LEARNING TIME VARYING GRAPHS

Vassilis Kalofolias, Andreas Loukas, Dorina Thanou, Pascal Frossard

Signal Processing Laboratory (LTS2/LTS4),
École Polytechnique Fédérale de Lausanne

Abstract

We consider the problem of inferring the hidden structure of high-dimensional time-varying data. We aim at capturing the dynamic relationships by representing data as valued nodes in a sequence of graphs. Our approach is motivated by the observation that imposing a meaningful graph topology can help solving the generally ill-posed and challenging problem of structure inference.

We introduce a new prior that asserts that the graph edges change smoothly in time. We propose a primal-dual optimization algorithm that scales linearly with the number of allowed edges and can be easily parallelized. Our new algorithm is shown to outperform standard graph learning and other baseline methods both on a synthetic and a real dataset.

Background

Smooth signals on a graph. Data matrix $X \in \mathbb{R}^{N\times T}$, containing columns x_k as time samples of graph signals. Smoothness quantified by the *Dirichlet energy*

$$\frac{1}{2}\sum_{k,i,j} W_{ij}\|X_{i,k} - X_{j,k}\|^2$$

where $W_{ij} \in \mathbb{R}_+$ is the weight of the edge (i,j).

Learning a graph from smooth signals. Solve the general problem

$$\min_{W\in\mathcal{W}} \|W \circ Z\|_{1,1} + f(W),$$

where $Z_{ij} = \|x_i - x_j\|^2$ and $\mathcal{W}$ denotes the set of valid adjacency matrices (positive and symmetric).

State-of-the-art methods assume different models for $f(W)$:

- $f(W) = -\alpha\mathbf{1}^\top \log(W\mathbf{1}) + \frac{\alpha}{2}\|W\|_F^2$ [Kalofolias]
- $f(W) = \alpha\|W\mathbf{1}\|^2 + \alpha\|W\|_F^2 + \mathbb{1}\{\|W\|_{1,1} = N\}$, [Hu, Dong]

Why graph learning?

Unless the number of samples is large, learning a graph "explains" the data structure (variance) better.

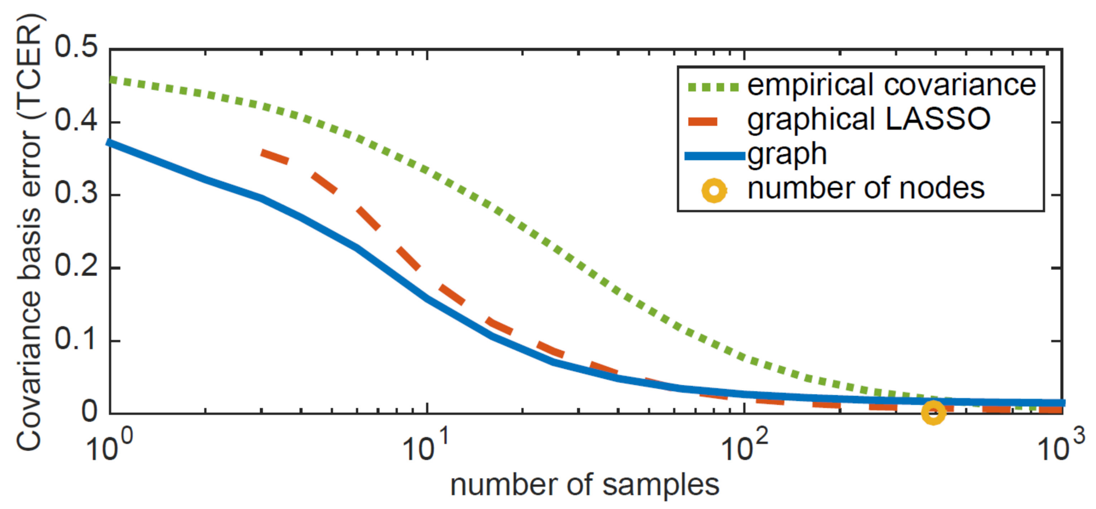


Figure 1. *Pixel-wise structure of the MNIST image dataset obtained by a) graph learning [Kalofolias], b) empirical covariance estimation, and c) the more computationally expensive sparse inverse covariance estimator also known as graphical LASSO[Banerjee]. The vertical axis intuitively quantifies how efficiently the learned structure captured the data variance, whereas the horizontal axis corresponds to the number of images used for training.*

Measuring the quality of a graph

Definition 1. ***Total cumulative energy residual (TCER):*** *Given a data distribution $p(X)$ with mean μ and covariance C, and a sorted orthogonal basis $Q = [q_1, \dots q_N]$,*

$$TCER\{p(X), Q\} = 1 - \frac{\mathbb{E}\left[\sum_{r=1}^N w_r\|q_r^\top X\|^2\right]}{\max_{Q\in\mathbb{O}}\mathbb{E}\left[\sum_{r=1}^N w_r\|q_r^\top X\|^2\right]}$$

where $w_r = N+1-r$ and the denominator is simply $\sum_{r=1}^N w_r\sigma_r^2$, for σ_r the rth singular value of $C + \mu\mu^\top$.

Learning time-varying graphs

We consider the case of a graph that changes slowly over time.

Optimization problem. Discretize time in K windows and denote by $W^{(k)}$ for $k = 1, \dots, K$ the adjacency matrix of the k-th window. Solve:

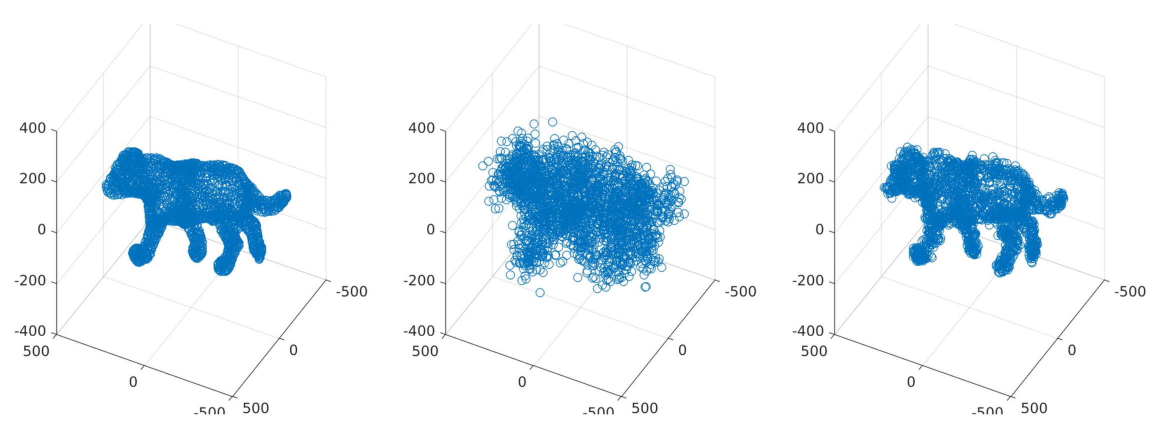
$$\min_{\{W^{(k)}\in\mathcal{W}\}} \sum_{k=1}^K \left[\left\|W^{(k)} \circ Z^{(k)}\right\|_{1,1} + f(W^{(k)})\right] + \cdots$$
$$\gamma\sum_{k=2}^K f_{\text{time}}\left(W^{(k)}, W^{(k-1)}\right).$$

- $f_{\text{time}} = \|W^{(k)} - W^{(k-1)}\|_F^2$ enforces that the graph edges change smoothly over time.
- $f_{\text{time}} = \|W^{(k)} - W^{(k-1)}\|_{1,1}$ better suited to graphs that are expected to have switching edges.

Complexity. $\mathcal{O}(N^2K)$ per iteration, and the number of iterations is typically within the hundreds.

Dynamic point-cloud de-noising

Goal is to overcome coordinate noise and point registration error, even if the geometry evolves over time.

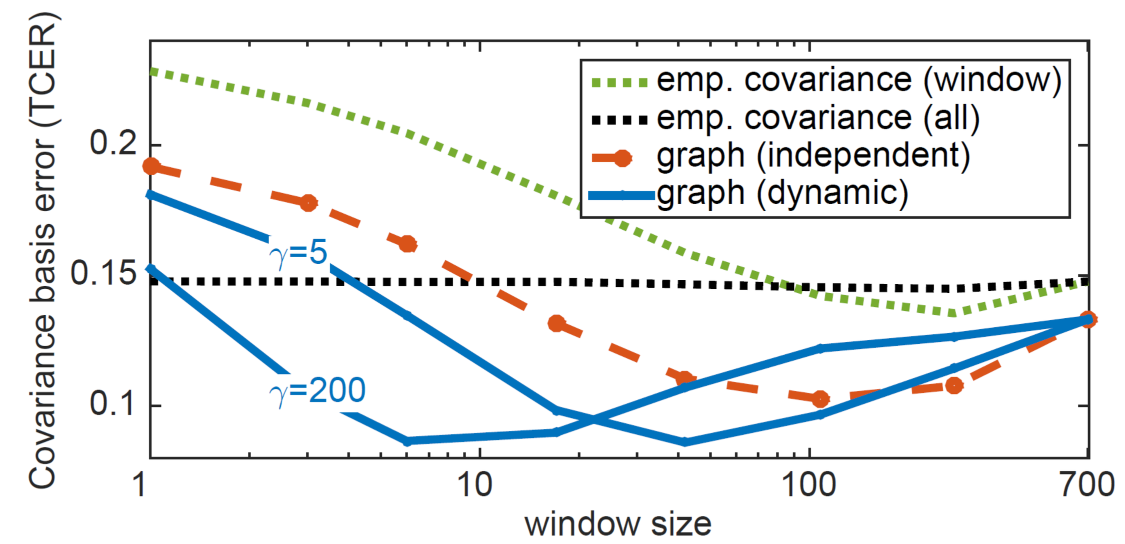


	Noisy data	1 graph (static)	12 graphs (independent)	12 graphs (dynamic)
SNR (dB)	14.03	17.58	18.77	**18.95**

Gaussian Markov chain graph estimation

Signals sampled from Gaussian Markov chain:

$$p\left(x_t | x_{t-1}, W^{(t)}\right) = \mathcal{N}\left(x_t | 0, (L^{(t)} + \sigma^2 I)^{-1}\right)\mathcal{N}\left(x_t | x_{t-1}, \frac{1}{\mu}I\right)$$



Random waypoint graph model

Figure 2. *Learning a graph from each window independently ($\gamma = 0$), performs better than the covariance matrix with fewer samples. The new prior helps when the number of samples are few, achieving the best result for $\gamma = 200$ (6 samples per window).*

Conclusions

New framework for learning dynamically changing graphs from smooth data observations:

- A time smoothness prior imposes that the graphs learnt in successive windows change smoothly over time.
- We achieve a good trade-off between temporal resolution and computation cost.
- Our experiments show that the new model can outperform classical graph learning and other baseline methods.

References

1. Vassilis Kalofolias, How to learn a graph from smooth signals, Conference on Artificial Intelligence and Statistics (AISTATS), 2016
2. Xiaowen Dong, Dorina Thanou, Pascal Frossard, and Pierre Vandergheynst, Learning Laplacian matrix in smooth graph signal representations, Trans. on Signal Processing, 2016
3. Chenhui Hu, Lin Cheng, Jorge Sepulcre, Georges El Fakhri, Yue M, Lu, and Quanzheng Li. A graph theoretical regression model for brain connectivity learning of Alzheimer's disease. Symposium on Biomedical Imaging, 2013
4. Onureena Banerjee, Laurent El Ghaoui, and Alexandre d'Aspremont, "Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data," Journal of Machine Learning Research, 2008.