



Preprint-Series: Department of Mathematics - Applied Mathematics

Graphical Lasso Granger method with 2-levels-thresholding for recovering causality networks

Sergiy Pereverzyev Jr. and Katerina Hlavackova-Schindler



Technikerstraße 21a - 6020 Innsbruck - Austria
Tel.: +43 512 507 53803 Fax: +43 512 507 53898
<https://www.applied-math.uibk.ac.at>

Graphical Lasso Granger method with 2-levels-thresholding for recovering causality networks

Sergiy Pereverzyev Jr.¹ and Kateřina Hlaváčková-Schindler²

October, 2013

¹ *Applied Mathematics Group, Department of Mathematics,
University of Innsbruck, Austria*
e-mail: sergiy.pereverzyev@uibk.ac.at

² *Chair of Bioinformatics, University of Natural Resources and Life Sciences,
Vienna, Austria*
e-mail: katerina.schindler@gmail.com

Abstract

The recovery of the causality networks with a number of variables is an important problem that arises in various scientific contexts. For detecting the causal relationships between variables in the network, the concept of the so called multivariate Granger causality has been proposed. Its application to the networks with a big number of variables requires a variable selection procedure. The Lasso is a well known example of such a procedure, and the method for reconstructing causality networks using the multivariate Granger causality with the Lasso is called Graphical Lasso Granger (GLG) method. It is widely believed that the GLG-method tends to overselect causal relationships. In this paper, we propose a thresholding strategy for the GLG-method, which we call 2-levels-thresholding, and we show that with this strategy the variable overselection of the GLG-method may be overcome. Moreover, we demonstrate that the GLG-method with the proposed thresholding strategy may become superior to other methods that were proposed for the recovery of the causality networks.

1 Introduction

Causality is a relationship between a cause and its effect (its consequence). One can say that *inverse problems* solving, where one would like to discover unobservable features of the cause from the observable features of an effect [6], i.e., searching for the cause of an effect, can in general be seen as a causality problem.

A *causality network* is a directed graph with nodes, which are variables $\{x^j, j = 1, \dots, p\}$, and directed edges, which are the causal influences between variables. We write $x^i \leftarrow x^j$ if the variable x^j has a causal influence on the variable x^i . Causality networks arise in various scientific contexts.

For example, In Cell Biology one considers causality networks which involve sets of active genes of a cell. An active gene produces a protein. It has been observed that the amount of the protein which is produced by a given gene may depend on, or may be *causally* influenced by, the amount of the proteins which are produced by other genes. In this way, causal relationships between genes and the corresponding causality network arise. These causality networks are also called gene regulatory networks. An example of such a network is presented in Figure 1. This network is achieved from the biological experiments in [11], and it can be found in the BioGRID database. This network has been used in several works [17, 13, 19] as a test network.

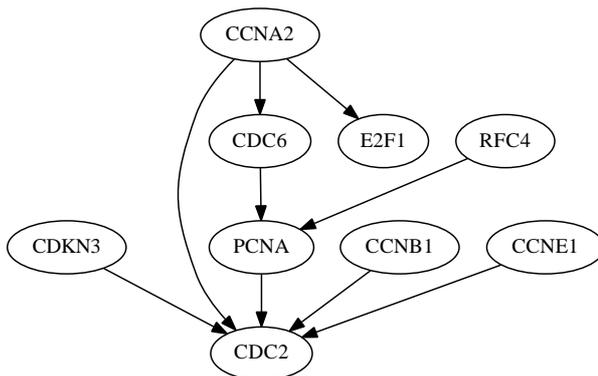


Figure 1: Causality network of the human cancer cell HeLa genes from the BioGRID database (www.thebiogrid.org).

Knowledge of the correct causality networks is important for changing them. In Cell Biology, these networks are used in the research of the causes of genetic diseases. For example, the network in Figure 1 consists of genes that are active in the human cancer cell HeLa [24]. If one wants to suppress the genes expression in this network, then the primary focus of the suppression therapy should be on the causing genes.

Causality networks are also important in Econometrics, where one is interested in the causal relationships between prices of different products and stocks [3]. In Geosciences, causality networks are used for modeling climate teleconnection mechanisms [23, 5, 20].

The principal questions of Neuroscience concern the functioning of the human brain. Here, the causality networks are used for expressing the interactions between

various regions of the brain. These networks can help in understanding of the human cognition as well as in the investigation of the neurological diseases [15, 18, 14].

How can causality network be recovered? In practice, the first information that can be known about the network is the time evolution (time series) of the involved variables $\{x_t^j, t = 1, \dots, T\}$. How can this information be used for inferring causal relationships between the variables?

The statistical approach to the derivation of the causal relationships between a variable y and variables $\{z^j, j = 1, \dots, p\}$ using the known time evolution of their values $\{y_t, z_t^j, t = 1, \dots, T, j = 1, \dots, p\}$ consists in considering a model of the relationship between y and $\{z^j, j = 1, \dots, p\}$. As a first step, one can consider a linear model of this relationship:

$$y_t \approx \sum_{j=1}^p \beta^j z_t^j, \quad t = 1, \dots, T.$$

The coefficients $\{\beta^j, j = 1, \dots, p\}$ can be specified using the least-squares method. Then, in Statistics [25] by fixing the value of a threshold parameter $\beta_{\text{tr}} > 0$, one says that there is a causal relationship $y \leftarrow z^j$ if $|\beta^j| > \beta_{\text{tr}}$.

For detecting causal relationships between variables $\{x^j, j = 1, \dots, p\}$ the concept of the so called *multivariate Granger causality* has been proposed. This concept originated in the work of Clive Granger [8], who was awarded the Nobel Prize in Economic Sciences in 2003. Based on the intuition that the cause should precede its effect, in Granger causality one says that a variable x^i can be potentially caused by the past versions of the involved variables $\{x^j, j = 1, \dots, p\}$.

Then, in the spirit of the statistical approach and using a linear model for the causal relationship, we consider the following approximation problem:

$$x_t^i \approx \sum_{j=1}^p \sum_{l=1}^L \beta_l^j x_{t-l}^j, \quad t = L + 1, \dots, T, \quad (1)$$

where L is the so called *maximal lag*, which is the maximal number of the considered past versions of the variables. The coefficients $\{\beta_l^j\}$ can be determined by the least-squares method. As in the statistical approach, one can now fix the value of the threshold parameter $\beta_{\text{tr}} > 0$ and say that

$$x^i \leftarrow x^j \quad \text{if} \quad \sum_{l=1}^L |\beta_l^j| > \beta_{\text{tr}}. \quad (2)$$

It is well known that for a big number of genes p , as it is pointed out for example in [13], the causality network, which is obtained from the approximation problem (1), is not satisfactory. First of all, it cannot be guaranteed that the solution of the corresponding minimization problem is unique. Another issue is connected with the number of the causality relationships that is obtained from (1). This number is typically very big, while one expects to have a few causality relationships with a given gene. To address this issue, various *variable selection procedures* can be

employed. The Lasso [21] is a well known example of such a procedure. In the regularization theory, this approach is known as l_1 -Tikhonov regularization. It has been extensively used for reconstructing the sparse structure of an unknown signal. We refer the interesting reader to [4, 16, 9, 7, 12] and the references therein.

The causality concept that is based on the Lasso was proposed in [1] and is named *Graphical Lasso Granger (GLG)* method. However, it is stated in the literature that the Lasso suffers from the variable overselection. And therefore, in the context of the gene causality networks several Lasso modifications were proposed. In [13], the so called *group Lasso* method was considered for recovering gene causality networks using the multivariate Granger causality. The corresponding method can be named Graphical group Lasso Granger (GgrLG) method. And in [19], the *truncating Lasso* method was proposed. The resulting method can be named Graphical truncating Lasso Granger (GtrLG) method.

Nevertheless, it seems that an important tuning possibility of the Lasso, namely an appropriate choice of the *threshold parameter* β_{tr} , has been overlooked in the literature devoted to the recovery of the gene causality networks. In this paper, we are going to show that the GLG-method, which is equipped with an appropriate *thresholding strategy* and appropriate *regularization parameter choice rule*, may become a superior method in comparison to other methods that were proposed for the recovery of the gene causality networks.

The paper is organized as follows. In section 2, we recall the GLG-method. The quality measures of the graphical methods are presented in section 3. In section 4, we use the network from Figure 1 to compare the performance of the known graphical methods with the ideal version of the GLG-method, which we call *optimal* GLG-estimator. Such a comparison demonstrates the potential of the GLG-approach. In section 5, we propose a thresholding strategy for the GLG-method that allows its automatic realization, which we describe in section 6. Then again we use the network from Figure 1 to compare the performance of the proposed version of the GLG-method with other graphical methods. It turns out that the proposed method has a superior quality compared to the known methods. The paper is finished with the conclusion and outlook in section 7.

2 Graphical Lasso Granger Method

Let us specify the application of the least-squares method to the approximation problem (1). For this purpose, let us define the vectors

$$Y^i = (x_{L+1}^i, x_{L+2}^i, \dots, x_T^i)',$$

$$\boldsymbol{\beta} = (\beta_1^1, \dots, \beta_L^1, \beta_1^2, \dots, \beta_L^2, \dots, \beta_1^p, \dots, \beta_L^p)',$$

and the matrix

$$X = \left((x_{t-1}^1, \dots, x_{t-L}^1, x_{t-1}^2, \dots, x_{t-L}^2, \dots, x_{t-1}^p, \dots, x_{t-L}^p); t = L + 1, \dots, T \right).$$

Then, in the least-squares method, one considers the following minimization problem:

$$\|Y^i - X\beta\|^2 \rightarrow \min_{\beta}, \quad (3)$$

where $\|\cdot\|$ denotes the l_2 -norm.

As it was mentioned in the introduction, the solution of (3) defines unsatisfactory causal relationships and various variable selection procedures should be employed instead. A well-known example of such procedures is the Lasso [21]. In this procedure, one considers the following minimization problem:

$$\|Y^i - X\beta\|^2 + \lambda\|\beta\|_1 \rightarrow \min_{\beta}. \quad (4)$$

Solution of (4) for each variable $\{x^i, i = 1, \dots, p\}$ with the causality rule (2) defines an estimator of the causality network between the variables $\{x^i\}$, and in this way one obtains the Graphical Lasso Granger (GLG) method [1].

3 Quality measures of the graphical methods

A *graphical method* is a method that reconstructs the causality network, which is a directed graph, with the variables $\{x^j\}$. The *quality* of a graphical method can be estimated from its performance on a *known* causality network. The network in Figure 1 has been used for testing methods' quality in several publications [17, 13, 19]. What measures can be used for estimating the quality of a graphical method?

First of all, let us note that a causality network can be characterized by the so called *adjacency matrix* $A = \{A_{i,j} | \{i, j\} \subset \{1, \dots, p\}\}$ with the following elements:

$$\begin{aligned} A_{i,j} &= 1 && \text{if } x^i \leftarrow x^j, \\ A_{i,j} &= 0 && \text{otherwise.} \end{aligned}$$

The adjacency matrix for the causality network in Figure 1 is presented in Figure 2.

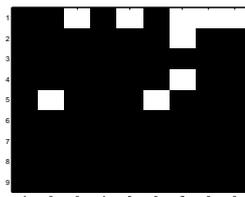


Figure 2: The adjacency matrix for the causality network in Figure 1. The white squares correspond to $A_{i,j} = 1$; the black squares — to the zero-elements. The genes are numbered in the following order: CDC2, CDC6, CDKN3, E2F1, PCNA, RFC4, CCNA2, CCNB1, CCNE1.

Now, imagine that there is a *true* adjacency matrix A^{true} of the true causality network, and there is its *estimator* A^{estim} , which is produced by a graphical method. The elements of the adjacency matrix A^{estim} can be classified as follows.

	P	R	F_1
GgrLG	0.24	0.44	0.3
GtrLG	0.3	0.33	0.32
CNET	0.36	0.44	0.4

Table 1: Quality measures of the known graphical methods.

- If $A_{i,j}^{\text{estim}} = 1$ and $A_{i,j}^{\text{true}} = 1$, then $A_{i,j}^{\text{estim}}$ is called true positive (TP).
- If $A_{i,j}^{\text{estim}} = 0$ and $A_{i,j}^{\text{true}} = 0$, then $A_{i,j}^{\text{estim}}$ is called true negative (TN).
- If $A_{i,j}^{\text{estim}} = 1$ and $A_{i,j}^{\text{true}} = 0$, then $A_{i,j}^{\text{estim}}$ is called false positive (FP).
- If $A_{i,j}^{\text{estim}} = 0$ and $A_{i,j}^{\text{true}} = 1$, then $A_{i,j}^{\text{estim}}$ is called false negative (FN).

Let $\#S$ denote the number of elements in a set S . Then, on the base of the comparison between A^{true} and A^{estim} the following quality measures of the estimator can be defined.

- *Precision* (also called positive predictive value) of A^{estim} :

$$P = \frac{\#\{\text{TP-elements of } A^{\text{estim}}\}}{\#\{\text{TP-elements of } A^{\text{estim}}\} + \#\{\text{FP-elements of } A^{\text{estim}}\}},$$

$$0 \leq P \leq 1.$$

- *Recall* (also called sensitivity) of A^{estim} :

$$R = \frac{\#\{\text{TP-elements of } A^{\text{estim}}\}}{\#\{\text{TP-elements of } A^{\text{estim}}\} + \#\{\text{FN-elements of } A^{\text{estim}}\}},$$

$$0 \leq R \leq 1.$$

Since it is possible to have a high precision and low recall, and vice versa, one considers also an average between these two measures. The so called F_1 -score is defined as the harmonic mean of precision and recall:

$$\frac{1}{F_1} = \frac{1/P + 1/R}{2}.$$

As it was already mentioned, the causality network in Figure 1 has been already used for testing quality of graphical methods. In particular, in [19] one finds the above mentioned quality measures for the following methods: GgrLG, GtrLG and CNET. CNET is a graph search-based algorithm that was introduced in [17]. The data $\{x_t^j\}$ is taken from the third experiment of [24] consisting of 47 time points, and the maximal lag L is taken to be equal to 3. The quality measures from [19] are presented in Table 1.

As it is seen from the table, CNET has the highest F_1 -score. However, CNET is the most computationally expensive among the considered methods that does not allow its application to large networks. GgrLG has a good recall but a poor precision, and thus, GtrLG can be considered as a better method among the considered methods.

4 Optimal GLG-estimator

As we have seen in the previous section, the graphical methods, which are based on the Lasso modifications, were tested on the network in Figure 1 (Table 1). However, the application of the graphical method that is based on the pure Lasso (GLG) to the network in Figure 1 has not been reported. Moreover, it seems that the possibility of varying the threshold parameter β_{tr} in GLG has also not been considered in the literature devoted to the reconstruction of the causality networks.

Assume that the true causality network with the variables $\{x^j\}$ is given by the adjacency matrix A^{true} . Assume further that the observation data $\{x_t^j\}$ is given. What is the *best* reconstruction of A^{true} that can be achieved by the GLG-method? The answer to this question is given by, what we call, the *optimal* GLG-estimator. Let us specify its construction.

Let $\beta_i(\lambda)$ denote the solution of the minimization problem (4) in the GLG-method, and $\beta_i^j(\lambda) = (\beta_{1,i}^j, \dots, \beta_{L,i}^j)$. Then, the GLG-estimator $A^{\text{GLG}}(\lambda, \beta_{\text{tr}})$ of the adjacency matrix A^{true} is defined as follows:

$$\begin{aligned} A_{i,j}^{\text{GLG}}(\lambda, \beta_{\text{tr}}) &= 1 \quad \text{if } \|\beta_i^j(\lambda)\|_1 > \beta_{\text{tr}}, \\ A_{i,j}^{\text{GLG}}(\lambda, \beta_{\text{tr}}) &= 0 \quad \text{otherwise.} \end{aligned}$$

Let $A_{i,*}^{\text{GLG}}(\lambda, \beta_{\text{tr}})$ denote the i -th row of the GLG-estimator. For the given regularization parameter λ , let $\beta_{\text{tr}} = \beta_{\text{tr}}^i(\lambda)$ be the threshold parameter that minimizes the number of false entries in the row $A_{i,*}^{\text{GLG}}(\lambda, \beta_{\text{tr}})$, i.e. the threshold parameter that solves the following minimization problem:

$$\|A_{i,*}^{\text{true}} - A_{i,*}^{\text{GLG}}(\lambda, \beta_{\text{tr}})\|_1 \rightarrow \min_{\beta_{\text{tr}}}.$$

Then, we consider the minimization of the number of false entries with respect to the regularization parameter λ , i.e. let $\lambda_{\text{opt},i}$ solve

$$\|A_{i,*}^{\text{true}} - A_{i,*}^{\text{GLG}}(\lambda, \beta_{\text{tr}}^i(\lambda))\|_1 \rightarrow \min_{\lambda}.$$

In this way, we obtain, what we call, the *optimal* GLG-estimator $A^{\text{GLG,opt}}$ of the true adjacency matrix A^{true} :

$$A_{i,j}^{\text{GLG,opt}} = A_{i,j}^{\text{GLG}}(\lambda_{\text{opt},i}, \beta_{\text{tr}}^i(\lambda_{\text{opt},i})).$$

Note that the optimal GLG-estimator minimizes the following quality measure, which we call *Fs*-measure:

$$Fs = \frac{1}{p^2} \|A^{\text{true}} - A^{\text{estim}}\|_1, \quad 0 \leq Fs \leq 1.$$

Fs-measure represents the number of *false* elements in the estimator A^{estim} that is scaled with the total number of elements in A^{estim} .

The optimal GLG-estimator of the adjacency matrix for the causality network in Figure 1 is presented in Figure 3. We used the same data $\{x_t^j\}$ as in [17, 13, 19].

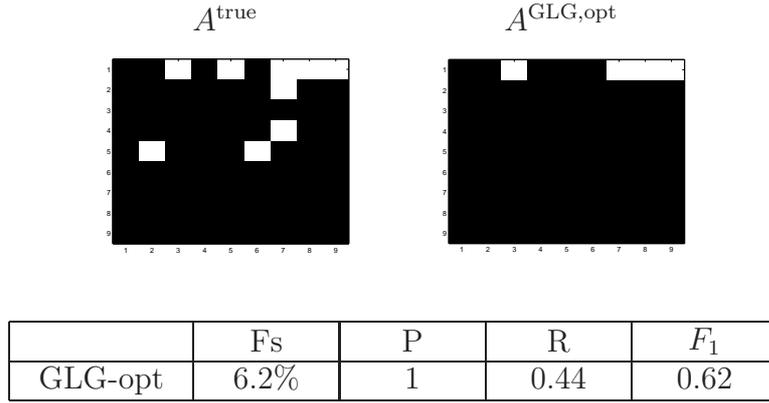


Figure 3: The optimal GLG-estimator and its quality measures.

Also, as in [13, 19], we take the maximal lag $L = 3$. As one can see, the optimal GLG-estimator reconstructs almost completely the causing genes of the most caused gene in the network. The recall of $A^{\text{GLG,opt}}$ is equal to the highest recall in Table 1, but precision and F_1 -score are considerably higher.

Of course, $A^{\text{GLG,opt}}$ is given by the ideal version of the GLG-method, where we essentially use the knowledge of A^{true} . How close can we come to $A^{\text{GLG,opt}}$ without such a knowledge? To answer this question, let us first decide about the choice of the threshold parameter β_{tr} .

5 Thresholding strategy

The purpose of the threshold parameter β_{tr} is to cancel the causal relationships $x^i \leftarrow x^j$ with *small* $\|\beta_i^j(\lambda)\|_1$. When can we say that $\|\beta_i^j(\lambda)\|_1$ is small? We propose to consider the following *guideindicators* of smallness:

$$\begin{aligned} \beta_{\min}^i(\lambda) &= \min\{ \|\beta_i^j(\lambda)\|_1, j = 1, \dots, p \mid \|\beta_i^j(\lambda)\|_1 \neq 0 \}, \\ \beta_{\max}^i(\lambda) &= \max\{ \|\beta_i^j(\lambda)\|_1, j = 1, \dots, p \}. \end{aligned} \quad (5)$$

In particular, we propose to consider the threshold parameter of the following form:

$$\beta_{\text{tr},\alpha}^i(\lambda) = \beta_{\min}^i(\lambda) + \alpha(\beta_{\max}^i(\lambda) - \beta_{\min}^i(\lambda)). \quad (6)$$

As a default value we take $\alpha = 1/2$.

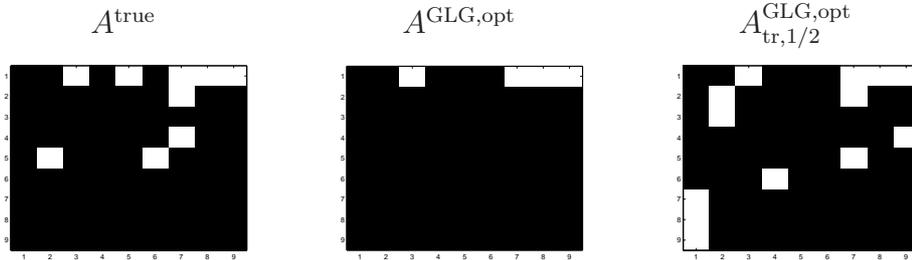
The optimal GLG-estimator with the threshold parameter $\beta_{\text{tr},1/2}^i$ can be defined as follows. Let $\lambda_{\text{opt},i}^{\text{tr},1/2}$ solve the minimization problem:

$$\|A_{i,*}^{\text{true}} - A_{i,*}^{\text{GLG}}(\lambda, \beta_{\text{tr},1/2}^i(\lambda))\|_1 \rightarrow \min_{\lambda}.$$

Then, the corresponding optimal GLG-estimator is

$$A_{\text{tr},1/2}^{\text{GLG,opt}}(i, j) = A_{i,j}^{\text{GLG}}(\lambda_{\text{opt},i}^{\text{tr},1/2}, \beta_{\text{tr},1/2}^i(\lambda_{\text{opt},i}^{\text{tr},1/2})).$$

For the causality network in Figure 1, this estimator is presented in Figure 4. One observes that although there is some quality decrease in comparison to $A^{\text{GLG,opt}}$, the quality measures are still higher than for the methods in Table 1. However, can this quality be improved?



	Fs	P	R	F_1
GLG-opt; tr, 1/2	14.8%	0.38	0.56	0.45

Figure 4: The optimal GLG-estimator that is obtained with the threshold $\beta_{\text{tr},1/2}^i$.

The choice of the threshold parameter $\beta_{\text{tr},1/2}^i$ rises the following issue. With such a choice we always assign a causal relationship, unless the solution of (4) $\beta_i(\lambda)$ is identically zero. But how strong are these causal relationships compared to each other? The norm $\|\beta_i^j(\lambda)\|_1$ can be seen as a *strongness indicator* of the causal relationship $x^i \leftarrow x^j$.

Let us now construct a matrix $A_{\text{tr},1/2}^{\text{GLG,opt};\beta}$, similarly to the adjacency matrix $A_{\text{tr},1/2}^{\text{GLG,opt}}$, where instead of the element 1 we put the norm $\|\beta_i^j(\lambda)\|_1$, i.e.

$$A_{\text{tr},1/2}^{\text{GLG,opt};\beta}(i, j) = \|\beta_i^j(\lambda_{\text{opt},i}^{\text{tr},1/2})\|_1 \quad \text{if} \quad \|\beta_i^j(\lambda_{\text{opt},i}^{\text{tr},1/2})\|_1 > \beta_{\text{tr},1/2}^i,$$

$$A_{\text{tr},1/2}^{\text{GLG,opt};\beta}(i, j) = 0 \quad \text{otherwise.}$$

This matrix is presented in Figure 5. One observes that the false causal relationships of the estimator $A_{\text{tr},1/2}^{\text{GLG,opt}}$ are actually weak. This observation suggests to use a second thresholding that is done on the network, or adjacency matrix, level.

We propose to do the thresholding on the network level similarly to the thresholding on the gene level. Namely, let us define the guideindicators of smallness on the network level similarly to (5):

$$A_{\min} = \min\{ A_{\text{tr},1/2}^{\text{GLG,opt};\beta}(i, j) \neq 0 \},$$

$$A_{\max} = \max\{ A_{\text{tr},1/2}^{\text{GLG,opt};\beta}(i, j) \}.$$

And similarly to (6), define the threshold on the network level as follows:

$$A_{\text{tr},\alpha_1} = A_{\min} + \alpha_1 (A_{\max} - A_{\min}). \tag{7}$$

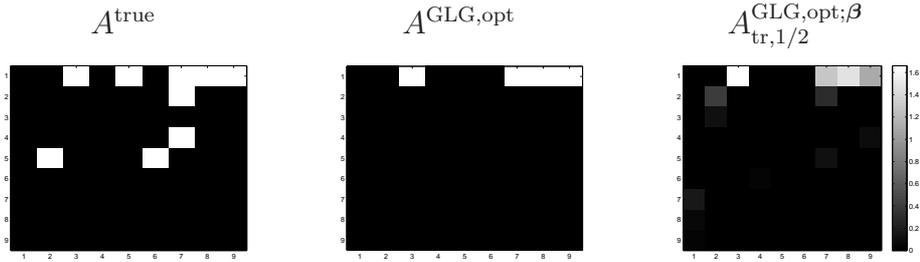


Figure 5: The matrix $A_{\text{tr},1/2}^{\text{GLG,opt};\beta}$ (zero-elements are black, nonzero-elements are gray with the intensity corresponding to their values) in comparison to the adjacency matrices A^{true} and $A^{\text{GLG,opt}}$.

We find it suitable to call the described combination of the two thresholdings on the gene and network levels as *2-levels-thresholding*. The adjacency matrix obtained by this thresholding strategy is the following:

$$A_{\text{tr},1/2;\alpha_1}^{\text{GLG,opt}}(i, j) = 1 \quad \text{if } A_{\text{tr},1/2}^{\text{GLG,opt};\beta}(i, j) > A_{\text{tr},\alpha_1},$$

$$A_{\text{tr},1/2;\alpha_1}^{\text{GLG,opt}}(i, j) = 0 \quad \text{otherwise.}$$

It turns out that with $\alpha_1 = 1/4$ in (7) the optimal GLG-estimator can be fully recovered (Figure 6).

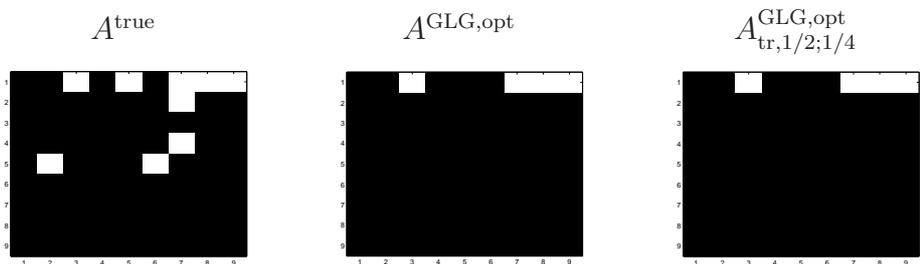


Figure 6: The optimal GLG-estimator that is obtained with the 2-levels-thresholding.

6 An automatic realization of the GLG-method

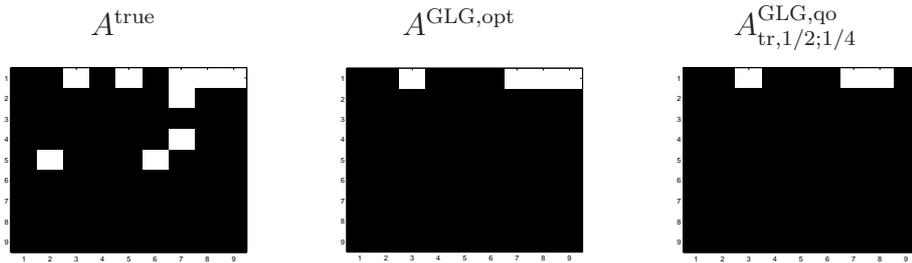
For an automatic realization of the GLG-method, i.e. a realization that does not rely on the knowledge of the true adjacency matrix A^{true} , in addition to a thresholding strategy one needs a choice rule for the regularization parameter λ in (4). For such a choice, we propose to use the so called quasi-optimality criterion [22, 2, 10]. In this criterion, one considers a set of regularization parameters

$$\lambda_k = \lambda_0 q^k, \quad q < 1, \quad k = 0, 1, \dots, n_\lambda,$$

and for each λ_k the corresponding solution of (4) $\beta_i(\lambda_k)$ is computed. Then, the index of the regularization parameter is selected as follows:

$$k_{\text{qo}}^i = \underset{k}{\operatorname{argmin}} \{ \|\beta_i(\lambda_{k+1}) - \beta_i(\lambda_k)\|_1 \}.$$

The reconstruction obtained by the GLG-method with the 2-levels-thresholding and quasi-optimality criterion is presented in Figure 7. One observes that there is a little decrease in recall in comparison to the optimal GLG-method; however, this recall is the same as for the GtrLG-method (Table 1). But due to the highest precision, the F_1 -score remains to be higher than for the methods in Table 1. Thus, one may say that the proposed realization of the GLG-method outperforms the methods in Table 1.



	Fs	P	R	F_1
GLG-qo; tr, 1/2; 1/4	7.4%	1	0.33	0.5

Figure 7: The estimator that is obtained by the proposed automatic realization of the GLG-method.

Nevertheless, one may still wonder, why the proposed realization of the GLG-method captures only the causal relationships of the most caused gene. It appears that the value of the maximal lag L plays an important role in the selection of the causal relationships.

In the modifications of the GLG-method the authors of [13, 19] considered $L = 3$. All results presented so far were also obtained with $L = 3$. It turns out that for $L = 4$ the optimal GLG-estimator in Figure 8 delivers a much better reconstruction of the causality network. In particular, two more caused genes are recovered.

The proposed automatic realization of the GLG-method with $L = 4$ in Figure 9 recovers an additional caused gene in comparison to the realization with $L = 3$. Also, all considered quality measures for our automatic realization of the GLG-method with $L = 4$ are considerably higher than for the methods in Table 1. We would like to stress that no use of the knowledge of A^{true} and no readjustment of the design parameters α, α_1 is necessary for obtaining the results displayed in Figures 7 and 9.

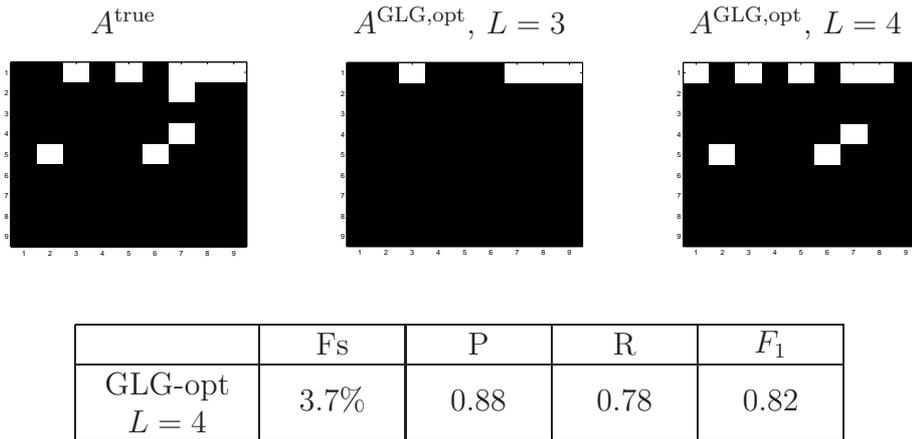


Figure 8: The optimal GLG-estimator with the maximal lag $L = 4$.

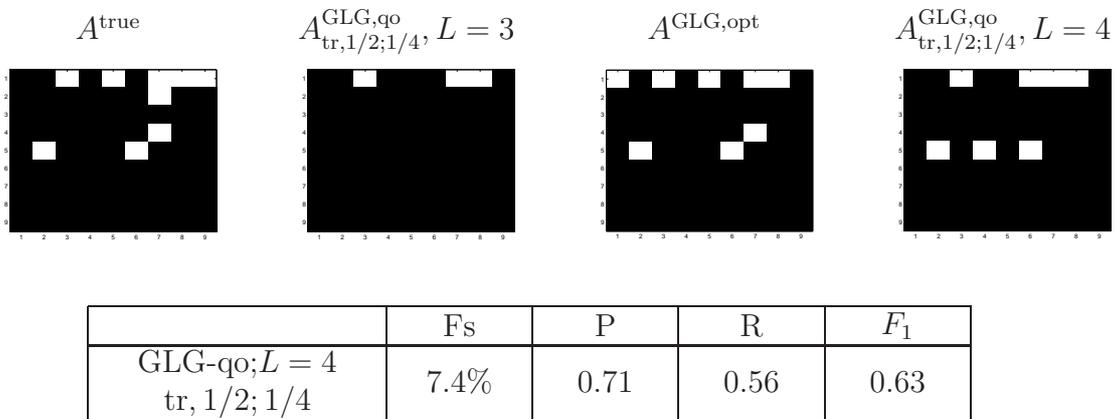


Figure 9: The estimator that is obtained by the proposed automatic realization of the GLG-method with the maximal lag $L = 4$.

7 Conclusion and outlook

The proposed realization of the Graphical Lasso Granger method with 2-levels-thresholding and quasi-optimality criterion for the choice of the regularization parameter shows a considerable improvement of the reconstruction quality in comparison to other graphical methods. So, the proposed realization is a very promising method for recovering causality networks. Further tests and developments of the proposed realization are worthwhile. In particular, applications to larger causality networks are of interest.

As an open problem for the future, one could consider a study of the choice of the maximal lag and its possible variation with respect to the caused and causing genes.

References

- [1] A. Arnold, Y. Liu, and N. Abe. Temporal causal modeling with graphical Granger methods. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 66–75, USA, 2007. ACM New York.
- [2] F. Bauer and M. Reiß. Regularization independent of the noise level: an analysis of quasi-optimality. *Inverse Probl.*, 24(5):16 p., 2008.
- [3] P. Caraiani. Using complex networks to characterize international business cycles. *PLoS ONE*, 8(3):e58109, 2013.
- [4] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.*, 57(11):1413–1457, 2004.
- [5] I. Ebert-Uphoff and Y. Deng. Causal discovery for climate research using graphical models. *J. Clim.*, 25:5648–5665, 2012.
- [6] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*. Kluwer Academic Publishers, Dordrecht, 1996.
- [7] M. Fornasier, editor. *Theoretical foundations and numerical methods for sparse recovery*. Berlin: de Gruyter, 2010.
- [8] C. Granger. Investigating causal relations by econometric models and crossspectral methods. *Econometrica*, 37:424–438, 1969.
- [9] M. Grasmair, M. Haltmeier, and O. Scherzer. Sparse regularization with l^q penalty term. *Inverse Probl.*, 24(5):13, 2008.
- [10] S. Kindermann and A. Neubauer. On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization. *Inverse Probl. Imaging*, 2(2):291–299, 2008.
- [11] X. Li and et al. Discovery of time-delayed gene regulatory networks based on temporal gene expression profiling. *BMC Bioinformatics*, 7(26), 2006.
- [12] D. A. Lorenz, P. Maass, and Q. M. Pham. Gradient descent for Tikhonov functionals with sparsity constraints: theory and numerical comparison of step size rules. *Electronic Transactions on Numerical Analysis*, 39:437–463, 2012.
- [13] A. C. Lozano, N. Abe, Y. Liu, and S. Rosset. Grouped graphical Granger modeling for gene expression regulatory networks discovery. *Bioinformatics*, 25:110–118, 2009.
- [14] D. Marinazzo, M. Pellicoro, and S. Stramaglia. Causal information approach to partial conditioning in multivariate data sets. *Computational and Mathematical Methods in Medicine*, 2012.

- [15] M. Palus and et al. Synchronization and information flow in EEGs of epileptic patients. *IEEE Engineering in Medicine and Biology Magazine*, 20(5):65–71, 2001.
- [16] R. Ramlau and G. Teschke. A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints. *Numer. Math.*, 104(2):177–203, 2006.
- [17] F. Sambo, B. D. Camillo, and G. Toffolo. CNET: an algorithm for reverse engineering of causal gene networks. In *NETTAB2008*, Varenna, Italy, 2008.
- [18] A. K. Seth. Causal connectivity of evolved neural networks during behavior. *Network-Computation in Neural Systems*, 16(1):35–54, 2005.
- [19] A. Shojaie and G. Michailidis. Discovering graphical Granger causality using the truncating lasso penalty. *Bioinformatics*, 26:i517–i523, 2010.
- [20] K. Steinhäuser, A. R. Ganguly, and N. V. Chawla. Multivariate and multiscale dependence in the global climate system revealed through complex networks. *Clim. Dyn.*, 39:889–895, 2012.
- [21] R. Tibshirani. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. B*, 58:267–288, 1996.
- [22] A. N. Tikhonov and V. B. Glasko. Use of the regularization method in non-linear problems. *USSR Comp. Math. Math. Phys.*, 5:93–107, 1965.
- [23] A. A. Tsonis, K. L. Swanson, and P. J. Roebber. What do networks have to do with climate? *Bull. Am. Meteorol. Soc.*, doi:10.1175/BAMS-87-5-585, 2006.
- [24] M. L. Whitfield and et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell*, 13:1977–2000, 2002.
- [25] Wikipedia. Causality — Wikipedia, The Free Encyclopedia, 2013. [Online; accessed 11-October-2013].