GRAPH LEARNING FROM GAUSSIAN AND STATIONARY GRAPH SIGNALS

Andrei Buciulea and Antonio G. Marques*

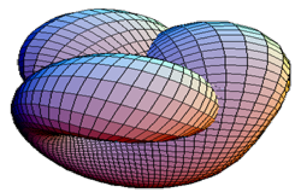
Dept. of Signal Theory and Communications, King Juan Carlos University, Madrid, Spain

Introduction

- Graph SP: models the irregular structure of a dataset using a graph
  - ⇒ The data structure carries critical information about the nature of the data
  - ⇒ Leverages the graph topology to process the data

Interpolate a brain signal from local observations

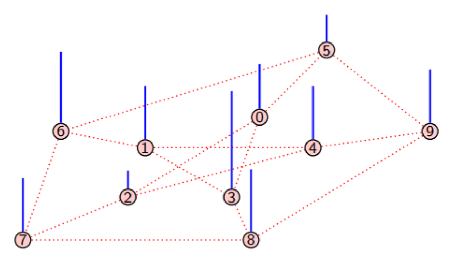
Compress a signal in an irregular domain



Localize the source of a rumor

Preliminaries of Graph Signal Processing

- A **graph** $\mathcal{G}$: N nodes and links connecting them
  - ⇒ $\mathcal{G} \equiv (\mathcal{V}, \mathcal{E}, \mathbf{A})$, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, $\mathbf{A} \in \mathbb{R}^{N \times N}$
- Define a **signal** $\mathbf{x} \in \mathbb{R}^N$ on top of the graph
  - ⇒ x_i = Signal value at node i

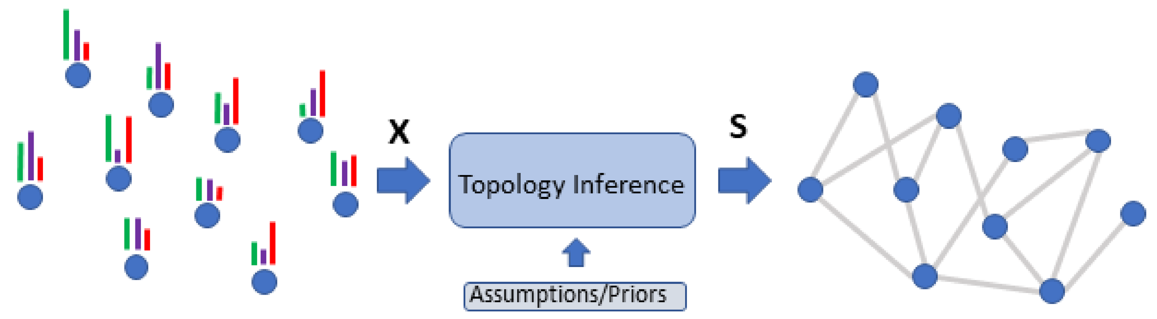


- Associated with $\mathcal{G} \rightarrow$ **graph-shift operator** (GSO) $\mathbf{S} \in \mathbb{R}^{N \times N}$
  - ⇒ $S_{ij} \neq 0$ if and only if $i = j$ or $(i, j) \in \mathcal{E}$
  - ⇒ Represents local structure in $\mathcal{G}$, in this work $\mathbf{S} = \mathbf{A}$
- **Graph filters** → Linear operators of the form $\mathbf{H} := \sum_{p=0}^{P-1} h_p \mathbf{S}^p$ [Sandryhaila13]
- Random signal $\mathbf{x}$ is **stationary** in GSO $\mathbf{S}$ if $\mathbf{x} = \mathbf{H}\mathbf{w}$, with $\mathbf{w}$ white
  - ⇒ $\mathbf{\Sigma} = \mathbb{E}\left[\mathbf{H}\mathbf{w}(\mathbf{H}\mathbf{w})^T\right] = \mathbf{H}\mathbb{E}\left[\mathbf{w}\mathbf{w}^T\right]\mathbf{H}^T = \mathbf{H}^2 \rightarrow \mathbf{\Sigma}$ is a polynomial on $\mathbf{S}$

Network Topology Inference

Network topology inference from nodal observations

"Given a collection $\mathbf{X} := [\mathbf{x}_1, ..., \mathbf{x}_R]$ of graph signal observations supported on the unknown graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{A})$ find an optimal $\mathbf{S}$"



- Ill posed problem: optimality, priors, regularizations
  - ⇒ Test Pearson correlations, partial correlations and conditional dependence
  - ⇒ Sparsity [Friedman07] and consistency [Meinshausen06]
  - ⇒ Graph Signal Processing (GSP) [Dong17,Mei17,Segarra17,Mateos19]
- This work uses graphical models and GSP to infer the network topology
  - ⇒ Assume each $\mathbf{x}_r$ is a i.i.d realization of $\mathcal{N}(\mathbf{0},\mathbf{\Sigma})$ and stationary in $\mathbf{S}$
- **General Goal:** Infer the GSO (precision matrix $\mathbf{\Theta}$ or $\mathbf{S}$) from $\hat{\mathbf{\Sigma}}$
  - ⇒ With $\hat{\mathbf{\Sigma}} = \frac{1}{R}\mathbf{X}\mathbf{X}^T$ being the sample covariance matrix and $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_R] \in \mathbb{R}^{N \times R}$
  - ⇒ Assuming: 1) Gaussian or 2) Stationary signals on the graph

Graph Topology Inference from Gaussian Signals (GL) [Friedman08]

- Let $\mathbf{X} \in \mathbb{R}^{N \times R}$ be R i.i.d samples of $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$

$$\hat{\mathbf{\Theta}} = \underset{\mathbf{\Theta} \succeq 0, \mathbf{\Theta} \in \mathcal{S}}{\text{argmin}} \; -\log\det(\mathbf{\Theta}) + \text{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Theta}) + \rho \text{h}(\mathbf{\Theta})$$

- General properties ⇒ Good performance in low-sample scenarios
  - ⇒ Specific covariance model $\mathbf{\Sigma}_{MRF} = (\sigma\mathbf{I} + \delta\mathbf{S})^{-1}$

Graph Topology Inference from Stationary Signals (GSR) [Segarra17]

- Let $\mathbf{X}$ be stationary w.r.t $\mathbf{S} \rightarrow \mathbf{x}_r = (\sum_{p=0}^{P-1} h_p \mathbf{S}^p)\mathbf{w}_r$

$$\hat{\mathbf{S}} = \underset{\mathbf{S}}{\text{argmin}} \; \|\mathbf{S}\|_0 \quad \text{s. to} \quad \hat{\mathbf{\Sigma}}\mathbf{S} = \mathbf{S}\hat{\mathbf{\Sigma}}, \quad \mathbf{S} \in \mathcal{S}$$

Problem Statement

- Given the sample covariance matrix $\hat{\mathbf{\Sigma}}$
- Find the sparsest solution $\mathbf{S}$ related to the graph structure considering:
  - ⇒ (AS1): $\{\mathbf{x}_r\}_{r=1}^R$ are i.i.d realizations of $\mathcal{N}(\mathbf{0}, \mathbf{I})$
    PDF in terms of precision matrix $\mathbf{\Theta}$ and log-likelihood function

$$f_{\mathbf{\Theta}}(\mathbf{x}) = (2\pi)^{-N/2} \cdot det^{\frac{1}{2}}(\mathbf{\Theta}) \cdot e^{-\frac{1}{2}\mathbf{x}^T\mathbf{\Theta}\mathbf{x}}, \;\; \mathcal{L}(\mathbf{X}|\mathbf{\Theta}) = \sum_{r=1}^{R} \log(f_{\mathbf{\Theta}}(\mathbf{x}_r))$$

  - ⇒ (AS2): $\{\mathbf{x}_r\}_{r=1}^R$ are stationary in $\mathbf{S}$
    $\mathbf{S}$ can be expressed as a polynomial on Θ ⇒ $\mathbf{S} = poly(\mathbf{\Theta})$
    $\mathbf{\Theta}$ can be expressed as a polynomial on $\mathbf{S}$ ⇒ $\mathbf{\Theta} = poly'(\mathbf{S})$

Problem I: Minimize $-\mathcal{L}(\mathbf{X}|\mathbf{\Theta})$ under (AS2)

$$\hat{\mathbf{\Theta}}, \hat{\mathbf{S}} = \underset{\mathbf{\Theta} \succeq 0, \mathbf{S} \in \mathcal{S}}{\text{argmin}} \; -\log(\det(\mathbf{\Theta})) + \text{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Theta}), \quad \text{s. to} \quad \|\mathbf{S}\|_0 \leq \kappa \;\text{ and }\; \mathbf{\Theta}\mathbf{S} = \mathbf{S}\mathbf{\Theta}.$$

- Sparsity constraint $\|\mathbf{S}\|_0 \leq \kappa \rightarrow$ Non-convex
- Commutativity constraint $\mathbf{\Theta}\mathbf{S} = \mathbf{S}\mathbf{\Theta} \rightarrow$ Bi-linear term

Graphical Models with Stationary Signals: Proposed Solution

- Reformulate Problem I adding:
  - ⇒ The ℓ_1 norm as a convex relaxation of the ℓ_0 norm
  - ⇒ A new optimization variable $\mathbf{\Theta}_2$

$$\hat{\mathbf{\Theta}}_1, \hat{\mathbf{\Theta}}_2, \hat{\mathbf{S}} = \underset{\mathbf{\Theta}_1, \mathbf{\Theta}_2 \succeq 0; \mathbf{S} \in \mathcal{S}}{\text{argmin}} \; \text{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Theta}_1) - \log\det(\mathbf{\Theta}_2) + \rho\|\mathbf{S}\|_1 \quad \text{s. to} \quad \mathbf{\Theta}_1\mathbf{S} = \mathbf{S}\mathbf{\Theta}_1 \;\text{ and }\; \mathbf{\Theta}_1 = \mathbf{\Theta}_2$$

- Problem II: Min. the augmented Lagrangian with Lagrange multipliers ⇒ $\mathbf{Y}$ and $\mathbf{Z}$

$$\hat{\mathbf{\Theta}}_1, \hat{\mathbf{\Theta}}_2, \hat{\mathbf{S}} = \underset{\mathbf{\Theta}_1, \mathbf{\Theta}_2 \succeq 0; \mathbf{S} \in \mathcal{S}}{\text{argmin}} \; \text{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Theta}_1) - \log\det(\mathbf{\Theta}_2) + \rho\|\mathbf{S}\|_1 + \langle \mathbf{Z}, \mathbf{\Theta}_1 - \mathbf{\Theta}_2 \rangle + \frac{\lambda}{2}\|\mathbf{\Theta}_1 - \mathbf{\Theta}_2\|_F^2 + \langle \mathbf{Y}, \mathbf{\Theta}_1\mathbf{S} - \mathbf{S}\mathbf{\Theta}_1 \rangle + \frac{\lambda}{2}\|\mathbf{\Theta}_1\mathbf{S} - \mathbf{S}\mathbf{\Theta}_1\|_F^2$$

Iterative Block Successive Upper-bound Minimization algorithm (GGSR) [Hong16]

- Subproblem I: given $\mathbf{\Theta}_2^{(t)}$, $\mathbf{S}^{(t)}$, $\mathbf{Y}^{(t)}$, and $\mathbf{Z}^{(t)}$ estimate $\mathbf{\Theta}_1^{(t+1)}$

$$\mathbf{\Theta}_1^{(t+1)} = \underset{\mathbf{\Theta}_1 \succeq 0}{\text{argmin}} \; \text{tr}(\mathbf{\Sigma}\mathbf{\Theta}_1) + \langle \mathbf{Z}^{(t)}, \mathbf{\Theta}_1 \rangle + \frac{\lambda}{2}\|\mathbf{\Theta}_1 - \mathbf{\Theta}_2^{(t)}\|_F^2 + \langle \mathbf{Y}^{(t)}, \mathbf{\Theta}_1\mathbf{S}^{(t)} - \mathbf{S}^{(t)}\mathbf{\Theta}_1 \rangle + \frac{\lambda}{2}\|\mathbf{\Theta}_1\mathbf{S}^{(t)} - \mathbf{S}^{(i)}\mathbf{\Theta}_1\|_F^2.$$

- Subproblem II: given $\mathbf{\Theta}_1^{(t+1)}$ and $\mathbf{Z}^{(t)}$ estimate $\mathbf{\Theta}_2^{(t+1)}$

$$\mathbf{\Theta}_2^{(t+1)} = \underset{\mathbf{\Theta}_2 \succeq 0}{\text{argmin}} \; -\log\det(\mathbf{\Theta}_2) + \langle \mathbf{Z}^{(t)}, \mathbf{\Theta}_2 \rangle + \frac{\lambda}{2}\|\mathbf{\Theta}_1^{(t+1)} - \mathbf{\Theta}_2\|_F^2.$$

- Subproblem III: given $\mathbf{\Theta}_2^{(t+1)}$ and $\mathbf{Y}^{(t)}$ estimate $\mathbf{S}^{(t+1)}$

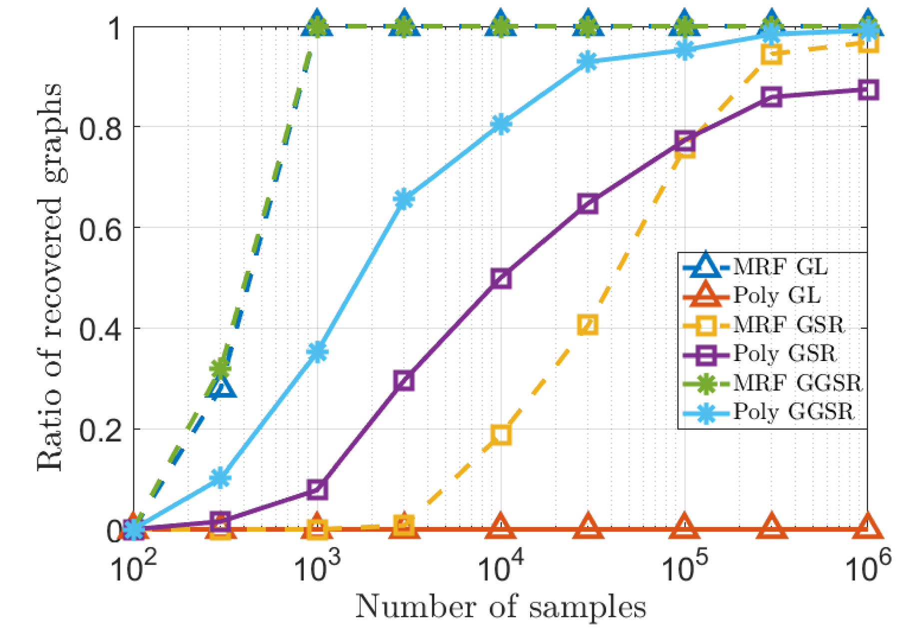
$$\mathbf{S}^{(t+1)} = \underset{\mathbf{S} \in \mathcal{S}}{\text{argmin}} \; \rho\|\mathbf{S}\|_1 + \langle \mathbf{Y}^{(t)}, \mathbf{\Theta}_2^{(t+1)}\mathbf{S} - \mathbf{S}\mathbf{\Theta}_2^{(t+1)} \rangle + \frac{\lambda}{2}\|\mathbf{\Theta}_2^{(t+1)}\mathbf{S} - \mathbf{S}\mathbf{\Theta}_2^{(t+1)}\|_F^2.$$

- Update $\mathbf{Y}$ and $\mathbf{Z}$ given $\mathbf{\Theta}_1^{(t+1)}$, $\mathbf{\Theta}_2^{(t+1)}$ $\mathbf{S}^{(t+1)}$, $\mathbf{Y}^{(t)}$, and $\mathbf{Z}^{(t)}$

$$\mathbf{Y}^{(t+1)} = \mathbf{Y}^{(t)} + \lambda\left(\mathbf{\Theta}_2^{(t+1)}\mathbf{S}^{(t+1)} - \mathbf{S}^{(t+1)}\mathbf{\Theta}_2^{(t+1)}\right),$$

Numerical Results - Influence of the number of samples

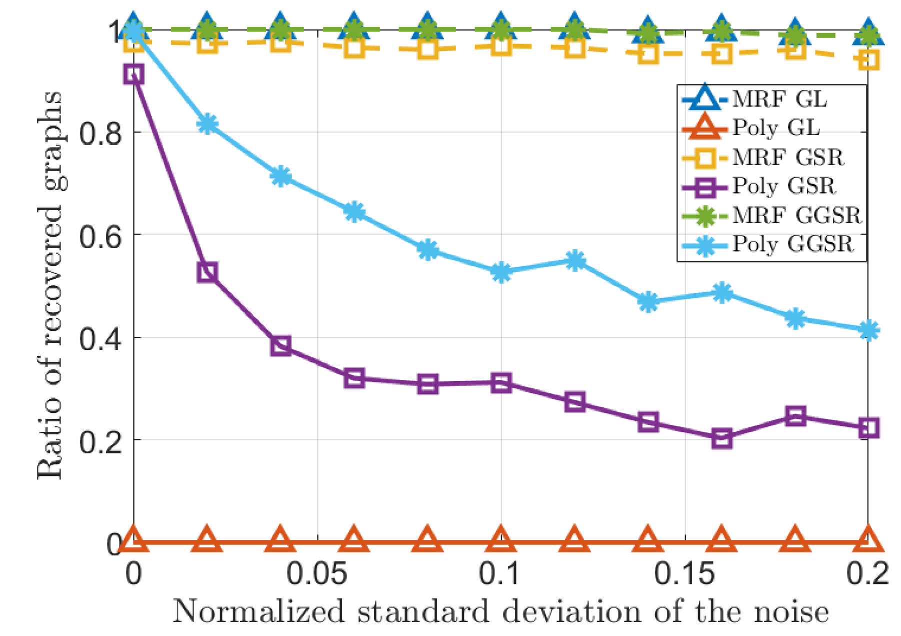
- Recovery performance for different algorithms from sample covariance considering 2 setups $\mathbf{\Sigma} = (\sigma\mathbf{I} + \delta\mathbf{S})^{-1}$ (MRF) and $\mathbf{\Sigma} = \sum_{p=0}^{2P-2} c_p \mathbf{S}^p$ (Poly) using ER graphs with N=20



  - ⇒ Similar performance results for GGSR and GL with MRF setup
  - ⇒ Significant improvement for GGSR compared to GSR with Poly setup

Numerical Results - Influence of noisy samples

- Recovery performance achieved by different algorithms for ER graphs when varying the noise power present in the signals ($R = 10^6$)



  - ⇒ The impact of the noise is more noticeable for Poly than for MRF setup
  - ⇒ For Poly setup GGSR is less sensitive to noise than GSR

Financial data experiment

- Time-varying graph learning for investment strategies [Cardoso20]
  - ⇒ How good the graph estimates are? ⇒ No real ground truth
  - ⇒ Approach: Using the graph to design an investment strategy