# A short note on how to estimate the covariance and precision matrix using linear regression

Xiangyu Wang

Department of Statistical Science

Duke University

wwrechard@gmail.com

**Abstract**

We present a method to fit covariance and precision matrix of a Gaussian graphical model via linear regression.

## 1 Method

When the data $x = (x_1, x_2, \cdots, x_p)^T$ is drawn from a joint Gaussian distribution $N(0, \Sigma)$, it is straightforward to estimate the covariance matrix and thus the precision matrix, i.e.,

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} x^{(i)} x^{(i)T} \quad \text{and} \quad \hat{\Omega} = \hat{\Sigma}^{-1}$$

However, when the model gets complicated (for example, time series model) or the dimension gets higher, there are chances where the above equation becomes harder to use, but fitting linear regression $x_i \sim x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_p$ remains simple. In this note, we deduct the equation that recovers the covariance and precision matrix.

For a given dimension $x_i$, we denote $\Sigma_i$ as its covariance, $\Sigma_{-i}$ as the covariance of the remaining dimensions and $\Sigma_{i,-i}$ as the correlation between $x_i$ and the rest dimensions. We have similar notions for $\Omega$. It is obvious that the conditional distribution of $x_i$ follows $x_i \sim N(m_i, C_i)$ where

$$m_i = \Sigma_{i,-i}\Sigma_{-i}^{-1}x_{-i} = -\Omega_{i,-i}\Omega_i^{-1}x_{-i}$$

and

$$C_i = \Sigma_i - \Sigma_{i,-i}\Sigma_{-i}^{-1}\Sigma_{-i,i} = \Omega_i^{-1}$$

Therefore, we can rewrite the conditional distribution in a linear regression form

$$x_i = x_1\beta_1^{(i)} + x_2\beta_2^{(i)} + \cdots + x_p\beta_p^{(i)} + N(0, \sigma_i^2)$$

and we can relate $\beta^{(i)}$ and $\sigma_i$ to the precision matrix as

$$\beta^{(i)} = -\Omega_{i,-i}\Omega_i^{-1}x_{-i} \quad \text{and} \quad \Omega_i^{-1} = \sigma_i^2$$

Thus, by fitting a linear regression $x_i \sim x_{-i}$, we obtain that

$$\hat{\Omega}_{i,-i} = -\hat{\beta}^{(i)}/\hat{\sigma}_i^2 \quad \text{and} \quad \hat{\Omega}_i = \hat{\sigma}_i^{-2}$$

and the precision estimator is

$$\hat{\Omega} = \begin{pmatrix} \hat{\sigma}_1^{-2} & -\hat{\sigma}_2^{-2}\hat{\beta}_1^{(2)} & -\hat{\sigma}_3^{-2}\hat{\beta}_1^{(3)} & \cdots & -\hat{\sigma}_p^{-2}\hat{\beta}_1^{(p)} \\ -\hat{\sigma}_1^{-2}\hat{\beta}_2^{(1)} & \hat{\sigma}_2^{-2} & -\hat{\sigma}_3^{-2}\hat{\beta}_2^{(3)} & \cdots & -\hat{\sigma}_p^{-2}\hat{\beta}_2^{(p)} \\ -\hat{\sigma}_1^{-2}\hat{\beta}_3^{(1)} & -\hat{\sigma}_2^{-2}\hat{\beta}_3^{(2)} & \hat{\sigma}_3^{-2} & \cdots & -\hat{\sigma}_p^{-2}\hat{\beta}_3^{(p)} \\ \vdots & \vdots & \vdots & & \vdots \\ -\hat{\sigma}_1^{-2}\hat{\beta}_p^{(1)} & -\hat{\sigma}_2^{-2}\hat{\beta}_p^{(2)} & -\hat{\sigma}_3^{-2}\hat{\beta}_p^{(3)} & \cdots & \hat{\sigma}_p^{-2} \end{pmatrix} \tag{1}$$

Thus, the covariance matrix will be

$$\hat{\Sigma} = \hat{\Omega}^{-1}.$$

## 2 Numerical experiments

We verify this result via some simple numerical examples. We consider three different covariance structure

Case i $\Sigma = I_p$ is an identity matrix

Case ii $\Sigma_{ij} = 0.9^{|i-j|}$ is a power decay matrix

Case iii $\Sigma = sprandsym(p, 0.2) + aI_p$ where *sprandsym(p, q)* is a matlab function to produce random sparse matrix with $p$ dimension and $q$ density. $a = 8$ for $p = 10, 50$ and $a = 20$ for $p = 200$.

We compare the linear regression method to the direct method shown in the beginning and demonstrate the result in terms of $\|\hat{\Sigma} - \Sigma\|_F^2$ and $\|\hat{\Omega} - \Omega\|_F^2$ where $\|\cdot\|_F$ is the Frobenious norm. The sample size is set to be $n = 1000$ and the dimension is set to be $p = 10, 50, 200$. The results are summarized in

Table 1: The results for Case i

| | | $\|\hat{\Sigma} - \Sigma\|_F^2$ | $\|\hat{\Omega} - \Omega\|_F^2$ |
|---|---|---|---|
| $p = 10$ | Direct estimate | 0.12 | 0.15 |
| | Linear regression | 0.11 | 0.15 |
| $p = 50$ | Direct estimate | 2.65 | 3.41 |
| | Linear regression | 2.65 | 3.41 |
| $p = 200$ | Direct estimate | 39.49 | 89.15 |
| | Linear regression | 39.45 | 89.07 |

Table 2: The results for Case ii

| | | $\|\hat{\Sigma}-\Sigma\|_F^2$ | $\|\hat{\Omega}-\Omega\|_F^2$ |
|---|---|---|---|
| $p = 10$ | Direct estimate | 0.66 | 8.11 |
| | Linear regression | 0.66 | 7.87 |
| $p = 50$ | Direct estimate | 1.60 | 271.18 |
| | Linear regression | 1.63 | 271.39 |
| $p = 200$ | Direct estimate | 39.45 | 8465 |
| | Linear regression | 39.31 | 8433 |

Table 3: The results for Case iii

| | | $\|\hat{\Sigma}-\Sigma\|_F^2$ | $\|\hat{\Omega}-\Omega\|_F^2$ |
|---|---|---|---|
| $p = 10$ | Direct estimate | 8.61 | 0.01 |
| | Linear regression | 8.55 | 0.01 |
| $p = 50$ | Direct estimate | 167.76 | 0.14 |
| | Linear regression | 167.67 | 0.14 |
| $p = 200$ | Direct estimate | 16056 | 0.36 |
| | Linear regression | 16031 | 0.35 |

# 3 Application

One application lies in the time series models. When the the observation is multidimensional, a joint modeling of the $p$-dimension observation would result in many problems. First, the Kalman filter becomes fairly complicated. For example, each dimension could have different latent structures, then the joint Kalman filter could be bad. Second, the computation burden is high. One have to invert an $p \times p$ matrix every step when conducting forward filtering. This might result in bad computational performance when $p$ is large.

Instead, we can make use of the estimation method in this article to estimate the covariance matrix between the $p$ observations. We simply treat the $p$-dimensional observation as $p$ separate time series. When fitting one particular dimension, we treat the remaining dimensions (using filtered results) as the control variables. Thus, the fitting can be parallelizable and is very efficient. Like EM, We then repeat this approach for certain amount of iterations to update the filtered results at each dimension to finalize the output. Finally, we recover the covariance or precision matrix using (1).