1 Proof Strategy

1.1 Sparsity Condition 1, Thresholding

Assume that there is a mapping $L : (i, j) \mapsto \{0, 1\}$, representing the "locations" of large elements. For $L_{ij} := L(i, j) = 1$ or 0, the sparsity condition are different, i.e.,

$$\sum_j^p |\sigma_{ij}|^{q_0} \mathbf{1}_{L_{ij}=0} < c_0(p)$$

and

$$\sum_j^p |\sigma_{ij}|^{q_1} \mathbf{1}_{L_{ij}=1} < c_1(p)$$

where $c_0(p), c_1(p) \longrightarrow \infty$ as $p \longrightarrow \infty$, but might have different rate of divergence. The belief is when $L_{ij} = 1$, although σ_{ij} might be large, the number of such σ_{ij} must be small, so the sparsity condition can hold with a smaller q. Although there potentially are many $L_{ij} = 0$, but each σ_{ij} will be small, and hence we can allow a relatively large q.

Then following bickel, consider a hard thresholding estimator with adaptive thresholds,

$$T_{t,L}(\sigma_{ij}) = \begin{cases} \sigma_{ij} & \text{if} \quad \sigma_{ij} > t_1 \quad \text{and} \quad L_{ij} = 1 \\ \sigma_{ij} & \text{if} \quad \sigma_{ij} > t_0 \quad \text{and} \quad L_{ij} = 0 \\ = 0 & \text{otherwise} \end{cases}$$

In practice we might estimate L with $\hat{L}$, We have the decomposition:

$$\left\| T_{t,\hat{L}}(\hat{\Sigma}) - \Sigma \right\| \leq \left\| T_{t,\hat{L}}(\Sigma) - \Sigma \right\| + \left\| T_{t,\hat{L}}(\hat{\Sigma}) - T_{t,\hat{L}}(\Sigma) \right\| = \mathbf{I} + \mathbf{II}$$

The first term $\mathbf{I}$ can be bounded by the L^1 norm,

$$\mathbf{I} \leq \max_i \sum_j^p |\sigma_{ij}| \left(\hat{L}^0_{ij} \mathbf{1}_{|\sigma_{ij}| \leq t_0} + \hat{L}^1_{ij} \mathbf{1}_{|\sigma_{ij}| \leq t_1} \right)$$

where $L^0_{ij} := \mathbf{1}_{L_{ij}=0}$, the location where $L_{ij} = 0$.

$$
\begin{aligned}
\mathbf{I} &\le \max_i \sum_j |\sigma_{ij}| \left(L^0_{ij} \mathbf{1}_{|\sigma_{ij}| \le t_0} + L^1_{ij} \mathbf{1}_{|\sigma_{ij}| \le t_1} \right) + \max_i \sum_j |\sigma_{ij}| \left(\mathbf{1}_{|\sigma_{ij}| \le t_0} (\hat{L}^0_{ij} - L^0_{ij}) \right) \\
&+ \max_i \sum_j |\sigma_{ij}| \mathbf{1}_{|\sigma_{ij}| \le t_1} \left(\hat{L}^1_{ij} - L^1_{ij} \right) \\
&= \mathbf{I}_1 + \mathbf{I}_2 + \mathbf{I}_3
\end{aligned}
$$

Assume that $\max_{ij} \left| \hat{L}^0_{ij} - L^0_{ij} \right| = O_p(k_n)$. Then we have

$$
\mathbf{I} \lesssim t_0^{1-q_0} c_0 + t_1^{1-q_1} c_1 + \mathbf{I}_2 + \mathbf{I}_3
$$

For $\mathbf{I}_2$, we have

$$
\mathbf{I}_2 = \max_i \sum_j |\sigma_{ij}| \left(\mathbf{1}_{|\sigma_{ij}| \le t_0} (\hat{L}^0_{ij} - L^0_{ij}) \right) \lesssim t_0^{1-q_1} c_1(p) k_n
$$

and for $\mathbf{I}_3$ we have,

$$
\mathbf{I}_3 = \max_i \sum_j |\sigma_{ij}| \left(\mathbf{1}_{|\sigma_{ij}| \le t_1} (\hat{L}^1_{ij} - L^1_{ij}) \right) \lesssim t_1^{1-q_0} c_0(p) k_n
$$

Hence, $I \lesssim c_1(t_1^{1-q_1} + k_n t_0^{1-q_1}) + c_0(t_0^{1-q_0} + k_n t_1^{1-q_0})$.

For Item **II**, we can decompose it like in Bickel and Levina, (handwriting):

$$
\mathbf{II} \lesssim c_1 \left(t_1^{-q_1} \sqrt{\frac{\log p}{n}} + k_n t_0^{-q_1} \sqrt{\frac{\log p}{n}} + t_1^{1-q_1} + k_n t_0^{1-q_1} \right) + c_0 \left(t_0^{-q_0} \sqrt{\frac{\log p}{n}} + k_n t_1^{-q_0} \sqrt{\frac{\log p}{n}} + t_0^{1-q_0} + k_n t_1^{1-q_0} \right)
$$

The thing is that if $k_n = o(1)$, both t_1, t_0 are of order $\sqrt{\log p / n}$.

1.2 Sparsity Condition 2, Banding + Thresholding Estimator

Suppose we don't shrink the $\hat{\sigma}_{ij}\hat{L}^1_{ij}$. And put conditions where $q_1 = 0$, i.e., like banding but with estimated locations, then we can define the operator as

$$T_{t,L}\left(\sigma_{ij}\right) = \begin{cases} \sigma_{ij} & \text{if} \quad L_{ij} = 1 \\ s_t(\sigma_{ij}) & \text{if} \quad L_{ij} = 0 \end{cases}$$

where s_t is the generalized thresholding operator.

Then we can show the convergence rate of $\left\|T_{t,\hat{L}}\left(\hat{\Sigma}\right) - \Sigma\right\|$ similar to Bickel, et al 2008 with banding.

Define $B_L(\Sigma) = \left[\sigma_{ij}L^1_{ij}\right]$, then for $(i, j) : L^1_{ij}$, there are ways to control the maximal difference $\max_{ij} \hat{\sigma}_{ij} - \sigma_{ij}$. The rest can be