SUPPLEMENT TO "ASYMPTOTIC NORMALITY AND OPTIMALITIES IN ESTIMATION OF LARGE GAUSSIAN GRAPHICAL MODEL"

By Zhao Ren [‡], Tingni Sun[§], Cun-Hui Zhang[†,¶] and Harrison H. Zhou[*,‖]

University of Pittsburgh[‡], *University of Maryland*[§], *Rutgers University*[¶] *and Yale University*[‖]

In this supplement we collect proofs of Theorems 1-3 in Section 2, Theorems 6, 8 in Section 3 and Theorems 10-11 as well as Proposition 1 in Section 4.

APPENDIX A: PROOF OF THEOREMS 1-3

A.1. Proof of Theorems 2-4. We will only prove Theorems 2 and 3. The proof of Theorem 4 is omitted as it is similar to that of Theorems 2 and 3.

A.1.1. *Proof of Theorem 2.* We first prove (i). As θ_{ii} and θ_{jj} are uniformly bounded, the large deviation probability in (20) follows from (18) for θ_{ii}^{ora} and θ_{jj}^{ora}. We then need only to consider the entry θ_{ij}^{ora}. Recall that $\overline{\mathbf{D}} = \text{diag}(\mathbf{X}^T\mathbf{X}/n)$ and $\mathbf{X}_{A^c}$ is independent of ϵ_A. It follows that $(\mathbf{X}\overline{\mathbf{D}}^{-1/2})_k^T \epsilon_m/n \sim \mathcal{N}(0, \theta_{mm}/n)$ for all $m \in A$, so that

$$\mathbb{P}\left\{\left\|(\mathbf{X}\overline{\mathbf{D}}^{-1/2})_{A^c}^T \epsilon_m/n\right\|_\infty > \sqrt{\delta\theta_{mm}(2/n)\log p}\right\} \le \frac{p^{-\delta}(p-2)}{\sqrt{2\delta\log p}}$$

[*]The research of Harrison H. Zhou was supported in part by NSF Career Award DMS-0645676 and NSF FRG Grant DMS-0854975.

[†]The research of Cun-Hui Zhang was supported in part by the NSF Grants DMS-11-06753 and DMS-12-09014 and NSA Grant H98230-11-1-0205.

by the union bound. Thus, it follows from (16) and (17) that

$$
\begin{aligned}
\left|\hat{\theta}_{ij}-\theta_{ij}^{ora}\right| &= \left|\hat{\epsilon}_i^T\hat{\epsilon}_j/n-\epsilon_i^T\epsilon_j/n\right| \\
&= \left|\left(\epsilon_i+\mathbf{X}_{A^c}\left(\beta_i-\hat{\beta}_i\right)\right)^T\left(\epsilon_j+\mathbf{X}_{A^c}\left(\beta_j-\hat{\beta}_j\right)\right)/n-\epsilon_i^T\epsilon_j/n\right| \\
&\leq \left\|(\mathbf{X}\overline{\mathbf{D}}^{-1/2})_{A^c}^T\epsilon_i/n\right\|_\infty\left\|\overline{\mathbf{D}}_{A^c}^{1/2}(\beta_j-\hat{\beta}_j)\right\|_1 \\
&\quad+\left\|(\mathbf{X}\overline{\mathbf{D}}^{-1/2})_{A^c}^T\epsilon_j/n\right\|_\infty\left\|\overline{\mathbf{D}}_{A^c}^{1/2}(\beta_i-\hat{\beta}_i)\right\|_1 \\
&\quad+\left\|\mathbf{X}_{A^c}\left(\beta_i-\hat{\beta}_i\right)\right\|\cdot\left\|\mathbf{X}_{A^c}\left(\beta_j-\hat{\beta}_j\right)\right\|/n \\
&\leq 2\sqrt{\delta\theta_{mm}(2/n)\log p}\,C_0 s\sqrt{\delta(\log p)/n}+C_0 s\delta(\log p)/n \\
&= C_1 s\delta(\log p)/n
\end{aligned}
$$

with at least probability $1-2p^{-\delta+1}\epsilon_\Omega-2p^{-\delta+1}/\sqrt{2\log p}$, and (20) follows.

As $\Theta_{A,A}$ has a bounded spectrum, the functional $\zeta_{kl}\left(\Theta_{A,A}\right)=\left(\Theta_{A,A}^{-1}\right)_{kl}$ is Lipschitz in a neighborhood of $\Theta_{A,A}$ for $k,l\in A$, so that (21) is an immediate consequence of (20).

For part (ii), we note that the regression model (7) has Gaussian error and Gaussian design, and the complexity of $\beta_{A^c,m}$, $m\in A$, is controlled by $s_\lambda(\Omega)\leq s\leq c_0 n/\log p$ up to a constant factor. Moreover, as the spectrum of the population covariance matrix is contained in $[1/M,M]$, the noise level and the spectrum of the population Gram matrix $\mathbb{E}\mathbf{X}_{A^c}^T\mathbf{X}_{A^c}/n$ are all contained in $[1/M,M]$ in the linear model. Thus, part (ii) follows from Theorem 10 (i), Theorem 11 (ii) and Proposition 1.

For part (iii), define random vector $\eta^{ora}=\left(\eta_{ii}^{ora},\eta_{ij}^{ora},\eta_{jj}^{ora}\right)$, where $\eta_{kl}^{ora}=\sqrt{n}\frac{\theta_{kl}^{ora}-\theta_{kl}}{\sqrt{\theta_{kk}\theta_{ll}+\theta_{kl}^2}}$. The following result is a multidimensional version of KMT quantile inequality: there exist some constants D_0, $\vartheta\in(0,\infty)$ and random Gaussian vector $Z=(Z_{ii},Z_{ij},Z_{jj})\sim\mathcal{N}\left(0,\breve{\Sigma}\right)$ with $\breve{\Sigma}=\mathrm{Cov}(\eta^{ora})$ such that whenever $|Z_{kl}|\leq\vartheta\sqrt{n}$ for all kl, we have

$$
\left\|\eta^{ora}-Z\right\|_\infty\leq\frac{D_0}{\sqrt{n}}\left(1+Z_{ii}^2+Z_{ij}^2+Z_{jj}^2\right). \tag{82}
$$

See Proposition [KMT] in Mason and Zhou (2012) for one dimensional case and consult Einmahl (1989) for multidimensional case. Note that $\sqrt{n}\eta^{ora}$ can be written as a sum of n i.i.d. random vectors with mean zero and covariance matrix $\breve{\Sigma}$, each of which is sub-exponentially distributed. Hence the assumptions of KMT quantile inequality in literature are satisfied. With a slight abuse of notation, we define $\Theta=(\theta_{ii},\theta_{ij},\theta_{jj})$. To prove the desired coupling inequality (23), we use the Taylor expansion of the function $\omega_{ij}\left(\Theta\right)=$

$-\theta_{ij}/\left(\theta_{ii}\theta_{jj}-\theta_{ij}^2\right)$ to obtain

$$\begin{aligned} & \omega_{ij}^{ora}-\omega_{ij} \\ = \; & \left\langle \nabla\omega_{ij}\left(\Theta\right),\Theta^{ora}-\Theta\right\rangle+\sum_{|\beta|=2}R_{\beta}\left(\Theta^{ora}\right)\left(\Theta-\Theta^{ora}\right)^{\beta}. \end{aligned} \tag{83}$$

The multi-index notation of $\beta=(\beta_1,\beta_2,\beta_3)$ is defined as $|\beta|=\sum_k\beta_k$, $x^{\beta}=\prod_k x_k^{\beta_k}$ and $D^{\beta}f\left(x\right)=\frac{\partial^{|\beta|}f}{\partial x_1^{\beta_1}\partial x_2^{\beta_2}\partial x_3^{\beta_3}}$. The derivatives can be easily computed. To save the space, we omit their explicit formulas. The coefficients in the integral form of the remainder with $|\beta|=2$ have a uniform upper bound $\left|R_{\beta}\left(\Theta_{A,A}^{ora}\right)\right|\leq 2\max_{|\alpha|=2}\max_{\Theta\in B}D^{\alpha}\omega_{ij}\left(\Theta\right)\leq C_2$, where B is some sufficiently small compact ball with center Θ when Θ^{ora} is in this ball B, which is satisfied by picking a sufficiently small value ϑ in our assumption $\left\|\eta^{ora}\right\|_{\infty}\leq\vartheta\sqrt{n}$. Recall that κ_{ij}^{ora} and η^{ora} are standardized versions of $\left(\omega_{ij}^{ora}-\omega_{ij}\right)$ and $\left(\Theta-\Theta^{ora}\right)$. Consequently there exist some deterministic constants h_1,h_2,h_3 and D_{β} with $|\beta|=2$ such that we can rewrite (83) in terms of κ_{ij}^{ora} and η^{ora} as follows,

$$\kappa_{ij}^{ora}=h_1\eta_{ii}^{ora}+h_2\eta_{ij}^{ora}+h_3\eta_{jj}^{ora}+\sum_{|\beta|=2}\frac{D_{\beta}R_{\beta}\left(\Theta^{ora}\right)}{\sqrt{n}}\left(\eta^{ora}\right)^{\beta},$$

which, together with Equation (82), completes our proof of Equation (23),

$$\left|\kappa_{ij}^{ora}-Z'\right|\leq\left(\sum_{k=1}^{3}|h_k|\right)\left\|Z-\eta^{ora}\right\|_{\infty}+\frac{C_3}{\sqrt{n}}\left\|\eta^{ora}\right\|^2\leq\frac{D_1}{\sqrt{n}}\left(1+Z_{ii}^2+Z_{ij}^2+Z_{jj}^2\right),$$

where constants $C_3,D_1\in(0,\infty)$ and $Z'=h_1Z_1+h_2Z_2+h_3Z_3\sim\mathcal{N}\left(0,1\right)$. The last inequality follows from $\left\|\eta^{ora}\right\|^2\leq C_4\left(Z_{ii}^2+Z_{ij}^2+Z_{jj}^2\right)$ for some large constant C_4, which can be shown using (82) easily. ∎

A.1.2. *Proof of Theorem 3.* The triangle inequality gives

$$\begin{aligned} \left|\hat{\omega}_{ij}-\omega_{ij}\right| & \leq \left|\hat{\omega}_{ij}-\omega_{ij}^{ora}\right|+\left|\omega_{ij}^{ora}-\omega_{ij}\right|, \\ \left\|\hat{\Omega}_{A,A}-\Omega_{A,A}\right\|_{\infty} & \leq \left\|\hat{\Omega}_{A,A}-\Omega_{A,A}^{ora}\right\|_{\infty}+\left\|\Omega_{A,A}^{ora}-\Omega_{A,A}\right\|_{\infty}. \end{aligned}$$

From Equation (21) we have

$$\mathbb{P}\left\{\left\|\hat{\Omega}_{A,A}-\Omega_{A,A}^{ora}\right\|_{\infty}>C_1s\frac{\log p}{n}\right\}=o\left(p^{-\delta+1}\right).$$

Now we give a tail bound for $\left|\omega_{ij}^{ora}-\omega_{ij}\right|$ and $\left\|\Omega_{A,A}^{ora}-\Omega_{A,A}\right\|_{\infty}$ respectively. Let $\Phi(t)$ be the $\mathcal{N}\left(0,1\right)$ distribution function. For the constant $C>0$, we apply (23) to obtain

$$\mathbb{P}\left\{\left|\kappa_{ij}^{ora}\right|>C\right\}$$

$$\begin{aligned}&\le \mathbb{P}\left\{\max\{|Z_{kl}|\} > \vartheta\sqrt{n}\right\} + \bar{\Phi}\left(\frac{C}{2}\right) + \mathbb{P}\left\{\frac{D_1}{\sqrt{n}}\left(1+Z_{ii}^2+Z_{ij}^2+Z_{jj}^2\right) > \frac{C}{2}\right\}\\ &\le o(1) + 2\exp\left(-C^2/8\right),\end{aligned}$$

according to the inequality $\bar{\Phi}(x) \le 2\exp\left(-x^2/2\right)$ for $x > 0$ and the union bound of three Gaussian tail probabilities. This immediately implies that for large C_4 and large n,

$$\mathbb{P}\left\{\left|\omega_{ij}^{ora} - \omega_{ij}\right| > C_4\sqrt{\frac{1}{n}}\right\} \le \frac{3}{4}\epsilon,$$

which, together with (21), yields that for $C_2 > C_1 + C_4$,

$$\mathbb{P}\left\{|\hat{\omega}_{ij} - \omega_{ij}| > C_2 \max\left\{s\frac{\log p}{n}, \sqrt{\frac{1}{n}}\right\}\right\} \le \epsilon.$$

Similarly, Equation (23) implies

$$\begin{aligned}\mathbb{P}\left\{\left|\kappa_{ij}^{ora}\right| > C\sqrt{\log p}\right\} &\le \mathbb{P}\left\{\max\{|Z_{kl}|\} > \vartheta\sqrt{n}\right\} + \bar{\Phi}\left(\frac{C\sqrt{\log p}}{2}\right)\\ &\quad + \mathbb{P}\left\{\frac{D_1}{\sqrt{n}}\left(1+Z_{ii}^2+Z_{ij}^2+Z_{jj}^2\right) > \frac{C\sqrt{\log p}}{2}\right\}\\ &= O\left(p^{-C^2/8}\right),\end{aligned}$$

where the first and last components in the first inequality are negligible due to $\log p \le c_0 n$ with a sufficiently small $c_0 > 0$, which follows from the assumption $s \le c_0 n/\log p$. That immediately implies that for C_5 large enough,

$$\mathbb{P}\left\{\left\|\Omega_{A,A}^{ora} - \Omega_{A,A}\right\|_\infty > C_5\sqrt{\frac{\log p}{n}}\right\} = o(p^{-\delta}),$$

which, together with (21), yields that for $C_3 > C_1' + C_5$.

$$\mathbb{P}\left\{\left\|\hat{\Omega}_{A,A} - \Omega_{A,A}\right\|_\infty > C_3 \max\left\{s\frac{\log p}{n}, \sqrt{\frac{\log p}{n}}\right\}\right\} = o\left(p^{-\delta+1}\right).$$

Thus we have the following union bound over all $\binom{p}{2}$ pairs of (i, j),

$$\mathbb{P}\left\{\left\|\hat{\Omega} - \Omega\right\|_\infty > C_3 \max\left\{s\frac{\log p}{n}, \sqrt{\frac{\log p}{n}}\right\}\right\} = p^2/2 \cdot o\left(p^{-\delta+1}\right) = o\left(p^{-\delta+3}\right).$$

Write

$$\sqrt{n}\left(\hat{\Omega}_{A,A} - \Omega_{A,A}\right) = \sqrt{n}\left(\hat{\Omega}_{A,A} - \Omega_{A,A}^{ora}\right) + \sqrt{n}\left(\Omega_{A,A}^{ora} - \Omega_{A,A}\right).$$

Under the assumption $s = o\left(\frac{\sqrt{n}}{\log p}\right)$, noting that $\omega_{ii}\omega_{jj} + \omega_{ij}^2$ is bounded, we have

$$\sqrt{n}\left\|\hat{\Omega}_{A,A} - \Omega_{A,A}^{ora}\right\|_\infty = o_p(1),$$

which together with Equation (23) further implies

$$\sqrt{n/\left(\omega_{ii}\omega_{jj}+\omega_{ij}^2\right)}\left(\hat{\omega}_{ij}-\omega_{ij}\right)\overset{D}{\sim}\sqrt{n/\left(\omega_{ii}\omega_{jj}+\omega_{ij}^2\right)}\left(\omega_{ij}^{ora}-\omega_{ij}\right)\overset{D}{\to}\mathcal{N}\left(0,1\right).$$

As an immediate consequence, $\hat{F}_{ij}$ is a consistent estimator of F_{ij}, which is bounded above and below by some positive constants. Thus we obtain $\hat{F}_{ij}/F_{ij}\to 1$. ∎

A.2. Proof of Theorem 1. The probabilistic results (i) and (ii) as well as (3) are the immediate consequences of Theorems 2 and 5. We only need to show the minimax rate of convergence result (2). According to the probabilistic lower bound result (35) in Theorem 5, we immediately obtain that

$$\inf_{\hat{\omega}_{ij}}\sup_{\mathcal{G}_0(M,k_{n,p})}\mathbb{E}\left|\hat{\omega}_{ij}-\omega_{ij}\right|\geq c_1\max\left\{C_1\frac{k_{n,p}\log p}{n},C_2\sqrt{\frac{1}{n}}\right\}.$$

Thus it is enough to show there exists some estimator of ω_{ij} such that it attains this upper bound. More precisely, we have defined a truncated estimator based on the $\hat{\omega}_{ij}$ in (10) to control the small event in which $\hat{\Theta}_{A,A}$ is nearly singular:

$$\breve{\omega}_{ij}=sgn(\hat{\omega}_{ij})\cdot\min\left\{\left|\hat{\omega}_{ij}\right|,\log p\right\}.$$

Define the event $G=\left\{\left|\hat{\omega}_{ij}-\omega_{ij}^{ora}\right|\leq C_1\frac{k_{n,p}\log p}{n},\left|\omega_{ij}^{ora}\right|\leq 2M\right\}$. Note that the Equations (20) and (23) in Theorem 2 imply $\mathbb{P}\left\{G^c\right\}\leq C\left(p^{-\delta+1}+\exp\left(-cn\right)\right)$ for some constants C and c. Now according to the variance of inverse Wishart distribution, we pick $\delta\geq 2\xi+1$ to complete our proof as follows:

$$\begin{aligned}\mathbb{E}\left|\breve{\omega}_{ij}-\omega_{ij}\right| &\leq \mathbb{E}\left(\left|\breve{\omega}_{ij}-\omega_{ij}^{ora}\right|1\left\{G\right\}\right)+\mathbb{E}\left(\left|\breve{\omega}_{ij}-\omega_{ij}^{ora}\right|1\left\{G^c\right\}\right)+\mathbb{E}\left|\omega_{ij}^{ora}-\omega_{ij}\right|\\ &\leq C_1\frac{k_{n,p}\log p}{n}+\left(\mathbb{P}\left\{G^c\right\}\mathbb{E}\left(\log p+\left|\omega_{ij}^{ora}\right|\right)^2\right)^{1/2}+\left(\mathbb{E}\left(\omega_{ij}^{ora}-\omega_{ij}\right)^2\right)^{1/2}\\ &\leq C_1\frac{k_{n,p}\log p}{n}+C_2p^{-\frac{\delta+1}{2}}\log p+C_3\frac{1}{\sqrt{n}}\\ &\leq C'\max\left\{\frac{k_{n,p}\log p}{n},\sqrt{\frac{1}{n}}\right\},\end{aligned}$$

where C_2, C_3 and C' are some constants and the last equation follows from the assumption $n=O\left(p^{\xi}\right)$.

APPENDIX B: PROOF OF THEOREMS IN APPLICATIONS

B.1. Proof of Theorem 6. When $\delta>3$, from Theorem 2 it can be shown that the following three results hold:

(i) For any constant $\varepsilon > 0$, we have

$$\mathbb{P}\left\{\sup_{(i,j)}\left|\frac{\hat{\omega}_{ii}\hat{\omega}_{jj}+\hat{\omega}_{ij}^2}{\omega_{ii}\omega_{jj}+\omega_{ij}^2}-1\right|>\varepsilon\right\}\to 0; \tag{84}$$

(ii) There is a constant $C_1 > 0$ such that

$$\mathbb{P}\left\{\sup_{(i,j)}\left|\omega_{ij}^{ora}-\hat{\omega}_{ij}\right|>C_1 s\frac{\log p}{n}\right\}\to 0; \tag{85}$$

(iii) For any constant $2 < \xi_1$, we have

$$\mathbb{P}\left\{\sup_{(i,j)}\frac{\left|\omega_{ij}^{ora}-\omega_{ij}\right|}{\sqrt{\omega_{ii}\omega_{jj}+\omega_{ij}^2}}>\sqrt{\frac{2\xi_1\log p}{n}}\right\}\to 0. \tag{86}$$

In fact, under the assumption $\delta \geq 3$, Equation (21) in Theorem 2 and the union bound over all pair (i, j) imply the second result (85), which further shows the first result (84) because that $\hat{\omega}_{ij}$ and $\hat{\omega}_{ii}$ are consistent estimators and $\omega_{ii}\omega_{jj} + \omega_{ij}^2$ is bounded below and above. For the third result, we apply Equation (23) from Theorem 2 and pick $2 < \xi_2 < \xi_1$ and $a = \sqrt{\xi_1} - \sqrt{\xi_2}$ to show that

$$\begin{aligned}
\mathbb{P}\left\{\left|\kappa_{ij}^{ora}\right|>\sqrt{2\xi_1\log p}\right\} &\leq \mathbb{P}\left\{\max\left\{|Z_{kl}|\right\}>\vartheta\sqrt{n}\right\}+\bar{\Phi}\left(\sqrt{2\xi_2\log p}\right)\\
&\quad +\mathbb{P}\left\{\frac{D_1}{\sqrt{n}}\left(1+Z_{ii}^2+Z_{ij}^2+Z_{jj}^2\right)>a\sqrt{2\log p}\right\}\\
&= O(1)p^{-\xi_2}\sqrt{\frac{1}{\log p}},
\end{aligned}$$

where the last inequality follows from $\log p = o(n)$. The third result (86) is thus obtained by the union bound with $2 < \xi_2$.

As the proof of (41) and (42) are nearly identical to each other, we only prove that (42) in Theorem 6 is just a simple consequence of results (i), (ii) and (iii). Set $\varepsilon > 0$ sufficiently small and $\xi \in (2, \xi_0)$ sufficiently close to 2 such that $2\sqrt{2\xi_0} - \sqrt{2\xi_0(1+\varepsilon)} > \sqrt{2\xi}$ and

$\xi_0(1-\varepsilon) > \xi$, and $2 < \xi_1 < \xi$. We have

$$
\begin{aligned}
& \mathbb{P}\left(\mathcal{S}(\hat{\Omega}_{thr}) = \mathcal{S}(\Omega)\right) \\
= \; & \mathbb{P}\left(\hat{\omega}_{ij}^{thr} \neq 0 \text{ for all } (i,j) : \omega_{ij} \neq 0\right) + \mathbb{P}\left(\hat{\omega}_{ij}^{thr} = 0 \text{ for all } (i,j) : \omega_{ij} = 0\right) \\
= \; & \mathbb{P}\left\{ |\hat{\omega}_{ij}| > \sqrt{\frac{2\xi_0\left(\hat{\omega}_{ii}\hat{\omega}_{jj} + \hat{\omega}_{ij}^2\right)\log p}{n}} \text{ for all } (i,j) : \omega_{ij} \neq 0 \right\} \\
& + \mathbb{P}\left\{ |\hat{\omega}_{ij}| \leq \sqrt{\frac{2\xi_0\left(\hat{\omega}_{ii}\hat{\omega}_{jj} + \hat{\omega}_{ij}^2\right)\log p}{n}} \text{ for all } (i,j) : \omega_{ij} = 0 \right\} \\
\geq \; & \mathbb{P}\left\{ \sup_{(i,j)} \frac{|\hat{\omega}_{ij} - \omega_{ij}|}{\sqrt{\omega_{ii}\omega_{jj} + \omega_{ij}^2}} \leq \sqrt{\frac{2\xi\log p}{n}} \right\} - \mathbb{P}\left\{ \sup_{(i,j)} \left| \frac{\hat{\omega}_{ii}\hat{\omega}_{jj} + \hat{\omega}_{ij}^2}{\omega_{ii}\omega_{jj} + \omega_{ij}^2} - 1 \right| > \varepsilon \right\},
\end{aligned}
$$

which is bounded below by

$$
\mathbb{P}\left\{ \sup_{(i,j)} \frac{\left|\omega_{ij}^{ora} - \omega_{ij}\right|}{\sqrt{\omega_{ii}\omega_{jj} + \omega_{ij}^2}} \leq \sqrt{\frac{2\xi_1\log p}{n}} \right\} - \left[\begin{array}{c} \mathbb{P}\left\{ \sup_{(i,j)} \left|\omega_{ij}^{ora} - \hat{\omega}_{ij}\right| > C_1 s \frac{\log p}{n} \right\} + \\ \mathbb{P}\left\{ \sup_{(i,j)} \left| \frac{\hat{\omega}_{ii}\hat{\omega}_{jj} + \hat{\omega}_{ij}^2}{\omega_{ii}\omega_{jj} + \omega_{ij}^2} - 1 \right| > \varepsilon \right\} \end{array} \right] = 1 + o(1),
$$

where $s = o\left(\sqrt{n/\log p}\right)$ implies $s\frac{\log p}{n} = o\left(\sqrt{(\log p)/n}\right)$. ∎

B.2. Proof of Theorem 8. Due to the limit of space, we follow the line of the proof of Theorems 2 and 3, but only give necessary details when the proof is different. As we explained before the statement of the theorem, the coefficient vectors in regressing a pair of observed variables against other observed variables are not sparse enough in the latent variable graphical model for direct application of Theorems 2 and 3. Our strategy is to decompose the coefficients into two parts,

$$\beta_{O\backslash A,A} = S_{O\backslash A,A}\Omega_{A,A}^{-1} - L_{O\backslash A,A}\Omega_{A,A}^{-1} = \beta_{O\backslash A,A}^S - \beta^L, \tag{87}$$

with $\beta_{O\backslash A,A}^S = S_{O\backslash A,A}\Omega_{A,A}^{-1}$ and $\beta_{O\backslash A,A}^L = L_{O\backslash A,A}\Omega_{A,A}^{-1}$, and define a biased model

$$\mathbf{X}_A = \mathbf{X}_{O\backslash A}\beta_{O\backslash A,A}^S + \left(\epsilon_A - \mathbf{X}_{O\backslash A}\beta_{O\backslash A,A}^L\right) = \mathbf{X}_{O\backslash A}\beta_{O\backslash A,A}^S + \epsilon_A^S, \tag{88}$$

with $\epsilon_A^S = \epsilon_A - \mathbf{X}_{O\backslash A}\beta_{O\backslash A,A}^L$. We then define two oracle estimators of $\Theta_{A,A}$ as

$$\Theta_{A,A}^{ora} = \epsilon_A^T\epsilon_A/n, \quad \Theta_{A,A}^{ora,S} = \left(\epsilon_A^S\right)^T\left(\epsilon_A^S\right)/n. \tag{89}$$

For $m \in A$, we treat $\beta_{O\backslash A,m}^S$ as a target regression coefficient vector. As the ℓ_2 size of the bias is bounded by $\|\beta_{O\backslash A,m} - \beta_{O\backslash A,m}^S\| = \|\beta_{O\backslash A,m}^L\| \lesssim (a_n/n)\log p$ with $a_n \to 0$ by

(48), Theorem 10 (iii), Theorem 11 (ii) and Proposition 1 can be used to obtain (16), (17) and (18) with $\{A^c, \beta_{A^c,A}, \hat{\beta}_{A^c,A}\}$ replaced by $\{O \setminus A, \beta^S_{O\setminus A,A}, \hat{\beta}_{O\setminus A,A}\}$. Moreover, by (48) and the concentration inequality for χ^2_n, we have

$$\mathbb{P}\left\{\|\mathbf{X}_{O\setminus A}\beta^L_{O\setminus A,m}/n^{1/2}\| > C_1\lambda\right\} = o(p^{1-\delta}),$$

so that by the union bound

$$\begin{aligned} \left\|\mathbf{X}^T_{O\setminus A}\epsilon^S_A/n\right\|_\infty &\leq \left\|\mathbf{X}^T_{O\setminus A}\epsilon_A/n\right\|_\infty + \left\|\mathbf{X}^T_{O\setminus A}\mathbf{X}_{O\setminus A}\beta^L_{O\setminus A,A}/n\right\|_\infty \\ &\leq C_0\lambda + \left\|\mathbf{X}_{O\setminus A}\beta^L_{O\setminus A,A}/n^{1/2}\right\| \\ &\leq C_1\lambda \end{aligned}$$

happens with at least probability $1 - o(p^{1-\delta})$ as in the proof of Theorem 2 (i) and the proof of (74) in Proposition 1. Thus, as in the proof of Theorem 2 (i), we have

$$\mathbb{P}\left\{\left\|\hat{\Theta}_{A,A} - \Theta^{ora,S}_{A,A}\right\|_\infty > C_1 k_{n,p}\delta(\log p)/n\right\} = o(p^{1-\delta}).$$

Conditionally on $\mathbf{X}_{O\setminus A}\beta^L_{O\setminus A,m}$ with $m \in A$, $\epsilon^T_m\mathbf{X}_{O\setminus A}\beta^L_{O\setminus A,m}$ has the Gaussian distribution with mean 0 and variance $\theta_{mm}\|\mathbf{X}_{O\setminus A}\beta^L_{O\setminus A,m}\|^2$. It follows that

$$\mathbb{P}\left\{\left|\epsilon^T_m\mathbf{X}_{O\setminus A}\beta^L_{O\setminus A,m}/n\right| > C_1\sqrt{2\delta(\log p)/n}\lambda\right\} = o(p^{1-\delta}).$$

Consequently, due to $\epsilon_A - \mathbf{X}_{O\setminus A}\beta^L_{O\setminus A,A} = \epsilon^S_A$, we have

$$\mathbb{P}\left\{\left\|\Theta^{ora,S}_{A,A} - \Theta^{ora}_{A,A}\right\|_\infty > 3C_1\lambda^2\right\} = o\left(p^{-\delta+1}\right).$$

By triangle inequality, we further obtain

$$\mathbb{P}\left\{\left\|\hat{\Theta}_{A,A} - \Theta^{ora}_{A,A}\right\|_\infty > 3C_1\lambda^2 + C_1 k_{n,p}\delta(\log p)/n\right\} = o\left(p^{-\delta+1}\right).$$

Then following the proof of Theorem 3 exactly, we establish Theorem 8. ☐

APPENDIX C: PROOF OF RESULTS IN LINEAR REGRESSION

C.1. Proof of Theorem 10. (i) This part of the theorem is a direct consequence of Theorems 1 and 2 of Sun and Zhang (2012a). Specifically, we have $\mathbb{P}\left\{\|\widetilde{\mathbf{Z}}\|_\infty > \lambda^*\right\} \leq 2\widetilde{\epsilon}_1$ by Lemma 17 of Sun and Zhang (2013) for the correlation vector in (59).

(ii) We modify the proof as follows. Let $\lambda_{*,0} = L_{n-3/2}(k/\widetilde{p})$, $\varepsilon_2 \in [\varepsilon_1, \varepsilon]$ and

$$J = \{j : |\widetilde{Z}_j| > (1+\varepsilon_2)\lambda_{*,0}\} \cup K \tag{90}$$

with the set K in (60). Consider the Lasso estimator at an oracle penalty level $\sigma^{ora}\lambda$,

$$\hat{\gamma}(\lambda) = \arg\min_{\gamma}\left\{\frac{\|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\gamma\|^2}{2n} + \sigma^{ora}\lambda\|\gamma\|_1\right\},$$

with $\lambda > (1+\varepsilon_2)\lambda_{*,0}$. The Karush-Kuhn-Tucker conditions assert that

$$\widetilde{\mathbf{X}}_j^T\big(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\hat{\gamma}(\lambda)\big)/n \begin{cases} = \sigma^{ora}\lambda\,\mathrm{sgn}(\hat{\gamma}_j(\lambda)) & \hat{\gamma}_j \neq 0 \\ \in \sigma^{ora}\lambda[-1,1] & \forall j. \end{cases}$$

Let $\mathbf{h} = (\gamma^{target} - \hat{\gamma}(\lambda))/\sigma^{ora}$, $b = \lambda + \lambda^*$, $c = 2\lambda\sqrt{(2/n)\log\widetilde{p}}(s_1 - |K|)$ and $\xi_\lambda = b/(\lambda - (1+\varepsilon_2)\lambda_{*,0})$. Multiplying $\mathbf{h}$ to both sides of the KKT conditions yields

$$\|\widetilde{\mathbf{X}}\mathbf{h}\|^2/n \le (1+\varepsilon_2)\lambda_{*,0}\|\mathbf{h}_{J^c}\|_1 + \lambda^*\|\mathbf{h}_J\|_1 + \lambda\|\hat{\gamma}(\lambda)/\sigma^{ora}\|_1 - \lambda\|\gamma^{target}/\sigma^{ora}\|_1$$

when $\|\widetilde{\mathbf{Z}}\|_\infty \le \lambda^*$. Under Cond_1 in (60), $2\lambda\|\gamma^{target}_{J^c}/\sigma^{ora}\|_1 \le c$, so that

$$\|\widetilde{\mathbf{X}}\mathbf{h}\|^2/n + (b/\xi_\lambda)\|\mathbf{h}_{J^c}\|_1 \le c + b\|\mathbf{h}_J\|_1. \tag{91}$$

This matches inequality (A1) in Sun and Zhang (2012a) with $\mathbf{h} = \hat{\beta} - w$. As the proof of Theorems 1 and 2 of Sun and Zhang (2012a) is based on their (A1), their proof still yields (56), (57) and (58) with $s = s_1 + s_2$, when $\mathbb{P}\{|J \setminus K| \ge s_2\} \le \widetilde{\epsilon}_1$ and (61) holds with $\alpha \ge \sqrt{2}\xi_{\lambda_0}$. Let $\varepsilon_2 = \varepsilon_1$. The condition on α certainly holds as

$$\begin{aligned}\sqrt{2}\xi_{\lambda_0} &= \sqrt{2}\frac{(1+\varepsilon)L_{n-3/2}(k/\widetilde{p}) + L_{n-3/2}(\widetilde{\epsilon}_1/\widetilde{p})}{(1+\varepsilon)L_{n-3/2}(k/\widetilde{p}) - (1+\varepsilon_1)L_{n-3/2}(k/\widetilde{p})} \\ &= \frac{\sqrt{2}}{\varepsilon - \varepsilon_1}\Big(1 + \varepsilon + \frac{L_1(\widetilde{\epsilon}_1/\widetilde{p})}{L_1(k/\widetilde{p})}\Big).\end{aligned}$$

For the condition on $|J|$, Proposition 10 of Sun and Zhang (2013) with $m = s_2$ yields

$$\begin{aligned}\mathbb{P}\Big\{|J \setminus K| \ge s_2 \Big| \mathrm{Cond}_3\Big\} &\le \mathbb{P}\Big\{\max_{|J'|\le s_2}\sum_{j\in J'}(|\widetilde{Z}_j| - \lambda_{*,0})_+^2 \ge \varepsilon_1^2\lambda_{*,0}^2 s_2 \Big| \mathrm{Cond}_3\Big\} \\ &\le e^{1/(4n-6)^2}\widetilde{\epsilon}_1. \end{aligned} \tag{92}$$

(iii) For $\gamma^{target} \neq \gamma$, we need to change the proof of (ii) to bound $\widetilde{\mathbf{Z}}^{target}$, where

$$\widetilde{\mathbf{Z}}^{target} = \widetilde{\mathbf{X}}^T(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\gamma^{target})/(\sqrt{n}\|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\gamma^{target}\|).$$

More precisely, we need to bound $\|\widetilde{\mathbf{Z}}^{target}\|_\infty$ and the size of

$$J = \{j : |\widetilde{Z}_j^{target}| > (1+\varepsilon_2)\lambda_{*,0}\} \cup K.$$

Note that $\sigma^{ora} = \|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\gamma^{target}\|/\sqrt{n}$ here. When $\|\widetilde{\mathbf{Z}}\|_\infty \le \lambda^*$ and Cond$_4$ holds,

$$\begin{aligned}
\left\|\widetilde{\mathbf{Z}}^{target} - \widetilde{\mathbf{Z}}\right\|_\infty &\le \left|1 - \frac{\|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\gamma\|}{\|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\gamma^{target}\|}\right| \|\widetilde{\mathbf{Z}}\|_\infty + \frac{\|\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}(\gamma^{target} - \gamma)\|_\infty}{\sqrt{n}\|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\gamma^{target}\|} \\
&\le \|\widetilde{\mathbf{X}}(\gamma^{target} - \gamma)\|\{\sqrt{n}\sigma^{ora}\}^{-1}(\lambda^* + 1) \\
&\le (2/C_4)\sqrt{n^{-1}\log(\widetilde{p}/\widetilde{\epsilon}_1)} \\
&\le \min(\sqrt{2} - 1, \varepsilon_2 - \varepsilon_1)\lambda_{*,0}.
\end{aligned}$$

The last inequality above is a consequence of the condition on C_4 and the definition of $\lambda_{*,0}$. This leads to (91) with $b = \lambda + \sqrt{2}\lambda^*$ instead of $b = \lambda + \lambda^*$. However, we still have

$$\sqrt{2}\xi_{\lambda_0} = \frac{\sqrt{2}}{\varepsilon - \varepsilon_2}\Big(1 + \varepsilon + \frac{\sqrt{2}L_1(\widetilde{\epsilon}_1/\widetilde{p})}{L_1(k/\widetilde{p})}\Big) \le \alpha$$

with the modified α. For $|J'| \le s_2$, the bound $\|\widetilde{\mathbf{Z}}^{target} - \widetilde{\mathbf{Z}}\|_\infty \le (\varepsilon_2 - \varepsilon_1)\lambda_{*,0}$ gives

$$\sqrt{\sum_{j\in J'}(|\widetilde{Z}_j^{target}| - \lambda_{*,0})_+^2} \le \sqrt{\sum_{j\in J'}(|\widetilde{Z}_j| - \lambda_{*,0})_+^2} + (\varepsilon_2 - \varepsilon_1)\lambda_{*,0}\sqrt{s_2},$$

so that $\mathbb{P}\Big\{|J \setminus K| \ge s_2, \|\widetilde{\mathbf{Z}}\|_\infty \le \lambda^*, \text{Cond}_4 \Big| \text{ Cond}_3\Big\} \le e^{1/(4n-6)^2}\widetilde{\epsilon}_1$ by (92). This completes the proof. □

C.2. Proof of Theorem 11. Let $\hat{\mathbf{P}}$ be the orthogonal projection to the linear span of $\{\widetilde{\mathbf{X}}_k, k \in \hat{S}\}$. We have $\hat{\sigma}^2 - \big(\hat{\sigma}^{lse}\big)^2 = \big\|\hat{\mathbf{P}}\big(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\hat{\gamma}\big)\big\|^2/n = \big\|\widetilde{\mathbf{X}}\big(\hat{\gamma}^{lse} - \hat{\gamma}\big)\big\|^2/n$, which implies the identity in (66). Moreover, the KKT conditions for the lasso give

$$\widetilde{\mathbf{X}}_k^T\widetilde{\mathbf{X}}\big(\hat{\gamma}^{lse} - \hat{\gamma}\big)/n = \widetilde{\mathbf{X}}_k^T\big(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\hat{\gamma}\big)/n = \hat{\sigma}\lambda_0 \operatorname{sgn}(\hat{\gamma}_k)$$

for all $k \in \hat{S}$. Consequently, we have

$$\phi^2_{\text{comp}}\Big(0, \hat{S}, \widetilde{\mathbf{X}}\Big)\, \|\hat{\gamma}^{lse} - \hat{\gamma}\|_1^2/|\hat{S}| \le \big\|\widetilde{\mathbf{X}}\big(\hat{\gamma}^{lse} - \hat{\gamma}\big)\big\|^2/n \le \hat{\sigma}\lambda_0\|\hat{\gamma}^{lse} - \hat{\gamma}\|_1,$$

which implies the inequalities in (66) and (67).

For $C_0 s\delta(\log\widetilde{p})/n \le (\varepsilon - \varepsilon_3)/(1 + \varepsilon)$, the oracle inequality in (58) give

$$\frac{\hat{\sigma}\lambda_0}{\sigma^{ora}} \ge (1 - C_0 s\delta(\log\widetilde{p})/n)(1 + \varepsilon)\lambda_{*,0} \ge (1 + \varepsilon_3)\lambda_{*,0}.$$

Let J be as in (90) and $K' \subseteq \hat{S} \setminus J$. For $k \in K'$, the KKT conditions guarantee

$$\left|\frac{\widetilde{\mathbf{X}}_k^T(\widetilde{\mathbf{X}}\gamma^{target} - \widetilde{\mathbf{X}}\hat{\gamma})}{n\sigma^{ora}}\right| = \left|\frac{\widetilde{\mathbf{X}}_k^T(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\hat{\gamma})}{n\sigma^{ora}} - \widetilde{Z}_k\right| \ge \frac{\hat{\sigma}\lambda_0}{\sigma^{ora}} - (1 + \varepsilon_2)\lambda_{*,0} > (\varepsilon_3 - \varepsilon_2)\lambda_{*,0}.$$

Thus, for $|K'| \le s_3$, Cond_3 in (62) and the oracle inequality in (56) imply that

$$|K'|(\varepsilon_3 - \varepsilon_2)^2 \lambda_{*,0}^2 < \sum_{k \in K'} \left| \frac{\widetilde{\mathbf{X}}_k^T (\widetilde{\mathbf{X}} \gamma^{target} - \widetilde{\mathbf{X}} \hat{\gamma})}{n \sigma^{ora}} \right|^2 \le C_3 C_0 s \delta (\log \widetilde{p})/n.$$

As $\lambda_{*,0} = L_{n-3/2}(k/\widetilde{p}) = (n-3/2)^{-1/2} L_1(k/\widetilde{p})$, it follows that

$$|K'| < \frac{C_3 C_0 s \delta (\log \widetilde{p})/n}{(\varepsilon_3 - \varepsilon_2)^2 \lambda_{*,0}^2} \le \frac{C_3 C_0 s \delta (\log \widetilde{p})}{(\varepsilon_3 - \varepsilon_2)^2 L_1^2(k/\widetilde{p})} \le s_3$$

This proves $|K'| < s_3$ for all $K' \subseteq \hat{S} \setminus J$ satisfying $|K'| \le s_3$, so that $|\hat{S} \setminus J| \le s_3$. Consequently, (68) follows from the bound $|J \setminus K| \le s_2$ in the proof of Theorem 10, as $s_2 + |K| \le s$. If in addition (69) holds, then the conclusions of Theorem 10 hold for $\{\hat{\gamma}^{lse}, \hat{\sigma}^{lse}\}$ by (66), (67), (68) and (58). □

C.3. Proof of Proposition 1. We need the following tail bound for the chi-squared distribution with n degrees of freedom,

$$\mathbb{P}\left\{ \left| \frac{\chi_{(n)}^2}{n} - 1 \right| \ge t \right\} \le 2 \exp\left(-nt\left(t \wedge 1\right)/8\right), \ \forall t > 0. \tag{93}$$

As $\mathrm{diag}(\mathbf{\Sigma})^{-1}\overline{\mathbf{D}}$ has $\chi_{(n)}^2/n$ diagonal elements, (93) directly implies (71). Similarly, as

$$\frac{\|\widetilde{\mathbf{X}}(\gamma^{target} - \gamma)\|^2}{(\gamma^{target} - \gamma)^T \mathbf{\Sigma} (\gamma^{target} - \gamma)} \sim \chi_{(n)}^2, \ \frac{n(\sigma^{ora})^2}{\mathbb{E}(\sigma^{ora})^2} \sim \chi_{(n)}^2,$$

(93) also implies (74) and justifies the replacement of σ^{ora} by $\sqrt{\mathbb{E}(\sigma^{ora})^2}$ or C_* in (56) and (57). It remains to prove (72) and (73).

It is well-known that for fixed α, $\delta > 1$ and sufficiently small $c_0 > 0$, the compatibility constant $\phi_{\mathrm{comp}}(\alpha, J, \mathbf{X})$ is no smaller than a positive constant with high probability $1 - o(p^{-\delta})$ under the assumption $|J| + \delta \le c_0 n / \log p$ for the Gaussian design $\mathbf{X}$ under the specified condition. For a complete proof, please refer to Corollary 1 in Raskutti, Wainwright and Yu (2010), where the conclusion holds for the restricted eigenvalue, a lower bound of the compatibility constant by its definition. See also Theorem 6 in Rudelson and Zhou (2013) for an extension to design matrices with sub-Gaussian marginals. For standardized sub-design matrix $\widetilde{\mathbf{X}}$, we just need to adjust the dimension from p to $\widetilde{p}$ and apply (93) to address the effect of standardization of design vectors. Thus, (72) holds.

The proof of (73) is simpler as the concentration inequality for the largest singular value of the standard Gaussian matrix can be directly applied. See for example Theorem II.13 of Davidson and Szarek (2001) and Proposition 2 of Zhang and Huang (2008). □

APPENDIX D: PROOF OF A LEMMA

D.1. Proof of Lemma 2. Now we establish the lower bound (81) for the total variation affinity. Since the affinity $\int q_0 \wedge q_1 d\mu = 1 - \frac{1}{2}\int |q_0 - q_1|\, d\mu$ for any two densities q_0 and q_1, Jensen's Inequality implies

$$\left[\int |q_0 - q_1|\, d\mu\right]^2 = \left(\int \left|\frac{q_0 - q_1}{q_0}\right| q_0 d\mu\right)^2 \le \int \frac{(q_0 - q_1)^2}{q_0} d\mu = \int \frac{q_1^2}{q_0} d\mu - 1.$$

Hence $\int q_0 \wedge q_1 d\mu \ge 1 - \frac{1}{2}\left(\int \frac{q_1^2}{q_0} d\mu - 1\right)^{1/2}$. To establish (81), it thus suffices to show that

$$\Delta = \int \frac{\left(\frac{1}{m_*}\sum_{m=1}^{m_*} f_m\right)^2}{f_0} - 1 = \frac{1}{m_*^2}\sum_{m,l}\int\left(\frac{f_m f_l}{f_0} - 1\right) \to 0.$$

The following lemma is used to calculate the term $\int (f_m f_l / f_0 - 1)$ in Δ. Let g_s be the density function of $\mathcal{N}(0, \Sigma_s)$, $s = 0, m$ or l. Then

$$\int \frac{g_m g_l}{g_0} = \left[\det\left(I - \Sigma_0^{-1}(\Sigma_m - \Sigma_0)\Sigma_0^{-1}(\Sigma_l - \Sigma_0)\right)\right]^{-1/2}. \tag{94}$$

Let $\Sigma_m = \Omega_m^{-1}$ for $0 \le m \le m_*$. It follows from (94) that

$$\int \frac{f_m f_l}{f_0} = \left(\int \frac{g_m g_l}{g_0}\right)^n = \left[\det\left(I - \Omega_0(\Sigma_m - \Sigma_0)\Omega_0(\Sigma_l - \Sigma_0)\right)\right]^{-n/2}.$$

Let $J(m,l)$ be the number of overlapping a between Σ_m and Σ_l in the first row. Recall the simple structures of Ω_0 (76) and $\Sigma_m - \Sigma_0$ by our construction. Elementary calculations yield that

$$\det\left(I - \Omega_0(\Sigma_m - \Sigma_0)\Omega_0(\Sigma_l - \Sigma_0)\right) = \left(1 - \frac{1+b^2}{(1-b^2)^2} J a^2\right)^2,$$

which is 1 when $J = 0$. Now we set $d = \frac{1+b^2}{(1-b^2)^2} > 1$ to simplify our notation. It is easy to see that the total number of pairs (Σ_m, Σ_l) such that $J(m,l) = j$ is $\binom{p-2}{k_{n,p}-2}\binom{k_{n,p}-2}{j}\binom{p-k_{n,p}}{k_{n,p}-2-j}$. Hence,

$$\begin{aligned}
\Delta &= \frac{1}{m_*^2}\sum_{0\le j\le k_{n,p}-2}\ \sum_{J(m,l)=j}\int\left(\frac{f_m f_l}{f_0} - 1\right) \\
&= \frac{1}{m_*^2}\sum_{0\le j\le k_{n,p}-2}\ \sum_{J(m,l)=j}\left((1 - dja^2)^{-n} - 1\right) \\
&\le \frac{1}{m_*^2}\sum_{1\le j\le k_{n,p}-2}\binom{p-2}{k_{n,p}-2}\binom{k_{n,p}-2}{j}\binom{p-k_{n,p}}{k_{n,p}-2-j}(1 - dja^2)^{-n}. \qquad (95)
\end{aligned}$$

Note that

$$(1 - dja^2)^{-n} \le (1 + 2dja^2)^n \le \exp\left(n2dja^2\right) = p^{2d\tau_1 j}$$

where the first inequality follows from the fact that $dja^2 \le dk_{n,p}a^2 \le \frac{1+b^2}{(1-b^2)^2}\tau_1 C_0 < 1/2$. Hence,

$$\begin{aligned}
\Delta &\le \sum_{1\le j\le k_{n,p}-2} \frac{\binom{k_{n,p}-2}{j}\binom{p-k_{n,p}}{k_{n,p}-2-j}}{\binom{p-2}{k_{n,p}-2}} p^{2d\tau_1 j} \\
&= \sum_{1\le j\le k_{n,p}-2} \frac{1}{j!} \frac{\left(\frac{(k_{n,p}-2)!}{(k_{n,p}-2-j)!}\right)^2}{\frac{(p-2)!(p-2k_{n,p}+2+j)!}{[(p-k_{n,p})!]^2}} p^{2d\tau_1 j} \\
&\le \sum_{1\le j\le k_{n,p}-2} \left(\frac{k_{n,p}^2 p^{2d\tau_1}}{p-k_{n,p}}\right)^j,
\end{aligned}$$

where the last inequality follows from the facts that $\frac{(k_{n,p}-2)!}{(k_{n,p}-2-j)!}$ is a product of j terms with each term less than $k_{n,p}$ and $\frac{(p-2)!(p-2k_{n,p}+2+j)!}{[(p-k_{n,p})!]^2}$ is bounded below by a product of j terms with each term greater than $(p-k_{n,p})$. Recall the assumption (32) $p \ge k_{n,p}^v$. So for large enough p, we have $p - k_{n,p} \ge p/2$ and

$$\begin{aligned}
k_{n,p}^2 \frac{p^{2d\tau_1}}{p-k_{n,p}} &\le 2p^{2/\nu} \cdot \frac{p^{2d\tau_1}}{p} \\
&\le 2p^{-(v-2)/(2v)},
\end{aligned}$$

where the last step follows from the fact that $\tau_1 \le (\nu-2)/(4\nu d)$. Thus

$$\Delta \le 2 \sum_{1\le j\le k_{n,p}-2} p^{-j(v-2)/(2v)} \to 0,$$

which immediately implies (81). ∎

REFERENCES

DAVIDSON, K. and SZAREK, S. (2001). Local operator theory, random matrices and Banach spaces. In *Handbook on the Geometry of Banach Spaces*, **1**.

EINMAHL, U. (1989). Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *Journal of multivariate analysis* **28** 20–68.

MASON, D. and ZHOU, H. H. (2012). Quantile Coupling Inequalities and Their Applications. *Probability Surveys* **9** 439-479.

RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2010). Restricted eigenvalue properties for correlated Gaussian designs. *The Journal of Machine Learning Research* **11** 2241–2259.

RUDELSON, M. and ZHOU, S. (2013). Reconstruction From Anisotropic Random Measurements. *Information Theory, IEEE Transactions on* **59** 3434-3447.

SUN, T. and ZHANG, C.-H. (2012). Scaled Sparse Linear Regression. *Biometrika* **99** 879–898.

SUN, T. and ZHANG, C.-H. (2013). Sparse matrix inversion with scaled lasso. *The Journal of Machine Learning Research* **14** 3385–3418.

ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the Lasso selection in high-dimensional linear regression. *Annals of Statistics* **36** 1567–1594.

DEPARTMENT OF STATISTICS
UNIVERSITY OF PITTSBURGH
PITTSBURGH, PENNSYLVANIA 15260
USA
E-MAIL: zren@pitt.edu

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742
USA
E-MAIL: tingni@umd.edu

DEPARTMENT OF STATISTICS AND BIOSTATISTICS
HILL CENTER, BUSCH CAMPUS
RUTGERS UNIVERSITY
PISCATAWAY, NEW JERSEY 08854
USA
E-MAIL: cunhui@stat.rutgers.edu

DEPARTMENT OF STATISTICS
YALE UNIVERSITY
NEW HAVEN, CONNECTICUT 06511
USA
E-MAIL: huibin.zhou@yale.edu