Joint Signal Recovery and Graph Learning from Incomplete Time-Series

Amirhossein Javaheri[†⋆], Arash Amini[⋆], Farokh Marvasti[⋆], Daniel P. Palomar[†]
sajavaheri@connect.ust.hk, aamini@sharif.edu, marvasti@sharif.edu, palomar@ust.hk

[†]The Hong Kong University of Science and Technology
[⋆]Sharif University of Technology

2024 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2024)

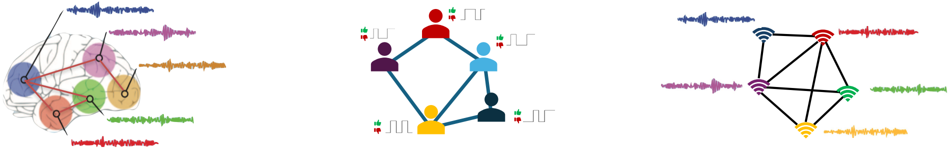
April, 2024

Outline

Outline

What Is a Graph Signal?

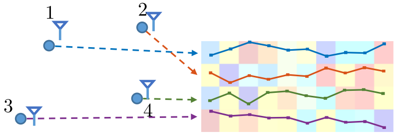
- Graphs provide efficient representation tools for data in a variety of applications in signal processing, machine learning, finance, etc [Dong, Thanou, Rabbat, *et al.* 2019; Marques *et al.* 2020].
- A weighted graph is denoted with $G = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$ ($\mathcal{V}$ vertex set, $\mathcal{E}$ edge set, and $\mathbf{W}$ (weighted) adjacency matrix).
- For an undirected graph (symmetric $\mathbf{W}$), one may represent the graph with edge weights vector $\mathbf{w}$. There are equivalent representations via adjacency/Laplacian operator: $\mathbf{W} = \mathcal{A}(\mathbf{w})$, $\mathbf{L} = \mathrm{Diag}(\mathbf{W}\mathbf{1}) - \mathbf{W} = \mathcal{L}(\mathbf{w})$ [Kumar *et al.* 2020].
- A (time-varying) graph signal $\mathbf{x}_t = f(\mathcal{V}; t)$ is a time series with spatio-temporal (vertex/time domain) correlations.

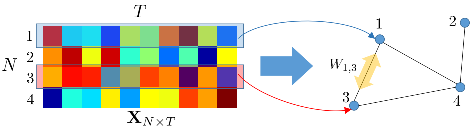


Brain Network Social Network Sensor Network

How to Represent Data Matrices with Graphs?

- Suppose, we are given N measurements of a (time-varying) graph signal $\mathbf{x}_t \in \mathbb{R}^N$ as $\mathbf{X} = [\mathbf{x}_1 \ldots, \mathbf{x}_T] \in \mathbb{R}^{N \times T}$.
- Each row of $\mathbf{X}$ is a time-series (time samples) corresponding to a vertex of the graph.
- An example: $\mathbf{x}_t$ is the prices of N stocks in a financial market and T is the number of daily measurements.
- A weighted undirected graph can model similarity (correlations) between elements (the higher $W_{i,j}$, the more similar (correlated) the time series at vertices i and j will be).

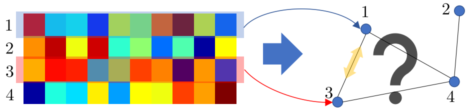




Graph Learning

Problem

- Given complete data, the goal is to find a graph structure that models inter-connected similarities/dependencies.



Graph learning given the (complete) data

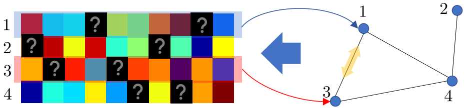
Related Works

- Algorithms that use probabilistic methods via a Gaussian Markov Random Field (GMRF) model, e.g., [Egilmez *et al.* 2017; Kumar *et al.* 2020; Lake & Tenenbaum 2010; Zhao *et al.* 2019], or deterministic regularization criteria such as smoothness [Kalofolias 2016] or stationarity [Segarra *et al.* 2016].

Graph Signal Recovery

Problem

- Given the underlying graphical model, the goal is to recover (impute) the signal.



Recovery of the data given the underlying graph structure

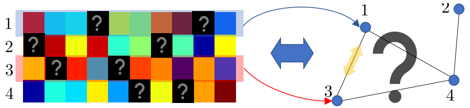
Related Works

- Incorporating properties such as least total-variation [Chen *et al.* 2015], stationarity [Perraudin & Vandergheynst 2017], spatio-temporal smoothness [Qiu *et al.* 2017], sparsity [Safavi *et al.* 2018] of the signal in a graph representation domain for imputation.

Joint Graph Learning and Signal Recovery

Problem

- The goal is to simultaneously impute the signal and infer the the underlying graphical models.



Joint signal recovery and graph learning

Related Works

Stochastic approaches to joint undirected graph learning and signal denoising using smoothness [Dong, Thanou, Frossard, *et al.* 2016] (GL-SigRep) and long-short term characteristics [Liu *et al.* 2020] (GL-LRSS) Or deterministic approaches for joint directed graph learning and signal recovery via Vector Autoregressive (VAR) model [Ioannidis *et al.* 2019] (JISG)

Outline

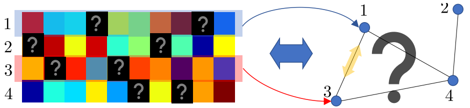
Proposed Method

Assumptions:

- Assume a connected undirected graph that models similarity in temporal variations of the signal elements. The larger $W_{i,j}$, the more similar the i-th and j-th components of the signal vary in time.
- The observations of the original signal have missing entries

Goal:

- Graph learning from missing data or semi-blind recovery of graph signal (no graph prior)
- Learn the graph and recover (impute) the signal in a jointly fashion.



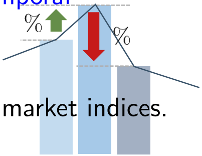
Joint signal recovery and graph learning (the investigated problem)

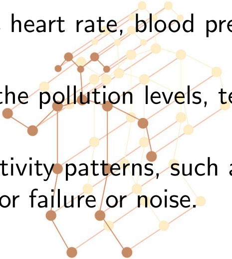
Applications:

- Our model can be applied for graph signals (time-series) where the temporal evolution/variation is of importance

Examples:

- **Finance**: Modelling the change (rate of return) in the stock prices or market indices. Missing entries occur due to trading halts, suspensions, holidays, etc.
- **Healthcare**: Monitoring changes in vital signs such as heart rate, blood pressure, etc.. Missing values due to sensor failure or noise.
- **Environmental Monitoring**: Modelling variations in the pollution levels, temperature, etc. Missing values due to sensor failure or noise.
- **Security and Surveillance**: Monitoring changes in activity patterns, such as motion detection, sound level, etc. Missing values due to sensor failure or noise.





Proposed Method: Intuition

Why spatio-temporal smoothness?

- Assume a first order VAR model with (spatially) non-white innovations with the graph Laplacian $\mathcal{L}(\mathbf{w})$ as the precision matrix

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t \quad 1 \leq t \leq T \qquad (\mathbf{x}_0 = \mathbf{0})$$

$$p_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_t) \propto (\det{}^* \mathcal{L}(\mathbf{w}))^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\epsilon}_t^T \mathcal{L}(\mathbf{w})\boldsymbol{\epsilon}_t\right)$$

- Assume we have noisy (AWGN) and missing observations of the original signal

$$\mathbf{y}_t = \mathbf{m}_t \odot (\mathbf{x}_t + \mathbf{n}_t) \quad \mathbf{n}_t \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I})$$

- The MAP estimation of the signal and graph ($\mathbf{X}$ and $\mathbf{w}$) with some sparsity-promoting graph prior ($p(\mathbf{w}) \propto \exp(-\gamma \|\mathbf{w}\|_1)$) gives:

$$\mathbf{X}^\star, \mathbf{w}^\star = \underset{\mathbf{X}, \mathbf{w} \geq \mathbf{0}}{\operatorname{argmin}} \frac{1}{\sigma_n^2} \|\mathbf{Y} - \mathbf{M} \odot \mathbf{X}\|_F^2 + S_T(\mathbf{X}, \mathbf{w}) - T \log \det{}^* \mathcal{L}(\mathbf{w}) + \gamma \|\mathbf{w}\|_1$$

- The term $S_T(\mathbf{X}, \mathbf{w}) \triangleq \sum_{t=1}^{T} (\mathbf{x}_t - \mathbf{x}_{t-1})^\top \mathcal{L}\mathbf{w} (\mathbf{x}_t - \mathbf{x}_{t-1})$ is called spatio-temporal smoothness.

Proposed Method: Intuition

Why spatio-temporal smoothness?

- Assume i.i.d. random samples of a zero-mean GMRF with the graph Laplacian $\mathcal{L}(\mathbf{w})$ as the precision matrix

$$\mathbf{x}_t \sim \mathcal{N}\left(\mathbf{0}, \mathcal{L}(\mathbf{w})^{\dagger}\right), \qquad \mathbf{x}_{t-1} \sim \mathcal{N}\left(\mathbf{0}, \mathcal{L}(\mathbf{w})^{\dagger}\right)$$

- The difference $\mathbf{z}_t = \mathbf{x}_t - \mathbf{x}_{t-1}$ is still a zero-mean GMRF

$$\mathbf{z}_t \sim \mathcal{N}\left(\mathbf{0}, 2\mathcal{L}(\mathbf{w})^{\dagger}\right)$$

- Then for $T \to \infty$, simple (spatial) graph smoothness $S(\mathbf{X}, \mathbf{w}) = \sum_t \mathbf{x}_t^\top \mathcal{L}(\mathbf{w})\mathbf{x}_t$ would be only a factor of the spatio-temporal smoothness $S_T(\mathbf{X}, \mathbf{w})$

$$\begin{aligned}\frac{1}{T}S(\mathbf{X}, \mathbf{w}) &= \frac{1}{T}\sum_t \mathbf{x}_t^\top \mathcal{L}(\mathbf{w})\mathbf{x}_t \approx \mathrm{Tr}\left(\mathcal{L}(\mathbf{w})\mathbb{E}[\mathbf{x}_t\mathbf{x}_t^\top]\right) = \mathrm{Tr}\left(\mathcal{L}(\mathbf{w})\mathcal{L}(\mathbf{w})^{\dagger}\right) \\ &= \frac{1}{2}\mathrm{Tr}\left(\mathcal{L}(\mathbf{w})2\mathcal{L}(\mathbf{w})^{\dagger}\right) = \mathrm{Tr}\left(\mathcal{L}(\mathbf{w})\mathbb{E}[\mathbf{z}_t\mathbf{z}_t^\top]\right) \approx \frac{1}{2}\frac{1}{T}\sum_t \mathbf{z}_t^\top \mathcal{L}(\mathbf{w})\mathbf{z}_t = \frac{1}{2T}S_T(\mathbf{X}, \mathbf{w})\end{aligned}$$

Conclusion

The spatio-temporal smoothness assumption works for both i.i.d. and time-dependent signals

Proposed Method: Problem Formulation

Problem Formulation

$$\mathbf{X}^\star, \mathbf{w}^\star = \underset{\mathbf{X}, \mathbf{w} \geq \mathbf{0}}{\operatorname{argmin}} f(\mathbf{X}, \mathbf{w})$$

$$f(\mathbf{X}, \mathbf{w}) \triangleq \|\mathbf{Y} - \mathbf{M} \odot \mathbf{X}\|_F^2 + \alpha S_T(\mathbf{X}, \mathbf{w}) - \beta R(\mathbf{w}) + \gamma \|\mathbf{w}\|_1$$

- $\|\mathbf{Y} - \mathbf{M} \odot \mathbf{X}\|_F^2$: Fidelity measure (similarity to the observation)
  - $\mathbf{M}$: The missing/sampling mask
  - $\mathbf{Y}$: Observations of the original signal: $\mathbf{y}_t = \mathbf{m}_t \odot \mathbf{x}_t$ or $\mathbf{Y} = \mathbf{M} \odot \mathbf{X}$
- $S_T(\mathbf{X}, \mathbf{w})$: Spatio-temporal smoothness measure: $S_T(\mathbf{X}, \mathbf{w}) = \operatorname{Tr}\left(\mathcal{L}(\mathbf{w})\Delta(\mathbf{X})\Delta(\mathbf{X})^\top\right)$
  - $\mathbf{D}$: The (first order) difference matrix: $\mathbf{D} = \sum_{t=1}^{T} \mathbf{e}_{t-1}\mathbf{e}_t^\top$
  - $\Delta(\mathbf{X})$: The (first order) difference signal with columns $\Delta_t(\mathbf{X}) := \mathbf{x}_t - \mathbf{x}_{t-1}$ $(\Delta(\mathbf{X}) := \mathbf{X} - \mathbf{D}\mathbf{X})$
- $R(\mathbf{w})$: Regularization term to enforce connected graph structure:

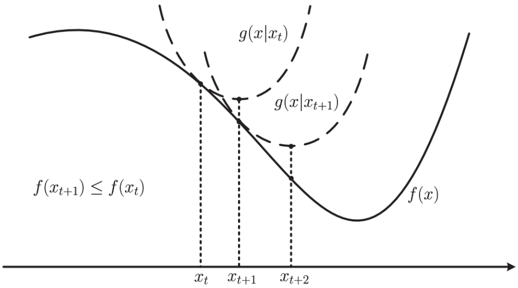
$$R(\mathbf{w}) = -\log\det(\mathcal{L}(\mathbf{w}) + \mathbf{J}), \qquad \mathbf{J} = (1/N)\mathbf{1}\mathbf{1}^\top$$

- $\|\mathbf{w}\|_1$: Sparsity promoting term (need to apply threshold to be effective)

Proposed Method: Solution

Optimization Algorithm:

- We use block Majorization-Minimization (MM) [Sun *et al.* 2017] or the Block Successive Upperbound Minimization (BSUM) [Razaviyayn *et al.* 2013] to solve the problem.
- We have two (block) variables $\mathbf{X}$ and $\mathbf{w}$ $\rightarrow$ we have two update steps.
- In each update step fix one (block) variable, and minimize a majorizer over the other (block) variable.



Solution: Signal ($\mathbf{X}$) Update

$\mathbf{X}$-subproblem

$$\mathbf{X}^\star = \underset{\mathbf{X}}{\mathrm{argmin}}\, f_{\mathbf{X}}(\mathbf{X})$$

$$f_{\mathbf{X}}(\mathbf{X}) = \mathrm{Tr}\left((\mathbf{Y} - \mathbf{M} \odot \mathbf{X})(\mathbf{Y} - \mathbf{M} \odot \mathbf{X})^\top\right) + \alpha\, \mathrm{Tr}\left(\mathcal{L}(\mathbf{w})\Delta(\mathbf{X})\Delta(\mathbf{X})^\top\right) + \text{const.}$$

$\mathbf{X}$-update steps:

- Vectorization: Restate the $f_{\mathbf{X}}(\mathbf{X})$ in vectorized form

$$f_{\mathbf{X}}(\mathbf{X}) = \mathrm{vec}(\mathbf{X})^\top \mathbf{G}\mathrm{vec}(\mathbf{X}) - 2\mathrm{vec}(\mathbf{X})^\top \mathbf{b} + \text{const}$$

$$\mathbf{G} = \mathrm{Diag}(\mathrm{vec}(\mathbf{M})) + \alpha \mathbf{H}^\top(\mathbf{I}_T \otimes \mathcal{L}(\mathbf{w}))\mathbf{H}, \quad \mathbf{H} = \mathbf{I}_{NT} - \mathbf{D}^\top \otimes \mathbf{I}_N$$

- Majorization: Find a majorizer for less complex solution (compared to inverting $\mathbf{G}$)

$$f^S_{\mathbf{X}}(\mathbf{X};\mathbf{X}_0) = f_{\mathbf{X}}(\mathbf{X}) + \mathrm{vec}(\mathbf{X} - \mathbf{X}_0)^\top(\theta \mathbf{I}_{NT} - \mathbf{G})\mathrm{vec}(\mathbf{X} - \mathbf{X}_0) \geq f_{\mathbf{X}}(\mathbf{X})$$

A sufficient condition for this upperbound to hold is if $\theta > 1 + 4\alpha \left\|\mathcal{L}(\mathbf{w})\right\| \geq \left\|\mathbf{G}\right\|$

- Minimization: Minimize $f^S_{\mathbf{X}}(\mathbf{X};\mathbf{X}_0)$ for $\mathbf{X}_0 = \mathbf{X}^j$ to obtain $\mathbf{X}^{(j+1)}$

$$\mathbf{X}^{(j+1)} = \underset{\mathbf{X}}{\mathrm{argmin}}\, f^S_{\mathbf{X}}(\mathbf{X};\mathbf{X}^{(j)}) = \mathbf{X}^{(j)} - \frac{1}{2\theta}\frac{\partial}{\partial \mathbf{X}} f_{\mathbf{X}}(\mathbf{X}^{(j)})$$

Solution: Graph ($\mathbf{w}$) Update

$\mathbf{w}$-subproblem

$$\mathbf{w}^{\star} = \underset{\mathbf{w}}{\text{argmin}}\, f_{\mathbf{w}}(\mathbf{w})$$

$$f_{\mathbf{w}}(\mathbf{w}) = \text{Tr}(\mathcal{L}(\mathbf{w})\mathbf{K}) - \log\det(\mathcal{L}(\mathbf{w}) + \mathbf{J}), \qquad \mathbf{K} = \frac{1}{\beta}\left(\alpha\Delta(\mathbf{X})\Delta(\mathbf{X})^{\top} + \gamma/2\mathbf{H}_{\text{off}}\right)$$

$\mathbf{w}$-update steps:

- First Majorization: Linear approximation of the concave function $\log\det((\mathcal{L}(\mathbf{w}) + \mathbf{J})^{-1})$

$$-\log\det(\mathcal{L}(\mathbf{w}) + \mathbf{J}) \leq \text{Tr}\left(\mathbf{F}_0(\mathbf{G}\text{Diag}(\tilde{\mathbf{w}})\mathbf{G}^{\top})^{-1}\right) - \log\det(\mathcal{L}(\mathbf{w}_0) + \mathbf{J}) - N$$

Here $\mathbf{F}_0 = \mathcal{L}(\mathbf{w}_0) + \mathbf{J}$, $\mathbf{G} = [\mathbf{E}, \mathbf{1}]$, $\tilde{\mathbf{w}} = [\mathbf{w}^{\top}\ 1/N]^{\top}$ and $\mathbf{w}_0$ is a fixed (previous) point
Also $\mathbf{E} = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N(N-1)/2}] \in \mathbb{R}^{N\times N(N-1)/2}$ consists of vectors $\boldsymbol{\xi}_k$ for $k = i - j + \frac{j-1}{2}(2N - j)$, $i > j$, each of which has a $+1$ at the j-th position, a -1 at the i-th position, and zeros elsewhere.

Solution: Graph ($\mathbf{w}$) Update

$\mathbf{w}$-update steps:

- Double Majorization!: Use an inequality from linear algebra

$$\begin{aligned}\mathrm{Tr}\left(\mathbf{F}_0(\mathbf{G}\mathrm{Diag}(\tilde{\mathbf{w}})\mathbf{G}^\top)^{-1}\right) &\leq \mathrm{Tr}\left(\mathbf{F}_0^{-1}\mathbf{G}\mathrm{Diag}(\tilde{\mathbf{w}}_0^{\circ 2} \oslash \tilde{\mathbf{w}})\mathbf{G}^\top\right) \\ &= \langle \mathbf{w}_0^{\circ 2} \oslash \mathbf{w}, \mathcal{L}^\star(\mathbf{F}_0^{-1})\rangle + \mathrm{Tr}\left(\mathbf{F}_0^{-1}\mathbf{J}\right)\end{aligned}$$

- Final Majorization!!: Add $\sum_i \tau q_i {w_0}_i^2 h(w_i/{w_0}_i)$ with $h(x) = x + \frac{1}{x} - 2 \geq 0$ for $x > 0$

$$\begin{aligned}f_{\mathbf{w}}(\mathbf{w}) \leq f_{\mathbf{w}}^S(\mathbf{w};\mathbf{w}_0) \triangleq \tau\langle \mathbf{q} \odot \mathbf{w}_0^{\circ 2}, \mathbf{w} \oslash \mathbf{w}_0 + (\mathbf{w}_0 + 1/\tau) \oslash \mathbf{w} - 2\rangle + \\ \langle \mathbf{w}, \mathbf{r}\rangle + \mathrm{Tr}\left((\mathcal{L}(\mathbf{w}_0) + \mathbf{J})^{-1}\mathbf{J}\right) - \log\det(\mathcal{L}(\mathbf{w}_0) + \mathbf{J}) - N\end{aligned}$$

Here $\mathbf{r} = \mathcal{L}^*(\mathbf{K})$, $\mathbf{q} = \mathcal{L}^*((\mathcal{L}\mathbf{w}_0 + \mathbf{J})^{-1})$, and $\tau > 0$ is a constant.

- Minimization: Minimize $f_{\mathbf{w}}^S(\mathbf{w};\mathbf{w}^{(j)})$ for $\mathbf{w}_0 = \mathbf{w}^j$ to obtain $\mathbf{w}^{(j+1)}$

$$\begin{aligned}\mathbf{w}^{(j+1)} &= \underset{\mathbf{w}}{\mathrm{argmin}}\ f_{\mathbf{w}}^S(\mathbf{w};\mathbf{w}^{(j)}) \\ &= \mathbf{w}^{(j)} \odot \sqrt{(\tau\mathbf{w}^{(j)} \odot \mathbf{q} + \mathbf{q}) \oslash (\tau\mathbf{w}^{(j)} \odot \mathbf{q} + \mathbf{r})}.\end{aligned}$$

Outline

1 Overview
- What Is a Graph Signal?
- Graph Learning
- Graph Signal Recovery
- Joint Graph Learning and Signal Recovery

2 Proposed Approach
- Proposed Model
- Problem Formulation
- Solution
  - Signal Update
  - Graph Update

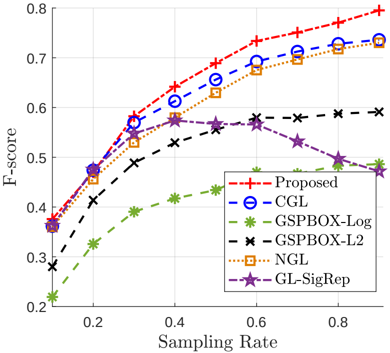
3 Numerical Results
- Synthetic Data
- Real Data

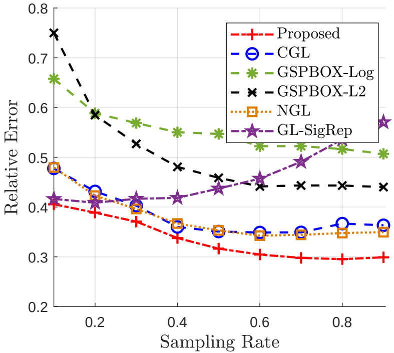
Numerical Results (Synthetic Data)

Graph Learning

- Evaluating our model for graph Laplacian $\mathbf{L} = \mathcal{L}(\mathbf{w})$ estimation from synthetic data.
- The F-score and Relative Error are used as performance metrics.

$$\text{F-score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}, \qquad \text{RelErr} = \frac{\|\mathbf{L}^\star - \widehat{\mathbf{L}}\|_F}{\|\mathbf{L}^\star\|_F}.$$



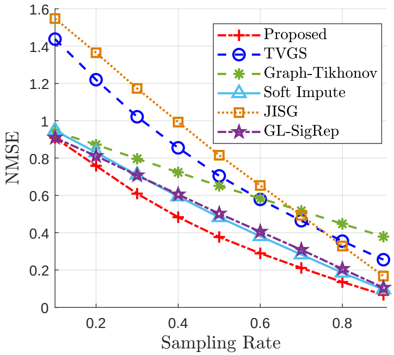


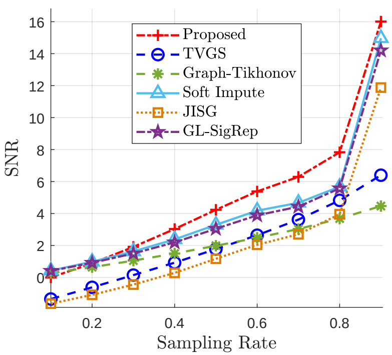
Numerical Results (Synthetic Data)

Signal Recovery

- Evaluating our model for graph signal $\mathbf{X}$ recovery from synthetic data.
- The SNR and NMSE are used as performance metrics.

$$\text{NMSE} = \frac{1}{T}\sum_{i=1}^{T}\frac{\|\mathbf{x}_i^\star - \hat{\mathbf{x}}_i\|^2}{\|\mathbf{x}_i^\star\|^2}, \quad \text{SNR} = 20\log_{10}\left(\frac{\|\mathbf{X}^\star\|_F}{\|\mathbf{X}^\star - \widehat{\mathbf{X}}\|_F}\right).$$



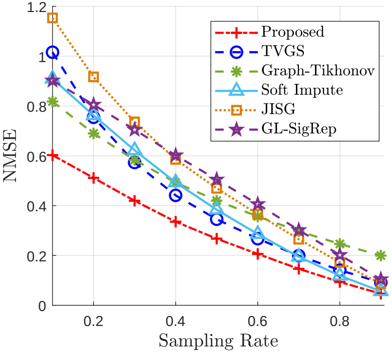


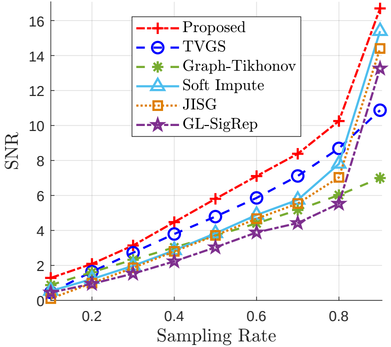
Numerical Results (Real Data)

Signal Recovery

- Evaluating our model for graph signal $\mathbf{X}$ recovery from real (US temperature) data.
- The SNR and NMSE are used as performance metrics.

$$\text{NMSE} = \frac{1}{T}\sum_{i=1}^{T}\frac{\|\mathbf{x}_i^\star - \hat{\mathbf{x}}_i\|^2}{\|\mathbf{x}_i^\star\|^2}, \quad \text{SNR} = 20\log_{10}\left(\frac{\|\mathbf{X}^\star\|_F}{\|\mathbf{X}^\star - \widehat{\mathbf{X}}\|_F}\right).$$

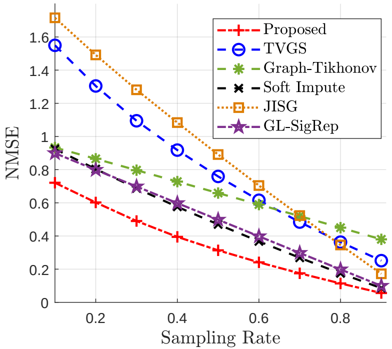


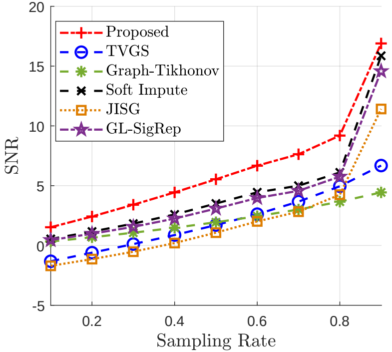


Signal Recovery

- Evaluating our model for graph signal $\mathbf{X}$ recovery from real (S&P500 stock data) data.
- The SNR and NMSE are used as performance metrics.

$$\text{NMSE} = \frac{1}{T}\sum_{i=1}^{T}\frac{\|\mathbf{x}_i^\star - \hat{\mathbf{x}}_i\|^2}{\|\mathbf{x}_i^\star\|^2}, \quad \text{SNR} = 20\log_{10}\left(\frac{\|\mathbf{X}^\star\|_F}{\|\mathbf{X}^\star - \widehat{\mathbf{X}}\|_F}\right).$$





Thanks!

Thanks for listening. For more information visit

www.danielpalomar.com
github.com/convexfi

1. Chen, S., Sandryhaila, A., Moura, J. M. F. & Kovačević, J. Signal Recovery on Graphs: Variation Minimization. *IEEE Transactions on Signal Processing* **63.** arXiv: 1411.7414, 4609–4624. ISSN: 1053-587X, 1941-0476. http://arxiv.org/abs/1411.7414 (2021) (Sept. 2015).

2. Dong, X., Thanou, D., Frossard, P. & Vandergheynst, P. Learning Laplacian Matrix in Smooth Graph Signal Representations. *IEEE Transactions on Signal Processing* **64,** 6160–6173. ISSN: 1053-587X, 1941-0476. http://ieeexplore.ieee.org/document/7552590/ (2021) (Dec. 2016).

3. Dong, X., Thanou, D., Rabbat, M. & Frossard, P. Learning graphs from data: A signal representation perspective. *IEEE Signal Processing Magazine* **36.** arXiv: 1806.00848, 44–63. ISSN: 1053-5888, 1558-0792. http://arxiv.org/abs/1806.00848 (2021) (May 2019).

4. Egilmez, H. E., Pavez, E. & Ortega, A. Graph Learning From Data Under Laplacian and Structural Constraints. *IEEE Journal of Selected Topics in Signal Processing* **11.** Conference Name: IEEE Journal of Selected Topics in Signal Processing, 825–841. ISSN: 1941-0484 (Sept. 2017).

5. Ioannidis, V. N., Shen, Y. & Giannakis, G. B. Semi-Blind Inference of Topologies and Dynamical Processes Over Dynamic Graphs. *IEEE Transactions on Signal Processing* **67,** 2263–2274. ISSN: 1053-587X, 1941-0476. https://ieeexplore.ieee.org/document/8662630/ (2021) (May 2019).

6. Kalofolias, V. *How to Learn a Graph from Smooth Signals*. en. in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics* ISSN: 1938-7228 (PMLR, May 2016), 920–929. https://proceedings.mlr.press/v51/kalofolias16.html (2023).

7. Kumar, S., Ying, J., Cardoso, J. V. d. M. & Palomar, D. P. A Unified Framework for Structured Graph Learning via Spectral Constraints. *Journal of Machine Learning Research* **21,** 1–60. http://jmlr.org/papers/v21/19-276.html (2020).

8. Lake, B. & Tenenbaum, J. *Discovering structure by learning sparse graphs*. in *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (Cognitive Science Society, Inc., Portland, Oregon, United States, Aug. 2010), 778–784.

9. Liu, Y. *et al.* Graph Learning for Spatiotemporal Signals with Long- and Short-Term Characterization. *IEEE Transactions on Signal and Information Processing over Networks* **6.** arXiv: 1911.08018, 699–713. ISSN: 2373-776X, 2373-7778. http://arxiv.org/abs/1911.08018 (2021) (2020).

10. Marques, A. G., Kiyavash, N., Moura, J. M., Van De Ville, D. & Willett, R. Graph Signal Processing: Foundations and Emerging Directions [From the Guest Editors]. *IEEE Signal Processing Magazine* **37,** 11–13. ISSN: 1053-5888, 1558-0792. https://ieeexplore.ieee.org/document/9244183/ (2022) (Nov. 2020).

11. Perraudin, N. & Vandergheynst, P. Stationary signal processing on graphs. *IEEE Transactions on Signal Processing* **65.** arXiv: 1601.02522, 3462–3477. ISSN: 1053-587X, 1941-0476. http://arxiv.org/abs/1601.02522 (2021) (July 2017).

12. Qiu, K. *et al.* Time-Varying Graph Signal Reconstruction. *IEEE Journal of Selected Topics in Signal Processing* **11,** 870–883. ISSN: 1932-4553, 1941-0484. http://ieeexplore.ieee.org/document/7979523/ (2021) (Sept. 2017).

References II

13. Razaviyayn, M., Hong, M. & Luo, Z.-Q. A Unified Convergence Analysis of Block Successive Minimization Methods for Nonsmooth Optimization. en. *SIAM Journal on Optimization* **23,** 1126–1153. ISSN: 1052-6234, 1095-7189. http://epubs.siam.org/doi/10.1137/120891009 (2022) (Jan. 2013).

14. Safavi, S. H., Khatua, M., Cheung, N.-M. & Torkamani-Azar, F. On Sparse Graph Fourier Transform. *arXiv:1811.08609 [eess].* arXiv: 1811.08609. http://arxiv.org/abs/1811.08609 (2022) (Nov. 2018).

15. Segarra, S., Marques, A. G., Mateos, G. & Ribeiro, A. Network Topology Inference from Spectral Templates. *arXiv:1608.03008 [physics].* arXiv: 1608.03008. http://arxiv.org/abs/1608.03008 (2021) (Aug. 2016).

16. Sun, Y., Babu, P. & Palomar, D. P. Majorization-Minimization Algorithms in Signal Processing, Communications, and Machine Learning. *IEEE Transactions on Signal Processing* **65,** 794–816. ISSN: 1053-587X, 1941-0476. http://ieeexplore.ieee.org/document/7547360/ (2021) (Feb. 2017).

17. Zhao, L., Wang, Y., Kumar, S. & Palomar, D. P. Optimization Algorithms for Graph Laplacian Estimation via ADMM and MM. *IEEE Transactions on Signal Processing* **67,** 4231–4244. ISSN: 1053-587X, 1941-0476. https://ieeexplore.ieee.org/document/8747497/ (2021) (Aug. 2019).