neq i} \Big| \sum_{k\notin\{i,j\}} A_{jk}\big(\psi(\theta_j^* - \theta_k^*) - \psi(\theta_j^* - \widehat{\theta}_k^{(i)})\big)\Big|$$

$$\leq \max_{i\in[n]}\|\widehat{\theta}^{(i)} - \theta_{-i}^*\|_\infty \cdot \max_{i\in[n]} \max_{j:j\neq i} \sum_{k\notin\{i,j\}} A_{jk}$$

$$\lesssim \sqrt{\frac{\log n}{npL}} \cdot np = \sqrt{\frac{np \log n}{L}},$$

with the prescribed probability. Putting together the estimates concludes the first inequality. For the second inequality, note that again by Lemma 14, $\max_{i\in[n]} \sum_{j:j\neq i}^n (I_{ij})^2 \lesssim n^2 p/L$. On the other hand, by Taylor expansion,

$$(II_{ij}) = \sum_{k\notin\{i,j\}} A_{jk}\psi'(\theta_j^* - \xi_{ijk})(\widehat{\theta}_k^{(i)} - \theta_k^*)$$

for some ξ_{ijk} between $\widehat{\theta}_k^{(i)}$ and θ_k^*. Hence using $\sup_{u\in\mathbb{R}} |\psi'(u)| \lesssim 1$, we have

$$\max_{i\in[n]} \sum_{j:j\neq i} (II_{ij})^2 = \max_{i\in[n]} \sum_{j:j\neq i} \Big(\sum_{k\notin\{i,j\}} A_{jk}\psi'(\theta_j^* - \xi_{ijk})(\widehat{\theta}_k^{(i)} - \theta_k^*)\Big)^2$$

$$\lesssim \|A\|_{\mathrm{op}}^2 \cdot \max_{i\in[n]}\|\widehat{\theta}^{(i)} - \theta_{-i}^*\|^2 \overset{(*)}{\lesssim} (np)^2 \cdot \frac{1}{pL} = \frac{n^2 p}{L},$$

where $(*)$ is by Lemmas 15 and 16.

(2) This follows directly from Lemma 13 and the fact that $\psi'(\theta_i - \theta_j) \asymp 1$ for any $\theta \in \Theta(\kappa)$.

(3) Fix $i \in [n]$ and then j such that $j \neq i$. Recall the definition of the leave-two-out MLE $\widehat{\theta}^{(i,j)}$ in (5.12). Then using analogous arguments for the leave-one-out MLE $\widehat{\theta}^{(i)}$, we have the following estimates for the leave-two-out MLE $\widehat{\theta}^{(i,j)}$ (cf. Lemma 16):

$$\max_{i,j:j\neq i}\|\widehat{\theta}^{(i,j)} - \theta_{-i,-j}^*\| \lesssim \sqrt{\frac{1}{pL}}, \quad \max_{i,j:j\neq i}\|\widehat{\theta}^{(i,j)} - \theta_{-i,-j}^*\|_\infty \lesssim \sqrt{\frac{\log n}{npL}}$$

and

$$\max_{i,j:j\neq i}\|\widehat{\theta}^{(i,j)} - \widehat{\theta}_{-j}^{(i)}\| \lesssim \sqrt{\frac{1}{npL}}.$$

Using the above estimates and Lemma 13, we have

$$
\left| g^{(-i,j)}(\theta_j^* | \widehat{\theta}_{-j}^{(i)}) - g^{(-i,j)}(\theta_j^* | \theta_{-i,-j}^*) \right|
$$

$$
= \Big| \sum_{k \notin \{i,j\}} A_{jk} \big(\psi'(\theta_j^* - \theta_k^*) - \psi'(\theta_j^* - \widehat{\theta}_k^{(i)}) \big) \Big|
$$

$$
\lesssim \sum_{k \notin \{i,j\}} A_{jk} |\widehat{\theta}_k^{(i)} - \theta_k^*| \leq \sum_{k \notin \{i,j\}} A_{jk} |\widehat{\theta}_k^{(i)} - \widehat{\theta}_k^{(i,j)}| + \sum_{k \notin \{i,j\}} A_{jk} |\widehat{\theta}_k^{(i,j)} - \theta_k^*|
$$

$$
\leq \Big(\sum_{k \notin \{i,j\}} A_{jk} \Big)^{1/2} \|\widehat{\theta}_{-j}^{(i)} - \widehat{\theta}^{(i,j)}\| + p \cdot \sum_{k \notin \{i,j\}} |\widehat{\theta}_k^{(i,j)} - \theta_k^*| + \sum_{k \notin \{i,j\}} (A_{jk} - p) |\widehat{\theta}_k^{(i,j)} - \theta_k^*|
$$

$$
\leq \sqrt{np} \sqrt{\frac{1}{npL}} + p\sqrt{n} \|\widehat{\theta}^{(i,j)} - \theta_{-i,-j}^*\| + \sqrt{p \log n} \|\widehat{\theta}^{(i,j)} - \theta_{-i,-j}^*\| + \log n \cdot \|\widehat{\theta}^{(i,j)} - \theta_{-i,-j}^*\|_\infty
$$

$$
\lesssim \sqrt{\frac{1}{L}} + \sqrt{\frac{np}{L}} + \sqrt{\frac{\log n}{L}} + \sqrt{\frac{(\log n)^3}{npL}} \asymp \sqrt{\frac{np}{L}} + \sqrt{\frac{(\log n)^3}{npL}}.
$$

The above bound is uniform over i, j so the proof is complete. □

5.3.2 *Bounds for $\delta^{(i)}$*

We first establish the following analogue of Proposition 11 for $\|\delta^{(i)}\|$. The proof is similar so we only sketch the steps.

LEMMA 4. Suppose $\kappa = \mathcal{O}(1)$ and $np \gg \log n$. Then the following holds with probability at least $1 - \mathcal{O}(n^{-10})$.

$$
\max_{m \in [n]} \|\delta^{(m)}\| \lesssim \frac{1}{\sqrt{pL}} \cdot \left[\sqrt{\frac{\log n}{np}} + \left(\frac{\log n}{npL} \right)^{1/4} \right] = \mathfrak{o}\Big(\frac{1}{\sqrt{pL}} \Big).
$$

Proof of Lemma 4 Following the arguments in Proposition 11, we arrive at the following expansion analogous to (5.6):

$$
(H^{(m)} + D^{(m)}) \delta^{(m)} = R_1^{(m)} + R_2^{(m)} + R_3^{(m)} + R_4^{(m)}. \tag{5.13}
$$

Here $H^{(m)}, D^{(m)}$, and $R_1^{(m)}$-$R_4^{(m)}$ are leave-one-out versions of H, D, and R^1-R^4 defined by

- $H^{(m)} \in \mathbb{R}^{(n-1) \times (n-1)}$ with $H_{ii}^{(m)} = \sum_{j:j \notin \{m,i\}} A_{ij} \psi'(\theta_i^* - \theta_j^*)$ for $i \neq m$ and $H_{ij}^{(m)} = -A_{ij} \psi'(\theta_i^* - \theta_j^*)$ for $j \notin \{m, i\}$.

- $D^{(m)} \in \mathbb{R}^{(n-1) \times (n-1)}$ is defined by $D^{(m)} = \mathrm{diag}(D_1^{(m)}, \ldots, D_{m-1}^{(m)}, D_{m+1}^{(m)}, \ldots, D_n^{(m)})$ with $D_i^{(m)} = g^{(-m,i)}(\theta_i^* | \widehat{\theta}_{-i}^{(m)}) - g^{(-m,i)}(\theta_i^* | \theta_{-m,-i}^*)$.

- $R_1^{(m)}$ - $R_4^{(m)}$ are defined by: for each $i \neq m$,

$$R_{1,i}^{(m)} = -\sum_{j \notin \{m,i\}} A_{ij} \psi'(\theta_i^* - \theta_j^*) \frac{f^{(-m,j)}(\theta_j^*|\theta_{-m,-j}^*)}{g^{(-m,j)}(\theta_i^*|\theta_{-m,-j}^*)},$$

$$R_{2,i}^{(m)} = -\frac{1}{2}\sum_{j \notin \{m,i\}} A_{ij} \psi''(\theta_i^* - \xi_{m,j})(\widehat{\theta}_j^{(m)} - \theta_j^*)^2,$$

$$R_{3,i}^{(m)} = f^{(-m,i)}(\theta_i^*|\widehat{\theta}_{-i}^{(m)})\Big(\frac{g^{(-m,i)}(\theta_i^*|\widehat{\theta}_{-i}^{(m)})}{g^{(-m,i)}(\theta_i^*|\theta_{-m,-i}^*)} - 1\Big),$$

$$R_{4,i}^{(m)} = g^{(-m,i)}(\theta_i^*|\widehat{\theta}_{-i}^{(m)})\big(\widehat{\theta}_i^{(m)} - \bar{\theta}_i^{(m)}\big).$$

Note that: (i) $\mathbf{1}_{n-1}$ is in the null space of $H^{(m)}$ and the smallest non-null eigenvalue satisfies $\min_{x \in \mathbb{R}^{n-1}: \|x\|=1, \mathrm{ave}(x)=0} x^\top H^{(m)} x \gtrsim np$ by Lemma 15; (ii) $\|D^{(m)}\|_{\mathrm{op}} = \mathfrak{o}(np)$ by Lemma 3-(2). Hence, continuing with the arguments in Proposition 11 leads to

$$\|\delta^{(m)}\| \lesssim \sqrt{n}|\mathrm{ave}(\delta^{(m)})| + (np)^{-1} \cdot \big(\|R_1^{(m)}\| + \|R_2^{(m)}\| + \|R_3^{(m)}\| + \|R_4^{(m)}\|\big).$$

The quantities on the right side can be bounded as follows:

- Note that by definition $\mathrm{ave}(\widehat{\theta}^{(m)}) = \mathrm{ave}(\theta_{-m}^*)$, hence by analogous arguments as in Lemma 1, we have

$$|\mathrm{ave}(\delta^{(m)})| = \Big|\frac{1}{n-1}\sum_{i:i\neq m} \frac{f^{(-m,i)}(\theta_i^*|\theta_{-m,-i}^*)}{g^{(-m,i)}(\theta_i^*|\theta_{-m,i}^*)}\Big| \lesssim \frac{1}{\sqrt{npL}} \cdot \sqrt{\frac{\log n}{n}}.$$

- By analogous arguments as in Lemma 2, we have $\|R_1^{(m)}\| \vee \|R_2^{(m)}\| \vee \|R_3^{(m)}\| \lesssim \sqrt{n\log n/L}$.
- By analogous arguments as in Lemma 19, we have

$$\big|\widehat{\theta}_i^{(m)} - \bar{\theta}_i^{(m)}\big| \lesssim \Big(\frac{\big|f^{(-m,i)}(\theta_i^*|\widehat{\theta}_{-i}^{(m)})\big|}{np}\Big)^{3/2}. \tag{5.14}$$

The bound $\|R_4^{(m)}\| \lesssim \big(n\sqrt{np\log n}/L^{3/2}\big)^{1/2}$ now follows from subsequent arguments in Lemma 2 upon using Lemma 3.

Collecting the estimates yields the desired claim. □

The following control of $\|\delta^{(m)}\|_\infty$ is also needed in the bound for $\|\delta\|_\infty$ below.

LEMMA 5. Under the conditions $\kappa = \mathcal{O}(1)$ and $np \gg \log n$, the following holds with probability at least $1 - \mathcal{O}(n^{-10})$.

$$\max_{m \in [n]} \|\delta^{(m)}\|_\infty \lesssim \sqrt{\frac{\log n}{npL}}.$$

Proof of Lemma 5 By combining (5.14) and Lemma 3-(1), we have

$$|\widehat{\theta}_i^{(m)} - \bar{\theta}_i^{(m)}| \lesssim \Big(\frac{|f^{(-m,i)}(\theta_i^* | \widehat{\theta}_{-i}^{(m)})|}{np}\Big)^{3/2} \lesssim \Big(\frac{\log n}{npL}\Big)^{3/4}. \tag{5.15}$$

Hence, for any $i \neq m$, we have

$$\begin{aligned}
|\delta_i^{(m)}| &\leq \Big|\frac{f^{(-m,i)}(\theta_i^* | \widehat{\theta}_{-i}^{(m)}) - f^{(-m,i)}(\theta_i^* | \theta_{-m,-i}^*)}{g^{(-m,i)}(\theta_i^* | \widehat{\theta}_{-i}^{(m)})}\Big| + |\widehat{\theta}_i^{(m)} - \bar{\theta}_i^{(m)}| \\
&\quad + \Big|f^{(-m,i)}(\theta_i^* | \theta_{-m,-i}^*)\Big(\frac{1}{g^{(-m,i)}(\theta_m^* | \theta_{-i}^*)} - \frac{1}{g^{(-m,i)}(\theta_m^* | \widehat{\theta}_{-i}^{(m)})}\Big)\Big| \\
&\overset{(*)}{\leq} (np)^{-1} \sum_{j \notin \{m,i\}} A_{ij} |\widehat{\theta}_j^{(m)} - \theta_j^*| + \Big(\frac{\log n}{npL}\Big)^{3/4} \\
&\quad + \sqrt{\frac{np \log n}{L}} (np)^{-2} \Big(\sqrt{\frac{np}{L}} + \sqrt{\frac{(\log n)^3}{npL}}\Big) \\
&\lesssim (np)^{-1} \Big(\sum_{j \notin \{m,i\}} A_{ij}\Big) \cdot \|\widehat{\theta}^{(m)} - \theta_{-m}^*\|_\infty + \Big(\frac{\log n}{npL}\Big)^{3/4} + \frac{\sqrt{\log n}}{npL} + \frac{(\log n)^2}{(np)^2 L} \\
&\overset{(**)}{\lesssim} \sqrt{\frac{\log n}{npL}} + \Big(\frac{\log n}{npL}\Big)^{3/4} + \frac{\sqrt{\log n}}{npL} + \frac{(\log n)^2}{(np)^2 L} \asymp \sqrt{\frac{\log n}{npL}}.
\end{aligned}$$

Here $(*)$ follows from (5.15) and Lemma 3, and $(**)$ follows from Lemma 16. The proof is complete. $\qquad\square$

5.3.3 *Main proof*

Proof of Proposition 10 Recall the following decomposition of δ in (5.4): for each $m \in [n]$,

$$\begin{aligned}
\delta_m &= \frac{f^{(m)}(\theta_m^* | \theta_{-m}^*) - f^{(m)}(\theta_m^* | \widehat{\theta}_{-m})}{g^{(m)}(\theta_m^* | \widehat{\theta}_{-m})} + (\widehat{\theta}_m - \bar{\theta}_m) \\
&\quad + f^{(m)}(\theta_m^* | \theta_{-m}^*)\Big(\frac{1}{g^{(m)}(\theta_m^* | \theta_{-m}^*)} - \frac{1}{g^{(m)}(\theta_m^* | \widehat{\theta}_{-m})}\Big).
\end{aligned}$$

Recall the definition of the leave-one-out MLE $\widehat{\theta}^{(m)}$ in (5.9). Then δ_m can be further decomposed as

$$\delta_m = \frac{\sum_{i:i \neq m} A_{mi}\big(\psi(\theta_m^* - \widehat{\theta}_i^{(m)}) - \psi(\theta_m^* - \widehat{\theta}_i)\big)}{g^{(m)}(\theta_m^* | \widehat{\theta}_{-m})}$$

$$+ \frac{\sum_{i:i \neq m} A_{mi}\big(\psi(\theta_m^* - \theta_i^*) - \psi(\theta_m^* - \widehat{\theta}_i^{(m)})\big)}{g^{(m)}(\theta_m^* | \widehat{\theta}_{-m})} + (\widehat{\theta}_m - \bar{\theta}_m)$$

$$+ \frac{f^{(m)}(\theta_m^* | \theta_{-m}^*)}{g^{(m)}(\theta_m^* | \theta_{-m}^*)} \Big(1 - \frac{g^{(m)}(\theta_m^* | \theta_{-m}^*)}{g^{(m)}(\theta_m^* | \widehat{\theta}_{-m})}\Big)$$

$$\equiv \delta_{1,m} + \delta_{2,m} + \delta_{3,m} + \frac{f^{(m)}(\theta_m^* | \theta_{-m}^*)}{g^{(m)}(\theta_m^* | \theta_{-m}^*)} \Big(1 - \frac{g^{(m)}(\theta_m^* | \theta_{-m}^*)}{g^{(m)}(\theta_m^* | \widehat{\theta}_{-m})}\Big).$$

By Lemma 18, we have $1 - g^{(m)}(\theta_m^* | \theta_{-m}^*)/g^{(m)}(\theta_m^* | \widehat{\theta}_{-m}) = \mathfrak{o}(1)$ under the condition $np \gg \log n$. On the other hand, it follows from Lemma 6 that under the condition $np \gg (\log n)^{3/2}$, we have $\max_{m \in [n]} |\delta_{1,m}| \vee |\delta_{2,m}| = \mathfrak{o}(1/\sqrt{npL})$ and

$$|\delta_{3,m}| \lesssim \varepsilon_m \Big|\frac{f^{(m)}(\theta_m^* | \theta_{-m}^*)}{g^{(m)}(\theta_m^* | \theta_{-m}^*)}\Big| + \mathfrak{o}(\frac{1}{\sqrt{npL}}),$$

where $\|\varepsilon\|_\infty = \mathfrak{o}(1)$ and the $\mathfrak{o}(1/\sqrt{npL})$ is uniform over $m \in [n]$. The proof is now complete by collecting the estimates. □

LEMMA 6. Recall the definition of $\delta_{1,m}$-$\delta_{3,m}$ from the proof of Proposition 10. Suppose $\kappa = \mathcal{O}(1)$ and $np \gg \log n$. Then the following holds with probability at least $1 - \mathcal{O}(n^{-10})$ uniformly over $m \in [n]$.

$$|\delta_{1,m}| \lesssim \frac{1}{\sqrt{npL}}\sqrt{\frac{1}{np}},$$

$$|\delta_{2,m}| \lesssim \frac{1}{\sqrt{npL}} \cdot \Big(\sqrt{\frac{\log n}{np}} + \Big(\frac{\log n}{npL}\Big)^{1/4} + \sqrt{\frac{(\log n)^3}{(np)^2}}\Big),$$

$$|\delta_{3,m}| \lesssim \varepsilon_m \Big|\frac{f^{(m)}(\theta_m^* | \theta_{-m}^*)}{g^{(m)}(\theta_m^* | \theta_{-m}^*)}\Big| + \frac{1}{\sqrt{npL}} \cdot \Big[\Big(\frac{\log n}{npL}\Big)^{1/4} + \frac{(\log n)^{7/4}}{(np)^{5/4}L^{1/4}}\Big].$$

Here $\varepsilon \in \mathbb{R}^n$ satisfies $\|\varepsilon\|_\infty = \mathfrak{o}(1)$. In particular, if $np \gg (\log n)^{3/2}$, we have $\max_{m \in [n]} |\delta_{1,m}| \vee |\delta_{2,m}| = \mathfrak{o}(1/\sqrt{npL})$ and $|\delta_{3,m}| \lesssim \varepsilon_m |f^{(m)}(\theta_m^* | \theta_{-m}^*)/g^{(m)}(\theta_m^* | \theta_{-m}^*)| + \mathfrak{o}(1/\sqrt{npL})$, where $\|\varepsilon\|_\infty = \mathfrak{o}(1)$ and $\mathfrak{o}(1/\sqrt{npL})$ is uniform over $m \in [n]$.

Proof of Lemma 6 (**Bound for** $\delta_{1,m}$) By Lemma 16, we have

$$|\delta_{1,m}| \lesssim (np)^{-1} \sum_{i:i\neq m} A_{mi} |\widehat{\theta}_i - \widehat{\theta}_i^{(m)}| \leq (np)^{-1} \Big(\sum_{i:i\neq m} A_{mi} \Big)^{1/2} \|\widehat{\theta}^{(m)} - \widehat{\theta}_{-m}\|$$

$$\lesssim (np)^{-1} \cdot \sqrt{np} \cdot \sqrt{\frac{1}{npL}} = \frac{1}{\sqrt{npL}} \sqrt{\frac{1}{np}}.$$

(**Bound for** $\delta_{2,m}$) Using Taylor expansion and the expansion in (5.11) for each $\widehat{\theta}_i^{(m)}$, $i \neq m$, we have

$$g^{(m)}(\theta_m^*|\widehat{\theta}_{-m})\delta_{2,m} = \sum_{i:i\neq m} A_{mi} \Big[\psi'(\theta_m^* - \theta_i^*)(\widehat{\theta}_i^{(m)} - \theta_i^*) - \frac{1}{2} \psi''(\theta_m^* - \xi_{m,i})(\widehat{\theta}_i^{(m)} - \theta_i^*)^2 \Big]$$

$$= - \sum_{i:i\neq m} A_{mi} \psi'(\theta_m^* - \theta_i^*) \frac{f^{(-m,i)}(\theta_i^*|\theta_{-m,-i}^*)}{g^{(-m,i)}(\theta_i^*|\theta_{-m,-i}^*)}$$

$$+ \sum_{i:i\neq m} A_{mi} \psi'(\theta_m^* - \theta_i^*)\delta_i^{(m)} - \frac{1}{2} \sum_{i:i\neq m} A_{mi} \psi''(\theta_m^* - \xi_{m,i})(\widehat{\theta}_i^{(m)} - \theta_i^*)^2$$

$$\equiv R_1^{(m)} + R_2^{(m)} + R_3^{(m)}.$$

The proof for $\delta_{2,m}$ is now complete by using Lemma 7 and the lower bound $g^{(m)}(\theta_m^*|\widehat{\theta}_{-m}) \gtrsim np$.

(**Bound for** $\delta_{3,m}$) It follows from Lemma 19 that

$$|\delta_{3,m}| = |\widehat{\theta}_m - \bar{\theta}_m| \lesssim \Big(\frac{|f^{(m)}(\theta_m^*|\widehat{\theta}_{-m})|}{np} \Big)^{3/2}.$$

By Lemma 18-(1), we have $\big|f^{(m)}(\theta_m^*|\widehat{\theta}_{-m})\big|/(np) \lesssim \sqrt{\log n/(npL)}$. Furthermore, we have by Lemma 18-(2)

$$\big|f^{(m)}(\theta_m^*|\widehat{\theta}_{-m})\big| = \big|f^{(m)}(\theta_m^*|\theta_{-m}^*)\big| + \mathscr{O}\Big(\sqrt{\frac{np}{L}} + \sqrt{\frac{(\log n)^3}{npL}} \Big).$$

Hence we have

$$|\delta_{3,m}| \lesssim \Big(\frac{\log n}{npL} \Big)^{1/4} \cdot \Big[\frac{|f^{(m)}(\theta_m^*|\theta_{-m}^*)|}{np} + \mathscr{O}\Big(\sqrt{\frac{1}{npL}} + \sqrt{\frac{(\log n)^3}{(np)^3 L}} \Big) \Big]$$

$$= \mathfrak{o}(1) \Big| \frac{f^{(m)}(\theta_m^*|\theta_{-m}^*)}{g^{(m)}(\theta_m^*|\theta_{-m}^*)} \Big| + \frac{1}{\sqrt{npL}} \cdot \Big[\Big(\frac{\log n}{npL} \Big)^{1/4} + \frac{(\log n)^{7/4}}{(np)^{5/4} L^{1/4}} \Big],$$

as desired. The proof is complete. □

LEMMA 7. Recall $R_1^{(m)}$-$R_3^{(m)}$ defined in the proof of Lemma 6. Suppose $\kappa = \mathcal{O}(1)$ and $np \gg \log n$. Then the following hold with probability $1 - \mathcal{O}(n^{-10})$.

$$\max_{m\in[n]} |R_1^{(m)}| \lesssim \sqrt{\frac{\log n}{L}}, \quad \max_{m\in[n]} |R_2^{(m)}| \lesssim \sqrt{\frac{\log n}{L}} + \frac{(np \log n)^{1/4}}{L^{3/4}} + \sqrt{\frac{(\log n)^3}{npL}},$$

$$\max_{m\in[n]} |R_3^{(m)}| \lesssim \frac{\sqrt{\log n}}{L} + \frac{(\log n)^2}{npL}.$$

Proof of Lemma 7 (**Bound of $R_1^{(m)}$**) Using a similar argument for the control of $R_{1,m}$ as in Lemma 2, we have $\max_{m\in[n]} |R_1^{(m)}| \lesssim \sqrt{\log n/L}$ with the prescribed probability.

(**Bound of $R_2^{(m)}$**) Note that by definition, $\delta^{(m)}$ is independent from data involving the mth individual. Hence, by applying Bernstein's inequality conditioning on data without the mth observation, we have

$$R_2^{(m)} = \sum_{i:i\neq m} A_{mi} \psi'(\theta_m^* - \theta_i^*) \delta_i^{(m)}$$

$$= p \cdot \sum_{i:i\neq m} \psi'(\theta_m^* - \theta_i^*) \delta_i^{(m)} + \sum_{i:i\neq m} (A_{mi} - p) \psi'(\theta_m^* - \theta_i^*) \delta_i^{(m)}$$

$$\lesssim p\sqrt{n}\|\delta^{(m)}\| + \sqrt{p \log n}\|\delta^{(m)}\| + \log n \cdot \|\delta^{(m)}\|_\infty$$

$$\overset{(*)}{\lesssim} p\sqrt{n} \cdot \frac{1}{\sqrt{pL}} \cdot \left[\sqrt{\frac{\log n}{np}} + \left(\frac{\log n}{npL}\right)^{1/4}\right] + \log n \cdot \sqrt{\frac{\log n}{npL}}$$

$$= \sqrt{\frac{\log n}{L}} + \frac{(np \log n)^{1/4}}{L^{3/4}} + \sqrt{\frac{(\log n)^3}{npL}},$$

with the prescribed probability. Here $(*)$ follows from Lemmas 4 and 5.

(**Bound for $R_3^{(m)}$**) By definition, we have

$$|R_3^{(m)}| \lesssim \left(\sum_{i:i\neq m} A_{mi}|\widehat{\theta}_i^{(m)} - \theta_i^*|\right) \cdot \|\widehat{\theta}^{(m)} - \theta_{-m}^*\|_\infty.$$

The first term satisfies, using the independence between $\{A_{mi}\}$ and $\widehat{\theta}^{(m)}$ and Bernstein's inequality,

$$\sum_{i:i\neq m} A_{mi}|\widehat{\theta}_i^{(m)} - \theta_i^*| = p \cdot \sum_{i:i\neq m} |\widehat{\theta}_i^{(m)} - \theta_i^*| + \sum_{i:i\neq m} (A_{mi} - p)|\widehat{\theta}_i^{(m)} - \theta_i^*|$$

$$\lesssim p\sqrt{n}\|\widehat{\theta}^{(m)} - \theta_{-m}^*\| + \log n \cdot \|\widehat{\theta}^{(m)} - \theta_{-m}^*\|_\infty$$

$$\overset{(*)}{\lesssim} p\sqrt{n}\sqrt{\frac{1}{pL}} + \log n\sqrt{\frac{\log n}{npL}} = \sqrt{\frac{np}{L}} + \sqrt{\frac{(\log n)^3}{npL}},$$

where $(*)$ follows from Lemma 16. Hence,

$$|R_3^{(m)}| \lesssim \left(\sqrt{\frac{np}{L}} + \sqrt{\frac{(\log n)^3}{npL}} \right) \cdot \sqrt{\frac{\log n}{npL}} = \frac{\sqrt{\log n}}{L} + \frac{(\log n)^2}{npL}.$$

This completes the proof.   □

6. Proof of the main expansion for spectral method

The goal of this section is to prove Theorem 3. We give a proof outline and introduce some preliminaries in Section 6.1, followed by two main steps of the proof in Sections 6.2 and 6.3, respectively. We then complete the proof in Section 6.4.

6.1 *Preliminary*

First note that by a simple Taylor expansion, we can identify the following main term of $\widetilde{\theta}_i - \theta_i^*$:

$$\Delta_i \equiv \frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} - \frac{1}{n} \sum_{k=1}^n \frac{\widehat{\pi}_k - \pi_k^*}{\pi_k^*}. \tag{6.1}$$

To find a close proxy of Δ that is tractable for analysis, we note the following property of $\widehat{\pi}_i$:

$$\widehat{\pi}_i = \frac{\sum_{j:j \neq i} A_{ij} \bar{y}_{ij} \widehat{\pi}_j}{\sum_{j:j \neq i} A_{ij} \bar{y}_{ji}},$$

which leads to the proxy choice (analogue of $\bar{\theta}$ for MLE)

$$\bar{\pi}_i \equiv \frac{\sum_{j:j \neq i} A_{ij} \bar{y}_{ij} \pi_j^*}{\sum_{j:j \neq i} A_{ij} \bar{y}_{ji}}. \tag{6.2}$$

The remainder vector δ for the spectral estimator is then defined by

$$\Delta_i = \frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} - \frac{1}{n} \sum_{k=1}^n \frac{\widehat{\pi}_k - \pi_k^*}{\pi_k^*} \equiv \frac{\bar{\pi}_i - \pi_i^*}{\pi_i^*} + \delta_i. \tag{6.3}$$

Now that the main term $(\bar{\pi}_i - \pi_i^*)/\pi_i^*$ is tractable for analysis, the goal is control $\|\delta\|_\infty$. Similar to the proof of its counterpart Proposition 10, an essential intermediate step is to give a tight bound for $\|\delta\|$, which we now discuss in detail.

6.2 *Control of $\|\delta\|$*

The goal of this subsection is to prove the following ℓ_2 bound of δ.

PROPOSITION 12. Suppose that $\kappa = \mathcal{O}(1)$ and $np \gg \log n$. Then it holds with probability $1 - \mathcal{O}(n^{-10})$ that $\|\delta\| = \mathfrak{o}(1/\sqrt{pL})$.

Proof of Proposition 12Fix $m \in [n]$. By the definitions of δ in (6.3) and $\bar{\pi}_m$ in (6.2), we have

$$\delta_m = \frac{\widehat{\pi}_m - \bar{\pi}_m}{\pi_m^*} - \frac{1}{n}\sum_{i=1}^n \frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} = \frac{\sum_{i:i\neq m} A_{im}\bar{y}_{mi}(\widehat{\pi}_i - \pi_i^*)}{\pi_m^* \cdot \sum_{i:i\neq m} A_{im}\bar{y}_{im}} - \frac{1}{n}\sum_{i=1}^n \frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*}.$$

Using again the expansion (6.3) for each $\widehat{\pi}_i - \pi_i^*$, we have

$$\delta_m = \frac{\sum_{i:i\neq m} A_{im}\bar{y}_{mi}(\bar{\pi}_i - \pi_i^*)}{\pi_m^* \cdot \sum_{i:i\neq m} A_{im}\bar{y}_{im}} + \frac{\sum_{i:i\neq m} A_{im}\bar{y}_{mi}\pi_i^*\delta_i}{\pi_m^* \cdot \sum_{i:i\neq m} A_{im}\bar{y}_{im}}$$

$$+ \frac{1}{n}\sum_{i=1}^n \frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} \cdot \frac{\sum_{i:i\neq m} A_{im}(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*)}{\pi_m^* \cdot \sum_{i:i\neq m} A_{im}\bar{y}_{im}}.$$

In matrix form, the above display is equivalent to $L\delta = R_1 + R_2$, where L is defined by $L_{ij} = -A_{ij}\bar{y}_{ij}\pi_j^*$ for $i \neq j$ and $L_{ii} = \sum_{j:j\neq i} A_{ij}\bar{y}_{ji}\pi_i^*$, and $R_1, R_2 \in \mathbb{R}^n$ are defined by

$$R_{1,m} = \sum_{i:i\neq m} A_{im}\bar{y}_{mi}(\bar{\pi}_i - \pi_i^*),$$

$$R_{2,m} = \frac{1}{n}\sum_{i=1}^n \frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} \cdot \sum_{i:i\neq m} A_{im}(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*).$$

Note that $\mathbb{E}(L_{ij}|A) = -A_{ij}\psi(\theta_i^* - \theta_j^*)\pi_j^* = -A_{ij}\psi(\theta_j^* - \theta_i^*)\pi_i^*$ and $\mathbb{E}(L_{ii}|A) = \sum_{j:j\neq i} A_{ij}\psi(\theta_j^* - \theta_i^*)\pi_i^*$, and hence $\mathbb{E}(L|A)$ is a symmetric Laplacian matrix. Hence,

$$\|L\delta\| \geq \|\mathbb{E}(L|A)\delta\| - \|L - \mathbb{E}(L|A)\|_{\text{op}}\|\delta\|$$

$$= \|\mathbb{E}(L|A)\big(\delta - \text{ave}(\delta)\mathbf{1}_n\big)\| - \|L - \mathbb{E}(L|A)\|_{\text{op}}\|\delta\|$$

$$\geq \lambda_{\min,\perp}(\mathbb{E}(L|A))\|\delta - \text{ave}(\delta)\mathbf{1}_n\| - \|L - \mathbb{E}(L|A)\|_{\text{op}}\|\delta\|$$

$$\geq \big[\lambda_{\min,\perp}(\mathbb{E}(L|A)) - \|L - \mathbb{E}(L|A)\|_{\text{op}}\big]_+ \|\delta\| - \sqrt{n} \cdot \lambda_{\min,\perp}(\mathbb{E}(L|A)) \cdot |\text{ave}(\delta)|.$$

Here for any $M \in \mathbb{R}^{n\times n}$, $\lambda_{\min,\perp}(M) \equiv \min_{x\in\mathbb{R}^n:\mathbf{1}_n^\top x=0,\|x\|=1} x^\top Mx$. By Lemma 8, $\|L - \mathbb{E}(L|A)\|_{\text{op}} \leq \lambda_{\min,\perp}(\mathbb{E}(L|A))/2$ and $\lambda_{\min,\perp}(\mathbb{E}(L|A)) \gtrsim p$ with the prescribed probability, hence rearranging the terms yields that

$$\|\delta\| \lesssim \sqrt{n}|\text{ave}(\delta)| + p^{-1} \cdot (\|R_1\| + \|R_2\|).$$

The proof is now complete by plugging in the estimates of $|\text{ave}(\delta)|$ in Lemma 9 and of $\|R_1\|$ and $\|R_2\|$ in Lemma 10. $\qquad\square$

LEMMA 8. Recall the matrix L defined in the proof of Proposition 12. Suppose that $\kappa = \mathscr{O}(1)$ and $np \gg \log n$, then with probability at least $1 - \mathscr{O}(n^{-10})$,

$$\lambda_{\min,\perp}(\mathbb{E}(L|A)) \geq cp, \quad \|L - \mathbb{E}(L|A)\|_{\mathrm{op}} \leq C\Big(\sqrt{\frac{p\log n}{nL}} \vee \frac{\log n}{n}\Big)$$

for some positive $c = c(\kappa)$ and $C = C(\kappa)$. Consequently, $\|L - \mathbb{E}(L|A)\|_{\mathrm{op}} \leq \lambda_{\min,\perp}(\mathbb{E}(L|A))/2$ for large enough n.

Proof of Lemma 8 Recall that $\mathbb{E}(L|A)$ is a symmetric Laplacian matrix with $\mathbb{E}(L_{ij}|A) = -A_{ij}\psi(\theta_j^* - \theta_i^*)\pi_i^*$ for $i \neq j$ and $\mathbb{E}(L_{ii}|A) = \sum_{j:j\neq i} A_{ij}\psi(\theta_j^* - \theta_i^*)\pi_i^*$. Hence, the first claim follows from Lemma 15 by noting that $\psi(\theta_j^* - \theta_i^*)\pi_i^* \asymp n^{-1}$.

Next we establish the concentration of L. Note that L can be written as $L = \sum_{i<j} S_{ij}$, where

$$S_{ij} \equiv A_{ij}\big(-\bar{y}_{ij}\pi_j^* \cdot e_i e_j^\top - \bar{y}_{ji}\pi_i^* \cdot e_j e_i^\top + \bar{y}_{ji}\pi_i^* \cdot e_i e_i^\top + \bar{y}_{ij}\pi_j^* \cdot e_j e_j^\top\big),$$

and S_{ij} are independent across $i < j$. Hence, by the Bernstein inequality for asymmetric matrices [38, Theorem 1.6], we have

$$\mathbb{P}\Big(\|L - \mathbb{E}(L|A)\|_{\mathrm{op}} \geq t|A\Big) \leq 2n \cdot \exp\Big(-\frac{Ct^2}{\sigma^2 + Rt}\Big) \tag{6.4}$$

for some universal $C > 0$, where R and σ^2 are such that

$$\max_{i<j}\|S_{ij}\|_{\mathrm{op}} \leq R, \quad \sigma^2 \equiv \max\Big\{\big\|\sum_{i<j}\mathbb{E}\big(S_{ij}S_{ij}^\top|A\big)\big\|_{\mathrm{op}}, \big\|\sum_{i<j}\mathbb{E}\big(S_{ij}^\top S_{ij}|A\big)\big\|_{\mathrm{op}}\Big\}.$$

Obviously, R can be taken to be $\mathscr{O}(n^{-1})$. For σ^2, we have by direct calculation that

$$\mathbb{E}\big(S_{ij}S_{ij}^\top|A\big) = \frac{A_{ij}}{L}\psi'(\theta_i^* - \theta_j^*)\big[(\pi_i^*)^2 + (\pi_j^*)^2\big](e_i e_i^\top + e_j e_j^\top - e_i e_j^\top - e_j e_i^\top).$$

This implies that, with $w_{ij} \equiv A_{ij}\psi'(\theta_i^* - \theta_j^*)\big[(\pi_i^*)^2 + (\pi_j^*)^2\big]/L$, $W \equiv \sum_{i<j}\mathbb{E}\big(S_{ij}S_{ij}^\top|A\big)$ is a Laplacian matrix with $W_{ij} = -w_{ij}$ for $i \neq j$ and $W_{ii} = \sum_{j:j\neq i} w_{ij}$. Hence, using the fact that $w_{ij}/A_{ij} \asymp (n^2 L)^{-1}$, Lemma 15 implies that $\|W\|_{\mathrm{op}} \lesssim (n^2 L)^{-1} \cdot (np) = p/(nL)$. A similar estimate for $\|\sum_{i<j}\mathbb{E}\big(S_{ij}^\top S_{ij}|A\big)\|_{\mathrm{op}}$ concludes that σ^2 can be taken to be $\mathscr{O}(p/(nL))$.

Finally, plugging the estimates of R and σ^2 into (6.4) yields that with the prescribed probability,

$$\|L - \mathbb{E}(L|A)\|_{\mathrm{op}} \lesssim \sqrt{\log n} \cdot \sigma + \log n \cdot R \lesssim \sqrt{\frac{p\log n}{nL}} \vee \frac{\log n}{n}.$$

The fact that $\|L - \mathbb{E}(L|A)\|_{\mathrm{op}} \leq \|\mathbb{E}(L|A)\|_{\mathrm{op}}/2$ (for large enough n) holds under the condition $np \gg \log n$. The proof is complete. $\square$

LEMMA 9. Suppose that $\kappa = \mathcal{O}(1)$ and $np \gg \log n$. Then, it holds with probability at least $1 - \mathcal{O}(n^{-10})$ that

$$|\mathrm{ave}(\delta)| \le C\sqrt{\frac{1}{npL}} \cdot \left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log n}{npL}}\right)$$

for some positive $C = C(\kappa)$. Consequently, $|\mathrm{ave}(\delta)| = \mathfrak{o}(1/\sqrt{npL})$.

Proof of Lemma 9 By the decomposition (6.3), we have

$$
\begin{aligned}
-\mathrm{ave}(\delta) = \mathrm{ave}\left(\frac{\bar{\pi} - \pi^*}{\pi^*}\right) &= \frac{1}{n}\sum_{m=1}^{m}\frac{\bar{\pi}_m - \pi_m^*}{\pi_m^*} \\
&= \frac{1}{n}\sum_{m=1}^{n}\frac{\sum_{i:i\neq m}A_{im}(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*)}{\pi_m^* \cdot \sum_{i:i\neq m}A_{im}\bar{y}_{im}} \\
&= \frac{1}{n}\sum_{m=1}^{n}\frac{\sum_{i:i\neq m}A_{im}(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*)}{\pi_m^* \cdot \sum_{i:i\neq m}A_{im}\psi(\theta_i^* - \theta_m^*)} \\
&\quad + \frac{1}{n}\sum_{m=1}^{n}\frac{\sum_{i:i\neq m}A_{im}(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*)}{\pi_m^*} \cdot \left(\frac{1}{\sum_{i:i\neq m}A_{im}\bar{y}_{im}} - \frac{1}{\sum_{i:i\neq m}A_{im}\psi(\theta_i^* - \theta_m^*)}\right) \\
&\equiv (I) + (II).
\end{aligned}
$$

To control (I), let $w_{im} \equiv A_{im}/\left(\pi_m^* \cdot \sum_{k:k\neq m}A_{km}\psi(\theta_k^* - \theta_m^*)\right)$. Then

$$n \cdot (I) = \sum_{i\neq m}w_{im}(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*) = \frac{1}{n}\sum_{i<m}(w_{im} - w_{mi})(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*).$$

It can be readily checked that the summands $\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*$ are independent across the indices $i < m$, centered, and sub-Gaussian with variance proxy bounded by $L^{-1}\|\pi^*\|_\infty^2 \lesssim (n^2 L)^{-1}$. Hence, conditioning on the graph A, Hoeffding's inequality yields that

$$\mathbb{P}(n \cdot |(I)| \ge t|A) \le 2\exp\left(-\frac{Ct^2}{(n^2 L)^{-1} \cdot \sum_{i<m}(w_{im} - w_{mi})^2}\right).$$

By Lemma 13 and the lower bound $\min_{m\in[n]}\pi_m^* \gtrsim n^{-1}$, we have $\sum_{i<m}(w_{im} - w_{mi})^2 \lesssim \sum_{i<m}(w_{im}^2 + w_{mi}^2) \lesssim n^2 p/p^2 = n^2/p$. Hence, by choosing $t \asymp \sqrt{\log n/(pL)}$, the above display yields that with the prescribed probability,

$$|(I)| \lesssim \frac{1}{n}\sqrt{\frac{\log n}{pL}} = \sqrt{\frac{1}{npL}} \cdot \sqrt{\frac{\log n}{n}}.$$

To control (*II*), using the lower bound $\min_{m \in [n]} \pi_m^* \gtrsim n^{-1}$ and standard concentration on the graph A, we have

$$
\begin{aligned}
|(II)| &\leq (np)^{-2} \times \max_{m \in [n]} \Big| \sum_{i:i\neq m} A_{im} \big(\bar{y}_{mi} \pi_i^* - \bar{y}_{im} \pi_m^* \big) \Big| \\
&\quad \times \sum_{m=1}^{n} \Big| \sum_{i:i\neq m} A_{im} (\bar{y}_{im} - \psi(\theta_i^* - \theta_m^*)) \Big| \\
&\leq (np)^{-2} \sqrt{n} \times \max_{m \in [n]} \Big| \sum_{i:i\neq m} A_{im} \big(\bar{y}_{mi} \pi_i^* - \bar{y}_{im} \pi_m^* \big) \Big| \\
&\quad \times \Big[\sum_{m=1}^{n} \Big| \sum_{i:i\neq m} A_{im} (\bar{y}_{im} - \psi(\theta_i^* - \theta_m^*)) \Big|^2 \Big]^{1/2} \\
&\overset{(*)}{\lesssim} (np)^{-2} \sqrt{n} \cdot \sqrt{\frac{p \log n}{nL}} \cdot \sqrt{\frac{n^2 p}{L}} = \sqrt{\frac{1}{npL}} \cdot \sqrt{\frac{\log n}{npL}},
\end{aligned}
$$

where $(*)$ follows from Lemma 14. The proof is complete. $\qquad\square$

LEMMA 10. Recall the definitions of R_1 and R_2 in the proof of Proposition 12. Suppose that $\kappa = \mathcal{O}(1)$ and $np \gg \log n$. Then the following holds with probability at least $1 - \mathcal{O}(n^{-10})$ for some $C = C(\kappa) > 0$:

$$
\|R_1\|_\infty \vee \|R_2\|_\infty \leq C \frac{\sqrt{\log n}}{n\sqrt{L}}, \quad \|R_1\| \vee \|R_2\| = \mathfrak{o}\Big(\sqrt{\frac{p}{L}}\Big).
$$

Proof of Lemma 10 It suffices to bound $\|R_1\|_\infty$ and $\|R_2\|_\infty$. We first bound $\|R_2\|_\infty$. By definition, we have

$$
\begin{aligned}
\|R_2\|_\infty^2 &\lesssim \Big(\sqrt{n} \|\widehat{\pi} - \pi^*\| \Big)^2 \cdot \max_{m \in [n]} \Big(\sum_{i:i\neq m} A_{im} (\bar{y}_{mi} \pi_i^* - \bar{y}_{im} \pi_m^*) \Big)^2 \\
&\overset{(*)}{\lesssim} \frac{1}{npL} \cdot \frac{p \log n}{nL} = \frac{\log n}{n^2 L} \cdot \frac{1}{L}.
\end{aligned}
$$

Here $(*)$ follows from Lemmas 17 and 14.

Next we bound $\|R_1\|_\infty$. By definition of $\bar{\pi}$ in (6.2), we have

$$
\begin{aligned}
R_{1,m} &= \sum_{i:i\neq m} A_{im}\bar{y}_{mi}\frac{\sum_{j:j\neq i}A_{ij}(\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^*)}{\sum_{j:j\neq i}A_{ij}\bar{y}_{ji}}\\
&= \sum_{i:i\neq m} A_{im}\psi(\theta_m^* - \theta_i^*)\frac{\sum_{j:j\neq i}A_{ij}(\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^*)}{\sum_{j:j\neq i}A_{ij}\psi(\theta_j^* - \theta_i^*)}\\
&\quad + \sum_{i:i\neq m} A_{im}\big(\bar{y}_{mi} - \psi(\theta_m^* - \theta_i^*)\big)\frac{\sum_{j:j\neq i}A_{ij}(\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^*)}{\sum_{j:j\neq i}A_{ij}\psi(\theta_j^* - \theta_i^*)}\\
&\quad + \sum_{i:i\neq m} A_{im}\bar{y}_{mi}\Big[\sum_{j:j\neq i}A_{ij}(\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^*)\Big(\frac{1}{\sum_{j:j\neq i}A_{ij}\bar{y}_{ji}} - \frac{1}{\sum_{j:j\neq i}A_{ij}\psi(\theta_j^* - \theta_i^*)}\Big)\Big]\\
&\equiv R_{1,m}^1 + R_{1,m}^2 + R_{1,m}^3.
\end{aligned}
$$

(**Bound of** $\|R_1^1\|_\infty$) Fix $m \in [n]$. Define $w_{ij} \equiv A_{im}A_{ij}\psi(\theta_m^* - \theta_i^*)/\big(\sum_{j:j\neq i}A_{ij}\psi(\theta_j^* - \theta_i^*)\big)$. Then $R_{1,m}^1$ can be written as

$$
\begin{aligned}
R_{1,m}^1 &= \sum_{i,j:i\neq m,i\neq j} w_{ij}\big(\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^*\big)\\
&= \sum_{i,j:i,j\neq m,i\neq j} w_{ij}\big(\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^*\big) + \sum_{i:i\neq m} w_{im}\big(\bar{y}_{im}\pi_m^* - \bar{y}_{mi}\pi_i^*\big)\\
&= \sum_{i,j:i,j\neq m,i<j} (w_{ij} - w_{ji})\big(\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^*\big) + \sum_{i:i\neq m} w_{im}\big(\bar{y}_{im}\pi_m^* - \bar{y}_{mi}\pi_i^*\big).
\end{aligned}
$$

Now conditioning on the graph A, $\{w_{ij}\}$ are deterministic and the above two sums are independent across their summands, which are sub-Gaussian with variance proxy bounded by $L^{-1}\|\pi^*\|_\infty^2 \lesssim (n^2L)^{-1}$. Hence, Hoeffding's inequality yields that with the prescribed probability,

$$
\Big|\sum_{i,j:i,j\neq m,i<j}(w_{ij} - w_{ji})\big(\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^*\big)\Big| \lesssim \sqrt{\frac{\log n}{n^2L}\cdot\sum_{i,j:i,j\neq m,i<j}(w_{ij} - w_{ji})^2},
$$

$$
\Big|\sum_{i:i\neq m} w_{im}\big(\bar{y}_{im}\pi_m^* - \bar{y}_{mi}\pi_i^*\big)\Big| \lesssim \sqrt{\frac{\log n}{n^2L}\cdot\sum_{i:i\neq m}w_{im}^2}.
$$

Now by Lemma 13, we have

$$\sum_{i,j:i,j\neq m,i<j}(w_{ij}-w_{ji})^2 \lesssim \sum_{i,j:i,j\neq m,i<j}\Big(\frac{A_{im}A_{ij}\psi(\theta_m^*-\theta_i^*)}{\sum_{j\neq i}A_{ij}\psi(\theta_j^*-\theta_i^*)}\Big)^2$$

$$\lesssim (np)^{-2}\sum_{i,j:i,j\neq m,i<j}A_{im}A_{ij} \lesssim 1,$$

$$\sum_{i:i\neq m}w_{im}^2 \lesssim \sum_{i:i\neq m}\Big(\frac{A_{im}\psi(\theta_m^*-\theta_i^*)}{\sum_{j\neq i}A_{ij}\psi(\theta_j^*-\theta_i^*)}\Big)^2 \lesssim (np)^{-1}.$$

Combining the two estimates yields that $\max_{m\in[n]}|R_{1,m}^1| \lesssim \sqrt{\log n/(n^2 L)}$.

(**Bound of** $\|R_1^2\|_\infty$) Fix $m \in [n]$. Note that $R_{1,m}^2$ can be further decomposed as

$$R_{1,m}^2 = \sum_{i:i\neq m}A_{im}(\bar{y}_{mi}-\psi(\theta_m^*-\theta_i^*))\frac{\sum_{j:j\neq i,m}A_{ij}(\bar{y}_{ij}\pi_j^*-\bar{y}_{ji}\pi_i^*)}{\sum_{j:j\neq i}A_{ij}\psi(\theta_j^*-\theta_i^*)}$$

$$+ \sum_{i:i\neq m}A_{im}(\bar{y}_{mi}-\psi(\theta_m^*-\theta_i^*))\frac{A_{im}(\bar{y}_{im}\pi_m^*-\bar{y}_{mi}\pi_i^*)}{\sum_{j:j\neq i}A_{ij}\psi(\theta_j^*-\theta_i^*)}$$

$$\equiv R_{1,m}^{2,1}+R_{1,m}^{2,2}.$$

To deal with $R_{1,m}^{2,1}$, note that

$$R_{1,m}^{2,1} = \sum_{i,j:i,j\neq m,i\neq j}\sum_{\ell=1}^{L}w_{ij}z_{ij\ell} = \sum_{i,j:i,j\neq m,i<j}\sum_{\ell=1}^{L}(w_{ij}-w_{ji})z_{ij\ell}$$

where

$$w_{ij} = \frac{A_{im}A_{ij}(\bar{y}_{mi}-\psi(\theta_m^*-\theta_i^*))}{\sum_{k:k\neq i}A_{ik}\psi(\theta_k^*-\theta_i^*)}, \quad z_{ij\ell} = \frac{1}{L}(y_{ij\ell}\pi_j^*-y_{ji\ell}\pi_i^*).$$

Now conditioning on A and all comparisons involving the mth individual, $\{w_{ij}\}$ is deterministic, and $z_{ij\ell}$ are independent across $i < j$ and sub-Gaussian with variance proxy of the order $(nL)^{-2}$; hence, Hoeffding's inequality applies to conclude that with the prescribed probability,

$$|R_{1,m}^{2,1}| \lesssim \sqrt{\frac{\log n}{n^2 L}\cdot\sum_{i,j:i,j\neq m,i\neq j}w_{ij}^2} + \frac{\log n}{nL}\cdot\max_{i,j:i,j\neq m,i\neq j}|w_{ij}|.$$

By definition, we have $\max_{i,j:i,j\neq m,i\neq j} |w_{ij}| \lesssim (np)^{-1}$ by Lemma 13, and by Lemma 14 we have

$$\sum_{i,j:i,j\neq m,i\neq j} w_{ij}^2 \lesssim (np)^{-2} \cdot np \cdot \sum_{i:i\neq m} A_{im}\big(\bar{y}_{mi} - \psi(\theta_m^* - \theta_i^*)\big)^2 \lesssim \frac{1}{L}.$$

Plugging in the above estimates yields that $\max_{m\in[n]} |R_{1,m}^{2,1}| \lesssim \sqrt{\log n}/(nL)$.

To deal with $R_1^{2,2}$, note that $R_{1,m}^{2,2} = \sum_{i:i\neq m} w_i z_i$, where

$$w_i = \frac{A_{im}}{\sum_{j:j\neq i} A_{ij}\psi(\theta_j^* - \theta_i^*)}, \quad z_i = \big(\bar{y}_{mi} - \psi(\theta_m^* - \theta_i^*)\big)\big(\bar{y}_{im}\pi_m^* - \bar{y}_{mi}\pi_i^*\big).$$

Further decompose

$$R_{1,m}^{2,2} = \sum_{i:i\neq m} w_i\mathbb{E}z_i + \sum_{i:i\neq m} w_i(z_i - \mathbb{E}z_i) \equiv (I) + (II).$$

Since $\mathbb{E}z_i = -(\pi_i^* + \pi_m^*)\psi'(\theta_m^* - \theta_i^*)/L$ so that $|\mathbb{E}z_i| \lesssim (nL)^{-1}$, Lemma 13 yields that with the prescribed probability,

$$|(I)| \lesssim \sum_{i:i\neq m} w_i|\mathbb{E}z_i| \lesssim (nL)^{-1} \cdot \sum_{i:i\neq m} \frac{A_{im}}{\sum_{j:j\neq i} A_{ij}\psi(\theta_j^* - \theta_i^*)} \lesssim (nL)^{-1}.$$

For (II), note that $\{z_i\}_{i:i\neq m}$ are independent, and each z_i is the product of two sub-Gaussian terms with variance proxies L^{-1} and $(n^2L)^{-1}$, respectively, so that z_i is sub-exponential with norm $K \lesssim (nL)^{-1}$. Therefore, Bernstein's inequality yields that

$$\mathbb{P}\Big(|(II)| \geq t\Big|A\Big) \leq \exp\Big(-\frac{Ct^2}{\|w\|^2 K^2 + \|w\|_\infty K}\Big).$$

By Lemma 13, we have $\|w\|_\infty \lesssim (np)^{-1}$ and $\|w\|^2 \lesssim (np)^{-1}$ with the prescribed probability. Hence, by choosing $t \asymp (nL)^{-1} \cdot \sqrt{\log n/(np)}$, the above estimate yields that $|(II)| \lesssim (nL)^{-1} \cdot \sqrt{\log n/(np)}$ with the prescribed probability. This concludes that $\|R_1^{2,2}\|_\infty \lesssim 1/(nL)$. Combining the estimates for $R_1^{2,1}$ and $R_1^{2,2}$ yields that $\|R_1^2\|_\infty \lesssim \sqrt{\log n}/(nL)$ with the prescribed probability.

(**Bound of** $\|R_1^3\|_\infty$) By definition, we have

$$|R_{1,m}^3| \lesssim (np)^{-2} \cdot \sum_{i:i\neq m} A_{im}\Big|\sum_{j:j\neq i} A_{ij}\big(\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^*\big)\Big|\Big|\sum_{j:j\neq i} A_{ij}\big(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\big)\Big|$$

$$\lesssim (np)^{-2}\Big(\sum_{i:i\neq m} A_{im}U_i^2\Big)^{1/2} \cdot \Big(\sum_{i:i\neq m} A_{im}V_i^2\Big)^{1/2},$$

where

$$U_i \equiv \sum_{j:j\neq i} A_{ij}(\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^*), \quad V_i \equiv \sum_{j:j\neq i} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)).$$

Now we bound $\sum_{i:i\neq m} A_{im}U_i^2$ for each $m \in [n]$. Since $\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^* = (\pi_i^* + \pi_j^*)(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))$, we have

$$\left(\sum_{i:i\neq m} A_{im}U_i^2\right)^{1/2} = \left[\sum_{i:i\neq m} A_{im}\left(\sum_{j:j\neq i} A_{ij}(\pi_i^* + \pi_j^*)(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))\right)^2\right]^{1/2}$$

$$= \sup_{u\in\mathscr{U}} \sum_{i:i\neq m} A_{im}u_i \cdot \sum_{j:j\neq i} A_{ij}(\pi_i^* + \pi_j^*)(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)),$$

where $\mathscr{U} \equiv \{\sum_{u\in\mathbb{R}^n} : \sum_{i:i\neq m} A_{im}u_i^2 \leq 1\}$ is defined conditioning on A. Let $\bar{\mathscr{U}}$ be a 1/2-covering of $\mathscr{U}$ in the sense that for any $u \in \mathscr{U}$, there exists some $u' \in \bar{\mathscr{U}}$ such that $\sqrt{\sum_{i:i\neq m} A_{im}(u_i - u_i')^2} \leq 1/2$. Since covering $\mathscr{U}$ is equivalent to covering the unit ball in dimension $\sum_{i:i\neq m} A_{im}$, we can find such a $\bar{\mathscr{U}}$ with $|\bar{\mathscr{U}}| \leq \exp(C\sum_{i:i\neq m} A_{im}) \leq \exp(C'np)$ with the prescribed probability. Then a standard covering argument yields that

$$\sqrt{\sum_{i:i\neq m} A_{im}U_i^2} \leq 2 \sup_{u\in\bar{\mathscr{U}}} \sum_{i:i\neq m} A_{im}u_i \cdot \sum_{j:j\neq i} A_{ij}(\pi_i^* + \pi_j^*)(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))$$

$$= 2 \sup_{u\in\bar{\mathscr{U}}} \left\{\frac{1}{L}\sum_{\ell=1}^{L} \sum_{i,j:i\neq m,j\neq m,i\neq j} A_{im}A_{ij}u_i(\pi_i^* + \pi_j^*)(y_{ij\ell} - \psi(\theta_i^* - \theta_j^*))\right.$$

$$\left. + \frac{1}{L}\sum_{\ell=1}^{L} \sum_{i:i\neq m} A_{im}u_i(\pi_i^* + \pi_m^*)(y_{im\ell} - \psi(\theta_i^* - \theta_m^*))\right\}.$$

Now by Hoeffding's inequality as in the analysis of R_1^1, we have for any fixed $u \in \bar{\mathscr{U}}$ and $t > 0$

$$\mathbb{P}\left(\left|\frac{1}{L}\sum_{\ell=1}^{L} \sum_{i,j:i\neq m,j\neq m,i\neq j} A_{im}A_{ij}u_i(\pi_i^* + \pi_j^*)(y_{ij\ell} - \psi(\theta_i^* - \theta_j^*))\right| \geq t\big|A\right)$$

$$\leq \exp\left(-\frac{CLt^2}{\sum_{i,j:i\neq m,j\neq m,i\neq j} A_{im}A_{ij}u_i^2(\pi_i^* + \pi_j^*)^2}\right).$$

Hence, by a union bound, Lemma 13, and choosing $t \asymp p/\sqrt{L}$, we have with the prescribed probability

$$\sup_{u\in\bar{\mathscr{U}}} \left|\frac{1}{L}\sum_{\ell=1}^{L} \sum_{i,j:i\neq m,j\neq m,i\neq j} A_{im}A_{ij}u_i(\pi_i^* + \pi_j^*)(y_{ij\ell} - \psi(\theta_i^* - \theta_j^*))\right| \lesssim \frac{p}{\sqrt{L}}.$$

Using a similar analysis, we have

$$
\sup_{u \in \widetilde{\mathscr{U}}} \left| \frac{1}{L} \sum_{i:i \neq m} A_{im} u_i (\pi_i^* + \pi_m^*)(y_{im\ell} - \psi(\theta_i^* - \theta_m^*)) \right| \lesssim \sqrt{\frac{p}{nL}} = \mathfrak{o}\left(\frac{p}{\sqrt{L}}\right).
$$

Putting together the two estimates yields that

$$
\max_{m \in [n]} \left(\sum_{i:i \neq m} A_{im} U_i^2 \right)^{1/2} \lesssim \frac{p}{\sqrt{L}}.
$$

A similar calculation yields that $\max_{m \in [n]} \left(\sum_{i:i \neq m} A_{im} V_i^2 \right)^{1/2} \lesssim np/\sqrt{L}$, so combining the estimates yields that $\max_{m \in [n]} |R_{1,m}^3| \lesssim (nL)^{-1}$ with the prescribed probability. Combining the estimates for $\|R_1^1\|_\infty$ - $\|R_1^3\|_\infty$ concludes the proof. □

6.3 *Entrywise expansion for the main term Δ*

Recall from Section 6.1 that Δ_i defined in (6.1) is the main term of $\widetilde{\theta}_i - \theta_i^*$. The goal of this subsection is to prove the following expansion.

PROPOSITION 13. Suppose that $\kappa = \mathscr{O}(1)$ and $np \gg (\log n)^{3/2}$. Then the following expansion holds with probability $1 - \mathscr{O}(n^{-10})$.

$$
\Delta_i = \left(1 + \varepsilon_{1,i} \right) \frac{\sum_{j:j \neq i} A_{ij}(\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^*)}{\pi_i^* \cdot \sum_{j:j \neq i} A_{ij}\psi(\theta_j^* - \theta_i^*)} + \varepsilon_{2,i},
$$

where $\varepsilon_1, \varepsilon_2 \in \mathbb{R}^n$ satisfy $\|\varepsilon_1\|_\infty = \mathfrak{o}(1)$ and $\|\varepsilon_2\|_\infty = \mathfrak{o}(1/\sqrt{npL})$.

As in Section 5.3 for the MLE, we will first perform some preliminary analysis in Section 6.3.1, and the main proof of Proposition 13 will be given in Section 6.3.2.

6.3.1 *Some preliminary analysis*
We first introduce a leave-one-out version of the spectral estimate $\widehat{\pi}$. Fix an index $m \in [n]$ to be left out, and define a new transition probability matrix $P^{(m)} \in \mathbb{R}^{n \times n}$ as follows. For off-diagonal elements, let

$$
P_{ij}^{(m)} \equiv \begin{cases} P_{ij}, & i \neq m \text{ and } j \neq m, \\ \mathbb{E}P_{ij} = \frac{p}{d}\psi(\theta_j^* - \theta_i^*), & i = m \text{ or } j = m. \end{cases}
$$

This leads to the choice of diagonal elements $P_{ii}^{(m)} \equiv 1 - \sum_{j:j \neq i} P_{ij}^{(m)}$. Note that in the case of $i = m$ or $j = m$, we have taken unconditional expectation of P_{ij} so that $P^{(m)}$ is independent of the data involving the mth individual, including the comparison indicators $\{A_{mi}\}_{i \neq m}$. With the above definition, let $\widehat{\pi}^{(m)}$ be the stationary measure of $P^{(m)}$, i.e. $\widehat{\pi}^{(m)}$ is defined by

$$
(\widehat{\pi}^{(m)})^\top P^{(m)} = (\widehat{\pi}^{(m)})^\top, \tag{6.5}
$$

which, after some similar manipulations for $\widehat{\pi}$, leads to

$$\widehat{\pi}_i^{(m)} = \frac{\sum_{j:j\neq i} \widehat{\pi}_j^{(m)} P_{ji}^{(m)}}{\sum_{j:j\neq i} P_{ij}^{(m)}} = \frac{\sum_{j:j\neq i,m} A_{ij}\bar{y}_{ij}\widehat{\pi}_j^{(m)} + \mathbb{I}_{i\neq m} \cdot p\psi(\theta_i^* - \theta_m^*)\widehat{\pi}_m^{(m)}}{\sum_{j:j\neq i,m} A_{ij}\bar{y}_{ji} + \mathbb{I}_{i\neq m} \cdot p\psi(\theta_m^* - \theta_i^*)}.$$

Note that different from the analogue $\widehat{\theta}^{(m)} \in \mathbb{R}^{n-1}$ in (5.9) for the MLE, $\widehat{\pi}^{(m)}$ is still in $\mathbb{R}^n$. Analogous to $\bar{\pi}$ in (6.2), define $\bar{\pi}^{(m)} \in \mathbb{R}^n$ as

$$\bar{\pi}_i^{(m)} = \frac{\sum_{j:j\neq i,m} A_{ij}\bar{y}_{ij}\pi_j^* + \mathbb{I}_{i\neq m} \cdot p\psi(\theta_i^* - \theta_m^*)\pi_m^*}{\sum_{j:j\neq i,m} A_{ij}\bar{y}_{ji} + \mathbb{I}_{i\neq m} \cdot p\psi(\theta_m^* - \theta_i^*)}.$$

Lastly, define the leave-one-out analogue of δ in (6.3) as

$$\frac{\widehat{\pi}_i^{(m)} - \pi_i^*}{\pi_i^*} - \frac{1}{n}\sum_{k=1}^n \frac{\widehat{\pi}_k^{(m)} - \pi_k^*}{\pi_k^*} \equiv \frac{\bar{\pi}_i^{(m)} - \pi_i^*}{\pi_i^*} + \delta_i^{(m)}. \tag{6.6}$$

Note that by construction, all of $\widehat{\pi}^{(m)}$, $\bar{\pi}^{(m)}$ and $\delta^{(m)}$ do not depend on the mth individual. We have the following estimates for $\delta^{(m)}$.

LEMMA 11. Suppose that $\kappa = \mathcal{O}(1)$ and $np \gg \log n$. Then the following holds with probability at least $1 - \mathcal{O}(n^{-10})$ for some $C = C(\kappa) > 0$.

$$\max_{m\in[n]} \|\delta^{(m)}\| = \mathfrak{o}(1/\sqrt{pL}) \quad \text{and} \quad \max_{m\in[n]} \|\delta^{(m)}\|_\infty \leq C\sqrt{\log n/(npL)}.$$

Proof of Lemma 11 The ℓ_2 bound follows from analogous arguments as in Proposition 12, except now using the estimate $\max_{m\in[n]} \|\widehat{\pi}^{(m)} - \pi^*\| \lesssim 1/\sqrt{n^2 pL}$ (instead of $\|\widehat{\pi} - \pi^*\| \lesssim 1/\sqrt{n^2 pL}$) in Lemma 10. The ℓ_∞ bound follows from the following simple estimate: for any $i \in [n]$,

$$|\delta_i^{(m)}| \lesssim \frac{\|\widehat{\pi}^{(m)} - \pi^*\|_\infty}{\pi_i^*} + \frac{1}{n}\sum_{k=1}^n \left|\frac{\widehat{\pi}_k^{(m)} - \pi_k^*}{\pi_k^*}\right| \lesssim \sqrt{\frac{\log n}{npL}},$$

using Lemma 17 in the last inequality. □

REMARK 5. In contrast to its counterpart Lemma 4 for the MLE, we do not need to analyze a leave-two-out version of $\widehat{\pi}$ here for the spectral estimator. The reason is that in the proof of Lemma 4, we need a leave-two-out analysis for the control of $g^{(-m,i)}(\theta_i^*|\widehat{\theta}_{-i}^{(m)})$ therein, which does not appear in the analysis of the spectral estimator. On the other hand, if we were to perform a leave-k-out analysis to improve the exponent in the regime (2.2), a corresponding leave-k-out version of $\widehat{\pi}$ would become necessary for the spectral method as well.

6.3.2 *Main proof of Proposition 10*

Proof of Proposition 13 Fix $m \in [n]$. Recall the definition of δ in (6.3). Then we have

$$
\begin{aligned}
\Delta_m &= \frac{\sum_{i:i \neq m} A_{mi}(\bar{y}_{mi} \pi_i^* - \bar{y}_{im} \pi_m^*)}{\pi_m^* \cdot \sum_{i:i \neq m} A_{mi} \bar{y}_{im}} + \delta_m \\
&= \frac{\sum_{i:i \neq m} A_{mi}(\bar{y}_{mi} \pi_i^* - \bar{y}_{im} \pi_m^*)}{\pi_m^* \cdot \sum_{i:i \neq m} A_{mi} \psi(\theta_i^* - \theta_m^*)} \Big(1 - \frac{\sum_{i:i \neq m} A_{mi}(\bar{y}_{im} - \psi(\theta_i^* - \theta_m^*))}{\sum_{i:i \neq m} A_{mi} \bar{y}_{im}}\Big) + \delta_m.
\end{aligned}
$$

Using standard concentration in Lemma 14, we have

$$
\max_{m \in [n]} \Big| \frac{\sum_{i:i \neq m} A_{mi}(\bar{y}_{im} - \psi(\theta_i^* - \theta_m^*))}{\pi_m^* \cdot \sum_{i:i \neq m} A_{mi} \bar{y}_{im}} \Big| \lesssim \frac{\sqrt{np \log n / L}}{np} = \sqrt{\frac{\log n}{npL}} = \mathfrak{o}(1).
$$

Hence, it remains to show that $\|\delta\|_\infty = \mathfrak{o}(1/\sqrt{npL})$ with the prescribed probability. To this end, recall the definition of $\widehat{\pi}^{(m)}$ in (6.5). Then we have

$$
\begin{aligned}
\delta_m &= \frac{\sum_{i:i \neq m} A_{im} \bar{y}_{mi}(\widehat{\pi}_i - \pi_i^*)}{\pi_m^* \cdot \sum_{i:i \neq m} A_{im} \bar{y}_{im}} - \frac{1}{n} \sum_{k=1}^n \frac{\widehat{\pi}_k - \pi_k^*}{\pi_k^*} \\
&= \Big(\frac{\sum_{i:i \neq m} A_{im} \bar{y}_{mi}(\widehat{\pi}_i - \widehat{\pi}_i^{(m)})}{\pi_m^* \cdot \sum_{i:i \neq m} A_{im} \bar{y}_{im}} - \frac{1}{n} \sum_{k=1}^n \frac{\widehat{\pi}_k - \widehat{\pi}_k^{(m)}}{\pi_k^*} \Big) \\
&\quad + \Big(\frac{\sum_{i:i \neq m} A_{im} \bar{y}_{mi}(\widehat{\pi}_i^{(m)} - \pi_i^*)}{\pi_m^* \cdot \sum_{i:i \neq m} A_{im} \bar{y}_{im}} - \frac{1}{n} \sum_{k=1}^n \frac{\widehat{\pi}_k^{(m)} - \pi_k^*}{\pi_k^*} \Big) \\
&\equiv \delta_{1,m} + \delta_{2,m}.
\end{aligned}
$$

Using Lemma 17, the two terms inside $\delta_{1,m}$ can be bounded as follows:

- The first term satisfies

$$
\begin{aligned}
\Big| \frac{\sum_{i:i \neq m} A_{im} \bar{y}_{mi}(\widehat{\pi}_i - \widehat{\pi}_i^{(m)})}{\pi_m^* \cdot \sum_{i:i \neq m} A_{im} \bar{y}_{im}} \Big| &\lesssim \frac{1}{p} \sum_{i:i \neq m} A_{im} |\widehat{\pi}_i - \widehat{\pi}_i^{(m)}| \\
&\leq \frac{1}{p} \Big(\sum_{i:i \neq m} A_{im} \Big)^{1/2} \cdot \|\widehat{\pi}^{(m)} - \widehat{\pi}\| \lesssim \frac{1}{p} \cdot \sqrt{np} \cdot n^{-1} \sqrt{\frac{\log n}{npL}} = \mathfrak{o}\Big(\frac{1}{\sqrt{npL}}\Big).
\end{aligned}
$$

- Using $\pi_m^* \gtrsim n^{-1}$, the second term satisfies

$$
\Big| \frac{1}{n} \sum_{k=1}^n \frac{\widehat{\pi}_k - \widehat{\pi}_k^{(m)}}{\pi_k^*} \Big| \lesssim \sqrt{n} \|\widehat{\pi}^{(m)} - \widehat{\pi}\| \lesssim \sqrt{\frac{\log n}{n}} \sqrt{\frac{1}{npL}} = \mathfrak{o}\Big(\frac{1}{\sqrt{npL}}\Big).
$$

We hence conclude $\|\delta_1\|_\infty = \mathfrak{o}(1/\sqrt{npL})$. For $\delta_{2,m}$, using the definition of $\delta^{(m)}$ in (6.6), we have

$$\Big(\pi_m^* \cdot \sum_{i:i\neq m} A_{im}\bar{y}_{im}\Big)\delta_{2,m} = \sum_{i:i\neq m} A_{im}\bar{y}_{mi}(\bar{\pi}_i^{(m)} - \pi_i^*) + \sum_{i:i\neq m} A_{im}\bar{y}_{mi}\pi_i^*\delta_i^{(m)} +$$

$$\sum_{i:i\neq m} A_{im}(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*) \cdot \frac{1}{n}\sum_{k=1}^n \frac{\widehat{\pi}_k^{(m)} - \pi_k^*}{\pi_k^*}$$

$$\equiv R_{1,m}^{(m)} + \sum_{i:i\neq m} A_{im}\bar{y}_{mi}\pi_i^*\delta_i^{(m)} + R_{2,m}^{(m)}.$$

By analogous arguments as in the proof of Lemma 10, we have

$$\max_{m\in[n]} \frac{|R_{1,m}^{(m)}| + |R_{2,m}^{(m)}|}{\pi_m^* \cdot \sum_{i:i\neq m} A_{im}\bar{y}_{im}} \lesssim \sqrt{\frac{\log n}{np}} \cdot \frac{1}{\sqrt{npL}} = \mathfrak{o}(\frac{1}{\sqrt{npL}}),$$

under the condition $np \gg \log n$. Lastly, for the middle term, by Lemma 11 and Bernstein's inequality applied conditionally to $\{A_{im}\}_{i\neq m}$ (note its independence from $\delta^{(m)}$ by construction),

$$\Big|\sum_{i:i\neq m} A_{im}\bar{y}_{mi}\pi_i^*\delta_i^{(m)}\Big| \leq \Big|p \cdot \sum_{i:i\neq m} \bar{y}_{mi}\pi_i^*\delta^{(m)}\Big| + \Big|\sum_{i:i\neq m}(A_{im} - p)\bar{y}_{mi}\pi_i^*\delta_i^{(m)}\Big|$$

$$\lesssim (p/n)\sqrt{n}\|\delta^{(m)}\| + \sqrt{p\log n \cdot \sum_{i:i\neq m}(\delta_i^{(m)}\pi_i^*)^2} + \log n \cdot n^{-1}\|\delta^{(m)}\|_\infty$$

$$= \mathfrak{o}(\sqrt{\frac{p}{nL}}) + \sqrt{\frac{p}{nL}} \cdot \frac{(\log n)^{3/2}}{np}.$$

Putting together the pieces, we conclude that $\|\delta\|_\infty = \mathfrak{o}(1/\sqrt{npL})$ under the condition $np \gg (\log n)^{3/2}$. The proof is complete. $\square$

6.4 *Completion of the proof*

Proof of Theorem 3 Fix $i \in [n]$. Recall the definition of Δ_i in (6.1). Then by definition of θ^* and the condition $\mathbf{1}_n^\top\theta^* = 0$, we have

$$\widetilde{\theta}_i - \theta_i^* = \big(\log\widehat{\pi}_i - \log\pi_i^*\big) - \text{ave}\big(\log\widehat{\pi} - \log\pi^*\big)$$

$$= \log\Big(\frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} + 1\Big) - \frac{1}{n}\sum_{k=1}^n \log\Big(\frac{\widehat{\pi}_k - \pi_k^*}{\pi_k^*} + 1\Big) \equiv \Delta_i + R_i.$$

Here with $h(x) \equiv \log(x + 1) - x$ for $x \geq -1$,

$$R_i = h\left(\frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*}\right) - \frac{1}{n}\sum_{k=1}^{n} h\left(\frac{\widehat{\pi}_k - \pi_k^*}{\pi_k^*}\right).$$

Define the event

$$\mathscr{A} \equiv \left\{ \max_{k \in [n]} \left|\frac{\widehat{\pi}_k - \pi_k^*}{\pi_k^*}\right| \leq \frac{1}{4} \right\}. \tag{6.7}$$

Then it follows from Lemma 17 that $\mathscr{A}$ holds with probability $1 - \mathscr{O}(n^{-10})$. On the event $\mathscr{A}$, using $|h(x)| \leq x^2$ for $x \in (-1/2, 1/2)$ and Lemma 17 again, we have

$$\frac{1}{n}\sum_{k=1}^{n} h\left(\frac{\widehat{\pi}_k - \pi_k^*}{\pi_k^*}\right) \leq \frac{1}{n}\sum_{k=1}^{n}\left(\frac{\widehat{\pi}_k - \pi_k^*}{\pi_k^*}\right)^2 \lesssim \frac{1}{npL} = \mathfrak{o}\left(\frac{1}{\sqrt{npL}}\right).$$

On the other hand, by definition of Δ_i in (6.1), we have

$$h\left(\frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*}\right) \leq \left(\frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*}\right)^2 = \frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} \cdot \left(\Delta_i + \frac{1}{n}\sum_{k=1}^{n}\frac{\widehat{\pi}_k - \pi_k^*}{\pi_k^*}\right)$$

$$= \mathfrak{o}(\Delta_i) + \mathfrak{o}\left(\frac{1}{\sqrt{npL}}\right)$$

by the estimates in Lemma 17, and thus, we have $|R_i| = \mathfrak{o}(\Delta_i) + \mathfrak{o}(1/\sqrt{npL})$. The proof is now completed by invoking Proposition 13. □

7. Proofs of Application I

In this section, we provide proofs for the application in Section 4.1.

Proofs of Propositions 4 *and* 5As the two proofs are similar, we only present the proof for the MLE. We start by proving the CLT with $\bar{\theta} = \theta^*$. By the expansion in Theorem 2, we have for each $i \in [n]$

$$\rho_i(\theta^*)(\widehat{\theta}_i - \theta_i^*) = (1 + \varepsilon_{1,i})f_i + \varepsilon_{2,i}'.$$

Here $\varepsilon_{2,i}' = \rho_i(\theta^*)\varepsilon_{2,i}$ satisfies $\|\varepsilon_{2,i}'\|_\infty = \mathfrak{o}(1)$ with probability $1 - \mathscr{O}(n^{-10})$ since standard concentration yields that $\max_{i \in [k]} \rho_i(\theta^*) = \mathscr{O}(\sqrt{npL})$ with the same probability, and the sequence $\{f_i\}$ is given by

$$f_i = \rho_i(\theta^*)\frac{b_i}{d_i} = \frac{\sqrt{L} \cdot \sum_{j:j \neq i} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))}{\sqrt{\sum_{j:j \neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)}}.$$

Let $Z \stackrel{d}{=} \mathcal{N}_k(0, I_k)$. By the Cramér–Wold theorem, it suffices to prove that for any fixed $a \in \mathbb{R}^k$,

$$a^\top (f_1, \ldots, f_k) \rightsquigarrow a^\top Z.$$

Let $\mathscr{A}_n$ be the event $\{A : \min_{m \in [n]} \sum_{i:i \neq m} A_{mi} \psi'(\theta_m^* - \theta_i^*) \geq c_0 np\}$ for some small enough constant c_0 such that $\mathscr{A}$ holds with probability $1 - \mathcal{O}(n^{-10})$ by Lemma 13. We first prove a conditional version of the above result by verifying the Lindeberg–Feller condition. With

$$U_{\ell i} \equiv \frac{a_\ell \sqrt{L} A_{\ell i} (\bar{y}_{\ell i} - \psi(\theta_\ell^* - \theta_i^*))}{\sqrt{\sum_{m:m \neq \ell} A_{\ell m} \psi'(\theta_\ell^* - \theta_m^*)}}.$$

we have

$$
\begin{aligned}
a^\top (f_1, \ldots, f_m) &= \sum_{\ell=1}^k a_\ell \frac{\sqrt{L} \cdot \sum_{i:i \neq \ell} A_{\ell i} [\bar{y}_{\ell i} - \psi(\theta_\ell^* - \theta_i^*)]}{\sqrt{\sum_{m:m \neq \ell} A_{\ell m} \psi'(\theta_\ell^* - \theta_m^*)}} \\
&\equiv \sum_{\ell=1}^k \sum_{i=1, i \neq \ell}^n U_{\ell i} = \sum_{\ell=1}^k \sum_{i=k+1}^n U_{\ell i} + \sum_{1 \leq \ell < i \leq k} (U_{\ell i} + U_{i\ell}) \\
&\equiv \sum_{\ell=1}^k \sum_{i=k+1}^n V_{\ell i} + \sum_{1 \leq \ell < i \leq k} V_{\ell i}.
\end{aligned}
$$

Here $\{V_{\ell i}\}$ with index set $\{\ell \leq k, k+1 \leq i \leq n\} \cup \{1 \leq \ell < i \leq k\}$ are defined by: if $\ell \leq k, k+1 \leq i \leq n$,

$$V_{\ell i} = U_{\ell i} = \frac{a_\ell \sqrt{L} A_{\ell i} (\bar{y}_{\ell i} - \psi(\theta_\ell^* - \theta_i^*))}{\sqrt{\sum_{m:m \neq \ell} A_{\ell m} \psi'(\theta_\ell^* - \theta_m^*)}},$$

and if $1 \leq \ell < i \leq k$,

$$V_{\ell i} = \sqrt{L} A_{\ell i} (\bar{y}_{\ell i} - \psi(\theta_\ell^* - \theta_i^*)) \cdot \left(\frac{a_\ell}{\sqrt{\sum_{m:m \neq \ell} A_{\ell m} \psi'(\theta_\ell^* - \theta_m^*)}} - \frac{a_i}{\sqrt{\sum_{m:m \neq i} A_{im} \psi'(\theta_i^* - \theta_m^*)}} \right).$$

Note that $\{V_{\ell i}\}$ are independent across its index set, centered when conditioning on A, and satisfy

$$
\begin{aligned}
&\left(\sum_{\ell=1}^k \sum_{i=k+1}^n + \sum_{1 \leq \ell < i \leq k} \right) \mathbb{E}(V_{\ell i}^2 | A) \\
&= \|a\|^2 + \sum_{1 \leq \ell < i \leq k} \frac{2 a_i a_\ell}{\sqrt{\sum_{m:m \neq \ell} A_{\ell m} \psi'(\theta_\ell^* - \theta_m^*)} \sqrt{\sum_{m:m \neq i} A_{im} \psi'(\theta_i^* - \theta_m^*)}} \\
&\rightarrow \|a\|^2 = \mathbb{E}(a^\top Z)^2,
\end{aligned}
$$

where the convergence holds under the condition $\min_{m \in [n]} \sum_{i:i \neq m} A_{mi} \psi'(\theta_m^* - \theta_i^*) \geq c_0 np$ on the event $\mathscr{A}_n$. On the other hand, the Lindeberg–Feller condition holds trivially since $\{V_{\ell,i}\}$ are bounded. Hence, the Lindeberg–Feller CLT applies to conclude that for any $A \in \mathscr{A}_n$,

$$a^\top (f_1, \ldots, f_k)|A \rightsquigarrow a^\top Z.$$

For the unconditional version, it holds for every $t \in \mathbb{R}$ that

$$\begin{aligned}
&\mathbb{P}(a^\top (f_1, \ldots, f_k) \leq t) \\
&= \mathbb{E}\Big[\mathbb{P}(a^\top (f_1, \ldots, f_k) \leq t|A)\mathbb{I}_{A \in \mathscr{A}_n}\Big] + \mathbb{E}\Big[\mathbb{P}(a^\top (f_1, \ldots, f_k) \leq t|A)\mathbb{I}_{A \in \overline{\mathscr{A}_n}}\Big] \\
&\to \mathbb{P}(a^\top Z \leq t) + 0 = \mathbb{P}(a^\top Z \leq t)
\end{aligned}$$

by the dominated convergence theorem. This concludes the claimed CLT for $\bar{\theta} = \theta^*$. The claim for $\bar{\theta} = \widehat{\theta}$ follows from the fact that $\rho_k(\widehat{\theta}) = (1 + \mathfrak{o}_P(1))\rho_k(\theta^*)$ uniformly over $k \in [n]$. Indeed, using the lower bound $\min_{i \in [n]} \rho_i(\theta^*) \wedge \rho_i(\widehat{\theta}) \gtrsim \sqrt{npL}$ with probability $1 - \mathscr{O}(n^{-10})$, we have

$$\begin{aligned}
\big|\rho_i(\theta^*) - \rho_i(\widehat{\theta})\big| &= \frac{\big|\rho_i^2(\theta^*) - \rho_i^2(\widehat{\theta})\big|}{\rho_i(\theta^*) + \rho_i(\widehat{\theta})} \lesssim \sqrt{\frac{L}{np}} \cdot \Big|\sum_{j:j \neq i} A_{ij}(\psi'(\theta_i^* - \theta_j^*) - \psi'(\widehat{\theta}_i - \widehat{\theta}_j))\Big| \\
&\lesssim \sqrt{\frac{L}{np}}\Big(\sum_{j:j \neq i} A_{ij}\Big) \cdot \|\widehat{\theta} - \theta^*\|_\infty \lesssim \sqrt{\frac{L}{np}} \cdot np \cdot \sqrt{\frac{\log n}{npL}} \\
&= \sqrt{\log n} = \mathfrak{o}(\sqrt{npL}).
\end{aligned}$$

The case where $\{A_{ij}\}$ are replaced by p can be dealt with similar arguments so the proof is complete. $\square$

8. Proofs of Application II

In this section, we provide proofs for the application in Section 4.2.

Proof of Proposition 6 It can be readily verified that the event $\{r(1) \notin [n_1 + 1, n - n_2]\}$ is contained in the event $\{\theta_i^* \notin \mathscr{C}_i \text{ for some } i\}$, whose probability can be bounded by

$$\mathbb{P}\Big(\{\theta_i^* \notin \mathscr{C}_i \text{ for some } i\}\Big) \leq \mathbb{P}(\theta_1^* \notin \mathscr{C}_1) + \sum_{i \neq 1} \mathbb{P}(\theta_i^* \notin \mathscr{C}_i).$$

Since $\mathbb{P}(\theta_1^* \notin \mathscr{C}_1) \to \alpha$ by construction of $\mathscr{C}_1$ and the CLT in Proposition 4, it suffices to show the second probability is vanishing.

Using preliminary estimates in Proposition 1, it is not hard to see that $\rho_i(\widehat{\theta})$ are uniformly close to their populations: $\rho_i(\widehat{\theta}) = (1 + \mathfrak{o}(1))\rho_i(\theta^*)$ with probability $1 - \mathscr{O}(n^{-10})$ and $\mathfrak{o}(1)$ uniform over $i \in [n]$. Note that $\rho_i(\theta^*) \asymp \sqrt{npL}$. Hence, by the expansion in Theorem 2, for each $i \neq 1$, we have for large

enough n

$$\mathbb{P}(\theta_i^* \notin \mathscr{C}_i) = \mathbb{P}\left(\left|(1+\varepsilon_{1,i})\frac{b_i}{d_i} + \varepsilon_{2,i}\right| \geq \tau_i\right)$$

$$\leq \mathbb{P}\left(\left|\frac{b_i}{d_i}\right| \geq \left(1+\frac{c_0}{2}\right)\sqrt{2\log n \cdot \rho_i^{-2}(\theta^*)}\right) + \mathscr{O}(n^{-10}).$$

Now note that $b_i/d_i = \sum_{j:j\neq i}\sum_{\ell=1}^{L} Z_{ij\ell}$, where $Z_{ij\ell} \equiv A_{ij}(y_{ij\ell} - \psi(\theta_i^* - \theta_j^*))/\left(L \cdot \sum_{k:k\neq i} A_{ik}\psi'(\theta_i^* - \theta_k^*)\right)$ are centered and independent when conditioning on A, and satisfy

$$\sum_{j:j\neq i}\sum_{\ell=1}^{L} \mathbb{E}(Z_{ij\ell}^2|A) = \rho_i^{-2}(\theta^*), \quad \max_{j:j\neq i}\max_{1\leq\ell\leq L} |Z_{ij\ell}| \leq 2\rho_i^{-2}(\theta^*).$$

Hence, for large enough n, it follows from Bernstein's inequality (with the prescribed constants in [2, Theorem 2.10]) applied conditionally on A that $\mathbb{P}(|b_i/d_i| \geq \widetilde{\tau}_i) \leq 2n^{-(1+c_0/2)}$. Take the union bound to complete the proof. $\qquad\square$

9. Proofs of Application III

In this section, we provide proofs for the application in Section 4.3.

9.1 *Proof of Proposition 5*

Proof of Proposition 7 Recall the definition of δ in (5.3), and the terms b_i, d_i in Theorem 2. Then for any $\varepsilon > 0$, with the prescribed probability, we have

$$\|\widehat{\theta} - \theta^*\|^2 \leq (1+\varepsilon)\sum_{m=1}^{n}\left(\frac{b_m}{d_m}\right)^2 + (1+\varepsilon^{-1})\|\delta\|^2.$$

By Proposition 11, we have $\|\delta\|^2 = \mathfrak{o}((pL)^{-1})$. For the main term, let $D \equiv \text{diag}(d_1,\ldots,d_n)$ be a diagonal matrix, which is invertible with the prescribed probability, and when this happens,

$$\sum_{m=1}^{n}\left(\frac{b_m}{d_m}\right)^2 = \|D^{-1}b\|^2 = \mathbb{E}(\|D^{-1}b\|^2|A) + (\|D^{-1}b\|^2 - \mathbb{E}(\|D^{-1}b\|^2|A)). \tag{9.1}$$

The second term in the above display can be bounded by Lemma 12 as follows. Note that $b_m = \sum_{i:i\neq m} A_{mi}Z_{mi}$ with $Z_{mi} = \bar{y}_{mi} - \psi(\theta_m^* - \theta_i^*)$ therein satisfying the anti-symmetry condition $Z_{mi} = -Z_{im}$. Moreover, up to anti-symmetry, Z_{mi} are independent sub-Gaussian variables with variance proxy bounded by a constant multiple (only depending on κ) of L^{-1}. Hence, with the prescribed probability,

Lemma 12 implies

$$\left(\|D^{-1}b\|^2 - \mathbb{E}\big(\|D^{-1}b\|^2|A\big)\right) \lesssim L^{-1}\sqrt{\log n}\Big(\min_{i\in[n]} d_i\Big)^{-2}\sqrt{n}\Big(\max_{i\in[n]}\sum_{j:j\neq i} A_{ij}\Big)$$

$$\lesssim \sqrt{\log n}\,\frac{\sqrt{n}np}{(np)^2 L} = \mathfrak{o}((pL)^{-1}),$$

where we use Lemma 13 in the second inequality. Next, for the first term in (9.1), it holds by direct calculation that

$$\mathbb{E}\big(\|D^{-1}b\|^2|A\big) = \frac{\text{tr}(D^{-1})}{L} \overset{(*)}{=} \big(1 + \mathfrak{o}(1)\big)\frac{\text{tr}\big((\mathbb{E}D)^{-1}\big)}{L}$$

$$= \big(1 + \mathfrak{o}(1)\big)\frac{1}{pL}\sum_{m=1}^{n}\frac{1}{\sum_{i:i\neq m}\psi'(\theta_m^* - \theta_i^*)},$$

where $(*)$ follows from the fact that uniformly over $i \in [n]$, $d_i \asymp np$ and $\|D - \mathbb{E}D\|_{\text{op}} = \max_i |d_i - \mathbb{E}d_i| \lesssim \sqrt{np\log n}$ by standard concentration. Putting together the two estimates, we have

$$\|\widehat{\theta} - \theta^*\|^2 \leq \frac{(1+\varepsilon)}{pL}(1 + \mathfrak{o}(1)) \cdot \sum_{i=1}^{n}\Big(\sum_{j:j\neq i}\psi'(\theta_i^* - \theta_j^*)\Big)^{-1} + C_\kappa'(1 + \varepsilon^{-1})\mathfrak{o}\Big(\frac{1}{pL}\Big).$$

The proof for the upper bound is now complete by choosing $\varepsilon \downarrow 0$ slow enough (e.g. $\varepsilon \asymp (\log n/n)^{1/4}$) such that the second term is still $\mathfrak{o}\big((pL)^{-1}\big)$. The proof for the lower bound is analogous by noting that for any $\varepsilon > 0$,

$$\|\widehat{\theta} - \theta^*\|^2 \geq (1-\varepsilon)\sum_{m=1}^{n}\Big(\frac{b_m}{d_m}\Big)^2 - (\varepsilon^{-1} - 1)\|\delta\|^2,$$

and then using the same estimates established in the upper bound proof. □

LEMMA 12. Let $A \in \{0,1\}^{n\times n}$ be a symmetric matrix and $D \in \mathbb{R}^{n\times n}$ be a diagonal matrix with positive diagonal entries $d_1, \ldots, d_n$. Define $b \in \mathbb{R}^n$ such that $b_i = \sum_{j:j\neq i} A_{ij}Z_{ij}$, where $\{Z_{ij}\}_{i,j:i<j}$ are a sequence of independent sub-Gaussian random variables with variance proxy τ^2 and $Z_{ij} = -Z_{ji}$ for any $i > j$. Then there exists some universal $C > 0$ such that

$$\big|\|D^{-1}b\|^2 - \mathbb{E}\|D^{-1}b\|^2\big| \leq C\tau^2\sqrt{\log n}\Big(\min_{i\in[n]} d_i\Big)^{-2}\sqrt{n}\Big(\max_{i\in[n]}\sum_{j:j\neq i} A_{ij}\Big),$$

with probability at least $1 - \mathcal{O}(n^{-10})$.

Proof. Using the symmetry of A and the anti-symmetry of $\{Z_{ij}\}_{i,j:i<j}$, we have

$$
\begin{aligned}
\|D^{-1}b\|^2 &= \sum_{i=1}^n \left(\frac{b_i}{d_i}\right)^2 = \sum_{i=1}^n d_i^{-2}\left(\sum_{j:j\neq i} A_{ij}Z_{ij}\right)^2 \\
&= \sum_{i=1}^n d_i^{-2}\left(\sum_{j:j>i} A_{ij}Z_{ij} + \sum_{j:j<i} A_{ji}(-Z_{ji})\right)^2 \\
&= \sum_{i=1}^n d_i^{-2}\left(\sum_{j,j':j,j'>i} A_{ij}A_{ij'}Z_{ij}Z_{ij'} + \sum_{j,j':j,j'<i} A_{ji}A_{j'i}Z_{ji}Z_{j'i} - \sum_{j,j':j,j'>i} A_{ij}A_{j'i}Z_{ij}Z_{j'i}\right),
\end{aligned} \tag{9.2}
$$

which can be written as a quadratic form of $\{Z_{ij}\}_{i,j:i<j}$. Let $\mathscr{I}$ be the set of double indices defined by $\mathscr{I} \equiv \{(i,j) : i,j \in [n], i < j\}$ with $N \equiv |\mathscr{I}| = \binom{n}{2}$. Let $Z \equiv (Z_{ij})_{(i,j)\in\mathscr{I}} \in \mathbb{R}^N$ such that it has independent sub-Gaussian coordinates with variance proxy τ^2. Then there exists a matrix $S = (S_{(i,j),(i',j')}) \in \mathbb{R}^{N\times N}$ such that $\|D^{-1}b\|^2 = Z^\top S Z = \sum_{(i,j),(i',j')\in\mathscr{I}} S_{(i,j),(i',j')}Z_{(i,j)}Z_{(i',j')}$ where S is defined as

$$
S_{(i,j),(i',j')} = \begin{cases}
A_{ij}A_{i'j'}(d_i^{-2} + d_j^{-2}), & \text{if } i' = i, j' = j, \\
A_{ij}A_{i'j'}d_i^{-2}, & \text{if } i' = i, j' \neq j, \\
A_{ij}A_{i'j'}d_j^{-2}, & \text{if } i' \neq i, j' = j, \\
-A_{ij}A_{i'j'}d_i^{-2}, & \text{if } j' = i, \\
-A_{ij}A_{i'j'}d_j^{-2}, & \text{if } i' = j, \\
0, & \text{otherwise.}
\end{cases}
$$

That is, the entry in S is zero if the two corresponding edges $(i,j), (i',j')$ have no joint vertex, as the matrix A can be interpreted as an adjacency matrix of a graph. Let $h_{\max} = \max_{i\in[n]} \sum_{j:j\neq i} A_{ij}$, which be understood as the maximum degree of the graph.

To apply the Hanson–Wright inequality (cf., [33, Theorem 1.1]) to control the deviation of $Z^\top S Z$, we now relate the quantities $\|S\|_F^2$ and $\|S\|_{\mathrm{op}}^2$. We first bound $\|S\|_F^2$. Denote $d_{\min} = \min_{i\in[n]} d_i$. We have

$$
\begin{aligned}
\|S\|_F^2 &\leq 4d_{\min}^{-4} \sum_{(i,j),(i',j')\in\mathscr{I}} A_{ij}A_{i'j'}\mathbb{I}\left\{\{i,j\}\cap\{i',j'\}\neq\emptyset\right\} \\
&\leq 8d_{\min}^{-4} \sum_{(i,j)\in\mathscr{I}} A_{ij}h_{\max} \leq 8d_{\min}^{-4}nh_{\max}^2.
\end{aligned} \tag{9.3}
$$

For $\|S\|_{\mathrm{op}}$, let $v = (v_{(i,j)}) \in \mathbb{R}^N$ be any vector such that $\|a\|^2 = \sum_{i<j} a_{(i,j)}^2 \leq 1$. Note that S is symmetric and in the above derivation of S we have (9.2) being equal to $Z^\top S Z$, which holds without any

condition on Z. Hence,

$$
\begin{aligned}
v^\top S v &= \sum_{i=1}^{n} d_i^{-2} \Big(\sum_{j:j>i} A_{ij} v_{ij} + \sum_{j:j<i} A_{ji}(-v_{ji}) \Big)^2 \\
&\leq 4 \sum_{i=1}^{n} d_i^{-2} \Big(\Big(\sum_{j:j>i} A_{ij} \Big) \Big(\sum_{j:j>i} v_{ij}^2 \Big) + \Big(\sum_{j:j<i} A_{ij} \Big) \Big(\sum_{j:j<i} v_{ij}^2 \Big) \Big) \\
&\leq 4 d_{\min}^{-2} h_{\max} \sum_{i=1}^{n} \Big(\Big(\sum_{j:j>i} v_{ij}^2 \Big) + \Big(\sum_{j:j<i} v_{ij}^2 \Big) \Big) \\
&\leq 8 d_{\min}^{-2} h_{\max}.
\end{aligned}
$$

Then we have

$$
\|S\|_{\mathrm{op}} \leq 8 d_{\min}^{-2} h_{\max}. \tag{9.4}
$$

Since each Z_{jk} is sub-Gaussian with variance proxy τ^2, the Hanson–Wright inequality now applies to conclude that for any $t \geq 0$,

$$
\mathbb{P}\big(|Z^\top S Z^\top - \mathbb{E}(Z^\top S Z^\top)| \geq t \big) \leq 2 \exp\Big(-C \frac{t^2}{\tau^4 \|S\|_F^2} \wedge \frac{t}{\tau^2 \|S\|_{\mathrm{op}}} \Big)
$$

for some universal $C > 0$. So by choosing $t = \tau^2 \cdot \mathcal{O}\big(\sqrt{\log n} \|S\|_F \vee (\log n) \|S\|_{\mathrm{op}} \big)$ and plugging in the estimates in (9.3) and (9.4), we obtain

$$
\begin{aligned}
|\|D^{-1}b\|^2 - \mathbb{E}\|D^{-1}b\|^2| &\lesssim \tau^2 \Big(\sqrt{\log n} \, d_{\min}^{-2} \sqrt{n} h_{\max} + (\log n) d_{\min}^{-2} h_{\max} \Big) \\
&\leq 2\tau^2 \sqrt{\log n} \, d_{\min}^{-2} \sqrt{n} h_{\max},
\end{aligned}
$$

with the prescribed probability. $\qquad\square$

9.2 *Proof of Proposition 6*

Proof of Proposition 8 We work on the event $\mathscr{A}$ defined in (6.7), which holds with the prescribed probability. Using $x - x^2 \leq \log(1+x) \leq x$ for $x \in (-1/2, 1/2)$ and the fact that

$$
\widetilde{\theta}_m - \theta_m^* = \log \Big(\frac{\widehat{\pi}_m - \pi_m^*}{\pi_m^*} + 1 \Big) - \frac{1}{k} \sum_{k=1}^{n} \log \Big(\frac{\widehat{\pi}_k - \pi_k^*}{\pi_k^*} + 1 \Big),
$$

we have (recall the definition of Δ in (6.1))

$$
\Delta_m - \Big(\frac{\widehat{\pi}_m - \pi_m^*}{\pi_m^*} \Big)^2 \leq \widetilde{\theta}_m - \theta_m^* \leq \Delta_m + \frac{1}{n} \sum_{i=1}^{n} \Big(\frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} \Big)^2.
$$

This implies that for any $\varepsilon > 0$,

$$\|\widetilde{\theta} - \theta^*\|^2 \leq (1 + \varepsilon)\|\Delta\|^2 +$$

$$\mathcal{O}(\varepsilon^{-1}) \cdot \max\left\{ \sum_{m=1}^{n} \left(\frac{\widehat{\pi}_m - \pi_m^*}{\pi_m^*}\right)^4, n\left(\frac{1}{n} \sum_{m=1}^{n} \left(\frac{\widehat{\pi}_m - \pi_m^*}{\pi_m^*}\right)^2\right)^2\right\}$$

$$= (1 + \varepsilon)\|\Delta\|^2 + \mathcal{O}(\varepsilon^{-1}) \cdot \sum_{m=1}^{n} \left(\frac{\widehat{\pi}_m - \pi_m^*}{\pi_m^*}\right)^4.$$

Using Lemma 17 and the lower bound $\min_{m \in [n]} \pi_m^* \gtrsim n^{-1}$, we have

$$\sum_{m=1}^{n} \left(\frac{\widehat{\pi}_m - \pi_m^*}{\pi_m^*}\right)^4 \leq n^4 \cdot \left\|\frac{\widehat{\pi} - \pi^*}{\pi^*}\right\|_\infty^2 \cdot \left\|\frac{\widehat{\pi} - \pi^*}{\pi^*}\right\|^2$$

$$\lesssim n^4 \cdot n^{-2} \frac{\log n}{npL} \cdot n^{-2} \frac{1}{pL} = \mathfrak{o}\left(\frac{1}{pL}\right).$$

Hence, the upper bound follows from Proposition 14 by choosing $\varepsilon \downarrow 0$ slow enough, e.g. of the order $\sqrt{\log n/(npL)}$. The lower bound proof is analogous so the proof is complete. $\square$

PROPOSITION 14. Recall the definition of Δ in (6.1). Suppose that $\kappa = \mathcal{O}(1)$ and $np \gg \log n$. Then the following holds with probability $1 - \mathcal{O}(n^{-10})$.

$$\|\Delta\|^2 = \frac{1 + \mathfrak{o}(1)}{pL} \cdot \sum_{m=1}^{n} \frac{\sum_{i:i \neq m}(e^{\theta_i^*} + e^{\theta_m^*})^2 \psi'(\theta_i^* - \theta_m^*)}{\left(\sum_{i:i \neq m}(e^{\theta_i^*} + e^{\theta_m^*})\psi'(\theta_i^* - \theta_m^*)\right)^2}$$

Proof of Proposition 14 Recall the definition of $\bar{\pi}_m$ in (6.2). By (6.3), we have

$$\|\Delta\|^2 \leq (1 + \varepsilon) \sum_{m=1}^{n} \left(\frac{\bar{\pi}_m - \pi_m^*}{\pi_m^*}\right)^2 + \mathcal{O}(\varepsilon^{-1})\|\delta\|^2.$$

By Proposition 12, we have $\|\delta\| = \mathfrak{o}(1/\sqrt{pL})$ with the prescribed probability. For the main term, we can further decompose it as

$$
\begin{aligned}
\frac{\bar{\pi}_m - \pi_m^*}{\pi_m^*} &= \frac{\sum_{i:i\neq m} A_{mi}\big(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*\big)}{\pi_m^* \cdot \sum_{i:i\neq m} A_{mi}\bar{y}_{im}} \\
&= \frac{\sum_{i:i\neq m} A_{mi}\big(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*\big)}{\pi_m^* \cdot \sum_{i:i\neq m} A_{mi}\psi\big(\theta_i^* - \theta_m^*\big)} + \\
&\qquad \sum_{i:i\neq m} \frac{A_{mi}\big(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*\big)}{\pi_m^*}\Big(\frac{1}{\sum_{i:i\neq m} A_{mi}\bar{y}_{im}} - \frac{1}{\sum_{i:i\neq m} A_{mi}\psi\big(\theta_i^* - \theta_m^*\big)}\Big) \\
&\equiv (I)_m + (II)_m.
\end{aligned}
$$

For $(II)_m$ in the above display, standard concentration and the lower bound $\min_{m\in[n]} \pi_m^* \gtrsim n^{-1}$ yield that

$$
|(II)_m| \lesssim n \cdot (np)^{-2} \cdot \Big| \sum_{i:i\neq m} A_{mi}\big(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*\big)\Big| \cdot \Big| \sum_{i:i\neq m} A_{mi}\big(\bar{y}_{im} - \psi(\theta_i^* - \theta_m^*)\big)\Big|.
$$

Hence, by Lemma 14,

$$
\begin{aligned}
\sum_{m=1}^{n} (II)_m^2 &\lesssim \frac{1}{n^2 p^4} \cdot \max_{m\in[n]} \Big| \sum_{i:i\neq m} A_{mi}\big(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*\big)\Big|^2 \times \\
&\qquad \sum_{m=1}^{n} \Big| \sum_{i:i\neq m} A_{mi}\big(\bar{y}_{im} - \psi(\theta_i^* - \theta_m^*)\big)\Big|^2 \\
&\lesssim \frac{1}{n^2 p^4} \cdot \frac{p\log n}{nL} \cdot \frac{n^2 p}{L} = \frac{\log n}{npL} \cdot \frac{1}{pL} = \mathfrak{o}\Big(\frac{1}{pL}\Big).
\end{aligned}
$$

For $(I)_m$, we have by definition of $\bar{\pi}_m$ that

$$
(I)_m = \frac{\sum_{i:i\neq m} A_{mi}\big(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*\big)}{\pi_m^* \cdot \sum_{i:i\neq m} A_{mi}\psi\big(\theta_i^* - \theta_m^*\big)} = \big(D^{-1}b\big)_m,
$$

where $D = \mathrm{diag}(\widetilde{d}_1, \ldots, \widetilde{d}_n)$ with $\widetilde{b}, \widetilde{d} \in \mathbb{R}^n$ defined in Theorem 3. This implies

$$
\|D^{-1}\widetilde{b}\|^2 = \mathbb{E}(\|D^{-1}\widetilde{b}\|^2 | A) + \big(\|D^{-1}\widetilde{b}\|^2 - \mathbb{E}(\|D^{-1}\widetilde{b}\|^2 | A)\big). \tag{9.5}
$$

The second term in the above display can be bounded by Lemma 12 as follows. Note that $\widetilde{b}_m = \sum_{i:i\neq m} A_{mi} Z_{mi}$ with $Z_{mi} = \bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*$, therein satisfying the anti-symmetry condition $Z_{mi} = -Z_{im}$. Moreover, up to anti-symmetry, Z_{mi} are independent sub-Gaussian variables with variance proxy bounded by a constant multiple (only depending on κ) of $(n^2 L)^{-1}$. Hence, with the prescribed

probability, Lemma 12 implies

$$\left(\|D^{-1}\widetilde{b}\|^2 - \mathbb{E}(\|D^{-1}\widetilde{b}\|^2|A)\right) \lesssim \frac{1}{n^2 L}\sqrt{\log n}\left(\min_{i\in[n]}\widetilde{d}_i\right)^{-2}\sqrt{n}\left(\max_{i\in[n]}\sum_{j:j\neq i}A_{ij}\right)$$

$$\lesssim \sqrt{\log n}\frac{\sqrt{n}np}{(p)^2 n^2 L} = \mathfrak{o}((pL)^{-1}),$$

where we use Lemma 13 in the second inequality.

Lastly, by direct calculation, the main term $\mathbb{E}(\|D^{-1}\widetilde{b}\|^2|A)$ in (9.5) satisfies

$$\mathbb{E}(\|D^{-1}\widetilde{b}\|^2|A) = \frac{1}{L}\sum_{m=1}^{n}\frac{\sum_{i:i\neq m}A_{im}(\pi_i^* + \pi_m^*)^2\psi'(\theta_i^* - \theta_m^*)}{\left(\pi_m^* \cdot \sum_{i:i\neq m}A_{im}\psi(\theta_i^* - \theta_m^*)\right)^2}$$

$$= \frac{1}{L}\sum_{m=1}^{n}\frac{\sum_{i:i\neq m}A_{im}(e^{\theta_i^*} + e^{\theta_m^*})^2\psi'(\theta_i^* - \theta_m^*)}{\left(\sum_{i:i\neq m}A_{im}(e^{\theta_i^*} + e^{\theta_m^*})\psi'(\theta_i^* - \theta_m^*)\right)^2}.$$

It remains to show that both the numerator and the denominator concentrate around their mean value with respect to A. For the denominator, it can be readily verified that $\mathbb{E}\left(\sum_{i:i\neq m}A_{im}(e^{\theta_i^*} + e^{\theta_m^*})\psi'(\theta_i^* - \theta_m^*)\right) \asymp np$, and Bernstein's inequality yields that with the prescribed probability,

$$\left|\sum_{i:i\neq m}(A_{im} - p)(e^{\theta_i^*} + e^{\theta_m^*})\psi'(\theta_i^* - \theta_m^*)\right| \lesssim \sqrt{np\log n} + \log n = \mathfrak{o}(np).$$

A similar calculation holds for the numerator. In summary, we have for any $\varepsilon > 0$,

$$\|\Delta\|^2 \leq (1+\varepsilon)\frac{(1+\mathfrak{o}(1))}{pL}\sum_{m=1}^{n}\frac{\sum_{i:i\neq m}(e^{\theta_i^*} + e^{\theta_m^*})^2\psi'(\theta_i^* - \theta_m^*)}{\left(\sum_{i:i\neq m}(e^{\theta_i^*} + e^{\theta_m^*})\psi'(\theta_i^* - \theta_m^*)\right)^2} + \mathcal{O}(\varepsilon^{-1})\mathfrak{o}(\frac{1}{pL}).$$

A lower bound can be similarly proved, and hence, the proof is complete by choosing $\varepsilon \downarrow 0$ slow enough (e.g. $\varepsilon \asymp (\log n/n)^{1/4}$). $\qquad\square$

9.3 *Proof of Lower bound (Theorem 3)*

Proof of Theorem 9 Let r_n be a sequence such that $(npL)^{-1/2} \ll r_n \ll \varepsilon_n \wedge 1$. For each $i \in [n]$, let $\pi_i(\cdot)$ be a prior density defined by $\pi_i(x) = r_n K\big((x - \theta_i^*)/r_n\big)$, where $K(\cdot)$ is a C^∞ kernel function supported on $[-1, 1]$ such that $K(\pm 1) = 0$.

By definition of $\Theta(\kappa)$, we have

$$\inf_{\widehat{\theta}}\sup_{\theta\in B(\theta^*,\varepsilon_n)\cap\Theta(\kappa)}\mathbb{E}_\theta\|\widehat{\theta} - \theta\|^2 = \inf_{\widehat{\theta}}\sup_{\theta\in B(\theta^*,\varepsilon_n)\cap\Theta(\kappa)}\mathbb{E}_\theta\|\big(\widehat{\theta} - \mathrm{ave}(\widehat{\theta})\mathbf{1}_n\big) - \big(\theta - \mathrm{ave}(\theta)\mathbf{1}_n\big)\|^2.$$

Then using the fact that $\|\theta - \mathrm{ave}(\theta)\mathbf{1}_n\|^2 = n^{-1} \cdot \sum_{i<j}(\theta_i - \theta_j)^2$ for any $\theta \in \mathbb{R}^n$, we have

$$\inf_{\widehat{\theta}} \sup_{\theta \in B(\theta^*, \varepsilon_n) \cap \Theta(\kappa)} \mathbb{E}_\theta \| (\widehat{\theta} - \mathrm{ave}(\widehat{\theta})\mathbf{1}_n) - (\theta - \mathrm{ave}(\theta)\mathbf{1}_n) \|^2$$

$$= \frac{1}{n} \inf_{\widehat{\theta}} \sup_{\theta \in B(\theta^*, \varepsilon_n) \cap \Theta(\kappa)} \sum_{i<j} \mathbb{E}_\theta \big[(\widehat{\theta}_i - \widehat{\theta}_j) - (\theta_i - \theta_j) \big]^2$$

$$\overset{(*)}{\geq} \frac{1}{n} \inf_{\widehat{\theta}} \sup_{\theta \in B(\theta^*, \varepsilon_n/2) \cap \widetilde{\Theta}(\kappa)} \sum_{i<j} \mathbb{E}_\theta \big[(\widehat{\theta}_i - \widehat{\theta}_j) - (\theta_i - \theta_j) \big]^2$$

$$\overset{(**)}{\geq} \frac{1}{n} \inf_{\widehat{\theta}} \int_{\theta \in [\theta^* - r_n \mathbf{1}_n, \theta^* + r_n \mathbf{1}_n]} \sum_{i<j} \mathbb{E}_\theta \big[(\widehat{\theta}_i - \widehat{\theta}_j) - (\theta_i - \theta_j) \big]^2 \Big(\prod_{i=1}^n \pi_i(\theta_i) \Big) \Big(\prod_{i=1}^n \mathrm{d}\theta_i \Big)$$

$$= \frac{1}{n} \inf_{\widehat{\theta}} \sum_{i<j} \int \Big(\prod_{k:k\neq i,j} \pi_k(\theta_k) \Big) \Big(\prod_{k:k\neq i,j} \mathrm{d}\theta_k \Big) \Big[\int \mathbb{E}_\theta \big[(\widehat{\theta}_i - \widehat{\theta}_j) - (\theta_i - \theta_j) \big]^2 \pi_i(\theta_i)\pi_j(\theta_j) \mathrm{d}\theta_i \mathrm{d}\theta_j \Big]$$

$$\overset{(***)}{\geq} \frac{1}{n} \sum_{i<j} \int \Big(\prod_{k:k\neq i,j} \pi_k(\theta_k) \Big) \Big(\prod_{k:k\neq i,j} \mathrm{d}\theta_k \Big) \Big[\inf_{\widetilde{T}_{ij}} \int \mathbb{E}_\theta \big[\widetilde{T}_{ij} - (\theta_i - \theta_j) \big]^2 \pi_i(\theta_i)\pi_j(\theta_j) \mathrm{d}\theta_i \mathrm{d}\theta_j \Big]$$

$$\equiv \frac{1}{n} \sum_{i<j} \int \Big(\prod_{k:k\neq i,j} \pi_k(\theta_k) \Big) R_{ij}(\theta_{-i,-j}) \Big(\prod_{k:k\neq i,j} \mathrm{d}\theta_k \Big). \tag{9.6}$$

Here $(*)$ holds as follows with $\widetilde{\Theta}(\kappa) \equiv \{\theta' \in \mathbb{R}^n : \max_i \theta'_i - \min_i \theta'_i \leq \kappa\}$ dropping the centering condition. For any $\widehat{\theta}$, with

$$\theta^* \equiv \mathrm{argmax}_{\theta \in B(\theta^*, \varepsilon_n/2) \cap \widetilde{\Theta}(\kappa)} \mathbb{E}_\theta \big[(\widehat{\theta} - \mathrm{ave}(\widehat{\theta})\mathbf{1}_n) - (\theta - \mathrm{ave}(\theta)\mathbf{1}_n) \big]^2$$

and $\widetilde{\theta} \equiv \theta^* - \mathrm{ave}(\theta^*)\mathbf{1}_n$, we have by the translation invariance of the model

$$\mathbb{E}_{\theta^*} \big[(\widehat{\theta} - \mathrm{ave}(\widehat{\theta})\mathbf{1}_n) - (\theta^* - \mathrm{ave}(\theta^*)\mathbf{1}_n) \big]^2 = \mathbb{E}_{\widetilde{\theta}} \big[(\widehat{\theta} - \mathrm{ave}(\widehat{\theta})\mathbf{1}_n) - (\widetilde{\theta} - \mathrm{ave}(\widetilde{\theta})\mathbf{1}_n) \big],$$

and since $\theta^* \in B(\theta^*, \varepsilon_n/2)$ and $\theta^* \in \Theta(\kappa)$,

$$|\mathrm{ave}(\theta^*)| = \Big| \frac{1}{n} \sum_{i=1}^n (\theta_i^* - \theta_i^*) \Big| \leq \|\theta^* - \theta^*\|_\infty \leq \varepsilon_n/2.$$

This entails that $\|\widetilde{\theta} - \theta^*\|_\infty \leq \varepsilon_n$, and hence, $\widetilde{\theta} \in B(\theta^*, \varepsilon_n) \cap \Theta(\kappa)$, which concludes the proof of $(*)$. Next $(**)$ follows with $[\theta^* - r_n \mathbf{1}_n, \theta^* + r_n \mathbf{1}_n] \equiv \times_{i=1}^n [\theta_i^* - r_n, \theta_i^* + r_n]$ by bounding the (local) maximal risk by the Bayes risk induced by the product prior $\prod_{i=1}^n \pi_i(\theta_i)$ on $[\theta^* - r_n \mathbf{1}_n, \theta^* + r_n \mathbf{1}_n]$. Note that in $(**)$ we have $B(\theta^*, \varepsilon_n/2) \subset \widetilde{\Theta}(\kappa)$ for large enough n due to the condition $\max_i \theta_i^* - \min_i \theta_i^* \leq \kappa/2$. Lastly, the infimum over $\widetilde{T}_{ij}$ in $(***)$ is taken over all estimators of $\theta_i - \theta_j$ with the knowledge of $\theta_{-i,-j}$.

Let $\mu \equiv (\theta_i, \theta_j) \in \mathbb{R}^2$ and $T_{ij} \equiv T_{ij}(\mu) \equiv \theta_i - \theta_j \in \mathbb{R}$. Now by the multivariate van Tree's inequality in [13, Equation (11)], each $R_{ij}(\theta_{-i}, \theta_{-j})$ can be lower bounded by

$$R_{ij}(\theta_{-i}, \theta_{-j}) \geq \frac{1}{n} \int \mathscr{J}_{ij}(\mu; \theta_{-i,-j}) \pi_i(\theta_i) \pi_j(\theta_j) d\theta_i d\theta_j - \frac{1}{n^2} \widetilde{\mathscr{I}}_{ij}(\pi_i, \pi_j), \tag{9.7}$$

where with $\mathscr{I}(\mu)$ being the Fisher information matrix of $\mu = (\theta_i, \theta_j)$ defined below:

$$\mathscr{J}_{ij}(\mu; \theta_{-i,-j}) \equiv \Big(\frac{\partial T_{ij}}{\partial \mu}\Big)^{\top} \mathscr{I}(\mu)^{-1} \Big(\frac{\partial T_{ij}}{\partial \mu}\Big),$$

and $\widetilde{\mathscr{I}}_{ij}(\pi_i, \pi_j)$ is also defined below after we introduce a few other notations. We now estimate the two terms on the right side of (9.7).

(**First term in (9.7)**) The fisher information of the parameter μ can be calculated as

$$\mathscr{I}(\mu) \equiv \mathbb{E}_{\mu}\Big[\Big(\frac{\partial \ell_n(\mu; \theta_{-i,-j})}{\partial \mu}\Big)^{\top} \Big(\frac{\partial \ell_n(\mu; \theta_{-i,-j})}{\partial \mu}\Big)\Big]$$

$$= (pL) \cdot \begin{bmatrix} \sum_{k:k\neq i} \psi'(\theta_i - \theta_k) & -\psi'(\theta_i - \theta_j) \\ -\psi'(\theta_i - \theta_j) & \sum_{k:k\neq j} \psi'(\theta_j - \theta_k) \end{bmatrix}.$$

This entails that, with $D_{ij} \equiv \sum_{k:k\neq i,j} \big(\psi'(\theta_i - \theta_k) + \psi'(\theta_j - \theta_k)\big)$,

$$\mathscr{J}_{ij}(\mu; \theta_{-i,-j}) \equiv \Big(\frac{\partial T_{ij}}{\partial \mu}\Big)^{\top} \mathscr{I}(\mu)^{-1} \Big(\frac{\partial T_{ij}}{\partial \mu}\Big)$$

$$= \frac{1}{pL} \cdot \frac{D_{ij}}{\big(\sum_{k:k\neq i,j} \psi'(\theta_i - \theta_k)\big)\big(\sum_{k:k\neq i,j} \psi'(\theta_j - \theta_k)\big) + \psi'(\theta_i - \theta_j)D_{ij}}$$

$$= (1 + \mathfrak{o}(1)) \cdot \frac{1}{pL} \cdot \frac{D_{ij}}{\big(\sum_{k:k\neq i,j} \psi'(\theta_i - \theta_k)\big)\big(\sum_{k:k\neq i,j} \psi'(\theta_j - \theta_k)\big)}$$

$$= (1 + \mathfrak{o}(1)) \cdot \frac{1}{pL} \cdot \Big(\frac{1}{\sum_{k:k\neq i,j} \psi'(\theta_i - \theta_k)} + \frac{1}{\sum_{k:k\neq i,j} \psi'(\theta_j - \theta_k)}\Big),$$

where the $\mathfrak{o}(1)$ term is uniform over all $\theta \in \Theta(\kappa)$ and i, j. Moreover, by definition of $\pi_i(\cdot)$ and the fact that $r_n = \mathfrak{o}(1)$, direct calculation yields that

$$\int \frac{1}{\sum_{k:k\neq i,j} \psi'(\theta_i - \theta_k)} \pi_i(\theta_i) \pi_j(\theta_j) d\theta_i d\theta_j$$

$$= \int \frac{1}{\sum_{k:k\neq i,j} \psi'(\theta_i - \theta_k)} K_{r_n}(\theta_i - \theta_i^*) d\theta_i = \frac{1 + \mathfrak{o}(1)}{\sum_{k:k\neq i,j} \psi'(\theta_i^* - \theta_k)}$$

uniformly overall $\theta \in \Theta(\kappa)$ and i, j. With a similar bound for the term involving index j, we have

$$
\int \mathscr{J}_{ij}(\mu; \theta_{-i,-j}) \pi_i(\theta_i) \pi_j(\theta_j) \mathrm{d}\theta_i \mathrm{d}\theta_j
$$
$$
= (1 + \mathfrak{o}(1)) \cdot \frac{1}{pL} \cdot \left(\frac{1}{\sum_{k:k\neq i,j} \psi'(\theta_i^* - \theta_k)} + \frac{1}{\sum_{k:k\neq i,j} \psi'(\theta_j^* - \theta_k)} \right). \tag{9.8}
$$

Here the $\mathfrak{o}(1)$ is uniform overall $\theta \in \Theta(\kappa)$ and i, j.

(**Second term in (9.7)**) On the other hand, let $\Delta \equiv \det \left((pL)^{-1} \mathscr{I}(\mu) \right)$. Then

$$
C(\mu) \equiv \left(\frac{\partial T_{ij}(\mu)}{\partial \mu} \right)^\top \mathscr{I}^{-1}(\mu) = \frac{1}{pL\Delta} \left(\sum_{k:k\neq i,j} \psi'(\theta_j - \theta_k), -\sum_{k:k\neq i,j} \psi'(\theta_i - \theta_k) \right).
$$

Hence, by definition of $\pi_i(\cdot)$ and $\pi_j(\cdot)$, we have

$$
\frac{\partial \left(C_{11}(\mu) \pi_i(\theta_i) \pi_j(\theta_j) \right)}{\partial \theta_i}
$$
$$
= \frac{\left(\sum_{k:k\neq i,j} \psi'(\theta_j - \theta_k) \right) \cdot K_{r_n}(\theta_j - \theta_j^*)}{pL\Delta^2} \cdot \left[r_n^{-1} K'_{r_n}(\theta_i - \theta_i^*) \Delta - K_{r_n}(\theta_i - \theta_i^*) \frac{\partial \Delta}{\partial \theta_i} \right].
$$

With a similar bound for $\partial \left(C_{12}(\mu) \pi_i(\theta_i) \pi_j(\theta_j) \right) / \partial \theta_j$, and the simple bounds that $\Delta \gtrsim n^2$ and $|\partial \Delta / \partial \theta_i| \vee |\partial \Delta / \partial \theta_j| = \mathcal{O}(n^2)$ uniformly over all $\theta \in \Theta(\kappa)$ and i, j, we have

$$
\widetilde{\mathscr{I}}_{ij}(\pi_i, \pi_j) \equiv \int \frac{1}{\pi_i(\theta_i) \pi_j(\theta_j)} \left[\frac{\partial \left(C_{11}(\mu) \pi_i(\theta_i) \pi_j(\theta_j) \right)}{\partial \theta_i} + \frac{\partial \left(C_{12}(\mu) \pi_i(\theta_i) \pi_j(\theta_j) \right)}{\partial \theta_j} \right]^2 \mathrm{d}\theta_i \mathrm{d}\theta_j
$$
$$
\lesssim \frac{1}{n^2 p^2 L^2} \left(1 \vee \frac{1}{r_n^2} \int \frac{(K'(x))^2}{K(x)} \mathrm{d}x \right) \lesssim \frac{1}{n^2 p^2 L^2 r_n^2}. \tag{9.9}
$$

Now apply (9.7) with (9.8) and (9.9) to yield that

$$
R_{ij}(\theta_{-i,-j}) \geq (1 + \mathfrak{o}(1)) \cdot \frac{1}{pL} \cdot \left(\frac{1}{\sum_{k:k\neq i,j} \psi'(\theta_i^* - \theta_k)} + \frac{1}{\sum_{k:k\neq i,j} \psi'(\theta_j^* - \theta_k)} \right)
$$

under the condition $r_n \ll (npL)^{-1/2}$. By (9.6) and the above lower bound, the local minimax risk is lower bounded by

$$
\frac{1 + \mathfrak{o}(1)}{npL} \cdot \sum_{i<j} \int \left(\frac{1}{\sum_{k:k\neq i,j} \psi'(\theta_i^* - \theta_k)} + \frac{1}{\sum_{k:k\neq i,j} \psi'(\theta_j^* - \theta_k)} \right) \left(\prod_{k:k\neq i,j} \pi_k(\theta_k) \right) \mathrm{d}\theta_{-i,-j},
$$

where the $\mathfrak{o}(1)$ is uniform over all i, j. Note that for any $\theta_{-i,-j}$ in the support of the product prior, we have

$$\sum_{k:k \neq i,j} \psi'(\theta_i^* - \theta_k) = (1 + \mathfrak{o}(1)) \sum_{k:k \neq i,j} \psi'(\theta_i^* - \theta_k^*) \quad \text{as} r_n = \mathfrak{o}(1),$$

where the $\mathfrak{o}(1)$ is uniform over all i, j. Hence, the local minimax risk is further lower bounded by

$$\frac{1 + \mathfrak{o}(1)}{npL} \cdot \sum_{i<j} \left(\frac{1}{\sum_{k:k \neq i,j} \psi'(\theta_i^* - \theta_k^*)} + \frac{1}{\sum_{k:k \neq i,j} \psi'(\theta_j^* - \theta_k^*)} \right)$$

$$= \frac{1 + \mathfrak{o}(1)}{npL} \cdot \sum_{i<j} \left(\frac{1}{\sum_{k:k \neq i} \psi'(\theta_i^* - \theta_k^*)} + \frac{1}{\sum_{k:k \neq j} \psi'(\theta_j^* - \theta_k^*)} \right)$$

$$= \frac{1 + \mathfrak{o}(1)}{pL} \cdot \sum_{i=1}^{n} \frac{1}{\sum_{k:k \neq i} \psi'(\theta_i^* - \theta_k^*)},$$

as desired. □

Data Availability Statement

No new data were generated or analyzed in support of this review.

Acknowledgments

The authors thank the anonymous reviewers and the AE for their valuable suggestions.

Funding

REFERENCES

1. BENAYCH-GEORGES, F., BORDENAVE, C. & KNOWLES, A. (2020) Spectral radii of sparse random matrices. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, **56**, 2141–2161.
2. BOUCHERON, S., LUGOSI, G. & MASSART, P. (2013) *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford: Oxford University Press With a foreword by Michel Ledoux.
3. CHATTERJEE, S., DIACONIS, P. & SLY, A. (2011) Random graphs with a given degree sequence. *Ann. Appl. Probab.*, **21**, 1400–1435.
4. CHEN, P., GAO, C. & ZHANG, A. Y. (2022) Optimal full ranking from pairwise comparisons. *Ann. Stat.*, **50**, 1775–1805.
5. CHEN, P., GAO, C. & ZHANG, A. Y. (2022) Partial recovery for top-k ranking: optimality of MLE and suboptimality of the spectral method. *Ann. Stat.*, **50**, 1618–1652.
6. CHEN, Y., FAN, J., MA, C. & WANG, K. (2019) Spectral method and regularized MLE are both optimal for top-*K* ranking. *Ann. Statist.*, **47**, 2204–2235.

7. CHEN, Y., FAN, J., MA, C. & YAN, Y. (2019) Inference and uncertainty quantification for noisy matrix completion. *Proc. Natl. Acad. Sci. U. S. A.*, **116**, 22931–22937.

8. CHEN, Y. & SUH, C. (2015) Spectral mle: top-k rank aggregation from pairwise comparisons. *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15.JMLR.org, pp. 371–380.

9. COSSOCK, D. & ZHANG, T. (2006) Subset ranking using regression. *Learning Theory*, vol. **4005** of Lecture Notes in Comput. Sci. Berlin: Springer, pp. 605–619.

10. DWORK, C., KUMAR, R., NAOR, M. & SIVAKUMAR, D. (2001) Rank aggregation methods for the web. *Proceedings of the 10th International Conference on World Wide Web*, WWW' 01. New York, NY, USA: Association for Computing Machinery, pp. 613–622.

11. FARRELL, M. H., LIANG, T. & MISRA, S. (2021) Deep neural networks for estimation and inference. *Econometrica*, **89**, 181–213.

12. FORD Jr., L. R. (1957) Solution of a ranking problem from binary comparisons. *Am. Math. Monthly*, **64**, 28–33.

13. GILL, R. D. & LEVIT, B. Y. (1995) Applications of the Van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, **1**, 59–79.

14. HAN, R., YE, R., TAN, C. & CHEN, K. (2020) Asymptotic theory of sparse Bradley-Terry model. *Ann. Appl. Probab.*, **30**, 2491–2515.

15. HOLLAND, P. W. & LEINHARDT, S. (1981) An exponential family of probability distributions for directed graphs. *J. Am. Stat. Assoc.*, **76**, 33–65.

16. HUNTER, D. R. (2004) MM algorithms for generalized Bradley-Terry models. *Ann. Statist.*, **32**, 384–406.

17. JAVANMARD, A. & MONTANARI, A. (2014) Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.*, **15**, 2869–2909.

18. LANGE, K., HUNTER, D. R. & YANG, I. (2000) Optimization transfer using surrogate objective functions. *J. Comput. Graph. Statist.*, **9**, 1–59 With discussion, and a rejoinder by Hunter and Lange.

19. LIU, Y., FANG, E. X. & JUNWEI, L. (2022) Lagrangian inference for ranking problems. *Operations Research*, 0(0).

20. DUNCAN LUCE, R. (1959) *Individual choice behavior: A theoretical analysis*. London: John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd.

21. MCFADDEN, D. (1973) Conditional logit analysis of qualitative choice behaviour. *Frontiers in Econometrics*. (P. ZAREMBKA ed). New York, New York, NY, USA: Academic Press, pp. 105–142.

22. MONARD, F., NICKL, R. & PATERNAIN, G. P. (2021) Statistical guarantees for Bayesian uncertainty quantification in nonlinear inverse problems with Gaussian process priors. *Ann. Statist.*, **49**, 3255–3298.

23. MOTEGI, S. & MASUDA, N. (2012) A network-based dynamical ranking system for competitive sports. *Sci. Rep.*, **2**, 1–7.

24. NEGAHBAN, S., SEWOONG, O. & SHAH, D. (2017) Rank centrality: ranking from pairwise comparisons. *Oper. Res.*, **65**, 266–287.

25. NEWMAN, M. E. J. (2003) The structure and function of complex networks. *SIAM Rev.*, **45**, 167–256.

26. NEYKOV, M., NING, Y., LIU, J. S. & LIU, H. (2018) A unified theory of confidence regions and testing for high-dimensional estimating equations. *Statist. Sci.*, **33**, 427–443.

27. NICKL, R. & van de GEER, S. (2013) Confidence sets in sparse regression. *Ann. Statist.*, **41**, 2852–2876.

28. NING, Y. & LIU, H. (2017) A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann. Statist.*, **45**, 158–195.

29. PARK, J. & NEWMAN, M. E. J.. Statistical mechanics of networks. *Phys. Rev. E (3)*, **70**:066117, 13, 2004.

30. PLACKETT, R. L. (1975) The analysis of permutations. *J. Roy. Statist. Soc. Ser. C*, **24**, 193–202.

31. REN, Z., SUN, T., ZHANG, C.-H. & ZHOU, H. H. (2015) Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist.*, **43**, 991–1026.

32. ROBINS, G., SNIJDERS, T., WANG, P., HANDCOCK, M. & PATTISON, P. (2007) Recent developments in exponential random graph (p*) models for social networks. *Soc. Networks*, **29**, 192–215.

33. RUDELSON, M. & VERSHYNIN, R. (2013) Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Comm. Probab.*, **18**, 9.

34. SHA, L., LUCEY, P., YUE, Y., CARR, P., ROHLF, C. & MATTHEWS, I. (2016) Chalkboarding: a new spatiotemporal query paradigm for sports play retrieval. *Proceedings of the 21st International Conference on Intelligent User Interfaces*. Association for Computing Machinery, New York, NY, United States, pp. 336–347.

35. SIMONS, G. & YAO, Y.-C. (1998) Approximating the inverse of a symmetric positive definite matrix. *Linear Algebra Appl.*, **281**, 97–103.

36. SIMONS, G. & YAO, Y.-C. (1999) Asymptotics when the number of parameters tends to infinity in the Bradley-Terry model for paired comparisons. *Ann. Statist.*, **27**, 1041–1060.

37. SUR, P. & CANDÈS, E. J. (2019) A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. U. S. A.*, **116**, 14516–14525.

38. TROPP, J. A. (2012) User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, **12**, 389–434.

39. TROPP, J. A. (2015) An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.*, **8**, 1–230.

40. VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. & DEZEURE, R. (2014) On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.*, **42**, 1166–1202.

41. VAN DER PAS, S., SZABÓ, B. & VAN DER VAART, A. (2017) Uncertainty quantification for the horseshoe (with discussion). *Bayesian Anal.*, **12**, 1221–1274.

42. VAN DER VAART, A. Asymptotic statistics, volume **3** of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

43. VIGNA, S. (2016) Spectral ranking. *Netw. Sci.*, **4**, 433–445.

44. ZERMELO, E. (1929) Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrschein-lichkeitsrechnung. *Math. Z.*, **29**, 436–460.

45. ZHANG, C.-H. & ZHANG, S. S. (2014) Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **76**, 217–242.

A. Additional auxiliary results

LEMMA 13. Suppose that $np \gg \log n$. Then there exist some universal positive constants c, C such that the following hold with probability $1 - \mathcal{O}(n^{-10})$.

$$cnp \leq \min_{i \in [n]} \sum_{j:j \neq i} A_{ij} \leq \max_{i \in [n]} \sum_{j:j \neq i} A_{ij} \leq Cnp,$$

$$\max_{m \in [n]} \sum_{i,j:i<j,i,j \neq m} A_{ij} A_{mi} \lesssim C(np)^2.$$

Proof. For any fixed $i \in [n]$, we have $cnp \leq \sum_{j:j \neq i} A_{ij} \leq Cnp$ with the prescribe probability by Hoeffding's inequality. The first claim now follows by taking the union bound over $i \in [n]$. The second claim follows since

$$\max_{m \in [n]} \sum_{i,j:i<j,i,j \neq m} A_{ij} A_{mi} \leq \max_{m \in [n]} \sum_{i:i \neq m} A_{mi} \cdot \max_{i \in [n]} \sum_{j:j \neq i} A_{ij} \lesssim (np)^2$$

by the first claim. □

The following lemma gives some standard concentration results in the analysis of both MLE and the spectral method.

LEMMA 14. Assume $np \geq c_0 \log n$ for some sufficiently large $c_0 > 0$ and $\kappa = \mathcal{O}(1)$. Then the following hold with probability $1 - \mathcal{O}(n^{-10})$ uniformly over $\theta^* \in \Theta(\kappa)$.

$$\sum_{i=1}^{n} \left(\sum_{j:j \neq i} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)) \right)^2 \leq C\frac{n^2 p}{L},$$

$$\max_{i \in [n]} \left| \sum_{j:j \neq i} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)) \right|^2 \leq C\frac{np \log n}{L},$$

$$\max_{m \in [n]} \left| \sum_{i:i \neq m} A_{im}(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*) \right|^2 \leq C\frac{p \log n}{nL}.$$

Here $C = C(\kappa) > 0$.

Proof of Lemma 14 The first two claims are proved in [5, Lemma 7.4]. For the third claim, note that the summands $\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*$ are independent, centered, and sub-Gaussian with variance proxy bounded by $L^{-1}\|\pi^*\|_\infty^2 \lesssim (n^2 L)^{-1}$. Hence the claim follows from Hoeffding's inequality

$$\mathbb{P}\left(\left| \sum_{i:i \neq m} A_{im}(\bar{y}_{mi}\pi_i^* - \bar{y}_{im}\pi_m^*) \right| \geq t | A \right) \leq 2 \exp\left(-\frac{Ct^2}{(n^2 L)^{-1} \cdot \sum_{i:i \neq m} A_{im}} \right),$$

the bound $\max_{m \in [n]} \sum_{i:i \neq m} A_{mi} \lesssim np$ by Lemma 13, and a union bound. □

The following lemma controls the spectrum of Laplacian matrices corresponding to a weighted Erdős-Rényi graph.

LEMMA 15. Let A be the Erdoős-Rényi graph adjacency matrix with connection probability p, and $\{w_{ij}\}_{i \neq j}$ be a non-negative weight sequence such that $w_{ij} = w_{ji}$. Let H be a weighted Laplacian matrix defined by $H_{ij} \equiv -A_{ij}w_{ij}$ for $i \neq j$ and $H_{ii} \equiv \sum_{j \neq i} A_{ij}w_{ij}$. Assume that $p \geq c_0 \log n/n$ for some sufficiently large $c_0 > 0$ and $\kappa_1 \leq \min_{i \neq j} w_{ij} \leq \max_{i \neq j} w_{ij} \leq \kappa_2$ for some $\kappa_1, \kappa_2 > 0$. Then for some positive c, C only depending on c_0,

$$(c\kappa_1)np \leq \lambda_{\min,\perp}(H) \leq \lambda_{\max,\perp}(H) \leq (C\kappa_2)np$$

with probability at least $1 - \mathcal{O}(n^{-10})$. Here $\lambda_{\min,\perp}$ stands for the smallest eigenvalue along directions orthogonal to $\mathbf{1}_n$.

Proof of Lemma 15 Let $\mathscr{L}_A$ be the graph Laplacian induced by A. For any unit vector $c \in \mathbb{R}^n$ such that $\mathbf{1}_n^\top c = 0$, we have

$$c^\top Hc = \sum_{i \neq j} \frac{1}{2}(c_i - c_j)^2 A_{ij} w_{ij} \geq \kappa_1 \cdot \sum_{i \neq j} \frac{1}{2}(c_i - c_j)^2 A_{ij} \geq \kappa_1 \cdot \lambda_{\min,\perp}(\mathscr{L}_A).$$

The claim now follows from standard results of the Erdoős-Rényi graph [39]. The upper bound proof is similar. □

The following lemma provides rate-optimal bounds for the global MLE $\widehat{\theta}$ and the leave-one-out MLE $\widehat{\theta}^{(m)}$ in (5.9) for each $m \in [n]$.

LEMMA 16. Suppose that $\kappa = \mathcal{O}(1)$ and $np \gg \log n$. There exists some $C = C(\kappa) > 0$ such that the following hold with probability $1 - \mathcal{O}(n^{-10})$ uniformly over all $\theta^* \in \Theta(\kappa)$.

1. We have $\max_{m \in [n]} \|\widehat{\theta}^{(m)} - \theta^*_{-m}\|^2 \leq C(pL)^{-1}$ and $\max_{m \in [n]} \|\widehat{\theta}^{(m)} - \theta^*_{-m}\|_\infty^2 \leq C \log n/(npL)$.

2. We have $\max_{m \in [n]} \|\widehat{\theta}^{(m)} - \widehat{\theta}_{-m}\|^2 \leq C(npL)^{-1}$.

Proof. These are essentially proved in [5, Lemmas 8.6 and 8.7]. Note that there are two differences between our leave-one-out MLE $\widehat{\theta}^{(m)}$ in (5.9) and the one in Equation (57) of [5]: (i) our (5.9) does not have the ℓ_∞ constraint $\|\theta_{-m} - \theta^*_{-m}\|_\infty \leq 5$ in its definition; (ii) our (5.9) satisfies $\text{ave}(\theta_{-m}) = \text{ave}(\theta^*_{-m})$ by definition. The above bounds continue to hold with the prescribed probability since for (i), the constraint is satisfied with the prescribed probability by (a leave-one-out version of) [5, Proposition 8.1], and for (ii), Lemmas 8.6 and 8.7 in [5] hold for $\widehat{\theta}^{(m)}$ only after the additional centering $\widehat{\theta}^{(m)} - \text{ave}(\widehat{\theta}^{(m)} - \theta^*_{-m})$, which coincides with the leave-one-out MLE in our definition (5.9). □

The following lemma summarizes some rate-optimal bounds of the spectral estimate $\widehat{\pi}$.

LEMMA 17. Lemma 9.1 of [5], see also [6]Suppose that $\kappa = \mathcal{O}(1)$ and $np \gg \log n$. Then the following hold with probability at least $1 - \mathcal{O}(n^{-10})$ for some positive $C = C(\kappa)$.

1. We have $\max_{m \in [n]} \|\widehat{\pi}^{(m)} - \widehat{\pi}\| \leq Cn^{-1}\sqrt{\log n/(npL)}$.

2. We have $\max_{m \in [n]} \|\widehat{\pi}^{(m)} - \pi^*\|_\infty \leq Cn^{-1}\sqrt{\log n/(npL)}$ and $\|\widehat{\pi} - \pi^*\|_\infty \leq Cn^{-1}\sqrt{\log n/(npL)}$.

3. We have $\max_{m \in [n]} \|\widehat{\pi}^{(m)} - \pi^*\| \leq Cn^{-1}\sqrt{1/(pL)}$ and $\|\widehat{\pi} - \pi^*\| \leq Cn^{-1}\sqrt{1/(pL)}$.

The following lemma summarizes some useful estimates regarding $f^{(m)}(\theta_m|\theta_{-m})$ and $g^{(m)}(\theta_m|\theta_{-m})$ defined in (5.1) for each $m \in [n]$.

LEMMA 18. Suppose $\kappa = \mathscr{O}(1)$ and $np \gg \log n$. Then the following hold with probability $1 - \mathscr{O}(n^{-10})$ for each $m \in [n]$.

1. We have $\max_{m \in [n]} \left| f^{(m)}(\theta_m^* | \widehat{\theta}_{-m}) \right| \lesssim \sqrt{np \log n / L}$ and $\sum_{m=1}^n \left(f^{(m)}(\theta_m^* | \widehat{\theta}_{-m}) \right)^2 \lesssim n^2 p / L$.

2. We have $\max_{m \in [n]} \left| f^{(m)}(\theta_m^* | \widehat{\theta}_{-m}) - f^{(m)}(\theta_m^* | \theta_{-m}^*) \right| \lesssim \sqrt{np/L} + \sqrt{(\log n)^3 / (npL)}$.

3. We have $g^{(m)}(\theta_m^* | \theta_{-m}^*) \asymp np$.

4. We have $\left| g^{(m)}(\theta_m^* | \widehat{\theta}_{-m}) - g^{(m)}(\theta_m^* | \theta_{-m}^*) \right| \lesssim \sqrt{np/L} + \sqrt{(\log n)^3 / (npL)}$. Consequently, $g^{(m)}(\theta_m^* | \widehat{\theta}_{-m}) = (1 + \varepsilon_m) g^{(m)}(\theta_m^* | \theta_{-m}^*)$ under the condition $np \gg \log n$ for some $\varepsilon \in \mathbb{R}^n$ such that $\|\varepsilon\|_\infty = \mathfrak{o}(1)$.

Proof of Lemma 18 Claims (1)(3)(4) follow from analogous arguments as in the proof of Lemma 3, and it remains to prove Claim (2). By definition, we have

$$f^{(m)}(\theta_m^* | \widehat{\theta}_{-m}) - f^{(m)}(\theta_m^* | \theta_{-m}^*) = \sum_{i:i \neq m} A_{mi} \big(\psi(\theta_m^* - \widehat{\theta}_i) - \psi(\theta_m^* - \theta_i^*) \big).$$

Recall that $\widehat{\theta}^{(m)}$ is the leave-one-out MLE defined in (5.9). Then we have

$$\sum_{i:i \neq m} A_{mi} \big(\psi(\theta_m^* - \widehat{\theta}_i) - \psi(\theta_m^* - \theta_i^*) \big)$$

$$= \sum_{i:i \neq m} A_{mi} \big(\psi(\theta_m^* - \widehat{\theta}_i^{(m)}) - \psi(\theta_m^* - \theta_i^*) \big)$$

$$+ \sum_{i:i \neq m} A_{mi} \big(\psi(\theta_m^* - \widehat{\theta}_i) - \psi(\theta_m^* - \widehat{\theta}_i^{(m)}) \big) \equiv \Delta_{m,1} + \Delta_{m,2}.$$

The term $\Delta_{m,1}$ is controlled with the prescribed probability by

$$|\Delta_{m,1}| \leq p \cdot \left| \sum_{i:i \neq m} \big(\psi(\theta_m^* - \theta_i^*) - \psi(\theta_m^* - \widehat{\theta}_i^{(m)}) \big) \right|$$

$$+ \left| \sum_{i:i \neq m} (A_{mi} - p) \big(\psi(\theta_m^* - \theta_i^*) - \psi(\theta_m^* - \widehat{\theta}_i^{(m)}) \big) \right|$$

$$\overset{(*)}{\lesssim} p\sqrt{n} \|\widehat{\theta}^{(m)} - \theta_{-m}^*\| + \sqrt{p \log n} \|\widehat{\theta}^{(m)} - \theta_{-m}^*\| + \log n \cdot \|\widehat{\theta}^{(m)} - \theta_{-m}^*\|_\infty$$

$$\overset{(**)}{\lesssim} p\sqrt{n} \cdot \sqrt{\frac{1}{pL}} + \log n \cdot \sqrt{\frac{\log n}{npL}} = \sqrt{\frac{np}{L}} + \sqrt{\frac{(\log n)^3}{npL}}. \tag{A1}$$

Here $(*)$ follows from Bernstein's inequality applied conditioning on data without the mth observation, and $(**)$ follows from Lemma 16.

On the other hand, $\Delta_{m,2}$ is controlled by

$$|\Delta_{m,2}| \leq \sum_{i:i\neq m} A_{mi}|\widehat{\theta}_i^{(m)} - \widehat{\theta}_j| \leq \Big(\sum_{i:i\neq m} A_{mi} \Big)^{1/2} \|\widehat{\theta}^{(m)} - \widehat{\theta}_{-m}\|$$

$$\overset{(*)}{\lesssim} \sqrt{np} \cdot \sqrt{\frac{1}{npL}} = \sqrt{\frac{1}{L}},$$

where $(*)$ follows by Lemma 16. Since both estimates are uniform over $m \in [n]$, the proof is complete. $\qquad\square$

Recall $\bar{\theta}$ defined in (5.2) and its motivation of local quadratic expansion in (2.9). The following lemma quantifies the closeness between $\bar{\theta}$ and $\widehat{\theta}$.

LEMMA 19. Suppose that $\kappa = \mathscr{O}(1)$ and $np \gg \log n$. Then there exists some $C = C(\kappa) > 0$ such that the following holds with probability $1 - \mathscr{O}(n^{-10})$.

$$\big|\widehat{\theta}_m - \bar{\theta}_m\big|^2 \leq C\Big(\frac{\big|f^{(m)}(\theta_m^*|\widehat{\theta}_{-m})\big|}{np}\Big)^3, \quad \forall m \in [n].$$

Proof of Lemma 19 This follows directly from the estimates before Equation (78) of [5]. $\qquad\square$