BLIND IDENTIFICATION OF GRAPH FILTERS WITH SPARSE INPUTS

Santiago Segarra*, Gonzalo Mateos†, Antonio G. Marques‡, and Alejandro Ribeiro* E-mail: ssegarra@seas.upenn.edu

*Electrical & Systems Engineering, University of Pennsylvania †Electrical & Computer Engineering, University of Rochester ‡Signal Theory & Communications, King Juan Carlos University

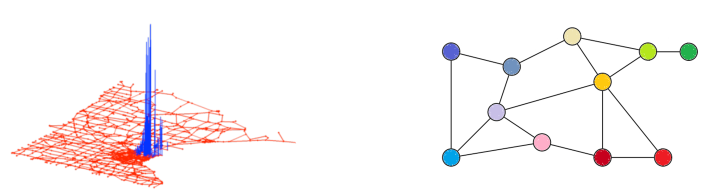
Penn UNIVERSITY of PENNSYLVANIA

Abstract

We postulate that diffusion processes can be modeled as outputs of graph filters. Leveraging recent advances in graph signal processing and classical blind deconvolution, we propose a convex algorithm for blind identification of graph filters with sparse inputs. This task amounts to finding the sources and diffusion coefficients that gave rise to an observed network state.

Graph signal processing - 101

- Network as graph $G = (\mathcal{V}, \mathcal{E}, W)$: encode pairwise relationships
- Interest here not in G itself, but in data associated with nodes in $\mathcal{V}$
  - ⇒ The object of study is a graph signal
- Ex: Opinion profile, buffer congestion levels, neural activity, epidemic

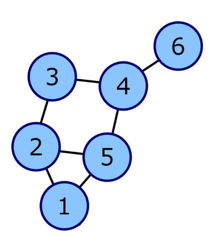


$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_{|\mathcal{V}|} \end{bmatrix} = \begin{bmatrix} 0.6 \\ \vdots \\ 0.7 \end{bmatrix}$$

- Graph SP: need to broaden classical SP results to graph signals
  - ⇒ Our view: GSP well suited to study network processes

Graph signals and graph-shift operator

- (Node) graph signals are mappings $x : \mathcal{V} \to \mathbb{R}$
  - ⇒ May be represented as a vector $\mathbf{x} \in \mathbb{R}^N$ (with $|\mathcal{V}| = N$)
- Graph G is endowed with a graph-shift operator $\mathbf{S}$
  - ⇒ Matrix $\mathbf{S} \in \mathbb{R}^{N \times N}$ satisfying: $S_{ij} = 0$ for $i \neq j$ and $(i, j) \notin \mathcal{E}$



$$\mathbf{S} = \begin{pmatrix} S_{11} & S_{12} & 0 & 0 & S_{15} & 0 \\ S_{21} & S_{22} & S_{23} & 0 & S_{25} & 0 \\ 0 & S_{23} & S_{33} & S_{34} & 0 & 0 \\ 0 & 0 & S_{43} & S_{44} & S_{45} & S_{46} \\ S_{51} & S_{52} & 0 & S_{54} & S_{55} & 0 \\ 0 & 0 & 0 & S_{64} & 0 & S_{66} \end{pmatrix}$$

$\mathbf{S}$ captures local structure in G

- Ex: Adjacency $\mathbf{A}$, Degree $\mathbf{D}$ and Laplacian $\mathbf{L}$
  - ⇒ Time-shift operator when $\mathbf{S} = \mathbf{A}_{dc}$ for G a directed cycle

Locality of S and frequency-domain representation

- $\mathbf{S}$ is a local linear operator ⇒ If $\mathbf{y} = \mathbf{S}\mathbf{x}$, $y_i = \sum_{j \in \mathcal{N}_i} S_{ij} x_j$ ⇒ 1-hop info
- Spectrum of $\mathbf{S}$ useful to analyze $\mathbf{x}$
  - ⇒ Consider diagonalizable shifts $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$
- Leverage $\mathbf{S}$ to define graph Fourier transform (GFT) and iGFT

$$\tilde{\mathbf{x}} = \mathbf{V}^{-1}\mathbf{x}, \qquad \mathbf{x} = \mathbf{V}\tilde{\mathbf{x}} \qquad \text{(Ex: DFT, PCA)}$$

- Key message: the two basic elements of GSP are $\mathbf{x}$ and $\mathbf{S}$

Linear (shift-invariant) graph filter

- A graph filter $H : \mathbb{R}^N \to \mathbb{R}^N$ is a map between graph signals
  - ⇒ Focus on linear filters ⇒ $N \times N$ matrix
- Filter $\mathbf{H}$ is a polynomial in $\mathbf{S}$ of degree L, with coeffs. $\mathbf{h} = [h_0, \dots, h_L]^T$

$$\mathbf{H} := h_0\mathbf{S}^0 + h_1\mathbf{S}^1 + \ldots + h_L\mathbf{S}^L = \sum_{l=0}^{L} h_l \mathbf{S}^l$$

- Key properties: shift-invariance and distributed implementation
  - ⇒ Satisfies $\mathbf{H}(\mathbf{S}\mathbf{x}) = \mathbf{S}(\mathbf{H}\mathbf{x})$, only L-hop information to form $\mathbf{y} = \mathbf{H}\mathbf{x}$

Frequency response of a graph filter

- Using $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, filter is $\mathbf{H} = \sum_{l=0}^{L} h_l \mathbf{S}^l = \mathbf{V}\left(\sum_{l=0}^{L} h_l \mathbf{\Lambda}^l\right)\mathbf{V}^{-1}$
- Since $\mathbf{\Lambda}^l$ are diagonal, the GFT-iGT can be used to write $\mathbf{y} = \mathbf{H}\mathbf{x}$ as

$$\tilde{\mathbf{y}} = diag(\tilde{\mathbf{h}})\tilde{\mathbf{x}}$$

  - ⇒ Output at frequency k depends only on input at frequency k
- Frequency response of filter $\mathbf{H}$ is $\tilde{\mathbf{h}} = \mathbf{\Psi}\mathbf{h}$, where $\mathbf{\Psi}$ is Vandermonde

$$\mathbf{\Psi} := \begin{pmatrix} 1 & \lambda_1 & \dots & \lambda_1^L \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_N & \dots & \lambda_N^L \end{pmatrix}$$

- Note that GFT for signals ($\tilde{\mathbf{x}} = \mathbf{V}^{-1}\mathbf{x}$) and filters ($\tilde{\mathbf{h}} = \mathbf{\Psi}\mathbf{h}$) is different
  - ⇒ If $\mathbf{S} = \mathbf{A}_{dc}$ (periodic signal), both $\mathbf{\Psi}$ and $\mathbf{V}^{-1}$ equal the DFT

Diffusion processes as graph filter outputs

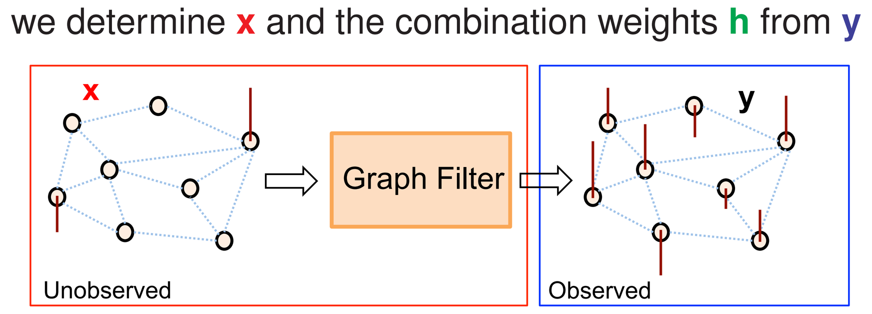
- Q: Upon observing a graph signal $\mathbf{y}$, how was this signal generated?
- Postulate the following generative model
  - ⇒ An originally sparse signal $\mathbf{x} = \mathbf{x}^{(0)}$
  - ⇒ Diffused via linear graph dynamics $\mathbf{S}$ ⇒ $\mathbf{x}^{(l)} = \mathbf{S}\mathbf{x}^{(l-1)}$
  - ⇒ Observed $\mathbf{y}$ is a linear combination of the diffused signals $\mathbf{x}^{(l)}$

$$\mathbf{y} = \sum_{l=0}^{L} h_l \mathbf{x}^{(l)} = \sum_{l=0}^{L} h_l \mathbf{S}^l \mathbf{x} = \mathbf{H}\mathbf{x}$$

- View few elements in $\text{supp}(\mathbf{x}) =: \{i : x_i \neq 0\}$ as sources or seeds

Motivation and problem formulation

- Global opinion profile formed by spreading a rumor
  - ⇒ What was the rumor? Who started it?
  - ⇒ How do people combine the opinions heard to form their own?
- Q: Can we determine $\mathbf{x}$ and the combination weights $\mathbf{h}$ from $\mathbf{y} = \mathbf{H}\mathbf{x}$?



- Problem: Blind identification of graph filters with sparse inputs
  - ⇒ Generalizes classic blind deconvolution to graphs
- Ill-posed ⇒ $(L + 1) + N$ unknowns and N observations
  - ⇒ Assume $\mathbf{x}$ is S-sparse i.e., $\|\mathbf{x}\|_0 := |\text{supp}(\mathbf{x})| \leq S$

"Lifting" the bilinear inverse problem

- Leverage the frequency response of graph filters ($\mathbf{U} := \mathbf{V}^{-1}$)

$$\mathbf{y} = \mathbf{V}\text{diag}(\mathbf{\Psi}\mathbf{h})\mathbf{U}\mathbf{x}$$

  - ⇒ $\mathbf{y}$ is a bilinear function of $\mathbf{h}$ and $\mathbf{x}$
- Blind graph filter identification ⇒ Non-convex feasibility problem

$$\text{find } \{\mathbf{h}, \mathbf{x}\}, \quad \text{s. to} \quad \mathbf{y} = \mathbf{V}\text{diag}(\mathbf{\Psi}\mathbf{h})\mathbf{U}\mathbf{x}, \quad \|\mathbf{x}\|_0 \leq S.$$

- Key observation: Using the Khatri-Rao product $\odot$ can write $\mathbf{y}$ as

$$\mathbf{y} = \mathbf{V}(\mathbf{\Psi}^T \odot \mathbf{U}^T)^T \text{vec}(\mathbf{x}\mathbf{h}^T) \tag{1}$$

  - ⇒ Reveals $\mathbf{y}$ is a linear combination of the entries of $\mathbf{Z} := \mathbf{x}\mathbf{h}^T$
- Matrix $\mathbf{Z}$ is of rank-1 and row-sparse ⇒ Linear matrix inverse problem

$$\min_{\mathbf{Z}} \text{rank}(\mathbf{Z}), \quad \text{s. to } \mathbf{y} = \mathbf{V}(\mathbf{\Psi}^T \odot \mathbf{U}^T)^T \text{vec}(\mathbf{Z}), \quad \|\mathbf{Z}\|_{2,0} \leq S$$

  - ⇒ Pseudo-norm $\|\mathbf{Z}\|_{2,0}$ counts the non-zero rows of $\mathbf{Z}$

Algorithmic approach via convex relaxation

- Rank minimization s. to row-cardinality constraint is NP-hard. Relax!
  - ⇒ Nuclear norm $\|\mathbf{Z}\|_* := \sum_k \sigma_k(\mathbf{Z})$ a convex proxy of rank
  - ⇒ ℓ_2/ℓ_1 mixed norm $\|\mathbf{Z}\|_{2,1} := \sum_{i=1}^{N} \|\mathbf{z}_i^T\|_2$ surrogate of $\|\mathbf{Z}\|_{2,0}$
- Convex relaxation

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_* + \alpha\|\mathbf{Z}\|_{2,1}, \quad \text{s. to} \quad \mathbf{y} = \mathbf{V}\left(\mathbf{\Psi}^T \odot \mathbf{U}^T\right)^T \text{vec}(\mathbf{Z})$$

  - ⇒ Scalable algorithm using method of multipliers
- Refine estimates via iteratively-reweighted optimization
  - ⇒ Weights $\alpha_i(k) = (\|\mathbf{z}_i(k)^T\|_2 + \delta)^{-1}$ per row i, per iteration k

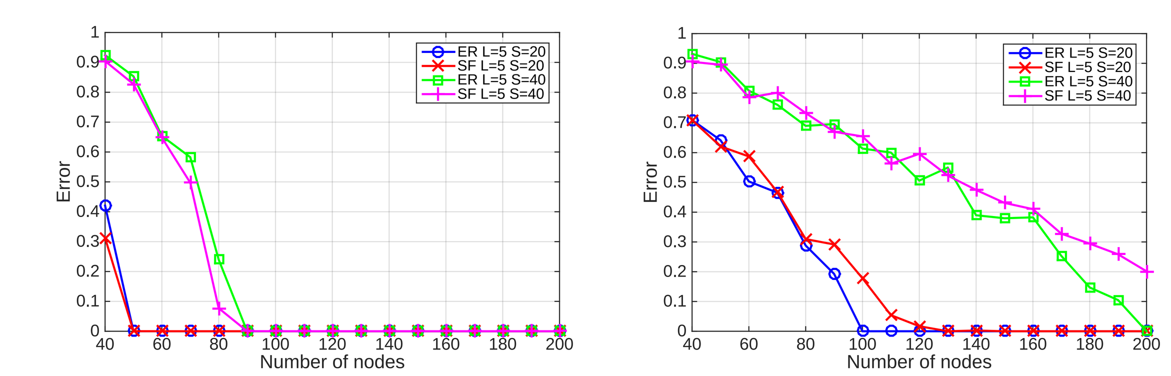
Multiple output signals

- Leverage multiple output signals to aid the blind identification task
- We have access to a collection of P output signals $\{\mathbf{y}_p\}_{p=1}^P$
  - ⇒ Corresponding to different sparse inputs $\mathbf{x}_p$ but a *common* filter $\mathbf{H}$
- Consider the stacked vectors $\bar{\mathbf{y}} := [\mathbf{y}_1^T, ..., \mathbf{y}_P^T]^T$ and $\bar{\mathbf{x}} := [\mathbf{x}_1^T, ..., \mathbf{x}_P^T]^T$
- Define the rank-one matrices $\mathbf{Z}_p := \mathbf{x}_p\mathbf{h}^T$, $p = 1, ..., P$, and stack them:
  - ⇒ (i) Vertically in $\bar{\mathbf{Z}}_v := [\mathbf{Z}_1{}^T, ..., \mathbf{Z}_P{}^T]^T = \bar{\mathbf{x}}\mathbf{h}^T \in \mathbb{R}^{NP \times L}$
  - ⇒ (ii) Horizontally in $\bar{\mathbf{Z}}_h := [\mathbf{Z}_1, ..., \mathbf{Z}_P] \in \mathbb{R}^{N \times PL}$.
- Note that $\bar{\mathbf{Z}}_v$ is a rank-one matrix and $\bar{\mathbf{Z}}_h$ is row-sparse

$$\min_{\{\mathbf{Z}_p\}_{p=1}^P} \|\bar{\mathbf{Z}}_v\|_* + \tau\|\bar{\mathbf{Z}}_h\|_{2,1}, \quad \text{s. to} \quad \bar{\mathbf{y}} = \left(\mathbf{I}_P \otimes \left(\mathbf{V}\left(\mathbf{\Psi}^T \odot \mathbf{U}^T\right)^T\right)\right) \text{vec}(\bar{\mathbf{Z}}_h)$$

Numerical tests: Known support, random graph models

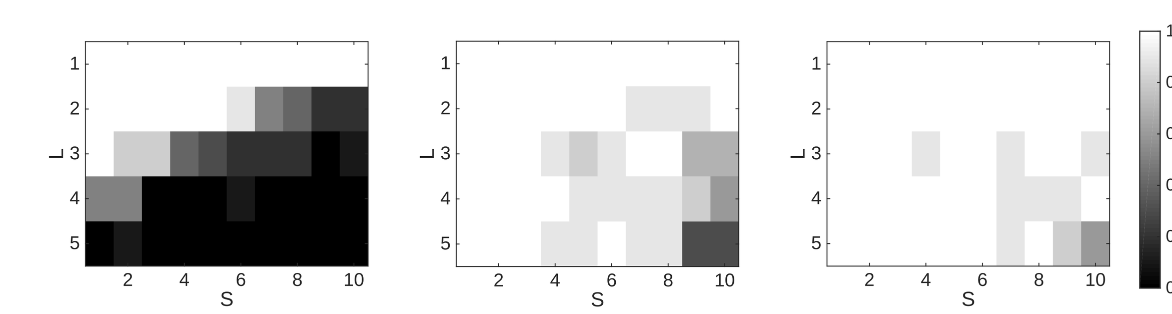
- Performance in Erdős-Rényi and scale-free graphs of varying size
  - ⇒ Assume known $\text{supp}(\mathbf{x})$ ⇒ $\mathbf{x} = [\bar{\mathbf{x}}^T, \mathbf{0}]^T$
  - ⇒ Error quantified as $\|\bar{\mathbf{x}}^*\mathbf{h}^{*T} - \bar{\mathbf{x}}\mathbf{h}^T\|_F$
  - ⇒ Two settings (i) $L = 5$, $S = 20$; and (ii) $L = 5$, $S = 40$
  - ⇒ Nuclear norm (left) vs. naive least-squares of (1) (right)



- Rank minimization achieves perfect recovery when $N \geq 2(L + S)$
  - ⇒ Well-below $N_0 := L \times S$ needed for least-squares to succeed
  - ⇒ Rank minimization is more robust to the type of graph

Recovery rate in random graphs: unknown support

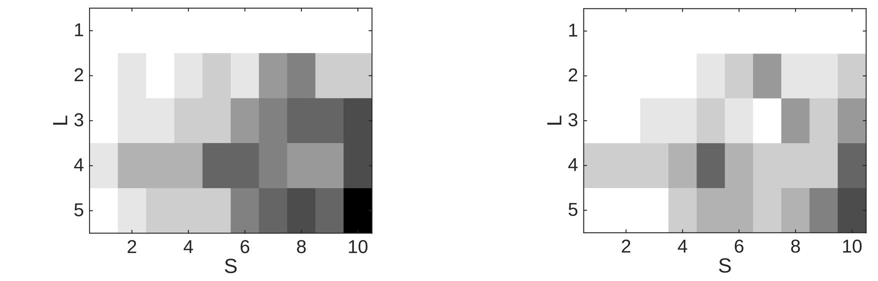
- Recovery rates on Erdős-Rényi graphs ($N = 50$) for varying L and S
- $P = 1$ (left), $P = 1$ + reweighted $\ell_{2,1}$ (mid), $P = 5$ + reweighted $\ell_{2,1}$ (right)



- Exact recovery over non-trivial (L, S) region
  - ⇒ Iteratively-reweighted optimization markedly improves recovery
  - ⇒ Multiple outputs further increase recovery success

Recovery rate in a brain graph: unknown support

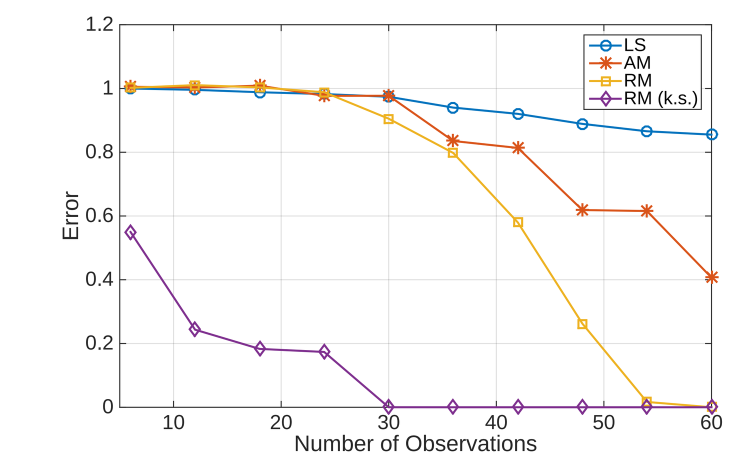
- Consider a brain structural graph ($N = 66$) [Hagmann]
- $P = 1$ + reweighted $\ell_{2,1}$ (left), $P = 5$ + reweighted $\ell_{2,1}$ (right)



- Encouraging results even for real-world graphs
  - ⇒ Gradual performance decay for increasing L and S

Performance comparison with alternative methods

- Human brain graph of $N = 66$ brain regions, $L = 6$ and $S = 6$



- Proposed method outperforms alternating-minimization and LS solvers
  - ⇒ Unknown $\text{supp}(\mathbf{x}) \approx$ Need twice as many observations

Discussion and road ahead

- Identifiability conditions
  - ⇒ Q: When is $\{\mathbf{x}, \mathbf{h}\}$ the unique solution (up to scaling)?
  - ⇒ Deterministic or probabilistic model assumptions
- Exact recovery conditions
  - ⇒ Q: When does the convex relaxation succeed?
  - ⇒ Lower bound on N to guarantee recovery for given L and S
  - ⇒ Depends on algebraic features of the graph-shift $\mathbf{S}$
  - ⇒ Some graphs are more amenable to blind identification that others
- Unknown shift $\mathbf{S}$ ⇒ Network topology inference
- Envisioned application domains
  - ⇒ Opinion formation in social networks
  - ⇒ Identify sources of epileptic seizure
  - ⇒ Event-driven information cascades
  - ⇒ Trace "patient zero" for an epidemic outbreak

References

- D. I. Shuman, et al., "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, Mar. 2013.
- A. Sandryhaila and J.M.F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.
- S. Segarra, A. G. Marques, and A. Ribeiro, "Distributed implementation of linear network operators using graph filters," in *Proc. 53rd Allerton Conf. on Commun. Control and Computing*, Monticello, IL, Sept. 30- Oct. 2, 2015.
- A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1711–1732, Mar. 2014.
- P. Hagmann, et al., "Mapping the structural core of human cerebral cortex," *PLoS Biol*, vol. 6, no. 7, pp. e159, 2008.
- S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," arXiv preprint, 2015.