Submitted to the Annals of Statistics

SUPPLEMENT TO "OPTIMAL RATES OF CONVERGENCE FOR SPARSE COVARIANCE MATRIX ESTIMATION"

BY T. TONY CAI † AND HARRISON H. ZHOU†

University of Pennsylvania and Yale University

In this supplement we prove the additional technical lemmas used in the proof of Lemma 6.

1. Proof of Lemma 8 (ii). Jensen's inequality yields that for any two densities q_0 and q_1 with respect to a common dominating measure μ,

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

Equation (40) implies

$$\tilde{\mathbb{E}}_{(\gamma_{-1},\lambda_{-1})} \left\{ TV^2 \left(\bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})},\ \bar{\mathbb{P}}_{(1,1,\gamma_{-1},\lambda_{-1})} \right) \right\} \leqslant c_2^2,$$

where $TV(\mathbb{P}, \mathbb{Q})$ denotes the total variation distance between two distributions $\mathbb{P}$ and $\mathbb{Q}$, which then yields

$$\tilde{\mathbb{E}}_{(\gamma_{-1},\lambda_{-1})} \left\{ TV \left(\bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})},\ \bar{\mathbb{P}}_{(1,1,\gamma_{-1},\lambda_{-1})} \right) \right\} \leqslant c_2, \tag{58}$$

due to the simple fact $(\text{Average}\{a_i\})^2 \leqslant \text{Average}\{a_i^2\}$. Note that the total variation affinity $\|\mathbb{P} \wedge \mathbb{Q}\| = 1 - TV(\mathbb{P}, \mathbb{Q})$ for any two probability distributions $\mathbb{P}$ and $\mathbb{Q}$. Equation (58) immediately implies

$$\tilde{\mathbb{E}}_{(\gamma_{-1},\lambda_{-1})} \left\{ \left\| \bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})} \wedge \bar{\mathbb{P}}_{(1,1,\gamma_{-1},\lambda_{-1})} \right\| \right\} \geqslant 1 - c_2 > 0.$$

Thus we have

$$\left\| \bar{\mathbb{P}}_{1,0} \wedge \bar{\mathbb{P}}_{1,1} \right\| \geqslant \tilde{\mathbb{E}}_{(\gamma_{-1},\lambda_{-1})} \left\{ \left\| \bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})} \wedge \bar{\mathbb{P}}_{(1,1,\gamma_{-1},\lambda_{-1})} \right\| \right\} \geqslant 1 - c_2 > 0$$

following from Lemma 4. ∎

*The research of Tony Cai was supported in part by NSF FRG Grant DMS-0854973.

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

2. Proof of Lemma 10. Write

$$
\begin{aligned}
\Sigma_1 - \Sigma_0 &= \begin{pmatrix} 0 & \mathbf{v}_{1\times(p-1)} \\ \left(\mathbf{v}_{1\times(p-1)}\right)^T & \mathbf{0}_{(p-1)\times(p-1)} \end{pmatrix} \\
\Sigma_2 - \Sigma_0 &= \begin{pmatrix} 0 & \mathbf{v}^*_{1\times(p-1)} \\ \left(\mathbf{v}^*_{1\times(p-1)}\right)^T & \mathbf{0}_{(p-1)\times(p-1)} \end{pmatrix}
\end{aligned}
$$

where $\mathbf{v}_{1\times(p-1)} = (v_j)_{2\leqslant j\leqslant p}$ satisfies $v_j = 0$ for $2 \leqslant j \leqslant p - r$ and $v_j = 0$ or 1 for $p - r + 1 \leqslant j \leqslant p$ with $\|\mathbf{v}\|_0 = k$, and $\mathbf{v}^*_{1\times(p-1)} = \left(v^*_j\right)_{2\leqslant j\leqslant p}$ satisfies a similar property. Without loss of generality we consider only a special case with

$$
\begin{aligned}
v_j &= \begin{cases} 1, & p - r + 1 \leqslant j \leqslant p - r + k \\ 0, & \text{otherwise} \end{cases} \\
v^*_j &= \begin{cases} 1, & p - r + k - J \leqslant j \leqslant p - r + 2k - J \\ 0, & \text{otherwise} \end{cases} .
\end{aligned}
$$

It is easy to see that

$$
q_{ij} = \begin{cases} J\epsilon^2_{n,p}, & i = j = 1 \\ \epsilon^2_{n,p}, & p - r + 1 \leqslant i \leqslant p - r + k, \text{ and } p - r + k - J \leqslant j - 1 \leqslant p - r + 2k - J \\ 0, & \text{otherwise} \end{cases} .
$$

It is clear that the rank of $(\Sigma_1 - \Sigma_0)(\Sigma_2 - \Sigma_0)$ is 2. A straightforward calculation shows that the characteristic polynomial

$$
\det\left[\lambda I_{p\times p} - (\Sigma_1 - \Sigma_0)(\Sigma_2 - \Sigma_0)\right] = \left(\lambda - J\epsilon^2_{n,p}\right)^2 \lambda^{p-2}
$$

which implies $(\Sigma_1 - \Sigma_0)(\Sigma_2 - \Sigma_0)$ has two identical nonzero eigenvalues $J\epsilon^2_{n,p}$.

Note that this special case corresponds to

$$
I_r = \{j : p - r + 1 \leqslant j \leqslant p - r + k\}
$$

and

$$
I_c = \{j : p - r + k - J \leqslant j \leqslant p - r + 2k - J\} .
$$

Hence, $I_r \cap I_c = \{j : p - r + k - J \leqslant j \leqslant p - r + k\}$ with $\mathrm{Card}(I_r \cap I_c) = J$. The general case can be reduced to the special case by matrix permutations. ∎

3. Proof of Lemma 11. Let

$$A = \left[I - (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)\right]^{-1} \left(\Sigma_0^{-2} - I\right)(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2), \tag{59}$$

and

$$R_{1,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}} = -\log\det(I - A).$$

Note that

$$\begin{aligned} R_{\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}} &= -\log\det\left[I - (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2) - \left(\Sigma_0^{-2} - I\right)(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)\right] \\ &= -\log\det\{[I - A] \cdot [I - (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)]\} \\ &= -\log\det\left[I - (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)\right] - \log\det(I - A) \\ &= -2\log\left(1 - J\epsilon_{n,p}^2\right) + R_{1,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}} \qquad (60) \end{aligned}$$

where the last equation follows from Lemma 10.

Now we establish Equation (46). It is important to observe that $\operatorname{rank}(A) \leqslant 2$ due to the simple structure of $(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)$. Let ϱ be an eigenvalue of A. It is easy to see that

$$\begin{aligned} |\varrho| &\leqslant \|A\| \leqslant \left\|\Sigma_0^{-2} - I\right\| \|\Sigma_0 - \Sigma_1\| \|\Sigma_0 - \Sigma_2\| / \left(1 - \|\Sigma_0 - \Sigma_1\| \|\Sigma_0 - \Sigma_2\|\right) \\ &\leqslant \left(\left(\frac{3}{2}\right)^2 - 1\right) \frac{1}{3} \cdot \frac{1}{3} / \left(1 - \frac{1}{3} \cdot \frac{1}{3}\right) = \frac{5}{32} < \frac{1}{6} \qquad (61) \end{aligned}$$

since $\|\Sigma_1 - \Sigma_0\| \leqslant \|\Sigma_1 - \Sigma_0\|_1 = 2k\epsilon_{n,p} < 1/3$ and $\lambda_{\min}(\Sigma_0) \geqslant 1 - \|I - \Sigma_0\| \geqslant 1 - \|I - \Sigma_0\|_1 > 2/3$ from Equation (22). Note that

$$|\log(1 - x)| \leqslant 2|x|, \text{ for } |x| < \frac{1}{6},$$

which implies

$$R_{1,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}} \leqslant 4\|A\|,$$

i.e.,

$$\exp\left(\frac{n}{2} R_{1,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}}\right) \leqslant \exp(2n\|A\|).$$

Note that

$$\|I - \Sigma_0\| \leqslant \|I - \Sigma_0\|_1 = 2k\epsilon_{n,p} < \frac{1}{3} < 1$$

and

$$\|(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)\| \leqslant \frac{1}{3} \cdot \frac{1}{3} < 1.$$

We can write

$$\begin{aligned}\Sigma_0^{-2} - I &= (I - (I - \Sigma_0))^{-2} - I = \left(I + \sum_{k=1}^{\infty} (I - \Sigma_0)^k \right)^2 - I \\ &= \left[\sum_{m=0}^{\infty} (m+2)(I - \Sigma_0)^m \right] (I - \Sigma_0) \qquad (62)\end{aligned}$$

where

$$\left\| \sum_{m=0}^{\infty} (m+2)(I - \Sigma_0)^m \right\| \leqslant \sum_{m=0}^{\infty} (m+2) \left(\frac{1}{3} \right)^m < 3.$$

Define

$$A_* = (I - \Sigma_0)(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2) \qquad (63)$$

then

$$\begin{aligned}\|A\| &\leqslant \left\| [I - (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)]^{-1} \right\| \left\| \sum_{m=0}^{\infty} (m+2)(I - \Sigma_0)^m \right\| \|A_*\| \\ &< 3 \cdot \frac{1}{1 - \frac{1}{3} \cdot \frac{1}{3}} \cdot \|A_*\| = \frac{27}{8} \|A_*\| \leqslant \frac{27}{8} \max\{\|A_*\|_1, \|A_*\|_\infty\}\end{aligned}$$

from Equations (59) and (62). It is then sufficient to show

$$\tilde{\mathbb{E}}_{(\lambda_1,\lambda_1')|J} \left[\tilde{\mathbb{E}}_{(\gamma_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')} \exp\left(\frac{27}{2} n \max\{\|A_*\|_1, \|A_*\|_\infty\} \right) \right] \leqslant \frac{3}{2}, \qquad (64)$$

where $\|A_*\|$ depends on the values of λ_1, λ_1' and $(\gamma_{-1}, \lambda_{-1})$. We dropped the indices λ_1, λ_1' and $(\gamma_{-1}, \lambda_{-1})$ from A to simplify the notations.

Let $A_* = \left(a_{ij}^* \right)_{1 \leqslant i,j \leqslant 1}$. Then $\|A_*\|_1 = \max_{1 \leqslant m \leqslant p} \sum_j \left| a_{mj}^* \right|$ and $\|A_*\|_\infty = \max_{1 \leqslant m \leqslant p} \sum_i |a_{im}^*|$. We will show that for every non-negative integer t and every absolute row sum we have

$$\tilde{\mathbb{P}} \left(\sum_j \left| a_{mj}^* \right| \geqslant 2t \cdot \epsilon_{n,p} \cdot k\epsilon_{n,p}^2 \right) \leqslant \left(\frac{k^2}{p/8 - 1 - k} \right)^t$$

and the same tail bound holds for every absolute column sum, which immediately implies

$$\tilde{\mathbb{P}} \left(\max\{\|A_*\|_1, \|A_*\|_\infty\} \geqslant 2t \cdot \epsilon_{n,p} \cdot k\epsilon_{n,p}^2 \right) \leqslant 2p \left(\frac{k^2}{p/8 - 1 - k} \right)^t.$$

For each row m, define $E_m = \{1, 2, \ldots, r\} \setminus \{1, m\}$. Note that for each column of λ_{E_m}, if the column sum of λ_{E_m} is less than or equal to $2k-2$, then the other two rows can still freely take values 0 or 1 in this column, because the total sum will still not exceed $2k$. Let $n_{\lambda_{E_m}}$ be the number of columns of λ_{E_m} with column sum at least $2k-1$, and define $p_{\lambda_{E_m}} = r - n_{\lambda_{E_m}}$. Without loss of generality we assume that $k \geqslant 3$. Since $n_{\lambda_{E_m}} \cdot (2k-2) \leqslant r \cdot k$, the total number of 1's in the upper triangular matrix by the construction of the parameter set, we thus have $n_{\lambda_{E_m}} \leqslant r \cdot \frac{3}{4}$, which immediately implies $p_{\lambda_{E_m}} = r - n_{\lambda_{E_m}} \geqslant \frac{r}{4} \geqslant \frac{p}{8} - 1$. Recall that the distribution of $(\gamma_{-1}, \lambda_{-1})$ given (λ_1, λ_1') is uniform over $\Theta_{-1}(\lambda_1, \lambda_1')$. Thus from Lemma 10 we have

$$\tilde{\mathbb{P}}\left(\sum_j |a_{mj}^*| \geqslant 2t \cdot \epsilon_{n,p} \cdot k\epsilon_{n,p}^2 | \lambda_{E_m}\right) \leqslant \frac{\binom{k}{t}\binom{p_{\lambda_{E_m}}}{k-t}}{\binom{p_{\lambda_{E_m}}}{k}} \leqslant \left(\frac{k^2}{p/8-1-k}\right)^t$$

for every non-negative integer t as shown in Equation (47), which immediately implies

$$\tilde{\mathbb{P}}\left(\sum_j |a_{mj}^*| \geqslant 2t \cdot \epsilon_{n,p} \cdot k\epsilon_{n,p}^2\right) \leqslant \left(\frac{k^2}{p/8-1-k}\right)^{t-1} \quad \text{for every } t > 2.$$

For any random variable $X \geqslant 0$ and constant $a \geqslant 0$ it is known that

$$\begin{aligned} EX &= \int_{x\geqslant 0} P(X > x)\,dx = \int_{x\leqslant a} P(X > x)\,dx + \int_{x>a} P(X > x)\,dx \\ &\leqslant a + \int_{x>a} P(X > x)\,dx. \end{aligned}$$

Setting $X = \exp\left(\frac{27}{2} n \max\{\|A_*\|_1, \|A_*\|_\infty\}\right)$ and $a = \exp\left(27n \cdot \frac{2\beta}{\beta-1} \cdot \epsilon_{n,p} \cdot k\epsilon_{n,p}^2\right)$, where $\beta > 1$ and $p > n^\beta$ as defined in Section 1, we have

$$\begin{aligned} &\tilde{\mathbb{E}}_{(\lambda_1,\lambda_1')|J}\left[\tilde{\mathbb{E}}_{(\gamma_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')} \exp\left(\frac{27}{2} n \max\{\|A_*\|_1, \|A_*\|_\infty\}\right)\right] \\ \leqslant\ & \exp\left(27n \cdot \frac{2\beta}{\beta-1} \cdot \epsilon_{n,p} \cdot k\epsilon_{n,p}^2\right) \\ &+ \int_{t>\frac{2\beta}{\beta-1}} 27nk\epsilon_{n,p}^3 \cdot \exp\left(\frac{27}{2} n \cdot 2t \cdot \epsilon_{n,p} \cdot k\epsilon_{n,p}^2\right) 2p \left(\frac{k^2}{p/8-1-k}\right)^{t-1} dt \\ \leqslant\ & \exp\left(\frac{54\beta}{\beta-1} \cdot c_{n,p}\epsilon_{n,p}^{3-q}\right) \\ &+ \int_{t>\frac{2\beta}{\beta-1}} \exp\left[\log(2p) - (t-1)\log\frac{p/8-1-k}{k^2} + 27n(t+1)k\epsilon_{n,p}^3\right] dt. \qquad (65) \end{aligned}$$

Note that $\epsilon_{n,p} = \upsilon\sqrt{\frac{\log p}{n}}$, $k = \left\lceil \frac{1}{2} c_{n,p}\epsilon_{n,p}^{-q} \right\rceil - 1$, $\upsilon^2 < \frac{\beta-1}{54\beta}$ and $M\upsilon^{1-q} < \frac{1}{3}$ as defined in Section 3, and $c_{n,p} \leqslant Mn^{\frac{1-q}{2}} (\log p)^{-\frac{3-q}{2}}$ from Equation (3). These facts imply

$$\exp\left(\frac{54\beta}{\beta-1} \cdot c_{n,p}\epsilon_{n,p}^{3-q}\right) \leqslant \exp\left(\frac{54\beta}{\beta-1}\upsilon^2 \cdot M\upsilon^{1-q}\right) \leqslant e^{\frac{1}{3}} < \frac{3}{2}, \tag{66}$$

and also

$$\begin{aligned} 27nk\epsilon_{n,p}^3 &\leqslant 27M\upsilon^{3-q} \\ \left(1+\frac{1}{\beta}\right)\log p &= \frac{\beta+1}{\beta-1} \cdot \left(1-\frac{1}{\beta}\right)\log p \leqslant \left(\frac{2\beta}{\beta-1} - 1\right)\log\frac{p/8-1-k}{k^2}. \end{aligned} \tag{67}$$

The last two equations yield

$$\int_{t>\frac{2\beta}{\beta-1}} \exp\left[\log(2p) - (t-1)\log\frac{p/8-1-k}{k^2} + 27n(t+1)k\epsilon_{n,p}^3\right] dt = o(1)$$

since $\log p \to \infty$. Then Equation (65) is bounded from above by $\frac{3}{2}$, which immediately implies (64) and thus establishes Lemma 11. ∎

Remark 1 Under the assumption $c_{n,p} \leqslant Mn^{\frac{1-q}{2}} (\log p)^{-\frac{3-q}{2}}$ we obtain a finite upper bound in Equations (66) and (67). It is not clear to us if this assumption can be weakened to $c_{n,p} \leqslant Mn^{\frac{1-q}{2}} (\log p)^{-\frac{1-q}{2}}$.

Department of Statistics
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104.
E-mail: tcai@wharton.upenn.edu
URL: http://www-stat.wharton.upenn.edu/ tcai

Department of Statistics
Yale University
New Haven, CT 06511.
E-mail: huibin.zhou@yale.edu
URL: http://www.stat.yale.edu/ hz68