

# Sparse graphical models and the problem of precision matrix estimation

Michał Makowski<sup>1</sup>

<sup>1</sup>) Uniwersytet Wrocławski  
michalmakowski(at)outlook(dot)com

Relationships between variables in large data sets could be represented as graphs, where nodes represent variables, and edges connect variables which are conditionally dependent, given all other variables. In such graphs number of edges going out from a given node can be used as the measure of the importance of a given variable.

Very often quantitative variables can be transformed so that their distribution resembles the normal distribution, and their joint distribution can be modelled using the multivariate normal distribution. In related Gaussian graphical models edges correspond to nonzero elements of a precision matrix, which is an inverse of a covariance matrix. In a case when the number of observations in a database is comparable to or smaller than the number of variables, classical maximum likelihood estimates of the precision matrix do not exist or have the very large variance.

We cover the problem of a sparse precision matrix estimation, where sparsity pertains to number of nonzero elements of the precision matrix mentioned earlier. We analyze and compare two different regularization methods, gLasso and gSLOPE, which stabilize the performance of MLE. The first is the well-known method for this kind of problems, specifically, gLasso uses  $\ell_1$  norm penalty to shrink the estimates to zero, the second is a novel method, which uses *Sorted  $\ell_1$  Penalized Regression* (SLOPE) to estimate matrix coefficients. The main motivation behind SLOPE is possibility of achieving higher power than Lasso with the same FDR control.

SLOPE was introduced by Bogdan et al. in 2013 as a new estimator for a vector of coefficients in a linear model, gSLOPE was introduced by Piotr Sobczyk. We developed ADMM algorithm for these two approaches. ADMM for gSLOPE is our main result, as gSLOPE was not implemented efficiently earlier. We compare the performance of the obtained solutions using synthetic experiments.

The simulations showed that gSLOPE systematically outperforms gLASSO with respect to ROC curves, which illustrates the compromise between the specificity and sensitivity in discovering the graph structure.