APPENDIX B: PROOF OF THEOREM 5.3

Proof. First of all, to simplify our notation, we write \( \Omega \) as a vector in the following way: divide the indexes of \( \Omega_0 = \{(\omega_{0ij}), i, j = 1, \ldots, p \} \) to two parts: \( A = \{(i, j), \omega_{0ij} \neq 0 & i \leq j \} \) and \( B = \{(i, j), \omega_{0ij} = 0 & i \leq j \} \). Denoting \( \Omega \) in a vector format, we write \( \beta = (\beta_1, \beta_2) \), where \( \beta_1 = (\omega_{ij}, (i, j) \in A) \) and \( \beta_2 = (\omega_{ij}, (i, j) \in B) \). As a result, \( \beta \) has the length of \( d = p(p + 1)/2 \).

In this way, \( \Omega \) can be considered as a function of \( \beta \): \( \Omega = \Omega(\beta) \). Denote the true value of \( \beta \) as \( \beta_0 = (\beta_{10}, \beta_{20}) = (\beta_{10}, 0) \), where the nonzero part \( \beta_{10} \) has the length of \( s \).

In the adaptive LASSO penalty setting, we define

\[
Q(\beta) = L(\beta) - n\lambda_n(\bar{\beta}^{-\gamma})^T|\beta|,
\]

where \( L(\beta) = \sum_{i=1}^{n} l_i(\Omega(\beta)) = \frac{n}{2} \log |\Omega| - \frac{n}{2} \log(2\pi) - \sum_{i=1}^{n} \frac{1}{2} x_i^T \Omega x_i \) is the log-likelihood function and \( \bar{\beta} = (\bar{\beta}_1, \bar{\beta}_2, \cdots, \bar{\beta}_d) \) is a \( a_n \)-consistent estimator of \( \beta \), i.e., \( a_n(\bar{\beta} - \beta_0) = O_p(1) \). In addition, we denote \( I(\beta) = E\{[\frac{\partial}{\partial \beta} l(\beta)][\frac{\partial}{\partial \beta} l(\beta)]^T\} \) 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

\[
P \left\{ \sup_{||u||=C} Q(\beta_0 + \tau_n u) < Q(\beta_0) \right\} \geq 1 - \epsilon
\]

This implies that with probability at least \( 1 - \epsilon \) that there exists a local maximum in the ball \( \{\beta_0 + \tau_n u : ||u|| \leq C\} \). Hence there exists a local maximizer such that \( ||\hat{\beta} - \beta_0|| = O_p(\tau_n) \).
The asymptotic normality of the estimator can be derived from Fan and Li (2001). For any $a_n$ satisfying $a_n |\tilde{\beta}_j| = O_p(1)$ as $n \to \infty$. Again, by $a_n$ consistency of $\tilde{\beta}$, 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