Scalable Inference of Sparsely-changing Gaussian Markov Random Fields

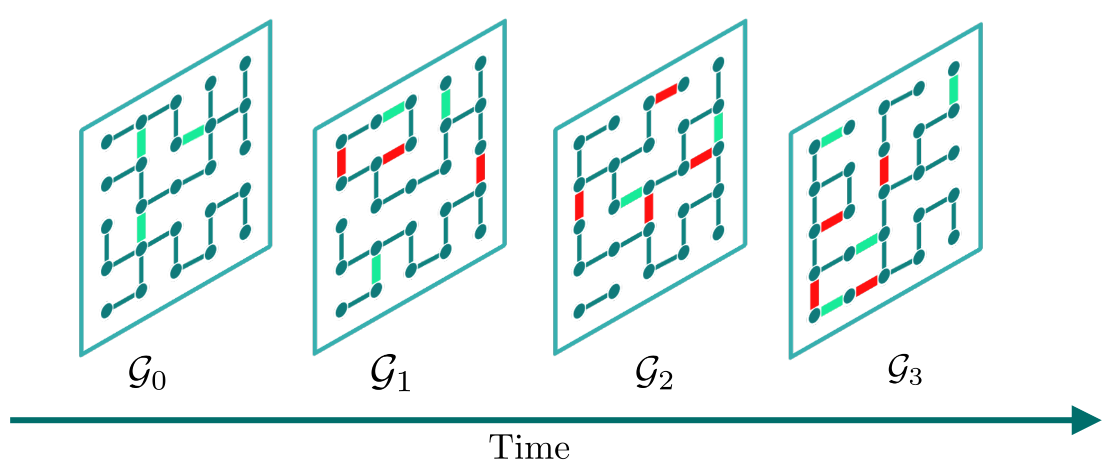
Salar Fattahi, University of Michigan, fattahi@umich.edu
Andres Gomez, University of Southern California, gomezand@usc.edu

NeurIPS 2021

Motivation

Motivation: Modern networked systems are ***massive-scale***, with ***time-varying*** and ***unknown*** topologies. The behavior of these systems can be captured via time-varying graphical models.

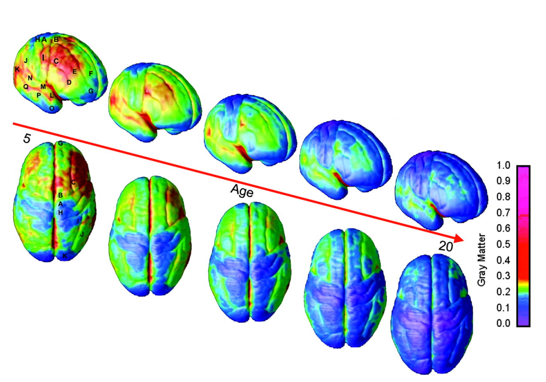
Goal: Estimate the time-varying graphical model based on limited number of observations.



➢ The underlying graph-based structure can be described via **Time-varying Markov Random Fields (MRFs).**

Applications

Application 1: Brain Networks



Available Data: fMRI measurements.
Hidden Structure: Functional connectivity network.
Application: Brain pathology discovery (Schizophrenia).
Size: 200K nodes (voxels), 20B links.

Brain connectivity network change with age and maturity.

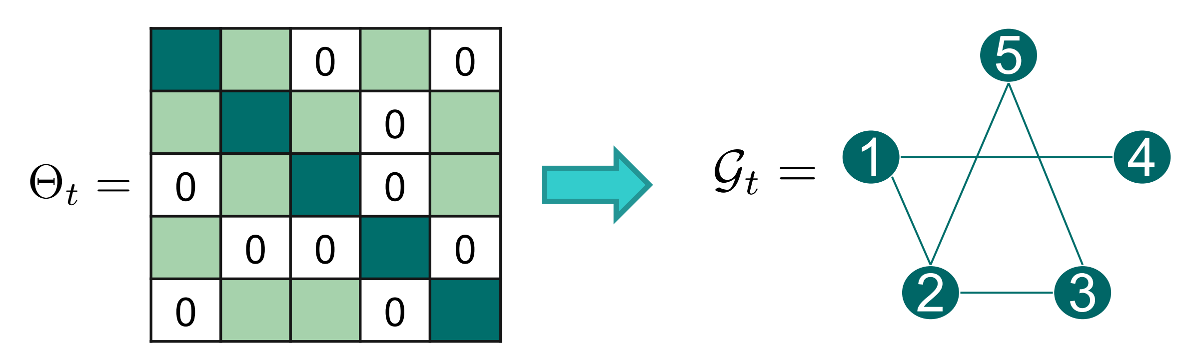
Application 2: Stock Correlation Network

Available Data: Stock prices
Hidden Structure: Stock correlation network
Application: Anomaly detection, portfolio optimization

Stock correlation network changes in response to global events.

Problem Statement

- **Sparsely-changing Gaussian MRF:** Data is generated from a sparsely-changing Gaussian distribution.
- **Goal:** Estimate sparsely-changing inverse covariance (precision) matrices from the observed data.



Common Approach: Maximum likelihood estimation

$$\{\widehat{\Theta}_t\}_{t=0}^T = \arg\min_{\Theta_t} \left(\sum_{t=0}^T \langle \Theta_t, \widehat{\Sigma}_t\rangle - \log\det(\Theta_t)\right) + \beta_1 \sum_{t=0}^T \|\Theta_t\|_0 + \beta_2 \sum_{t=1}^T \|\Theta_t - \Theta_{t-1}\|_0$$

$$\text{subject to:} \quad \Theta_t \succ 0 \quad \text{for all } t = 0, 1, \ldots, T$$

Nonconvex and intractable **Good statistical guarantees**

Convex relaxation

$$\{\widehat{\Theta}_t\}_{t=0}^T = \arg\min_{\Theta_t} \left(\sum_{t=0}^T \langle \Theta_t, \widehat{\Sigma}_t\rangle - \log\det(\Theta_t)\right) + \beta_1 \sum_{t=0}^T \|\Theta_t\|_1 + \beta_2 \sum_{t=1}^T \|\Theta_t - \Theta_{t-1}\|_1$$

$$\text{subject to:} \quad \Theta_t \succ 0 \quad \text{for all } t = 0, 1, \ldots, T$$

Convex and tractable **Inferior statistical guarantees**

Proposed Method

Given the approximate backward mapping $\widetilde{F}^*(\mu_t)$, solve:

$$\left\{\widehat{\theta}_t\right\}_{t=0}^T \in \arg\min_{\{\theta_t\}_{t=0}^T} \ (1-\beta)\sum_{t=0}^T \|\theta_t\|_0 + \beta \sum_{t=1}^T \|\theta_t - \theta_{t-1}\|_0$$

$$\text{subject to} \quad \left\|\theta_t - \widetilde{F}^*(\widehat{\mu}_t)\right\|_\infty \le \lambda_t \qquad \text{for} \quad t = 0, 1, \ldots, T$$

Absolute regularization — Temporal regularization — Distance from backward mapping

$$\widetilde{F}^*(\widehat{\Sigma}_t) = [\mathtt{ST}_\nu(\widehat{\Sigma}_t)]^{-1}, \qquad [\mathtt{ST}_\nu(M)]_{ij} = \begin{cases} M_{ij} - \mathrm{sign}(M_{ij})\nu & \text{if } i \neq j \\ M_{ij} & \text{if } i = j \end{cases}$$

$\widehat{\Sigma}_t =$

1	X	X	X	X	X
X	1	X	X	X	X
X	X	1	X	X	X
X	X	X	1	X	X
X	X	X	X	1	X
X	X	X	X	X	1

➡ $\mathtt{ST}(\widehat{\Sigma}_t) =$

1	0	0	X	0	X
0	1	X	0	X	0
0	X	1	X	0	0
X	0	X	1	0	0
0	X	0	0	1	0
X	0	0	0	0	1

➡ $[\mathtt{ST}(\widehat{\Sigma}_t)]^{-1} =$

1	X	X	X	X	X
X	1	X	X	X	X
X	X	1	X	X	X
X	X	X	1	X	X
X	X	X	X	1	X
X	X	X	X	X	1

Theoretical Results

Statistical guarantee

Suppose that

$$T \gtrsim (\log T \log d)^{1.5}, \quad \lambda_t \asymp \frac{\sqrt{\log T \log d}}{T^{1/3}}, \qquad \nu_t \asymp \frac{\sqrt{\log T \log d}}{T^{1/3}},$$

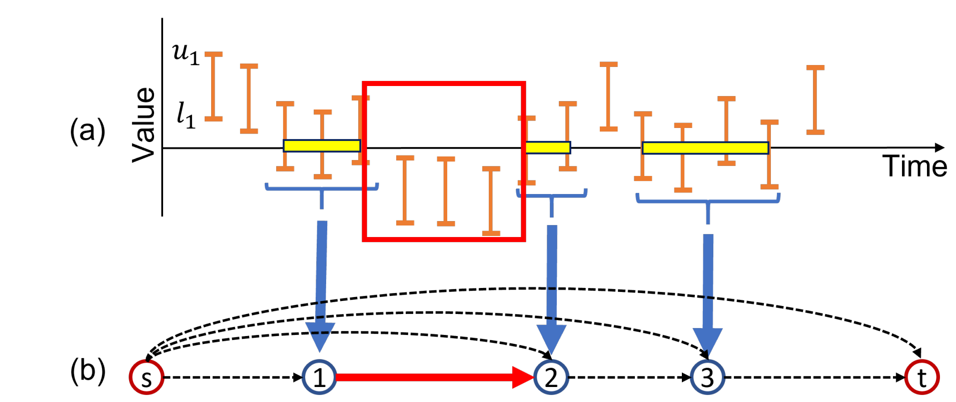
Then, with as few **as one sample per time**, we have *sparsistency* and

$$\|\widehat{\Theta}_t - \Theta_t^*\|_\infty \lesssim \frac{\sqrt{\log T \log d}}{T^{1/3}}, \qquad \|\widehat{\Theta}_t - \Theta_t^*\|_2 \lesssim \frac{k\sqrt{\log T \log d}}{T^{1/3}}$$

Computational guarantee

The proposed optimization problem can be solved in **near-linear time and memory.**

Key idea: Shortest path problem over DAG.

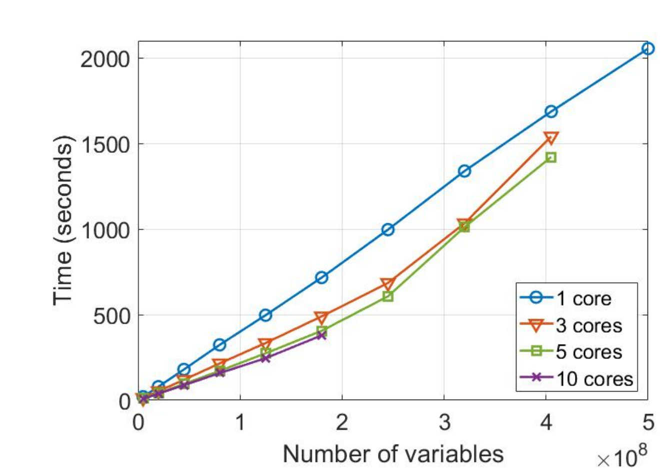


Simulations

Experiment 1: Massive-scale Datasets

*Instance with **500M** variables solved in less than an hour.*

Using 5 cores can reduce the runtime by 40%.



Experiment 2: Stock Correlation Network

Daily stock prices for 214 securities from 1990-2017

Goal: Infer the stock correlation network given the prices.