Joint Inference via Bi-Clustering on Graphs
Marius Cătălin Iordan, Vignesh Ramanathan, Arun Chaganty
EE364b: Convex Optimization II Class Project

Identifying Characters in Videos and Scripts
- Video content comprises ≥50% of all worldwide Internet traffic.
- Need an automatic way to label people, objects, and actions in videos.
- We would like to use movie scripts as training data for this task.
- Key problem: Joint tasks of face recognition and coreference resolution
  - Coreference resolution disambiguates coreferents like ‘she’ or ‘the man’.

Graph Bi-Clustering
- Joint task equivalent to performing simultaneous clustering on two graphs:
  - First graph: edges between similar faces
  - Second graph: edges between coreferent script characters (‘he’, ‘Chris’)
- We also learn a mapping $M$ between faces and names

Optimization problem:
\[
\begin{align*}
\text{minimize} & \quad tr(Y_1^T M_1 Y_1) + tr(Y_2^T M_2 Y_2) + \|Y_1 - M Y_2\|_F^2 \\
\text{subject to} & \quad A^T \begin{bmatrix} \vec(Y_1) \\ \vec(Y_2) \\ \vec(M) \end{bmatrix} - b \leq 0 \quad \text{and} \quad M \succeq 0 \\
& \quad Y_1 \in [0,1]^{n_1 \times d}, \quad Y_2 \in [0,1]^{n_2 \times d} \quad \text{and} \quad M \in [0,1]^{n_1 \times n_2}, \quad M_1 = 1, \quad Y_1 Y_2 = 1 \\
& \quad Y_1, Y_2: \text{cluster assignment matrices with } d \text{ clusters and } n_1, n_2 \text{ nodes} \\
& \quad M_{ij}: \text{indicator variable for whether the } i\text{-th point in the first graph is mapped to the } j\text{-th point in the second graph} \\
& \quad A, b: \text{linear constraints} \\
& \quad M_1: \text{mapping matrix} \\
& \quad \Pi_1 \subset \mathbb{R}^{n_1 \times n_1}, \quad \Pi_2 \subset \mathbb{R}^{n_2 \times n_2}: \text{grouping matrices given by problem data} \\
\end{align*}
\]

Numerical Example
- Mapping agreement term in objective not convex: $\|Y_1 - M Y_2\|_F^2$
- Experiment with approximate solutions on artificial dataset
- Constructed simulated instance of bi-clustering problem:
  - Generate random cluster assignment $Y_2$ and random mapping $M$
  - Set $Y_1 = MY_2$
  - Generate random features $X_1 \in \mathbb{R}^{n_1 \times 2}$, $X_2 \in \mathbb{R}^{n_2 \times 2}$ consistent with cluster assignments
  - Construct cluster matrices $\Pi_1$ from feature matrices $X_1$ as described in (Bach & Harchaoui, 2007)
  - Generate $A$ and $b$ such that the ordering constraint on the two graphs is satisfied ($A$ and $b$ also serve to seed the clustering of the second domain with weak supervision)
- Parameters: $n_1 = 20, n_2 = 40$ data points, $d = 3$ clusters, and 5 inequality constraints on $Y_2$

Solution 1: Alternating Minimization
- Optimization problem is bi-convex in $M$ and $(Y_1, Y_2)$
- Alternating minimization commonly used in practice
- Only guaranteed to converge to local minimum

Algorithm 1
- Given parameters $\Pi_1, \Pi_2, A, b$.
- Initialize $Y_1, Y_2$ by removing mapping term $\|Y_1 - MY_2\|_F^2$ and solving independently for $Y_1, Y_2$. Fix $Y_1$ and $Y_2$ found above and solve QP in $M$.
- Repeat
  - Fix $M$ and solve QP in $Y_1$ and $Y_2$.
  - Fix $Y_1$ and $Y_2$ and solve QP in $M$.
- Until local minimum found.

References

Solution 2: SDP Relaxation
- Problem can also be formulated as a non-convex QCQP.
- We solve the following SDP relaxation of the bi-clustering problem:

\[
\begin{align*}
\text{minimize} & \quad tr(Q_1 Y_1 Y_1^T) + tr(Q_2 Y_2 Y_2^T) + \|E\|_F^2 \\
\text{subject to} & \quad A^T \begin{bmatrix} \vec(Y_1) \\ \vec(Y_2) \\ \vec(M) \end{bmatrix} - b \leq 0 \\
& \quad M \succeq 0, \quad Q_1 \succeq 0, \quad Q_2 \succeq 0 \\
& \quad \frac{P_{ij}}{P_{ij}^T} - \frac{P_{ij}^T}{P_{ij}} \geq 0, \quad P_{ij} = \begin{bmatrix} M_{ij}^T \\ Y_{ij} \end{bmatrix}, \quad 1 \leq i \leq n_1, 1 \leq j \leq d \\
& \quad E_{ij} \leq \|Y_{ij} - tr(P_{ij})\|_F, \quad 1 \leq i \leq n_1, 1 \leq j \leq d \\
& \quad E_{ij} \geq 0, \quad 1 \leq i \leq n_1, 1 \leq j \leq d \\
& \quad \Pi_1 \succeq 0, \quad \Pi_2 \succeq 0, \quad \Pi_1 \succeq 0 \\
& \quad E \in \mathbb{R}^{n_1 \times d}$
\end{align*}
\]

The error in the mapping $\Pi_3 \succeq 0$ handles the non-convex quadratic constraint.

Results
- Simulated Data and Ground Truth Clusters
- Method Comparison

Conclusion
- SDP relaxation much slower and less accurate than alternating minimization for graph bi-clustering.
- Results on simulated data suggest alternating minimization may offer a reasonable solution to original joint inference problem.