Learning sparse models of diffusive graph signals

Shuyu Dong[1], Dorina Thanou[2], P.-A. Absil[1], Pascal Frossard[2] *

[1]ICTEAM Institute, Université catholique de Louvain
B-1348 Louvain-la-Neuve, Belgium

[2]Ecole polytechnique fédérale de Lausanne, LTS4
Route Cantonale, CH-1015 Lausanne, Switzerland

Abstract. Graph signals that describe data living on irregularly structured domains provide a generic representation for structured information in very diverse applications. The effective analysis and processing of such signals however necessitate good models that identify the most relevant signal components. In this paper, we propose to learn sparse representation models for graph signals that describe heat diffusion processes. This consists in learning a dictionary that incorporates spectral properties of an implicit graph diffusion kernel. The underlying formulation enables the identification of both sparse features and an adaptive graph structure from mere signal observations. Experiments on synthetic and real datasets show that the proposed dictionaries not only reflect the underlying diffusion process but also significantly reduce over-fitting of data in comparison to state-of-the-art methods.

1 Introduction

With the fast growing volume of network data in many application domains, we see an increasing amount of signals whose input space is represented by a graph that incorporates important structured information of the data [1]. Unfortunately, it is not always easy to choose good models in order to effectively process and analyze graph signals. In this paper, we are interested in learning a dictionary that carries information of a latent graph structure over the input space. The proposed framework is based on [2] and presents a novel data-driven approach for sparsely representing signals that live on a graph. The major challenge in this work is that the graph structure to be captured in the dictionary is assumed to be unknown. Recent works within a similar problem setting include attempts to learning graph Laplacian and graph shift matrices using smooth models [3, 4], spectral templates [5], graph diffusion models [6] as well as an autoregressive model [7].

In this work, we consider learning heat diffusion models in a dictionary learning framework. While dictionary learning has proven to be an efficient and adaptive approach for signal processing tasks, our work further enforces its capability by imposing a structure on the dictionary matrix in order to capture latent graph information of the network data. We focus on heat diffusion models for applications to a variety of network data that exhibit diffusive patterns.

*This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office.

Data recording the concentration of tourists or traffic jams over a network of geographical sites inside a city and measurements of pollution over a geographical sensor network are some examples. Besides efficient sparse representation, the learned dictionary also enables the graph identification and source localization tasks. More precisely, we propose to learn a structured dictionary that approximates a graph-based heat diffusion kernel of the form of $e^{-\tau L}$ [8], where L is the unknown graph Laplacian operator. Our approximate formulation results in a dictionary learning problem that is a relaxation of the graph learning problem specifically addressed in [6]. The proposed dictionary learning problem is solved by the alternating minimization procedure for sparse coding and dictionary update. The Nonnegative Orthogonal Matching Pursuit (OMP) [9] is used for solving the sparse coding step. In the dictionary update step, first-order methods in the framework of Riemannian optimization [10] are adopted to handle spectral constraints of the dictionary. Experimental results on synthetic and real datasets validates that the proposed dictionaries are efficient in representing diffusive graph signals and have proven to successfully reduce over-fitting of data compared to unstructured dictionaries such as the one learned with K-SVD [11].

2 A sparse representation model for diffusive graph signals

Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ with n vertices, equipped with a weighted adjacency matrix $W \in \mathcal{S}^n$, where $\mathcal{S}^n$ is the set of $n \times n$ real symmetric matrices. The combinatorial graph Laplacian is defined as $L = D - W$, where $D = \mathrm{Diag}(W\mathbf{1})$. The graph Laplacian is positive semidefinite thus can be decomposed as $L = U\Lambda U^\top$, where $U \in \mathcal{O}(n)$ is a real orthogonal matrix and $\Lambda = \mathrm{Diag}(\lambda_0, .., \lambda_{n-1}) \succeq 0$ is a diagonal matrix of the eigenvalues. We design a diffusion model based on the graph heat diffusion process. Given an initial source signal $x \in \mathbb{R}^n$, a diffusion process $y(t)$ on $\mathcal{G}$ is determined by

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -Ly, \quad y(t=0) = x, \tag{1}$$

which describes the flow of heat over the graph $\mathcal{G}$. The solution to the heat equation (1) is $y(t) = e^{-tL}x$, where the matrix exponential e^{-tL} is the heat diffusion kernel [8], which is also a low-pass filter on the graph [1]. We define a class of diffusive signals as snapshots of the diffusion process, that is signals observed at a fixed time τ, called a characteristic time, of the diffusion process. These settings give rise to the graph-based diffusion signal model, expressed as

$$y = e^{-\tau L}x + \epsilon, \tag{2}$$

where $x \in \mathbb{R}^n$ is the sparse source signal and $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2 I)$ is some Gaussian background noise. An important assumption in this signal model is the sparsity of x, since for many diffusion phenomena such as spreading of information over social networks, physical measurements on geographical networks and epidemic expansion, the number of sources is much smaller than the total number of vertices involved in a graph diffusion process. Moreover, we are particularly

interested in modeling x as non-negative, which is a natural assumption for diffusion sources.

3 The dictionary learning algorithm

Next, we show how we can learn the unknown variables of the model (2), *i.e.* the graph Laplacian L and the sparse codes x. Learning these latent variables by maximum-likelihood is difficult mainly because $e^{-\tau L}$ is nonlinear w.r.t. the latent variable L. To avoid this difficulty, we propose to relax this maximum-likelihood problem by considering the following dictionary learning problem. Instead of learning explicitly the latent variable L, we optimize $D \in \mathbb{R}^{n\times n}$ that approximates $e^{-\tau L}$ according to its spectral properties, as described next.

Given a set of m training signals $Y \in \mathbb{R}_+^{n\times m}$, the learning formulation (denoted GDL hereafter) is as follows:

$$\begin{aligned} \underset{X\in\mathbb{R}^{n\times m}, D\in\mathbb{R}^{n\times n}}{\text{minimize}} \quad & \|Y - DX\|_F^2 \\ \text{subject to} \quad & 0 \prec D \preceq I_n \\ & X_j \succeq 0, \|X_j\|_0 \leq T_0, \forall j = 1, .., m, \end{aligned} \tag{3}$$

where each column of X is assumed to be supported on just a few vertices due to the sparsity prior of each source signal in the model (2). The constraint $X_j \succeq 0$ captures the non-negativeness of the source signals in the diffusion model. The constraints for D are determined by spectral properties of the latent diffusion kernel $e^{-\tau L}$. It is necessarily positive definite and all its eigenvalues are contained in $]0, 1]$ since the spectrum of $e^{-\tau L} = Ue^{-\tau\Lambda}U^\top$ is $e^{-\tau\Lambda}$, whose diagonal entries are included in $]0, 1]$ as the graph Laplacian is positive semidefinite ($\Lambda \succeq 0$). Hence the search space for D, as stated in (3), is a spectral set [12] included in $\mathcal{S}_{++}^n$, which by definition is composed of matrices whose spectra belong to a subset of $\mathbb{R}_+^n$: in our case, $\mathcal{C} = \{D \in \mathbb{R}^{n\times n} : 0 \prec D \preceq I_n\} \subset \mathcal{S}_{++}^n$.

The objective function $\mathcal{F}(D, X) = \|Y - DX\|_F^2$ is non-convex w.r.t the tuple (D, X) but is convex w.r.t D and X respectively. Therefore we adopt an alternating minimization scheme between the update of X (the sparse coding step) and D (the dictionary update step) separately. For the minimization in X, we apply the nonnegative variant [9, 13] of the OMP algorithm, which is a state-of-the-art method for sparse coding. For the minimization in D, projected gradient methods are used: a retraction operator[1] [10] is designed to carry out projected gradient descents over the nonlinear search space $\mathcal{C}$. Starting from the initialization at $D_0 = I_n$, the steepest descent on $\mathcal{C}$ is achieved by a projective retraction [12]. A line-search technique is employed to optimize the step size through the Armijo condition while ensuring the constraint $D \succ 0$. During each steepest-descent iteration, the retraction (denoted $\mathcal{P}_\mathcal{C}$) acts on the line-searched candidate matrix $Q^{t+1} \in \mathcal{S}_{++}^n$ through spectral projections: based on the eigen-decomposition $Q^{t+1} = U\Lambda U^\top$ with $U \in \mathcal{O}(n)$ and $\Lambda = \text{Diag}(\lambda_0, .., \lambda_{n-1}) \succ 0$,

$$\mathcal{P}_\mathcal{C}(Q^{t+1}) = U\Pi_{[0,1]}(\Lambda)U^\top.$$

[1] A first-order approximation of the Riemannian exponential map.

The projected diagonal matrix is $\Pi_{[0,1]}(\Lambda) = \text{Diag}(\Pi_{[0,1]}(\lambda_0), .., \Pi_{[0,1]}(\lambda_{n-1}))$, where $\Pi_{[0,1]}(\lambda) = \min(\max(0, \lambda), 1)$. The retraction $\mathcal{P}_{\mathcal{C}}$ ensures steepest descent within the constrained set $\mathcal{C}$ in the dictionary update step.

4 Experiments

We quantify the performance of our algorithm on synthetic and real datasets. For all our experiments, we measure the performance of a trained dictionary D^* by the Peak Signal-to-Noise Ratio (PSNR), commonly used for evaluation of image recovery or compression tasks. The PSNR on a test dataset $Y_{\text{test}} \in \mathbb{R}^{n \times m'}$ is computed based on (D^*, X^*_{test}), where X^*_{test} is obtained through sparse coding given Y_{test} and D^*.

4.1 Synthetic data

We generate a random graph $\mathcal{G}$ of $n = 100$ vertices that simulates a sensor network using the GSP toolbox [14], with vertices in a 2-dimensional coordinated plane. Each vertex is linked to its nearest neighbors with a degree of around $5 \sim 7$. The edge weight of two connected vertices depends on the Gaussian radius basis function of their Euclidean distance, that is $W_{ij} = e^{-d(i,j)^2/\sigma^2}, \forall (i,j) \in \mathcal{E}$. The ground truth source signals x are generated on a few vertices drawn independently and uniformly from $\mathcal{V}$, with a sparsity level between $8\% \sim 20\%$, according to the signal model (2). Our experiments are based on synthetic datasets Y_τ corresponding to different values of τ, since the time-scale parameter τ determines the diffusive characteristics of data. The size of training and test data is 500 and 125 respectively. The sparsity threshold T_0 is fixed to a level that is slightly larger than $\max_j \|X_j\|_0$ of the ground-truth signals.

Based on each dataset Y_τ, dictionaries are trained using our learning algorithm and a graph-agnostic algorithm such as K-SVD [11]. The obtained results are presented in Table 1 and show that our proposed dictionary not only has comparable or even better PSNR performance on the training data compared to the unstructured dictionary but also outperforms the latter on the test data. These results confirm that the graph structure can indeed be beneficial not only for understanding the nature of the graph signals but also for reducing the overfitting caused by the limited number of training signals in the learning phase.

We have further examined the consistency (last row of Table 1) of the graph structure information captured by the learned dictionary with respect to the ground-truth graphs as follows. From the learned dictionary D^*, we identify a graph Laplacian matrix L^{valid} by projecting the matrix logarithm $-\log D^*$ onto the convex set of valid graph Laplacian matrices $\mathcal{S}_{Lap} := \{L \in \mathcal{S}^n, L_{ij} = L_{ji} \leq 0, L\mathbf{1} = 0\}$. This leads to a convex optimization problem and is solved by using the `CVX` toolbox [15]. The graph edges are recovered by thresholding L^{valid} with a sufficiently large value $\varepsilon > 0$: $L^\varepsilon_{ij} := L^{\text{valid}}_{ij} \delta_{L^{\text{valid}}_{ij} > \varepsilon}$. The graph recovery scores are then measured by the maximal F-measure according to the Precision-Recall curve of $(L^\varepsilon)_\varepsilon$ w.r.t. the ground truth. We observe that the F-measure score is

particularly high, when τ is small and tends to reduce when we increase τ. This is quite expected as a small τ leads to very localized processes on the graph, which makes the graph recovery process easier.

Dataset-τ		0.10	0.24	0.38	0.52	0.66	0.80	1.00
Training	KSVD	**100.40**	**85.34**	69.63	45.24	41.02	38.76	37.44
	GDL	75.78	72.84	**73.10**	**47.15**	**44.48**	**41.27**	**38.98**
Test	KSVD	**92.94**	69.03	59.87	42.01	37.55	35.64	32.52
	GDL	73.96	**69.50**	**70.80**	**46.01**	**41.48**	**39.23**	**35.78**
Graph Recovery	GDL	0.9197	0.8318	0.9697	0.8717	0.7805	0.8111	0.7252

Table 1: Upper part: PSNR results on training and test data using GDL and K-SVD. Last row: consistency of the recovered graphs by F-measure.

4.2 Real-world data

Our real-world data is extracted from the Alameda dataset (publicly available on `http://pems.dot.ca.gov`), which consists of daily traffic delay detected at 559 geographical points during 2772 days along several intersecting roads situated in Alameda, California. For simplicity, we consider observations on a subset of $n = 246$ detection points. Each of the 2772 samples is a vector of delay-time measurements over the n sites. The dictionaries are trained and tested on 862 and 215 samples respectively. The evaluation in terms of PSNR is presented in Table 2.

T_0	Training data		Test data	
	KSVD	GDL	KSVD	GDL
15	**41.25**	36.49	34.84	**35.00**
24	**44.97**	40.36	37.93	**38.51**
32	**48.16**	43.46	40.53	**41.56**
41	**52.24**	47.37	43.41	**45.48**
50	**56.37**	52.16	46.12	**50.38**

Table 2: Experiments on the Alameda traffic data: PSNR results on training and test data using GDL and K-SVD.

From these results, similarly to the synthetic case, we observe that our learning algorithm is much more efficient than purely numerical dictionary learning algorithms in representing test data that are not included in the training phase.

5 Conclusion

In this paper, we have presented a framework of dictionary learning for diffusive graph signals. Specifically, we have proposed an algorithm that promotes sparsity of signals in a graph diffusion dictionary. Experimental results on both

synthetic and real datasets show that the proposed dictionary is capable of capturing graph structured information of the input space in a data-driven manner. Moreover, the sparse representation model for graph signals is learned with high generalization power, which is otherwise limited in the case of purely numerical dictionary learning algorithms. Finally, the graph recovery performance shows high consistency of the recovered graphs with respect to the ground-truth ones.

References

[1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.

[2] S. Dong. Learning graph topologies for diffusion signal representation on graphs. Master's thesis, École polytechnique fédérale de Lausanne, Route Cantonale, 1015 Lausanne, Suisse; Télécom ParisTech, 46 Rue Barrault, 75013 Paris, France, 2016.

[3] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst. Laplacian Matrix Learning for Smooth Graph Signal Representation. In *Proceedings of IEEE ICASSP*, 2015.

[4] Y. Yankelevsky and M. Elad. Dual graph regularized dictionary learning. *IEEE Transactions on Signal and Information Processing over Networks*, 2(4):611–624, Dec 2016.

[5] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro. Network topology inference from spectral templates. *CoRR*, abs/1608.03008, 2016.

[6] D. Thanou, X. Dong, D. Kressner, and P. Frossard. Learning heat diffusion graphs. *IEEE Journal on Selected Topics in Signal Processing and IEEE Transactions on Signal and Information Processing over Networks, joint special issue on Graph Signal Processing*, 2016.

[7] J. Mei and J. M. F. Moura. Signal processing on graphs: Modeling (causal) relations in big data. *CoRR*, abs/1503.00173, 2015.

[8] J. Lafferty and R. I. Kondor. Diffusion Kernels on Graphs and Other Discrete Input Spaces. *ICML '02 Proceedings of the Nineteenth International Conference on Machine Learning*, pages 315—-322, 2002.

[9] A. Bruckstein, M. Elad, and M. Zibulevsky. On the Uniqueness of Nonnegative Sparse Solutions to Underdetermined Systems of Equations. *IEEE Transactions on Information Theory*, 54(11):4813–4820, 2008.

[10] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. 2008.

[11] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.

[12] P.-A. Absil and J. Malick. Projection-like retractions on matrix manifolds. *SIAM Journal on Optimization*, 22(1):22(1):135—-158, 2012.

[13] M. Yaghoobi Vaighan, D. Wu, and M. Davies. Fast non-negative orthogonal matching pursuit. *IEEE Signal Processing Letters*, 2015.

[14] N. Perraudin, J. Paratte, D. Shuman, L. Martin, V. Kalofolias, P. Vandergheynst, and D. K. Hammond. GSPBOX: A toolbox for signal processing on graphs. *ArXiv e-prints*, August 2014.

[15] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1, March 2014.