DATA-DRIVEN NETWORK INFERENCE FOR INFRASTRUCTURE SYSTEMS

GITHUB

Estéban Nocet-Binois, *supervised by Jürgen Hackl*

Department of Civil & Environmental Engineering | Princeton University, NJ, USA

MOTIVATION

The objective is to learn a basis that captures the dependencies and global modes of variation (called graph harmonics, or *frequencies*) present in the data. Analyzing how system states (like frequency deviations) decompose into these modes reveals which patterns are most likely to persist or amplify after a disturbance. Thus, given a *meaningful* basis, we can analyze how signals propagate, diffuse, or concentrate, and how oscillations or disturbances (such as power outages) manifest across the entire grid.

NOTATIONS

Graph Laplacian

- The Laplacian $\boldsymbol{L} \in \mathbb{R}^{n\times n}$ encodes how values at each node deviate from its neighborhood. We denote the eigendecomposition:

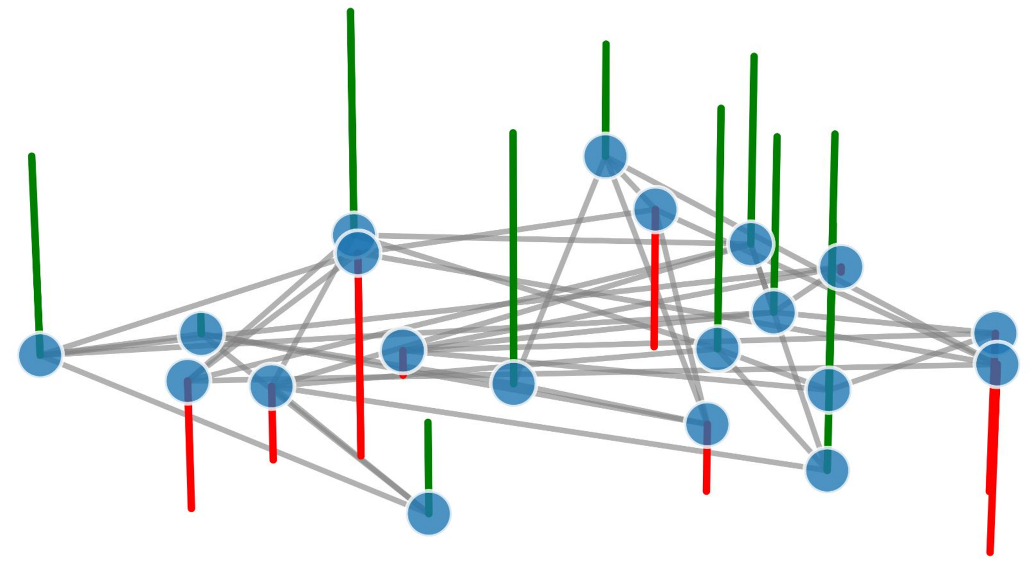
$$\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top,\ \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$$

- We note the pseudoinverse:

$$\boldsymbol{L}^\dagger = \sum_{k=1}^{n-1} \frac{1}{\lambda_k} \boldsymbol{u}_k \boldsymbol{u}_k^\top$$

Signal Matrix

- A graph signal is a n-valued tensor, collecting states across all nodes.



- Collection $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ of m signals (rows) over n nodes (columns)
- Sample covariance:

$$\boldsymbol{\Sigma} = \frac{1}{m}\boldsymbol{X}^\top\boldsymbol{X}$$

The objective is, given a series of graphs signals, and optional adjacencies, to infer the dependency structure as edge weights.

LAPLACIAN LEARNING

$\boldsymbol{L}^\dagger$ encodes graph topology via $\mathbb{E}[\Sigma]$.

Edge Interpretation

- $\boldsymbol{\Omega} = \boldsymbol{L}$: Precision matrix
- $\boldsymbol{\Omega}_{i,j} = -W_{i,j}$ (partial correlations)
- $\boldsymbol{\Omega}_{i,i} = D_i$ (degree)

Conditional Independence

$$x_i \perp x_j \mid \boldsymbol{X}_{\setminus\{i,j\}} \iff \boldsymbol{L}_{i,j} = 0$$

Maximum Likelihood

Promotes large eigenvalues (smooth surrogate for rank)

Penalizes deviation from data covariance Σ

$$\max_{\boldsymbol{L}\in\mathcal{L}}\ \log\det(\boldsymbol{L}) - \mathrm{tr}(\boldsymbol{L}\boldsymbol{\Sigma})$$

$$\mathcal{L} = \{\boldsymbol{L} :\ \boldsymbol{L} \succeq 0,\ \boldsymbol{L}\boldsymbol{1} = \boldsymbol{0},\ \boldsymbol{L}_{i\neq j} \leq 0\}$$

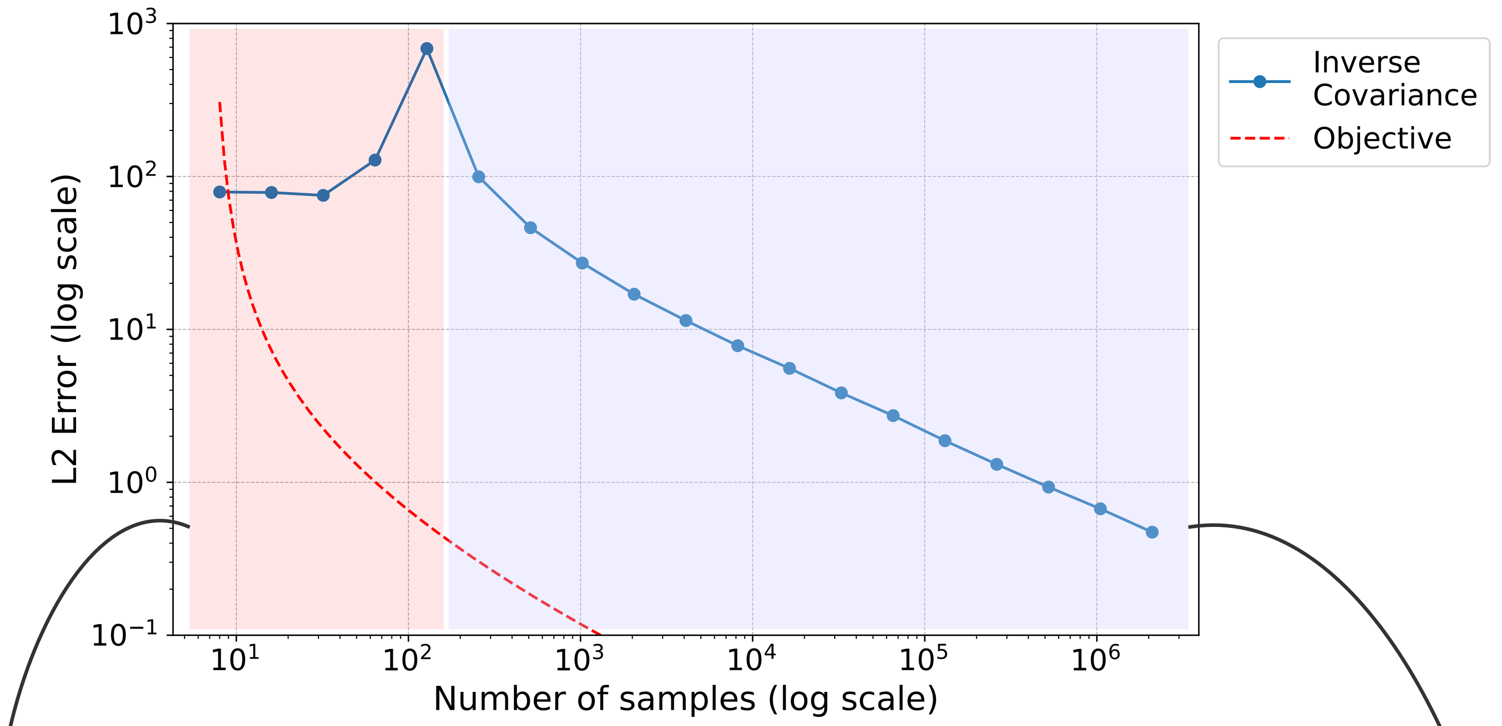
Off-diagonal entries are non-positive

Rows sum to zero (centering)

$\boldsymbol{L}$ is positive semidefinite

Note. We add the ridge term $\delta\boldsymbol{I}$ to ensure $\boldsymbol{L} + \delta\boldsymbol{I} \succ 0$.

Sampling Complexity



Spectral Relaxation

$$\hat{\boldsymbol{L}} = \arg\min_{\boldsymbol{L}\in\mathcal{L}} \left\{\mathrm{tr}(\boldsymbol{L}\boldsymbol{\Sigma}) + \alpha\|\boldsymbol{L}\|_F^2\right\}, \quad \alpha > 0$$

- Strongly convex due to Frobenius norm, while implicitly enforcing spectral properties.
- Stabilizes estimation when Σ is rank-deficient $(m < n)$.

Sparsity Penalty

$$\min_{\boldsymbol{L}\in\mathcal{L}} \left\{\mathrm{tr}(\boldsymbol{L}\boldsymbol{\Sigma}) + \alpha\psi(\boldsymbol{L}) + \beta\sum_{i\neq j}\rho(\boldsymbol{L}_{ij})\right\}, \quad \alpha, \beta > 0$$

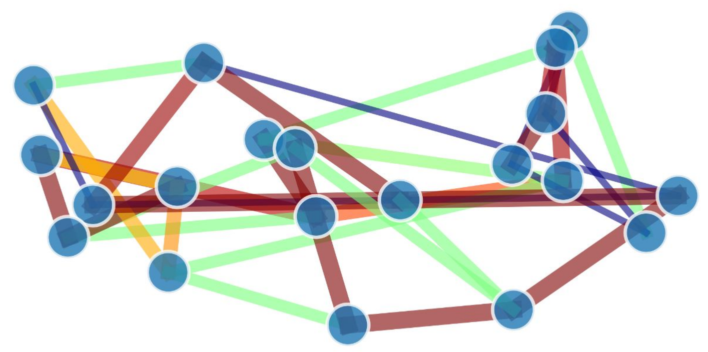
- The sparse penalty function $\rho(\cdot)$ (e.g. ℓ_1, SCAD, MCP or LCP) prevents edge weight shrinkage.

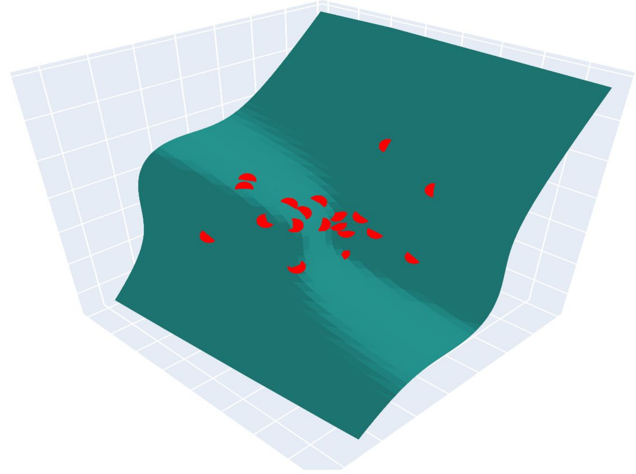
MANIFOLD STRUCTURE

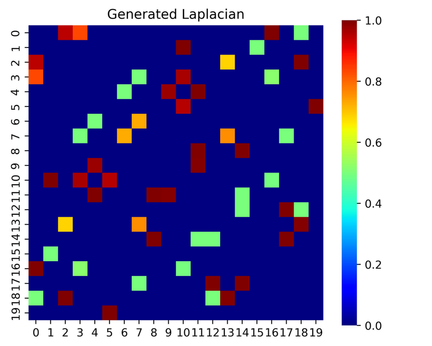
- As $n \to \infty$ and bandwidth $\epsilon \to 0$, $\boldsymbol{L} \to \Delta$ (Laplace-Beltrami) under uniform sampling.
- Dirichlet energy, whose minimization solves $\Delta f = 0$:

$$\mathcal{E}(f) = \int_{\mathcal{M}} \|\nabla f\|^2 \leftrightarrow \boldsymbol{x}^\top\boldsymbol{L}\boldsymbol{x}$$

- The Laplacian eigenvalues and eigenvectors provide a Fourier-like basis for graph signals, enabling frequency analysis.







The task shifts from learning a graph Laplacian that reflects data covariance, to identifying low-dimensional manifolds that preserve local neighborhoods.

- Diffusion maps learn a set of orthonormal basis functions—the eigenvectors of the diffusion operator H_t:

$$\frac{\partial \boldsymbol{f}}{\partial t} = -\boldsymbol{L}\boldsymbol{f} \iff \boldsymbol{f}(t) = \underbrace{e^{-t\boldsymbol{L}}}_{H_t}\boldsymbol{f}_0$$

FUTURE WORK

Two nodes with identical time series patterns but shifted in time, or with different overall magnitudes, will yield a larger Euclidian distance (hence smaller similarity). The question is: are these *meaningful*?

- Sampling rate: If node signals are sampled at different rates, it can artificially increase or decrease distance, since higher sampling rates in time series increase autocorrelation.
- Non-linear dependencies: For example, (constructive, or destructive) interference can lead to scale gaps that may not reflect actual dependencies.

Thus, the need to move to kernels that can capture non-linear dependencies:

- Can we generalize diffusion-based approaches to distortion-based ones?
- Can we go beyond i.i.d. time series?
- Can we have theoretical guarantees for nonlinear dynamics?

Application to Infrastructures

- Water flow in water distribution systems (WDSs)
- Water quality and propagation in sensor networks (over WDSs)
- Power grid stability