Large precision matrix estimation via pairwise tilting

Na Huang† and Piotr Fryzlewicz‡

Department of Statistics, London School of Economics, UK

E-mail: n.huang1@lse.ac.uk†, p.fryzlewicz@lse.ac.uk‡

Abstract

We propose a $tilting$-based method to estimate the precision matrix of a p-dimensional random variable, $\mathbf{X}$, when p is possibly much larger than the sample size n. Each 2×2 block indexed by (i, j) of the precision matrix can be estimated by the inversion of the pairwise sample conditional covariance matrix of X_i and X_j controlling for all the other variables. However, in the high dimensional setting, including too many or irrelevant controlling variables may distort the results. To determine the controlling subsets in high dimensional scenarios, the proposed method applies the $tilting$ technique to measure the contribution of each remaining variable to the variance-covariance matrix of X_i and X_j, and only puts the (hopefully) highly relevant remaining variables into the controlling subsets. We illustrate conditions under which it can successfully distinguish the highly relevant remaining variables from the rest. The simulation results will be presented under different scenarios for the underlying precision matrix. Comparison with other competing methods will also be given.

Introduction

- Importance and Challenging

Estimating covariance matrix Σ or precision matrix (inverse of covariance matrix) Σ^{-1} has always been an important part of multivariate analysis.

- Financial risk management, Linear discriminant analysis (LDA), Principal component analysis (PCA)
- Graphic networks, large portfolio selection, and so on

In high-dimensional cases, where the dimension p is of the same order as the sample size n, or even much larger than n, estimating covariance or precision matrix is much more challenging.

- When $p > n$, noninvertability of sample covariance matrix
- Even p is of the same order as n, parameters needed to be estimated is $p(p-1)/2$ can be much larger than n.

- Existing Methods

Most of the existing methods are based on assumption of the structure of underlying precision matrix.

- Assumption of ordering among variables: banding or tapering (Bickel and Levina, 2008)
- Assumption of Sparsity structure: Lasso (Levina *et al.*, 2008) or Thresholding (Bickel and Levina, 2008)
- Assumption of Factor model (Fan *et al.*, 2008)

Methods

- Pairwise Precision Matrix

- Let $\boldsymbol{X} = (X_1, X_2, \cdots, X_p)^T$ be a p-dimensional time series. For a given pair of i, j $(1 \le i < j \le p)$, we denote Pairwise Precision Matrix as below,
  - $p \times p$ precision matrix

$$\Sigma^{-1} = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \bullet & \cdot & \bullet \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \bullet & \cdot & \bullet \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix} \begin{matrix} \\ \leftarrow i \\ \\ \leftarrow j \\ \\ \end{matrix} \xrightarrow[inverse]{} \quad \Sigma \text{ covariance matrix}$$

  - 2×2 pairwise precision matrix

$$\Sigma^{-1}_{\underline{i,j}} = \begin{pmatrix} \bullet & \bullet \\ \bullet & \bullet \end{pmatrix} \xrightarrow[inverse]{} cov((X_i, X_j)^T | \Omega^{\{-i,j\}})$$

pairwise conditional covariance matrix

- It can be proved that given i and j, the pairwise precision matrix is the inverse of pairwise conditional covariance matrix.

- Pairwise Conditional Covariance Matrix

- Given i and j, regressing X_i and X_j on all the other variables $\Omega^{\{-i,j\}}$, we have

$$X_i = \sum_{k \neq i,j} \beta_{i,k} X_k + \epsilon_i \qquad (1)$$

$$X_j = \sum_{k \neq i,j} \beta_{j,k} X_k + \epsilon_j \qquad (2)$$

- the pairwise conditional covariance matrix

$$cov((X_i, X_j)^T | \Omega^{\{-i,j\}}) = cov((\epsilon_i, \epsilon_j)^T) = \begin{pmatrix} var(\epsilon_i) & cov(\epsilon_i, \epsilon_j) \\ cov(\epsilon_i, \epsilon_j) & var(\epsilon_j) \end{pmatrix} \qquad (3)$$

Lemma 1 *The inverse of covariance of $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{ij}}$ is a consistent estimator of the pairwise precision matrix, $\Sigma^{-1}_{\underline{i,j}}$, where $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{ij}} = (\hat{\epsilon}_i, \hat{\epsilon}_j)^T$ is OLS estimator of $\boldsymbol{\epsilon}_{\boldsymbol{ij}}$.*

- Pairwise Tilting

- In high dimensional cases, $cov((X_i, X_j)^T | \Omega^{\{-i,j\}})$ is not well-defined, as the controlling subset $\Omega^{\{-i,j\}}$ contains too many controlling variables.
- We want to replace $\Omega^{\{-i,j\}}$ by $\mathcal{C}_{ij}$, which only contains the (hopefully) highly relevant remaining variables.
- It becomes a variable selection problem in a pair of linear regression models.
- We assume, for each pair i, j, that we can decompose X_i and X_j as below,

$$X_i = \sum_{b \in \mathcal{B}} \beta_{i,b} X_b + \underline{\sum_{e_i \in \mathcal{E}_i} \beta_{i,e_i} X_{e_i}} + \underline{\underline{\sum_{e_j \in \mathcal{E}_j} \beta_{i,e_j} X_{e_j}}} + \underline{\underline{\sum_{n \in \mathcal{N}} \beta_{i,n} X_n}} + \epsilon_i \qquad (4)$$

$$X_j = \sum_{b \in \mathcal{B}} \beta_{j,b} X_b + \underline{\underline{\sum_{e_i \in \mathcal{E}_i} \beta_{j,e_i} X_{e_i}}} + \underline{\sum_{e_j \in \mathcal{E}_j} \beta_{j,e_j} X_{e_j}} + \underline{\underline{\sum_{n \in \mathcal{N}} \beta_{j,n} X_n}} + \epsilon_j \qquad (5)$$

where

$\mathcal{B} = \{b : |\beta_{i,b}| > C_1, \text{and } |\beta_{j,b}| > C_1\}, \ 0 \le C_1 \le 1$

$\mathcal{E}_i = \{b : |\beta_{i,e_i}| > C_1, \text{and } |\beta_{j,e_i}| < C_2\}, \ 0 \le C_2 \le C_1 \le 1$

$\mathcal{E}_j = \{b : |\beta_{i,e_j}| < C_2, \text{and } |\beta_{j,e_j}| > C_1\}, \ 0 \le C_2 \le C_1 \le 1$

$\mathcal{N} = \{b : |\beta_{i,n}| < C_2, \text{and } |\beta_{i,n}| < C_2\}, \ 0 \le C_2 \le 1$

C_1 is lower bound of regression coefficients for relevant variables;
C_2 is upper bound of regression coefficients for irrelevant variables.

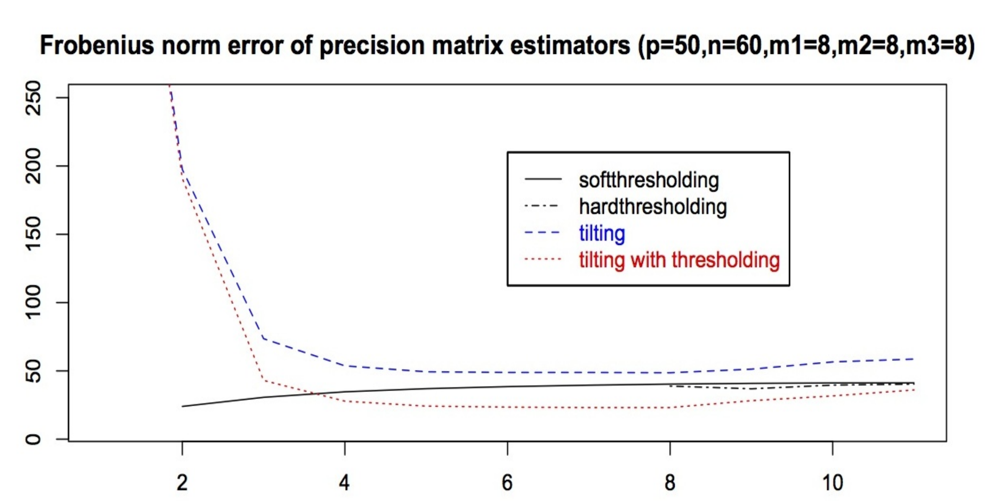
- 3 Types of Pairwise Tilting

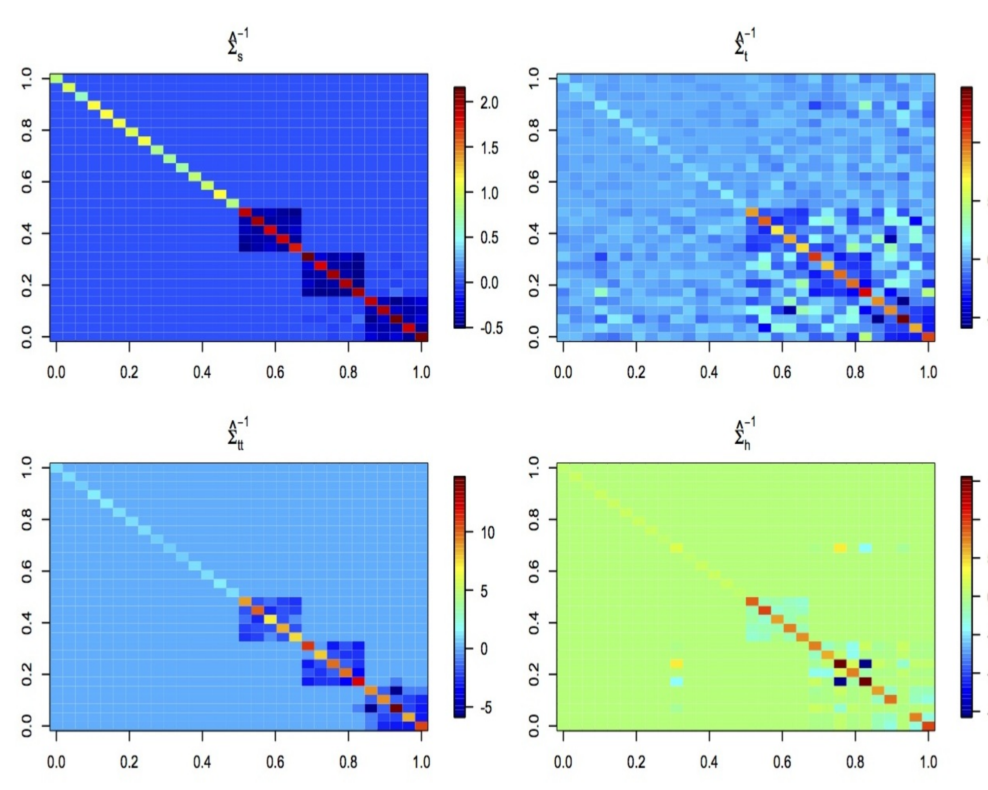
- Single Tilting:
  $\mathcal{C}_{ij} = \{X_k : |corr(X_k, X_i)| > \pi \ \text{ or } \ |corr(X_k, X_j)| > \pi)\}$
- Double Tilting:
  $\mathcal{C}_{ij} = \{X_k : |corr(X_k, X_i)| > \pi \ \text{ and } \ |corr(X_k, X_j)| > \pi)\}$
- Separate Tilting:
  $\mathcal{C}_i = \{X_{k_i} : |corr(X_{k_i}, X_i)| > \pi\}$,
  $\mathcal{C}_j = \{X_{k_j} : |corr(X_{k_j}, X_j)| > \pi\}$

Where π is a chosen thresholding.

Simulation study

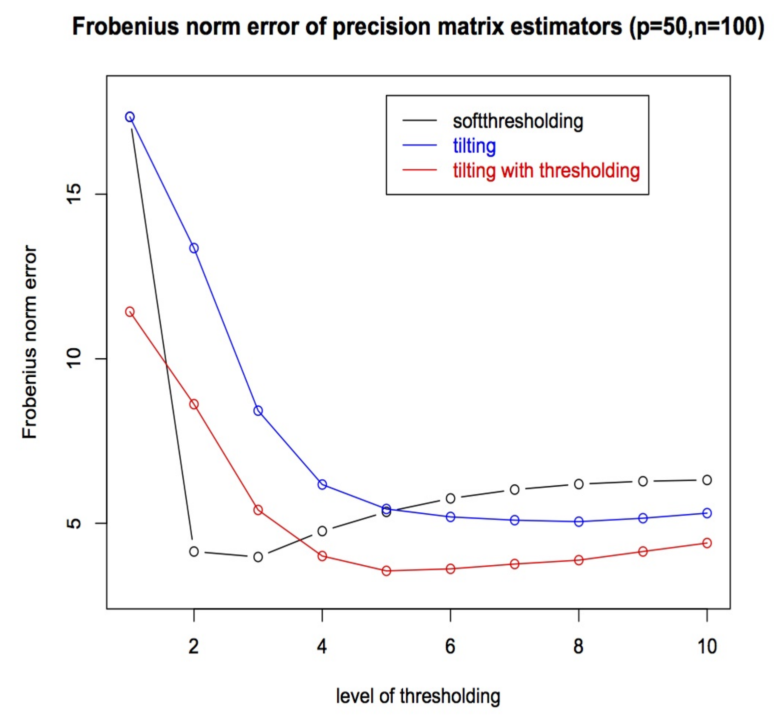
- Scenario 1: Block precision matrix

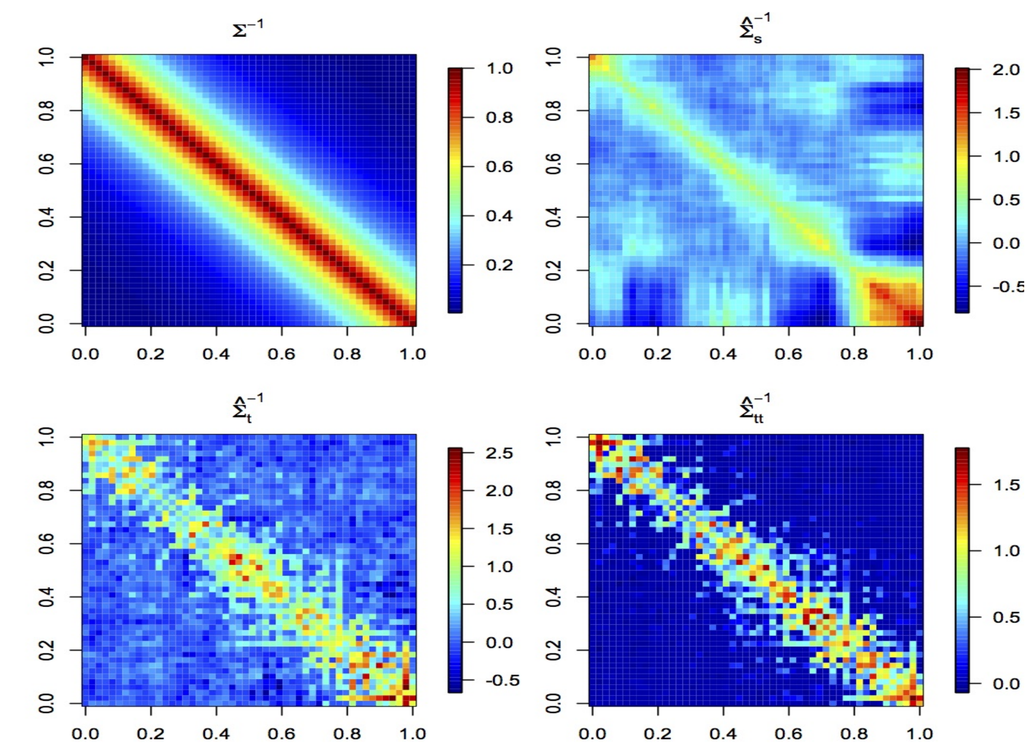




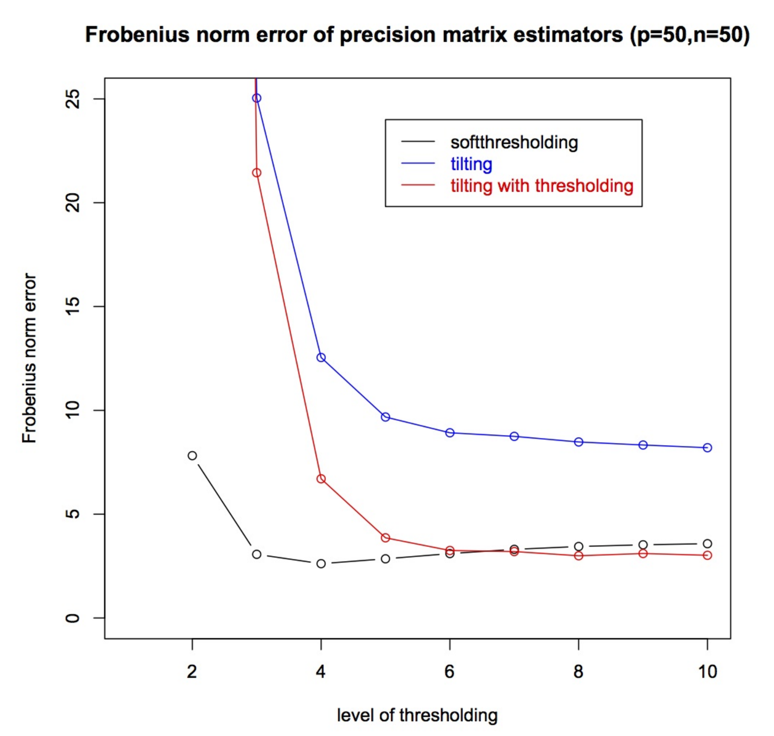
Where m1, m2, m3 are the dimensions of blocks.

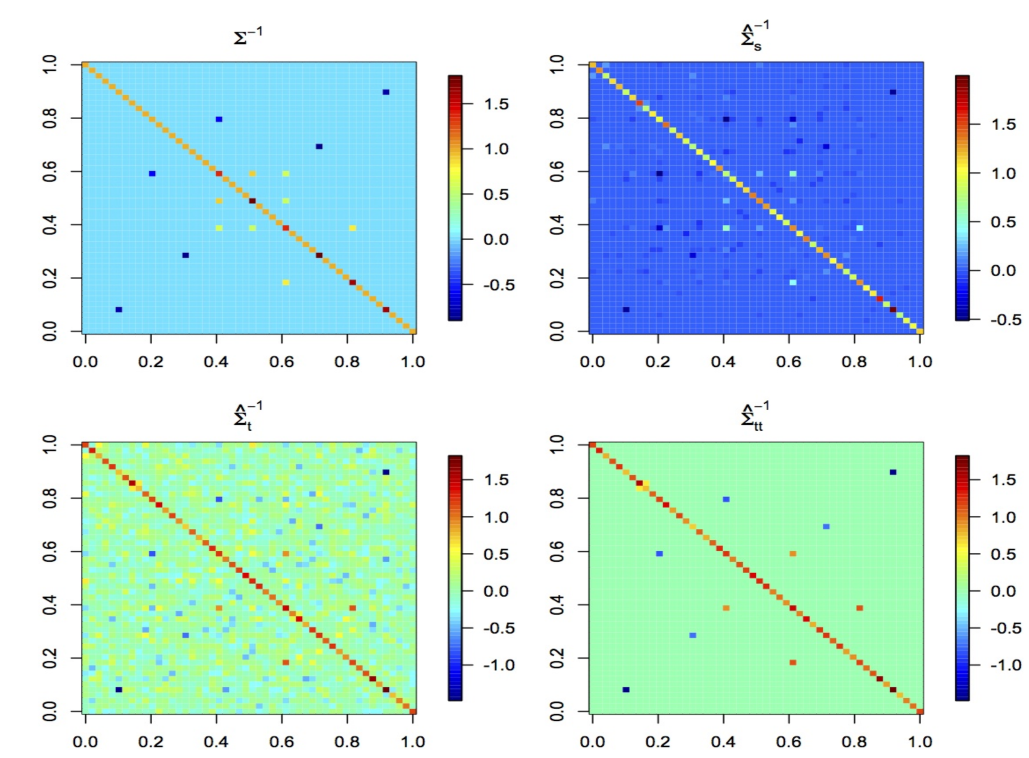
- Scenario 2: Banding precision matrix





- Scenario 3: Sparse precision matrix





References

[1] P. J. Bickel and E. Levina (2008) "Regularized estimation of large covariance matrices", *The Annals of Statistics*, **36**, 199-227.

[2] P. J. Bickel and E. Levina (2008) "Covariance Regularization by Thresholding", *The Annals of Statistics*, **36**, 2577-2604.

[3] E. Levina, A. J. Rothman, and J, Zhu. (2008) "High dimensional covariance matrix estimation using a factor model", *Journal of Econometrics*, **147(1)**,186-197.

[4] J. Fan, Y. Fan, J. Lv. (2008) "Sparse Estimation of Large Co-varianceMatrices via a Nested Lasso Penalty", *Annals of Applied Statistics*, **2**,245-263.

[5] H. Cho and P. Fryzlewicz. (2012) "High-dimensional variable selection via tilting". *Journal of the Royal Statistical Society: Series B (statistical methodology)*, online ISSN 1369-7412.