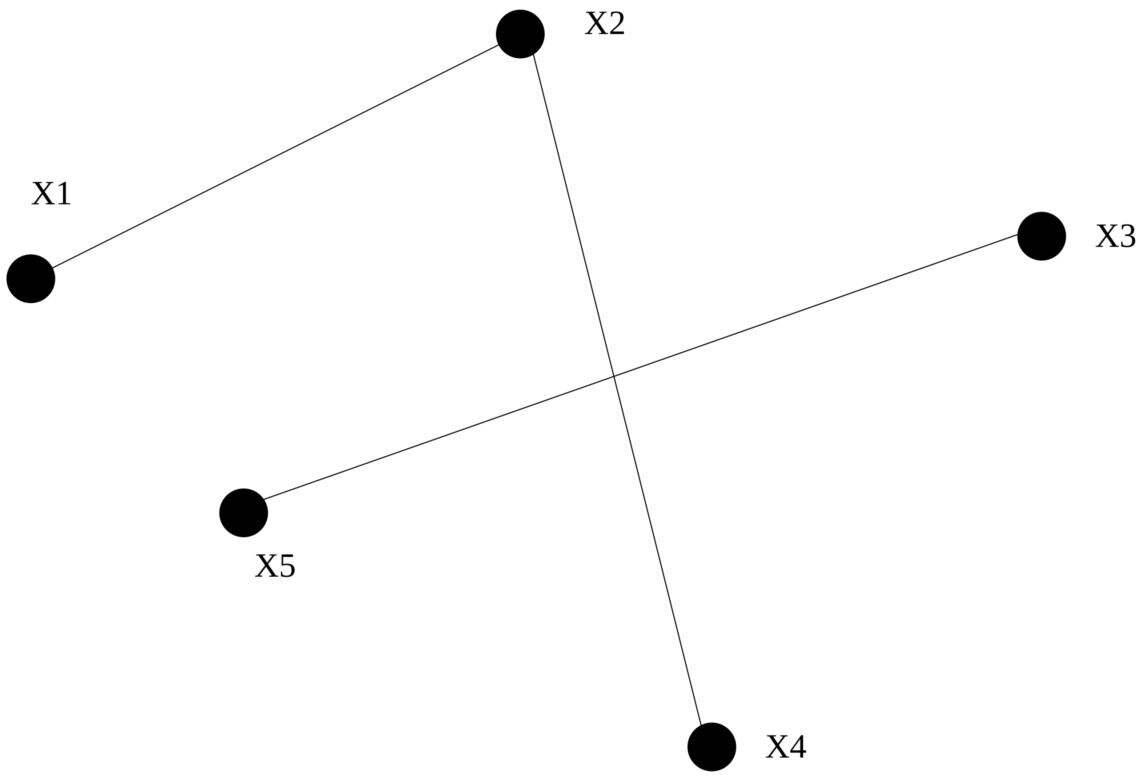
Estimation of sparse Gaussian graphical models:

The graphical lasso

Data: p quantitative variables measured on N observations. For example, $p = 5$ proteins measured in $N = 10,000$ cells

	protein1	protein2	protein3	protein4	protein5
cell 1	2.5	2.6	-1.3	2.8	4.7
cell 2	4.7	-3.3	1.8	3.3	3.4
cell 3	-1.2	-1.4	2.1	4.4	2.4
.					
.					
.					

Goal: estimate the best undirected graph on the variables: a missing link means that the variables are **conditionally independent**



The Lasso for regression

- $$\min \sum_i (y_i - \sum_j x_{ij}\beta_j)^2 \text{ subject to } \textstyle\sum_j |\beta_j| \le t$$
- For $t \ge 0$ sufficiently small, some $\hat{\beta}_j$ will be zero. This is a smooth form of subset selection. We have fast algorithms for solving the lasso.

  Lasso is a nice **convex** method for achieving sparsity.

Naive procedure for graphs

- The coefficient of the jth predictor in a regression measures the partial correlation between the response variable and the jth predictor
- Idea: apply lasso to graph problem by treating each node in turn as the response variable
- include an edge $i \leftrightarrow j$ in the graph if either coefficient of jth predictor in regression for x_i is non-zero, or ith predictor in regression for x_j is non-zero.

Better formulation

- Assume $\mathbf{x} \sim N(0, \Sigma)$ and let $\ell(\Sigma; \mathbf{X})$ be the log-likelihood. X_j, X_k are conditionally independent if $(\Sigma^{-1})_{jk} = 0$
- let $\Theta = \Sigma^{-1}$, and let S be the empirical covariance matrix, the problem is to maximize the penalized log-likelihood

$$\log\det\Theta - \mathrm{tr}(S\Theta) \;\text{ subject to } ||\Theta||_1 \leq t, \tag{1}$$

  over non-negative definite matrices Θ
- Convex problem! How to maximize?

$\texttt{lasso}(\mathbf{X}_{[,-j]},\ \mathbf{X}_j,\ t)$ Naive
$\texttt{lasso}((\mathbf{X}^T\mathbf{X})_{[-j,-j]},\ \mathbf{X}^T_{[-j,-j]}\mathbf{X}_j,\ t)$ Naive using inner products
$\texttt{lasso}(\Sigma_{[-j,-j]},\ \mathbf{X}^T_{[-j,-j]}\mathbf{X}_j,\ t)$ Exact!

Why does this work?

- Considered one row and column at a time, the log-likelihood and the lasso problem have the same gradient
- Resulting algorithm is a blockwise coordinate descent procedure

Details

Subgradient equation for maximization of the log-likelihood is

$$W - S - \rho \cdot \Gamma = 0, \tag{2}$$

using the fact that the derivative of $\log\det\Theta$ equals $\Theta^{-1} = W$, given in e.g (Boyd & Vandenberghe 2004), page 641. Here $\Gamma_{ij} \in \mathrm{sign}(\Theta_{ij})$; i.e. $\Gamma_{ij} = \mathrm{sign}(\Theta_{ij})$ if $\Theta_{ij} \neq 0$, else $\Gamma_{ij} \in [-1, 1]$ if $\Theta_{ij} = 0$.

The upper right block of equation (2) is

$$w_{12} - s_{12} - \rho \cdot \gamma_{12} = 0. \tag{3}$$

Lasso problem is

$$\min_\beta \left\{ \frac{1}{2} ||W_{11}^{1/2}\beta - b||^2 + \rho||\beta||_1 \right\}, \tag{4}$$

where $b = W_{11}^{-1/2} s_{12}$;

The sub-gradient equation from (4) works out to be

$$W_{11}\beta - s_{12} + \rho \cdot \nu = 0, \tag{5}$$

These are the same if we take $w_{12} = W_{11}\beta$.

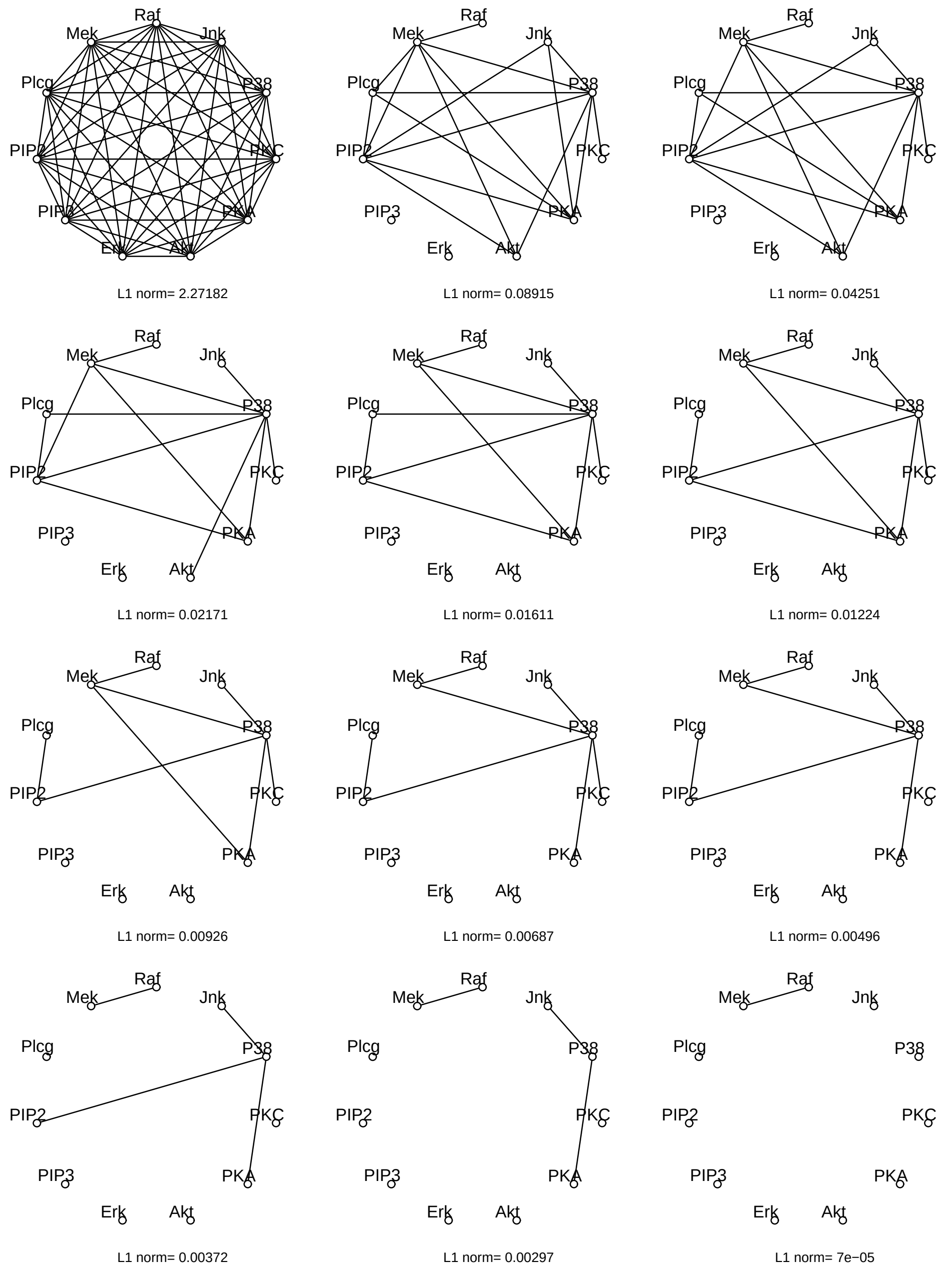
Graphical Lasso Algorithm

1. Start with $W = S + \rho I$. The diagonal of W remains unchanged in what follows.

2. For each $j = 1, 2, \ldots p, 1, 2, \ldots p, \ldots$, solve the lasso problem (4), which takes as input the inner products W_{11} and s_{12}. This gives a $p - 1$ vector solution $\hat{\beta}$. Fill in the corresponding row and column of W using $w_{12} = W_{11}\hat{\beta}$.

3. Continue until convergence

Cell signalling proteins


Raf
Mek
Jnk
Plcg
P38
PIP2
PKC
PIP3
PKA
Erk
Akt
L1 norm= 2.27182
L1 norm= 0.08915
L1 norm= 0.04251
L1 norm= 0.02171
L1 norm= 0.01611
L1 norm= 0.01224
L1 norm= 0.00926
L1 norm= 0.00687
L1 norm= 0.00496
L1 norm= 0.00372
L1 norm= 0.00297
L1 norm= 7e−05

The Punch line

Timings for $p = 400$ *variables*

Method	CPU time
State-of-the-art convex optimizer	27 min
Graphical lasso	6.2 sec

Take that, Steve Boyd!

Speed is also due to our use of the new **coordinate descent** procedures for lasso (Friedman, Hastie, Hoefling ,Tibshirani)

References

Boyd, S. & Vandenberghe, L. (2004), **Convex Optimization**, Cambridge University Press.