Maximum-Likelihood Estimation of Autoregressive Models with Conditional Independence Constraints

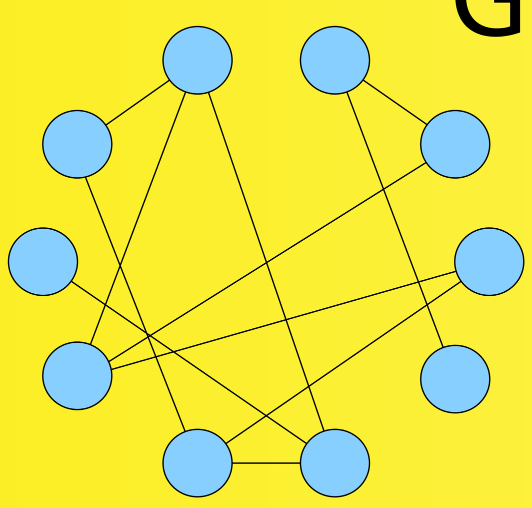
Jitkomut Songsiri
University of California, Los Angeles

Joachim Dahl
Aalborg University

Lieven Vandenberghe
University of California, Los Angeles

Introduction

Graphical Models



- Represent relations between random variables
- Applications in
  - economics (exchange rates, stock prices, etc.)
  - brain networks (functional connectivity between brain regions)
  - ...

- In conditional independence graph, nodes correspond to random variables X_i
- Link (i,j) is absent if X_i and X_j are **conditionally independent**

Characterization for Gaussian time series

$$X(t) = (X_1(t), X_2(t), \dots, X_n(t)) \ , \ t \in \mathbb{Z}$$

X_i and X_j are **conditionally independent** if $(S(\omega)^{-1})_{ij} = 0 \ , \ \forall \omega$

$S(\omega)$ is the spectral density of $X(t)$ (Brillinger (1996))

Goal

investigate a parametric spectrum estimation method that incorporates sparsity constraints in $S(\omega)^{-1}$

Multivariate AR Process

$$B_0 x(t) = -\sum_{k=1}^{p} B_k x(t-k) + v(t)$$

where $x(t) \in \mathbf{R}^n$ and $v(t) \sim N(0, I)$ is Gaussian white noise.

Conditional Independence in AR processes

$$S(\omega)^{-1} = Y_0 + \sum_{k=1}^{p} (e^{-jk\omega} Y_k + e^{jk\omega} Y_k^T)$$

$$Y_k = \sum_{i=0}^{p-k} B_i^T B_{i+k} \ , \ k = 0, 1, \dots, p$$

$$(S(\omega)^{-1})_{ij} = 0 \longleftrightarrow [Y_k]_{ij} = [Y_k]_{ji} = 0 \ , \ k = 0, \dots, p \longleftrightarrow \mathrm{P}(\mathrm{D}(B^T B)) = 0$$

where $B = [B_0 \quad B_1 \quad \dots \quad B_p]$, P : projection on the sparsity pattern

D returns sums along the block diagonals

$$\mathrm{D}_k(X) = \sum_{i=0}^{p-1} X_{i,i+k}, \quad k = 0, 1, \dots, p$$

Problem Formulation

$$\begin{array}{ll} \text{minimize} & -2 \log\det B_0 + \mathbf{tr}(C B^T B) \\ \text{subject to} & \mathrm{P}(\mathrm{D}(B^T B)) = 0. \end{array} \quad \text{(P1)}$$

variable: $B = [\ B_0 \quad B_1 \quad \cdots \quad B_p\ ]$

$\updownarrow$ $X = B^T B$

$$\begin{array}{ll} \text{minimize} & -\log\det X_{00} + \mathbf{tr}(CX) \\ \text{subject to} & \mathrm{P}(\mathrm{D}(X)) = 0 \\ & X \succeq 0 \end{array} \quad \text{(P2)}$$

variable: $X \in \mathbf{S}^{n(p+1)}$ with blocks X_{ij}

$\updownarrow$ Dual

$$\begin{array}{ll} \text{maximize} & \log\det W + n \\ \text{subject to} & \begin{bmatrix} W & 0 \\ 0 & 0 \end{bmatrix} \preceq C + \mathrm{T}(\mathrm{P}(Z)). \end{array} \quad \text{(D2)}$$

variables: $W \in \mathbf{S}^n$ and $Z = [\ Z_0 \ Z_1 \ \cdots \ Z_p\ ]$

Exactness of the Relaxation

If C is block-Toeplitz, the low-rank property of $X^\star$ follows from

$$C + \mathrm{T}(\mathrm{P}(Z^\star)) \succeq \begin{bmatrix} W^\star & 0 \\ 0 & 0 \end{bmatrix} \implies C + \mathrm{T}(\mathrm{P}(Z^\star)) \succ 0$$

and the complementary slackness condition

$$X^\star \left(C + \mathrm{T}(\mathrm{P}(Z^\star)) - \begin{bmatrix} W^\star & 0 \\ 0 & 0 \end{bmatrix} \right) = 0$$

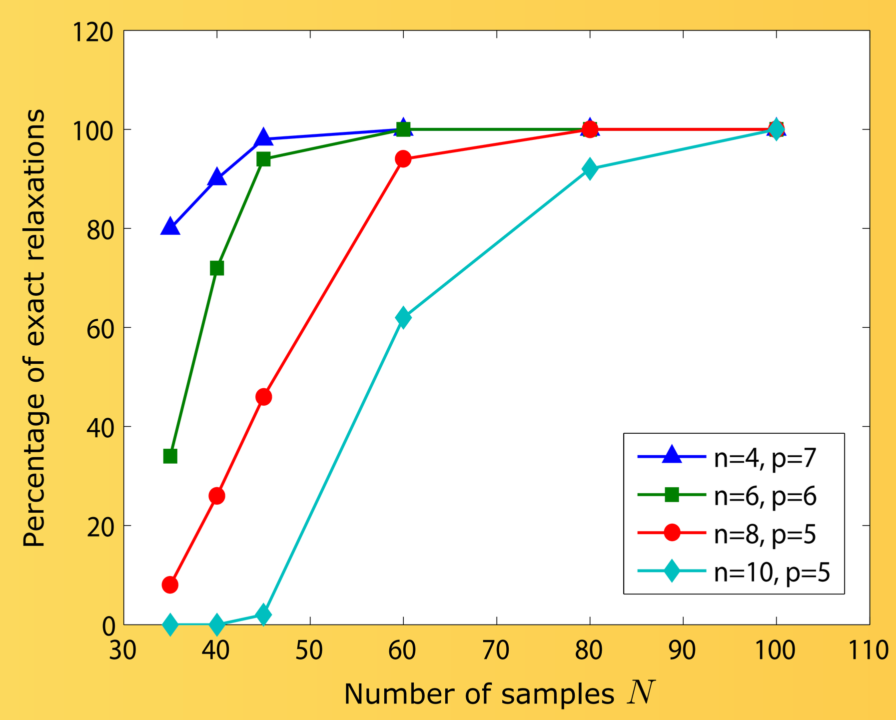
P1 Maximum-Likelihood Estimation

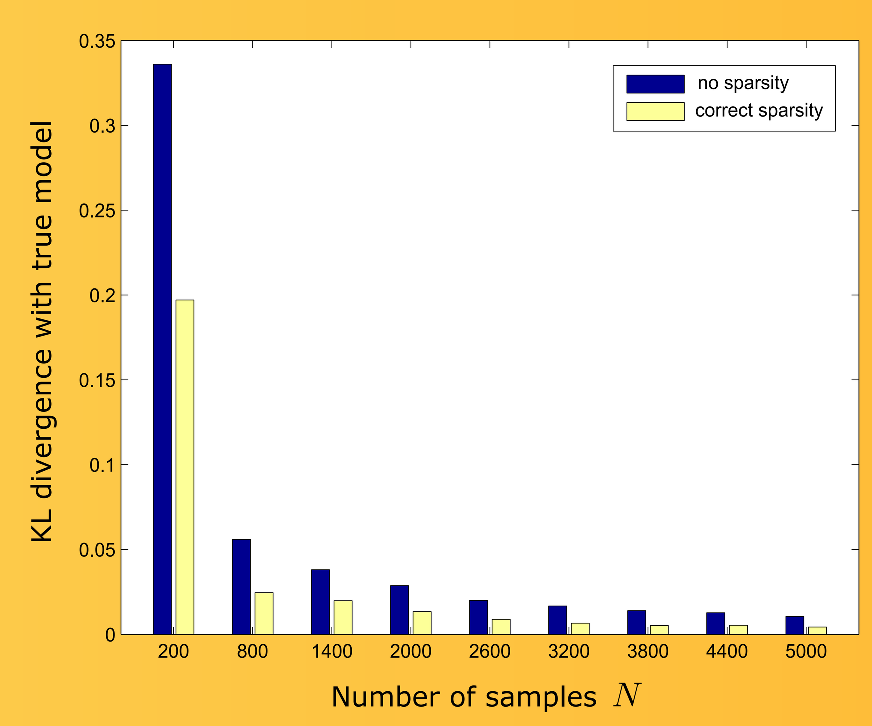
- Includes conditional independent constraints
- C is a sample covariance matrix computed by using the non-windowed estimate
- Nonconvex because of quadratic equality constraints

P2 Convex Relaxation

- If $X^\star$ has rank n, then by factorizing $X^\star = B^T B$, B must be optimal in (P1)
- The relaxation is exact if $X^\star$ has rank n
- The low-rank property of $X^\star$ can be proved for block-Toeplitz and positive definite C
- For ML problem, C is close to a block-Toeplitz matrix when the sample size $N \to \infty$

Example



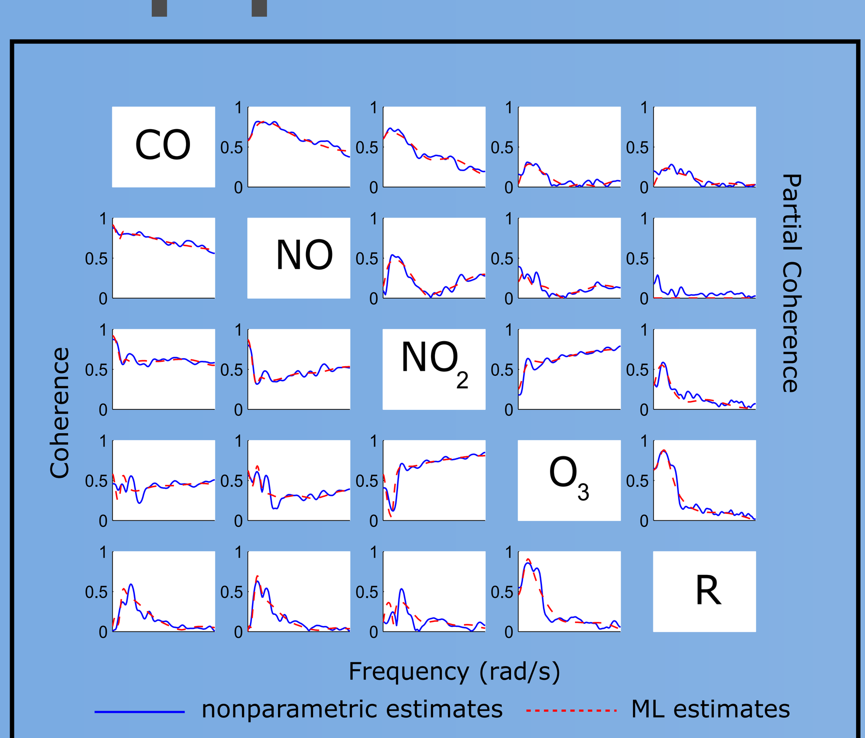


- Solve (P2) with different sample covariance matrices C (not block-Toeplitz)
- The relaxation is often exact for moderate values of N, even when C is not block-Toeplitz

ML estimate without sparsity constraints gives a model with substantially larger values of KL when N is small

Application

Model Selection



- Relations among air pollutants, CO, NO, NO_2 , O_3 , and solar intensity (R)
- Enumerate different models (p and topology)
- Calculate BIC scores (Bayes information criterion)

$$\mathrm{BIC} = -2L + k \log N$$

k : number of effective parameters
L : maximized log-likelihood
N : sample size

- Select the model with the lowest BIC score

Conclusions

Graphical Models of Gaussian AR processes

- Maximum-likelihood estimation leads to a nonconvex problem
- The convex relaxation solves the ML problem if C is block-Toeplitz
- In practice, the relaxation is often exact even if C is not block-Toeplitz
- The method is useful for model selection problems in combination with AIC, BIC scores