Graphical continuous Lyapunov models

Incontro di Statistica Matematica, Sestri Levante

Gherardo Varando, joint work with Niels Richard Hansen
Department of Mathematical Sciences

27/01/2020

Ornstein–Uhlenbeck process

Consider the SDE,

$$dX_t = B(X_t - a)dt + DdW_t,$$

where $B, D \in \mathbb{R}^{p \times p}$, $X_t, a \in \mathbb{R}^p$ and W_t is a standard Brownian motion in $\mathbb{R}^p$.

- We assume that the matrix B is **stable**, that is all its eigenvalues have negative real part
- Then X_t admit an invariant distribution $N(a, \Sigma)$ where the invariant covariance satisfies the **continuous time Lyapunov equation**:

$$B\Sigma + \Sigma B^t + C = 0$$

where $C = DD^t$

Graphical representation

For a mixed graph $\mathcal{G} = ([p], E = E_1 \cup E_2)$ We say that the matrices B and C are $\mathcal{G}$ compatible if:

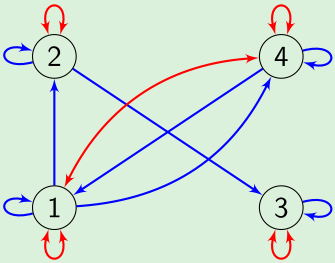
$$B_{i,j} \neq 0 \Rightarrow (j,i) \in E_1 \quad (j \to i)$$

$$C_{i,j} \neq 0 \Rightarrow (i,j),(j,i) \in E_2 \quad (i \leftrightarrow j)$$

Example

$$C = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 4 \end{pmatrix}$$

$$B = \begin{pmatrix} -2 & 0 & 0 & 1 \\ 1 & -5 & 0 & 0 \\ 0 & -1 & -1 & 0 \\ 2 & 0 & 1 & -3 \end{pmatrix}$$



Graphical continuous Lyapunov models

For a mixed graph $\mathcal{G}$ we define

$$\mathcal{M}_{\mathcal{G}} = \{\Sigma \text{ s.t. } B\Sigma + \Sigma B^t + C = 0 \text{ with } (B, C)\, \mathcal{G}\text{-compatible}\}$$

- A *trek* from i to j, denoted $i \rightsquigarrow j$, is a walk of the form

$$\tau : \underbrace{i \leftarrow \cdots \leftarrow i_1}_{n(\tau)} \leftarrow k \leftrightarrow l \rightarrow \underbrace{j_1 \rightarrow \cdots \rightarrow j}_{m(\tau)}.$$

- If $\Sigma \in \mathcal{M}_{\mathcal{G}}$ and there is no trek from i to j in $\mathcal{G}$ then $\Sigma_{ij} = 0$.

Ambition

We want to investigate if it is possible to estimate B (and/or C) with $(B, C) \in$ from observations of the invariant distribution.

In general the model is not identifiable, even if we fix the matrix C (e.g. a known diagonal matrix). That means that there are B_1, B_2 stable with the same pattern of zeros, such that

$$\Sigma(B_1, C) = \Sigma(B_2, C)$$

- We can hope that imposing enough sparsity on the matrix B we can estimate it from observations of the invariant distribution (from the empirical covariance matrix $\hat{\Sigma}$)

DAG models with fixed C

If the directed part of $\mathcal{G}$ is a DAG, then B can be reordered such that it is lower-triangular.

Identifiability of DAG

For every covariance matrix Σ there exist one and only one lower triangular matrix B, such that

$$B\Sigma + \Sigma B^t + C = 0$$

We are interested in the set,

$$\mathbf{B}(\Sigma) = \{B \in \mathbb{R}^{p \times p} | B\Sigma + \Sigma B^t + C = 0\}$$

- All matrices in $B(\Sigma)$ are stable
- Lyapunov equation $\Rightarrow B\Sigma + (B\Sigma)^t = -C$,

$$B\Sigma = W - \frac{1}{2}C \quad W \in \mathrm{Skew}_p$$

$$\mathbf{B}(\Sigma) = \left\{ \left(W - \frac{1}{2} C \right) \Sigma^{-1} | W \in \mathsf{Skew}_p \right\}$$

Imposing B to be a triangular matrix :

$$\mathbf{B}(\Sigma) \cap \{B | B_{i,j} = 0 \text{ if } j > i\}$$

To prove identifiability we need to show that the above intersection has just one element. We can write the corresponding linear equations for $W_{i,j}$, for $j > i$:

$$-\sum_{k<i} W_{k,i} P_{k,j} + \sum_{k>i} W_{i,k} P_{k,j} = \frac{1}{2} C_{i,i} P_{i,j}$$

where $P = \Sigma^{-1}$

$$-\sum_{k<i} W_{k,i} P_{k,j} + \sum_{k>i} W_{i,k} P_{k,j} = \frac{1}{2} C_{i,i} P_{i,j}$$

The $\frac{p(p-1)}{2}$ equations over the $\frac{p(p-1)}{2}$ elements of W form a block triangular system of linear equations.

$$\left(\begin{array}{ccc|cc|c} P_{2,2} & P_{3,2} & P_{4,2} & 0 & 0 & 0 \\ P_{2,3} & P_{3,3} & P_{4,3} & 0 & 0 & 0 \\ P_{2,4} & P_{3,4} & P_{4,4} & 0 & 0 & 0 \\ \hline -P_{1,3} & 0 & 0 & P_{3,3} & P_{4,3} & 0 \\ -P_{1,4} & 0 & 0 & P_{3,4} & P_{4,4} & 0 \\ \hline 0 & -P_{1,4} & 0 & -P_{2,4} & 0 & P_{4,4} \end{array}\right) \begin{pmatrix} W_{1,2} \\ W_{1,3} \\ W_{1,4} \\ W_{2,3} \\ W_{2,4} \\ W_{3,4} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} C_{1,1} P_{1,2} \\ C_{1,1} P_{1,3} \\ C_{1,1} P_{1,4} \\ C_{2,2} P_{2,3} \\ C_{2,2} P_{2,4} \\ C_{3,3} P_{3,4} \end{pmatrix}$$

- Since P is positive definite the leading minors are definite positive and the system has just one solution $W(P)$
- $\mathbf{B}(\Sigma) \cap \{B | B_{i,j} = 0 \text{ if } j > i\} = \{\left(W(P) - \frac{1}{2}C\right) P\}$
- If we fix an order of the variables, we can find the MLE of B from the empirical covariance matrix $\hat{\Sigma}$
- To find $W(P)$ we need to solve p systems of linear equations, where the coefficient matrices are the leading sub-matrix of P (size $p - i$)
- We can compute the inverses of the leading sub-matrices of P using the Woodbury matrix identity (or the formula for the inverse of a block matrix)

$$P^{-1}_{-1,-1} = \Sigma_{-1,-1} - \frac{1}{\Sigma_{1,1}} \Sigma_{-1,1} \Sigma_{1,-1}$$

- We can also use the parametrization

$$\mathbf{B}(\Sigma) = \left\{ (W - \frac{1}{2}C)\Sigma^{-1} | W \in \mathsf{Skew}_p \right\}$$

  to study identifiability in the more general case (also with non fixed C)
- A necessary condition for identifiability is that the number of parameters of the model (non-zero entries in B and in C) is less than $\frac{p(p+1)}{2}$

The Jacobian of the Lyapunov equation

We consider now the matrix valued function that maps every stable matrix B and symmetric C to the solution of the associated Lyapunov equation

$$\Sigma(B, C) : \text{Stab} \times \mathbb{S} \to \mathbb{S}$$

$$\Sigma(B, C) = \text{Solution of the Lyapunov eq.} \quad B\Sigma + \Sigma B^t + C = 0$$

We can now differentiate the Lyapunov equation with respect to $B_{i,j}$ and we obtain:

$$B\frac{\partial \Sigma(B,C)}{\partial B_{i,j}} + \frac{\partial \Sigma(B,C)}{\partial B_{i,j}} B^t + E_{(i,j)}\Sigma(B,C) + \Sigma(B,C)E^t_{(i,j)} = 0,$$

where $E_{(i,j)}$ is the (i,j) element of the canonical base of $\mathbb{R}^{p \times p}$.

The elements of the Jacobian $\frac{\partial \Sigma(B.C)}{\partial B_{i,j}}$ are the solution of a Lyapunov equation:

$$\frac{\partial \Sigma(B,C)}{\partial B_{i,j}} = \Sigma(B, Q_{(i,j)}), \quad Q_{(i,j)} = E_{(i,j)}\Sigma(B,C) + \Sigma(B,C)E_{(j,i)},$$

- Similarly for $\frac{\partial \Sigma(B,C)}{\partial C_{i,j}}$

Gradient of log-likelihood

Thanks to the Jacobian of $\Sigma(B, C)$ we can obtain the gradient of functions of $\Sigma(B, C)$, for example if we consider the Gaussian log-likelihood,

$$\ell(B, C) = -\log\det(\Sigma(B, C)) - \operatorname{trace}(\Sigma(B, C)^{-1}\hat{\Sigma})$$

We have that the gradient of ℓ is given by,

$$\begin{aligned}\frac{\partial \ell\,(\Sigma(B, C))}{\partial B_{ij}} &= \operatorname{tr}\left(\frac{\partial \Sigma(B, C)}{\partial B_{ij}}\frac{\partial \ell(\Sigma(B, C))}{\partial \Sigma}\right) \\ &= \operatorname{tr}\left(\Sigma(B, Q_{(i,j)})\Sigma^{-1}(\Sigma - \hat{\Sigma})\Sigma^{-1}\right)\end{aligned}$$

Efficient gradient computation

Computing the gradient of a loss function with

$$\frac{\partial L(\Sigma(B, C))}{\partial B_{ij}} = \operatorname{tr}\left(\Sigma(B, Q_{(i,j)}) \nabla L\right)$$

require computing the solution of p^2 CLEs.

Instead we show that it is possible to obtain directly the gradient solving only one additional CLE.

We note that, for each fixed stable B, $\Sigma(B, \cdot)$ is a linear operator on symmetric matrices with adjoint operator given by $\Sigma(B^t, \cdot)$, that is,

$$\operatorname{tr}\left(\Sigma(B, C) D\right) = \operatorname{tr}\left(C \Sigma(B^t, D)\right)$$

$$\frac{\partial L(\Sigma(B,C))}{\partial B_{ij}} = \operatorname{tr}\left(Q_{(i,j)}\Sigma\left(B^t,\nabla L\right)\right) = \left(2\Sigma(B,C)\Sigma(B^t,\nabla L)\right)_{ij},$$

$$\frac{\partial L(\Sigma(B,C))}{\partial C_{ij}} = \operatorname{tr}\left(E_{(i,j)} + E_{(j,i)}\Sigma\left(B^t,\nabla L\right)\right) = \left(2\Sigma(B^t,\nabla L)\right)_{ij}.$$

Penalized log-likelihood

$$\begin{array}{ll} \text{minimize} & -\ell(B, C) + \lambda\rho_1(B) \\ \text{subject to} & \text{B stable} \end{array}$$

where $\rho_1(B) = \sum_{i \neq j} |B_{ij}|$.

- We can use a proximal gradient method as in the classical LASSO problem
- At each iteration we move in the direction of the negative gradient of ℓ and we then apply the soft-thresholding.

Require: $\hat{\Sigma}, C \in \mathrm{PD}(p)$, $B_0 \in \mathrm{Stab}(p)$, $M \in \mathbb{N}$, $\varepsilon, \lambda > 0$, $\alpha \in (0,1)$

$B = B_0$

$s = 1$

$I = \{(i,j) \in [p] \times [p] : B_{0,ij} \neq 0\}$

$\Sigma = \Sigma(B, C)$

repeat

  $f = L(\Sigma) + \lambda \rho_1(B)$

  $D = 2\Sigma(B, C)\Sigma(B^t, \nabla L)$

  $D_{ij} = 0$ for $(i,j) \notin I$

  loop

    $B' = \mathcal{S}_{s\lambda}(B - sD)$

    $\Sigma' = \Sigma(B', C)$

    $f' = L(\Sigma') + \lambda \rho_1(B')$

    if $B' \in \mathrm{Stab}(p)$ **and**
    $f' \leq f + \frac{1}{2s}||B - B'||_F^2 - \mathrm{tr}((B - B')D)$ **then**

      break

    else

      $s = \alpha s$

    end if

  end loop

  $\delta = \frac{f - f'}{|f|}$

  $\Sigma = \Sigma', B = B', f = f'$

  $I = \{(i,j) \in [p] \times [p] : B_{ij} \neq 0\}$

until $k > M$ **or** $\delta < \varepsilon$

Ensure: B, Σ such that $\Sigma = \Sigma(B, C)$

Implementation

- We implemented the algorithm in FORTRAN
- To solve the Lyapunov equation we use a version of the Bartels and Stewart algorithm partially implemented in LAPACK - $\mathbf{O}(\mathbf{p}^3)$
- Each iteration $\mathbf{O}(\mathbf{p}^3)$
- We actually compute all the regularization path (as in LASSO and GLASSO) using warm start of the proximal gradient.
- Code available `https://github.com/gherardovarando/clggm`

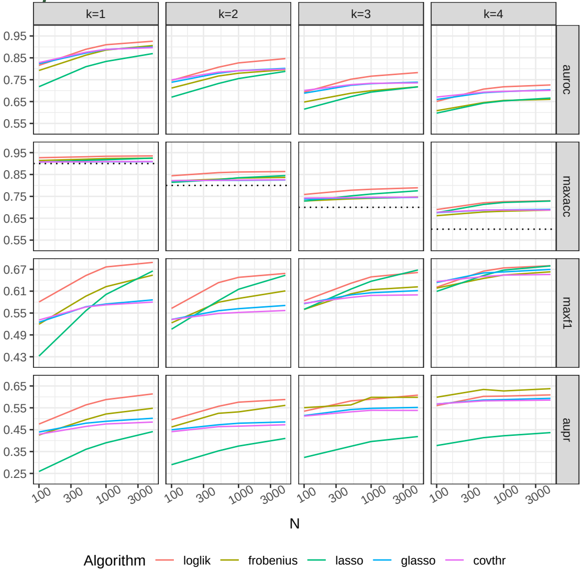
The direct Lasso Path

Fitch (2019) suggests estimating B as a sparse, approximate solution to the Lyapunov equation for Σ fixed and equal to the empirical covariance matrix, $\hat{\Sigma}$. For fixed λ the estimate is the solution to the lasso problem

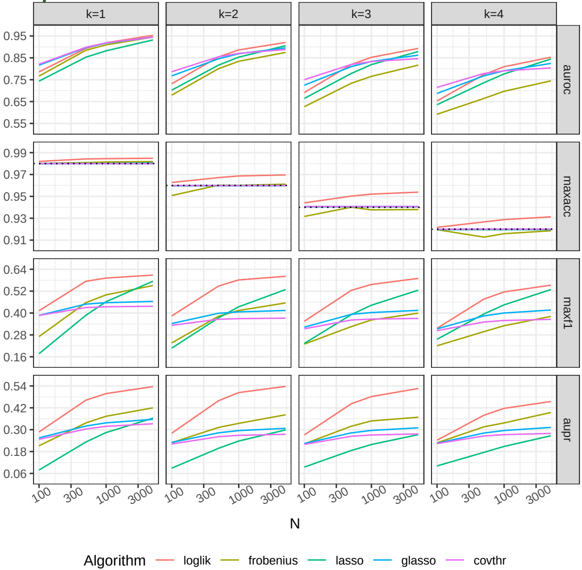
$$\text{minimize} \quad \|B\hat{\Sigma} + \hat{\Sigma}B^t - C\|_F^2 + \lambda\rho_1(B). \qquad (1)$$

for a fixed C.

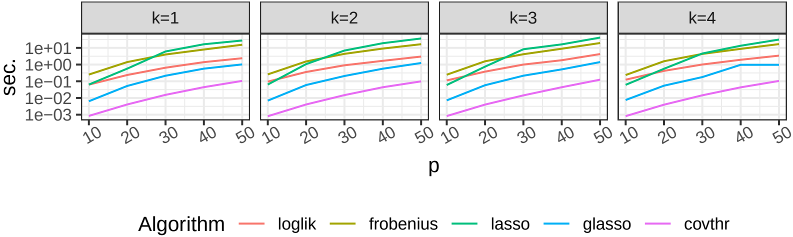
Results $p = 10$


k=1
k=2
k=3
k=4
auroc
maxacc
maxf1
aupr
0.95
0.85
0.75
0.65
0.55
0.67
0.61
0.55
0.49
0.43
0.65
0.55
0.45
0.35
0.25
100
300
1000
3000
N
Algorithm
loglik
frobenius
lasso
glasso
covthr

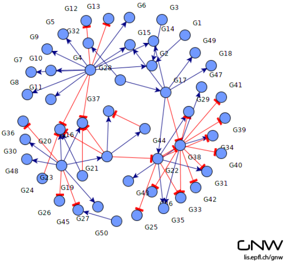
Results $p = 50$


k=1
k=2
k=3
k=4
auroc
maxacc
maxf1
aupr
0.95
0.85
0.75
0.65
0.55
0.99
0.97
0.95
0.93
0.91
0.64
0.52
0.40
0.28
0.16
0.54
0.42
0.30
0.18
0.06
100
300
1000
3000
N
Algorithm
loglik
frobenius
lasso
glasso
covthr

Results, execution time



Applications



- Gene regulatory networks
- Protein interaction networks
- Chemical reaction networks

Approaches usually restrict to graphs with no cycles (DAG, Bayesian networks) or consider correlation-based approaches where directions of the edges are not recovered.

Protein-Signaling Network

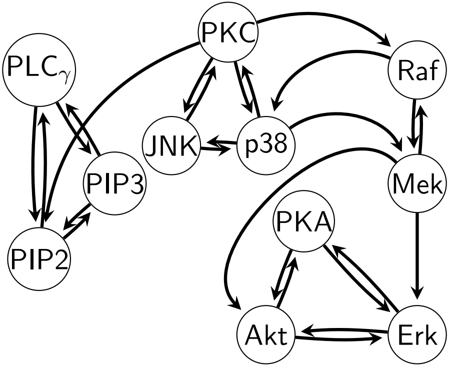


Figure: Estimated graph from data in Sachs et al. (2005). Self loops and bidirected edges are not plotted.

Future work

- Nesterov accelerated methods
- Characterizzation of identifiable models
- Application to time-series dataset

Graphical continuous Lyapunov models

Incontro di Statistica Matematica, Sestri Levante

Gherardo Varando, joint work with Niels Richard Hansen
Department of Mathematical Sciences

Katherine Fitch. Learning directed graphical models from Gaussian data. *arXiv:1906.08050*, 2019.

Karen Sachs, Omar Perez, Dana Pe'er, Douglas A. Lauffenburger, and Garry P. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.