Granger Causality for Ill-Posed Problems

Methods, Ideas and Application in Life Sciences

Kateřina Hlaváčková-Schindler

Department of Adaptive Systems
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Prague, Czech Republic

Statistics and Causality, University of Vienna
Austria, May 23-24, 2014

Outline

1. Introduction
2. Causality problems in neurology (brain research) and biology (gene regulatory networks)
3. Brief introduction to inverse problems
4. Granger causality and multivariate Granger causality
5. Our two-level thresholding method
6. Application to a gene regulatory network and comparison to other methods
7. Conclusion and outlook

Introduction

THE ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

Concept of causality by Granger 1969

1. an econometric concept
2. a probabilistic concept, relating to the ability of one time series to predict another one, conditional on a given information set;

Concept of causality by Pearl 2000

1. a direct causality concept
2. addresses interventions rather of functional than of probabilistic dependence

White, Chalak, Lu, 2011:

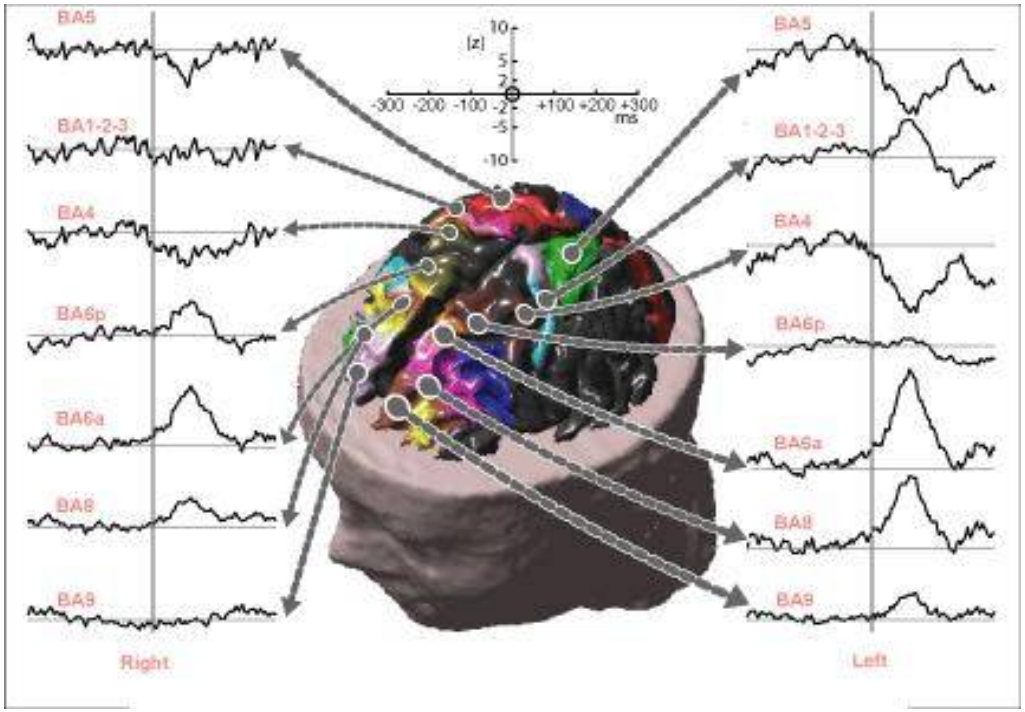
Granger and Pearl's concepts are in fact closely linked: straighforward practical methods to test direct causality in Pearl's sense using Granger causality were proposed.

Bivariate Granger causality; multivariate Granger causality:

Modeling interactions among regions of interest (ROI)

For analysis of causal interactions within human brain regions based on EEG or MR signal (time series), these models are in practice used:

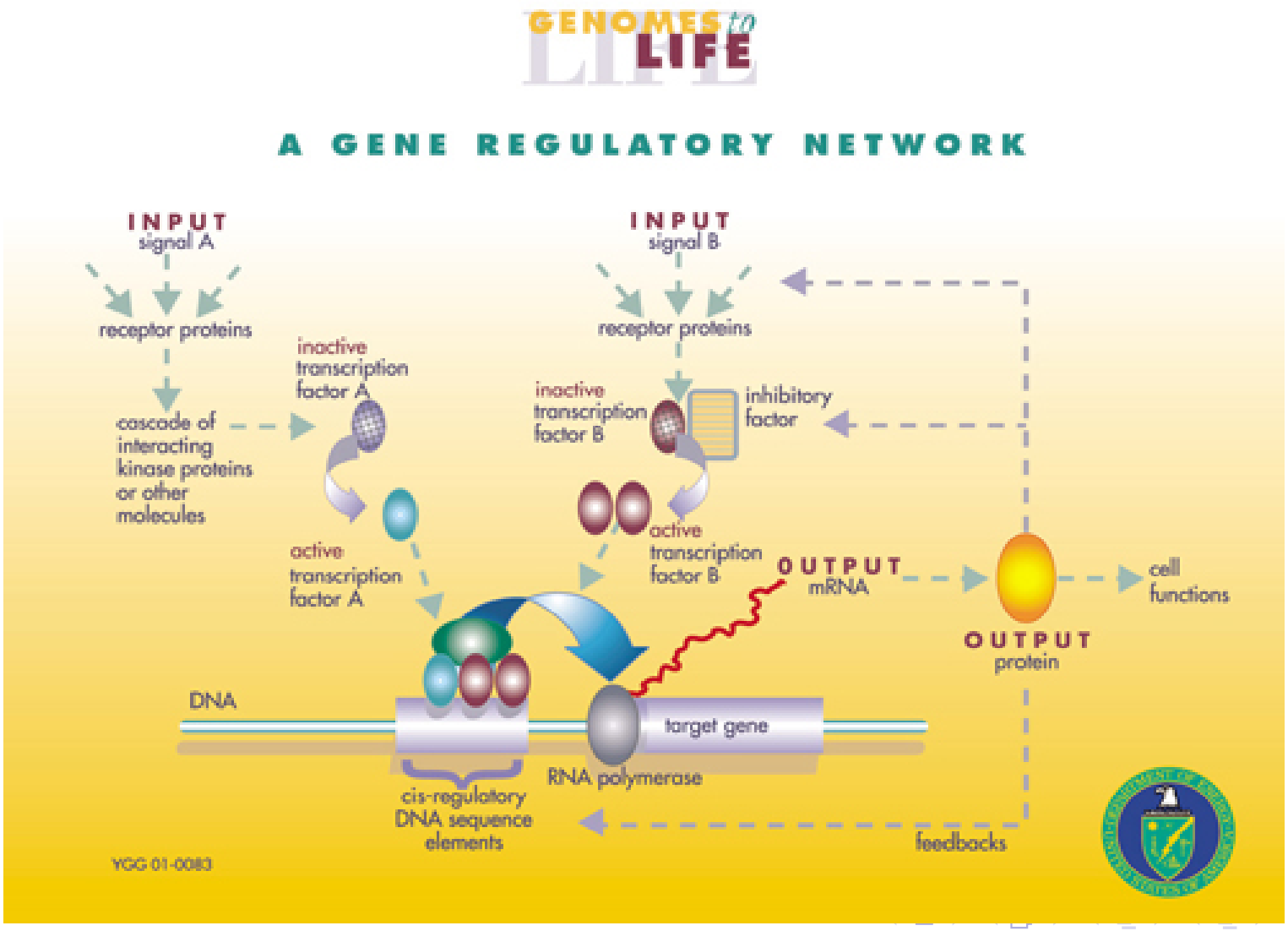
Granger causal model Granger 1969 and its variations, **spectral G-causality** Geweke 1982 or **transfer entropy** Schreiber 2000



A gene regulatory network or genetic regulatory network (GRN)

- is a collection of DNA segments in a cell which interact with each other indirectly (through their RNA and protein expression products) and with other substances in the cell, thereby governing the expression levels of mRNA and proteins.
- Biological cells can be thought of as "partially mixed bags" of biological chemicals: these chemicals are mostly the mRNAs and proteins that arise from gene expression. These mRNA and proteins interact with each other with various degrees of specificity - gene regulatory networks. (Source: Wikipedia)
- Gene expression is the process by which the information encoded in a gene is used to direct the assembly of a protein molecule. These products are often proteins.

Structure of a gene regulatory network (source: Wikipedia)



Application in genetics

A gene regulatory network - a formal simplification

1. directed graph:
   - nodes - genes
   - edges - interactions between genes
2. Data:
   each node is given by a time series of gene expressions (expressions of proteins)
3. Unknown - the directed edges

Gene regulatory networks

Methods for recovery of gene regulatory networks:

- Dynamic Bayesian networks, structural equation models, probabilistic Boolean networks, petri nets, graphical Gaussian models, fuzzy controls and differential equations. - A good modeling of small regulatory networks for which biological information is available (local kinetics).
- For large networks: Due to small data sets (observations) and high dimensionality of the microarrays (a collection of microscopic DNA spots attached to a solid surface) these models suffer from curse of dimensionality (exp. search space) and the related parameter estimation problems are ill-posed.

An alternative which overcomes these deficits:
Graphical Granger models

A short introduction to Inverse Problems

Definition by Engl, Hanke, Neubauer 1996:
Inverse problems are concerned with determining causes for a desired or an observed effect.
The inverse problem (IP) is considered the "inverse" to the direct problem which relates the model parameters to the data that we observe:

direct problem

$\Rightarrow$

CAUSE	EFFECT
parameters, unknown solution of IP	data, observations

$\Leftarrow$

inverse problem

Direct and Inverse Problems

The classification as direct or inverse is in the most cases based on the well/ill-posedness of the associated problems:

$$\begin{array}{c} \text{stable} \\ \Rightarrow \\ \text{Cause} \quad \text{Effect} \\ \Leftarrow \\ \text{unstable} \end{array}$$

Inverse problems $\rightsquigarrow$ Ill-posed/(Ill-conditioned) problems

Well posed and Ill posed problems

- A central feature of inverse problems is their ill-posedness
- Well-Posedness in the sense of Hadamard 1923
  - Existence of a solution (for all admissible data)
  - Uniqueness of a solution
  - Continuous dependence of solution on the data
- Well-Posedness in the sense of Nashed 1987
  A problem is well posed if the set of data/observations is a closed set. (The range of the forward operator is closed).

Abstract inverse problem:

Solve equation for $x \in X$ (Banach, Hilbert space), given data $y \in Y$ (Banach, Hilbert space)

$$F(x) = y$$

where F^{-1} does not exist or is not continuous.

F is the forward operator and we want to find $x^\dagger$ the (generalized) solution so that

$$x^\dagger = F^{-1}(y).$$

If the forward operator is linear then it is a linear inverse problem.

Well-posed and Ill-posed problems

- Problems that are not well-posed in the Hadamard sense are ill-posed.
- Inverse problems are often ill-posed
  (the solution is highly sensitive to changes in data).
- Continuum models must often be discretized in order to obtain a numerical solution. While solutions may be continuous w. r. t. the initial conditions, they may suffer from numerical instability when solved with finite precision, or with errors in the data.
- Even if a problem is well-posed, it may still be ill-conditioned
  (i.e. a small error in the initial data can result in much larger errors in the answers.)
- An ill-conditioned problem is indicated by a large condition number.

Methods to solve Ill-posed problems

- If the problem is well-posed then it stands a good chance of solution on a computer using a stable algorithm.
- If it is not well-posed, it needs to be reformulated for numerical treatment. Typically this involves including additional assumptions, such as smoothness of solution.
- This process is known as regularization.
- The most applied regularization methods are Tikhonov regularization (ridge regression), elastic nets, Lasso or bridge regression (l_q norms).

Penalty as a stabilizator of the least means square problem (LMS):
If parameters (the solution) of LMS are unconstrained, they have very high variance. To control variance, we might regularize the coefficients by imposing the penalty

- l_2 penalty - Tikhonov regularization
  (ridge regression in statistics)

$$\|y - X\beta\|^2 + \lambda\|\beta\|_2^2 \to \min_{\beta}. \tag{1}$$

  where $\|\beta\|_2^2 = \sum_j \beta_j^2$ and λ is a tuning parameter controling the amount of regularization.

- l_1 penalty - Lasso penalty
  (sometimes called l_1-Tikhonov regularization)

Lasso Penalty

- l_1 penalty - Lasso penalty
  (Least absolute shrinkage and selection operator)
  introduced by Tibshirani in 1996
- LASSO coefficents are the solutions to the l_1 optimization problem:

$$\|y - X\beta\|^2 + \lambda\|\beta\|_1 \to \min_{\beta}. \tag{2}$$

  where $\|\beta\|_1 = \sum_j |\beta_j|$ and λ is a tuning parameter controling the amount of regularization.
- Often we believe that many of the β_j's should be 0.
- Hence we look for a set of sparse solutions.
- Large enough λ will set some coefficents exactly equal to 0.
- Lasso is therefore a model selection method. Its tendency to overselect variables will be addressed by our two-level thresholding algorithm.

Back to genetics

Consider a **gene regulatory network** of 19 genes, playing a substantial role at the human ovarian cancer, with names: PCNA, NPAT, E2F1, CCNE1, CDC25A, CDKN1A, BRCA1, CCNF, CCNA2, CDC20, STK15, BUB1B, CKS2, CDC25C, PLK1, CCNB1, CDC25B, TYMS, DHFR.

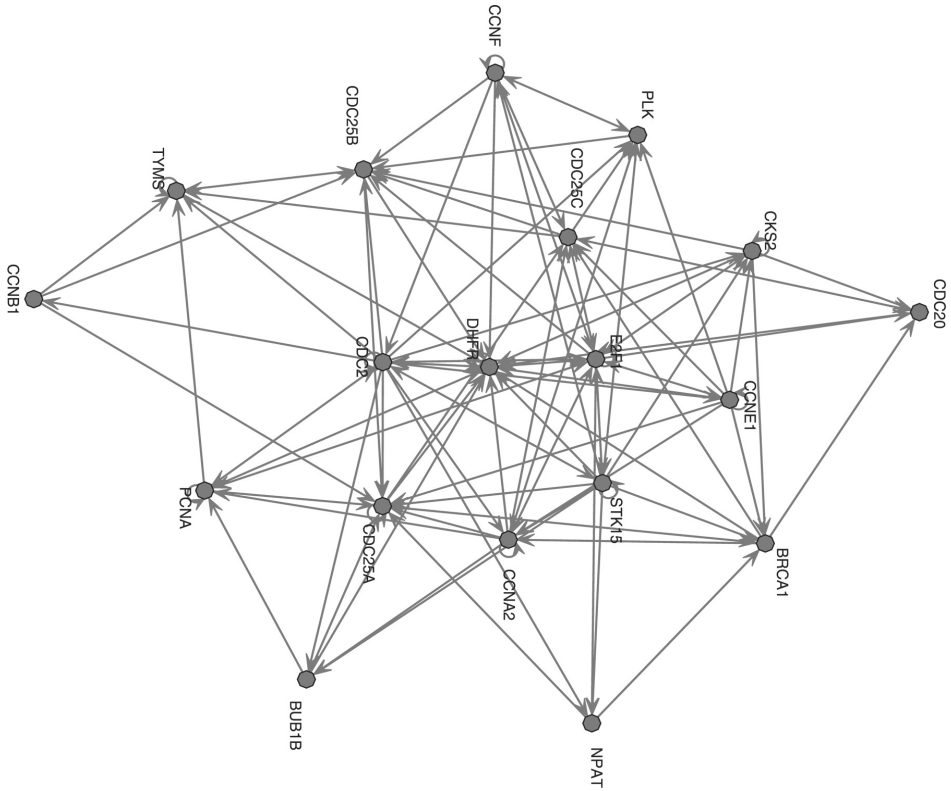


Figure : The causal structure for these genes identified by biological experiments of Li et al. 2006.

Gene regulatory network (GRN)

1. For each of 19 genes (graph nodes) given gene expressions for 48 time observations (data).
   Problem: Find the values of edges of the graph on the nodes.
2. The problem is ill-posed - Multivariate Granger causality (MVAR) will not suffice
3. Graphical Granger models = multivariate Granger causality model with various regularizations:
   - Lozano et al. 2009 Grouped Granger method
     - uses group elastic net (combination of l_1 and l_2 penalties)
   - Arnold et al 2007 (Lasso)
   - Shojaie 2010, 2012 Truncated Lasso, Adaptive thresholding 2010, 2012
4. Our approach:
   Two-level thresholding for reconstructing gene regulatory networks
   Pereverzyev, Hlaváčková-Schindler, 2013

THE ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

Granger Causality

based on the probabilistic notion of causality, is defined:

An event X is a cause to the event Y if

(i) X occurs before Y,

(ii) likelihood of X is non zero, and

(iii) likelihood of occurring Y given X is more than the likelihood of Y occurring alone.

It is said that the process X_t Granger causes another process Y_t if future values of Y_t can be better predicted using the past values of X_t and Y_t rather then only past values of Y_t.

The standard test of GC is based on
a linear vector auto-regressive model (VAR)

$$Y_t = a_o + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \sum_{k=1}^{L} b_{2k} X_{t-k} + \xi_t, \tag{3}$$

where ξ_t are uncorrelated random variables with zero mean and variance σ^2, L is the specified number of time lags, and time $t = L+1, \ldots, N$.
The null hypothesis that X_t does not Granger cause Y_t is supported when $b_{2k} = 0$ for $k = 1, \ldots, L$, reducing Eq. (3) to

$$Y_t = a_o + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \tilde{\xi}_t. \tag{4}$$

This model leads to two test statistics, the Granger-Sargent and the Granger-Wald test.

Extension of GC to more than two time series:
A multivariate vector autoregressive model MVAR: graphical Granger models.
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 $\{x^j,\ j = 1, \ldots, p\}$.
Then, in the spirit of the statistical approach we consider the following approximation problem:

$$x_t^i \approx \sum_{j=1}^{p} \sum_{l=1}^{L} \beta_l^j x_{t-l}^j, \quad t = L+1, \ldots, T, \tag{5}$$

where L is the so called *maximal lag*, which is the maximal number of the considered past versions of the variables.

Multivariate Granger causality threshold parameter

The approximation problem $x_t^i \approx \sum_{j=1}^{p} \sum_{l=1}^{L} \beta_l^j x_{t-l}^j, t = L+1, \ldots, T$, can be specified using the least-squares approach:

$$\sum_{t=L+1}^{T} (x_t^i - \sum_{j=1}^{p} \sum_{l=1}^{L} \beta_l^j x_{t-l}^j)^2 \to \min_{\beta_l^j}.$$

Then, the coefficients $\{\beta_l^j\}$ can be determined from a system of linear equations. As in the statistical approach, one can fix the value of the threshold parameter $\beta_{tr} > 0$ and say that

$$x^i \leftarrow x^j \quad \text{if} \quad \sum_{l=1}^{L} |\beta_l^j| > \beta_{tr}. \tag{6}$$

Multivariate Granger causality threshold parameter

- For a big number of genes p, the causality network, obtained from the approximation problem (5), is not satisfactory.
- First, it cannot be guaranteed that the solution of the corresponding minimization problem is unique.
- Second, the number of the causality relationships obtained from (5) is typically very big, while one expects to have a few causality relationships with a given gene.
- To address these issues, various variable selection procedures can be employed.

Multivariate Granger causality

- Lasso penalty Tibshirani et al 1996
  is extensively used for reconstructing the sparse structure of an unknown signal.
- The causality concept based on Lasso was proposed by Arnold et al. 2007: Graphical Lasso Granger (GLG) method.
- However, literature refers about that the Lasso suffers from the variable overselection. Therefore, in the context of the gene causality networks several Lasso modifications proposed:
- E.g. graphical Group Lasso Granger model Lozano et al, 2009 (GgrLG), using elastic net or
- Graphical truncating Lasso Granger method was proposed by Shojaie et al 2010 (GtrLG).

1. An important tuning possibility of the Lasso, namely an appropriate choice of the threshold parameter β_{tr} in (4), has been overlooked in the literature on recovery of the gene causality networks.
2. We show on a practical example 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 proposed for the recovery of the gene causality networks, namely GgrLG, GtrLG, Copula Granger and ODE-DBN model.

Define the vectors $Y^i = (x^i_{L+1}, x^i_{L+2}, \ldots, x^i_T)^T$,
$\boldsymbol{\beta} = (\beta^1_1, \ldots, \beta^1_L, \ldots, \beta^p_1, \ldots, \beta^p_L)^T$, and the matrix

$$X = \left((x^1_{t-1}, \ldots, x^1_{t-L}, \ldots, x^p_{t-1}, \ldots, x^p_{t-L});\ t = L+1, \ldots, T \right).$$

Consider the following minimization problem, $\|\cdot\|$ denotes the l_2-norm:

$$\|Y^i - X\boldsymbol{\beta}\|^2 \to \min_{\boldsymbol{\beta}}, \tag{7}$$

Solution of (7) defines unsatisfactory causal relationships
$\Rightarrow$ we apply a variable selection procedure with Lasso:

$$\|Y^i - X\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_1 \to \min_{\boldsymbol{\beta}}. \tag{8}$$

Solution of (8) for each x^i, $i = 1, \ldots, p$ with the causality rule

"$x^i \leftarrow x^j$ if $\sum_{l=1}^{L} |\beta^j_l| > \beta_{tr}$" defines an estimator of the causality among $\{x^i\}$,$\Rightarrow$ one obtains the Graphical Lasso Granger (GLG) from Arnold.

Quality measures of the graphical methods

THE ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

- The quality of a graphical method can be estimated from its performance on the known causality network. A causality network is characterized by the adjacency matrix $A = \{A_{i,j} \mid i,j = 1,\ldots,p\}$: $A_{i,j} = 1$ if $x^i \leftarrow x^j$, $A_{i,j} = 0$ otherwise.
- Denote true adjacency matrix A^{true} of the true causality network and its estimator A^{estim}.
  The elements of A^{estim} can be classified as follows:
  - If $A^{\text{estim}}_{i,j} = 1$ and $A^{\text{true}}_{i,j} = 1$, then $A^{\text{estim}}_{i,j}$ is called true positive (TP).
  - If $A^{\text{estim}}_{i,j} = 0$ and $A^{\text{true}}_{i,j} = 0$, then $A^{\text{estim}}_{i,j}$ is called true negative (TN).
  - If $A^{\text{estim}}_{i,j} = 1$ and $A^{\text{true}}_{i,j} = 0$, then $A^{\text{estim}}_{i,j}$ is called false positive (FP).
  - If $A^{\text{estim}}_{i,j} = 0$ and $A^{\text{true}}_{i,j} = 1$, then $A^{\text{estim}}_{i,j}$ is called false negative (FN).

Quality measures of the graphical methods

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

$$P = \frac{TP}{TP + FP}, 0 \leq P \leq 1.$$

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

$$R = \frac{TP}{TP + FN}, 0 \leq R \leq 1.$$

- Classification accuracy of A^{estim}:

$$CA = \frac{TP + TN}{(TP + TN + FP + FN)}, 0 \leq CA \leq 1.$$

Optimal GLG-estimator

Varying the threshold parameter β_{tr} in GLG model
- not considered in the literature yet.
Given the true causality network with $\{x^j\}$ by A^{true}
and the observations $\{x_t^j\}$.
What is the best reconstruction of A^{true} that can be achieved by the GLG-method?
Let $\boldsymbol{\beta}_i(\lambda)$, $\lambda \in (0, \lambda_{max})$ denote the solution of the problem

$$\|Y^i - X\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_1 \to \min_{\boldsymbol{\beta}}$$

in the GLG-method, and $\boldsymbol{\beta}_i^j(\lambda) = (\beta_{1,i}^j, \ldots, \beta_{L,i}^j)$.
Then, the GLG-estimator $A^{\mathrm{GLG}}(\lambda, \beta_{\mathrm{tr}})$ of matrix A^{true} with $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_p)$, $\boldsymbol{\beta}_{tr} = (\beta_{tr,1}, \beta_{tr,2}, \ldots, \beta_{tr,p})$ is defined by:

$$A_{i,j}^{\mathrm{GLG}}(\lambda, \beta_{\mathrm{tr}}) = 1 \text{ if } \quad \|\boldsymbol{\beta}_i^j(\lambda_i)\|_1 > \beta_{tr,i},$$

$$A_{i,j}^{\mathrm{GLG}}(\lambda, \beta_{\mathrm{tr}}) = 0 \text{ otherwise.}$$

Optimal GLG-estimator

Denote $A^{\mathrm{GLG}}_{i,*}(\lambda, \beta_{\mathrm{tr}})$ the i-th row of the GLG estimator.
For given reg. parameter λ, let $\beta^i_{\mathrm{tr}}(\lambda)$ be the threshold parameter that minimizes the number of false entries in the row $A^{\mathrm{GLG}}_{i,*}(\lambda, \beta_{\mathrm{tr}})$,

$$\|A^{\mathrm{true}}_{i,*} - A^{\mathrm{GLG}}_{i,*}(\lambda, \beta_{\mathrm{tr}})\|_1 \to \min_{\beta_{\mathrm{tr}}}. \tag{9}$$

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

$$\|A^{\mathrm{true}}_{i,*} - A^{\mathrm{GLG}}_{i,*}(\lambda, \beta^i_{\mathrm{tr}}(\lambda))\|_1 \to \min_{\lambda}. \tag{10}$$

In this way, we obtain, what we call, the optimal Graphical Lasso Granger estimator $A^{\mathrm{GLG,opt}}$ of the true adjacency matrix A^{true}:

$$A^{\mathrm{GLG,opt}}_{i,j} = A^{\mathrm{GLG}}_{i,j}(\ \lambda_{\mathrm{opt},i}, \beta^i_{\mathrm{tr}}(\lambda_{\mathrm{opt},i})\ ).$$

Thresholding strategy: the first thresholding

- $A^{\mathrm{GLG,opt}}$ is given by the ideal version of the GLG-method, where A^{true} is known.
- How close can we come to $A^{\mathrm{GLG,opt}}$ without knowledge of A^{true}?

First decide about the choice of the threshold parameter β_{tr}.
The purpose of the threshold parameter β_{tr} is to cancel the causal relationships $x^i \leftarrow x^j$ with small $\|\hat{\beta}_i^j(\lambda)\|_1$. We propose to consider $\beta_{min,i}(\lambda)$, $\beta_{max,i}(\lambda)$ as guideindicators of smallness:

$$\begin{aligned} \beta_{min,i}(\lambda) &= \min\{\ \|\hat{\beta}_i^j(\lambda)\|_1,\ j = 1,\dots,p \mid \|\hat{\beta}_i^j(\lambda)\|_1 \neq 0\ \}, \\ \beta_{max,i}(\lambda) &= \max\{\ \|\hat{\beta}_i^j(\lambda)\|_1,\ j = 1,\dots,p\ \}. \end{aligned} \tag{11}$$

In particular, we propose to consider the threshold parameter as:

$$\beta^i_{\mathrm{tr},\alpha}(\lambda) = \beta_{min,i}(\lambda) + \alpha(\ \beta_{max,i}(\lambda) - \beta_{min,i}(\lambda)\ ). \tag{12}$$

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

Thresholding strategy: the first thresholding

The optimal GLG-estimator for each gene i with the threshold parameter $\beta^i_{\mathrm{tr},1/2}$ can be defined as follows.

Let $\lambda^{\mathrm{tr},1/2}_{\mathrm{opt},i}$ solve the minimization problem for each gene i:

$$\|A^{\mathrm{true}}_{i,*} - A^{\mathrm{GLG}}_{i,*}(\lambda, \beta^i_{\mathrm{tr},1/2}(\lambda))\|_1 \to \min_{\lambda}.$$

Then, the corresponding optimal GLG-estimator (i.e. for all genes) is

$$A^{\mathrm{GLG,opt}}_{\mathrm{tr},1/2}(i,j) = A^{\mathrm{GLG}}_{i,j}(\ \lambda^{\mathrm{tr},1/2}_{\mathrm{opt},i}, \beta^i_{\mathrm{tr},1/2}(\lambda^{\mathrm{tr},1/2}_{\mathrm{opt},i})\ ).$$

The second thresholding: on the network level

The choice of the threshold parameter $\beta^i_{\mathrm{tr},1/2}$ rises the following issue. A gene is assigned always a causal relationship, unless the solution of $\|Y^i - X\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_1 \to \min_{\boldsymbol{\beta}}$, $\boldsymbol{\beta}_i(\lambda)$ is identically zero.

But how strong are these causal relationships compared to each other?

The norm $\|\boldsymbol{\beta}^j_i(\lambda)\|_1$ can be seen as a strongness indicator of the causal relationship $x^i \leftarrow x^j$.

Let us now construct a matrix $A^{\mathrm{GLG,opt};\boldsymbol{\beta}}_{\mathrm{tr},1/2}$, similarly to the adjacency matrix $A^{\mathrm{GLG,opt}}_{\mathrm{tr},1/2}$, where instead of 1 we put the norm $\|\boldsymbol{\beta}^j_i(\lambda)\|_1$, i.e.

$$\begin{aligned} A^{\mathrm{GLG,opt};\boldsymbol{\beta}}_{\mathrm{tr},1/2}(i,j) &= \|\boldsymbol{\beta}^j_i(\lambda^{\mathrm{tr},1/2}_{\mathrm{opt},i})\|_1 \quad \text{if} \quad \|\boldsymbol{\beta}^j_i(\lambda^{\mathrm{tr},1/2}_{\mathrm{opt},i})\|_1 > \beta^i_{\mathrm{tr},1/2}, \\ A^{\mathrm{GLG,opt};\boldsymbol{\beta}}_{\mathrm{tr},1/2}(i,j) &= 0 \quad \text{otherwise.} \end{aligned} \tag{13}$$

The second thresholding: on the network level

We propose to do the thresholding on the network level similarly to the thresholding on the gene level. Define the guideindicators of smallness on the network level:

$$A_{\min} = \min\{\, A^{\mathrm{GLG,opt};\boldsymbol{\beta}}_{\mathrm{tr},1/2}(i,j) \neq 0 \,\}, A_{\max} = \quad \max\{\, A^{\mathrm{GLG,opt};\boldsymbol{\beta}}_{\mathrm{tr},1/2}(i,j) \,\}.$$

And similarly to (12), define the threshold on the network level as follows:

$$A_{\mathrm{tr},\alpha_1} = A_{\min} + \alpha_1(\, A_{\max} - A_{\min}\,). \tag{14}$$

We call the described combination of the two thresholdings on the gene and network levels as two-level thresholding. The adjacency matrix obtained by this thresholding strategy is the following:

$$\begin{aligned} A^{\mathrm{GLG,opt}}_{\mathrm{tr},1/2;\alpha_1}(i,j) &= 1 \quad \text{if} \quad A^{\mathrm{GLG,opt};\boldsymbol{\beta}}_{\mathrm{tr},1/2}(i,j) > A_{\mathrm{tr},\alpha_1}, \\ A^{\mathrm{GLG,opt}}_{\mathrm{tr},1/2;\alpha_1}(i,j) &= 0 \quad \text{otherwise.} \end{aligned} \tag{15}$$

An automatic realization of GLG-method

- Case when the true adjacency matrix A^{true} is not known
- In addition to a thresholding strategy one needs a choice rule for the regularization parameter λ in $\|Y^i - X\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_1 \to \min_{\boldsymbol{\beta}}$.

For such a choice, we propose to use the so called quasi-optimality criterion Tikhonov, Glasko 1965, Bauer, Reiss, 2008. One considers a set of regularization parameters

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

and for each λ_k the corresponding solution of (7) $\boldsymbol{\beta}_i(\lambda_k)$ is computed. Then, the index of the regularization parameter is selected as follows:

$$k^i_{\text{qo}} = \operatorname*{argmin}_k \{\ \|\boldsymbol{\beta}_i(\lambda_{k+1}) - \boldsymbol{\beta}_i(\lambda_k)\|_1\ \}. \tag{16}$$

Summary:

1. for $i = 1, 2, \ldots, p$ do
   1. for $k = 0, 1, \ldots, n_\lambda$ do
   2. compute the solution $\beta_i(\lambda_k)$of $\|Y^i - X\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_1 \to \min_{\boldsymbol{\beta}}$;
   3. if $k > 0$, compute the norm $\|\boldsymbol{\beta}_i(\lambda_{k+1}) - \boldsymbol{\beta}_i(\lambda_k)\|_1$; end for k
2. compute $k^i_{\rm qo} = \operatorname{argmin}_k\{\ \|\boldsymbol{\beta}_i(\lambda_{k+1}) - \boldsymbol{\beta}_i(\lambda_k)\|_1\ \}$; denote $\lambda_{qo} := \lambda_{k^i_{\rm qo}}$;
3. compute the corresp. solution of $\|Y^i - X\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_1 \to \min_{\boldsymbol{\beta}}$
   $\beta_{i,qo} := \boldsymbol{\beta}_i(\lambda_{qo})$;
4. with threshold $\beta^i_{tr,1/2}(\lambda_{qo}) = \beta_{min,i}(\lambda) + \alpha(\beta_{max,i}(\lambda) - \beta_{min,i}(\lambda))$
   compute the i-th row of matrix $A^{GLG,qo,\beta}_{tr,1/2}$ as
   in $A^{\mathrm{GLG,opt};\boldsymbol{\beta}}_{\mathrm{tr},1/2}(i,j) = \|\boldsymbol{\beta}^j_i(\lambda^{\mathrm{tr},1/2}_{\mathrm{opt},i})\|_1$ if $\|\boldsymbol{\beta}^j_i(\lambda^{\mathrm{tr},1/2}_{\mathrm{opt},i})\|_1 > \beta^i_{\mathrm{tr},1/2}$,
   $A^{\mathrm{GLG,opt};\boldsymbol{\beta}}_{\mathrm{tr},1/2}(i,j) = 0$ otherwise. end for i
5. with threshold $A_{tr,1/4}$ in $A_{tr,\alpha_1} = A_{min} + \alpha_1(A_{max} - A_{min})$ computed for matrix $A^{GLG,qo,\beta}_{tr,1/2}$, compute finally the adjacency matrix $A^{GLG,qo}_{tr,1/2;1/4}$ as in (15).

Application: Reminder of the gene regulatory network

Consider a **gene regulatory network** of 19 genes (slide 17)

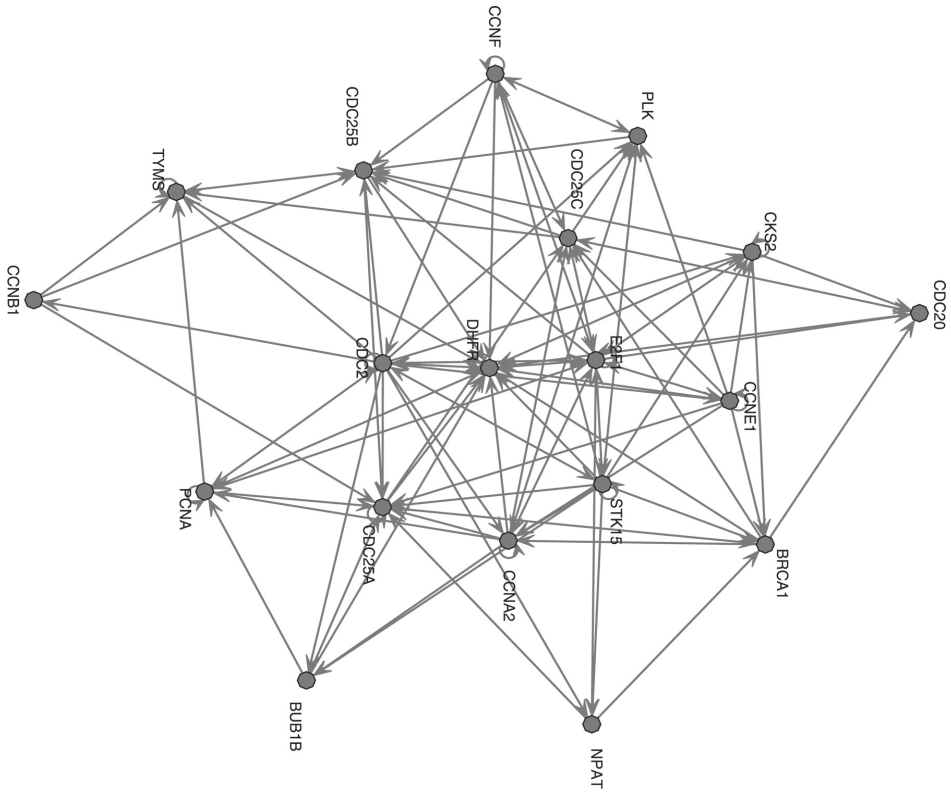


Figure : The causal structure for these genes identified by biological experiments of Li et al. 2006.

Application and comparison

For the network of 19 genes we compared the following methods:

- GP4GRN Dynamic Bayesian network with ordinary differential equations from Äijö and Lähdesmäki 2009
- CG Copula Granger Method from Bahadori and Liu 2013
- LG Lasso Granger method Arnold et al 2007
- Alternatives of our method of two-level thresholding:
  - LG1 Lasso Granger method with optimized threshold and regularization parameter
  - LG2 Lasso Granger method with optimized threshold $\beta^i_{\mathrm{tr},\alpha}(\lambda) = \beta_{min,i}(\lambda) + \alpha(\ \beta_{max,i}(\lambda) - \beta_{min,i}(\lambda)\ )$. with $\alpha = 1/4$ and optimized regularization parameter
  - LG3 Automatic realization of the Lasso Granger method without the knowledge of the true adjacency matrix and with threshold $\beta^i_{\mathrm{tr},\alpha}(\lambda) = \beta_{min,i}(\lambda) + \alpha(\ \beta_{max,i}(\lambda) - \beta_{min,i}(\lambda)\ )$. with $\alpha = 1/4$ and reg. parameter chosen by quasi-optimality criterion

Data and Experiments

- 19 genes, each given by gene expressions for 48 time observations
- Given the gene regulatory network, achieved by experiments from Li et al. 2006
- Matlab codes
- TP - the number of true positives, CA - the classification accuracy were tested
- LG1, LG2, LG3 gave best results for $\alpha = 1/4$.
- real time: GP4GRN - 38 minutes for 19 genes,
  other methods a few seconds (PC workstation with 64-bit processor);
  CG better precision than GP4GRN but worse than all LGi methods.

Comparison

	GP4GRN	LG	CG	LG1	LG2	LG3
CA	0.7507	0.5789	0.80066	0.8753	0.8532	0.8116
TP	95	38	58	63	51	42

  Table : Abbreviations and quality measures for the considered methods

- Classification accuracy of A^{estim}:

$$CA = (TP + TN)/(TP + TN + FP + FN), 0 \leq CA \leq 1.$$

- Computational complexity of the methods

GP4GRN	LG	CG	LGi, $i = 1, \ldots, 3$
exp	$O(nd^2p^2)$	polyn	$O(N_{tr}.N_{\lambda}nd^2p^2)$

Table : n is the number of observations, p the number of genes, d the order of VAR model; We consider a set with N_{tr} values for β_{tr}, and a set with N_{λ} values for λ

Results

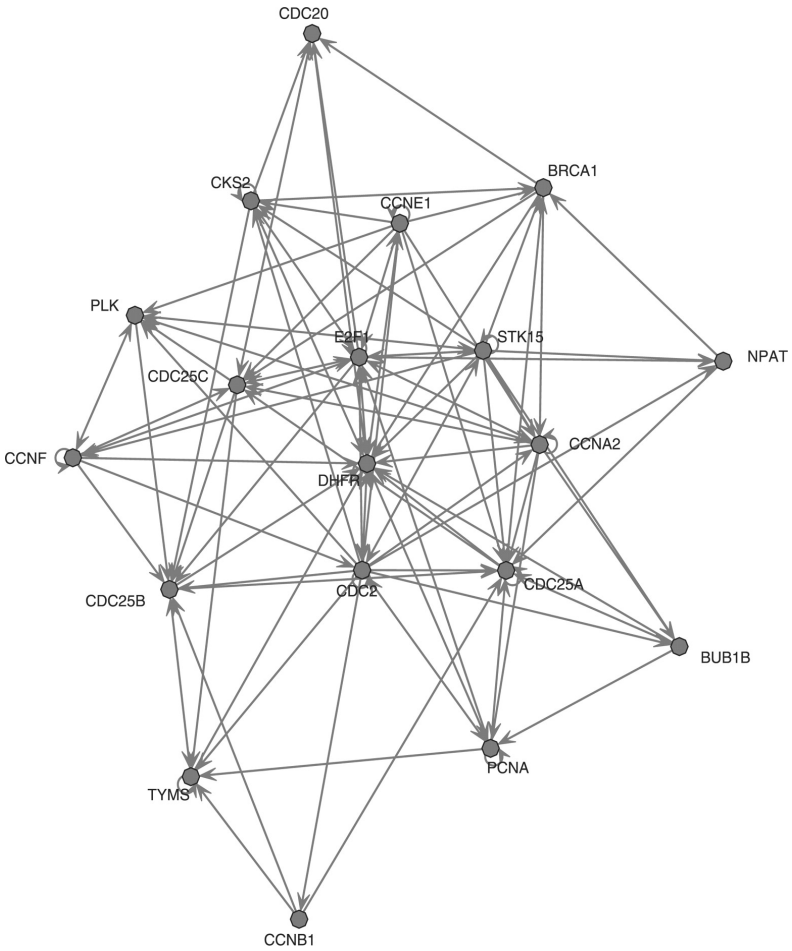


Figure : Causal structure from biol. experiments of Li et al. 2006.

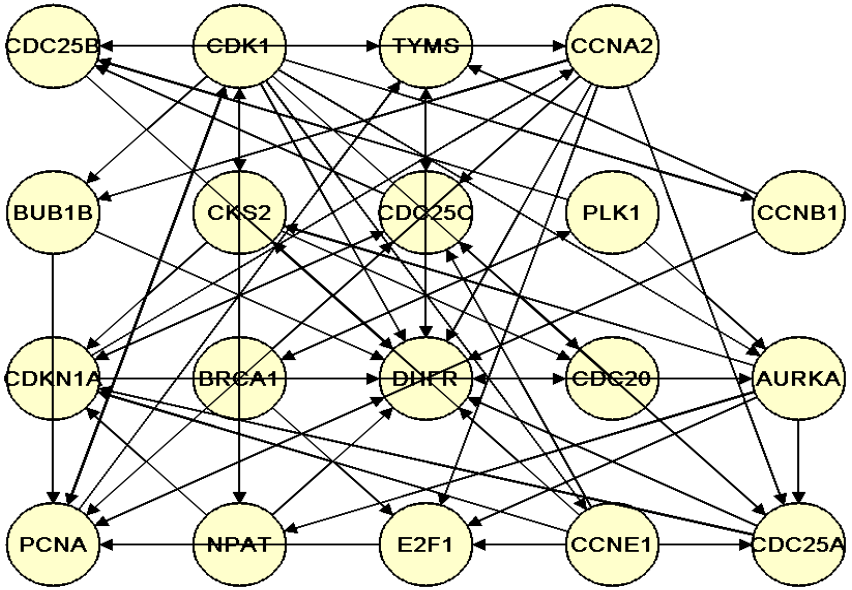


Figure : Causal structure achieved by Lasso Granger method with optim. threshold and regularization parameter

Conclusion and outlook

- The best method (w.r.t. precision and comp. costs) - Lasso Granger LG1 with opt. threshold and reg. parameter.
- It gave higher classiffication accuracy and good TP quality measures in comparison to CG, (regular) LG and GP4GRN method even for its automatic realization.
- All LGi methods - a polynomial comput. complexity.
- Low precision of GP4GRN (overfitting) and high comp. costs - GP4GRN (DBN-ODN) unfeasible for large gene regulatory networks.
- Future development: applications to larger causality networks; the role of the time lag parameter L for reconstruction of gene regulatory networks.
- The discussed algorithms can be also applied to other real-world problems dealing with interactions in a multi-agent system.