# SUPPLEMENTARY MATERIAL FOR "NETWORK EXPLORATION VIA THE ADAPTIVE LASSO AND SCAD PENALTIES"

By Jianqing Fan, Yang Feng and Yichao Wu

## APPENDIX B: PROOF OF THEOREM 5.3

Proof. First of all, to simplify our notation, we write $\mathbf{\Omega}$ as a vector in the following way: divide the indexes of $\mathbf{\Omega}_0 = \{(\omega_{0ij}), i, j = 1, \cdots, p\}$ to two parts: $\mathcal{A} = \{(i,j), \omega_{0ij} \neq 0 \ \& \ i \leq j\}$ and $\mathcal{B} = \{(i,j), \omega_{0ij} = 0 \ \& \ i \leq j\}$. Denoting $\mathbf{\Omega}$ in a vector format, we write $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, where $\boldsymbol{\beta}_1 = (\omega_{ij}, (i,j) \in \mathcal{A})$ and $\boldsymbol{\beta}_2 = (\omega_{ij}, (i,j) \in \mathcal{B})$. As a result, $\boldsymbol{\beta}$ has the length of $d = p(p+1)/2$. In this way, $\mathbf{\Omega}$ can be considered as a function of $\boldsymbol{\beta}$: $\mathbf{\Omega} = \mathbf{\Omega}(\boldsymbol{\beta})$. Denote the true value of $\boldsymbol{\beta}$ as $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}) = (\boldsymbol{\beta}_{10}, \mathbf{0})$, where the nonzero part $\boldsymbol{\beta}_{10}$ has the length of $s$.

In the adaptive LASSO penalty setting, we define

$$Q(\boldsymbol{\beta}) = L(\boldsymbol{\beta}) - n\lambda_n(|\tilde{\boldsymbol{\beta}}|^{-\gamma})^T|\boldsymbol{\beta}|,$$

where $L(\boldsymbol{\beta}) = \sum_{i=1}^n l_i(\mathbf{\Omega}(\boldsymbol{\beta})) = \frac{n}{2}\log|\mathbf{\Omega}| - \frac{n}{2}\log(2\pi) - \sum_{i=1}^n \frac{1}{2}\boldsymbol{x}_i^T\mathbf{\Omega}\boldsymbol{x}_i$ is the log-likelihood function and $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_1, \tilde{\beta}_2, \cdots, \tilde{\beta}_d)$ is a $a_n$-consistent estimator of $\boldsymbol{\beta}$, *i.e.*, $a_n(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = O_p(1)$. In addition, we denote $I(\boldsymbol{\beta}) = E\{[\frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta})][\frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta})]^T\}$ be the Fisher information matrix.

Let $\tau_n = n^{-1/2}$, we want to show that for any given $\epsilon > 0$, there exists a large constant $C$ such that

$$P\left\{\sup_{\|\boldsymbol{u}\|=C} Q(\boldsymbol{\beta}_0 + \tau_n\boldsymbol{u}) < Q(\boldsymbol{\beta}_0)\right\} \geq 1 - \epsilon \tag{B.1}$$

This implies that with probability at least $1-\epsilon$ that there exists a local maximum in the ball $\{\boldsymbol{\beta}_0 + \tau_n\boldsymbol{u} : \|u\| \leq C\}$. Hence there exists a local maximizer such that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p(\tau_n)$.

From the fact that only the first $s$ elements of $\boldsymbol{\beta}_0$ are non-zero, we have

$$\begin{aligned} D_n(\boldsymbol{u}) &= Q(\boldsymbol{\beta}_0+\tau_n\boldsymbol{u})-Q(\boldsymbol{\beta}_0) \\ &\leq L(\boldsymbol{\beta}_0+\tau_n\boldsymbol{u})-L(\boldsymbol{\beta}_0)-n\lambda_n\sum_{j=1}^{s}|\tilde{\beta}_j|^{-\gamma}(|\beta_{j0}+\tau_n\boldsymbol{u}|-|\beta_{j0}|) \\ &= \tau_n L'(\boldsymbol{\beta}_0)^T\boldsymbol{u}-\frac{1}{2}n\tau_n^2\boldsymbol{u}^T\boldsymbol{I}(\boldsymbol{\beta}_0)\boldsymbol{u}\{1+o_p(1)\}-n\lambda_n\tau_n\sum_{j=1}^{s}|\tilde{\beta}_j|^{-\gamma}\mathrm{sgn}(\beta_{j0})u_j \\ &= n^{-1/2}L'(\boldsymbol{\beta}_0)^T\boldsymbol{u}-\frac{1}{2}\boldsymbol{u}^T\boldsymbol{I}(\boldsymbol{\beta}_0)\boldsymbol{u}\{1+o_p(1)\}-n^{1/2}\lambda_n\sum_{j=1}^{s}|\tilde{\beta}_j|^{-\gamma}\mathrm{sgn}(\beta_{j0})u_j \qquad \text{(B.2)} \end{aligned}$$

Note that $n^{-1/2}L'(\boldsymbol{\beta}_0)=O_p(1)$. Thus the first term on the right hand side of (B.2) is on the order $O_p(1)$. For the third term of (B.2), we have $|\tilde{\beta}_j|^{-\gamma}=O_p(1)$ for $j=1,\cdots,s$ since $\tilde{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}_0$ and $\beta_{j0}\neq 0$. Thus, the third term is also on the order of $O_p(1)$ from the assumption that $n^{1/2}\lambda_n=O_p(1)$. By choosing a sufficiently large $C$, the second term dominates the first term and the third term uniformly in $\|u\|=C$. Then (B.1) holds.

Now, we want to show that with probability tending to 1 as $n\to\infty$, for any $\boldsymbol{\beta}_1$ satisfying $\boldsymbol{\beta}_1-\boldsymbol{\beta}_{10}=O_p(n^{-1/2})$ and any constant $C$,

$$Q\left\{\begin{pmatrix}\boldsymbol{\beta}_1\\ \mathbf{0}\end{pmatrix}\right\}=\max_{\|\boldsymbol{\beta}_2\|\leq Cn^{-1/2}}Q\left\{\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}\right\}. \qquad \text{(B.3)}$$

Denote $\boldsymbol{\beta}^*=\binom{\boldsymbol{\beta}_1}{\mathbf{0}}$, and $\boldsymbol{\beta}=\binom{\boldsymbol{\beta}1}{\boldsymbol{\beta}_2}=\boldsymbol{\beta}^*+n^{-1/2}\boldsymbol{u}$, where $\|\boldsymbol{u}\|\leq C$ and $u_j=0$ for all $j=1,\cdots,s$. Follow the same reasoning before,

$$\begin{aligned} &Q(\boldsymbol{\beta}^*+n^{-1/2}\boldsymbol{u})-Q(\boldsymbol{\beta}^*) \\ &= n^{-1/2}L'(\boldsymbol{\beta}^*)^T\boldsymbol{u}-\frac{1}{2}\boldsymbol{u}^T\boldsymbol{I}(\boldsymbol{\beta}^*)\boldsymbol{u}\{1+o_p(1)\}-n^{1/2}\lambda_n\sum_{j=s+1}^{d}|\tilde{\beta}_j|^{-\gamma}|u_j| \qquad \text{(B.4)} \end{aligned}$$

Since $C$ is a fixed constant, the second term on the right hand side of (B.4) will be at the order of $O_p(1)$. For $j=s+1,\cdots,d$, we have $\beta_{j0}=0$. Again, by $a_n$ consistency of $\tilde{\boldsymbol{\beta}}$, we have $a_n|\tilde{\beta}_j|=O_p(1)$ as $n\to\infty$. Thus, the order of the third term of (B.4) is $n^{1/2}\lambda_n a_n^{\gamma}\to\infty$ as $n\to\infty$ by our assumption. Hence (B.3) holds. This completes the proof of the sparsity part. The asymptotic normality of the estimator can be derived from Fan and Li (2001). □

## REFERENCES

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96** 1348–1360.