Define the gap that quantifies the difference for clustering k-th layer into cluster a to cluster b. Define the precision matrix Ω_k

$$\Omega_k(a) = \Theta_0 + u_k\Theta_a$$

$$\Delta_k(a,b) := \Omega_k(a) - \Omega_k(b) = u_k(\Theta_a - \Theta_b), \quad \Delta_{\min} = \min_{k\in[d]} \min_{a\neq b\in[R]} \|\Delta_k(a,b)\|_F.$$

$$(\hat{\mu}^{\text{oracle}}, \hat{\Theta}^{\text{oracle}}) = \min_{\boldsymbol{\Omega}\in\mathcal{P}(r)} Q(\Omega; z^{\text{true}}) + \lambda \text{Pen}(\Theta)$$

$$\text{Err}_{\text{oracle}}(\delta) = \sum_{k\in[d]} \sum_{b\in[r]/z_k} \|\Delta_k(z_k,b)\|_F^2 \mathbb{1}\left\{\|\Delta_k(z_k,b)\|_F^2 \leq -(1+\delta)\left\langle \varepsilon_k,\ \hat{\Delta}_k^{\text{oracle}}(z_k,b)\right\rangle\right\}$$

The following lemma shows that, once the true clustering is provided, the oracle gap statistics has weak correlation with the noise.

Lemma 1. For any δ sufficiently small. Under the exclusive events, with probability $1-\exp(-\Delta_{\min})$,

$$\text{Err}_{\text{oracle}}(\delta) \lesssim d\exp(-(1-\delta)\Delta_{\min}^2).$$

Proof.

$$\begin{aligned}
&\mathbb{P}\left\{\left\langle \varepsilon_k,\ \hat{\Delta}_k^{\text{oracle}}(z_k,b)\right\rangle \leq -(1-\delta)\|\Delta_k(z_k,b)\|_F^2\right\}\\
\leq\ &\underbrace{\mathbb{P}\left\{\left\langle \varepsilon_k,\ \Delta_k(z_k,b)\right\rangle \leq -(1-2\delta)\|\Delta_k(z_k,b)\|_F^2\right\}}_{:=I} + \underbrace{\mathbb{P}\left\{\left\langle \varepsilon_k,\ \hat{\Delta}_k^{\text{oracle}}(z_k,b) - \Delta_k(z_k,b)\right\rangle \leq -\delta\|\Delta_k(z_k,b)\|_F^2\right\}}_{:=II}.
\end{aligned}$$

For the first term, based on the uniform concentration of sample covariance, we have

$$I \leq \mathbb{P}\left\{\|\varepsilon_k\|_\infty \geq \frac{1-2\delta}{p}\|\Delta_k(a,b)\|_F\right\} \lesssim \exp\left(\frac{n(1-2\delta)^2}{p^2}\|\Delta_k(a,b)\|_F^2\right).$$

For the second term,

$\square$