Nonconvex Sparse Graph Learning under Laplacian Constrained Graphical Model

Jiaxi Ying, José Vinícius de M. Cardoso, and Daniel P. Palomar

For more information visit: https://www.danielpalomar.com; R Package: https://github.com/mirca/sparseGraph

1 Graphical Model

A graph $G = (V, E, W)$ is an intuitive way to represent relationships between entities.

- Nodes: $V = \{1, 2, \ldots, n\}$ corresponding to the entities.
- Edges: $E = \{1, 2, \ldots, m\}$ encodes conditional dependence between entities.
- Weights: $W$ is a weight matrix with $W_{ij}$ the graph weight between node $i$ and node $j$.

1.1 Laplacian Constrained Gaussian Graphical Model

Graph Laplacian: $L = D - W$, where $D$ is the degree matrix with $D_{ii} = \sum_{j \neq i} W_{ij}$.

Theorem 2 shows that a large regularization parameter of the $\ell_1$-norm does not work.

$\lambda = 1$ and $\lambda'$ are monotone and Lipschitz continuous for $\lambda \in [0, \infty)$.

$\Theta^\dagger$ satisfies the following conditions:
1. $\Theta_j^\dagger = 0$, and $\Theta_j^\dagger$ is monotone and Lipschitz continuous for $\lambda \in [0, \infty)$.
2. There exists a $\gamma > 0$ such that $\lambda_j \geq \gamma$.
3. $\lambda_j \geq \lambda_{j+1}$ for $j \in [n]$, where $n = |V|$.
4. $\lambda_j \geq \lambda_{j+1}$ for $j \in [n]$, where $n = |V|$.
5. $\lambda_j \geq \lambda_{j+1}$ for $j \in [n]$, where $n = |V|$.
6. $\lambda_j \geq \lambda_{j+1}$ for $j \in [n]$, where $n = |V|$.
7. $\lambda_j \geq \lambda_{j+1}$ for $j \in [n]$, where $n = |V|$.
8. $\lambda_j \geq \lambda_{j+1}$ for $j \in [n]$, where $n = |V|$.
9. $\lambda_j \geq \lambda_{j+1}$ for $j \in [n]$, where $n = |V|$.
10. $\lambda_j \geq \lambda_{j+1}$ for $j \in [n]$, where $n = |V|$.

The minimal nonzero graph weight satisfying $L_{ij} = 0$ is a constant.

2 $\ell_1$-norm Analysis and Proposed Method

2.1 $\ell_1$-norm Does Not Work

The $\ell_1$-norm regularized maximum likelihood estimation of Laplacian constrained precision matrices $\{1, 2\}$ can be formulated as

$$\min_{\Theta \in \mathbb{R}^{n \times n}} -\log det(\Theta + J) + \ell(\Theta) = \sum_{i,j} \lambda_j \Theta_{ij},$$

where $\lambda_j > 0$ is a nonnegative sparsity-promoting function such as SCAD and MCP.

To handle the constraint $\Theta \in S_+^n$, we introduce a linear operator $L \{3\}$ that maps a vector $w$ to a Laplacian matrix $L_w$.

$$L_w = \sum_{i,j} \frac{W_{ij}}{\lambda_j} \Theta_{ij},$$

where $\lambda_j > 0$ is a nonnegative sparsity-promoting function such as SCAD and MCP.

The optimization (4) can be reformulated as

$$\min_{\Theta \in \mathbb{R}^{n \times n}} -\log det(\Theta + J) + \ell(\Theta) = \sum_{i,j} \lambda_j \Theta_{ij},$$

where $\lambda_j > 0$ is a nonnegative sparsity-promoting function such as SCAD and MCP.

We establish a sequence $w(\ell_k)$ by solving a sequence of sub-problems

$$w^{(\ell_k)} = \arg \min_{w \in \mathbb{R}^{n \times n}} -\log det(\Theta + J) + \ell(\Theta) = \sum_{i,j} \lambda_j \Theta_{ij},$$

The optimization (6) can be solved by a projected gradient descent algorithm with backtracking line search.

Algorithm 1 Nonconvex Graph Learning (NGL)

Input: Sample covariance $S$, $\lambda$, $w(0)$;

$\ell = 1$;

while Stopping criteria not met do

1. Update $w^{(\ell)} = L_w(\lambda - 1/\ell)I_{n \times n}$, for $\ell = 1, \ldots, p - 1$;

2. Update $w^{(\ell)} = w^{(\ell)} - \lambda_j \Theta_{ij}$ for $j \in [n]$;

3. Update $w^{(\ell)} = w^{(\ell)} - \lambda_j \Theta_{ij}$ for $j \in [n]$;

4. $\ell = \ell + 1$;

5. end while

Output: $w^{(\ell)}$.

2.2 Proposed Method

The penalized maximum likelihood of the precision matrix with Laplacian structural constraints can be formulated as

$$\min_{\Theta \in \mathbb{R}^{n \times n}} -\log det(\Theta + J) + \ell(\Theta) = \sum_{i,j \in S_+} \lambda_j \Theta_{ij},$$

where $h_\gamma$: is the minimal nonzero graph weight satisfies $L_{ij} = 0$ and the estimated graph is fully connected.

The labels represent the life status of patients, alive (green) or no longer alive (red).

Figure 2: Performance measures (a) Number of positive edges, (b) Relative error and (c) $F$-score as a function of regularization parameter $\lambda$.

3 Experimental Results

3.1 Synthetic Data

The data matrix $X \in \mathbb{R}^{p \times n}$ with each column independently sampled from GMRF. The ground-truth graph is a random Barabási-Albert graph with $M$ nodes, and the weights are randomly sampled from $U(0, 1)$. The compared methods include the state-of-the-art GLE-ADMM algorithm [2] and the baseline projected gradient descent with $\ell_1$-norm.

Figure 3: Performance measures (a) Number of positive edges, (b) Relative error and (c) $F$-score as a function of the sample size ratio $n/p$.

3.2 Real-world Data

The data set is 2019-nCoV [4] from 98 Chinese patients affected by the outbreak of 2019-nCoV on early February 2020. The features include age, gender, and location. The labels represent the life status of patients, alive (green) or no longer alive (red).

Figure 4: The learned graphs using the 2019-nCoV data set by (a) GLE-ADMM, (b) NGL-SCAD (proposed method), and (c) NGL-MCP (proposed method).

References


