Nonconvex Sparse Graph Learning under Laplacian-structured Graphical Model

a talk by

Jiaxi Ying, José Vinícius de M. Cardoso, and Daniel P. Palomar
The Hong Kong University of Science and Technology

Thirty-fourth Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada

December, 2020

Learning Sparse Undirected Connected Graphs

- data generating process: Laplacian-constrained Gaussian Markov Random Field (L-GMRF) with rank $p-1$
- its $p \times p$ precision matrix $\boldsymbol{\Theta}$ is modeled as a combinatorial graph Laplacian
- state of the art (Egilmez *et al.* 2017)[1], (Zhao *et al.* 2019)[2]:

$$\begin{array}{ll} \underset{\boldsymbol{\Theta} \succeq \mathbf{0}}{\text{minimize}} & \operatorname{tr}(\mathbf{S}\boldsymbol{\Theta}) - \log \det^{\star}\left(\boldsymbol{\Theta} + \boldsymbol{J}\right) + \lambda \|\boldsymbol{\Theta}\|_{1,\text{off}}, \\ \text{subject to} & \boldsymbol{\Theta}\mathbf{1} = \mathbf{0}, \Theta_{ij} = \Theta_{ji} \leq 0 \end{array} \tag{1}$$

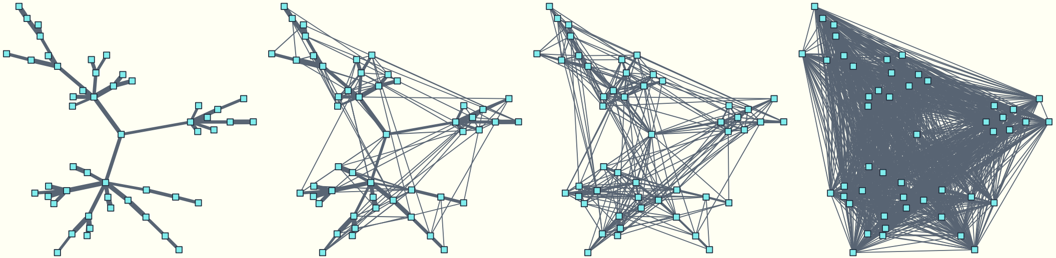
where $\boldsymbol{J} = \frac{1}{p}\mathbf{1}\mathbf{1}^{\top}$, $\|\boldsymbol{\Theta}\|_{1,\text{off}} = \sum_{i>j} |\Theta_{ij}|$ is the entrywise ℓ_1-norm, and $\lambda \geq 0$

[1] HE Egilmez *et al.* Graph learning from data under Laplacian and structural constraints. *IEEE Journal of Selected Topics in Signal Processing* 11 (6), 825-841.

[2] L Zhao *et al.* Optimization algorithms for graph laplacian estimation via ADMM and MM. *IEEE Transactions on Signal Processing* 67 (16), 4231-4244.

Are sparse solutions recoverable via ℓ_1-norm?

- TL;DR: **they aren't**
- empirically:



(a) ground-truth (b) $\lambda = 0$ (c) $\lambda = 0.1$ (d) $\lambda = 10$

Are sparse solutions recoverable via ℓ_1-norm?

- theoretically:

Theorem

Let $\hat{\boldsymbol{\Theta}} \in \mathbb{R}^{p \times p}$ be the global minimum of (1) with $p > 3$. Define $s_1 = \max_k S_{kk}$ and $s_2 = \min_{ij} S_{ij}$. If the regularization parameter λ in (1) satisfies $\lambda \in [(2 + 2\sqrt{2})(p+1)(s_1 - s_2), +\infty)$, then the estimated graph weight $\hat{W}_{ij} = -\hat{\Theta}_{ij}$ obeys

$$\hat{W}_{ij} \geq \frac{1}{(s_1 - (p+1)s_2 + \lambda)p} > 0, \quad \forall\, i \neq j.$$

Proof

Please refer to our supplementary material 😄

Our framework for sparse graphs

- nonconvex formulation:

$$\underset{\boldsymbol{w} \geq \boldsymbol{0}}{\text{minimize}} \quad \text{tr}(\boldsymbol{S}\mathcal{L}\boldsymbol{w}) - \log\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) + \textstyle\sum_i h_\lambda(w_i) \tag{2}$$

$\mathcal{L}$ is the Laplacian operator and $h_\lambda(\cdot)$ is a nonconvex regularizer such as

  - *Minimax Concave Penalty (MCP)*
  - *Smoothly Clipped Absolute Deviation (SCAD)*

Our framework for sparse graphs

Algorithm 0: Connected sparse graph learning

Data: Sample covariance $\boldsymbol{S}$, $\lambda > 0$, $\hat{\boldsymbol{w}}^{(0)}$

Result: Laplacian estimation: $\mathcal{L}\hat{\boldsymbol{w}}^{(k)}$

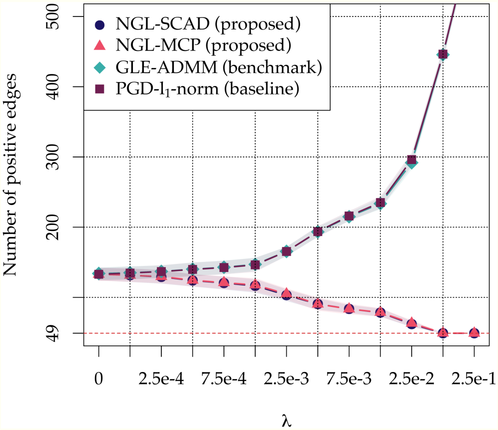
$k \leftarrow 1$

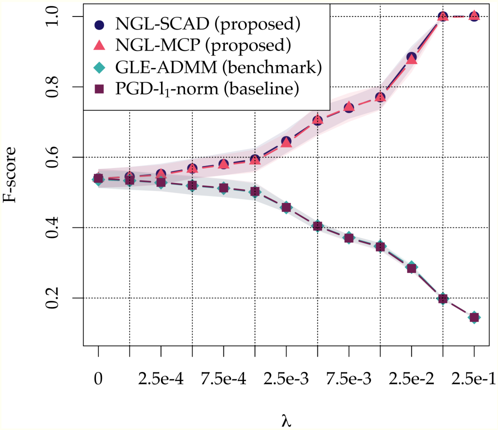
while *stopping criteria not met* **do**

- $\triangleright$ update $z_i^{(k-1)} = h'_\lambda(\hat{w}_i^{(k-1)})$, for $i = 1, \ldots, p(p-1)/2$
- $\triangleright$ update $\hat{\boldsymbol{w}}^{(k)} = \arg\min_{\boldsymbol{w} \geq \boldsymbol{0}} - \log\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) + \mathrm{tr}(\boldsymbol{S}\mathcal{L}\boldsymbol{w}) + \sum_i z_i^{(k-1)} w_i$
- $\triangleright$ $k \leftarrow k + 1$

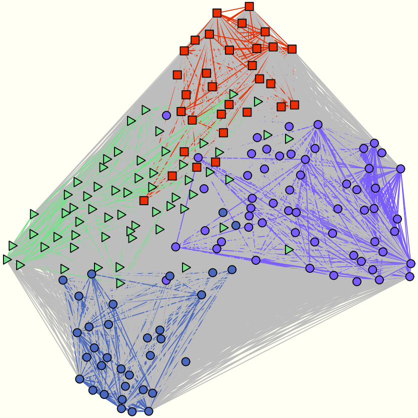
end

Sneak peek on the results: synthetic data

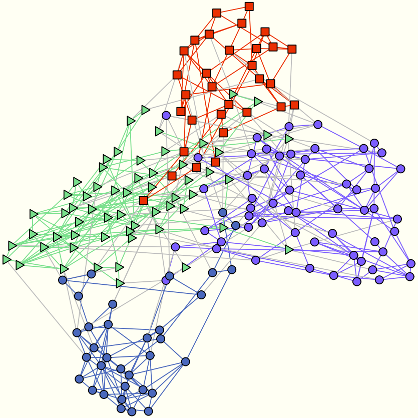




Sneak peek on the results: S&P 500 stocks



(a) GLE-ADMM (benchmark) $\lambda = 0$



(b) NGL-MCP (proposed) $\lambda = 0.5$

Reproducibility

- The code for the experiments can be found at `https://github.com/mirca/sparseGraph`
- Convex Research Group at HKUST: `https://www.danielppalomar.com`