A Completely Tuning-Free and Robust Approach to Sparse Precision Matrix Estimation

Chau Tran (UC Santa Barbara)

Joint work with Guo Yu (UC Santa Barbara)

Gaussian Graphical Models

- For an undirected graph $G = (\{X_1, \cdots, X_d\}, E)$, if $(i,j) \notin E$, then

$$X_i \perp\!\!\!\perp X_j | X_{\text{rest}}$$

- Suppose $(X_1, \cdots, X_d) \sim N_d(0, \Sigma)$, then conditional independence is encoded in $\Omega = \Sigma^{-1}$. In particular,

$$(i,j) \notin E \iff \Omega_{ij} = 0.$$

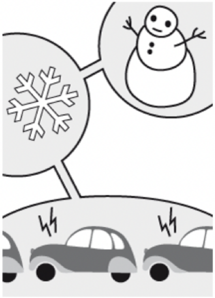


Figure: Conditional independence [1]

[1] Christophe Giraud. "Introduction to high-dimensional statistics". 2021.

Popular methods

- Penalized likelihood methods: Graphical Lasso [Yuan and Lin '07, Friedman et. al. '07, Banerjee et. al. '08], etc.

- Column-by-column estimation methods: Neighborhood Selection [Meinshausen and Bühlman '06], graphical Dantzig selector [Yuan '10], CLIME [Cai. et. al '11], SCIO [Liu and Luo '12], Scaled Lasso [Sun and Zhang '13], TIGER [Liu and Wang '17], etc.

- In practice, computationally intensive tuning procedures are used to choose proper regularization level.

Graphical Rank Lasso estimator

With the Rank Loss function [Wang et. al '20] for $1 \le j \le d$

$$Q_j(\beta) = [n(n-1)]^{-1} \sum_{k=1}^{n} \sum_{m \neq k} \times |(X_{kj} - X_{mj}) - (X_{k,-j} - X_{m,-j})\beta|$$

$$\hat{\beta}^{(j)} = \operatorname*{argmin}_{\beta \in \mathbb{R}^{d-1}} \{ Q_j(\beta) + \lambda_j \|\beta\|_1 \}$$

$$\hat{\sigma}_j^2 = n^{-1} \|X_{*,j} - X_{*,-j} \hat{\beta}^{(j)}\|_2^2$$

$$\hat{\Omega}_{jj} = 1/\hat{\sigma}_j^2, \qquad \hat{\Omega}_{-j,j} = -\hat{\Omega}_{jj} \hat{\beta}^{(j)}$$

$$\hat{\Omega} = \begin{pmatrix} \hat{\Omega}_{11} & \cdots & \hat{\Omega}_{1j} & \cdots & \hat{\Omega}_{1d} \\ \hat{\Omega}_{21} & \cdots & \hat{\Omega}_{2j} & \cdots & \hat{\Omega}_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\Omega}_{d1} & \cdots & \hat{\Omega}_{dj} & \cdots & \hat{\Omega}_{dd} \end{pmatrix}$$

Comparison of ℓ_1-regularized estimators

Method	$\|\cdot\|_1$ Minimax	$\|\cdot\|_F$ Minimax	Tuning-free
CLIME	$\times$	$\times$	$\times$
GLasso	$\surd$	$\surd$	$\times$
Yuan '10	$\surd$	$\times$	$\times$
SCIO	$\surd$	$\surd$	$\times$
TIGER	$\surd$	$\surd$	Asymptotic
gRankLasso	$\surd$	$\times$	Complete

Table: Theoretical properties of popular ℓ_1-regularized methods.

- GLasso and SCIO require the Irrepresentable Condition.
- TIGER and CLIME consider a larger matrix class when only conditional number of Ω is bounded.

Second-stage estimator with non-convex penalty

- A second-stage improvement with $\hat{\beta}^{(j)}$ as an initial value[1]:

$$\tilde{\beta}^{(j)} = \underset{\beta \in \mathbb{R}^{d-1}}{\operatorname{argmin}} \{ Q_j(\beta) + \sum_{i=1}^{d-1} p'_\eta(|\hat{\beta}_i^{(j)}|)|\beta_i| \}$$

SCAD, MCP, etc.

- Similarly, we can estimate $\tilde{\Omega}$ after $\tilde{\beta}^{(j)}, j = 1, ..., d$ are obtained:

$$\tilde{\sigma}_j^2 = n^{-1} \| X_{*,j} - X_{*,-j} \tilde{\beta}^{(j)} \|_2^2,$$
$$\tilde{\Omega}_{jj} = 1/\tilde{\sigma}_j^2, \qquad \tilde{\Omega}_{-j,j} = -\tilde{\Omega}_{jj} \tilde{\beta}^{(j)}.$$

- Under mild assumptions, the second-stage estimator $\tilde{\Omega}$ achieves the oracle property (i.e. as if it knows the true sparsity pattern of Ω) and faster convergence rates.

[1] Zou and Li '08, Wang et.al. '20

Simulation studies: General comparison

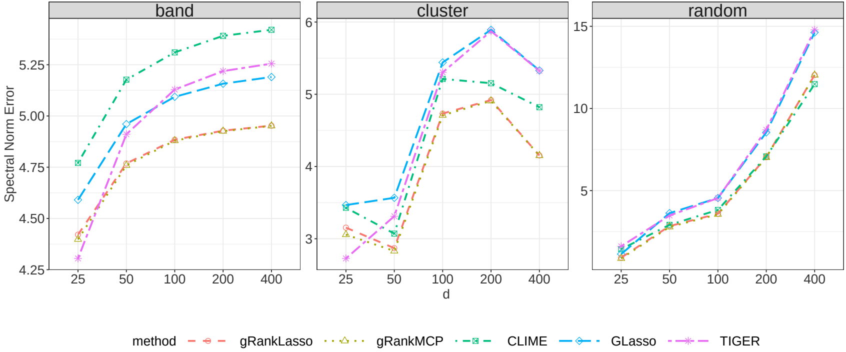


Figure: Average spectral error $\|\hat{\Omega} - \Omega\|_2$ using sparse graphs.

Simulation studies: Tuning-free methods

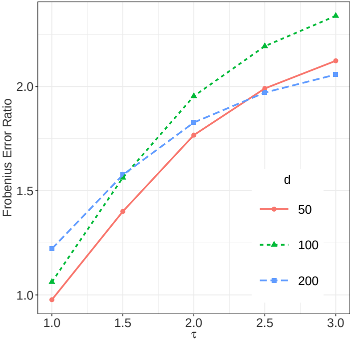


Figure: Average Frobenius error ratio $\frac{\|\hat{\Omega}_{\text{TIGER}} - \Omega\|_F}{\|\hat{\Omega}_{\text{gRankLasso}} - \Omega\|_F}$ using random graph.

Simulation studies: Benefits of bias reduction

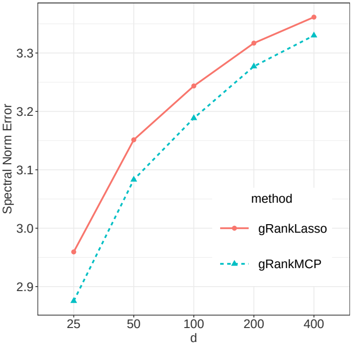


Figure: Average spectral error $\|\hat{\Omega} - \Omega\|_2$ using dense graph.

Simulation studies: Heavy-tailed setting

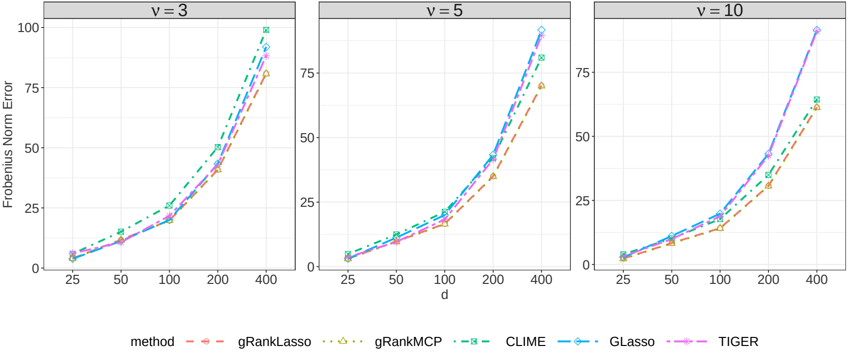


Figure: Average Frobenius error $\|\hat{\Omega} - \Omega\|_F$ using random graph and multivariate t-distribution $t_\nu(0, \Omega^{-1})$.

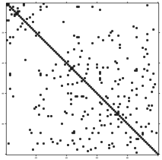
Real data example: Human gene network

Data: $d = 100$ most variable probes corresponding to different Illumina TargetID transcripts with $n = 60$.

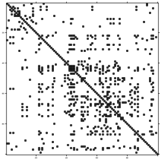
Goal: learn the significant associations among the chosen traits.

Method	Precision
GLasso	0.255
TIGER	0.368
gRankLasso	0.456
gRankMCP	0.518

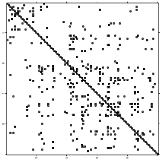
Table: Edge recovery results



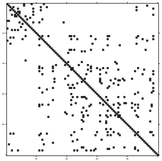
(a) Significant edges



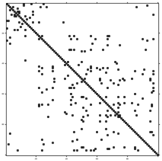
(b) GLasso



(c) TIGER



(d) gRankLasso



(e) gRankMCP