Research Article

Vilda Purutçuoğlu*, Ezgi Ayyıldız and Ernst Wit

Comparison of two inference approaches in Gaussian graphical models

Gaussian grafiksel modelinde iki tahmin yaklaşıımının karşılaştırılması

DOI 10.1515/tjb-2016-0298
Received March 30, 2016; accepted August 9, 2016; previously published online May 17, 2017

Abstract

Introduction: The Gaussian Graphical Model (GGM) is one of the well-known probabilistic models which is based on the conditional independency of nodes in the biological system. Here, we compare the estimates of the GGM parameters by the graphical lasso (glasso) method and the threshold gradient descent (TGD) algorithm.

Methods: We evaluate the performance of both techniques via certain measures such as specificity, F-measure and AUC (area under the curve). The analyses are conducted by Monte Carlo runs under different dimensional systems.

Results: The results indicate that the TGD algorithm is more accurate than the glasso method in all selected criteria, whereas, it is more computationally demanding than this method too.

Discussion and conclusion: Therefore, in high dimensional systems, we recommend glasso for its computational efficiency in spite of its loss in accuracy and we believe than the computational cost of the TGD algorithm can be improved by suggesting alternative steps in inference of the network.

Keywords: Gaussian graphical model; Graphical lasso; threshold gradient descent algorithm; Model performance criteria; Biological system.

Özet

Giriş ve amaç: Gaussian grafiksel modeli (GGM), biyolojik sistemde düğümlerin koşullu bağımsızlığına dayanan, en çok bilinen olasılık modellerinden biridir. Bu çalışmada, GGM’in parametre tahminlerini grafiksel lasso metodu ve eşik meyil iniş algoritması kullanarak karşılaştırmaktayız.

Yöntem ve gereçler: Her iki teknin performanslarını belirginlik, F-ölçüsü ve AUC (eğri altında kalan alan) ölçüleri kullanarak değerlendirilir. Analizler, Monte Carlo simulasyonları ile farklı boyutlardaki sistemler kullanılarak yapılmıştır.

Bulgular: Sonuçlar TGD algoritmamın, seçilen bütün kriterler için glasoonun yontemine göre daha doğru sonuçlar verdiği, buna karşın, hesaplama süresinde artışın da glasso’ya göre daha uzun zaman gerektirdiğini göstermektedir.

Tartışma ve sonuç: Bu nedenle, büyük boyutlu sistemlerde, doğrulanktan kaybetmesine rağmen, glasso’yı hesaplamadaki hızlandırma tavsıye ederiz ve inanıyoruz ki TGD algoritmamın hesaplama maliyeti, sistemin tahmininde, alternatif basamaklar önerilerek iyileştirilebilir.

Anahtar kelimeler: Gaussian grafiksel modeli; Grafiksel lasso; Eşik meyil iniş algoritması; Model performers kriterleri.

Introduction

There are different mathematical modelling approaches which can describe the structure of a biological system. Among many alternatives, the main modelling approaches can be listed as the Boolean model, the differential equation model and the stochastic model [1].
Among these choices, the differential equation (DE) model is more common than other approaches since the Boolean model, can merely give very introductory information about the system such as on/off positions. On the contrary, the stochastic model needs very detailed knowledge about the system and mostly such knowledge has not been available yet. Moreover, the calculation is very costly. Hence, in this study, we deal with an alternative model of DE, called the Gaussian graphical model (GGM), which can explain the steady-state behavior of the system by using the changes in concentrations of the species.

GGM is a probabilistic approach where the states of the system are described under the multivariate normal distribution and the interactions between genes are indicated via the precision, i.e. inverse of the covariance, matrix, whose elements are estimated under the assumption of the conditionally independency between genes. Moreover, the entries of the precision indicate the strength of interactions without the directional knowledge of genes. There are several methods to estimate the entries of the precision matrix in GGM. For instance, Schafer and Strimmer [2] suggest to test the distribution of the observed partial correlations across edges by the false discovery rate so that the statistical significance of each estimated edge, i.e. interaction, in the network, can be controlled. But the most two common approaches are called as the neighborhood selection algorithm with the lasso regression [3] and the graphical lasso (glasso) approach [4]. The first method suggests a lasso model for each gene in the system by regressing other genes as predictors. In the calculation, an optimization procedure based on a convex function is applied consecutively for each gene in the system. On the other hand, the second method proposes a blockwise coordinate descent approach which depends on the maximization of the penalized likelihood for solving the lasso regression. Here, the $l_1$-penalty is applied on the entries of the precision matrix.

In this study, we aim to compare two major inference approaches based on GGM. We evaluate the performance of glasso with the threshold gradient descent algorithm in different dimensional and sparse systems based on various criteria. From the Monte Carlo runs and real systems’ applications, we assess their accuracies and computational demand. Hereby, in the organization of the study, we present the description of GGM with two suggested methods with model performance criteria in Materials and methods. Application part is dedicated to the analyses and the results. Finally, we summarize the outputs and suggest the future works in Conclusion.

### Materials and methods

In a biological system, the interactions between components can be described by using graphical models that consist of a $p$-dimensional set of nodes representing the components of the system such as genes or proteins via a state vector $Y=(Y^{(0)},...,Y^{(p)})$, and a set of edges displaying the interactions among these components. A most widely used graphical model is the Gaussian graphical model (GGM). This model makes the assumption that the states of a $p$-dimensional system have a multivariate normal distribution with a $p$-dimensional mean vector $\mu$ and a $(p \times p)$ - dimensional covariance matrix $\Sigma$. In GGM, the conditional independencies of two nodes are indicated with the absence of the undirected edge between corresponding nodes and are shown by zero entries in the inverse of the covariance matrix that is also called the precision or concentration matrix.

In genomic data, since we typically have a large number of genes $p$ and few samples or observations per gene, $n$, the sample covariance estimator $\hat{\Sigma}=S$ is inferred via different approaches. One of these methods is to apply the regression method in such a way that the regression functions for each node against all remaining nodes give the conditional independent structure of the model. Accordingly, when we divide the Gaussian state vector $Y$ into the two parts such that $Y=(Y^{(-p)},...,Y^{(p)})$, where $Y^{(-p)}=(Y^{(0)},...,Y^{(p-1)})$ denotes the vector of all nodes except for the last one, we can obtain the conditional distribution for the node $Y^{(p)}$ on all the remaining nodes via

$$Y^{(p)} | Y^{(-p)} = y - N(\mu_{-p} + (y - \mu_{p})^T \Sigma_{-p,p}^{-1} \Sigma_{-p,p}, \Sigma_{-p,p}^{-1}) \quad (1)$$

In Equation (1), $(A)^T$ denotes the transpose of $A$. Hereby, the mean $\mu$ and the covariance $\Sigma$ can be partitioned as

$$\mu = \begin{bmatrix} \mu_{-p} \\ \mu_{p} \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{-p,-p} & \Sigma_{-p,p} \\ \Sigma_{-p,p}^T & \Sigma_{p,p} \end{bmatrix} \quad (2)$$

where $\mu_{-p}$ presents the mean vector of all nodes except for the last one and $\mu_{p}$ indicates the mean of the last node. Furthermore, $\Sigma_{-p,-p}$ is the $(p-1) \times (p-1)$-dimensional variance-covariance matrix of all nodes apart from the last node. Accordingly, $\sigma_{-p,p}$ refers to the $(p-1) \times 1$-dimensional covariance vector of $Y^{(-p)}$ and $Y^{(p)}$, and $\sigma_{p,p}$ denotes the variance of the last node $Y^{(p)}$.

Thereby, the structure of the conditional dependence is identified by the regression coefficient $\beta = \sum_{-p,p} \Sigma_{-p,p} \Sigma_{p,p}$.
Hence, \(Y^{(p)}\) and \(Y^{(-p)}\) \((i = 1, \ldots, p - 1)\) are conditionally independent when \(\beta_i = 0\). Hereby, the idea of the conditional independency is combined under the lasso regression via

\[
Y^{(p)} = \beta Y^{(-p)} + \epsilon
\]

(3)

where \(Y^{(p)}\) and \(Y^{(-p)}\) present the state of the \(p\)th node and the states of all other nodes except the node \(p\), respectively. Finally, \(\epsilon\) describes the independent and normally distributed error term with the mean vector and the covariance matrix as given in Equation (2).

**Graphical lasso method**

In inference of \(\Theta\), if \(f(y)\) presents the joint density function of the observation \(y\) for the vector \(Y\) by

\[
f(y, \mu, \Sigma) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (y - \mu)^T \Sigma^{-1} (y - \mu) \right\},
\]

(4)

the log-likelihood \(l(\Theta)\) is shown as

\[
l(\Theta) = \frac{n}{2} \log |\Theta| - \frac{n}{2} \text{Tr}(S\Theta)
\]

(5)

by replacing the covariance matrix by the precision matrix while \(\text{Tr}(\cdot)\) indicates the trace of the associated matrix and \(S = S_y\) denotes the following sample covariance matrix.

\[
s_{ij} = \frac{1}{n} \sum_{k=1}^{n} (y_{ik}^{(k)} - \bar{y}(i))(y_{jk}^{(k)} - \bar{y}(j)).
\]

(6)

Now, by maximizing the log-likelihood \(l(\Theta)\) under the zero constraints, the maximum likelihood estimator (MLE) of the precision matrix equals to \(\hat{\Theta} = S^{-1}\). In general, although the MLE method is successful when full data are available and \(\Theta\) is nonsingular, it cannot be applicable for non-invertible \(\Theta\) [5]. But under these challenges, the lasso-regression model computes the \(l_1\)-penalty for \(\beta\) such that \(\|\beta\|_1 = \sum_i |\beta_i| < \lambda\) [5-8].

In the lasso-regression approach, another important issue is the selection of the penalty parameter \(\lambda\), \((i = 1, \ldots, n)\). The penalized likelihood idea is the most common approach for this challenge as an alternative of the ordinary least square estimation method with low bias and large variance [7] and here, the \(l_1\)-norm is more preferable due to its easy interpretation [8]. In this well-known method, the penalized likelihood optimization can be written by the Lagrangian dual form as

\[
\max_{\Theta} \left\{ \log |\Theta| - \text{Tr}(S\Theta) - \lambda \|\Theta\|_1 \right\}.
\]

(7)

Banerjee et al. [9] suggest the block coordinate descent and the Nestrov’s first order method to decide on \(\lambda\) in Equation (7) by gaining from computational demand for large systems. Witten et al. [6] consider to write the estimated \(\Theta\) as the block diagonal form in order to speed up the computation. But the necessary and sufficient condition of the Karush-Kuhn-Tucher condition [6] is required so that the estimated \(\Theta\) can be block diagonal. Indeed in the selection of the penalty constant, several choices can be considered. For instance, the smoothly clipped absolute deviation method [10] suggests an asymmetric, non-concave and singular precision at the origin to get sparse solutions. The adaptive lasso [11] proposes different weights for penalizing distinct coefficients in the \(l_1\)-penalty. Alternatively, the fused lasso [12] considers the sparsity of the coefficients and their local constancies via the sparsity of their differences. Finally, the elastic net approach [13] can successfully capture the sparsity of the system when its dimension is much bigger than the number of observations. But LARS [14] performs better than the elastic net under the lasso regression. We can also choose the \(k\)-cross validation approach for simplicity. But in this study, we use a set of penalty values both in glasso and the threshold gradient descent algorithm for the comparative analyses [15, 16].

**Threshold gradient descent algorithm**

As an alternative inference method of the \(l_1\)-lasso is the threshold gradient descent algorithm (TGDA). In TGDA, the sparsity problem of the system is solved by penalizing the estimation of the off-diagonal elements of \(\Theta\). In the calculation, first of all, \(\Theta\) is splitted into diagonal and off-diagonal vectors as \(\theta^1 = \theta_{ii}, \ldots, \theta^p\) representing the vector of diagonal elements of \(\Theta\) and \(\theta^{ij} = (\theta_i)_i \neq j\) describing the vector of off-diagonal elements of \(\Theta\), respectively. Then, the likelihood function \(l\) is described as follows.

\[
l(\theta^1, \theta^2) = \frac{n}{2} \log |\Theta| - \frac{1}{2} \sum_{k=1}^{n} Y^{(k)'} \Theta Y^{(k)},
\]

(8)

where \(Y^{(k)}\) denotes the vector of the \(k\)th observation. To deal with the sparsity of the precision matrix, the negativity of the log-likelihood function in Equation (8) can be defined as a loss function.

On the other hand, the gradient of the loss function with respect to \(\Theta\) can be written as

\[
\frac{1}{\Theta} = \frac{n}{2} \Theta^{-1} - \frac{1}{2} \sum_{k=1}^{n} Y^{(k)'} \Theta Y^{(k)}.
\]

(9)
After initializing $\theta^d$ to a unit vector and $\theta^o$ to a zero vector, the off-diagonal elements of $\Theta$, are updated by using the gradient descent step via

$$\hat{\Theta}^o(\gamma + \Delta \gamma) = \hat{\Theta}^o(\gamma) + \Delta \gamma d(\gamma).$$

(10)

In this expression, $\hat{\Theta}^o(\gamma)$ denotes the value of $\theta^o$ for the current $\gamma$. $\Delta \gamma$ shows a very small increment ($\Delta \gamma > 0$) and $d(\gamma)$ describes the direction of the tangent vector which corresponds to a descent direction at each step. Friedman and Popescu [17] define $d(\gamma)$ as the multiplication of the two functions such that

$$d(\gamma) = w_j(\gamma) \cdot g_j(\gamma)$$

(11)

for $j = 1, ..., q$ in which

$$w_j(\gamma) = I(|g_j(\gamma)| \geq \lambda \cdot \max_{1 \leq i \leq q} |g_i(\gamma)|).$$

(12)

In Equation (12), $I[\cdot]$ stands for an indicator function and $\lambda$ shows a threshold parameter in the range $[0, 1]$ to adjust the diversity of the values of $w(\gamma)$. Thus, as the diversity increases, $\lambda$ is closer to 1 meaning that $\lambda = 1$ implies the most sparse graph.

On the other side, the entries of $\Theta$ are maximized via an iterative method with respect to the log-likelihood function in Equation (8) to update its diagonal elements. Here, we choose the Newton-Raphson approach among alternative iterative methods.

In the following part, we implement these two methods in inference of small and moderately large systems.

**Applications**

In the literature, there are many model performance criteria used to evaluate the accuracy of the binary classification. In this study, we apply all these well-known criteria, listed as the precision, recall, specificity, false positive rate, false discovery rate, accuracy, F-measure and the Matthews correlation coefficient. In the comparison, we initially evaluate the performance of these criteria in GGM and once we select the most appropriate criterion for the assessment of the estimation via the Monte Carlo runs, we implement it for the analyses of simulated and real datasets.

**Comparison of model performance criteria**

To generate true precision matrices for the evaluation, we apply two different scenarios under three different numbers of observations per gene. The first scenario considers that $\Theta$ has only positive entries corresponding to positive relations between genes and in the second scenario, we accept negative entries representing inhibitory relations between genes. For both plans, we compute precision matrices under $(5 \times 5)$, $(10 \times 10)$, $(20 \times 20)$, $(40 \times 40)$ and $(50 \times 50)$ dimensions, and consider $40\%$, $70\%$, $85\%$, $92.5\%$, and $94\%$ sparsity percentages, respectively. To produce the sparsity, we use the band matrix that is typically preferred in graphical models. In this study, our $\Theta$’s have a special form, similar to the tri-diagonal band matrix.

In order to compute the model performance criteria from the estimated $\Theta$’s, we initially convert the true and estimate $\Theta$’s into a binary form via a threshold value. In the calculation, we take this value as 0.01 since in most of the biological studies with a very sparse number of observations per gene and a very limited information about the system of interest, the expression values higher than zero are typically taken as the candidate significant genes of the network. Moreover, by setting all the estimated values less than 0.01 to zero, and by allowing the remaining as 1, we enforce the system to be highly connected since most of the realistic biochemical systems have such dense structures [18]. Finally, we iterate the algorithm for 10,000 times and take the mean of these Monte-Carlo outputs. In every iteration, we choose the best score for each measure under its optimal $\lambda$.

On the other side, for the comparative analyses of two inference methods, we observe from previous studies [19] that the penalties between 0 and 1 are the most suitable range for the data generated from the multivariate normal distribution. Because the upper boundary for $\lambda$ differs for each sample covariance matrix. Moreover, distinct $\lambda$’s for each run and the inference method cause non-comparable results. Hence, we select common and the widest range for $\lambda$. Then, we divide this range into 7 equal widths to generate our set of penalties, which equals to 0.10, 0.25, 0.40, 0.55, 0.7, 0.85 and 1, in both inference methods and evaluate all results from each penalty.

In all computational works, we carry in the R programme language with the version 2.15.1 and execute our originally developed codes on the Core i5 3.10 GHz processor.

Furthermore, in order to detect the effects of the number of observations per gene, we check the outputs under 5, 10 and 30 numbers of observations for each gene. From the assessment in Tables 1 and 2 it is seen that both the specificity and the accuracy are close to 1 when the dimension of matrices increases in the first and the second scenarios, and this conclusion is invariant from
Table 1: Comparison of model performance criteria under 10,000 Monte-Carlo runs and under both positive off-diagonal entries (the first scenario, S1) and negative off-diagonal entries (the second scenario, S2) with 10 numbers of observations per gene.

| p  | Precision | Recall | Specificity | Accuracy | FDR | FPR | F-measure | MCC |
|----|-----------|--------|-------------|----------|-----|-----|-----------|-----|
| S1 | 5         | 0.9734 | 0.9180      | 0.9668   | 0.8038 | 0.0266 | 0.0332    | 0.8374 | NA  |
|    | 10        | 0.9635 | 0.8968      | 0.9911   | 0.8385 | 0.0365 | 0.0089    | 0.6917  | 0.6098 |
|    | 20        | 0.9703 | 0.8554      | 0.9975   | 0.9077 | 0.0297 | 0.0025    | 0.6021  | 0.5896 |
|    | 40        | 0.9817 | 0.7939      | 0.9995   | 0.9515 | 0.0183 | 0.0005    | 0.5511  | 0.5813 |
|    | 50        | 0.9852 | 0.7712      | 0.9996   | 0.9611 | 0.0148 | 0.0004    | 0.5399  | 0.5795 |
| S2 | 5         | 0.9973 | 0.9192      | 0.9973   | 0.7825 | 0.0027 | 0.0027    | 0.8225  | NA  |
|    | 10        | 0.9946 | 0.8967      | 0.9989   | 0.8364 | 0.0054 | 0.0011    | 0.6862  | 0.6054 |
|    | 20        | 0.9944 | 0.8545      | 0.9996   | 0.9072 | 0.0056 | 0.0003    | 0.5987  | 0.5871 |
|    | 40        | 0.9960 | 0.7941      | 0.9999   | 0.9513 | 0.0040 | 0.0000    | 0.5489  | 0.5792 |
|    | 50        | 0.9968 | 0.7707      | 0.9999   | 0.9609 | 0.0031 | 0.0000    | 0.5378  | 0.5779 |

Perfection level: 1 1 1 1 0 0 1 1

The number of observations per gene. But the performance of the specificity is better than the accuracy, whereas, the recall and the F-measure decrease substantially while the dimension of the systems increases. Moreover, the precision is not affected by the changes in dimensions. Finally, the behaviors of the false discovery rate and the false positive rate are similar to the precision and the specificity, respectively. Thus, we see that the specificity measure is observed as the best model performance criterion for the selection of the optimal \( \lambda \) and this conclusion is irrelevant from the number of observations per gene. On the other hand, since there is a trade-off between specificity and sensitivity, we also calculate F-measure and AUC besides specificity so that we can evaluate the effects of both rates simultaneously.

GGM via glasso

In the application of glasso, we calculate the mean values of the specificity and F-measure under 1000 Monte Carlo runs based on different dimensional matrices, and we consider distinct sparsity percentages. Finally, for each matrix, the number of observations per gene is taken as 10.

From Table 3, the results show that the specificity values differ regarding penalty values and become closer to 1 when the penalties increase. Additionally, the specificity values increase when the dimension of \( \Theta \)’s increases. On the other side, the values of F-measure increase when the penalties increase except the \((5 \times 5)\)-dimensional matrices. However, the F-measures decrease while the dimensions rise.

GGM via TGDA

In order to run TGDA, the increment \( \Delta \gamma \) in Equation (10) is taken as \( 10^{-3} \) in all Monte-Carlo runs to grid the estimation space comprehensively. Additionally, the identity matrix is chosen as an initial matrix in the calculation. Then, its off-diagonal elements are updated by using the TGD steps and the Newton-Raphson algorithm is implemented to estimate diagonal elements of \( \Theta \). Finally, the estimated
Θ’s are converted to the binary form with the threshold value 0.01.

From the findings in Table 4, we see that the specificity values become closer to 1 when the penalties increase. Also, the specificity values increase when the dimensions increase. Whereas, the F-measures increase when the penalties rise except the (5 × 5)-dimensional matrices regarding the results of the glasso method.

Furthermore, from the comparison of the specificity via glasso and TGDA, it is detected that TGDA with the threshold value 1 gives better results than glasso. To make the comparison between these two methods, we also calculate the area under the curve (AUC) method which refers to the area under the receiver operating characteristic (ROC) curve. As seen in Table 5, the TGD algorithm has greater AUC values with respect to the glasso method for all dimensions. Moreover, if we compare the computational times of both approaches, we observe a big difference between two methods. The time required for TGDA is approximately 20 times longer than glasso for (20 × 20), and (40 × 40)-dimensional precision matrices. We consider that the reason of this great computational demand for TGDA depends on the Newton Raphson algorithm which is applied to estimate diagonal entries of Θ. Because if Θ is non-positive definite, the algorithm turns back to the calculation and changes the initial values until the requirement is satisfied. Thus, we suggest that despite of its loss in accuracy, glasso is more convenient for high dimensional systems due to its advantage in the computational time.

### Application via real datasets

To evaluate the results under realistic systems, we implement methods to a simulated dataset of a complex
JAK/STAT pathway. Moreover, to assess the suggested approaches in real data, we select two different systems, namely, the MAPK/ERK pathway and the lipid biosynthesis pathway of the oil palm.

The JAK/STAT pathway, whose major components JAK and STAT, is an important signaling pathway, in particular, for the immune system of living cells and it is widely worked on the treatment of illnesses such as the hepatitis B [20, 21].

In this study, we generate a time-course dataset for this system by using the Gillespie algorithm [22]. To get the simulated data, we initialize the number of molecules for each protein and the reaction rate constants of the pathway as given in the study of Maiwald et al. [21] by using their quasi reaction list and the quasi true structure of the pathway defined by 38 proteins. The underlying quasi interactions have been also validated in different studies [20–24]. Then, we gather 10 observations for each protein as presented in the study of Ayyıldız [19].

Finally, we apply both inference approaches to estimate the structure of JAK/STAT. Once the precision matrices are estimated by both approaches, we convert these matrices into the binary form via the same threshold value. The results of TGDA and glasso are shown in Table 6 and it is seen that the specificity values of TGDA are greater than the corresponding values of glasso. Furthermore, when the penalty values increase, the specificity values also increase since the JAK/STAT system is sparse. Whereas, the F-measures are low for both inference methods.

As the application via real data, we use the MAPK/ERK pathway, which affects the growth control and the cell death of all eukaryotes [25, 26], and thereby, are widely worked on oncogenic researches. In this study, we estimate the MAPK/ERK pathway by using a real time-course dataset [26, 27] which is collected under 8 time-points for 6 proteins, namely, Ras.GTP, Ras.GDP, Raf.Am, MEK.p2, ERK.p2. The quasi true interactions between these proteins have been already validated in various laboratories [25–29]. Thus, in our analyses, we compare our estimated systems via both approaches with this true structure and from the results in Table 6, it is observed that TGDA outperforms glasso.

Finally, we analyze the changes in the lipid content of developing the oil palm mesocarp. The mesocarp is the outer pulp containing the palm oil which is the highest oil producing plant. Accordingly, the lipid biosynthesis is one of the primary metabolizes of the oil palm and understanding this pathway for the oil palm is necessary in order to improve the quality of the oil [30].

For the analyses, the real time-course dataset which contains five time points for five components showing the weight of the lipid classes is taken from Oo et al. [31].

Table 6: Comparison of the specificity and F-measure computed via the TGD algorithm and the glasso method for the JAK/STAT, the MAPK/ERK pathways and the pathway of the Oil palm based on set of common penalties.

| Penalty value | Specificity | F-measure | Specificity | F-measure |
|---------------|-------------|-----------|-------------|-----------|
| JAK/STAT pathway | 0.10 0.8280 0.1060 | 0.7020 0.1910 |
| 0.25 0.9300 0.0860 | 0.6610 0.1860 |
| 0.40 0.9500 0.0400 | 0.6630 0.1950 |
| 0.55 0.9590 0.0430 | 0.6150 0.1800 |
| 0.70 0.9650 0.0220 | 0.6360 0.1940 |
| 0.85 0.9670 0.0400 | 0.6480 0.1940 |
| 0.10 0.9700 0.0400 | 0.6580 0.1870 |
| 0.25 0.4780 0.5720 | 1.0000 0 |
| MAPK/ERK Pathway | 0.40 0.6960 0.6200 | 1.0000 0 |
| 0.55 0.8260 0.5220 | 0.9130 0.5720 |
| 0.70 0.9130 0.5720 | 0.9130 0.5720 |
| 0.85 0.9130 0.5720 | 0.9130 0.5720 |
| 1 0.9130 0.5720 | 0.9130 0.5720 |
| 0.10 0.5450 0.5920 | 0.2730 0.6670 |
| 0.25 0.9090 0.5720 | 0.2730 0.6670 |
| The Oil palm Pathway | 0.40 0.9090 0.5720 | 0.2730 0.6670 |
| 0.55 0.9090 0.5720 | 0.2730 0.6670 |
| 0.70 0.9090 0.5720 | 0.3640 0.6450 |
| 0.85 0.9090 0.5720 | 0.3640 0.6450 |
| 1 0.9090 0.5720 | 0.3640 0.6450 |
and the true interactions between these five major components have been reported in Oo et al. [31] and Quek et al. [32]. From Table 6 obtained from the comparison of biologically validated links with estimated ones, it is detected that the specificities of glasso are lower than the corresponding values of TGDA. Whereas, there is no difference in the F-measure of two inference methods.

**Conclusion**

In this study, glasso and TGDA methods have been implemented to estimate the biological networks. Both methods have been widely used in the literature, but have not been applied in the GGM context under such comprehensive simulation scenarios. In the comparison of these approaches, we have initially listed outputs of different model performance criteria and have selected the specificity measures among alternatives as the best criterion for the model selection. But due to the trade-off between the specificity and the sensitivity, we have also computed the corresponding sensitivities (F-measure) and AUC for all analyses.

From the findings, we have concluded that TGDA is more accurate than glasso based on all measures of Θ. However, its computational demand is high. Therefore, in high dimensional matrices, we recommend glasso to gain in computation in spite of its loss in accuracy.

On the other side, as the future study, we consider to improve the performance of TGDA in computational cost by proposing alternative steps for the Newton-Raphson algorithm in the calculation of non-singular precisions. Furthermore, we think to model the system via the logistic algorithm in the calculation of non-singular precisions. By proposing alternative steps for the Newton-Raphson algorithm in the calculation of non-singular precisions. Instead of using the original algorithm, we improve the performance of TGDA in computational cost in spite of its loss in accuracy.

**Acknowledgements:** V. Purutcuoglu and E. Ayyildiz thank to SysPatho EU FP-7 Collaborative Project (Project No: 260429) for the support. Moreover, all authors thank to the editors and anonymous referees for their valuable comments which improve the quality of the paper.

**Conflict of interest statement:** The authors have no conflict of interest.

**References**

1. Bower JM, Bolouri H. Computational modeling genetic and biochemical networks. Cambridge, Massachusetts: Institute of Technology Press, 2001.

2. Strimmer K, Schafer J. A shrinkage approach to large-scale covariance matrix estimation and implication for functional genomics. Stat Appl Genet Mol Biol 2005;4:Article 32.

3. Meinshausen N, Bühmann P. High dimensional graphs and variable selection with the lasso. Ann Stat 2006;34:1436–62.

4. Friedman JH, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 2008;9:432–41.

5. Whittaker J. Graphical models in applied multivariate statistics. New York: John Wiley and Sons, 1990.

6. Witten D, Friedman J, Simon N. New insights and faster computations for the graphical lasso. J Comput Graph Stat 2011;20:892–900.

7. Tibshirani R. Regression shrinkage and selection via the lasso. J Royal Stat Soc 1996;58:267–88.

8. Goeman JJ. Penalized estimation in the cox proportional hazards model. Biometrical J 2010;52:70–84.

9. Banerjee O, Ghaoui L, D'Aspremont A. Model selection through sparse maximum likelihood estimation. J Mach Learn Res 2008;9:485–516.

10. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc 2001;96:1348–60.

11. Zou H. The adaptive lasso and its oracle properties. J Am Stat Assoc 2006;101:1418–29.

12. Tibshirani R, Saunders M, Rosset S, Zhu J, Knight K. Sparsity and smoothness via the fused lasso. J Royal Stat Soc 2005;67:91–108.

13. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc 2005;67(Part 2):301–20.

14. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. Ann Stat 2004;32:407–99.

15. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: prediction, inference and data mining, 2nd ed. New York: Springer-Verlag, 2009.

16. Wit E, Vinceti V, Purutcuoglu V. Statistics for biological networks. 25th International Biometric Conference Short Course Notes, Florianopolis, Brazil. 2010.

17. Friedman JH, Popescu BE. Gradient directed regularization. Technical Report, Stanford University, 2004.

18. Barabasi AL, Oltvai, ZN. Network biology: understanding the cell’s functional organization. Nat Rev Genet 2004;5:101–13.

19. Ayyildiz, E. Gaussian graphical approaches in estimation of biological systems. Master Thesis, Graduate School of Natural and Applied Sciences, Department of Statistics, METU, Ankara, Turkey, 2013.

20. Shuai K, Liu B. Regulation of JAK/STAT signalling in the immune system. Nat Rev Immuno 2003;3:900–11.

21. Maiwald T, Schneider A, Busch H, Sahle S, Gretz N, Weiss TS, et al. Combining theoretical analysis and experimental data generation reveals IRF9 as a crucial factor for accelerating interferon λ-induced early antiviral signalling. FEBS J 2010;277:4741–54.

22. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem 1977;81:2340–61.

23. Yamada S, Shiono S, Joo A, Yoshimura A. Control mechanism of JAK/STAT signal transduction pathway. FEBS Lett 2003;534:190–6.

24. Zi Z, Cho KH, Sung MH, Xia X, Zheng J, Sun Z. In silico identification of the key components and steps in IFN-γ induced JAK-STAT signaling pathway. FEBS Lett 2005;579:1101–8.
25. Kolch W, Calder M, Gilbert D. When kinases meet mathematics: the systems biology of MAPK signalling. FEBS Lett 2005;579:1891–5.

26. Purutçuoğlu V, Wit E. Estimating network kinetics of the MAPK/ERK pathway using biochemical data. Math Probl Eng 2012;2012:1–34.

27. Yeung K, Seitz T, Li S, Janosch P, McFerran B, Kaiser C, et al. Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP. Nature 1999;401:173–7.

28. Hornberg JJ, Binder B, Bruggeman FJ, Schoeberl B, Heinrich R, Westerhoff HV. Control of MAPK signalling: from complexity to what really matters. Oncogene 2005;24:5533–42.

29. Kolch W. Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. Biochem J 2000;351:289–305.

30. Sambanthamurthi R, Sundram K, Tan Y. Chemistry and biochemistry of palm oil. Prog Lipid Res 2000;39:507–58.

31. Oo KC, Teh SK, Khor HT, Ong SH. Fatty acid synthesis in the oil palm (Elaeis guineensis): incorporation of acetate by tissue slice of the developing fruits. Lipids 1985;20:205–10.

32. Quek EM, Purutçuoğlu V, Sambanthamurthi R, Weber GW. Modeling lipid biosynthesis pathways of oil palms by boolean and graphical approaches. Proceeding of the 6th International Symposium on Health, Informatics and Bioinformatics 2011; İzmir, Turkey.