Supplementary Material

6 Auxiliary Lemmas: Proof of Lemma 3

Proof. We can rewrite (6) as an optimization problem over the ℓ_1/ℓ_2 ball of radius C for some $C(\lambda_n) < \infty$. Since $\lambda_n > 0$, by KKT conditions, $\left\|\widetilde{\Theta}_{\backslash r}\right\|_{1,2} = C$ for all optimal primal solution $\widetilde{\Theta}_{\backslash r}$.

By definition of the ℓ_1/ℓ_2 subdifferential, we know that for any column $u \in V\backslash\{r\}$, we have $\left\|\left(\hat{Z}_{\backslash r}\right)_u\right\|_2 \leq 1$. Considering the necessary optimality condition $\nabla\ell\left(\hat{\Theta}_{\backslash r}\right) + \lambda_n \hat{Z}_{\backslash r} = 0$, by complementary slackness condition, we have $\left\langle \widetilde{\Theta}_{\backslash r}, \hat{Z}_{\backslash r} \right\rangle - C = \left\langle \widetilde{\Theta}^T_{\backslash r}, \hat{Z}_{\backslash r} \right\rangle - \left\|\widetilde{\Theta}_{\backslash r}\right\|_{1,2} = 0$. Now if for an arbitrary column $u \in V\backslash\{r\}$, we have $\left\|\left(\hat{Z}_{\backslash r}\right)_u\right\|_2 < 1$ and $\left(\widetilde{\Theta}_{\backslash r}\right)_u \neq 0$ then this would contradict the condition that $\left\langle \widetilde{\Theta}_{\backslash r}, \hat{Z}_{\backslash r} \right\rangle = \left\|\widetilde{\Theta}_{\backslash r}\right\|_{1,2}$.

For this restricted problem, if the Hessian sub-matrix is positive definite, then the problem is strictly convex and it has a unique solution.

□

7 Derivatives of the Log-Likelihood Function

In this section, we point out the key properties of the gradient, Hessian and derivative of the Hessian for the log-liklihood function. These properties are used to prove the concentration lemmas.

7.1 Gradient

By simple derivation, we have

$$\frac{\partial}{\partial\theta^*_{rt;\ell k}}\ell^{(i)}(\Theta_{\backslash r}; D) = \mathcal{I}\left[x^{(i)}_t = k\right]\left(\mathcal{I}\left[x^{(i)}_r{=}\ell\right] - \mathbb{P}_{\Theta^*_{\backslash r}}\left[X_r{=}\ell \,|\, X_{\backslash r}{=}x^{(i)}_{\backslash r}\right]\right).$$

It is easy to show that $\mathbb{E}_{\Theta^*_{\backslash r}}\left[\frac{\partial}{\partial\theta^*_{rt;\ell k}}\ell^{(i)}(\Theta_{\backslash r}; D)\right] = 0$ and $\mathrm{Var}\left(\frac{\partial}{\partial\theta^*_{rt;\ell k}}\ell^{(i)}(\Theta_{\backslash r}; D)\right) \leq \frac{1}{4}$. With i.i.d assumption on drawn samples, we have $\mathrm{Var}\left(\frac{\partial}{\partial\theta^*_{rt;\ell k}}\ell(\Theta_{\backslash r}; D)\right) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta^*_{rt;\ell k}}\ell^{(i)}(\Theta_{\backslash r}; D)\right) \leq \frac{1}{4n}$. Hence, for a fixed $t \in V\backslash\{r\}$ by Jensen's inequality,

$$\begin{aligned}\mathbb{E}_{\Theta^*_{\backslash r}}\left[\left\|\frac{\partial}{\partial\theta^*_{rt;\ell k}}\ell(\Theta_{\backslash r}; D)\right\|_2\right] &\leq \sqrt{\mathbb{E}_{\Theta^*_{\backslash r}}\left[\left\|\frac{\partial}{\partial\theta^*_{rt;\ell k}}\ell(\Theta_{\backslash r}; D)\right\|^2_2\right]} \\ &\leq \frac{m-1}{2\sqrt{n}}.\end{aligned}$$

Considering the terms associated with $\theta^*_{rt;\ell k}$'s in the gradient vector of the log-likelihood function, for a fixed $t \in V\backslash\{r\}$, only $m-1$ (out of $(m-1)^2$) values are non-zero. By a simple calculation, we get

$$\max_{t\in V\backslash\{r\}}\left\|\frac{\partial}{\partial\theta^*_{rt;\ell k}}\ell^{(i)}(\Theta_{\backslash r}; D)\right\|_2 \leq \sqrt{2} \qquad \forall i.$$

By Azuma-Hoeffding inequality, we get

$$\mathbb{P}\left[\left\|\frac{\partial}{\partial\theta^*_{rt;\ell k}}\ell(\Theta_{\backslash r}; D)\right\|_2 > \frac{m-1}{2\sqrt{n}} + \epsilon\right] \leq 2\exp\left(-\frac{\epsilon^2}{4}n\right),$$

for all $t \in V\backslash\{r\}$. Using the union bound, we get

$$\mathbb{P}\left[\max_{t\in V\backslash\{r\}}\left\|\frac{\partial}{\partial\theta^*_{rt;\ell k}}\ell(\Theta_{\backslash r}; D)\right\|_2 > \frac{m-1}{2\sqrt{n}} + \epsilon\right] \leq 2\exp\left(-\frac{\epsilon^2}{4}n + \log(p-1)\right). \tag{12}$$

7.2 Hessian

For the Hessian of the log-likelihood function, we have

$$\frac{\partial^2\,\ell^{(i)}(\Theta_{\backslash r}; D)}{\partial\theta^*_{rt_2;\ell_2k_2}\,\partial\theta^*_{rt_1;\ell_1k_1}} = \mathcal{I}\left[x^{(i)}_{t_1}{=}k_1\right]\mathcal{I}\left[x^{(i)}_{t_2}{=}k_2\right]\eta_{\ell_1\ell_2}\left(x^{(i)}\right),$$

where,

$$\eta_{\ell_1\ell_2}\left(x^{(i)}\right) := \mathbb{P}_{\Theta^*_{\backslash r}}\left[X_r = \ell_1 \,\middle|\, X_{\backslash r} = x^{(i)}_{\backslash r}\right] \left(\mathcal{I}\left[x^{(i)}_r{=}\ell_1\right]\mathcal{I}\left[x^{(i)}_r{=}\ell_2\right] - \mathbb{P}_{\Theta^*_{\backslash r}}\left[X_r{=}\ell_2 \middle| X_{\backslash r}{=}x^{(i)}_{\backslash r}\right]\right).$$

Consider the zero-mean random variable

$$Z^{(i)}_{t_1\ell_1k_1;t_2\ell_2k_2} := \frac{\partial^2\,\ell^{(i)}(\Theta_{\backslash r}; D)}{\partial\theta^*_{rt_2;\ell_2k_2}\,\partial\theta^*_{rt_1;\ell_1k_1}} - \mathbb{E}\left[\frac{\partial^2\,\ell(\Theta_{\backslash r}; D)}{\partial\theta^*_{rt_2;\ell_2k_2}\,\partial\theta^*_{rt_1;\ell_1k_1}}\right].$$

Notice that $\mathrm{Var}\left(Z^{(i)}_{t_1\ell_1k_1;t_2\ell_2k_2}\right) \leq 1$ and consequently, by i.i.d assumption, $\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n Z^{(i)}_{t_1\ell_1k_1;t_2\ell_2k_2}\right) \leq \frac{1}{n}$.

Hence, for fixed values t_1, ℓ_1, k_1 and $t_2 \in S_2 \subseteq V\backslash\{r\}$, we have

$$\begin{aligned}\mathbb{E}_{\Theta^*_{\backslash r}}\left[\left\|\frac{1}{n}\sum_{i=1}^{n} Z^{(i)}_{t_1\ell_1 k_1; t_2\ell_2 k_2}\right\|_2\right] &\leq \sqrt{\mathbb{E}_{\Theta^*_{\backslash r}}\left[\left\|\frac{1}{n}\sum_{i=1}^{n} Z^{(i)}_{t_1\ell_1 k_1; t_2\ell_2 k_2}\right\|_2^2\right]} \\ &\leq \sqrt{\frac{|S_2|}{n}}.\end{aligned} \tag{13}$$

This radom variable, for fixed values t_1, ℓ_1, k_1 and a fixed t_2, is bounded and in particular, $\left\|\frac{1}{n}\sum_{i=1}^{n} Z^{(i)}_{t_1\ell_1 k_1; t_2\ell_2 k_2}\right\|_2 \leq 2$. By Azuma-Hoeffding inequality and the union bound,

$$\begin{aligned}&\mathbb{P}\left[\left\|Q^n_{S_rS_r} - Q^*_{S_rS_r}\right\|_{\infty,2} > \frac{\sqrt{d_r}}{\sqrt{n}} + \epsilon\right] \\ &\qquad \leq 2\exp\left(-\frac{\epsilon^2}{8}n + \log\left((m-1)^2 d_r\right)\right). \\ &\mathbb{P}\left[\left\|Q^n_{S^c_rS_r} - Q^*_{S^c_rS_r}\right\|_{\infty,2} > \frac{\sqrt{d_r}}{\sqrt{n}} + \epsilon\right] \\ &\qquad \leq 2\exp\left(-\frac{\epsilon^2}{8}n + \log\left((m-1)^2(p-d_r-1)\right)\right).\end{aligned} \tag{14}$$

Similar analysis as (13) combined with the ineqality $\Lambda_{\max}(\cdot) \leq \|\cdot\|_{\infty,2}$, shows that

$$\begin{aligned}&\mathbb{P}\left[\Lambda_{\max}\left(Q^n_{S_rS_r} - Q^*_{S_rS_r}\right) > \frac{\sqrt{d_r}}{\sqrt{n}} + \epsilon\right] \\ &\qquad \leq 2\exp\left(-\frac{\epsilon^2}{8}n + \log\left((m-1)^2 d_r\right)\right).\end{aligned} \tag{15}$$

We also need a control over the deviation of the inverse sample Fisher information matrix from the inverse of its mean. We have

$$\begin{aligned}&\Lambda_{\max}\left(\left(Q^n_{S_rS_r}\right)^{-1} - \left(Q^*_{S_rS_r}\right)^{-1}\right) \\ &= \Lambda_{\max}\left(\left(Q^*_{S_rS_r}\right)^{-1}\left(Q^*_{S_rS_r} - Q^n_{S_rS_r}\right)\left(Q^n_{S_rS_r}\right)^{-1}\right) \\ &\leq \Lambda_{\max}\left(\left(Q^*_{S_rS_r}\right)^{-1}\right)\Lambda_{\max}\left(Q^*_{S_rS_r} - Q^n_{S_rS_r}\right) \\ &\qquad\qquad \Lambda_{\max}\left(\left(Q^n_{S_rS_r}\right)^{-1}\right) \\ &\leq \frac{\sqrt{d_r}}{C_{\min}\sqrt{n}}\Lambda_{\max}\left(\left(Q^n_{S_rS_r}\right)^{-1}\right).\end{aligned}$$

By part (B1) in Lemma 1, we have

$$\begin{aligned}&\mathbb{P}\left[\Lambda_{\max}\left(\left(Q^n_{S_rS_r}\right)^{-1}\right) > \frac{1}{C_{\min}} + \epsilon\right] \\ &\leq 2\exp\left(-\frac{\left(\frac{C_{\min}\epsilon\sqrt{n}}{1+C_{\min}\epsilon} - \sqrt{d_r}\right)^2}{8} + \log\left((m-1)^2 d_r\right)\right).\end{aligned} \tag{16}$$

Hence, we get,

$$\begin{aligned}&\mathbb{P}\left[\Lambda_{\max}\left(\left(Q^n_{S_rS_r}\right)^{-1} - \left(Q^*_{S_rS_r}\right)^{-1}\right) > \frac{\sqrt{d_r}}{C^2_{\min}\sqrt{n}} + \epsilon\right] \\ &\leq 4\exp\left(-\frac{\left(\frac{C_{\min}\epsilon\sqrt{n}}{1+C_{\min}\epsilon} - \sqrt{d_r}\right)^2}{8} + \log\left((m-1)^2 d_r\right)\right).\end{aligned} \tag{17}$$

7.3 Derivative of Hessian

We want to bound the rate of the change for the elements of Hessian matrix. Let

$$\begin{aligned}&\nabla Q^{(i)}_{t_2\ell_2 k_2; t_1\ell_1 k_1} \\ &\quad := \frac{\partial}{\partial\Theta_{\backslash r}}\frac{\partial^2 \ell^{(i)}(\Theta_{\backslash r}; D)}{\partial\theta^*_{rt_2;\ell_2 k_2}\,\partial\theta^*_{rt_1;\ell_1 k_1}} \\ &\quad = \mathcal{I}\left[x^{(i)}_{t_1} = k_1\right]\mathcal{I}\left[x^{(i)}_{t_2} = k_2\right]\frac{\partial}{\partial\Theta_{\backslash r}}\eta_{\ell_1\ell_2}\left(x^{(i)}\right).\end{aligned}$$

Recall the definition of $\eta(\cdot)$ from section 7.2. We have

$$\begin{aligned}\frac{\partial\eta_{\ell_1\ell_2}\left(x^{(i)}\right)}{\partial\theta_{rt_3;\ell_3k_3}} &= \mathcal{I}\left[x^{(i)}_{t_3} = k_3\right]\mathbb{P}_{\Theta^*_{\backslash r}}\left[X_r = \ell_1 \middle| X_{\backslash r} = x^{(i)}_{\backslash r}\right] \\ &\left(\eta_{\ell_2\ell_3}\left(x^{(i)}\right) - \frac{\eta_{\ell_1\ell_2}\left(x^{(i)}\right)\eta_{\ell_1\ell_3}\left(x^{(i)}\right)}{\mathbb{P}_{\Theta^*_{\backslash r}}\left[X_r = \ell_1 \middle| X_{\backslash r} = x^{(i)}_{\backslash r}\right]^2}\right).\end{aligned}$$

For any $t_3 \in V\backslash\{r\}$, each entry is bounded by $\frac{1}{2}$ and there are only $m-1$ non-zero entries for each k_3. Hence, for any t_3, one can colculde that $\left\|\frac{\partial}{\partial\theta_{rt_3;\ell_3k_3}}\eta_{\ell_1\ell_2}\left(x^{(i)}\right)\right\|_2 \leq \frac{m-1}{\sqrt{2}}$ for all i. Finally, for all ℓ_1 and ℓ_2 we have

$$\max_{t_3\in V\backslash\{r\}}\left\|\frac{\partial}{\partial\theta_{rt_3;\ell_3k_3}}\eta_{\ell_1\ell_2}\left(x^{(i)}\right)\right\|_2 \leq \frac{m-1}{\sqrt{2}}. \tag{18}$$

8 Proof of Lemma 1

(B1) By variational representation of the smallest eigenvalue, we have

$$\begin{aligned}\Lambda_{\min}\left(Q^*_{S_rS_r}\right) &= \min_{\|x\|_2=1} x^T Q^*_{S_rS_r} x \\ &\leq y^T Q^n_{S_rS_r} y + y^T\left(Q^*_{S_rS_r} - Q^n_{S_rS_r}\right)y,\end{aligned}$$

for all $y \in \mathbb{R}^{(m-1)^2 d_r}$ with $\|y\|_2 = 1$ and in particular for the unit-norm minimal eigenvalue of $Q^n_{S_r S_r}$. Hence,

$$\Lambda_{\min}\left(Q^n_{S_r S_r}\right) \geq \Lambda_{\min}\left(Q^*_{S_r S_r}\right) - \Lambda_{\max}\left(Q^*_{S_r S_r} - Q^n_{S_r S_r}\right).$$

By (15), we get

$$\begin{aligned}
&\mathbb{P}\left[\Lambda_{\min}\left(Q^n_{S_r S_r}\right) < C_{\min} - \epsilon\right] \\
&\quad \leq \mathbb{P}\left[\Lambda_{\max}\left(Q^*_{S_r S_r} - Q^n_{S_r S_r}\right) > \epsilon\right] \\
&\quad \leq 2\exp\left(-\frac{(\epsilon\sqrt{n} - \sqrt{d_r})^2}{8} + \log\left((m-1)^2 d_r\right)\right).
\end{aligned}$$

(B2) We can write

$$\begin{aligned}
Q^n_{S_r^c S_r}\left(Q^n_{S_r S_r}\right)^{-1} &= \underbrace{Q^*_{S_r^c S_r}\left(Q^*_{S_r S_r}\right)^{-1}}_{T_0} \\
&+ \underbrace{Q^*_{S_r^c S_r}\left(\left(Q^n_{S_r S_r}\right)^{-1} - \left(Q^*_{S_r S_r}\right)^{-1}\right)}_{T_1} \\
&+ \underbrace{\left(Q^n_{S_r^c S_r} - Q^*_{S_r^c S_r}\right)\left(Q^*_{S_r S_r}\right)^{-1}}_{T_2} \\
&+ \underbrace{\left(Q^n_{S_r^c S_r} - Q^*_{S_r^c S_r}\right)\left(\left(Q^n_{S_r S_r}\right)^{-1} - \left(Q^*_{S_r S_r}\right)^{-1}\right)}_{T_3}.
\end{aligned}$$

Considering assumption (A3), $\|T_0\|_{\infty,2} < \frac{1-2\alpha}{\sqrt{d_r}}$ and hence, it suffices to show that $\|T_i\|_{\infty,2} < \frac{\alpha}{3\sqrt{d_r}}$ for $i = 1, 2, 3$. For the first term, we have

$$\begin{aligned}
&\left\|Q^*_{S_r^c S_r}\left(\left(Q^n_{S_r S_r}\right)^{-1} - \left(Q^*_{S_r S_r}\right)^{-1}\right)\right\|_{\infty,2} \\
&= \left\|Q^*_{S_r^c S_r}\left(Q^*_{S_r S_r}\right)^{-1}\left(Q^*_{S_r S_r} - Q^n_{S_r S_r}\right)\left(Q^n_{S_r S_r}\right)^{-1}\right\|_{\infty,2} \\
&\leq \left\|Q^*_{S_r^c S_r}\left(Q^*_{S_r S_r}\right)^{-1}\right\|_{\infty,2} \Lambda_{\max}\left(Q^*_{S_r S_r} - Q^n_{S_r S_r}\right) \\
&\qquad\qquad \Lambda_{\max}\left(\left(Q^n_{S_r S_r}\right)^{-1}\right) \\
&\leq \frac{1-2\alpha}{\sqrt{d_r}} \frac{\sqrt{d_r}}{\sqrt{n}} \frac{1}{C_{\min}}.
\end{aligned}$$

The last inequality follows from (14) and (16) with high probability. Setting $\bar{C}_{\min} = \min(C_{\min}, 1)$, by applying the union bound,

$$\begin{aligned}
&\mathbb{P}\left[\left\|Q^*_{S_r^c S_r}\left(\left(Q^n_{S_r S_r}\right)^{-1} - \left(Q^*_{S_r S_r}\right)^{-1}\right)\right\|_{\infty,2} > \epsilon\right] \\
&\leq 4\exp\left(-\frac{\left(\bar{C}_{\min}\epsilon\sqrt{n} - \sqrt{d_r} - \frac{1-2\alpha}{C_{\min}}\right)^2}{8} + \log\left((m-1)^2 d_r\right)\right).
\end{aligned}$$

For the second term, we have

$$\begin{aligned}
&\left\|\left(Q^n_{S_r^c S_r} - Q^*_{S_r^c S_r}\right)\left(Q^*_{S_r S_r}\right)^{-1}\right\|_{\infty,2} \\
&\qquad \leq \left\|Q^n_{S_r^c S_r} - Q^*_{S_r^c S_r}\right\|_{\infty,2} \Lambda_{\max}\left(\left(Q^*_{S_r S_r}\right)^{-1}\right) \\
&\qquad \leq \frac{\sqrt{d_r}}{\sqrt{n}} \frac{1}{C_{\min}}.
\end{aligned}$$

The last inequality follows from (14) with high probability. Hence, we have

$$\begin{aligned}
&\mathbb{P}\left[\left\|\left(Q^n_{S_r^c S_r} - Q^*_{S_r^c S_r}\right)\left(Q^*_{S_r S_r}\right)^{-1}\right\|_{\infty,2} > \epsilon\right] \\
&\leq 2\exp\left(-\frac{\left(\epsilon\sqrt{n} - \frac{(1+C_{\min})\sqrt{d_r}}{C_{\min}}\right)^2}{8} + \log\left((m-1)^2(p-1-d_r)\right)\right).
\end{aligned}$$

For the third term, we have

$$\begin{aligned}
&\left\|\left(Q^n_{S_r^c S_r} - Q^*_{S_r^c S_r}\right)\left(\left(Q^n_{S_r S_r}\right)^{-1} - \left(Q^*_{S_r S_r}\right)^{-1}\right)\right\|_{\infty,2} \\
&\leq \left\|Q^n_{S_r^c S_r} - Q^*_{S_r^c S_r}\right\|_{\infty,2} \Lambda_{\max}\left(\left(Q^n_{S_r S_r}\right)^{-1} - \left(Q^*_{S_r S_r}\right)^{-1}\right) \\
&\leq \frac{\sqrt{d_r}}{\sqrt{n}} \frac{\sqrt{d_r}}{C^2_{\min}\sqrt{n}} = \frac{d_r}{C^2_{\min} n}
\end{aligned}$$

The last inequality follows from (14) and (17). Hence, we have

$$\begin{aligned}
&\mathbb{P}\left[\left\|\left(Q^n_{S_r^c S_r} - Q^*_{S_r^c S_r}\right)\left(\left(Q^n_{S_r S_r}\right)^{-1} - \left(Q^*_{S_r S_r}\right)^{-1}\right)\right\|_{\infty,2} > \epsilon\right] \\
&\leq 6\exp\Bigg(-\frac{\left(\bar{C}_{\min}\epsilon\sqrt{n} - \left(1 + \frac{\sqrt{d_r}}{C^2_{\min}\sqrt{n}}\right)\sqrt{d_r}\right)^2}{8} \\
&\qquad\qquad + \log\left((m-1)^2(p-1-d_r)\right)\Bigg).
\end{aligned}$$

The result follows by substituting ϵ with $\frac{\alpha}{3\sqrt{d_r}}$.

(B3) We can write

$$\begin{aligned}
&\mathbb{P}\left[\Lambda_{\max}\left(\mathcal{J}^n\right) > D_{\max} + \epsilon\right] \\
&\qquad \leq \mathbb{P}\left[\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{J}^{(i)} - \mathcal{J}^*\right)\right\|_F > \epsilon\right].
\end{aligned}$$

Consequently, same analysis as part (B1) gives the result.

This concludes the proof of the Lemma.

9 Sufficiency Lemmas for Pairwise Dependencies

Lemma 5. *The constructed candidate primal-dual pair* $\left(\hat{\Theta}_{\setminus r}, \hat{Z}_{\setminus r}\right)$ *satisfy the conditions of the Lemma 3*

with probability $1-c_1\exp(-c_2 n)$ *for some positive constants* $c_1, c_2 \in \mathbb{R}$.

Proof. Using the mean-value theorem, for some $\bar{\Theta}_{\backslash r}$ in the convex combination of $\hat{\Theta}_{\backslash r}$ and $\Theta^*_{\backslash r}$, we have

$$\begin{aligned}
&\nabla^2 \ell\left(\Theta^*_{\backslash r}; D\right)\left[\hat{\Theta}_{\backslash r} - \Theta^*_{\backslash r}\right] \\
&= \nabla \ell\left(\hat{\Theta}_{\backslash r}; D\right) - \nabla \ell\left(\Theta^*_{\backslash r}; D\right) \\
&\quad + \left(\nabla^2 \ell\left(\Theta^*_{\backslash r}; D\right) - \nabla^2 \ell\left(\bar{\Theta}_{\backslash r}; D\right)\right)\left[\hat{\Theta}_{\backslash r} - \Theta^*_{\backslash r}\right] \\
&= -\lambda_n \hat{Z}_{\backslash r} - \underbrace{\nabla \ell\left(\Theta^*_{\backslash r}; D\right)}_{W^n_{\backslash r}} \\
&\quad + \underbrace{\left(\nabla^2 \ell\left(\Theta^*_{\backslash r}; D\right) - \nabla^2 \ell\left(\bar{\Theta}_{\backslash r}; D\right)\right)\left[\hat{\Theta}_{\backslash r} - \Theta^*_{\backslash r}\right]}_{R^n_{\backslash r}}.
\end{aligned}$$

We can rewrite these set of equations as two sets of equations over S_r and S_r^c. By Lemma 1, the Hessian sub-matrix on S_r is invertible with high probability and thus we get

$$\begin{aligned}
Q^n_{S_r^c S_r}\left(Q^n_{S_r S_r}\right)^{-1}&\left(-\lambda_n\left(\hat{Z}_{\backslash r}\right)_{S_r} - \left(W^n_{\backslash r}\right)_{S_r} + \left(R^n_{\backslash r}\right)_{S_r}\right) \\
&= -\lambda_n\left(\hat{Z}_{\backslash r}\right)_{S_r^c} - \left(W^n_{\backslash r}\right)_{S_r^c} + \left(R^n_{\backslash r}\right)_{S_r^c}.
\end{aligned}$$

Equivalently, we get

$$\begin{aligned}
\left(\hat{Z}_{\backslash r}\right)_{S_r^c} &= \frac{1}{\lambda_n}\left[\left(W^n_{\backslash r}\right)_{S_r^c} - \left(R^n_{\backslash r}\right)_{S_r^c}\right] \\
&\quad - \frac{1}{\lambda_n} Q^n_{S_r^c S_r}\left(Q^n_{S_r S_r}\right)^{-1}\left(\left(W^n_{\backslash r}\right)_{S_r} - \left(R^n_{\backslash r}\right)_{S_r}\right) \\
&\quad + Q^n_{S_r^c S_r}\left(Q^n_{S_r S_r}\right)^{-1}\left(\hat{Z}_{\backslash r}\right)_{S_r}.
\end{aligned}$$

Notice that $\left\|\left(\hat{Z}_{\backslash r}\right)_{S_r}\right\|_{\infty,2} = 1$. Thus, we can establish the following bound

$$\begin{aligned}
&\left\|\left(\hat{Z}_{\backslash r}\right)_{S_r^c}\right\|_{\infty,2} \\
&\quad \le \left(1 + \left\|Q^n_{S_r^c S_r}\left(Q^n_{S_r S_r}\right)^{-1}\right\|_{\infty,2}\sqrt{d_r}\right) \\
&\qquad \left[\frac{\left\|W^n_{\backslash r}\right\|_{\infty,2}}{\lambda_n} + \frac{\left\|R^n_{\backslash r}\right\|_{\infty,2}}{\lambda_n} + 1\right] - 1 \\
&\quad \le (2-\alpha)\left(\frac{\alpha}{4(2-\alpha)} + \frac{\alpha}{4(2-\alpha)} + 1\right) - 1 \\
&\quad = 1 - \frac{\alpha}{2} < 1.
\end{aligned}$$

The second inequality holds with high probability according to Lemma 1 and Lemma 6.

□

Lemma 6. *For quantities defined in the proof of Lemma 5, the following inequalities hold:*

$$\begin{aligned}
&\mathbb{P}\left[\frac{\left\|W^n_{\backslash r}\right\|_{\infty,2}}{\lambda_n} \ge \frac{\alpha}{4(2-\alpha)}\right] \\
&\quad \le 2\exp\left(-\frac{\left(\frac{\alpha}{4(2-\alpha)}\lambda_n\sqrt{n} - \frac{m-1}{2}\right)^2}{4} + \log(p-1)\right) \\
&\mathbb{P}\left[\frac{\left\|R^n_{\backslash r}\right\|_{\infty,2}}{\lambda_n} > \frac{\alpha}{4(2-\alpha)}\right] \\
&\quad \le 2\exp\left(-\frac{\left(\frac{\alpha}{4(2-\alpha)}\lambda_n\sqrt{n} - \frac{m-1}{2}\right)^2}{4} + \log(p-1)\right).
\end{aligned}$$

Proof. The first inequality follows directly from (12), for $\epsilon = \frac{\alpha}{4(2-\alpha)}\lambda_n - \frac{m-1}{2\sqrt{n}}$, provided that $\lambda_n \ge \frac{2(2-\alpha)}{\alpha}\frac{m-1}{\sqrt{n}}$. This probability goes to zero, if $\lambda_n \ge \frac{8(2-\alpha)}{\alpha}\left(\sqrt{\frac{\log(p-1)}{n}} + \frac{m-1}{4\sqrt{n}}\right)$.

Before we proceed, we want to point out a technical fact that we will use it through the rest of the proof. For λ_n achieves the lower bound mentioned above, any positive value K and $n \ge \frac{1}{K^2}\frac{64(2-\alpha)^2}{\alpha^2}\left(\sqrt{\log(p-1)} + \frac{m-1}{4}\right)^2 d_r^2$, we have $\lambda_n d_r \le K$. Hence, we can assume $\lambda_n d_r$ is less than any *fixed* constant K for sufficiently large n.

In order to bound $R^n_{\backslash r}$, we need to bound $\left\|\left(\hat{\Theta}_{\backslash r}\right)_{S_r} - \left(\Theta^*_{\backslash r}\right)_{S_r}\right\|_{\infty,2}$, using the technique used in Rothman et al. [28]. Let $G : \mathbb{R}^{(m-1)^2 d_r} \to \mathbb{R}$ be a function defined as

$$\begin{aligned}
G\Big((U)_{S_r}\Big) &:= \ell\left(\left(\Theta^*_{\backslash r}\right)_{S_r} + (U)_{S_r}; D\right) - \ell\left(\left(\Theta^*_{\backslash r}\right)_{S_r}; D\right) \\
&\quad + \lambda_n\left(\left\|\left(\Theta^*_{\backslash r}\right)_{S_r} + (U)_{S_r}\right\|_{1,2} - \left\|\left(\Theta^*_{\backslash r}\right)_{S_r}\right\|_{1,2}\right).
\end{aligned}$$

By optimality of $\hat{\Theta}_{\backslash r}$, it is clear that $\left(\hat{U}\right)_{S_r} = \left(\hat{\Theta}_{\backslash r}\right)_{S_r} - \left(\Theta^*_{\backslash r}\right)_{S_r}$ minimizes G. Since $G(\mathbf{0}) = 0$ by construction, we have $G\Big(\left(\hat{U}\right)_{S_r}\Big) \le 0$. Suppose there exist an ℓ_∞/ℓ_2 ball with radius B_r such that for any $\left\|(U)_{S_r}\right\|_{\infty,2} = B_r$, we have that $G\Big((U)_{S_r}\Big) > 0$. Then, we can claim that $\left\|\left(\hat{U}\right)_{S_r}\right\|_{\infty,2} \le B_r$; because if, in contrary, we assume that $\left(\hat{U}\right)_{S_r}$ is outside the ball,

then for an appropriate choice of $t \in (0,1)$, the point $t\left(\hat{U}\right)_{S_r} + (1-t)\mathbf{0}$ lies on the boundary of the ball. By convexity of G, we have

$$G\left(t\left(\hat{U}\right)_{S_r} + (1-t)\mathbf{0}\right) \leq t\, G\left(\left(\hat{U}\right)_{S_r}\right) + (1-t)G\left(\mathbf{0}\right) \leq 0.$$

This is a contradiction to the assumption of the positivity of G on the boundary of the ball.

Let $(U)_{Sr} \in \mathbb{R}^{(m-1)^2 d_r}$ be an arbitrary vector with $\left\|(U)_{S_r}\right\|_{\infty,2} = \frac{5}{C_{\min}}\lambda_n$. Applying mean value theorem to the log liklihood function, for some $\beta \in [0,1]$, we get

$$\begin{aligned} G\Big((U)_{S_r}\Big) &= \left\langle \left(W_{\setminus r}\right)_{S_r}, (U)_{S_r} \right\rangle \\ &+ \left\langle (U)_{S_r}, \nabla^2 \ell\left(\left(\Theta^*_{\setminus r}\right)_{S_r} + \beta\,(U)_{S_r}; D\right)(U)_{S_r} \right\rangle \\ &+ \lambda_n\left(\left\|\left(\Theta^*_{\setminus r}\right)_{S_r} + (U)_{S_r}\right\|_{1,2} - \left\|\left(\Theta^*_{\setminus r}\right)_{S_r}\right\|_{1,2}\right). \end{aligned} \tag{19}$$

We bound each of these three terms individually. By Cauchy-Schwartz inequality, we have

$$\begin{aligned} \left|\left\langle \left(W_{\setminus r}\right)_{S_r}, (U)_{S_r} \right\rangle\right| &\leq \left\|\left(W_{\setminus r}\right)_{S_r}\right\|_{\infty,2} \left\|(U)_{S_r}\right\|_{1,2} \\ &\leq \frac{\alpha}{4(2-\alpha)}\lambda_n d_r \frac{5}{C_{\min}}\lambda_n \\ &\leq \frac{5}{4C_{\min}} d_r \lambda_n^2. \end{aligned}$$

Moreover, by triangle inequality,

$$\begin{aligned} \lambda_n\left(\left\|\left(\Theta^*_{\setminus r}\right)_{S_r} + (U)_{S_r}\right\|_{1,2} - \left\|\left(\Theta^*_{\setminus r}\right)_{S_r}\right\|_{1,2}\right) &\geq -\lambda_n \left\|(U)_{S_r}\right\|_{1,2} \\ &\geq -\frac{5}{C_{\min}} d_r \lambda_n^2. \end{aligned}$$

To bound the other term, notice that by Tailor expansion, we get

$$\begin{aligned} &\Lambda_{\min}\left(\nabla^2 \ell\left(\left(\Theta^*_{\setminus r}\right)_{S_r} + \beta\,(U)_{S_r}; D\right)\right) \\ &\geq \min_{\beta\in[0,1]} \Lambda_{\min}\left(\nabla^2 \ell\left(\left(\Theta^*_{\setminus r}\right)_{S_r} + \beta\,(U)_{S_r}; D\right)\right) \\ &\geq \Lambda_{\min}\left(Q^*_{S_r S_r}\right) \\ &\quad - \max_{\beta\in[0,1]} \Lambda_{\max}\left(\left\langle \left.\frac{\partial \nabla^2 \ell\left(\Theta_{S_r}; D\right)}{\partial \Theta_{S_r}}\right|_{\left(\Theta^*_{\setminus r}\right)_{S_r} + \beta(U)_{S_r}}, (U)_{S_r} \right\rangle\right) \\ &\geq C_{\min} - \left(\max_{t_3 \in V\setminus\{r\}} \left\|\frac{\partial}{\partial \theta_{rt_3;\ell_3 k_3}} \eta_{\ell_1 \ell_2}\left(x^{(i)}\right)\right\|_2 \sqrt{d_r}\right) \\ &\qquad \Lambda_{\max}(\Im^*)\sqrt{d_r}\left\|(U)_{S_r}\right\|_{\infty,2}, \end{aligned} \tag{20}$$

where, $\eta(\cdot)$ is defined in Section 7.2. We know that $\Lambda_{\max}(\Im^*) = \Lambda_{\max}(\mathcal{J}^*)$ as a property of Kronecher product. By (18) and assumption on the maximum eigenvalue of $\mathcal{J}^*$, we have

$$\begin{aligned} &\Lambda_{\min}\left(\nabla^2 \ell\left(\left(\Theta^*_{\setminus r}\right)_{S_r} + \beta\,(U)_{S_r}; D\right)\right) \\ &\quad \geq C_{\min} - \frac{m-1}{\sqrt{2}}\, d_r D_{\max}\left\|(U)_{S_r}\right\|_{\infty,2} \\ &\quad \geq C_{\min} - \frac{m-1}{\sqrt{2}}\, d_r D_{\max}\frac{5}{C_{\min}}\lambda_n \\ &\quad \geq \frac{C_{\min}}{2} \qquad \left(\lambda_n d_r \leq \frac{C^2_{\min}}{\sqrt{50}(m-1)D_{\max}}\right). \end{aligned}$$

Hence, from (19), we get

$$G\Big((U)_{S_r}\Big) \geq d_r \frac{5}{C_{\min}}\lambda_n^2\left(-\frac{1}{4} + \frac{5}{2} - 1\right) > 0.$$

We can colclude that

$$\left\|\left(\hat{\Theta}_{\setminus r}\right)_{S_r} - \left(\Theta^*_{\setminus r}\right)_{S_r}\right\|_{\infty,2} \leq \frac{5}{C_{\min}}\lambda_n. \tag{21}$$

with high probability. With similar analysis on the maximum eigenvalue of the derivative of Hessian as in (20), it is easy to show that

$$\begin{aligned} &\frac{\left\|R^n_{\setminus r}\right\|_{\infty,2}}{\lambda_n} \\ &\quad \leq \frac{1}{\lambda_n}\frac{m-1}{\sqrt{2}}\, d_r D_{\max}\left\|\left(\hat{\Theta}_{\setminus r}\right)_{S_r} - \left(\Theta^*_{\setminus r}\right)_{S_r}\right\|^2_{\infty,2} \\ &\quad \leq \frac{m-1}{\sqrt{2}}\, d_r D_{\max}\frac{25}{C^2_{\min}}\lambda_n \\ &\quad \leq \frac{\alpha}{4(2-\alpha)}, \end{aligned}$$

provided that $\lambda_n d_r \leq \frac{C^2_{\min}}{50\sqrt{2}(m-1)D_{\max}}\frac{\alpha}{2-\alpha}$. □

10 Proof of Lemma 4

(D1) By variational representation of the smallest eigenvalue, we have

$$\begin{aligned}
&\Lambda_{\min}\left(\left[\nabla^2\ell\Big(\bar{\Theta}^*_P;D\Big)\right]_{S_rS_r}\right)\\
&\geq \Lambda_{\min}\left(\left[\nabla^2\ell\Big(\bar{\Theta}^*_{\setminus r};D\Big)\right]_{S_rS_r}\right)\\
&\quad-\Lambda_{\max}\left(\left[\nabla^2\ell\Big(\bar{\Theta}^*_{\setminus r};D\Big)\right]_{S_rS_r}-\left[\nabla^2\ell\Big(\bar{\Theta}^*_P;D\Big)\right]_{S_rS_r}\right)\\
&\geq C_{\min}(1+\gamma)\\
&\quad-\Lambda_{\max}\left(\left[\nabla^2\ell\Big(\bar{\Theta}^*_{\setminus r};D\Big)\right]_{S_rS_r}-\left[\nabla^2\ell\Big(\bar{\Theta}^*_P;D\Big)\right]_{S_rS_r}\right).
\end{aligned}$$

In the second inequality, we used the result of Lemma 1, i.e., the inequality holds with probability stated in Lemma 4. By Tailor expansion, for some $\beta\in[0,1]$, and by (23), we get

$$\begin{aligned}
&\Lambda_{\max}\left(\left[\nabla^2\ell\Big(\bar{\Theta}^*_{\setminus r};D\Big)\right]_{S_rS_r}-\left[\nabla^2\ell\Big(\bar{\Theta}^*_P;D\Big)\right]_{S_rS_r}\right)\\
&\leq \Lambda_{\max}\left(\left\langle\frac{\partial\left[\nabla^2\ell\Big(\bar{\Theta};D\Big)\right]_{S_rS_r}}{\partial\bar{\Theta}}\bigg|_{\bar{\Theta}^*_{\setminus r}-\beta\bar{\Theta}^*_{P^c}},\bar{\Theta}^*_{P^c}\right\rangle\right)\\
&\leq \left\|\nabla\eta_{\ell_1\ell_2}\left(x^{(i)}\right)\right\|_\infty D_{\max}\left\|\bar{\Theta}^*_{P^c}\right\|_1\\
&=\gamma C_{\min}.
\end{aligned}$$

Note that $\left\|\nabla\eta_{\ell_1\ell_2}\left(x^{(i)}\right)\right\|_\infty\leq 1$ for $\eta(\cdot)$ defined in section 7.3. The last inequality holds as a result of Lemma 1 with the probability stated in Lemma 4. Hence, the result follows.

(D2) We can write

$$\nabla^2\ell\left(\bar{\Theta}^*_P;D\right)_{S_r^cS_r}\Big(\nabla^2\ell\left(\bar{\Theta}^*_P;D\right)_{S_rS_r}\Big)^{-1}=\sum_{i=0}^{3}T_i,$$

where,

$$\begin{aligned}
T_0&=\nabla^2\ell\left(\bar{\Theta}^*_{\setminus r};D\right)_{S_r^cS_r}\Big(\nabla^2\ell\left(\bar{\Theta}^*_{\setminus r};D\right)_{S_rS_r}\Big)^{-1}\\
T_1&=\nabla^2\ell\left(\bar{\Theta}^*_{\setminus r};D\right)_{S_r^cS_r}\\
&\qquad\left(\Big(\nabla^2\ell\left(\bar{\Theta}^*_P;D\right)_{S_rS_r}\Big)^{-1}-\Big(\nabla^2\ell\left(\bar{\Theta}^*_{\setminus r};D\right)_{S_rS_r}\Big)^{-1}\right)\\
T_2&=\left(\nabla^2\ell\left(\bar{\Theta}^*_P;D\right)_{S_r^cS_r}-\nabla^2\ell\left(\bar{\Theta}^*_{\setminus r};D\right)_{S_r^cS_r}\right)\\
&\qquad\qquad\Big(\nabla^2\ell\left(\bar{\Theta}^*_{\setminus r};D\right)_{S_rS_r}\Big)^{-1}\\
T_3&=\left(\nabla^2\ell\left(\bar{\Theta}^*_P;D\right)_{S_r^cS_r}-\nabla^2\ell\left(\bar{\Theta}^*_{\setminus r};D\right)_{S_r^cS_r}\right)\\
&\qquad\left(\Big(\nabla^2\ell\left(\bar{\Theta}^*_P;D\right)_{S_rS_r}\Big)^{-1}-\Big(\nabla^2\ell\left(\bar{\Theta}^*_{\setminus r};D\right)_{S_rS_r}\Big)^{-1}\right).
\end{aligned}$$

By Lemma 1, we have that $\|T_0\|_{\infty,1}\leq\frac{1-\tau}{\sqrt{d_r}}$ with the probability stated in Lemma 4. For the second term, we have

$$\begin{aligned}
&\|T_1\|_{\infty,2}\\
&\leq\|T_0\|_{\infty,2}\Lambda_{\max}\left(\underbrace{\nabla^2\ell\left(\bar{\Theta}^*_{\setminus r};D\right)_{S_rS_r}-\nabla^2\ell\left(\bar{\Theta}^*_P;D\right)_{S_rS_r}}_{T_{12}}\right)\\
&\qquad\Lambda_{\max}\left(\underbrace{\Big(\nabla^2\ell\left(\bar{\Theta}^*_P;D\right)_{S_rS_r}\Big)^{-1}}_{T_{13}}\right)\\
&\leq\frac{1-\tau}{\sqrt{d_r}}\gamma C_{\min}\frac{1}{C_{\min}}\ =\frac{1-\tau}{\sqrt{d_r}}\gamma.
\end{aligned}$$

We used the result of (D1) for $\Lambda_{\max}\left(T_{13}\right)\leq\frac{1}{C_{\min}}$.

For the third term, we have

$$\begin{aligned}
\|T_2\|_{\infty,2}&\leq\left\|\underbrace{\nabla^2\ell\left(\bar{\Theta}^*_{\setminus r};D\right)_{S_r^cS_r}-\nabla^2\ell\left(\bar{\Theta}^*_P;D\right)_{S_r^cS_r}}_{T_{21}}\right\|_{\infty,2}\\
&\qquad\Lambda_{\max}\left(\underbrace{\Big(\nabla^2\ell\left(\bar{\Theta}^*_{\setminus r};D\right)_{S_rS_r}\Big)^{-1}}_{T_{22}}\right)\\
&\leq\gamma C_{\min}\frac{1}{C_{\min}(1+\gamma)}\\
&=\frac{\gamma}{1+\gamma}.
\end{aligned}$$

For the fourth term, we have

$$\begin{aligned}
\|T_3\|_{\infty,2}&\leq\|T_{21}\|_{\infty,2}\,\Lambda_{\max}\left(T_{22}\right)\Lambda_{\max}\left(T_{12}\right)\Lambda_{\max}\left(T_{13}\right)\\
&\leq\gamma C_{\min}\frac{1}{C_{\min}(1+\gamma)}\gamma C_{\min}\frac{1}{C_{\min}}\\
&\leq\frac{\gamma^2}{1+\gamma}.
\end{aligned}$$

Putting all piences together, we get the result.

(D3) The result follows directly from Lemma 1.

This concludes the proof of Lemma.

11 Sufficiency Lemmas for Higher Order Dependencies

Lemma 7. *The constructed candidate primal-dual pair* $\left(\hat{\Theta}_{\setminus r},\hat{Z}_{\setminus r}\right)$ *satisfy the conditions of the Lemma 3*

with probability $1-c_1\exp(-c_2 n)$ *for some positive constants* $c_1, c_2 \in \mathbb{R}$.

Proof. Using the mean-value theorem, for some $\bar{\Theta}_{\backslash r}$ in the convex combination of $\hat{\Theta}_{\backslash r}$ and $\bar{\Theta}^*_P$, we have

$$\begin{aligned}
&\nabla^2 \ell\left(\bar{\Theta}^*_P; D\right)\left[\hat{\Theta}_{\backslash r} - \bar{\Theta}^*_P\right] \\
&= \nabla \ell\left(\hat{\Theta}_{\backslash r}; D\right) - \nabla \ell\left(\bar{\Theta}^*_P; D\right) \\
&\quad + \left(\nabla^2 \ell\left(\bar{\Theta}^*_P; D\right) - \nabla^2 \ell\left(\bar{\Theta}_{\backslash r}; D\right)\right)\left[\hat{\Theta}_{\backslash r} - \bar{\Theta}^*_P\right] \\
&= -\lambda_n \hat{Z}_{\backslash r} - \underbrace{\nabla \ell\left(\bar{\Theta}^*_P; D\right)}_{\bar{W}^n_{\backslash r}} \\
&\quad + \underbrace{\left(\nabla^2 \ell\left(\bar{\Theta}^*_P; D\right) - \nabla^2 \ell\left(\bar{\Theta}_{\backslash r}; D\right)\right)\left[\hat{\Theta}_{\backslash r} - \bar{\Theta}^*_P\right]}_{\bar{R}^n_{\backslash r}}.
\end{aligned}$$

We can rewrite these set of equations as two sets of equations over S_r and S_r^c. By Lemma 4, the Hessian sub-matrix on S_r is invertible with high probability and thus we get

$$\begin{aligned}
&\nabla^2 \ell\left(\bar{\Theta}^*_P; D\right)_{S_r^c S_r}\left(\nabla^2 \ell\left(\bar{\Theta}^*_P; D\right)_{S_r S_r}\right)^{-1} \\
&\left(-\lambda_n \left(\hat{Z}_{\backslash r}\right)_{S_r} - \left(\bar{W}^n_{\backslash r}\right)_{S_r} + \left(\bar{R}^n_{\backslash r}\right)_{S_r}\right) \\
&\qquad = -\lambda_n \left(\hat{Z}_{\backslash r}\right)_{S_r^c} - \left(\bar{W}^n_{\backslash r}\right)_{S_r^c} + \left(\bar{R}^n_{\backslash r}\right)_{S_r^c}.
\end{aligned}$$

Notice that $\left\|\left(\hat{Z}_{\backslash r}\right)_{S_r}\right\|_{\infty,2} = 1$ and hence, we get

$$\begin{aligned}
&\left\|\left(\hat{Z}_{\backslash r}\right)_{S_r^c}\right\|_{\infty,2} \\
&\leq \left(1 + \left\|\nabla^2 \ell\left(\bar{\Theta}^*_P; D\right)_{S_r^c S_r}\left(\nabla^2 \ell\left(\bar{\Theta}^*_P; D\right)_{S_r S_r}\right)^{-1}\right\|_{\infty,2} \sqrt{d_r}\right) \\
&\qquad \left[\frac{\left\|\bar{W}^n_{\backslash r}\right\|_{\infty,2}}{\lambda_n} + \frac{\left\|\bar{R}^n_{\backslash r}\right\|_{\infty,2}}{\lambda_n} + 1\right] - 1 \\
&\leq (2-\alpha)\left(\frac{\alpha}{4(2-\alpha)} + \frac{\alpha}{4(2-\alpha)} + 1\right) - 1 \\
&= 1 - \frac{\alpha}{2} < 1.
\end{aligned}$$

The second inequality holds with high probability according to Lemma 4 and Lemma 8.

□

Lemma 8. *For quantities defined in the proof of Lemma 7, the following inequalities hold:*

$$\begin{aligned}
&\mathbb{P}\left[\frac{\left\|\bar{W}^n_{\backslash r}\right\|_{\infty,2}}{\lambda_n} > \frac{\alpha}{4(2-\alpha)}\right] \\
&\leq 2\exp\left(-\frac{\left(\left(\frac{\alpha}{4(2-\alpha)}\lambda_n - \frac{1}{2}\left\|\bar{\Theta}^*_{P^c}\right\|_1\right)\sqrt{n} - \frac{m-1}{2}\right)^2}{4} \right. \\
&\qquad \left. + \log(p-1)\right) \\
&\mathbb{P}\left[\frac{\left\|\bar{R}^n_{\backslash r}\right\|_{\infty,2}}{\lambda_n} > \frac{\alpha}{4(2-\alpha)}\right] \\
&\leq 2\exp\left(-\frac{\left(\left(\frac{\alpha}{4(2-\alpha)}\lambda_n - \frac{1}{2}\left\|\bar{\Theta}^*_{P^c}\right\|_1\right)\sqrt{n} - \frac{m-1}{2}\right)^2}{4} \right. \\
&\qquad \left. + \log(p-1)\right).
\end{aligned}$$

Proof. By simple derivation, we have

$$\begin{aligned}
\frac{\partial}{\partial \bar{\theta}^*_{rt;\ell k}} \ell^{(i)}(\bar{\Theta}_P; D) &= \mathcal{I}\left[x_t^{(i)} = k\right] \\
&\left(\mathcal{I}\left[x_r^{(i)} = \ell\right] - \mathbb{P}_{\bar{\Theta}^*_P}\left[X_r = \ell \mid X_{\backslash r} = x_{\backslash r}^{(i)}\right]\right).
\end{aligned}$$

It is easy to show that

$$\begin{aligned}
&\mathbb{E}_{\bar{\Theta}^*_{\backslash r}}\left[\frac{\partial}{\partial \bar{\theta}^*_{rt;\ell k}} \ell^{(i)}(\bar{\Theta}_P; D)\right] \\
&= \mathbb{P}_{\bar{\Theta}^*_{\backslash r}}\left[X_r = \ell \mid X_t = k, X_{\backslash r,t} = x_{\backslash r,t}\right] \\
&\qquad - \mathbb{P}_{\bar{\Theta}^*_P}\left[X_r = \ell \mid X_t = k, X_{\backslash r,t} = x_{\backslash r,t}\right] \\
&\leq \left\|\bar{\Theta}^*_{P^c}\right\|_1 \\
&\quad \max_{\beta \in [0,1]} \left\|\nabla \mathbb{P}_{\bar{\Theta}^*_{\backslash r} - \beta \bar{\Theta}^*_{P^c}}\left[X_r = \ell \,|\, X_t = k, X_{\backslash r,t} = x_{\backslash r,t}\right]\right\|_\infty \\
&\leq \frac{1}{4}\left\|\bar{\Theta}^*_{P^c}\right\|_1,
\end{aligned}$$

where, with abuse of notation $\bar{\Theta}^*_{\backslash r} - \beta\bar{\Theta}^*_{P^c}$ represents the matrix $\bar{\Theta}^*_{\backslash r}$ purturbed only on the entries corresponding to $\bar{\Theta}^*_{P^c}$. Also, one can show that $\mathrm{Var}\left(\frac{\partial}{\partial \bar{\theta}^*_{rt;\ell k}} \ell^{(i)}(\Theta_{\backslash r}; D)\right) \leq \frac{1}{4}$. Consequently, with i.i.d assumption on drawn samples, we have $\mathrm{Var}\left(\frac{\partial}{\partial \bar{\theta}^*_{rt;\ell k}} \ell(\Theta_{\backslash r}; D)\right) \leq \frac{1}{4n}$. For a fixed $t \in V \backslash \{r\}$

by Jensen's inequality,

$$\begin{aligned}\mathbb{E}_{\Theta^*_{\backslash r}}\left[\left\|\frac{\partial}{\partial\bar{\theta}^*_{rt;\ell k}}\ell(\Theta_{\backslash r};D)\right\|_2\right] &\leq \sqrt{\mathbb{E}_{\Theta^*_{\backslash r}}\left[\left\|\frac{\partial}{\partial\bar{\theta}^*_{rt;\ell k}}\ell(\Theta_{\backslash r};D)\right\|_2^2\right]}\\ &\leq \frac{1}{2}\sqrt{\frac{(m-1)^2}{n}+\left\|\bar{\Theta}^*_{P^c}\right\|_1^2}\\ &\leq \frac{m-1}{2\sqrt{n}}+\frac{1}{2}\left\|\bar{\Theta}^*_{P^c}\right\|_1.\end{aligned}$$

We have $\max_{t\in V\backslash\{r\}}\left\|\frac{\partial}{\partial\bar{\theta}^*_{rt;\ell k}}\ell^{(i)}(\Theta_{\backslash r};D)\right\|_2\leq\sqrt{2}$ for all i and hence, by Azuma-Hoeffding inequality and the union bound, we get

$$\begin{aligned}\mathbb{P}\left[\left\|\frac{\partial}{\partial\bar{\theta}^*_{rt;\ell k}}\ell(\Theta_{\backslash r};D)\right\|_{\infty,2}>\frac{m-1}{2\sqrt{n}}+\frac{1}{2}\left\|\bar{\Theta}^*_{P^c}\right\|_1+\epsilon\right]\\ \leq 2\exp\left(-\frac{\epsilon^2}{4}n+\log(p-1)\right).\end{aligned}$$

For $\lambda_n\geq\frac{8(2-\alpha)}{\alpha}\left(\frac{m-1}{4\sqrt{n}}+\frac{1}{4}\left\|\bar{\Theta}^*_{P^c}\right\|_1\right)$, the result follows.

In order to bound $\bar{R}^n_{\backslash r}$, we need to control the estimation error $\left(\hat{\Theta}_{\backslash r}\right)_{S_r}-\left(\bar{\Theta}^*_P\right)_{S_r}$. Let $H:\mathbb{R}^{(m-1)^2d_r}\to\mathbb{R}$ be a function defined as

$$\begin{aligned}H(U_{S_r}):=\ell\left(\left(\bar{\Theta}^*_P\right)_{S_r}+U_{S_r};D\right)-\ell\left(\left(\bar{\Theta}^*_P\right)_{S_r};D\right)\\ +\lambda_n\left(\left\|\left(\bar{\Theta}^*_P\right)_{S_r}+U_{S_r}\right\|_{1,2}-\left\|\left(\bar{\Theta}^*_P\right)_{S_r}\right\|_{1,2}\right).\end{aligned}$$

By optimality of $\hat{\Theta}_{\backslash r}$, it is clear that $U^*=\left(\hat{\Theta}_{\backslash r}\right)_{S_r}-\left(\bar{\Theta}^*_P\right)_{S_r}$ minimizes H. Since $H(\mathbf{0})=0$ by construction, we have $H(U^*)\leq 0$. Suppose there exist an ℓ_∞/ℓ_2 ball with radius B_r such that for any $\|U\|_{\infty,2}=B_r$, we have that $H(U)>0$. Then, we can claim that $\|U^*\|_{\infty,2}\leq B_r$. See proof of Lemma 6 for more discussion on this proof technique. Let $U_0\in\mathbb{R}^{(m-1)^2d_r}$ be an arbitrary vector with $\|U_0\|_{\infty,2}=\frac{5}{C_{\min}}\lambda_n$. We have

$$\begin{aligned}H(U_0):=\ell\left(\left(\bar{\Theta}^*_P\right)_{S_r}+U_0;D\right)-\ell\left(\left(\bar{\Theta}^*_P\right)_{S_r};D\right)\\ +\lambda_n\left(\left\|\left(\bar{\Theta}^*_P\right)_{S_r}+U_0\right\|_{1,2}-\left\|\left(\bar{\Theta}^*_P\right)_{S_r}\right\|_{1,2}\right).\end{aligned}\tag{22}$$

We bound each of these three terms individually. Applying mean value theorem to the log liklihood function, for some $\beta\in[0,1]$, we get

$$\begin{aligned}&\ell\left(\left(\bar{\Theta}^*_P\right)_{S_r}+U_0;D\right)-\ell\left(\left(\bar{\Theta}^*_P\right)_{S_r};D\right)\\ &=\left\langle\left(\bar{W}^n_{\backslash r}\right)_{S_r},U_0\right\rangle+\left\langle U_0,\nabla^2\ell\left(\left(\bar{\Theta}^*_P\right)_{S_r}+\beta U_0;D\right)U_0\right\rangle.\end{aligned}$$

Note that $\frac{\alpha}{4(2-\alpha)}\lambda_n\leq\frac{1}{4}\lambda_n$ and hence, by our bound on $\bar{W}^n_{\backslash r}$ and Cauchy-Shwartz inequality, we have

$$\begin{aligned}\left|\left\langle\left(\bar{W}^n_{\backslash r}\right)_{S_r},U_0\right\rangle\right|&\leq\left\|\left(\bar{W}^n_{\backslash r}\right)_{S_r}\right\|_{\infty,2}\|U_0\|_{1,2}\\ &\leq\frac{\lambda_n}{4}d_r\|U_0\|_{\infty,2}\\ &\leq\frac{5}{4C_{\min}}\lambda_n^2d_r.\end{aligned}$$

To bound the other term, by Tailor expansion, we get

$$\begin{aligned}&\Lambda_{\min}\left(\nabla^2\ell\left(\left(\bar{\Theta}^*_P\right)_{S_r}+\beta U_0;D\right)\right)\\ &\geq\min_{\beta\in[0,1]}\Lambda_{\min}\left(\nabla^2\ell\left(\left(\bar{\Theta}^*_P\right)_{S_r}+\beta U_0;D\right)\right)\\ &\geq\Lambda_{\min}\left(\nabla^2\ell\left(\left(\bar{\Theta}^*_P\right)_{S_r};D\right)\right)\\ &\quad-\max_{\beta\in[0,1]}\Lambda_{\max}\left(\left\langle\left.\frac{\partial\nabla^2\ell\left(\left(\bar{\Theta}_P\right)_{S_r};D\right)}{\partial\left(\bar{\Theta}_P\right)_{S_r}}\right|_{\left(\bar{\Theta}^*_P\right)_{S_r}+\beta U_0},U_0\right\rangle\right)\\ &\geq C_{\min}\\ &\quad-\max_{t_3\in V\backslash\{r\}}\left\|\frac{\partial\eta_{\ell_1\ell_2}\left(x^{(i)}\right)}{\partial\bar{\theta}_{rt_3;\ell_3k_3}}\right\|_2d_r\ \Lambda_{\max}(\Im^*)\|U_0\|_{\infty,2}\\ &\geq C_{\min}-\frac{m-1}{\sqrt{2}}\ d_rD_{\max}\|U_0\|_{\infty,2}\\ &\geq\frac{C_{\min}}{2}\qquad\left(\lambda_nd_r\leq\frac{C^2_{\min}}{\sqrt{50}(m-1)D_{\max}}\right).\end{aligned}\tag{23}$$

Here, we used the fact that $\Lambda_{\max}(\Im^*)=\Lambda_{\max}(\mathcal{J}^*)$ as a property of Kronecher product and also our assumption on the maximum eigenvalue of $\mathcal{J}^*$. By triangle inequality,

$$\begin{aligned}\lambda_n\left(\left\|\bar{\Theta}^*_P+U_0\right\|_{1,2}-\left\|\bar{\Theta}^*_P\right\|_{1,2}\right)&\geq-\lambda_n\|U_0\|_{1,2}\\ &\geq-\lambda_nd_r\|U_0\|_{\infty,2}\\ &\geq-\frac{5\lambda_n^2d_r}{C_{\min}}.\end{aligned}$$

Hence, from (22), we get $H(U_0)\geq\frac{5\lambda_n^2d_r}{4C_{\min}}>0$ and hence,

$$\left\|\left(\hat{\Theta}_{\backslash r}\right)_{S_r}-\left(\bar{\Theta}^*_P\right)_{S_r}\right\|_{\infty,2}\leq\frac{5}{C_{\min}}\lambda_n,\tag{24}$$

with high probability. With similar analysis as in 23,

we have

$$\begin{aligned}
&\frac{\left\|\bar{R}^n_{\backslash r}\right\|_{\infty,2}}{\lambda_n} \\
&\quad\leq \frac{1}{\lambda_n}\frac{m-1}{\sqrt{2}}\, d_r D_{\max} \left\|\left(\hat{\Theta}_{\backslash r}\right)_{S_r} - \left(\Theta^*_{\backslash r}\right)_{S_r}\right\|^2_{\infty,2} \\
&\quad\leq \frac{m-1}{\sqrt{2}}\, d_r D_{\max} \frac{25}{C^2_{\min}}\lambda_n \\
&\quad\leq \frac{\alpha}{4(2-\alpha)},
\end{aligned}$$

provided that $\lambda_n d_r \leq \frac{C^2_{\min}}{50\sqrt{2}(m-1)D_{\max}}\frac{\alpha}{2-\alpha}$.

□