# Sorted L-One penalized estimator for graphical models and large concentration matrices

*Malgorzata Bogdan*, Giovanni Bonnacolto, Ivan Hejny, Philipp Kremer, Sandra Paterlini, Ricardo Riccobello, Piotr Sobczyk, Jonas Wallin

Lund University, University of Wroclaw, Kore University of Enna, EBS Universitaet fur Wirtschaft and Recht, Trento University, Wroclaw University of Science and Technology

19th of October, 2023

# Outline

- Introduction and Motivation
- Gaussian Graphical Model
- Graphical LASSO
- Sorted L-One Penalty and Graphical SLOPE
- Tlasso and Tslope
- Pattern recovery by Tlasso and Tslope

# Covariance and Correlation

$$Var(X) = \sigma_1^2, \quad Var(Y) = \sigma_2^2$$

# Covariance and Correlation

$$Var(X) = \sigma_1^2, \quad Var(Y) = \sigma_2^2$$

$\rho(X, Y)$ - correlation coefficient between $X$ and $Y$

# Covariance and Correlation

$$Var(X) = \sigma_1^2, \quad Var(Y) = \sigma_2^2$$

$\rho(X, Y)$ - correlation coefficient between $X$ and $Y$

$$Cov(X, Y) = \rho(X, Y)\sigma_1\sigma_2$$

# Covariance and Correlation

$$Var(X) = \sigma_1^2, \quad Var(Y) = \sigma_2^2$$

$\rho(X, Y)$ - correlation coefficient between $X$ and $Y$

$$Cov(X, Y) = \rho(X, Y)\sigma_1\sigma_2$$

$$Y = \alpha X + \varepsilon, \quad Var(\varepsilon) = \sigma^2$$

# Covariance and Correlation

$$Var(X) = \sigma_1^2, \quad Var(Y) = \sigma_2^2$$

$\rho(X, Y)$ - correlation coefficient between $X$ and $Y$

$$Cov(X, Y) = \rho(X, Y)\sigma_1\sigma_2$$

$$Y = \alpha X + \varepsilon, \quad Var(\varepsilon) = \sigma^2$$

$$Cov(X, Y) = Cov(\alpha X + \varepsilon, X) = \alpha Var(X) = \alpha\sigma_1^2$$

# Covariance and Correlation

$$Var(X) = \sigma_1^2, \quad Var(Y) = \sigma_2^2$$

$\rho(X, Y)$ - correlation coefficient between $X$ and $Y$

$$Cov(X, Y) = \rho(X, Y)\sigma_1\sigma_2$$

$$Y = \alpha X + \varepsilon, \quad Var(\varepsilon) = \sigma^2$$

$$Cov(X, Y) = Cov(\alpha X + \varepsilon, X) = \alpha Var(X) = \alpha\sigma_1^2$$

$$\rho(X, Y) = \frac{\alpha\sigma_1^2}{\sigma_1\sigma_2} = \frac{\alpha\sigma_1}{\sqrt{\alpha^2\sigma_1^2 + \sigma^2}}$$

# Covariance and Precision Matrices

$(X_1, \ldots, X_p)$ - p dimensional random vector

# Covariance and Precision Matrices

$(X_1, \ldots, X_p)$ - p dimensional random vector

$\Sigma_{p \times p}$ - covariance matrix

# Covariance and Precision Matrices

$(X_1, \ldots, X_p)$ - p dimensional random vector

$\Sigma_{p \times p}$ - covariance matrix

$$\Sigma_{i,j} = Cov(X_i, X_j)$$

# Covariance and Precision Matrices

$(X_1, \dots, X_p)$ - p dimensional random vector

$\Sigma_{p \times p}$ - covariance matrix

$$\Sigma_{i,j} = Cov(X_i, X_j)$$

$\Omega = \Sigma^{-1}$ - precision matrix

# Motivation: Global Minimum Variance Portfolio

$R = [R_1 \ \cdots \ R_K]'$ - a random vector of returns

$$\mu = [\mathbb{E}[R_1] \ \cdots \ \mathbb{E}[R_K]]'$$

$$\Sigma = \mathbb{E}\left[(R-\mu)(R-\mu)'\right].$$

$w \in R^K$ - portfolio weights

$$Var(w'R) = w'\Sigma w$$

$$w^\star = \arg\min_{w \in \mathbb{R}^K} w'\Sigma w \ \text{ subject to } w'\mathbf{1}_K = 1,$$

$$w^\star = \frac{\Sigma^{-1}\mathbf{1}_K}{\mathbf{1}_K'\Sigma^{-1}\mathbf{1}_K}$$

# Graphical Model

$X = (X_1, \ldots, X_p)$ - a random vector

# Graphical Model

$X = (X_1, \ldots, X_p)$ - a random vector
Graph: $G = (V, E)$
Vertices (V): indices (names) of components of $X$: $V = \{1, \ldots, p\}$
Edges (E): $(u, v) \notin V \iff X_u \perp\!\!\!\perp X_v | X_{V \setminus \{u,v\}}$

# Graphical Model

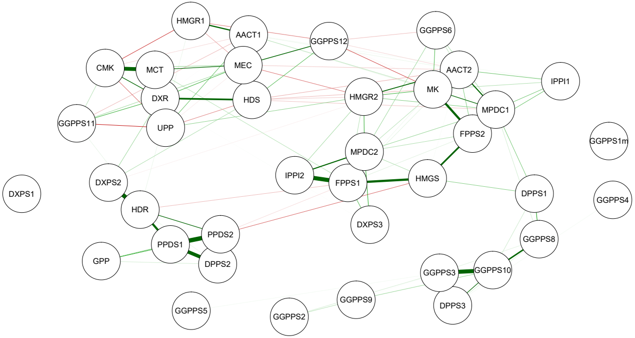
$X = (X_1, \ldots, X_p)$ - a random vector

Graph: $G = (V, E)$

Vertices (V): indices (names) of components of $X$: $V = \{1, \ldots, p\}$

Edges (E): $(u, v) \notin V \iff X_u \perp\!\!\!\perp X_v | X_{V\setminus\{u,v\}}$

Glasso



# Gaussian Graphical Model

$$X = (X_1, \ldots, X_p) \sim N(0, \Sigma), \quad \Omega = \Sigma^{-1} \quad \text{- precision matrix}$$

# Gaussian Graphical Model

$$X = (X_1, \ldots, X_p) \sim N(0, \Sigma), \;\; \Omega = \Sigma^{-1} \quad \text{- precision matrix}$$

$$X_i \perp\!\!\!\perp X_j | X_{V \setminus \{u,v\}} \Longleftrightarrow \Omega_{ij} = 0$$

# Gaussian Graphical Model

$$X = (X_1, \ldots, X_p) \sim N(0, \Sigma), \;\; \Omega = \Sigma^{-1} \;\; \text{- precision matrix}$$

$$X_i \perp\!\!\!\perp X_j | X_{V \setminus \{u,v\}} \iff \Omega_{ij} = 0$$

Example: for $j = 2, \ldots, p$ $X_j = X_1 + \varepsilon_j$, where $Var(\varepsilon_j) = 2$

# Gaussian Graphical Model

$$X = (X_1, \ldots, X_p) \sim N(0, \Sigma), \;\; \Omega = \Sigma^{-1} \;\; \text{- precision matrix}$$

$$X_i \perp\!\!\!\perp X_j | X_{V \setminus \{u,v\}} \Longleftrightarrow \Omega_{ij} = 0$$

Example: for $j = 2, \ldots, p$ $X_j = X_1 + \varepsilon_j$, where $Var(\varepsilon_j) = 2$

for $i \neq j$, $Cov(X_i, X_j) = \sigma_1^2 = 2$

# Gaussian Graphical Model

$$X = (X_1, \ldots, X_p) \sim N(0, \Sigma), \;\; \Omega = \Sigma^{-1} \quad \text{- precision matrix}$$

$$X_i \perp\!\!\!\perp X_j | X_{V \setminus \{u,v\}} \Longleftrightarrow \Omega_{ij} = 0$$

Example: for $j = 2, \ldots, p$ $X_j = X_1 + \varepsilon_j$, where $Var(\varepsilon_j) = 2$

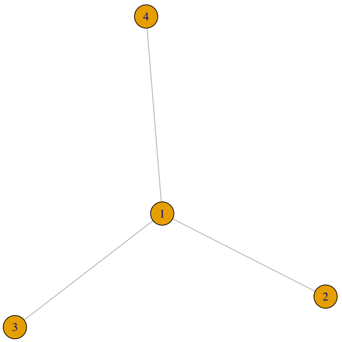
for $i \neq j$, $Cov(X_i, X_j) = \sigma_1^2 = 2$

for $j \geq 2$, $Var(X_j) = 40$

# Hub

$$\Sigma = \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 4 & 2 & 2 \\ 2 & 2 & 4 & 2 \\ 2 & 2 & 2 & 4 \end{bmatrix}, \Omega = \begin{bmatrix} 2 & -0.5 & -0.5. & -0.5 \\ -0.5 & 0.5 & 0 & 0 \\ -0.5 & 0 & 0.5 & 0 \\ -0.5 & 0 & 0 & 0.5 \end{bmatrix}$$

**Example 1**



# Minimizing the cross entropy function

$$S = \frac{1}{n}X'X\text{- sample covariance matrix}$$

# Minimizing the cross entropy function

$$S = \frac{1}{n}X'X \text{- sample covariance matrix}$$

$$ES = \Sigma$$

# Minimizing the cross entropy function

$$S = \frac{1}{n}X'X\text{- sample covariance matrix}$$

$$ES = \Sigma$$

$$\hat{\Omega}_1 = S^{-1}$$

# Minimizing the cross entropy function

$$S = \frac{1}{n} X'X \text{- sample covariance matrix}$$

$$ES = \Sigma$$

$$\hat{\Omega}_1 = S^{-1}$$

$$\text{Negative cross-entropy:} L(\Omega, X) = C + \frac{n}{2} \log \det \Omega - \frac{n}{2} tr(S\Omega).$$

# Large $p$ scenario

When $S$ is invertible, $\hat{\Omega}_{MLE} = S^{-1}$

# Large $p$ scenario

When $S$ is invertible, $\hat{\Omega}_{MLE} = S^{-1}$

$$ES^{-1} = \frac{n}{n-p-1}\Omega$$

# Large $p$ scenario

When $S$ is invertible, $\hat{\Omega}_{MLE} = S^{-1}$

$$ES^{-1} = \frac{n}{n-p-1}\Omega$$

$S$ is not invertible when $p > n$.

# Large $p$ scenario

When $S$ is invertible, $\hat{\Omega}_{MLE} = S^{-1}$

$$ES^{-1} = \frac{n}{n-p-1}\Omega$$

$S$ is not invertible when $p > n$.

$\hat{\Omega}_{MLE}$ is dense (inverse Wishart distribution)

# Large $p$ scenario

When $S$ is invertible, $\hat{\Omega}_{MLE} = S^{-1}$

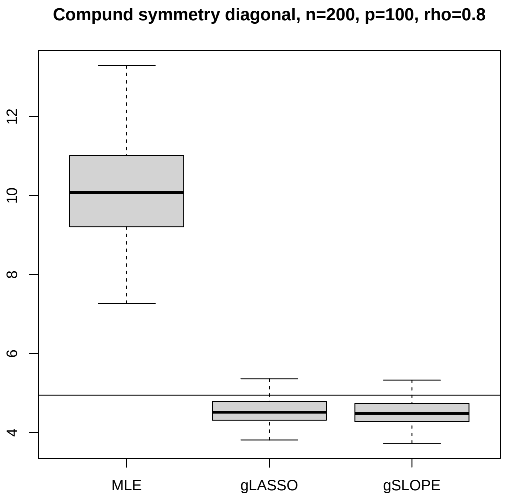
$$ES^{-1} = \frac{n}{n-p-1}\Omega$$

$S$ is not invertible when $p > n$.

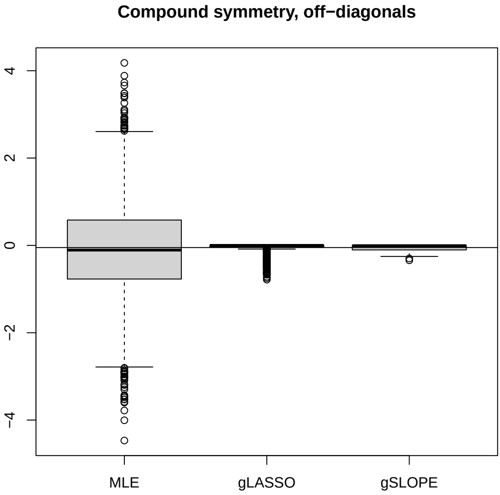
$\hat{\Omega}_{MLE}$ is dense (inverse Wishart distribution)

For large $p$, $MSE = E||\hat{\Omega} - \Omega||_F^2$ is large

# Simulation example (1)



# Simulation example (2)



# Graphical LASSO, Friedman et al. (2008)

$$L(\Omega, X) = C + \frac{n}{2} \log \det \Omega - \frac{n}{2} tr(S\Omega).$$

# Graphical LASSO, Friedman et al. (2008)

$$L(\Omega, X) = C + \frac{n}{2} \log\det\Omega - \frac{n}{2} tr(S\Omega).$$

$$\text{gLASSO:}\widehat{\Omega}_L = \arg\max_{\Omega \in Sym_+^p} \left[\log\det\Omega - tr(S\Omega) - \lambda ||\Omega||_1\right] \quad ,$$

$$||\Omega||_1 = \sum_{i<j} |\omega_{ij}| \quad .$$

Example: $n = 50$, $p = 30$, $\Sigma$ - block diagonal with 3 blocks of dimension $10 \times 10$, correlation within blocks $\rho = 0.8$

# Graphical LASSO, Friedman et al. (2008)

$$L(\Omega, X) = C + \frac{n}{2}\log\det\Omega - \frac{n}{2}tr(S\Omega).$$

$$\text{gLASSO:}\widehat{\Omega}_L = \arg\max_{\Omega\in Sym_+^p}\left[\log\det\Omega - tr(S\Omega) - \lambda||\Omega||_1\right] \quad ,$$

$$||\Omega||_1 = \sum_{i<j}|\omega_{ij}| \quad .$$

Example: $n = 50,\ p = 30$, $\Sigma$ - block diagonal with 3 blocks of dimension $10 \times 10$, correlation within blocks $\rho = 0.8$
$||\Omega - \hat{\Omega}_{MLE}||_F = 10800$

# Graphical LASSO, Friedman et al. (2008)

$$L(\Omega, X) = C + \frac{n}{2}\log\det\Omega - \frac{n}{2}tr(S\Omega).$$

$$\text{gLASSO:}\widehat{\Omega}_L = \arg\max_{\Omega\in Sym_+^p}\left[\log\det\Omega - tr(S\Omega) - \lambda||\Omega||_1\right] \quad ,$$

$$||\Omega||_1 = \sum_{i<j}|\omega_{ij}| \quad .$$

Example:$n = 50$, $p = 30$, $\Sigma$ - block diagonal with 3 blocks of dimension $10 \times 10$, correlation within blocks $\rho = 0.8$

$||\Omega - \hat{\Omega}_{MLE}||_F = 10800$

Banerjee and d'Aspremont (2008), FWER control for block diagonal matrices with standardized entries:

$$\lambda_\alpha^{Banerjee} = \frac{t_{n-2}\left(1 - \frac{\alpha}{2p^2}\right)}{\sqrt{n - 2 + t_{n-2}^2\left(1 - \frac{\alpha}{2p^2}\right)}} \quad , \tag{1}$$

# Graphical LASSO, Friedman et al. (2008)

$$L(\Omega, X) = C + \frac{n}{2}\log\det\Omega - \frac{n}{2}tr(S\Omega).$$

gLASSO:$\widehat{\Omega}_L = \arg\max_{\Omega \in Sym_+^p}\left[\log\det\Omega - tr(S\Omega) - \lambda||\Omega||_1\right]$ ,

$$||\Omega||_1 = \sum_{i<j}|\omega_{ij}| \quad .$$

Example:$n = 50,\ p = 30$, $\Sigma$ - block diagonal with 3 blocks of dimension $10 \times 10$, correlation within blocks $\rho = 0.8$
$||\Omega - \hat{\Omega}_{MLE}||_F = 10800$
Banerjee and d'Aspremont (2008), FWER control for block diagonal matrices with standardized entries:

$$\lambda_\alpha^{Banerjee} = \frac{t_{n-2}\left(1 - \frac{\alpha}{2p^2}\right)}{\sqrt{n - 2 + t_{n-2}^2\left(1 - \frac{\alpha}{2p^2}\right)}} \quad , \qquad (1)$$

Perfect graph discovery, $||\Omega - \hat{\Omega}_{gLASSO}||_F = 468$

# Sorted L-One PEnalization (SLOPE)

- High dimensional multiple regression:

$$Y = X\beta + \varepsilon, \;\; \varepsilon \sim N(0, I) \;\; .$$

- SLOPE (B., van den Berg, Su, Candès, arxiv 2013, B.,van den Berg, Sabatti, Su, Candès, AoAS, 2015) penalizes larger coefficients more stringently

$$\hat{\beta}_{SLOPE} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \, \frac{1}{2}|Y - X\beta|^2 + \sum_{j=1}^{p} \lambda_j |\beta|_{(j)},$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and $|\beta|_{(1)} \geq |\beta|_{(2)} \geq \cdots \geq |\beta|_{(p)}$.

# False discovery rate (FDR) control

- Let $\widetilde{\beta}$ be estimate of $\beta$
- We define:
  - the number of all discoveries, $R := \left|\left\{i : \ \widetilde{\beta}_i \neq 0\right\}\right|$
  - the number of false discoveries, $V := \left|\left\{i : \ \beta_i = 0, \quad \widetilde{\beta}_i \neq 0\right\}\right|$
  - false discovery rate - expected proportion of false discoveries among all discoveries

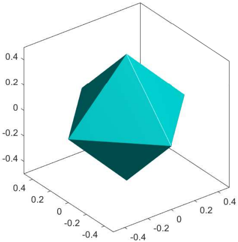
$$FDR := \mathbb{E}\left[\frac{V}{\max\{R, 1\}}\right]$$

Theorem (B,van den Berg, Su and Candès (2013))
*When $X^T X = I$ SLOPE with*

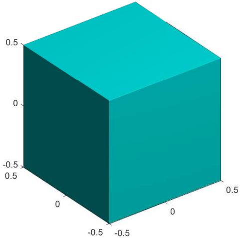
$$\lambda_i^{BH} := \sigma\Phi^{-1}\left(1 - i \cdot \frac{q}{2p}\right)$$

*controls FDR at the level $q\frac{p_0}{p}$ .*

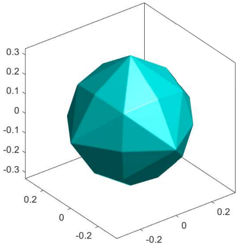
# Unit balls for different SLOPE sequences by D.Brzyski




((a)) (2,2,2)




((b)) (2,0,0)




((c)) (3,2,1)

# Clustering properties of SLOPE

- Schneider and Tardivel (JMLR, 2022) - definition of the SLOPE pattern.
- Skalski, B., Graczyk, Kołodziejek, Tardivel, Wilczyński, arxiv, 2022 - conditions under which SLOPE can identify the pattern.

# gSLOPE

Vectorize the upper triangle of $\Omega$, creating a vector

$$\omega = [\omega_1 \cdots \omega_m]$$

$$J_\lambda(\Omega) = \sum_{i=1}^{m} \lambda_i \, |\omega|_{(i)}$$

$$\widehat{\Omega}_{Gslope} = \arg\max_{\Omega} \, \{\log|\Omega| - (\Omega S) - J_\lambda(\Omega)\}\,.$$

# FWER control by Gslope

$$\lambda_k^{\text{Holm}} = \frac{t_{n-2}(1 - \frac{\alpha}{2(m+1-k)})}{\sqrt{n - 2 + t_{n-2}^2(1 - \frac{\alpha}{2(m+1-k)})}}$$

$C_l$ - the connectivity component of the $l^{th}$ node

Theorem

*Under mild regularity conditions Gslope with $\lambda^{Holm}$ satisfies*

$$P\left(\forall k \in \{1, \ldots, p\} : \hat{C}_k^{\lambda} \subset C_k\right) \geq 1 - \alpha \ .$$

# Distant FDR control

$$\lambda_k^{\mathrm{BH}} = \frac{t_{n-2}(1 - \frac{\alpha k}{2m})}{\sqrt{n - 2 + t_{n-2}^2(1 - \frac{\alpha k}{2m})}}$$

# Distant FDR control

$$\lambda_k^{\text{BH}} = \frac{t_{n-2}(1 - \frac{\alpha k}{2m})}{\sqrt{n - 2 + t_{n-2}^2(1 - \frac{\alpha k}{2m})}}$$

$V_d$: the number of falsely identified edges between different connectivity components

# Distant FDR control

$$\lambda_k^{\mathrm{BH}} = \frac{t_{n-2}(1 - \frac{\alpha k}{2m})}{\sqrt{n - 2 + t_{n-2}^2(1 - \frac{\alpha k}{2m})}}$$

$V_d$: the number of falsely identified edges between different connectivity components

$R$ : the total number of identified edges

# Distant FDR control

$$\lambda_k^{\text{BH}} = \frac{t_{n-2}(1 - \frac{\alpha k}{2m})}{\sqrt{n - 2 + t_{n-2}^2(1 - \frac{\alpha k}{2m})}}$$

$V_d$: the number of falsely identified edges between different connectivity components

$R$ : the total number of identified edges

$$dFDR = E\left(\frac{V_d}{\max(R,1)}\right) \quad .$$

# Distant FDR control

$$\lambda_k^{\mathrm{BH}} = \frac{t_{n-2}(1-\frac{\alpha k}{2m})}{\sqrt{n-2+t_{n-2}^2(1-\frac{\alpha k}{2m})}}$$

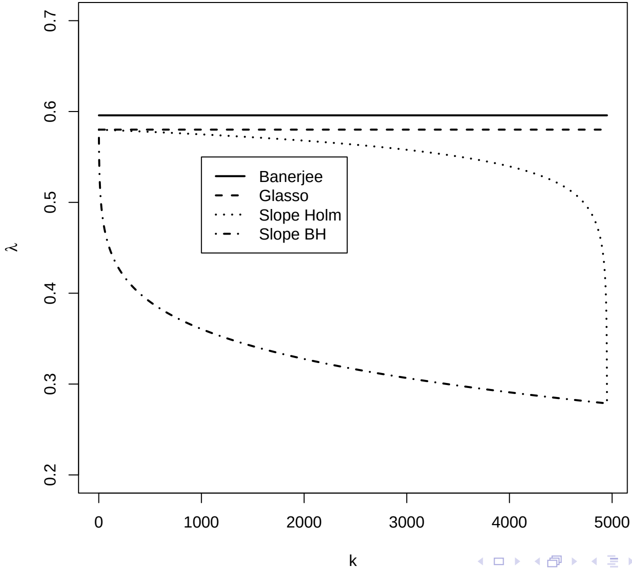
$V_d$: the number of falsely identified edges between different connectivity components

$R$ : the total number of identified edges

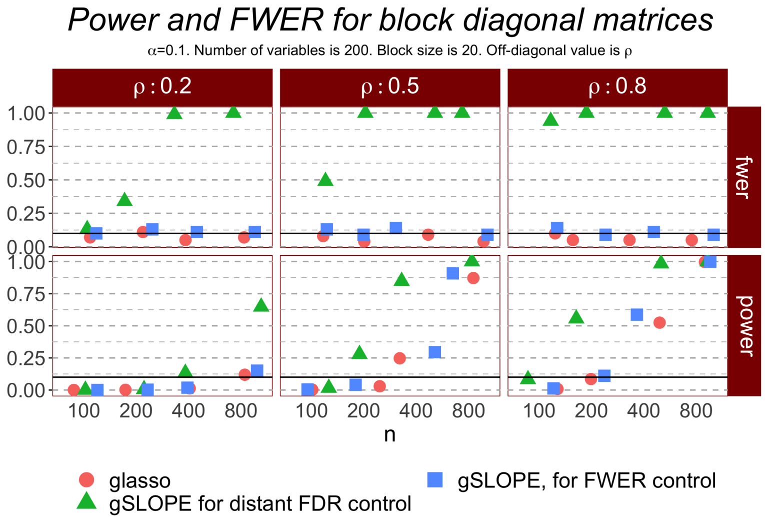
$$dFDR = E\left(\frac{V_d}{\max(R,1)}\right) \quad .$$

**Conjecture:** gSLOPE controls *dFDR* in cases when the BH procedure on the sample correlation coefficients controls FDR

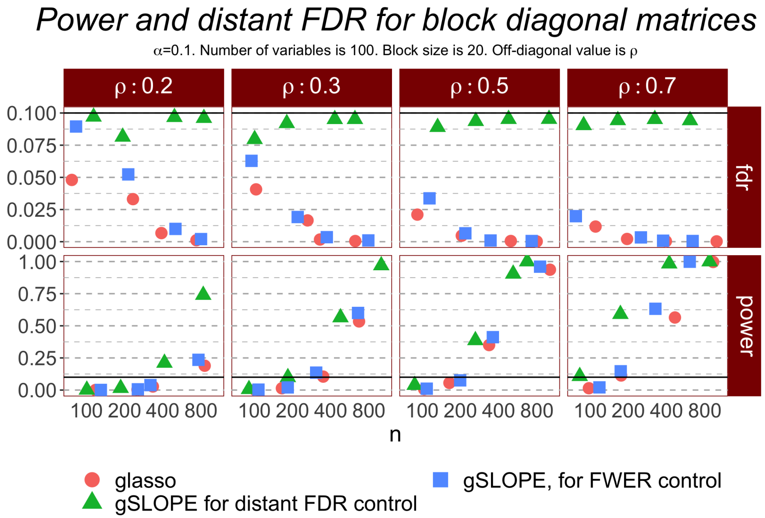
# Different tuning sequences, $p = 100$ $(m = 4950)$, $n = 50$



# FWER control



# Distant FDR Control



# TSlope, Riccobello

Extension of TLasso [Finegold and Drton (2011)] for the heavy tailed t-distributed variables. Important e.g. for financial applications.

# TSlope, Riccobello

Extension of TLasso [Finegold and Drton (2011)] for the heavy tailed t-distributed variables. Important e.g. for financial applications.

$$\boldsymbol{W} \sim \mathcal{N}_p(\mathbf{0}, \Psi), \tau \sim \Gamma(\nu/2, \nu/2)$$

# TSlope, Riccobello

Extension of TLasso [Finegold and Drton (2011)] for the heavy tailed t-distributed variables. Important e.g. for financial applications.

$$\boldsymbol{W} \sim \mathcal{N}_p(\mathbf{0}, \Psi), \tau \sim \Gamma(\nu/2, \nu/2)$$

$$\boldsymbol{X} = \mu + \frac{\boldsymbol{W}}{\sqrt{\tau}} \sim t_p(\mu, \Psi, \nu).$$

# TSlope, Riccobello

Extension of TLasso [Finegold and Drton (2011)] for the heavy tailed t-distributed variables. Important e.g. for financial applications.

$$\boldsymbol{W} \sim \mathcal{N}_p(\mathbf{0}, \Psi), \tau \sim \Gamma(\nu/2, \nu/2)$$

$$\boldsymbol{X} = \mu + \frac{\boldsymbol{W}}{\sqrt{\tau}} \sim t_p(\mu, \Psi, \nu).$$

$$f_X\left(x \mid \tau\right) \sim \mathcal{N}_p\left(\mu, \Psi/\tau\right).$$

# TSlope, Riccobello

Extension of TLasso [Finegold and Drton (2011)] for the heavy tailed t-distributed variables. Important e.g. for financial applications.

$$\boldsymbol{W} \sim \mathcal{N}_p(\mathbf{0}, \Psi), \tau \sim \Gamma(\nu/2, \nu/2)$$

$$\boldsymbol{X} = \mu + \frac{\boldsymbol{W}}{\sqrt{\tau}} \sim t_p(\mu, \Psi, \nu).$$

$$f_X(x \mid \tau) \sim \mathcal{N}_p(\mu, \Psi/\tau).$$

$$\Omega = \Psi^{-1}$$

# EM algorithm

E-step:

$$E(\tau|X) = \frac{\nu + p}{\nu + \delta_X(\mu, \Omega)}$$

$$\text{with} \ \ \delta_X(\mu, \Omega) = (X - \mu)'\Omega(X - \mu)$$

$$\tau_i^{(t+1)} = \frac{\nu + p}{\nu + \delta_{X_i}(\mu^{(t)}, \Omega^{(t)})}$$

# EM algorithm

E-step:

$$E(\tau|X) = \frac{\nu + p}{\nu + \delta_X(\mu, \Omega)}$$

$$\text{with} \quad \delta_X(\mu, \Omega) = (X - \mu)'\Omega(X - \mu)$$

$$\tau_i^{(t+1)} = \frac{\nu + p}{\nu + \delta_{X_i}(\mu^{(t)}, \Omega^{(t)})}$$

M-step:

$$\mu^{(t+1)} = \frac{\sum_{i=1}^{n} \tau_i^{(t+1)} X_i}{\sum_{i=1}^{n} \tau_i^{(t+1)}}$$

$$S^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} \tau_i^{(t+1)} [X_i - \mu^{(t+1)}][X_i - \mu^{(t+1)}]'$$

$$\Omega^{(t+1)} = \arg\max_{\Omega} \left\{ \log|\Omega| - \left(\Omega S^{(t+1)}\right) - J_\lambda(\Omega) \right\}.$$

# Application for GMV portfolio

"Sparse Graphical Modelling for Minimum Variance Portfolios", R Riccobello, G Bonaccolto, PJ Kremer, S Paterlini, M Bogdan, Available at SSRN 4383314.

Tslope outperforms other methods for the Standard & Poor's 100 and 500 indices (many assets and heavy tailed return distribution)

# General eliptical distributions

$$x = \Sigma^{1/2} U R$$

Direction $U \sim Unif \mathbb{S}^{p-1} \subset \mathbb{R}^p$ is independent from the radius $R \geq 0$.

$$\mathbb{E}[x] = 0$$

$$Cov(X) = \mathbb{E}[R^2] p^{-1} \Sigma$$

We assume that $Cov(x) = \Sigma$, or equivalently $\mathbb{E}[R^2] = p$.

# gSLOPE

Let $\Lambda \in \mathbb{R}^{p(p-1)/2}$, $\Lambda_{21} \geq \Lambda_{31} \geq .. \geq \Lambda_{pp} \geq 0$, be the penalty vector for the lower diagonal entries and $J_\Lambda(\Psi) := \sum_{i>j} \Lambda_{ij}|\Psi|_{(ij)}$.

For the diagonal entries, let $\lambda \in \mathbb{R}^p$ such that $\lambda_1 \geq .. \geq \lambda_p \geq 0$ and set $J_\lambda(\Psi) = \sum_{i=1}^{p} \lambda_i |\Psi|_{(ii)}$.

# gSLOPE pattern convergence

### Theorem

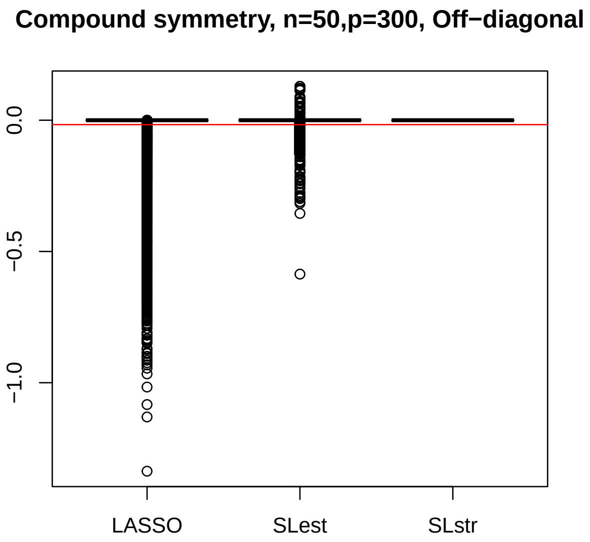
*Assume $x$ follows a centered elliptical distribution with $Cov(x) = \Sigma = \Theta_0^{-1}$. Let $\Lambda^n/\sqrt{n} \to \Lambda$ and $\lambda^n/\sqrt{n} \to \lambda$. Then $\sqrt{n}(\widehat{\Psi}_n - \Psi_0)$ converges weakly and in pattern to the minimizer of $V : \mathbb{R}^{p(p+1)/2} \to \mathbb{R}$;*

$$V(u) = \frac{1}{2} u^T \tilde{C} u - u^T \tilde{W} + J'_{\Lambda,\lambda}(\Psi_0; u),$$

*where $\tilde{W} \sim \mathcal{N}(0, C_\triangle)$, and;*

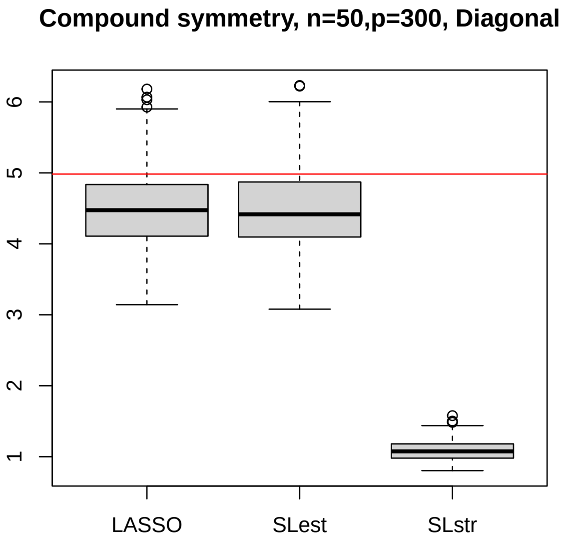
$$\tilde{C} = \frac{1}{2} D_p^T (\Theta_0^{-1} \otimes \Theta_0^{-1}) D_p,$$
$$C_\triangle = D_p^T Cov(vec(xx^T)) D_p/4 \quad ,$$

*where $D_p$ is the duplication matrix satisfying $D_p vech(A) = vec(A)$.*

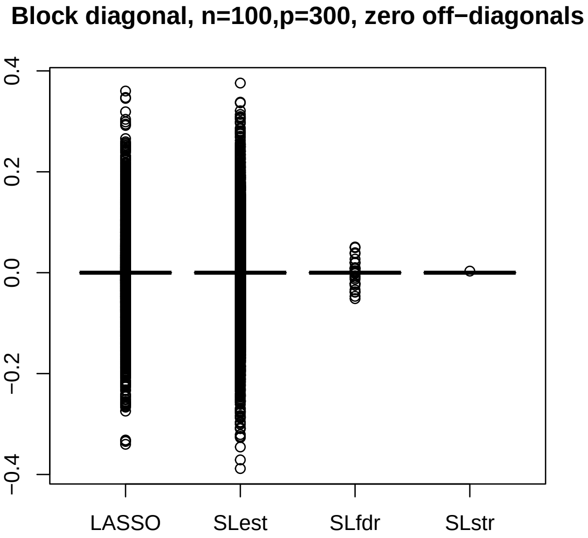


Compound symmetry, n=50,p=300, Off−diagonal
0.0
−0.5
−1.0
LASSO
SLest
SLstr

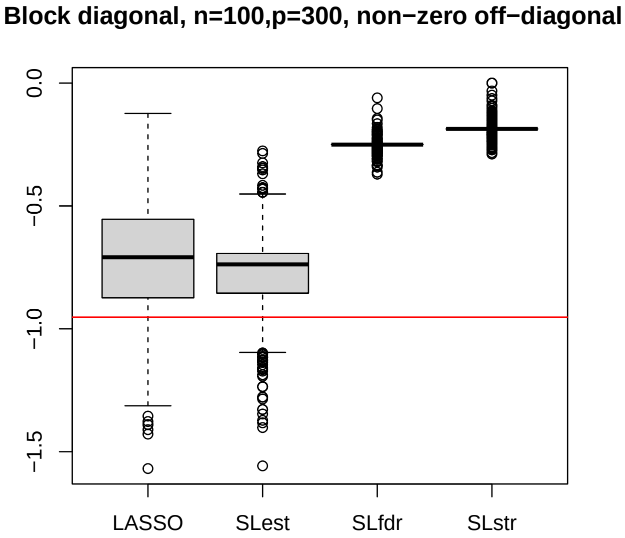
# Compound symmetry (2)



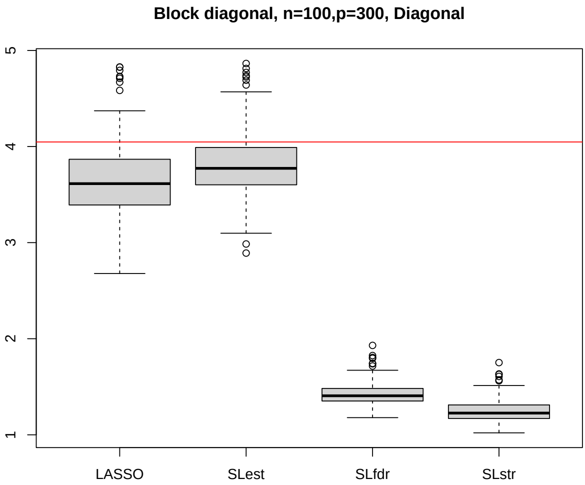
# Block diagonal (1)



# Block diagonal (2)



# Block diagonal (3)



# Conclusions

GSlope and TSlope outperform the respective versions of LASSO if the precision matrix is highly structure like e.g. in VAR (vector autoregressive) models

# Conclusions

GSlope and TSlope outperform the respective versions of LASSO if the precision matrix is highly structure like e.g. in VAR (vector autoregressive) models

Different sequences of tuning parameters need to be used for optimizing the estimation accuracy and identifying the structure

# Conclusions

GSlope and TSlope outperform the respective versions of LASSO if the precision matrix is highly structure like e.g. in VAR (vector autoregressive) models

Different sequences of tuning parameters need to be used for optimizing the estimation accuracy and identifying the structure

This can be addressed by adaptive gSLOPE versions - ongoing research.

# Conclusions

GSlope and TSlope outperform the respective versions of LASSO if the precision matrix is highly structure like e.g. in VAR (vector autoregressive) models

Different sequences of tuning parameters need to be used for optimizing the estimation accuracy and identifying the structure

This can be addressed by adaptive gSLOPE versions - ongoing research.

Bias on the diagonal can be reduced by penalizing the partial correlation matrix instead of the precision matrix - ongoing research.