Weekly Report

WU Zihan

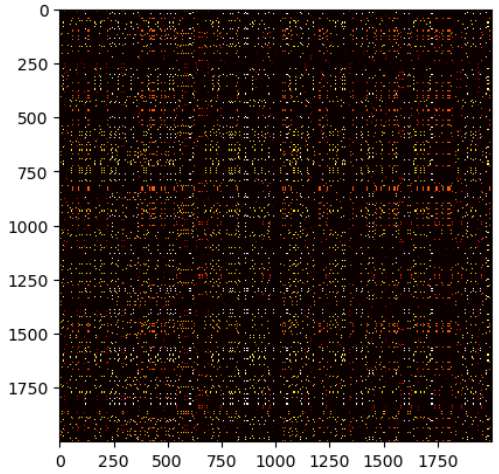
September 7, 2023

Results for simulated data matrix

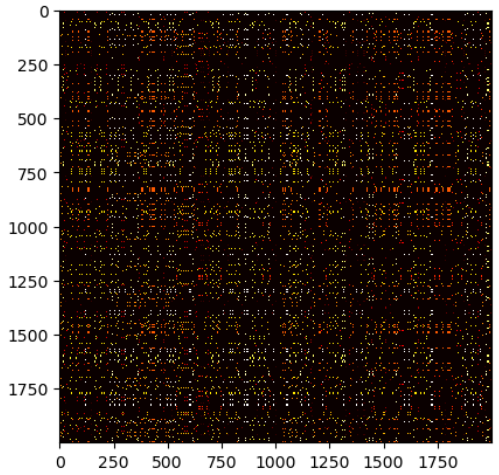
Experiment settings

- $A_{M\times N}$, $M = N = 2000$
- Biclusters number $k = 15$
- Bicluster height m_i and width n_i subject to uniform distribution $U(50, 200)$
- Noise level $\sigma_{\text{Noise}} = 10^{-5} \times \max(A)$

Experiment results 100% on both precision and recall



(a) Ground truth



(b) Biclusters detected

Figure: Comparison of ground truth and detected biclusters

Methods

Score Function for Biclustering

- $$\langle x, y \rangle = \exp(-\frac{||x-y||_1^2}{2\alpha||x||_1||y||_1})$$
- Use the inner product above to substitute for the co-variance

Selection of S_{th}

- For a bicluster $B \in \mathbb{R}^{T_m \times T_n}$
- Suppose $\max(||B||_1, ||B^\top||_1) \leq B_{\max}$
- If B is a bicluster with $\beta = 1\%$ tolerance, then
- $$S \leq \frac{1 - \exp^{-\frac{1}{2\alpha \min(T_m, T_n)^2}}}{\min(T_m, T_n) - 1}$$
- In the experiment, $\alpha = 1, S_{\text{th}} = 0.05$