# Sparse Inverse Covariance Estimation with the Graphical Lasso - Graphical Learning of Sparse and Structured Biological Networks

**Ioannis Maris**
University of Crete, Heraklion
math5666@math.uoc.gr

## Contents

# Abstract

This project presents a comprehensive overview of the Graphical Lasso algorithm, drawing inspiration from the influential work of R. Tibshirani, T. Haiste, and J. Friedman (2008) [2]. It includes a practical application on RNA-Seq. gene expression levels of patients with five distinct types of tumors: Breast Invasive Carcinoma, Kidney Renal Clear Cell Carcinoma, Colon Adenocarcinoma, Lung Cancer, and Prostate Adenocarcinoma, sourced from the UCI Machine Learning Repository. Due to the high number of predictors (genes), feature selection and dimensionality reduction techniques have been employed to make the application of the Graphical Lasso algorithm feasible.

**Keywords**: Probabilistic Graphical Models, Markov Networks, Biological Structure Learning, Lasso, Graphical Lasso, Penalized Maximum Likelihood Estimation, $\ell^1$ Regularization, Sparse Inverse Covariance Estimation, Blockwise Coordinate Descent, Convex Optimization, Genomics, High Dimensional Data.

# Introduction

## 0.1 Motivation & Related Work

The paper titled "Graphical Lasso: Estimation of Gaussian Graphical Models" by Jerome Friedman, Trevor Hastie, and Robert Tibshirani, discusses a method for estimating sparse graphs using an $\ell^1$ regularization.

The authors start by introducing Gaussian graphical models and their use in representing the conditional independence structure between variables in a Gaussian distribution. They then discuss the problem of estimating the inverse covariance matrix, which encodes this structure.

The motivation behind this paper is the need for an efficient and accurate method to estimate sparse graphs, particularly in high-dimensional settings. Traditional methods for estimating the inverse covariance matrix, which encodes the conditional independence structure between variables in a Gaussian distribution, often result in dense matrices. This is problematic as it does not reflect the true sparsity present in many real-world datasets, and can be computationally challenging in high-dimensional settings.

The authors then present simulation studies to demonstrate the performance of the Graphical Lasso algorithm. They compare it with other methods, such as the covariance thresholding and the adaptive lasso, and show that the Graphical Lasso performs well in terms of both speed and accuracy.

Related works mentioned in the paper:

◇ Banerjee *and others* (2006): Proposed a method for estimating the inverse covariance matrix using an $\ell^1$ penalty. However, their method can be computationally expensive due to the requirement of solving a semidefinite program.

◇ Meinshausen and Bühlmann (2006): Suggested an approach where a lasso model is fitted to each variable against all others, providing an approximation to the inverse covariance matrix. This method is computationally more feasible.

## 0.2 Definition of the Graphical Lasso

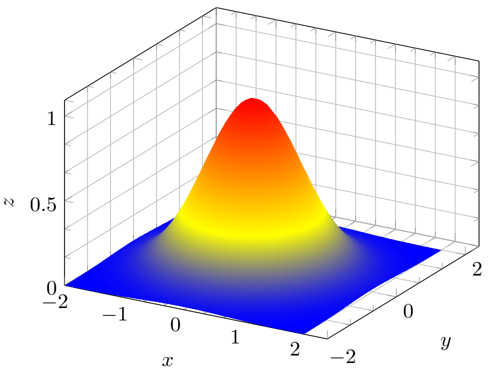
The Graphical Lasso algorithm, an efficient method for estimating a sparse inverse covariance matrix using an $\ell^1$ penalty. The algorithm is based on coordinate descent and blockwise coordinate descent, and the authors provide a detailed explanation of its implementation. The algorithm operates under the assumption that the inverse covariance matrix is sparse

### 0.3 Applications

The Graphical Lasso algorithm has potential applications in various fields where understanding the relationships between a large number of variables is crucial. Here are a few examples:

1. **Genomics**: In genomics, the Graphical Lasso can be used to infer gene networks based on gene expression data. By understanding the conditional dependencies between different genes, researchers can gain insights into the complex biological processes and pathways.
2. **Proteomics**: Similar to genomics, the Graphical Lasso can be used in proteomics to understand the complex interactions between proteins. This can help in identifying key proteins in various biological processes and diseases.
3. **Financial Econometrics**: In financial econometrics, the Graphical Lasso can be used to understand the relationships between different financial variables, such as stock prices, exchange rates, interest rates, etc. This can help in portfolio optimization, risk management, and understanding the dynamics of financial markets.
4. **Neuroscience**: In neuroscience, the Graphical Lasso can be used to understand the complex network of neurons in the brain. This can help in understanding brain diseases and developing treatments.

## 1. Optimizing Penalized Maximum Likelihood Estimation on Multivariate Gaussian Distributions



### 1.1 MLE & Penalized MLE

Consider a random vector $\mathbf{x} \sim \mathcal{N}(0, \Sigma)$ with probability density:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}\det(\Sigma)^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right\} \propto \det(\Theta)^{1/2} \exp\left\{-\frac{1}{2}\mathbf{x}^T\Theta\mathbf{x}\right\},$$

where $\Theta = \Sigma^{-1}$, also known as **precision matrix** and $\Sigma = \mathbb{E}[\mathbf{x}\mathbf{x}^T] > 0$ is the **covariance matrix**.

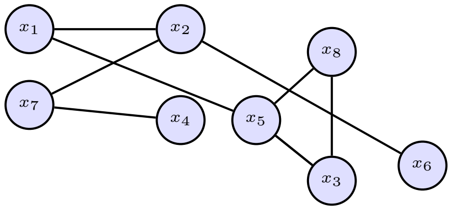
The following **theorem** is crucial: Consider a Gaussian vector $\mathbf{x} \sim \mathcal{N}(0, \Sigma)$. Then,

$$\forall\, i, j : \; x_i \perp\!\!\!\perp x_j \mid \mathbf{x}_{V \setminus \{i,j\}} \iff \Theta_{i,j} = 0.$$

For example, consider the following precision matrix:

$$\Theta = \begin{bmatrix} * & * & 0 & 0 & * & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 & * & * & 0 \\ 0 & 0 & * & 0 & * & 0 & 0 & * \\ 0 & 0 & 0 & * & 0 & 0 & * & 0 \\ * & 0 & * & 0 & * & 0 & 0 & * \\ 0 & * & 0 & * & 0 & 0 & * & 0 \\ 0 & * & 0 & * & 0 & 0 & * & 0 \\ 0 & 0 & * & 0 & * & 0 & 0 & * \end{bmatrix}$$

Then, the corresponding Markov graph is the following:



Now, let's consider `iid` observations $X_1, \dots, X_n$ from multivariate Gaussian distribution $X \sim \mathcal{N}(0, \Sigma)$, then, the log-likelihood is:

$$\ell(\Theta) = \frac{1}{n}\sum_{i=1}^{n} \log f(X_i) = \frac{1}{2}\log\det(\Theta) - \frac{1}{2n}\sum_{i=1}^{n} X_i^T \Theta X_i = \frac{1}{2}\log\det(\Theta) - \frac{1}{2}\langle S, \Theta\rangle,$$

where $S = \dfrac{1}{n}\sum_{i=1}^{n} X_i X_i^T$ (Sample covariance matrix or Empirical) and $\langle S, \Theta\rangle = \text{tr}(S\Theta)$.

### 1.2 MLE Convex Optimization

**Goal**:

$$\underset{\Theta \geq 0}{\texttt{maximize}} \quad \frac{1}{2}\log\det(\Theta) - \frac{1}{2}\langle S, \Theta\rangle \tag{1}$$

**Penalized Log-Likelihood**: Use $\ell^1$ regularization to promote sparsity!

$$\hat{\Theta} = \underset{\Theta \geq 0}{\operatorname{argmin}} \quad \Bigg( \underbrace{\text{tr}(S\Theta)}_{\langle S, \Theta\rangle} - \log\det(\Theta) + \rho \underbrace{\sum_{j \neq k} |\Theta_{jk}|}_{\|\Theta\|_1} \Bigg) \tag{2}$$

Banerjee *and others*, show that the optimization problem is convex and consider estimation of $\Sigma$ (rather than $\Sigma^{-1}$). Assume $W$ is an estimate for $\Sigma$. They proved that the problem can be solved by optimizing each row and column of $W$ separately, using a method called **block coordinate descent**.

$$W = \begin{bmatrix} W_{11} & w_{12} \\ w_{21}^T & w_{22} \end{bmatrix}, \; S = \begin{bmatrix} S_{11} & s_{12} \\ s_{21}^T & s_{22} \end{bmatrix}$$

they show that the solution for $w_{12}$ satisfies:

$$w_{12} = \underset{y}{\operatorname{argmin}} \left\{ y^T W_{11}^{-1} y : \|y - s_{12}\| \leq \rho \right\} \tag{3}$$

### 1.3 The Dual Problem

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

where $b = W_{11}^{-1/2} s_{12}$, if $\beta$ solves (4), then $w_{12} = W_{11}\beta$ solves (3).

**Algorithm 1** *Graphical Lasso*

```
Initialize $W = S + \rho I$. The diagonal of $W$ remain unchaged in what follows.
Repeat for $j = 1, 2, \dots, p, \; 1, 2, \dots, p, \dots$ until convergence:
    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 $w_{12} = W_{11}\hat{\beta}$.
```

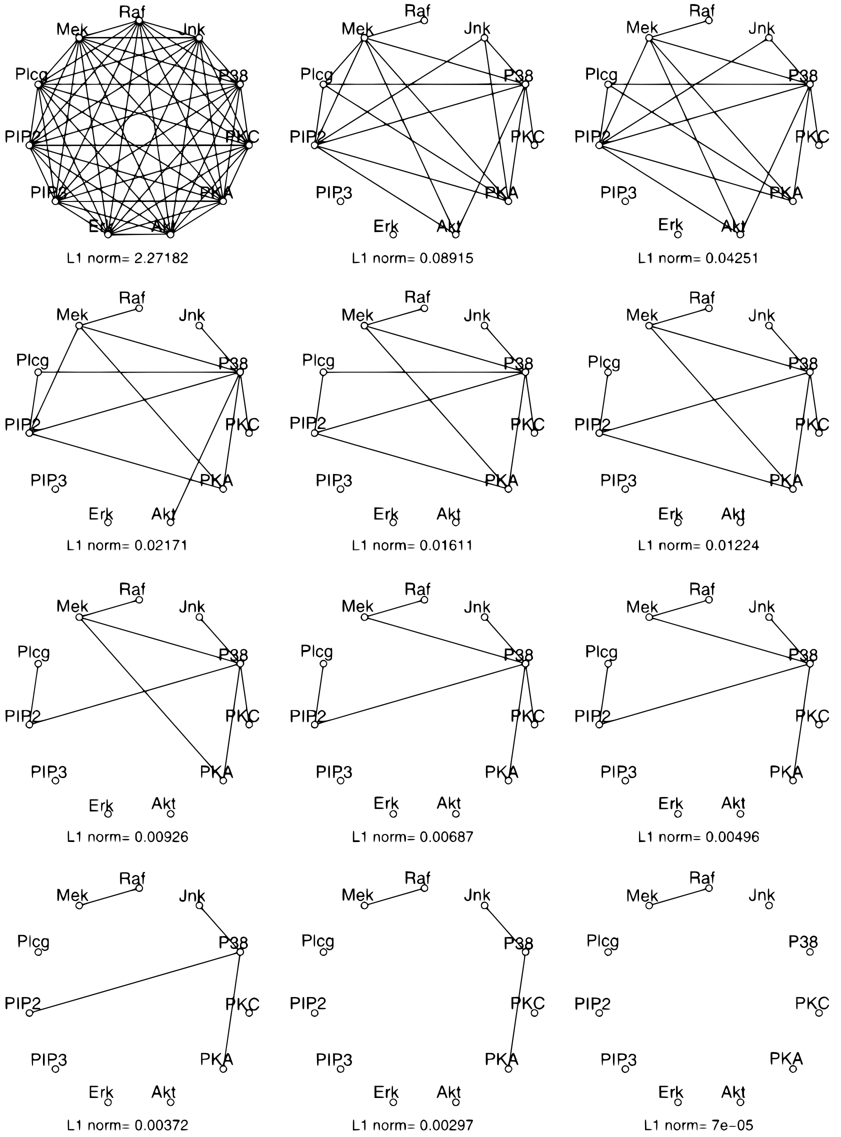


Figure 1: Cell-signaling data: Undirected graphs from graphical lasso with different values of the penalty parameter $\rho$

### 1.4 Application in Cell-Signaling

The authors applied their algorithm to a `cell-signaling` dataset. This dataset consisted of measurements of protein activities in human immune system cells in response to various stimuli. The goal was to infer the network of interactions between these proteins.

The authors performed 10-fold cross-validation using two methods: "regression" and "likelihood". The "regression" approach predicted protein values in the validation set, while the "likelihood" approach evaluated the log-likelihood over the validation set. The unregularized model performed best in both methods, with the "likelihood" approach showing less variability.

They also compared the cross-validated sum of squares of the exact Graphical Lasso approach to the MeinhausenBuhlmann approximation. The exact approach had a clear advantage for lightly regularized models.

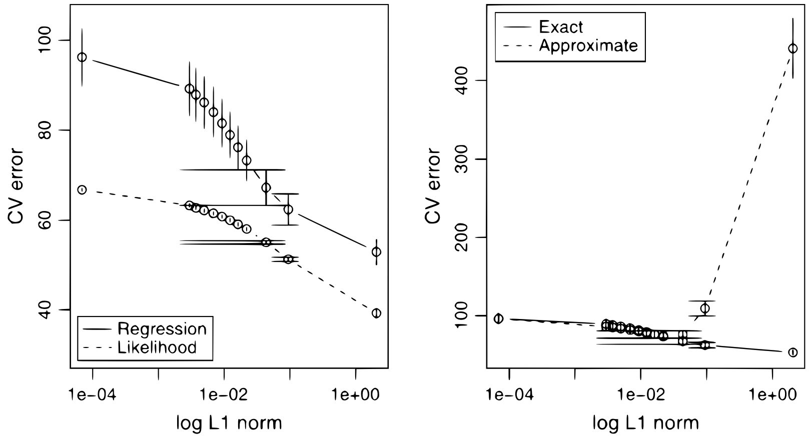


Figure 2: `CV-Error` - `log-scale` $\ell^1$ penalty

- Left panel shows 10-fold `CV` using both regression and likelihood approaches.
- Right panel compares the regression sum of squares of the exact graphical lasso approach to the `Meinhausen–Buhlmann` approximation.

### 1.5 Software

- `sklearn.covariance.GraphicalLasso` in `Python`
- `sklearn.covariance.GraphicalLassoCV` in `Python`
- `glasso` (Graphical Lasso) package in `R`
- `JGL` (Joint Graphical Lasso) package in `R`

### 1.6 Motivating Application: Analyzing Gene Regulatory Interactions in Cancer

The `'Gene Expression Cancer RNA-Seq'` (clickable) dataset from the `UCI Machine Learning Repository` is a comprehensive collection of gene expression data. It contains RNA-Seq gene expression profiles of fresh-frozen primary tumor samples from **801** patients. The dataset includes **20531** genes for each patient sample.

The dataset is high-dimensional, with the number of features (genes) significantly exceeding the number of observations (patients). The primary objective of this dataset is to classify patients into one of the five cancer types, namely BRCA, KIRC, COAD, LUAD and PRAD. This makes it a multi-class classification problem. The dataset is also useful for exploring various aspects of cancer genomics and developing new computational methods for analyzing high-dimensional genomic data.

Graphical Lasso allows us to uncover potential relationships between genes, which can be useful for understanding the underlying biological processes and for identifying potential targets for treatment. However, due to the high dimensionality of the data, applying the Graphical Lasso directly may be computationally challenging. Dimensionality reduction techniques or feature selection methods may be needed to make the problem more tractable. Hence, we employ lasso regression. The figure below shows a pairplot of random genes and their distributions, which closely resemble the Gaussian distribution.

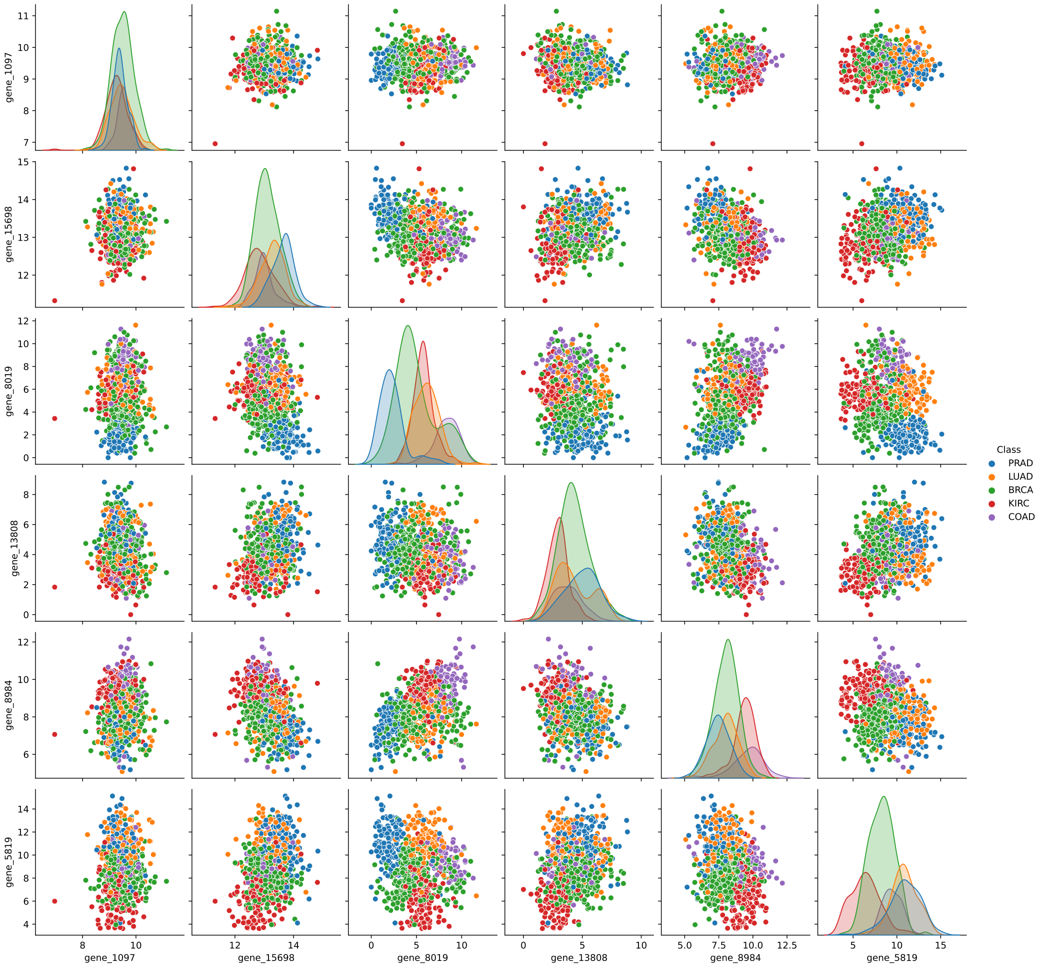


Figure 3: Genes pairplot

Certainly, the Graphical Lasso is incapable of accommodating the analysis of **20,531** genes. In our case, Lasso regression is a highly feasible option for feature selection. After applying lasso regression, **2,240** genes remain. After tuning the parameter $\rho$ (each time with a different set of rhos) we get the following graphs:

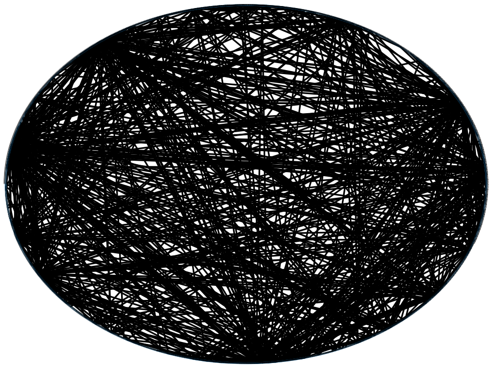


Figure 4: $\rho = 10$ (11min 27s)

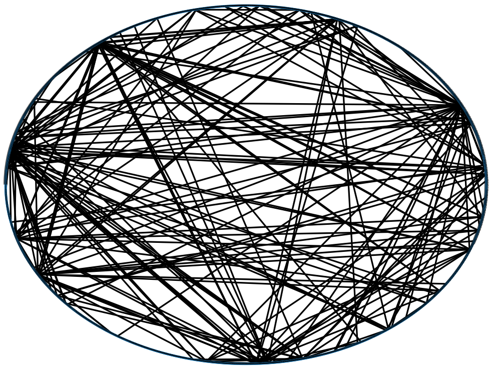


Figure 5: $\rho = 13$

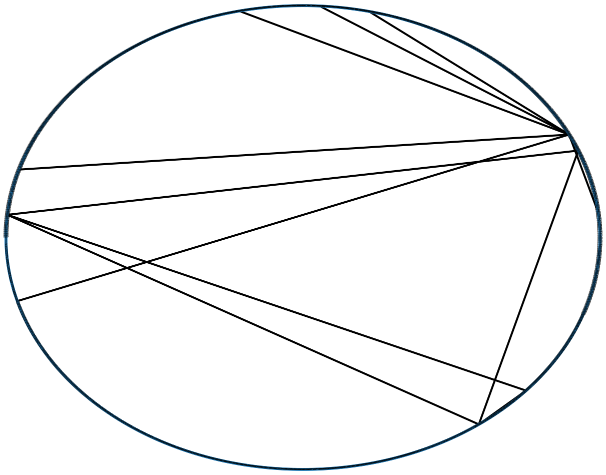


Figure 6: $\rho = 20$

## 2. Conclusions

In conclusion, the Graphical Lasso algorithm provides an efficient and accurate method for estimating sparse inverse covariance matrices. Its effectiveness is demonstrated through simulations and an application to a cell-signaling dataset. The algorithm's potential applications span various fields, including genomics and proteomics, highlighting its versatility and utility in high-dimensional data analysis.

The code can be found here: Github

## References

[1] Patrick Danaher, Pei Wang, and Daniela M Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 76(2):373, 2014.

[2] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[3] Rahul Mazumder and Trevor Hastie. The graphical lasso: New insights and alternatives. *Electronic journal of statistics*, 6:2125, 2012.

[4] Daniela M Witten, Jerome H Friedman, and Noah Simon. New insights and faster computations for the graphical lasso. *Journal of Computational and Graphical Statistics*, 20(4):892–900, 2011.