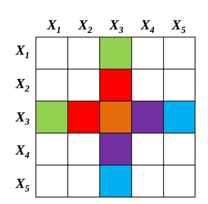
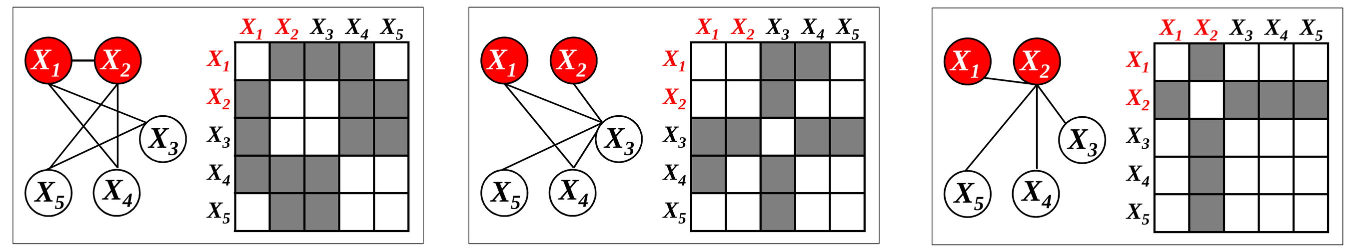


# STRUCTURED ESTIMATION OF GAUSSIAN GRAPHICAL MODELS

KARTHIK MOHAN (karna@uw.edu), MICHAEL JAE-YOON CHUNG (mjyc@cs.washington.edu), SEUNGYEOP HAN (syhan@cs.washington.edu),
DANIELA WITTEN (dwitten@uw.edu), SU-IN LEE (suinlee@cs.washington.edu), MARYAM FAZEL (mfazel@uw.edu)
University of Washington, Seattle

## PROBLEM

- We consider the problem of jointly estimating multiple *Gaussian graphical models*.
- **Setting:** High-dimensional setting with more variables than samples.
- **Prior knowledge:** We assume prior knowledge on the structure of the GGMs - Specifically that the GGMs differ in *node-perturbations*. A node is said to be *perturbed* between two networks if the node has a high degree in the difference of networks.
- Applications in gene-regulatory networks. Regulatory genes play a prominent role in gene-regulatory networks. Detection of regulatory genes that differ between brain cancer and lung cancer gene-regulatory networks is an important application.
- **Main contribution:** Propose a novel convex optimization based approach to detect *node-based* perturbations in GGMs along with an efficient *alternating direction method of multipliers* (ADMM) algorithm.



a) Network 1 b) Network 2 c) Difference of Networks

## PAST WORK

- *Graphical Lasso* [3] - Single network estimation

$$\underset{\Theta\in S^p_{++}}{\text{maximize}}\,\{n(\log\det\Theta - \text{trace}(\mathbf{S}\Theta)) - \lambda\|\Theta\|_1\},$$

where $\mathbf{S}$ is sample-covariance matrix given by $\mathbf{S} = \mathbf{X}\mathbf{X}^T/n$ where $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n \sim$ i.i.d $\mathcal{N}(0,\Sigma)$ and $\Sigma \in S^p_{++}$.

- *Fused Graphical Lasso* (FGL) [4] - Multiple network estimation based on Edge based perturbations

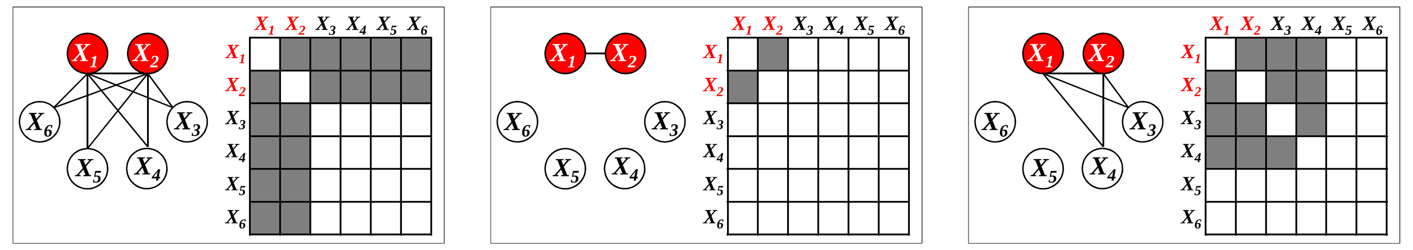
$$\underset{\Theta^1\in S^p_{++},\ldots,\Theta^K\in S^p_{++}}{\text{maximize}}\left\{L(\Theta^1,\ldots,\Theta^K) - \lambda_1\sum_{k=1}^K\|\Theta^k\|_1 - \lambda_2\sum_{k\neq l}\|\Theta^k - \Theta^l\|_1\right\},$$

where, $L(\Theta^1,\ldots,\Theta^K) = \sum_{k=1}^K n_k\left(\log\det\Theta^k - \text{trace}(\mathbf{S}^k\Theta^k)\right)$. $\mathbf{S}^k = \mathbf{X}^k(\mathbf{X}^k)^T/n$ where $\mathbf{X}^k_1, \mathbf{X}^k_2, \ldots, \mathbf{X}^k_n \sim$ i.i.d $\mathcal{N}(0,\Sigma^k)$ and $\Sigma^k \in S^p_{++}$.

## NAIVE APPROACH

$$\underset{\Theta^1\in S^p_{++},\Theta^2\in S^p_{++}}{\text{maximize}}\left\{L(\Theta^1,\Theta^2) - \lambda_1\|\Theta^1\|_1 - \lambda_1\|\Theta^2\|_1 - \lambda_2\sum_{j=1}^p\|\Theta^1_j - \Theta^2_j\|_2\right\},$$

- Support of difference of estimates expressed as *complement of the union of groups* instead of *union of groups*.
- Figure depicts the true difference of networks and estimated difference of networks.



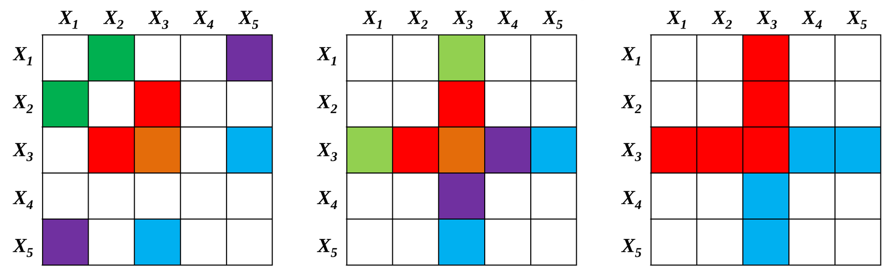
a) True Difference b) Naive difference c) Naive difference with smaller $\lambda_2$

## ROW-COLUMN OVERLAP NORM

**Definition.** *The row-column overlap norm (RCON) induced by a matrix norm $f$ is defined as*

$$\Omega_f(\mathbf{A}) = \min_{\mathbf{V}:\mathbf{A}=\mathbf{V}+\mathbf{V}^T} f(\mathbf{V})$$

- Variant of Overlap norm introduced by Jacob et al [2]
- Accounts for symmetry elegantly.
- Generalizes $\ell_1/\ell_2$ norm used in [2] to arbitrary norms $f$.



$f = \ell_1/\ell_1$ $f = \ell_1/\ell_2$ $f = \ell_1/\ell_\infty$

## CONVEX FORMULATION -PNJGL

*Perturbed node joint graphical lasso* (PNJGL):

$$\underset{\Theta^1\in S^p_{++},\Theta^2\in S^p_{++}}{\text{maximize}}\left\{L(\Theta^1,\Theta^2) - \lambda_1\|\Theta^1\|_1 - \lambda_1\|\Theta^2\|_1 - \lambda_2\Omega_f(\Theta^1-\Theta^2)\right\}.$$

$f(\Theta) = \sum_{i=1}^p\|\Theta_i\|_q$ known as the $\ell_1/\ell_q$ norm. FGL special case of PNJGL with $q=1$.

## POSSIBLE ALGORITHMS

- Proximal operator computations for overlapping group lasso penalties [5] don't apply since the RCON penalty promotes union of overlapping groups instead of the complement. Proximal operator for RCON has no closed-form. Also PSD constraints complicate computation.
- Projected Subgradient: The subgradient computation for RCON is non-trivial.
- Second order methods such as interior-point methods are expensive: $O(p^6)$.
- Our approach: ADMM. Per-iteration complexity: $O(p^3)$.

## ADMM APPROACH

Consider the following simple optimization problem,

$$\begin{array}{ll}\underset{X}{\text{minimize}} & g(X)+h(X)\\ \text{subject to} & X\in\mathcal{X}\end{array}$$

The ADMM approach [1] is as follows:

1. Decide parts of the objective to decouple. Here we decouple g and h by introducing new variable $Y$ and constraining $X=Y$. The resulting optimization problem is given by,

$$\begin{array}{ll}\underset{X,Y}{\text{minimize}} & g(X)+h(Y)\\ s.t. & X\in\mathcal{X}, X=Y\end{array}$$

2. Form the augmented Lagrangian to (1) by first forming the Lagrangian and then *augmenting* it with a quadratic function of equality constraints. Lagrangian given by $L(X,Y,\Lambda) = g(X)+h(Y)+\langle\Lambda, X-Y\rangle$. Augmented Lagrangian: $L(X,Y,\Lambda)+\frac{\rho}{2}\|X-Y\|_F^2$.
3. Next minimize in turn each primal variable, keeping all other variables fixed. The dual variables get updated using a dual-ascent update.

## ADMM ALGORITHM

- Reformulation:

$$\begin{array}{ll}\underset{\Theta^1\in\mathcal{S}^p_{++},\Theta^2\in\mathcal{S}^p_{++},\mathbf{Z}_1,\mathbf{Z}_2,\mathbf{V},\mathbf{W}}{\text{minimize}} & \left\{-L(\Theta^1,\Theta^2) + \lambda_1\|\mathbf{Z}_1\|_1 + \lambda_1\|\mathbf{Z}_2\|_1 + \lambda_2\sum_{j=1}^p\|\mathbf{V}_j\|_q\right\}\\ \text{subject to} & \Theta^1-\Theta^2 = \mathbf{V}+\mathbf{W}, \mathbf{V}=\mathbf{W}^T, \Theta^1=\mathbf{Z}_1, \Theta^2=\mathbf{Z}_2.\end{array}$$

- Augmented Lagrangian:

$$\begin{aligned}&-\ L(\Theta^1,\Theta^2) + \lambda_1\|\mathbf{Z}_1\|_1 + \lambda_1\|\mathbf{Z}_2\|_1 + \lambda_2\sum_{j=1}^p\|\mathbf{V}_j\|_q + \langle\mathbf{F},\Theta^1-\Theta^2-(\mathbf{V}+\mathbf{W})\rangle\\ &+\ \langle\mathbf{G},\mathbf{V}-\mathbf{W}^T\rangle + \langle\mathbf{Q}_1,\Theta^1-\mathbf{Z}_1\rangle + \langle\mathbf{Q}_2,\Theta^2-\mathbf{Z}_2\rangle + \frac{\rho}{2}\|\Theta^1-\Theta^2-(\mathbf{V}+\mathbf{W})\|_F^2\\ &+\ \frac{\rho}{2}\|\mathbf{V}-\mathbf{W}^T\|_F^2 + \frac{\rho}{2}\|\Theta^1-\mathbf{Z}_1\|_F^2 + \frac{\rho}{2}\|\Theta^2-\mathbf{Z}_2\|_F^2.\end{aligned}$$

- $\text{Expand}(\mathbf{A},\rho,n_k) = \underset{\Theta\in S^p_{++}}{\text{argmin}}\left\{-n_k\log\det(\Theta)+\rho\|\Theta-\mathbf{A}\|_F^2\right\}$
$= \frac{1}{2}\mathbf{U}\left(\mathbf{D}+\sqrt{\mathbf{D}^2+\frac{2n_k}{\rho}\mathbf{I}}\right)\mathbf{U}^T,$

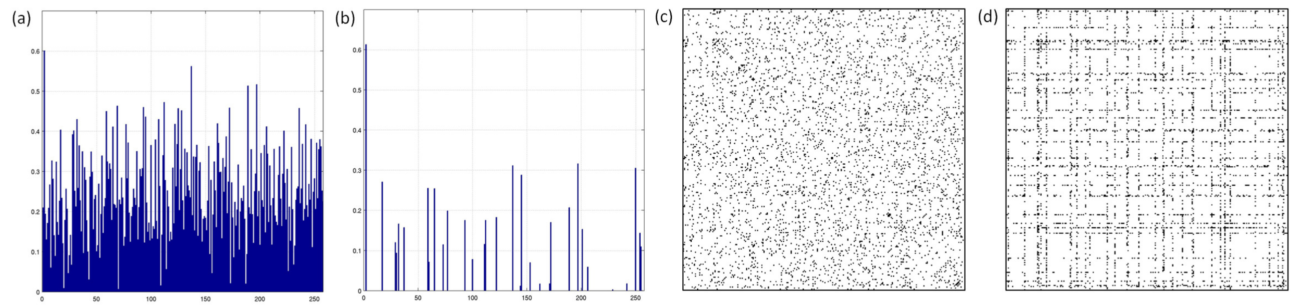
  where $\mathbf{U}\mathbf{D}\mathbf{U}^T$ is the eigenvalue decomposition of $\mathbf{A}$.

- Proximal operator to $\ell_1/\ell_q$ norm: $\mathcal{T}_q(\mathbf{A},\lambda) = \underset{\mathbf{X}}{\text{argmin}}\left\{\frac{1}{2}\|\mathbf{X}-\mathbf{A}\|_F^2 + \lambda\sum_{j=1}^p\|\mathbf{X}_j\|_q\right\}.$

**Algorithm 1**: ADMM algorithm for the PNJGL optimization problem

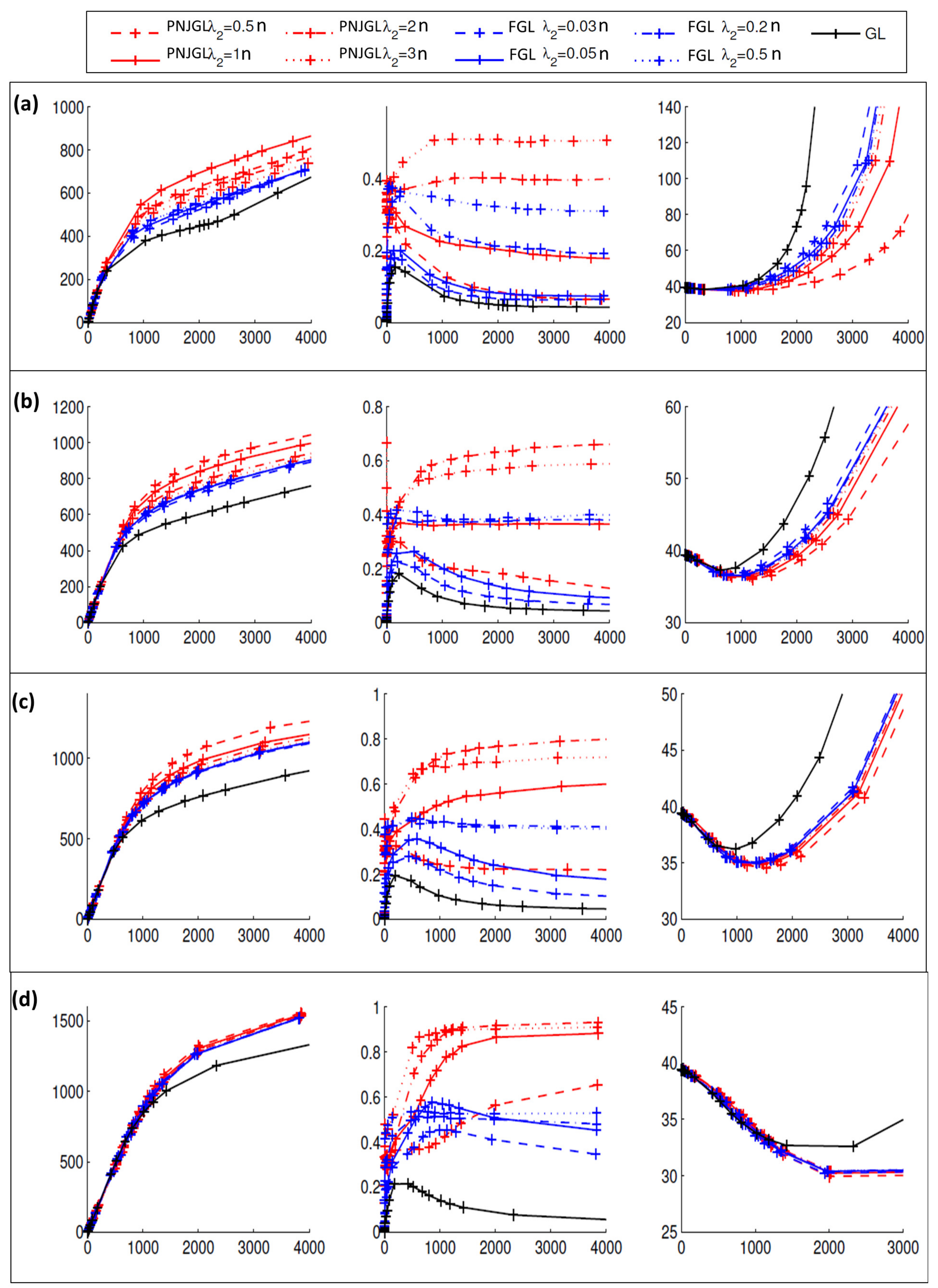
**input**: $\rho>0, \mu>1, t_{\max}>0, \epsilon>0$;
**for** *t = 1:*$t_{\max}$ **do**
  $\rho\leftarrow\mu\rho$;
  **while** *Not converged* **do**
    $\Theta^1\leftarrow\text{Expand}\left(\frac{1}{2}(\Theta^2+\mathbf{V}+\mathbf{W}+\mathbf{Z}_1)-\frac{1}{2\rho}(\mathbf{Q}_1+n_1\mathbf{S}_1+\mathbf{F}),\rho,n_1\right)$;
    $\Theta^2\leftarrow\text{Expand}\left(\frac{1}{2}(\Theta^1-(\mathbf{V}+\mathbf{W})+\mathbf{Z}_2)-\frac{1}{2\rho}(\mathbf{Q}_2+n_2\mathbf{S}_2-\mathbf{F}),\rho,n_2\right)$;
    $\mathbf{Z}_i\leftarrow\mathcal{T}_1\left(\Theta^i+\frac{\mathbf{Q}_i}{\rho},\frac{\lambda_1}{\rho}\right)$ for $i=1,2$;
    $\mathbf{V}\leftarrow\mathcal{T}_q\left(\frac{1}{2}(\mathbf{W}^T-\mathbf{W}+(\Theta^1-\Theta^2))+\frac{1}{2\rho}(\mathbf{F}-\mathbf{G}),\frac{\lambda_2}{2\rho}\right)$;
    $\mathbf{W}\leftarrow\frac{1}{2}(\mathbf{V}^T-\mathbf{V}+(\Theta^1-\Theta^2))+\frac{1}{2\rho}(\mathbf{F}+\mathbf{G}^T)$;
    $\mathbf{F}\leftarrow\mathbf{F}+\rho(\Theta^1-\Theta^2-(\mathbf{V}+\mathbf{W}))$;
    $\mathbf{G}\leftarrow\mathbf{G}+\rho(\mathbf{V}-\mathbf{W}^T)$;
    $\mathbf{Q}_i\leftarrow\mathbf{Q}_i+\rho(\Theta^i-\mathbf{Z}_i)$ for $i=1,2$

## NUMERICAL RESULTS - REAL DATA



PNJGL with $q=2$ and FGL were performed on the brain cancer data set corresponding to 258 genes in patients with Proneural and Mesenchymal subtypes. (a)-(b): $NP_j$ is plotted for each gene, based on (a) the FGL estimates and (b) the PNJGL estimates. (c)-(d): A heatmap of $\hat{\Theta}^1-\hat{\Theta}^2$ is shown for (c) FGL and (d) PNJGL; zero values are in white, and non-zero values in black.

## NUMERICAL RESULTS - SYNTHETIC DATA




Simulation study results for PNJGL with $q=2$, FGL, and the graphical lasso (GL), for (a) $n=10$, (b) $n=25$, (c) $n=50$, (d) $n=200$, when $p=100$. Within each panel, each line corresponds to a fixed value of $\lambda_2$ (for PNJGL with $q=2$ and for FGL). Each plot's $x$-axis denotes the number of edges estimated to be non-zero. The $y$-axes are as follows. *Left:* Number of edges *correctly* estimated to be non-zero. *Center:* Number of edges *correctly* estimated to differ across networks, *divided by* the number of edges estimated to differ across networks. *Right:* The Frobenius norm of the error in the estimated precision matrices.

## REFERENCES

[1] S.P. Boyd and N. Parikh and E. Chu and B. Peleato and J. Eckstein. *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in ML (2010).
[2] L. Jacob, G. Obozinski and J.P. Vert. *Group Lasso with Overlap and Graph Lasso*, Proceedings of the 26th International Conference on Machine Learning (2009)
[3] Yuan, M. and Lin, Y. *Model selection and estimation in the Gaussian graphical model*. Biometrika 94(10) (2007).
[4] P. Danaher and P. Wang and D. Witten. *The joint graphical lasso for inverse covariance estimation across multiple classes*. 2012.
[5] S. Mosci and S. Villa and A. Verri and L. Rosasco. *A primal-dual algorithm for group sparse regularization with overlapping groups*. Proceedings of Neural Information Processing Systems (2010).